This paper is published in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6504005, doi: 10.1109/LGRS.2021.3132935.

Uncertainty-aware Forward Correction for Weakly Supervised Solar Panel Mapping from High-resolution Aerial Images

Jue Zhang, *Student Member, IEEE,* Xiuping Jia, *Fellow, IEEE,* Jiankun Hu, *Senior Member, IEEE*

***Abstract*—Solar panel mapping from high resolution aerial images is becoming increasingly crucial to grid planning and operation, where weakly supervised approach has been explored. To cope with the noisy nature of pseudo labels (PLs) generated by weakly supervised object localization, we propose an effective uncertainty-aware forward correction (UA-FC) method to learn clean predictions from the noisy PLs. The proposed method consists of two steps: heteroscedastic uncertainty estimation and forward correction procedure. The purpose of the first step is to produce uncertainty as an indicator for the instance-dependent noise. The second step includes a target mapping network to produce clean predictions and a transition function to model the relationship between clean predictions and noisy PLs. As estimating every probability of one class flipped into another is difficult and time-consuming, we introduce heteroscedastic uncertainty as a measurement and propose an uncertainty-based transformation function to map clean predictions into noisy ones. By minimizing the errors between the noisy predictions and noisy labels, the target mapping network is able to offer clean predictions close to the actual objects. Extensive experiments on an aerial data set reveal that the proposed method outperforms other state-of-the-art methods by a large margin, especially in recovering the boundary of the objects.**

***Index Terms*—Remote sensing, solar panel mapping, weakly supervised learning, uncertainty estimation, forward correction**

I. Introduction

DUE to the rapid and continual growth in global solar deployment, collecting detailed information about the distributed solar photovoltaic (PV) systems is becoming increasingly crucial to grid planning and operation [1]. In recent years, the availability of frequently updated high-resolution aerial images (ground sampling distance less than 50 cm) makes it possible to inspect solar panels with the aid of machine learning techniques.

The goal of automatic solar panel mapping is to provide the label for each pixel in an image, thereby identifying pixels belonging to solar panels and the background. Traditionally, works on this topic construct a classification system [2], by resorting to hand-crafted features, and classical classifiers. For the past few years, convolutional neural networks (CNNs) have become a popular solution to many tasks in the community of computer vision and remote sensing. Concerning solar panel mapping, most existing CNN-based methods are carried out in a fully supervised manner [3–5], which require a large number of training samples with manually annotated pixel-wise ground-truth. The performance of the methods aforementioned mainly depends on a large number of accurate pixel-wise annotations, which requires expert knowledge and is inevitably time-consuming.

To reduce the labeling workload, weakly supervised learning (WSL) has been introduced to extract objects of interest [6]. Instead of relying heavily on accurate pixel-wise annotations, weakly supervised learning makes use of labels in weaker forms such as image-level labels, and scribbles to provide supervision. Learning with weak labels, however, generally suffer from performance degradation. To minimize the performance gap, a broad range of strategies have been proposed. For example, self-training scheme [7, 8] is one of the most popular and effective strategies in WSL, which generates pseudo labels (PLs) to construct predictive models. Popular methods to obtain PLs include class activation mapping (CAM) [9], and gradient-weighted class activation mapping (Grad-CAM) [10]. Despite the improvement achieved, the self-training scheme has difficulties in giving complete regions with precise edges, which can be attributed to the inherent and noisy nature of PLs. With such identification, the idea of learning with noisy labels has recently drawn lots of attention. Label-noise-based models aim to reduce negative effects of the label noise by optionally modeling its distribution [11, 12].

For weakly supervised solar panel mapping (WS-SPM), recent efforts consider challenges brought by misleading details from high-resolution remote sensing images [13, 14], e.g. intricate spatial details and ambiguities caused by shadows and occlusion. To date, most studies are developed with the self-training scheme [15–17], with label correction strategy and regularizer terms utilized to further boost the performance.

In this paper, we are inspired by the state-of-the-art performance of learning with noisy labels and develop an uncertainty-aware forward correction (UA-FC) model to disentangle a clean prediction from noisy PLs. The proposed model consists of two parts: heteroscedastic uncertainty estimation to approximate the uncertainty level and forward correction procedure to produce clean predictions from noisy labels. Our contribution is twofold:

- We assume that the samples where the model may fail to give confident predictions are more likely to be affected by a higher noise rate. With this assumption, we propose to use the heteroscedastic uncertainty to indicate the

This work was supported by the Australian Research Council (ARC) under Grant DP190103660 and DP200103207. *(Corresponding author: Jue Zhang.)*

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

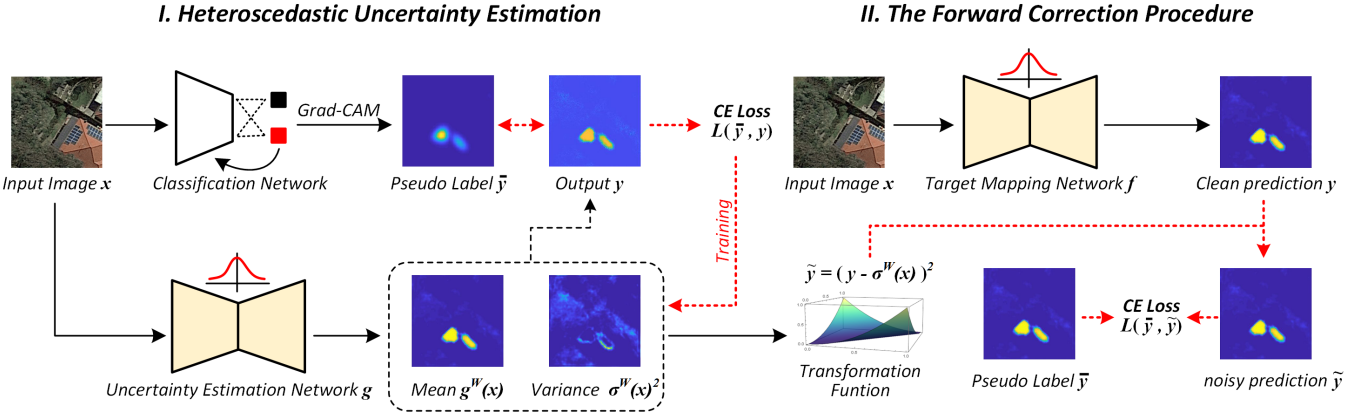

Fig. 1. The flowchart of the proposed uncertainty-aware forward correction method. The red broken line represents the calculation of loss functions and the training phase. Each noisy pseudo label $\bar{y}$ is represented as the clean prediction y transformed by a transition function $T(x)$. The heteroscedastic uncertainty is produced by the uncertainty estimation network (UEN) as a measurement of the instance-dependent noise for each pixel in an image. After the clean prediction y is transformed by a newly proposed mapping function, the target mapping network is trained by minimizing the error between the noisy prediction $\tilde{y}$ and noisy pseudo labels $\bar{y}$.

noise level varying with the observation data rather than estimating every probability of one class flipped into another accurately.

- We propose an effective transformation function to approximate the instance-dependent transition with the estimated uncertainty and then develop a forward correction procedure to train a target mapping network under noisy labels.

II. Methodology

To tackle WS-SPM from a label noise perspective, we propose an uncertainty-aware forward correction model to separate clean prediction from noisy labels. The proposed method first estimates instance-dependent uncertainty and then develops an uncertainty-based transition function to implement a forward correction procedure. The flowchart of the proposed method is shown in Fig.1.

A. Preliminaries

We formulate the weakly supervised solar panel mapping by considering the noise in the PLs. In supervised C-class classification, let $\bar{\mathcal{D}}$ be the noisy joint distribution of two random variables X and $\bar{Y}$. We have $(X, \bar{Y}) \in \mathcal{X} \times \bar{\mathcal{Y}}$, where $\mathcal{X} \subseteq \mathbb{R}^d$, $\bar{\mathcal{Y}} = \{1, 2, ..., C\}$. Suppose that $\mathbf{x}$ is a data point sampled from X, the goal of learning with noisy labels is to find a distribution $\Phi_{\bar{Y}}(\mathbf{x}) := P(\bar{Y} \mid X = \mathbf{x})$ over the the latent clean label Y. Assume there is an instance-dependent transition matrix $T(\mathbf{x}) = (T_{i,j}(\mathbf{x}))_{i,j=1}^{C} \in [0,1]^{C \times C}$, $P(\bar{Y} = j \mid X = \mathbf{x})$ over Y can be expressed as

$$P(\bar{Y} = j \mid X = \mathbf{x}) = \sum_{i=1}^{C} T_{i,j}(\mathbf{x}) \cdot P(Y = i \mid X = \mathbf{x}), \quad (1)$$

$$T_{i,j}(\mathbf{x}) = P(\bar{Y} = j \mid Y = i, X = \mathbf{x}), i, j = \{1, ..., C\}. \quad (2)$$

With the input sample $\mathbf{x}$, the transition matrix $T_{i,j}(\mathbf{x})$ indicates the probability of class i being flipped into class j. For object segmentation and mapping tasks, it is noteworthy that the instance dependence rises the difficulty of estimating $T(\mathbf{x})$: if the input image has the size of $w \times h$, $T(\mathbf{x})$ is characterized by $C^2 \times w \times h$ functions.

To approximate the clean distribution $P(Y \mid X)$, we try to train a predictor f with the parameter set W^*, which is required to satisfy the following equation:

$$T(\mathbf{x}) \cdot f(\mathbf{x}; W^*) \approx P(\bar{Y} \mid X). \quad (3)$$

As the noisy posterior $P(\bar{Y} \mid X)$ can be easily estimated by training a noisy predictor, herein lies two crucial steps: the first is how to estimate transition matrix $T(\mathbf{x})$ and then how to train the predictor f.

B. Heteroscedastic Uncertainty Estimation

In general, estimating an instance-dependent transition matrix is intractable. In our work, we propose to alleviate this issue by measuring heteroscedastic uncertainty in the noisy PLs. Uncertainty of a deep model refers to the case where the prediction results given by the model are potentially inaccurate and cannot be trusted blindly. Although it is crucial to know when the model is unable to make confident decisions, quantifying uncertainty is not an easy task. Bayesian deep learning offers a solution to this issue by assuming the parameters of deep models obeying a prior distribution such as a Gaussian distribution [18], i.e., $\mathbf{W} \sim \mathcal{N}(0, I)$. By doing that, the uncertainty of the model can be quantitatively measured by approaching $P(\mathbf{W}|\mathbf{X}, \mathbf{Y})$. According to the Bayesian modeling, heteroscedastic uncertainty is decided by the observation data and is supposed to change with input potentially having varying degrees of noise.

Motivated by these properties, we assume that the samples where the model may fail to give confident predictions are more likely to be affected by higher noise rates. With this assumption, we propose to use the heteroscedastic uncertainty to indicate the noise level varying with the observation data and develop an uncertainty-aware measurement for the transition function.

Given the input sample $\mathbf{x} \in \mathbb{R}^{w \times h}$, we denote the uncertainty predictor as g. In Bayesian modeling, the logit vector of the model output $\hat{y}$ is supposed to follow a Gaussian distribution:

$$\hat{y}|\mathbf{x} \sim \mathcal{N}(g^{\mathbf{W}}(\mathbf{x}), \sigma^{\mathbf{W}}(\mathbf{x})^2). \tag{4}$$

This means the predictive output $\hat{y}$ is produced with the prediction $g^{\mathbf{W}}(\mathbf{x}) \in \mathbb{R}^{C \times w \times h}$ corrupted by an error following the normalized Gaussian distribution (as Eq.(5) shows). $\sigma^{\mathbf{W}}(\mathbf{x})^2 \in \mathbb{R}^{C \times w \times h}$ is the variance, which represents the uncertainty level regarding the sample $\mathbf{x}$.

$$\hat{y} = softmax(g^{\mathbf{W}}(\mathbf{x}) + \sigma^{\mathbf{W}}(\mathbf{x}) \cdot \epsilon), \quad \epsilon \sim \mathcal{N}(0, I). \tag{5}$$

We can expect that the predictor g can not only minimise the predicative errors (as Eq.(6) shows), but also give the uncertainty level for the input.

$$-log\mathbb{E}_{\mathcal{N}(\hat{y}; g^{\mathbf{W}}(\mathbf{x}), \sigma^{\mathbf{W}}(\mathbf{x})^2)}(\hat{y}). \tag{6}$$

C. The Forward Correction Procedure

To approximate the conditional distribution of clean label $P(Y \mid X)$, we consider the target mapping network f transforming training sample x into a clean prediction y, and assume that corresponding noisy predictive logit vector $\widetilde{y}$ can be written as :

$$\widetilde{y}_j = P(\widetilde{Y} = j | X = \mathbf{x}) = \sum_{i=1}^{C} T_{i,j}(\mathbf{x}) \cdot y_i, \tag{7}$$

$$y_i = f_i(\mathbf{x}), \tag{8}$$

where $\widetilde{y}_j$ is the noisy prediction with respect to the j^{th} class. y_i is the clean prediction produced by f for the i^{th} class.

By minimising the loss between noisy prediction $\widetilde{y}$ and PLs, the predictor f will be capable of giving parameters separating clean prediction y from noisy labels $\bar{y}$. This is also called the forward correction procedure [12]. Taking cross-entropy as an example, the corrected loss function is shown as follows:

$$\begin{aligned} \mathcal{L} &= -\sum_{i=1}^{C} \mathbf{1}\left\{\overline{y}_i = 1\right\} \cdot log(\widetilde{y}_i) \\ &= -\sum_{i=1}^{C} \mathbf{1}\left\{\overline{y}_i = 1\right\} \cdot log(\sum_{s=1}^{C} T_{s,i}(x) \cdot f_s(\mathbf{x})), \end{aligned} \tag{9}$$

where $\mathbf{1}\left\{\overline{y}_i = 1\right\}$ denotes indicator function.

From the definition of heteroscedastic uncertainty, we can know that a higher $\sigma^{\mathbf{W}}$ means the model is much more likely to give a wrong prediction, and vice versa. For a multi-classification problem, there is still a gap between the level of uncertainty and the transition functions, as every probability of one class flipped into another is required to be estimated. For a binary case, however, the situation would be easier to cope with. We only need to find a mapping function to mimic the rules aforementioned, rather than estimate every $T_{i,j}(x)$. That is, with a large $\sigma^{\mathbf{W}}$, the clean prediction of the target mapping network y should be shifted towards the opposite direction, which may lead to a flip into the other class.

Following these rules, a simple but effective transformation function is developed:

$$\widetilde{y} = (y - \sigma^{\mathbf{W}}(\mathbf{x}))^2. \tag{10}$$

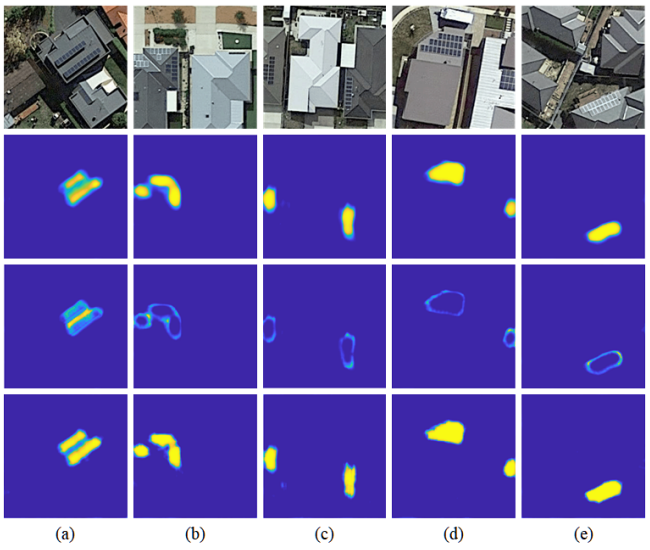

Fig. 2. Examples of results from the uncertainty-aware forward correction. First row: aerial images. Second row: mapping results from the UEN. Third row: uncertainty (σ) from UEN. Fourth row: final mapping results.

III. Experiments

A. Experimental Setup

1) Data Set: From Google Static Map API, the GMS-ACT data set was collected over the city, Canberra, ACT, Australia, with the spatial resolution varying from 0.15m to 0.3m. Each image has RGB bands with the size of 256×256 pixels. The training set contains 437 positive samples and 414 negative samples while the testing set consists of 98 positive samples and 92 negative samples. The image-level annotations are made manually.

2) Implementation Details: The proposed method contains three networks: classification network to generate PLs via Grad-CAM, uncertainty estimation network as the uncertainty predictor, and target mapping network to produce the clean predictions. For the classification network, image-level annotations are utilized in the training phase, which reflect the existence of solar panels in images. The configuration of the network is the same as that in [17]. For uncertainty estimation network and target mapping network, we take RAN [16] as the backbone. Regarding the uncertainty estimation network, PLs generated by the former classification network are regarded as noisy pixel-wise labels. Two branches, which compute the output and uncertainty, respectively, are constructed based on the basic network architecture of RAN. The cross-entropy loss function is used to train RAN with the Adam optimizer. For both networks, weights of all trainable layers are randomly initialized with a truncated normal distribution. The batch size is 8. The initial learning rate is 10^{-3} and decreased by a factor of 0.9 after every epoch. The learning rate remains unchanged after it reaches 10^{-4}.

The proposed method was developed with PyTorch and executed on NVIDIA Tesla V100 GPU with 32GB memory.

3) Evaluation Metrics: To evaluate the proposed method quantitatively, we calculate five metrics [14], i.e., *Accuracy* (**AC**), *Precision* (**P**), *Recall* (**R**), F_1 score and *IoU* score.

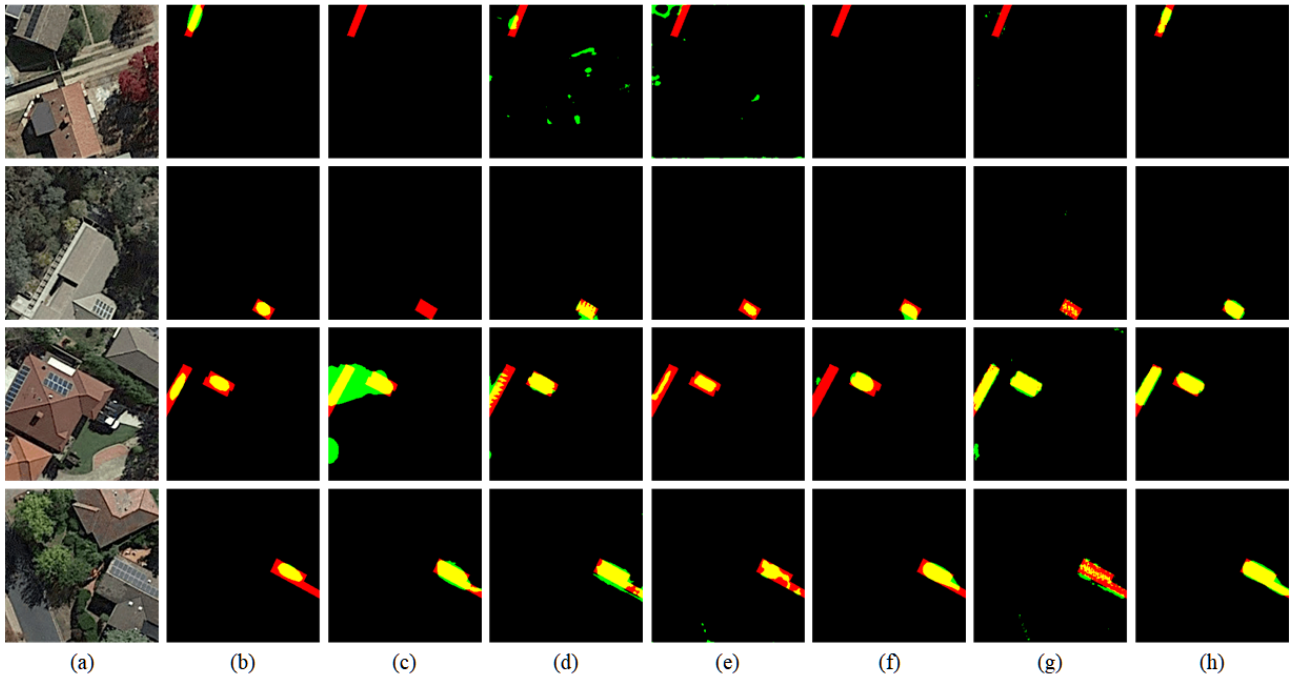

Fig. 3. Examples of mapping results on the GMS-ACT data set. Red pixels, yellow pixels,and green pixels represent the false negative, true positive and false positive. (a) Aerial image. (b) Grad-CAM. (c) WS-SOD. (d) HWSL. (e) PSL. (f) DeepSolar. (g) PS-CNNLC. (h) UA-FC (Ours).

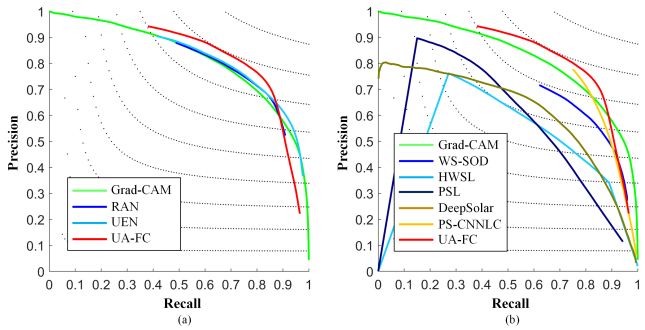

Fig. 4. PRCs of the different mapping methods. (a) Ablation study. (b) Comparison with the state of the art.

TABLE I
ABLATION STUDY ON THE GMS-ACT DATASET

Method	AC	P	R	F_1	IoU
Grad-CAM	0.9875	0. **9036**	0.4238	0.5770	0.4055
RAN	0.9900	0.7629	0.7305	0.7463	0.5953
UEN	0.9896	0.7141	**0.8022**	0.7556	0.6072
UA-FC (Ours)	**0.9912**	0.7929	0.7621	**0.7772**	**0.6356**

Notably, higher $Accuracy$, F_1 score and IoU score suggest superior overall performance. We also use the precision-recall curve (PRC) to compare the quantitative performance.

B. Ablation Study

To reveal the validation of the uncertainty estimation and the forward correction procedure, we performed ablation study and obtained both visual and quantitative results (see Fig.4 (a), and Table I).

1) Effect of the uncertainty estimation: The uncertainty estimation network g is constructed based on the Bayesian learning and expected to give predictions as well as their uncertainty level. Compared with the baseline, RAN, the uncertainty estimation network helps improve the coverage of the mapping results significantly at the cost of increasing false alarms. It also brings slight improvement in the overall performance (0.9% in F_1 score and 1.2% in IoU score).

2) Effect of the forward correction procedure: The benefits of the forward correction are presented both qualitatively and quantitatively. In Fig.2, we show a few examples of mapping results and uncertainty produced by the uncertainty estimation network, and corresponding clean predictions of the target mapping network. As can be seen from Fig.2, the boundary of objects and pixels difficult to distinguish are marked with high uncertainty levels. After the forward correction, the final results are much closer to the actual layout of the object. Numerical results in Table I support the advantages of the proposed method. We can observe that UA-FC outperforms the method aforementioned by a wide margin, especially in improving the completeness of the objects and reducing the background interference. Compared with the baseline RAN, the employment of only uncertainty estimation helps a 7.2% increase in $Recall$ but also leads to a 4.8% drop in $Precision$. By taking advantage of both uncertainty estimation and forward correction simultaneously, UA-FC offers a further improvement in F_1 and IoU score. Fig. 4(a) also shows the superiority of the proposed UA-FC.

TABLE II
QUANTITATIVE COMPARISON ON THE GMS-ACT DATASET

Method	AC	P	R	F_1	IoU
Grad-CAM	0.9875	0. **9036**	0.4238	0.5770	0.4055
WS-SOD	0.9848	0.5864	0.8178	0.6830	0.5186
HWSL	0.9829	0.5688	0.6047	0.5862	0.4146
PSL	0.9796	0.4915	0.4955	0.4934	0.3275
DeepSolar	0.9856	0.6517	0.6023	0.6260	0.4556
PS-CNNLC	0.9879	0.6568	**0.8315**	0.7339	0.5796
UA-FC (Ours)	**0.9912**	0.7929	0.7621	**0.7772**	**0.6356**

C. Comparison with the state of the art

For an extensive evaluation, we compared the proposed method with six state-of-the-art weakly supervised methods, i.e., WS-SOD [8], Grad-CAM [10], HWSL [13], PSL [14], DeepSolar [15] and PS-CNNLC [17].

Fig.3 shows several examples of mapping results produced by different methods. It can be observed that Grad-CAM can only highlight the most discriminative part of the objects. WS-SOD misses several objects with small sizes. HWSL makes relatively accurate predictions, although it is affected by more background interference. The performance of PSL is quite similar to that of HWSL, with less false positives. DeepSolar misses several objects of interest and offers fewer false alarms, mainly because it is developed based on class activation mapping. PS-CNNLC fails to recognize a few objects with large variations while it offers much better boundary maintenance. By contrast, UA-FC can accurately identify dispersively distributed objects with sharp edges close to the actual layout.

Quantitative results in Table II and Fig.4(b) show that UA-FC achieves a better trade-off in $Precision$ and $Recall$, and remarkably performs better in overall performance.

IV. CONCLUSION

In this paper, we approached weakly supervised object extraction from a label noise perspective and proposed an uncertainty-aware forward correction method for WS-SPM. The proposed method assumes that each noisy PL can be transformed from a clean prediction with a transition function, and contains two effective sub-models: uncertainty estimation network and target mapping network. The former network gives the uncertainty level of each pixel in the image, which is subsequently combined with a mapping function to approximate the transition function. With the uncertainty-aware transition function, a forward-correction procedure is introduced to minimize the error between the noisy PLs and the noisy predictions transformed from the clean results given by the target mapping network. Both ablation study and comprehensive experiments on an aerial data set suggest that the proposed method has a superior ability to recognize objects with large variations as well as providing precise boundaries.

REFERENCES

[1] N. M. Haegel, R. Margolis, T. Buonassisi, D. Feldman, A. Froitzheim, R. Garabedian, M. Green, S. Glunz, H.-M. Henning, B. Holder *et al.*, "Terawatt-scale photovoltaics: Trajectories and challenges," *Science*, vol. 356, no. 6334, pp. 141–143, 2017.

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

[3] 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)*. IEEE, 2017, pp. 874–877.

[4] 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, 2018.

[5] 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.*, p. 106283, 2020.

[6] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.

[7] L. Chen, W. Wu, C. Fu, X. Han, and Y. Zhang, "Weakly supervised semantic segmentation with boundary exploration," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 347–362.

[8] J. Zhang, X. Yu, A. Li, P. Song, B. Liu, and Y. Dai, "Weakly-supervised salient object detection via scribble annotations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 12 546–12 555.

[9] 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)*, 2016.

[10] 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 *Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.

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

[12] G. Patrini, A. Rozza, A. Krishna 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)*, 2017, pp. 1944–1952.

[13] 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, 2020.

[14] L. Zhang and J. Ma, "Salient object detection based on progressively supervised learning for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–15, 2021.

[15] 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, 2018.

[16] 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. (IGARSS)*. IEEE, 2021.

[17] ——, "Pseudo supervised solar panel mapping based on deep convolutional networks with label correction strategy in aerial images," in *2020 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2020, pp. 1–8.

[18] K. Posch and J. Pilz, "Correlated parameters to accurately measure uncertainty in deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1037–1051, 2020.