



动态掩码卷积的图像分类网络

姜文涛, 由卓丞, 张晟翀

引用本文:

姜文涛, 由卓丞, 张晟翀. 动态掩码卷积的图像分类网络[J]. *智能系统学报*, 2026, 21(2): 423-434.

JIANG Wentao, YOU Zhuocheng, ZHANG Shengchong. Dynamic mask convolution for image classification networks[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(2): 423-434.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202503019>

您可能感兴趣的其他文章

基于孪生变分自编码器的小样本图像分类方法

A small-sample image classification method based on a Siamese variational auto-encoder

智能系统学报. 2021, 16(2): 254-262 <https://dx.doi.org/10.11992/tis.201906022>

改进MobileNet的图像分类方法研究

Research on the improved image classification method of MobileNet

智能系统学报. 2021, 16(1): 11-20 <https://dx.doi.org/10.11992/tis.202012034>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740-749 <https://dx.doi.org/10.11992/tis.201910039>

基于竞争性协同表示的局部判别投影特征提取

Competitive collaborative representation-based local discriminant projection for feature extraction

智能系统学报. 2019, 14(5): 974-981 <https://dx.doi.org/10.11992/tis.201809020>

基于卷积神经网络的遥感图像分类研究

Classification of remote-sensing image based on convolutional neural network

智能系统学报. 2018, 13(4): 550-556 <https://dx.doi.org/10.11992/tis.201706078>

一种基于联合表示的图像分类方法

Syncretic representation method for image classification

智能系统学报. 2018, 13(2): 220-226 <https://dx.doi.org/10.11992/tis.201611036>

DOI: 10.11992/tis.202503019

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20260203.1747.004>

动态掩码卷积的图像分类网络

姜文涛¹, 由卓丞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: 针对复杂场景下传统图像分类方法存在的特征适应性弱、多尺度信息捕捉能力有限以及细节特征表达能力不足的问题, 提出了一种基于动态掩码卷积的图像分类网络。1) 设计多分支掩码卷积融合模块, 将多分支结构与动态掩码机制相结合, 以实现不同尺度信息的融合, 并根据输入图像的上下文信息动态选择和强化关键特征, 从而提升网络的特征提取能力。2) 在残差学习中引入自适应增强模块, 通过整合像素级与通道级注意力机制自适应调整特征权重, 精准地捕捉图像中重要的细节信息。在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Imagewoof 数据集上的实验, 分别达到了 96.85%、82.39%、97.88%、93.35%、85.93% 的分类准确率, 显著优于传统图像分类方法, 该网络能够在面对多样化的图像特征和复杂的场景时, 表现出优异和稳定的分类性能, 为深度学习在图像分类领域的应用提供了新的思路。

关键词: 图像分类; 掩码机制; 残差网络; 动态掩码卷积; 膨胀卷积; 注意力机制; 特征融合; 特征提取

中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2026)02-0423-12

中文引用格式: 姜文涛, 由卓丞, 张晟翀. 动态掩码卷积的图像分类网络 [J]. 智能系统学报, 2026, 21(2): 423-434.

英文引用格式: JIANG Wentao, YOU Zhuocheng, ZHANG Shengchong. Dynamic mask convolution for image classification networks[J]. CAAI transactions on intelligent systems, 2026, 21(2): 423-434.

Dynamic mask convolution for image classification networks

JIANG Wentao¹, YOU Zhuocheng¹, ZHANG Shengchong²

(1. College of Software, Liaoning Technology University, Huludao 125105, China; 2. Science and Technology on Electro-Optical Information Security Control Laboratory, Tianjin 300308, China)

Abstract: Aiming at the problems of traditional image classification methods in complex scenes, such as weak feature adaptability, limited ability to capture multi-scale information, and insufficient ability to express detailed features, an image classification network based on dynamic mask convolution is proposed. Firstly, the multi-branch mask convolution fusion module is designed, which combines the multi-branch structure with the dynamic mask mechanism to realize the fusion of different scale information, and dynamically selects and strengthens the key features according to the context information of the input image, so as to improve the feature extraction ability of the network. Secondly, the adaptive enhancement module is introduced in the residual learning, and the feature weights are adaptively adjusted by integrating the pixel-level and channel level attention mechanisms to accurately capture the important details in the image. Through experiments on CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewoof datasets, the classification accuracy of 96.85%, 82.39%, 97.88%, 93.35% and 85.93% respectively, which is significantly better than the traditional image classification methods. The network can show excellent and stable classification performance in the face of diverse image features and complex scenes, and provides a new idea for the application of deep learning in the field of image classification.

Keywords: image classification; masking mechanism; residual networks; dynamic mask convolution; dilated convolution; attention mechanism; feature fusion; feature extraction

收稿日期: 2025-03-12. 网络出版日期: 2026-02-04.

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).

通信作者: 姜文涛. E-mail: Intuwulue@163.com.

传统的图像分类方法依赖于手工提取特征, 并使用经典的机器学习算法进行分类, 随着深度学习技术的发展, 卷积神经网络 (convolutional

neural networks, CNN) 的应用在图像分类任务中取得了显著的性能提升, 经典图像分类网络如 AlexNet (alexnet convolutional neural network)^[1]、VGG (visual geometry group)^[2] 和 GoogLeNet (google inception network)^[3] 等通过堆叠多个卷积层和池化层, 能够有效地提取图像的特征, 特别是 Res-Net (residual network)^[4] 的提出, 通过引入残差连接有效解决了神经网络中的梯度消失问题, 为深度学习的发展奠定了基础。然而, 传统卷积操作往往受限于通过增加网络深度来扩大感受野, 不仅会消耗大量的计算资源, 而且难以有效提取局部特征, 尤其在捕捉长程依赖方面表现不足。

为解决上述问题, Zagoruyko 等^[5] 提出 WRN (wide residual networks) 通过增加网络中每一层的宽度来增强网络性能; Abdi 等^[6] 提出 MSRN (multi-scale residual network) 通过引入多分支结构, 在不同尺度上提取特征, 增强网络的特征表达能力; 为了同时兼顾计算效率和性能, Wang 等^[7] 提出 LSN (large-small network) 通过结合大核感知和小核聚合的 LS (large-small) 卷积, 高效捕获全局信息并实现精细特征聚合; Yang 等^[8] 提出风车形卷积 PConv (pinwheel-shaped convolution) 突破传统标准卷积的对称性限制, 能更精确捕捉小目标的边缘和纹理特征并扩大感受野; Tan 等^[9] 在 EfficientNetV2 中提出复合缩放策略, 通过在深度、宽度和分辨率进行精细的缩放, 能够在保证高效性的同时实现更高的精度。Yu 等^[10] 在 ConvNeXt (convolutional network next)^[11] 的基础上提出 InceptionNeXt (inception depthwise convolution), 通过结合多分支结构大核卷积允许网络在不增加计算量的情况下在多个尺度上捕捉更丰富的特征信息。Luo 等^[12] 提出 HO-ResNet (higher-order residual network) 在传统的残差网络的基础上引入高阶残差模块, 使网络更好地捕捉到特征之间的复杂关系。许新征等^[13] 提出轻量化 VGG16 模型利用多分支结构和征膨胀卷积模块利用特征复用思想复用模型中的特征。

对于深层网络难以捕捉长程依赖的问题, Vaswani 等^[14] 提出 Transformer 架构通过自注意力机制能够更好地捕捉全局特征和长距离依赖; Dai 等^[15] 提出的 CoAtNet (convolutional attention network) 同时利用卷积和自注意力机制的优势高效地捕捉图像中的局部和全局信息; 为了更加高效地捕捉图像中的局部和全局信息, Cao 等^[16] 提出 GCNet (global context network) 引入全局上下文

注意力机制, 有效地捕捉图像中的局部和全局信息, 具有较强的自适应性; 赵凤等^[17] 提出 CTCMN (cross temporal coherence mining network) 通过融合多尺度 CNN 与 Transformer 实现跨层多尺度局部和全局特征信息的有效提取; Wu 等^[18] 在 DSwinIR (degradation-aware swin image restoration network) 中提出可变形滑动窗口注意力并引入内容自适应的动态感受野, 解决了边界信息割裂的问题, 显著提升了特征交互的连贯性; 刘万军等^[19] 提出 FFRMA (fast fourier residual multi-scale attention network) 将特征增强的前景目标与区域掩码自注意力网络相结合学习多样化的特征信息; Kang 等^[20] 提出 CPAM (cross-layer partial attention module) 引入一种不需要降维的注意力机制, 保持通道间的依赖建模的同时, 能够避免在空间维度上引入过多的计算, 有效地提升了模型的效率。在 MixDehazeNet (mix-strategy dehazing network) 中, Lu 等^[21] 提出像素注意力机制来强调图像中的重要部分, 更好地处理图像中的局部细节。

受上述思想启发, 本文提出动态掩码卷积的图像分类网络 (dynamic mask convolution for image classification networks, DMCNet), 主要创新如下: 1) 设计多分支掩码卷积融合 (multi-branch mask convolutional fusion, MMCF) 模块, 该模块并行多分支结构从多个角度处理输入图像, 每个分支采用不同的卷积核尺寸和掩码策略进行特征提取, 增强网络对多尺度信息的感知能力和对复杂内容的适应性。2) MMCF 中设计动态掩码卷积 (dynamic mask convolution, D_M_Conv), 其动态掩码机制不仅能够调整通道和卷积核的权重, 还可以根据上下文特征生成条件掩码, 对特征图进行精确的处理。通过动态掩码机制, 网络能够提升特征表达的多样性, 增强特征提取能力。3) 设计自适应增强 (adaptive enhancement, AE) 模块, 该模块协同优化了像素级与通道级注意力机制, 像素注意力机制关注局部的关键区域, 通道注意力机制则自适应调整通道的权重。AE 自动聚焦于关键区域并调整特征通道的权重, 在局部细节与全局语义之间实现更优的平衡, 显著提升了网络在复杂图像分类任务中的表现。

1 动态掩码卷积的图像分类网络

1.1 网络结构

DMCNet 的整体网络结构如图 1 所示, 主要由多分支掩码卷积融合模块和自适应增强模块组成, 整个网络被划分为 4 个阶段。

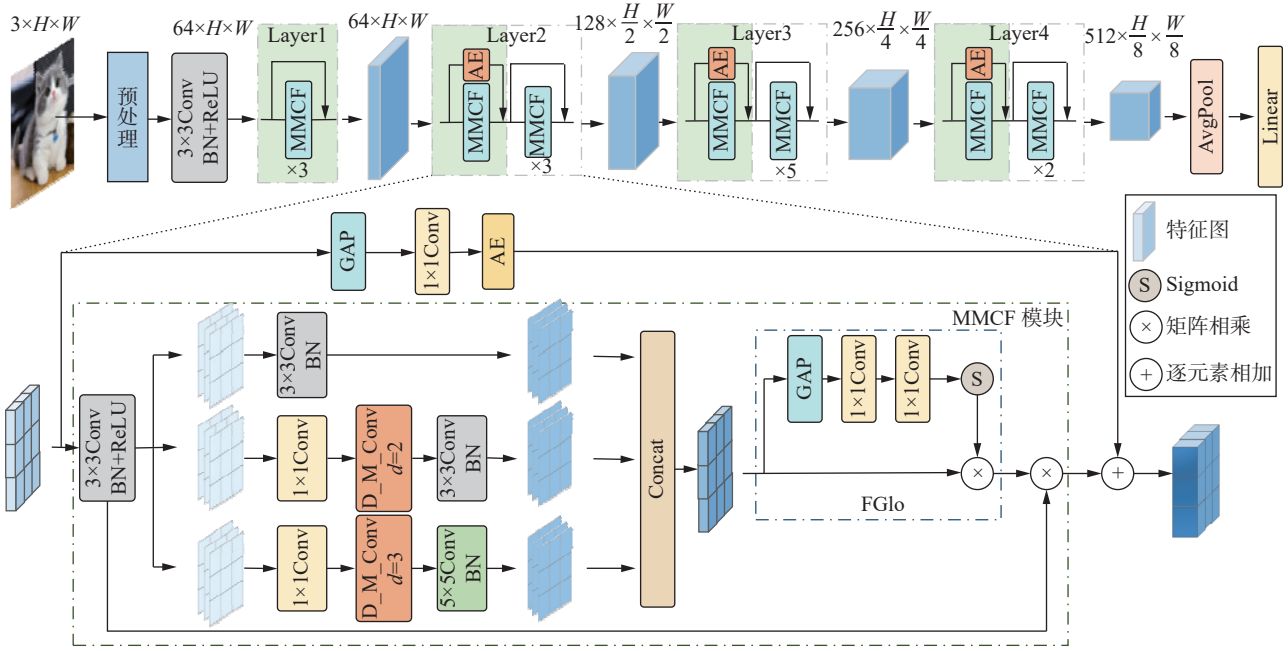


图 1 DMCNet 的整体网络结构

Fig. 1 Overall network structure of DMCNet

1) 预处理阶段。对输入图像进行图像增强和标准化操作, 通过图像增强 (如 AutoAugment^[22]、随机裁剪、随机翻转和 RandomErase^[23]), 在训练过程中接触多样性的样本, 在面对复杂的场景和数据时表现得更稳定, 有效地防止过拟合, 提高模型在未知数据上的泛化能力。标准化操作加快训练过程的收敛速度, 使每个特征通道的分布保持一致, 还能有效防止梯度消失问题。

2) 低级特征提取阶段。该阶段主要关注捕获图像中的低级特征, 为后续的深层特征提取提供丰富的低级信息。在小样本数据集 (CIFAR10、CIFAR100 和 SVHN) 中, 网络通过 3×3 卷积层和非线性激活函数 ReLU(rectified linear unit) 提取输入图像的基础特征, 而 Imagenette 和 Imagewoof 图像尺寸较大, 初级特征提取时使用 7×7 卷积层和最大池化操作, 可以更好地适应大尺寸图像。

3) 深层特征提取阶段。该阶段的网络由多个多分支掩码卷积融合模块和自适应增强模块组成, 通过多分支架构, 结合动态掩码卷积操作从多个尺度上提取特征, 使得网络能够聚焦于图像的重要部分, 有效地提取不同层次的语义信息。每个层由多个模块组成, Layer1 在保持低级特征信息的同时逐步加强特征的表达能力, Layer2 使网络继续学习更复杂的图像特征, Layer3 进一步加深网络结构, 通过多层次的卷积层提高特征提取的能力, Layer4 进一步细化特征提取以提取图像中的高级语义特征。这种层级结构使网络能够逐层从低级特征逐渐过渡到更高阶的语义特征,

从而使网络能够有效地增强对复杂图像的分类能力, 为后续的分类任务提供更加丰富的特征信息。

4) 分类阶段。经过多阶段特征提取后, 多个特征图会融合成一个全局特征表示, 其中包含了图像的丰富语义信息, 为分类提供了充分的支持。再通过全连接层将提取到的特征映射到输出的类别空间, 生成对应的预测结果。

1.2 残差结构

MMCF 模块的 3 种残差结构如图 2 所示, 其中, 图 2(a) 基础残差块能够有效地解决维度不匹配的问题, 其采用了一个简单的快捷连接 (shortcut), 在步长不为 1 或者输入输出通道数不相同的情况下, 通过卷积操作 (步长为 2) 调整输入的尺寸和通道数, 确保下采样时输入和输出之间的维度一致。但步长为 2 的卷积会导致空间分辨率的减小, 可能丧失某些细节信息, 尤其在下采样时卷积层的参数量大大增加, 会增加模型的计算复杂度。

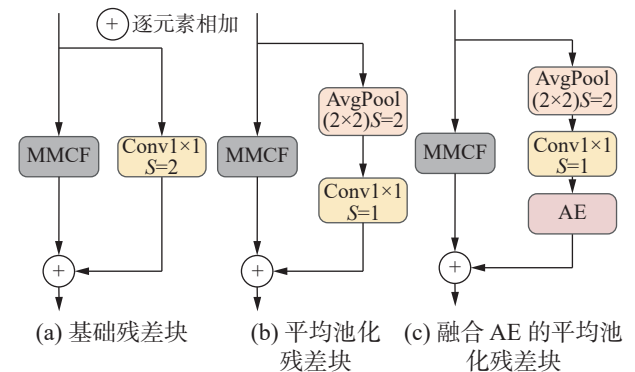


图 2 MMCF 的 3 种残差结构

Fig. 2 3 residual structures for MMCF

图 2(b) 平均池化残差块则在快捷连接部分加入一个平均池化层代替卷积进行下采样操作, 同时卷积操作的步长调整为 1, 通过对空间维度进行池化, 可以更好地保留全局信息。平均池化操作能够平滑特征图中的空间信息, 相较于卷积, 池化操作的计算开销较小, 同时避免基础残差块中的过度下采样问题, 但面对复杂的图像特征时, 平均池化的表达能力较弱, 可能无法捕捉到一些较细致的特征。

本文所使用的残差块如图 2 (c) 所示, 即在平均池化残差块的基础上加入 AE。AE 通过像素注意力和通道注意力机制使得模型能够更好地捕捉

复杂的空间和通道关系, 显著增强模型对复杂特征的学习能力, 帮助网络细致地捕捉图像中的重要信息, 避免信息丢失。与图 2(a)、(b) 相比, 融合 AE 的平均池化残差块在复杂任务中展现了显著的优势。

1.3 多分支掩码卷积融合模块

1.3.1 动态掩码机制

动态掩码卷积中的动态掩码机制如图 3 所示, 根据输入的特征和上下文信息对卷积核、输入特征进行动态调整, 使网络选择性地学习特征, 能够有效地增强网络对复杂场景的适应能力、提取多样化特征, 进而提高网络的泛化能力。

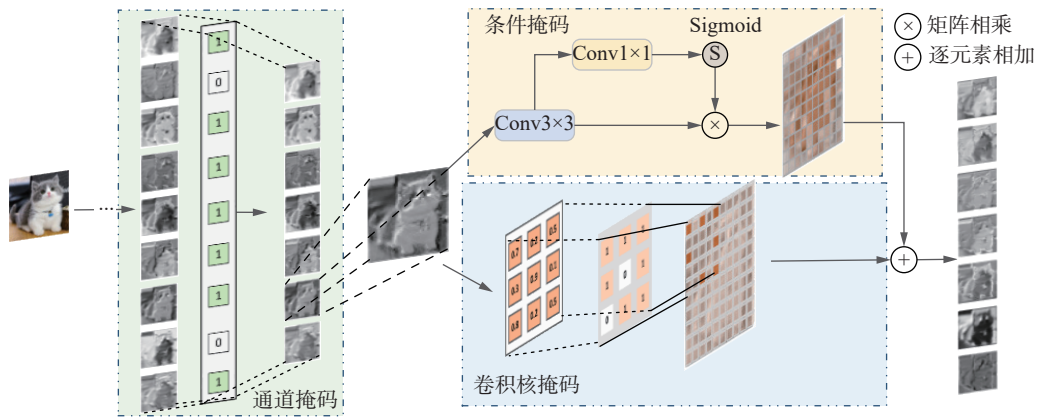


图 3 动态掩码机制

Fig. 3 Dynamic masking mechanism

动态掩码机制由 3 种掩码组成: 通道掩码、卷积核掩码和条件掩码。通道掩码用于在输入图像的通道维度上进行动态调整, 随机将部分输入通道的值设为零, 从而实现通道上的稀疏化处理。假设输入特征图 X 为 $C \times H \times W$, C 为通道数, H 和 W 分别为高度和宽度, 掩码中的元素为 rand 函数从均匀分布 $[0,1]$ 中生成随机数, 如果大于预设的阈值 m , 则返回 True 表示该通道在这一位置被激活, 否则返回 False 表示该通道被屏蔽, 然后将布尔类型的张量转换为浮点数类型, 从而得到一个二进制的掩码矩阵 M_1 , 在输入特征图 X 的每个通道上乘以掩码得到 X_{ch} 。其计算过程为

$$m_i = \begin{cases} \text{True}, & r_i < m \\ \text{False}, & r_i \geq m \end{cases}$$

$$X_{ch} = X \otimes M_1(\text{float}(m_i))$$

式中: r_i 为均匀分布中生成的随机数, M_1 表示掩码矩阵, m_i 为 M_1 中第 i 个元素, $\text{float}(\cdot)$ 为浮点数类型转换, \otimes 为按元素相乘。

该掩码使某些输入通道在向前传播的过程中不会对最终输出产生影响, 增强网络对不同通道特征的选择能力。

卷积核掩码在卷积操作中动态地调整卷积核权重, 该掩码作用于膨胀卷积, 其通过增加卷积核中元素间的距离扩大感受野, 设输入张量 $x \in \mathbf{R}^{h \times w}$, 卷积核尺寸为 k , 膨胀率为 d , 步幅为 s , 填充为 p , 则输出特征图的尺寸高 H 和宽 W 分别为

$$H = (h + 2p - d(k - 1) - 1) / s + 1$$

$$W = (w + 2p - d(k - 1) - 1) / s + 1$$

卷积核掩码与通道掩码相似, 最终会生成与卷积核权重的形状一致的掩码张量 M_2 , 将生成的掩码与卷积核的权重逐元素相乘得到特征图 X_{ke} , 其计算过程为

$$m_{c,h,w} = \begin{cases} \text{True}, & r_{c,h,w} < m \\ \text{False}, & r_{c,h,w} \geq m \end{cases}$$

$$X_{ke} = X_{ch} \otimes M_2(\text{float}(m_{c,h,w}))$$

式中: $r_{c,h,w}$ 为均匀分布中生成的随机数, M_2 为卷积核掩码矩阵, $m_{c,h,w}$ 为通道 c 中 M_2 第 h 行 w 列元素。

通过结合卷积核掩码, 可以有效增强卷积对上下文信息的感知能力, 卷积核掩码会遮蔽某些不重要的上下文信息以降低计算成本。

条件掩码的生成依赖于输入图像的具体内容

或上下文信息, 首先, 输入张量 X_{ch} 通过一个 3×3 卷积层提取上下文特征, 接着使用 1×1 卷积生成条件掩码, 使用 Sigmoid 激活函数将上下文特征映射到 $[0, 1]$ 范围内得到 M_3 , 这一过程可以表示为

$$M_3 = \text{Sigmoid}(\text{Conv}1 \times 1(\text{Conv}3 \times 3(X_{ch})))$$

式中: $\text{Conv}3 \times 3(\cdot)$ 为 3×3 卷积, $\text{Conv}1 \times 1(\cdot)$ 为 1×1 卷积, $\text{Sigmoid}(\cdot)$ 为 Sigmoid 激活函数, 其计算公式为

$$\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)}$$

式中 $\exp(-x)$ 为以 e 为底的指数函数的倒数形式。

特征图中物体的中心区域, 会需要更多的上下文信息, 而对于边缘或细节部分, 则需要较小的感受野进行细致的处理。最后, 将条件掩码 M_3 应用到通道掩码的输出 X_{ch} 上得到 X_{co} , 可以表示为

$$X_{co} = X_{ch} \otimes M_3$$

与完全随机生成的掩码不同, 条件掩码会根据特征图的上下文信息决定激活或抑制哪些区域, 这种方法与卷积操作中的膨胀率调整和感受野大小的自适应设计相吻合, 能根据不同的输入调整网络的响应, 更好地满足不同区域的需求。

最终, 将卷积核掩码的输出 X_{ke} 和条件掩码的输出 X_{co} 相加融合得到 X_{out} , 可以表示为

$$X_{out} = X_{co} + X_{ke}$$

1.3.2 MMCF 模块设计

多分支掩码卷积融合模块通过多分支结构和动态掩码卷积进行特征提取和融合, 从而有效地增强网络的特征表达能力。首先, 输入特征图 $F \in \mathbf{R}^{c \times h \times w}$ 通过卷积、批归一化和 ReLU 激活函数进行初步特征提取得到 F_0 , 可表示为

$$F_0 = \text{ReLU}(\text{BN}(\text{Conv}3 \times 3(F)))$$

式中: $\text{BN}(\cdot)$ 为批归一化, $\text{ReLU}(\cdot)$ 为 ReLU 激活函数, 其计算公式为

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

初步处理后将输出特征图通道分组, 共分为 3 组, 每个分支分别采用不同的操作, 表示为

$$F_0 = [F_1, F_2, F_3]$$

式中: F_1 、 F_2 、 F_3 分别表示 3 个分支的特征通道, 每组通道数为 $c/3$ 。

分支 1 负责提取图像的低层次、局部的基础特征, 输入 F_1 会经过一个标准卷积和批归一化操作, 得到特征图 F'_1 , 可以表示为

$$F'_1 = \text{BN}(\text{Conv}3 \times 3(F_1))$$

分支 2 负责捕获图像中的中等尺度的细节特征 (如物体的边缘、形状等), 动态掩码卷积的引入使模型能够选择性地关注图像中的重要区域, 而非无差别地处理所有区域, 从而提升模型在复杂场景下的敏感度, 尤其对于关键信息的提取。首先 F_2 通过 1×1 卷积进行通道数的压缩, 然后采用动态掩码卷积 (膨胀率为 2) 捕捉图像的中等尺度特征, 最后通过 3×3 卷积进一步提取信息输出 F'_2 , 可以表示为

$$F'_2 = \text{Conv}3 \times 3(\text{D_M_Conv}(\text{Conv}1 \times 1(F_2)))$$

式中 $\text{D_M_Conv}(\cdot)$ 为动态掩码卷积, 分支 2 能够在减少冗余计算的同时保持特征表达能力。

分支 3 通过结合动态掩码卷积 (膨胀率为 3) 和 5×5 卷积能够有效地捕获图像中远程依赖特征, 尤其适用于场景中较大、较远的背景信息。与分支 2 类似, 分支 3 首先通过 1×1 卷积压缩通道数, 再通过动态掩码卷积和 5×5 卷积来捕捉远程依赖特征得到最终的特征图 F'_3 , 可以表示为

$$F'_3 = \text{Conv}5 \times 5(\text{D_M_Conv}(\text{Conv}1 \times 1(F_3)))$$

式中 $\text{Conv}5 \times 5(\cdot)$ 为 5×5 卷积。

3 个分支协同工作能够有效增强模型的特征表达能力。随后, 将 3 个分支的输出在通道维度上拼接, 形成一个新的特征图 F_{cat} , 可以表示为

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

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

然后通过 FGlo (feature global) 模块对融合后的特征进行全局特征建模与增强, 作为一个全局特征加权模块, FGlo 模块使网络在多分支结构融合后的特征图中能够更加关注重要的特征通道, 减少冗余和无关信息, 有效地抑制不重要的特征, 提高网络的全局感知能力。

FGlo 模块首先对每个通道进行全局平均池化计算每个通道的全局信息, 接着进行卷积操作学习如何加权全局信息, 再通过 Sigmoid 函数得到每个通道的加权系数, 根据每个通道的贡献调整其重要性, 最终通过逐通道加权动态调整每个通道的输出强度, 这一过程可以表示为

$$F_T = \text{Sigmoid}(\text{Conv}1 \times 1(\text{Conv}1 \times 1(\text{GAP}(F_{cat}))))$$

式中: $\text{GAP}(\cdot)$ 表示全局平均池化。

最终融合部分, 输入特征图 F 会与经过加权后的特征图 F_T 进行加权求和, 得到最终的输出特征图 F_{out} , 可以表示为

$$F_{out} = s \times F_{cat} + F_{in}$$

式中: s 为缩放因子, 控制原始输入和处理后特征图之间的平衡。该模型 s 设为 0.5, 能够保留输入特征的细节信息, 而避免过度依赖卷积后的高级

特征,保证了深层网络梯度流动的顺畅性。

1.4 自适应增强模块

为提升模型的特征提取能力,解决在深层

网络中梯度消失问题,本文提出自适应增强(AE)模块并将其应用于残差连接中,其结构如图4所示。

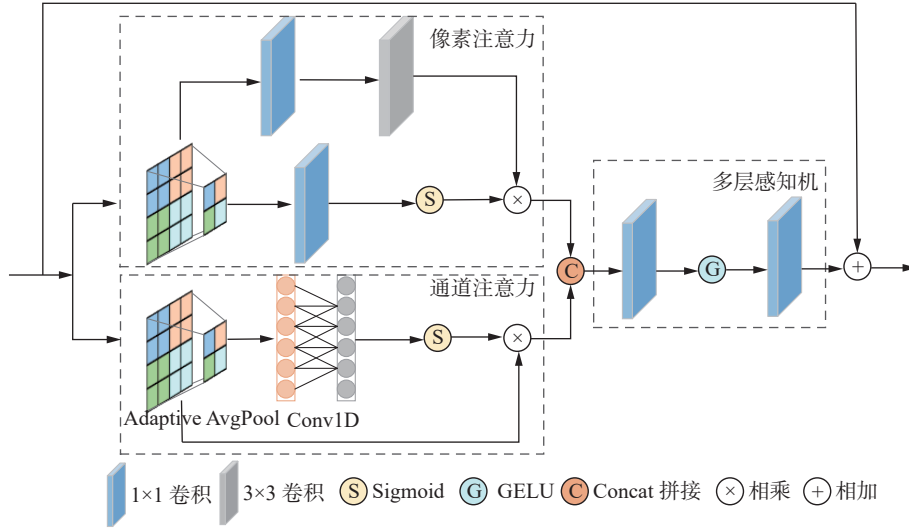


图 4 自适应增强模块结构

Fig. 4 Adaptive enhancement module structure

像素注意力机制通过对空间维度的加权,确保模型在空间维度上也能有效关注到关键位置。首先,通过卷积操作对输入特征图 $S \in R^{c \times h \times w}$ 进行处理,得到空间级别的加权特征 S_v ,可表示为

$$S_v = \text{Conv}3 \times 3(\text{Conv}1 \times 1(S))$$

接着,对输入特征图 S 进行自适应平均池化处理得到每个通道的全局信息征 S_p :

$$S_p = \text{AAP}(S) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W S_{c,h,w}$$

式中: $\text{AAP}(\cdot)$ 表示自适应平均池化操作, H 和 W 分别为特征图的高度和宽度, $S_{c,h,w}$ 表示特征图在 c 通道 (h, w) 位置处的值。

S_p 进行 1×1 卷积操作后,应用 Sigmoid 激活函数,生成每个像素位置的权重 W_p ,可表示为

$$W_p = \text{Sigmoid}(\text{Conv}1 \times 1(S_p))$$

最后,将像素权重 W_p 与空间加权后的特征 S_v 相乘,得到最终的空间增强特征图 S_{spa} ,可表示为

$$S_{spa} = S_v \otimes W_p$$

通道注意力机制可以自适应地调整特征图中每个通道的重要性,避免无关通道的干扰。首先对输入特征图 S 进行自适应平均池化操作,得到每个通道的全局信息后,进行 1D 卷积操作计算每个通道的权重,最后通过 Sigmoid 激活函数将卷积结果 W_a 压缩到 $[0, 1]$ 的范围内,可表示为

$$W_a = \text{Sigmoid}(\text{Conv}1d(\text{AAP}(S)))$$

式中 $\text{Conv}1d(\cdot)$ 为 1D 卷积,使用 $\text{Conv}1d$ 作为通道

注意力机制的卷积操作不仅可以有效地减少计算量,还可以更好地聚焦于通道之间的信息关系。

将通道权重 W_a 乘以原始输入 S 得到加权后的输出 S_{att} ,可表示为

$$S_{att} = S \otimes W_a$$

将通道注意力机制和像素注意力机制的输出结果进行拼接得到 S_{cat} ,可表示为

$$S_{cat} = \text{Concat}(S_{spa}, S_{att})$$

将拼接后的特征通过多层感知机进一步处理,由两层 1×1 卷积和激活函数 GELU(Gaussian error linear unit) 组成,可表示为

$$S_{mlp} = \text{Conv}1 \times 1(\text{GELU}(\text{Conv}1 \times 1(S_{cat})))$$

式中 $\text{GELU}(\cdot)$ 为 GELU 激活函数。

最后,通过残差连接将输入特征图 S 与 S_{mlp} 相加,得到最终的输出 S_{out} ,可表示为

$$S_{out} = S + S_{mlp}$$

在传统的残差连接结构中,特征信息经过多层卷积和激活函数的处理后会受到压制,当网络较深时会难以捕捉到有用的高级特征,因此,AE 能够有效增强网络在通道和空间维度上的特征表达能力,保证关键信息在网络中的有效传递。

2 实验结果与分析

2.1 实验配置

2.1.1 实验数据集与预处理

在本研究中,采用 CIFAR-10、CIFAR-100、SVHN 和 ImageNet 的子集 (Imagenette 和 Image-

woof) 数据集上进行评估, 5 种数据集的详细信息如表 1 所示。在数据预处理阶段, 为了增加数据的多样性, 防止模型过拟合, 对数据集使用数据增强, 包括 AutoAugment 和 RandomErase。AutoAugment 通过强化学习算法能够为给定的数据集自搜索最优的增强策略, 提升模型的泛化能力。RandomErase 随机擦除技术能够增强模型在面对部分丢失图像信息时的鲁棒性, 其中擦除概率 p 为 0.5。此外, 为了加速模型的训练, 对图像进行标准化处理。

表 1 实验数据集
Table 1 Experimental datasets

数据集	尺寸	类别数	训练集数	测试集数
CIFAR-10	32×32	10	50 000	10 000
CIFAR-100	32×32	100	50 000	10 000
Imagenette	224×224	10	9 469	3 925
Imagewoof	224×224	10	9 025	3 929
SVHN	32×32	10	73 257	26 032

2.1.2 实验环境及参数

本文实验的操作系统为 Ubuntu 22.04, 使用 PyTorch 作为深度学习框架, 版本为 PyTorch2.2.0。GPU 为 NVIDIA RTX 4090 拥有 60 GB 内存。

本文采用分类准确率作为模型性能的标准。模型训练中, 使用以下超参数设置确保模型的最佳性能, 其中学习率 (Learning Rate) 为 0.1, 批次大小 (Batch Size) 为 128, 训练轮数 (Epochs) 为 200 轮。采用动量为 0.9、权重衰减为 5×10^{-4} 的 SGD(stochastic gradient descent) 优化器。学习率调度器先采用 Warm-up 学习率预热线性升温学习率至 0.1, 剩余轮次中, 采用余弦退火策略 (CosineAnnealingLR)^[24] 调整学习率, 最小学习率设置为 1×10^{-6} 。

2.2 不同掩码组合对网络性能的影响

本实验中, 为了探究不同的掩码组合对网络分类性能的影响, 使用掩码卷积核掩码、通道掩码和条件掩码的不同组合分别在表 1 中的数据集上进行对比实验。设定了 8 个不同的掩码配置, 如表 2 所示, mask1 至 mask8 代表各种掩码的组合。8 个掩码组合在 5 个数据集上的对比实验结果如表 3 所示。实验结果表明, 当将 3 个掩码共同作用于网络时, 分类准确率达到最优值, 这表明适当的掩码机制不仅可以改善模型的抗干扰能力, 还能在不同层次上进行有效的特征选择和增强, 最终实现更高的分类准确率, 在多种数据集上的实验结果充分验证了这一组合策略的有效性。

表 2 不同掩码组合
Table 2 Different mask combinations

掩码	卷积核掩码	通道掩码	条件掩码
mask1	×	×	×
mask2	√	×	×
mask3	×	√	×
mask4	×	×	√
mask5	√	√	×
mask6	√	×	√
mask7	×	√	√
mask8	√	√	√

表 3 不同掩码组合对网络性能的实验对比
Table 3 Experimental comparison of different mask combinations on network performance %

掩码	CIFAR-10	CIFAR-100	SVHN	Image-nette	Image-woof
mask1	95.67	78.80	97.27	91.11	84.35
mask2	96.22	80.73	97.57	91.90	84.75
mask3	95.82	80.14	97.40	91.77	85.06
mask4	96.38	81.02	97.56	92.38	85.11
mask5	96.38	81.16	97.41	92.08	85.14
mask6	96.51	81.42	97.63	92.59	85.88
mask7	96.64	82.24	97.80	92.92	85.37
mask8	96.85	82.39	97.88	93.35	85.93

注: 加粗字体为每列最优值。

2.3 AE 不同位置对网络性能的影响

本实验研究了 AE 嵌入残差块的位置与数量对网络分类性能的影响。AE 使模型能够更加专注于重要特征的学习, 提升分类精度的同时弥补特征降维时损失的特征信息。不同残差块中 AE 的 8 种嵌入方式如图 5 所示, 其中嵌入 AE 的 DMC(dynamic mask convolution) 残差块简记为 DMCBLOCK_AE。AE 的 8 种嵌入方式的对比实验结果如表 4 所示, 方式 H 的分类准确率达到最高值, 因此将 AE 同时嵌入到 DMCBLOCK2、DMCBLOCK3 和 DMCBLOCK4 的残差连接中分类效果最佳。实验中, 单独将 AE 嵌入 DMCBLOCK2 时分类准确率较低, 原因在于该层主要负责提取图像的低级特征, 其局限性可能会限制 AE 模块中注意力机制的效果。将 AE 嵌入到 DMCBLOCK4 时分类性能提升, 虽然未达到最佳性能, 但可以证明 AE 处理高级特征中复杂的语义信息的效果更佳。当 AE 同时嵌入 DMCBLOCK2、3 和 4 时, 模型能够在多层自适应地优化特征, 增强低级到高级特征之间的相互作用, 显著提升模型的性能。

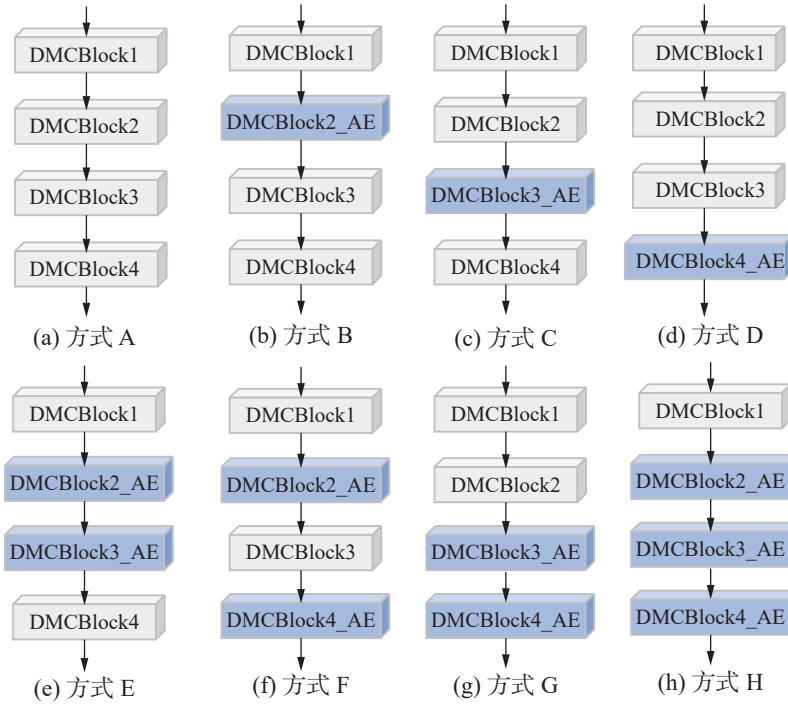


图 5 AE 的 8 种嵌入方式

Fig. 5 8 insertion methods for adaptive enhancement

表 4 残差块的不同嵌入方式对网络性能的影响

Table 4 Impact of different embedding methods for residual blocks on network performance %

嵌入方式	CIFAR-10	CIFAR-100	SVHN	Image-nette	Image-woof
A	96.56	81.83	97.28	92.46	85.29
B	96.35	80.93	97.47	92.15	85.11
C	96.43	80.94	97.62	92.48	85.26
D	96.59	81.46	97.60	93.07	85.44
E	95.93	81.06	97.63	92.97	85.37
F	96.28	82.04	97.67	93.25	85.54
G	96.60	82.32	97.75	93.22	85.75
H	96.85	82.39	97.88	93.35	85.93

注: 加粗字体为每列最优值。

2.4 超参数配置对网络性能的影响

2.4.1 不同网络层数对网络性能的影响

模型的深度直接影响特征表达能力和模型的学习效果, 为了探究不同网络层数对网络性能的具体影响, 表 5 总结了 4 个不同深度模型 的分类准确率和交叉熵损失值, 交叉熵定义公式为

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c})$$

式中: N 为样本数量, C 为类别总数, $y_{i,c}$ 为第 i 个样本的真实标签, $p_{i,c}$ 为模型第 i 个样本属于 c 类的概率。在 CIFAR-100 中的准确率折线图如图 6 所示。

表 5 不同网络层数的参数对比

Table 5 Comparison of parameters for different network layers

网络	深度/层	准确率/%	交叉熵
DMCNet18	18	80.84	0.2932
DMCNet34	34	82.39	0.2377
DMCNet50	50	80.45	0.2650
DMCNet101	101	79.43	0.2402

注: 加粗字体为每列最优值。

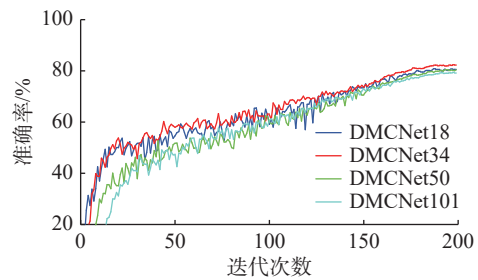


图 6 4 个网络的分类准确率变化趋势

Fig. 6 Trends in classification accuracy for 4 networks

从实验结果可知, DMCNet34 的分类准确率 达到最高值, 且交叉熵损失值最低, 表明 34 层 的网络深度有助于提升模型 的分类能力, 相比之下, 18 层的网络较浅, 可能难以有效学习复杂特征。随着网络深度进一步增加, DMCNet50 和 DMCNet101 的准确率逐渐下降, 表明单纯增加网络深度并不能保证性能的提升, 更深的网络可能会捕捉到训练数据中的噪声导致过拟合, 影响模型的泛化能力。

2.4.2 不同初始学习率对网络性能的影响

初始学习率的设置对于模型训练和性能表现起着关键的作用, 为探讨不同初始学习率对网络性能的影响, 5 种不同的初始学习率, 在 CIFAR-10、CIFAR-100 和 SVHN 数据集上训练 200 轮的结果如图 7 所示。

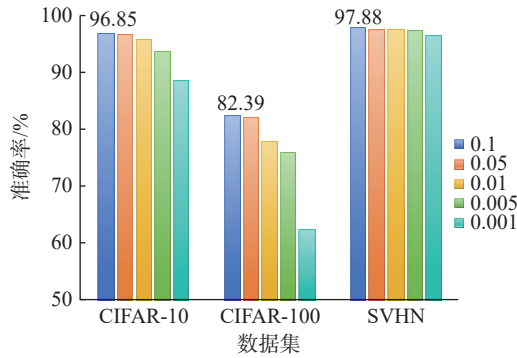


图 7 不同初始学习率的网络性能对比

Fig. 7 Comparison of network performance with different initial learning rates

由图 7 可知, 当初始学习率为 0.1 时分类准确率均达到最高值。证明较高的初始学习率 (如 0.1) 在保持收敛速度的同时能够有效避免过拟合, 尤其在训练的初期使模型能够获得强有力的梯度更新, 可以克服训练过程中的噪声和不确定因素, 充分利用输入数据中的特征信息, 还能够帮助模型在损失面中跳出局部最优点, 从而有机会找到更优解, 而较低的学习率 (如 0.001 或 0.005) 则可能导致模型在次优解附近徘徊, 难以达到全局最优的表现。

2.4.3 不同预热轮次对网络性能的影响

为了缓解高学习率带来的不稳定性, 本文通过学习率预热逐步提高学习率使模型在初期稳定

学习, 从而有效提高模型的收敛速度和准确率。为深入探讨不同预热轮次对网络分类性能的影响, 在 CIFAR-10、CIFAR-100 和 SVHN 数据集上进行对比实验, 如图 8 所示。

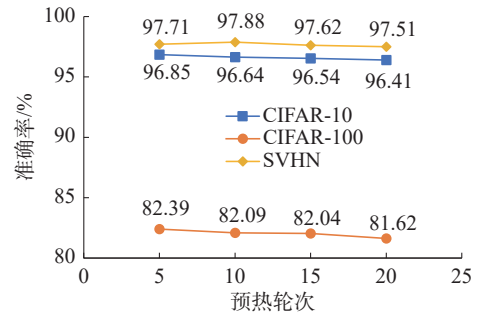


图 8 不同预热轮次的网络性能对比

Fig. 8 Comparison of network performance for different warm-up rounds

从图 8 可知, CIFAR-10/100 在预热轮次为 5 时分类准确率最高, 相较于 CIFAR 数据集, SVHN 数字识别数据集图像更简单且易于分类, 因此延长预热周期平稳地提升学习率, 可以获得更好的性能。当预热轮次为 15 或 20 时, 3 个数据集的分类准确率均出现下降趋势, 证明预热轮次过长可能会降低模型在初期阶段的学习效率, 对训练样本的关注度下降, 增加过拟合风险。预热期结束后采用余弦退火策略调整学习率, 如果预热期过长, 会导致初期学习率不能及时适应后续的训练, 影响收敛速度。因此, 合理设置预热轮次有助于模型训练的稳定性, 提升网络性能。

2.5 对比实验

为证明 DMCNet 的有效性, 表 6 给出了本文与一些先进网络在表 1 的 5 个数据集上的实验对比。

表 6 不同图像分类网络实验对比

Table 6 Experimental comparison of different image classification networks

网络	CIFAR-10/%	CIFAR-100/%	SVHN/%	Imagenette/%	Imagewoof/%	参数量/ 10^6	计算量/ 10^9
ResNet-34 ^[4]	88.87	71.49	95.79	87.82	78.44	21.34	3.68
WideResnet-28-10 ^[5]	95.87	80.50	96.58	88.34	78.71	36.54	5.96
MSRN ^[6]	94.65	78.68	94.58	87.69	81.26	15.80	1.85
GhostNet ^[25]	94.92	77.17	93.87	87.83	78.35	5.47	0.19
QKFormer ^[26]	96.21	80.29	97.15	88.43	81.75	64.96	1.29
AugLocal ^[27]	93.56	79.23	96.90	87.96	80.91	29.88	4.78
ATONet ^[28]	94.51	78.54	96.65	86.77	80.29	13.44	2.32
RTSA Net-101 ^[29]	96.12	81.60	96.67	—	—	49.92	8.24
SSCNet ^[30]	96.72	80.63	97.43	88.75	82.09	27.42	1.86
FDPRNet ^[31]	96.17	81.73	97.09	—	—	32.23	5.22
DMCNet(本文)	96.85	82.39	97.88	93.35	85.93	20.59	3.33

注: 加粗字体为每列准确率最优值, “—”表示未知结果。

主要网络模型包括: ResNet-34^[4]、WideResnet-28-10^[5]、MSRN^[6]、GhostNet^[25]、QKFormer^[26]、AugLocal^[27]、ATONet^[28]、RTSA Net-101^[29]、SSCNet^[30]和 FDPRNet^[31]。其中,对于未开源代码的网络,优先采用原论文中所提供的实验结果进行对比,对于已提供开源代码的网络,则使用 2.1 小节的实验环境复现来获得相应的数据。

与现有模型的性能对比,DMCNet 不仅在小尺寸图像任务上表现优异,在大尺寸图像任务中也有更出色的性能,大尺寸图像(如 Imagenette、Imagewoof)包含了更多的细节和更复杂的背景,而 DMCNet 能够从复杂的背景和细节中有效地提取图像中的细节特征,获得了较其他模型更高的分类准确率。因此,DMCNet 能够在复杂的视觉任务中有效处理图像中的细节信息,展

示了其在图像分类任务中的领先优势,证明了在复杂场景中的有效性。

2.6 消融实验

为了验证 DMCNet 中 MMCF 模块和 AE 模块对网络性能的作用,本文进行一系列消融对比实验。共有 4 种消融方式:1)删除 AE 模块,保留 MMCF 模块,记为 Net1。2)删除 MMCF 模块,保留 AE 模块,记为 Net2。3)删除 MMCF 模块和 AE 模块,记为 Net3。4)删除 MMCF 模块和 AE 模块的同时,将初始卷积层中的 3×3 卷积改为 7×7 卷积,记为 Net4。每种网络结构在表 1 中 5 个数据集上的表现如表 7 和图 9 所示,其中 Imagenette 和 Imagewoof 属于大尺寸样本,初始卷积层为 7×7 卷积,则未参与 Net3 消融实验。

表 7 消融对比实验
Table 7 Comparative ablation experiments

网络	CIFAR-10/%	CIFAR-100/%	SVHN/%	Imagenette/%	Imagewoof/%	参数量/10 ⁶	计算量/10 ⁹	训练时间/h
DMCNet	96.85	82.39	97.88	93.35	85.93	20.59	3.33	3.14
Net1	96.56	81.83	97.28	92.46	85.29	18.86	3.17	2.59
Net2	96.32	80.46	97.18	91.54	83.89	23.07	3.83	2.88
Net3	95.65	78.61	96.72	—	—	21.33	3.60	2.39
Net4	88.87	71.49	95.79	87.82	78.44	21.34	3.68	2.21

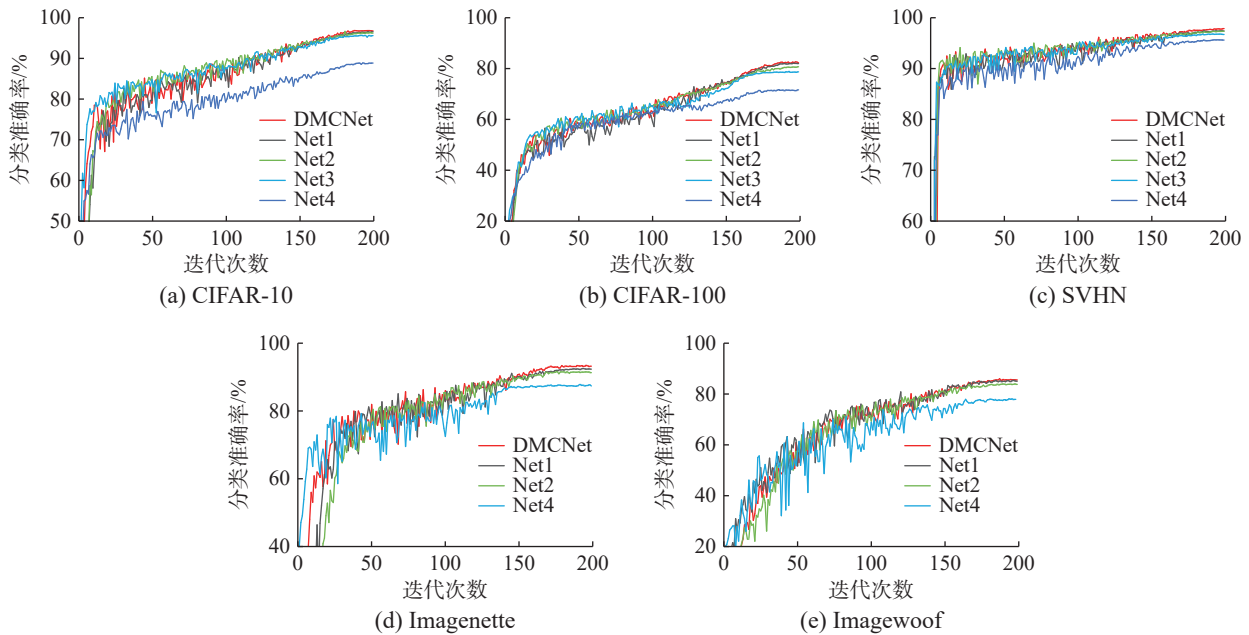


图 9 消融实验的分类准确率折线图

Fig. 9 Line graph of classification accuracy of ablation experiments

由实验结果可知,DMCNet 中 MMCF 模块和 AE 模块在提升网络性能上发挥了重要作用。Net1 在删除 AE 模块后性能略有下降,表明 AE 模块在

网络的特征学习中起到了积极的作用,能够帮助模型更好地捕捉关键特征并减少信息损失,从而提升了分类效果。Net2 在删除 MMCF 模块后性

能显著下降, 表明 MMCF 模块在 DMCNet 中是一个核心组件, 其作用在于能够更有效地整合和处理不同层次的特征信息, 从而提高了网络的整体性能。在 Net3 和 Net4 中, 将 MMCF 模块和 AE 模块同时删除后, 分类准确率大幅度降低, 进一步证明了两个模块在网络中的重要性, 其中 Net4 实验还得出在面对小尺寸样本数据集 (CIFAR-10、CIFAR-100 和 SVHN) 时, 使用较大的卷积核在处理特征图时会降低特征提取的精细度, 而下采样时过于依赖最大池化处理反而会限制对关键细节的捕捉。

3 结束语

本文提出了一种动态掩码卷积的图像分类网络。首先, 提出多分支掩码卷积融合模块, 其通过多分支结构结合动态掩码机制, 能够有效地提取不同尺度的特征, 增强网络在面对复杂图像时的鲁棒性。此外, 动态掩码机制的引入能够增强卷积操作的灵活性, 能够帮助网络自适应地处理特征图不同区域的特征信息, 避免冗余特征的影响, 使模型更关注于重要的部分。其次, 提出自适应增强模块, 将像素注意力机制与通道注意力机制结合, 动态调整不同区域和不同通道的特征权重, 不仅保持了计算效率, 还能够增强网络对细节信息的捕捉能力。本文提出的网络在 CIFAR-10、CIFAR-100、SVHN、Imagenette 和 Image-woof 图像数据集上都取得了优异的分类性能, 分类准确率分别为 96.85%、82.39%、97.88%、93.35%、85.93%, 与分类精度次优的模型对比, 分别提升 0.13、0.66、0.45、4.6、3.84 个百分点。实验结果证明了 DMCNet 的有效性。尽管 DMCNet 在图像分类任务中有着优异的表现, 但在处理高分辨率图像时, 其计算效率和计算开销上仍应进一步改进。因此, 要在保证精度的同时提升网络的效率, 如何优化模型结构是未来一个重要的研究和改进方向。此外, 本文提出的方法如何在其他领域中的应用需进一步探索, 可以在目标检测和图像分割等不同的任务中进一步评估模型的适应性。

参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 412–420.
- [3] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1–9.
- [4] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [5] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL]. (2016–05–23) [2025–03–12]. <https://arxiv.org/abs/1605.07146>.
- [6] ABDI M, NAHAVANDI S. Multi-residual networks: improving the speed and accuracy of residual networks [EB/OL]. (2016–09–19) [2025–03–12]. <https://arxiv.org/pdf/1609.05672.pdf>.
- [7] WANG Ao, CHEN Hui, LIN Zijia, et al. LSNet: see large, focus small[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025: 9718–9729.
- [8] YANG Jiangnan, LIU Shuangli, WU Jingjun, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2025, 39(9): 9202–9210.
- [9] TAN Mingxing, LE Q V. EfficientNetV2: smaller models and faster training[C]//International Conference on Machine Learning. Virtual: PMLR, 2021: 13–24.
- [10] YU Weihao, ZHOU Pan, YAN Shuicheng, et al. InceptionNeXt: when inception meets ConvNeXt[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 5672–5683.
- [11] LIU Zhuang, MAO Hanzhi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966–11976.
- [12] LUO Zhengbo, SUN Zitang, ZHOU Weilian, et al. Rethinking ResNets: improved stacking strategies with high-order schemes for image classification[J]. *Complex & intelligent systems*, 2022, 8(4): 3395–3407.
- [13] 许新征, 李杉. 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.
- [14] XU Xinzhen, LI Shan. Research of lightweight convolution neural network based on feature expansion convolution[J]. *Acta electronica sinica*, 2023, 51(2): 355–364.
- [15] 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] DAI Zihang, LIU Hanxiao, LE Q V, et al. CoAtNet: marrying convolution and attention for all data sizes[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2022, 44(9): 3201–3212.

- [16] CAO Yue, XU Jiarui, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 1971–1980.
- [17] 赵凤, 耿苗苗, 刘汉强, 等. 卷积神经网络与视觉 Transformer 联合驱动的跨层多尺度融合网络高光谱图像分类方法[J]. *电子与信息学报*, 2024, 46(5): 2237–2248.
ZHAO Feng, GENG Miaomiao, LIU Hanqiang, et al. Convolutional neural network and vision transformer-driven cross-layer multi-scale fusion network for hyperspectral image classification[J]. *Journal of electronics & information technology*, 2024, 46(5): 2237–2248.
- [18] WU Gang, JIANG Junjun, JIANG Kui, et al. DSwinIR: rethinking window-based attention for image restoration [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2025: 1–18.
- [19] 刘万军, 赵思琪, 曲海成, 等. 结合前景特征增强与区域掩码自注意力的细粒度图像分类[J]. *智能系统学报*, 2022, 17(6): 1134–1144.
LIU Wanjun, ZHAO Siqi, QU Haicheng, et al. Combining foreground feature reinforcement and region mask self-attention for fine-grained image classification[J]. *CAAI transactions on intelligent systems*, 2022, 17(6): 1134–1144.
- [20] KANG Ming, TING C M, TING F F, et al. ASF-YOLO: a novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. *Image and vision computing*, 2024, 147: 105057.
- [21] LU Liping, XIONG Qian, XU Bingrong, et al. MixDehazeNet: mix structure block for image dehazing network[C]//2024 International Joint Conference on Neural Networks. Yokohama: IEEE, 2024: 1–10.
- [22] CUBUK E D, ZOPH B, SHLENS J, et al. AutoAugment: learning augmentation policies from data[C]//International Conference on Machine Learning. Los Angeles: PMLR, 2019: 874–883.
- [23] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2020, 34(7): 13001–13008.
- [24] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[C]//International Conference on Learning Representations. Toulon: OpenReview.net, 2017: 1–16.
- [25] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 1577–1586.
- [26] ZHOU Chenlin, ZHANG Han, ZHOU Zhaokun, et al. QKFormer: query-key interaction for efficient vision Transformers[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1700–1709.
- [27] MA Chenxiang, WU Jibin, SI Chenyang, et al. Scaling supervised local learning with augmented auxiliary networks[C]//International conference on learning representations. Vienna: OpenReview.net, 2024: 1–18.
- [28] WU Xidong, GAO Shangqian, ZHANG Zeyu, et al. Auto-train-once: controller network guided automatic network pruning from scratch[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 16163–16173.
- [29] 邱云飞, 张家欣, 兰海, 等. 融合张量合成注意力的改进 ResNet 图像分类模型[J]. *激光与光电子学进展*, 2023, 60(6): 97–106.
QIU Yunfei, ZHANG Jiabin, LAN Hai, et al. Improved ResNet image classification model based on tensor synthesis attention[J]. *Laser & optoelectronics progress*, 2023, 60(6): 97–106.
- [30] 姜文涛, 陈晨, 张晟翀. 空间位置矫正的稀疏特征图像分类网络[J]. *光电工程*, 2024, 51(5): 240050.
JIANG Wentao, CHEN Chen, ZHANG Shengchong. Sparse feature image classification network with spatial position correction[J]. *Opto-electronic engineering*, 2024, 51(5): 240050.
- [31] 袁姮, 刘杰, 姜文涛, 等. 特征重排列注意力机制的双池化残差分类网络[J]. *中国图象图形学报*, 2025, 30(1): 110–129.
YUAN Heng, LIU Jie, JIANG Wentao, et al. Double-pooling residual classification network based on feature reordering attention mechanism[J]. *Journal of image and graphics*, 2025, 30(1): 110–129.

作者简介:



姜文涛, 副教授, 博士, 主要研究方向为图像与视觉信息计算。主持预研基金项目、辽宁省教育厅科学技术项目和辽宁省自然科学基金面上项目, 发表学术论文 35 篇。E-mail: Intuwulue@163.com。



由卓丞, 硕士, 主要研究方向为深度学习与图像处理、模式识别与人工智能。E-mail: 1046491150@qq.com。



张晟翀, 高级工程师, 硕士, 主要研究方向为数字信号处理, 发表学术论文 10 余篇。E-mail: zsc417@126.com。