

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)08-2790-18

论文引用格式: Jiang W T, You Z C and Yuan H. 2025. Image classification network with random dilated convolution. Journal of Image and Graphics, 30(8):2790-2807(姜文涛, 由卓丞, 袁姮. 2025. 随机空洞卷积的图像分类网络. 中国图象图形学报, 30(8):2790-2807)[DOI:10.11834/jig.240746]

随机空洞卷积的图像分类网络

姜文涛, 由卓丞*, 袁姮

辽宁工程技术大学软件学院, 葫芦岛 125105

摘要: 目的 针对图像分类任务中对于细粒度特征提取困难, 同时背景噪声和不相关区域影响网络对目标特征学习的问题, 本文提出随机空洞卷积的图像分类网络(image classification network with random dilated convolution, RDCNet)。方法 RDCNet 网络以 ResNet-34(residual network-34)为基线网络。首先, 提出多分支随机空洞卷积(multi-branch random dilated convolution, MRDC)模块, 通过多个分支的卷积操作和随机膨胀卷积核的设计, 实现了从不同尺度和感受野上对细粒度特征的有效捕捉。通过引入细粒度特征增强(fine-grained feature enhancement, FGFE)模块, 实现对全局信息的学习和局部特征的增强, 提升了网络局部特征提取和全局上下文理解能力。同时引入随机掩码机制动态地遮蔽部分输入特征和卷积核权重, 不仅可以通过多样化的特征组合学习更加健壮和鲁棒性的表示, 还能够有效减少过拟合, 提升对噪声和不相关区域的适应能力。最后, 提出上下文激励(context excitation, CE)模块, 通过引入上下文信息并动态调整特征通道的权重, 增强网络对关键特征的专注能力, 抑制背景噪声的干扰, 提升了特征的表达能力。结果 本文方法在 CIFAR-10(Canadian institute for advanced research 10)、CIFAR-100、SVHN(street view house number)、Imagenette 和 Imagenet 数据集上均有良好的分类准确率, 相比于性能第2的模型, 分类准确率分别提高了 0.02%、1.12%、0.18%、4.73% 和 3.56%。实验结果表明, RDCNet 具有较高的分类性能。结论 随机空洞卷积的图像分类网络具有更强的细粒度特征敏感度, 能够在多尺度和上下文中提取丰富的特征信息, 较好地关注关键特征, 对复杂背景下目标具有更优秀的辨识能力, 从而在分类任务中表现出优秀的分类性能。

关键词: 图像分类; 残差网络; 空洞卷积; 随机空洞卷积; 细粒度特征; 随机掩码机制

Image classification network with random dilated convolution

Jiang Wentao, You Zhuocheng*, Yuan Heng

College of Software, Liaoning Technical University, Huludao 125105, China

Abstract: Objective Image classification tasks are particularly common in modern computer vision tasks. However, with the continuous development of deep learning methods, effectively extracting fine-grained features, suppressing noise interference, and dealing with target features in complex backgrounds remain difficult to be solved. In particular, although residual networks exhibit strong feature learning capabilities, they often struggle to efficiently learn fine-grained features due to the diversity of training data, background noise, and differences between target objects at different scales. Conventional convolutional operations typically ignore detailed information at different scales when dealing with such complex prob-

收稿日期: 2024-12-12; 修回日期: 2025-01-01; 预印本日期: 2025-01-08

* 通信作者: 由卓丞 1046491150@qq.com

基金项目: 国家自然科学基金项目(61601213); 辽宁省自然科学基金项目(20170540426); 辽宁省教育厅重点基金项目(LJYL049)

Supported by: National Natural Science Foundation of China (61601213); Natural Science Foundation of Liaoning Province, China (20170540426); Key Project of Education Department of Liaoning Province (LJYL049)

lems. They are susceptible to overfitting when dealing with noisy or irrelevant regions, which leads to performance degradation. This study proposes image classification network with random dilated convolution (RDCNet) to address the abovementioned challenges. The network aims to solve the problems of difficulty in fine-grained feature extraction, background noise interference, and overfitting. Several innovative designs have enabled the extraction of key features from complex backgrounds and improved the classification ability of the network. **Method** RDCNet is based on the classical ResNet-34 as the backbone network, which utilizes its powerful residual connectivity property to enhance the training depth and stability of the network. In this study, several innovative modules are proposed on this basis to enhance the feature extraction capability of the network. First, the multi-branch random dilated convolution module is proposed, which realizes the effective capture of fine-grained features from different scales and sensory fields by the convolution operation of multiple branches and the design of randomly inflated convolution kernel. Compared with the traditional convolution operation, the expansion convolution can expand the receptive field without increasing the computational effort. Thus, it enhances the capture of the multi-scale information in the image. The design of the randomized expansion rate ensures that the network can adapt to different target scales and variations by diversifying the convolution kernel structure, which further improves the ability of fine-grained feature extraction. The Fine-Grained Feature Enhancement module is introduced to enhance the sensitivity of the network to small objects and detailed features by fusing global information with local features, which in turn improves the representation of local features. The module extracts global features of an image using a global average pooling operation and models them jointly with local features to help the network better understand contextual information. With this feature enhancement mechanism, the network can classify small differences more accurately, which boosts the ability to recognize tiny targets. At the same time, the introduction of random masking mechanism dynamically masks part of the input features and convolutional kernel weights. This process not only enables learning of more robust representations through diverse feature combinations but also effectively reduces overfitting and improves the ability to adapt to noise and unknown inputs. This mechanism is similar to the Dropout technique, but the difference is that the random masking mechanism is based on the dynamic masking of features and convolutional kernels, which has higher flexibility to cope with various complex input data. Finally, the Context Excitation module is proposed to enhance the ability of the network to focus on key features by introducing contextual information and dynamically adjusting the weights of feature channels. In image classification tasks, features in certain regions are commonly critical to the final classification results, while background noise may negatively affect the results. This module helps the network focus on the critical features that contribute to the classification task by adaptively adjusting the importance of each feature channel while suppressing interference from irrelevant regions. **Result** This study conducts experiments on the CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewoof datasets to verify the effectiveness of RDCNet in image classification tasks. The experimental results show that RDCNet exhibits a significant performance improvement on all the datasets. RDCNet compares with the model with the second highest performance in the CIFAR-10 dataset by 0.02%, which confirms its superiority in fine-grained feature extraction. The classification accuracy of RDCNet on the CIFAR-100 dataset is enhanced by 1.12%, which demonstrates its superiority in dealing with large-scale classification tasks. The classification accuracy of RDCNet on the SVHN dataset is increased by 0.18%, which highlights its superiority in dealing with the Street View Digit Recognition problem. The classification accuracy of RDCNet on the Imagenette dataset is improved by 4.73%, which verifies its ability to recognize target features in more complex backgrounds. The classification accuracy of RDCNet on the Imagewoof dataset is raised by 3.56%, which further proves the excellent performance of the network in different scenarios. In addition, to deeply analyze the roles of each innovative module in RDCNet, this study conducts ablation experiments to further demonstrate their unique contributions. These experiments involve gradually removing key modules from the network to evaluate their impact on the overall performance, which demonstrates how the synergy between the modules enables the network to maintain superior performance when dealing with complex tasks. Through these experiments, this study demonstrates the superiority of RDCNet on multiple datasets, especially in handling complex background, noise, and fine-grained features. **Conclusion** We propose RDCNet to effectively solve the fine-grained feature extraction, noise interference suppression, and overfitting problems in image classification tasks. This network introduces an innovative multi-branch randomized cavity convolution module, a fine-grained feature enhancement module, a randomized masking mechanism, and a contextual excitation module. The experimental results show that the

classification performance of RDCNet on multiple standard datasets is significantly improved, especially its target recognition ability in complex background. The key contributions of RDCNet are stronger fine-grained feature sensitivity through multi-scale feature extraction, fine-grained feature enhancement, and contextual information modeling. It excels in extracting rich feature information in multi-scales and contexts. It focuses on key features and discriminates targets in complex contexts, which lead to excellent performance in classification tasks. In addition, the introduction of the random masking mechanism effectively enhances the robustness of the network and reduces the risk of overfitting, which enable it to perform more stably in complex real-world scenarios. Future research can further explore the application of this method in other visual tasks and its integration with other advanced techniques to boost its performance.

Key words: image classification; residual network; dilated convolution; random dilated convolution; fine-grained feature; random masking mechanism

0 引言

图像分类是计算机视觉领域的核心任务,其目的是根据图像中的对象或场景,将图像划分到预定义的类别中。基于深度学习的分类方法已成为主流,其中卷积神经网络(convolutional neural network, CNN)是应用最广泛的网络架构之一。Lecun 等人(1998)提出 LeNet 通过多个卷积层和池化层从原始图像中提取高级特征表示,减少了手动特征设计的需求。Krizhevsky 等人(2017)提出 AlexNet 通过增加网络深度,使网络能够学习到更复杂和更高级别的特征表示。随后 VGGNet (Visual Geometry Group network) (Simonyan 和 Zisserman, 2015) 和 GoogLeNet (Szegedy 等, 2015) 的提出进一步证明了增加网络深度的优势,同时 GoogLeNet 中 Inception 模块实现了多尺度特征提取,但也暴露了深层网络中梯度消失和梯度爆炸等问题。He 等人(2016)通过引入深度网络中的残差连接提出了 ResNet(residual network),有效地解决了上述梯度问题。同时 WRN (wide residual network) (Zagoruyko 和 Komodakis, 2017)通过增加残差块的通道数提高了模型的表现力;DenseNet(Huang 等, 2017)采用密集连接结构,有效地缓解梯度消失的问题;MobileNet(Howard 等, 2017)引入深度可分离卷积等实现了网络轻量化设计。

上述分类方法通过堆叠 3×3 卷积主要聚焦于局部特征提取,虽然随着网络深度增加会扩大感受野,但对于大尺度特征的学习仍显不足,无法充分利用输入数据的空间上下文信息。为解决上述问题, Wang 等人(2022)提出 HRNet (high-resolution network),在保持高分辨率特征和多尺度特征融合方面

表现出色,但缺点在于其高计算和内存开销。Abdi 和 Nahavandi(2017)提出 Multi-ResNet,通过多分辨率特征提取有效提升了模型的性能,但多分辨率特征的提取和融合需要更多的计算资源导致训练速度下降。Luo 等人(2022)提出 HO-ResNet(hybrid-order resnet),通过高阶特征提取和多尺度特征融合机制改进了深度学习模型,但对于不同位置关键信息的捕捉仍然存在不精确性。Chen 等人(2018)提出 DeepLab,其中的 ASPP(atrous spatial pyramid pooling)模块通过并行不同膨胀率的空洞卷积拓宽感受野,获得多尺度上下文信息,但在细粒度特征处理上有所欠缺。Liu 等人(2018)提出 RFBNet(receptive field block network),通过结合不同大小的卷积核和空洞卷积更好地捕捉图像中的多尺度特征,但仍存在学习特征的灵活性不足、特征提取的过程对输入噪声敏感。黄港等人(2023)提出 CADF(complementary attention diversity feature fusion),利用分层感受野的差异学习细粒度特征。同时, FADC(frequency adaptive dilated convolution)(Chen 等, 2024)和 SFH-former (spatial-frequency hybrid transformer) (Jiang 等, 2025)针对增强细粒度特征提取能力做了进一步改进。

传统 CNN 通过增加卷积层深度建立远程依赖关系,然而长距离位置之间的信息传递受到限制会导致效率低下且优化困难,持续堆叠卷积层会导致卷积核性能退化。为此, Vaswani 等人(2017)提出了 Transformer 架构。Wang 等人(2018)随后提出 Non-Local 模块,利用自注意力(self-attention)机制使网络有效地捕获远程信息,增强特征表示能力,但 Non-Local 计算量巨大。CCNet(criss-cross network) (Huang 等, 2019)和 ANN(asymmetric non-local neural network) (Zhu 等, 2019)针对减少 Non-Local 的计

算量做了优化。在Non-Local和SENet(squeeze-and-excitation network)(Hu等, 2018)的基础上, Cao等人(2019)提出GCNet(global context network), 能够有效地捕捉全局上下文信息, 同时具有较强的自适应性。

受上述思想启发, 本文提出随机空洞卷积的图像分类网络(image classification network with random dilated convolution, RDCNet)。主要贡献如下: 1) 提出多分支随机空洞卷积模块(multi-branch random dilated convolution, MRDC), 将MRDC模块替换BasicBlock中的 3×3 卷积, 通过多分支结构和不同空洞率的空洞卷积提取特征, 有效地扩大感受野, 显著提升了模型的特征提取能力。2) 引入细粒度特征增强模块(fine-grained feature enhancement, FGFE), 使模型在决策时考虑更多的上下文信息, 增强模型局部特征的表达能力, 提高对空间信息的感知。3) 在空洞卷积中引入随机掩码机制, 动态地调整卷积核的权重, 使模型能够更好地适应不同的输入特征, 实现对输入特征的选择性屏蔽, 关注到更有用的特征通道, 降低对噪声的敏感性, 提升模型在复杂环境中的鲁棒性。4) 提出上下文激励模块(context excitation, CE)。CE模块通过软最大化机制更好地捕捉输入特征中的上下文信息, 结合特征重标定, 使特征加权更为灵活和动态, 使网络在不同输入下能有更好的表现。

1 随机空洞卷积的图像分类网络

RDCNet首先修改ResNet34浅层特征提取卷积模块; 设计MRDC模块嵌入残差块中, 显著提升模型的特征提取能力; 最后设计CE模块嵌入深层特征提取阶段后, 使模型在全局平均池化前更好地关注有用的特征, 下文将详细介绍各个模块。

1.1 浅层特征提取阶段

在基线网络ResNet34中, 原始浅层特征提取阶段采用 7×7 的卷积层和 3×3 最大池化层用于初级特征提取。为了优化模型, 本文调整卷积核的大小为 3×3 , 且步长为1, 填充为1, 同时删除最大池化层。通过卷积层和池化层后的图像尺寸为

$$W_{\text{out}} = \frac{(W_{\text{in}} + 2p - k)}{s} + 1 \quad (1)$$

$$H_{\text{out}} = \frac{(H_{\text{in}} + 2p - k)}{s} + 1 \quad (2)$$

式中, W_{in} 和 H_{in} 为输入特征图的宽度和高度, p 为填充的大小, k 为卷积核的大小, s 为步长, W_{out} 和 H_{out} 为输出特征图的宽度和高度。

假设图像输入尺寸为 32×32 像素, 经过 7×7 卷积层和最大池化层处理后图像输出尺寸变为 8×8 像素, 若仅通过一个 3×3 卷积层处理, 图像输出尺寸则为 32×32 像素。因此, 大卷积核和池化层虽然减少了特征图的空间分辨率, 但会导致细节信息丢失, 尤其在浅层特征提取阶段会忽视小的但重要的特征。本文使用 3×3 小卷积核能够逐层捕捉细微的图像特征, 使输出特征图的空间分辨率保持不变且有效防止特征图的极度下采样。删除最大池化层使模型直接处理原始特征图, 有助于模型更全面地捕捉输入图像中的细节信息, 学习更丰富的特征。

1.2 多分支随机空洞卷积模块

为了从多尺度中有效地捕捉细粒度特征, 增强网络在处理不同类型图像时的表达能力, 本文提出多分支随机空洞卷积(MRDC)模块。

1.2.1 空洞卷积结合随机掩码机制

在深度神经网络中, 空洞卷积(又称膨胀卷积)是一种扩展传统卷积的技术, 通过引入一个超参数dilation(空洞率, 又称膨胀系数)控制卷积核元素之间的间隔, 使得卷积核在不增加参数量的情况下扩大感受野。传统 3×3 卷积如图1(a)所示, 卷积核的元素相邻, 感受野固定为 3×3 。若将空洞卷积的dilation设置为2和3, 其感受野尺寸分别变为 5×5 和 7×7 , 如图1(b)(c)所示。空洞卷积图像输出尺寸为

$$w_{\text{out}} = \frac{w_{\text{in}} + 2p - d(k - 1) - 1}{s} + 1 \quad (3)$$

$$h_{\text{out}} = \frac{h_{\text{in}} + 2p - d(k - 1) - 1}{s} + 1 \quad (4)$$

式中, w_{in} 和 h_{in} 为输入特征图的宽度和高度, d 为空洞率的大小, w_{out} 和 h_{out} 为输出特征图的宽度和高度。假设图像输入尺寸为 32×32 像素, 在 $d = 2, k = 3, s = 1, p = 0$ 的情况下输出尺寸将变为 28×28 像素, 无填充的卷积操作会直接在图像边缘进行, 导致感受野的扩展无法完全补偿使输出尺寸变小。本文选择将填充值 p 始终等于空洞率 d , 填充是对输入图像的边缘进行扩展, 补充零值, 可以确保卷积操作能够

覆盖图像的边缘区域。填充后如图1(d)所示,在 $d=2, k=3, s=1, p=2$ 的情况下,图像输出尺寸仍然为 32×32 ,有助于抵消空洞卷积带来的感受野扩大效应。

尽管空洞卷积能够有效地捕捉更大的上下文信息,但其稀疏采样的特性容易引发空洞效应(grid-

ding effect),导致图像的细节和边缘信息被忽略或丢失。为减轻空洞效应并提升网络的表达能力,本文将空洞卷积与随机掩码机制相结合,通过随机屏蔽部分输入通道和卷积核元素,增强网络的多样性、鲁棒性和泛化能力,有效地缓解空洞效应,提升网络在细节和边缘特征上的提取能力。

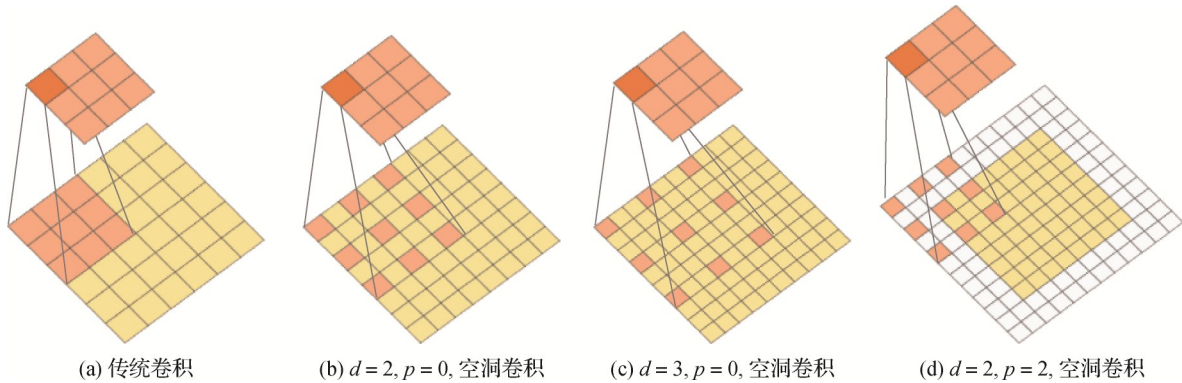


图1 传统卷积和不同参数的空洞卷积示意图

Fig. 1 Schematic diagram of traditional convolution and dilated convolution with different parameters ((a) conventional convolution; (b) $d=2, p=0$, dilation convolution; (c) $d=3, p=0$, dilation convolution; (d) $d=2, p=2$, dilation convolution)

随机掩码机制首先在特征输入通道引入通道掩码对输入特征图的某些通道随机遮挡,再对卷积核引入卷积核掩码对卷积核的权重随机遮挡。随机掩码操作过程如图2所示。假设输入特征图 X 的大小为 (C, H, W) ,其中 C 为通道数, H 和 W 分别为特征图的高度和宽度。通道掩码的目标是生成一个 $(C, 1, 1)$ 的掩码张量,选择是否保留每个通道的特征信息,若掩码值为1,输入特征图的对应位置保留原始值,若掩码值为0,则其对应位置将被屏蔽。通道掩码的生成计算为

$$m_c = \begin{cases} 1 & r_c < mask \\ 0 & r_c \geq mask \end{cases} \quad (5)$$

$$X_c = X \otimes M_c \quad (6)$$

式中, M_c 表示通道掩码矩阵, m_c 为 M_c 的第 c 个元素,值为0或1, r_c 为均匀分布中采样的随机数。 $mask$ 为 $[0, 1]$ 之间的阈值,表示通道被保留的概率。 $\otimes$ 代表按元素相乘, X_c 为经过通道掩码后的特征图。

卷积核掩码是大小为 (C, H, W) 的三维张量,其计算为

$$m_{c,h,w} = \begin{cases} 1 & r_{c,h,w} < mask \\ 0 & r_{c,h,w} \geq mask \end{cases} \quad (7)$$

$$X_k = X_c \otimes M_k \quad (8)$$

式中, M_k 表示卷积核掩码矩阵, $m_{c,h,w}$ 为第 c 个通道的

掩码矩阵中第 h 行 w 列元素, $r_{c,h,w}$ 为从 $[0, 1]$ 区间均匀分布中采样的随机数。 X_k 为经过卷积核掩码后的特征图。

随机掩码机制通过随机屏蔽输入通道和卷积核的部分元素,能够有效地减轻空洞效应,特别是减少图像中心区域的过度关注。然而空洞卷积本质上会导致图像边缘的信息提取较少,结合随机掩码机制后会进一步降低对图像边缘信息的提取。为解决这一问题,本文引入填充值与空洞率相等的策略,填充后的图像尺寸增大,空洞卷积操作在更大的区域内提取信息而不受零填充影响,确保原始图像的边缘区域能够得到充分的特征提取,从而改善了边缘信息的表示能力。

在利用空洞卷积扩大感受野、增强网络对远程上下文信息的捕捉能力的基础上,本文提出一种结合空洞卷积与随机掩码机制的方法,使模型能够在不同的特征组合下进行训练,避免过度依赖某一部分的特征,有助于减少过拟合,提升模型的泛化能力。掩码操作基于逐元素乘法,没有增加额外的计算开销,使结合随机掩码机制的空洞卷积可以在不增加计算量的情况下,获得更多信息。

1.2.2 细粒度特征增强模块

FGFE模块作为MRDC模块中的一条分支,不仅

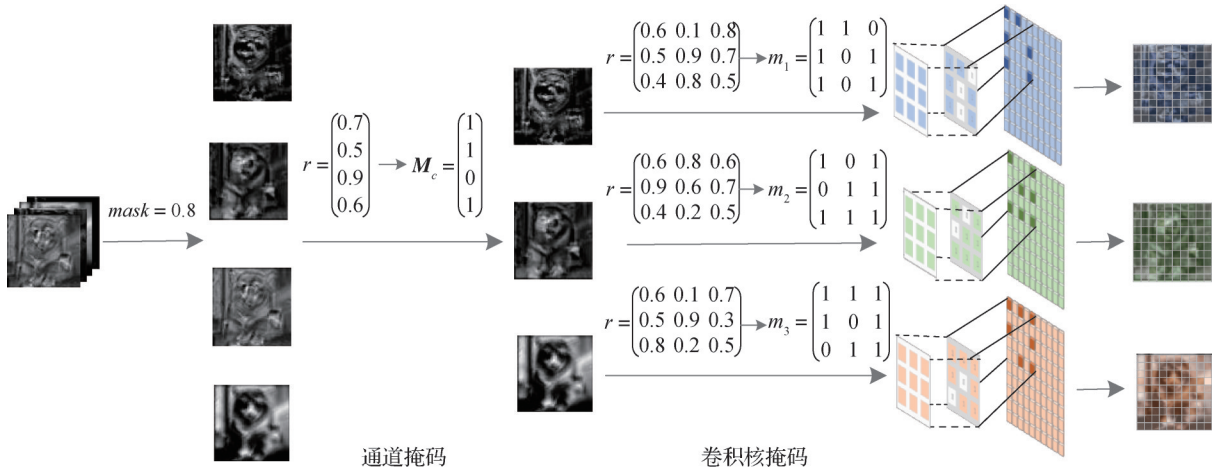


图2 随机掩码结合空洞卷积示意图

Fig. 2 Schematic diagram of random mask combined with dilated convolution

能够协同其他分支共同提取特征信息,还能将全局信息转化为局部特征增强。FGFE模块帮助网络更好地捕捉图像中的微小变化和细节,弥补了其他分支可能忽略的局部信息,有效避免了信息丢失和特征模糊的问题,不仅提高了模型的性能,还增强了对细节特征的敏感性。

FGFE模块输入特征图为 $F_4 \in \mathbf{R}^{c \times h \times w}$ 。首先应用全局平均池化(global average pooling, GAP)通过计算每个通道内所有空间位置的平均值,将每个通道的空间维度压缩至 1×1 的尺寸,得到特征图 $F_p \in \mathbf{R}^{c \times 1 \times 1}$ 表示每个通道的全局信息,计算为

$$F_p = \text{GAP}(F_4) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W F_{c,h,w} \quad (9)$$

式中, H 和 W 分别是特征图的高度和宽度, $F_{c,h,w}$ 表示通道 c 处的空间位置 (h, w) 的特征值, $\text{GAP}(\cdot)$ 表示全局平均池化操作。

将全局平均池化后的特征图通过 1×1 卷积将全局信息映射到局部空间,再通过批归一化层标准化每一层的输出,随后应用 ReLU (rectified linear unit) 激活函数引入非线性变换,增强模型的表达能力,得到输出后的特征图 $F_l \in \mathbf{R}^{c \times 1 \times 1}$, 可表示为

$$F_l = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(F_p)))$$

式中, $\text{Conv}_{1 \times 1}(\cdot)$ 为 1×1 卷积, $\text{BN}(\cdot)$ 为批量归一化, $\text{ReLU}(\cdot)$ 为 ReLU 激活函数,计算为

$$\text{ReLU}(x) = \max(0, x) \quad (10)$$

这一过程通过全局信息引导局部特征,增强局部特征的细节表达能力,有效克服了传统卷积层局部特征信息不足的问题,使网络能够捕捉到图像中

微小的细节特征。最后,将特征图上采样到输入特征图的尺寸,得到 $F'_4 \in \mathbf{R}^{c \times h \times w}$, 可以表示为

$$F'_4 = \delta(F_l)$$

式中, $\delta(\cdot)$ 为双线性插值操作,其计算过程如下。

首先,使用关于 x 的单线性插值分别计算 R_1 和 R_2 的像素值,具体为

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (11)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (12)$$

然后,使用关于 y 方向的双线性插值计算 P 点的像素值,具体为

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (13)$$

式中, P 表示目标插值点, R 表示水平插值的中间结果, Q 表示已知的4个邻域像素, $f(P)$ 表示目标像素值, $f(Q_{11})$ 、 $f(Q_{12})$ 、 $f(Q_{21})$ 、 $f(Q_{22})$ 表示目标像素周围的4个邻域像素值, $f(R_1)$ 、 $f(R_2)$ 表示目标像素在水平方向与左右邻近像素的相对位置进行插值得到的结果。

双线性插值处理图像示意图如图3所示,通过在二维空间上对目标像素(P)周围4个邻近像素(Q_{11} 、 Q_{12} 、 Q_{21} 、 Q_{22})的加权平均进行插值,能够平滑地过渡各个像素的值,不会产生明显的锯齿效应,插值质量和性能可满足大多数分类任务中的上采样需求。

1.2.3 MRDC模块总体结构

MRDC模块总体结构如图4所示,首先对输入特征图 $F_{in} \in \mathbf{R}^{c \times h \times w}$ 的通道分组,在每组内并行计算可有效减少计算量,同时增强网络对不同类型特征的

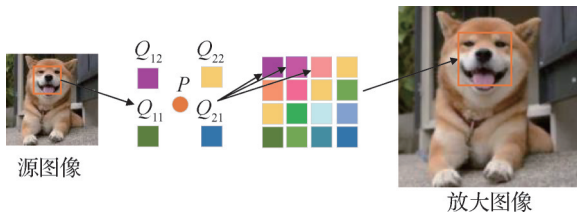


图3 双线性插值图像处理示意图

Fig. 3 Schematic diagram of bilinear interpolation image processing

学习能力。特征通道共分为4组,可表示为

$$\mathbf{F}_{in} = [\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4] \quad (14)$$

式中, $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$ 分别表示4个分支的特征通道,每组通道数为 $c/4$ 。

通道分组后的特征图 $\mathbf{F}_1$ 进入第1个分支,此分支保留原始 BasicBlock 中的第2个 3×3 卷积,确保局部特征的捕捉不被过度简化或忽略。同时将原始 3×3 卷积保留可以确保网络训练时的稳定性,即使在其他分支引入复杂的操作,这一分支仍然提供一个稳定的结构和正常的梯度流。第1个分支经过卷积操作的输出为 $\mathbf{F}'_1$,可以表示为

$$\mathbf{F}'_1 = \text{Conv}_{3 \times 3}(\mathbf{F}_1) \quad (15)$$

式中, $\text{Conv}_{3 \times 3}(\cdot)$ 为 3×3 卷积。

第2个分支中,特征图 $\mathbf{F}_2$ 首先经过空洞率为2的随机掩码空洞卷积得到 $\mathbf{F}_{2r}$,可表示为

$$\mathbf{F}_{2r} = M_DConv(\mathbf{F}_2) \quad (16)$$

式中, $M_DConv(\cdot)$ 为随机掩码空洞卷积。空洞卷积可以在卷积过程中扩大感受野,保留特征图的空间分辨率,不会引入过多的计算负担,从而更好地捕捉到图像中长距离依赖关系。使用随机掩码会在卷积过程中对某些区域进行屏蔽,使网络集中处理更重要的信息,减小噪声的影响。

接着,经过 3×3 卷积得到 $\mathbf{F}'_2$,可以表示为

$$\mathbf{F}'_2 = \text{Conv}_{3 \times 3}(\mathbf{F}_{2r}) \quad (17)$$

经过随机掩码空洞卷积后,特征图信息可能变得抽象和多样化,再通过 3×3 卷积层进行细化,可以进一步提取局部特征并融合多尺度信息。

第3个分支与第2个分支类似,特征图 $\mathbf{F}_3$ 经过空洞率为3的随机掩码空洞卷积得到 $\mathbf{F}_{3r}$,再经过 1×5 卷积和 5×1 卷积细化得到 $\mathbf{F}'_3$,可以表示为

$$\mathbf{F}'_3 = \text{Conv}_{5 \times 1}(\text{Conv}_{1 \times 5}(M_DConv(\mathbf{F}_3))) \quad (18)$$

式中, $\text{Conv}_{1 \times 5}(\cdot)$ 和 $\text{Conv}_{5 \times 1}(\cdot)$ 分别为 1×5 卷积和 5×1 卷积。

通过空洞率为3的随机掩码空洞卷积可以进一步扩大卷积的感受野,捕捉到更大范围的上下文信息。接着通过 1×5 卷积和 5×1 卷积可以捕捉水平方向和垂直方向的特征,帮助细化上下文信息的提取,增强在处理长条形特征或不规则形状特征的感知能力。其次这两个卷积相较于直接使用一个 5×5 卷积有着更低的参数量和计算复杂度,使得网络能够以较少的计算资源捕获到丰富的特征信息。

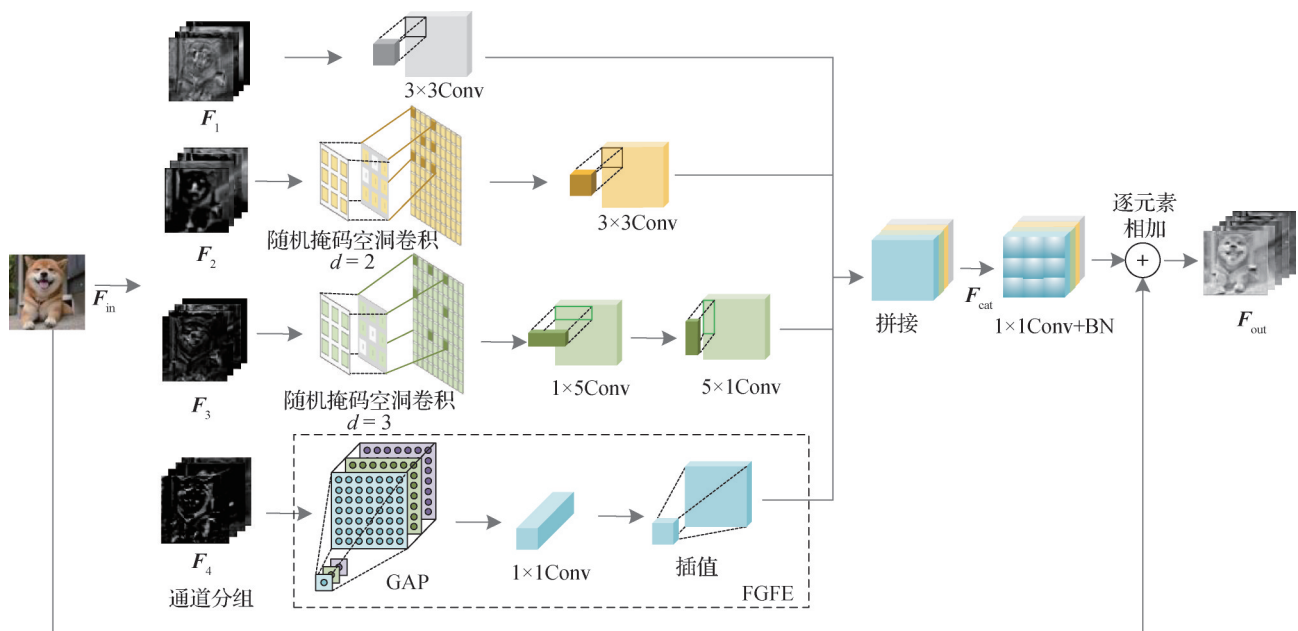


图4 MRDC 总体结构

Fig. 4 General structure of MRDC

将以上3个分支的输出特征图 F'_1, F'_2, F'_3 与经过 FGFE 模块的输出特征图 F'_4 沿通道维度拼接, 将不同特征的信息融合得到 F_{cat} , 表示为

$$F_{cat} = \text{Concat}(F'_1, F'_2, F'_3, F'_4) \quad (19)$$

式中, $\text{Concat}(\cdot)$ 为通道拼接 (concatenate) 操作。

最后, 将通道拼接后的特征图经过 1×1 卷积层确保输出特征图的通道数与输入特征图 F_{in} 保持一致, 再经过批量归一化操作 (batch normalization, BN) 和跳跃连接后的加权求和操作, 得到最终的输出 F_{out} , 可以表示为

$$F_{out} = \text{scale} \times \text{BN}(\text{Conv}_{1 \times 1}(F_{cat})) + F_{in} \quad (20)$$

式中, $\text{BN}(\cdot)$ 是批量归一化操作, scale 是一个缩放因子, 控制主路径输出与残差连接之间的权重。

1.3 上下文激励模块

对于卷积神经网络中局部信息表达不足和上下文信息流动不充分等问题, 本文受自注意力机制的启发, 提出上下文激励模块 (CE), 如图 5 所示。

CE 模块的输入特征图为 $F \in \mathbf{R}^{c \times h \times w}$ 。在上下文建模部分, 首先通过 1×1 卷积使特征图中每一个像素点都被映射为单一的标量值, 作为特征图中每个像素的“key”, 为依赖关系的学习提供基础, F 经过卷积操作的输出为 $F_K \in \mathbf{R}^{1 \times h \times w}$, 表示为

$$F_K = \text{Conv}_{1 \times 1}(F) \quad (21)$$

接着, 对 F_K 进行 softmax 归一化处理, 使每个像素的响应值处于 0~1 范围内, 最终生成一个响应图 $F_S \in \mathbf{R}^{1 \times h \times w}$, 使得模型能够在全局范围自适应增强对重要区域的关注, 其计算式为

$$F_S = \text{softmax}(F_K)$$

式中, $\text{softmax}(\cdot)$ 为 softmax 激活函数, 其计算式为

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} \quad (22)$$

式中, x_i 为向量 x 中第 i 个元素, $\sum_{i=1}^n e^{x_i}$ 为所有类别得分的指数函数之和。

然后, 分别将特征图 F 和 F_S 进行展平操作, 将特征图 F 展平成一个 c 行, $h \times w$ 列的矩阵, 将特征图 F_S 展平成一个 $h \times w$ 行, 1 列的列向量。将两个矩阵进行乘法操作得到 $F_m \in \mathbf{R}^{c \times 1}$, 表示为

$$F_m = \text{matmul}(F, F_S) \quad (23)$$

式中, $\text{matmul}(\cdot, \cdot)$ 为两个矩阵之间的乘法运算, 目的为将展平的空间信息和通道信息进行结合, 从而得到每个通道在整个空间中的加权值。

将每个通道的加权信息进一步处理, 首先使用 1×1 卷积将输入通道数映射到较小的通道数, 降低空间维度。经过 ReLU 激活后, 再通过 1×1 卷积恢复原始通道数, 对输入特征图的每个通道进行加权处理, 实现更细致的特征调节。最后 sigmoid 激活函数对每个通道特征加权调节, 得到最终加权后的特征图 $F_T \in \mathbf{R}^{c \times 1 \times 1}$ 。这种方式使网络能够为各个通道分配不同的权重, 从而增强对关键特征的关注能力。其计算过程可以表示为

$$F_T = \text{sigmoid}(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(F_m)))) \quad (24)$$

式中, $\text{sigmoid}(\cdot)$ 为 sigmoid 激活函数, 计算为

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (25)$$

式中, e^{-x} 为以 e 为底的指数函数的倒数形式。

在最终融合部分, 输入特征图 F 会与经过 value 加权后的特征图 F_T 融合, 通过逐元素乘法得到输出特征图 $F_{out} \in \mathbf{R}^{c \times h \times w}$, 其计算过程可以表示为

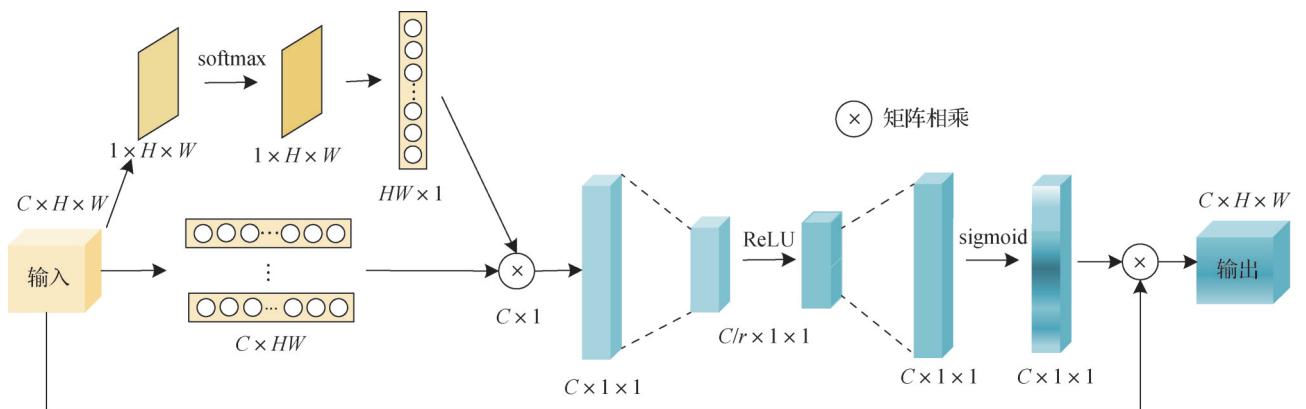


图5 上下文激励模块

Fig. 5 Context excitation block

$$F_{\text{out}} = F_T \cdot F \quad (26)$$

式中,“ $\cdot$ ”表示逐元素乘法。

加权后的特征图 F_T 中的权重能够反映每个特征的重要性,通过逐元素乘法实现特征融合,既聚焦于重要的上下文信息,又保持了特征的稀疏性,使其动态地调整特征图中的信息流。

综上所述,CE模块通过全局上下文建模和键值对机制捕获像素点之间的相似性,有效地捕获了空间位置间的相互关系。再通过通道重加权自适应地调整输入特征图每个通道的响应,使网络能够有效地聚焦于重要的区域,提高特征表达能力。

1.4 融合MRDC模块的残差结构

ResNet的核心思想是通过引入残差学习解决深层神经网络中的退化问题,其核心模块为残差块(residual block),如图6(a)所示,包含两个 3×3 的卷积和 1×1 的卷积的跳跃连接,残差块的核心思想是通过恒等映射(identity mapping)帮助信息在网络中传递。ResNet的残差映射关系为

$$H(x) = F(f(x)) + x \quad (27)$$

式中, x 是输入, $H(x)$ 是输出, $f(x)$ 是通过第一层卷积层、激活函数和归一层处理后的函数, $F(f(x))$ 是经过 $f(x)$ 处理后的特征进一步处理后的残差函数。

平均池化残差块(average pooling residual block, AP-block)将全局平均池化与残差块结合,如图6(b)所示,AP-block在原始ResNet跳跃连接的基础上做一个解耦,将下采样任务交给池化层,使网络在卷积操作时提取更多的特征信息,同时通过平均池化减少特征图的尺寸,避免 1×1 卷积下采样时可能导致的信息丢失。AP-block的残差映射关系为

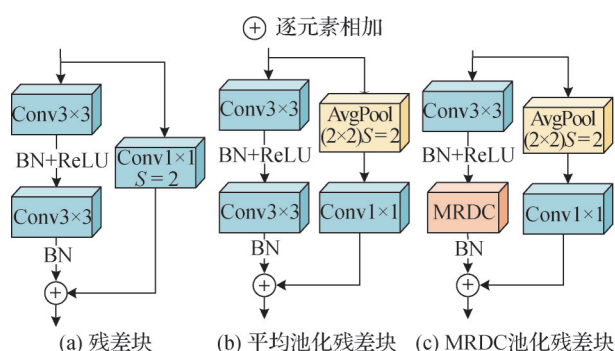


图6 3种残差块结构

Fig. 6 Three residual block structures ((a) residual block; (b) average pooling residual block; (c) MRDC pooling residual block)

$$H(x) = F(f(x)) + g(x) \quad (28)$$

式中, $g(x)$ 表示经过平均池化和卷积层的输入函数。

MRDC池化残差块(MRDC pooling residual block, MRDC-block)如图6(c)所示,MRDC-block以AP-block为基础,再通过不同卷积核和空洞卷积的组合捕捉多尺度信息,同时通过掩码机制引入随机性,降低过拟合风险。MRDC-block的残差映射关系为

$$H(x) = M(f(x)) + g(x) \quad (29)$$

式中, $M(f(x))$ 表示经过 $f(x)$ 处理后的特征通过MRDC后的函数。

1.5 RDCNet总体结构

RDCNet结构如图7所示,共分为5个阶段:

1)预处理阶段。对输入图像进行增强,通过随机翻转、填充后随机剪切和随机擦除(Zhong等, 2020)操作提示样本的多样性,降低过拟合的风险。同时进行图像归一化稳定训练,加快收敛速度。

2)初级特征提取阶段。将预处理后的图像针对不同尺寸图像调整结构,在Imagenette、Imagewoof大尺寸数据集中使用 7×7 卷积和最大池化操作。在CIFAR-10、CIFAR-100和SVHN小尺寸数据集中使用 3×3 卷积,可以更好地适应小尺寸图像,保留细节特征的同时降低过拟合风险。

3)深层特征提取阶段。采用多分支随机空洞卷积的残差模块,通过逐层增加特征图的深度和通道数,实现从低级到高级特征的逐步学习。每一层提取的特征都会在下一层进一步处理,形成多层次、多尺度的特征表示。

4)全局特征整合阶段。CE模块接收深层特征提取阶段的输出特征图,通过上下文信息动态调整通道的特征权重。再通过全局平均池化将特征图的空间维度压缩,生成具有固定尺寸的全局特征向量。

5)分类阶段。将全局平均池化后得到的输出张量展平,通过全连接层映射到最终的结果。

总体而言,RDCNet针对ResNet-34的改进如下:

1)为了增强网络的特征提取能力,通过多个卷积分支(包括不同空洞率的随机掩码空洞卷积和局部特征增强模块)捕捉多尺度信息,提高模型对多样化特征的捕捉能力。使模型在处理复杂图像时更灵活,能够提取到更丰富的上下文信息。2)为了增强通道间的关系进一步整合全局特征,采用CE模块学习特

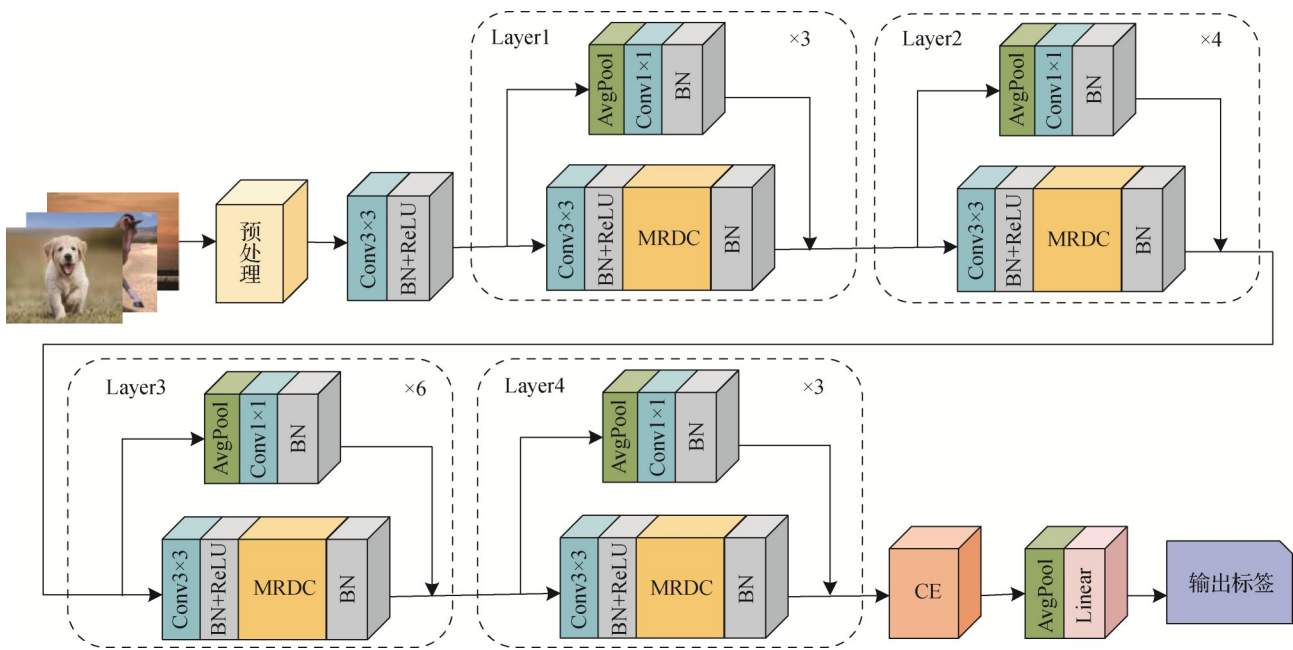


图7 RDCNet 总体结构

Fig. 7 Overall structure of RDCNet

征图的上下文信息以动态调整通道的重要性,从而更关注于关键特征。

2 实验结果及分析

2.1 实验环境

本文实验的编程语言为Python,采用Pytorch框架进行网络搭建,GPU采用NVIDIA RTX 3090显卡。软件环境为PyTorch2.3.0、CUDA11.8、CUDNN8.7.0。为评估RDCNet的分类性能,选择CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof数据集作为实验数据集。其中,自然图像数据集CIFAR-10和CIFAR-100分别包含10种和100种类别图像,用于评估模型在不同类别数和分类难度下的表现;街牌号数据集SVHN包含10种类别图像,用于测试模型在数字识别任务中的泛化能力;Imagenette、Imagewoof数据集包含10种类别图像,均为ImageNet数据集中的子集,用于评估模型在大规模图像分类任务中的表现。数据集信息如表1所示。

实验采用图像分类的分类准确率(accuracy)作为评价指标。在训练过程中,迭代次数设置为200轮,优化器选择随机梯度下降(stochastic gradient descent,SGD),标签平滑为0.1,动量为0.9,权重衰减为 5×10^{-4} 。初始学习率设置为0.1,并采用余弦

退火策略(CosineAnnealingLR)调整学习率。训练期间使用随机翻转、填充后随机剪切和随机擦除操作进行数据增强。CIFAR数据集和SVHN数据集的输入图像尺寸调整为 32×32 像素,批次大小为128,Imagenette数据集和Imagewoof数据集的输入图像尺寸调整为 224×224 像素,批次大小为64。

表1 实验数据集

Table 1 Experimental datasets

数据集	尺寸/像素	类	训练集	测试集
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

2.2 影响MRDC模块性能的参数

影响MRDC模块性能的主要参数包括MRDC模块中掩码策略的不同和MRDC模块中特征融合权重的缩放因子(scale)的大小。

2.2.1 不同掩码策略对MRDC模块性能的影响

随机掩码机制使每次前向传播时动态选择不同的特征组合,捕捉更多样的特征,减少对某些特定特征的依赖,提高网络面对输入噪声时的鲁棒性。将空洞卷积结合掩码机制,可以在扩大感受野的同时

引导模型随机关注重要的特征,不仅帮助减少冗余信息的干扰,还可以降低模型的计算量和存储需求,提升运行效率。为深入研究不同掩码策略对MRDC模块性能的影响,分别采用4种掩码策略在5个数据集上进行对比实验,实验结果如表2所示,其中,不使用任何掩码记为null_mask,使用卷积核掩码记为k_mask,使用通道掩码记为c_mask,结合卷积核掩码和通道掩码记为c_k_mask。

由表2可以得出,将卷积核掩码和通道掩码结合的随机掩码机制在5个数据集分类的效果最佳,由此可见,两种掩码的结合促使模型在卷积核和输入通道之间形成更复杂的特征交互,这种协同学习有助于增强模型的特征表达能力和鲁棒性,使其在处理复杂数据时可以展现出更强的泛化能力。

表2 不同掩码策略对MRDC模块性能的影响
Table 2 Impact of different masking strategies on MRDC module performance

掩码策略	分类准确率/%				
	CIFAR-10	CIFAR-100	SVHN	Inette	Iwoof
null_mask	96.08	79.98	96.92	90.93	83.84
k_mask	96.24	80.41	97.23	91.72	84.93
c_mask	96.16	80.76	97.19	91.11	85.03
c_k_mask	96.38	81.16	97.41	92.58	85.14

注: Inette: Imagenette; Iwoof: Imagewoof。加粗字体表示各列最优结果。

2.2.2 scale值对MRDC模块性能的影响

为深入研究MRDC模块中的特征融合权重值对网络性能的影响,本节对MRDC模块scale的值展开研究,在未嵌入CE模块的基础上,scale取值0.1~1范围内在5种数据集上的实验结果如表3所示。

由表3可知,当scale值为0.4时SVHN数据集上的准确率最高;当scale值为0.5时CIFAR-10、CIFAR-100、Imagenette数据集上的准确率最高;当scale值为0.6时Imagewoof数据集上的准确率最高。该实验证明在不同数据集上scale的取值不同会对RDCNet的分类准确率有一定影响,但5个数据集中scale参数的最优取值都在0.5左右。

scale参数决定经过多分支处理的特征图和原始路径的输出在最终结果中的相对重要性。实验表明,scale参数的最优取值适中会显著提高RDCNet

表3 scale值对MRDC模块性能的影响
Table 3 Impact of scale value on MRDC module performance

scale	分类准确率/%				
	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
0.1	96.09	80.30	97.23	92.25	84.35
0.2	96.16	80.78	97.25	92.15	84.19
0.3	96.22	81.09	97.26	91.90	83.76
0.4	96.27	80.95	97.41	92.23	84.91
0.5	96.38	81.16	97.39	92.58	85.09
0.6	95.89	80.84	97.34	92.48	85.14
0.7	95.78	80.68	97.38	92.20	84.58
0.8	95.86	80.31	97.29	92.03	84.68
0.9	95.72	79.70	97.18	91.54	83.89
1	95.75	80.01	97.13	91.46	84.34

注:加粗字体表示各列最优结果。

的分类准确率。scale值过小时,网络会更依赖于原始输入特征,这些特征相对简单且不够丰富,无法有效捕捉复杂的上下文信息,同时会引起梯度流失或无法有效传递重要特征信息,从而影响模型的表达能力;scale值过大时,网络会过度依赖较为复杂的语义特征,增加对噪声或不相关信息的敏感性,从而导致过拟合,使分类准确率下降。如果scale设置适中,网络可以在简单特征和复杂特征之间取得平衡,避免两者都过于依赖,提供了一个较好的中间点,使得网络既能够保留低层次的特征信息,又能够有效地利用多分支提取的高级特征。较为适中的scale值能让模型在训练初期较为稳定,同时在后期逐渐优化特征的加权组合,从而避免了过于复杂的特征变换造成的训练困难和过拟合风险。

2.3 CE模块嵌入残差块的位置与数量对网络性能的影响

RDCNet的网络主干分为4个layer层,当CE模块放于不同层时,需要调节的特征图的内容和通道的依赖性也不同,某些层中强化特征的上下文信息可能产生更显著的效果。为深入研究CE模块嵌入RDCNet残差块的数量和位置对分类准确率的影响,将融合MRDC模块的layer层简称为Mlayer1、Mlayer2、Mlayer3、Mlayer4,在Mlayer中嵌入的CE模块简称为C-Block,在确定MRDC残差块最佳scale

参数值和最佳掩码策略的基础上设计了 8 种 CE 模块的嵌入方式,如图 8 所示。

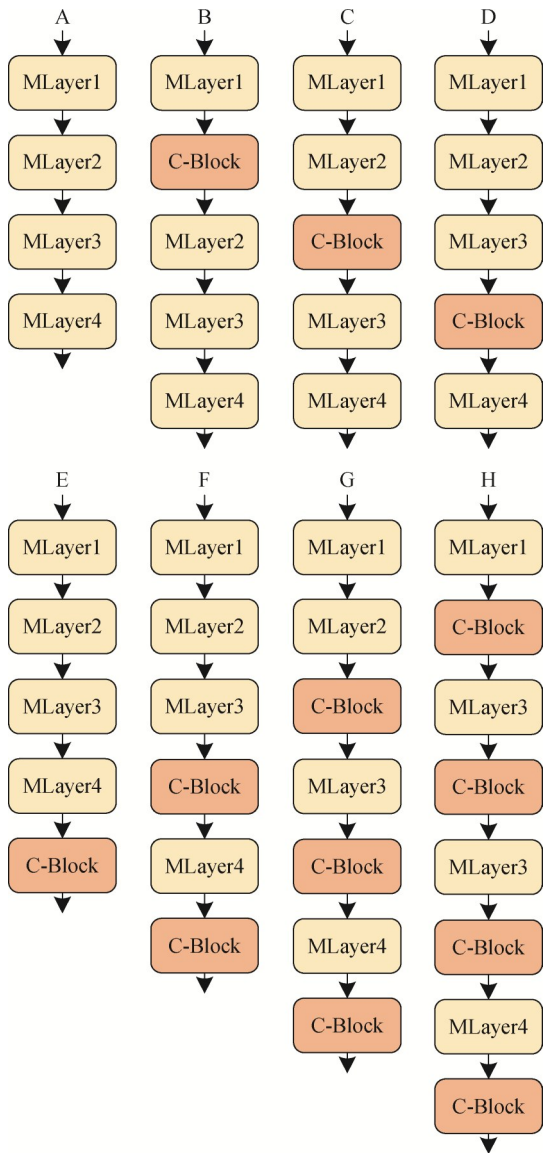


图 8 嵌入 C-Block 的 8 种排列方式

Fig. 8 Eight permutations embedded in C-Block

5 个数据集上不同位置和数量的 CE 模块的分类准确率如表 4 所示。由表 4 可得出,排列方式 E 在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Image-woof 数据集上的分类准确率均高于其他组合方式,因此将 CE 模块嵌入 Mlayer4 后的效果最显著。实验表明,将 CE 模块放在较低的网络层时,网络会基于较简单的特征来加权,但低层次的特征通常并未含有太多语义信息,因此提升效果不明显。而将 CE 模块放在较高的网络层时,特征已经更加抽象且语义信息丰富,特别在 Mlayer4 之后的特征图包含了物体

的高级语义信息。随着层数的增加,特征图的空间维度小、深度大,为 CE 模块提供了更多的上下文信息,能够更加有效地对高级语义特征进行处理和调整,从而显著提升模型的分类能力。

表 4 不同位置和数量 C-Block 对分类准确率的影响

Table 4 The effects of different locations and number of C-Blocks on classification accuracy

	/%				
方式	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
A	96.38	81.16	97.41	92.58	85.14
B	96.58	80.15	97.64	92.18	84.39
C	96.47	80.87	97.71	91.26	84.65
D	96.26	80.48	97.73	92.38	85.44
E	96.71	81.75	97.78	93.48	85.65
F	96.18	81.05	97.74	91.89	85.46
G	96.30	80.12	97.71	92.76	85.11
H	96.40	79.81	97.57	92.25	84.98

注:加粗字体表示各列最优结果。

2.4 CE 模块与其他模块的性能对比

为了进一步验证 CE 模块在提升 RDCNet 网络模型性能的有效性,本文选取了 4 种不同的模块,分别为 SE (squeeze-and-excitation)、NLB (non-local block)、CCA (criss-cross attention) 和 GCB (global context block) 与 CE 模块进行对比实验,分别将 4 种模块替换 CE 模块嵌入到 RDCNet 网络 Mlayer4 后,在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Image-woof 数据集上进行训练,训练过程与 3.1 节一致,实验结果如图 9 和表 5 所示。由实验结果可知,CE 模块在 5 个数据集上的分类准确率均为最高值,表明 CE 模块相比于其他模块能以更高效的方式捕捉全局和局部的上下文信息,同时避免计算瓶颈和过度依赖全局信息,提升分类性能。

SE 模块通过对每个通道特征进行自适应重标定实现通道之间的动态权重调整,但缺乏空间信息的建模。相比之下,CE 模块不仅考虑通道间的权重调整,还能够捕捉到图像中空间位置之间的相互关系,能够通过更复杂的上下文信息调整特征。

NLB 模块通过计算全局自注意力捕捉空间位置之间的依赖关系,对于长距离依赖关系表现良好,但计算每个像素和其他像素的相似性会导致高昂的计

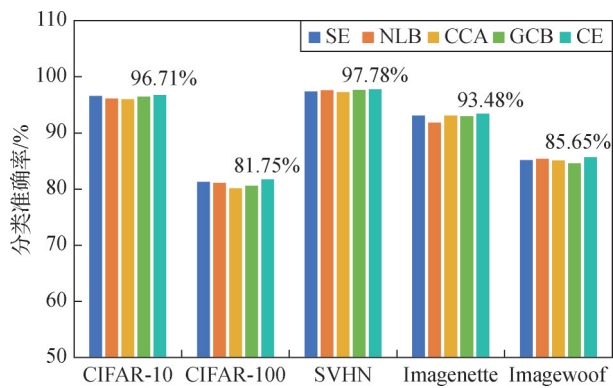


图9 5个模块在RDCNet的分类准确率
Fig. 9 Classification accuracy of 5 modules in RDCNet

表5 5种模块的性能对比
Table 5 Comparison of the performance of the five modules

模块	分类准确率/%				
	CIFAR-10	CIFAR-100	SVHN	Inette	Iwoof
+ SE	96.55	81.28	97.35	93.09	85.14
+ NLB	96.13	81.10	97.58	91.80	85.39
+ CCA	96.00	80.15	97.26	93.07	85.11
+ GCB	96.48	80.59	97.66	92.97	84.58
+ CE	96.71	81.75	97.78	93.48	85.65

注: Inette: Imagenette; Iwoof: Imagewoof。加粗字体表示各列最优结果。

算开销。CE模块通过简化的键值对机制和局部注意力机制增强模块的上下文感知能力,避免NLB模块中每个像素与其他像素之间的全连接操作,保持较高性能的同时也避免了NLB模块的高计算成本。

CCA模块利用水平和垂直方向的注意力机制捕捉特征图中不同空间位置之间的依赖关系。然而,CCA模块在捕获全局依赖关系方面存在一定的局限性。而CE模块不仅能关注局部的空间依赖,还能根据图像内容自适应地捕捉全局上下文信息,在捕捉长距离依赖时比CCA模块更有优势。

GCB模块通过全局池化操作提取图像的全局上下文信息,并利用这些信息对每个位置进行相应的调整。CE模块没有完全依赖全局上下文,而是通过上下文激励机制在通道和空间维度上动态调整信息,不仅能提取全局信息,也能兼顾局部信息的细节,从而避免GCB模块可能出现的过度全局化问题。

2.5 不同网络层数和学习率对网络性能的影响

为了验证不同的网络层数对RDCNet性能的影响,本文选择RDCNet18、RDCNet34、RDCNet50、RDCNet101 4种不同的网络层进行对比,选择CIFAR-100数据集并采用2.1小节中的实验环境进行复现。4种网络的分类准确率、F1-Score和X-entropy如表6所示,4种网络的分类准确率折线图如图10所示。

表6 4种网络的参数对比

Table 6 Comparison of parameters for four modules

网络	分类准确率/%	F1-score	X-entropy
RDCNet18	79.39	0.793 7	0.231
RDCNet34	81.75	0.817 2	0.151
RDCNet50	80.11	0.800 9	0.234
RDCNet101	79.43	0.794 0	0.230

注:加粗字体表示各列最优结果。

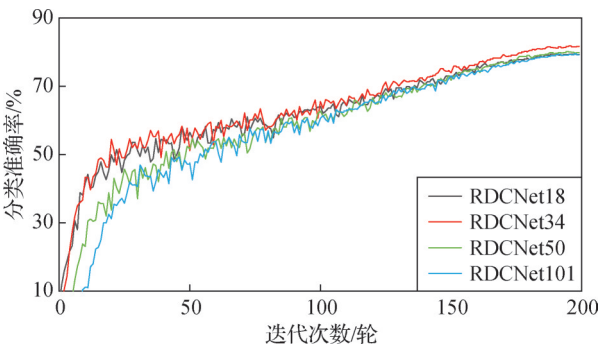


图10 网络层数不同对分类准确率的影响

Fig. 10 The impact of different network layers on classification accuracy

从表6和图10得知,在CIFAR-100数据集上,RDCNet在34层的网络层数(即RDCNet34)上取得最佳分类准确率,为81.75%,分别比RDCNet18、RDCNet50和RDCNet101高2.36%、1.64%和2.32%。实验表明,在特定的网络设计中,34层网络平衡了模型复杂度与学习能力,能够有效提取数据中的特征。18层的网络可能过于简单,尤其在更深层次特征提取的问题上不能充分学习复杂数据特征,导致其表现不佳。对于50层和101层的网络,尽管可以学习更多的特征,但增加的参数量使得模型的复杂度增加,会增大模型过拟合的几率,导致模型在测试集上的性能较训练集下降。因此RDCNet34代表了最佳的网络设计层数。

模型的训练过程及其最终的分类性能都会受到学习率的影响。本文实验的学习率调度器选择余弦退火策略,在训练后期逐渐减小学习率使模型能够在接近最优解时精细地调整参数。为了探究初始学习率的差异在余弦退火策略中对网络性能的影响,选择在 CIFAR-10、CIFAR-100 和 SVHN 数据集上进行对比实验,将初始学习率 $lr_1 \sim lr_5$ 分别设置为 0.1、0.05、0.01、0.005 和 0.001,不同初始学习率对分类准确率的影响如图 11 所示。由图 11 可知,当初始学习率为 0.1 时,在 CIFAR-10、CIFAR-100、SVHN 数据集的分类准确率均为最优值,表明采用余弦退火策略时,初始学习率设置为 0.1 时达到最佳效果。

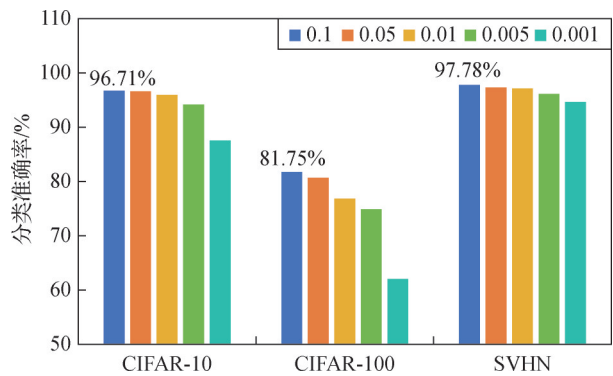


图 11 初始学习率对分类准确率的影响

Fig. 11 The effect of different initial learning rates on classification accuracy

当学习率为 0.1 时,模型初期能够快速适应数据特征,加速收敛并更快地找到最优解的方向,尤其在复杂任务中,较大步长能够使模型更快地跨越损失函数的多个局部最小值,接近全局最优解,后期通过余弦退火的减小策略能够进一步优化模型性能。而较小的初始学习率会导致模型收敛变慢,会限制模型在优化的过程步长的大小,可能导致参数停滞在某些局部最优解附近,无法有效跨越这些局部最优区域,从而导致模型性能下降。

2.6 对比实验

为进一步验证 RDCNet 的有效性,本文选择在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Imagewoof 数据集上进行对比实验,对比的分类网络包括 ResNet-34、HO-ResNet、WideResnet-28-10、CAPR-DenseNet、Multi-ResNet、EfficientNet (Tan 和 Le, 2020)、GhostNet (Han 等, 2020)、Couplformer (Lan 等, 2021)、FAVOR+ (fast attention via positive orthogonal

random features) (Choromanski 等, 2020)、MMA-CCT-7/3 $\times$ 2 (multi-manifold attention with cascaded cross-scale transformers) (Konstantinidis 等, 2023)、ATONet (auto-train-once network) (Wu 等, 2024)、QKFormer (query-key spiking transformer) (Zhou 等, 2024)、DAMSNet (double-branch multi-attention mechanism based sharpness-aware classification network) (姜文涛等, 2023) 和 SSCNet (sparse feature image classification network with spatial position correction) (姜文涛等, 2024)。

14 种不同分类方法数据来源有两种:1)未开源的代码优先采用原论文中分类准确率的最高值。2)开源的代码使用本文 2.1 小节的实验环境进行复现。表 7 列出 15 种网络在 5 个数据集上的分类准确率。通过对比实验结果,发现 RDCNet 在所有 5 个数据集上的分类准确率均表现最佳。

表 7 不同网络在 5 个数据集上的分类准确率

Table 7 Classification accuracy of different networks on 5 datasets

网络	/%				
	CIFAR-10	CIFAR-100	SVHN	Inette	Iwoof
ResNet-34	88.87	71.49	95.51	87.72	78.04
HO-ResNet	96.32	77.12	95.69	87.23	79.74
WideResnet-28-10	95.87	80.50	96.58	88.34	78.71
CAPR-DenseNet	94.24	78.85	94.97	87.72	80.78
Multi-ResNet	94.56	78.69	94.58	87.71	81.26
EfficientNet	94.02	75.98	93.34	88.06	77.95
GhostNet	94.92	77.17	93.86	87.83	78.35
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 $\times$ 2	94.74	77.51	94.29	-	-
ATONet	94.51	78.56	95.21	86.68	80.21
QKFormer	96.18	80.27	97.13	88.42	81.67
DAMSNet	96.51	80.50	97.60	-	-
SSCNet	96.69	80.63	97.43	88.75	82.09
RDCNet(本文)	96.71	81.75	97.78	93.48	85.65

注: Inette: Imagenette; Iwoof: Imagewoof。加粗字体表示各列最优结果。“-”表示相关工作未提供数据。

为了进一步验证 RDCNet 的分类性能,本文进行分类混淆矩阵实验反映 RDCNet 和 ResNet-34 在 CIFAR-10、Imagenette 数据集上的分类性能。如图 12 所示,通过对比混淆矩阵中的对角线数值表

明,RDCNet 相较于 ResNet-34 在两个数据集中表现更为良好,有着更多的正确分类样本数量和更少的错误分类样本数量,进一步验证了 RDCNet 在复杂视觉任务中的优势,表现出较强的分类性能。

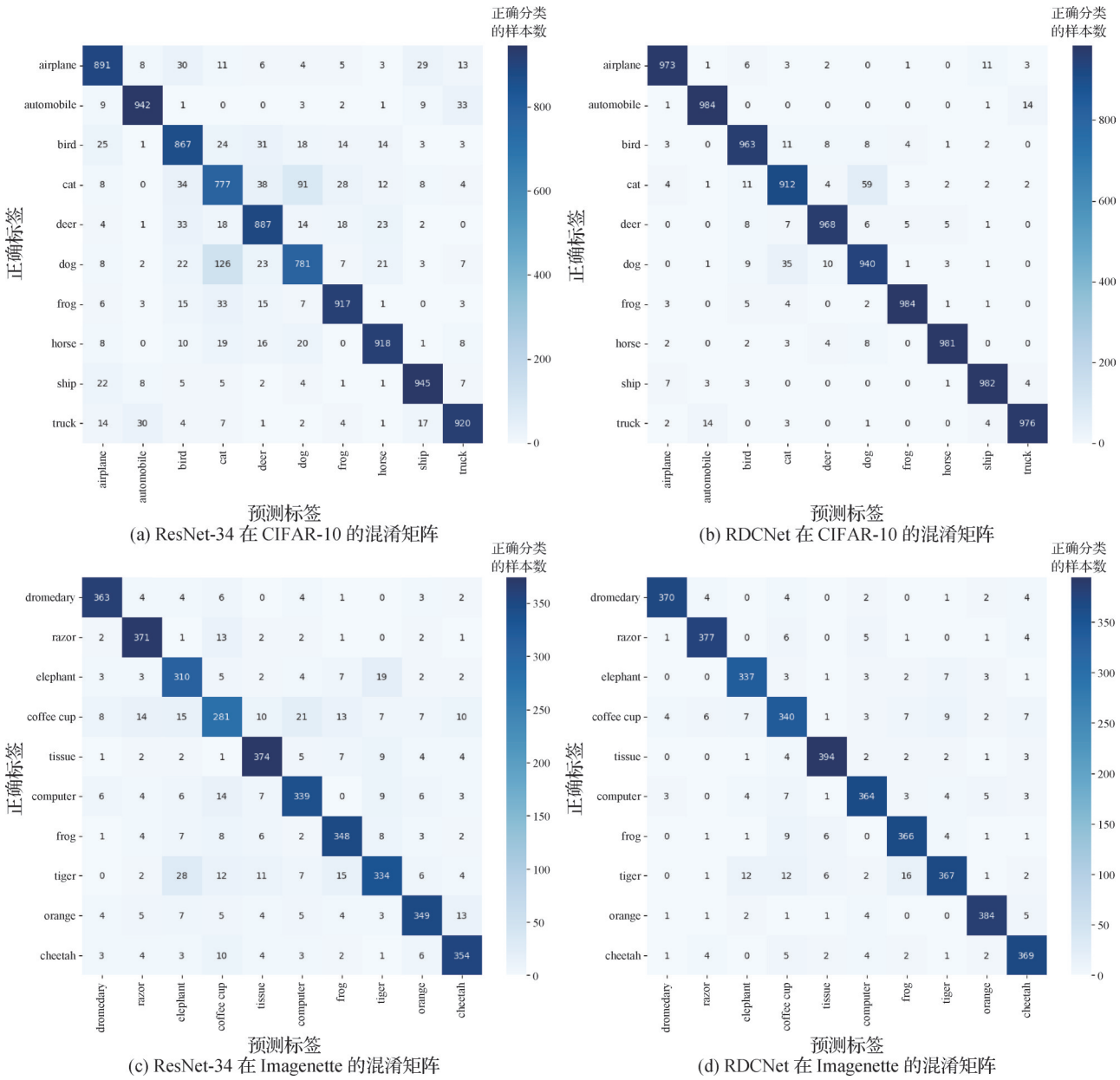


图 12 ResNet-34与 RDCNet 在 CIFAR-10 和 Imagenette 数据集上的混淆矩阵

Fig. 12 Confusion matrix of ResNet-34 and RDCNet on CIFAR-10 and Imagenette datasets

((a) ResNet-34 with CIFAR-10; (b) RDCNet with CIFAR-10; (c) ResNet-34 with Imagenette; (d) RDCNet with Imagenette)

2.7 消融实验

为了评估 RDCNet 网络中各模块的作用,进行多个消融实验分析各模块对 RDCNet 网络整体性能的贡献。共有 4 种消融方式:1)去除 RDCNet 网络中的 CE 模块,记为 Net₁;2)将 RDCNet 网络中 MRDC 模

块替换为 3 × 3 卷积,记为 Net₂;3)去除 RDCNet 网络中的 CE 模块,同时将 MRDC 模块替换为 3 × 3 卷积,记为 Net₃。4)在 RDCNet 中不修改初始卷积层普通卷积核的尺寸,保持 7 × 7 卷积核和最大池化层,记为 Net₄。表 8 展示了各网络在表 1 所列的 5 个数据集

上的分类准确率,图 13 呈现了消融实验中不同网络 分类准确率的变化趋势折线图。

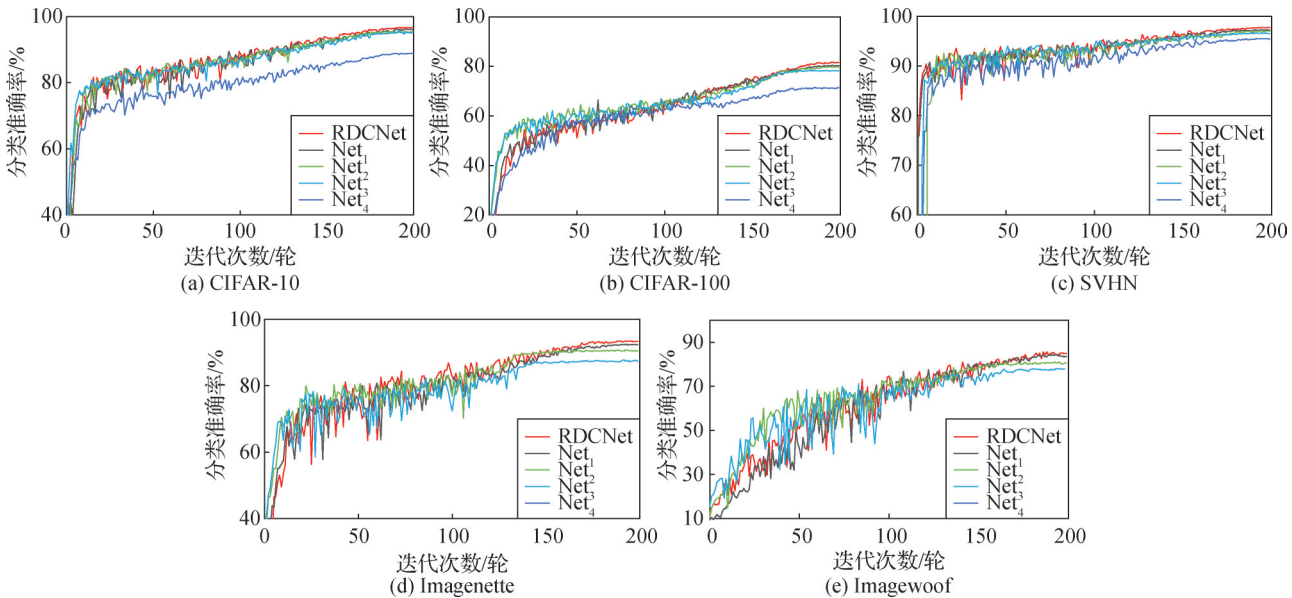


图 13 各个网络的分类准确率折线图

Fig. 13 Line graph of classification accuracy for each network
((a) CIFAR-10; (b) CIFAR-100; (c) SVHN; (d) Imagenette; (e) Imgewoof))

表 8 RDCNet 的消融实验结果
Table 8 Results of ablation experiments on RDCNet

网络	CIFAR-10	CIFAR-100	SVHN	Inette	Iwoof
RDCNet(本文)	96.71	81.75	97.78	93.48	85.65
Net ₁	96.38	81.16	97.41	92.58	85.14
Net ₂	95.46	79.78	97.03	90.75	81.06
Net ₃	95.23	78.35	96.65	—	—
Net ₄	88.87	71.49	95.51	87.72	78.04

注: Inette: Imagenette; Iwoof: Imgewoof。加粗字体表示各列最优结果。“—”表示相关工作未提供数据。

由表 8 和图 13 可得出 RDCNet 的分类准确率最优。去除 CE 模块后模型的准确率下降,验证了 CE 模块提取的上下文特征对最终预测结果至关重要。删除 MRDC 模块后模型的准确率显著下降,验证了 MRDC 模块的引入显著改善了模型的特征提取能力,通过多尺度与随机空洞卷积相结合,该模块使得模型在面对复杂数据时,能够更为有效和灵活地提取关键特征,进而提升整体的分类效果。此消融实验表明,在改变初始卷积核尺寸的基础上,将 Basic-Block 中的卷积替换为 MRDC 模块,同时结合 CE 模块,可以显著提高网络的分类性能。

3 结 论

针对传统图像分类网络模型在细粒度特征提取方面的挑战,以及背景噪声和不相关区域对目标特征学习的干扰问题,本文提出随机空洞卷积的图像分类网络(RDCNet)。其中多分支随机空洞卷积模块(MRDC)通过结合多空洞率的空洞卷积、掩码机制、多尺度特征融合以及细粒度特征增强模块(FGFE),有效扩大感受野,提升特征表达的多样性并增强信息融合的充分性,显著提高模型的分类性能。此外,提出上下文激励模块(CE),通过增强特

征图的多样性、提高信息融合能力、促进全局特征的交互,进一步提升图像分类的性能。综上所述,RDCNet通过结合MRDC模块和CE模块,能够更好地应对细粒度特征提取和背景噪声干扰问题,从而在图像分类任务中实现显著的性能提升。

RDCNet与原ResNet-34相比,在CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof数据集上的分类准确率分别提升7.84%、10.26%、2.27%、5.76%和7.61%。实验结果表明,RDCNet网络在多个图像分类任务中都取得了优异的性能,展示了其良好的泛化能力和强大的特征提取能力。

尽管RDCNet在图像分类的精度上取得了显著的提升,但在计算复杂度上仍存在瓶颈,有进一步提升的空间。特别是在多分支随机空洞卷积模块的实现上,在推理速度和内存占用方面可能成为限制因素。此外,虽然模型在特征提取方面表现突出,但在捕捉局部特征和全局上下文信息的平衡方面仍有改进的空间,特别是在背景干扰较强的图像中,模型可能难以有效地区分目标区域与噪声信息。针对这些问题,未来的工作将围绕优化计算效率、提高局部特征捕捉能力、加强信息融合以及增强跨任务适应性等方向展开,进一步提升模型在实际应用中的价值和性能。

参考文献(References)

- Abdi M and Nahavandi S. 2017. Multi-residual networks: improving the speed and accuracy of residual networks [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1609.05672.pdf>
- Cao Y, Xu J R, Lin S, Wei F Y and Hu H. 2019. GCNet: non-local networks meet squeeze-excitation networks and beyond//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE: 1971-1980 [DOI: 10.1109/ICCVW.2019.00246]
- Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Chen L W, Gu L, Zheng D Z and Fu Y. 2024. Frequency-adaptive dilated convolution for semantic segmentation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 3414-3425 [DOI: 10.1109/CVPR52733.2024.00328]
- Choromanski K, Likhoshesterov V, Dohan D, Song X Y, Kane A, Sarlos T, Hawkins P, Davis J, Mohiuddin A, Kaiser L, Belanger D, Colwell L and Weller A. 2020. Rethinking attention with performers [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2009.14794.pdf>
- Han K, Wang Y H, Tian Q, Guo J Y, Xu C J and Xu C. 2020. GhostNet: more features from cheap operations//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1577-1586 [DOI: 10.1109/cvpr42600.2020.00165]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1704.04861.pdf>
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Huang G, Zheng Y L, Liao K Y, Lin G F, Cao C J and Song X F. 2023. Mutual attention diversity feature fusion network-relevant fine-grained classification. Journal of Image and Graphics, 28(8): 2420-2431 (黄港, 郑元林, 廖开阳, 蔺广逢, 曹从军, 宋雪芳. 2023. 互补注意多样性特征融合网络的细粒度分类. 中国图象图形学报, 28(8): 2420-2431) [DOI: 10.11834/jig.220295]
- Huang Z L, Wang X G, Huang L C, Huang C, Wei Y C and Liu W Y. 2019. CCNet: criss-cross attention for semantic segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 603-612 [DOI: 10.1109/iccv.2019.00069]
- Jiang W T, Chen C and Zhang S C. 2024. Sparse feature image classification network with spatial position correction. Opto-Electronic Engineering, 51(5): #240050 (姜文涛, 陈晨, 张晟翀. 2024. 空间位置矫正的稀疏特征图像分类网络. 光电工程, 51(5): #240050) [DOI: 10.12086/oe.2024.240050]
- Jiang W T, Zhao L L and Tu C. 2023. Double-branch multi-attention mechanism based sharpness-aware classification network. Pattern Recognition and Artificial Intelligence, 36(3): 252-267. (姜文涛, 赵琳琳, 涂潮. 2023. 双分支多注意力机制的锐度感知分类网络. 模式识别与人工智能, 36(3): 252-267) [DOI: 10.16451/j.cnki.issn1003-6059.202303005]
- Jiang X Y, Zhang X H, Gao N and Deng Y. 2025. When fast fourier

- transform meets transformer for image restoration//Proceedings of 2025 European Conference on Computer Vision. Milan, Italy: Springer: 381-402 [DOI: 10.1007/978-3-031-72995-9_22]
- Konstantinidis D, Papastratis I, Dimitropoulos K and Daras P. 2023. Multi-manifold attention for vision transformers. *IEEE Access*, 11: 123433-423444 [DOI: 10.1109/ACCESS.2023.3329952]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lan H, Wang X H, Shen H, Liang P D and Wei X. 2021. Couplformer: rethinking vision transformer with coupling attention//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE [DOI: 10.1109/WACV56688.2023.00641]
- Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324 [DOI: 10.1109/5.726791]
- Liu S T, Huang D and Wang Y H. 2018. Receptive field block net for accurate and fast object detection//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 404-419 [DOI: 10.1007/978-3-030-01252-6_24]
- Luo Z B, Sun Z T, Zhou W L, Wu Z Z and Kamata S I. 2022. Rethinking ResNets: improved stacking strategies with high-order schemes for image classification. *Complex and Intelligent Systems*, 8(4): 3395-3407 [DOI: 10.1007/s40747-022-00671-3]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2024-12-12]. <http://arxiv.org/pdf/1409.1556.pdf>
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Tan M X and Le Q V. 2020. EfficientNet: rethinking model scaling for convolutional neural networks [EB/OL]. [2024-12-12]. <http://arxiv.org/pdf/1905.11946.pdf>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010 [DOI: 10.5555/3295222.3295349]
- Wang J, Yang Q P, Yang S Q, Chai X L and Zhang W J. 2022. Dual-path processing network for high-resolution salient object detection. *Applied Intelligence*, 52(10): 12034-12048 [DOI: 10.1007/s10489-021-02971-6]
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE [DOI: 10.1109/CVPR.2018.00813]
- Wu X D, Gao S Q, Zhang Z Y, Li Z Z, Bao R X, Zhang Y F, Wang X Q and Huang H. 2024. Auto-train-once: controller network guided automatic network pruning from scratch//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 16163-16173 [DOI: 10.1109/CVPR52733.2024.01530]
- Zagoruyko S and Komodakis N. 2017. Wide residual networks [EB/OL]. [2024-12-12]. <http://arxiv.org/pdf/1605.07146.pdf>
- Zhong Z, Zheng L, Kang G L, Li S Z and Yang Y. 2020. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7): 13001-13008 [DOI: 10.1609/aaai.v34i07.7000]
- Zhou C L, Zhang H, Zhou Z K, Yu L T, Huang L W, Fan X P, Yuan L, Ma Z Y, Zhou H H and Tian Y H. 2024. QKFormer: hierarchical spiking transformer using Q-K attention [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2403.16552.pdf>
- Zhu Z, Xu M D, Bai S, Huang T T and Bai X. 2019. Asymmetric non-local neural networks for semantic segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 593-602 [DOI: 10.1109/ICCV.2019.00068]

作者简介

姜文涛,男,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: jiangwentao@lntu.edu.cn

由卓丞,通信作者,男,硕士研究生,主要研究方向为图像处理、模式识别与人工智能。E-mail: 1046491150@qq.com

袁姮,女,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: yuanheng@lntu.edu.cn