

背景感知机制的图像分类网络

袁 姮¹, 冉 超^{1*}, 张晟翀²

(1. 辽宁工程技术大学软件学院, 辽宁葫芦岛 125105; 2. 光电信息控制和安全技术重点实验室, 天津 300308)

摘要: 针对图像分类方法缺乏对复杂场景的有效理解, 导致模型对关键特征的捕捉能力受限, 进而影响分类精度等问题, 本文提出背景感知机制的图像分类网络(image classification Network of Background Perception Mechanism, BPMNet). 首先, 提出背景感知(Background Perception, BP)模块, 通过双分支结构分别处理前景与背景信息, 动态调整输入特征的贡献度, 强化背景信息对前景特征的上下文支撑作用, 以增强模型对背景信息的感知能力; 然后, 结合BP模块, 设计了背景感知注意力(Background Perceptual Attention, BPA)模块, 考虑局部特征信息、长程依赖关系的同时关注图像前景与背景之间的关系, 动态调控背景信息对主体目标特征的影响程度, 增强关键目标特征的判别性和定位能力. 最后, 将背景感知模块与背景感知注意力模块嵌入残差块中, 实现从浅层细节到深层语义的特征传递, 结合局部细节与全局语义, 增强复杂场景下前景目标的特征表示能力. 在图像数据集CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof上, BPMNet分别达到了96.95%、80.85%、97.68%、90.10%、81.70%的分类准确率, 与其他主流网络相比分别平均提高了2.39%、3.17%、2.36%、2.30%、2.67%. 与当前先进的网络模型相比, 本文方法能够增强模型对复杂场景的理解, 提高关键区域表达能力, 从而更有效地提取关键特征, 进一步提高模型的鲁棒性和泛化能力.

关键词: 图像分类; 背景感知机制; 背景感知注意力; 局部特征; 神经网络

基金项目: 国家自然科学基金(No.61601213); 国防预研基金(No.172068); 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(2025)08-2779-15

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20250028

Image Classification Network of Background Perception Mechanism

YUAN Heng¹, RAN Chao^{1*}, ZHANG Sheng-chong²

(1. School of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;

2. Key Laboratory of Optoelectronic Information Control and Security Technology, Tianjin 300308, China)

Abstract: In view of the lack of effective understanding of complex scenes in image classification methods, which leads to the limited ability of models to capture key features and thus affects the classification accuracy, this paper proposes an image classification network of background perception mechanism (BPMNet). Firstly, the background perception (BP) module is proposed. Through a dual-branch structure, the foreground and background information are processed respectively, the contribution degree of the input features is dynamically adjusted, and the context support role of the background information on the foreground features is strengthened to enhance the model's perception ability of background information. Then, combined with the BP module, the background perception attention (BPA) module is designed. While considering the local feature information and long-range dependency relationship, it also pays attention to the relationship between the foreground and background of the image, and dynamically regulates the influence degree of the background information on the features of the subject target and enhances the discriminability and positioning ability of key target features. Finally, the background perception module and the background perception attention module are embedded in the residual block to achieve feature transfer from shallow details to deep semantics, and the feature representation ability of foreground targets in complex scenes is enhanced by combining local details and global semantics. Compared with other mainstream networks, the classification accuracy of BPMNet achieved on the image data sets such as CIFAR-10, CIFAR-100, SVHN, Imagenette and Imagewoof, are 96.95%, 80.85%, 97.68%, 90.10% and 81.70%, respectively, which increased by 2.39%, 3.17%, 2.36%, 2.30% and 2.67% on average. Compared with the current advanced network models, the proposed method can enhance the model's understanding of complex scenes, improve the ability to express key regions, extract key features more effectively,

and further improve the robustness and generalization ability of the model.

Key words: image classification; background perception mechanism; background perception attention; local feature; neural network

Foundation Item(s): National Natural Science Foundation of China (No.61601213); National Defense Preliminary Research Fund (No.172068); 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的核心任务之一,其目标是通过输入图像进行分析与识别,将图像归类到预定义的类别中^[1]. 由于图像分类面临着图像多样、背景复杂等挑战,容易导致特征提取不足,进而影响分类的准确性. 如何有效应对这些问题并提高图像分类精度,一直是该领域的研究重点和难点^[2].

基于深度学习的图像分类任务通常使用卷积神经网络(Convolutional Neural Network, CNN)来实现高效的特征提取^[3]. 例如,LeCun等^[4]提出的LeNet,通过卷积操作来提取局部特征,并通过池化操作进行特征降维,从而有效地减小计算量并提高模型的泛化能力. 基于LeNet, Krizhevsky等^[5]提出AlexNet,通过增加网络层数和引入非线性激活函数ReLU,并采用Dropout进行正则化,显著降低了过拟合风险. Simonyan等^[6]提出了VGGNet,通过堆叠多个3×3的小卷积核来增加网络的深度,同时保持卷积核大小一致,使网络结构更加统一,提升了模型的表现力. Szegedy等^[7]提出GoogLeNet,在VGGNet的基础上引入Inception模块,通过并行应用多个大小不同的卷积核,实现多尺度特征提取.

尽管上述网络通过增加网络深度有效提高了模型分类准确率,但在反向传播过程中,梯度可能会随层数的增加而迅速衰减或放大,进而导致梯度消失或梯度爆炸等问题. 为了解决这些问题,He等^[8]提出ResNet,通过引入残差学习的概念来简化深层网络的训练过程,通过跳跃连接让每一层学习输入的残差,缓解退化问题,解决网络训练时梯度消失和梯度爆炸等问题. 基于ResNet, Zagoruyko等^[9]提出了Wide ResNet,通过增加每层的特征通道数来提升网络的表达能力,宽度的增加使得网络能够学习到更丰富的特征表示,从而在图像分类等任务中取得更好的性能. Han等^[10]提出GhostNet,通过设计Ghost模块生成特征图的一部分来降低计算成本,仅需少量的计算即可生成有代表性的高效特征,提升模型的效率和性能,适用于资源受限的应用场景. Huang等^[11]提出DenseNet,通过连接每一层与前面所有层,增加特征的复用性和传播能力,提升了信息流动及反向传播速率,提供了更紧凑的特征传递方法. 结合ResNet和DenseNet, Chen等^[12]提出DPN(Dual Path Networks),使用双路径来增强特征的复用性和流动性,在提取特征的同时考虑长短期信息,使

得模型能够提取更加有效的特征.

上述卷积神经网络在处理图像分类任务时具有较好的效果,但均存在关键特征提取不足以及无法有效聚焦重要特征等问题. 针对该问题, Hu等^[13]提出基于通道注意力的SENet,通过全局平均池化提取通道信息,再利用两个全连接层生成权重,动态调整通道特征的表达. 受SENet的启发, Wang等^[14]提出ECA-Net,通过高效的一维卷积进行通道特征提取,实现了轻量化和高效的通道注意力计算. Qin等^[15]提出FcaNet,通过将全局平均池化扩展为频域特征分解的形式,同时引入多光谱分量来增强通道注意力的表达能力,提高模型在分类任务中的性能.

Transformer最初由Vaswani等^[16]在2017年提出,与传统注意力机制不同的是,Transformer通过自注意力机制实现了对长序列的建模,并且能够同时捕捉输入序列中各个元素之间的相互关系以及长距离依赖关系,克服传统注意力在处理长序列时的局限性. 在Transformer的基础上, Choromanski等^[17]提出Performer自注意力机制,通过引入正态化的核方法实现自注意力的高效近似,优化了模块的计算效率和可扩展性. Lan等^[18]提出了自注意力机制Couplformer,通过局部与全局信息的有机结合使得模型在提取关键视觉特征时,能够有效地捕捉到不同尺度的信息. 张峰等^[19]通过同时建模相邻图像块之间的多尺度特征表达,使模型学习到输出特征在不同维度下的语义信息.

上述基于注意力机制的网络能在一定程度上提高特征提取的有效性和整体性能,具有较好的图像分类能力. 但这些注意力机制在进行图像内容分析和特征提取过程中,往往只聚焦于图像的主体信息,忽略了背景信息,而图像中的主体信息与背景信息是相互依赖、互为支持的,忽视背景信息会导致模型在复杂场景下因上下文信息缺失,难以准确区分前景类别,进而限制其对关键区域的表达能力. 基于此,本文提出背景感知机制的图像分类网络(image classification Network of Background Perception Mechanism, BPMNet). BPMNet中的背景感知(Background Perception, BP)模块和背景感知注意力(Background Perceptual Attention, BPA)模块,同时处理输入图像的前景信息和背景信息,考虑背景信息对前景信息的支持和依赖作用,动态调整背景特征对主体目标特征的影响程度,以增强模型对关键目

标特征的感知和捕捉能力。

2 背景感知机制

在神经网络中,单个神经元通常接收多个输入信号,每个输入信号对应一个权重,表示该输入对神经元输出的重要性,同时神经元还包含一个偏置,用于调整神经网络的响应速度和输出幅度,使模型具有更好的拟合能力,其计算过程如式(1)所示:

$$z = \sum_i w_i x_i + b \quad (1)$$

其中, x_i 表示单个输入信号; w_i 对应的权重为 w_i ; b 为偏置项, z 为所有输入信号的加权和,表示神经元总的输入。

将神经元总的输入信号 z 经过非线性变换后,得到神经元的输出,其计算过程如式(2)所示:

$$y = f(z) \quad (2)$$

其中, y 表示信号输出, $f(\cdot)$ 表示非线性变换的激活函数。

在神经元中,输入信号 z 的符号为负时通常表示输入图像的背景信息,如图1中的草地、树叶、树枝等元素,部分负值可能对应前景中低响应区域,如小狮子身上的均匀毛发;而输入信号 z 的符号为正时则通常表示输入图像的前景信息,如图1中正值的输入信号往往对应小狮子的耳朵、鼻子、胡须、眼睛和嘴巴等主体特征。



图1 输入信号为正值与负值对应特征图对比

从图像内容分析的角度来看,通常会认为图像的主体信息是主要关注的对象,而背景信息则被视为无关且应当舍弃的部分。然而实际上,某些背景信息可能暗示图像主体的存在,并在一定程度上为主体信息提供支持;此外,背景与主体之间的边缘区域往往包含有价值的特征信息,两者之间不仅存在相互依赖关系,还在一定程度上相互支持。例如,城市街道的背景暗示了行人、自行车和汽车的高概率出现,海洋背景则会暗示海洋生物的高概率出现。这种对背景的感知能力能够

增强模型对复杂场景的理解,基于此,本文将背景信息与前景信息之间的相互支持与依赖关系称为背景感知机制。

2.1 背景感知模块

传统卷积神经网络对前景信息和背景信息的无差别处理,使得模型无法有效利用背景中的有用信息,且易受无效背景信息影响,从而限制模型对复杂场景的理解。此外,ReLU激活函数对输入信号的非线性变换会将所有负值输入信号直接置零,该特性会导致有价值的背景信息被直接丢弃,进而造成特征空间的上下文信息损失。

为了关注背景信息对目标主体的支持作用,增强模型对背景信息的感知能力,并提高其对复杂场景的理解,本文基于背景感知机制,提出背景感知(BP)模块。该模块采用双分支结构,分别提取图像的前景与背景特征,并通过可学习比例系数 α 自适应调整前景与背景信息的贡献度,从而增强模型对复杂场景的理解。

如图2所示,BP模块设计了两个独立的分支:前景分支 M 和背景分支 N 。其中,前景分支专注于提取显著的前景信息,而背景分支则适当地保留背景信息,从而增强模型对背景信息的感知能力,提高对目标主体和背景的综合理解。

在前景分支中,BP模块首先对输入特征图 $\mathbf{x} \in \mathbf{R}^{C \times H \times W}$ 逐元素应用ReLU激活函数筛选正值,即

$$M(\mathbf{x}) = \text{ReLU}(\mathbf{x}) = \begin{cases} 0, & \mathbf{x}_{c \times h \times w} < 0 \\ \mathbf{x}_{c \times h \times w}, & \mathbf{x}_{c \times h \times w} > 0 \end{cases} \quad (3)$$

其中, $\mathbf{x}_{c \times h \times w}$ 为输入特征图中第 c 个通道在高度 h 和宽度 w 处的值。随后,通过TopK操作选择特征通道中前 K 个最大值的特征通道,表示为

$$\mathbf{X}_1 = \text{TopK}(M(\mathbf{x}), K)$$

其中, $K = \alpha \times C$, α 为比例系数, C 为通道总数。

背景分支专注于处理输入特征图中的负值信息。通过TopK操作选择特征通道中前 L 个具有显著背景特征的通道,表示为

$$\mathbf{X}_2 = \text{TopK}(N(\mathbf{x}), L)$$

其中, $N(\mathbf{x}) = \begin{cases} \mathbf{x}_{c \times h \times w}, & \mathbf{x}_{c \times h \times w} \leq 0 \\ 0, & \mathbf{x}_{c \times h \times w} > 0 \end{cases}$,用于筛选负值信息, $L =$

$(1 - \alpha) \times C$, $1 - \alpha$ 为比例系数, C 为通道总数。

α 初始值根据数据集特性设定,以平衡不同场景下背景信息的利用强度。在训练过程中, α 通过反向传播算法动态优化,其更新公式为

$$\alpha = \alpha - \eta \frac{\partial G}{\partial \alpha}$$

其中, η 为学习率, G 为损失函数。

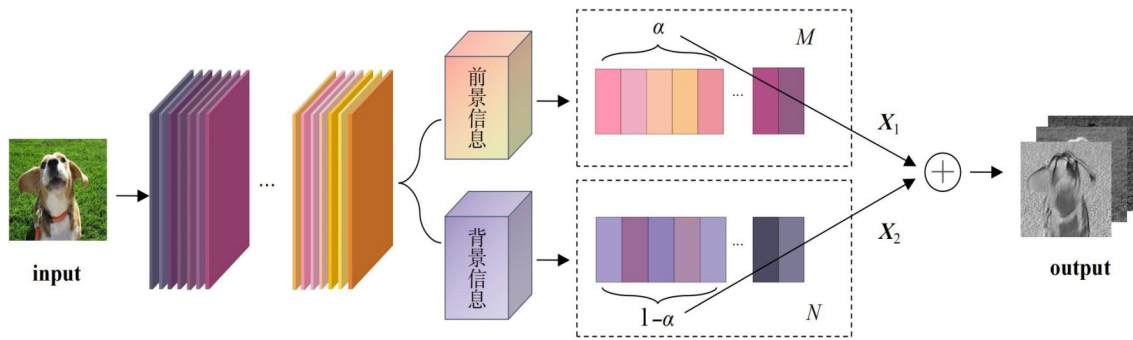


图2 BP模块结构图

最后,将两分支输出的特征进行相加,生成包含前景与背景综合信息的特征图,其过程如下所示:

$$\text{output} = X_1 + X_2 = \text{TopK}(M(x), K) + \text{TopK}(N(x), L)$$

为实现特征的动态选择,本模块引入比例系数 α 作为可学习参数.相比传统的线性修正单元,模块中参数 α 的动态调节,能够有效控制对背景信息的关注程度: α 值越小,保留的负值特征比例越大,模块对背景信息的关注度就越高;反之, α 值越大,则更强调前景特征的重要性.过度关注背景信息可能会削弱模型对主体特征的学习能力,而过于忽视背景信息则会限制其在图像分析中的作用.为了在前景特征提取与背景信息利用之间实现平衡, α 初始取值的选择尤为关键.因此, α 初始取值的优化需结合具体应用场景进行调整,针对不同数据集,最优初始取值可能有所差异,具体选择将在第四章实验部分进行详细说明.

通过这种“前景主导,背景支持”的特征表达模式,模型能够有效捕捉复杂场景中前景与背景的上下文关联.在减少无关信息干扰的同时,充分利用有效背景信

息来丰富特征表达,使得模型在面对不同场景时更具适应性,确保对目标特征的有效关注,显著提升模型对关键特征的提取能力和分类性能.

2.2 背景感知注意力模块

CNN虽然能够自动提取图像特征,但在处理复杂场景和捕捉局部细节特征时,传统卷积层往往忽视背景与主体之间的依赖关系,导致背景信息无法得到有效利用,进而影响模型对复杂场景的理解和关键区域的表达能力.为了解决这一问题,结合BP模块,本文提出背景感知注意力(BPA)模块,综合考虑上下文信息和局部特征之间的交互和依赖关系,建立更准确的前景-背景关联,并根据其贡献度赋予权重,确定关键特征区域,增强模型在复杂场景和局部视觉任务中的表现.

BPA结构如图3所示,输入特征图 $M \in \mathbf{R}^{C \times H \times W}$,其中, C 为特征图的通道数, $H \times W$ 为特征图的空间维度.BPA同时推导两个分支的二维特征矩阵,对输入图像提取不同空间维度上的特征.

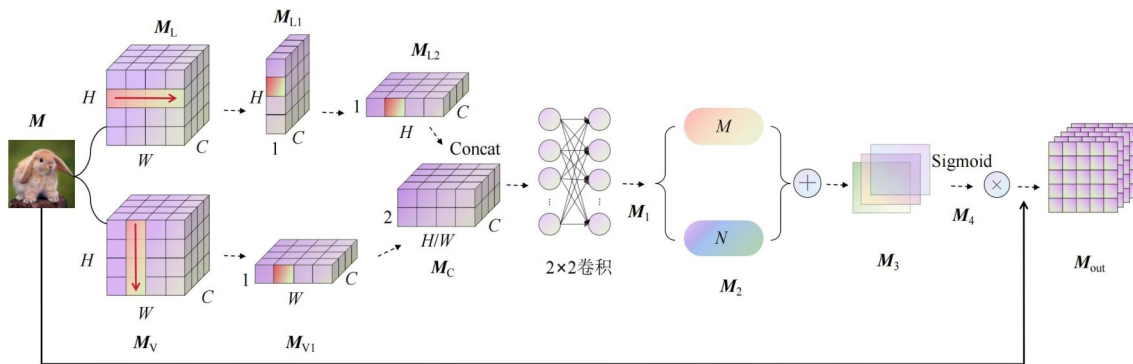


图3 BPA模块结构图

首先,给定输入特征图 M ,使用池化核 $(H, 1)$ 和 $(1, W)$ 分别沿水平方向和垂直方向提取每个通道的特征信息,如图3中的 $M_L \in \mathbf{R}^{C \times H \times W}$ 表示水平方向计算特征, $M_V \in \mathbf{R}^{C \times H \times W}$ 表示垂直方向计算特征,然后获得两个独立的特征图 $M_{L1} \in \mathbf{R}^{C \times H \times 1}$ 和 $M_{V1} \in \mathbf{R}^{C \times 1 \times W}$.每个特征图的全程依赖沿着一个空间方向,因此,可以在所生

成的注意力图中保留相应的空间方向信息.

在水平方向使用平均池化操作如图4(a)所示, K 表示池化窗口大小, S 表示步长,其中宽度被压缩为1.其公式可以表示为

$$M_c^h(h) = \frac{1}{W} \sum_{i=1}^W M(c, h, w) \quad (4)$$

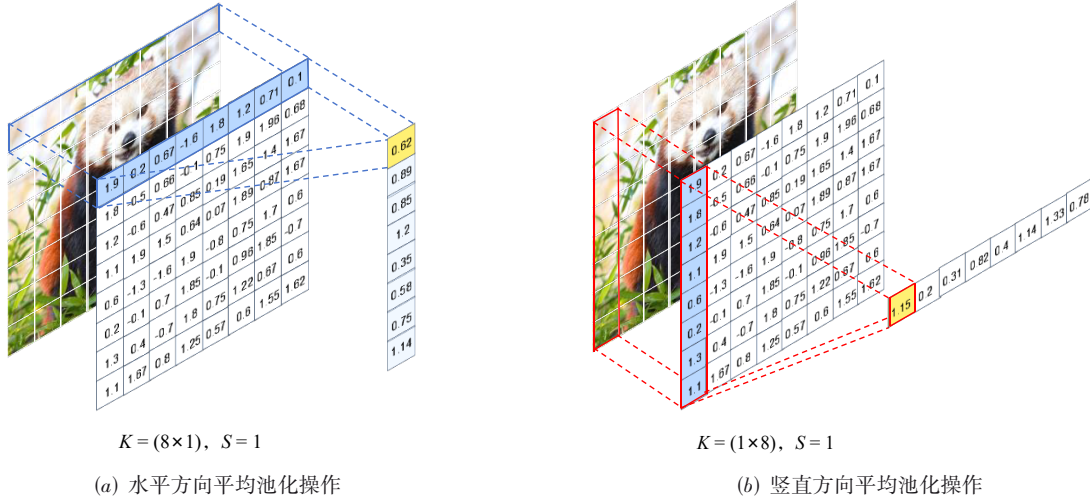


图4 平均池化操作示意图

其中, $M_c^h(h)$ 为平均池化后第 c 个通道在高度 h 处的输出, $M(c, h, w)$ 为输入特征图中第 c 个通道在高度 h 和宽度 w 处的值, W 为输入特征图的宽度. 类似地, 在竖直方向上应用平均池化操作如图 4(b) 所示, K 表示池化窗口大小, S 表示步长, 其中高度被压缩为 1. 其公式可以表示为

$$M_c^w(w) = \frac{1}{H} \sum_{h=1}^H M(c, h, w) \quad (5)$$

其中, $M_c^w(w)$ 为平均池化后第 c 个通道在宽度 w 处的输出, $M(c, h, w)$ 为输入特征图中第 c 个通道在高度 h 和宽度 w 处的值, H 为输入特征图的高度.

上述两种变换分别沿着高和宽两个维度聚合特征, 生成一对具有不同方向特征信息的特征图, 这不仅捕获跨通道信息, 还能提取来自不同空间维度上的细节特征, 从而互补地应用于输入特征图, 提升目标对象的表达能力. 通过对不同空间维度的特征进行建模, 能够有效提升模型的表达效率与计算性能.

接着, M_{L1} 经过翻转操作得到与 M_{V1} 相同维度的 M_{L2} , 这个过程可以表示为

$$M_{L2} = \text{Transpose}(M_{L1}, 2, 3)$$

其中, $\text{Transpose}(M_{L1}, 2, 3)$ 表示将 M_{L1} 的第 2 维和第 3 维进行交换.

通过将不同空间维度上的信息嵌入到通道注意力中, 使感受野增大的同时提升模型的计算效率, 弥补了通道注意力只聚焦于通道层面的局限性.

将两个特征图在高度上进行拼接得到 $M_C \in \mathbf{R}^{C \times 2 \times H/W}$, 其中 H/W 表示高度或宽度, 其操作可以表示为

$$M_C = \text{Concat}(M_{L2}, M_{V1})$$

将高度方向上的特征和宽度方向上的特征拼接在一起时, 实际上是在进行特征融合. 这种融合使得网络

得到来自两个不同空间维度上的信息, 从而学习到更加丰富的特征表示. 然后, 将所得到的 M_C , 通过二维卷积操作进行局部特征提取, 可表示为

$$M_1 = \text{Conv2d}(M_C)$$

其中, $M_1 \in \mathbf{R}^{C \times 1 \times H/W}$ 为经过卷积之后的特征图, $\text{Conv2d}(\cdot)$ 表示二维卷积操作. 与普通 2D 卷积操作有所不同, 由于拼接后的特征图高度仅为 2, 卷积核只在一个维度上滑动, 不仅能够在感受野大的情况下捕捉到更多局部细节特征, 还能有效提取来自不同维度之间的交互特征信息, 使得模型能够在更大范围内整合上下文信息, 从而提升模型的表现力.

为了更好地平衡前景信息和背景信息的贡献, 增强上下文之间的联系, 将 BP 模块直接嵌入到 2×2 卷积操作之后. 经过 2×2 卷积操作后的特征图 M_1 包含了整合的前景信息和背景信息, 接下来通过 BP 模块将 M_1 分为前景和背景两个分支进行分别处理, 使得在生成通道特征权重时有选择地增强重要的前景特征, 同时动态地调整背景信息对权重生成的影响, 提升模型对复杂背景的鲁棒性和对局部特征的捕捉能力. 这个过程可表示为

$$M_3 = X_1 + X_2$$

最后通过 Sigmoid 激活函数得到通道特征权重矩阵 M_4 , 可表示为

$$M_4 = \text{Sigmoid}(M_3)$$

Sigmoid 公式如式 (6) 所示

$$S(x) = \frac{1}{1 + \exp(-x)} \quad (6)$$

该函数的输出值范围在 0 到 1 之间, 用于评估不同通道的相对重要性. 较高的概率值表明该通道对特征表达的贡献更大, 而较低的概率则意味着该通道的重要性相对较低, 有助于更好地辨识通道中的关键部分.

最后,将通道特征权重矩阵 M_4 与输入图像 M 进行逐元素相乘,使之能够充分考虑上下文信息,从而在决策时综合更多相关因素并有效提取关键特征,得到背景感知注意力模块的输出图像 M_{out} ,可以表示为

$$M_{out} = M \times M_4 = M \times \text{Sigmoid}(M(M_{L1}) + N(M_{L1}))$$

其中,

$$M_{L1} = \text{Conv 2d}(\text{Concat}(\text{Transpose}(M_{L1}, 2, 3), M_{L1}))$$

3 背景感知机制的图像分类网络

针对图像分类方法缺乏对复杂场景的有效理解,导致模型对关键特征的捕捉能力受限,进而影响分类精度等问题,本文提出背景感知机制的图像分类网络(BPMNet).

3.1 浅层特征提取模块

BPMNet以ResNet-34为基线模型.在ResNet-34残差网络中,初始特征提取通常使用 7×7 的卷积层和最大池化层完成. 7×7 卷积核虽然具备较大的感受野,可以捕获全局特征,但其对所有像素的无差别处理可能引入大量无效特征,影响分类效果.因此,需要在保留关键全局信息的同时,减少无效特征干扰,确保模型对关键信息的有效捕获.

本文将BP模块置于首层卷积之后,在初始特征提取阶段对前景和背景信息进行分别处理.低层特征包含大量背景细节(如纹理、光照),这些细节中既存在有助于分类的关键线索,也包含会干扰判断的冗余信息,通过前景-背景分离可初步筛选关键特征,减少低层边缘与背景无关信息的干扰.前景分支聚焦于显著的目标特征提取,而背景分支通过动态调整有效保留与主体相关的上下文信息,从而在减少无效特征干扰的同时筛选出图像的关键信息.

浅层特征提取模块结构如图5所示,将普通卷积层输出的特征图经过BP模块的前景与背景分支进行分别处理,得到显著的前景信息与更具表达性的背景信息.最后,将两分支输出的特征进行相加,生成包含前景与背景综合信息的特征图,为后续残差块提供更加

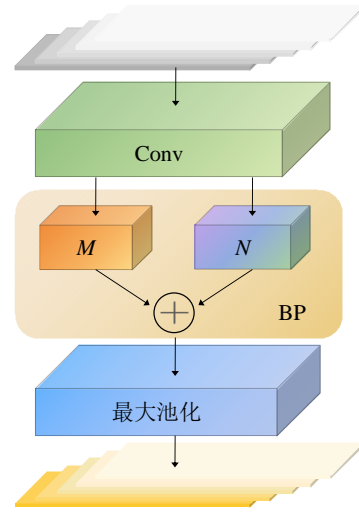


图5 浅层特征提取模块示意图

平衡和丰富的输入特征.

3.2 融合BP和BPA的残差模块

ResNet的残差块通过跳跃连接将输入直接与卷积层的输出相加,从而在保留原始特征的基础上学习残差部分,用于解决深层网络中的梯度消失和退化问题.其结构如图6(a)所示.

本文将BP模块和BPA模块嵌入残差块的主分支中,其结构如图6(b)所示,该残差块主干包含两个 3×3 卷积用于提取局部特征,两个BP模块分别嵌入两个卷积层之后,并且通过双分支结构对前景与背景特征进行分别处理,针对高阶语义特征,能够优化前景-背景的语义关联,抑制无关背景对分类决策的干扰,进一步提取关键特征.然后进入BPA模块,BPA模块能够聚焦关键区域,在保留高效特征提取能力的同时,又通过背景感知机制增强模型对复杂场景的理解,从而提升重点区域的权重.

残差模块内部卷积操作后均使用BN层以稳定梯度,提升网络训练效果,并利用残差连接将输入特征与处理后的特征进行相加,既保留了原始特征,又增强了高阶语义信息表达能力,从而提升模型分类精度和鲁棒性.

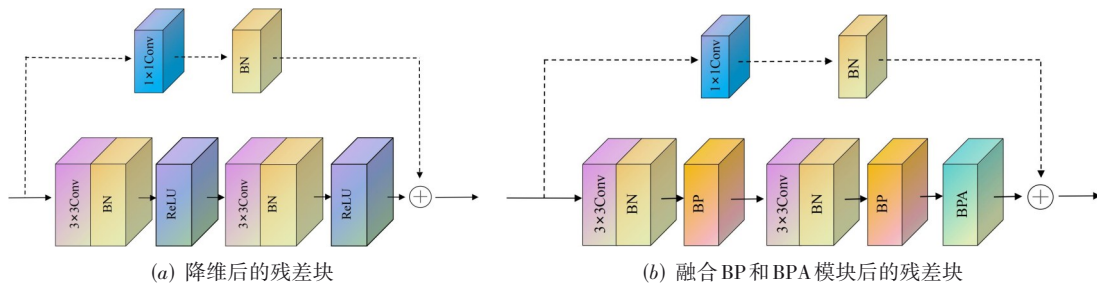


图6 两种残差块

ResNet的残差映射关系如式(7)所示:

$$L(x) = F(x) + x \quad (7)$$

其中, $L(x)$ 为待拟合对象, $F(x)$ 为卷积层学习的残差映射, x 为输入特征图.

融合 BP 模块和 BPA 模块的残差映射关系如式(8)所示:

$$L(x) = H(x) + x \quad (8)$$

其中, $H(x)$ 表示经过 BP 模块和 BPA 模块后学习的残差映射.

通过融合 BP 模块和 BPA 模块优化残差映射, 构建从浅层到深层的背景语义传递路径, 促进局部细节与全局上下文信息的协同表达, 保留关键前景-背景交互特征, 使模型能够更高效地学习输入数据的关联信息, 提高模型对复杂场景的关键特征提取能力.

3.3 BPMNet 总体结构

BPMNet 主要包含三个模块, 分别为背景感知模块, 融合背景感知模块的背景感知注意力模块, 融合背景感知模块和背景感知注意力的残差模块. 网络整体结构如图 7 所示.

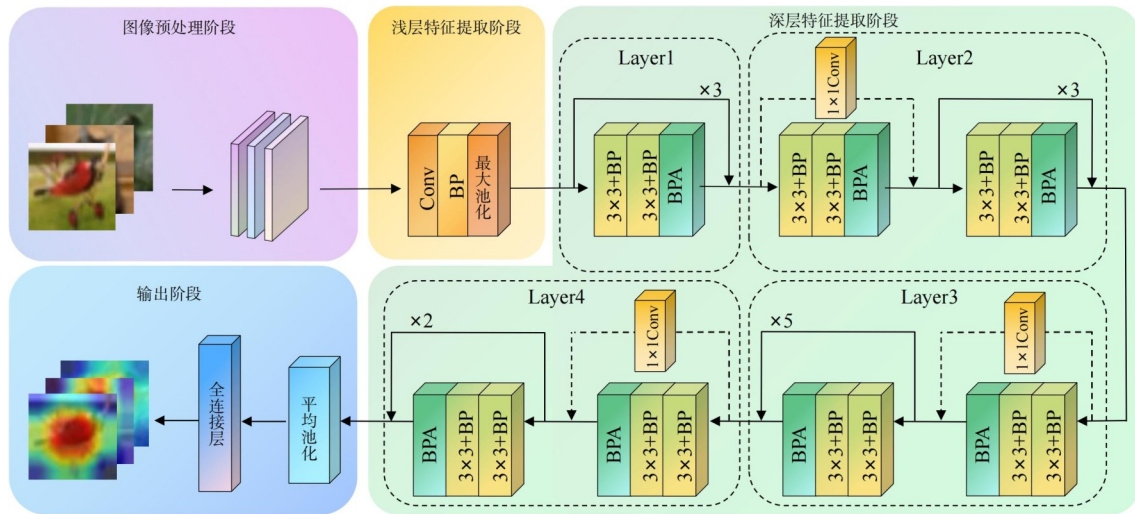


图 7 BPMNet 总体结构

整个网络结构分为以下四个阶段.

(1) 图像预处理阶段. 对输入的图像进行数据增强处理, 包括随机裁剪、随机翻转以及 Cutout 操作, 通过增加数据多样性来提升模型在不同场景下的鲁棒性和泛化能力.

(2) 浅层特征提取阶段. 经过预处理后的图像输入首层卷积层以及背景感知模块, 增强模型对背景感知能力, 为后续阶段提供更具判别性的特征表示.

(3) 深层特征提取阶段. 将输出的特征图输入融合背景感知模块和背景感知注意力的残差模块, 进一步提取图像中的深层次特征, 增强模型对细节和抽象特征的理解能力.

(4) 输出阶段. 将提取到的深层次特征通过平均池化层进行降维, 再经过全连接层进行分类, 输出最终的分类结果.

4 实验与结果分析

4.1 实验环境

实验环境的搭建基于 Python 编程语言, 操作系统使用 Windows 11, 硬件环境包括 NVIDIA Tesla P100 显卡、60 GB 内存. 选择 5 个数据集作为实验数据集, 分别

为 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Imgewoof. CIFAR-10 是一个广泛应用于图像分类任务的数据集, 涵盖了 10 个类别, 如: 飞机、汽车、鸟等. 相比之下, CIFAR-100 扩展了类别数量, 共有 100 个类别, 每类样本较少, 任务更具挑战性. SVHN 是一个数字识别数据集, 来自街景门牌号的实际照片, 包含 10 类. Imagenette 是 ImageNet 的一个子集, 包含 10 个较易分类的类别, 如: 吉他、草莓等. Imgewoof 是 ImageNet 的另一个子集, 专注于犬类的 10 个类别. 由于类别间相似性较高, 任务更具挑战性. 各数据集具体信息如表 1 所示.

在实验过程中, 数据集批次大小为 128, 训练周期设置为 200 (Epochs), 采用 SGD 优化器, 初始学习率设

表 1 实验数据集

数据集	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
Imgewoof	224×224	10	9 025	3 929

置为0.1,权重衰减设为 5×10^{-4} .

本文实验均在基本的数据增强后的数据集上进行,采用图像分类常用指标验证集的分类准确率作为评价指标.

4.2 影响网络性能的因素

影响BMNet性能的主要因素有:BP模块的参数 α 的选择、BP嵌入BPA的数量及位置、BPA嵌入残差块主干的数量及位置、首层卷积核尺寸.

4.2.1 BP模块的参数 α 的选择对BPMNet性能的影响

为深入研究BP模块对BPMNet的性能影响,针对BP中的可学习比例系数 α 初始取值展开实验.对于背景信息来说,由于占据过低的比例系数(如接近0)可能会导致模型对背景信息过度忽视,从而降低准确率;占据过高的比例系数(接近1)可能会导致背景信息对模型的干扰,降低对目标信息的关注.

因此,在其他基本参数一致的情况下,针对CIFAR-10、CIFAR100、SVHN、Imagenette和Imagewoof数据集,采用 $(\alpha, 1-\alpha)$ 的8种初始值组合方式进行实验,分别为A(0.74, 0.26)、B(0.72, 0.28)、C(0.70, 0.30)、D(0.68, 0.32)、E(0.66, 0.34)、F(0.64, 0.36)、G(0.62, 0.38)、H(0.60, 0.40),其中 α 表示前景信息所占输出特征图的比例, $1-\alpha$ 表示背景信息所占输出特征图的比例.对于每组实验, α 的初始值被设定为固定比例,但在训练过程中, α 通过反向传播动态优化,以自适应调整前景与背景特征的贡献权重.通过这种方式,可验证 α 的可学习性如何在不同初始化条件下引导模型在不同数据集收敛至更优状态,实验结果如图8所示.

由图8可知,在CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof数据集上的分类准确率与 $(\alpha, 1-\alpha)$ 的取值有关.当 $(\alpha, 1-\alpha)$ 初始取值为(0.70, 0.30)时,在SVHN数据集上的准确率最高;初始取值为(0.66, 0.34)时,在CIFAR-10和CIFAR-100数据集上的准确率最高;初始取值为(0.64, 0.36)时,在Imagenette和Imagewoof数据集上的准确率最高.该实验验证了在不同数据集上, $(\alpha, 1-\alpha)$ 的初始取值对BPMNet的准确率有一定影响,使得模型在一定程度上感知背景信息,但不会过多地让无关信息影响结果.数据集SVHN、CIFAR-10和CIFAR-100、Imagenette和Imagewoof在背景信息复杂程度上逐渐递增,因此 α 初始取值是逐渐递减的,这同时也验证2.1小节中“当 α 的值越小时, $1-\alpha$ 越大,背景感知模块对图像背景的关注程度会越高”思想以及 α 的动态学习能力及其对复杂场景的适应性.

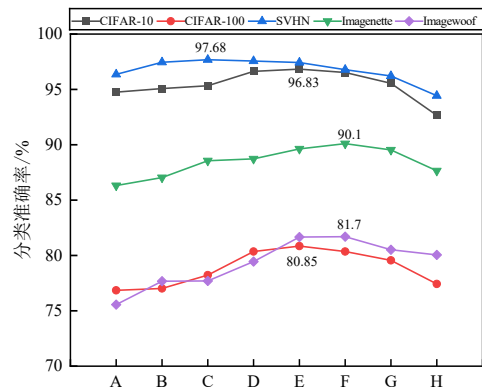


图8 不同比例组合方式对分类准确率的影响

4.2.2 BP嵌入BPA的数量及位置对BPMNet性能的影响

背景感知注意力模块通过特征聚合与卷积操作,并结合背景感知模块,在捕获长程依赖关系的同时,能够有效理解背景信息与目标主体之间的关联,提高关键区域表达能力,以更低的计算开销实现复杂场景下的高效特征提取.

为更好地分析BP嵌入BPA的数量和位置对分类准确率的影响,在其他参数保持一致的情况下设计了8种不同的BPA组合方式,如图9所示,其中A_C-Block表示双分支池化和拼接操作. A类为不在BPA内嵌入BP,其他类分别为加入不同数量和位置BP的BPA.其在5个数据集上对分类准确率的影响如图10所示.

实验结果表明,组合方式D在数据集CIFAR-10、CIFAR100、SVHN、Imagenette和Imagewoof上分别取得了96.95%、80.85%、97.68%、90.10%和81.70%的最高分类准确率.因此,在 2×2 卷积层后接入单层BP能够显著提高模型的分类能力,达到最优的性能表现.若继续增加BP的层数,随着层数的增多,模型的复杂度也会相应增加.这种过高的复杂度容易导致过拟合现象的出现,进而影响模型的泛化能力.因此,过度增加BP层数并不能带来持续的性能提升,反而可能引发泛化能力的下降.

4.2.3 BPA嵌入残差主干的数量及位置对BPMNet性能的影响

注意力机制在卷积神经网络中的嵌入位置和数量对分类任务的性能会产生不同程度的影响,嵌入位置和数量的变化能够影响特征的表达和信息流动,从而影响分类准确率.为了深入研究BPA在残差网络主干中嵌入不同位置和数量对分类准确率的影响,在不修改其他模块位置和参数的情况下,设计6种不同的嵌入方案,如图11所示.

这些方案通过调整BPA的嵌入位置(浅层、深层或多层组合),以测试注意力机制在不同层次上的效果.

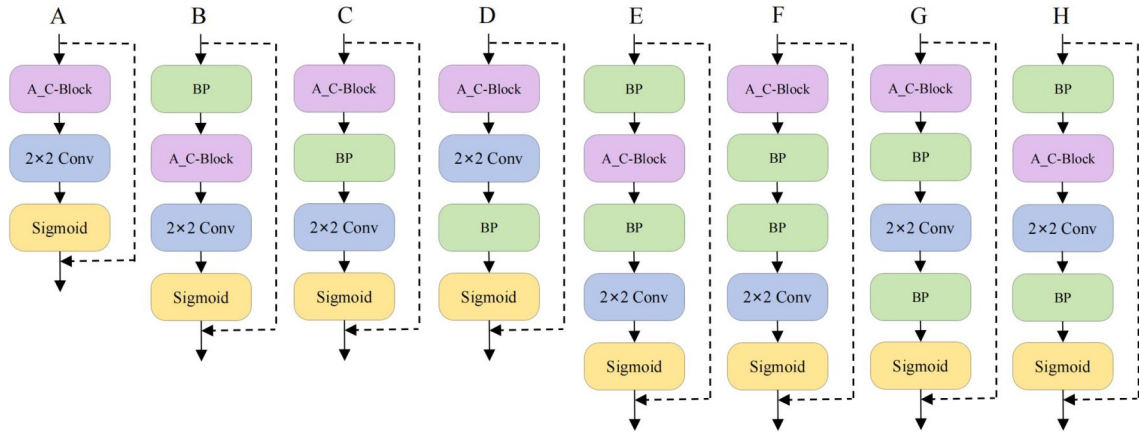


图9 BP嵌入位置和数量的排列结构图

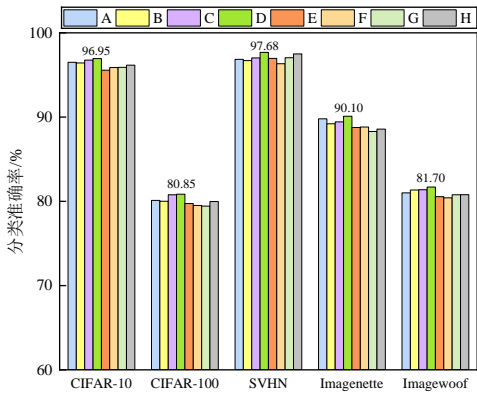


图10 不同位置和数量BP对分类准确率的影响

嵌入不同位置及不同数量的注意力机制在5个数据集上的分类准确率如图12所示。

由图12可以看出,方案C在CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof数据集上分别取得了96.95%、80.85%、97.68%、90.10%和81.70%的最高分类准确率。这表明,深层网络通常具有较为丰富的抽象特征表达能力,而单个注意力机制能够更好地捕捉到这些深层次的关键信息。通过这种方式,网络能够在训练过程中自适应地关注到更加重要的特征区域,而忽略不重要或冗余的特征。因此,将BPA模块放置于方案C的位置,可以有效地提高网络对关键特征的捕捉能

力,并通过增强特征表达能力来获得最佳的分类性能。

为进一步观察融合BPA的残差块对模型复杂度的影响,实验对比了修改首层卷积尺寸后的原网络(ResNet-34,简称 R_N)与加入BPA后的残差网络(简称 S_N)在五个数据集上的训练时间变化。

实验结果如表2所示。结合图12和表2可知,在所有数据集上, S_N 网络的性能相较于 R_N 网络均有显著提升,且性能提升的幅度远大于时间开销的增加,表明BPA模块能够以较小的计算代价换取显著的性能提升,并具备较高的实用价值。

4.2.4 首层卷积核尺寸对BPMNet性能的影响

网络输入层的特征图大小对模型的性能具有重要影响。为研究首层卷积核大小对BPMNet分类性能的影响,本文在5个数据集上分别采用 3×3 、 5×5 、 7×7 、 9×9 和 11×11 大小的卷积核进行实验。对于CIFAR-10、CIFAR-100和SVHN这三个分辨率较低的图像数据集,为避免特征损失,移除了最大池化层;而对于Imagenette和Imagewoof两个 224×224 大小的图像数据集,由于图像尺寸较大,选用 7×7 、 9×9 和 11×11 大小的卷积核,并保留最大池化层以减少特征图的空间维度。实验结果如图13所示。

从图13中可以看出,较小的卷积核在小尺寸图像数据集上具有更好的分类性能,其中 3×3 卷积核在 32×32

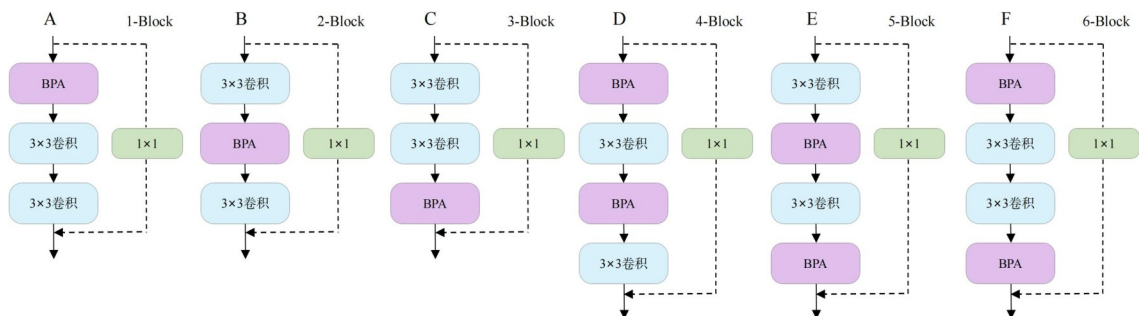


图11 BPA-Block位置和数量的排列结构图

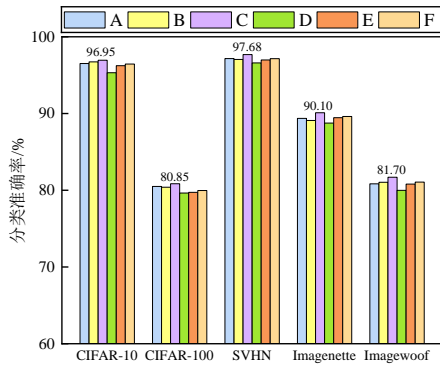


图 12 不同位置和数量BPA对分类准确率的影响

表 2 BPA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	41	42	61	23	25
S_N	67	77	99	32	32

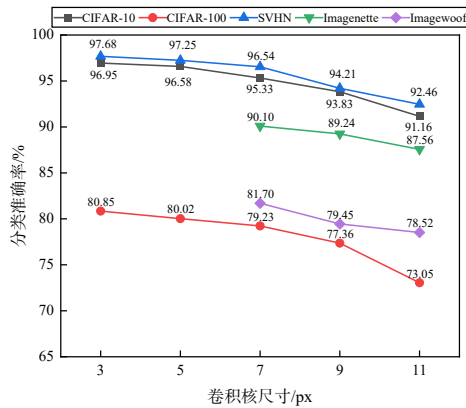


图 13 首层卷积核尺寸对分类准确率的影响

图像数据集上取得最高准确率;而对于 224×224 的图像数据集, 7×7 卷积核取得最高准确率. 进一步分析表明, 简单增大首层卷积核的尺寸并未有效提升分类精度, 甚至可能降低对关键特征的捕捉能力. 这是因为较大的卷积核在提取高分辨率图像的局部细节时容易导致信息的过度平滑, 从而影响网络对重要特征的捕捉能力, 最终降低了分类精度.

4.3 对比实验

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

对比分类方法如下: ResNet-34^[8], EfficientNet^[20], GhostNet^[10], CAPR-DenseNet^[11], QKFormer^[21], TLENet^[22], Couplformer^[18], FAVOR+ (Fast Attention via Positive Orthogonal Random Features Approach)^[17], MMA-CCT-7/3 \times 2^[23], WideResnet-28-10^[9], DAMSNet^[24],

SSLLNet^[25], GAC-SNN^[26], UPANets^[27], ATONet^[28], CCT-6/3 \times 1^[29]. 本文中对分类准确率的对比实验数据来源如下: (1) 对于未开源的代码, 优先采用对比网络所对应论文提供的实验结果; (2) 通过使用论文提供的开源代码进行复现实验.

各网络在 5 个数据集上的分类准确率如表 3 所示, 表中加粗字体表示最优值, “—”表示未知结果.

通过对比表 3 中所有模型分类准确率可知, BPMNet 在不同的分类数据集上均优于其他 16 种网络模型. 其最高准确率分别为 96.95%、80.85%、97.68%、90.10%、81.70%. 对比其他 16 种网络, 平均提升了 2.39%、3.17%、2.36%、2.30%、2.67%, 验证了本文所提出的 BPMNet 在处理图像分类任务方面具有一定的实用性和有效性.

为进一步验证 BPMNet 的有效性, ResNet-34 和 BPMNet 在 CIFAR-10 (32×32) 和 Imagewoof (224×224) 数据集上的分类混淆矩阵如图 14 所示, 列出了各个类别的正确与错误样本数量. 从图 14 中可以看出, 与 ResNet-34 相比, BPMNet 在两个数据集上都拥有更多的正确分类样本和更少的错误分类样本. 这表明, BPMNet 在类别区分能力上优于 ResNet-34, 展现出更强的分类性能.

4.4 消融实验

为了验证本文网络中各模块对分类准确率提升的贡献, 在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Imagewoof 数据集上进行消融实验, 通过移除或替换特定模块, 评估其对模型性能的实际影响.

定义如下网络: Net_1 表示在 BPMNet 中去掉残差主干中的 BP 模块, Net_2 表示在 BPMNet 中去掉残差主干中的 BPA 模块, Net_3 表示在 BPMNet 中不在首层卷积后添加 BP 模块, Net_4 表示在 BPMNet 中不修改首层卷积核尺寸. 各网络具体消融实验结果如表 4 所示, 表中加粗字体表示最优值.

由图 15 可知, 在经过 BPMNet、 Net_1 、 Net_2 、 Net_3 、 Net_4 5 种不同的消融实验之后分类准确率依次下降, 从而验证在残差主干中嵌入 BP 模块能够有效感知背景信息, 提升前景-背景关联性, 从而增强模型对复杂场景的理解. 在残差主干中使用 BPA 模块, 可以自适应地关注图像中的重点区域, 增强关键区域特征的贡献, 进一步强化网络对重要信息的捕捉能力. 将首层普通卷积后的结构替换为 BP 模块, 可以使网络能够更早地捕捉关键特征, 使得后续层能够处理更精确的特征表示, 从而提升模型的分类精度和对复杂环境的鲁棒性. 修改首层卷积核尺寸, 可以更精确地提取细节特征, 避免特征丢失, 有效保留图像原始信息.

表 3 各网络在 5 个数据集上的分类准确率

单位:%

网络	CIFAR-10	CIFAR100	SVHN	Imagenette	Imagewoof
ResNet-34	88.10	72.35	91.74	87.24	77.82
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
QKFormer	96.18	80.26	97.13	88.32	81.65
TLENet	95.46	78.42	96.83	87.62	80.57
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	—	—	—
WideResnet-28-10	95.83	79.50	95.21	88.34	78.71
DAMNet	96.51	80.50	97.60	—	—
SSLNet	95.51	79.23	96.91	87.93	80.89
GAC-SNN	96.00	80.23	96.17	—	—
UPANets	95.32	80.69	96.64	—	—
ATONet	94.51	78.54	95.21	86.67	80.19
CCT-6/3×1	95.29	77.31	96.83	—	—
BPMNet	96.95	80.85	97.68	90.10	81.70

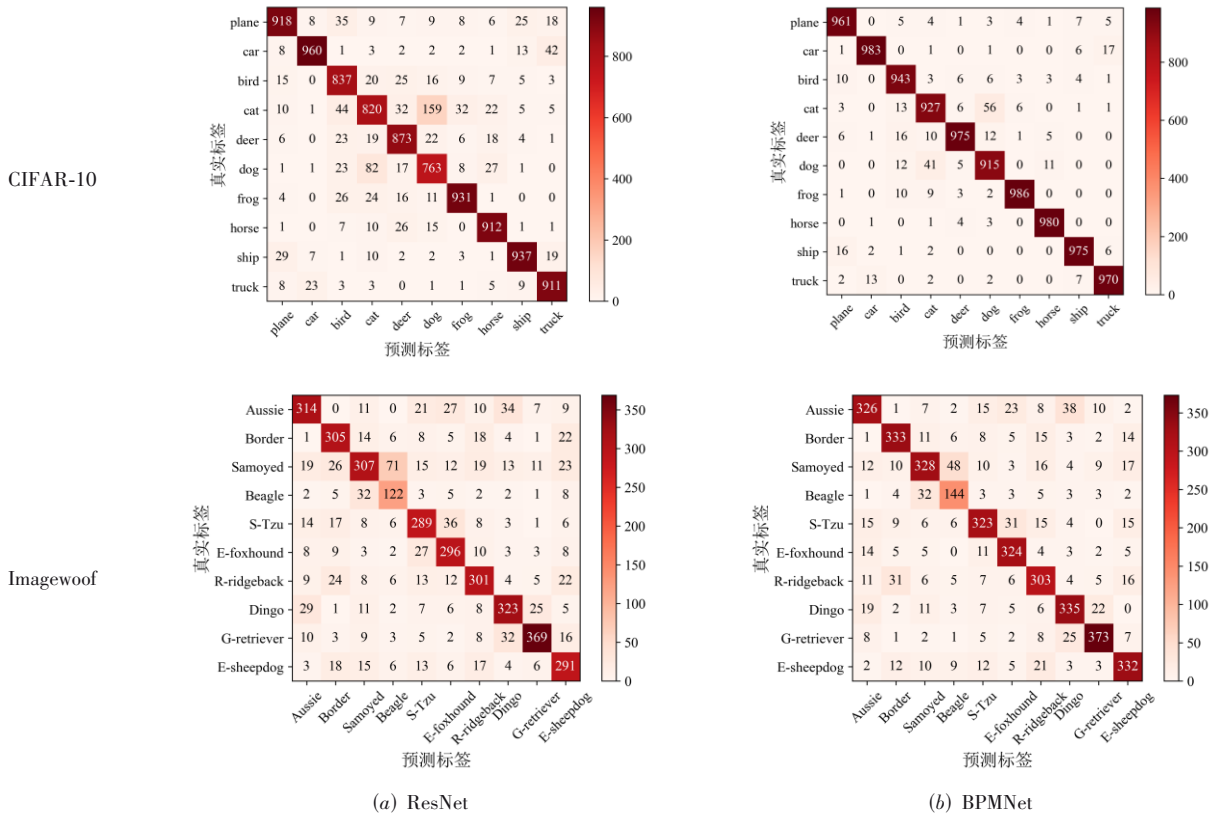


图 14 ResNet-34 和 BPMNet 在 2 个数据集上的混淆矩阵

表 4 BPMNet 的消融实验结果 单位: %

网络	分类准确率				
	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
BPMNet	96.95	80.85	97.68	90.10	81.70
Net ₁	96.23	79.83	96.52	89.24	79.67
Net ₂	94.44	78.54	95.74	88.15	78.57
Net ₃	91.32	77.79	94.45	87.24	77.82
Net ₄	88.10	72.35	91.74		

实验表明,修改首层卷积核尺寸、在第首层卷积后添加BP模块、在残差主干中嵌入BP及BPA能够有效提升网络性能。

4.5 可视化分析

4.5.1 BP 模块可视化分析

为进一步验证背景感知模块在处理复杂背景中的优势,图 16 展示了对同一张图片在相同通道上未使用 BP 模块处理的特征图和使用 BP 模块处理后的特征图

对比. 为了直观地感受不同图片之间的差异,所有图片均采用了灰度处理。

从图 16(a)中可以看出,未使用BP模块时,在某些通道会产生大量负值特征被直接置零的情况(如图 16(a)中黑色区域),导致背景信息(如草地、远处树木)及前景中低响应区域(如小狮子均匀毛发)的语义信息完全丢失;而在图 16(b)中,经过BP模块处理后的特征图则显示出明显的改善. 首先,小狮子均匀毛发(低响应区域)的负值特征被部分保留,避免因完全置零导致的细节丢失. 其次,与图 16(a)中消失的背景细节相比,BP模块通过动态调整背景特征的权重,使其与前景特征有效互补,在灰度图像中,特征图中的细节部分更加突出,边缘更加清晰,同时也能更好地捕捉到图像的上下文关系,表现出更为丰富的语义信息,提升模型对图像内容的理解. 更为重要的是,BP模块不仅保留一定的背景信息,还增强了前景与背景之间的关联性,提升了特征图的整体信息密度。

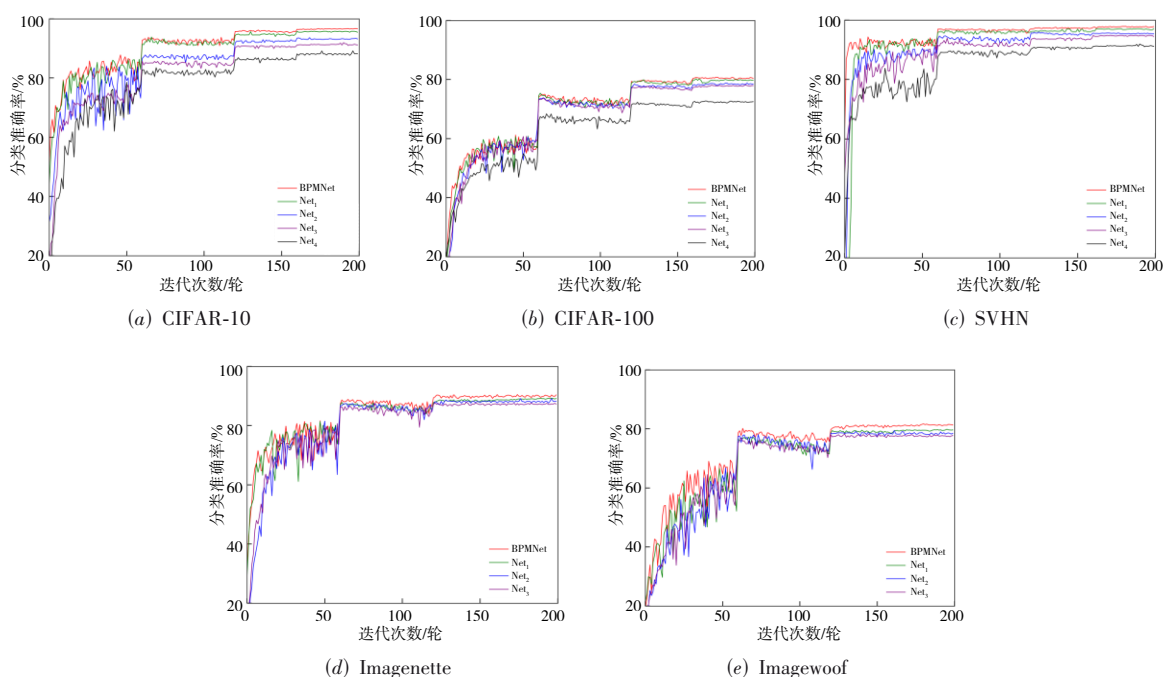


图 15 各网络在 5 个数据集上的分类准确率



(a) 未经过 BP 模块处理的图像

(b) 经过 BP 模块处理的图像

图 16 经过和未经过 BP 模块处理的图像

4.5.2 BPA 可视化分析

为进一步观察 BPA 对本文网络的有效性,选择如下四种注意力机制进行对比分析:EMA(Efficient Multi-Scale Attention)^[30], FCA(Frequency Channel Attention)^[15], SA(Split Attention)^[31], LA(Lightning Attention-2)^[32].

如图 17 所示,展示了不同注意力机制在 CIFAR-10 和 Imagewoof 数据集上的可视化效果.其中前三张图片来自分辨率为 32×32 的数据集 CIFAR-10,后三张图片来自分辨率为 224×224 的数据集 Imagewoof.

根据图 17 的可视化结果,在 CIFAR-10 数据集中,

BPA 表现出更高的专注度,与其他注意力机制相比, BPA 能够更全面地捕捉图像中的关键特征,同时对细节部分表现出更高的敏感性.不仅能够突出主体区域,还能有效提取猫的耳朵等边缘和细微特征,从而实现更精准的特征提取和特征表达.在 Imagewoof 数据集中, BPA 同样能够显著聚焦于狗的眼睛、鼻子等重要特征,还能细致地提取耳朵、尾巴等边缘区域的细节特征,关注区域较大.此外, BPA 还能够对背景中的关联性信息进行合理建模,避免忽略可能有助于分类的场景特征,使得模型在处理复杂场景时,能够更加全面地提取有效信息,显著提升模型分类性能和泛化能力.

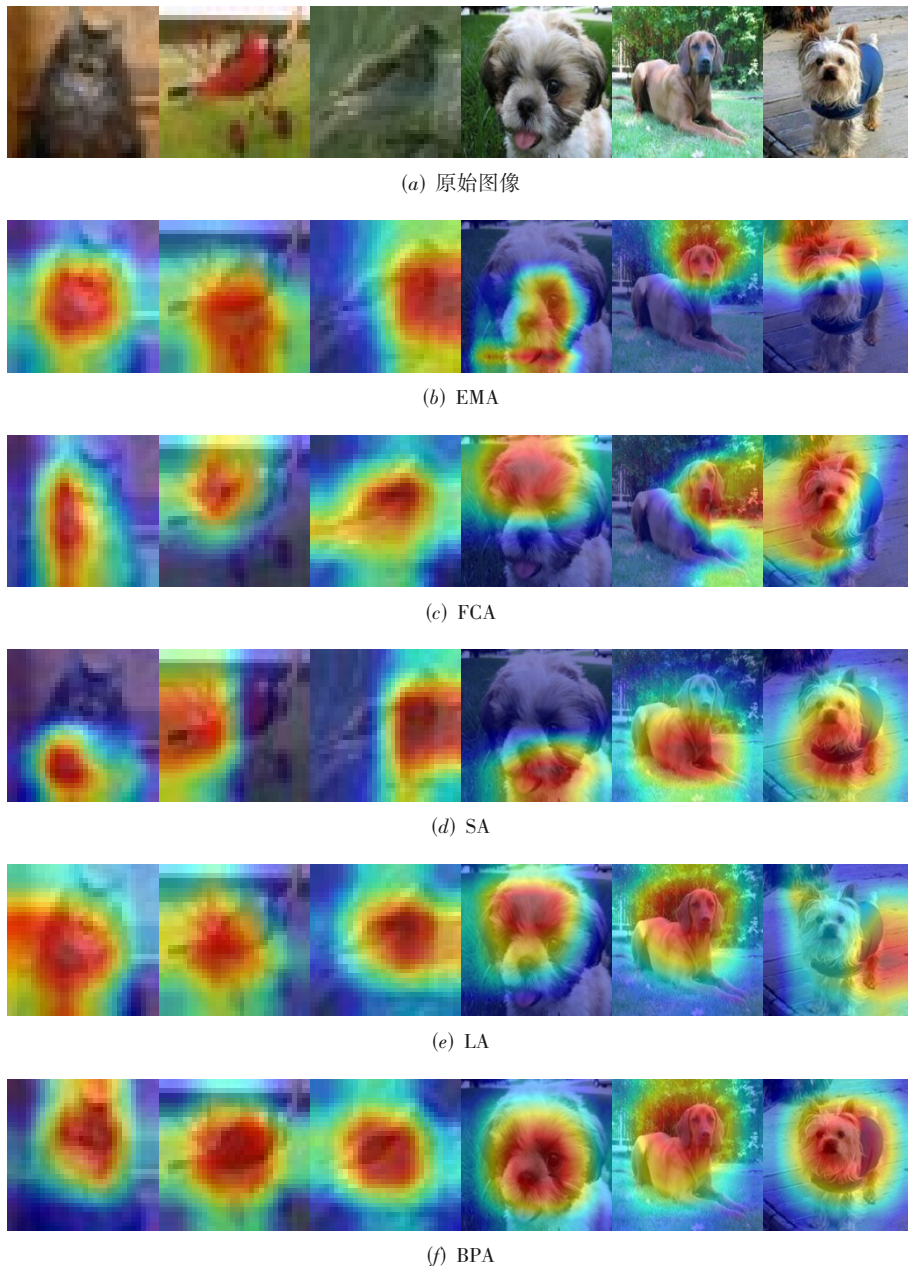


图 17 不同注意力机制的热力图对比

5 结论

针对目前图像分类方法缺少背景信息支持,缺乏对复杂场景的有效理解,导致模型对关键特征的捕捉能力受限,进而影响分类精度等问题,本文提出背景感知机制的图像分类网络(BPMNet)。BPMNet基于背景感知机制,设计了BP模块和BPA模块,充分利用有效背景信息来丰富特征表达,建立更准确的前景-背景关联,确保模型对目标特征的有效关注;同时,为了促进背景信息在网络中的有效传递,调整了残差块内部的网络结构,通过残差块结构优化,构建背景信息与前景特征的高阶语义交互联系,实现背景上下文的高效传递以及前景关键区域特征表示的优化,增强模型对复杂场景的理解能力,从而提高模型的分类精度和鲁棒性。

BPMNet虽然取得了显著效果,但也存在不足之处,在BPA中过多地嵌入BP模块,会导致背景特征重叠,在一定程度上降低模型的性能。在接下来的研究中,将对现有的BPA模型进行优化,通过增加网络深度和改进模块设计,以提高模型的准确性和泛化能力。

参考文献

- [1] 杨传广, 陈路明, 赵二虎, 等. 基于图表征知识蒸馏的图像分类方法[J]. 电子学报, 2024, 52(10): 3435-3447.
YANG C G, CHEN L M, ZHAO E H, et al. Graph-based representation knowledge distillation for image classification[J]. Acta Electronica Sinica, 2024, 52(10): 3435-3447. (in Chinese)
- [2] 姜文涛, 高原, 袁姮, 等. 门控机制的图像分类网络[J]. 电子学报, 2024, 52(7): 2393-2406.
JIANG W T, GAO Y, YUAN H, et al. Image classification network of gating mechanism[J]. Acta Electronica Sinica, 2024, 52(7): 2393-2406. (in Chinese)
- [3] JIANG W T, YUAN H, LIU W J. Neuron signal attenuation activation mechanism for deep learning[J]. Patterns, 2025, 6(1): 101117.
- [4] 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.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2025-01-13]. <https://arxiv.org/pdf/1409.1556>.
- [7] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 1-9.
- [8] 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.
- [9] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL]. (2016-05-23)[2025-01-13]. <https://arxiv.org/pdf/1605.07146>.
- [10] 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: 1577-1586.
- [11] 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: 2261-2269.
- [12] CHEN Y, LI J, XIAO H, et al. Dual path networks[J]. Advances in Neural Information Processing Systems, 2017, 1: 4470-4478.
- [13] 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.
- [14] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11534-11542.
- [15] 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: 763-772.
- [16] 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.
- [17] CHOROMANSKI K, LIKHOSHERSTOV V, DOHAN D, et al. Rethinking attention with performers[EB/OL]. (2020-09-30)[2025-01-13]. <https://arxiv.org/pdf/2009.14794>.
- [18] 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: 6464-6473.
- [19] 张峰, 黄仕鑫, 花强, 等. 基于Depth-wise卷积和视觉Transformer的图像分类模型[J]. 计算机科学, 2024, 51(2):

196-204.

ZHANG F, HUANG S X, HUA Q, et al. Novel image classification model based on depth-wise convolution neural network and visual transformer[J]. Computer Science, 2024, 51(2): 196-204. (in Chinese)

- [20] 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.
- [21] ZHOU C, ZHANG H, ZHOU Z, et al. QKFormer: Hierarchical spiking transformer using QK attention[EB/OL]. (2024-03-25) [2025-01-13]. <https://arxiv.org/pdf/2403.16552v1>.
- [22] SHIN H, CHOI D W. Teacher as a lenient expert: Teacher-agnostic data-free knowledge distillation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(13): 14991-14999.
- [23] KONSTANTINIDIS D, PAPASTRATIS I, DIMITROPOULOS K, et al. Multi-manifold attention for vision transformers[J]. IEEE Access, 2023, 11: 123433-123444.
- [24] 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)
- [25] MA C, WU J, SI C, et al. Scaling supervised local learning with augmented auxiliary networks[EB/OL]. (2024-

02-27)[2025-01-13]. <https://arxiv.org/pdf/2402.17318>.

- [26] QIU X, ZHU R J, CHOU Y, et al. Gated attention coding for training high-performance and efficient spiking neural networks[EB/OL]. (2024-06-04)[2025-01-13]. <https://arxiv.org/pdf/2308.06582>.
- [27] TSENG C H, LEE S J, FENG J N, et al. UPANets: Learning from the universal pixel attention networks[J]. Entropy, 2022, 24(9): 1243.
- [28] WU X D, GAO S Q, ZHANG Z Y, et al. Auto-train-once: controller network guided automatic network pruning from scratch[EB/OL]. (2024-03-21) [2025-01-13]. <https://arxiv.org/pdf/2403.14729>.
- [29] HASSANI A, WALTON S, SHAH N, et al. Escaping the big data paradigm with compact transformers[EB/OL]. (2022-06-07)[2025-01-13]. <https://arxiv.org/pdf/2104.05704>.
- [30] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2023: 1-5.
- [31] 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: 2735-2745.
- [32] QIN Z, SUN W, LI D, et al. Lightning attention-2: A free lunch for handling unlimited sequence lengths in large language models[EB/OL]. (2024-01-15) [2025-01-13]. <https://arxiv.org/pdf/2401.04658>.

作者简介



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



冉 超 女, 2000年10月出生于辽宁省朝阳市. 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: 1487194401@qq.com



张晟翀 男, 1973年4月出生于安徽省合肥市. 高级工程师. 主要研究方向为数字信号处理、模式识别与人工智能.
E-mail: zsc417@126.com