High-Resolution Land Cover Mapping Through Learning With Noise Correction

Runmin Dong, *Graduate Student Member, IEEE*, Weizhen Fang, Haohuan Fu, *Member, IEEE*, Lin Gan, *Member, IEEE*, Jie Wang, and Peng Gong

***Abstract*— High-resolution land cover mapping over large areas is a challenging task due to the lack of high-quality labels. A potential solution is to leverage the existing knowledge contained in the freely available lower-resolution land cover products. However, the relatively low resolution and low accuracy of the products lead to numerous inaccurate labels, which harms the performance of the neural network. This article addresses the challenge by jointly optimizing the network parameters and correcting the noisy labels with a novel online noise correction approach and a synergistic noise correction loss. By incorporating the information entropy as a measurement to determine the probable correct labels, the proposed noise correction approach learns to make effective correction of the noisy labels during training and eventually boosts the performance with a training set containing less noisy labels. Experimental results show that the proposed method can effectively correct the noisy labels and reduce their negative impact on network training. By employing the proposed method, we produce a refined high-resolution (3-m) land cover map from a lower-resolution (10-m) product in China and improve the accuracy from 74.96% (10-m) to 81.32% (3-m). Such an approach that can effectively learn from noisy data sets leads to many potential opportunities for using and magnifying existing knowledge and results.**

***Index Terms*— Deep learning, high-resolution imagery, noisy label, semantic segmentation.**

I. Introduction

HIGH-RESOLUTION land cover maps provide necessary information for detailed national land resource survey, spatial planning, and many sustainability-related applications [1]. Supervised deep learning methods have been a widely adopted approach for this task [2]–[5]. However, training a deep neural network (DNN) requires a vast number of accurately annotated images. As high-resolution satellite image interpretation is labor-intensive, time-consuming, and it especially demands a high level of expertise, existing high-resolution benchmark land cover segmentation data sets are scarce over large areas [6].

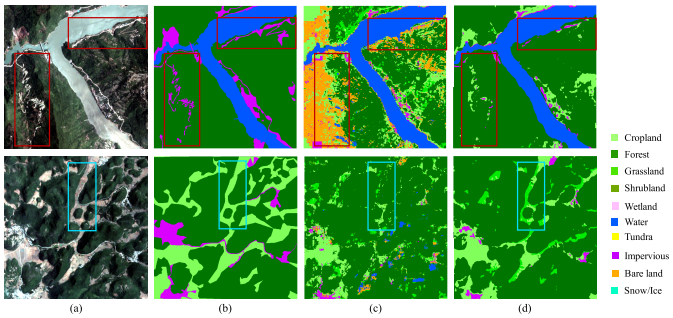


Fig. 1. Examples of a large-scale data set built by existing public land cover products and corresponding land cover mapping results of existing work [7]. (a) Image. (b) Ground truth. (c) Original noise label. (d) Result of baseline [7].

To address the above issue, some studies begin to utilize existing public land cover products, produced by automated and semiautomated processes, to build large-scale data sets [6], [7]. For example, the 2020 Data Fusion Contest [8] use the SEN12MS data set, which integrates the Sentinel 1/2 imagery and moderate resolution imaging spectroradiometer (MODIS)-derived low-resolution land cover maps. This challenge aims to train classification models for high-resolution land cover prediction from low-resolution annotations. However, as summarized by Grekousis *et al.* [9], the overall accuracies of the global and regional land cover products are between 64% and 88%. Besides, these large-scale land cover products with lower resolutions (range from 10 to 1000-m) need to be upsampled to a higher resolution to build pairing segmentation data for high-resolution land cover mapping. As a result, large-scale data sets made by existing public land cover products could contain plenty of inaccurate labels.

As shown in Fig. 1(a) and (c), Dong *et al.* [7] built a large-scale data set by using the 3-m resolution satellite images and 10-m resolution land cover product.

Manuscript received August 14, 2020; revised December 16, 2020 and February 27, 2021; accepted March 19, 2021. Date of publication April 6, 2021; date of current version December 16, 2021. This work was supported in part by the National Key Research and Development Plan of China under Grant 2017YFA0604500, Grant 2017YFB0202204, Grant 2017YFA0604401, and Grant 2020YFB0204700; and in part by the National Natural Science Foundation of China under Grant U1839206. *(Corresponding authors: Haohuan Fu; Peng Gong.)*

Runmin Dong is with the Department of Earth System Science, Tsinghua University, Beijing 100084, China, and also with the SenseTime Group Ltd., Beijing 100084, China (e-mail: drm17@mails.tsinghua.edu.cn).

Weizhen Fang is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: fangweizhen@ict.ac.cn).

Haohuan Fu and Peng Gong are with the Department of Earth System Science, Tsinghua University, Beijing 100084, China (e-mail: haohuan@tsinghua.edu.cn; penggong@tsinghua.edu.cn).

Lin Gan is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: lingan@tsinghua.edu.cn).

Jie Wang is with the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China (e-mail: wangjie@radi.ac.cn).

Digital Object Identifier 10.1109/TGRS.2021.3068280

Although Dong *et al.* [7] have demonstrated that training a DNN directly on the noisy data set can achieve improved results compared with the original low-resolution products [as shown in Fig. 1(c) and (d)], the results still suffer from some confusion between different land cover types. The reason is that DNNs can learn or memorize on any training data set [10]. Thus, the network is under the risk of overfitting to the noisy data. In other words, the performance of DNN can be further improved by handling the noisy label [11].

The traditional noise correction and boundary smoothing methods, such as the fully connected conditional random fields (CRFs) [12], can reduce the local noise and optimize the boundary between different categories. However, the type of noise in our application is quite different from the pure local noise. The noise in our application comes from not only the missing details in lower-resolution labels (e.g., narrow roads and rivers) but also the confusion between different land cover types over a vast region. As a result, the noise in our case can be regarded as class-dependent and global noise. Therefore, traditional methods cannot correct such global noise in this task.

In the computer vision domain, there are many studies focusing on handling noisy labels on classification tasks [10], [13], where the large-scale data sets are collected from websites. However, existing methods that deal with noisy labels on the classification task cannot be directly applied to or cannot obtain the same effect on the segmentation task [14]. The reason is that we should consider the inner-connection between labels in each image and the demand for fine boundaries on the segmentation task. Meanwhile, the handling of noisy labels is also involved in the methods of weak supervision and coarse labeling problems on the segmentation task. However, confronted with different real-world challenges, researchers often find the optimal solutions by utilizing the characteristics of the specific noise scenarios.

In this work, we use a large-scale data set built by combining the latest 10-m resolution land cover product (with an overall accuracy (OA) of 72.8% at the global scale) [15] with the 3-m resolution satellite images [7]. The data set provides a real-world challenge of training with noisy labels. We propose a workflow with a novel online noise correction approach that can correct the noisy labels during the training stage to obtain a relatively clean training set. By employing the proposed approach, we train the network with the corrected labels to achieve higher performance. The contributions of our work are summarized as follows.

1) In the land cover mapping application, we propose a pixel-level noise correction approach, which alleviates the influence of noisy labels during training and facilitates better segmentation performance. Our approach can be regarded as a data pre-processing mechanism and is compatible with other noise handling methods.
2) We demonstrate that the proposed approach can iteratively correct the noisy labels by introducing an uncertainty estimation module, an adaptive noise correction module, and a synergistic noise correction loss.
3) We validate the effectiveness of our proposed approach in both the real-world scenario and the simulated scenarios. We also produce the 3-m resolution land cover maps for the whole of China, and part of the detailed results is published at https://rs.sensetime.com/land.html.

II. Related Work

A. Use of Large-Scale Public Data With Noisy Labels in Land Cover Mapping

With the rapid development in land cover mapping studies over the past few decades, many large-scale or even global-scale land cover products have become available [16]–[19]. Although these products contain plenty of noisy labels, they are free of charge and contain a lot of existing knowledge. Therefore, some studies begin to utilize these publicly available data to avoid annotating a vast amount of training data. Kaiser *et al.* [20] demonstrated that the large-scale data in spite of low accuracy can replace a substantial part (85% in this case) of the manually annotated high-quality data, with which the network can still achieve reasonable performance. Lee *et al.* [21] applied an improvement of the Bayesian Updating of Land Cover (BULC) algorithm sharpening the land cover product from a 300 to 30-m classification. Schmitt *et al.* [22] built a large-scale data set fusing the high-resolution (10-m) images and low-resolution (250 to 1000-m) land cover products. They apply off-the-shelf deep learning and machine learning models to demonstrate the challenges and opportunities of this data set [6]. However, the above studies directly train on the noisy data set and do not address the negative impact of a large amount of noise in the data set.

Some studies utilize these large-scale "imperfect" data sets while trying to reduce the impact of the noise. For example, Maggiori *et al.* [23] initialized the network with a large amount of possibly inaccurate reference data, and then refine the network on a small amount of accurately labeled data. However, in most cases, the annotation of the high-resolution clean data sets is usually limited to certain regions, which limits these methods' application scope. Malkin *et al.* [24] presented a label super-resolution network using the joint distribution between the low-resolution and high-resolution labels for super-resolving the coarse labels. Based on this noise-robust method, Robinson *et al.* [3] fused a 30-m resolution public product and the 1-m resolution high-quality labels to improve the generalization ability of the model. Maas *et al.* [25] proposed a label noise-tolerant random forest for the classification of remote sensing data. Damodaran *et al.* [26] proposed a classification loss with entropic optimal transport (CLEOT) to learn robust DNNs under label noise in remote sensing. However, although the noise-robust method aims to reduce the impact of noise, it is difficult to completely avoid the influence of the noisy labels when keeping those labels during training. Dong *et al.* [7] combined the state-of-the-art 10-m resolution land cover product and 3-m resolution satellite images, and automatically choose the relatively high-quality data and remove the low-quality data based on the similarities. However, this method only evaluates the quality at the image level and cannot handle the noise at the pixel level.

In this work, we use the 10-m resolution land cover product and up-sample it to 3-m resolution as a starting data set with noisy labels, and then propose a pixel-level noise correction approach to use the data set for 3-m resolution land cover mapping.

B. Learning With Noisy Labels in the Computer Vision Domain

Label noise is a significant problem in the computer vision domain. Frénay and Verleysen [27] surveyed relatively early methods. Rolnick *et al.* [28] demonstrated that DNNs are promising in noise handling. Deep learning methods have achieved state-of-the-art results in recent years. Therefore, we mainly review the existing noise handling methods in deep learning on both classification and segmentation tasks.

1) On Classification Tasks: Noise handling methods can be roughly divided into two categories: noise-cleansing-based methods [29], [30] and noise-robust methods [10], [31]. Some studies attempt to detect noisy labels and then prune potential noisy labels or reduce the impact of noise. For example, Northcutt *et al.* [30] used the predicted probabilities to determine the uncertainty of each label and prune noisy images. Huang *et al.* [32] proposed a noisy label detection approach by the normalized average loss of a sample. Alternatively, some studies gradually correct noisy labels by the predictions of DNN to improve the quality of the raw labels [13], [33]. For example, Sukhbaatar and Fergus [34] used a clean data set to estimate the noisy data and correct them by DNN. Yi and Wu [13] iteratively updated both the network parameters and probability distributions of the labels. Our proposed approach is inspired by this work but is designed for the pixel-level noise correction instead of the image-level one. To deal with the pixel-level noise and storage consumption of probability distributions, we propose an online noise correction approach that is more effective for segmentation tasks.

Most of the noise-robust methods modify the loss function to achieve noise-robust classification. Ghosh *et al.* [35], [36] proved the mean absolute error (MAE) is robust against noisy labels and used it for noise-tolerant loss function. The generalized cross-entropy (GCE) loss is a generalization of MAE and categorical cross-entropy loss for both noisy robustness and reduced the difficulty in training [31]. The symmetric cross-entropy loss (SCE) is created to address both the insufficient training and overfitting problem of cross-entropy on the noisy data set [10]. Noise-robust methods can be easily applied to segmentation tasks. Therefore, we compare our proposed method with the GCE and SCE methods in this article.

2) On Segmentation Tasks: Pixel-level noise optimization efforts mostly focus on weak labeling, coarse labeling, and incomplete labeling [37], [38]. Researchers apply different prior knowledge and construct a particular model to different scenarios. For example, Lu *et al.* [39] cast the weakly supervised semantic segmentation problem into a noise reduction problem and propose a super-pixel label noise reduction model. Ibrahim *et al.* [40] proposed a semisupervised approach to utilize both a fine-labeled data set and a weakly labeled data set. This approach uses the fine-labeled data set and an ancillary model to correct the noisy labels. Vicente *et al.* [41] addressed incomplete labeling in the shadow detection task by jointly learning a shadow region classifier and recovering the labels in the training set. Damodaran *et al.* [42] proposed Wasserstein Adversarial Regularization (WAR) for both classification and semantic segmentation problems, which use distances between word embeddings of the class names to derive a semantic ground cost. However, due to the complexity of scenes in different applications, noise label removal in segmentation is still underexplored. Without a fine-labeled data set or reliable foreground annotation, we propose an adaptive noise correction approach that is different from the previous ones.

III. Approach

Assume that a training set of high-resolution satellite imagery and noisy labels is given. The set of N training images is denoted as $X = \{x_i | x_i \in \mathbb{R}^{H\times W\times C}, i = 1, 2, \ldots, N\}$, where each image x_i has a height of H, a width of W, and a channel depth of C. The associated label set is denoted as $Y = \{y_i | y_i \in [1, \ldots, L]^{H\times W}, i = 1, 2, \ldots, N\}$, where L is the number of classes. Generally, the optimization problem on clean data involves minimizing a standard loss function $\mathcal{L}$, such as the cross-entropy loss, with respect to the network parameters θ, i.e., $\min_{\theta} \mathcal{L}(\theta|X, Y)$. However, in our task with the noisy labels, models trained with such a standard loss function are subject to being misled by the incorrect labels.

In this work, we jointly optimize network parameters and noisy labels, i.e., $\min_{\theta,Y} \mathcal{L}(\theta, Y|X)$. We define the predicted distribution as $\hat{Y}^d = f(X; \theta)$, where f is the model's prediction processed by the softmax function, and the updated label set as $\hat{Y} = \{\hat{y}_i | \hat{y}_i \in [1, \ldots, L]^{H\times W}, i = 1, 2, \ldots, N\}$. At the start of the training, $\hat{Y}$ is initialized by Y, i.e., $\hat{Y} = Y$. In the training epoch t, we obtain $\hat{Y}^{(t)}$ by updating the label $\hat{Y}^{(t-1)}$ from last epoch $t-1$, of which the details are presented in Section III-B. The final updated label set is denoted as $\hat{Y}^{(T)}$, where T is the number of the total training epochs.

The noise correction framework is shown in Fig. 2. The convolutional neural network (CNN) backbone is trained to predict the class probability distribution of each pixel. The uncertainty estimation module calculates the uncertainty map based on the class probability distributions. Then, an adaptive noise correction module is introduced to determine the portion of labels that needs to be updated. We update the network parameters θ and label $\hat{y}_i$ through forward computation and backward propagation in each mini-batch step. To facilitate more accurate label correction, we apply a synergistic noise correction loss to make use of the original correct labels, while benefitting from the updated labels in the iterative process. In Sections III-A–III-C, we elaborate on each module and the overall procedures of the proposed approach.

A. CNN Backbone and Loss Function

We use the semantic segmentation method to estimate the class probability distribution. Various commonly adopted segmentation models can be applied as the CNN backbone,

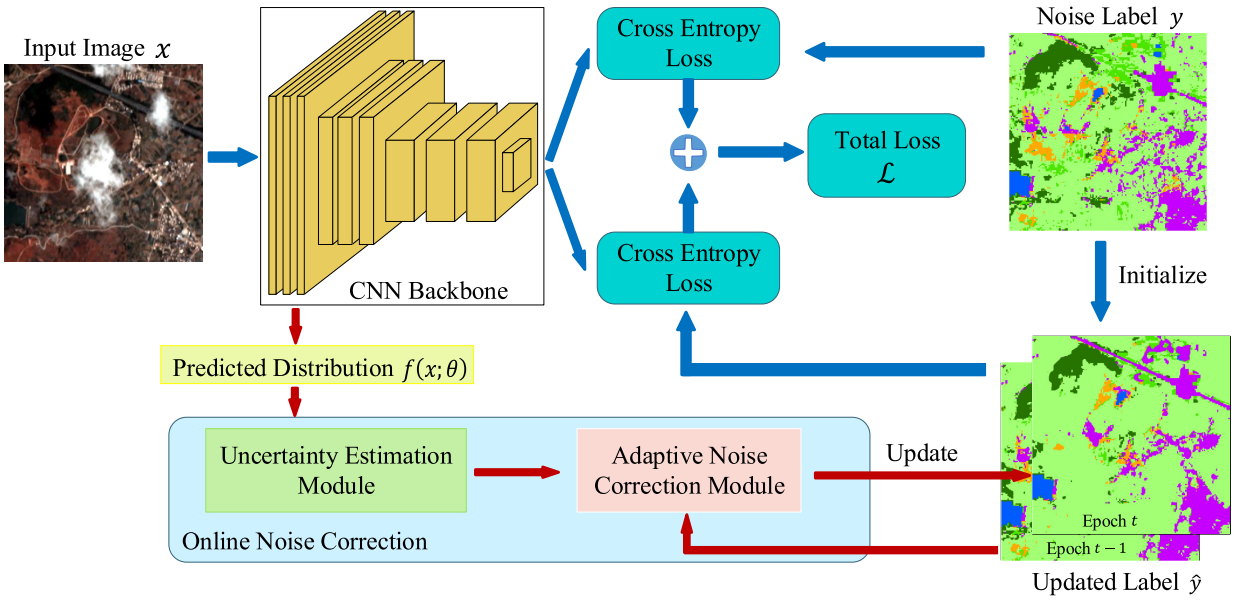


Fig. 2. Noise correction learning framework. It jointly optimizes network parameters and noisy labels through an online noise correction approach and a synergistic noise correction loss.

such as the U-Net [43] and fully convolutional denseNets (FC-DenseNet) [44]. We apply the high-resolution network (HRNet) [45] as the CNN backbone, as it can maintain strong high-resolution representations and achieves favorable performance against state-of-the-art methods in land cover mapping. We also adopt the same generalization strategy as Dong *et al.* [7] to maintain stable performance in different areas or different satellite image sources. The batch normalization (BN) [46] layer after the first convolution layer is replaced with the instance normalization (IN) [47].

The loss function is the cross-entropy loss between the predicted class probability distribution and the updated labels from the last epoch. In this way, the network can iteratively utilize better annotated data. However, to produce more reliable label correction, the original labels are also utilized for training because they contain a decent proportion of correct labels. Therefore, we propose a noise correction loss function $\mathcal{L}(\theta, Y|X)$, which consists of two terms and is formulated as

$$\mathcal{L}(\theta, Y|X) = \mathcal{L}_{\text{ce}}\left(\theta, \hat{Y}\right) + \alpha \mathcal{L}_{\text{ce}}(\theta, Y) \tag{1}$$

where $\mathcal{L}_{\text{ce}}(\theta, \hat{Y})$ and $\mathcal{L}_{\text{ce}}(\theta, Y)$ denote a cross-entropy loss with the updated label and a cross-entropy loss with the original noisy label, respectively. α is a hyperparameter that balances the two loss terms during training.

The term $\mathcal{L}_{\text{ce}}(\theta, \hat{Y})$ is the primary loss which guides the update of the network parameters θ. It is noted that the updated labels used in each epoch are obtained from the last epoch through the noise correction module. $\mathcal{L}_{\text{ce}}(\theta, \hat{Y})$ is defined as

$$\mathcal{L}_{\text{ce}}\left(\theta, \hat{Y}\right) = -\sum_{m=1}^{H\times W}\sum_{j=1}^{L} \hat{y}_{m,j} \log f_m(x_m;\theta). \tag{2}$$

To prevent increasingly coarse boundary, we also keep the original noisy label Y in use, $\mathcal{L}_{\text{ce}}(\theta, Y)$ is defined as

$$\mathcal{L}_{\text{ce}}(\theta, Y) = -\sum_{m=1}^{H\times W}\sum_{j=1}^{L} y_{m,j} \log f_m(x_m;\theta). \tag{3}$$

B. Online Noise Correction Approach

As the image quality and scenario complexity of each satellite image are different, we update the labels of each image independently rather than rely on the statistical information of all images. The independent update strategy makes it possible to correct the labels in every mini-batch training step rather than after every training epoch, which improves the training efficiency by avoiding an extra inference process on the entire training data set. After each mini-batch training step, we only record the single label with the maximal probability rather than a class probability distribution for each pixel, which facilitates the reduction of the memory consumption on a large data set. Note that all labels on the entire data set are updated in every training epoch.

1) Uncertainty Estimation: For each image, we utilize the output of the CNN backbone, i.e., the class probability distribution $\hat{Y}^d$, to estimate the uncertainty of the prediction by using the entropy as the measurement. We choose the largest and the second largest probability values of each pixel to calculate the uncertainty. The reason why we do not use all probability values of the distribution is that the distribution of other classes could introduce redundant information for determining whether the pixel label needs to be updated. Then, we normalize the two probability values. The uncertainty value u_i at a pixel is defined as

$$u_i = -(\hat{y}_{\text{max}} \log \hat{y}_{\text{max}} + \hat{y}_{\text{sec}} \log \hat{y}_{\text{sec}}). \tag{4}$$

We define an uncertainty map U composed of the uncertainty values of all pixels on an image, where $u_i \in (0, \log 2)$.

A smaller value of u_i indicates less uncertainty, and the corresponding prediction label is more likely to be ground truth.

2) Adaptive Noise Correction: We utilize the uncertainty map U obtained from the uncertainty estimation module to determine the pixels that need to be updated. Due to the different scenario complexity of each image, using a fixed threshold for label updates is unreasonable, which may lead to insufficient or excessive updates on different images. Therefore, an adaptive threshold v_i is adopted for each image. We calculate the mean value of each uncertainty map as its updated threshold, which empirically facilitates effective updates on different images. To avoid insufficient updates using relatively low mean values in simple scenarios, thresholds below K are truncated to K. The adaptive noise correction threshold for each image is formulated as

$$v_i = \begin{cases} \text{mean}_{H\times W}(c_i), & \text{if } \text{mean}_{H\times W}(c_i) > K \\ K, & \text{if } \text{mean}_{H\times W}(c_i) \leq K. \end{cases} \tag{5}$$

The predicted labels at different pixels on an image are treated as the correct predicted labels if the corresponding uncertainty values are below the threshold. If the current labels are inconsistent with the correct predicted labels, they will be considered as noisy labels and corrected to the correct predicted labels.

C. Overall Procedure of the Proposed Approach

The overall training procedure consists of three phases. In the first phase, we obtain the initial network parameters for the next phase of noise correction by training the backbone network from scratch with only the cross-entropy loss $\mathcal{L}_{\text{ce}}(\theta, Y)$ between the predictions and original labels. To avoid overfitting to the noisy labels, we use a fixed learning rate and train only a few epochs. In the second phase, we perform the correction of labels and obtain a relatively clean label set. We jointly optimize network parameters and noisy labels, in which a fixed learning rate is adopted to prevent overfitting to updated labels. In the third phase, we use the final updated label set $\hat{Y}^{(T)}$ to train the network from scratch with a gradually reduced learning rate as in standard network training. Note that other noise handling methods can be easily integrated into the third step of our pipeline to boost the performance even further.

D. Implementation Details

The size of the input image is 513×513 pixels, which is an empirically optimal input size considering the receptive field and GPU memory in the high-resolution land cover mapping scenario. All networks are trained using stochastic gradient descent (SGD) with a momentum of 0.9, and a weight decay of 10^{-4}. We use the largest mini-batch that can fit in GPU memory (i.e., a batch size of 2 for a single GPU). We implement the algorithms using synchronized BN over 16 NVIDIA 1080Ti GPUs, with a total batch size of 32. In the first phase of our approach, we train the network for 10 epochs with a fixed learning rate of 0.01. In the second phase, we use the network parameters obtained in the last phase as initialization and set the fixed learning rate to 0.01. The network is trained for 30 epochs until no obvious changes of the labels are observed. The hyperparameter α in the loss function is set to 0.2. We set $K = 0.1$ to truncate the noise correction thresholds. In the final phase, the network is trained from scratch for 50 epochs with an initial learning rate of 0.01 on the corrected labels obtained in the last phase. We schedule the learning rate using the "poly" policy, in which the learning rate is scaled by $(1 - (\text{iter}/\text{total\_iter}))^{0.9}$ [48].

IV. Experiments

A. Data Sets

The evaluation is performed on two data sets, which are the planet image data set (PID) [7] and the GaoFen image data set (GID) [4]. PID is a large-scale noise data set of land cover mapping in China. GID is a well-annotated land cover segmentation data set in China with various scenes and large distribution [4]. To gain more insights into the effect of the noise correction approach in the land cover task, we simulate the noisy scene and conduct the noise-controlled experiments on the training data set of GID.

1) Planet Image Data Set: We fuse the 3-m resolution Planet satellite images and the latest public 10-m resolution land cover product, Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC10), to build the training data set. The Planet images were acquired from Planet satellites with four bands [R, G, B, near-infrared response (NIR)] in June, 2017. Planet satellite images are downloaded through Planet application programming interface (API) (https://www.planet.com/products/platform), which has a screening function for clouds. The Radiometric correction has been applied to the data. The Planet images that we download are less than 15% cloud cover. FROM-GLC10 was produced by a random forest classifier for global land cover mapping in 2017. FROM-GLC10 includes ten land cover types (i.e., Cropland, Forest, Grassland, Shrubland, Wetland, Water, Tundra, Impervious, Bare land, and Snow/Ice) and achieves an OA of 72.8%. The training data set contains 210 000 images of 513×513 pixels, covering about 5% of China (i.e., 500 000 km^2).

Planet image test data set (PITD) is a human-annotated land cover segmentation data set. PITD is collected from China and contains 165 manually labeled 1024×1024 images. PITD includes six land cover types (i.e., Cropland, Forest, Grassland, Water, Impervious, and Bare land). Note that the PITD is a pixel-level annotated data set, instead of a point-level annotated test data set in [7]. Fig. 3 shows the distribution and examples of PITD, and Table I lists the proportion of each land cover type of PITD.

2) GaoFen Image Data Set: GID is a public land cover data set collected from China. The images are acquired from Gaofen-2 satellites. Gaofen-2 images are 4-m resolution with four bands (R, G, B, and NIR). GID contains 120 training images and 30 validation images, each with a size of 6800×7200 pixels. GID includes five land cover types (i.e., Farmland, Forest, Meadow, Water, and Built-up). Note that our land cover

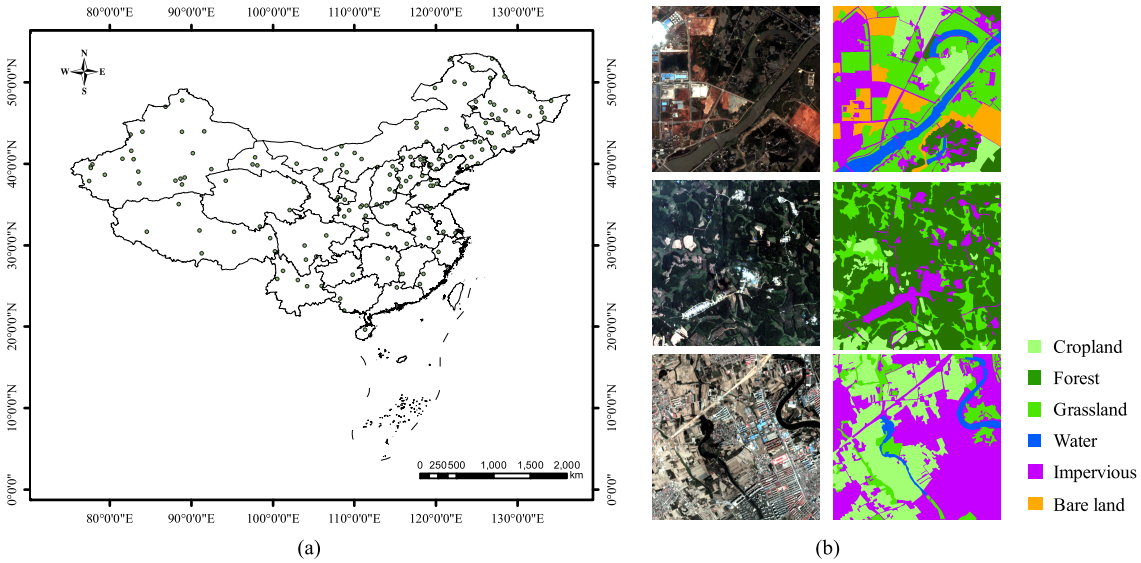


Fig. 3. Distribution and examples of PITD. (a) Distribution of PITD. (b) Examples of PITD.

TABLE I
PERCENTAGE OF EACH LAND COVER TYPE ON PITD

Land cover type	Cropland	Forest	Grassland	Water	Impervious	Bare land
Proportion (%)	33.84	25.86	9.17	2.12	10.96	18.05

mapping results consist of ten types, which cover all the types of GID.

B. Results on PID

PID is a large-scale land cover segmentation data set. Although PID contains a certain degree of noise, DNN can still learn the existing knowledge contained in PID and obtain reasonable results. Therefore, we first analyze the baseline model which is trained without the noise correction. Then, we compare our proposed method with the baseline method and two noise-robust methods.

1) Baseline Method: We use the cross-entropy loss $\mathcal{L}_{\text{ce}}(\theta, Y)$ to train the baseline model without noise correction. The baseline model is trained for 50 epochs with an initial learning rate of 0.01 and the "poly" policy for scheduling the learning rate scaling. Note that all the experimental settings are the same as those in the final phase of our proposed method except for the labels. We use 40 human-annotated images as the validation set to collect the model. For evaluation, we adopt the Overall Accuracy (OA) and the mean-Intersection-over-Union (mIoU) as the performance metrics. The results of the baseline model are presented in Table II. Compared with the 10-m resolution land cover product, which is treated as noisy labels in this work, the baseline method can improve the OA from 74.96% to 79.74% on the PITD. Thanks to the rich texture information in higher resolution satellite images, even though trained on the noisy data, the DNN can still achieve reasonable results. However, there are still some shortcomings with the results of the baseline model. Specifically, the vegetation classes (e.g., Cropland and Grassland) are easily confused, as shown in the blue rectangles in Fig. 1. The narrow roads (belonging to impervious type) are difficult to be identified, as shown in the red rectangles in Fig. 1. These problems are caused by the noise in the original labels, e.g., the disappearance of details and the serious confusion between vegetation types. To tackle these problems, we can gradually correct the original noise labels using the proposed method and hence, reduce the negative impact of the noisy labels.

TABLE II
EVALUATION ON THE PITD AND GID DATA SETS. THE MODELS ARE TRAINED ON THE PID. THE AVERAGE SCORE OF FIVE TRIALS IS REPORTED

Method	PITD		GID	
	OA (%)	mIoU (%)	OA (%)	mIoU (%)
Original labels	74.96	55.52	-	-
Baseline [7]	79.74	61.50	86.15	64.48
GCE [31]	80.00	60.65	85.92	64.90
SCE [10]	80.32	62.07	86.46	65.86
Our method	**81.32**	**63.02**	**89.45**	**69.72**

2) Comparison Results: We train the model using the proposed method on the PID and evaluate the performance on both the PITD and the GaoFen image validation data set. As the noise-robust methods can also be applied to the segmentation problem, we evaluate the GCE loss [31] and SCE loss [10] with the backbone and training strategy that is the same as the baseline on the segmentation task. For the hyperparameters in these two noise-robust methods, we use noise-robust coefficient $q = 0.7$ for the GCE loss, and loss weight $\alpha = 1$ of cross-entropy, loss weight $\beta = 0.025$ of

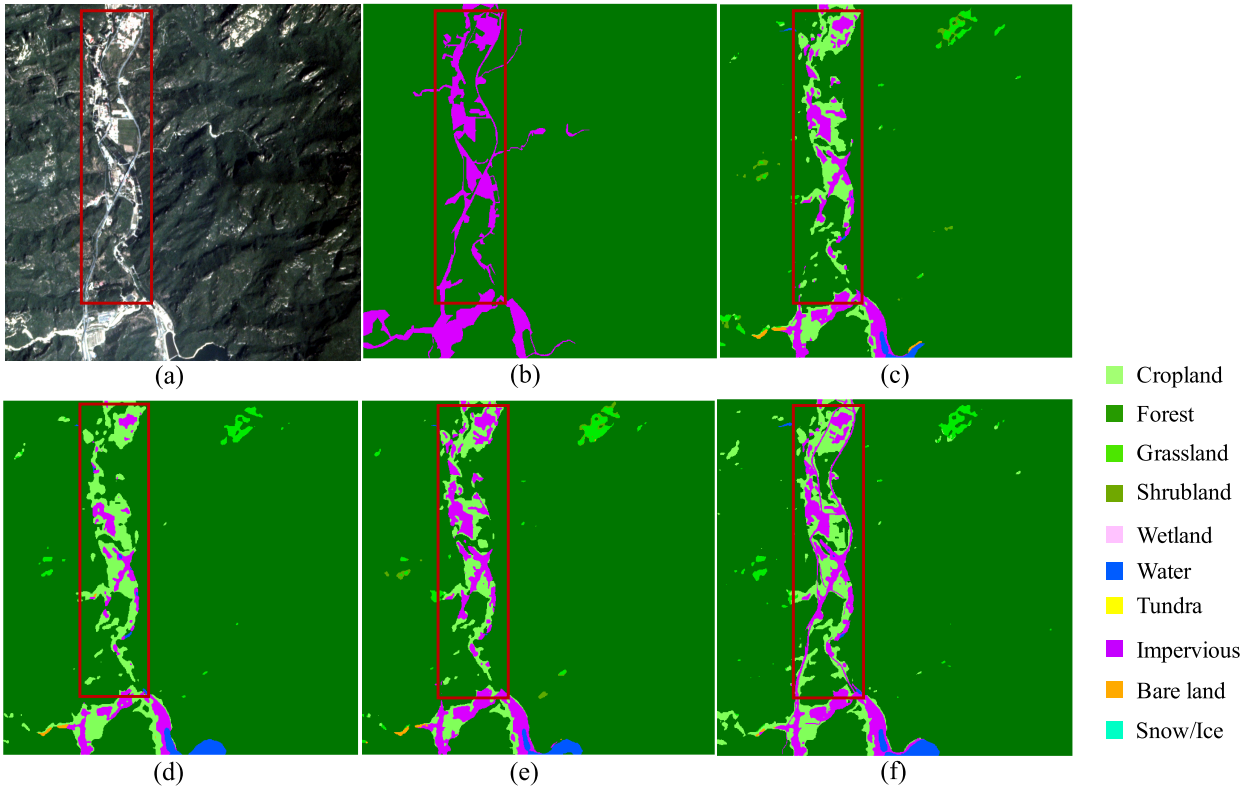


Fig. 4. Visualization of the land cover mapping results on the PITD. The models are trained on the PID. (a) Image. (b) Ground truth. (c) Baseline. (d) GCE. (e) SCE. (f) Ours.

reverse cross-entropy, and hyperparameter $A = -4$ for the SCE loss. The implementation is based on the released codes of these two works, and the parameter tuning is performed according to the original parameter settings. The experimental results are shown in Table II. The GCE loss achieves similar accuracy compared to the baseline model, while the SCE loss can obtain slightly better accuracy in both the PITD and GID data sets. Although the GCE loss and SCE loss can improve the robustness against the noisy labels, the impact of the label noise remains serious especially when the noise rate is high, leading to only a slight improvement of the accuracy. However, the proposed method achieves significantly better results than both the GCE and SCE methods on this task. As shown in Figs. 4 and 5, the proposed method can produce a more accurate segmentation of the linear objectives (e.g., road) and reduce the confusion between vegetation classes (e.g., Cropland and Grassland).

3) Results for Each Land Cover Type on PITD: To further understand the effectiveness of the proposed method, we analyze the results for each land cover type on PITD. Table III lists the Precision, Recall, and F1-score for each type. It can be seen that the effectiveness of our method for each type is related to its original accuracy. The original F1-scores of cropland, water, and impervious types are between 50%–80%, and our method can increase the F1-score by 1.5%–2% compared with the baseline. The original F1-scores of forest and bare land types are higher than 80%, and the improvements of F1-score are about 0.6%. However, the original F1-score of grassland is 38.31%, and the performance of our method is reduced by 1% compared with the baseline. The reason is that the initial model cannot provide a reasonable estimation of grassland from the original noise label.

C. Results on GID

To evaluate the performance of the proposed approach under different noise levels and types, we corrupt the training set labels of the GID by applying various percentages of pixel-level and object-level noise. Specifically, for the pixel-level noise, we first use the original training data of GID to train the segmentation model, i.e., without adding label noise. Then, each image of the training data is evaluated by the resulting model to produce the class prediction for each pixel. The easily-confused class for each pixel is recorded as asymmetric label noise. We define the easily-confused class for each pixel as the predicted class for the incorrect prediction or the class with the second-largest probability for the correct prediction. Then, we randomly select 40%, 50%, and 60% pixels of each image, respectively, to change the original label to the easily-confused class. For the object-level noises, we first use a simple linear iterative cluster (SLIC) algorithm [49] to segment the image. SLIC adopts a k-means clustering approach to generate superpixels for each image, and the resulting clusters can be regarded as objects. We randomly select 20%, 30%, and 40% objects of each image, respectively, to flip labels.

We use the same CNN backbone and training strategy as introduced in Section III-C. The experimental results are

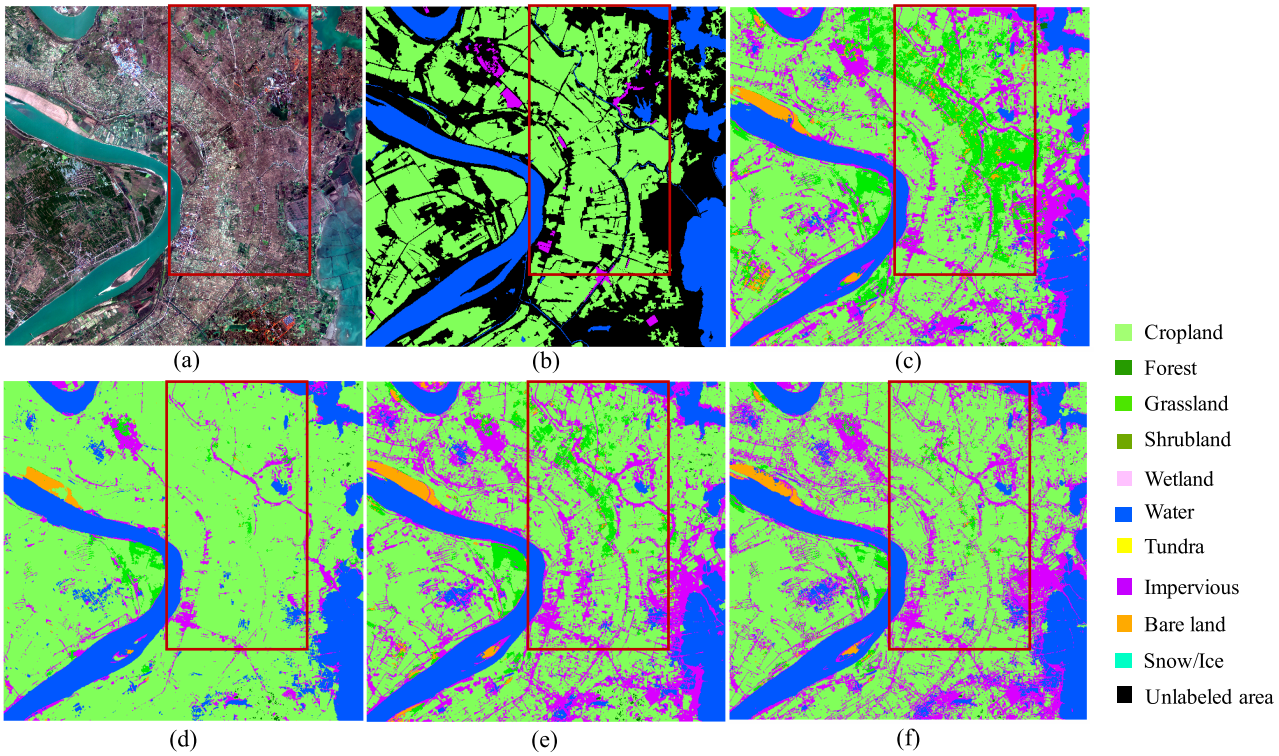


Fig. 5. Visualization of the land cover mapping results on the Gaofen image validation data set. The models are trained on the PID. (a) Image. (b) Ground truth. (c) Baseline. (d) GCE. (e) SCE. (f) Ours.

TABLE III

SUMMARY OF THE EXPERIMENTAL RESULTS (%) FOR EACH CLASS ON PITD. THE MODELS ARE TRAINED ON THE PID

Method	Original labels			Baseline			Our method		
Land cover type	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Cropland	76.79	80.44	78.47	79.75	85.72	82.63	81.45	88.01	**84.61**
Forest	83.94	85.43	84.68	89.48	86.54	87.99	85.45	91.96	**88.59**
Grassland	33.36	44.98	38.31	43.86	51.46	**47.36**	47.65	45.00	46.29
Water	64.40	83.14	72.58	67.83	82.71	74.54	71.54	82.31	**76.55**
Impervious	78.07	38.98	52.00	70.33	60.83	65.23	75.39	59.92	**66.77**
Bare land	90.09	85.80	87.89	95.65	84.32	89.63	96.21	84.90	**90.20**
OA (%)		74.96			79.74			**81.32**	

TABLE IV

EXPERIMENTAL RESULTS (%) OF DIFFERENT MODELS ON THE GID WITH VARIOUS PERCENTAGES OF PIXEL-LEVEL AND OBJECT-LEVEL NOISY LABELS

Method	without noise		Pixel-level noise percentage						Object-level noise percentage					
			40%		50%		60%		20%		30%		40%	
	OA	mIoU	OA	mIoU	OA	mIoU	OA	mIoU	OA	mIoU	OA	mIoU	OA	mIoU
Baseline [7]	98.06	90.85	97.41	88.54	92.25	76.85	53.97	30.99	94.67	84.93	93.92	82.06	92.15	74.98
GCE [31]	-	-	**97.90**	89.71	93.73	80.66	57.92	36.59	95.11	84.17	95.05	82.82	93.03	78.03
SCE [10]	-	-	97.62	89.42	92.87	77.39	50.82	31.25	95.77	84.93	94.75	84.45	93.73	78.09
Our method	**-**	**-**	97.80	**90.12**	**95.57**	**84.68**	**71.43**	**49.03**	**97.65**	**87.72**	**96.19**	**87.14**	**94.00**	**78.29**

shown in Table IV. The baseline results of different noise level data show that label noise degrades segmentation performance. However, the proposed method performs consistently better compared with the baseline method. As the percentage of pixel-level noise increases, the proposed method achieves more significant performance gains than the noise-robust methods (i.e., GCE and SCE). For the object-level noises, when the training set contains 30% object-level noises, the results of our method obtain the maximum benefit. Besides, we also train the model on the training set with 70% pixel-level noise or 50% object-level noise. Both the baseline experiments and the proposed method fail in those settings. The reason is that the noise correction is based on the probability estimate from the baseline model. A baseline model obtained from a training set with too much label noise will not produce reliable enough class estimations adopted for the noise correction approach. These findings shed light on the limits of the proposed noise correction method and verify that it can work

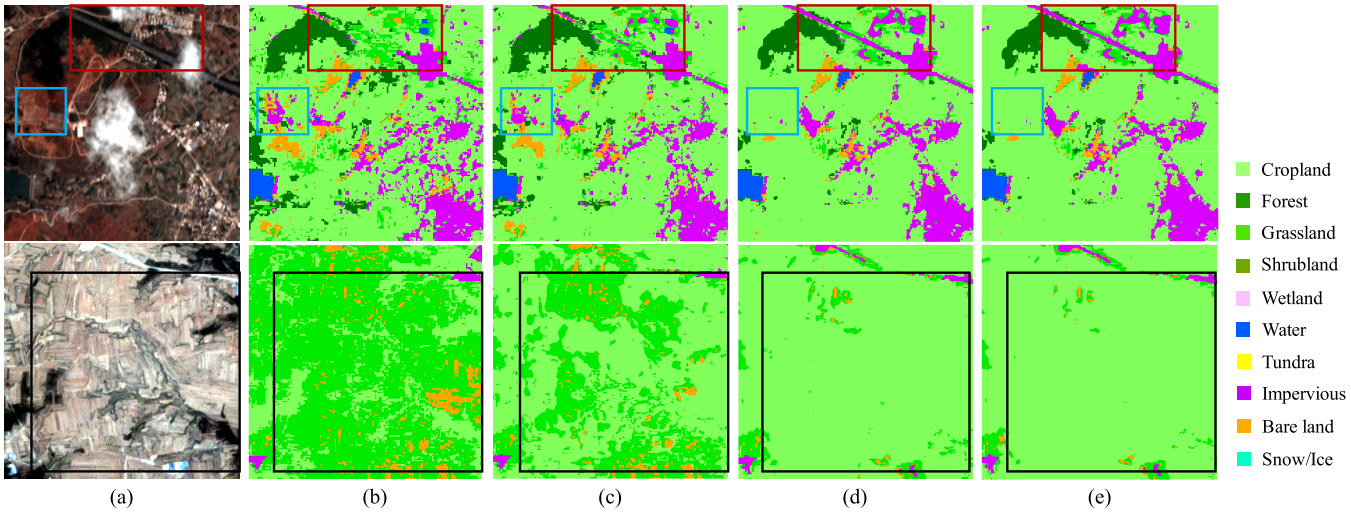


Fig. 6. Examples of the noise correction process on the training data set. Each row shows a sample image. (c)–(e) Demonstrate the iteratively updated labels at different epochs in the noise correction process. Note that due to the high uncertainty value estimated by our method in the cloud coverage area, the corresponding labels will not be updated. (a) Image. (b) Original noise label. (c) Epoch 1. (d) Epoch 15. (e) Epoch 25.

TABLE V

COMPARISON OF DIFFERENT THRESHOLDING STRATEGIES. THE MODELS ARE TRAINED ON THE PID. WE EVALUATE THE RESULTS OF THE FIXED THRESHOLD AND THE ADAPTIVE THRESHOLD ON BOTH THE PITD AND THE GAOFEN IMAGE VALIDATION DATA SET. THE AVERAGE SCORE OF FIVE TRIALS IS REPORTED

Method	PITD		GID	
	OA (%)	mIoU (%)	OA (%)	mIoU (%)
Baseline	79.74	61.50	86.15	64.48
Fixed threshold	81.04	62.78	88.01	68.18
Adaptive threshold	**81.32**	**63.02**	**89.45**	**69.72**

effectively when the pixel-level noise is no more than 60%, or the object-level noise is no more than 40%.

D. Effectiveness of the Noise Correction

In this section, we present a more detailed analysis of the noise correction approach on the PID. The ablation studies are conducted to investigate the importance of the adaptive threshold. We also visualize and present statistics for the noise correction process.

1) Fixed Threshold Versus Adaptive Threshold: We compare the use of the fixed threshold and the adaptive threshold on the noise correction module. The fixed threshold is empirically set to 0.1, which is decided based on the ablation study in Section V-A.2. As shown in Table V, our method obtains consistently better performance than the baseline model for both the fixed threshold and the adaptive threshold. The results demonstrate that the noise correction approach effectively corrects the noisy labels and thus maintains good performance despite the noise. Moreover, using the adaptive threshold achieves 0.28% and 1.44% improvement in OA and 0.24% and 1.54% improvement in mIoU on the two test data sets, respectively, compared with using a fixed threshold. The reason is that the adaptive threshold considers that different images have different levels of uncertainty, enabling a more reasonable selection of the labels that need to be corrected.

2) Visualization and Analysis of the Noise Correction: Example training images for the noise correction are shown in Fig. 6. As shown in the red rectangles, a clear road and buildings (both belonging to the impervious class) are gradually emerging during the noise correction phase although they are initially confused with the grassland and cropland. Also, the mislabeled impervious is corrected into the cropland, as shown in the blue rectangles. The black rectangles show that large tracts of pixels, mislabeled as grassland, are gradually updated to the correct cropland. Therefore, these examples reveal that the proposed noise correction approach can effectively eliminate the noise. Besides, we empirically find that the results become stable after 25 epochs during the noise correction phase. In the last few epochs, only minor improvement is observed for the updated results, which means that the noise correction phase is gradually converging.

3) Statistics of the Noise Correction: We calculate the total noise correction ratio on the training data set as 6.03%. Fig. 7 reports the transfer matrix of the label updates between classes, where each number is calculated by (the number of label updates from class A to B)/(total number of label updates from class A). Each row is interpreted as the distribution of the new classes that the original class is transferred to. Larger values in the transfer matrix indicate that the two classes are more likely to be confused in original noise labels.

E. Land Cover Mapping for China

As an application of the proposed approach, we produce the first 3-m resolution land cover map for the whole of China, as shown in Fig. 8(a). The time distribution of Planet satellite images used for land cover mapping in China is shown in Fig. 8(b). The Planet satellite images are collected in June, 2018, and March, 2019. There is a total of 26 000 image tiles with a size of 8000 $\times$ 8000 pixels in China. The prediction of each image tile using the proposed method takes 15 min on a single NVIDIA TITAN Xp GPU. We produce the land

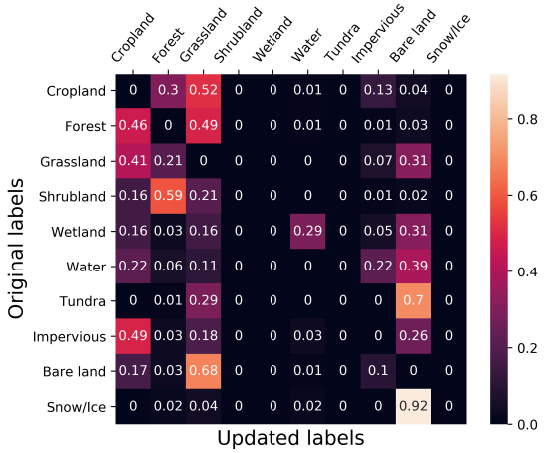


Fig. 7. Transfer matrix of the updated labels between classes. Each row is interpreted as the distribution of the new classes that the original class is transferred to. We ignore no updated labels and set the value on the diagonal to 0. The transfer distribution for each class is normalized.

TABLE VI

COMPARISON RESULTS OF DIFFERENT LOSS WEIGHTS ON GID. THE AVERAGE SCORE OF FIVE TRIALS IS REPORTED

Loss weight α	OA (%)	mIoU (%)
$\alpha = 0$	94.58	82.01
$\alpha = 0.2$	95.57	84.68
$\alpha = 0.5$	95.83	84.04
$\alpha = 1$	95.59	83.76

cover map for the whole of China within 3 days on 64 NVIDIA TITAN Xp GPUs. Here, we show the more detailed land cover map of Beijing in Fig. 9 as an example. We can see that, even though the training starts from a faulty lower-resolution result, our proposed approach can effectively produce refined 3-m resolution land cover maps. Compared with the latest GlobeLand30 2020 data set [50], we can see the advantages of the 3-m resolution land cover map.

V. DISCUSSIONS

A. Discussion of Hyperparameters

In this section, we discuss the setting of hyperparameters, i.e., the loss weight and threshold truncation K, of the proposed method.

1) Loss Weight α: α is a hyperparameter that balances the two loss terms, i.e., $\mathcal{L}_{\text{ce}}(\theta, \hat{Y})$ and $\mathcal{L}_{\text{ce}}(\theta, Y)$, in (1) during training. $\mathcal{L}_{\text{ce}}(\theta, \hat{Y})$ and $\mathcal{L}_{\text{ce}}(\theta, Y)$ denote the cross-entropy loss with the updated label and the cross-entropy loss with the original noise label, respectively. Therefore, a large α indicates that network training is more dependent on the original labels. We experiment with different loss weights on GID with 50% pixel-level noise, and the comparison results are shown in Table VI. When $\alpha = 0$, i.e., we only use the cross-entropy loss with the updated label, the performance drops significantly compared with other settings. Therefore, it is necessary to employ both the original labels and the updated labels for training since plenty of correct labels are contained in the original labels. It also shows that an appropriate weight of the cross-entropy loss with the original labels is favorable for the optimization of the model, while an excessive weight reduces the benefit of the label correction process.

TABLE VII

COMPARISON RESULTS OF DIFFERENT TRUNCATIONS ON GID. THE AVERAGE SCORE OF FIVE TRIALS IS REPORTED

Threshold truncation K	OA (%)	mIoU (%)
$K = 0$	95.48	83.74
$K = 0.1$	95.57	84.68
$K = 0.3$	95.42	83.87
$K = 0.5$	94.50	83.86

TABLE VIII

RESULTS (%) OF OUR METHOD WITH DIFFERENT INITIAL MODEL. THE MODELS ARE TRAINED ON THE PID

Method	PITD		GID	
	OA	mIoU	OA	mIoU
Original labels	74.96	55.52	-	-
Baseline1	79.74	61.50	86.15	64.48
Baseline2 (with drop-out strategy)	80.68	62.05	87.12	66.50
Our method (based on Baseline1)	81.32	63.02	89.45	69.72
Our method (based on Baseline2)	81.65	63.33	90.14	70.00

2) Threshold Truncation K: Due to the variations of the uncertainty for different images, we use adaptive thresholds to update the labels. A proper value of truncation K is needed to alleviate the insufficient update problem when the inputs are from simple scenes, leading to low thresholds. As shown in Table VII, when $K = 0$, i.e., we do not use the threshold truncation strategy, the mIoU is reduced by 0.94% compared with $K = 0.1$. The best performance is observed when $K = 0.1$, which is the setting we adopt in the experiments of Section IV.

B. Discussion of the Initial Model

It is essential to provide a reasonable initial model to estimate the uncertainty at the second phase. Therefore, in the first phase, we use early stopping and fixed learning rate strategies to prevent the initial model from overfitting to the noisy labels. The early stopping criterion is to obtain the highest accuracy on the validation set in the first phase. There are some regularization techniques, e.g., dropout strategy, which can be used to improve the performance of the initial model. As shown in Table VIII, when the performance of the first phase model is improved, our method can achieve better results. Therefore, our method can be used in conjunction with other regularization techniques to improve the performance.

C. Discussion of Spatial Heterogeneity in the Label Update Process

In the second phase (i.e., noise correction phase), we use the loss weight, threshold truncation K, and the number of training epoch to control the label update. As a result, most of the label updates are reasonable, as shown in Fig. 6. There is inevitably a small number of errors in the update, which may harm the spatial heterogeneity of the updated labels. However, the small spatial heterogeneity loss of training data in the noise

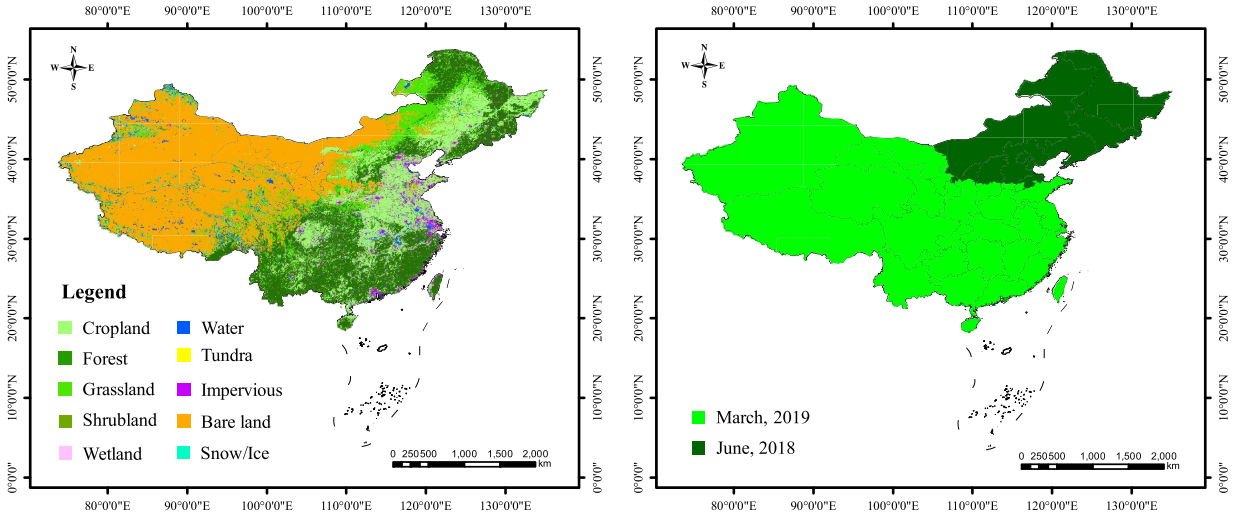


Fig. 8. 3-m resolution land cover map of China produced by the proposed approach. (a) 3-m resolution land cover map of China. (b) Time distribution of the 3-m resolution land cover map.

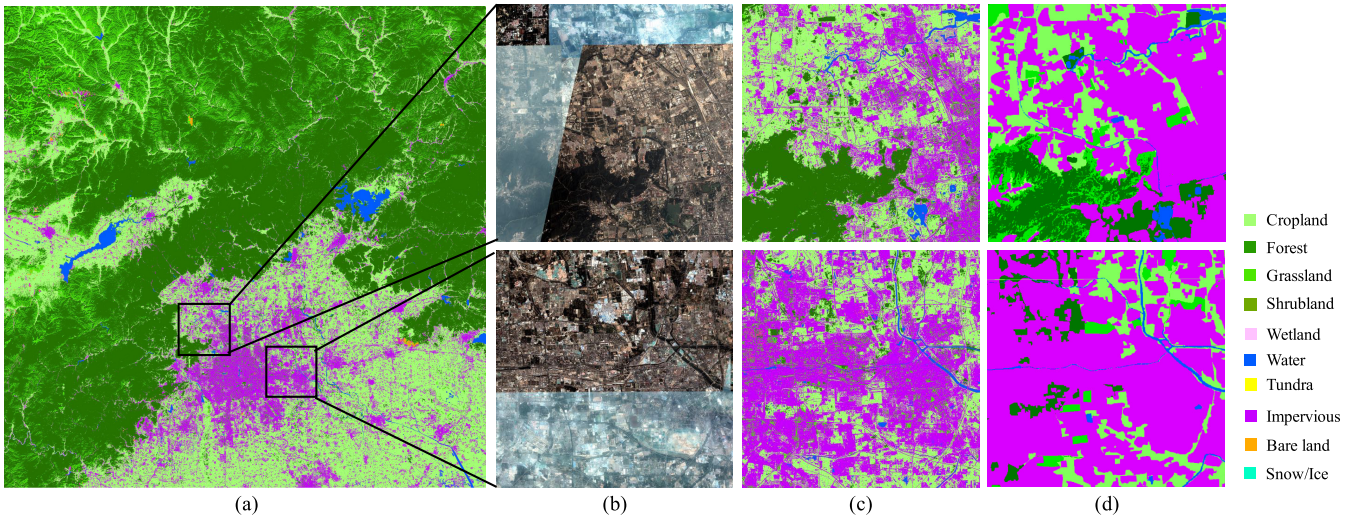


Fig. 9. 3-m resolution land cover map of Beijing in 2018 and the comparison results of GlobeLand30 in 2020. (a) 3-m resolution land cover mapping results of Beijing. (b) Planet image. (c) 3-m resolution land cover mapping results (ours). (d) GlobeLand30 in 2020.

correction phase has a minor effect on the spatial heterogeneity on the final results obtained by the third phase (i.e., training the network from scratch using the updated labels), because the most updated labels are correct. Specifically, the total noise correction ratio on the training data set is 6.03%, and most of the correction can reduce the confusion between different types over a vast region. As shown in Figs. 4 and 5(f) is the result obtained by using the updated training data, and Figs. 4 and 5(c) is the result obtained by using the original training data. Figs. 4 and 5(f) not only have better spatial heterogeneity than Figs. 4 and 5(c) (e.g., the road and water), but also has a further improvement in accuracy.

D. Weaknesses and Potential Strategies for Further Research

In this section, we analyze the weaknesses of the proposed method and introduce potential strategies for further research. The first weakness involves the hyperparameter tuning. Loss weight α controls the dependence on the original labels in this work. We will encode this assumption into the network, for example, in the last layer of the network or using a separate branch to estimate α dynamically. This can reduce the cost of the hyperparameter tuning and obtain a more reasonable dynamic α during the training.

The second weakness concerns the applicability of our approach. As the noise correction depends on the probability estimation of the model obtained in the first phase, it requires the baseline model to extract the useful information from the original noise labels. If the original labels contain too much noise (e.g., 70% noise), it is difficult for the noise correction approaches to take effect. Therefore, we assume that the excessive resolution difference (e.g., over ten times) between the land cover products and high-resolution images is not suitable for the noise correction methods. Similarly, if some

types individually contain too much noise in their original labels, e.g., Shrubland, Wetland, Tundra, and Snow/Ice in this work, it is hard to correct the labels for these types. In future work, we will add a small number of manually labeled samples of these types and employ few-shot learning methods to improve the performance.

VI. Conclusion

In this article, we propose a noise correction approach for large-scale land cover mapping. We demonstrate the proposed method is effective on both the real-world noise data set and the simulated noise data sets. Using the proposed approach with the existing 10-m resolution land cover product, we produce the refined 3-m resolution land cover maps without any human-labeled data. The noise correction approach would lead to lots of potential opportunities to use existing knowledge and results in remote sensing scenarios, such as road extraction using Open Street Map (OSM) data. In future research, we will further explore the characteristics of the noisy labels in the land cover scenario, then utilize them as prior knowledge to improve the results.

References

[1] X. Tong, W. Zhao, J. Xing, and W. Fu, "Status and development of China high-resolution Earth observation system and application," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3738–3741.

[2] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019.

[3] C. Robinson *et al.*, "Large scale high-resolution land cover mapping with multi-resolution data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12726–12735.

[4] X.-Y. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111322.

[5] R. Dong *et al.*, "Oil palm plantation mapping from high-resolution remote sensing images using deep learning," *Int. J. Remote Sens.*, vol. 41, no. 5, pp. 2022–2046, Mar. 2020.

[6] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping—Challenges and opportunities," 2020, *arXiv:2002.08254*. [Online]. Available: http://arxiv.org/abs/2002.08254

[7] R. Dong *et al.*, "Improving 3-m resolution land cover mapping through efficient learning from an imperfect 10-m resolution map," *Remote Sens.*, vol. 12, no. 9, p. 1418, 2020.

[8] N. Yokoya, P. Ghamisi, R. Haensch, and M. Schmitt, "2020 IEEE GRSS data fusion contest: Global land cover mapping with weak supervision [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 1, pp. 154–157, Mar. 2020.

[9] G. Grekousis, G. Mountrakis, and M. Kavouras, "An overview of 21 global and 43 regional land-cover mapping products," *Int. J. Remote Sens.*, vol. 36, no. 21, pp. 5309–5335, Nov. 2015.

[10] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 322–330.

[11] Y. Hua, S. Lobry, L. Mou, D. Tuia, and X. X. Zhu, "Learning multi-label aerial image classification under label noise: A regularization approach using word embeddings," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2020.

[12] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.

[13] K. Yi and J. Wu, "Probabilistic end-to-end noise correction for learning with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7017–7025.

[14] H. Zhu, J. Shi, and J. Wu, "Pick-and-learn: Automatic quality evaluation for noisy-labeled image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 576–584.

[15] P. Gong *et al.*, "Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017," *Sci. Bull.*, vol. 64, no. 6, pp. 370–373, Mar. 2019.

[16] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.

[17] *Land Cover CCI Product User Guide Version 2*, ESA, Libin, Belgium, 2017.

[18] M. Friedl and D. Sulla-Menashe, "MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 global 500m SIN grid V006 [data set]," in *Proc. NASA EOSDIS Land Processes DAAC*, vol. 10, 2015, pp. 1–6.

[19] C. Jun, Y. Ban, and S. Li, "Open access to earth land-cover map," *Nature*, vol. 514, no. 7523, p. 434, 2014.

[20] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6054–6068, Nov. 2017.

[21] J. Lee, J. Cardille, and M. Coe, "BULC-U: Sharpening resolution and improving accuracy of land-use/land-cover classifications in Google Earth Engine," *Remote Sens.*, vol. 10, no. 9, p. 1455, Sep. 2018.

[22] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS—A curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion," 2019, *arXiv:1906.07789*. [Online]. Available: http://arxiv.org/abs/1906.07789

[23] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[24] K. Malkin *et al.*, "Label super-resolution networks," in *Proc. Int. Conf. Learn. Represent.* 2018.

[25] A. E. Maas, F. Rottensteiner, and C. Heipke, "A label noise tolerant random forest for the classification of remote sensing data based on outdated maps for training," *Comput. Vis. Image Understand.*, vol. 188, Nov. 2019, Art. no. 102782.

[26] B. B. Damodaran, R. Flamary, V. Seguy, and N. Courty, "An entropic optimal transport loss for learning deep neural networks under label noise in remote sensing images," *Comput. Vis. Image Understand.*, vol. 191, Feb. 2020, Art. no. 102863.

[27] B. Frenay and M. Verleysen, "Classification in the presence of label noise: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 845–869, May 2014.

[28] D. Rolnick, A. Veit, S. Belongie, and N. Shavit, "Deep learning is robust to massive label noise," 2017, *arXiv:1705.10694*. [Online]. Available: http://arxiv.org/abs/1705.10694

[29] K.-H. Lee, X. He, L. Zhang, and L. Yang, "CleanNet: Transfer learning for scalable image classifier training with label noise," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5447–5456.

[30] C. G. Northcutt, L. Jiang, and I. L. Chuang, "Confident learning: Estimating uncertainty in dataset labels," 2019, *arXiv:1911.00068*. [Online]. Available: http://arxiv.org/abs/1911.00068

[31] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8778–8788.

[32] J. Huang, L. Qu, R. Jia, and B. Zhao, "O2U-net: A simple noisy label detection approach for deep neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3326–3334.

[33] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5552–5560.

[34] S. Sukhbaatar, J. Bruna, M. Paluri, L. Bourdev, and R. Fergus, "Training convolutional networks with noisy labels," 2014, *arXiv:1406.2080*. [Online]. Available: http://arxiv.org/abs/1406.2080

[35] A. Ghosh, N. Manwani, and P. S. Sastry, "Making risk minimization tolerant to label noise," *Neurocomputing*, vol. 160, pp. 93–107, Jul. 2015.

[36] A. Ghosh, H. Kumar, and P. Sastry, "Robust loss functions under label noise for deep neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 1919–1925.

[37] Z. Yu *et al.*, "Simultaneous edge alignment and learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 388–404.

[38] D. Acuna, A. Kar, and S. Fidler, "Devil is in the edges: Learning semantic boundaries from noisy annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11075–11083.

[39] Z. Lu, Z. Fu, T. Xiang, P. Han, L. Wang, and X. Gao, "Learning from weak and noisy labels for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 486–500, Mar. 2017.
[40] M. S. Ibrahim, A. Vahdat, M. Ranjbar, and W. G. Macready, "Semi-supervised semantic image segmentation with self-correcting networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12715–12725.
[41] T. F. Y. Vicente, M. Hoai, and D. Samaras, "Noisy label recovery for shadow detection in unfamiliar domains," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3783–3792.
[42] B. B. Damodaran, K. Fatras, S. Lobry, R. Flamary, D. Tuia, and N. Courty, "Wasserstein adversarial regularization (WAR) on label noise," 2019, *arXiv:1904.03936*. [Online]. Available: https://arxiv.org/abs/1904.03936
[43] 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*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[44] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19.
[45] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5693–5703.
[46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[47] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*. [Online]. Available: https://arxiv.org/abs/1607.08022
[48] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
[49] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Sásstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[50] C. Jun. *Remote Sensing Mapping of Global Land Cover*. Accessed: Nov. 2020. [Online]. Available: http://www.globallandcover.com

Runmin Dong (Graduate Student Member, IEEE) received the bachelor's degree in information and computing science from the Department of Science, Beijing Jiaotong University, Beijing, China, in 2017. She is pursuing the Ph.D. degree in ecology with the Department of Earth System Science, Tsinghua University, Beijing.

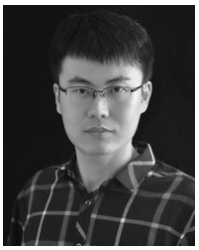
Her research interests include remote sensing image processing, deep learning, land cover mapping, and image super-resolution reconstruction.

Weizhen Fang received the M.S. degree in photogrammetry and remote sensing from the Institute of Remote Sensing and GIS, Peking University, Beijing, China, in 2020.

He is working as an Engineer with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. His research interests include big earth data, signal processing, smart city and the combination of remote sensing, and artificial intelligence (AI).

Haohuan Fu (Member, IEEE) received the Ph.D. degree in computing from the Imperial College London, London, U.K, in 2009.

He is a Professor with the Ministry of Education, Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China. He is also the Deputy Director of the National Supercomputing Center, Wuxi, China. His research interests include design methodologies for highly efficient and highly scalable simulation applications that can take advantage of emerging multi-core, many-core, and reconfigurable architectures, and make full utilization of current Peta-Flops and future Exa-Flops supercomputers; and intelligent data Management, analysis, and data mining platforms that combine the statistics methods and machine learning technologies.



Lin Gan (Member, IEEE) is an Associate Researcher with the Department of Computer Science, Tsinghua University, Beijing, China, and the Assistant Director with the National Supercomputing Center, Wuxi, China. His research interests include high-performance solutions to scientific applications based on state-of-the-art platforms such as CPU, field-programmable gate arrays (FPGAs), and GPUs.

Dr. Gan was a recipient of the 2016 ACM Gordon Bell Prize, the 2017 ACM Gordon Bell Prize Finalist, the 2018 IEEE-CS TCHPC Early Career Researchers Award for Excellence in HPC, the Most Significant Paper Award in 25 Years awarded by IEEE FPL 2015, the 2017 Tsinghua-Inspur Computational Earth Science Young Researcher Award, and so on.

Jie Wang is an Associate Professor with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His major research interest is machine learning for land monitoring and multisource data analysis.

Peng Gong built the Department of Earth System Science and served as the Dean for the School of Sciences, Tsinghua University, Beijing, China. He also served as the founding Director for the Tsinghua Urban Institute, Beijing. He had previously taught at the University of Calgary, Calgary, AB, Canada, and the University of California, Berkeley, Berkeley, CA, USA. He is the Chair Professor of Global Sustainability with the University of Hong Kong, Hong Kong. He has authored or coauthored more than 600 articles and 8 books. He chaired/co-chaired eight Lancet Commission reports on climate change and health, and healthy cities in China. His major research interests include mapping, monitoring and modeling of global environmental change, and modeling of environmentally related infectious diseases such as schistosomiasis, avian influenza, dengue and COVID-19, and healthy and sustainable cities.