An Improved Land Use Classification Method Based on DeepLab V3+ Under GauGAN Data Enhancement

Kang Zheng, *Student Member, IEEE*, Haiying Wang, Feng Qin, Changhong Miao, and Zhigang Han

***Abstract*—Land use is a reflection of human activities in surface space, and classifying it helps better understand the relationship between human activities and the spatial environment. However, the imbalance in land use datasets acquired through remote sensing images has become a major obstacle to improving the accuracy of land use classification. To maintain the balance in the samples of the land use dataset and improve the accuracy of land use classification, this article proposes an improved model based on the DeepLab V3+ network under the GauGAN data enhancement strategy. First, regarding the data imbalance problem, this article proposes an attention optimization mechanism to enhance the learning ability of the generator of GauGAN for contextual semantic information, and adds spectral normalization to the discriminator to induce stable model training. Thus, the model can synthesize excellent small-sample feature data. For the land use classification model, this article improves the DeepLab V3+ network by modifying the expansion rate of the ASPP module and adding the proposed feature fusion module to enhance the ability of the model for combining high- and low-level semantic features. Finally, this article implements both of the proposed improvements to achieve high-precision land use classification. The results showed the following. The land use data synthesized by improved GauGAN contained more complex semantic information and detailed features than the synthetic results of other models, and thus better represented the features of land use. The improved DeepLab V3+ model outperformed the U-Net, FPN, DeepLab V3+, MANet, and TransUNet.**

***Index Terms*—Convolutional neural network, data enhancement, deep learning, generative adversarial network, land use classification.**

Manuscript received 5 October 2022; revised 8 March 2023; accepted 18 May 2023. Date of publication 22 May 2023; date of current version 26 June 2023. This work was supported in part by the Young Key Teacher Training Plan of Henan under Grant 2020GGJS028, in part by the Natural Resources Science and Technology Innovation Project of Henan Province under Grant 202016511, in part by the Key Scientific Research Project Plans of Higher Education Institutions of Henan under Grant 21A170008, in part by the National Major Project of High-Resolution Earth Observation System under Grant 80-Y50G19-9001-22/23, and in part by the Technology Development Plan Project of Kaifeng under Grant 2003009. *(Corresponding author: Haiying Wang.)*

The authors are with the College of Geography and Environment Science, Henan University, Kaifeng 475004, China, and with the Key Laboratory of Geospatial Technology for Middle and Lower Yellow River Regions, Ministry of Education, Kaifeng 475004, China, and with the Henan Technology Innovation Center of Spatio-Temporal Big Data, Henan University, Zhengzhou 450046, China, and with the Henan Industrial Technology Academy of Spatio-Temporal Big Data, Henan University, Zhengzhou 450046, China, and also with the Key Research Institute of Yellow River Civilization and Sustainable Development, Henan University, Kaifeng 475004, China (e-mail: zkang@henu.edu.cn; whyhdgis@henu.edu.cn; qinfen@henu.edu.cn; chhmiao@henu.edu.cn; zghan@henu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3278862

I. Introduction

LAND use is the high-level expression of human activities on land [1], and is the basis for understanding the dynamic changes and socio-ecological linkages on the Earth's surface [2]. Land use classification is an important means of obtaining information on land use, and plays a key role in urban and regional management, government decision-making, and monitoring the activities of the population [3]. With continual developments in remote sensing technology, the spatial and temporal resolutions of remote sensing images have improved and they contain rich spectral information. Data on remote sensing images can be obtained in various ways, and provide a basis for highly precise land use classification [4]. However, due to the complexity and diversity of features of land use in remote sensing images, land use classification is often significantly challenging.

Traditional methods of land use classification can be classified into pixel-based and object-oriented approaches, depending on the scale of the processing unit. Pixel cell-based methods classify individual pixels in images based on their spectral reflectance without considering the relationships among adjacent pixels [5]. The accuracy of classification of these methods is often limited due to the noise and interclass variation in remotely sensed images. To overcome the shortcomings of pixel cell-based methods, some scholars have proposed postprocessing for optimization [6]. However, postprocessing tends to ignore small-scale characteristics of the features in images, such as patches of grass and houses in small areas. Object-oriented methods of land use classification are based on segmented objects as the basic unit, and can reduce the amount of unnecessary detail while introducing contextual information on the objects. These methods include the application of co-occurrence, neighborhood-graph-based and geometric measure [7], [8], [9], [10]. The greatest shortcoming of these methods is that the size of the segmented objects is unsuitable for capturing different features. If the segmentation scale is too large, other small-scale features are easily ignored. If it is too small, it will reduce the computational efficiency.

As generally known, many land use products were obtained by using traditional methods like random forest [11], [12]. Land use classification cannot satisfy the demands of various applications by relying only on the above-mentioned traditional methods of classification in the current era of big data. To respond to this challenge, deep learning technology has emerged as a reliable solution for the efficient extraction of features of land use from remote sensing images [13]. Owing to rapid technological development, methods of land use classification based on deep learning are widely used [14]. Helber et al. [15] proposed a patch-based land use classification method based on Sentinel-2, which can be used to improve the mapping capability. Weng et al. [16] combined convolutional neural networks and extreme learning machine to achieve high accuracy land use classification. Xiong et al. [17] combined a full convolutional network with a generative adversarial network to form a Bayesian semantic segmentation network for land use classification. However, because none of these methods of classification considers the problem of data imbalance, they are often unable to accurately classify small features. Therefore, solving the problem of data imbalance is crucial for improving the accuracy of land use classification.

The problem of data imbalance is often solved by modifying the loss function, and through image enhancement. The typical loss functions used in this context include weighted cross-entropy, focal loss, and dice loss [18], [19], [20]. A drawback of the loss function is that the process of tuning it is complicated and it is difficult to determine the optimal hyperparameters for it. The commonly used methods of image enhancement include the affine transform, information removal, image fusion, and the generative adversarial network (GAN) [21], [22], [23], [24]. Affine transformation, information deletion, and image fusion are often used as general methods of image enhancement. A drawback of these methods is that they tend to lead to overfitting, which is mainly caused by constantly learning the same data.

The GAN can overcome the shortcomings of the above-mentioned methods, and increases the diversity and richness of the data on land use by synthesizing new features through the deep learning of their characteristics. The GAN is a generative model that was proposed by Goodfellow et al. in 2014 [21]. It is a popular area of research in Artificial Intelligence. The basic idea underlying it is derived from the two-person zero-sum game in game theory that involves a generator and a discriminator, both of which are trained by adversarial learning [25], [26]. The GAN is widely used for image enhancement, obtaining a super-resolution, and converting the image style, and is applied to estimate the potential distribution of the data to generate new samples. Among them, generative adversarial networks for data generation include conditional and unconditional methods. Regarding the unconditional data generation methods, they include StyleGAN2 [27], BigGAN [28], and ReACGAN [29]. However, such methods lack the consideration of labels and increase the cost of expensive data annotation.

In the context of image enhancement to conditional generate new data, Cui et al. [30] used noise and conditional information as inputs to generate data while controlling the results by means of the conditional GAN (CGAN). Isola et al. [31] proposed pix2pix based on the CGAN. It can learn the mapping relationship between input and output images to generate new data based on the graph of edges of the images. Wang et al. [32] improved pix2pix to propose pix2pixHD, which can generate high-resolution images. Park et al. [33] claimed that the commonly used normalization layers tend to lose semantic information in images, and proposed the GauGAN network, which contains new normalization layers, to solve this problem. All these methods can synthesize new data on land use. However, remote sensing data have a high resolution and rich semantic information. Therefore, a suitable GAN network needs to be selected and improved in order to synthesize data based on a small number of samples, which is the key to solve this problem.

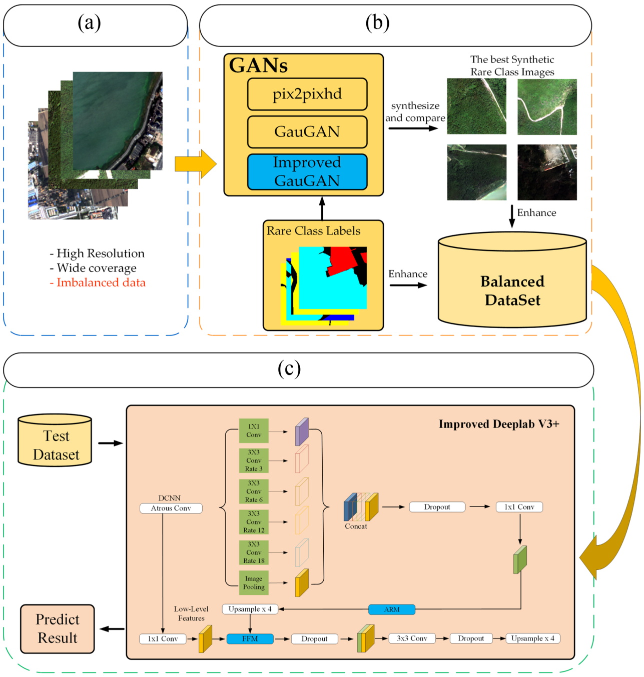


Fig. 1. Flow of the proposed method of land use classification. GID preprocessing, data enhancement, and classification experiments on the improved method based on the GauGAN and DeepLab V3+ network. (a) GID DataSet. (b) Data Enhancement. (c) Landuse Classification.

In summary, this article investigates land use classification in terms of dataset and classification method, respectively, and proposes an improved method of land use classification by using DeepLab V3+ with data enhancement based on the GauGAN. In the dataset problem, the generator and discriminator of GauGAN are improved to synthesize new feature data with few samples and solve the sample imbalance problem. Based on this, the DeepLab V3+ method is improved by designing a feature fusion module to effectively enhance the learning ability of the model for both high- and low-level features, thus enhancing the land use classification accuracy.

II. Research Materials and Methods

The process of land use classification studied here is illustrated in the following. The Gaofen image dataset (GID) [34] is first input to pix2pixHD, GauGAN, and improved GauGAN for training. New sample data are generated by using a small sample of ground class labels, and the results of the synthetic data generated by the three GAN networks are compared. The better result is selected to expand the entire GID dataset to form a dataset with a balanced number of samples from each class. Finally, this dataset is fed to the improved DeepLab V3+ network for land use classification (see Fig. 1).

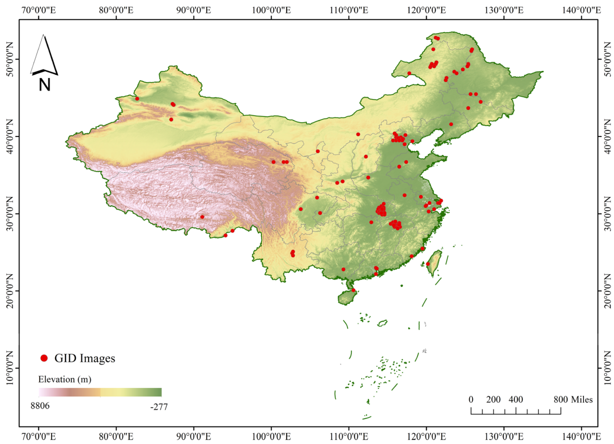


Fig. 2. Map of distribution of the GID. The map is a visualization of the geospatial distribution of GID data, which cover the major cities and types of landforms in China.

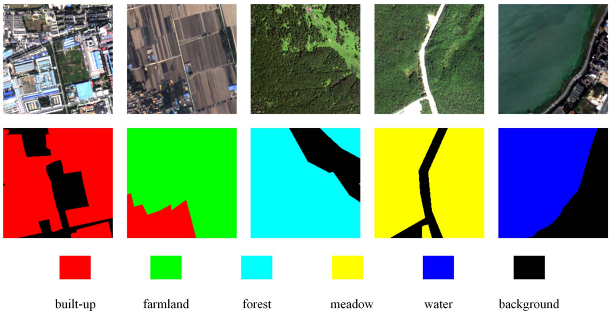


Fig. 3. GID dataset. The dataset contained five types of features: Built-up land, farmland, forest, meadow, and water. Their RGB compositions were (255, 0, 0), (0, 255, 0), (0, 255, 255), (255, 255, 0), (255, 255, 0), and (0, 0, 255).

A. Research Materials

The GID dataset (see Fig. 2) is a high-resolution land use dataset based on remote sensing images from the Gaofen-2 satellite. The dataset was collected from December 2014 to October 2016 in more than 60 cities in China. It has a geographical range of more than 50 000 km^2, and contains more than 150 remote sensing images, each with a size of 6800 × 7200 pixels, and a spatial resolution of 1 m. The images are in the RGB and NIR+RGB formats.

We used a small part of the GID dataset to demonstrate the validity of the proposed method. Images in the dataset were cropped to a size of 224 × 224 (see Fig. 3), with a total of 1391 pictures. The categories of land use considered here were based on the Chinese Land Use Classification Criteria (GB/T21010-2017) as a reference. Five categories of features were considered: built-up land, farmland, meadow, water, and forest. The pixel percentages of each category in the dataset are shown in Fig. 4
.

B. GAN Data Enhancement Methods

pix2pixHD (see Fig. 5) was proposed based on the CGAN framework. It is mainly used for image-to-image translation, and can generate images with a high resolution. It includes generators and discriminators. The generator consists of a residual network that maps the labels on the acquired image and uses them to synthesize a new image. The discriminator has multiple scales, and is used to distinguish between real and synthesized images. The loss function consists of GAN loss, feature matching loss, and content loss.

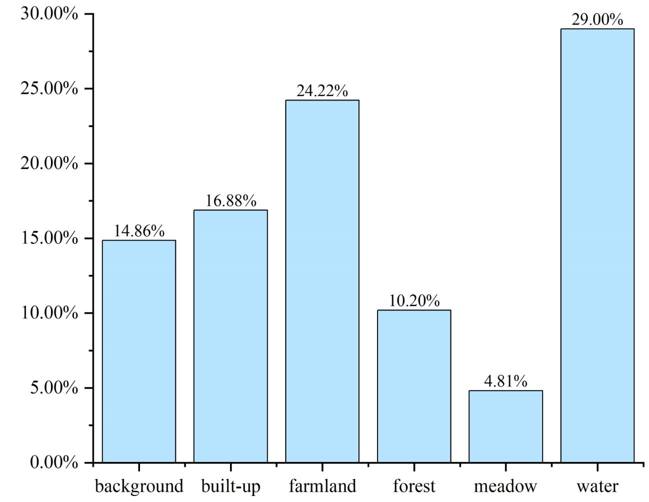


Fig. 4. Share of types of land in the sample of the GID dataset. The problem of data imbalance is clear, with fewer features for meadows and forests than for the other categories of land use.

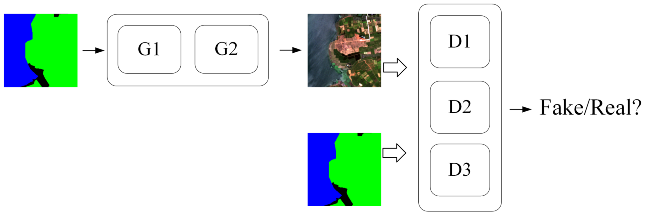


Fig. 5. pix2pixHD.

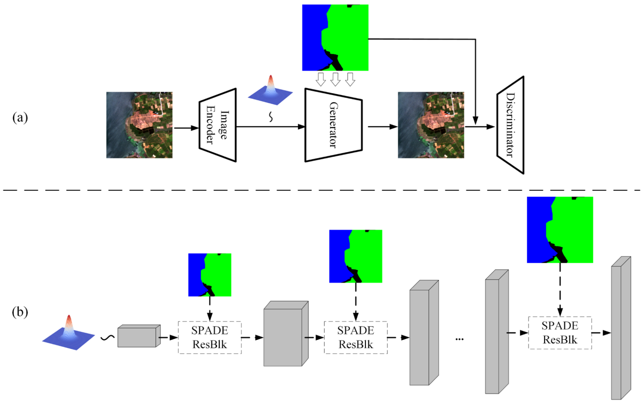


Fig. 6. Training process of (a) GauGAN and (b) SPADE generator.

GauGAN is a modified GAN on top of the pix2pixHD network that is used to generate high-fidelity images based on semantic labels. The network consists of three parts: a generator, a discriminator, and an image encoder as shown in Fig. 6. The generator is a fully convolutional encoder consisting of the SPADE ResBlk (residual block, see Fig. 7), and the input to it is a feature map obtained after normalization. Upsampling

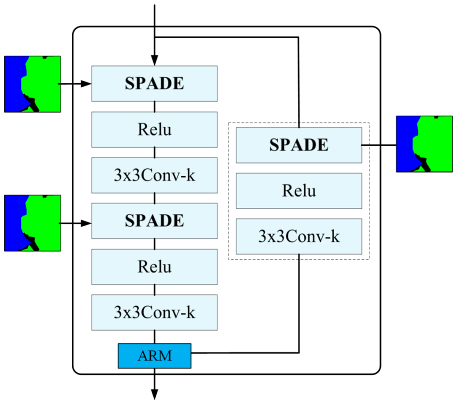


Fig. 7. SPADE ResBlk.

is subsequently performed through computation in the SPADE ResBlk. Finally, training is completed by continuously using the annotated data to enhance the semantic information during image generation.

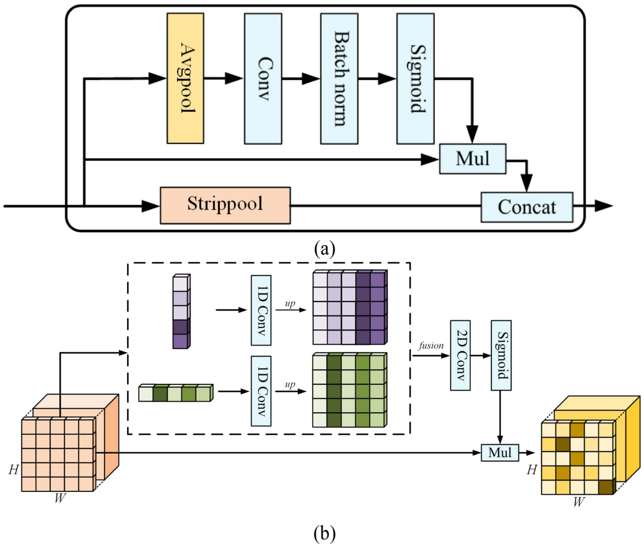
SPADE normalization is the main difference between GauGAN and pix2pixHD. The latter uses an unconditional method of normalization that leads to the loss of semantic information. SPADE achieves better normalization by modifying the parameters of batch normalization to all pixels of the feature map, instead of setting the parameters only for channels of the feature map.

C. Improvements to GauGAN

The SPADE method of GauGAN effectively improves the network's ability to stabilize synthetic images, but the ability of the generator to learn semantic information of images and the discriminator both need to be improved. Therefore, in this article, we propose ARM (Attention Refinement Module) and incorporate it in the SPADE ResBlk of the generator to optimize the ability of the generator to learn semantic features.

ARM is mainly used to guide feature learning with the help of global average pooling and strip pooling, which enables the GauGAN generator to synthesize excellent small-sample features, as shown in Fig. 8 . First, the feature maps are pooled globally on average, and then the output is point multiplied with the original feature maps by convolution, normalization, and Sigmoid calculation. Finally, the output feature map is concatenated with the result of stripe pooling. Global average pooling is to calculate an average value for each channel of the input features, which is beneficial to integrate global spatial information and improve model robustness. Stripe pooling is divided into vertical and horizontal stripe pooling, which aims to increase the long-range contextual information learning capability of the generator in these two directions.

The discriminator of GauGAN is mainly designed with pix2pixHD network. However, the discriminator is based on Patch-GAN, which has difficulty in recognizing high-resolution images with complex features. Therefore, this article proposes to incorporate Spectral Normalization (SN) into the discriminator (see Fig. 9), which can improve the ability of the discriminator to capture image features as well as induce stable training of the model.



(a)

(b)

Fig. 8. Structure of (a) ARM and (b) strip pooling.

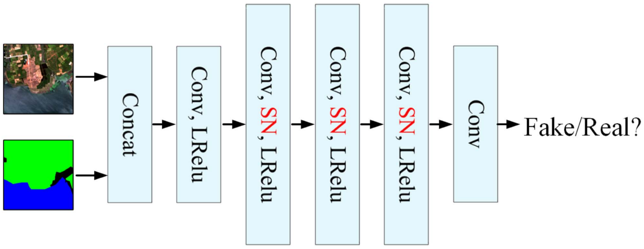


Fig. 9. Discriminator for adding SN.

D. Improvements to DeepLab V3+

The main structure of DeepLab V3+ is shown in the following. It contains two parts: an encoder and a decoder. The encoder is a deep convolutional network with dilated convolution, followed by a spatial pyramid pooling layer with dilated convolution that can learn multiscale information. In contrast to the previous generation of the DeepLab network, it contains a simple and effective decoder module to fuse low-level features with high-level features to improve the accuracy of edge segmentation. We make targeted modifications to the void rate in the atrous spatial pyramid pooling (ASPP) [35] module, fuse the FFM (Feature Fusion Module) and ARM with the network to improve the generalization capability of the model. The overall structure of the network is shown in the following.

ASPP is a method of feature extraction (Fig. 11) that can extract dense features by learning feature maps at different scales. The convolution method uses the hole convolution to enhance the network's ability to extract dense features. The principle is to use different rates of expansion of the hole convolution and pooling layers through multiple parallel connections, and then fuse them to extract multiscale information.

The higher is the rate of expansion of the dilated convolution in the ASPP module, the larger is the feature size, and vice

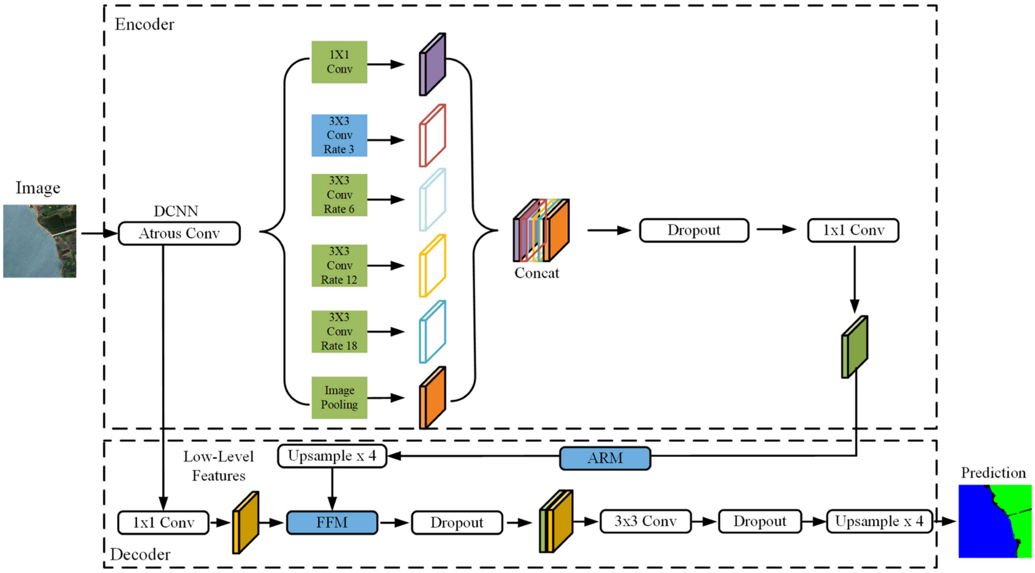


Fig. 10. Improvements to the DeepLab V3+ network.

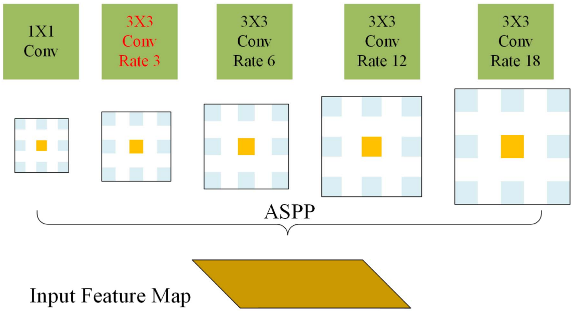


Fig. 11. Improvements to the ASPP module.

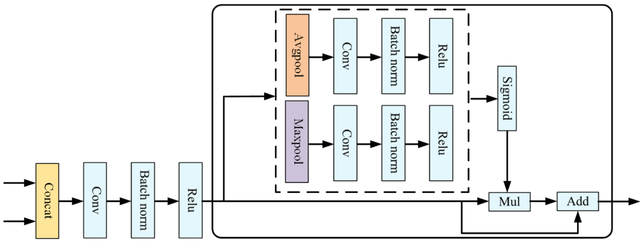


Fig. 12. Structure of FFM.

versa. The rates of expansion of the original ASPP module are six, 12, and 18. To improve the ability of the model to learn small features of land use, we add a cavity convolution with a rate of expansion of three to the ASPP module. A new feature map is thus obtained by dimension reduction in the dropout layer to obtain a 1 × 1 convolution layer. Finally, we also added the ARM module to optimize the output characteristics of the model at that stage before upsampling.

DeepLab V3+ simply sums high- and low-level features in the decoder. However, the two features are not the same and it is not possible to simply sum these. Therefore, we propose the FFM module to enhance the ability of the model for feature selection and fusion. Therefore, we propose the FFM module to enhance the ability of the model for feature selection and fusion. First, two feature maps are combined, and after convolution operation, they are subjected to average pooling and maximum pooling methods, respectively. The two are transformed into weight vectors by Sigmoid function. Last, the final output is obtained by dot product and summation.

E. Accuracy Evaluation

We assessed the accuracy of the model along two dimensions: qualitative and quantitative. The qualitative aspects included the comparison of the new data samples generated by pix2pixHD, GauGAN, and improved GauGAN in terms of the characteristics and clarity of the features. In addition, the synthesis results were evaluated for quality using SSIM (structural similarity index), PSNR (peak signal-to-noise ratio), and LPIPS (learned perceptual image patch similarity) [36], [37], [38].

SSIM is a measure of image similarity in terms of luminance, contrast, and structure, respectively. A larger SSIM value indicates better image quality, based on the following formula:

$$SSIM\,(x, y) \;=\; \frac{(2\mu_x\mu_y + c_1)\,(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\,\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \qquad (1)$$

where c_1 and c_2 are two constants to avoid the occurrence of 0, μ_x is the mean of x, μ_y is the mean of y, μ_x^2 is the variance of x, μ_y^2 is the variance of y, and σ_{xy} is the covariance of x and y.

The PSNR is calculated based on the mean square error between the corresponding pixel points and is a widely used metric. The larger the PSNR value, the less distortion in the image. The specific formula is as follows:

$$PSNR \;=\; 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \qquad (2)$$

where MAX_I^2 is the maximum possible pixel value in the image, MSE is the mean square error of the two images.

LPIPS is a metric for generating results based on learned perceptual image patch similarity. This metric differs from the above-mentioned two metrics in that it is more in line with human perception. The smaller the LPIPS value is, the better the image quality is. The specific formula is as follows:

$$d(x,x_0)=\sum_l \frac{1}{H_lW_l}\sum_{h,w}\left\|w_l\odot\left(\hat{y}^l_{hw}-\hat{y}^l_{0hw}\right)\right\|_2^2 \qquad (3)$$

$$\mathcal{L}PIPS\ (x,x_0,x_1,h) = -h\log\mathcal{G}\left(d\left(x,x_0\right),d\left(x,x_1\right)\right) - (1-h)\,log\left(1-\mathcal{G}\left(d\left(x,x_0\right),d\left(x,x_1\right)\right)\right) \qquad (4)$$

where x and y are from the real image, x_0 and y_0 are from the synthetic image, d is the distance between the real image and the synthetic image, and h is the mapping score of the top training network.

The results of land use classification were evaluated based on subjective experience, such as whether the features were continuous and whether the edge contours were clear. Quantitatively, the results of land use classification were analyzed by using recall, precision, the F1-score, overall accuracy (OA), and the MIoU. True positive (TP) is a result where the model correctly predicts the positive category. False negative (FN) is a result where the model incorrectly predicts the negative category. False positive (FP) is a result where the model incorrectly predicts the positive category. And true negative is a result where the model correctly predicts the negative category.

Recall is the ratio of the number of correctly classified positive category to the total number of positive samples

$$Recall\ =\frac{TP}{TP+FN}. \qquad (5)$$

Precision is the ratio of the number of correctly classified positive category to the number of all positive category in the results of classification

$$Precision\ =\frac{TP}{TP+FP}. \qquad (6)$$

The F1-score is based on recall and precision, and is an overall evaluation of these metrics

$$F1score\ =\ 2\cdot\frac{Recall\cdot Precision}{Recall+Precision}. \qquad (7)$$

The OA is the ratio of the number of correctly classified samples to the total number of samples

$$OA\ =\frac{TP+TN}{TP+FN+FP+TN}. \qquad (8)$$

MIoU is the ratio of the intersection and concatenation of the true and the predicted values, and is a global evaluation of the results of semantic segmentation

$$MIoU\ =\frac{1}{k+1}\sum_{i\,=\,0}^{k}\frac{TP}{FN+FP+TP}. \qquad (9)$$

III. Experiments and Results

The environment for the GAN-based data enhancement consisted of the PyTorch platform and that of DeepLab V3+ was based on the TensorFlow platform. The CPU used was an Intel(R) Xeon(R) Gold 5218, the graphics card was NVIDIA GeForceRTX 2080TI, and the programming language was Python3.6.

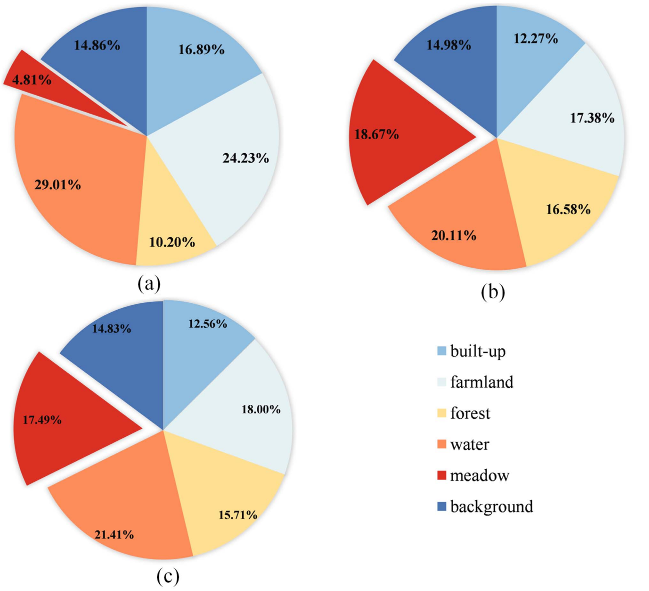


Fig. 13. Proportional distribution of features in the dataset after enhancement by different methods. (a) Original GID dataset. (b) Dataset obtained by augmenting the GID with translation, rotation and scaling. (c) Dataset obtained by augmenting the GID with GAN.

A. Experiment Datasets

Fig. 13 shows that the number of features representing meadows in the GID was small, accounting for only 4.81% of the total. Features representing water accounted for 29.01%, and the number of features representing forests was also relatively small. We performed normal image enhancement and GAN-based image enhancement using the GID. The number of samples added was kept approximately equal (see Fig. 10) to balance the data among the different features. The normal methods of image enhancement included rotation, translation, and scaling. The GAN-based methods of image enhancement involved comparing the results of pix2pixHD and GauGAN, and then choosing images that were more vivid and realistic.

B. Comparison of Results of GAN-Based Image Generation

Both pix2pixHD and GauGAN used the Adam optimizer on the PyTorch platform. The GID was used as the training set. We used feature maps with inputs and outputs of size 224 × 224, a learning rate of 0.0002, and 200 training epochs.

The overall results of the images generated by pix2pixHD, GauGAN, and our model are shown earlier (see Fig. 14). The quality of the images generated by ours was superior to those generated by pix2pixHD and GauGAN. The pix2pixHD not only lacked fine-grained textural features, but also exhibited a prominent phenomenon of collapse. In the first and second sets of the generated images, only meadows were distinguished from the features representing other forms of land use, and the features were blurred. In the third and fourth sets of the

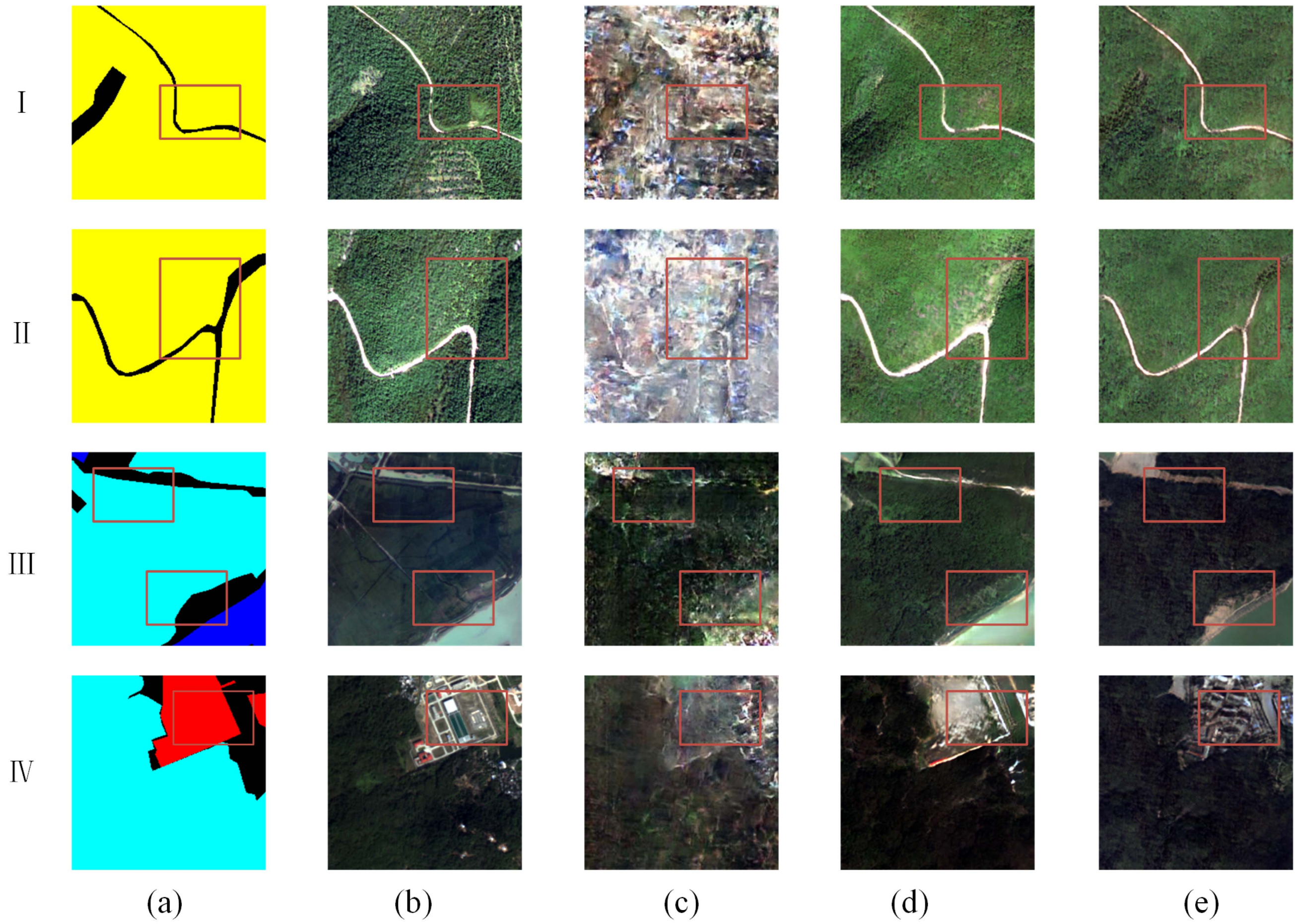


Fig. 14. Results of GAN-based generation. (a) labels. (b) images. (c) pix2pixhd. (d) GauGan. (e) Ours.

generated images, the phenomenon of collapse was mitigated. But when multiple features were synthesized at the same time, pix2pixHD did not perform well—for example, the synthesized images representing water and built-up land were not as good as expected.

Compared to images synthesized using GauGAN, our method generates images with more intricate and rich feature details while maintaining high image quality. For example, in the second set of synthesized images, the road in the top-left corner is more consistent with the corresponding land use feature distribution in the label. In the third set of synthesized images, the transition between the forest and background boundary appears more natural. In the fourth set of synthesized images, the building synthesis result contains more details, enabling the identification of more houses.

According to Table I, it is evident that our proposed method exhibits outstanding performance in various image quality metrics and significantly outperforms other methods. In particular, our method outperforms other methods by a large margin in terms of SSIM and LPIPS metrics, indicating that the synthesized results of our method have a certain similarity with the original data in terms of structure and are more consistent with human perception. Regarding the PSNR metric, there is not a substantial difference between our proposed method and GauGAN. Nonetheless, our method achieves a value much closer to 30, which suggests that the distortion loss of image quality in our synthesized images falls within an acceptable range.

TABLE I
Quality Evaluation of the Synthesis Results

Models	SSIM	PSNR	LPIPS
pix2pixHD	0.071	7.090	0.649
GauGAN	0.535	24.484	0.449
Ours	0.734	29.825	0.116

Therefore, our improved GauGAN method is better suited for synthesizing land use features and can effectively address the issue of imbalanced samples.

C. Evaluation of Results of Improved GauGAN

In this study, we utilized the improved GauGAN method to perform data augmentation on the GID dataset, resulting in an augmented dataset named "Ours-GID." To demonstrate the effectiveness of our method for land use classification, we compared the classification accuracy of our method with GauGAN-GID, GID, and NIE-GID (normal image enhance-GID).

TABLE II
RESULTS OF CLASSIFICATION OF DIFFERENT DATASETS

Dataset	Category	Recall	Precision	F1-Score	OA	MIoU
GID	Built-up	81.44%	63.59%	71.42%	79.74%	61.13%
	Farmland	74.30%	63.88%	68.69%		
	Forest	**95.56%**	77.94%	85.86%		
	Water	**78.54%**	81.76%	80.12%		
	Meadow	66.61%	77.36%	71.58%		
NIE-GID	Built-up	83.03%	60.04%	69.69%	77.24%	61.30%
	Farmland	53.70%	69.79%	60.69%		
	Forest	86.00%	77.90%	81.75%		
	Water	74.09%	89.50%	81.07%		
	Meadow	**92.11%**	76.90%	**83.82%**		
GauGAN-GID	Built-up	81.75%	71.23%	76.13%	82.70%	66.86%
	Farmland	75.67%	67.53%	71.37%		
	Forest	95.47%	82.70%	88.63%		
	Water	77.97%	80.91%	79.41%		
	Meadow	83.57%	83.74%	83.66%		
Ours-GID	Built-up	**94.40%**	**97.84%**	**96.09%**	**85.74%**	**77.81%**
	Farmland	**82.18%**	**76.52%**	**79.25%**		
	Forest	94.74%	**92.63%**	**93.67%**		
	Water	78.45%	**91.87%**	**84.64%**		
	Meadow	76.61%	**88.16%**	81.98%		

Note: The bold font represents the value with the highest classification accuracy.

TABLE III
ABLATION EXPERIMENT OF THE PROPOSED IMPROVED DEEPLAB V3+

Models	F1-score					OA	MIoU
	Built-up	Farmland	Forest	Meadow	Water		
ASPP+ARM+FFM+DeepLab V3+	91.95%	**89.11%**	94.82%	78.80%	**92.18%**	**85.79%**	**81.98%**
ASPP+ARM+DeepLab V3+	93.04%	85.49%	**96.66%**	87.21%	82.57%	84.88%	80.56%
ASPP+FFM+DeepLab V3+	93.34%	84.79%	93.25%	87.45%	82.92%	85.75%	79.40%
ARM+FFM+DeepLab V3+	94.33%	84.80%	90.93%	86.01%	83.09%	85.49%	78.56%
FFM+DeepLab V3+	93.83%	78.89%	97.30%	**88.60%**	80.81%	84.97%	79.12%
ARM+DeepLab V3+	93.07%	81.18%	93.10%	87.26%	82.48%	85.34%	78.01%
ASPP+DeepLab V3+	92.58%	84.23%	91.04%	82.95%	86.46%	84.98%	77.90%
DeepLab V3+	**96.09%**	79.25%	93.67%	84.64%	81.98%	85.74%	77.81%

Note: The bold font represents the value with the highest classification accuracy.

For land use classification, we employed the classical DeepLab V3+ network, which was implemented on the TensorFlow framework platform using the Adam optimizer with a learning rate of 3e-4 and a training period of 70 epochs. The test set was not included in the training data and remained unchanged throughout subsequent experiments. Finally, the classification results are presented in Table II.

In terms of recall accuracy, Ours-GID shows the best performance in classifying built-up and farmland, with a significant improvement in meadow accuracy without any obvious low accuracy for any land use classes. For precision accuracy, the proposed method shows the best performance among all classes, which compensates for the low classification accuracy of built-up and farmland in GauGAN-GID. In terms of

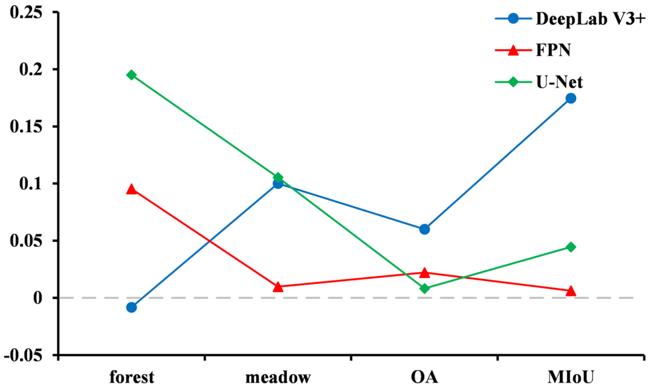


Fig. 15. Difference in model accuracy before and after enhancement.

F1-Score accuracy, Ours-GID exhibits the best performance for all land use classes except meadow, where it is outperformed by NIE-GID. With respect to OA and MIoU, Ours-GID achieved 85.74% and 77.81%, respectively, indicating a significant improvement in overall classification accuracy compared to GauGAN.

D. Influence of Improved GauGAN on Different Classification Models

To examine whether Ours-GID is applicable to other models, the GID datasets before and after enhancement were input to the FPN [39], U-Net [40], and DeepLab V3+ network models. The results were evaluated by using recall, the OA and MIoU.

Fig. 15 shows that all accuracies in the FPN network have increased, with the meadow accuracy and MIoU increasing relatively less. This is mainly due to the fact that the feature fusion method in FPN can effectively learn shallow features and high-level features, alleviating the problem of pixel misalignment caused by convolutional operations, thus making the gain from data augmentation relatively low. The accuracy improvement of U-Net and FPN in the forest class is notable, which can be attributed to the initial confusion of the network in distinguishing forest from other classes prior to data augmentation. After augmentation, U-Net exhibits the highest improvement in land use classification accuracy for both forest and meadow classes. Due to the similarity between meadow and forest, the increase in the number of meadow samples significantly reduced the network's ability to recognize both, leading to a decrease in the accuracy of forest in DeepLab V3+. This also highlights the need to improve the ability of the DeepLab V3+ network to learn the forest class. However, the overall classification accuracy of the model has significantly improved after augmentation.

In general, the proposed improved GauGAN method is shown to be well-suited for land use classification and effective in addressing sample imbalance issues, as semantic segmentation models exhibit improved classification accuracy after augmentation.

E. Ablation of Improved DeepLab V3+ Network

Building upon addressing the issue of sample imbalance, this article proposes improvements to the DeepLab V3+ network to further enhance land use classification accuracy. Specifically, modifications to the ASPP module and integration of the FFM and ARM modules into the network are proposed to effectively improve the model's ability to learn and fuse contextual semantic information. To demonstrate the effectiveness of the proposed improvements, a series of ablation experiments were conducted, and the results are presented in the following (as in Table III).

After increasing the dilation rate in the ASPP module, the model's learning ability for farmland and water has been improved, and when combined with the FFM and ARM modules, respectively, the accuracy of large-scale land use in the classification results has significantly increased. This also indicates that increasing the dilation rates can help the model learn semantic information of larger-scale features. Although the ARM module did not perform exceptionally well in different land cover classifications, overall, there are no poorly classified land use categories, and the MIoU accuracy has improved. The ARM module effectively improves the segmentation ability of the model by optimizing the output feature maps of the ASPP module. When the FFM module was integrated into the network, the accuracy of forest and meadow was significantly improved, demonstrating its ability to enhance the model's recognition capability of these two types. However, it still lacks the ability to learn large-scale land use such as farmland. Therefore, by effectively combining these three methods, their respective deficiencies can be complemented to achieve high-precision land cover classification.

F. Assessing the Results of Classification of Improved DeepLab V3+ Network

To demonstrate the superiority of the improved DeepLab V3+, this article conducted a series of comparative experiments on the model's efficiency and classification accuracy, building upon the solution of the sample imbalance problem in land use classification. The experiment used the Adam optimizer with a learning rate of 0.0003. The learning rate decay strategy was gradient decay and the training epoch were 70. The classification models include U-Net, FPN, DeepLab V3+, ISANet [41], MANet [42], TransUNet [43], and Ours.

Model efficiency is evaluated by calculating the number of images that the model can process per second, measured in frames per second (FPS). As shown in Fig. 16, classical convolutional neural networks have relatively fast processing efficiency. By incorporating attention mechanisms, the model parameters increase, resulting in decreased efficiency, as observed in ISANet, MANet, and Ours. TransUNet has the lowest model efficiency, mainly due to the Transformer. The computational complexity and number of parameters of the Transformer are much larger than those of convolutional neural networks, resulting in higher computational costs and slower inference speeds [44]. The model efficiency is relatively low due to the incorporation of attention mechanisms to enhance model performance in this study. Therefore, the model structure needs further lightweight design.

Regarding the comparison of classification accuracy (as in Table IV), our proposed model outperforms others in terms

TABLE IV
RESULTS OF CLASSIFICATION OF DIFFERENT MODELS

Models	F1-score					OA	MIoU
	Built-up	Farmland	Forest	Meadow	Water		
U-Net	89.66%	70.33%	95.08%	83.69%	87.81%	80.66%	75.27%
FPN	93.73%	75.00%	91.55%	86.54%	81.90%	82.14%	75.65%
DeepLab V3+	**96.09%**	79.25%	93.67%	**84.64%**	81.98%	85.74%	77.81%
ISANet	74.83%	77.83%	74.64%	80.08%	86.03%	70.19%	65.06%
MANet	82.34%	74.87%	82.24%	80.36%	75.72%	76.34%	65.55%
TransUNet	51.94%	65.11%	**95.38%**	77.03%	70.36%	66.52%	58.28%
Ours	91.95%	**89.11%**	94.82%	78.80%	**92.18%**	**85.79%**	**81.98%**

Note: The bold font represents the value with the highest classification accuracy.

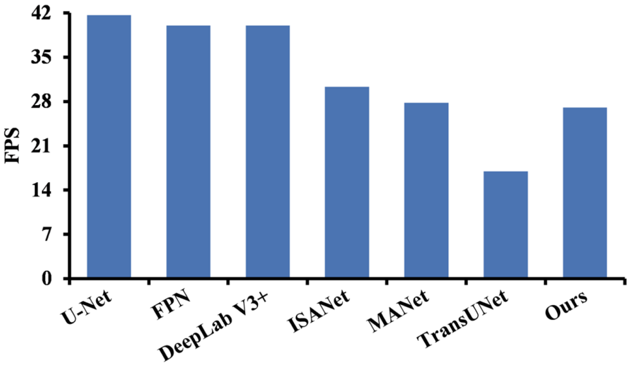


Fig. 16. Comparison of the efficiency of different models.

of OA and MIoU, as well as F1-Score for all land use types. TransUNet performs poorly in classification accuracy, which can be attributed to the fact that Transformer-based models require a large amount of data to learn effectively, otherwise leading to underfitting issues. As shown in Table IV, TransUNet performs well only for the forest land cover type, and poorly for other types. U-Net, FPN, and DeepLab V3+ models all exhibit higher accuracy than ISANet and MANet, potentially due to excessive use of attention mechanisms that require significant computational resources and may overlook the relationships and details among land use types. In addition, DeepLab V3+ has the best classification performance for built-up and meadow, but the overall classification accuracy is still lower than that of our proposed model, and the ability to learn about farmland still needs further improvement. In summary, our proposed method improves the DeepLab V3+ network model to some extent, enabling it to better perform land use classification.

IV. DISCUSSION

In this article, we proposed improved GauGAN to solve the problem of dataset imbalance. Based on the earlier, we enhanced the accuracy of land use classification by improving the DeepLab V3+. The quality of land use images generated by our method was significantly better than those obtained by other network as they were more consistent with the real samples. Our proposed method not only outperforms normal image augmentation methods but also adapts well to semantic segmentation networks for land use classification. Furthermore, our improved DeepLab V3+ model, although exhibiting average classification efficiency, achieves superior overall land use classification accuracy compared to other models.

A. *Choosing the Method for Image Enhancement*

The problem of sample imbalance in remote sensing datasets hinders improvements in the accuracy of land use classification. In addition to modifying the loss function, the amount of data can be increased to solve this problem. For example, Jiang used a StyleGAN-based method to increase the number of features in the DIOR [45], Tekerek and Yapici [46] used CycleGAN to increase the amount of data in BIG2015, and Ding et al. [47] used pix2pixHD to synthesize high-resolution dermoscopic images. These studies show that a GAN-based approach can solve the problem of data imbalance. However, the StyleGAN and CycleGAN networks cannot synthesize high-resolution land use data with rich semantic information. The GauGAN network used in this article can accomplish this task. The pix2pixHD network synthesizes data with a relatively high resolution, but is prone to collapse, and the features generated by it are significantly inconsistent with real features. Moreover, the process by which the GauGAN network synthesizes images of features with few samples is stable and not prone to collapse. This is because the SPADE normalization in the generator can compensate for the loss of semantic information in the other normalization layers. However, GauGAN also has some limitations, as can be seen from Fig. 14, its ability to synthesize complex land cover such as buildings is still weak. Therefore, it is proposed to improve the generator and discriminator of GauGAN, further enhancing the network's ability to synthesize land use, and thus achieving image enhancement of small-sample land use and addressing the issue of data imbalance.

In addition, general methods of image enhancement for semantic segmentation include cropping, rotation, and scaling, and can significantly improve accuracy with a small amount of data (as in Table II). Compared with the improved GauGAN method of enhancement, general image enhancement induces the model to focus to a greater extent on learning features represented by few samples, but it tends to ignore the learning of other features. This results in an overall decrease in its accuracy.

B. Improved DeepLab V3+ in Comparison With Other Networks

Once the problem of data imbalance in GID dataset has been solved, the improvement in classification accuracy is limited if only the classical network for semantic segmentation is used [48], [49]. In addition, targeted improvements to the classical network are needed according to different classification tasks. For example, Wu et al. [50] classified hydroponic lettuce by modifying the backbone network of the DeepLab V3+ network, and Li et al. [51] modified U-Net by using the residual module as the coding module for segmenting images of defects in steel surfaces. This improved the learning capability of the model. However, in land use classification, it is also necessary to consider the model's learning ability for different scale and contextual semantic information in order to achieve better classification performance.

Therefore, this article proposes the use of FFM to enhance the model's ability to select relevant features for land use classification, guided by the ARM to optimize the learning of these features. We also modified the ASPP module by increasing the dilation rates to enhance the model's learning ability for objects of different scales. Compared to U-Net, FPN, DeepLab V3+, MANet, ISANet, and TransUNet, the proposed method with our improvements is more conducive to land use classification.

C. Shortcomings of the Study

The GauGAN model used here requires a relatively large amount of data for adversarial learning, which is not suitable for data enhancement with few samples. In addition, the GauGAN model and the DeepLab V3+ model have a large number of parameters, and are not proposed based on land use classification under remote sensing images, which needs to be targeted for improvement and innovation.

V. Conclusion

The problem of data imbalance in land use classification in the presence of few features leads to poor performance. To solve this problem, we propose to improve GauGAN at the data level to synthesize high-resolution and detail-rich small-sample land use objects, and further improve the accuracy of land use classification by enhancing DeepLab V3+. Specifically, we incorporate the proposed ARM module into the generator of GauGAN for optimization, and modify the normalization method to spectral normalization in the discriminator to facilitate stable model training. Later, we proposed FFM to enhance the fusion ability of high-order and low-order features in the DeepLab V3+ network, modified the dilation rate of the ASPP module to improve the learning ability for different-scale objects, and used ARM to optimize the model's feature selection ability. Finally, high-precision land use classification is achieved based on the aforementioned methods.

In future work, we plan to further improve the proposed model by reducing the number of its parameters while maintaining its performance in training. It is also important to collect more feature-related data to expand land use datasets as this can help the GAN network generate land use data containing rich semantic information. Finally, high-precision land use classification and intelligent mapping need to be realized under different spatial and temporal resolutions.

References

[1] B. Chen, B. Xu, and P. Gong, "Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities," *Big Earth Data*, vol. 5, no. 3, pp. 410–441, Jul. 2021, doi: 10.1080/20964471.2021.1939243.

[2] J. E. Patino and J. C. Duque, "A review of regional science applications of satellite remote sensing in urban settings," *Comput., Environ., Urban Syst.*, vol. 37, pp. 1–17, Jan. 2013, doi: 10.1016/j.compenvurbsys.2012.06.003.

[3] X. Liu et al., "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017, doi: 10.1080/13658816.2017.1324976.

[4] B. Zhao, Y. Zhong, and L. Zhang, "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 73–85, Jun. 2016, doi: 10.1016/j.isprsjprs.2016.03.004.

[5] P. H. Verburg, K. Neumann, and L. Nol, "Challenges in using land use and land cover data for global change studies," *Glob. Change Biol.*, vol. 17, no. 2, pp. 974–989, 2011, doi: 10.1111/j.1365-2486.2010.02307.x.

[6] R. E. McRoberts, "Post-classification approaches to estimating change in forest area using remotely sensed auxiliary data," *Remote Sens. Environ.*, vol. 151, pp. 149–156, Aug. 2014, doi: 10.1016/j.rse.2013.03.036.

[7] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 581–589, Mar. 2005, doi: 10.1109/TGRS.2004.839547.

[8] M. Voltersen, C. Berger, S. Hese, and C. Schmullius, "Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level," *Remote Sens. Environ.*, vol. 154, pp. 192–201, Nov. 2014, doi: 10.1016/j.rse.2014.08.024.

[9] I. Walde, S. Hese, C. Berger, and C. Schmullius, "From land cover-graphs to urban structure types," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 3, pp. 584–609, Mar. 2014, doi: 10.1080/13658816.2013.865189.

[10] X. Huang, H. Liu, and L. Zhang, "Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3639–3657, Jul. 2015, doi: 10.1109/TGRS.2014.2380779.

[11] J. Sun, H. Wang, Z. Song, J. Lu, P. Meng, and S. Qin, "Mapping essential urban land use categories in Nanjing by integrating multi-source big data," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 2386, doi: 10.3390/rs12152386.

[12] S. Chang, Z. Wang, D. Mao, K. Guan, M. Jia, and C. Chen, "Mapping the essential urban land use in Changchun by applying random forest and multi-source geospatial data," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 2488, doi: 10.3390/rs12152488.

[13] A. Alem and S. Kumar, "Deep learning methods for land cover and land use classification in remote sensing: A review," in *Proc. IEEE 8th Int. Conf. Rel., Infocom Technol., Optim. (Trends Future Directions)*, 2020, pp. 903–908, doi: 10.1109/ICRITO48877.2020.9197824.

[14] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, Jan. 2020, Art. no. 2495, doi: 10.3390/rs12152495.

[15] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019, doi: 10.1109/JSTARS.2019.2918242.

[16] Q. Weng, Z. Mao, J. Lin, and W. Guo, "Land-use classification via extreme learning classifier based on deep convolutional features," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 704–708, May 2017, doi: 10.1109/LGRS.2017.2672643.

[17] D. Xiong, C. He, X. Liu, and M. Liao, "An end-to-end Bayesian segmentation network based on a generative adversarial network for remote sensing images," *Remote Sens.*, vol. 12, no. 2, pp. 216–237, Jan. 2020, doi: 10.3390/rs12020216.

[18] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1937–1949, Oct. 2019, doi: 10.1007/s11063-018-09977-1.

[19] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification," *J. Ambient Intell. Humanized Comput.*, to be published, doi: 10.1007/s12652-020-01773-x.

[20] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2020, pp. 465–476, doi: 10.18653/v1/2020.acl-main.45.

[21] E. J. Bjerrum, "SMILES enumeration as data augmentation for neural network modeling of molecules," 2017, *arXiv:1703.07076*.

[22] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 13001–13008, doi: 10.1609/aaai.v34i07.7000.

[23] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear independent components estimation," 2015, *arXiv:1410.8516*.

[24] I. J. Goodfellow et al., "Generative adversarial networks," 2014, *arXiv:406.2661v1*.

[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[26] K. Wang, C. Gou, Y.-J. Duan, L. Yilun, and X.-H. Zheng, "Generative adversarial networks: The state of the art and beyond," *Zidonghua Xuebao/Acta Automatica Sinica*, vol. 43, pp. 321–332, Mar. 2017, doi: 10.16383/j.aas.2017.y000003.

[27] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8107–8116, doi: 10.1109/CVPR42600.2020.00813.

[28] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," presented at the Int. Conf. Learn. Representations, Jan. 2023. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm

[29] M. Kang, W. J. Shim, M. Cho, and J. Park, "Rebooting ACGAN: Auxiliary classifier GANs with stable training," presented at the Adv. Neural Inf. Process. Syst., Nov. 2021. [Online]. Available: https://openreview.net/forum?id=r7UC-b67YkO

[30] Y. R. Cui, Q. Liu, C. Y. Gao, and Z. Su, "FashionGAN: Display your fashion design using conditional generative adversarial nets," *Comput. Graph. Forum*, vol. 37, no. 7, pp. 109–119, 2018, doi: 10.1111/cgf.13552.

[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.

[32] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[33] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Los Alamitos, CA, USA, 2019, pp. 2332–2341, doi: 10.1109/CVPR.2019.00244.

[34] 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, doi: 10.1016/j.rse.2019.111322.

[35] 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. 2018, doi: 10.1109/TPAMI.2017.2699184.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[37] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369, doi: 10.1109/ICPR.2010.579.

[38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595, doi 10.1109/CVPR.2018.00068.

[39] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[41] L. Huang, Y. Yuan, J. Guo, C. Zhang, X. Chen, and J. Wang, "Interlaced sparse self-attention for semantic segmentation," 2019, *arXiv:1907.12273*.

[42] R. Li et al., "Multiattention network for semantic segmentation of fine-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jan. 2022, doi: 10.1109/TGRS.2021.3093977.

[43] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[44] C. Yang et al., "Lite vision transformer with enhanced self-attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11988–11998, doi: 10.1109/CVPR52688.2022.01169.

[45] Y. Jiang and B. Zhu, "Data augmentation for remote sensing image based on generative adversarial networks under condition of few samples," *Laser Optoelectron. Prog.*, vol. 58, no. 8, pp. 238–244, 2021, doi: 10.3788/LOP202158.0810022.

[46] A. Tekerek and M. M. Yapici, "A novel malware classification and augmentation model based on convolutional neural network," *Comput. Secur.*, vol. 112, Jan. 2022, Art. no. 102515, doi: 10.1016/j.cose.2021.102515.

[47] S. Ding et al., "High-resolution dermoscopy image synthesis with conditional generative adversarial networks," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102224, doi: 10.1016/j.bspc.2020.102224.

[48] A. Y. Wenya LIU, "Urban green space extraction from GF-2 remote sensing image based on DeepLabv3+ semantic segmentation model," *Remote Sens. Natural Resour.*, vol. 32, no. 2, pp. 120–129, Jun. 2020, doi: 10.6046/gtzyyg.2020.02.16.

[49] C. Yang, Z. Kong, Q. Xie, and J. Du, "Image recognition method for transmission line based on the DeepLab v3+ deep convolutional network," *Electric Power Eng. Technol.*, vol. 40, no. 4, pp. 189–194, 2021, doi: 10.12158/j.2096-3203.2021.04.027.

[50] Z. Wu, R. Yang, F. Gao, W. Wang, L. Fu, and R. Li, "Segmentation of abnormal leaves of hydroponic lettuce based on DeepLabV3+ for robotic sorting," *Comput. Electron. Agriculture*, vol. 190, Nov. 2021, Art. no. 106443, doi: 10.1016/j.compag.2021.106443.

[51] Y. Li, Y. Li, J. Liu, H. Fan, and Q. Wang, "Research on segmentation of steel surface defect images based on improved res-UNet network," *J. Electron. Inf. Technol.*, vol. 44, no. 5, pp. 1513–1520, May 2022, doi: 10.11999/JEIT211350.