空间约束注意力机制的图像分类网络

姜文涛, 王鑫杰, 张晟翀

引用本文:
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.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202505025

DOI: 10.11992/tis.202505025
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20251029.1756.002

空间约束注意力机制的图像分类网络

姜文涛[1], 王鑫杰[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: 针对分类网络中低阶特征提取不充分和特征图空间位置加权不足的问题, 本文提出了一种空间约束注意力机制的图像分类网络(spatially constrained attention mechanism for image classification network, SCAM-Net)。SCAM-Net 网络以 WideResnet-28-10 残差网络为基础架构。本文提出空间约束注意力机制(spatial constrained attention mechanism, SCA), 通过引入空间约束机制和动态加权策略, 显著增强了特征图的空间位置感知能力, 使网络能够更精准地聚焦于关键区域, 从而优化特征表示质量, 提升模型在复杂场景下的判别能力。提出了边缘感知卷积(edge aware convolution, EAConv), 通过融合 Sobel 算子和不同尺寸的卷积核, 实现了对跨层次信息的整合, 解决了原模型中首层卷积对边缘特征提取能力不足的问题。实验结果表明, 在 CIFAR-100、CIFAR-10、SVHN 和 GTSRB 4 种数据集上, SCAM-Net 相较于基线模型 WideResnet-28-10 在分类准确率上分别提升了 2.43%、0.93%、0.14% 和 0.91%; 同时, 相比于性能排名第 2 的 QKFormer 网络在 4 种数据集上的分类准确率分别提高了 0.13%、0.10%、0.12% 和 0.34%。空间约束注意力机制和边缘感知卷积相互协作, 使得 SCAM-Net 在处理图像时能够更准确地关注图像中的复杂细节, 有效提升图像分类精度。

关键词: 图像分类; 空间约束注意力机制; 边缘感知卷积; 随机池化; 空间信息; 边缘特征; 特征融合; 残差网络
中图分类号: TP391　　**文献标志码**: A　　**文章编号**: 1673−4785(2025)06−1444−17

中文引用格式: 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.
英文引用格式: 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.

Spatially constrained attention mechanism for image classification network

JIANG Wentao[1], WANG Xinjie[1], ZHANG Shengchong[2]
(1. School of Software, Liaoning Technical University, Huludao 125105, China; 2. Key Laboratory of Optoelectronic Information Control and Security Technology, Tianjin 300308, China)

Abstract: This paper addresses two major issues in image classification networks: insufficient low-level feature extraction and inadequate spatial weighting of feature maps. A novel image classification network named SCAM-Net (spatially constrained attention Mechanism for Image Classification Network) is proposed. SCAM-Net is built upon the WideResNet-28-10 architecture. First, a Spatial-Constrained Attention (SCA) mechanism is introduced. By incorporating a spatial constraint strategy and a dynamic weighting approach, SCA significantly enhances the network’s ability to perceive spatial positions in feature maps. This enables the model to focus more precisely on critical regions and improves the quality of feature representation, leading to better discrimination in complex scenarios. Second, an Edge-Aware Convolution (EAConv) is developed. EAConv integrates Sobel operators with convolutions of multiple kernel sizes to capture multi-level edge information, thereby compensating for the weak edge feature extraction capability in the original first convolutional layer. Experimental results demonstrate that SCAM-Net outperforms the baseline WideResNet-28-10 by 2.43%, 0.93%, 0.14%, and 0.91% on CIFAR-100, CIFAR-10, SVHN, and GTSRB datasets, respectively. Compared with the second-best model QKFormer, SCAM-Net achieves 0.13%, 0.10%, 0.12%, and 0.34% higher classification accuracy on the same datasets. These results confirm that the collaboration between the spatial-constrained attention mechanism and the edge-aware convolution allows SCAM-Net to better capture fine-grained visual details and effectively improve image classification performance.

Keywords: image classification; spatially constrained attention mechanism; edge aware convolution; stochastic pooling; spatial information; edge features; feature fusion; residual network

图像分类是计算机视觉领域中的核心任务, 旨在根据图像内容将其划分到预定义的类别中。通常, 分类网络对输入图像进行特征提取和分类, 以识别其中的物体、场景或模式。

收稿日期: 2025−05−27.　**网络出版日期**: 2025−10−30.
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: lntuwulue@163.com.

以 ResNet[1] 为代表的深度残差网络在图像分类任务中取得了巨大成功, 奠定了现代卷积神经网络架构设计的基础。近年来, 国内外在图像分类网络领域取得了显著进展。文献 [2-4] 提出的网络通过减少模型参数量和计算复杂度, 降低内存占用和计算需求, 从而提升模型在实际场景中的运行速度和资源利用效率, 使得神经网络更适用于边缘设备与移动端部署。文献 [5-8] 提出的网络通过引入多模态特征融合与对齐机制, 提升了模型在异构数据间语义一致性建模与协同感知方面的能力, 有效缓解了模态间信息不对齐或冗余干扰的问题, 显著增强了模型对复杂输入的判别能力。文献 [9-11] 提出的网络通过增强模型对局部与全局特征的协同建模能力, 有效提升了网络对复杂视觉场景中多尺度语义信息的理解与表达能力, 改善了模型在分类任务下的泛化性能。

随着计算机视觉的应用不断深入, 模型处理复杂图像任务时需要更强的特征聚焦和提取能力。因此, 人们提出了注意力机制, 使模型能够自适应地关注输入数据中的关键特征, 提升其在图像分类、目标检测等任务中的性能, 尤其是在面对高维数据和多模态信息时, 注意力机制展现出了显著的优势。近年来, 注意力机制的发展较为迅速, 文献 [12-16] 提出的注意力机制主要针对图像退化场景, 通过增强模型对关键信息区域的响应能力, 有效提升了低质量图像重建中的鲁棒性和上下文恢复效果, 展现出在信息受限条件下保持全局一致性与细节还原的优势。文献 [17] 所提出的注意力机制则引入概率建模思想, 进一步提升了模型对关键区域的自适应感知能力。该方法通过为各特征分配概率权重, 灵活调节响应强度, 有效优化了模型在多样化任务中的精度与稳定性, 尤其在存在较高不确定性的环境下表现出较强的适应能力与效率。文献 [18-20] 所提出的注意力机制则更侧重于增强特征间的信息交互与语义表达能力, 结合结构设计优化, 提升了模型在复杂场景中的适应性, 以及对关键区域的感知与建模能力。

尽管上述网络和注意力机制在模型轻量化、多尺度建模及关键区域感知等方面取得了诸多成果, 极大地推动了图像分类技术的发展, 但在应对更高挑战性任务时仍存在一定局限。一方面, 现有网络在细粒度边界建模与小目标识别能力方面表现不足, 尤其在结构复杂或语义边界模糊的图像中, 容易忽略关键特征; 另一方面, 主流注意力机制在空间信息保留与跨尺度特征聚合方面仍存在优化空间, 难以充分挖掘图像中细节信息与空间结构特征。

为了解决上述这些问题, 本文在宽残差网络 WideResnet-28-10[21] 的基础上提出了基于空间约束注意力机制的图像分类网络 (spatially constrained attention mechanism for image classification network, SCAM-Net)。本文创新点为: 1) 提出了 EAConv 卷积。通过结合 Sobel 边缘检测算子和卷积操作, 能够有效提取图像中的边缘、局部和深度局部特征, 并通过特征融合操作整合不同尺度的信息, 从而增强网络对细节和复杂结构的识别能力。2) 本文提出了 SCA (spatial constrained attention mechanism) 注意力机制。通过在 SCA 注意力机制内部引入通道注意力机制来增强关键特征, 确保网络能够自动聚焦于重要的局部细节和全局信息。利用动态池化选择策略, 使网络能够在不同的尺度上捕获多样化的特征信息。同时, SCA 注意力机制利用位置编码保留了图像的空间结构, 有效地捕捉长程依赖, 使得网络能够更好地理解空间关系并提高图像分类的准确性。SCA 注意力机制和 EAConv 卷积的结合, 使得网络能够在处理图像时更加精确地抓住细节和边缘信息, 提升了分类性能, 尤其在复杂场景下表现得尤为突出。

1　相关工作

1.1　注意力机制

在图像分类任务中, 注意力机制已经发展出多种方向。通道注意力通过在不同特征通道之间分配权重, 能够突出判别性特征, 从而提升特征选择能力, 但往往对空间结构的保持不足。空间注意力则侧重于在特征图上生成权重分布, 以强化显著区域的表达, 虽然提升了对关键位置的感知, 但在建模长程依赖和保持全局结构方面仍有局限。自注意力方法能够捕获远距离的依赖关系, 适合处理细粒度特征, 但在计算开销和轻量化应用上存在一定挑战。多尺度或混合注意力机制则通过融合通道与空间信息、引入尺度变化来增强特征表达, 但通常结构较为复杂, 对边界模糊和小目标的处理效果也有限。基于以上不足, 本文提出了一种空间约束注意力机制, 在位置编码的帮助下为特征提供稳定的空间参考, 并结合特征打乱操作来减弱模型对局部位置信息的过度依赖, 同时利用随机池化增强多尺度特征建模能力。通过这种设计, 模型在保持空间结构约束的同时, 能够更好地感知关键区域与细节特征, 从而带来整体分类性能的提升。

1.2　卷积运算

在图像分类任务中，卷积运算作为最基础的特征提取方式，随着研究的深入也不断演化出不同的分支。小卷积核的堆叠能够在控制参数量的同时实现深层次特征建模，但受限于感受野而难以捕获全局信息；大卷积核设计则在增强全局依赖的同时带来了额外的计算和优化开销；可变形或动态卷积能够适应形变与复杂结构，但在轻量化与稳定性方面仍存在不足；分组卷积与深度可分离卷积则在提升效率的同时削弱了通道间交互，限制了特征表达的充分性。总体来看，现有卷积运算虽在不同方向上取得了进展，但在对细节与边界区域的敏感性方面仍显不足。

针对这一问题，本文提出了一种边缘感知卷积。该方法将传统卷积与 Sobel 算子相结合，显式提取边缘特征，并进一步融合多尺度卷积结构，以同时获取局部与深层局部特征。通过这种设计，边缘感知卷积在保持常规卷积计算效率的同时，显著增强了模型对细节和边界的感知能力，并为后续注意力机制提供了更丰富的特征输入，从而有效提升了整体分类性能。

2　空间约束注意力机制的图像分类网络

2.1　网络结构

在图像分类任务中，本文提出的图像分类网络模型整体结构如图 1 所示，主要由边缘感知卷积（edge aware convolution, EAConv）、3 个残差块（Layer1、Layer2、Layer3）以及空间约束注意力机制（SCA）构成。首先，输入图像经过 EAConv 提取低阶边缘特征；随后，三层残差块逐步提取并融合多层次语义特征。在此基础上，在第 3 个残差块后引入 SCA 模块，在其内部通过位置编码、特征打乱与随机池化等操作实现空间结构约束与多尺度特征增强，从而突出关键信息区域并抑制冗余特征。上述各组件相互衔接，形成自底向上的特征提取与增强流程，为最终分类提供有效支持。

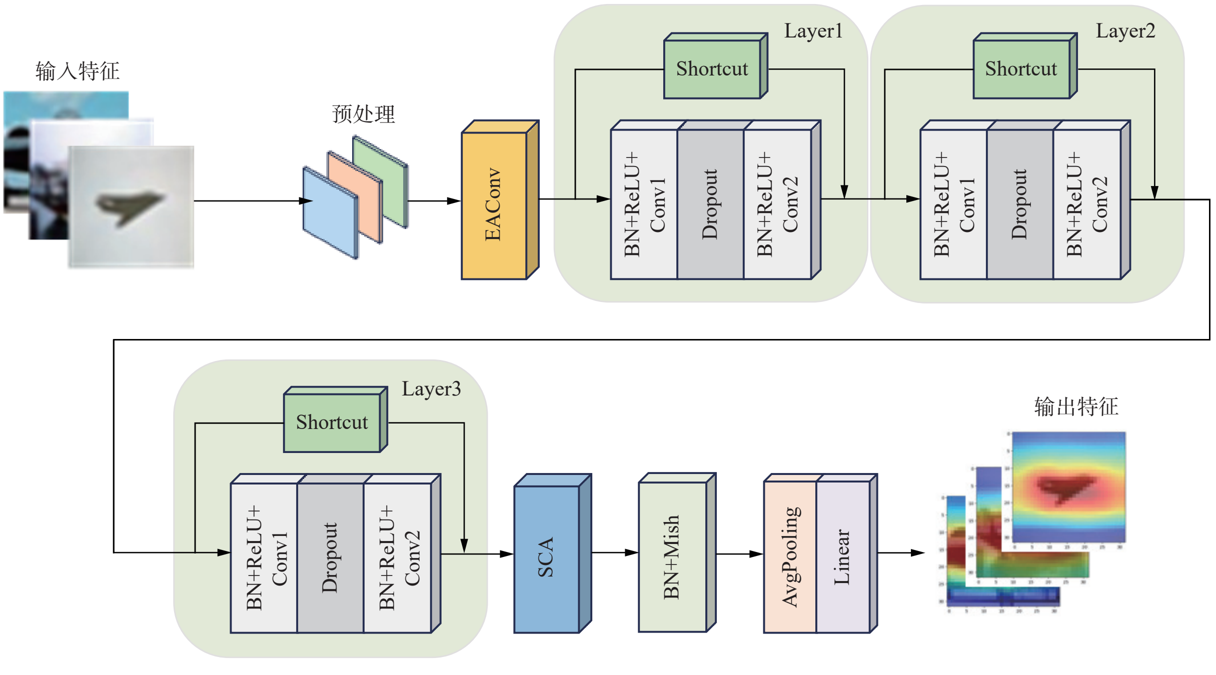


图 1　SCAM-Net 整体结构

Fig. 1　Overall architecture of SCAM-Net

2.2　空间约束注意力机制模块

WideResnet-28-10 网络虽然可以通过宽层深度有效提升特征表达能力，但对于空间位置信息提取不足，无法更好地捕捉空间中的相对位置关系，本文提出了空间约束注意力机制即 SCA 注意力机制，SCA 注意力机制的结构如图 2 所示。本节仅探讨 SCA 的设计理论，SCA 注意力机制在整体网络模型中的相对位置是通过第 3.2 节的 SCA 注意力机制插入位置实验具体得出的。

输入特征图 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$，其中 C 为输入通道数，$H\times W$ 为特征图空间大小，H 为特征图高度，W 为特征图宽度。

图 2 中符号 $\otimes$ 为特征之间的乘法操作，符号 $\oplus$ 为特征之间的相加操作。

SCA 注意力机制由非线性通道增强模块、二维位置编码模块、AttnConv 模块、特征打乱模块和随机池化模块几部分组成。

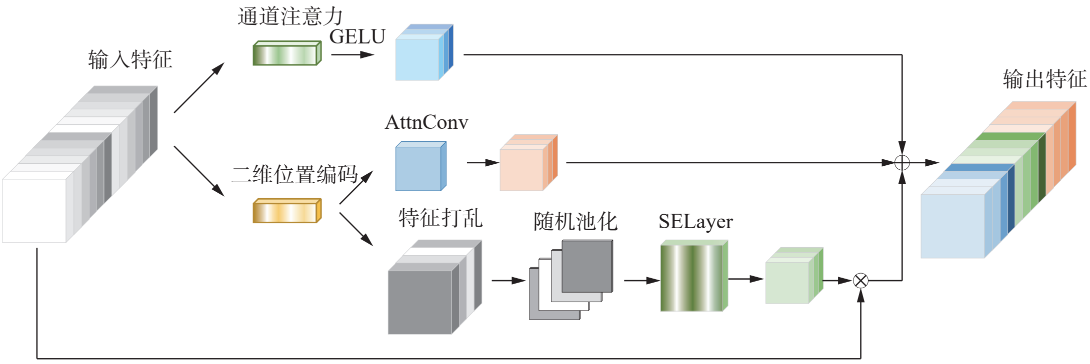


图 2　SCA 注意力机制模块

Fig. 2　SCA attention mechanism module

2.2.1　非线性通道增强模块

SCA 主要由 3 个分支组成, 其中在这一分支主要是由通道注意力机制 (channel attention mechanism)[22] 和 GELU(Gaussian error linear unit) 激活函数组成。输入特征 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$, 首先经过通道注意力机制得到特征 $\boldsymbol{F}_1\in\mathbf{R}^{C\times1\times1}$, 在此处引入通道注意力是利用其可以显著提高输入特征的表达能力。通道注意力的核心思想是根据不同通道的重要性对输入特征进行加权, 帮助模型自动学习哪些通道对当前任务更为关键, 从而强化有用信息的传递, 抑制冗余或不相关的特征。这种加权操作能够使得网络更加聚焦于具有判别力的特征通道, 提升网络在复杂任务中的表现。

随后对通道注意力的输出特征 $\boldsymbol{F}_1\in\mathbf{R}^{C\times1\times1}$ 应用GELU激活函数得到特征 $\boldsymbol{F}_2\in\mathbf{R}^{C\times1\times1}$, 可表示为

$$\boldsymbol{F}_2=\mathrm{GELU}(\boldsymbol{F}_1)$$

式中: GELU 激活函数公式为

$$\mathrm{GELU}(x)=\frac{1}{2}x\left(1+\tanh\left(\sqrt{\frac{2}{\pi}}\,(x+0.044\,715x^3)\right)\right)\tag{1}$$

式中: 三次项 $0.044\,715x^3$ 的引入使得在输入为负值时, GELU 激活函数的输出不再完全归零, 而是保留一个较小的负值, 从而在负半轴区域维持一定的响应, 避免了“神经元失活”现象的出现; tanh 函数是双曲正切函数, 将输入有效压缩至 $[-1,1]$ 区间, 实现平滑的非线性变换。使用 GELU 激活函数可以在复杂的深度网络中能够更好地捕捉非线性特征, 提升模型性能。

通道注意力与 GELU 激活函数的结合旨在提升模型的特征表达与训练稳定性。通道注意力通过全局特征聚合、自适应筛选和加权机制, 使模型能够自动关注关键通道, 增强有效特征提取并抑制冗余信息。GELU 作为平滑的非线性激活函数, 避免了 ReLU (rectified linear unit) 在 0 处的不连续问题, 使梯度传递更稳定, 减轻梯度消失或爆炸风险。同时, GELU 的概率性特征选择机制既能突出重要特征, 又能平滑弱特征影响, 从而加快收敛并提升泛化性能。

在深度学习中, 每个通道可视为不同尺度或特征空间的信号响应。通道注意力通过全局统计自适应调整权重, 突出关键通道并抑制冗余; GELU 则以平滑的非线性变换保留细微特征, 避免直接置零带来的信息损失。先施加通道注意力再接入 GELU, 可确保加权基于完整特征, 优化信息流动并提升非线性变换效率; 若顺序倒置, 则特征分布被改变, 影响注意力的准确性。该设计在视觉任务中不仅增强了特征提取能力, 还提高了训练的效率与稳定性, 从而优化整体性能。

2.2.2　二维位置编码模块

在 SCA 的第 2 个分支中, 将输入特征 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$ 传递给二维位置编码模块[23]。二维位置编码模块动态获取输入特征中的高度 H 和宽度 W。针对高度 H 和宽度 W 生成张量 $\boldsymbol{p}$, 将输入特征 $\boldsymbol{F}$ 和张量 $\boldsymbol{p}$ 进行相加, 得到了一个包含位置信息的特征图 $\boldsymbol{F}_3\in\mathbf{R}^{C\times H\times W}$, 可表示为

$$\boldsymbol{F}_3=\boldsymbol{F}+\boldsymbol{p}$$

二维位置编码模块应用于初始输入特征, 为模型提供空间位置信息, 帮助其学习图像中各位置的相对关系。卷积操作擅长提取局部特征, 但无法直接感知全局位置信息, 而位置编码通过为每个像素附加唯一位置信息, 弥补了这一不足, 尤其在像素级任务中表现突出。将位置编码置于输入特征后, 可确保后续计算时融合空间和内容信息, 从而提升模型对空间结构和特征关系的捕捉能力, 增强特征提取与注意力计算的精度。

位置编码模块通过在输入特征中注入空间位置信息, 解决了注意力机制对输入顺序不敏感的问题, 使模型能够理解图像不同区域的空间关系。其基于正余弦函数对高度和宽度进行编码, 既保留绝对位置感知, 又隐含相对位置信息, 从

而建立关键的空间关系建模能力。相比卷积在深层网络中因池化等操作导致位置信息丢失，位置编码能有效弥补这一不足，增强模型的空间感知能力。

在此基础上，位置编码为 AttnConv 提供了全局空间上下文，使其在关注特征内容的同时兼顾空间位置，提高不同位置特征的加权效果。即便特征在后续处理中被打乱，位置编码仍能帮助模型保持对相对关系的理解，从而提升整体性能与鲁棒性。

2.2.3　AttnConv 模块

使用 AttnConv 模块对位置编码的输出特征进行更高层次的处理，可以使网络更好地感知特征在空间中的相对位置关系。AttnConv 提出将自注意力机制 (self-attention)[24] 和卷积相结合，AttnConv 模块图如图 3 所示。

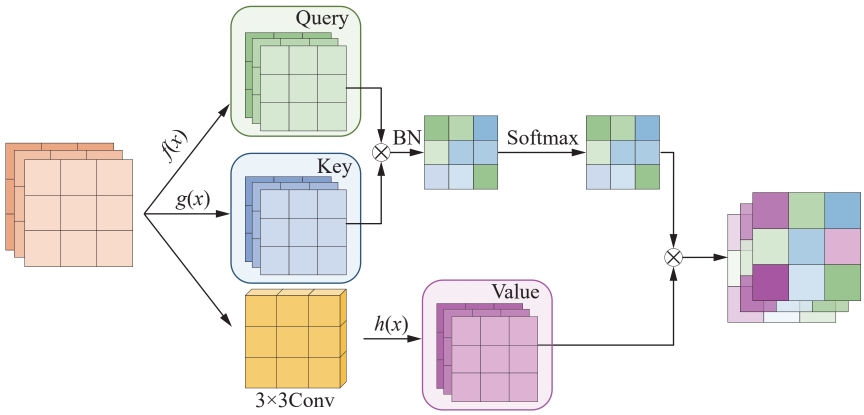


图 3　AttnConv 模型

Fig. 3　AttnConv model

通过输入特征 $\boldsymbol{F}_3 \in \mathbf{R}^{C\times H\times W}$，可得到张量 $\boldsymbol{Q}$(query) 和张量 $\boldsymbol{K}$(key)，对这 2 个张量通过 bmm (batch matrix multiplication) 函数进行相乘操作，将输出的结果进行归一化和标准化操作得到注意力矩阵 $\boldsymbol{e}$。将矩阵 $\boldsymbol{e}$ 经过 Softmax 激活函数，将注意力权重标准化为一个概率分布，得到矩阵 $\boldsymbol{T}$，可表示为

$$\boldsymbol{T} = \mathrm{Softmax}(\boldsymbol{e})$$

其中 Softmax 的公式为

$$\mathrm{Softmax}(x_i) = \frac{\mathrm{e}^{x_i}}{\sum_{j=1}^{n} \mathrm{e}^{x_j}},\quad i = 1,2,\cdots,n \tag{2}$$

式中：e^{x_i} 表示对输入的第 i 个元素进行指数变换，以确保其取值为正，分母则为对所有输入元素经指数变换后的结果进行求和。整体公式计算得到的是第 i 个输入在所有输入中的归一化比重，其输出构成一个概率分布，且所有元素之和恒为 1。Softmax 激活函数通过指数运算在数值上放大较大输入并压缩较小输入，从而增强了特征间的相对差异性，使注意力机制能够更加集中于显著区域，实现对关键信息的有效建模。

将特征 $\boldsymbol{F}_3 \in \mathbf{R}^{C\times H\times W}$ 传入尺寸为 3×3 的卷积得到特征 $\boldsymbol{F}_4 \in \mathbf{R}^{C\times H\times W}$，将特征 $\boldsymbol{F}_4$ 展平得到特征序列 $\boldsymbol{V}$(value)，将特征序列 $\boldsymbol{V}$ 与调整形状后的矩阵 $\boldsymbol{T}$ 传入 bmm 函数进行矩阵乘法操作，得到 AttnConv 模块输出特征 $\boldsymbol{F}_5 \in \mathbf{R}^{C\times H\times W}$。

在 AttnConv 模块中，卷积特征映射经自注意力权重自适应重校正，实现局部结构保真与全局特征动态建模的平衡。其置于位置编码层之后，使注意力建立在明确的空间坐标系上，增强位置信息表达，避免低级特征中的位置模糊，确保空间关系的有效建模。

AttnConv 模块结合了卷积的局部特征提取能力和自注意力机制的全局信息建模能力，从而增强特征表示能力。卷积有效提取局部特征，但受限于感受野；自注意力机制通过位置间相似度计算捕捉全局上下文。结合两者后，AttnConv 模块在提取局部信息的同时利用全局上下文，提升了模型的表现能力。

2.2.4　特征打乱模块

首先，将二维位置编码的输出特征进行特征打乱操作。特征打乱操作如图 4 所示。

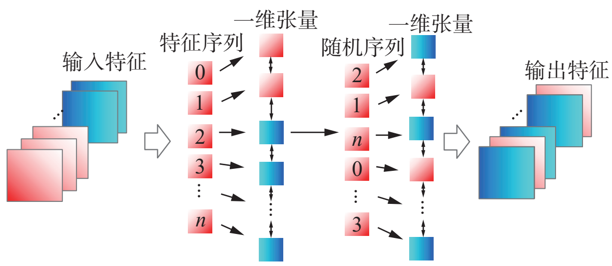


图 4　特征打乱模块

Fig. 4　Feature shuffling module

先将输入特征$\boldsymbol{F}_3$展开生成的张量$\boldsymbol{F}' \in \mathbf{R}^{C\times H\times W}$，并根据张量$\boldsymbol{F}'$生成相应的从 0 开始的索引序列。随后将索引序列打乱生成随机索引序列 $\boldsymbol{I}$，并根据这个随机的索引序列重新排布特征，得到一个新的张量$\boldsymbol{F}'' \in \mathbf{R}^{C\times H\times W}$。再将张量$\boldsymbol{F}''$的形状恢复至和输入特征 $\boldsymbol{F}_3 \in \mathbf{R}^{C\times H\times W}$ 的形状相同，得到特征 $\boldsymbol{F}_6 \in \mathbf{R}^{C\times H\times W}$，特征打乱的过程可表示为

$$\boldsymbol{F}_6 = \mathrm{reshape}\left(\boldsymbol{F}'\left[\ :,\quad :,\quad \boldsymbol{I}\ \right]\right)$$

式中：reshape 是对一维向量的形状进行调整的操作。$\boldsymbol{F}'\left[\ :,\quad :,\quad \boldsymbol{I}\ \right]$表示仅对最后一个维度的索引进行随机排列，而前两个维度保持不变。$\boldsymbol{I}$ 为前文提到的索引序列。

传统注意力机制易过度依赖局部空间模式，带来两方面问题：一是将偶然的空间相关性误判为语义关联，引发过拟合；二是对几何变换和遮挡过于敏感，在复杂场景下性能下降。为此，特征打乱模块通过全局重排特征图的空间布局，削弱模型对绝对位置的依赖，促使其关注更具语义价值的特征关系。这种空间解耦策略不仅降低了对局部纹理和背景噪声的敏感性，还增强了模型在遮挡与形变下的鲁棒性。通过随机置换空间信息，网络被引导学习跨布局保持一致的特征表征，从而更好地捕捉远程依赖并提升全局信息整合能力。

位置编码在特征打乱中充当空间锚点，通过提供绝对坐标基准，在破坏原始结构的同时保持几何一致性。特征打乱削弱模型对绝对位置的依赖，引导其学习基于内容的关联，而位置编码则维持必要的空间约束，实现局部与全局关系的协同建模，从而提升模型在复杂场景下的鲁棒性与泛化能力。如图 5 所示。

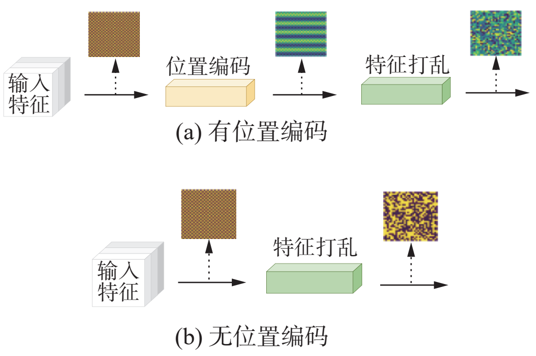


图 5　位置编码对于特征打乱的影响

Fig. 5　Impact of position encoding on feature shuffling

由图 5 可知，图 5(a) 与图 5(b) 输入相同，均为棋盘格图像。经过位置编码后，输出特征呈现明显的横向条纹，说明模型的空间位置敏感性和空间感知能力得到增强。输入原本仅由两个离散值构成，而位置编码使特征分布变得连续，信息更丰富，不仅包含内容，还嵌入了位置信息。这种增强的连续性提升了特征表达能力，并为特征打乱提供了位置信息补偿，使模型在结构扰动下依然能够有效建模空间关系。

需要指出的是，本文的特征打乱与 ShuffleNet 的通道打乱存在本质差异。本文方法在同一通道内随机扰动空间位置，旨在实现全局空间解耦与鲁棒性增强；而 ShuffleNet 则在通道维度进行规则化重排，通过分组卷积后的张量变换实现跨组信息交互，同时保持空间结构完整。二者在操作维度、结构保持性和应用目的上均显著不同。

2.2.5　随机池化模块

对打乱后的特征$\boldsymbol{F}_6 \in \mathbf{R}^{C\times H\times W}$进行随机池化操作得到特征$\boldsymbol{F}_7 \in \mathbf{R}^{C\times 1\times 1}$。随机池化模块如图 6 所示。

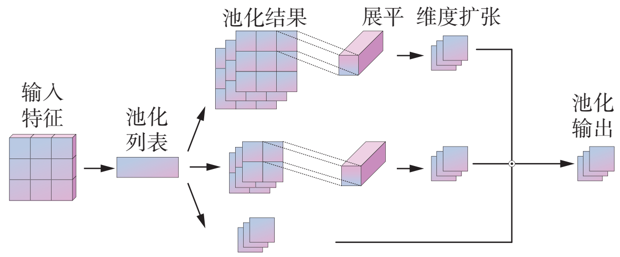


图 6　随机池化模块

Fig. 6　Randomized pooling module

随机池化模块中首先定义一个池化列表，池化列表由几个不同的整数构成，这些整数就是池化目标输出尺寸，针对每一个池化目标大小创建一个自适应平均池化层。动态地随机选择一个整数作为自适应平均池化的尺寸。但如果选择的池化尺寸不是 1，则先将特征展平成一维向量，从空间维度中随机选择一个位置，提取该位置元素，并将其作为新的张量，再将张量的维度进行扩张得到1×1尺寸的特征图。

特征打乱与位置编码共同作用于随机池化模块，构建了一种动态平衡的调控机制。前者通过扰乱输入特征的空间连续性，促使随机池化弱化对局部结构的依赖，更关注具备泛化能力的语义关联；而后者则为扰乱后的特征提供稳定的空间参考，确保池化过程在抽象特征时仍能保留关键的几何信息。两者协同引导随机池化，在适应结构扰动的同时，实现多尺度特征的有效提取，并协同建模语义与几何信息。

随机池化模块通过不同尺度的池化操作捕捉多尺度上下文信息，小尺度池化聚焦局部细节，大尺度池化提取全局特征，增强模型对整体结构的理解。多尺度池化丰富特征表示，提高注意力图的精确性与多样性，使网络更灵活地关注关键

信息，从而提升模型的准确性和鲁棒性，尤其在复杂任务中表现更优。

将特征 $\boldsymbol{F}_7$ 进行上采样操作后传入 SELayer 得到特征 $\boldsymbol{F}_8 \in \mathbf{R}^{C\times H\times W}$，SELayer 是将 1×1 尺寸的卷积、ReLU激活函数和Sigmoid激活函数堆叠起来组成的模块，具体过程可表示为

$$\boldsymbol{F}_8 = \mathrm{Sigmoid}(\mathrm{Conv}(\mathrm{ReLU}(\mathrm{Conv}(\boldsymbol{F}_7))))$$

式中ReLU激活函数为

$$\mathrm{ReLU}(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & \boldsymbol{x} \geqslant 0 \\ 0, & \boldsymbol{x} < 0 \end{cases} \tag{3}$$

ReLU激活函数对输入 $\boldsymbol{x}$ 的处理方式为：当输入为负时输出置零，当输入为正时则保持不变。该机制在函数中引入非线性，使模型能够突破线性映射的限制，从而提升神经网络对复杂模式与高维特征的建模能力。

Sigmoid激活函数为

$$\mathrm{Sigmoid}(\boldsymbol{x}) = \frac{1}{1+\mathrm{e}^{-\boldsymbol{x}}} \tag{4}$$

Sigmoid激活函数的主要特性在于将输入映射到 (0,1) 区间，其输出具有概率意义，因此常用于二分类任务。在通道注意力建模中，Sigmoid函数能够通过不同的概率值反映各通道的重要性，从而突出关键通道特征并抑制冗余信息。作为一种典型的非线性函数，Sigmoid还能够增强网络对复杂特征模式的学习与表达能力。

Conv 是尺寸为 1×1 的卷积操作。使用 1×1 的卷积对通道进行调整，将通道数在输入通道 C 和中间通道 M 之间进行转换，M 的计算公式为

$$M = \max\left(\left\lceil \frac{C}{4} \div 8 \right\rceil \times 8, 32\right) \tag{5}$$

式中：C 为输入通道数，4 为压缩因子，其作用在于将输入通道数缩减至原始规模的四分之一，通过向上取整操作保证调整后的通道数既为 8 的整数倍，又不小于理论计算结果，从而满足硬件计算友好性与模型结构约束。利用max函数限定通道数下限为 32，以避免中间通道数过小导致关键信息的损失。采用中间通道数 M 进行特征变换，不仅能够有效降低数据维度与计算复杂度，还能够在压缩的同时保留和提炼输入特征中的重要信息。

最后，将特征 $\boldsymbol{F}_8 \in \mathbf{R}^{C\times H\times W}$ 与输入特征 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$ 进行逐元素相乘后的结果与特征 $\boldsymbol{F}_5 \in \mathbf{R}^{C\times H\times W}$ 和经过上采样后的特征 $\boldsymbol{F}_2' \in \mathbf{R}^{C\times H\times W}$ 进行相加求和操作，得到空间约束注意力机制的输出特征 $\boldsymbol{F}_{\mathrm{out}} \in \mathbf{R}^{C\times H\times W}$，可表示为

$$\boldsymbol{F}_{\mathrm{out}} = \boldsymbol{F}_8 \cdot \boldsymbol{F} + \boldsymbol{F}_5 + \boldsymbol{F}_2'$$

这种输出形式分别从空间选择性、几何结构保持与语义增强的角度对特征进行建模，并通过残差连接维持原始信息的完整性，构建出相互协同且互补正交的特征表达体系。

2.3　边缘感知卷积模块

在整个 SCAM-Net 中，EAConv 卷积是用来代替原 WideResnet-28-10 网络中首层卷积核尺寸为 3×3 的卷积。在保留原卷积将 RGB 通道转化为 16 通道的功能的基础上，增强了对低阶特征提取的能力，实现了对局部特征、深度局部特征和边缘特征的特征融合。通过多层次特征提取，并结合特征融合和残差连接，为后续网络提供了更丰富的特征，使得被原始尺寸为 3×3 的卷积核忽略掉的边缘特征重新利用了起来。EAConv 卷积结构如图 7 所示。

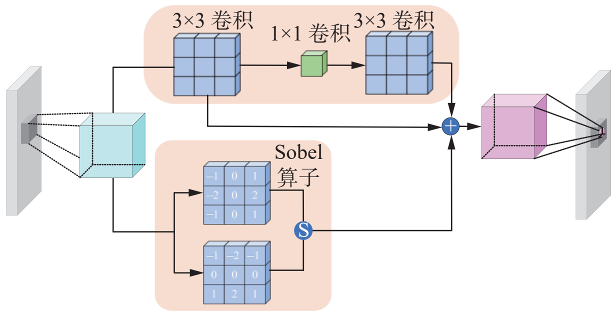


图 7　EAConv 模块结构

Fig. 7　EAConv module architecture

对于 EAConv 的输入特征 $\boldsymbol{x}$ 为 $\boldsymbol{x}:(B,C_{\mathrm{in}},H,W)$，其中 C_{in} 为输入通道数，H 为特征图高度，W 为特征图宽度。EAConv 拥有 2 个分支，第一个分支进行局部特征提取和深度局部特征提取，第二个分支进行边缘特征提取。

首先，在第一个分支中通过 3×3 卷积、BN (batch normalization) 操作、LeakyReLU 激活函数进行局部特征提取，输入特征 $\boldsymbol{x}$ 经过局部特征提取模块，得到输出特征 $\boldsymbol{X}_1 \in \mathbf{R}^{C\times H\times W}$，其中 C 为第一个残差块 Layer1 的输入通道数，整个过程具体可表示为

$$\boldsymbol{X}_1 = \mathrm{LeakyReLU}(\mathrm{BN}(\mathrm{Conv}(\boldsymbol{x})))$$

其中 LeakyReLU 激活函数公式为

$$\mathrm{LeakyReLU}(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & \boldsymbol{x} > 0 \\ \alpha\boldsymbol{x}, & \boldsymbol{x} \leqslant 0 \end{cases} \tag{6}$$

式中：α 是负斜率，可根据具体任务需求进行调整。LeakyReLU 激活函数通过在负半轴区域保留非零梯度，有效缓解了传统 ReLU 激活函数可能导致的“神经元失活”问题，从而提升训练过程的稳定性。该函数还能够在保留更多特征信息的基础上增强模型的非线性表达能力。

将输出的局部特征$\boldsymbol{X}_1 \in \mathbf{R}^{C\times H\times W}$依次输入到$1\times1$卷积，Dropout 操作，$3\times3$卷积。其中每个卷积操作中还包含了 BN 操作和 LeakyReLU 激活函数，得到深度局部特征$\boldsymbol{X}_2 \in \mathbf{R}^{C\times H\times W}$，可表示为

$$\boldsymbol{X}_2 = \mathrm{Conv}_2\left(\mathrm{Dropout}\left(\mathrm{Conv}_1\left(\boldsymbol{X}_1\right)\right)\right)$$

其中Conv_1和Conv_2分别指代了卷积核尺寸为1×1和卷积核尺寸为3×3的卷积。每个卷积的内部还包含了 BN 操作和 LeakyReLU 激活函数。

在传递给第 1 个分支的同时，还传递给第 2 个分支进行边缘特征的提取。在这一部分利用 Sobel 算子[25] 提取图像的边缘特征，其中卷积核$\boldsymbol{S}_x$、$\boldsymbol{S}_y$分别来检测图像中水平方向的边缘和垂直方向的边缘。卷积核$\boldsymbol{S}_x$、$\boldsymbol{S}_y$分别表示为

$$\boldsymbol{S}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \boldsymbol{S}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

对于输入图像$\boldsymbol{x}$使用这两个卷积核进行卷积，得到水平方向的梯度$\boldsymbol{G}_x$和垂直方向的梯度$\boldsymbol{G}_y$，计算过程可分别表示为

$$\boldsymbol{G}_x = \boldsymbol{S}_x(\boldsymbol{x})$$

$$\boldsymbol{G}_y = \boldsymbol{S}_y(\boldsymbol{x})$$

利用水平方向的梯度$\boldsymbol{G}_x$和垂直方向的梯度$\boldsymbol{G}_y$可得到边缘特征$\boldsymbol{X}_3 \in \mathbf{R}^{C\times H\times W}$，过程可表示为

$$\boldsymbol{X}_3 = \sqrt{\boldsymbol{G}_x^2 + \boldsymbol{G}_y^2}$$

提取完局部特征、深度局部特征和边缘特征后，进入特征融合阶段。通过 Concat 操作在通道维度上对特征$\boldsymbol{X}_1 \in \mathbf{R}^{C\times H\times W}$、$\boldsymbol{X}_2 \in \mathbf{R}^{C\times H\times W}$、$\boldsymbol{X}_3 \in \mathbf{R}^{C\times H\times W}$进行拼接，将不同层次的特征进行融合得到特征$\boldsymbol{X}_4 \in \mathbf{R}^{3C\times H\times W}$，可表示为

$$\boldsymbol{X}_4 = \mathrm{Concat}\left(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\right)$$

此时是在通道维度上进行拼接，对于特征$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$均有相同的通道数C。在经过 Concat 操作后通道数变为$3C$。

特征$\boldsymbol{X}_4$通过卷积核尺寸为1×1的卷积，实现了对通道的调整，将通道数重新调整至输出要求的通道数C，得到特征$\boldsymbol{X}_5 \in \mathbf{R}^{C\times H\times W}$，此阶段可表示为

$$\boldsymbol{X}_5 = \mathrm{Conv}\left(\boldsymbol{X}_4\right)$$

式中Conv为通过卷积核尺寸为1×1的卷积实现通道调整操作。将特征融合后的通道调整至规定的输出通道，以实现 EAConv 模块对通道数转化的功能。

最后，对初始特征x的通道调整至C，得到特征$\boldsymbol{X}_6 \in \mathbf{R}^{C\times H\times W}$，将特征$\boldsymbol{X}_6$和调整过通道数的融合特征$\boldsymbol{X}_5$进行特征相加得到 EAConv 模块的输出特征$\boldsymbol{X}_{\mathrm{out}} \in \mathbf{R}^{C\times H\times W}$，可表示为

$$\boldsymbol{X}_{\mathrm{out}} = \boldsymbol{X}_5 + \boldsymbol{X}_6$$

式中：通过 Sobel 算子显式提取高频边缘特征，结合局部卷积特征和深度局部特征，形成丰富的多尺度表示。而 WideResnet-28-10 的宽结构进一步放大了这种优势，使网络在早期就能聚焦关键轮廓信息。这种设计弥补了传统卷积神经网络早期层对边缘响应不足的问题，使得神经网络的分类效果有更进一步的提升，尤其是在纹理丰富的图像分类中表现更为突出。

3 实验与结果分析

3.1 实验环境

本实验采用的操作系统为 Ubuntu 22.04，其他环境参数分别为：RTX4090 显卡，24 GB 显存，60 GB 内存，10 核 CPU。软件环境为：PyTorch2.2.0-Ubuntu 22.04。实验训练共 200 轮，batch_size 设置为 128，初始学习率设置为 0.1，在 60、120、160 轮将学习率缩减为原来的 0.2，使用的是 SGD 优化器。训练过程中的相关超参数如表 1 所示。

表 1 实验相关超参数相关信息

Table 1 Details of experimental hyperparameters

超参数	设定值
lr	0.1
net_type	‘wide-resnet’
depth	28
widen_factor	10
dropout	0.3
batch_size	128
epoch	200
optim_type	‘SGD’
momentum	0.9
weight_decay	5×10^{-4}
learning_rate_schedule	分段衰减

实验分别在 4 个公共数据集上进行测验，分别是数据集 CIFAR-100[26]、CIFAR-10[26]、SVHN[27] 和 GTSRB[28]。

本实验是在随机剪裁、水平翻转和图像的标准化处理操作的基础上进行的，未采用其他额外的数据增强手段。实验的评价指标为模型对数据集图像的分类准确率。

3.2 SCA 注意力机制插入位置实验

在分类网络中，注意力机制位置的不同也可能会导致分类网络的分类准确率不同。在浅层位置，它更倾向于增强模型对局部特征的捕获能

力，例如边缘和纹理信息；而在深层位置，它更能聚焦全局语义相关的重要区域，提升对复杂模式的理解能力。这种差异的原因在于网络中特征表达的层次性，浅层特征主要包含低级的空间信息，而深层特征逐渐聚合出高级语义和全局上下文。此外，特征的维度和分布在不同层次中也有所变化，这影响了注意力机制对重要特征的权重分配，从而导致其在不同位置展现出不同的效果。因此，SCA 注意力机制插入位置实验的设定对确定网络的整体模型尤为重要。本文设计了 7 种不同的实验方案，如图 8 所示。

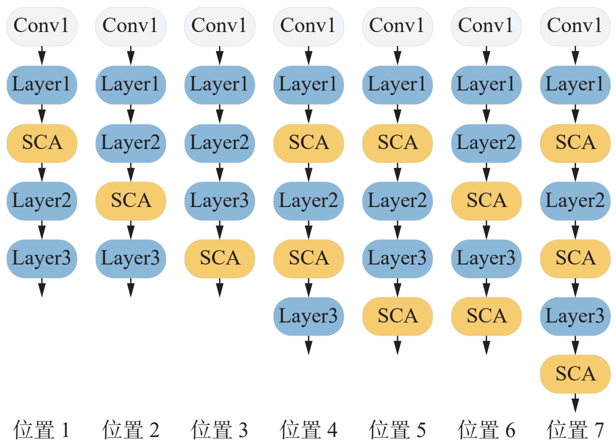


图 8　SCA 注意力机制位置和数量排列结构

Fig. 8　Arrangement structure of SCA attention mechanism positions and quantities

为了保证实验的准确性，本次实验将其他参数都设置相同，在上述 4 种数据集上进行实验，仅测试不同位置和不同数量的 SCA 注意力机制对分类网络的分类准确率的影响。实验结果如表 2 所示。

表 2　不同位置对准确率的影响

Table 2　Impact of different positions on accuracy　%

位置	CIFAR100	CIFAR10	SVHN	GTSRB
位置1	76.23	95.06	97.02	96.90
位置2	78.86	95.34	96.91	97.04
位置3	79.38	96.02	97.15	97.57
位置4	77.23	95.44	97.05	96.74
位置5	77.22	95.20	96.98	97.21
位置6	77.98	95.87	97.10	97.30
位置7	77.59	94.73	96.89	97.47

从实验结果上看，SCA 注意力机制放在位置 3 的时候分类准确率最高，在 4 个数据集上准确率分别达到了 79.38%、96.02%、97.15%、97.57%。说明 SCA 注意力机制插在深层网络中效果较好，通过前几层网络的提取，特征已经高度抽象，特征图已经提取出丰富的深层次的语义信息，包含丰富的上下文关系和抽象特征，此时再利用 SCA 注意力机制强化了局部相关性和空间细节，使得分类网络能够在语义特征丰富的基础上进一步提升特征质量，从而增强分类任务的性能。因此将 SCA 注意力机制加在位置 3 时，能够提升分类的准确率，使得整个分类网络的性能进一步提升。

3.3　注意力机制对比实验

为了证实 SCA 注意力机制对 SCAM-Net 网络分类效率提升的有效性，本次实验选取了 4 种不同的注意力机制与本文提出的 SCA 注意力机制进行对比。因此针对 4 种不同的注意力机制提出 4 种不同的方案。方案 1 是文献 [29] 提出的 SE(squeeze-and-excitation) 注意力，主要用于提升卷积神经网络的特征表达能力。方案 2 是文献 [30] 提出的 SK(selective kernel attention) 注意力，是一种动态通道注意力机制。方案 3 是文献 [22] 提出的 CA(channel attention) 注意力，通道注意力机制可以应用于多种网络架构中，并且它是 SE 注意力的核心思想之一。方案 4 是文献 [31] 提出的 CBAM(convolutional block attention module) 注意力，是 CA 注意力的加强版本，融合了 CA 注意力和 SA(spatial attention) 两个部分。

本次实验是将这 4 种注意力机制替换 SCAM-Net 网络中 SCA 注意力机制，将这 4 种注意力模块添加在前文位置 3 的地方，EAConv 卷积内部的卷积核尺寸使用方案 1，其他参数保持不变，在 CIFAR-100、CIFAR-10、SVHN、GTSRB 数据集上进行 200 轮迭代，实验结果如表 3 所示。

表 3　不同注意力机制的实验结果

Table 3　Experimental results of different attention mechanisms　%

注意力	CIFAR100	CIFAR10	SVHN	GTSRB
SE	79.14	96.24	97.05	97.19
CA	79.89	95.63	97.15	97.08
CBAM	79.24	95.92	97.15	97.12
SK	79.87	96.16	96.97	97.00
SCA	80.40	96.28	97.25	97.87

实验结果表明，在 CIFAR-100 和 SVHN 数据集上加入 CA 与 SK 注意力均能显著提升准确率，但整体效果仍弱于 SCA。这是因为 CA 与 SK 更偏重多尺度或通道间特征加权，而未像 SCA 那样专注于精细特征的捕捉。SE 在 GTSRB 数据集上的准确率仅次于 SCA，其通过通道缩放增强有意义的特征，在位置 3 处的高层语义信息上能发挥

作用，但因缺乏对局部与精细空间特征的深入建模，表现略低于 SCA。CBAM 在 CIFAR-100 和 SVHN 上也取得不同程度提升，仅次于 SCA，其通过通道与空间注意力动态调整特征重要性，强化有用信息并抑制噪声，虽在全局特征建模上效果良好，但对深层复杂特征的细粒度挖掘仍不足。

SCA 注意力机制和其他注意力机制相比，在 4 个数据集上分别平均提高了 0.87%、0.29%、0.17%、0.77%，说明本文提出的空间约束注意力能够在几种不同的数据集上提供更好的分类准确率。SCA 注意力机制能够实现在深层次的网络结构中对精细的特征选择与空间关系建模的能力。并且，本文提出的 EAConv 卷积可以提取更细致的局部特征，SCA 注意力模块在深层网络中的加入能够进一步提升这些细节特征的表达。SCA 注意力机制通过在这些基础上进行加权，增强了局部特征的关注，尤其是当这些特征在深层次网络中变得更加抽象时，SCA 注意力机制能够精细地增强和优化这些局部特征，使得网络在深层能够更加精准地处理复杂的空间关系和特征交互。EAConv 卷积和 SCA 注意力机制相互协同作用，实现了“1+1>2”的作用，使神经网络对分类精度得到进一步提升。

3.4　卷积对比实验

为了进一步验证 EAConv 卷积的有效性，选取不同的卷积与本文所提出的 EAConv 卷积进行对比。本次实验选取 3 种卷积作为对比分析模块，3 种卷积分别是由文献 [32] 提出的深度可分离卷积 DSConv(depthwise separable convolution)，文献 [33] 提出的可变形卷积 DConv(deformable convolution)，文献 [34] 提出的条件卷积 CondConv(conditional convolutions)。

实验的具体操作为，使 SCA 注意力机制在前文位置 3 的位置保持不变，只将首层的 EAConv 卷积替换为这 3 种卷积，控制其他参数保持一致，在 CIFAR-100、CIFAR-10、SVHN、GTSRB 数据集进行 200 轮迭代。实验结果如表 4 所示。

表 4　不同卷积对分类的影响
Table 4　Impact of different convolutions on classification　%

卷积种类	CIFAR100	CIFAR10	SVHN	GTSRB
DSConv	78.46	95.81	97.02	97.43
DConv	79.34	95.43	97.20	97.34
CondConv	79.28	95.70	97.10	97.07
EAConv	80.40	96.28	97.25	97.87

由实验结果分析可知，在 CIFAR-100 和 SVHN 数据集上采用 DConv 和 CondConv 在分类精度上都有较大幅度的提升。DConv 通过灵活调整卷积核的采样位置，使网络能够在输入数据中精确地捕捉到更加丰富的空间信息和局部细节，从而有效补充了注意力机制的空间选择能力，使得网络能够在更复杂的输入数据中更加精确地识别和提取高质量的特征，从而在精度上取得了提升；而 CondConv 通过动态调整卷积核参数，使得网络能够针对不同输入生成适合的卷积核，从而捕捉更多样化的特征；与 DConv 和 CondConv 不同的是，在 CIFAR-100 数据集上将初始卷积替换为 DSConv 后，神经网络对于图像分类的精度的提升效果不明显，主要是因为 DSConv 通过将标准卷积拆解为深度卷积和逐点卷积来降低计算复杂度，这虽然提高了效率，但可能会损失一些信息流的表达能力，尤其在特征提取的初期阶段，DSConv 可能未充分保留这些信息，导致进入后续层的特征不够丰富，从而使得分类效果并不理想。

与其他 3 种卷积相比，EAConv 卷积在 4 种数据集上分别平均提升了 1.70%、0.63%、0.14%、0.59%。说明本文提出的 EAConv 卷积在局部和边缘等细节特征上提取效果较好，拥有较好的整合不同层次信息的能力，可以将更高质量的抽象特征提供给注意力机制，进而使得神经网络对图像分类精度有更进一步的提升。

3.5　消融实验

为了更进一步证明 SCA 注意力机制和 EAConv 卷积对分类网络的影响性，本文设计了分类网络的消融实验。该消融实验设置了 4 组不同的方案。方案 1：WideResnet-28-10 基线模型；方案 2：WideResnet-28-10 和 EAConv 卷积；方案 3：WideResnet-28-10 和 SCA 注意力机制；方案 4：WideResnet-28-10 和 EAConv 卷积和 SCA 注意力机制。消融实验在 4 种不同的数据集上进行，4 种数据集分别是 CIFAR-100、CIFAR-10、SVHN、GTSRB。为了控制变量，在本次消融实验中，将 SCA 注意力置于前文所提到的位置 3，将 EAConv 放置于 WideResnet-28-10 网络的首层卷积位置，以确保分类拥有较高的准确率。在数据预处理阶段仅采用随机剪裁、水平翻转、图像的标准化处理等数据增强。在 4 种数据集上分别进行 200 轮迭代，实验结果如表 5 所示。

由实验结果分析可得，与原模型 WideResnet-28-10 相比，在 CIFAR-100、CIFAR-10、SVHN、GTSRB 4 种数据集上方案 2 的分类准确率分别提升

了 0.5%、0.6%、0.09%、0.16%；方案 3 的分类准确率分别提升了 1.41%、0.67%、0.04%、0.61%；方案 4 的分类准确率分别提升了 2.43%、0.93%、0.14%、0.91%。与原 WideResnet-28-10 模型相比，方案 4 对分类准确率的提升最大。由于 CIFAR-100 是分类数量较多的数据集，模型可以获取更丰富的特征，因此模型在 CIFAR-100 上所获得的提升最大。CIFAR-10、SVHN、GTSRB 这些数据集的分类数较少，且分类精度普遍较高，使得提升不大。SCA 注意力机制和 EAConv 卷积对网络的分类准确率都有提升，且同时加入 SCA 注意力机制和 EAConv 卷积对网络的分类准确率提升最大。

表 5　消融实验的分类准确率
Table 5　Classification accuracy in ablation studies　%

方案	CIFAR100	CIFAR10	SVHN	GTSRB
方案1	77.97	95.35	97.11	96.96
方案2	78.47	95.95	97.20	97.12
方案3	79.38	96.02	97.15	97.57
方案4	80.40	96.28	97.25	97.87

3.6　对比实验

为了验证 SCAM-Net 的有效性，将本文网络模型和常见的分类网络模型 WideResnet-28-10、ResNet-34、PreActResNet18[35]、DFR-DenseNet[36]、EfficientNets、VGG11B[37]、CT-6/3x1[38]、QKFormer[39]、FAVOR+[40]、Couplformer[41]、MDTC[42]、CN2Conv[43]、MHTNA-ResNet[44]、MSMDFormer[45]、MPCNet[46]。在 CIFAR-100、CIFAR-10、SVHN、GTSRB 数据集上进行对比实验。本文中对各分类网络的准确率比较实验数据来源如下：对于未提供开源代码的网络，优先采用对比网络所对应论文提供的实验结果；对于提供开源代码的网络，分类准确率通过对代码进行复现实验得出。各网络的分类准确率如表 6 所示，未知准确率用符号“—”表示。

通过表 6 可知，本文提出的 SCAM-Net 网络在分类准确率上优于表格中提到的其他 15 种分类网络，在 4 个数据集上分别达到了 80.40%、96.28%、97.25%、97.87%。

在分类准确率上表现较好的 3 个分类网络分别是 VGG11B、PreActResNet18 和 QKFoemer。VGG11B 通过深度卷积层与小卷积核的堆叠设计，逐层累积感受野，有效捕捉图像细粒度纹理，因此在 CIFAR-100 等类别众多且差异细微的数据集上表现良好。然而，深度堆叠带来较大计算开销，且在数据量有限或数据增强不足时易过拟合。实验表明，VGG11B 的精度提升主要依赖局部细节建模，但在全局依赖和空间关系刻画上仍不足，这限制了其在复杂场景下的性能。

表 6　其他网络的分类准确率
Table 6　Classification accuracy of other networks　%

网络	CIFAR100	CIFAR10	SVHN	GTSRB
文献[21]	77.97	95.35	97.11	96.96
文献[1]	69.41	87.89	91.51	95.10
文献[35]	79.48	95.81	96.89	94.35
文献[36]	78.15	94.53	—	—
文献[3]	75.97	94.00	95.32	92.18
文献[37]	79.91	94.02	96.93	95.06
文献[38]	77.33	95.30	96.80	96.73
文献[39]	80.27	96.18	97.13	97.53
文献[40]	72.56	91.42	93.21	—
文献[41]	73.92	93.54	94.26	95.84
文献[42]	77.30	95.64	95.89	—
文献[43]	75.69	95.02	—	—
文献[44]	75.31	93.65	—	—
文献[45]	77.46	95.65	96.87	—
文献[46]	66.54	90.61	—	—
SCAM-Net	80.40	96.28	97.25	97.87

PreActResNet18 在残差块中引入预激活结构，将批归一化和激活置于卷积前，改善梯度传播并增强正则化，使训练更稳定、收敛更快，有效缓解深层网络的梯度消失问题并提升复杂特征表达能力，因此在 CIFAR-10 和 SVHN 上表现良好。同时，预激活在保持低计算量的同时增强泛化能力，使网络高效学习判别特征。但由于深度和宽度有限，PreActResNet18 在 CIFAR-100 等类别众多、特征复杂的数据集上仍存在不足。

QKFormer 基于查询−键注意力机制，通过聚焦关键区域并抑制冗余信息，增强了对判别特征的捕捉能力，在 CIFAR-100 等类别众多、类间差异细微的数据集上显著提升类间可分性。实验表明，其性能提升主要依赖全局建模与关键区域选择，而非深度或宽度扩展，使其在计算量可控的情况下取得高分类精度。但由于依赖注意力计算，在大规模或高分辨率数据上可能面临效率挑战。

3.7　SCAM-Net 网络性能

3.7.1　计算量 FLOPs 和参数量 Params 相关实验

参数量 Params 是指模型中所有可训练参数的总数，计算量 FLOPs 是指完成一次前向传播所

需的浮点运算次数。本次实验将使用 SCAM-Net 与 WideResnet-28-10、EfficientNet-B7、VGG11B 网络对计算量和参数量进行对比分析，相关网络的计算量和参数量如表 7 所示。

表 7　不同网络的 FLOPs 和 Params 对比

Table 7　Comparison of FLOPs and Params for different networks

网络	浮点运算速度/10^9 s^{-1}	参数量/10^6
WideResnet-28-10	5.96	36.5
EfficientNet-B7	37.00	66.0
VGG11B	15.50	132.9
SCAM-Net	6.02	39.0

在上述网络中，EfficientNet-B7 的计算量最高，这是由于其复合缩放策略同时提升了网络的深度、宽度与分辨率。但得益于 MBConv 和 SE 机制的优化，其参数量控制在 $66×10^6$。VGG11B 的参数量最高，主要源于其全卷积结合全连接的结构设计，尽管计算量低于 EfficientNet-B7，仍远高于 WideResNet-28-10。WideResNet-28-10 通过增加宽度而非深度，在实现较低计算量的同时保持了 $36.5×10^6$ 的参数规模，达到了计算与存储的较好平衡。而 SCAM-Net 在 WideResnet-28-10 的基础上增加了卷积和注意力机制，使其在提升分类精度的同时增加了少量的计算量和参数量，实现了性能与资源消耗之间的权衡。

随着参数量和计算量的增加，模型的时间复杂度相应提升。表 8 列出了 SCAM-Net 与 WideResNet-28-10 在 CIFAR-100、CIFAR-10、GTSRB 及 SVHN 4 个数据集上的单轮运行时间，所有实验均在 P100 GPU 环境下完成。

表 8　SCAM-Net 和 WideResNet-28-10 运行单轮次时间对比

Table 8　Comparison of single-epoch running time between SCAM-Net and WideResNet-28-10

网络	CIFAR100	CIFAR10	GTSRB	SVHN
WideResnet-28-10	2 min 38 s	2 min 38 s	1 min 36 s	3 min 50 s
SCAM-Net	3 min 5 s	3 min 5 s	1 min 51 s	4 min 36 s

SCAM-Net 基于 WideResNet-28-10，通过引入注意力机制并替换部分卷积结构构建而成。尽管计算量与参数量有所上升，导致训练时间相应增加，但其有效提升了特征表达能力，最终带来了更优的分类性能。

3.7.2　Recall 和 F_1-Score 及相关实验

召回率（Recall）是衡量模型对正类样本的识别能力的参数，表示所有真实正类样本中被正确预测的比例。较高的 Recall 说明模型在正类样本的检测上具有较强的覆盖能力。F_1-Score 结合了精确率（Precision）和召回率，通过它们的调和平均数提供整体性能评估。在分类任务中，F_1-Score 较高意味着模型在正确识别正类样本的同时，也能有效减少误报。本次实验在 CIFAR-100、CIFAR-10、GTSRB 3 个数据集上对基线模型 WideResnet-28-10 和 SCAM-Net 的 Recall 和 F_1-Score 进行分析对比，实验结果如表 9 和表 10 所示。

表 9　SCAM-Net 和 WideResnet-28-10 在 3 个数据集上的 Recall 对比

Table 9　Comparison of Recall between SCAM-Net and WideResNet-28-10 on three datasets

网络	CIFAR100	CIFAR10	GTSRB
SCAM-Net	0.8033	0.9622	0.9784
WideResnet-28-10	0.7707	0.9459	0.9608

表 10　SCAM-Net 和 WideResnet-28-10 在 3 个数据集上的 F_1-Score 对比

Table 10　Comparison of F_1-Score between SCAM-Net and WideResNet-28-10 on three datasets

网络	CIFAR100	CIFAR10	GTSRB
SCAM-Net	0.8038	0.962	0.9781
WideResnet-28-10	0.7748	0.9457	0.9578

由表 9 和表 10 可知，在 3 个数据集上 SCAM-Net 的 Recall 和 F_1-Score 均高于 WideResnet-28-10，这表明 SCAM-Net 在分类任务中的整体表现优于 WideResNet-28-10，能够在保持较高召回率的同时，确保误分类率较低，说明该模型对于目标类别的判别更加精准。

PR（Precision-Recall）曲线是用于评估分类模型性能的曲线，它通过精确率（Precision）和召回率（Recall）的关系，直观地展示模型在不同阈值下的分类能力。曲线下的面积越大，表示模型的分类能力越强。本次实验分别在 CIFAR-100、CIFAR-10 和 GTSRB 3 个数据集上对基线模型 WideResNet-28-10 和 SCAM-Net 的 PR 曲线进行可视化分析，实验结果如图 9 所示。

在绘制 PR 曲线时聚焦于置信度较高的区域进行分析，在进行 GTSRB 数据集的实验时为了使实验结果更易于分析，在数据预处理阶段加入少量高斯噪声。由图 9 可知，在 CIFAR-100、CIFAR-10 和 GTSRB 3 个数据集上 SCAM-Net 的曲线下的面积更大，表明在 Precision 与 Recall 两项指标上均取得提升，进一步验证了模型在分类任务中的性能优势。

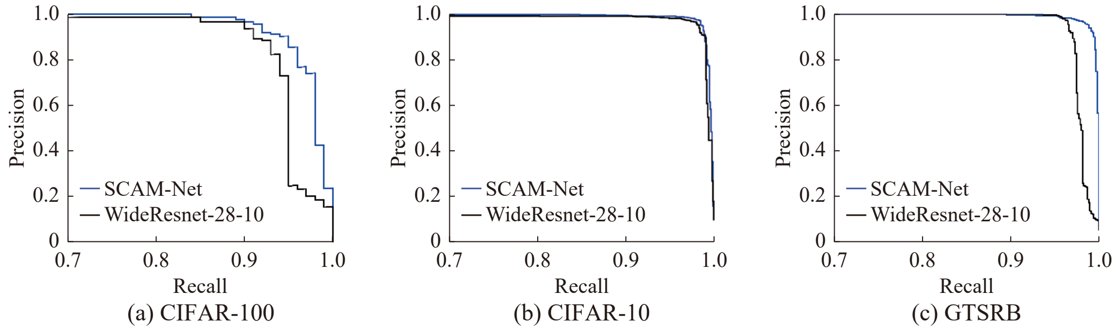


图 9　WideResnet-28-10 和 SCAM-Net 在两个数据集上的 PR 曲线对比

Fig. 9　Comparison of PR curves between WideResNet-28-10 and SCAM-Net on two datasets

3.7.3　收敛性验证实验

为了验证本文提出的 SCAM-Net 图像分类网络的有效性和训练稳定性，设计了收敛性验证实验。通过分析模型训练过程中的 ACC(accuracy) 曲线评估模型是否能够高效、稳定地收敛到最优解。SCAM-Net 和 WideResNet-28-10 在 CIFAR-100、CIFAR-10、GTSRB 和 SVHN 数据集上的 ACC 曲线如图 10 所示。

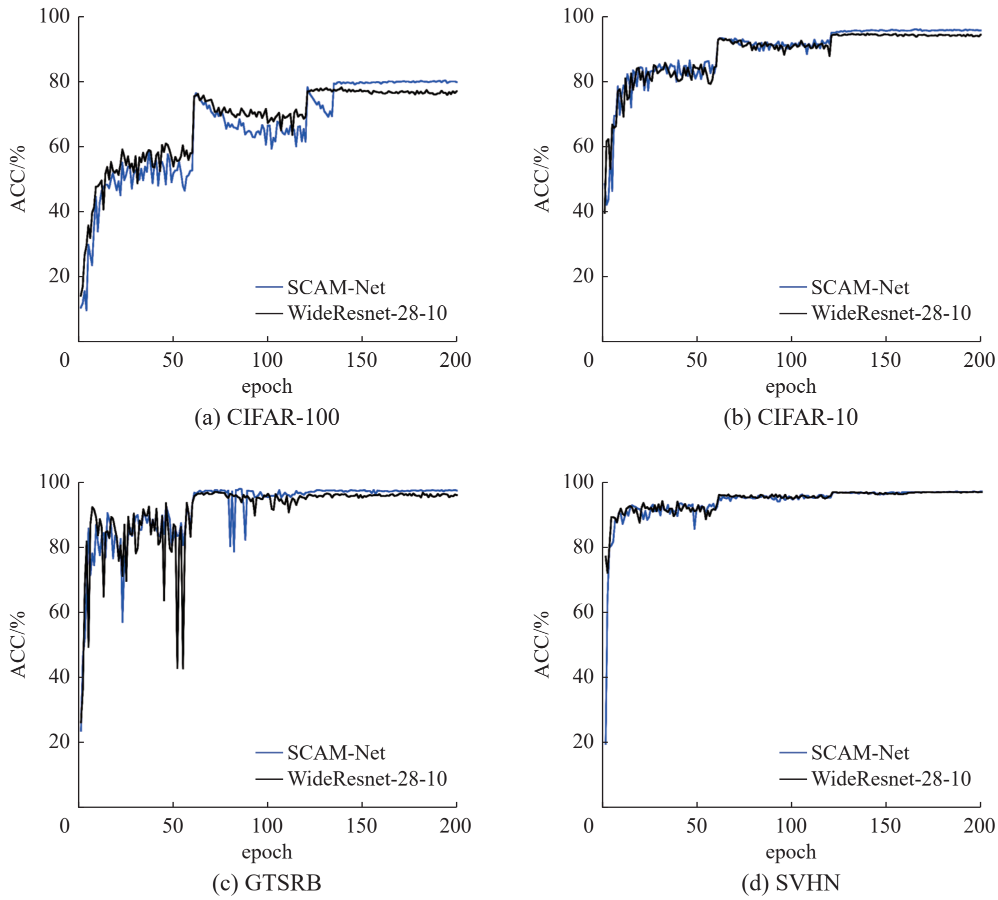


图 10　4 个数据集下 SCAM-Net 和 WideResnet-28-10 的 ACC 曲线

Fig. 10　ACC curves of SCAM-Net and WideResNet-28-10 under four datasets

由图 10 的 ACC 曲线可以看出，SCAM-Net 在第 60 轮和第 120 轮均出现明显的上升，这主要受学习率调整的影响。在 4 个数据集上，SCAM-Net 均于第 120 轮后曲线趋于平稳，表明模型已基本学得主要特征，逐步收敛。

3.8　可视化分析实验

3.8.1　边缘特征可视化实验

为了进一步证明 EAConv 模块对边缘特征提取的有效性，分别对 EAConv 模块和 WideResnet-28-10 网络原初始卷积层3×3卷积进行对比，随机选择

5 张图片, 对其提取边缘特征的效果进行可视化分析。如图 11 所示, 不同的卷积对边缘的特征提取效果不同, 经过可视化后的图片如果边缘清晰, 则证明对边缘特征提取的效果良好, 反之效果较差。

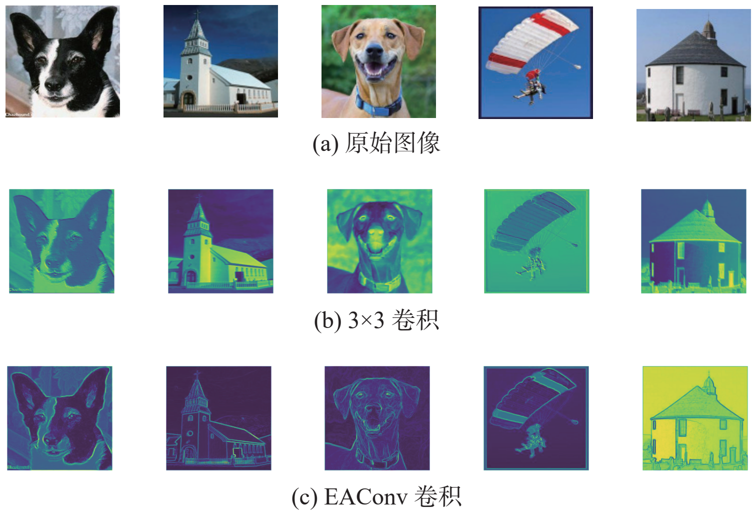


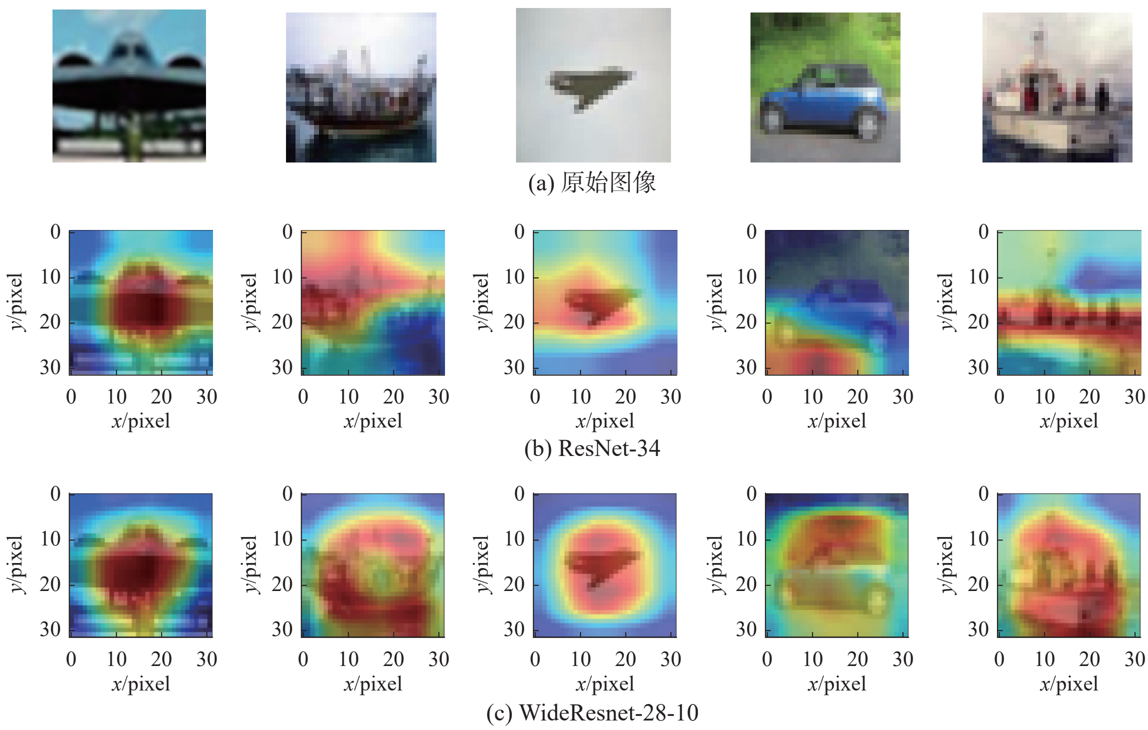
图 11　边缘特征可视化对比

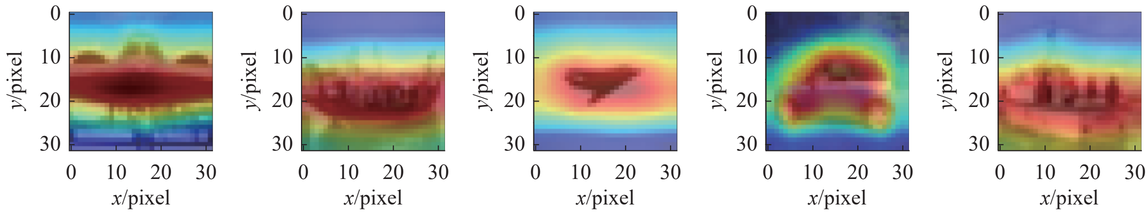
Fig. 11　Comparison of edge feature visualization

由图 11 可知, 3×3卷积对边缘特征的提取效果不如 EAConv 卷积对边缘特征的提取效果。这是因为普通3×3卷积核在初始阶段是随机的, 需要通过训练才能学习到边缘特性, 且容易受低频干扰。而结合 Sobel 边缘特征、局部特征和深度局部特征的 EAConv 卷积可以显著增强模型对图像细节的理解和区分能力, 使得在边缘特征的提取上拥有更好的效果。

3.8.2　网络特征可视化实验

为了进一步说明 SCAM-Net 对特征提取的效果, 选择与 ResNet-34、WideResnet-28-10 进行对比, 在 CIFAR-10 数据集中随机选择 5 张图片, 分别对各个网络的最后一个卷积层进行 Grad-CAM (gradient-weighted class activation mapping) 特征可视化, 如图 12 所示。不同颜色的区域代表网络对不同区域的关注度不同, 红色区域表示网络对这部分区域更加关注, 而蓝色区域表示网络对这部分区域的关注度较低。图 12 中横坐标表示图像水平方向的像素位置, 纵坐标表示图像垂直方向的像素位置, 单位均为像素 (pixel)。





(d) SCAM-Net

图 12　不同网络的特征可视化对比

Fig. 12　Comparison of feature visualizations across different networks

通过图 12 可以看出，3 种网络几乎都可以准确地定位到图像中关键的区域，但是 SCAM-Net 的效果最好。在图 12(d) 中，每个图像的红色区域几乎覆盖了原始图像中的关键区域，每个图像的整体和轮廓都有较好的显示效果，在汽车的显示效果最为明显，在路面和草丛背景的干扰下，依旧可以提取到清晰的边缘特征和关键区域的全局特征。这再一次证明了 SCAM-Net 在图像分类领域有较好的效果。

4　结束语

本文针对分类网络对低阶特征提取不充分和在深层网络中对重要特征关注度不足的问题提出了 SCAM-Net 网络。首先提出 SCA 注意力机制，使得分类网络对特征图的空间位置有更好的了解，通过 SCA 注意力机制使得分类网络更加关注特征图中重要的信息。然后提出 EAConv 卷积，并将该卷积置于分类网络首层的位置，通过 EAConv 卷积提取并融合边缘特征、局部特征等不同尺度的特征，为后续残差网络的处理提供了丰富的特征信息。SCAM-Net 在 CIFAR-100、CIFAR-10、SVHN 和 GTSRB 4 种数据集上，相较于基线模型 WideResnet-28-10 分别提升了 2.43%、0.93%、0.14%、0.91%。尽管 SCAM-Net 在图像分类任务的准确率上有着优异的表现，但是未能在提升准确率的同时减少计算量和时间复杂度，这也是 SCAM-Net 日后改进的主要方向。此外，关于 SCAM-Net 在图像分割等其他计算机视觉任务中的潜在应用价值，尚需后续研究进行系统性探索和评估。

参考文献：

[1] 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.

[2] DING Xiaohan, ZHANG Xiangyu, HAN Jungong, et al. Scaling up your kernels to 31 × 31: revisiting large kernel design in CNNs[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11953−11965.

[3] TAN Mingxing, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019−05−28)[2020−09−11]. http://arxiv.org/pdf/1905.11946.pdf.

[4] LIU Xinyu, PENG Houwen, ZHENG Ningxin, et al. EfficientViT: memory efficient vision Transformer with cascaded group attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 14420−14430.

[5] 姜文涛, 张大鹏. 优化分类的弱目标孪生网络跟踪研究[J]. 智能系统学报, 2023, 18(5): 984−993.
JIANG Wentao, ZHANG Dapeng. Research on weak object tracking based on Siamese network with optimized classification[J]. CAAI transactions on intelligent systems, 2023, 18(5): 984−993.

[6] 刘晓敏, 余梦君, 乔振壮, 等. 面向多源遥感数据分类的尺度自适应融合网络[J]. 电子与信息学报, 2024, 46(9): 3693−3702.
LIU Xiaomin, YU Mengjun, QIAO Zhenzhuang, et al. Scale adaptive fusion network for multimodal remote sensing data classification[J]. Journal of electronics & information technology, 2024, 46(9): 3693−3702.

[7] 刘佳, 宋泓, 陈大鹏, 等. 非语言信息增强和对比学习的多模态情感分析模型[J]. 电子与信息学报, 2024, 46(8): 3372−3381.
LIU Jia, SONG Hong, CHEN Dapeng, et al. A multimodal sentiment analysis model enhanced with non-verbal information and contrastive learning[J]. Journal of electronics & information technology, 2024, 46(8): 3372−3381.

[8] 王柳, 梁铭炬. 融合深度信息的室内场景分割算法[J]. 计算机系统应用, 2024, 33(3): 111−117.
WANG Liu, LIANG Mingju. Indoor scene segmentation algorithm based on fusion of deep information[J]. Computer systems and applications, 2024, 33(3): 111−117.

[9] ZHAO Youpeng, TANG Huadong, JIANG Yingying, et al. Parameter-efficient vision Transformer with linear at-

tention[C]//2023 IEEE International Conference on Image Processing. Kuala Lumpur: IEEE, 2023: 1275−1279.

[10] SARKAR R, LIANG Hanxue, FAN Zhiwen, et al. Edge-MoE: memory-efficient multi-task vision Transformer architecture with task-level sparsity via mixture-of-experts [C]//2023 IEEE/ACM International Conference on Computer Aided Design. San Francisco: IEEE, 2023: 1−9.

[11] WANG Wenxiao, CHEN Wei, QIU Qibo, et al. CrossFormer: a versatile vision Transformer hinging on cross-scale attention[J]. IEEE transactions on pattern analysis and machine intelligence, 2024, 46(5): 3123−3136.

[12] 姜文涛, 孟庆姣. 自适应时空正则化的相关滤波目标跟踪[J]. 智能系统学报, 2023, 18(4): 754−763.
JIANG Wentao, MENG Qingjiao. Correlation filter tracking for adaptive spatiotemporal regularization[J]. CAAI transactions on intelligent systems, 2023, 18(4): 754−763.

[13] YANG Jian, LI Chen, LI Xuelong. Underwater image restoration with light-aware progressive network[C]//ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes Island: IEEE, 2023: 1−5.

[14] LI Zixuan, WANG Yuangen. Optimizing Transformer for large-hole image inpainting[C]//2023 IEEE International Conference on Image Processing. Kuala Lumpur: IEEE, 2023: 1180−1184.

[15] CHEN Xiangyu, WANG Xintao, ZHANG Wenlong, et al. Hat: hybrid attention Transformer for image restoration [EB/OL]. (2023−09−11)[2025−10−01]. https://arxiv.org/abs/2309.05239.

[16] JI Jiahuan, ZHONG Baojiang, SONG Weigang, et al. Learning multi-scale features for jpeg image artifacts removal[C]//2023 IEEE International Conference on Image Processing. Kuala Lumpur: IEEE, 2023: 1565−1569.

[17] LIU Yifeng, TIAN Jing. Probabilistic attention map: a probabilistic attention mechanism for convolutional neural networks[J]. Sensors, 2024, 24(24): 8187.

[18] POLANSKY M G, HERRMANN C, HUR J, et al. Boundary attention: learning curves, corners, junctions and grouping[EB/OL]. (2024−01−01)[2025−10−01]. https://arxiv.org/abs/2401.00935.

[19] XIAO Da, MENG Qingye, LI Shengping, et al. Improving Transformers with dynamically composable multi-head attention[EB/OL]. (2024−05−17)[2025−10−01]. https://arxiv.org/abs/2405.08553.

[20] YU Xiang, GUO Hongbo, YUAN Ying, et al. An improved medical image segmentation framework with Channel-Height-Width-Spatial attention module[J]. Engineering applications of artificial intelligence, 2024, 136: 108751.

[21] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL]. (2016−05−23)[2025−10−01]. https://arxiv.org/abs/1605.07146.

[22] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11531−11539.

[23] PARMAR N, VASWANI A, USZKOREIT J, et al. Image Transformer[C]//International Conference on Machine Learning. Stockholm: PMLR, 2018: 4055−4064.

[24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 5998−6008.

[25] CHIEN Y. Pattern classification and scene analysis[J]. IEEE transactions on automatic control, 1974, 19(4): 462−463.

[26] SHARMA N, JAIN V, MISHRA A. An analysis of convolutional neural networks for image classification[J]. Procedia computer science, 2018, 132: 377−384.

[27] NETZER Y, WANG T, COATES A, et al. The street view house numbers (SVHN) dataset[EB/OL]. (2011−12−12)[2023−05−04]. http://ufldl.stanford.edu/housenumbers/.

[28] STALLKAMP J, SCHLIPSING M, SALMEN J, et al. The German traffic sign recognition benchmark [EB/OL]. (2012−03−16)[2023−05−04]. http://benchmark.ini.rub.de/?section=gtsrb&subsection=news.

[29] HU Jie, SHEN Li, SAMUEL A, et al. Squeeze-and-excitation networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 42(8): 1.

[30] LI Xiang, WANG Wenhai, HU Xiaolin, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 510−519.

[31] 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: Springer, 2018: 3−19.

[32] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1800−1807.

[33] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 764−773.

[34] YANG B, BENDER G, LE Q V, et al. CondConv: conditionally parameterized convolutions for efficient inference

[EB/OL]. (2019−04−10)[2024−10−12]. https://arxiv.org/abs/1904.04971.

[35] LUU M L, HUANG Zeyi, XING E P, et al. Expeditious sali ency-guided mix-up through random gradient Threshold ing[EB/OL]. (2022−12−09)[2024−10−12]. https://arxiv.org/abs/2212.04875.

[36] 郭玉荣, 张珂, 王新胜, 等. 端到端双通道特征重标定 DenseNet 图像分类[J]. 中国图象图形学报, 2020, 25(3): 486−497.
GUO Yurong, ZHANG Ke, WANG Xinsheng, et al. Image classification method based on end-to-end dual feature reweight DenseNet[J]. Journal of image and graphics, 2020, 25(3): 486−497.

[37] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015−04−10)[2024−10−12]. https://arxiv.org/abs/1409.1556.

[38] HASSANI A, WALTON S, SHAH N, et al. Escaping the big data paradigm with compact Transformers[EB/OL]. (2022−06−07)[2024−10−12]. https://arxiv.org/abs/2104.05704.

[39] ZHOU C L, ZHANG H, ZHOU Z K, et al. QKFormer: hierarchical spiking Transformer using Q-K attention [EB/OL]. (2024−03−25)[2024−10−08]. https://arxiv.org/abs/2403.16552.

[40] CHOROMANSKI K, LIKHOSHERSTOV V, DOHAN D, et al. Rethinking attention with performers[EB/OL]. (2020−09−30)[2024−01−05]. https://arxiv.org/pdf/2009.14794.pdf.

[41] 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. Waikoloa: IEEE, 2023: 6464−6473.

[42] 谢奕涛, 苏鹭梅, 杨帆, 等. 面向目标类别分类的无数据知识蒸馏方法[J]. 中国图象图形学报, 2024, 29(11): 3401−3416.
XIE Yitao, SU Lumei, YANG Fan, et al. Data-free knowledge distillation for target class classification[J]. Journal of Image and Graphics, 2024, 29(11): 3401−3416.

[43] 柴智, 丁春涛, 郭慧, 等. CN2Conv: 面向物联网设备的强鲁棒 CNN 设计方法[J]. 计算机应用研究, 2025, 42(7): 2154−2160.
CHAI Zhi, DING Chuntao, GUO Hui, et al. Combined non-linearity convolution kernel generation: strong robust CNN design method based on IoT[J]. Application research of computers, 2025, 42(7): 2154−2160.

[44] 宫智宇, 王士同. 面向重尾噪声图像分类的残差网络学习方法[J/OL]. 计算机应用. [2025−10−02]. https://doi.org/10.11772/j.issn.1001-9081.2024101407.
GONG Zhiyu, WANG Shitong. Residual network learning method for image classification under heavy-tail noise [J/OL]. Computer applications. [2025−10−02]. https://doi.org/10.11772/j.issn.1001-9081.2024101407.

[45] 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. Multi-scale and multi-directional Transformer-based image recognition[J]. Chinese journal of computers, 2025, 48(2): 249−265.

[46] 朱秋慧, 杨靖, 黄若愚, 等. 基于部分卷积的多尺度特征卷积神经网络模型[J/OL]. 无线电通信技术. [2025−05−21]. http://kns.cnki.net/kcms/detail/13.1099.TN.20250310.1707.012.html.
Zhu Qiuhui, Yang Jing, Huang Ruoyu, et al. Partial convolution-based multi-scale feature convolutional neural network model[J/OL]. Radio communications technology. [2025−05−21]. http://kns.cnki.net/kcms/detail/13.1099.TN.20250310.1707.012.html.

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。主持国防预研基金项目、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 余篇。E-mail：lntuwulue@163.com。

王鑫杰，硕士研究生，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：2585178999@qq.com。

张晟翀，硕士研究生，高级工程师，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。