

https://dspace.org/

Deep Learning for HighSpeed Lightning FootageA Semantic Segmentation Network Comparison

Tyson Cross; Jason Smit; Carina Schumann; T Warner; Hugh Hunt

MDPI AG
https://hdl.handle.net/10539/40442
WOS

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

atmosphere

MDPI

Article

Deep Learning for High-Speed Lightning Footage—A Semantic Segmentation Network Comparison

Tyson Cross [1], Jason R. Smit [1], Carina Schumann [1], Tom A. Warner [2] and Hugh G. P. Hunt [1,*]

[1] The Johannesburg Lightning Research Laboratory, School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2000, South Africa
[2] ZTResearch, Rapid City, SD 57701, USA
* Correspondence: hugh.hunt@wits.ac.za

Abstract: We present a novel deep learning approach to a unique image processing application: high-speed (>1000 fps) video footage of lightning. High-speed cameras enable us to observe lightning with microsecond resolution, characterizing key processes previously analyzed manually. We evaluate different semantic segmentation networks (DeepLab3+, SegNet, FCN8s, U-Net, and AlexNet) and provide a detailed explanation of the image processing methods for this unique imagery. Our system architecture includes an input image processing stage, a segmentation network stage, and a sequence classification stage. The ground-truth data consists of high-speed videos of lightning filmed in South Africa, totaling 48,381 labeled frames. DeepLab3+ performed the best (93–95% accuracy), followed by SegNet (92–95% accuracy) and FCN8s (89–90% accuracy). AlexNet and U-Net achieved below 80% accuracy. Full sequence classification was 48.1% and stroke classification was 74.1%, due to the linear dependence on the segmentation. We recommend utilizing exposure metadata to improve noise misclassifications and extending CNNs to use tapped gates with temporal memory. This work introduces a novel deep learning application to lightning imagery and is one of the first studies on high-speed video footage using deep learning.

Keywords: DeepLabV3+; fully convolutional networks; high-speed video footage; lightning; semantic segmentation; SegNet

check for updates

Citation: Cross, T.; Smit, J.R.; Schumann, C.; Warner, T.A.; Hunt, H.G.P. Deep Learning for High-Speed Lightning Footage—A Semantic Segmentation Network Comparison. *Atmosphere* **2024**, *15*, 873. https://doi.org/10.3390/atmos15080873

Academic Editor: Olaf Scholten

Received: 28 May 2024
Revised: 4 July 2024
Accepted: 17 July 2024
Published: 23 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Modern high-speed cameras, with frame rates in excess of 1000 fps, allow us to observe the phenonmenon of lightning as it has never been seen before (Figure 1). Before the arrival of these cameras, the physical processes involved in the lightning phenonmenon could only be understood and estimated through remote electromagnetic field measurements and rare, specific occasions of direct current measurement [1–3]. While the photography of lightning has been ongoing for several years, it is really the work of Saba, Warner and Ballaroti that pioneered the study of lightning through high-speed camera observations, characterizing not only cloud-to-ground and cloud-to-cloud lightning, but other distinctions such as the direction of leader propagation (upward or downward lightning), polarity (a clear, visual distinction can be made between positive lightning and negative lightning) and cases of multiple ground contact points, among other key physical observations [4–12]. Until now, such analyses have only been done manually, requiring hours of painstaking work on the part of the researcher to carefully watch each frame of a single video-lightning events typically occur within 500 ms, meaning at least 500 frames of careful analysis-and record the visual characteristics seen in the frame. It is clearly work in need of automation but still requires a large degree of knowledge about what patterns indicate which physical processes. The main contributions of this paper are: the use of machine learning in performing the task of characterising lightning events from high-speed video footage by comparing a number of semantic segementation techniques (Deeplabv3+ [13], SegNet [14],

FCN8s [15], U-net [16] and AlexNet [17]). This is not only a unique contribution to the field of lightning research, but provides a novel approach to image processing, being one of the first attempts at applying pattern recognition and machine learning techniques to high frame rate video footage.

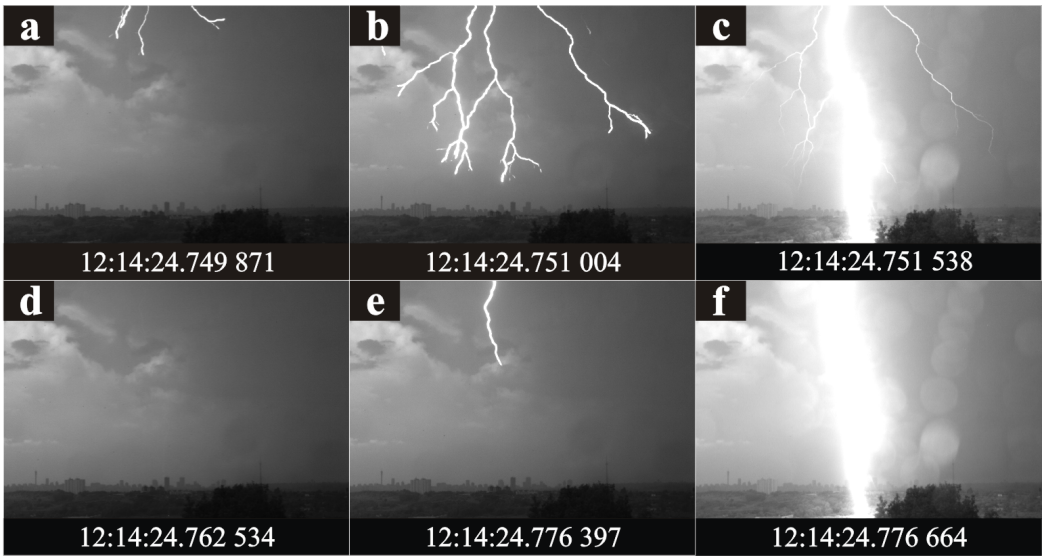


Figure 1. High-speed image of downward negative lightning. (**a**–**f**) show different stages of the progression of a downward negative leader.

2. Background

In the 1880's, the photographer William Jennings captured the first ever photograph of a lightning flash. The first 'high-speed' footage was actually captured in 1926 by Sir Basil Schonland using a Boys camera and Schonland was able to see the phenomenon known as strokes (repeated current impulses in a lightning channel which give rise to the 'flickering' effect one notices if viewing lightning with the naked eye) [18–20]. Subsequent photographic studies of lightning involving low frame rate cameras have been carried out from the 1930s through to the 1990s, with early 'higher'-speed studies being conducted at 200 frames per second [21–25]. It was not until the advent of truly high-speed (more than 1000 frames per second) cameras, and the pioneering work of Saba, Warner and Ballaroti, did such studies provide genuine contributions to the field [4–11]. Since then, high-speed cameras have become an essential tool in studying lightning [26–29].

Figure 1 shows six frames from a high-speed video capture of a lightning flash, captured on the 1 February 2017 at 12:42:14 UTC in Johannesburg, South Africa [30–32]. This shows a downward negative lightning event and some of the characteristics we can learn from high-speed footage of lightning. Firstly, in (a-b), the propagation of the leader is visible. A leader is a self-propagating electron avalanche ('spark') which generates its own electric field and is one of the key mechanisms behind lightning, particularly cloud-to-ground lightning. It is only visible through high-speed footage (notice the time difference between frame (a) and (b) is 33 microseconds). The continuous development of the branching seen in frame (b) is typical of negative propagation. In frame (c), we have attachment to ground and this is known as the return stroke and is what we see with the naked eye [4,6,7,33,34]. This is where large impulsive current flows (peaks in the kiloampere range) leading to the brightness.

In frame (d), we see that the luminosity of the channel has died down after approximately 10 ms but then another leader following the same path appears 10 ms later in (e)—this is a dart leader and it follows the original channel resulting in a subsequent return stroke seen in frame (f)—another impulsive current that flows to the ground. On average, lightning events such as this have 2–3 subsequent strokes but can have as many as 25. Other characteristics such as recoil leaders associated with positive events, upward

leader propagation from tall towers, continuing current durations (lower peak currents but not impulsive that follow return strokes), ICC pulses and M-components as well as the structure of lightning channels can also be seen.

Machine learning has seen major improvements in accuracy, number of learnable variables and speed as the network topologies improve [35]. Machine learning is applied in several fields of study with much success in image resolution [36–39], however most of the prior work is focused on slow speed (less than 1000 frames per second) footage. Machine learning has been applied in the field of lightning however these applications have been focused on lightning prediction and weather pattern effects [40–51]. Smit et al. [52] proposed an initial approach to lightning image segmentation where they found that high-speed footage could not be processed in large amounts due to the memory requirements and it was important to perform segmentation on individual images. However, other than this initial investigation, no work has been done on applying machine learning and image segmentation techniques to lightning footage, or high frame rate video footage for that matter.

3. Approach Taken

The chosen approach segments input videos of lightning activity into separate events, both spatially and temporally, and then counts important metrics within each event. An overview is shown in Figure 2 (The implementation can be found online at https://github.com/TysonCross/Project-Raiden).

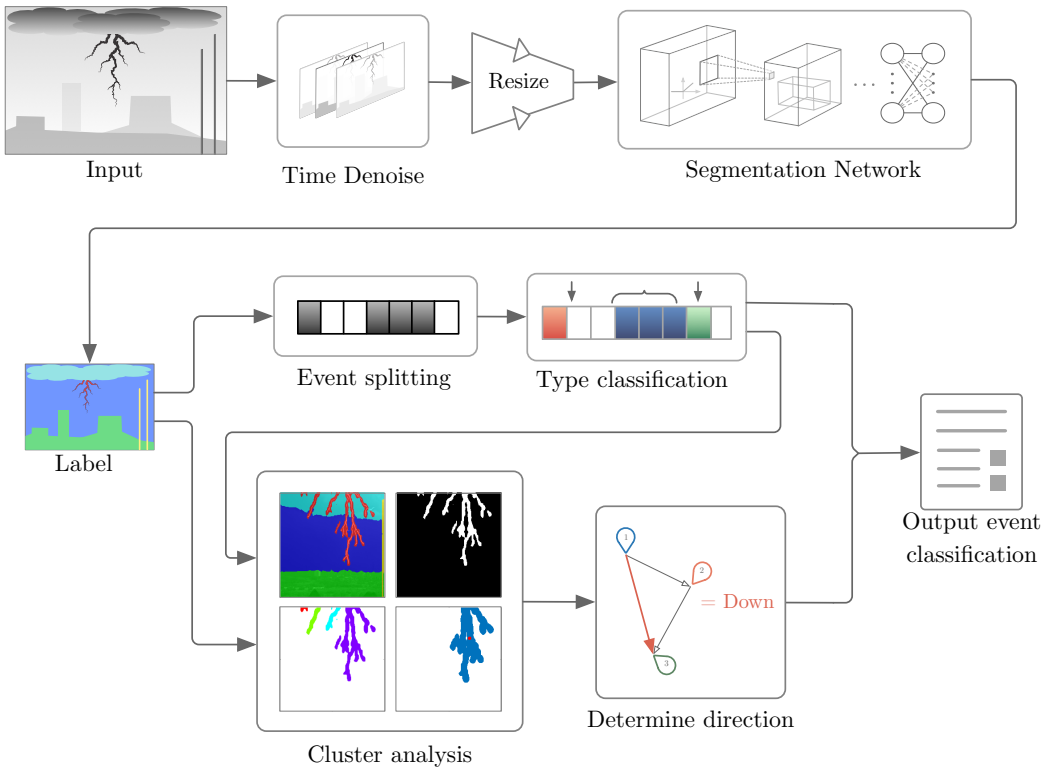


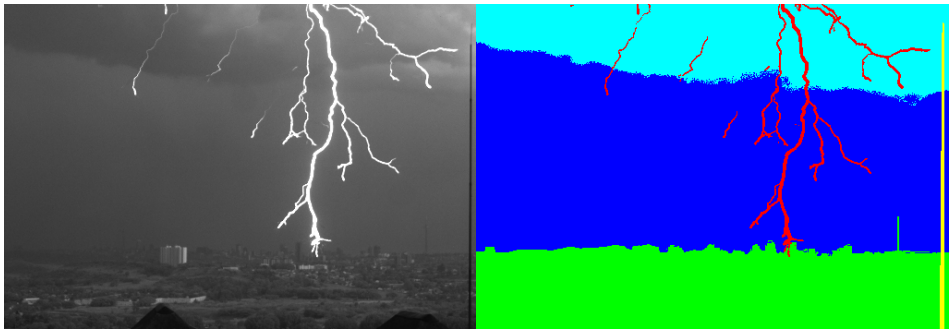
Figure 2. Simplistic representation of the system sub-components.

Input image sequences are first de-noised (over the time series frames), then resized to fit a Convolutional Neural Network (CNN), which produces a labelled output for each frame. Sequences of labelled frames are then classified in time as separate lightning 'events'. An event is defined by a sequence of frames showing spatially coherent, visible lightning activity, and ends with frames where no visible lightning is present. Each 'event' is classified as either an 'attempted leader' (leader activity without attachment) or an attachment event (where a return stroke occurs). The time of the initial return stroke, the

number of subsequent strokes, stroke duration, total event duration, and direction of leader propagation (downward or upward) are calculated.

3.1. Input Image Processing

Rather than manually labeling individual frames, a node-based procedural compositing software generates image 'mattes' or 'masks' for each event sequence, exploiting the static nature of most scene elements. Background, clouds, ground, and other elements in a frame are labeled on a single frame and then extended to the entire sequence. Per-frame luminance-based mattes isolate the bright lightning against the dimmer background. Figure 3 shows an example of an input frame and the corresponding labeled output matte. Details of this process are discussed in Section 4.



(**a**) Input image (**b**) Output Matte

Figure 3. (**a**) Input image and (**b**) corresponding labeled 'Matte' or 'Mask'.

3.2. Semantic Segmentation Network

A transfer learning approach was applied, using existing performant deep learning networks for new image classification problems in specialized domains without retraining from scratch [53]. A new 6-label feature set specific to the input footage was defined, discussed further in Section 4. Semantic segmentation is performed on single-frame input images and on the 'masks' rather than the pixel images.

Choice of Networks

The deep learning CNNs selected and their capacities are shown in Table 1. The networks are Deeplabv3+ [13], SegNet [14], FCN8s [15], U-net [16], and AlexNet [17]. Key aspects of each network are discussed below.

Table 1. Network size comparison.

Network	Layers	Input Size	Total Learnables	Size
deeplabv3+	101	256 × 256	20,611,166	61.6 MB
alexnet	27	227 × 227	57,040,269	121 MB
fcn8s	51	227 × 227	134,300,132	501.2 MB
segnet	91	256 × 256	29,446,482	110.7 MB
u-net	58	256 × 256	31,032,068	116 MB

- **DeepLabv3+** was developed by Chen et al., implementing several innovations and refinements on previous DeepLab models [13]. It follows an encoder/decoder CNN topology, with depthwise-separable convolutional operations to extract features, and incorporates novel features such as Atrous Spatial Pyramid Pooling, which expands the retinal sensitivity of the network with fewer convolutional operations and preserves feature edges. Residual connections from early links in the network help recover spatial information during the decoder portion. A simplified diagram is shown in Figure 4.

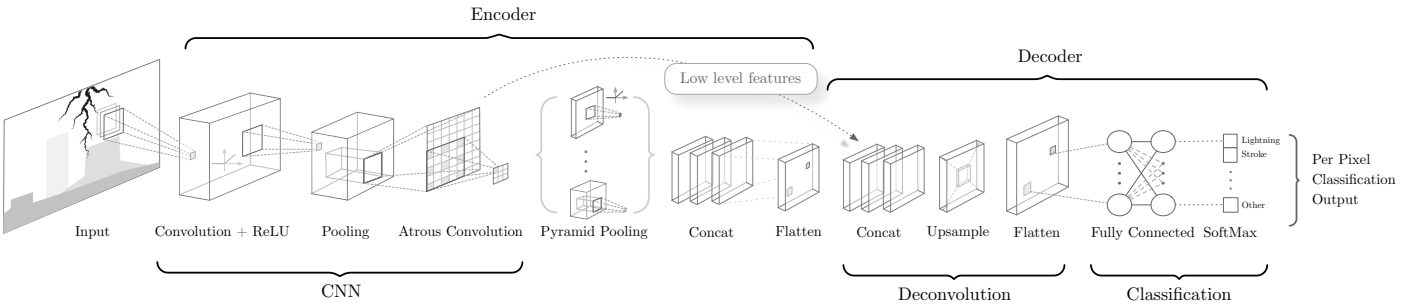


Figure 4. Simplified DeepLabv3+ schematic, adapted from [13].

- **SegNet** was developed by Badrinarayanan et al. at the University of Cambridge primarily for road scene segmentation [14]. It follows an encoder/decoder CNN architecture with pixel-wise classification, incorporating pooling indices in the decoder portions when upsampling, and a matched hierarchy of decoders for each encoder.
- **FCN8s** was selected for its superior retention of spatial information and edge recovery. Shelhamer et al. replaced fully-connected layers with 1×1 convolutional layers and upsampling the features revealed by the initial encoding layers, achieving accurate pixel-wise classification [15]. They demonstrated that 'fusing' with skipped layers produced superior spatial results and an increase in training speed.
- **U-net** was developed by Ronneberger et al. for biomedical image applications [16]. It produces accurate segmentation results with fewer input images, using upsampling and successive convolutional layers to preserve initial features. The shape of the symmetric feature-encoding and decoding portions of the network gave rise to the network name.
- **AlexNet** was created by Krizhevsky et al. in 2012 and is considered a hugely influential network in computer vision. The CNN incorporates multiple stacked convolutional operations and fully-connected layers, followed by ReLUs [17]. It includes overlapping max-pooling layers to reduce accuracy errors and dropout layers to reduce overfitting.

3.3. Sequence Segmentation

Smit et al. developed a technique to split and classify image sequences based on the network output of image labels [52]. Each frame is examined for the labels contained in the image. If pixels labeled *stroke* are present in any frames, the event is considered an attachment event otherwise it is considered an attempted leader. A frame can only be labelled *stroke* if the frame itself saturates (Figure 7d is a good example). The number of strokes present in an event and the duration of each stroke are counted based on the presence of labeled pixels above a threshold percentage (0.2%) of the total frame pixel count, to mitigate false positives. The initial portion of the event frames before a stroke (or the first half of the event in the case of an attempted leader) are marked as candidate frames for analyzing the direction of the lightning.

Direction Classification

The candidate label frames for directional classification are analyzed using a weighted clustering method. All pixels labeled *stroke* are converted to a sparse 2D point cloud. A region of interest is defined using a bounding box. These points are then clustered using the DBSCAN algorithm to exclude outlier clusters and group larger, connected lines of lightning [54]. The centers of the largest clusters are found, and the overall center is calculated, weighted by the size of the cluster, as shown in Figure 5. This was not applied to intra-cloud lightning events, as defining the direction for these is more complex.



(**a**) Label (**b**) Clusters (**c**) Resultant vector

Figure 5. Direction analysis.

4. Data and Image Processing

4.1. High-Speed Footage

The high-speed lightning footage is from a study of cloud-to-ground lightning events over Johannesburg, South Africa, conducted by Schumann et al. [30,55]. The footage was captured with a Vision Phantom camera (model v310) from February 2017 to December 2018. The greyscale footage ranges from 7000 to 23,000 frames per second (fps), predominantly from a fixed location offering a raised vantage point with good views of urban and peri-urban lightning activity. Cameras used various exposure settings and positions, recording footage at 512 × 256 or 512 × 352 resolution. The high contrast between the bright lightning and dark background allows for easy mask separation using luminance-based keying operations. However, the dataset exhibits high variance in pixel intensities, both within sequences and across frames, complicating normalization and making it difficult to establish a baseline 'average' image [56]. The footage contains significant noise from rain, low-light exposure settings and sensor noise. Powerful discharges during attachment events resulted in multiple frames being clipped to pure white, while frames without lightning activity occasionally had pixels close to pure black.

4.2. Image Processing

Digital compositing software has revolutionized post-production and visual effects in film and television, with Foundry's Nuke being ubiquitous in these industries [57–59]. The software's maturity and widespread use make sophisticated image processing operations highly accessible and optimized for large image sequences in high resolution. Figure 6 shows the basic image operations used in Nuke to create the sequence of masks.

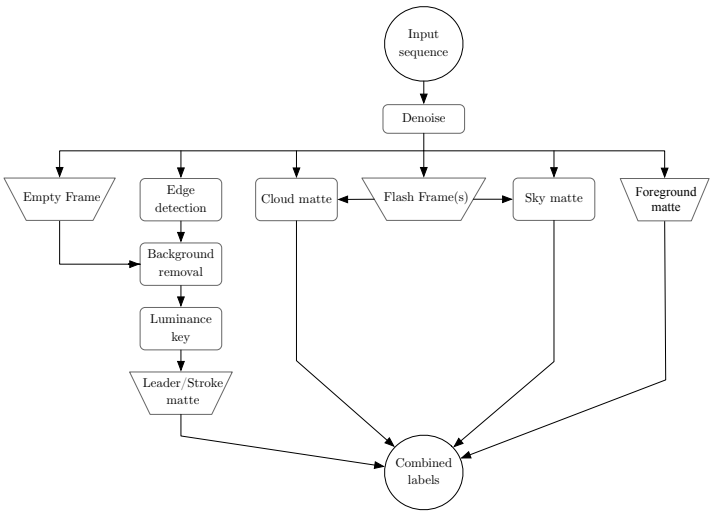


Figure 6. Image processing graph.

4.2.1. Input Sequence

Video footage in Vision Research's proprietary CINE format was imported into Black Magic's DaVinci Resolve editing software [60–62]. The footage was roughly cut into sequences excluding 'empty' frames (containing no lightning events) and exported as .tif image sequences. Per-sequence compositing scripts were created in Nuke, utilizing its image processing operations. The procedural nature of the software, which uses a directed acyclic graph (DAG), allowed easy adaptation of scripts to different input sequences with minimal adjustments.

4.2.2. Noise Reduction

Most video footage was extremely noisy due to rain and camera sensor noise. A denoising operation, creating a noise profile through sampling small sections of the footage, reduced digital sensor noise. Optical flow analysis converted the resulting vectorized color channels to specific 'motion' channels, allowing further noise reduction using predominantly static sky and cloud elements. The image was then slightly blurred with a Gaussian blur operation, masked by a luminance mask to restrict the filter to areas without lightning. The edges of the lightning were preserved without artifacts using a merge/dissolve node to blend the lightning edges back into the main composite.

4.2.3. Edge Detection

A Laplacian edge detect operation produced a sharp, precise white shape matching the lightning but added fine detail noise to the remainder of the image. This operation was significantly sharper than equivalent Sobel or Prewitt edge detection filters.

4.2.4. Background Removal

A frame hold operation fixed the output image on a frame without lightning, which was subtracted from the main composite, removing bright city lights, illuminated edges of buildings, background clouds, or bright sky. The output composite consisted of lightning or stroke elements and noise introduced by the Laplacian operation.

4.2.5. Radial Mask

A large, soft radial mask was manually positioned over the approximate area where lightning occurs. This mask, blurred to further soften the area of lightning, was used to exclude any areas not under the radial shape. This process, subject to human error and judgment, is specific to the exact lightning event recorded.

4.2.6. Luminance Keying

The input images underwent luminance keying to restrict visible pixels to a specific intensity range. Two keys were applied: a narrow key for the brightest pixels, capturing the main center of the leaders, and a broader key for the dimmer, softer lightning branches. The image was graded to lower the 'black level', offsetting the darkest pixel luminance values and shifting the overall luminance range. The alpha channel was then thresholded in a binary alpha operation, producing a pixel value of 1 where there was lightning and 0 elsewhere.

4.2.7. Rotoscoping

For non-lightning image elements such as *cloud*, *sky*, and *ground*, a single frame mask was painted or drawn with a bezier-curve rotoscoping tool and extended throughout the sequence. Masks were reused across sequences with minor adjustments to match movement between events in the high-speed footage. These masks were cached to disk and read back as still images, speeding up rendering time without re-evaluating the rotoscope shapes on every frame.

4.2.8. Manual Keyframing

Manual keyframing was restricted to the shape and position of the radial mask to exclude noise introduced by the Laplacian operation in areas away from the lightning and to manually switch between *stroke* and *leader* frames. The keyframe interpolation for lightning type was set to a stepped/constant setting, holding the type of lightning (*stroke/leader*) until a change occurred. A decision was made based on pixel intensity dimming during continuous current on what constituted the end of a *stroke* and an established channel.

4.2.9. Image Labels

Figure 7a–i show selected frames from a sequence of combined masks. The labels are: *leader*, the bright branching edges of the unattached lightning and the established channel of brightly lit continuing current; *stroke*, the extremely bright, often blown-out areas caused by initial or subsequent return strokes; *sky*, flat areas without cloud cover; *ground*, including the city skyline; *cloud*, dark areas of cloud cover silhouetted during a brightly lit 'flash-frame'; and *other*, a label to mask out tall lightning rods visible in several sequences. These masks were created using keys and rotoscoping, with some manual painting if a clean matte could not be pulled. A threshold was applied to set the value of non-zero pixels to the maximum intensity of 1, producing binary masks. This process resulted in 111 per-pixel labeled image sequences, representing the 'ground-truth' for the associated video footage, totaling 48,381 frames of labeled data.

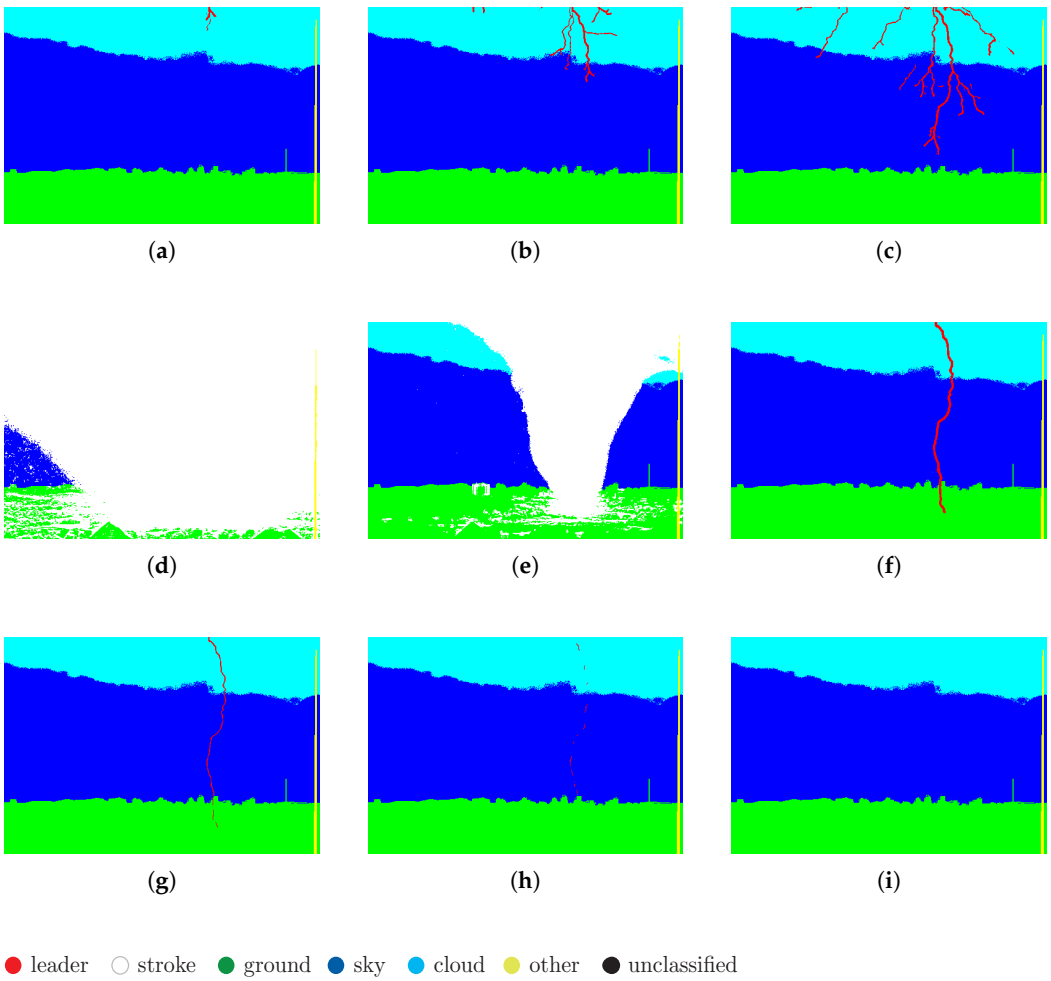


Figure 7. (**a**–**i**) Example frame sequence of binary masked labels for an attachment event based on the original video shown in Figure 3.

5. Performance Evaluation

5.1. Data Partitioning

The dataset was divided into training, validation, and test sets with a 70/15/15 percentage scheme, respectively. This scheme allowed a large number of test sequences for evaluation while providing sufficient validation data during training. The test set was created by randomly selecting entire sequences whose frame counts collectively represented 15% of the total labeled data. Training and validation images were then randomly shuffled to mitigate over-fitting. Important events such as the initial establishment of a leader and discharges during attachment events are less common than the lengthy periods of an established channel. This temporal variance was compensated for by discarding a percentage of frames without *stroke* and *leader* labels. All stroke frames and most leader frames were included in the training set. To supplement the dataset, input images were augmented with random horizontal and vertical translations, and versions of the input data were generated with exposure variations of ± 2 f-stops of brightness (the luminance was multiplied by $2^{\pm 2}$).

Despite these techniques, the pixel counts of individual labels were not uniformly represented in the training set, as shown in Table 2. To compensate, the training used the normalized inverse frequency of the label incidence for weighting, ensuring sufficient sensitivity to important features such as strokes and leaders.

Table 2. Per-label pixel label count. Tables should be placed in the main text near to the first time they are cited.

Feature	PixelCount	ImagePixelCount
leader	2.3662×10^6	1.2495×10^9
stroke	2.8696×10^6	4.6858×10^7
sky	7.2953×10^8	1.6890×10^9
cloud	7.8179×10^8	1.4996×10^9
ground	1.6991×10^8	1.6894×10^9
other	2.9131×10^6	3.6890×10^8

5.2. Training

The networks were trained on a workstation with a single GPU with a CUDA compute capability of 7.5 and 8 GB of NVRAM. 16 GB of system memory was supplemented with a 64 GB swap partition on an internal SSD, allowing larger batch sizes at the cost of slower training. The networks trained until a validation criterion was met (a ValidationPatience of 4) or the maximum number of epochs was completed. The outermost layers for each network were replaced with input and output layers for the required data sizes and features. The results incorporate the best-tested hyper-parameters and highest accuracy from training.

5.2.1. Training Hyperparameters

The networks were trained using stochastic gradient descent with momentum loss minimization (SGDM) with a momentum of 0.9. This produced consistent and comparable results across varying batch sizes and network topologies compared to adaptive loss functions such as ADAM or RMSProp. The batch size for each network was set as large as possible. Initial learning rates were set to 0.001 with a scheduled reduction varying per network. These parameters were chosen through iterative comparison and evaluation on the test sets. All selected networks were pre-trained using ImageNet [63].

5.2.2. Training Results

Table 3 shows the training results for the five different networks with some variation in parameters. The best-performing network was DeepLabv3+ trained on a smaller set of images. However, it was concluded to be overfitting and did not perform well on unseen

data, likely due to the similarity of test data to training data. Additionally, this network performed poorly on dark input images, which were underrepresented in the training set. Once a wider set of sequences was labeled, increasing the range of test data candidates, the reported accuracy decreased to 93.2%, but stroke detection Intersection over Union (IoU) increased to 53.7%. The network trained with a batch size of 100 and 20,982 images was more sensitive to segmenting lightning and strokes in dim footage and performed better in the directional classifier. However, this sensitivity also introduced false detections of lightning on vertical poles and high-intensity pixels on cloud edges. Figure 8 shows the training loss graph for the DeepLabv3+ network.

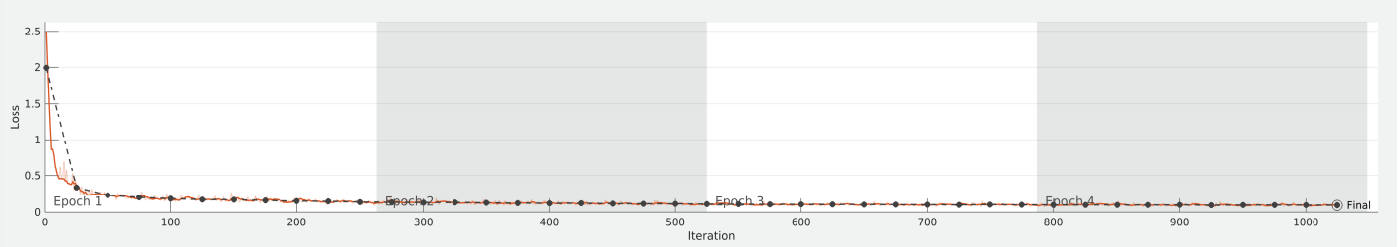


Figure 8. Training loss graph for selected DeepLabv3+.

Table 3. Networks training performance.

Network	Accuracy (%)			IoU (%)		Training Time [min]	Learning Rate		Minibatch	Epochs	Number of Images		
	Validation	Global	Mean	*Leader*	*Stroke*		Initial	End			Training	Validation	Test
Alexnet	83.6	83.9	55	1	<1	479	0.001	0.000125	120	17	20,801	5200	5429
Alexnet	89	90.4	74.1	2.5	66	191	0.001	0.005	120	7	20,801	5200	4718
DeepLabv3+	95.2	95.7	88.3	13	41	106	0.01	0.01	50	4	11,512	2878	2659
DeepLabv3+	93.7	89.9	87.5	9	51	320	0.001	0.0005	120	7	20,622	5156	5422
DeepLabv3+	**93.2**	**92.1**	**83.9**	**7.9**	**53.7**	**322**	**0.001**	**0.0005**	**100**	**4**	**20,982**	**5245**	**4511**
FCN8s	89.5	79.1	82.8	3.7	9.2	838	0.01	0.01	10	2	11,628	2907	2451
FCN8s	90.7	90.6	65.2	7	23	869	0.001	0.001	10	1	20,697	5174	4861
SegNet	95.5	94	76.3	11.9	0	966	0.01	0.01	10	2	11,512	2878	2659
SegNet	92.1	88.6	60.4	5	0	1159	0.01	0.01	10	1	20,806	5201	5429
U-net	52.6	54.3	29	9.2	<1	733	0.001	0.001	10	1	20,806	5201	4718
U-net	70.7	69.5	51.2	4.2	<1	1262	0.001	0.001	10	1	20,959	5240	4540

5.3. Network Evaluation

The following metrics were used for evaluation: Validation Accuracy, Average Accuracy (mean of all label accuracy), Global Accuracy (ratio of correctly classified pixels to total pixels), IoU for *stroke* and *leader* labels in the test set, and BFscore (BFS), a contour matching score measuring Boundary F1 matching between predicted and actual labels. Figure 9 shows the performance of the different semantic segmentation networks. The best-performing network was DeepLabv3+ trained with 20,982 training images and a batch size of 100. SegNet and U-Net were unable to detect *strokes* despite good accuracy for *leaders*.

Table 4 shows the results of evaluating the output event classifications against the unseen test data sequences with DeepLabv3+ as the segmentation network, obtained by comparing the classified output with the actual sequences using manual observance of the imagery. Successful direction classification for the 27 test sequences was less than 50%, although correct stroke frame detection was 74.1%. The system was only able to correctly segment 51.9% of the separate events in the test sequences.

Table 4. System Accuracy with Deeplab3+ set as segmentation network.

Event					Stroke	
Num	**Start**	**Duration**	**Type**	**Direction**	**Start**	**Duration**
51.9%	48.1%	48.1%	81.5%	80.0%	74.1%	66.7%

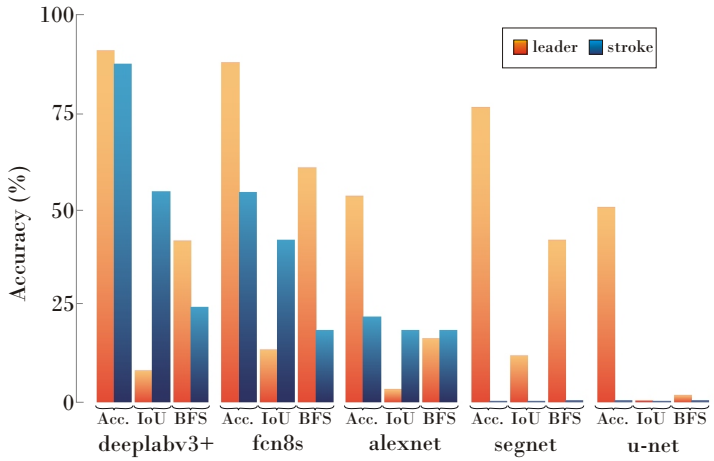


Figure 9. CNN lightning labels accuracy.

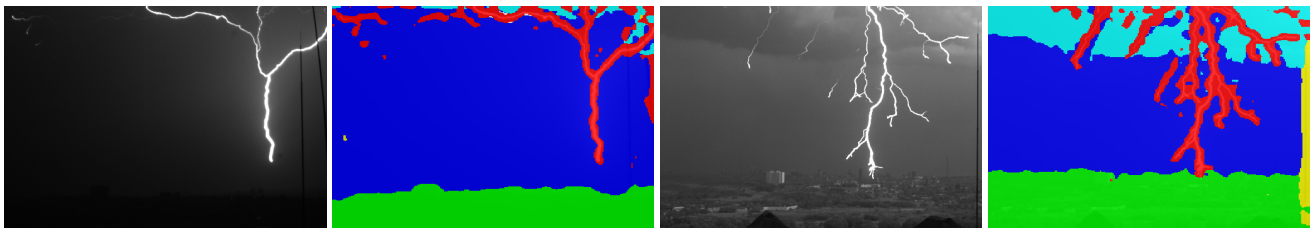
6. Discussion

6.1. Training Observations

Early training showed excellent accuracy percentages, but the limited size of the input data caused overfitting, leading to poor performance on unseen footage. Doubling the labeled input data reduced reported accuracy but improved directional and event segmentation performance. Networks with larger batch sizes consistently performed better. SegNet required a higher initial learning rate (0.01) for accurate training. The Training/Validation/Test partitioning was adjusted from 80/10/10 to 70/15/15 to provide a larger test set and avoid overfitting. Early tests with relaxed validation patience criteria were tightened to 4–6 to ensure early stopping and avoid overfitting. Networks with additional minibatch normalization nodes, such as DeepLabv3+, trained well on the input data.

6.2. Network Performance-Output Labels

Some typical input frames and their results are shown in Figure 10. These images demonstrate the structural similarity in the composition of elements across the input data, such as the flat ground horizon line in the lower third of the image. This repeated arrangement restricts the generality of the trained network, suggesting the system may not perform well on footage from different locations or with different foreground shapes. The network's failure to capture detail in the skyline of the horizon is likely due to image distortion when resized to the network's input size (512 × 352 resized to 256 × 256) and the loss of spatial information inherent in the convolutional technique of DeepLabv3+.

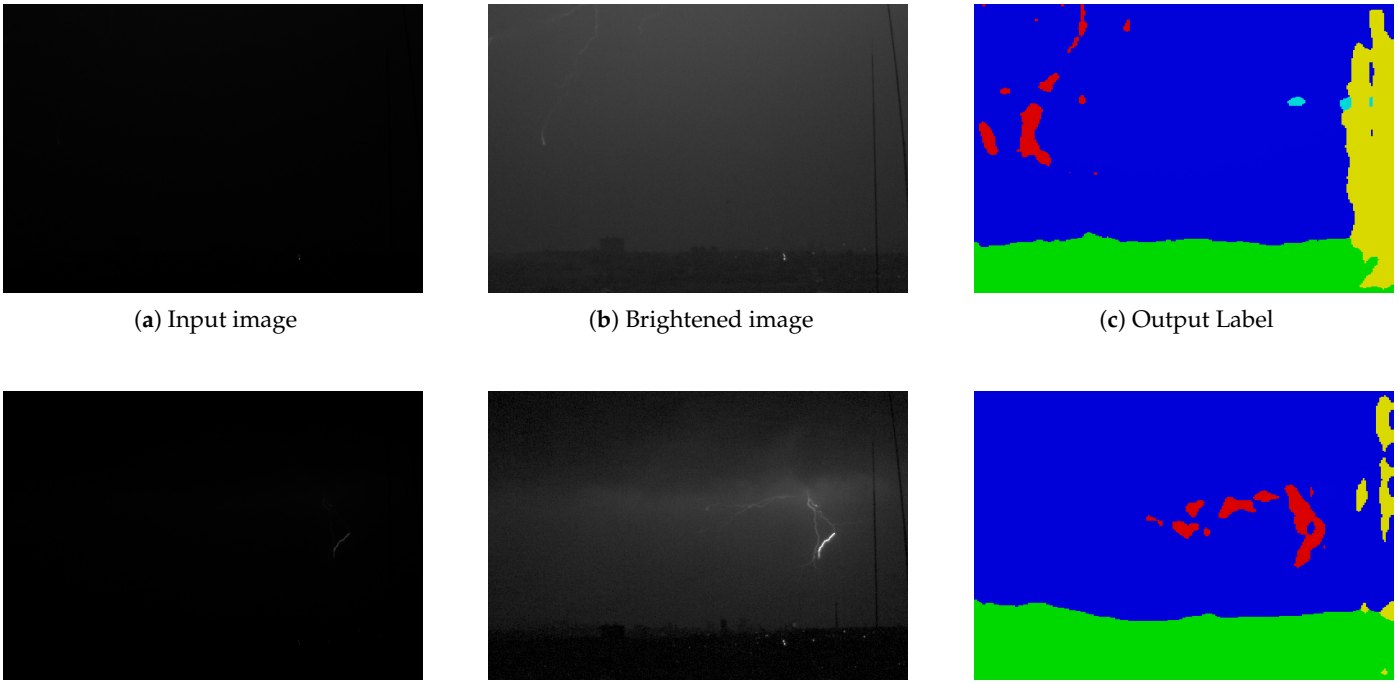


(**a**) Input image (**b**) Output Label (**c**) Input image (**d**) Output Label

Figure 10. Example result images.

6.3. Segmentation of Dark Footage

Figure 11 shows successful examples of semantic segmentation on very dark frames. Although the input frames and lightning leaders are dim, the system can label the presence of lightning in footage that might be missed by humans. However, this heightened sensitivity also introduces false positives, as clusters of noise are labeled as *leader*, reducing event classification accuracy if they occur in candidate frames for direction-finding.



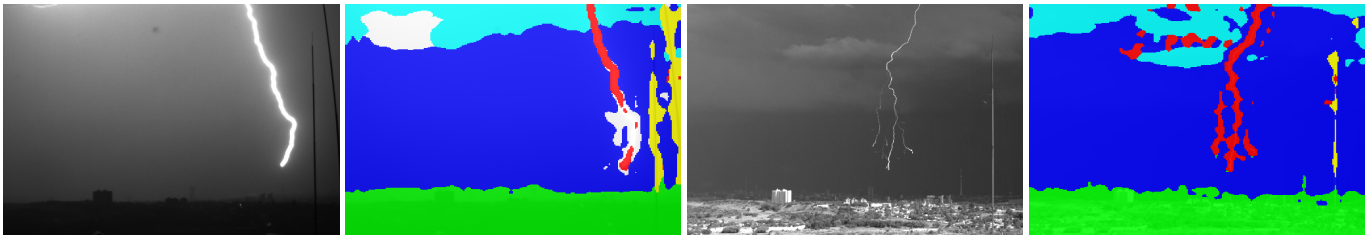
(**a**) Input image (**b**) Brightened image (**c**) Output Label

(**d**) Input image (**e**) Brightened image (**f**) Output Label

Figure 11. Examples with very dark input images.

6.4. Misclassification

Figure 12 shows examples of label misclassification by the trained system. In Figure 12b, the poles on the right of the frame include a small patch labeled *ground* close to the skyline, and the top sky portion is misclassified as *cloud*. Ancillary label misclassification is not critical since the crucial labels are *leader* and *stroke*. In Figure 12d, a portion of brightly lit cloud is misclassified as *leader*, which affects directional classification. Despite these errors, the segmentation is broadly correct, and the crucial labels cover the lightning activity. The events these frames were taken from correctly classified the direction of the lightning despite the inaccuracies.



(**a**) Input image (**b**) Output Label (**c**) Input image (**d**) Output Label

Figure 12. Example misclassifications.

6.5. Recommendations

The overall low accuracy (Table 4) indicates that the implementation cannot currently replace human metric collection. The linear dependence of the classification system on the semantic segmentation network and sensitivity to noise in low-light footage leads to accumulating classification errors. However, the semantic segmentation labels are useful for identifying low luminance lightning events in dim footage or emphasizing lightning during result collection. Improving the quality of input data by using a high-speed camera with a higher dynamic range or better low-light performance would be beneficial. Capturing a larger portion of the electromagnetic spectrum is a potential area for further research. Reducing noise could be achieved by profiling the camera under different temperature conditions to create model-specific sensor noise patterns, which could be used in denoising techniques with software like OpenCV, Nuke, or MATLAB.

Utilizing exposure settings recorded as metadata for automated exposure correction could help normalize sequences across different input footage. Developing a full CINE reader to remove manual exporting of image sequences and maintain a link between video metadata and CNN input footage would be useful. Incorporating additional data sources from extra sensors could improve data classification. Evaluating temporal luminance curves of lightning footage, masked by CNN output *leader* labels, might help classify attachment events by correlating luminance to current profiles. Extending CNNs to use tapped gates with temporal memory, such as LSTMs or RNNs, would introduce the time domain into segmentation, helping stabilize noise and reduce misclassifications. Using multiple neural network architectures in parallel or averaging the output of several networks could reduce dependence on single network performance, though it requires more computing power. Increasing the training set by labeling more footage would likely improve semantic segmentation results and reduce the risk of overfitting. Finally, a full assessment of the direction classification and different networks would yield an interesting investigation as well as training the models on intra-cloud lightning as well.

7. Conclusions

This study introduces a novel approach to analyzing high-speed lightning footage using deep learning, focusing on semantic segmentation networks. We evaluated DeepLabv3+, SegNet, FCN8s, U-Net, and AlexNet. DeepLabv3+ achieved the highest accuracy (93–95%), followed by SegNet and FCN8s. Despite their performance, overall system accuracy for event classification remains insufficient to replace manual analysis due to accumulating classification errors and sensitivity to noise in low-light footage. Improving input data quality with high-speed cameras offering higher dynamic ranges or better low-light performance is essential. Incorporating exposure metadata and developing a full CINE reader to streamline data processing would enhance workflow efficiency. Using extra sensors and evaluating temporal luminance curves could refine event classification.

Future work should explore integrating temporal memory networks like LSTMs or RNNs to stabilize noise and reduce misclassifications. Implementing multiple neural network architectures in parallel or averaging outputs could reduce dependence on a single network, though it requires more computational resources. Expanding the training dataset with more labeled footage would likely improve semantic segmentation results and mitigate overfitting. In conclusion, while our deep learning approach shows promise, especially in detecting low luminance events, further improvements in data quality, processing techniques, and network architectures are needed to achieve reliable and accurate classification for practical applications.

Author Contributions: Conceptualization, C.S. and H.G.P.H.; Data curation, T.C., J.R.S., C.S. and H.G.P.H.; Formal analysis, T.C. and J.R.S.; Funding acquisition, T.A.W. and H.G.P.H.; Methodology, T.C. and J.R.S.; Project administration, C.S., T.A.W. and H.G.P.H.; Resources, C.S., T.A.W. and H.G.P.H.; Software, T.C. and J.R.S.; Supervision, C.S. and H.G.P.H.; Validation, T.C., J.R.S. and C.S.; Visualization, T.C.; Writing—original draft, T.C. and J.R.S.; Writing—review & editing, C.S. and H.G.P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work is based on the research supported in part by the National Research Foundation of South Africa and their support of the research through the Thuthuka programme (Unique Grant No.: TTK23030380641) and by DEHNAFRICA and their support of the Johannesburg Lightning Research Laboratory.

Data Availability Statement: The datasets presented in this article are not readily available because of size and transfer limitations. Requests to access the datasets should be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results. ZT Research had no role in the design of the study; in the collection, analyses, or interpretation of data; in writing of the manuscript, or in the decision to publish results.

References

1. Geldenhuys, H.; Eriksson, A.; Bourn, G. Fifteen years' data of lightning current measurements on a 60 m mast. *Trans. S. Afr. Inst. Electr. Eng.* **1989**, *80*, 98–103.
2. Diendorfer, G.; Pichler, H.; Mair, M. Some Parameters of Negative Upward-Initiated Lightning to the Gaisberg Tower (2000–2007). *IEEE Trans. Electromagn. Compat.* **2009**, *51*, 1–12. [CrossRef]
3. Rakov, V.A.; Uman, M.A. *Lightning Physics and Effects*; Cambridge University Press: Cambridge, UK, 2003.
4. Saba, M.; Ballarotti, M.; Pinto, O., Jr. Negative cloud-to-ground lightning properties from high-speed video observations. *J. Geophys. Res. Atmos.* **2006**, *111*. [CrossRef]
5. Saba, M.; Schulz, W.; Warner, T.; Campos, L.; Orville, R.; Krider, E.; Cummins, K.; Schumann, C. High-speed video observations of positive lightning flashes. *J. Geophys. Res. Atmos.* **2010**, *115*. [CrossRef]
6. Ballarotti, M.; Saba, M.; Pinto, O., Jr. High-speed camera observations of negative ground flashes on a millisecond-scale. *Geophys. Res. Lett.* **2005**, *32*. [CrossRef]
7. Ballarotti, M.G.; Medeiros, C.; Saba, M.M.F.; Schulz, W.; Pinto, O., Jr. Frequency distributions of some parameters of negative downward lightning flashes based on accurate-stroke-count studies. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]
8. Warner, T.A. Observations of simultaneous upward lightning leaders from multiple tall structures. *Atmos. Res.* **2012**, *117*, 45–54. [CrossRef]
9. Warner, T.A.; Cummins, K.L.; Orville, R.E. Upward lightning observations from towers in Rapid City, South Dakota and comparison with National Lightning Detection Network data, 2004–2010. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]
10. Warner, T.A.; Saba, M.M.F.; Schumann, C.; Helsdon Jr., J.H.; Orville, R.E. Observations of bidirectional lightning leader initiation and development near positive leader channels. *J. Geophys. Res. Atmos.* **2016**, *121*, 9251–9260. [CrossRef]
11. Saba, M.M.; Schumann, C.; Warner, T.A.; Ferro, M.A.S.; de Paiva, A.R.; Helsdon, J., Jr.; Orville, R.E. Upward lightning flashes characteristics from high-speed videos. *J. Geophys. Res. Atmos.* **2016**, *121*, 8493–8505. [CrossRef]
12. Schumann, C.; Saba, M.M.F.; Warner, T.A.; Ferro, M.A.S.; Helsdon, J.H.; Thomas, R.; Orville, R.E. On the Triggering Mechanisms of Upward Lightning. *Sci. Rep.* **2019**, *9*, 9576. [CrossRef] [PubMed]
13. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
14. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
15. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2015; pp. 234–241.
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
18. BOYS, C.V. Progressive Lightning. *Nature* **1926**, *118*, 749–750. [CrossRef]
19. Schonland, B. Progressive lightning IV-The discharge mechanism. *Proc. R. Soc. Math. Phys. Eng. Sci.* **1938**, *164*, 132–150.
20. Chapter 5 Stepped Leader. In *The Lightning Discharge*; Uman, M.A., Ed.; Academic Press: Cambridge, MA, USA, 1987; Volume 39, pp. 82–98. [CrossRef]
21. Waldteuffl, P.; Metzger, P.; Boulay, J.L.; Laroche, P.; Hubert, P. Triggered lightning strokes originating in clear air. *J. Geophys. Res. Ocean.* **1980**, *85*, 2861–2868. [CrossRef]
22. Brantley, R.D.; Tiller, J.A.; Uman, M.A. Lightning properties in Florida thunderstorms from video tape records. *J. Geophys. Res.* **1975**, *80*, 3402–3406. [CrossRef]

23. Jordan, D.M.; Rakov, V.A.; Beasley, W.H.; Uman, M.A. Luminosity characteristics of dart leaders and return strokes in natural lightning. *J. Geophys. Res. Atmos.* **1997**, *102*, 22025–22032. [CrossRef]
24. Winn, W.P.; Aldridge, T.V.; Moore, C.B. Video tape recordings of lightning flashes. *J. Geophys. Res.* **1973**, *78*, 4515–4519. [CrossRef]
25. Moreau, J.P.; Alliot, J.C.; Mazur, V. Aircraft lightning initiation and interception from in situ electric measurements and fast video observations. *J. Geophys. Res. Atmos.* **1992**, *97*, 15903–15912. [CrossRef]
26. Jiang, R.; Qie, X.; Wu, Z.; Wang, D.; Liu, M.; Lu, G.; Liu, D. Characteristics of upward lightning from a 325-m-tall meteorology tower. *Atmos. Res.* **2014**, *149*, 111–119. [CrossRef]
27. Mazur, V.; Ruhnke, L.H. Physical processes during development of upward leaders from tall structures. *J. Electrost.* **2011**, *69*, 97–110. [CrossRef]
28. Flache, D.; Rakov, V.A.; Heidler, F.; Zischank, W.; Thottappillil, R. Initial-stage pulses in upward lightning: Leader/return stroke versus M-component mode of charge transfer to ground. *Geophys. Res. Lett.* **2008**, *35*. [CrossRef]
29. Wang, D.; Takagi, N.; Watanabe, T.; Sakurano, H.; Hashimoto, M. Observed characteristics of upward leaders that are initiated from a windmill and its lightning protection tower. *Geophys. Res. Lett.* **2008**, *35*. [CrossRef]
30. Schumann, C.; Hunt, H.G.; Tasman, J.; Fensham, H.; Nixon, K.J.; Warner, T.A.; Saba, M.M. High-speed video observation of lightning flashes over Johannesburg, South Africa 2017–2018. In Proceedings of the 2018 34th International Conference on Lightning Protection (ICLP), Rzeszow, Poland, 2–7 September 2018; pp. 1–7.
31. Fensham, H.; Hunt, H.G.; Schumann, C.; Warner, T.A.; Gijben, M. The Johannesburg Lightning Research Laboratory, Part 3: Evaluation of the South African Lightning Detection Network. *Electr. Power Syst. Res.* **2023**, *216*, 108968. [CrossRef]
32. Guha, A.; Liu, Y.; Williams, E.; Schumann, C.; Hunt, H., Lightning Detection and Warning. In *Lightning: Science, Engineering, and Economic Implications for Developing Countries*; Gomes, C., Ed.; Springer: Singapore, 2021; pp. 37–77. [CrossRef]
33. Saba, M.M.F.; Campos, L.Z.S.; Krider, E.P.; Pinto Jr., O. High-speed video observations of positive ground flashes produced by intracloud lightning. *Geophys. Res. Lett.* **2009**, *36*. [CrossRef]
34. Ding, Z.; Rakov, V.A.; Zhu, Y.; Kereszy, I.; Chen, S.; Tran, M.D. Propagation Mechanism of Branched Downward Positive Leader Resulting in a Negative Cloud-To-Ground Flash. *J. Geophys. Res. Atmos.* **2024**, *129*, e2023JD039262. [CrossRef]
35. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*; Adaptive Computation and Machine Learning series; MIT Press: Cambridge, MA, USA, 2018.
36. Cust, E.E.; Sweeting, A.J.; Ball, K.; Robertson, S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *J. Sport. Sci.* **2019**, *37*, 568–600. [CrossRef]
37. Pickup, L.C. Machine Learning in Multi-Frame Image Super-Resolution. Ph.D. Thesis, Oxford University, Oxford, UK, 2007.
38. Gandhi, T.; Trivedi, M.M. Computer vision and machine learning for enhancing pedestrian safety. In *Computational Intelligence in Automotive Applications*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 59–77.
39. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]
40. Essa, Y.; Ajoodha, R.; Hunt, H.G. A LSTM Recurrent Neural Network for Lightning Flash Prediction within Southern Africa using Historical Time-series Data. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 16–18 December 2020; pp. 1–6. [CrossRef]
41. Essa, Y.; Hunt, H.G.; Ajoodha, R. Short-term Prediction of Lightning in Southern Africa using Autoregressive Machine Learning Techniques. In Proceedings of the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 21–24 April 2021; pp. 1–5. [CrossRef]
42. Essa, Y.; Hunt, H.G.P.; Gijben, M.; Ajoodha, R. Deep Learning Prediction of Thunderstorm Severity Using Remote Sensing Weather Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4004–4013. [CrossRef]
43. Qiu, T.; Zhang, S.; Zhou, H.; Bai, X.; Liu, P. Application study of machine learning in lightning forecasting. *Inf. Technol. J.* **2013**, *12*, 6031–6037. [CrossRef]
44. Booysens, A.; Viriri, S. Detection of lightning pattern changes using machine learning algorithms. In Proceedings of the International Conference on Communications, Signal Processing and Computers, Guilin, China, 18 December 2014.
45. Leal, A.F.; Rakov, V.A.; Rocha, B.R. Compact intracloud discharges: New classification of field waveforms and identification by lightning locating systems. *Electr. Power Syst. Res.* **2019**, *173*, 251–262. [CrossRef]
46. Leal, A.F.R.; Rakov, V.A. Characterization of Lightning Electric Field Waveforms Using a Large Database: 1. Methodology. *IEEE Trans. Electromagn. Compat.* **2021**, *63*, 1155–1162. [CrossRef]
47. Leal, A.F.R.; Matos, W.L.N. Short-term lightning prediction in the Amazon region using ground-based weather station data and machine learning techniques. In Proceedings of the 2022 36th International Conference on Lightning Protection (ICLP), Cape Town, South Africa, 16 November 2022; pp. 400–405. [CrossRef]
48. Leal, A.F.R.; Ferreira, G.A.V.S.; Matos, W.L.N. Performance Analysis of Artificial Intelligence Approaches for LEMP Classification. *Remote Sens.* **2023**, *15*, 5635. [CrossRef]
49. Mansouri, E.; Mostajabi, A.; Tong, C.; Rubinstein, M.; Rachidi, F. Lightning Nowcasting Using Solely Lightning Data. *Atmosphere* **2023**, *14*, 1713. [CrossRef]
50. Mostajabi, A.; Mansouri, E.; Rubinstein, M.; Tong, C.; Rachidi, F. Machine Learning Based Lightning Nowcasting using Single-Site Meteorological Observations and Lightning Location Systems Data. In Proceedings of the 36th International Conference on Lightning Protection (ICLP), Johannesburg, South Africa, 2–7 October 2022.

51. Zhu, Y.; Bitzer, P.; Rakov, V.; Ding, Z. A Machine-Learning Approach to Classify Cloud-to-Ground and Intracloud Lightning. *Geophys. Res. Lett.* **2020**, *48*, e2020GL091148. [CrossRef]
52. Smit, J.; Schumann, C.; Hunt, H.; Cross, T.; Warner, T. Generation of metrics by semantic segmentation of high speed lightning footage using machine. In Proceedings of the SAUPEC/RobMech/PRASA 2020, Potchefstroom, South Africa, 27–29 January 2020.
53. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
54. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Kdd, 1996; Volume 96, pp. 226–231.
55. Fensham, H.; Schumann, C.; Hunt, H.; Tasman, J.; Nixon, K.; Warner, T.; Gijben, M. Performance evaluation of the SALDN using highspeed camera footage of ground truth lightning events over Johannesburg, South Africa. In Proceedings of the 2018 34th International Conference on Lightning Protection (ICLP), Rzeszow, Poland, 25 October 2018; pp. 1–5.
56. Lanier, L. *Digital Compositing with Nuke*; Routledge: London, UK, 2012.
57. Okun, J.A.; Zwerman, S. *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures*; Taylor & Francis: Abingdon, UK, 2010.
58. Abler, J. *Rendering and Compositing for Visual Effects*; East Tennessee State University: Johnson City, TN, USA, 2015.
59. Nuke | VFX Software | Foundry. 2019. Available online: https://www.foundry.com/products/nuke (accessed on 9 March 2019).
60. Research, V. *Cine File Format*, 705 v12.0.705.0 ed.; Ametek: Wayne, NY, USA, 2011.
61. Phantom High Speed. Available online: https://www.phantomhighspeed.com/ (accessed on 9 March 2019).
62. DaVinci Resolve 16 | Blackmagic Design. Available online: https://www.blackmagicdesign.com/sa/products/davinciresolve/ (accessed on 9 March 2019).
63. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.