

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 局部注意力引导下的全局池化残差分类网络

姜文涛, 董睿, 张晟翀

#### 引用本文:

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.

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.

<https://doi.org/10.12086/oe.2024.240126>

收稿日期: 2024-05-28; 修改日期: 2024-08-03; 录用日期: 2024-08-05

### 相关论文

#### 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

*光电工程* 2024, 51(5): 240050 doi: [10.12086/oe.2024.240050](https://doi.org/10.12086/oe.2024.240050)

#### 基于BiLevelNet的实时语义分割算法

吴马靖, 张永爱, 林珊玲, 林志贤, 林坚普

*光电工程* 2024, 51(5): 240030 doi: [10.12086/oe.2024.240030](https://doi.org/10.12086/oe.2024.240030)

#### 融合多分辨率特征的点云分类与分割网络

陶志勇, 李衡, 豆淼森, 林森

*光电工程* 2023, 50(10): 230166 doi: [10.12086/oe.2023.230166](https://doi.org/10.12086/oe.2023.230166)

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



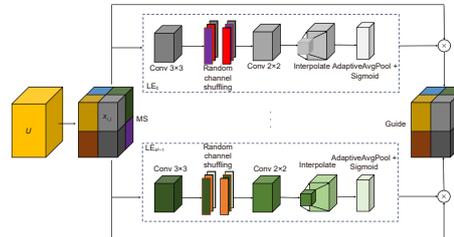
 OE\_Journal



Website

DOI: 10.12086/oe.2024.240126

# 局部注意力引导下的全局池化残差分类网络

姜文涛<sup>1</sup>, 董睿<sup>1\*</sup>, 张晟翀<sup>2</sup><sup>1</sup> 辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105;<sup>2</sup> 光电信息控制和安全技术重点实验室, 天津 300308

**摘要:** 大部分注意力机制虽然能增强图像特征, 但没有考虑局部特征的关联性影响特征整体的问题。针对以上问题, 本文提出局部注意力引导下的全局池化残差分类网络 (MSLENet)。MSLENet 的基线网络为 ResNet34, 首先改变首层结构, 保留图像重要信息; 其次提出多分割局部增强注意力机制 (MSLE) 模块, MSLE 模块将图像整体分割成多个小图像, 增强每个小图像的局部特征, 通过特征组交互的方式将局部重要特征引导到全局特征中; 最后提出池化残差 (PR) 模块来处理 ResNet 残差结构丢失信息的问题, 提高各层之间的信息利用率。实验结果表明, MSLENet 通过增强局部特征的关联性, 在多个数据集上均有良好的效果, 有效地提高了网络的表达能力。

**关键词:** 图像分类; 注意力机制; 残差结构; 局部特征; 全局特征; 关联性

**中图分类号:** TP391

**文献标志码:** A

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

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

## Global pooling residual classification network guided by local attention

Jiang Wentao<sup>1</sup>, Dong Rui<sup>1\*</sup>, Zhang Shengchong<sup>2</sup><sup>1</sup> College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;<sup>2</sup> Science and Technology on Electro-Optical Information Security Control Laboratory, Tianjin 300308, China

**Abstract:** Most attention mechanisms, while enhancing image features, do not consider the impact of local feature interaction on overall feature representation. To address this issue, this paper proposes a global pooling residual classification network guided by local attention (MSLENet). The baseline network for MSLENet was ResNet34. First, the initial layer structure was modified to retain important image information. Second, a multiple segmentation local enhancement attention mechanism (MSLE) module was introduced. The MSLE module first segmented the image into multiple small images, then enhanced the local features of each small image, and finally integrated these important local features into the global features through feature group interaction. Lastly, a pooling residual (PR) module was proposed to address the information loss problem in the ResNet residual structure and improve the information utilization between layers. The experimental results show that by enhancing the interaction of local

收稿日期: 2024-05-28; 修回日期: 2024-08-03; 录用日期: 2024-08-05

基金项目: 国防预研基金资助项目 (172068); 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: 董睿, 2669213806@qq.com。

版权所有©2024 中国科学院光电技术研究所

features, MSLENet achieves good performance on multiple datasets and effectively improves the expressive ability of the network.

**Keywords:** image classification; attention mechanism; residual structure; local features; global features; interaction

## 1 引言

图像分类在计算机视觉任务中扮演着非常重要的角色, 是计算机视觉任务中重点研究对象, 它的作用域广泛, 比如医疗影像中对癌症的分类、自动驾驶中快速识别目标物体、统计小物品种类数量等。图像分类算法大多数是基于机器学习算法进行研究, 在机器学习算法中, 两种算法最为常见, 一种是传统的机器学习算法, 例如 K 邻近算法 (K-nearest neighbor, KNN)、朴素贝叶斯算法 (Naive Bayes, NB)、回归算法等, 这些算法容易理解、操作简单、算法轻量, 但是它们泛化能力相对较弱, 表达能力较差, 对于现实世界的不同情况不能准确把握; 第二种是深度学习算法, 它通过神经网络的多神经元排列组合而成, 结合梯度下降算法, 从而使网络结构具有学习能力<sup>[1-2]</sup>。

在深度学习算法中, 运用最广泛的网络结构为卷积神经网络 (CNNs), 在卷积神经网络的发展过程中 Lecun 等<sup>[3]</sup> 发布 LeNet 了网络, 它通过五层网络结构拨开了卷积神经网络的面纱, 限制于当时的硬件条件, LeNet 并没有被很好地发掘, 直到 2012 年 Krizhevsky 等<sup>[4]</sup> 人提出了 AlexNet, 并且首次在 GPU 上面运行模型, 加深了网络结构, 添加 ReLU 激活函数, 由于它良好的表达能力, 使得卷积神经网络有了新的突破。随后 Simonyan 等<sup>[5]</sup> 提出 VGG, 通过扩展通道数使模型具有更好的泛化能力。Szegedy 等<sup>[6]</sup> 提出 GoogLeNet, 它不仅提高了网络宽度, 同时加深了网络深度, 使模型的表达能力更强。但是实验表明, 更深的网络层次并不能带来更好的表达效果, 会出现过拟合, 梯度消失等问题。2015 年由 He 等<sup>[7]</sup> 提出的 ResNet, 成功地改善了上述问题, 它通过引入残差结构, 从而有效地加深了网络层数, 在各种视觉任务中均具有非常好的表现, 但是 ResNet 仍然会存在部分信息丢失的问题。

为了解决部分信息丢失的问题, Huang 等<sup>[8]</sup> 提出 DenseNet, 它使用多分支的残差连接保留上下文信息的方法。Abdi 等<sup>[9]</sup> 提出 Multi-ResNet, 提高每个位置的残差数量来保证信息的传递。但是以上网络在处理图像问题时, 仅仅从保存各网络层之间信息的角度去

提高模型的性能, 但是却忽略了信息之间的冗余和信息无用性, 为了使每层输出的特征能够得到更好的体现, 同时使网络更轻量化, Howar 等<sup>[10]</sup> 提出 MobileNet, 它改变原有卷积的方式, 使用深度卷积和点卷积使得网络变得轻量化, 并且在一些移动设备上具有很好的效果。Zhang 等<sup>[11]</sup> 提出 ShuffleNet, 它讨论了组卷积和点卷积的优缺点, 提出使用通道洗牌 (shuffle channel) 的方式来处理组卷积带来的通道交互受损和点卷积带来的计算复杂度提高的问题。但是, 以上网络只是不断地积累网络层从而优化特征提取, 却忽视了特征信息的利弊, 以及不同特征信息对网络层的影响。

为了解决上述问题, 人们从生物的视觉系统中得到启发, 提出注意力机制。Google DeepMind<sup>[12]</sup> 团队提出了空间域注意力机制 (spatial transformer networks), 它通过对图像的旋转、裁剪、伸缩等变化挑选特征图像中的重要信息, 使网络表达能力变得更强。Hu 等<sup>[13]</sup> 提出了 SENet, 通过通道之间的交互关系, 从而增强图像中的特征表达。百度和 Sydney<sup>[14]</sup> 提出了 GCTnet, 它通过门控的方式来改善通道之间的竞争和协作关系, 使注意力机制更轻量化。CA 注意力机制使用坐标注意力和坐标信息嵌入到网络中, ECA 则使用通道注意力机制提升重要信息的关注度, Zhang 等<sup>[15]</sup> 改进 Transformer 模型, 使用深度卷积提取块与块之间的信息。是上述模型均从图像通道去考虑特征信息, 只把握通道特征之间的关系, 并没有考虑特征图像局部信息之间的关系, 上述问题可能会导致图像的一部分细节信息被忽略, 从而使模型的表达能力减弱, 为了表达局部空间特征的交互关系, 使网络更有效地观察到图像的重要特征, 本文提出了局部注意力引导下的全局池化残差分类网络 (global pooling residual classification network guided by local attention, MSLENet), 它能关注图像空间的局部重要信息, 并提升局部信息之间的关联性, 最后将这些信息引导并整合到整体的信息当中。

MSLENet 的主要模块为多分割局部增强注意力机制 (multiple segmentation local enhancement attention

mechanism) 模块, 简称为 MSLE 模块, MSLE 模块包含三个部分, 分别称为多分割 (multiple segment, MS) 模块、LE (local enhancement) 模块、Guide 模块: 1) MS 模块的作用是将图像按照行列均分成具有不同特征的同尺寸小图像, 从而凸显分割后的特征信息来提高网络的表达能力; 2) LE 模块的作用是增强局部特征的重要信息, 改善局部信息之间的关系, 增强有利信息的传递; 3) Guide 模块将经过 LE 模块的特征拼接在一起形成特征组, 增强特征组的关联性, 同时将局部信息引导入全局信息当中。MSLENet 以 ResNet34 为基线网络, 首先改变网络首层结构, 提高图像信息的完整性, 然后改变残差结构, 提出池化残差 (pooling residual, PR) 模块, 最后加入 MSLE 模块增强重要信息的利用率, 从而提高模型的表达能力。

本文的主要贡献如下: 1) 提出 MSLE 模块, MSLE 通过 MS 模块和 LE 模块处理局部和局部以及局部和整体之间的信息关系, 提高重要特征的利用率, 然后使用 Guide 模块提升局部信息的关联性, 引导局部重要信息到全局特征中, 从而增强网络的表达能力。2) 提出 SCFnet, SCFnet 以 ResNet34 为基线网络, 改变首层网络结构, 同时提出 PR 模块, 增强信息的完整性, 提高信息的利用率。3) MSLE 模块从多角度处理了空间维度的局部特征之间的关系, 增强局部与局部之间的特征关系, 从局部关系考虑整体特征与局部之间的联系, 从而增强网络对重要信息的利用率。

## 2 理论推导

MSLENet 以 ResNet34 为基线网络, 在网络的主干部分, 首先修改 ResNet34 输入层的结构; 然后改变 ResNet34 的残差部分, 防止信息丢失; 最后加入 MSLE 模块进行特征提取, 提高信息利用率。

在 MSLE 模块中, 首先提出 MS 模块, MS 模块通过分割函数将每个特征矩阵分割成多个特征小块, 分成的小块进入 LE 模块做特征提取; LE 通过卷积、随机通道打乱机制 (random channel shuffle)、上采样操作来促进局部特征的表达, 增强局部特征之间的联系; 最后通过 Guide 模块将每个特征块拼接在一起, 实现局部与全局特征之间的引导交互, 下文将详细介绍每个模块。

### 2.1 调整网络输入层

在基线网络 ResNet34 中, 为了迎合大尺寸图片

的特征映射, 强化后续网络层的感受野, 采用步长 2, 填充为 3 的 7×7 的卷积, 并且采用 3×3 的最大池化操作进一步缩小图像尺寸来增强特征。但是对于尺寸较小的图片, 过早地使用大卷积核来提取特征会造成图片部分信息的丢失, 所以 MSLENet 将卷积核尺寸改为 3×3, 步长为 1, 同时减少填充的大小, 填充改为 1, 并且删除最大池化层, 从更细节的角度提取输入层输入图像的特征信息, 保证原有图像信息的完整性, 防止信息进一步的丢失。通过卷积后图像尺寸可以表示为

$$N = (W - K + 2P) / (S + 1), \quad (1)$$

其中:  $N$  表示为卷积后输出图像的尺寸,  $W$  表示卷积前输入图像的尺寸,  $K$  表示卷积核的尺寸,  $P$  表示填充的大小,  $S$  表示卷积过程中窗口移动的步长。假设输入图像尺寸  $W$  为  $V_{10 \times 10}$ , 步长  $K$  等于 2, 填充大小  $P$  等于 3, 求得输出图像尺寸  $N$  为  $3 \times 3$ 。如果通过改变卷积核尺寸  $7 \times 7$  变为  $3 \times 3$ , 降低步长和填充, 那么输出图像尺寸  $N$  为  $10 \times 10$ , 可以得出, 如果过早地进行大尺寸卷积, 图像会忽视掉大部分信息, 从而阻碍了后续层去提取重要信息的过程。在 ResNet34 中过早地使用最大池化操作, 会进一步影响输入图像的信息, 最大池化操作忽略了其它权值对图像特征的影响, 导致模型出现表达能力受限、空间感知力下降的问题, 最大池化后图像尺寸公式可以表达为

$$H = (W - K + 2P) / (S + 1), \quad (2)$$

其中:  $H$  表示为最大池化后输出图像的尺寸,  $W$  表示最大池化前输入图像的尺寸,  $K$  表示最大池化窗口的尺寸,  $P$  表示最大池化中填充的大小,  $S$  表示最大池化过程中窗口移动的步长, 如果将  $K$  设置成 3,  $S$  等于 2,  $P$  等于 1, 那么输出图像尺寸  $H$  会减半。所以, 网络如果过早地使用最大池化会进一步衰减图像的信息, 导致模型整体泛化能力变弱。对于输入的图像来说, 删除最大池化保留原图像的特征能够增加网络提取图像特征的能力。

### 2.2 池化残差模块

ResNet 通过残差结构有效地防止过拟合、梯度消失的问题, 从而加深了网络的层数, 提升了网络的表达能力。在 ResNet 残差结构中, 为了保证上一层输出的维度大小和残差块输出的维度大小相同, 使用步长为 2 的  $1 \times 1$  卷积来控制通过残差块的特征矩阵的维度, 以此来保证残差块主干的输出和残差块中残差结构的输出尺寸相同, 但是  $1 \times 1$  步长为 2 的卷积操作

会造成一半的信息被忽略掉, 这些信息不能被确定为无用的信息, 为了改善上述信息丢失的问题, 本文提出池化残差模块, 如果输出尺寸和输入尺寸相同, 那么直接将输出特征和输入特征相加。如果尺寸不同, 将残差结构的卷积步长改变为 1, 以此来最大程度保留上层信息, 但是, 改变步长后会导致上一层输出的特征矩阵尺寸是经过残差结构的特征矩阵尺寸的二倍, 为了保证上述两个输出的尺寸一致, 使用窗口大小为  $2 \times 2$ , 步长为 2 的平均池化, 平均池化更有利于减少信息的冗余。修改后的残差结构如图 1 所示。

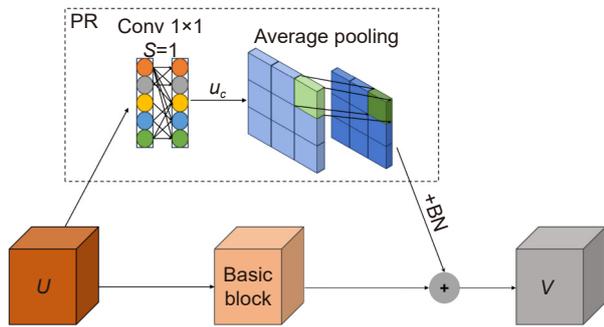


图 1 池化残差结构  
Fig. 1 Pooling residual structure

设上一层的输出为  $U$ ,  $U \in R^{C \times H \times W}$ , 首先  $U$  经过一层  $1 \times 1$  的卷积映射后的可以表示为

$$u_c = Conv2d(U), \quad (3)$$

其中:  $Conv2d(\cdot)$  表示  $1 \times 1$  步长为 1 的卷积,  $u_c$  表示

经过卷积后的特征矩阵,  $u_c \in R^{C \times H \times W}$ , 然后将  $u_c$  经过平均池化, 表示为

$$u'_c = Avgpool(u_c), \quad (4)$$

其中:  $Avgpool(\cdot)$  表示  $2 \times 2$  步长为 2 的平均池化操作,  $u'_c \in R^{C \times \frac{H \times W}{4}}$ , 通过平均池化操作使得输出  $U$  能够保留大部分的信息, 之后通过 BN 层得到  $u''_c$ , 可以表示为

$$V = U_c + u''_c. \quad (5)$$

然后将上一层的输出  $U$  通过 Basic block 基本块后得到  $U_c$ ,  $U_c \in R^{C \times \frac{H \times W}{4}}$ , 最后  $U_c$  与上述中得到的  $u''_c$  相加,  $V \in R^{C \times \frac{H \times W}{4}}$ ,  $V$  也就是 PR 模块与 Basic block 交互的特征矩阵。

PR 模块可以让网络能最大程度保留不同层的信息, 不同层信息互相利用的时候几乎没有造成信息丢失, 同时增加了网络层之间的信息流通性, 从而让特征信息能够在网络层之间相互利用, 增强模型的泛化能力, 并且 PR 模块并没有增强任何参数量, 也几乎没有增加计算的复杂性。

### 2.3 MSLE 模块

为了更好地提升局部重要特征的交互性, 利用局部有效信息引导增强全局有效信息, 本文提出多分割局部增强注意力机制 (MSLE) 模块, MSLE 模块如图 2 所示。首先将输入特征送入 MS 模块, MS 模块

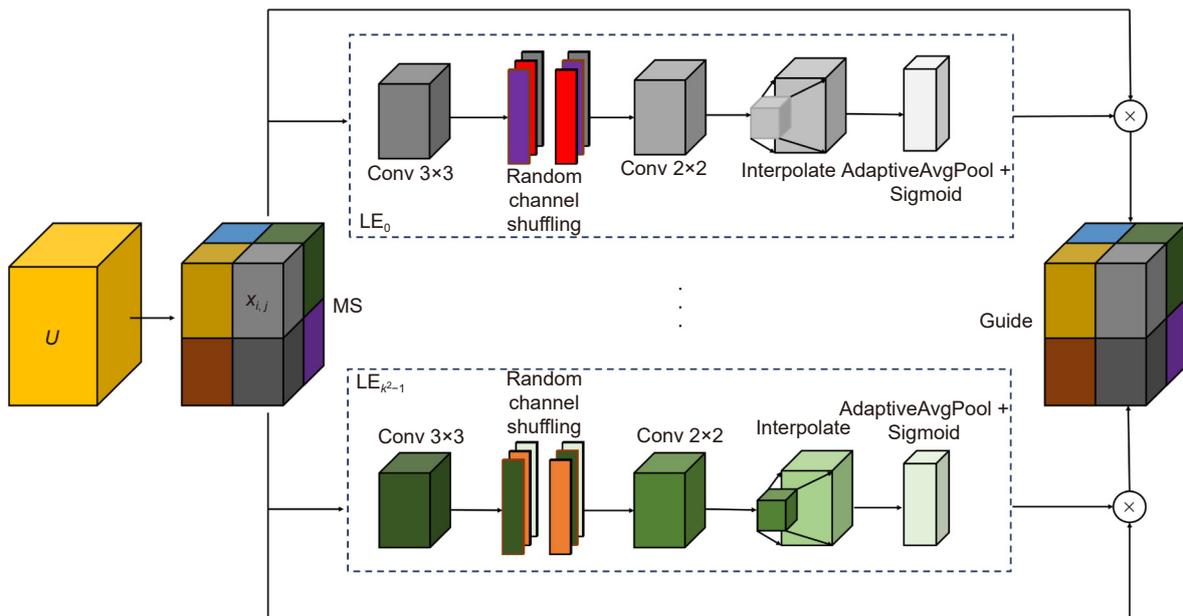


图 2 MSLE 结构图  
Fig. 2 MSLE structure diagram

通过超参数  $k$  进行空间分割, 分割成  $k^2$  个特征小块, 实现局部特征之间的交互; 然后将每个特征小块送入 LE 模块, 它通过增强扩大局部信息来表达全局特征的有效信息; 最后将 LE 模块得到的小特征矩阵送入 Guide 模块, Guide 模块将局部特征整合在一起, 增加上述小特征矩阵的关联性, 从而实现局部引导全局有利信息的表达。以下会详细介绍 MSLE 模块的细节部分。

### 2.3.1 MS 模块

MS 模块作用是将每一层得到的特征矩阵在空间上做均匀分割, 通过分割函数把特征矩阵分割成多个小的特征矩阵, 分割前后示意图如图 3 所示, 左边为分割前图像, 右边为分割后图像。首先设分割前的特征矩阵为  $X \in R^{C \times H \times W}$ , 并设置分割超参数为  $k$ ,  $k \in (N^+ \cap [0, H-1])$ , 其中  $N^+$  代表正整数集合,  $[0, H-1]$  表示参数  $k$  的取值在 0 到  $H-1$  之间, 包括 0 和  $H-1$ 。  $H$  表示分割前特征矩阵的行数, 然后设行分割函数为  $F_H(\cdot)$ , 列分割函数为  $F_W(\cdot)$ ,  $F_H(\cdot)$  与  $F_W(\cdot)$  分别为在  $dim = 2$  和  $dim = 3$  的 Chunk 函数, 由于 MS 模块是均匀分割, 所以分割前后的公式为

$$x_{i,j} = F_W(F_H(X)), \quad (6)$$

其中:  $x_{i,j}$  表示分割后在  $i$  行  $j$  列的特征矩阵小块,  $i$  表示小特征矩阵  $x$  在横坐标 (行) 轴的第  $i$  个位置,  $j$  表示小特征矩阵  $x$  在竖坐标轴 (列) 的第  $j$  个位置, 每个  $x_{i,j}$  的尺寸相同, 根据上述表达得出:

$$x_{i,j} \in R^{C \times \frac{H \times W}{k^2}}. \quad (7)$$

由式 (6) 和式 (7) 联合最终得出  $i, j$  的区间范围:

$$i, j \in (N^+ \cap [0, k-1]). \quad (8)$$

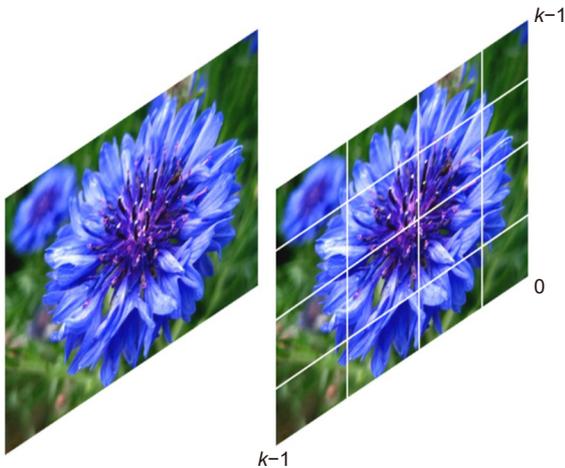


图 3 分割前后示意图

Fig. 3 Schematic diagram before and after segmentation

最后, 通过 MS 模块后得到的特征矩阵小块  $x_{i,j}$  的总数为  $k^2$  个。MS 通过行列分割的方式使后续的网络层更能凸显图像的局部特征, 也为 LE 模块增强局部重要特征提供基础。

### 2.3.2 LE 模块

为了提高局部特征之间的联系, 增强局部特征的重要信息, 本文提出 LE 模块。将通过 MS 模块分割后得到的小特征矩阵  $x_{i,j}$  送入到 LE 模块中, 由于  $x_{i,j}$  的总数为  $k^2$  个, 首先将每一个  $x_{i,j}$  按顺序送入到一个列表中, 得到  $x_m = x_{i,j}$ , 其中  $m$  为送入列表后每个  $x_{i,j}$  在列表中的位置, 即表示为

$$x_m \in [x_{0,1}, x_{0,2}, x_{0,3} \cdots x_{i,j}, x_{i,j+1} \cdots x_{k-1,k-2}, x_{k-1,k-1}], \quad (9)$$

其中:  $m \in [0, 1 \cdots k^2 - 1]$ , 然后将  $k^2$  个  $x_m$  放入  $k^2$  个 LE 模块中, 进行局部的最优特征提取。

为了方便表达 LE 模块, 将 LE 模块分为三部分, 第一部分称为特征提取部分, 将其设为  $F_{s1}(\cdot)$ , 作用是增强局部特征。第二部分称为扩大特征部分, 设为  $F_{s2}(\cdot)$ , 作用是扩大第一部分提取出的特征信息, 增强局部特征之间的关联性。第三部分是特征表达部分, 设为  $F_{s3}(\cdot)$ , 它通过池化和激活函数来实现对局部重要特征的表达。

1) 特征提取部分。为了清晰表达特征提取部分的过程, 特征提取结构图如图 4 所示。

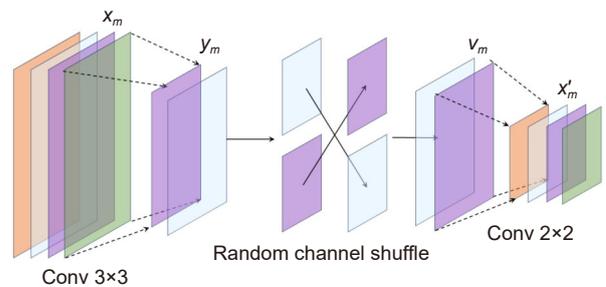


图 4 特征提取结构图

Fig. 4 Feature extraction structure diagram

特征提取部分首先使用  $3 \times 3$ , 步长为 1 的卷积,  $3 \times 3$  的卷积更有利于小特征的表达, 因为  $3 \times 3$  的卷积可以通过更小的权重矩阵来获取输入特征更大的感受野, 同时在卷积的过程中使用超参数  $l$  来压缩通道数, 提升通道间的交互。表达方式为

$$y_m = Conv2d(x_m), \quad (10)$$

其中:  $Conv2d(\cdot)$  表示卷积操作,  $y_m$  表示经过卷积后得到的特征矩阵, 所以压缩通道数后,  $y_m \in \frac{C}{l} \times$

$\frac{(H-2) \times (W-2)}{k^2}$ 。为了提升通道之间的竞争和交互关系, 使用随机通道打乱机制 (random channel shuffle), 其作用是使通道之间能够相互合作, 相互竞争, 从而增强不同通道之间的交互性, 有利于通道之间的信息流通。因为  $y_m$  的通道数变为  $C/l$ , 所以  $y_m$  按照通道数  $C/l$  来分割, 表示为

$$z_n = \frac{y_m}{C/l}, \quad (11)$$

其中:  $z_n$  表示通道分割后每一个特征矩阵, 其中  $n \in [0, 1, \dots, \frac{C}{l} - 1]$ ,  $z_n$  的通道数变为 1, 随后将所有分割后特征矩阵  $[z_0, z_1 \dots z_{C/l-1}]$  进行随机打乱, 例如将  $[z_0, z_1 \dots z_{C/l-1}]$  变为  $[z_2, z_{C/l-1}, z_{C/l-3} \dots z_1]$ , 打乱后对所有的特征矩阵进行通道维度的 *Concat* 拼接, 从而得到最后的结果  $y_m$ , 所以随机通道打乱机制可以表示为

$$v_m = f_{rs}(z_n) = \text{Concat}(\text{Random}_{\text{dim}=1}(z_n)), \quad (12)$$

其中:  $\text{Random}_{\text{dim}=1}(\cdot)$  表示在通道维度上的随机打乱,  $v_m$  的尺寸为  $\frac{C}{l} \times \frac{(H-2) \times (W-2)}{k^2}$ ,  $f_{rs}(z_n)$  表示随机通道打乱机制表达式。通过随机打乱通道机制后, 最后使用  $2 \times 2$ , 步长为 1 的卷积来进一步提取图像的特征,  $2 \times 2$  的卷积通过细节捕捉的方式来捕捉到输入图像的特征信息, 从而使网络能够抓取大部分细节信息, 其表达式为

$$x'_m = \text{Conv2d}(v_m), \quad (13)$$

其中:  $x'_m$  表示经过特征提取部分得到的特征矩阵, 其尺寸为  $\frac{C}{l} \times \frac{(H-3) \times (W-3)}{k^2} \times l$ 。所以特征提取部分整体过程表示为

$$x'_m = F_{s1}(x_m), \quad (14)$$

$$F_{s1}(x_m) = \text{Conv2d}(f_{rs}(\text{Conv2d}(x_m))). \quad (15)$$

从整体上看, 特征提取部分之所以用小卷积核, 是因为在 MS 模块已经将图像通过参数  $k$  进行分割, 分割后的特征矩阵尺寸相对较小, 为了表达这些小特征矩阵  $x_m$  的有利因子, 使用尺寸小的卷积核来进行特征提取, 从而使每个  $x_m$  都能表达它们的细节信息。在  $3 \times 3$  卷积和  $2 \times 2$  的卷积之后, 本文都引入了 Tanh 激活函数, Tanh 激活函数公式表示为

$$T(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (16)$$

首先 Tanh 能够更好地保存输入特征的信息, 更好地解决了梯度爆炸的问题; 然后对于小特征的表达, Tanh 能够更好地增强小特征矩阵之间的非线性关系;

最后, Tanh 会将输入矩阵中的元素值控制在  $-1$  到  $1$  之间, 使得矩阵元素更平滑。

2) 特征扩大部分。为了使特征提取部分提取的关键特征在空间范围内扩大, 本文使用最邻近插值上采样算法, 最邻近插值上采样操作图像如图 5 所示。

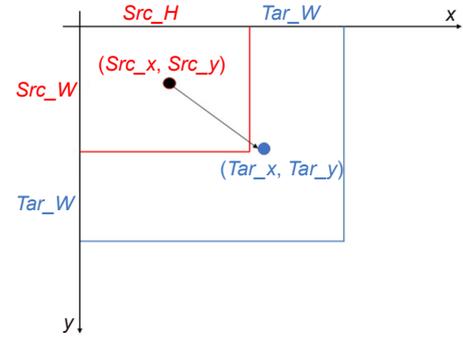


图 5 最邻近插值上采样操作图像

Fig. 5 Nearest neighbor interpolation upsampling operation image

图 5 中 *Src* 表示原特征矩阵, *Tar* 表示经过最邻近插值上采样算法后的目标矩阵, *Src\_H*、*Src\_W*、*Tar\_H*、*Tar\_W* 分别表示原特征矩阵的高度和宽度以及目标特征矩阵的高度和宽度, 从图中可以看出, *Src* 中的某个元素 (*Src\_x*, *Src\_y*) 经过上采样后坐标被放大到更远的位置 (*Tar\_x*, *Tar\_y*), 所以不论是原坐标还是目标坐标都是特征矩阵经过上采样操作后的坐标, 两者具有相同的坐标值, 所以原特征矩阵的元素值被放大成了两倍, 通过上述过程来放大特征提取部分提取的重要特征。将特征提取部分得到的  $x'_m$  经过最邻近插值上采样的表达式为

$$x''_m = F_{s2}(x'_m) = \text{interpolate}_{\text{nearest}}(x'_m), \quad (17)$$

其中:  $x''_m$  为经过上采样函数得到的特征矩阵, 其尺寸  $x''_m \in R^{C \times \frac{H \times W}{k^2}}$ ; *interpolate* 为上采样函数, *nearest* 为上采样函数的最邻近插值方式。通过上采样函数来扩大特征矩阵的尺寸, 使得与分割后的特征矩阵的尺寸相同, 同时使用 *nearest* 插值方式, 来扩大特征提取部分获取重要特征信息, 从而使得分割后的特征矩阵在空间上充分利用有益信息, 移除网络中不重要的边缘信息, 增强每个  $x_{i,j}$  的特征元素之间的联系, 从而实现局部信息之间的交流协作。

3) 特征表达部分。特征表达部分总体表示为

$$x'''_m = F_{s3}(x''_m) = \text{Sigmoid}(F_{\text{avg}}(x''_m)) \cdot x''_m, \quad (18)$$

其中:  $x'''_m$  表示特征表达部分输出的特征矩阵, 第三部分具体过程如下, 首先将每一个  $x''_m \in R^{C \times \frac{H \times W}{k^2}}$  通过全局平均池化操作, 使用  $p, q$  来表示  $x''_m$  特征矩阵中每

个元素的位置, 全局平均池化操作可以表示成:

$$y_{\text{avg}} = F_{\text{avg}}(x_m'') = \frac{1}{(H/k) \times (W/k)} \sum_p^k \sum_q^k x_m''(p, q), \quad (19)$$

其中:  $y_{\text{avg}}$  表示  $x_m''$  通过全局平均池化后的特征矩阵,  $y_{\text{avg}} \in R^{C \times 1 \times 1}$ ,  $x_m''(p, q)$  表示  $x_m''$  中在  $(p, q)$  位置的元素, 使用全局平均池化可以将每个通道的特征描述成一种集合, 这些局部描述的统计量可以表示每个  $x_m''$  的通道特征, 从而体现通道间的特征映射关系。为了使局部上下文信息能够更好的嵌入, 最后使用 Sigmoid 函数来体现这些信息, 通过 Sigmoid 函数输出的特征与最初进入 LE 层的小特征矩阵相乘, 实现对局部特征的表达, 可以表示成:

$$x_m''' = S(y_{\text{avg}}) = \frac{x_m''}{(1 + e^{-y_{\text{avg}}})}, \quad (20)$$

其中:  $x_m'''$  是通过第三部分的输出的特征矩阵,  $x_m''' \in R^{C \times \frac{H \times W}{k^2}}$ ,  $S(\cdot)$  是 Sigmoid 函数和原特征矩阵  $x_m''$  构造的特征表达函数。通过 Sigmoid 函数将得到的特征集合  $y_{\text{avg}}$  控制在  $[0, 1]$  的范围内, 可以将线性变化的输入值映射到一个概率密度分布上, 并且它的平滑性能能够有自觉地减轻梯度消失的问题, 然后输入和输出相乘, 实现对局部特征的注意力机制, 从而获取局部特征的重要信息, 让网络能够通过局部之间的关系来增强网络表达能力。

### 2.3.3 Guide 模块

为了将局部信息引导到全局中, 使得网络继续通过全局去提取特征, 提出 Guide 模块, 为了更清楚地表达引导特征信息的过程, Guide 模块示意图如图 6 所示。

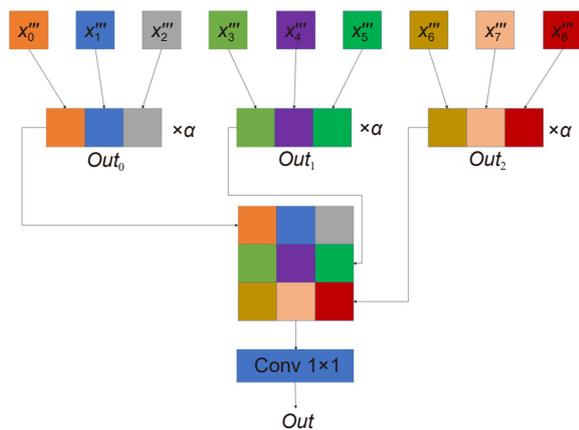


图 6 引导特征信息示意图  
Fig. 6 Guided feature information diagram

首先将 LE 模块得到的  $x_m''$  且属于同一行的  $x_m''$  整合在一起, 使用 Concat 函数来整合, 将整合后的特征

组均乘以可学习参数  $\alpha$  来提高每个特征组中  $x_m''$  间的关联性, 从而实现将局部小特征引导到部分局部特征中, 可表示为

$$out_i = \alpha \cdot \text{Concat}_{\text{dim}=3}([x_i'', x_{i+1}'', \dots, x_{i+k-1}'']), \quad (21)$$

其中:  $out_i$  表示整合后的特征矩阵, 每个  $out_i$  的尺寸可以表示为  $out_i \in R^{C \times H \times (W/k^2)}$ ,  $i$  表示整合后的行数,  $i \in [0, 1, 2, \dots, k^2 - 1]$ ,  $\text{dim} = 3$  表示在列维度整合。然后将  $k^2$  个  $out_i$  合并在一起, 合并过程表示为

$$out = \text{Concat}_{\text{dim}=2}([out_0, out_1, \dots, out_{k^2-1}]), \quad (22)$$

其中:  $out$  表示整合后的特征矩阵, 其尺寸可以表示为  $out \in R^{C \times H \times W}$ ,  $\text{dim} = 2$  表示在行维度整合, 通过两种整合方式, 使  $out$  的尺寸与进入 MSLE 模块的特征矩阵  $X$  尺寸保持一致。最后使用  $1 \times 1$  的卷积再次提升局部特征之间的关系, 使局部特征相互连接进入整体特征当中, 从而实现将局部重要信息引导到整体特征当中。

总体而言, MSLE 模块的本质是通过小特征矩阵在空间维度之间的排列组合和通道维度之间交互来实现局部信息引导全局信息进行表达, 为了更清楚地表达 MSLE 的过程, MSLE 过程可视化示意图如图 7 所示。首先特征图经过 MS 进行分割, 然后进入 LE 模块, LE 模块增强并提取特征中的有利信息, 同时扩大提取的有利信息, 让每个小特征图像的有利信息均能相互交流, 增强其关联性。最后通过 Guide 模块将局部交互的信息引导到全局特征中, 从而实现扩大并增强全局重要信息。在视觉任务中, 输入的图像往往都是随机的、非线性的, 如果只通过整体来提取图像的有利信息, 大部分只能获取到中心信息, 进而减弱了边缘信息的有效性。但是如果从局部就开始关注有效信息, 那么网络就会保留中心有效信息的同时保留边缘有效信息, 网络就更能注意到图像重要信息的位置, 进而来增强模型的表达能力。

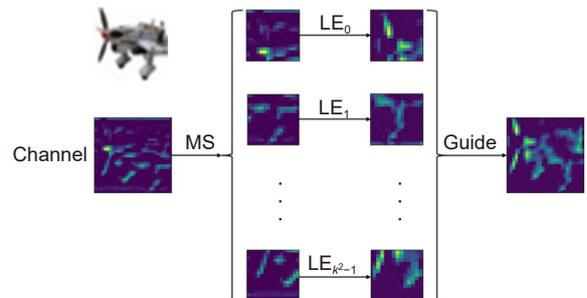


图 7 MSLE 过程可视化示意图  
Fig. 7 Visualization diagram of the MSLE process

## 2.4 总体网络结构

### 2.4.1 网络层总体结构

ResNet 通过跳跃连接的方式, 实现了上一层输出的信息保留到下一层, 使得每层之间的信息能够有效传递, 从而防止了网络层过深导致梯度消失和过拟合的问题。ResNet 常规残差块如图 8(a) 所示, 简称为 block, 主干上面使用  $3 \times 3$  的卷积, 跳跃连接部分使用  $1 \times 1$  的卷积。

融合多分割局部增强注意力机制 (MSLE) 模块的残差块, 简称为 (M-block) 如图 8(b) 所示, 跳跃连接部分与常规残差块 (block) 一致, 主干上半部分继续使用 2 个  $3 \times 3$  的卷积, 中间包含 BN 和 ReLU, 下半部分融合 MSLE 模块, 目的是区分局部特征之间的利弊, 关注局部特征在全局特征中的位置, 增强网络的泛化能力。

在常规残差块 (block) 中添加多分割局部增强注意力机制 (MSLE) 模块和池化残差 (PR) 模块的基础块如图 8(c), 融合 MSLE 模块和 PR 模块的基础块简称为 MP-block, MP-block 能够最大可能地保存上一层传递的信息, 有效地防止梯度消失等问题, 使网络更容易优化和训练, 融合了多分割局部增强注意力机制模块以及池化残差模块的函数可以表示为:

$$H(\mathbf{x}) = M(F(\mathbf{x})) + P(\mathbf{x}), \quad (23)$$

其中:  $F(\mathbf{x})$  表示残差块主干的特征输出函数,  $\mathbf{x}$  表示为上一层传递的特征矩阵,  $P(\mathbf{x})$  表示经过 PR 模块的

特征函数,  $M(F(\mathbf{x}))$  表示残差块主干上输出的特征通过 MSLE 后的表达函数。

### 2.4.2 MSLENet 总体结构

MSLENet 由多个融合 MSLE 模块和 PR 模块的残差块 (MP-block) 连接组成, 网络结构如图 9 所示。网络结构以 ResNet34 作为基线网络, 总体网络结构共分为 3 个部分:

1) 将原 ResNet34 的  $7 \times 7$  卷积改为  $3 \times 3$ , 并删除最大池化操作。用于最大保留输入端的特征信息。

2) 输入端得到的特征信息, 进入 Layer 层, Layer 层共四个结构来最大化提取特征的信息, 提高模型泛化能力。每个 Layer 层中的 MP-block 个数为 [3, 4, 6, 3], 与 ResNet34 保持一致。较浅 Layer 层的作用是提取特征的细节信息, 同时保留原有特征的完整性, 增强通道间的互动。较深 Layer 层作用是提取网络中的抽象信息, 增加网络的泛化能力。

3) 经过 Layer 层得到的特征, 通过平均池化和全连接层提取 Layer 层的特征映射信息, 实现网络的分类结果。

总体而言, MSLENet 首先将  $7 \times 7$  卷积改为  $3 \times 3$ , 并删除最大池化操作, 更好地增强网络中的有利信息; 然后提出 MSLE 模块, 通过关注局部的特征信息, 从而增强且扩大局部有利信息, 提升局部信息之间的交互性, 引导全局去提取重要信息; 最后, 为了改善跳跃连接结构中信息的丢失问题, 进一步改进 ResNet34 中的跳跃连接部分, 加入平均池化来更好地保留上一

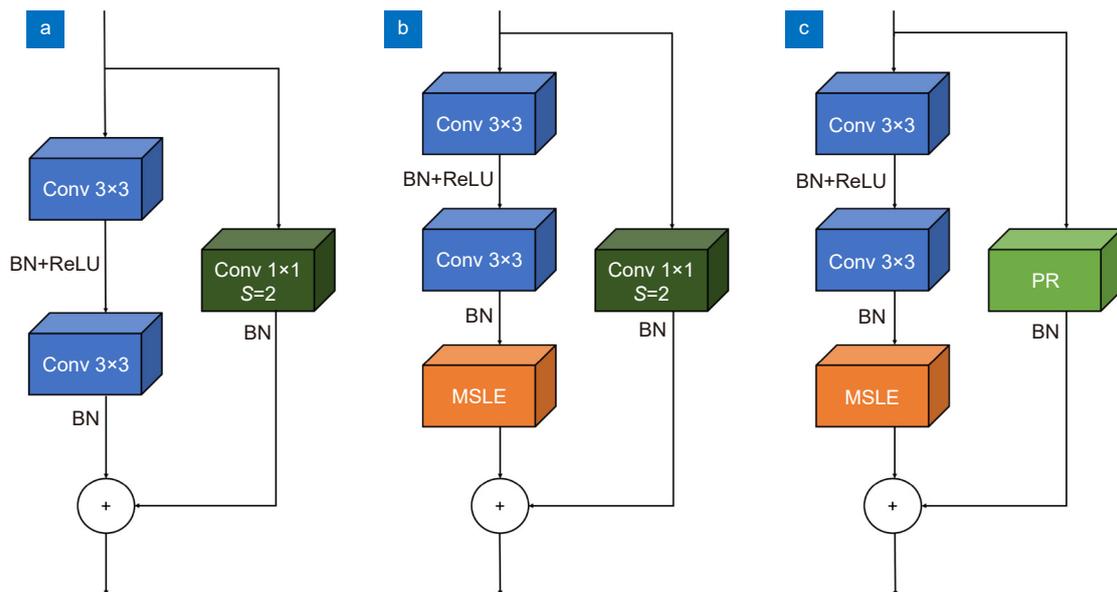


图 8 三种模块结构。(a) 常规残差块; (b) 融合 MSLE 残差块; (c) 融合 MSLE 和 PR 残差块

Fig. 8 Three module structures. (a) Block; (b) M-block; (c) MP-block

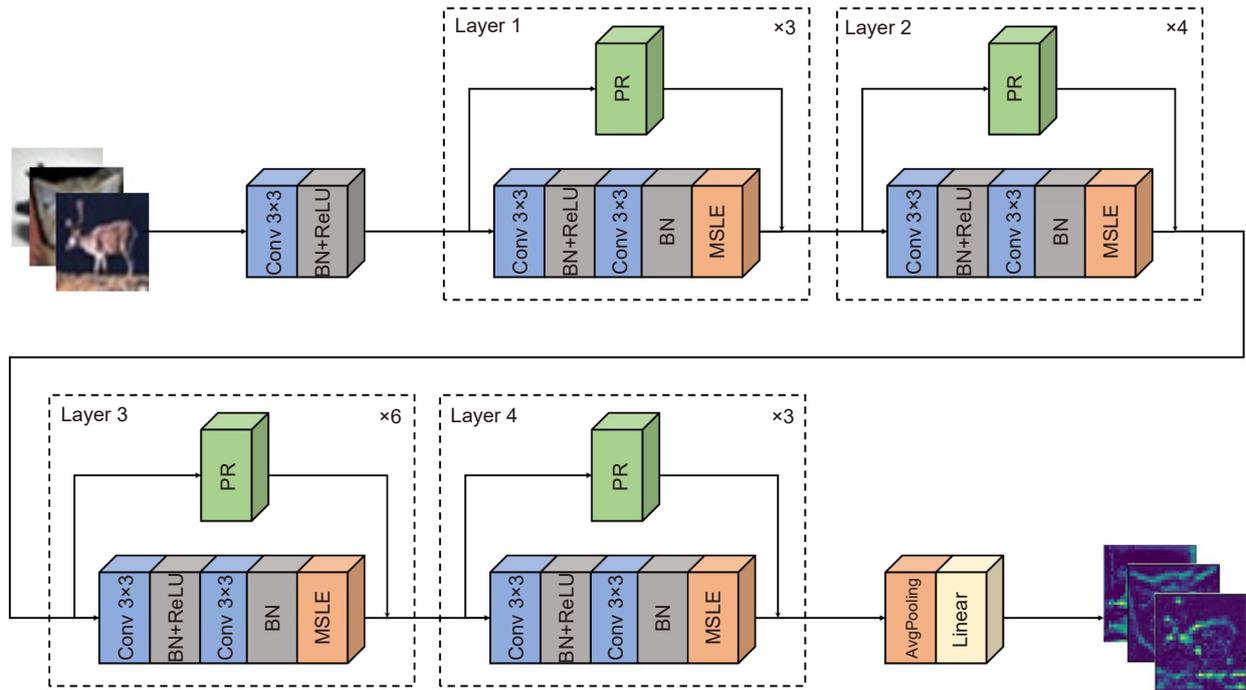


图 9 MSLENet 总体结构图

Fig. 9 Overall structure of MSLENet

层传递的信息, 改善每一层信息之间的交互流通, 从而提高模型的表达能力。

### 3 实验与结果分析

#### 3.1 数据集及实验环境

为了评估 MSLENet 的性能, 在后面的实验中分别使用 7 种数据集来验证, 七种数据集分别是 CIFAR-10、CIFAR-100、SVHN、GTSRB、STL-10、Imagenette、NWPU-RESISC45。其中 CIFAR-10 包含 10 种类别, 总计 60000 张  $32 \times 32$  的彩色图片, 每个类别包含 6000 张图片。CIFAR-100 是 CIFAR-10 数据集的变体, 含有 100 个类别, 每个类别包含 600 张  $32 \times 32$  的图片。SVHN 是一种用于数字识别的数据集, 它包含 10 个类别, 其中训练集包含 73257 张图片, 测试集包含 26032 张图片, 图像尺寸为  $32 \times 32$ 。

GTSRB 是交通标志的 43 分类的数据集, 其尺寸为  $32 \times 32$ , 训练集有 39209 个图片, 测试集含有 12630 张图片。STL-10 是 10 分类图像, 其中共有 5000 张训练图像和 8000 张测试图像, 是  $96 \times 96$  的彩色图像。Imagenette 是 image 子集, 图像尺寸为  $320 \times 320$ 。NWPU-RESISC45 是 45 分类的不同物品的数据集, 包括房、车、高速公路灯, 每种数据集的详细参

数如表 1 所示。在后续的实验, 对于不同的模块使用不同的数据集进行验证。

表 1 数据集

Table 1 Dataset

名称	图像尺寸	分类数	训练集数量	测试集数量
CIFAR-10	$32 \times 32$	10	50000	10000
CIFAR-100	$32 \times 32$	100	50000	10000
SVHN	$32 \times 32$	10	73257	26032
GTSRB	$32 \times 32$	43	39209	12630
STL-10	$96 \times 96$	10	5000	8000
Imagenette	$320 \times 320$	10	7000	3000
NWPU-RESISC45	$256 \times 256$	45	27000	4500

本文实验环境采用的系统 Linux 5.15.133, 训练所用的硬件以及参数为: NVIDIA Tesla P100 显卡、60 GB 内存; 软件环境为: Pytorch 2.0、CUDA 11.4; 编程环境为 Python 3.8。

#### 3.2 三种结构的对比

为了验证 MSLE 模块和 PR 模块在 MSLENet 中的作用, 使用三种网络结构来测试两种模块的性能, 第一种网络是替换输入层  $7 \times 7$  卷积为  $3 \times 3$  卷积, 删除池化层的 ResNet34, 简记为 ResNet34-c; 第二种网络在 ResNet34-c 的基础上融合了 MSLE 模块, 称为 M-

MSLENet; 第三种是在第二种的基础上融合了 PR 模块, 简称为 MP-MSLENet。ResNet34-c、M-MSLENet、MP-MSLENet 的基础块分别为 Block, M-block, MP-block, 三种网络结构如图 10 所示。

为了验证 MSLE 模块和 PR 模块对网络的影响, 三种网络使用相同的训练方案, 使用的数据集为 CIFAR-10、CIFAR-100、SVHN。对于数据预处理, 首先添加随机裁剪进行数据增强, 裁剪大小为  $32 \times 32$ , 并在图像周围填充 4 个像素, 达到增强图像边界信息的作用, 然后执行随机水平翻转, 翻转概率为 0.5, 最后进行图像标准化, 加快网络的收敛速度。在优化器的选择上, 使用 SGD 执行优化过程, 动量为 0.9, 批处理大小 (batch-size) 为 128, 衰减比为  $5e-4$ 。对于 MP-MSLENet 和 MP-MSLENet 中 MSLE 模块带来超参数  $k$  和  $l$ , 设每个 Layer 层中  $k$  值为 4、2、0、0,

$l$  为 16。对于学习率, 设置初始学习率为 0.1, CIFAR-10、CIFAR-100、SVHN 分别设置训练次数为 200、250、200, CIFAR-10 和 SVHN 的学习率衰减策略分别到 60、120、160 轮衰减, 衰减率为 0.2, CIFAR-100 是在每 60、120、160、200、230 轮衰减, 衰减率为 0.2。三种网络在 CIFAR-10、CIFAR-100、SVHN 数据集下的迭代准确率和迭代损失如图 11、图 12 所示。从三种数据集的准确率和损失来看, 加入 MSLE 模块的 M-MSLENet 相比于 ResNet-c 有更高的准确率和更快的收敛速度, 因为 MSLE 通过运用局部注意力机制的方式引导全局去提取有利信息, 从而增强网络提取重要信息的能力, 提高网络的准确率。而加入 PR 模块的 MP-MSLENet 使得网络拥有的更快的收敛速度和更高的准确率, 因为 PR 模块降低了原 ResNet 残差结构带来的信息损失, 保留的信息能够

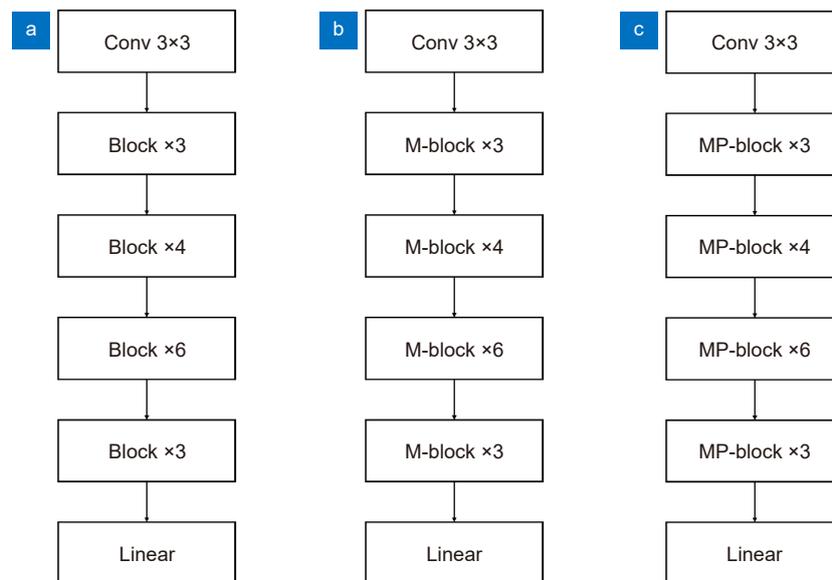


图 10 三种网络结构图。(a) 改变首层的 ResNet; (b) 融合 MSLE 的 MSLENet; (c) 融合 MSLE 和 PR 的 MSLENet

Fig. 10 Structure diagrams of three types of network. (a) ResNet34-c; (b) M-MSLENet; (c) MP-MSLENet

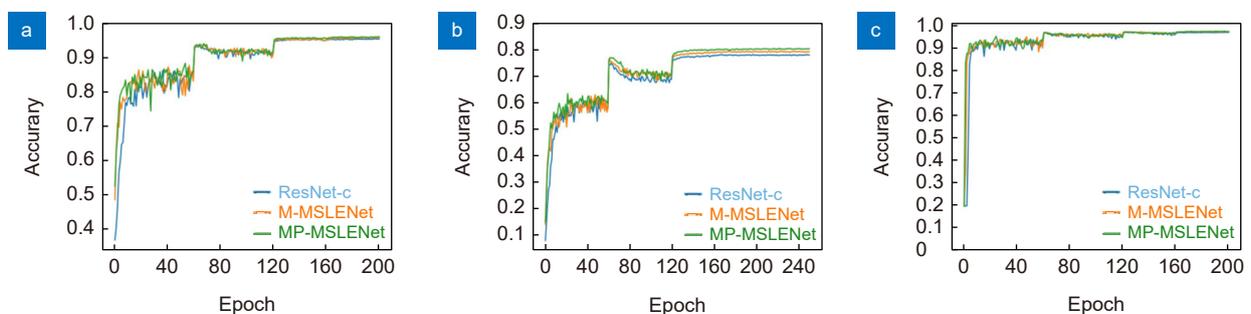


图 11 三种数据集下三种网络的迭代准确率。(a) CIFAR-10; (b) CIFAR-100; (c) SVHN

Fig. 11 Three type of network iteration accuracies under three datasets.

(a) CIFAR-10; (b) CIFAR-100; (c) SVHN

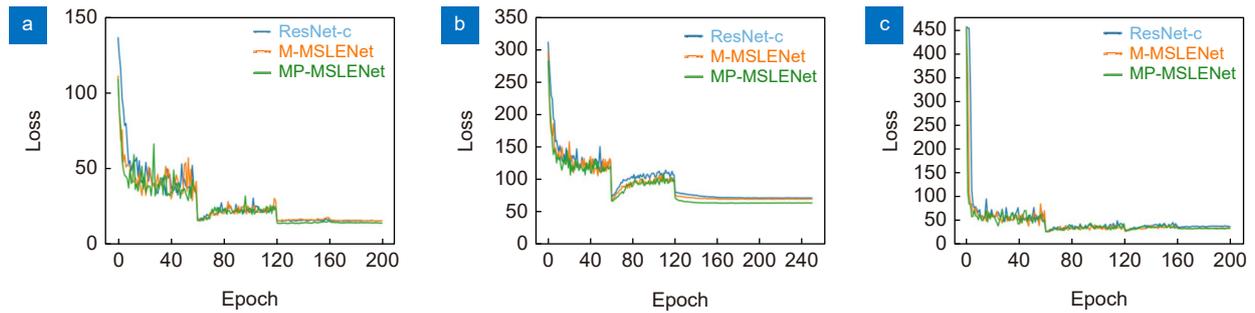


图 12 三种数据集下三种网络的迭代损失。(a) CIFAR-10; (b) CIFAR-100; (c) SVHN  
Fig. 12 Three type of network iteration loss under three datasets. (a) CIFAR-10; (b) CIFAR-100; (c) SVHN

加固 MSLE 模块对特征信息的提取, 从而使得网络的泛化能力更强。为了更直观地看到三种网络在三种数据集下的准确率, 三种网络在三种数据集下的最高准确率如表 2 所示。在不同数据集下 MP-MSLENet 均达到最高的准确率, 根据三种网络在不同数据集下的准确率来看, MSLE 模块和 PR 模块的相结合对于网络性能的提升是较大的, 并且有效地提高了网络的泛化能力, 所以本文将 MP-MSLENet 作为最终的网络模型, 后文中出现的 MSLENet 均代表 MP-MSLENet。

表 2 三种网络在三种数据集下的准确率  
Table 2 Accuracy of three networks under three datasets

网络	CIFAR-10/%	CIFAR-100/%	SVHN/%
ResNet-c	95.38	78.02	96.63
M-MSLENet	95.78	79.33	96.89
<b>MP-MSLENet</b>	<b>96.02</b>	<b>80.42</b>	<b>96.94%</b>

### 3.3 不同网络层性能对比

为了验证网络层对 MSLENet 性能影响, 选用 MSLENet18、MSLENet34、MSLENet50、MSLENet101 四种网络层进行对比, 四种网络的基线网络为 ResNet18、ResNet34、ResNet50、ResNet101, 四种网络在 CIFAR-100 数据集下的准确率、平均损失、F1-Score 的对比如表 3 所示, 表中黑体为最佳参数, 从表 3 中得知 MSLENet34 的性能最佳。MSLENet18 网络层较少, 参数量较少, 训练过程可能出现欠拟合的情况, 所以网络性能较低; MSLENet50 和 MSLENet101 虽然网络层比较深, 参数量比较大, 但是过深的网络层很可能造成网络的过拟合现象, 同时过深的网络层很可能过分识别图像的抽象信息, 使图像重要特征失效, 从而准确率较低; MSLENet34 的

网络层正好适用于当前训练过程, 网络层和参数量都到达了一种平衡, 所以性能最佳。

表 3 四种网络的参数对比  
Table 3 Comparison of parameters for four modules

网络	准确率/%	F1-score	Xentropy
MSLENet18	79.14	0.7916	0.012
<b>MSLENet34</b>	<b>80.42</b>	<b>0.8065</b>	<b>0.008</b>
MSLENet50	78.66	0.7883	0.010
MSLENet101	79.65	0.7965	0.008

### 3.4 各个模块性能对比

为了验证 MSLE 模块是否拥有先进的性能, 本文将 MSLE 模块与其它的注意力机制相比较, 比较的对象分别是 CA<sup>[16]</sup>、ECA<sup>[17]</sup>、GCT、SE, 使用四种数据集进行比较, 四种数据集分别是 CIFAR-100、STL-10、Imagenette、NWPU-RESISC45。训练过程的迭代次数均为 250 次, MSLE 超参数  $k$  设置为 [4,2,0,0], 使用的基线网络均为 ResNet34, 训练过程不做任何增强, 其余训练过程与 3.2 节一样。

五种模块在 CIFAR-100、STL-10、Imagenette、NWPU-RESISC45 下的准确率、F1-score、平均损失 (Xentropy) 对比如表 4~7 所示, 黑体为表中最佳值。根据表中可以得出在不同的数据集下几种模块的最佳参数均有所不同, 对于 CIFAR-100、STL-10 这种多分类数据集来说, ECA 的性能相比于其余三种更加强势, 因为 ECA 注意力机制保持原有通道的特性的同时对原通道特征进行的特征提取, 缺点是对整体通道的全局平均池化并不能代表该通道下的特征矩阵的整体, 可能会忽略细节部分。对于 Imagenette 和 NWPU-RESISC45 这类尺寸较大的数据集, CA、ECA、SE 各有优势, 因为它们均是通过通道注意力机制提高重要信息的利用率, 但是它们都忽略了图像

局部空间特征也会影响全局特征的问题, 而 MSLE 模块不论是准确率、F1-score、平均损失相比与其余四种更好, 因为 MSLE 注意到特征的细节部分, 将重要信息引导入全局当中, 从局部角度和全局角度都能注意到重要特征, 所以加入 MSLE 模块相比于其它模块有更好的性能, 准确率更高。

四种数据集下五种模块在不同迭代次数下的准确率对比如图 13 所示, 从图中可见, 在四种数据集下, 加入 MSLE 模块的网络, 不论是收敛速度还是准确率相比于其它均有提升, 可见 MSLE 模块拥有较好的性能。因为 MSLE 模块细节抓取图像特征的重要信息, 提升每一层的感受野, 从而减少了网络中信息的丢失, 所以 MSLE 模块能够提升网络的泛化能力, 提高网络的鲁棒性。

为了更有效地观测 MSLE 模块的效果, 将图像通过第一层网络得到特征矩阵进行通道分解, 随机挑

选 12 种单通道下的特征矩阵进行可视化, 并与不同的模块进行对比, 对比的模块有 CA、ECA、GCT、SE。五种模块经过第一层网络得到的单通道特征矩阵可视化如图 14 所示。由图 14 可以清晰表达出, CA、ECA、GCT、SE 均从整体关注图像信息, 12 种通道下的图像均忽视掉部分重要的边缘信息, 图像的散度相对较差, 例如, ECA、CA 的有些通道更关注边缘的无用信息。而加入 MSLE 模块不论是中心信息还是部分边缘细节信息均得到有效利用, 并且每个通道的特征均关注到不同程度的信息, 散度更强, 局部信息利用更完善, 从而引导全局特征的效果更强, 能够实现局部信息到全局信息的引导。

### 3.5 MSLENet 网络性能

为了验证 MSLENet 网络的性能, 本文将使用不同的数据集测试 MSLENet 的最佳性能, 分别使用 CIFAR-10、CIFAR-100、SVHN、GTSRB, 四个数据

表 4 五种模块在 CIFAR-100 的参数对比

Table 4 Comparison of parameters for five modules on CIFAR-100

网络	准确率/%	F1-score	Xentropy
+CA	78.55	0.7881	0.010
+ECA	79.33	0.7949	0.011
+GCT	77.08	0.7810	0.012
+SE	79.25	0.7932	0.010
<b>+MSLE</b>	<b>80.37</b>	<b>0.8054</b>	<b>0.009</b>

表 6 五种模块在 Imagenette 的参数对比

Table 6 Comparison of parameters for five modules on Imagenette

网络	准确率/%	F1-score	Xentropy
+CA	89.14	0.8924	0.003
+ECA	88.87	0.8903	0.004
+GCT	87.34	0.8745	0.002
+SE	89.07	0.8907	0.003
<b>+MSLE</b>	<b>89.70</b>	<b>0.8991</b>	<b>0.002</b>

表 5 五种模块在 STL-10 的参数对比

Table 5 Comparison of parameters for five modules on STL-10

网络	准确率/%	F1-score	Xentropy
+CA	72.25	0.7228	0.005
+ECA	72.40	0.7241	0.006
+GCT	69.26	0.6931	0.008
+SE	70.32	0.7040	0.007
<b>+MSLE</b>	<b>72.78</b>	<b>0.7282</b>	<b>0.003</b>

表 7 五种模块在 NWPU-RESISC45 的参数对比

Table 7 Comparison of parameters for five modules on NWPU-RESISC45

网络	准确率/%	F1-score	Xentropy
+CA	93.00	0.9302	0.012
+ECA	95.33	0.9533	<b>0.003</b>
+GCT	94.20	0.9421	0.007
+SE	95.13	0.9513	<b>0.003</b>
<b>+MSLE</b>	<b>95.40</b>	<b>0.9540</b>	0.004

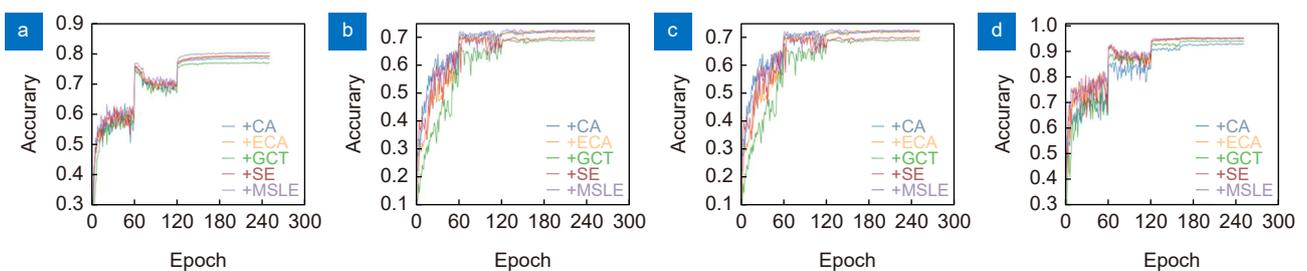


图 13 五种模块不同迭代次数下的准确率。(a) CIFAR-100; (b) STL-10; (c) Imagenette; (d) NWPU-RESISC45

Fig. 13 Accuracy of five modules at different iterations. (a) CIFAR-100; (b) STL-10; (c) Imagenette; (d) NWPU-RESISC45

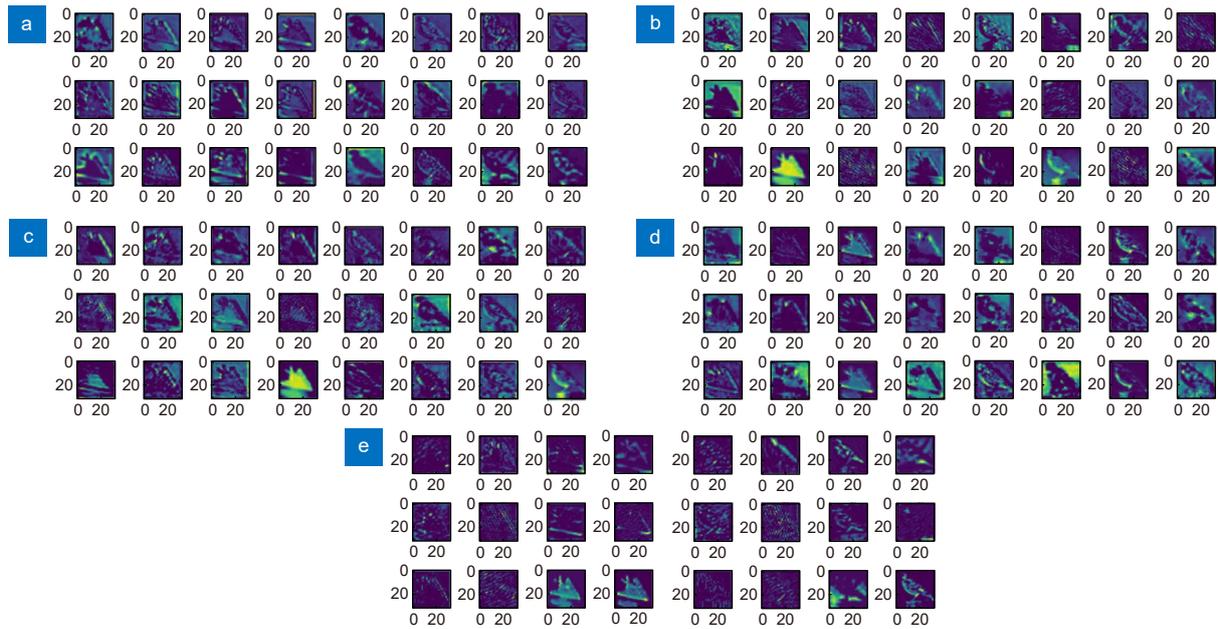


图 14 不同模块下的通道可视化图像。(a) CA; (b) ECA; (c) GCT; (d) SE; (e) M-APC

Fig. 14 Channel visualizations under different modules. (a) CA; (b) ECA; (c) GCT; (d) SE; (e) M-APC

集分别对 MSLENet 训练。图像预处理阶段使用一些图像增强的方法。使用  $32 \times 32$  的随机裁剪 (RandomCrop), 填充值为 4, 并对图像矩阵进行标准化; 使用随机水平翻转 (RandomHorizontalFlip), 随机翻转概率值为 0.5; 使用随机擦除<sup>[18]</sup> (RandomErasing), 随机擦除概率等于 0.25; 超参数设置中, 迭代次数设置为 300 次; 所有数据集在训练中初始学习率 (lr) 均为 0.1, 学习率分别在 60、120、160、200、240、270 轮的时候下降原来的五分之一; 批处理大小 (batch size) 为 128, 使用 SGD 优化器, 动量 (momentum) 设置 0.9, 权重衰减 (weight decay) 为  $5e-4$ ; 设置 MSLE 模块中的分割数  $k$  在每一个 Layer 层的值分别为 4、2、0、0, 超参数  $l$  设置为 16; 使用 Label Smoothing<sup>[19]</sup> (标签平滑) 辅助训练过程; 使用 Mixup<sup>[20]</sup> 和 EMA<sup>[21]</sup> 进一步提高 MSLENet 的性能。预处理中使用的数据增强方法及对应的超参数, 以及训练过程中设置的超参数和技术方法等如表 8 所示。经过多次实验, 在使用上述训练过程后, CIFAR-10、CIFAR-100、SVHN、GTSRB 的分类准确率分别达到 96.93%、82.28%、97.22%、97.39%, 四个数据均为分类最高值。

### 3.6 各网络对比实验

为了验证 MSLENet 的性能, 使用 MSLENet 与其它网络在 CIFAR-10、CIFAR-100、SVHN 数据集下进行对比实验, 对比的分类网络包括: VGG-16、

ResNet34、DenseNet-121、Multi-ResNet、MobileNetV2、ShuffleNet、SENet、CAPR-DenseNet<sup>[22]</sup>、EfficientNet<sup>[23]</sup>、SSE-GAN<sup>[24]</sup>、Couplformer<sup>[25]</sup>、FAVOR+<sup>[26]</sup>、ResNet50+SPAM<sup>[27]</sup>、ResNet-CE<sup>[28]</sup>、MMA-CCT-7/3 $\times$ 2<sup>[29]</sup>、CaiT<sup>[30]</sup>、Swin-T<sup>[31]</sup>, 由于不同的训练过程对于不同的网络均有不同的性能影响, 并不能保证不同的增强对于网络均有良好效果, 有些论文也并没有提供训练过程, 所以分类数据集准确度来源如下: 1) 首先使用原论文提供的最高分类准确率; 2) 如若原文没有提供分类准确率, 则使用提供的开源代码进行复现, 同时训练过程与 3.5 小节一致。各个分类网络在 CIFAR-10、CIFAR-100、SVHN 数据集下的分类准确率结果如表 9 所示。由表 9 加粗黑体得出 MSLENet 在 3 种数据集下均达到最高的准确率。

VGG-16、MobileNetV2、ShuffleNet 均为轻量化网络, 它们主要为了表现在移动端的分类网络的开发, 所以网络层数较少, 并且使用的参数量相对较少, 所以对于三种数据集来说, 并不会体现出分类准确率的优势, 而且这三种网络的复杂度较低, 对于识别和提取图像特征来说存在一定的困难, 也会导致特征信息的丢失。所以它们的分类准确率相对其它网络来说较低。

DenseNet-121、CAPR-DenseNet、Multi-ResNet 为密集类残差网络, 因为密集类网络的网络模型复杂度高, 层与层之间的连接相对较多, 图像特征的重复

表 8 训练过程中超参数的设定值

Table 8 Setting values of hyperparameters during training process

超参数	设定值
Input size	32×32
RandomCrop	4
RandomHorizontalFlip	0.5
RandomErasing	0.2
epochs	300
优化器	SGD
lr	0.1
lr decay	0.2
batch size	128
Momentum	0.9
Weight decay	5e-4
Mixup	0.2
EMA	0.9
Label Smoothing	0.1
$k$	[4,2,0,0]
$l$	16

表 9 各网络在三种数据集下的分类准确率

Table 9 Classification accuracy of each network under three datasets

网络	CIFAR-10/%	CIFAR-100/%	SVHN/%
VGG-16	91.79	67.84	-
SENet	95.22	73.22	87.06
DenseNet-121	94.55	77.01	95.83
CAPR-DenseNet	94.24	78.84	94.95
MobileNetV2	93.37	68.08	-
ShuffleNet	89.40	70.06	-
ResNet34	87.89	69.41	91.51
Multi-ResNet	94.65	78.68	-
EfficientNet	94.01	75.96	93.32
SSE-GAN	85.14	-	92.92
Couplformer	93.54	73.92	94.26
ResNet50+SPAM	-	80.53	-
FAVOR+	91.42	72.56	93.21
ResNet-CE	94.27	76.15	-
MMA-CCT-7/3×2	94.74	77.5	94.26
CaiT	94.91	79.89	--
Swin-T	94.46	78.07	--
<b>MSLENet</b>	<b>96.93</b>	<b>82.28</b>	<b>97.22</b>

利用率也较高,所以在分类准确率上有所提升,但是同时也会带来复杂的计算成本,由于层数的堆积和连接,导致它们均存在信息堆积量大,造成信息冗余,从而使网络并不能有效地识别特征中的有利信息。

如加入通道注意力机制的 SENet、EfficientNet 以及基于 Transformer 的 Couplformer、Swim-T、Cai-T 在分类准确率提升相对较多,这是因为通过注意力机制的参与,网络识别到图像特征中的有利信息,阻碍图像特征的不利信息,从而提高网络的准确率。但是通道注意力机制只是单纯从每个通道的特征矩阵的整体出发,对每个完整通道使用全局平均池化会造成通道中有些细节信息被忽视掉,通道的整体平均池化有时并不能代表整个通道,可能会阻碍注意力的感受野。对于自注意力而言,它通过元素序列来进行计算,这导致了计算成本的显著提升,虽然它能体现特征矩阵元素之间的映射关系,但是它过于依赖元素的顺序信息,如果元素间信息的不相关性较大,则网络的性能可能会被阻碍。

为了提升网络性能,必须考虑将全局特征和局部特征相结合,充分利用特征之间的流通关系,从而提高模型的泛化能力。对比发现融合了 MSLE 模块和 PR 模块的 MSLENet 的网络性能最佳。同时为了测试 MSLENet 的计算量 (FLOPs) 和参数量 (params) 的

优越性,使用 MSLENet 与以下经典网络的计算量和参数量进行对比,对比网络有 Wide-ResNet、ConvNeXt<sup>[32]</sup>、EfficientNet、Swim-T、Muti-ResNet。各个网络的计算量和参数量如表 10 所示,MSLENet 的计算量和参数量对网络负担是最小的。虽然 EfficientNet 计算量较低,但是它对网络的负担较重,因为不停的卷积和自注意力机制都会带来额外的开销,并且层数较低时性能较差;Wide-ResNet、Muti-ResNet 和 Swim-T 由于网络层数过多,所以参数量和计算量均较大;ConvNeXt 类似于 ResNet,所以参数量和计算量较低,但是它使用增加了卷积层的同时改变卷积方式来提高性能,因此也提高了参数量和计算量。综合来看,MSLENet 不论是计算量还是参数量均达到一个较为平衡的结果,并没有过度增加网络负担,网络性能提升也较大。

### 3.7 消融实验

为了验证 MSLE 模块中分割数  $k$  的有效分割值,本文在四个数据集下进行验证,四个数据集分别是 CIFAR-10、CIFAR-100、GTSRB、NWPU-RESISC45,训练过程使用 3.5 小节相同训练过程。定义网络 Net\_2\_2\_0\_0、Net\_4\_2\_0\_0、Net\_4\_2\_2\_0、Net\_8\_4\_2\_0 分别代表  $k$  取值 [2,2,0,0]、[4,2,0,0]、[4,2,2,0]、[8,4,2,0],例如, $k$  取 [4,2,0,0] 表示 MSLENet 中的四个 Layer 层中的分割数  $k$  按顺序取 4、2、0、0,其

中 0 代表不分割, 即原特征矩阵直接进入一个 LE 层。

各个网络的消融实验结果的最高分类准确率如表 11 所示, 从表中可知, 对于 CIFAR-10 和 GTSRB 的小分类数据集, 当  $k$  取 [4,2,0,0] 时达到最高的准确率, 分别是 96.93% 和 97.39%。对于多分类数据集 CIFAR-100,  $k$  取 [8,4,2,0] 时达到最高的准确率, 准确率为 82.51%。对于高分辨率数据集 NWPU-RESISC45,  $k$  取 [4,4,2,0] 时达到最高的准确率, 准确率为 95.44%。黑体字表示各个网络的最优准确率, 可知取不同的  $k$  值, 对网络性能均存在一定的影响。对于小分类数据集而言, 接近网络输入层取较小的  $k$  值会相对提升网络的性能, 但是  $k$  值不能太小, 因为太小的  $k$  值并不能更好地提取局部特征。对于多分类数据集而言, 在靠近输入端的 Layer 层取较大的  $k$  值会对网络性能提升较大。对于高分辨率数据集而言, 输入端存在较多的重要特征, 取较大的  $k$  值更有可能获取细节信息, 但是  $k$  值不宜过大, 因为分辨率高的数据集在输入端的特征之间存在较大的联系, 过度分割会造成重要信息被分离, 高分辨率图像接近输出端的图像尺寸也相对较大, 抽象信息较多, 所以要使用适当的  $k$  值进行分割, 以获取抽象细节特征为目的。

总而言之,  $k$  的取值在接近输入端的时候取值相对大一些, 越接近输出端时,  $k$  的取值应该越小。因为对于接近输入端的网络层应尽量提取网络的细节部

分, 而接近输出端的网络层需要整体把握特征信息, 从抽象的角度获取细节特征。

## 4 结论

为了更好地利用图像局部信息来引导整体实现特征提取, 提升局部特征之间的关联性的同时, 使整体特征信息得到有效利用, 本文提出了局部注意力引导下的全局池化残差分类网络 (MSLENet)。MSLENet 以 ResNet34 为基线网络, 首先改变首层网络结构, 保证输入图像的特征信息更加完整; 然后改变残差结构, 提出 PR 模块, PR 模块减少了特征信息的丢失, 强化不同层之间信息的可利用性, 同时改善了网络过拟合, 梯度消失等问题; 最后提出 MSLE 模块, MSLE 模块对每个空间局部特征和不同通道的局部特征均提取不同的重要信息, 并扩大上述局部重要信息, 从而增强局部特征之间的关联性, 同时去除局部不利的边缘信息, 进而把握局部和整体的交互性, 并将局部重要特征引导入全局特征当中, 提高网络的泛化能力。实验结果表明, MSLENet 在 CIFAR-10、CIFAR-100、SVHN、GTSRB、NWPU-RESISC45 准确率分别为 96.93%、82.51%、97.22%、97.39%、95.44%, 显然 MSLENet 能够提高分类网络的准确率, 提高网络的表达能力。但是 MSLEnet 受到  $k$  值的影响, 虽然较大的  $k$  值能够提升网络性能, 但是随之而来的结果是网络的参数量增加。在随后的工作中, 我们会尽

表 10 各个网络的计算量和参数量

Table 10 FLOPs and params of various networks

网络	Params/M	FLOPs/G
Wide-ResNet	37.16	5.96
ConvNext	27.80	1.45
EfficientNet	52.98	1.49
Swim-T	86.78	4.25
Multi-ResNet	51.23	3.13
MSLENet	22.35	1.20

表 11 各个网络实验结果

Table 11 Experimental results of various networks

网络	CIFAR-10/%	CIFAR-100/%	GTSRB/%	NWPU-RESISC45/%
Net_2_2_0_0	96.81	79.65	97.26	95.17
Net_4_2_0_0	<b>96.93</b>	82.28	<b>97.39</b>	95.40
Net_4_2_2_0	96.80	82.23	97.13	<b>95.44</b>
Net_8_4_2_0	96.90	<b>82.51</b>	97.27	95.15

可能地在提高分类结果的同时简化 MSLE 中带来的参数量。

**利益冲突:** 所有作者声明无利益冲突

## 参考文献

- [1] Robbins H, Monro S. A stochastic approximation method[J]. *Ann Math Statist*, 1951, **22**(3): 400–407.
- [2] Yang H, Li J. Label contrastive learning for image classification[J]. *Soft Comput*, 2023, **27**(18): 13477–13486.
- [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] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*Proceedings of the 3rd International Conference on Learning Representations*, San Diego, 2015.
- [6] 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>.
- [7] 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>.
- [8] 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>.
- [9] Abdi M, Nahavandi S. Multi-residual networks: improving the speed and accuracy of residual networks[Z]. arXiv: 1609.05672, 2017. <https://arxiv.org/abs/1609.05672>.
- [10] Howar A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[Z]. arXiv: 1704.04861, 2017. <https://arxiv.org/abs/1704.04861>.
- [11] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 6848–6856. <https://doi.org/10.1109/CVPR.2018.00716>.
- [12] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, 2015: 2017–2025.
- [13] 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>.
- [14] Yang Z X, Zhu L C, Wu Y, et al. Gated channel transformation for visual recognition[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 11794–11801. <https://doi.org/10.1109/CVPR42600.2020.01181>.
- [15] Zhang F, Huang S X, Hua Q, et al. Novel image classification model based on depth-wise convolution neural network and visual transformer[J]. *Comput Sci*, 2024, **51**(2): 196–204. 张峰, 黄仕鑫, 花强, 等. 基于 Depth-wise 卷积和视觉 Transformer 的图像分类模型[J]. *计算机科学*, 2024, **51**(2): 196–204.
- [16] 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>.
- [17] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 11531–11539. <https://doi.org/10.1109/CVPR42600.2020.01155>.
- [18] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, 2020. <https://doi.org/10.1609/aaai.v34i07.7000>.
- [19] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016: 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>.
- [20] Zhang H Y, Cissé M, Dauphin Y N, et al. mixup: beyond empirical risk minimization[C]//*Proceedings of the 6th International Conference on Learning Representations*, Vancouver, 2018.
- [21] Polyak B T. Some methods of speeding up the convergence of iteration methods[J]. *USSR Comput Math Math Phys*, 1964, **4**(5): 1–17.
- [22] Zhang K, Guo Y R, Wang X S, et al. Channel-wise and feature-points reweights densenet for image classification[C]//*Proceedings of 2019 IEEE International Conference on Image Processing*, Taipei, China, 2019: 410–414. <https://doi.org/10.1109/ICIP.2019.8802982>.
- [23] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//*Proceedings of the 36th International Conference on Machine Learning*, Long Beach, 2019: 6105–6114.
- [24] Fu X, Shen Y T, Li H W, et al. A semi-supervised encoder generative adversarial networks model for image classification[J]. *Acta Autom Sin*, 2020, **46**(3): 531–539. 付晓, 沈远彤, 李宏伟, 等. 基于半监督编码器生成对抗网络的图像分类模型[J]. *自动化学报*, 2020, **46**(3): 531–539.
- [25] Choromanski K M, Likhoshesterov V, Dohan D, et al. Rethinking attention with performers[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.
- [26] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//*Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, 2011: 315–323.
- [27] Wang F, Qiao R P. SPAM: spatially partitioned attention

- module in deep convolutional neural networks for image classification[J]. *J Xi'an Jiaotong Univ*, 2023, 57(9): 185–192.
- 王方, 乔瑞萍. 用于图像分类的深度卷积神经网络中的空间分割注意力模块[J]. *西安交通大学学报*, 2023, 57(9): 185–192.
- [28] Yang M L, Zhang W S. Image classification algorithm based on classification activation map enhancement[J]. *J Front Comput Sci Technol*, 2020, 14(1): 149–158.
- 杨萌林, 张文生. 分类激活图增强的图像分类算法[J]. *计算机科学与探索*, 2020, 14(1): 149–158.
- [29] Konstantinidis D, Papastratis I, Dimitropoulos K, et al. Multi-manifold attention for vision transformers[J]. *IEEE Access*, 2023, 11: 123433–123444.
- [30] Touvron H, Cord M, Sablayrolles A, et al. Going deeper with image transformers[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, 2021: 32–42. <https://doi.org/10.1109/ICCV48922.2021.00010>.
- [31] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, 2021: 9992–10002. <https://doi.org/10.1109/ICCV48922.2021.00986>.
- [32] Woo S, Debnath S, Hu R H, et al. ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 2023: 16133–1614. <https://doi.org/10.1109/CVPR52729.2023.01548>.

## 作者简介



姜文涛(1986-), 男, 博士, 副教授, 硕士生导师, 2010年于辽宁工程技术大学获得学士学位, 2012年、2015年于辽宁工程技术大学分别获得硕士和博士学位, 主要从事嵌入式系统、人工智能、计算机视觉等方面的研究。

E-mail: [jiangwentao@lntu.edu.cn](mailto:jiangwentao@lntu.edu.cn)



【通信作者】董睿(2000-), 男, 辽宁阜新人, 硕士, 2022年于辽宁工程技术大学获得学士学位, 主要从事人工智能、计算机视觉等方面的研究。

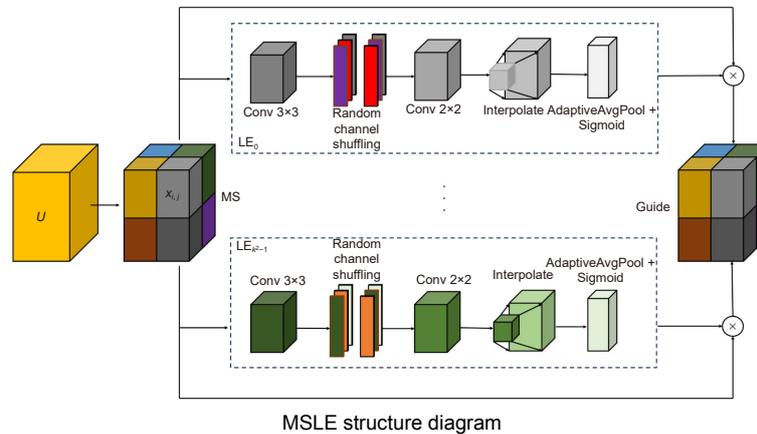
E-mail: [2669213806@qq.com](mailto:2669213806@qq.com)



扫描二维码, 获取PDF全文

# Global pooling residual classification network guided by local attention

Jiang Wentao<sup>1</sup>, Dong Rui<sup>1\*</sup>, Zhang Shengchong<sup>2</sup>



MSLE structure diagram

**Overview:** In image classification tasks, it has been demonstrated through various experiments that attention mechanisms can significantly enhance a model's generalization ability. However, most attention mechanisms only focus on enhancing the importance of local or global features, without considering that the interrelationships between local features can also affect the overall image features. To address this issue and improve the model's generalization ability, this paper proposes a global pooling residual classification network guided by local attention (MSLENet). MSLENet uses ResNet34 as its baseline network. It first modifies the initial convolution structure by replacing the convolution method and removing the pooling layer, allowing the network to retain the basic information of the image and enhance the utilization of detailed information. Secondly, this paper introduces a multiple segmentation local enhancement attention mechanism (MSLE) module, which enhances the information relationship between local and global features and amplifies local key information. The MSLE module consists of three sequential components: the multiple segmentation (MS) module, the local enhancement (LE) module, and the guide module. The MS module uniformly segments the image to fully utilize local information. The LE module enhances the local features of each segmented image and amplifies the local important information of the enhanced segments, thereby improving the interaction among local features and increasing the utilization of local key information. The guide module directs important local features into global features through the interaction between feature layers and different feature groups, thus enhancing the global important features and the network's expressiveness. Finally, to address the issue of information loss in the residual structure of ResNet, the pooling residual (PR) module is proposed. The PR module modifies the residual structure of ResNet34 by replacing the convolution method in the residual structure with pooling operations, thereby improving the information utilization between layers and reducing the network's overfitting. Experimental results show that MSLENet achieves accuracy rates of 96.93%, 82.51%, 97.22%, 72.82%, 97.39%, 89.70%, and 95.44% on the CIFAR-10, CIFAR-100, SVHN, STL-10, GTSRB, Imagenette, and NWPU-RESISC45 datasets, respectively. Compared to other networks or modules, MSLENet demonstrates improved performance, proving that the interaction between local and global features, the comprehensive utilization of both local and global information, and the guidance of important local features to global features effectively enhance the network's accuracy.

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; DOI: 10.12086/oe.2024.240126

Foundation item: Project supported by National Defense Pre-Research Foundation of China (172068), Liaoning Provincial Natural Science Foundation of China (20170540426), and Key Fund of Liaoning Provincial Department of Education (LJYL049)

<sup>1</sup>College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China; <sup>2</sup>Science and Technology on Electro-Optical Information Security Control Laboratory, Tianjin 300308, China

\* E-mail: 2669213806@qq.com