Gully Identification of Debris Flow Disaster Based on Knowledge Distillation

Kunxiang LIU[a] and Baoyun WANG[b,1]

[a] *School of Information, Yunnan Normal University, Yunnan, China*
[b] *School of Mathematics, Yunnan Normal University, Yunnan, China*

Abstract. Convolutional neural networks have been applied in the field of remote sensing image classification. For convolutional neural networks with shallower layers and simpler structures, the accuracy of the recognition of debris flow gully images is not ideal, while the number of layers is deeper and the structure is relatively simple. More complex neural networks often consume a lot of system resources and are difficult to deploy on the user side. Aiming at this problem, an optimized convolutional neural network method is proposed. First, through multiple sets of comparative experiments, select Resnet101 and Resnet18 models with good image classification performance; then, use the characteristics of debris flow gully images to pre-train the deeper and more complex Resnet101 model; finally, through the method of knowledge distillation, The trained "knowledge" is extracted into the Resnet18 model to achieve the effect of improving accuracy while reducing system resource occupation. Experimental data shows that after using knowledge distillation, the accuracy and sensitivity of the Resnet18 model are increased by 2.36 and 1.72 percentage points, respectively, and the image processor occupancy is reduced by 37 percentage points compared with the Resnet101 model.

Keywords. Resnet101, Resnet18, knowledge distillation, DEM data, debris flow disaster.

1. Introduction

Debris flow disasters occur frequently in my country, especially in Yunnan, which is mostly mountainous [1]. In recent years, many scholars have applied neural networks to the analysis, prediction, and classification of debris flow disasters. For example: Debris flow risk assessment combining T-S fuzzy model and neural network [2]; Debris flow prediction based on neural network [3]; BP neural network analysis of economic losses caused by debris flow disasters [4].

Among the numerous researches on neural networks, the classification of images is more widely used. Many researchers use existing neural networks to improve classification results. Zhao Jingxia and others applied the improved LeNet-5 network to the diagnosis of breast diseases [5]. Shi Cuiping et al. based on the improved AlexNet neural network to recognize facial expressions [6]. Wu Siyu et al. achieved efficient classification of flowers through an improved VGGNet model [7]. In terms of prediction and classification of geological disasters, neural networks have also been

---

[1] Corresponding Author, Baoyun WANG, School of Mathematics, Yunnan Normal University, China;
Email: wspbmly@163.com.

widely used. Some researchers have applied them to earthquakes [8-10], landslides [11], floods [12-13], debris flows [14-17] and other geological disasters. Prediction and analysis. Du Xing [18] et al. based on MPL neural network to evaluate and predict sand liquefaction under earthquake, the accuracy rate can reach 96%; Liu Shiyang [19] et al. used convolutional neural network and time series prediction method to study tunnel boring machine driving The possibility of landslide section; Yu Guoqiang [20] and others used BP neural network and support vector machine model to predict Jiangjiagou debris flow data, and analyzed the sensitivity of each influencing factor of average velocity, and established the average velocity of debris flow Sensitivity factor prediction model. Although the performance of modern deep neural networks has been significantly improved, it is still necessary to extract effective information from huge and redundant data during model training. Under normal circumstances, the real-time requirements are not considered, and the model parameters obtained in the final training are large. Due to the limitations of computing resources and latency, deep neural networks face many obstacles in practical applications.

Hinton et al. first proposed the idea of knowledge distillation [21]. The core idea is to use a large and accurate teacher model to guide a small and fast student model. Based on this idea, this research attempts to select a teacher model and a student model suitable for debris flow gully data from the 6 commonly used convolutional neural networks. Through knowledge distillation, the model obtained can reduce the parameter amount of the network model and the model running time while ensuring the accuracy. On the basis of sacrificing a little accuracy, the cost of deploying the model can be greatly reduced.

2. Research Methods and Principles

2.1. Knowledge Distillation Model

With the rapid development of neural networks, the number of layers of the network is getting higher and higher, and the parameters that need to be trained are getting larger and larger. This has higher requirements for computer hardware. Therefore, Hinton et al. proposed the idea of knowledge distillation in this context.Knowledge distillation is inspired by teachers and students, and guides a streamlined, low-performance student network learning by using a complex, large-scale but high-performance teacher network. Through training, the student network can also achieve performance equivalent to that of the teacher network. The specific flow chart of the knowledge distillation model is shown in figure 1.
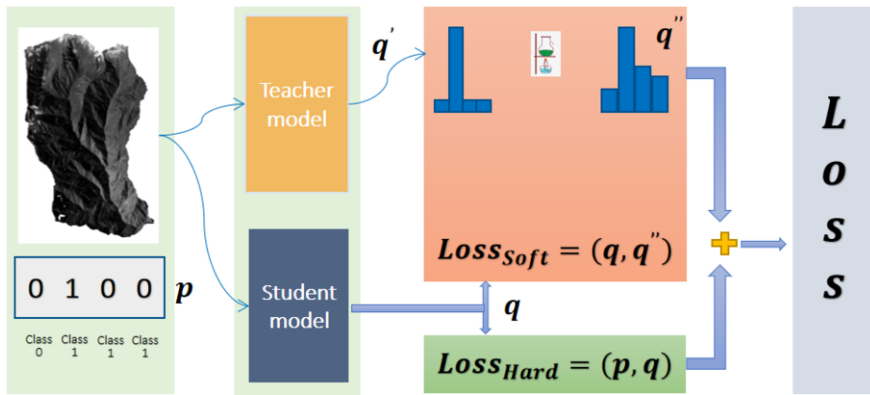
Figure 1. Flow chart of knowledge distillation

(1) Input samples with hard label p into the teacher network and student network for training at the same time;

(2) Train the teacher network and use the hard label p to get the output result q';

(3) In the teacher network, q' is "distilled" through the softmax function with temperature T to obtain a smoother classification probability value q'', which is used as the soft label of the student network;

(4) Train the student network, the output result is q, and calculate the loss of hard label p and output result q as hard label loss ($Loss_{hard}$);

(5) Calculate the loss of the student network output q and the soft label q'' as the soft label loss ($Loss_{soft}$);

(6) The weighted sum of hard label loss and soft label loss is taken as the loss of the entire network model of knowledge distillation, and the student network is trained by optimizing the loss function;

(7) Finally use the trained student network to make predictions.

2.2. Data Source and Processing

This paper is mainly to classify the mud-rock flow gully data set of Yunnan Province that I built. In order to verify the reliability of the model and experiment, the public data set RSSCN7 is trained and tested before the model is determined. Here is a brief introduction to these two data sets.

　　1) RSSCN7 DataSet remote sensing image data set

　　RSSCN7 Dataset is the remote sensing data produced by Wuhan University in 2015. The data contains 2800 remote sensing images, each with a pixel size of 400*400, and a total of 7 categories. These images come from 7 typical scene categories—grassland, forest, farmland, parking lot, residential area, industrial area, and river and lake. Each category contains 400 images, which are sampled on 4 different scales. In this dataset, the diversity of scene images makes it more challenging. These images come from different seasons and weather changes and are sampled at different scales. Figure 2 shows some pictures of RSSCN7 Dataset.

| Grass | Field | Industry | RiverLake |

| Forest | Resident | Parking |

Figure 2. RSSCN7 data set display

2) Debris flow data set

The second data in this article is the valleys and valleys where debris flow disasters occurred during 2005-2017 in Yunnan Province and some valleys that did not occur. The data is obtained through ArcGIS software processing of DEM images and Gaofen No. 1 remote sensing data. A total of 560 pieces of original data were obtained. This data set mainly contains two kinds of data, one is DEM data, the other is Gaofen-1 remote sensing data, among which Gaofen-1 remote sensing data is composed of 4 bands (red light band, green light band, blue light band) And near infrared bands). The purpose of this experiment is to efficiently identify the probability of a debris flow occurring in a certain place through intelligent means. Therefore, this paper classifies the extracted debris flow valleys according to the number of disasters, and divides them into category 0, category 1, category 2, and category 3. Type 0 means that no debris flow has been recorded in the valley, Type 1 means that 1 debris flow has been recorded in the valley, Type 2 means that there have been 2 debris flows recorded in the valley, and Type 3 means that there have been 3 times recorded in the valley. Above the mudslide. Figure 3 shows the 0, 1, 2, 3 data of DEM, red band, green band, blue band and near-infrared wave.
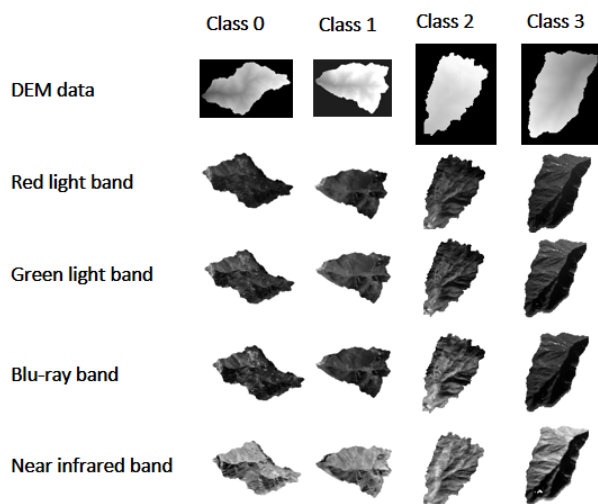
Figure 3. Debris flow data set display

In the training of the deep learning model, a large number of data sets are needed to prevent overfitting and obtain a network with strong generalization ability. However, in debris flow research, it is difficult to obtain a large number of complex and labeled remote sensing images of debris flow valleys as training data sets. Data enhancement can increase the amount of data and improve the overall performance of the neural network model. Therefore, the method of data enhancement is adopted to perform horizontal and vertical flipping, random angle rotation, random color jitter, random noise, and contrast, brightness, and color enhancement of remote sensing images of debris flow valleys to solve the problem of insufficient data. After data enhancement, a total of 1,240 debris flow valley images were obtained. The number of remote sensing images of the four types of debris flow valleys were 300, 290, 320, and 330, according to the 6:2:2 ratio of the training set, validation set and test set. Divide the data set.

2.3. Debris Flow Gully Classification Model Design

Aiming at the task characteristics, this article first compares the currently commonly used 6 network models AlexNet[22], VGG16[23], Googlenet, Resnet18, Resnet50, and Resnet10. Table 1 shows the comparison of their performance indicators such as the number of network layers, model memory, and the amount of parameters.

Table 1. Performance comparison of each model

Model	Depth layer	memory/MB	Parameter/10^6	Calculation amount/10^6	top-1 err	top-5 err
AlexNet	8	219	62.3	720	36.7	15.3
VGG16	16	512	138.3	15300	24.4	7.1
Googlenet	22	46	6.8	1550	29	9.2
Resnet18	18	45	33	1800	27.88	13.9

| Resnet50 | 50 | 99 | 46 | 3800 | 20.74 | 5.25 |
| Resnet101 | 101 | 163 | 85 | 7600 | 19.87 | 4.6 |

It can be seen that the Resnet101 model has a greater advantage than other models in evaluating the Top-1 error rate of the parameter of the recognition accuracy index. Its error rate is reduced by nearly 1/3 compared with the AlexNet model, and it also has different degrees of decline compared with other models. Although the accuracy rate of the Resnet101 model ranks first, the model size and calculation amount are second only to VGG16, which is very unfriendly to the deployment of the model. The purpose of knowledge distillation is to use a teacher model with larger memory and better performance to teach a small model with poor performance, so that the small model can be as close as possible to the learning ability of the teacher model, and finally a small memory and Model with better classification effect. From these six models, the choice of teacher and student network cannot be determined subjectively. It can be seen from the table that the Alexnet model not only occupies a larger memory and ranks second, but also has the worst classification performance, so it is not considered. In addition, although the VGG16 model performs well in classification performance and ranks third, it is not the best choice due to the large memory of the model. After excluding these two models, choose from the remaining four models.

2.4. Determine the Teacher and Student Model

In order to determine the best teacher-student group sum, this paper conducts pair-wise training on the remaining four network models. In order to ensure the reliability of the experiment, the comparative experiment is based on the RSSCN7 data set. The experimental results are shown in table 2.

Table 2. Comparison of knowledge distillation of different networks

Student model	Resnet18		Googlenet	
Teacher model	Resnet50	Resnet101	Resnet50	Resnet101
memory/MB	42.7	42.7	45.9	45.9
Time	16m53s	17m1s	17m	17m19s
Parameter/10^6	11.18	11.18	11.99	11.99
ACC/%	82.54%	84.92%	81.75%	82.14%

It can be seen from table 2 that the teacher network after knowledge distillation retains the model size of the original network. By distilling the knowledge of the teacher network, the student network can also achieve a higher accuracy rate. The initial accuracy of the student network Resnet18 was 81.25%. Through the learning guidance of the Resnet101 teacher network with an accuracy of 85.25%, the final test

result can also reach 84.92%, an increase of 3.73% on the original basis. Although compared with the teacher network, the distilled student network is still inferior to the teacher network, but the distilled student network has a 3.67% improvement on its original basis, and the effect is significant, which further demonstrates the feasibility of knowledge distillation. According to the comparison of knowledge distillation between the two student networks of Resnet18 and Googlenet and the two teacher networks of Resnet50 and Resnet101, the combination of Resnet18 and Resnet101 is significantly better than the other three groups of knowledge distillation models in terms of model size and test accuracy. Therefore, the final teacher and student models in this article are Resnet18 and Resnet101. The model after knowledge distillation is named kd_101-18 model.

3. Experimental Design

3.1. Experimental Platform

In this experiment, the learning rate of the Resnet101 model is set to 0.001, the learning rate decay rate is 0.98, and the weight decay rate is 0.00004. The classification function is a softmax function, the loss function uses a cross entropy function (Cross Entropy Loss), and the optimization method uses a stochastic gradient descent (SGD) algorithm. Although the commonly used optimization method is Adam, it also has the problem of too fast convergence to reach the optimal solution. Through comparison in this experiment, it is found that using SGD can achieve better results. In the training process of the Resnet18 model, the learning rate is set to 0.0005, the decay rate is set to 0.98, and the weight decay rate is 0.00004. The loss function uses the cross entropy function, the classification function uses softmax, and the optimization method also uses the stochastic gradient descent algorithm.

3.2. Evaluation Index

In multi-classification tasks, commonly used evaluation indicators include accuracy, precision, recall, F1 value, ROC + AUC, etc. In order to evaluate the classification task of debris flow gully, this paper uses the above five evaluation indicators to analyze the knowledge distillation model. Among them, the accuracy rate, precision rate, recall rate, and F1 value are all calculated based on the confusion matrix. Therefore, the confusion matrix of the test data needs to be obtained before calculating these indicators.

4. Experimental Results and Analysis

4.1. Classification Results of Each Channel

The Gaofen-1 remote sensing data used in this paper has four bands of red, green, blue and near-infrared. The red light band is mainly used to measure the chlorophyll absorption rate of plants and to classify vegetation. In areas where cities and vegetation are mixed, buildings can be well distinguished from vegetation. The green light band is mainly used to detect the green reflectivity of healthy plants and reflect underwater

characteristics. Many information characteristics of the water body are well reflected. The blue light band can obtain the boundary information of the intersection of the ground features, and the boundary information of the intersection of the ground features can be obtained very refreshingly from the image. The near-infrared band is used to distinguish between land and water lines and crop distribution areas.

From the above, it can be seen that different bands correspond to different characteristic information. Based on the knowledge distillation model, the DEM data and the debris flow valley data of the four bands are separately trained, and the influence of different data and channels on the classification of debris flow valleys is compared and analyzed. Experimental results as shown in table 3.

Table 3. Classification results of each channel

	Class 0	Class 1	Class 2	Class 3	Overall acc
Red light band	69.15	68.87	70.32	71.00	69.835
Green light band	67.90	69.08	70.24	70.13	69.34
Blu-ray band	71.06	70.50	72.47	71.18	71.30
Near infrared band	73.3	72.41	73.44	71.21	72.58
DEM data	72.45	70.09	71.66	71.07	71.32

Through the analysis of various experimental results of the four sets of data, it is found that the accuracy of various classifications is between 67% and 74%. The classification results of each type of each waveband are red, blue, and green. The classification results are better; the classification results of the 0th and 2nd classes in the near-infrared band and DEM are better. Overall, the near-infrared band has the best classification effect, reaching 72.58%, followed by the DEM map and the green band, at about 71.3%, and the red and blue bands are poor, less than 70%. The experimental results show that the near-infrared band in remote sensing images has the best classification effect on debris flow valleys, indicating that there is richer geographic information in the near-infrared band, so that the network model can learn more category features, thereby improving the classification performance. In order to make the results of the experiment more referential, in the subsequent classification experiments, the near-infrared band of the remote sensing image is used as the training and test data.

4.2. The Influence of Different Temperatures

The temperature parameter T plays a key role in knowledge distillation. Temperature determines the effect of knowledge distillation and directly affects the classification results. table 4 shows the effects of different temperatures on the experimental results. Here, 1, 10, 20, 30, 40, and 50 are selected as the experimental parameters. When T=1, it means that no knowledge distillation occurs, so the result is poor; when T is 40, 50, the accuracy rate begins to decrease, indicating that the temperature should not be set too high; when the temperature is 10, 20, and 30, the distillation effect is better. The experiment finally chooses T=10, which has the lowest loss, and performance indicators such as accuracy, specificity and sensitivity are also higher than the results at

other temperatures.

Table 4. Comparison of results with different temperature values

Parameter_T	Acc/%	Loss	Sensitivity/%	Specificity/%
1	70.97	0.0239	70.99	89.89
10	72.58	0.0228	72.59	91.10
20	71.37	0.0235	71.35	90.46
30	72.17	0.0235	72.38	90.97
40	70.56	0.0249	71.19	90.83
50	70.97	0.0258	71.04	90.31

When training the student model Resnet18, in addition to the normal training result-the hard target, the same method as above was used to obtain a soft target. Combine the two when calculating the loss function, the calculation formula is as follows.

$$L = \alpha L_{soft} + \beta L_{hard}$$

$$Loss_{soft} = -\sum_{N}^{i} p_i^T log(q_i^T)$$

$$Loss_{hard} = -\sum_{N}^{i} c^i log(q_i^1)$$

Among them: L is the final loss used by the training student model, and respectively represent the proportion of soft label loss and hard label loss in the total loss, where the sum is 1;, respectively represent the teacher and student network under the condition of temperature T = T Softmax outputs the value on the i-th category;, represents the logical output of the teacher and student network output respectively; is the real label on the i-th category, and N is the total number of labels.
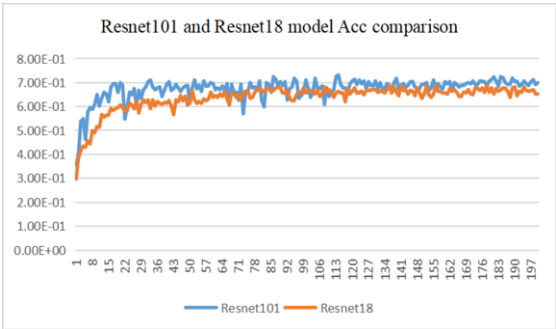
4.3. Result Analysis

Table 5 shows the comparison between the kd_101-18 model after knowledge distillation and the Resnet18 and Resnet101 models. Compared with the unoptimized Resnet18 model, the kd_101-18 model has an improved index accuracy rate of 2.26% compared with the latter, and the training time has been reduced by 3m11s. In the process of identification and classification of debris flow disasters, timely detection and timely preventive measures are vital to the safety of nearby villagers and national property. The higher the sensitivity, the lower the error rate of the model for determining debris flow. The sensitivity of the kd_101-18 model is 1.72% higher than the Resnet18 model and 0.99% higher than the Resnet101 model. In terms of model deployment, compared with the teacher model Resnet101, the kd_101-18 model is reduced to 1/4 of the original model memory, which enables the kd_101-18 model to be more flexibly deployed on computers with different sizes of video memory; at the same
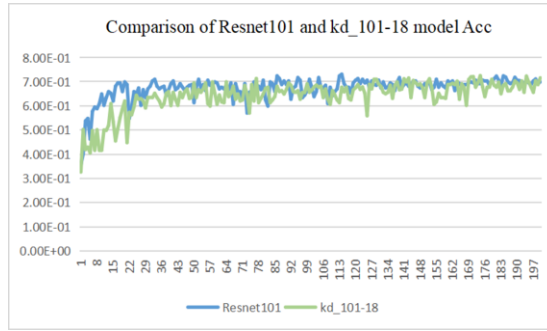
time; , The GPU utilization rate has dropped by 37%, so that the kd_101-18 model can be applied to computers with weak GPU computing power, and the portability of neural networks is greatly improved. Figure 5 shows the accuracy comparison curve between the three models of Resnet101, Resnet18 and kd_101-18. From Figure 5(a), we can see that there is still a certain difference in accuracy performance between the teacher network and the student network; Figure 5(b) is the comparison between the distilled kd_101-18 model and the teacher model. Although the distilled model is still Not as good as the teacher model, but it can be seen that the optimized model is as close to the teacher model as possible, which verifies the reliability of the knowledge distillation method; Figure 5(b) shows the comparison between the kd_101-18 model and the student model. The kd_101-18 model always stays on top of the teacher model, indicating that the kd_101-18 model has been greatly improved on the basis of the teacher model. Based on the experimental results, it can be seen that the knowledge distillation method can effectively enhance the recognition ability of the student model and improve its classification performance.

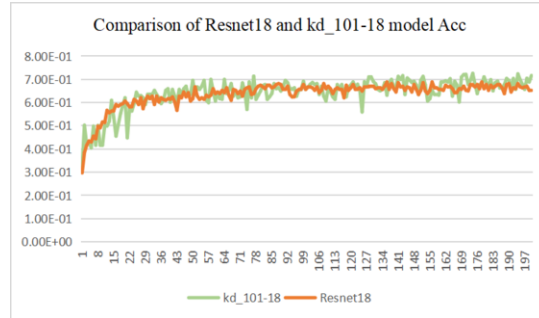**Table 5.** Comparison of models before and after distillation

Model	memory/M	GPU Occupy/%	Time	Sensitivity/%	Acc/%
Resnet101	160	88	16m30s	70.61	72.77
Resnet18	45	51	13m6s	69.88	70.32
Kd_101-18	45	51	10m17s	71.6	72.58

(a)ACC curve of Resnet101 and Resnet18 models

(b)ACC curve of Resnet101 and kd_101-18 model

(c)ACC curve of Resnet18 and kd_101-18 model

Figure 4. Accuracy comparison curve of Resnet101, Resnet18 and kd_101-18 models

4.4. Confusion Matrix

Figure 5 shows the 6 mixing matrices obtained by the experiment of model kd_101-18 at different temperatures, where the horizontal axis represents the predicted value, and the vertical axis represents the value of the real label. Through comparison, it is found that 2 of these 6 groups of experiments show that the test results of the second and third types are the best, the 0th type is the second, and the 1st type is the worst; there are 3 groups of the 2nd and 3rd types with the best results; There are 5 groups of experiments that show the worst results of the first category, only when T=20, the first category predicts two more correct numbers than the second category. It can be seen that the model has a higher recognition rate for categories 0, 2, and 3, and a slightly worse recognition rate for category 1. In addition, the debris flow gully data is classified according to the statistics of debris flow disasters. The more the number of debris flow disasters, the more obvious the characteristics of the gully. The results of the six confusion matrices in figure 4 also verify this point and further prove it. The effectiveness of the classification model. In the results of the second and third types, the second type is slightly better than the third type. It may be that in the data collection, the second type of raw data is more than the third type of data, so the second type of data is more informative in the model training. sensitive.
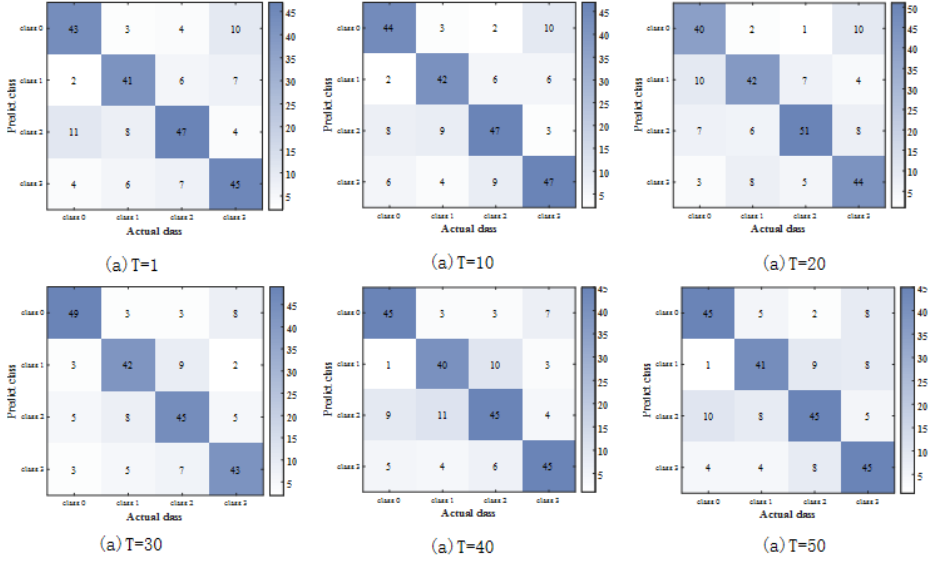
Figure 5. Confusion matrix at different temperatures

Through comparative experiments at different temperatures, it is concluded that when T=10, the distillation effect is the best. Therefore, the analysis of various indicators in this section is based on the model results of T=10. Next, calculate the accuracy, recall, precision, F1 value and ROC curve according to the confusion matrix of T=10 in figure 5.

1) Accuracy

$$accuacy_{T=10} = \frac{T00 + T11 + T22 + T33}{N} = 72.58\%$$

2) Recall rate

$$Recall_{T=10} = \frac{R_0 + R_1 + R_2 + R_3}{4} = \frac{73.3 + 72.41 + 73.44 + 71.21}{4} = 72.59\%$$
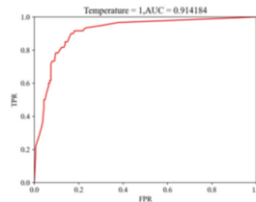
3) Accuracy

$$Precision_{T=10} = \frac{P_0 + P_1 + P_2 + P_3}{4} = \frac{76.67 + 72.41 + 70.15 + 71.21}{4} = 72.51\%$$

4) F1 value

$$F1_{T=10} = \frac{2P*R}{P+R} = \frac{2*72.59*72.51}{72.59+72.51} = 0.7255$$

5) ROC curve

When T=10, the total data of the test set is 248, of which there are 60 in category 0, 58 in category 1, 64 in category 2, and 66 in category 3. This is also the number of true labels of various types. The confusion matrix shows that there are 44 correct predictions for type 0, 42 correct predictions for type 1, 47 correct predictions for type 2, and 47 correct predictions for type 3. The correct numbers of the four categories are not much different. There are more types 2 and 3 in the real label, so the test is relatively more correct. The accuracy rate obtained by calculation is 72.58%, which means that the probability of correct prediction in all test samples is 72.58%; the recall rate is 72.59%, which means that the probability of correct prediction in all positive labels is 72.58%; the accuracy is 72.51 %. The F1 value is the harmonic mean value of the precision rate and the recall rate, which can avoid the extreme case that the precision rate or the recall rate is 1 and the other is 0. The F1 value is calculated to be 0.7255. It can be seen from the ROC curve that the curve is located at the upper left, a small distance is close to the (0, 1) point, and the AUC value is 0.9142. It can be seen that the classification model has a balanced comparison of various classification effects, and this model can be further improved.

5. Conclusion

Aiming at the characteristics of remote sensing image classification, this paper proposes a method of using convolutional neural network to realize the classification of mud-rock flow valleys in Yunnan Province. Using two neural network models with different structures and depths, Resnet101 and Resnet18, the optimized model kd_101-18 is obtained using the knowledge distillation method. While the classification performance is close to the Resnet101 model, the model's demand for system resources is greatly reduced. In the future work, while further improving the performance of the teacher model, reducing the demand for system resources of the student model can be the direction of future research.

References

[1]   Sun Xianchen, Wang Baoyun, Liu Kunxiang, Peng Juan, Kong Yan, Zhang Zhuhong. Analysis of the impact factors of debris flow disasters in Yunnan Province[J/OL]. People's Yangtze River: 1-10[2020-11-23].
[2]   Cao Lulai, Xu Linrong, Chen Shuyang, Jie Xiaofeng. Debris flow risk assessment based on fuzzy neural network [J]. Hydrogeology and Engineering Geology, 2014, 41(02): 143-147.
[3]   Wang Qian, Li Guangjie, Zheng Baigong, Yang Lijian. Debris flow prediction in Panshi City based on artificial neural network [J]. People's Yangtze River, 2009, 40(03): 46-48.
[4]   Ji Xiaoling. Rough BP neural network analysis of economic losses from landslide and debris flow disasters [D]. Southwestern University of Finance and Economics, 2012.

[5]   Zhao Jingxia, Qian Yurong, Zhang Meng, Du Jiao. Breast disease diagnosis method based on improved convolutional neural network LeNet-5 [J]. Journal of Northeastern Normal University (Natural Science Edition), 2019, 51(02)): 65-70.

[6]   Shi Cuiping, Tan Cong, Zuo Jiang, Zhao Kexin. Facial expression recognition based on improved AlexNet convolutional neural network[J]. Telecommunications Technology, 2020, 60(09): 1005-1012.

[7]   Wu Siyu, Feng Ji. Flower recognition based on improved VGGNet convolutional neural network [J]. Journal of Chongqing Normal University (Natural Science Edition), 2020, 37(04): 124-131.

[8]   Lv Junchao, Pan Jianping, Yu Shexin. Prediction of liquefaction side shift of gentle slope site based on cyclic neural network[J]. China Work Safety Science and Technology, 2021, 17(08): 18-23.

[9]   Tian Yajun, Gao Jinghuai, Wang Daxing, Chen Daoyu. Seismic strong reflection stripping method based on deep neural network [J]. Chinese Journal of Geophysics, 2021, 64(08): 2780-2794.

[10] Liu Zhiyuan, Yang Chunfeng, Meng Xiangcui, Jin Yunyun, Zhuang Yipeng, Chang Wenxin. Reliability analysis of artificial neural network waveform classification for oil and gas favorable zone division——Taking Xunyi Upper Paleozoic as an example[J]. Petroleum Geology and Engineering, 2021,35(01):34-39+46.

[11] Liu Shiyang, Chen Zuyu, Zhang Yunyang, Li Xu, Zhao Shengjie. Inversion analysis of TBM collapse section based on convolutional neural network[J]. Chinese Journal of Solid Mechanics, 2021, 42(03): 287-301.

[12] Zhang Ke, Niu Jiefan, Li Xi, Chao Lijun. Application comparison of intelligent flood forecasting models in semi-arid and semi-humid regions of China[J]. Water Resources Protection, 2021, 37(01): 28-35+60.

[13] Li Guyuan, Luo Yihua, Li Huanming, Gong Yuming, Zhou Ying, Lin Xinyi. Wenshanli flood forecast model based on BP neural network[J]. Henan Water Conservancy and South-to-North Water Diversion, 2020, 49(08): 28-29.

[14] Ma Jiehua, Sun Jianqi, Wang Jun, Yu Entao, He Shuangshuang. Prediction of extreme precipitation and landslide and debris flow disasters in my country in the summer of 2018[J]. Journal of Atmospheric Sciences, 2019, 42(01): 93-99.

[15] Zhu Wei, Cai He, Tang Wen, Feng Yulin. Study on the distribution characteristics and disasters of volcanic debris flow at Tianchi Volcano in Changbai Mountain[J]. Geology and Resources, 2017, 26(06): 608-615.

[16] Zhu Junxing, Liu Rui. Analysis of the cause mechanism of landslide and debris flow in dumping site and determination of critical rainfall [J]. Non-ferrous metals (mining part), 2017, 69(05): 75-79+91.

[17] Zhang He. Debris flow trend forecasting in Xiuyan County[J]. Agricultural Science and Technology and Equipment, 2017(03): 43-44.

[18] Du Xing, Sun Yongfu, Song Yupeng, Song Binghui, Zhao Xiaolong, Song Shasha, Wang Yue. Evaluation and prediction of sand liquefaction under earthquake action based on MPL neural network[J]. Journal of Engineering Geology, 2020, 28(06): 1425-1432.

[19] Liu Shiyang, Chen Zuyu, Zhang Yunyang, Li Xu, Zhao Shengjie. Inversion analysis of TBM collapse section based on convolutional neural network[J]. Chinese Journal of Solid Mechanics, 2021, 42(03): 287-301.

[20] Yu Guoqiang, Zhang Maosheng, Wang Genlong. Research on the prediction model and sensitive factors of the average velocity of debris flow[J]. Journal of Engineering Geology, 2014, 22(03): 355-360.

[21] Hinton G, Vinyals O, Dean J. Distilling he knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38-39.

[22] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[23] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv, 2014.