Learning Remote Sensing Aleatoric Uncertainty for Semi-Supervised Change Detection

Jinhao Shen, Cong Zhang, Mingwei Zhang, Qiang Li, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Significant progress has been recently achieved in the field of remote sensing image change detection based on data-driven deep learning. Fully supervised models have limitations on the availability of massive annotated training data, while Semi-Supervised Change Detection (SSCD) has garnered increasingly widespread attention. Nevertheless, existing SSCD methods do not categorize the types of Remote Sensing Aleatoric Uncertainty (RSAU), let alone investigate the impact of uncertainty on performance. To this end, we define RSAU for SSCD and introduce the Progressive Uncertainty-aware and -guided Framework (PUF). It consists of two crucial components to perceive and guide the RSAU in the training stage. The first component, *i.e.*, Progressive Uncertainty-Aware Learning (PUAL), decodes and quantifies the uncertainty inherent in the samples from the weak branch. The second one, *i.e.*, Uncertainty-guided Multi-view Learning (UML), generates multiple image pairs designed for distortion and mixing for the strong branch. UML utilizes the uncertainty values derived from PUAL to offer guidance throughout the training process, which discerns and learns discriminative features from high-quality samples. Extensive experiments are conducted on three multi-class and building change detection benchmarks, *i.e.*, CDD, SYSU, and LEVIR-CD. Furthermore, we propose a small dataset to enhance the understanding of aleatoric uncertainty, namely LEVIR-AU. The proposed PUF consistently achieves state-of-the-art performance. The dataset and codes are available at https://github.com/shenjh0/PUF.**

***Index Terms*—Remote sensing, change detection, semi-supervised learning, uncertainty.**

I. Introduction

WITH the growth of earth observation satellites like Sentinel2 [1], GeoEye [2], and GF, the quantity of Remote Sensing Images (RSI) has grown exponentially. Change Detection (CD) aims to identify pixel-level changes in RSI pairs across different periods. This is a continually evolving field of research with diverse applications involving environmental monitoring [3], urban expansion [4], [5] and resource management [6].

Thanks to advancements in deep learning [7], [8], CD approaches have attained impressive performance. Currently, mainstream CD methods focus on Convolutional Neural Networks (CNN) [9], [10], [11]. Nonetheless, these approaches substantially rely on extensive manually annotated change maps. For instance, the full mask labels cost 239.7 seconds per image on Pascal VOC [12]. Thus, labeling change maps at the pixel level is highly time-consuming and labor-intensive. As a result, the effectiveness of fully-supervised methods is significantly constrained when a limited amount of labeled RSI is available. Semi-Supervised Change Detection (SSCD) is a promising solution in such circumstances. It effectively exploits labeled data while actively extracting valuable information from unlabeled images. SSCD categorizes RSI pairs into labeled and unlabeled based on the presence of annotations. For labeled images, most SSCD methods follow the same supervised training strategy, which is rarely explored. In the case of unlabeled images, some SSCD methods employ the paradigm of adversarial learning [13]. These methods primarily use feature perturbation to encourage the models to learn valid representations between the predicted change maps of unlabeled and the actual change maps of labeled RSI. However, they introduce a significant amount of noise and uncertainty into the input images. The mainstream SSCD methods mainly consider the consistency in unlabeled RSI pairs [14], [15], [16]. These methods split the training pipeline into weak and strong branches to acquire the invariant change information by consistency learning. Recently, consistency

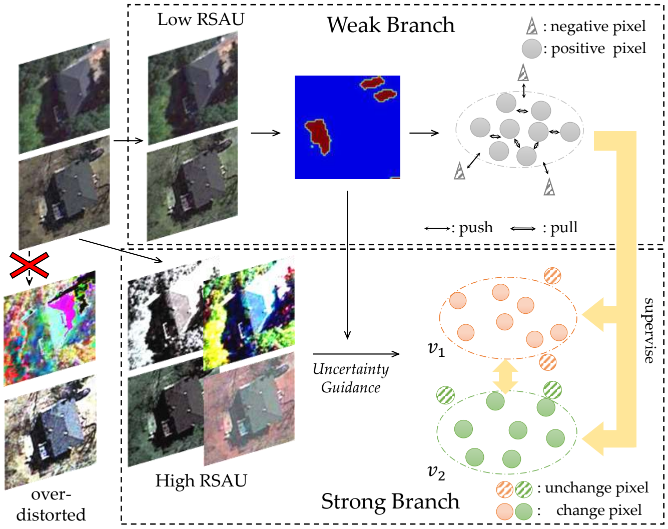

Fig. 1. The motivation and task of the PUF. It perceives the RSAU from the weakly augmented image. After generating multiple views by applying the original RSI pairs with moderate distortion. The predictions from multiple views are trained with pseudo labels based on the RSAU value in the weak branch.

Jinhao Shen, Mingwei Zhang, Qiang Li, and Qi Wang are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China. (e-mail: jinhaoshen00@gmail.com, dlaizmw@gmail.com, liqmges@gmail.com, crabwq@gmail.com)

Cong Zhang is with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong. (e-mail: congclarence.zhang@connect.polyu.hk)

This work was supported in part by the National Natural Science Foundation of China under Grant U21B2041.

Corresponding author is Qi Wang.

learning-based methods have attained outstanding performance and achieved the primary State-Of-The-Art (SOTA) methods.

The uncertainty severely restricts the performance of aerial tasks in the literature. Although it is still not extensively researched in SSCD, it performs a vital role in the training process. It is commonly categorized into two forms based on their underlying causes [17]: epistemic uncertainty and aleatoric uncertainty (*i.e.*, data uncertainty). Epistemic uncertainty [18] is closely associated with the model itself. Aleatoric uncertainty [19], [20] is primarily caused by intrinsic data corruption. We take a deep dive into the aerial data and assume that temporal variation uncertainty [21] and imaging uncertainty [22], [23] are fundamental components of the aleatoric uncertainty. The imaging uncertainty stems from diverse forms of complex noise within the remote sensing image imaging process. Additionally, temporal variation uncertainty arises from variations in light intensity, shadows, and resolution across different images. This study concentrates on these two categories of uncertainty, referring to them as Remote Sensing Aleatoric Uncertainty (RSAU).

Recent superior SSCD methods typically divide the training pipeline into weak and strong branches. From the uncertainty perspective, the weak branches generate valid pseudo-labels and learn valid representations, preserving the RSAU associated with the original RSI pairs. In contrast, the strong branch is compelled to incorporate temporal variation uncertainty and imaging uncertainty. In the real world, the RSAU varies among different samples in the training set. Previous methods [24] overlook this variation and treat all unlabeled training data uniformly. However, as different samples contain varying degrees of valid changes, it becomes necessary to adaptively remove relatively inefficient samples. Besides, most SSCD methods [25] directly inherit the data augmentations in semi-supervised segmentation methods [26]. However, natural images exhibit minimal imaging uncertainty, whereas RSI pairs employed for change detection encompass extensive and intricate remote sensing aleatoric uncertainty. The over-distorted augmentations for a single image will hurt the data distribution, misleading the model for training.

To this end, we take a deep dive into the guidance of remote sensing aleatoric uncertainty in the training process and quantify the uncertainty inherent in the data. Moreover, we devise multiple views to prevent excessive distortion of images while preserving a high level of uncertainty. Fig. 1 illustrates the content of various data RSAU along with our primary concepts. We take a dive into the weak-strong pipelines and propose the Progressive Uncertainty-aware and -guided Framework (PUF). For the weak branch, we introduce an uncertainty-aware modeling scheme. It designs an uncertainty-aware group decoder to generate distinctive logits and quantifies the prediction map's uncertainty level. The decoder groups change features to automatically sense uncertainty, further outputs a predictive change map. We contend that these logits encompass not only confidence information but also the uncertainty distributions. Hence, we devised the uncertainty-based probabilistic value to quantify the uncertainty values within the weak prediction results, aiming to inform and guide the subsequent training process.

For the strong branch, we design an Uncertainty-guided Multi-view Learning (UML) strategy. It observes diverse scenarios by adjusting moderate and multiple distortions while mitigating the degradation caused by over-distorted augmentations. Leveraging strongly augmented RSI pairs, this approach addresses some limitations of existing methods by incorporating confidence maps and a patch mixing part. We serve the uncertainty-based probabilistic value as coefficients to guide multi-view RSI pairs with various remote sensing aleatoric uncertainties. Additionally, the patch mixing strategy guided by confidence in positive predictions blends images at the patch level, altering the uncertainty distribution. In this way, our proposed method enhances the resilience and performance of semi-supervised change detection in complex environmental conditions. We conduct comprehensive experiments within the SSCD setting to validate this method on three well-established change detection datasets. These datasets encompass the multi-class change datasets (CDD, SYSU) and the building dataset (LEVIR). The building change detection dataset shows lower aleatoric uncertainty than multi-class datasets because it only has one change type and simple scenarios. It is noteworthy that we annotate thousands of image pairs featuring building variations. These are employed to enhance the data uncertainty related to buildings, which increases the task's difficulty. This dataset is referred to as LEVIR-AU. The contributions of this work can be summarized as follows:

1) We define the Remote Sensing Aleatoric Uncertainty (RSAU), encompassing imaging uncertainty and temporal variation uncertainty, for SSCD. Furthermore, we introduce the Progressive Uncertainty-aware and -guided Framework.
2) We propose Progressive Uncertainty-Aware Learning (PUAL) for the weak branch, which contains the uncertainty-aware group decoder and quantifies the uncertainty in prediction maps.
3) The Uncertainty-guided Multi-view Learning (UML) is designed for the strong branch. It adaptively learns from multiple RSI pairs with moderate distortions and excludes the impact of samples exhibiting various uncertainties.
4) We annotate nearly a thousand RSI pairs with building changes to enrich the data uncertainty of the LEVIR dataset, namely LEVIR-AU. Experimental results demonstrate that our method achieves SOTA performance.

II. Related Work

This section provides a concise overview of the relevant prior research, encompassing the fully-supervised change detection, the semi-supervised change detection, and the uncertainty in remote sensing.

A. Fully-supervised Change Detection

Recently, fully-supervised change detection methods have garnered significant attention due to their exceptional performance. Within the realm of deep learning-based approaches, two primary branches emerge: image-level and feature-level strategies.

The former approach involves aggregating input data and directing it through a singular segmentation network for change detection. Current research focuses on the latter, specifically on feature-level strategy. In the work [27], De et al. concatenate two bitemporal images and input them into a convolutional-based network. Fang et al. [28] enhance shallow-layer features to avoid missing small objects. Additionally, Li et al. [29] introduce the Densely Attentive Refinement Network (DARNet) to facilitate the learning of multiscale global features.

The aforementioned methods showcase the efficacy of CNNs and attention mechanisms on various fully-labeled datasets. Nevertheless, a notable limitation is their heavy reliance on a substantial volume of labeled data, and their robustness in scenarios with missing labels has yet to be established.

B. Semi-supervised Change Detection

Given the significant cost associated with manual pixel-level labeling, there is a growing interest in semi-supervised change detection (SSCD) methods that make efficient use of label-free data. Wang et al. [30] introduce a novel reliable contrastive learning (RCL) by designing a contrastive loss based on changed areas to enhance feature extraction, and utilizing uncertainty-based selection of reliable pseudo labels for improved pseudo label quality. Peng et al. [13] introduce an end-to-end semi-supervised convolutional network called SemiCDNet, which designs a lightweight attention module to generate initial change maps for pairs of RSI.

Recent studies have demonstrated significant performance gains with consistency learning-based Semi-Supervised Change Detection (SSCD) methods. These approaches leverage the framework of knowledge distillation [7], [31], [32] to facilitate a thorough investigation of intrinsic change information by constructing both weak and strong branches. Bandara et al. [15] employ random perturbations on the feature difference map and minimize dissimilarity using Mean Square Error (MSE) loss for change probability prediction. Zhang et al. [25] propose an innovative progressive SSCD framework with a feature-prediction alignment strategy. Additionally, Zhang et al. [16] introduce Joint Self-Training and Rebalanced Consistency Learning (ST-RCL) to enhance model robustness against imbalanced distribution and rotation consistency.

C. Uncertainty in Remote Sensing

To our knowledge, there is essentially no research on uncertainty in SSCD. Therefore, we primarily refer to related studies in computer vision and remote sensing tasks. Kendall et al. [17] identify two kinds of uncertainty: epistemic and aleatoric uncertainty. To be specific, epistemic uncertainty pertains to uncertainties associated with model weights, while aleatoric uncertainty encompasses variations arising from noise in the data. The quantification and analysis of uncertainty play a crucial role in ascertaining the confidence level associated with inherent errors in observations. Consequently, this approach has been widely applied in numerous remote sensing tasks [33], [34].

Huang et al. [18] define epistemic uncertainty and elucidate the oversights in the detector. However, they overlook the data uncertainty within the pipeline. Asadi et al. [19] highlight that evaluating data-dependent uncertainty for a regression neural network involves applying a Gaussian prior over the output from the last hidden layer. Nevertheless, it still fails to consider the temporal variational uncertainty inherent in dual-temporal remote sensing images. Saberi et al. [20] introduce a method that perturbs the predicted logit by utilizing a CNN model to draw samples from a Gaussian distribution, providing predictions with associated uncertainty information. Chen et al. [22] propose a Bayesian CNN incorporating variational inference for ice–water detection. This model furnishes uncertainty maps, predictions, and valuable feedback on regions, contributing to the enhancement of scene interpretation. Zhang et al. [23] introduce an innovative semi-supervised method for SAR target detection. This method specifically selects highly confident samples and incorporates a strong branch marked by substantial data perturbation, such as rand-augment. The uncertainty it defines falls under the category of imaging uncertainty. As has been stated, existing research has focused exclusively on one aspect of RSAU, leading to certain limitations. It is noteworthy that the significance of RSAU in the training pipeline is not taken into account in semi-supervised change detection.

III. METHOD

In this section, we provide a detailed exposition of the Progressive Uncertainty-aware and -guided Framework (PUF) for semi-supervised change detection. The framework comprises two principal modules: 1) Progressive Uncertainty-aware Learning (PUAL) for the weak branch and 2) Uncertainty-guided Multi-view Learning (UML) for the strong branch. PUAL incorporates an uncertainty-aware group decoder and systematically quantifies uncertainty within prediction maps. Further, UML augments the quantity of image pairs while diminishing the uncertainty content in a single pair. It strategically utilizes uncertainty values to direct the training process. In conclusion, the algorithm is formulated as Algorithm 1.

In subsequent sections, we present a comprehensive overview of our algorithm. The section III.A introduces the overarching framework, providing a high-level understanding. Following this, the section III.B delves into PUAL's intricate design details. Immediately thereafter, the section III.C meticulously elucidates the principles underlying UML.

A. Overview

Here, we provide a detailed overview of our training pipeline. During the training stage, the train images are divided into two subsets, we set them as a labeled dataset $D_l = \{\{l_{pre,i},\ l_{post,i}\}, y_i\}_{i=1}^{B}$ and an unlabeled dataset $D_u = \{u_{pre,j},\ u_{post,j}\}_{j=1}^{B}$. $l_{pre,i}$ and $l_{post,i}$ denote the image before and after change respectively. B represents the number of samples in each batch.

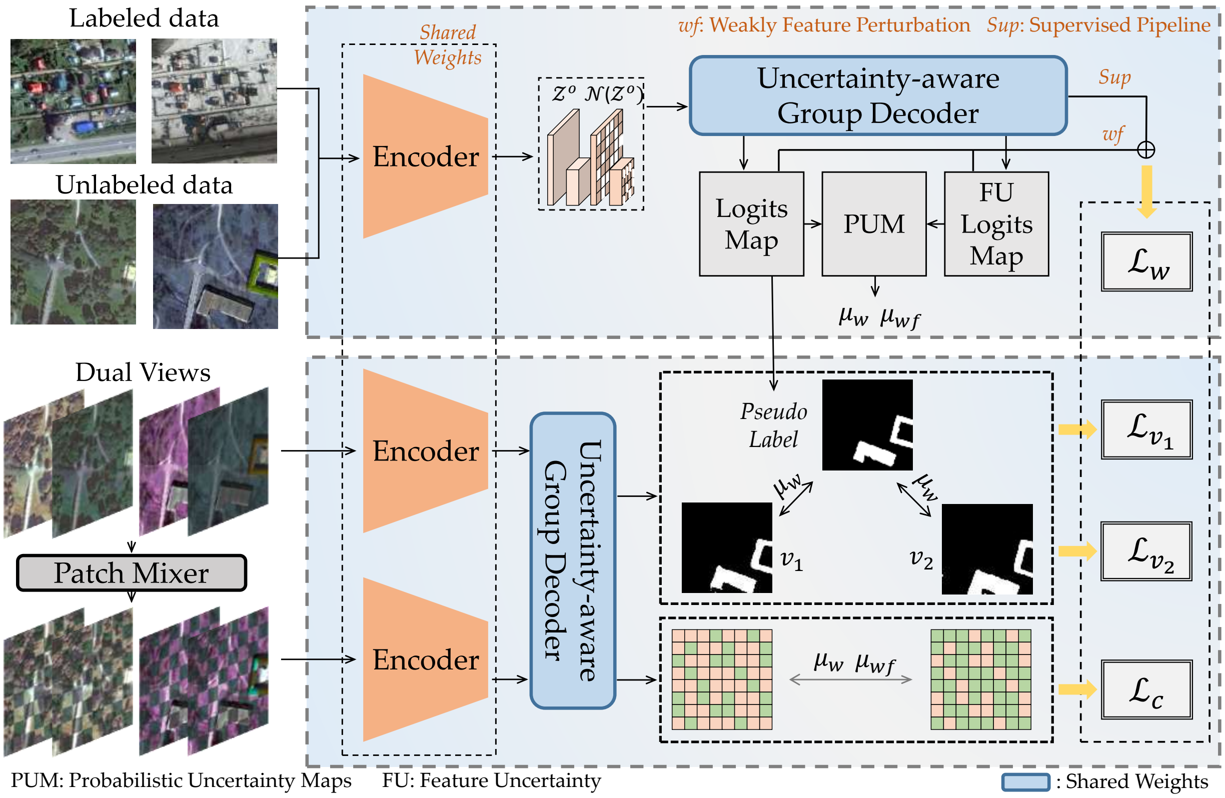

Fig. 2. The overview of the proposed method. The upper and lower segments correspond to the weak and strong branches, respectively. The weak branch computes the dissimilarity between labeled and weakly augmented unlabeled data, characterized by lower uncertainty. In contrast, the strong branch produces suitably distorted multi-view RSI pairs, leveraging information from the weak branch during training.

Following [26], each unlabeled image is weakly and strongly augmented (denoted as A_w and A_s) to w and v. The w contains the original uncertainties of the image, and v is augmented with a large number of uncertainties. Given a batch of RSI pairs, we adopt a standard encoder network $f(\cdot)$ to extract change features z. Concretely, z_l^o, z_w^o, and z_s represent the features of labeled, weakly augmented, and strongly augmented RSI pairs.

For features from labeled and weakly augmented batches, we employ a straightforward mixing operation that utilizes feature uncertainty to amalgamate z_l^o and z_w^o into a unified representation denoted as z^m. This operation is essential in effectively reducing the inherent uncertainty in the dataset. Subsequently, we develop a novel uncertainty-aware group decoder, denoted as ψ_g, which selectively incorporates emphasized features. The outputs of $\psi_g(\cdot)$ encompass the logits of change maps $\hat{y}_w$ and $\hat{y}_x$, further transform them into the probabilistic uncertainty maps. They are categorized into $\hat{p}_w$ and $\hat{p}_x$. Additionally, the confidence maps are generated by selecting pixels from the predicted logits based on fixed thresholds.

For the strongly augmented batches, we employ random intensity-based augmentations to enhance the remote sensing aleatoric uncertainties. These two image pairs are denoted as v_1 and v_2. After obtaining z_{v_1} and z_{v_2} by the same encoder $f(\cdot)$, we feed them into ψ_g to generate two predicted change images, which are denoted as $\hat{y}_{v_1}$ and $\hat{y}_{v_2}$, respectively. Furthermore, we mix v_1 and v_2 with original confidence maps, denoted as v_{12} and v_{21}. They are fed into the model to obtain two pairs of change detection maps, denoted as $\hat{y}_{v_{12}}$ and $\hat{y}_{v_{21}}$. The uncertainty between the two prediction maps should be consistent, so we use L_c to calculate their losses.

Algorithm 1 Pseudo Codes

```
Input:
    Labeled and unlabeled data: $\mathbf{D_l}$, $\mathbf{D_u}$
    Weak and strong augmentation: $\mathcal{A}_w$, $\mathcal{A}_s$
    Encoder model: $f$
    Uncertainty-aware group decoder: $\psi_g$

TRAIN(Weak Branch)
select randomly $l,\ y_x \subset \mathbf{D_l},\ w \subset \mathcal{A}_w(\mathbf{D_u})$
# inject uncertainty to the change features
$z_l^o,\ z_w^o = f(l,\ w),\ \mathcal{Z}^o = Cat\{z_l^o,\ z_w^o\}$
$\hat{y}_x,\ \hat{y}_w,\ \hat{y}_{xf},\ \hat{y}_{wf} = \psi_g(Cat\{\mathcal{Z}^o,\ \mathcal{N}(\mathcal{Z}^o)\})$
$\mu_w,\ \mu_{wf} \leftarrow \mathcal{H}_c(X_w),\ \mathcal{H}_c(X_{wf}) \leftarrow \hat{y}_w,\ \hat{y}_{wf}$
$\mathcal{L}_w \leftarrow \mu_{wf},\ y_x,\ \hat{y}_x,\ \hat{y}_w, \hat{y}_{wf}$
TRAIN(Strong Branch)
select randomly $v_1,\ v_2 \subset \mathcal{A}_s(\mathbf{D_u})$
$v_{12},\ v_{21} \leftarrow \mathcal{M}(v_1,\ v_2)$
$\hat{y}_{v_1},\ \hat{y}_{v_2} = \psi_g(f(v_1,\ v_2))$
$\mathcal{L}_{v_1} + \mathcal{L}_{v_2} \leftarrow \mu_w,\ \hat{y}_{v_1},\ \hat{y}_{v_2}$
# mix the dual views
$\hat{y}_{v_{12}},\ \hat{y}_{v_{21}} = \psi_g(f(v_{12},\ v_{21}))$
$\mathcal{L}_c \leftarrow \mu_w,\ \mu_{wf},\ \hat{y}_{v_{12}},\ \hat{y}_{v_{21}}$
$\mathcal{L} \leftarrow \mathcal{L}_w,\ \mathcal{L}_{v_1},\ \mathcal{L}_{v_2},\ \mathcal{L}_c$
# labeled RSI output: $\hat{y}_x$, weak branch output: $\hat{y}_w$
# strong branch output: $\hat{y}_{v_1},\ \hat{y}_{v_2}$
Output: $\hat{y}_x, \hat{y}_w, \hat{y}_{v_1}, \hat{y}_{v_2}$
```

B. Progressive Uncertainty-aware Learning

The supervised and weak enhanced unlabeled images encapsulate both the primary spatial location and color information from the original RSI pairs. Therefore, we jointly fed them into the CNNs. After acquiring the encoded features of an individual image, we perform feature subtraction between different images, resulting in the extraction of z_l^o and z_w^o.

The features of labeled and weakly augmented images are computed at once and denoted as $\mathcal{Z}^o = Cat\{z_l^o, z_w^o\}$. The [15] introduces a variety of feature noise, but according to our analysis, its effective nature is a modest amount of feature uncertainty. Unlike them, we treat feature uncertainty as a regularization term as follows:

$$\mathcal{Z}^m = Cat(\mathcal{Z}^o, \mathcal{N}(\mathcal{Z}^o)) \tag{1}$$

where $\mathcal{N}(\cdot)$ denote the feature uncertainty type.

After incorporating feature regularization uncertainty, we design the uncertainty-aware group decoder to refine the semantic features with precision. It deviates from the vanilla decoder [35] by incorporating a dynamic uncertainty-aware module. Specifically, the module employs several heads to compress the features into multiple groups. Each head diminishes the size of the middle layer through a pooling layer, further capturing global feature relationships with stacked linear-relu blocks. The fused maps are employed as weights, multiplicatively applied to the original features of the middle layer. The output can be partitioned into groups according to the number of headers, which encapsulates comprehensive feature information for both applied and unmodified remote sensing aleatoric uncertainty. After stacking, multiple groups are fed into the classifier to generate the final predictions. The uncertainty-aware group decoder depicted in Fig. 3 can be represented by the following mathematical formula:

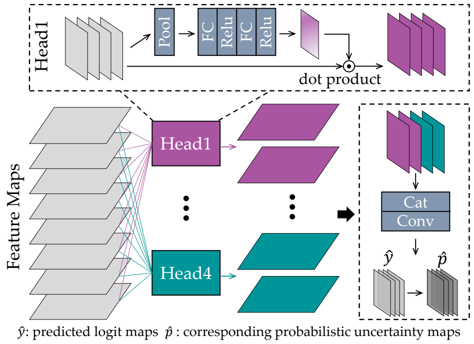
Head1
Pool
FC
Relu
FC
Relu
dot product
Feature Maps
Head1
Head4
Cat
Conv
$\hat{y}$
$\hat{p}$
$\hat{y}$: predicted logit maps $\hat{p}$: corresponding probabilistic uncertainty maps

Fig. 3. The design of the dynamic module in the uncertainty-aware group decoder. It fuses the intermediate features through each of the four prediction heads, then outputs predicted logit maps along with corresponding probabilistic uncertainty maps.

$$\hat{y}_x,\ \hat{y}_w,\ \hat{y}_{xf},\ \hat{y}_{wf} = \psi_g(\mathcal{Z}^m) \tag{2}$$

Here, the generated logits can be divided into four categories: $\hat{y}_x$ for labeled RSI, $\hat{y}_w$ for weakly augmented RSI, $\hat{y}_{xf}$ for labeled RSI with feature uncertainty, and $\hat{y}_{wf}$ for weakly augmented RSI with feature uncertainty.

Previous work [16], [30] have confirmed the presence of reliable confidence information within the logits. However, these logits contain more crucial uncertainty information, as they can reflect the model's degree of ambiguity for each pixel. We first use softmax to transform four types of logits into probabilistic uncertainty maps, denoted as $\hat{p}(x_i)$. After this, the relative relationships between different classes in the predicted probability maps will remain unchanged. For binary change detection tasks, the number of channels in confidence maps typically equals 2, representing the predicted probabilities for unchanged and changed pixels. In the channel responsible for predicting changes, as the probability values approach 1, the confidence in a change occurrence increases. When pixel probabilities approach 0.5, this signifies that the algorithm encounters difficulty in discerning the presence of a change, resulting in heightened uncertainty. During the initial phases of model training, the predicted probabilities for all categories lack reliability. Consequently, the probabilities for different classes at a given point remain unnormalized. The probabilistic uncertainty maps preserves the initial pixel confidence information. Essentially, pixels with lower probabilities exhibit reduced confidence and heightened uncertainty.

Drawing from the aforementioned analysis, we introduce the uncertainty-based probabilistic value to quantify the remote-sensing aleatoric uncertainty of change maps. The formula is as follows:

$$\mathcal{H}_c(X) = \mathbb{E}[-\log \hat{p}(x_n)] + \mathbb{E}(1 - \mathrm{TopK}(\hat{p}(x_n))) \tag{3}$$

where x_n represents the n^{th} pixel of the probabilistic uncertainty map in the batch x. The TopK operation allows the model to focus only on the most reliable pixels in the

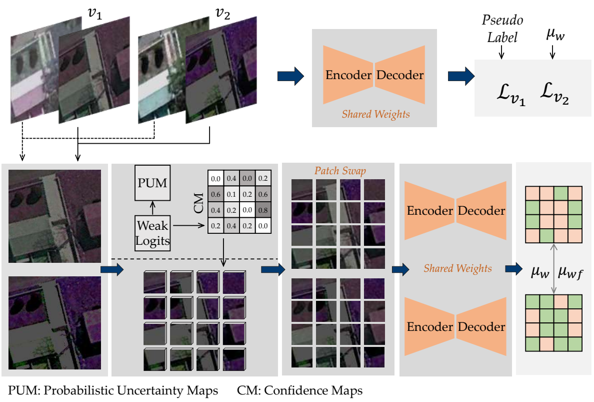

Fig. 4. The design of Uncertainty-guided Multi-view Learning. It exchanges each patch tending to change in dual-views, v_1 and v_2, based on the confidence maps generated from the weak logits.

unlabeled batch. We ultimately calculate the uncertainty-based probabilistic value for the logits from the weakly augmented RSIs, denoted as $\mathcal{H}_c(X_w)$ and $\mathcal{H}_c(X_{wf})$.

Furthermore, we compute the loss with the logits $\hat{y}_w$ and $\hat{y}_{wf}$. It boosts the capacity to recognize feature uncertainty. Thus, the total loss of this branch is written as:

$$\mathcal{L}_w = \mathcal{CE}(\hat{y}_x, y_x) + \mu_{wf} \cdot \mathcal{CE}(\hat{y}_w, \hat{y}_{wf}) \tag{4}$$

where $\mathcal{L}_w$ represents the loss of the labeled and weakly augmented branch, $\mathcal{CE}$ represents the cross entropy loss function.

C. Uncertainty-guided Multi-view Learning

The Uncertainty-guided Multi-view Learning (UML) focuses on the strong branch of unlabeled images. One of the goals of SSCD methods is to optimize the utilization of strongly augmented images. Following the method proposed by [26], mainstream SSCD methods employ random strong augmentations to boost the differences between RSI pairs. The core purpose of strong augmentations in SSCD is to generate different views with various uncertainties from the same RSI to simulate diverse perturbations in RSI pairs. The majority of SSCD methods directly incorporate data augmentation strategies from semi-supervised methods in computer vision. However, Yuan et. al [36] have demonstrated that excessively distorted augmentations can adversely impact data distribution and degrade Semi-Supervised Segmentation (SSS) performance. Previous SSCD methods have not noted this standpoint. Further, we argue that the large amount of RSAU in RSI pairs aggravates the multiple uncertainties of the image, leading to an more uncontrollable training process. Several auto-augmentation techniques, notably the simplified AugSeg [37], have seen widespread use in perturbing unlabeled samples. Based on above statement, an effective idea is to increase the quantity of unlabeled RSI pairs while introducing a measured level of uncertainty interference.

Hence, we simulate the variation of uncertainty with multi views by adjusting color jitting such as brightness and contrast before and after changes. The moderately-distorted multi views allow the model to observe a more diverse scenarios. To further ensure reasonable exploitation of uncertain information within the image, we incorporate probabilistic uncertainty maps into the training of this branch. Specifically, we randomly adopt strong intensity-based augmentations for each unlabeled image w to generate dual views, v_1 and v_2. This process is formulated as:

$$v_1, v_2 = \mathcal{A}_s(w) \tag{5}$$

where $\mathcal{A}_s$ consists of horizontal clip, cutmix, and color jitting operations.

The dual-view RSI pairs preserve the change information in corresponding w, yet exhibit more temporal variation uncertainty following the application of $\mathcal{A}_s$. After that, v_1 and v_2 are feed into the network to generate predicted logits $\hat{y}_{v_1}$ and $\hat{y}_{v_2}$. Previous studies establish a direct association between these predictions and the utilization of pseudo-labeling for loss computation. Nevertheless, it is imperative to acknowledge the considerable variance in the reliability of information inherent in distinct unlabeled images. The model's performance is certainly restricted when all images are treated uniformly.

Therefore, we quantify the reliability of each sample based on the uncertainty According to the predictions from the weak branch, we employ the uncertainty-based probabilistic value $\mathcal{H}_c(X_w)$ and $\mathcal{H}_c(X_{wf})$ as coefficients to adjust the loss values. It is noteworthy that each component within $\mathcal{H}_c$ falls within the range of (0, 1), thereby establishing the overall range of $\mathcal{H}_c$ within (0, 2). The algorithm's loss demonstrates a negative correlation with the uncertainty associated with the samples. Therefore, these two coefficients can be expressed as:

$$\mu_w = 1 - \frac{1}{2} \cdot \mathcal{H}_c(X_w) \tag{6}$$

$$\mu_{wf} = 1 - \frac{1}{2} \cdot \mathcal{H}_c(X_{wf}) \tag{7}$$

With the coefficients mentioned above, we compute the loss concerning the pseudo-change map for each view of bitemporal RSI pairs as:

$$\mathcal{L}_{v_1} + \mathcal{L}_{v_2} = \frac{1}{B} \sum_{i=1}^{B} (\mu_w \cdot \mathcal{CE}(\hat{y}_{w,i}, \hat{y}_{v_1,i}) + \mu_w \cdot \mathcal{CE}(\hat{y}_{w,i}, \hat{y}_{v_2,i})) \tag{8}$$

where $\mathcal{L}_{v_1}$ and $\mathcal{L}_{v_2}$ represent the cross-entropy with pseudo-change maps.

In emulation of CutMix [38], our intention is to extend the amalgamation of images from distinct views at the image level. This deliberate fusion aims to alter the uncertainty distribution of RSI pairs, thereby enhancing the capability of foreground targets. Hence, we design a patch mixer based on the probabilistic uncertainty maps from the weak branch. In specific terms, we partition the multi-view RSI pairs into non-overlapping patches, which are similar to chessboards. Then, we turn our attention to each individual patch and automatically adjust the RSI. Further, we mix each patch of dual-view images as follows:

$$v_{12} = \sum_{j=1}^{J} \mathbb{1}(\mathcal{M}^j\{v_1, v_2\}); \tag{9}$$

$$v_{21} = \sum_{j=1}^{J} \mathbb{1}(\mathcal{M}^j\{v_2, v_1\}) \tag{10}$$

where v_{12} and v_{21} represent the images from mixed v_1 and v_2. $\mathbb{1}(\cdot)$ is the indicator function. J denotes the number of patch sequences. $\mathcal{M}^j$ represents a adaptive mixing function with original confidence maps from the weak branch to mix the after-change images from v_1 and v_2. The same operation is applied to the images before the change.

In contrast to MixMIM [39], which utilizes a random mixing mask for self-supervised training, we introduce a novel mixing strategy with confidence maps. It fully leverages the trustworthiness information associated with the pseudo labels to combine foreground targets across different units. Specifically, we adaptively mix each patch according to the original confidence maps of positive predictions from the weak branch as follows:

$$\mathcal{M}^j = ReLU(\sum_{n=1}^{N_j} \hat{c}_n^j / N_j > \tau_c) \tag{11}$$

Here, n represents the pixel within patch j, τ_c is a constant threshold, and N_j denotes the total number of pixels within the j^{th} patch. $\hat{c}_n^j$ represents the prediction confidence of patch j in the corresponding weakly augmented image. $\hat{c}_n^j$ is obtained by filtering the logits $\hat{y}_n^j$ from the weak branch with a threshold set to 0.95.

During the training phase, we input mixed images v_{12} and v_{21} into the encoder-decoder framework network, resulting in the generation of the change map denoted as $\hat{y}_{v_{12}}^j$ and $\hat{y}_{v_{21}}^j$. The v_{12} and v_{21} primarily revise partial temporal variation uncertainty, aiming to retain the inherent change information. Consequently, $\hat{y}_{v_{12}}^j$ is expected to share a similar probability distribution space with $\hat{y}_{v_{21}}^j$.

In the end, we introduce a mixing consistency loss, namely L_c, which is defined as follows:

$$\mathcal{L}_c = 0.5 \cdot (\mu_w + \mu_{wf}) \cdot \mathcal{L}_{cu}(v_{12}, v_{21}) \tag{12}$$

Here, we choose the Smooth L1 loss function as $\mathcal{L}_{cu}$. Therefore, we derive the ultimate loss function for our method, denoted as $\mathcal{L}_v$:

$$\mathcal{L}_{total} = \mathcal{L}_w + \mathcal{L}_{v_1} + \mathcal{L}_{v_2} + \mathcal{L}_c \tag{13}$$

where the μ_c is the mixing temperature factor. It is employed to balance the emphasis among various losses in UML.

IV. EXPERIMENTS

In this section, we begin by introducing the datasets, evaluation metrics, and the experimental setup for our study. Subsequently, we conduct ablation experiments to analyze the core modules of our method and investigate the impact of major hyperparameters. Finally, we present the results of comparative experiments conducted on multi-class generic change detection datasets, *i.e.*, CDD, SYSU, and LEVIR. we conduct ablation experiments to analyze the core modules of our method and investigate the impact of hyperparameters.

A. Datasets

In real-world scenarios, diverse and dynamically changing targets present a more challenging and practical environment. Following it, our experimental focus is on the multi-class binary change detection datasets CDD and SYSU.

1) CDD: It is a multi-class binary change detection dataset, which is presented in 2018 by [40] with various change types. It comprises 16,000 images and they are partitioned into training, testing, and validation sets, with 10,000, 3,000, and 3,000 images. These images are derived from seven pairs of season-varying images with dimensions of 256×256 pixels, each originally sized at 4725×2700 pixels and obtained from Google Earth.

2) SYSU: This is another public multi-class binary change detection dataset that is collected in Hong Kong. It comprises a total of 20,000 orthographic aerial image pairs, each with a spatial size of 256 × 256 pixels and a spatial resolution of 0.5 meters. For the distribution of the dataset into training, validation, and testing sets, we allocate 12,000, 4,000, and 4,000 samples, respectively.

3) LEVIR: Differing from the above two multi-class change datasets, it is a building change detection dataset. It consists of 637 very high-resolution image pairs, encompassing over 30000 geospatial objects. These RSI pairs possess a spatial resolution of 0.5 meters, each measuring 1024 by 1024 pixels. Since it is one of the most widely employed datasets, we keep our dataset settings the same as the standard configurations [24].

4) LEVIR-AU: We collect and annotate 2,848 RSI pairs of disasters in Turkey from the open data of Maxar, each with a resolution of 256x256. These images exhibit significant differences from the building change types found in the LEVIR dataset. Therefore, we incorporate these images into the LEVIR dataset, maintaining the same split scale and enhancing the aleatoric uncertainty in LEVIR. This processed dataset is referred to as LEVIR-AU.

B. Metrics

We performed a comprehensive and objective evaluation of the model's performance by aggregating key assessment metrics from the tasks of change detection and object segmentation. The metrics include change intersection over union (IoU^c), precision, recall, F1 score (F_1), Kappa coefficient (K_c).

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{PRE} = \frac{(\text{TP} + \text{FN}) \times (\text{TP} + \text{FP})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})^2}$$

$$Precision(P) = \frac{\text{TP}}{\text{TP} + \text{FP}}, \; Recall(R) = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$K_c = \frac{\text{OA} - \text{PRE}}{1 - \text{PRE}}, \; F_1 = \frac{2 \times P \times R}{P + R}$$

$$IoU^c = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

Here, TP and TN stand for the count of correctly identified changed and unchanged pixels, respectively. Conversely, FP signifies the number of unchanged pixels erroneously classified as changed, while FN represents the number of changed pixels mistakenly identified as unchanged.

In change detection tasks, evaluation metrics are typically expressed as percentages, ranging between 0% and 100%. All metrics, except for K_c, fall within this percentage range, while the value of K_c ranges from -1 to 1. We maintain a recording precision of three significant digits for all metrics.

C. Experimental Settings

To ensure a fair comparison, we evaluate the SSCD performance of our method against SOTA methods. The experiments are implemented in PyTorch, and the training is conducted on an NVIDIA RTX 3090. Following prior work, we utilize DeepLabV3++ as the base model, with the ASPP serving as the vanilla decoder, and the encoder implemented using ResNet50. In the training scheme, we employ the SGD optimizer with an initial learning rate of 0.01, a weight decay of 10^{-4}, and a momentum of 0.9. For the weak branch, we apply weak data augmentations, including random flip, random crop, and random rescale. For the strong branch, we apply some intensity-based augmentations to add remote sensing aleatoric uncertainty, which contains cutmix, brightness, contrast, saturation, and hue. In addition, all methods undergo training for 80 epochs, with a fixed batch size of 4 across total datasets. During the training of the semi-supervised method, we execute identical data splits independently for each dataset. We randomly sample subsets of 1%, 5%, 10%, and 20% RSI pairs from the training set as labeled data, utilizing the remaining data as unlabeled training data. To ensure a fair comparison, we compare all SSCD methods with the same split training data.

D. Ablation Studies

To verify the effectiveness of each component of our methods, we conduct thorough and comprehensive ablation experiments. We present the outcomes of ablation experiments conducted on the two principal modules of our method to showcase the synergy and efficacy. Subsequently, a meticulous analysis of the primary settings for each module is undertaken. Table I presents the ablation experiments' outcomes on the primary modules. The sup baseline represents the test results obtained when training with only 5% of labeled images. The addition of each module results in a significant improvement of more than 7 points in terms of the IoU^c metric. Remarkably, by combining image- and feature-level mixing strategies, our method achieves peak performance, surpassing the baseline by 14.9 and 11.7 points in IoU^c and F_1, respectively. We add ablation experiments on the mixing consistency loss function $\mathcal{L}_{cu}$, as depicted in Table II. When designing $\mathcal{L}_{cu}$, we notice that remarkable performance can be attained with straightforward paradigm functions. Hence, there's no need for a complex design of the consistency loss to ensure the model's adaptability to diverse scenarios. The outstanding performance serves as a confirmation of the versatility and practicality of our method.

TABLE I
ABLATION STUDIES FOR VARIOUS MODULES ON CDD DATASET WITH 5% LABELED IMAGES. SUP. DENOTES THE SUPERVISED METHOD

UML	PUAL	IoU^c ↑	F_1 ↑
Sup. baseline		61.0	74.6
✔		69.2	81.8
	✔	68.7	81.5
✔	✔	**75.9**	**86.3**

TABLE II
ABLATION STUDIES FOR THE MIXING CONSISTENCY LOSS ON CDD DATASET WITH 5% LABELED IMAGES.

$\mathcal{L}_{cu}$	IoU^c	*Precision*	*Recall*	OA
SmoothL1	75.9	84.0	88.8	96.8
L1	74.4	82.7	88.5	96.6
L2	73.8	82.9	87.0	96.5

TABLE III
ABLATION STUDIES FOR THE FEATURE UNCERTAINTY TYPE ON CDD DATASET WITH 10% LABELED IMAGES.

$\mathcal{N}(\cdot)$	IoU^c	F_1	K_c	OA
Uniform	80.4	89.1	0.872	97.5
Dropout	80.8	89.4	0.878	97.5
Gaussian	79.1	88.4	0.866	97.3
Salt-pepper	79.7	88.7	0.870	97.4

TABLE IV
ABLATION STUDIES FOR HYPERPARAMETERS OF UML ON CDD DATASET WITH 5% LABELED IMAGES.

N	τ_c	IoU^c	*Precision*	F_1	Kc
2x2	0.2	74.1	83.2	85.1	0.829
	0.3	74.2	84.6	85.2	0.831
	0.4	72.4	82.6	84.0	0.816
	0.5	73.2	81.8	84.5	0.820
4x4	0.2	74.4	85.3	85.8	0.831
	0.3	**75.9**	**86.3**	**84.0**	**0.841**
	0.4	72.8	80.9	84.2	0.816
	0.5	73.5	83.3	84.7	0.825
8x8	0.2	74.0	82.1	85.1	0.826
	0.3	73.6	82.1	84.8	0.823
	0.4	73.5	82.0	84.8	0.822
	0.5	73.7	83.7	84.9	0.827

We showcase the results of ablation experiments concerning the feature uncertainty type $\mathcal{N}(\cdot)$ in Table III. During the design of $\mathcal{N}(\cdot)$, we observed that remarkable performance could be attained with straightforward paradigm functions. The findings in Table III underscore the achievement of outstanding performance with simple $\mathcal{N}(\cdot)$ types. Among these different uncertainty types, the addition of Gaussian uncertainty results in a more pronounced degradation of the algorithm. We attribute this phenomenon to the substantial presence of Gaussian and salt-pepper uncertainties in the original RSI pairs. The random addition further disrupts the distribution of the

TABLE V
ABLATION STUDIES WITH DIFFERENT UNCERTAINTY LEVELS ON 5% CDD RSI PAIRS.

MI reduction (%) ↓	IoU^c	F_1	K_c
3±1	75.9	86.3	0.841
6±1	75.6	86.1	0.835
10±1	72.9	83.6	0.805
15±1	71.6	83.4	0.796

associated noise. In contrast, the injection of Dropout enables the model to gain a broader perspective on the situation, such as occlusion, thereby enhancing its robustness.

The Table IV illustrates the impact of the two primary parameters, N and τ_c, in the UML. We utilize three variations of N to create horizontal and vertical image partitions, yielding 2x2, 4x4, and 8x8 patches, respectively. Furthermore, we linearly choose τ over a predetermined interval. As τ decreases, our method exhibits a tendency to swap strongly augmented RSI with dual views. Consequently, the performance demonstrates a upward trajectory, providing evidence for the effectiveness of our modules.

Within UML, the excessively distorted RSI pairs can exert adverse effects by introducing unwarranted uncertainty, thereby disrupting the distribution of the images. Drawing upon information entropy theory, we employ the measure of Mutual Information (MI) reduction between the original and mixed RSI to quantify the rise of uncertainty. Results are demonstrated in Table V. With the growth of distortion in image pairs, the algorithm's performance exhibits a notable deterioration. Hence, the phenomenon of abrupt changes presented in The Table V provides compelling evidence that excessive distortion of a single image results in performance degradation. Combined with Table I, the results illustrate the effectiveness of dual views.

E. Comparison Experiments

To benchmark our approach against the advanced methods across diverse datasets, we incorporate a selection of late SOTA methods, encompassing s4GAN [41], SemiSANet [14], SemiCD [15], RCL [30], FPA [25], and Unimatch [24]. In addition, we complement the algorithm's performance with labeled data only, namely, Sup. baseline. These comparative experimental results of the aforementioned methods are presented in Table VI, VII, VIII, and IX.

The quantitative experiments on the CDD dataset are presented in Table VI. All of these methods use the same data split to ensure the fairness of the experiments. The experimental results demonstrate that our method surpasses other approaches in various cases, affirming the efficacy of uncertainty perception. When training with only 1% labeled RSI pairs, most methods may face challenges in effectively distinguishing and mitigating the uncertainty within the dataset. By partitioning the training process into multiple stages and effectively discarding unreliable unlabeled RSI pairs, RCL [30] exhibits excellent performance, with only 1% labeled RSI pairs. However, its approach to selecting reliable samples is not robust. It may even result in removing discriminative images, as experiments indicate its effectiveness is less than satisfactory in other cases.

As shown in Table VII, we illustrate the comparative experiments on the SYSU dataset. Likewise, our approach achieves SOTA to this general change detection dataset. Notably, RCL exhibits slightly superior performance compared to Unimatch. We acknowledge variations in annotation quantity and quality across different generic datasets, resulting in distinct data domains within SYSU and CDD. These datasets (CDD and SYSU) present a higher level of complexity compared to a single-target variation dataset.

Building change is a typical category among various changes, and numerous methods have primarily been evaluated using the building change dataset LEVIR. We present the results of our experiments in Table VIII. The performance of our method in this experiment notably exceeds that of CDD and SYSU, primarily due to the presence of only one change category. Our method is class-aware, allowing it to excel in handling specific change types. Compared with other methods, ours demonstrates superior performance across all labeled ratios, indicating its robustness and efficacy across different datasets. We augment the data collected in different cases in Turkey to obtain LEVIR-AU, taking into account the relatively homogeneous change types within the LEVIR dataset. Table IX demonstrates the results of the experiment with complex building uncertainties. It's obvious that the efficiency of all evaluated methods has dropped. FPA and SemiCD both decrease by approximately ten points. When there are only a few samples (1% labeled data), PUF executes effectively because it measures the samples' reliability. Our approach demonstrates its outstanding robustness and generalization performance, though it also suffers a decline.

The visualization results for different datasets are presented in Fig. 5, 6, 7, and 8. Through our experiments, we observed that superior detection results can be attained by utilizing merely 10% of labeled data. A selection of test samples is presented in Fig. 5. In Progressive Uncertainty-aware Learning, the utilization of probabilistic uncertainty maps holds significance in the training of subsequent algorithms. Thus, we visualize these maps across different channels in Fig. 6. The first channel represents the probability of belonging to the no-change region, while the second channel signifies the probability of belonging to the change region. These probabilities correspond to columns 4 and 5 in Fig. 6 of each RSI pair, respectively. Moreover, UML introduces robust perturbations to weakly augmented samples, intensifying the uncertainty of the image, and consequently, diminishing the mutual information of bitemporal image pairs. Fig. 7 illustrates the images subjected to various distortions. It is evident that the divergence in color gamut between the bitemporal images becomes increasingly pronounced as the mutual information decreases. Fig. 8 illustrates the trend of coefficients μ_w over an epoch across two datasets. The values in the CDD dataset are higher compared to SYSU. This discrepancy between the two curves can be attributed to SYSU primarily comprising forests, simple urban buildings, and scenes lacking complexity. The soft coefficient can adapt to various data distributions, adaptively assign difficulty to samples, and exhibit excellent

TABLE VI
COMPARISON RESULTS ON CDD DATASET.
THE BEST SCORE IS IN **RED**, THE SECOND BEST SCORE IS IN **BLUE**.

Method	1%			5%			10%			20%		
	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc
Sup. baseline	45.0	60.0	0.473	60.9	74.6	0.680	69.5	80.8	0.769	78.0	86.9	0.847
s4GAN [41]	3.4	6.5	0.054	46.1	63.1	0.591	67.4	80.6	0.780	79.6	88.6	0.870
SemiSANet [14]	44.9	62.0	0.575	66.3	79.2	0.766	73.9	84.7	0.824	80.3	88.7	0.872
SemiCD [15]	34.3	51.1	0.465	66.7	80.0	0.774	74.6	85.5	0.835	80.9	89.5	0.879
RCL [30]	52.9	**69.3**	**0.673**	68.9	81.6	0.798	75.8	86.2	0.847	80.4	89.1	87.7
FPA [25]	47.5	64.4	0.597	69.6	82.1	0.797	75.4	86.0	0.840	79.2	88.4	0.868
Unimatch [24]	**53.4**	68.1	0.607	**73.2**	**83.8**	**0.810**	**79.2**	**87.8**	**0.855**	**86.2**	**92.2**	**0.910**
Ours	**53.1**	**69.4**	**0.637**	**75.9**	**86.3**	**0.841**	**80.8**	**89.4**	**0.878**	**86.3**	**92.4**	**0.912**

TABLE VII
COMPARISON RESULTS ON SYSU DATASET.

Method	1%			5%			10%			20%		
	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc
Sup. baseline	54.3	70.4	0.621	61.8	76.4	0.690	63.8	77.9	0.709	64.4	78.4	0.720
SemiSANet [14]	51.8	68.2	0.588	58.4	73.7	0.654	59.4	74.6	0.668	61.0	75.8	0.683
SemiCD [15]	45.8	62.8	0.515	56.7	72.4	0.652	60.5	75.5	0.694	61.1	75.9	0.700
RCL [30]	49.5	66.2	0.588	65.1	78.9	0.711	64.9	78.7	0.734	66.9	80.2	0.753
FPA [25]	**60.1**	**75.1**	**0.679**	**66.3**	**79.7**	**0.736**	**66.7**	**80.0**	**0.744**	66.9	80.2	**0.747**
Unimatch [24]	56.1	72.3	0.610	65.2	79.7	0.726	65.5	79.9	0.729	**67.5**	**81.0**	0.741
Ours	**61.4**	**77.0**	**0.696**	**66.6**	**80.0**	**0.736**	**67.6**	**81.6**	**0.756**	**68.5**	**81.3**	**0.751**

TABLE VIII
COMPARISON RESULTS ON LEVIR DATASET.

Method	1%			5%			10%			20%		
	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc	IoU^c	F_1	Kc
Sup. baseline	55.1	71.1	0.673	70.1	82.4	0.793	76.8	86.9	0.853	78.7	88.1	0.868
s4GAN [41]	26.4	41.7	0.404	55.9	71.7	0.706	65.2	79.0	0.780	76.2	86.5	0.859
SemiSANet [14]	56.9	72.5	0.714	72.3	83.9	0.832	76.5	86.7	0.861	78.7	88.1	0.875
SemiCD [15]	61.4	76.0	0.751	75.8	86.3	0.855	78.0	87.6	0.870	80.0	88.9	0.884
RCL [30]	62.3	76.7	0.751	73.1	84.5	0.833	75.5	86.0	0.855	76.1	86.4	0.857
FPA [25]	57.2	72.8	0.717	75.8	86.3	0.856	78.7	88.1	0.875	79.6	88.7	0.881
Unimatch [24]	**65.5**	**80.3**	**0.762**	**80.3**	**89.2**	**0.882**	**81.4**	**90.0**	**0.891**	**81.9**	**90.0**	**0.892**
Ours	**77.3**	**87.2**	**0.858**	**82.0**	**90.1**	**0.892**	**83.2**	**90.8**	**0.901**	**83.4**	**91.0**	**0.903**

TABLE IX
COMPARISON RESULTS ON LEVIR-AU DATASET.

Method	1%		5%		10%		20%	
	IoU^c	F_1	IoU^c	F_1	IoU^c	F_1	IoU^c	F_1
SemiCD [15]	47.4	64.3	65.2	78.9	68.6	81.4	71.0	83.0
RCL [30]	53.8	69.9	62.9	77.2	66.4	79.8	69.4	82.0
FPA [25]	49.2	65.9	68.1	81.0	69.9	82.3	71.5	83.4
Unimatch [24]	**57.0**	**72.6**	**71.9**	**83.6**	**72.5**	**83.9**	**73.9**	**85.0**
Sup. baseline	40.1	57.3	61.7	76.4	66.2	79.6	69.6	82.1
Ours	**61.7**	**76.3**	**73.0**	**84.4**	**74.6**	**85.5**	**76.1**	**86.5**

generalization performance.

V. CONCLUSION

This paper investigates the uncertainty which hinders the efficiency of the model, which we refer to as Remote Sensing Aleatoric Uncertainty (RSAU). This form of RSAU is primarily composed of imaging uncertainty and temporal variation uncertainty. Leveraging these insights, we introduce a semi-supervised change detection method, namely, Progressive Uncertainty-aware and -guided Framework (PUF). It is primarily composed of Progressive Uncertainty-aware Learning (PUAL) and Uncertainty-guided Multi-view Learning (UML) for the weak and strong branches, respectively. PUL critically revisits the integration of uncertainty guidance in the training process through both the weak and strong branches. PUAL systematically quantifies the inherent uncertainty in prediction maps of unsupervised samples by designing a uncertainty-aware group decoder. Furthermore, UML simulates a variety of scenarios by generating pairs of images with multiple views and various distortions. The creation of these multi-view RSI pairs is guided by the uncertainty-based probabilistic values from PUAL. We evaluate PUL with ablation experiments on essential elements, supporting our comparative analysis against several state-of-the-art methods on the CDD, SYSU,

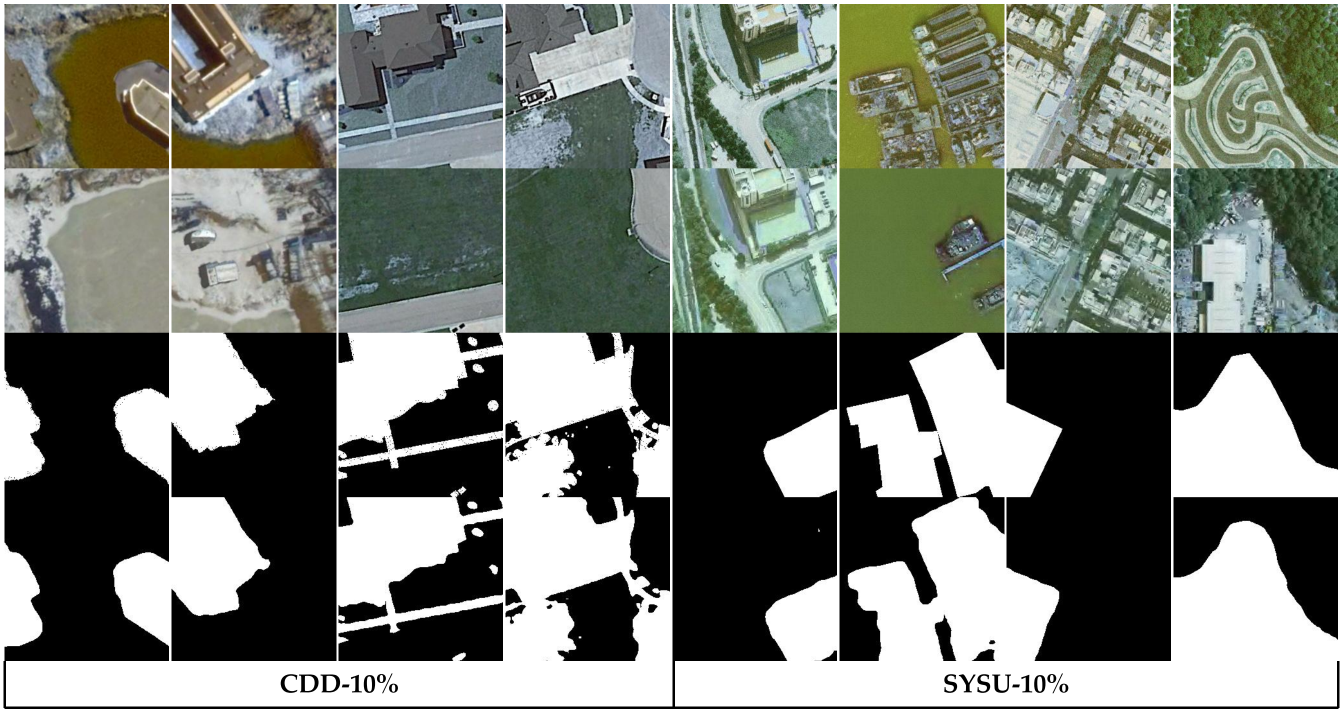

Fig. 5. The visualization of our method on the test dataset with 10% labeled training data. For each RSI pair, the presentation consists of four rows of images depicting pre- and post-change images, predictions, and ground truth.

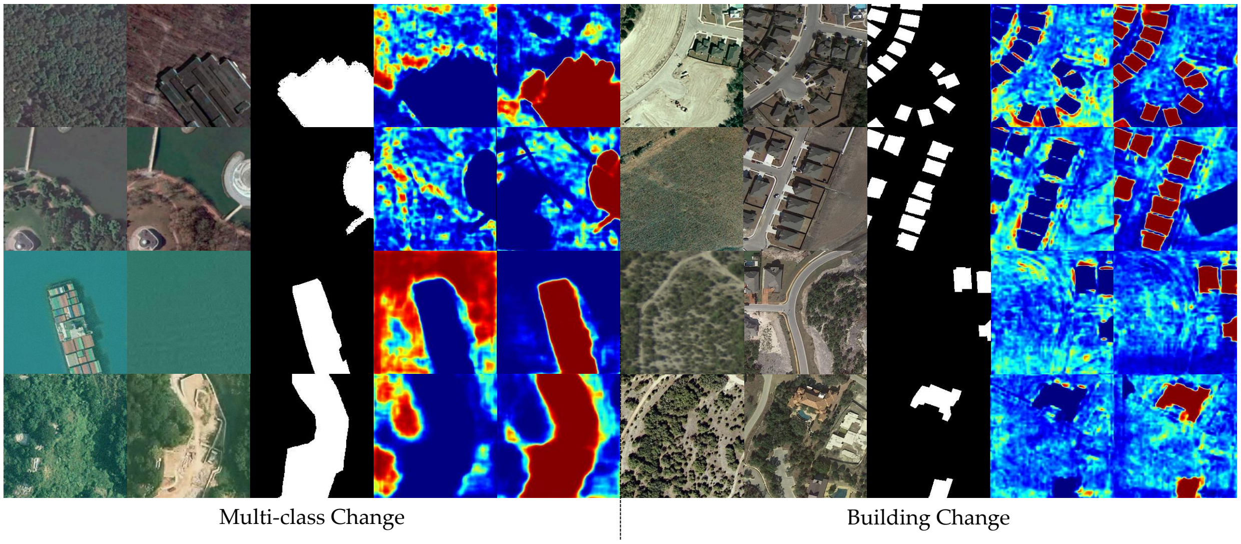

Fig. 6. The visualization of the probabilistic uncertainty maps. The content is organized into three primary panels from multi-class and building datasets of 10% labeled images, respectively. Within each image pair, the presentation comprises five columns, including pre- and post-change images, ground truth, heat maps without change, and heat maps with change.

and LEVIR datasets. To boost the uncertainty of data, we annotate multiple images of building changes, which we refer to as LEVIR-AU. The comprehensive experimental results explicitly confirm PUF's efficacy.

In future work, there is potential for enhancing the introduced PUL by conducting a more thorough exploration of the uncertainty associated with features in the strong branch. Moreover, optimization of the uncertainty-based probabilistic values as a pixel-level map can be pursued to guide subsequent model training more effectively.

References

[1] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14, p. 2291, 2020.

[2] M. Aguilar, M. Saldaña, and F. Aguilar, "Geoeye-1 and worldview-2 pan-sharpened imagery for object-based classification in urban environments," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2583–2606, 2013.

[3] C. Mucher, K. Steinnocher, F. Kressler, and C. Heunks, "Land cover characterization and change detection for environmental monitoring of pan-europe," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1159–1181, 2000.

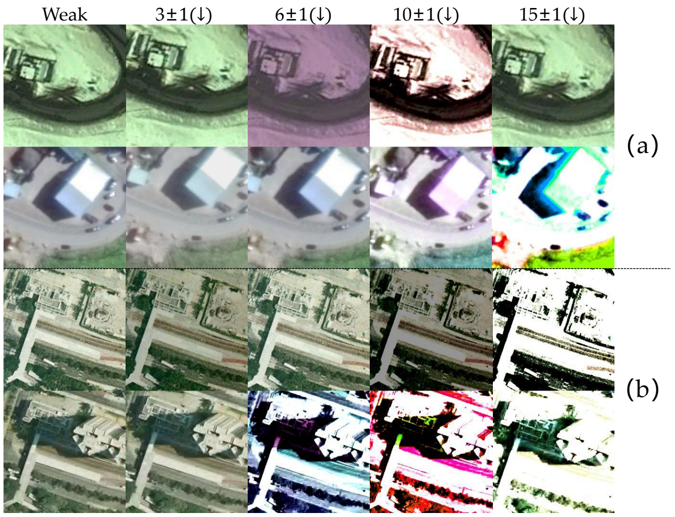

Fig. 7. The visualization of enhanced RSI pairs in the strong branch. The uncertainty of the data increases sequentially from left to right.

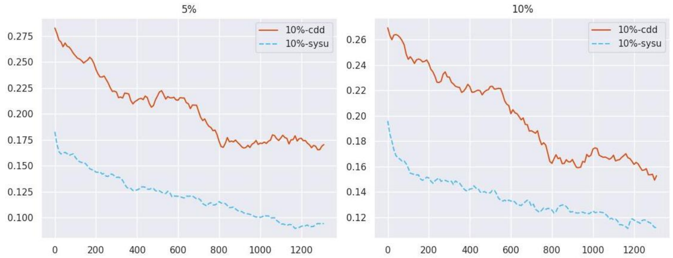

Fig. 8. The tendency of coefficients in the training pipeline with 5% and 10% labeled data.

[4] K. Basavaraju, N. Sravya, S. Lal, J. Nalini, C. S. Reddy, and F. Dell'Acqua, "Ucdnet: A deep learning model for urban change detection from bi-temporal multispectral sentinel-2 satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[5] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 197–212, 2019.

[6] L. Giustarini, R. Hostache, P. Matgen, G. J.-P. Schumann, P. D. Bates, and D. C. Mason, "A change detection approach to flood mapping in urban areas using terrasar-x," *IEEE transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2417–2430, 2012.

[7] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[8] Q. Li, Y. Yuan, and Q. Wang, "Multi-scale factor joint learning for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] Y. Liu, Z. Xiong, Y. Yuan, and Q. Wang, "Transcending pixels: Boosting saliency detection via scene understanding from aerial imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023.

[11] Q. Li, Y. Yuan, X. Jia, and Q. Wang, "Dual-stage approach toward hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 31, pp. 7252–7263, 2022.

[12] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–338, 2010.

[13] 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 Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5891–5906, 2020.

[14] C. Sun, J. Wu, H. Chen, and C. Du, "Semisanet: A semi-supervised high-resolution remote sensing image change detection model using siamese networks with graph attention," *Remote Sensing*, vol. 14, no. 12, p. 2801, 2022.

[15] W. G. C. Bandara and V. M. Patel, "Revisiting consistency regularization for semi-supervised change detection in remote sensing images," *arXiv preprint arXiv:2204.08454*, 2022.

[16] X. Zhang, X. Huang, and J. Li, "Joint self-training and rebalanced consistency learning for semi-supervised change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[17] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *Advances in neural information processing systems*, vol. 30, 2017.

[18] Z. Huang, Y. Liu, X. Yao, J. Ren, and J. Han, "Uncertainty exploration: Toward explainable sar target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[19] N. Asadi, K. A. Scott, A. S. Komarov, M. Buehner, and D. A. Clausi, "Evaluation of a neural network with uncertainty for detection of ice and water in sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 247–259, 2020.

[20] N. Saberi, K. A. Scott, and C. Duguay, "Incorporating aleatoric uncertainties in lake ice mapping using radarsat–2 sar images and cnns," *Remote Sensing*, vol. 14, no. 3, p. 644, 2022.

[21] S. Mayr, I. Klein, M. Rutzinger, and C. Kuenzer, "Determining temporal uncertainty of a global inland surface water time series," *Remote Sensing*, vol. 13, no. 17, p. 3454, 2021.

[22] X. Chen, K. A. Scott, L. Xu, M. Jiang, Y. Fang, and D. A. Clausi, "Uncertainty-incorporated ice and open water detection on dual-polarized sar sea ice imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[23] X. Zhang, Y. Luo, and L. Hu, "Semi-supervised sar atr via epoch- and uncertainty-aware pseudo-label exploitation," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[24] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7236–7246.

[25] X. Zhang, X. Huang, and J. Li, "Semisupervised change detection with feature-prediction alignment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[26] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[27] P. P. De Bem, O. A. de Carvalho Junior, R. Fontes Guimarães, and R. A. Trancoso Gomes, "Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks," *Remote Sensing*, vol. 12, no. 6, p. 901, 2020.

[28] S. Fang, K. Li, J. Shao, and Z. Li, "Snunet-cd: A densely connected siamese network for change detection of vhr images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[29] Z. Li, C. Yan, Y. Sun, and Q. Xin, "A densely attentive refinement network for change detection based on very-high-resolution bitemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[30] J.-X. Wang, T. Li, S.-B. Chen, J. Tang, B. Luo, and R. C. Wilson, "Reliable contrastive learning for semi-supervised change detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[31] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Transactions on Cybernetics*, vol. 51, no. 8, pp. 3889–3900, 2021.

[32] Y. Liu, Z. Xiong, Y. Yuan, and Q. Wang, "Distilling knowledge from super-resolution for efficient remote sensing salient object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023.

[33] P. Ortiz, M. Orescanin, V. Petković, S. W. Powell, and B. Marsh, "Decomposing satellite-based classification uncertainties in large earth science datasets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[34] L. Alagialoglou, I. Manakos, M. Heurich, J. Červenka, and A. Delopoulos, "A learnable model with calibrated uncertainty quantification for estimating canopy height from spaceborne sequential imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[35] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmen-

tation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[36] J. Yuan, Y. Liu, C. Shen, Z. Wang, and H. Li, "A simple baseline for semi-supervised semantic segmentation with strong data augmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8229–8238.

[37] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 350–11 359.

[38] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[39] J. Liu, X. Huang, Y. Liu, and H. Li, "Mixmim: Mixed and masked image modeling for efficient visual representation learning," *arXiv preprint arXiv:2205.13137*, 2022.

[40] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 565–571, 2018.

[41] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 4, pp. 1369–1379, 2019.

Jinhao Shen is currently pursuing the M.S. degree with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include remote sensing and deep learning.

Cong Zhang received the M.S. degree from the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. He is currently pursuing a Ph.D. with the Department of Electronic and Information Engineering at The Hong Kong Polytechnic University in Hong Kong. His research interests include computer vision and machine learning.

Mingwei Zhang received the B.E. degree in automation from Zhengzhou University, Zhengzhou, China, in 2021, and the M.S. degree from the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China, in 2024. He is currently pursuing the Ph.D. degree with the School of Computer Science, Northwestern Polytechnical University, Xi'an, China. His research interests include remote sensing image acquisition and processing.

Qiang Li received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China in 2022. He is currently a postdoc with the School of Electronic Engineering, Xidian University, Xi'an. His research interests include remote sensing image processing and computer vision.

Qi Wang (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition and remote sensing.