ISSN 1003-501X CN 51-1346/O4

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

空间位置矫正的稀疏特征图像分类网络

姜文涛，陈晨，张晟翀

引用本文:

姜文涛，陈晨，张晟翀. 空间位置矫正的稀疏特征图像分类网络[J]. 光电工程，2024， **51**(5): 240050.

Jiang W T, Chen C, Zhang S C. Sparse feature image classification network with spatial position correction[J]. *Opto-Electron Eng*, 2024, **51**(5): 240050.

https://doi.org/10.12086/oee.2024.240050

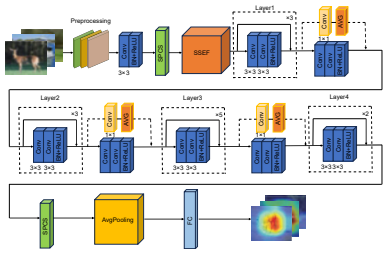
收稿日期: 2024-03-06; 修改日期: 2024-04-23; 录用日期: 2024-04-24

Opto-Electronic Engineering
光 电 工 程

Article
2024 年, 第 51 卷, 第 5 期

DOI: 10.12086/oee.2024.240050

空间位置矫正的稀疏特征图像分类网络



姜文涛[1]，陈　晨[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： 为稀疏语义并加强对重点特征的关注，增强空间位置和局部特征的关联性，对特征空间位置进行约束，本文提出空间位置矫正的稀疏特征图像分类网络 (SSCNet)。该网络以 ResNet-34 残差网络为基础，首先，提出稀疏语义强化特征模块 (SSEF)，SSEF 模块将深度可分离卷积 (DSC) 和 SE 相融合，在稀疏语义的同时增强特征提取能力，并能够保持空间信息的完整性；然后，提出空间位置矫正对称注意力机制 (SPCS)，SPCS 将对称全局坐标注意力机制加到网络特定位置中，能够加强特征之间的空间关系，对特征的空间位置进行约束和矫正，从而增强网路对全局细节特征的感知能力；最后，提出平均池化残差模块 (APM)，并将 APM 应用到网络的每个残差分支中，使网络能够更有效地捕捉全局特征信息，增强特征的平移不变性，延缓网络过拟合，提高网络的泛化能力。在多个数据集中，SSCNet 相比于其它高性能网络在分类准确率上均有不同程度的提升，证明了其在兼顾全局信息的同时，能够更好地提取局部细节信息，具有较高的分类准确率和较强的泛化性能。

关键词： 图像分类；特征提取；空间位置矫正；稀疏语义；对称注意力；全局感知

中图分类号： TP391.4　　**文献标志码：** A

姜文涛，陈晨，张晟翀. 空间位置矫正的稀疏特征图像分类网络 [J]. 光电工程，2024，**51**(5): 240050
Jiang W T, Chen C, Zhang S C. Sparse feature image classification network with spatial position correction[J]. *Opto-Electron Eng*, 2024, **51**(5): 240050

Sparse feature image classification network with spatial position correction

Jiang Wentao[1], Chen Chen[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: To sparse semantics and enhance attention to key features, enhance the correlation between spatial and local features, and constrain the spatial position of features, this paper proposes a sparse feature image classification network with spatial position correction (SSCNet) for spatial position correction. This network is based on the ResNet-34 residual network. Firstly, a sparse semantic enhanced feature (SSEF) module is proposed, which combines depthwise separable convolution (DSC) and SE to enhance feature extraction ability while maintaining the integrity of spatial information; Then, the spatial position correction symmetric attention mechanism (SPCS) is proposed. SPCS adds the symmetric global coordinate attention mechanism to specific positions in the network,

收稿日期：2024-03-06；修回日期：2024-04-23；录用日期：2024-04-24
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 (LJYL049)
*通信作者：陈晨，867428188@qq.com。

which can strengthen the spatial relationships between features, constrain and correct the spatial positions of features, and enhance the network's perception of global detailed features; Finally, the average pooling module (APM) is proposed and applied to each residual branch of the network, enabling the network to more effectively capture global feature information, enhance feature translation invariance, delay network overfitting, and improve network generalization ability. In the CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewood datasets, SSCNet has shown varying degrees of improvement in classification accuracy compared to other high-performance networks, proving that SSCNet can better extract local detail information while balancing global information, with high classification accuracy and strong generalization performance.

Keywords: image classification; feature extraction; space position correction; sparse semantics; symmetric attention; global perception

1 引　言

图像分类旨在将输入的图像分为不同的预定义类别，其主要目的是通过训练一个模型，使其能够自动识别和理解图像中的内容，并将其归到相应的类别中。图像分类在计算机视觉和人工智能领域，如农林园艺、医学影像分析、自动驾驶等，均具有广泛应用。然而，由于视觉变化、类内差异、类间相似等因素，使得同一个物体在不同图像中的外观表现不同，增加了图像分类的难度。

传统图像分类方法依赖手工设计的特征提取器，如尺度不变特征变换 (SIFT)、方向梯度直方图 (HOG) 和局部二值模式 (LBP) 等，这些方法依赖领域专家的经验，难以捕捉到图像中的高层语义信息。而基于深度学习的图像分类方法，通过神经网络自动学习图像的特征表示，能够捕捉更丰富、更准确的图像特征，自动理解图像高层语义信息，在图像分类中取得了显著的成果[1-2]。

深度学习图像分类方法由于其卓越的性能，已逐渐成为主流图像分类方法。如 Lecun 等[3] 提出 LeNet，引入了卷积和池化层的结构，通过局部感知和参数共享提取图像特征，然而，LeNet 适用性局限并且网络相对较浅，难以学习复杂特征。Krizhevsky 等[4] 提出 AlexNet，通过较深的网络结构和大规模的训练数据，实现了更好的特征学习和表达能力。然而，AlexNet 容易过拟合，对于小样本数据表现不佳，且网络结构相对较简单，无法有效地处理一些复杂的视觉任务。He 等[5] 提出 ResNet (residual network)，通过引入残差连接来解决深层网络训练中的梯度消失和模型退化问题。残差连接有助于信息流动，提高网络的收敛速度和准确性，但模型较复杂，需要更多的计算资源和参数，在一些较小的数据集上容易过拟合。基于此，在基础残差网络上进行更改[6]，Wang 等[7] 提出一种多分辨率的卷积神经网络结构 HRNet (high-resolution network)，在保留高分辨率和低分辨率特征信息的同时，通过逐级融合来提取多尺度的特征表示，然而模型的计算复杂度较高。Xue 等[8] 提出 IX-ResNet (interpolation-extended residual network)，采用了一种新的碎片化多尺度融合特征转换的策略，将多个大型同构模块堆叠成网络，每个大型模块由多个小型异构模块组成，这样设计能够提高网络的表达能力和学习能力，但会导致网络结构的复杂性增加，使得模型难以训练和优化。Jiang 等[9] 提出了 ADD-ResNet (aggregated decentralized down-sampling-based residual network)，一种新的聚合分散下采样的策略，将未参与卷积运算的区域重新卷积，并堆叠加到向前传播层和短路层的深度信息上，保证特征映射的逐渐收敛，避免了特征信息的丢失，然而该网络需要额外的卷积计算，导致计算复杂度和内存消耗的增加。Luo 等[10] 提出了 HO-ResNet (high order residual network)，在广泛使用的 CV 基准上进行充分的实验来验证假设，性能得到稳定而显著的提高，收敛性和鲁棒性也得到改善，但存在对不同位置关键信息捕捉不精确等问题。Jafar 等[11] 提出 HOD-ResNet (high-speed hyperparameter optimization deep residual network)，为具有不同层数的 ResNet 模型构建了一个超参数优化方法，并表明网络结构的优化显著提高网络性能，但无法捕捉不同位置的重要信息。

为提升模型的表示能力，捕捉更丰富、更准确的特征表示，提高模型对关键信息的感知能力，研究者采用注意力机制与网络结构相结合来提升网络的性能[12-14]。Hu 等[15] 提出 SE (squeeze-and-excitation)，通过学习通道之间的关系，实现通道的自适应缩放，从而提升模型的表达能力和性能。Ying 等[16] 提出 PSE

(PSigmoid SE)，不仅以信道方式抑制特征，还增强特征提取能力，提高分类准确率，但是在空间感知性方面存在不足。Hou 等[17] 提出 CA (coordinate attention)，用于处理图像或图像序列的注意力机制，通过引入绝对位置信息来增强模型对图像中不同位置关系之间的建模能力，提升模型的性能和对空间结构的理解能力。Ji 等[18] 提出 LAM (lightweight attention module)，用于轻量级卷积神经网络，有效地集成注意力机制，在空间模块中使用元素加法和更小的卷积核，避免梯度消失的问题。Zhong 等[19] 提出在 CBS (Conv batch normalization SiLU) 模块中合并不同的注意力机制，发现与 CA 合并后的网络模块准确率最高，然而该模块忽略了全局上下文信息，限制了对全局信息的利用。Qi 等[20] 提出了一种新的坐标注意力模块来提高卷积神经网络的分类精度，并且证明了该方法的有效性，提高了分类准确率，但处理复杂数据或特定任务时受限。

现有方法由于模型结构的限制，在处理复杂语义信息时感知能力不足，无法捕捉到更加细致的特征；重点关注局部特征之间的关系，缺乏对全局空间信息的有效利用，对全局细节感知能力不足；存在过拟合风险，全局空间信息的保留有限，影响分类准确性。针对以上问题，本文提出空间位置矫正的稀疏语义图像分类网络。主要贡献有:

1) 设计即插即用的稀疏语义强化特征模块 SSEF (sparse semantic enhanced feature module)，SSEF 模块将 DSC 和 SE 注意力机制进行融合，可以依据通道内特征信息的重要性来分配权重，在稀疏语义的同时，提取更加细致、关键的特征，并保持空间信息的完整性。

2) 提出空间位置矫正对称注意力机制 (spatial position correction symmetric attention mechanism, SPCS)，SPCS 通过调整不同位置的权重，对特征图中的空间位置进行校正和调整，使网络更准确地处理重要的特征区域，并加强特征之间的空间关系，突出重要空间位置，从而更好地理解图像的结构和空间布局，提高对全局细节的感知能力。

3) 提出平均池化残差模块 (average pooling module, APM)，APM 是在残差分支中添加平均池化，引入正则化效果，在残差路径中增强特征传递能力。加入 APM 的网络在增强特征的平移不变性的同时，能够更有效地捕捉全局特征信息，提高网络的泛化能力。

2 SSCNet 网络结构

为增强特征的空间关系，加强对重要特征的提取，提高特征表达能力，增强网络泛化能力，在保持高性能的同时提高计算效率，本文提出空间位置矫正的稀疏特征图像分类网络 (sparse feature image classification network with spatial position correction, SSCNet)，网络结构如图 1 所示。

本文网络包括如下 5 个部分：

第 1 部分：在网络浅层阶段，图像输入存在冗余信息，影响提取特征的准确性。因此，将输入的图像经过预处理、3×3 卷积层和 BN[21] 层，提取图像精细的局部特征，提高浅层特征提取的精确度，减小信息损失，提高模型性能。

第 2 部分：传统的浅层特征提取方法难以充分捕捉图像中的关键信息，对图像全局空间位置的感知性较差，降低了模型的鲁棒性。因此，将处理后的图像经过空间位置矫正对称注意力机制的首层，强化图像中关键信息的感知能力，提高空间建模能力，提升模型的鲁棒性并减少冗余信息。

第 3 部分：网络模型层数加深导致网络参数量的增加，降低网络计算效率，且存在冗余信息，导致对关键特征信息判别不足，降低模型的准确率。因此，将处理后的图像经过 SSEF 模块，从而降低特征维度，减少模型参数，并自适应地调整特征通道的重要性，更好地捕捉关键信息并保持特征的空间信息，进而有助于提高模型的感受野和特征的判别性。

第 4 部分：网络层数加深不仅导致参数量的增加，还容易出现梯度爆炸的问题，使网络的学习能力受限，对全局信息提取不充分。因此，将处理后的图像输入到 Layer1、Layer2、Layer3、Layer4 中，进行深层特征提取。在残差分支中通过 APM 模块降低空间维度、提取全局信息并改善模型的平移不变性。

第 5 部分：在网络的训练和学习过程中，没有对深层特征的权重进行学习和重新分配，不利于网络捕捉关键信息。因此，将输出特征图输入到空间位置矫正对称注意力机制的尾层，网络进一步调整特征的权重，更好地捕捉特征的关键信息，形成矫正后的空间位置，并且将特征图输入到平均池化和全连接层，获得特征降维并输出结果。

2.1 浅层特征提取模块

在传统 ResNet-34 残差网络中，使用 7×7 的卷积

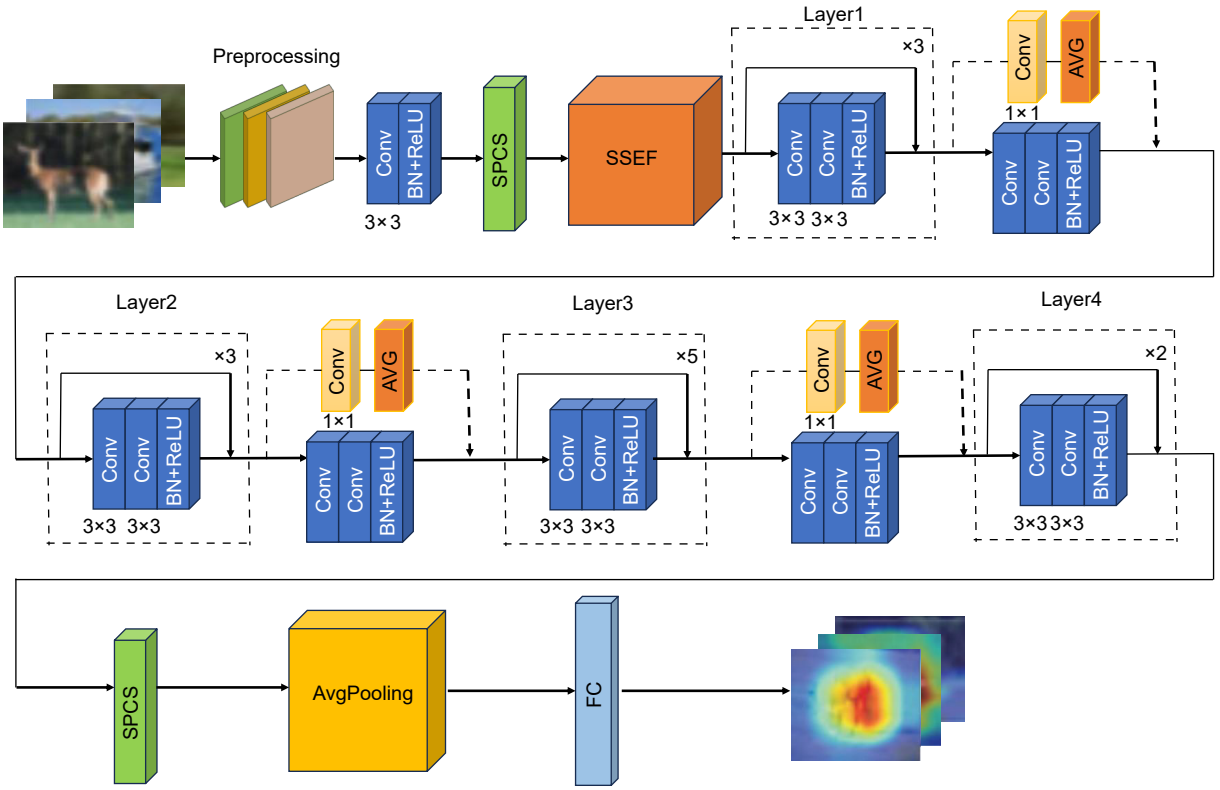


图 1 SSCNet 网络结构

Fig. 1 SSCNet network structure

层和最大池化层来提取图像的浅层特征，然而，这种方法会导致原始图像信息丢失。为避免原始图像浅层信息丢失，对网络浅层特征提取模块进行修改，以更好地平衡特征表达和保留图像信息。

为提取更加细微的特征，采用 3×3 的小卷积核替代首层中 7×7 卷积核，并删除最大池化层，减小卷积核的步长和填充大小，保持输出特征图的尺寸与输入图像相同，从而保留原始图像的信息。通过改进浅层特征提取方式，在有效提取底层特征的同时保留原始图像的重要信息。首层卷积核尺寸前后对比如图 2 所示。

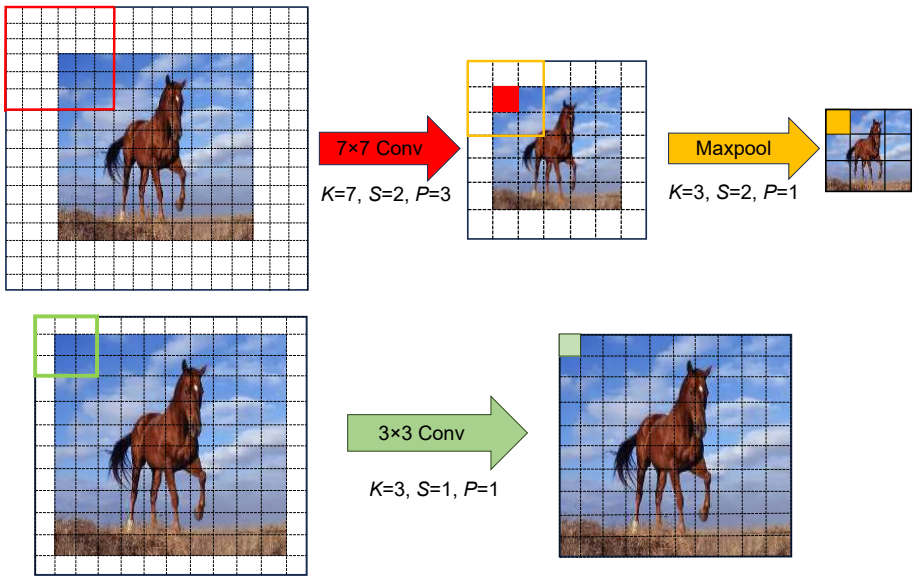


图 2 修改首层卷积核尺寸前后卷积操作对比

Fig. 2 Comparison of convolution operations before and after modifying the size of the first layer convolution kernel

以尺寸 10×10 的图像为例，进行浅层特征提取对比，卷积后的尺寸大小表示为

$$output\_size = \frac{input\_size - kernel\_size + 2 \times padding}{stride + 1}, \tag{1}$$

其中：*output_size*表示输出特征图尺寸，*input_size*表示输入特征图尺寸，*kernel_size*表示卷积核尺寸，*stride*表示步长，*padding*表示填充。由式 (1) 可得，10×10 的图像经过未更改的浅层特征提取，得出特征图尺寸为 3×3，对后续深层网络的特征提取不便。经过更改后的浅层特征提取模块，输出的特征仍然为 10×10，保留了更完整的原始图像信息，便于后续特征提取。

2.2 稀疏语义强化特征模块

为增强模型对通道特征的关注，更好地捕捉重要特征信息，减少参数数量并提高模型效率，本文提出稀疏语义强化特征模块 SSEF，SSEF 结构如图 3 所示。

本文设计的 SSEF 模块由 1×1 卷积和 DSC 融合 SE 模块组合而成。通过 1×1 卷积学习通道间的相关性，调整特征图的通道数，减少冗余信息并降低维度，实现语义稀疏。进入 DSC 初始阶段，将输出的特征图根据通道维度分成 N 组，每一组通道对应 3×3 大小的卷积独立学习空间中的相关性，之后进行逐点卷积独立学习通道间的相关性，通过分解卷积操作来减少参数量，模块更加关注特征的语义信息，减少冗余参数，实现语义稀疏的效果。再将这 N 组送入压缩激励模块，根据通道内特征信息的重要性来分配权重系数，从而增强对重要特征的提取，再次进行稀疏语义减少冗余信息。将得到的特征再次经过 1×1 卷积，对不同通道的特征进行线性组合实现特征的融合，调整通道数量进一步减少冗余信息，从而增强语义稀疏的效果，提高模型效率。最后，将特征经过 BN 层和 ReLU 激活函数，缓解网络中梯度消失，增强模型的学习能力，从而更有效地提取关键特征。将特征图分成 N 组，N 值的不同会导致通道提取特征的不同，选取不同的 N 值来观察模型准确率 (ACC)，结果如表 1 所示。由表 1 可知，当 N 取值为 64 时，可以提取到更加丰富的特征，模型的准确率最高。

表 1　不同 N 值下的准确率

Table 1　Accuracy under different N values

N	24	32	64	128
ACC/%	72.19	76.53	78.91	77.78

特征图通过 1×1 卷积学习通道间的相关性并提取特征，调整特征图的通道数与后续模型维度相匹配，减少参数量和冗余信息并降低维度，实现语义稀疏。输入的特征F_{in}经过 1×1 卷积处理后得到F_1：

$$F_1 = Conv1(F_{\text{in}}), \tag{2}$$

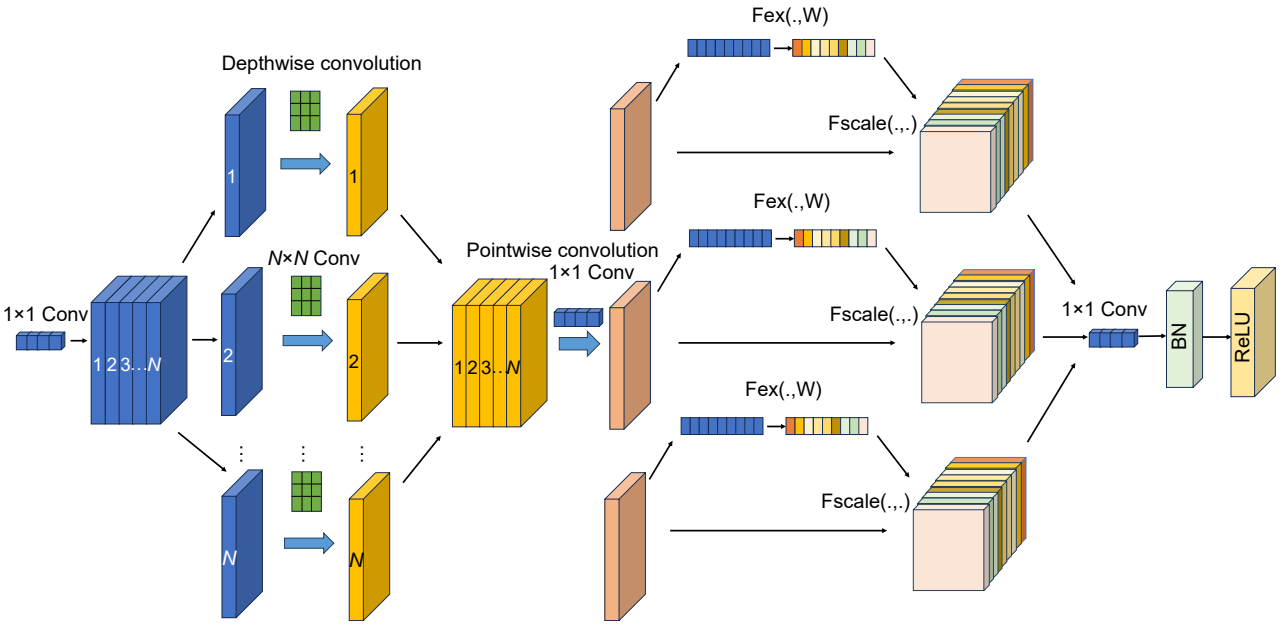


图 3　SSEF 模块

Fig. 3　SSEF module

其中：$F \in R^{H\times W\times R}$表示特征图，Conv1 表示进行 1×1 卷积操作。特征经过 DSC 并融合 SE 模块。DSC 由两层构成：深度卷积和点卷积。深度卷积为每个输入通道应用一个过滤器，每个输入通道滤波器的深度卷积表示为

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m}\,, \tag{3}$$

其中：$\hat{K}$为大小为$D_K \times D_K \times M$的深度卷积核，$\hat{K}$中的第 M 个滤波器应用于 F 中的第 M 个信道，得到滤波后的输出特征映射$\hat{G}$的第 M 个信道。深度卷积计算代价为：$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$，深度卷积相对于标准卷积是非常有效的。但只过滤输入通道，没有将它们组合起来创建新特性。因此，需要通过 1×1 卷积来计算深度卷积输出的线性组合。深度可分卷积的代价：$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$，是深度卷积和 1×1 点卷积的和，$N$ 为输出通道数，将卷积表示为滤波和组合的两步过程，计算量减少为

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}\,. \tag{4}$$

可以看出 DSC 参数量比起标准卷积大大减少。假设输出 256 通道，卷积核尺寸为 3×3，DSC 计算量只有标准卷积的 12%。由此可知，该模块具有稀疏语义减少模型参数的效果。特征经过 DSC 处理后得到 P，其中$Depthwise(\cdot)$表示深度卷积，$Pointwise(\cdot)$表示点卷积。P 的计算式为

$$P = (Pointwise(Depthwise(Conv1(F_{\text{in}})))\,. \tag{5}$$

在 DSC 的逐点卷积中融合 SE。首先，将处理后的特征经过 Squeeze，通过全局平均池化将特征在空间维度上进行压缩，得到通道的全局特征，特征图从尺寸为$W \times H \times C$转换成$1 \times 1 \times C$。然后，特征经过 Excitation，通过全连接层 FC1、SeLU 激活函数、全连接层 FC2 和 sigmoid 激活函数，学习不同通道间的特征信息，对全局特征进行激发，得到不同通道间的权重，输出的权重数目和输入特征图的通道数相同，即$1 \times 1 \times C$。最后，通过 Scale，将前面得到的归一化权重加权到每个通道的特征上，输出特征图尺寸大小为$W \times H \times C$。输入的特征 P 经过 DSC 融合 SE 处理后得到

$$S_n = (Squeeze(P) + Excitation(P)) * Scale(P)\,. \tag{6}$$

将所得到的特征经过 1×1 卷积、BN 层和 ReLU 激活函数得最终输出F_{out}，对不同通道特征进行线性组合实现特征融合，提高模块稳定性，更有效地提取关键特征。其中$\gamma(\cdot)$表示 BN 层，$\delta(\cdot)$表示 ReLU 激活函数。F_{out} 表达式为

$$F_{\text{out}} = \delta(\gamma(Conv1((Squeeze(P) + Excitation(P)) * Scale(P)))\,. \tag{7}$$

2.3 空间位置矫正对称注意力模块

传统的 ResNet-34 残差网络缺乏位置感知性，对全局信息理解受限，无法准确地捕捉到重要的局部特征，导致特征提取不充分。为解决上述问题，本文提出空间位置矫正对称注意力机制 SPCS，结构如图 4 所示。

坐标注意力机制同时考虑通道维度和空间维度上的注意力，通过学习自适应的通道权重，使模型更加关注重要的通道信息，但由于在对整个特征图进行注意力权重计算时，缺乏对全局信息的整合，导致模型在处理全局上下文和长距离依赖时存在一定的局限性。本文提出的 SPCS 模块由两个全局坐标注意力机制 (max coordinate attention mechanism, MCA) 构成，两个 MCA 分别对称地放入网络的浅层位置和残差层尾部，对空间位置进行矫正，可以有效增强网络在通道维度和空间维度上的感知能力并提高网络上下文信息

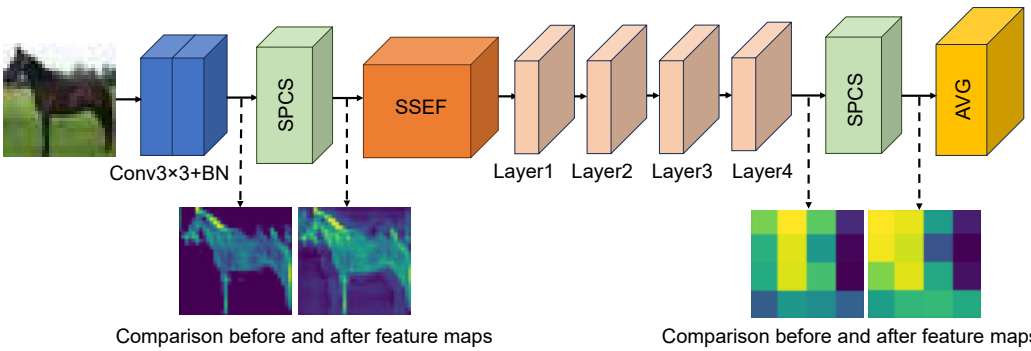


图 4 空间位置矫正对称注意力

Fig. 4 Spatial position rectification symmetric attention

的交互能力。原始的坐标注意力机制与 MCA 的内部结构图分别如图 5 和图 6 所示。传统的坐标注意力机制内部结构应用平均池化操作会导致特征图模糊，无法明确分清重要特征和次要特征，导致重要细节信息丢失，并且对空间位置感知不足。改进后的全局坐标注意力机制应用全局池化操作来突出图像中显著特征并捕捉局部空间的变化和细节信息，增强空间位置的敏感性。

首先，将 SPCS 应用于网络的浅层位置，通过对输入特征图的通道维度和空间维度进行建模，以捕捉全局和局部的关键特征。在通道维度上，SPCS 自适应地调整通道的重要性，使网络更加关注重要特征；在空间维度上，SPCS 捕捉到不同位置之间的关联性，允许网络充分利用空间信息，更大范围感知网络的上下文信息。对称性体现在对每个通道的权重调整和每个空间位置的关联性计算上，确保通道和空间的对称处理，将感知到的上下文信息根据空间位置的关联性计算紧密地联系起来，提取更充分的上下文空间信息，从而提高对空间位置的矫正能力。然后，在网络的尾部再次使用 SPCS，此处被选定是因为在网络的较深层级中，特征图的表达能力和抽象程度更高，对特征的加权和整合变得尤为重要。通过此处应用 SPCS，网络进一步调整特征权重，更好地捕捉输入特征的关键信息。在浅层处 SPCS 捕捉输入图像局部细节信息，根据空间位置信息自适应地调整各个位置的注意力权重，使网络捕捉不同层级的上下文信息，在尾部处 SPCS 增强网络对整体上下文感知能力，整合全局上下文信息和位置关系提高模型空间感知能力。对称性体现在整体网络使用 SPCS 应用上，确保对称的特征加权和整合，更加准确地利用上下文信息，对特征的空间位置进行矫正。由图 4 可知，无论是在网络的浅层位置还是残差层尾部，通过 SPCS 后的特征图与未加 SPCS 的特征图相比明显变亮，特征更加明显，说明提取到更加细致、更加重要的特征。

坐标注意力的全局池通常用于通道注意力以全局编码空间信息，但它将全局空间信息压缩到通道中，难以保存位置信息，为使注意力模块用精确的位置信息在空间上捕捉长距离的交互，本文将全局池分解为等式。给定输入 x，使用池核 $(H, 1)$ 或 $(1, W)$ 的两个空间范围分别沿水平坐标和垂直坐标对每个通道进行编码。因此，高度 h 处的第 c 通道的输出和宽度 w 处的第 c 通道的输出分别为式 (8)、(9) 所示：

$$Z_c^h(h) = \frac{1}{W}\sum_{0\leqslant i\leqslant W} x_c(h, j)\,, \tag{8}$$

$$Z_c^w(w) = \frac{1}{H}\sum_{0\leqslant i\leqslant H} x_c(j, w)\,. \tag{9}$$

上述两个变换分别沿着两个空间方向聚合特征，生成一对方向感知特征图。这两种变换允许注意力模块捕获沿一个空间方向的长距离依赖关系，并沿另一个方向保留精确的位置信息，有助于网络提取更重要

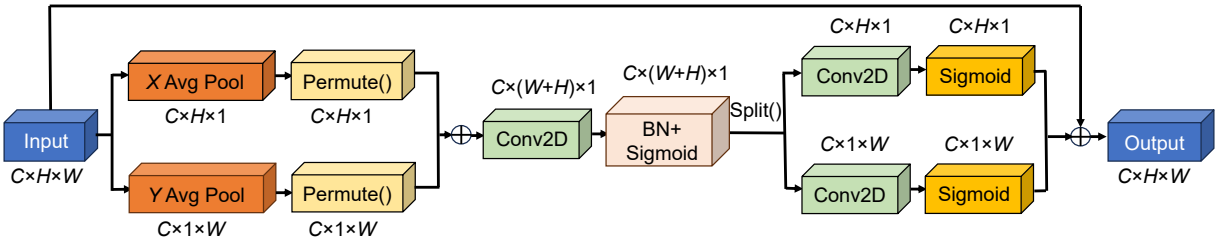


图 5 坐标注意力结构

Fig. 5 Coordinate attention structure

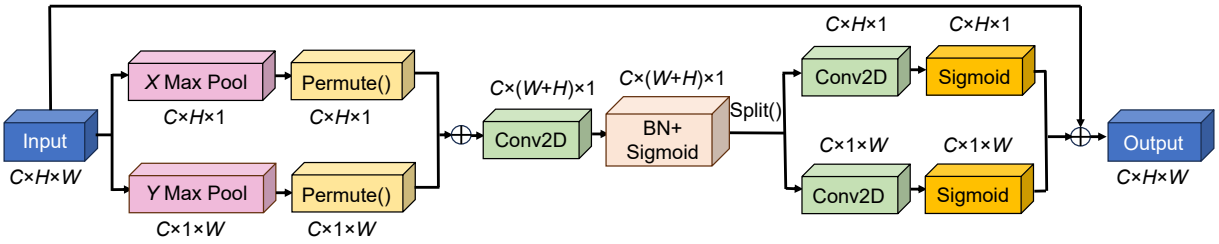


图 6 全局坐标注意力结构

Fig. 6 Max coordinate attention structure

的特征信息。将式 (8) 和式 (9) 连接起来，然后将其发送到共享的 1×1 卷积变换函数F_1，得到：

$$f = \delta(F_1([Z^h, Z^w])), \tag{10}$$

其中：$[\cdot,\cdot]$表示沿着空间维度的级联操作，δ是非线性激活函数，$f \in R^{C/r\times(H+W)\times 1}$是在水平方向和垂直方向上编码空间信息的中间特征图。然后，沿着空间维度将f分成两个独立的张量$\boldsymbol{f}^h \in R^{C/r\times H\times 1}$和$\boldsymbol{f}^w \in R^{C/r\times 1\times W}$，另外两个 1×1 卷积变换$F_h$和$F_w$用于将$\boldsymbol{f}^h$和$\boldsymbol{f}^w$分别变换为与输入 X 具有相同通道数的张量，从而得到g^h和g^w，输出g^h和g^w被扩展并分别用作注意力权重，最后可得出$y_c(i,j)$：

$$\boldsymbol{g}^h = \delta(F_h(\boldsymbol{f}^h)), \tag{11}$$

$$\boldsymbol{g}^w = \delta(F_w(\boldsymbol{f}^w)), \tag{12}$$

$$y_c(i,j) = x_c(i,j)\times g_c^h(i)\times g_c^w(j). \tag{13}$$

坐标注意力还考虑对空间信息进行编码。如上所述，沿水平和垂直方向的注意力同时应用于输入张量，两个注意力图中的每个元素都反映了感兴趣的对象是否存在于相应的行和列中。这种编码过程允许坐标注意力更准确地定位感兴趣对象的准确位置，从而帮助整个模型更好地识别。总之，在 ResNet-34 残差网络中引入 SPCS，确保在通道和空间维度上的特征权重调整是相互对应的，这种对称性设计使网络能够充分利用通道和空间信息，提高模型对空间位置的理解以及增强上下文感知能力，增强空间位置矫正能力，并提高网络的表达能力和性能，空间位置矫正使得特征加权和整合过程在不同位置和层级上具有一致性，为网络提供强大且准确的特征表示能力。

2.4 平均池化残差模块

在图像分类中，随着网络模型层次深度的增加，传统卷积神经网络会出现梯度消失和梯度爆炸等问题，导致模型训练难度大，错误率高。针对以上问题，He 等[5] 提出采用跳跃残差连接来建立网络模型，从而加深网络，缓解梯度消失、梯度爆炸等问题。ResNet-34 残差网络在深度神经网络中取得了成功，但也存在不足，如 ResNet-34 残差网络的残差分支中特征图维度保持不变，在后续卷积操作需处理较大的特征图，增加了计算量，降低网络运行速度；残差分支中没有降采样操作，特征图的感受野相对较小，导致处理全局上下文信息时性能不佳；残差分支中使用普通卷积操作，会引入大量参数，增加模型过拟合风险。为解决上述问题，本文提出平均池化残差模块 APM，如图 7 所示。

APM 是由 2 个 3×3 卷积以及包含 1 个 1×1 卷积和平均池化的残差分支构成，APM 在残差分支中降低特征图的空间维度，保留主要特征减少冗余信息，减少后续卷积层的计算负担。将 SSEF 模块处理后的信息在每一次残差中进行筛选，再次除去冗余信息，使得模型具有更高的计算效率。APM 能增大感受野，捕捉残差中更多上下文信息，增强残差中特征的感知能力，为尾处 SPCS 提供更大范围的上下文信息，加强全局特征的感知能力。APM 在保持平移不变性的同时引入正则化效果，减少过拟合的风险，增强模型泛化能力。

3 实验及结果分析

3.1 实验环境

本文实验中采用的操作系统为 Ubuntu 18.04 硬件环境为：NVDIA Tesla P100 显卡、60 GB 内存。 软件环境为：Pytorch 1.9.7、CUDA 12.1、CUDNN 8.3。编程语言为 Python 3.9.13。

本文采用 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集作为实验数据集。CIFAR-10 数据集包含 10 个类别共计 60000 张彩色图

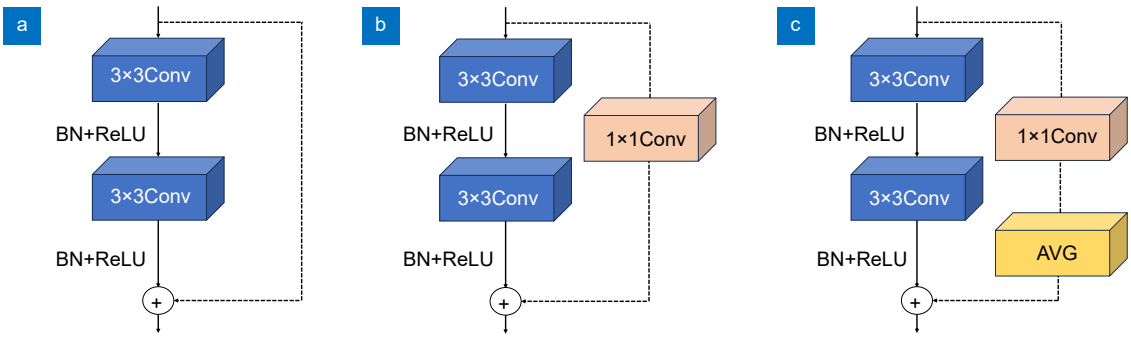


图 7 三种残差块比较。(a) 残差块；(b) 降维后的残差块；(c) 平均池化残差模块

Fig. 7 Three types of residual blocks. (a) Basic block; (b) Residual block; (c) APM

像，每张图像的分辨率为 32×32 像素，该数据集常用于测试和比较图像分类算法的性能。CIFAR-100 数据集是 CIFAR-10 的扩展，共包含 100 个类别，其中每个类别包含 600 张图像，这些类别分为 20 个大类，每个大类包含 5 个小类，每张图像的分辨率也是 32×32 像素，CIFAR-100 数据集更具挑战性，用于评估对更细粒度图像分类任务的模型性能。SVHN 数据集包含真实世界中的街景房屋号码图像，它包含来自 Google 街景图像的数字图像，用于识别房屋门牌上的数字，SVHN 数据集中的图像分为训练集、测试集和额外的训练集，每张图像包含一个或多个数字，并且图像的分辨率也比 CIFAR 数据集更高。Imagenette、Imagewoof 数据集是 ImageNet 中提取的小规模子集，在 ImageNet 数据集的基础上进行了精简和调整。具体数据集信息表如表 2 所示。

表 2　实验数据集

Table 2　Experimental datasets

Dataset	Size	Classification	Trainset	Testset
CIFAR-10	32×32	10	50000	10000
CIFAR-100	32×32	100	50000	10000
SVHN	32×32	10	73257	26032
Imagenette	224×224	10	9469	3925
Imagewoof	224×224	10	9025	3929

3.2 参数对网络性能的影响

3.2.1 不同模块的数量和位置对网络性能的影响

本文 SSCNet 网络性能受以下参数的影响：APM 模块嵌入的位置与数量、SSEF 模块所放置的位置、残差网络第一层卷积核尺寸 k 和学习率 lr。针对 32×32 分辨率图像，设置第一层卷积核尺寸为 3，学习率为 0.1，迭代次数为 300 轮。对于 224×224 分辨率图像，与上述不同的是卷积核尺寸为 7，不删除最大池化层，迭代次数为 260 轮，此时分析各种情况下对分类准确率的影响，确定好最优的网络结构，分析不同模块所放位置、不同模块的数量对分类准确率的影响，选择最优的网络参数。ResNet-34 残差网络模型共有 3 处降维残差块 (basic-block) 记为 LB，使用降维残差块会导致信息丢失和模糊，以及对特征的全局表达能力有限，为弥补这些不足，本文采用 APM 模块 (APM-Block) 记为 LA，在整体的 ResNet-34 残差网络中加入 SSEF 模块和 SPCS 机制，来提升网络的性能和表达能力。为更好研究 APM-Block 和 SSEF 模块的位置和数量对分类准确率的影响，本文设计如图 8 所示组合方式。

不同位置的 SSEF 模块和不同数量及不同位置的 APM-Block 在 3 个数据集上对分类准确率的影响如表 3 所示。从表 3 中得出，排列方式 H 在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上的分类准确率分别为 96.72%、80.63%、97.43%、88.75%、82.09%，均为最高值。因此，在 SPCS 后加入 SSEF 模块并且在 3 个特征降维处分别嵌入 APM-Block 效果最好。

3.2.2 卷积核尺寸和学习率对网络性能的影响

网络第一层卷积核尺寸对于特征提取能力至关重要，它决定了网络如何对原始图像进行特征提取。在基于 SSCNet 的实验中，对比了使用不同尺寸的卷积

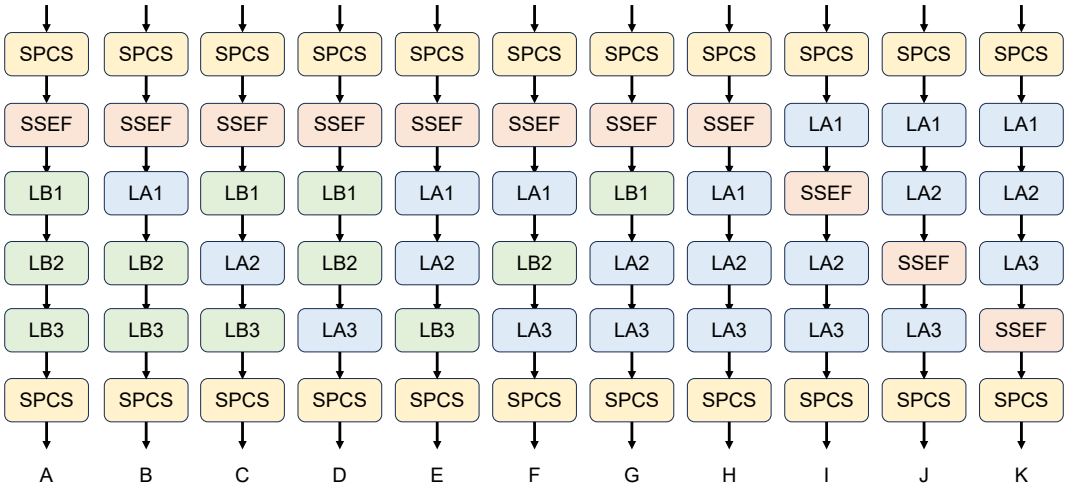


图 8　APM-Block 和 SSEF 模块位置和数量的排列方式

Fig. 8　The arrangement of APM-Block and SSEF module positions and quantities

表 3 不同位置和数量 SSEF 模块与 APM-Block 对分类准确率的影响

Table 3 Influence of different positions and numbers of SSEF modules and APM-Blocks on classification accuracy

	CIFAR-10/%	CIFAR-100/%	SVHN/%	Imagenette/%	Imagewoof/%
A	94.79	77.22	95.63	86.76	80.56
B	95.19	77.86	96.37	87.17	81.17
C	95.13	77.63	96.25	87.09	80.86
D	95.27	78.03	96.35	87.12	81.06
E	95.89	78.79	96.97	87.79	81.35
F	95.76	78.57	96.56	87.64	81.29
G	95.69	78.63	96.89	87.72	81.41
H	**96.72**	**80.63**	**97.43**	**88.75**	**82.09**
I	96.37	80.12	97.26	88.39	81.71
J	96.13	79.81	97.19	88.27	81.63
K	96.46	80.52	97.31	88.51	81.76

核 (3×3、5×5、7×7、9×9 和 11×11) 的影响，其它参数保持不变。图 9 展示了不同尺寸卷积核对分类准确率的影响。根据图 9 的结果，观察到随着卷积核尺寸的增大，分类准确率并没有提升，反而在尺寸为 3×3 的卷积核上达到最优值。进一步增大卷积核尺寸导致特征提取能力退化，使网络分类准确率下降，因此，尺寸为 3×3 的卷积核在特征提取能力和分类准确率方面表现最佳。

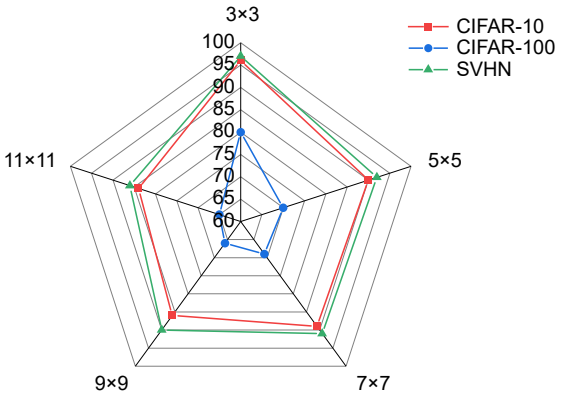


图 9 不同卷积核尺寸对分类准确率的影响

Fig. 9 Influence of different convolutional kernel sizes on classification accuracy

学习率的不同会对模型的权重更新有影响，进而对模型的收敛有影响，本文在其它因素不变的情况下，选择一个较大的学习率范围。在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上，初始学习率设为 0.01~0.5，在训练过程中逐渐降低学习率，对于 CIFAR-10、CIFAR-100、SVH N 在 60、120、160 轮降低至原来的 0.2 倍，针对 Imagenette、Imagewoof 在 100、180 轮降低至原来的 0.2 倍，同时观察分类精度的变化。不同学习率对分类准确率的影响如图 10 所示，可知学习率 LR5 的初始学习率为 0.1 时，在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上分类准确率分别 96.72%、80.61%、97.43%、88.75%、82.09%，均为最优值。

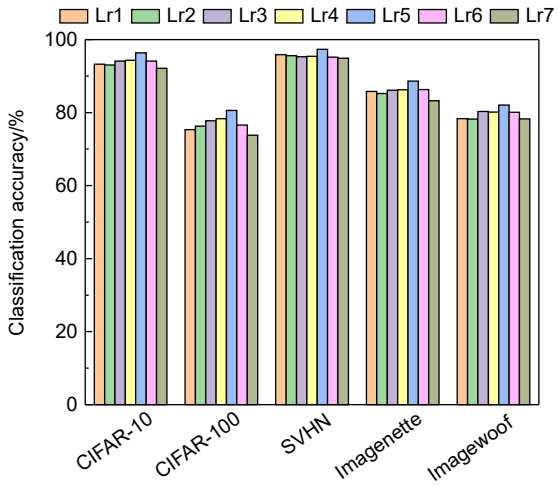


图 10 不同学习率对分类准确率的影响

Fig. 10 Influence of different learning rates on classification accuracy

3.2.3 减少模型参数量对网络性能的影响

本文提出的 SSEF 模块是 SSCNet 网络参数减少的主要原因，通过比较不同模块的参数量和准确率，来验证 SSEF 模块减少模型参数、降低特征维度、强化特征提取并提高准确率的性能。

采用数据集 CIFAR-100，在 ResNet34 网络残差层之前相同位置添加模块 S1、S2、S3，分别将 SSEF 模块中的 DSC 换成普通卷积、膨胀卷积 (dilated convolution, DC) 和混合卷积 (mixed depthwise convolutional kernels, MDC)，实验结果如表 4 所示。由表 4 可知，SSEF 模型的参数量和损失值最少并且准确率最高，可验证出 1×1 卷积在初始阶段调整特征图通道数从而降低维度减小计算量，经过 DSC 与 SE 融合阶段，减小模型的参数量并提取通道中重要的特征，在减少参数的同时提高模型的准确率。

表 4 SSEF 模块对参数和准确率的影响

Table 4 The impact of SSEF module on parameters and accuracy

Module	Param	ACC/%	Loss
S1	45696	78.12	0.82
S2	45696	78.53	0.79
S3	49920	78.19	0.81
SSEF	13504	78.96	0.76

为验证模型参数量减少对网络计算效率的影响，在相同的实验环境和数据集下，采用多个指标来验证不同网络模型的性能，其中 Speed 表示网络运行速度，能够衡量网络运行的快慢，Params 表示参数量，能够衡量网络的空间复杂度，FLOPs 表示浮点运算次数，能够衡量网络的计算复杂度。实验结果如表 5 所示。

表 5 不同网络下参数量减少对计算效率的影响

Table 5 The impact of parameters reduction on computational efficiency under different networks

Network	Speed/(f/s)	Params/M	FLOPs/G	ACC/%
Multi-ResNet[22]	1.62	51.23	37.93	78.68
ResNet-PSE[16]	2.23	40.56	27.56	72.81
ResNeXt-PSE[16]	2.07	47.29	31.34	77.32
SSLLNet[23]	2.57	31.57	20.86	79.23
ATONet[24]	2.88	30.12	16.91	78.54
SSCNet	3.05	21.36	11.71	80.63

由表 5 可知，SSCNet 网络的 Speed 均高于其它网络，且 Params 均低于其它网络，说明在减少参数量的同时提高了模型的运行速度，SSCNet 网络的 FLOPs 的值均低于其它网络，并且 ACC 的值达到最高，证明了本文所提 SSEF 模块的有效性，SSEF 模块不仅能够提高网络模型的计算效率，且能提高网络模型的准确性。

3.3 特征图对比分析

为更好地观察 SPCS 模块在网络中的效果，通过对 SPCS 模块前后的特征图进行可视化，并对其进行对比分析，通过直观感受特征图来验证 SPCS 模块的效果。从 10 种类别的测试集中分别随机选取 1 张经过不同位置的特征图，不同位置所对应的不同特征图结果如图 11 所示。由图 11 可以清楚地观察到，未经过 SPCS 模块首尾部的特征图提取细微特征的能力较差，对全局有效特征的位置信息感知能力不足，缺乏对上下文的感知能力，导致对图像的识别能力有限。观察经过 SPCS 模块首尾部的特征图可看出，重要特征区域明显变亮，例如马的鬃毛和臀部、鹿的腹部和鹿角以及深层特征提取中的关键像素位置由暗变亮，说明通过 SPCS 模块增强特征图对全局空间位置的理解，加强对重要细微特征的提取，依靠对称性更有效地处理上下文关键信息，对空间位置进行约束排除冗余信息，提取关键位置的特征。

3.4 消融实验

为验证提出模块的有效性，本文在不同数据集上对 SSEF 模块、SPCS 模块、APM 模块进行消融实验。在 CIFAR-10、CIFAR-100、SVHN、Imagenette 数据集上的准确率分别记为 ACC1、ACC2、ACC3、ACC4，模型运行速度记为 Speed，浮点运算次数记为 FLOPs。第 1 组 Net_1 表示在 SSCNet 中去除 SPCS 模块，第 2 组 Net_2 表示在 SSCNet 中去除 SSEF 模块，第 3 组 Net_3 表示在 SSCNet 中去除 APM 模块，第 4 组 Net_4 是本文 SSCNet 网络。消融实验结果如表 6 所示，表中加粗字体为最优值。在四个数据集上的消融实验结果如图 12 所示。

由表 6 和图 12 知，SSCNet 网络的分类准确率、模型运算速度和浮点运算次数最优，进而验证在 SSCNet 网络采用 SPCS 模块实现了多层次的空间建模，强化重要位置和特征，对特征空间位置进行矫正，增强了模型对细节的感知能力，提升模型性能。SSCNet 网络模型的运行速度最快，浮点运算次数最低，进而验证 SSCNet 网络采用 SSEF 模块在减少参数数量的同时强化特征的提取能力，提高网络计算效率。SSCNet 网络的残差分支中使用 APM 模块，能够降低特征图的空间维度，减少冗余信息提高模型计算效率。由此可知，在 SSCNet 网络中同时使用 SSEF 模块、SPCS 模块和 APM 模块，能有效地提高网络的性能。

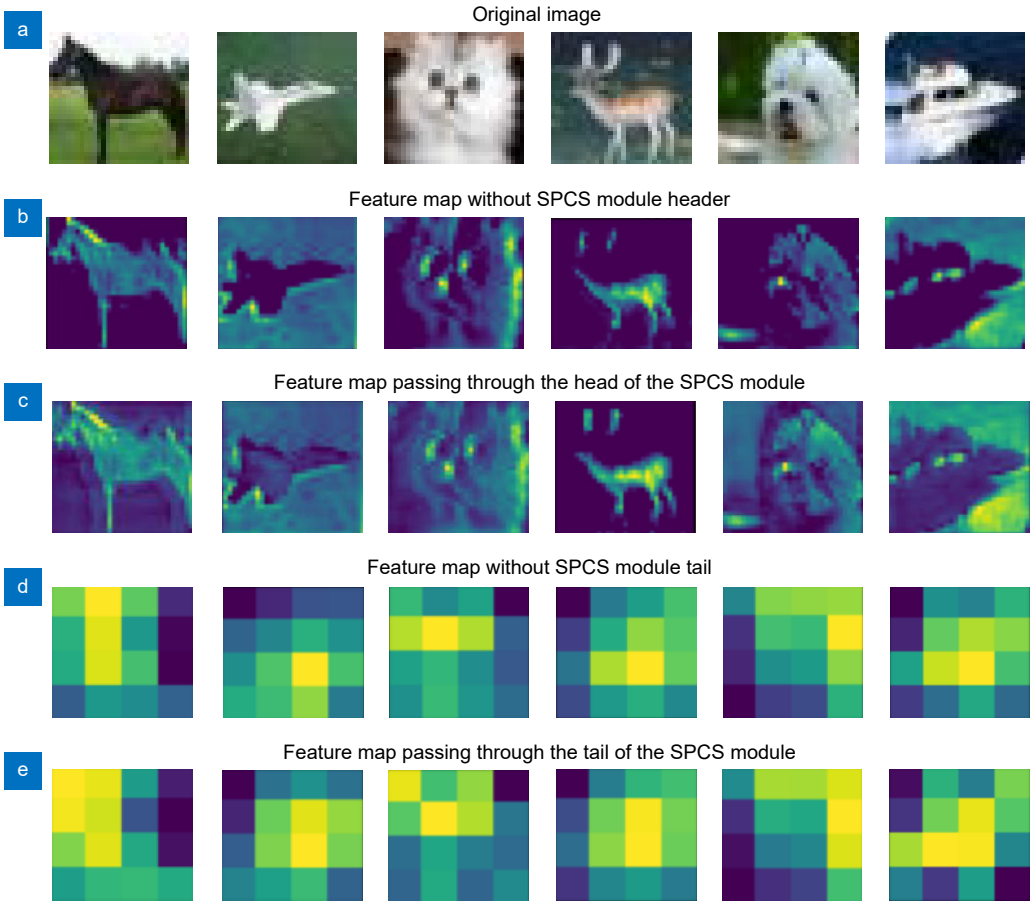


图 11 经过 SPCS 模块的特征图前后对比

Fig. 11 Comparison of feature maps before and after SPCS module

表 6 各模块之间在不同数据集上的消融实验

Table 6 Ablation experiments between different modules on different datasets

Group	SPCS	SSEF	APM	ACC1/%	ACC2/%	ACC3/%	ACC4/%	Speed/ (f/s)	FLOPs/G
1	-	√	√	90.23	69.63	92.89	83.67	2.85	13.93
2	√	-	√	92.36	75.51	94.17	85.23	2.32	26.67
3	√	√	-	95.67	77.34	96.56	87.09	2.96	12.36
4	√	√	√	**96.72**	**80.63**	**97.43**	**88.75**	**3.05**	**11.17**

3.5 网络模型对比实验

为证实 SSCNet 网络的先进性，在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上进行对比实验。实验选择对比分类方法有：ResNet-34、HO-ResNet、CAPRDenseNet[25]、MobileNet-LAM、Multi-ResNet[22]、Couplformer[26]、ResNet-PSE、ResNeXt-PSE、ATONet[24]、QKFormer[27]、TLENet[28] 和 SSLENet[23] 网络。本文分类准确率的对比实验数据分别来自于对应文献提供的实验结果和开源代码的复现，加粗的网络名称为开源代码的复现，引用文献实验结果的网络已给出具体的参考文献，加粗实验指标为最优值。

更改 ResNet 的网络结构，如 Multi-ResNet 和 HO-ResNet，此网络在 Cifar10、Cifar100、SVHN、Imagenette 和 Imagewoof 上的准确率分别为 94.56%、78.54%、94.58%、87.69%、81.21 和 96.32%、77.12%、95.69%、86.23%、79.64%。Multi-ResNet 网络增加残差块中残差函数的数量，并提出模型并行技术，将残差块计算分配给多个处理器，提高计算的准确性。HO-ResNet 网络提出一种基于高阶数值方法的网络堆叠策略，这种策略不仅提高网络的准确性，还改善网络的收敛性和鲁棒性。虽然在更改网络模型处取得了

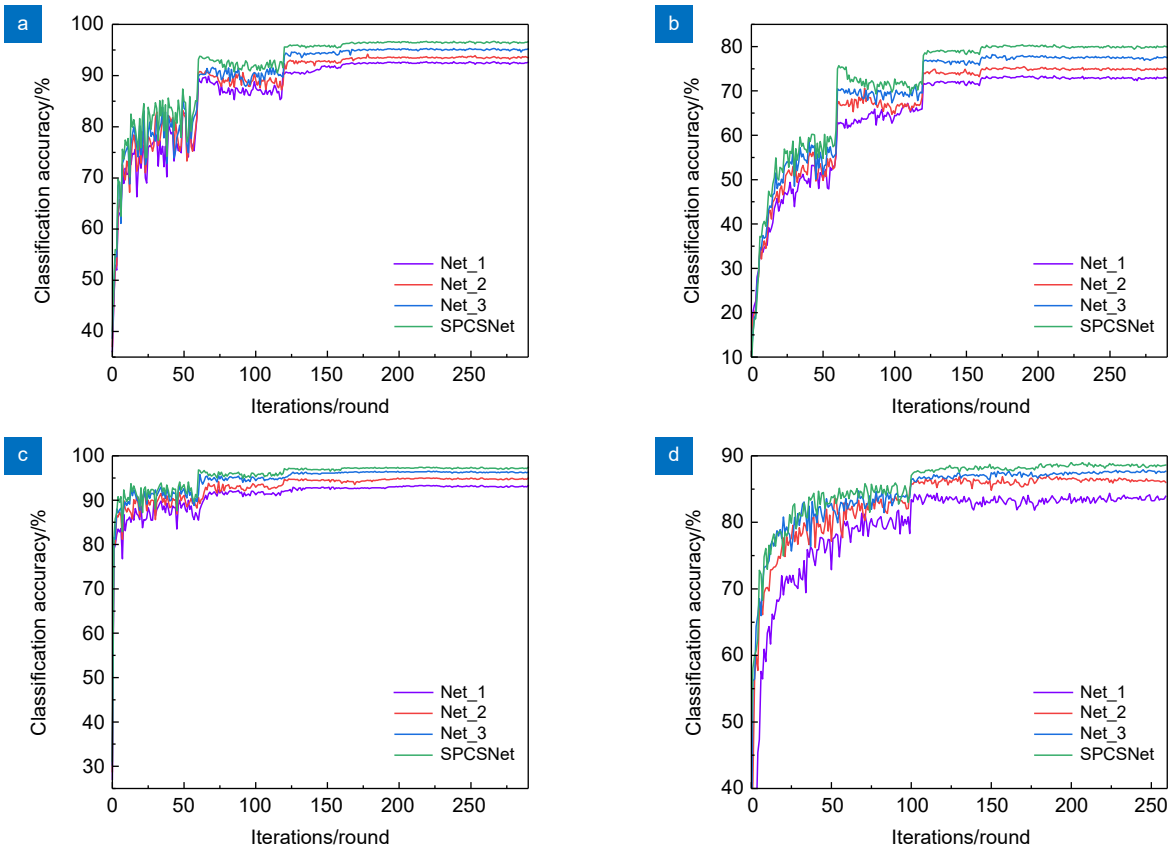


图 12 不同数据集上各网络分类准确率。(a) CIFAR-10; (b) CIFAR-100; (c) SVHN; (d) Imageneete

Fig. 12 Classification accuracy of each network on different datasets. (a) CIFAR-10; (b) CIFAR-100; (c) SVHN; (d) Imageneete

进步，但是随着网络层次的加深，Params 和 FLOPs 指标大大增加，分别达到 52.06 M、51.23 M 和 35.69 G、37.93 G，远高于 SSCNet，且梯度消失、梯度爆炸、网路退化和计算资源需求高的问题还未得到有效的解决。

加入注意力机制的网络，如在 MobileNet 中加入 LAM，MobileNet-LAM 在 Cifar10、Cifar100 的准确率为 89.37%、68.09%。LAM 模块同时考虑通道注意力模块和空间注意力模块，利用图像在通道和空间维度上的特征，提升模型的表达能力和性能。QKFormer 提出 Q-K 注意力机制，通过使用具有线性复杂度的二进制向量来有效地对通道维度进行建模，捕捉网络中重要信息。在 Cifar10、Cifar100、SVHN、Imagenette 和 Imagewoof 上的准确率分别为 96.18%、80.26%、97.13%、88.32% 和 81.65%。在 ResNet 和 ResNeXt 中加入 PSE 模块，引入了 PSigmoid 作为 SE 的激活函数，通过引入通道参数和共享参数的结合，提高激活函数的表达能力。在 Cifar10、Cifar100、SVHN、Imagenette 和 Imagewoof 上的准确率为 93.22%、73.03%、96.14%、85.09%、79.13% 和 94.25%、77.48%、96.54%、86.27%、80.66%。加入注意力机制的网络能够增强特征表达、提高模型的鲁棒性和泛化能力，增强局部和全局信息的融合，并提供可解释性和可视化能力，但是也面临计算复杂度增大和过拟合的风险。

计算效率可以反映出网络的性能，ATONet 提出 Auto-Train-Once 方法训练目标模型，指导目标模型权重的学习，避免陷入局部最优的问题，提高模型计算效率。TLENet 提出 TA-DFKD 方法提高样本的多样性，增强模型鲁棒性。SSLENet 提出一种增强的局部学习规则，通过在每个隐藏层的辅助网络中选择一部分来自其后续网络层，建立每个局部层的辅助网络，提高模型准确率。ATONet、TLENet 和 SSLENet 在 Speed 指标上的值均低于 SSCNet 网络，在 Params 和 FLOPs 的值均高于 SSCNet 网络，说明 SSCNet 网络在减少参数量的同时提高网络的计算效率，具有更高的运算速度，并且 SSCNet 网络在准确率指标上达到最优。

在 CIFAR-10、CIFAR-100、SVHN、Imagenette、Imagewoof 数据集上，SSCNet 网络与其它网络的准确率、运行速度、参数量和浮点运算次数对比分析结果如表 7 所示。由表 7 可知，与基线网络 ResNet-34 相比，SSCNet 网络的参数量和浮点运算次数与基线网络 ResNet-34 网络基本持平，但准确率大幅提升。与其它网络相比，SSCNet 网络的所有指标均达到最好，优于其它网络。由实验结果可知，SSCNet 网络在减少参数量的同时，提高了模型的运行速度，并且提升了模型的准确率。

3.6 不同网络的热力图可视化

为进一步验证 SSCNet 的有效性，不同网络在 CIFAR-10 数据集上的热力图可视化效果如图 13 所示，可清楚地观察到，MobileNet 网络由于轻量级的缘故，

表 7　各网络在 5 个数据集上的不同指标

Table 7　Different metrics for each network on five datasets

Network	CIFAR10	CIFAR-100/%	SVHN	Imagenette/%	Imagewoof/%	Speed/(f/s)	Params/M	FLOPs/G
ResNet-34[5]	87.82	68.92	91.39	84.91	78.86	3.02	21.32	11.63
HO-ResNet[10]	96.32	77.12	95.69	86.23	79.64	1.93	50.26	35.69
CAPRDenseNet[25]	94.24	78.84	94.95	87.56	80.79	2.86	25.51	17.73
MobileNet-LAM[18]	89.37	68.09	------	------	------	------	------	------
Multi-ResNet[22]	94.56	78.68	94.58	87.69	81.21	1.62	51.23	37.93
Couplformer[26]	93.54	73.92	94.26	85.13	79.08	2.73	27.63	14.29
ResNet-PSE[16]	92.89	72.81	96.14	85.09	79.13	2.23	40.56	27.56
ResNeXt-PSE[16]	93.92	77.32	96.54	86.27	80.66	2.07	47.29	31.34
ATONet[24]	94.51	78.54	95.21	86.67	80.19	2.88	30.12	16.91
QKFormer[27]	96.18	80.26	97.13	88.32	81.65	2.36	35.62	26.39
TLENet[28]	95.46	78.42	96.83	87.62	80.57	2.19	46.67	30.57
SSLLNet[23]	95.51	79.23	96.91	87.93	80.89	2.57	31.57	20.86
SSCNet	96.72	80.63	97.43	88.75	82.09	3.05	21.36	11.71

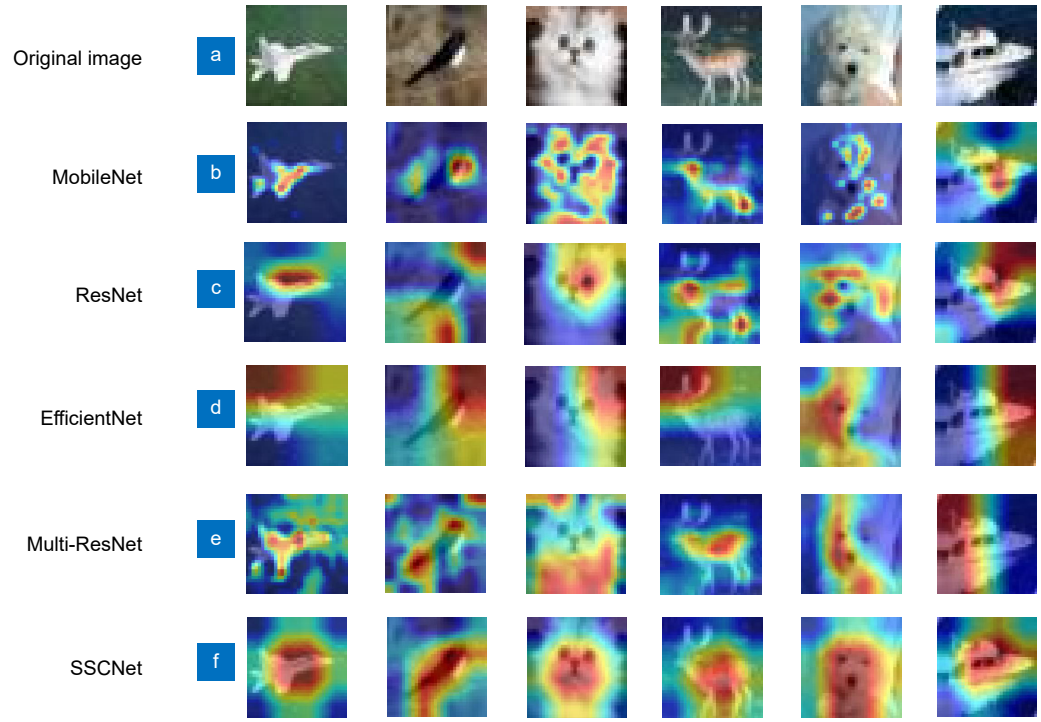


图 13　不同网络的热力图可视化图像

Fig. 13　Visualization images of heat maps for different networks

提取深层次特征的能力有限，分类效果不佳。ResNet 网络可捕捉更深层次特征但未能考虑局部特征和全局特征，故效果欠佳。EfficientNet[29] 网络不能关注重要的特征区域，易受外界因素的干扰。Multi-ResNet 可以提取到更加细致的特征，但泛化能力差，效果不稳定。SSCNet 网络可增强空间特征和局部特征的关联性，加强对重点特征的关注，能够更加有效地利用关键特征，例如，SSCNet 网络能够注意到鸟的嘴巴和腿部、飞机的机翼等细节特征。相对于其它网络，SSCNet 网络能够更加细致、全面地提取到图像的特性，增加网络分类结果的可信度。

4 结 论

针对图像分类时网络模型对全局上下文信息感知能力较弱、疏忽特征的空间位置关系、无法关注图像关键区域和难以有效提取细微特征，导致模型泛化能力较差的问题，本文提出空间位置矫正的稀疏特征图像分类网络 SSCNet。本文主要贡献如下： 1) 提出 SSEF 模块，增强特征表达能力、减少参数数量、提升计算效率和强化通道注意力； 2) 提出 SPCS 模块，通过调整不同位置的权重，增强重要区域的特征提取，加强特征间的空间关系，对空间位置进行矫正； 3) 提出 APM 模块加入在残差网络中减少过拟合、增强全局感受野和增强鲁棒性。SSCNet 网络能够结合全局和局部信息，在减少参数量的同时提取更加丰富细致的特征，能够提高分类准确率并增强泛化能力。在 CIFAR-10、 CIFAR-100、 SVHN、 Imagenette、Imagewoof 数据集上的实验表明，相比于其它先进网络模型，SSCNet 网络在不同数据集上取得了更优的效果，对于大量图像，增强了对重点特征的提取并减少冗余信息，注重特征的全局空间位置，提高了网络性能。然而，改进的网络在参数量方面仍具有进步空间。

参考文献

[1] Yang H, Li J. Label contrastive learning for image classification[J]. *Soft Comput*, 2023, **27**(18): 13477−13486.

[2] 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.

[3] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proc IEEE*, 1998, **86**(11): 2278−2324.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Commun ACM*, 2017, **60**(6): 84−90.

[5] 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.

[6] Xu S J, Jing Y, Li H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. *Opto-Electron Eng*, 2023, **50**(7): 230052.
徐胜军, 荆扬, 李海涛, 等. 渐进式多粒度ResNet车型识别网络[J]. 光电工程, 2023, **50**(7): 230052.

[7] Wang J, Yang Q P, Yang S Q, et al. Dual-path processing network for high-resolution salient object detection[J]. *Appl Intell*, 2022, **52**(10): 12034−12048.

[8] Xue T, Hong Y. IX-ResNet: fragmented multi-scale feature fusion for image classification[J]. *Multimed Tools Appl*, 2021, **80**(18): 27855−27865.

[9] Jiang Z W, Ma Z J, Wang Y N, et al. Aggregated decentralized down-sampling-based ResNet for smart healthcare systems[J]. *Neural Comput Appl*, 2023, **35**(20): 14653−14665.

[10] 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.

[11] Jafar A, Lee M. High-speed hyperparameter optimization for deep ResNet models in image recognition[J]. *Cluster Comput*, 2023, **26**(5): 2605−2613.

[12] Chen L, Zhang J L, Peng H, et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, **50**(4): 220232.
陈龙, 张建林, 彭昊, 等. 多尺度注意力与领域自适应的小样本图像识别[J]. 光电工程, 2023, **50**(4): 220232.

[13] Liang L M, Jin J X, Feng Y, et al. Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction[J]. *Opto-Electron Eng*, 2024, **51**(1): 230276.
梁礼明, 金家新, 冯耀, 等. 融合坐标感知与混合提取的视网膜病变分级算法[J]. 光电工程, 2024, **51**(1): 230276.

[14] Ye Y C, Chen Y. Single image rain removal based on cross scale attention fusion[J]. *Opto-Electron Eng*, 2023, **50**(10): 230191.
叶宇超, 陈莹. 跨尺度注意力融合的单幅图像去雨[J]. 光电工程, 2023, **50**(10): 230191.

[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] Ying Y, Zhang N B, Shan P, et al. PSigmoid: improving squeeze-and-excitation block with parametric sigmoid[J]. *Appl Intell*, 2021, **51**(10): 7427−7439.

[17] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.

[18] Ji Q W, Yu B, Yang Z W, et al. LAM: lightweight attention module[C]//*15th International Conference on Knowledge Science, Engineering and Management*, Singapore, 2022: 485–497. https://doi.org/10.1007/978-3-031-10986-7_39.

[19] Zhong H M, Han T T, Xia W, et al. Research on real-time teachers' facial expression recognition based on YOLOv5 and attention mechanisms[J]. *EURASIP J Adv Signal Process*,

2023, **2023**(1): 55.

[20] Qi F, Wang Y L, Tang Z. Lightweight plant disease classification combining GrabCut algorithm, new coordinate attention, and channel pruning[J]. *Neural Process Lett*, 2022, **54**(6): 5317−5331.

[21] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//*32nd International Conference on Machine Learning*, Lille, 2015: 448–456.

[22] Abdi M, Nahavandi S. Multi-residual networks: improving the speed and accuracy of residual networks[Z]. arXiv: 1609.05672, 2016. https://doi.org/10.48550/arXiv.1609.05672.

[23] Ma C X, Wu J B, Si C Y, et al. Scaling supervised local learning with augmented auxiliary networks[Z]. arXiv: 2402.17318, 2024. https://doi.org/10.48550/arXiv.2402.17318.

[24] Wu X D, Gao S Q, Zhang Z Y, et al. Auto-train-once: controller network guided automatic network pruning from scratch[Z]. arXiv: 2403.14729, 2024. https://doi.org/10.48550/arXiv.2403.14729.

[25] 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.

[26] Lan H, Wang X H, Shen H, et al. Couplformer: rethinking vision transformer with coupling attention[C]//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 2023: 6464–6473. https://doi.org/10.1109/WACV56688.2023.00641.

[27] Zhou C L, Zhang H, Zhou Z K, et al. QKFormer: hierarchical spiking transformer using Q-K attention[Z]. arXiv: 2403.16552, 2024. https://doi.org/10.48550/arXiv.2403.16552.

[28] 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*, Vancouver, 2024: 14991–14999. https://doi.org/10.1609/aaai.v38i13.29420.

[29] Tan M X, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks[C]//*36th International Conference on Machine Learning*, Long Beach, 2019: 6105–6114.

作者简介

姜文涛 (19896-)，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

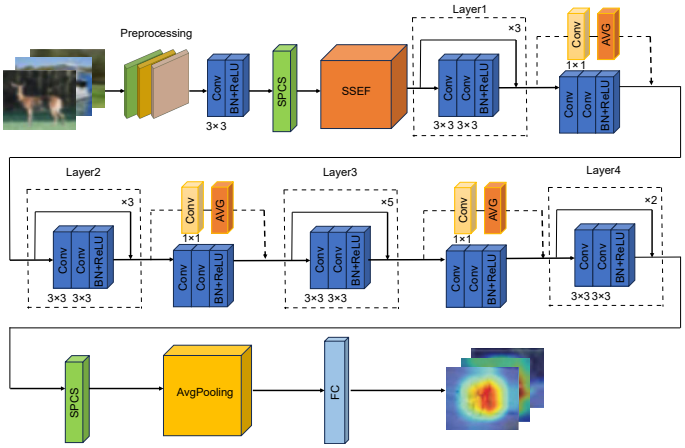
【通信作者】陈晨 (2000-)，男，硕士研究生，主要研究方向为图像处理、模式识别、人工智能。

E-mail:867428188@qq.com

扫描二维码，获取PDF全文

Sparse feature image classification network with spatial position correction

Jiang Wentao[1], Chen Chen[1*], Zhang Shengchong[2]



SSCNet network structure

Overview: To sparse semantics and enhance attention to key features, enhance the correlation between spatial and local features, and constrain the spatial position of features, this paper proposes a Sparse Feature Image Classification Network with Spatial Position Correction (SSCNet) for spatial position correction. Firstly, a Sparse Semantic Enhanced Feature Module (SSEF) module is proposed, which combines Depth Separable Convolution (DSC) and SE (Squeeze and Excitation) modules to enhance feature extraction ability while maintaining spatial information integrity; Then, the Spatial Position Correction Symmetric Attention Mechanism (SPCS) is proposed. SPCS adds the symmetric coordinate attention mechanism to specific positions in the network, which can strengthen the spatial relationships between features, constrain and correct their spatial positions, and enhance the network's perception of global detailed features; Finally, the Average Pooling Module (APM) is proposed and applied to each residual branch of the network, enabling the network to more effectively capture global feature information, enhance feature translation invariance, delay network overfitting, and improve network generalization ability. This article used CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewood datasets as experimental datasets. The CIFAR-10 dataset contains a total of 60000 color images from 10 categories, each with a resolution of 32×32 pixels. This dataset is commonly used to test and compare the performance of image classification algorithms. The CIFAR-100 dataset is more challenging and used to evaluate model performance for finer grained image classification tasks. The SVHN dataset contains real-world street view house number images, which contain digital images from Google Street View images used to recognize numbers on house signs. The images in the SVHN dataset are divided into training, testing, and additional training sets, each containing one or more numbers, and the resolution of the images is also higher than that of the CIFAR dataset. The Imagenette and Imagewof datasets are small scale subsets extracted from ImageNet, which have been streamlined and adjusted based on the ImageNet dataset. This article compares the network model with 12 other network models on 5 datasets. In the CIFAR-10, CIFAR-100, SVHN, Imagenette, and Imagewood datasets, the classification accuracy of SSCNet is 96.72%, 80.63%, 97.43%, 88.75%, and 82.09%. Compared with other methods, SSCNet in this paper can better extract local detail information while balancing global information, and has higher classification accuracy and strong generalization performance.

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; DOI: 10.12086/oee.2024.240050

Foundation item: Project supported by National Natural Science Foundation of China (61172144), Liaoning Provincial Natural Science Foundation of China (20170540426), and Key Fund of Liaoning Provincial Department of Education (LJYL049)

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

* E-mail: 867428188@qq.com