

https://dspace.org/

Deep learning models for defect detection in electroluminescence images of solar PV modules

Pratt, Lawrence

2024-05-29
University of the Witwatersrand
https://hdl.handle.net/10539/42163
Pratt, Lawrence. (2024). Deep Learning Models for Defect Detection in Electroluminescence Images of Solar PV Modules [PhD thesis, University of the Witwatersrand, Johannesburg]. WireDSpace.https://hdl.handle.net/10539/42163

Downloaded from a DSpace repository. https://dspace.org/

SOLAR CELL DEFECT DETECTION

Deep Learning Models for Defect Detection in Electroluminescence Images of Solar PV Modules

School of Computer Science & Applied Mathematics
University of the Witwatersrand

Lawrence Pratt
2300722

Supervised by
Professor Richard Klein

May 29, 2024



A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the Doctor of Philosophy in Computer Science

Abstract

This thesis introduces multi-class solar cell defect detection (SCDD) in electroluminescence (EL) images of PV modules using semantic segmentation. The research is based on experimental results from training and testing existing deep-learning models on a novel dataset developed specifically for this thesis. The dataset consists of EL images and corresponding segmentation masks for defect detection and quantification in EL images of solar PV cells from monocrystalline and multi-crystalline silicon wafer-based modules. While many papers have already been published on defect detection and classification in EL images, semantic segmentation is new to this field. The prior art was largely focused on methods to improve EL image quality, classify cells into normal or defective categories, statistical methods and machine learning models for classification, object detection, and some binary segmentation of cracks specifically. This research shows that multi-class semantic segmentation models have the potential to provide accurate defect detection and quantification in both high-quality lab-based EL images and lower-quality field-based EL images of PV modules. While most EL images are collected in factory and lab settings, advancements in imaging technology will lead to an increasing number of EL images taken in the field. Thus, effective methods for SCDD must be robust to various images taken in the labs and the real world, in the same way that deep-learning models for autonomous vehicles that navigate the city streets in some parts of the world today must be robust to real-world environments. The semantic segmentation of EL images, as opposed to image classification, yields statistical data that can then be correlated to the power output for large batches of PV modules.

This research evaluates the effectiveness of semantic segmentation to provide a quantitative analysis of PV module quality based on qualitative EL images. The raw EL image is translated into tabular datasets for further downstream analysis. First, we developed a dataset that included 29 classes in the ground truth masks in which each pixel was coloured according to the class. The classes were grouped into intrinsic "features" of solar cells and extrinsic "defects." Next, a fully-supervised U-Net trained on the small dataset showed that SCDD using semantic segmentation was a viable approach. Next, additional fully-supervised deep-learning models (U-Net, PSPNet, DeepLabV3, DeepLabV3+) were trained using equal, inverse, and custom class weights to identify the best model for SCDD. A benchmark dataset was published along with benchmark performance metrics. The model performance was measured using mean recall, mean precision, and the mean intersection over union (mIoU) for a subset of the most common defects (cracks, inactive areas, and gridline defects) and features (ribbon interconnects and cell spacing) in the dataset. This work focused on developing a deep-learning method for SCDD independent of the imaging equipment, PV module design, and image quality that would be broadly applicable to EL images from any source.

The initial experiment showed that semantic segmentation was a viable method for SCDD. The U-Net trained on the initial dataset with 108 images in the training dataset produced good representations of the features common to most of the cells and good representations of the defects with a reasonable sample size. Other defects with only a few examples in the training dataset were not effectively detected in this model. The U-Net results also showed that the mIoU measured higher for the features compared to the defects across all models, which correlated with the size of the large features compared to the small defects that each class occupies in the images.

The next set of experiments showed that the DeepLabv3+ trained with custom class weights scored the highest in terms of mIoU for the selected defects and features when compared to the alternative fully-supervised models. While the mIoU for cracks was still low (25%), the recall was high (86%). While increasing the recall substantially, the many long, narrow defects (e.g. cracks and gridlines) and features (e.g. ribbon interconnects and spacing) in the dataset were challenging to segment, especially at the borders. The custom class weights also tended to dilate the long, narrow features, which led to low precision. However, the resulting representations reliably located these defects in the complex images with both large and small objects, and the dilation proved effective at visually highlighting the long-narrow defects when the cell-level images were combined into module-level images. Therefore, the model proved useful in the context of detecting critical defects and quantifying the relative size of the defects in EL images of PV cells and modules despite the relatively low mIoU. The dataset was also published along with this paper.

The final set of experiments focused on semi-supervised and self-supervised models. The results suggested that supervised training on a large out-of-domain (OOD) dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and semi-supervised pretraining (CCT) were statistically equivalent as measured by the mIoU on a subset of critical defects and features. A new state-of-the-art (SOTA) for SCDD was achieved, exceeding the mIoU from the DeeplabV3+ with custom weights. The experiments also demonstrated that certain pretraining schemes resulted in the ability to detect and quantify underrepresented classes, such as the round ring defect.

The unique contributions from this work include two benchmark datasets for multi-class semantic segmentation in EL images of solar PV cells. The smaller dataset consists of 765 images with corresponding ground truth masks. The larger dataset consists of more than 20,000 unlabelled EL images. The thesis also documents the performance metrics from various deep-learning models based on fully-supervised, semi-supervised, and self-supervised architectures.

Publications

Portions of this work have been published or are under review:

[Pratt et al. 2021] Lawrence Pratt, Devashen Govender, and Richard Klein. Defect detection and quantification in electroluminescence images of solar PV modules using U-Net semantic segmentation. Renewable Energy, 178:1211–1222, 2021.

[Pratt et al. 2023] Lawrence Pratt, Jana Mattheus, and Richard Klein. A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation. Systems and Soft Computing, page 200048, 2023.

[Torpey et al.] David Torpey, Lawrence Pratt, and Richard Klein. A Large-scale Evaluation of Pretraining Paradigms for the Detection of Defects in Electroluminescence Solar Cell Images. Under review.

Declaration

I, Lawrence Pratt, hereby declare the contents of this research proposal to be my own work. This thesis is submitted for the degree of Doctor of Philosophy in Computer Science at the University of the Witwatersrand. This work has not been submitted to any other university or for any other degree.



Lawrence Pratt

May 29, 2024

Acknowledgements

I wish to thank Dr. Richard Klein for his expert guidance and support as my PhD advisor.

I wish to thank Dr. Kittessa Roro for encouraging and supporting the PhD as my Research Group Leader at the Council for Scientific and Industrial Research (CSIR).

I wish to thank the CSIR management for technical support, measurement equipment, test samples, and financial support for the PhD.

I wish to thank CFV Labs and ARTsolar for providing EL images and the team of labellers who annotated them: Nandi Bau, Sibusiso Mgidi, Rifumo Mzimba, Siyathandana Nontolwana, Kian Reddy, and Kyle Wootton from the University of the Witwatersrand. A special note of thanks goes out to Keketso Moletsane, who worked tirelessly to clean up errors in the historical dataset and to expand the dataset with images of new defects.

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number: 118075).

The authors also acknowledge the Centre for High-Performance Computing (CHPC), South Africa, for providing computational resources to this research project.

Motivation

The nexus between renewable energy and computer science inspired this thesis. In 2000, I earned an MSc. in Applied Statistics from the University of New Mexico in the United States while working as a graduate student at the Los Alamos National Labs, the birthplace of the nuclear bomb. I gained a strong appreciation for the importance of data-driven decisions and computing in scientific and engineering endeavours that have the power to change lives all across the world. My passion for data analysis and visualisation was born. As someone once wrote in an article about statistical science, "You don't have to be 'aweful' to be a good statistician, but it helps." The spelling of 'aweful' was intentional and highlighted that computers and statistical methods could extract useful information from large datasets in awe-inspiring ways. I developed those passions as a Yield Engineer at the Intel Corporation, another data-driven organisation developing and producing microprocessors for the computers and smartphones that have changed lives in every corner of the world. In 2006, I joined a start-up solar PV module company across the Rio Grande River from Intel Corporation in Albuquerque, New Mexico. There, my passion for renewable energy was born. In 2017, I joined the Council for Scientific and Industrial Research (CSIR) in Tshwane, South Africa. I was recruited to establish the first PV module quality and reliability lab in Southern Africa. At the encouragement of CSIR management and policy, I contacted the University of the Witwatersrand with some ideas for a Ph.D. thesis. Fortunately, I found my way to the office of Dr. Richard Klein, where this fantastic opportunity to explore deep-learning applications in the solar PV industry was born. As much as I have been an 'aweful' statistician throughout my career, I now hope to double down as an 'aweful' computer scientist inspired by the field of deep learning in computer vision with application in the renewable energy industry.

Contents

Preface **i**
Abstract . . . ii
Publications . . . iv
Declaration . . . v
Acknowledgements . . . vi
Motivation . . . vii
Table of Contents . . . viii
List of Figures . . . xi
List of Tables . . . xvi

1 Introduction **1**
1.1 Executive Summary . . . 1
1.2 Contributions . . . 3
1.3 Thesis structure . . . 5

2 Solar Energy and PV module Characterisation **6**
2.1 Renewable Energy Industry . . . 6
2.2 Solar Energy . . . 8
2.3 Electroluminescence imaging . . . 10
2.4 EL imaging in the lab or factory . . . 15
2.5 EL and IR imaging in the field . . . 16
2.6 Pre-processing - cropping the cells from EL images of full modules . . . 18
2.7 Pre-processing - image quality . . . 20
2.8 PV module performance measurements . . . 21
2.9 Correlating EL defects to PV module power . . . 22
2.10 Summary . . . 24

3 Related Work **25**
3.1 SCDD on wafers and bare cells . . . 25
3.2 SCDD - image classification . . . 28
3.3 SCDD - object detection and localisation . . . 32
3.4 SCDD - binary segmention . . . 33
3.5 SCDD multiclass segmentation using Semantic Segmentation . . . 40
3.6 Semi-Supervised Learning and Self-Supervised Learning . . . 42
3.7 Vision Transformers . . . 44

3.8 Conclusions . . . 45

4 General Methods **46**
4.1 Deep-learning models evaluated for SemSeg . . . 46
4.2 Loss Function . . . 49
4.3 Class weights . . . 50
4.4 Statistical design of experiments for validation of class weights . . . 51
4.5 Performance metrics for models . . . 52
4.6 Statistical methods for model comparisons . . . 55
4.7 Linear Regression . . . 59

5 Datasets for SCDD **61**
5.1 Public datasets . . . 61
5.2 New labelled dataset for SCDD . . . 66
5.3 New unlabelled dataset for SCDD . . . 76
5.4 Comparison of SCDD vs UCF datasets . . . 77
5.5 Conclusions . . . 79

6 Proof of concept using Semantic Segmentation for SCDD **80**
6.1 Introduction . . . 80
6.2 Dataset . . . 81
6.3 Method . . . 81
6.4 Results . . . 82
6.5 Conclusions . . . 86
6.6 Next steps . . . 86

7 Fully supervised deep-learning models **87**
7.1 Introduction . . . 87
7.2 Dataset . . . 87
7.3 Method . . . 87
7.4 Results . . . 90
7.5 Conclusion . . . 97
7.6 Next steps . . . 98

8 Semi- and Self-supervised deep-learning models **99**
8.1 Introduction . . . 99
8.2 Dataset . . . 100
8.3 Method . . . 100
8.4 Experiments . . . 102
8.4.1 Experimental Setup . . . 102
8.4.2 Dataset . . . 102
8.5 Results . . . 104
8.5.1 Defect Analysis . . . 105
8.5.2 OOD Analysis . . . 107
8.6 Conclusion . . . 109
8.7 Next steps . . . 109

9 Correlating EL defects and PV module performance **110**
9.1 Introduction . . . 110
9.2 Method . . . 111
9.2.1 Experimental Setup . . . 111
9.2.2 Dataset . . . 112
9.3 Results . . . 113
9.3.1 Sensitivity Analysis . . . 114
9.4 Conclusion . . . 116
9.5 Next steps . . . 118

10 Conclusion **119**
10.1 Summary . . . 119
10.2 Future Work . . . 121

A Appendix **123**
A.1 Library of features and defects . . . 123
A.1.1 Defects . . . 123
A.1.2 Features . . . 129
A.2 Colour coding scheme for ground truth masks and predictions . . . 134

List of Figures

1.1 Web-based tool for SCDD under development at the CSIR which translates raw qualitative data contained in the EL image (left) to a prediction mask (centre) to a structured qualitative dataset for post-prediction analysis and correlation studies (right) . . . 2

2.1 Energy flow diagram for South Africa in 2017 moving from primary energy supply (left) to energy conversion (middle) to end users (right) [Dlamini *et al.* 2022] . . . 7
2.2 Optical (a) and EL image (b) of a PV module made from mono-crystalline cells with severely cracked solar cells . . . 11
2.3 In-situ images of a 4 x 16 array of PV modules taken with an optical (top), infrared (middle), and EL (bottom) camera [Buerhop *et al.* 2022] . . . 12
2.4 EL image of a single cell exhibiting three types of cracks: line cracks (Type A), partially disconnected regions (Type B), and cracks creating electrically disconnected regions (Type C) . . . 12
2.5 Example of an I-V and EL characterisation station typical of a PV module manufacturer . . . 15
2.6 Test equipment at the CSIR PQRL for (a) recording EL imaging and (b) measuring module power inside the sun simulator [halm. 2024] . . . 16
2.7 Example of a hotspot cell (red square) and the thermal signature for the junction boxes on each module (green circle) in an IR image taken at the CSIR . . . 18
2.8 Examplees of (a) mono-c-si solar cells with rounded corners and (b) multi-c-si solar cells with 90°corners sampled from the dataset developed for this research. Cells with and without defects are included. . . . 19
2.9 Example I-V and power curve created from the IV pairs recorded for a PV module measured at the CSIR . . . 21

3.1 (a) mono-c-si wafer (top) and optical image of a finished solar cell (bottom) with rounded corners, (b) multi-c-si wafer (top) and optical image of a finished solar cell (bottom) with square corners, and (c) PV module assembled from multi-c-si solar cells [Fisher *et al.* 2012] . . . 26
3.2 Examples from Qian *et al.* [2018] showing the original images (a), the crack detection results (b), and the ground truth (c) . . . 27
3.3 EL images and predictions for (a,b) Type A cracks, (c,d) Type B cracks, and (e,f) gridline defects [Tsai *et al.* 2012] . . . 34

3.4 Image segmentation results comparing the proposed and standard segmentation techniques. (a) Original image, (b) Otsu's thresholding, (c) Sobel edge detector, (d) Canny's hysteresis, (e) LoG filter, (f) FIR, (g) the proposed method, and (h) ground truth images. [Anwar and Abdullah 2014] 35
3.5 Example of micro-crack detection and identification in a standard 60-cell mc-Si PV module: red - dendritic or cracks with multiple orientations; blue - cracks parallel to the busbars; magenta-parallel branches in a dendritic crack; yellow - other types of cracks [Spataru *et al.* 2016a] 36
3.6 Crack predictions for six EL images and all the binary segmentation methods evaluated [Chen *et al.* 2019] 37
3.7 EL image of a solar cells with a crack and belt mark defect taken a three different time exposures and the final binary prediction that only poorly captures the crack [Dhimish and Mather 2019] 38
3.8 Input image, intermediate heat maps, and predictions of two selected images with cracks [Mayr *et al.* 2019] 39
3.9 Predictions for four selected images from several methods evaluated for binary segmentation, including the MAU-Net [Rahman and Chen 2020] 40

4.1 Predictions for one cell when trained using a DeepLabv3+ architecture with (a) class weights for gridlines = 5 and ribbons = 11 (Model 220180-10) and (b) gridlines = 82 and ribbons = 1.5 (Model 220218-12) 51
4.2 Mean IoU for cracks from the 16-run DOE for class weights, including the custom class weights, shows as an asterisk. 52
4.3 One example showing (a) the original EL image, the ground truth mask, and the prediction, and (b) the confusion matrices for precision and recall 53
4.4 Crack layer for CFVS-00035-r8-c4 showing the prediction for cracks from the U-Net model (green) and the underlying ground truth mask (white). The IoU metric provides a numerical estimate of the overlap between the two. 54
4.5 95% confidence intervals for cIoU on crack defects by regime 57
4.6 IoU for cracks by model over five training runs for six selected images (A-F) from the test dataset without block-centring (a) and with block-centring (b) . . 57
4.7 CIs from Tukey's HSD method comparing model performance for crack detection using cIoU (left) and $IoU_{c,bc}$ (right) 58
4.8 CIs from Tukey's HSD method comparing model performance using $mIoU$, $wIoU$, $mIoU_{bc}$, and $wIoU_{bc}$ 59
4.9 Linear regression model of module FF versus average of cell-level brightening defect predicted by the model 60

5.1 Sample images of solar panels contained in the Google Open Image database . 62
5.2 Sample EL images from the ELPV benchmark dataset: (a) multi-ci-si with two ribbon interconnections and a long crack, (b) multi-ci-si with three ribbon interconnections and gridline defects, (c) mono-ci-si with three ribbon interconnections, cracks, and inactive areas, and (d) mono-ci-si with three ribbon interconnections and no defects [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2021] 62

5.3 Sample EL images of multi-c-si cells with four ribbon interconnection in the SDLE benchmark dataset: (a) no defects, (b) long, narrow crack, and (c) corroded interconnection ribbons [French *et al.* 2020] 63
5.4 Sample EL images of thin films modules in the PEARL benchmark dataset: (a) droplets and (b) shunts [Sovetkin *et al.* 2020] 63
5.5 Sample EL images from the RAPTOR dataset: (a) module with relatively cool regions at the top and bottom , (b) module with two hotspots along the bottom region, (c) module with a cooler region along the right edge, and (d) module with a cooler region along the top [Millendorf *et al.* 2020] 64
5.6 Sample EL images of multi-c-si cells with four ribbon interconnection in the PVEL-AD benchmark dataset and the corresponding bounding boxes [su binyi 2022] 65
5.7 Sample EL images from the UCF dataset: (a) mono-c-si cell with five ribbon interconnects and severe cracks, (b) mono-c-si cell with four ribbon interconnects and one heavy crack, (c) multi-c-si cell with three ribbon interconnects, heavy cracks, and inactive areas, and (d) mono-c-si cell with two ribbon interconnects and faint gridline defects [Fioresi *et al.* 2022] 65
5.8 Example from the UCF dataset showing (a) the EL image and (b) the ground truth mask with one large conglomeration [Fioresi *et al.* 2022] 66
5.9 Examples of cell level EL images (top) and corresponding ground truth masks (bottom), one from each source: (a) ARTsolar multi-si cell cropped from the edge of a full module, (b) ELPV single multi-si cell with padding on four sides, (c) CFV Labs mono-si cell cropped from the centre of a full module, Dd) CSIR mono-si cell cropped from the center of a full module, and (e) SDLE single multi-si cell with padding on four sides. 67
5.10 Number of images containing each class in the Feature group (left) and Defect group (right) 68
5.11 Profiles created from the row-wise and column-wise average intensities with peaks corresponding to cell spacing regions used to crop cells from the full module image. 69
5.12 Examples of cell images cropped from module-level EL images, including the surrounding cells and white space padding for edge cells and corner cells . . . 70
5.13 (a) Mono-c-si cell with rounded corners and (b) multi-c-si cell with square corners 70
5.14 Examples of cell images cropped from the left edge of a module image built from (a) full-sized, square cells and (b) half-cut cells, including the white space padding to maintain a square dimension without losing the aspect ratio of the cell features and defects 71
5.15 Examples of single-cell images not cropped from module-level EL images including the white space padding around all four edges 72
5.16 A random sample of ground truth masks from the SCDD dataset 73
5.17 Precision and recall distributions on a set of calibration images created to identify and reduce inter-rater variability 74
5.18 Crack layer for CFVS-00035-r8-c4 showing the prediction from the U-Net model (green) and the underlying ground truth mask manually coloured (white). 75

5.19 Example showing the pre-processing and augmentation of one original ground truth mask: a) original GT, b) resized GT, c) coded, d) flipped horizontal, e) flipped vertical, and f) rotated 180 deg to produce four ground truth masks from each original . . . 76
5.20 EL images and ground truth masks from (a) the UCF dataset [Fioresi *et al.* 2022] and (b) the SCDD dataset [Pratt *et al.* 2023] . . . 78
5.21 (a) EL image from the UCF dataset, (b) the corresponding ground truth mask from UCF, and (c) the ground truth mask as annotated by the method used for SCDD dataset . . . 79

6.1 Example of semantic segmentation for (a) an EL image of a solar cell, (b) the ground truth annotated by a human, and (c) the prediction from a model [Pratt *et al.* 2021a] . . . 81
6.2 Examples of cell level EL images (top) and corresponding predictions (bottom) from a U-Net model trained on 108 images . . . 82
6.3 Distributions of model performance metrics on 30 test images using the U-Net model trained on a small dataset . . . 83
6.4 EL images of a single module taken at several stages of an accelerated stress test sequence showing initial cell cracks after SMLT and gradual darkening of damaged areas with subsequent tests. . . . 84
6.5 Impact of accelerated stress testing as seen in the original EL images (top) and prediction masks from the U-Net model (bottom) for cell r4-c4 at (a) initial, (b) post SMLT, (c) post HF10, and (d) post HF20 . . . 84
6.6 Least squares means for the cracks (left) and inactive areas (right) . . . 85

7.1 Graphical summary of the method . . . 88
7.2 Trends of median IoU and loss during training for (a) U-Net-12, (b) U-Net-25, (c) PSPNet, and (d) DeepLabv3+. Each subplot overlays the trends for [a] equal class weights, [b] inverse class weights, and [c] custom class weights . . . 90
7.3 Prediction masks for cell CFVS 00035-r8-c4 from 12 models: U-Net-12 (row 1), U-Net-25 (row 2), PSPNet (row 3), DeepLabv3+ (row 4), equal class weights (column a), inverse class weights (column b), and custom class weights (column c) . . . 91
7.4 IoU and mIoU for mono-crystalline (blue circles), multi-crystalline (red) circles), and mIoU (solid lines) for selected defects and features in 50 test images across 12 models . . . 93
7.5 Two selected images and corresponding predictions from the DeepLabV3+ with custom weights (Model 4c), (a) one multi-c-si cell with some grain boundaries misclassified as cracks (white pixels) in the bottom left corner and (b) one mono-c-si cell with no similar errors in crack predictions . . . 94
7.6 (a) Median IoU, (b) median recall, and (c) median precision for five classes across 12 models and the average for selected features and defects . . . 94
7.7 EL, ground truth mask, and prediction mask from DeepLabv3+ with custom class weights (4c) for CFVS-00035-r8-c4 highlighting the dilated cracks and gridline defects in the prediction . . . 96

7.8 Bottom left corner of ARTS-00020-r7-c3 showing the labelling errors along the edges of ribbons and cell spacing in the ground truth image leading to errors in precision and recall . . . 96
7.9 EL, ground truth, and prediction mask from DeepLabv3+ with custom class weights (4c) for SDLE-00513 showing accurate predictions for the corrosion defect (dark green) around the ribbon interconnects (bright green) . . . 97

8.1 Number of annotated images containing each defect class (top) and feature class (bottom) . . . 103
8.2 (a) mIoU and (b) wIoU by class type (extrinsic defect and intrinsic feature) . . 105
8.3 Block-centred IoU average and 95% confidence intervals for crack detection by regime showing R2 and R4 perform statistically significantly better on crack detection compared to other models . . . 106
8.4 EL image, ground truth, and predictions for CSIR_00067_r3_c2 with ring defect and dogbone spacing . . . 106
8.5 Tukey's HSD for (a) ring defect showing R4 is statistically significantly better than all other models and (b) dogbone spacing showing 7 out of 10 models are statistically equivalent . . . 107

9.1 Methodology used for the regression analysis of I-V characteristics on EL defects 111
9.2 Module 19040-12 (a) EL image before PID stress at 100% Isc, (b) EL image after PID stress at 100% Isc, (c) EL image after PID stress at 10% Isc and brightened for visual evaluation, (d) EL image after PID stress at 10% Isc as input to the prediction model, and (e) the prediction mask for the EL image after PID stress at 10% Isc . . . 113
9.3 Module 19058-03 (a) EL image before TC stress, (b) prediction mask before TC stress, (c) EL image after TC600, and (d) EL prediction mask after TC600 . 114
9.4 Regression analysis between power loss and percentage of dark cells predicted by SCDD . . . 114
9.5 Regression analysis between power loss and percentage of crack cells predicted by SCDD . . . 115
9.6 Percentage of dark cell pixels pre- and post- PID at different levels of bias current 115
9.7 SASC-00145 EL images post PID (a) original, (b) low contrast, (c) high contrast, (d) low brightness, and (e) high brightness . . . 117
9.8 SASC-00145 predictions post PID and the average of the percentage of dark cells (a) original - 1.5%, (b) low contrast - 0.4%, (c) high contrast - 2.6%, (d) low brightness - 5.9%, and (e) high brightness - < 0.1% . . . 117
9.9 SASC-00172 EL images post TC600 (a) original, (b) low contrast, (c) high contrast, (d) low brightness, and (e) high brightness . . . 117
9.10 SASC-00172 predictions post TC600 (a) original - 0.48%, (b) low contrast - 0.45%, (c) high contrast - 0.48%, (d) low brightness - 0.48%, and (e) high brightness - 0.52% . . . 118

List of Tables

1.1 Contributions from each author towards Paper 1 [Pratt *et al.* 2021a], Paper 2 [Pratt *et al.* 2023], and Paper 3 [Torpey *et al.* 2024] by Lawrence Pratt (LP), Richard Klein (RK), Devashen Govender (DG), Jana Mattheus (JM), and David Torpey (DT) 4

2.1 Examples of extrinsic defects with descriptions, EL image, and ground truth (GT) from the benchmark dataset developed for this research. 13
2.2 Examples of intrinsic features with descriptions, EL image, and ground truth (GT) from the benchmark dataset developed for this research. 14

4.1 Class weights for a selected set of classes. 50

5.1 Summary of publicly available datasets with EL images 66

7.1 Model ID and descriptions 89

8.1 Regime descriptions evaluated for SCDD using SemSeg. 100
8.2 Results on the EL dataset for the SCDD using semantic segmentation 104
8.3 Results over all models by whether an OOD dataset was used for pretraining. . 108
8.4 Results for OOD and in-distribution pretraining on the Cityscapes and EL datasets (as per R6). 108

9.1 Sample records of the dataset used for the correlation analysis 112

A.1 Description of defects 123
A.2 Library of defects 127
A.3 Description of Features 129
A.4 Library of features 132
A.5 CSV Data Table 134

Chapter 1

Introduction

Chapter 1 provides an executive summary of the thesis, the motivation for exploring semantic segmentation of electroluminescence (EL) images of PV modules, contributions from co-authors to related published articles, and the chapter layout.

1.1 Executive Summary

This thesis introduces multi-class solar cell defect detection (SCDD) in electroluminescence (EL) images of solar photovoltaic (PV) modules using semantic segmentation (SemSeg). The research is based on experimental results from training and testing existing deep-learning models on a novel dataset explicitly developed for this thesis. The dataset consists of EL images and corresponding segmentation masks for defect detection and quantification in EL images of solar PV cells from a broad sample of mono-crystalline and multi-crystalline silicon wafer-based modules. While many papers have already been published on defect detection and classification in EL images, semantic segmentation is new to this domain. The prior art was primarily focused on methods to improve EL image quality, classify cells into normal or defective categories, statistical methods and machine learning models for classification, object detection, and some binary segmentation of cracks specifically. This research shows that multi-class semantic segmentation models have the potential to provide accurate defect detection and quantification in both high-quality lab-based EL images and lower-quality field-based EL images of PV modules. While most EL images are collected in factory and lab settings, advancements in imaging technology will lead to an increasing number of EL images taken in the field. Thus, practical methods for SCDD must be robust to various images taken in the different labs and the real world, in the same way, deep-learning models for autonomous vehicles that navigate the city streets in some parts of the world today must be robust to real-world environments.

Semantic segmentation (SemSeg) is applied to SCDD and published in the scientific literature for the first time. SemSeg has several potential advantages over the classification and bounding-box approaches. Specifically, SemSeg can detect, locate, and quantify the size of multiple classes on a single image by counting the number of pixels associated with each class. This leads to increased resolution compared to image classification or bounding box methods. Semantic segmentation of EL images yields detailed statistical data that can be correlated to the power output for large batches of PV modules, and this is a primary motivation for conducting

the research. While PV modules are priced based on the power output, the quality of the EL images can provide additional insights for stakeholders in the PV industry. A semantic segmentation model translates the qualitative data to quantitative data that can provide insights at various stages of the PV module lifetime. Certain defects captured in EL images may correlate to power output measured at the manufacturer, while other EL defects may correlate to power loss in PV plants over time. The multi-class defect detection and quantification at the pixel level generated by a semantic segmentation model generates a rich dataset to draw such important correlations. Cell-level image classification enables correlation studies between counts of cells with defects and module power output. Object detection enables a richer dataset for analysis that includes counts of multiple defects per image and some indication of each defect's size based on the bounding box size. On the other hand, semantic segmentation generates the richest data for correlation analysis because it includes a precise estimate of the size of every defect detected on each cell. For example, the size of an inactive area defect or a brightening defect in a solar cell correlates to power loss.

Figure 1.1 shows a snippet from a web-based tool for SCDD under development at the CSIR based on the outcome of this research. EL images are provided as input, and the statistical summary and visualisation of all the defects detected in a batch of modules are generated as the outputs. The batch statistics can be used as input for linear regression methods and other machine learning models to correlate the percentage of pixels associated with specific defects to the electrical output of the PV module.

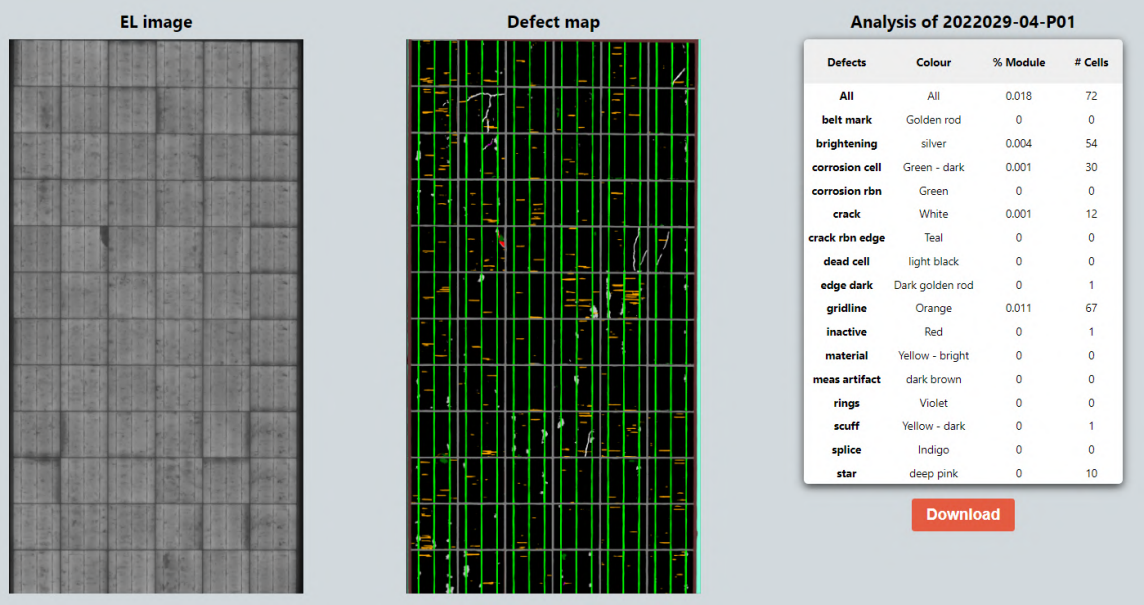


Figure 1.1: Web-based tool for SCDD under development at the CSIR which translates raw qualitative data contained in the EL image (left) to a prediction mask (centre) to a structured qualitative dataset for post-prediction analysis and correlation studies (right)

Modern, deep learning techniques for computer vision provide an effective method for converting terabytes of raw data contained in EL images to valuable statistical information for stakeholders all across the PV industry value chain. Stakeholders throughout the PV industry can use statistical information to assess the quality of the PV modules in a given batch of PV modules. Defect detection and quantification in EL images can improve the efficiency and

reliability of PV modules at the factory and during PV plant operations and maintenance by identifying and quantifying defect types and faulty modules.

1.2 Contributions

The contributions of this work include two published journal articles [Pratt *et al.* 2021a 2023], one article under review [Torpey *et al.* 2024], and several benchmark datasets to enable other researchers to develop methods for SCDD using semantic segmentation. A summary of each paper and the benchmark datasets follow.

The article published in Renewable Energy [Pratt *et al.* 2021a] was the first published work on SCDD using semantic segmentation on the novel dataset developed for this research. Earlier work in the field was focused on image classification and object detection using bounding boxes. In this research, we successfully trained a semantic segmentation model to detect multiple defects and features in EL images of solar cells using a U-Net model with a VGG16 backbone. We also demonstrated that the defects could be numerically quantified and correlated to changes in the EL images following a sequence of accelerated stress tests conducted in a PV module lab. The co-authors included Devashen Govender and Professor Richard Klein. I was the lead author.

The second article published in Systems and Soft Computing [Pratt *et al.* 2023] focused on evaluating alternative fully-supervised models and the impact of class weights on predictions. This article presented the expanded benchmark dataset and results for SCDD using deep learning models trained on 24 defects and features in EL images of crystalline silicon solar cells. Four deep learning models (U-Net-12, U-Net-25, PSPNet, and DeepLabv3+) were trained using equal class weights, inverse class weights, and custom class weights for twelve sets of predictions for each of the 50 test images. The class weights were added to address the class imbalance in the dataset. The DeepLabv3+ with custom class weights scored the highest in terms of mean Intersection over Union (mIoU) for the selected defects in this dataset. While the mIoU for cracks was low (25%) even for the DeepLabv3+, the recall was high (86%), and the resulting prediction masks reliably located the defects in complex images with both large and small objects. The co-authors included Jana Mattheus and Professor Richard Klein. I was the lead author.

A third article is under review by the Journal of Electronic Imaging. The research evaluated semi-supervised and self-supervised models for SCDD. We performed a large-scale evaluation of various pretraining methods for SCDD in EL images. We also experimented with both in-distribution and out-of-distribution (OOD) pretraining to observe how this affects downstream performance. The results suggested that supervised training on a large OOD dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and semi-supervised pretraining (CCT) all yield statistically equivalent performance for mIoU. We achieved a new state-of-the-art for SCDD and demonstrated that specific pretraining schemes resulted in superior performance in underrepresented classes. Additionally, we curated an unlabelled dataset of 150,000 EL images for further research in developing self- and semi-supervised training techniques in this domain. David Torpey was the lead author, and I was the second author.

The new benchmark datasets were published on GitHub.[1] The repository contains directories for each version release. The first version (2021-11-04) stores the data for the second article [Pratt *et al.* 2023]. The second version (2022-07-23) stores the data used for the third article [Torpey *et al.* 2024] submission. The third version (2022-10-08) stores the most current version, containing 695 EL images and corresponding ground truth masks. The images comprise mono-c-si and multi-c-si cells, including samples from older modules no longer in production and the latest half-cut solar cells on the market. The variety in the dataset was consciously designed so that the model would have a broad range of applicability for all types of modules in the market and the field. A dataset of unlabelled EL images was also published using a link provided on the github repository. This dataset includes over 150,000 EL images for further research into self- and semi-supervised learning in this field.

Table 1.1 summarises the contributions from each author to each paper following the contributor role taxonomy (CRediT) [Brand *et al.* 2015]. Lawrence Pratt, the PhD candidate, was the lead author of two papers and the second author of the third paper. He curated the dataset and generated the initial set of images in the EL dataset as a proof of concept for the fully-supervised models. He subsequently led a university student team organised by Professor Klein to expand on the dataset. He selected representative images for annotation, wrote work instructions, assigned images to team members, addressed inter-rater variability, and reviewed the final versions for approval into the benchmark dataset. Finally, he created and published the unlabelled dataset with more than 150,000 EL images for further research into semi- and self-supervised learning models.

Table 1.1: Contributions from each author towards Paper 1 [Pratt *et al.* 2021a], Paper 2 [Pratt *et al.* 2023], and Paper 3 [Torpey *et al.* 2024] by Lawrence Pratt (LP), Richard Klein (RK), Devashen Govender (DG), Jana Mattheus (JM), and David Torpey (DT)

Term	**Paper 1**	**Paper 2**	**Paper 3**
Conceptualization	LP, RK	LP, RK	DT, RK
Methodology	LP, RK	LP, RK	DT, RK, LP
Software	LP, DG	LP	DT, LP
Validation	LP, DG	LP	DT, LP
Formal analysis	LP, RK	LP, RK	DT, LP, RK
Investigation	LP, DG	LP, JM	DT, LP
Resources	LP, RK	LP, RK	DT, LP, RK
Data Curation	LP, DG	LP	LP, DT
Writing - Original Draft	LP	LP	DT, LP
Writing - Review and Editing	LP, RK	LP, RK, JM	DT, LP, RK
Visualization	LP	LP	DT, LP
Supervision	RK	RK	RK
Project administration	LP, RK	LP, RK	DT, RK
Funding acquisition	LP, RK	LP, RK	DT, RK, LP

[1] https://github.com/TheMakiran/BenchmarkELimages

1.3 Thesis structure

The paper is organised into ten chapters and an appendix. Chapter 1 provides an introduction to the research topic and the motivation. Chapter 2 provides general background on solar energy, electroluminescence imaging, EL image preprocessing, and PV module performance measurements. Chapter 3 summarises the published work on SCDD in EL images and deep-learning models. Chapter 4 summarises the general methods for training, testing, and evaluating the experimental results throughout this research. Chapter 5 summarises the datasets developed for this research project to enable training for fully supervised, self-supervised, and semi-supervised semantic segmentation models. Chapters 6-8 summarise the separate journal articles related to this research project that were submitted to scientific journals for publication and conferences for presentation. Chapter 9 provides an illustrative example of a regression analysis on the power output of PV modules versus the semantic segmentation output to illustrate a potential use case for the output of a semantic segmentation model. Chapter 10 provides the conclusions. The appendix consists of images representing the defects and features with corresponding descriptions.

Chapter 2

Solar Energy and PV module Characterisation

Chapter 2 provides general background information on renewable energy, solar energy, electroluminescence (EL) imaging, EL image preprocessing, infrared imaging, and PV module performance measurements. This information provides context for the research and how it might be used to improve fault detection in solar PV modules.

2.1 Renewable Energy Industry

Modern human civilisation relies on primary energy that comes in many forms, including chemical, electrical, radiant, mechanical, nuclear, and thermal. The primary fuels are converted to other forms of energy to serve a purpose. For example, burning wood converts chemical energy stored in the bonds of the wood into thermal energy (heat) and radiant energy (light). Burning fossil fuel in a generator converts chemical energy from the fuel into mechanical energy that turns an electromagnet that converts the mechanical energy into electrical energy. In 2021, global primary energy consumption reached 176 431 terra-watt-hours (TWh) [Ritchie, Hannah and Roser, Max 2023]. Wind energy and solar energy can also be converted into electrical energy. According to Ritchie, Hannah and Roser, Max [2023], wind and solar accounted for 4.2% of global energy consumption, while natural gas, oil, and coal combined accounted for 77.1% of the energy consumption. The transition from fossil to renewable fuels is moving slowly across the entire energy landscape.

Primary energy can be used directly or converted into other forms of energy before end use. In South Africa, the total primary energy supply (TPES) is dominated by coal, as shown in Figure 2.1 [Dlamini *et al.* 2022]. Approximately 33% of the coal is exported, and another 40% of the energy from coal is lost in conversion efficiencies and transmission. Some primary energy is used directly by consumers in the form of natural gas and biomass, while the remaining supply is converted into electricity and petroleum products. Industrial users and the transportation sector dominate the total final energy consumption (TFEC). Wind and solar make only a small contribution to the primary energy supply.

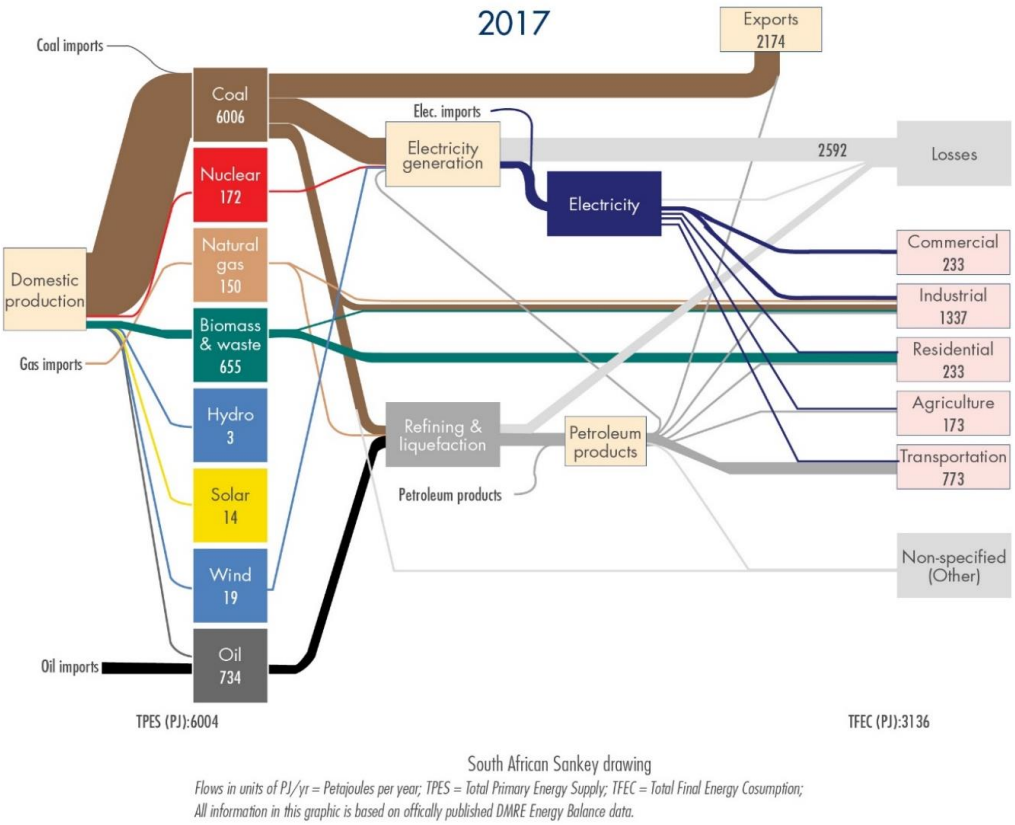


Figure 2.1: Energy flow diagram for South Africa in 2017 moving from primary energy supply (left) to energy conversion (middle) to end users (right) [Dlamini *et al.* 2022]

Renewable energy forms a subset of the primary energy supply derived from natural sources that are replenished at a higher rate than consumed, according to the definition on the United Nations Climate Action website [United Nations 2023]. Renewable energy sources include wind, solar, geothermal, hydropower, ocean, and bioenergy. Renewable energy is not new to human civilisation, however. Humans have been actively harnessing renewable energy since first passively warming themselves in the sun and cooking with fire, discovered over 200,000 years ago. Since the dawn of the planet, renewable solar energy from the sun has sustained plant and animal life on Earth. Over millions of years, that same solar energy was converted into the fossil fuels we use today. The chemical bonds and the carbon in plants convert radiant energy from the sun to chemical energy through photosynthesis. That chemical energy became fossil fuels after millions of years under pressure. In that sense, even the fossil fuels that society depends on today were originally a form of renewable energy from the sun. However, the rate at which society consumes these fossil fuels today far exceeds the rate at which they are replenished.

Electricity is critical to modern living. Contrary to the broader energy landscape, global electrical energy generation is rapidly shifting from fossil fuels to renewables. Renewables are predicted to account for over 90% of global electricity capacity expansion between 2022 and 2027 [Agency 2022b]. However, renewables will still provide less than 20% of the worldwide power generation due to the relatively low capacity factors. The capacity factor is a measure used in the energy industry to assess the actual output or utilisation of a power generation facility relative to its maximum potential output over a given period. A solar-powered electricity generator has a relatively low capacity factor compared to a coal or nuclear power plant because the fuel from the sun that powers the solar PV generator is only available for 8-12 hours per day, while the fuel that supplies a coal or nuclear generator is available 24 hours per day. Still, the low capacity factor has not hindered deployment in many parts of the world. China is predicted to install almost half the new global renewable energy capacity [Agency 2022b] by 2027.

2.2 Solar Energy

Solar photovoltaic (PV) systems drive growth in the renewable electrical energy industry. Distributed and utility-scale solar PV generators account for more than 50% of the renewable annual net capacity additions globally (Figure 1.5 in Agency [2022b]). Cumulative solar PV capacity is predicted to reach 1,500 GW globally by 2027, surpassing natural gas and coal combined [Agency 2022b]. For reference, South Africa had a nominal generating capacity of just 54 GW from all sources at the end of 2022, including 2.3 GW of utility-scale solar PV [Pierce, Warrick and Le Roux, Monique 2023]. Furthermore, China added more than 51 gigawatts of small-scale solar power in 2022 alone [Bloomberg], nearly equal to the total South African grid generating capacity. The trend in solar growth is driven by declining costs for PV systems, rising costs of fossil fuels, and the need to combat climate change by reducing CO_2 emissions. Solar PV systems are also quick to build, scalable, and adaptable to both urban rooftops for residential and commercial users and open space for ground-mounted, utility-scale power producers. Coupled with storage systems, a hybrid PV system can also mitigate the impacts of load-shedding in regions where the national grid supply is prone to outages, such as South Africa.

PV systems typically consist of three major components. Firstly, all PV systems include PV modules. PV modules are made from individual solar cells that convert radiant energy into electrical energy. Second, most PV systems include inverters that convert the direct current (DC) electricity from the PV module into alternating current (AC) that powers most gadgets and appliances used in everyday life. Third, the off-grid and hybrid systems include some form of battery storage or small fossil-fuel generator to provide electrical energy when the grid is down or the sun is not shining. Older off-grid systems relied on the lead-acid battery for storage, but the lithium-ion battery is a common storage medium in modern hybrid PV systems. These systems will also include some form of energy management system to control the energy balance from the grid, the generators, and the storage components.

PV systems generate electricity by converting the radiant energy from the sun in the form of photons into electrical energy in the form of electrons. The electrons are then used to power anything from the lights in our homes to electric cars on our roads. The basic operating principle of a solar cell, known as the 'photovoltaic effect,' was discovered by Edmond Becquerel in 1839 [Becquerel 1839]. The solar cell harnesses energy from the sun in the form of photons to generate a voltage, thus the term photo-voltaic, or PV, for short. The solar cells that make up the PV modules can be made from crystalline silicon wafers, thin films deposited onto glass, or a combination of both. They can be designed to operate under one sun intensity or under concentrated sunlight using dishes, hyperbolic troughs, Fresnel lenses, and solar trackers. A Fresnel lens is a series of concentric rings or steps, each with a slightly different curvature, that concentrates the sunlight passing through the lens on a smaller area Xie *et al.* [2011]. Most of the PV systems in the world are powered by silicon wafer-based solar cells assembled into rugged PV modules that can be mounted on virtually any surface to generate DC electricity in any environment where the sun shines.

The quality and reliability of solar PV modules are critical to ensure the security of the electricity grid as the contribution from solar energy continues to grow. Poor quality and reliability of the PV modules will have an immediate and long-term impact on the safety, performance, and financial return on investment from the PV plant. PV modules constitute roughly 25-35% of the overall cost of utility-scale solar PV projects in the 5-100 MW range, and the module cost remains the single biggest cost item for PV systems regardless of the scale [Fu *et al.* 2017]. Manufacturers are under constant pressure to deliver PV modules at lower prices, which can conflict with the needs of consumers who expect a high-quality and reliable product. Independent analysis and defect detection in EL images is one means for buyers to hold manufacturers accountable for quality and reliability in a cost-competitive market.

PV modules made from crystalline silicon cells are susceptible to cracking, and cracked cells lead to decreased electricity generation over time [Köntges *et al.* 2010]. Cracks form during module manufacturing, shipping, installation, and heavy stresses induced by wind, snow, and human traffic during routine operations and maintenance. Electroluminescence (EL) images are widely used in the industry to detect cracks and other defects in solar PV modules. The sheer volume of solar modules made from individual solar cells poses a challenge for monitoring their quality and reliability, making SCDD a critical tool for the present and future energy industry increasingly dependent on solar PV generators.

2.3 Electroluminescence imaging

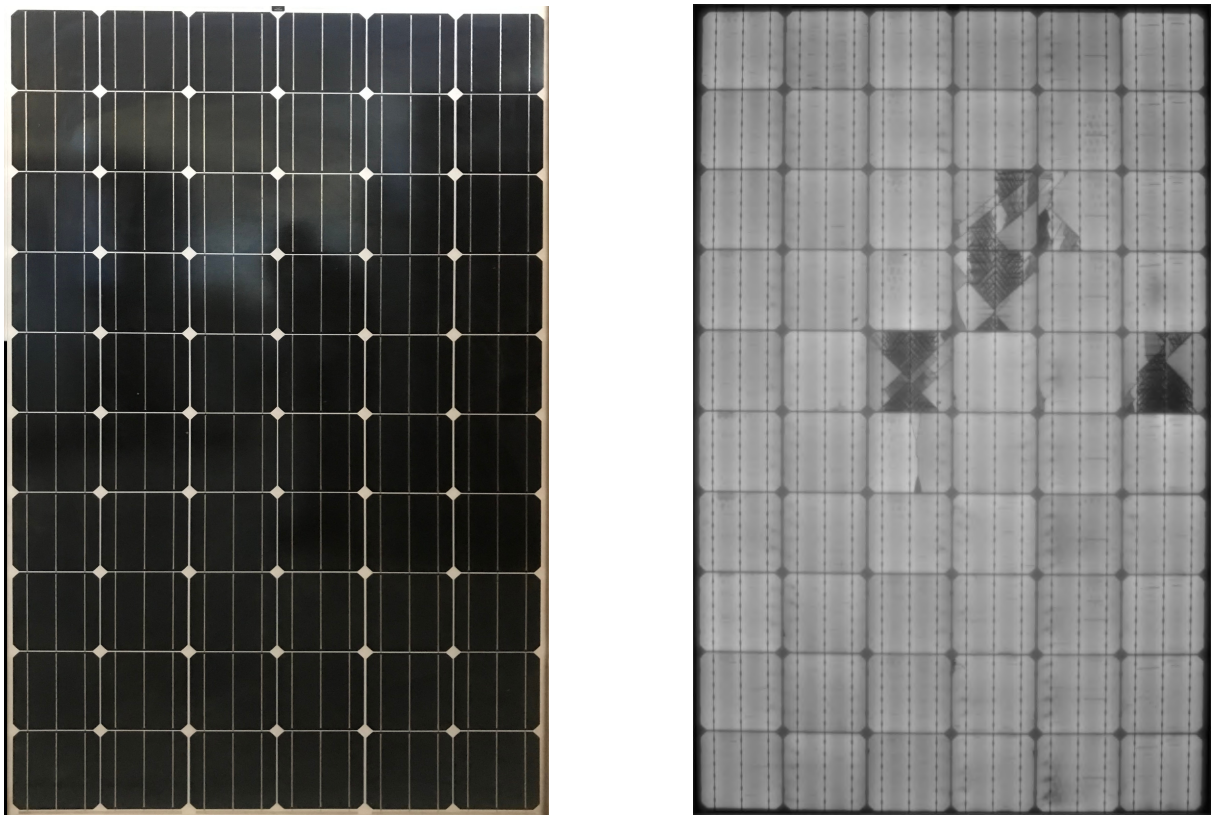
Electroluminescence (EL) imaging of silicon solar cells was first introduced by Fuyuki *et al.* [2005] to map minority carrier diffusion length in multi-c-si solar cells using a photographic survey method. Long minority carrier diffusion lengths correlate to high collection efficiency and, ultimately, the output power of the finished solar cell. Longer diffusion lengths increase the probability of the minority carrier collection at the nearest printed metal circuit designed to extract current from the cell before recombination occurs. Minority carrier diffusion length is measured in the cell manufacturing line on a sampling basis using different methods that are more expensive and time-consuming than EL. Fuyuki *et al.* [2005] showed how the spatial distribution of light intensity across a multi-c-si solar cell as measured by conventional methods correlated to the spatial distribution of light intensity across a solar cell as captured by an EL image. EL was proven to provide a semi-quantitative analysis of minority carrier diffusion lengths in finished cells and modules, enabling 100% EL imaging of PV modules manufactured daily across the globe. Fuyuki *et al.* [2007] later showed how the ratio of the light intensity from EL images could be used to estimate the open circuit voltage (Voc) of a PV cell relative to a reference cell.

An EL image is generated by recording the light emissions from a PV module that occur when electron-hole pairs recombine. The electron-hole pairs are generated by forcing a direct current through the PV module with an external power supply. The photon emissions peak near 1150 nm wavelengths for c-si cells, beyond the range of electromagnetic radiation visible to the human eye, roughly 400-700 nm wavelengths. A modified digital camera captures the photon emissions and creates a grey-scale map of the photon emissions that is perceptible to the human eye. In general, brighter areas in EL images correspond to the higher current generation, and uniform intensity across the module is desirable.

Electroluminescence (EL) imaging is an effective method to detect micro-cracks in PV modules made from silicon cells [Köntges *et al.* 2014]. EL images enable defect detection in solar PV modules that are otherwise invisible to the naked eye, much like an X-ray allows a doctor to detect cracks and fractures in bones. An EL image is generated by forward biasing the PV module with an electrical current, which creates electron-hole pair recombination and photon emissions peaking near 1150 nm wavelengths for c-si cells, beyond the range of electromagnetic radiation that is visible to the human eye, roughly 400-700 nm wavelengths. A modified digital camera captures the photon emissions and creates a grey-scale map of the photon emissions that is perceptible to the human eye. Brighter areas in the EL image are associated with higher current generation under normal operation, and darker areas are associated with defects or intrinsic features of the cell that reduce the current generation. Higher current generation leads to a higher power output of the PV module, so generally speaking, a typical solar cell should generate a bright, uniform intensity map, as seen in the EL image.

Figure 2.2 shows a PV module made from mono-crystalline silicon wafers as seen with the human eye (Figure 2.2a) and the corresponding EL image (Figure 2.2b). The module comprises 60 cells arranged in a 6 x 10 grid with uniform spacing around each cell. The cells are typically electrically connected in series with multiple interconnecting ribbons on each cell. In this case, the cells are connected by the three ribbons running vertically over the top of each cell, while more modern PV modules often use 5-10 interconnect ribbons. The PV module looks perfectly fine to the naked eye, but the EL image reveals several cracked cells that were not detected

during a visual inspection of the module itself. Five severely cracked cells are visible near the centre of the EL image and show dark regions corresponding to cracks and inactive areas. This level of damage to the solar cells almost certainly occurred after leaving the factory because EL images are used for quality control to screen modules with severely cracked cells at the factory.



(a) Optical image

(b) EL image

Figure 2.2: Optical (a) and EL image (b) of a PV module made from mono-crystalline cells with severely cracked solar cells

The IEC 60904-13 [International Electrotechnical Commission 2018] and the IEA PVPS Report T13-10:2018 [Jahn *et al.* 2018b] have been published to cover best practices for EL imaging, pre-processing images, and analysis. Jahn *et al.* [2018b] compares infrared (IR) imaging and EL imaging. The authors concluded that IR imaging is helpful for identifying small temperature gradients within a module and across modules installed co-planar in a PV plant. The low-resolution temperature gradients observed in IR images often correlate to disconnected PV modules or strings, PID effects or shunted cells, and short-circuited bypass diodes. On the other hand, EL images provide high-resolution images of defects in modules. Figure 2.3 illustrates the difference between optical, infrared, and el images of an array of modules recorded in the field [Buerhop *et al.* 2022]. While the EL images provide more detail regarding defects, they are more time-consuming to collect. Specifically, the IEA PVPS Report T13-10:2018 states that a ground-based IR scan of a 1 MW PV plant could be completed in less than 10 hours, whereas a ground-based EL inspection of the same plant could take up to 100 hours. The report also notes that the IR images tend to identify faults that impact power output, whereas

defects identified in EL images often do not correlate to power output. In practice, the guideline recommends both techniques for characterising defects in PV modules.

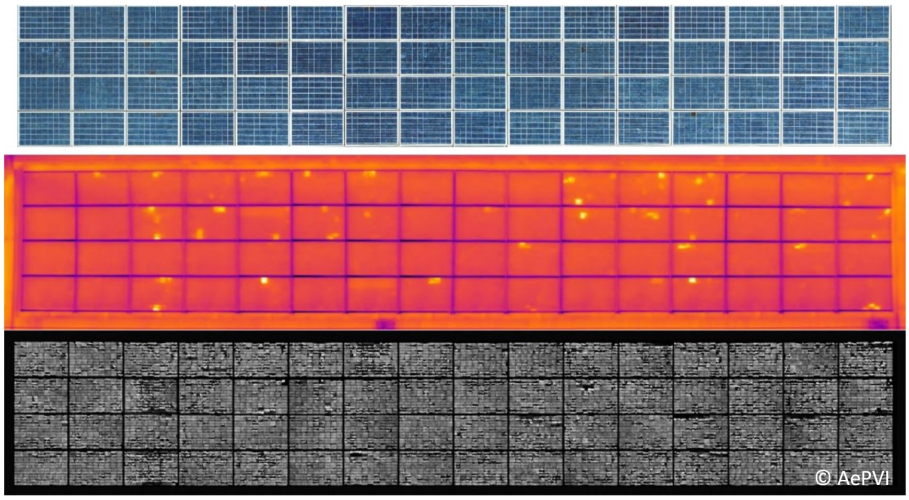


Figure 2.3: In-situ images of a 4 x 16 array of PV modules taken with an optical (top), infrared (middle), and EL (bottom) camera [Buerhop *et al.* 2022]

The IEC 60904-13 provides a library of defects and guidance on quantifying solar cell cracks using EL images. For example, Figure 2.4 shows an EL image from the SCDD dataset with three types of cracks defined in the standard: Type A, B, and C. Type A micro-cracks appear as line defects in EL images and do not remove active cell area from the cell or cause significant cell power loss. Type B cracks appear as grey regions delineate partially electrically disconnected regions. Type C cracks appear as black regions and delineate essentially electrically disconnected regions from the remaining module electrical circuit.

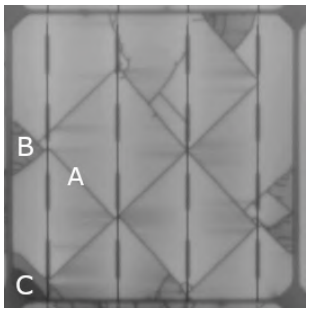


Figure 2.4: EL image of a single cell exhibiting three types of cracks: line cracks (Type A), partially disconnected regions (Type B), and cracks creating electrically disconnected regions (Type C)

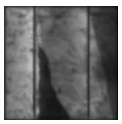
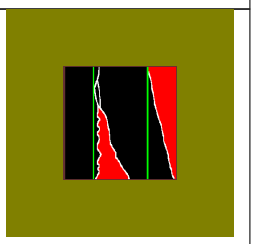
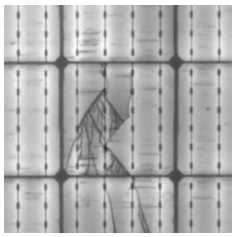
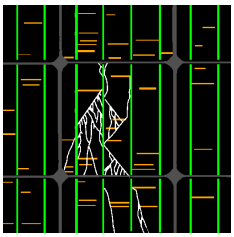
EL imaging is widely used throughout the solar industry to assess the quality and reliability of PV modules. Most PV modules sold today consist of solar cells made from crystalline silicon wafers, susceptible to cracks and other defects. Micro-cracks and other defects can form during cell processing, module assembly, transportation, installation, and environmental stresses that occur in the field over the twenty-five-year lifetime of a PV module. The cracks and defects can negatively impact performance in the field [Köntges *et al.* 2014]. An estimated 1.7 million EL images are produced every day across the globe in solar PV module manufacturing

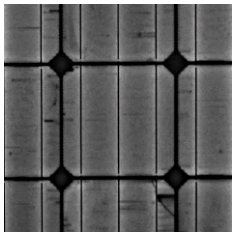
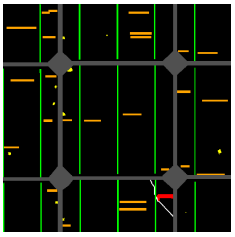
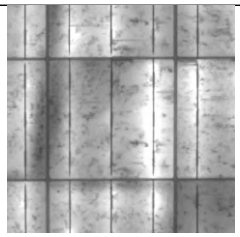
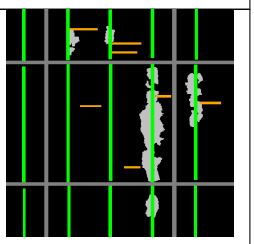
facilities alone, assuming 250 GW of annual production at an average of 400 W per panel. The prevalence of multiple defects (e.g., micro-cracks, inactive regions, gridline defects, and material defects) in the EL images must be detected and accurately quantified to evaluate the quality and reliability of PV modules properly.

PV modules are susceptible to cracks and many other defects which can also be seen in EL images. Much of the published work focuses on crack detection and inactive areas since these are primary concerns with respect to quality and reliability. Cracks appear as dark lines, singular in nature or forming dendritic patterns as seen in Figure 2.2. Inactive areas appear as dark, irregular-shaped regions where cell sections are electrically isolated from the external circuit due to Type C cracks. Gridline defects appear as dark lines running perpendicular to the ribbon interconnections and correspond to regions of the electrical circuit with abnormally high resistance. Other defects with origins in manufacturing and environmental stress may also be observed in some images, such as belt marks, dark edges along one or two sides of the cell, corrosion along the ribbon interconnects, and dead cells.

A complete library of defects and features for the novel dataset created for this research is included in the appendix A.1. The labels and crack classifications as defined in International Electrotechnical Commission [2018] were not incorporated into the library for this new dataset. Standardisation of defect labels going forward is desirable, and the IEC standard is a good place to start. However, there are many defects in SCDD dataset that do not exist in the standard. Some examples of extrinsic defects in this new benchmark dataset are shown in Table 2.1 along with a brief description. The ID # corresponds to the number from the full list of defects and features. Most defects are identified as dark regions with a characteristic shape in the EL image. The brightening defect is the exception to the rule because the model is trained to find relatively bright regions.

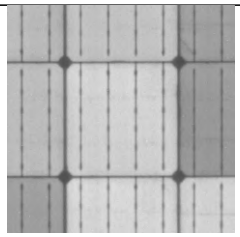
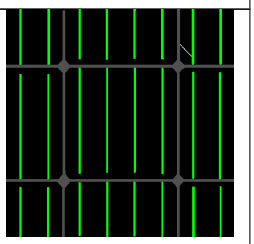
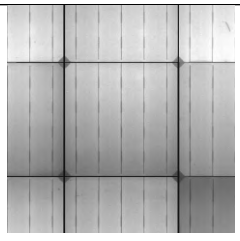
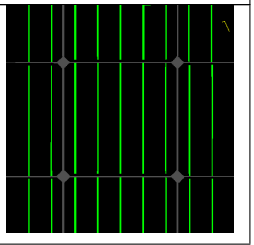
Table 2.1: Examples of extrinsic defects with descriptions, EL image, and ground truth (GT) from the benchmark dataset developed for this research.

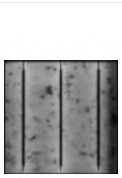
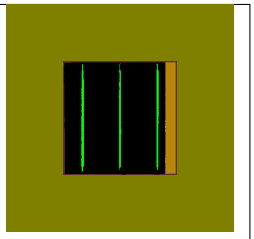
ID #	Description	Color	EL	GT
11	**Inactive regions** identify dead areas often caused by cracks in the cell.	Red		
14	**Crack defects** identify microcracks often resulting from mishandling that occurs primarily after the module has left the factory. Factories typically screen for cracks that occur during manufacturing before shipping.	White		

15	**Gridline defects** or 'finger defects' identify screen-printed silver lines with elevated resistance. The silver lines carry current from the cell surface to the ribbon interconnects.	Orange		
27	**Brightening defect** identifies areas where current crowding occurs, typically the result of increased resistance over the other darker regions.	Silver		

Some examples of intrinsic cell features are shown in Table 2.2. The ID # corresponds to the number from the full list of defects and features. Some features, such as ribbon interconnection regions, are printed on top of the incoming bare wafers during cell manufacturing. Other features, such as busbars and junction boxes, are added during module assembly. Features serve various functions in the PV module, including current extraction from the point of electron-hole pair generation, cell interconnection, and extraction of the electrical energy from the module through busbars and external leads. Features are also associated with dark regions of the EL image but with different characteristic shapes compared to defects. While the cell spacing and ribbons are intrinsic to a PV module, the padding layer is not. The padding layer shown below is the result of the image pre-processing method that ensures a full cell is always at the centre of each image. The edge and corner cells images cropped for a full modules will have a corresponding padding layer, but it should not be considered an extrinsic defect or an intrinsic feature.

Table 2.2: Examples of intrinsic features with descriptions, EL image, and ground truth (GT) from the benchmark dataset developed for this research.

ID #	Label	Color	EL	GT
2	**Spacing** identifies the area of the module that is necessary to maintain electrical isolation between adjacent cells.	Dark grey		
4	**Ribbons** identify the narrow metal wires that serve to interconnect the cells in series to build up the module voltage and extract the electrical energy.	Lime green		

7	**Padding** identifies the padded areas added around the cell images strictly for processing. The padding is neither a feature or a defect.	Olive		

2.4 EL imaging in the lab or factory

EL images are collected on every module in every factory worldwide before shipping to customers. Figure 2.5 shows an example of a power measurement and EL characterisation station typical of a PV module manufacturer. At the end of the production line, modules are loaded onto a conveyor belt and moved into a dark chamber where the power output is measured and the EL images are recorded. Given the volume of images and the need for analysis, automatic defect detection and classification in EL images of solar cells has been the subject of many scientific publications, as described in the following sections. Some forms of computer-aided vision systems are typically incorporated into the inline inspection equipment to assist human operators in identifying faulty cells by highlighting potential defects with a bounding box around the cell. Computer vision models based on semantic segmentation could replace the bounding box methods and provide higher-resolution data for monitoring production quality within the factory.

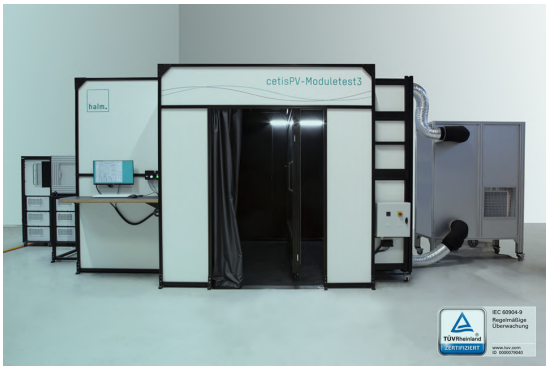


Figure 2.5: Example of an I-V and EL characterisation station typical of a PV module manufacturer

Figure 2.6 shows test equipment typical of a PV module test lab. Figure 2.6a shows the EL imaging studio at the CSIR PV Module Quality and Reliability Lab (PQRL). The setup includes a digital camera modified to allow light in wavelengths above the visible spectrum, a power supply bank to supply the electrical current, a data logger to record module temperature, a computer to store the image, and a dark room with temperature control. Figure 2.6b shows the



(a) EL imaging studio



(b) I-V measurement chamber

Figure 2.6: Test equipment at the CSIR PQRL for (a) recording EL imaging and (b) measuring module power inside the sun simulator [halm. 2024]

sun simulator used to flash the module with simulated light and record the power output of the module. Together, the EL studio and the sun simulator generate high-quality data to quantify PV module performance. However, correlating EL images to PV module power measurements remains a challenge due to the complexity of the qualitative data embedded in the EL image.

2.5 EL and IR imaging in the field

In-situ EL imaging has been the subject of many scientific publications. Most EL images are taken in controlled environments such as manufacturing facilities and PV test labs where temperature and light levels can be managed. However, the industry requires field tests on large-scale PV plants during their operational lifetime. In-situ monitoring is preferable to lab tests post-installation because it avoids the costs and risks of removing the modules from operation, packing, and shipping to a lab.

Solar Zentrum developed the ground-based DaySy system that enabled string-level EL imaging in the field during the daytime [Reuter *et al.* 2015]. This device is expensive but effective. The DaySy requires a large power supply to energise an entire string of PV modules instead of one module at a time, which increases the cost. However, images of PV module strings can be taken in the field during the day rather than at night, saving the time and resources required to remove modules from the racks for imaging indoors. Indoor EL imaging in the factory or lab is typically conducted in a dark room to minimise the impact of stray light, one module at a time.

Hobbs *et al.* [2018] deployed special rigging equipment to enable ground-based EL imaging of single modules in the field at night. A comparison of defect detection in EL images taken in the field versus those taken in the lab yielded nearly identical results. The field EL images captured 93% of the defects observed in the lab images. However, the authors note that some of the missed defects may be due to additional damage that occurred during handling and transport to the lab. The in-situ EL imaging techniques also eliminate the risk of additional damage to PV modules during handling.

Drone-based EL imaging has the potential to generate millions of EL images taken on fielded PV modules. This is an exciting and challenging development for SCDD as the volume of

images will increase substantially, as will the need for automated tools to assess these images. However, with the increase in volume from in-situ measurements, the quality of the images may decrease compared to indoor, lab-based EL images. Drone-based methods for EL imaging are still being developed, with only a few companies worldwide offering a service. K.G.Bedrich [2021] filed a patent for EL imaging using an aerial vehicle.

Koch *et al.* [2016] compared EL images captured in the field using a ground-based system and a drone-based system. The ground-based system deployed a cherry picker to elevate an operator with a camera a few meters above the PV modules. The drone-based system also required a specially designed switch box that allowed fast switching between the power supply and the strings to maximise the number of images collected during the short, 60-minute flights. The ground-based system also required an external power supply, but the fast switching was not a requirement, given the slow pace at which the ground-based images were collected. Proprietary software was developed to analyse the drone-based images and identify patterns of defects in the EL images. The software was trained to detect cracks within cells and dark cells that were likely impacted by potential induced degradation (PID). The cell images were then aggregated to the batch level for pattern detection across the modules. The analysis revealed a higher likelihood of cracks at the centre of modules and PID degradation at the edges of the modules. The quality and resolution of the EL images were not addressed, but the pattern detection was sufficient for the spatial analysis conducted at the module level. This article provides a good example highlighting the importance of batch statistics and spatial analysis of defects detected in batches of EL images.

Alves dos Reis Benatto *et al.* [2020] published detailed EL images captured on a single PV module with a drone in full daylight using AC and DC pulse modulation to generate EL and background images. The post-processing of images included background subtraction. The images were clear enough to detect large inactive areas, but the image quality was still inferior to indoor, lab-quality images. The results also demonstrate that the quality of images taken in motion was lower than those taken while stationary. This work suggests that high-volume, rapid EL scans of an entire PV plant remain challenging.

By contrast, drone-based infrared (IR) imaging of PV plants is well-established and effective at detecting hotspots. Figure 2.7 shows an example of an IR scan recorded by the author at the CSIR in 2021 using a handheld camera. The IR image shows 42 modules mounted on dual-axis Tracker 11. The red square highlights an anomaly for one cell in one module operating at a higher temperature than the surrounding cells. Further investigation is required to understand the root cause in this case. The green circle highlights an elevated temperature over a small area at the edge of one module. This location corresponds to the location of the junction boxes, and the elevated temperature can be seen on every module in the array. The IR scans effectively capture defects that lead to hotspots, but cracks rarely lead to hotspots. However, IR images do not provide the same level of resolution as EL images [Koch *et al.* 2016], so the research and development into in-situ EL imaging continues.

There are two major challenges to drone-based EL imaging that do not impact drone-based IR scans. First, the modules need to be energised for EL imaging by an external power supply. This requires that the PV modules or strings of PV modules be disconnected from the system and connected to an external power supply that is large enough to power a single module or a string of modules during imaging. IR images, by contrast, are recorded during the day while

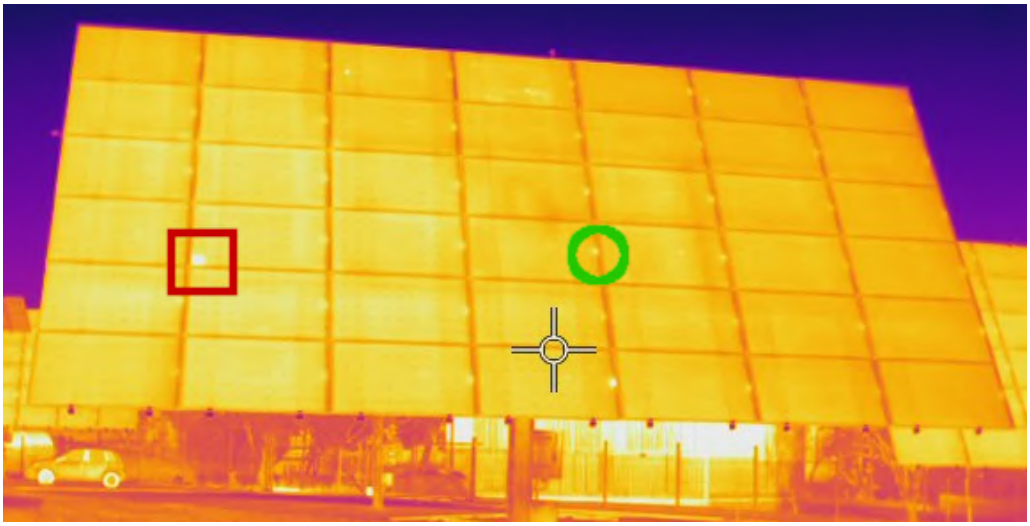
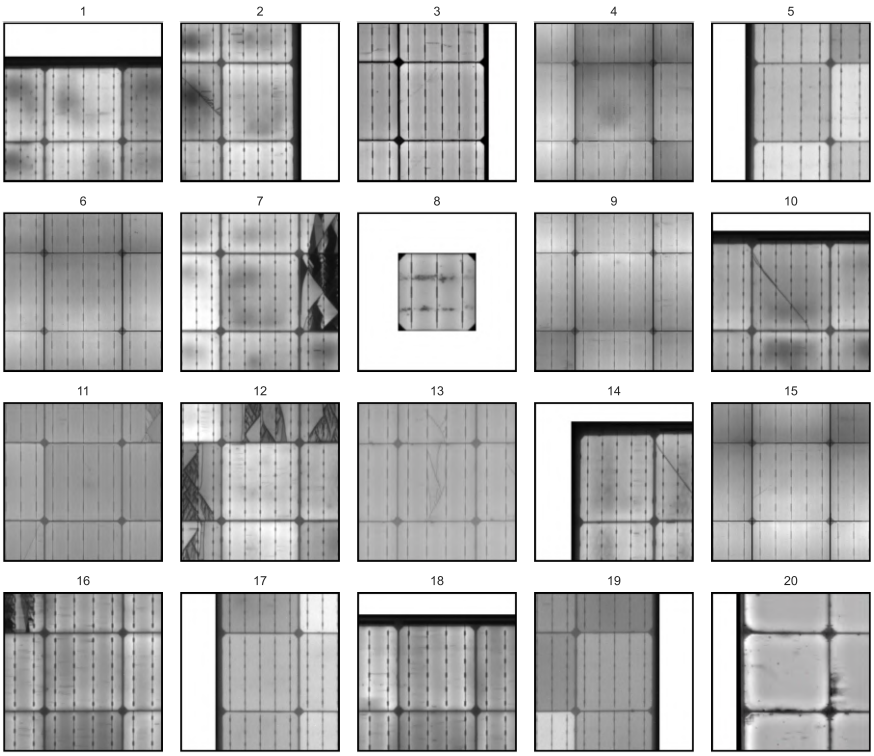


Figure 2.7: Example of a hotspot cell (red square) and the thermal signature for the junction boxes on each module (green circle) in an IR image taken at the CSIR

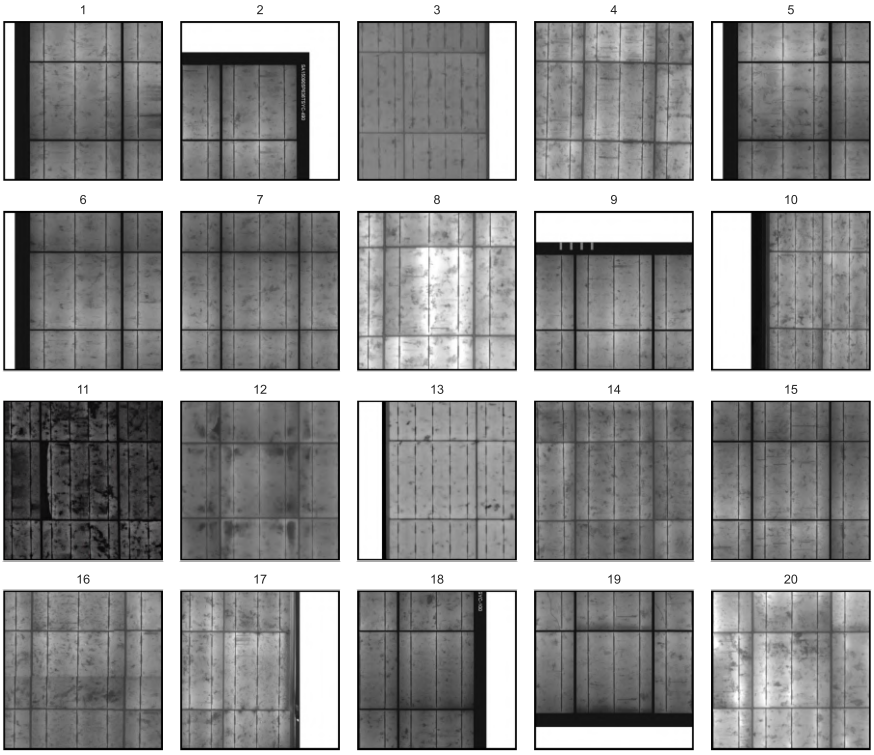
the PV modules are under normal operating conditions, so an entire PV plant can be scanned without having to make any electrical disconnections. Second, EL imaging requires tight focus control compared to IR scans. A bit of blurriness due to loss of focus will have little impact on a thermal scan, but blurriness may entirely obscure the small defects hidden in the EL images.

2.6 Pre-processing - cropping the cells from EL images of full modules

EL images of PV modules are typically cropped to evaluate one solar cell at a time when using computer vision models. Figure 2.8 shows examples of cell-level EL images for mono-crystalline silicon (mono-c-si) and multi-crystalline silicon (multi-c-si) cells samples from the SCDD dataset. These images reveal various extrinsic defects of interest to the trained analyst: cracks, inactive regions, brightening, horizontal gridline defects, and cell corrosion. The images also capture many of the intrinsic features of solar cells that can be seen in the optical images, such as the interconnect ribbons that appear as dark lines in the EL running from top to bottom and the related solder bond pads underneath the ribbons that are not visible in the optical image. The images also reveal a key difference between the cells made from mono-c-si wafers (Figure 2.8a) and multi-c-si wafers (Figure 2.8b). A mono-c-si cell without defects exhibits a clear background, whereas a multi-c-si cell without defects exhibits irregular dark regions. The dark regions arise from the various crystal orientations that form during the casting of ingots from which multi-c-si wafers are sliced, thus complicating object detection and classification on multi-si cells. Mono-c-si wafers are sliced from single crystalline ingots with uniform orientation; thus, the grain boundaries are absent. In the optical images, grain boundaries can be difficult to see, while in EL images the grain broundaries are clearly eveident.



(a) mono-c-si cells



(b) multi-c-si cells

Figure 2.8: Examplees of (a) mono-c-si solar cells with rounded corners and (b) multi-c-si solar cells with 90°corners sampled from the dataset developed for this research. Cells with and without defects are included.

Several papers describe methods for cropping module-level EL images into component cells. The cell-level images cropped from module-level EL images yield larger training datasets because typical module images have 60 or 72 cells per module. The smaller cell-level images are also less computationally expensive to process. However, EL images are typically recorded for a full PV module and increasingly for strings containing twenty or more PV modules in series. Some articles use the term 'segmentation' to describe methods for extracting cell-level images from module-level EL images. In this thesis, the term 'cropping' is used instead of 'segmentation' to avoid confusion with the term 'semantic segmentation.' The detailed and lengthy pipelines are useful for generating high-quality images of solar cells for subsequent defect detection. For example, Deitsch *et al.* [2018] proposed a method to crop cell-level images from lab-quality images of PV modules. The pipeline comprises 17 steps that can take over six minutes to process a single PV module. Given the time and complexity, this approach is untenable for pre-processing hundreds and thousands of images.

Sovetkin and Steland [2019] proposed a simpler method to crop cells based on finding the peak intensities in the column and row sums after correcting the image for rotation and perspective distortion. Mantel *et al.* [2019] developed a similar method using the number of rows and columns of cells in the module and the sum of the pixel intensities along each axis to identify the cell spacing. The intersection of the row and column spacing identified the corners of each cell to be used for cell extraction. This solution provides a simple, elegant solution that should be readily adaptable to any PV module design. A similar method was developed independently for cropping EL images used in this thesis.

2.7 Pre-processing - image quality

This section summarises the research and developments on preprocessing EL images to improve image quality before applying defect detection methods. The preprocessing methods may apply to module-level and/or cell-level EL images recorded in a lab, a factory, or in the field.

Bedrich *et al.* [2016] proposed methods to improve the quality of EL images by improving focus, correcting lens distortion, removing erroneous pixel-level artefacts, and correcting for rather larger perspective transformations. The ability to correct large perspective distortions is important, especially for EL images recorded in the field when camera positioning may be far from normal to the surface of the module.

Mantel *et al.* [2018] proposed two methods to improve image quality, one to correct for rotation and another to correct for rotation and perspective simultaneously. The first method consisted of rotating the image by various angles and selecting the most probable based on the projection on the X and Y axes of the image. The local maxima in peak intensities corresponding to the spacing between rows and columns of cells are identified during the rotation from -45 °C to +45 °C. The second method uses a Sobel gradient, erosion, dilation, and a direct linear transformation to correct for rotation and perspective on an example with a high degree of perspective distortion.

Sovetkin and Steland [2019] proposed another method for correcting rotation and perspective distortions in the PV module image before cropping. The rotation correction method is based

on a function that optimises the pixel intensities summed along the rows and columns. The perspective transform was based on a Hough transform.

Karimi *et al.* [2019] developed a preprocessing pipeline of the original raw EL images, including lens distortion correction, filtering, thresholding, convex hull, regression fitting, and perspective transformation to produce planar indexed module and single-cell images.

2.8 PV module performance measurements

Figure 2.9 shows the current-voltage (I-V) curve of a PV module overlaid with the power curve. The I-V curve is used to characterise the DC electrical output of a PV module. The I-V curve of a PV module is best measured on an indoor Class AAA solar simulator with controlled environmental conditions. Class AAA indicates Class A uniformity of the source light, Class A spectral match to natural sunlight, and Class A temporal stability of light intensity while the I-V curve is generated. The maximum power output (Pmp) is determined where the product of the current and voltage reaches a peak along the curve. The corresponding point on the curve determines the current at maximum power (Imp) and the voltage at maximum power (Vmp). The I-V curve also measures the short-circuit current (Isc) and open-circuit voltage (Voc), where the curves cross the corresponding X and Y axes. The price of the module paid in the market is based on the Pmp output measured from the I-V curve. The Isc and Voc are used as input parameters when designing and sizing other components of the PV system, such as cabling and inverters.

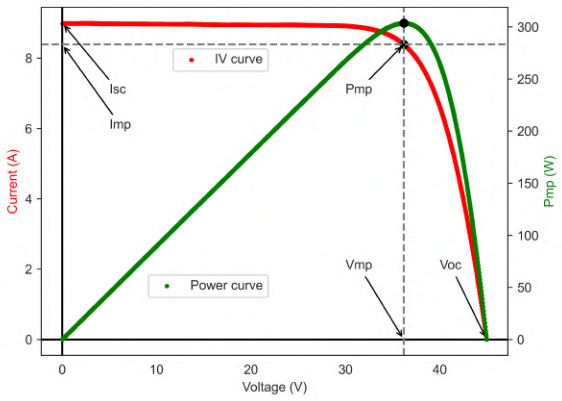


Figure 2.9: Example I-V and power curve created from the IV pairs recorded for a PV module measured at the CSIR

The electrical output of the PV modules is critical to the success or failure of the PV system over the expected lifetime. Modules are bought and sold based on a price per watt as measured at the factory or independent test lab under standard test conditions (STC). STC is defined as 1000 W/m^2 of irradiance, 25 °C cell temperature, and a spectral distribution from the light source that is consistent with natural sunlight at air mass = 1.5.

The electrical output of a PV module is impacted by a number of factors, some of which can be detected in an EL image. For example, large inactive areas seen as dark regions under EL may

correlate to lower Imp, which leads to a reduction in Pmp. Non-uniform resistance across the electrical circuit of a solar cell may be due to gridline defects which may decrease the Vmp. Non-uniform pixel intensity across broader regions of a cell may also correlate to lower Vmp, which leads to a reduction in Pmp. The statistical output from the SCDD model can be used to correlate defects with the electrical characteristics measured on the I-V curve.

2.9 Correlating EL defects to PV module power

This section summarises research conducted to quantify the impact of solar cell defects on module power and research to correlate features of EL images to module power.

Köntges *et al.* [2011] analysed the direct impact of micro cracks on the module power and the consequences after artificial ageing. The experimental results showed that artificially initiated micro-cracks in the silicon wafer do not reduce the power generation of a PV module by more than 2.5% relative if the crack does not harm the electrical contact between the cell fragments. However, the electrical resistance between cracked cell fractions increases with accelerated stress testing, leading some Type A micro-cracks (no electrically isolated regions) to become Type B micro-cracks (increased resistance across the crack) and even inactive areas (Type C cracks). The research showed that the number of cracked cells correlates with the power degradation after the accelerated ageing test, and the power loss shows almost a linear dependence on time.

Wu *et al.* [2018] evaluated three image processing tools to analyse defects and predict the performance of the photovoltaic modules using infrared thermography, electroluminescence, and ultraviolet-induced fluorescent images of the modules. For EL image analysis, the pixel intensity of a full module was binned into one of five levels, and each level was assigned a weight. The weighted average of pixel intensities in each bin for a module was correlated to the electrical performance of the module. Electroluminescence image processing demonstrated a linear correlation between the dark areas and the fill factor with an r^2 of 86.5%. This method is simple and effective, provided the EL images were taken under similar conditions to control electrical bias settings, stray light, temperature, camera setup, etc. The relative brightness of an EL image can also be changed by a user with simple image processing tools to increase or decrease brightness to facilitate visual analysis, and that will impact the correlation to power.

Bedrich *et al.* [2018] presented an empirical method for estimating relative power losses caused by potential-induced degradation (PID) for p-type solar cells and modules using quantitative electroluminescence (EL) analysis (QELA). PID degradation leads to interesting checkerboard patterns in PV modules where entire cells become inactive and dark under EL imaging, typically around the frame edge. The researchers estimated module power loss from the EL images by averaging the ratios of corrected EL intensities in images taken at Month 0 and again at Month 12 after PID degradation was detected. The modules were also measured on an indoor sun simulator to measure the power loss. A strong linear correlation was evident between the measured power from the sun simulator and the estimated power from the EL images with a root mean squared error (RMSE) of 3%. The method also relies on EL images taken in a consistent manner over time, and the results may not translate to EL images recorded by different labs or under different conditions.

Bdour *et al.* [2020] presented a summary of data collected from different projects in Jordan to describe the effect of microcracks on power loss. The article contains EL images showing cracks, PID degradation, and other defects in modules from multiple locations. I-V data was also collected on each module, and the delta to nameplate power was recorded along with the number of cracked cells and the shape of the cracks per module. However, there was no evidence of a rigorous analysis to correlate the cracks with electrical power loss relative to the nameplate rating. Perhaps the correlation between pixel intensity and power loss did not hold across multiple module types.

Karimi *et al.* [2020] developed models to predict PV module I-V features from EL image characteristics, especially ribbon corrosion. 15 modules were exposed to 3000 hours of damp heat and measured every 500 hours. 15 modules were exposed to 600 cycles of thermal cycling and measured every 200 cycles. 11,700 EL images of solar cells were derived from the 30 modules measured over the course of the accelerated stress testing. Four hand-crafted features were extracted from each cell-level image, including the busbar corrosion ratio (BBCR). A CNN was trained to classify each cell into one of five different classes corresponding to the level of busbar corrosion. The cells were then averaged to obtain the BBCR. Wider busbars in the EL images correlated to more corrosion around the busbars/ribbon interconnect. The BBCR and the other three features were then used to build statistical models for predicting the electrical performance of the module, quantified with an I-V curve, particularly the series resistance (Rs) and the maximum power (Pmp) of the module. The correlation between EL features and I-V was weak until 2500 hours of damp heat exposure, which is well beyond the standard 1000 hours of exposure for PV module certification. However, the quadratic model for regressing Rs on BBCR provided a good fit to the data with an r^2 of 73%. A third-order polynomial provided a good fit when regressing Pmp on BBCR with an r^2 of 81%, while a third-order polynomial fit of Pmp on median pixel intensity achieved an even higher r^2 of 88%. This paper was the first to show that features extracted from EL images could be correlated to I-V characteristics.

Sovetkin *et al.* [2021] published results attempting to correlate power loss to EL defects on thin film modules. A series of methods based on deep neural networks were evaluated for semantic segmentation of EL images of thin-film modules. The researchers presented a correlation plot between the electrical performance of the modules versus the number of detected shunts from a semantic segmentation model. Shunts typically change the shape of the I-V curves, specifically the slope of the I-V curve at short-circuit current (Isc), and those changes impact the output power. However, the scatterplot of the slope at Isc and the number of detected shunts resembled a shotgun blast with no correlation.

Dhimish and Hu [2022a] investigated the impact of cracks and fractural defects in solar cells and their cause for output power losses and the development of hotspots. Ten solar cells with cracks ranging from 1% to 58% of the cell area were characterised. The cells were carefully extracted from the PV module circuit so that each cell could be measured independently. The researchers claim the crack patterns were not altered during the process. I-V measurements were taken on each cell over various irradiance levels and correlated to the crack area. The researcher presented a near 1:1 relationship between the percentage of the cell impacted by the crack and the percentage of power loss, once the impacted area reached 11% or more. The cell with 58% of the area impacted by the crack showed a 60% power loss at both high and

low irradiance. The researchers also quantified the impact of the crack on hotspots, noting that the peak temperatures reached up to 45°C on a cell with a 25% crack percentage, 20°C above the operating temperature of cells with only minor crack percentages. As the crack percentage increased above the 25% level, the operating temperatures decreased significantly. Thus, cracks can lead to hotspots known to accelerate module degradation over time in the field.

2.10 Summary

SCDD in EL images of solar cells is a challenging and vital field, given the widespread uptake of solar systems worldwide. Solar modules are critical components of solar systems and are expected to operate for 25 years in the field. However, hidden defects in solar cells, such as cracks and inactive areas, can lead to higher degradation rates than expected, which can significantly impact the performance of PV modules. EL images provide meaningful information about the quality and reliability of the PV modules at the time of manufacturing and after years in the field. Still, they are difficult to analyse visually, given the complexity and volume of images. Modern computer vision methods and deep learning models can solve this problem and improve methods for PV module characterisation and fault detection.

Chapter 3

Related Work

Chapter 3 summarises the published literature on SCDD of solar cells using machine learning models. The first section covers SCDD on incoming wafers and solar cells in the factory before lamination into modules. The remaining sections cover research on defect detection in EL images of finished modules beginning with the most straightforward image classification methods through to the most complex form of semi-supervised models for semantic segmentation. The methods include image classification, defect detection, binary segmentation, multi-class semantic segmentation, self-supervised, and semi-supervised models.

3.1 SCDD on wafers and bare cells

The methods described in this section apply to defect detection in the factory before the bare wafers are made into cells or before the finished cells are soldered together into strings and laminated into modules. SCDD on bare wafers and solar cells is essential for monitoring PV module production lines before module assembly. Therefore, defect detection on incoming wafers and finished solar cells before lamination has been the topic of several papers. Wafers are the round, square, pseudo-square, or rectangular silicon substrates on which a solar cell is built. Figure 3.1 shows square multi-c-si and pseudo-square mono-c-si wafer substrates used to manufacture older generation solar cells with only two interconnection ribbons [Fisher *et al.* 2012]. However, some of the newest, large-area mono-c-si cells sold today exhibit less rounded corners. In contrast, the earliest generation of mono-c-si wafers were entirely rounded.

Tsai and Luo [2010] proposed a method for detecting fingerprints and contamination defects in optical images of multi-c-si wafers. The algorithm is based on a mean-shift technique that moves each data point to the mode of the data based on a kernel density estimator. This technique removes defect-free grain boundaries and noise from the image. The areas of high entropy in edge gradients that correlate to fingerprints are more easily identified. However, the method will not work for crack detection, which has more consistent edge directions than a fingerprint, as noted by the authors.

Lydia *et al.* [2017] proposed a method using Particle Swarm Optimisation for crack detection in images taken in the visible and infrared spectrum. Unfortunately, this paper is light on the

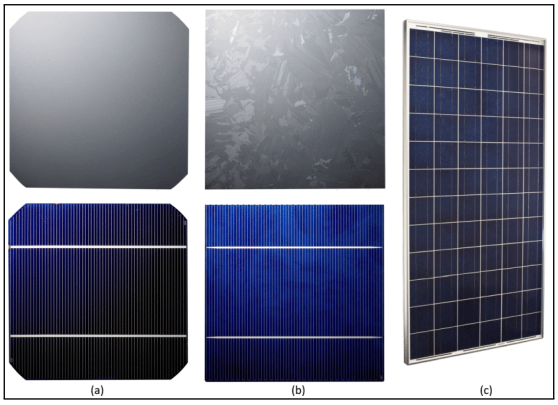
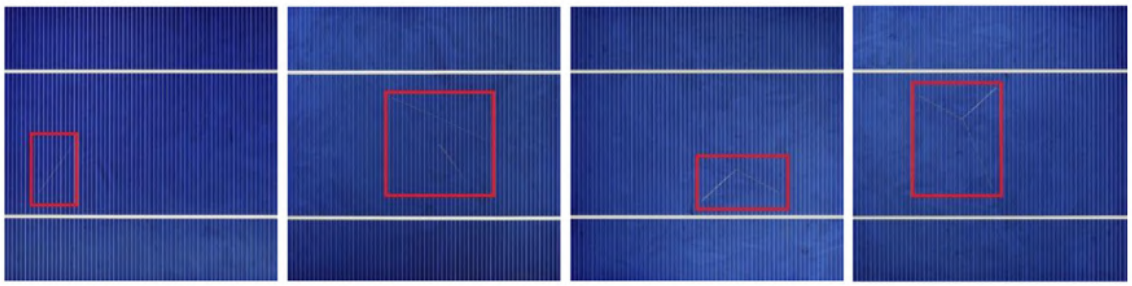


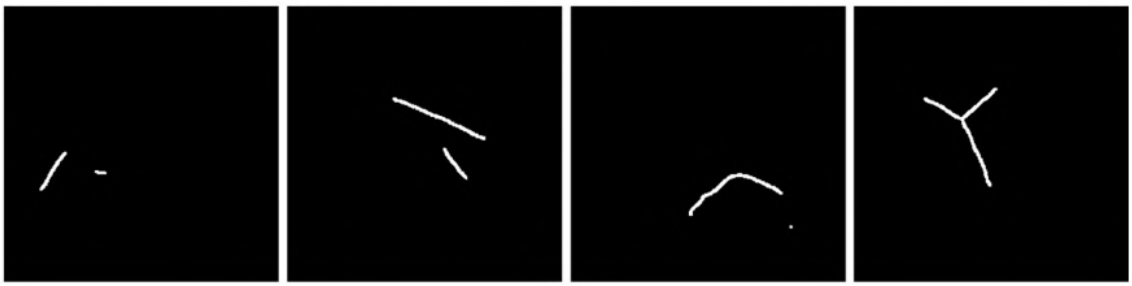
Figure 3.1: (a) mono-c-si wafer (top) and optical image of a finished solar cell (bottom) with rounded corners, (b) multi-c-si wafer (top) and optical image of a finished solar cell (bottom) with square corners, and (c) PV module assembled from multi-c-si solar cells [Fisher *et al.* 2012]

details and the results. There were no examples showing the results of the method for detecting cracks nor any indication of performance metrics.

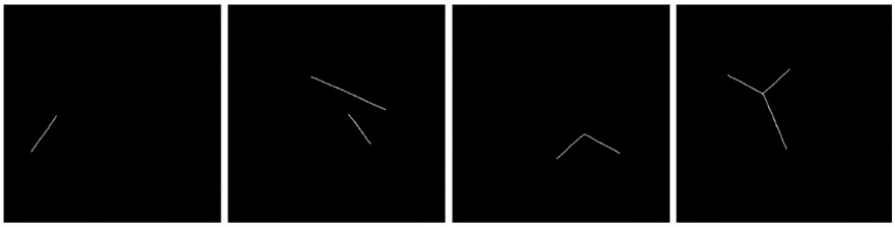
Qian *et al.* [2018] proposed a method for detecting cracks in optical images of finished solar cells using self-learning features. First, the optical image was denoised using a median filter and anisotropic diffusion, and the white-coloured busbars and gridlines were filtered out of the image based on average pixel intensities along the columns and rows. Second, an initial prediction of the crack was derived from a self-learning scheme using a series of feature templates. The feature templates were derived from a series of image patches and thus referred to as 'self-learning'. Third, a low-rank matrix recovery was applied to extract the preliminary estimate of the crack from the feature matrix. Fourth, the preliminary image was further processed using superpixel segmentation to locate the defects precisely. Finally, the image was processed using morphological operators to fine-tune the result. Figure 3.2 shows examples of well-localised crack detection as binary segmentation masks [Qian *et al.* 2018]. The method had a high recall (75%), low precision (21%), and an F-score of 32%, averaged over 80 test samples. While this method produces excellent binary segmentation masks for micro-cracks, it does not segment multiple classes.



(a) Original optical image



(b) Crack detection results



(c) Ground truth

Figure 3.2: Examples from Qian *et al.* [2018] showing the original images (a), the crack detection results (b), and the ground truth (c)

Two publications describe methods using acoustic techniques for defect detection in solar cells. Li *et al.* [2019] proposed a method using ultrasonic lamb waves to detect cracks, and Emirov *et al.* [2011] proposed a method using resonance ultrasonic vibrations to detect pinholes less than 1 mm in length that can lead to cracks. These methods may help detect small cracks in wafers and cells in the factory so that they may be segregated before further processing. However, the methods do not apply to defect detection in finished PV modules because they rely on detecting deviations in acoustic waveforms generated from individually tested wafers and cells.

3.2 SCDD - image classification

This section summarises the published literature for classifying EL images into two or more defect bins. Many papers rely on the ELPV benchmark dataset [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2018]. The classification models face challenges with multiple defects on a single image. EL images of solar cells often have multiple defects, which means some information contained in the EL image is lost when classifying a cell into one fault bin or another. Semantic segmentation, by contrast, performs pixel-level classification, so information about all the defects may be preserved.

The ELPV dataset was the first benchmark dataset for SCDD in EL images of solar cells [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2018]. The dataset consists of 2,624 EL images of solar cells, mono-c-si, and multi-c-si, with 0, 2, or 3 ribbon interconnects. Researchers working on classification and binary segmentation methods have widely cited this dataset. However, the dataset consists of largely outdated solar cell architectures cropped from only 18 unique mono-c-si modules and 26 unique multi-c-si modules [Demirci *et al.* 2019]. Notably, images from modules currently on the market are absent from the dataset because significant developments in solar cell engineering have occurred since publication in 2018. Specifically, modules made from half-cut cells using multi-ribbon interconnects are missing, and the adoption of this module type is growing rapidly in the market.

Banda and Barnard [2018] from Nelson Mandela University in South Africa proposed a deep-learning model for classifying solar cells into two bins: normal and defective. The experiments were conducted in Python using the Keras and Tensorflow libraries. Three model architectures were trained on 600 EL images and validated on 48 holdouts: a simplified LeNet [LeCun and others 2015], CifarCNN, and GoogleNet [Szegedy *et al.* 2015]. The test dataset consisted of 10 images. The LeNet and GoogleNet architecture achieved 98% binary classification accuracy on the training dataset.

Akram *et al.* [2019a] proposed a light convolutional neural network architecture for classifying solar cells from the ELPV dataset [Buerhop-Lutz *et al.* 2018] into normal and defective categories. The model consisted of four convolutional layers and two fully connected layers. The researchers trained several convolutional neural networks, including VGG 19, VGG 16, VGG 13, and VGG 11 [Simonyan and Zisserman 2014], using OpenCV, NumPy, TensorFlow, TFLearn, and PIL libraries in Anaconda Spyder (Python) integrated development environment. The classifier achieved 93.0% accuracy when classifying into two categories.

Sun *et al.* [2018] proposed a simple CNN to classify EL images into one of five categories: defect-free, linear crack, cross crack, flaky crack, and broken crack. A shallow CNN with

five layers (two convolutional layers, two pooling layers, and one fully connected layer) was trained on a dataset with augmentation. The authors report a classification accuracy of 98.4%. The article does highlight the challenge of classification in the presence of many different types of defects. In this study, only different kinds of cracks were detected, and many of the images had more than one type present with associated probabilities. For example, images with broken cracks also contained linear cracks. This result motivates the need for object detection using semantic segmentation to identify multiple defects in a single image.

Deitsch *et al.* [2019] proposed an SVM and a deep CNN method for classifying solar cells into four categories of defect probabilities. The methods were developed on the ELPV dataset. The SVM was trained on hand-crafted features derived from popular combinations of keypoint detectors and feature extractors from the literature. The SVM included masking to remove the background features of the cell, but the authors note this has only a marginal impact on model performance. The CNN was based on the VGG19 network with pre-trained weights from ImageNet and tuned to predict defect probabilities for four classes for each cell: 0%, 33%, 66%, and 100%. The output included a module-level map with defect probabilities for each cell. The accuracy of the SVM classifier achieved 82% accuracy, while the CNN accuracy achieved 88% due to the improved performance when classifying multi-c-si cells. Both models performed equally well on mono-c-si cells.

Demirci *et al.* [2019] proposed deep CNNs for classifying solar cells from the ELPV dataset into the four defect classes included in the dataset. The researchers investigated AlexNet, GoogleNet, MobileNetv2, and SqueezeNet architectures with transfer learning. The classification accuracy on the test dataset was the lowest for SqueezeNet (76%) and highest for AlexNet (97%).

Karimi *et al.* [2019] investigated three machine learning algorithms to classify EL images of solar cells into three categories over a series of accelerated stress tests in a damp heat environment. The researchers trained an SVM, a Random Forest, and a CNN on EL images of cells cropped from 15 modules, three modules from each of five different models, and downsized to 50×50 pixels. The output categories were 'good,' 'cracked,' or 'corroded.' The CNN performed best for this classification problem, achieving an F-score of 98%.

Mantel *et al.* [2019] demonstrated the accuracy of SVM and RF models for classifying solar cells into four categories: finger failures and three types of cracks based on their respective severity levels (A, B, and C). The dataset was collected from EL images on 982 multi-c-si modules imaged at night time in the field. Experts identified and annotated 1,847 faults to serve as a ground truth. After pre-processing the images, two region extraction methods used to identify potential defective regions within the cell were implemented. The first method was based on a Hough approach to identify potential regions with cracks. The second method was based on the percentiles of pixel intensities to identify potential regions with finger failures. 25 features were extracted for each region of interest, and the two machine learning models were trained. The best results were achieved by SVM using the percentile method for region extraction, which gave an accuracy of 99.7%, a recall of 27.4%, and a precision of 2.8%. The high accuracy, low recall, and low precision can only be explained by the fact that 669,201 regions were extracted from the dataset, and roughly 99.7% of those regions were predicted to be without fault.

Tang *et al.* [2019] presented a method for data augmentation first and then designed a CNN-based model for defect detection and classification in EL images of mono-c-si cells. 1,000 EL images of solar cells were curated for this research and classified into one of four categories: defect-free, mini-crack, finger interruption, or break. The data augmentation consisted of flipping, rotating, scaling, contrast variations, and noise disturbance. A simple CNN with six layers was trained for 40,000 epochs. The accuracy for mini-crack detection on the test dataset achieved 92%. The researchers also show that the accuracy of the simple model is similar to the accuracy of a model based on the VGG16 architecture, but the computation time is reduced by 96%. The data augmentation also improved the accuracy for each defect type. Specifically, the classification accuracy for mini-cracks increased from 85% to 94% with a six-fold increase in the sample size from data augmentation.

Tang *et al.* [2020] proposed a CNN with only eight layers for defect detection and classification into one of four classes: no defect, micro-cracks, finger interruption, and break. The research paper included a method for augmenting EL images using generative adversarial networks (GANs) to increase the number of training data available. The generator creates new EL images by adding random noise to the existing EL images, and the discriminator attempts to learn the difference between real and fake images. The network results in a more computationally intensive method compared to conventional data augmentation based on translations and scaling, but the experimental results show a significant improvement in classification accuracy. The test dataset was curated from ELPV and JinkoPower Company, a Chinese manufacturer of PV modules. The authors trained models using VGG16, ResNet50, Inception V3, and MobileNet architectures with and without data augmentation to classify images. The data indicate a significant improvement in classification accuracy from 47% for micro-cracks without data augmentation to 82% with the proposed data augmentation. The results also indicate that the model accuracy is equivalent to a deep-learning model based on a VGG16 backbone. The proposed CNN model with 12.9 million trainable parameters was 83% accurate, while the VGG16 model with 25,1 million trainable parameters was 82% accurate. The authors also highlight the importance of detecting multiple coexisting defects in future work.

Demirci *et al.* [2021] proposed a Deep Feature-Based (DFB) method for classifying solar cells into one of two or four classes. The experiments were conducted using the ELPV dataset containing 2,624 EL images of solar cells [Buerhop-Lutz *et al.* 2018]. First, feature vectors were obtained using DarkNet-19, ResNet-50, VGG-16, and VGG-19 pre-trained deep neural networks and combined. Next, a minimum Redundancy Maximum Relevance (mRMR) algorithm was employed to reduce the number of features used for classification. Finally, the feature vectors were classified with Decision Tree, Random Forest, K-Nearest Neighborhood, Naïve Bayes, and Support Vector Machines. Classification scores using the SVMs achieved 90.6% and 94.5% for the dataset in the 4-class and 2-class experiments, respectively.

Ge *et al.* [2021] proposed a novel architecture named Hybrid Fuzzy Convolutional Neural Network (HFCNN), which integrates fuzzy logic and convolution operations. The models were trained and tested on the ELPV dataset for classification into one of four defect probability classes. The authors trained two versions of a VGG8 and two versions of a CNN8 as the base cases. The proposed model replaces the conventional convolutional layers with a fuzzy convolution block. The experimental results indicate a small improvement in classification accuracy and F1-score using the fuzzy convolutional block compared to the base models. The VGG16

and VGG19 model accuracies ranged from 86.3% to 87.4%, while the HFCNN model accuracies ranged from 85.9% to 88.2%. However, the models with fuzzy convolutional blocks haver fewer model parateters to train.

Huang *et al.* [2022] proposed an algorithm-based deep convolutional neural network (DCNN) pruning method called PSOPruner to classify cells into one of five categories. The pruning problem was solved by a particle swarm optimisation (PSO) algorithm. A ResNet18-based model was trained on the EL images from the ELPV dataset. The PSOPruner was applied sequentially over several rounds to simplify the model regarding trainable parameters and floating point operations per second (FLOPs) to reduce training time without compromising accuracy. The accuracy for classifying finger interruptions and cracks decreased by a few percentage points after five pruning rounds, while the classification accuracy increased by double digits for cell disconnection and material defects. The authors also show that the PSOPruner can reduce model complexity when applied to other deep-learning models without compromising accuracy and f1-scores. The researchers demonstrate that the PSOPruner can find a pruning scheme to construct an extreme light network for automatic PV defects classification on mobile and embedded devices. However, the supporting evidence is based on training time, not prediction time. While the computational resources for training models are a concern, deploying a fully-trained prediction machine on embedded devices is not likely a significant limitation from the computational complexity standpoint.

Tang *et al.* [2022] proposed an algorithm combined with traditional image processing technology, deep learning, transfer learning, and deep clustering to classify solar cells into one of five categories. The algorithm can also automatically recognise unknown or unlabelled defects in the original dataset and update the model when a sufficient number of examples are collected. Cells are classified as unknown defects if the probability of belonging to any other known defect class is below a predetermined threshold. Then the unknown defects are then grouped into similar classes based on a k-means clustering algorithm. Five different CNNs were trained on EL images collected from a drone during an aerial inspection of a 100 MW PV plant in Gonghe, Qinghai Province, China, to detect normal cells, cracks, finger interruptions, black cores, and busbars corrosion. Unfortunately, the details on how the aerial images were captured and how the cells were cropped from the module-level images are missing. The optimised CNN was trained and evaluated after a 10-fold cross-validation experiment. The mean classification accuracy for all four defects was 93% or higher. Next, the researchers trained the model without the busbar corrosion or black core defects. Busbar corrosion and black core defects were viewed as unknown defects by the model for this experiment and classified by k-mean clustering. Unfortunately, the researchers do not include any measure of classification accuracy for the unknown defects.

Zhao *et al.* [2023] proposed an algorithm for SCDD in EL images based on a high-resolution network (HRNet) with an identification algorithm adapted to the image model, called the self-fusion network (SeFNet), in place of the conventional classification layer. A GAN was also implemented for data augmentation of the original ELPV dataset. The HRNet generates multiple feature maps of different resolutions in parallel using convolutions, downsampling, and upsampling. The SeFNet and an improved asymmetric convolution were developed to fuse the feature maps generated at different resolutions by the HRNet to improve the final classification. The researchers presented the results of the classification accuracy for mono-c-si and multi-c-si

cells separately and compared them against classifiers based on ResNet18, ResNet50, VGG16, and InceptionV3. The classification accuracy using HRNet and SeF-HRNet was significantly higher for mono-c-si cells compared to multi-c-si cells and significantly higher than the benchmark models for mono-c-si. The classification accuracy for multi-c-si cells was fairly equal, ranging from 79% to 82%.

3.3 SCDD - object detection and localisation

This section summarises defect detection and localisation research using bounding box methods on EL images. These methods are an improvement over image classification because they enable the identification of multiple defects in a single cell, which preserves more information contained in the EL image compared to classifying an image to a single bin or defect. Bounding boxes also enable some estimation of defect size, which may be important when correlating EL images to PV module power output.

Zhang *et al.* [2020b] proposed a method for defect detection and localisation with bounding boxes. The method fuses the output of two models: the Region Proposal-based object detection framework (R-CNN) and the region-based, fully convolutional networks (R-FCN). The fused model improves the object detection rate while maintaining high-accuracy positioning by utilising feature maps that are more sensitive to the location. Each model produces a set of bounding boxes, and the box is kept and improved if it appears in both models with a minimum IoU score. The models were trained on a private dataset generated for this research to detect broken cells, cracks, and unsoldered areas. The proposed method generally improved the detection accuracy compared to using either of the models independently by reducing the false negative rate of the R-CNN and the false-positive rate of the R-FCN. The detection accuracy rates was 91.3% for faster R-CNN, 94.7% for the R-FCN, and 98.3% for the fused model averaged across all four classes. The authors point to estimating the size of defects for future work.

Su *et al.* [2020] proposed a novel Complementary Attention Network (CAN) for adaptive background feature suppression and defect features highlighting. The CAN is formed by connecting the novel channel-wise attention subnetwork with the spatial attention subnetwork sequentially. The CAN is embedded into the region proposal network of a faster R-CNN (convolution neural network) to extract more refined defective region proposals. The model was developed and deployed to detect cracks, finger interruptions, and the 'black core' defect in the production line to remove the defective cells before module assembly automatically. The model was trained and tested on 3,629 EL images using AlexNet, VGG16, ResNet50, and DenseNet121 architectures. The classification effectiveness was evaluated on finger interruption patches of 128 × 128 pixels only. The experimental results showed a small increase in the F1-score of 1-2 percentage points for each of the four architectures evaluated. In one example figure, the top 50 proposal regions using the embedded CAN network appeared to align with the crack defect more effectively than the top 50 proposal regions without the CAN.

Zhao *et al.* [2021] developed a deep learning-based automatic detection method for multitype defects in photovoltaic modules with application in a real PV module production line. The team built a database of 5,983 labelled EL images of 144 half-cut cell modules with 9 ribbon interconnects and the corresponding ground truth images with 19 defects created using La-

belMe software. Of the 19 defects identified, the researchers focused on the 14 defects with the highest occurrence in the library. Over-sampling techniques were implemented to handle the class imbalance, and data augmentation was implemented to increase the dataset size. A Mask R-CNN with ResNet-101-FPN backbone was trained for 300,000 iterations in the Detecton2 platform. The mean average precision over all IoU thresholds (mAPall) was used for evaluating the model performance. However, only AP values on bounding boxes with at least 50% accuracy were reported. The AP values for the pixel-level segmentations were not reported. The best model was tested in the production for automatic screening after specific thresholds were determined. For example, finger interruptions were not used to reject modules because the frequency of occurrence was high, and the impact on cell efficiency and appearance was deemed negligible. Rejecting modules based on finger interruptions would have resulted in a significant reduction in the production line output. The researchers also measured the 'compensation rate' to quantify modules in which some of the defects were missed but compensated by other correct predictions in the same module. The model was deployed on the production line for three days, and the normal/defective binning process based on the model was compared with the human-based approach. The F1-scores achieved 95%, 96%, and 97% over the three days. The researchers also noted that the 'copy' method for data augmentation was detrimental and that scratches and finger interruptions were the two most difficult defects to detect accurately.

3.4 SCDD - binary segmention

This section summarises research into binary segmentation of EL images. The researchers focus on a limited number of defects. Crack detection received the most attention, while two articles also included gridline defects. These methods improve upon the object detection methods by providing pixel-level classification for a small number of critical defects. However, these methods segment only one defect class or a conglomeration of defect classes from background pixels.

Tsai *et al.* [2012] proposed a method based on a Fourier image reconstruction technique to detect small cracks, breaks, and finger interruptions in EL images taken on bare cells before lamination. The method produces spatially accurate representations of cracks, breaks, and gridline defects (Figure 3.3) by differencing the original EL image and the reconstructed image from the spectral representation of the original with the defect removed. A statistical control limit is used to segment defects from the differenced images. However, the method does not differentiate between the different defect types. The method segments a conglomeration of defects from background noise using binary segmentation.

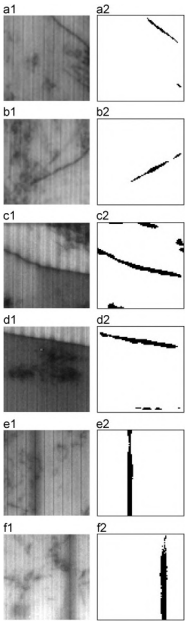


Figure 3.3: EL images and predictions for (a,b) Type A cracks, (c,d) Type B cracks, and (e,f) gridline defects [Tsai *et al.* 2012]

Anwar and Abdullah [2014] proposed an algorithm featuring an improved anisotropic diffusion filter and advanced image segmentation technique for detecting cracks in EL images of multi-c-si cells. The filtering was based on the Fourier transformation and reconstruction, similar to the work by Tsai *et al.* [2012]. The binary image segmentation included two stages of thresholding to enhance the crack further. The recall, or 'completeness' in the words of the authors, averaged 71.9% across the test dataset. The precision, or 'correctness', averaged 4.6% across the test dataset. Anwar and Abdullah [2014] used the resulting shapes to train a Support Vector Machine (SVM) to classify cells as cracked. Figure 3.4 shows results for four selected EL images using various segmentation techniques. The classification sensitivity (recall) and specificity (precision) averaged 97.7% and 80.2% over 600 EL images using the proposed method.

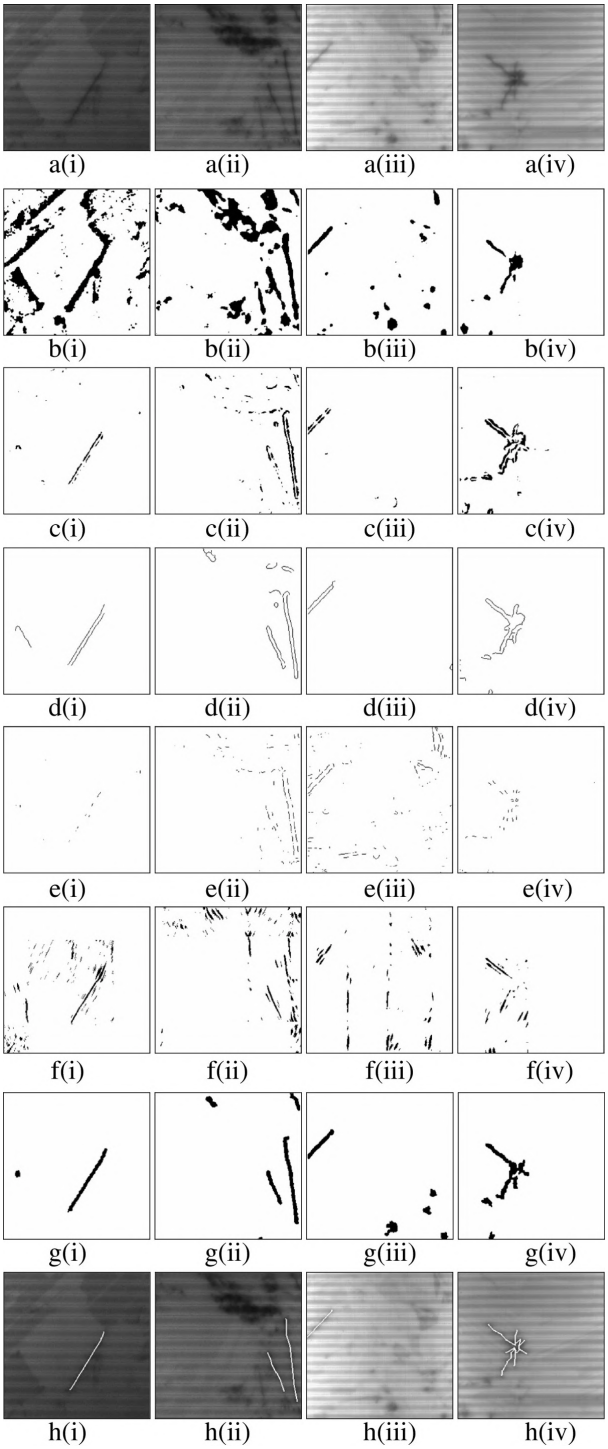


Figure 3.4: Image segmentation results comparing the proposed and standard segmentation techniques. (a) Original image, (b) Otsu's thresholding, (c) Sobel edge detector, (d) Canny's hysteresis, (e) LoG filter, (f) FIR, (g) the proposed method, and (h) ground truth images. [Anwar and Abdullah 2014]

Spataru *et al.* [2016a] proposed a method for detecting and classifying microcracks in EL images of solar cells using two-dimensional matched filters. The method also provides a means

to estimate crack lengths and orientations. Multiple microcrack templates were convolved with the EL image to find the best-matched filter for a given crack segment. The convolutions were aggregated for all crack regions to produce a binary segmentation for the whole cell, estimate the shape of the cracks, and estimate the length of the cracks. The method was evaluated on one multi-c-si module with 60 cells and shown to detect 90% of cracks larger than 3 cm, with few false positives. Figure 3.5 shows a module-level EL image with an overlay of the crack defects, coloured by crack type as defined by the authors. While this paper focuses on cracks only, it presents a module-level view of the output and includes some statistical analysis by generating a histogram of the number of cells with a given crack length.

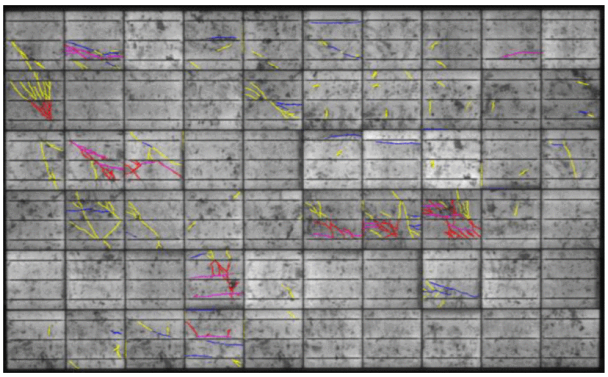


Figure 3.5: Example of micro-crack detection and identification in a standard 60-cell mc-Si PV module: red - dendritic or cracks with multiple orientations; blue - cracks parallel to the busbars; magenta-parallel branches in a dendritic crack; yellow - other types of cracks [Spataru *et al.* 2016a]

Chen *et al.* [2019] proposed a robust crack detection and classification method using steerable evidence filtering in EL images of multi-c-si solar cells specifically to handle the grain boundaries that occur in normal multi-c-si cells. The researchers differentiated 'pure' cracks that occur away from grain boundaries and cracks that are 'submerged' by grain boundaries. The crack detection scheme included three steps: crack saliency map generation by steerable evidence filtering, segmentation-based crack extraction by local threshold and minimum spanning tree, and finally, crack location by skeleton extraction. The dataset consisted of 372 defective solar cells and 200 defect-free solar cells imaged at the Tianjin Yingli solar cell production line. The performance of the proposed method compared favourably to previous work from Chiou *et al.* [2011], Tsai *et al.* [2012], and Anwar and Abdullah [2014], especially for 'submerged' cracks and cracks in dark EL images. Figure 3.6 shows the predictions for six EL images for all evaluated methods and ground truth. The F1 scores ranged from 91% to 98.6% depending on the type of crack. The SEF method show excellent predictions for the cracks connected throughout the length of the crack, even across grain boundaries, and not dilated. False positives were rare, leading to high precision and, ultimately, high F1 scores.

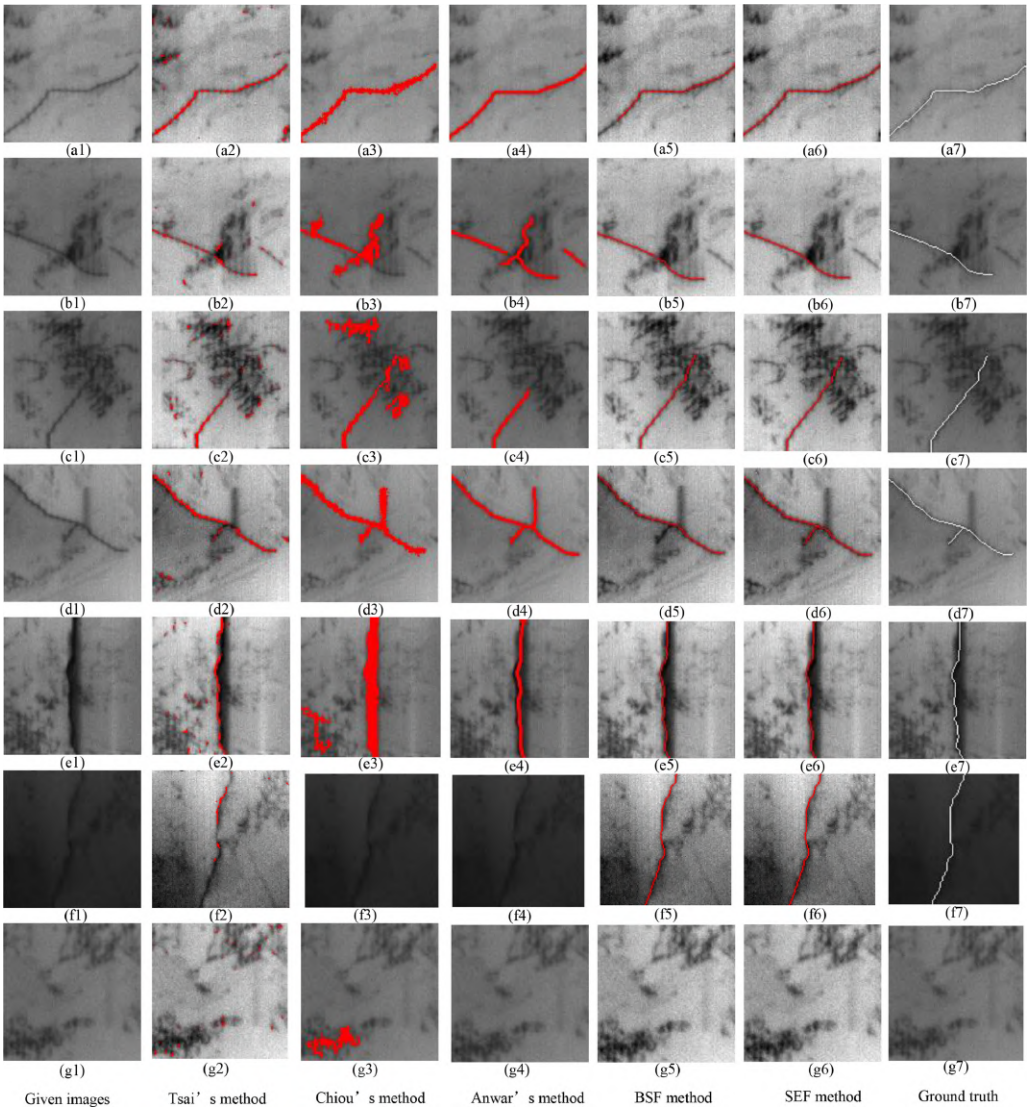


Figure 3.6: Crack predictions for six EL images and all the binary segmentation methods evaluated [Chen *et al.* 2019]

Dhimish and Mather [2019] demonstrated an EL defect detection system integrated into a cell manufacturing line. The finished cells were fed to the EL imaging station using robotics, and a Labview program analysed the output image to accept or reject the cell based on detected cracks. The method compares the EL image of the cell to a reference cell with no defects at each point on a grid and assigns a '1' if they are different and a '0' if they are the same. The resulting binary image was used to evaluate the pass/fall decision. While the method was intended to detect cracks in the cells, the example images indicated that the method was effective at detecting any defective region that resulted in a dark area in the EL image of the cell compared to the reference cells. In one example, a crack was evident in the same region as a 'belt mark' defect. The binary image output captured more of the belt mark defect than the crack (Figure 3.7), which was the target of the detection method. In any case, this method only works with a corresponding reference cell with identical cell architecture.

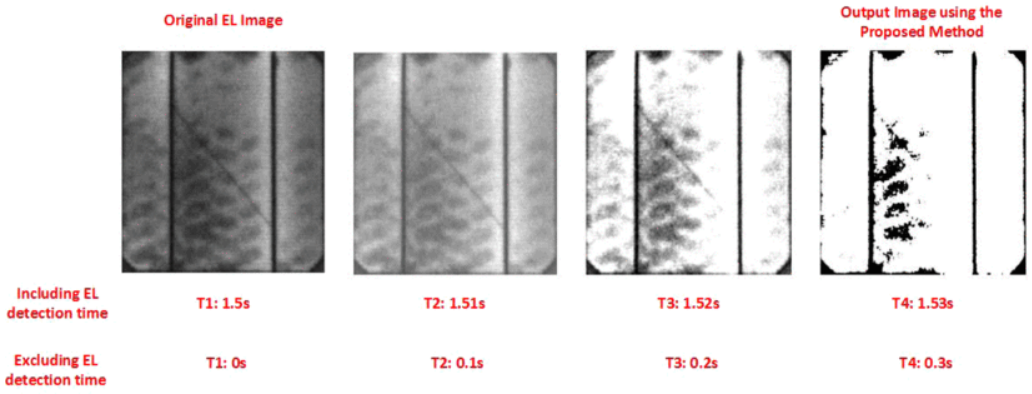


Figure 3.7: EL image of a solar cells with a crack and belt mark defect taken a three different time exposures and the final binary prediction that only poorly captures the crack [Dhimish and Mather 2019]

Mayr *et al.* [2019] proposed a weakly supervised learning strategy that only uses image-level annotations to obtain a method that is capable of segmenting cracks on EL images of solar cells. A modified ResNet-50 was trained on the ELPV dataset after the researchers added image-level annotations of cracks only. During the classification process, EL images were translated to heat maps used to identify cracks, and those heat maps also served as a coarse binary segmentation of the crack. The best-case classification model achieved only 94%, but the binary segmentation maps were coarse at best. Figure 3.8 shows the input image, intermediate heat maps, and predictions of two selected images with cracks.

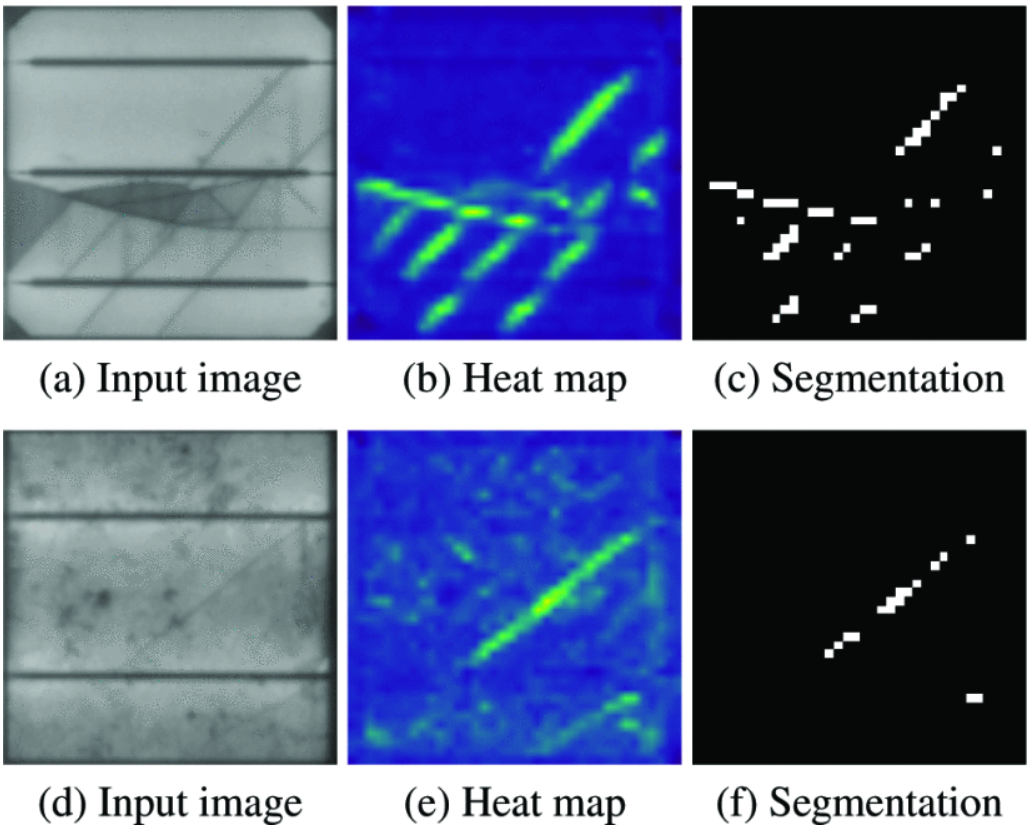


Figure 3.8: Input image, intermediate heat maps, and predictions of two selected images with cracks [Mayr *et al.* 2019]

Rahman and Chen [2020] proposed a modified U-Net that incorporates a multi-attention network (MAU-Net) to efficiently extract the most important features and neglect the nonessential features during training. The channel attention network emphasises defect regions, while the spatial attention network suppresses the background noise. The model was trained with a hybrid loss function to detect cracks and gridline defects in EL images of mult c-si cells before lamination into PV modules. The dataset consisted of 828 images and corresponding binary masks as the ground truth. Figure 3.9 shows the predictions for four selected images from several methods evaluated, including the MAU-Net. The experimental results produced excellent binary predictions for cracks and gridline defects aggregated together as a 'defect' class but could not differentiate one from the other. The background noise from grain boundaries and ribbon interconnects were also effectively eliminated from the predictions. The MAU-Net showed a significant increase in the mIoU compared to previous work and a significant improvement over the original U-Net. The mIoU was 60.9% for the original U-Net and 69.9% for the MAU-Net. There are several key differences between Rahman and Chen [2020] and this thesis. First, the EL images come from a single production line, which means the images have a greater similarity. Second, the EL images were taken before lamination so no glass or encapsulant is obscuring the image, and there is no need to crop images from EL images of full modules. Third, the model was trained to detect only two defects, and they were lumped together as one during segmentation. Fourth, the model removes the features from the seg-

mentation map, resulting in a binary segmentation. However, there may be some advantages to incorporating the MAU-Net into our model trained for multi-class semantic segmentation.

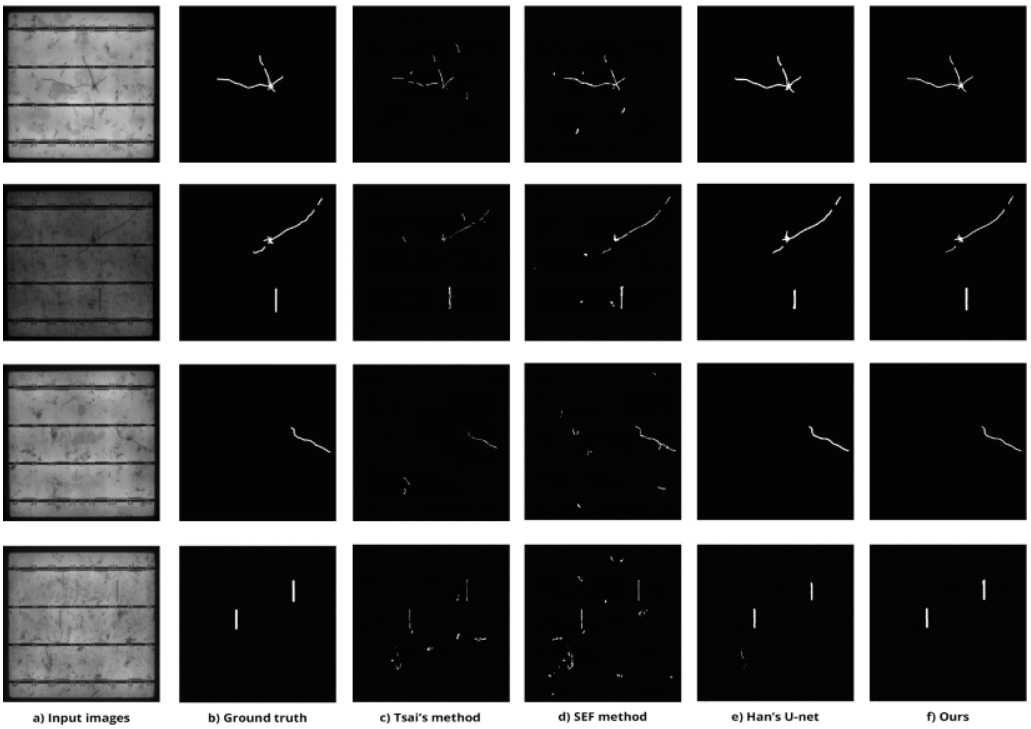


Figure 3.9: Predictions for four selected images from several methods evaluated for binary segmentation, including the MAU-Net [Rahman and Chen 2020]

3.5 SCDD multiclass segmentation using Semantic Segmentation

This section covers general research on multi-classification using semantic segmentation (SemSeg) and two papers specifically that applied SemSeg on EL images of solar modules and cells. One paper evaluated SemSeg on EL images of thin film modules [Sovetkin *et al.* 2021], while another evaluated SemSeg on EL images of crystalline-silicon solar cells [Fioresi *et al.* 2020]. Sample images from both papers are included in Chapter 5. The application of SemSeg on EL images of solar cells evolved naturally from image-level classification beginning in 2018 to object detection with bounding boxes to binary segmentation to pixel-level multi-class segmentation in 2022. This thesis advances and enables further research using SemSeg for SCDD in EL images.

The U-Net for semantic segmentation was first introduced in 2015 and evaluated on biomedical images of cell membranes [Ronneberger *et al.* 2015]. The U-Net model includes a series of blocks comprised of convolutional layers and max pooling layers that encode or learn the features in the image. The U-Net then decodes the result from the encoder and upsamples the output to recover the original image size with pixel-level classification.

SemSeg, more broadly, has been used to evaluate road scenes [Badrinarayanan *et al.* 2017], brain tumors [Dong *et al.* 2017], the cityscapes dataset [Chen *et al.* 2018a], MRI images of the brain [Cui *et al.* 2018], the common objects in context (COCO) dataset [He *et al.* 2017], advanced scanning transmission electron microscopy (STEM) images of steel [Roberts *et al.* 2019], and other datasets [Yuan *et al.* 2020]. Zhang *et al.* [2020a] published a method for detecting thin cracks in images of pavement combining U-nets and GANs focused on addressing the challenges in detecting thin cracks with mainly background pixels that could prove helpful for crack detection in EL images.

Sovetkin *et al.* [2021] evaluated a series of image segmentation methods based on deep neural networks in order to perform semantic segmentation of EL images of thin-film modules. The researchers trained various combinations of encoder-decoder models with pre-trained weights from ImageNet to detect shunts and droplets in the EL image of copper indium gallium diselenide (CIGS) thin-film modules. Shunts and droplets are common defects in thin film modules. The selection of encoders consisted of Mobile-Net, ResNet, VGG-net, and U-Net. The decoders consisted of PSP-net, SegNet, U-net, and FCN-net. The researchers identified two separate models, one for each defect type. The precision and recall for the shunts achieved 0.35 and 0.55, respectively. The precision and recall for the droplets achieved 0.61 and 0.68, respectively. The researchers also presented a correlation plot between the electrical performance of the modules versus the number of detected shunts. Shunts are expected to change the shape of the I-V curves, specifically the slope of the I-V curve at short-circuit current (Isc). However, the scatterplot of the slope at Isc and the number of shunts resembled a shotgun blast with no correlation. A heatmap of all 6000 segmentation maps for droplets was presented, and a clear arc-shaped pattern can be seen over the batch. The results show that the systematic segmentation of large images can reveal subtle features that cannot be inferred from studying individual images.

Fioresi *et al.* [2020] reported results on EL images from a semantic segmentation model based on a ResNet-101 backbone in a one-page summary. The model segments two defects in EL images: cracks and contact defects. The initial work was conducted under the Research Experiences for Undergraduates (REU) program in the USA. The researchers created a private dataset with annotations for 14 defects, three related to cracks and five related to contact/gridline defects. The dataset was later published along with Fioresi *et al.* [2022]. See Chapter 5 for more details and some examples of the annotations. Notably, the annotation method created 'conglomerations' of defects. For example, multiple gridline line defects may be aggregated into one larger block, or a large cell area may be marked, not just the crack line. This method leads to larger defects that should be easier to detect. The model achieved F1-scores of 95%, 75%, and 56% for no-defect, crack, and contact defects, respectively.

Fioresi *et al.* [2022] published a journal article with results from the semantic segmentation model initially reported under the Research Experiences for Undergraduates (REU) program in the USA Fioresi *et al.* [2020]. This article has the same lead author as the 2020 publication [Fioresi *et al.* 2020] but different co-authors and very little new content. The model segments two types of solar cell defects in electroluminescence images: cracks and contact defects. The authors trained a fully connected network (FCN) with a ResNet-101 backbone. They achieved a weighted f1-score of 96%, a pixel-level global accuracy of 95.7%, and a mean intersection over union (IoU) score of 75.8%.

3.6 Semi-Supervised Learning and Self-Supervised Learning

This section summarises recent publications on semi-supervised (semi-SL) and self-supervised (self-SL) learning methods for semantic segmentation. These methods have been shown to improve predictions in other domains compared to fully-supervised models. However, none of the published articles on semi-SL and sefl-SL for semantic segmentation focus on applying these methods to SCDD in EL images.

Chen *et al.* [2020a] presented SimCLR: a simple framework for contrastive learning of visual representations with experimental results based on the ImageNet dataset. SimCLR learns representations by maximizing agreement between differently augmented views of the same data example via a contrastive loss. The researchers used random cropping, colour distortions, and Gaussian blur for data augmentation and a ResNet as the base encoder in this work. The contrastive loss is not applied to the network representations but rather to the output of a small multi-layer perceptron (MLP) network that takes the network representations as inputs. For all positive pairs of augmented representations in the mini-batch, the training algorithm computed the normalized temperature-scaled entropy loss and updated the weights to minimize the contrastive loss. The experimental results showed that SimCLR classification accuracy achieved 69.3% and outperformed other self-supervised methods using the ResNet50 encoder by at least 5.5% absolute. SimCLR classification accuracy achieved 76.5% using a ResNet-50 (4x) encoder and outperformed other methods based on different encoders by a similar margin. However, the researchers stressed that the choice of augmentation composition methods, the project head, the batch size, and the number of iterations were all critical to achieving good results.

In He *et al.* [2020a], researchers from Facebook AI Research (FAIR) presented Momentum Contrast (MoCo) for unsupervised visual representation learning. From a perspective on contrastive learning as a dictionary look-up, the researchers built a dynamic dictionary with a queue and a moving-averaged encoder. In this context, unsupervised learning trains encoders to perform a dictionary look-up: an encoded "query" should be similar to its matching key and dissimilar to others. A query matches a key if they are encoded views (e.g., different crops) of the same image (positive pairs). The negative pairs corresponding to the query are stored in the queue. The keys are continuously updated with new keys added to the queue while the oldest key drops out of the queue. The loss function includes both the positive and negative pairs. The momentum parameter smoothes the learning rate as the updates are averaged over training updates. The MoCo was evaluated and compared against other self-supervised models for a classification task and object detection. The models were pre-trained according to each method and frozen. Then the features were frozen and used to train a linear classifier. The MoCo model achieved best-in-class accuracy for each model size evaluated, from the smallest model with 24 million parameters to the largest model with 375 million parameters. MoCo also outperformed a fully supervised model for object detection on the ImageNet dataset by 4.2 percentage points on the AP and by 7.4 percentage points on the AP75.

In Chen *et al.* [2020c], researchers from Facebook AI Research (FAIR) verified the effectiveness of two of SimCLR's design improvements by implementing them in the Momentum Contrast (MoCo) framework for self-supervised learning. Specifically, the researchers con-

firmed improvement in classification accuracy for the MoCo framework by adding the MLP projection head and stronger data augmentation. They also showed that the large batch sizes required for SimCLR training (4000-8000 images per batch) could be greatly reduced to 256. The researchers showed that the improved model MoCo V2 achieved 71.1% classification accuracy on ImageNet compared to the original MoCo V1, which achieved 60.6% classification accuracy after 200 epochs. After extended training, MoCo V2 achieved 71.1% classification accuracy versus 69.3% for SimCLR.

Purushwalkam and Gupta [2020] presented quantitative experiments based on analysis of ImageNet and MSCOCO to demystify recent gains in self-supervised learning models owing to data augmentation and access to clean object-centric datasets like ImageNet. The results showed that self-supervised models had improved occlusion invariance compared to supervised models, i.e., the model classification accuracy was more robust on images in which part of the object is hidden behind some other object in the image. The other five variants tested were worse for the self-supervised models. The researchers posit that the improved occlusion invariance was likely due to the random cropping applied during data augmentation when generating positive pairs because the model learned to classify both crops as the same even though the crop may represent a different part of the entire object. The object-centric nature of ImageNet, as opposed to a more scene-centric dataset, was also found to be critical in the success of self-supervised models. For example, random crops from the image of a chair would all likely contain some portion of the legs, the seat, and the back support, assuming a reasonably large crop, so these positive pairs would have similar features that are sensible to compare using a contrastive loss function. However, random crops from a scene, such as a kitchen, may include a kitchen table, a stove, or a refrigerator. The model using contrastive loss would be forced to view these objects as the same and learn the wrong weights. The researchers posit that aggressive cropping is harmful in object discrimination and does not lead to the right representation learning objective unless trained on an object-centric dataset.

Ouali *et al.* [2020a] presented a novel method for cross-consistency training (CCT) using semi-supervised learning for semantic segmentation. A shared encoder and a main decoder are trained in a supervised manner using the available labelled examples. For the unlabelled examples, a consistency between the main decoder's predictions and those of the auxiliary decoders was enforced over different types of perturbations applied to the inputs of the auxiliary decoders. The effectiveness of consistency training depends heavily on the behavior of the data distribution, i.e., the cluster assumption, where the classes must be separated by low-density regions, which is why the researchers proposed to enforce the consistency over different forms of perturbations applied to the encoder's output and not the input images. The auxiliary decoders have a negligible number of parameters compared to the main decoder. The authors also show that enforcing consistency across different domains can push the model toward better generalisation, even in the extreme case of non-overlapping label spaces. The model was trained by alternating across multiple domains at each iteration by sampling from one domain for an iteration and then sampling from another domain for the successive iteration. The domains are only partially or fully non overlapping labels. The experiments were conducted using a ReNet50 pretrained on ImageNet and a PSPnet. The PASCAL VOC dataset was used for training. The results indicated improved mIoU when using the CCT model compared to baseline, especially when the number of labelled data is 20. As the number of labelled data increased, the mIoU increase for all models and the gains over the fully-supervised baseline model decreased. With

n = 100 labelled, images, the mIoU measured 2-4% above the baseline, depending on the type of perturbation applied.

Wang *et al.* [2022a] proposed a semi-supervised semantic segmentation framework for Using Unreliable Pseudo-Labels (U^2PL) by including unreliable pseudo-labels into training. The method classifies the pixels from pseudo-labels for unlabelled images as reliable or unreliable. The pseudo-labels are the pixel-level labels from the prediction model, so the pseudo-labels become more accurate as the training evolves. The reliable pseudo-labels are those located away from class boundaries, while unreliable pseudo-labels are found at the borders. The researchers utilised the entropy of every pixel's probability distribution to filter high-quality pseudo-labels for further supervision. If the pixel level entropy was less than an adaptive threshold, then the pixel was labelled reliable, otherwise unreliable. The unreliable pseudo labels were stored in memory banks, one for each class, and used as negative samples for the class to improve learning. The unreliable labels were included in the training at a rate of 20% initially and gradually decreased with each training iteration. As training evolves, the pseudo-labels become more reliable, so fewer unreliable pseudo-labels are necessarily included in the training. The experiments were conducted on PASCAL VOC 2012 with 20 semantic classes of objects and Cityscapes using ResNet-101 pre-trained on ImageNet [11] as the backbone and DeepLabv3+ [6] as the decoder. Evaluations were measured on the mIoU of the validation dataset. The researchers compared the U^2PL results with four semi-supervised models and a fully supervised model, all trained with 92, 183, 366, 732, and 1464 labelled images in the training dataset. The mIoU for the U^2PL model trained on the classic PASCAL VOC 2012 measured approximately 22%, 14%, 8%, 5%, and 7% higher than the fully supervised model as the number of labelled images in the training datasets increase. The U^2PL outperforms the next best semi-supervised model by 11%, 3%, 4%, 2%, 3% as the number of labelled images increases. On the blender PASCAL VOC 2012 validation dataset, modest gains between 1-3% were achieved compared to the next best semi-supervised model under the different partition protocols. This dataset was much larger, so the smallest partition included 662 labelled images, up from the 92 labelled images used for training on the classic PACAL dataset. On the Cityscape dataset, modest gains of 0.1-1% were reported for U^2PL compared to the next best semi-supervised model. Ablation experiments indicated that including unreliable labels improved mIoU by 1-2% compared to the U^2PL without pseudo-labels. The qualitative examples showed improvement in segmenting border areas between classes, specifically around the long, narrow defects such as the leg of humans on bikes or horses and the dog's tail compared to the fully-supervised model.

3.7 Vision Transformers

Vision transformers (ViT) emerged in recent years as an alternative architecture to convolutional neural networks (CNNs) for computer vision applications using semantic segmentation. Thisanke *et al.* [2023] reviewed and compared the performance of many ViT architectures designed for semantic segmentation using benchmarking datasets such as ADE20k, Cityscapes, and PACAL-Context. ViTs for image classification, object detection, and semantic segmentation evolved following the success of transformers applied to natural language processing. However, ViTs rely on large datasets and carry high computational costs for accurate predictions. While ViTs were not evaluated as part of this thesis, future work in SCDD should evaluate this architecture for both fully-supervised and self-supervised models.

3.8 Conclusions

Computer vision methods have proven effective in classifying EL images of solar cells as normal or defective. Statistical methods, support vector machines (SVMs), convolutional neural networks (CNNs), and others have all been developed for classifying cells into a limited number of defect types typically focused on cracks, inactive areas, and gridline defects. Multiple themes were detected from the literature review.

First, the statistical methods and machine learning models proved effective for classifying solar cells into a limited number of bins. Still, most of the prior research on SCDD focused on the classification of cell-level images, not on object detection or defect segmentation.

Research was often carried out on private datasets. The publication of the ELPV dataset [Buerhop-Lutz *et al.* 2018] provided a benchmark for subsequent researchers working on classification methods to compare results. The publication of the UCF dataset in 2022 Fioresi *et al.* [2022] will hopefully have a similar impact on research into semantic segmentation.

Prior work using binary segmentation was successful [Tsai *et al.* 2012; Anwar and Abdullah 2014; Chen *et al.* 2019; Rahman and Chen 2020; Qian *et al.* 2018; Tsai *et al.* 2012]. However, most experiments were conducted on a small number of classes or defects, particularly cracks. Semantic segmentation for multi-class SCDD only emerged in 2021, but public datasets were still unavailable. Results from semi-supervised and self-supervised models for SCDD in EL images have not been published. In general, the performance gains from semi-supervised and self-supervised models are most significant when the labelled dataset is small. Data augmentation methods are critical in the performance of semi-supervised and self-supervised models.

Semantic segmentation promises to improve upon the earlier methods by generating pixel-level classification, which enables object detection, classification, and quantification of multiple defects that may be present in a single cell. The pixel-level classification preserves the maximal amount of information embedded in the EL image compared to image classification and object detection. With pixel-level classification, PV module quality can be analysed for statistical differences in batches from multiple suppliers of different PV modules and over time for the same batch of PV modules. This information would be useful during procurement, operations, maintenance, and warranty claims.

Chapter 4

General Methods

Chapter 4 summarises the general methods used in the subsequent chapters, which cover the experimental methods and results from this research. The sections include a description of the deep-learning model implemented, loss functions, class weights, performance metrics for model evaluation, statistical comparison of model metrics, and a section on correlating the output of the semantic segmentation models to PV module power output.

4.1 Deep-learning models evaluated for SemSeg

Deep-learning methods fall under the umbrella of artificial intelligence and within the broader family of machine-learning methods. They form a class of models that learn features from the data using multiple layers of mathematical abstraction. Those features are then mapped to the output, providing useful information extracted from the raw data.

The deep-learning approach differs significantly from earlier machine-learning models that relied on hand-crafted features created by a human analyst to extract useful information from the raw data. As an example related to SCDD, a hand-crafted feature may be a single threshold value derived to distinguish good versus defective solar cells when analysing histograms of pixel intensities [International Electrotechnical Commission 2018]. In deep learning, the models automatically learn features from the data that are used to predict pixels that are likely to be inactive areas or cracks without relying on simple, hand-crafted features such as threshold values for evaluating histograms. In this way, a deep-learning model can learn to detect multiple defects within a single image and assign each pixel to a specific defect or feature. SemSeg models are deep learning models based on fully convolutional neural networks (FCNs) that assign a prediction or label to each pixel in the input image. The FCNs are more complex models than the convolutional neural networks (CNNs) used to classify the entire image into one class or another with some degree of probability.

Fully-supervised semantic segmentation models

This section provides a brief description of the fully-supervised SemSeg models in general and additional details on the fully-supervised models used in this thesis: the U-Net model, Pyramid Scene Parsing Network (PSPNet), DeepLabV3, and DeepLabV3+. The U-Net model was trained and evaluated in Chapter 6. All four fully-supervised models were trained and

compared in Chapter 7. Fully supervised SemSeg models require pairs of images; one is the actual image of interest, and the other is a corresponding ground truth annotation. The pairs of images are used to train the models to optimise the loss function and test the models using performance metrics.

Fully-supervised SemSeg has been used to evaluate images in many different domains. Examples of fully-supervised SemSeg in the literature include analysis of road scenes [Badrinarayanan *et al.* 2017], brain tumours [Dong *et al.* 2017], the cityscapes dataset [Chen *et al.* 2018a], MRI images of the brain [Cui *et al.* 2018], the everyday objects in context (COCO) dataset [He *et al.* 2017], advanced scanning transmission electron microscopy (STEM) images of steel [Roberts *et al.* 2019], and other datasets [Yuan *et al.* 2020]. Zhang *et al.* [2020a] published a method for detecting cracks in images of pavement. The authors successfully combined U-Nets and GANs to address the challenges in detecting thin cracks with mainly background pixels.

The U-Net model is a widely-used semantic segmentation model initially developed and trained for biological image segmentation [Ronneberger *et al.* 2015]. The fully-supervised model consists of a contracting encoder and an expanding decoder that enable precise pixel-level segmentation. The U-Net model includes a series of blocks comprised of convolutional layers and max pooling layers that learn the features in the image. The U-Net then decodes the result from the encoder and upsamples the output to recover the original image size with pixel-level classification. The original paper has been cited over 60,000 times and applied in diverse fields. The authors demonstrate how the model is quick to train and can learn on a small sample of ground truth data. Their original training dataset contained only 30 images of white cells with black membranes. For these reasons, the U-Net was selected for the initial experiments on SCDD using SemSeg.

The Pyramid Scene Parsing Network (PSPNet) is another widely used semantic segmentation model that achieved state-of-the-art performance on benchmark datasets in 2016 [Zhao *et al.* 2017]. In this work, the images or scenes are more complex compared to the biological images used in the original U-Net paper. The scenes contain more classes, and the global scenes represent different contexts, such as an outdoor rural setting or an indoor bedroom. One of the unique contributions of PSPNet to semantic segmentation is its use of a pyramid pooling module to capture multi-scale contextual information. This module concatenates the feature map generated by a CNN from the original image with feature maps generated at different scales and resolutions, allowing the network to reason about objects and regions at multiple levels of detail. The authors claim that a boat is less likely to be incorrectly classified as a car if the model can utilise global scene category clues such as the presence of a boathouse and water in the scene. This model was included in the experiments to quantify the impact of pyramid pooling in SCDD. While the scenes do not vary in the same way as scenes in the natural world, the pyramid pooling may prove useful in distinguishing multi-c-si from mono-c-si cells.

DeepLabV3 and DeepLabV3+ models were developed by a research team from Google [Chen *et al.* 2018b]. These models promise an improved solution for SemSeg. DeepLabV3 combines both Atrous Spatial Pyramid Pooling (ASPP) to capture the contextual information similar to the PSPnet and the encoder-decoder structure similar to the U-Net to obtain sharp object boundaries. The atrous convolution applies a convolution filter that includes holes in the matrix with zero-valued weights to expand the receptive field and learn from a larger context. Atrous

convolution is often referred to as dilated convolution because the holes dilate the area of the filter. The DeepLabV3+ includes a simple extension to DeepLabV3 that the authors claim improves performance, especially along object boundaries. This separable convolution applies a standard convolution filter to each input channel separately. The DeepLabv3+ achieved a new state-of-the-art performance level on PASCAL VOC 2012 and Cityscapes datasets in 2018. The mIoU for PASCAL VOC 2012 attained 89% using DeepLabv3, compared to 85.4% using the PSPNet. The DeepLabV3 was included in the experiments to quantify the impact on border areas, which are crucial to accurate predictions when the objects have a long and narrow aspect ratio. However, it's unclear if the channel-wise convolution will add any value on greyscale images.

Self-supervised models

This section introduces self-supervised learning (Self-SL) models. Ground truth images are not required to train such models. Additional details on the models selected for the experiments described in Chapter 8 are also provided.

Self-SL is a class of algorithms in which the supervision signal or label is derived from the data. By contrast, fully-supervised models require a set of hand-labelled ground truth masks to provide the supervision signal, one ground truth mask for each image in the dataset. The Self-SL algorithms are easily scaled in terms of dataset size since they remove the requirement of the human annotation bottleneck required for fully-supervised learning. Contrastive learning and reconstruction are two common approaches to Self-SL in the computer vision domain. In contrastive learning, the model learns to discriminate between similar and dissimilar pairs of data points. In reconstruction, the model learns to reconstruct an image from a distorted or corrupted version of the original. Models from these two families were included in the experiments to quantify the impact of including the thousands of unlabelled images in the dataset.

The Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [Chen *et al.* 2020a] and Momentum Contrast (MoCov2) [Chen *et al.* 2020c] are both examples of contrastive learning models that were tested on SCDD in this work. SimCLR uses the InfoNCE [Sohn 2016a] objective to train a network to discriminate between different views of the same or different images obtained through random data augmentation. SimCLR performs best with large batch sizes that are needed to obtain a sufficiently large amount of negative samples when computing the loss. MoCo [He *et al.* 2020a] is another contrastive algorithm that minimizes InfoNCE. However, it contains a Siamese architecture in which one network (encoder) is updated via backpropagation. In contrast, the other momentum encoder is updated via an exponential moving average of the encoder weights, which ensures the consistency of the representations of negative samples drawn from a memory bank. MoCov2 improves on MoCo by a projection head and additional data augmentation to generate random views.

Semi-supervised models

This section introduces semi-supervised learning (Semi-SL) models. Semi-SL is a class of algorithms incorporating labeled and unlabelled data into the training process. Some Semi-SL models have been developed for semantic segmentation. Additional details on the models selected for the experiments described in Chapter 8 are also provided.

The Cross-Consistency Training (CCT) [Ouali *et al.* 2020a] and Using Unreliable Pseudo-Labels (U^2PL) [Wang *et al.* 2022a] are two Semi-SL frameworks that were tested on SCDD in this work. In CCT, a shared encoder and main decoder are trained in a supervised fashion, while consistency with the output of various auxiliary decoders is simultaneously enforced. Consistency is enforced by making the model's predictions invariant to small perturbations in the input. U^2PL approaches semi-supervised semantic segmentation via pseudo-labeling [Lee and others 2013] rather than consistency training. Typically, the only pseudo-labels used for training are those the model is confident about. However, U^2PL aims to make use of these unreliable pseudo-labels, as they may still prove helpful for discrimination. This is done by using so-called anchor pixels (or query pixels), selecting associated negative and positive samples for each anchor pixel, and computing a contrastive loss. This contrastive loss is part of the overall loss function, which includes additional supervised and unsupervised terms.

4.2 Loss Function

The loss function is a mathematical representation of the learning objective for a machine-learning model. In linear regression models, for example, the root-mean-squared-error (RMSE) typically serves as the loss function. The RMSE measures the difference between the actual data points and the predictions based on the least-squares fit of the data that minimises the RMSE. The model learns the slope and intercept for a line from the data that minimises the RMSE across all observations. The concept of a loss function also applies to more complex deep-learning models.

Cross-entropy loss was selected to train the fully supervised models in this research. Jadon [2020] compared fourteen loss functions on semantic segmentation tasks and found that no single loss function performed best for all tasks. However, the authors noted that cross-entropy is widely used for semantic segmentation. The cross-entropy loss function in a semantic segmentation context forces the model to learn coefficients that minimise the classification error between the predicted and the ground truth for each pixel. The equation for cross entropy loss in binary classification is attributed to Niculescu-Mizil and Caruana [2005] and nicely defined in Yeung *et al.* [2022] for class imbalanced medical image segmentation. The cross-entropy loss function for binary classification can be written as shown in Equation 4.1. For multi-classification SemSeg, the binary cross-entropy is computed for each class, and then the weighted average is computed across all classes.

$$CrossEntropyLoss = -(y * log(p) + (1 - y) * log(1 - p)) \tag{4.1}$$

where:

y represents the true label (0 or 1),
p represents the predicted probability of the positive class.

For the self- and semi-supervised models, the loss functions were defined by the researchers to achieve the object of each specific model.

4.3 Class weights

The class weights are an important parameter to consider when training the SemSeg model. This research included experiments on three sets of class weights for the fully-supervised SemSeg models. The weights were all equal in the proof of concept using the U-Net. The equal class weights (Set A) assigned a value of '1' to all classes ($wA_1 = wA_2 = \ldots = wA_c = 1$) where wA_c is the equal class weight for the cth class.

In the second publication [Pratt *et al.* 2023], each model was trained with three different sets of class weights. Set A assigned an equal value of 1 to all class weights. Set B assigned a value for each class equal to the inverse of the median percentage of pixels in each class across all images in the training dataset (Equation 4.2).

$$wB_c = \frac{1}{median(p_{c1}, p_{c2}, ..., p_{cn})} \tag{4.2}$$

where:

wB_c is the inverse class weight for the cth class,
p_{cn} is the percentage of pixels in the cth class for the nth image,
n is the number of images in the test dataset.

Set C class weights were based on engineering judgment and assigned unique weights only to specific features and defects of interest. The weights for the remaining classes were all set to 1. A ratio of 100:1 for the crack class to the background class was set to enhance crack detection and minimize the impact of noise from the grain boundaries in the multi-c-si solar cells. The class weights for the critical gridline and inactive regions were set to ten to improve detection relative to other defects. The class weight for ribbons was set to 3 to improve the detection of this feature so that it could be used to classify the cell type with respect to the number of interconnect ribbons in future work.

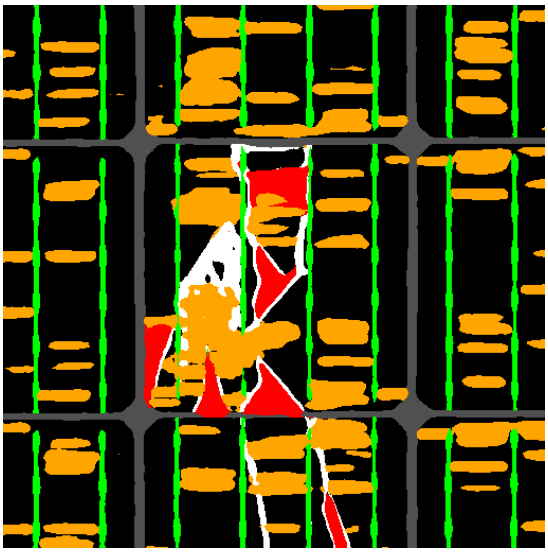
Table 4.1 shows three sets of class weights for selected classes. Note that Set B and Set C class weights are linearly correlated ($r^2 = 0.85$).

Table 4.1: Class weights for a selected set of classes.

Class	Set A	Set B	Set C
Background	1	1.3	0.2
Ribbons	1	22.4	3.0
Gridline	1	164.4	10.0
Inactive	1	84.2	10.0
Crack	1	644.0	20.0
other	1	wB_c	1.0

4.4 Statistical design of experiments for validation of class weights

A statistical design of experiment (DOE) was conducted to explore the impact of class weights on predictions from the DeeplabV3+ semantic segmentation model. The DOE proved that the choice of class weights significantly impacts the predictions when training a multi-class deep learning model. Figure 4.1 shows the impact on the predictions for one cell when the class weights for the gridlines and ribbons were changed from 5 and 11 (Figure 4.1a) to 82 and 1.5 (Figure 4.1b), respectively. The weights for all remaining classes were the same when training these two models. Gridlines and even cracks were all but absent in predictions from Model 220218-10. Gridlines appeared, although dilated, and the cracks correctly appeared around the inactive areas in Model 220218-12 predictions.



(a) Model 220218-10

(b) Model 220218-12

Figure 4.1: Predictions for one cell when trained using a DeepLabv3+ architecture with (a) class weights for gridlines = 5 and ribbons = 11 (Model 220180-10) and (b) gridlines = 82 and ribbons = 1.5 (Model 220218-12)

A two-level screening design was selected for the experiment because it is ideal for exploring many factors with a minimum of experimental runs. Each factor is set to a high and a low value to examine the main effects and interactions. An experiment with two factors requires $2^2 = 4$ runs, while an experiment with five factors requires $2^5 = 32$ runs. This experiment focused on five classes in the dataset: background, ribbons, cracks, gridline defects, and inactive areas.

Figure 4.2 shows the mIoU values from all 16 runs in the DOE and the custom weights used in this work. The values along the x-axis show the high and low settings for each set of class weights repeated for three selected defects: cracks, gridlines, and inactive areas. The low values were set to 50% of the custom class weights, and the high values were set to 50% of the inverse class weights. The y-axis shows the mean value for the IoUs for each of the three defect classes. The choice of class weights has a significant impact on the predictions. Notably, the custom class weights shown as an asterisk resulted in IoUs being among the best for these three defect classes and performing consistently well across other classes.

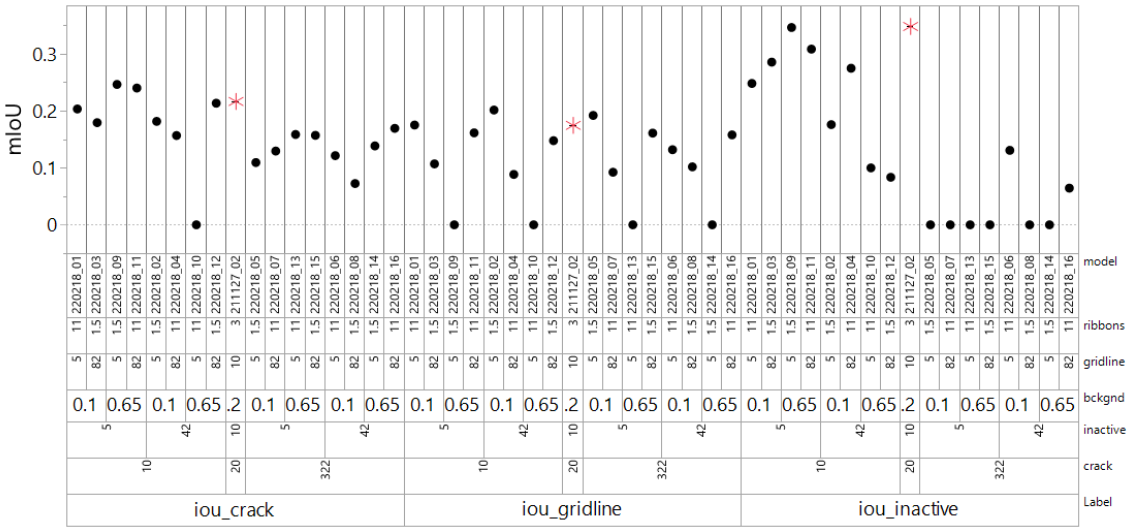


Figure 4.2: Mean IoU for cracks from the 16-run DOE for class weights, including the custom class weights, shows as an asterisk.

4.5 Performance metrics for models

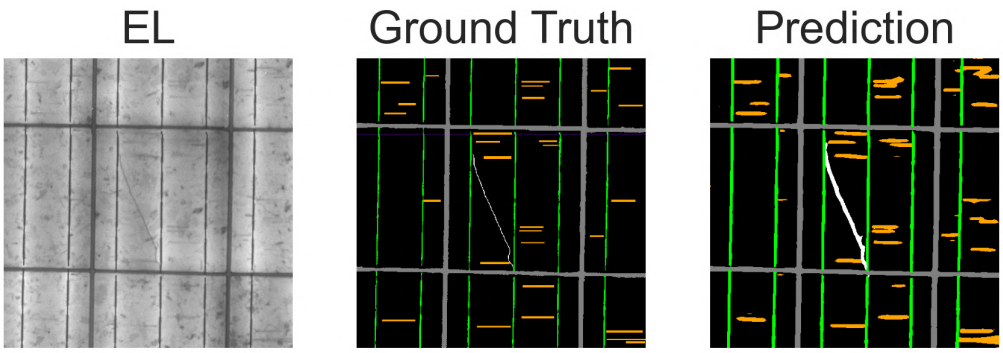
Precision and recall were calculated to evaluate the performance of each model and compare the performance of various models. Precision is the ratio of true positive (TP) predictions to the total number of positive predictions made by the model. It focuses on the correctness of the positive predictions and punishes a model for too many false positives (FP). Recall is the ratio of true positive (TP) predictions to the total number of actual positive samples. It focuses on capturing as many true positive samples as possible and punishes a model for too many false negatives (FN). Equations for both metrics were published in Davis and Goadrich [2006], cited over 6,800 times.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4.3}$$

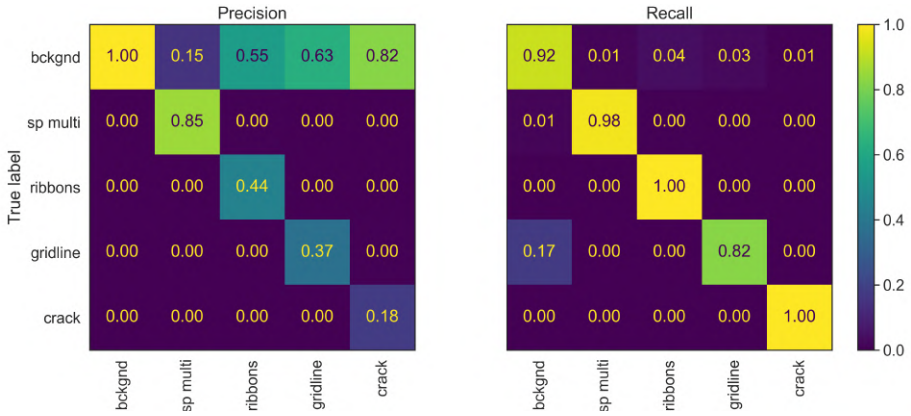
$$\text{Recall} = \frac{TP}{TP+FN}, \tag{4.4}$$

where TP, FP and FN represent the True Positives, False Positives and False Negatives, respectively.

Figure 4.3 shows the output for the precision and recall of a single cell. Figure 4.3a shows the input EL image, ground truth mask, and predictions mask. Figure 4.3b shows the confusion matrices for the precision and recall for selected defects. A perfect SemSeg model would show all ones along the diagonal of both matrices with zeros in the off-diagonal cells.



(a) El, ground truth, and prediction



(b) Confusion matrix

Figure 4.3: One example showing (a) the original EL image, the ground truth mask, and the prediction, and (b) the confusion matrices for precision and recall

The precision and recall matrices quantify how well the prediction masks match the ground truth masks. For example, the row labelled ‘sp multi’ shows how well the computer model predicted the pixels that were labelled as ‘sp multi’ and coloured grey in the ground truth mask. The precision matrix indicates that 85% of the pixels classified as ‘sp multi’ by the computer model were also labelled ‘sp multi’ in the ground truth image (high precision). The recall matrix indicates that 98% of the pixels labelled as ‘sp multi’ in the ground truth mask were also labelled as ‘sp multi’ in the prediction mask (high recall). Similarly, the row labelled as ‘cracks’ in the precision matrix shows that 18% of the pixels labelled as ‘cracks’ by the computer model were also labelled ‘cracks’ in the ground truth mask (low precision). The remaining 82% of the pixels were mistakenly labelled as background by the model. By contrast, the row labelled as ‘cracks’ in the recall matrix shows that 100% of the pixels labelled as cracks in the ground truth image were also predicted as cracks by the model (perfect recall). The low precision and high recall for the crack defect in this image can be explained by the dilated representation of the crack in the prediction. Clearly, the model located the crack within the cell accurately, but the width of the crack was too wide, i.e., dilated. While low precision is undesirable in general, the high recall covering the crack in the correct location proves that SemSeg can be a useful tool for SCDD. Defect dilation was also an issue with human raters. Some raters labelled more pixels than others when annotating ground truth images, especially long, narrow defects like

cracks and gridline defects. This also has an impact on precision and recall metrics. Measures were put in place to minimise the impact of human rater differences, as described in a later section.

The Intersection over Union (IoU) was also used to evaluate the performance of each model and compare the performance of various models. The IoU provides a measure of both precision and recall. For a given class, the IoU is the ratio of the number of pixels in the intersection of the ground truth and the prediction divided by the area of the union, as shown in Equation 4.5. A high IoU score can only be achieved if the precision and the recall are both high.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} = \frac{\text{Area in GT} \cap \text{Area in Prediction}}{\text{Area in GT} \cup \text{Area in Prediction}} \tag{4.5}$$

Figure 4.4 illustrates the concept of IoU using an example of a crack observed in cell CFVS-00035-r8-c4 from our dataset. This image shows the overlay of the crack layer from a ground truth mask (white) and the crack layer from the prediction mask (green) from the first U-Net model with the class weight = 1. The intersection is the number of pixels that are both white in the ground truth and green in the corresponding location in the prediction. The union is the sum of the pixels that are white in the ground truth or green in the prediction mask. For the crack layer, precision = 60%, recall = 48%, and IoU = 36%. If the class weight were increased, then the prediction tends to be dilated, leading to higher recall and lower precision, as seen in Figure 4.3a above. This crack prediction provided additional proof of concept for the potential of SCDD using SemSeg. The model did a good job of locating a long crack defect that was only a few pixels in width.

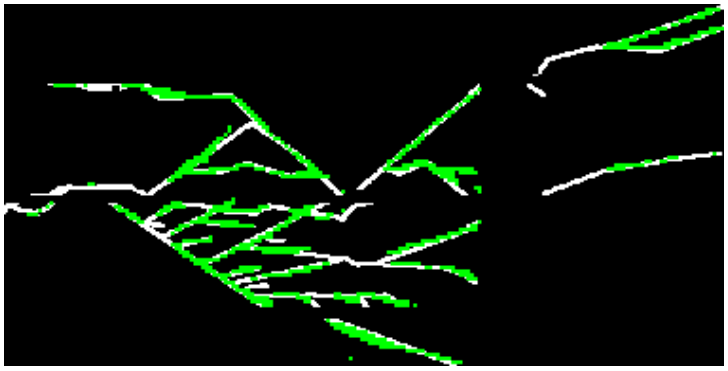


Figure 4.4: Crack layer for CFVS-00035-r8-c4 showing the prediction for cracks from the U-Net model (green) and the underlying ground truth mask (white). The IoU metric provides a numerical estimate of the overlap between the two.

Equation 4.6 shows the computation for the mean IoU (mIoU) for a single image. The mIoU is the arithmetic average of the IoUs computed for each class in the image. The mIoU applies equal weight to all classes, so prediction errors in the background class have the same penalty as the prediction error in the crack class.

$$\text{mIoU} = \frac{1}{n_c} \sum_{i=1}^{n_c} IoU_i \tag{4.6}$$

where:

n_c is the number of classes in the image, and
IoU_i is the IoU for the i^{th} class in the image.

Equation 4.7 shows the computation for the weighted average IoU (wIoU) for a single image. The wIoU is the weighted average of the IoUs computed for each class in the image given the set of class weights used to train the model. The wIoU applies a greater penalty to classes with higher weights during training, so the class weigths can be used to focus the learning on specific classes by assigning a higher weight. The wIoU is a useful metric when training models with class imbalance. Similarly, the class weights can be applied to the mIoU averaged across a set of images for each class.

$$\text{wIoU} = \sum_{i=1}^{n_c} \frac{w_i}{\sum_{i=1}^{n_c} w_i} IoU_i \tag{4.7}$$

where:

n_c is the number of classes in the image, and
IoU_i is the IoU for the i^{th} class in the image, and
w_i is the class weight for the i^{th} class.

Equation 4.8 shows the computation for the mean IoU for a single class (cIoU). The cIoU is the arithmetic average of the IoUs computed for a single class across all images in the dataset. This statistic was used to compare the performance of multiple models for a specific class of interest, such as cracks, gridlines, and defects.

$$\text{cIoU} = \frac{1}{n} \sum_{i=1}^{n} IoU_i \tag{4.8}$$

where:

n is the number of images in the dataset with a given class, and
IoU_i is the IoU for the i^{th} image in the dataset.

4.6 Statistical methods for model comparisons

Multiple models were trained and evaluated over the course of this research. In many cases, the differences in IoU were quite small, so a statistical method was needed to determine which models were significantly better or worse than others in the presence of random variation resulting from training. For example, the cIoU for cracks in the test dataset will vary from one round of training to another on the same dataset. To make a valid statistical comparison across models, each model was trained over five rounds, resulting in five estimates of the average IoU

for each class from each model in the same test dataset. Those data were then subjected to an analysis of variance (ANOVA) method to determine if any model was better or worse than the average performance. The ANOVA method for equality of means is achieved by analysing estimates of variance based on differences in group means to the grand mean of the data, giving rise to the method name 'analysis of variance.' Following the ANOVA, Tukey's Honest Significant Difference (HSD) method was implemented to compare the IoU of each pair of models. The HSD method increases the confidence interval width for each pairwise comparison to account for the multiple comparisons and therefore controls the experiment-wise error rate. In many cases, several models would not show a statistically significant difference in performance, while other models show significantly better or worse performance.

The 'statsmodels' python library included methods for implementing the ANOVA and HSD methods. The library also included a method for visualising statistically significant differences when comparing the average values from multiple groups [Seabold and Perktold 2010]. For example, Figure 4.5 shows the mean and 95% confidence intervals (CI) for IoU on crack detection from multiple models (R0, R1, ..., R9), i.e. for the cIoU for cracks. The CIs were computed from multiple training runs of the same model and a pooled estimate of variance. Model R2 was designated as the reference model in this example, so the 95% CI for the cIoU is shown in blue. The choice of the reference model is arbitrary and has no bearing on the outcome of the statistical comparison. The choice of reference model does impact the colour for each model depending on whether or not there is a statistically significant difference relative to the reference model. R2 was selected in this case as the reference model because it was the top performer. Model R4 and R7 appear in grey because the 95% CIs for the cIoU overlap the CI for the reference model R2. The overlap means there is no statistically significant difference at the p = 0.05 level between the cIoU for these three models. The remaining CIs are all coloured red because the cIoUs are significantly different from the reference model. All other pairwise comparisons can also be made by looking for overlapping CIs. For example, the graph shows there is no statistically significant difference among models R0, R3, R5, R6, R7, and R9 at the p = 0.05 level because of the overlapping CIs. In addition, the chart also shows that the cIoUs for model R1 and R8 are both significantly lower than all other models because of the non-overlapping CIs.

A block-centred IoU ($IoU_{c,bc}$) as shown in Equation 4.9 was computed for each class to reduce the variance in the response and increase the power of the statistical comparison based on Tukey's HSD method when comparing across models. The variability from one image to the next within the same class will lead to an increase in the overall variability and reduce the power of the test to detect a statistically significant difference across models unless explicitly accounted for in the analysis. The $IoU_{c,bc}$ effectively eliminates the variability in the distribution of the IoUs attributed to the variability from one image to another for a given class by subtracting the IoU for each image averaged across all models from the IoU for the image. Figure 4.6 provides a visual illustration of the impact of block-centred IoUs. The IoU for cracks on selected images from the dataset from multiple deep-learning models over five training runs is shown before (Figure 4.6a) and after (Figure 4.7b) subtracting the average IoU for each image. Note that the IoU varies only slightly across models for any single image but varies significantly from one image to the next. Figure 4.6a shows the IoU for predictions are quite high for Image E across all models and quite low for Image F across all models. Figure 4.7b shows that the $IoU_{c,bc}$ decreases the variance in the distribution of IoUs after subtracting each IoU from

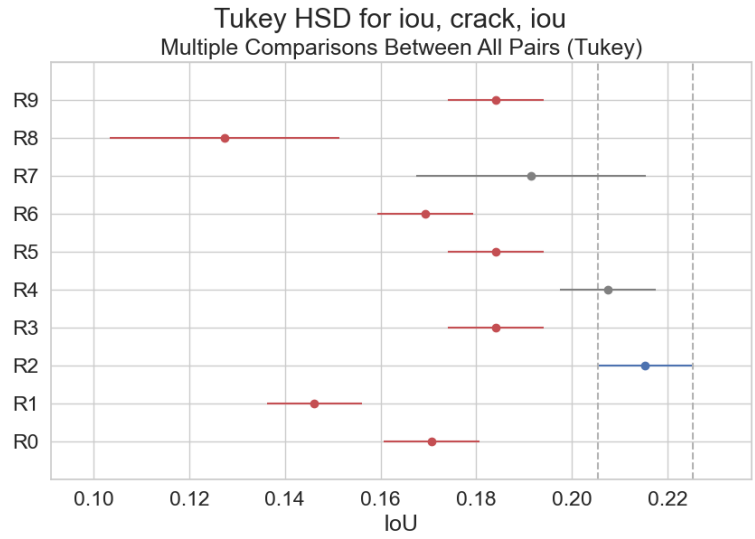
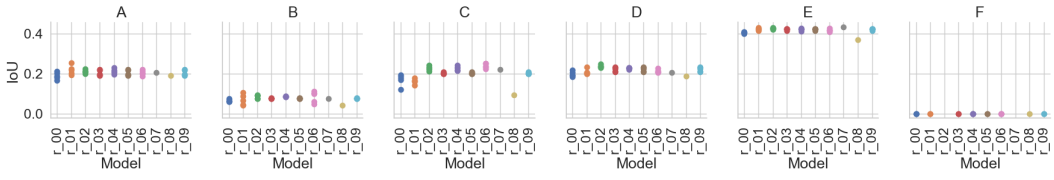
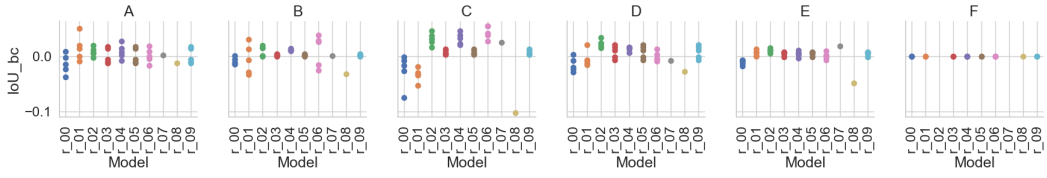


Figure 4.5: 95% confidence intervals for cIoU on crack defects by regime

the average image value across all models, thus centring the differences attributed to the model around zero for all images. This reduction in variance increases the statistical power to detect differences across models. Approximately 85% of the variability in the distribution of IoU for cracks is attributed to the variability across images before subtracting the averages. A simple two-way ANOVA could be modelled to include effects from both the 'model' and 'image' as predictors (two-way ANOVA), but that method is not compatible with the multiple comparison plots used to identify which specific models are statistically better or worse than other models. Thus, the block-centring on the average IoU for each image had to be handled before calling the method.



(a) IoU without block-centring



(b) IoU with block-centring

Figure 4.6: IoU for cracks by model over five training runs for six selected images (A-F) from the test dataset without block-centring (a) and with block-centring (b)

Equation 4.9 shows the computation of the block-centred IoU for a specific image, model, and class ($IoU_{i,m,c,bc}$). The average value of IoUs for all images from a specific class and set of models is subtracted from the specific IoU to compute the block-centred IoU. Graphically,

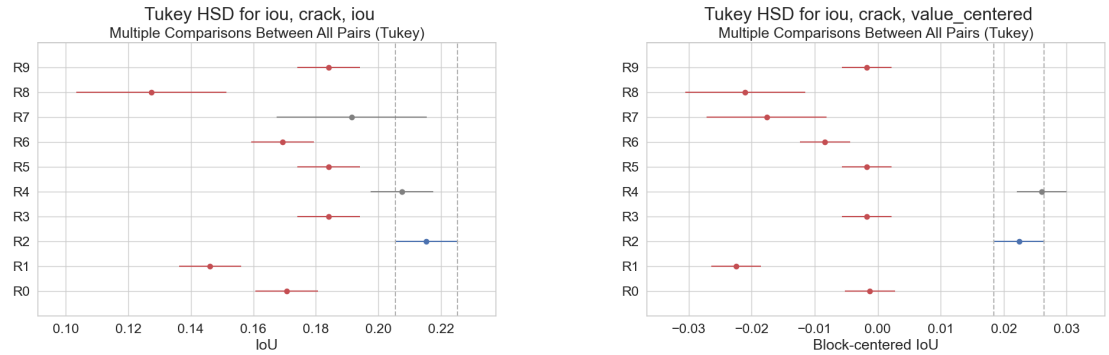
the mean of each subplot in Figure 4.6 is subtracted from each IoU in the same subplot so that all the $IoU_{i,m,c,bc}$ subplots are centred at zero. This method is easily extended to include the repeated values from over multiple training runs. After this normalisation process, the only variability remaining in the distribution is attributed to the main effect ('model') and the random variability from multiple training runs.

$$IoU_{i,m,c,bc} = IoU_{i,m,c} - \frac{1}{n_{models}} * \frac{1}{n_{images}} \sum_{m=1}^{n_{models}} \sum_{i=1}^{n_{images}} IoU_{i,m,c} \tag{4.9}$$

where:

$IoU_{i,m,c,bc}$ is the block-centred IoU for a specific image, model, and class, and
$IoU_{i,m,c}$ is the IoU computed from the prediction for a specific image, model, and class, and
n_{models} is the number of models in the evaluation, and
n_{images} is the number of images with the specific class

Figure 4.7 compares 95% CIs for cracks based on the cIoU to the $IoU_{c,bc}$ with respect to the power of the test to detect differences in performance. The graph on the left compares the CIs for cIoU from multiple models, and the graph on the right compares the $IoU_{c,bc}$ from multiple models. The reduction in variance leads to tighter CIs for each model and greater statistical power when using the $IoU_{c,bc}$. Note that the CIs for all models using the cIou range from 0.1 to 0.225 (0.122), while the CIs for all models using $IoU_{c,bc}$ range from -0.03 to +0.035 (0.065). In this example, the conclusion regarding the performance of model R7 changes from statistically equivalent to the best-in-class performance to significantly worse than the best-in-class.



(a) cIou (b) $IoU_{c,bc}$

Figure 4.7: CIs from Tukey's HSD method comparing model performance for crack detection using cIoU (left) and $IoU_{c,bc}$ (right)

Tukey's HSD method also extends to the mIoU and the wIoU. Figure 4.8 compares all the defect classes from nine models using $mIoU$, $wIoU$, $mIoU_{bc}$, and $wIoU_{bc}$. Based on this graph, the model comparisons change only slightly with the choice of metric when averaging across all the defects. Model R0 tends to separate more from the pack when using the block-centred IoUs.

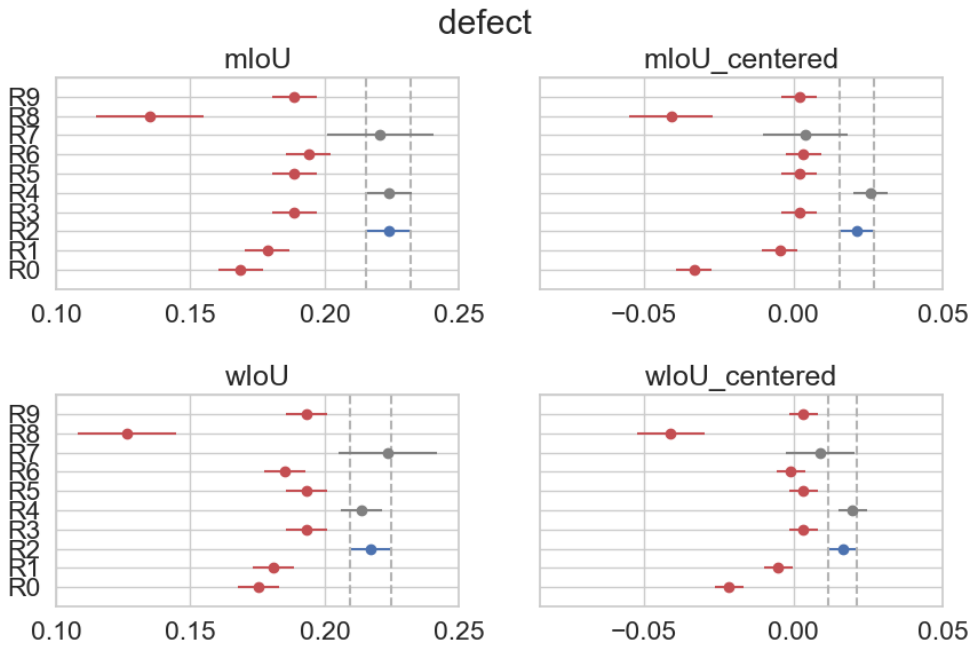


Figure 4.8: CIs from Tukey's HSD method comparing model performance using $mIoU$, $wIoU$, $mIoU_{bc}$, and $wIoU_{bc}$

4.7 Linear Regression

This section summarises the analysis conducted on a dataset generated at the PV Module Quality and Reliability Laboratory at the CSIR. The details of this work are outside the scope of the thesis, which is focused on the prediction machine and not the subsequent analysis of the output. However, this analysis is included to illustrate the potential value of the SCDD tool based on SemSeg to the PV industry. Further details on the work are provided in Chapter 9.

Figure 4.9 shows the correlation between the fill factor (FF) and the 'brightening' defect predicted by one of the deep-learning models trained in this research. FF is a function of four points from the I-V curve, and a higher FF is desirable. Each point in the plot represents a PV module. The y-axis represents the measured FF for a module. The x-axis represents the number of pixels from each cell in the module associated with the brightening defect as a percentage of the total number of pixels from each cell. The output of the deep-learning model prediction determined the number of pixels. The brightening defect quantifies the degree to which the regions closest to the ribbon interconnects on the solar cell are brighter than the regions further from the ribbon interconnects. Based on this linear regression model, 42% of the variability in FF can be explained by the extent of brightening defects (r^2 = 0.42).

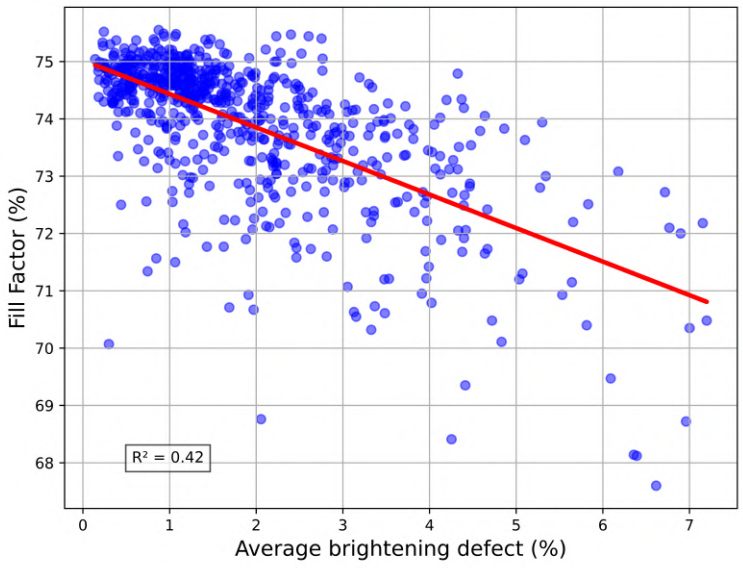


Figure 4.9: Linear regression model of module FF versus average of cell-level brightening defect predicted by the model

Chapter 5

Datasets for SCDD

5.1 Public datasets

Many researchers developed methods for SCDD over the years using the ELPV dataset [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2021]. Benchmark datasets for SCDD in EL images using multi-class semantic segmentation (SemSeg) did not exist until 2020 when Sovetkin *et al.* [2020] published EL images of thin film module to segment two classes of defects for the PEARL project. A benchmark dataset for EL images of silicon wafer-based PV modules was not publicly available until the University of Central Florida published the UCF dataset in 2022 [Fioresi *et al.* 2022]. Examples of both datasets are provided in this section. Multi-class SemSeg models for EL images have received relatively little attention in the research to date, and that is likely due to a lack of available datasets with ground truths to use for testing and training models. While the PEARL dataset became public in 2020, the thin film PV module market captures less than 5% of the overall PV market globally so interest may be low. Hopefully, the new UCF dataset published in 2022 and the benchmark dataset published as part of this thesis will promote further research in SemSeg on EL images of silicon wafer-based modules.

Public datasets of natural and man-made objects are available to train deep-learning models for classification and object detection. The ImageNet [Russakovsky *et al.* 2015] dataset consists of 1000 object classes, including animals, food, and man-made objects found in everyday modern life. The original COCO [Lin *et al.* 2014] dataset consists of 91 objects in 12 categories also common in modern life. The Cityscapes [Cordts *et al.* 2016] dataset consists of 30 object classes commonly found in modern urban life, including roads, buildings, vehicles, and people. The Google Open Images V7 dataset contains approximately 9 million images spanning over 20,000 categories. The CIFAR10 [Krizhevsky *et al.* 2009] dataset consists of ten classes, including animals, airplanes, ships, and trucks. The Google Open Images database does include seven categories containing the word ‘solar’ and over 5,000 corresponding images. The images range from small PV panels for battery charging to large, ground-mounted PV systems and space applications. Some examples are shown in Figure 5.1. The ‘solar vehicle’ category contains many interesting and novel applications of solar PV on land and water. ImageNet1000 also includes three categories with the word ‘solar’ in the title. However, none of these datasets includes EL images of solar cells, which require specialized equipment to obtain.

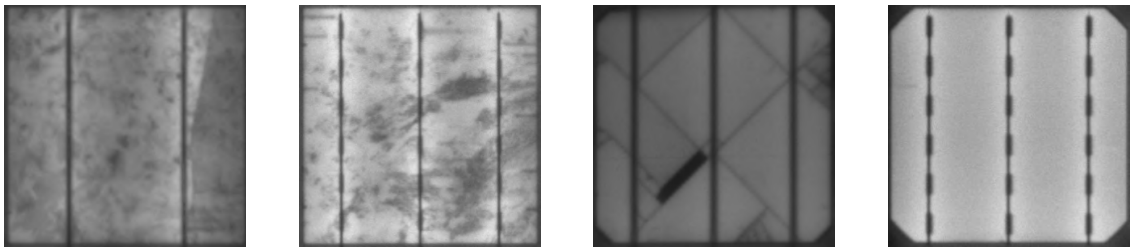


(a) (b) (c) (d) (e)

Figure 5.1: Sample images of solar panels contained in the Google Open Image database

Publicly available datasets with EL images of solar cells also exist. However, the datasets published originally did not support semantic segmentation models for SCDD due to a lack of necessary ground truth images.

In 2018, ZAE Bayern, a non-university research centre in Germany, published the ELPV dataset with 2,624 EL images and corresponding four defect probabilities for analysis [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2021].[1] This dataset has been widely cited by researchers working on classification and binary segmentation methods. However, the dataset consists of largely outdated solar cell architectures cropped from only 18 unique mono-c-si modules and 26 unique multi-c-si modules [Demirci *et al.* 2019]. Figure 5.2 shows some examples of EL images in the ELPV dataset with only two or three interconnect ribbons, consistent with older generation solar cell architectures.

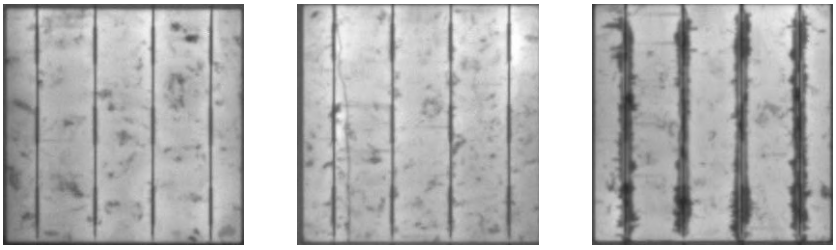


(a) (b) (c) (d)

Figure 5.2: Sample EL images from the ELPV benchmark dataset: (a) multi-ci-si with two ribbon interconnections and a long crack, (b) multi-ci-si with three ribbon interconnections and gridline defects, (c) mono-ci-si with three ribbon interconnections, cracks, and inactive areas, and (d) mono-ci-si with three ribbon interconnections and no defects [Buerhop-Lutz *et al.* 2018; Deitsch *et al.* 2019 2021]

In 2020, Case Western Reserve University Solar Durability and Lifetime Extension (SDLE) published the SDLE dataset with 1,028 EL images of solar cells for the classification of good, cracked, and cells with corroded interconnect ribbons [French *et al.* 2020].[2] The SDLE dataset consists entirely of multi-c-si cells with four ribbon interconnections. Figure 5.3 shows examples of EL images from the SDLE dataset.

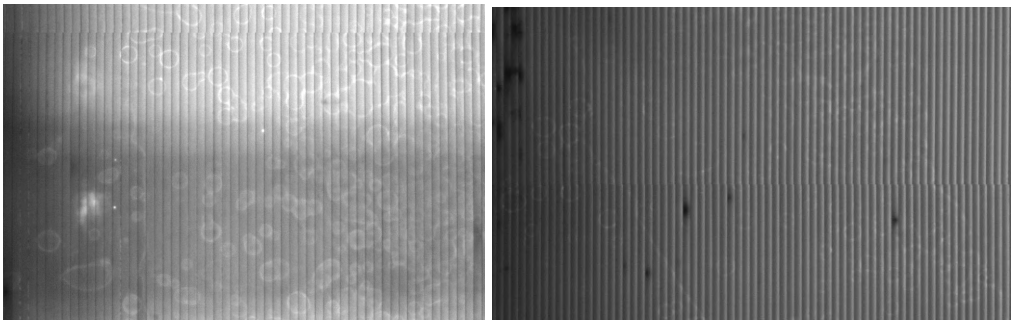
[1] https://github.com/zae-bayern/elpv-dataset/tree/master/images
[2] https://osf.io/4qrtv/



(a) (b) (c)

Figure 5.3: Sample EL images of multi-c-si cells with four ribbon interconnection in the SDLE benchmark dataset: (a) no defects, (b) long, narrow crack, and (c) corroded interconnection ribbons [French *et al.* 2020]

In 2020, the PEARL project published 6000 EL images and annotations for SemSeg on thin-film modules Sovetkin *et al.* [2020].[3] The models were trained to segments two classes of defects: shunts and droplets. Thin-film modules have a completely different design than silicon-wafer-based modules. In crystalline silicon modules, individual solar cells are connected together in series to form a module. In thin film solar modules, thin layers of semiconductor materials are deposited onto a glass substrate or superstrate. Lasers are used at various stages to scribe long, horizontal lines from one end of the module to the other to create the individual cells. The resulting 'solar cells' are typically less than one cm wide by 120 cm long, depending on the glass size. Figure 5.4 shows examples of EL images in the PEARL dataset.



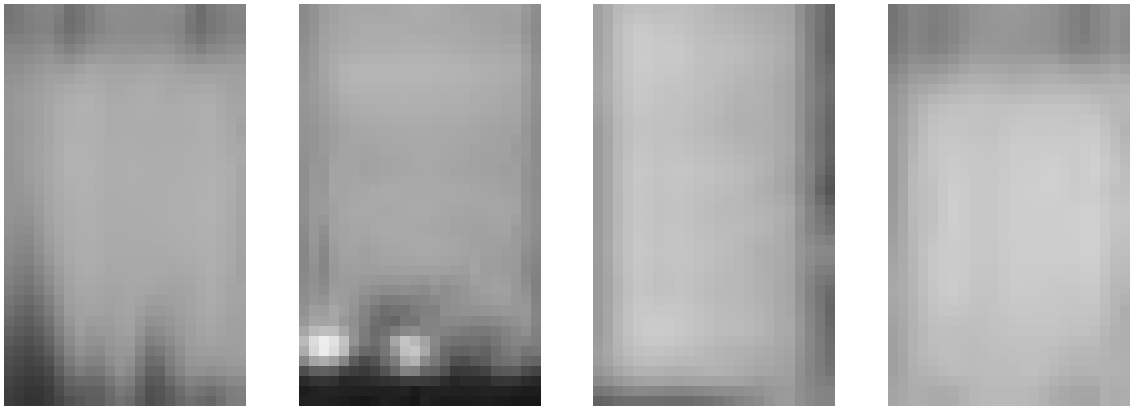
(a) Droplets (b) Shunts

Figure 5.4: Sample EL images of thin films modules in the PEARL benchmark dataset: (a) droplets and (b) shunts [Sovetkin *et al.* 2020]

In 2020, Raptor Maps Inc. published a dataset of infrared (IR) images of PV modules to support research in defect classification based on IR images.[4] The dataset consists of 20,000 low-resolution infrared images (24 by 40 pixels) for image classification. There are 12 defined classes of solar modules presented with 11 classes of different anomalies, and the remaining class being normal. Figure 5.5 shows examples of images from the RAPTOR dataset. These are noticeably different from typical EL images concerning resolution and detail. The highly pix-

[3] https://b2share.fz-juelich.de/records/e19fe0f0bc0e4f1b8e05b28e977934c9
[4] https://github.com/RaptorMaps/InfraredSolarModules

elated images reveal hot spots within the module corresponding to the white pixels and cooler regions of the module corresponding to darker regions. However, the information embedded in these IR iamges is significantly less than the information embedded in a typical EL image.



(a) (b) (c) (d)

Figure 5.5: Sample EL images from the RAPTOR dataset: (a) module with relatively cool regions at the top and bottom , (b) module with two hotspots along the bottom region, (c) module with a cooler region along the right edge, and (d) module with a cooler region along the top [Millendorf *et al.* 2020]

In 2022, Beihang University published the PV EL Anomaly Detection (PVEL-AD) database to support research in object detection. The dataset consists of 36, 543 EL images from multi-crystalline silicon cells with ground truth bounding boxes [su binyi 2022].[5] Figure 5.6 shows examples of the defect classes and corresponding bounding boxes. These EL images were taken before the cells were assembled into PV modules, so current injection requires custome hardware configured to match the solar cell design. The current injection occurs through a series of pogo pins aligned to each ribbon interconnect region. The pogo pins are visible in the EL images.

In 2022, researchers from the University of Central Florida (UCF) published a dataset for Sem-Seg on wafer-based cells. The dataset includes 11,851 EL images, the corresponding ground truth images for training, and an additional 17,000 images for testing [Fioresi *et al.* 2022].[6] Figure 5.7 shows examples of EL images from the UCF dataset. The work focused on the semantic segmentation of ten different defect classes associated with contacts, interconnections, and cracks. Four types of contact defects were identified: gridline, ribbon corrosion, belt mark, and near solder pad. Three types of interconnect defects were identified: disconnected, highly resistive, and bright spot. Three types of crack defects were identified: closed, resistive, and isolated. These classifications follow the Type A, B, and C crack definitions described in the IEC technical specification [International Electrotechnical Commission 2018]. There are similarities and overlaps between the UCF and the dataset developed for this thesis with respect to defects. However, the UCF dataset does not label any of the intrinsic features in a solar module,

[5] https://github.com/binyisu/PVEL-AD
[6] https://github.com/ucf-photovoltaics/UCF-EL-Defect

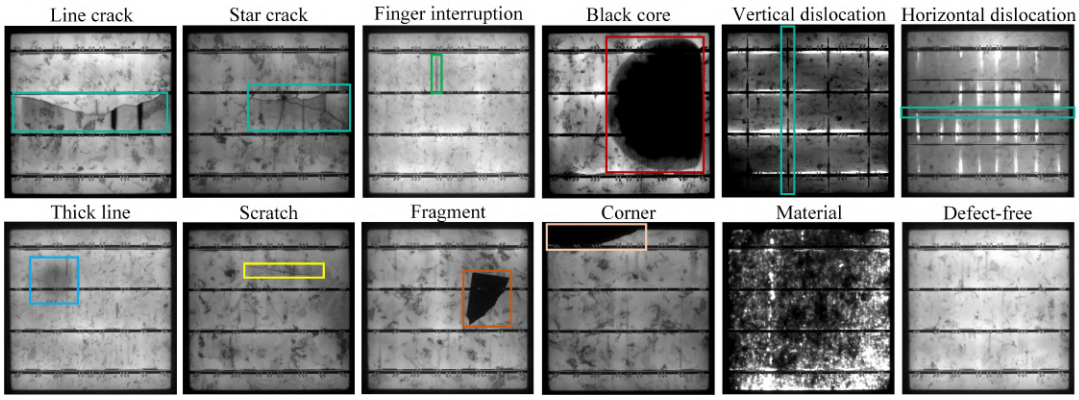
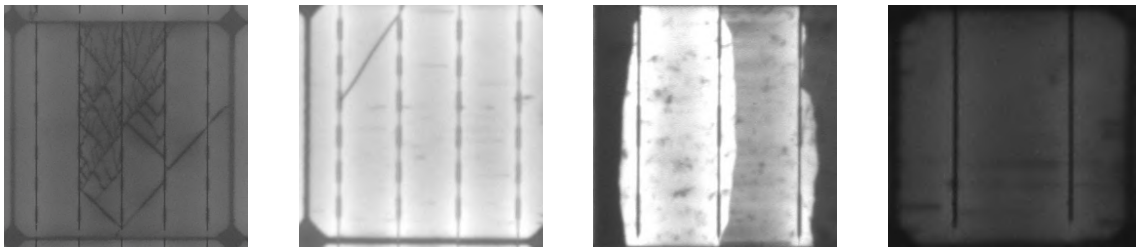


Figure 5.6: Sample EL images of multi-c-si cells with four ribbon interconnection in the PVEL-AD benchmark dataset and the corresponding bounding boxes [su binyi 2022]

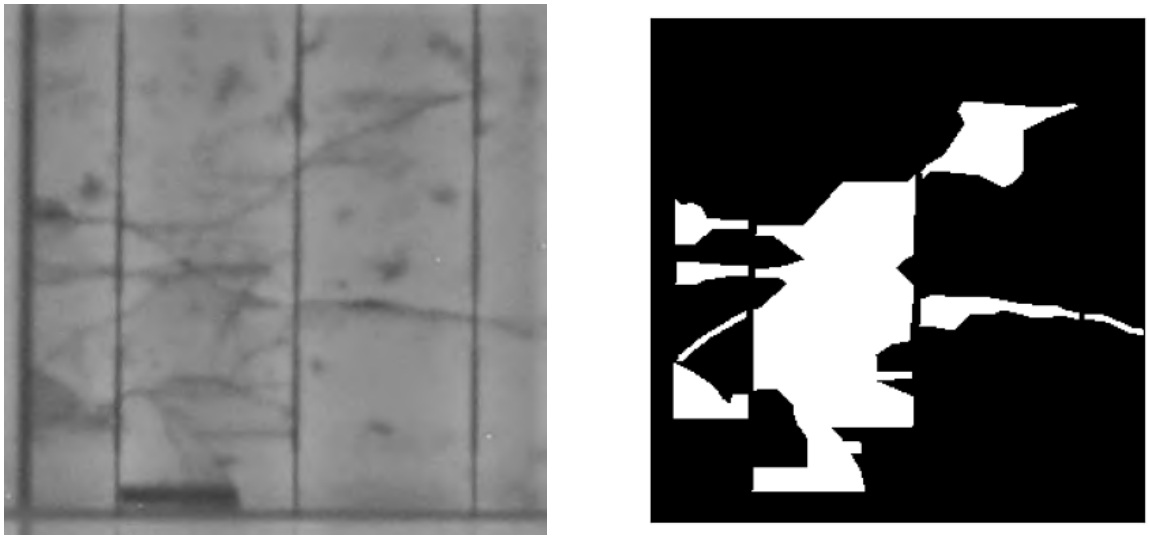
such as the spacing between the cells, the ribbon interconnects, or the busbars. Furthermore, the cracks and gridline defects in the UCF dataset are often aggregated into larger regions rather than labelled individually. The authors refer to such a region as a 'conglomeration'.



(a) (b) (c) (d)

Figure 5.7: Sample EL images from the UCF dataset: (a) mono-c-si cell with five ribbon interconnects and severe cracks, (b) mono-c-si cell with four ribbon interconnects and one heavy crack, (c) multi-c-si cell with three ribbon interconnects, heavy cracks, and inactive areas, and (d) mono-c-si cell with two ribbon interconnects and faint gridline defects [Fioresi *et al.* 2022]

Figure 5.8 shows an example of one EL image and corresponding ground truth mask from the UCF dataset. Several defects are visible in Figure 5.8a, including gridline defects, cracks, and inactive areas. The features present include the cell spacing and the ribbon interconnect. In Figure 5.8b, none of the features were labelled, and all defect types were aggregated into one large 'conglomeration,' as termed by the authors. The larger conglomeration should be easier to label and detect, but the conglomeration does not provide the same level of resolution that individual classes can provide.



(a) EL image (b) Ground truth

Figure 5.8: Example from the UCF dataset showing (a) the EL image and (b) the ground truth mask with one large conglomeration [Fioresi *et al.* 2022]

Table 5.1 summarises the prior public datasets for SCDD in EL images.

Table 5.1: Summary of publicly available datasets with EL images

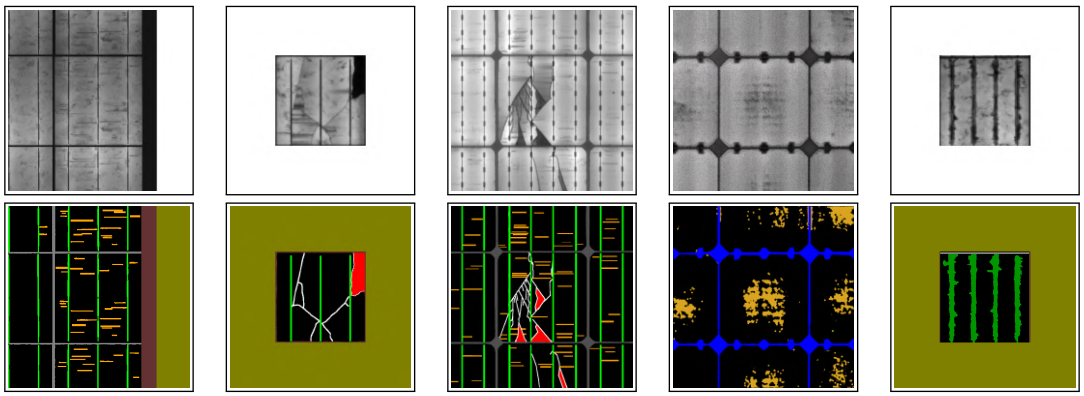
Name	**Year**	**Task**	**Technology**	**Count**
ELPV	2018	Classification	c-Si	2,624
SDLE	2020	Classification	c-Si	1,028
PEARL	2020	SemSeg	Thin Film	6,000
PVEL-AD	2022	Classification	c-Si	36,543
UCF	2022	SemSeg	c-Si	11,851

5.2 New labelled dataset for SCDD

In 2023, the benchmark dataset for SCDD experiments supporting this thesis and the benchmark metrics from various semantic segmentation models was published in Pratt *et al.* [2023].[7] The SCDD dataset consists of EL images and the corresponding ground truth masks with two categories of classes: defects and features. There are 16 classes in the defect category, which includes objects that correspond to extrinsic problems with solar cells, such as cracks, inactive areas, and gridline defects. There are 13 classes in the feature category, which include objects that correspond to intrinsic features of solar cells as designed by the engineers, such as busbars, interconnect ribbons, and cell spacing. The dataset was developed over a series of iterations as the dataset expanded to include more images and classes while simultaneously resolving errors in the annotations. A calibration dataset was deployed for training the larger team of

[7] https://github.com/TheMakiran/BenchmarkELimages

annotators, but thereafter each image was assigned to one annotator. A single expert annotator reviewed and corrected each team member's resulting ground truth masks to reduce inter-rater variability. As the PV expert, I conducted a final semi-automated review of the output from the expert annotator and gave instructions for further edits before final approval and inclusion into the benchmark dataset. Figure 5.9 shows some examples of the EL images and corresponding ground truth masks in the dataset.



(a) (b) (c) (d) (e)

Figure 5.9: Examples of cell level EL images (top) and corresponding ground truth masks (bottom), one from each source: (a) ARTsolar multi-si cell cropped from the edge of a full module, (b) ELPV single multi-si cell with padding on four sides, (c) CFV Labs mono-si cell cropped from the centre of a full module, Dd) CSIR mono-si cell cropped from the center of a full module, and (e) SDLE single multi-si cell with padding on four sides.

A complete library of all features and defects is provided in the annex, along with the RGB colour code used in the ground truth mask. This dataset is the most detailed semantic segmentation dataset currently available, and the benchmark represents the current state of the art. The performance of the models trained and tested in the thesis demonstrates the power of SemSeg to reliably detect and quantify multiple defects in the EL images of solar cells.

Sourcing the EL images

Development of the SCDD dataset began in 2020. The dataset was intentionally designed to include images from all available sources and balanced across cell type (mono-c-si vs multi-c-si) so that the model will be optimised for general applications. EL images were gathered from public and private sources. Cell-level EL images were downloaded from the ELPV dataset published by ZAE Bayern and the SDLE dataset published by Case Western Reserve University. Module-level EL images were curated from 1) CFV Labs, a PV module testing and certification lab in the USA; 2) ARTsolar, a PV module manufacturer in South Africa; 3) the CSIR, a research institution in South Africa; and 4) Brightspot, an equipment company based in the USA. The EL images were assigned unique IDs and organized in a structured directory.

Figure 5.10 shows the number of images containing each defect and feature by source represented in the most current dataset. There are 765 images and ground truth masks in the 20230603 database. Each image is resized to $512x512$ pixels. Some classes are well represented for the purposes of training and statistical analysis of the test results. Some classes were present in only a few images which hindered the training process and analysis. All images have

a ‘background’ layer, so the distribution of the ‘background’ class provides an indication of the distribution across sources. The majority of images were curated from the CSIR, followed by CFVS and ARTS. Relatively few images were curated from the remaining sources. The figure also highlights the imbalance between the feature and defect groups of classes. The feature classes, for the most part, represent things that are supposed to be in the module by design, such as cell spacing. There are exceptions, however. For example, not all cells have interconnect ribbons that can be seen in an EL image because some modules are designed with a back-contact architecture. The defect classes have lower counts by comparison because these objects are not supposed to be in the module by design. During the curation and annotation process, the majority of cells were selected for inclusion because they contain at least one key defect of interest. Other cells were included despite a lack of defects so that some good cells would be available for testing. To date, only four defect classes have a reasonable representation across multiple sources: cracks, gridline defects, inactive areas, and material defects.

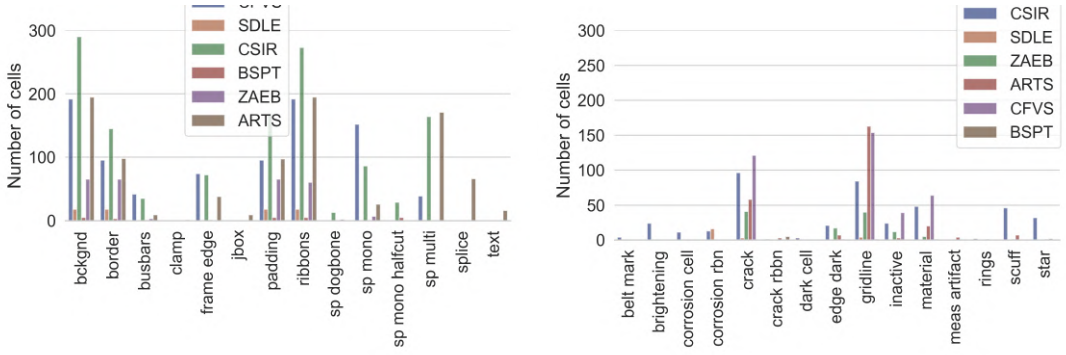


Figure 5.10: Number of images containing each class in the Feature group (left) and Defect group (right)

The sources include module-level EL images of many different sizes. The physical dimensions of a PV module are determined by the number of cells, the cell configuration, and the size of the individual cells. The original dataset included modules with 60 cells or 72 cells, typically. However, a small number of modules had 96 cells (12 rows x 8 columns) or 128 cells (16 rows x 8 columns). More recent module-level images with 144 half-cut cells were added to 20230603 . Modules made with a greater number of cells connected in series have a higher voltage output, while modules made with large cells tend to have a higher current output. Regardless of the module design, the individual solar cell image can be extracted from the module-level image for analysis. Cell-level images form the basic unit of analysis in SCDD, following the lead of previous authors [Mantel *et al.* 2019; Karimi *et al.* 2019; Spataru *et al.* 2016b; Akram *et al.* 2019b; Deitsch *et al.* 2019].

Cropping the cell-level images

A simple method was developed to crop cell-level images from the module-level EL images regardless of the module size. Previously published work focused on detailed methods for precisely cropping the solar cell image from the module-level EL images, but the general applicability was of concern. Our SCDD dataset was purposefully designed for broad applicability, so the pre-processing and cell cropping methods were kept simple and easily generalised. The method was also developed to be independent of the solar cell dimensions. The EL images

were cropped using a custom Python function based on the PeakUtils library [Negri and Vestri 2017]. This approach takes advantage of regularly spaced rows and columns of solar cells that make up a full crystalline silicon module. The spaces between cells are inactive areas of the PV modules, so they do not emit light when the module is energised for EL imaging. The inactive spaces create characteristic dark rows and columns spaced evenly through the EL image of the module. The row-wise and column-wise average pixel intensities create a characteristic profile in two dimensions, with peaks and valleys corresponding to the darker columns dominated by the cell spacing. PeakUtils provides a convenient method for locating the 'peaks' by analysing the inverse of the intensity profiles. The custom Python script includes a method to identify the intersection of those peaks which correspond to the four corners of each cell. A similar step was included in the pipeline proposed in previous work [Sovetkin and Steland 2019; Mantel *et al.* 2019].

Figure 5.11 illustrates the method used to crop cell-level images from the module-level EL image. These methods are similar to those proposed by [Sovetkin and Steland 2019] and [Mantel *et al.* 2019]. The figure includes the module-level EL image and the corresponding profiles for the inverse of the average pixel intensity along the rows and columns. The peaks correlate with the row and column cell spaces. The intersection of the peaks corresponds to the four corners of each cell in the module-level image. The lesser peaks seen in the column profile correspond to the ribbon interconnects. The ribbon interconnects are metal strips soldered onto the cell surface to connect the top layer of each cell to the bottom layer of the next cell in the series. The area of the cell covered with ribbon interconnects does not emit light during EL imaging, so the average pixel intensities corresponding to these columns are low and the inverse of the pixel intensity creates a peak. The peaks corresponding to the ribbon interconnects are lower than the peaks corresponding to the cell spacing because they are narrower. The row-wise profile shows strong peaks corresponding to the row spacing. The lesser peaks do not appear because the interconnect ribbons only run in the vertical direction.

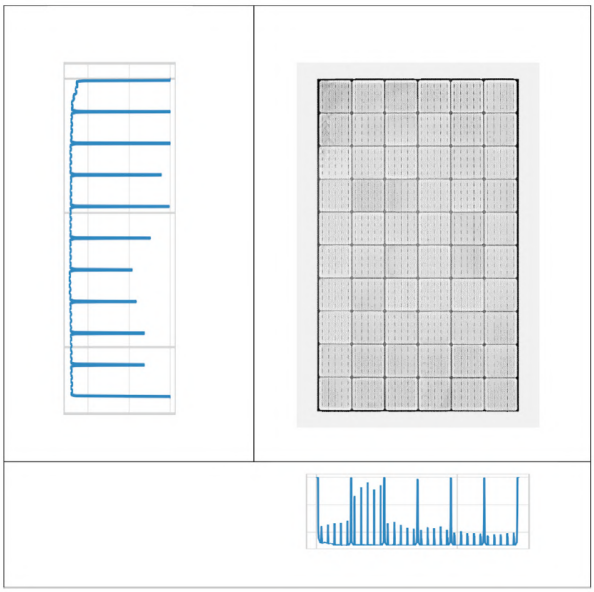
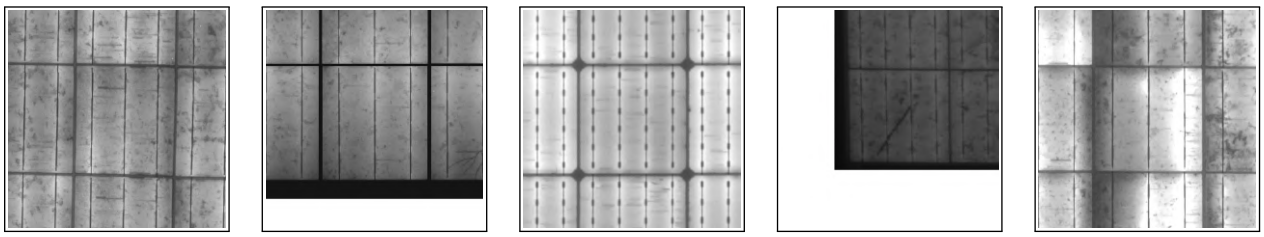


Figure 5.11: Profiles created from the row-wise and column-wise average intensities with peaks corresponding to cell spacing regions used to crop cells from the full module image.

Expanding cell crops

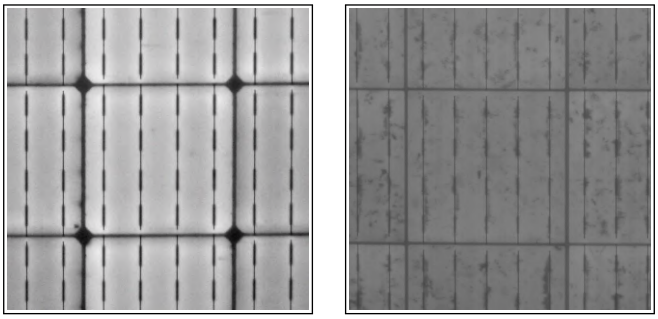
The cropping method was designed to encode additional information from the surrounding area of each cell. Rather than crop the cell-level images exactly at the four corners of the row and column spacing, an additional area around each cell was included. Other researchers focus on methods to crop the solar cell images at the cell-spacing boundary. In contrast, the SCDD cell-level images include the cell of interest, the cell spacing, and the half-cell area around the cell. Figure 5.12 shows examples of cell-level EL images in the SCDD dataset. The additional area includes the neighbouring cells for cells cropped from the centre of the module or 'padding' for cells cropped from the edges of the module.



(a) (b) (c) (d) (e)

Figure 5.12: Examples of cell images cropped from module-level EL images, including the surrounding cells and white space padding for edge cells and corner cells

The additional area surrounding the cells was included in the cropped image for three reasons. First, the model can be trained to differentiate mono-c-si versus multi-c-si cell types by the shape of the spacing object. Mono-c-si wafers are cut from round ingots that form 'pseudo-squares' with rounded corners, which create a diamond shape feature in the space where cell 'corners' converge. By contrast, multi-c-si wafers are cut from rectangular blocks that form square wafers with 90° corners, which do not create the diamond shape. Figure 5.13 shows the difference between a mono-c-si cell with rounded corners and a multi-c-si cell with square corners.

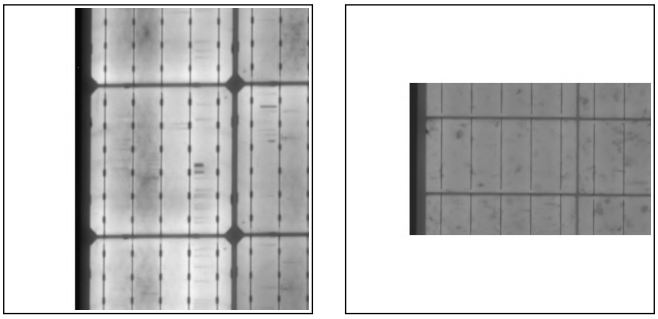


(a) Mono-c-si (b) Multi-c-si

Figure 5.13: (a) Mono-c-si cell with rounded corners and (b) multi-c-si cell with square corners

The second reason for including additional area surrounding the cells was to minimise the risk of clipping off part of the cell during the cropping procedure, as sometimes occurs when cropping just the cell itself. There are many examples in the literature where the cell edge is

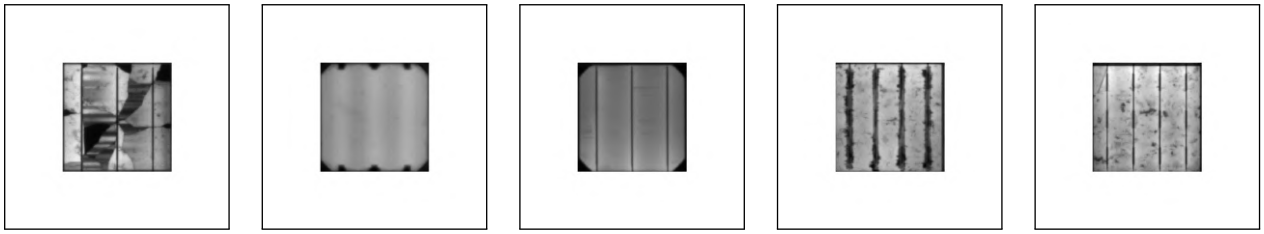
clipped. For cells along the edge of the module, a 'padding' layer was added and filled with white pixels to maintain the cell of interest at the center of the image regardless of where the cell was cropped. Third, the cropping method provides flexibility to ensure that each cell-level image has the same square dimensions without losing the aspect ratio for the objects in the cell-level image. For example, modules made from half-cut solar cells have entered the market in recent years. The half-cut solar cells are rectangular in shape, not square like the predecessors. Using the padding layer, the rectangular half-cut cells were cropped from the module-level EL images using the PeakUtils method after some relatively minor adjustments and then padded to create a square image that maintains the original aspect ratio for the cell. Figure 5.14 shows an example of a cell cropped from the left edge of a module built from full-sized, square cells (left) and a module built from half-cut cells (right), including the white space padding to maintain a square dimension without losing the aspect ratio of the cell features and defects. This method allowed solar cells from modules with different shapes and sizes can be combined into one dataset.



(a) Full cell (b) Half-cut cell

Figure 5.14: Examples of cell images cropped from the left edge of a module image built from (a) full-sized, square cells and (b) half-cut cells, including the white space padding to maintain a square dimension without losing the aspect ratio of the cell features and defects

The cropping method also created additional padding around each single-cell image downloaded from the ELPV and SDLE datasets. Padding was added to maintain the cell at the center of the image with a similar resolution as the cells cropped from the module-level images. The padding for these cells consisted of white space around all four edges to maintain the cell at the centre of the final image. In retrospect, this padding may not have been necessary. Resizing these images may be equally effective. Figure 5.15 shows examples of these single-cell images not cropped from the module-level EL images, including the white space padding around all four edges.



(a) (b) (c) (d) (e)

Figure 5.15: Examples of single-cell images not cropped from module-level EL images including the white space padding around all four edges

Annotating the ground truth images

A method for constructing ground truth masks was developed and refined for building the SCDD dataset. The ground truth masks represent the 'true' classification for each pixel in the EL image. The ground truth masks, together with the corresponding EL images, form the critical pairs needed for the training, testing, and validation of the model. The ground truth masks were generated by manually colouring a copy of the corresponding EL images so that the colour of each pixel corresponds to a specific defect or feature in the corresponding EL image.

A team of individuals from the CSIR, Wits University, and personal networks was assembled to create the ground truth dataset. A method for training and calibration of the annotators was developed. The steps are summarised as follows:

1. Collaborate with a CSIR colleague to train initial images
2. Write a work instruction, including a library of defects and classes
3. Convene an initial meeting with the larger team
4. Review the work instruction
5. Issue a set of common 'calibration' images for everyone to annotate
6. Review the annotations for the calibration images, focusing on the inter-rater differences
7. Revise the calibration images and review
8. Update the work instructions and discuss inconsistencie with annotators
9. Assign random images to each annotator, including one or two calibration images in each round
10. Develop a method using Python to identify mislabelled colours in each image
11. Annotate 400 images
12. Disband the full team
13. Designate one person as an expert annotator to review and edit all the images, as needed
14. All images were reviewed, and final versions were approved by me

15. Create new EL images and ground truth masks to grow the number and diversity of images in the dataset

16. Continuous editing on existing images, as needed

The team used GIMP 2.10.14 to colour the EL images according to a prescribed colour code. Many online tools are available for annotation [Rizzoli 2015], but none of the tools provided the same depth and flexibility as GIMP at zero cost. While the learning curve was steep for GIMP, the flexibility and broad scope of functions included in the tool were invaluable when fine-tuning the ground truth images. Figure 5.16 shows a random sample of ground truth masks from the SCDD dataset. Each pixel in the ground truth mask was classified as one of the intrinsic features or extrinsic defects by a human and coloured accordingly. The intrinsic features of the cell consist of the regions that are part of the solar module by design, for example, spacing between cells, ribbon interconnects, bus bars, and module borders. The extrinsic defects include cracks, inactive regions, belt mark defects, gridline defects, and corrosion. The padded regions were coloured olive, and any pixels not identified as one of the features or defects were coloured black for background.

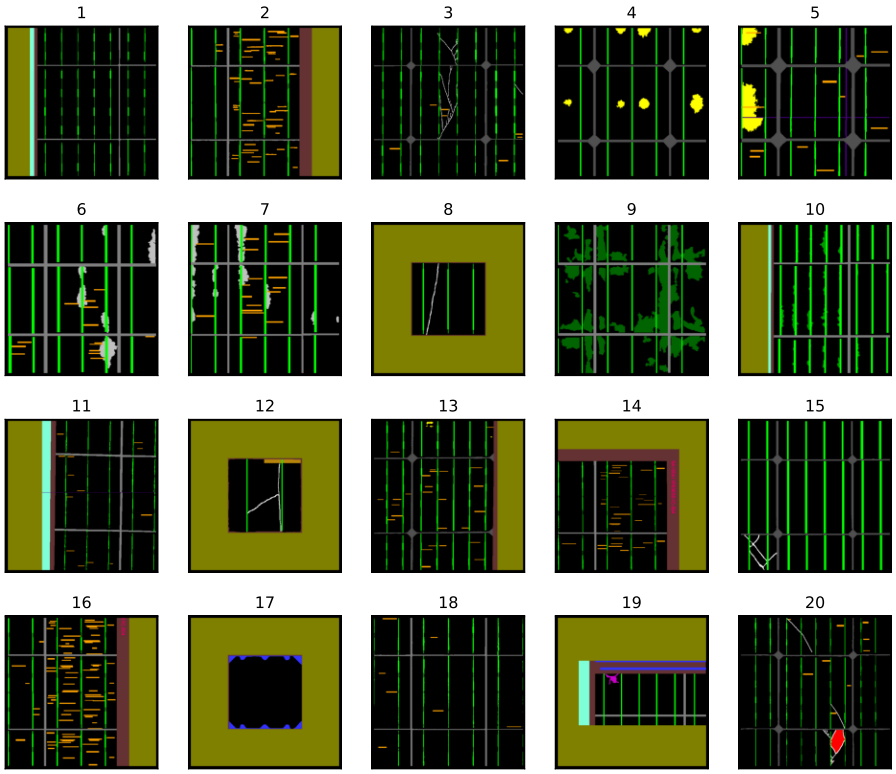


Figure 5.16: A random sample of ground truth masks from the SCDD dataset

Inter-rater variability in ground truth images

The ten team members were assessed for inter-rater variability to highlight differences and reduce variability in the ground truth masks. Each team member was assigned a common set of calibration images for annotation, and the outputs were reviewed in meetings. The calibration images were randomly selected from the database of cell-level images and distributed to all members of the team. As a PV expert with over 15 years of experience analysing EL images, I annotated each calibration image and calculated the precision and recall for each image and each team member. The review meeting then focused on the features and defects with high inter-rater variability. Individual guidance was also given to team members at the low end of distributions. Figure 5.17 shows the results from one round of calibration reviews. The inter-rater variability was low for the larger features like background and padding, as expected. However, the variability in the spacing layer for multi-c-si was also low relative to the variability in the spacing layer for the mono-c-si layer. This was rather unexpected, so the meeting focused on standardising the method for annotating the mono-c-si layer. Some labellers did not include the rounded corners in the cell spacing layer, which contributed to the higher variability for mono-c-si spacing. The figure also highlights the relatively low score for recall of the two defects listed compared to the precision. The analysis highlighted that many labellers were only annotating the pixels at the centre of the object and not annotating the pixels at the border of the object where the greyscale was transitioning from dark to light. The PV expert did include the transition pixels at the borders for all objects, so guidance was then given to all team members to include the additional border pixels. The anti-aliasing setting in GIMP was also a source of error, creating a blurred edge for labellers who selected the option.

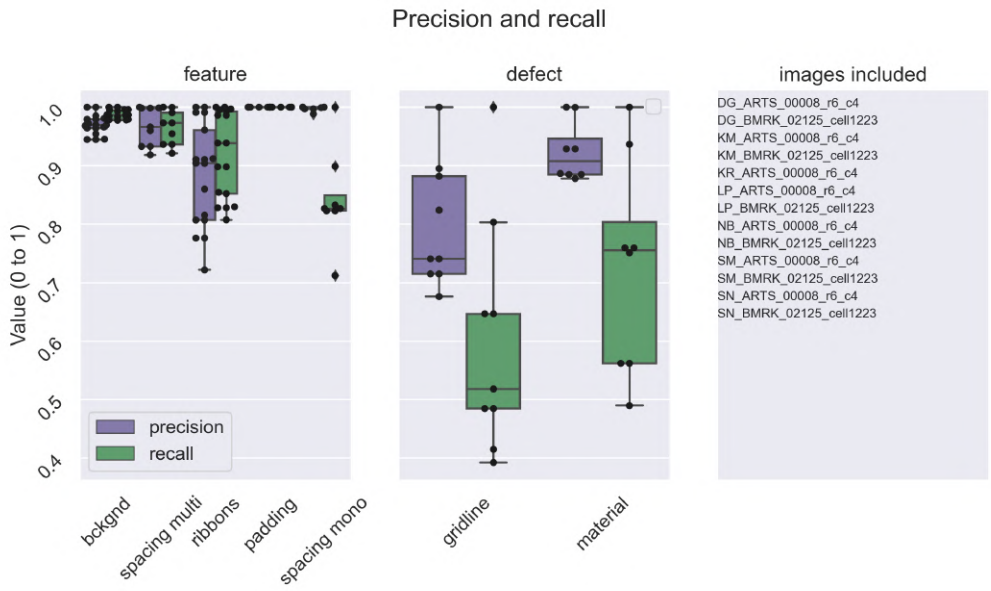


Figure 5.17: Precision and recall distributions on a set of calibration images created to identify and reduce inter-rater variability

The precision and recall metrics for each team member also depended on the quality of the ground truth masks produced by the PV expert. Accurate ground truth masks are critical for success. For example, Figure 5.18 shows the crack layer (white pixels) cropped from a full cell. The prediction mask from the U-Net is coloured green. The remaining white pixels indicate

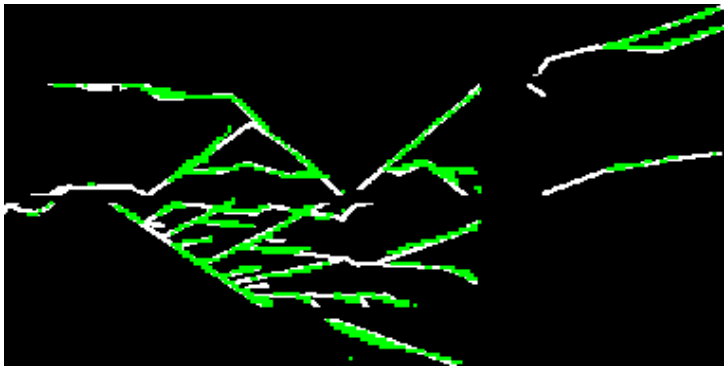


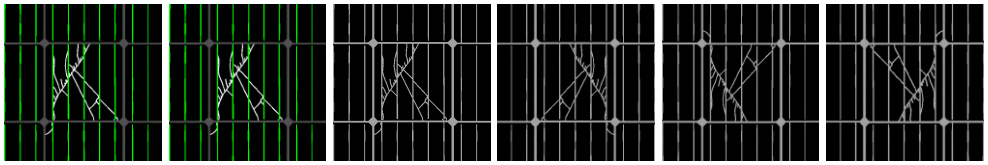
Figure 5.18: Crack layer for CFVS-00035-r8-c4 showing the prediction from the U-Net model (green) and the underlying ground truth mask manually coloured (white).

cracks not detected by the model, thus contributing to low recall (48%). A closer inspection of this image also uncovered issues with the accuracy of the ground truth mask. The annotation in the ground truth mask did not always cover the crack defect seen in the EL images. Annotation for the crack layer was a tedious process, and sometimes the ground truth label did not follow precisely the crack lines, leading to some misalignment. In these cases, the computer-generated prediction mask was sometimes more accurate than the human-generated ground truth mask, leading to a slight misalignment between the model prediction and the ground truth. Errors in the ground truth masks were updated regularly as they were discovered.

Train, test, and validation datasets

Finally, all the cell images and ground truth masks were resized and coded. The EL images and the corresponding ground truth masks were resized to a standard shape of 512 pixels x 512 pixels. The shape of the original images varied depending on the source. The resized images created another challenge when downsizing. The cell-level images cropped from module-level images were saved with 1200 x 1200 resolution and annotated. Upon downsizing, crack layers in the mask sometimes became disconnected or disjointed as white pixels were replaced with black background pixels. Dilating the original crack layer in GIMP by one or two pixels to include more border transition pixels resolved this issue. The ground truth masks were finally coded so that each pixel was assigned a unique number from 0 to n-1, where n is the number of unique classes. A detailed list of the classes is provided in the annexe.

The cell-level EL images and corresponding ground truth masks were then divided into three separate groups: training, validation, and testing. The training and validation datasets are used during model training, and the test dataset is used to measure the accuracy of the final model. Each group consists of matched pairs of one EL image and the corresponding ground truth mask. The split into training, validation, and testing is typically done during the training using a random selection that allocates 80% to training, 10% to validation, and 10% to testing. In the SCDD, however, the test dataset was created from a random selection of all the available cells with cracks or inactive regions. In the first version of the SCDD dataset, only cracks and inactive regions had a reasonable sample size for training and statistical analysis of the test dataset. Gridline defects were widely represented, so no special selection criteria were included for gridline defects. Following the creation of the test dataset, the remaining cells were randomly assigned to the training and validation datasets. The split for the 20221008



(a) (b) (c) (d) (e) (f)

Figure 5.19: Example showing the pre-processing and augmentation of one original ground truth mask: a) original GT, b) resized GT, c) coded, d) flipped horizontal, e) flipped vertical, and f) rotated 180 deg to produce four ground truth masks from each original

version of the dataset was 79.6% training, 10.1% validation, and 10.3% testing. The split remained constant so that the models would be trained and tested on an identical datasets.

Data augmentation

The training images and masks were augmented after the split. Augmentation refers to a standard practice in machine learning that takes the original image and creates variations based on the original to increase the size of the training dataset in a meaningful and useful way. In this experiment, each training image and corresponding ground truth mask was augmented using a horizontal flip, a vertical flip, and a 180° rotation which is equivalent to a horizontal flip followed by a vertical flip. Figure 5.19 shows a sequence of images for one representative cell from the training dataset, including the original ground truth mask, resized, coded, and augmented. The augmentation occurred only after the split to avoid data leakage, i.e., none of the augmented images from the training dataset appeared in the validation or test datasets. The validation and test sets were not augmented. The split and the augmentation were handled separately from the training so that the same dataset would be used across all the models tested. Each EL image and corresponding ground truth mask yielded a total of four images and masks for training.

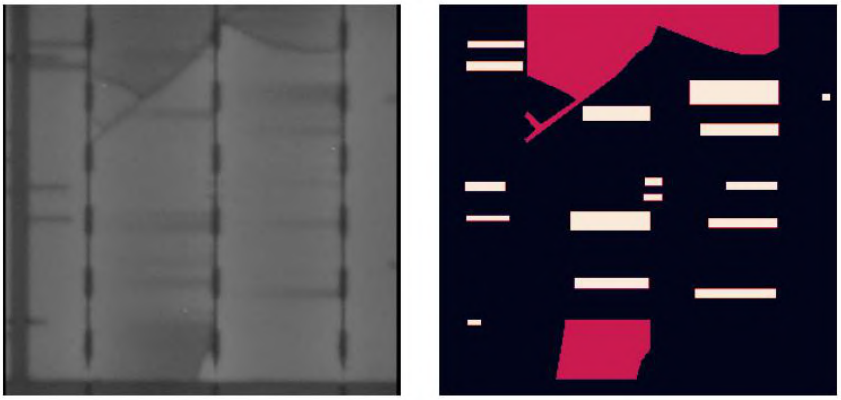
5.3 New unlabelled dataset for SCDD

The CSIR manages a database of EL images. The database contains 263,510 cell-level images as of October 21, 2023, and the database continues to grow every year. The 765 annotated images were selected from the full dataset. 209,843 images (80%) were recorded at the PV Module Quality and Reliability Laboratory in Pretoria, South Africa. More than 150,000 images unlabelled images were published along with the annotated SCDD dataset previously published on github[8] for future research focused on advancing self- and semi-supervised training techniques in this field. The remaining images were not published due to confidentiality agreements.

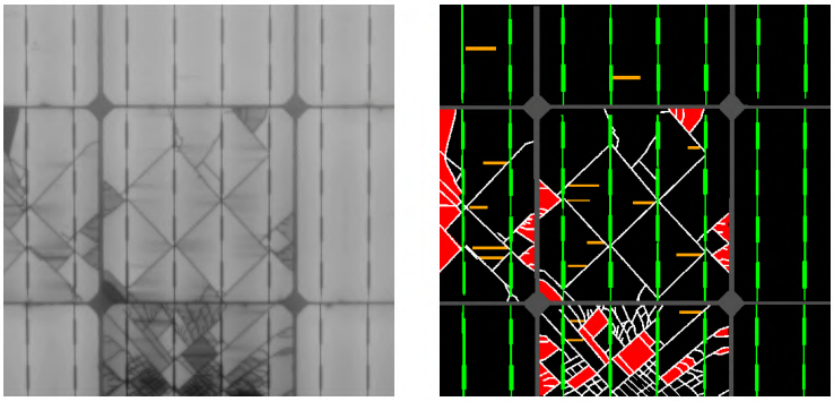
[8] https://github.com/TheMakiran/BenchmarkELimages

5.4 Comparison of SCDD vs UCF datasets

Figure 5.20 shows examples of EL images and ground truth masks from UCF [Fioresi *et al.* 2022] and the SCDD dataset [Pratt *et al.* 2023]. The UCF images exclusively label the extrinsic defects. In contrast, the ground truth images in the SCDD dataset include the extrinsic defects and the intrinsic features of the solar cells, such as cell spacing (grey) and ribbon interconnects (green). By having the intrinsic features in the masks, the potential regions for defect detection are reduced because the feature areas are essentially masked. The model is trained to see the full image, not just the defects, mimicking the human eye-brain activity when viewing an EL image. The additional features also provide more context for the model to learn about the scene. For example, gridline defects will always run perpendicular to the ribbon interconnects, and that context should improve module predictions provided that the model has enough depth. The ground truth images in the SCDD dataset also tend to show finer resolution for cracks (thin white lines) and gridline/contact defects compared to the UCF dataset. The gridline/contact defects are represented by the thin orange rectangles in the SCDD dataset, while they appear as larger rectangular regions in the UCF ground truth. Finer features are more challenging to detect than larger features because they occupy a smaller region in the overall image and a higher percentage of border pixels.



(a) UCF dataset



(b) SCDD dataset

Figure 5.20: EL images and ground truth masks from (a) the UCF dataset [Fioresi *et al.* 2022] and (b) the SCDD dataset [Pratt *et al.* 2023]

Figure 5.21 shows an EL image from the UCF dataset and the published ground truth mask. The EL image was also annotated according to procedures for developing the SCDD dataset. The colours used in the UCF were assigned to the UCF to match the colours used in the SCDD dataset to represent cracks (white) and gridline defects (orange) for easy comparison. In the UCF paper, the gridline defects are labelled as 'contact' defects. The UCF ground truth masks often include one large region encompassing all the cracks in what the authors call a 'conglomeration.' By contrast, the SCDD ground truth annotates the crack lines individually, resulting in the thin white lines in the ground truth mask. The orange rectangles represent the gridline/contact defects. The gridline/contact defects are similar in both datasets, although they tend to be narrower with a higher aspect ratio in the SCDD ground truth mask. Along with the prediction mask, the UCF system also outputs a json file with the percentage of pixels associated with each defect in the prediction mask for each cell, which can be helpful for further processing the output of the predictions. The ground truth masks in the SCDD dataset provide a higher resolution for defects, leading to more accurate analysis when quantifying the defects during downstream tasks.



(a) EL image (b) GT it UCF dataset (c) GT in SCDD dataset

Figure 5.21: (a) EL image from the UCF dataset, (b) the corresponding ground truth mask from UCF, and (c) the ground truth mask as annotated by the method used for SCDD dataset

5.5 Conclusions

Specific findings from this chapter include:

- A new dataset was required to enable research in semantic segmentation models for the EL images of solar cells. The UCF dataset was under development in parallel, but not available publicly at the start of this research.
- The new SCDD dataset contains more detail with respect to the size and shape of cracks and gridline defects compared to the UCF dataset.
- Creating the ground truth images in GIMP results in high-resolution masks, but it does require 30-60 minutes per image.
- Consistency in labelling across multiple labellers was a challenge but mitigated with the use of calibration images and a single advanced labeller who reviewed and edited all images.

Chapter 6

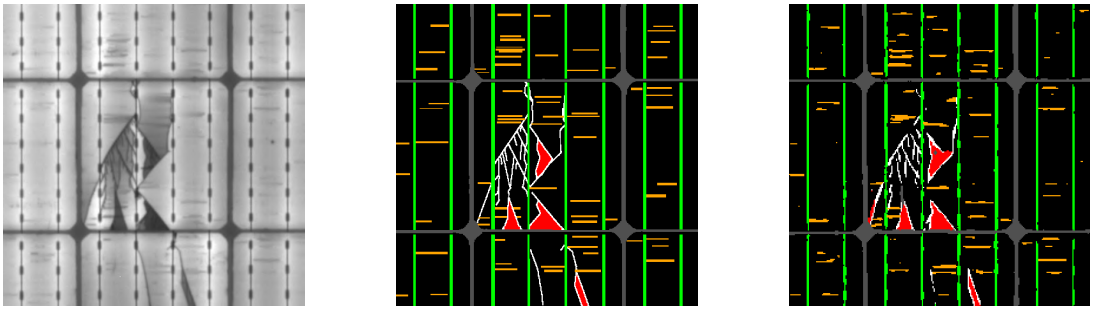
Proof of concept using Semantic Segmentation for SCDD

6.1 Introduction

Chapter 6 summarises the research and results from the first published article in which we proposed semantic segmentation (SemSeg) for SCDD [Pratt *et al.* 2021a]. The key contribution from SemSeg lies in the pixel-level classification of images [Hao *et al.* 2020; Ulku and Akagunduz 2019], which enables the translation from an unstructured image to a structured dataset. The resulting pixel-level classification captures more detailed information embedded in EL images than image classification or object detection. Time will tell if the additional complexity of a semantic segmentation model provides a significant improvement for correlating defects in EL images to module power output. Based on the results from this chapter, the predictions capture enough information from the EL images to correlate with module power loss.

SCDD from EL images using multi-class semantic segmentation was a novel idea in 2020 when this research began. This article was the first in a series of publications stemming from the research conducted for this thesis. The research aimed to develop a semantic segmentation model to detect defects in EL images from a broad range of cell architectures that was independent of the equipment used to generate the EL images, the wafer-based module design, and the image quality. To that end, we developed an initial training dataset, experimented with a U-Net model, and correlated the output to PV module power output.

SemSeg has been the topic of several publications across multiple domains. Figure 6.1 shows an example of semantic segmentation in the EL image domain[Pratt *et al.* 2021a]. Each pixel in Figure 6.1a is assigned to a specific class by the SemSeg model and represented by a unique colour in the output prediction 6.1c. The ground truth image in Figure 6.1b is the target representation the model tries to learn during training on the full set of training images and corresponding ground truth images.



(a) EL image (b) Ground truth (c) Prediction

Figure 6.1: Example of semantic segmentation for (a) an EL image of a solar cell, (b) the ground truth annotated by a human, and (c) the prediction from a model [Pratt *et al.* 2021a]

6.2 Dataset

The solar cell defect detection (SCDD) dataset was curated from five sources so the model would learn from a broad range of cell architectures, imaging equipment, pre-processing steps, and defect types. The dataset consisted of only 148 cell-level EL images with ground truth masks at this early stage. 108 images (73%) were assigned for training, 10 images (7%) were assigned for validation, and 30 images (20%) were assigned for testing. Each training image and corresponding ground truth mask was then augmented using a horizontal flip, a vertical flip, and 180° rotation. The annotated dataset included classes for twelve (12) intrinsic features of solar cells and eleven (11) extrinsic defects coloured by a human using GIMP software. The intrinsic features of the solar cells were included to provide more context for model training to minimise confusion between features and defects. A more detailed description of how the dataset was constructed is included in Chapter 5.

6.3 Method

The U-Net model was selected for experimentation because of its success across a broad range of domains with limited samples [Ronneberger *et al.* 2015]. The model was adapted for EL images based on computer code published on Github [Gupta D 2021]. The model consisted of a VGG-16 encoder and the U-Net decoder. The model was trained using the adaptive moment (Adam) optimiser with a learning rate = 0.001 and the categorical cross-entropy loss function as implemented in TensorFlow 2.8.0 [Abadi *et al.* 2015] and Keras [Chollet and others 2015]. The model hyperparameters were not optimised specifically for this new domain. The class weights were equal for all defects and features. The initial weights were transferred from ImageNet and then fine-tuned on the SCDD dataset. The model was trained up to 500 epochs, beyond the point at which the model stopped learning based on trends in the loss and IoU graphs. The best training epochs were used for further testing and comparisons.

6.4 Results

The experimental results proved that high-fidelity prediction masks for multi-class defect detection were possible using a U-net architecture over a broad sample of cell architectures. Precision and recall were higher for feature detection than defect detection for multi- and mono-c-si cells. Precision and recall were the lowest for defect detection on multi-c-si cells. The research also showed that correlations could be established between the change in the number of pixels associated with cracks, inactive areas, and gridline defects with the EL images taken during a sequential accelerated test designed to provoke failures that can be seen in the EL images. Future work points towards improving the prediction machine and correlating the statistical data extracted from the semantic segmentation prediction masks to electrical performance on large batches of PV modules, as described in Chapter 9

The U-Net produced reasonably accurate predictions for the targeted features and defects in EL images of solar cells despite the small dataset and many classes. Figure 6.2 shows examples of cell-level EL images and corresponding predictions from the U-Net model. The background (black), padding (olive), ribbon (green), and cell spacing (two shades of grey) classes were all detected with reasonable accuracy. However, the dogbone spacing class (blue) was often confused with the other spacing layers. Note that it was also underrepresented in the training dataset compared to the spacing layer for multi-c-si and mono-c-si. The inactive area (red) and the gridline defects (orange) were also predicted with reasonable accuracy. The cracks (white) predictions were in the correct locations, but the lines were thin and had discontinuities, leading to low recall.

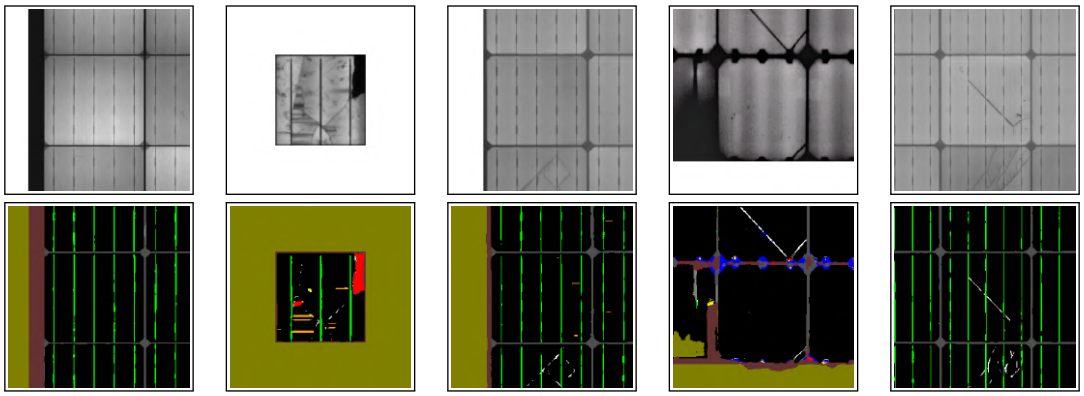


Figure 6.2: Examples of cell level EL images (top) and corresponding predictions (bottom) from a U-Net model trained on 108 images

Figure 6.3 shows the distributions of the quality metrics for predictions for the 30 test images. Several observations can be made based on this figure. First, the quality metrics for the features tend to be higher than the defects. This may be due to the larger sample sizes or because the features tend to be larger objects. Second, the precision and recall tend to match for a given class and cell type, e.g., the precision and recall distribution for the ribbon class are nearly identical. Third, the quality metrics for most features are similar for mono-c-si and multi-c-si cells. The background, padding, and ribbon classes are well-matched, but the spacing class is lower for multi-c-si. The defect classes are significantly lower for multi-c-si compared to

mono-c-si. The grain boundaries in the multi-c-si cells may impact the model performance. Fourth, the distributions of IoU scores are lower than the F1 scores across all classes.

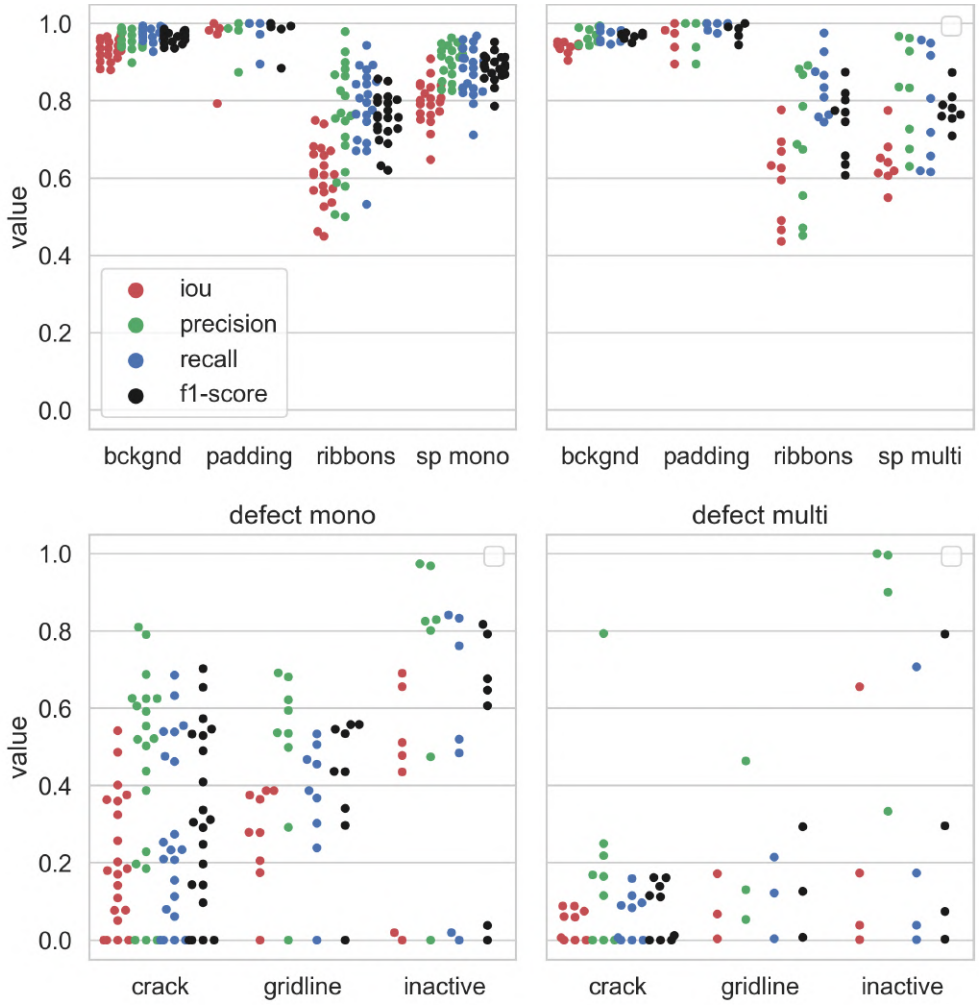


Figure 6.3: Distributions of model performance metrics on 30 test images using the U-Net model trained on a small dataset

The U-Net model was applied to a series of EL images recorded on modules exposed to accelerated stress tests. Figure 6.4 presents EL images of a single module exposed to a sequential accelerated stress test at CFV Test Labs in Albuquerque, New Mexico. The first image shows the state of the module before any stress testing. The following image shows the module's state after the static mechanical load test (SMLT), designed to induce micro-cracks in the cells. The subsequent images show the same module after thermal cycling and humidity freeze test exposures.

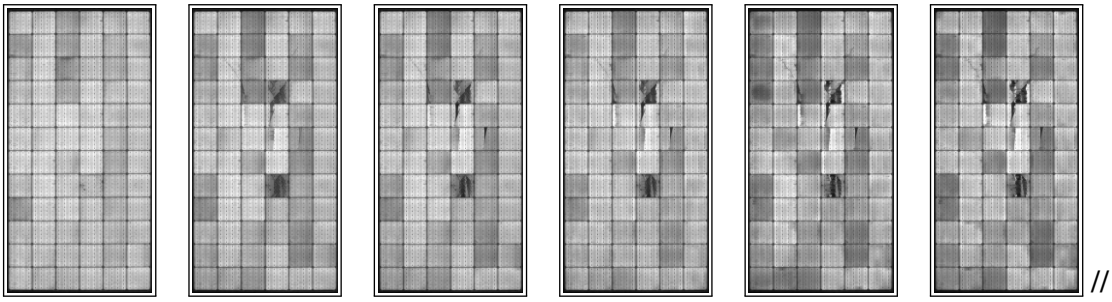
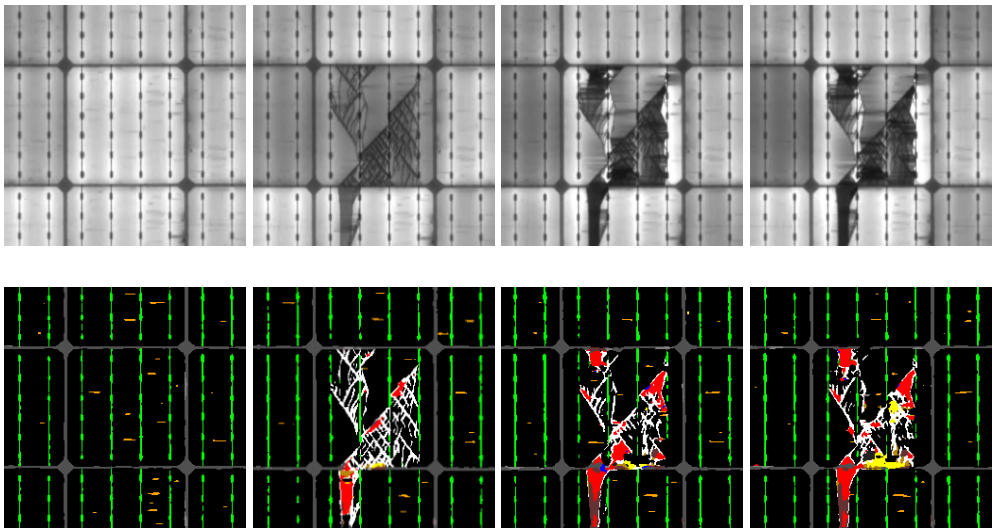


Figure 6.4: EL images of a single module taken at several stages of an accelerated stress test sequence showing initial cell cracks after SMLT and gradual darkening of damaged areas with subsequent tests.

A visual inspection of the images in Figure 6.4 shows cracks and inactive areas developed in several cells after the SMLT. The subsequent images show a darkening of the damaged areas with subsequent tests due to the thermal expansion and contraction from the environmental stresses. Figure 6.5 illustrates the changes in the prediction masks for one of the most damaged cells cropped from row 4, column 4. The cracks (white) and inactive areas (red) seen in the prediction mask post SMLT (Figure 6.5b) are good representations of the cracks and inactive areas seen in the corresponding EL image. The subsequent prediction mask in Figure 6.5c shows an increase in inactive areas (red) corresponding to darkening in the corresponding EL image. However, the prediction mask in Figure 6.5d shows increased material defects (yellow) instead of the original cracks and inactive areas from the previous prediction.



(a) EL image (b) Post SMLT (c) Post HF10 (d) Post HF20

Figure 6.5: Impact of accelerated stress testing as seen in the original EL images (top) and prediction masks from the U-Net model (bottom) for cell r4-c4 at (a) initial, (b) post SMLT, (c) post HF10, and (d) post HF20

The summary statistics extracted from the U-Net model predictions correlated with the observations made in the visual inspection. Figure 6.6 shows the least squares mean estimates for the cracks and inactive areas over the sequential test sequence. JMP statistical software was used to

fit a two-way analysis of variance (ANOVA) model, including module ID and cell ID as independent variables, for each defect and generate the least squares means plots. The y-axis shows the percentage of pixels predicted for each class, averaged across the 72 cell-level predictions in each module ID. The x-axis shows the module ID assigned to the image after each sequential test, although the physical module is the same. The LS means plot shows a statistically significant increase in the percentage of pixels associated with cracks after the SMLT, correlating to the change seen in a visual inspection of the EL image. The crack percentage then decreases with subsequent testing. By contrast, the percentage of pixels associated with inactive areas decreased after SMLT and then increased thereafter. The initial decrease was not expected and was rather likely due to false positives for inactive areas picked up on the initial prediction. The increase in inactive areas thereafter is expected and correlates to the subtle changes observed during the visual inspection, especially for the cells that the SMLT most impacted.

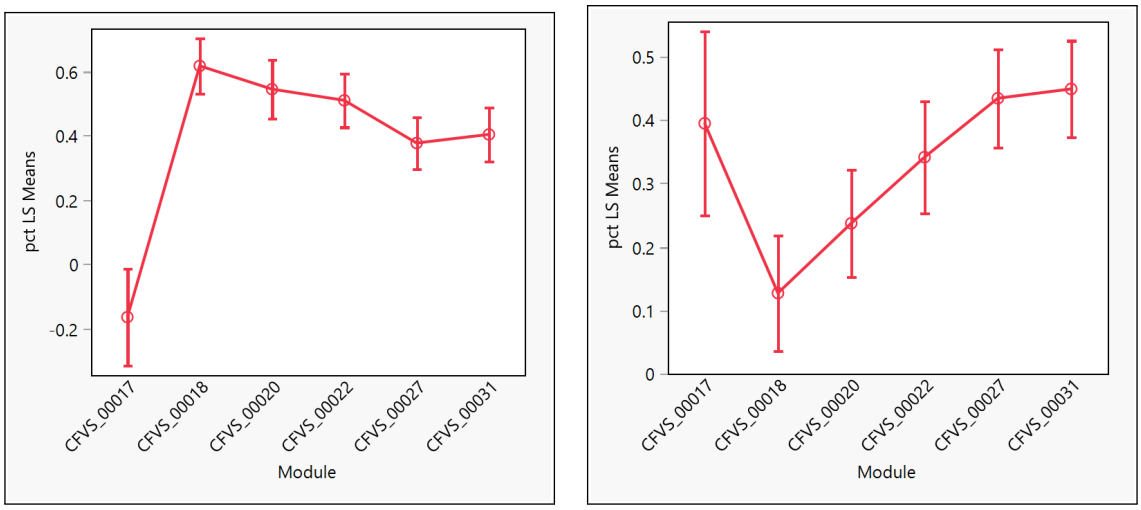


Figure 6.6: Least squares means for the cracks (left) and inactive areas (right)

The statistical summary of the model predictions was used to quantify the changes in EL images of modules subjected to accelerated stress tests in a PV module lab. The work focused on developing and testing a deep learning method that is independent of the equipment used to generate the EL images, independent of the wafer-based module design, and independent of the image quality for broad applicability on out-of-domain images. Prior work in the research field was largely limited by the variety of cell architectures, the extensive pipelines for detailed pre-processing, and the number of defects in the multi-class detection model.

The output of the SemSeg model can be used to quantify the extent of each defect by counting the pixels classified to each defect class. These statistics can be used to analyse EL images from accelerated stress tests in the lab, to rank batches of PV modules for quality metrics, and for plant-level inspection analysis. Pratt *et al.* [2021a], which originated from this work, was the first published work to use SemSeg for multi-class SCDD. Fully-supervised, semi-supervised, and self-supervised models were trained and evaluated on a new benchmark dataset.

6.5 Conclusions

Semantic segmentation showed potential for detecting, classifying, and quantifying features and defects in EL images of PV modules made from mono- and multi-crystalline silicon cells. The model performed better on large intrinsic features, such as cell spacing and ribbon interconnections, compared to the smaller extrinsic defects, such as cracks and gridline defects. For defects, the model performed better on the cells made from mono-c-si cells compared to cells made from multi-c-si cells. The recall on inactive, crack, and gridline defects for cells made from mono-crystalline silicon wafers reached as high as 84%, 69%, and 53%, respectively, with minimal false positives. The output statistics from the model also correlated with changes observed in a sequence of EL images taken over several stages of an accelerated stress test experiment conducted at CFV Test Labs.

Semantic segmentation of EL images is well suited for SCDD and large-scale statistical analysis of defects in millions of PV modules. Automation of the process fundamentally changes the amount and type of information that can be extracted from EL images, which opens opportunities for new analytics to correlate defects to electrical performance, compare module quality from different suppliers, and quantify changes in defect patterns over time. Semantic segmentation can also be applied to other renewable energy technologies because of the model's ability to convert any visual image to a numerical array that can be further analysed by engineers and scientists in their respective fields to address other research questions and challenges.

Specific findings from this chapter include:

- Semantic segmentation shows significant promise for the field of SCDD.
- The U-Net provided a good starting point despite a relatively small dataset and no hyperparameter optimisation.
- The output from the U-Net can be used to quantify meaningful changes in EL images taken over the course of an accelerated stress test sequence because the actual size of a defect can be accurately quantified.
- The accuracy of the ground truth images impacted precision and recall, so the images in the test dataset should get extra special attention with respect to the accuracy of the labels.

6.6 Next steps

The next steps will focus on expanding the SCDD dataset, experimenting with class weights, and evaluating additional fully-supervised models for SCDD.

Chapter 7

Fully supervised deep-learning models

7.1 Introduction

Chapter 7 summarises the research and results from the second published article based on an expanded benchmark dataset and experimental results for automatic detection and classification using four selected fully-supervised deep learning models [Pratt *et al.* 2023]. A new benchmark dataset specifically for SemSeg and benchmark metrics was published. This research aimed to evaluate additional fully-supervised models using different class weights on the expanded SCDD dataset to improve predictions. We also introduced class weights to assess the impact on smaller defects and features.

7.2 Dataset

The original SCDD dataset was expanded to include 582 images and ground truth masks. The number of labelled images exceeded 600, but some were excluded following a quality control review. Fifty images with a prevalence of cracks, gridline defects, and inactive areas were assigned to testing, evenly split for mono-c-si and multi-c-si wafers. Fifty-four images were randomly selected from the remaining images for validation. The remaining images were assigned for training and augmented using a 180° rotation, mirror, and flip, yielding 896 mono-c-si cell images and 1016 multi-c-si cell images for a total of 1912 images in the training dataset. Overall, 478 images (82%) were assigned to training, 54 images (9%) were assigned to validation, and 50 images (9%) were assigned to test. The dataset included the original set of 12 defect classes and 12 feature classes.

7.3 Method

Figure 7.1 summarises the methodology followed for this chapter. The EL images and ground truth labels were curated, split into subsets, augmented, and trained using four different architectures and three sets of class weights. The models were used to generate a prediction mask for each image in the test dataset in which each pixel was labelled from 0 to 23, corresponding to the predicted class. The mask was coloured to facilitate the visual analysis of each mask.

The median IoU was computed for each image/class combination and plotted for comparison. The EL images and the corresponding ground truth masks were made public on github.

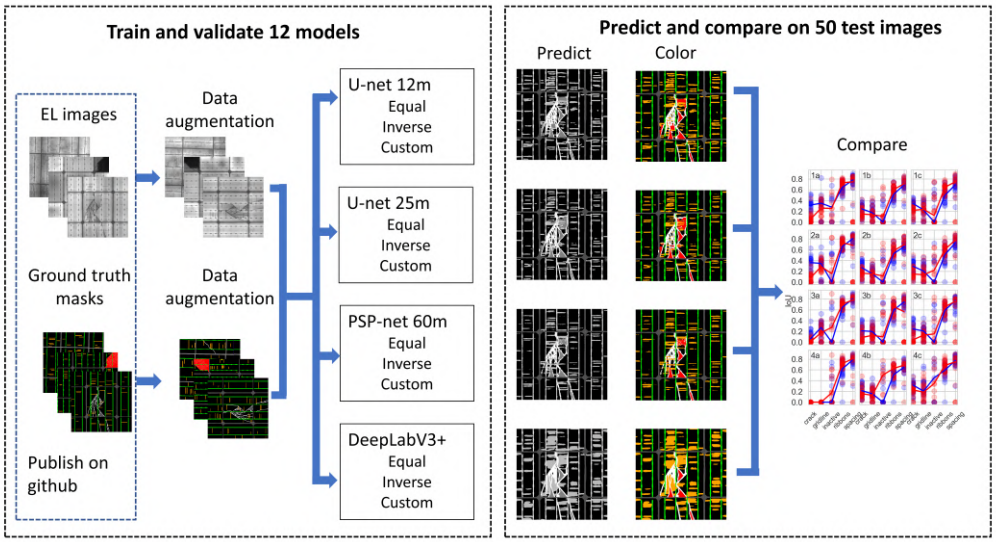


Figure 7.1: Graphical summary of the method

Multiple deep-learning models were trained on the dataset. Table 7.1 summarises the 12 fully supervised deep learning models trained and tested to identify the best-performing architecture for this benchmark dataset. The model ID is used in the subsequent analysis to reference each of the 12 models. The Python code for training each model was adapted from code published on the respective GitHub repositories [Gupta D 2021; Kamikawa Y 2020; Tomar N 2022; iakubovskii P 2020; Kawakita S 2021]. The U-Net was included because it performs well with a small set of labelled data [Ronneberger *et al.* 2015], and it pioneered the symmetric encoder-decoder model with skip connections [Hao *et al.* 2020], which forms the basis of many current deep learning models. A second, deeper U-Net was included to determine if the additional trainable parameters would result in better predictions than the smaller U-Net. The PSPNet was selected for the depth and complexity of the model, which includes a novel pyramid pooling module to improve complex scene understanding [Zhao *et al.* 2017]. DeepLabv3+ was selected because it was expected to show improved predictions, especially along the object borders [Chen *et al.* 2018b].

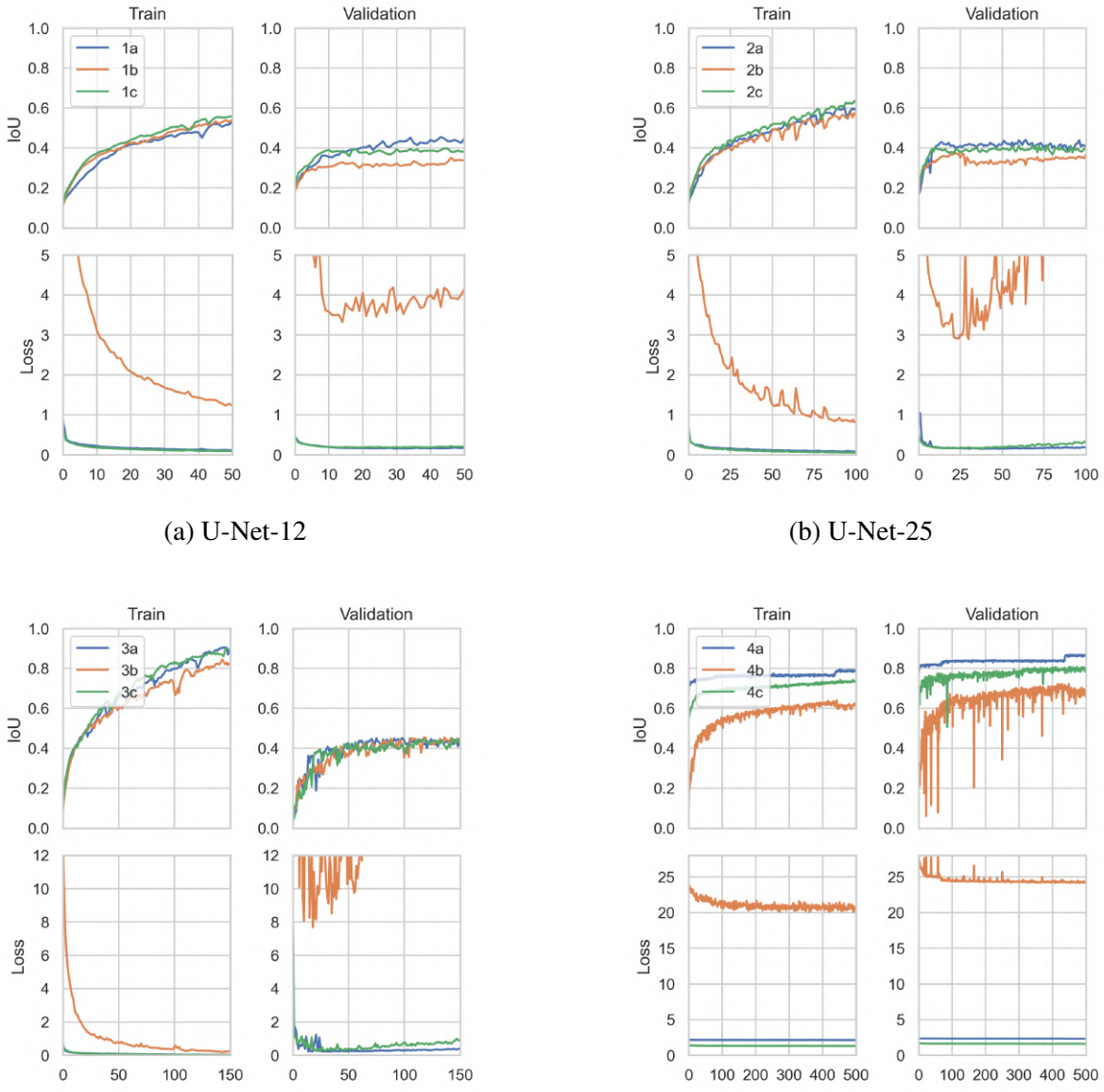
Table 7.1: Model ID and descriptions

ID	Class Weights	Architecture	Github Authors	Trainable Parameters
1a 1b 1c	Equal Inverse Custom	U-Net	Divam	12,333,720
2a 2b 2c	Equal Inverse Custom	U-Net	Tomar	25,858,887
3a 3b 3c	Equal Inverse Custom	PSPNet	Kamikawa	58,038,784
4a 4b 4c	Equal Inverse Custom	DeepLabv3+	Yakubovskiy and Kawakita	22,443,368

The models were trained with the same default settings, rather than optimise hyperparameters for each model individually. All models were trained with the adaptive moment (Adam) optimiser with a learning rate = 0.001 and the categorical cross-entropy loss function. The U-Net and PSPNet models were implemented in TensorFlow 2.8.0 [Abadi *et al.* 2015] and Keras [Chollet and others 2015]. The DeepLabV3+ model was implemented in PyTorch 1.10.0+cu113 [Paszke *et al.* 2019]. The ImageNet pre-trained weights were used to initialise the models.

Three different sets of class weights were evaluated on each model: A) equal, B) inverse, and C) custom. Set A assigned an equal value of 1 to all class weights. Set B assigned a value for each class equal to the inverse of the median percentage of pixels in each class across all images in the training dataset, which gives a higher weight to smaller defects. Set C class weights were based on engineering judgment and assigned unique weights only to specific features and defects of interest. Additional details and equations for class weights were described in detail in Chapter 4. A statistical design of experiments was run to evaluate the performance of the class weights, also described in Chapter 4.

Figure 7.2 plots the training and validation logs from all twelve models. Both U-Net models finished training within the first 65 epochs. The PSPNet models finished learning within 150 epochs, while the DeepLabv3+ models trained for up to 850 epochs before converging on the best weights. The training and validation losses were highest for the inverse class weights for all four architectures.



(a) U-Net-12

(b) U-Net-25

(c) PSPNet

(d) DeepLabv3+

Figure 7.2: Trends of median IoU and loss during training for (a) U-Net-12, (b) U-Net-25, (c) PSPNet, and (d) DeepLabv3+. Each subplot overlays the trends for [a] equal class weights, [b] inverse class weights, and [c] custom class weights

7.4 Results

Figure 7.3 presents the prediction masks for the EL image of one cell (CFVS-00035-r8-c4) from all 12 models. This cell was selected for the visual evaluation because it contains the three defects of primary concern: cracks, gridline defects, and inactive areas. The cell ID indicates that the cell was cropped from Module 00035, row 8, and column 4. The rows in Figure 7.3 correspond to the model: U-Net-12 (row 1), U-Net-25 (row 2), PSPNet (row 3), and DeepLabv3+ (row 4). The columns correspond to class weights: (a) equal class weights, (b) inverse class weights, and (c) custom class weights. All four models generate similar prediction

masks across the three sets of class weights, i.e., the four rows look similar. They all detect cell spacing (grey) and ribbons (green) well. However, the models with equal class weights predict fewer and smaller cracks compared to the models with inverse class weights and custom class weights. The prediction masks generated from inverse class weights and custom class weights tend to show dilated cracks and gridline defects. The dilation leads to a reduction in the precision and IoU. However, the dilation leads to higher recall and allows the user to visualise the cracks more easily when scanning a large batch of predictions.

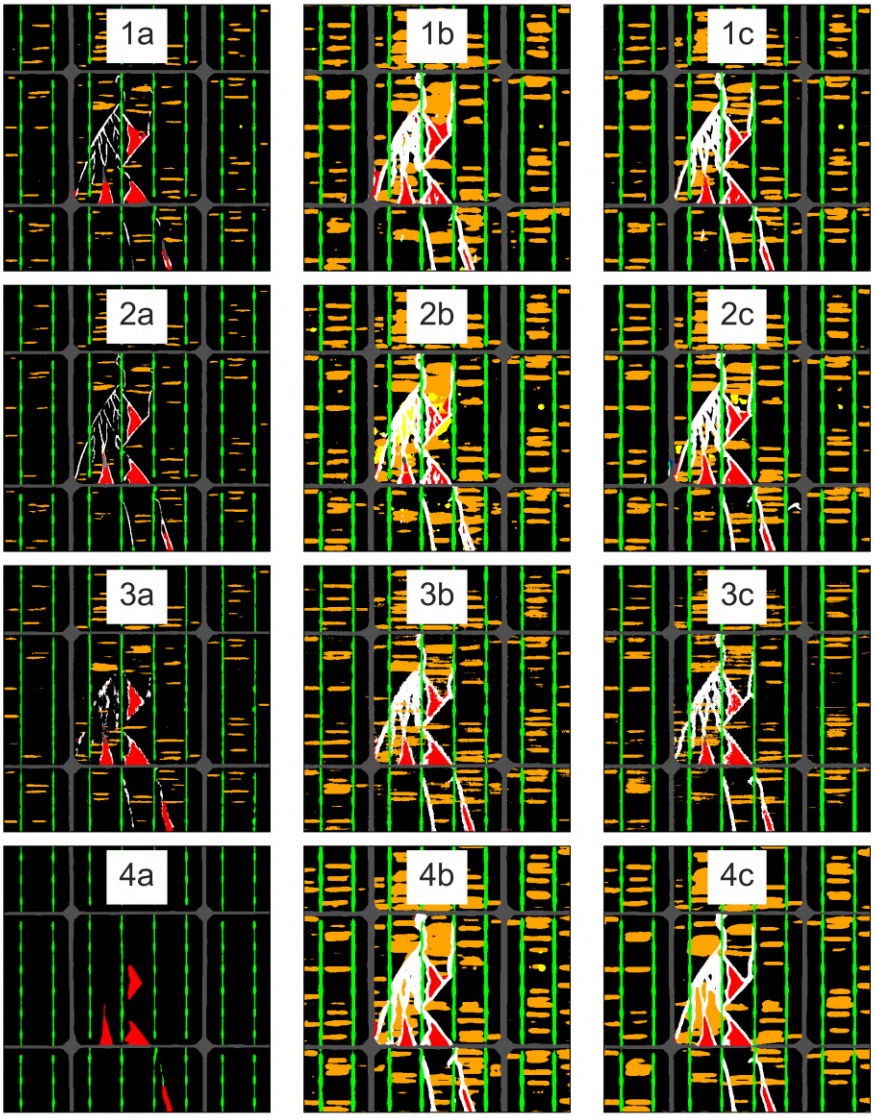


Figure 7.3: Prediction masks for cell CFVS 00035-r8-c4 from 12 models: U-Net-12 (row 1), U-Net-25 (row 2), PSPNet (row 3), DeepLabv3+ (row 4), equal class weights (column a), inverse class weights (column b), and custom class weights (column c)

Figure 7.4 shows the IoU and mIoU for three defects and two features on the predictions from the 12 models for the entire test dataset. The grid layout follows the layout in Figure 7.3 to facilitate a visual comparison of one indicative sample image and the combined results for the test dataset. The lines connect the mIoU for each class, and the colours correspond to

mono-c-si (blue) and multi-c-si (red) wafers. Generally, the models perform equally well for mono-c-si and multi-c-si cells when predicting ribbons and spacing. All models detect the features (spacing and ribbons) better than the defects (cracks, gridlines, and inactive areas). This correlates to the relative size of the features and defects. The features account for a greater number of pixels in the images than the defects. Based on the median value for the number of pixels in the test dataset, the cracks account for 557 pixels, the inactive areas account for 1288 pixels, and the gridlines account for 1750 pixels. The spacing accounts for 12,529 pixels, and the ribbons account for 12,871 pixels. Thus, the features are roughly 10 to 20 times larger than the defects, which may explain the higher mIoU for features compared to defects. Alternatively, the improved performance for larger features may also be explained by the sample size. While every image in the dataset has a cell spacing layer, only some images have cracks and inactive areas, i.e., cracks are underrepresented compared to cell spacing. The best results for both multi-c-si and mono-c-si were observed on the DeepLabv3+ with custom weights.

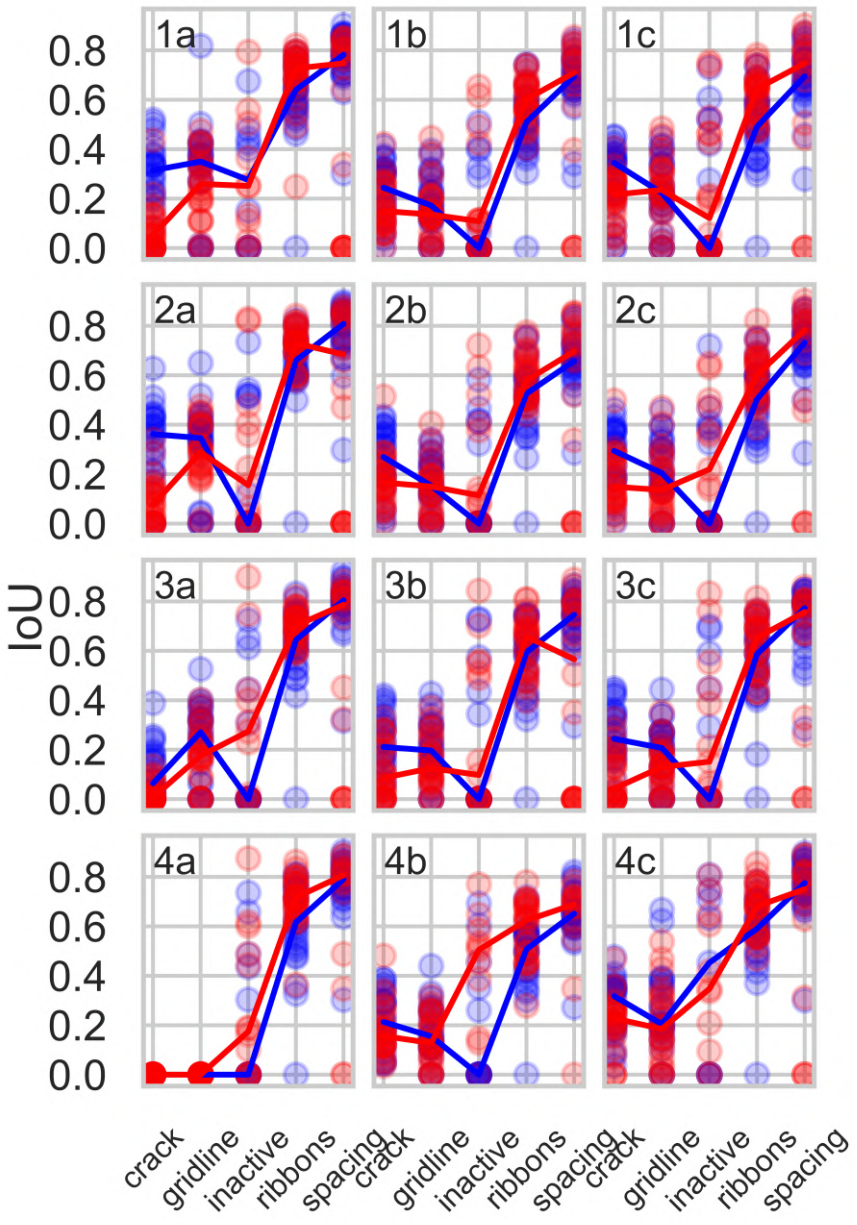
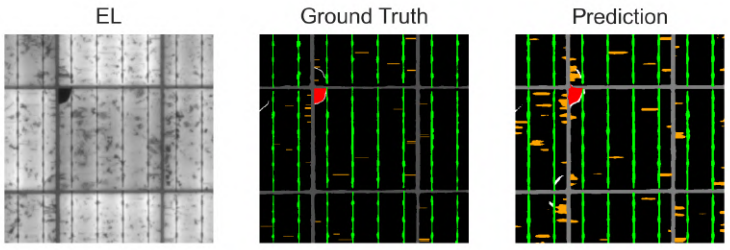
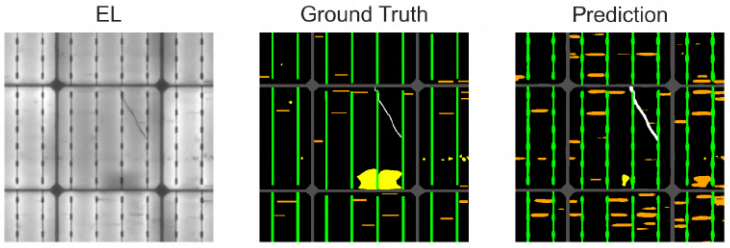


Figure 7.4: IoU and mIoU for mono-crystalline (blue circles), multi-crystalline (red) circles), and mIoU (solid lines) for selected defects and features in 50 test images across 12 models

Figure 7.4 also shows the mIoU for cracks and gridline defects is generally higher on mono-c-si silicon compared to multi-c-si wafers. Figure 7.5 compares two selected images and corresponding predictions from the DeepLabV3+ with custom weights (Model 4c), one multi-c-si cell and one mono-c-si cell. The models tended to confuse grain boundaries intrinsic to the multi-c-si wafers for cracks and gridline defects. Note the additional cracks (white) in the predictions mask for the multi-c-si cell in the bottom left corner. Since mono-c-si cells do not have intrinsic grain boundaries, there is less opportunity for confusion and error. In this mono-c-si example, the crack detection was located correctly but still dilated.



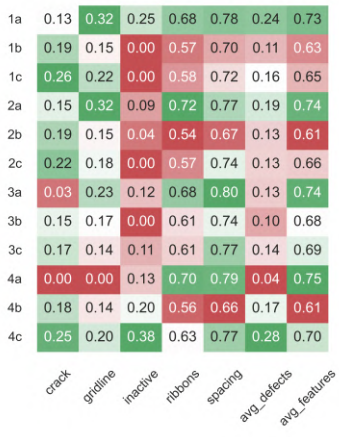
(a) multi-c-si CFVS-00180-r11-c4



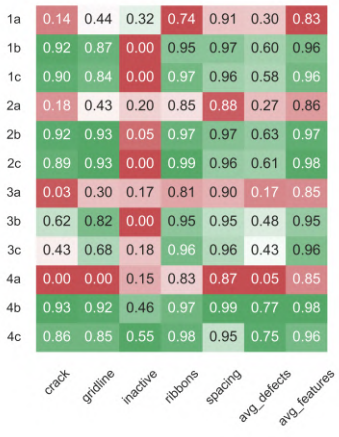
(b) Mono-c-si CFVS-00082-r7-c3

Figure 7.5: Two selected images and corresponding predictions from the DeepLabV3+ with custom weights (Model 4c), (a) one multi-c-si cell with some grain boundaries misclassified as cracks (white pixels) in the bottom left corner and (b) one mono-c-si cell with no similar errors in crack predictions

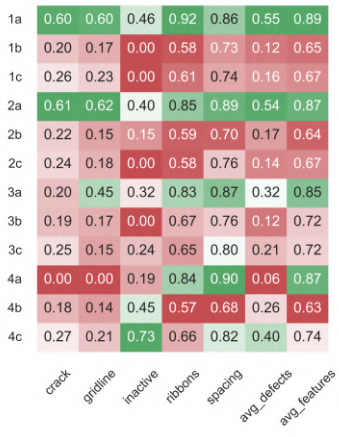
Figure 7.6 shows a heatmap for the mIoU, mRecall, and mPrecision for the five selected classes across 12 models. The two columns on the right side of each table show the average values for the three defects (crack, gridline, inactive) and two features (ribbons, spacing), respectively. The colour scale was applied independently to each column. The darker green cells correspond to high mIoU in each column, and red cells correspond to low mIoU.



(a) median IoU



(b) median Recall



(b) median Precision

Figure 7.6: (a) Median IoU, (b) median recall, and (c) median precision for five classes across 12 models and the average for selected features and defects

Figure 7.6 shows that the models with equal class weights (1a, 2a, 3a, 4a) have relatively high mIoU for the larger features (0.73–0.75) and relatively low mIoU for defects (0.04–0.24). The mIoU for the larger features compares reasonably well to the IoU values reported in Lin *et al.* [2016] for detecting objects such as bikes (0.41), boats (0.68), chairs (0.43), tables (0.65), sofas (0.64), and TVs (0.73). However, the examples in Lin *et al.* [2016] also illustrate the challenges around detecting long, narrow objects such as bird feet, table legs, chair legs, cat tails, and tires which form part of a small part of the larger object. These smaller features were often reduced, dilated, or missing altogether in the predictions presented in Lin *et al.* [2016]. The mean IoU scores in Lin *et al.* [2016] ranged from 62.2% to 78% across a range of objects despite the fact that many of the smaller features were missing in the predictions. The mIoU for cracks and gridlines reported in Figure 7.4 is much lower, likely because the cracks and gridlines consist exclusively of long, narrow features which are challenging to locate with pixel-level resolution precisely. The highest average mIoU for defects (28%) was generated by the DeepLabv3+ with custom class weights (4c), making it the best choice for detecting two of the critical defects in this dataset.

Figure 7.6 also shows a heatmap for the mRcl and the corresponding averages for the three defects and two features. The models using equal class weights (1a, 2a, 3a, 4a) show the lowest average mRcl for defects (0.05–0.3). The pattern is highlighted by the red hue in the corresponding row for each of these models. The models using inverse class weights (1b, 2b, 3b, 4b) show the highest average mRcl for defects (0.48–0.77). The pattern is highlighted by the green hue in the corresponding row for each of these models. The rows corresponding to the custom class weights (1c, 2c, 3c, 4c) show a similar green hue as the inverse class weights, indicating similar performance. The statistics are consistent with the images in Figure 7.3, which show dilated cracks and gridline defects in predictions from models when the inverse class weights were applied compared to models with equal class weights. The models using custom class weights (1c, 2c, 3c, 4c) also show relatively high average mRcl for defects and visually dilated defects in Figure 7.3. The DeepLabv3+ with custom class weights (4c) is among the top performers for the average of the mRcl for both features and defects. The mRcl for cracks (0.86) and gridline defects (0.85) from model 4c also compares favourably to the recall for cracks (0.82) and gridline/contact defects (0.78) reported in Fioresi *et al.* [2022]. With respect to recall, the finer defects in the SCDD ground truth masks perform slightly better than the conglomerations in the UCF ground truth masks. Refer to Chapter 5 Section 5.4 for additional comparisons between the two datasets.

While the DeepLabv3+ with custom weights (4c) shows a high recall averaged across the three defect classes, the mIoU remains relatively low for cracks and gridlines due to the low precision, as seen in 7.6. Cracks and gridline defects are both long, narrow features compared with the inactive area defect. Figure 7.7 provides some insight to explain the low precision. The ground truth mask accurately reflects the location and width of cracks and gridline defects in the corresponding EL image. The prediction mask also accurately locates the defects in the EL images. However, the defects are dilated in the prediction mask relative to the ground truth mask, leading to low precision. In this example, the gridlines are especially dilated relative to the ground truth masks and the original EL images. This may be mitigated by decreasing the class weights for gridline defects. If the cracks and gridlines were annotated as conglomerations as in the UCF dataset, the precision and IoU would likely measure higher.

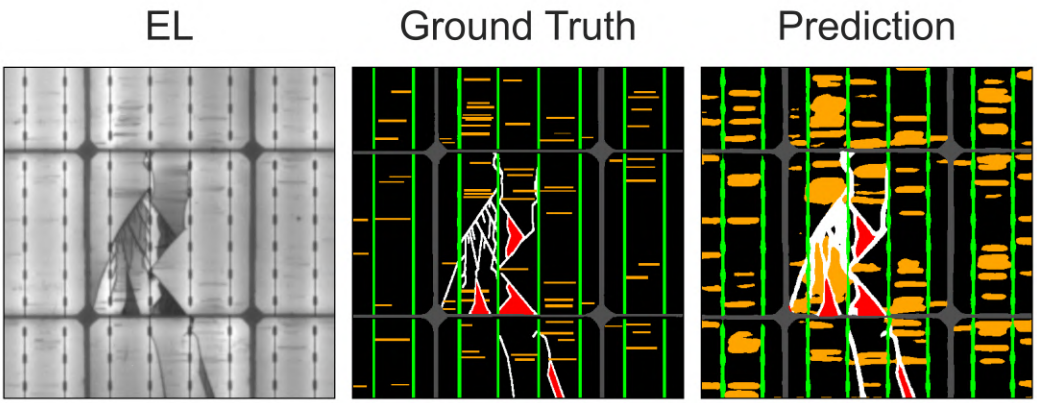


Figure 7.7: EL, ground truth mask, and prediction mask from DeepLabv3+ with custom class weights (4c) for CFVS-00035-r8-c4 highlighting the dilated cracks and gridline defects in the prediction

The accuracy of the ground truth image can also impact the IoU and recall during testing. Some mismatch between the EL and the ground truth masks is inevitable despite efforts to mitigate this issue during the labelling process. Figure 7.8 shows the bottom left corner of ARTS-00020-r7-c3 to illustrate the potential for error, especially for long, narrow objects. The spacing layer in the prediction mask (grey) covers the corresponding feature in the EL image more completely than the ground truth mask. The spacing layer in the ground truth mask shows some dark pixels along the border of the spacing layer that were inaccurately left to the background layer instead of correctly labelled as spacing. Similarly, the ribbons layer in the prediction mask (green) may reflect the corresponding feature in the EL image better than the ground truth mask. Errors in labelling along the borders of cracks and gridline defects have a particularly big impact on the mIoU due to the higher aspect ratios of those defects. These defects tend to be long and narrow, sometimes only a few pixels wide, so incorrectly labelling the edge pixels along the long border can greatly impact the performance metrics.

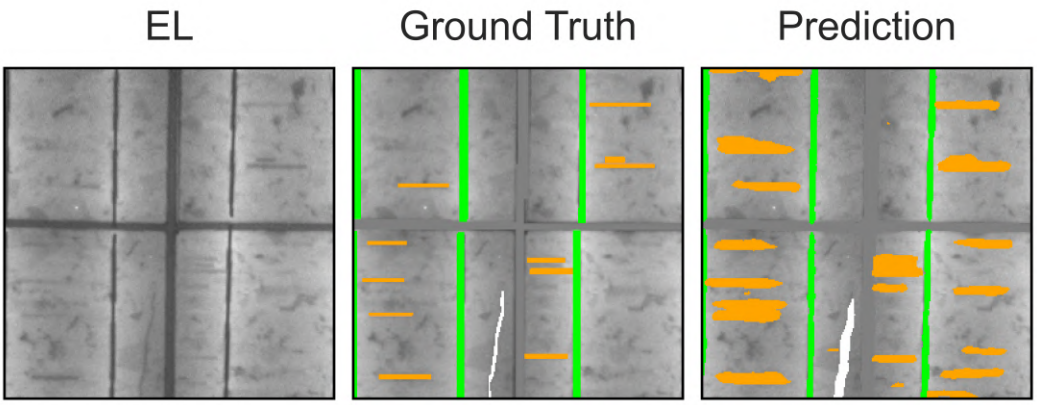


Figure 7.8: Bottom left corner of ARTS-00020-r7-c3 showing the labelling errors along the edges of ribbons and cell spacing in the ground truth image leading to errors in precision and recall

The low mIoU values for defects are driven by the low precision. The low precision correlates to the dilated predictions of the correct features and defects rather than the erroneous detection of spurious features and defects in the wrong location. However, the extent of the dilation and spurious detections was deemed acceptable at this stage. The impact of the dilation could be mitigated by adapting the class weights, resulting in a trade-off between recall and precision. As seen in Figure 7.3, the dilation of features and defects is minimal when the class weights are equal. Further optimisation on class weights could lead to higher precision and mIoU.

In addition to the key defects and features analysed above, the models also learned to detect other features and defects despite relatively small sample sizes. Figure 7.9 shows the EL image, ground truth, and prediction mask from the DeepLabv3+ model (4c) for an image sourced from Karimi *et al.* [2019], one of the public datasets used for training. The authors originally trained models to classify images as 'good,' 'cracked,' or 'corroded.' The DeepLabv3+ model (4c) detected and localised the corrosion defect in the image despite having only 17 images with the corrosion defect in the training dataset. The corrosion defect (dark green) was correctly detected around the ribbon feature (bright green). Ten out of twelve models trained on this dataset detected the corrosion defect with more or less accuracy. The only models that failed to detect the corrosion were the DeepLabv3+ models with equal and inverse class weights. The large olive padding feature added around the cell during pre-processing was also clearly predicted, as was the smaller border feature (brown) surrounding the cell. Some of the gridline defects were detected, but the crack was not. This is a multi-c-si cell, and Figure 7.4 shows that the cracks have lower mIoU in multi-c-si cells than mono-c-si cells. The model also learned to ignore the grain boundaries prevalent in multi-c-si cells correctly.

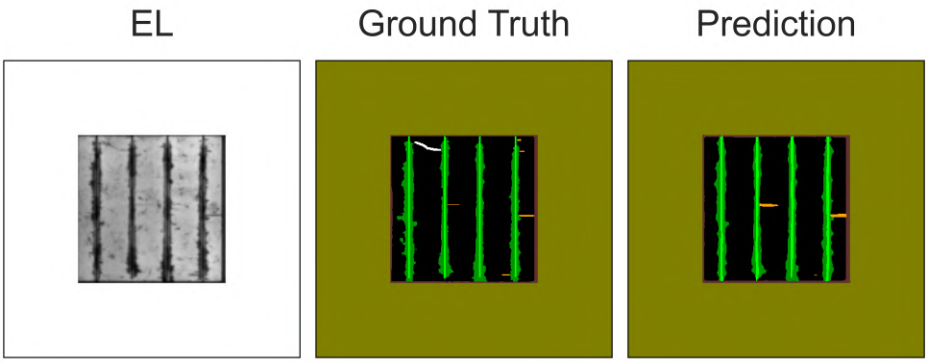


Figure 7.9: EL, ground truth, and prediction mask from DeepLabv3+ with custom class weights (4c) for SDLE-00513 showing accurate predictions for the corrosion defect (dark green) around the ribbon interconnects (bright green)

7.5 Conclusion

Four deep learning models (U-Net-12, U-Net-25, PSPNet, and DeepLabv3+) were trained using equal class weights, inverse class weights, and custom class weights for twelve sets of predictions for each of the 50 test images. The models were trained to simultaneously detect 24 classes in EL images of solar PV cells using semantic segmentation. Twelve classes correspond to intrinsic features of a solar cell, and twelve classes correspond to extrinsic defects. This section focused on detecting three critical defects and two common features in crystalline silicon

solar cells. The DeepLabv3+ with custom class weights was found to have the highest mIoU averaged across three critical defects identified in this dataset. The mIoU for the features was significantly higher than the mIoU for the defects for all models tested. Notably, the models achieved lower mIoU and mRcl on the small, narrow defects such as cracks and gridlines compared to large features such as spacing and ribbons. The low IoU is likely driven more by the dilated features in the predication masks and less due to spurious misclassifications. The models tested are effective in detecting, localising, and quantifying multiple features and defects in EL images of solar cells.

The unique contributions from this work comprise the published benchmark dataset and the benchmark performance metrics for semantic segmentation on EL images of solar PV cells. While the SCDD dataset originated from the work published in Pratt *et al.* [2021a], the dataset was expanded and made public in parallel with Pratt *et al.* [2023].

Specific findings from this chapter include:

- The DeepLabv3+ model showed the highest mIoU and mRcl for defects.
- Defect detection is marginally better for mono-c-si compared to mult-c-si for all models.
- The mIoU is significantly higher for predicting larger features versus smaller defects.
- The predictions for cracks and gridline defects when training with inverse class weights and custom class weights resulted in dilated predictions, leading to lower precision and IoU.
- Accurate ground truth masks are critical for assessing model performance.
- Long, narrow objects are especially susceptible to annotation errors along the border pixels.

7.6 Next steps

The next steps will focus on the evaluation of semi- and self-supervised models. Less than 600 EL images are annotated for training fully-supervised models, while over 150,000 unlabelled images are available.

Chapter 8

Semi- and Self-supervised deep-learning models

8.1 Introduction

Chapter 8 focuses on methods for Semi-Supervised Learning (Semi-SL) and Self-Supervised Learning (Self-SL) to use unlabelled data for SCDD in EL images. We found no published literature for Semi-SL or Self-SL in the EL imaging domain. We evaluated existing semantic segmentation models using fully-supervised (DeepLabv3 [Chen *et al.* 2017] and DeepLabv3+ [Chen *et al.* 2018b]), semi-supervised (CCT [Ouali *et al.* 2020b] and U^2PL [Wang *et al.* 2022b]), and two self-supervised methods (SimCLR [Chen *et al.* 2020b] and MoCov2 [Chen *et al.* 2020d]). No new models were developed at this research stage. A journal article with methods and results of our work in this space is currently under review [Torpey *et al.* 2024].

Pretraining has improved performance in many domains, including semantic segmentation, especially in domains with limited training data. This research aimed to evaluate the impact on prediction accuracy for SCDD by incorporating unlabelled data into the training. We evaluated various pretraining methods for Solar Cell Defect Detection (SCDD) in EL images, a field with limited labelled datasets. We also experimented with both in-distribution and out-of-distribution (OOD) pretraining and observed how this impacts downstream performance. The results suggest that supervised pretraining on a large OOD dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and semi-supervised pretraining (CCT) all yield statistically equivalent performance for mean Intersection over Union (mIoU). We improve on our state-of-the-art results for SCDD and demonstrate that specific pretraining schemes result in superior performance in underrepresented classes. Additionally, we provide an unlabelled EL image dataset of 150,000 images for further research in developing self- and semi-supervised training techniques in this domain.

8.2 Dataset

The data comprised the expanded SCDD dataset with annotations plus an additional 5,000 unlabelled images from the same public and private sources from which the annotated dataset was sampled.

8.3 Method

Four methods for pre-training the models were defined and evaluated for SCDD. Table 8.1 summarises nine models or regimes that fall into one of the four pre-training methods. The methods for pre-training consisted of no pre-training, fully-supervised, self-supervised, and semi-supervised. The fully-supervised regimes consisted of DeepLabV3 and DeepLabV3+. The self-supervised regimes consisted of SimCLR [Chen *et al.* 2020b] and MoCov2 [Chen *et al.* 2020d]. The semi-supervised regimes consisted of CCT and U^{2}PL. We formalise the model description with the following terms. First, we define the SCDD dataset as $X = \{(x_i, y_i)\}$, where $x_i \in X$ is the EL image and $y_i \in Y$ is the corresponding ground truth mask. Next, we define an in-distribution unlabelled set of images X_{ul}, which is the set of EL images in the CSIR database from which $X = \{(x_i, y_i)\}$ was selected and annotated. Finally, we define X_{ul}^{OOD} as the set of unlabelled images from ImageNet. We define a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that takes an EL image $\mathcal{X}$ as input and predicts the segmentation mask $\mathcal{Y}$. The function f can be decomposed into an encoder and decoder: $f = e \circ d$, where e is the encoder and d is the decoder.

Table 8.1: Regime descriptions evaluated for SCDD using SemSeg.

Regime	**Pre-training**	**Weights**	**Model**
R0	Fully-supervised	ImageNet	f trained using DeepLabv3+
R1	NA	NA	f trained using DeepLabv3
R2	Fully-supervised	COCO SemSeg	f trained using DeepLabv3
R3	Fully-supervised	ImageNet	e pretrained on ImageNet
R4	Self-supervised	ImageNet	e pretrained on X_{ul}^{OOD} using SimCLR
R5	Self-supervised	ImageNet	e pretrained on X_{ul}^{OOD} using MoCov2
R6	Self-supervised	EL	e pretrained on X_{ul}^{EL} using SimCLR
R7	Semi-supervised	CCT	f trained on X_{ul} and X using CCT
R8	Semi-supervised	U^{2}PL	f trained on X_{ul} and X using U^{2}PL
R9	Semi-supervised	CCT+	f trained with R7 followed by X

The state-of-the-art (SOTA) regime (R0) was the same model described in Section 7 of this thesis and further defined in Pratt *et al.* [2023]. This is a DeepLabV3+ architecture with a ResNet34 backbone initialised with pre-trained weights from ImageNet. For all other regimes, we used a ResNet50 [He *et al.* 2016] backbone with a DeepLabV3 segmentation model unless otherwise stated. For R1 (Random), we used random weight initialisation.

The first new group consisted of two fully-supervised models (R1 and R2). For regime R1, f is a randomly initialised DeepLabV3 model without pretraining. For regime R2, f is initialised by supervised pretraining on COCO SemSeg weights, an OOD semantic segmentation dataset, followed by fine-tuning on X.

The second group consisted of three models using self-supervised learning. In regime R3, we consider the case where the encoder e is first estimated by pretraining on ImageNet in a supervised fashion. For the following two regimes, we evaluate a model where e is first estimated offline by performing self-supervised pretraining using X_{ul}^{OOD} from ImageNet with SimCLR (R4) and MoCov2 (R5), and using X_{ul} with SimCLR (R6). Sim-CLR is a contrastive learning method that uses the InfoNCE [Sohn 2016a] objective to train a network to discriminate between different views (obtained through random data augmentation) of the same or different images. SimCLR performs best with large batch sizes to obtain a sufficiently large amount of negative samples when computing the loss. MoCo is an algorithm that also minimises InfoNCE, however, it contains a Siamese architecture in which one network (encoder) is updated via backpropagation, while the other (momentum encoder) is updated via an exponential moving average of the encoder weights (which is used to ensure the consistency of the representations of negatives samples drawn from a memory bank). MoCov2 improves on MoCo by a projection head and additional data augmentation to generate random views.

The third group consisted of three models using semi-supervised regimes. Here f is estimated using both X_{ul} and X simultaneously. In this regime, we consider estimating f using CCT (R7) and U^{2}PL (R8). CCT is a consistency-based semi-supervised technique for semantic segmentation tasks. A shared encoder and main decoder are trained in a supervised fashion, while consistency with the output of various auxiliary decoders is simultaneously enforced. Consistency is enforced by making the model's predictions invariant to small perturbations in the input. On the other hand, U^{2}PL approaches semi-supervised semantic segmentation via pseudo-labelling [Lee and others 2013] rather than consistency training. Typically, the only pseudo-labels used for training are those the model is confident about. However, U^{2}PL uses these unreliable pseudolabels, as they may still prove useful for discrimination in the task at hand. This is done by using so-called anchor pixels (or query pixels), selecting associated negative and positive samples for each anchor pixel, and computing a contrastive loss. This contrastive loss is part of the overall loss function, which includes additional supervised and unsupervised terms. Finally, we consider the case where the encoder estimated from R7 is used for fine-tuning a DeepLabV3 model using X (R9). We use CCT and U^{2}PL as these are both recent state-of-the-art methods explicitly tailored for semi-supervised semantic segmentation tasks. Additionally, they provide a simple and easy way to standardise the backbone networks to align with the other regimes under consideration.

For the training of SimCLR in R6, we perform a hyperparameter sweep over the learning rate and weight decay, where the metric is IoU on the validation subset of X. The parameters used in estimating f for all other regimes are aligned with those reported in the original works [Pratt *et al.* 2023; Ouali *et al.* 2020b; Wang *et al.* 2022b].

While we use various initialisation schemes for the encoder, we use random initialisations for the decoder, d, in all models except R2 (COCO), where the pre-trained decoder weights from torchvision are used.

It should be noted that in our experiments, $|X_{ul}^{\text{OOD}}| \gg |X_{ul}| \gg |X|$, where $|\cdot|$ represents the number of images in the dataset. We use ImageNet and COCO datasets for X_{ul}^{OOD}. Additionally, we leverage the SCDD dataset with 642 EL images and the corresponding ground truth masks. Furthermore, we constructed an unlabelled EL image dataset of 64,000 images from commercial partners.

Due to limitations on computing capacity, we used a random subset $X_{ul}^s \subset X_{ul}, |X_{ul}^s| = 5000$ in the estimation of f in regimes R7, R8, and R9. For R6 we use $|X_{ul}^s| = 32000$. We found this works better empirically for our EL dataset.

8.4 Experiments

8.4.1 Experimental Setup

For R0 (SOTA), we used the same setup as the current state-of-the-art for segmentation-based SCDD [Pratt *et al.* 2023]. This is a DeepLabV3+ architecture with a ResNet34 backbone initialised with pre-trained weights from ImageNet. For all other regimes, we use a ResNet50 [He *et al.* 2016] backbone with a DeepLabV3 segmentation model unless otherwise stated. For R1 (Random), we use random weight initialisation. For R2 (COCO) and R3 (ImageNetSup), we use the pre-trained models within torchvision [TorchVision maintainers and contributors 2016] where the models were pre-trained on COCO (supervised semantic segmentation) and ImageNet (supervised image classification), respectively. For R4 (ImageNet SimCLR) and R5 (ImageNet MoCov2), we use the models available within the VISSL model zoo [Goyal *et al.* 2021] where the backbone models were pre-trained using SimCLR and MoCov2 on ImageNet, respectively. For R6 (EL SimCLR), we train a SimCLR model on the entire unlabelled in-distribution dataset X_{ul} for 500 epochs with the LARS [You *et al.* 2017] optimiser and a batch size of 32 to get the initial weights. The 500 epochs were selected based on recommendations in the literature. Then, we train the full model for 1000 epochs using those initial weights. We use an initial learning rate of 0.09, weight decay of 1.4e-07, and ten warmup epochs. The learning rate is decayed using cosine decay. The default set of data augmentations is used as per the original SimCLR work [Chen *et al.* 2020b].

For R7 (CCT), R8 (U^{2}PL), and R9 (CCT+), we train at a resolution of 512x512 with the SGD optimiser with a learning rate of 0.01 (with polynomial decay), weight decay of 0.0001, and momentum of 0.9. However, we only train CCT for 400 epochs and U^{2}PL for 500 epochs, as there was no improvement in mIoU past these points for the respective models.

All DeepLabV3 models were trained with the same hyperparameters and custom class weights as the previous SOTA model [Pratt *et al.* 2023]. We train all DeepLabV3 models for 1000 epochs, having converged on the best weights in less than 1000 epochs.

For all models, we report results on the test set for the best model found in terms of mIoU on a validation dataset. We report two metrics: mIoU (mean intersection-over-union) and a weighted version of IoU, which we term wIoU. This is simply IoU weighted according to the frequency of the different classes. The wIoU was defined in Chapter 4, Section 4.5. The wIoU is used to include a metric that lessens the impact of underrepresented classes, which is important in certain parts of our analyses and appropriate in the SCDD domain when there is a significant class imbalance.

8.4.2 Dataset

The experiments were conducted on the SCDD dataset for SemSeg of EL images. The dataset consists of 642 EL images of solar cells with labelled ground truth masks, including 15 de-

fect classes representing potential extrinsic problems with solar cells and 13 feature classes representing intrinsic features of solar cells. This dataset was bigger than the dataset used for the initial work on the U-Net [Pratt *et al.* 2021a]. Figure 8.1 shows the counts of defects and features represented in the dataset. Some classes are well represented for training and statistical analysis of the test results. Some classes are present in only a few images, which hinders the training process and analysis for those classes. The background class bar corresponds to the total number of images in the dataset since the background class dominates all the solar cells. Nearly all images have ribbon interconnects, and roughly 50% have the border class and padding. The number of multi-c-si and mono-c-si cells is roughly equivalent. The remaining features are present in fewer than 150 images. The defects section of the graph shows that only cracks, gridline defects, and material defects were represented by 100 images or more. Other critical defects such as scuffs, brightening, and dead cells remain highly underrepresented. Data augmentation was implemented for all images, as described in 5.2. The data augmentation method increased the dataset size by a factor of four but did not reduce the class imbalance. The class imbalance in the dataset was mitigated with custom class weights. Future work will include identifying existing images with underrepresented defects and adding them to the ground truth dataset. The existing SCDD models can identify good candidate images from the thousands of available images.

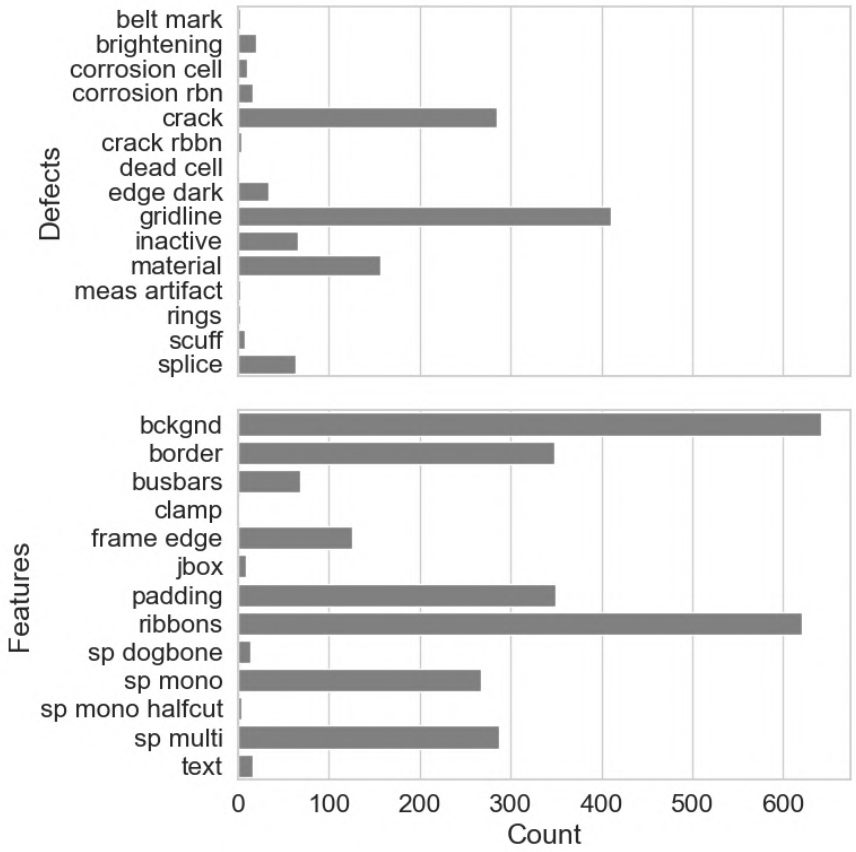


Figure 8.1: Number of annotated images containing each defect class (top) and feature class (bottom)

8.5 Results

Table 8.2 shows that the semi-supervised method CCT (R7) performs best in terms of mIoU, while a COCO pre-trained DeepLabV3 model (R2) performs best in terms of wIoU. The top 3 models for mIoU metric (R7, R2, and R4) all incorporate supervised pretraining in their regime. The differences among the top three models are not statistically significant when comparing the average mIoU across five runs. The same 3 models are also the top performing for wIoU, just ordered differently. This suggests that the superiority of self-supervised pretraining over supervised pretraining described in Ericsson *et al.* [2020] does not necessarily generalise to this SCDD domain. We argue that modern self-supervised learning techniques' objectives are not designed to allow the network's embeddings to effectively capture relevant information for downstream tasks that involve images from a distribution that significantly differs from the one used to train the self-supervised model. However, the supervision signal derived from image classification-based supervised training on ImageNet has been widely shown to perform well on a host of downstream tasks, even where the data is notably different from ImageNet. A typical evaluation of self-supervised learning versus supervised learning will focus on SemSeg for images that are comprised of common objects and natural scenes [Ericsson *et al.* 2020], which is very different to the EL dataset. R2 and R7 also show a statistically significant increase in mIoU compared to R0 SOTA ($p < 0.03$). The differences in average mIoU between SOTA and the other models are not statistically significant except for the improved performance of SOTA versus U^{2}PL (R8) ($p < 0.002$).

Table 8.2: Results on the EL dataset for the SCDD using semantic segmentation

Model ID	**mIoU**	**wIoU**	**OOD**
R0 (SOTA)	0.506	0.535	✓
R1 (Random)	0.486	0.542	-
R2 (COCO)	0.544	**0.589**	✓
R3 (ImageNet Sup)	0.508	0.562	✓
R4 (ImageNet SimCLR)	0.534	0.578	✓
R5 (ImageNet MoCov2)	0.508	0.562	✓
R6 (EL SimCLR)	0.517	0.571	✗
R7 (CCT)	**0.550**	0.580	✗
R8 (U^{2}PL)	0.448	0.515	✗
R9 (CCT+)	0.508	0.562	✗

Notably, the other semi-supervised algorithm, U^{2}PL, performs worst – even worse than random weights ($p \leq 0.01$) for all pairwise comparisons. This phenomenon was noticed in the original U^{2}PL work and occurs when the pseudo-label pixels are not adequately filtered out when using the contrastive loss term. It can lead to worse results than simply training on the supervised data alone. It should also be noted that in regime R9, training the decoder of a trained CCT model in a supervised manner for additional epochs within DeepLabV3 does not improve performance. In fact, both mIoU and wIoU decreased. This suggests that the CCT architecture alone is sufficient to perform SemSeg for SCDD.

Figure 8.2 shows the mIoU and wIoU for each model and class type. Performance on extrinsic defects is consistently worse than on intrinsic features. The poor performance on the defects is

likely two-fold. Firstly, many of the defects are underrepresented, so the models may not have sufficient data to learn the features to segment defects. The wIoU is slightly higher than the IoU for defects, which may be due to the higher weights assigned to the defects with greater representation. Secondly, defects are generally smaller in size than features. This makes them harder to detect for conventional semantic segmentation algorithms that are typically optimised for and benchmarked on datasets that contain natural scenes and common objects, such as COCO [Lin *et al.* 2014], Pascal VOC [Everingham *et al.* 2010], and CamVID [Brostow *et al.* 2009]. The average size of a defect is 1.5%, and the average size of a feature is 24%, measured as a percentage of the image. Therefore, features are roughly 16 times larger than defects.

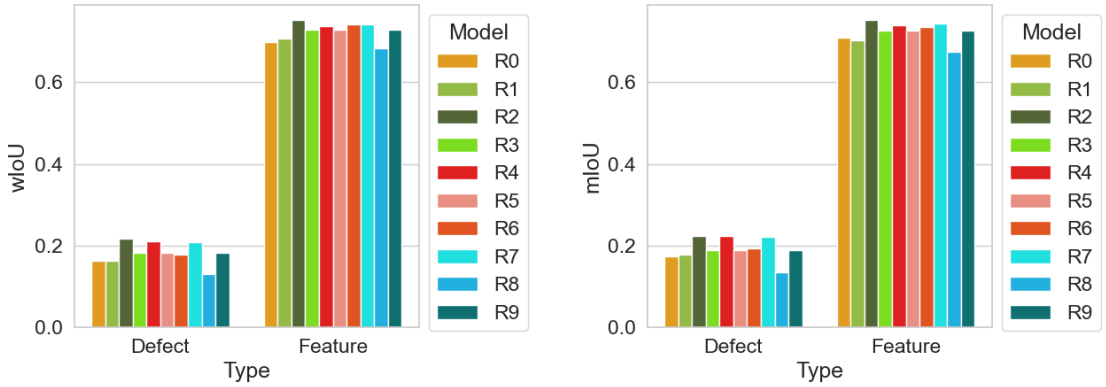


Figure 8.2: (a) mIoU and (b) wIoU by class type (extrinsic defect and intrinsic feature)

Nevertheless, it is clear that the trends from Table 8.2 hold when analysing the results by group (defect versus feature), i.e., models R2, R4, and R7 perform best, while U^{2}PL performs the worst.

8.5.1 Defect Analysis

While 15 defects are present in the dataset, crack detection remains paramount in real-world applications. A statistical analysis of the block-centred IoU for cracks indicates that R2 and R4 outperform all other models ($p \leq 0.0001$ for all pairwise comparisons), including R7 CCT. Figure 8.3 shows the mean and 95% confidence intervals for crack detection based on Tukey's Honest Significant Difference (HSD) as implemented in the 'statsmodels' python library [Seabold and Perktold 2010]. To improve the signal-to-noise ratio, the image-level IoU values were block-centred to eliminate the variability from image-to-image differences, as described in Chapter 4, Section 4.6. The block-centred IoU values were then averaged within each regime, leading to a sample size of five for each regime with multiple runs. The simultaneous confidence intervals calculated by 'statsmodels' are based on the overall variance divided by the number of observations in each group [Perktold *et al.* 2023], so the models with five runs have equally narrow confidence intervals. Models R0, R7, and R8 were only run once, so the confidence intervals are wider. R7 and R8 were run only once due to high computational costs.

Many of the features and defects in the dataset are underrepresented, so training and statistical analysis are challenging. However, several of the models perform well on underrepresented defects, even for defects not detected by the R0 SOTA model. The 'ring' defect and 'dog-

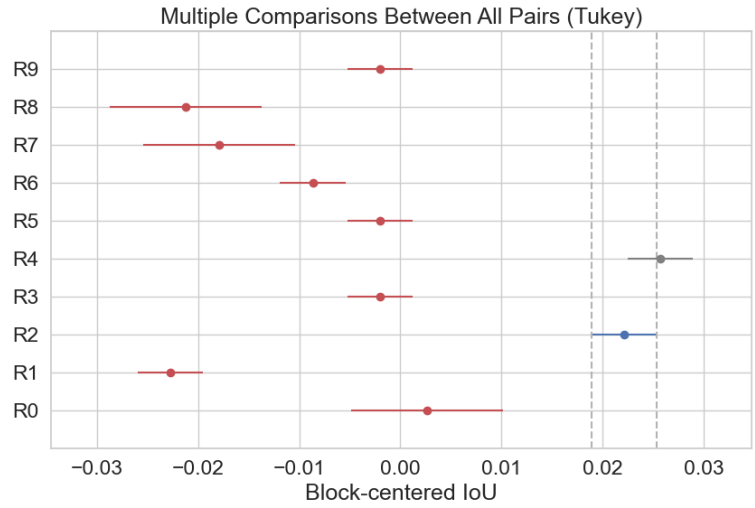


Figure 8.3: Block-centred IoU average and 95% confidence intervals for crack detection by regime showing R2 and R4 perform statistically significantly better on crack detection compared to other models

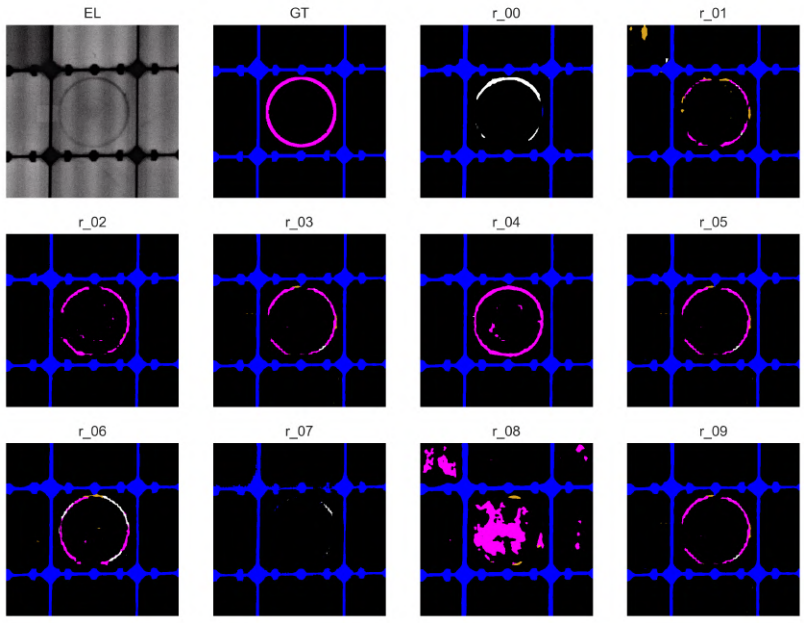
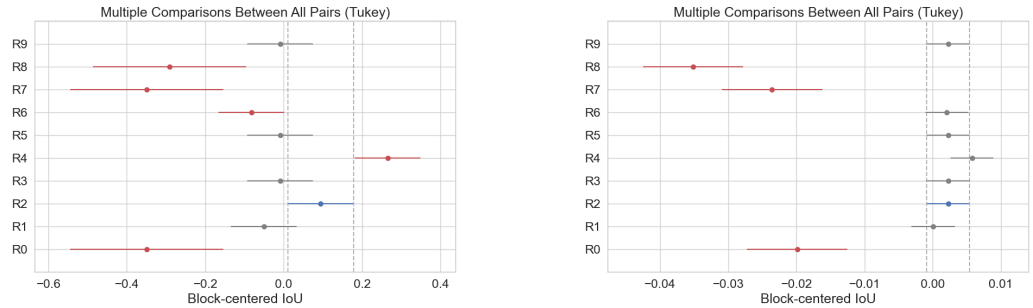


Figure 8.4: EL image, ground truth, and predictions for CSIR_00067_r3_c2 with ring defect and dogbone spacing

bone spacing' feature are two examples of underrepresented classes. Figure 8.4 shows the EL, ground truth, and predictions from the R0 SOTA model and the nine new models for the ring defect and dogbone spacing feature. The ring defect (pink) was present in only three images from the database, and the dogbone spacing feature (blue) was only present in 14 images. A single image with both ring defect and dogbone spacing was reserved for testing, so the training dataset included no more than two images for the ring defect and 13 images for the dogbone spacing. While the R7 CCT model performed among the best overall for all defects combined, the ring defect was only partially detected and incorrectly classified as a crack by this model. The predictions from models R2 and R4 for ring defects are impressive, especially given the small sample size.

Figure 8.5 shows Tukey's HSD for the ring defect and the spacing dogbone feature. The value of a statistical comparison on the ring defect is negligible, given that only one sample in the test dataset. However, the predictions for rings from R4 for that image were significantly better for all five runs compared to all other models ($p < 0.04$). Several other models also made reasonable predictions for the ring defect image. R0, R7, and R8 performed the worst quantitatively, consistent with qualitative analysis in Figure 8.4. The predictions for dogbone spacing from 7 out of 10 models were statistically equivalent. R0, R7, and R8 performed significantly worse, although that is not apparent from the qualitative analysis of the one test image in 8.4.



(a) ring defect (b) dogbone spacing

Figure 8.5: Tukey's HSD for (a) ring defect showing R4 is statistically significantly better than all other models and (b) dogbone spacing showing 7 out of 10 models are statistically equivalent

8.5.2 OOD Analysis

Table 8.3 shows the mIoU and wIoU averaged across regimes by group. One group included OOD pretraining, and the other group did not. The regimes that included OOD pretraining measured better overall performance. The previous R0 SOTA model is excluded from this table in order to standardise on models implemented and analysed in this work. It is interesting to note that this result generalises to the SCDD domain, even in the case where the in-domain unlabelled dataset for pretraining X_{ul} contains a non-trivial number of examples – approximately 30k (although still less than the OOD dataset). The results suggest that the scale of the pretraining dataset (in our experiments, ImageNet, and to a lesser degree, COCO) matters more for final performance than whether the dataset is in-distribution or OOD. We posit that the fea-

tures learned when pretraining at scale result in richer, more useful embeddings and backbone networks than in-distribution pretraining on a smaller dataset.

Table 8.3: Results over all models by whether an OOD dataset was used for pretraining.

OOD	mIoU	wIoU
✗	0.505	0.557
✓	**0.530**	**0.573**

Self-SL pretraining within regimes R4 and R6 provides more evidence for the advantage of pretraining on large OOD datasets. The ImageNet-pretrained SimCLR model (R4) performed better than the EL-pretrained SimCLR model (R6). We verify this result by performing an analogous experiment on the Cityscapes dataset. We take 90% of the training set for SimCLR-based Self-SL pretraining (with the same training parameters as R6) and the remaining 10% for fine-tuning in a DeepLabV3 model. The results of this can be found in Table 8.4.

Table 8.4: Results for OOD and in-distribution pretraining on the Cityscapes and EL datasets (as per R6).

	OOD	mIoU	wIoU
Cityscapes	✗	0.189	0.252
	✓	**0.243**	**0.319**
EL (R4 & R6)	✗	0.516	0.571
	✓	**0.534**	**0.578**

A primary reason for this gap in performance when pretraining on the in-distribution dataset is likely due to an observation that Self-SL methods perform significantly better when trained on object-centric, clean datasets [Purushwalkam and Gupta 2020], such as ImageNet. Semantic segmentation datasets typically contain multi-object images or scenes, as Cityscapes and EL indeed do. Thus, they are likely less suitable for Self-SL pretraining. In contrastive self-supervised methods such as SimCLR, the model learns based on an objective whereby the embeddings from two random views of the same image attract, and two different views repel. However, in a non-object-centric setting such as the EL or Cityscapes datasets, two random views from the same image may contain different objects/defects/features. For example, one random crop might contain a crack or inactive area where the other random crop form the same image might be crack-free. The network's objective would essentially be encouraging the crop with and without the crack to be the 'same' (i.e., represented by the same embedding). Our results suggest that this property results in less useful representations for the downstream semantic segmentation task.

Further, the particularly good performance of pretraining on the COCO dataset in a supervised manner (R2) – even outperforming supervised pretraining on ImageNet in many cases – is likely due to the fact that COCO is pre-training on the task of semantic segmentation while ImageNet is pre-training on image classification. The COCO pre-training results in a pre-trained network more tailored towards the semantic segmentation task of SCDD.

8.6 Conclusion

The experimental results suggest that supervised training on a large OOD dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and in-domain semi-supervised pretraining (CCT) all yield statistically equivalent performance for mIoU. The semi-supervised pretraining (CCT) using the relatively small domain-specific images did not improve the model performance compared to OOD pretraining, contrary to recent results in more object-centric domains. However, some DeepLabV3 models with a ResNet50 backbone achieved a new state-of-the-art for SCDD using semantic segmentation.

The results also demonstrate that specific models result in a statistically significant improvement for crack detection and underrepresented classes compared to state-of-the-art based on a DeepLabV3+ model with a ResNet34 backbone. We posit that the multiview nature of various Self-SL/Semi-SL techniques is not well suited to the EL semantic segmentation task. Future work in this field may yield improved techniques to pre-train such models using large amounts of unlabelled EL data. To this end, we also contribute a new EL image dataset of 120,000 unlabelled images for further research in developing machine learning models for semantic segmentation in the EL image segmentation domain.

Specific highlights from this chapter:

- This is the first and only large-scale analysis detailing how pretraining affects SCDD.
- A new state-of-the-art performance benchmark for SCDD was achieved.
- We show that supervised training on a large OOD dataset (COCO) (Model R2), self-supervised pretraining on a large OOD dataset (ImageNet) (Model R4), and semi-supervised pretraining (CCT) (Model 7) all yield statistically equivalent performance for mIoU on this dataset.
- We show that self-supervised pretraining in the SCDD domain does not yield a statistically significant improvement over supervised training, contrary to recent results [Ericsson *et al.* 2020].
- We publish an unlabelled dataset with over 150,000 EL images of solar cells.
- Long, narrow defects such as cracks and gridline defects remain difficult to detect, suggesting a new SemSeg architecture model should be investigated.

8.7 Next steps

Future work will focus on increasing the training dataset size for fully-supervised models. Specific attention will be given to increasing the samples for underrepresented defects and features. Models that focus on small object detection will be evaluated. Composite models will also be considered. A composite model would train separate models for specific groups of features and defects and then combine them into one prediction mask.

Chapter 9

Correlating EL defects and PV module performance

9.1 Introduction

Chapter 9 summarises work that was presented at the SASEC23 conference in November 2023 at Nelson Mandela University in Port Elizabeth, South Africa [Zandamela *et al.* 2023]. The Solar PV Team completed the lab work in 2020 under my direct supervision and individual contributions. I completed the technical work on cell cropping, semantic segmentation modelling, and regression analysis in 2023. Frank Zandamela was listed as the lead author partly due to his contributions to developing a perspective transform tool for pre-processing the module-level EL images, compiling the literature review, writing the original draft, and presenting the work. More importantly, as his line manager, I wanted to support his strong interest and proven track record in computer vision by assisting in his career ladder advancement through publication and conference presentations. The aim of the research was to apply the semantic segmentation model to EL images collected prior to the research and establish a correlation to PV module power output. We also investigated the prediction model sensitivities to the image brightness and pre-processing methods.

We used a simple linear regression model to correlate EL defects predicted by SemSeg and I-V curve characteristics of PV modules exposed to accelerated stress testing at the CSIR PV Module Quality and Reliability Laboratory in Pretoria, South Africa. The data was collected in 2019 during the commissioning of the PV lab that I was hired to start-up. Four module manufacturers donated modules to the CSIR in exchange for free test results. Each module type had a specific bill of materials (BOM), i.e. manufacturer, cell technology, power class, back sheet type, etc. Not all PV modules are built with the same materials, just as not all cars or televisions are built with the same materials. Specifically, we correlate power loss to dark cell defects following a potential-induced degradation (PID) stress test and power loss to cracks following a thermal cycling stress test. Results suggest a significant correlation between the predicted defects and the power loss ($r^2 = 55\%$), depending on the application and the pre-processing. Results indicate the predictions for dark cell pixels are sensitive to changes in brightness and contrast before making the predictions and the current bias applied when

recording the EL image. Predictions for cracks appear robust to changes in brightness and contrast.

9.2 Method

Figure 9.1 summarises the methodology. I-V curves and EL images were recorded for all modules initially, although that step was omitted from the graphic. Next, three modules from each BOM were subjected to various stress sequences according to international testing protocols. I-V curves and EL images were recorded again after each stress until the sequences were completed. The EL images were post-processed and passed through the SCDD model to generate pixel-level predictions for each cell in each EL image. The output from the SCDD model was averaged for each module at each step in the sequence for each defect. Finally, a simple linear regression model was fit to the power loss of each module relative to the initial power measurement versus the percentage of defective pixels. The strength of the linear correlations was quantified by the square of Pearson's Correlation Coefficient (r^2). Note that correlation does not imply causation, however.

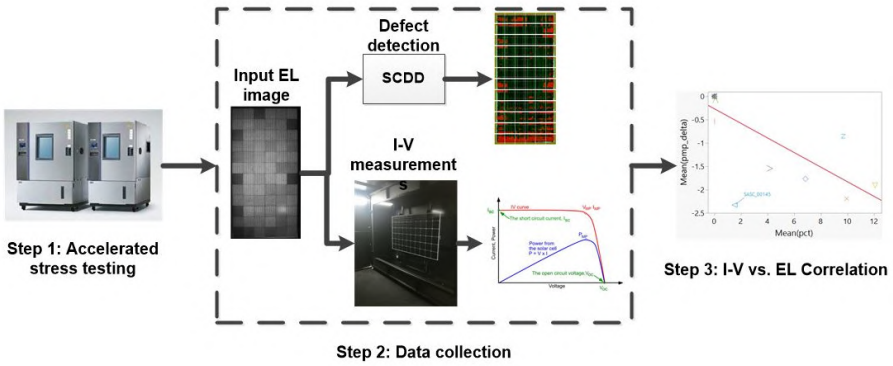


Figure 9.1: Methodology used for the regression analysis of I-V characteristics on EL defects

The sensitivity of the SCDD predictions to image processing was also analysed for dark cells and cracks. A simple design of experiments was conducted to vary the brightness and contrast of the EL images before making the predictions to determine if the percentage defective would change significantly. A significant change in the predictions will impact the regression analysis if the percentage defective changes depending on the image processing method.

9.2.1 Experimental Setup

The regression analysis was conducted on a batch of PV modules subjected to an extended reliability program conducted in 2019-2020 at the CSIR PVQRL [Pratt *et al.* 2021b]. The batch of PV modules consisted of 10-15 modules from four different 'bill-of-materials' (BOMs), i.e. four different manufacturers/models. The PV modules were subjected to a series of accelerated stress tests according to the methods described in the IEC 61215:2016 international standard for PV module design qualification and type approval and in the PV Module Testing Protocol for Quality Assurance Programs described in the ANSI C450-18 (C450). The C450 is a public standard designed for long-term reliability testing of PV modules based on the IEC 61215

series, in which the certification tests are conducted repeatedly. For example, a certification test per IEC 61215 requires 200 thermal cycles and the C450 requires 600 thermal cycles with a characterisation sequence every 200 cycles. The characterisation sequence includes EL images and I-V measurements. The characterisation sequence was conducted on each module as received at the lab and again after each step in the stress testing sequence, generating nearly 200 EL images and I-V curves. The accelerated stress tests are designed to simulate real-world conditions in a controlled lab environment that can lead to typical field failures in sub-standard modules, such as power loss due to various defects, some of which can also be detected in EL images.

9.2.2 Dataset

The dataset consisted of matched pairs for each module at each characterisation step. The pairs consisted of the I-V characteristic for the PV module and module level average of the cell-level predictions for each class from the SCDD model. Table 9.1 shows a subset of the data used in this study. Row 1 shows a record for the 'dark cell' defect predictions on Module 1 at the initial inspection. The initial maximum power point (Pmp) serves as the reference point for subsequent I-V measurement, so the delta to initial = 0. The percentage defective shows the output of the SCDD model averaged over all the cells in the module. At this stage, the SCDD model did not detect any 'dark cell' defects in any of the cells from this module. The second row shows the results on the same module after the PID stress test. The module power decreased by 6.8%, and the SCDD model predicted 1.77% of the pixels were likely dark cells, averaged over all cells in the module. The bottom half of the table shows a similar example for one module after the thermal cycling sequence. After 600 thermal cycles (TC600), the module power decreased by 1.1%, and the percentage of pixels predicted as cracks increased to 0.52%. This study focused on Pmp versus dark cells after PID and Pmp versus cracks after thermal cycling (TC).

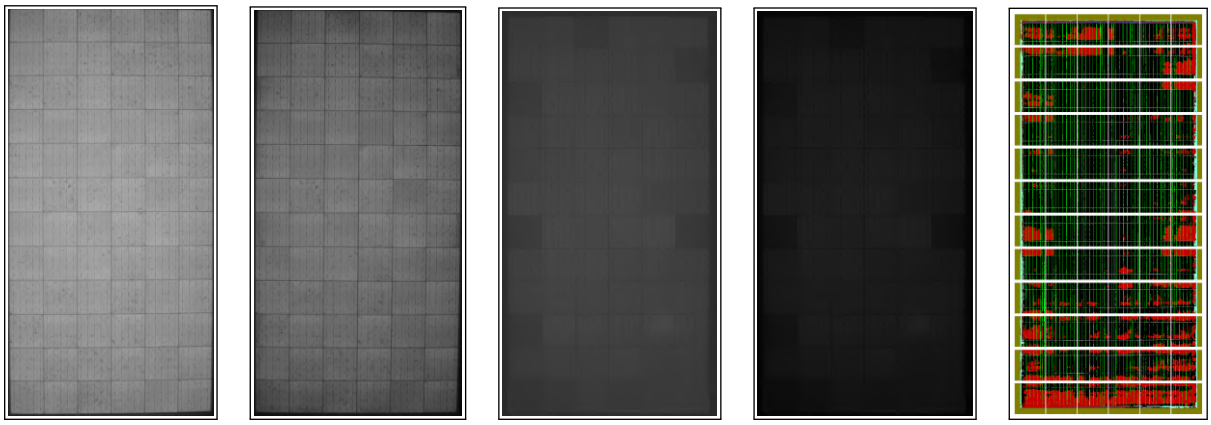
Defect	Module ID	Sequence	Pmp Delta to Initial (%)	Percentage Defective (%)
Dark cell	1	Initial	0.0	0.00
Dark cell	1	Post PID	6.8	-1.77
Dark cell	2	Initial	0.0	0.00
Dark cell	2	Post PID	9.9	-2.19
Crack	3	Initial	0.0	0.06
Crack	3	Post TC200	-0.2	0.07
Crack	3	Post TC400	-0.5	0.48
Crack	3	Post TC600	-1.1	0.52

Table 9.1: Sample records of the dataset used for the correlation analysis

The I-V curves were generated on an indoor sun simulator at the CSIR PVQRL. The sun simulator is designed to measure electrical characteristics over a range of temperatures and irradiance levels. Each module is loaded into a chamber with temperature and irradiance controls. The temperature chamber has a glass door to allow light from a xenon arc lamp to hit the module, and the irradiance level is monitored by a calibrated reference cell co-planar with the PV module. The irradiance level is varied by adjusting the current to the lamp, and feedback from the reference cell is used to tune the simulator to the desired irradiance level. During a C450

reliability sequence, most I-V curves are conducted at standard test conditions (STC) defined as 1000 W/m2, 25 °C cell temperature, and a light spectrum consistent with the natural sunlight at an airmass of 1.5.

Figure 9.2 shows a set of images for a single PV module in the PID dataset. The set consists of EL images and the prediction mask recorded at 10% of Isc current after the PID stress. The PID dataset includes a post-stress EL image recorded at 10% of Isc current, but the corresponding image is missing from the pre-stress characterisation sequence. The missing data limits the validity of the regression analysis. However, the dataset still provides insight into the importance of both EL imaging methods and image post-processing methods applied before the prediction model is run for dark cells. Both the EL imaging methods and the image post-processing methods impact the output of the prediction model for dark cells.



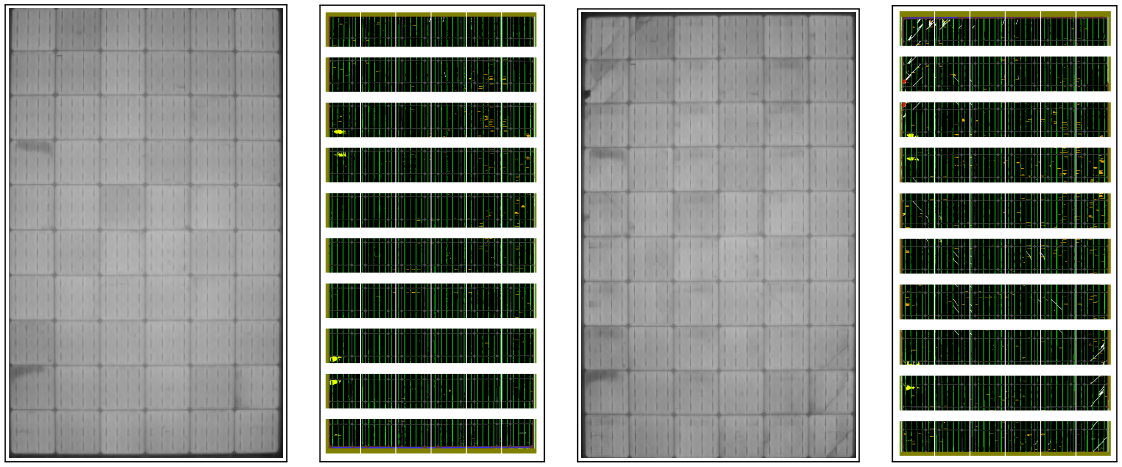
(a) (b) (c) (d) (e)

Figure 9.2: Module 19040-12 (a) EL image before PID stress at 100% Isc, (b) EL image after PID stress at 100% Isc, (c) EL image after PID stress at 10% Isc and brightened for visual evaluation, (d) EL image after PID stress at 10% Isc as input to the prediction model, and (e) the prediction mask for the EL image after PID stress at 10% Isc

Figure 9.3 shows a set of images for a single PV module from the TC dataset. The set consists of EL images and prediction masks for each PV module taken before and after the TC600 stress test. In this dataset, all images were recorded at a high current bias equal to Isc current, so the regression analysis results are valid for this dataset. Cracks appear in the EL image and prediction mask (white lines) taken after the stress test but not before the stress test. This was unexpected since thermal cycling does not typically cause cracks, and none of the modules from the other three BOMs developed significant cracks. Interestingly, the cracks showed a similar pattern of long, 45°diagonal cracks developing in the top left and bottom right corners across all three modules.

9.3 Results

Figure 9.4 shows the regression analysis between module power loss and the mean percentage of dark cells after PID with an r^2 value of 55%. This indicates that the increase in the average dark cell pixel percentage can explain 55% of the decrease in Pmp. However, recall that correlation does not imply causation and the validity of this analysis is questionable because



(a) (b) (c) (d)

Figure 9.3: Module 19058-03 (a) EL image before TC stress, (b) prediction mask before TC stress, (c) EL image after TC600, and (d) EL prediction mask after TC600

pre-stress images were recorded at 100% of Isc current. In contrast, post-stress images were recorded at 10% of Isc current.

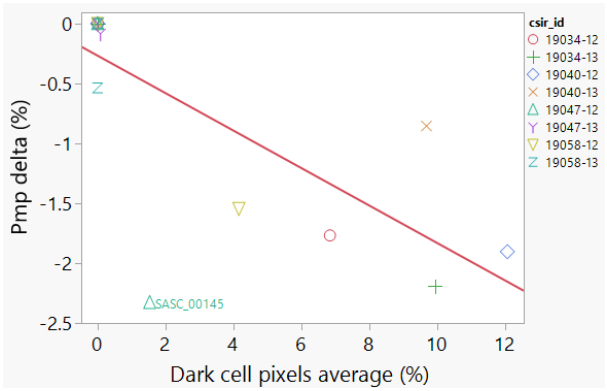


Figure 9.4: Regression analysis between power loss and percentage of dark cells predicted by SCDD

Figure 9.5 shows the regression analysis between module power loss and the mean percentage of crack cells on three modules from one BOM that were subjected to thermal cycling. An r^2 value of 55% was computed. This indicates that the increase in the cracked cell percentage can explain 55% of the decrease in Pmp.

9.3.1 Sensitivity Analysis

In principle, the two regression models prove that power loss correlates with the output from the SCDD tool for certain defects and stressors. However, two important sensitivities must be considered when interpreting the regression analysis results. Both sensitivities relate to the quality of the data encoded in the EL image and subsequent impact on the model predictions. One issue relates to the method used when recording the EL images, and the other relates to the method used for pre-processing of EL images.

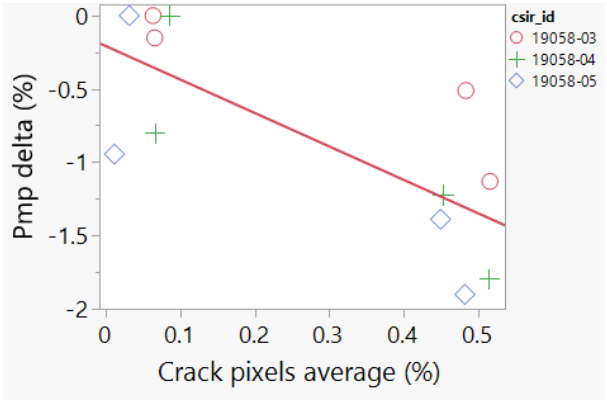


Figure 9.5: Regression analysis between power loss and percentage of crack cells predicted by SCDD

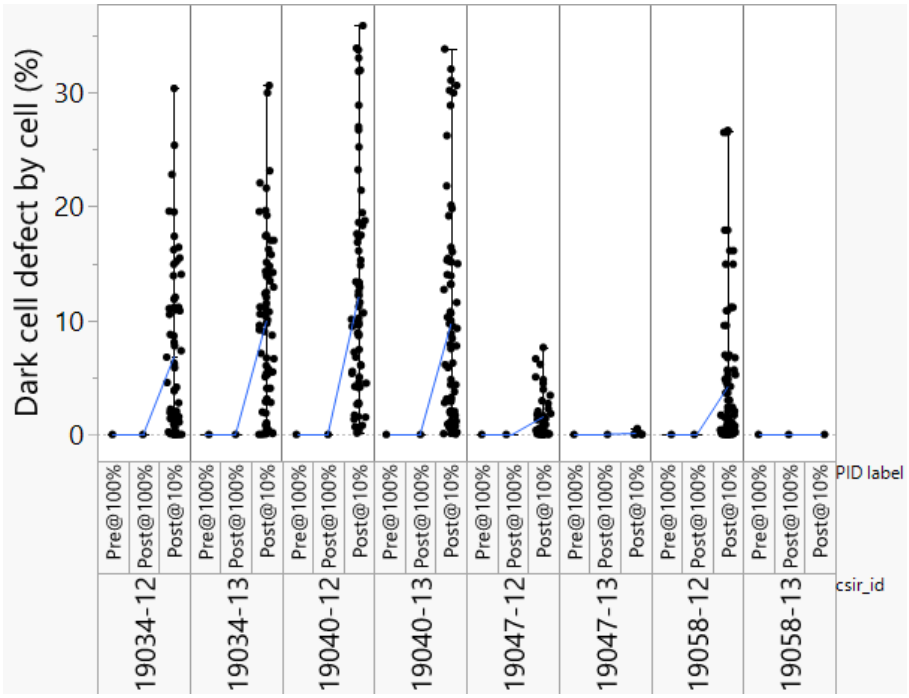


Figure 9.6: Percentage of dark cell pixels pre- and post- PID at different levels of bias current

First, the method used when recording the EL image can have an impact on the predictions. Figure 9.6 shows the change in the percentage of dark cell pixels predicted in the pre-stress and post-stress EL images. The pre-stress dataset only includes EL images recorded at 100% of Isc current, while the post-stress dataset includes EL images at 10% and 100% of Isc current. At 100% of Isc current, the prediction model does not classify any of the pixels as dark cells, even after the PID stress. The percentage of pixels predicted as dark cells increases significantly when the current bias during EL imaging is decreased from 100% of Isc current to 10% of Isc current, as per the international standard. The IEC TS 62804-1 describes the test methods for the detection of potential-induced degradation. In that technical specification, the test sequence includes EL imaging at both 100% of Isc current and 10% of Isc current because the PID degradation is more easily seen to the human observer at the low bias setting. It follows that the SCDD model would also detect more dark cell pixels in images taken at low bias. Unfortunately, the pre-stress EL images at 10% of Isc were not available to assess the percentage of dark cell pixels at that bias current prior to the stress. However, this result indicates that dark cell predictions and subsequent regression analysis are sensitive to the current bias applied to the module when capturing the EL image. Furthermore, none of the EL images in the training dataset were taken at 10% of the Isc current.

Second, the method used for pre-processing the EL images can have an impact on the predictions. Pre-processing may include steps to improve the image's visual appearance by adjusting sharpness, contrast, vignette, and brightness. For example, a dark EL image taken at 10% of Isc may be brightened before being included in a report for a client.

Figure 9.7 shows five different versions of a single EL image from the PID dataset recorded at 100% of Isc current. The image was manipulated in GIMP to generate four additional images: low contrast (-25), high contrast (+25), low brightness (-50), and high brightness (+50).

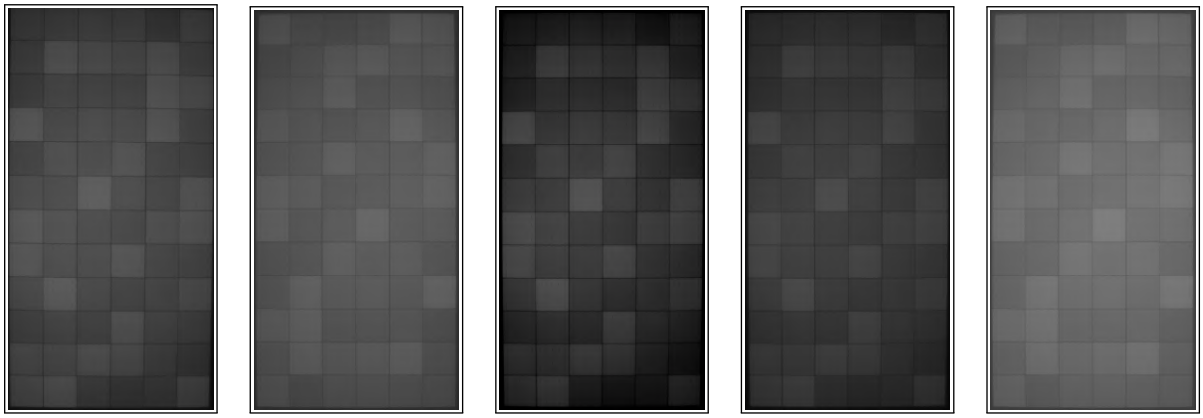
Figure 9.8 shows the resulting prediction masks for the original and manipulated images. The predicted percentage of dark cell pixels (dark red) varied significantly (0.1 - 5.9%) depending on the changes in brightness and contrast.

Figure 9.9 shows five different versions of a single EL image from the TC dataset recorded at 100% of Isc current. The image was manipulated in GIMP to generate four additional images: low contrast (-25), high contrast (+25), low brightness (-50), and high brightness (+50).

Figure 9.10 shows the resulting prediction masks for the original and manipulated images. The predicted percentage of crack pixels (white) remained relatively constant (0.45 - 0.52%) regardless of the changes in brightness and contrast.

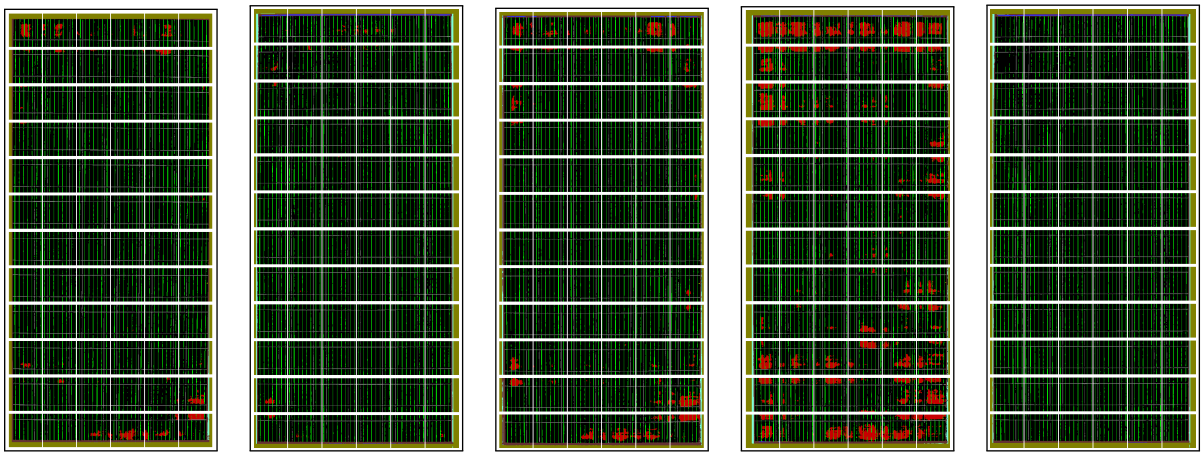
9.4 Conclusion

These experiments demonstrate the potential for using the SCDD tool to gain insights into PV module power loss. This work investigated linear correlations between the output of a semantic segmentation model trained to detect defects in EL images and the corresponding I-V characteristics of PV modules exposed to accelerated stress testing. The output power of PV modules that were subjected to damp heat and thermal cycling was correlated to the percentage of dark cells and the percentage of cracks, respectively. The simple linear regression models explain 55% of the decrease in module power by the increase in dark cell pixels after the PID



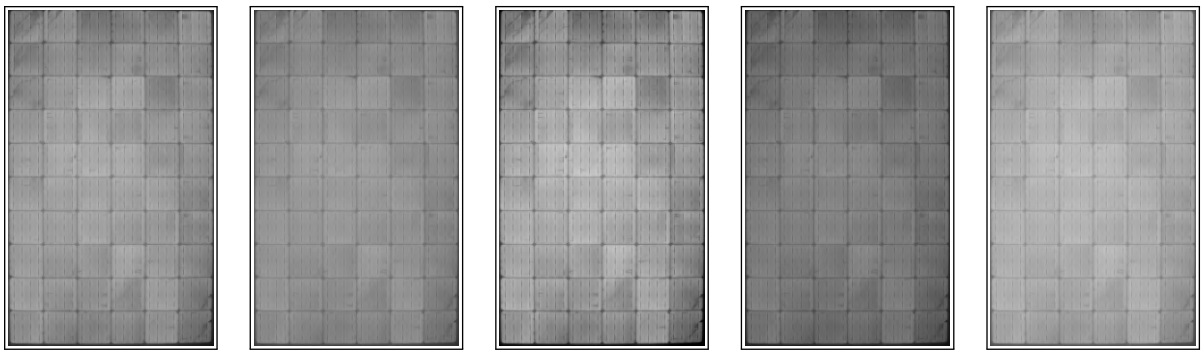
(a) (b) (c) (d) (e)

Figure 9.7: SASC-00145 EL images post PID (a) original, (b) low contrast, (c) high contrast, (d) low brightness, and (e) high brightness



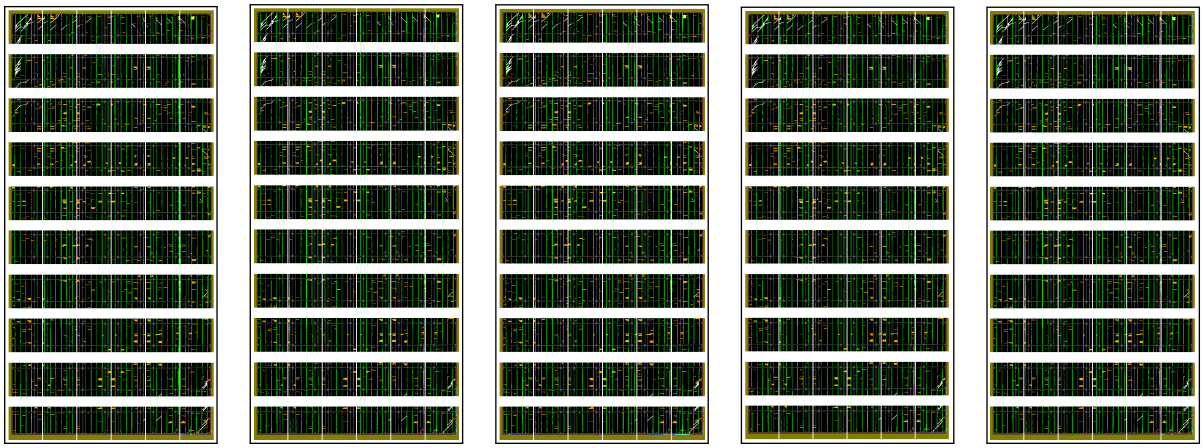
(a) (b) (c) (d) (e)

Figure 9.8: SASC-00145 predictions post PID and the average of the percentage of dark cells (a) original - 1.5%, (b) low contrast - 0.4%, (c) high contrast - 2.6%, (d) low brightness - 5.9%, and (e) high brightness - < 0.1%



(a) (b) (c) (d) (e)

Figure 9.9: SASC-00172 EL images post TC600 (a) original, (b) low contrast, (c) high contrast, (d) low brightness, and (e) high brightness



(a) (b) (c) (d) (e)

Figure 9.10: SASC-00172 predictions post TC600 (a) original - 0.48%, (b) low contrast - 0.45%, (c) high contrast - 0.48%, (d) low brightness - 0.48%, and (e) high brightness - 0.52%

stress and by the increase in crack pixels after the TC stress. However, the regression models are sensitive to the characteristics of the EL image. EL images taken at a low electrical current bias resulted in a higher percentage of dark cell pixel predictions compared to EL images taken at a higher electrical current bias. The results also show that predictions for dark cell pixels are sensitive to changes in brightness and contrast. Predictions for crack pixels appear robust to changes in brightness and contrast.

9.5 Next steps

Future work will focus on increasing the training dataset size for fully-supervised models. Specific attention will be given to increasing the samples for underrepresented defects such as dark cells. Further investigations on the sensitivity of some defects to image preprocessing are also necessary to support the development of a web-based tool designed to analyse EL images. EL images uploaded to the web-based tool may have a wide range of pre-processing and imaging techniques. For instance, EL images taken in the field may yield significantly different predictions compared to EL images taken in the lab under controlled conditions.

Chapter 10

Conclusion

10.1 Summary

Solar PV systems are driving rapid growth in the renewable electrical energy industry. PV systems generate electricity when PV modules convert the radiant energy from the sun, in the form of photons, into electrical energy, in the form of electrons. The quality and reliability of solar PV modules are critical to ensuring the security of the future electricity grid as the contribution from solar generators continues to grow.

Electroluminescence (EL) imaging effectively detects micro-cracks in PV modules made from silicon cells. EL images enable defect detection in solar PV modules that are otherwise invisible to the naked eye, much like an X-ray allows a doctor to detect cracks and fractures in bones. EL imaging is widely used throughout the solar industry to assess the quality and reliability of PV modules. PV modules made from crystalline silicon cells are susceptible to cracking, and cracked cells lead to decreased electricity generation over time.

Computer vision methods have proven effective for detecting and classifying solar cells as defective or normal based on features in the EL images of solar cells. Statistical methods, support vector machines (SVMs), and convolutional neural networks (CNNs) have been used for object detection and localisation of various defect types, typically focused on cracks, inactive areas, and gridline defects. Deep-learning methods have also been applied to classify solar cell images into single-defect bins or develop methods to locate different defect classes within a bounding box overlaid on the image.

This research evaluated semantic segmentation models for SCDD for the first time in the published literature. Modern, deep learning techniques for computer vision provide an effective method for converting terabytes of raw data contained in EL images into valuable statistical information for stakeholders all across the PV industry value chain. The contributions of this work included two published articles [Pratt *et al.* 2021a] [Pratt *et al.* 2023], one article under review [Torpey *et al.* 2024], and several benchmark datasets to support further research in SCDD

using semantic segmentation.[1] The benchmark dataset for fully supervised models is one of only two publically available for multi-class semantic segmentation of crystalline silicon cells.

Specific contributions and findings from Chapters 5-8 are compiled in the following list.

- A new dataset was required to enable research in semantic segmentation models for the EL images of solar cells. The UCF dataset was under development in parallel but not available publicly at the start of this research. Thus, the SCDD dataset was initiated and expanded throughout the research.
- The new SCDD dataset for semantic segmentation contains more detail concerning the size and shape of cracks and gridline defects compared to the UCF dataset. While the UCF dataset combined defects into 'conglomerations', the SCDD provide a finer resolution for each defect type. The SCDD dataset also included features intrinsic to solar cells, so additional information about the cell architecture can be learned.
- Ground truth images created in GIMP resulted in high-resolution masks but require 30-60 minutes per image. GIMP is a feature-rich image processing tool, and it's free to use. Key features such as 'growing', 'shrinking' and 'anti-aliasing' proved helpful when fine-tuning the ground truth images.
- Consistency in labelling across multiple labellers was mitigated with calibration images and a single advanced labeller who reviewed and edited all images. Having a single person review all the annotations contributed by the team also helped provide consistency.
- Semantic segmentation shows significant promise for the field of SCDD. The U-Net provided a good starting point despite a relatively small dataset and no hyperparameter optimisation. The predictions from an 'off the shelf' model were impressive enough to motivate further work.
- The output from the U-Net can be used to quantify meaningful changes in EL images taken throughout an accelerated stress test sequence because the actual size of a defect can be accurately quantified. Changes visually observed in the EL images were correlated to changes in the U-Net predictions.
- The accuracy of the ground truth images impacted precision and recall, so the images in the test dataset should get extra special attention concerning the accuracy of the labels. Long-narrow defects were especially susceptible to annotation errors along the edges.
- The DeepLabv3+ model showed the highest mIoU and mRcl for defects. The DeepLabv3+ model was included because it promised to perform better along the edges.
- Defect detection was marginally better for mono-c-si than multi-c-si for all fully-supervised models evaluated. This is consistent with findings from other published research and is likely due to the grain boundaries intrinsic to the multi-c-si cells.

[1] https://github.com/TheMakiran/BenchmarkELimages

- The mIoU was significantly higher for predicting larger features versus smaller defects. The features tend to be larger, so the edge pixels comprise a smaller percentage of features than defects, such as cracks, gridlines, and ribbon corrosion.
- The predictions for cracks and gridline defects when training with inverse class weights and custom class weights resulted in dilated predictions, leading to lower precision and IoU. While the dilated predictions may be useful for practical fault detection in an application meant to minimise false negatives, the model performance metrics suffer because the precision suffers.
- Long, narrow objects are especially susceptible to annotation errors along the border pixels. This seems true for any feature with a relatively high percentage of edge pixels. Edge pixels are difficult to annotate because the edge of a feature is not black-and-white in an EL image. The boundary is a transition from light to dark, so choosing where to draw the edge of the feature is a bit subjective.
- We published the first and only large-scale analysis detailing results from semi-supervised and self-supervised models trained for SCDD in EL images.
- Three different architectures achieved a new state-of-the-art performance benchmark for SCDD. Supervised training on a large OOD dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and semi-supervised pretraining (CCT) all yield statistically equivalent performance for mIoU on this dataset. This suggests that different architectures evaluated have some fundamental commonality in how they learn to detect defects and features in EL images.
- We show that self-supervised pretraining in the SCDD domain does not yield a statistically significant improvement over supervised training, contrary to recent results published for other domains. This result may have been due to the nature of EL images or the data augmentation steps that excluded important methods such as occlusion.
- We published an unlabelled dataset with over 150,000 EL images of solar cells. This dataset enables further research into SCDD in EL images using semi-sl and self-sl methods.
- Long, narrow defects such as cracks and gridline defects remain challenging to detect, suggesting alternative models should be investigated or a custom model should be designed.

10.2 Future Work

Future work includes increasing the size of the training dataset for fully-supervised models. Specific attention will be given to increasing the number of images for underrepresented defects and features. New defects will also be added to the dataset as they arise. For example, the 'star' defect has already been added since the writing of this thesis.

Models that focus on small object detection will be evaluated. This research exposed a challenge in detecting long, narrow defects such as cracks and gridline defects in EL images. Similar issues were noted in benchmark datasets such as COCO and ImageNet. For example, while

a deep-learning model may provide accurate predictions for a bird's body or a stop sign, the bird's legs and the signpost often need to be more accurately predicted. The legs and signposts have a long, narrow shape, which can be challenging to detect. The UCF researchers sidestep this challenge by labelling the cracks and defects as regions or 'conglomerations' that are easier to detect but have far less resolution than the SCDD dataset developed for this research. Vision transformers (ViTs), a promising new architecture in computer vision, may also improve predictions for both fully-supervised and self-supervised models in EL images.

The research and development will extend to training machine learning models that can correlate the predictions from SCDD to module-level I-V characteristics, as noted in the introduction. One paper showcasing this work was already presented at the South African Solar Energy Conference (SASEC) in November 2023 [Zandamela *et al.* 2023]. The paper presented the correlations between predictions from the DeepLabv3 model and the module-level I-V measurements taken throughout a long-term accelerated stress test sequence conducted at the CSIR. A cloud-based tool is also being developed at the CSIR that will enable users to upload batches of EL images of PV modules, visualise the predictions, and download reports. The reports include tabular data to quantify defects at the cell level.

Appendix A

Appendix

A.1 Library of features and defects

The most current dataset includes 29 classes divided into two groups: defects and features. Features correspond to an intrinsic part of the PV module design, and they are needed to meet the form and function of the module. Defects correspond to extrinsic faults in the PV module and may negatively impact module performance now or in the future. This section of the appendix includes a brief description of each class, an EL image example, and the corresponding ground truth mask. Many of the definitions were paraphrased from the International Electrotechnical Commission [2018].

A.1.1 Defects

Table A.1 describes each defect. The table includes the coded label, importance, and source for each defect listed. The coded label was used when coding the ground truth masks. The importance is either high, medium or low, corresponding to the impact of the defect on module performance now or over time. The source is either a factory, field, unknown, or measurement artefact corresponding to the likely defect source.

Table A.1: Description of defects

Tags	**Label** and Description
#10 High Factory	**Crack ribbon edge** identifies a short crack with origins at the edge of the cell where the interconnect ribbon folder over the cell edge. The crack has a minimal impact at this stage, but it is likely to increase over time leading to longer cracks and potentially inactive areas that impact performance. A model that can detect this small defect would be highly valuable to the manufacturer so that the source of the cracks can be rectified.

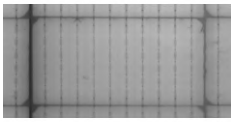
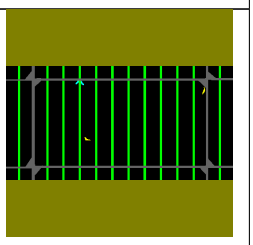
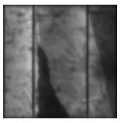
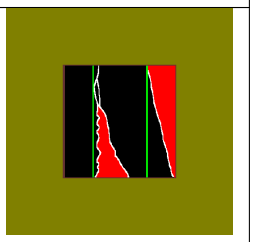
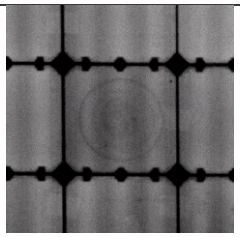
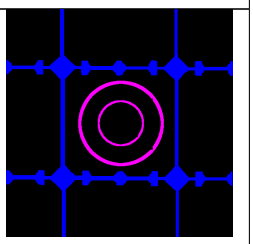
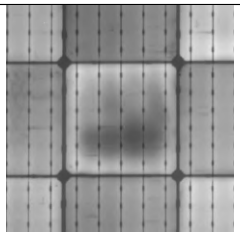
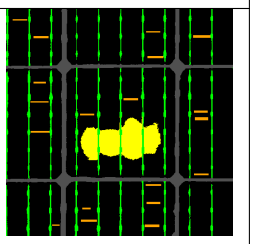
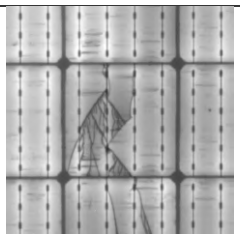
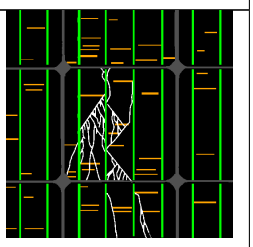
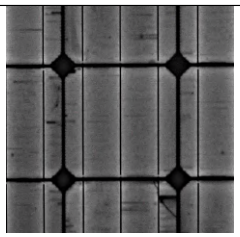
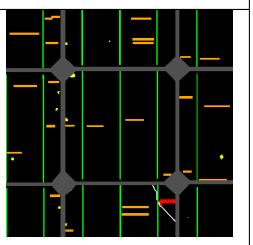
#11 High Field	**Inactive region** identifies a region of the solar cell that has been completely disconnected from the rest of the cell due to a severe crack. The inactive region will continue to generate electron-hole pairs when exposed to the sun. Still, the electrons recombine in place because there are no alternative current paths to the cell circuit in the module. This type of cracking may lead to hot spots.
#12 Low Factory	**Rings** identify a region of the solar cell impacted by handling during manufacturing. The rings are formed by contaminated suction cups used to move the cells in the factory.
#13 Medium Unknown	**Material** defects identify a dark area of the cell that is not readily assignable to other classes. This is a catch-all, or 'other', category of defects.
#14 High Field	**Cracks** identify micro-cracks in the silicon solar cell. Cracks can lead to degraded module performance by creating hotspots and inactive areas. Cells with cracks running perpendicular to the interconnect ribbons are susceptible to further degradation if the cracks expand, but less severely than cells with cracks running parallel to the interconnect ribbons. Cracks parallel to the interconnect ribbon are more likely to lead to inactive regions. The risk of cracks leading to large inactive regions is reduced in modern PV module architectures assembled with more interconnect ribbons than in earlier generations assembled with two or three interconnect ribbons.
#15 Medium Factory	**Gridline defects** identify a region of the cell with missing, broken, or delaminated grid finger lines. Missing or broken grid fingers are generally stable, and their influence is captured in the cell's efficiency. Grid finger adhesion may degrade over time, degrading the module performance.
#16 Low Artifact	**Splice** is not a defect nor a feature, but it was assigned to the defect group early on in the development of the dataset. The splice identifies a long, narrow line in the EL image corresponding to an artefact of the measurement. Some EL imaging stations will use multiple cameras and stitch them together with software to provide a module-level image. The splice identifies the regions where the separate images come together. They were included in the training to minimise confusion with other classes such as spacing, gridlines, and interconnect ribbons with long, narrow shapes.
#17 Low Factory	**Dark cell** identifies cell-wide lower minority carrier lifetime. This may occur from a lower minority carrier lifetime based on the position of the wafer in the ingot from which it came and the impurities in the wafer that are inherently contained. Such defects may be alternatively associated with cell shunting.

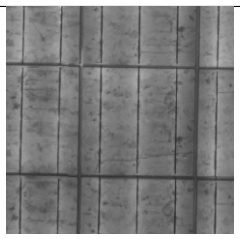
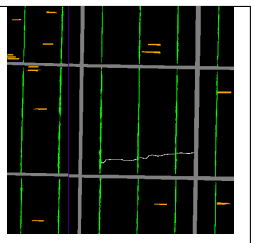
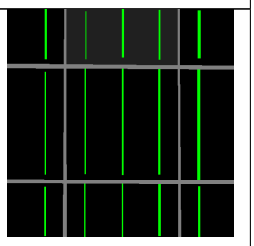
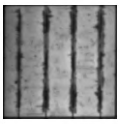
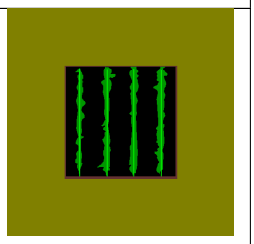
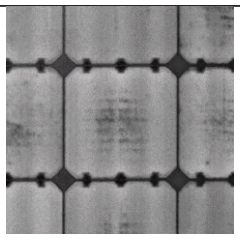
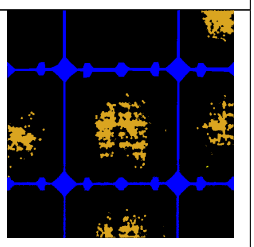
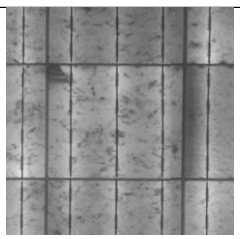
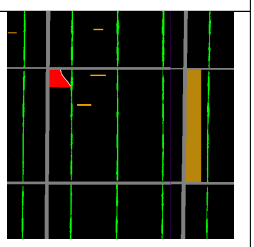
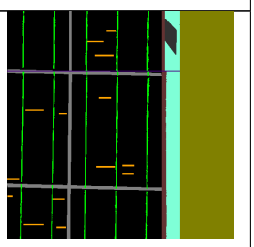
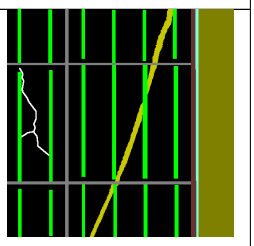
#18 High Factory	**Corrosion ribbon** identifies regions around the interconnect ribbons impacted by solder flux interaction with the grid finger metallization-silicon interface. This may indicate a substandard soldering process leading to higher series resistance, hot spots, module degradation, and failure. These defects, which degrade performance, increase over time with elevated humidity.
#19 Medium Factory	**Belt mark** identifies dark regions of the cell that developed during cell firing. After the frontside metallisation is printed onto the solar cell, the cell is placed on a conveyor belt that carries it through a high-temperature furnace to harden the silver paste and bond it with the silicon. This defect often resembles a tyre track or herringbone pattern that mirrors the conveyor belt. Belt mark defects are generally stable over time, and the impact is captured in the initial efficiency of the cell and module.
#20 Medium Factory	**Dark edge** identifies cell regions with reduced lifetime from the silicon casting process. They occur over particular grains or regions of the ingot that contain elevated defect or impurity concentrations. They are generally stable over time, and their influence is largely captured by the initial efficiency of the cell and module.
#23 Low Artifact	**Measurement artefact** identifies a bright area of the cell that is not readily assignable to other classes. Bright spots can occur in the EL image if the light is not properly controlled during the imaging process. This is a catch-all, or 'other', category of defects for bright regions. The dataset contains only a few images with this defect.
#25 High Field	**Scuff** identifies a region of the module that was damaged during handling, especially modules with polymer backsheets. Polymer back sheets are rather pliable, so when an object is dragged along the backside of a module with sufficient pressure, the scuff defect appears in the EL image. This defect may also exhibit micro-cracks emanating from the scuff if the pressure on the backsheet was especially high. The scuff defect may also carry over to adjacent cells.
#26 High Field	**Corrosion cell** identifies the region of a cell that has been corroded from the outside edges inward. This defect can occur along all four edges and may be worse on corner cells where two edges of the cell are close to the module perimeter. This corrosion may be due to humidity that has passed through the polymer backsheet between two adjacent cells or cracks.

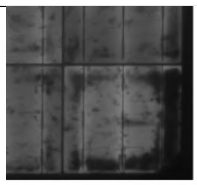
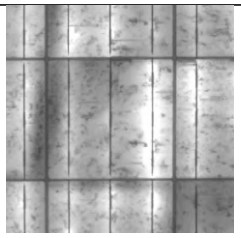
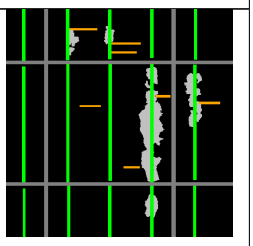
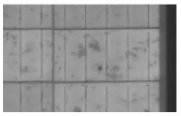
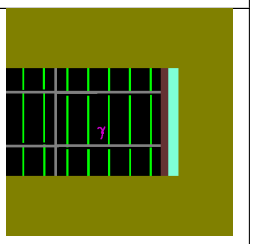
#27 High Factory	**Brightening** identifies regions of the cell that are significantly brighter along some interconnect ribbons than others. The defective region of the cell corresponds to the darker regions, although the model was trained to find the brighter regions. The dark regions indicate improper interconnection or tab ribbon bonding. Series resistance is elevated over the cell regions associated with improper interconnection or tab ribbon bonding. The increased resistance leads to a higher current forced through other regions, which yields a higher probability of hot spots and module failure.
#28 High Field	**Star** identifies a small cell region impacted by a point force. Similar to the scuff defect, the point defect results from an object that hits the backsheet once compared to an object dragged across the surface, leading to a scuff defect. The star defect may also exhibit cracks emanating from the centre, giving it a star-like appearance.

Table A.2 provides an example of each defect.

Table A.2: Library of defects

ID#	Label	Colour	RGB	EL	GT
10	Crack at the ribbon and cell edge	Teal	[0,255,255]		
11	Inactive region	Red	[255,0,0]		
12	Rings	Violet	[255,0,255]		
13	Material	Bright yellow	[255,255,0]		
14	Crack	White	[255,255,255]		
15	Gridline	Orange	[255,165,0]		

16	Splice	Indigo	[75,0,130]		
17	Dead cell	Light black	[32,32,32]		
18	Corrosionn ribbon	Dark green	[0,150,0]		
19	Belt mark	Goldenrod	[218,165,32]		
20	Dark edge	Dark golden rod	[184,134,11]		
23	Measurement artifact	Dark brown	[50,50,50]		
25	Scuff	Dark yellow	[200,200,0]		

26	Corrosion cell	Dark green	[0,100,0]		
27	Brightening	Silver	[192,192,192]		
28	Star	Deep pink	[200,0,200]		

A.1.2 Features

Table A.3 provides a description of each feature. The table includes the coded label, the label in bold text, and the description. The coded label was used when coding the ground truth masks.

Table A.3: Description of Features

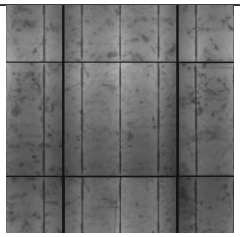
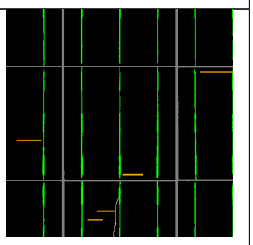
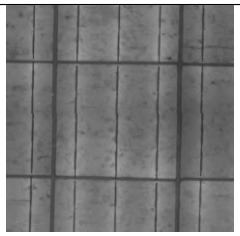
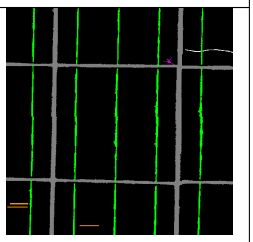
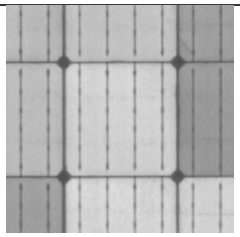
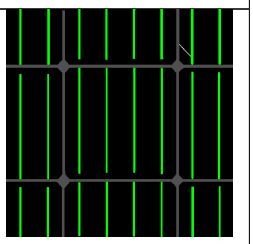
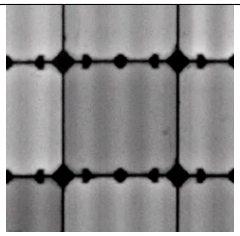
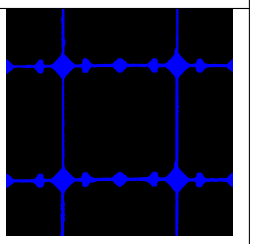
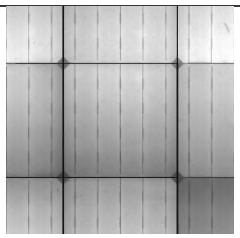
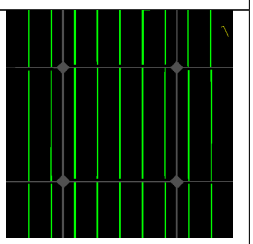
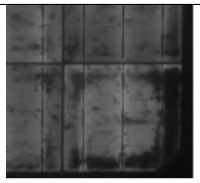
Tags	**Label** and Description
#0	**Background** identifies all the spaces in the EL image not associated with other defects and features. The background class accounts for the majority of pixels in every image. Importantly, the grain boundaries seen in the multi-c-si wafers are labeled as background. Mono-c-si wafers do not show these grain boundaries characteristic of multi-c-si wafers.
#1	**Spacing multi** identifies the gaps between adjacent multi-c-si cells assembled into the PV module. Multi-c-si wafers are cut from square ingots, so the corners of each cell have a sharp corner.
#2	**Spacing mono** identifies the gaps between adjacent mono-c-si cells assembled into the PV module. Mono-c-si wafers are cut from round ingots, so the corners of each cell tend to have a rounded shape. In some modern mono-c-si cells, the rounded corners have been reduced making it difficult to be sure about the wafer type. The surface of mono-c-si wafers is uniform, lacking any dark regions associated with difference crystallographic regions.

#3	**Spacing dogbone** identifies the gaps between adjacent back-contact cells that have been assembled with 'dogbone' shaped interconnect. Adjacent cells are typically connected with interconnect ribbons that can be seen on the frontside of every cell in the module. Back-contact cells do not have any metal on the frontside, and the dogbone interconnect is used instead of the ribbons. The dogbone shape only appears between the series connected cells separated by a horizontal space, not between the adjacent cells separated by a vertical space.
#4	**Ribbons** identifies the interconnect wires that connect the adjacent cells in series. These ribbons typically run in a vertical orientation from the top to the bottom of the module. Each ribbon connects the top side of one cell to the bottom side of the next cell in series, creating a positive-to-negative-to-positive series that strings all the cells in the module together. In this way, the individual cell voltages of approximately 0.6V build to approximately 36 V for a 60-cell module and 43 V for a 72-cell module. The ribbons bend at the edges of adjacent cells creating stress at the edges which can lead to the crack ribbon edge defect when not assembled correctly.
#5	**Border** identifies the space between the edge of a cell and the outside edge of the module. The is similar to the cell spacing layer, except there is no adjacent cell on the other side of the border. The space between the cell and the edge of a module is typically wider than the space between adjacent cells due to the design requirements of the international standards for PV module type approval. A minimum distance must be maintained between the active area of the cell and edge of the laminate to minimize the risk of shock through creepage pathways.
#6	**Text** identifies text that may be scribed into the glass or otherwise marked on the module to provide a unique identifier.
#7	**Padding** identifies the region around the cell that was added during the pre-processing. This is not an actual feature of a PV module, nor is it any kind of defect. The padding is adding during pre-processing to maintain the full cell at the centre of every image.
#8	**Clamp** identifies a region at the module edge that is used to hold the module in place during EL imaging. If the clamp extends over the active area of the cell, it will create a dark region in the EL image. This is not a feature or a defect, but rather a measurement artifact.
#9	**Busbars** identifies the region of the module where adjacent columns of cells are connected with a thick metal busbar. Each column of a crystalline silicon wafer-based module consists of series-connected cells that are isolated from the cells in the columns to the left and right. The busbar connects two adjacent columns at the top of the module and the bottom of the module which allows all the cells to be connected in series as one long string of cells.
#21	**Frame edge** identifies the four edges of the module frame. PV modules are typically framed in aluminum to provide rigidity and protection during handling. The frame edge is often seen in the module-level EL images as the top lip of the frame wraps over the border region.

#22	**Junction box** identifies the location where the busbars are connected to the standard leads. Technically, the junction box is a polymer enclosure inside which the busbars are connected to leads that protrude from the junction box. Since all the electrically active parts of the enclosure are covered, they should not show any light regions. The images with junction box features in this dataset were taken before the junction box was attached or the lid was closed. The standard leads that protrude from the junction are used to connect adjacent modules together in series or parallel to build up the string voltage or current, respectively.
#24	**Spacing mono half-cut** identifies the gaps between adjacent mono-c-si half-cut cells assembled into the PV module. These cells are cut from round ingots that exhibit rounded corners on all four corners creating a pqeudo-square shape. When the pseudo-square is cut in half, then two corners are rounded and two corners are square. By contrast, the half-cut cells made from multi-c-si wafers are square on all four corners jsut like the full-sized wafers cut from multi-c-si ingots.

Table A.4 provides an example of each feature.

Table A.4: Library of features

ID#	Label	Colour	RGB	EL	GT
0	Background	Black	[0,0,0]		
1	Spacing multi	Light grey	[128,128,128]		
2	Spacing mono	Dark grey	[80,80,80]		
3	Spacing dogbone	Bright blue	[0,0,255]		
4	Ribbons	Lime green	[0,255,0]		
5	Border	Brown	[100,50,50]		

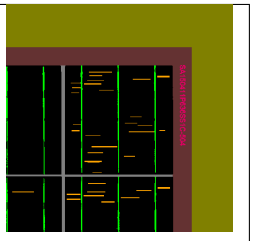
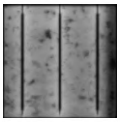
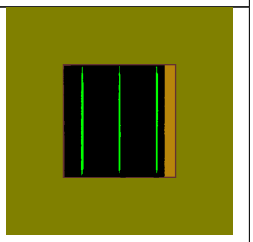
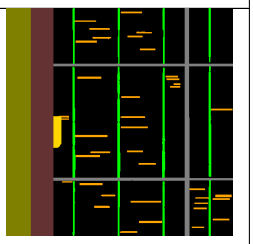
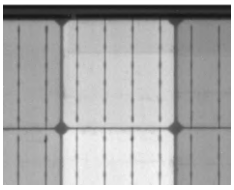
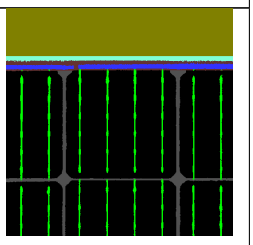
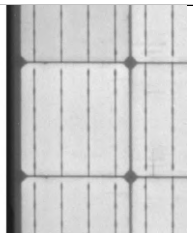
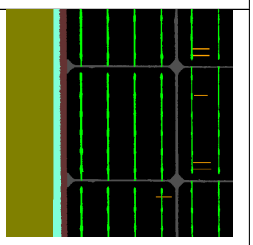
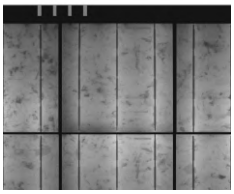
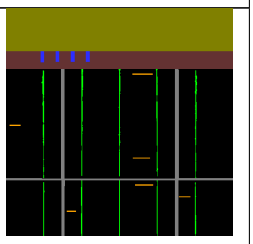
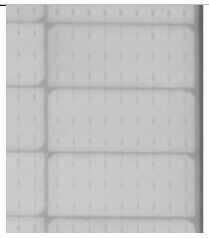
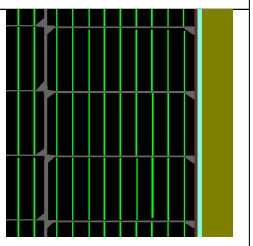
6	Text	Pink	[255,0,100]		
7	Padding	Olive	[128,128,0]		
8	Clamp	Gold	[255,215,0]		
9	Busbars	Blue	[50,50,255]		
21	Frame edge	Aquamarinee	[127,255,215]		
22	Junction box	Blue	[45,45,255]		
24	Spacing mono half-cut	Medium grey	[100,100,100]		

Table A.5: CSV Data Table

Label	Class	Description	Color	Red	Green	Blue	Gray
0	feature	bckgnd	Black	0	0	0	0
1	feature	sp multi	gray light	128	128	128	128
2	feature	sp mono	gray dark	80	80	80	80
3	feature	sp dogbone	Blue - bright	0	0	255	29
4	feature	ribbons	Green - lime	0	255	0	150
5	feature	border	Brown	100	50	50	65
6	feature	text	Pink	225	0	100	79
7	feature	padding	Olive	128	128	0	113
8	feature	clamp	Gold	255	215	0	202
9	feature	busbars	Blue	50	50	255	73
10	defect	crack rbn edge	Teal	0	255	255	179
11	defect	inactive	Red	255	0	0	76
12	defect	rings	Violet	255	0	255	105
13	defect	material	Yellow - bright	255	255	0	226
14	defect	crack	White	255	255	255	255
15	defect	gridline	Orange	255	165	0	173
16	defect	splice	Indigo	75	0	130	37
17	defect	dead cell	light black	32	32	32	32
18	defect	corrosion rbn	Green	0	150	0	88
19	defect	belt mark	Golden rod	218	165	32	166
20	defect	edge dark	Dark golden rod	184	134	11	135
21	feature	frame edge	Aqua marine	127	255	215	212
22	feature	jbox	blue	45	45	255	69
23	defect	meas artifact	dark brown	50	50	50	50
24	feature	sp mono halfcut	gray medium	100	100	100	100
25	defect	scuff	Yellow - dark	200	200	0	177
26	defect	corrosion cell	Green - dark	0	100	0	59
27	defect	brightening	silver	192	192	192	192
28	defect	star	deep pink	200	0	200	83

A.2 Colour coding scheme for ground truth masks and predictions

Table A.5 lists the defects and features in the SCDD dataset. The label IDs for the coded masks and the RGB values are included.

References

[Abadi *et al.* 2015] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. Software available from tensorflow.org.

[Abdelhamid *et al.* 2014] Mahmoud Abdelhamid, Rajendra Singh, and Mohammed Omar. Review of microcrack detection techniques for silicon solar cells. *IEEE Journal of Photovoltaics*, 4(1):514–524, 2014.

[Agency 2022a] International Energy Agency. *Electricity Sector, IEA*. https://www.iea.org/reports/electricity-sector, 2022.

[Agency 2022b] International Energy Agency. *Renewables 2022, IEA*. https://www.iea.org/reports/renewables-2022, 2022.

[Agency 2022c] International Energy Agency. *Solar PV, IEA*. https://www.iea.org/reports/solar-pv, 2022.

[Akram *et al.* 2019a] M Waqar Akram, Guiqiang Li, Yi Jin, Xiao Chen, Changan Zhu, Xudong Zhao, Abdul Khaliq, M Faheem, and Ashfaq Ahmad. Cnn based automatic detection of photovoltaic cell defects in electroluminescence images. *Energy*, 189:116319, 2019.

[Akram *et al.* 2019b] M Waqar Akram, Guiqiang Li, Yi Jin, Xiao Chen, Changan Zhu, Xudong Zhao, Abdul Khaliq, M Faheem, and Ashfaq Ahmad. Cnn based automatic detection of photovoltaic cell defects in electroluminescence images. *Energy*, 189:116319, 2019.

[Alpher and Fotheringham-Smythe 2003] Alvin Alpher and Ferris P. N. Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003.

[Alpher *et al.* 2004] Alvin Alpher, Ferris P. N. Fotheringham-Smythe, and Gavin Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.

[Alpher 2002] Alvin Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002.

[Alves dos Reis Benatto *et al.* 2020] Gisele Alves dos Reis Benatto, Claire Mantel, Sergiu Spataru, Adrian Alejo Santamaria Lancia, Nicholas Riedel, Sune Thorsteinsson, Pe-

ter Behrensdorff Poulsen, Harsh Parikh, Søren Forchhammer, and Dezso Sera. Drone-based daylight electroluminescence imaging of pv modules. *IEEE Journal of Photovoltaics*, 10(3):872–877, 2020.

[Anwar and Abdullah 2014] Said Amirul Anwar and Mohd Zaid Abdullah. Micro-crack detection of multicrystalline solar cells featuring an improved anisotropic diffusion filter and image segmentation technique. *EURASIP Journal on Image and Video Processing*, 2014:1–17, 2014.

[Badrinarayanan *et al.* 2017] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[Banda and Barnard 2018] P Banda and Lynette Barnard. A deep learning approach to photovoltaic cell defect classification. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, pages 215–221, 2018.

[Bandyopadhyay] Avimanyu Bandyopadhyay. *Avimanyu786SVG version: Tukijaaliwa - File:AI-ML-DL.png*. https://commons.wikimedia.org/w/index.php?curid=90131352.

[Bdour *et al.* 2020] Mathhar Bdour, Zakariya Dalala, Mohammad Al-Addous, Ashraf Radaideh, and Aseel Al-Sadi. A comprehensive evaluation on types of microcracks and possible effects on power degradation in photovoltaic solar panels. *Sustainability*, 12(16):6416, 2020.

[Becquerel 1839] ME Becquerel. Mémoire sur les effets électriques produits sous l'influence des rayons solaires. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 9:561–567, 1839.

[Bedrich *et al.* 2016] Karl G Bedrich, Martin Bliss, Thomas R Betts, and Ralph Gottschalg. Electroluminescence imaging of pv devices: Camera calibration and image correction. In *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, pages 1532–1537. IEEE, 2016.

[Bedrich *et al.* 2018] Karl G. Bedrich, Wei Luo, Mauro Pravettoni, Daming Chen, Yifeng Chen, Zigang Wang, Pierre J. Verlinden, Peter Hacke, Zhiqiang Feng, Jing Chai, Yan Wang, Armin G. Aberle, and Yong Sheng Khoo. Quantitative electroluminescence imaging analysis for performance estimation of pid-influenced pv modules. *IEEE Journal of Photovoltaics*, 8(5):1281–1288, 2018.

[Berg *et al.* 2014] Thomas Berg, Jiongxin Liu, Seung Woo Lee, Michelle L. Alexander, David W. Jacobs, and Peter N. Belhumeur. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[Bharathi and Venkatesan 2022] S Bharathi and P Venkatesan. Enhanced classification of faults of photovoltaic module through generative adversarial network. *IJEER*, 10(3):579–584, 2022.

[Bloomberg] Bloomberg. *China added solar power capacity as big as Eskom in one year*. https://mybroadband.co.za/news/energy/485549-china-added-solar-power-capacity-as-big-as-eskom-in-one-year.html.

[Brand *et al.* 2015] Amy Brand, Liz Allen, Micah Altman, Marjorie Hlava, and Jo Scott. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2):151–155, 2015.

[Brostow *et al.* 2009] Gabriel J. Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognit. Lett.*, 30(2):88–97, 2009.

[Buerhop *et al.* 2022] Claudia Buerhop, Lukas Bommes, Jan Schlipf, Tobias Pickel, Andreas Fladung, and Ian Marius Peters. Infrared imaging of photovoltaic modules: a review of the state of the art and future challenges facing gigawatt photovoltaic power stations. *Progress in Energy*, 4(4):042010, 2022.

[Buerhop-Lutz *et al.* 2018] Claudia Buerhop-Lutz, Sergiu Deitsch, Andreas Maier, Florian Gallwitz, Stephan Berger, Bernd Doll, Jens Hauch, Christian Camus, and Christoph J. Brabec. A benchmark for visual identification of defective solar cells in electroluminescence imagery. In *European PV Solar Energy Conference and Exhibition (EU PVSEC)*, 2018.

[Chen *et al.* 2017] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[Chen *et al.* 2018a] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[Chen *et al.* 2018b] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[Chen *et al.* 2019] Haiyong Chen, Huifang Zhao, Da Han, and Kun Liu. Accurate and robust crack detection using steerable evidence filtering in electroluminescence images of solar cells. *Optics and Lasers in Engineering*, 118:22–33, 2019.

[Chen *et al.* 2020a] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[Chen *et al.* 2020b] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 1597–1607. PMLR, 13–18 Jul 2020.

[Chen *et al.* 2020c] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

[Chen *et al.* 2020d] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

[Chiou *et al.* 2011] Yih-Chih Chiou, Jian-Zong Liu, and Yu-Teng Liang. Micro crack detection of multi-crystalline silicon solar wafer using machine vision techniques. *Sensor Review*, 31(2):154–165, 2011.

[Chollet and others 2015] François Chollet et al. *Keras*. https://keras.io, 2015.

[Cordts *et al.* 2016] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[Cui *et al.* 2018] Shaoguo Cui, Lei Mao, Jingfeng Jiang, Chang Liu, and Shuyu Xiong. Automatic semantic segmentation of brain gliomas from mri images using a deep cascaded neural network. *Journal of healthcare engineering*, 2018, 2018.

[Davis and Goadrich 2006] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240, 2006.

[Deitsch *et al.* 2018] S Deitsch, C Buerhop-Lutz, E Sovetkin, A Steland, A Maier, F Gallwitz, and C Riess. Segmentation of photovoltaic module cells in electroluminescence images. arxiv 2018. *arXiv preprint arXiv:1806.06530*, 2018.

[Deitsch *et al.* 2019] Sergiu Deitsch, Vincent Christlein, Stephan Berger, Claudia Buerhop-Lutz, Andreas Maier, Florian Gallwitz, and Christian Riess. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*, 185:455–468, June 2019.

[Deitsch *et al.* 2021] Sergiu Deitsch, Claudia Buerhop-Lutz, Evgenii Sovetkin, Ansgar Steland, Andreas Maier, Florian Gallwitz, and Christian Riess. Segmentation of photovoltaic module cells in uncalibrated electroluminescence images. *Machine Vision and Applications*, 32(4), 2021.

[Demirci *et al.* 2019] MY Demirci, N Beşli, and A Gümüşçü. Defective pv cell detection using deep transfer learning and el imaging. *Proceedings Book*, 311, 2019.

[Demirci *et al.* 2021] Mustafa Yusuf Demirci, Nurettin Beşli, and Abdülkadir Gümüşçü. Efficient deep feature extraction and classification for identifying defective photovoltaic module cells in electroluminescence images. *Expert Systems with Applications*, 175:114810, 2021.

[Dhimish and Hu 2022a] Mahmoud Dhimish and Yihua Hu. Rapid testing on the effect of cracks on solar cells output power performance and thermal operation. *Scientific Reports*, 12(1):1–11, 2022.

[Dhimish and Hu 2022b] Mahmoud Dhimish and Yihua Hu. Rapid testing on the effect of cracks on solar cells output power performance and thermal operation. *Scientific Reports*, 12(1):1–11, 2022.

[Dhimish and Mather 2019] Mahmoud Dhimish and Peter Mather. Ultrafast high-resolution solar cell cracks detection process. *IEEE Transactions on Industrial Informatics*, 16(7):4769–4777, 2019.

[Dlamini *et al.* 2022] T Dlamini, R Chikwamba, M Mokonyama, and C Carter-Brown. Impact of the increasing fuel prices on the economy and possible alternatives and/or considerations in addressing increases in fuel prices. March 2022.
Unpublished CSIR internal report

[Dong *et al.* 2017] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21*, pages 506–517. Springer, 2017.

[Du *et al.* 2017] Bolun Du, Ruizhen Yang, Yunze He, Feng Wang, and Shoudao Huang. Nondestructive inspection, testing and evaluation for si-based, thin film and multi-junction solar cells: An overview. *Renewable and sustainable energy reviews*, 78:1117–1151, 2017.

[Emirov *et al.* 2011] Yu Emirov, A Belyaev, D Cruson, I Tarasov, A Kumar, H Wu, S Melkote, and S Ostapenko. Pinhole detection in si solar cells using resonance ultrasonic vibrations. In *2011 37th IEEE Photovoltaic Specialists Conference*, pages 002161–002163. IEEE, 2011.

[Ericsson *et al.* 2020] Linus Ericsson, Henry G. R. Gouk, and Timothy M. Hospedales. How well do self-supervised models transfer? *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5410–5419, 2020.

[Everingham *et al.* 2010] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[Fioresi *et al.* 2020] Joseph Fioresi, Rohit Gupta, Mengjie Li, Mubarak Shah, and Kristopher Daivs. *Deep Learning for Defect Classification of Photovoltaic Module Electroluminescence Images*. https://www.crcv.ucf.edu/wp-content/uploads/2018/11/Joe_F_Report.pdf, 2020. Accessed on 28th June 2023.

[Fioresi *et al.* 2022] Joseph Fioresi, Dylan J. Colvin, Rafaela Frota, Rohit Gupta, Mengjie Li, Hubert P. Seigneur, Shruti Vyas, Sofia Oliveira, Mubarak Shah, and Kristopher O. Davis. Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images. *IEEE Journal of Photovoltaics*, 12(1):53–61, 2022.

[Fisher *et al.* 2012] Graham Fisher, Michael R. Seacrist, and Robert W. Standley. Silicon crystal growth and wafer technologies. *Proceedings of the IEEE*, 100(Special Centennial Issue):1454–1474, 2012.

[French *et al.* 2020] Roger H French, Ahmad M Karimi, and Jennifer L Braid. *Electroluminescent (EL) Image Dataset of PV Module Under Step-wise Damp Heat Exposures*. osf.io/4qrtv, Oct 2020.

[Fu *et al.* 2017] Ran Fu, David Feldman, Robert Margolis, Mike Woodhouse, and Kristen Ardani. *US solar photovoltaic system cost benchmark: Q1 2017*. Technical report, EERE Publication and Product Library, Washington, DC (United States), 2017.

[Fuyuki *et al.* 2005] Takashi Fuyuki, Hayato Kondo, Tsutomu Yamazaki, Yu Takahashi, and Yukiharu Uraoka. Photographic surveying of minority carrier diffusion length in polycrystalline silicon solar cells by electroluminescence. *Applied Physics Letters*, 86(26), 06 2005.

[Fuyuki *et al.* 2007] Takashi Fuyuki, Hayato Kondo, Yasue Kaji, Akiyoshi Ogane, and Yu Takahashi. Analytic findings in the electroluminescence characterization of crystalline silicon solar cells. *Journal of Applied Physics*, 101(2), 01 2007. 023711.

[Ge *et al.* 2020] Chunpeng Ge, Zhe Liu, Liming Fang, Huading Ling, Aiping Zhang, and Changchun Yin. A hybrid fuzzy convolutional neural network based mechanism for photovoltaic cell defect detection with electroluminescence images. *IEEE Transactions on Parallel and Distributed Systems*, 32(7):1653–1664, 2020.

[Ge *et al.* 2021] Chunpeng Ge, Zhe Liu, Liming Fang, Huading Ling, Aiping Zhang, and Changchun Yin. A hybrid fuzzy convolutional neural network based mechanism for photovoltaic cell defect detection with electroluminescence images. *IEEE Transactions on Parallel and Distributed Systems*, 32(7):1653–1664, 2021.

[Goyal *et al.* 2021] Priya Goyal, Quentin Duval, Jeremy Reizenstein, Matthew Leavitt, Min Xu, Benjamin Lefaudeux, Mannat Singh, Vinicius Reis, Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Ishan Misra. *VISSL*. https://github.com/facebookresearch/vissl, 2021.

[Gupta D 2021] Gupta D. *image-segmentation-keras*. https://github.com/divamgupta/image-segmentation-keras, 2021. Accessed on 10th June 2023.

[halm. 2024] halm. *halm sun simulator*. https://www.halm.de/products/module-lab, 2024. Accessed on 10 May 2024.

[Hao *et al.* 2020] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406:302–321, 2020.

[He *et al.* 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[He *et al.* 2017] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[He *et al.* 2020a] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition, pages 9729–9738, 2020.

[He *et al.* 2020b] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735, 2020.

[Hobbs *et al.* 2018] William B. Hobbs, Babak Hamzavy, C. Birk Jones, Cara Libby, and Olga Lavrova. In-field electroluminescence imaging: Methods, comparison with indoor imaging, and observed changes in modules over one year. In *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU PVSEC)*, pages 3257–3260, 2018.

[Huang *et al.* 2022] Chao Huang, Zhiying Zhang, and Long Wang. Psopruner: Pso-based deep convolutional neural network pruning method for pv module defects classification. *IEEE Journal of Photovoltaics*, 12(6):1550–1558, 2022.

[iakubovskii P 2020] iakubovskii P. *segmentation_models.pytorch*. https://github.com/qubvel/segmentation_models.pytorch, 2020. Accessed on 10th June 2023.

[International Electrotechnical Commission 2018] International Electrotechnical Commission. *Photovoltaic devices – Part 13: Electroluminescence of photovoltaic modules*. Edition 1.0, 2018. IEC Standard No. TS 60904-13.

[Jadon 2020] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–7. IEEE, 2020.

[Jahn *et al.* 2018a] U. Jahn, M. Herz, Kontges M., and et al. *IEA PVPS T13-10 2018 Review on infrared and electroluminescence imaging for PV field applications: International Energy Agency Photovoltaic Power Systems Programme: IEA PVPS Task 13, Subtask 3.3: report IEA-PVPS T13-12:2018*. https://iea-pvps.org/wp-content/uploads/2020/01/Review_on_IR_and_EL_Imaging_for_PV_Field_Applications_by_Task_13.pdf, 2018. Accessed on 18th June 2023.

[Jahn *et al.* 2018b] Ulrike Jahn, Magnus Herz, David Parlevliet, Marco Paggi, Ioannis Tsanakas, Joshua Stein, Karl Berger, Samuli Ranta, Roger French, Mauricio Richter, et al. *Review on infrared and electroluminescence imaging for PV field applications*. International Energy Agency, 2018.

[Kamikawa Y 2020] Kamikawa Y. *tf-keras-PSPNet*. https://github.com/ykamikawa/tf-keras-PSPNet, 2020. Accessed on 10th June 2023.

[Karimi *et al.* 2019] Ahmad Maroof Karimi, Justin S Fada, Mohammad Akram Hossain, Shuying Yang, Timothy J Peshek, Jennifer L Braid, and Roger H French. Automated pipeline for photovoltaic module electroluminescence image processing and degradation feature classification. *IEEE Journal of Photovoltaics*, 9(5):1324–1335, 2019.

[Karimi *et al.* 2020] Ahmad Maroof Karimi, Justin S. Fada, Nicholas A. Parrilla, Benjamin G. Pierce, Mehmet Koyutürk, Roger H. French, and Jennifer L. Braid. Generalized and mechanistic pv module performance prediction from computer vision and machine learning on electroluminescence images. *IEEE Journal of Photovoltaics*, 10(3):878–887, 2020.

[Kawakita S 2021] Kawakita S. *multiclass-segmentation*. https://github.com/shirokawakita/multiclass-segmentation/blob/main/example_camvid_multiclassB_quita.ipynb, 2021. Accessed on 10th June 2023.

[K.G.Bedrich 2021] Y. Wang K.G.Bedrich, Y.S. Khoo. *Method, system, and image processing device for capturing and/or processing electroluminescence images, and an aerial vehicle*. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2021137764, 2021. Accessed on 18th June 2023.

[Kingma and Ba 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Koch *et al.* 2016] Simon Koch, Thomas Weber, Christian Sobottka, Andreas Fladung, Patrick Clemens, and Juliane Berghold. Outdoor electroluminescence imaging of crystalline photovoltaic modules: Comparative study between manual ground-level inspections and drone-based aerial surveys. In *32nd European Photovoltaic Solar Energy Conference and Exhibition*, pages 1736–1740, 2016.

[Köntges *et al.* 2010] M Köntges, I Kunze, S Kajari-Schröder, X Breitenmoser, and B Bjørneklett. Quantifying the risk of power loss in pv modules due to micro cracks. In *25th European Photovoltaic Solar Energy Conference, Valencia, Spain*, pages 3745–3752, 2010.

[Köntges *et al.* 2014] Marc Köntges, Sarah Kurtz, CE Packard, Ulrike Jahn, Karl A Berger, Kazuhiko Kato, Thomas Friesen, Haitao Liu, Mike Van Iseghem, John Wohlgemuth, et al. *Review of failures of photovoltaic modules*, 2014.

[Krizhevsky *et al.* 2009] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *Technical Report TR-2009*, 2009.

[Köntges *et al.* 2011] M. Köntges, I. Kunze, S. Kajari-Schröder, X. Breitenmoser, and B. Bjørneklett. The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks. *Solar Energy Materials and Solar Cells*, 95(4):1131–1137, 2011.

[LeCun and others 2015] Yann LeCun et al. Lenet-5, convolutional neural networks. *URL: http://yann. lecun. com/exdb/lenet*, 20(5):14, 2015.

[Lee and others 2013] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[Lee 2013] Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*, 07 2013.

[Li *et al.* 2019] Yongkun Li, Cunfu He, Yan Lyu, Guorong Song, and Bin Wu. Crack detection in monocrystalline silicon solar cells using air-coupled ultrasonic lamb waves. *NDT & E International*, 102:129–136, 2019.

[Lin *et al.* 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[Lin *et al.* 2016] Guosheng Lin, Chunhua Shen, Anton Van Den Hengel, and Ian Reid. Efficient piecewise training of deep structured models for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2016.

[Lydia *et al.* 2017] MD Dafny Lydia, K Sri Sindhu, and K Gugan. Analysis on solar panel crack detection using optimization techniques. *Journal of Nano-and Electronic Physics*, 9(2):2004–1, 2017.

[Mantel *et al.* 2018] Claire Mantel, Sergiu Spataru, Harsh Parikh, Dezso Sera, Gisele A. dos Reis Benatto, Nicholas Riedel, Sune Thorsteinsson, Peter B. Poulsen, and Søren Forchhammer. Correcting for perspective distortion in electroluminescence images of photovoltaic panels. In *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU PVSEC)*, pages 0433–0437, 2018.

[Mantel *et al.* 2019] Claire Mantel, Frederik Villebro, Gisele Alves dos Reis Benatto, Harsh Rajesh Parikh, Stefan Wendlandt, Kabir Hossain, Peter Poulsen, Sergiu Spataru, Dezso Sera, and Søren Forchhammer. Machine learning prediction of defect types for electroluminescence images of photovoltaic panels. In *Applications of Machine Learning*, volume 11139, page 1113904. SPIE, 2019.

[Mayr *et al.* 2019] Martin Mayr, Mathis Hoffmann, Andreas Maier, and Vincent Christlein. Weakly supervised segmentation of cracks on solar cells using normalized l p norm. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1885–1889. IEEE, 2019.

[Millendorf *et al.* 2020] Matthew Millendorf, Edward Obropta, and Nikhil Vadhavkar. Infrared solar module dataset for anomaly detection. In *Proc. Int. Conf. Learn. Represent*, 2020.

[Name 2014a] Full Author Name. *The Frobnicatable Foo Filter*, 2014. Face and Gesture submission ID 324. Supplied as additional material `fg324.pdf`.

[Name 2014b] Full Author Name. *Frobnication Tutorial*, 2014. Supplied as additional material `tr.pdf`.

[Negri and Vestri 2017] Lucas Hermann Negri and Christophe Vestri. *lucashn/peakutils: v1.1.0*. https://doi.org/10.5281/zenodo.887917, 2017.

[Niculescu-Mizil and Caruana 2005] Alexandru Niculescu-Mizil and Rich Caruana. A unified view of performance metrics: translating threshold choice into expected classification loss. *Journal of Machine Learning Research*, 6(Apr):973–1022, 2005.

[Ouali *et al.* 2020a] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020.

[Ouali *et al.* 2020b] Yassine Ouali, Celine Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[Owen-Bellini *et al.* 2020] Michael Owen-Bellini, Dana B. Sulas-Kern, Greg Perrin, Hannah North, Sergiu Spataru, and Peter Hacke. Methods for in situ electroluminescence imaging of photovoltaic modules under varying environmental conditions. *IEEE Journal of Photovoltaics*, 10(5):1254–1261, 2020.

[Paszke *et al.* 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[Perktold *et al.* 2023] Josef Perktold, Skipper Seabold, and Jonathan Taylor. *Source code for statsmodels.sandbox.stats.multicomp*. https://www.statsmodels.org/dev/_modules/statsmodels/sandbox/stats/multicomp.html, 2023. Accessed on 2023-03-07.

[Pierce, Warrick and Le Roux, Monique 2023] Pierce, Warrick and Le Roux, Monique. *Statistics of utility-scale power generation in South Africa*. Available at: https://www.csir.co.za/documents/statistics-power-sa-2022-csirpdf, 2023. Accessed: 2023-05-20.

[Pratt *et al.* 2021a] Lawrence Pratt, Devashen Govender, and Richard Klein. Defect detection and quantification in electroluminescence images of solar pv modules using U-Net semantic segmentation. *Renewable Energy*, 178:1211–1222, 2021.

[Pratt *et al.* 2021b] Lawrence E Pratt, Manjunath Basappa Ayanna, Siyasanga I May, Hlaluku W Mkasi, Elijah L Maweza, and Kittessa T Roro. Pv module reliability scorecard round 1. In *SASEC 2021*, 2021.

[Pratt *et al.* 2023] Lawrence Pratt, Jana Mattheus, and Richard Klein. A benchmark dataset for defect detection and classification in electroluminescence images of pv modules using semantic segmentation. *Systems and Soft Computing*, page 200048, 2023.

[Purushwalkam and Gupta 2020] Senthil Purushwalkam and Abhinav Gupta. Demystifying contrastive self-supervised learning: Invariances, augmentations and dataset biases. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural*

Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.

[Qian *et al.* 2017] Xiaoliang Qian, Heqing Zhang, Huanlong Zhang, Yuanyuan Wu, Zhihua Diao, Qing-E Wu, and Cunxiang Yang. Solar cell surface defects detection based on computer vision. *International Journal of Performability Engineering*, 13(7):1048, 2017.

[Qian *et al.* 2018] Xiaoliang Qian, Heqing Zhang, Cunxiang Yang, Yuanyuan Wu, Zhendong He, Qing-E Wu, and Huanlong Zhang. Micro-cracks detection of multicrystalline solar cell surface based on self-learning features and low-rank matrix recovery. *Sensor Review*, 38(3):360–368, 2018.

[Rahman and Chen 2020] Muhammad Rameez Ur Rahman and Haiyong Chen. Defects inspection in polycrystalline solar cells electroluminescence images using deep learning. *IEEE Access*, 8:40547–40558, 2020.

[Reuter *et al.* 2015] M. Reuter, L. Stoicescu, and J.H. Werner. *PV module electroluminescence: enlightening defects*. https://www.solarzentrum-stuttgart.com/uploads/file/platzhalter_vortrag_spezial_hagelschaden_DaySy_April2015_04.pdf, 2015. Accessed on 18th June 2023.

[Ritchie, Hannah and Roser, Max 2023] Ritchie, Hannah and Roser, Max. *Energy mix*. Available at: = https://ourworldindata.org/energy-mix, 2023. Accessed: 2023-05-20.

[Rizzoli 2015] Alberto Rizzoli. *13 Best Image Annotation Tools of 2023 [Reviewed]*, 2015. Accessed on 10 May 2024.

[Roberts *et al.* 2019] Graham Roberts, Simon Y Haile, Rajat Sainju, Danny J Edwards, Brian Hutchinson, and Yuanyuan Zhu. Deep learning for semantic segmentation of defects in advanced stem images of steels. *Scientific reports*, 9(1):12744, 2019.

[Ronneberger *et al.* 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[Russakovsky *et al.* 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[Seabold and Perktold 2010] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[Simonyan and Zisserman 2014] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[Sizkouhi *et al.* 2022] AM Moradi Sizkouhi, SM Esmailifar, M Aghaei, and M Karimkhani. Robopv: An integrated software package for autonomous aerial monitoring of large scale pv plants. *Energy Conversion and Management*, 254:115217, 2022.

[Sohn 2016a] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems*, 29, 2016.

[Sohn 2016b] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

[Sovetkin and Steland 2019] Evgenii Sovetkin and Ansgar Steland. Automatic processing and solar cell detection in photovoltaic electroluminescence images. *Integrated Computer-Aided Engineering*, 26(2):123–137, 2019.

[Sovetkin *et al.* 2020] Evgenii Sovetkin, Elbert Jan Achterberg, Thomas Weber, and Bart E Pieters. Encoder–decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules. *IEEE Journal of Photovoltaics*, 11(2):444–452, 2020.

[Sovetkin *et al.* 2021] Evgenii Sovetkin, Elbert Jan Achterberg, Thomas Weber, and Bart E. Pieters. Encoder–decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules. *IEEE Journal of Photovoltaics*, 11(2):444–452, 2021.

[Spataru *et al.* 2016a] Sergiu Spataru, Peter Hacke, and Dezso Sera. Automatic detection and evaluation of solar cell micro-cracks in electroluminescence images using matched filters. In *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, pages 1602–1607. IEEE, 2016.

[Spataru *et al.* 2016b] Sergiu Spataru, Peter Hacke, and Dezso Sera. Automatic detection and evaluation of solar cell micro-cracks in electroluminescence images using matched filters. In *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, pages 1602–1607. IEEE, 2016.

[Su *et al.* 2020] Binyi Su, Haiyong Chen, Peng Chen, Guibin Bian, Kun Liu, and Weipeng Liu. Deep learning-based solar-cell manufacturing defect detection with complementary attention network. *IEEE Transactions on Industrial Informatics*, 17(6):4084–4095, 2020.

[su binyi 2022] su binyi. *Photovoltaic cell anomaly detection*. https://kaggle.com/competitions/pvelad, 2022. Accessed on 2023-02-18.

[Sun *et al.* 2018] Mingjian Sun, Shengmiao Lv, Xue Zhao, Ruya Li, Wenhan Zhang, and Xiao Zhang. Defect detection of photovoltaic modules based on convolutional neural network. In *Machine Learning and Intelligent Communications: Second International Conference, MLICOM 2017, Weihai, China, August 5-6, 2017, Proceedings, Part I 2*, pages 122–132. Springer, 2018.

[Szegedy *et al.* 2015] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[Tang *et al.* 2019] Wuqin Tang, Qiang Yang, and Wenjun Yan. Deep learning based model for defect detection of mono-crystalline-si solar pv module cells in electroluminescence images using data augmentation. In *2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1–5. IEEE, 2019.

[Tang *et al.* 2020] Wuqin Tang, Qiang Yang, Kuixiang Xiong, and Wenjun Yan. Deep learning based automatic defect identification of photovoltaic module using electroluminescence images. *Solar Energy*, 201:453–460, 2020.

[Tang *et al.* 2022] Wuqin Tang, Qiang Yang, and Wenjun Yan. Deep learning-based algorithm for multi-type defects detection in solar cells with aerial el images for photovoltaic plants. *CMES-Computer Modeling in Engineering & Sciences*, 130(3), 2022.

[Thisanke *et al.* 2023] Hans Thisanke, Chamli Deshan, Kavindu Chamith, Sachith Seneviratne, Rajith Vidanaarachchi, and Damayanthi Herath. Semantic segmentation using vision transformers: A survey. *Engineering Applications of Artificial Intelligence*, 126:106669, 2023.

[Tomar N 2022] Tomar N. *Semantic-Segmentation-Architecture*. https://github.com/nikhilroxtomar/Semantic-Segmentation-Architecture, 2022. Accessed on 10th June 2023.

[TorchVision maintainers and contributors 2016] TorchVision maintainers and contributors. *TorchVision: PyTorch's Computer Vision library*. https://github.com/pytorch/vision, 2016.

[Torpey *et al.* 2024] David Torpey, Lawrence Pratt, and Richard Klein. A large-scale evaluation of pretraining paradigms for the detection of defects in electroluminescence solar cell images. *under review by the Journal of Electronic Imaging*, 2024.

[Tsai and Luo 2010] Du-Ming Tsai and Jie-Yu Luo. Mean shift-based defect detection in multicrystalline solar wafer surfaces. *IEEE Transactions on Industrial Informatics*, 7(1):125–135, 2010.

[Tsai *et al.* 2012] Du-Ming Tsai, Shih-Chieh Wu, and Wei-Chen Li. Defect detection of solar cells in electroluminescence images using fourier image reconstruction. *Solar Energy Materials and Solar Cells*, 99:250–262, 2012.

[Ulku and Akagunduz 2019] I Ulku and E Akagunduz. A survey on deep learning-based architectures for semantic segmentation on 2d images. *Journal of Applied Artificial Intelligence*, 2019.

[United Nations 2023] United Nations. *What is renewable energy*. Available at: https://www.un.org/en/climatechange/what-is-renewable-energy, 2023. Accessed: 2023-05-20.

[Wang *et al.* 2022a] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022.

[Wang *et al.* 2022b] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[Wu *et al.* 2018] Jiawei Wu, Eric Chan, Raginee Yadav, Hamsini Gopalakrishna, and GovindaSamy TamizhMani. Durability evaluation of pv modules using image processing tools. In *New Concepts in Solar and Thermal Radiation Conversion and Reliability*, volume 10759, pages 141–156. SPIE, 2018.

[Xie *et al.* 2011] WT Xie, YJ Dai, RZ Wang, and K Sumathy. Concentrated solar energy applications using fresnel lenses: A review. *Renewable and Sustainable Energy Reviews*, 15(6):2588–2606, 2011.

[Yeung *et al.* 2022] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95:102026, 2022.

[You *et al.* 2017] Yang You, Igor Gitman, and Boris Ginsburg. *Large Batch Training of Convolutional Networks*, 2017.

[Yuan *et al.* 2020] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 173–190. Springer, 2020.

[Zandamela *et al.* 2023] Frank Zandamela, Lawrence Pratt, Siyasanga May, Wisani Mkasi1, and Thabang Mabeo. Towards improved solar PV module characterisation: Correlating electroluminescence image defects with i-v curve characteristics using a semantic segmentation-based multi-defect detection algorithm. *Renewable Energy*, 2023.

[Zhang *et al.* 2020a] Kaige Zhang, Yingtao Zhang, and Heng-Da Cheng. Crackgan: Pavement crack detection using partially accurate ground truths based on generative adversarial learning. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1306–1319, 2020.

[Zhang *et al.* 2020b] Xiong Zhang, Yawen Hao, Hong Shangguan, Pengcheng Zhang, and Anhong Wang. Detection of surface defects on solar cells by fusing multi-channel convolution neural networks. *Infrared Physics & Technology*, 108:103334, 2020.

[Zhao *et al.* 2017] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[Zhao *et al.* 2021] Yang Zhao, Ke Zhan, Zhen Wang, and Wenzhong Shen. Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line. *Progress in Photovoltaics: Research and Applications*, 29(4):471–484, 2021.

[Zhao *et al.* 2023] Xiaolong Zhao, Chonghui Song, Haifeng Zhang, Xianrui Sun, and Jing Zhao. Hrnet-based automatic identification of photovoltaic module defects using electroluminescence images. *Energy*, page 126605, 2023.