

全局依赖引导的特征重构图像分类网络

袁 姮¹, 武景瑞^{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: 针对卷积神经网络在图像分类任务中存在的长距离依赖关系建模不足的问题, 本文提出全局依赖引导的特征重构图像分类网络(Global Dependency guided Feature Reconstruction for image classification, GDFRNet)。GDFRNet通过设计新颖的特征重构模块(Feature Reconstruction Module, FRM)和特征优化分支, 构建了一个协同工作的双路径架构, 实现了长距离依赖建模和细节特征增强。FRM通过引入并行的水平与垂直方向全局平均池化, 分别在两个空间维度上压缩特征并获取具有全局视野的上下文向量, 将其重新映射到二维特征空间, 从而建立起跨越图像全域的长距离特征依赖关系, 同时结合转置卷积等操作实现对特征空间的重构, 抑制无关背景噪声, 强化目标主体的连贯语义表达。特征优化分支通过设计细粒度特征捕捉模块(Fine-Grained feature Capture Module, FGCM)和特征优化模块(Feature Optimization Module, FOM)提炼并融合细节信息, 减少网络抽象过程中细节信息的损失。FGCM通过引入高斯-拉普拉斯卷积, 捕获图像中易于丢失的细节信息; FOM负责在高分辨率特征池中对FRM处理后的全局语义特征图与FGCM提取的丰富细节特征进行自适应融合与优化。两条路径形成了“全局轮廓-局部细节”互补的工作机制, FRM提供的全局语义特征图为细节增强提供指导, 确保细节强化不偏离整体语义。同时, 特征优化分支所提炼的丰富底层细节又为FRM的特征重构过程提供了必要的细微信息反馈, 形成了良性的优化循环。这种互补机制使网络最终融合了经过重构的语义信息与经过增强的局部细节信息, 生成判别性更强的图像表征, 整体上强化了模型对图像整体结构的理解能力, 并提升了特征空间的判别性。论文所提模型在CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof五个数据集上与当前最先进模型(State-Of-The-Art, SOTA)进行了对比实验, GDFRNet均表现出了卓越的性能, 与其他先进方法相比, GDFRNet在以上五个数据集上的分类精度依次平均提升了2.39%、3.73%、2.35%、3.33%和2.92%, 证明了GDFRNet的有效性和先进性。

关键词: 图像分类; 全局依赖; 特征重构; 细节增强; 特征优化; 卷积神经网络

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

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2026)01-0291-17

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20250619

Global Dependency Guided Feature Reconstruction for Image Classification

YUAN Heng¹, WU Jingrui^{1*}, ZHANG Shengchong²

(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: To address the inadequacy of modelling long-range dependencies in convolutional neural networks for image classification tasks, this paper proposes the global dependency guided feature reconstruction for image classification (GDFRNet). GDFRNet constructs a synergistic dual-path architecture through the design of a novel feature reconstruction module (FRM) and a feature optimization branch, achieving both long-range dependency modelling and fine-grained feature enhancement. The FRM introduces parallel horizontal and vertical global mean pooling to compress features across two spatial dimensions. This extracts context vectors with global vision, remapping them into a two-dimensional feature space to establish long-range feature dependencies spanning the entire image. Concurrently, operations such as transposed convolutions reconstruct the feature space, suppressing irrelevant background noise while reinforcing coherent semantic representations of the target subject. The feature optimization branch refines and fuses detail information through the fine-grained feature capture module (FGCM) and feature optimization module (FOM), reducing the loss of detail information during the network abstraction process. FGCM employs Gaussian-Laplacian convolution to focus on capturing easily lost fine details within images. The FOM performs adaptive fusion and optimization of the global semantic feature map provided by the

FRM and the rich detail features extracted by the FGCM within a high-resolution feature pool. These two pathways establish a complementary “global contour-local detail” working mechanism, the global semantic map provided by the FRM guides detail enhancement, ensuring that detail reinforcement does not deviate from the overall semantic context; simultaneously, the rich underlying details refined by the feature optimization branch provide essential fine-grained feedback and task-relevant guidance for FRM’s feature reconstruction process, establishing a virtuous optimization cycle. This complementary mechanism enables the network to ultimately fuse reconstructed semantic information with enhanced local detail, generating more discriminative image representations. This holistically strengthens the model’s understanding of overall image structure and significantly enhances the discriminative power of the feature space. Comparative experiments between the proposed model and state-of-the-art (SOTA) models were conducted across five benchmark datasets: CIFAR-10, CIFAR-100, SVHN, Imagenette and Imagewoof. GDFRNet demonstrated outstanding performance across all datasets. Compared with other advanced methods, GDFRNet achieved average improvements in classification accuracy of 2.39%, 3.73%, 2.35%, 3.33%, and 2.92% on the five datasets mentioned above, demonstrating the effectiveness and advancement of GDFRNet.

Keywords: image classification; global dependency; feature reconstruction; detail enhancements; feature optimization; CNN

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

0 引言

图像分类是计算机视觉的一项基本任务,是遥感土地覆盖分类、医疗影像分析、智能交通系统等应用的基本感知组件^[1]。卷积神经网络(Convolutional Neural Network, CNN)凭借局部连接、权值共享、端到端训练以及泛化能力强的优势在图像分类方法中占据了核心地位^[2]。随着网络深度的增加,梯度消失和梯度爆炸成为神经网络无法避免且亟待解决的问题,ResNet^[3]通过引入快捷连接有效缓解了这一问题,但其在图像特征的长程建模方面仍存在不足。在此基础上,文献[4-6]通过增加卷积层和激活函数,缓解了长距离依赖建模不足的问题。但是这些改进仍存在特征提取不充分、难以有效捕捉图像关键信息的问题。

为了克服上述问题,研究者们提出了一系列建模长程依赖关系的方法。文献[7]结合深度可分离卷积和SE(Squeeze-and-Excitation)注意力,增强通道间全局特征交互能力。文献[8]中改进的可变形卷积继承了卷积的局部稀疏性和平移不变性,具备一定的长距离依赖建模能力。Kim等人^[9]通过点插值实现参数高效的大核,平衡了长距离依赖建模与计算效率。此外,多尺度特征融合的方式也得到了广泛关注。文献[10]利用不同空洞率的空洞卷积提取多尺度特征信息,使上下文信息得到充分利用。文献[11]通过提取不同尺度的特征信息,获得图像的多层次特征,最大限度地提高了特征利用率。文献[12]将ResNet网络提取到的浅层特征在通道维度上均匀分组处理并拼接,增强了网络多尺度特征表示能力。文献[13]通过多尺度提取图像特征,并有效融合不同尺度的特征,提高网络的性能。

通过注意力机制建立长距离依赖的研究主要分为两类:一类是基于卷积的注意力机制,如文献[14-17]从通道和空间两个维度处理特征信息,增强特征表达能力,提升网络性能。另一类是结合Transformer的注意力机制,如文献[18-20]有效整合Transformer的全局信息捕获能力和CNN的局部特征提取能力,使网络更好地理解上下文信息。

上述方法虽然通过捕获特征长距离依赖,帮助网络专注关键特征,但是单分支模型表达能力有限,无法捕捉到复杂场景下的重要信息。为了提高网络对复杂场景中特征的提取能力,研究者们采用辅助分支结构捕获全局信息和局部信息,有效学习图像特征,提升网络性能。Shi等人^[21]结合辅助分支的思想,学习底层特征的同时通过先验信息引导模型关注任务相关的特定区域,提高了模型在多种视觉任务中的性能和泛化能力。文献[22]通过构建每个隐藏层的辅助网络增强局部学习的效果,帮助模型在局部学习的同时保持与全局学习的有效联系。文献[23]引入CTR3模块作为辅助分支学习显著特征,加强目标信息和位置感知的关联性,为模型提供归纳偏置,提高模型的泛化能力。

受上述两种思路的启发,针对卷积神经网络难以建模长距离特征依赖的瓶颈,本文提出了全局依赖引导的特征重构图像分类网络(Global Dependency guided Feature Reconstruction for image classification, GDFRNet)。

(1)为建立长距离特征依赖,提出了特征重构模块(Feature Reconstruction Module, FRM)。结合转置卷积与条带池化捕捉远距离特征间的关联性,重建特征空间,在减少冗余的同时实现了有效的全局依赖建模。

(2)为了高效学习上下文信息,将FRM与跳跃连接结合,构建特征重构残差块(FRM Residual Block, FRRB),以此堆叠形成网络主干,在保持特征表达能力的同时建模特征间的长依赖关系。

(3)为了使网络在解析全局依赖的同时,保持对细节信息的敏感度,提出了结合特征优化分支的策略。通过设计细粒度特征捕捉模块(Fine-Grained feature Capture Module, FGCM)和特征优化模块(Feature Optimization Module, FOM),在保持特征图高分辨率

的情况下增强细节信息,并根据像素的重要性有选择地融合全局信息和细节信息,与FRM形成“全局-局部”协同机制,进而增强表征学习能力。

1 全局依赖引导的特征重构图像分类网络

1.1 GDFRNet总体结构

GDFRNet主要包含两部分:由FRRB堆叠组成的网络主干和由FGCM、FOM组成的特征优化分支,网络整体结构如图1所示。

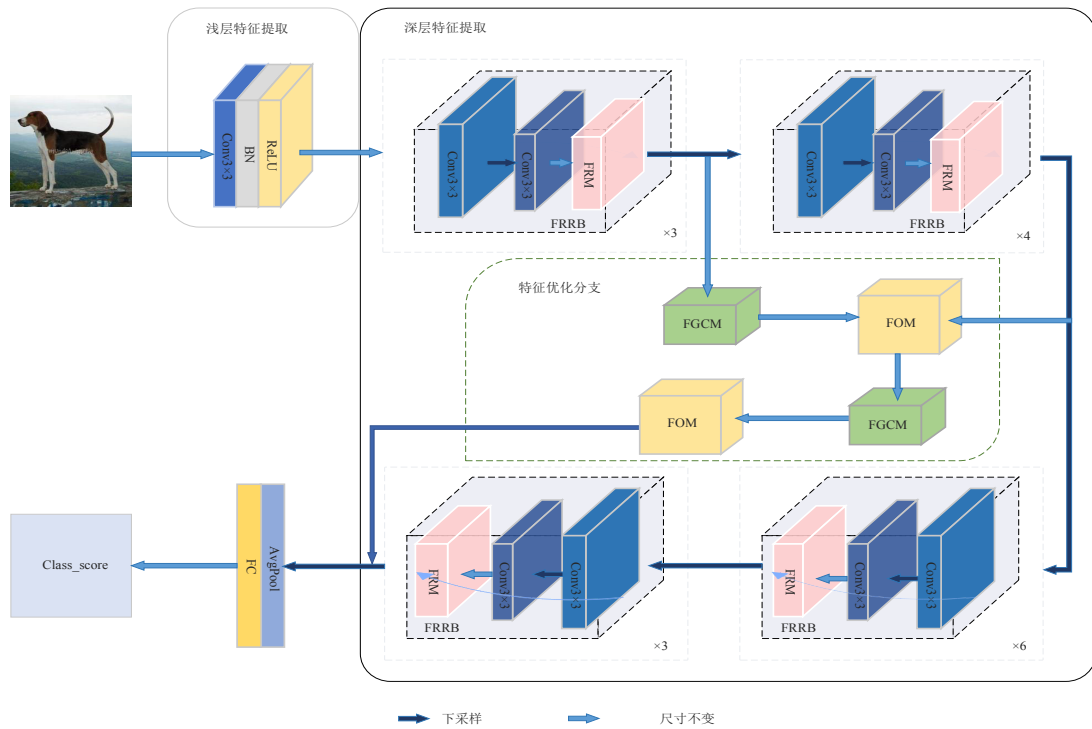


图1 GDFRNet框架图

Figure 1 Architecture of GDFRNet

GDFRNet模型的训练分为如下五个阶段:

(1)数据预处理。对输入图像进行数据增强预处理,结合填充图像、随机裁剪、水平翻转、归一化操作,改善图像质量,为网络特征提取做准备。

(2)浅层特征提取。在处理尺寸为 32×32 的图像时,采用 3×3 滤波器提取输入图像的特征,然后采用批归一化、ReLU激活函数对提取到的浅层特征进行调整,为后续深层网络学习做准备;处理尺寸为 256×256 的图像时,用 7×7 滤波器提取输入图像特征,接着用最大池化处理,提取图像中的显著特征。

(3)深层特征提取和增强。FRM结合跳跃连接,构成FRRB。将浅层特征输入FRRB中,解析特征之间的长距离依赖关系,重构特征空间,引导FGCM优先增强关键区域的边缘细节。

(4)特征优化。结合FGCM和FOM搭建特征优化

分支,与FRRB并行工作,在保持特征图高分辨率的同时增强细节信息,并将带有全局依赖的抽象特征根据其重要性融合到高分辨率特征图中,完善特征学习图谱的构建,使网络在学习全局关联特征的同时保留细节信息,进一步优化FRM的长距离建模精度,提升网络分类性能。

(5)输出。将学习到的特征输入自适应平均池化层下采样,经过全连接层计算分类得分结果。

1.2 全局依赖引导的特征重构机制

GDFRNet的核心贡献是提出了特征重构模块FRM,将转置卷积和深度卷积结合,以促进跨尺度特征交互,增强空间信息重构能力,同时引入方向感知池化策略,捕获方向敏感的长距离依赖,动态校准特征响应,共同完成跨尺度特征交互与全局上下文建模的联合优化,显著增强了网络的表征能力。

1.2.1 特征重构模块

现有的卷积神经网络通常依赖局部卷积运算,虽能有效捕捉短程空间依赖,但在建模长距离依赖关系和捕捉全局信息时存在局限性。为此设计了特征重构模块 FRM,通过方向感知的全局池化生成空间注意力权重,动态校准特征响应;同时结合深度卷积和转置卷积融合多分辨率信息,重构特征空间并有效建模长距离像素间的语义关联。如图 2 所示,FRM 主要

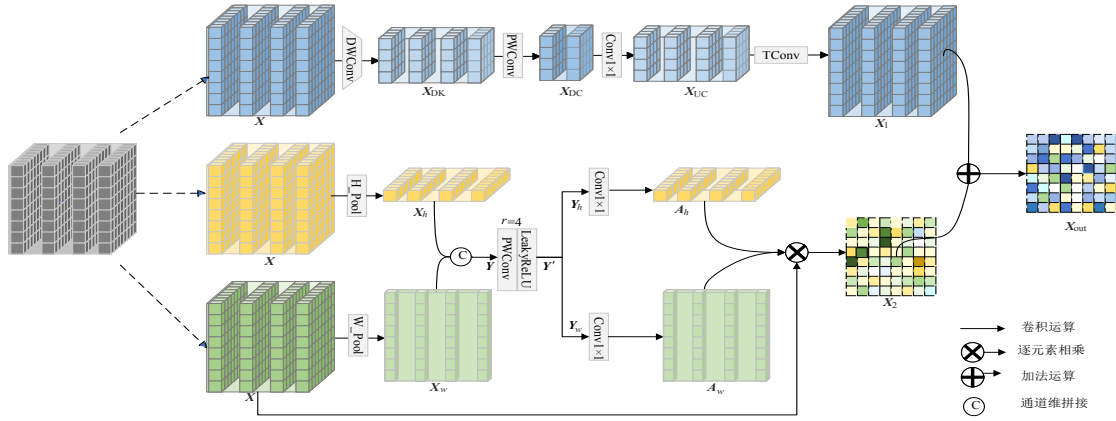


图2 FRM 结构图

Figure 2 Structural diagram of FRM

第一层分支中,首先大小为 3×3 的深度卷积 DWConv 对特征图 X 下采样,进行空间维度压缩,以捕获更大范围的上下文信息,其过程可以表示为

$$X_{DK} = \text{ReLU}(\text{BN}(\text{DWConv}(X))) \in \mathbf{R}^{B \times C \times \frac{H}{2} \times \frac{H}{2}} \quad (2)$$

其中, X_{DK} 表示输出特征图;BN 表示批归一化操作;ReLU 表示 ReLU 激活函数,下文符号含义相同。接着逐点卷积 PWConv 对特征图通道维度压缩,聚焦关键特征,实现跨通道的信息交互与特征重标定,增强特征表达能力,得到特征图 X_{DC} :

$$X_{DC} = \text{ReLU}(\text{BN}(\text{PWConv}(X_{DK}))) \in \mathbf{R}^{B \times \frac{C}{2} \times \frac{H}{2} \times \frac{H}{2}} \quad (3)$$

随后, 1×1 卷积和尺寸为 4×4 、步长为 2,填充为 1 的转置卷积 TConv 对 X_{DC} 上采样,重建在深度卷积过程中可能损失的空间结构,灵活捕捉特征间的复杂关系和结构信息,得到承载了从输入特征中提取并重建的多尺度上下文信息的特征图 X_1 :

$$X_{UC} = \text{ReLU}(\text{BN}(\text{Conv}1 \times 1(X_{DC}))) \in \mathbf{R}^{B \times C \times \frac{H}{2} \times \frac{W}{2}} \quad (4)$$

$$X_1 = \text{ReLU}(\text{BN}(\text{TConv}(X_{UC}))) \in \mathbf{R}^{B \times C \times H \times W} \quad (5)$$

与传统全局平均池化将全局空间信息压缩为单一向量,从而导致空间结构信息丢失不同,FRM 将二维的全局空间关系分解为水平与垂直两个一维的全局编码。第二层分支通过垂直平均池化 H_Pool 捕获垂直方向的信息,建模纵向长距离依赖,生成

分为三条并行分支,整体数学表达式为

$$X_{\text{out}} = F_{\text{FRM}}(X) = F_{\text{branch1}}(X) + F_{\text{branch2}}(X) \odot F_{\text{branch3}}(X) \odot X \quad (1)$$

其中, $X \in \mathbf{R}^{B \times C \times H \times W}$ 表示输入图像, B 为批量大小, C 为通道数, (H, W) 为特征图的空间维度; \odot 表示特征图逐元素相乘,下文相同。 $F_{\text{branch1}}(X)$ 、 $F_{\text{branch2}}(X)$ 、 $F_{\text{branch3}}(X)$ 分别表示第一层、第二层、第三层分支对输入图像 X 的操作。

$X_h \in \mathbf{R}^{B \times C \times 1 \times W}$ 。第三层分支通过水平平均池化 W_Pool 捕获水平方向的信息,建模横向长距离依赖,生成 $X_w \in \mathbf{R}^{B \times C \times H \times 1}$,接着将两个一维全局特征编码通道维度拼接,得到特征图 $Y \in \mathbf{R}^{B \times C \times (H+W) \times 1}$,过程可以描述为

$$X_h = \text{H_Pool}(X) \quad (6)$$

$$X_w = \text{W_Pool}(X) \quad (7)$$

$$Y = \text{Concat}(X_h, X_w) \quad (8)$$

将 Y 输入逐点卷积 PWConv,按照预设的缩放因子 $r=4$,在通道维度压缩,促进水平与垂直全局信息的交互与融合,再通过批归一化层标准化特征分布以稳定训练,采用斜率为 0.2 的 LeakyReLU 激活函数在保留负值信息的同时引入非线性表达,增强模块对方向敏感特征的表达能力,得到特征图 $Y' \in \mathbf{R}^{B \times \frac{C}{4} \times (H+W) \times 1}$,其过程可以表示为

$$Y' = \text{LeakyReLU}(\text{BN}(\text{PWConv}(Y)), 0.2) \quad (9)$$

LeakyReLU 的数学表达式为

$$\text{LeakyReLU}(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (10)$$

其中, x 表示输入; α 表示一个小于 1 的正数。

将 Y' 拆分回垂直特征 $Y_h \in \mathbf{R}^{B \times C \times 1 \times W}$ 和水平特征 $Y_w \in \mathbf{R}^{B \times C \times H \times 1}$,并通过逐点卷积与 Sigmoid 激活函数生成垂直方向和水平方向的空间注意力权重图

$A_h \in \mathbf{R}^{B \times C \times 1 \times W}$ 和 $A_w \in \mathbf{R}^{B \times C \times H \times 1}$ 。最后通过将这两个一维注意力图以广播方式与输入特征 X 进行逐元素相乘,实现对特征图中不同位置进行方向感知的、差异化的重要性加权,从而增强与任务相关的关键区域特征,抑制无关背景信息,输出特征 $X_2 \in \mathbf{R}^{B \times C \times H \times W}$,这个过程可以表示为

$$X_2 = X \odot A_w \odot A_h \quad (11)$$

最后,将包含丰富多尺度语义信息的特征 X_1 和包含精确空间权重的特征 X_2 互补融合,得到先验图 $X_{out} \in \mathbf{R}^{B \times C \times H \times W}$,增强模型在复杂场景下的特征表征:

$$X_{out} = X_1 + X_2 \quad (12)$$

FRM 通过重构特征空间并建立远距离像素间的语义关联,增强网络对特征的提取能力和对全局信息的理解能力。

FRM 结合跳跃连接,构成特征重构残差块 FRRB,其结构如图 3(b) 所示,通过动态特征选择与融合,重构每一层的特征空间分布,打破传统残差块的局部性限制,使网络在全局范围内重新校准特征响应,有效捕捉特征之间的长距离依赖,进而增强网络的学习能力。

1.2.2 特征重构残差块

普通残差块通过跳跃连接在直接传递信息的同时降低了信息优化的难度,使深层网络在避免梯度消失和爆炸的同时有效训练并加速收敛,其结构如图 3(a) 所示。

如图 3(b) 所示, FRRB 包含两层标准的 3×3 卷积,用于从输入特征中提取空间信息,每个卷积层后面均跟随批归一化层,对每个通道提取到的特征进行规范化,从而使得训练过程更加稳定。在第二个 3×3 卷积层后嵌入 FRM,确保网络深度特征提取的每层之间都会进行信息交互,动态生成先验图,重构特征空间中的权重分布,抑制冗余特征并增强判别性区域的响应,从而在深层网络中建立显式的特征间长程依赖关系,结合残差连接,提高网络表征学习能力的同时缓解了深度网络训练中的梯度消失问题。

如图 4 所示,跳跃连接通过直接将输入信息与输出信息相加,解决了梯度消失或梯度爆炸的问题,提高了训练的稳定性和收敛速度。FRRB 通过集成 FRM,不仅有效建模了特征之间的长距离依赖,而且通过保留残差连接保障了深层网络的稳定训练。

普通残差映射和 FRRB 映射分别如式 (13) 和式 (14):

$$H(x) = F(x) + x \quad (13)$$

$$H(x) = M(F(x)) + x \quad (14)$$

1.3 特征优化

FRRB 通过建立长程空间依赖,构建了以语义对象为中心的粗粒度表征,但损失了部分高频细节信

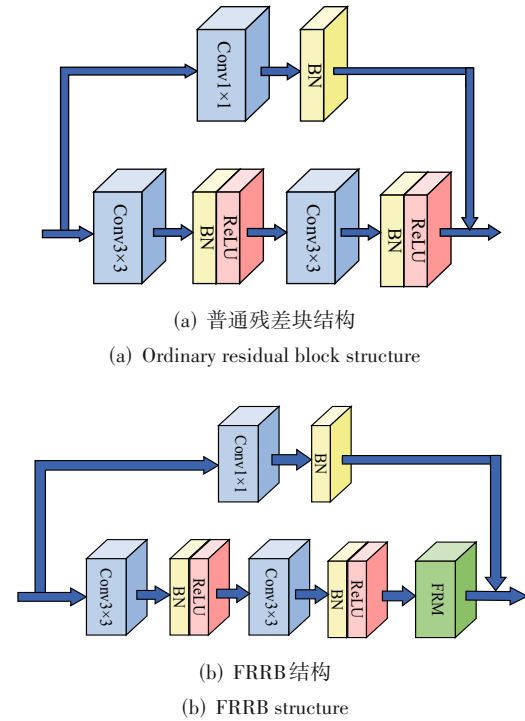


图 3 两种残差块的结构对比图

Figure 3 Structural comparison of two residual blocks

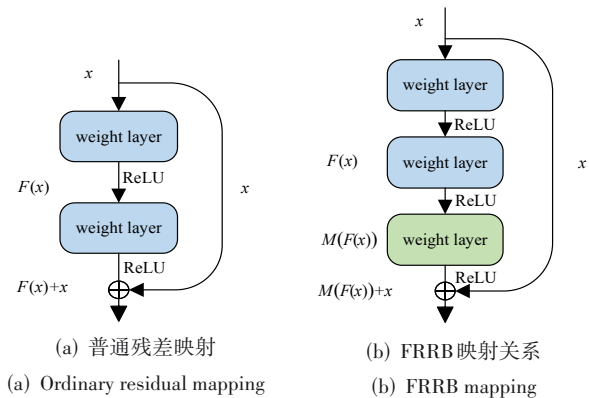


图 4 两种残差映射对比

Figure 4 Comparison of two residual maps

息。为此设计了特征优化分支,在继承全局依赖特征重构的基础上,通过局部细粒度表征增强机制有效补充丢失的细节信息,实现全局语义与局部特征的优化。

将高斯卷积、拉普拉斯卷积以及逐点卷积相结合设计了细粒度特征捕捉模块 FGCM,加强特征之间的类内一致性和类间差异性,在保持高分辨率特征的同时增强细节信息,为后续特征融合做准备。

引入矩阵相乘的思想,设计了一个特征优化模块 FOM,根据 FGCM 捕捉到的高频特征和经过 FRM 重构

的全局特征生成概率图,将两者按照概率图映射权重求和,丰富提取到的特征信息图谱,实现特征空间的完整覆盖。

1.3.1 拉普拉斯-高斯复合卷积

高斯卷积是一种归一化低通滤波器,通过对输入特征 X 进行加权平均实现特征平滑,在此基础上,结合拉普拉斯卷积可以高效提取图像边缘特征。FGCM 采用高斯卷积和拉普拉斯卷积提取图像边缘信息。

高斯平滑为

$$X_{\text{smooth}}(x,y) = \sum_{i=-r}^r \sum_{j=-r}^r G(i,j) \cdot X(x+i,y+j) \quad (15)$$

其中, $X_{\text{smooth}}(x,y)$ 是平滑后图像在位置 (x,y) 处的像素值; $X(x+i,y+j)$ 是原始图像在位置 $(x+i,y+j)$ 处的像素值; r 表示半径,若高斯卷积核 G 大小为 5 , r 为 2 ; $G(i,j)$ 是高斯核在偏移 (i,j) 处的权重; x,y,i,j 是空间坐标,高斯核 G 被定义为

$$G(i,j) = \frac{1}{2\pi\sigma^2} e^{-\frac{(i^2+j^2)}{2\sigma^2}} \quad (16)$$

高斯核根据距核中心的距离远近 σ 分配权重,距离中心越远,权重越小。高斯核的傅里叶变换为

$$F\{G\}(\omega) = e^{-\frac{\sigma^2\omega^2}{2}} \quad (17)$$

其能有效抑制高频噪声。

对于同一类别的两个样本 I_1 和 I_2 , 其差异 $\Delta I = I_1 - I_2$ 可分解为信号差异 ΔI_s 和噪声差异 ΔI_n , 高斯卷积可分解为 $G*\Delta I = G*\Delta I_s + G*\Delta I_n$, 其中 $*$ 表示卷积操作。假设两者不相关,高斯卷积后,由于 ΔI_n 是高频成分,受到显著抑制, $\text{Var}(G*\Delta I) < \text{Var}(\Delta I)$, 即类内方差减小,一致性提升。

拉普拉斯算子作为二阶微分算子,对图像边缘和纹理突变高度敏感。在连续域,二维拉普拉斯算子定义为

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (18)$$

对于阶跃边缘 $f(x) = H(x)$ (Heaviside 阶跃函数), 其一阶导数为 $\delta(x)$, 二阶导数为 $\delta'(x)$, 在 $x=0$ 处产生强响应。

拉普拉斯算子的离散形式采用 8 邻域近似: $L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ 。对于不同类别 A 和 B 的图像 I_A 和 I_B , 在边缘区域 Ω 内, 定义原始差异和拉普拉斯差异:

$$D_{\text{orig}} = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} |I_A(x,y) - I_B(x,y)| \quad (19)$$

$$D_{\text{laplace}} = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} |\nabla^2 I_A(x,y) - \nabla^2 I_B(x,y)| \quad (20)$$

由于拉普拉斯算子对边缘敏感,在 Ω 内通常有

$D_{\text{laplace}} > D_{\text{orig}}$, 即类间差异被放大。

拉普拉斯-高斯复合卷积定义为

$$\text{LoG}(X) = \nabla^2 (G*X) = (\nabla^2 G)*X \quad (21)$$

其中, $\nabla^2 G = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2+y^2}{2\sigma^2} \right] e^{-\frac{x^2+y^2}{2\sigma^2}}$; LoG 在频域的响应为 $F\{\text{LoG}\}(\omega) = -4\pi^2\omega^2 e^{-2\pi^2\omega^2\sigma^2}$, 同理, LoG 能够在抑制类内噪声的同时增强类间边缘差异,实现类内一致性与类间差异性的优化。

1.3.2 细粒度特征捕捉模块

细粒度特征捕捉模块 FGCM 旨在从高分辨率特征图中提取具有判别力的局部细节信息,为实现这一目标引入高斯卷积平滑特征并抑制噪声,利用拉普拉斯卷积突出图像细节,并通过逐点卷积调整通道,增强细节特征,进一步探索局部语义细节。

FGCM 结构如图 5 所示,首先通过尺寸为 5×5 、标准差为 1.0 的卷积核进行高斯卷积计算,抑制高频噪声,平滑噪声,提升类内一致性。接着采用拉普拉斯卷积增强纹理特征,突出不同类别的细微结构差异,从而提升类间差异性,实现先平滑去噪后锐化增强的级联处理,在抑制类内噪声的同时增强类间差异。该过程数学描述为

$$F_h = L*G*X_{\text{high}} = (\nabla^2 G)*X_{\text{high}} \quad (22)$$

其中, X_{high} 表示输入的高分辨率特征图; F_h 表示输出的高分辨率特征图。

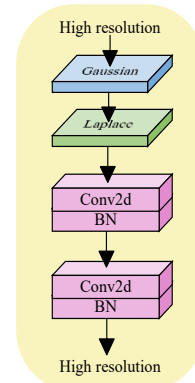


图5 FGCM 结构图

Figure 5 Structure of FGCM

高分辨率特征 F_h 通过两层逐点卷积加权重重要特征通道,进一步增强了具有判别性的细节特征,得到高质量的细节特征 F_{out} , 其过程可以表示为

$$F_{\text{out}} = \text{BN} \left(\text{Conv1} \times 1 \left(\text{BN} \left(\text{Conv1} \times 1 (F_h) \right) \right) \right) \quad (23)$$

1.3.3 特征优化模块

为了有效融合 FGCM 提取的细节特征和网络主干提取的全局信息,提出了特征优化模块 FOM, 该模块的核心思想是根据全局信息和细节信息生成概率

图,并根据概率图的相似性对输入特征图的像素进行加权融合,以达到整合细节信息和全局信息的目的。

如图6所示,输入高分辨率特征图 $F_h \in \mathbf{R}^{B \times C \times H \times W}$ 和低分辨率特征图 $F_l \in \mathbf{R}^{B \times C \times h \times w}$, 其中 $h < H, w < W$ 。FOM 将 F_h 和 F_l 分别通过 3×3 的滤波器卷积, 去除冗余信息, 保留相关信息, 得到 $F_h^* \in \mathbf{R}^{B \times \frac{C}{2} \times H \times W}$ 和 $F_l^* \in \mathbf{R}^{B \times \frac{C}{2} \times h \times w}$, 接着将低分辨率特征图 F_l^* 进行上采样,

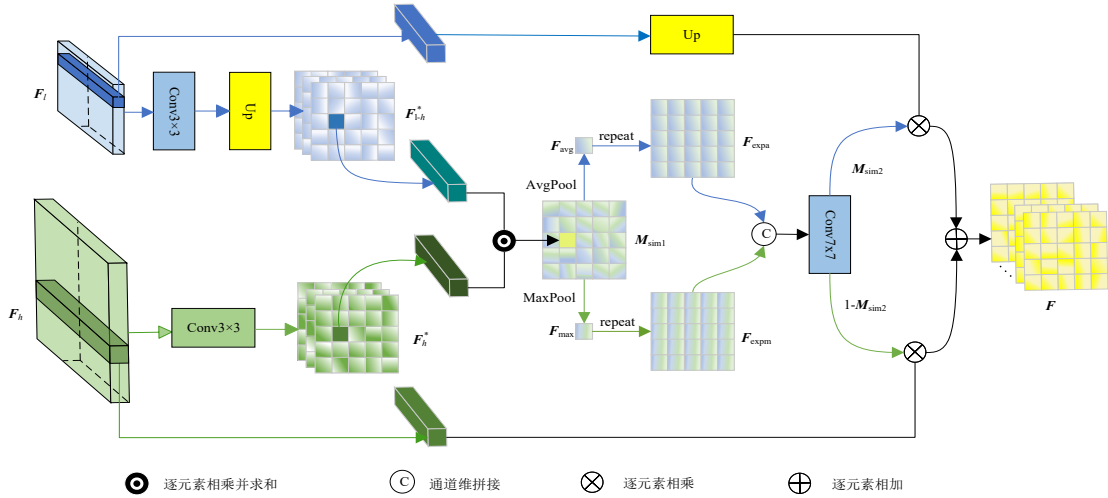


图6 FOM结构图

Figure 6 Structure of FOM

为了获得更精确的融合权重,对 M_{sim1} 进行空间细化。首先计算全局平均池化和全局最大池化:

$$F_{avg} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W M_{sim1}[:, :, i, j] \in \mathbf{R}^{B \times 1 \times 1 \times 1} \quad (25)$$

$$F_{max} = \max_{1 \leq i \leq H, 1 \leq j \leq W} M_{sim1}[:, :, i, j] \in \mathbf{R}^{B \times 1 \times 1 \times 1} \quad (26)$$

为保留空间结构信息,并将两个特征图在空间维度上扩展为与原特征图相同尺寸:

$$F_{expa} = \text{repeat}(F_{avg}, H, W) \in \mathbf{R}^{B \times 1 \times H \times W} \quad (27)$$

$$F_{expm} = \text{repeat}(F_{max}, H, W) \in \mathbf{R}^{B \times 1 \times H \times W} \quad (28)$$

然后将两者拼接得到尺寸为 $[B, 2, H, W]$ 的先验图,通过 7×7 卷积层融合空间上下文,经 Sigmoid 函数得到细化的相似性映射 $M_{sim2} \in \mathbf{R}^{B \times 1 \times H \times W}$ 。

最后,使用细化的相似性映射 M_{sim2} 为 F_h 和上采样后的 F_l 分配权重,得到 $F \in \mathbf{R}^{B \times C \times H \times W}$ 。因为全局下采样提取到的信息更全面且语义丰富,若概率图相应位置权重值较大,则偏重 F_l ,反之,偏重 F_h ,即

$$F = M_{sim2} \odot \text{Upsample}(F_l) + (1 - M_{sim2}) \odot F_h \quad (29)$$

1.3.4 特征优化分支

特征优化分支由 FGCM、FOM 结合深度卷积组成,其核心思想是通过在高分辨特征图中不断补充网络下采样学习到的信息,集成全局语义依赖并增强细

得到特征图 $F_{1-h}^* \in \mathbf{R}^{B \times \frac{C}{2} \times H \times W}$ 。将 F_{1-h}^* 与 F_h^* 逐元素相乘,得到多分辨率特征交互图,尺寸为 $[B, C/2, H, W]$,在通道维度上对特征交互图求和,再经过 Sigmoid 激活函数生成单通道相似性映射:

$$M_{sim1} = \text{Sigmoid} \left(\sum_{c=1}^{C/2} (F_{1-h}^* \odot F_h^*)_c \right) \quad (24)$$

其中, $(\cdot)_c$ 表示特征交互图的第 c 个通道, $M_{sim1} \in \mathbf{R}^{B \times C \times H \times W}$ 表示整体特征的匹配度。

节特征,其结构如图7所示,可以分为如下3个阶段:

(1) stage1。主干网络 layer1 解析得到带有全局依赖的特征图,使用 FGCM 增强其细节信息,用 FOM 对 layer2 处理后的全局信息和 FGCM 提取到的细节信息进行权重分析,生成重要特征概率图,引导全局信息和细节信息融合,生成带有全局信息和局部信息的高分辨率特征图,增强细节特征,为之后的特征提取提供更丰富的信息。

(2) stage2。对 stage1 处理后的特征进一步提取细节信息,基于 FOM 动态生成的概率图的引导,将全局信息补充到带有丰富细节信息的高分辨率特征图,提升特征的丰富度并增强全局语义依赖。

(3) stage3。使用深度卷积对 stage2 生成的带有丰富信息的高分辨率特征图选择关键特征学习并逐层降维,保留样本的分布特性,并通过梯度回传,使得 FRM 能够在其后续层级中解析出更准确、更具判别性的长距离依赖关系,提高网络的学习能力。

特征优化分支利用 FGCM 提取细节信息,通过 FOM 将局部细节信息和全局信息融合,既保留图像全局上下文信息又增强对局部变化的敏感性,使网络学习更完整的特征,提升网络对不同尺度图像和对象的学习能力,进而提升网络的性能。

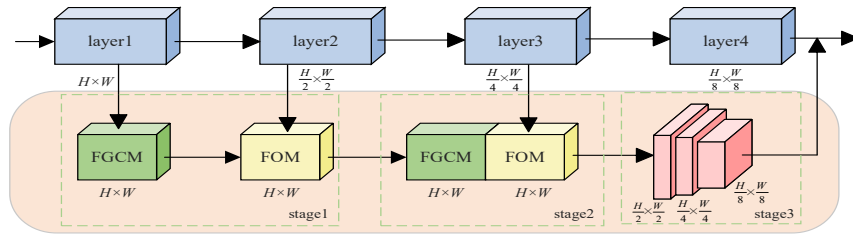


图7 特征优化分支结构图

Figure 7 Structure of feature optimization branch

2 实验与结果分析

2.1 实验设置

2.1.1 实验环境

本文实验采用Python作为编程语言,使用Pytorch框架构建模型,配备NVIDIA RTX 3090显卡在Windows 11操作系统上运行,显存容量为24 GB。

2.1.2 数据集概述

为了测试GDFRNet的性能,选择CIFAR-10、CIFAR-100、SVHN、GTSRB、Imagenette和Imagewoof数据集作为评估数据集,表1统计了六个数据集的基本参数信息。

表1 实验所用数据集参数

Table 1 Parameters of the dataset used in the experiment

数据集名称	总量	类别	样本尺寸	训练集	测试集
CIFAR-10	60 000	10	32 × 32	50 000	10 000
CIFAR-100	60 000	100	32 × 32	50 000	10 000
SVHN	600 000	10	32 × 32	73 257	26 032
GTSRB	51 839	43	32 × 32	39 209	12 630
Imagenette	6 000	10	256 × 256	9 469	3 925
Imagewoof	6 000	10	256 × 256	9 025	3 929

本文实验数据集均在数据增强后使用,采用图像分类常用指标分类准确率(Accuracy)作为评价指标。

2.1.3 实验细节

训练过程和模型对比实验中使用的参数如表2所示,其中“通用超参数”为实验中各类对比方法及本文所提出的方法共同使用的超参数,“FRM参数”为本文提出的方法在实验中定量时使用的参数。

2.2 FRM中影响网络性能的因素

FRM中影响网络性能的因素有FRM的缩放因子 r 、转置卷积核大小 k 以及FRM在FRRB中的嵌入位置,此外,本节还将探讨FRM中使用转置卷积上采样的优势。

2.2.1 FRM的缩放因子 r 对GDFRNet性能的影响

FRM通过缩放因子 r 控制缩放的通道数,促使网络跨通道特征交互,间接影响特征重构。为了研究缩放因子 r 对网络性能的影响,设计了 $r=2$ 、 $r=4$ 、 $r=8$ 、 $r=16$ 、 $r=32$ 五种测试方案,并在CIAFR-10、CIFAR-100、

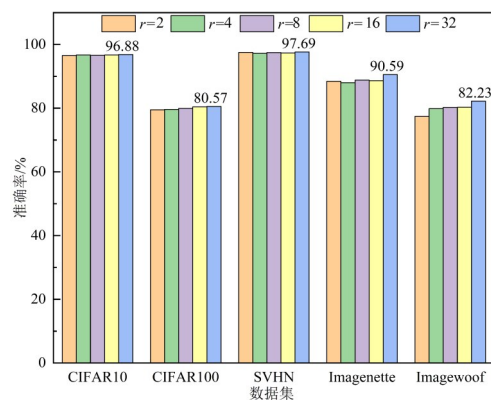
表2 训练过程和模型对比实验中使用的参数表

Table 2 Parameters used in training and model comparison experiments

类别	参数	代表符号	取值/范围
通用超参数	学习率	Lr	0.1
	学习率衰减	Ld	0.2
	迭代次数	Epoch	200
	优化器	optimizer	SGD
	标签平滑	Ls	0.1
	动量	momentum	0.9
	权重衰减率	weight_decay	5×10^{-4}
	大尺寸数据集批次	L_batch_size	64
	小尺寸数据集批次	S_batch_size	128
FRM参数	缩放因子	r	32
	卷积核尺寸	k	4

SVHN、Imagenette和Imagewoof数据集上展开实验。

如图8所示,整体上GDFRNet的性能和缩放因子呈正相关。缩放因子 r 为32时,网络整体性能高于其他方案,证明此时FRM可以有效减少冗余信息,同时保留关键特征。

图8 FRM的缩放因子 r 对GDFRNet性能的影响Figure 8 Effect of FRM scaling factor r on GDFRNet performance

2.2.2 FRM中转置卷积对网络性能的影响

转置卷积可以通过学习权重使得网络提取到更多关键特征,但是不恰当的卷积方式会加重棋盘效应,不利于网络学习语义信息。为了实现将特征图扩大两倍的目标,FRM的转置卷积步长设置为2,此时

卷积核大小为其整数倍,可以缓解棋盘效应带来的影响。为了探索转置卷积核 T_Conv 的大小对模型性能的影响,设计了五种方案,如表 3 所示。

表 3 FRM 的 5 种转置卷积方案

Table 3 Five transposed convolution methods for FRM

实验方案	kernel_size	stride	padding
1	2	2	0
2	4	2	1
3	6	2	2
4	8	2	3
5	10	2	4

结果如图 9 所示,五种方案在不同数据集上表现均有所差别。在 Imagenette 和 Imagewoof 上差别较明显,方案 2 的准确度相对较高,即卷积核大小为 4 时, GDFRNet 的性能最优;卷积核大小为 2 时,网络在 CIFAR-100、SVHN、Imagenette 数据集上表现较好,但在 Imagewoof 上性能较差;卷积核大小为 6 时,网络在 Imagewoof 上性能较好,但是在 CIFAR-100、SVHN、Imagenette 上性能有所下降。较大的卷积核可以覆盖更多的邻域信息,有助于捕捉更广泛的上下文信息,在特征还原时能够更好地恢复细节和纹理信息,然而卷积核过大可能会增加模型的复杂度,引入更多的噪声,不利于网络表征学习,导致网络性能下降。

2.2.3 FRM 不同上采样方式的对比

上采样方法主要有转置卷积和图像处理常用的插

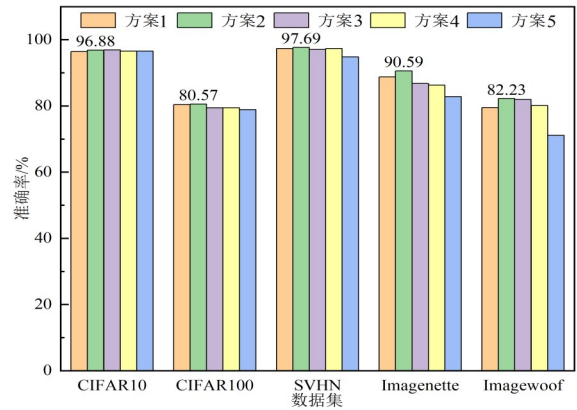


图 9 T_Conv 尺寸对模型性能的影响

Figure 9 Impact of T_Conv size on model performance

值法,故将两种方法在 GTSRB、Imagenette、Imagewoof 三个数据集上进行比较,结果如图 10 所示,net1 表示采用双线性插值的网络,net2 表示采用转置卷积的网络。

整体而言,net2 的准确率曲线比 net1 更稳定,在 epoch=150 后,net2 的准确率都保持在 net1 之上浮动,证明结合转置卷积逆向重构特征空间,可以有效捕捉特征的全局依赖关系,提高网络的表征学习能力。如图 10(a)所示,net2 的准确率曲线在 epoch=150 之前都低于 net1,但经过迭代学习 net2 网络的分类能力逐渐超越 net1,表明使用转置卷积上采样可以通过训练逐渐提高网络对目标的拟合能力。

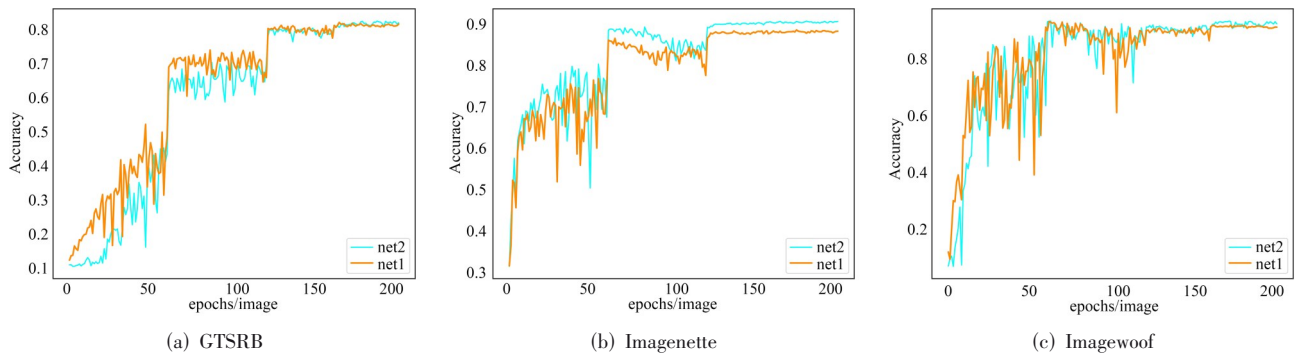


图 10 FRM 中转置卷积和双线性对模型性能的影响

Figure 10 Impact of transposed convolution and bilinear mapping on model performance in FRM

2.2.4 FRM 在 FRRB 中的位置对网络性能的影响

FRM 是 FRRB 中的核心引擎,为了探索 FRM 嵌入位置对网络性能的影响,如图 11 所示,设计了四种方案,在 CIFAR-10、CIFAR-100、SVHN、Imagenette 以及 Imagewoof 数据集上进行实验。其中,C_Conv 表示普通卷积层、批归一化层和激活函数的组合,D_Conv 表示降维操作,实验结果如表 4 所示。

如表 4 所示,方案 C 的嵌入位置对于 GDFRNet 的

性能提升最显著,FRM 通过解析 FRRB 中两层卷积处理后的特征语义关联,动态重构特征分布空间,建立跨层、跨区域的长距离依赖关系,同时,跳跃连接避免梯度消失和网络退化的问题。方案 D 的效果次之,FRM 处于跳跃连接分支时,与 FRRB 中的卷积层并联,FRM 通过对前一层网络的输出信息交互得到先验图,依据先验信息调整特征权重,可能由于未直接处理卷积层之间的特征交互,导致长距离依赖的建模

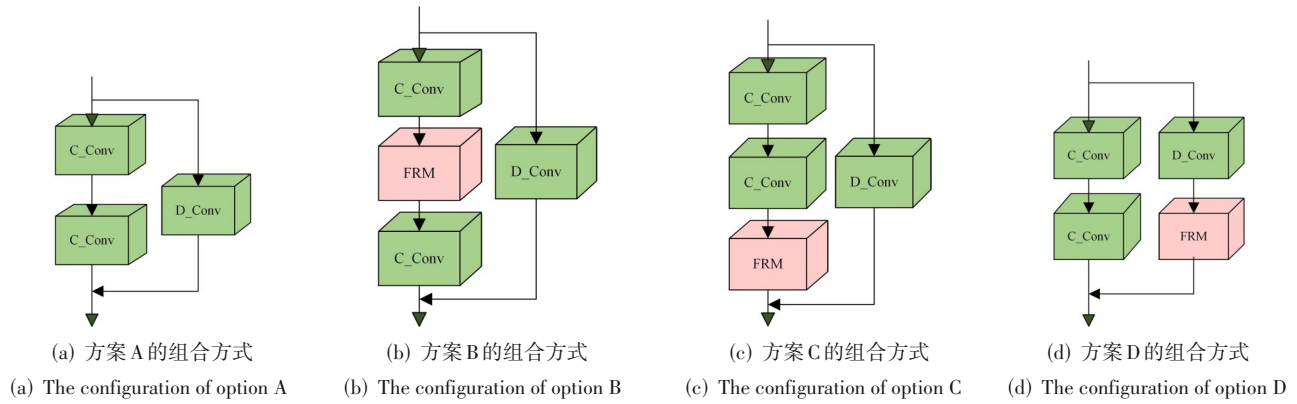


图 11 FRM嵌入位置的4种组合方式

Figure 11 Four combination methods for FRM embedding locations

能力较弱,性能提升有限。方案B中FRM处理第一层卷积输出的特征,在CIFAR-100和SHVN数据集上提升,然而在CIFAR-10、Imagenette、Imagewoof数据集上性能下降,这可能与数据集的复杂背景以及对象在图像中的占比大小有关。

表 4 FRM不同嵌入位置的实验结果 单位:%

Table 4 Experimental results for different FRM embedding positions

unit: %

实验方案	CIFAR-10	CIFAR-100	SHVN	Imagenette	Imagewoof
A	96.45	79.01	97.39	88.28	76.66
B	96.17	79.17	97.49	85.66	74.83
C	96.88	80.57	97.69	90.59	82.23
D	96.50	80.02	97.26	89.10	82.34

注:黑色加粗字体表示效果最优。

2.3 特征优化分支中影响网络性能的因素

细节信息的提取在网络表征学习中起到了举足轻重的作用。特征优化分支中FGCM的数量、排列位置和特征优化分支的层数均会对网络的性能造成影响,本节在CIFAR-100、GTSRB、SVHN、Imagenette和Imagewoof五个数据集上进行实验,探究特征优化分支中影响网络性能的因素。

2.3.1 FGCM的布局对模型性能的影响

FGCM主要作用是提取并增强边缘等细节信息,以备后续网络敏感信息检索。为了研究FGCM不同的摆放方式对GDFRNet性能的影响,设计了四种方案,在CIFAR-100、GTSRB、Imagenette和Imagewoof四个数据集上进行实验,设计方案如图12所示。

实验结果如表5所示,对于四种不同类型的数据集,四种方式排列布局带来的效果各有不同。对于CIFAR-100数据集而言,方案A的准确率最高,达到80.57%,方案C的准确率次之,方案B的准确率最低,

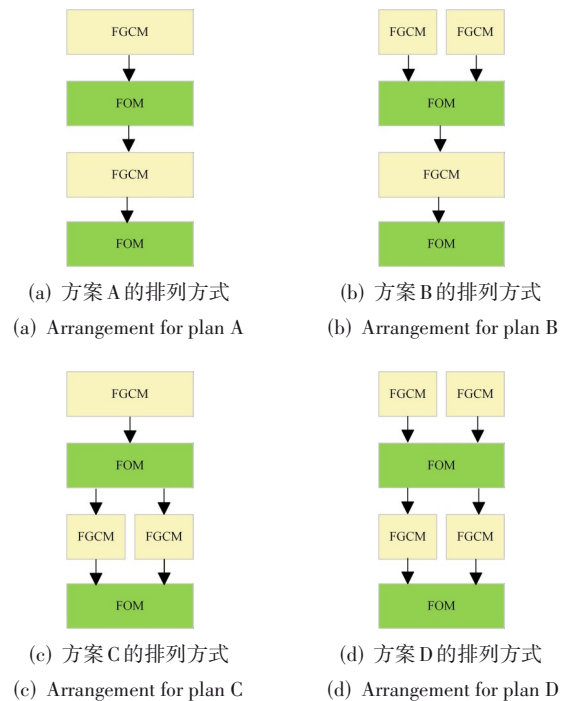


图 12 FGCM模块的4种组合方式

Figure 12 Four configuration options for the FGCM module

可能第一次细节信息提取阶段细节特征过多,噪声增加,导致模型学习到无用的特征,影响模型性能。对于GTSRB数据集而言,方案D的表现最优,由于交通标志比较规范,细节信息的提取与保留对模型一直保持积极影响。对于Imagenette而言,方案C性能最好,随着特征图丰富度的增加,FGCM通过细粒度提取,提炼特征中的细节信息,可以减少无关信息对模型的干扰,提升模型的性能,但是当按照方案D组合时会产生噪声,降低模型性能。对于Imagewoof数据集而言,方案B的性能最好,在高分辨率图像中提取足够的细

节信息, 可以结合全局信息在复杂的背景中区分目标对象, 并且关注到小的目标, 从而提升网络性能。

细节的提取对于比较规范的图像可以详细地进行表征学习, 提升网络性能; 但是对复杂场景的图像而言, 过多关注细节信息会加重噪声对网络的影响, 干扰网络特征选择, 降低网络性能。

表5 不同FGCM的组合方式对GDFRNet性能的影响 单位:%

Table 5 Impact of different FGCM combination methods unit: %

实验方案	CIFAR-100	GTSRB	Imagenette	Imagewoof
A	80.57	94.89	90.59	82.23
B	79.95	95.47	90.90	83.95
C	80.09	95.30	91.38	82.83
D	80.02	96.09	90.14	81.95

注: 黑色加粗字体表示性能最优。

2.3.2 特征优化分支层数对GDFRNet性能的影响

特征优化分支的主要任务是保持高分辨率的的同时补充细节特征, 辅助网络完善特征的学习。特征优化分支的层数少会无法捕捉足够的细节信息, 特征优化分支过多会导致噪声影响表征学习, 带来不利影响。为了探究特征优化分支的层数对性能的影响, 设计四种实验方案, 如图13所示, 其中画出关键模块, DRM包括FGCM和FOM。本次实验在CIFAR-100、SVHN、Imagenette、Imagewoof四个数据集上进行。

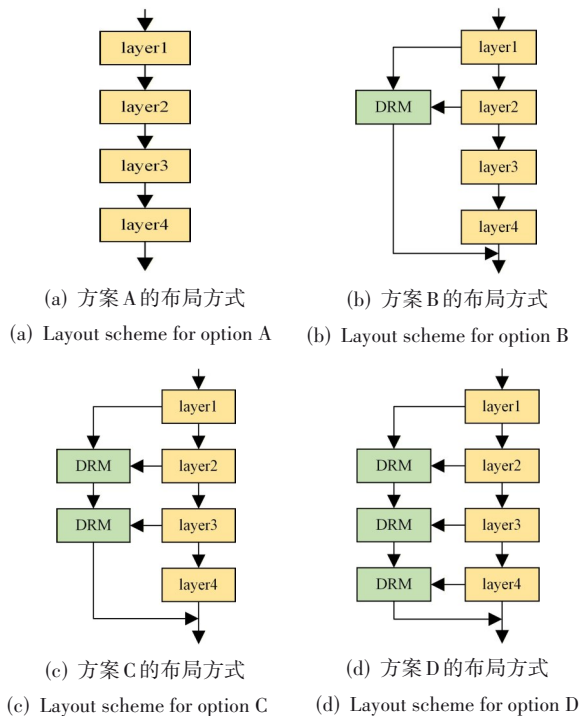


图13 GDFRNet特征优化分支的4种方案

Figure 13 Four schemes for the feature optimisation branch of GDFRNet

实验结果如表6所示, 方案B的准确率均比方案A高, 方案C的准确率均比方案B高, 表明特征优化分支层数的增加会促使网络性能提高, 但是当网络的特征优化分支增加到三层时, 由方案D的数据可知, 网络性能下降, 这可能是由于特征优化分支过多, 增加噪声, 干扰了网络学习。

表6 细节分支层数对GDFRNet性能的影响 单位:%

Table 6 Impact of detail branch levels on GDFRNet performance unit: %

方案	CIFAR-100	SVHN	Imagenette	Imagewoof
A	80.28	97.54	88.88	82.16
B	80.27	97.41	88.94	82.17
C	80.57	97.69	90.59	82.23
D	79.04	97.63	90.34	82.34

注: 黑色加粗字体表示性能最优。

2.4 对比实验与结果分析

为了验证GDFRNet的有效性, 将其与四类当前最先进模型(State-Of-The-Art, SOTA)进行了对比实验: (1) 基于卷积神经网络的经典分类模型ResNet-34^[3]和CAPR-DenseNet^[24] (Dense convolutional Network); (2) 基于卷积建模长依赖关系的网络模型Multi-ResNet^[4] (Multi-Residual Networks)、ChostNet^[5]、WideResnet-28-10^[6]、EfficientNet^[7]和SCAM-Net^[25] (Spatially Constrained Attention Mechanism for image classification Network); (3) 基于Transformer注意力机制的分类模型, 包括Couplformer^[26]、FAVOR+^[27] (Fast Attention Via positive Orthogonal Random features approach)、MMA-CCT-7/3×2^[20] (Multi-manifold Multihead Attention) 和QKFormer^[28]; (4) 采用辅助分支的分类方法, 包括TA-DFKD^[23] (the Teacher-Agnostic Data-Free Knowledge Distillation)、SpaceshipNet^[29]、ATONet^[30]、SLLNet^[22]和MSMDFormer^[31]。实验结果如表7所示。

如表7所示, GDFRNet在五个数据集上测试效果显著。与ResNet-34和CAPR-DenseNet比较, GDFRNet准确率在CIFAR-10数据集上平均提升4.96%, 在Imagenette数据集上提升3%左右, 证明在层之间增加信息的交互以重建特征空间以及特征优化分支的策略是有效的。与SLLNet和MSMDFormer相比, GDFRNet均有优势, 有力地证明通过FRM建模特征间的长依赖关系和特征优化分支补充细节特征, 网络表现出更优的效果, 进一步验证这两方面的改进带来了准确率的提升。

虽然GDFRNet在与SOTA准确率对比实验中表现良好, 但是运行复杂度仍需要更全面地评估, 为此, 进一步分析了GDFRNet与SOTA的参数数量和计算量, 实验结果见表8。

表7 对比实验结果
Table 7 Comparative experimental results

单位:%
unit: %

SOTA 模型	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
ResNet-34	89.60	63.29	96.21	87.54	79.04
CAPR-DenseNet	94.24	78.84	94.95	87.72	77.91
ChostNet	94.92	77.15	93.86	87.83	78.22
EfficientNet	94.01	75.96	93.32	88.01	77.93
Muti-ResNet	94.56	78.68	94.58	87.69	81.21
WideResNet-28-10	95.83	79.50	95.21	88.34	78.71
SCAM-Net	96.28	80.40	97.25	—	—
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	94.29	—	—
QKFormer	96.18	80.26	97.13	88.32	81.65
TA-DFKD	95.46	78.42	96.83	87.62	80.57
SpaceshipNet	95.39	77.41	—	80.59	—
ATONet	94.51	78.54	95.21	86.67	80.19
SLLNet	95.51	79.23	96.91	87.93	80.89
MSMDFormer	95.65	77.46	96.87	—	—
GDFRNet	96.88	80.57	97.69	90.59	82.23

注:黑色加粗字体表示最优值,实验中SOTA数据采用论文报告结果(未开源)或代码复现结果(已开源)。

表8 参数量和计算量对比结果

Table 8 Comparison of parameter quantities and computational complexity

模型	GDFRNet(ours)	ResNet-34	Multi-ResNet	SLLNet	TA-DFKD	EfficientNet	WideResNet-28-10	QKFormer
模型参数/M	56.45	21.28	51.23	31.57	40.97	52.98	36.48	6.74
模型计算量/G	2.56	1.16	3.13	2.08	1.85	1.49	5.25	6.09

如表8所示,GDFRNet与ResNet以及GhostNet等轻量级模型相比,参数复杂度更高,但计算量显著低于部分大模型,如Multi-ResNet和WideResnet以及QKFormer,这一对比证明更高的参数量带来了精度突破,同时避免了计算冗余,一定程度平衡了特征解析能力、空间信息保留和计算效率。

为了进一步分析模型性能,分别绘制了ResNet-34和GDFRNet在CIFAR-10、Imagenette以及Imagewoof数据集上的混淆矩阵和受试者工作特征曲线(Receiver Operating Characteristic curve, ROC),如图14和图15所示。

如图14所示,可以明显看出GDFRNet将真实为正的样本预测为正样本的数量较多,验证了GDFRNet的性能较好。

如图15所示,在三个数据集上,GDFRNet的曲线下面积(Area Under the Curve, AUC)均高于ResNet-34,并且其ROC曲线更陡峭,表明前者在较低的假阳性率下就能获得较高的真阳性率。在Imagenette数据集上,GDFRNet的ROC曲线波动更小,表现更稳定,

故GDFRNet的性能更好。

2.5 消融实验

为了进一步探究GDFRNet中各模块对整体性能提升的有效性和模块之间的协作效果,本文展开了消融实验,在大量实验数据中,选取以下五组数据(即net1、net2、net3、net4、net5分别在CIFAR-10、CIFAR-100、SVHN、Imagenette和Imagewoof五个数据集上的实验结果)展开进一步的分析,如表9所示。

net1是baseline模型,net2是修改浅层特征提取部分后的网络,即将首层 7×7 卷积层改为 3×3 卷积,并删去最大池化操作,net3是消去特征优化分支的网络,net4是不使用FRRB的网络,net5是堆叠FRRB并增加特征优化分支的完整网络,即本文提出的GDFRNet。

为了更好地表现数据信息和实验效果,画出了五组数据的折线图,如图16所示,主要可以得出以下结论。

在CIFAR-10、CIFAR-100、SVHN数据集上,net2的曲线趋势显著优于net1,由此现象可得,通过将网络首层 7×7 卷积替换为 3×3 卷积并删除最大池化,网络性能得到显著提升。

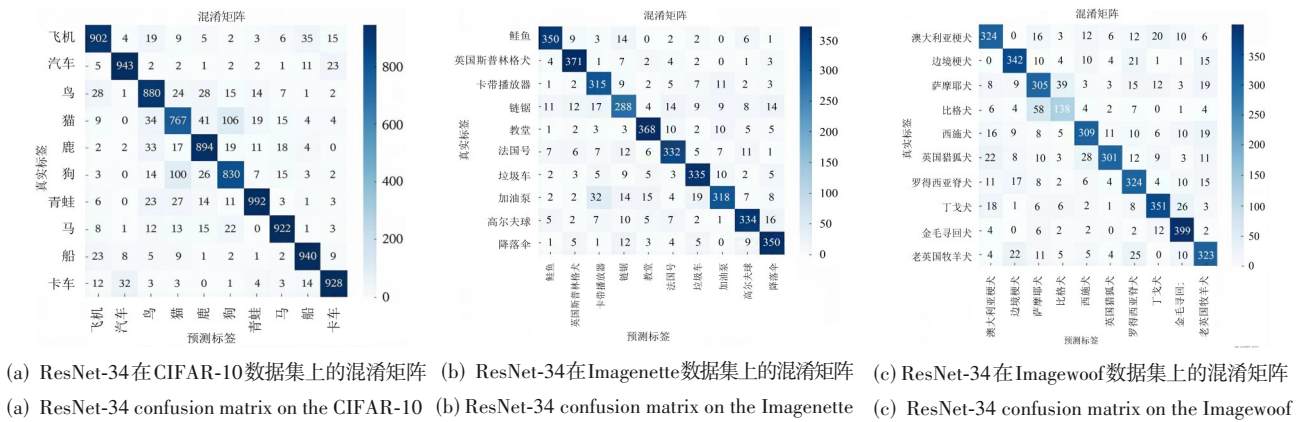


图 14 ResNet-34和GDFRNet在3个数据集上的混淆矩阵

Figure 14 Confusion matrices for ResNet-34 and GDFRNet across three datasets

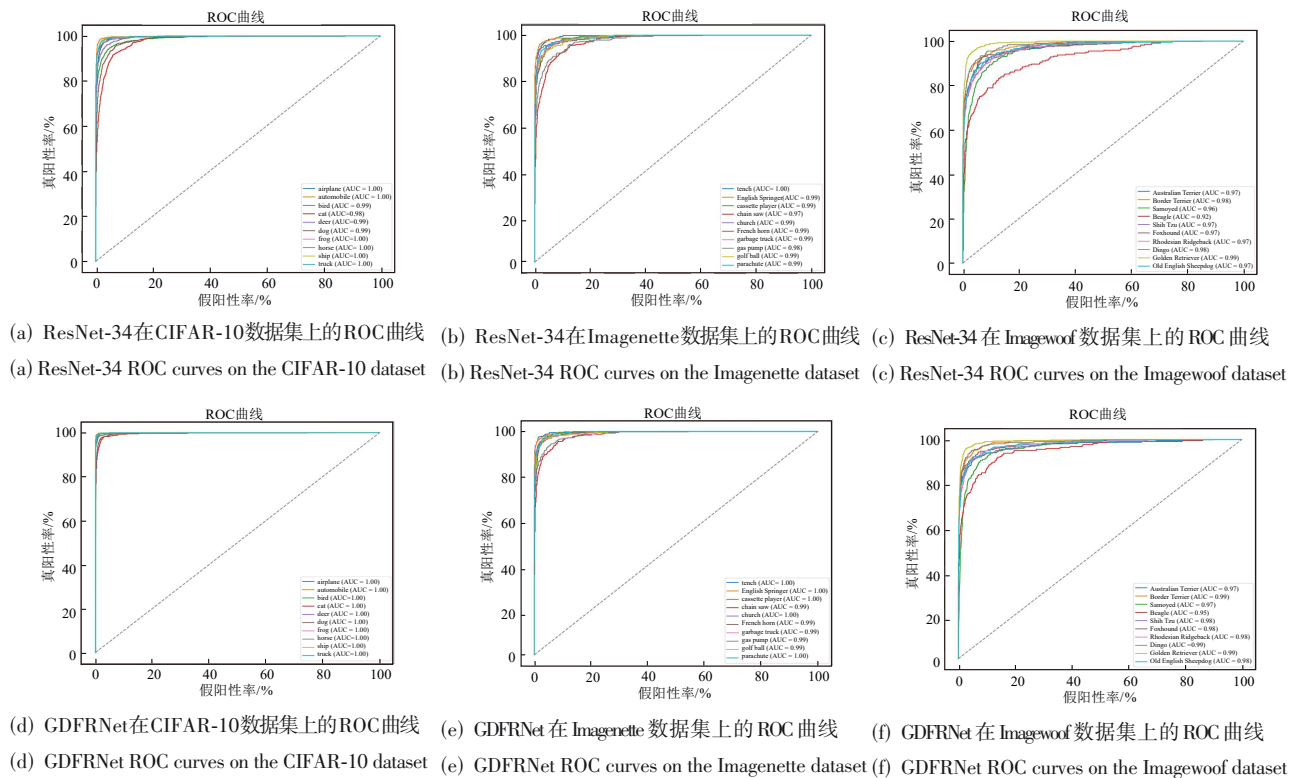


图 15 ResNet-34和GDFRNet在3个数据集上的ROC曲线对比

Figure 15 Comparison of ROC curves for ResNet-34 and GDFRNet across three datasets

表9 GDFRNet消融实验结果表 单位:%
Table 9 GDFRNet ablation experiment results unit: %

模型	CIFAR-10	CIFAR-100	SVHN	Imagenette	Imagewoof
net1	89.6	63.29	96.21	87.54	79.04
net2	96.46	78.14	97.10		
net3	96.58	80.28	97.54	88.88	82.16
net4	96.56	79.01	97.39	89.19	82.06
net5	96.88	80.57	97.69	90.59	82.23

注:黑色加粗字体表示最优值。

使用FRRB提取特征的net3在五个数据集上的曲线效果一直呈现较高准确率的状态,有力地证明了FRM的有效性,FRM加强网络学习过程中的信息交互能力,减少冗余特征,对特征空间进行调整,有效建立特征间的全局依赖。但是在Imagewoof数据集上

FRM可能会导致模型容易出现过拟合的现象,且与net5相比,分类效果不佳,证明了特征优化分支提供细节信息的必要性。

如图16(e)所示,net4曲线整体波动较小,epoch=100后的浮动相对稳定,表明特征优化分支很好地缓解了这个问题。当增加特征优化分支辅助时,网络可以更完整地学习特征,提高网络的稳定性和表征学习能力。如图16(b)所示,net4与net5相比,可以证明FRM解析长距离依赖对细节特征提取有促进作用。

当FRRB和特征优化分支相结合时,网络对五个数据集的分类准确率均有所提升,验证了两个方法相互协作的有效性。net5曲线整体比较稳定,有力地证明了FRRB和特征优化分支的结合提升了网络的稳定性和分类性能。

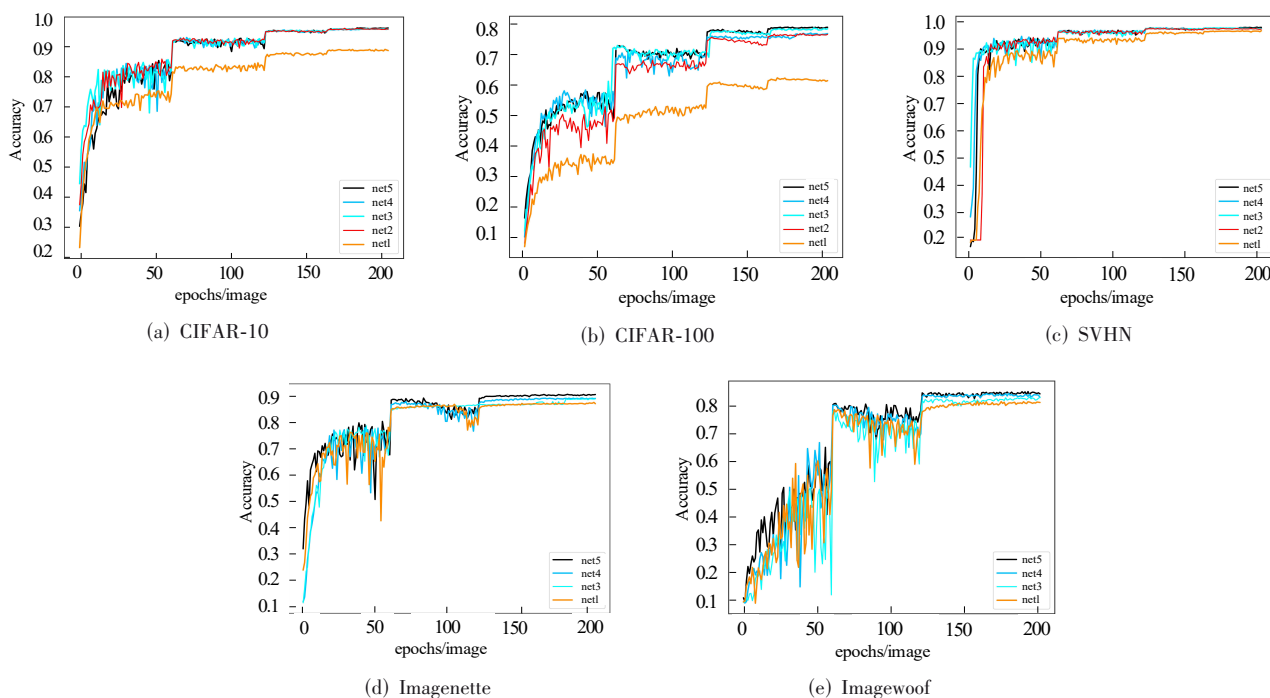


图16 各网络在5个数据集的分类准确率

Figure 16 Classification accuracy of various networks across five datasets

2.6 可视化实验

本节将GDFRNet不同部分提取到的特征进行可视化展示,从而分析FRRB中特征重构模块FRM和特征优化分支的有效性,如图17所示。

图17(b)和图17(c)的对比可以清楚地观察到FRM能有效区分目标和背景信息。图17(b)中小狗周围的背景存在大量离散且高强度的激活响应,这些语义无关的响应构成了显著的特征噪声。图17(c)中小狗的背景相对较暗,表明背景噪声被有效抑制,小

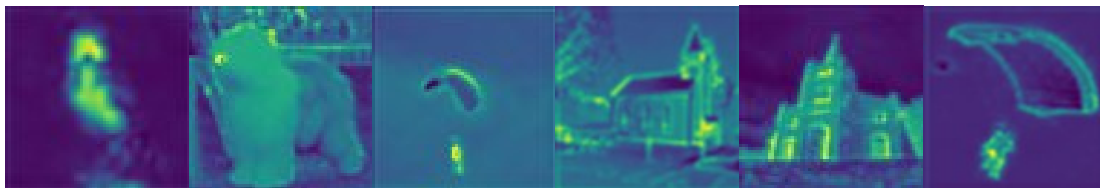
狗的颜色更亮,目标更突出,头部特征分布更集中,且捕捉到小狗的纹理,各部位之间形成了语义连贯的激活区域;房子和降落伞的颜色都较亮,轮廓都较突出,背景颜色更暗,相较普通残差块,FRRB通过建立长距离依赖关系,动态筛选关键特征,抑制背景噪声,实现对特征空间的重构。与图17(c)相比,图17(d)中第二列小狗的毛发轮廓、第五列房屋的窗户结构均得到清晰呈现,验证了FGCM能有效捕捉并增强重要的局部细节特征。图17(b)中小狗的轮廓较模糊,图17(e)

融合了图 17(c)的全局语义和图 17(d)的局部细节,小鸟的面部、背部以及尾部的纹理信息得到了显著增强,目标的激活更加集中和连贯,可以看到小鸟和背景的背景的边界,说明 FRM 建立特征间长距离依赖关系,可以引导 FGCM 增强关键的局部细节,在 FOM 融合

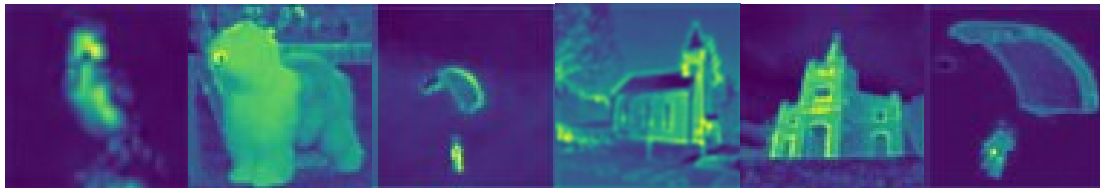
FRRB 得到的全局信息和 FGCM 提取到的细节信息阶段,目标语义信息被强化,并且保留了丰富的细节信息,背景信息被进一步抑制。上述渐进式可视化分析表明,网络通过捕获全局语义与局部细节的协同表征,显著增强了特征判别力,进而提升了模型的性能。



(a) 原图
(a) Original image



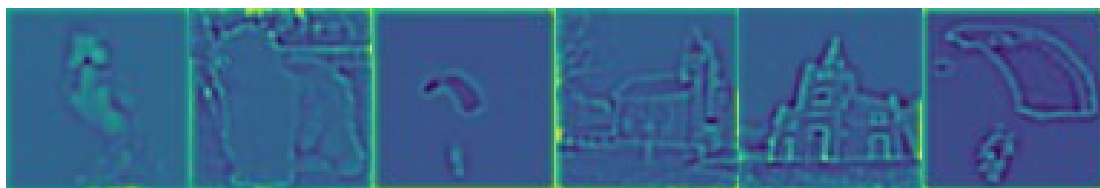
(b) 普通残差块layer1 卷积层提取到的特征图
(b) Feature maps extracted from the layer1 convolutional layer of the ordinary residual block



(c) FRRB 提取到的特征图
(c) Feature maps extracted by FRRB



(d) 特征优化分支提取到的含有边缘信息的细节特征图
(d) Detail feature map containing edge information extracted by the feature optimisation branch



(e) FRRB 处理后的全局信息和细节信息结合的特征图
(e) Feature map combining global and local information processed by the FRRB

注:图中特征图使用热力图可视化,颜色明度表示激活强度,越亮的区域代表该特征响应越强。

图 17 GDFRNet 模型主要模块特征可视化

Figure 17 Visualisation of key features in GDFRNet model

3 结束语

针对卷积神经网络对特征间长距离依赖关系建模不足的问题,本文提出了全局依赖引导的特征重构图像分类网络GDFRNet。该网络通过设计特征重构模块和特征优化分支,建立“全局-局部”双流架构,在全局依赖特征的引导下强化细节特征,使网络在保持对整体结构认知的同时,捕捉细微的判别性特征,实现特征空间在语义完整性和细节丰富性上的最优平衡,增强模型对复杂数据的解析能力,进而增强网络表征学习能力和图像分类性能。在实验验证中,选取CIFAR-10、CIFAR-100、SVHN、GTSRB、Imagenette以及Imagewoof六个具有代表性的数据集,通过准确率、混淆矩阵以及ROC曲线对GDFRNet进行了评估,通过可视化分析验证了提出的FRM、FGCM以及FOM的有效性。

虽然GDFRNet在图像分类任务中表现出较好的性能,但仍有很大的改进空间,未来可以结合动态稀疏注意力设计轻量化特征重构模块,增强网络全局依赖建模的同时降低计算量。

参考文献

- [1] 刘颖,薛家昊,张伟东,等.基于坐标重要性池化和解耦类别对齐蒸馏的图像分类算法[J].电子学报,2025,53(3): 962-973.
Liu Ying, Xue Jiahao, Zhang Weidong, et al. Image classification algorithm based on coordinate importance pooling and decoupled class alignment distillation[J]. Acta Electronica Sinica, 2025, 53(3): 962-973. (in Chinese)
- [2] Jiang Wentao, Yuan Heng, Liu Wanjun. Neuron signal attenuation activation mechanism for deep learning[J]. Patterns, 2025, 6(1): 101117.
- [3] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [4] Abdi M, Nahavandi S. Multi-residual networks: Improving the speed and accuracy of residual networks[PP/OL]. V2. arXiv (2017-03-15) [2025-07-07]. <https://doi.org/10.48550/arXiv.1609.05672>.
- [5] 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. Piscataway: IEEE, 2020: 1577-1586.
- [6] Zagoruyko S, Komodakis N. Wide residual networks[PP/OL]. V2. arXiv(2017-06-14) [2025-07-07]. <https://arxiv.org/pdf/1605.07146>.
- [7] Tan Mingxing, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[PP/OL]. V2. arXiv(2020-09-11) [2025-07-07]. <https://arxiv.org/abs/1905.11946>.
- [8] Xiong Yuwen, Li Zhiqi, Chen Yuntao, et al. Efficient deformable ConvNets: Rethinking dynamic and sparse operator for vision applications[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 5652-5661.
- [9] Kim S, Park E. SMPConv: Self-moving point representations for continuous convolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 10289-10299.
- [10] 孙航,付秋月,李勃辉,等.基于跨层注意力特征交互和多尺度通道注意力的单幅图像去雾网络[J].电子学报,2024,52(11): 3711-3726.
Sun Hang, Fu Qiuyue, Li Bohui, et al. Cross-layer attention feature interaction and multi-scale channel attention network for single image dehazing[J]. Acta Electronica Sinica, 2024, 52(11): 3711-3726. (in Chinese)
- [11] Hao Chaofan, Li Yanfeng, Sun Jia, et al. MFEN: Multi-scale feature expansion network for visible-infrared person re-identification[C]//Proceedings of the International Conference on Computer Vision and Deep Learning. New York: ACM, 2024: 3653824.
- [12] Gao Shanghua, Cheng Mingming, Zhao Kai, et al. Res2Net: A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.
- [13] Jeong J, Zou Yang, Kim T, et al. WinCLIP: Zero-/few-shot anomaly classification and segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 19606-19616.
- [14] Hou Qibin, Zhou Daquan, Feng Jiashi. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 13708-13717.
- [15] 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 Wentao, Zhao Linlin, Tu Chao. Double-branch multi-attention mechanism based sharpness-aware classification network[J]. Pattern Recognition and Artificial Intelligence, 2023, 36(3): 252-267. (in Chinese)
- [16] 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 Wentao, Gao Yuan, Yuan Heng, et al. Image classification network of gating mechanism[J]. Acta Electronica Sinica, 2024, 52(7): 2393-2406. (in Chinese)
- [17] Wang Yan, Liu Yi, Zhao Shijie, et al. CAMixerSR: Only details need more “attention” [C]//2024 IEEE/CVF Con-

- ference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 25837-25846.
- [18] Yao Zhiyang, Liu Shuyang, Yuan Xiaoyun, et al. SPeCAT: SPatial-spEctral cumulative-attention transformer for high-resolution hyperspectral image reconstruction[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 25368-25377.
- [19] Zhang Shuoxi, Liu Hanpeng, Lin S, et al. You only need less attention at each stage in vision transformers[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 6057-6066.
- [20] Konstantinidis D, Papastratis I, Dimitropoulos K, et al. Multi-manifold attention for vision transformers[J]. IEEE Access, 2023, 11: 123433-123444.
- [21] Shi Baifeng, Darrell T, Wang Xin. Top-down visual attention from analysis by synthesis[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 2102-2112.
- [22] Ma Chenxiang, Wu Jibin, Si Chenyang, et al. Scaling supervised local learning with augmented auxiliary networks[PP/OL]. V2.arXiv(2024-02-27)[2025-07-07]. <https://arxiv.org/abs/2402.17318>.
- [23] 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.
- [24] Huang Gao, Liu Zhuang, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2261-2269.
- [25] 姜文涛, 王鑫杰, 张晟翀. 空间约束注意力机制的图像分类网络[J]. 智能系统学报, 2025, 20(6): 1444-1460.
- Jiang Wentao, Wang Xinjie, Zhang Shengchong. Spatially constrained attention mechanism for image classification network[J]. CAAI Transactions on Intelligent Systems, 2025, 20(6): 1444-1460. (in Chinese)
- [26] Lan Hai, Wang Xihao, Shen Hao, et al. Couplformer: Rethinking vision transformer with coupling attention[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 6464-6473.
- [27] Choromanski K M, Likhoshesterov V, Dohan D, et al. Rethinking attention with performers[PP/OL]. V2.arXiv(2022-11-19)[2025-07-07]. <https://arxiv.org/abs/2009.14794>.
- [28] Zhou Chenlin, Zhang Han, Zhou Zhaokun, et al. QKFormer: Hierarchical spiking transformer using Q-K attention[C]//Proceedings of the 38th Annual Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2024.
- [29] Yu Shikang, Chen Jiachen, Han Hu, et al. Data-free knowledge distillation via feature exchange and activation region constraint[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 24266-24275.
- [30] 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. Piscataway: IEEE, 2024: 16163-16173.
- [31] 杨育婷, 李玲玲, 刘旭, 等. 基于多尺度-多方向 Transformer 的图像识别[J]. 计算机学报, 2025, 48(2): 249-265. Yang Yuting, Li Lingling, Liu Xu, et al. Multiscale and multidirectional transformer-based image recognition[J]. Chinese Journal of Computers, 2025, 48(2): 249-265. (in Chinese)

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、硕士生导师。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: Intuyuanheng@163.com



武景瑞 女, 2000年7月出生于山西省吕梁市。现为辽宁工程技术大学软件学院硕士研究生。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: wujingrui0810@163.com



张晟翀 男, 1973年4月出生于安徽省合肥市。现为高级工程师。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