

用于图像分类的深度卷积神经网络中的 空间分割注意力模块

王方, 乔瑞萍

(西安交通大学信息与通信工程学院, 710049, 西安)

摘要: 针对现有注意力机制常采用融合或压缩的方式来获取所需信息, 会导致空间或通道维度损失的信息过多的问题, 提出了一种有效的轻量级注意力模块 SPAM, 它可以在不经过通道融合或压缩的情况下获取注意力。对于输入的中间特征图, SPAM 首先会自适应地采用平均池化和最大池化来进行特征提取; 采用空间上局部的块特征代替点特征以减少计算量, 并利用实例标准化 (IN) 层与深层卷积来捕获全局空间注意力; 通道维度信息的重建通过分组卷积来完成; 最终使用插值操作获得整体注意力, 对输入特征图进行加权。SPAM 可以方便地嵌入到各种主流卷积神经网络架构中, 只需增加微量参数和计算量, 就可以显著提高网络性能。为了证明 SPAM 的有效性, 在 ImageNet-1K、CIFAR-100 和 Food-101 图像分类数据集上进行了大量实验, 并使用 Grad-CAM 可视化了网络的关注区域。实验结果表明, 在 ImageNet-1K、CIFAR-100 和 Food-101 数据集上, SPAM 分别将基线网络的准确率最多提高了约 1.08%、2.46% 和 1.09%。研究结果表明, 嵌入 SPAM 的网络的性能都有较大提升; 且相较于其他常用的轻量级注意力机制, SPAM 的效果始终更好; SPAM 使网络更关注目标对象所在区域, 确切提高了网络的表达能力。

关键词: 卷积神经网络; 注意力机制; 图像分类; 特征提取

中图分类号: TP391.4 **文献标志码:** A

DOI: 10.7652/xjtub202309019 **文章编号:** 0253-987X(2023)09-0185-08

SPAM: Spatially Partitioned Attention Module in Deep Convolutional Neural Networks for Image Classification

WANG Fang, QIAO Ruiping

(School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Existing attention mechanisms often use fusion or compression to obtain the required information, but this leads to a large quantity of information lost in the spatial or channel dimension. In order to solve this problem, the Spatially Partitioned Attention Module (SPAM), a really effective and lightweight attention module that can help obtain attention without channel fusion or compression, was proposed in the paper. For the input intermediate feature map, the SPAM first adaptively used average pooling and maximum pooling features for feature extraction, replaced the point feature with the local block feature in space to reduce the amount of calculation and used the IN layer and depthwise convolution to capture global spatial attention. Meanwhile, the reconstruction of channel dimension information was directly completed by group convolu-

tion. Finally, the interpolation operation was used to obtain overall attention and weight the input feature map. Notably, the SPAM can be easily embedded in various mainstream CNN architectures, and network performance can be significantly improved by increasing a few microparameters and calculations. To demonstrate the effectiveness of the SPAM, numerous experiments were conducted on the ImageNet-1K, CIFAR-100, and Food-101 datasets, and the network's regions of interest were visualized using Grad-CAM. On the ImageNet-1K, CIFAR-100, and Food-101 datasets, the SPAM improved the accuracy of the baseline network by up to about 1.08%, 2.46%, and 1.09%, respectively. The results show that the performance of the network embedded with the SPAM components is greatly improved; compared to other commonly used lightweight attention mechanisms, the SPAM always works better; the SPAM can really induce the networks to pay more attention to the target object regions and accurately improve the expression ability of the networks.

Keywords: convolutional neural networks; attention mechanisms; image classification; feature extraction

卷积神经网络是一种著名的深度学习架构,其灵感来自于生物的自然视觉感知机制^[1-2]。2012年,Alex等使用8层的深度卷积神经网络以绝对优势获得了ImageNet竞赛冠军^[3],卷积神经网络在图像识别中达到了前所未有的准确度。卷积神经网络从此广泛应用于图像分类、目标检测、语义分割等计算机视觉领域^[4]。

灵长类动物的视觉系统接受了大量的感官输入,这些感官输入远远超过了大脑能够完全处理的程度。然而,并非所有刺激的影响都是相等的。意识的聚集和专注使灵长类动物能够在复杂的视觉环境中将注意力引向感兴趣的物体^[5]。人类可以自然有效地在复杂场景中找到突出区域,受这一观察的启发,注意力机制被引入计算机视觉^[6]。

注意力机制可以使卷积神经网络关注于感兴趣的信息,同时抑制不必要的信息,基于输入特征完成权重的动态调整。注意力机制在许多视觉任务中取得了巨大成功^[7-11]。研究表明:在网络中添加注意力机制可以有效地提高网络的表达能力^[12]。

注意力机制可以分为通道注意力机制、空间注意力机制、通道空间混合注意力机制等。通道注意力机制旨在对卷积网络中不同通道赋予不同的重要性,其中经典算法就是挤压和激励注意力(SE)^[13]模块,它通过建模特征通道之间的相互依赖关系,自适应地学习每个特征通道的重要程度^[14]。空间注意力机制旨在对卷积网络的空间域不同位置赋予不同的重要性,例如借鉴非局部均值方法的非局部神经网络(Non-local Neural Networks)^[15]。空间注意力机制和通道注意力机制可以同时加入到网络中,

如卷积注意力机制模块(CBAM)^[16],也取得了优异的效果。

纵观这些注意力机制方法,人们总是考虑压缩某一维度以获得所需要的信息,例如通道注意力模块一般是通过在特征图上进行空间维度压缩以得到不同通道的重要程度,而空间注意力模块在通道维度进行压缩以得到不同位置的重要程度。卷积神经网络的不同通道通常代表着不同模式^[17],特征图对不同输入的响应也应该不同。考虑了对每一张特征图分别获取其重要性,本文提出了一种新的轻量级空间分割注意力模块(SPAM),可以对每一张特征图分别进行建模,增强网络本身的特征表达能力。

1 空间分割注意力模块

1.1 相关工作

深度学习通过使用反向传播算法来发现大型数据集中的复杂结构,以指示机器应该如何改变其内部参数。一定程度上,卷积神经网络可以自适应地建模图像空间和通道维度的关系,当然这并不意味着它总是可以找到最佳方式。卷积神经网络会自适应地学习一些关系,注意力机制更像是加入了额外的约束,使网络向期待的方向训练优化。

注意力机制可以认为是对某些维度的建模,自适应地预测潜在的关键特征,例如SE在通道维度去强化重要的通道,抑制无用的通道。注意力机制近几年也有许多相关工作,例如有效通道注意力(ECA)^[18]一文中,Wang等认为避免降维对于学习通道注意力非常重要,故而提出了一种不降维的局部跨信道交互策略;坐标注意力(CA)^[19]一文中,

Hou 等认为位置信息对于生成空间选择性非常重要,因此将通道注意力分解为两个一维特征编码过程,分别沿两个空间方向聚合特征。

通道注意力机制总是直接压缩空间维度的信息以获取不同通道间的关联性,这会损失大量的空间信息,而使得在某些任务中效果较差。关注空间重要性的注意力机制中,例如非局部神经网络利用特征图中所有位置的特征加权信息计算出每一个位置的响应,计算量非常大,这也导致像素级的建模方法难以实际应用;CBAM 等混合注意力机制在获取空间上重要性时先进行通道融合或通道压缩来减小计算量,再去获取特征图不同位置的重要性;CA 则采取分别对 X、Y 方向建模的方法,再获取重要性。

通道融合或压缩时也会损失一些信息,同时每一幅特征图的空间上的注重点可能不同,本文考虑在不进行通道融合或压缩的情况下,同时简化像素级建模方法的运算,获取不同位置 and 不同通道特征的重要性。在空间金字塔池化(SPP)^[20]网络中,SPP 块采用了固定池化窗口数,改变池化窗口大小的池化方式,从而让任意尺寸的输入产生固定维度的输出;DC-SPP-YOLO^[21]中提出了 SPP-YOLO,其中的 SPP 块没有将特征图调整为具有固定维度的特征向量,而是使用固定尺寸的滑动窗口池化来提取并融合多尺度局部区域特征。可以认为特征图在某一部分上包含的特征数是固定值,而用一定的提取特征去表征它。本文通过将特征图压缩来获得特征,也即获取注意力时通过自适应的池化来提取空间上的特征。

网络获取空间注意力的部分大多是通过通道融合或压缩得到的,但不同通道通常表征不同模式,必然会有不同之处,这样的处理方式过于粗糙。本文方法考虑对每一张特征图分别进行建模,不考虑每一像素和其他所有像素之间的关系,而是考虑它在这张特征图中的重要程度,去突出空间中较为重要的部分,也即是特征所在位置。本文提出了一种简单有效的注意力模块 SPAM。SPAM 使用池化操作去提取空间块特征,并将其作为空间点特征,缩小特征图尺寸从而减小计算量。SPAM 在不进行通道上的融合或其他操作的情况下,获得所有特征图的空间位置的重要性,通过对每张特征图进行空间位置的建模以最大程度地保留空间上的信息,同时对噪声等特异像素点进行抵抗。

1.2 SPAM 获取注意力方法

SPAM 可以分为特征提取、特征标定、通道重

建和特征聚合 4 个步骤。SPAM 结构示意图如图 1 所示,输入特征图经过一系列处理后会重新加权,重要的特征更加明显,与大多数常见注意力机制相同,它可以方便地嵌入到多种 CNN 架构中。

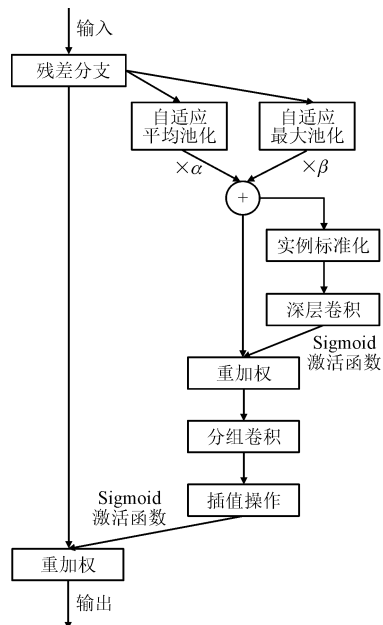


图 1 空间分割注意力模块结构示意图

Fig. 1 Schematic of SPAM

对于输入张量的中间特征图 F , 尺寸为 $C \times H \times W$, 其中 C 表示特征通道数, H 表示特征图的高度, W 表示特征图的宽度, SPAM 的处理过程如下。

步骤 1 通过自适应池化提取原始特征。认为最大池化和平均池化特征所关注的部分并不同, 提取的特征也有各自的作用, 所以同时使用了最大池化和平均池化特征。此外, 为了使网络可以学习到更需要的特征, 使用了两个可学习参数 α 和 β 控制此处的融合。这两个可训练参数不仅产生了最大池化和平均池化特征之间的自适应机制, 而且丰富了图像特征提取过程中的特征。输入特征图分别经过平均自适应池化和最大自适应池化操作得到 F_{Avg} 、 F_{Max} 两组特征图, 池化后的特征图尺寸由 $C \times H \times W$ 变为 $C \times h \times w$, 其中 $h = \lfloor \sqrt{H} \rfloor$, $w = \lfloor \sqrt{W} \rfloor$, $\lfloor \cdot \rfloor$ 表示向下取整操作。通过自适应加权提取得到所需特征 F_1 。该自适应模块的内部过程如下

$$F_1 = \alpha F_{\text{Avg}} + \beta F_{\text{Max}} \quad (1)$$

步骤 2 使用实例标准化(IN)层获取注意力特征图。一般认为对网络响应值更高的地方更为重要, 此处可以认为在原始特征图中的块状区域寻找特征是否会出现, 或出现概率的高低, 让网络关注到更重要的地方; 同时 IN 层在全局上进行强化或抑制, 可

以进一步区分特征。接下来使用深度可分离卷积层学习空间位置上的重要性,由于 IN 层破坏了通道上的关系,所以此处不再采用普通卷积,同时可以节省参数和计算量。最终使用 Sigmoid 激活函数将学习到的权重归一化,并乘以输入 F_1 ,得到 F_2 。处理过程如下式

$$F_2 = \sigma(\text{DW-Conv}(\text{IN}(F_1)))F_1 \tag{2}$$

式中 σ 表示 Sigmoid 激活函数;DW-Conv 表示深层卷积操作;IN 表示实例标准化操作。

步骤 3 使用分组卷积重建通道维度。分组卷积的组数为 $\lfloor C/16 \rfloor$,通过 Sigmoid 激活函数获得归一化权重,得到注意力权重图 F_3 。处理过程如下式

$$F_3 = \sigma(\text{GroupConv}(F_2)) \tag{3}$$

式中 GroupConv 表示分组卷积操作。

步骤 4 插值操作。当前的特征图 F_3 大小为 $C \times h \times w$ 。针对每一通道,在像素所在块的位置上得到了注意力权重。为了对原始特征图进行重标定,此处选择使用最近邻插值方式将其恢复成尺寸为 $C \times H \times W$ 的特征图,并与输入特征图 F 相乘得到最终标定后的输出特征图,处理过程如下式

$$F_4 = \text{Interpolate}(F_3)F \tag{4}$$

2 实验结果及分析

为了评估 SPAM,在常用分类数据集 ImageNet-1K^[22]、CIFAR-100、Food-101 上进行了实验。本节描述了实验设置,进行了一系列消融实验,以证明所提出的 SPAM 中每个组件对性能的贡献,从而证明了 SPAM 设计方案的有效性。将本文方法与之前的一些轻量级注意力机制进行了比较,获得

了更好的性能,这证明了 SPAM 的强大。还在其他不同大小的数据集上进行了验证,以证明 SPAM 的鲁棒性。最后,采用 Grad-CAM^[23] 工具直观地观察模型在预测时关注的图像区域,从而更好地理解网络的决策过程和预测依据。所有的实验网络均在 PyTorch^[24] 框架下实现。

2.1 消融实验

在这些消融实验中,使用的是 ImageNet-1K 数据集,并选择 ResNet50^[25] 作为注意力模块的主干模型。在训练期间,使用衰减和动量为 0.9 的标准 SGD 优化器来训练所有模型。权重衰减设置为 0.000 01,初始学习率为 0.02。使用 1 个 NVIDIA GPU 进行训练,批量大小设置为 64。在没有额外声明时将 ResNet50 作为基线网络,所有模型都训练了 90 个 Epoch。对于数据增强,使用与 ResNet^[25] 中相同的方法。

在消融实验中,验证了 SPAM 中不同组件的有效性,并探索了特征提取方式和重建通道维度的方法对网络的影响,以促进 SPAM 达到最佳性能。

首先探究了不使用池化、在原始特征图上获取注意力的方法,也即去掉最开始特征提取和最后插值部分,直接生成注意力去标定特征图;IN 层所起作用是在每张特征图上区分特征重要性,可以使网络获得全局性,此处也显示了去除 IN 层的结果;探究了去除通道重建部分的 SPAM,结果是网络性能大幅下降。此处将这几种不同构建方法分别命名为 SPAM(无池化)、SPAM(无 IN 层)、SPAM(无通道重建),实验结果见表 1。注意力模块 SPAM 的组件每一部分都有其作用,可以使网络性能更佳。

表 1 SPAM 组件消融实验结果

Table 1 Results of SPAM component ablation experiment

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
Baseline	25.56	4.11	75.984
Baseline+SPAM(无池化)	25.95	4.25	77.006
Baseline+SPAM(无 IN 层)	25.95	4.12	76.886
Baseline+SPAM(无通道重建)	25.71	4.11	74.016
Baseline+SPAM	25.95	4.12	77.062

在最初的特征提取模块中,使用了两个可训练参数 α 和 β ,用于平均池化特征和最大池化特征之间的自适应提取,这两个特征之间的自适应机制可以丰富特征图,并允许它们具有完整的关系。与只使用最大池化、只使用平均池化、组合无参数(平均池化和最大池化的特征图权重都取 0.5)的方法作了比较。为了

便于标记,这几种方法分别命名为 SPAM(最大池化)、SPAM(平均池化)、SPAM(组合无参数)。不同池化方式的 SPAM 加入网络后进行分类实验的结果见表 2,可以看到,在使用不同的池化特征时,虽然相较于基线网络都有一定的提升,但 SPAM 所采用的自适应池化达到了最高的精度。

在通道重建部分,采用分组卷积来简单重建通道之间的依赖关系。此处也使用了类似 SE 中的压缩激励方法,记为 SPAM-SE。这两种方法的作用相同,参数也相差不大。对比了这两种方法在不同的组数 g 或压缩比率 r 时的情况。不同通道重建方式的 SPAM 加入网络后进行分类实验的结果见表 3, SPAM 使用简单分组卷积增加了少量参数,即可带来网络性能提升。

在整个消融实验中,确定了 SPAM 的最终方案, SPAM 采用所有配置中效果最佳的组件和参数。在下面的实验中,将使用 SPAM 与其他的注意力机制进行比较,以验证它的有效性。

2.2 与其他注意力机制的比较

将 SPAM 与其他轻量级注意力机制进行了比较,包括常用的 SE、ECA、CBAM 和 CA。在 ImageNet-1K

数据集上进行图像分类实验,使用相同的网络和配置来严格评估注意力模块。遵循消融实验(第 2.1 节)中提到的训练规则,将不同的注意力模块嵌入网络架构。ImageNet-1 数据集分类结果见表 4,由表 4 可以看出,嵌入 SPAM 模型的网络相比基线网络有明显的提升,准确率提升了约 1.08%,这表明 SPAM 确实可以提高网络的表示能力。此外,添加了这几种注意力机制的网络增加的参数与计算量都较小,在可接受的范围内,效果相比基线网络都有一定提升,这也体现了注意力机制的作用。在这几种注意力机制中,参数量和计算量最少的 ECA 对网络的提升效果最低,而紧随其后的 SPAM 增加了微量参数和计算量,达到了最好的效果,超出 CA、SE、CBAM 等方法。使用 SPAM 的模型达到了最高的精度,这表明本文提出的方法是有效的。

表 2 不同池化方式的 SPAM 加入网络后进行分类实验的结果

Table 2 Experimental results of different pooling methods

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
基线网络	25.56	4.11	75.984
基线网络+SPAM(最大池化)	25.95	4.12	76.888
基线网络+SPAM(平均池化)	25.95	4.12	76.686
基线网络+SPAM(组合无参数)	25.95	4.12	76.922
基线网络+SPAM	25.95	4.12	77.062

表 3 不同通道重建方式的 SPAM 加入网络后进行分类实验的结果

Table 3 Experimental results of different channel reconstruction methods

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
基线网络	25.56	4.11	75.984
基线网络+SPAM-SE($r=8$)	30.74	4.18	76.654
基线网络+SPAM-SE($r=16$)	28.22	4.15	76.634
基线网络+SPAM-SE($r=32$)	26.97	4.13	76.624
基线网络+SPAM($g=\lfloor C/8 \rfloor$)	25.83	4.12	76.880
Baseline+SPAM($g=\lfloor C/16 \rfloor$)	25.95	4.12	77.062
Baseline+SPAM($g=\lfloor C/32 \rfloor$)	26.19	4.12	76.894

表 4 ImageNet-1K 数据集分类结果

Table 4 ImageNet-1K dataset classification results

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
基线网络	25.56	4.11	75.984
基线网络+SE	28.07	4.12	76.874
基线网络+ECA	25.56	4.12	76.678
基线网络+CBAM	28.07	4.14	76.710
基线网络+CA	27.47	4.14	76.734
基线网络+SPAM	25.95	4.12	77.062

2.3 在其他数据集上的分类结果

为了更全面地评估本文所提算法,也在其他的数据集上进行了图像分类实验,以下显示了在 CI-

FAR-100 和 Food-101 数据集上的分类结果。

CIFAR-100 数据集是由 100 个类、60 000 张 3 通道 RGB 彩色图像组成,每张图片尺寸为 32×32

像素,每类含有 600 张图像,其中含有 50 000 张训练图像和 10 000 张测试图像。采用基线网络为 ResNet50(第一层使用的卷积核的大小为 3×3 、步长为 1,并去掉了第一个最大池化层),在训练期间,使用衰减和动量为 0.9 的标准 SGD 优化器来训练所有模型。权重衰减设置为 0.000 5,初始学习率为 0.01,批量大小设置为 128,共训练 160 轮次。CIFAR-100 数据集分类结果见表 5,可以看到 SPAM 在此数据集上相较基线网络准确率提升了约 2.46%,相较其他注意力机制取得了最好的结果。

Food-101 数据集是由 101 个类,101 000 张 3 通道 RGB 彩色图像组成,每类含有 1 000 张图像,其中共含有 75 750 张训练图像和 25 250 张测试图像。此处所采用的基线网络为 ResNet50,在训练期间,使用衰减和动量为 0.9 的标准 SGD 优化器来训练所有模型。权重衰减设置为 0.000 5,初始学习率为 0.01,批量大小设置为 8,共训练 100 轮次。

Food-101 数据集分类结果见表 6,可以看到 SPAM 在此数据集上相较基线网络准确率提升了约 1.09%,相较其他注意力机制也取得了最好的结果。

RegNet 是 Radosavovic 等^[26]于 2020 年提出的一种卷积神经网络架构,它结合了手动设计与神经架构搜索,最终得到的网络相比之前的在性能上有较大的提升。RegNet 具有 RegNetX 和 RegNetY 两种不同架构,这两种架构的区别在于 RegNetY 中加入了 SE 模块,效果也更好一些。此处采用了和 50 层网络的 ResNet50 计算量最相近的浮点运算次数约为 4.0×10^9 的架构 RegNetY-4.0GF 作为基线网络,在 Food-101 数据集上进行训练与评估,并使用其他的注意力机制直接替换 RegNetY-4.0GF 中的 SE 模块以进行对比。改用其他注意力机制的网络分别记为 RegNetY-ECA、RegNetY-CBAM、RegNetY-CA 和 RegNetY-SPAM,实验结果见表 7。

表 5 CIFAR-100 数据集分类结果
Table 5 CIFAR-100 dataset classification results

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
ResNet50	23.71	1.31	78.070
ResNet50+SE	26.22	1.32	79.400
ResNet50+ECA	23.71	1.31	79.250
ResNet50+CBAM	26.24	1.32	80.150
ResNet50+CA	25.62	1.33	79.650
ResNet50+SPAM	24.10	1.32	80.530

表 6 Food-101 数据集分类结果
Table 6 Food-101 dataset classification results

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
ResNet50	23.71	4.13	81.216
ResNet50+SE	26.23	4.14	81.921
ResNet50+ECA	23.72	4.14	81.826
ResNet50+CBAM	26.25	4.14	82.095
ResNet50+CA	25.63	4.16	81.754
ResNet50+SPAM	24.11	4.14	82.305

表 7 使用 RegNetY 的 Food-101 数据集分类结果
Table 7 Food-101 dataset classification results using RegNetY

网络	网络参数量/MB	浮点运算次数/ 10^9	分类准确率/%
RegNetY-4.0GF	19.67	4.02	83.885
RegNetY-ECA	17.18	4.02	84.190
RegNetY-CBAM	17.90	4.02	83.347
RegNetY-CA	17.75	4.03	83.901
RegNetY-SPAM	17.43	4.02	84.602

RegNet 是通过神经架构搜索得到的,与 ResNet 等手动设计的网络有很大不同,例如很多时候RegNet 的通道数增加并不是成倍增长,每阶段中的层数也不尽相同。从表 7 可见,RegNetY-4.0GF 相较 Res-Net50 所使用的参数量和计算量都稍少一些,最终的准确率却有相当大的增长,这体现了 RegNet 的强大性。相较原本使用 SE 模块的 RegNetY-4.0GF 网络,准确率增长较多的是使用了 ECA 和 SPAM 的新架构,因为它们都采用了不降维的方法,能够更好地适应网络的变化。SPAM 相较缺少位置信息的 ECA,提升效果更好,对于准确率已相当高的 RegNetY-4.0GF 网络,使用 SPAM 直接替换其中的 SE 模块仍能增加约0.72%的准确率。

表 5、表 6 实验结果与表 4 得到的分布基本相符,而在表 7 中由于所采用基线网络的变化,不降维的方法更好。在这些不同的实验中,加入 SPAM 的网络都取得了最好的表现。在这几种不同大小数据集上以及应用了不同网络时的分类结果,证明了 SPAM 的鲁棒性良好,能够应用于不同情况。

2.4 Grad-CAM 可视化

为了能够更直接地反映 SPAM 的优势,使用 Grad-CAM 作为可视化工具,在最后一个构建块中生成使用了不同注意力机制的模型特征图的可视化。Grad-CAM 可以解释 CNN 如何通过使用梯度来计算卷积层的重要区域来进行分类决策。基于 ImageNet-1K 数据集中的图片将不同注意力机制的可视化结果进行了比较,结果如图 2 所示,在每张原始图像的顶部显示了它的真实标签。SPAM 的可视化结果在覆

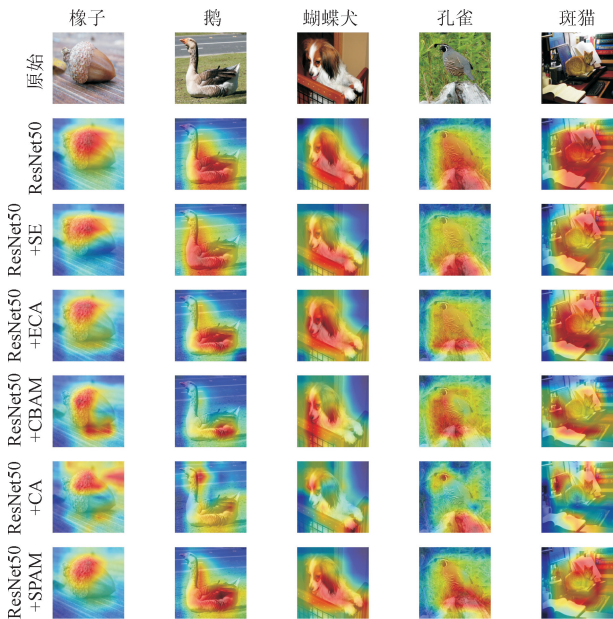


图 2 Grad-CAM 可视化结果

Fig. 2 The visualization results by Grad-CAM

盖目标类别区域方面优于其他网络。这说明了 SPAM 确实可以使网络学会专注于目标所在区域,且相较其他注意力方法能更精确地定位感兴趣的目标。可以推测,SPAM 模块可以引导网络更加关注重要的特征,而忽略不重要的特征。

3 结 论

本文对注意力机制进行了一些研究,提出了一种新的轻量级、有效的注意力模块 SPAM,它通过池化操作先减少计算量,可以在不经过通道融合或压缩的情况下获取所有特征图的注意力。SPAM 可以方便地嵌入 CNN 架构中,以显著提高网络性能。在特征提取方面,同时使用了平均池化和最大池化操作,并添加了自适应机制,以提取更为丰富、全面、有效的特征;通过先池化以局部块特征代替点特征实现空间注意力的获取,使用 IN 层和深层卷积捕获全局空间注意力;使用了分组卷积完成通道维度重建,插值恢复获得整体注意力。最终注意力模块可以使网络有效地关注目标所在区域,提高网络的性能。在 ImageNet-1K 数据集上进行的消融实验,证明了设计方案的有效性。为了评估 SPAM 的效果,将其与 SE、ECA 和 CBAM、CA 等轻量级注意力模块进行了比较,在 ImageNet-1K、CIFAR-100、Food-101 数据集上进行验证,并使用 Grad-CAM 可视化了网络的关注区域。实验结果表明,SPAM 有效提高了 CNN 的表达能力,且对不同的分类数据集和不同的网络架构具有鲁棒性。SPAM 可以促使网络更加关注目标对象所在区域,这也是注意力机制的意义所在。SPAM 有望作为一个重要组件应用于计算机视觉任务。

参考文献:

[1] GU Jiuxiang, WANG Zhenhua, KUEN J, et al. Recent advances in convolutional neural networks [J]. Pattern Recognition, 2018, 77: 354-377.

[2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[4] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.

[5] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述 [J]. 计算机工程与应用, 2021, 57(20): 64-72.

ZHANG Chenjia, ZHU Lei, YU Lu. Review of atten-

- tion mechanism in convolutional neural networks [J]. Computer Engineering and Applications, 2021, 57(20): 64-72.
- [6] GUO Menghao, XU Tianxing, LIU Jiangjiang, et al. Attention mechanisms in computer vision: a survey [J]. Computational Visual Media, 2022, 8(3): 331-368.
- [7] WANG Fei, JIANG Mengqing, QIAN Chen, et al. Residual attention network for image classification [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 6450-6458.
- [8] ZHU Minghao, JIAO Licheng, LIU Fang, et al. Residual spectral-spatial attention network for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 449-462.
- [9] HUANG Zilong, WANG Xinggang, HUANG Lichao, et al. CCNet: criss-cross attention for semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2019: 603-612.
- [10] 刘文祥, 舒远仲, 唐小敏, 等. 采用双注意力机制 Deeplabv3+算法的遥感影像语义分割 [J]. 热带地理, 2020, 40(2): 303-313.
- LIU Wenxiang, SHU Yuanzhong, TANG Xiaomin, et al. Remote sensing image segmentation using dual attention mechanism Deeplabv3+ algorithm [J]. Tropical Geography, 2020, 40(2): 303-313.
- [11] LE V T, KIM Y G. Attention-based residual autoencoder for video anomaly detection [J]. Applied Intelligence, 2023, 53(3): 3240-3254.
- [12] PU Chunyu, HUANG Hong, YANG Liping. An attention-driven convolutional neural network-based multi-level spectral-spatial feature learning for hyperspectral image classification [J]. Expert Systems with Applications, 2021, 185: 115663.
- [13] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [14] 张珂, 冯晓晗, 郭玉荣, 等. 图像分类的深度卷积神经网络模型综述 [J]. 中国图象图形学报, 2021, 26(10): 2305-2325.
- ZHANG Ke, FENG Xiaohan, GUO Yurong, et al. Overview of deep convolutional neural networks for image classification [J]. Journal of Image and Graphics, 2021, 26(10): 2305-2325.
- [15] WANG Xiaolong, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7794-7803.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Berlin, Germany: Springer International Publishing, 2018: 3-19.
- [17] CHEN Long, ZHANG Hanwang, XIAO Jun, et al. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 6298-6306.
- [18] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2020: 11531-11539.
- [19] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2021: 13708-13717.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [21] HUANG Zhanchao, WANG Jianlin, FU Xuesong, et al. DC-SPP-YOLO: dense connection and spatial pyramid pooling based YOLO for object detection [J]. Information Sciences, 2020, 522: 241-258.
- [22] RUSSAKOVSKY O, DENG Jia, SU Hao, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [C]//2017 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2017: 618-626.
- [24] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library [C]//Advances in Neural Information Processing Systems. San Francisco, CA, USA: Curran Associates, Inc., 2019: 8024-8035.
- [25] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [26] RADOSAVOVIC I, KOSARAJU R P, GIRSHICK R, et al. Designing network design spaces [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2020: 10425-10433.

(编辑 武红江 杜秀杰)