

## 单参数通道注意力模块

姚亮亮<sup>1,2,3</sup>, 张太红<sup>1,2,3\*</sup>, 张洋宁<sup>1,2,3</sup>, 温钊发<sup>1,2,3</sup>

(1. 新疆农业大学 计算机与信息工程学院, 新疆 乌鲁木齐 830052;

2. 智能农业教育部工程研究中心, 新疆 乌鲁木齐 830052;

3. 新疆农业信息化工程技术研究中心, 新疆 乌鲁木齐 830052)

**摘要:**随着深度学习的发展,通道注意力在卷积神经网络上的表征能力上发挥了巨大的作用。为了进一步加强通道注意力模块在深度神经网络中的作用,针对通道注意力的参数量方面,提出了一种单参数通道注意力(APA)模块。首先,APA模块在图像通道特征的求和向量上加单参数。然后,通过度量通道向量和求和向量在方向上的关系,求得通道注意力权重。最后,经过激活函数(Sigmoid)激活注意力权重,使其分布更平稳。与其他通道注意力模块相比,该模块只有微量参数,且该模块的代码实现非常简单。在数据集CIFAR-10与CIFAR-100上,使用APA模块嵌入到MobileNet, ResNet系列主干,与同类方法压缩激励模块(SE)、有效的通道注意力模块(ECA)进行了实验对比,验证了APA模块的有效性。

**关键词:**卷积神经网络;深度学习;通道注意力;图像分类;计算量

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2023)12-0215-06

doi:10.3969/j.issn.1673-629X.2023.12.030

## Single-parameter Channel Attention Module

YAO Liang-liang<sup>1,2,3</sup>, ZHANG Tai-hong<sup>1,2,3\*</sup>, ZHANG Yang-ning<sup>1,2,3</sup>, WEN Zhao-fa<sup>1,2,3</sup>

(1. School of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China;

2. Intelligent Agriculture Ministry of Education Engineering Research Center, Urumqi 830052, China;

3. Xinjiang Agricultural Information Engineering Technology Research Center, Urumqi 830052, China)

**Abstract:** With the development of deep learning, channel attention has played a huge role in the representation ability of convolutional neural networks. In order to further strengthen the role of channel attention module in deep neural network, a single parameter channel attention (APA) module is proposed for the parameter quantity of channel attention. First, the APA module adds a single parameter to the image channel feature summation vector. Then, the channel attention weight is obtained by measuring the relationship between the channel vector and the summation vector in the direction. Finally, the attention weight is activated by the activation function (Sigmoid) to make its distribution more stable. Compared with other channel attention modules, the proposed module has only trace parameters, and the code implementation of the module is largely simple. By embedding the APA module into the MobileNet and ResNet series backbones on the data sets CIFAR-10 and CIFAR-100, it is compared with the similar method Squeeze-and-Excitation module (SE) and Effective Channel Attention Module (ECA), which shows the effectiveness of the APA module.

**Key words:** convolutional neural network; deep learning; channel attention; image classification; amount of calculation

### 0 引言

从计算机视觉任务诞生起始,方法从机器学习主导转为深度学习主导,而卷积神经网络也成为计算机视觉中的主要方法。以卷积神经网络为主干模型的任务包括图像分类<sup>[1-2]</sup>、目标检测<sup>[3-4]</sup>、图像分割任务<sup>[5-6]</sup>等。而现存采用较为广泛的主干模型包括 ResNet<sup>[7]</sup>系列、MobileNet<sup>[8]</sup>系列等。而这些模型的成功离不开卷积神经网络优秀的结构设计。卷积神经网络结构的

设计大致分为两方面,一个是模型整体<sup>[7-8]</sup>的设计,另一个是模型中即插即用<sup>[9-10]</sup>的模块。因此,网络设计是深度卷积神经网络中一个必不可少的因素。

当前,卷积神经网络的结构变得越来越复杂,模型层数的增加,或者是模型宽度的增加。但是其基本操作都是围绕卷积为核心,归一化和激活函数为辅助构成的。随着网络深度越来越深,宽度越来越宽,深度卷积神经网络的表征能力也越来越强,尤其是在大数据

收稿日期:2023-02-14

修回日期:2023-06-15

基金项目:科技创新2030——“新一代人工智能”重大项目(2022ZD0115805);新疆维吾尔自治区重大科技专项(2022A02011)

作者简介:姚亮亮(1997-),男,硕士研究生,研究方向为计算机视觉;通信作者:张太红(1965-),男,博导,博士,研究方向为人工智能。

上的表现超乎其想象。但是,在小数据集上越深的网络或越大的网络,对其效果越不好。大网络模型由于其神经元超多,也就是参数超多,对待小数据集就会陷入过拟合。所以,为了减少参数量,避免由于神经元过多而产生过饱和这样的现象,模型的发展往往朝着轻量化发展,同时也是为了适应工业的发展,加快模型的推理速度。伴随着这一需求的出现,卷积神经网络的设计成为一个极具挑战性的问题。

除去卷积神经网络本身发展的更迭,一种方便提升卷积神经网络的表征能力的模块<sup>[9-10]</sup>也发展得尤为火热,例如通道注意力模块。这种模块独立于卷积神经网络,不受神经网络的约束,可以根据各种神经原理、数学原理方便地计算通道的注意力权重。且这种模块适用范围广,可以方便地应用到计算机视觉的各种任务。

为了更好地改进通道注意力,使之更为精巧,该文主要关注使用一个参数更好地提取到通道的注意力权重。主要贡献如下:

(1)提出一种即插即用的而非卷积神经网络新架构的通道注意力模块(APA)。

(2)在数据集 CIFAR-10 与 CIFAR-100 使用该模块在多个流行的卷积神经网络主干上进行图像分类,取得了良好的性能。

## 1 相关工作

在这部分,将叙述卷积神经网络的设计和卷积神经网络中注意力模块的代表性工作。

### 1.1 卷积神经网络设计的发展

20 世纪 60 年代,加拿大科学家从猫的视觉中枢发现了感受野、双眼视觉等视觉功能,这是神经网络发展的里程碑事件。1980 年左右,出现了类似卷积神经网络的神经认知网络,用于手写字符识别和其他识别任务。1998 年, Yann LeCun 等人<sup>[11]</sup>提出了基于梯度下降的学习算法—LeNet,这是卷积神经网络后续发展的开端。LeNet 网络由简单卷积层和全连接层组成。2012 年,随着 AlexNet<sup>[1]</sup>的出现,深度卷积神经网络从此主导了计算机视觉任务,当然计算机硬件资源的提升也功不可没。AlexNet 的突出特点是采用了非线性激活函数 Relu 和 Dropout 层,旨在解决梯度消失问题,防止过拟合。然后在 2014 年, VGG<sup>[2]</sup>网络使用了大量的 3×3 卷积和 2×2 池来扩展卷积神经网络的感受野。同年出现了 GoogLeNet<sup>[12]</sup>,整个卷积神经网络由初始模块组成,其本质是将信道分成几个感受野不同的信道,从而得到不同的感受野。初始模块使用了大量的 1×1 卷积,其主要作用是降低数据的维数,节省计算量。2015 年, ResNet<sup>[7]</sup>网络的出现,让卷积网络更加

深入。ResNet 主要利用剩余模块组成整个网络,其跳转连接结构在缓解深度网络梯度消失的同时,极大地提高了网络特征提取的能力。2016 年, DenseNet<sup>[13]</sup>重用了网络中各层的特征图,旨在缓解梯度消失问题,提高网络表示能力,减少网络参数。2018 年, MobileNet<sup>[8]</sup>采用深度可分卷积作为网络的基本模块,减少了参数数量和计算量,使模型更轻便。

### 1.2 通道注意力模块

在人类的视觉系统中,人眼总是对感兴趣的事物关注度比较高,对不感兴趣的事物关注度低,或完全不关注。例如,在电影院中,为了让观众获得更好的体验,往往采取关灯的方式,让观众只关注电影屏幕。同样在卷积神经网络中,为了让卷积神经网络获得更好的性能,研究者会采用更好的方式去抑制或提高图像特征块中通道的重要性,给每个通道赋予一个注意力权重。在深度卷积神经网络提取特征的过程中,特征图的不同通道和不同空间位置通常是不同的部分。在计算机视觉的任务中,不同部分的相对重要性是不同的。为了使网络更有效地学习特征,注意力机制可以通过自适应学习对不同部分赋予不同的权重,这可以看作是一个特征选择的过程。通道注意力的通用形式可以表述为:

$$X_o = X_i * f(X_i) \quad (1)$$

式中,  $X_o$  表示输出特征向量,  $X_i$  表示输入特征向量,  $f(\cdot)$  表示计算注意力权重的方法。从公式(1)中可以看出计算通道注意力的过程分为两步:

(1)根据输入特征向量利用计算通道注意力的方法计算出通道注意力权重,也就是式中的  $f(X_i)$ 。

(2)将计算出的通道注意力权重乘到输入特征向量的对应通道中,也就是  $X_i * f(X_i)$ 。

当前,卷积神经网络中最具有代表性的通道注意力模块当属压缩激励模块(SE)<sup>[9]</sup>。在压缩激励网络这篇文章中首先提出了通道注意力的概念,压缩激励模块计算通道注意力权重的方式分为两部分:压缩与激励。压缩部分是通过全局平均池化计算通道的全局上下文信息,以获得通道的依赖关系。激励部分是通过全连接层去调整通道的依赖关系,以使获得的通道注意力权重更加准确。至此,每年的计算机视觉顶级会议中都会出现一些通道注意力模块相关的设计,不过大多数都是压缩激励模块方法的改进。其中,最具有吸引力的是有效的注意力模块(ECA)<sup>[14]</sup>,这种模块改进压缩激励模块中的激励部分,认为获得更合适的注意力权重,不应该使用降维的方式,遂采取一维卷积来代替激励部分的全连接层,该模块简单有效,且参数量极少。由 Gao 等人<sup>[15]</sup>提出的 GSoP-Net 在压缩模块中使用全局二阶池(GSoP)来捕获全局信息,以提高

其捕获高阶统计信息的能力。Lee 等人<sup>[16]</sup>提出的 SRM 模块采用与信道宽度相同的 style pooling 和 full connection layer (CFC) 分别代替压缩激励模块 (SE) 中的压缩和激励部分。Yang 等人<sup>[17]</sup>提出的 GCT 通过范数捕获信道的全局信息,并应用一个可学习的向量 ( $\alpha$ ) 来缩放该特征。然后采用通道规范化的竞争机制进行通道间的互动。像其他常见的正常归一化方法一样,可学习的尺度参数 ( $\gamma$ ) 和部分  $\beta$  用于调整归一化。最后,  $\tanh$  激活用于控制注意力载体。同时国内有很多关于通道注意力的应用,如 Jiang 等人<sup>[18]</sup>应用压缩激励模块 (SE) 进行表情识别。Zhang 等人<sup>[19]</sup>利用通道注意力提升表情识别任务的性能。

## 2 方法

在这部分先介绍单参数通道注意力 (APA) 模块的原理和代码实现,再讨论卷积神经网络中具有代表性的通道注意力模块压缩激励模块 (SE)<sup>[9]</sup> 和有效的通道注意力模块 (ECA)<sup>[14]</sup> 与 APA 模块在操作上的差异以及参数量的对比。文中方法和传统的通道注意力模块在计算通道注意力权重的方式上大有不同,但是最终的目的都是为了计算通道注意力权重。

### 2.1 APA 参数设置与代码实现

从减少通道注意力模块参数的角度, 该文提出了一种只有单参数通道注意力模块 (APA)。APA 在计算通道注意力权重是基于简单的数学原理。流程如图 1 所示。

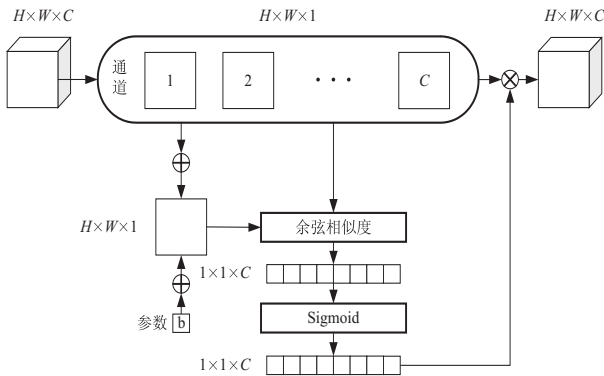


图 1 单参数的通道注意力 (APA) 模块流程

计算通道注意力的方式同传统的通道注意力一样,也分为两部分。设输入特征向量为  $F_{[C,H,W]}$ , 其中  $C, H, W$  分别代表输入特征向量的通道数以及高和宽。第一部分如公式 (2), 先将  $F_{[C,H,W]}$  的通道求和, 然后在得到的向量上加一个参数 ( $b$ ) 得到特征向量 ( $F_{[1,H,W]}$ )。第二部分如公式 (3), 将第一部分得到的特征向量和输入特征向量的每个通道求取余弦相似度, 然后通过 Sigmoid 激活函数进行非线性激活, 最后得到了通道注意力权重。再将得到的通道注意力权重

乘到输入通道向量的对应通道中, 如公式 (4)。

$$F_{[1,H,W]} = \sum_{i=0}^C F_{[i,H,W]} + b \quad (2)$$

$$\text{Attention} = \text{Sigmoid}(\cos(F_{[1,H,W]}, F_{[i,H,W]})) = \text{Sigmoid}\left(\frac{\sum_{j=0}^{H \times W} F_{[1,j]} \times F_{[i,j]}}{\sqrt{\sum_{j=0}^{H \times W} (F_{[1,j]})^2} \times \sqrt{\sum_{j=0}^{H \times W} (F_{[i,j]})^2}}\right) \quad (3)$$

$i = 1, 2, \dots, C$

$$F_{\text{output}} = \text{Attention} \times F_{[C,H,W]} \quad (4)$$

接下来, 解释一下这样做的原因。之所以对通道求和是因为求和这项操作会将通道的依赖关系隐式映射到  $F_{[1,H,W]}$  上, 而添加  $b$  是因为要利用余弦相似度去度量  $F_{[C,H,W]}$  与  $F_{[1,H,W]}$  的相似度, 余弦相似度是为了度量两个向量在方向上的关系。 $b$  的改变会导致  $F_{[1,H,W]}$  在方向上发生改变, 如图 2 所示, 从而达到动态调整通道依赖关系的目的, 以使得到的通道注意力权重更加适合各个卷积神经网络以及各类视觉任务。

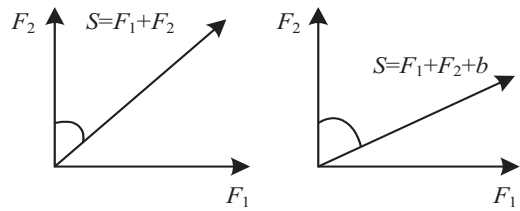


图 2 参数 ( $b$ ) 对向量 ( $S$ ) 在方向上的影响

为了说明 APA 的简单性, 图 3 给出了 APA 模块基于 Pytorch 框架的实现代码。

```

APA 代码
def APA(self, input)
    # input_size: [B, C, H, W], b_size: [1]
    S = torch.sum(input, dim = 1) + b
    # cos
    output = Cosine Similarity (S, input, dim = C)
    return sigmoid(output)
    
```

图 3 “APA 代码”形式表现

### 2.2 APA 与 SE, ECA 的差异

表 1 展现了文中模块 (APA) 和压缩激励 (SE)、有效的注意力模块 (ECA) 在操作上的差异以及参数量的对比。

表 1 SE, ECA, APA 模块操作以及参数的对比

注意力模块	操作	参数
SE <sup>[9]</sup>	GAP, FC, ReLU, Sigmoid	$2 C^2 / r$
ECA <sup>[14]</sup>	GAP, C1D, Sigmoid	$k$
APA	+, cos, Sigmoid	1

注: GAP, FC, C1D 分别代表全局平均池化、全连接层、一维卷积;  $C, r, k$  分别表示通道数、压缩激励模块中衰减系数、卷积核的尺寸。

### 2.3 APA 插入卷积神经网络主干的位置

APA 是一个即插即用的灵活模块,可以插入到当前流行卷积神经网络模型中,以此来增强模型对图像通道特征的代表能力。APA 在 ResNet 系列及其改进系列(PreResNet, WideResNet),插入其卷积网络模块的示意如图 4 所示。

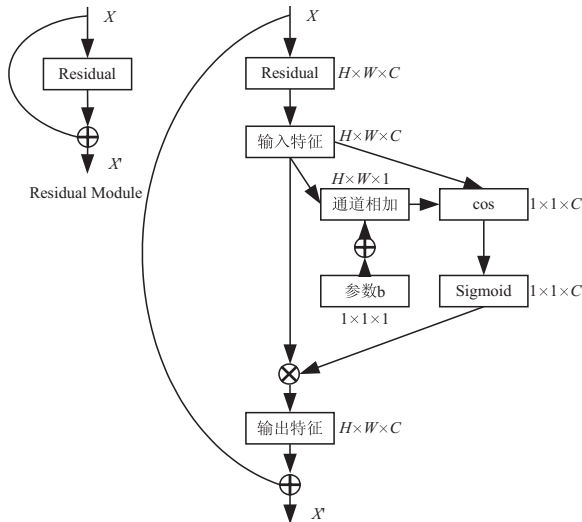


图 4 单参数通道注意力模块(APA)嵌入 ResNet 系列的示意图

## 3 实验结果

在这部分,先叙述一下实验环境与数据集情况,再

表 2 不同通道注意力模块的实验结果。

方法	TOP-1 准确率/% (CIFAR-10)	TOP-1 准确率/% (CIFAR-100)	参数/M	额外参数	推理速度 (FPS)
ResNet-20 <sup>[7]</sup>	92.33±0.19	68.88±0.15	0.272 5	0	256
+SE <sup>[9]</sup>	92.42±0.14	69.45±0.11	0.280 5	0.010 8 M	204
+ECA <sup>[14]</sup>	92.35±0.35	68.89±0.57	0.272 5	40	222
+APA	92.63±0.18	68.58±0.29	0.027 5	9	206
ResNet-56 <sup>[7]</sup>	93.58±0.20	72.24±0.37	0.855 8	0	129
+SE <sup>[9]</sup>	93.69±0.14	72.84±0.51	0.880 0	0.024 2 M	100
+ECA <sup>[14]</sup>	93.68±0.14	72.45±0.38	0.855 8	98	102
+APA	94.15±0.21	74.21±0.26	0.855 8	18	99
ResNet-110 <sup>[7]</sup>	94.51±0.31	75.54±0.24	1.730 7	0	77
+SE <sup>[9]</sup>	94.68±0.22	76.56±0.30	1.779 1	0.038 4 M	51
+ECA <sup>[14]</sup>	94.72±0.15	76.33±0.65	1.730 7	212	52
+APA	94.81±0.14	75.75±0.22	1.730 7	36	51
PreResNet-20	92.14±0.25	68.70±0.30	0.272 6	0	308
+SE <sup>[9]</sup>	92.24±0.06	68.70±0.21	0.280 6	0.014 0 M	206
+ECA <sup>[14]</sup>	92.16±0.25	68.31±0.46	0.272 6	40	214
+APA	92.29±0.20	68.76±0.09	0.272 6	9	203

讨论一下文中模块在当前流行的卷积神经网络主干上进行图像分类的实验结果。

### 3.1 实验环境

该文的参数设置遵循所有模型的标准训练流程<sup>[20]</sup>。具体来说,每张图像进行零填充,每侧 4 个像素,并从该填充图像或其水平翻转中随机裁剪一个 32 × 32 的图像供训练使用。评价过程中,所有模型均使用原始测试集图像进行测试。优化通过动量为 0.9,批量大小为 128,权重衰减为 0.000 5 的 SGD<sup>[21]</sup> 优化器完成。所有网络都在单个 GPU 上进行训练。学习率以 0.1 开始,在 32 000 和 48 000 次迭代(在 64 000 次迭代时停止划分)时除以 10。

该文使用作者提供的公共代码,因为所有方法具有随机性,所以实验均是进行 3 次重复后取平均值得出的结果。

实验显卡为一块 24G 的 NVIDIA GeForce RTX 3090。

### 3.2 数据集介绍

所采用的数据集是图像分类中的公开数据集 CIFAR-10<sup>[22]</sup> 与 CIFAR-100<sup>[22]</sup>。CIFAR-10 源自一个共有 8 000 万张微小图像的数据集。CIFAR-10 共有 6 万张 32×32 的彩色图像,其中 5 万张图像用于训练,1 万张图像用于测试。

续表 2

方法	TOP-1 准确率/% (CIFAR-10)	TOP-1 准确率/% (CIFAR-100)	参数/M	额外参数	推理速度 (FPS)
PreResNet-56	93.71±0.24	71.83±0.23	0.590 4	0	136
+SE <sup>[9]</sup>	93.57±0.15	72.57±0.32	0.848 5	0.258 1 M	75
+ECA <sup>[14]</sup>	93.78±0.14	72.43±0.45	0.590 4	98	80
+APA	93.97±0.06	74.11±0.20	0.590 4	18	72
PreResNet-110	94.22±0.18	75.95±0.22	1.146 8	0	69
+SE <sup>[9]</sup>	94.40±0.18	76.68±0.33	1.662 9	0.516 1 M	39
+ECA <sup>[14]</sup>	94.70±0.31	76.11±0.54	1.146 8	212	42
+APA	94.87±0.18	76.60±0.15	1.146 8	36	40
WideResNet-28×10	95.78±0.10	81.31±0.39	36.479 2	0	155
+SE <sup>[9]</sup>	96.24±0.04	80.31±0.08	37.554 4	1.075 2 M	126
+ECA <sup>[14]</sup>	96.12±0.15	80.30±0.22	36.479 2	156	134
+APA	96.09±0.09	80.67±0.32	36.479 2	25	125
MobileNet-V2 <sup>[24]</sup>	91.86±0.12	71.32±0.09	2.296 9	0	145
+SE <sup>[9]</sup>	91.79±0.20	71.54±0.31	2.410 8	0.103 9 M	105
+ECA <sup>[14]</sup>	92.34±0.23	71.24±0.45	2.296 9	59	116
+APA	92.22±0.15	73.15±0.06	2.296 9	19	100

共有 10 个类别,每个类别 6 000 张图像。CIFAR-100 同 CIFAR-10 一样是 8 000 万微小图像微小数据集的子集。CIFAR-100 也有 6 万张 32×32 的彩色图像,其中 5 万张用于训练,1 万张用于测试,不同的是 CIFAR-100 有 100 个类,每个类有 600 张图像。

### 3.3 评价指标

实验所采用的评价指标有 TOP-1 准确率(预测概率排名第一的类别与实际标签相符的准确率)、插入到卷积网络模型的模型参数、相比较基础模型的额外参数、推理速度(FPS)。推理速度(FPS)表示每秒预测图像的张数,实验结果中的推理速度是采用 NVIDIA GeForce RTX 3060 的 6G 显存版本在 1 万张图像上平均所得。

### 3.4 结果分析

文中模块(APA)在 CIFAR-10 与 CIFAR-100 上进行了图像分类任务,并与同类方法压缩激励模块(SE)与有效的通道注意力模块(ECA)进行比较。由于文中方法是卷积神经网络即插即用的模块,所以要嵌入到卷积神经网络主干中进行实验。实验的卷积神经网络主干包括 ResNet-20, ResNet-56, ResNet-110, PreResNet-20, PreResNet-56, PreResNet-110, WideResNet-20×10, MobileNet-v2<sup>[23]</sup>。从实验结果可以看出,APA 以绝对的准确率优势战胜基线模块,分别在 CIFAR-10 与 CIFAR-100 中以 ResNet-20, ResNet-56, ResNet-110, PreResNet-20, PreResNet-56,

PreResNet-110, WideResNet-28×10, MobileNet-v2 为主干提升了 0.3 百分点、-0.3 百分点、0.57 百分点、1.97 百分点、0.3 百分点、0.21 百分点、0.15 百分点、0.06 百分点、0.26 百分点、2.28 百分点、0.65 百分点、0.65 百分点、0.31 百分点、-0.64 百分点、0.36 百分点、1.83 百分点,表明注意力模块在卷积神经网络中是非常可取的。且与同类方法相比,该文以绝对参数优势取得了较好的精确率,较基线只有 9, 18, 36, 9, 18, 36, 25, 19 的额外参数,也说明了文中模块是简单且有用的通道注意力模块。从推理速度(FPS)上看 APA 模块的推理速度和 SE 模块相近,略低于 ECA 模块。但是,SE 模块和 ECA 模块具有复杂的超参数设置,需要经过大量的实验去寻找相对于各类任务的最优超参数。而 APA 模块没有额外的超参数需要设置,不需要大量时间调优去适应视觉任务。

## 4 结束语

该文提出了一种只有单参数注意力模块(APA),相比较同类模块,模块参数量极少,例如在 ResNet-20 上只有 9 个额外参数,MobileNet-v2 上只有 19 个附加参数,而在更深的网络 ResNet-110 上也只有 36 个额外参数。且 APA 模块在准确率上和同类方法相比也有绝对的优势。综上所述,APA 是一个在结构上简单、在模型上有效的通道注意力模块,且该模块有更多的机会应用到更多的计算机视觉任务中,如目标检测、

图像分割。

参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2014, 60(6): 84–90.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv:1409.1556, 2014.
- [3] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137–1149.
- [4] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440–1448.
- [5] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE international conference on computer vision. Venice: IEEE, 2017: 2980–2988.
- [6] KE L, TAI Y W, TANG C K. Deep occlusion-aware instance segmentation with overlapping bilayers [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. [s. l.]: IEEE, 2021: 4019–4028.
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770–778.
- [8] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [J]. arXiv:1404.04861, 2014.
- [9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 7132–7141.
- [10] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 3–19.
- [11] 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.
- [12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 1–9.
- [13] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2014: 4700–4708.
- [14] LEE H J, KIM H E, NAM H. Srm: a style-based recalibration module for convolutional neural networks [C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 1854–1862.
- [15] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 11534–11542.
- [16] GAO Z, XIE J, WANG Q, et al. Global second-order pooling convolutional networks [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 3024–3033.
- [17] YANG Z, ZHU L, WU Y, et al. Gated channel transformation for visual recognition [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 11494–11803.
- [18] 姜丽莉, 黄承宁. 融合注意力机制改进残差网络的表情识别方法 [J]. 计算机技术与发展, 2022, 32(5): 42–46.
- [19] 张波, 兰艳亭, 李大威, 等. 基于卷积网络通道注意力的人脸表情识别 [J]. 无线电工程, 2022, 52(1): 148–153.
- [20] LEE C Y, XIE S, GALLAGHER P, et al. Deeply-supervised nets [C]//Artificial intelligence and statistics. San Diego: JMLR, 2015: 562–570.
- [21] LECUN Y, BOTTOU L, ORR G B. Neural networks: tricks of the trade [J]. Canadian Journal of Anaesthesia, 2012, 41(7): 421–436.
- [22] KRIZHEVSKY A, HINTON G. Convolutional deep belief networks on cifar-10 [J]. Unpublished Manuscript, 2010, 40(7): 1–9.
- [23] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 4510–4520.