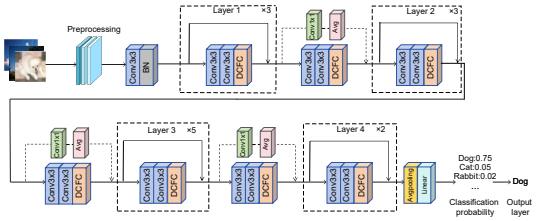
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 5 月, 第 53 卷, 第 5 期

动态通道特征校准的池化残差图像分类网络



姜文涛, 赵文雯*

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105

摘要：针对通道特征交互能力不足，导致难以筛选关键特征的问题，本文提出了动态通道特征校准的池化残差分类网络(pooling residual image classification network with dynamic channel feature calibration, DCPRNet)。DCPRNet 以 ResNet-34 残差网络为基础，首先修改残差网络的首层，将 7×7 卷积替换为 3×3，删除最大池化，提高深层网络特征的表达能力。其次提出分块加权注意力 (chunked weighted attention, CWA) 机制，通过特征分块、评分生成等操作，对通道特征分块的重要性权重进行校准，抑制干扰信息，提高关键区域的贡献度。然后，提出动态通道特征校准(dynamic channel feature calibration, DCFC) 模块，该模块由随机子组置换模块和分块加权注意力操作组成，使通道动态交互，增强跨组信息交流，校准特征权重，筛选关键特征。最后，在动态通道特征校准模块的残差分支上，加入平均池化操作，实现平滑下采样，保留更多有效特征。本文方法在数据集 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 上分别达到了 96.41%、80.36%、96.97%、91.59%、80.61% 的分类准确率，对比现有的主流模型，该网络能提高通道间的信息交互，增强特征表达能力和对关键特征的提取能力，提升模型的分类性能。

关键词：图像分类；注意力机制；特征校准；残差网络

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250309 | CSTR: 32245.14.oee.2026.250309 | 中图分类号：TP391 | 文献标识码：A

姜文涛, 赵文雯. 动态通道特征校准的池化残差图像分类网络 [J]. 光电工程, 2026, **53**(5): 250309

Jiang W T, Zhao W W. Pooling residual image classification network with dynamic channel feature calibration[J]. *Opto-Electron Eng*, 2026, **53**(5): 250309

1 引　言

图像分类是计算机视觉的核心基础任务，是学术研究的热点，长期以来备受关注，广泛应用于医疗诊断、安防监控、遥感影像分析等重要领域[1]。该任务的根本目标是将输入图像自动划分到预定义的语义类别中，是实现机器理解视觉的关键一步。近年来，随着深度学习技术的飞速发展，图像分类性能显著提升，逐步成为诸多智能化系统中的基础支撑技术[2-3]。

传统的图像分类方法严重依赖于研究者的智慧和经验来设计，如 SIFT[4]、HOG[5] 等，其表示能力有限，难以应对真实场景中复杂的视觉变化深度，直到 Krizhevsky 团队[6] 在 2012 年提出 AlexNet，首次使用 GPU 加速训练，使得训练如此深的模型在时间上变得可行。同时使用 ReLU 激活函数，有效缓解梯度消失问题，加速了训练过程的收敛。2014 年 Simonyan 等[7] 提出 VGGNet，结构简单，通过堆叠多个小卷积核 (3×3) 构建深层网络，模块化设计便于扩展，对于大规模图像分类

收稿日期：2025-10-14；修回日期：2026-01-18；录用日期：2026-01-22；网络出版日期：2026-05-25

基金项目：国家自然科学基金资助项目 (61172144)；辽宁省自然科学基金资助项目 (20170540426)；辽宁省教育厅重点基金资助项目 (LJ2017QL034)

*通信作者：赵文雯，1435998539@qq.com

版权所有©2026 中国科学院光电技术研究所

任务增加深度也可以取得显著性能。但是参数量极大，远超于 AlexNet。为解决参数量的问题，Szegedy 等[8]在 2014 年提出了 GoogLeNet，设计创新的 Inception 模块和巧妙使用 1×1 卷积，增加了网络的深度和宽度，提高了计算资源的利用效率，总参数量大幅减少，使用多尺度特征融合，增强模型表达能力。2015 年 He 等[9]提出 ResNet，提出残差连接，残差结构的设计使得梯度可以直接反向传播到浅层，有效缓解了深度网络中的梯度消失问题，同时还提出 ResNet-101、ResNet-152 等深层网络结构。2017 年 Xie 等[10]提出 ResNeXt，保留深度和宽度，使用分组卷积增加并行路径，性能与复杂度权衡更优，设计简洁规整。同年 Zagoruyko 等[11]提出 Wide ResNet，显著增加每一层卷积的输出通道数，更宽的结构通常比更深的结构能更有效地利用参数，获得更好的性能。2018 年，Huang 等[12]提出的 DenseNet 突破了传统残差连接的范式，通过建立层间的密集连接，使每一层都能直接利用所有前层的特征输出，显著提升特征复用效率，减少参数冗余。2018 年 Zhang 等[13]提出的 ShuffleNet，运用通道混洗操作，有效缓解了分组卷积导致的组间信息隔离问题，在轻量化设计中取得了显著成效。然而，该方法本质上仍是一种静态、均匀的特征重排策略，难以根据输入图像的内容特性进行自适应调整，无法在复杂场景下精准地强化关键特征并抑制冗余信息。

上述网络虽在特征提取与分类性能上有所突破，却未能解决特征信息丢失、难以动态选择有效特征的核心问题。针对这一现状，得益于灵长类生物视觉系统对重点区域的选择性关注启发，注意力机制[14]被逐步应用于计算机视觉领域，通过模拟生物视觉的注意力分配逻辑，提升模型对关键特征的捕捉能力。2018 年 Hu 等[15]提出 SENet，通过全局平均池化与全连接层实现通道注意力，提升模型的分类能力。2018 年 Woo 等[16]提出 CBAM，是一种轻量级注意力机制模块，优化特征的通道维度与空间维度表达，强化关键通道特征、抑制冗余通道。2018 年 Wang 等[17]提出 NLNet，通过捕捉全局上下文信息，增强模型对图像中不同部分相互关系的理解，打破局部感受野限制，增强全局建模能力。随后 Hou 等[18]提出一个高效且轻量的 CA 注意力机制，采用水平和垂直两个方向的特征向量拼接起来，送入一个共享的 1×1 卷积和非线性激活函数，进行信息融合和降维。在极小的计算开销下，建模通道间的关系，让网络定位特征。

近年来，Transformer 架构在计算机视觉领域的发展非常迅速。2020 年，Khan 等[19]提出的 ViT 首次将纯 Transformer 架构应用于图像分类，通过将图像分割为固定大小的图像块，并将其转化为序列输入，利用自注意力机制捕捉全局特征依赖。ViT 在大规模数据集上的性能超越传统 CNN，但在中小规模数据集上易过拟合，且对图像局部细节的捕捉能力较弱。2021 年，Liu 等[20]提出的 Swin Transformer 引入滑动窗口自注意力机制，将全局自注意力分解为局部窗口内的注意力计算，在降低计算复杂度的同时，提升了对局部细节的建模能力。汪澳等[21]对 Swin Transformer 模型进行改进，实现低剂量纳米 CT 的少量投影高质量重建，提升图像信噪比。2024 年，Zhu 等[22]提出 Vision Mamba，是一种用于序列建模的新型架构，将输入图像分割再展平为一个序列，然后将二维的图像序列按照两个方向分别重新排列成一维序列，状态空间模型与选择性机制的结合，并将其成功适配到视觉领域。

尽管现有注意力机制与网络架构为特征表达优化提供了有效路径，但仍未彻底解决通道间特征孤立、跨通道信息流动不充分的核心问题[23-24]，制约了模型对关键特征的捕捉与表征能力。针对上述局限，本文提出一种动态通道特征校准池化残差图像分类网络 (DCPRNet)。首先，设计随机子组置换模块，通过双层分组与动态重组的理论设计，解决了静态混洗的固有缺陷，实现更充分的跨通道动态交互。其次，提出分块加权注意力 (CWA) 机制，该机制通过分块处理与全局聚合相结合的策略，实现对交互后特征的自适应权重校准，精准筛选关键信息并抑制冗余。最后，在残差分支中嵌入基于平均池化的下采样结构，以无参方式减少信息丢失并控制计算复杂度。通过上述模块的协同作用，DCPRNet 在增强特征表达能力的同时，有效提升了模型的分类性能与泛化能力。

2 理论推导

本文设计了动态通道特征校准网络 (DCPRNet)，修改网络首层，减少特征信息丢失。随机子组置换操作促进不同子通道间信息交互，再用分块加权注意力机制，计算平均聚合分数，全局权重融合，筛选关键特征，使模型高效融合跨通道信息，提升特征的表达能力。最后将池化残差操作融入到残差块中，减少计算量，保留特征信息，提升网络的泛化能力。

2.1 修改网络首层

在传统 ResNet-34 残差网络的输入层架构中，依次为 7×7 卷积层、批量归一化 (BN) 层、ReLU 激活函数与最大池化层，形成标准化的浅层特征处理流程。通过多组件协同作用完成输入图像的尺寸压缩，并初步提取边缘、纹理等底层视觉特征。具体来看，7×7 卷积层通常搭配较大步幅，其设计初衷在于借助大尺寸卷积核的覆盖范围与大步幅的下采样能力，快速扩大网络感受野，从而高效捕捉图像中较大范围的全局特征。然而，大尺寸卷积核计算成本高，易导致模型在小数据集上过拟合，增加调参的依赖。大卷积核配合大步长可能导致下采样过于激进，丢失细粒度特征，倾向于捕获全局模式，但对局部细节的提取能力较弱，可能影响后续层级特征的构建。以 10 pixel×10 pixel 的小尺寸图像为例如图 1(a) 所示，根据卷积输出尺寸计算公式：

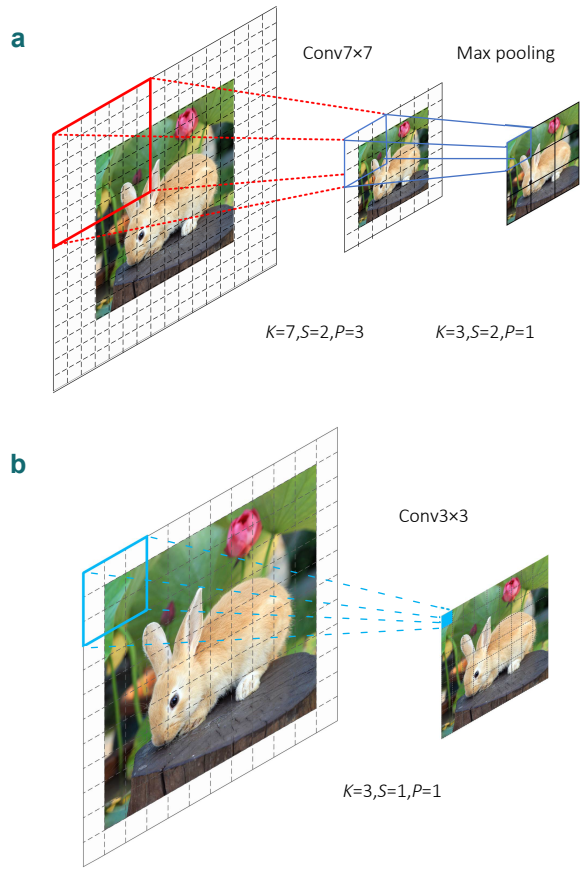


图 1 网络首层卷积对比。(a) ResNet-34 首层积；(b) DCPR 首层卷积

Fig. 1 Comparison of the first layer convolution in the network. (a) The first layer convolution of ResNet-34; (b) The first layer convolution of DCPR

$$N=\frac{W-F+2P}{S+1}, \tag{1}$$

式中：N 表示输出后特征图尺寸；W 为输入特征图像尺寸；F 为卷积核尺寸；P 为填充大小；S 为步幅。根据式 (1) 可完成特征图尺寸计算：对于一个 10 pixel×10 pixel 的输入图像，经该层处理后输出的特征图尺寸会急剧缩减至 3 pixel×3 pixel。压缩过程丢失大量空间细节信息，限制特征图的表达能力，难以支撑深层网络对丰富特征表示的需求，影响整个模型的性能。若将首层替换后，如图 1(b) 所示，输出的尺寸仍为 10 pixel×10 pixel。

该设计能够最大限度保留原始图像的空间结构信息，为后续深层网络提供更充分、更细致的特征输入。小卷积核在相同下采样率下更渐进地压缩特征图，减少信息损失。最大池化会丢失背景与细节信息。删除该层可保留更多原始特征，增强对细微纹理与边缘的捕捉能力，从而提升细粒度[25] 任务的表现。

2.2 动态通道特征校准模块

本文提出动态通道特征校准模块 (DCFC)，其核心创新在于将跨组动态交互与特征动态校准深度融合，通过结构化设计打破静态机制的固有约束。该模块由随机子组置换模块和分块加权注意力 (CWA) 操作紧密结合而成，其中随机子组置换模块通过双层分组与可控随机重组，实现通道关联的动态重构。分块加权注意力则基于特征语义重要性生成动态权重，完成关键特征的精准筛选与增强。二者协同作用，使通道交互与特征校准均能随输入数据分布自适应调整，既解决了传统方法通道交互不足、校准缺乏针对性的问题，又通过有约束的动态设计保障了性能稳定性，最终为后续残差模块提供更优质的特征输入，提升模型泛化能力。

2.2.1 随机子组置换模块

通过子组划分加随机置换的双重机制，在保持分组卷积高效性的同时，实现了跨组特征融合，从而提升特征表达能力和泛化能力如图 2 所示。该模块具体流程如下。设输入特征为 $\boldsymbol{X}$，$\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，首先对输入特征 $\boldsymbol{X}$ 的通道维度进行分组，设组数为 g，设分组函数为 $Group(\cdot)$，所以分组后可以表示为

$$\boldsymbol{X}_i=Group(\boldsymbol{X}), \tag{2}$$

式中：H 和 W 分别为高和宽；$\boldsymbol{X}_i$ 为 $\boldsymbol{X}$ 分组后的通道特征，表示 $\boldsymbol{X}$ 分组的第 i 组特征矩阵，$\boldsymbol{X}_i\in\mathbb{R}^{\frac{C}{g}\times H\times W}$，所以 $\boldsymbol{X}_i$ 包含于 $[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_g]$。

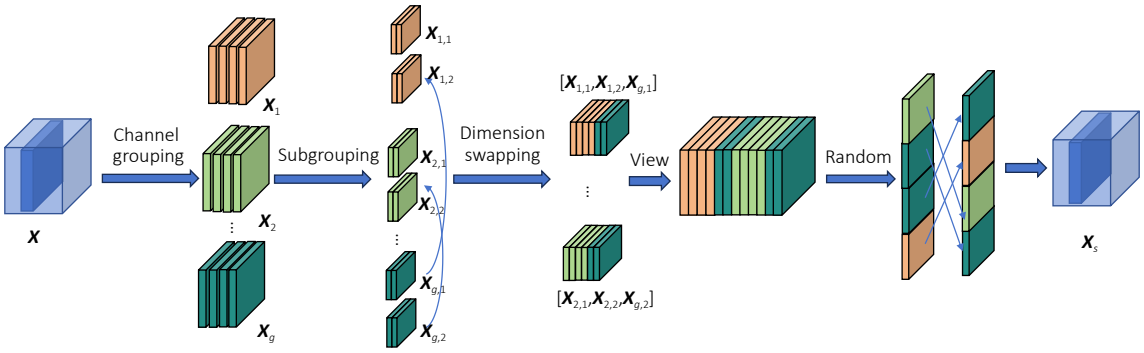


图 2 随机子组置换模块

Fig. 2 Random subgroup permutation module

然后将分组后的通道特征在进一步分子组，其目的是提供更精细的特征交互控制，相比单层分组，双层分组保持了局部特征连贯性。设每一个$\boldsymbol{X}_i$的子组数g_c，分组后的输出特征可以表示为

$$\boldsymbol{X}_{i,j} = Group(\boldsymbol{X}_i), \tag{3}$$

式中：$\boldsymbol{X}_{i,j}$表示第 i 组的第 j 个子组特征矩阵，$j \in [1, g_c]$，$\boldsymbol{X}_{i,j}$的尺寸可以表示为 $\boldsymbol{X}_{i,j} \in \mathbb{R}^{\frac{C}{g_c \times g} \times H \times W}$。其中，经过两次分组，输入特征为 $\boldsymbol{X}$ 的尺寸变为 $g \times g_c \times \dfrac{C}{g_c \times g} \times H \times W$。

其次为了增强通道间交互关系，将维度的第一、二两个维度进行重组交替，即可以表示为

$$\boldsymbol{X}_{\mathrm{p}} = Permute(\boldsymbol{X}, [1,0,2,3,4]), \tag{4}$$

式中：$Permute(\cdot)$为转置函数，经过重组交替后，$\boldsymbol{X}_{\mathrm{p}}$的尺寸变为$g_c \times g \times \dfrac{C}{g_c \times g} \times H \times W$。重组后，来自不同大组的相关子组特征相邻分布，后续卷积层可直接捕捉跨大组的子组关联，解决传统分组卷积仅通过组内卷积提取局部特征，组间信息孤立的核心问题。随后将重组后的所有$\boldsymbol{X}_{\mathrm{p}}$进行通道维度的合并，可以表示：

$$\boldsymbol{X}_{\mathrm{v}} = V(\boldsymbol{X}_{\mathrm{p}}, [C, H, W]), \tag{5}$$

式中：$V(\cdot)$表示对前三个维度的合并；$\boldsymbol{X}_{\mathrm{v}} \in \mathbb{R}^{C \times H \times W}$。将跨大组交互后的子组特征拼接为完整通道维度，使动态交互的结果被保留在特征张量中，避免后续操作破坏交互信息。

最后对$\boldsymbol{X}_{\mathrm{v}}$进行通道维度的随机打乱，可以表示为

$$\boldsymbol{X}_{\mathrm{s}} = random_{dim=0}(\boldsymbol{X}_{\mathrm{v}}), \tag{6}$$

式中：$random_{dim=0}(\cdot)$表示对通道维度的随机打乱函数，经过打乱后，原本归属于同一组的通道会被分散至不同的新组，且重排过程中保留的子组间的语义关联，能有效避免盲目重排可能导致的特征结构破坏，同时在通道层面构建起动态交互基础。上述随机子组置换模块通过双层分组、约束动态重组实现了细粒度跨组交互，为进一步明确理论创新性，下文将与分组卷积领域中经典的 ShuffleNet 通道混洗机制进行针对性理论对比，剖析二者在核心设计与特征学习逻辑上的本质差异。

ShuffleNet 采用单层分组策略，其通道混洗的前提是将输入通道均匀划分为 G 个独立大组，这种分组方式决定了其混洗操作仅能在大组层面实现整体重排，数学表达为

$$\boldsymbol{X}_{\text{shuffle}} = Permute(\boldsymbol{X}_{\text{grouped}}, order = fixed), \tag{7}$$

式中：$order = fixed$表示所有输入图像共享预设的固定通道重排规则。从操作底层逻辑来看，ShuffleNet 的 Permute 操作仅实现维度顺序的整体置换，其作用对象是组级整体维度，核心目标是完成组间通道的均匀分配，无法打破组内子结构的关联边界。而本文模块的 Permute 操作大组与子组两个维度进行交叉转置，作用对象是子组级细粒度维度，通过维度交换使不同大组的同位置子组通道直接相邻，实现跨大组的子组交叉重组。这种操作逻辑的差异，直接导致 ShuffleNet 只能实现粗粒度组间交互，而本文模块可达成细粒度子组交互，为特征学习提供更丰富的关联信息。

ShuffleNet 的通道混洗是与输入无关的静态操作，依赖固定重排规则，即便输入图像的语义类别、背景复杂度不同，也无法调整交互策略，如细粒度分类任务中难以针对性强化关键通道关联，限制了复杂场景适配能力。本文模块通过带语义约束的随机打乱操作 (式 (6)) 实现动态重组，其随机性并非无规则盲目重排，而是受限于子组语义关联约束，确保重组过程中保留子组内通道的固有语义关联性。训练时每批次通道交互路径动态

调整，迫使网络学习不依赖固定通道顺序的鲁棒特征依赖关系，这种动态性本质是一种结构正则化手段，能有效提升模型泛化能力，同时推理阶段可沿用训练学到的最优模式，对不同输入产生自适应响应，弥补了ShuffleNet静态混洗的输入无关性缺陷。

综上，ShuffleNet的单层固定机制在交互粒度、动态适应性和特征优化上存在局限，本文提出的随机子组置换模块通过创新设计实现突破，为通道交互机制提供更优思路。

2.2.2 分块加权注意力

为了实现分块感知、全局决策的自适应特征校准，更精准的筛选出特征的关键信息，本文设计了CWA模块，如图3所示。经过通道随机子组置换后的特征$\boldsymbol{X}_{\mathrm{s}} \in \mathbb{R}^{C\times H\times W}$沿通道维度被均匀分割为$K$个分块，每个分块的通道数为$C_k = C/K$，将输入数据分割为多个分块，独立处理以提升计算效率

$$chunks = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_K\}, \tag{8}$$

式中：$chunks$为特征分块集合，$\boldsymbol{X}_i \in \mathbb{R}^{C_k\times H\times W}$表示第$i$个分块特征，再通过一个3×3卷积层对分块特征$\boldsymbol{X}_i$进行通道维度的压缩，实现通道映射，同时保留空间信息。

$$\boldsymbol{F}_1 = Cond_{3\times3}(\boldsymbol{X}_i) \in \mathbb{R}^{C_k\times H\times W}. \tag{9}$$

该卷积层的输入通道数为C_k，输出通道数为$C_k/4$，卷积核大小为3×3，步长为1。通过减少通道数降低计算复杂度，同时3×3卷积的感受野能捕捉局部空间相关性，为后续注意力权重的生成提供空间上下文信息。压缩后的特征$\boldsymbol{F}_1$经过tanh激活函数进行非线性映射：

$$\boldsymbol{F}_2 = \tanh(\boldsymbol{F}_1) \in \mathbb{R}^{(C_k/4)\times H\times W}, \tag{10}$$

式中：tanh函数的表达式为$\tanh(x) = \dfrac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}}$，其输出范围为[−1,1]。其目的是增强模型对复杂特征的表达能力，通过正负值区分特征的重要性方向，其中正值可视为增强线索，负值可视为抑制线索。

最后通过一个1×1卷积层将特征通道数恢复至分块原始通道数C_k，

$$\hat{\boldsymbol{X}}_i = Conv_{1\times1}(\tanh(Conv_{3\times3}(\boldsymbol{X}_i))). \tag{11}$$

该过程通过先压缩后恢复的通道变换策略，在降低计算成本的同时，结合局部空间感受野和非线性激活，使输出特征既保留了原始分块的通道维度，又融入了空间上下文和通道交互信息，为后续生成精准的分块重要性分数奠定基础。

增强后的分块特征$\hat{\boldsymbol{X}}_i \in \mathbb{R}^{C_k\times H\times W}$需通过一个可学习的评分函数$g_\theta(\cdot)$将其映射为一个标量重要性分数$s_i$，该过程是实现推理动态性的核心。其函数关系可定义为

$$s_i = g_\theta(\hat{\boldsymbol{X}}_i). \tag{12}$$

具体而言，函数$g_\theta(\cdot)$封装了从特征到分数的完整变换，包括全局空间信息压缩和通道信息聚合两个核心步骤，首先通过自适应平均池化(GAP)对增强特征$\hat{\boldsymbol{X}}_i$的空间维度进行全局压缩：

$$\boldsymbol{F}_{\mathrm{gap}} = GAP(\hat{\boldsymbol{X}}_i) \in \mathbb{R}^{C_k\times H\times W}, \tag{13}$$

式中：$GAP(\cdot)$为自适应平均池化对特征图的每个通道单独进行空间维度的平均池化，无论输入空间尺寸(H,W)如何，均输出固定为1×1的空间尺寸。对于第i个分块的第c个通道，其计算方式为

$$\boldsymbol{F}_{\mathrm{gap}}[c,1,1] = \frac{1}{H\times W}\sum_{x=1}^{H}\sum_{y=1}^{W}\hat{\boldsymbol{X}}_i[c,x,y], \tag{14}$$

式中：c为通道索引；(x,y)为空间坐标。将每个通道的

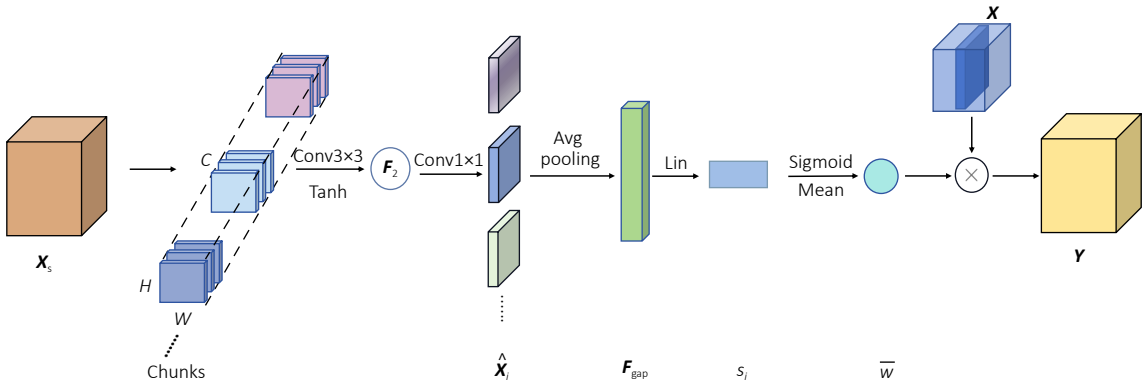


图3 分块加权注意力模块结构图

Fig. 3 CWA module structure diagram

空间信息压缩为一个全局的均值统计量，同时保留通道维度的特征分布，实现空间信息全局化。

为适配后续线性层的输入格式，需将$GAP(\cdot)$输出的四维特征$\boldsymbol{F}_{\text{gap}}$展平为二维向量：

$$\boldsymbol{F}_{\text{flat}} = Flatten(\boldsymbol{F}_{\text{gap}}) \in \mathbb{R}^{C_k}, \tag{15}$$

式中：$Flatten(\cdot)$核心是通过维度重排完成展平，将形状从$(C_k,1,1)$转换为(C_k)，即：

$$\boldsymbol{F}_{\text{flat}} = [\boldsymbol{F}_{\text{gap}}[0,1,1], \boldsymbol{F}_{\text{gap}}[1,1,1], \cdots, \boldsymbol{F}_{\text{gap}}[C_k-1,1,1]]. \tag{16}$$

将通道维度的全局统计量转换为向量形式，为线性层的通道信息聚合做准备。展平后的特征向量$\boldsymbol{F}_{\text{flat}}(\cdot)$经线性层$Lin_i$映射为单个标量分数：

$$s_i = Lin(\boldsymbol{F}_{\text{flat}}) \in \mathbb{R}^1. \tag{17}$$

线性层的权重矩阵为$\boldsymbol{W} \in \mathbb{R}^{1\times C_k}$，偏置为$\boldsymbol{B}_0 \in \mathbb{R}^1$，计算方式为

$$s_i = \boldsymbol{W} \cdot \boldsymbol{F}_{\text{flat}}[B,:]^{\text{T}} + \boldsymbol{B}_0 = \sum_{C=0}^{C_k} \boldsymbol{W}[0,c] \cdot \boldsymbol{F}_{\text{flat}}[\boldsymbol{B},c] + \boldsymbol{B}_0, \tag{18}$$

式中：$Lin(\cdot)$为线性映射函数，$\boldsymbol{W}[0,c]$表示线性层对第c个通道的权重，用于衡量该通道在分块重要性评分中的贡献。至此，式 (11) 至 (16) 共同详细刻画了评分函数$s_i = g_\theta(\hat{\boldsymbol{X}}_i)$的内部实现。该函数$g_\theta$的参数$\theta$即线性层的权重$\boldsymbol{W}$与偏置$\boldsymbol{B}_0$。该过程通过 GAP 实现空间信息的全局压缩，通过线性层实现通道信息的加权聚合，最终将分块特征$\hat{\boldsymbol{X}}_i$映射为一个批次维度的标量分数s_i。这一分数不仅包含了分块内所有通道的全局信息，还通过线性层的权重学习自动区分不同通道的重要性，为后续跨分块的权重聚合提供了精细化的评分依据。

所有分块的分数拼接为矩阵后，通过均值运算与 Sigmoid 激活得到最终权重：

$$S = [s_1, s_2, \cdots, s_i] \in R^K, \tag{19}$$

$$\bar{w} = Sigmoid\left(\frac{1}{K}\sum_{i=1}^{K} s_{:,i}\right) \in R^1, \tag{20}$$

式中：$s_{:,i}$表示第i列分数，均值计算实现跨分块信息聚合；Sigmoid函数将权重约束在$[0,1]$区间。值得注意的是，权重$\bar{w}$可以视为一个关于所有分块分数的函数。由于每个分块分数s_i本身又是输入特征$\boldsymbol{X}$的函数，因此最终权重本质上是一个依赖于整体输入$\boldsymbol{X}$的动态标量，即$\bar{w} = w(\boldsymbol{X})$。这种输入依赖性正是 CWA 模块实现推理动态性的基础。

最终，权重作用于原始输入特征$\boldsymbol{X}$，实现自适应特征调整：

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \bar{w} \in \mathbb{R}^{C\times H\times W}, \tag{21}$$

式中：$\bar{w}$表示将权重从$\mathbb{R}^1$扩展为$\mathbb{R}^{1\times1\times1}$，通过广播匹配原始特征的空间与通道维度，完成逐样本的全局特征加权。

综上所述，CWA 机制通过特征分块、独立评分、全局聚合这一系列精细化操作，构建了一个动态的特征校准流程。该机制使模型能够依据输入内容自适应地筛选关键特征，从而显著增强对细微判别性信息的捕捉能力，为网络后续处理提供更高质量的特征输入。

2.3 残差平均池化

在经典 ResNet 残差结构中，为实现残差块主干输出与残差分支输出的维度匹配，采用步长为 2 的 1×1 卷积调整残差分支特征矩阵维度。未优化特征图空间尺寸，使后续卷积层处理高分辨率图，计算量剧增，主干与残差分支依赖 3×3 局部感受野，难以捕捉跨区域上下文，特征信息丢失，在小规模数据上易过拟合。

为了解决特征表达不完整以及计算量大的问题，本文设计了融合平均池化与批量归一化的增强型残差模块，如图 4 所示。池化路径先以 1×1 卷积将输入特征通道数减半，再进行 2×2 自适应平均池化，输出前加 BN 层与 ReLU 激活确保训练稳定。相比传统残差模块，残差分支使特征图尺寸减半，降低后续卷积计算量，主干参数减少，提高计算效率。平均池化操作能根据不同区域特征分布灵活池化窗口，精准保留关键信息。依据图像整体特征分布扩大感受野，捕获图像的全局上下文信息，进而提升特征表达对不同数据的适配性，增强模型泛化能力。

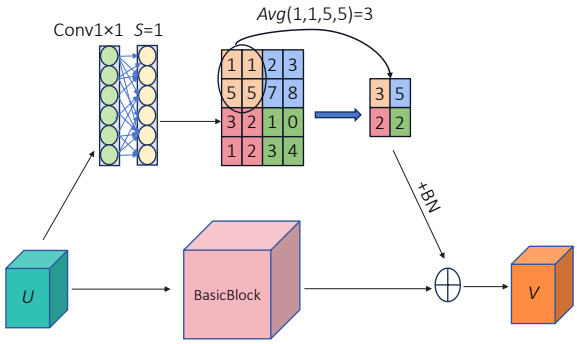


图 4 池化残差模块

Fig. 4 Pooling residual module

2.4 融合动态通道特征校准和池化残差模块

传统 ResNet-34 残差块通过主干两个 3×3 卷积和跳跃连接 1×1 卷积实现特征传递，当输入输出通道维度不一致时，采用 1×1 卷积进行维度匹配，确保残差连接的有效性，如图 5(a) 所示，但存在通道特征交互不足、提取关键特征能力弱的问题。

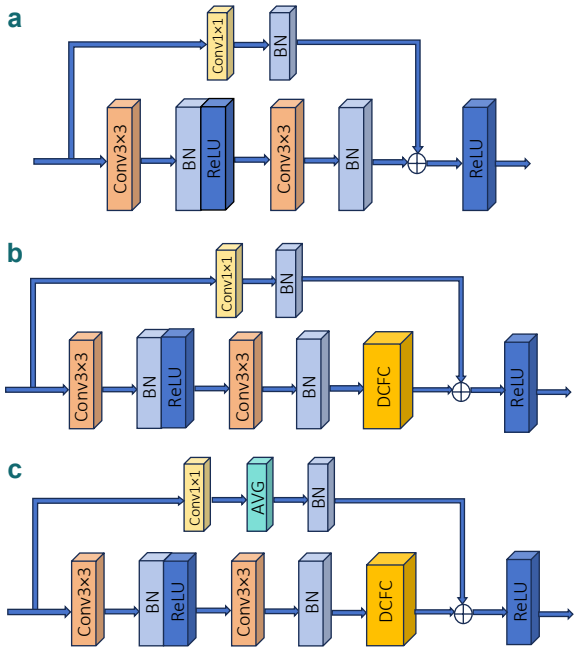


图 5 三种残差块。(a) 常规残差块; (b) 融合 DCFC 残差块; (c) 融合 DCFC 和池化残差块

Fig. 5 Three types of residual blocks. (a) Conventional residual block; (b) DCFC-integrated block; (c) DCFC- and pooling-integrated block

为了提高通道间的信息交流，实现对特征权重的校准，将动态通道特征校准 (DCFC) 机制融入常规残差块中，下文简称为 D-block，其结构如图 5(b) 所示。在主干结构上，其左半部分延续常规残差块的双 3×3 卷积配置，且在卷积层之间嵌入 BN 层与 ReLU 激活函数，实现基础特征的提取与非线性变换。主干右半部分则串联了 DCFC 机制，通过该机制对特征进行动态校准，提升网络对复杂样本的适应能力与泛化性能。将 DCFC 融合到残差块中，能够打破通道特征单一，促进跨组信息交互，优化特征权重分配。

本文在 D-block 基础上，融合动态通道特征校准机制与池化残差模块，如图 5(c) 所示，动态通道特征校准的池化残差模块 (DRPR-block)，采用特征校准、高效下采样、残差增强的策略，在 DRPR-block 的残差分支中嵌入池化残差模块，用 1×1 卷积和 2×2 平均池化替代传统的步长卷积，结合 BN 操作实现无参下采样，避免整数步长偏移导致的边缘特征丢失，最大程度保留细粒度信息。

2.5 网络总体结构

DCPRNet 的网络架构以串联残差块为基础，每个残差块均融合了动态通道特征校准 (DCFC) 模块与池化残差模块，使网络在特征处理过程中既能促进通道信息交互，又能保留有效信息。总体网络结构如图 6 所示。本网络共分为以下 5 个部分：

1) 对图片进行预处理，采用随机裁剪、翻转等策略，

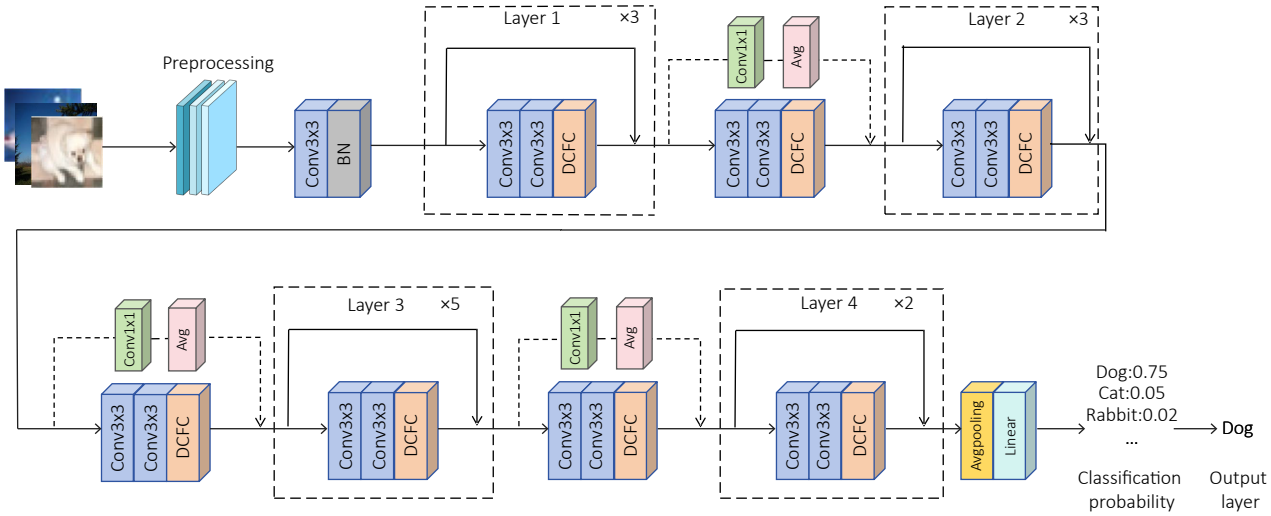


图 6 DCPRNet 总体结构图

Fig. 6 Overall structure diagram of DCPRNet

将原始图像裁剪至目标尺寸，使模型能够学习到目标在不同空间位置与方向下的特征表征，减少因数据分布单一导致的过拟合风险。

2) 对输入层进行优化，将原 ResNet-34 输入层的 7×7 卷积核替换为 3×3 卷积核，在保持等效感受野的同时，显著降低参数量，减少计算冗余并增强特征非线性表达能力，避免导致的边缘、纹理等细粒度特征丢失，确保输入特征保留完整的空间分辨率。

3) 将输入层优化后的特征图进入特征提取层，分别经过 4 个 Layer 层结构 (记为 Layer1~Layer4)，通过随机子组置换实现跨通道交互，增强底层特征多样性，DCFC 机制更细粒度的分组设置模块的动态权重分配，抑制噪声通道并突出目标相关关键特征。

4) 同时各 Layer 层中 DCPRNet 模块的残差分支采用 1×1 卷积、2×2 平均池化和 BN，替代传统步长卷积下采样，在实现空间尺寸压缩的同时保留局部区域全局统计信息，提升深层特征训练稳定性。

5) 最后，特征提取层输出的抽象特征进入分类输出层，通过全局平均池化将特征图压缩，保留通道特征分布，再经过全连接层，完成图像分类任务。整体而言，DCPRNet 通过输入层优化实现特征保留、特征提取层动态校准与高效下采样协同、分类输出层特征聚合，又通过无参池化、通道压缩等设计控制参数量增幅，兼顾模型性能与部署效率。

3 实验及结果分析

3.1 实验环境与数据集

本文实验采用的操作系统为 Ubuntu 18.04。硬件环境为：NVIDIA Tesla P100 显卡、60 GB 内存。软件环境为：Pytorch 1.9.1、CUDA 11.1、CUDNN 8.05、编程语言为 Python 3.8。模型的训练效果与 CNN 神经网络参数的设置密不可分。为保证实验可复现性，实验参数统一设置如下：优化器采用 SGD (动量 0.9，权重衰减 5e-4)，训练 200 个周期 (SVHN 为 150、ImageNet100 为 100)，批量大小 128。初始学习率 0.1，分别在第 60、120、160 周期衰减至原 0.2 倍，损失函数为交叉熵损失。

其中，数据集 CIFAR-10 包含 10 种类别图像，包含常见物体有鸟类、船只、卡车等，每一个类别有 6000 张彩色图像。CIFAR-100 数据集包含 100 种类别图像，包含类别有动物、交通工具等，每个类别含 600 张图像。街牌号数据集 SVHN 包含不同光照条件、不同视角拍摄以及数字等 10 种类别的图像。Imagewoof、Imagenett 是 ImageNet 的精选子集数据集。Flowers-102 细粒度图像分类数据集，包含 102 种常见花卉的图像。STL-10 数据集包含 10 个类别，分别是飞机、鸟、汽车、猫等，训练集为 5000 张，每个类别 500 张，还有 100000 张未标记图像用于无监督学习。ImageNet100 是完整 ImageNet-1K 数据集的一个标准化子集，包含了 100 个类别，继承了 ImageNet-1K 的真实世界图像多样性，包含物体、动物、场景等多种类别。数据集属性如表 1 所示。数据预处理方面，CIFAR-10、CIFAR-100、SVHN 等 32×32 小型数据集，训练集采用 32 pixel×32 pixel 随机裁剪、4pixels 零填充，测试集中心裁剪至 32 pixel×32 pixel 后归一化。Flowers-102、STL-10等中大型数据集统一调整输入尺寸为 224 pixel×224 pixel，训练集采用 0.08 比例随机裁剪并缩放，测试集短边缩放至 256 pixels 后中心裁剪至 224 pixel×224 pixel。所有数据集均执行 p=0.5 随机水平翻转和通道归一化。在后续的实验中，对于不同的模块使用不同的数据集进行验证。

表 1 实验数据集
Table 1 Experimental dataset

Dataset	Size/pixels	Class	Trainset	Testset
CIFAR-10	32×32	10	50000	10000
CIFAR-100	32×32	100	50000	10000
SVHN	32×32	10	73257	26032
Imagewoof	224×224	10	9025	3929
Imagenette	224×224	10	9469	3925
Flowers-102	96×96	102	14740	15153
STL-10	96×96	10	5000	8000
ImageNet100	224×224	100	130000	5000

3.2 参数对网络性能的影响

3.2.1 模块的数量和位置对网络性能的影响

动态通道特征校准 (DCFC) 模块是 DCPRNet 实现特征校准的核心，通过动态分组与随机子组置换，可增强特征跨通道交互能力、抑制冗余信息。DCFC 模块嵌入的位置与数量直接影响网络性能，嵌入位置偏浅层，易受底层纹理噪声干扰，导致校准出现偏差。嵌入位置过深，抽象特征已丢失细粒度信息，模块难以发挥校准价值。嵌入数量过多，会显著增加参数量与计算开销，引

发小规模数据集过拟合风险。为明确 DCFC 模块位置与数量对网络分类性能的影响，插入位置和数量结构如图 7 所示。在其它训练参数确定的条件，使用 CIFAR-10、CIFAR-100 和 SVHN 数据集在 ResNet-34 网络模型中进行训练和验证。不同位置与数量 DCFC 模块的分类准确率结果如表 2 和图 8 所示。

从实验结果上来看，位置 2 是 DCFC 模块的最优嵌入位置，在三类数据集上均取得最高分类准确率，最高分别提升 1.81%、1.18%、0.32%。从特征处理逻辑看，位置 2 将 DCFC 模块嵌入残差块主干末端，此时主干经两层 3×3 卷积完成特征提取 (第一层捕捉浅层细节，第二层挖掘目标语义关联)，且经 BN1、BN2 标准化大幅降低噪声，为 DCFC 提供高质量校准输入。位置 1 作为未嵌入 DCFC 模块的基线组，仅完成了输入层的基础优化，保留了细粒度特征但未解决通道交互不足的核心问题。靠输入层优化虽能为特征提取奠定基础，但缺乏动态校准机制时，通道特征仍处于相对孤立状态，冗余信息无法有效抑制，关键特征的辨识度不足，导致分类准确率低于其他组别。为后续 DCFC 模块的有效性验证提供了明确参照。DCFC 通过随机子组置换增强通道交互、分块加权注意力动态调整权重，既能强化目标关键特征、抑制冗余背景，又可避免位置 4 或 5 因过早嵌入出现受原始噪声干扰、特征未充分抽象导致校准不彻底的问题，最终实现特征质量与分类性能最优提升。对比其他位置，位置 4 因模块介入过早，未过滤的原始特征易受背景噪声干扰，三类数据集准确率最低。位置 5、6、7 多位置嵌入会引入计算冗余，且多次权重调整可能过度抑制有效特征，准确率低于位置 2。位置 3、5 校准时机早于位置 2，特征未完成深层抽象，校准无法充分覆盖目标语义信息，性能同样较低。

综上，位置 2 能最大程度发挥 DCFC 模块的动态校准价值，实现特征提取与优化的协同增效，显著提升网

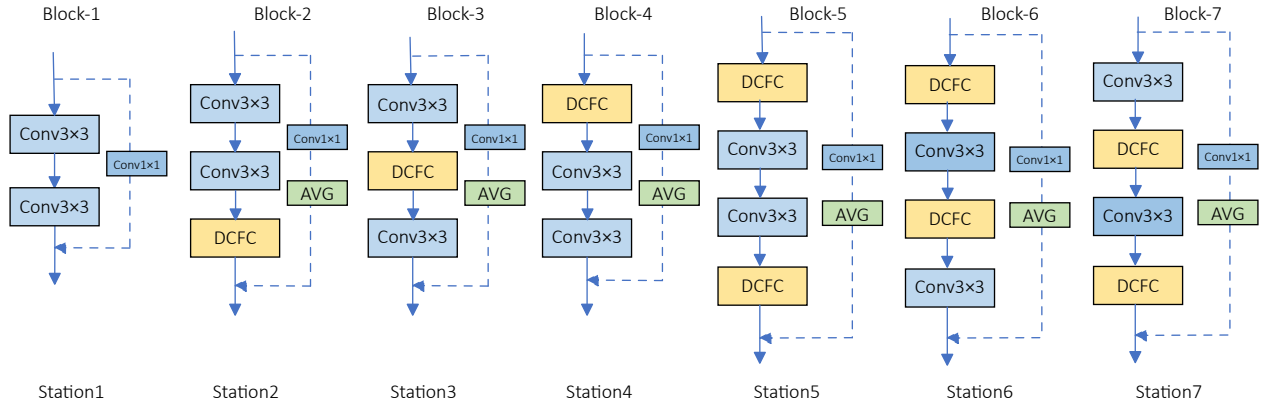


图 7 位置和数量嵌入结构图

Fig. 7 Structure diagram of position and quantity embedding

表 2 不同位置在三种数据集上的准确率/%

Table 2 Accuracy rates of different positions on three datasets/%

Station	CIFAR-100	CIFAR-10	SVHN
1	78.21	95.29	96.62
2	**80.36**	**96.41**	**96.97**
3	78.77	95.40	96.78
4	78.55	95.23	96.65
5	79.83	95.90	96.89
6	78.92	95.62	96.72
7	79.58	95.51	96.82

注：加粗字体表示各列最优结果，下表同。

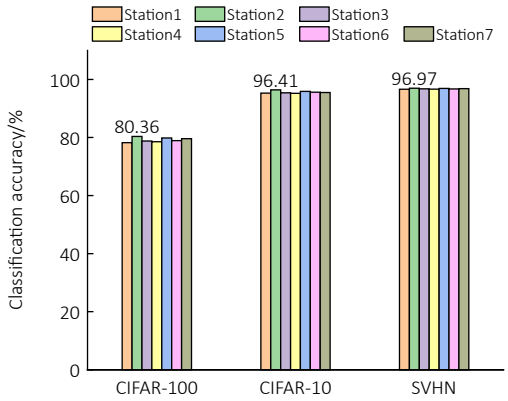


图 8 DCPR 模块不同位置与数量的准确率

Fig. 8 Accuracy rates of the DRPR module

络分类性能与泛化能力。

3.2.2 不同分组数对实验的影响

DCFC 模块的性能依赖随机子组置换的双层分组逻辑 (大组划分+子组划分)，大组分组数 (groups) 决定跨组交互的基础粒度，子组分组数 (subgroups) 影响交互的精细度，二者协同决定通道特征的分布均衡性与信息关联性。为全面验证分组参数对网络性能的影响，本实验分两阶段开展，先固定子组分组数 (*subgroups*=2)，验证大组分组数的最优区间，再补充子组分组数鲁棒性实验，排除子组参数对核心结论的干扰，确保分组参数设置的严谨性。

本实验基于 3.2.1 节中已验证的 DCFC 模块最优嵌入位置 (位置 1)，在数据集 CIFAR-100、CIFAR-10、SVHN 上开展对比试验。针对网络中的 4 个 Layer 层 (Layer1~Layer4，通道数分别为 64、128、256、512) 设计 5 组不同分组方案，分别为方案 1 (1-1-1-1) 不进行分组、方案 2 (2-4-8-16) 每组 32 通道、方案 3 (4-8-16-32) 每组 16 通道、方案 4 (8-16-32-64) 每组 8 通道、方案 5 (16-32-64-128) 每组 4 通道，实验结果如表 3 和图 9 所示。

表 3 不同分组数在三种数据集上的准确率/%

Table 3 Accuracy rates of different grouping numbers on three datasets/%

Scheme	CIFAR-100	CIFAR-10	SVHN
1	78.52	95.86	96.23
2	78.95	95.93	96.68
3	**80.36**	**96.41**	**96.97**
4	79.12	96.03	96.75
5	78.88	95.84	96.51

根据上述实验结果来看，分组参数并非越大越好或越小越好，而是存在最优区间，方案 1 未进行有效分组，通道间仍处于信息隔离状态，随机子组置换模块无法发挥跨通道交互作用，导致模型对复杂数据的特征表达能力不足，在 CIFAR-100 数据集上准确率最低。方案 3 在三类数据集上均取得最优性能，该方案中每组 16 通道确保每组内特征具有足够多样性，避免方案 4、5 因每组通道数过少导致的特征碎片化、特征缺乏差异性与关联性的问题，通过合理分组数实现高效跨组信息交互。方案 3 的优势源于分组粒度与 Layer 层通道数的适配性和随机子组置换策略的协同作用，分组粒度适配，浅层 Layer1、Layer2 需保留细粒度特征，4~8 组的分组数可

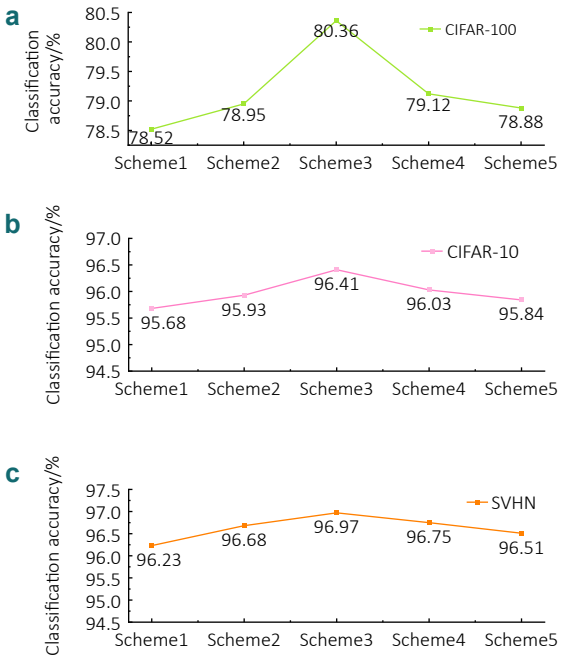


图 9 不同分组数在三种数据集上的准确率。(a) CIFAR-100；(b) CIFAR-10；(c) SVHN

Fig. 9 Accuracy rates of different grouping numbers on the three datasets. (a) CIFAR-100; (b) CIFAR-10; (c) SVHN

避免过度拆分导致的细节丢失。深层 Layer3、Layer4 需强化抽象特征交互，16~32 组的分组数可提升跨通道关联挖掘能力。随机子组置换通过二次分组和随机打乱，进一步减少模型对通道顺序的敏感性，使不同组间通道既能充分交互，又不丢失基础关联，有效提升模型泛化能力，在 SVHN 数据集上，较方案 2 准确率提升 0.29%。

为排除子组分组数 (subgroups) 影响大组分组数最优结论的潜在干扰，本阶段固定大组分组数为核心实验验证的最优值 (方案 3)，测试不同子组分组数对性能的影响，实验结果如表 4 所示。

表 4 不同子组分组数在三类数据集上的准确率/%

Table 4 Accuracy of different numbers of subgroups on three datasets/%

Subgroups	CIFAR-100	CIFAR-10	SVHN
1	78.84	94.89	96.25
2	**80.36**	**96.41**	**96.97**
4	79.73	95.78	96.34

在子组分组方面，测试 *subgroups*=1、2、4 的性能差异，*subgroups*=1 因无实质子组划分，仅单层大组分组无法实现子组交错，CIFAR-100 数据集准确率较低。*subgroups*=4 因子组过细，组内通道数过少，特征

关联性减弱且计算冗余增加，导致分类准确率为低于 *subgroups*=2。而 *subgroups*=2 既能通过从大组到子组维度置换实现不同大组的子组交错，充分打破通道隔离，又能确保子组通道数足够，保留特征关联性，与大组分组协同实现最优跨通道交互效果，且参数量与计算效率均衡。综上，DCFC 模块大组最佳分组为 4-8-16-32 (每组 16 通道)、子组最佳分组为 *subgroups*=2，二者协同作用可最大化发挥跨通道交互与动态校准功能，为 CWA 模块提供优质输入特征，显著提升网络分类性能。

3.2.3 不同网络层数性能对比

为探究网络层数对 DCPRNet 特征提取的影响，本实验以 ResNet-18、ResNet-34、ResNet-50、ResNet-101 为基线，构建对应 DCPRNet 变体，在 CIFAR-100 数据集上开展对比实验，结果如表 5 和图 10 所示。

表 5　四种网络的参数对比

Table 5　Comparison of parameters of four networks

Networks	Accuracy/%	Params/M	Xentropy
DCPRNet-18	78.92	11.60	0.013
DCPRNet-34	**80.36**	**29.42**	**0.008**
DCPRNet-50	79.15	35.50	0.011
DCPRNet-101	79.63	44.40	0.009

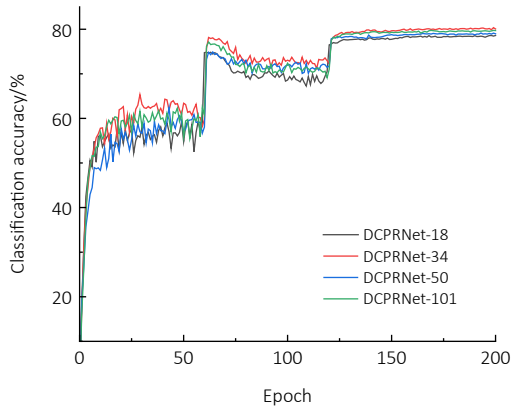


图 10　四种网络准确率对比图

Fig. 10　Comparison of accuracy rates of four networks

在此情况下，DCFC 模块因特征抽象程度不足而难以有效建立跨通道依赖关系，CWA 机制也因缺乏充分的语义基础而无法精准筛选关键特征，最终导致模型欠拟合、性能下降。DCPRNet-50 与 DCPRNet-101 层数深、参数多，虽理论上可提升特征抽象能力，但 CIFAR-100 样本有限易过拟合，且参数量冗余增加存储与时间成本，准确率均未超 DCPRNet-34。DCPRNet-34 层数与参数量适配核心模块需求，既通过足够层数激活随机子组置换跨通道交互与 CWA 动态权重分配功能，又借 3×3 卷积替代 7×7 卷积、无参平均池化下采样控制参数量，避免欠拟合与冗余。

综上，过浅的网络性能不够好，根本原因是用来提取特征的层数太少，没办法捕捉到数据里深层次的语义关联，导致在分类任务里很容易出现特征提取不充分的情况。太深的网络又会带来很高的成本，这种成本不只体现在参数量和计算量的大幅增加，还容易因为梯度弥散、冗余特征堆积，引发模型过拟合等问题。DCPRNet-34 的平衡设计验证了动态通道校准与池化残差模块有效性，从而实现最佳的性能和效率。

3.2.4 随机性对网络稳定性的影响

随机子组置换模块是 DCPRNet 实现动态通道交互的核心创新之一，其通过在训练阶段引入可控的随机性，打破通道间的固定连接模式，从而增强模型的泛化能力。为了验证随机子组置换模块带来的性能提升是稳定且可靠的，而非增加结果的不确定性，本实验进行了严格的随机种子鲁棒性测试。实验使用 5 个不同的随机种子 24，42，123，3407，2024 分别初始化并训练 DCPRNet 模型。这些种子覆盖了常见的实验设置，确保了测试的广泛性。训练在 CIFAR-100 和 CIFAR-10 数据集上进行，记录每个随机种子下模型在测试集上的最高分类准确率，并计算其均值与标准差，结果如表 6 所示。

表 6　不同随机种子下的模型准确率/%

Table 6　Model accuracy under different random seeds /%

Random seeds	CIFAR-100	CIFAR-10
24	80.28	96.35
42	80.41	96.39
123	80.35	96.41
3407	80.30	96.38
2024	80.39	96.40

从表 6 的实验结果可以看出在不同随机种子下，DCPRNet 在 CIFAR-100 和 CIFAR-10 数据集上的分类准确率波动极小。5 个随机种子的准确率均值分别为 80.35% 和 96.39%，表明模型的最终性能不受随机种子选择的显著影响。准确率的标准差分别为 0.05% 和 0.02% 远低于 0.1%，这一微小的波动范围充分证明，随机子组置换模块引入的随机性并未导致训练结果的不确定性，模型能够稳定地收敛到性能优良且一致的解。值得注意

的是，该操作在已完成重组交替的特征之上，此时跨组语义信息已实现初步融合，随机打乱在此优化基础上产生的扰动具备方向性约束，有效保障通道动态交互，降低完全随机可能引发的性能震荡风险。同时这一设计可有效防止网络对固定通道关联模式产生过拟合，促进模型收敛至更平坦且泛化能力更强的损失区域，进一步保障了动态交互带来的收益稳定性。

综上所述，随机种子鲁棒性实验证实了 DCPRNet 中随机性设计的稳健性，消除了对其可能引发性能不稳定的疑虑，该实验结果强有力地论证了随机子组置换模块的有效性和可靠性。

3.2.5 不同注意力对网络参数的影响

为验证本文提出的分块加权注意力 (CWA) 模块对 ResNet-34 网络性能提升的有效性，明确其在特征校准能力上的优势，选取 SE (squeeze and excitation networks)、ECA[26] (efficient channel attention)、CA (coordinate attention)、CBAM (convolutional block attention module) 四种已广泛应用的注意力机制作为对比对象，在多类数据集上开展性能评估实验，以覆盖不同尺寸、不同任务场景的图像分类需求。

将五种注意力模块分别嵌入网络特征提取层 (Layer1~Layer4)，确保各模块嵌入位置、数量完全一致。数据预处理根据数据集尺寸适配调整，针对 CIFAR-100 (32 pixel×32 pixel)、SVHN (32 pixel×32 pixel) 等小尺寸数据集，沿用本文优化的 ResNet-34 首层。针对 Flowers-102 (224 pixel×224 pixel)、STL-10 (96 pixel×96 pixel) 等中大型尺寸数据集，将首层恢复 7×7 卷积并添加最大池化，以适配大尺寸图像的特征提取需求，确保实验变量仅为注意力模块差异。测试数据集选取覆盖小尺寸通用图像 (CIFAR-100)、街景图像 (SVHN)、花卉细粒度图像 (Flowers-102)、提供标注与无标注数据的小样本图像 (STL-10) 及犬类图像 (Imagewoof)，以分类准确率作为核心评价指标，全面验证不同注意力模块在多样数据场景下的适配能力，实验结果如表 7 和图 11 所示。

综合实验数据，嵌入 CWA 模块的 ResNet-34 在所有数据集上均获最高分类准确率，验证了其特征校准有效性与场景适配性。在 STL-10、Flowers-102、CIFAR-100、Imagewoof、SVHN 数据集上，较次优的 CBAM 均有提升。尤其在 Flowers-102 这类类别多、类间相似度高的细粒度数据集上，准确率提升显著，表明 DCPRNet 的 DCFC 模块和 CWA 机制具有优异的特征校准能力。它能有效地从高度相似的花朵图像中区分出微小的局部特征 (如花瓣纹理、颜色梯度)，这是传统 ResNet 难以做到的。对比其他注意力模块，SE、ECA 仅调通道权重，缺乏空间与动态交互。相比之下，DCPRNet 的 DCFC 模块通过随机子组置换和分块加权，实现更复杂的动态跨通道信息交流。CA 虽引入空间信息，但特征判别力未显著提升，小样本场景易受噪声干扰。CBAM 虽结合通道与空间注意力，但串行执行易产生干扰，校准精度不及 CWA。CWA 注意力通过分块再卷积、全局平均池化、均值聚合、Sigmoid 归一化一系列操作实现全局决策式特征校准。3×3 卷积捕捉局部空间相关性，适配细粒度特征提取。均值聚合与 Sigmoid 抑制异常值，提升校准稳定性。通过随机子组置换促进

表 7　不同注意力机制在 5 种数据集上的准确率/%

Table 7　Accuracy rates of different attention mechanisms on five datasets/%

Attention	CIFAR-100	SVHN	Imagewoof	Flowers-102	STL-10
+ECA	79.58	96.85	79.85	77.15	90.13
+SE	79.25	96.78	79.72	76.89	89.82
+CA	79.83	96.89	80.10	78.56	90.37
+CBAM	80.02	96.93	80.35	78.91	90.45
+CWA	**80.36**	**96.97**	**80.60**	**80.62**	**91.91**

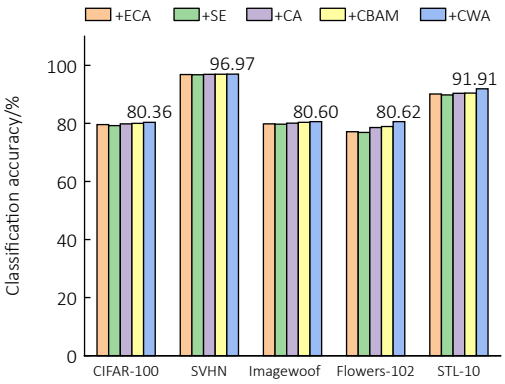


图 11　不同注意力机制对分类准确率影响

Fig. 11　The impact of different attention mechanisms on classification accuracy

跨通道交互，适配不同尺寸数据集，高效完成特征重校准，助力网络分类性能提升。

3.2.6 权重聚合方式对比

为了验证权重聚合方式对实验结果的影响，基于DCPRNet完整模型替换权重聚合方式，构建4组对比模型。其中第一组为分块评分取均值，本文原始设计，全局平均分块评分，Sigmoid约束权重至[0,1]、第二组分块评分取最大值、第三组分块评分加权求和、第四组分块评分Softmax拼接不做全局均值，按分块维度单独归一化，忽视分块间全局关联性，四组实验分别在CIFAR-100、CIFAR-10、SVHN数据集上测试性能，准确率如表8和图12所示。

表 8　四种权重聚合方式准确率/%

Table 8　Accuracy of four weight aggregation methods/%

Aggregation methods	CIFAR-10	CIFAR-100	SVHN
Mean value	**96.41**	**80.36**	**96.97**
Max value	95.12	78.05	95.33
Chunk-wise summation	96.25	80.12	96.75
Independent normalization	95.58	78.89	95.87

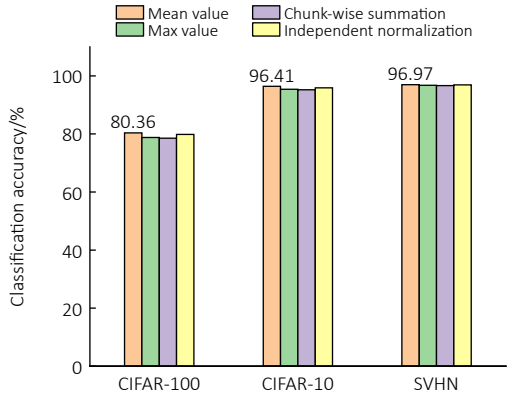


图 12　四种权重聚合方式对准确率影响

Fig. 12　The impact of four weight aggregation methods on accuracy rate

在DCPRNet的分块加权注意力模块中，权重聚合环节的分块评分均值(mean value)和Sigmoid归一化是实现精准筛选关键特征的核心。最大值(Max value)仅选取所有分块评分中的最大值作为全局权重基准，极易受到个别包含噪声或偶然性强特征的分块影响，导致权重分配不稳定且具有高方差。模型因此会过度聚焦于可能是偶然出现的局部特征，而忽视其他分块中稳定、互补的判别性信息，最终损害了特征的鲁棒性与模型的泛化能力，这在三个数据集上的准确率较低且波动较大。

加权求和(Chunk-wise summation)与独立Softmax归一化取得了接近但略逊于均值法的结果。加权求和法对所有分块评分进行简单累加，其权重未经过Sigmoid压缩的线性组合。虽然考虑了所有分块，但未对评分进行平滑约束，使得权重容易偏向数值较大的分块，且输出尺度不稳定，可能在前向传播中引发梯度异常。独立Softmax法在每个分块内部进行归一化，使得分块之间失去可比较的全局尺度。虽然每个分块内部的相对重要性得以保留，但模型失去了对全局视角评估的能力，破坏了特征校准所需的整体观。均值加Sigmoid的方案在抗干扰性、决策性以及局部到全局信息整合方面达到了最佳平衡。这一设计保障了CWA模块能够稳定、精准地筛选并增强关键特征，抑制噪声，是DCPRNet实现卓越分类性能的重要环节，也为注意力机制中的权重聚合提供了稳健且高效的解决方案。

3.2.7 不同网络对分类准确率的影响

为全面验证DCPRNet在图像分类任务中的综合性能优势，选取12种覆盖不同技术路线的主流网络作为对比模型，ResNet-34、SENet-34、DenseNet、GhostNet[27]、MobileNetV2[28]、Wide ResNet 28-10[11]、EfficientNet[29]、Couplformer[30]、Swin-T[20]、FAVOR+[31]、MMA-CCT-7/3[32]、ATO (auto-train-once)[33]在CIFAR-10、CIFAR-100、SVHN、Imagewoof、Imagenette和ImageNet100六类具有代表性的数据集上开展分类准确率对比实验。在本文针对不同网络的分类准确率对比实验中，数据来源遵循以下原则以保障可靠性与一致性，对于未公开源代码的对比网络，实验数据优先选取其对应研究论文中所公布的实验结果，而对于论文已提供开源代码的对比网络，则通过复现其开源代码来获取分类准确率数据，如表9所示。

从实验结果来看，DCPRNet在六类数据集上的分类准确率均优于所有对比网络，其性能优势源于自身核心创新模块的协同作用，且在大规模复杂数据集ImageNet100上的优异表现，充分验证了模型更强的泛化能力。以传统残差网络ResNet-34为基线，DCPRNet通过输入层优化保留细粒度特征，DCFC模块打破通道固化实现跨组信息交互，池化残差模块以平均池化替代步长卷积减少下采样信息丢失，三者共同解决传统ResNet-34的特征表达局限，使得DCPRNet在CIFAR-100与Imagenette数据集准确率均提升较多，在ImageNet100数据集上也实现78.5%至80.01%的性能

表 9 不同网络在 6 种数据集上的分类准确率/%

Table 9 Classification accuracy rates of different networks on six datasets/%

Model	CIFAR-10	CIFAR-100	SVHN	Imagewoof	Imagenette	ImageNet100
ResNet-34	95.29	78.21	96.62	79.36	87.20	78.50
DenseNet	94.24	78.85	94.97	77.97	87.72	79.98
GhostNet	94.92	77.16	93.86	78.22	87.83	—
EfficientNet	94.01	75.96	93.32	77.95	88.06	—
MobileNetV2	93.37	68.09	—	—	—	—
Wide ResNet 28-10	95.83	79.50	95.21	78.71	88.34	79.3
SENet-34	95.22	73.22	87.06	—	—	79.54
MMA-CCT-7/32	94.74	77.50	94.26	—	—	—
Couplformer	93.54	73.92	94.26	77.89	87.91	—
FAVOR+	91.42	72.56	93.21	77.57	88.16	—
Swin-T	94.46	78.08	—	—	—	—
ATO	94.51	78.54	95.51	80.19	86.67	—
DCPRNet (Ours)	**96.41**	**80.36**	**96.97**	**80.61**	**91.59**	**80.01**

跨越，展现出对大规模、高多样性数据的适配能力。

相较于基于 ResNet-34 改进且加入通道注意力的 SENet-34，DCPRNet 的 CWA 模块通过分块再卷积与全局均值聚合实现通道级的动态权重校准，在 CIFAR-100 数据集准确率提升 1.40%，在 ImageNet100 数据集上较 SENet-34 提升 0.47%，充分验证动态加权注意力对单一通道注意力的优化价值，以及在复杂数据场景下的特征校准有效性。

在轻量化网络对比中，GhostNet 依靠特征复用、MobileNetV2 采用深度可分离卷积以降低参数量，但为追求效率牺牲特征交互充分性，导致准确率显著低于 DCPRNet。在 CIFAR-100 数据集，DCPRNet 较 MobileNetV2 和 GhostNet 均有提升，在中小规模数据集的性能差距可知，DCPRNet 的动态校准与无参池化设计在保障效率的同时，实现了更高的特征利用效率，避免了轻量化网络常见的性能妥协，使其在大规模数据集上仍能保持优势。

高效缩放网络 EfficientNet 通过宽度、深度、分辨率的复合缩放平衡性能与效率，但其缩放策略未针对通道交互不足问题优化，在 CIFAR-100 数据集准确率低于 DCPRNet，且未在 ImageNet100 上展现出超越 DCPRNet 的性能，证明仅靠网络规模缩放难以弥补核心特征交互机制的缺陷。WideResNet 28-10 通过拓宽通道提升特征多样性，但因参数量较高且缺乏动态校准机制，在 CIFAR-100 和 ImageNet100 上均低于 DCPRNet，说明单纯增加通道宽度无法替代动态特征校准对复杂数据的适配作用。

此外，MMA-CCT-7/32、Couplformer、Favor +等特定架构网络，受限于自身设计缺陷，在各类数据集上的准确率均低于 DCPRNet，其中 FAVOR +在 CIFAR-100 数据集准确率较 DCPRNet 低 7.8%，其在大规模数据集上的性能短板进一步凸显了 DCPRNet 核心机制的优越性。Transformer 类网络 Swin-T 通过滑动窗口自注意力平衡全局与局部建模，在 CIFAR-100 数据集准确率达 78.08%，但仍低于 DCPRNet，且其依赖的注意力机制在局部特征捕捉上不及 DCPRNet 的卷积与动态校准结合架构，尤其在 Imagewoof 等中尺寸数据集和 ImageNet100 这类大规模数据集上，DCPRNet 的准确率显著更高，体现出其对图像局部细节、细粒度特征以及大规模复杂特征的更优捕捉能力。

ATO 提出了一种控制器网络引导的自动剪枝训练范式，尽管 ATO 在自动化剪枝和结构优化方面表现出色，但在 CIFAR-10、CIFAR-100、SVHN 等数据集上的分类准确率仍低于 DCPRNet，这一差距源于 ATO 更侧重于网络结构的自动稀疏化与参数效率优化，而 DCPRNet 则通过 DCFC 和池化残差模块，在特征交互、细粒度保留和全局上下文建模方面进行了更深层次的增强，使其能够更好地适应大规模数据集的复杂特征分布，泛化能力更具优势。

在工业界实际部署场景中，模型的推理效率与内存访问开销直接决定其适用性，而 DCFC 模块中的置换和注意力操作可能引入额外内存访问成本，导致理论

FLOPs 与实际运行速度存在偏差。MAC (Multiply-accumulate) 代表网络的理论计算复杂度，是评估模型计算开销与硬件资源占用的重要指标。为此，本文在 NVIDIA Tesla P100 GPU 平台上，引入 MAC 作为关键指标，结合批大小 *BS*=1 时的单张推理延迟、*BS*=128 时的吞吐量，对 DCPRNet 及对比模型开展推理效率基准测试，结果如表 10 所示。

由表 10 可知，DCPRNet 的推理延迟仅略高于基线 ResNet-34，这一差异源于 DCFC 模块新增的无参化通道重组与注意力校准操作，虽增加了少量计算步骤，但未引入额外卷积或全连接层，因此延迟增幅控制在 17% 以内。而 Swin-T、EfficientNet 延迟显著更高，核心是 Transformer 架构的多头自注意力机制存在大量非连续内存访问，EfficientNet 的复合缩放策略则导致网络宽度与深度同步增加，均大幅提升了计算与内存开销。吞吐量 (throughout) 方面，DCPRNet 在 *BS*=128 时达到 370.4 f/s，虽低于 ResNet-34，但远超 Transformer 类模型和宽网络，这是因为 DCPRNet 的池化残差模块采用 1×1 卷积和平均池化的无参下采样设计，在压缩特征图尺寸的同时未降低并行计算效率，而 Swin-T 的滑动窗口注意力需频繁进行窗口划分与合并，Wide ResNet 28-10 的通道数拓宽则导致单批次计算量激增，均严重制约了吞吐量提升。

通过 MAC 值量化分析，DCPRNet 总 MAC 为 6.5 G，仅比 ResNet-34 增加 35.4%，远低于 Swin-T 和 Wide ResNet 28-10。拆解来看，随机子组置换靠无参化维度重组降低开销，CWA 注意力通过分块压缩与 3×3 卷积降维减少内存占用，池化残差模块因平均池化无需中间存储效率更高。而 Swin-T 需存储全局关联矩阵、Wide ResNet 28-10 宽通道设计增加特征图占用，二者 MAC 值显著更高。

综上，DCPRNet 的效率优异，核心得益于三大模块的协同优化随机子组置换采用无参化 Permute 操作，内存访问模式连续，避免了分组卷积常见的非连续内存读取开销。CWA 模块通过分块处理和局部卷积替代全局注意力，在精准校准特征权重的同时降低了通道维度的计算与内存冗余。池化残差模块采用无参平均池化替代步长卷积，减少了下采样过程中的内存读写次数。三者共同作用，使 DCPRNet 在引入动态通道校准功能的前提下，仍保持了与基线模型接近的推理效率，实现了性能与部署效率的平衡。

3.2.8 可视化分析

为从特征捕捉与利用的直观层面，进一步验证 DCPRNet 的特征重校准有效性，选择 MobileNet、ShuffleNet、ResNet-34 作为对比模型，在 CIFAR-10、CIFAR-100、Imagewoof、Imagenette 数据集中随机选取 5 幅不同类别图像为实验样本，如图 13(a) 所示，对各网络最后一个卷积层的特征响应进行热力图分析，直观对比不同网络对关键特征的聚焦能力与细粒度信息捕捉效果，从各网络热力图表现来看，不同架构网络的特征提取精度存在显著差异，MobileNet 的热力图如图 13(b) 所示，呈现零散、模糊的特征响应，效果欠佳。其轻量级架构在追求高效的同时，深层特征提取能力受限，导致模型难以形成对目标整体结构的连贯性理解，只能激活部分边缘或纹理等局部特征，无法精准聚焦于判别性最强的核心区域。ShuffleNet 的热力图如图 13(c) 相比 MobileNet 有所改善，但高响应区域仍分散且缺乏针对性。这也印证了只有通道打乱这一核心机制的局限性，其打乱策略是静态和均匀的，对所有通道一视同仁，这种均质化的、非自适应的通道交互方式，虽能提升信息流动性，但难以实现对关键特征的精准筛选与增强。ResNet 热力图如图 13(d) 所示，其能捕捉目标的整体轮廓，显示出深层网络的优势。但其响应区域较为粗糙和宽泛，会将目标物体的大部分区域均等地激活。反映了传统残差网络在缺乏有效注意力机制时的问题，能够提取强语义特征，但缺乏对细粒度关键部位的精准定位能力，特征校准能力不足。DCPRNet 的特征聚焦与细节保留效果最优如图 13(e)，虽然 DCPRNet 的热力图呈现红色的关键区域能精准聚焦细粒度细节、抑制冗余信息，

表 10 推理效率对比

Table 10 Comparison of inference efficiency

Model	Params/M	FLOPs/G	Latency/ms (*BS*=1)	Throughout/(f/s) (*BS*=128)	MAC/G
ResNet-34	21.3	1.2	2.3	435.7	4.8
Wide ResNet 28-10	37.2	5.9	8.1	158.2	19.6
Swin-T	86.7	4.3	12.5	95.3	15.2
EfficientNet	52.9	2.5	4.7	265.9	8.9
DCPRNet	29.4	2.1	2.7	370.4	6.5

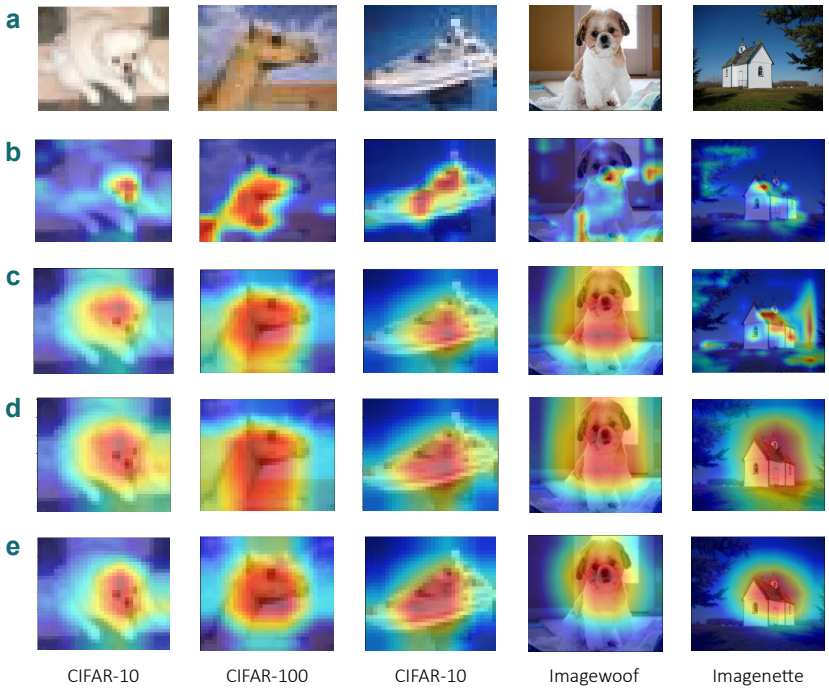


图 13 不同网络的可视化图像。(a) Original image；(b) MobileNet；(c) ShuffleNet；(d) ResNet；(e) DCPRNet

Fig. 13 Visualized images of different networks. (a) Original image; (b) MobileNet; (c) ShuffleNet; (d) ResNet; (e) DCPRNet

提高关键区域的贡献度，进而提高网络的分类准确度。

综上所述，图 13 的热力图对比从特征响应层面证实，DCPRNet 中的随机子组置换动态重组，打破了固定交互模式，为模型学习了更鲁棒的通道依赖关系，奠定了特征响应连贯且完整的基础。而分块加权注意力 (CWA) 则在此基础上实现了动态、细粒度的权重校准，使模型自适应地聚焦最具判别性的局部特征，从而实现了对关键区域的精准锁定。

3.3 消融实验

为验证 DCPRNet 中各核心创新模块的有效性，本实验以 ResNet-34 为基线网络，通过逐步移除关键模块构建 4 组对比模型，第一组为替换首层中 7×7 卷积为 3×3 卷积，删除最大池化的 ResNet-34。第二组为在 ResNet-34 上加入 CWA 模块，记为 Net1。第三组为在 ResNet-34 上融合随机子组置换模块记为 Net2，第四组加入池化残差模块为 DCPRNet 整体模型。将上述 4 种网络分别在 CIFAR-100、CIFAR10、SVHN、Imagewoof 和 Imagenette 数据集上进行性能测试，实验结果如表 11 和图 14 所示。实验保持数据预处理、优化器及超参数设置同 3.1 小节。由实验结果可知，相较于 ResNet-34 模型，各创新模块的加入均能显著提升网络性能，且完整模型 DCPRNet 在所有数据集上均取得最优结果。

输入层优化是提升网络性能的基础，能够有效保留细粒度特征，为后续模块提供优质初始输入。DCFC 模块中的随机子组置换与分块加权注意力协同作用，可打破通道隔离，实现跨组信息交互与动态特征校准，显著提升模型对复杂数据的适配性。动态通道特征校准的池化残差模块通过无参下采样与残差增强，在减少信息丢失

表 11 各模块在不同数据集上的消融实验结果 (分类准确率/%)

Table 11 Ablation experiments among different modules on different datasets (classification accuracy rates/%)

Group	CIFAR-100	CIFAR-10	SVHN	Imagewoof	Imagenette
ResNet-34	78.21	95.29	96.62	79.36	87.20
Net1	79.33	95.68	96.76	79.69	88.79
Net2	80.11	95.75	96.83	80.08	90.40
DCPRNet	**80.36**	**96.41**	**96.97**	**80.61**	**91.59**

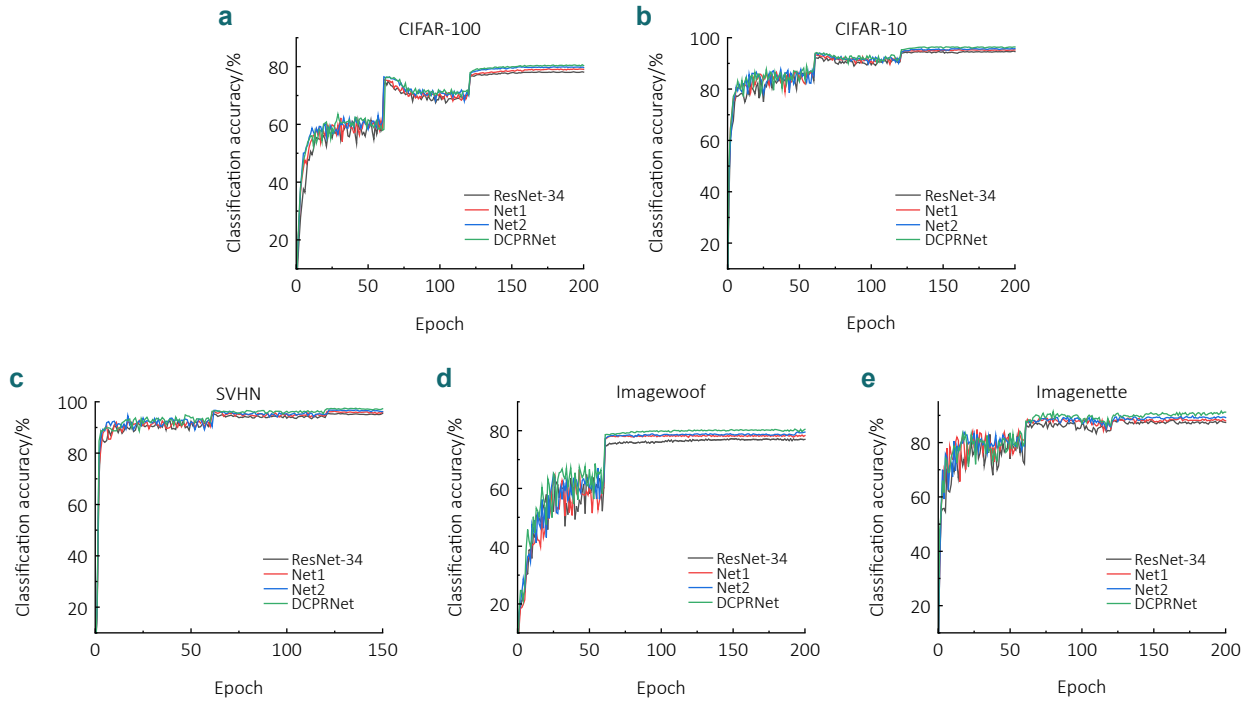


图 14 各模块之间在不同数据集上准确率图。(a) CIFAR-100；(b) CIFAR-10；(c) SVHN；(d) Imagewoof；(e) Imagenette

Fig. 14 Accuracy rates of each module on different datasets. (a) CIFAR-100; (b) CIFAR-10; (c) SVHN; (d) Imagewoof; (e) Imagenette

的同时降低计算量，兼顾性能提升与推理效率。DCPRNet 通过各模块的协同集成，在分类准确率、收敛速度与计算效率上均表现最优，验证了本文提出的动态通道特征校准与池化残差设计的有效性，为图像分类任务提供了高效可行的网络架构。

4 结 论

本文围绕传统 ResNet 网络在图像分类任务中存在的通道特征交互能力不足导致精准地筛选关键特征差异的问题，设计并提出动态通道特征校准的池化残差分类网 (DCPRNet)。该网络以 ResNet-34 为基线架构，在输入层，将原 7×7 卷积替换为 3×3 卷积并删除最大池化层，在降低参数量的同时保留图像更多信息。在特征处理层，设计动态通道特征校准 (DCFC) 模块，结合随机子组置换与分块加权注意力 (CWA)，解决传统分组卷积的通道交互不足问题，实现跨组信息交互与特征权重校准，有效抑制异常值干扰。在残差结构层，加入池化残差模块池化，以平均池化替代步长卷积完成下采样，在减少计算量的同时避免边缘特征丢失，增强模型泛化能力。

多数据集实验结果显示，DCPRNet 在 CIFAR-10、CIFAR-100、和 SVHN 等不同类型数据集上均展现出优异分类性能。消融实验进一步验证，输入层优化、DCFC 模块与池化残差模块的协同作用是性能提升的关键，且 DCPRNet-34 在参数量与分类精度间实现最优平衡。DCPRNet 单张推理延迟低、批量吞吐量高，内存开销合理，兼顾性能与部署效率，且能精准聚焦目标细粒度特征、规避背景干扰，特征校准能力优于传统网络。DCPRNet 通过结构与机制创新，解决传统残差网络缺陷，低开销下提升分类性能，适用于通用图像分类及医疗影像、安防监控等场景。未来可探索大规模高分辨率数据集泛化性，拓展垂直领域应用，优化模块实现更优轻量化部署。

作者贡献

赵文雯：研究资料的收集、实验、分析，论文写作；姜文涛：指导论文选题、论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Yang H, Li J. Label contrastive learning for image classification[J]. *Soft Comput*, 2023, **27** (18): 13477−13486.
2. Jiang W T, Dong R, Zhang S C. Global pooling residual classification network guided by local attention[J]. *Opto-Electron Eng*,

2024, **51** (7): 240126.
姜文涛, 董睿, 张晟翀. 局部注意力引导下的全局池化残差分类网络[J]. 光电工程, 2024, **51** (7): 240126.
3. Jiang W T, Yuan H, Liu W J. Neuron signal attenuation activation mechanism for deep learning[J]. *Patterns*, 2025, **6** (1): 101117.
4. Zheng L, Yang Y, Tian Q. SIFT meets CNN: a decade survey of instance retrieval[J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, **40** (5): 1224−1244.
5. Albiol A, Monzo D, Martin A, et al. Face recognition using HOG–EBGM[J]. *Pattern Recogn Lett*, 2008, **29** (10): 1537−1543.
6. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classifica- tion with deep convolutional neural networks[J]. *Commun. ACM*, 2017, **60**(6): 84−90.
7. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[Z]. arXiv: 1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.1556.
8. Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015: 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
9. He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.
10. Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5987–5995. https://doi.org/10.1109/CVPR.2017.634.
11. Zagoruyko S, Komodakis N. Wide residual networks[C]//*Proceedings of the British Machine Vision Conference*, York, 2016: 87. https://doi.org/10.5244/C.30.87.
12. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017: 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
13. Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 6848–6856. https://doi.org/10.1109/CVPR.2018.00716.
14. Lin G J, Gong T, Lin T. A survey of attention mechanisms in image classification tasks[J]. *Telecommun Eng*, 2025, **65** (8): 1349−1362.
林国军, 龚涛, 林通. 图像分类任务中的注意力机制研究综述[J]. 电讯技术, 2025, **65** (8): 1349−1362.
15. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
16. Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*, Munich, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
17. Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 7794–7803. https://doi.org/10.1109/CVPR.2018.00813.
18. Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.
19. Khan S, Naseer M, Hayat M, et al. Transformers in vision: a survey[J]. *ACM Comput Surv*, 2022, **54** (10s): 200.
20. Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 2021: 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986.
21. Wang A, Peng Z, Wang J, et al. Low-dose nano-CT reconstruction image denoising based on Swin Transformer and convolutional neural network[J]. *Opto-Electron Eng*, 2025, **52** (6): 240306.
汪澳, 彭真, 汪俊, 等. 基于 Swin Transformer 和卷积神经网络的低剂量纳米 CT 重建图像去噪[J]. 光电工程, 2025, **52** (6): 240306.
22. Zhu L H, Liao B C, Zhang Q, et al. Vision mamba: efficient visual representation learning with bidirectional state space model[C]//*Proceedings of the 41st International Conference on Machine Learning*, Vienna, 2024: 2584.
23. Yuan H, Liu J, Jiang W T, et al. Double-pooling residual classification network based on feature reordering attention mechanism[J]. *J Image Graph*, 2025, **30** (1): 110−129.
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.
24. Gao Y, Han X T, Wang X, et al. Channel interaction networks for fine-grained image categorization[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, 2020: 10818–10825. https://doi.org/10.1609/aaai.v34i07.6712.
25. Bai S W, Wang M Y, Hu J, et al. Multi-region attention network for fine-grained image classification[J]. *Comput Eng*, 2024, **50** (1): 271−278.
白尚旺, 王梦瑶, 胡静, 等. 多区域注意力的细粒度图像分类网络[J]. 计算机工程, 2024, **50** (1): 271−278.
26. Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.
27. Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 2020: 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165.
28. Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.
29. Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//*Proceedings of the 36th International Conference on Machine Learning*, Long Beach, 2019: 6105–6114.
30. Lan H, Wang X H, Shen H, et al. Couplformer: rethinking vision transformer with coupling attention[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 2023: 6464–6473. https://doi.org/10.1109/WACV56688.2023.00641.
31. Choromanski K M, Likhosherstov V, Dohan D, et al. Rethinking attention with performers[C]//*International Conference on Learning Representations*, Vienna, 2021.
32. Konstantinidis D, Papastratis I, Dimitropoulos K, et al. Multi-manifold attention for vision transformers[J]. *IEEE Access*, 2023, **11**: 123433−123444.

33. Wu X D, Gao S Q, Zhang Z Y, et al. Auto- Train-Once: controller network guided automatic network pruning from scratch[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2024: 16163–16173. https://doi.org/10.1109/CVPR52733.2024.01530.

作者简介



姜文涛 (1986-)，男，博士，副教授，主要研究方向为模式识别与人工智能、图像与视觉信息计算。

E-mail：jiangwentao@lntu.edu.com



【通信作者】赵文雯 (1999-)，女，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、人工智能。

E-mail：1435998539@qq.com

Pooling residual image classification network with dynamic channel feature calibration

Jiang Wentao, Zhao Wenwen*

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

Abstract

Objective Image classification, a foundational core task in computer vision, supports diverse critical applications including clinical medical diagnosis, intelligent security surveillance, remote sensing image interpretation, and autonomous driving scene recognition. Deep learning models such as ResNet, DenseNet, and ViT have notably advanced classification performance in recent years, but traditional networks still bear inherent flaws that restrict their practical applicability in complex real-world scenarios. ResNet, a widely adopted architecture for image classification tasks due to its gradient preservation advantage, lacks sufficient channel feature interaction, which directly impedes the effective selection of key discriminative features. Its initial 7×7 convolution and max-pooling layers trigger excessive downsampling at the early feature extraction stage, resulting in the loss of fine-grained spatial details that are vital for accurate classification of complex images with overlapping objects or subtle differences. Fixed channel grouping in traditional convolutions further isolates cross-channel feature information, limiting the model's ability to integrate multi-dimensional feature cues and capture inter-channel correlations. Existing attention mechanisms like SE and CBAM only focus on single dimensions—either channel or spatial—failing to achieve dynamic and comprehensive feature calibration varying image contents. This study addresses these critical limitations, enhances cross-channel information interaction, strengthens feature representation and key feature extraction capabilities, and ultimately improves the overall classification performance and generalization ability of the model.

Methods To achieve this objective, a pooling residual image classification network with dynamic channel feature calibration (DCPRNet) was proposed, which is optimized and improved based on the ResNet-34 architecture to address its inherent deficiencies. Three core targeted improvements were integrated into the network to overcome the drawbacks of traditional models while maintaining computational feasibility. First, the initial layer was optimized: the 7×7 convolution was replaced with a 3×3 convolution to reduce parameter complexity by nearly half while retaining robust feature extraction capability, and the max-pooling layer was removed to fully preserve fine-grained spatial information for subsequent feature processing, fusion, and calibration. Second, a dynamic channel feature calibration (DCFC) module was designed, which combines a random subgroup shuffle module and chunked weighted attention (CWA) mechanism. The random subgroup shuffle module breaks the isolation between channel groups via dual-layer grouping and random shuffling, effectively promoting cross-group information interaction and dynamic channel communication without introducing excessive computation. The CWA mechanism calibrates the weights of feature chunks through feature chunking, local importance score generation, mean pooling-based global information aggregation, and sigmoid normalization, thereby suppressing redundant background information and enhancing the contribution of critical feature regions related to target classification. Third, an average pooling-based residual module was added to the DCFC branch, using 1×1 convolution and 2×2 average pooling instead of traditional stride convolution to realize low-cost smooth

Foundation item: National Natural Science Foundation of China (61172144), Natural Science Foundation of Liaoning Province (20170540426), and the Key Fund Project of Liaoning Provincial Department of Education (LJ2017QL034)

*Correspondence: Zhao W W, E-mail: 1435998539@qq.com

downsampling, reduce feature distortion, and better preservation of effective feature information.

Results and Discussions Comprehensive experiments were conducted on five representative datasets (CIFAR-10, CIFAR-100, SVHN, Imagenette, Imagewoof) that cover simple object, complex category, low-resolution, and real-world natural image scenarios, ensuring the model's performance is validated under diverse conditions. DCPRNet achieved classification accuracies of 96.41%, 80.36%, 96.97%, 91.59%, and 80.61% on these datasets respectively, showing stable and reliable performance across different data types, complexity levels, and feature distributions. Ablation studies further confirmed that input layer optimization, the DCFC module, and the average pooling residual module each boost model performance significantly, and the synergistic effects among the three components yield the optimal overall result by complementing each other's advantages. Comparative experiments against 12 mainstream models, including ResNet-34, DenseNet-121, ViT-B/16, and MobileNet-V2, demonstrated DCPRNet's consistent superiority: it achieved a 2.15% accuracy gain over the baseline ResNet-34 on the CIFAR-100 dataset and a 4.39% improvement on the ImageNet dataset, while maintaining comparable computational cost. Heatmap-based visual analysis further verified that DCPRNet can accurately focus on critical fine-grained features of target objects, effectively avoiding interference from irrelevant background regions and exhibiting excellent feature calibration capability. These results confirm that the proposed improvements effectively enhance cross-channel information interaction, resolve the inherent limitations of traditional networks, and strengthen the model's ability to extract discriminative features.

Conclusions DCPRNet was successfully developed by optimizing the ResNet-34 architecture and introducing the DCFC module and average pooling residual module. The model effectively enhances cross-channel information interaction, significantly strengthens feature representation ability and key feature extraction efficiency, and achieves a balanced trade-off between classification accuracy, computational cost, and model robustness. Experimental results and visual analysis fully validate its superiority over mainstream models on multiple datasets, confirming its good robustness and adaptability to different image scenarios and data distributions. By addressing the core flaws of traditional image classification networks—such as insufficient channel interaction, excessive feature loss during downsampling, and limited feature calibration capability—DCPRNet provides a reliable technical solution for image classification tasks in multiple fields including medical imaging, intelligent surveillance, and remote sensing detection. It also offers valuable theoretical and practical reference for the further optimization of deep learning-based image classification models, laying a solid foundation for the development of more efficient, accurate, and lightweight computer vision algorithms suitable for edge computing devices.

Keywords: image classification; attention mechanism; feature calibration; residual network

Jiang W T, Zhao W W. Pooling residual image classification network with dynamic channel feature calibration[J]. *Opto-Electron Eng*, 2026, **53**(5): 250309; DOI: 10.12086/oee.2026.250309

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文