Adaptive Self-Paced Collaborative and 3-D Adversarial Multitask Network for Semantic Change Detection Using Zhuhai-1 Orbita Hyperspectral Remote Sensing Imagery

Dawei Wen, Xin Huang, *Senior Member, IEEE*, Qiquan Yang, and Jianqin Tang

***Abstract*—In recent years, numerous change detection methodologies have been proposed, with a predominant focus on binary change detection. Furthermore, there exists a paucity of research addressing semantic change detection in scenarios where solely binary change labels are available. This article introduces a multitask network for semantic change detection. First, 3-D ResUnet model is employed to generate initial multitemporal land cover results through postclassification comparison. Subsequently, the multitask network, encompassing two subtasks—binary change detection and multitemporal semantic segmentation—is proposed. Specifically, the shared branch of the network employs 3-D residual blocks to extract joint spectral-spatial features. In the subsequent task-specific branch, a 3-D GAN is incorporated for the binary change detection task to enhance the discrimination ability of latent high-level features for changes. Novel adaptive self-paced learning and certainty-weighted focal loss are proposed for multitemporal semantic segmentation to mitigate adverse effects from noisy semantic labels by considering sample complexity and reliability in the network optimization process. Experiments conducted on the Orbita Hyperspectral dataset in the Xiong'an New Area demonstrate the superior performance of the proposed method, achieving 99.28% and 76.60% for overall accuracy and kappa, respectively. This outperformance is notable when compared to other methods, such as Str4 and Bi-SRNet, showing an increase of 39.82% and 54.17% for kappa. Moreover, comparative experiments on SECOND further confirm the advantage of the proposed method, achieving 54.62% for kappa and outperforming other comparative methods, such as Bi-SRNet (47.61%).**

Manuscript received 9 September 2023; revised 11 November 2023 and 30 November 2023; accepted 20 December 2023. Date of publication 1 January 2024; date of current version 12 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 41901279, and in part by the Science Foundation Research Project of Wuhan Institute of Technology of China under Grant K202239. *(Corresponding author: Jianqin Tang.)*

Dawei Wen and Jianqin Tang are with the School of Computer Science and Engineering, Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China (e-mail: daweiwen_whu@163.com; tangjq@wit.edu.cn).

Xin Huang is with the School of Remote Sensing and Information Engineering, The Hubei Luojia Laboratory, Wuhan University, Wuhan 430079, China, and also with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

Qiquan Yang is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China, and also with the State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Macau 999078, China (e-mail: qiquanyang@tongji.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3348572

***Index Terms*—Change detection, hyperspectral remote sensing, multitask learning, pseudolabel, semantic change detection.**

I. Introduction

CHANGE detection stands as a cornerstone application in remote sensing technology. The rapid evolution of remote sensing technology has empowered the acquisition of high-spatial-resolution hyperspectral imagery (HSI), enabling us to monitor Earth's surface with unprecedented detail and precision. Nevertheless, hyperspectral change detection remains an exceptionally challenging task owing to various intricate factors, such as spectral variability, data redundancy, high dimensionality, limited annotated data, noise, anomalies, and the complexities inherent in real-word scenarios [1].

Over the preceding decades, the evolution of change detection algorithms has closely paralleled the development of machine learning and remote sensing technologies [2]. The initial phase of the change detection development focused on comparing pixel-level spectral or textual values using techniques, such as image differencing [3], change vector analysis [4], principal component analysis [5], and multivariate alteration detection [6]. In the late 20th century, the launch of satellites with very high spatial resolution enabled the capture of finer spatial details. However, this progress often came at the expense of lower spectral separability, given the inherent tradeoff between spatial and spectral resolution, particularly considering the limitations in imaging capabilities of that era. In response to reduced spectral separability, spatial features (e.g., textures, deep features, object-based features, and angular features) and machine learning methods (e.g., support vector machines, random forests, neural networks, K-means, deep learning models) were employed [7].

In recent years, the rapid evolution of remote sensors has made it feasible to acquire high-spatial-resolution HSI. For high-spatial-resolution HSI, both spectral and spatial information are imperative and should be fully exploited. Among various spatial features, deep features have been extensively studied and applied in change detection tasks since these features, tailored for the specific task, are learned directly from data rather than being handcrafted. Consequently, deep learning has achieved state-of-the-art performance in hyperspectral image change detection. For instance, Wang et al. [8] introduced a general 2-D

convolutional neural network (CNN) for hyperspectral image change detection, capable of learning discriminative features at a higher level. In another study, Li et al. [9] embedded a cross-band feature extraction module into a spatial-spectral self-attention module to jointly learn pixel-level spectral differences and spatial correlations of adjacent pixels in HSI change detection. In [10], bitemporal spectral-spatial feature maps were extracted by 3-D convolution layers, and temporal information of these features was learned via convolutional long short-term memory.

Despite the superior performance of deep learning in HSI change detection, it requires a substantial volume of high-quality labeled data, which is time-consuming and labor-intensive [11]. Consequently, we must address scenarios where limited labeled data are available. One promising solution in the literature involves training deep networks with weak supervision information, such as incomplete, inaccurate, and inexact labels [12], [13]. For instance, Liu et al. [14] proposed a semisupervised change detection framework with a multilayer cascade screening strategy that leverages both spatial information and active learning to select highly reliable unlabeled samples to augment the training sets. Gong et al. [15] introduced a semisupervised strategy combing traditional unsupervised and supervised algorithms when labeled data is insufficient. Tong et al. [16] proposed a method tailored to scenarios with limited training samples in a single-time image, involving unsupervised binary change detection, classification of pretime images, classification of posttime images via transfer learning, and generation of multiclass change map through postclassification comparison (PCC).

On the other hand, change detection can be categorized as binary change detection and semantic change detection, contingent upon the final outputs [17]. Despite the rich spectral and spatial information embedded in HSI, previous research has predominantly focused on traditional binary change detection, with limited emphasis on identifying change type, known as semantic change detection. Semantic change detection can be accomplished through PCC or direct multidate classification, provided multitemporal ground reference data are available [1]. However, as mentioned earlier, the availability of such ground reference data is either scarce or nonexistent. Furthermore, open-source hyperspectral image change detection datasets for semantic change detection are exceedingly scarce in terms of data availability. In the domain of remote sensing, land cover information is more easily acquired and accessed compared to temporal information. Even at a global scale, there are several existing land cover products, e.g., the global land-cover product with fine classification system at 30- [18] and 10-m finer resolution observation and monitoring of the global land cover (FROM-GLC) [19]. However, the generated land cover information is noisy and monotemporal. To date, there has been relatively few research works into semantic change detection in scenario where only noisy monotemporal land cover labels and binary change labels are available, which is of great importance to promote practical applications of change detection.

Therefore, the objective of this study is to present a semantic change detection method based on a binary change detection dataset and a monotemporal land-cover map. Under this scenario, we propose a novel multitask change detection network based on adversarial learning and self-paced learning (MASNet). Initially, using monotemporal land cover labels, a 3-D Residual UNet (ResUNet) model is adopted to generate initial multitemporal land cover results with certainty scores (cs) via PCC. Subsequently, to exploit the potential of noisy land-cover labels and true binary change labels, a multitask paradigm that integrates binary change detection and multitemporal semantic segmentation tasks is proposed. A 3-D generative adversarial network (GAN) for binary change detection task and self-paced learning for semantic change detection task are integrated into this multitask paradigm. These designs guide the network in learning classwise change information, even in the presence of inaccurate pseudolabels. Comprehensive comparisons and ablations studies have been carried out to demonstrate the effectiveness of the proposed method.

In summary, the novelties of this article can be concluded as follows.

1) We present a novel MASNet tailored for semantic change detection. This new method only uses binary change samples and a monotemporal land-cover map. To the best of our knowledge, MASNet is the first end-to-end multitask network designed for semantic change detection leveraging a binary change detection dataset. This innovation holds substantial practical values in real-world applications.
2) MASNet includes three modules: A 3-D encoder to extract joint spectral-spatial features, a binary change detection branch with a 3-D discriminator to improve the discrimination of change in high-level latent features, and a semantic segmentation branch with novel adaptive self-paced learning (ASPL) and certainty-weighted focal loss. Under this multitask framework, true binary change information and noisy PCC results are learned in an end-to-end manner, boosting the reported accuracy of semantic change detection. In addition, both sample complexity and reliability are considered in the network optimization process.

The rest of this article is organized as follows. Section II analyzes related work, including multitask change detection and self-paced learning. In Section III, the proposed method is described in detail. Section IV introduces the experimental datasets, implementation details, and comparative methods. Subsequently, results and analysis are provided in Section V. The discussions of results are presented in Section VI. Finally, Section VII concludes this article.

II. Related Work

A. Semantic Change Detection

The conventional semantic change detection methods involve the application of PCC or direct classification techniques. In the context of PCC, the key points are generating highly accurate classification outputs for individual monotemporal images, contingent upon the availability of multitemporal land cover or land use samples. However, the accumulation of errors poses a significant challenge for PCC-based semantic change detection [20]. On the other hand, the direct classification methods treat change

detection as a classification problem with semantic change types as the desired output [21]. Nevertheless, the presence of extreme sample imbalance within semantic change types presents a formidable obstacle for direct classification methods [22].

A recent advancement in the field of semantic change detection involves the adoption of multitask learning, which facilitates the synthesize of domain-specific information from related tasks, resulting in enhanced generalizability and improved performance [23]. In the domain of semantic change detection, multitask structures have gained prominence and are presently considered the prevailing approach [24]. For instance, in a study by Zhao et al. [25], a change detection model coupled with domain adaption as an auxiliary task was proposed to mitigate the impact of domain shifts and irrelevant changes. Similarly, in a study focused on building change detection, the change detection loss function is constrained by auxiliary building detection tasks, thereby leveraging intrinsically correlated features within building footprints for the detection of fine-grained building changes [26]. Moreover, in a change detection framework utilizing time-series images, an additional decoding branch was added to conduct semantic segmentation on the available semantic categories from different input dates [27]. Notably, Caye Daudt et al. [28] introduced a multitask semantic change detection method based on two fully CNNs. These CNNs were trained for binary change detection and land-cover mapping, respectively, and the land cover related features were incorporated into the binary change detection task. Recent works, such as those in [29], [30], [31], [32], also decoupled semantic change detection as two tasks, namely classification and binary change detection, and have demonstrated the effectiveness of this paradigm. Nonetheless, it is important to note that collecting labeled data for semantic changes is typically more challenging compared to binary change data. Consequently, the practical semantic change detection applications are very limited. Therefore, the development of an effective multitask semantic change detection method tailored to binary change detection datasets is imperative.

B. Self-Paced Learning

Self-paced learning is a form of joint learning that exerts control over the training process by gradually incorporating increasingly complex samples, starting from the simplest ones [33]. Self-paced learning is useful in various types of computer vision and pattern recognition tasks, including long-term tracking [34], natural language processing [35], image classification [36]. A theoretical study proved that self-paced learning is robust to noisy samples and can address local optimum problem [37]. Its application has extended to the domain of change detection. For example, Li et al. [38] incorporated self-paced learning into a convolutional network to address change detection in heterogeneous images. Shang et al. [39] proposed an unsupervised algorithm aiming at change detection in synthetic aperture radar images based on self-paced learning. In addition, many variations of self-paced learning have been proposed and applied in change detection. Gong et al. [40] proposed a group self-paced learning for unsupervised change detection, wherein sample complexity is determined based on loss values, and group information is integrated to prevent the inclusion of training samples coming from the homogeneous regions [33]. Li et al. [41] presented a cost-sensitive self-paced learning for change detection using time-series images, where the cost-sensitive strategy involved assigning different cost values to false-negative and false-positive errors. In these methods, sample complexity based on loss values is utilized as a criterion deciding whether each sample is selected or not. However, this strategy is not applicable well in samples without true labels since label errors are not considered in the training process. Therefore, sample reliability is incorporated in the self-paced learning process to deal with noisy samples.

III. Methodology

The proposed approach aims to generate a semantic change detection result using bitemporal images $IM(t_1)$ and $IM(t_2)$, a monotemporal land-cover map Y, and binary change samples. It consists of two main components: a multitemporal classification network to extract pseudomultitemporal semantic labels and a multitask segmentor for binary change detection and multitemporal semantic segmentation.

A. Multitemporal Classification Network

The ResUNet is a relatively straightforward yet highly effective encoder–decoder structure commonly used for semantic segmentation tasks. Incorporating residual blocks within the ResUNet architecture promotes training stability by addressing the issue of vanishing gradients in deep convolutional networks [42]. Moreover, the inclusion of residual connections plays a crucial role in preserving spatial resolution and retaining high-frequency information from prior layers [43]. Consequently, this network architecture serves as the fundamental framework for multitemporal classification.

However, when dealing with HSI, the conventional 2-D version of the ResUNet fails to fully exploit the inherent interdependencies present between spatial dimensions and spectrum bands. This limitation arises from the fact that HSI is presented as a 3-D hypercube, characterized by both spatial and spectral continuity [44]. In this research article, we propose the utilization of a 3-D ResUNet to fully harness the joint spectral-spatial features present in high-resolution hyperspectral images. By doing so, we aim to overcome the limitations of the traditional 2-D ResUNet in effectively capturing the complex relationships within HSI data.

Furthermore, to mitigate the potential impact of varying imaging conditions of bitemporal images (e.g., differences in illumination condition and season effects), as well as the temporal gap between IM(t_1)/IM(t_2) and Y, PCC is employed. Specifically, bitemporal images are classified separately, and the resulting classifications are subsequently compared to generate pseudomultitemporal semantic labels [45]. Unfortunately, PCC is susceptible to the inherent issue of errors accumulation. To address this concern, we introduce the concept of confidence (c) for pseudomultitemporal samples, which is computed by multiplying the cs associated with bitemporal classifications.

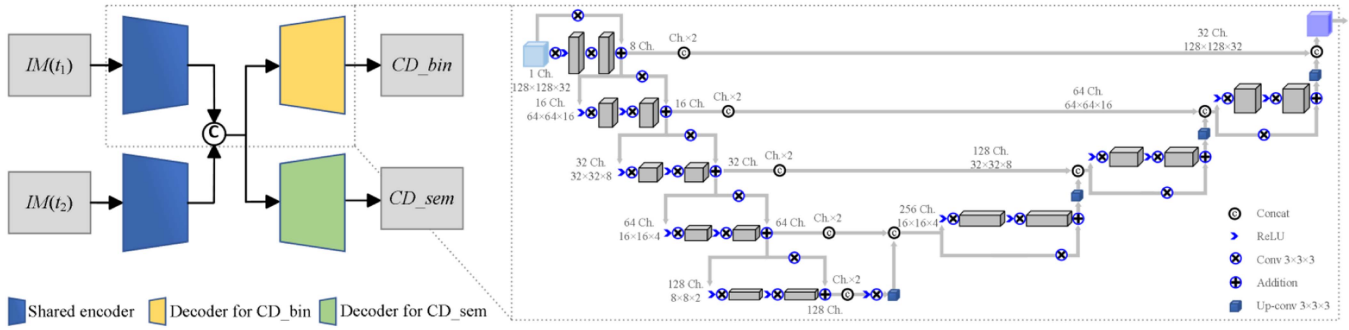

Fig. 1. Multitask segmentor for binary change detection and multitemporal semantic segmentation.

This measure of confidence is then utilized to assign weights to individual samples.

The cs for pixel j in IM(t_i) ($i = 1, 2$) classification are calculated as follows:

$$\mathrm{cs}\left(IM\left(t_i\right), j\right) = \sum_{k=1}^{K-1} \left(\vec{p}_k^{\,i}\left(j\right) - \vec{p}_{k+1}^{\,i}\left(j\right)\right) \times \frac{1}{k}, \qquad i = 1, 2 \tag{1}$$

$$c\left(j\right) = \mathrm{cs}\left(IM\left(t_1\right), j\right) \times \mathrm{cs}\left(IM\left(t_2\right), j\right) \tag{2}$$

where $\vec{p}_1^{\,i}(j), \ldots, \vec{p}_k^{\,i}(j), \ldots, \vec{p}_K^{\,i}(j)$ represents the multiclass probabilistic output arranged in descending order. c ranges between zero and one. The higher the c, the more reliable the corresponding multitemporal semantic label is.

B. Multitask Segmentor

Multitask learning has the advantage of leveraging interrelated subtasks, leading to improved performance and generalizability compared to individual-task models [23]. Within this investigation, we employ a multitask paradigm that integrates binary change detection (CD_bin) and multitemporal semantic segmentation (CD_sem) tasks using an encoder–decoder segmentor architecture, as visually depicted in Fig. 1. The shared encoder module captures latent feature representations pertaining to change information, while the two decoder modules extract higher level features specifically relevant to binary or semantic change detection, respectively. The encoder module comprises a series of 3-D residual blocks, with each block containing two 3×3×3 convolutional layers. The skip connections in the ordinary 2-D residual networks are replaced by a 3×3×3 convolutional layer as well. In addition, batch normalization and rectified linear unit (ReLU) functions are applied. The multitask network concludes with a fully connected layer activated using sigmoid/softmax activation function for binary change detection and multitemporal semantic segmentation.

In order to concurrently model true binary change labels and pseudomultitemporal semantic labels within a multitask learning framework, we integrate adversarial learning and ASPL methodologies. These learning strategies are implemented with the aim of mitigating the adverse effects stemming from noisy

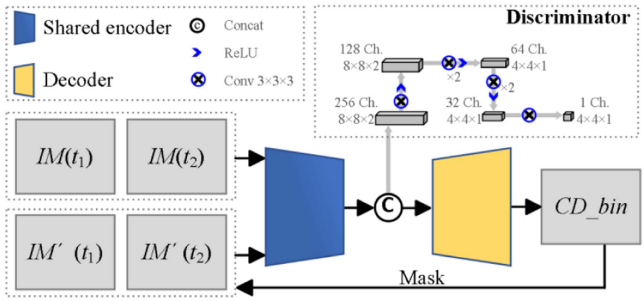

Fig. 2. Proposed discriminator network.

pseudolabels, ultimately guiding the model towards the precise extraction of salient change features. Specifically, for the binary change detection task, we incorporate adversarial learning, wherein the shared encoder within the segmentor structure serves as the generator component for a 3-D GAN. The primary objective of this generator is to map concatenated multitemporal image pairs, i.e., concatenated IM(t_1) and IM(t_2) (referred to as X), and X masked with generated binary change detection result (referred to as X′) to the latent change features (Z and Z,′ respectively). Subsequently, Z and Z' are fed into a discriminator, consisting of a sequence of seven 3-D convolution layers activated by the ReLU function, as illustrated in Fig. 2. The discriminator serves to transform the latent change features, Z and Z', each possessing dimensions of $8 \times 8 \times 2 \times 128$, into a 4×4 array as outputs (denoted as YG). Within this array, the individual elements YG_{ij} signifies whether the subpatch (i, j) of the image pairs is real or fake. Through the interactive process of adversarial learning, as the generator progressively learns the latent features and consequently enables the segmentor to accurately discriminate between changed and unchanged pixels, the task of the discriminator to differentiate between the masked changed pairs and real unchanged ones becomes increasingly intricate.

In the context of the multitemporal semantic segmentation task, we integrate the proposed ASPL to dynamically select a collection of reliable and challenging samples. This strategy is deemed notably efficacious in eliminating the adverse influence of noisy samples, while simultaneously exploiting discriminant information of samples near decision boundaries. The original hard version of ASPL, as proposed by Guo et al. [46], has been

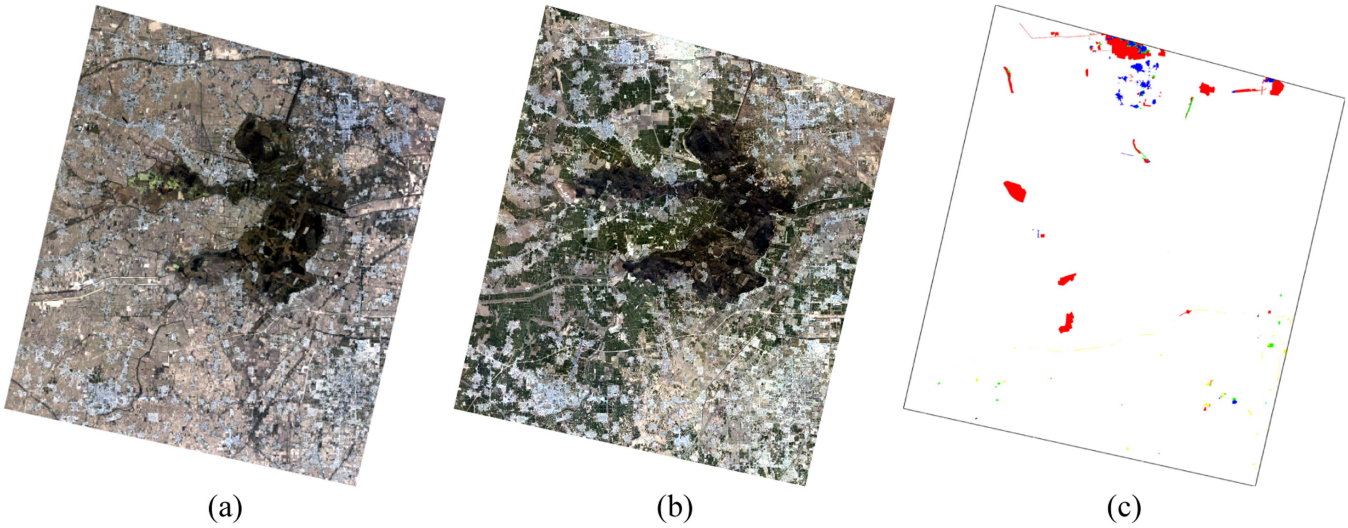

Fig. 3. OHS dataset. (a) Image acquired on October 10, 2018. (b) Image acquired on April 4, 2020. (c) Reference image.

deprecated within the scope of our study due to its susceptibility to the presence of noisy samples. Consequently, we have proposed a modified ASPL as expounded in the following:

$$v_i = \begin{cases} c(i), & \text{if } L_i > \lambda \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

$$\lambda = \mu\left(L^t\right) + \frac{t}{T}\sigma\left(L^t\right) \tag{4}$$

where $c(i)$ is the certainty score for pixel i. Values of c(i) smaller than the threshold of 0.5 are set as 0. v_i denotes pixelwise sample weights in the loss function calculation. $\mu(L^t)$ and $\sigma(L^t)$ represent average and standard deviation of losses during the tth epoch, respectively. Here, T stands for the total number of epochs. It is pivotal to underscore that the dynamically calculated threshold λ, as outlined in (4), essentially informs the selection criteria for samples based on their loss magnitudes. This calibrated threshold is tailored to adaptively accommodate the evolution of loss dynamics across epochs, ensuring the selection of informative samples while effectively mitigating the deleterious effects of noise. Furthermore, to tackle the extreme class imbalance issues in the semantic change detection network, multiclass focal loss [47] is employed. Since semantic change samples are noisy, we proposed a novel sample-weighted focal loss. There, the loss function for multitemporal semantic segmentation branch is defined as follows:

$$L_{wf} = -\frac{1}{N}\sum_{i=1}^{N} W(1-P_i)^{\gamma} T_i \log(P_i)\, v_i \tag{5}$$

where T_i and P_i are one-hot class label and predicted probability value for pixel i, respectively. W represents the class weight vector, and the values are set inversely proportional to class frequencies. N is the number of samples per-batch. Finally, the multiclass focal loss is multiplied by pixelwise certainty score v_i. In this way, both sample reliability (denoted as v_i) and complexity (indicated by dynamically calculated threshold λ) are considered in the network optimization process.

TABLE I
PIXELS AND PERCENTS FOR THE DIFFERENT SEMANTIC CHANGE TYPES

Change type	Number (pixels)	Percent (%)
Unchanged	137 96 794	98.28
Cropland	170 278	1.21
Vegetation	15 964	0.11
Impervious surface	43 262	0.31
Bareland	12 359	0.09

IV. EXPERIMENTS

A. Dataset

The evaluation of MASNet's performance utilized the Orbita Hyperspectral (OHS) dataset. OHS satellites offer 256 spectral bands (400–1000 nm) with a 10 m spatial resolution. Due to constraints in data compression and storage, we opted for the default selection of 32 bands. The dataset illustrates an urban–rural region in the Xiong'an New Area, China, with the size of 4322 × 4709 pixels (2035 km^2), within which both changes related to land-cover transitions and seasonal effects can be observed. The Xiong'an New Area is a state-level new area established on April 1, 2017. After its establishment, land cover has undergone significant land cover change, making it an ideal study area for change detection. Fig. 3 visually depicts bitemporal true color composite images and the corresponding semantic change map. Furthermore, the statistics on different change types is presented in Table I. Notably, it is observed that a substantial 98.28% of unchanged pixels exist, whereas the remaining four changed categories exhibit extreme imbalances (0.09%–1.21%). This dataset can be aptly characterized as extremely unbalanced for semantic change detection, especially in scenarios with limited semantic labels during model training. In the preprocessing phase, bitemporal images were

directly cropped into patches with the size of 128 × 128 pixels. In addition, a monotemporal classification map is required to provide land cover information. According to the comparison of three 10-m land cover products [48], FROM-GLC in 2017 [19] was chosen for its reported highest accuracy.

B. Implementation Details

Our model is implemented on the TensorFlow framework, leveraging the computational power of a single NVIDIA RTX 4090 GPU. For optimization, we employed the Adam optimizer with a learning rate 1e-4 and decay rate set at 1e-5 to minimize the loss function. Due to computational constraints imposed by the GPU, a batch size of 8 and a total of 200 epochs were set to facilitate model convergence.

To comprehensively evaluate the proposed methods, we utilized two widely adopted metrics: overall accuracy (OA) and kappa coefficient. These metrics are defined as follows:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{6}$$

$$P_e = \frac{(\mathrm{TN} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP}) + (\mathrm{FP} + \mathrm{TP})(\mathrm{FN} + \mathrm{TP})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2} \tag{7}$$

$$\mathrm{KC} = \frac{\mathrm{OA} - P_e}{1 - P_e} \tag{8}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

C. Comparative Methods

To substantiate the superiority of MASNet, we conducted comparative analyses with five state-of-the-art semantic change detection methods. These methods are introduced briefly as follows.

1) *Direct comparison of bitemporal land-cover maps (Str1)* [28]*:* In this approach, an encoder–decoder architecture with skip connections is trained to generate bitemporal land-cover maps. Subsequently, these bitemporal land-cover maps are directly compared to derive semantic change types.
2) *Direct semantic change detection (Str2)* [28]*:* For direct semantic change detection, a semantic segmentation network, structured similarly to the first approach, is utilized. However, it takes bitemporal image pairs as input to generate semantic change types.
3) *Separate binary and semantic change detection (Str3)* [28]*:* In this approach, two networks are employed: one for binary change detection and another for land-cover mapping. These networks are trained independently to produce a change map and bitemporal land-cover maps. Subsequently, when a pixel is identified as changed in the change map, change types are assigned based on the bitemporal land-cover maps.
4) *Integrated change detection and land-cover mapping network (Str4)* [28]*:* In the integrated version, the bitemporal encoder features within the land-cover mapping branch and the encoder features within the binary change detection branch are fused through a difference skip connection. These fused features are then fed into the decoder of the binary change detection network.
5) *Bitemporal semantic reasoning (Bi-SRNet)* [49]*:* A triple-branch convolution neural network containing two temporal branches and a change branch was presented. Semantic reasoning blocks are proposed to reason both single-temporal and cross-temporal semantic correlations and introduced in the Bi-SRNet.

TABLE II
QUANTITATIVE RESULTS OF MASNET AND FIVE COMPARATIVE SEMANTIC CHANGE DETECTION METHODS

Methods	OA (%)	Kappa ×100 (%)
Str1	91.49	3.82
Str2	97.50	1.15
Str3	<u>98.36</u>	15.14
Str4	98.32	<u>36.78</u>
Bi-SRNet	95.79	22.43
MASNet	**99.28**	**76.60**

V. RESULTS AND ANALYSIS

A. Performance Evaluation

The quantitative results of our experiments are presented in Table II. Bold values indicate the best accuracies, and the second-best are underlined. Notably, methods Str1 and Str2, relying solely on noisy semantic labels, exhibited the poorest performance. In contrast, Str3 and Str4, which incorporated manual binary change labels in sperate and integrated manners, demonstrated significantly superior performance, as indicated by considerably higher kappa values. Among the five comparative methods, Str4 performed the best due to its utilization of multitask learning and semantic-temporal feature fusion. Bi-SRNet performed the second best among the comparative methods in terms of kappa, and had the second lowest OA value, i.e., 95.79%. Our proposed MASNet, which additionally incorporates joint spectral-spatial feature representation, ASPL, and adversarial learning, surpasses Str4 significantly, achieving a remarkable increase of 39.82% for kappa.

The change detection results of the OHS dataset are presented in Fig. 4. This graphical representation elucidates that the proposed method obtains commendable change detection performance. Four representative scenes marked by purple boxes are selected for detailed visualization. In Scenes #1 and #4, there are cropland changes with linear and patch-like shape, respectively. On the other hand, Scenes #2 and #3 encompass various change types, including cropland change, vegetation change, and impervious surface change. The changed regions within these scenes exhibit notable variations in size and shape. The figure

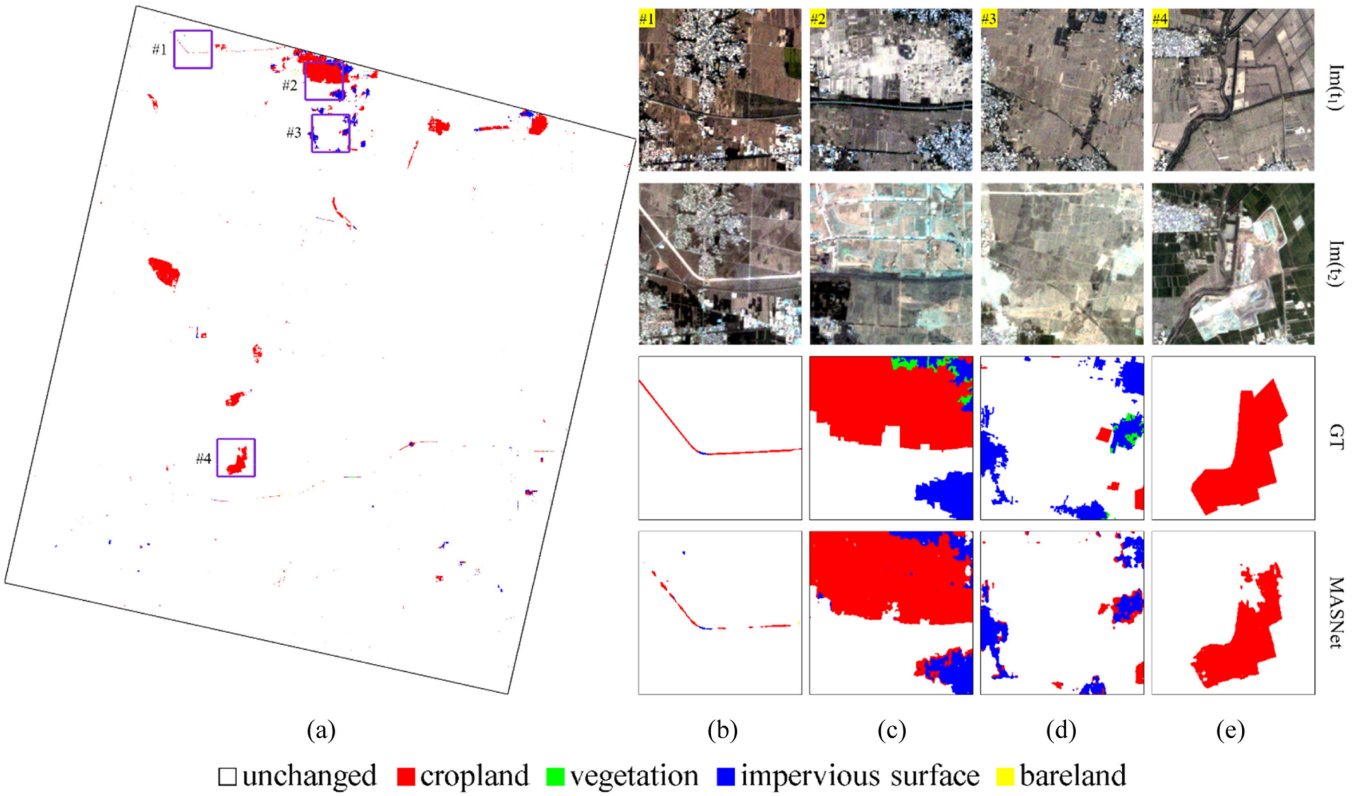

Fig. 4. Change detection results over the OHS dataset. (a) Entire region. (b) Scene #1. (c) Scene #2. (d) Scene #3. (e) Scene #4. The purple boxes in (a) are four representative scenes to demonstrate the visual details. GT denotes ground truth.

shows that the change detection results have good agreement in change boundaries and types with the ground truth. In addition, the change detection results of the compared methods over these four scenes are shown in Fig. 5. In general, the comparative results demonstrate that only MASNet accurately identifies changed regions and distinguishes change types, yielding the fewest false positives and false negatives. In contrast, other methods generally fall short. Str1, Str2, and Str3 tend to produce scattered and small patches in large change regions. In addition, Str1 misclassifies certain edge pixels of unchanged regions as changed. Str4, while exhibiting relatively complete contours, misidentifies pixels within contours as unchanged, assigning incorrect semantic change types. Scene #3 further illustrates that detecting small-scale changes and minority classes (e.g., vegetation change) poses a more formidable challenge. Bi-SRNet tends to overdraw semantic features of different temporal images. In sceneries with noisy semantic labels, it will result in poor results.

B. Ablation Study

In order to systematically assess the individual contributions of each component within our proposed method, we undertake an ablation study. In specific, we deconstruct our method into distinct combinations of its components and rigorously evaluate their impact. The various configurations used in the ablation study are delineated as follows.

1) Baseline Method: As the starting point, we took the multitemporal classification network as our baseline method. This method primarily focuses on extracting pseudomultitemporal semantic labels through PCC. As presented in Table III, the baseline method attains the highest OA at 99.72%. However, it yields the lowest values for kappa, i.e., 3.68%. This observation can be attributed to two primary factors: 1) the extreme imbalance within the semantic change detection dataset, featuring a scarcity of changed samples and varying proportions of different changed types, and 2) the susceptibility of PCC to error accumulation contending with noisy multitemporal semantic labels, leading to challenges in distinguishing between semantic change types.

2) Importance of Different Learning Strategies: To verify the efficacy of different learning strategies within the MASNet, we progressively introduced multitask (M), adversarial (A), and

TABLE III
ABLATION STUDY FOR THE DIFFERENT LEARNING STRATEGIES

		PCC	M	MA	MS	MAS
Methods	Baseline	√				
	Multitask		√	√	√	√
	Adversarial			√		√
	Self-paced				√	√
Metrics	OA	**99.72**	99.22	99.24	99.22	99.28
	Kappa	3.68	74.31	75.60	73.54	**76.60**

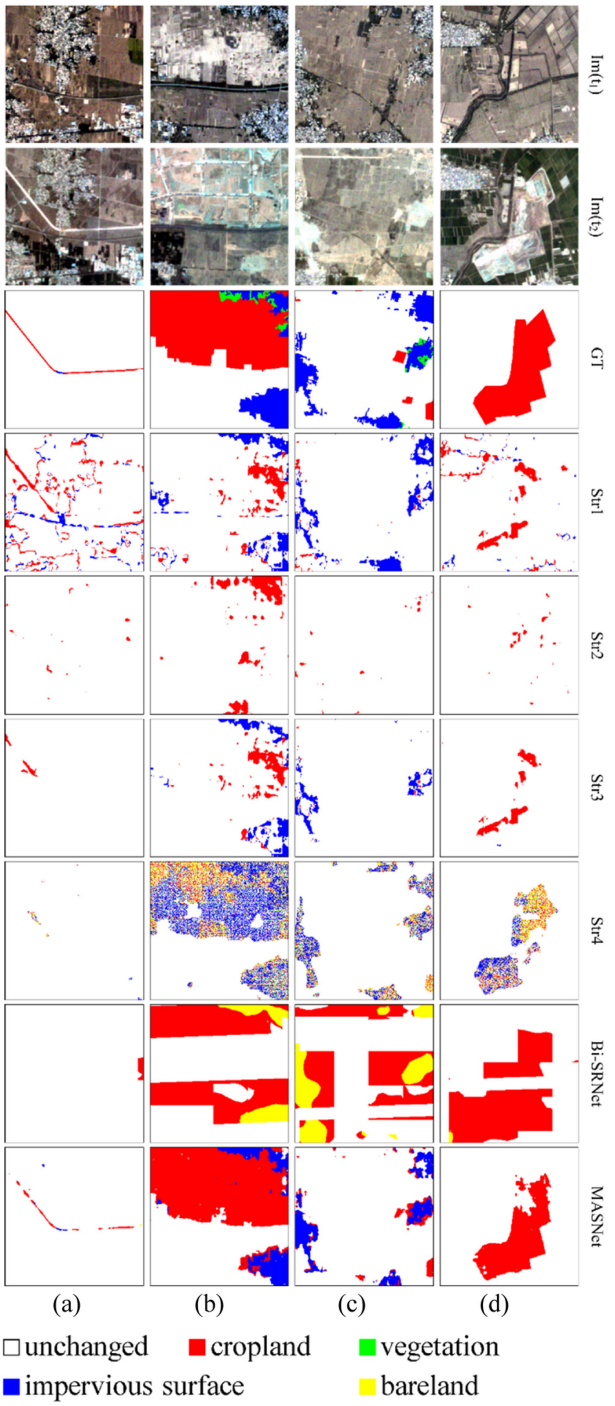

Fig. 5. Visualization of semantic change detection results of MASNet and five comparative methods over the four representative scenes. (a) Scene #1. (b) Scene #2. (c) Scene #3. (d) Scene #4. GT denotes ground truth.

self-paced (S) learning components. The results are delineated in Table III.

As illustrated in Table III, the incorporation of multitask learning, compared to the baseline, yields a noteworthy gain of 70.63% for kappa. This underscored that multitask learning significantly improves semantic change detection performance by leveraging true binary change labels and concurrently training multiple tasks. The subsequent incorporation of adversarial learning within the multitask network results in a modest 1.29% increase in kappa, demonstrating that improved feature representation of change can be achieved via 3-D GAN. The inclusion of self-paced learning within the multitask network leads to a decline of 0.77% in kappa due to negative impact of noisy labels in the self-paced network optimization without guidance of change features. The joint utilization of adversarial learning and self-paced learning within the multitask network, i.e., MASNet, surpasses all others by achieving the best results in kappa with improvements of 2.29%, compared to the multitask network. The joint utilization of adversarial learning and self-paced learning within the multitask network strikes a harmonious balance between the exploitation of noisy semantic labels and true binary change labels, thereby enhancing semantic change detection performance.

TABLE IV
ABLATION STUDY FOR 3-D CONVOLUTION. "CONV." IS AN ABBREVIATION OF CONVOLUTION

%	2-D Conv.	3-D Conv.
OA	99.00	**99.28**
Kappa	64.06	**76.60**

3) Importance of 3-D Convolution: To ascertain the significance of 3-D convolution, we conducted an ablation study by substituting the 3-D convolution filters in the MASNet with traditional 2-D filters. Table IV provides the results highlighting the importance of 3-D convolution. It is evident that the inclusion of 3-D convolution yields improvements in OA and kappa by 0.28% and 12.54%, respectively. This finding substantiates the effectiveness of joint spectral-spatial feature representation for hyperspectral change detection.

VI. DISCUSSIONS

A. Semantic Change Analysis

To meet the imperative of alleviating noncapital functions and reducing the urban expansion rate in Beijing, the Xiong'an New Area was established in 2017 [50]. Urban planning in the initial phase (2017–2020) involved creating an external transportation network and initiating infrastructure and public service facilities construction within the "starting zone" [51]. Our study period (2018–2020) coincided with the first phase of urban planning. Consequently, semantic change analysis in this study offers insights into the characteristics of this phase-one development.

Based on the change detection results in Fig. 4, we can observe that most changes occurred on the northwest side of the Baiyangdian Wetland, i.e., starting zone, and the major change types are cropland and impervious surface loss in this region. Meanwhile, there are also significant changes in the south part of the study area, characterized by the loss of cropland due to road construction and impervious surface changes because of urban renewal. These findings are consistent with the underlying urban planning objectives. Furthermore, in accordance with the

TABLE V
QUANTITATIVE RESULTS OF MASNET AND FIVE COMPARATIVE SEMANTIC CHANGE DETECTION METHODS ON THE SECOND

Methods	OA (%)	Kappa ×100 (%)
Str4	69.29	3.82
Bi-SRNet	81.71	47.61
MASNet	83.87	54.62

change analysis of the Xiong'an New Area by Luo et al. [52] cropland and impervious surface decreased from 2017 to 2020. This phenomenon implies that the development of Xiong'an New Area occupied farmland, and the demand for urban capacity was met through a combination of urban expansion and urban renewal. It is noteworthy that, in contrast to the historically quantity-oriented urbanization observed in many Chinese cities, three rigid redlines of ecological preservation, prime farmland, and urban development were delineated in the Xiong'an New Area to protect ecological space and ensure food security [53]. Therefore, semantic change analysis undertaken in the Xiong'an New Area merits attention within the existing body of research. Continuous monitoring of land cover, ecological quality, and cropland quantity and quality in the Xiong'an New Area is necessary in the further work.

B. Additional Dataset

To further evaluate the proposed method, another semantic change detection dataset was added. However, it should be noted that semantic change detection datasets are rare, let alone hyperspectral one. In this regard, we utilized an open-source semantic change detection datasets, SECOND [54]. Currently, this dataset provides 2968 pairs of bitemporal images of size 512 × 512 pixels with RGB bands. It has six land cover types, i.e., water, ground, low vegetation, tree, building, and playground. The labels are manually labeled with high accuracy. To make the SECOND applicable to scenarios where only true binary change labels and noisy land cover labels are available, we used the classification outputs as land cover labels and the compared bitemporal results as semantic change labels. In the experiment, the 2968 image pairs were divided into 2375 and 593 pairs for training and testing.

Based on this dataset, on one hand, the performance of the proposed method can be further evaluated and compared; on the other hand, the applicability of the method in large-scale multispectral high-resolution images can be tested. The generated semantic change labels yielded an OA and kappa of 47.10% and 17.17%, respectively. Table V reports the semantic change detection performance of our MASNet and two comparative methods, i.e., Str4 and Bi-SRNet, on the SECOND. The proposed method achieves the highest accuracy with OA of 83.87% and kappa of 54.62%. In addition, Fig. 6 visually displays the semantic change detection results of these three approaches on the SECOND. It can be seen that our method demonstrates

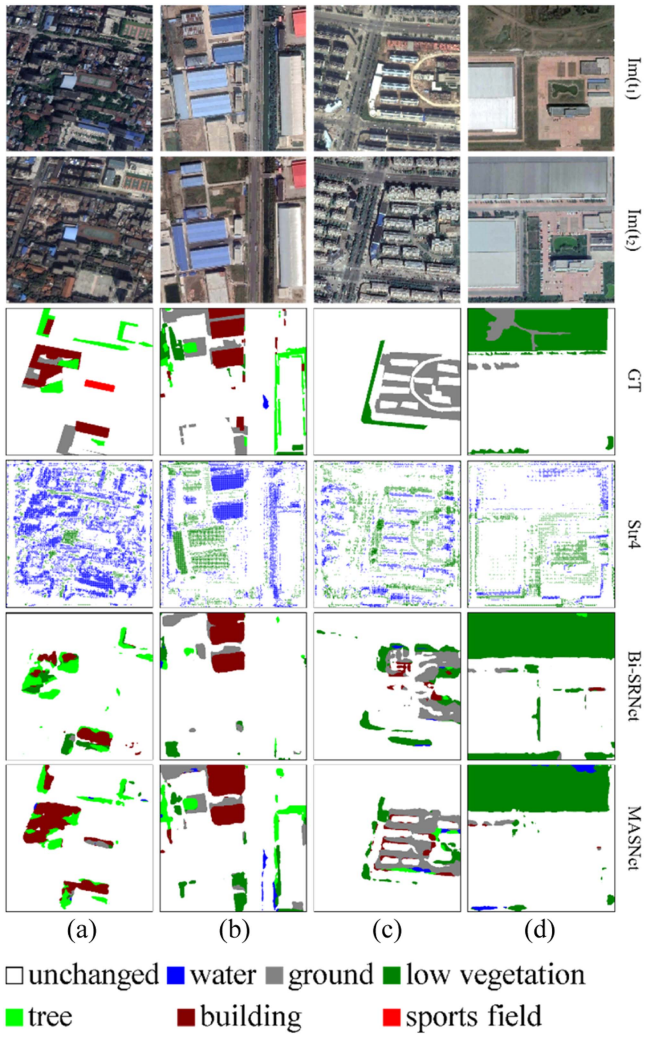

Fig. 6. Visualization of semantic change detection results of Str4, Bi-SRNet, and MASNet, over the four representative scenes in the SECOND. (a) Scene #1. (b) Scene #2. (c) Scene #3. (d) Scene #4. GT denotes ground truth.

greater consistency with ground truth maps compared to the two alternatives. In specific, our method can detect small targets in complex scenarios and provide more precise boundaries. Str4 produces serious errors, such as misclassifying shadows as water changes and erroneously identifying changes, especially in low vegetation, owing to different spectral features in bitemporal RGB images. While Bi-SRNet can identify most changes but fail to capture the small-scale ones. In conclusion, our method shows better performance compared with previous approaches in scenarios with noisy labels, which agrees with the quantitative result. More importantly, our method designed for hyperspectral semantic change detection tasks can also deal with large-scale RGB images.

C. Comparison of Multispectral and Hyperspectral Bands

We evaluated the efficacy of change detection using both multispectral (RGB + NIR) and hyperspectral (all 32) bands. The specific bands for RGB + NIR were #13 (R), #6 (G), #2 (B), and

TABLE VI
COMPUTATIONAL COMPLEXITY COMPARISON IN TERMS OF MODEL PARAMETERS (PARAMS.) AND FLOPS

Methods	Params. (M)	FLOPs (G)
Str1	1.11	0.84
Str2	1.11	0.93
Str3	2.22	1.77
Str4	2.21	2.64
Bi-SRNet	23.38	11.88
MASNet	15.96	190

#25 (NIR) within the 32-band configuration. Compared to kappa of 76.60% for change detection using all bands, the use of only four bands decreased to 62.98%. The hyperspectral bands seem to perform better in change detection than four-bands images, which signify that rich spectral information are important in distinguishing between different change types.

D. Model Computational Complexity

For a comprehensive comparison of model computational complexity, number of parameters (Params.) and floating points of operations (FLOPs) are calculated and reported in Table VI. The size of the images used for complexity comparison is 128×128 pixels. One can observe that Str1 and Str2 have smaller Params. and FLOPs, but exhibit poor performance in change detection. Our proposed method MASNet has the largest FLOPs (190 G) but the second largest Params. (15.96 M) after Bi-SRNet (23.38 M). It needs to be acknowledged that our method has higher computational complexity due to the use of 3-D convolutions in the network.

E. Advantages and Limitations

In this study, we made full use of true binary change samples and monotemporal land-cover map to extract semantic change types. The proposed approach offers two attractive advantages. First, owing to the free availability of 10-m land cover dataset, the semantic change detection approach proposed in this study can be easily migrated to other binary change detection datasets. Second, error accumulation in PCC is quantitatively assessed via multitemporal sample certainty. Subsequently, the change detection network is established to exploit certainty-weighted semantic change guided by the binary change detection task. It is proven that our method has significantly improve the performance compared to the conventional PCC. Therefore, this workflow provides a practical and effective way to overcome the limitations of PCC.

Nevertheless, it is imperative to acknowledge certain limitations inherent in our methodology. First and foremost, the accuracy of our change detection results depends on the quality of land-cover map. Despite the partial mitigation of the influence of noisy samples through sample weighting, pseudolabel refinement techniques are encouraged to improve the change detection performance in further works. Second, it is crucial to recognize that the proposed method is incapable of identifying change types that deviate from the predefined class categories within existing land-cover maps. Finally, it is crucial to underscore that our method is unsuitable for very-high-resolution semantic change detection since land-cover map with the same spatial resolution can be hardly acquired. Nevertheless, finer scale change detection with relatively coarse labels is an interesting research direction.

VII. CONCLUSION

In this study, we proposed a multitask network called MASNet for semantic change detection in high-resolution hyperspectral imagery. Our methodology is developed to operate within scenarios where only noisy monotemporal land cover labels and true binary change labels are available. This dataset configuration holds significant practical values as the availability of semantic change detection datasets remains limited. Our proposed MASNet leverages ASPL, adversarial training, and multitask learning to enhance change detection accuracy. In specific, 3-D ResUnet serves for joint spectral-spatial feature extraction from hyperspectral imagery, serving as the baseline network for our multitask framework. For the binary change detection branch, 3-D GAN is integrated to enhance the discrimination ability of high-level spectral-spatial features in detecting changes. To mitigate the adverse effects of noisy labels within the multitemporal semantic segmentation branch, a novel soft version of self-paced learning is proposed to consider both sample reliability and complexity in the network optimization. Experimental results demonstrate its effectiveness over representative semantic change detection networks, and the comprehensive ablation study indicate that components in the proposed network are combined together to improve the change detection accuracy. Based on the change detection results in the Xiong'an New Area, it is revealed that major changes are cropland and impervious surface change, and some urban expansion was achieved by occupying croplands. Finally, the advantages and limitations of our method are analyzed in detail. Future research will further improve the change detection results by pseudolabel refinement techniques and extend the proposed approach for finer scale change detection with relatively coarse labels.

REFERENCES

[1] 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.

[2] 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.

[3] M. K. Ridd and J. Liu, "A comparison of four algorithms for change detection in an urban environment," *Remote Sens. Environ.*, vol. 63, pp. 95–100, Feb. 1998.

[4] F. Bovolo and L. Bruzzone, "A novel theoretical framework for unsupervised change detection based on CVA in polar domain," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, 2006, pp. 379–382.

[5] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, pp. 4823–4838, Aug. 2008.

[6] P. R. Marpu, P. Gamba, and M. J. Canty, "Improving change detection results of IR-MAD by eliminating strong changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 799–803, Jul. 2011.

[7] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.

[8] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[9] Y. Li et al., "CBANet: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513011.

[10] A. Song, J. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3D fully convolutional networks," *Remote Sens.*, vol. 10, 2018, Art. no. 1827.

[11] 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, Jan. 2023, Art. no. 113371.

[12] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, pp. 44–53, 2018.

[13] Y. Li, X. Li, Y. Zhang, D. Peng, and L. Bruzzone, "Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing Big Data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 120, Jun. 2023, Art. no. 103345.

[14] L. Liu, D. Hong, L. Ni, and L. Gao, "Multilayer cascade screening strategy for semi-supervised change detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1926–1940, 2022.

[15] M. Gong et al., "A spectral and spatial attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521614.

[16] X. Tong et al., "A novel approach for hyperspectral change detection based on uncertain area analysis and improved transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2056–2069, 2020.

[17] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, Aug. 2015.

[18] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery," *Earth Syst. Sci. Data*, vol. 13, pp. 2753–2776, 2021.

[19] P. Gong, X. Li, and W. Zhang, "40-year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing," *Sci. Bull.*, vol. 64, pp. 756–763, Jun. 2019.

[20] 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.

[21] D. Wang, F. Zhao, C. Wang, H. Wang, F. Zheng, and X. Chen, "Y-Net: A multiclass change detection network for bi-temporal remote sensing images," *Int. J. Remote Sens.*, vol. 43, pp. 565–592, Jan. 2022.

[22] S. Tian, Y. Zhong, Z. Zheng, A. Ma, X. Tan, and L. Zhang, "Large-scale deep learning based binary and semantic change detection in ultra high resolution remote sensing imagery: From benchmark datasets to urban application," *Int. Soc. Photogrammetry Remote Sens. J. Photogrammetry Remote Sens.*, vol. 193, pp. 164–186, Nov. 2022.

[23] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, pp. 41–75, Jul. 1997.

[24] S. Tian, X. Tan, A. Ma, Z. Zheng, L. Zhang, and Y. Zhong, "Temporal-agnostic change region proposal for semantic change detection," *Int. Soc. Photogrammetry Remote Sens. J. Photogrammetry Remote Sens.*, vol. 204, pp. 306–320, Oct. 2023.

[25] C. Zhao et al., "High-resolution remote sensing bitemporal image change detection based on feature interaction and multitask learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5511514.

[26] Y. Sun, X. Zhang, J. Huang, H. Wang, and Q. Xin, "Fine-grained building change detection from very high-spatial-resolution remote sensing images based on deep multitask learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8000605.

[27] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, "A deep multitask learning framework coupling semantic segmentation and fully convolutional LSTM networks for urban change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7651–7668, Sep. 2021.

[28] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understanding*, vol. 187, Oct. 2019, Art. no. 102783.

[29] Z. Zheng, A. Ma, L. Zhang, and Y. Zhong, "Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15173–15182.

[30] Z. Zheng, Y. Zhong, S. Tian, A. Ma, and L. Zhang, "ChangeMask: Deep multi-task encoder-transformer-decoder architecture for semantic change detection," *Int. Soc. Photogrammetry Remote Sens. J. Photogrammetry Remote Sens.*, vol. 183, pp. 228–239, Jan. 2022.

[31] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "FCCDN: Feature constraint network for VHR image change detection," *Int. Soc. Photogrammetry Remote Sens. J. Photogrammetry Remote Sens.*, vol. 187, pp. 101–119, May 2022.

[32] F. Cui and J. Jiang, "MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 118, Apr. 2023, Art. no. 103294.

[33] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, pp. 1189–1197, 2010.

[34] J. S. Supancic III and D. Ramanan, "Self-paced learning for long-term tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2379–2386.

[35] X. He, Y. Tao, S. Yang, H. Dai, and H. Lin, "voxel2vec: A natural language processing approach to learning distributed representations for scientific data," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 10, pp. 4296–4311, Oct. 2023.

[36] W. Hou, N. Chen, J. Peng, W. Sun, and Q. Du, "Pyramidal dilation attention convolutional network with active and self-paced learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1503–1518, 2023.

[37] D. Meng, Q. Zhao, and L. Jiang, "A theoretical understanding of self-paced learning," *Inf. Sci.*, vol. 414, pp. 319–328, Nov. 2017.

[38] H. Li, M. Gong, M. Zhang, and Y. Wu, "Spatially self-paced convolutional networks for change detection in heterogeneous images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4966–4979, 2021.

[39] R. Shang, Y. Yuan, L. Jiao, Y. Meng, and A. M. Ghalamzan, "A self-paced learning algorithm for change detection in synthetic aperture radar images," *Signal Process.*, vol. 142, pp. 375–387, Jan. 2018.

[40] M. Gong, Y. Duan, and H. Li, "Group self-paced learning with a time-varying regularizer for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2481–2493, Apr. 2020.

[41] H. Li, J. Li, Y. Zhao, M. Gong, Y. Zhang, and T. Liu, "Cost-sensitive self-paced learning with adaptive regularization for classification of image time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11713–11727, 2021.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[43] R. A. Naqvi and W. K. Loh, "Sclera-Net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network," *IEEE Access*, vol. 7, pp. 98208–98227, 2019.

[44] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7570–7588, 2021.

[45] O. Ahlqvist, "Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 U.S. National Land Cover Database changes," *Remote Sens. Environ.*, vol. 112, pp. 1226–1241, Mar. 2008.

[46] X. Guo et al., "Adaptive self-paced deep clustering with data augmentation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 9, pp. 1680–1693, Sep. 2020.

[47] S. Deepak and P. M. Ameer, "Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion," *Neurocomputing*, vol. 520, pp. 94–102, Feb. 2023.

[48] J. Kang et al., "Comparison of three ten meter land cover products in a drought region: A case study in Northwestern China," *Land*, vol. 11, p. 427, 2022.

[49] L. Ding, H. Guo, S. Liu, L. Mou, J. Zhang, and L. Bruzzone, "Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5620014.

[50] D. Liu, X. Zheng, H. Wang, C. Zhang, J. Li, and Y. Lv, "Interoperable scenario simulation of land-use policy for Beijing–Tianjin–Hebei region, China," *Land Use Policy*, vol. 75, pp. 155–165, Jun. 2018.

[51] Y. Zou, Z. Chen, N. Zhong, and W. Zhao, "Urban planning as a way to pursue quality-oriented urbanization: Anatomy of the urban planning of Xiong'an new area, China," *J. Urban Affairs*, vol. 45, pp. 1418–1433, Sep. 2023.
[52] J. Luo, X. Ma, Q. Chu, M. Xie, and Y. Cao, "Characterizing the up-to-date land-use and land-cover change in Xiong'an new area from 2017 to 2020 using the multi-temporal sentinel-2 images on Google Earth Engine," *Int. Soc. Photogrammetry Remote Sens. Int. J. Geo-Inf.*, vol. 10, p. 464, 2021.
[53] B. Xu, "China's National New Areas in the ecological transition," *Environ., Develop. Sustainability*, vol. 25, pp. 3747–3770, Apr. 2023.
[54] K. Yang et al., "Asymmetric siamese networks for semantic change detection in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609818.

Dawei Wen received the B.S. degree in surveying and mapping from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2018.

She is currently a Lecturer with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China. Her research interests include urban remote sensing, high-resolution image processing, and change detection.

Xin Huang (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, where he is also a Full Professor and Teaches remote sensing and image interpretation. He is the Head with the Institute of Remote Sensing Information Processing, School of Remote Sensing and Information Engineering, Wuhan University. He has been supported by the National Program for Support of Top-notch Young Professionals in 2017, the China National Science Fund for Excellent Young Scholars in 2015, and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. He has authored or coauthored more than 200 peer-reviewed articles (SCI papers) in the international journals. His research interests include remote sensing image processing methods and applications.

Dr. Huang was the recipient of the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) in 2010, the John I. Davidson President's Award from ASPRS in 2018, and the National Excellent Doctoral Dissertation Award of China in 2012. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as the Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the winner of the IEEE GRSS Data Fusion Contest in the years of 2014 and 2021. He was the Lead Guest Editor of the special issue for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, *Photogrammetric Engineering and Remote Sensing*, and *Remote Sensing*. He was an Associate Editor for the *Photogrammetric Engineering and Remote Sensing*, from 2016 to 2019, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, from 2014 to 2020, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING, from 2018 to 2022. Since 2022, he has been an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. Since 2019, he has been an Editorial Board Member of the Remote Sensing of Environment

Qiquan Yang received the B.S. degree in survey and mapping and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016 and 2021, respectively.

He is currently a Postdoctoral Researcher affiliated with the Macau University of Science and Technology, Macau, China, and Tongji University, Shanghai, China. He has authored more than ten peer-reviewed papers in the international journals. He is research interests mainly include infrared remote sensing data processing and analysis, surface thermal environment modeling, and planetary surface water infrared detection.

Jianqin Tang received the B.S. degree in computer science and technology from Wuhan University, Wuhan, China, in 2003.

She is currently a Senior Experimentalist with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China. Her research interests include soft engineering and image processing.