An Image Recognition Technology Based on Deformable and CBAM Convolution Resnet50

Chao Chen, Bin Wu, Hongying Zhang

Abstract—**Due to many uncertain factors in the real environment, the accuracy of current target recognition is low. In order to obtain more salient features, we present an improved Resnet-50 Convolutional Neural Network (CNN). Firstly, we add a deformable convolution to adapt for the change of object shape automatically, so as to obtain the feature expression ability and enlarge the receiving field. So the accuracy of object recognition accuracy will be improved greatly. Secondly, the activation function in the improved deformable convolution is RELU6, which is used to capture more features of overlapping target, crowded target and small target. Finally, after the every residual Block layer feature of resnet50, we follow Convolutional Block Attention Module(CBAM) which could help convolutional neural network to obtain more discriminative features. Compared with the classical Resnet50 model, on Cifar10 and FashionMNIST dataset, our improved Resnet50 reaches over better accuracy while keeping favorable speed.**

Index Terms—**Resnet50, CBAM, deformable convolution, RELU6 active function**

I. Introduction

After VGG[1] raised the top1 accuracy of ImageNet classification to above 70%, there have been many innovations in making ConvNets complicated for high performance, e.g., the contemporary GoogLeNet[2] and later Inception models[3-5] adopted elaborately designed multi-branch architectures, ResNet[6] proposed a simplified two-branch architecture, and DenseNet[7] made the topology more complicated by connecting lower-level layers with numerous higher-level ones. Except for the inconvenience of implementation, the complicated models may reduce the degree of parallelism[8] hence slow down the inference. An initialization method was proposed to train extremely deep_plain ConvNets[9]. A recent work combined several techniques including Leaky RELU[10-11], max-norm and careful initialization. It is easy to cause the gradient to disappear or explode, and eventually the extracted features are not obvious enough. It is necessary to build an excellent model with reasonable depth and favorable accuracy-speed trade-off, which only involving the most common components(e.g., regular conv and BN). The Resnet50 structure consists only of a stack of 3×3 convolution, 1×1 convolution and sigmoid active functions[10-11]. Aiming at the above-mentioned goals, inspired by deformed and dilated convolution techniques[12-13] and deformable ConvNets [14-15], we add a deformable convolution to enlarge the receiving field. While inspired by some space attention techniques[16-19] and some channel attention [20-24], coordinate attention [25-26] in recent years, we introduce the attention mechanisms after the every residual Block layer feature in CNN so that enhance the feature representation. Based on above analysis, contributions of the current work are summarized as follows. Firstly, in order to fuse the discriminative features, a deformable convolution is added. Secondly, for the purpose of fusing the discriminative features, we modified the RELU to RELU6 active function in deformed convolution. Thirdly, after the every residual block layer feature of resnet50, we follow a CBAM which could help CNN to obtain more Characteristics of interactions. Finally, we carry on the experiment on Cifar10 data set and FashionMNIST data set. Our improved Resnet50 reaches over better accuracy keeping favorable speed. The effectiveness and efficiency of improved Resnet50 on image classification is shown.

Compared with the classical Resnet50 model, on Cifar10, our improved Resnet50 reaches over 94.84%, acc_top1 accuracy, but our model's training and evaling time increase around 1.3%, relevant parameters only increase 1.5%, FLOPs only increase 0.35%, respectively. On FashionMNIST, the accuracy of acc_top1 improved by around 5.6%, but training and testing time was increased by around 1.3% than the Resnet50CBAM. Compared to the classical models like ResNet 50, our improved resnet50 obtained favorable accuracy speed trade-off.

II. Related work

A. Deformable Convolution

Some techniques[27] have been increased to increase the convolution receptive field of deformable convolution, and the convolution effect would be finally increased. In order to increase the characteristic region of convolution, deformable convolution is introduced here.

During the training process, the convolution kernel have an extra parameter, so that the deformable convolution's accept field can be extended to a large range. Figure 1 shown Deformable Convolution. Figure 1. (a) is a image. Figure 1. (b) is a conventional standard convolution kernel with a size of 3×3 (black dots in the figure); Figure 1. (c) is a deformable convolution with a direction vector added to the parameters of each

Manuscript received May 25, 2022; revised November 07, 2022.

This work was supported in part by the Application Basic Research Plan Project of Sichuan Province(No.2021JY0108) and the Science and Technology Research Plan Project of Sichuan Province (No.NJFH20-003).

Chao Chen is a PhD candidate of Southwest University of Science and Technology, Mianyang, Qinglong Avenue 59, P. R. China. (e-mail: ch1050 3@ 126.com).

Bin Wu is a dean of graduate school of Southwest University of Science and Technology, Mianyang, Qinglong Avenue 59, P. R. China. (phone: 139-0901-8585; e-mail: wubin@swust.edu.cn).

Hongying Zhang is a manager of Robotics Technology Laboratory of Southwest University of Science and Technology, Mianyang, Qinglong Avenue 59, P. R. China.(e-mail: zhywyd @163.com).

convolution kernel (the light black arrow in Figure 1. (c) and Figure 1. (d) enable the convolution kernel to change into any shape, Figure 1. (c) and Figure 1. (d) are special forms of deformable convolution. Figure 1. (d) deformable convolution is used in this paper [14-15].

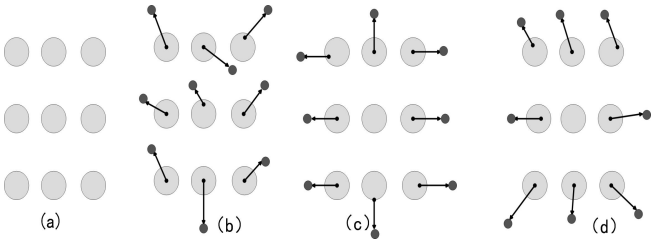

Fig. 1. the differences between standard convolution and deformable convolution.

For a feature map of input, it is assumed that the original convolution operation is 3×3, so in order to learn the offset. We define another 3×3 convolution layer, and the output dimension is actually the size of the original feature map, and the number of channels is equal to 2N. The following deformable convolution can be viewed as an interpolation operation at offset, and then the ordinary convolution can be performed. It basically can cover targets of different sizes. Therefore, through deformable convolution, we can better extract the complete features of the objects we are interested in, and the effect is very good. The operation of deformable Convolution is shown in Figure 2[14-15].

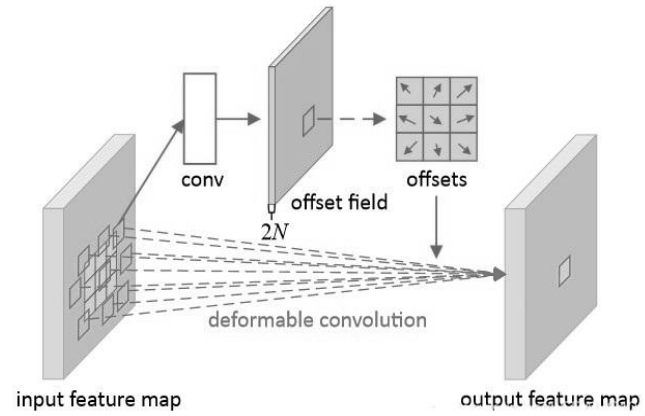

Fig. 2. the operation of deformable convolution.

The differences between Standard Convolution and Deformable Convolution is shown in Figure 3[14-15].

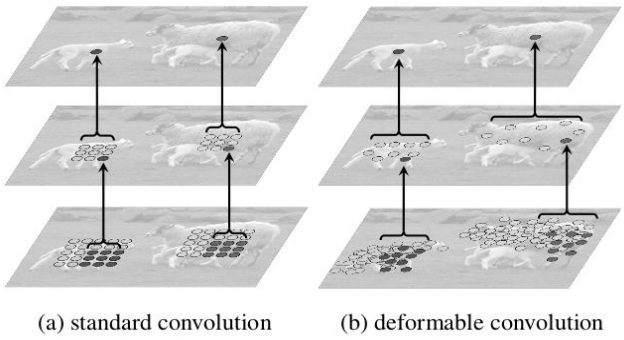

Fig. 3. the differences between Standard Convolution and Deformable Convolution.

B. Attention mechanism

In order to strengthen the interaction between each spatial feature layer. The spatial domain information in the picture is to do the corresponding spatial transformation, so as to extract the key information. Mask generation and scoring of space is the representative work of Spatial Attention Module. It is similar to add a weight to the signals on each channel to represent the relevance of the channel with key information. The greater the weight, the higher the relevance. The mask of Channel was generated and scored. The representative works are SE_NET and Channel Attention Module. The attention mechanism expresses the critical degree of each part in the characteristic data and learns and trains it. The essence of attention mechanism is to use the relevant feature map to carry out the weight of learning, and then apply the weight of learning to the original feature map to carry out the weighted sum, and then get the enhanced feature. According to the different attention domains, the attention mechanisms in computer vision(CV) can be divided into two categories, namely spatial domain, channel domain. This paper uses CBAM attention mechanisms to interpret. And it is shown in Figure 4. CBAM is a light weight universal module, which can be integrated into any classic CNN backbone, and can carry out end-to-end training with the backbone.

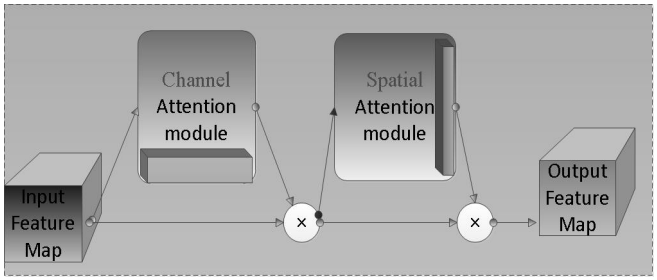

Fig. 4. the operation of deformable convolution.

In our paper, after the feature map output of each stage, the inter channel attention mechanism and spatial attention mechanism are added.

C. Improved Activate function

Because the Sigmoid function is much more expensive than the swish calculation; Disadvantages of RELU function is the gradient is 0 when the input is negative, the problem of gradient disappearance will occur. Leaky RELU function solves the problem of gradient disappearance caused by the RELU function when the input is negative. RELU6 can learn sparse features earlier and can prevent numerical explosion. The activation function was replaced with RELU6. The graphs of the common activation functions and corresponding derivative are shown in Figures 5-6.

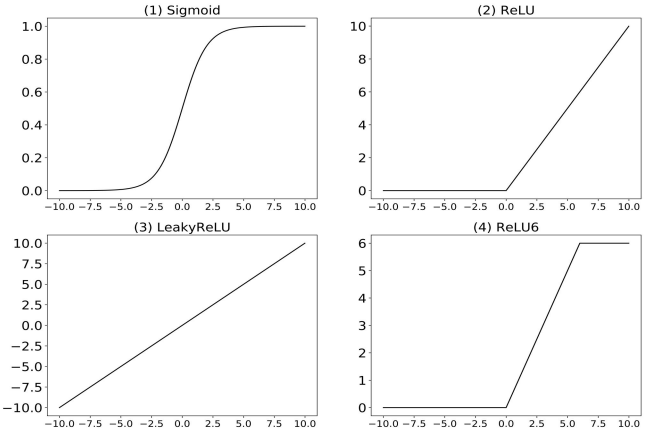

Fig. 5. four Common activation function

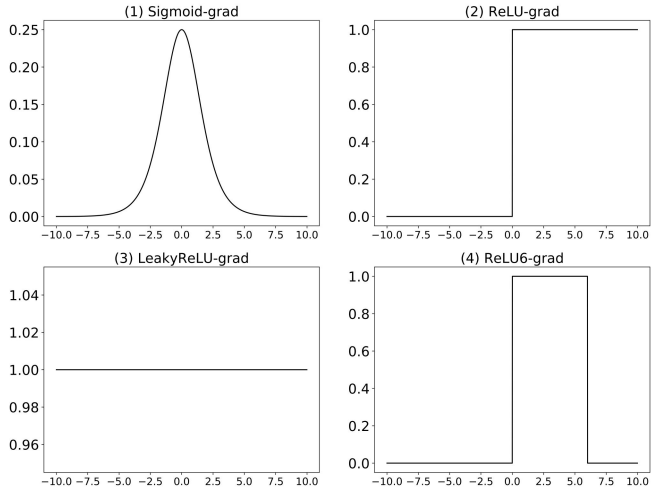

Fig. 6. activation functions' derivative

D. Improved Deformable Convolution

The standard convolution unit samples the input feature map at a fixed position, which results in that the receptive field size of the activation unit at the same CNN layer is the same. This is not desirable for shallow neural networks that encode location information, because different locations may have objects of different scales or different deformations, and these layers need methods that can automatically adjust the scales or sensing fields. Inspired by the dilated Convolution [13], this step of the deformable convolution is set to 2, increasing the receptive field. Especially for non-grid objects, this is not optimal. In the CIFAR10 data set and the Fashion-MNIST data set, there are a lot of irregular targets such as people and animals, in order to identify these targets more accurately. Based on deform Convolution, we design a new convolution,which namely Deformed Convolution which is shown in Figure 7 [14, 15, 27, 28, 29].

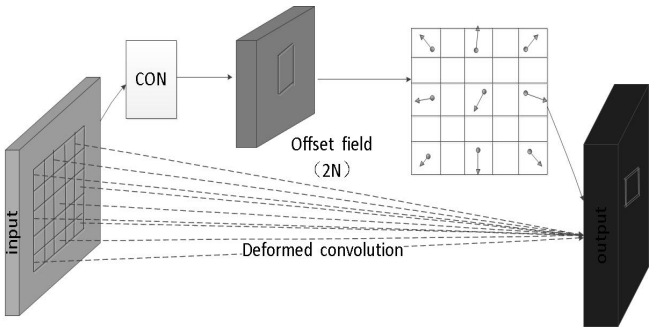

Fig. 7. the operation of our Deformed Convolution

E. Our Improved Resnet50 Framework

Combined with the above three effective modules, we designed an improved Resnet50. The detailed steps are as follows: after the 7×7 convolution, we add a deformable convolution to enlarge the receiving field. And modify the activation function to the leak_RELU; In the bottle block, we add a 1×1 convolution to fuse the discriminative features. After the first layer feature of resnet50, we add CBAM which can obtain more discriminative features. It is shown in Figure 8.

III. Experiment

A. Cifar10 and FashionMNIST Date Set

1*) Cifar10 Date Set*

CIFAR10 is a small data set collated by Alex Krizhevsky and Ilya Sutskever for identifying pervasive objects. A total of 10 categories of RGB color images: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck. The image size is 32×32, and the data set consists of 50000 training images and 10000 test images. The sample picture of Cifar10 is shown in Figure 9.

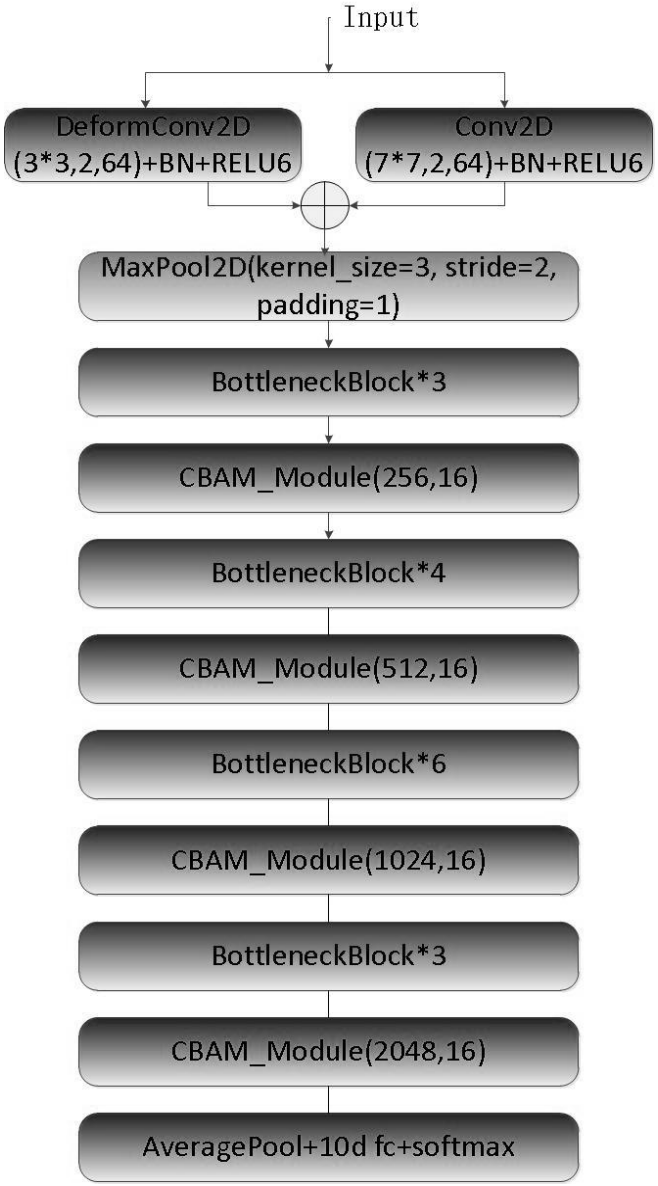

Fig. 8. our improved Resnet50 network architecture with 4 attention mechanism

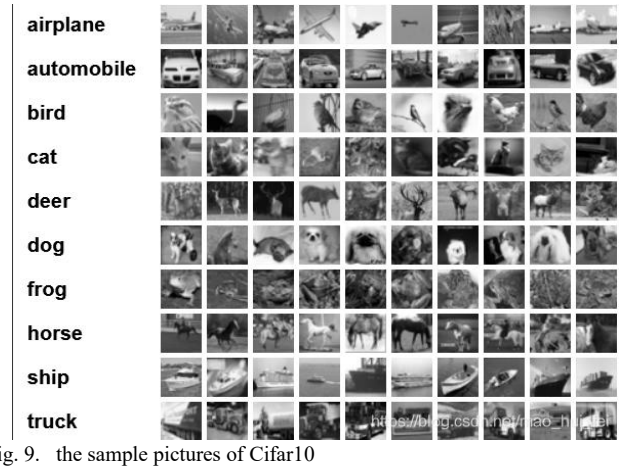

Fig. 9. the sample pictures of Cifar10

Compared with MNIST dataset, Cifar10 has the following differences:
a. Cifar10 is a 3-channel color RGB image, while MNIST is a grayscale image.
b. The image size of Cifar10 is 32×32, while the image size of MNIST is 28×28, which is slightly larger than MNIST.
c. Compared with hand written characters, Cifar10 contains real objects in the real world, which not only has a lot of noise, but also has different proportions and characteristics of objects. So this will bring great difficulties for recognition.

2) FashionMNIST data set

To verify the recognition speed and recognition accuracy of our algorithm, comparative experiments are performed on the FashionMNIST dataset, and the Fashion MNIST sample picture is shown in Figure 10.

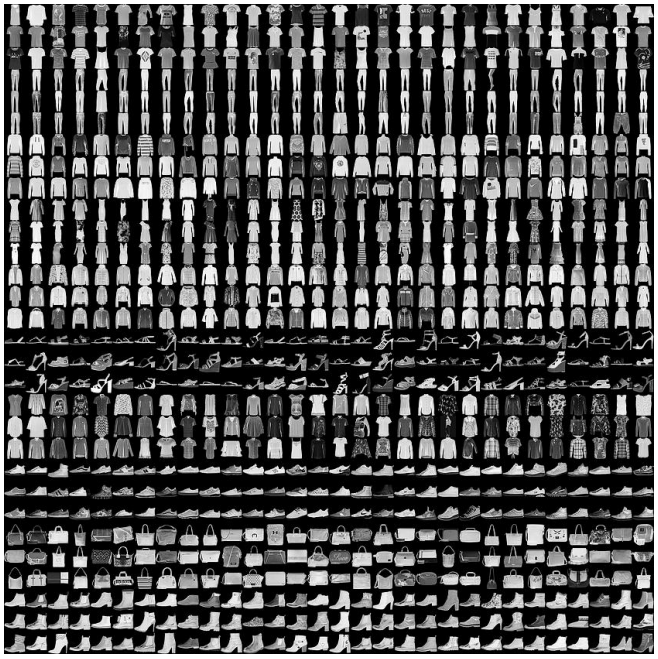

Fig. 10. sample image of FashionMNIST

Compared with the MNIST dataset, the FashionMNIST dataset has the following differences: the FashionMNIST also has an image size of 28×28, but features significantly more than the MNIST. Compared to the MNIST dataset, Fashion MNIST contains real-world real objects, not only with a lot of noise, but also with the proportion and characteristics of objects, causing great difficulty in identification. 60,000 images for training and 10,000 for testing. The model has expanded in size. FashionMNIST is necessities (clothing) in 10 categories, as shown in Table I.

TABLE I
CORRESPONDING CATEGORIES OF FASHIONMNIST

serial number	name	serial number	name
0	T-shirt	5	Sandal
1	Trouser	6	Shirt
2	Pullover	7	Sneaker
3	Dress	8	Bag
4	Coat	9	Ankle boot

In conclusion, identifying FashionMNIST dataset is much more difficult, so it is used to test the effectiveness of the improved algorithm.

3) Experimental environment

Our improved Resnet50 is trained with synchronized SGD over one V100 16GB GPU with a total of 128 images per mini_batch and one 32GB CPU. The training environment is Windows 10 professional edition(64bit) and CUDA10.2; Programming language is python 3.7(64 bit); The framework is AI studio PaddlePaddle 2.0.2(64 bit). The initial learning rate is 0.01. To stabilize the training at the beginning, we extend the number of warmup iterations form the first epoch to the second epoch. For other hyperparasitism, we set the Epoch = 100, boundaries = [60,70], momentum = 0.9, learning_rate = paddle.Optimizer.lr.Piecewise Decay(boundaries = bound aries, values = values), learning_rate = paddle. optimizer. lr. Linear Warmup, paddle.optimizer.Adam, Cross Entropy Loss. Every five epoch, we eval the training situation model.

4) Data preprocessing

Because the Cifar10 and FashionMNIST data set are very simple. And we mainly test the effectiveness of the neural network, this training only uses, Random Resized Crop, ColorJitter, Random Horizontal Flip, Normalize, and does not use excessive data enhancement.

B. Relevant metric

1) Cross Entropy Loss

The most commonly used classification problem is Cross Entropy Loss. This Loss comes from Shannon's information theory. For two objection are be classed, there are only two cases that need to be predicted in the end of the model. For each objection, the probability of our prediction is $\boldsymbol{p}$, this loss is as follows:

$$loss = y_i \log p + (1 - y_i) \log(1 - p) \tag{1}$$

where, $\boldsymbol{y_i}$ is the label of the $\boldsymbol{i}$ sample's label, the positive class is 1 and the negative class is 0. $\boldsymbol{p_i}$ is the probability that the $\boldsymbol{i}$ sample is predicted to be a positive class. In fact, the situation of multi classification is an extension of two classification.The calculation formula is as follows:

$$loss = \sum_{i}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \tag{2}$$

2) Accuracy rate

Accuracy rate (acc) refers to the proportion of the correctly classified records to the total number of records in the classification by using the eval set. The results show that the proportion of the correct forecast quantity in the total quantity is higher, the model effect is better. The calculation formula is as follows:

$$acc = \frac{TP}{TP + FP} \tag{3}$$

where, *TP* represents the number of correctly classified records and *FP* represents the number of wrongly classified test data. As we know, ImageNet has about 1000 categories, and when the model predicts a certain picture, it will give 1000 categories ranking from high to low in probability. The so-called top1 Accuracy refers to the accuracy rate that the first category in the ranking is consistent with the actual result. Top5 Accuracy refers to the Accuracy of actual results included in the top five categories.

3) Parameter quantity and floating point number calculation quantity

At the same time, the actual situation of parameter quantity and floating-point number calculation quantity are analyzed. The parameter quantities (PARAMs) in the convolution network correspond to the spatial complexity. PARAMs can be computed by conv_param = (kernel_size × in_channel + bias)

× out_channel, where kernel_size indicates the convolution kernel size, in_Channel indicates the number of input channels, out_Channel indicates the number of output channels and bias indicates the number of offset. Floating point calculations (FLOPs) correspond to time complexity. It refers to floating point operands and is understood as computational quantities. Therefore, these two quantities can be used to measure the complexity of the algorithm/model. That is, the amount of network parameters is closely related to the video memory, and the amount of floating-point calculations is related to the computing speed of the GPU. It refers to floating point operands and is understood as computational quantities. Therefore, these two quantities can be used to measure the complexity of the algorithm/model.

4) Cifar10 *experimental results*

In order to evaluate the recognition performance of the proposed method, the experimental results are analyzed quantitatively and qualitatively. The results are compared with other mainstream recognition algorithms. The category confidence of the visualization results is greater than or equal to 0.5, and different shapes of line are used to indicate no same target category. Experimental results show that our proposed method can effectively improve the performance of object recognition. In order to improve the actual effect of Cifar10, we normalized images to 224×224. Cifar10's eval_ loss is shown as Figure 11.

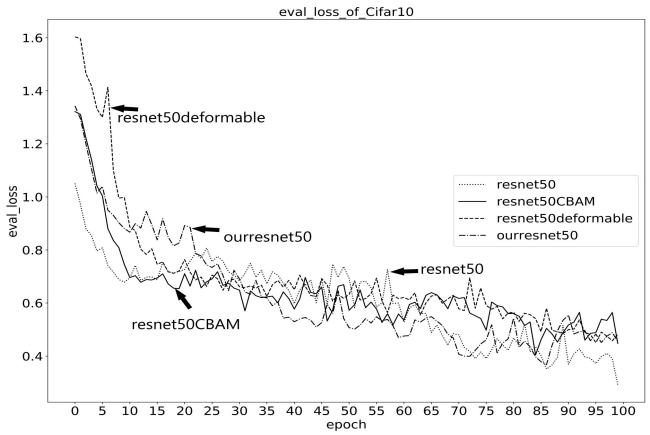

Fig. 11. eval Loss of Cifar10

In the eval process of iteration, it can be seen that the loss corresponding to the our Resnet50 is continuously declining without rebound phenomenon, and the convergence speed is faster than that of the other four algorithms. At the same time, the actual effect can be observed by the accuracy is shown as Figures 12 and 13.

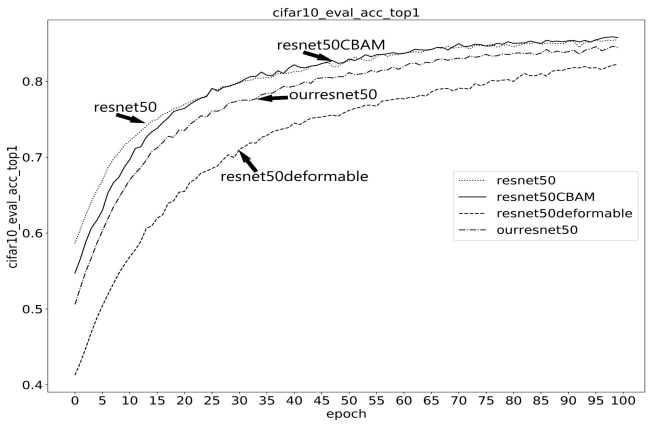

Fig. 12. eval acc_top1 of Cifar10

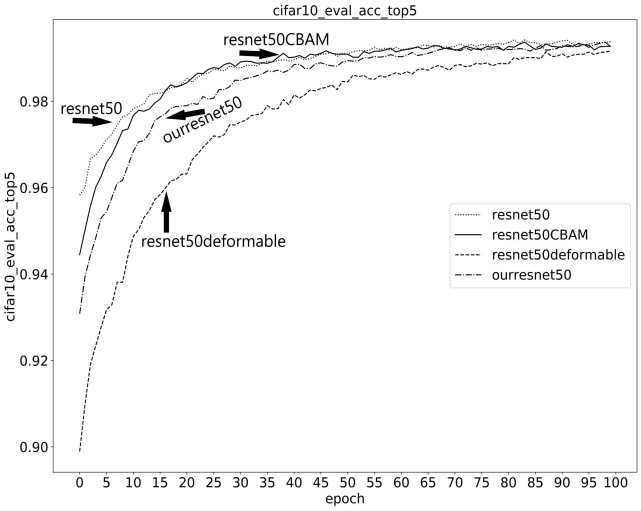

Fig. 13. eval acc_top5 of Cifar10

In the eval process of iteration, the acc_top1 and acc_top5 of our Resnet50 continue to rise at a relatively steady trend. It exceed the accuracy of the Resnet50, and obviously better than the other three algorithms. During training, the corresponding train experimental results are shown in Figures 14-1 6.(For clear display, the training situation is displayed 2000 times per iteration.)

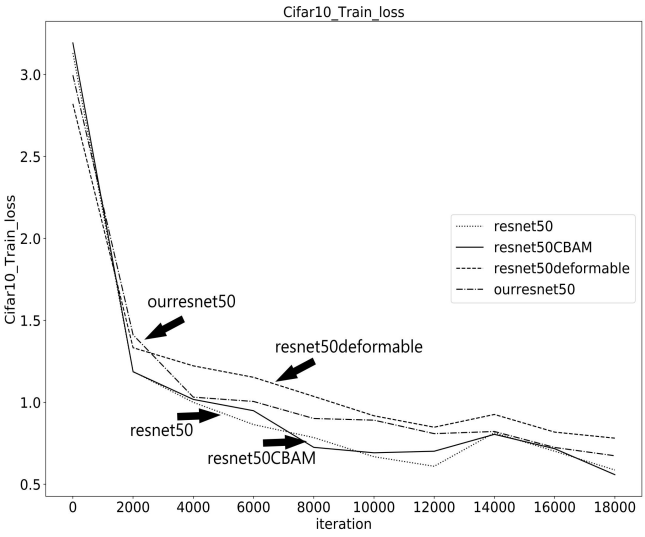

Fig. 14. train loss of Cifar10

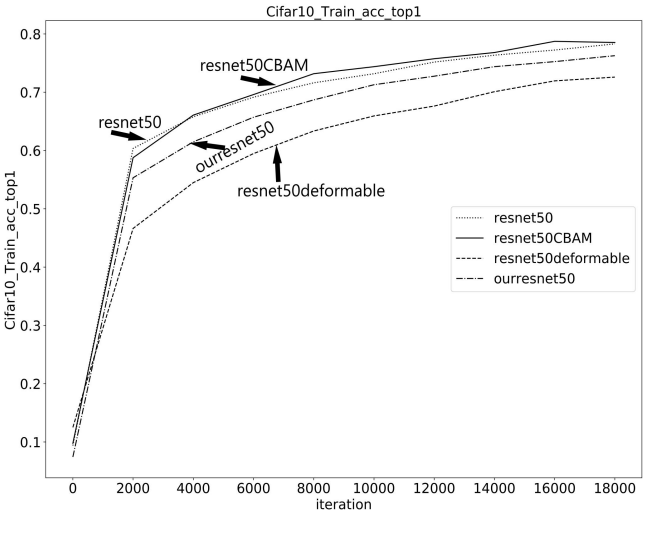

Fig. 15. train acc_top1 of Cifar10

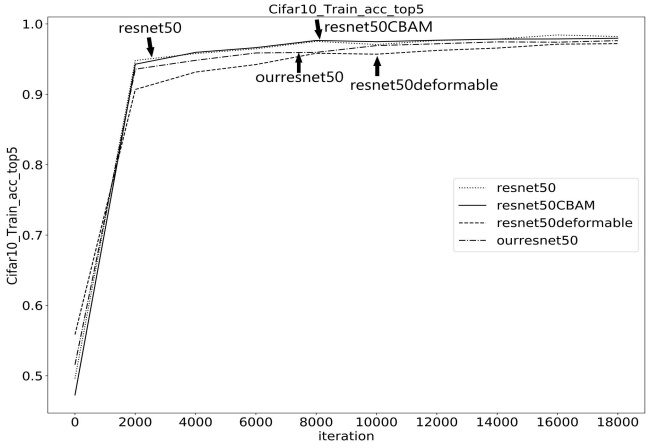

Fig. 16. train acc_top5 of Cifar10

In terms of accuracy, it is significantly better than another Resnet50. At the same time, we have analyzed the number of parameters, calculation and other aspects of the contrast effect. And The details are shown in Table II. For the simplicity(A: Resnet50, B: Resnet50_CBAM, C: Resnet50_ Deformable, D: Resnet50_CBAM_Deform.).

TABLE II
Cifar10's parameters of the four algorithms

	Time(hour)	PARAM	FLOP
A	14:36:29	23581386	4107881472
B	14:48:42	24282438	4112809472
C	14:41:58	23584190	4118703680
D	**14:59:40**	**24284986**	**4122292736**

As can be seen from Table II, compared with the classical Resnet50 model, on Cifar10, our improved Resnet50 reaches over 94.84% acc_top1 accuracy. The actual accuracy and detection speed are shown in Table III.

In particular, compared with the classical Resnet50 model on Cifar10, our model's training and evaluating time increase around 1.0%, relevant parameters only increase 1.5%, FLOP only increase 0.35%, respectively. And favorable higher accuracy and faster speed compared to the classical models are shown like ResNet50. The experimental results show that the RELU6 activation function is better than the traditional active function, because the RELU6 activation can avoid the vanishing gradient and can effectively solve the problem of unbalanced positive and negative samples in object recognition problem.

5) FashionMNIST Experimental Result

To save training time, normalize all the images to the size of 32×32 here. FashionMNIST's eval loss, acc_top1 and acc_ top5 are shown as Figures 17-19.

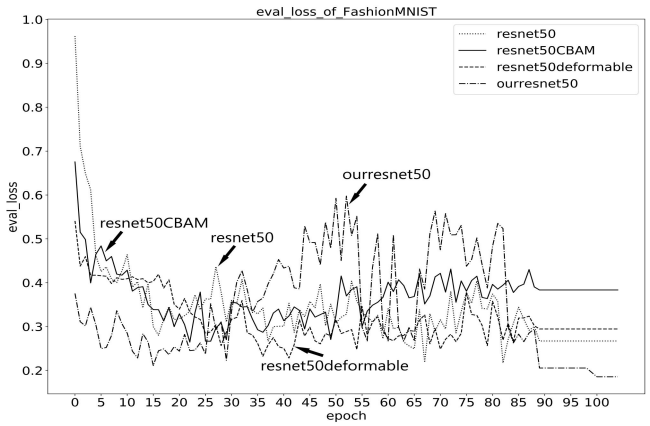

Fig. 17. eval loss of FashionMNIST

TABLE III
Cifar10's results of the four algorithms

	loss	Acc_top1(%)	Acc_top5(%)	train/eval(ms)
A	0.2418	94.66	99.27	2000/415
B	0.2675	93.97	99.89	2000/7800
C	0.3316	94.38	99.92	2000/432
D	**0.13915**	**94.84**	**99.72**	**2000/4650**

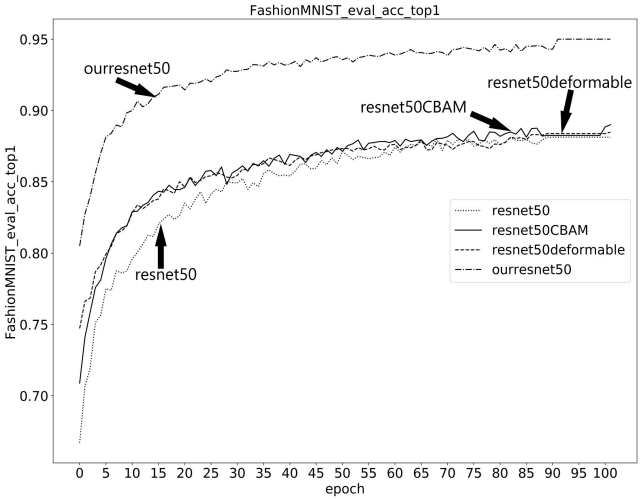

Fig. 18. eval acc_top1 of FashionMNIST

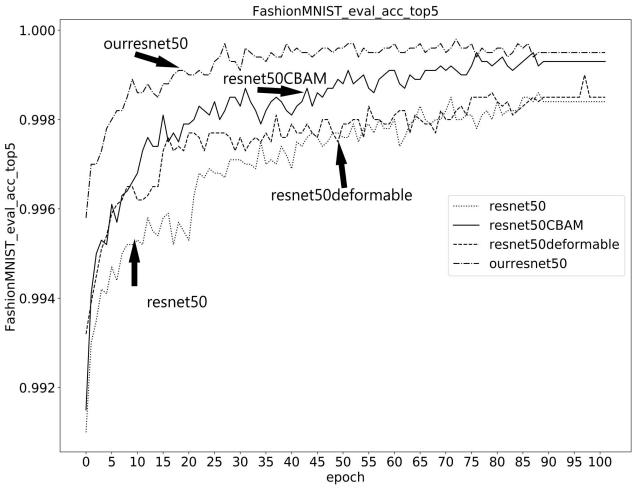

Fig. 19. eval acc_top5 of FashionMNIST

FashionMNIST's train loss , acc_top1 and acc_top5 are shown as Figures 20-22.(For clear display, the training situation is displayed 4000 times per iteration.)

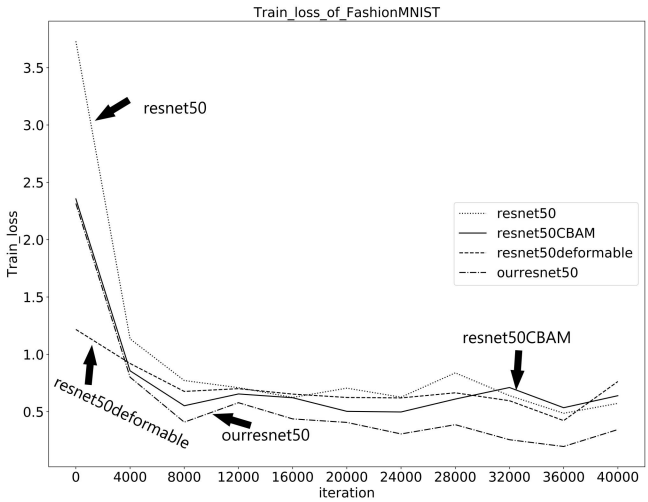

Fig. 20. train loss of FashionMNIST

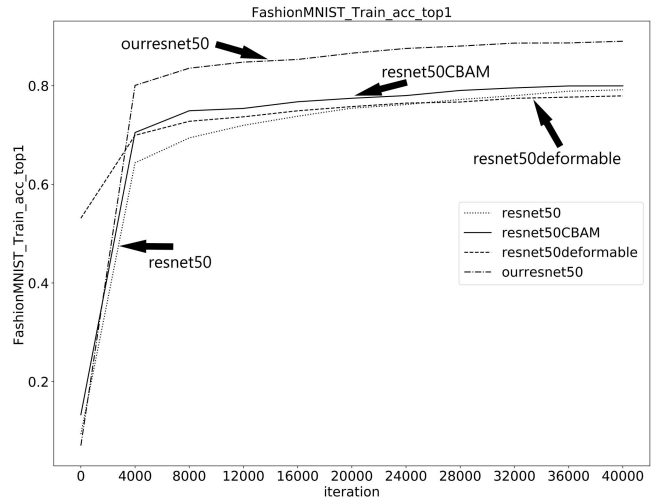

Fig. 21. train acc_top1 of FashionMNIST

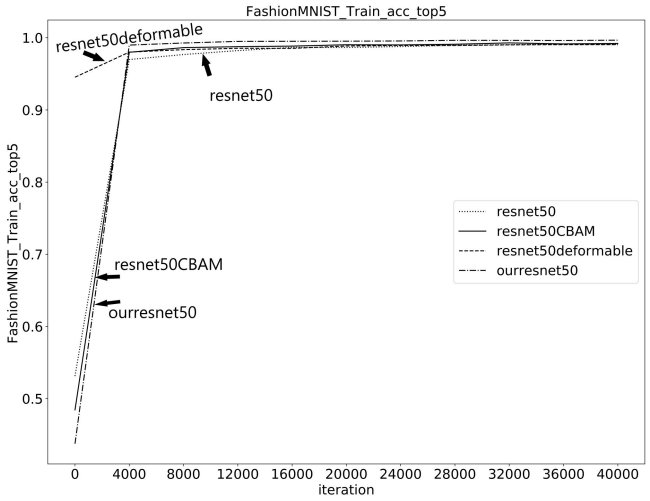

Fig. 22. train acc_top5 of FashionMNIST

And the details are shown in Table IV and V. For the simplicity(A: Resnet50, B: Resnet50_CBAM, C: Resnet50_Deformable, D: Resnet50_CBAM_Deform.).

TABLE IV
FashionMNIST's relevant parameters of the four algorithms

	Time(hour)	PARAM	FLOP
A	2:27	23581642	81669632
B	2:47	24282438	83113408
C	2:27	23574782	81674240
D	**2:35**	**24284986**	**83306944**

TABLE V
FashionMNIST's results of the four algorithms

	loss	Acc_top1(%)	Acc_top5(%)	train/eval(ms)
A	0.1897	88.13	99.83	191/119
B	0.3969	88.76	99.94	195/122
C	0.4124	88.38	99.98	192/120
D	**0.3455**	**94.61**	**99.93**	**201/124**

From Table IV and V, compared the resnet50, the improved model's training and evaluating time increase around 0.05% and 0.04%, relevant parameters only increase 3%, FLOP only increase 2%, respectively. In particular, on FashionMNIST, the accuracy of acc_top1 improved by around 6.3%, but training and testing time was increased by around 3% than the Resnet50CBAM. Improved resnet50 obtained better results compared classics models like ResNet50. Our improved RELU6 activation function performs well in solving the problem of uneven positive and negative samples in object detection. Experimental results show that our proposed method can effectively improve the performance of object recognition and doesn't add much extra time overhead.

IV. Conclusions

We have proposed a method to add a improved deformable convolution involving our modified RELU activation function. After every residual module layer feature of resnet50, we follow a CBAM which could help CNN to obtain more Characteristics of interactions between channels and Spaces. On Cifar10 and FashionMNIST data set, there's a good trade-off between accuracy and speed using our improved Resnet-50. In the future work, new backbone networks may be modified to extract better features, or loss functions may be modified to find the optimal value.

Acknowledgment

We are thankful to our colleagues(Hua Kong, Yi Liu, Fang liu, Ju Wu and Li Zhang) for their help.

References

[1] K. Simonyan and A. Zisserman, “Very deep convolutional networks for largescale image recognition”.ICLR,(2015), pp. 1 – 14. Doi://dx.doi. org/10.4236.

[2] C. Szegedy, W. Liu, Y. Jia, et al, ” Going Deeper with Convolutions”, In Proceedings of the IEEE conference on computer vision and pattern recognition,(2015), pp. 1 – 9. http://arxiv.org/abs/14 09.4842.

[3] I. Sergey and S. Christian, ” Batch normalization: Accelerating deep network training by reducing internal covariate shift”, in International Conference on Machine Learning ,(2015).http://arXiv:1502.03167v3.

[4] C. Szegedy, S. Ioffe, V. Vanhoucke,et al,” Inception v4, Inception ResNet and the Impact of Residual Connections on Learning ”, In AAAI Conference on Artificial Intelligence, (2016), pp. 1 – 121/16. htt p://arXiv:1602.07261v2.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, et al, ” Rethinking the inception architecture for computer vision ” ,In Proceedings of the IEEE conference on computer vision and pattern recognition, (2015), pp. 2818-2826. http://arXiv:1512.00567v3.

[6] K. He, X. Zhang,S. Ren and J. Sun, ” Deep residual learning for image recognition ”, In Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), pp. 770 – 778.http://arXiv:1512. 03385v1.

[7] G. Huang, Z. Liu, V. Laurens, et al,” Densely Connected Convolutional Networks ”, In Proceedings of the IEEE conference on computer vision and pattern recognition , (2017), pp. 2261 – 2269.http://arXiv:-512.03385v1.

[8] N. Ma, X. Zhang, H. Zheng and J. Sun. ” ShuffleNet V2:Practical Guidelines for Efficient CNN Architecture Design”, European Conference on Computer Vision, Springer, Cham , (2018).http://ar Xiv:1807. 11164v1.

[9] X. Li,Y. Bahri,J. Sohl – Dickstein, et al,” Dynamical isometry and a mean field theory of cnns: How to train 1000-layer vanilla convolutional neural networks”, Proceedings of the International Conference on Machine Learning,(2018).

[10] O. K. Oyedotun, A. Shabayek, D. Aouada, et al,” Going deeper with neural networks without skip connections”, In 2020 IEEE International Conference on Image Processing(ICIP)(2020), pp. 1756 – 1760. DOI: 10.1109/ICIP40778.2020.9191356.

[11] X. Ding, X. Zhang,N. Ma, et al, ” RepVGG: Making VGG style Con vNets Great Again ”, In Computer Vision and Pattern Recognition, (2021). https: //arxiv. org/abs/2101.03697.

[12] Y. Ma, H. Shuai and W. Cheng,” Spatiotemporal Dilated Convolution with Uncertain Matching for Video-based Crowd Estimation”, IEEE Transactions on Multimedia(2021) ,(2021), pp. 261 – 273.DOI:10.11 09/TMM.2021.3050059.

[13] H. Cheng, et al, ” Millimeter Wave Path Loss Modeling for 5G Communications Using Deep Learning with Dilated Convolution and Attention ” IEEE Access,(2021),V99 pp. 1-7. DOI: 10.1109/ACCESS.2021.3070711.
[14] F. Chen, et al, “Adaptive deformable convolutional network ” , Neurocomputing ,(2020): pp. 853 – 8646/2020. DOI:10.1016 /j.neucom.2020.06.128.
[15] X. Zhu, ”Deformable ConvNets V2: More Deformable, Better Results, Deformable ConvNets V2: More Deformable” , Better Results,(2018).https://arxiv.org/abs/1811.11168.
[16] K. He, X. Zhang, S. Ren, et al, ” Deep residual learning for image recognition ” , In Proceedings of the IEEE conference on computer vision and pattern recognition ,92(2016). https://doi.org/10.1109/CVPR.2016.90.
[17] [17] M. Goyal, N. Reeves, et al, ” DFUNet: Convolutional Neural Networks for Diabetic Foot Ulcer Classification” , IEEE Transactions on Emerging Topics in Computational Intelligence 2020 ,(2020):pp. 728 – 739. DOI:10.1109/TETCI.2018.2866254.
[18] H. Zhang and C. Wu, ” ResNeSt:Split – Atten tion Networks Resnest: Split-attention networks” , (2020). https://arXiv:2004.08955 v2.
[19] S. Wang, Y. Liu, Y. Qing, et al,” Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network” , IEEE Access(2020),8(2020), 184841 – 184850. DOI:10.1109 /ACCESS.2020.3029857.
[20] A. Brock, S. De, S. L. Smith, et al, ”High – Performance Large – Scale Image Recognition Without Normalization”,(2021). http://arXiv: 2102.06171v1.
[21] S. Woo,J. Park,J. Lee, et al, ” CBAM: convolutional block attention module ” , IEEE European Conference on Computer Vision (ECCV),(2018),pp. 3 – 19. http://arXiv:1807. 06521v2.
[22] Y. Cao, J. Xu, S. Lin, et al, ” Gcnet: Non – local networks meet squeeze – excitation networks and beyond” , newblock In Proceedings of the IEEE/CVF International Conference on Computer Vision(ICCV)Workshops, (2019). https://arxiv.org/abs/1904.11492.
[23] I. Bello, B. Zoph, A. Vaswani, et al, ” Attention augmented convoluti onal networks” , (2019).http://arXiv: 1904. 09925.
[24] O. Ronneberger, P. Fischer and T.Brox. ”U – net: Convolutional networks for biomedicalimage segmentation” , (2015): pp. 234 – 241.http://arXiv:1505.04597v1.
[25] Q. Hou, D. Zhou and J. Feng. ”Coordinate attention for efficient mobile network design” ,(2021), pp. 770 – 778. http://arXiv:2103.02907.
[26] Q. Cheng, H. Li, Q. Wu, et al, ” BAM: A Batch Aware Attention Module for Image Classification” , (2021). http://arXiv:2103.15099.
[27] L. Guo, X. L. Zhao, X. M. Gu, et al, ” Three – dimensional fractional total variation regularized tensor optimized model for image deblurring” , Applied Mathematicsand Computation, V. 404, N11/21. (2021). https://doi.org/10.1016/j.amc.2021.126224.
[28] Y. Luo, Y. C. Zeng, R. Z. Lv, et al,”Dual – stream VO: Visual Odometry Based on LSTM Dual – Stream Convolutional Neural Network”,Engineering Letters, vol.30, no.3, pp.926 – 934,2022. (2022).https://doi.org/10.48550/arXiv.2105.14734.
[29] J. Gao, and Y. M. Chen, ” Finite-time and Fixed time Synchronization for Inertial Memristive Neural Networks with Timevarying Delay and Linear Coupling” ,IAENG International Journal of Applied Mathematics, vol.52, no.3, pp. 534 – 540, 2022.(2022). https://doi.org/10.1007/s11571-017-9455-z.

Chao Chen is a lecturer in Key Laboratory of Numerical Simulation in Sichuan University, Dongtong Road 1124, P. R. China. He has published 8 core papers and have applied for 23 computer software Copyrights.
Bin Wu is a professor and doctoral supervisor in Southwest University of Science and Technology; He is also a part-time doctoral supervisor of China Academy of Engineering Physics, an outstanding expert with outstanding contributions in Sichuan Province and an academic and technical leader in Sichuan Province.
Hongying Zhang is a professor and doctoral supervisor in Southwest University of Science and Technology. She is also a candidate of academic and technical leader of Sichuan Province. Her research interests include visual information processing technology, big data visualization technology, image restoration and reconstruction technology, etc.