COUD: Continual Urbanization Detector for Time Series Building Change Detection

Yitao Zhao, *Graduate Student Member, IEEE*, Heng-Chao Li, *Senior Member, IEEE*, Sen Lei, *Member, IEEE*, Nanqing Liu, *Member, IEEE*, Jie Pan, and Turgay Celik, *Senior Member, IEEE*

***Abstract*—Building change detection on remote sensing images is an important approach to monitoring the urban expansion and sustainable development of natural resources. In conventional building change detection tasks, only changed regions between two time phases are typically concerned. The relevance and trend of spatiotemporal changes between multiple time phases are neglected in most cases. In this article, we propose a two-stage continual urbanization detector (COUD) for time series urban building change detection task. The COUD method employs self-supervised pretraining for feature refinement, and performs optimization through temporal distillation approach. Consequently, multitemporal feature extraction and changing regions localization of urban building complexes are conducted. Considering the gap in available dataset for time series change detection task, we produce and release a time series dataset named "TSCD". Chengdu region of China is selected as the study area in this research, which is partially covered by the proposed TSCD dataset. By applying the proposed COUD method to the selected study area for exploring the changing pattern from 2016 to 2022, a comprehensive analysis is conducted in conjunction with actual planning policies published by the management department. Extensive experimental results confirm the reliability of our proposed method.**

***Index Terms*—Change detection (CD), remote sensing, temporal distillation, time series, urbanization.**

I. Introduction

FOR management departments, an accurate understanding of the urbanization process is of great significance for urban development decisions [1], [2]. After decades of rapid economic and industrial development, China, a developing country, is undergoing tremendous evolution [3], [4]. Including the expansion of new city complexes, the demolition and renovation of old city buildings, and the renewal and evolution of agricultural and forestry land [5], [6], [7]. With the growth of the urbanization rate, the proportion of China's urban population has exceeded 60% [8], [9], [10].

This greatly promotes the transformation and upgrading of the industrial structure, while simultaneously providing an obvious boost to per capita GDP. However, urbanization inevitably leads to a series of challenges to the maintenance of the ecological environment at the same time. For example, the generation of the urban heat island effect, the shortage of groundwater, the destruction of the ecological environment around urban areas, and geological disasters caused by terrain reconstruction, etc. [11], [12], [13]. Without external constraints on urban development, limited natural resources will be severely eroded. Therefore, macroassessment and regulation of the urbanization changing trends are the top priority of urban management.

To prevent such situation, it is necessary for the management departments to implement a series of corresponding measures to discipline and facilitate the general direction of urbanization [14]. One of the most intuitive ways to reflect the urbanization process is building change detection (CD) [15], [16]. CD collects multiple information from two or more time phases of the same geo-location, and utilizes a series of analysis methods to extract the differential features [17], [18], [19]. By this approach, new construction and demolition of urban building complex can be located. Initial CD tasks heavily rely on manual walk-through survey, but the inefficiency of human counting could not match the update rate of urban building complex. With the rapid emergence of remote sensing imaging technology, numerous medium-resolution and high-resolution Earth observation platforms have been established to expand the perspective, such as WorldView, Quickbird, ZY Series, and GF Series satellites [20], [21]. In addition to on-board optical sensors, multisource Earth observation platforms, such as TerraSAR, Sentinel Series, COSMO-SkyMed satellites are available to fill the gap in data acquisition under severe climates [22], [23], [24], [25], [26]. The proliferate of remote sensing data shifts the form of CD from manual scattered investigation to large-scale interpretation.

Nevertheless, it is inefficient to rely on expert knowledge-based interpretation on single-temporal image and subsequent comprehensive judgments [27], [28]. Some researchers proposed various of automatic algorithms to improve the efficiency of CD by algebra-based methods, such as CVA method [29], ratio methods [30], [31], and normalized difference index

Received 20 July 2024; revised 18 September 2024; accepted 4 October 2024. Date of publication 17 October 2024; date of current version 8 November 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271418 and Grant 61871335, and in part by the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC0030, and in part by the Fellowship of China National Postdoctoral Program for Innovative Talents under Grant BX20240291. *(Corresponding author: Heng-Chao Li.)*

Yitao Zhao, Heng-Chao Li, Sen Lei, and Nanqing Liu are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: ytzhao@my.swjtu.edu.cn; hcli@swjtu.edu.cn; senlei@swjtu.edu.cn; lansing163@163.com).

Jie Pan is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100045, China (e-mail: panjie@aircas.ac.cn).

Turgay Celik is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China, and also with the School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2017, South Africa.

The code is available https://github.com/MarsZhaoYT/TSCD-Dataset

Digital Object Identifier 10.1109/JSTARS.2024.3482559

methods [32], [33]. In addition, feature transformation-based methods were also proposed, such as MAF method [34], tasselled cap transformation method [35], and PCA method [36], [37], [38]. However, the aforementioned methods are usually limited in specific scenarios and depend on data preprocessed. Therefore, such kind of algorithms is not capable of handling the massive and multiple formats of remote sensing data.

The emergence of deep learning-based computer vision technology brings large-scale remote sensing CD methods into a new era [39], [40]. Daudt et al. [41] proposed three networks with fully convolutional U-Net structure to handle the CD on bitemporal images. Chenxiao et al. [42] exploited a two-stream image fusion network in which effective features were obtained by the fusion of deep feature and original samples with the difference discrimination module. In [43], a dual-branch U-Net++ architecture with ECAM attention mechanism was utilized to extract and densely connect the hierarchical features. Chen et al. [44] proposed an efficient transformer-based method for bitemporal CD, in which the original embedding was split into several tokens to model image contexts. In [45], a deep siamese PCFN model was proposed to detect semantic changes by leveraging temporal correlation. Lei et al. [46] adopted a lightweight architecture with multiscale decoupled convolution and feature cooperation to improve the performance of high-resolution CD task. For building CD, Cao and Huang [47] proposed a full-level fused transfer-learning method which contains a building extraction network, a pseudolabel generation module and a CD module.

In the conventional CD pipeline, bitemporal features are extracted individually and fused differentially. Then, the feature distances are measured to optimize the model with a fully supervised loss function [48], [49], [50]. However, the fully supervised CD suffers from the drawback of data-driven limitation, which requires pixel-level manual labeling of the change region for each pair of bitemporal samples. Considering the fact that intensive manual labeling is very tedious and time-consuming, researchers turn the interest to the application of semisupervised and self-supervised approaches in CD tasks, promoting the model to discover the latent features with unlabeled datasets. Peng et al. [51] proposed an end-to-end architecture with lightweight attention applied to improve the segmentation performance. In [52], a self-supervised framework considering contrast on seasonal factor was proposed to deal with the domain gap in remote sensing images. Akiva et al. [53] designed a self-supervised method to obtain better representation on the material and texture of ground objects. Noh et al. [54] offered a distinct pipeline of unsupervised CD via domain transferring method. Single temporal image was transferred by a CycleGAN [55] model to obtain the pseudochange pairs, then the reconstruction loss were measured between the pairs. For semantic representation, Chen et al. [56] proposed a triple-branch contrastive framework for semisupervised CD. A novel CD method leveraging dual-frequency signals and interactive fusion was proposed in [57]. Jiang et al. [58] applied a SimSiam [59] structure to self-supervised contrastive pretraining for CD.

Although both the prosperity of deep learning and massive remote sensing data provide dual additives to remote sensing CD tasks, there still remain several limitations as follows.

1) The vast majority of the existing deep learning-based CD methods are restricted in bitemporal scenarios. However, urbanization is a continuous and long-term process and the analysis of change patterns of time series data is inevitable.
2) Expert knowledge is required to interpret the complex scenarios of remote sensing data, which greatly limits the efficiency of manual labeling. As a consequence, manual labeling is not able to fulfill the gap of comprehensive processing for massive remote sensing data. In addition, environmental factors, such as illumination and shadows, are common among data from various platforms, which also cause negative impacts on feature extraction.

However, few research works with deep learning-based methods for time series CD are conducted [60], [61]. The majority of the existing exploration on time series CD task concentrate on statistical analysis method [2], [62] and unsupervised learning method [63]. These methods suffer from shortcomings in effectiveness and generalization, as well as the limited capability of feature extraction from the massive unlabeled remote sensing data.

Overall, the main challenge for time series CD is that: *How to fully leverage the available labeled data to capture temporal context information to design a continual time series CD framework?*

Motivated by this, we perform an investigation on deep learning-based time series CD of remote sensing images. In this article, we consider the CD of building complex in the urbanization process as our main task and propose a time series-based continual urbanization detector (COUD). The COUD method is organized in a two-stage approach, containing pretraining for COUD framework and temporal distillation supported fine-tuning. Unlabeled RS data are utilized for the pretraining stage for COUD framework. And the temporal distillation supported fine-tuning, driven by the pretrained encoder, is assigned to sequentially perform temporal optimization between a series of individual models by temporal distillation.

The main contributions can be summarized as follows.

1) To the best of our knowledge, it is the first to propose a self-supervised pretraining framework "COUD" to fully leverage the spatiotemporal patterns in massive unlabeled time series RSI data.
2) Two adaptive pretraining schemes are provided in our proposed framework for various data formats, which greatly expands the applicable scenarios. Contrast-COUD is suitable for discrete RSI data, while TS-COUD utilizes grouped times series data for optimization.
3) We originally annotate and release a challenging benchmark (TSCD) of Chengdu, China from 2016 to 2022 to fill the gap in time series CD dataset. Extensive experiments and analysis are performed on this dataset with the proposed method.
4) In the downstream fine-tuning, we design a continual CD approach optimized with temporal distillation to endow the model with the capability to perceive the temporal context.

The rest of this article is organized as follows. Section II expounds study site and data involved this article. The details of

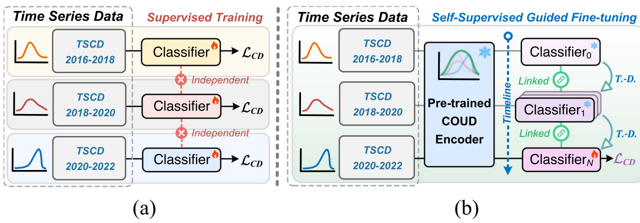

Fig. 1. **Existing CD methods versus our COUD.** Illustration of the utilization of temporal context from pretrained encoder in our COUD method. (a) Existing CD methods. (b) Proposed COUD method.

TABLE I
SPECIFIC PARAMETERS OF THE SOURCE IMAGES OF TSCD DATASET

Original Image	Acquisition Date	Resolution	Number of Bands
2016	2016.2.8	0.5 m/pixel	3 (RGB)
2018	2018.11.11		
2020	2020.1.25		
2022	2022.8.22		

the COUD method are described in Section III. Extensive experimental results are presented in Section IV. Finally, Section V concludes this article.

II. STUDY SITE AND DATA COLLECTION

A. Study Site

Since the "Western Development" strategy has been proposed by the Chinese government in 2000, comprehensive construction covering the economy, technology, and ecological environment has been carried out in the western region of China. As a significant central city in western China, Chengdu, Sichuan Province is selected as the study area in this article to evaluate the proposed method described later.

Chengdu, the capital of Sichuan Province, is located in the southeast of China. With a permanent population of more than 20 million, the city covers an area of about 14 355 km^2. As shown in the Fig. 2, the specific study area is located at Tianfu New Area Science City in the south of Chengdu, near the Xinglong Lake. During 2016 and 2022, this region has undergone a significant evolution of ground objects (e.g., buildings, vegetation, cultivated), which is suitable for performing experiments as a typical study area.

B. Data Collection

The source images of study area are acquired by the WorldView 02 Satellite with the resolution of approximately 0.5 m/pixel, which are shown in the Table I. Considering the continuity of the surface change process, we set two years as the sampling interval and collect images for four periods of 2016, 2018, 2020, and 2022, respectively. As shown in Fig. 3, to alleviate the impact of external factors when image acquisition, image co-registration by manual control points selection and resampling are performed to ensure a uniform coordinate frame.

TABLE II
SETTING OF THE PROPOSED TSCD DATASET

Dataset	Temporal of Subset	Train/Val/Test
TSCD	2016-2018	900/257/128
	2018-2020	900/257/128
	2020-2022	900/257/128

The number of train/val/test is pairs of bitemporal image samples.

After that, we conduct a dense labeling of building objects on the images for each time phase to acquire the corresponding building distribution maps. Three changing labels (2016–2018, 2018–2020, 2020–2022) are obtained by differential operations on the adjacent building maps. The final TSCD dataset is generated by uniformly cropping and dataset partition on the original images/labels. Both building polygon labels in ".tif" format and vertex coordinates in ".json" format are provided in the TSCD dataset. The specific setting of the proposed TSCD dataset is listed in the Table II. For better illustration, we present the multilevel metadata and two series of temporal samples in the TSCD dataset are shown in Fig. 4.

III. PROPOSED METHOD

In this section, the key components of the proposed COUD method will be introduced in detail. Afterward, the timing-aware temporal distillation for fine-tuning is also explained.

A. Overview

The overall workflow of COUD is illustrated in Fig. 5. In contrast to conventional CD methods, our proposed method pays more attention to continual CD task in a time series context. We customize two keys steps for time series processing: **pretraining for coud framework** and **temporal distillation supported fine-tuning**. In the pretraining stage, massive unlabeled data are exploited to optimize the COUD framework, in which the encoder will be reused in the fine-tuning stage. Following this, a temporal distillation updating approach for continual CD is introduced to align multitemporal classifiers. In comparison with conventional CD methods illustrated in Fig. 1, our method focuses more on the time dimension in which various temporal phases are connected and unified by incorporating temporal contextual.

B. Pretraining for COUD Framework

Conventional CD paradigm requires extensive labels for optimization. However, in label-limited scenarios (e.g., time series CD), the model may suffer from low robustness and instability. Therefore, we provide two pretraining paradigms with massive unlabeled data for different scenarios.

1) Contrast-COUD Pretraining: As shown in Fig. 5(a), Contrast-COUD is applicable to ungrouped RSI data and is organized in a concise architecture. For each unlabeled single-temporal sample $I \in \mathbb{R}^{H\times W\times C}$ in the unlabeled dataset, view augmentations are performed to generate paired views $\mathcal{V} = \{Q, K\}$ with distinct styles. Afterward, the two generated views

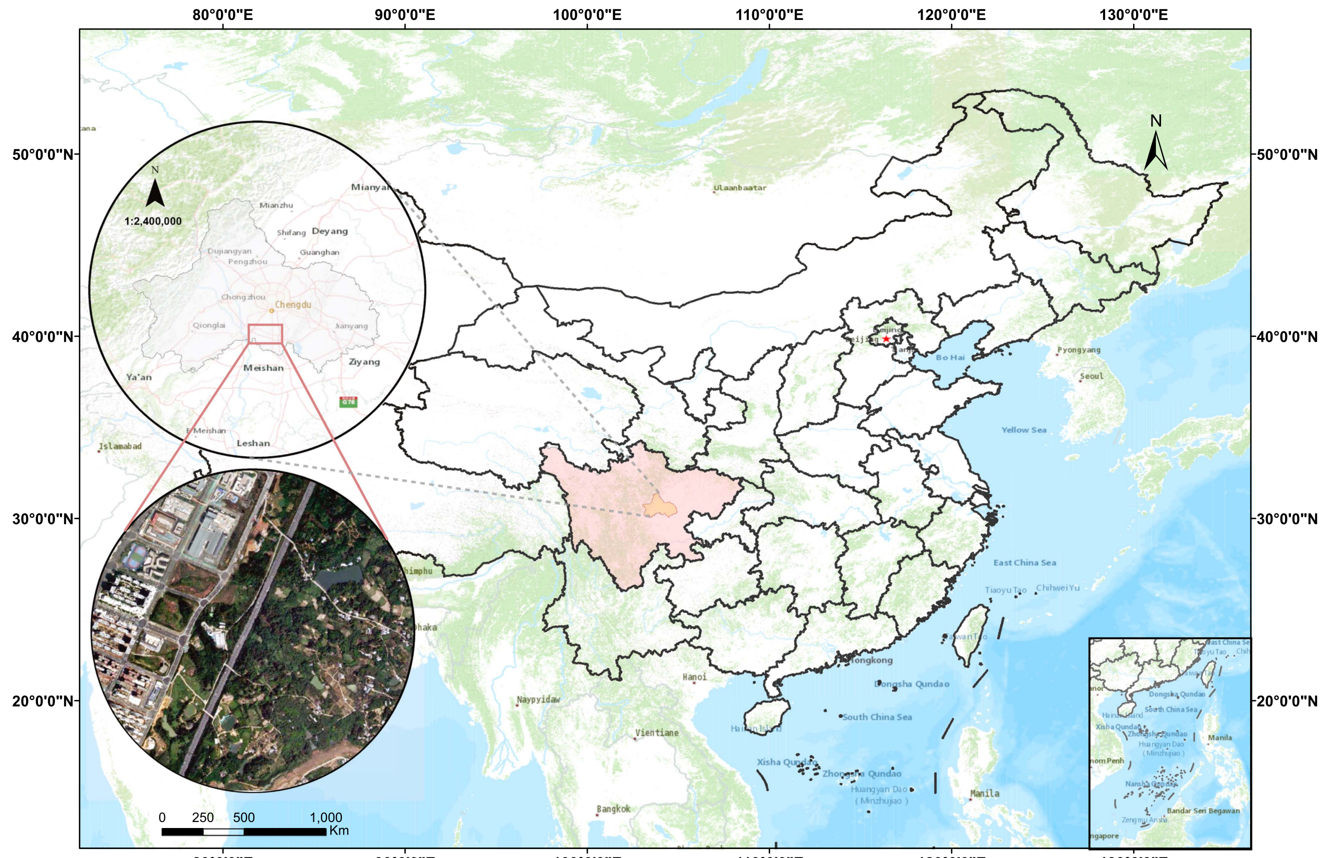

Fig. 2. Study site of this article. The specific region is in the south of Chengdu, Sichuan Province. A thumbnail of the area is presented in the upper left corner, and a sample of satellite imagery covering the area is displayed in the lower left corner.

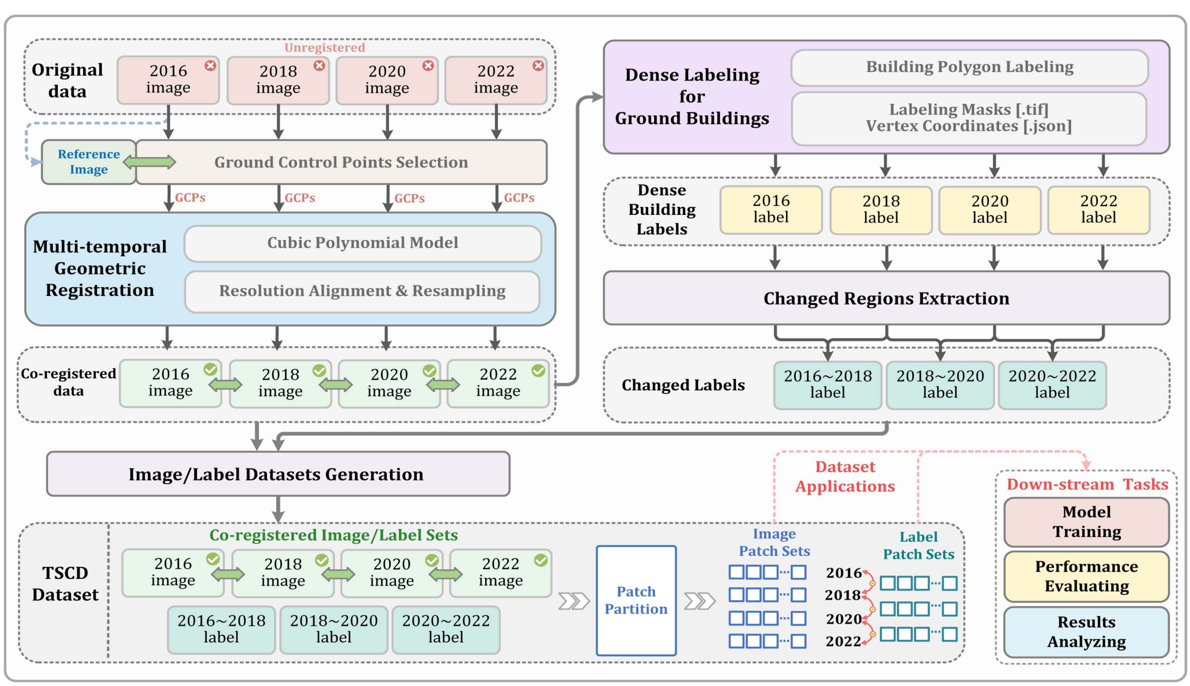

Fig. 3. Workflow of the preprocessing and dense labeling for TSCD dataset.

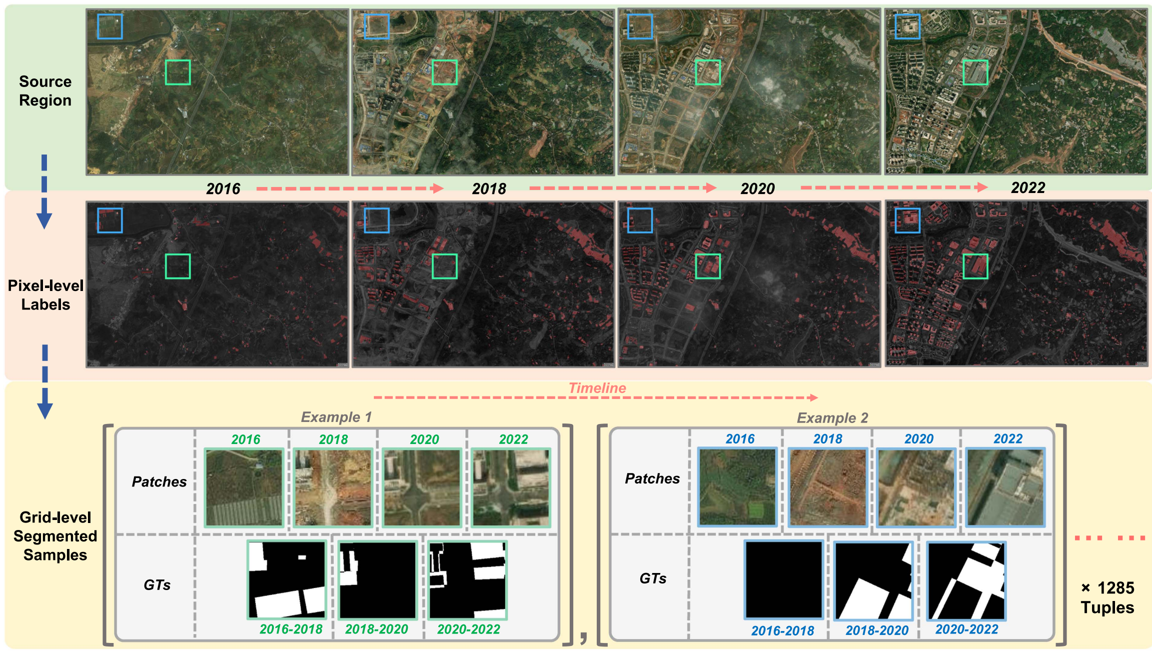

Fig. 4. Overview of the source data and segmented samples in TSCD dataset.

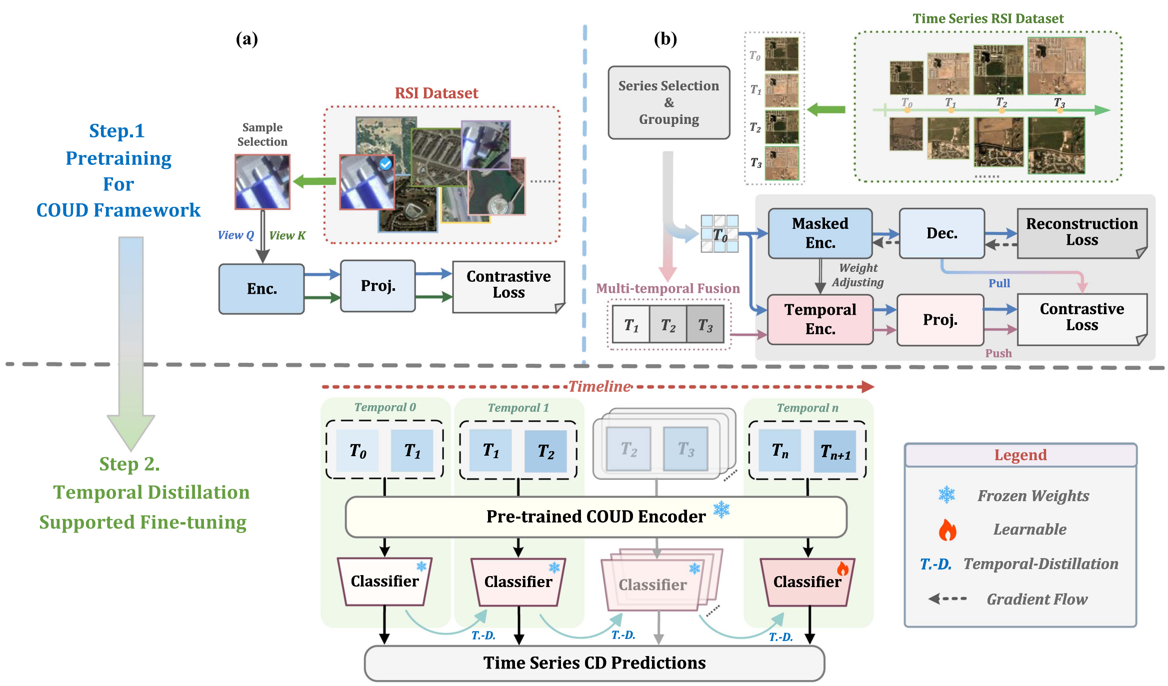

Fig. 5. Workflow of the proposed COUD framework for time series CD task. The whole framework is composed of pretrained stage and fine-tuning stage. Two pretraining paradigms are proposed for discrete RSI data and time series RSI data. Then, the continual CD is performed with temporal distillation to correlate the temporal context. (a) Contrast-COUD. (b) TS-COUD.

Q/K are fed into the lightweight encoder for feature extraction. A projector is attached to process the embedding features from the encoder. Both the hidden features $\mathcal{H}$ and the final semantic vectors $\mathcal{S}$ are acquired through the contrastive MLP. The forward calculation is as follows:

$$\{\mathcal{H}, \mathcal{S}\} = \mathbf{P}\left[\mathbf{E}(\mathcal{V})\right] \tag{1}$$

where $\mathbf{E}$ represents the encoder, and $\mathbf{P}$ is the projector in Contrast-COUD. $\mathcal{H}$ and $\mathcal{S}$ are the hidden features and output semantic vectors, respectively.

In the proposed Contrast-COUD paradigm, the specific model architecture is listed as follows.

- *Lightweight encoder:* U-Net architecture with ResNet-18 downsampling stages. Five upsampling blocks (Deconv →

BN→ReLU) are utilized to restore the feature size as the input samples.

- *Contrastive MLP:* Three-layer MLP with BN layers. The dimensions of hidden features and output semantic vectors are 2048-dim.

In this stage, only single-temporal images without any labels are fed into the Contrast-COUD framework. The pretraining loss function to pull the features together is as follows:

$$\mathcal{L}_C = \frac{1}{N} \sum_{(i,j) \in \mathcal{B}, i \neq j} \boldsymbol{\mathcal{D}} \left[\mathcal{H}_i, \mathcal{S}_j \right] \tag{2}$$

where $\mathcal{H}_i$ and $\mathcal{S}_i$ are the ith element of the hidden features and semantic vectors, respectively. $\mathcal{B}$ represents the binary selective set of $\{0, 1\}$. N is the length of the output list, and $\mathcal{D}$ is the cosine similarity [64].

2) TS-COUD Pretraining: As shown in Fig. 5(b), we also propose an alternative pretraining paradigm to fully leverage the temporal pattern of unlabeled time series RSI data.

Masking branch: For a group of time series data $\{I_{T_0}, I_{T_1}, I_{T_2}, I_{T_3}\}$, the initial state I_{T_0} is fed into an masked encoder with random masking, and the decoder reconstructs the latent feature to the input sample. The forward calculation process is as follows:

$$r_{T_0} = \mathbf{Dec_M} \left[\mathbf{Enc_M}(I_{T_0}) \right] \tag{3}$$

$$\mathcal{L}_M = \frac{1}{N_M} \sum\nolimits_{i=1}^{N_M} ||r_{T_0} - I_{T_0}||^2 \tag{4}$$

where $\mathbf{Enc_M}$ and $\mathbf{Dec_M}$ are the masked encoder and decoder, respectively. N_M are the patch number for reconstruction.

Temporal branch: Then, the subsequent states $\{I_{T_1}, I_{T_2}, I_{T_3}\}$ are fused according to the temporal distance α_t to the initial state

$$I_{T_{\text{fuse}}} = I_{T_1} + \alpha_t * I_{T_2} + (1 - \alpha_t) * I_{T_3}. \tag{5}$$

Simultaneously, the initial state I_{T_0} and the fused state are processed in parallel by the temporal encoder and projector to obtain corresponding prediction

$$p_{T_0} = \mathbf{Proj_T} \left[\mathbf{Enc_T}(I_{T_0}) \right] \tag{6}$$

$$p_{T_{\text{fuse}}} = \mathbf{Proj_T} \left[\mathbf{Enc_T}(I_{T_{\text{fuse}}}) \right] \tag{7}$$

where $\mathbf{Enc_T}$ and $\mathbf{Proj_T}$ are the temporal encoder and projector, respectively.

The specific model architecture of TS-COUD is listed as follows.

1) *Masked/temporal encoder*: Modified ViT-L-16 with 1024-dim embedding.
2) *Decoder*: Eight-layer multihead self-attention blocks with 512-dim embedding.
3) *Projector*: Four-layer MLP with BN layers. The dimensions of hidden features and output semantic vectors are 1024.

The objective function of temporal branch in TS-COUD is as follows:

$$\mathcal{L}_T = -\log \frac{\exp(\rho^+/\tau)}{\exp(\rho^+/\tau) + \sum_{j=1}^{K-1} (\exp(\rho^-/\tau))} \tag{8}$$

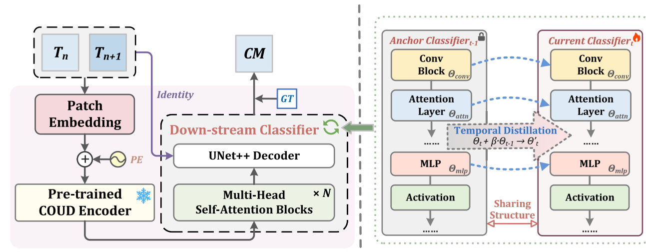

Fig. 6. Illustration of temporal distillation supported fine-tuning. The parameters of later classifier are updated based on the former anchor classifier.

where ρ^+ is the cosine similarity between r_{T_0} and p_{T_0}, and ρ^- is the cosine similarity between r_{T_0} and $p_{T_{\text{fuse}}}$.

The overall learning target is the weighted combination of loss for masking branch and temporal branch

$$L = \mathcal{L}_M + \mathcal{L}_T. \tag{9}$$

It is worth noting that in the backward propagation, the stop-gradient operation [59], [65] is conducted to avoid the mode collapse. The weight of temporal encoder is adjusted with the masked encoder by EMA strategy. Through the COUD pretraining, the model is promoted to acquire robustness to image background and sensitiveness to temporal pattern.

C. Temporal Distillation Supported Fine-Tuning

Downstream fine-tuning: After the pretrained stage, the weight of masked encoder is frozen and assigned to assist the downstream fine-tuning. As shown in the Fig. 6, the bitemporal embeddings pass through the pretrained COUD encoder to acquire the semantic tokens. Afterward, the original bitemporal embeddings is sent into the downstream CD classifier with the semantic tokens. Given a pair of bitemporal image pairs as I_{T_n} and $I_{T_{n+1}}$, the aforementioned calculation process is summarized as follows:

$$f_{T_n} = \mathbf{Enc}^\star(I_{T_n}, \mathbf{pe}) \tag{10}$$

$$\text{token}_{T_n} = \mathbf{MHSA}(\text{token}_{T_n}) \tag{11}$$

$$f_{T_{n+1}} = \mathbf{Enc}^\star(I_{T_{n+1}}, \mathbf{pe}) \tag{12}$$

$$\text{token}_{T_{n+1}} = \mathbf{MHSA}(\text{token}_{T_{n+1}}) \tag{13}$$

$$\text{CM} = \mathbf{Dec}\left([I_{T_n}, I_{T_{n+1}}], [\text{token}_{T_n}, \text{token}_{T_{n+1}}] \right) \tag{14}$$

where $\mathbf{Enc}^\star$ is the pretrained COUD encoder, $\mathbf{pe}$ represents positional encoding, $\mathbf{MHSA}$ is the multihead self-attention blocks, and $\mathbf{Dec}$ is the UNet++ decoder in downstream classifier.

Temporal distillation: Compared to conventional CD tasks, the proposed COUD is designed to pay more attention to the temporal context related to the continual changes in the urbanization process. Therefore, we formulate a temporal distillation approach, namely "T.-D.," to bridge the neighboring CD classifiers with temporal characteristic.

As shown in the Fig. 6, the anchor classifier is the well trained classifier in the previous time phase, as current classifier is the classifier to be optimized. Supposing that the anchor classifier

$(t-1)$ share the same architecture with the current classifier (t), we stipulate that each parameter is updated with two weighted components. Denoting the parameter sets of the previous classifier and current classifier as Θ_{t-1} and Θ_t, respectively. The optimization of the latest classifier parameter Θ_t' is calculated by

$$\Theta_t' \leftarrow \Theta_t + \beta_{\mathrm{TD}} * \Theta_{t-1} \tag{15}$$

where β_{TD} is the weight parameter for distillation.

Through the T.-D. updating, the current training classifier inherits the temporal-aware context of the former model by the timeline. Then, the ground truth is introduced to the classifier training by supervised fine-tuning. In the implementation, we adopt a concise classifier of an five-layer U-Net++ architecture to participate in the fine-tuning.

The experimental results of temporal distillation supported fine-tuning can be found in Section V-A.

IV. Experiments

A. Experimental Settings

1) Datasets: In this section, we introduce the datasets related to our experiments. Since the proposed COUD is a two-stage framework containing pretraining and downstream continual CD stages, which have been elaborated in Sections III-B and III-C in detail. Therefore, we present these two parts, respectively.

1) *Pretraining dataset of Contrast-COUD:* We crop the whole CDD dataset into 256 × 256 pixels and keep the original dataset partition of train/val/test with 10 000/3000/3000 pairs. Only the unlabeled single-temporal images are preserved to exploited for training, and the labels in the CDD dataset are removed.
2) *Pretraining dataset of TS-COUD:* We select four time phase of 2018-01, 2018-07, 2019-01, 2019-07 from the Space-Net-7 dataset to perform the TS-COUD pretraining. The images are cropped into 12 812 patched with 256 × 256 pixels.
3) *Downstream time series dataset:* In the downstream task, we fine-tune and evaluate the performance on the proposed TSCD dataset shown in Table II. We name the three subsets in TSCD by imaging year as *TSCD-2016–2018*, *TSCD-2018–2020*, *TSCD-2020–2022*. The comparative experiments are conducted on each subset.

2) Evaluation Metrics: In the evaluation procedure, three metrics of $F1$-score, IoU, and OA are reported, which are commonly utilized in CD tasks. Moreover, considering the temporal characteristic of the TSCD dataset, we additionally introduce three metrics related to temporal dimension: T-$F1$, T-IoU, and T-OA. The specific calculation methods for above-mentioned three metrics are as follows:

$$T\text{-}F1 = \frac{1}{M}\sum_{i=1}^{M} F1_i = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{2\times \mathrm{TP}}{2\times \mathrm{TP}+\mathrm{FN}+\mathrm{FP}}\right] \tag{16}$$

$$T\text{-IoU} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{IoU}_i = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}\right] \tag{17}$$

TABLE III
Comparison Results on the *TSCD-2016–2018* Dataset

Dataset	Method	F1-Score	IoU	OA
TSCD-2016-2018	FC-EF	61.40	44.30	97.93
	FC-Siam-Conc	55.53	38.44	97.21
	FC-Siam-Diff	57.72	40.57	97.60
	SNUNet-CD	68.50	52.90	98.47
	USSFCNet	59.02	41.86	98.13
	A2Net	59.67	42.53	98.11
	SEIFNet	64.79	47.92	98.05
	Contrast-COUD	**69.38**	**53.12**	**98.38**
	TS-COUD	**69.12**	**52.82**	**98.59**

$$T\text{-OA} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{OA}_i = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FN}+\mathrm{FP}}\right] \tag{18}$$

where TP, TN FN, and FP are the main elements of confusion matrix, and M is the temporal number of the dataset.

3) Implementation Details and Comparative Methods: For a fair exploration of the native representation capabilities, we remove all the pretrained weights from the selected methods. We would like to declare that although our downstream classifier employs a well-trained encoder, no labels are involved during the self-supervised stage. The model is motivated to automatically extract the latent features without any supervision from labels.

FC-EF: A single-branch architecture with fully convolutional U-Net [41].

FC-Siam-Conc: A dual-branch fully convolution network with feature concatenation during the feature fusion stage [41].

FC-Siam-Diff: A dual-branch CD network with a difference map calculation module before feature fusion [41].

SNUNet-CD: Modified U-Net++ structure for bitemporal remote sensing images with attention mechanism utilized in the decoder [43].

USSFCNet: A lightweight design for effective CD. Spatial–spectral feature interaction strategy is adopted to feature extraction process [46].

A2Net: A lightweight method which identifies changes by progressive aggregation and supervised attention [66].

SEIFNet: A Siamese CD network with spatiotemporal enhancement and interlevel fusion [67].

All training and testing are performed on an NVIDIA RTX 3090 GPU. Each baseline is trained with a batch size of 8 for 100 epochs and optimized by Adam optimizer with momentum of 0.999 and weight decay of 1e-2. The initial learning rate is set to 1e-3, which decreases to 90% every eight epochs.

B. Quantitative Analysis

In this part, we perform comparative experiments on the three time series subsets of TSCD dataset listed in Table II.

TABLE IV
COMPARISON RESULTS ON THE *TSCD-2018–2020* DATASET

Dataset	Method	F1-Score	IoU	OA
TSCD-2018-2020	FC-EF	31.97	18.90	97.75
	FC-Siam-Conc	13.35	7.15	97.79
	FC-Siam-Diff	9.37	4.91	97.54
	SNUNet-CD	47.17	30.86	98.06
	USSFCNet	37.50	23.08	97.90
	A2Net	35.31	21.44	97.98
	SEIFNet	45.91	29.79	98.13
	Contrast-COUD	**48.45**	**31.97**	**98.25**
	TS-COUD	**50.52**	**33.80**	**98.21**

TABLE V
COMPARISON RESULTS ON THE *TSCD-2020–2022* DATASET

Dataset	Method	F1-Score	IoU	OA
TSCD-2020-2022	FC-EF	66.99	50.37	95.41
	FC-Siam-Conc	55.65	38.55	94.79
	FC-Siam-Diff	51.87	35.01	94.29
	SNUNet-CD	73.99	58.72	96.29
	USSFCNet	70.52	54.46	95.87
	A2Net	64.49	47.59	95.34
	SEIFNet	70.40	54.32	96.06
	Contrast-COUD	**75.21**	**60.27**	**96.52**
	TS-COUD	**76.07**	**61.38**	**96.90**

The experimental results of the comparison are shown in the Tables III–V.

From the results on TSCD-2016–2018 subset shown in Table III, we observe that the proposed Contrast-COUD and TS-COUD methods achieve relative higher performance than the selected CD baselines in the two main metrics of $F1$-score and IoU. In terms of OA metric, COUD also outperforms all the baselines.

It should be noted that the TSCD-2016–2018 is the first changing time phase of TSCD dataset, so we remove the temporal distillation in the training of CD. Only the pretrained COUD encoder is preserved for feature refinement. Despite the absence of distillation weights from the pretemporal phase, the proposed COUD method still obtains a superiority of 1.38% and 1.03% over the comparative baseline in the $F1$ and IoU metrics.

On the TSCD-2018–2020 subset, TS-COUD achieves the optimal performance in $F1$ and IoU metrics. Specifically, TS-COUD and Contrast-COUD receive joint guidance from the distillation weights of the previous 2016–2018 time phase. By vertical comparison, we observe that the best method surpasses the comparative method by 3.35% and 2.94% in the $F1$ and IoU metrics, which is a further improvement compared to the performance in the previous time phase.

TABLE VI
PERFORMANCE OF TEMPORAL METRICS ON THE TSCD DATASET

Dataset	Method	T-F1	T-IoU	T-OA
TSCD	FC-EF	53.39	37.86	97.03
	FC-Siam-Conc	41.51	28.05	96.60
	FC-Siam-Diff	39.65	26.83	96.48
	SNUNet-CD	63.22	47.22	97.61
	USSFCNet	55.68	39.80	97.30
	A2Net	53.16	37.19	97.14
	SEIFNet	60.37	44.01	97.41
	Contrast-COUD	**64.35**	**48.45**	**97.72**
	TS-COUD	**65.24**	**49.33**	**97.90**

In addition, we notice a significant decay in the performance of all baselines. The reason for such decay is the presence of a certain thickness and area of cloud cover in the original image of the TSCD-2018–2020 subset, which raises a great challenge to the feature extraction capability of the baseline. Therefore, we consider this subset as hard samples for exploratory evaluation. As demonstrated by the results, despite the extreme case of difficult sample detection, the COUD series methods are still able to achieve acceptable performance.

On the TSCD-2020–2022 subset, TS-COUD outperforms the comparative baseline by 2.08% and 2.66% in the $F1$ and IoU metrics. Along the timeline, we can observe an positive trend in the performance improvement of COUD series methods relative to the suboptimal comparative baseline. This is partially due to the fact that in temporal distillation, the two previous changing time phases provide weighted gains over the current time phase, making our proposed methods more concerned with the temporal context during feature extraction. Therefore, the COUD series methods are naturally suitable for CD tasks with time series data.

Meanwhile, the pretrained encoder could provide constant guidance on feature refinement without any labels. For continual CD in urbanization process, the emergence of massive unlabeled data can be directly pretrained with the COUD series methods, thereby leveraging limited labeled data for downstream fine-tuning. Compared to the high time cost of "full labeling, full supervised training" in conventional CD, the two-stage paradigm of COUD is an advanced endeavor.

Performance on temporal related metrics: According to the three temporal related metrics introduced in Section IV-A2, an additional temporal evaluation is performed through the TSCD dataset. Compared to the conventional metrics (e.g., $F1$andIoU) utilized in bitemporal CD tasks, we believe the temporalized metrics (T-$F1$, T-IoU, and T-OA) reflect the temporal correlation between time phases, which is therefore more applicable to the time series CD tasks. The experimental results from Table VI illustrate that the COUD series methods could capture the temporal features to improve the performance in continual vision tasks.

From the perspective of model architecture, SEIFNet [67] and SNUNet-CD [43] adopt lightweight design, which effectively

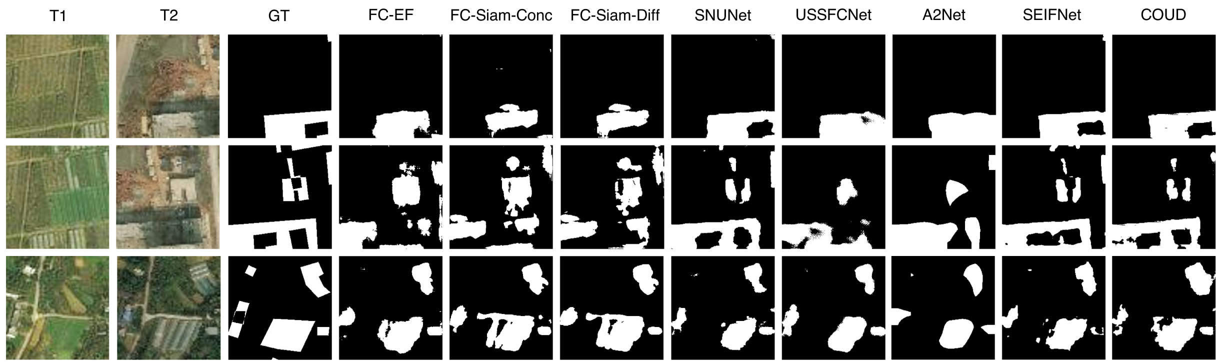

Fig. 7. Visualization of the comparative experiments on TSCD-2016–2018 subset.

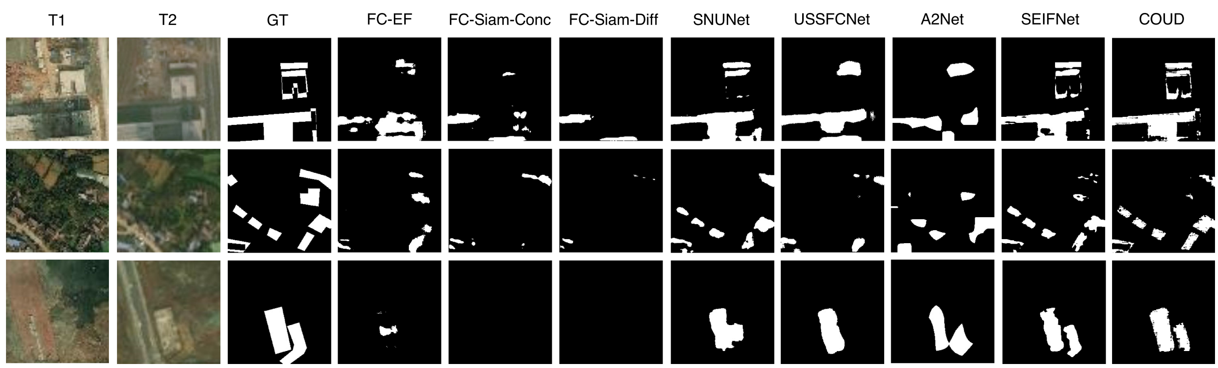

Fig. 8. Visualization of the comparative experiments on TSCD-2018–2020 subset.

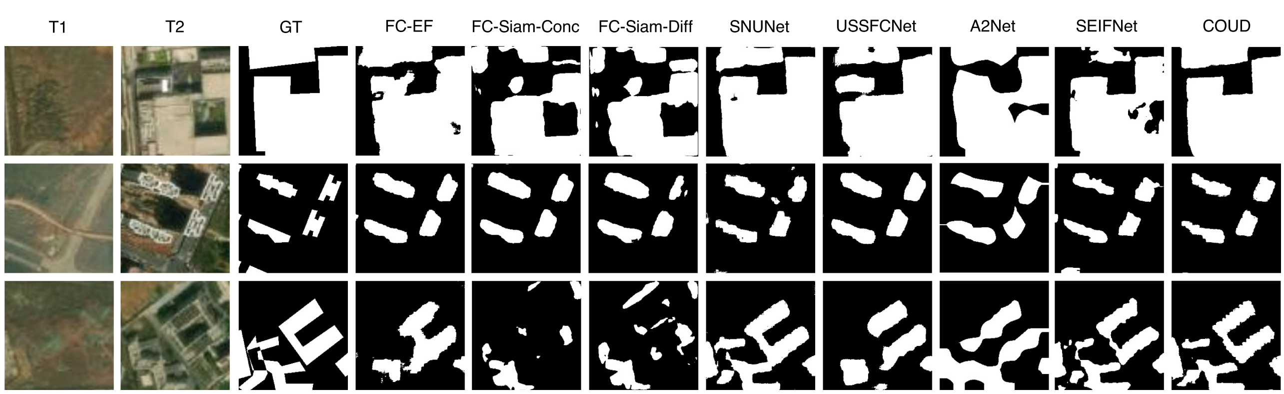

Fig. 9. Visualization of the comparative experiments on TSCD-2020–2022 subset.

reduces the model parameter counts while achieving satisfying performance. In contrast, our proposed COUD method contains two separate stages, i.e., self-supervised pretraining and downstream fine-tuning. As such, the scale of the underlying parameters in our COUD model is relatively large than that of other comparative methods. The experimental analysis about computational cost will be discussed in Section V. Discussion—V-D.

C. Visualization Analysis

We visualize the detection results for all CD baselines on the three subsets of TSCD for more intuitive discrimination. The visualization results are presented in Figs. 7–9, respectively. The selected experimental scenarios mainly contain urban buildings and cultivated land. Through the comparison, we find that the proposed COUD method is more competitive in capturing the changing region, especially in the segmentation for the geometric features of the object's boundaries.

The vertical comparison of the three time series subsets demonstrates that the COUD method exhibits more accurate performance in the object segmentation with an increasing time phase, which confirms the effectiveness of the proposed temporal distillation approach.

Furthermore, we notice that despite the large background and illumination variations between T_1 and T_2 embeddings, COUD can still stably extract the change region in contrast to

TABLE VII
ABLATION EXPERIMENTS OF MODEL ARCHITECTURE ON TSCD-2020–2022 DATASET

Method	+ COUD Enc.	+ T.-D.	F1-Score	IoU
TS-COUD	✗	✗	73.99	58.72
	✓	✗	74.24	59.04
	✗	✓	74.81	59.76
	✓	✓	**76.07**	**61.38**

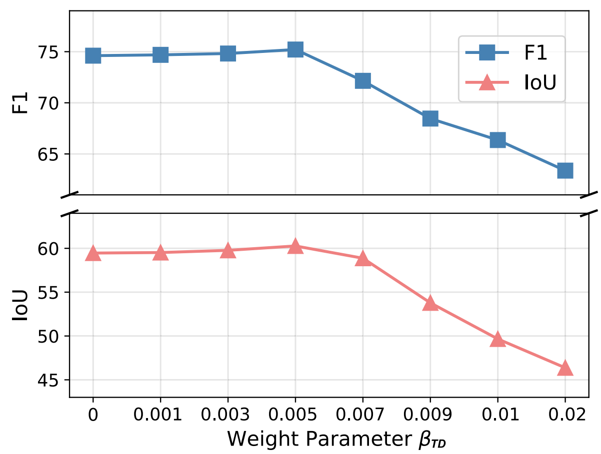

Fig. 10. Performances of different parameter $\beta_{\rm TD}$ in temporal distillation.

other baselines. Since the pretrained encoder receives numerous patches with distinct styles for discrimination in the pretraining stage, which introduces robustness against the background and illumination variations, thereby providing continual guidance for feature refinement.

V. DISCUSSION

In this section, we further investigate the main components and configurations of the proposed COUD method.

A. Ablation Experiments

In this part, we perform several ablation experiments to verify the effectiveness of the key components in the proposed method. The ablation experiments contain two main parts as follows.

1) *Classifier architecture*: To evaluate the performance of pretrained COUD encoder and temporal distillation, we discard the two components in fine-tuning, respectively. The corresponding performance are reported in the Table VII. It is proved that the application of both COUD encoder and temporal distillation provide benefits to the COUD framework.
2) *Parameter β_{TD} in temporal distillation:* The effect of the weight parameter β is also explored and the specific results are illustrated in Fig. 10. Through the comparative experiment, the optimal parameter $\beta_{\rm TD}$ is set as 0.005.
3) *Effect of the COUD pretraining strategy:* In Table VIII, we conduct ablation study to analyze the effect of COUD pretraining strategy on downstream classifier, by removing one of the masking and temporal branches and pretraining the COUD framework. The experimental results demonstrate the utilization of both branches can effectively boost the classifier performance.

TABLE VIII
ABLATION STUDY OF COUD PRETRAINING STRATEGY (MASK. AND TEMPO. ARE THE MASKING BRANCH AND TEMPORAL BRANCH MENTIONED IN SECTION III-B2)

Method	+ Mask.	+ Tempo.	F1-Score	IoU
COUD	✗	✓	75.21	60.27
	✓	✗	75.78	61.01
	✓	✓	76.07	61.38

TABLE IX
THE TIME COST OF COUD PRE-TRAINING, FINE-TUNING, AND INFERENCE.

Stage	Total Time	Epochs	Data Volume	Speed
Pre-training	4.45 h	200	12812	160 FPS
Fine-tuning	2.08 h	100	900	12 FPS
Inference	11 s	1	128	11.6 FPS

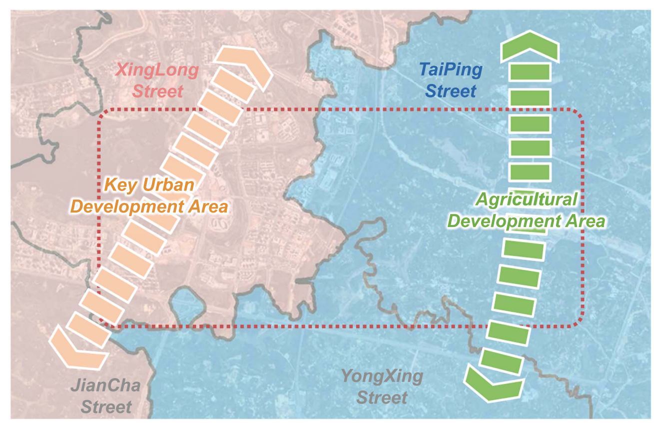

Fig. 11. Illustration of the overall development plan for Chengdu. The red dotted box shows the approximate geographical scope of this study area.

B. Time Series Analysis of Region-Based Change

To achieve a detailed overview of the global urbanization process in the study area, a time series analysis of region-based change is performed. We set 128 m as the size of grid cell and partition the whole study area into a 15 × 24 region matrix. On the basis of the predictions of the COUD method for full-region change evaluation, we calculate the changing ratio of the building complex in each cell, thereby present the region-based change map according to the statistics.

According to the overall development plan for Chengdu [68] illustrated in Fig. 11, the study area in this article spans two areas: "Xinglong Street" and "Taiping Street," which are assigned the development goals of "key urban development area" and "agricultural development area," respectively.

From the region-based change map shown in Fig. 12, the major changes between 2016 and 2018 are concentrated in the

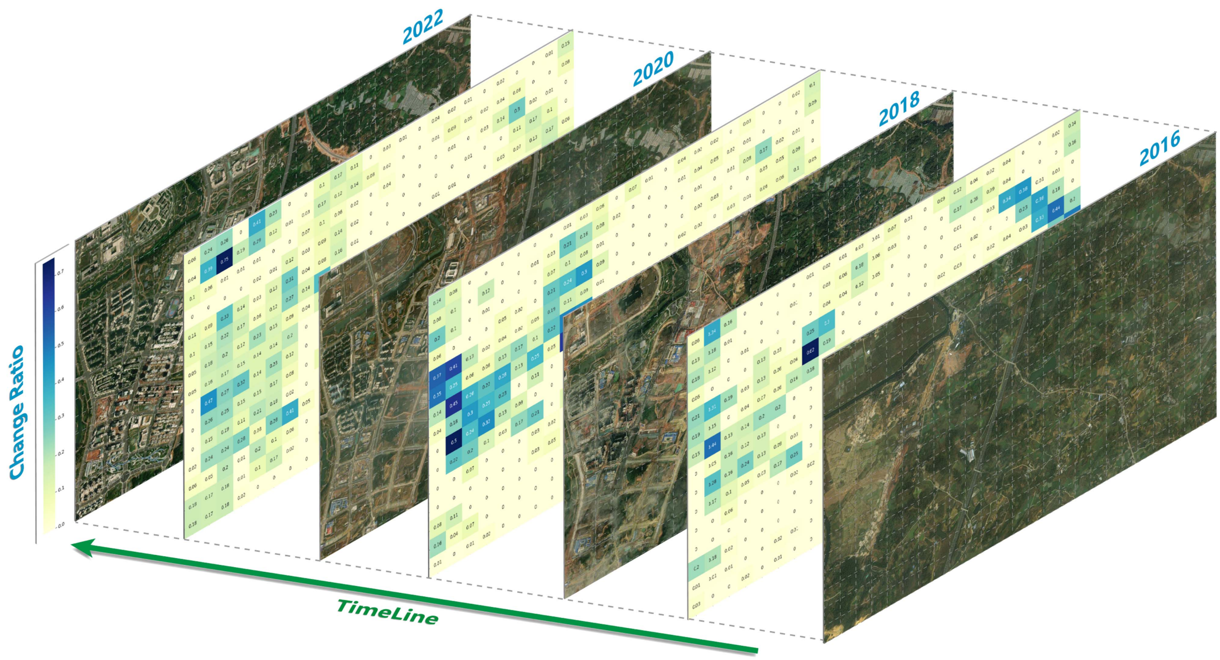

Fig. 12. Region-based change maps from 2016 to 2022.

western and northeastern parts of the study area. In the western part, several residential buildings and manufactories have been constructed, where the highest change ratio in the grid cells reaches 62%. In the northeast regions, the presence of numerous new planting greenhouses confirms the development plan for agriculture in the official documents.

It is also illustrated that the urbanization process in the western of the study area rapidly increased from 2018 to 2020, and the emergence of massive building complexes and infrastructures are highlighted in the dark blue cells in the region-based change map. Meanwhile, subtle changes have occurred in some cultivated regions, mainly reflected in the alternation of crop types.

From 2020 to 2022, we notice a boom in the number of buildings in the southwestern of the study area, where multiple residential settlements are under construction. In the meantime, changes in the eastern agricultural development area show scattered characteristics, with numerous grid cells experiencing changes to a certain extent. Specifically, such changes are manifested in the reclamation of new agricultural land. Moreover, a newly constructed high-speed railway appears in the northeast of the study area, leading to the change in spatial and temporal distribution of the cropland.

Overall, from 2016 to 2022, the development orientation in the study area gradually shifts southward, accompanied by a simultaneous development in urbanization and agriculture.

In addition, we display two regions to explore the continual change trends from 2016 to 2022. As shown in the Fig. 13(a), a building complex is gradually being constructed. In Fig. 13(b), a large shopping mall is emerging in the left of the scenario. We notice that despite the significant background and illumination differences between the time series original embeddings, the COUD method is still able to accurately localize the changed building objects and extract the corresponding changed pixels.

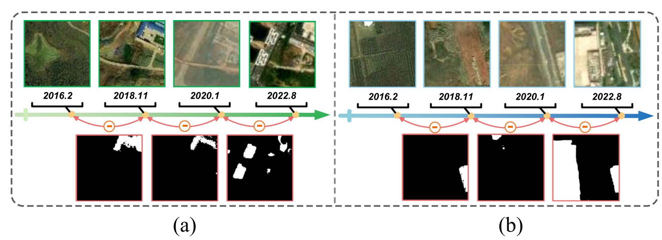

Fig. 13. Time series visualization of changed objects from 2016 to 2022.

C. Multilevel Change Ratio Evaluation

We conduct a multilevel evaluation of the building object evolution in the whole study area, in which the quantitative classification is performed by establishing various criteria. On the basis of the change ratio calculated by the previous section, the change status of each time phase is classified into the following four levels.

Lv.1. No change: ratio $= 0\%$.

Lv.2. Minor change: $0\% <$ ratio $\leqslant 10\%$.

Lv.3. Moderate change: $10\% <$ ratio $\leqslant 30\%$.

Lv.4. Major change: ratio $> 30\%$.

The statistical results shown in Fig. 14 indicate that among the three change time phases, the "*no change*" and "*minor change*" levels account for the majority of change situations, while "*moderate change*" and "*major change*" occupy a relatively minor proportion. By horizontal comparison, the first two change levels dominate significantly within the 2016–2018 and 2018–2020 time phases. Combined with the region-based change maps in Figs. 12, the changes in these two periods mostly show concentration characteristics, which are reflected in the concentrated construction of buildings or planting greenhouses. And in the last period of 2020–2022, the percentage of "*moderate change*" rises

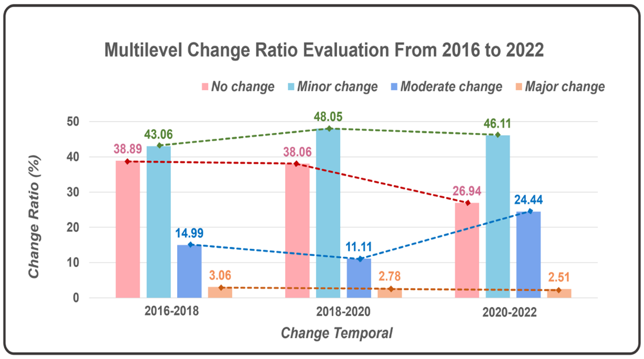

Fig. 14. Multilevel change ratio evaluation from 2016 to 2022.

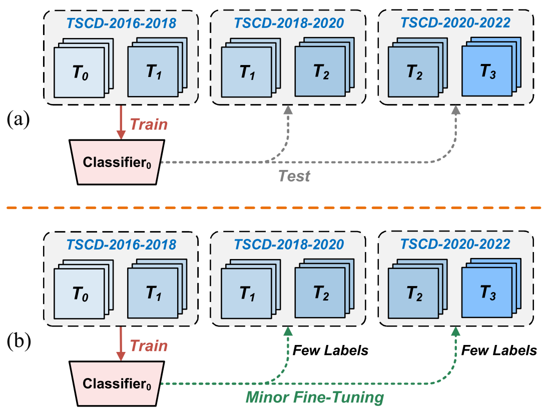

Fig. 15. Illustration of the two evaluation paradigms on temporal generalization. (a) Classifier is directly transferred to the other subsets in TSCD dataset. (b) Classifier will be evaluated after minor fine-tuning with few labels from the target subsets.

sharply to 24.44%, which corresponds to the new construction of extensive building complexes appearing in Fig. 12.

D. Model Computational Cost

a) *Time cost:* The time cost of the pretraining, fine-tuning, and inference stages are tested on the NVIDIA RTX 3090 GPU, respectively, listed in Table IX.
1) Pretraining stage: In this stage, we use part of the SpaceNet-7 dataset to optimize the model for 200 epochs. From Table IX, the pretraining will cost about 4.45 h with the speed of 160 FPS, which is basically acceptable. Moreover, due to the two-stage architecture of COUD, the pre-training is allowed to be performed in an offline mode.
2) Fine-tuning and inference stages: Here, the classifier update the parameters with triplet guidance: i) Pretrained encoder; ii) Previous T_{i-1} classifier; iii) Backward propagation. Therefore, the fine-tuning efficiency might be slightly affected (e.g., 12 FPS), but the total optimization time still remains at a relatively low level of 2.08 h. In addition, the 11 s inference process can satisfy the requirement in real-world scenarios.

TABLE X
INTERTEMPORAL CROSS-VALIDATION ON THE CLASSIFIER TRAINED ON THE **TSCD-2016–2018** SUBSET

Testing Subset	Testing Paradigm	Method	F1	IoU
TSCD-2018-2020	Direct Transferring	SNUNet-CD	47.17 → 17.27	30.86 → 9.45
		Δ *(Decay)* ↓	-29.90	-21.41
		COUD	48.45 → **24.57**	31.97 → **14.01**
		Δ *(Decay)* ↓	-23.88	-17.96
	Minor Fine-Tuning	SNUNet-CD	17.27 → 29.18	9.45 → 17.08
		Δ *(Gain)* ↑	+11.91	+7.63
		COUD	24.57 → **45.19**	14.01 → **29.19**
		Δ *(Gain)* ↑	+20.62	+15.18
TSCD-2020-2022	Direct Transferring	SNUNet-CD	73.99 → 27.34	58.72 → 15.83
		Δ *(Decay)* ↓	-46.65	-42.89
		COUD	75.21 → **39.43**	60.27 → **24.55**
		Δ *(Decay)* ↓	-35.78	-35.72
	Minor Fine-Tuning	SNUNet-CD	27.34 → 64.74	15.83 → 47.87
		Δ *(Gain)* ↑	+37.40	+32.04
		COUD	39.43 → **71.17**	24.55 → **55.24**
		Δ *(Gain)* ↑	+31.74	+30.69

b) *Model parameter counts:* The parameter counts of the total COUD framework are evaluated to be 468.25M. Specifically, the majority of the parameter counts (i.e., 403.01M) are contained in the pretraining stage, while the fine-tuning classifier only accounts for a minor portion (i.e., 65.18M). Although the pretrained encoder is reused for downstream fine-tuning, the parameters are frozen in this stage. As such, the learnable parameters of the downstream classifier could be optimized with high efficiency. It is worth noting that the COUD framework adopts the ViT basic blocks, such that its underlying parameter counts might be larger than that of the related works in CNN architecture. Meanwhile, inspired by SEIFNet [67], the lightweight COUD framework could be the future research.

E. Exploration on Model Generalization

Considering the model generalization capability as an essential scalability for deep learning-based methods, we additionally explore the temporal generalization of the proposed COUD method based on the TSCD time series dataset. Ideal method is expected to possess strong robustness to temporal changes. In this article, the temporal generalization is primarily reflected in the intertemporal cross-validation performance on unseen samples between the three subsets of TSCD.

Despite the fact that for data-driven methods, direct transferring to an unknown dataset typically leads to unavoidable performance degradation, we still record the degradation extent. With consideration of the self-supervised paradigm adopted by the proposed method, we also investigate the adaptability on unknown dataset by performing minor fine-tuning for comprehensive evaluation on the model generalization. According to the results in Section IV-B, we select the suboptimal method SNUNet-CD [43] for comparison.

The results of intertemporal cross-validation are listed in the Tables X–XII. The comparative results are divided into two portions: performance decay and fine-tuning gain. The decay

TABLE XI
INTERTEMPORAL CROSS-VALIDATION
ON THE CLASSIFIER TRAINED ON THE **TSCD-2018–2020** SUBSET

Testing Subset	Testing Paradigm	Method	F1	IoU
TSCD-2016-2018	Direct Transferring	SNUNet-CD	68.50 → 25.96	52.09 → 14.91
		Δ (*Decay*) ↓	-42.54	-37.18
		COUD	69.38 → **39.10**	53.12 → **24.30**
		Δ (*Decay*) ↓	-30.28	-28.82
	Minor Fine-Tuning	SNUNet-CD	25.96 → 53.66	14.91 → 36.67
		Δ (*Gain*) ↑	+27.70	+21.76
		COUD	39.10 → **60.39**	24.30 → **43.26**
		Δ (*Gain*) ↑	+21.29	+18.96
TSCD-2020-2022	Direct Transferring	SNUNet-CD	73.99 → 45.12	58.72 → 29.13
		Δ (*Decay*) ↓	-28.87	-29.59
		COUD	75.21 → **50.26**	60.27 → **33.56**
		Δ (*Decay*) ↓	-24.95	-26.71
	Minor Fine-Tuning	SNUNet-CD	45.12 → 70.79	29.13 → 54.78
		Δ (*Gain*) ↑	+25.67	+25.65
		COUD	50.26 → **72.71**	33.56 → **57.12**
		Δ (*Gain*) ↑	+22.45	+23.56

TABLE XII
INTERTEMPORAL CROSS-VALIDATION
ON THE CLASSIFIER TRAINED ON THE **TSCD-2020–2022** SUBSET

Testing Subset	Testing Paradigm	Method	F1	IoU
TSCD-2016-2018	Direct Transferring	SNUNet-CD	68.50 → 30.62	52.09 → 18.08
		Δ (*Decay*) ↓	-37.88	-34.01
		COUD	69.38 → **43.43**	53.12 → **27.74**
		Δ (*Decay*) ↓	-25.95	-25.38
	Minor Fine-Tuning	SNUNet-CD	30.62 → 60.57	18.08 → 43.44
		Δ (*Gain*) ↑	+29.95	+25.36
		COUD	43.43 → **62.66**	27.74 → **45.63**
		Δ (*Gain*) ↑	+19.23	+17.89
TSCD-2018-2020	Direct Transferring	SNUNet-CD	47.17 → 21.87	30.86 → 12.28
		Δ (*Decay*) ↓	-25.30	-18.58
		COUD	48.45 → **27.02**	31.97 → **15.61**
		Δ (*Decay*) ↓	-21.43	-16.36
	Minor Fine-Tuning	SNUNet-CD	21.87 → 40.63	12.28 → 25.49
		Δ (*Gain*) ↑	+18.76	+13.21
		COUD	27.02 → **46.38**	15.61 → **30.19**
		Δ (*Gain*) ↑	+19.36	+14.58

value reflects the performance maintenance capability when the model is directly transferred to other unseen temporal subsets. In contrast, the gain value indicates the capability of efficient performance reconstruction when minor fine-tuning is performed on the model.

The proposed COUD method exhibits lower performance decay on unseen temporal samples compared to the existing methods. Benefiting from the temporal context of the temporal distillation approach, the COUD possesses superior performance maintenance capability. In terms of fine-tuning gain, we perform minor fine-tuning for 20 epochs to observe the performance reconstruction. Experimental results demonstrate that the COUD method is able to rapidly restore and approach to the original performance without heavy training. In summary, the proposed COUD method could provide satisfying temporal generalization when dealing with time series data.

VI. CONCLUSION

In this article, we proposed a COUD to handle the CD task in time series remote sensing images. The proposed two-stage method contains the pretraining of COUD framework and the temporal distillation supported fine-tuning. We also release a challenging TSCD dataset covering Chengdu region of China, fulfilling the absence of time series CD dataset. The experimental results demonstrate that our proposed method is capable of leveraging the temporal pattern in time series CD task. Comprehensive evaluation related to the actual planning documents released by the management department on the selected study area demonstrates the effectiveness and exactness of the COUD method.

We will further explore robust methods under extreme time series conditions of occlusion and shadowing, as well as multitemporal interaction and model lightweight compression. We expect that the COUD method could provide an efficient tool to the administrative departments for urbanization monitoring in the future.

REFERENCES

[1] Z. Wang, Y. Gao, X. Wang, Q. Lin, and L. Li, "A new approach to land use optimization and simulation considering urban development sustainability: A case study of Bortala, China," *Sustain. Cities Soc.*, vol. 87, 2022, Art. no. 104135.

[2] X. Huang, Y. Cao, and J. Li, "An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images," *Remote Sens. Environ.*, vol. 244, 2020, Art. no. 111802.

[3] C. Lv, B. Bian, C.-C. Lee, and Z. He, "Regional gap and the trend of green finance development in China," *Energy Econ.*, vol. 102, 2021, Art. no. 105476.

[4] B. Yu, "Ecological effects of new-type urbanization in China," *Renewable Sustain. Energ. Rev.*, vol. 135, 2021, Art. no. 110239.

[5] B. Yang, T. Xu, and L. Shi, "Analysis on sustainable urban development levels and trends in China's cities," *J. Cleaner Prod.*, vol. 141, pp. 868–880, 2017.

[6] S. Wu, "Smart cities and urban household carbon emissions: A perspective on smart city development policy in China," *J. Cleaner Prod.*, vol. 373, 2022, Art. no. 133877.

[7] C. Yu et al., "Regional integration and city-level energy efficiency: Evidence from China," *Sustain. Cities Soc.*, vol. 88, 2023, Art. no. 104285.

[8] T. Huo, X. Li, W. Cai, J. Zuo, F. Jia, and H. Wei, "Exploring the impact of urbanization on urban building carbon emissions in China: Evidence from a provincial panel data model," *Sustain. Cities Soc.*, vol. 56, 2020, Art. no. 102068.

[9] L. Fang, J. Huang, Z. Zhang, and V. Nitivattananon, "Data-driven framework for delineating urban population dynamic patterns: Case study on Xiamen Island, China," *Sustain. Cities Soc.*, vol. 62, 2020, Art. no. 102365.

[10] H. Chen et al., "A new method for building-level population estimation by integrating LiDAR, nighttime light, and POI data," *J. Remote Sens.*, vol. 2021, pp. 1–17, 2021.

[11] W. Shan et al., "Ecological environment quality assessment based on remote sensing data for land consolidation," *J. Cleaner Prod.*, vol. 239, 2019, Art. no. 118126.

[12] C. Yang, W. Zeng, and X. Yang, "Coupling coordination evaluation and sustainable development pattern of geo-ecological environment and urbanization in Chongqing municipality, China," *Sustain. Cities Soc.*, vol. 61, 2020, Art. no. 102271.

[13] T. Shi, S. Yang, W. Zhang, and Q. Zhou, "Coupling coordination degree measurement and spatiotemporal heterogeneity between economic development and ecological environment–Empirical evidence from tropical and subtropical regions of China," *J. Cleaner Prod.*, vol. 244, 2020, Art. no. 118739.

[14] H. Ni, L. Yu, P. Gong, X. Li, and J. Zhao, "Urban renewal mapping: A case study in Beijing from 2000 to 2020," *J. Remote Sens.*, vol. 3, 2023, Art. no. 0072.

[15] Y.-C. Li, S. Lei, N. Liu, H.-C. Li, and Q. Du, "IDA-SiamNet: Interactive- and dynamic-aware Siamese network for building change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5628213.

[16] K. Chen, K. Fu, X. Gao, M. Yan, X. Sun, and H. Zhang, "Building extraction from remote sensing images with deep learning in a supervised manner," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1672–1675.

[17] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[18] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.

[19] Z. Du, X. Li, J. Miao, Y. Huang, H. Shen, and L. Zhang, "Concatenated deep-learning framework for multitask change detection of optical and SAR images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 719–731, 2024.

[20] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3633–3643, May 2020.

[21] N. Liu, X. Xu, T. Celik, Z. Gan, and H.-C. Li, "Transformation-invariant network for few-shot object detection in remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5625314.

[22] H. Jiang et al., "A survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1552.

[23] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 871.

[24] X. Zhang, L. Liu, J. Wang, T. Zhao, W. Liu, and X. Chen, "Automated mapping of Global 30-m tidal flats using time-series Landsat imagery: Algorithm and products," *J. Remote Sens.*, vol. 3, 2023, Art. no. 0091.

[25] Y. Zhao, T. Celik, N. Liu, and H.-C. Li, "A comparative analysis of GAN-based methods for SAR-to-optical image translation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 3512605.

[26] P. Guo, T. Celik, N. Liu, and H.-C. Li, "Break through the border restriction of horizontal bounding box for arbitrary-oriented ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4005505.

[27] E. J. Parelius, "A review of deep-learning methods for change detection in multispectral remote sensing images," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 2092.

[28] J. He, Q. Yuan, J. Li, Y. Xiao, and L. Zhang, "A self-supervised remote sensing image fusion framework with dual-stage self-learning and spectral super-resolution injection," *ISPRS J. Photogramm. Remote Sens.*, vol. 204, pp. 131–144, 2023.

[29] R. D. Johnson and E. Kasischke, "Change vector analysis: A technique for the multispectral monitoring of land cover and condition," *Int. J. Remote Sens.*, vol. 19, no. 3, pp. 411–426, 1998.

[30] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[31] J. Ma, M. Gong, and Z. Zhou, "Wavelet fusion on ratio images for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1122–1126, Nov. 2012.

[32] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006.

[33] M. Forkel, N. Carvalhais, J. Verbesselt, M. D. Mahecha, C. S. Neigh, and M. Reichstein, "Trend change detection in NDVI time series: Effects of inter-annual variability and methodology," *Remote Sens.*, vol. 5, no. 5, pp. 2113–2144, 2013.

[34] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.

[35] T. Han, M. A. Wulder, J. C. White, N. C. Coops, M. Alvarez, and C. Butson, "An efficient protocol to process Landsat images for change detection with tasselled cap transformation," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 1, pp. 147–151, Jan. 2007.

[36] J. Deng, K. Wang, Y. Deng, and G. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.

[37] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-Means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.

[38] L. I. Kuncheva and W. J. Faithfull, "PCA feature extraction for change detection in multidimensional unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 69–80, Jan. 2014.

[39] Q. Zhu et al., "Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 184, pp. 63–78, 2022.

[40] Y. Zhao, H.-C. Li, N. Liu, and R. Wang, "Toward distortion-aware change detection in realistic scenarios," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2024, pp. 10319–10322.

[41] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[42] Z. Chenxiao et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, 2020.

[43] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8007805.

[44] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5607514.

[45] H. Xia, Y. Tian, L. Zhang, and S. Li, "A deep Siamese postclassification fusion network for semantic change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5622716.

[46] T. Lei et al., "Ultralightweight spatial–spectral feature cooperation network for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4402114.

[47] Y. Cao and X. Huang, "A full-level fused cross-task transfer learning method for building change detection using noise-robust pretrained networks on crowdsourced labels," *Remote Sens. Environ.*, vol. 284, 2023, Art. no. 113371.

[48] J. Wang, F. Gao, J. Dong, S. Zhang, and Q. Du, "Change detection from synthetic aperture radar images via graph-based knowledge supplement network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 1823–1836, 2022.

[49] C. Han, C. Wu, H. Guo, M. Hu, and H. Chen, "HANet: A hierarchical attention network for change detection with bitemporal very-high-resolution remote sensing images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 3867–3878, 2023.

[50] J. Chen et al., "DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 1194–1206, 2021.

[51] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

[52] O. Manas, A. Lacoste, X. Giró-i Nieto, D. Vazquez, and P. Rodriguez, "Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 9394–9403.

[53] P. Akiva, M. Purri, and M. Leotta, "Self-supervised material and texture representation learning for remote sensing tasks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8193–8205.

[54] H. Noh, J. Ju, M. Seo, J. Park, and D.-G. Choi, "Unsupervised change detection based on image reconstruction loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1351–1360.

[55] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[56] H. Chen, W. Li, S. Chen, and Z. Shi, "Semantic-aware dense representation learning for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5630018.

[57] C. Xu et al., "Rethinking building change detection: Dual-frequency learnable visual encoder with multiscale integration network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 6174–6188, 2024.

[58] F. Jiang, M. Gong, H. Zheng, T. Liu, M. Zhang, and J. Liu, "Self-supervised global–local contrastive learning for fine-grained change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4400613.

[59] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15745–15753.

[60] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 213–247, Dec. 2022.

[61] C. Tao, J. Qi, M. Guo, Q. Zhu, and H. Li, "Self-supervised remote sensing feature learning: Learning paradigms, challenges, and future works," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5610426.

[62] N. Wang, W. Li, R. Tao, and Q. Du, "Graph-based block-level urban change detection using Sentinel-2 time series," *Remote Sens. Environ.*, vol. 274, 2022, Art. no. 112993.

[63] Y. Chen and L. Bruzzone, "Unsupervised CD in satellite image time series by contrastive learning and feature tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5401813.

[64] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Inf. Sci.*, vol. 307, pp. 39–52, 2015.

[65] Y. Zhao, T. Celik, N. Liu, F. Gao, and H.-C. Li, "SSLChange: A self-supervised change detection framework based on domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, 2024, doi: 10.1109/TGRS.2024.3489615.

[66] Z. Li et al., "Lightweight remote sensing change detection with progressive feature aggregation and supervised attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5602812.

[67] Y. Huang, X. Li, Z. Du, and H. Shen, "Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5609414.

[68] StateCouncil, "State council approves overall development plan for Chengdu," 2015. [Online]. Available: https://english.www.gov.cn/policies/latest_releases/2015/12/04/content_281475247840920.htm

Yitao Zhao (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in photogrammetry and remote sensing from Southwest Jiaotong University, Chengdu, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in information and communication engineering with the School of Information Science and Technology, Southwest Jiaotong University.

His research interests include change detection and self-supervised learning.

Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in information and communication engineering from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree in information and communication engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently a Full Professor with the School of Information Science and Technology, Southwest Jiaotong University. From 2013 to 2014, he was a Visiting Scholar with Prof. William J. Emery with the University of Colorado Boulder, Boulder, CO, USA. His research interests include statistical analysis of SAR images, remote sensing image processing, and pattern recognition.

Sen Lei (Member, IEEE) received the B.S. degree in detection, guidance, and control technology and Ph.D. degree in pattern recognition and intelligent systems from Beihang University, Beijing, China, in 2015 and 2021, respectively.

He is currently an Assistant Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. His research interests include computer vision, deep learning, and remote sensing image processing.

Nanqing Liu (Member, IEEE) received the B.Sc. degree in electronic information engineering from the Wuhan University of Science and Technology, Wuhan, China, in 2016. He is currently working toward the Ph.D. degree in information and communication engineering with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China.

His research interests include label-efficient object detection and optical remote-sensing image processing.

Jie Pan received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2000, and the Ph.D. degree in communication and information system from the Chinese Academy of Sciences, Beijing, China, in 2019.

She is currently the Director of the Airborne Remote Sensing Center, Chinese Academy of Sciences. Her research interests include operation and engineering management of airborne remote sensing system.

Turgay Celik (Senior Member, IEEE) received the Ph.D. degree in engineering from the University of Warwick, Coventry, U.K., in 2011.

He is currently a Professor with University of the Witwatersrand, Johannesburg, South Africa. His research interests include signal and image processing, computer vision, machine intelligence, robotics, data science, and remote sensing.

Dr. Celik is an Associate Editor for *IET ELL*, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL), IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (IEEE JSTARS), and *Springer SIVP*.