ABSTRACT It has been demonstrated that adding an attention mechanism to a convolutional neural network enhances network performance. However, in practice, the skewness error during the extraction process affects the feature information owing to the implementation of the global average pooling operation, which eventually lowers the performance of the network design. We think that when the weight is activated, the spatial information is effectively captured while the extraction range of the channel information is narrowed. We also think that because the weight is activated using multiple local extraction substitutions, the influence of skewness error can be effectively reduced. As a result, we suggest dimensionality reduction attention (RA) in this study as a new type of attention mechanism. Dimensionality reduction attention is a method of combining spatial information and channel information into a new feature distribution by aggregating dimensionality reduction of feature information, in contrast to other attentions that act on distinct feature tensors in space and channel. Under the premise of capturing long-distance context information and guaranteeing feature coordinates, it may realize the locality of spatial information through this operation and identify the local information of each channel. To completely display the important feature information, the produced feature maps are then encoded into two complimentary attention maps along the spatial and channel directions, respectively. Our dimension reduction focus is a straightforward and all-encompassing module. We run tests on various datasets using various deep architectures and various class designs. The outcomes of the trial demonstrate that our attention-based approach has clear benefits.

INDEX TERMS Convolutional neural network, attention mechanism, feature extraction network.
and its core function is still global feature extraction with embedded position information. The skewness created during the process does not significantly modify the characteristics. In order to enhance the application performance of the network architecture, we suggest an unique attention mechanism in this study, as seen in Figure 1(d). This mechanism was created with the goal of modifying the skewness error. We divide each channel into two one-dimensional tensors with complementary information based on the dependencies between the contexts in the channels’ connection since the two-dimensional global pooling operation has a significant skewness value. The channel connection minimizes the dimension of the decomposed tensor, hence reducing the skewness. Our approach specifically makes use of two stripe pooling modules to aggregate data in accordance with the relationships across contexts and produce two-dimensional feature maps with dependencies from various viewpoints. Dimensionality reduction is the process of redistributing each 1D feature map along the channel direction to create a 2D image. After dimensionality reduction, the two 2Ds are combined to create a 1D. By differentiating the channel characteristics while preserving the global spatial information, the skewness value is kept within an acceptable range of errors. We refer to the suggested attention approach as dimensionality reduction attention since the dimensionality reduction operation plays a significant role in it.

Our focus on dimensionality reduction maintains the skewness error in an unaffected region, which significantly boosts the network architecture’s performance. The benefits of other attention, such the location information suggested by the architecture in [3], are also included into the dimensionality reduction attention, which is likewise effective. Additionally, our attention approach may be easily integrated into various network designs to highlight advantageous aspects. We draw the conclusion that our dimensionality reduction focus provides a large performance advantage for pretrained models based on the experimental findings in Section IV.

II. RELATED WORK

A. CONVOLUTIONAL NEURAL NETWORKS

The most typical network structure in deep learning technologies is the convolutional neural network. Deep learning only became feasible with the development of convolutional neural networks. LeNet in [4], [5], [6], and [7] was the first to include backpropagation into a convolutional neural network design, enabling deep learning. Since then, the convolutional neural network’s performance has been enhanced based on LeNet. For instance, to deepen the model, the AlexNet presented in [8] and [9] uses ReLU as the activation function and is based on LeNet. In addition, AlexNet is an older model that may be used with a GPU. Then, [10] introduced Vggnet, which could better match features and prevent gradient disappearance since it had a set number of convolution kernels compared to AlexNet.

Inception v1 is described in [11] as an improvement over Vggnet, but it also demonstrates that Inception v1 starts with the issue at hand, enhances the depth and breadth of the whole model, and significantly decreases the number of parameters. The addition of the BN layer and the use of Vggnet to shrink the convolution kernel in the Inception v2 in [12] and [13] decreased the number of parameters and sped up computation. References [12] and [14] proposed the Inception v3 convolutional network, which divides the convolution kernel into a one-dimensional size, to further speed up the model’s calculation.

To speed up and enhance performance, Inception v4 in [15] is integrated with Resnet’s structural design in [13]. The extended network Xception network is suggested as a model for the Inception series ([16]), and it is a famous illustration of effective model parameter usage. However, the MobileNet series suggested by [9], [17], and [18] employs depthwise separable convolution to increase speed as opposed to the Xception network’s intended usage. In addition to these enhancements made to the convolutional neural network model by utilizing its depth, breadth, and cardinality, we think it is preferable to boost its performance by creating a module that reduces the complexity of the distribution.

B. ATTENTION MECHANISM

In a mathematical sense, attention is an algorithm that enlarges relevant information and reduces irrelevant information. Residual Attention, an attention-aware function produced by several attention modules, was introduced in 2017 by [19]. However, [1], [17], [22] developed a straightforward SE attention mechanism in 2018, which is a channel attention mechanism based on the compression of spatial information. The BAM attention method, which makes use of global max pooling to further boost channel attention through dimensionality reduction, was put out by [20] the same year.

Only performing channel attention is insufficient, therefore [2], [23] suggest CBAM, which combines $7 \times 7$ convolution with max pooling and average pooling to produce spatial attention while using channel attention. A dual attention method was developed by [21], employing a location attention module to learn the spatial interdependence of features and constructing a channel attention module to model the channel interdependence. Location information is essential for spatial information. However, this strictly only takes into account information about the immediate area. In 2021, [3], [24] presented coordinate attention, which decomposes spatial data into two feature maps in opposite directions using two 1D global pooling procedures, and then activates by convolution to create weights and actualize location information.

The aforementioned attention mechanism does not really take skewness into account. There is no particular process for the distortion of feature extraction, despite the regularization of the allocated weights. Additionally, the skewness impact is controlled by our suggested dimensionality reduction attention through the dimensionality reduction operation. To actualize the correlation between the space and the channel and lessen feature information distortion during feature
extraction, the channel and space are fused in accordance with the interdependence between the contexts. Our attention strategy outperforms other attention strategies.

III. DIMENSIONALITY REDUCTION ATTENTION MECHANISM

Our attention mechanism design mainly consists of two steps: feature tensor dimensionality reduction and weight generation. Take \( x = [x_1, x_2, \ldots, x_c] \in \mathbb{R}^{c \times h \times w} \) as the input tensor, and output \( y = [y_1, y_2, \ldots, y_c] \) with the same shape as \( x \) after the operation of the dimensionality reduction attention mechanism. Instead of directly addressing the skewness issue, this attention technique reduces the dimension of the feature information to create a feature distribution map that corresponds to the space and the channel. The weights for the spatial features and local channel features are then enabled in accordance with the feature map. As shown in Figure 1, the following are the detailed steps.

A. DIMENSIONALITY REDUCTION OF PARALLEL FEATURES

In order to achieve the goal of feature fusion and simplify nonlinear activation, we employ degrading the 3D relationship between the \( x \)-channels to 2D. Taking the spatial attention as the starting point, we borrow the spatial information mentioned in [3] to locate the position, and decompose the spatial information into \( Z^h_c \in \mathbb{R}^{c \times h \times 1} \) and \( Z^v_c \in \mathbb{R}^{c \times 1 \times w} \) from the horizontal and vertical directions. Based on \( Z^h_c \) and \( Z^v_c \), the series operation is carried out in the horizontal and vertical directions respectively, and a new classification is performed for the original features, and two parallel planes \( F_1 \in \mathbb{R}^{1 \times h \times c} \) and \( F_2 \in \mathbb{R}^{1 \times c \times w} \) are generated. The calculation formula is:

\[
F_1 = [z^h_i, z^v_i] \quad i \in (1, 2, \ldots, c - 1)
\]

\[
F_2 = [z^v_j, z^h_j] \quad j \in (1, 2, \ldots, c - 1)
\]

where \([\cdot]\) represents the concatenation operation, \( F_1 \) and \( F_2 \) are two parallel planes positioned based on spatial information.

Discussion: By switching from a broad extrusion operation on many planes to a precise operation on one plane, the conversion of 3D to 2D enhances the flexibility of feature information location and decreases the difficulty of operation. This is a channel and model space rearrangement.

B. GENERATION OF FEATURE WEIGHTS

To exploit the above generated expressive features, we propose a second transformation, called dimensionality reduction attention generation. Our design follows the following two criteria. First, it has to be easy to operate. Second, it can learn the mutual exclusivity of spatial information and the non-exclusivity between channels. In order to meet these criteria, we performed the following operations. For the classified feature surfaces \( F_1 \) and \( F_2 \), we need to integrate the interaction and information between the two feature surfaces. Before integration, we need to transpose \( F_2 \) to achieve two The layout of the feature maps is the same. Then we used \( 1 \times 1 \) convolution \( G \) to operate. Get

\[
F_3 = G(F_1, F_2)
\]

where \( F_3 \) is the feature map after the convolution operation. In order to enable the feature distribution to optimize the distribution under the condition of maintaining its logical law to prepare for the next step of nonlinearization, we use the BN layer to normalize it according to the characteristics of the feature map to obtain

\[
F_4 = BN(F_3)
\]

Discussion: By switching from a broad extrusion operation on many planes to a precise operation on one plane, the conversion of 3D to 2D enhances the flexibility of feature information location and decreases the difficulty of operation. This is a channel and model space rearrangement.
sigmoid function for the spatial attention to mutually exclude the feature information of the space and excite it in the vertical direction. The following formula is obtained:

\[
M^c = \text{softmax}(F_4) \quad (5)
\]

\[
M^h = \text{sigmoid}(F_4) \quad (6)
\]

\(M^c \in R^{1 \times h \times c}\) and \(M^h \in R^{1 \times h \times c}\) are the channel attention weights and the spatial attention weights with the same shape as \(F_4 \in R^{1 \times h \times c}\), respectively. Then we split \(M^c\) and \(M^h\) along the horizontal and vertical directions, respectively, to convert the plane from 2D to 3D. Broadcast the feature weights of each channel to have the same shape as each channel of \(x\) to get \(M_1 \in R^{c \times h \times w}\) and \(M_2 \in R^{c \times h \times w}\). Finally, the \(y\) output by the dimensionality reduction attention mechanism can be written as

\[
y = M_1 \otimes x + M_2 \otimes x \quad (7)
\]

**Discussion:** Activation as a weight designating the significance of information about X characteristic. The channel and the space are entirely linked as a result of the dimensionality reduction attention mechanism’s total integration of the input parallel features from various angles, which also realizes skewness control of the feature information, increasing the accuracy of the feature weight.

**IV. EXPERIMENTAL PART**

The experimental design is initially presented in this section, after which we conduct a number of ablation experiments to assess the performance impact of each component of the dimensionality reduction attention mechanism on the system as a whole. We compare our approach to other attention-based approaches after completing ablation experiments. We conclude by summarizing the outcomes of our approach and other attention-based techniques for object and picture detection.

**A. EXPERIMENTAL SETUP**

We first implement each of our experiments using PyTorch. All of the models were trained using the standard SGD optimizer. For this optimizer, the values for decay and momentum, weight decay, and starting learning rate are set at 0.9, \(5 \times e^{-4}\), and 0.01, respectively. Until the model training is finished after 70 epochs, we employ two NVIDIA GPUs to train models using the Resnet-34 and Resnet-50 baselines in batch 128.

**B. ABLEATION EXPERIMENT**

We conduct a series of ablation tests using Resnet-34 as the baseline to prove the effectiveness of our suggested dimensionality reduction approach. The associated results are displayed in Table 1. Under the premise that Resnet-34 + space means we only examine dimensionality reduction space attention, Resnet-34 + channel implies we only consider dimensionality reduction channel attention. The top-1 is 91.80 when the attention mechanism is not introduced, and it is 91.89 when we merely insert the spatial attention for dimensionality reduction, as shown in Table 1, demonstrating that the performance of the network design can be enhanced by the addition of the attention mechanism. The top-1 is 91.99 after the dimensionality reduction channel is inserted in the attention part, showing that the skewness of the channel attention portion has improved further. When spatial and channel attention are combined into the network design, the top-1 is 92.10, demonstrating that the outcomes are most effective when the feature information is refined.

**C. COMPARISON WITH OTHER ATTENTION METHODS**

1) **TAKE THE Resnet-34 NETWORK AS THE BASELINE**

We use Resnet-34 as the baseline and compare dimensionality reduction attention to different attention techniques. SE attention, CBAM, and CA are often utilized in Table 2. Comparing it to the baseline Top-1, it was discovered that after adding SE attention, it improved by 0.09 percent. The fact that it contributes very nothing to CBAM compared to SE attention suggests that the skewness effect is not much altered by CBAM. The Top-1 of the model is improved by 0.19 percent once CA attention is added, but when we examine CA’s makeup, we see that it has no effect on the distortion brought on by skewness. Instead, it improves feature identification accuracy by embedding position information. The network architecture’s dimensionality reduction focus was where we saw the best outcomes. Evidently, dimensionality reduction attention is more able to mine valuable characteristics than other attention techniques. The skewness coefficient is kept in the effective range by the dimensionality reduction attention by discriminatively activating the local characteristics of the channel, and the location information is also preserved. This is why the dimensionality reduction attention may produce superior results. As a result, the dimensionality reduction attention mechanism works quite well.

2) **STRONG BASELINE**

We use Resnet-50 as the baseline to show that dimensionality-reduced attention still has the largest benefit under strong baselines with rising computation and parameter values. The outcomes are displayed in Table 3 after inserting the dimensionality reduction attention and other attention techniques into the baseline for comparison. We provide the training graph, as shown in Figure 2, so that you may more clearly see the changes that occur during the training process. The outcomes of the dimensionality reduction effort are still good enough. This demonstrates that dimensionality reduction attention still functions effectively even with greater processing and parameters.

**TABLE 1.** With Resnet-34 as the baseline, the importance of each function is highlighted according to the functional analogy.

| Settings   | Param. (M) | PLOP (M) | Top-1(%) |
|------------|------------|----------|----------|
| Resnet-34  | 21.797     | 77.24    | 91.80    |
| Resnet-34+channel | 21.798    | 77.30    | 91.89    |
| Resnet-34+space | 21.798    | 77.41    | 91.89    |
| Resnet-34+RA  | 21.797    | 77.38    | 92.10    |

Z. Wang, L. Wang: Attention Approach for Dimensionality Reduction That Can Correct Feature Collection Skewness
### TABLE 2. With Resnet-34 as the baseline, the importance of each function is highlighted according to the functional analogy.

| Settings     | Param.(M) | FLOPs(M) | Top-1(%) |
|--------------|-----------|----------|----------|
| Resnet-34    | 21.797    | 77.724   | 91.80    |
| Resnet-34+SE | 21.798    | 77.30    | 91.89    |
| Resnet-34+CA | 21.798    | 77.41    | 91.99    |
| Resnet-34+RA | 21.797    | 77.38    | 92.10    |

### TABLE 3. Using Resnet-50 as the baseline to compare the fitness of attention methods.

| Settings     | Param.(M) | FLOPs(M) | Top-1(%) |
|--------------|-----------|----------|----------|
| Resnet-50    | 25.557    | 87.98    | 92.00    |
| Resnet-50+CBAM | 25.558  | 88.22    | 92.10    |
| Resnet-50+SE | 25.558    | 88.05    | 92.23    |
| Resnet-50+CA | 25.558    | 88.15    | 92.21    |
| Resnet-50+RA | 25.557    | 88.13    | 92.51    |

### TABLE 4. Comparing the effectiveness of various attention-training techniques using SSD300 as a baseline.

| Settings     | Param.(M) | FLOPs(M) | mAP   |
|--------------|-----------|----------|-------|
| SSD300       | 26.15     | 31.35    | 0.775 |
| SSD300+RA    | 26.15     | 31.35    | 0.781 |
| SSD300+CBAM  | 26.18     | 31.35    | 0.778 |
| SSD300+SE    | 26.18     | 31.35    | 0.776 |
| SSD300+CA    | 26.20     | 31.35    | 0.778 |

3) DIFFERENT NETWORK ARCHITECTURES

We compare dimensionality reduction attention with other attention methods using SSD300 as the baseline to show that our attention approach is as sophisticated across various network designs. As can be seen in Table 4, the performance of the architecture is improved since the mAP increases from 0.775 when the attention approach is not used to 0.776 when it is. However, mAP was 0.778 when both CBAM and CA were inserted, indicating that under some circumstances, the global squeezing and maxima through space in CBAM might be similar to localization of location. Because the effects of skewness was controlled during the feature information collecting stage and the position information was also kept in the activation weights, RA produces the best results when it is used. This demonstrates that paying attention to dimensionality reduction has significant benefits in many designs.

### D. APPLICATION

To investigate the transfer learning potential of dimensionality reduction attention in comparison to other attention techniques, we perform picture classification and object recognition studies in this section.

1) IMAGE CLASSIFICATION

**a: IMPLEMENTATION DETAILS**

Based on PyTorch and Resnet-34, our code. We include the attention approach into the first layer of the convolutional network while maintaining its integrity. We employ the common SGD optimizer while training with cifar-10, with the initial learning rate, weight decay parameter, momentum parameter, and decay and momentum parameters all set to 0.9, 5 * e – 4, and 0.01, respectively. We repeated the training in 70 epoch units with a batch size of 128. We will repeat the training in increments of 100 epochs while using the imagenet dataset.

**b: CIFAR-10 RESULTS**

We monitor the Top-1 and Top-5 accuracy in this investigation. Table 5 displays the detection outcomes for the Cifar-10
After adding various attention techniques, the top-ranked category of detection results is improved by all inserted attention techniques, but dimensionality reduction attention shows the greatest increase of 0.3 percent. The findings are the same whether SE and CBAM are included, demonstrating that raising the maximum value of CBAM has no impact on categorization. However, only RA considerably enhances detection results in the top five categories, whereas CA and CBAM decrease accuracy, proving that RA is highly helpful to model transfer learning for skewness processing in the first data collection. Figure 3 of our presentation displays trends in the error rate and accuracy rate for training and test outcomes.
to other attention approaches, detection experiments on the cifar-10 dataset demonstrate.

**c: IMAGENET RESULTS**
The picture size was set to $84 \times 84$, and the types of image net datasets were changed to 100. We concluded by displaying the image net training results in Table 6. We discovered that when various mechanisms for attention are compared to the impact of dimensionality reduction, as the complexity of the data distribution rises, the attention method’s benefit does not diminish, demonstrating that the transferability of dimensionality reduction attention to classification models does not degrade. As illustrated in Figure 4, we also provide a trend graph along with other attention techniques that may be studied in further depth.

### 2) OBJECT DETECTION

**a: IMPLEMENTATION DETAILS**

Our code is based on Pytorch and SSD300. When training PascalVoc for training, the batch size is set to 8 and the standard SGD optimizer is used with an initial learning rate of 0.001, momentum of 0.9, and weight decay of $5 \times 10^{-4}$. The paper trains for 120,000 iterations with a batch size of 32, decays after 80,000 and 100,000 iterations.

**b: PASCALVOC RESULTS**

For the purposes of this experiment, we use the value of mAP as the foundation for precise detection. Table 7 displays the outcomes of model training using the Pascal VOC 2007 dataset. Under the assumption of the same amount of calculation, it is obvious that adding dimensionality reduction attention to VGG16 may significantly enhance the detection results, and the parameters of the dimensionality reduction attention model can be ignored. The effect of CBAM and CA on the target detection model can be seen in the table, which also demonstrates that the position information is not crucial for the model’s detection but is highly helpful for skewness control when extracting the feature’s starting information. We provide the representations of various attention techniques in Figure 5. This demonstrates the superior transfer learning capability of the convolutional model with dimensionality reduction attention.

**c: MS COCO2017 RESULTS**

We employ a more complicated data set to confirm that the performance enhancement of the network by dimensionality reduction attention is still superior to other attention
techniques, demonstrating that the superior performance of this attention approach in application is unaffected by the data set. Table 8 contains the operational outcomes of the various focus techniques. The results demonstrate that dimensionality reduction attention is still the most accurate, whereas other attention methods’ accuracy will change depending on the distribution of the data. This also demonstrates that the stability of dimensionality reduction attention will not alter the network’s performance impact as a result of the different dataset distribution. For easier viewing, we also display their representations, as seen in Figure 6.

V. CONCLUSION

In order to enhance feature extraction in convolutional neural networks, we suggest a unique strategy called a dimensionality reduction attention mechanism. Through a dimensionality reduction procedure, our attention technique merges features. The performance of the network design is enhanced by limiting the impact of skewness during the feature extraction process. For experimental verification, we have employed the pre-trained models Resnet-34, Resnet-50, and VGG with various parameter settings. All of these models have produced perfect results. We apply the parameters cifar-10 and PascalVoc with various distributions for verification in order to further strengthen our verifiability, and we see that our dimensionality reduction focus still has clear benefits over other types. Finally, we anticipate that, in the case of continuous improvement, the dimensionality reduction attention mechanism can support the feature extraction of convolutional neural networks.

REFERENCES

[1] 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.
[2] S. Woo, J. Park, and J. Y. Lee, CBAM: Convolutional Block Attention Module. Cham, Switzerland: Springer, 2018.
[3] Q. Hou, D. Zhou, and J. Feng, “Coordinate attention for efficient mobile network design,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 13713–13722.
[4] B. Wu et al., “SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud,” in Proc. Int. Conf. Robot. Automat. (ICRA). IEEE, 2019, pp. 4376–4382.
[5] C. Xu, B. Wu, Z. Wang, W. Zhan, P. Vajda, K. Keutzer, and M. Tomizuka, “SqueezeSegV3: Spatially-adaptive convolution for efficient point-cloud segmentation,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, Aug. 2020, pp. 1–19.
[6] A. Milioio, I. Vizzo, J. Behley, and C. Stachniss, “RangeNet++: Fast and accurate LiDAR semantic segmentation,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Nov. 2019, pp. 4213–4220.
[7] J. Behley et al., “SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 9297–9307.
[8] F. N. Indola et al., “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size,” 2016, arXiv:1602.07360.
[9] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” 2017, arXiv:1704.04861.
[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.
[11] C. Szegedy et al., “Going deeper with convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9.
[12] C. Szegedy, V. Vanhoucke, and S. Ioffe, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., May 2016, pp. 2818–2826.
[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Apr. 2016, pp. 770–778.
[14] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. PMLR, May 2019, pp. 6105–6114.
[15] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-V4, inception-ResNet and the impact of residual connections on learning,” in Proc. AAAI Conf. Artif. Intell., Feb. 2017, vol. 31, no. 1, pp. 1–10.
[16] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1251–1258.
[17] 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. 1–5.
[18] A. Howard et al., “Searching for MobileNetV3,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 1314–1324.
[19] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, “Residual attention network for image classification,” 2017, arXiv:1704.06904.
[20] J. Park et al., “BAM: Bottleneck attention module,” 2018, arXiv:1807.06514.
[21] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3146–3154.
[22] N. Ma, ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. Cham, Switzerland: Springer, 2018.
[23] X. Tang, “Residual attention network for image classification,” 2017, arXiv:1704.06904.
[24] D. Ruan, “Gaussian context transformer,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 15129–15138.

ZIYUAN WANG received the B.S. degree in computer science and technology from the Agricultural University of Hebei, Baoding, China, in 2020. He is currently pursuing the master’s degree with the Liaoning University of Science and Technology. His research interest includes computer vision.

LI WANG received the Ph.D. degree from Tianjin University, Tianjin, China, in 2004. She is currently a Professor with the Liaoning University of Science and Technology. She has authored or coauthored more than 50 publications in conferences and journals. Her research interests include image processing and computer vision.

TABLE 8. The results of model training on the Ms coco2017 dataset.

| Settings     | Param (M) | FLOPs(M) | mAP  |
|--------------|-----------|----------|------|
| SSD300       | 26.15     | 31.35    | 0.368|
| SSD300+RA    | 26.15     | 31.35    | 0.373|
| SSD300+CBAM  | 26.18     | 31.35    | 0.370|
| SSD300+SE    | 26.18     | 31.35    | 0.370|
| SSD300+CA    | 26.20     | 31.35    | 0.372|