Weakly Supervised Solar Panel Mapping via Uncertainty Adjusted Label Transition in Aerial Images

Jue Zhang, *Graduate Student Member, IEEE*, Xiuping Jia, *Fellow, IEEE*, Jun Zhou, *Senior Member, IEEE*, Junpeng Zhang, *Member, IEEE*, and Jiankun Hu, *Senior Member, IEEE*

***Abstract*— This paper proposes a novel uncertainty-adjusted label transition (UALT) method for weakly supervised solar panel mapping (WS-SPM) in aerial Images. In weakly supervised learning (WSL), the noisy nature of pseudo labels (PLs) often leads to poor model performance. To address this problem, we formulate the task as a label-noise learning problem and build a statistically consistent mapping model by estimating the instance-dependent transition matrix (IDTM). We propose to estimate the IDTM with a parameterized label transition network describing the relationship between the latent clean labels and noisy PLs. A trace regularizer is employed to impose constraints on the form of IDTM for its stability. To further reduce the estimation difficulty of IDTM, we incorporate uncertainty estimation to first improve the accuracy of noisy dataset distillation and then mitigate the negative impacts of falsely distilled examples with an uncertainty-adjusted re-weighting strategy. Extensive experiments and ablation studies on two challenging aerial data sets support the validity of the proposed UALT.**

***Index Terms*— Weakly supervised learning, label noise, solar panel mapping, uncertainty estimation, aerial images.**

I. Introduction

ENERGY demand has grown significantly in recent decades and solar energy has been recognized as one of the best energy sources for the future world [1]. Solar panels are the key component in solar photovoltaic systems, converting sunlight into electricity. The installation of solar panels on households' rooftops shows an exponential increase over the past several years. Collecting comprehensive deployment information including the locations, sizes, and power capacity can help the government and power companies gain a quantitative understanding of solar energy utilization [2].

Aiming at low-cost monitoring, automatic solar panel mapping leverages high-resolution remote sensing imagery and machine learning techniques to localize solar panels and produce a binary prediction for each pixel in the input image. A collection of research has been devoted to this area in the past decade. Most pioneer works [3], [4], [5] take advantage of hand-made features and classical classifiers including support vector machine and random forest to implement classification. Due to the significant progress made by the deep convolutional neural networks (CNNs), deep learning-based approaches [6], [7], [8], [9], [10] greatly surpass traditional methods in performance. These CNN-based methods are mostly developed under the paradigm of fully supervised learning, which requires a huge number of hand-labeled pixel-wise ground truth data which are extremely laborious and time-consuming to get [11]. To reduce the labeling workload, weakly supervised data with weaker forms of annotations are adopted [12], [13], [14], [15]. However, due to the absence of precise object locations and shapes, learning predictive models in this scenario is much more challenging. To bridge the gap between image-level labels and pixel-wise labels, the alternative training scheme provides a solution: weakly supervised object localization (WSOL) [16], [17], [18] is firstly utilized to propagate image-level annotations to pixel-level labels, which are subsequently taken as pseudo labels (PLs) to build predictive models. Fig. 1 illustrates the pipeline of the alternative training scheme. The quality of PLs, however, is far from perfect, as most WSOL methods focus on discriminative features and may fail to provide integral object regions with precise boundaries. As reported in many studies [13], [19], such inaccurate supervision may lead to performance degradation when the alternative training scheme is applied.

Recently, weakly supervised solar panel mapping (WS-SPM) has gained attention from the remote sensing community [20], [21], [22]. The majority of the studies achieve progress by modifying WSOL methods [20] or resort to label correction and regularizers with the alternative training scheme [21], [23]. Despite the improvement achieved,

Manuscript received 7 November 2022; revised 8 August 2023 and 5 October 2023; accepted 15 November 2023. Date of publication 8 December 2023; date of current version 19 January 2024. This work was supported by the Australian Research Council (ARC) under Grant DP190103660 and Grant DP200103207. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yannick Berthoumieu. *(Corresponding author: Jue Zhang.)*

Jue Zhang and Xiuping Jia are with the School of Engineering and Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: jue.zhang@student.unsw.edu.au; x.jia@adfa.edu.au).

Jun Zhou is with the School of Information and Communication Technology, Griffith University, Nathan, QLD 4111, Australia (e-mail: jun.zhou@griffith.edu.au).

Junpeng Zhang is with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, International Research Center of Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: junpengzhang@xidian.edu.cn).

Jiankun Hu is with the School of Systems and Computing, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: j.hu@adfa.edu.au).

Digital Object Identifier 10.1109/TIP.2023.3336170

WS-SPM faces three challenges: inaccurate object coverage, severe object mismatch, and inadequate ability to separate co-occurring objects. The unavoidable noise in PLs greatly degenerates the accuracy and reliability of predictive models, particularly the deep CNN-based ones, due to the memorization effects and over-fitting issues.

From the perspective of combating noisy PLs, most works for WSL in remote sensing images focus on how to mitigate the side effects of the label noise in PLs instead of modeling it explicitly. Studies in the field of image classification [24], [25], [26] show that modeling the label noise, as well as the generation process of noisy labels, are more effective in developing a theoretically guaranteed predictive model. With the assumptions on the label noise distribution, the noisy labels are assumed to be clean labels corrupted by the noise. Modeling the noise in a more realistic way, the instance-dependent label transition matrix (IDTM) describes the mislabeled probability affected by both the class labels and the instance features and has been widely adopted to reveal the label corruption process. Learning IDTM, however, is an ill-posed problem because the clean class posterior is latent and unobservable [27]. One feasible way is to learn it on distilled examples [25] with reasonable assumptions such as noise rate upper bound. Although the noise rate upper bound assumption and noisy dataset distillation enable the estimation of IDTM, there are limitations to be addressed: due to the over-fitting problem during the training process of the noisy classifier, the collection of distilled examples is inevitably biased, which will lead to increased estimation error. Another issue worthy of attention is the high degree of freedom of IDTM. Particularly, for the dense prediction task, each pixel has its corresponding IDTM and each element in the IDTM can be presented as a function of the instance feature. How to reduce the computational complexity is the key point in approximating IDTM accurately.

To cope with the problems, we introduce the predictive uncertainty to indicate the pixels more likely to be affected by the label noise and propose an uncertainty-adjusted label transition (UALT) method for WS-SPM from high-resolution aerial images. The goal of our work is to learn an accurate mapping model by estimating IDTM in a parametric way. Specifically, we consider the PLs generated by GradCAM as noisy labels with instance-dependent label noise (IDLN). Under the noisy data, an uncertainty estimation network (UEN) is constructed to generate the initial mapping results coupled with the predictive uncertainty levels. Then, we collect the distilled examples from the initial mapping results to learn a label transition network (LTN). To make CNN-based IDTM estimable, we propose an effective trace regularizer and utilize an uncertainty-adjusted (UA) re-weighting strategy to alleviate the negative impacts of falsely labeled pixels. Finally, with the well-estimated IDTM, a target mapping network can be developed. The contributions of the proposed method are summarized as follows:

1) To cope with the performance degeneration caused by label noise in PLs, we formulate the WS-SPM as a label-noise learning problem and propose a UALT method to learn a statistically consistent mapping model by approximating the instance-dependent label transition between the clean labels and noisy PLs with a parameterized LTN.
2) We develop an uncertainty estimation network based on Monte Carlo dropout to give initial mapping results with corresponding uncertainty levels under PLs. By introducing the stochastic process, UEN is capable of alleviating the over-fitting issue and improving mapping accuracy. The improved initial results will benefit the collection of distilled examples as the initial predictions with higher uncertainty levels will be assigned with relatively lower prediction values by UEN.
3) We propose a UA re-weighting strategy to adjust the contributions of pixels with different uncertainty levels. The uncertainty levels quantitatively indicate the likelihood of UEN predictions being influenced by labeling errors in PLs, i.e., label noise. Pixels with higher uncertainty levels are supposed to have a stronger probability of being impacted by the label noise and wrongly inferred. By UA re-weighting, the detrimental impact of potentially mislabeled distilled examples will be further mitigated.
4) To reduce the degree of freedom of the IDTM, we propose a trace regularizer to impose constraints on the form of IDTM. By maximizing the trace of IDTM, the trace regularizer encourages the LTN to focus on ambiguous areas such as object boundaries. The trace regularizer significantly reduces the complexity of estimating IDTM and makes the training process of LTN more stable.

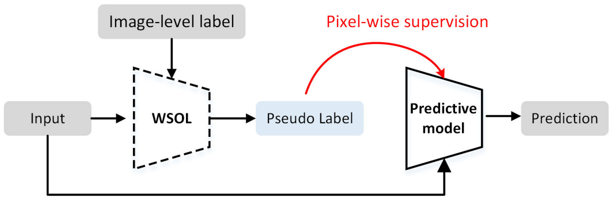


Fig. 1. The pipeline of the alternative training scheme. WSOL methods are used to propagate the image-level labels into pixel level and generate pixel-wise annotations, which subsequently serve as the pseudo labels for the training processing of a predictive model.

The rest of this paper is organized as follows. Section II presents a brief review of solar panel mapping, weakly supervised learning, and learning with label noise. Section III provides problem formulation and a detailed introduction to the proposed UALT. The experiment and analysis are presented in Section IV. The conclusion is drawn in Section V.

II. Related Work

A. Solar Panel Mapping

Pioneer works [3], [4], [5], [28] draw inspirations from classical object detection algorithms, which first generate hand-made features and then utilize classifiers including support vector machine and random forest to implement classification. To support the study on solar panel mapping, data sets [29] collected from satellite images are also created. Due to the significant progress made by the deep convolutional neural networks (CNNs), deep learning-based approaches

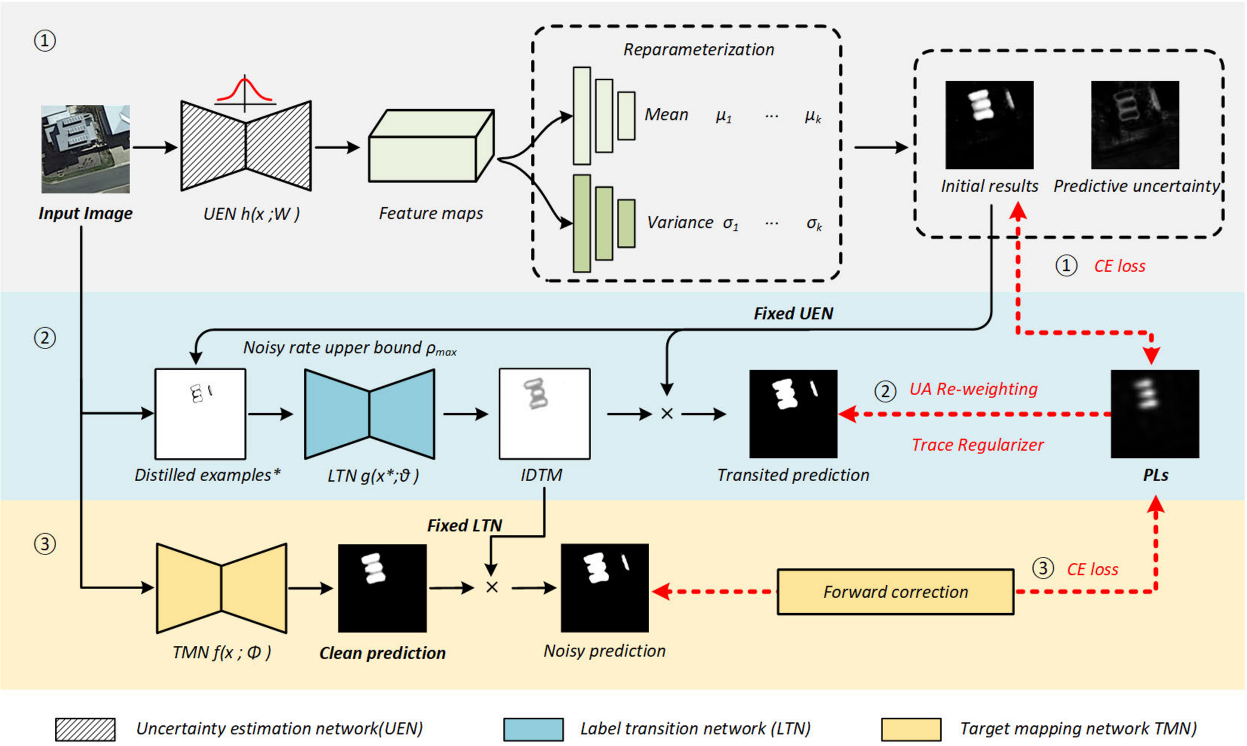


Fig. 2. An illustration of the proposed uncertainty adjusted label transition. The black solid lines represent the feed-forward process, and the dashed lines represent the back-propagation procedure. The purpose of the proposed UALT is to learn the clean predictor $f(\mathbf{x}, \phi)$ under only noisy PLs. The training process of the proposed method involves three CNNs, i.e., UEN, LTN, and TMN, trained in sequence. In the inference phase, only the well-trained TMN will be used for forward propagation to provide mapping results.

greatly surpass traditional methods in performance. Yuan et al. first employed the ConvNet method to extract the spatial content of solar panels [6]. Then, classical deep CNN architectures such as VGG [30], SegNet [31], Unet [32] and DeepLabv3+ have been introduced to identify solar panels at pixel level and made dense predictions [7], [8], [9], [10]. These methods aforementioned are developed under the paradigm of fully supervised learning.

B. Weakly Supervised Learning

The aim of weakly supervised learning (WSL) is to take advantage of weak supervision, such as image-level labels [12], [13], [14], point-level labels [33], bounding-box-based labels [34] and scribble-based labels [15], [35] to build predictive models with performance close to fully supervised ones. The main difficulty in WSL is how to propagate annotations in weak forms to densely annotated labels. The alternative training scheme is a popular scheme in WSL: class-specific object localization generated by WSOL methods such as class activation mapping (CAM) [16], gradient-weighted class activation mapping (GradCAM) [17] and LayerCAM [18] can be subsequently taken as PLs to train predictive models. Considering the unstable and inconsistent quality of PLs, strategies such as gated structure-aware loss [15], conditional random field-based regularizer [36], and label refinement [37], [38] are developed to address inaccurate object localization and poor boundary preservation.

Most of the WSL methods for remote sensing images follow the typical alternative training scheme [39], [40], [41] and seek to achieve superior performance by refining the PLs [42], introducing robust learning strategies [43] and designing consistency regularizer [44]. For example, Fang et al. [42] proposed a new adversarial climbing strategy to assist CAM to generate better building pseudo masks that are further refined with pairwise semantic affinities. Zhang and Ma [43] incorporated curriculum learning with the feedback saliency analysis network to produce reliable predictions for residential areas. Xu and Ghamisi [44] started from point-level annotations, iteratively expanded annotated areas by selecting confident pixels, and finally leveraged the consistency loss to cope with the potential misguide from the inaccurate expanded annotations.

WSL for solar panel mapping from high-resolution aerial images has not been extensively explored. Yu et al. [20] proposed the first WSL-based solar panel mapping method named “DeepSolar” with only image-level labels, which modified CAM with a greedy layer-wise training mechanism to produce mapping results with clear boundaries. Zhang et al. [23] proposed a residual aggregated network (RAN), with the network architecture specially designed for preserving detailed object boundaries by residual feature aggregation. Zhang et al. [21] considered the inconsistent quality of PLs and proposed a confidence-aware loss and self-paced label correction strategy to adjust the contribution of quality-varying PLs.

C. Learning with Label Noise

Learning with label noise has gained significant attention in the machine learning community recently, with most of the work focusing on the image classification task. The algorithms combating noisy labels generally begin with placing assumptions on the label noise distribution. There are numerous label noise models of great popularity, such as random classification noise [45], class-conditional label noise [46], and IDLN [24], [25], [47], [48]. As a more realistic case, IDLN can be described by the IDTM, which is characterized by the mislabeled probability affected by both the clean class and the instance features. With the noise model, the IDTM describing the corruption process has been widely exploited to show the relationship between the input, latent clean labels, and noisy labels. It is noteworthy that given only noisy data, it is ill-posed to estimate the IDTM, as the true labels are latent in this case. A number of studies attempted to mitigate this issue by learning with noisy dataset distillation [24], [25]. The noisy dataset distillation refers to a process of collecting distilled examples out of the noisy data, i.e., examples whose labels are identical to the one assigned by the Bayes optimal classifier under the clean distribution [25]. With the label noise upper bound assumption [25], the noisy dataset distillation makes it possible to learn a classifier that can converge to the Bayes optimal classifier with clean labels. Furthermore, involving high computational complexity, estimation of the IDTM is a non-trivial task. Making reasonable assumptions on the transition matrix [26] has revealed an effective way to mitigate this issue.

III. Methodology

In the proposed UALT, we formulate the WS-SPM in a label-noise learning framework and develop a statistically consistent predictive model by estimating the IDTM in a parametric way. As shown in Fig. 2, the proposed UALT consists of three parts: uncertainty estimation network, label transition network, and target mapping network. The uncertainty estimation network is specifically designed to generate initial mapping results coupled with predictive uncertainty. A label transition network is then developed to approximate the IDTM by learning from the distilled examples. To effectively reduce the estimation errors as well as the estimation complexity, an uncertainty-adjusted re-weighting strategy, and the trace regularizer are proposed. Finally, with the estimated IDTM, the target mapping network is optimized by the forward correction to obtain the mapping results.

A. The Label-Noise Learning Framework

In this paper, we assume that weakly supervised solar panel mapping can be formulated as a label-noise learning problem, and the PLs given by GradCAM are noisy labels transformed from clean labels randomly corrupted with the instance-dependent noise. Given two random variables $\mathbf{X}$ and $\widetilde{\mathbf{Y}}$ representing variables for the input and corresponding PL, we denote $\widetilde{D}$ as the noisy joint distribution of $(\mathbf{X}, \widetilde{\mathbf{Y}}) \in \mathcal{X} \times \widetilde{\mathcal{Y}}$. With d-dimension input feature in a binary case, we have feature space $\mathcal{X} \subseteq \mathbb{R}^d$, and label space $\widetilde{\mathcal{Y}} = \left\{\boldsymbol{e}^i : i \in \{-1, +1\}\right\}$. In our work, $\boldsymbol{e}^{+1}$ denotes the positive label and $\boldsymbol{e}^{-1}$ is the negative one. $(\mathbf{x}, \widetilde{\mathbf{y}})$ is an observation sample drawn from the noisy joint distribution $\widetilde{D}$. D denotes the clean distribution of pair-wise variables $\mathbf{X}$ and corresponding clean label $\mathbf{Y}$, which is latent and unobservable. The goal of the label-noise learning is to disentangle clean labels $\mathbf{y}$ from the noisy labels $\widetilde{\mathbf{y}}$ with the noisy data set $\widetilde{\mathcal{D}}$: $\{(\mathbf{x}, \widetilde{\mathbf{y}})^n\}_{n=1}^N$. N is the total number of training samples.

We consider the label noise in an instance-dependent setting, where each clean label $\mathbf{y}$ is supposed to flip into noisy label $\widetilde{\mathbf{y}}$ randomly with the probability $P(\widetilde{\mathbf{Y}} \mid \mathbf{Y}, \mathbf{X})$. Different from class-conditional noise, instance-dependent noise is a more general approximation of corruption in the real world, assuming the flip rate varies with the actual instance feature.

The instance-dependent transition matrix $T(\mathbf{x})$ is defined to bridge the posterior probabilities of the noisy and clean joint distribution. The element $T_{i,j}(\mathbf{x})$ in $T(\mathbf{x})$ represents the probability of the clean label $\mathbf{y} = \boldsymbol{e}^i$ flipped into the noisy one $\widetilde{\mathbf{y}} = \boldsymbol{e}^j$:

$$\forall i, j \quad T_{i,j}(\mathbf{x}) = P(\widetilde{\mathbf{Y}} = \boldsymbol{e}^j \mid \mathbf{Y} = \boldsymbol{e}^i, \mathbf{X} = \mathbf{x}). \tag{1}$$

With transition matrix $T(\mathbf{x}) = \left\{T_{i,j}(\mathbf{x})\right\} \in [0, 1]^{2\times 2}$, the noisy class-posterior probability $P(\widetilde{\mathbf{Y}} \mid \mathbf{X})$ can be computed as follows:

$$P(\widetilde{\mathbf{Y}} = \boldsymbol{e}^j \mid \mathbf{X} = \mathbf{x}) = \sum_i T_{i,j}(\mathbf{x}) \cdot P(\mathbf{Y} = \boldsymbol{e}^i \mid \mathbf{X} = \mathbf{x}). \tag{2}$$

As the noisy class-posterior probability $P(\widetilde{\mathbf{Y}} = \boldsymbol{e}^j \mid \mathbf{X} = \mathbf{x})$ can be estimated by exploring the noisy data set, the clean class-posterior probability is expected to be inferred with an accurate $T(\mathbf{x})$. By minimizing the empirical risk on the noisy predictions $\bar{\mathbf{y}}$ and PLs $\widetilde{\mathbf{y}}$, we can obtain clean distribution $P(\mathbf{Y}|\mathbf{X} = \mathbf{x})$ by learning a mapping $f(\mathbf{x}; \phi)$ with an estimator for IDTM $T(\mathbf{x}; \theta)$ parameterized by θ:

$$\begin{aligned}\min_{\phi,\theta} R(\phi, \theta) = &- \sum_{\mathbf{x}\in\widetilde{\mathcal{D}}} \sum_i (\mathbf{1}\left\{\widetilde{\mathbf{y}}_i = 1\right\} \\ &\cdot log(\sum_s T_{s,i}(\mathbf{x}; \theta) \cdot f_s(\mathbf{x}; \phi)),\end{aligned} \tag{3}$$

where $\widetilde{\mathbf{y}}_i$ denotes the probability for class i in $\widetilde{\mathbf{y}}$. $f_s(\mathbf{x}; \phi)$ denotes the learned clean posterior probability for class s. ϕ is the parameter set for the estimator f. $\mathbf{1}\left\{\widetilde{\mathbf{y}}_i = 1\right\}$ is the indicator function defined as follows:

$$\mathbf{1}\left\{\widetilde{\mathbf{y}}_i = 1\right\} := \begin{cases} 1, & if \quad \widetilde{\mathbf{y}}_i = 1 \\ 0, & if \quad \widetilde{\mathbf{y}}_i = 0 \end{cases} \tag{4}$$

This is also called the forward correction [46]. Intuitively, when minimizing the object function in Eq.(3), we first learn an estimator $g(\mathbf{x}; \theta)$ for the IDTM and train a target mapping network $f(\mathbf{x}; \phi)$ with $g(\mathbf{x}; \theta)$ fixed.

B. Predictive Uncertainty Estimation

Uncertainty of a predictive model usually refers to the occasions where the predictions of this model are not always accurate and cannot be trusted blindly. This is a common case in deep learning-based models, which are not sufficiently robust to over-fitting. Hence, it is of great importance to understand when the models will give uncertain predictions. In the context of modeling, the uncertainty can be categorized into either aleatory or epistemic [49]. Aleatory uncertainty presents the intrinsic noise in the observations while epistemic uncertainty is defined as that being caused by the model. Under the theory of Bayesian deep learning, uncertainty estimation makes it possible to quantify the noise inherent in both training data and the model. Although various methods were proposed to measure these two kinds of uncertainty separately [50], [51]. In the setting of WSL under noisy data, modeling both in a synthesized manner can be more meaningful. Inspired by Kendall's work [51], we integrate the measurement of both aleatory and epistemic uncertainty in a unified uncertainty estimation network $h(\mathbf{x}; \mathbf{W})$ under the noisy data set $\widetilde{\mathcal{D}}$.

To capture the noise in the training data, we place a Gaussian distribution over the UEN output and construct two branches producing mean and variance. Let μ be the output of the mean branch and σ^2 be the output of the variance branch:

$$\left\langle \mu, \sigma^2 \right\rangle = h(\mathbf{x}; \mathbf{W}), \tag{5}$$

where $h(\mathbf{x}; \mathbf{W})$ denotes the UEN with input $\mathbf{x}$ and parameter set $\mathbf{W}$.

By reparameterization, we have $\hat{\mathbf{p}}$:

$$\hat{\mathbf{p}}|\mathbf{W} \sim \mathcal{N}(\mu, \sigma^2), \tag{6}$$

$$\hat{\mathbf{p}} = \mu + \sigma \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{7}$$

where $\mathcal{N}(0, \mathbf{I})$ is a standard Gaussian distribution.

To model the epistemic uncertainty, the network architecture of UEN follows the Bayesian convolutional neural network with Monte-Carlo dropout, which adds dropout after all convolutional layers and averages the results of stochastic feed-forward at the testing time. Gal and Ghahramani [50] revealed that Monte Carlo (MC) dropout in CNN-based models can be interpreted as an approximation to the well-known Gaussian process in Bayesian deep learning. By repeating the random forward process K times, we approximate the output of the network by averaging the results from the mean branch:

$$\hat{\mu} \approx \frac{1}{K}\sum_{k=1}^{K} \mu_k, \tag{8}$$

where the sub-script k denotes the k th forward process.

We use the predictive variance to estimate the corresponding uncertainty:

$$\psi \approx \frac{1}{K}\sum_{k=1}^{K} \sigma_k^2 + \frac{1}{K}\sum_{k=1}^{K} \mu_k^2 - \left(\frac{1}{K}\sum_{k=1}^{K} \mu_k\right)^2, \tag{9}$$

where the first term is the average of the predictive variance, indicating the potential noise in the predictions. The second term measures the uncertainty of the model parameters, i.e. epistemic uncertainty.

The loss of the UEN is defined as:

$$\log \mathbb{E}_{\hat{\mathbf{p}} \sim \mathcal{N}(\mu, \sigma^2)} \left[\hat{\mu}\right]. \tag{10}$$

In practice, to avoid the training degeneration with infinite uncertainty, we combine the cross entropy loss with a regularization term to impose constraints on the value of uncertainty ψ:

$$\mathcal{L}_h = \mathcal{L}_{ce}(\widetilde{\mathbf{y}}, \hat{\mu}) + \tau \cdot \psi^2, \tag{11}$$

where τ is the weight for the regularization term.

In the testing phase, we run the stochastic feed-forward pass K times, and obtain the initial mapping results $\hat{\mu}$ and the predictive uncertainty ψ by Eqs. (8) and (9).

C. Uncertainty Adjusted Label Transition

1) Learning With Distilled Examples: To reduce the difficulties in estimating IDTM, a reasonable assumption for the IDN is the noise rates upper bounds [25]:

$$\rho_i(\mathbf{x}) = P(\widetilde{\mathbf{Y}} \neq \boldsymbol{e}^i \mid \mathbf{Y} = \boldsymbol{e}^i, \mathbf{X} = \mathbf{x}), \tag{12}$$

$$0 \leq \rho_i(\mathbf{x}) \leq \rho_{imax} < 1, i = \{+1, -1\}, \tag{13}$$

$$0 \leq \rho_{+1}(\mathbf{x}) + \rho_{-1}(\mathbf{x}) < 1. \tag{14}$$

where $\rho_i(\mathbf{x})$ is the noise rate for class i. Given the input $\mathbf{X} = \mathbf{x}$, it is defined as the probability that the clean label $\mathbf{Y} = \boldsymbol{e}^i$ flips into the corrupted one $\widetilde{\mathbf{Y}} \neq \boldsymbol{e}^i$. In our work, $\rho_{+1}(\mathbf{x})$ and $\rho_{-1}(\mathbf{x})$ denote the noise rate for the foreground and background classes, respectively. This indicates that the noise rate is dependent on both the true label $\mathbf{Y}$ and the input $\mathbf{X}$. $\rho_{imax}, i = \{+1, -1\}$ denotes the upper bounds of noise rates for class i. The $\rho_{+1} + \rho_{-1} < 1$ requirement means that the dataset is not fully corrupted, and there is as least something left for the model to learn, which is a standard analysis in the class-conditional setting [52].

With this assumption, distilled examples satisfying the following rules can be collected out of noisy data:

$$\begin{cases} \tilde{\eta}(\mathbf{x}) < \dfrac{1 - \rho_{+1\,\max}}{2}, \\ \tilde{\eta}(\mathbf{x}) > \dfrac{1 + \rho_{-1max}}{2}, \end{cases} \tag{15}$$

where $\tilde{\eta}(\mathbf{x}) = P_D(\widetilde{\mathbf{Y}} = e^{+1} \mid \mathbf{X} = \mathbf{x})$. Although $\tilde{\eta}(\mathbf{x})$ is inaccessible practically, we can approximate it by training a noisy classifier with the noisy class posterior probability $P_{\widetilde{D}}(\widetilde{\mathbf{Y}} \mid \mathbf{X} = \mathbf{x})$. Then, the inferred label will be assigned to distilled examples $\mathbf{x}^*$ as the Bayes optimal label $\check{\mathbf{y}}^*$. The details about the theoretical guarantee can be found in Cheng's and Yang's work [24], [25]. In our work, we take UEN as an approximation to the $\tilde{\eta}(\mathbf{x})$. The initial mapping results of UEN are leveraged to obtain the distilled examples.

In the training phase of the LTN, we utilize the distilled subset $\mathcal{D}^* := \left\{(\mathbf{x}^*, \widetilde{\mathbf{y}}^*, \check{\mathbf{y}}^*)^m\right\}_{m=1}^{M}$ and aim to train an estimator g parameterized by θ. With Eq. (2), by minimizing the loss between the predicted noisy results $\bar{\mathbf{y}}^*$ and noisy labels $\widetilde{\mathbf{y}}^*$, the LTN can model the transition from clean labels to noisy PLs by a corrected cross-entropy loss $\mathcal{L}_{cce}(\widetilde{\mathbf{y}}^*, T(\mathbf{x}^*; \theta), \check{\mathbf{y}}^*)$:

$$\begin{aligned} \min_{\theta} \quad & \mathcal{L}_{cce} = -\sum_{\mathbf{x}^* \in \mathcal{D}^*} \sum_{i} \mathbf{1}\left\{\widetilde{\mathbf{y}}_i^* = 1\right\} \cdot \log(\bar{\mathbf{y}}_i^*), \\ s.t. \quad & \sum_{j} T_{i,j}(\mathbf{x}^*, \theta) - 1 = 0, \end{aligned} \tag{16}$$

where $\mathbf{1}\{\cdot\}$ denotes indicator function. $T(\mathbf{x}^*; \theta)$ is the predicted IDTM given by LTN with input $\mathbf{x}^*$ and parameter set θ.

The restriction term $\sum_j T_{i,j}(\mathbf{x}^*, \theta) - 1 = 0$ comes from the definition of the transition matrix T. For brevity, here we use $T_{i,j}$ to denote $T_{i,j}(\mathbf{x}^*, \theta)$. In the binary class, we have:

$$\forall\ i,\ \ T_{i,+1} + T_{i,-1} = 1. \tag{17}$$

The noisy prediction $\bar{\mathbf{y}}^*$ can be computed as:

$$\bar{\mathbf{y}}^* = T^\top \times \check{\mathbf{y}}^* = \begin{bmatrix} T_{+1,+1} & T_{-1,+1} \\ T_{+1,-1} & T_{-1,-1} \end{bmatrix} \times \begin{bmatrix} \check{\mathbf{y}}^*_{+1} \\ \check{\mathbf{y}}^*_{-1} \end{bmatrix} \tag{18}$$

where $T^\top$ denotes the transpose of the transition matrix $T(\mathbf{x}^*, \theta)$. $\check{\mathbf{y}}^*$ is the class posterior probability given by the well-trained UEN for the distilled example $\mathbf{x}^*$.

2) Re-Weighting by Predictive Uncertainty: The process aforementioned, however, is hardly achieved and inevitably biased, due to two reasons. As estimators such as CNN-based classifiers are sensitive to the overfitting issue, the approximation of $\tilde{\eta}(\mathbf{x}) \approx P_{\widetilde{\mathcal{D}}}(\widetilde{\mathbf{Y}} \mid \mathbf{X} = \mathbf{x})$ may not be as accurate as expected, which means the results of the noisy classifier cannot be trusted blindly. Another issue is that the selection of distilled examples ignores the ambiguous observations located in the range $\left[\frac{1-\rho_{+1\max}}{2}, \frac{1+\rho_{-1\max}}{2}\right]$. These observations are usually closer to the decision boundary and contain information valuable for avoiding false alarms and keeping clear object outlines.

Our strategy to address these problems is to first assume lower $\rho_{\max}$ to make more valuable pixels involved in the training process and then introduce predictive uncertainty to mitigate the negative impacts caused by falsely labeled pixels. Specifically, with the initial mapping results $\hat{\mu}$, the distilled examples can be collected as:

$$\mathbf{x}^* = \left\{\mathbf{x} \in \mathcal{X} \mid \hat{\mu}_{+1} \in \left(\frac{1+\rho_{-1\max}}{2}, 1\right] \bigcup \left[0, \frac{1-\rho_{+1\max}}{2}\right)\right\}. \tag{19}$$

By assigning lower classification scores to pixels potentially affected by the overfitting issue, the UEN is capable of providing distilled examples with higher accuracy. To further take advantage of the ambiguous observations, we propose to set a lower upper bound $\rho_{\pm 1max}$, thus a growing number of ambiguous pixels will be involved in the training of the LTN. This means the distribution of the distilled sub-set is more likely to match the clean distribution. The participation of ambiguous pixels, however, will make the sample selection bias worse and cause larger approximation errors. To cope with this issue, we perform importance re-weighting by adjusting the contributions of pixels with different uncertainty levels. The uncertainty-adjusted loss is defined as:

$$\mathcal{L}_{ua} = \exp(-\psi) \cdot \mathcal{L}_{cce} \tag{20}$$

By adopting the UA re-weighting, pixels with higher uncertainty will contribute less to the training process and those with lower risks will dominate this process.

3) Trace Regularizer: Despite the assumptions aforementioned, estimating IDTM with a parameterized CNN-based model is challenging. To further reduce the solution space of $T(\mathbf{x}; \theta)$, we assume that pixels susceptible to label noise only make up a small proportion of PLs, and propose a trace regularizer to force the LTN focus on the pixels with high label noise. More specifically, by maximizing the trace of the estimated IDTM, LTN will pay less attention to the pixels that have sufficient confidence belonging to one class. The trace regularizer is defined as:

Algorithm 1 Uncertainty Adjusted Label Transition (UALT)

```
Data: Noisy dataset $\widetilde{\mathcal{D}} : \{(\mathbf{x}, \widetilde{\mathbf{y}})^n\}_{n=1}^N$
Result: Ouput classifier $f(\cdot; \phi)$
(1) Train UEN $h(\mathbf{x}; \mathbf{W})$ on noisy dataset $\widetilde{\mathcal{D}}$ with Eq.(11);
(2) Compute the initial result $\mu(\mathbf{x})$ and predictive uncertainty $\psi(\mathbf{x})$;
(3) Collect distilled set $\mathcal{D}^* := \left\{(\mathbf{x}^*, \widetilde{\mathbf{y}}^*, \check{\mathbf{y}}^*)^m\right\}_{m=1}^M$ with the initial result $\mu(\mathbf{x}^*)$ with Eq.(15);
// LTN training with distilled examples;
while $epoch < EP_{max}$ do
    (4) Compute noisy predictions $\bar{\mathbf{y}}^*$ with $T(\mathbf{x}^*; \theta)$ and $\check{\mathbf{y}}^*$ with Eq.(18);
    (5) Calculate the UA reweighted loss $\mathcal{L}_{ua}$ with uncertainty levels $\psi$ and corrected CE loss $\mathcal{L}_{cce}$ by Eq.(20);
    (6) Add the trace regularizer $\mathcal{L}_{tr}$ by Eq.(21);
    (7) Train LTN $g(\mathbf{x}^*; \theta)$ by minimizing the objective function $\mathcal{L}_g$ in Eq.(22);
end
// TMN training with predicted IDTM;
(8) Compute IDTM $T(\mathbf{x}; \theta)$ for each sample in the noisy dataset $\widetilde{\mathcal{D}}$ with the well-trained LTN;
(9) Train the TMN $f(\mathbf{x}; \phi)$ by minimzing Eq.(3);
(10) Output $f(\cdot; \phi)$;
```

$$\mathcal{L}_{tr} = \sum_i T_{i,i}(\mathbf{x}^*; \theta). \tag{21}$$

where $T(\mathbf{x}^*; \theta)$ is the output of LTN, with the input instance $\mathbf{x}^*$ and the parameter set θ. $T_{i,i}(\mathbf{x}^*; \theta)$ denotes the i th diagonal element in $T(\mathbf{x}^*; \theta)$.

Combing the UA re-weighting and the trace regularizer, the loss for training LTN is shown as follows:

$$\begin{aligned} \mathcal{L}_g = & \sum_{\mathbf{x}^* \in \mathcal{D}^*} \exp(-\psi(\mathbf{x}^*)) \cdot \mathcal{L}_{cce}(\widetilde{\mathbf{y}}^*, T(\mathbf{x}^*; \theta), \check{\mathbf{y}}^*) \\ & - \gamma \cdot \sum_i T_{i,i}(\mathbf{x}^*; \theta), \end{aligned} \tag{22}$$

with restriction

$$\sum_j T_{i,j}(\mathbf{x}^*, \theta) - 1 = 0, \tag{23}$$

where γ is a trade-off parameter controlling the impacts of the regularizer term.

The process of IDTM estimation is summarised in Algorithm 1.

TABLE I

QUANTITATIVE EVALUATION RESULTS ON THE GSM-ACT AND GSM-BRIS DATA SETS

Databases	Metrics	GradCAM	LayerCAM	WS-SOD	PSL	HWSL	MFNet	SCW	DeepSolar	PS-CNNLC	RAN	SP-RAN	UALT
GMS-ACT	$AC \uparrow$	0.9840	0.9696	0.9811	0.9785	0.9761	0.9797	0.9819	0.9712	0.9795	0.9790	0.9781	**0.9892**
	$F_\beta \uparrow$	0.5872	0.5365	0.6904	0.5791	0.5747	0.6051	0.6374	0.5894	0.7009	0.6667	0.6260	**0.8016**
	$IoU \uparrow$	0.4156	0.3666	0.5272	0.4076	0.4032	0.4338	0.4678	0.4179	0.5396	0.5000	0.4556	**0.6688**
	$MAE \downarrow$	0.0203	0.0475	0.0193	0.0225	0.2143	0.0205	0.0180	0.0565	0.0294	0.0154	0.0269	**0.0108**
	$S_\alpha \uparrow$	0.8319	0.7985	0.8509	0.7730	0.6592	0.8325	0.8254	0.8024	0.8481	0.8933	0.8521	**0.9077**
GMS-Bris	$AC \uparrow$	0.9865	0.9847	0.9775	0.9857	0.9808	0.9795	0.9775	0.9689	0.9769	0.9866	0.9818	**0.9888**
	$F_\beta \uparrow$	0.7153	0.7434	0.6644	0.6301	0.6342	0.5813	0.5685	0.5354	0.5596	0.7497	0.7112	**0.7795**
	$IoU \uparrow$	0.5567	0.5916	0.4974	0.4600	0.4644	0.4097	0.3972	0.3655	0.3885	0.5997	0.5518	**0.6387**
	$MAE \downarrow$	0.0212	0.0352	0.0232	0.0208	0.2123	0.0211	0.0229	0.0811	0.0292	0.0160	0.0270	**0.0122**
	$S_\alpha \uparrow$	0.7568	0.7251	0.7718	0.5814	0.5198	0.7069	0.6918	0.6047	0.6733	0.8157	0.7331	**0.8468**

IV. EXPERIMENT AND ANALYSIS

A. Data Sets and Experimental Settings

1) Data Sets: We collected two aerial data sets from Google Static Map API, GMS-ACT, and GMS-BRIS, for performance evaluation of the proposed method. Training data in both data sets are annotated manually with image-level labels. The GMS-ACT data set was captured over the Australian Capital Territory, Australia. The training set consists of 3927 positive samples and 2524 negative samples while the test set contains consists of 98 positive samples and 87 negative samples. All the images have a size of 256 × 256 pixels with a spatial resolution of 0.15 m. GMS-BRIS data set was collected over Brisbane, Australia with the spatial resolution varying from 0.15 m to 0.3 m. The training set is composed of 2086 positive samples and 3024 negative samples. The test set contains 90 positive samples and 117 negative samples. Each image in both data sets has RGB bands available in the size of 256 × 256 pixels. All the samples in the test sets were annotated with polygons by experts in the field of remote sensing manually. The polygons for the localization of solar panels were then transformed into pixel-level binary masks for testing.

2) Methods for Comparison: In this paper, for comprehensive comparison, we adopted 11 state-of-the-art WSL-based methods including GradCAM [17], LayerCAM [18], WS-SOD [15], PSL [43], HWSL [53], MFNet [37], SCW [38], DeepSolar [20], PS-CNNLC [54], RAN [23], SP-RAN [21].

3) Evaluation Criteria: To measure the mapping performance of our method quantitatively, we adopted five evaluation metrics including Accuracy (AC), F-measure (F_β), Intersection over Union (IoU), Mean Absolute Error (MAE), and S-measure (S_α) [55] to generate quantitative results. PR curves and E-measure curves [56] are also used for evaluation.

AC evaluates the proportion of correctly classified pixels over the image. When it comes to the small-scale foreground object, the differences on AC may not be significant.

F_β estimates the relation between precision and recall, as shown in Eq. (24). With a varying threshold in obtaining Precision and Recall, we can obtain the PR curves presenting the trade-off between precision and recall. A greater area under the curve indicates superior performance.

$$F_\beta = \frac{(1+\beta^2)\cdot TP}{(1+\beta^2)\cdot TP + \beta^2 \cdot FN + FP}, \tag{24}$$

where *TP*, *FP*, and *FN* are true positives, false positives and false negatives, respectively. β is a weight reflecting the importance of precision and recall. In our work, we calculated *TP*, *FP*, and *FN* on the entire test set to include negative samples in the performance evaluation. We set $\beta = 1$.

IoU is a measurement used to describe the overlap of two closed regions. If the predictive results perfectly match the GT, the IoU score will be equal to 1. We computed it by measuring the overlap between detected foreground objects and ground-truth, and then dividing it by the union of these two regions. In our work, the calculation was performed on the entire test set.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}. \tag{25}$$

MAE calculates the mean absolute error between the results and ground truth, providing an intuitional estimation of the pixel-wise difference.The definition of MAE is shown below:

$$MAE = \frac{1}{h \times w}\left|Y - G\right|, \tag{26}$$

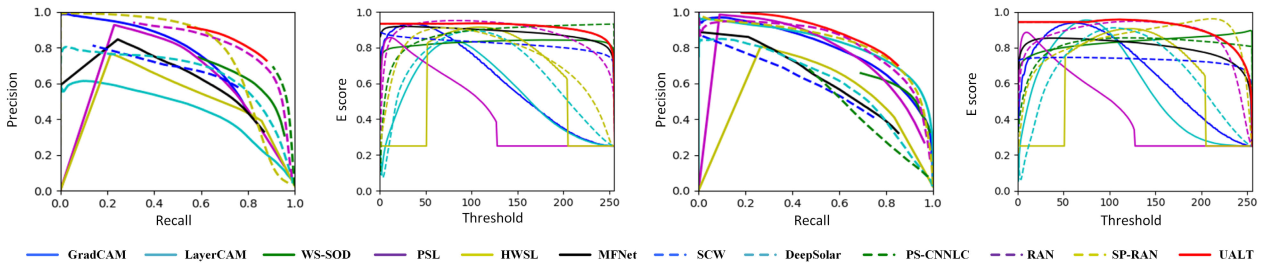
where Y is the mapping result. G is the ground truth. $h \times w$ calculates the total number of pixels in the image.

S_α is a recently proposed metric aiming to provide an evaluation from the aspect of structural similarity. Considering the region (S_r) and object (S_o) perspectives, S_α measures not only the structural similarity but also the foreground-background contrast. S_α is defined as:

$$S_\alpha = \alpha \cdot S_o + (1-\alpha)\cdot S_r, \tag{27}$$

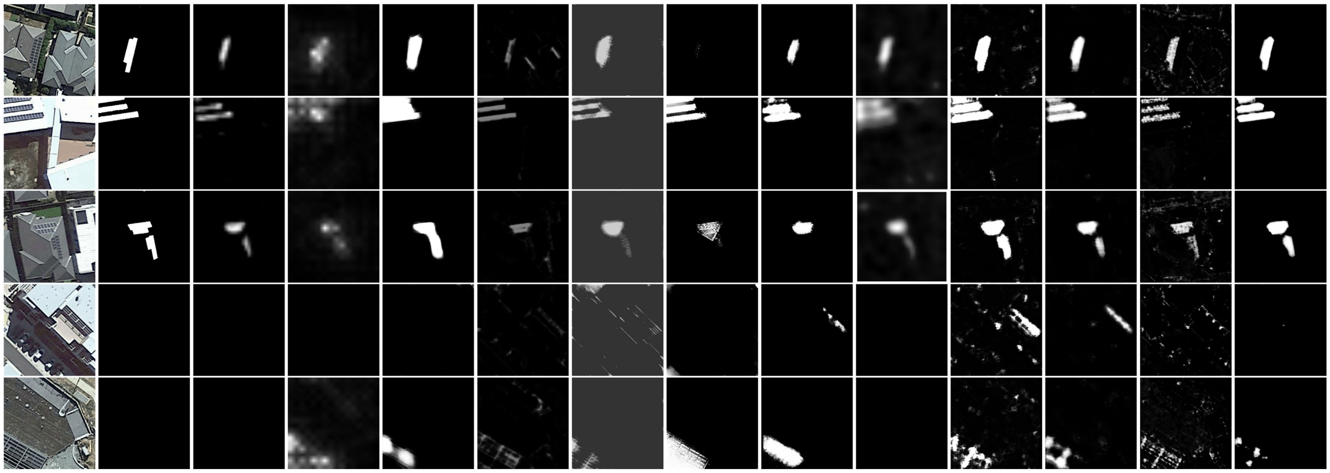
where α shows different contributions of region similarity and object similarity, which is set to 0.5 as default. S_α closer to 1 indicates better performance.

E-measure considers both the pixel-level errors and image-level errors between the binary foreground maps and ground truth. Compared with IoU and F_β measure, E-measure is capable of capturing the global and local pixel matching information simultaneously. E-measure curves are generated



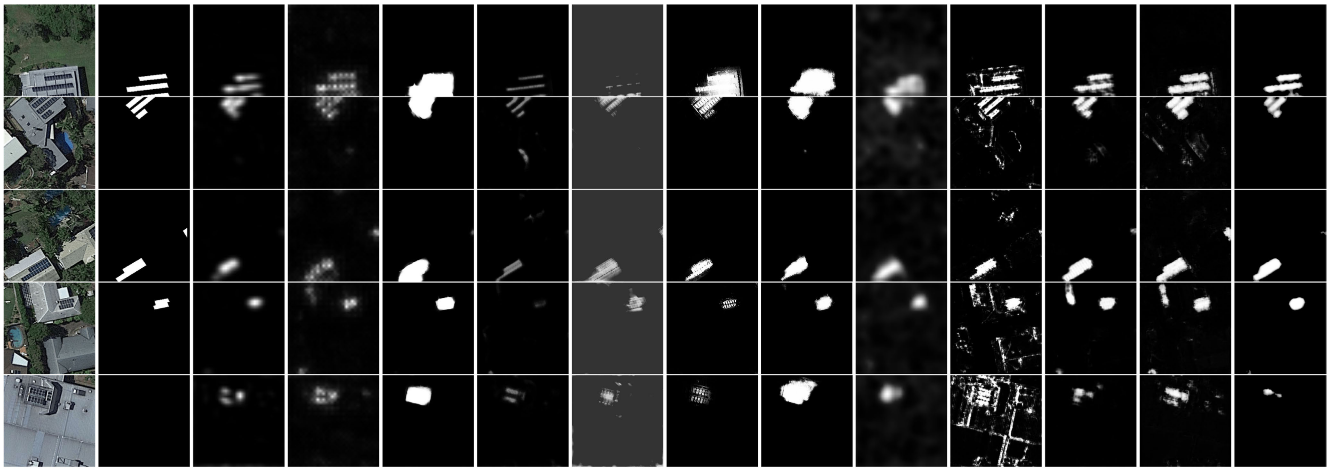
(a) (b)

Fig. 3. PR curves and E-measure curves on two aerial data sets. (a) GMS-ACT. (b) GMS-BRIS. The solid line in red is the proposed method. It is observed that our method produces stable improvement over PR curve and E score.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n)

Fig. 4. Predicted mapping results of competing methods and our method on the GMS-ACT data set. First to third row: positive samples; fourth to fifth row: negative samples. (a) Test images. (b) GT. (c) GradCAM. (d) LayerCAM. (e) WS-SOD. (f) PSL. (g) HWSL. (h) MFNet. (i) SCW. (j) DeepSolar. (k) PS-CNNLC. (l) RAN. (m) SP-RAN. (n) UALT(Ours).



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n)

Fig. 5. Predicted mapping results of competing methods and our method on the GMS-BRIS data set. First to fourth row: positive samples; fifth row: negative sample. (a) Test images. (b) GT. (c) GradCAM. (d) LayerCAM. (e) WS-SOD. (f) PSL. (g) HWSL. (h) MFNet. (i) SCW. (j) DeepSolar. (k) PS-CNNLC. (l) RAN. (m) SP-RAN. (n) UALT(Ours).

by changing the thresholds when producing the binary foreground maps with predicted results. Higher E scores indicate better performance.

4) *Implementation Details:* In this section, we will introduce the implementation details including how to generate PLs with GradCAM and how to train UEN, LTN, and TMN. The configurations of the training procedure of each network will be also clarified.

For the classification network utilized in GradCAM, we used the entire training set with image-level annotations. VGG16 [30] is adopted as the backbone. To avoid the challenges of training VGG16 from scratch, we opted for a fine-tuning approach, initializing the weights of its convolutional layers with a pre-trained model trained on the ImageNet dataset [57]. During the training process, a batch size of 16 was utilized, and the learning rate

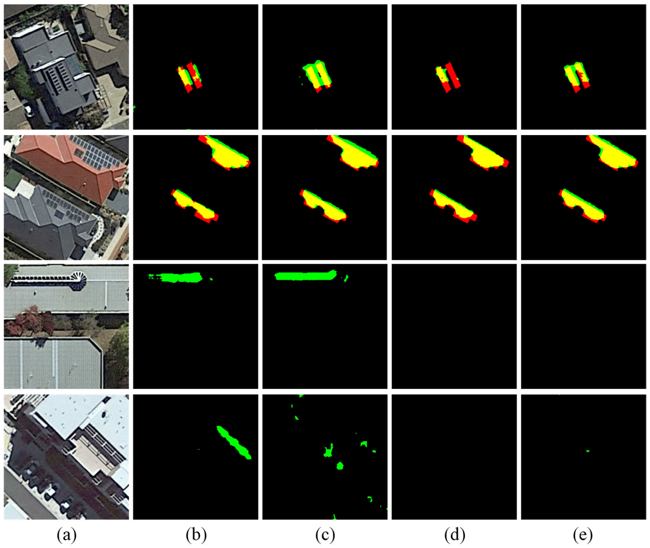


Fig. 6. Visual comparison for the ablation study on the GMS-ACT data set. (a) Aerial images. Results are given by: (b) RAN. (c) UEN. (d) + Trace regularizer. (e) + UA re-weighting. True positives, true negatives, false positives, and false negatives are marked in yellow, black, green, and red, respectively.

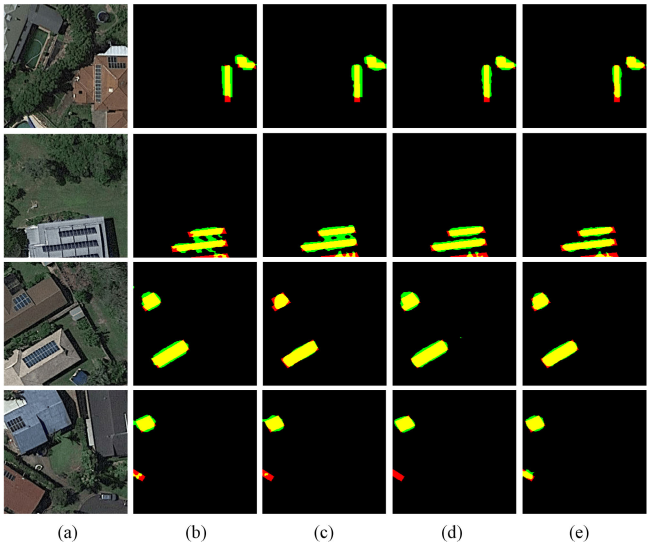


Fig. 7. Visual comparison for the ablation study on the GMS-BRIS data set. (a) Aerial images. Results are given by: (b) RAN. (c) UEN. (d) + Trace regularizer. (e) + UA re-weighting. True positives, true negatives, false positives, and false negatives are marked in yellow, black, green, and red, respectively.

was set to 10^{-4}. The training phase was terminated after 10 epochs.

After the preparation of PLs, the implementation of the proposed method can be divided into three steps as illustrated in Fig. 2. The three convolutional neural networks proposed, i.e., UEN, LTN, and TMN are trained in sequence. The objective of the proposed method is to learn an accurate TMN for target mapping, with the aid of IDTM estimation. In the initial stage, UEN aims to generate initial mapping results accompanied by uncertainty levels, leveraging the noisy dataset. Once the UEN is stabilized and fixed, we proceed to perform noisy data distillation, extracting the distilled set, which subsequently serves as the training set for LTN. Finally, the training process of the TMN involves the entire set of noisy training data and corresponding IDTM generated by fixed LTN. In the inference phase, only the well-trained TMN will be used for forward propagation to provide mapping results.

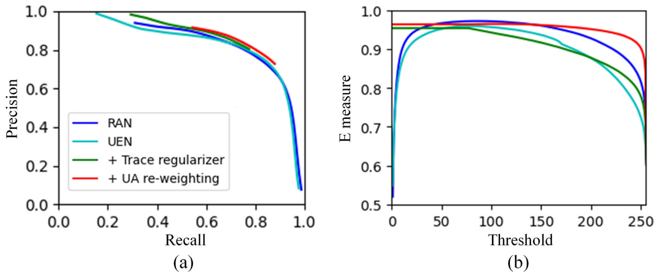


Fig. 8. PR curves and E-measure curves for the ablation study on the GMS-ACT data set. (a) PR curves. (b) E-measure curves.

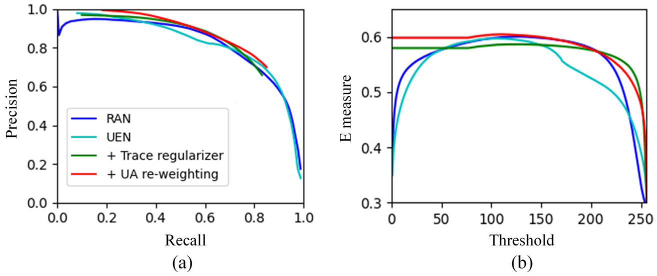


Fig. 9. PR curves and E-measure curves for the ablation study on the GMS-BRIS data set. (a) PR curves. (b) E-measure curves.

All the steps were trained with PyTorch on a PC with a single NVIDIA GeForce RTX 3090 GPU. The backbone for UEN, LTN, and TMN is the residual aggregated network, which is revealed effective in WSL [23]. Adam optimizer was utilized with a training batch size of 8 in the proposed method. For the training of UEN, we set the dropout probability $p = 0.2$ for both data sets. The stochastic forward process was implemented for $K = 3$ times in both the training and testing phase. The initial learning rate is 10^{-3}, and was decayed by 0.9 after each epoch. The training phase ended after 10 epochs for the GMS-ACT, and 40 epochs for the GMS-BRIS data set. Regarding LTN, the whole training takes 30 epochs for GMS-ACT and 50 epochs for the GMS-BRIS data set. The learning rate was initialized as 10^{-3}, and was decreased by 10 percent after every 3 epochs. The training of TMN took 4 epochs, with the initial learning rate 10^{-4} decayed by 0.9 after each epoch.

The code and the dataset of our work will be publicly available at https://github.com/zhangjue1993/Uncertanty-adjusted-label-transtion.

B. Comparison with The State of The Art

1) Quantitative Comparison: We report the quantitative performance of the proposed method and methods for comparison on the two aerial data sets in Table I. Results in Table I reveal the advantages of the proposed method, which consistently achieves the best overall performance on both data sets. Compared with the competing methods, we can observe that our method remarkably improves the scores of F_{β}, IoU and S_{α}. The proposed method outperforms the state-of-art WOSL method LayerCAM by around 0.27 and 0.04 on the F_{β} scores and achieves 0.18 and 0.07 F_1 performance boost when compared with the recently proposed SP-RAN. This demonstrates that the proposed method provides predictions much closer to the GT. We also show the PR curves and E-measure curves of the proposed method and competing methods in Fig.3. We can observe that the quality of predictions given by GradCAM, LayerCAM, PSL, HWSL, and DeepSolar changes dramatically when the threshold varies. By contrast, the proposed method (red solid lines in Fig.3) stably produces high-quality foreground maps and achieves the best performance on both data sets.

2) Qualitative Comparison: In Figs. 4 and 5, we present five examples for visual comparison with 11 competing methods. In these examples, we can find that solar panels are small-scale objects scattered in complex residential areas. Sometimes, they share similar colors and textures as the rooftops, which adds to the mapping difficulty. Another issue is that there may exist some objects that can cause confusion, such as the last two rows in Fig. 4 and Fig.5. Constructed on the binary images classification network, GradCAM and LayerCAM are two state-of-the-art WSOL methods, which only highlight the most discriminative regions with blurry boundaries. WS-SOD successfully locates the target objects but presents a deficient ability to separate them from the complex background. PSL and HWSL, originally proposed for residential area extraction in remote sensing images, fail to provide enough contrast between the foreground objects and the background. On the GMS-ACT data set, MFNet shows limited detection performance. SCW can discover the desired objects but may fail to provide accurate localization. On the GMS-BRIS data set, the predictions of MFNet are not complete. SCW is prone to mix desired objects and rooftops in proximity and suffers from false detections. The mapping results given by Deepsolar are blurry. SP-RAN provides predictions with good shapes, but the object regions are not complete. For the GMS-BRIS dataset, PS-CNNLC can find the object boundaries close to the actual object outline but suffers from high background interference and false detection. The results on the GMS-ACT dataset show that despite the advantage of keeping precise boundaries, PS-CNNLC suffers from background interference. RAN provides a moderate mapping performance, with considerable ability to suppress the background, but is not always able to remove false detection. By contrast, the proposed ULAT exhibits two noteworthy advantages. Firstly, it excels in identifying desired objects from visually analogous counterparts, especially those in low-contrast areas, as evidenced by the results depicted in the last two rows of Fig. 4. Secondly, it consistently demonstrates its capability to accurately differentiate desired foreground objects, i.e., solar panels from the roof situated in close proximity, which can be observed from the second row in Fig. 4 and the first row in Fig. 5.

C. Ablation Study

We further evaluate the contribution of each part of the proposed method in this section. The baseline of our approach is RAN. UEN is the uncertainty estimation network based on the Monte-Carlo dropout with predictive uncertainty. In the training phase of LTN, we studied the influence of the Trace regularizer and UA re-weighting. We denote "+ Trace regularizer" as training the LTN with the following loss:

$$\mathcal{L}_g = \mathcal{L}_{ce} + \gamma \cdot \mathcal{L}_{tr} \tag{28}$$

"+ UA re-weighting" denotes the LTN training by the proposed loss shown in Eq. (22). The quantitative comparison is reported in Tables II and III. To see the discrepancy clearly, we also provide the visual and quantitative results for the ablation study in Figs. 6 to 10. In Figs. 6 and 7, hard samples with confusing areas are presented to show the effectiveness of the proposed method in challenging scenes.

From Tables II and III, we can see that UEN outperforms RAN by a considerable margin, with the increase reaching around 0.034 and 0.016 for F_1 scores on GMS-ACT and GMS-BRIS data sets, respectively. This observation supports the effectiveness of the predictive uncertainty in boosting the mapping accuracy, which is achieved by introducing the stochastic forward pass both in the training and testing phases. Another contribution of the UEN is to help improve the accuracy of distilling examples by separating noise from the predictions. From Figs. 6 and 7, we can see that compared with RAN, UEN can discover more object regions. By incorporating the trace regularizer, the performance sees an advantage of removing the false detection. For the GMS-ACT data set, the increase is around 0.01 and 0.013 on F_1 score and IoU score. For the GMS-BRIS data set, we can also observe a considerable increase. The employment of the UA re-weighting also plays a crucial role in the superiority of our approach. From the quantitative results, we can see a stable improvement in both data sets, with the F_β score reaching 0.8016 and 0.7795, respectively. The comparison of PR curves and E-measure curve Figs. 8 and 9 also show the proposed UEN, trace regularizer, and UA-reweighting are able to bring varying degrees of gains in performance improvement.

From the examples in Figs. 6 and 7, the discrepancy is explicit: the results in the fourth-row show superior performance in localizing target objects correctly as well as providing object boundaries closer to the real object outlines. The baseline, RAN is capable of discovering the target object, with the average performance in suppressing the background interference. The mapping results are contaminated by falsely classified pixels, such as pixels located on the rooftop close to the solar panels. UEN reduces false alarms to a certain extent and provides more accurate segments, but we can still observe a large number of errors. With the trace regularizer and UA re-weighting, the mapping results are further refined. The confusion caused by the complex background is well suppressed and the detected object boundaries are becoming increasingly close to the actual object outlines.

To further reveal the validity of the trace regularizer and UA re-weighting, we show several examples of the IDTM estimated by "+ Trace regularizer" and "+ UA re-weighting" in Fig. 10. We can observe that the predictive uncertainty in column (d) assigns proper uncertainty levels to the predicted results in column (c). With the UA re-weighting, the estimated IDTM T is less deterministic and more accurate. In the third row in Fig. 10, we can observe that although there are falsely detected regions in the initial mapping results, with the high

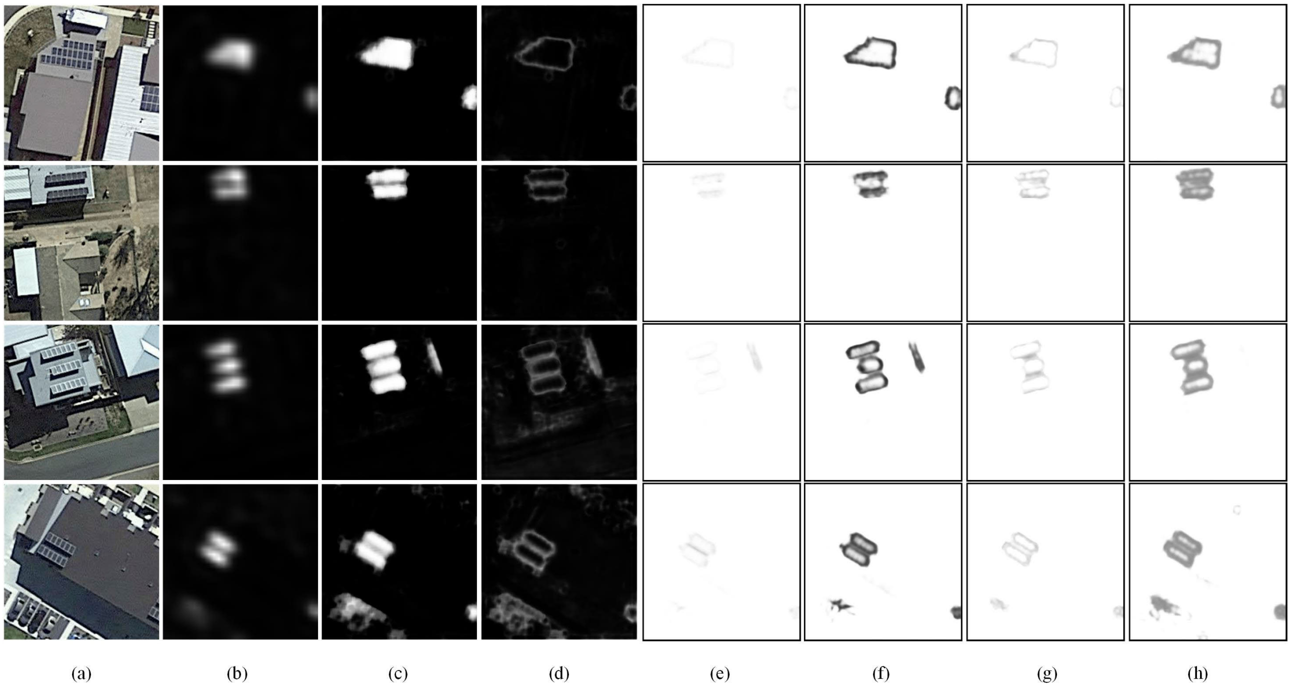


Fig. 10. Prediction results of LTN. (a) Training samples. (b) PLs. (c) Initial results by UEN. (d) Predictive uncertainty. w/o UA re-weighting: (e) T_{00}, (f) T_{11}. w UA re-weighting: (g) T_{00}, (h) T_{11}.

uncertainty levels, the negative impacts of interference on the IDTM are successfully removed.

D. Statistical Significance Analysis

We also conducted a statistical significance analysis, i.e., one-tailed t-test, on the quantitative results of UEN, +Trace Regularizer, and +UA re-weighting, to reveal whether the improvement is significant or not. To make the analysis feasible, we conduct it on the positive samples, so the F_1 score and IoU score can be computed for each sample. Then, we can obtain the mean values and standard deviations for statistical significance analysis. The statistical results are shown in Tables V to VIII. On both data sets, we implemented one-tailed t-test on two collections of positive samples with different numbers. The t-test results for GMS-ACT and GMS-BRIS data sets are summarized in Tables VII and VIII. Cases I II and III denote the method pairs employed in the t-test, UEN v.s. “+Trace regularize”, UEN v.s. “+UA re-weighting”, and “+Trace regularizer” v.s. “+UA re-weighting”. In the one-tailed t-test, the null hypothesis is that methods with the proposed strategies will not bring any improvement in the F_1 score and IoU score on positive samples.

On the GMS-ACT data set, we can observe that “+UA re-weighting” outperforms UEN and “+Trace regularizer” with a notable increase in F_1 score and IoU score. “+ UA re-weighting” also helps reduce the variation of results. Compared to UEN, “+Trace regularizer” sees a decline. All the method pairs except UEN v.s. “+Trace regularizer” exhibits significant improvement with $p < 0.05$ when a large number of positive samples (385) are employed. In case I, there are no significant improvements between the pairs, which is understandable as the main purpose of the Trace regularizer is to make the IDTM estimation stable.

TABLE II
ABLATION STUDY OF THE PROPOSED METHOD ON THE GMS-ACT DATA SET

Method	$AC\uparrow$	$F_\beta\uparrow$	$IoU\uparrow$	$MAE\downarrow$	$S_\alpha\uparrow$
Baseline					
RAN	0.9790	0.6667	0.5000	0.0154	0.8933
UEN	0.9816	0.7002	0.5387	0.0169	0.8860
LTN training					
+ Trace regularizer	**0.9897**	0.7916	0.6550	0.0108	0.8911
+ UA re-weighting	0.9892	**0.8016**	**0.6688**	**0.0108**	**0.9077**

TABLE III
ABLATION STUDY OF THE PROPOSED METHOD ON THE GMS-BRIS DATA SET

Method	$AC\uparrow$	$F_\beta\uparrow$	$IoU\uparrow$	$MAE\downarrow$	$S_\alpha\uparrow$
Baseline					
RAN	0.9866	0.7497	0.5997	0.0160	0.8157
UEN	0.9875	0.7655	0.6201	0.0179	0.8037
LTN training					
+ Trace regularizer	0.9883	0.7686	0.6241	**0.0120**	0.8380
+ UA re-weighting	**0.9888**	**0.7795**	**0.6387**	0.0122	**0.8468**

On the GMS-BRIS data set, we can see that compared with UEN, “+ UA re-weighting” also contributes to a stable increase in F_1 score and IoU score as well as variation suppression. Compared to UEN, “+ Trace regularizer” sees a slight decline on the set with number = 90. With a larger number of positive samples (272), the improvement between UEN and “+ UA re-weighting” is revealed as significant.

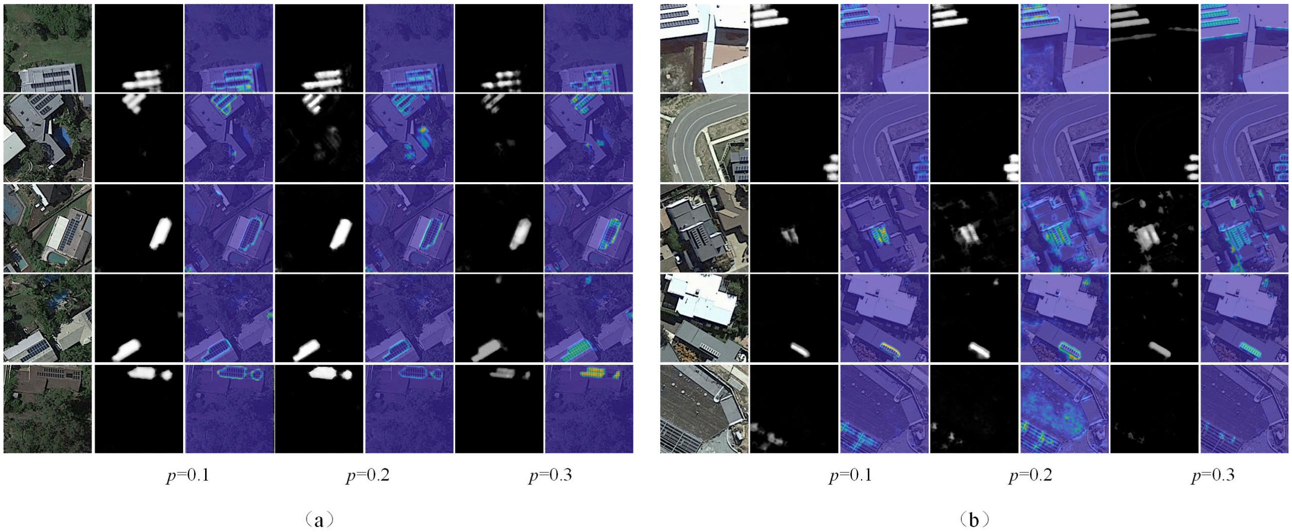


Fig. 11. Illustrating the difference when the dropout probability p changes. (a) GMS-ACT data set. (b) GMS-BRIS data set. The first column in each sub-figure shows the aerial images. For each value of p, the initial mapping results and corresponding predictive uncertainty are exhibited.

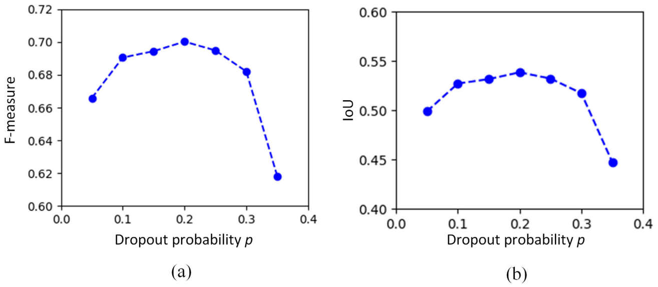


Fig. 12. Impacts of the dropout probability p on the performance of UEN on the GMS-ACT data set. (a) F-measure. (b) IoU scores.

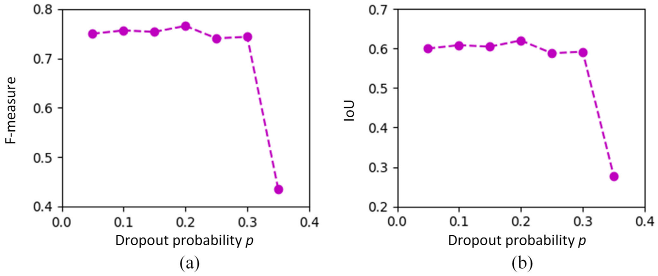


Fig. 13. Impacts of the dropout probability p on the performance of UEN on the GMS-BRIS data set. (a) F-measure. (b) IoU scores.

E. Parameter Analysis

In this section, to explain how the parameters in the proposed method would influence the model performance, we report the quantitative results with different parameter values.

1) Dropout Probability p for the UEN: The purpose of UEN is to produce initial results with accurate uncertainty levels. As the UEN is built based on Monte Carlo dropout, it is important to know the impacts of the dropout probability p on the performance of UEN. Figs. 12 and 13 show the visual comparison with the dropout probability p changing from 0.05 to 0.35. Theoretically, as p grows, it will be increasingly difficult for UEN to converge, and UEN will add more uncertainty levels to its predictions. We can observe that for both data sets, with the growing dropout probability p, the quantitative performance sees a steady increase and reaches the peak when $p = 0.2$. After that, the figures decrease quickly and we can see a huge performance degeneration with $p = 0.35$. The produced predictive uncertainty is also presented to show the discrepancy visually in Figs. 11 with $p = 0.1, 0.2$ and 0.3. The uncertainty level exhibits a notable increase in the object boundaries and co-occurring objects similar in color and texture. When p gets larger, e.g. $p = 0.3$, the UEN is more likely to assign higher uncertainty levels to its predictions, although the predictions are correct, which leads to performance degeneration. This is reasonable as a larger dropout probability means a more dramatic stochastic feed-forward process would happen and it will be increasingly difficult for the model to converge. Hence, in our work, we set p=0.2.

2) Trade-off Parameter γ for the Trace Regularizer: The trade-off parameter γ help imposes different levels of constraints on the trace of the IDTM. In this part, we studied the impacts of γ in two situations: "w UA re-weighting" and "w/o UA re-weighting". Quantitative results are reported in Table IV. We can observe that without UA re-weighting, a smaller weight γ help achieves the best results: setting γ=0.1 for GMS-ACT and GMS-BRIS data sets, the achieved F_β scores reach 0.7916 and 0.7693. As γ grows, there is a slight degeneration of the model performance. On the GMS-ACT data set, the figures first notably decrease to 0.7761 and then see a considerable increase to 0.7820. On the GMS-BRIS data set, the quantitative results are quite close when γ varies. With the UA re-weighting, we can see a different trend. Firstly, compared with "w/o UA re-weighting", the proposed uncertainty-based modification provides a further improvement when γ has different values. The best overall performance is achieved at γ=0.5 and 0.2 for GMS-ACT and GMS-BRIS data sets, respectively, and the F_β scores reach their peak at 0.8016 and 0.7795. This observation also supports the validity of the employment of predictive uncertainty. With the re-weighting guided by uncertainty, our method becomes more stable against the parameter change of γ. Considering

TABLE IV
PARAMETER ANALYSIS FOR γ W/O AND W UA RE-WEIGHTING ON THE GMS-ACT AND GMS-BRIS DATA SETS

Method	GMS-ACT						GMS-BRIS					
	γ	$AC \uparrow$	$F_\beta \uparrow$	$IoU \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$	γ	$AC \uparrow$	$F_\beta \uparrow$	$IoU \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$
w/o UA re-weighting	0.1	0.9897	**0.7916**	**0.6550**	**0.0108**	**0.8911**	0.05	**0.9883**	0.7686	0.6241	**0.0120**	**0.8380**
	0.2	0.9893	0.7785	0.6373	0.0114	0.8837	0.1	0.9875	**0.7693**	**0.6251**	0.0125	0.8292
	0.3	0.9893	0.7761	0.6342	0.0117	0.8809	0.15	0.9875	0.7690	0.6246	0.0124	0.8312
	0.4	**0.9895**	0.7820	0.6420	0.0113	0.8843	0.2	0.9875	0.7573	0.6094	0.0128	0.8231
	0.5	0.9895	0.7806	0.6402	0.0113	0.8837						
w/ UA re-weighting	0.1	0.9893	0.7984	0.6645	0.0108	0.9035	0.05	0.9882	0.7730	0.6299	0.0121	0.8435
	0.2	0.9893	0.7943	0.6587	0.0110	0.8999	0.1	0.9882	0.7709	0.6272	0.0121	0.8389
	0.3	**0.9894**	0.7968	0.6623	0.0108	0.9015	0.15	**0.9893**	0.7791	0.6381	**0.0119**	0.8442
	0.4	0.9893	0.7916	0.6551	0.0111	0.8976	0.2	0.9888	**0.7795**	**0.6387**	0.0122	**0.8468**
	0.5	0.9892	**0.8016**	**0.6688**	**0.0108**	**0.9077**						

TABLE V
MEAN AND STANDARD DEVIATION OF F SCORE AND IOU SCORE ON GMS-ACT POSITIVE SAMPLES

Number	Metric	UEN	+ Trace regularizer	+ UA re-weighting
98	F_β	0.7267±0.2156	0.6900±0.2252	0.7637±0.1768
	IoU	0.6046±0.2020	0.5627±0.2114	0.6420±0.1723
385	F_β	0.6815±0.2113	0.6360±0.2499	0.7229±0.1684
	IoU	0.5479±0.1978	0.5069±0.2249	0.5877±0.1670

TABLE VI
MEAN AND STANDARD DEVIATION OF F SCORE AND IOU SCORE ON GMS-BRIS POSITIVE SAMPLES

Number	Metric	UEN	+ Trace regularizer	+ UA re-weighting
90	F_β	0.7426±0.1436	0.7327±0.1735	0.7639±0.1036
	IoU	0.6075±0.1512	0.6015±0.1724	0.6278±0.1186
272	F_β	0.7102±0.2208	0.7302±0.2091	0.7400±0.1727
	IoU	0.5808±0.1970	0.6075±0.2014	0.6107±0.1733

TABLE VII
STATISTICAL SIGNIFICANCE ANALYSIS (p-VALUE) ON GMS-ACT POSITIVE SAMPLES

Number	98		385	
Metric	F_β	IoU	F_β	IoU
I	0.87	0.92	0.99	0.99
II	0.09	0.08	0.001	0.001
III	0.006	0.002	1.1×10^{-8}	1.1×10^{-8}

[1] In one-tailed t-test, setting the significance level $\alpha = 0.05$, we reject null hypothesis when $p < \alpha$.

TABLE VIII
STATISTICAL SIGNIFICANCE ANALYSIS (p-VALUE) ON GMS-BRIS POSITIVE SAMPLES

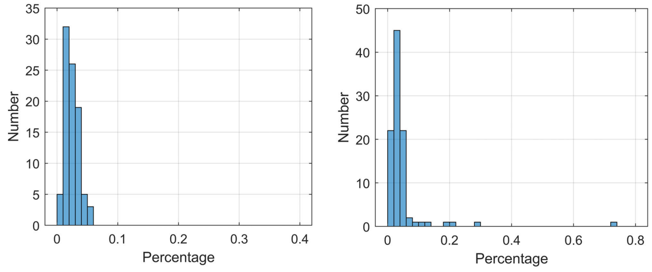
Number	90		272	
Metric	F_β	IoU	F_β	IoU
I	0.6613	0.608	0.12	0.06
II	0.12	0.16	0.03	0.03
III	0.07	0.11	0.27	0.42

[1] In one-tailed t-test, setting the significance level $\alpha = 0.05$, we reject null hypothesis when $p < \alpha$.

the subtle discrepancy of model performance with UA re-weighting, we set $\gamma = 0.5$ and 0.2 for GMS-ACT and GMS-BRIS data sets in our work.

F. Performance on Sub-Sets With Varying-Size Objects

To reveal the high imbalance in the solar panel data sets, we calculated the percentage of positive pixels in testing sets. Fig.14 exhibits the histogram of ratios of positive pixels in every image to statistically show the distributions of solar panels with varying sizes. The target class and the background have an uneven distribution of observation, with most solar panels accounting for less than 5% of the total pixels in a single image. For the GMS-ACT and GMS-BRIS test sets, the ratios of positive/negative pixels in the test set are approximately 1/38 and 1/93. Furthermore, the size of the solar panels also varies greatly. In this case, to validate the effectiveness of our method, we separated the positive samples in the test sets into three subsets, with varying percentages of positive pixels, i.e. 0∼5%, 5%∼25%, and 25%∼40%. The quantitative results are reported in Tables IX and X. On GMS-BRIS test subsets, as there are no positive samples with object



(a) (b)

Fig. 14. Distributions of images with different object sizes in the test set. (a) GMS-ACT data set. (b) GMS-BRIS data set. The x-axis is the ratio of the positive pixels in each image. The y-axis is the number of images.

TABLE IX
Performance of The Proposed Method on The GMS-ACT Test Subset

Percentage	$AC\uparrow$	$F_\beta\uparrow$	$IoU\uparrow$	$MAE\downarrow$	$S_\alpha\uparrow$
0∼5%	0.9880	0.7981	0.6640	0.0122	0.8254
5%∼25%	0.9851	0.8163	0.6896	0.0279	0.8597
25%∼1	0.9817	0.8182	0.6924	0.1751	0.6778

TABLE X
Performance of The Proposed Method on The GMS-BRIS Test Subset

Percentage	$AC\uparrow$	$F_\beta\uparrow$	$IoU\uparrow$	$MAE\downarrow$	$S_\alpha\uparrow$
0∼5%	0.9888	0.7758	0.6338	0.0115	0.8433
5%∼25%	0.9889	0.7800	0.6394	0.0233	0.8717
25%∼1	-	-	-	-	-

percentages above 25%, we only show the results on the subsets 0∼5% and 5%∼25%. We can see that on both data sets, the proposed method achieves consistent increases on F_1 score, IoU score, and S_α as the object size increases from 0∼5% to 5%∼ 25%.There is a decline with MAE when the object size grows, but the F_1 score and IoU score still support the validity of our method.

V. Discussion and Conclusion

This paper proposed a novel UALT method for WS-SPM using only image-level annotations. By taking the PLs as noisy labels, we formulate the WSL problem as a label noise problem. The purpose of the proposed method is to develop a parameterized model to first estimate IDTM and then produce accurate predictions. Approximating IDTM with a CNN is quite challenging due to two reasons: inaccurate selection of distilled examples and the high degree of freedom of IDTM. To solve these problems, we propose to use predictive uncertainty to mitigate the over-fitting issue, which helps improve the accuracy of collecting distilled examples. In the training of LTN, we also propose a UA re-weighting strategy to adaptively modify the contributions of pixels with varying uncertainty levels. To reduce the complexity of estimation IDTM, we propose a trace regularizer to increase the stability of LTN training. Quantitative and visual inspection of the mapping results on two aerial data sets for solar panel detection demonstrated the superiority of the proposed method. We also performed an ablation study to explain the contribution of each part of our method. With a larger number of positive samples, the improvement by employing UA re-weighting is consistently significant.

In the proposed UALT, there are two crucial assumptions: the noise rate upper bound assumption helps guarantee the feasibility of using distilled examples to estimate IDTM; in the proposed trace regularizer, we assume that pixels susceptible to label noise only make up a small proportion of PLs. PLs with high noise levels may not satisfy these two assumptions. In future work, we will investigate the impacts of varying noise levels in PLs on the feasibility and performance of the label noise learning framework.

References

[1] M. K. H. Rabaia et al., "Environmental impacts of solar energy systems: A review," *Sci. Total Environ.*, vol. 754, Jan. 2021, Art. no. 141989.

[2] N. M. Haegel et al., "Terawatt-scale photovoltaics: Trajectories and challenges," *Science*, vol. 356, no. 6334, pp. 141–143, Apr. 2017.

[3] J. M. Malof, R. Hou, L. M. Collins, K. Bradbury, and R. Newell, "Automatic solar photovoltaic panel detection in satellite imagery," in *Proc. Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Nov. 2015, pp. 1428–1431.

[4] J. M. Malof, K. Bradbury, L. M. Collins, and R. G. Newell, "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery," *Appl. Energy*, vol. 183, pp. 229–240, Dec. 2016.

[5] Y. Bazi and F. Melgani, "Convolutional SVM networks for object detection in UAV imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3107–3118, Jun. 2018.

[6] J. Yuan, H. L. Yang, O. A. Omitaomu, and B. L. Bhaduri, "Large-scale solar panel mapping from aerial images using deep convolutional networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2703–2708.

[7] J. M. Malof, L. M. Collins, and K. Bradbury, "A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 874–877.

[8] R. Castello, S. Roquette, M. Esguerra, A. Guerra, and J.-L. Scartezzini, "Deep learning in the built environment: Automatic detection of rooftop solar panels using convolutional neural networks," *J. Phys., Conf. Ser.*, vol. 1343, no. 1, Nov. 2019, Art. no. 012034.

[9] L. Zhuang, Z. Zhang, and L. Wang, "The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106283.

[10] M. V. C. V. D. Costa et al., "Remote sensing for monitoring photovoltaic solar plants in Brazil using deep semantic segmentation," *Energies*, vol. 14, no. 10, p. 2960, May 2021.

[11] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

[12] X. Feng, J. Han, X. Yao, and G. Cheng, "Progressive contextual instance refinement for weakly supervised object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8002–8012, Nov. 2020.

[13] Q. Lai, T. Zhou, S. Khan, H. Sun, J. Shen, and L. Shao, "Weakly supervised visual saliency prediction," *IEEE Trans. Image Process.*, vol. 31, pp. 3111–3124, 2022.

[14] G. Cheng, J. Yang, D. Gao, L. Guo, and J. Han, "High-quality proposals for weakly supervised object detection," *IEEE Trans. Image Process.*, vol. 29, pp. 5794–5804, 2020.

[15] J. Zhang, X. Yu, A. Li, P. Song, B. Liu, and Y. Dai, "Weakly-supervised salient object detection via scribble annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12543–12552.

[16] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[18] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, "LayerCAM: Exploring hierarchical class activation maps for localization," *IEEE Trans. Image Process.*, vol. 30, pp. 5875–5888, 2021.

[19] Y.-F. Li, L.-Z. Guo, and Z.-H. Zhou, "Towards safe weakly supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 334–346, Jan. 2021.

[20] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States," *Joule*, vol. 2, no. 12, pp. 2605–2617, Dec. 2018.

[21] J. Zhang, X. Jia, and J. Hu, "SP-RAN: Self-paced residual aggregated network for solar panel mapping in weakly labeled aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5612715.

[22] J. Zhang, X. Jia, and J. Hu, "Uncertainty-aware forward correction for weakly supervised solar panel mapping from high-resolution aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[23] J. Zhang, X. Jia, and J. Hu, "Weakly supervised solar panel mapping using residual aggregated network for aerial images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 2787–2790.

[24] S. Yang et al., "Estimating instance-dependent Bayes-label transition matrix using a deep neural network," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2022, pp. 25302–25312.

[25] J. Cheng, T. Liu, K. Ramamohanarao, and D. Tao, "Learning with bounded instance and label-dependent label noise," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1789–1799.

[26] D. Cheng et al., "Instance-dependent label-noise learning with manifold-regularized transition matrix estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16609–16618.

[27] Z. Zhu, T. Liu, and Y. Liu, "A second-order approach to learning with instance-dependent label noise," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10108–10118.

[28] M. Wang, Q. Cui, Y. Sun, and Q. Wang, "Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 100–111, Jul. 2018.

[29] K. Bradbury et al., "Distributed solar photovoltaic array location and extent dataset for remote sensing object identification," *Sci. Data*, vol. 3, no. 1, pp. 1–9, Dec. 2016.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[33] J.-H. Lee, C. Kim, and S. Sull, "Weakly supervised segmentation of small buildings with point labels," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7386–7395.

[34] Y. Liu, P. Wang, Y. Cao, Z. Liang, and R. W. H. Lau, "Weakly-supervised salient object detection with saliency bounding boxes," *IEEE Trans. Image Process.*, vol. 30, pp. 4423–4435, 2021.

[35] Y. Xu, X. Yu, J. Zhang, L. Zhu, and D. Wang, "Weakly supervised RGB-D salient object detection with prediction consistency training and active scribble boosting," *IEEE Trans. Image Process.*, vol. 31, pp. 2148–2161, 2022.

[36] A. Obukhov, S. Georgoulis, D. Dai, and L. Van Gool, "Gated CRF loss for weakly supervised semantic image segmentation," 2019, *arXiv:1906.04651*.

[37] Y. Piao, J. Wang, M. Zhang, and H. Lu, "MFNet: Multi-filter directive network for weakly supervised salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4116–4125.

[38] Y. Piao, J. Wang, M. Zhang, Z. Ma, and H. Lu, "To be critical: Self-calibrated weakly supervised learning for salient object detection," 2021, *arXiv:2109.01770*.

[39] Z. Li, X. Zhang, P. Xiao, and Z. Zheng, "On the effectiveness of weakly supervised semantic segmentation for building extraction from high-resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3266–3281, 2021.

[40] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[41] X. Yan, L. Shen, J. Wang, X. Deng, and Z. Li, "MSG-SR-Net: A weakly supervised network integrating multiscale generation and superpixel refinement for building extraction from high-resolution remotely sensed imageries," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1012–1023, 2022.

[42] F. Fang et al., "Improved pseudomasks generation for weakly supervised building extraction from high-resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1629–1642, 2022.

[43] L. Zhang and J. Ma, "Salient object detection based on progressively supervised learning for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9682–9696, Nov. 2021.

[44] Y. Xu and P. Ghamisi, "Consistency-regularized region-growing network for semantic segmentation of urban scenes with point-level annotations," *IEEE Trans. Image Process.*, vol. 31, pp. 5038–5051, 2022.

[45] B. Biggio, B. Nelson, and P. Laskov, "Support vector machines under adversarial label noise," in *Proc. Asian Conf. Mach. Learn.*, 2011, pp. 97–112.

[46] G. Patrini, A. Rozza, A. K. Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2233–2241.

[47] X. Xia et al., "Part-dependent label noise: Towards instance-dependent label noise," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 33, 2020, pp. 1–15.

[48] P. Chen, J. Ye, G. Chen, J. Zhao, and P.-A. Heng, "Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 13, 2021, pp. 11442–11450.

[49] A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?" *Struct. Saf.*, vol. 31, no. 2, pp. 105–112, Mar. 2009.

[50] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," 2015, *arXiv:1506.02158*.

[51] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 30, 2017, pp. 5574–5584.

[52] A. K. Menon, B. Van Rooyen, and N. Natarajan, "Learning from binary labels with instance-dependent noise," *Mach. Learn.*, vol. 107, nos. 8–10, pp. 1561–1595, 2018.

[53] L. Zhang, J. Ma, X. Lv, and D. Chen, "Hierarchical weakly supervised learning for residential area semantic segmentation in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 117–121, Jan. 2020.

[54] J. Zhang, X. Jia, and J. Hu, "Pseudo supervised solar panel mapping based on deep convolutional networks with label correction strategy in aerial images," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Mar. 2020, pp. 1–8.

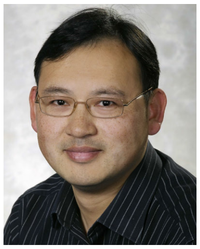
[55] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4548–4557.

[56] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1–12.

[57] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

Jue Zhang (Graduate Student Member, IEEE) received the B.S. degree in electronic science and technology and the M.S. degree in communication and information systems from Beijing Normal University, Beijing, China, in 2014 and 2019, respectively. She is currently pursuing the Ph.D. degree in computer science with the School of Engineering and Technology, University of New South Wales, Canberra, ACT, Australia. Her research interests include remote sensing, weakly supervised learning, and deep learning. She is serving as the Chair for the IEEE UNSW Canberra Student Branch and the IEEE GRSS UNSW Canberra Student Chapter.

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from the University of New South Wales, Australia, in 1996. She is currently an Associate Professor with the School of Engineering and Technology, University of New South Wales, Canberra, Australia. She has authored and coauthored more than 270 publications, addressing various topics, including data correction, feature reduction, and image classification. She is the coauthor of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, Third Edition (1999) and Fourth Edition (2006)]. She is the author of *Field Guide to Hyperspectral/Multispectral Image Processing* (SPIE, 2022). Her research interests include remote sensing, machine learning, and spatial data analysis. She is the Editor-in-Chief of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



Jun Zhou (Senior Member, IEEE) received the B.S. degree in computer science and the B.E. degree in international business from the Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively, the M.S. degree in computer science from Concordia University, Montreal, Canada, in 2002, and the Ph.D. degree from the University of Alberta, Edmonton, Canada, in 2006. He is currently a Professor with the School of Information and Communication Technology, Griffith University, Nathan, Australia. Previously, he was a Research Fellow with the Research School of Computer Science, Australian National University, Canberra, Australia, and a Researcher with the Canberra Research Laboratory, NICTA, Australia. His research interests include pattern recognition, computer vision, and hyperspectral image processing with their applications to remote sensing and environmental informatics. He is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *Pattern Recognition*.

Junpeng Zhang (Member, IEEE) received the B.Sci. and master's degrees in surveying engineering from the China University of Mining and Technology, Xuzhou, China, in 2013 and 2016, respectively, and the Ph.D. degree in electrical engineering from the University of New South Wales, Australia, in 2021. He is currently an Associate Professor with Xidian University. His research interests include object detection and tracking in remote sensing imaginary. He received the "DSTG Best Contribution to Science Award" in the 2018 Digital Image Computing: Techniques and Applications (DICTA 2018). He served as the Chair for the IEEE GRSS UNSW Canberra Student Chapter in 2020.

Jiankun Hu (Senior Member, IEEE) received the Ph.D. degree in control engineering from the Harbin Institute of Technology, China, in 1993, and the master's degree in computer science and software engineering from Monash University, Australia, in 2000. He was a Research Fellow with Delft University, The Netherlands, from 1997 to 1998, and The University of Melbourne, Australia, from 1998 to 1999. He is currently a Full Professor of cyber security with the School of Systems and Computing, University of New South Wales, Canberra, Australia. His research interests include cyber security, including biometrics security, where he has published many papers in high-quality conferences and journals, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He has served on the editorial boards for up to seven international journals and served as the Security Symposium Chair for the IEEE Flagship Conferences of IEEE ICC and IEEE GLOBECOM. He has obtained nine Australian Research Council (ARC) Grants. He served at the prestigious Panel of Mathematics, Information, and Computing Sciences, ARC Excellence in Research for Australia (ERA) Evaluation Committee in 2012.