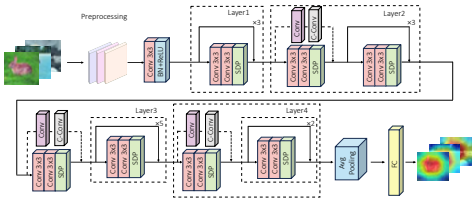
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 6 月, 第 53 卷, 第 6 期

空间动态感知的图像分类网络



姜文涛[1], 纪佳莹[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: 为了增强对重点特征的动态感知能力, 增强空间和通道间的交互作用, 提出空间动态感知的图像分类网络 (spatial dynamic perception image classification network, SDPNet)。此网络以 ResNet-34 残差网络为基础, 首先引入负值指数抑制模块 (negative exponent suppression, NES) 强化对图像空间特征的感知, 增强网络对判别性特征的学习能力; 然后采用了空间动态感知模块 (spatial dynamic perception, SDP), 将负值指数抑制模块与并行结构的特征旋转 (feature rotation, FR) 与动态扩张 (dynamic expansion, DE) 相融合, 捕捉图像空间对称性, 动态选择空间特征中的判别性信息, 实现空间特征增强。最后引入了双维度融合残差模块 (dual-dimension fusion residual, DFR), 将 SDP 模块嵌入到残差块中并将条件卷积嵌入到降维后的残差分支中, 动态调整空间和通道的交互作用, 优化特征表达, 减少网络过拟合, 提高网络的泛化能力。在 CIFAR-100、CIFAR-10、SVHN、Flowers-102、NWPU-RESISC45 数据集上准确率分别达到了 80.63%、96.95%、97.66%、90.34%、92.46%, 与当前先进的方法相比分别平均提高了 3.60%、2.52%、2.09%、7.29%、2.36%, 与现有主流网络模型比较, SDPNet 能够对原始特征图中重要的空间位置和通道位置进行增强, 提高关键特征的贡献度, 有效提高神经网络的图像分类能力。

关键词: 图像分类; 特征选择性抑制; 特征旋转; 动态扩张; 残差网络

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

DOI: 10.12086/oee.2026.250376 | CSTR: 32245.14.oee.2026.250376 | 中图分类号: TP391 | 文献标识码: A

姜文涛, 纪佳莹, 张晟翀. 空间动态感知的图像分类网络 [J]. 光电工程, 2026, **53**(6): 250376

Jiang W T, Ji J Y, Zhang S C. Spatial dynamic perception image classification network[J]. *Opto-Electron Eng*, 2026, **53**(6): 250376

1 引 言

图像分类是计算机视觉领域的核心研究方向之一, 旨在通过算法模型自动识别并标注输入图像的类别[1]。作为人工智能技术的重要应用基础, 该任务的核心挑战在于如何从像素级数据中提取特征, 并建立鲁棒的分类决策边界。传统方法主要依赖于手工设计特征[2]与浅层分类器[3]的结合, 其特征表达能力有限, 难以适应复杂的视觉场景变化, 随着数据量的增加和任务复杂度的提升, 传统方法难以满足日益增长的图像分类任务需求, 因此激发了后续深度学习方法的蓬勃发展[4]。在深度学习[5]模型领域应用最为广泛的是卷积神经网络[6], 卷积神经网络能够从原始图像中自动提取特征, 并且通过多层卷积和池化操作[7], 自动学习图像的低级特征和高级特征, 从而减少传统方法中对于手动特征设计的依赖。

随着卷积神经网络技术的兴起, 图像分类任务迎来了革命性的突破。2012 年, AlexNet[8]首次证明了深度卷积神经网络的优越性。该模型采用 5 层卷积和 3 层全连接结构, 有效缓解了梯度消失问题, 但由于其网络结构相对简单, 所以特征提取能力十分有限。2014 年, 牛

收稿日期: 2025-12-04; 修回日期: 2026-02-22; 录用日期: 2026-02-28; 网络出版日期: 2026-06-25
基金项目: 国家自然科学基金 (61601213); 辽宁省自然科学基金 (20170540426); 辽宁省教育厅重点基金 (LJ2017QL034)
*通信作者: 纪佳莹, 826796020@qq.com
版权所有©2026 中国科学院光电技术研究所

津大学提出的 VGGNet[9] 系统研究了网络深度的影响，其采用连续的 3×3 小卷积核堆叠，相比大卷积核既能减少参数量，又能通过多层非线性变换增强特征表达能力。同年，GoogleNet[10] 提出革命性的 Inception模块，该结构采用多分支并行设计，通过不同感受野捕捉多尺度特征，但多分支设计显著增加了内存访问开销，辅助分类器的梯度回传效果随网络加深而衰减。He 等[11] 提出 ResNet，通过残差学习解决了深层网络梯度消失难题。其核心的跳跃连接允许原始输入直接跨层传输，使梯度能够畅通无阻地反向传播[12]，但残差连接对噪声和对抗样本的放大效应导致恒等映射在深层网络中的特征被稀释，瓶颈设计中 1×1 卷积占比过高，需要更多的设计资源。

为提升计算效率，解决传统神经网络不能进行动态权重分配、可解释性与灵活适配性不高等问题，研究人员提出注意力机制与网络结构相结合的方法来重塑视觉领域。SENet[13] 创新性地在通道维度引入注意力机制，通过全局平均池化和全连接层生成通道权重，实现特征通道的自适应校准。CBAM[14] 在 SENet 的基础上实现了对通道和空间上双注意力协同工作，但它的空间注意力仅基于浅层池化特征，对复杂关系建模不足。Non-Local Networks[15] 通过引入自注意力，计算所有像素点的相似度权重来实现全局上下文建模，然而其计算复杂度随分辨率平方增长，难以处理高分辨率图像。Vision Transformer (ViT)[16] 突破性地将图像拆分为 16 pixel×16 pixel 的图块序列，通过多头自注意力机制实现全局建模，成为视觉 Transformer 的里程碑工作，但由于串口划分破坏平移不变性，位置代码需预设最大分辨率[17]。

现有注意力机制在全局性、灵活性和可解释性上均展现出多方面的优势，但是由于模型结构的限制，在处理复杂信息时总是过度强调全局关系，牺牲了对局部细微特征和纹理的感知[18]，难以适应不同尺度的物体和空间关系的变化，不能充分整合全局空间上下文信息，由于模型对全局信息的保留不够完整，使得分类准确性受到影响，使模型有过拟合的风险。

为了解决以上问题，本文提出空间动态感知的图像分类网络 SDPNet。该网络以 ResNet-34 为基础，首先提出了负值指数抑制模块 (negative exponent suppression, NES) 强化对图像空间特征的感知，其非线性特征变换方法可以增强网络对判别性特征的学习能力。然后提出了 SDP 注意力机制，通过负值指数抑制模块结合并行结构的 FR 与 DE 捕捉图像空间对称性，提升旋转鲁棒性，动态聚焦空间特征中的判别性信息，增强对空间特征的感知能力。最后提出了 DFR，该模块由融合 SDP 注意力机制的残差块和嵌入条件卷积到降维后的残差分支构成，增强空间和通道的交互作用，优化特征表达。

2 空间动态感知的图像分类网络

为了增强对重点特征的动态感知能力，增强空间和通道间的交互作用，提升模型对物体方向变换的鲁棒性，增强网络的泛化能力，本文提出了空间动态感知的图像分类网络 (SDPNet)，总体结构如图 1 所示。

本研究的网络主要包括以下几个部分：

1) 对输入图像进行预处理操作。进行预处理操作后，将尺寸归一化，满足模型输入层的张量形状要求，避免因尺寸不匹配导致运行时出现错误，除了确保输入维度匹配外，在图像的预处理操作中还能将像素标准化，像素标准化处理过的图像可以加速收敛，兼容预训练权重。

2) 将通过预处理后的图像输入网络的首层特征提取模块。将原有的 7×7 首层卷积改为 3×3 首层卷积可以更好地关注局部特征，能够保留更多的细节。

3) 图像通过首层卷积后，再将图像传输到由双维度融合残差模块构成的 Layer1、Layer2、Layer3、Layer4 中，通过 SDP 动态选择空间特征中的判别性信息，强化对图像空间特征的感知，实现空间特征增强。在降维后的残差分支中加入条件卷积，增强空间和通道的交互作用，提高网络的泛化能力。

4) 将特征图传输到平均池化层中，目的是将每个特征图的全局空间信息压缩为单一标量，减少模型复杂度，忽略物体在图像中的具体位置，只关注是否存在某类特征。再通过全连接层将高级语义特征转换为类别标签，让模型自动学习从原始数据到目标的完整映射，输出最终的分类结论。

2.1 首层特征提取模块

ResNet-34 网络采用 7×7 卷积配合最大池化操作来完成初级特征的提取。虽然 7×7 卷积核的初始感受野较大，能够较快捕捉更多的特征，但是大卷积核会过度模糊局部细节。以 10 pixel×10 pixel 尺寸输入图像为例，经过 ResNet-34 网络初始特征提取层处理后，输出特征图的尺寸可根据式 (1) 计算：

$$T=\frac{S-K+2P}{D}+1\,, \tag{1}$$

其中：S 为输入特征图像尺寸；K 为卷积核尺寸；P 为填充大小；T 为输出特征图像尺寸；D 为步长。

根据式 (1) 可以得出输出特征图尺寸减小为 3 pixel×3 pixel，经过卷积后的图像尺寸变小，物体轮廓变模糊，空间关系失真。

为了解决这一问题，本研究将首层的 7×7 卷积核替换为 3×3 的小卷积核，首层卷积核尺寸对比如图 2 所示。

小卷积核能够改善模型的计算效率，还能增强局部特征的提取能力，避免过大感受野引入的无关背景干扰。同时删除最大池化层，最大限度保留图像的原始信息，设置合理的填充大小来保持输出图像与输入图像尺寸一致。

综上所述，本研究提出的首层特征提取模块是通过采用较小尺寸的卷积核实现的，对于小尺寸输入图像，该方法不仅能够有效提取底层视觉特征，还能改善模型的计算效率，提升分类结果的准确率。

2.2 负值指数抑制模块

为了强化对图像空间特征的感知能力，有效过滤噪声和冗余信息，使网络能够学习更具区分度的特征表示，增强特征的判别性。本研究提出了负值指数抑制模块 (negative exponent suppression，NES)，具体结构图如图 3 所示。

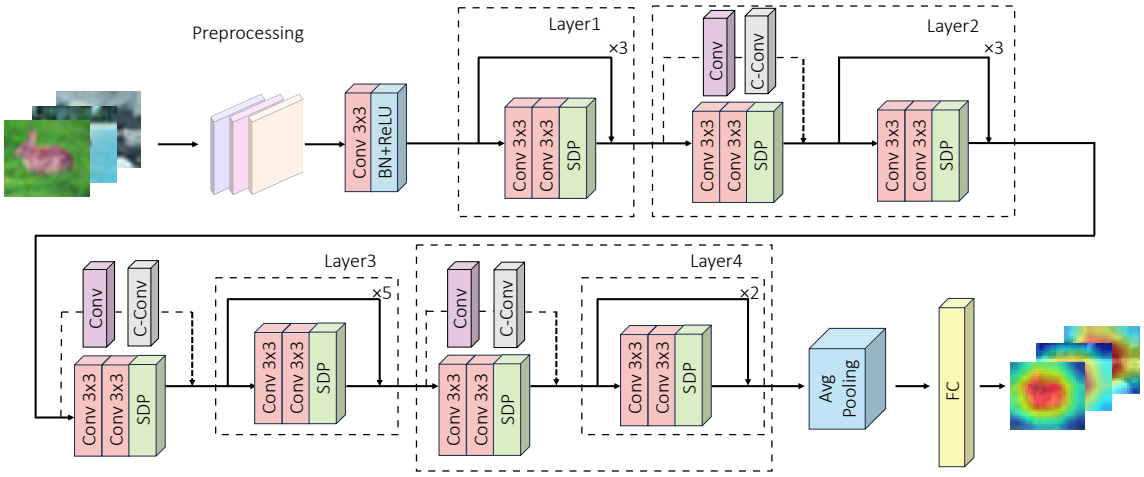


图 1 SDPNet 总体结构图

Fig. 1 Overall structure of SDPNet

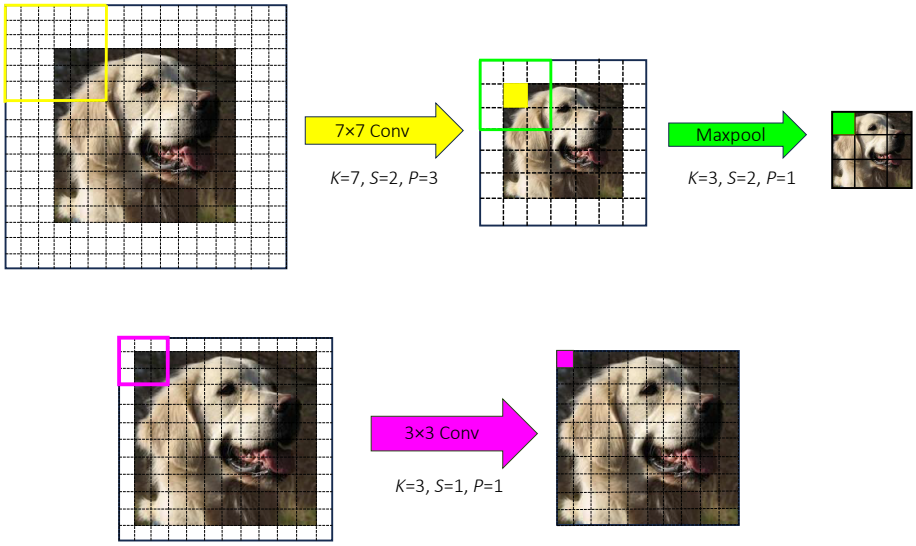


图 2 网络首层卷积修改前后对比图

Fig. 2 Comparison chart of network first-layer convolution layer before and after modification

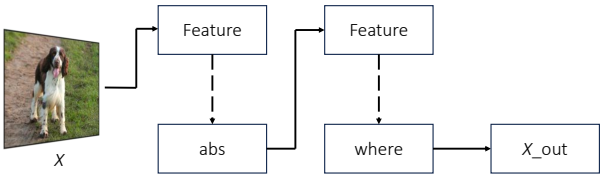


图 3 负值抑制结构图

Fig. 3 Negative value suppression structure diagram

负值指数抑制部分是一种新型非线性特征变换方法，由绝对值变换操作 (abs) 和指数衰减抑制操作 (where) 组成，对于给定的输入特征图$X \in \mathbb{R}^{H\times W\times C}$，其中，$H\times W$ 为特征图的空间大小；C 为通道数，特征图将先经过一个绝对值变换层，具体可以表示为

$$X_\mathrm{abs} = |X|, \tag{2}$$

其中：$X_$abs 表示对特征图进行绝对值变换。绝对值变换在保持特征连续性的同时，可以去除正负符号带来的极性差异，增强特征对比度，突出边缘信息。

然后进行指数衰减抑制操作，对原始特征中的负值进行特殊处理，用其绝对值的指数衰减值替代原始特征中的负值，再通过指数衰减因子使得负值的贡献快速衰减，形成非线性响应，具体可以表示为

$$X_\mathrm{out} = where(X > 0, X, X_\mathrm{abs} \odot \mathrm{e}^{-6}), \tag{3}$$

其中：$X_$out 为输出的变换后特征图；⊙表示逐元素乘法；e^{-6} 为衰减因子，衰减因子将负响应强度压缩至原始值的 0.25% ($\mathrm{e}^{-6}\approx0.0025$)。

为进一步观察 NES 模块的效果，图 4 给出了 NES 模块使用前后特征可视化对比图，由图 4 可知，该模块通过负值指数抑制技术，以指数衰减方式弱化负响应所携带的边缘信息，并通过正向激活区域的响应强度提升特征图的整体对比度，使关键目标的轮廓和细节更加突出，在保留信息与增强判别性特征中取得平衡，提高了特征的对比度，使网络聚焦于更具有鉴别力的区域，进一步加强了网络对图像特征的学习能力。

2.3 空间动态感知模块

传统的卷积神经网络因其静态且局部的特性，在处理复杂空间结构、几何变换、多尺度目标等任务时存在瓶颈。为了增强重点特征的动态感知能力，强化对图像空间特征感知的同时提升模型对几何变换的鲁棒性，实现空间特征增强，本文提出了 SDP 模块，将负值指数抑制模块与并行结构的特征旋转、动态扩张以及通道混洗



图 4 使用 NES 模块前后特征图像对比

Fig. 4 Comparison of feature images before and after using the NES module

部分方法相融合，动态选择空间特征中的判别性信息，提高关键区域敏感性，从而提升网络性能。

特征图在经过负值指数抑制模块处理后，对特征图进行 FR、DE、通道调整 (channel shuffle, CS) 三部分并行处理。

1) FR：具体结构如图 5 所示。由图可知，先通过对特征图进行 180°中心对称旋转，在特征空间实现了隐式数据增强，有效缓解了标准卷积核在旋转变化下捕捉特征的局限性。然后将旋转后的特征图经由 1×1 深度可分离卷积处理，具体如式 (4) 所示：

$$X_{\mathrm{rot}} = \tanh(W_2 \cdot rot(W_1 \cdot X)), \tag{4}$$

其中：W_1 为第一个 1×1 卷积；rot 为旋转操作；W_2 为第二个 1×1 卷积；tanh 为双曲正切激活函数；$X_$rot 为最终的输出。

卷积从旋转后的视角提取特征，有效提升了特征空间的多样性分布，参数量较标准卷积大幅减少，在提高模型对物体方向变化的鲁棒性的同时实现图像分类任务中的几何不变性增强与动态特征校准。

最后使用 tanh 激活函数对卷积后的特征图进行处理，tanh 激活函数公式如下：

$$\tanh(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}}. \tag{5}$$

由式 (5) 可知，通过 tanh 激活函数可以将任意输入压缩到 [−1,1]，有利于旋转特征的梯度回传。同时，tanh 的非线性特性会在抑制微弱特征的同时突出较大值的显著特征，对比度增强有助于网络聚焦于更重要的特征。

综上，tanh 在此处的设计既稳定了数值，又通过非线性变换优化了特征表达，最终使后续特征图融合更有效，确保模型在处理旋转等几何变换时网络训练的稳定性。

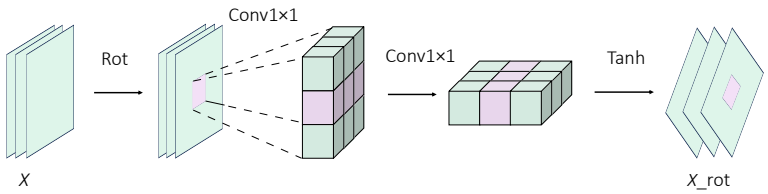


图 5 FR 部分结构图

Fig. 5 FR structure diagram

2) DE：由两步操作组成，首先是特征提取操作，其具体结构如图 6 所示。

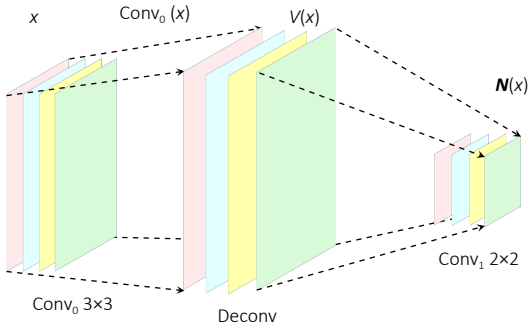


图 6 特征提取操作结构图

Fig. 6 Feature extraction operation structure diagram

首先对特征图采用卷积核大小为 3×3 的卷积，通过降低空间分辨率来扩大感受野，同时在卷积过程中使用反卷积来恢复特征图空间信息，其核心公式如下：

$$\boldsymbol{V}(x)=Deconv(Conv_0(x)), \tag{6}$$

其中：$Conv_0$ 表示卷积核为 3×3 的卷积；$Conv\ (x)$ 表示经过卷积后得到的特征矩阵；$Deconv\ (x)$ 表示反卷积；$\boldsymbol{V}\ (x)$ 表示经过反卷积后的特征矩阵。反卷积的具体工作原理为：先在输入特征图的每个像素之间中插入零值，将插入的零值作为占位符，确保原始像素在输出中的位置关系不变，从而达到恢复空间信息的目的。

特征图通过反卷积操作后，再使用步长为 1，卷积核大小为 2×2 的卷积对特征图进行局部特征提取，以逐像素级滑动扫描实现密集覆盖，捕捉输入图像的局部特征，从而增强网络对细节信息的感知能力。其表达公式为

$$\boldsymbol{N}(x)=Conv1(\boldsymbol{V}(x)), \tag{7}$$

其中：$Conv1$ 表示卷积核为 2×2 的卷积；$\boldsymbol{N}\ (x)$ 表示经过卷积后的特征矩阵。

进行特征提取部分操作后，为了确保特征图的尺寸与原始输入一致，使得后续逐元素加法顺利进行，本文对特征图像实行最近邻插值操作强制调整特征图尺寸，最邻近插值操作图像如图 7 所示。

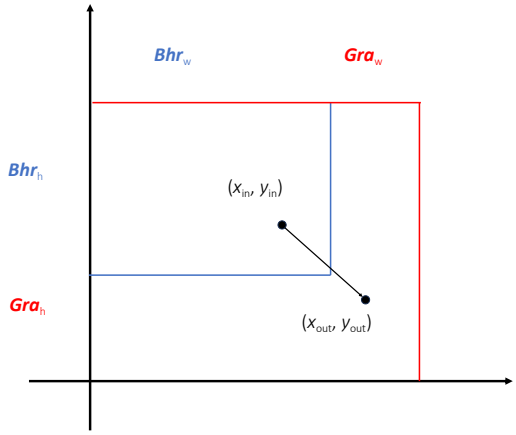


图 7 最邻近插值操作图像

Fig. 7 Nearest neighbor interpolation operation image

图中 ***Bhr*** 表示原特征矩阵；***Gra*** 表示经过最近邻插值上采样算法后的目标矩阵；$\boldsymbol{Bhr}_{\mathrm{h}}$、$\boldsymbol{Bhr}_{\mathrm{w}}$ 表示原特征矩阵的尺寸参数，分别为高度宽度信息；$\boldsymbol{Gra}_{\mathrm{h}}$、$\boldsymbol{Gra}_{\mathrm{w}}$ 表示目标特征矩阵的参数，分别为高度和宽度信息，$(x_{\mathrm{in}}, y_{\mathrm{in}})$ 为原特征矩阵中某个元素坐标；$(x_{\mathrm{out}}, y_{\mathrm{out}})$ 为经过最邻近插值上采样操作后的同一个元素坐标，由图可见，原坐标随着上采样操作和特征矩阵被等比放大，⌊·⌋表示向下取整，其计算坐标映射关系为

$$y_{\mathrm{out}}=\left\lfloor y_{\mathrm{in}}\cdot\frac{\boldsymbol{Bhr}_{\mathrm{h}}}{\boldsymbol{Gra}_{\mathrm{h}}}\right\rfloor, x_{\mathrm{out}}=\left\lfloor x_{\mathrm{in}}\cdot\frac{\boldsymbol{Bhr}_{\mathrm{w}}}{\boldsymbol{Gra}_{\mathrm{w}}}\right\rfloor. \tag{8}$$

由式 (8) 可知，最近邻插值上采样操作的核心原理就是，对于输出中的每个像素，找到输入中几何位置最近的像素，直接复制其值。因此最邻近插值上采样操作不引入新颜色或渐变，无复杂计算，因其仅依赖原始数据，所以内存占用低，无随机性，能够保持稳定输出。

3) CS：由两个 1×1 卷积和一个通道打乱操作组成，具体如图 8 所示。

首先，将经过前序处理的特征图输入到一个 1×1 卷积层中进行通道维度的变换，1×1 卷积可以促进通道之间的信息融合，改变通道维度；接着，对卷积后的特征图执行通道混洗操作，通过随机化通道顺序，强制不同通道组之间进行信息交互，通道之间的信息交互可以打破通道间的固定连接模式，增强特征多样性；最后，混洗后的特征再通过第二个 1×1 卷积层处理，将通道数恢复到原始输入尺寸，并基于重组后的通道关系学习更有效的特征表示。

整个流程通过三步操作，实现了高效的通道注意力建模，既能捕捉通道间的重要依赖关系，又保持了较低的计算开销，从而实现通道交互和重组，网络动态抑制冗余通道、增强有用通道的目标。

SDP 模块由上述四部分组成，其具体结构图如图 9 所示。

首先对特征图进行 NES 操作，强化对图像空间特征的感知，突出关键特征，然后对经过 NES 操作的特征图并行执行三步操作，通过 FR 操作，捕捉图像空间的对称性，增强模型对物体方向变化的鲁棒性；通过 DE 操作，同时捕捉细节和整体轮廓，动态选择空间特征中的判别性信息；FR 操作提供几何不变性特征，DE 操作保留多尺度空间信息，二者通过相加融合，提升特征表达能力，再对相加后的特征进行空间平均池化操作，实现空间特征增强。在通道上，通过 CS 操作打破固定通道关系，配合后续卷积挖掘跨通道交互，抑制冗余通道；将经过 CS 操作后的特征图送入通道平均池化层中，聚焦特征重要性；将空间平均池化后的特征图与通道平均池化后的特征图相加，再通过 tanh 激活函数将注意力归一化，增加特征矩阵之间的关联性，实现对输入特征进行自适应增强，从而提升模型对关键特征的敏感性。最后将处理结果与原始特征图像相乘，防止原数据丢失，并输出最终结果。

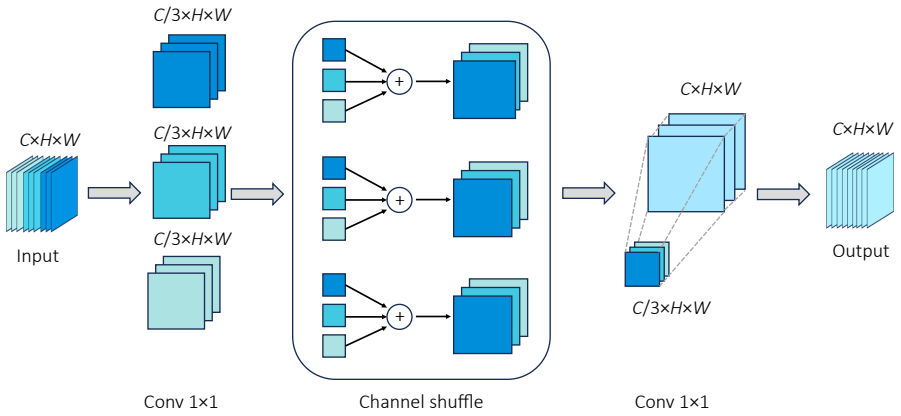


图 8 CS 部分结构图

Fig. 8 CS structure diagram

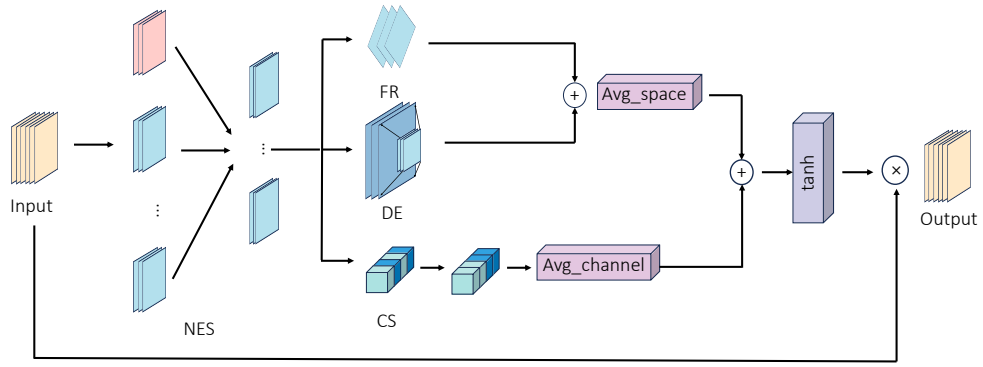


图 9 SDP 模块结构图

Fig. 9 SDP structure diagram

2.4 双维度融合残差模块

在图像分类任务中，随着传统卷积神经网络层数的不断增加，模型容易出现不稳定现象，如梯度爆炸和梯度消失等问题，这提升了训练过程的复杂度，更直接导致了分类错误率的升高，为了解决这一问题，He 等[11]创新性地提出了基于跳跃残差连接的网络结构，如图 10(a) 所示。通过建立跨层连接来加深网络，缓解梯度爆炸和梯度消失等问题。当输入与输出特征图的维度不匹配时，采用 1×1 卷积核的跳跃连接实现维度对齐，如图 10(b) 所示。但是降维后的残差模块通过固定的卷积核权重处理输入特征，其输出的特征图的所有空间特征和通道特征具有均等的贡献度，这种无差别的特征融合方式忽视了不同区域或通道对任务目标的差异性，导致模型可能无法有效抑制噪声或冗余特征，从而降低特征判别性。为解决这一问题，通过将 SDP 注意力机制融入残差模块中，如图 10(c) 所示，在通道和空间上对注意力权重进行分配，增强对重点特征的关注度，强化对图像空间和通道特征的感知，并将条件卷积嵌入到降维后的残差分支中构成 DFR 模块，如图 10(d) 所示。

条件卷积的公式为

$$y=\sum_{i=1}^{N}\alpha_i(x)\cdot(W_i*x)+b, \tag{9}$$

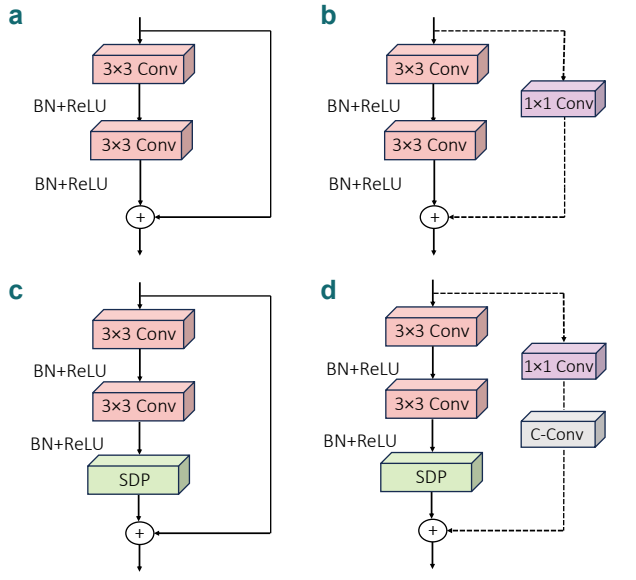


图 10 4 种残差块比较。(a) 残差块；(b) 降维后的残差块；(c) 融合 SDP 的残差块；(d) 双维度融合残差模块

Fig. 10 Four types of residual blocks. (a) Basic block; (b) Residual block; (c) Residual block fused with SDP; (d) DFR

其中：W_i 为卷积核权重；b 为偏置项；$\alpha_i(x)$ 为由路由函数生成的第 i 个权重。路由函数是一个由输入特征图驱动的、动态的卷积核混合器，使卷积操作由静态变为动态，实现条件化。

DFR 能够调整感受野，有效增强模型对空间特征的提取能力，动态调整通道交互。在空间维度上，DFR 通过多尺度特征融合策略，将不同尺度的空间特征进行有效整合，显著提升了模型对复杂空间结构的表征能力。在通道维度方面，DFR 引入了动态通道交互机制，通过自适应权重分配强化重要特征通道的贡献，在缓解信息丢失的同时有效去除冗余信息，增强特征图的图像信息与空间信息的平衡，以适应不同尺度的目标，进一步优化特征表达，使特征表达在保持丰富图像细节的同时，实现了信息的高效压缩，从而提升模型鲁棒性。

3 实验及结果分析

3.1 实验环境

3.1.1 硬件、软件平台以及数据集

实验采用的硬件与软件平台配置如下：操作系统为 Windows11，显卡为 NVIDIA RTX 3090 显卡 (24 GB 显存)，软件框架包括 Pytorch 1.9.7、CUDA 12.1 与 CUDNN8.3。编程语言为 Python 3.9.13。所使用的数据集包括 CIFAR-10、CIFAR-100、SVHN、Flowers-102 及 NWPU-RESISC45，以覆盖多种图像分类场景。具体数据集信息如表 1 所示。

表 1 实验数据集

Table 1 Experimental datasets

Name	Size	Num. of classes	Number of training samples	Num. of test samples
CIFAR-10	32×32	10	50000	10000
CIFAR-100	32×32	100	50000	10000
SVHN	32×32	10	73257	26032
Flowers-102	96×96	102	14740	15153
NWPU-RESISC45	256×256	45	27000	4500

3.1.2 参数初始化与训练准备

网络参数的初始化对模型训练的稳定性至关重要，为确保实验的可复现性和模型训练的稳定性，使所有对比模型在公平一致的条件下进行训练，本研究对所有网络参数采用以下初始化策略。

网络主体卷积层采用 Kaiming 正态初始化，该方法是适配 ReLU 激活函数的经典初始化策略，保持在前向传播过程中激活值的分布稳定，有利于梯度在反向传播中有效流动；所有残差块末尾的批量归一化层进行零初始化，以稳定深层网络训练；SDP 模块中的相应卷积层采用了 Xavier 初始化以适配 tanh 函数，避免因初始化不当导致梯度饱和或消失问题，使 SDP 模块从训练初期就能正常参与学习。

3.2 SDPNet 整体性能分析

为全面评估 SDPNet 的有效性，在 CIFAR-100、CIFAR-10、SVHN 数据集上，从分类准确率与模型效率两个维度，将其与多种主流模型进行整体对比。

3.2.1 SDPNet 分类准确率

为评估 SDPNet 的有效性，本研究将 SDPNet 与多种主流图像分类网络模型 ResNet-34[11]、EfficientNet[19]、GhostNet[20]、MobileNet-LAM[21]、QKFormer[22]、TLENet[23]、ATONet[24]、Couplformer[25]、SSLLNet[26]、Multi-ResNet[27]、DCDENet[28] 在 CIFAR-100、CIFAR-10、SVHN 数据集上进行了对比实验。实验所采用的分类准确率数据均严格参照以下来源：对于未公开代码的模型，直接引用其原始论文中报告的实验结果；开源模型的结果则通过官方代码复现验证。各网络模型在三个数据集上的分类准确率如表 2 所示。

表 2　不同网络在 3 种数据集上的分类准确率/%

Table 2　The classification accuracy rates of different networks on three datasets/%

Method	CIFAR-100	CIFAR-10	SVHN
ResNet-34[11]	78.14	95.80	96.82
EfficientNets[19]	75.96	94.01	93.32
GhostNet[20]	77.15	94.92	93.86
MobileNet-LAM[21]	68.09	89.37	—
QKFormer[22]	80.26	96.18	97.13
TLENet[23]	78.42	95.46	96.83
ATONet[24]	78.47	93.83	94.61
Couplformer[25]	73.92	93.54	94.26
SSLLNet[26]	79.32	95.79	97.04
Multi-ResNet[27]	78.68	94.65	—
DCDENet[28]	80.08	96.56	97.55
SDPNet (Ours)	**80.63**	**96.95**	**97.66**

“—”表示未知结果，加粗字体为最优实验。综合表 2 中各模型的分类准确率可知，SDPNet(本文网络) 在多个数据集上的分类性能均优于其他 11 种网络模型。其中 SDPNet 的最高准确率分别为 80.63%、96.95%、97.66%，对比其他网络分类准确率平均提升了 3.60%、2.52%、2.09%。

性能仅次于 SDPNet 的网络是 DCDENet 和 QKFormer，准确率分别为 80.08%、96.56%、97.66% 与 80.26%、96.18%、97.13%。DCDENet 同 SDPNet 一样，都是在 ResNet 残差网络的基础上加以改进，DCDENet 采用双分支多注意力机制模块，抑制了非关键特征的干扰，提升了模型对目标区域的定位效率。QKFormer 的核心贡献在于提出了 Q-K 注意力机制，该机制通过线性复杂度的二进制向量编码策略，实现了通道维度的高效特征建模，使其具备优异的关键信息筛选能力。但是上述网络仅仅只是在捕捉重要信息方面有所提升，在局部特征捕捉方面不够完善，因此分类效果低于 SDPNet。

EfficientNets 在 CIFAR100、CIFAR-10、SVHN 三个数据集上的准确率分别为 75.96%、94.01%、93.32%，它通过联合调整深度、宽度和分辨率，而非传统单一维度的缩放，实现较优的性能与计算量的平衡，但是复合缩放依赖高分辨率输入，在小尺寸图像上分辨率缩放策略则会失效，分类准确率受到影响。

GhostnNet 和 Couplformer 都是通过结构重构实现高效计算，尽管在某些数据集上较 MobileNet-LAM 有所提升，但是轻量化设计被迫牺牲对细微判别特征的捕捉能力，导致网络泛化能力变差。

其他网络虽然在上述三个数据集中性能表现尚可，但是深入比较后发现，对于动态特征校准以及几何变换后模型鲁棒性降低的问题，SDPNet 效果最佳，因此，SDPNet 在增强局部特征提取能力的同时，保证了网络的泛化性，能够准确聚焦关键区域和特征，提升分类准确率。

3.2.2 SDPNet 参数量、计算量及运行速度

为明确 SDPNet 在实际应用中的系统可用性和可维护性，对各网络的参数量、计算量以及运行速度进行对比分析，具体结果如表 3 所示。

由表 3 可知，与上述网络相比，SDPNet 的参数量与 ResNet-34 的参数量相近，仅相差 0.58 M，与 EfficientNets 的 53.68 M、Multi-ResNet 的 52.03 M 以及 HO-ResNet[29] 的 51.22 M 相比，SDPNet 的参数量最低，这是因为 SDPNet 未采用大核卷积，降低了网络的参数量，引入了动态计算，使模型更轻量高效。

SDPNet 的计算量为 7.50 G，高于 SSLLNet 的 6.36 G、

表 3 各网络的参数量、计算量及运行速度

Table 3 The parameter count, FLOPs and FPS of each network

Method	Params/M	FLOGs/G	FPS
ResNet-34[11]	21.35	3.67	3.01
EfficientNets[19]	53.68	4.79	—
Couplformer[25]	27.74	4.47	2.83
Multi-ResNet[27]	52.03	11.68	1.60
TLENet[23]	47.03	9.42	2.28
ATONet[24]	29.97	9.83	1.81
SSLLNet[26]	31.62	6.36	2.12
HO-ResNet[29]	51.22	10.98	1.93
SDPNet (Ours)	21.93	7.50	2.55

EfficientNets 的 4.79 G 以及 Couplformer 的 4.47 G。虽不是最低值，但与 Multi-ResNet 的 11.68 G 相比，仍处于较低水平，表明 SDPNet 通过接受适度的计算开销来增加模型性能的提升，体现了更优的计算效率。

SDPNet 的运行速度为 2.55 f/s，优于 Multi-ResNet 的 1.60 f/s 和 TLENet 的 2.28 f/s，在上述网络中处于中等水平。尽管速度并非最快，但 SDPNet 的速度仍处于可比范围，满足实际部署所需的推理效率，在精度提升的同时，并未导致实时性严重下降。

综上所述，SDPNet 在参数量、计算量以及运行速度方面均表现较为优异，较低的参数量和较快的运行速度，为其在资源受限场景下的实际部署奠定了坚实基础，增强实际应用中的系统可用性和可维护性。

3.3 消融实验

为了评估 SDP 模块和 DFR 模块对 SDPNet 性能的贡献，本研究基于 ResNet-34 设计了消融实验，通过单独及组合加入这两种模块，并与原始 ResNet-34 进行对比分析。为确保实验可比性，在五种不同规模的数据集上开展验证：其中 CIFAR-100、CIFAR-10、SVHN 为小尺寸图像数据集，NWPU-RESISC45 为大尺寸图像数据集，Flowers-102 则包含多种尺寸的自然图像。为统一输入尺度，本研究将 Flowers-102 数据集的图像在预处理阶段统一裁剪为 96 pixel×96 pixel，鉴于 Flowers-102 和 NWPU-RESISC45 数据集尺寸大于 CIFAR-100、CIFAR-10、SVHN 三种数据集尺寸，为适应大尺寸图像的特征提取需求，我们将 ResNet-34 的初始特征提取层恢复为 7×7 卷积核结构，并增加最大池化层。对应的消融实验结果如表 4 所示。

由表 4 的消融实验结果可知，在多个数据集上，引入 SDP 与 DFR 模块均对 ResNet-34 的分类性能产生了显著提升。具体而言，ResNet-34+SDP 在 CIFAR-100、CIFAR-10、SVHN、Flowers-102 和 NWPU-RESISC45 数据集上的分类准确率分别提高了 1.72%、1.03%、1.61%、4.14% 和 1.99%；ResNet-34+DFR 分别提升了 1.40%、0.13%、0.12%、2.65% 和 1.66%；而同时融合 SDP 与 DFR 模块的完整模型 (SDPNet) 提升最为显著，准确率分别提高了 2.11%、0.53%、0.39%、7.59% 和 2.38%。SDPNet 相对于原模型提升幅度最大，尤其在类别数量较多的 CIFAR-100、Flowers-102 和 NWPU-RESISC45 数据集上表现更为突出。这表明模型能够从更丰富的类别中学习更具判别性的特征表示，从而带来更显著的性能增益。相比之下，CIFAR-10 和 SVHN 数据集虽然样本规模较大，但类别相对较少，本身基线准确率已较高，因此各类模型的提升幅度有限。

综上所述，无论数据集规模大小或类别数量多少，DFR 与 SDP 模块均能有效提升 ResNet-34 的分类准确率，并且两者组合构成的 SDPNet 在所有测试条件下均表现出最优的综合性能。

3.4 注意力机制对比实验

为验证 SDP 模块对 ResNet-34 的性能提升效果，本研究选取了五种主流注意力机制进行对比，包括 SE、SCConv (Li 等[30]，2023)、CA (Huo 等[31]，2021)、CBAM (Woo 等[14]，2018) 与 ADWM (Huang 等[32]，2025) 采用的注意力机制简介汇总如表 5，将 5 种注意力机制分别

表 4 引入不同模块的实验结果：分类准确率 (单位：%)

Table 4 The experimental results of introducing different modules: the classification accuracy rates (Unit: %)

Method	CIFAR-100	CIFAR-10	SVHN	Flowers-102	NWPU-RESISC45
ResNet-34	78.14	95.80	96.82	82.73	90.07
ResNet-34+SDP	79.86	96.83	97.23	86.87	92.06
ResNet-34+DFR	79.54	95.93	96.94	85.38	91.73
ResNet-34+SDP +DFR	**80.25**	**96.33**	**97.21**	**90.32**	**92.45**

嵌入到 ResNet-34 网络中，并在 CIFAR-100、Flowers-102、STL-10 三个数据集上进行训练，以对比分析其性能差异。

由于 Flowers-102 和 STL-10 的原始图像尺寸大于 CIFAR-100，针对大尺寸图像的特征提取需求，本研究将 ResNet-34 的首层特征提取层调整为 7×7 卷积核结构并补回最大池化层，目的是扩大感受野以增强特征捕获能力。实验结果如表 6 所示，其中，SDP 模块的性能最优，在上述三个数据集上的分类准确率高于其他 5 种注意力机制模块。

由实验结果可知，SE 模块通过引入通道注意力机制，使网络模型的准确率明显提升，通过增强通道的自适应选择能力，从而提升特征的判别性与模型的整体表达能力。但是其综合分类性能仍弱于本文提出的 SDP 模块，这是因为 SE 通过全连接层建模所有通道间的全局依赖关系，但并非所有通道两两之间均存在关联，这种跨通道的交互本质上是非线性且高阶的，因其复杂特性无法被简单的 SE 有效捕捉；引入 SCConv 模块的网络在 Flowers-102 数据集上取得了优异的分类性能，其准确率仅次于 SDP，通过动态空间注意力聚焦局部关键区域，抑制复杂背景干扰，再通过通道重建技术，增强判别性通道，有助于模型提升对细粒度分类的能力，SCConv 的双重建结构虽然轻量，但与普通卷积或简单注意力机制相比，仍会引入额外的计算开销；ADWM 通过特征双重自适应加权机制，有效抑制特征图中的噪声，使最终特征更具判别性，通过动态调整感受野，使网络自适应地决定依赖局部细节还是全局上下文，从而更好的处理不同大小的物体，但结合 ADWM 的残差块要比标准残差块更复杂，会增加计算量和参数，需要更精细的学习率调度和更长的训练时间。CA 与 CBAM 均融合了通道与空间注意力机制，但是相比于 SDP 在三个数据集上准确率分别低 1.18%、1.16%、2.26% 和 1.13%、2.48%、0.11%，这是因为以上两种注意力均为空间注意力和通道注意力协同引入机制，当注意力模块加入网络中时，空间注意力和通道注意力会形成复杂的相互耦合关系，当空间特征和通道特征的贡献度存在差异时，空间注意力和通道注意力会互相干扰从而导致性能下降。

与其他 5 种注意力机制相比，SDP 在三个数据集上分别平均提升 0.6%、1.27%、1.56%。说明本文提出的空间动态感知模块 SDP 能够在各种数据集上提供更精准的分类准确率。SDP 实现了动态调整通道重要性，增强了特征多样性，一方面增强跨通道的信息提取，另一方面融合多尺度上下文，最终实现跨通道与跨空间的特征交互与增强。

表 5　多种注意力机制简介

Table 5　Introduction to various attention mechanisms

注意力机制	简介
SE	SE是一种经典的通道注意力机制，通过动态学习各通道的重要性权重，增强关键特征并抑制冗余信息。
SCConv	SCConv是一种创新的双维度重建注意力机制，通过解耦空间和通道维度的特征优化，实现了高效的特征选择。
ADWM	ADWM是一种复合型注意力机制，通过不同维度的特征间自适应加权，有效抑制噪声干扰，并动态捕捉特征间的互补信息，提升模型的鲁棒性。
CA	CA也是一种双维度注意力机制，通过分解式坐标信息捕获，同时建立通道关系和长距离空间依赖，在不增加计算量的情况下，提升了模型对空间结构的建模能力。
CBAM	CBAM是一种经典的双维度注意力机制，通过串联通道和空间两个维度的注意力机制，提升CNN的特征选择能力。

表 6　引入不同注意力机制的实验结果：分类准确率 (单位：%)

Table 6　Experimental results of introducing different attention mechanisms: the classification accuracy rates (Unit: %)

Method	CIFAR-100	Flowers-102	STL-10
ResNet-34	78.14	82.73	69.88
+SE	79.26	86.43	70.62
+SCConv	79.45	87.60	72.09
+ADWM	79.19	87.04	73.42
+CA	78.68	86.77	71.60
+CBAM	78.73	85.45	73.75
+SDP	**79.86**	**87.93**	**73.86**

3.5 首层特征提取模块实验

为了对比 ResNet-34 网络第一层卷积核尺寸大小对实验结果的影响，本文设计了首层特征提取模块实验，首先将 ResNet-34 网络首层卷积由原始的 7×7 大小的卷积核改为 3×3 大小的卷积核，并删除最大池化层，再测试原始模型与改进后的模型在 CIFAR-100 数据集上的准确率，如图 11 所示。

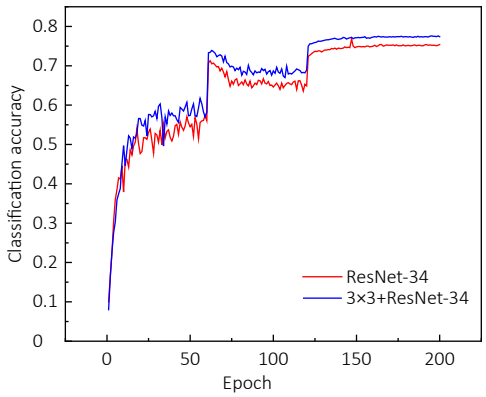


图 11 修改网络首层卷积对准确率的影响

Fig. 11 The impact of modifying the first layer convolution of the network on classification accuracy

由图 11 可知，改进后的模型准确率为 77.6%，原始模型的准确率为 75.4%，通过对比实验结果，发现较小尺寸卷积核能够增强局部特征的提取能力，避免过大感受野带来的无关上下文信息干扰，同时删除最大池化层能够最大程度保留更多的图像原始信息，从而提升分类结果的准确率。

3.6 SDP 模块实验

3.6.1 SDP 不同模块实验

SDP 模块中的影响实验结果准确率的参数为衰减因子和特征图的旋转角度。以下将分别对这两个参数进行参数影响实验。

1) 为了探究衰减因子对准确率的影响，对特征图赋上大小为 E 的衰减因子，试验不同的 E 值在 CIFAR100 数据集上的准确率 (accuracy，ACC)，如表 7 所示，当衰减因子 E 取值为 $e^{-6}\approx0.0025$ 时，模型能提取到更加丰富的特征，且分类准确率最高。

2) 为了探究旋转角度参数对网络模型准确率的影响，设置 5 组不同的旋转角度参数，分别为方案 1 不旋转、方案 2 旋转 90°、方案 3 旋转 180°、方案 4 旋转 270°、方案 5 可学习的旋转角度。以方案 3 为例，旋转 180°表示对经过负值指数抑制部分的特征图进行 180°的旋转，使特征图的拓扑结构与原图中心对称，这种操作帮助模型捕捉输入中的对称模式，实现空间对称性增强。为确定最优的旋转角度，采用上述不同的 5 种方案分别在 CIFAR-100、CIFAR-10、SVHN 数据集进行训练和测试，所有实验使用基本的数据增强并设置 200 个迭代轮次，最佳的旋转角度参数根据最高准确率取值。

表 8 为不同旋转角度参数对 SDPNet 网络模型在三个数据集上的实验结果，从中可以看出旋转角度不是越大越好或者越小越好。较大的旋转角度会导致语义被破坏，暴露更多图像边缘的填充区域，引入噪声，增加模型学习负担。较小的旋转角度会生成高度相似的特征，导致模型难以区分原始特征和旋转特征，反而增加收敛难度，使计算成本增加。

方案 3 旋转 180°，在 CIFAR-100、CIFAR-10、SVHN 数据集上分别取得 79.86%、96.83%、97.23% 的分类准确率，均为最高值。综上可得，180°为最优旋转角度，能在增强数据多样性和保持特征有效性之间取得最佳平衡，从而提高模型的效率和鲁棒性，实现图像分类任务中的几何不变性增强与动态特征校准。

3.6.2 SDP 部分方法实验

SDP 模块由负值指数抑制 (NES)、特征旋转 (FR)、动态扩张 (DE) 以及通道混洗 (CS) 部分方法构成，这些方法分别从特征抑制、空间变换、动态调整感受野以及通道变换的不同维度增强模型的表达能力，但各个部分方法的组合效应与各自的贡献尚不明确，为了明确核心有效方法，本文设计了从单一方法到全组合的 14 种不同部分方法排列组合，这些组合方式覆盖了所有可能的组合情形，能够有效分析出各方法的独立贡献和协同效应，具体组合方法如表 9 所示。

表 7 不同 E 值下的准确率

Table 7 Accuracy rates at different E values

E	$e^{-3}\approx0.0497$	$e^{-5}\approx0.0067$	$e^{-6}\approx0.0025$	$e^{-7}\approx0.0009$
ACC	73.29	75.93	78.96	76.32

采用上述 14 种组合分别在 CIFAR-100、CIFAR-10、SVHN 数据集上进行分类准确率测试，所有实验均采用相同的训练设置与超参数，以确保结果的可比性。具体实验结果如图 12 所示。

由图 12 可知，在单一方法中，单独加入 NES 方法的准确率分别为 78.79%、95.38%、95.92%，说明该模块的核心有效方法为 NES 方法，通过绝对值处理和指数衰减，抑制特征图中的负值响应，增强特征对比度，为后续所有分支处理打下基础；在组合方法中，全组合方法准确率分别为 79.86%、96.83%、97.23%，这种组合方法通过分阶段、多维度的动态空间感知，形成自适应的特征选择机制，从而提升模型的鲁棒性和泛化能力，提升分类结果的准确率。

表 8　不同旋转角度对分类准确率的影响 (单位：%)

Table 8　The impact of different rotation angles on classification accuracy (Unit: %)

Method	CIFAR-100	CIFAR-10	SVHN
Method1	78.96	96.22	96.93
Method2	79.39	96.34	97.05
Method3	**79.86**	**96.83**	**97.23**
Method4	79.42	96.60	96.58
Method5	79.09	96.72	96.47

表 9　各部分方法组合方式

Table 9　The combination methods of each part

Combination	NES	FR	DE	CS
1	√	—	—	—
2	—	√	—	—
3	—	—	√	—
4	—	—	—	√
5	√	√	—	—
6	√	—	√	—
7	√	—	—	√
8	—	√	√	—
9	—	√	—	√
10	—	—	√	√
11	√	√	√	—
12	√	—	√	√
13	—	√	√	√
14	√	√	√	√

3.6.3 SDP 插入位置实验

注意力机制在 CNN 中插入位置会对其识别率产生显著的影响，不同位置会引导模型聚焦于不同层次的特征表达，不恰当的插入位置除了可能会因注意力机制错误聚焦无关信息导致性能下降外，还可能会产生额外的参数量和计算开销，降低模型推理速度，这些影响互相关联，需要根据具体任务和资源限制来权衡注意力机制的插入位置。本文设计了 6 种不同插入方案来比较 SDP 注意力机制的位置以及数量对准确率的影响，如图 13 所示。

在控制训练参数一致的条件下，基于 ResNet-34 网络在 CIFAR-100、CIFAR-10、SVHN 中训练和验证，实验结果如表 10 所示。

实验结果表明，将 SDP 模块插入在位置 1 上时，在 CIFAR-100、CIFAR-10、SVHN 数据集上分类准确率最高，分别为 79.86%、96.83%、97.23%，该现象说明 SDP 模块与高层网络结合能最大化其效果，通过高层网络来实现更大的感受野和关键特征聚焦能力，实现从局部细节到全局上下文的层次化感知，进而 SDP 机制精准捕捉其中的复杂空间与通道交互，引导网络动态关注关键区域，从而增强特征的判别性。因此，将 SDP 模块置于位置 1 能够更充分地发挥其注意力引导作用，有效提升模型的整体准确率。

3.7 DFR 模块实验

为了研究 DFR 残差块嵌入位置与数量对 SDPNet 性能的影响，本研究在残差连接处嵌入 DFR 残差块 (简记为 D-Block)，依次替换原始 ResNet-34 结构中的残差连接模块 (简记为 C-Block) 设计了 8 种排列组合方式，如图 14 所示。

上述 8 种方案在 CIFAR-100、CIFAR-10、SVHN 数据集上的分类准确率测试结果如图 15 所示。由图 15 可知，组合方式 H 在 CIFAR-100、CIFAR-10、SVHN 数据集上的分类准确率分别为 80.63%、96.98%、97.63%，性能最优。这表明在残差连接处嵌入 3 个 D-Block 时能获得最佳的分类结果。

3.8 特征可视化分析

为了深入阐释 SDPNet 的特征提取过程，本研究对其进行了可视化分析，选择如下网络进行对比：MobileNet_V2、EfficientNet、ResNet，在 CIFAR-100 数据集上随

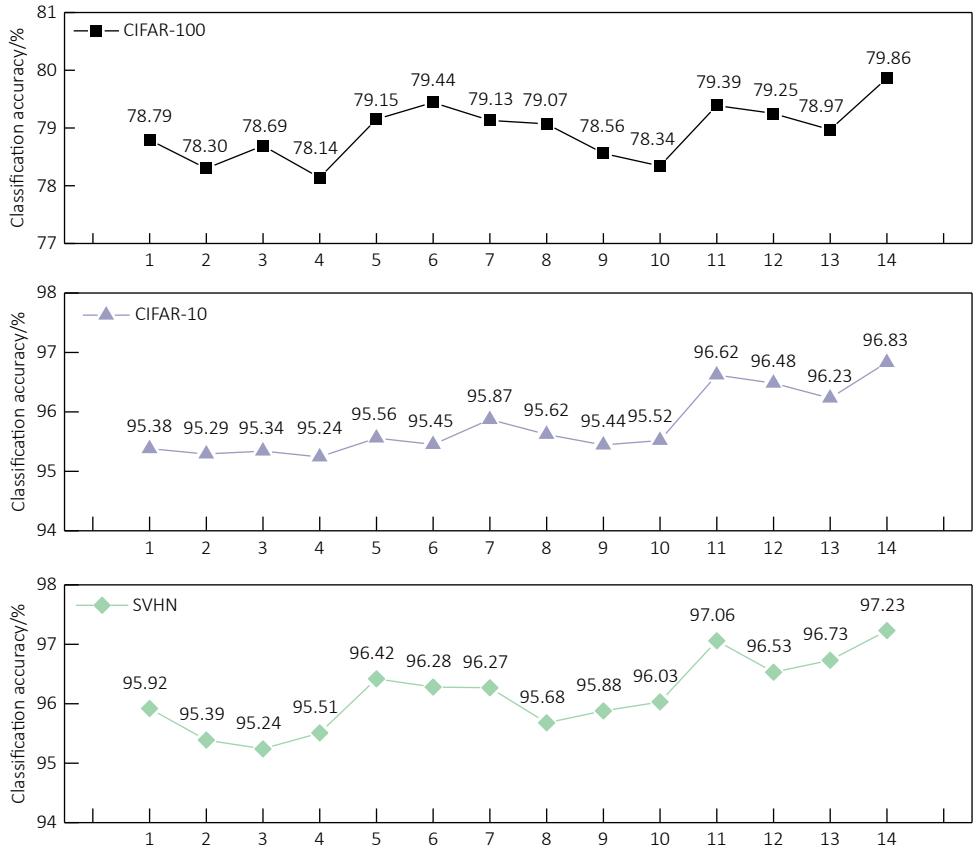


图 12 各部分方法组合方式对分类准确率的影响

Fig. 12 The influence of the combination methods of each part on the classification accuracy rate

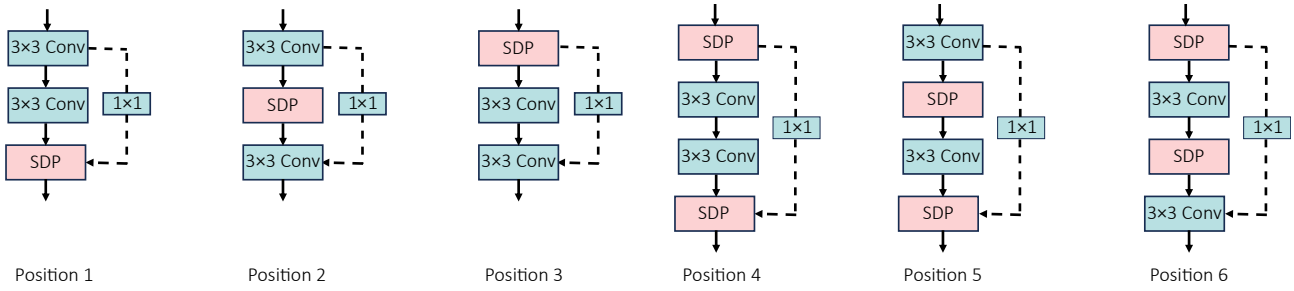


图 13 SDP-Block 位置和数量的排列图

Fig. 13 Arrangement structure diagram of SDP-Block position and quantity

表 10 不同位置对准确率的影响 (单位：%)

Table 10 The impact of different positions on accuracy (Unit: %)

Position	CIFAR-100	CIFAR-10	SVHN
Position1	**79.86**	**96.83**	**97.23**
Position2	78.97	95.62	96.71
Position3	79.05	96.26	96.93
Position4	79.82	96.33	97.20
Position5	79.64	96.04	96.89
Position6	78.72	95.41	96.73

机选取 5 张图像，其热力图可视化效果如图 16 所示。该网络模型能够有效学习不同类别图像的判别性特征实现准确分类。在图像分类任务中，热力图上的红色区域代表高响应权重，这些区域对模型的预测起关键作用，而蓝色区域则表示低响应权重，其对最终预测结果的贡献相对较小。

由图 16 可知，MobileNet_V2 受限于其轻量化的网络结构，对图像细微特征的捕捉能力较弱，导致其热力图中的红色区域仅能覆盖最具代表性的关键特征，而难以进一步细化至更精细的局部结构。无法捕捉到细微的特征。EfficientNet 由于其独特的复合缩放机制引发的特征稀疏效应，网络不能关注重要区域特征，容易受外界因素的干扰，所以热力图中红色区域呈现过度扩散现象。相较于 MobileNet_V2 和 EfficientNet，ResNet-34 的网络层数更深，可以提取到更细致的特征，但是在局部特征和全局特征的提取上能力欠佳，所以热力图中红色区域虽然能覆盖图像中的主要特征，但往往难以捕捉到更细微的局部细节，导致部分关键信息被忽略。

由特征可视化分析可以看到，SDPNet 可以抑制噪声和冗余信息，聚焦关键区域和特征，在强化对重点特征关注的同时，还能高效利用关键特征，相比于其他网络，SDPNet 展现出更精细和全面的特征提取能力，从而增强了分类结果的可信度。

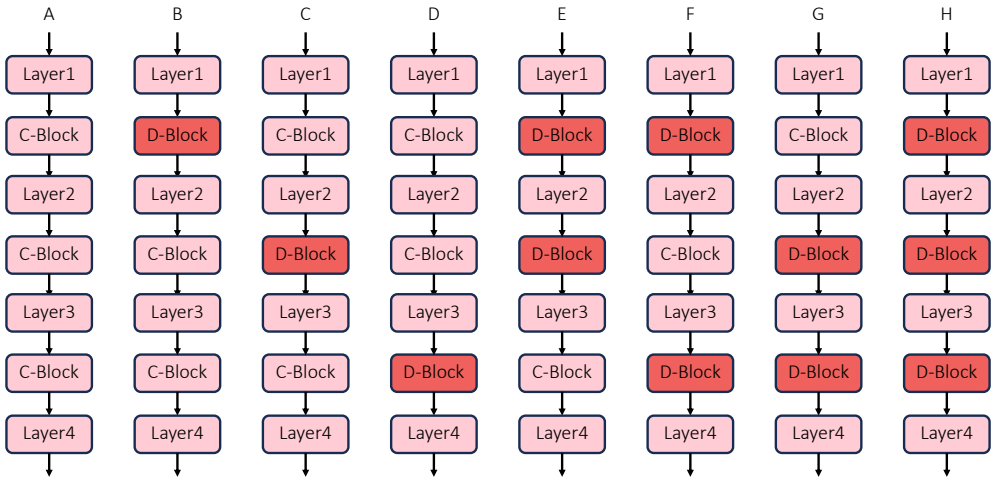


图 14 嵌入 D-Block 的 8 种组合方式

Fig. 14 Eight combinations of embedding D-Block

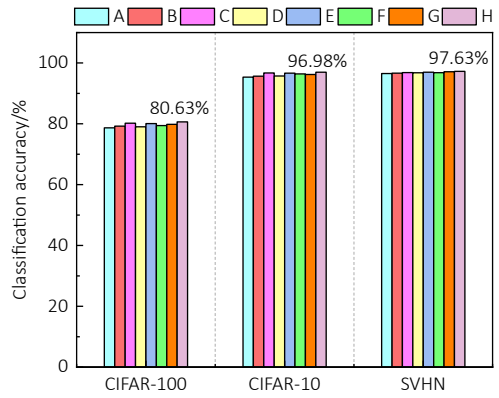


图 15 嵌入 D-Block 的 8 种组合方式对分类准确率的影响

Fig. 15 The impact of 8 combination methods of embedding D-Block on accuracy

4 结束语

针对图像分类时网络模型对重点特征的关注度较弱、空间和通道间的交互作用不够充分等问题，本文提出了空间动态感知的图像分类网络 SDPNet。主要贡献如下：1) 提出 NES 模块强化对图像空间特征的感知，其非线性特征变换方法可以增强网络对判别性特征的学习能力；2) 提出 SDP 模块，应用负值指数抑制技术以及空间旋转、动态扩张方法，捕捉空间特征中的判别性信息，实现空间特征增强； 3) 提出 DFR 模块，在缓解信息丢失的同时有效去除冗余信息，动态调整空间和通道的交互作用，优化特征表达，减少网络过拟合，提高网络的泛化性。

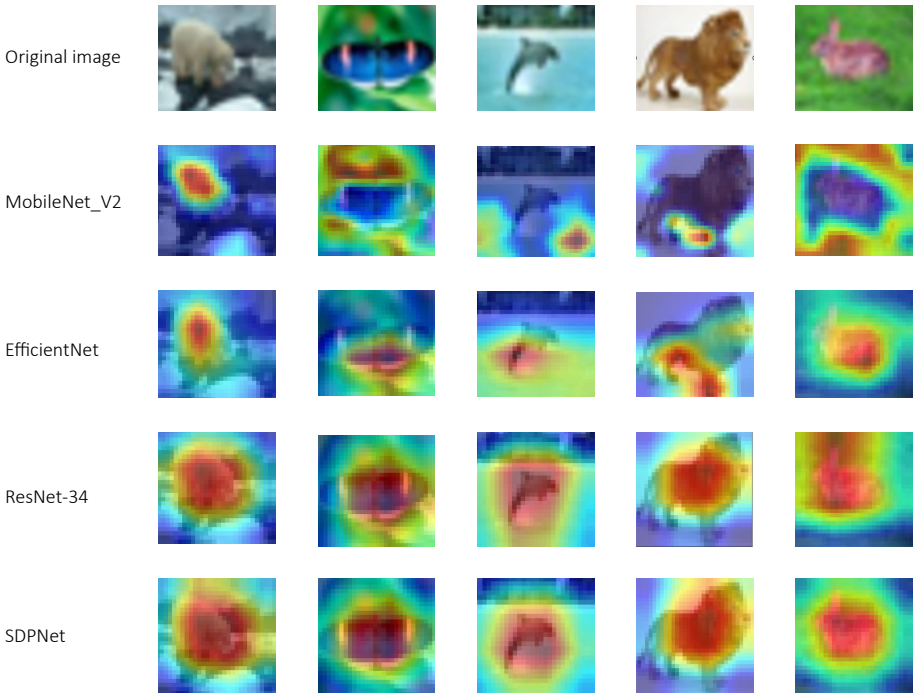


图 16 不同网络的可视化图像

Fig. 16 Visualization images of different networks

SDPNet 能够捕捉空间对称性，强化对图像空间特征的感知，在提升分类精度的同时还能稳定梯度。在 CIFAR-100、CIFAR-10、SVHN、Flowers-102、NWPU-RESISC45 多个数据集上的实验结果表明， SDPNet 在不同数据集上均表现优异，其性能优于其他对比模型。通过负值指数抑制、特征旋转、动态扩张以及通道混洗部分操作，使网络更关注对任务有用的特征，从而提升分类结果的准确率。由于空间动态感知的图像分类网络只限于空间域操作，下一步工作将结合频域感知进一步提升网络的分类性能。

作者贡献

姜文涛，张晟翀：指导论文选题、论文修改；纪佳莹：研究资料的收集、分析，实验的设计与完成，论文写作。

利益冲突

所有作者声明无利益冲突

参考文献

1. Yuan H, Liu J, Jiang W T, et al. Double-pooling residual classification network based on feature reordering attention mechanism[J]. *J Image Graphics*, 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.
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, Chen C, Zhang S C. Sparse feature image classification network with spatial position correction[J]. *Opto-Electron Eng*, 2024, **51** (5): 240050.
姜文涛, 陈晨, 张晟翀. 空间位置矫正的稀疏特征图像分类网络[J]. 光电工程, 2024, **51** (5): 240050.
4. Zhang J L, Ren J F, Zhang Q, et al. Spatial context-aware object-attentional network for multi-label image classification[J]. *IEEE Trans Image Process*, 2023, **32**: 3000−3012.
5. Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview[M]//Hassanien A E, Bhatnagar R, Darwish A. *Advanced Machine Learning Technologies and Applications*. Singapore: Springer, 2021: 599–608. https://doi.org/10.1007/978-981-15-3383-9_54.
6. Yuan H, Huo X R, Jiang W T. Image classification network of holographic gradient differential convolution[J]. *Acta Autom Sin*, 2025, **51** (9): 2106−2130.
袁姮, 霍欣燃, 姜文涛. 全息梯度差分卷积的图像分类网络[J]. 自动化学报, 2025, **51** (9): 2106−2130.
7. Liu Y, Xue J H, Zhang W D, et al. Image classification algorithm based on coordinate importance pooling and decoupled class alignment distillation[J]. *Acta Electron Sin*, 2025, **53** (3): 962−973.
刘颖, 薛家昊, 张伟东, 等. 基于坐标重要性池化和解耦类别对齐蒸馏的图像分类算法[J]. 电子学报, 2025, **53** (3): 962−973.
8. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with

deep convolutional neural networks[C]//*Proceedings of the 26th International Conference on Neural Information Processing Systems*, 2012: 1097–1105.

9. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2015.
10. Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
11. He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.
12. Chen H, Wang H W, Jin H B. Intrusion detection model combining improved self-encoder and residual network[J]. *Comput Eng*, 2024, **50** (2): 188−195.
陈虹, 王瀚文, 金海波. 融合改进自编码器和残差网络的入侵检测模型[J]. 计算机工程, 2024, **50** (2): 188−195.
13. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
14. 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)*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
15. Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7794–7803. https://doi.org/10.1109/CVPR.2018.00813.
16. 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*, 2021: 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986.
17. He F T, Wu Q Q, Yang Y, et al. Research on laser speckle image recognition technology based on transfer learning[J]. *Laser Technol*, 2024, **48** (3): 443−448.
贺锋涛, 吴倩倩, 杨祎, 等. 基于深度学习的激光散斑图像识别技术研究[J]. 激光技术, 2024, **48** (3): 443−448.
18. Jiang W T, Yuan H, Liu W J. Neuron signal attenuation activation mechanism for deep learning[J]. *Patterns*, 2025, **6** (1): 101117.
19. Tan M X, Le Q V. Efficientnet: rethinking model scaling for convolutional neural networks[C]//*Proceedings of the 36th International Conference on Machine Learning*, 2019: 6105–6114.
20. Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165.
21. Ji Q W, Yu B, Yang Z W, et al. LAM: lightweight attention module[C]//*Proceedings of the 15th International Conference on Knowledge Science, Engineering and Management*, 2022: 485–497. https://doi.org/10.1007/978-3-031-10986-7_39.
22. Zhou C L, Zhang H, Zhou Z K, et al. QKFormer: hierarchical spiking transformer using Q-K attention[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2024: 416.
23. Shin H, Choi D W. Teacher as a lenient expert: teacher-agnostic data-free knowledge distillation[C]//*Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 2024: 1672. https://doi.org/10.1609/aaai.v38i13.29420.
24. 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*, 2024: 16163–16173. https://doi.org/10.1109/CVPR52733.2024.01530.
25. Lan H, Wang X H, Wei X. Couplformer: rethinking vision transformer with coupling attention map[EB/OL]. arXiv: 2112.05425, 2021. https://doi.org/10.48550/arXiv.2112.05425.
26. Ma C X, Wu J B, Si C Y, et al. Scaling supervised local learning with augmented auxiliary networks[C]//*Proceedings of the 12th International Conference on Learning Representations*, 2024.
27. Abdi M, Nahavandi S. Multi-residual networks: improving the speed and accuracy of residual networks[EB/OL]. arXiv: 1609.05672, 2016. https://doi.org/10.48550/arXiv.1609.05672.
28. Yuan H, Fan T T, Gao Y. Dual-Frequency channel difference enhancement for image classification[J]. *Comput Eng Appl*, 2026, **62**(11): 259−271.
袁姮, 范桐桐, 高原. 双频通道差异增强的图像分类网络[J]. 计算机工程与应用, 2026, **62**(11): 259−271.
29. Luo Z B, Sun Z T, Zhou W L, et al. Rethinking ResNets: improved stacking strategies with high-order schemes for image classification[J]. *Complex Intell Syst*, 2022, **8** (4): 3395−3407.
30. Li J F, Wen Y, He L H. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 6153–6162. https://doi.org/10.1109/CVPR52729.2023.00596.
31. 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*, 2021: 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.
32. Huang J, Chen H R, Ren J X, et al. A general adaptive dual-level weighting mechanism for remote sensing pansharpening[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025: 7447–7456. https://doi.org/10.1109/CVPR52734.2025.00698.

作者简介



姜文涛 (1986-)，男，博士，副教授，主要研究方向为图像处理、模式识别、人工智能。

E-mail：jiangwentao@lntu.edu.cn



【通信作者】纪佳莹 (2000-)，女，硕士研究生，主要研究方向为图像处理、模式识别、人工智能。

E-mail：826796020@qq.com

Spatial dynamic perception image classification network

Jiang Wentao[1], Ji Jiaying[1*], Zhang Shengchong[2]

[1] College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;
[2] Key Laboratory of Optoelectronic Information Control and Security Technology, Tianjin 300308, China

Abstract

Objective Image classification serves as a fundamental task in computer vision, aiming to automatically assign semantic labels to input images through algorithmic models. Although deep learning, particularly convolutional neural networks (CNNs), has achieved remarkable progress in this domain, existing approaches still face significant challenges. Traditional attention mechanisms often overemphasize global relationships while sacrificing the perception of local fine-grained features and textures. Furthermore, these methods struggle to adapt to variations in object scale and spatial relationships, failing to fully integrate global spatial contextual information. The residual connections in deep networks can amplify noise and adversarial perturbations, leading to feature dilution in identity mappings. Additionally, most current architectures lack dynamic weight assignment capabilities, resulting in limited interpretability and flexible adaptability. To address these limitations, this paper proposes a Spatial Dynamic Perception Image Classification Network (SDPNet) that enhances the dynamic perception of critical features, strengthens spatial-channel interactions, improves robustness to geometric transformations, and increases overall network generalization capability.

Methods The proposed SDPNet is constructed upon the ResNet-34 architecture with three key innovations. First, the Negative Exponent Suppression (NES) module is introduced as a novel nonlinear feature transformation technique. This module applies an absolute value transformation to remove polarity differences while preserving feature continuity, followed by an exponential decay suppression operation that replaces negative values with their absolute values multiplied by a decay factor of exp(-6), compressing negative responses to approximately 0.25% of their original magnitude. This approach enhances feature contrast, highlights edge information, and strengthens the network's ability to learn discriminative features while maintaining a balance between information preservation and discriminative enhancement. Second, the Spatial Dynamic Perception (SDP) module is developed, integrating the NES module with three parallel processing branches: Feature Rotation (FR), Dynamic Expansion (DE), and Channel Shuffle (CS). The FR branch performs 180-degree center-symmetric rotation of feature maps followed by 1×1 depthwise separable convolutions and Tanh activation, capturing spatial symmetry and improving robustness to object orientation changes. The DE branch employs convolutional and deconvolutional operations with 3×3 and 2×2 kernels respectively, dynamically expanding receptive fields while preserving spatial resolution through nearest-neighbor interpolation upsampling, thereby capturing both fine details and holistic contours. The CS branch utilizes two 1×1 convolutions with channel shuffling between them to break fixed channel connection patterns, enhancing feature diversity and enabling cross-channel information exchange. These three branches are fused through spatial and channel average pooling followed by Tanh activation, enabling adaptive feature enhancement and dynamic selection of discriminative spatial information. Third, the Dual-Dimension Fusion Residual (DFR) module is proposed, which embeds the SDP attention mechanism into residual blocks and incorporates conditional convolutions into the dimension-reduced residual branches. The conditional convolution dynamically adjusts kernel weights based on input features through a routing function, transforming static convolution into dynamic, condition-dependent operations. This design enhances spatial and channel interactions, optimizes feature representation, reduces overfitting, and improves generalization capability. Additionally, the initial feature extraction layer is modified by replacing the 7×7 convolutional kernel with a 3×3 kernel and removing the max pooling layer to better preserve local details and original spatial information.

Results and Discussions Comprehensive experiments were conducted on five benchmark datasets: CIFAR-100, CIFAR-10, SVHN, Flowers-102, and NWPU-RESISC45. SDPNet achieved classification accuracies of 80.63%, 96.95%, 97.66%, 90.34%, and 92.46% respectively, outperforming eleven state-of-the-art models including ResNet-34, EfficientNets, GhostNet, QKFormer, TLENet, ATONet, Couplformer, SSLLNet, Multi-ResNet, DCDENet, and MobileNet-LAM. Compared with these competing methods, SDPNet demonstrated average accuracy improvements of 3.60%, 2.52%, 2.09%, 7.29%, and 2.36% across the respective datasets. The model's parameter count (21.93 M) remains comparable to ResNet-34 (21.35 M) while being substantially lower than EfficientNets (53.68 M) and Multi-ResNet (52.03 M). The computational complexity (7.5 GFLOPs) and inference speed (2.55 f/s) demonstrate practical deployability in resource-

Foundation item: National Natural Science Foundation of China (61601213), Liaoning Provincial Natural Science Foundation of China (20170540426), and Key Fund of Liaoning Provincial Department of Education (LJ2017QL034)

*Correspondence: Ji J Y, E-mail: 826796020@qq.com

constrained scenarios. Ablation studies confirm the complementary contributions of each module: SDP alone improves ResNet-34 by up to 4.14% on Flowers-102, DFR alone improves by up to 2.65%, and their combination yields synergistic gains of up to 7.59%, validating the effectiveness of the integrated design. Attention mechanism comparisons against SE, SCConv, CA, CBAM, and ADWM across CIFAR-100, Flowers-102, and STL-10 datasets show that SDP consistently achieves superior or comparable performance, with average improvements of 0.6%, 1.27%, and 1.56% respectively. The 180° rotation angle in the FR branch was empirically determined as optimal through systematic parameter analysis. Experiments on SDP module combinations reveal that the full NES+FR+DE+CS configuration achieves the highest accuracy, with NES identified as the core effective component. Position sensitivity analysis indicates that placing SDP at deeper network layers (position 1) maximizes its effectiveness by leveraging larger receptive fields for hierarchical spatial-channel interaction capture. Optimal DFR module placement involves embedding three D-Blocks within the residual connections. Visualization analysis through heatmaps demonstrates that SDPNet effectively suppresses noise and redundant information while focusing on critical regions, exhibiting more refined and comprehensive feature extraction capabilities compared to MobileNet_V2, EfficientNet, and ResNet-34.

Conclusions This paper presents SDPNet, a novel image classification network that addresses fundamental limitations in existing approaches through three synergistic contributions. The NES module enhances spatial feature perception through nonlinear feature transformation, the SDP module dynamically captures discriminative spatial information through integrated rotation, expansion, and shuffling operations, and the DFR module optimizes spatial-channel interactions through conditional convolutions. The comprehensive experimental validation across multiple datasets demonstrates SDPNet's superior classification performance, robust generalization capability, and practical efficiency. Future work will extend this spatial-domain approach by incorporating frequency-domain perception to further enhance classification performance.

Keywords: image classification; feature-selective inhibition; feature rotation; dynamic expansion; residual network

Jiang W T, Ji J Y, Zhang S C. Spatial dynamic perception image classification network[J]. *Opto-Electron Eng*, 2026, **53**(6): 250376; DOI: 10.12086/oee.2026.250376

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文