Optimization of Natural Landscape Images Using CNN and Improved U-Net Technology

Qinyi Yin[a]

Faculty of Science and Technology, Beijing Normal University,
Hong Kong Baptist University United International College, Zhuhai, Guangdong, 519000, China

Keywords: Convolutional Neural Networks, U-Net, Natural Landscape Image.

Abstract: Natural landscape images are of great significance in many fields such as computer vision and environmental protection, and have a high degree of diversity and complexity. Therefore, zoning optimization for these images is crucial. In this paper, the Convolutional Neural Networks (CNN) are used to classify the images and the improved U-Net is used to segment the images. This paper first introduces the processing method of CNN, introduces the basic concepts, introduces the structure diagram and other content. Secondly, it introduces the improvement method of U-Net and the optimized structure. Then it makes a comparative analysis of different methods. Through the comparative analysis, this paper finds that CNN can classify images more accurately and U-Net can segment images more clearly. However, it also points out the limitations of convolutional neural network for more complex images and the complexity of U-Net after improvement, such as increasing the consumption of computing resources. The experimental results show that the above summarized methods improve the accuracy of image classification and segmentation, and provide a good basis for the optimization of natural landscape image segmentation.

1 INTRODUCTION

The study of natural landscape images is of great significance in many fields such as computer vision, environmental protection, and so on. It not only records and reflects the diversity of nature, but also provides important data support for environmental detection, ecological research, and so on. However, for the diversity and complexity of seasonal changes, weather conditions, lighting differences, and other issues, it is necessary to optimize the natural landscape images by zoning. This not only enables the detection of changes in complex natural environments but also helps scientists understand the impact of climate change on ecosystems.

Image classification and image enhancement are two techniques to optimize the segmentation of natural landscape images. Image classification refers to classifying natural landscape images in different situations, such as according to different light. Image enhancement refers to adding elements such as color to the image to facilitate subsequent cutting and partitioning optimization. In recent years, with the development of deep learning technology, remarkable progress has been made in the analysis and processing of natural landscape images. Initially based on feature engineering methods and then machine learning classifiers, Zhang proposed an image classification method that combines a support vector machine (SVM) with a K-nearest neighbor algorithm (KNN). When tested on the Caltech-101 dataset, the classification accuracy of the proposed method reached 87.5%, which was 15% higher than the traditional KNN method (Zhang et al., 2006). Ye showed that the method based on matrix decomposition and low-rank representation performed well in dimensionality reduction tasks for MNIST datasets, and the dimensionality reduction error of the Principal Component Analysis (PCA) method was reduced by 12%, which further proved its effectiveness in machine learning (Ye, 2005). Jayalakshmi and Babu, in their study, demonstrated the image generation capability of generative adversarial networks (GANs) on CIFAR-10 datasets, and the resulting images outperformed convolutional neural network (CNN) -based methods in visual

[a] https://orcid.org/0009-0000-3128-194X

Yin and Q.
Optimization of Natural Landscape Images Using CNN and Improved U-Net Technology.
DOI: 10.5220/0013516700004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 334-339
ISBN: 978-989-758-754-2

quality, With an 8% improvement in F1 scores (Jayalakshmi, Babu, 2016). The CNNS proposed by Krizhevsky performs well on the ImageNet dataset, with a classification accuracy of 84.7%, significantly higher than traditional image classification methods (Krizhevsky et al., 2012).

These studies show that the methods based on SVM, KNN, PCA, GAN, and CNN models not only shorten the processing time but also significantly improve the accuracy of image classification and generation, effectively overcoming the limitations of traditional methods in computational complexity and performance.

However, there are still some problems in practice, such as complicated backgrounds and multi-angle photos, which cannot be accurately positioned. Therefore, picture classification and picture partition are very important for picture optimization. This paper summarizes the main applications of CNN and U-Net technology in image classification and segmentation discusses the current challenges, and looks forward to future development. This paper aims to provide a theoretical basis for the development of this field.

2 METHOD REVIEW

2.1 CNN

2.1.1 Adaptive Image Enhancement Algorithms

Based on the original image classification, the adaptive image enhancement algorithm adds the scope judgment of image brightness overflow value and image enhancement processing. It provides a good environment for further ethnic clothing image classification and improves classification accuracy. The basic flow of the algorithm is shown in Figure 1.

2.1.2 Image Classification Based on Convolutional Neural Networks

The convolutional layer is designed to extract the data features of the input image. Its convolution operation is shown in formula (1).

$$y_i = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_i \cdot x_i \quad (1)$$

Where y_i represents the result after convolution, w_i represents the parameters of the convolution kernel, and x_i represents the pixel value of the original image. CNN maps the process of input image convolution into the neural network, as shown in formula (2).

$$x_i^l = f\left(\sum_{i \in M_j} x_i^{l-1} \cdot w_{ij}^l + b_j^l\right) \quad (2)$$

Where w_{ij}^l represents the convolution kernel corresponding to the j feature graph of layers. b_j^l represents the offset term of the output feature graph, x_i^{l-1} represents the i feature graph of layer $l-1$, and f represents the activation function.

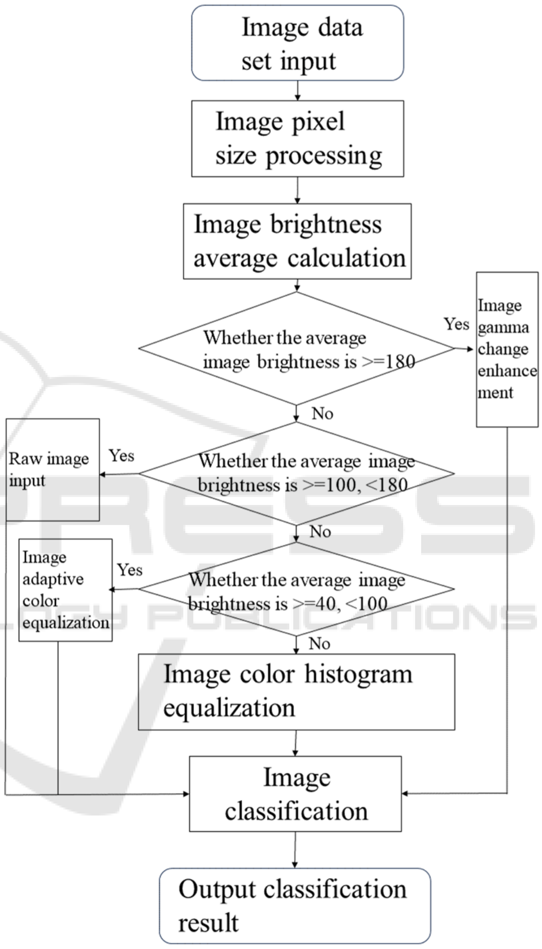

Figure 1: Flow chart of adaptive image enhancement algorithm (Hou et al., 2022)

Where represents the convolution kernel corresponding to the first feature map of the first layer, represents the offset term of the output feature map, represents the first feature map of the first layer, and represents the activation function.

The pooling layer is designed to reduce the data dimension and data volume while maintaining the statistical properties of image features and effectively avoiding overfitting phenomena. The calculation formula (3) of the pooling layer is shown below.

$$x_i^l = f\left(\beta_j^l down\left(x_i^{l-1}\right) + b_j^l\right) \quad (3)$$

Where, β_j^l represents the corresponding coefficient, b_j^l represents the bias term of the j feature map of the l layer. Down represents the sampling function

Input image data set, through the first convolution layer and call ReLU function activation, the output dimension is 28×28×20. After the first pooling layer processing, the output dimension is 14×14×20. After the second convolution layer and activation by calling the ReLU function, the output dimension is 10×10×40. After the second pooling layer processing, the output dimension is 5×5×40. The overfitting phenomenon is solved by two fully connected layers and calling Dropout. The final output layer consists of 5 features, representing the classification results of five types of natural scenery images respectively.

2.2 The Improved U-Net

For image segmentation, the sample picture has high background requirements when shooting, and the model effect is easy to be affected by shooting Angle, illumination, etc., and the edge contour is often unclear due to the complex background. Therefore, based on the U-Net model, residual module, multi-scale mechanism, and dual-channel attention mechanism are developed. Four aspects of the attention-oriented AG module and Dropout mechanism have been improved.

2.2.1 Residual Module

Since the U-Net network has few layers, the image information extracted is not enough, so I added the residual block to the original U-Net network and adopted a shortcut connection, which not only increased the network performance but also avoided the gradient disappearing.

2.2.2 Multi-Scale Mechanism

In practical projects, because the unmanned aerial vehicle will capture images at different scales, it brings certain difficulties to image segmentation. To obtain input images of different scales, firstly, a series of images with gradually reduced resolution are obtained after continuous downsampling of existing images to form an image pyramid. Secondly, the obtained multi-scale images are input into the constructed network to extract feature information under different scales. Finally, the Label chart under the corresponding size is used to deeply supervise the image segmentation results. This process is shown in Figure 2.

2.2.3 Two-Channel Attention Mechanism

Since it receives useful feature information while receiving useless feature information for the final segmentation target, the attention mechanism is set to enhance the segmentation performance of the model. There are two kinds of attention mechanisms, namely channel attention mechanism and the spatial attention mechanism.

The Channel Attention mechanism refers to extracting feature information from images through different convolution kernels after input. The design of the CAM module is shown in Figure 3.

$$CA = f(x, W) = \sigma\left(fc_2\left(\delta\left(fc_1(x, W_1)\right), W_2\right)\right) (4)$$

The above formula, W represents all parameters in the channel's attention module, σ represents the Sigmoid function, f represents the fully connected

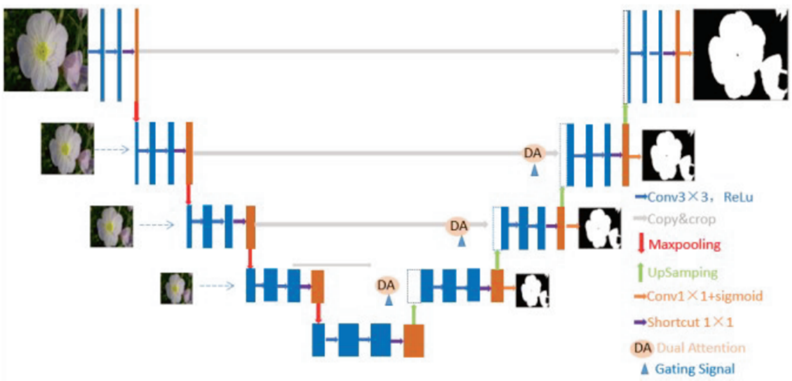

Figure 2: Improved U-Net segmentation model (Zhao et al., 2021)

layer, and c is the ReLu function. After obtaining the CA coefficient of the feature map of each layer, the final output feature value is obtained by weighting the feature map of each layer:

$$\widetilde{f_h} = CA \cdot x_h \tag{5}$$

The essence of the channel attention mechanism is to assign different weights to different features, to learn more useful feature channels for plant segmentation tasks.

The Spatial Attention mechanism is designed to clarify the boundary outline between the target area and the background, using two adjacent convolution nuclei, and finally using the Sigmoid function to map the final feature between [0,1]. The SAM module is shown in Figure 4.

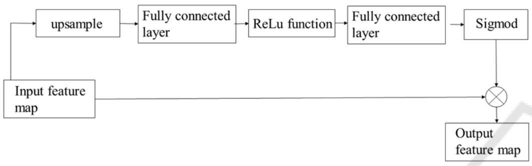

Figure 3: CAM structure (Zhao et al., 2021)

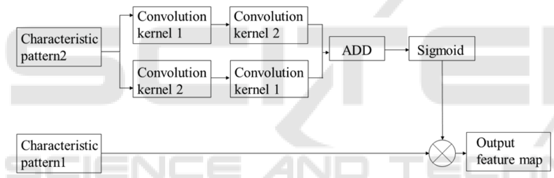

Figure 4: SAM structure (Zhao et al., 2021)

The Dual attention channel module is shown in Figure 5.

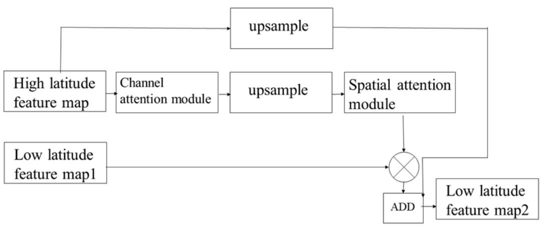

Figure 5: DAC structure (Zhao et al., 2021)

2.2.4 Attention-Directed AG Module and Dropout Mechanism

To further improve the feature extraction of images from different regions by the network model, the attention guide model (AG) is introduced, and its network architecture is shown in Figure 6.

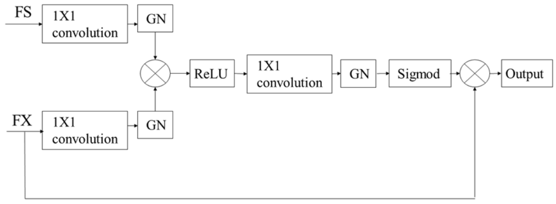

Figure 6: Schematic diagram of AG module (Chen et al., 2024)

At the same time, the multi-parameter dropout loss mechanism is also introduced in the upsampling part of the decoding area. That is, the lowest layer of the network downsampling Dropout parameter is 0.3, and the other four upsampling layers introduce two Dropout parameters of 0.2 and 0.1, respectively. In this way, the network model designed in this paper can avoid overfitting and improve the image segmentation accuracy of the network model.

3 DISCUSSION

3.1 Comparative Analysis

Hou selected 5,000 images of different female minority costumes, including Bai, Miao, Mongolian, Uyghur, and Tibetan, while randomly dividing the training set and test set in a 7:3 ratio to ensure the reliability of the data. The results of the comparative analysis of model performance are shown in Table 1.

Table 1: Comparison of KNN and CNN data (Hou et al., 2022)

	KNN			CNN		
Data set	Accuracy rate	Recall rate	F1 value	Accuracy	Recall rate	F1 value
White	0.5455	0.6667	0.6000	0.9355	0.8788	0.9063
Miao	0.5313	0.6296	0.5763	0.9063	0.8286	0.8657
Mongolian	0.5000	0.5385	0.5185	0.6970	0.8519	0.7667
Uyghurs	0.7241	0.6000	0.6563	0.9231	0.9231	0.9231
Tibetan	0.7143	0.5714	0.6349	0.9286	0.8966	0.9123

The overall error of CNN training on the dataset in this study gradually decreases and converges from the initial high value, and its error eventually converges to less than 0.0002. Table 1 shows the data processed by adaptive image enhancement and convolutional neural network algorithm, and the accuracy and other values are significantly improved. For example, compared with KNN, the accuracy rate of CNN is significantly improved, especially for Bai nationality.

The image of steel structure surface corrosion is taken as the data set. The image is from the Internet. There are 500 images of steel structure surface corrosion in the data set of rust images. Since there are few existing steel structure surface rust image datasets, to generalize the network to be further improved, the training set, training set, and test set are divided according to the ratio of 6:2:2, to reduce the possibility of overfitting of the model. On the other hand, Chen made the training set and verification more widespread than before after setting data set division by using eight data set expansion methods such as rotation and contrast enhancement, and did not expand the test set, to better evaluate the actual detection and segmentation accuracy of the network model. The results of the comparative analysis of model performance are shown in Table 2.

The improvement of the U-Net network model shows the better segmentation accuracy of corroded images to some extent, and the results of corroded image segmentation are more clear and coherent. The data pairs are shown in Table 2. In particular, the FLOP value is significantly reduced, which means that the calculation amount of the model is reduced and the model is more lightweight.

3.2 Limitations

3.2.1 Limitations of CNN and Suggestions for Improvement

Despite the use of multiple convolutional layers and pooling layers, the network structure is still relatively simple, and the accuracy may be insufficient for more complex image classification. In addition, the scalability of the model is limited because the input images are limited to the clothing images of the five ethnic groups. Some studies have shown that the lack of deeper layers or advanced techniques such as batch normalization and discarding of CNNS may lead to poor generalization, especially when the data sets are diverse or noisy (Lei et al., 2020).

In response to the above problems, this paper puts forward some suggestions to improve the structural complexity of CNNS. An image classification method that combines the advantages of CNN and Transformer can be adopted. One study points out that the spatial location attention mechanism is embedded in the channel when CNN extracts image feature information to extract the feature regions of interest in the image. Then, the Transformer encoder is used to assign large weights to regions of interest to focus on salient regions and salient features, thus improving the classification accuracy of the model (Jin et al., 2023).

3.2.2 Limitations and Improvement Suggestions of U-Net

Although U-Net has improved significantly in plant image segmentation, model structures such as multi-scale mechanism and dual-channel attention mechanism may lead to increased consumption of computing resources, and its training time is longer, which is not suitable for real-time or resource-constrained application scenarios. This has been proved in some studies (Su et al., 2023).

Given the above problems, this paper puts forward some suggestions to improve the efficiency of the U-Net model. The efficiency of the U-Net model can be optimized by introducing a lightweight and high-performance alien invasive plant recognition model (MobileNet-LW). In one study, for example, the citation (MobileNet-LW) model improved the judgment accuracy of alien invasive plant images to 86% (Wu et al., 2024). The image data was enhanced by rotation and Gaussian noise, etc., which improved the details of the image of invasive alien plants and increased the number of data sets.

Table 2: Comparison of corrosion segmentation performance of different network models (Chen et al., 2024)

Different net	Accuracy/%	Precision/%	IoU/%	Params/M	FLOPs/G	Time/ms
FCNs	93.82	83.26	72.66	18.64	19.52	103.56
UNet	90.50	81.92	68.00	17.27	30.77	169.91
Deeplabv3+	93.35	81.72	69.99	40.35	13.26	137.01
FAT_Net	93.36	83.71	82.47	29.62	42.80	186.00
Imporved U-Net	95.54	84.59	77.43	3.25	0.51	52.03

4 CONCLUSIONS

This paper introduces in detail the application and improvement of CNN and U-Net models in image classification and image segmentation. For image classification and other problems, the performance of the CNN model in ethnic clothing image classification has been significantly improved by introducing an adaptive image enhancement algorithm. For problems such as complex background and unclear edges, U-Net model improves by improving the residual module, multi-scale mechanism, dual-channel attention mechanism and Dropout mechanism. It has shown high precision and robustness in plant image segmentation. However, although these improvements effectively improve the performance of the model, there are still some limitations, such as the limitations of convolutional neural networks, and the increase in computing resource consumption caused by the complexity of the network structure.

To further improve the practicability and scope of application of the model, future research can be explored from the aspects of improving the complexity of CNN structure and optimizing the efficiency of the U-Net model. These improvements can not only improve the classification and segmentation performance of the existing model but also provide a more reliable solution for complex scenes in practical applications.

REFERENCES

Chen, F., Dong, H. and He, X., 2024. Improved U_Net network of the surface of the steel structure corrosion image segmentation method. Journal of Electronic Measurement and Instrument, 38(02), pp.49-57

Du, S., Jia, X. and Huang, Y., 2022. Design of efficient activation function for image classification of CNN model. Infrared and Laser Engineering, 51(03), pp.493-501.

Hou, H. T., Wang, W. and Shen, H. T., 2020. Research on image classification of ethnic costumes based on adaptive image enhancement and CNN. Modern Computer, 28(24), pp.29-35.

Jayalakshmi, D. S. and Ramesh Babu, D. R., 2016. A survey on image classification approaches and techniques. International Journal of Computer Science and Information Security, 14(3), pp.70-78.

Jin, C. and Tong, C., 2023. A remote sensing image classification method integrating CNN and Transformer structure. Advances in Laser and Optoelectronics, 60(20), pp.233-242.

Krizhevsky, A., Sutskever, I. and Hinton, G. E., 2012. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, pp.1097-1105.

Lei, F., Liu, X., Dai, Q., 2020. Shallow convolutional neural network for image classification. SN Applied Sciences, 2, 97.

Su, X., 2023. Dual Attention U-Net with feature infusion: Pushing the boundaries of multiclass defect segmentation. arXiv preprint arXiv:2312.14053.

Wu, H. F., Liu, W. X., Xian, X. Q., 2024. Research on lightweight alien invasive plant recognition model based on improved MobileNet. Plant Protection, 50(01), pp.85-96.

Ye, J., 2005. Generalized low rank approximations of matrices. Machine Learning, 61(1), pp.167-191.

Zhao, Y. J., Guo, X. L., Liu, Y., 2021. A plant image segmentation algorithm based on improved U-Net. Journal of Chinese Media University (Natural Science Edition), 28(3), pp.32-40.

Zhang, H., Berg, A. C., Maire, M. and Malik, J., 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2, pp.2126-2136.