# Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions

# 全球注意力机制：保留信息以增强通道-空间交互

Yichao Liu

刘毅超

Helmholtz-Zentrum Dresden-Rossendorf

德累斯顿-罗斯托克亥姆霍兹中心

Dresden, Germany

德国德累斯顿

y.liu@hzdr.de

Zongru Shao

邵宗儒

Helmholtz-Zentrum Dresden-Rossendorf

德累斯顿-罗斯托克亥姆霍兹中心

Dresden, Germany

德国德累斯顿

Center for Advanced Systems Understanding

高级系统理解中心

Görlitz, Germany

德国格尔利茨

z.shao@hzdr.de

Nico Hoffmann

尼科·霍夫曼

Helmholtz-Zentrum Dresden-Rossendorf

德累斯顿-罗斯托克亥姆霍兹中心

Dresden, Germany

德国德累斯顿

n.hoffmann@hzdr.de

# Abstract

# 摘要

A variety of attention mechanisms have been studied to improve the performance of various computer vision tasks. However, the prior methods overlooked the significance of retaining the information on both channel and spatial aspects to enhance the cross-dimension interactions. Therefore, we propose a global attention mechanism that boosts the performance of deep neural networks by reducing information reduction and magnifying the global interactive representations. We introduce 3D-permutation with multilayer-perceptron for channel attention alongside a convolutional spatial attention submodule. The evaluation of the proposed mechanism for the image classification task on CIFAR-100 and ImageNet-1K indicates that our method stably outperforms several recent attention mechanisms with both ResNet and lightweight MobileNet.

已经研究了许多注意力机制来提高各种计算机视觉任务的表现。然而，先前的方法忽视了在通道和空间方面保留信息的重要性，以增强跨维度的交互。因此，我们提出了一种全局注意力机制，通过减少信息减少和放大全局交互表示来提升深度神经网络的性能。我们引入了3D排列和多层感知器用于通道注意力，以及一个卷积空间注意力子模块。在CIFAR-100和ImageNet-1K图像分类任务上评估所提出的机制表明，我们的方法在ResNet和轻量级MobileNet上稳定地优于最近的一些注意力机制。

# 1 Introduction

# 1 引言

Convolutional neural networks (CNNs) have been widely used in many tasks and applications in the computer vision domain (Girshick et al. [2014], Long et al. [2015], He et al. [2016], Lampert et al. [2009]). Researchers have found that CNNs are performing well in extracting deep visual representations. With technological improvements related to CNNs, image classification on the ImageNet dataset (Deng et al. [2009]) has increased from to accuracy in the past nine years (Krizhevsky et al. [2012], Zhai et al. [2021]). Such an achievement also attributes to the complexity of the ImageNet dataset, which offers exceptional opportunities for related studies. Given the diversity and large scale of real-life scenes it covers, it has been benefitting studies for conventional image classification benchmarking, representation learning, transfer learning, etc. Particularly, it also brings challenges for the attention mechanisms.

卷积神经网络（CNNs）已经在计算机视觉领域（Girshick et al. [2014], Long et al. [2015], He et al. [2016], Lampert et al. [2009]）的许多任务和应用中得到了广泛应用。研究者们发现CNNs在提取深度视觉表示方面表现良好。随着与CNNs相关的技术改进，ImageNet数据集（Deng et al. [2009]）上的图像分类准确率在过去九年里从 提高到 （Krizhevsky et al. [2012], Zhai et al. [2021]）。这一成就也归功于ImageNet数据集的复杂性，它为相关研究提供了 exceptional 的机会。鉴于它覆盖的现实生活场景的多样性和大规模，它已经有益于传统图像分类基准测试、表示学习、迁移学习等研究。特别是，它也为注意力机制带来了挑战。

The attention mechanisms have been improving performance in multiple applications and attracted research interests in recent years (Niu et al. [2021]). Wang et al. [2017] used an encoder-decoder residual attention module to refine the feature maps to obtain better performance. Hu et al. [2018], Woo et al. [2018], Park et al. [2018] used spatial and channel attention mechanisms separately and achieved a higher accuracy. However, these mechanisms utilize visual representations from limited receptive fields due to information reduction and dimension separation. They lose global spatial-channel interactions in the process. Our research objective is to investigate attention mechanisms across the spatial-channel dimensions. We propose a "global" attention mechanism that reserves information to magnify the "global" cross-dimension interactions. Therefore, we name the proposed method Global Attention Mechanism (GAM).

注意力机制已经在多个应用中提高了性能，并在近年来吸引了研究兴趣（Niu et al. [2021]）。Wang et al. [2017] 使用了编码器-解码器残差注意力模块来细化特征图以获得更好的性能。Hu et al. [2018]，Woo et al. [2018]，Park et al. [2018] 分别独立地使用了空间和通道注意力机制，并实现了更高的准确率。然而，由于信息减少和维度分离，这些机制利用了有限感受野的视觉表示。在这个过程中，它们失去了全局的空间-通道交互。我们的研究目标是调查跨空间-通道维度的注意力机制。我们提出了一种“全局”注意力机制，保留信息以放大“全局”跨维度交互。因此，我们将提出的方法命名为全局注意力机制（GAM）。

# 2 Related Works

# 2 相关工作

There have been several studies focusing on performance improvements of attention mechanisms for image classification tasks. Squeeze-and-Excitation Networks (SENet) (Hu et al. [2018]) is the first to use channel attention and channel-wise-feature-fusion to suppress the unimportant channels. However, it is less efficient in suppressing unimportant pixels. The later-on attention mechanisms considered both spatial and channel dimensions. The convolutional block attention module (CBAM) (Woo et al. [2018]) places the channel and spatial attention operation sequentially, while bottleneck attention module (BAM) (Park et al. [2018]) did it in parallel. However, both of them ignore the channel-spatial interactions and lose the cross-dimension information consequently. Considering the significance of the cross-dimension interactions, the triplet attention module (TAM) (Misra et al. [2021]) boosts efficiency by utilizing the attention weights between each pair of the three dimensions - channel, spatial width, and spatial height. However, the attention operations are still applied on two of the dimensions each time instead of all three. To magnify cross-dimension interactions, we propose an attention mechanism that is capable of capturing significant features across all three dimensions.

已经进行了几项研究，关注于图像分类任务中注意力机制的性能改进。Squeeze-and-Excitation Networks (SENet)（Hu et al. [2018]）首次使用通道注意力和通道特征融合来抑制不重要的通道。然而，它在抑制不重要的像素方面效率较低。后续的注意力机制同时考虑了空间和通道维度。卷积块注意力模块（CBAM）（Woo et al. [2018]）依次执行通道和空间注意力操作，而瓶颈注意力模块（BAM）（Park et al. [2018]）则是并行执行。然而，它们都忽略了通道-空间交互，并因此丢失了跨维度信息。考虑到跨维度交互的重要性，三元注意力模块（TAM）（Misra et al. [2021]）通过利用三个维度（通道、空间宽度和空间高度）之间的注意力权重来提高效率。但是，注意力操作仍然每次只应用于两个维度，而不是全部三个。为了放大跨维度交互，我们提出了一种能够捕捉所有三个维度上显著特征的注意力机制。

# 3 Global Attention Mechanism (GAM)

# 3 全局注意力机制（GAM）

Our objective is to design a mechanism that reduces information reduction and magnifies global dimension-interactive features. We adopt the sequential channel-spatial attention mechanism from CBAM and redesign the submodules. The overall process is illustrated in Fig. 1 and formulated in Equation 1 and 2 (Woo et al. [2018]). Given the input feature map , the intermediate state and the output are defined as:

我们的目标是设计一种机制，减少信息降低并放大全局维度交互特征。我们采用了CBAM中的顺序通道-空间注意力机制并重新设计了子模块。整体流程如图1所示，并在方程1和2中进行了公式化（Woo et al. [2018]）。给定输入特征图 ，中间状态 和输出 定义为：

where and are the channel and spatial attention maps, respectively; denotes element-wise multiplication.

其中 和 分别是通道和空间注意力图； 表示逐元素乘法。

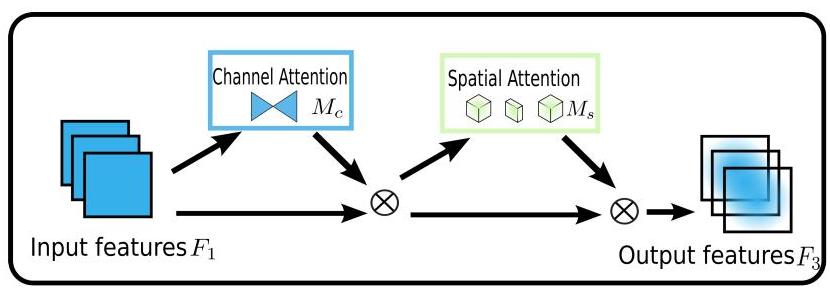


Figure 1: The overview of GAM

图1：GAM的概览

The channel attention submodule uses 3D permutation to retain information across three dimensions. It then magnifies cross-dimension channel-spatial dependencies with a two-layer MLP (multi-layer perceptron). (The MLP is an encoder-decoder structure with a reduction ratio , same as BAM.) The channel attention submodule is illustrated in Fig. 2

通道注意力子模块使用3D排列来保留三个维度上的信息。然后，它用一个双层MLP（多层感知器）放大跨维度的通道-空间依赖性（MLP是一个编码器-解码器结构，具有与BAM相同的降低比例 ）。通道注意力子模块在图2中有所说明。

In the spatial attention submodule, to focus on spatial information, we use two convolutional layers for spatial information fusion. We also use the same reduction ratio from the channel attention submodule, same as BAM. Meanwhile, max-pooling reduces the information and contributes negatively. We remove pooling to further retain the feature maps. As a result, the spatial attention module sometimes increase the number of parameters significantly. To prevent a notable increase of the parameters, we adopt group convolution with channel shuffle (Zhang et al. [2018]) in ResNet50. The spatial attention submodule without group convolution is shown in Fig. 3

在空间注意力子模块中，为了关注空间信息，我们使用了两个卷积层来进行空间信息融合。我们还使用了与通道注意力子模块相同的降低比例 ，与BAM相同。同时，最大池化减少了信息并产生负面影响。我们移除了池化以进一步保留特征图。因此，空间注意力模块有时会显著增加参数的数量。为了防止参数数量显著增加，我们在ResNet50中采用了带有通道混洗的组卷积（Zhang et al. [2018]）。没有组卷积的空间注意力子模块在图3中显示。

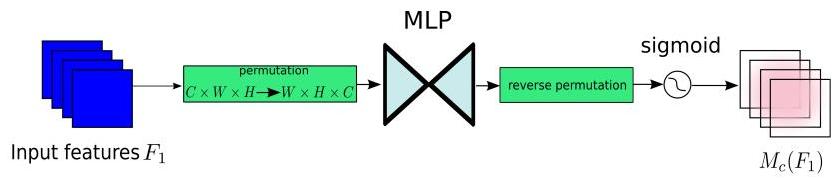


Figure 2: Channel attention submodule.

图2：通道注意力子模块。

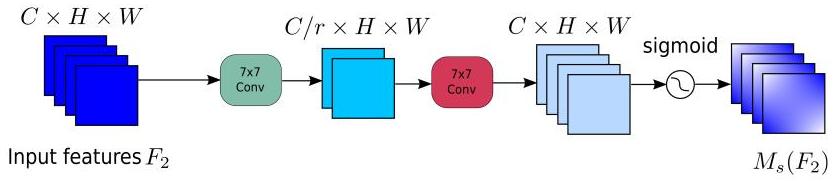


Figure 3: Spatial attention submodule.

图3：空间注意力子模块。

# 4 Experiment

# 4 实验部分

In this section, we evaluate GAM on both CIFAR-100 (Krizhevsky et al. [2009]) and ImageNet-1K datasets (Deng et al. [2009]) with classification benchmarking and two ablation studies. We use two datasets to verify method generalization. Note that both datasets are standard for classification. ImageNet-1K has a higher impact on real-life applications.

在本节中，我们通过分类基准测试和两个消融研究，评估了GAM在CIFAR-100（Krizhevsky et al. [2009]）和ImageNet-1K数据集（Deng et al. [2009]）上的表现。我们使用这两个数据集来验证方法的泛化能力。请注意，这两个数据集都是分类的标准数据集。ImageNet-1K对现实生活应用具有更高的影响力。

# 4.1 Classification on CIFAR-100 and ImageNet datasets

# 4.1 在CIFAR-100和ImageNet数据集上的分类

We evaluate GAM with both ResNet (He et al. [2016]) and MobileNet V2 (Sandler et al. [2018]) as (a) they are standard architectures for image classification (b) they represent the regular and the lightweight networks respectively. We compare GAM against SE, BAM, CBAM, TAM, and Attention Branch Network (ABN) (Fukui et al. [2019]). We re-implement the networks & mechanisms and evaluate them under the same conditions. All models are trained on four Nvidia Tesla V100 GPUs.

我们使用 ResNet（He et al. [2016]）和 MobileNet V2（Sandler et al. [2018]）来评估 GAM（a）因为它们是图像分类的标准架构（b）它们分别代表了常规网络和轻量级网络。我们将 GAM 与 SE、BAM、CBAM、TAM 以及 Attention Branch Network（ABN）（Fukui et al. [2019]）进行比较。我们重新实现了这些网络和机制，并在相同条件下进行评估。所有模型都在四块 Nvidia Tesla V100 GPU 上进行训练。

For CIFAR-100, we evaluate GAM with and without group convolution (gc). We train all networks for 200 epochs with a starting learning rate of 0.1 . Then, we drop the learning rate at the epochs of 60, 120, and 160. The results are shown in Table 1. It shows that GAM outperforms SE, BAM, and CBAM.

对于 CIFAR-100，我们评估了 GAM 在有和无组卷积（gc）的情况下的表现。我们使用初始学习率为 0.1 对所有网络进行 200 个周期的训练。然后，我们在第 60、120 和 160 个周期时降低学习率。结果如表 1 所示。它表明 GAM 的性能优于 SE、BAM 和 CBAM。

Table 1: Classification results on Cifar100

表 1：在 Cifar100 上的分类结果

| Architecture | Parameters | FLOPs | Top-1 Error (%) | Top-5 Error (%) |
| --- | --- | --- | --- | --- |
| ResNet 50 | 23.71M | 1.3G | 22.74 | 6.37 |
| ResNet 50 + SE | 26.22M | 1.31G | 20.29 | 5.18 |
| ResNet 50 + BAM | 24.06M | 1.33G | 19.97 | 5.03 |
| ResNet 50 + CBAM | 26.24M | 1.31G | 19.44 | 4.66 |
| ResNet 50 + GAM | 149.47M | 8.02G | 18.67 | 4.54 |
| ResNet 50 + GAM (gc\*) | 57.05M | 3.08G | 18.99 | 4.87 |

gc stands for group convolution (we set its hyper-parameter as 4).

gc 代表组卷积（我们将其超参数设置为 4）。

For ImageNet-1K, we pre-process the images to (He et al. [2016]). We include both ResNet18 and ResNet50 (He et al. [2016]) to verify method generalization on different network depths. For ResNet50, we include a comparison with group convolution to prevent a notable increase of the parameters. We set the starting learning rate as 0.1 and drop it for every 30 epochs. We use 90 training epochs in total. In the spatial attention submodule, we switch the first stride of the first block from 1 to 2 in order to match the size of the features. Other settings are preserved from CBAM for a fair comparison, including the use of max-pooling in the spatial attention submodule.

对于 ImageNet-1K，我们对图像进行预处理 （He et al. [2016]）。我们包括 ResNet18 和 ResNet50（He et al. [2016]）来验证方法在不同网络深度上的泛化性。对于 ResNet50，我们加入组卷积的比较，以防止参数数量显著增加。我们将初始学习率设置为 0.1，并且每 30 个周期降低一次。我们总共使用 90 个训练周期。在空间注意力子模块中，我们将第一个块的第一个步长从 1 切换到 2，以匹配特征的大小。其他设置保持与 CBAM 相同，以便公平比较，包括在空间注意力子模块中使用最大池化。

MobileNet V2 is one of the most efficient lightweight models for image classification. We use the same setup of ResNet for MobileNet V2 except using an initial learning rate of 0.045 and a weight decay of .

MobileNet V2 是图像分类中最有效的轻量级模型之一。我们为 MobileNet V2 使用与 ResNet 相同的设置，除了初始学习率设为 0.045 和权重衰减为 。

The evaluation on the ImageNet-1K is shown in Table 2. It shows that GAM stably enhances the performance across different neural architectures. Especially, for ResNet18, GAM outperforms ABN with fewer parameters and better efficiency.

在 ImageNet-1K 上的评估结果如表 2 所示。它表明 GAM 稳定地提升了不同神经网络架构的性能。特别是对于 ResNet18，GAM 以更少的参数和更好的效率超过了 ABN。

Table 2: Classification results on ImageNet-1K

表 2：ImageNet-1K 上的分类结果

| Architecture | Parameters | FLOPs | Top-1 Error (%) | Top-5 Error (%) |
| --- | --- | --- | --- | --- |
| ResNet 18 | 11.69M | 1.82G | 30.91 | 11.12 |
| ResNet 18 + SE | 11.78M | 1.82G | 30.07 | 10.59 |
| ResNet 18 + BAM | 11.71M | 1.82G | 30.18 | 10.77 |
| ResNet 18 + CBAM | 11.78M | 1.82G | 29.89 | 10.53 |
| ResNet 18 + TAM | 11.69M | 1.83G | 30.0 | 10.64 |
| ResNet 18 + ABN | 21.61M | 3.76G | 29.4 | 10.34 |
| ResNet 18 + GAM | 16.04M | 2.45G | 29.34 | 10.23 |
| ResNet 50 | 25.56M | 4.11G | 24.81 | 7.69 |
| ResNet 50 + SE | 28.07M | 4.12G | 23.56 | 6.82 |
| ResNet 50 + BAM | 25.92M | 4.19G | 24.0 | 7.01 |
| ResNet 50 + CBAM | 28.09M | 4.12G | 23.1 | 6.57 |
| ResNet 50 + TAM | 25.56M | 4.16G | 23.29 | 6.7 |
| ResNet 50 + ABN | 43.58M | 7.64G | 23.43 | 6.92 |
| ResNet 50 + GAM | 151.32M | 24.66G | 22.78 | 6.43 |
| ResNet 50 + GAM (gc) | 58.9M | 9.56G | 23.01 | 6.52 |
| MobileNet V2 | 3.51M | 0.31G | 30.52 | 11.20 |
| MobileNet V2 + SE | 3.53M | 0.32G | 29.77 | 10.65 |
| MobileNet V2 + BAM | 3.54M | 0.32G | 29.91 | 10.80 |
| MobileNet V2 + CBAM | 3.54M | 0.32G | 29.74 | 10.66 |
| MobileNet V2 + GAM | 4.93M | 0.47G | 29.31 | 10.43 |

# 4.2 Ablation studies

# 4.2 抽象研究

We conduct two ablation studies on ImageNet-1K with ResNet18. We first evaluate the contributions of spatial and channel attention separately. Then, we compare GAM against CBAM with and without max-pooling.

我们在 ImageNet-1K 上使用 ResNet18 进行了两个抽象研究。首先单独评估空间注意力和通道注意力的贡献。然后，我们比较了 GAM 与带有和不带有最大池化的 CBAM。

To better understand the contribution of spatial and channel attention separately, we conduct the ablation study by turning one on and the other off. For example, indicates the spatial attention is switched off and the channel attention is on. sp indicates the channel attention is turned off and the spatial attention is on. The results are shown in Table 3 . We could observe a boost of performance on both of the on-off experiments. It indicates that both spatial and channel attentions are contributing to the performance gain. Note that their combination improves the performance with a further step.

为了更好地理解空间注意力和通道注意力单独的贡献，我们通过开启一个并关闭另一个来进行抽象研究。例如， 表示空间注意力被关闭而通道注意力开启。sp 表示通道注意力被关闭而空间注意力开启。结果如表 3 所示。我们可以观察到在开启-关闭实验中性能的提升。这表明空间注意力和通道注意力都对性能提升有所贡献。注意，它们的组合进一步提高了性能。

Table 3: Ablation studies on ImageNet

表 3：ImageNet 上的抽象研究

| Architecture | Parameters | FLOPs | Top-1 Error (%) | Top-5 Error (%) |
| --- | --- | --- | --- | --- |
| ResNet 18+ | 11.69M | 1.82G | 30.91 | 11.12 |
| ResNet 18 + GAM (sp\*) | 15.95M | 2.45G | 29.61 | 10.41 |
| ResNet 18 + GAM (ch\*) | 11.78M | 1.83G | 30.25 | 10.97 |
| ResNet 18 + GAM (ch+sp) | 16.04M | 2.45G | 29.34 | 10.23 |

sp stands for spatial attention only. ch stands for channel attention only.

sp 代表仅空间注意力。ch 代表仅通道注意力。

+ same as Table 2

+ 与表 2 相同

It is possible for max-pooling to contribute negatively in spatial attention depends on the neural architecture (e.g., ResNets). Therefore, we conduct another ablation study that compares GAM

在某些神经网络架构（例如 ResNets）中，最大池化可能对空间注意力产生负面影响。因此，我们进行了另一个抽象研究，比较了 GAM 与 ResNet18 上带有和不带有最大池化的 CBAM。

against CBAM with and without max-pooling for ResNet18. The results are shown in Table 4. It is observed that our method outperforms CBAM under both conditions.

结果如表 4 所示。观察到我们的方法在两种条件下都超过了 CBAM。

Table 4: Ablation studies on ImageNet

表4：ImageNet上的消融研究

| Architecture | Parameters | FLOPs | Top-1 Error (%) | Top-5 Error (%) |
| --- | --- | --- | --- | --- |
| ResNet 18 + CBAM+ | 11.78M | 1.82G | 29.89 | 10.53 |
| ResNet 18 + GAM | 16.04M | 2.46G | 29.34 | 10.23 |
| ResNet 18 + CBAM (wmp\*) | 11.78M | 1.83G | 29.44 | 10.24 |
| ResNet 18 + GAM (wmp\*) | 16.05M | 2.47G | 28.57 | 9.83 |

wmp stands for without max pooling.

wmp代表没有最大池化。

† same as Table 2

† 与表2相同

# 5 Conclusion

# 5 结论

In this work, we proposed GAM to magnify salient cross-dimension receptive regions. Our experimental results indicate that GAM stably improves the performance for CNNs with different architectures and depths.

在这项工作中，我们提出了GAM来放大显著的跨维度感受区域。我们的实验结果表明，GAM稳定地提高了不同架构和深度的CNNs的性能。

CIFAR-100 and ImageNet-1K are benchmarked in our evaluation as proof of concept. They represent a scaling up with the number of classes and images. Therefore, our experiments imply that GAM is prone to data scaling capability and robustness. We consider the full ImageNet dataset serves better for applications in production. It is expensive for large-model training, especially the up-to-date top-tier solutions. Our evaluation with ResNet and MobileNet proves its feasibility on model scaling as well. We aim to investigate detailed scaling capability of GAM as the next step.

CIFAR-100和ImageNet-1K在我们的评估中作为概念验证。它们代表了随着类别和图像数量的增加而进行的扩展。因此，我们的实验表明GAM倾向于数据扩展能力和鲁棒性。我们认为完整的ImageNet数据集更适合生产环境中的应用。对于大型模型训练来说，尤其是最新的顶级解决方案，成本高昂。我们使用ResNet和MobileNet的评估也证明了它在模型扩展方面的可行性。我们计划在下一步详细研究GAM的扩展能力。

GAM obtains performance gain with an increase in the number of network parameters. In the future, we plan to investigate technologies that reduce the number of parameters for large networks, e.g., ResNet50, ResNet101, etc. Meanwhile, we also plan to explore other cross-dimension attention mechanisms that utilize parameter-reduction techniques.

GAM随着网络参数数量的增加而获得性能提升。未来，我们计划研究减少大型网络（例如ResNet50、ResNet101等）参数数量的技术。同时，我们还计划探索其他利用参数减少技术的跨维度注意力机制。

# References

# 参考文献

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580-587, 2014.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431-3440, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.

Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 951-958. IEEE, 2009.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248-255. Ieee, 2009.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25:1097-1105, 2012.

Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. arXiv preprint arXiv:2106.04560, 2021.

Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. Neurocomputing, 452:48-62, 2021.

Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3156-3164, 2017.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132-7141, 2018.

Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3-19, 2018.

Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Bam: Bottleneck attention module. arXiv preprint arXiv:1807.06514, 2018.

Diganta Misra, Trikay Nalamada, Ajay Uppili Arasanipalai, and Qibin Hou. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3139-3148, 2021.

Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 6848-6856, 2018.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mo-bilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4510-4520, 2018.

Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi. Attention branch network: Learning of attention mechanism for visual explanation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10705-10714, 2019.