**ACNet-based MobileNet for image classification**

# Abstract

In this paper, we propose a novel ACNet-based MobileNet(Adaptively Connected Neural Networks based MobileNet) for image classification. Google's MobileNet[1] gets a significant achievement in image classification on the mobile device platform in recent years. However, MobileNet has fewer model parameters, making its accuracy still not comparable to other large-scale network models. Previously, ACNet[2] proposed to improve the traditional convolutional neural networks (CNNs), can flexibly change the global and local reasoning in the internal feature performance, and it also enhances classification accuracy. We believe that ACNet can adequately compensate for the above-mentioned MobileNet problems. Therefore, our ACNet – based MobileNet have benefited is that while retaining the inverted residual architecture of the MobileNet model, the model parameters are small enough. It also could improve the accuracy of image classification and reduce training time.

# Introduction

Google's MobileNet significantly reduces the parameters of the model by using depthwise separable convolutions(DWS)[3], which makes a meaningful contribution to porting to mobile devices. Even though mobileNetV3 has been improved by 3.2% in accuracy compared with the mobileNetV2 in ImageNet classification through the inverted residual with linear bottleneck and squeeze and excitation structure, the accuracy is still not comparable to other large-scale network models, such as ResNet[4] and VGG[5]. How to improve the accuracy of MobileNet under the premise that the model volume is small enough has become the key to the successful application of deep learning in the field of mobile devices.

Simultaneously, more and more models use Convolutional neural network (CNN) as a vital part of the model with the large-scale application of deep learning in image classification and target detection. However, the limitations of CNN itself have also been continuously confirmed. Due to CNN only extracts information from local neighboring pixels, each layer in the convolutional network does not have an excellent global overturning ability. Therefore, the convolution operation cannot distinguish two similar objects well. ACNet - Adaptively Connected Neural Networks can effectively solve this problem. The author holds that the optimization and reconstruction of DWS in MobileNet by ACNet can effectively avoid CNN pays too much attention to the local reasoning phenomenon, to improve the accuracy. Wang et al. also proposed that using ACNet has the function of reducing the model training cycle. The training period of ACNet-based MobileNet after optimization in this article will raise more efficiency compared to mobileNetV3[1].

# Background and Literature review

Since LeCun's[6] paper on CNN was published, CNN's modern structure has been widely used in image classification. At the same time, large-scale datasets and the continuous improvement of computer computing power, deep CNN is continuously applied to image classification. The most famous is that Krizhevsky[7] proposed a famous AlexNet CNN structure, and gained a significant breakthrough in image recognition. AlexNet was a great success, setting off a research boom in convolutional neural networks. After this, the researchers put forward other improvement techniques. Based on the AlexNet model, a more layered and deeper VGGNet[5] model was proposed to solve image classification and achieve higher accuracy. GoogLeNet[8] adopts the idea of an Inception structure to enrich models' diversity. ResNet[4] took the lead in proposing the concept of residual network, which effectively suppressed the overfitting of the model. During this period, CNN's model is continuously developing in a more complex and deeper direction. ResNet's score in the ILSVRC 2015[9] competition is more than 20 times that of AlexNet and more than eight times that of VGGNet. By increasing the depth, the network can obtain a more accurate non-linear objective function, so that the model can better reflect the characteristics. However, increasing the network's overall complexity makes the model too large, which can only stay in the laboratory stage.

For this reason, how to apply deep learning models to real-life has attracted more and more attention. The Google team has made many contributions to this. The Google team has proposed MnasNet[8] and MobileNet. Among them, MobileNet uses technologies such as an inverted residual network and DWS to reduce the model parameters significantly. Although the model can be applied to mobile devices, the reduced model parameters also bring a side effect of decreasing accuracy. Simultaneously, another defect of CNN is also reflected on MobileNet: CNN only extracts information from local neighboring pixels, so each layer inside the convolutional network does not have an excellent global knockdown capability[9]. Therefore, the convolution operation cannot distinguish two similar objects well. The ability to distinguish between two similar objects on a mobile device is another major bottleneck for deep learning applications.

This paper's primary purpose is to build a deep learning model with sufficiently small parameter models and sufficient accuracy. Aiming at the shortcomings of CNN, we use ACNet's flexible parameter storage method to learn the ability to transform general data local and global reasoning and propose an ACNet-based MobileNet network model. In this way, a brand-new, mobile-friendly, and high-precision image classification model is obtained.

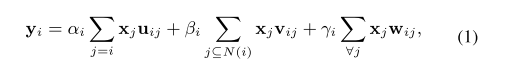
# Methodology

**Dataset**

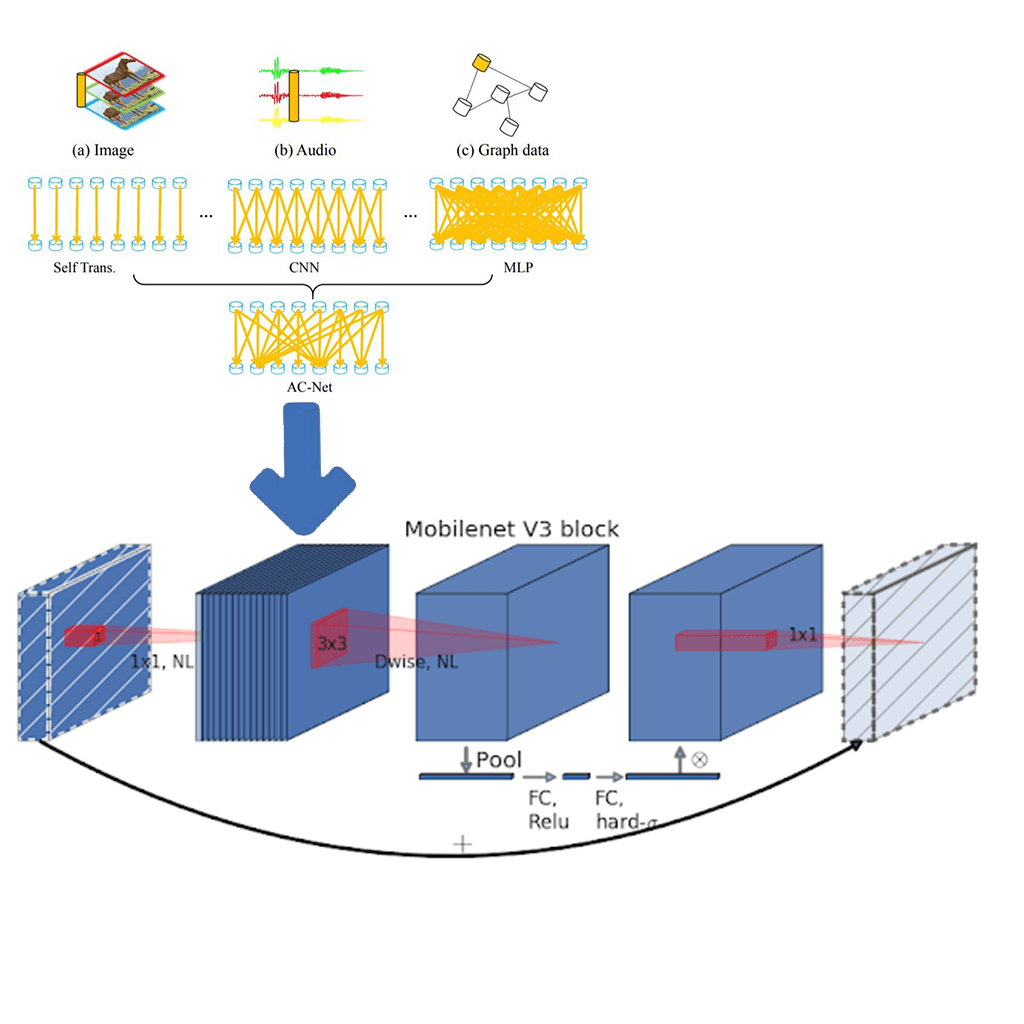
This article uses Cifar-100[10] dataset. The dataset has 100 classes, and each class includes 600 images. Among the 600 images, there are 500 training images and 100 testing images. According to a convention, two error rates will be provided: top-1 and top-5. The top-5 error rate means that the testing image's correct label is not among the five most likely notes considered by the model. In the data preprocessing stage, the data pictures are uniformly cropped to a fixed size of 224\*224 at the center point. We will mirror and flip the image to achieve the purpose of expanding the data set.

**Architecture**

In ACNet-based MobileNet, the 3x3 and 5x5 convolution operation in the inverted residual module will no longer be used. Instead, we propose an adaptive inverted residual module. The traditional CNN convolution is optimized by introducing formula (1). X is the input layer, Uij​,Wij​,Vij​ represent the learnable weights. Use the local transformation or the global transformation for self-inference through the three weight parameters of α,β,γ.



We can obtain an adaptive neural layer through the above formula, and then apply the layer model to the inverted residual module to obtain an adaptive inverted residual module(AIR). Universal adaptive inverted residual module to adaptively capture global and local dependencies. So that ACNet-based MobileNet has higher accuracy and global reasoning ability. Due to the use of the adaptive inverted residual module, the training model time is faster.



**Training**

We use multiple controlled trials to test ACNet-based MobileNet. The first is to verify the validity of the ACNet-based MobileNet theory. Use the MobileNetV3-Small structure mentioned in the mobilenetv3 paper for training on cifar100. Then use ACNet-based MobileNet (Table1)with the same structure and params to classify and predict cifar100. Use a single factor to verify the functionality of ACNet-based MobileNet. At the same time, we used the techniques in Tong's[11] paper to optimize the model parameters and tried to get the model with the highest accuracy.

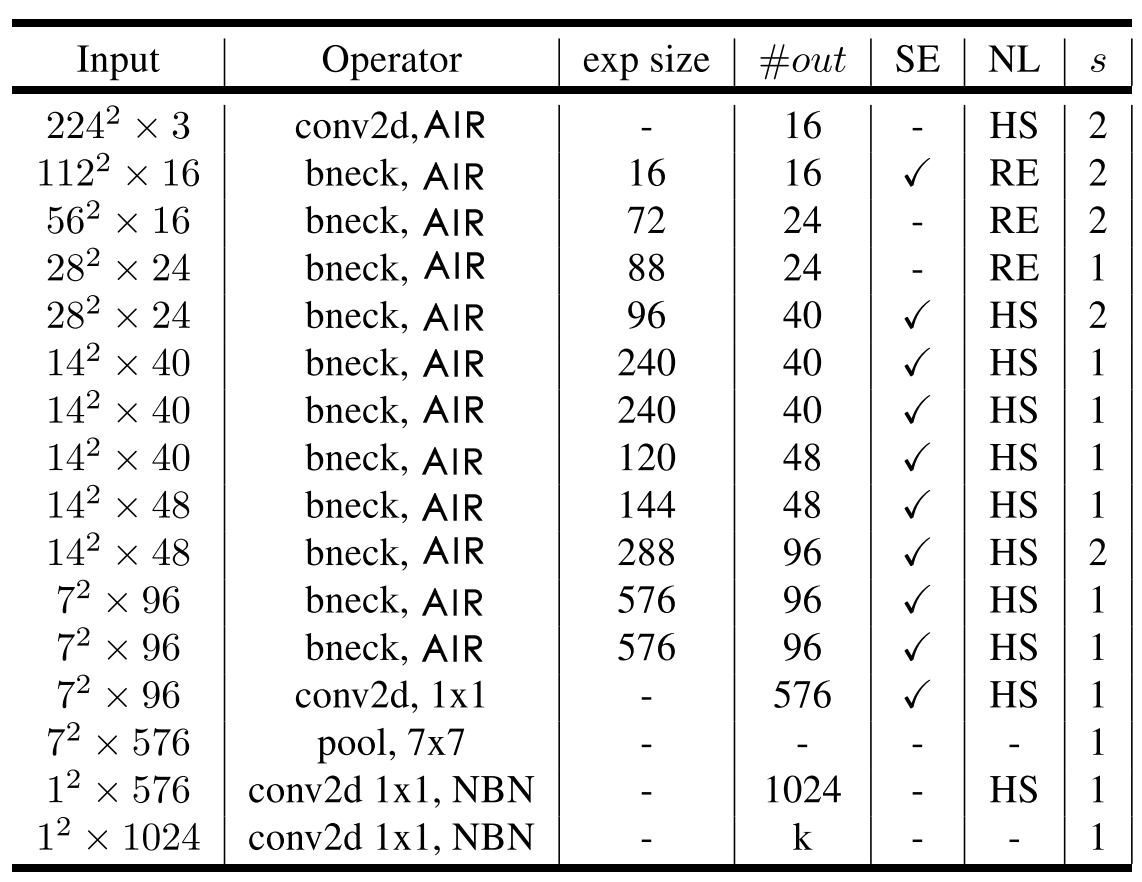


Table1: It uses the same architecture as mobilenetv3, which uses AIR instead of 3X3 and 5X5 convolution operations.

# Possible conclusions

The ACNet-based MobileNet model uses the AIR module, which theoretically can effectively overcome the weakness of CNN's weak global reasoning. The model can effectively distinguish objects with the same shape feature, thereby improving the model's accuracy. The accuracy of ACNet-based MobileNet is expected to be higher than that of MobileNetV3. At the same time, the use of the AIR module can reduce computational efficiency, appropriately reduce model parameters, and shorten training time. In contrast, ACNet-based MobileNet has more advantages than MobileNetV3 when applied to the mobile terminal.

# References

[1] A. Howard *et al.*, “Searching for mobileNetV3,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 1314–1324, 2019, doi: 10.1109/ICCV.2019.00140.

[2] G. Wang, K. Wang, and L. Lin, “Adaptively connected neural networks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 1781–1790, 2019, doi: 10.1109/CVPR.2019.00188.

[3] A. G. Howard and W. Wang, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2012.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, 1998, doi: 10.1109/5.726791.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[8] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June-2015, doi: 10.1109/CVPR.2015.7298594.

[9] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[10] M. B. McCrary, “Urban multicultural trauma patients.,” *ASHA*, vol. 34, no. 4, 1992.

[11] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, “Bag of tricks for image classification with convolutional neural networks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 558–567, 2019, doi: 10.1109/CVPR.2019.00065.