

Received April 20, 2021, accepted April 27, 2021, date of publication May 17, 2021, date of current version June 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079435

Multi-Attention Ghost Residual Fusion Network for Image Classification

XIAOFEN JIA^{1,2}, SHENGJIE DU¹, YONGCUN GUO², YOURUI HUANG¹, AND BAITING ZHAO¹

¹School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China

²State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China

Corresponding author: Yongcun Guo (ycguo2018@163.com)


This work was supported in part by the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-048 and GXXT-2020-54, in part by the Natural Science Research Projects of Colleges and Universities in Anhui Province under Grant KJ2018ZD008, in part by the National Key Research and Development Program under Grant 2016YFC0600908, and in part by the National Natural Science Foundation of China under Grant 61501006.

ABSTRACT In order to achieve high-efficiency and high-precision multi-image classification tasks, a multi-attention ghost residual fusion network (MAGR) is proposed. MAGR is formed by cascading basic feature extraction network (BFE), ghost residual mapping network (GRM) and image classification network (IC). The BFE uses spatial and channel attention mechanisms to help the MAGR extract low-level features of the input image in a targeted manner. The GRM is formed by cascading 4 multi-branch group convolutional ghost residual blocks (MGR-Blocks). Each MGR-Block is cascaded by a dimension reducer and several ghost residual sub-networks (GRSs). The GRS integrates ghost convolution and residual connection, and the use of ghost convolution can significantly reduce parameters and achieve high-efficient classification. The GRS is a parallel convolution structure with 32 branches, which ensures that GRM has enough width to extract advanced features and extract as much feature information as possible, so as to obtain high-precision classification. The IC completes the aggregation of high-dimensional channel feature information, and then achieves a significant improvement in the classification accuracy of MAGR, by fusing the effective channel attention mechanism, global average pooling and SoftMax layer. Simulation experiment shows that MAGR has excellent classification capability while achieving high efficiency and lightweight. Compare with VGG16, the parameters of MAGR on CIFAR-10 is reduced by 94.8% while the classification accuracy is increased by 1.18%. Compare with MobileNetV2, the parameters of MAGR on CIFAR-100 is reduced by 33.9% while the classification accuracy is increased by 15.6%.

INDEX TERMS Image classification, attention mechanism, ghost convolution, multi-branch group convolution, residual connection.

I. INTRODUCTION

Image classification is a technology that uses algorithms to determine the category of a given image. It is widely used in security (face recognition, pedestrian detection), traffic (vehicle counting, retrograde detection, license plate recognition), internet (image retrieval, photo album automatic classification) and other fields. Traditional image classification algorithms, such as AlexNet [1] and VggNet [2], perform well for simple classification tasks, but have low efficiency and poor precision classification result for images with severe interference or subtle differences. Therefore, one of the mainstream

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo .

directions of current scholars is to design neural networks with high accuracy and fast training.

In order to improve the classification accuracy of the model, He [3] *et al.* propose the ResNet residual network, which solves the gradient dispersion and explosion problems caused by network deepening through the residual connection structure. On the basis of ResNet, DenseNet [4] achieves better classification results than ResNet with the same depth. MobileNetV2 [5] is proposed by introducing residual structure of the ResNet network to MobileNetV1 to achieve higher classification accuracy. All the above methods adopt the idea of residual connection to increase the network depth, thereby achieving the improvement of classification effect. However, the increase of network depth will inevitably

increase computation cost and slow down the training speed. Therefore, it is necessary to explore methods to improve the classification accuracy without deepening the network depth. ResNeXt [6] improves the feature extraction ability by broadening the network width, and the classification effect is better than ResNet network under the same layers condition. Whether the increasing of the network depth or the network width, it will bring additional calculations. Compare with increasing the network depth, widening the network width will bring less computation. Therefore, widening the network width is more cost-effective to improve the network classification effect, but the performance of the model will not continue to increase with the increase of the network width, and the model can only perform best when the width of the model is suitable.

For reasons of accelerate the network training speed, MobileNetV1 [7] utilizes the concept of deep separable convolution. SqueezeNet [8] reduces the computational cost and improves the network speed by squeezing and expanding. ShuffleNet [9] uses point-by-point group convolution and channel reorganization to construct an extremely efficient CNN architecture. An automatic neural architecture search method for the mobile terminal model is utilized in MnasNet [10], its training speed is 1.5 times faster than MobileNetV1. Howard A [11] combines the advantages of MobileNetV1 and MnasNet to build MobileNetV3, which improves the training efficiency further. GhostNet [12] is a lightweight neural network, which greatly improves the network efficiency through ghost convolution. The above methods all effectively reduce model parameters and save calculation cost through the new type of convolution. However, due to the limitation of network depth or width, there is still much room for improvement in classification accuracy.

Focusing on extracting feature information useful for classification and eliminating redundant information can improve classification accuracy and efficiency. Therefore, in recent years, scholars have begun to explore ways to improve model performance by adding attention mechanisms to the network. Convolutional block attention module (CBAM) [13] is a convolution block based attention mechanism, which significantly improves the accuracy of image classification. SqueezeNet [8] uses an effective attention mechanism to learn attention of the channel and achieves good results. The efficient channel attention (ECA) [14] module focuses on the channel correlation of high-dimensional information, and effectively improves the classification accuracy without additional calculation. It can be seen that the classification effect can be further improved by adding an appropriate attention mechanism.

In summary, replacing traditional convolution with a new type of convolution can effectively reduce parameters and accelerate training of network. Broadening width can enhance the feature extraction capability and improve the classification effect of the network. The addition of attention mechanism can help the model to focus on the extraction of

more useful feature information for the final classification, so as to further improve the classification effect. However, the three effective methods can only improve the performance of a certain aspect respectively, and have not been integrated. We are committed to integrating the improvement of various capabilities to design a high-efficiency and high-precision classification model. Thus, we propose a multi-attention ghost residual fusion network (MAGR) for image classification. The main innovations include the following three aspects.

(1) Introduce ghost convolution into ghost residual mapping network (GRM). Establish the “ghost” mapping relationship between similar images by utilizing the linear operation of ghost module, to realize the full use of redundant feature information between similar images, thereby greatly reducing the model parameters and speeding network training.

(2) The residual connection is introduced into the multi-branch group convolution ghost residual blocks (MGR-Blocks) to broaden the network width, and enhance the feature extraction ability, thereby improving the classification accuracy.

(3) Integrate CBAM and ECA at different stages of the network to strengthen the attention to different channels and spatial feature information of images. CBAM helps to improve the ability of extracting basic feature information, while ECA is responsible for the information interaction between high-dimensional channels and helps the model obtain feature information that is conducive to improving the classification accuracy.

The paper is organized as follows. Section 2 introduces related work. Section 3 describes the proposed MAGR architecture. Section 4 analyzes the structure of MAGR and presents experimental results on three public datasets. Conclusion and prospect are drawn in Section 5.

II. RELATED WORK

A. GHOST CONVOLUTIONAL LAYER

Excellent CNN models, such as AlexNet [1], VggNet [2], ResNet [7], all have high classification accuracy, but there is redundancy of feature maps, which will inevitably affect the network speed. Few people consider the problem of redundancy in the model structure design. Kai Han *et al.* [12] start from the redundancy problem of feature maps, propose a structure -- ghost module, which can generate a large number of feature maps only through a small amount of cheap operations. Compare with the traditional convolution process, ghost convolution has obvious advantages of simplicity and speed. GhostNet [12] introduces ghost into the residual network for the first time, and realizes the extraction of feature information by replacing the traditional convolution of each layer in the residual block. Achieving rapid feature information extraction by using ghost convolution to replace the traditional convolution layer, this idea can be used to build new CNN classification models.

B. MULTI-BRANCH GROUP CONVOLUTIONAL RESIDUAL

The model’s fitting ability and expression ability of the complex objective function increase with the deepening of the network, however, the increase of the depth will bring about difficulty in model training and substantial increase in calculations. ResNeXt [6] is a multi-branch group convolutional network, which proves that widening the network width to improve the model classification effect is more cost-effective than deepening the network depth. This is because widening the network width can effectively enhance the feature extraction capabilities of the model, that is, the model can learn richer feature information, such as textures and colors in different directions and frequencies. If the network width is not enough, the information that each layer can capture is limited. In this case, even if the network is deep enough, it is impossible to extract enough information to transfer to the next layer. ResNext did not analyze the relationship between network width and depth. However, incorporating the residual structure into ResNeXt can help the model quickly implement feature extraction at each layer, while learning enough feature information. The residual connection directly connects the input and output of each multi-branch group convolution structure, effectively improves the flow of information within the network, and finally enables the network to transfer the basic feature information to the subsequent classifier as much as possible, which helps the classifier achieves better classification effect.

C. ATTENTION MECHANISM

The attention mechanism is a potential means to enhance the classification ability of CNN. Adding appropriate attention mechanisms at different locations of the network can strengthen the network’s attention to different degrees of image feature, thereby enabling the model to perform different levels of extraction according to the importance. SqueezeNet [8] proposes an effective mechanism to learn channel attention and achieves good results. However, a large number of attention mechanisms will inevitably lead to an increase in computation. The development of attention mechanism can be divided into two directions: one is to enhance feature aggregation, the other is to realize the combination of channel and spatial attention mechanisms, and the latter has attracted much attention from scholars. CBAM [13] not only considers the importance of different feature channels, but also considers the spatial importance of different positions of the same feature channel. CBAM improves the convolution layer’s attention and main feature extraction ability by organically combining channel attention mechanism and spatial attention mechanism, thereby achieves higher classification accuracy. ECA [14] is a local cross-channel interaction mechanism without dimension reduction. It captures local cross-channel interactions by considering each channel and its several neighbors.

In summary, ghost convolution can solve the problem of large computation cost of traditional convolution. The multi-branch convolution residual can enhance the capability of

feature extraction and improve the information flow within the network. CBAM helps to extract useful feature information with emphasis in the basic feature extraction stage. ECA facilitates the interaction of high-dimensional channel information, which improves the model classification accuracy while effectively reducing the complexity of the network. The above method is very effective for designing a classification network with high precision and fast training. It will inevitably produce adverse effects while a single improvement scheme brings favorable results to the network, so it is necessary to consider the combination of multiple schemes.

III. MULTI-ATTENTION GHOST RESIDUAL FUSION NETWORK

The structure of MAGR is shown in Fig.1, which uses a cascading structure, first extracts the basic feature information, then extracts the advanced feature information, and finally outputs the classification results. MAGR consists of three parts, i.e., basic feature extraction network (BFE), ghost residual mapping network (GRM) and image classification network (IC). The BFE extracts basic features of the input image by using the feature extraction layer and sends them to GRM. GRM uses a series of multi-branch group convolutional ghost residual blocks (MGR-Block) to extract advanced features. The IC judges the category according to all the extracted feature information and finally obtains the corresponding label of the input image.

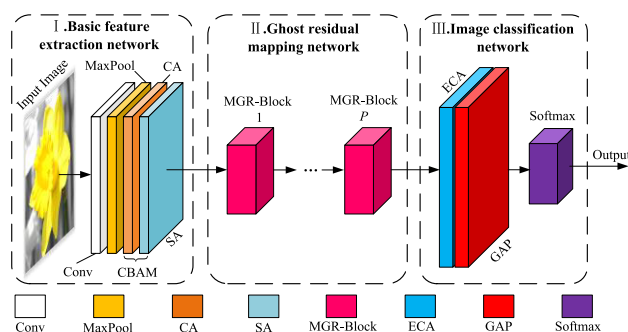


FIGURE 1. The structure of MAGR network.

A. THE STRUCTURE OF BFE

Traditional convolution extraction image feature will result in omission of detailed information. CBAM can help the convolution layer to extract feature information which is helpful for classification. The input image passes through a convolution layer and a maximum pooling layer in turn, then, the obtained feature map x is sent to CBAM. The structure diagram of CBAM is shown in Fig.2. In channel attention (CA) module, the height and width of the feature map x are sent to the shared multilayer perceptron (MLP) after global max pooling and global average pooling respectively. The output features of MLP are added based on element-wise, and then activated by sigmoid to generate $M_c(x)$. In the spatial attention (SA) module, channel-based global max pooling and global average pooling are carried out on the input x' ,

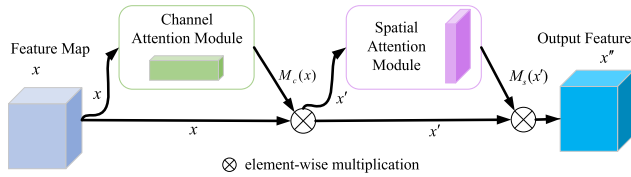


FIGURE 2. Overall structure diagram of CBAM.

and then, the two results are concatenated based on channel. After a convolution operation, the dimensions are reduced to one channel, then activated by sigmoid to generate.

The transfer process of feature map x in CBAM is that, sent x to CA module to get the weighted processing result $M_c(x)$, which is element-wise multiplied with x to obtain x' . Then, x' is weighted by SA to obtain $M_s(x')$, which is element-wise multiplied with x' to get x'' , that is the extracted basic feature information. The calculation process of feature extraction in CBAM as follows,

$$\begin{aligned} x' &= M_c(x) \otimes x \\ x'' &= M_s(x') \otimes x' \end{aligned} \quad (1)$$

where is $x'' \in R^{c \times h \times w}$ the basic feature information that extracted from the input image and will be input to GRM. c , h and w are channels number, height and width of the input image, respectively. \otimes represents the element wise operation, that is, multiplying the corresponding elements one by one.

The focus of CA is on how to learn what is more meaningful in the input image. CA extracts information between different channels with special emphasis by compressing the spatial dimension of the input feature map. SA is a supplement to CA, which focuses on the spatial information part of “where”, and helps the network to extract more useful information in the BFE stage.

B. THE STRUCTURE OF GRM

Ghost residual mapping network (GRM) is designed by combining ghost convolution, MGR-Block and residual connection, to realize advanced feature information extraction through ghost residual mapping.

1) GHOST CONVOLUTION

Ghost convolution layer consists of two parts. The first part generates feature maps with fewer channels by traditional convolution. The second part generates more feature maps by linear operation using the results of the first part. The two groups of feature maps are concatenated together to get the final output [12]. Ghost convolution can use fewer filters to generate more feature maps, and the realization process is shown in Fig.3. After the output x'' of BFE is sent to the ghost convolution layer, the ghost convolution layer first uses fewer feature information in x'' to generate the intrinsic feature maps, and then uses linear transformation to generate ghost feature maps similar to intrinsic feature maps. These ghost feature maps like the “ghost” of the intrinsic feature maps, which is also called ghost mapping.

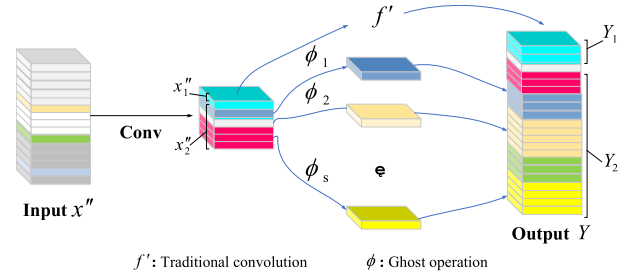


FIGURE 3. Ghost convolution implementation process.

Suppose $x'' = x''_1 + x''_2$ ($x''_1 < x''_2$), x''_1 and x''_2 are useful basic feature information and redundant basic feature information, respectively. x''_1 is used to generate m intrinsic feature maps Y_1 . For each intrinsic feature map, linear operation ϕ_j is used to generate s ghost feature maps. So, m intrinsic feature maps generates $n = m \times s$ ghost feature maps Y_2 . After ghost convolution operation, $m + n$ output feature maps Y is obtained. The mathematical model of ghost convolution operation is shown as follows,

$$\begin{aligned} Y_1 &= x''_1 * f' \\ Y_2 &= y_{ij} = \phi_j(y_i), \quad \forall i = 1, \dots, m, j = 1, \dots, s \\ Y &= Y_1 + Y_2 \end{aligned} \quad (2)$$

where $f' \in R^{c \times k \times k \times m}$ is the filter, $k \times k$ is the size of the convolution kernel, m is the number of intrinsic feature maps $Y_1 \in R^{h' \times w' \times n}$, h' and w' are height and width of the output feature map, respectively, and n is the number of ghost feature maps.

The biggest difference between ghost convolution and traditional convolution is that linear transformation is used by ghost convolution to replace most traditional convolutions, which greatly reduces the computation of convolution process and speeds up the network training.

2) THE STRUCTURE OF MGR-BLOCK

The GRM is cascaded by P MGR-Blocks. The structure of MGR-Block is shown in Fig.4, which is cascaded by a dimension reducer and M ghost residual sub-networks (GRS). Firstly, the dimension reducer is used to double the channels number, and reduce the length and width to one half of original. By setting the convolution parameters in GRS, the input and output feature maps of residual blocks are consistent in size, and then concatenating to avoid the gradient disappearance and degradation of deep network. GRS utilizes residual connection to transmit information, and each GRS is cascaded by 1×1 convolution, 3×3 convolution, and 1×1 convolution in turn. After each convolutional layer, the batch normalization layer (BN layer) and the ReLU activation function layer are sequentially added. The BN layer is usually added before the activation function. The main function is to normalize the input of the activation function, so as to prevent the deviation or enlargement. Every convolution layer is followed by a ReLU activation layer to increase the nonlinearity of the neural network model.

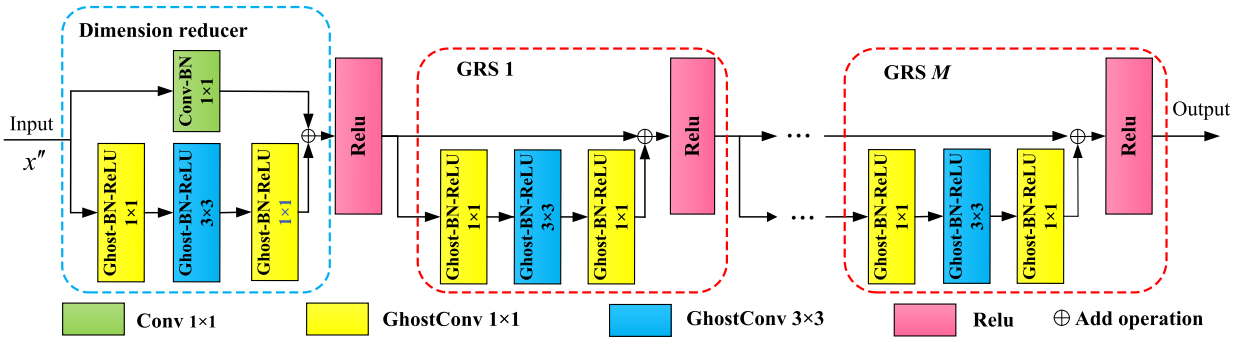


FIGURE 4. The structure of MGR-Block.

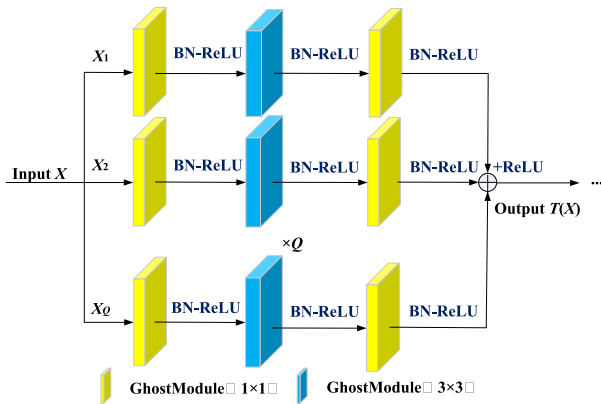


FIGURE 5. GRS internal convolution structure diagram.

The input of the first 1×1 convolution layer is directly connected with the output of the last 1×1 convolution layer, and input to the next GRS module after activated by ReLU, thus, circulates to the n -th GRS.

Each convolution layer of the GRS is divided into Q branches to form a multi-branch group convolution. The structure is helpful for the convolution layer to extract richer feature information from input X and improve classification accuracy of MAGR. The detailed structure of the convolution layer in GRS is shown in Fig.5. Its mathematical model as follows,

$$T(X) = \sum_{i=1}^Q T(X_i) \quad (3)$$

where the input $X = x''$, the input X is divided into Q , $Q \geq 1$ inputs X_i , $i = 1, \dots, Q$, and $T(X_i)$ represents the mapping result of the i branch.

According to Eq. (3), the mathematical model of the k , $k = 1, \dots, P$ MGR-Block can be obtained as follows,

$$T_k(X) = \sum_{j=1}^M \sum_{i=1}^Q T(X_i) \quad (4)$$

where $j = 1, \dots, M$, M is the number of GRS contained in the k -th MGR-Block.

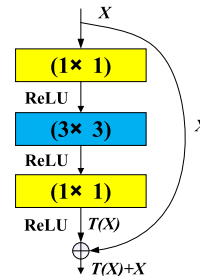


FIGURE 6. Residual connection.

Using MGR-Block to extract image features can help each ghost convolution layer learn enough image features to enhance the feature extraction capability of the network. Ensure that more complete image features are passed to the classification layer, thereby helping the classification layer complete the classification task more accurately.

3) RESIDUAL CONNECTION STRUCTURE

GRS uses residual connection internally to transfer input directly to the output layer. A GRS input is divided into Q inputs. On the one hand, each input is transmitted forward as shown in Fig.5, and on the other hand, it is directly transmitted to the output layer with the help of the residual connection shown in Fig. 6. Thus, the final mathematical model of GRS can be obtained as follows,

$$T_{Fin}(X) = \sum_{i=1}^Q (T(X_i) + X_i) \quad (5)$$

Then the final mathematical model of the k -th ($k = 1, \dots, P$) MGR-Block is,

$$T_{kFin}(X) = \sum_{j=1}^M \sum_{i=1}^Q (T(X_i) + X_i) \quad (6)$$

Thus, we can obtain the output of GRM,

$$T_{PFin}(X) = \sum_{j=1}^M \sum_{i=1}^Q (T_{(P-1)Fin}(X_i) + X_i) \quad (7)$$

With the help of ghost convolution, GRM uses redundant information between similar images to generate ghost mapping linearly, thus, realizes fast and comprehensive learning of the same category image's features. Within each independent residual module, convolution kernels of different sizes are used to extract feature information gradually, and variable step convolutions are utilized to achieve dimension increase or decrease of image channel information. The information flow in the front-back layer is supervised and strengthened by residual connection.

C. THE STRUCTURE OF IC

Fig.7 is the structure of IC, including ECA attention module, global average pooling (GAP) and SoftMax layer in turn. The ECA attention module is responsible for strengthening the connection between the high-dimensional channel information that output by the GRM, which helps the model to further extract useful feature information without increasing the calculation cost. GAP is used to replace the pooling of global feature information in the full connectional layer, which can not only prevent overfitting, but also reduce the calculation cost in the classification stage. The classifier SoftMax is utilized in the last layer of MAGR to perform the final classification task.

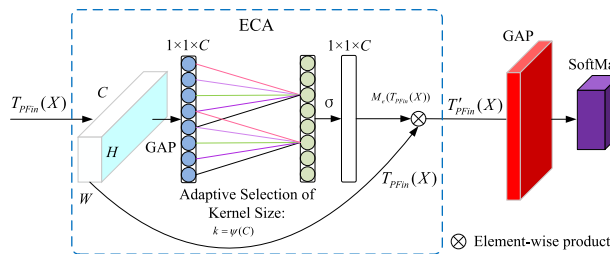


FIGURE 7. Image classification network.

Input GRM's output $T_{PFin}(X)$ to the IC, $T_{PFin}(X)$ enters the ECA module firstly, where global average pooling is performed channel by channel without reducing the dimension. A $1 \times 1 \times C$ feature vector is generated at the GAP layer, and then the cross-channel information interaction is completed through a one-dimensional convolution layer, to obtain the second $1 \times 1 \times C$ feature vector. The kernel size of one-dimensional convolution is determined by an adaptive function, which enables the layer with a large number of channels to conduct more cross-channel interaction. The calculation formula of the adaptive convolutional kernel size as follows,

$$k = \psi(C) = \frac{1}{2} \lceil 1 + \log_2(C) \rceil \quad (8)$$

where, C represents the channel dimension, which is used to determine the size of the convolution kernel. The kernel size k represents the coverage of network cross-channel interaction, and the coverage increases proportionately to the channel dimension.

For convenience, the mapping process of ECA in Figure 7 is represented by M_e , so the output $T'_{PFin}(X)$ of ECA module can be expressed as,

$$T'_{PFin}(X) = M_e(T_{PFin}(X)) \otimes T_{PFin}(X) \quad (9)$$

where, \otimes has the same meaning with Eq. (3).

Then $T'_{PFin}(X)$ is sent to gap layer to pool the global average value of each input feature map, thus, each feature map corresponds to a feature point. Finally, the feature vector composed of all feature points is sent to SoftMax layer to realize the final classification.

IV. SIMULATION EXPERIMENT ANALYSIS

We conduct experiments in the environment Pytorch 1.2.0 on the PC with NVIDIA GTX 2060. When training the MAGR classification model, the number of iterations is set to 120 epochs, the learning rate is initialized to 0.1. Using stochastic gradient descent and momentum methods training, the learning rate is attenuated to one-tenth of the original every 30 epochs.

We select three classical datasets, i.e., CIFAR-10 [15], CIFAR-100 [15] and UC-M [16] for our experiments. The datasets CIFAR-10 and CIFAR-100 are both 32×32 color images, containing 10 and 100 categories respectively. The dataset UC-M contains 21 categories, including 2100 remote sensing images of 256×256 . Parameters, floating point calculations (FLOPs) and accuracy are used as evaluation indicators of the model. The lower of the first two indexes is better, while the higher of the last index is better.

A. MODEL ANALYSIS

MAGR integrates the ghost module, MGR-Block and attention mechanism. It is necessary to determine the replacement scheme of ghost module, the number of MGR-Block cascades, the number of GRS branches and the addition scheme of attention mechanism. During the experiment, the batch size of MAGR is set to 128.

1) REPLACEMENT SCHEME OF GHOST MODULE

The MAGR model includes three parts, that is, BFE, GRM and IC. Among them, BFE contains one convolution and one maximum pooling, GRM contains multiple convolutions. There are two replacement strategies. One is to replace the convolution and maximum pooling in BFE with ghost convolution, the other is to replace all convolutions in GRM with ghost convolution. According to the combination of BFE and GRM, carry out four experiments, i.e., (a) Replace only BFE, (b) Replace only GRM, (c) Replace BFE and GRM, (d) Replace neither BFE nor GRM, the experiment results are shown in Table 1. The results show that on datasets CIFAR-10 and CIFAR-100, compared with (d), the classification accuracy of (a) is improved, but the amount of parameters and FLOPs are not changed. The parameters and FLOPs of (b) and (c) are equal and the lowest, but (b) obtain the highest classification accuracy. Therefore, the (b) scheme is selected for convolution layer replacement.

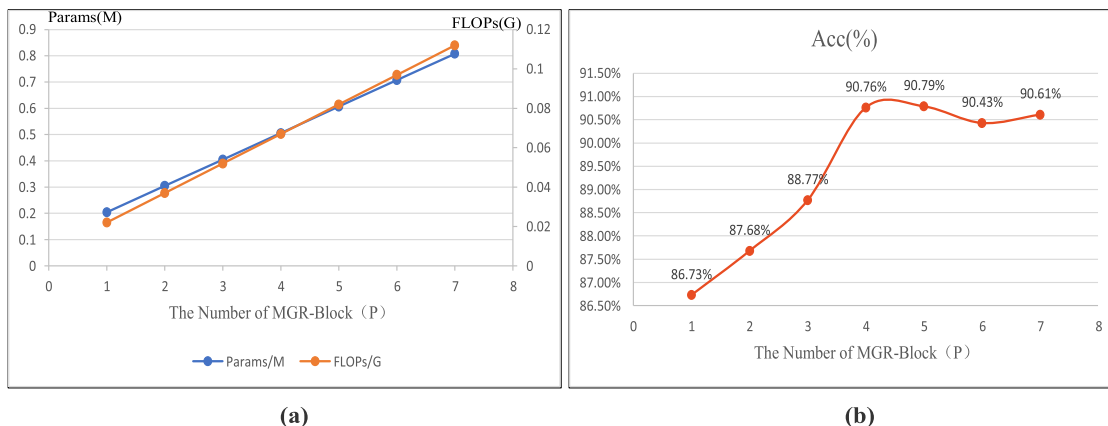


FIGURE 8. Comparison experiment of the number of MGR-Block cascades on dataset CIFAR10.

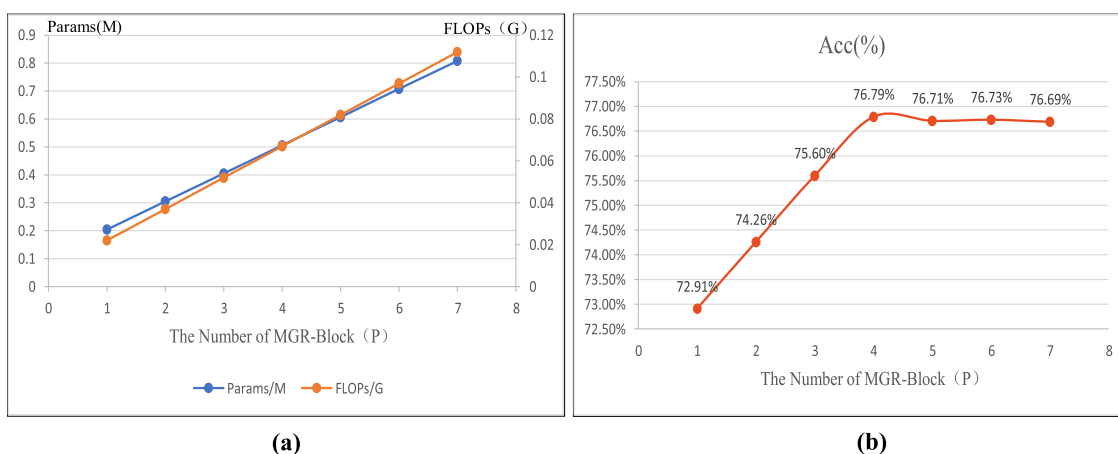


FIGURE 9. Comparison experiment of the number of MGR-Block cascades on dataset CIFAR100.

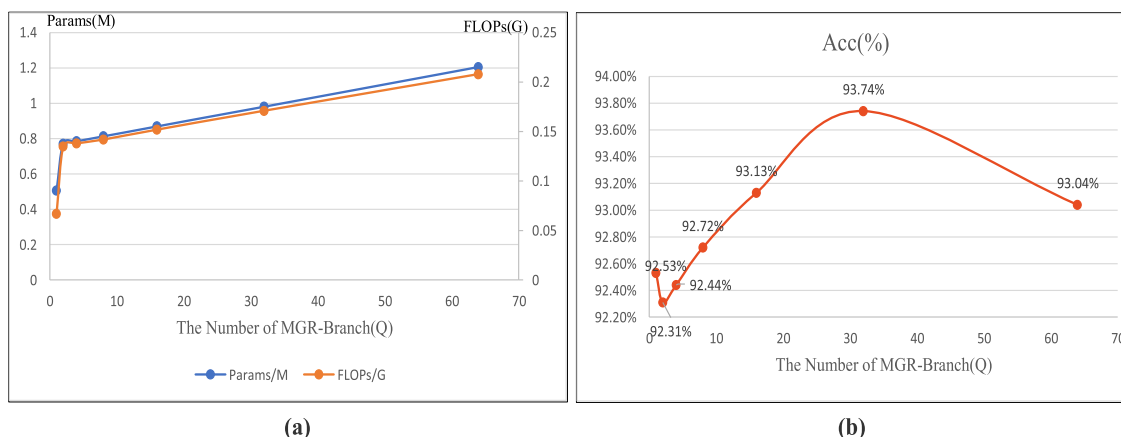


FIGURE 10. Comparison experiment of the number of GRS branches on dataset CIFAR10.

2) THE NUMBER OF MGR-BLOCK CASCADES AND GRS BRANCHES

GRM is formed by cascading P MGR-Blocks, and each GRS contains Q branches. P and Q represent the depth and width of MAGR network respectively. If the model is too deep or too wide, it will lead to training difficulties, which goes against the design intention of lightweight and efficient. Therefore,

we set the value range of P as [1], [7], and set the value range of Q as [1], [64].

First set Q as the default value 1, and P change from 1 to 7, test the values of parameters, FLOP and classification accuracy on datasets CIFAR-10 and CIFAR-100. The experimental results are shown in Figure 8 and Figure 9, where the abscissa is the cascade number P of the MGR-blocks.

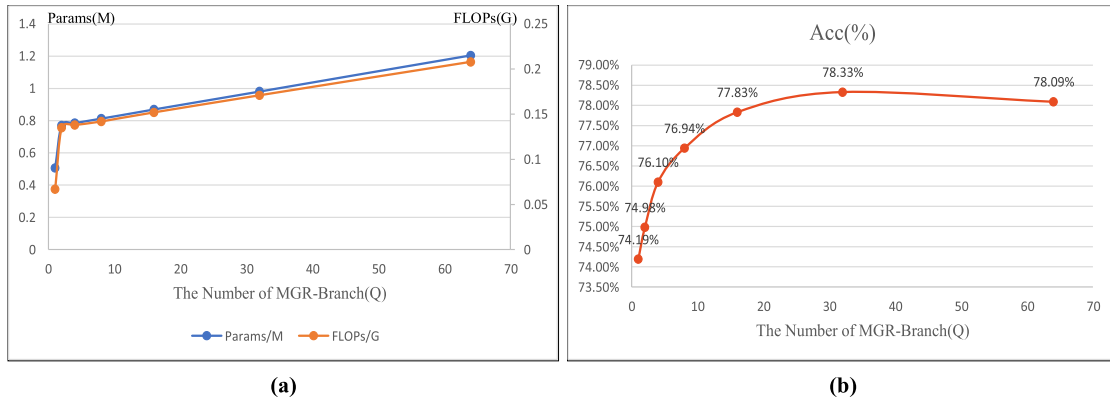


FIGURE 11. Comparison experiment of the number of GRS branches on dataset CIFAR100.

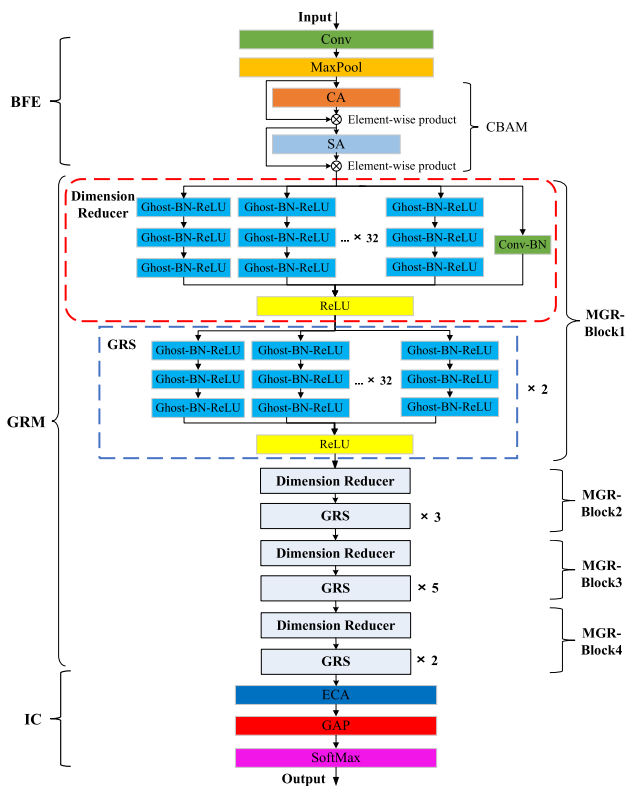


FIGURE 12. The final MAGR network.

The left and right ordinates in Figure 8 (a) and Figure 9 (a) represent parameters and FLOPs, respectively. The ordinates in Figure 8 (b) and Figure 9 (b) represent the classification accuracy. It can be seen that the number of parameters and FLOPs increase linearly with P from 1 to 7, the classification accuracy increases nonlinearly with P from 1 to 4, and decreases nonlinearly with P from 5 to 7. The overall performance of MAGR is the best when the number of MGR-Blocks is 4, therefore, we set $P = 4$ in the subsequent experiments. The number of GRS contained in each MGR-Block is not exactly the same. Motivating by the number of residual sub-blocks in ResNet50, we take 3, 4, 6, and 3 GRSs in turn in the 4 MGR-Blocks, where the

TABLE 1. MAGR performance of ghost convolution replacing convolution at different positions.

Data Result	CIFAR-10			CIFAR-100		
	Params/M	FLOPs/G	Acc/%	Params/M	FLOPs/G	Acc/%
(a)	25.55	0.083	91.47	25.55	0.083	61.68
(b)	0.50	0.067	92.73	0.50	0.067	62.37
(c)	0.50	0.067	91.78	0.50	0.067	62.13
(d)	25.59	0.083	91.44	25.59	0.083	61.19

TABLE 2. Attention feasibility comparison experiment.

Result	Dataset	CIFAR-10		
		Params/M	FLOPs/G	Acc
(a)		0.7842	0.1104	93.47%
(b)		0.7851	0.1105	93.61%
(c)		0.7848	0.1104	93.97%
(d)		0.7853	0.1106	94.73%

dimension reducer is regarded as a special GRS, to realize the establishment of GRM network.

To determine the value of Q , we set P as 4, the number of branch Q in GRS ranges from 1 to 64, and test the values of parameters, FLOP and accuracy on datasets CIFAR-10 and CIFAR-100. The experimental results are shown in Figure 10 and Figure 11, where the abscissa is the number of branches Q . The ordinate of Figure 10 and Figure 11 has the same meaning as that of Figure 8 and Figure 9. It can be seen that parameters and FLOPs increase linearly with Q from 1 to 64. Figure 10 (b) show that the classification accuracy decreases first and then increases and then decreases with the increase of Q . We can find from Figure 11 (b), the classification accuracy of MAGR decreases first and then increases. On datasets CIFAR-10 and CIFAR-100, MAGR achieves the highest classification accuracy when Q is 32. Therefore, Q set 32 in the following experiments.

Figure 8~11 show that with the increase of P and Q , the depth and width of the network increases gradually.

TABLE 3. Performance comparison with other models on CIFAR-10.

Model	Params(M)	FLOPs(M)	Acc(%)
VGG-16 ^[2]	15.0	313	93.6
MnasNet ^[10]	12.7	-	80.8
Ghost-VGG-16 ^[12]	7.7	158	93.7
Ghost-ResNet-56 ^[12]	0.43	63	92.7
R-MnasNet ^[17]	3.0	-	91.3
Dagger(VggNet) ^[18]	2.7	119	93.9
L1-VGG-16 ^[19]	5.4	206	93.4
L1-ResNet-56 ^[19]	0.73	91	92.5
SBP-VGG-16 ^[20]	-	136	92.5
DenseNet(k=24) ^[21]	27.2	-	94.2
VGG16+SD ^[22]	-	-	89.0
ResNet-56(CRA) ^[23]	0.92	126	94.3
ResNet-20(t=3) ^[24]	-	-	93.3
DNN-VGG16 ^[25]	-	-	93.1
MAGR _(our)	0.78	110	94.7

TABLE 4. Performance comparison with other models on CIFAR-100.

Model	Params(M)	FLOPs(M)	Acc(%)
ResNeXt-164 ^[6]	1.70	260	75.0
ShuffleNet ^[9]	0.91	161	69.0
SqueezeNet (RMAF) ^[26]	-	-	68.7
MobileNetV2 ^[27]	1.18	158	68.1
VGG16-half ^[28]	5.40	225	68.8
SSC-Net-6-9 ^[29]	1.15	-	75.9
DWConvXSepConv ^[30]	1.59	16.8	74.0
MFR-DenseNet-100 (k=8) ^[31]	6.29	31.7	76.3
VGG19-S-GD ^[32]	3.20	161	73.6
ResNet-164-S-GD ^[32]	0.66	92	77.4
FR-ResNets (135) ^[33]	1.70	-	75.1
MAGR _(our)	0.78	110	78.7

The increase of depth and width not only improve the classification accuracy, but also bring a larger number of parameters and FLOPs cost, thus increasing the time of model training. Considering the performance and calculation cost of MAGR, $P = 4$ and $Q = 32$ are determined.

3) THE EFFECT OF ATTENTION MECHANISM

One of the characteristics of MAGR networks is lightweight. Therefore, when adding attention, it is necessary to ensure that the performance of the model improves while avoiding the additional computational cost. There are two strategies for adding attention mechanism, namely adding CBAM attention in BFE module and adding ECA attention in IC module. According to the addition situation, four experiments are carried out, i.e., (a) No attention is added, (b) Add only

CBAM attention, (c) Add only ECA attention, and (d) Add CBAM and ECA attention.

The experimental results are listed in Table 2. It can be seen that the addition of attention mechanism has a weak influence on parameters and FLOPs, but has a great influence on the classification accuracy. Among the four tests, (d) obtains the highest classification accuracy, while (a) gets the lowest. Adding CBAM and ECA improves the classification accuracy by 1.33% compared with not adding them. This experiment proves the feasibility and effectiveness of adding CBAM and ECA attention mechanism to the model.

According to the results of model analysis, the final architecture of MAGR is shown in Figure 12. It is cascaded by BFE, GRM and IC. CBAM and ECA are added to BFE and IC, respectively. GRM is composed of 4 cascading

TABLE 5. Performance comparison with other models on UC-M.

Model	Acc(%)
VGG-16 ^[2]	95.7
MobileNet ^[7]	89.5
SqueezeNet ^[8]	76.1
AlexNet (Fine-tuning) ^[16]	95.7
GoogleNet (Fine-tuning) ^[16]	96.0
Improved CNN-SVM ^[34]	96.4
ResNet (FeatureRCG-SVM) ^[35]	93.8
VGG-VD-16 ^[36]	94.4
Inception_ResNet ^[37]	94.4
MAGR _(our)	96.7

MGR-blocks, each of which contains a dimension reducer. In addition, 2, 3, 5, and 2 GRSS in turn in the 4 MGR-Blocks are utilized. Each GRS contains 32 branches.

B. COMPARED WITH OTHER METHODS

After determining the model structure of MAGR, comparative experiments are carried out on three classic datasets to verify the classification effect of MAGR. Test MAGR on dataset CIFAR-10, and compares with the methods in literatures [2], [10], [12], [17]–[25], the results are shown in Table 3, where boldface indicates the best results, blue indicates the second-best results. It can be seen that the parameters and FLOPs of Ghost-ResNet-56 [12] are both the lowest, but the classification accuracy is not high. Both parameters and FLOPs of L1-ResNet-56 [19] rank second, but the classification accuracy is only 92.5%. MAGR's parameters and FLOPs indicators both rank third, but the classification accuracy is the highest, reaching 94.7%, which is 2.2% and 2.4% higher than Ghost-ResNet-56 and L1-ResNet-56, respectively.

Test MAGR on dataset CIFAR-100 and compares with the methods in literatures [6], [9] and [26~33], the results are shown in Table 4. Table 5 is the test results on dataset UC-M of MAGR and methods in literatures [2], [7], [8], [16] and [34~37]. Bold in Table 4~5 indicate the best results, and blue indicate the second-best results. As can be seen from Table 4, DWConvSepconv [30] obtains the best FLOPs, and RESNET-164-S-GD [32] gets the lowest parameters. MAGR achieves the second least parameters and the highest classification accuracy of 78.4%, which is 6.4% and 1.7% higher than DWConvXSEPCONV and RESNET-164-S-GD, respectively. It can be seen from Table 5 that MAGR achieves the best effect and the classification accuracy reaches 96.7%.

According to Table 3~5, on datasets CIFAR10, CIFAR100 and UC-M, although the parameters and FLOPs of MAGR are not the best, they are in the middle and upper class among the comparative methods, but the classification accuracy of MAGR is the highest. Therefore, the training efficiency of the

MAGR network is superior, and the classification accuracy is significantly more competitive than other CNN models.

V. SUMMARY AND PROSPECT

This paper proposes a multi-attention ghost residual fusion network, which can achieve high-efficiency and high-precision multi-image classification tasks. The introduction of ghost convolution can effectively degrade the computational complexity. Using the multi-branch group convolution residual mapping structure to broaden width can enhance feature extraction ability and generalization ability of the network. Residual connection can broaden the network width and reduce gradient dispersion and explosion problems without significantly increasing the computational cost. Adding CBAM in the BFE stage helps MAGR to effectively extract basic features. Adding ECA in the IC stage helps the classifier analyzes the relationship between the high-dimensional feature channels more effectively, so as to achieve better classification effect. The simulation results show that MAGR can effectively improve the classification effect while achieving network efficiency and lightweight.

There are some shortcomings in the content of this article, which are mainly divided into the following two points.

(1) In order to lighten the model, ghost convolution is used to replace all convolution layers in the GRM structure. There is also the possibility of retaining a part of the convolutional layer without replacing to achieve higher classification accuracy, which is not discussed in this paper.

(2) In the aspect of adding attention mechanism, MAGR only inserts the corresponding attention module at the key position of feature extraction. It is also possible that adding all positions will bring about the possibility of better performance improvement, but using the attention mechanism on large-scale will inevitably bring considerable computational costs, which is not discussed in detail.

ACKNOWLEDGMENT

The authors would like to thank the authors of the literatures compared to MAGR for providing their codes.

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [6] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.
- [7] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [8] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372).
- [9] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
- [10] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2815–2823.
- [11] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
- [12] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.
- [13] S. Woo, J. Park, and J. Y. Lee, *CBAM: Convolutional Block Attention Module*. Cham, Switzerland: Springer, 2018, pp. 3–19.
- [14] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
- [15] X. Liu, Y. Zhang, F. Bao, K. Shao, Z. Sun, and C. Zhang, "Kernel-blending connection approximated by a neural network for image classification," *Comput. Vis. Media*, vol. 6, no. 4, pp. 467–476, Dec. 2020.
- [16] N. K. Uba, "Land use and land cover classification using deep learning techniques," 2019, *arXiv:1905.00510*. [Online]. Available: <http://arxiv.org/abs/1905.00510>
- [17] P. Shah and M. El-Sharkawy, "R-MnasNet: Reduced MnasNet for computer vision," in *Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS)*, Sep. 2020, pp. 1–5.
- [18] X. Su, S. You, T. Huang, H. Xu, F. Wang, C. Qian, C. Zhang, and C. Xu, "Data agnostic filter gating for efficient deep networks," 2020, *arXiv:2010.15041*. [Online]. Available: <http://arxiv.org/abs/2010.15041>
- [19] Z. Liu, M. Sun, T. Zhou, G. Huang, and T. Darrell, "Rethinking the value of network pruning," 2018, *arXiv:1810.05270*. [Online]. Available: <http://arxiv.org/abs/1810.05270>
- [20] Y. He, X. Zhang, and J. Sun, "Channel pruning for accelerating very deep neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1398–1406.
- [21] O. K. Oyedotun, A. E. R. Shabayek, D. Aouada, and B. Ottersten, "Improved highway network block for training very deep neural networks," *IEEE Access*, vol. 8, pp. 176758–176773, 2020.
- [22] L. Qian, L. Hu, L. Zhao, T. Wang, and R. Jiang, "Sequence-dropout block for reducing overfitting problem in image classification," *IEEE Access*, vol. 8, pp. 62830–62840, 2020.
- [23] Y. Shen and Y. Wen, "Convolutional neural network optimization via channel reassessment attention module," 2020, *arXiv:2010.05605*. [Online]. Available: <http://arxiv.org/abs/2010.05605>
- [24] H. Jung, R. Lee, S.-H. Lee, and W. Hwang, "Active weighted mapping-based residual convolutional neural network for image classification," *Multimedia Tools Appl.*, pp. 1–15, Sep. 2020, doi: [10.1007/s11042-020-09808-3](https://doi.org/10.1007/s11042-020-09808-3).
- [25] Z. Shao, J. Yang, and S. Ren, "Increasing trustworthiness of deep neural networks via accuracy monitoring," 2020, *arXiv:2007.01472*. [Online]. Available: <http://arxiv.org/abs/2007.01472>
- [26] Y. Yu, K. Adu, N. Tashi, P. Anokye, X. Wang, and M. A. Ayidzoe, "RMAF: Relu-memristor-like activation function for deep learning," *IEEE Access*, vol. 8, pp. 72727–72741, 2020.
- [27] B. Singh, D. Toshniwal, and S. K. Allur, "Shunt connection: An intelligent skipping of contiguous blocks for optimizing MobileNet-V2," *Neural Netw.*, vol. 118, pp. 192–203, Oct. 2019.
- [28] C. Shen, X. Wang, Y. Yin, J. Song, S. Luo, and M. Song, "Progressive network grafting for few-shot knowledge distillation," 2020, *arXiv:2012.04915*. [Online]. Available: <http://arxiv.org/abs/2012.04915>
- [29] Y. Lu, G. Lu, and B. Zhang, "Super sparse convolutional neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 4440–4447.
- [30] J. Chen, Z. Lu, J.-H. Xue, and Q. Liao, "XSepConv: Extremely separated convolution," 2020, *arXiv:2002.12046*. [Online]. Available: <http://arxiv.org/abs/2002.12046>
- [31] B. Chen, T. Zhao, and J. Liu, "Multipath feature recalibration DenseNet for image classification," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 3, pp. 651–660, 2020.
- [32] Y. Liu, D. Wentzlaff, and S. Y. Kung, "Rethinking class-discrimination based CNN channel pruning," 2020, *arXiv:2004.14492*. [Online]. Available: <http://arxiv.org/abs/2004.14492>
- [33] F. Ren, W. Liu, and G. Wu, "Feature reuse residual networks for insect pest recognition," *IEEE Access*, vol. 7, pp. 122758–122768, 2019.
- [34] X. Sun, L. Liu, C. Li, J. Yin, J. Zhao, and W. Si, "Classification for remote sensing data with improved CNN-SVM method," *IEEE Access*, vol. 7, pp. 164507–164516, 2019.
- [35] M. Wang, X. Zhang, X. Niu, F. Wang, and X. Zhang, "Scene classification of high-resolution remotely sensed image based on ResNet," *J. Geovisualization Spatial Anal.*, vol. 3, no. 2, pp. 1–9, Dec. 2019.
- [36] C. He, B. He, X. Yin, W. Wang, and M. Liao, "Relationship prior and adaptive knowledge mimic based compressed deep network for aerial scene classification," *IEEE Access*, vol. 7, pp. 137080–137089, 2019.
- [37] L. Li, T. Tian, and H. Li, "Classification of remote sensing scenes based on neural architecture search network," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 176–180.

• • •