

## 基于轻量级分组注意力模块的图像分类算法

张盼盼<sup>1,2</sup>, 李其中<sup>2\*</sup>, 杨词慧<sup>2</sup>

(1. 南昌航空大学 信息工程学院, 南昌 330063; 2. 江西省图像处理与模式识别重点实验室(南昌航空大学), 南昌 330063)

(\* 通信作者电子邮箱 li\_qishen@163.com)

**摘要:**针对图像分类任务中现有神经网络模型对分类对象特征表征能力不足, 导致识别精度不高的问题, 提出一种基于轻量级分组注意力模块(LGAM)的图像分类算法。该模块从输入特征图的通道和空间两个方向出发重构特征图: 首先, 将输入特征图沿通道方向进行分组并生成每个分组对应的通道注意力权重, 同时采用阶梯型结构解决分组间信息不流通的问题; 然后, 基于各分组串联成的新特征图生成全局空间注意力权重, 通过两种注意力权重加权得到重构特征图; 最后, 将重构特征图与输入特征图融合得到增强的特征图。以分类Top-1错误率作为评估指标, 基于Cifar10和Cifar100数据集以及部分ImageNet2012数据集, 对经LGAM增强之后的ResNet、Wide-ResNet、ResNeXt进行对比实验。实验结果表明, 经LGAM增强之后的神经网络模型其Top-1错误率均低于增强之前1至2个百分点。因此LGAM能够提升现有神经网络模型的特征表征能力, 从而提高图像分类的识别精度。

**关键词:**注意力机制; 图像分类; 通道注意力; 空间注意力; 分组卷积

**中图分类号:** TP391.4 **文献标志码:** A

### Image classification algorithm based on lightweight group-wise attention module

ZHANG Panpan<sup>1,2</sup>, LI Qishen<sup>2\*</sup>, YANG Cihui<sup>2</sup>

(1. School of Information Engineering, Nanchang Hangkong University, Nanchang Jiangxi 330063, China;

2. Key Laboratory of Jiangxi Province for Image Processing and Pattern Recognition (Nanchang Hangkong University), Nanchang Jiangxi 330063, China)

**Abstract:** Aiming at the problem that the existing neural network models have insufficient ability to characterize the features of classification objects in image classification tasks and cannot achieve high recognition accuracy, an image classification algorithm based on Lightweight Group-wise Attention Module (LGAM) was proposed. The proposed module reconstructed the feature maps from the channel and space of the input feature maps. Firstly, the input feature maps were grouped along the channel direction, and channel attention weight corresponding to each group was generated. At the same time, ladder type structure was used to solve the problem that the information between the groups was not circulated. Secondly, the global spatial attention weight was generated based on the new feature maps concatenated by each group, and the reconstructed feature maps were obtained by weighting the two attention weights. Finally, the reconstructed feature maps were merged with the input feature maps to generate the enhanced feature maps. Experiments were performed on the Cifar10 and Cifar100 datasets and part of the ImageNet2012 dataset with using the classification Top-1 error rate as the evaluation indicator to compare the ResNet, Wide-ResNet and ResNeXt enhanced by LGAM. Experimental results show that the Top-1 error rates of the neural network models enhanced by LGAM are 1 to 2 percentage points lower than those of the models before enhancing. LGAM can improve the feature characterization ability of existing neural network models, thus improving the recognition accuracy of image classification.

**Key words:** attention mechanism; image classification; channel attention; spatial attention; group convolution

## 0 引言

图像分类是计算机视觉领域的一个重要研究方向, 其主要任务是通过计算机对图像中的目标进行自动识别, 并将其归属到对应的类别集合中<sup>[1-3]</sup>。传统的图像分类算法主要包含图像预处理、图像特征提取和分类器分类三个步骤, 其中, 特征提取是图像分类中最核心的步骤, 该步骤所提取的特征优劣直接影响到图像分类的性能。传统的图像分类算法提取

的特征具有较大的冗余性, 并且不具备较好的泛化性, 为此研究者们提出了很多优秀的神经网络模型<sup>[4-8]</sup>。

目前, 卷积神经网络因其结构在图像处理方面有着独特的优势, 因此被广泛应用于图像分类任务。然而, 由于图像经过多层卷积和池化操作, 丢失了大量的重要信息, 导致卷积神经网络及其改进模型所提取的特征对目标的表征能力不足。针对此类问题, 研究者们提出一系列特征增强模块, 以提高网络中间特征的健壮性。然而大多数特征增强模块不仅增加了

收稿日期: 2019-08-16; 修回日期: 2019-10-15; 录用日期: 2019-10-22。

基金项目: 国家自然科学基金资助项目(61662048); 南昌航空大学研究生专项创新资金资助项目(YC2018030)。

作者简介: 张盼盼(1993—), 女, 江苏宿迁人, 硕士研究生, 主要研究方向: 图像处理、模式识别、深度学习; 李其中(1975—), 男, 河北衡水人, 副教授, 博士, 主要研究方向: 图像处理、模式识别; 杨词慧(1979—), 男, 江西九江人, 副教授, 博士, 主要研究方向: 图像处理、模式识别。

原始神经网络模型的计算量,同时对精度的提升效果也不尽人意。

本文提出了一种基于轻量级分组注意力模块(Lightweight Group-wise Attention Module, LGAM)对现有的神经网络模型进行增强的算法。该模块结合输入特征图的通道和空间信息对输入特征图进行重构,以提高现有神经网络模型的特征表征能力。通过在Cifar10和Cifar100数据集以及部分ImageNet2012数据集上将LGAM增强后的网络模型和增强前的进行对比实验,证实了本文模块的通用性和有效性,同时还本文将本文模块与其他特征增强模块进行对比,证实了本文模块是一种更加轻量级的模块。

## 1 相关工作

### 1.1 注意力机制

近年来,随着深度学习的广泛使用,图像分类研究获得了快速发展,但图像分类任务中现有的神经网络模型仍然存在对分类对象特征表征能力不足等问题。注意力机制通过凸显特征图中具有较多有效信息的特征表示,同时抑制无用信息的干扰,有效提高神经网络模型的性能,该机制已成功应用于很多计算机视觉任务<sup>[9-12]</sup>。文献[13]提出采用残差注意力网络,通过引入沙漏形式的注意力模块实现跨空间和通道维度的全局重点;但整个操作在3D特征图上进行,没有对特征图通道和空间维度进行解耦,因此计算量较大。文献[14]提出的SE(Squeeze-and-Excitation)模块是一种有效的轻量级门控机制,通过学习通道重要性的方式自适应校准特征图。然而SE模块只关注了特征图通道方面的影响而忽略了空间维度的重要性。除了采用通道注意力机制之外,文献[15]提出了SGE(Spatial Group-wise Enhance)模块,该模块通过为每个语义组中的每个空间位置生成一个注意因子来调整每个子特征的重要性,其实质是一个空间注意力模块。文献[16-17]提出的CBAM(Convolutional Block Attention Module)和BAM(Bottleneck Attention Module)以类似的方式同时引入了通道注意力和空间注意力两种机制。文献[18]提出采用GC(Global Context)模块改善非局部网络计算量较大的缺点,同时与SE模块相结合设计了一个全局上下文模块。文献[19]提出一种动态内核选择机制,主要思想是通过注意力机制实现神经元动态调整感受野。

针对特征图耦合度较高、计算量较大等问题,本文将特征图通道和空间维度进行解耦,分别生成对应的注意力权重,即结合通道注意力机制和空间注意力机制,提高神经网络模型的图像分类性能。

### 1.2 分组卷积

目前卷积神经网络仍存在大量的冗余,因此研究者们设计了一种轻量级的网络结构,即使用分组卷积以避免冗余对计算和存储资源造成巨大的浪费。分组卷积将特征图沿通道方向分成并行的多组结构,以此减少计算量。

传统卷积结构如图1所示,该结构计算量公式为 $D_k \times D_k \times M \times N \times D_i \times D_i$ 。 $D_k \times D_k$ 表示卷积核的尺寸, $M$ 表示输入特征图的通道数, $N$ 表示输出特征图的通道数, $D_i \times D_i$ 表示输出特征图的尺寸。分组卷积结构如图2所示,该结构计算量公式为 $D_k \times D_k \times M \times N \times D_i \times D_i / G$ (组数用 $G$ 表示)。从该公式中可看出分组卷积的计算量减少为传统卷积计算量的 $1/G$ 。

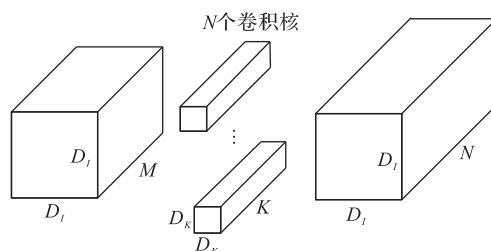


图1 传统卷积

Fig. 1 Traditional convolution

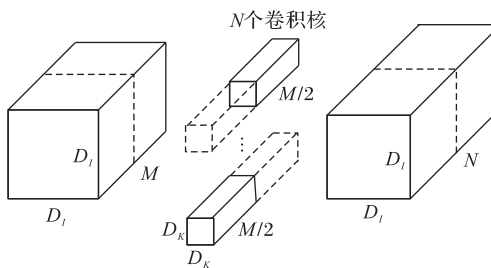


图2 分组卷积( $G=2$ )

Fig. 2 Group convolution ( $G=2$ )

分组卷积首次应用于文献[4],由于当时计算条件有限,研究者将卷积操作分成两组,采用双GPU并行运算的方式以提高计算效率。文献[20]提出基数的概念,即分组卷积的组数,通过大量实验验证,研究者们发现分组卷积不仅可以减少计算量,还可以一定程度上提高网络的识别精度。文献[21]中引入了深度可分离卷积,该卷积将普通卷积分解为深度卷积和逐点卷积两部分来减少计算量,其中深度可分离卷积属于分组卷积的特例。文献[22-23]提出了交错分组卷积以解决分组卷积带来的信息不流通的问题,并开发了许多紧凑模型。

由此,本文将分组卷积与注意力机制相结合提出一种轻量级分组注意力模块(LGAM),借鉴分组卷积的思想降低模块的计算量。与其他注意力模块相比,本文模块具有更少的参数以及更易解释的性质。

## 2 本文算法

现有的神经网络模型中大部分的计算是卷积操作,而卷积得到的特征图直接影响着网络模型的特征表征能力。特征图由通道和空间信息两个部分组成,本文探索了一种由通道注意力机制和空间注意力机制融合的注意力模块,从两个角度对输入特征图进行重构,提高现有神经网络模型的特征表征能力。

本文所提出的模块LGAM如图3所示。LGAM主要由两部分组成:通道注意力和空间注意力。首先,将特征图沿通道方向分解成多组,并生成相应的通道注意力权重;然后,将多个分组进行串联产生新的特征图,并在其基础上生成相应的空间注意力权重,两者进行加权得到重构的特征图;最后,将重构特征图和输入特征图两者融合。

图3中, $A$ 表示输入特征图, $x_k$ 表示各分组, $y_k$ 表示各分组产生的结果,CA(Channel Attention)表示通道注意力操作,SA(Spatial Attention)表示空间注意力操作。 $\oplus$ 表示逐元素相加, $\otimes$ 表示逐元素相乘。

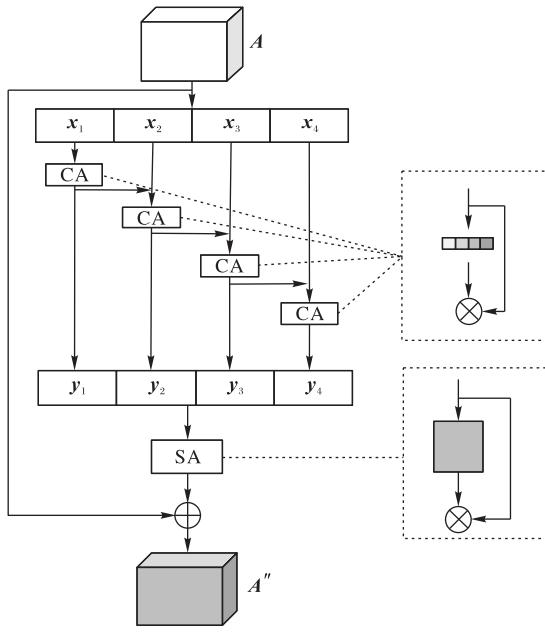


图3 轻量级分组注意力模块

Fig. 3 Lightweight Group-wise Attention Module (LGAM)

### 2.1 通道注意力(Channel Attention, CA)

假设给定一个输入特征图  $A \in \mathbb{R}^{H \times W \times C}$ , 将特征图沿通道维度分成  $G$  组, 用  $x_k \in \mathbb{R}^{H \times W \times C'}$  表示各个分组,  $k \in \{1, 2, \dots, G\}$ , 每个  $x_k$  包含  $C' = C/G$  个通道。然后将分组后的特征图转换成向量形式  $x_k = U = [u_1, u_2, \dots, u_{C'}]$ ,  $u_i \in \mathbb{R}^{H \times W}$ 。由于每个通道都包含特定的功能响应, 因此本文利用通道间的相互关系, 通过全局平均池化(Global Average Pooling, GAP)聚合每个通道中的特征映射生成对应的权重信息, 权重的每个元素都由特征图的空间维度  $H \times W$  进行压缩得到, 即对每个通道中的全局信息进行软编码。全局平均池化计算如式(1)所示:

$$s_{\text{chan}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_{C'}(i, j) \quad (1)$$

其中:  $s_{\text{chan}} \in \mathbb{R}^{1 \times 1 \times C'}$  表示聚合通道得到的权重;  $H \times W$  表示特征图空间维度的信息。

为了自适应生成通道间的注意力权重, 提高计算效率, 使用具有一个隐藏层的多层感知机(MultiLayer Perceptron, MLP), 隐含层单元数为  $C'/r$ ,  $r$  为缩放比例。在多层感知机后加入批归一化(Batch Normalization, BN)层调整输出比例。使用Sigmoid函数将得到的权重归一化。公式如下:

$$M_{\text{chan}} = \sigma(\delta(\text{MLP}(s_{\text{chan}}))) \quad (2)$$

其中:  $\delta$  表示ReLU函数,  $\sigma$  表示Sigmoid函数,  $M_{\text{chan}}$  表示每个分组对应生成的通道注意力权重。

最后将每组的输出采用残差结构以避免层数较深导致的深层特征值消失和性能下降等问题。因此每组通道注意力加权公式为:

$$x'_k = M_{\text{chan}} \otimes x_k \oplus x_k \quad (3)$$

其中:  $x'_k \in \mathbb{R}^{H \times W \times C'}$  表示每个分组经过注意力机制处理的输出结果。 $\oplus$  表示逐元素相加,  $\otimes$  表示逐元素相乘。通道注意力具体操作如图4所示。

LGAM的第一个步骤是将输入特征图分组, 该方式类似分组卷积的思想, 分组卷积计算量较少, 但不同分组的通道间

缺少信息流通。因此本文采用图3所示的阶梯型结构, 改善分组间信息不流通的问题。阶梯型结构计算如式(4)所示:

$$y_k = \begin{cases} F_{\text{chan}}(x_k), & k = 1 \\ F_{\text{chan}}(x_k + y_{k-1}), & 1 < k \leq G \end{cases} \quad (4)$$

其中:  $F_{\text{chan}}(\cdot)$  表示对  $x_k$  进行一次求通道注意力权重的操作。

最后将各个分组得到的输出  $y_k$  串联成新的特征图  $A' \in \mathbb{R}^{H \times W \times C}$ 。

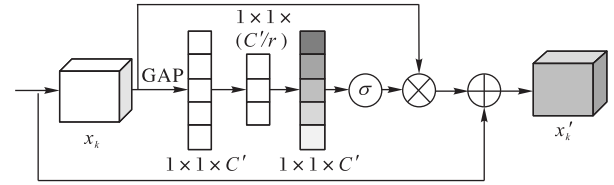


图4 通道注意力

Fig. 4 Channel attention

### 2.2 空间注意力(Spatial Attention, SA)

由于噪声的影响, 卷积神经网络模型通常难以获得良好分布的特征响应, 但特征图空间和通道信息之间有着诸多联系, 因此本文利用通道信息对空间信息进行增强处理, 目的是强调和抑制不同空间位置的特征。

本文提出的空间注意力模块通过对空间信息间的相互关系进行建模, 生成相应的空间注意力权重, 整个流程可以看作特征选取的操作。空间信息与通道信息互补, 为了得到有效的空间注意力图, 将输入特征图的通道信息进行压缩, 生成有效的特征描述符, 利用特征通道的信息来突出空间信息的方法已被证明是有效的<sup>[24]</sup>。

通过2.1节通道注意力计算后得到的新特征图为  $A' \in \mathbb{R}^{H \times W \times C}$ , 将该特征图转换成向量  $V = [v_1, v_2, \dots, v_s]$ ,  $v_i \in \mathbb{R}^C$ ,  $s = H \times W$ , 对通道信息进行压缩的计算公式为:

$$c_{\text{spat}} = \sum_{i=1}^C v_s(i) \quad (5)$$

其中:  $c_{\text{spat}}$  表示第  $s$  个位置生成的空间信息权重,  $v_s$  表示第  $s$  个位置的视觉特征。

为了防止样本数据的偏差太大, 影响网络的泛化性, 因此对空间信息的权重进行归一化处理。如式(6)所示:

$$M_{\text{spat}} = \sigma(\delta(c_{\text{spat}})) \quad (6)$$

其中:  $\delta$  表示ReLU函数,  $\sigma$  表示Sigmoid函数。

最后将重构后的特征图与输入特征图两者进行融合, 公式如下:

$$A'' = A' \otimes M_{\text{spat}} \oplus A \quad (7)$$

其中:  $A'' \in \mathbb{R}^{H \times W \times C}$  表示LGAM最后的输出,  $A'$  表示分组通道注意力的输出结果,  $M_{\text{spat}}$  表示生成的空间注意力权重,  $\oplus$  表示逐元素相加,  $\otimes$  表示逐元素相乘。

## 3 实验与分析

本文在两种公共基准的数据集上对LGAM进行实验测评, 分别为Cifar数据集和部分ImageNet 2012数据集。

对比实验均在Linux16.04操作系统, CPU(32核, 2.1 GHz), GPU TITAN Xp(显存12 GB)下进行。深度学习框架采用Pytorch。

### 3.1 Cifar数据集

本节采用公共基准的Cifar10数据集和Cifar100数据集来评估算法的有效性, 这两个数据集各包含60 000张彩色图片,

大小均为  $32 \times 32$ , 其中 50 000 张用于训练, 10 000 张用于测试。数据增强采用常规方法: 即裁剪和随机水平翻转。

为了保证实验的可比性, 设定实验选择相同的超参数初始化方法: 优化方法采用随机梯度下降 (Stochastic Gradient Descent, SGD); 动量设定为 0.9; 训练批次设定为 128; 测试批次设定为 100; 权重衰减设定为 0.0005; 学习率初始化为 0.1, 并采用自适应降低学习率算法; 迭代次数设定为 160 次。最终的结果取 5 次测试的平均值。

对于 ResNet 模型, 为使其更适合 Cifar 数据集, 本文在结构方面作一些细微调整: 原有的 ResNet 网络模型设计用于 ImageNet 数据集, 第一层采用  $7 \times 7$  的卷积核, 但 Cifar 数据集中的图片大小均为  $32 \times 32$ , 因此将第一层卷积核大小调整为  $3 \times 3$ , 同时去掉最大池化层。3.1 节实验中涉及到的 ResNet 框架均以调整后的框架为基准。

3.1.1 模块泛化性验证分析

该部分将提出的 LGAM 置于多个先进的神经网络模型中用于验证本文模块的泛化性。网络模型选取 ResNet18, ResNet34, Wide-ResNet (WRN16-10) 和 ResNeXt29 (32\*4d)。数据集采用 Cifar100 数据集。实验结果如表 1 所示, 表中显示信息分别为网络模型、模型参数量以及模型分类的 Top-1 错误率。

表 1 模块泛化性验证

Tab. 1 Module generalization verification

| 网络模型            | 参数量/ $10^6$ | Top-1 错误率/% |
|-----------------|-------------|-------------|
| ResNet 18       | 11.22       | 23.52       |
| ResNet 18+LGAM  | 11.25       | 22.94       |
| ResNet 34       | 21.33       | 22.45       |
| ResNet 34+LGAM  | 21.39       | 21.36       |
| WRN 16-10       | 17.43       | 22.54       |
| WRN 16-10+LGAM  | 17.48       | 21.55       |
| ResNeXt 29      | 24.26       | 22.59       |
| ResNeXt 29+LGAM | 24.34       | 21.47       |

从表 1 中的最后一列数据可以看出, 本文提出的 LGAM 模块与任意神经网络模型集成得到的 Top-1 错误率相较于原神经网络模型都降低了 1% 左右。LGAM 从特征图通道和空间两个方向应用注意力机制, 只增加了较少的参数有效地提升网络模型的性能, 上述分析充分证明本文模块的泛化性。

3.1.2 本文算法与其他注意力模块对比

对 LGAM 模块与其他的注意力模块做对比实验。将各种注意力模块分别嵌入到 ResNet18 网络框架的每个残差块中用于实验, 对比的注意力模块分别是 SE<sup>[14]</sup>、SGE<sup>[15]</sup>、CBAM<sup>[16]</sup>、BAM<sup>[17]</sup>。数据集采用 Cifar10 和 Cifar100 数据集。具体网络框架见表 2, 表 2 中 AM (Attention Module) 表示注意力模块。实验结果见表 3, 表 3 中显示信息分别为网络模型、模型的参数量以及模型分类的 Top-1 错误率。

LGAM 与其他注意力模块的对比结果如表 3 所示: 参数方面, 在 Cifar100 数据集上, 本文提出的 LGAM 模块与 ResNet18 模型结合后的模型参数量为  $11.25 \times 10^6$ , 而原始 ResNet18 模型参数量为  $11.22 \times 10^6$ , 仅增加了  $0.03 \times 10^6$ , 而 SE、CBAM、BAM 三种模块分别增加了  $0.09 \times 10^6$ 、 $0.09 \times 10^6$ 、 $0.07 \times 10^6$ , 因此得出结论: LGAM 占用的网络开销较小, 是一种轻量级的模块。LGAM 将输入特征图通道和空间进行解耦, 同时采用分组卷积的思想降低模块的开销, 提升网络的性能。虽然在模块的

计算量方面略高于 SGE 模块, 但 LGAM 模块的分类准确率高于 SGE 模块。分类性能方面, SE 模块是通道注意力机制, SGE 模块是空间注意力机制, 而本文采用两种注意力机制相结合的方式, 从结果上看出, LGAM 的两种注意力机制结合方式的分类效果优于 SE 和 SGE 的单注意力机制算法。上述分析证实, 本文将 LGAM 模块与 ResNet18 框架相结合, 得到了最低的分类错误率, 充分证明本文算法的有效性。

表 2 注意力模块+ResNet18 网络框架

Tab. 2 Network framework of attention module + ResNet18

| Layer           | Output Size    | Block Type                                                                  |
|-----------------|----------------|-----------------------------------------------------------------------------|
| Conv1           | $32 \times 32$ | $[3 \times 3, 64]$                                                          |
| Conv2_x         | $32 \times 32$ | $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$   |
| AM              | $32 \times 32$ | Attention $\times 1$                                                        |
| Conv3_x         | $16 \times 16$ | $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$ |
| AM              | $16 \times 16$ | Attention $\times 1$                                                        |
| Conv4_x         | $8 \times 8$   | $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$ |
| AM              | $8 \times 8$   | Attention $\times 1$                                                        |
| Conv5_x         | $4 \times 4$   | $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$ |
| AM              | $4 \times 4$   | Attention $\times 1$                                                        |
| Average Pooling | $1 \times 1$   | $4 \times 4$                                                                |
| FC, Softmax     | 10(100)        |                                                                             |

表 3 不同注意力模块结果对比 (Cifar)

Tab. 3 Comparison of different attention module results on Cifar

| 网络模型                           | Cifar10     |             | Cifar100    |             |
|--------------------------------|-------------|-------------|-------------|-------------|
|                                | 参数量/ $10^6$ | Top-1 错误率/% | 参数量/ $10^6$ | Top-1 错误率/% |
|                                |             |             |             |             |
| ResNet 18                      | 11.17       | 5.68        | 11.22       | 23.52       |
| ResNet 18-SE <sup>[14]</sup>   | 11.26       | 5.54        | 11.31       | 23.32       |
| ResNet 18-SGE <sup>[15]</sup>  | 11.17       | 5.33        | 11.22       | 23.35       |
| ResNet 18-CBAM <sup>[16]</sup> | 11.26       | 5.41        | 11.31       | 23.23       |
| ResNet 18-BAM <sup>[17]</sup>  | 11.24       | 5.25        | 11.29       | 23.21       |
| ResNet 18-LGAM                 | 11.20       | 5.16        | 11.25       | 22.94       |

3.2 ImageNet2012 数据集

本节采用 ImageNet2012 数据集, 该数据集共包含 128 万张训练图像, 5 万张验证图像, 共 1 000 个类别。由于计算时间原因这里随机抽取 20 类用于测试本文的模型。数据增强采用常规方法, 即随机裁剪和随机水平翻转, 同时采用平均通道减法用于标准化图像用于训练和测试。

为保证实验的可比性, 本部分实验选择相同的超参数初始化方法: 优化方法采用随机梯度下降 (Stochastic Gradient Descent, SGD); 动量设定为 0.9; 训练批次设定为 64; 测试批次设定为 100; 权重衰减设定为 0.0001; 初始化学率率为 0.1 并且每 30 次迭代次数降低 10 倍; 迭代次数设定为 90 次。最终的结果取 5 次测试的平均值。

3.2.1 本文算法与其他注意力模块对比

该部分将提出的 LGAM 模块与其他的注意力模块做对比实验。对比的注意力模块分别是 SE、SGE、CBAM、BAM。实验结果见表 4, 表中显示信息分别为网络模型、模型的参数量以及模型分类的 Top-1 错误率。

从表 4 中的数据可以看出, 本文提出的 LGAM 模块与

RseNet18 模型结合后的模型参数量为  $11.22 \times 10^6$ , 而原始 ResNet18 模型参数量为  $11.19 \times 10^6$ , 仅增加了  $0.03 \times 10^6$ , 并且都低于 SE、CBAM、BAM。因此可以得出 LGAM 占用的网络开销较小, 是一种轻量级的模块。同时在分类性能方面, LGAM 与 ResNet18 相结合得到了最优的分类效果, 与在 Cifar 数据集上实验得到的结论一致, 再一次证实本文算法的有效性。

表 4 不同注意力模块结果对比 (ImageNet2012)

Tab. 4 Comparison of different attention module results on ImageNet2012

| 网络模型                           | 参数量/ $10^6$ | Top-1 错误率/% |
|--------------------------------|-------------|-------------|
| ResNet 18                      | 11.19       | 25.57       |
| ResNet 18-SE <sup>[14]</sup>   | 11.27       | 25.26       |
| ResNet 18-SGE <sup>[15]</sup>  | 11.19       | 24.38       |
| ResNet 18-CBAM <sup>[16]</sup> | 11.27       | 23.74       |
| ResNet 18-BAM <sup>[17]</sup>  | 11.25       | 23.57       |
| ResNet 18-LGAM                 | 11.22       | 23.14       |

3.2.2 不同网络模型结果可视化

本文为方便定性分析注意力模块给网络模型带来的作用, 使用 Grad-CAM<sup>[25]</sup> 来可视化 ImageNet 2012 验证集的图片。Grad-CAM 是一种新型可视化方法, 主要利用目标类的梯度, 传播到最终卷积层来生成粗略定位图, 可视化的结果可以清楚地观察到网络如何选择预测类的重要区域, 借此来判断注意力模块带来的影响。

本文比较了 ResNet18 与 SE、SGE、CBAM、BAM、LGAM 分别相结合的可视化效果, 并从 ImageNet 2012 数据集中随机挑选了 6 个种类的图片分别进行可视化。如图 5 所示, 从第二列开始直到最后一列, 每张图片上的数据表示相应的输入图像在各个网络框架的预测分数。最后一列中 LGAM 模块集成的网络可视化的注意力掩膜比其他注意力模块更好地凸显目标对象区域。由此可以判断 LGAM 模块可以帮助神经网络模型更好地学习目标对象的特征信息, 提高目标正确类别的分数。

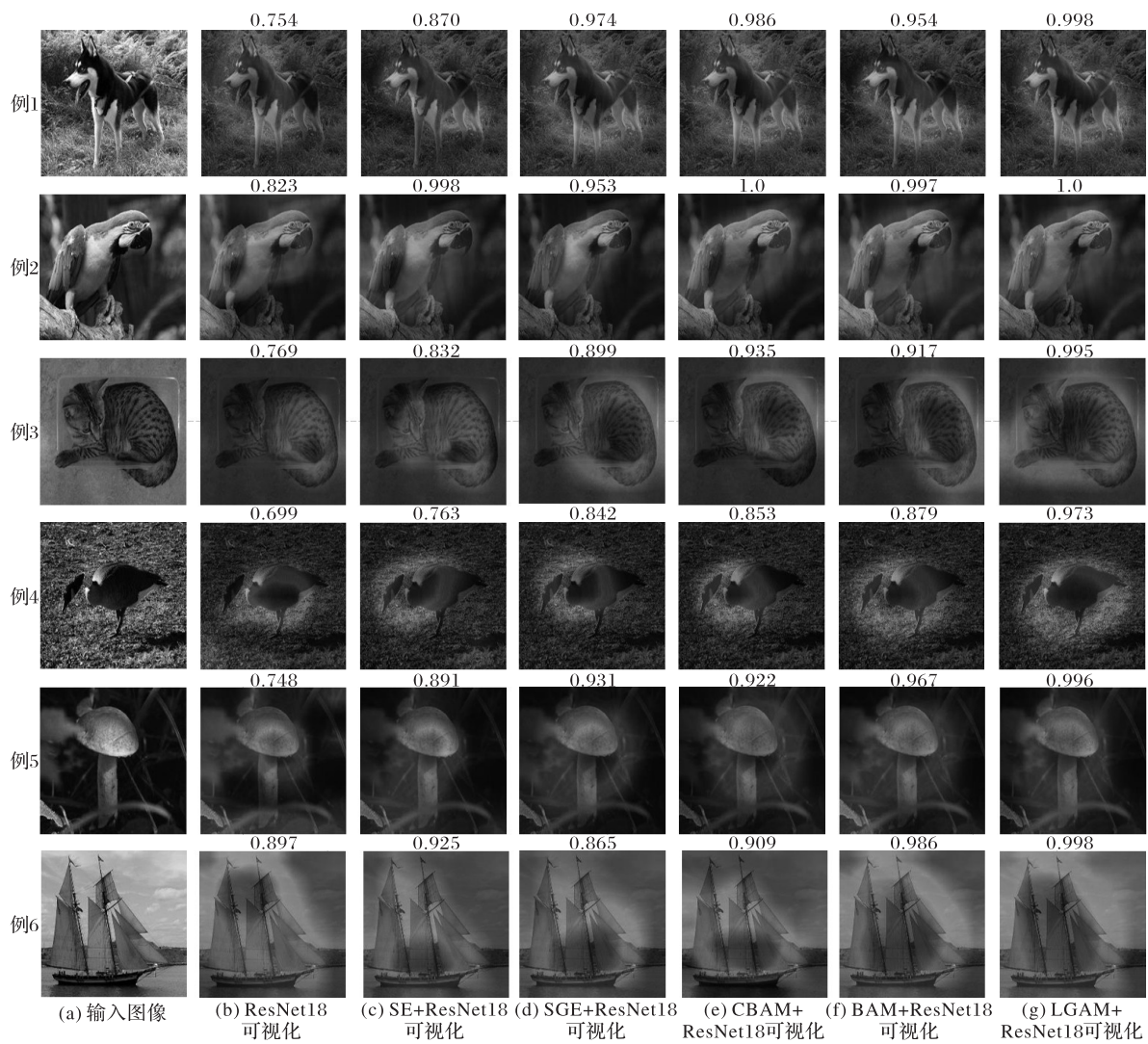


图 5 Grad-CAM 可视化结果

Fig. 5 Grad-CAM visualization result

4 结语

本文针对现有神经网络模型提取的特征表征能力不足导致识别精度不高的问题, 提出了一种基于注意力机制的图像

分类算法。通过设计一种轻量级的注意力模块 LGAM 与任意现有的神经网络模型相结合, 该模块将输入特征图的通道信息和空间信息同时用于加权特征图可以有效提高模型的特征表征能力, 并且利用分组的思想设计模块可以有效地降低模

块的开销。最后在Cifar10和Cifar100数据集以及部分ImageNet 2012数据集上通过实验证实了LGAM可以提高现有神经网络模型的识别精度。在后续的研究中,可以通过深度可分离卷积代替普通卷积,进一步减少网络模型和注意力模块的开销。另外,还可以采用将底层特征和高层特征相结合的方式进一步提高网络的识别精度。

#### 参考文献(References) (References)

- [1] 孙瑜阳. 深度学习及其在图像分类识别中的研究综述[J]. 信息技术与信息化, 2018 (1): 138-140. (SUN Y Y. A review of deep learning and its research in image classification and recognition[J]. Information Technology and Informatization, 2018 (1): 138-140. )
- [2] 杨真真, 匡楠, 范露, 等. 基于卷积神经网络的图像分类算法综述[J]. 信号处理, 2018, 34 (12): 1474-1489. (YANG Z Z, KUANG Y, FAN L, et al. Review of image classification algorithms based on convolutional neural networks[J]. Journal of Signal Processing, 2018, 34(12): 1474-1489. )
- [3] 高磊, 范冰冰, 黄穗. 基于残差的改进卷积神经网络图像分类算法[J]. 计算机系统应用, 2019, 19(7): 139-144. (GAO L, FAN B B, HUANG S. Improved CNN image classification algorithm based on residuals[J]. Journal of Computer Systems and Applications, 2019, 19(7): 139-144. )
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2012: 1097-1105.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2018-12-11]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [7] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.
- [8] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [EB/OL]. [2019-03-10]. <https://arxiv.org/pdf/1605.07146.pdf>.
- [9] 王文冠, 沈建冰, 贾云得. 视觉注意力检测综述[J]. 软件学报, 2019, 30(2): 416-439. (WANG W G, SHEN J B, JIA Y D. Review of visual attention detection[J]. Journal of Software, 2019, 30 (2): 416-439. )
- [10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2017: 6000-6010.
- [11] ZHANG Y, LI K, LI K, et al. Image superresolution using very deep residual channel attention networks [EB/OL]. [2018-11-10]. <https://arxiv.org/pdf/1807.02758.pdf>.
- [12] ANDERSON P, HE X, BUEHLER C, et al. Bottom-up and top-down attention for image captioning and visual question answering [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6077-6086.
- [13] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6450-6458.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [15] LI X, HU X, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks [EB/OL]. [2019-08-15]. <https://arxiv.org/pdf/1905.09646.pdf>.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11211. Cham: Springer, 2018: 3-19.
- [17] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module [EB/OL]. [2018-12-18]. <https://arxiv.org/pdf/1807.06514.pdf>.
- [18] CAO Y, XU J, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond [EB/OL]. [2019-06-12]. <https://arxiv.org/pdf/1904.11492.pdf>.
- [19] LI X, WANG W, HU X, et al. Selective kernel networks [C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 6120-6132.
- [20] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural network [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5987-5995.
- [21] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2019-06-12]. <https://arxiv.org/pdf/1704.04861.pdf>.
- [22] HLUCHYJ M G, KAROL M J. ShuffleNet: an application of generalized perfect shuffles to multihop lightwave networks [J]. Journal of Lightwave Technology, 1991, 9(10): 1386-1397.
- [23] ZHANG T, QI G J, XIAO B, et al. Interleaved group convolutions [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4383-4392.
- [24] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2018-05-20]. <https://arxiv.org/pdf/1612.03928.pdf>.
- [25] SELVARAJU R R, COGSWELL, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [J]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 618-626.

This work is partially supported by the National Natural Science Foundation of China (61662048), the Graduate Special Innovational Foundation of Nanchang Hangkong University (YC2018030).

**ZHANG Panpan**, born in 1993, M. S. candidate. Her research interests include image processing, pattern recognition, deep learning.

**LI Qishen**, born in 1975, Ph. D., associate professor. His research interests include image processing, pattern recognition.

**YANG Cihui**, born in 1979, Ph. D., associate professor. His research interests include image processing, pattern recognition.