门控机制的图像分类网络

姜文涛,高 原*,袁 姮,刘万军
(辽宁工程技术大学软件学院,辽宁葫芦岛 125105)

摘 要: 为了提取更具表达能力和区分度的重点特征,减少网络传递时关键特征的流失,提高神经网络图像分类能力,提出一种新的门控机制图像分类网络(image classification Network of Gating Mechanism,GMNet). 首先,使用门控卷积提取浅层特征,通过门控机制选择性地进行卷积操作,提高网络对原始图像关键特征的提取能力;其次,设计了一种插值门控卷积(Interpolation Gated Convolution,IGC)模块,利用Lanczos插值与门控卷积相结合,强化浅层特征的同时提取更具区分度的特征,提高特征的非线性表达能力;然后,设计了大核门控注意力机制(Large kernel Gated Attention Mechanism,LGAM)模块,将大核注意力与门控卷积相融合,实现了特征的选择性增强和选择性融合,提高关键区域特征的贡献度;最后,将大核门控注意力机制模块嵌入到残差分支中,让模型更有效地学习输入数据的特征和上下文信息,减少关键特征在网络信息传递时流失,提高网络的分类能力. 本文方法在图像数据集CIFAR-10、CIFAR100、SVHN、Imagenette、Imagewoof上分别达到了97.05%、83.68%、97.68%、90.60%、83.05%的分类准确率,与当前先进的方法相比分别平均提高了3.26%、7.08%、3.44%、2.65%、5.02%. 与现有主流网络模型相较,本文门控机制图像分类网络能够增强特征的非线性表达能力,提取更具表达能力和区分度的重点特征,减少关键特征流失,提高关键区域特征的贡献度,有效地提高神经网络图像分类能力.

关键词: 图像分类;门控机制;门控卷积;插值门控卷积;大核门控注意力;残差网络

基金项目: 国家自然科学基金(No.61601213);辽宁省自然科学基金(No.20170540426);辽宁省教育厅重点基金(No.LJYL049)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2024)07-2393-14

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20240104

Image Classification Network of Gating Mechanism

JIANG Wen-tao, GAO Yuan*, YUAN Heng, LIU Wan-jun
(*School of Software*, *Liaoning Technical University*, *Huludao*, *Liaoning* 125105, *China*)

Abstract: To extract more expressive and discriminative key features, reduce the loss of key features during network transmission, and improve the image classification ability of neural networks, a new image classification network of gating mechanism (GMNet) is proposed. Firstly, the shallow features are extracted using gated convolution, and the convolution operation is selectively performed through the gating mechanism to improve the network's ability to extract key features of the original image. Secondly, an interpolation gated convolution (IGC) module is designed, which combines Lanczos interpolation with gated convolution to enhance shallow features while extracting more discriminative features, improving the non-linear expression ability of features. Then, a large kernel gated attention mechanism (LGAM) module is designed, which combines large kernel attention with gated convolution to achieve selective enhancement and fusion of features, and improve the contribution of key region features. Finally, the large kernel gated attention mechanism module is embedded into the residual branch to enable the model to learn input data's features and contextual information more effectively, reduce the loss of key features during network information transmission, and improve the network's classification ability. The method achieved classification accuracy of 97.05%, 83.68%, 97.68%, 90.60%, and 83.05% on image datasets CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewoof, respectively, and improved on average by 3.26%, 7.08%, 3.44%, 2.65%, and 5.02% compared to current advanced methods. Compared with existing mainstream network models, the gated mechanism image classification network proposed in this paper can enhance the non-linear expression ability of features, extract more expressive and discriminative vital features, the loss of key features, improve the contribution of key region features,

收稿日期:2024-01-26;修回日期:2024-06-05;责任编辑:覃怀银
*通讯作者:高原

and effectively improve the image classification ability of neural networks.

Key words: image classification; gating mechanism; gated convolution; interpolation gated convolution; large kernel gated attention; residual network

Foundation Item(s): National Natural Science Foundation of China (No.61601213); Liaoning Provincial Natural Science Foundation (No.20170540426); Key Fund of Liaoning Provincial Department of Education (No.LJYL049)

1 引言

图像分类是计算机视觉领域研究热点之一,其利用计算机视觉和机器学习技术,对图像进行特征提取和识别,从而实现对图像的自动分类[1]. 由于图像分类过程中存在图像多样性、复杂背景、提取特征不充分等因素,容易发生特征丢失等问题,如何解决这些问题,提高图像分类的准确率,是图像分类领域的研究热点和难点[2].

随着卷积神经网络(Convolutional Neural Network,CNN)的出现,基于CNN的图像分类方法得到快速发展. 文献[3~5]验证了增加网络深度可以有效提高模型的准确率. 但是,随着CNN网络层数的加深,在反向传播过程中,梯度会被逐层乘以权重和偏差,当网络深度增加时,梯度可能会变得非常小或非常大,导致模型无法有效地更新权重和偏差,进而产生梯度消失和梯度爆炸等问题. He等[6]提出ResNet(Residual Network),通过构建深度网络中的残差连接,解决网络训练时梯度消失和梯度爆炸的问题,进一步提高模型的准确率. 文献[7~11]利用残差结构增加模型的表达能力,适应更加复杂和多样化的任务要求.

值得注意的是,上述分类方法处理图像分类问题时没有充分考虑通道之间的相关性以及长距离依赖关系,忽略通道维度适应性,导致网络不能有效地学习到不同通道的特征信息. 为了解决这一问题,文献[12,13]通过引入通道注意力和空间注意力对输入特征进行细粒度自适应调整,增强模型在处理复杂场景和学习显著特征的能力.

随着Transformer[14]在自然语言处理领域取得的显著成绩,文献[15~17]将Transformer应用到了计算机视觉任务中,Transformer采用自注意力机制,相比传统的注意力机制具有更好的捕捉长距离依赖关系、参数少、并行化计算和可解释性更强.

上述分类方法中均引入了注意力机制,以增强网络对图像中重要特征的关注度,从而提升对关键信息的捕获能力,然而,这些模型忽略了网络输入层获取的特征图对整个网络的重要性. 在网络输入层提取的特征图中,常常混入大量的背景和噪声因素,这些因素可能导致部分关键特征在后续网络传递过程中丢失,注意力机制难以有效解决背景和噪声对模型准确率产生的负面影响. 基于此,本文提出一种门控机制的图像分类网络(image classification Network of Gating Mechanism,GMNet).

2 插值门控卷积与大核门控注意力

本文将门控机制与插值操作相结合,设计了IGC(Interpolation Gated Convolution)模块,通过插值操作在放大图像时更好地保留图像的细节和特征,再结合门控卷积操作实现特征的还原和增强,提高特征的非线性表达能力.

将门控机制与大核注意力相融合,设计了LGAM模块,通过门控机制来提高关键区域的贡献度,减少关键特征在网络信息传递时流失,实现了特征的选择性增强和融合.

2.1 门控卷积

门控卷积是一种特殊的卷积操作,旨在通过引入门控机制,控制信息的流动,从而捕捉图像中相邻像素之间的高阶空间交互关系. Yu等[18]提出DeepFillv2,作者在文中提出一种gated convolution的特殊卷积方式,如图1所示.

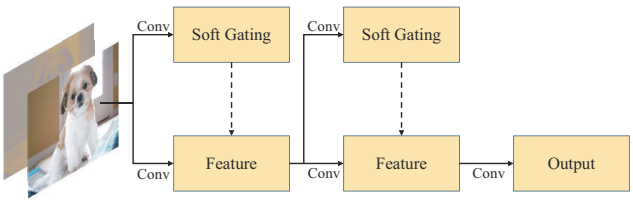


图1 门控卷积

卷积运算包含两个部分:(1)对特征图进行卷积操作后,通过激活函数进行非线性变换;(2)对相同的特征图进行卷积操作后,利用Sigmoid函数产生门控信号. 最后将两个部分输出的特征图进行逐像素相乘,其运算规则如式(1)~(3)所示:

$$\boldsymbol{G}_{x,y}=\sum\sum \boldsymbol{W}_g\cdot \boldsymbol{I} \tag{1}$$

$$\boldsymbol{F}_{x,y}=\sum\sum \boldsymbol{W}_f\cdot \boldsymbol{I} \tag{2}$$

$$\boldsymbol{O}_{x,y}=\varphi\left(\boldsymbol{F}_{x,y}\right)\odot\sigma\left(\boldsymbol{G}_{x,y}\right) \tag{3}$$

其中,$\boldsymbol{I}$是特征图,$\boldsymbol{W}_g$和$\boldsymbol{W}_f$是两个不同的卷积滤波器,φ是激活函数(例如,ReLU、ELU和Leaky ReLU),σ是Sigmoid函数.

门控卷积采用门控机制来提升模型的表达能力以及泛化能力,利用门控卷积,可以有效地学习输入数据

的关键特征,提高对原始图像关键特征的提取能力.

2.2 插值门控卷积模块

插值是一种常用的图像处理方法,用于图像的放大、缩小或重采样等处理. 图像插值操作是在图像已知像素点之间,通过数学计算得到新像素点的数值,从而达到调整图像尺寸或者图像分辨率的目的. 在深度学习中,通常使用卷积层来提取图像的特征,但这会导致图像的尺寸变小,而插值可以有效地恢复图像的原始尺寸,使得模型可以更好地处理原始图像数据.

考虑到在图像分类任务中,特征图的信息对分类效果有重要的影响,而随着网络的加深,特征图尺寸变小,会导致部分关键特征信息丢失而影响模型的分类精度. 因此,本文将门控机制与插值操作相结合,构建了IGC模块,通过插值操作放大图像,更好地保留图像的细节和特征,再结合门控卷积提取更具区分度的特征,提高特征的非线性表达能力,实现特征的还原和增强.

IGC结构如图2所示,首先,将浅层提取到的特征图通过一个3×3的门控卷积进行特征进一步提取,然后,通过Lanczos插值[19]算法将特征图尺寸放大到原来的2倍,最后再通过一个3×3的门控卷积进行特征提取,即

$$\boldsymbol{F}_{\mathrm{IGB}}=\mathrm{G\_Gonv}\left(\gamma\left(\mathrm{G\_Gonv}\left(\boldsymbol{F}_{\mathrm{input}}\right)\right)\right)$$

其中,$\boldsymbol{F}_{\mathrm{IGB}}$表示IGC模块输出特征,G_Gonv表示门控卷积操作,γ表示Lanczos插值操作,$\boldsymbol{F}_{\mathrm{input}}$表示输入特征.

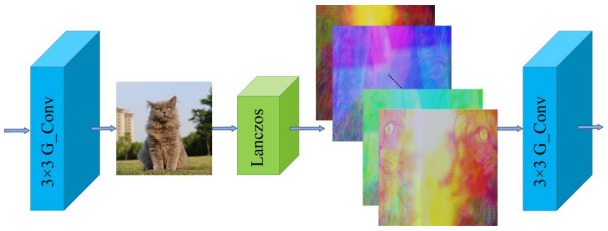


图2　插值门控卷积模块

由于近邻插值基本原理是假设每个像素的值与其最近的像素值相同,这种方法计算量较小,速度较快,但是缩放后的图像质量相对较低;双线性插值基本原理是根据两个已知点的斜率,计算出待插值点的近似值,这种方法在处理具有复杂形状的图像时可能无法准确地描述这些形状,导致图像失真;双三次插值基本原理是利用待采样点周围16个点作三次插值,虽然生成的图像质量较佳,但其算法复杂度偏高. 本文的插值门控卷积模块使用Lanczos插值如图3所示,其计算过程如式(4)~(6)所示.

$$L(x)=\begin{cases}1, & x=0\\ \dfrac{a\sin(\pi x)\sin(\pi x/a)}{\pi^2x^2}, & 0<|x|<a\\ 0, & \text{otherwise}\end{cases}\tag{4}$$

其中,a为滤波器尺寸参数,$L(x)$为一维Lanczos插值的重建核.

$$L(x,y)=L(x)L(y)\tag{5}$$

$$S(x,y)=\sum_{i=\lfloor x\rfloor-a+1}^{\lfloor x\rfloor+a}\sum_{j=\lfloor y\rfloor-a+1}^{\lfloor y\rfloor+a}s_{ij}L(x-i)L(y-j)\tag{6}$$

其中,x,y是原图像像素坐标,$L(x,y)$为二维Lanczos插值的重建核,s_{ij}为原图像(i,j)位置像素值,$S(x,y)$为目标插值结果.

在a为3时,Lanczos插值考虑了输入图像与输出图像像素点映射位置最邻近的8个像素点的像素值,并利用这8个像素值计算输出目标图像中像素点的像素值,因此,Lanczos插值在缩小和放大图像时能够更好地保留图像的细节和特征,提供更高的图像质量和更好视觉效果.

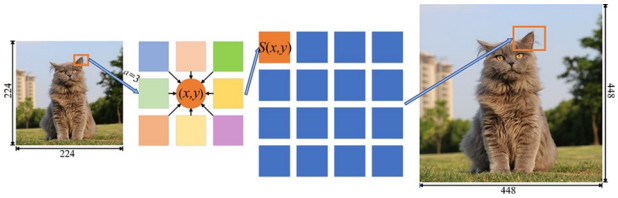


图3　Lanczos插值图像处理示意图

2.3 大核门控注意力机制模块

传统的卷积神经网络通过层级结构提取特征,但这种结构无法明确表示不同特征之间的交互和依赖关系,为了解决这个问题,受VAN(Visual Attention Network)[20]的启发,本文将门控卷积与大核注意力相融合,设计了大核门控注意力机制(LGAM)模块. LGAM模块在考虑了局部结构信息、长程依赖性的同时,关注通道维度的适应性,减少关键特征流失,并且通过门控机制来提高关键区域的贡献度,降低由于单一特征不足而导致识别或分类的不确定性,实现了特征的选择性增强和选择性融合.

LGAM结构如图4所示,LGAM的核心是通过两次跳跃连接将激活函数与大核门控卷积(LGA)模块相结合形成注意力机制. 其中,LGA主要由三部分组成:空间局部卷积(深度门控卷积)、空间长距离卷积(深度膨胀卷积)和通道卷积(1×1卷积). 深度门控卷积提取空间显著纹理特征,可以减少图像复杂背景对分类方法训练的影响;深度膨胀卷积在保证感受野扩大的同时降低计算量,可以提高模型训练效率以及捕捉局部特征的能力;1×1卷积可以实现特征提取和降维的效果.

如图4所示,在LGAM的第一个分支中,输入特征$\boldsymbol{F}$先进行GELU[21]激活函数,对输入数据进行非线性变换,得到输出特征:

$$\boldsymbol{F}_G=\mathrm{GELU}(\boldsymbol{F})$$

其中,$\boldsymbol{F}_G$表示第一分支输出特征,GELU(x)表示GELU

激活函数，其计算过程如式(7)所示：

$$\mathrm{GELU}(x)=0.5x\left[1+\tanh\left(\sqrt{\frac{2}{\pi}}\left(x+0.047\,715x^3\right)\right)\right] \tag{7}$$

由式(7)可知，GELU激活函数相比于一些传统的激活函数(如ReLU)更加平滑，并且是非线性的．这意味着在反向传播算法中，它有更强的梯度特性，有助于训练更深、更复杂的神经网络．

然后在LGAM的第二分支中，依次经过两层3×3的深度门控卷积，一层卷积核为5×5、dilation为3、padding为6的深度膨胀卷积，两层3×3的深度门控卷积，一层1×1的通道卷积，得到输出特征：

$$\begin{aligned}&\boldsymbol{F}_{\mathrm{GL}}\\&=\mathrm{Conv}_{1\times1}(\mathrm{DW\_G\_Conv}(\mathrm{DW\_G\_Conv}(\mathrm{DW\_D\_Conv}\\&\quad(\mathrm{DW\_G\_Conv}(\mathrm{DW\_G\_Conv}(\boldsymbol{F}_G))))))\end{aligned}$$

其中，$\boldsymbol{F}_{\mathrm{GL}}$表示第二分支输出特征，$\mathrm{Conv}_{1\times1}$表示1×1大小的通道卷积，DW_G_Conv表示深度门控卷积，DW_D_Conv表示深度膨胀卷积．

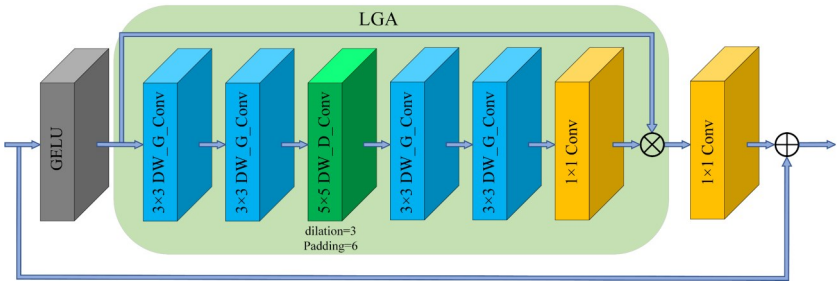


图4 LGAM模块结构图

将第二分支输出特征$\boldsymbol{F}_{\mathrm{GL}}$与第一分支输出特征$\boldsymbol{F}_G$相乘，得到第二分支注意力图：

$$\boldsymbol{F}_{A1}=\boldsymbol{F}_G\otimes\boldsymbol{F}_{\mathrm{GL}}$$

其中，$\boldsymbol{F}_{A1}$表示第二分支注意力图．

最后将输出的注意力图经过1×1卷积操作，以确保特征通道数与最终输出的通道数一致，然后与输入特征$\boldsymbol{F}$进行融合，获得注意力模块的最终输出特征，即

$$\boldsymbol{F}_{A2}=\boldsymbol{F}\oplus\mathrm{Conv}_{1\times1}\left(\boldsymbol{F}_{A1}\right)$$

其中，$\boldsymbol{F}_{A2}$表示最终的注意力图．

在卷积神经网络中，浅层参数的微弱变化经过多层线性变换与激活函数后会被放大，这种放大的效应会在网络反向传播时产生显著的影响，导致反向传播梯度较小，参数更新速度较慢，最终使网络训练难以收敛．针对该问题，本文在LGAM模块后采用批量归一化方法(Batch Normalization，BN)，缓解梯度消失，加快收敛．

3 门控机制的图像分类网络

为了提取更具表达能力和区分度的重点特征，减少关键特征在网络中流失，提高神经网络图像分类能力，本文提出门控机制的图像分类网络(GMNet)．

GMNet以ResNet-34残差网络为基础融合了本文设计的IGC模块和LGAM模块：

(1)将网络输入层改为门控卷积，修改卷积核尺寸，筛选图像的关键特征．再经过IGC模块，进一步增强图像的关键特征．

(2)将LGAM模块嵌入残差分支中，加强网络对关键特征的学习能力，减少关键特征在网络信息传递时流失，提高网络模型的学习能力．

3.1 浅层特征提取模块

在ResNet-34残差网络中，首先使用7×7的卷积层和最大池化层提取图像的浅层特征，但普通卷积在提取特征时会对所有像素作为有效像素进行提取，这必然会使特征图混入大量的无效特征进而影响模型的分类效果．而池化层的目的是通过下采样减少特征图的尺寸，从而间接增大卷积层的感受野，以确保获取更多的特征信息，但是，随着卷积核尺寸的增大，池化层可能会丢失原始图像的许多信息．因此，我们需要提取更加细微的关键特征以保持图像的关键信息．

受生物视觉系统的启发，门控卷积在特征提取时会重点提取关键特征信息，贴近生物视觉系统在观察外界时会重点关注生物关键特征的机制．因此，本文用门控卷积替换首层普通卷积，同时调整门控卷积的步长和填充大小，平衡网络参数的特征表达．

如图5所示，将普通卷积层输出的特征图$\boldsymbol{F}$通过Sigmoid激活函数产生门控信号，得到一个像素值全部限制在0~1之间的特征图$\boldsymbol{F}_S$，其计算过程如式(8)所示：

$$\boldsymbol{F}_S=\mathrm{Sigmoid}(\boldsymbol{F})=\frac{1}{1+\mathrm{e}^{(-\boldsymbol{F})}} \tag{8}$$

其中，$\mathrm{Sigmoid}(x)$表示Sigmoid激活函数．

再将普通卷积层输出的特征图$\boldsymbol{F}$通过斜率系数为0.01的LeakyReLU激活函数得到特征图$\boldsymbol{F}_R$，其计算方式为

$$\boldsymbol{F}_R=\mathrm{LeakyReLU}(\boldsymbol{F})$$

其中，$\mathrm{LeakyReLU}(x)$为LeakyReLU激活函数，其运算规

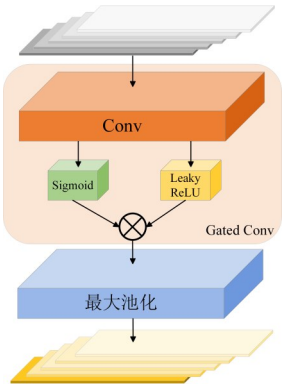


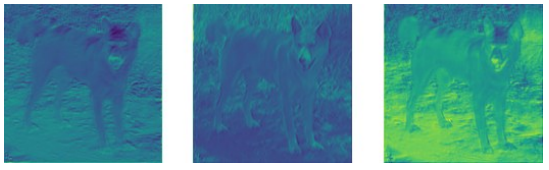
图5　浅层特征提取模块示意图

则如式(9)所示:

$$\mathrm{LeakyReLU}(x)=\begin{cases}x, & x>0\\ ax, & x\leqslant 0\end{cases} \tag{9}$$

最后,将$\boldsymbol{F}_S$与$\boldsymbol{F}_R$进行逐像素相乘更好地捕捉图像中相邻像素之间的高阶空间交互关系,减少信息丢失的可能性,增强模型对图像的敏感度.

为进一步观察门控机制的效果,图6给出了两种卷积操作下输出的特征图. 图6(*a*)和图6(*b*)分别是经过普通卷积和门控卷积输出的特征图,由图可知,门控卷积具有更好的特征提取能力,通过门控机制可以灵活地控制信息传播,更有效地从数据中提取关键特征.



(*a*) 普通卷积

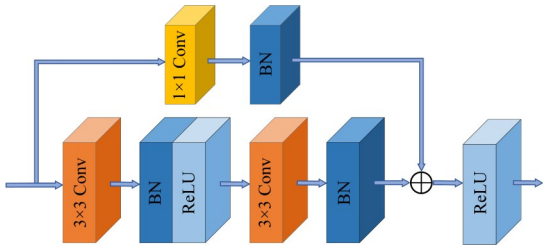
(*b*) 门控卷积

图6　两种卷积输出的特征图

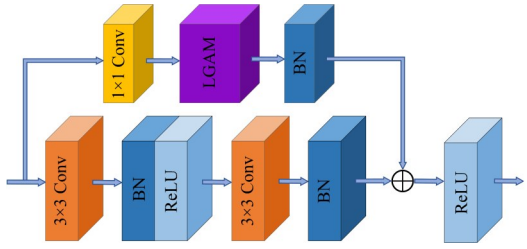
3.2　融合大核门控注意力机制的残差模块

ResNet通过残差块实现跨层连接,直接传递梯度,避免梯度在逐层传播过程中消失或爆炸,使得网络更容易训练,其结构如图7(*a*)所示.

本文将LGAM模块嵌入残差块的残差分支中,其结构如图7(*b*)所示,由2个3×3卷积和残差分支组成,残差分支由1×1卷积与LGAM模块相结合,考虑局部结构信息、长程依赖性的同时关注通道维度的适应性,减少关键特征流失,并且通过门控机制来提高关键区域的贡献度,降低由于单一特征不足而导致的识别或分类的不确定性,提高了网络模型对核心目标的聚焦能力. 在每次卷积后,输出的特征图像经过BN层和ReLU层,防止梯度消失,加快收敛,增强网络层间的特征关系.



(*a*) 降维后的残差块



(*b*) 融合LGAM的残差块

图7　两种残差块

为了提高网络模型对核心目标的聚焦能力,解决梯度消失和梯度爆炸等问题. 将LGAM嵌入残差分支中,考虑局部结构信息、长程依赖性的同时关注通道维度的适应性,减少关键特征在网络信息传递时候流失,让模型更有效地学习输入数据的特征和上下文信息,提高网络的分类能力. 改进前后的残差函数拟合关系如图8所示. ResNet的残差映射关系如式(10)所示:

$$H(x)=F(x)+x \tag{10}$$

其中,$H(x)$表示待拟合对象,$F(x)$表示残差函数,x表示学习输入.

融合LGAM模块的残差映射关系如式(11)所示:

$$H(x)=F(x)+L(x) \tag{11}$$

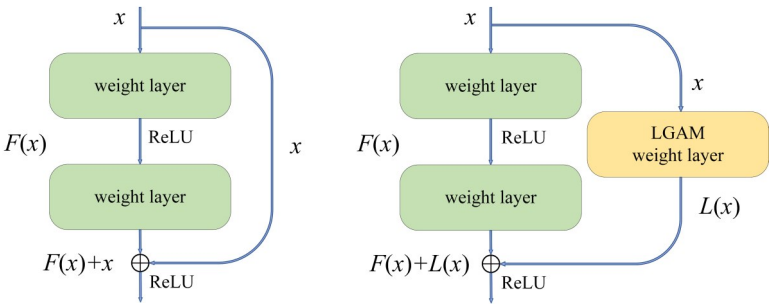
其中,$L(x)$表示LGAM注意力机制函数.

通过融合LGAM模块优化残差映射,提高网络模型对核心目标的聚焦能力,解决梯度消失和梯度爆炸等问题.

3.3　GMNet总体结构

GMNet主要包含3个模块:门控卷积模块、插值门控卷积模块和融合大核门控注意力机制的残差模块,网络整体结构如图9所示. 模型训练分为五个阶段:

(1)图像预处理阶段. 对输入图像进行图像增强预处理操作,即随机翻转、填充后随机剪切和cutout操作,提高图像的质量和可读性,使神经网络能够更好地提



(a) 残差映射关系 (b) 融合LGAM后的残差映射关系

图8 两种残差映射关系

取特征,提高模型的泛化能力和鲁棒性.

(2)浅层特征提取阶段. 将预处理后的图像输入网络的门控卷积层,利用门控机制提取浅层特征,再利用最大池化层对特征映射降维.

(3)特征增强阶段. 将输出的特征图经过插值门控卷积模块,提取更具表达力和区分度的特征,提高特征的非线性表达能力,实现特征还原和增强.

(4)深层特征提取阶段. 将输出的特征图输入融合大核门控注意力机制的残差模块,进行更深层特征提取,获取图像高层抽象特征.

(5)输出阶段. 将高层抽象特征输入平均池化层进行下采样,再经过全连接层生成图像分类结果.

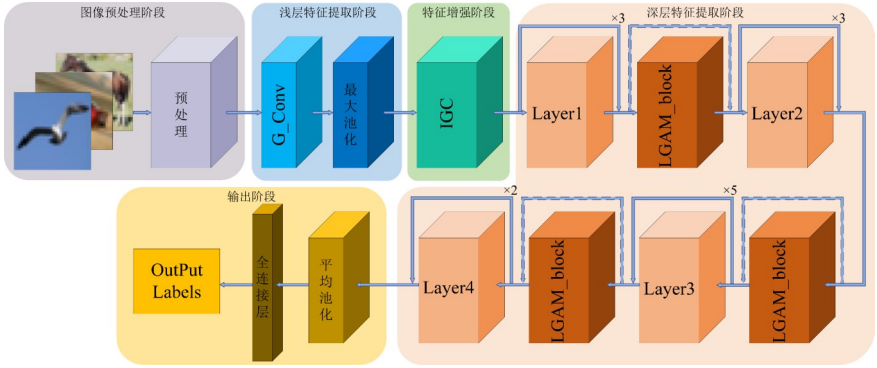


图9 GMNet总体结构

总之,GMNet针对ResNet-34的改进如下:

(1)为了提高网络输入层的特征提取能力,采用门控卷积替换普通卷积,调整卷积核、步长、填充大小,提高对原始图像关键特征的提取能力.

(2)为了进一步保留图像的细节和特征,采用IGC模块提取更具表达力和区分度的特征,实现特征的还原和增强.

(3)为了让模型更有效地学习输入数据的特征和上下文信息,在残差分支中嵌入LGAM模块,提高关键区域的贡献度,减少关键特征在网络信息传递时候流失,实现特征的选择性增强和融合.

4 实验与结果分析

4.1 实验环境

本文实验采用Python语言和Pytorch框架进行网络搭建,操作系统采用了Windows11系统,GPU采用NVIDIA RTX 3090显卡,显存容量24 GB. 为评估GMNet的性能,选择CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof数据集作为实验数据集,数据集属性如表1所示.

表1 实验数据集

名称	图像尺寸/px	分类数	训练集数量	测试集数量
CIFAR-10	32×32	10	50 000	10 000
CIFAR-100	32×32	100	50 000	10 000
SVHN	32×32	10	73 257	26 032
Imagenette	224×224	10	9 469	3 925
Imagewoof	224×224	10	9 025	3 929

在训练过程中,迭代次数设置为200,采用随机梯度下降(Stochastic Gradient Descent,SGD)优化器,标签平滑为0.1,动量为0.9,权重衰减为5×10^{-4},学习率为0.1,CIFAR数据集和SVHN数据集批次大小为128,Imagenette数据集和Imagewoof数据集批次大小为64.

本文实验均在数据增强后的数据集上进行,采用图像分类常用指标分类准确率(accuracy)作为评价指标.

4.2 影响网络性能的因素

影响GMNet性能的主要参数包括DW_G_Conv嵌入LGAM的数量及位置、LGAM中DW_D_Conv卷积核尺寸、融合LGAM残差块嵌入的位置与数量、IGC中Lanczos插值的缩放因子(scale_factor)大小、首层门控卷积核尺寸k.

4.2.1 嵌入LGAM的数量和位置对GMNet性能的影响

门控卷积与大核注意力相融合,考虑局部结构信息、长程依赖性的同时关注通道维度的适应性,减少关键特征流失,并且通过门控机制来提高关键区域的贡献度,提高了网络模型对核心目标的聚焦能力.

为深入研究DW_G_Conv嵌入LGAM的数量和位置对分类准确率的影响,在DW_D_Conv卷积核尺寸确定的情况下设计了9种LGA模块的组合方式,如图10所示.

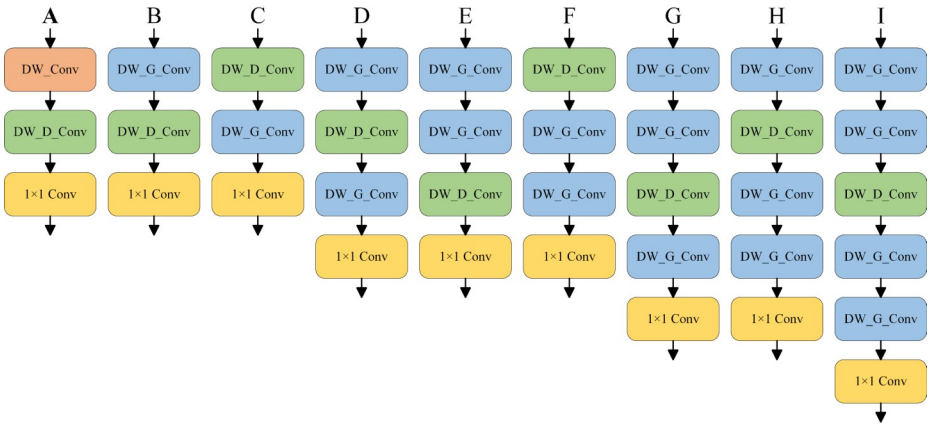


图10　DW_G_Conv位置和数量的9种组合方式

其中A类为原VAN,其他分别为加入不同数量和位置DW_G_Conv的LGAM,其在5个数据集上对分类准确率的影响如图11所示.

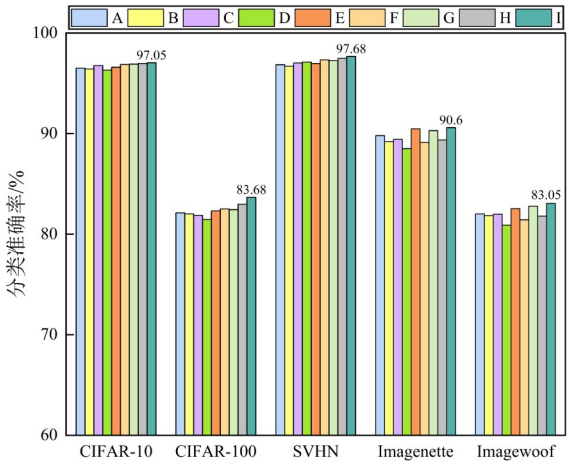


图11　不同位置和数量DW_G_Conv对分类准确率的影响

由图11可知,排列方式I在数据集CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof上分别取得了97.05%、83.68%、97.68%、90.60%、83.05%的分类准确率,均高于其他组合方式. 因此,在DW_D_Conv前后分别加入两层DW_G_Conv时网络的分类效果最佳,继续增加DW_G_Conv的层数可能会导致模型过于复杂,从而出现过拟合现象,降低模型的泛化能力.

4.2.2 LGAM中卷积核尺寸对GMNet性能的影响

DW_D_Conv卷积核尺寸决定了LGAM对图像核心目标的聚焦能力,因此,分别采用5种不同大小尺寸的卷积核及其他参数如表2所示,并分别在5个数据集进行实验,实验结果如图12所示.

表2　不同尺寸的DW_D_Conv　　单位:px

Method	Kernel size	Padding	Dilation
G_0	3×3	2	2
G_1	5×5	6	3
G_2	7×7	9	3
G_3	9×9	12	4
G_4	11×11	15	4

由图12可知,DW_D_Conv卷积核尺寸过大或过小都会影响网络模型对图像核心目标的聚焦能力,参数大小为G_1时,GMNet在5个数据集上准确率均最优.

4.2.3 融合LGAM残差块嵌入的位置与数量对GMNet性能的影响

本文以ResNet-34为基础,网络主干分为4个Layer部分,在4个Layer部分之间有3个残差连接模块(简记为R-Block),R-Block的设计目的是解决深度神经网络训练过程中的梯度消失和表示瓶颈问题. 通过引入残差连接,网络的训练变得更加稳定,可以更深层次地挖掘数据的特征.

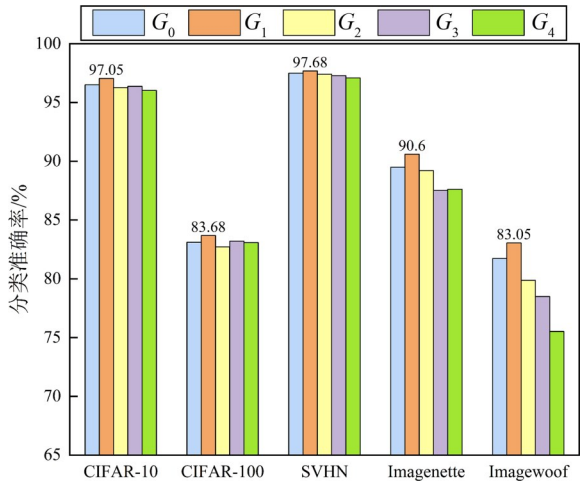


图 12 DW_D_Conv 卷积核尺寸对分类准确率的影响

由于在残差连接之前使用注意力机制，可能会在降维过程中损失更多的关键特征，为了减少关键特征流失，提高网络模型对核心目标的聚焦能力，本文在残差连接处嵌入融合 LGAM 的残差块（简记为 L-Block）.

为了研究融合 LGAM 残差块嵌入的位置与数量对 GMNet 性能的影响，本文设计 8 种排列组合方式，如图 13 所示.

嵌入 L-Block 的 8 种组合方式在 5 个数据集上的分类准确率如图 14 所示.

由图可知，组合方式 H 在数据集 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 上分别取得了 97.05%、83.68%、97.68%、90.60%、83.05% 的分类准确率，均高于其他组合方式. 因此，在 3 个残差连接处嵌入 3 个 L-Block 时网络的分类效果最佳.

为进一步观察融合 LGAM 残差块对模型复杂度的影响，如表 3 所示，分别列出在 5 个数据集上 LGAM 残差块对网络训练时间的影响. 其中，网络 R_N 为原 ResNet-34，网络 L_N 为在 3 个残差连接处嵌入融合 LGAM 的残差网络.

结合图 14 和表 3 可知，虽然在网络中引入 LGAM 会增加一定的开销，但其带来的分类性能的提升远大于这些额外的开销，所以，在网络中引入 LGAM 是一种有效的方法，能够有效提高分类效果.

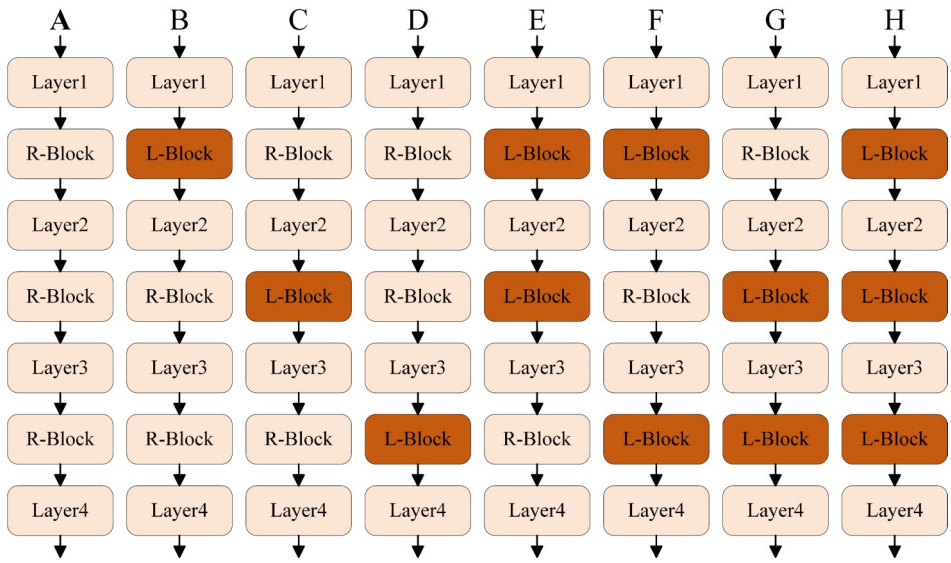


图 13 嵌入 L-Block 的 8 种组合方式

4.2.4 IGC 中 Lanczos 插值对 GMNet 性能的影响

IGC 通过插值操作在放大图像时更好地保留图像的细节和特征，再结合门控卷积提取更具区分度的特征，提高特征的非线性表达能力，实现特征的还原和增强. 因此 IGC 中 Lanczos 插值的 scale_factor 大小直接影响图像的细节和特征. 本文分别使用大小为 1、2、3 的 scale_factor 对 5 个数据集进行实验，其中，scale_factor 为 1 时，IGC 模块不使用 Lanczos 插值对图像进行放大，只进行了两层门控卷积进行特征提取；scale_factor 为 2 和 3 时，IGC 模块使用 Lanczos 插值对图像进行放大原来的 2 倍和 3 倍. 实验结果如表 4 所示.

由表 4 可见，5 个数据集均在 scale_factor 大小为 2 时取得了最优值，继续增大 scale_factor 可能会导致模型过于复杂，从而出现过拟合现象，降低模型的泛化能力.

4.2.5 首层门控卷积核尺寸 k 对 GMNet 性能的影响

网络输入层获取的特征图直接影响网络模型的分类能力. 本文分别使用大小为 3×3、5×5、7×7、9×9、11×11 的卷积核对 5 个数据集进行实验，其中 CIFAR-10、CIFAR-100、SVHN 数据集由 32×32 图像组成，为防止原始图像特征损失，删除最大池化层；而 Imagenette、

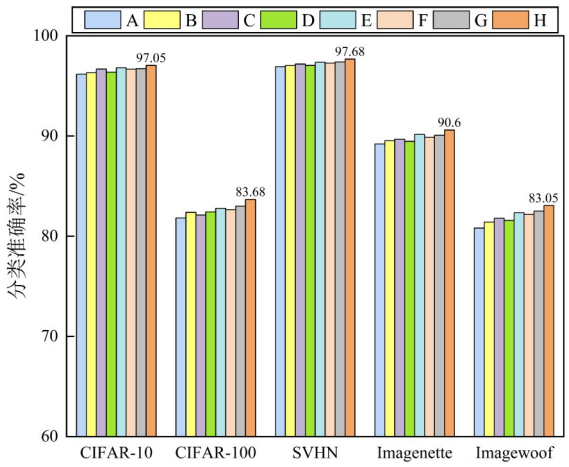


图 14　嵌入 L-Block 的 8 种组合方式对分类准确率的影响

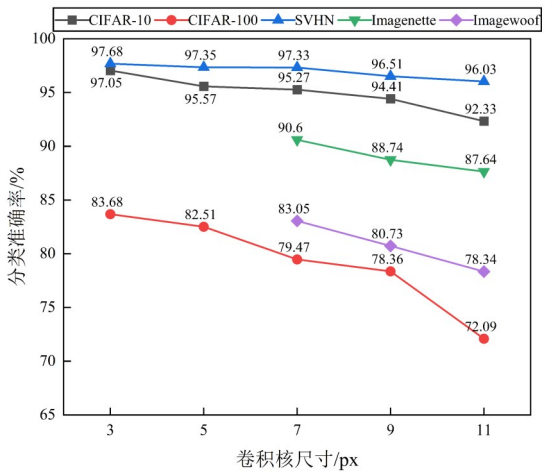


图 15　首层门控卷积核尺寸对分类准确率的影响

Imagewoof 数据集由 224×224 图像组成,由于图像尺寸过大,采用 7×7、9×9、11×11 的卷积核对其进行实验,并且为降低特征图空间大小,保留最大池化层. 首层门控卷积核尺寸对分类准确率的影响如图 15 所示.

由图 15 可知,分类准确率并未随卷积核尺寸的增大而提升,在图像大小为 32×32 数据集中尺寸为 3×3 的卷积核上分类效果最佳,图像大小为 224×224 数据集中尺寸为 7×7 的卷积核上分类效果最佳. 增大首层门控卷积核尺寸并不能提高对原始图像关键特征的提取能力,反而会出现特征提取能力减弱的现象,从而导致网络模型分类准确率下降.

4.3　对比实验

为验证 GMNet 的先进性,在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上进行对比实验.

对比分类方法如下:ResNet-34[6],Efficient-Nets[10],GhostNet[11],CAPR-DenseNet[8],Multi-ResNet[9],WideResnet-28-10[7],Couplformer[16],FAVOR+(Fast Attention via Positive Orthogonal Random Features Approach)[15],MMA-CCT-7/3×2[17],DAMSNet[22].

本文的分类准确率对比实验数据来源如下:

(1)未开源的代码优先采用对比网络对应论文提供的实验结果.

(2)通过论文提供的开源代码复现.

各网络在 5 个数据集上的分类准确率对比如表 5 所示,表中黑体数字表示最优值. 由表 5 可得出,GMNet 在 5 个数据集上的分类准确率最高.

表 3　LGAM 对网络训练时间的影响　　单位:s

网络	每轮训练时间				
	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
R_N	38	38	55	26	25
L_N	49	49	71	32	31

表 4　scale_factor 大小对分类准确率的影响　　单位:%

Scale factor	分类准确率				
	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
1	96.50	81.29	97.40	88.69	80.55
2	**97.05**	**83.68**	**97.68**	**90.60**	**83.05**
3	96.89	83.47	97.43	90.32	81.98

为进一步验证 GMNet 的有效性,ResNet-34 和 GMNet 在 32×32 数据集 CIFAR-10 和 224×224 数据集 Imagewoof 上的分类混淆矩阵如图 16 所示,图中列出了数据集各个类别的正确与错误的样本数量. 与 ResNet-34 相比 GMNet 在 CIFAR-10、Imagewoof 数据集上有更多的正确样本数量,以及更少的错误样本数量,因此 GMNet 具有出色的类别区分能力,表现出较强的分类能力.

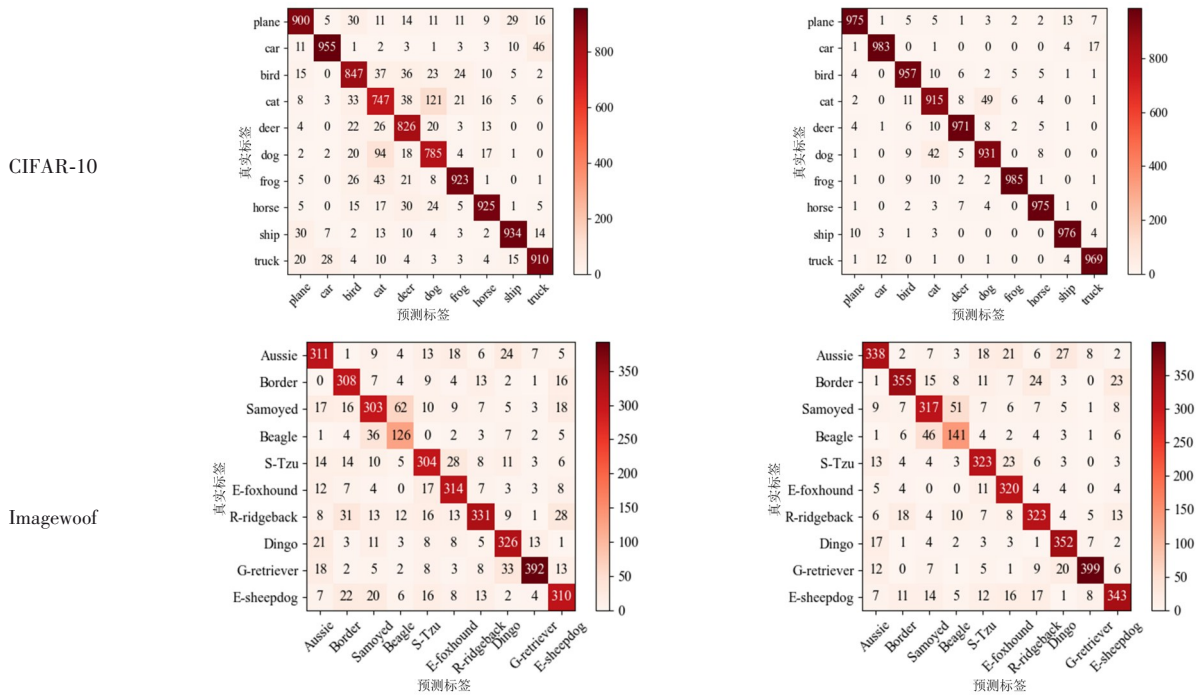
4.4　消融实验

为了验证本文网络各模块的有效性,在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上进行消融实验.

定义如下网络:Net_1 表示在 GMNet 中去掉残差分支中

表5 各网络在5个数据集上的分类准确率 单位:%

网络	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
ResNet-34	88.09	71.41	91.51	87.66	77.96
EfficientNets	94.01	75.96	93.32	88.01	77.93
GhostNet	94.92	77.15	93.86	87.83	78.22
CAPR-DenseNet	94.24	78.84	94.95	87.72	77.91
Multi-ResNet	94.65	78.68	—	—	—
WideResnet-28-10	95.83	79.50	95.21	88.34	78.71
Couplformer	93.54	73.92	94.26	87.91	77.89
FAVOR+	91.42	72.56	93.21	88.16	77.57
MMA-CCT-7/3×2	94.74	77.50	—	—	—
DAMSNet	96.51	80.50	97.60	—	—
GMNet	**97.05**	**83.68**	**97.68**	**90.60**	**83.05**



(*a*) ResNet (*b*) GMNet

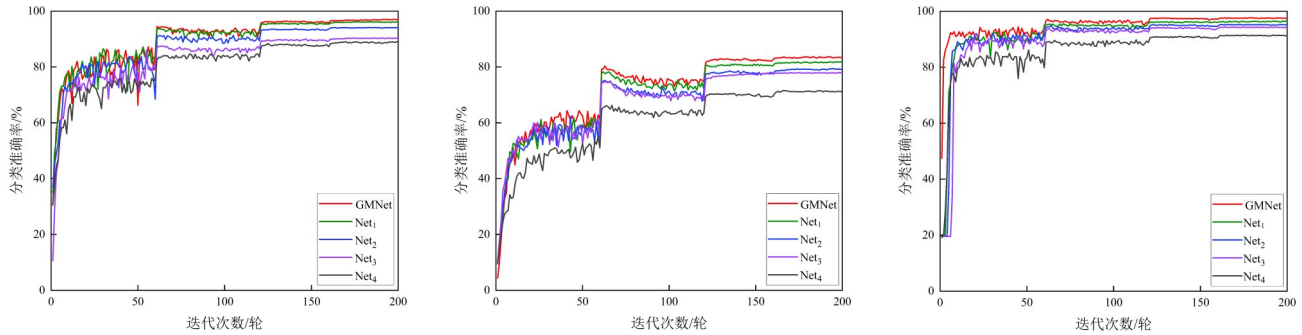
图16 ResNet-34和GMNet在2个数据集上的混淆矩阵

表6 GMNet的消融实验结果 单位:%

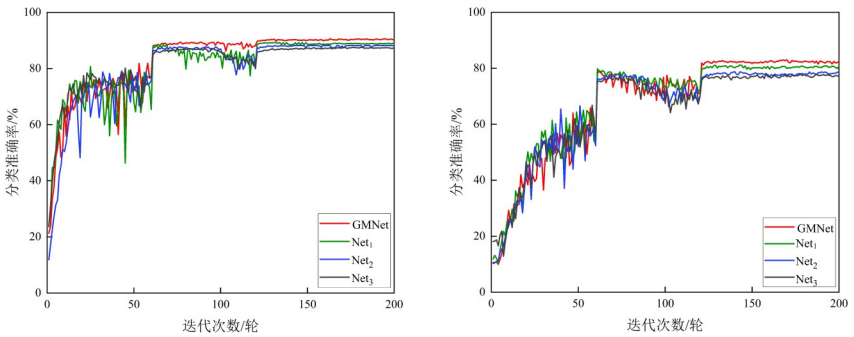
网络	分类准确率				
	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
GMNet	**97.05**	**83.68**	**97.68**	**90.60**	**83.05**
Net_1	96.18	81.82	96.92	89.21	80.81
Net_2	94.48	79.42	95.82	88.17	78.67
Net_3	90.86	77.34	94.80	87.66	77.96
Net_4	88.09	71.41	91.51		

使用的LGAM,Net_2表示在GMNet中去掉IGC,Net_3表示在GMNet中不将第1层普通卷积改为门控卷积,Net_4表示在GMNet中不修改第1层普通卷积核尺寸.各网络具体消融实验结果如表6所示.表中黑体数字表示最优值.

由图17可知,分类准确率由高到低依次为GMNet、Net_1、Net_2、Net_3、Net_4,从而验证在残差分支中使用



(a) CIFAR-10　(b) CIFAR-100　(c) SVHN



(d) Imagenette　(e) Imagewoof

图 17　各网络在 5 个数据集上的分类准确率

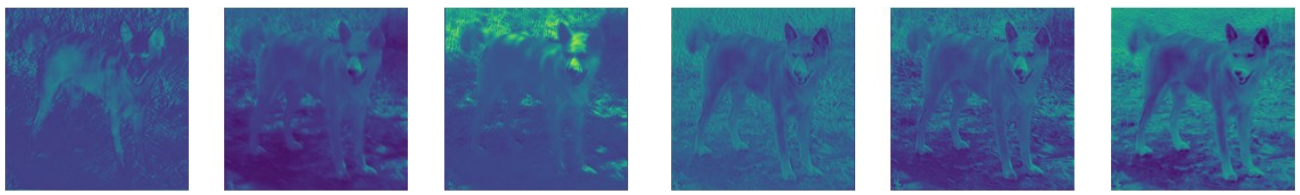
LGAM,可以有效实现特征的选择性增强和选择性融合,提高关键区域特征的贡献度,减少关键特征在网络信息传递中流失. 利用 IGC 可以强化浅层特征的同时提取更具区分度的特征,提高特征的非线性表达能力,实现特征的还原和增强. 将第 1 层普通卷积改为门控卷积能够有效提高网络对原始图像关键特征的提取能力. 修改第 1 层卷积核尺寸,能够有效减少原始图像特征损失,提取更加细微的关键特征.

实验表明,将首层普通卷积改为门控卷积并修改卷积核尺寸,利用 IGC 以及在残差分支中使用 LGAM 时,网络性能得到一定的提升.

4.5　可视化分析

4.5.1　IGC 可视化分析

为进一步观察 IGC 模块保留图像的细节和特征的效果,分别使用大小为 1 和 2 的 scale_factor 对图像进行特征提取. 其特征图如图 18 所示.



(a) scale_factor 为 1 时输出的特征图　(b) scale_factor 为 2 时输出的特征图

图 18　两种 scale_factor 输出的特征图

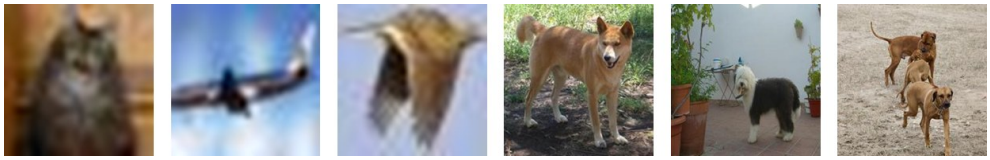
由图 18 可以观察到,利用 Lanczos 插值将特征图放大,再结合门控卷积进行特征提取,能够获得内容更加细致丰富的特征信息,避免细节丢失,提取更具区分度的特征.

4.5.2　LGAM 可视化分析

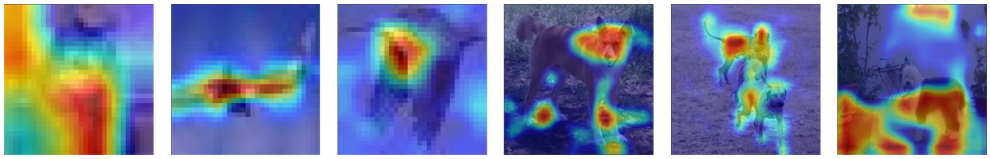
为进一步观察 LGAM 本文注意力机制的效果,选择如下注意力机制进行对比分析:FCA(Frequency Channel Attention)[23],CA(Coordinate Attention)[24],SA(Split Attention)[25],VA(Visual Attention)[20].

各注意力机制在 CIFAR-10、Imagewoof 数据集上的可视化图像如图 19 所示,其中前 3 张图片来自 32×32 数据集 CIFAR-10,后三张图片来自 224×224 数据集 Imagewoof.

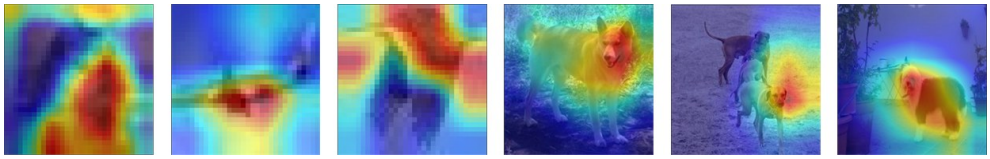
由图 19 可以观察到在 CIFAR-10 数据集上,虽然 LGAM 的关注区域小,但其注意力分布全部集中在图像



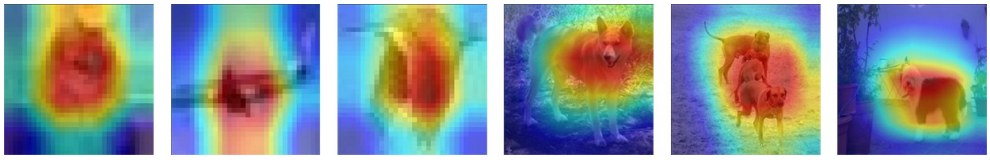
(*a*) 原始图像



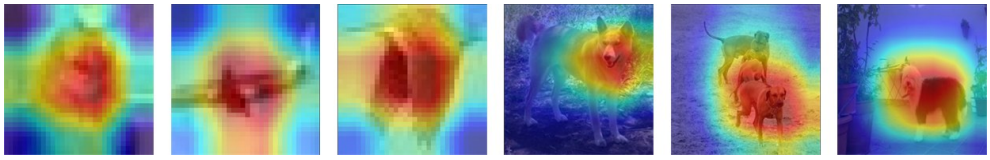
(*b*) FCA



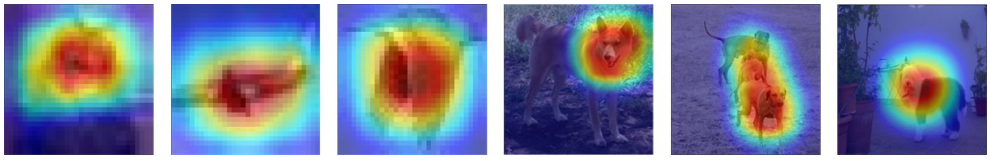
(*c*) CA



(*d*) SA



(*e*) VA



(*f*) LGAM

图 19 不同注意力机制的热力图对比

关键特征上;同样在Imagewoof数据集上,LGAM能够重点关注狗脸部的细节特征.相比其他注意力机制,LGAM能够提高关键区域的关注度,有效提取关键特征,从而提高网络的分类精度.

5 结论

针对图像分类时无法高效利用关键特征,特征提取不充分,易发生特征丢失等问题,本文提出门控机制的图像分类网络(GMNet).GMNet引入门控机制,设计了插值门控卷积模块IGC和大核门控注意力机制模块LGAM,增强特征的非线性表达能力,提取更具表达能力和区分度的重点特征,减少关键特征流失,提高关键区域特征的贡献度,有效地提高神经网络图像分类能力.

参考文献

[1] 刘颖, 庞羽良, 张伟东, 等. 基于主动学习的图像分类技术: 现状与未来[J]. 电子学报, 2023, 51(10): 2960-2984.
LIU Y, PANG Y L, ZHANG W D, et al. Active learning-based image classification technology: Status and future[J]. Acta Electronica Sinica, 2023, 51(10): 2960-2984. (in Chinese)

[2] 许新征, 李彬. 基于特征膨胀卷积模块的轻量化技术研

究[J]. 电子学报, 2023, 51(2): 355-364.
XU X Z, LI B. Research of lightweight convolution neural network based on feature expansion convolution[J]. Acta Electronica Sinica, 2023, 51(2): 355-364. (in Chinese)

[3] 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.

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

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2024-01-05]. https://arxiv.org/pdf/1409.1556.pdf.

[6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.

[7] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [EB/OL]. (2016-05-23) [2024-01-05]. https://arxiv.org/pdf/1605.07146.pdf.

[8] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 4700-4708.

[9] ABDI M, NAHAVANDI S. Multi-residual networks: Improving the speed and accuracy of residual networks[EB/OL]. (2016-09-19)[2024-01-05]. https://arxiv.org/pdf/1609.05672.pdf.

[10] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. San Diego: PMLR, 2019: 6105-6114.

[11] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1580-1589.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

[13] HU X F, ZHANG Z H, JIANG Z Y, et al. SPAN: Spatial pyramid attention network for image manipulation localization[C]//Computer Vision — ECCV 2020. Cham: Springer International Publishing, 2020: 312-328.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 6000-6010.

[15] CHOROMANSKI K, LIKHOSHERSTOV V, DOHAN D, et al. Rethinking attention with performers[EB/OL]. (2020-09-30)[2024-01-05]. https://arxiv.org/pdf/2009.14794.pdf.

[16] LAN H, WANG X H, SHEN H, et al. Couplformer: Rethinking vision transformer with coupling attention[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2023: 6475-6484.

[17] KONSTANTINIDIS D, PAPASTRATIS I, DIMITROPOULOS K, et al. Multi-manifold attention for vision transformers[EB/OL]. (2022-07-18) [2024-01-05]. https://arxiv.org/pdf/2207.08569.pdf.

[18] YU J H, LIN Z, YANG J M, et al. Free-form image inpainting with gated convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 4471-4480.

[19] 郭莹, 李伦, 王鹏. 基于Lanczos核的实时图像插值算法[J]. 通信学报, 2017, 38(6): 142-147.
GUO Y, LI L, WANG P. Real time interpolation algorithm based on Lanczos kernel[J]. Journal on Communications, 2017, 38(6): 142-147. (in Chinese)

[20] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733-752.

[21] HENDRYCKS D, GIMPEL K. Gaussian error linear units (GELUs)[EB/OL]. (2016-06-27)[2024-01-05]. https://arxiv.org/pdf/1606.08415.pdf.

[22] 姜文涛, 赵琳琳, 涂潮. 双分支多注意力机制的锐度感知分类网络[J]. 模式识别与人工智能, 2023, 36(3): 252-267.
JIANG W T, ZHAO L L, TU C. Double-branch multi-attention mechanism based sharpness-aware classification network[J]. Pattern Recognition and Artificial Intelligence, 2023, 36(3): 252-267. (in Chinese)

[23] QIN Z Q, ZHANG P Y, WU F, et al. FcaNet: Frequency channel attention networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 783-792.

[24] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR). Piscataway: IEEE, 2021: 13713-13722.
[25] ZHANG H, WU C R, ZHANG Z Y, et al. ResNeSt: Split-attention networks[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2022: 2736-2746.

作者简介

姜文涛 男,1986年10月出生于辽宁省大连市. 现为辽宁工程技术大学软件学院副教授. 主要研究方向为图像与视觉信息计算、模式识别与人工智能.
E-mail: lntuwulue@163.com

高 原 男,2000年4月出生于辽宁省沈阳市. 现为辽宁工程技术大学软件学院在读硕士. 主要研究方向为图像与视觉信息计算、模式识别与人工智能.
E-mail: 1422822508@qq.com

袁 姮 女,1988年2月出生于湖北省黄冈市. 现为辽宁工程技术大学软件学院副教授. 主要研究方向为图像与视觉信息计算、模式识别与人工智能.
E-mail: lntuyuanheng@163.com

刘万军 男,1959年10月出生于辽宁省北镇市. 现为辽宁工程技术大学软件学院教授. 主要研究方向为软件工程理论、图像与视觉信息计算、模式识别与人工智能.
E-mail: liuwanjun39@163.com