

# Feature Consistency Constraints-Based CNN for Landsat Land Cover Mapping

Xuemei Zhao<sup>ID</sup>, Luo Liang<sup>ID</sup>, Jun Wu<sup>ID</sup>, Haijian Wang, and Xingyu Gao<sup>ID</sup>

**Abstract**—The cascade of convolution layers and the end-to-end training process facilitate CNN feature extraction and transmission, and promote the success of CNN in image processing. However, the drawback of heavily relying on large-scale high-quality training samples restricts its applications. To avoid costly and unrealistic manual annotations for large-scale remote sensing images, existing land cover maps are considered as an alternative to manual annotations, in which noisy labels are inevitable. To alleviate the impact of noisy labels, this article proposes to improve the consistency feature learning ability of CNNs as a feasible solution in practical land cover mapping. First, an intraclass feature consistency constraint is introduced to maintain the consistency of CNN feature maps for the same class. Then, an inter-iteration feature consistency constraint is employed to guide the network to learn features that are consistent with the whole underlying distribution inside a minibatch. These two feature consistency constraints work in a cooperative and complementary manner with the traditional cross-entropy, and together improve the consistency feature learning ability of the proposed feature consistency constraints-based CNN (FCNet). Experimental results demonstrate the effectiveness of the proposed FCNet. Extensive experiments on different network structures validate the generalization of the proposed feature consistency constraints.

**Index Terms**—CNN, feature consistency constraint, land cover mapping, Landsat images, loss function.

## I. INTRODUCTION

**L**ANDSAT is one of the most commonly used data sources in large-scale land cover mapping due to its long-term and appropriate observation ability [1], [2], [3]. Benefitting from the manually collected labels, supervised classifiers have achieved great success in land cover mapping [4], [5]. Among them, CNN is becoming an increasingly relevant topic in land cover mapping since it has made many breakthroughs in computer

vision [6], [7]. These improvements promote its use in land cover mapping [8], [9], [10]. As an end-to-end classifier, CNNs transmit input remote sensing images to land cover mapping products without any human interaction. What it relies on are the large-scale high-quality training samples, the architectures, and the loss functions.

Benefitting from the large capacity of network structures, CNNs can learn various features of the same class. Actually, it learns a kind of feature through a convolution kernel in one layer and then fuses them through the stacking of convolution layers. In this process, a convolution kernel extracts and fuses image features inside a receptive field to make it robust to noisy labels, and the stacking of a large number of convolution kernels makes the network adaptable to various features. Yet, the multi-to-one process is irreversible, and there exists some confusing information, especially near the target boundary. In addition, noisy labels provide conflict information to the network, and thus, it is difficult for the network to learn consistent formation. The uncertainty brought by noisy labels may confuse CNNs and lead to unsatisfactory classification results [11], [12]. What's worse, the network cannot distinguish benefits from harmful information during training, and thus loses the controlling ability of the learning preference from the information point of view. This results in the phenomenon that the higher the quality of the training set is, the more accurate and comprehensive the information it provides. Consequently, the more general is the trained CNN. While for Landsat land cover mapping, enough high-quality training samples are difficult to access. Although some methods aim to increase the information transmission capability of CNNs by improving the network structures, but they still lack guarantee about the consistency of information. This makes it very important for the network to extract and transfer consistent information from existing imperfect training samples accurately and efficiently.

Improving architectures improve the feature extraction and transfer capability, but the end-to-end process hinders the introduction of additional constraints, which forces the network to learn consistent information. The loss function is used to evaluate the discrepancy between the predicted results and the ground truth. Thus, modifications are performed on it to control the learning preference of CNN [13], [14]. Some of the defects hidden in the training samples can be overcome if a proper loss function is used. For example, giving more weights to minority classes is efficient to deal with class imbalanced training samples [15], [16]. Image information-related constraints are introduced to deal with noisy labels [17], [18]. Due to the

Manuscript received 10 October 2022; revised 6 January 2023 and 21 February 2023; accepted 8 March 2023. Date of publication 16 March 2023; date of current version 12 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 42261061, 41801233, and 41761087, and in part by the Natural Science Foundation of Guangxi Province under Grant 2020GXNSFBA159012. (Corresponding author: Jun Wu.)

Xuemei Zhao, Luo Liang, and Jun Wu are with the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: 374010101@qq.com; ob0660@163.com; wujun93161@163.com).

Haijian Wang and Xingyu Gao are with the Guangxi Key Laboratory of Manufacturing System and Advanced Manufacturing Technology, School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: whj19870608@guet.edu.cn; gxy1981@guet.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3257836

cascading nature of CNNs, detailed information is lost, resulting in unsatisfactory results. Introducing boundary information to define loss function improves CNN's learning ability on detailed information [19]. However, the following problems still exist.

- 1) *Problems of the training set:* Affected by imaging illumination conditions, pixels representing the same class present different spectral features in Landsat images. The differences in spectral features in the same class increase the difficulty of recognizing different classes on the one hand and seriously affect the labeling of training samples on the other hand. For Landsat land cover mapping, both the variations spectral features and noisy label problems exist simultaneously.
- 2) *Problems in feature extraction:* The cascading of convolution layers will result in similar features in adjacent areas. The linear weighted sum essence of convolution layers may submerge some information that is important for land cover mapping, thus leading to similar CNN feature maps for different classes, which confuses the classifier. This inconsistency of CNN feature maps usually occurs near the object boundary, especially for Landsat images with 30-m resolution.
- 3) *Problems in network training:* Traditional CNNs use shuffled minibatches to approximate the whole dataset. Each minibatch provides various gradients for the network to learn general features, while the randomly sampled minibatches cannot well simulate the whole underlying distribution, due to the unbalanced natural distribution of land cover types. This means that the information provided by the minibatches is a biased estimation, which leads to the oscillation of the trained network.

Despite all this, the success of loss functions in controlling the learning preference of CNN gives us a chance to force the network to learn consistent information from noisy labels in Landsat images. Therefore, we propose to incorporate feature consistency constraints to the loss function to guide the learning preference of a CNN. The main contributions of the proposed FCNet are summarized as follows.

- 1) An intraclass feature consistency constraint is employed to minimize the discrepancy of learned features inside the same class to improve the learning ability of the network, with the presence of noisy labels.
- 2) An inter-iteration feature consistency constraint is introduced to guild the network to learn consistent features within a class among iterations when using minibatches to approximate the underlying distribution of the data.
- 3) We demonstrate the effectiveness and complementarity of the proposed feature consistency constraints and their extensibility on other network structures.

## II. RELEVANT WORK

### A. CNN Architecture

The design of the CNN architecture determines how image features are learned and transmitted in the network. So, early research focused on increasing or expanding the number of

network layers [20]. Along with the increase in model capacity, new problems such as the vanishing-gradient problem arise. ResNet, which takes skipping connection as the core idea, overcomes this drawback and allows the network to go deeper [21]. However, a deeper network does not mean stronger learning ability, especially on specific tasks. To fully utilize the difference between different CNN architectures, multibranch-based CNNs are proposed. Zhao et al. [22] used 1-D and 2-D convolution to learn spectral and spatial features, respectively, to improve the learning ability of CNN. Different representations also have a great impact on the learning ability of CNNs. Using each branch to learn information from a representation can also improve the performance [23], while the detailed information lost in the cascading of convolution layers cannot be compensated with multiple branches. [24] used dense connections to allow the information to be transmitted in nonadjacent layers. As demonstrated in [25], dense connections can not only strengthen the feature extraction but also alleviate the vanishing-gradient problem to some extent.

The attention mechanism is first proposed in natural language processing to give different weights to the input. It can be used in soft, hard, global, local, and other ways to capture long-range connections [26]. However, it cannot accurately describe the relationship between the source and the target. To overcome this drawback, the self-attention mechanism is proposed. In [27], the proposed multihead self-attention module completely constitutes the network architecture, called transformer. Even though it is computationally expensive, it outperforms other architectures when the training samples are sufficient [28]. To alleviate the dependency of GPU memory, Huang et al. [29] proposed a criss-cross attention module to capture the full image dependencies from all pixels and reduce GPU memory usage simultaneously. Jiang et al. [30] proposed an online attention accumulation strategy to obtain more integral object regions at different training phases.

### B. Loss Function

The loss function evaluates the differences between the predicted result and the ground truth, so as to provide the gradient of back-propagation. Accordingly, there are many contributions in this regard [31], [32]. Among all these loss functions, cross-entropy is the most widely used [33]. However, affected by the imbalanced and noisy labels, cross-entropy cannot always reach its optimum, in practical applications. Luo et al. [15] combined focal loss and cross-entropy to alleviate the impact of class imbalance and outliers. Potential noisy labels were automatically compensated by the asymmetric loss function proposed in [34]. The loss function can not only solve the imbalanced and noisy label problems but also deal with multiscale data conveniently. Considering the scale differences between small and large objects, Zhou et al. [35] used contrastive loss and cross-entropy loss to define a new loss function that is suitable for small sample target. Yang et al. [36] used discrete wavelet transforms to divide the image into patches with different sizes and then accumulated the information through the structure similarity loss function.

As previously discussed, detailed information will be lost during the cascading of convolution layers. To overcome this drawback, Borse et al. [19] proposed a boundary-aware loss function to learn the degree of parametric transformations between the predicted and the ground-truth boundaries. Along with the cross-entropy, it achieved satisfactory results. Similarly, Guarda et al. [37] proposed an adaptive distortion metric-based loss function to improve the rate-distortion performance inside a neighborhood. To improve the learning ability of networks on interest targets, Liu et al. [38] proposed a task-specific loss function. Choi and Kil [13] proposed to learn compact and discriminative high-level features by using a radial basis function kernel-based loss function. A pairwise Gaussian loss function is employed to address the intraclass compactness and ensure good interclass separability [39]. The learned knowledge about different objects is accumulated by a loss transferring method proposed in [40].

### C. Class-Related Information

Different from CNNs which extract information by automatically learning from the training set, traditional supervised and unsupervised methods fully utilize the class-related information to achieve optimum results. Zhang et al. [41] proposed to use information entropy, conditional entropy, and mutual information to construct class-specific regularizations to describe the internal relationship, which is an extension of their previous work [42]. Fisher information can also be used to describe class-related information by decreasing intraclass scatter and enlarging interclass scatter [43]. Besides the information used for evaluating the intra- and interclass similarity, the way to use this information also has a great impact on the performance of the method. Zhu et al. [44] used intravideo and intervideo distance metrics to propose a distance learning method. Leng et al. [45] proposed an intra-inter-scale discrimination index to balance the spectral difference inside a superpixel and between superpixels. However, the discrimination ability of the scale-based index is limited. To fully utilize consistent and diverse information, Se et al. [46] imposed a clustering structure constraint on the subspace self-representation and employed an exclusivity constraint term to enhance the diversity of specific representations. In subspace clustering methods, a sparse construction error is used to describe the interclass and intraclass relationship [47]. Class-level sparse and globally low-rank constraints are also important class-related information [48]. Rong et al. [49] proposed to use subdictionaries to capture class-specific information and class-shared dictionary to model class-shared information. Class-related information has also been introduced to CNN architectures. Fan et al. [50] proposed an intraclass discriminator to learn intraclass boundaries to improve the recognition ability of the objects. Alhuzali and Ananiadou [51] proposed a triplet center loss as an auxiliary task to cross-entropy loss. The triplet center loss is defined by the distance of a pixel to its center and the distance to other centers. In [52], an intraconcentration and interseparability-based loss function, which is defined by the distance of a pixel to its center and the distance between class centers, is used to make intraclass samples concentrate and interclass samples separable.

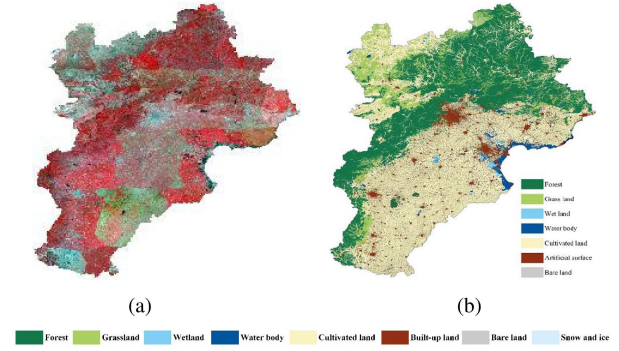


Fig. 1. Original Landsat images and the corresponding reference of Jing-Jin-Ji region. (a) Original Landsat images. (b) Reference.

## III. METHODOLOGY

In this article, we present a feature consistency constraints-based CNN (FCNet) to fully utilize the consistent information to improve the performance of CNN-based Landsat land cover mapping. First, data sources and corresponding preprocessing are introduced in Section III-A. Motivations inspired by the characteristics of the data sources are analyzed in Section III-B. On the basis of feature consistency theory, intraclass and inter-iteration consistency constraints are defined in Sections III-C and III-D. Then, Sections III-E and III-F introduce the overall loss function and the overall architecture of the proposed FCNet. Finally, the optimization of the FCNet is analyzed in Section II-I-G.

### A. Data Sources and Preprocessing

Jing-Jin-Ji region is selected as the study area since there exist many types of land surface coverage. The original Landsat images of this area are shown in Fig. 1(a), which are visually expressed by near-infrared, red, and green bands. Red, green, blue, near-infrared, and the other two mid-infrared bands are used to train the network. A large number of training samples are necessary to train a network. However, manually labeling Landsat images is labor-intensive, and labeling accuracy is difficult to be guaranteed. Existing high-accuracy land cover products can be considered as an alternative. Land Cover Map of the People's Republic of China (1:1000000) is a widely accepted product with 94% overall accuracy for the first-level classes and 86% overall accuracy for the second-level classes. According to the land surface coverage distribution and the recognition ability of Landsat images in the study area, a reference land cover map, shown in Fig. 1(b), is produced from the Land Cover Map of the People's Republic of China (1:1000000). In this study area, 1280 nonoverlapping image patches with  $512 \times 512$  pixels are sampled, in which 640 image patches construct the training set and another 640 images are used as the validation set.

### B. Motivation

Consistency of image features is the essence of human recognition and manually designed classifiers. However, the limitations of manually designed features and constraints heavily restrict the recognition ability of classifiers, resulting in their inability in recognizing partially well-defined features. Along with



the improvement of remote sensing image observation ability, image features show a trend of diversification. This diversity makes it difficult for traditional algorithms to effectively recognize classes with complex characteristics. The large capacity of CNNs provides a solution for complex image classification. However, existing CNN-based methods increase the learning ability by improving the network structure and introducing constraints in the loss function, and they commonly ignore the consistency of learned CNN feature maps. CNNs use a large number of convolution kernels to learn the diverse features of the same class. This process is heavily affected by noisy labels since they provide contrary labels for similar features. The label-based training method will also lead to inconsistent features learned by the network, with the presence of noisy labels. Actually, the learned features of the same class should be consistent. This can be considered as a constraint forcing the network to resist the impact of noisy labels. Consequently, this article proposes to use intraclass variance to constrain the network on learning consistency features. Despite the intraclass feature consistency, using minibatches to approximate the whole underlying distribution is another factor influencing the learning ability of the network. The diversity gradients of minibatches promote the network to learn general information while also causing the oscillation of the losses during training. This can be partially explained that there are obvious differences between the distributions of subsets and the whole dataset, due to the unbalanced natural distribution of land cover types. To address this problem, this article approximates the distribution with the KL divergence of minibatches and the whole underlying distribution, under the identity covariance matrix assumption.

### C. Intraclass Feature Consistency Constraint

From a class-wise perspective, the learned features representing the same class should be similar. This inspires us to define an intraclass feature consistency constraint to force the network to learn consistent features for the same class. As is well known, class variance is a simple yet effective index describing the variance of features inside a class. Therefore, this article proposes to use class variance to define the intraclass feature consistency constraint. Let  $x_i$  represent the output feature of the  $i$ th pixel in the CNN feature map, then the variance of the  $j$ th class is calculated as

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n_j} (x_i - \mu_j)^2}{n_j - 1}} \quad (1)$$

where  $n_j$  represents the number of pixels in the  $j$ th class, and  $\mu_j$  is the mean of the  $j$ th class, which can be defined as

$$\mu_j = \frac{\sum_{i=1}^{n_j} x_i}{n_j}. \quad (2)$$

To force the network to learn consistent intraclass information is to minimize the variance of each class. Therefore, the intraclass feature consistency constraint can be defined as

$$L_{\text{var}} = \frac{\sum_{j=1}^k \sigma_j}{k} \quad (3)$$

where  $k$  is the number of classes.

### D. Intraclass Feature Consistency Constraint

Consistent information facilitates the network to learn essential information of each class, while the unexplainability and complexity of the network make the gradient propagation methods difficult to converge to the global optimum. In addition, using minibatches to approximate the whole underlying distribution is a challenging task due to the various features of the same class. To make it flexible, this article assumes that there is no distribution shift when using minibatches to approximate the whole underlying distribution, and no feature variance inside a class, i.e., each class has an identity matrix as its covariance matrix in both minibatches and the whole distribution. We use KL divergence to evaluate the discrepancy between minibatch distributions  $p_m$  and the underlying distribution of the whole training set  $p_u$ :

$$\text{KL}(p_m||p_u) = p_m \log \left( \frac{p_m}{p_u} \right) \quad (4)$$

Taking Gaussian distribution into consideration, (4) can be rewritten as

$$\begin{aligned} \text{KL}(p_m||p_u) &= \int_x \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(x-\mu_m)^2}{2\sigma_m^2}} \log \frac{\frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(x-\mu_m)^2}{2\sigma_m^2}}}{\frac{1}{\sqrt{2\pi}\sigma_u} e^{-\frac{(x-\mu_u)^2}{2\sigma_u^2}}} dx \\ &= \log \frac{\sigma_u}{\sigma_m} - \frac{1}{2} + \frac{\sigma_m^2 + (\mu_m - \mu_u)^2}{2\sigma_u^2} \end{aligned} \quad (5)$$

where  $\mu_m, \mu_u, \sigma_m$ , and  $\sigma_u$  represent the means and covariances of  $p_m$  and  $p_u$ , respectively. Under the identity covariance matrix assumption, the KL divergence can be simplified as

$$\text{KL}(p_m||p_u) = \frac{(\mu_m - \mu_u)^2}{2}. \quad (6)$$

For a specific class, for example, the  $j$ th class,  $\mu_m$  can be easily calculated inside a minibatch according to (2):

$$\mu_m = \mu_j = \frac{\sum_{i=1}^{n_j} x_i}{n_j}. \quad (7)$$

Nevertheless,  $\mu_u$  is unknown. Actually, the network is learning from the whole training set with iteration. Therefore, this article proposes to use the accumulated mean instead of the mean of the whole underlying distribution. To be flexible, the accumulated mean of the  $j$ th class in the  $t$ th iteration is calculated as

$$\mu_u \approx \mu_{\text{cum-}j}^{(t)} = \frac{\mu_j^{(t)} + \mu_{\text{cum-}j}^{(t-1)}}{2} \quad (8)$$

where  $\mu_j^{(1)}$  is the mean of the first minibatch, and the accumulated mean is the average of the current minibatch and all the historical minibatches. Equation (8) means that the accumulated mean of the  $t$ th minibatch can be calculated by the mean of the  $t$ th minibatch and the accumulated mean of all the  $(t-1)$ th minibatches. Then, the KL divergence between the minibatches of the  $j$ th class in the  $t$ th iteration and the whole unknown underlying distribution can be approximated as

$$\text{KL}(p_m||p_u) \approx \frac{(\mu_j^{(t)} - \mu_{\text{cum-}j}^{(t)})^2}{2}. \quad (9)$$

$$\begin{aligned} \frac{d\text{Loss}_{\text{fc}}}{dx_i} &= \frac{\lambda_1}{k} \sum_{j=1}^k \frac{\frac{2}{n_j-1} \left( x_i - \frac{\sum_{i=1}^{n_j} x_i}{n_j} \right) \left( 1 - \frac{1}{n_j} \right)}{\sqrt{\frac{\sum_{i=1}^{n_j} \left( x_i - \frac{\sum_{i=1}^{n_j} x_i}{n_j} \right)^2}{n_j-1}}} \\ &\quad + \frac{\lambda_2}{C_k^2} \sum_{j=1}^k \left( \left( \frac{\sum_{i=1}^{n_j} x_i}{n_j} \right)^{(t)} - \mu_{\text{cum-}j}^{(t)} \right) \left( \frac{1}{n_j} \right) \\ &= \frac{2\lambda_1}{k(n_j)^2} \sum_{j=1}^k \frac{n_j x_i - \sum_{i=1}^{n_j} x_i}{\sigma_j} \\ &\quad + \frac{\lambda_2}{C_k^2 (n_j)^2} \left( \left( \sum_{i=1}^{n_j} x_i \right)^{(t)} - n_j \mu_{\text{cum-}j}^{(t)} \right). \quad (15) \end{aligned}$$

Equation (15) demonstrates that the proposed  $\text{Loss}_{fc}$  can be differentiated by  $x_i$ . As the output of the network,  $x_i$  can be differentiated by the parameters of the network. According to the gradient propagation theory,  $\text{Loss}_{fc}$  can also be differentiated by the parameters of the network. To further explore how the proposed  $\text{Loss}_{fc}$  guides the network to learn consistent features, the two terms in (15) are explained as follows. 1) The first term describes the differences between the estimated result, which is based on a pixel value and the number of pixels inside a class, and the true sum of this class, considering the variance within the class. This is equivalent to estimating the ground truth with the representative pixels. The resulted gradient can guide the network to evolve toward the ground truth inside each class, and thus force the network to learn intraclass compactness features. Consequently, the learned features tend to be intraclass consistent. 2) The second term describes the differences between the sum of intraclass features in the current iteration and the accumulated features with all the former iterations, which is estimated by the accumulated mean of a class and the number of pixels of the class. This is equivalent to approximating the accumulated features with features of the current minibatch. The resulted gradient can force the network to learn global features with image features inside a minibatch. Consequently, the network can learn consistent information inside a class through iterations.

Considering the output of the network as a whole, the derivatives of cross-entropy can be written as

$$\frac{d\text{Loss}_{ce}}{dx_i} = -\frac{d \sum_{i=1}^n l_i \log x_i}{dx_i} = -\frac{l_i}{x_i} \quad (16)$$

where  $l_i$  represents the ground-truth label of pixel  $x_i$  outputted by the network. The derivatives of  $x_i$  with respect to the network parameters are not discussed in this article, since they are the same for  $\text{Loss}_{ce}$  and  $\text{Loss}_{fc}$ . Compared with (15), (16) can only provide a gradient for the network to approach the ground-truth label. In addition, this approach is equivalent to each class, resulting in the accuracy scarifying in minority classes and classes with noisy labels. With the guidance of the gradients provided by (15), the proposed feature consistency constraints can significantly improve the learning ability of the network, especially with the presence of noisy labels.

#### IV. EXPERIMENTS AND RESULTS

In this section, we first introduce the backbones and implementation details in Section IV-A. Then, the proposed FCNet is compared with state-of-the-art methods focusing on improving the learning ability of the network through modifying the loss function in Section IV-B. Further, the proposed feature consistency constraints are performed on other networks to test their generality in Section IV-C. An ablation study is carried out in Section IV-D. The analysis of parameters and discussion about the proposed FCNet are shown in Sections IV-E and IV-F.

##### A. Backbones and Implementation Details

To assess the effectiveness of the proposed FCNet, three widely used backbones are employed, namely, FCN [53],

PSPNet [54], and DeepLab V3+ [55]. The proposed feature consistency constraints are introduced to the three backbones. Adam optimizer is used to train the network modules. All parameters are trained based on the models pretrained on ImageNet. The learning rate is set to be  $10^{-4}$ , and the weight decay is  $10^{-4}$ . The number of epochs is set to 300 on four NVIDIA Tesla V100 GPUs with a batch size of 52.

##### B. Comparison With the State of the Art

To verify the effectiveness of the proposed FCNet, it is performed on the study area shown in Fig. 1. Detailed classification results are displayed in Fig. 3. Classic frameworks such as FCN, PSPNet, and DeepLab V3+ tend to expand the built-up land to some extent. This indicates the weak control of the loss function on the learning preference. Because of the stacking of convolution layers, detailed information such as edges between different classes is lost. Without a better control of the learning preference, a network tends to classify the ambiguous areas as the most responsive class, in this case, the built-up land. Fig. 3(a6)–(a8) demonstrates the controlling ability of loss functions on the learning preference of the network, among which the proposed FCNet is the best. Similar phenomena arise in other areas within the study area, as shown in Fig. 3(b1)–(b8) and (d1)–(d8).

The linearly distributed water body is also a challenge for deep CNNs, impacted by the stacking of convolution layers. Existing networks either obtain thicker or thinner classification results, compared with the real river range. However, the proposed FCNet achieves better classification results not only on the water body but also on the nearby built-up land and grassland, as shown in Fig. 3(b8). The confusion between grassland and forest is an obvious problem in the training set, as shown in Fig. 3(c2). Accordingly, networks trained by the corresponding training set are easily affected, as shown in Fig. 3(c3)–(c5) and (c7). CEL-VTCL [51] employed a triplet center loss function to force the network to learn compact and discriminative features, and obtained satisfactory classification results, as shown in Fig. 3(c6). The proposed FCNet uses the feature consistency characteristic inside a class to enhance the connections among pixels representing the same class and obviously decreases the impact of noisy labels, as shown in Fig. 3(c8). As shown in Fig. 4, wetland occupies the least proportion in the training set. For a network that equally learns from each class of the training set, the information on minority classes is insufficient. Therefore, the classification results shown in Fig. 3(d3)–(d5) cannot recognize the wetland located in the upper part of the image. Although balancing the sample proportion or introducing corresponding features can improve the learning ability of the model, the minority classes should be known in advance, which is impractical in some applications I2CS [52]. The proposed FCNet can force the network to learn the compactness features of each class and thus improve the recognition ability of minority classes. Even though the effect may not be as good as balancing class proportions, there is no need for prior knowledge about minority classes.

To qualitatively evaluate the classification results in the whole study area, the accuracy of each class and the overall accuracies



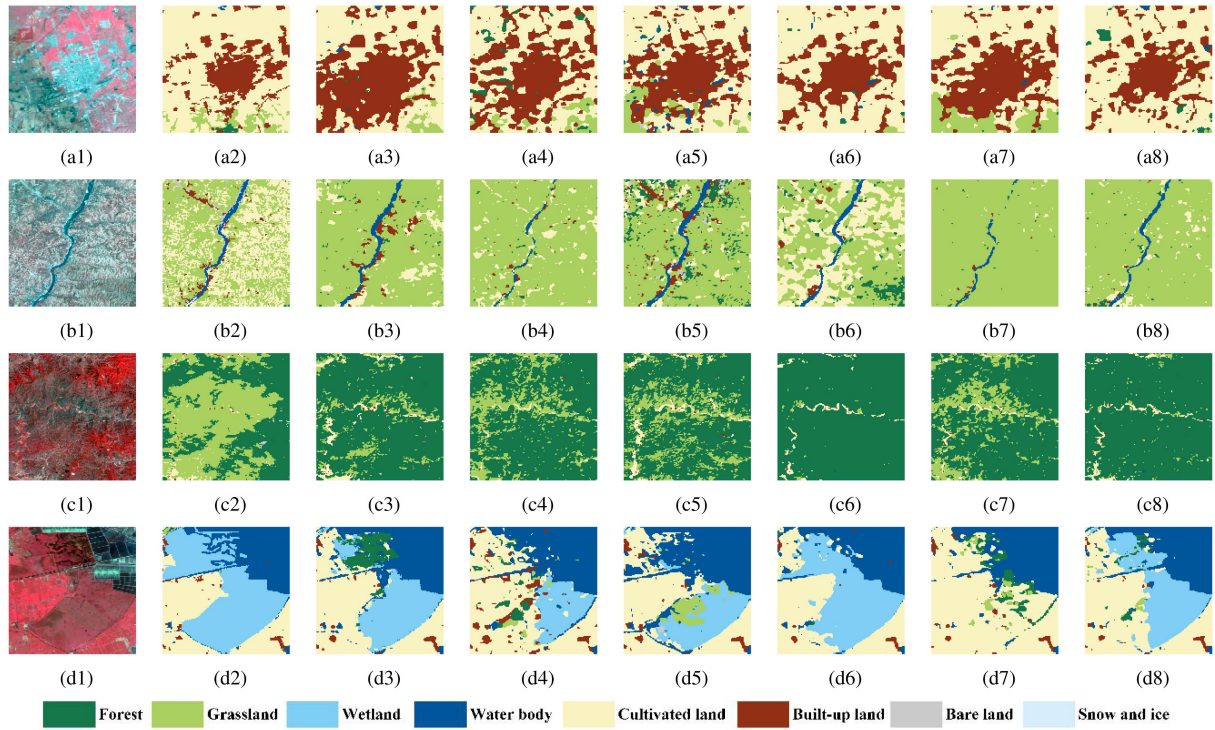


Fig. 3. Detailed classification results of the study area. (a1)–(d1) Original Landsat images. (a2)–(d2) Reference. (a3)–(d3) Classification results of FCN. (a4)–(d4) Classification results of PSPNet. (a5)–(d5) Classification results of DeepLab V3+. (a6)–(d6) Classification results of CEL-VICL. (a7)–(d7) Classification results of I2CS. (a8)–(d8) Classification results of the proposed FCNet.

TABLE I  
QUANTITATIVE EVALUATION OF SEGMENTATION RESULTS (%)

Networks (%)	Forest	Grass land	Wet land	Water body	Built-up land	Artificial surface	Bare land	Overall accuracy
FCN	81.31	51.84	18.05	72.13	80.08	61.08	<b>28.72</b>	73.77
PSPNet	76.21	66.06	9.19	54.00	68.42	<b>68.33</b>	26.89	69.25
DeepLab V3+	76.26	58.00	18.08	<b>73.89</b>	78.32	60.81	31.04	72.51
CEL-VTCL	<b>85.56</b>	39.96	<b>21.47</b>	69.68	80.02	57.78	7.51	72.67
I2CS	73.76	<b>70.93</b>	0.00	64.40	73.54	57.00	14.32	70.74
FCNet	83.96	51.81	21.20	69.90	<b>81.44</b>	57.53	28.50	<b>74.75</b>

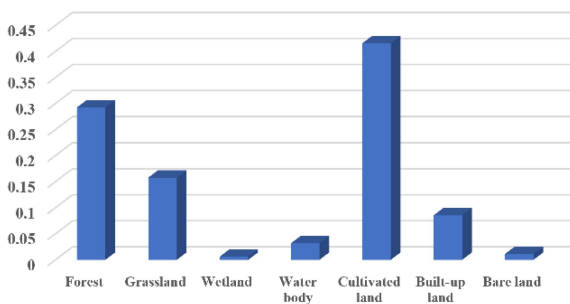


Fig. 4. Proportions of each class in the training set.

are listed in Table I, where the highest accuracies are shown in bold. In general, the accuracies of forests, grassland, and cultivated land are higher than those of the other classes. This coincidence with the proportions of each class in the training set is shown in Fig. 4. This means that networks tend to learn features of majority classes over minority ones to maintain global optimum. This is determined by the definition of loss function that controls the learning preference of the network.

Assigning different learning weights to different classes is a solution, but the effect is limited and the proportion of each class should be known as a prior. The proposed FCNet aims to improve the learning preference of minority classes and classes with noisy labels by compacting the features of each class. As can be seen in Table I, although the accuracy of each class of the proposed FCNet is not the highest, it is more balanced than those of other algorithms. So it can achieve the highest overall accuracy. This indicates that the effect of compacting the features of each class lies in balancing the learning preference of the network.

### C. Feature Consistency Constraints for Other Network Structures

To test the generality of the proposed feature consistency constraints, experiments on different network structures are conducted. In this experiment, four network structures, namely, FCN, DeepLab V3+, PSPNet, and FLANet [56], are employed to test the effectiveness of the proposed feature consistency constraints. Corresponding accuracies of the original network

TABLE II  
ACCURACIES OF DIFFERENT NETWORK STRUCTURES (%)

Networks (%)	Forest	Grass land	Wet land	Water body	Built-up land	Artificial surface	Bare land	Overall accuracy
FCN	81.31	51.84	18.05	72.13	80.08	61.08	28.72	73.77
FCFCN	80.85	49.91	16.80	70.14	82.11	59.71	28.47	74.13
DeepLabV3+	76.26	58.00	18.08	<b>73.89</b>	78.32	60.81	31.04	72.51
FCV3	75.04	59.08	17.67	63.41	<b>82.17</b>	48.12	23.20	72.70
PSPNet	76.21	<b>66.06</b>	9.19	54.00	68.42	<b>68.33</b>	26.89	69.25
FCNet	<b>83.96</b>	51.81	<b>21.20</b>	69.90	81.44	57.53	28.50	<b>74.75</b>
FLANet	75.61	66.49	21.00	66.96	65.17	58.79	29.13	67.16
FCFLANet	71.94	60.40	19.64	60.52	82.06	49.92	<b>31.90</b>	72.18

TABLE III  
ACCURACIES OF THE ABLATION STUDY (%)

Networks (%)	Forest	Grass land	Wet land	Water body	Built-up land	Artificial surface	Bare land	Overall accuracy
PSPNet	76.21	<b>66.06</b>	9.19	54.00	68.42	<b>68.33</b>	26.89	69.25
PSPNet- $\sigma$	83.38	51.10	<b>26.12</b>	<b>71.62</b>	79.44	64.52	<b>33.90</b>	74.29
PSPNet-dis	79.73	54.19	17.47	68.00	<b>82.24</b>	59.66	27.93	74.45
FCNet	<b>83.96</b>	51.81	21.20	69.90	81.44	57.53	28.50	<b>74.75</b>

structures and the ones with feature consistency constraints are listed in Table II, where the prefix “FC” represents the network structure with feature consistency constraints. Although the effect of feature consistency constraints is to improve the intraclass and inter-iteration compactness, it is clear from Table II that the feature consistency constraints cannot improve the accuracy of each class. The overall accuracies performed on the three network structures are all increased compared with the baselines. The improvement on DeepLab V3+ is the least among all the structures. This may be caused by the noisy labels contained in the training samples. DeepLab V3+ has used atrous convolution, atrous spatial pyramid pooling, Xception, and other modules to improve the learning ability of the network. With the effect of existing modules, the introduction of feature consistency constraints did not work as well as without these modules. On the contrary, the original PSPNet and FLANet are sensitive to noisy labels, and introducing feature consistency constraints improves their learning ability on consistent features. Therefore, the performance improvements of these networks are the most obvious. Although FLANet solved the attention missing problem, but these channel and spatial attention cannot alleviate the impact of noisy labels. Therefore, both accuracies of FLANet and FCFLANet are lower than those of PSPNet and FCNet. As for FCN, the original overall accuracy is 73.77%, with the consideration of feature consistency constraints, the overall accuracy of FCFCN is comparable with FCNet.

#### D. Ablation Study

Ablation experiments are carried out to prove the efficiency of the proposed FCNet with the backbone of PSPNet as an example. The accuracies are listed in Table III. It is clear that both the intraclass and inter-iteration feature consistency constraints make a significant improvement on the learning ability of PSPNet. The intraclass feature consistency constraint (PSPNet- $\sigma$ ) tends to improve the learning ability of PSPNet on classes with spectral consistency such as wetland, water body, and bare land. Actually, the classification accuracy of forests also improves from 76.21% to 83.38%. The inter-iteration feature consistency constraint tends to balance the learning ability of PSPNet. As shown in

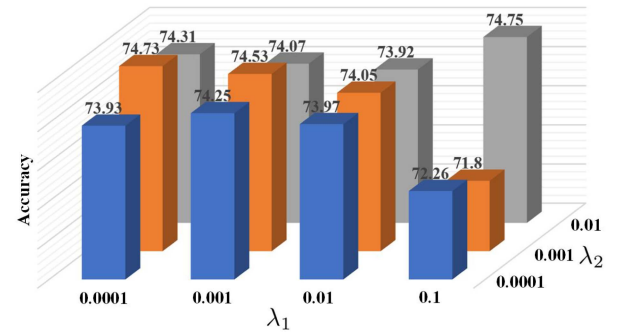


Fig. 5. Accuracy with the change of parameters.

Table III, the overall accuracy of PSPNet-dis is 74.45%, which is higher than both PSPNet and PSPNet- $\sigma$ , while only built-up land obtains the highest accuracy among the four networks. The proposed FCNet fully utilizes the advantages of the intraclass and inter-iteration feature consistency constraints and obtains the highest overall accuracy of 74.75%. From its class-wise accuracy, it is found that the FCNet tends to balance the learning preference of the network rather than improve the learning ability of a certain class.

#### E. Parameter Analysis

To validate the robustness of the coefficients, the changes of accuracies along with the changes of  $\lambda_1$  and  $\lambda_2$  are shown in Fig. 5. Actually, the loss function shown in (12) contains three parts, where the coefficient of cross-entropy is set to 1 in default. From Fig. 5, it is clear that the overall accuracies exceed 74% with the most combinations of the two coefficients. This means that the proposed FCNet is robust to changes in parameters in a relatively large range. This facilitates the utilization of the proposed feature consistency constraints.

#### F. Discussion

To illustrate the advantages of the proposed FCNet, the changes in average variance and average mean with iterations are shown in Fig. 6. It is clear that the average variance has



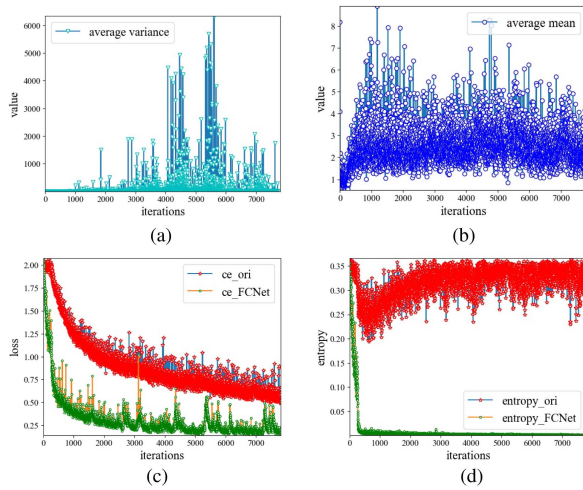


Fig. 6. Changes of the losses with iteration. (a) Average variance. (b) Average mean. (c) Cross-entropy. (d) Entropy.

a significant increase between the iterations 4000 and 6000, then it dramatically decreases. Compared with Fig. 6(c), which represents the decrease of cross-entropy with iterations, it can be inferred that the former 4000 iterations aim to learn general information from the training set with a notable decrease of the cross-entropy. Then, the average variance decreases along with the cross-entropy after 6000 iterations, indicating that it begins to influence the learning preference along with the cross-entropy. As shown in Fig. 6(b), the fluctuation of the average mean decreases with iteration. This means that the learned features of the network tend to be more stable.

To verify the effectiveness of the proposed feature consistency constraints in improving the learning ability of the networks, the changes of entropy along with the iteration are shown in Fig. 6(d). Without introducing the feature consistency constraints, the entropy decreases before the first 800 iterations and then increases to the former value. The entropy has not decreased significantly throughout the iteration, indicating that the certainty of the learned features remains the same. On the contrary, the entropy of the proposed FCNet decreases significantly in the former 400 iterations, and the variance of entropy further decreases in the following iterations. This means that the proposed FCNet can continuously improve the certainty of the learned features.

## V. CONCLUSION

The proposed FCNet introduces intraclass and inter-iteration feature consistency constraints to control the learning preference of the network. The proposed feature consistency constraints can improve the learning ability of different network structures as a plug-and-play operation. Experimental results show that the intraclass feature consistency constraint is expert in improving the network learning ability in classes with strong intraclass consistency, such as water bodies, and wetlands. However, the inter-iteration feature consistency constraint tends to balance the learning preference of the network. They are complementary and insensitive to the corresponding coefficients, which makes the

plug-and-play operation more convenient for different datasets. Further analysis of the loss and entropy verifies that the proposed feature consistency constraints can balance the learning preference of different network structures from the viewpoint of gradient descent.

## REFERENCES

- [1] J. Buchner et al., "Land-cover change in the Caucasus mountains since 1987 based on the topographic correction of multi-temporal Landsat composites," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111967.
- [2] W. Li, R. Dong, H. Fu, J. Wang, L. Yu, and P. Gong, "Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111563.
- [3] C. Jing, W. Zhou, Y. Qian, W. Yu, and Z. Zheng, "A novel approach for quantifying high-frequency urban land cover changes at the block level with scarce clear-sky Landsat observations," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112293.
- [4] F. Zhang and X. Yang, "Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection," *Remote Sens. Environ.*, vol. 251, 2020, Art. no. 112105.
- [5] A. Ghorbanian, M. Kakooei, M. Amani, S. Mahdavi, A. Mohammadzadeh, and M. Hasanlou, "Improved land cover map of Iran using sentinel imagery within Google Earth engine and a novel automatic workflow for land cover classification using migrated training samples," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 276–288, 2020.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [7] A. Saxe, S. Nelli, and C. Summerfield, "If deep learning is the answer, what is the question?," *Nature Rev. Neurosci.*, vol. 22, no. 1, pp. 55–67, 2021.
- [8] X.-Y. Tong et al., "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322.
- [9] S. Pan et al., "Land-cover classification of multispectral Lidar data using CNN with optimized hyper-parameters," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 241–254, 2020.
- [10] P. Gopal Singh, N. Bordu, D. Singh, H. Yahia, and K. Daoudi, "Per-mutated spectral and permuted spectral-spatial CNN models for PolSAR-multispectral data based land cover classification," *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 1096–1120, Feb. 2021.
- [11] 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, 2019.
- [12] Z. Chen, W. Fan, B. Zhong, J. Li, J. Du, and C. Wang, "Coarse-to-fine road extraction based on local Dirichlet mixture models and multiscale-high-order deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4283–4293, Oct. 2020.
- [13] Y. R. Choi and R. M. Kil, "Face video retrieval based on the deep CNN with RBF loss," *IEEE Trans. Image Process.*, vol. 30, pp. 1015–1029, 2021.
- [14] Z. Chen, C. Wang, J. Li, W. Fan, J. Du, and B. Zhong, "AdaBoost-like end-to-end multiple lightweight U-Nets for road extraction from optical remote sensing images," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 100, 2021, Art. no. 102341.
- [15] X. Luo, J. Li, M. Chen, X. Yang, and X. Li, "Ophthalmic disease detection via deep learning with a novel mixture loss function," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3332–3339, Sep. 2021.
- [16] Z. Chen et al., "Road extraction in remote sensing data: A survey," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 112, 2022, Art. no. 102833.
- [17] Y. Liu and H. Guo, "Peer loss functions: Learning from noisy labels without knowing noise rates," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 6226–6236.
- [18] X. Ma, H. Huang, Y. Wang, S. Romano, S. Erfani, and J. Bailey, "Normalized loss functions for deep learning with noisy labels," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 6543–6553.
- [19] S. Borse, Y. Wang, Y. Zhang, and F. Porikli, "InverseForm: A loss function for structured boundary-aware segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 5901–5911.
- [20] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

- [22] X. Zhao et al., "Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.
- [23] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13521–13531, Nov. 2020.
- [24] R. Shang, J. He, J. Wang, K. Xu, L. Jiao, and R. Stolkin, "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," *Knowl.-Based Syst.*, vol. 194, 2020, Art. no. 105542.
- [25] Z. Huang, J. Wang, X. Fu, T. Yu, Y. Guo, and R. Wang, "DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection," *Inf. Sci.*, vol. 522, pp. 241–258, 2020.
- [26] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.
- [27] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
- [28] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [29] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 603–612.
- [30] P.-T. Jiang, L.-H. Han, Q. Hou, M.-M. Cheng, and Y. Wei, "Online attention accumulation for weakly supervised semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7062–7077, Oct. 2022.
- [31] J. Xie, K. Hu, Y. Guo, Q. Zhu, and J. Yu, "On loss functions and CNNs for improved bioacoustic signal classification," *Ecological Informat.*, vol. 64, 2021, Art. no. 101331.
- [32] H. Shi, L. Wang, N. Zheng, G. Hua, and W. Tang, "Loss functions for pose guided person image generation," *Pattern Recognit.*, vol. 122, 2022, Art. no. 108351.
- [33] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, 2005.
- [34] G. Pezzano, V. Ribas Ripoll, and P. Radeva, "CoLe-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation," *Comput. Methods Programs Biomed.*, vol. 198, 2021, Art. no. 105792.
- [35] X. Zhou, T. Tang, Y. Cui, L. Zhang, and G. Kuang, "Novel loss function in CNN for small sample target recognition in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4018305.
- [36] H.-H. Yang, C.-H. H. Yang, and Y.-C. James Tsai, "Y-Net: Multi-scale feature aggregation network with wavelet structure similarity loss function for single image dehazing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 2628–2632.
- [37] A. F. R. Guarda, N. M. M. Rodrigues, and F. Pereira, "Neighborhood adaptive loss function for deep learning-based point cloud coding with implicit and explicit quantization," *IEEE MultiMedia*, vol. 28, no. 3, pp. 107–116, Jul.-Sep. 2021.
- [38] F. Liu, G. Lin, and C. Shen, "Discriminative training of deep fully connected continuous CRFs with task-specific loss," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2127–2136, May 2017.
- [39] Y. Qin, C. Yan, G. Liu, Z. Li, and C. Jiang, "Pairwise Gaussian loss for convolutional neural networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6324–6333, Oct. 2020.
- [40] C. Liang, H. Zhang, D. Yuan, and M. Zhang, "A novel CNN training framework: Loss transferring," *IEEE Trans. Circuits Syst. Video. Technol.*, vol. 30, no. 12, pp. 4611–4625, Dec. 2020.
- [41] X. Zhang, H. Yao, Z. Lv, and D. Miao, "Class-specific information measures and attribute reducts for hierarchy and systematicness," *Inf. Sci.*, vol. 563, pp. 196–225, 2021.
- [42] X. Zhang, J. Yang, and L. Tang, "Three-way class-specific attribute reducts from the information viewpoint," *Inf. Sci.*, vol. 507, pp. 840–872, 2020.
- [43] B.-Q. Yang, X.-P. Guan, J.-W. Zhu, C.-C. Gu, K.-J. Wu, and J.-J. Xu, "SVMs multi-class loss feedback based discriminative dictionary learning for image classification," *Pattern Recognit.*, vol. 112, 2021, Art. no. 107690.
- [44] X. Zhu, X.-Y. Jing, X. You, X. Zhang, and T. Zhang, "Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5683–5695, Nov. 2018.
- [45] Q. Leng, H. Yang, J. Jiang, and Q. Tian, "Adaptive multiscale segmentations for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5847–5860, Aug. 2020.
- [46] X. Si, Q. Yin, X. Zhao, and L. Yao, "Consistent and diverse multi-view subspace clustering with structure constraint," *Pattern Recognit.*, vol. 121, 2022, Art. no. 108196.
- [47] G. Zhang, J. Yang, Y. Zheng, Z. Luo, and J. Zhang, "Optimal discriminative feature and dictionary learning for image set classification," *Inf. Sci.*, vol. 547, pp. 498–513, 2021.
- [48] D. Liu, L. Liu, Y. Tie, and L. Qi, "Multi-task image set classification via joint representation with class-level sparsity and intra-task low-rankness," *Pattern Recognit. Lett.*, vol. 132, pp. 99–105, 2020.
- [49] Y. Rong, S. Xiong, and Y. Gao, "Double graph regularized double dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 29, pp. 7707–7721, 2020.
- [50] J. Fan, Z. Zhang, C. Song, and T. Tan, "Learning integral objects with intra-class discriminator for weakly-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4282–4291.
- [51] H. Alhuzali and S. Ananiadou, "Improving textual emotion recognition based on intra- and inter-class variation," *IEEE Trans. Affect. Comput.*, to be published, doi: [10.1109/TAFFC.2021.3104720](https://doi.org/10.1109/TAFFC.2021.3104720).
- [52] H. Peng and S. Yu, "Beyond softmax loss: Intra-concentration and inter-separability loss for classification," *Neurocomputing*, vol. 438, pp. 155–164, 2021.
- [53] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [54] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.
- [55] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 833–851.
- [56] Q. Song, J. Li, C. Li, H. Guo, and R. Huang, "Fully attentional network for semantic segmentation," in *Proc. 36th AAAI Conf. Artif. Intell.*, 2022, pp. 2280–2288.

**Xuemei Zhao** received the Ph.D. degree in photogrammetry and remote sensing from Liaoning Technical University, Fuxin, China, in 2017.

She is currently an Associate Professor with Guilin University of Electronic Technology, Guilin, China. Her research interests include deep learning, information geometry, and their application in image processing.

**Luo Liang** is currently working toward the master's degree in electronic science and technology with the Guilin University of Electronic Technology, Guilin, China, under the supervision of Xuemei Zhao.

His research interests include deep learning-based semantic segmentation and corresponding domain adaption.

**Jun Wu** received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2003.

He is currently a Professor with the Guilin University of Electronic Technology, Guilin, China. His research interests include computer vision and photoelectric information processing.

**Haijian Wang** received the Ph.D. degree in mechanical engineering from Liaoning Technical University, Fuxin, China, in 2017.

He is currently an Associate Professor with the Guilin University of Electronic Technology, Guilin, China. His research interests include machine vision and automation.

**Xingyu Gao** received the Ph.D. degree in testing and measurement technology and instruments from Tianjin University, Tianjin, China, in 2010.

He is currently a Professor with the Guilin University of Electronic Technology, Guilin, China. His research interests include machine vision detection technology and collaborative robot.