ASYMMETRIC FUZZY CLASSIFICATION NETWORKS FOR CONSTRUCTION LAND DETECTION IN HIGH RESOLUTION REMOTE SENSING IMAGES

Ruixin Fang [1], Zhaocong Wu [1,2,3] *, Xiaohui Song [4]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China - (fangrx, zcwoo)@whu.edu.cn
[2] Hubei Luojia Laboratory, China
[3] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China
[4] Henan Academy of Science, China - xhsong@foxmail.com

Commission III and IV

KEY WORDS: Construction land detection, urban expansion, scene classification, neural network, remote sensing image, urban dynamic cognition.

ABSTRACT:

Urbanization is an essential phase of a nation's economic development. A very effective way to examine urban growth is to look at how impervious surface changes over time, however impervious surface can only show the current situation in terms of urban development. Compared to the existing methods, the use of construction bare land to monitor urban growth has the following benefits over currently used techniques. First, it is possible to track the progress of structures being constructed as part of urban expansion. The second is to assess the city's development intensity and identify the inward expansion. Therefore, the detection of construction bare land is of great significance for the development of a more sophisticated urban dynamic perception technology. This paper proposes an asymmetric fuzzy classification network (AFCNet) for detection of construction bare land scenes. The generation of fuzzy sample sets, the backbone network, and the proposed fuzzy classification module make up the method's three primary components. The deep features of the scene are then extracted using the fuzzy classification network and converted into ambiguity. Finally, the ambiguity is converted into predicted probability using the fuzzy weight vector. The fuzzy sample set is generated to introduce more prior information into the network. High-level features are extracted using the backbone network. Fuzzy classification methods based on spectral features are used to improve the performance of scene classification. The results demonstrate that the OA of our method is higher than all other comparison methods.

1. INTRODUCTION

Today in a world of rapid urbanization, urban monitoring and dynamic cognition are crucial. Urbanization is an essential phase of a nation's economic development. Remote sensing can model and forecast how urban elements will change, which is useful for managing urban space (Feng et al., 2022).

As many cities continue to grow as a result of globalization and will continue to expand in the future, greater efforts are needed to improve urban planning solutions and tools. Urban monitoring is the basis of urban planning. Remote sensing monitoring of urban land use changes and spatial evolution mainly involves the identification, extraction and monitoring of urban land cover and driving factors, thus directly serving urban comprehensive decision-making and management. Based on remote sensing monitoring, it is possible to simulate and predict the changes of urban elements, which is an in-depth deduction for the changes of urban elements, and has a practical guiding role in urban space management, planning and development.

The majority of long-term land cover data used for current monitoring of urban growth is of relatively low resolution (Wang et al., 2022), including MODIS land cover at 500 m and 1 km resolution, ESACCI land cover at 300 m resolution, and GLASS land cover at 5 km resolution. High-resolution monitoring of urban growth is less researched. Liu et al. (2020) examined the urbanization process using a technique for spotting changes in impervious surface in remote sensing images. A very effective way to examine urban growth is to look at how impervious surface changes over time, however impervious surface can only show the current situation in terms of urban development. The expansion tendency and intermediate status of the city cannot be reflected in the state's performance. In order to examine the growth of cities, some people also use urban night light data (Zhou et al., 2015), however this method is less sensitive to changes occurring within the city itself and is more prone to observing the expansion of urban boundary contours.

Compared to the existing methods, the use of construction bare land to monitor urban growth has the following benefits over currently used techniques. First, it is possible to track the progress of structures being constructed as part of urban expansion. The second is to assess the city's development intensity and identify the inward expansion. Therefore, the detection of construction bare land is of great significance for the development of a more sophisticated urban dynamic perception technology.

The current monitoring of urban expansion mainly uses long-term land cover data. Unfortunately, existing continuous multi-temporal data are obtained at lower spatial resolutions, such as MODIS land cover at 500 m and 1 km resolution, ESACCI land cover at 300 m, GLASS land cover at 5 km, these land cover data contain a lot of mixed pixels with low spatial resolution. There are studies on monitoring the urbanization process by detecting changes in the impervious surface in remote sensing images (Liu et al., 2020). Some researchers also use urban night light data to

* Corresponding author

study the development of cities, but night light data is more inclined to observe the expansion of the urban boundary contour, and is not very sensitive to changes within the city.

The construction bare land is an intermediate state of urban expansion, which not only provides new ideas for studying the temporal and spatial evolution of urbanization, but also can be cross-validated with products based on other data such as impervious surfaces. Therefore, the detection of construction bare land is more refined for development. The remote sensing image urbanization monitoring technology is of great significance.

At present, there are few studies on the recognition direction of construction land scenes, and most of them still focus on the visual interpretation of remote sensing images, and manually interpret the construction sites in the images. The visual interpretation process has high requirements on the professional background knowledge of the interpreters. When performing inspection tasks on long-term and large-scale construction sites, the high-resolution images used are massive, and the use of manual interpretation is inefficient, strong subjectivity, high cost, and obviously cannot meet the requirements of information extraction tasks.

Traditional classification methods have a poor effect on the recognition of construction bare land scenes, because bare soil serves as the primary identifier for building barren land with other elements varied, and there isn't a distinguishing factor. Additionally, the spots' shapes are asymmetrical on the remote sensing image, and the arrangement regulation is quite complex. This paper proposes a deep learning neural network for detection of construction bare land scenes.

The deep learning neural network classification method is very suitable for the scene classification of remote sensing images. There are also some studies using deep learning methods to identify construction sites. Because of its powerful image feature extraction ability, convolutional neural networks are widely used in land cover or land use classification, such as Googlenet (Szegedy et al., 2014), Shufflenet (Ma et al., 2018), ResNet (He et al., 2015), Nasnet (Zoph et al., 2018). Among them, the ResNet network pre-trained on ImageNet is used to remote sensing image classification, and it is compared with classical machine learning models such as random forests. The results show that convolutional neural networks have better classification effects on remote sensing images.

However, deep learning requires a large number of training samples, and currently there are few public datasets about construction land, and manual labelling requires a lot of human and material resources. Aiming at the problem that deep learning requires a large number of training samples, this paper proposes a method to generate a sample dataset on a large scale using the spectral information of remote sensing images.

In summary, the main contributions of this paper are as follows.
1) A method for urban expansion monitoring is proposed. The use of construction bare land can show the intermediate state of urban expansion, which provides a new idea for studying the spatiotemporal evolution of urbanization.
2) A classification method AFCNet is proposed for the recognition of construction bare land scenes. Fuzzy samples can be used as auxiliary datasets to extract deep features of remote sensing images through fuzzy classification network.
3) A prior fuzzy sample set generation method based on spectral index is proposed. The soil index is used to generate fuzzy samples of suspected construction bare land from an easily available unlabeled data set to solve the problem of insufficient construction bare land samples.

2. METHODOLOGY

2.1 Overview

In this paper, the construction bare land (or the construction land) refers to the bare land with civil construction activities or pre-construction foundation work. Different from the general bare land, the construction bare land has obvious traces of human activities.

This paper proposes a construction bare land detection method, which uses the fuzzy set generated by the soil index and the manually labeled data set for joint training, then extracts the deep features of the scenes through a neural network and converts them into classification ambiguity, and finally uses the parameterized fuzzy weight vector to convert the classification ambiguity into a predicted probability, so as to achieve the effect of scene classification.

The AFCNet classification method consists of three main parts, namely, the generation of fuzzy sample sets, the backbone network and the proposed fuzzy classification module. The fuzzy sample set is generated to introduce more prior information into the network (refer to section 2.2 for details). The backbone network is used to extract high-level features (refer to section 2.3). Fuzzy classification methods based on spectral features are used to improve the performance of scene classification (refer to section 2.4). Furthermore, to fit the model, we design a two-stage loss function, which not only optimizes the training process of the network, but also significantly reduces detection dropouts (refer to section 2.5).

2.2 Generation of Fuzzy Sample Sets

In 1965, Zadeh expounded the fuzzy theory in his book "Fuzzy Sets", which was later widely used in the field of land use remote sensing classification. It is a multi-valued logic system that quantitatively states uncertainty. Boolean logic statement, any value between 0 and 1 can be used to represent the transition state between yes and no. By avoiding hard boundaries, fuzzy logic can describe the real world better than Boolean logic with binary semantics.

In this paper, the soil index is used to generate the fuzzy dataset required for joint training, and the ambiguity is initialized according to the change of the soil index of each sample.Studies have shown that soil has different spectral response characteristics in the near-infrared (NIR) and short-wave infrared (SWIR) bands from other ground objects. In order to quantitatively describe this characteristic, Nguyen et al. (2021) proposed a soil index called MBI. Its calculation method is as follows:

$$\mathrm{MBI} = \frac{SWIR1 - SWIR2 - NIR}{SWIR1 - SWIR2 - NIR} + f \qquad (1)$$

where MBI is the soil index, SWIR1 is the reflectivity in the short-wave infrared band with a center wavelength around 1600 nm, SWIR2 is the reflectivity in the short-wave infrared band with a center wavelength around 2200 nm, NIR is the reflectivity in the near-infrared band with a center wavelength around 850 nm, f is a parameter used to adjust the exponential distribution, which is artificially set through experiments.

Since the training datasets (GID, SYSU-CD, samples from Google Earth, etc.) used by the neural network in this paper do not have the short-wave infrared band (SWIR), the two-phase images of Landsat-8 and Sentinel-2 are used and sliced to obtain the difference of MBI. The slices are sorted and grouped by the quantile according to their difference of MBI in the histogram. A fuzzy set with initial ambiguity parameters is thus generated. The difference of the soil index is recorded as d-MBI.

Then, slice the result map generated by the above method, calculate the average d-MBI statistics of each slice to obtain a histogram, and then sort these slices according to the size of d-MBI, and divide them into four groups according to quantiles, and each group is divided into four groups according to the quantile. The number of digits marks the fuzzy interval μ(x).

$$\mu(x) = \begin{cases} \{\mu \mid a_1 \le \mu < b_1\} & x \ge Pct(c_1) \\ \{\mu \mid a_2 \le \mu < b_2\} & Pct(c_2) \le x < Pct(c_1) \\ \{\mu \mid a_3 \le \mu < b_3\} & Pct(c_3) \le x < Pct(c_2) \\ \{\mu \mid a_4 \le \mu < b_4\} & Pct(c_4) \le x < Pct(c_3) \end{cases} \quad (2)$$

where μ(x) represents the ambiguity interval of the sample group, $a_i(i = 1,2,3,4)$ represents the lower bound of the fuzzy interval of the corresponding sample group, and $b_i(i = 1,2,3,4)$ represents the ambiguity of the corresponding sample group, the upper bound of the interval. $Pct(c_i)(i = 1,2,3,4)$ represents the c_i percentile of the sample group means of d-MBI in all samples. In this way, a fuzzy set is generated using the spectral information of the image, and the fuzzy parameter is initialized.

2.3 Backbone

ResNet is a type of deep residual network. In order to create multi-scale hierarchical features, we design a backbone network based on ResNet as an encoder, remove the final fully connected layer, and perform a down sampling operation before integrating the AFC module suggested in this paper. To categorize the fuzzy sets, we make use of a fuzzy classification module as a decoder. The AFC module uses the initial value of the fuzzy quantile as the weight to train the network model to update these weight parameters, and finally, the result can be obtained after hard classification to the fuzzy data. The feature extraction network provides the probability that the training sample belongs to the sample group.

As DCNN consists of multiple trainable layers that can extract expressive features at different levels, including low-level and high-level feature information, high-level features are more suitable for the extraction of easily confused objects that have semantic characteristics. Low-level features mainly reflect spatial structural information of objects, such as the material or surface characteristics, and boundary information. Remote sensing image and natural image have similar low-level features; therefore, using this pretrained strategy can reduce overfitting to a certain extent. Different from other studies using pretrained networks, we removed the last fully connected layer and down-sampling operations.

First, we construct a backbone network to generate multi-scale deep features. The backbone network consists of a convolutional layer, a pooling layer, and three residual blocks. Fig. 3 shows the illustration of a residual block. Each residual block contains three convolutional layers. Batch normalization and rectified linear unit (ReLU) activation layer are followed by each convolution layer.

Let F(X) be the underlying function to be learned by a residual block. X denotes the input of the first layer of a residual block. The residual function is expressed as R(X)=F(X)−X, and F(X) can be written as F(X)=R(X)+X. The bias term in convolutional layers is not used. R(X) can be expanded as

$$R(X) = \sigma(\sigma(\sigma(X * W_1) * W_2) * W_3) \quad (3)$$

where W1, W2, and W3 represent the convolutional kernels of three convolutional layers of a residual block, and σ refers to the ReLU activation function.

For the last two residual blocks in the backbone network, the convolution kernel of the first convolutional layer has a stride of 2. The size of features is reduced by half after each residual block. Therefore, the backbone network generates a feature hierarchy consisting of feature maps at three scales with a scaling step of 2. The multiscale hierarchy features are fused in the latter fusion network.

2.4 Ambiguity Module

We apply the fuzzy classification method using spectral features to improve accuracy of scene classification. The classification procedure is described as follows.

$$F(x) = [p_1 \ p_2 \ p_3 \ p_4 \ p_5 \ p_6] \quad (4)$$

where p_i (i=1, 2, …, 6) represents the probability that the training sample given by the model belongs to the corresponding sample group. F(x) will be the input to the ASF module.
In the ASF module, we calculate

$$P(X) = F(x)G(x) \quad (5)$$

Where

$$G(x) = [w_1 \ w_2 \ w_3 \ w_4 \ w_5 \ w_6]^T \quad (6)$$

w_i (i=1, 2, …, 6) is a set of parameters initialized according to μ(x), which will be updated with the iteration of the model.
The final obtained P(x) is the ambiguity that the sample predicted by the model belongs to the first category.

2.5 Two-stage Loss

To fit the model, we design a two-stage loss function.

$$\mathcal{L} = a\mathcal{L}_1 + b\mathcal{L}_2 \quad (7)$$

When epoch$\leqslant e_0$ is called the first stage, at this time a=1, b=0; when epoch>e_0 is called the second stage, at this time a=0, b=1; e_0 is the hyperparameter of this network model.

In the first stage, we update F(X) using the cross-entropy loss function

$$\mathcal{L}_1 = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log(p_{ic}) \quad (8)$$

where M represents the number of sample groups, where y_{ic} represents the label of sample i, 1 when the predicted value falls within the fuzzy interval of the corresponding sample group, 0 otherwise, and p_{ic} represents the probability that sample i is predicted to be class c.

In the second stage, we lock the value of F(X) and update G(X) using the conditional cross-entropy loss function

$$\mathcal{L}_2 = \begin{cases} \mathcal{L}_1 & y_{ic} = 0 \ or \ 1 \\ 0 & y_{ic} \ne 0 \ or \ 1 \end{cases} \quad (9)$$

That is, when the sample is a fuzzy sample, the gradient is no longer updated.

3. EXPERIMENTS

We test our algorithm and analyze the experimental results in this section. For performance comparison, several neural network classification models that have demonstrated better performance are used to compare with our method. Section 3.1 presents the experimental data, Section 3.2 presents the experimental platform performance, and Section 3.3 presents the comparison

method and evaluation metrics. Section 3.4 tests our algorithm and compares it with several existing neural network classification models.

3.1 Experimental Data

At present, there are relatively few public datasets on the identification of construction bare land in high-resolution remote sensing images. Therefore, we processed Google Earth images, SYSU-CD dataset (Shi et al., 2022), and GID dataset (Tong et al., 2020) by manual interpretation, and finally 9060 images were obtained as the experimental data set of this paper, including 1089 image samples of the construction site and 7971 samples not including the construction site.

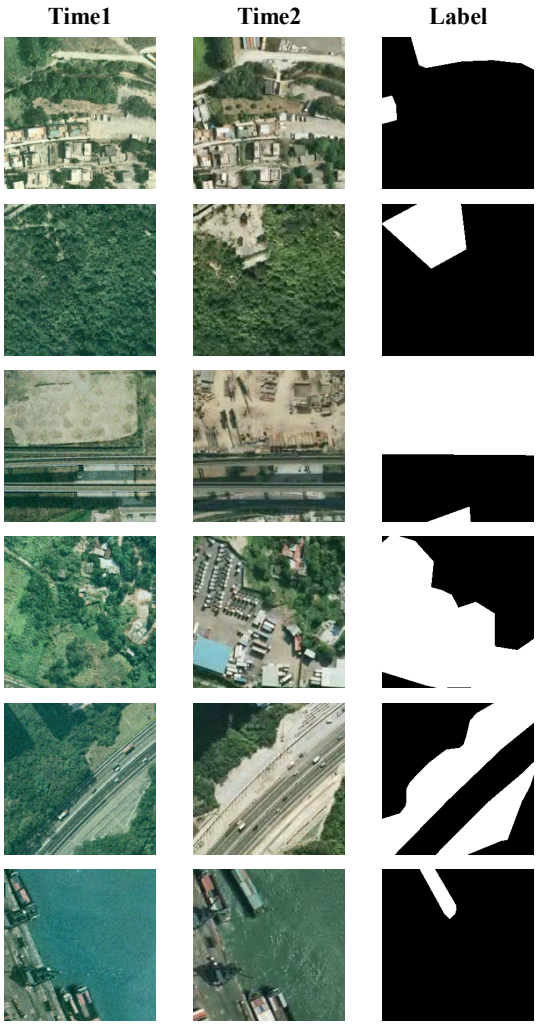


Figure 1 Example of the SYSU-CD dataset

The SYSU-CD dataset is a change detection dataset largely complements existing CD datasets in terms of image resolution, variation type, and data volume, and further provides a new benchmark for CD. The dataset contains 20,000 pairs of 0.5-meter aerial images, taken between 2007 and 2014 in Hong Kong, which has long been a prosperous and densely populated metropolis in southern China. In addition, from 2007 to 2014, the construction and maintenance of ports, shipping routes, and ocean coastal projects in Hong Kong and major shipping hubs in the international and Asia-Pacific regions increased rapidly. Develop coastal economy and maritime transport. Thus, our dataset is greatly complemented with high-rise building change instances that are difficult to label in HRI due to bias and shadow effects, as well as port-related change information compared to previous datasets.

A total of 20,000 pairs of aerial image patches of size 256×256. As shown in Figure 1, the main types of changes in the dataset include (a) new urban buildings; (b) suburban expansion; (c) pre-construction foundation work; (d) vegetation changes; (e) road expansions; (f) offshore construction. The above 6 categories are only a rough description of the main types of changes in the data set, and do not mean that the samples are clearly divided into the above 6 categories.

The GID dataset refers to the Chinese Land Use Classification Standard (GB/T21010-2017) to determine the classification system and labels five categories: buildings, farmland, forests, grasslands, and water bodies. Areas that do not belong to the above five categories and cluttered areas are marked as background, indicated in black. The fine-grained land cover classification set consists of 15 sub-categories: paddy field, irrigated land, dry land, garden land, arbor forest, shrub land, natural meadow, artificial meadow, industrial land, urban dwelling, rural dwelling, transportation land, river, lake and pond. Its training set contains 2000 samples per class, and validation images are labeled at the pixel level.

They constructed a large-scale land-cover dataset with Gaofen-2 (GF-2) satellite imagery. This new dataset, named Gaofen-2 Image Dataset (GID), outperforms existing land cover datasets due to its large coverage, wide distribution, and high spatial resolution. GID consists of two parts: a large-scale classification set and a fine-grained land-cover classification set. The large-scale classification set consists of 150 pixel-level annotated GF-2 images, and the fine classification set consists of 30,000 multi-scale image patches plus 10 pixel-level annotated GF-2 images. 15 classes of training and validation data are collected and relabeled based on 5 classes of training and validation images, respectively. In this paper, the second part is mainly used.

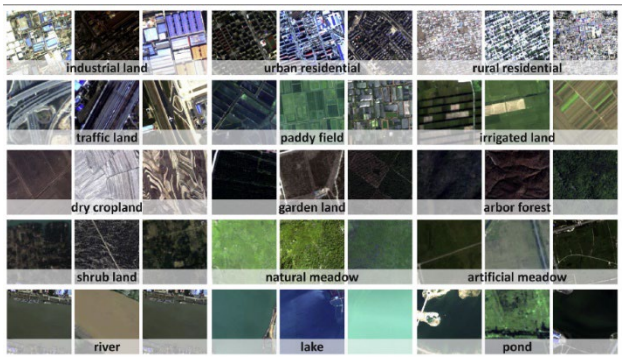


Figure 2 Example of the GID dataset (Tong et al., 2020)

3.2 Experimental Setup

All experiments were performed on a server computer with an AMD Ryzen 9 3950X 16-Core Processor CPU, 64GB RAM, and a Nvidia GeForce RTX 3090 GPU (24GB RAM). The implementation of the framework is based on the open source toolbox Pytorch.

3.3 Experimental Evaluation Indicators

This chapter evaluates the method based on the ISPRS standard evaluation method based on the computation of the confusion matrix on the test set, as well as the accumulated confusion matrix. Different metrics can be derived from the confusion

matrix: compute completeness (recall), correctness (precision) and F1-score (F1-score) for each class, overall accuracy (OA).

$$\text{Precision} = \frac{tp}{tp + fp} \tag{10}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{11}$$

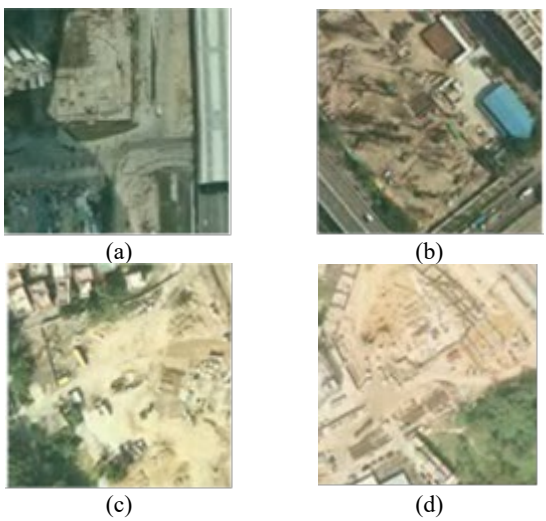
$$\text{F1} = \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

$$OA = \frac{\sum_{i=1}^{M} N_{ii}}{N} \tag{13}$$

Where tp, fp, fn represent the number of true positive, false positive, and false negative samples, respectively. N is the total number of samples, M is the total number of categories of the samples. $\sum_{i=1}^{M} N_{ii}$ is the sum of the number of samples on the diagonal of the confusion matrix, and N is the total number of the samples.

3.4 Experimental Results and Discussion

Based on the experimental evaluation indicators mentioned in Section 3.3, this section compares the final classification results from both qualitative and quantitative perspectives. Figure 3 and Table 1 show the qualitative comparison results of our method and other neural network classification methods, and Table 2 shows the quantitative results of the experiments.



(a) (b) (c) (d)

Figure 3 Qualitative analysis sample examples (all of them are construction bare land scenes)

Methods	(a)	(b)	(c)	(d)
resnet(He et al., 2015)				✓
googlenet(Szegedy, 2014)	✓	✓		✓
shufflenetv2(Ma et al., 2018)			✓	✓
nasnet(Zoph et al., 2018)	✓	✓		✓
densenet121(Huang et al., 2018)				✓
mobilenetv2(Sandler, 2019)				
squeezenet(Iandola, 2016)				
stochasticdepth18(Xie, 2017)	✓	✓		✓
wideres(Zagoruyko, 2017)				✓
AFCNet(Our Work)	✓	✓	✓	✓

Table 1 The prediction results for the sample in Figure 3

Qualitative comparison results are discussed. Figure 3 shows the excellent performance of our model. The 4 sample scenes given are all construction bare ground scenes. It can be seen from the figure that for the 4 typical construction bare ground samples, the method in this paper is correctly identified, and as the other models compared are not all correctly identified.

Methods	OA	P	R	F1
resnet	0.7500	0.7487	0.8938	0.8148
googlenet	0.8000	0.7673	0.9688	0.8564
shufflenetv2	0.8308	0.8671	0.8563	0.8616
nasnet	0.8000	0.7621	0.9813	0.8579
densenet121	0.8500	0.8457	0.9250	0.8836
mobilenetv2	0.8538	0.8280	0.9625	0.8902
squeezenet	0.8000	0.9355	0.7250	0.8169
stochasticdepth18	0.8423	0.8251	0.9438	0.8805
wideres	0.7654	0.7538	0.9188	0.8282
AFCNet	**0.8585**	0.7923	0.9810	0.8766

Table 2 The accuracy comparison of AFCNet and other neural network methods

Quantitative comparison results are discussed. As shown in Table 2, the results demonstrate that the OA of our method is higher than all other comparison methods. The OA, precision, recall and F1 scores of AFCNet are 85.9%, 79.2%, 98.1%, 0.8766, which are higher than the most typical model Nasnet's 80.0%, 76.2%, 98.1%, 0.8579.

Experiments show that our method achieves high classification performance, which proves that this method can be used for tasks such as dynamic perception of urban expansion.

4. CONCLUSION

This paper proposed a novel construction land scene detection method based on fuzzy classification network to resolve the problem that the identification features of construction bare land in remote sensing images are not obvious. Different from traditional scene classification methods and existing deep learning scene classification methods, the main contribution of the proposed method is to introduce spectral prior information to generate pseudo-label data sets, which greatly reduces the need for manual labeling of data. From the detection results, the experimental results show that the OA of our method is higher than all other comparison methods. The proposed method can be used for the dynamic perception of urban expansion, which provides a new idea for studying the spatiotemporal evolution of urbanization.

REFERENCES

Barakat, A., Ouargaf, Z., Khellouk, R., El Jazouli, A. and Touhami, F., 2019. Land Use/Land Cover Change and Environmental Impact Assessment in Béni-Mellal District (Morocco) Using Remote Sensing and GIS. Earth Systems and Environment, 3(1): 113-125.

Bi, Q., Qin, K., Zhang, H. and Xia, G., 2021. Local Semantic Enhanced ConvNet for Aerial Scene Recognition. IEEE Transactions on Image Processing, 30: 6498-6511.

Chen, S. et al., 2022. Remote Sensing Scene Classification via Multi-Branch Local Attention Network. IEEE Transactions on Image Processing, 31: 99-109.

Cheng, G., Han, J. and Lu, X., 2017. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE, 105(10): 1865-1883.

Cheng, G., Xie, X., Han, J., Guo, L. and Xia, G., 2020. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13: 3735-3756.

Deng, G., Wu, Z., Wang, C., Xu, M. and Zhong, Y., 2022. CCANet: Class-Constraint Coarse-to-Fine Attentional Deep Network for Subdecimeter Aerial Image Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-20.

Feng Yongjiu et al., 2022. Key technologies for remote sensing intelligent monitoring and simulation of typical urban elements. Chinese Journal of Surveying and Mapping, 51(04): 577-586.

He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition.

Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K.Q., 2018. Densely Connected Convolutional Networks.

Iandola, F.N. et al., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size.

Li, M., Lei, L., Tang, Y., Sun, Y. and Kuang, G., 2021. An Attention-Guided Multilayer Feature Aggregation Network for Remote Sensing Image Scene Classification. Remote Sensing, 13(16): 3113.

Li, Z., Wu, Q., Cheng, B., Cao, L. and Yang, H., 2022. Remote Sensing Image Scene Classification Based on Object Relationship Reasoning CNN. IEEE Geoscience and Remote Sensing Letters, 19: 1-5.

Liu, F. et al., 2021. Urban Expansion of China from the 1970s to 2020 Based on Remote Sensing Technology. Chinese Geographical Science, 31(5): 765-781.

Liu, X. et al., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. Nature Sustainability, 3(7): 564-570.

Ma, N., Zhang, X., Zheng, H. and Sun, J., 2018. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design.

Nguyen, C.T., Chidthaisong, A., Diem, P.K. and Huo, L., 2021. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. LAND, 10(3).

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L., 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks.

Shi, C., Wang, T. and Wang, L., 2020. Branch Feature Fusion Convolution Network for Remote Sensing Scene Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13: 5194-5210.

Shi, Q. et al., 2022. A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-16.

Szegedy, C. et al., 2014. Going Deeper with Convolutions.

Tang, X. et al., 2021. Attention Consistent Network for Remote Sensing Scene Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 2030-2045.

Tong, X. et al., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sensing of Environment, 237: 111322.

Wang, H., Zhang, X., Du, S., Bai, L. and Liu, B., 2022. Mapping Annual Urban Evolution Process (2001–2018) at 250 m: A normalized multi-objective deep learning regression. Remote Sensing of Environment, 278: 113088.

Wang, J., Zhong, Y., Zheng, Z., Ma, A. and Zhang, L., 2021. RSNet: The Search for Remote Sensing Deep Neural Networks in Recognition Tasks. IEEE Transactions on Geoscience and Remote Sensing, 59(3): 2520-2534.

Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks.

Zagoruyko, S. and Komodakis, N., 2017. Wide Residual Networks.

Zhang Hanchao, 2019. Research on the evolution of spatiotemporal pattern and sustainability evaluation of typical Chinese cities based on remote sensing monitoring, Wuhan University, 158 pp.

Zhao, Z., Li, J., Luo, Z., Li, J. and Chen, C., 2021. Remote Sensing Image Scene Classification Based on an Enhanced Attention Module. IEEE Geoscience and Remote Sensing Letters, 18(11): 1926-1930.

Zhou, N., Hubacek, K. and Roberts, M., 2015. Analysis of spatial patterns of urban growth across South Asia using DMSP-OLS nighttime lights data. Applied Geography, 63: 292-303.

Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V., 2018. Learning Transferable Architectures for Scalable Image Recognition.