ForestResNet: A Deep Learning Algorithm for Forest Image Classification

Yongqing Tang¹, Hao Feng¹*, Junyan Chen² and Yuan Chen³

¹ School of computer and information security, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
² School of computer and information security, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
³ School of computer and information security, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

*Corresponding author’s e-mail: fengh@guet.edu.cn

Abstract. Due to its large area and rugged terrain, the forest often fails to be detected in time and eventually causes severe losses[1]. Therefore, early detection of forest fires is significant for forest fire protection. The application of deep learning to the classification of smoke and fire in forest images can detect forest conditions more accurately. In this paper, a classification network, named ForestResNet, is proposed to efficiently detect forest conditions, which uses ResNet50[2] as a feature extraction network to achieve rapid and accurate extraction of image feature information. Experimental results show that the proposed network achieves excellent segmentation performance in terms of efficiency and accuracy.

1. Introduction
Forest fire is a kind of natural disaster that is easy to occur in the dry climate. Early detection of forest fire is essential for forest fire protection. In recent years, crewless aerial vehicles have been widely used for forest fire detection due to their high mobility and low cost. UAV fire detection mainly starts from a visual point of view and judges whether there is a fire by analyzing the color of the flame and the dynamic characteristics of the smoke [4,5]. However, the forest classification algorithm based on graphics cannot provide accurate and effective classification results for many forest images. Although the accuracy of machine learning is very high, it wastes a lot of resources and is inefficient. Therefore, it is urgent and necessary to study accurate and fast forest classification algorithms.

To achieve the accurate and effective classification of forest images, this paper first obtained 175 forest images from the Internet as an experimental data set. Based on the fire area characteristics presented by the pictures of the data set, a novel classification network ForestResNet was proposed. Experimental results show that this method is an effective automated forest classification tool.

2. Related work
Image classification is a task in computer vision. Its essence is to divide each pixel or area in an image or image into several categories based on the different characteristics of different categories reflected in the image information.

Forest image classification refers to the accurate classification of forest images based on different regional information (fire, smoke, and normal) presented in the forest image. Most traditional image
segmentation methods are complicated in operation and cannot be well adapted in many image fields. In recent years, the development of deep learning has been very rapid. Many experts and scholars have proved in practice that deep learning algorithms have significant advantages in image feature extraction. In 2015, Christian Szegedy et al. [6] proposed a more profound and broader GoogLeNet, which uses the Inception module to fuse the multi-scale feature information of the image under the same amount of calculation to achieve more efficient calculations[9]. Chen et al. [6] proposed a lightweight convolutional neural network MobileNet built using deep separable convolution. The deep separable convolution decomposes the standard convolution into the convolution of each channel and realizes the channel convolution. The product is combined with the point convolution, thus in the efficient image classification. However, although these two feature extraction methods are efficient, they cannot obtain accurate image feature information.

With better anti-fitting characteristics, it reduces vanishing gradients and dramatically reduces the number of parameters. Moreover, the related researches[8][10] showed that increasing the number of layers of a deep convolutional neural network can significantly improve the ability of extracting feature information. However, when the depth of the network increases to a certain number of layers, the network performance is saturated or even decreased. This problem is mainly caused by the gradient disappears during the training process, which is also called as the network degradation problem. In 2016, He et al. [8] proposed Residual Network (ResNet), which uses residual units to solve network degradation.

Inspired by the above-mentioned documents, ForestResNet uses ResNet50 as the feature extraction part so that the framework can extract deeper feature information of forest images. This paper uses 175 labeled forest images as the experimental data set for experiments. Experimental results show that this method can accurately and quickly classify forest images and can be used as an automated tool for forest fire detection.

3. Method

3.1. Data preprocessing
To better input the forest image data into the deep learning model for training, this paper converted the original forest image data into image data of the same size through a series of preprocessing operations. The data preprocessing was divided explicitly into the following five steps:

a. Randomly cropped the original forest image data to a size of 256×256 and rotated the image randomly at -15°~+15°.

b. Randomly flipped the cropped image horizontally with the default probability of 0.5.

c. Cut the image processed in step 2 to a size of 224*224 at the center.

d. Convert the cropped forest image with a size of 224×224 into a tensor of 0~1.

e. Normalize the tensor from 0 to 1 after conversion from -1 to 1.

3.2. Forest image classification algorithm
This paper used the 50-layer Residual Network (ResNet50) as the feature extraction part of the forest image classification algorithm. As shown in Figure 1, the residual unit used by the Residual Network consists of two convolution operations with a kernel size of 1×1, a convolution operation with a kernel size of 3×3, and a Shortcut Connection. Assuming that the mask of input feature was the x, the expected output is H(x), and its equivariant residual mapping was F(x) = H(x)-x. It was worth noting that the learning goal of the residual unit is no longer the complete output but the residual function F(x) between the output and the input. The original learning goal was H(x)=F(x)+x. When the input x was 0, the residual mechanism can achieve equivariant mapping, which can effectively avoid the degradation of deep convolutional neural networks. Moreover, the shortcut connection in the residual unit can enable the deep convolutional neural network to be solved by backpropagation at the existing depth without changing the size of the network or the number of parameters.
3.1. Residual unit structure.

The network of ResNet50 was shown in Figure 2, which consisted of 49 convolutional layers and a fully connected layer (FC layer). The 49 convolutional layers were divided into five groups. There were three kernel size convolutions: 7×7, 3×3, and 1×1, which extracted features of the input image. At the same time, the activation function of the convolutional layer was a nonlinear Relu activation function. Moreover, the ResNet50 network structure also uses a 7×7 average pooling and a 3×3 maximum pooling. The stride size of the pooling operation was set to 2, and each pooling operation would reduce its dimension to The original 1/2. Finally, the dimension of the fully connected layer of the ResNet50 network structure was set to 3, which was used to realize the classification of 3 categories of forest images.

3.3. Loss Function

The forest fire classification algorithm used the cross-entropy function as the loss function of the model. The specific Equation as follows Equation 1, supposed it is the fire category label; is the probability that the predicted category is the label of the current sample category; then it represented the probability that the predicted class was not the current sample category. From the perspective of maximum likelihood, the predicted class was The probability of the current label, and the probability of not being the current category label was integrated. Finally, it was introduced into the function. In the forest fire classification algorithm, the closer the predicted category was to the true sample label, the smaller the loss function; if the prediction output was incorrect, the loss function would become smaller.

\[
Loss = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})]
\]  

(1)

4. Experiment

4.1. Dataset

Since the forest fire data set has no public data set, the experiment obtained 175 forest images from the Internet as the experimental data set (150 training images, 25 test images). Among the 150 images in the forest image training set, there are 50 normal images, 50 smoke images, and 50 fire images, and their labels are set to 0, 1, and 2, respectively. The 25 images in the test set are three fire images randomly selected from the Internet. The experiment selects the best model through 25 random forest images chosen.

4.2. Implementation details

Deep learning is accompanied by many calculations, and the hardware configuration greatly affects floating-point numbers and matrix operations, so a powerful experimental platform is required. This experiment is run on the Nvidia 2080Ti RTX 24G graphics card of the Ubuntu 18.04 operating system, using the currently popular deep learning framework PyTorch to implement the deep learning model. During the experiment, we set the batch size to 4 and the iteration period to 140. When the loss function value is stable and does not continue to decrease, save the model.
4.3. Evaluation metrics
The experiment uses the accuracy rate as the evaluation index. This paper assumes two types of actual results: positive samples and negative samples, and the predicted results also have positive samples and negative samples. The classification results can be divided into the following four types, as shown in Table 1.

| predict result / label | Positive | Negative |
|------------------------|----------|----------|
| Positive               | TP       | FN       |
| Negative               | FN       | TN       |
| Total                  | P        | N        |

Specifically, the accuracy rate is defined as the proportion of the correct sample results among the sample results predicted by the model, as shown in equation (2).

$$ accuracy = \frac{TP + TN}{TP + TN + FP + FN} $$ (2)
4.4. Experiment analysis
The loss curve of the training set of the forest image classification algorithm is shown in Figure 4. It could be seen that when the number of iterations of the deep learning network is 140 rounds, the value of the loss function has stabilized and does not continue to decline, which indicates that the deep learning network has been point to the optimal solution.

The accuracy curve of the forest image classification algorithm is shown in Figure 5. The abscissa of the graph is the number of iterations, and the ordinate is the average accuracy. It could be seen that when the number of iterations of the deep learning network is 140 rounds, the forest image classification accuracy rate deliberately reaches more than 90%. In addition, this article also conducted experiments on the test set. Where the 25 forest images, 23 images were classified correctly, and the accuracy rate reached 92%.

5. Conclusion
Forest fire is a kind of natural disaster that was easy to occur in the dry climate. Real-time monitoring of forest conditions is an effective measure for forest fire protection. With the rapid development of deep learning, computer-aided forest protection had also made considerable progress. This paper
proposed a forest image classification network using ResNet50 as a feature extraction structure. Experimental results show that this network has achieved good results in the accuracy of forest image classification. Among them, the Accuracy evaluation index can reach 92%. The network can realize fast and accurate classification of forest image feature information to realize real-time monitoring of forest conditions. However, the diversity of the forest image data in the natural environment will bring challenges to the classification algorithm. Future research will further optimize the deep learning model for this series of problems.

Acknowledgment
This work was supported by the National Natural Science Foundation of China (No.61861013), Science and Technology Major Project of Guangxi (No.AA18118031), and Guangxi Natural Science Foundation (No.2018GXNSFAA281318).

References
[1] Ganesan P, Sathish B S, Sajiv G. A comparative approach of identification and segmentation of forest fire region in high resolution satellite images[C]//2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave). IEEE, 2016: 1-6.
[2] Chen H Y, Su C Y. An enhanced hybrid MobileNet[C]//2018 9th International Conference on Awareness Science and Technology (iCAST). IEEE, 2018: 308-312.
[3] Hu Y, Wen G, Luo M, et al. Competitive inner-imaging squeeze and excitation for residual network[J]. arXiv preprint arXiv:1807.08920, 2018.
[4] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[5] Yuan C, Liu Z, Zhang Y. UAV-based forest fire detection and tracking using image processing techniques[C]//2015 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2015: 639-643.
[6] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
[7] Chen H Y, Su C Y. An enhanced hybrid MobileNet[C]//2018 9th International Conference on Awareness Science and Technology (iCAST). IEEE, 2018: 308-312.
[8] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//European conference on computer vision. Springer, Cham, 2016: 630-645.
[9] Qiu L, Geng J, Zhang Y, et al. 3D EMSU-Net: A framework for automatic segmentation of brain tumors[C]//2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2021: 1049-1053.
[10] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016: 424-432.
[11] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.