E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2025)01-0110-20

论文引用格式:Yuan H, Liu J, Jiang W T and Liu W J. 2025. Double-pooling residual classification network based on feature reordering attention mechanism. Journal of Image and Graphics, 30(01):0110-0129(袁姮, 刘杰, 姜文涛, 刘万军. 2025. 特征重排列注意力机制的双池化残差分类网络. 中国图象图形学报, 30(01):0110-0129)[DOI:10.11834/jig.240061]

特征重排列注意力机制的双池化残差分类网络

袁姮,刘杰*,姜文涛,刘万军

辽宁工程技术大学软件学院,葫芦岛　125105

摘　要:**目的**　针对残差图像分类网络中通道信息交互不充分导致通道特征没有得到有效利用,同时残差结构使部分特征信息丢失的问题,本文提出特征重排列注意力机制的双池化残差分类网络(double-pooling residual classification network of feature reordering attention mechanism,FDPRNet)。**方法**　FDPRNet以ResNet-34(residual network)残差网络为基础,首先将第1层卷积核尺寸从7×7替换为3×3,保留更多的特征信息,增强网络非线性表达能力,同时删除最大池化层,提高局部细节捕捉能力;然后,提出特征重排列注意力机制(feature reordering attention mechanism,FRAM)模块,将特征图通道进行分组,对其进行组间和组内重排序,通过一维卷积提取各通道组合的特征并拼接,得到重排列特征的权重;最后,提出了双池化残差(double-pooling residual,DPR)模块,该模块使用最大池化和平均池化并行操作特征图,并对池化后的特征图进行逐元素相加和卷积映射,以提取关键特征,减小特征图尺寸。**结果**　在CIFAR-100(Canadian Institute for Advanced Research)、CIFAR-10和SVHN(street view house numbers)数据集上与其他11种图像分类网络进行比较,相比于性能第2的模型RTSA Net-101(residual Net-101 with tensor-synthetic attention),准确率分别提高了1.16%、1.01%和0.98%。实验结果表明FDPRNet显著提升了分类准确率。**结论**　本文提出的FDPRNet具有增强图像通道内部信息交流和减少特征丢失的能力,不仅在分类精度上表现出较高水平,而且显著提升了模型的泛化能力。

关键词:图像分类;特征重排列;注意力机制;残差网络;深度学习

Double-pooling residual classification network based on feature reordering attention mechanism

Yuan Heng, Liu Jie*, Jiang Wentao, Liu Wanjun

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

Abstract: **Objective** A residual classification network is a deep convolutional neural network architecture that plays an important role and has a considerable influence in the field of deep learning. It has become one of the commonly used network structures in various image classification tasks in the field of computer vision. To solve the problem of network degradation in deep networks, unlike the traditional method of simply stacking convolutional layers, residual networks innovatively introduce residual connections, which directly add input features to output features through skip connections, and pass the original features directly to subsequent network layers. It forms a shortcut path, thereby preserving and utilizing

收稿日期:2024-02-04;**修回日期**:2024-05-13;**预印本日期**:2024-05-20

*通信作者:刘杰　2269224774@qq.com

基金项目:国家自然科学基金项目(61601213);辽宁省自然科学基金项目(20170540426);辽宁省教育厅重点基金项目(LJYL049)

Supported by: National Natural Science Foundation of China (61601213); Natural Science Foundation of Liaoning Province, China (20170540426); Key Project of Education Department of Liaoning Province(LJYL049)

feature information better. Although residual classification network effectively solves the problems of gradient explosion and vanishing during deep network training, when the output dimension of the residual block does not match the input dimension, convolution maps are needed to ensure the same dimensions, which causes a large number of pixels on the channel matrix in the residual module to be skipped, resulting in the problem of feature information loss. In addition, correlation exists between image channels, and a fixed order of channels may lead to feature bias, making it difficult to fully utilize information from other channels and limiting the model's ability to express key features. In response to the above issues, this article proposes a double pooling residual classification network of feature reordering attention mechanism (FDPRNet). **Method** FDPRNet is based on the ResNet-34 residual network. First, the kernel size of the first convolutional layer is changed from 7 × 7 to 3 × 3. This change is made because, for relatively small images, larger convolutional kernels can cause the receptive field to become larger, capturing too much useless contextual information. Time, the maximum pooling layer is removed to prevent the feature map from shrinking further, retaining more image information, avoiding information loss caused by pooling operations, and making it easier for subsequent network layers to extract features better. Then, a feature reordering attention module (FRAM) is proposed to group the feature map channels and perform inter-group and intra-group reordering so that adjacent channels are no longer connected, and the intra-group channels are grouped in a sequence of equal differences with a step size of 1. This operation can not only disrupt the order of some original channels before and after but also preserve the relationship between some channels before and after, introducing a certain degree of randomness, allowing the model to comprehensively consider the interaction between different channels, and avoiding excessive dependence on specific channels. The features of each channel combination are extracted and spliced by one-dimensional convolution, and then the sigmoid activation function is used to obtain the weights of the rearranged features, which are multiplied element by element with the input features to obtain the feature map of the feature rearranged attention mechanism. Finally, a double pooling residual (DPR) module is proposed, which uses both maximum pooling and average pooling to perform parallel operations on feature maps. This module obtains both salient and typical features of the input images, enhancing the expressive power of features and helping the network capture important information better in the images, thereby improving model performance. Element-by-element summation and convolutional mapping on the after-pooling feature maps are performed to extract key features, reduce the size of the feature maps, and ensure that the channel matrices are capable of element-level summation operations in residual concatenation. **Result** In the CIFAR-100, CIFAR-10, SVHN, Flowers-102, and NWPU-RESISC45 datasets, compared with the original model ResNet-34, the accuracy of ResNet-34 with the addition of FRAM is improved by 1.66%, 0.19%, 0.13%, 4.28%, and 2.00%, respectively. The accuracy of ResNet-34 with the addition of DPR is improved by 1.7%, 0.26%, 0.12%, 3.18%, and 1.31%, respectively. The accuracy of FDPRNet, which is the combination of the FRAM and DPR modules, is improved by 2.07%, 0.3%, 0.17%, 8.31%, and 2.47%, respectively. Compared with four attention mechanisms—squeeze and excitation, efficient channel attention, coordinate attention, and convolutional block attention module, the accuracy of FRAM is improved by an average of 0.72%, 1.28%, and 1.46% in the CIFAR-100, Flowers-102, and STL-10 datasets. In summary, whether on small or large, less categorized, or more categorized datasets, both the FRAM and DPR modules contribute to the improvement of recognition accuracy in the ResNet-34 network. The combination of the two modules—FDPR—has the best effect on improving the recognition rate of the network and achieves a significant improvement in accuracy compared with other image classification networks. **Conclusion** The proposed FDPRNet can enhance the information exchange within the image channel and reduce feature loss. It not only shows high classification accuracy but also effectively enhances the network's feature learning ability and model generalization ability. The main contributions of this article are as follows: 1) FRAM is proposed, which breaks the connections between the original channels and groups them according to certain rules. Learning the weights of channel combinations in different orders ensures that the channels between different groups interact without losing the front and back connections between all channels, achieving information exchange and channel crossing within the feature map, enhancing the interaction between features, better capturing the correlation between contextual information and features, and improving the accuracy of model classification. 2) DPR is proposed, which replaces the skip connections in the original residual block with a DPR module, solving the problem of feature information loss caused by a large number of pixel points being skipped in the channel matrix during the skip connections in the

residual module. Using dual pooling to obtain salient and typical features of input images can not only enhance the expression ability of features but also help the network better capture important information in images and improve model classification performance. 3) The proposed FDPRNet inserts two modules—FRAM and DPR—into the residual network to enhance network channel interaction and feature expression capabilities, enabling the network model to capture complex relationships and strong generalization ability. It achieved high classification accuracy on some mainstream image classification datasets.

Key words: image classification; feature rearrangement; attention mechanism; residual network; deep learning

0 引 言

图像分类是计算机视觉领域的一项基本任务，其利用计算机技术对图像内容进行分析和识别，将图像分配到预定义标签或类别中，实现对图像的自动分类。图像分类在航拍图像分类、医学影像分析和行人属性识别等领域均有广泛应用价值。随着深度学习技术的快速发展，基于深度学习的图像分类方法取得诸多成果，为多种实际问题提供有效的解决方案。

基于深度学习的图像分类方法通常使用卷积神经网络(convolutional neural network，CNN)。CNN是一种高效的网络模型，通过堆叠的卷积层、池化层和全连接层逐步提取图像的特征信息。从20世纪40年代开始，人们一直没有停止对卷积神经网络的研究，提出多种深度卷积神经网络的模型。例如，Lecun等人(1998)提出的LeNet，采用一系列的卷积和池化层提取图像的特征，并在图像分类任务中取得不错的效果，为后续的卷积神经网络模型奠定了基础，推动了深度学习在图像处理领域中的发展。在LeNet的基础上，Krizhevsky等人(2017)提出AlexNet(imageNet classification with deep convolutional neural network)，通过更深的网络结构，引入ReLU(rectified linear unit)激活函数和dropout正则化技术，利用GPU(graphics processing unit)加速训练，显著提升了网络的表达能力。Simonyan和Zisserman(2015)提出VGGNet(Visual Geometry Group network)，通过统一网络架构，连续使用多个小卷积核替代一个大卷积核来提高网络的性能。Szegedy等人(2015)提出GoogLeNet(Google inception network)，采用了Inception模块，包含多个并行的卷积分支，使网络能够提取不同尺度的特征，提高了网络的表达能力。He等人(2016)提出ResNet(residual network)，提出残差的概念，残差块的设计解决了在深度网络中梯度消失和网络退化的问题，同时还提出了ResNet-101和ResNet-152使得在加深网络的同时还能够提高分类任务准确率。Xie等人(2017)提出的ResNeXt(residual neural network with cardinality)，在ResNet的基础上通过引入可拓展的网络结构，将多个路径并行组合在一起，拓宽网络的深度和宽度，提高了模型的表达能力。Zagoruyko和Komodakis(2017)提出Wide ResNet(wide residual network)，同样以ResNet为基础，通过宽度因子控制残差块内卷积层的通道数，更多的通道数可以提供更多的特征表示空间，使网络能够处理大规模的数据集。

虽然上述网络在处理图像分类任务时提高了模型的表达能力和性能，但是仍存在部分特征信息丢失的问题，导致模型无法有效利用关键特征，进而产生错误的推理和预测。针对该问题，人们从生物视觉角度出发，提出注意力机制，其原理为模仿图像通过人眼经过视神经的处理最后在人脑中成像的过程，强大的特征提取能力使其逐渐成为图像分类领域的研究热点。注意力机制的核心是将输入图像与一个权重向量相乘，该向量就代表了注意力权重，可以根据任务的需求自动调整注意力的分配，从而使网络能够关注图像信息，达到最优效果。在注意力机制的发展过程中，Hu等人(2018)提出SENet(squeeze-and-excitation networks)，引入了一种通道注意力机制，通过学习通道间的关系自适应地调整特征图中每个通道的权重，使网络关注重要的特征通道，减少无关通道的影响。Wang等人(2018)提出NLNet(non-local neural networks)，引入了一种空间注意力机制，允许网络在全局范围内建立像素之间的依赖关系，从而捕捉全局上下文信息。Zhou等人(2016)提出CAM(class activation mapping)，引入了一种可视化注意力机制，将网络的特征图与类别标签之间建立联系，有助于人们了解网络在图像中的

注意力分布。Huang 等人(2019)提出的 CCNet(criss-cross attention network)引入了交叉注意力机制,实现了像素之间的全局感知能力,更好地捕捉远距离位置的空间上下文信息。宋巍等人(2021)使用边卷积操作从动态图结构中提取局部特征,并通过注意力机制从邻域中聚合更具代表性的逐点特征。武斌等人(2024)提出一种结合空间结构卷积和注意力机制的点云分类网络,能够有效地实现特征适应性调整,具有较强的细粒度特征提取能力。上述注意力模块虽然可以帮助网络关注重要区域,提高网络捕捉特征的能力和模型的表现能力,但是均不能解决图像通道缺乏交互和联系的问题,限制了模型对关键特征的表达能力。

为了解决上述问题,本文提出特征重排列注意力机制的双池化残差分类网络 FDPRNet(double-pooling residual classification network of feature reordering attention mechanism)。FDPRNet 以 ResNet-34 作为基线模型,首先,针对图像通道之间固定的通道顺序可能导致特征偏重,无法充分利用其他通道信息的问题,提出特征重排列注意力机制(feature reordering attention mechanism,FRAM)模块,它实现了特征图内部之间通道交叉和信息交流的过程,更好地提取前后通道间的信息,同时增强特征之间的相关性;然后针对残差模块中通道矩阵上大量像素点被跳过所造成的特征信息丢失问题,提出双池化残差(double-pooling residual,DPR)模块,它通过使用平均池化和最大池化来提取输入图像的所有特征,能更好地捕捉图像的具体细节信息。

1 特征重排列注意力机制的双分支残差网络

1.1 整体网络结构

本文所提出的特征重排列注意力机制的双池化残差网络(FDPRNet)由 4 个融合 FRAM 的堆叠残差块(F-Block)和融合 DPR 的残差块(D-Block)串联组成,每个残差块内部由多个残差单元组成,网络结构如图 1 所示。主干网络部分包括如下几个阶段:1)对输入图像进行预处理操作,如随机裁剪、随机水平翻转图像等。将图像裁剪成指定大小,默认为 32 × 32像素,同时在裁剪过程中为保持图像大小进行边缘填充,然后通过反转图像增加数据多样性,使模型学习到不同位置和方向上的特征。2)将经过预处理后的图像输入网络的首层,通过 3 × 3 卷积在输入图像上进行卷积运算,从而提取图像中的局部特征,输入图像一般会有 3 个通道,首层卷积会将其拓展为更多的通道,有助于提供更为丰富的特征表示。3)将经过首层卷积输出的特征图分别经过 Layer1、Layer2、Layer3、Layer4 进行更为具体的特征提取和通道交互,Layer2、Layer3、Layer4 中再通过一组 D-Block 进行通道升维并缩小特征尺寸。4)利用平均池化层将上个阶段输出的特征图高度和宽度维度

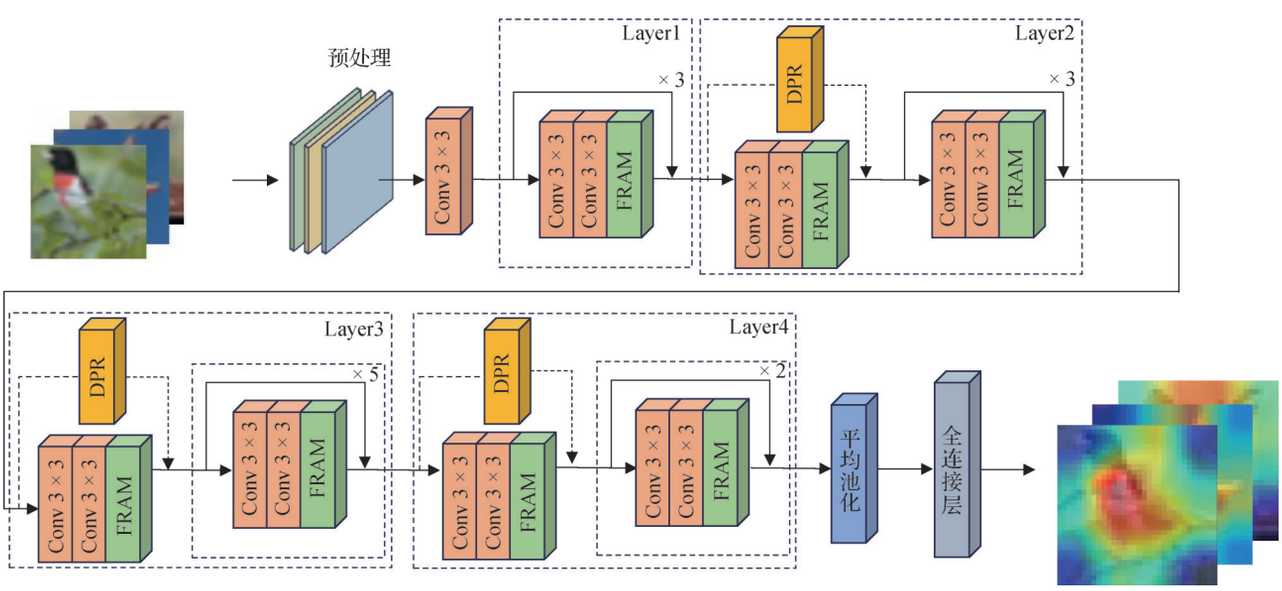


图 1 FDPRNet 总体结构

Fig. 1 Overall structure of FDPRNet

压缩为1，这样不仅能够减少参数量还能够保留每个特征图的整体特征信息。再使用全连接层将所得到的特征与类别之间建立映射关系，生成最终的分类结果。

1.2 首层特征提取模块

在ResNet-34网络中，对于大尺寸数据集来说，首先使用7×7大小卷积核和较大的步长，对输入图像进行初始特征提取，较大的步长可以更快地减少输入图像的尺寸，从而在更短的时间内捕捉到更多的特征。然后再通过最大池化层，保持特征图通道不变的情况下，减少特征图的空间维度。但是对于小尺寸数据集来说，由于使用过大的卷积核尺寸导致一些细节和空间信息可能会丢失，从而限制了最大池化层提取更高级别特征的能力。

以10×10像素的小尺寸图像举例，通过首层7×7卷积和卷积核大小为3的最大池化后，如图2(a)所示，卷积后的图像大小可表示为

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

式中，W表示输入特征图像尺寸，F表示卷积核尺寸，P表示填充大小，N表示输出特征图像尺寸。

根据式(1)可计算出输出特征图大小为3×3，相对于原始图像，经过卷积后所得到的图像尺寸缩小到原来的1/10。分辨率下降，导致无法捕捉到原始图像中微小细节和纹理信息，增加了过拟合的风险，从而影响模型的性能。

本文将首层卷积核大小为7×7的卷积替换为卷积核大小为3×3的卷积，如图2(b)所示，更小的卷积核可以更好地捕捉图像中的局部细节特征，缩小填充和步长意味着可以更细致地探测图像中的细微变化，更好地保留图像边缘的细节信息。删除最大池化层，意味着特征图尺寸不会再继续缩小，将会保留更多的空间信息，避免了因为池化操作而产生信息丢失，减少了计算量和模型的复杂性，避免过度平滑。同样对于10×10像素的小尺寸图像来说，根据式(1)可以计算出输出的图像大小还是10×10像素，极大地保留了图像的原始信息。

综上所述，对于原始图像较小的图像来说，大尺寸的卷积核并不能更好地进行特征提取，较大的卷积核会导致感受野变大，捕捉过多无用的上下文信息，这些信息对于较小的图像可能并不重要，同时影响池化层的特征提取。相反，较小尺寸的卷积核能够更好地捕捉局部特征，由于其感受野较小，能够精准地定位和提取图像中的信息，并且较小的卷积核还可以使输出的特征图保持较大的尺寸从而保留更多的特征信息，以便后续网络层更好地提取特征。

1.3 特征重排列注意力机制模块

为了增强特征之间的交互，更好地捕捉上下文信息和特征相关性，本文提出特征重排列注意力机制(FRAM)模块，FRAM结构如图3所示，输入特征图$\boldsymbol{F} \in \mathbf{R}^{C \times H \times W}$，其中，$C$为通道数，$H \times W$为特征图的空间大小。

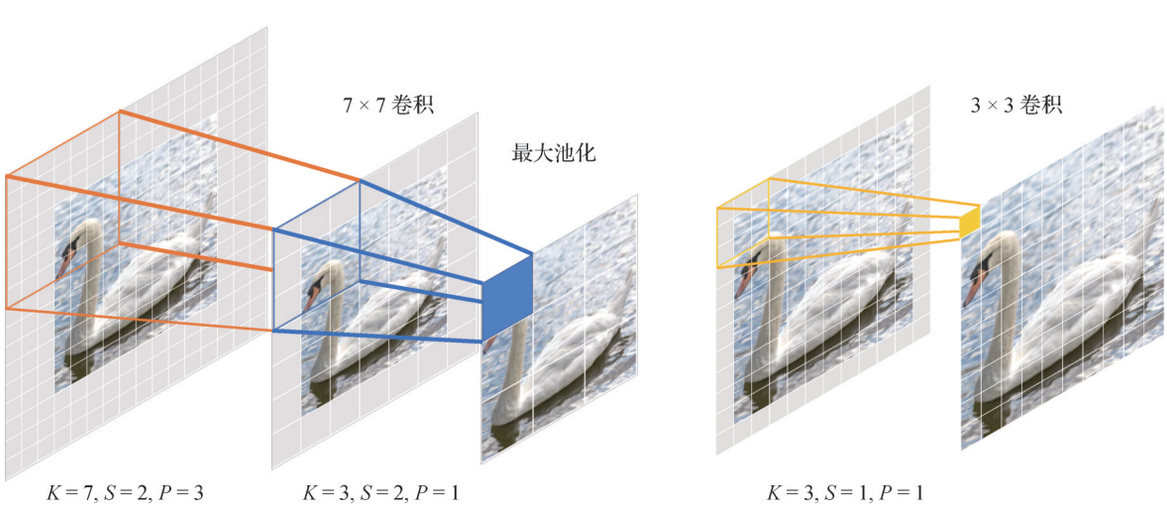


图2 网络首层卷积对比示意图

Fig. 2 Schematic diagram of network first layer convolution comparison

((a) ResNet-34 first layer convolution; (b) FDPRNet first layer convolution)

首先，将特征图通道进行通道分组，分组后可以表示为 $\boldsymbol{F}_1=\left[\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_3,\cdots,\boldsymbol{f}_n\right]$。其中，$\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_3,\cdots,\boldsymbol{f}_n$ 分别表示每组特征通道，$\boldsymbol{F}_1$ 表示分组后每组特征通道的集合，得到 n 组通道。每一组通道特征表示为 $\boldsymbol{f}_i\in\mathbf{R}^{\frac{C}{n}\times H\times W}$，其中，$i$ 表示 $\boldsymbol{F}_1$ 中某一个通道特征在集合中的位置，$\frac{C}{n}$ 表示 $\boldsymbol{f}_i$ 的通道数。通道分组操作对模块起到降低参数量，提高推理和训练速度的作用，并且不同组内的卷积核可以专注不同的特征子空间，有助于强化特征提取。其次，对已经分组完成的通道集合 $\boldsymbol{F}_1$ 进行随机打乱，调整通道间的前后顺序得到 $\boldsymbol{F}_2$，打乱过程表示为

$$\boldsymbol{F}_2=Random\left(\boldsymbol{F}_1\right)=Random\left(\left[\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_3,\cdots,\boldsymbol{f}_n\right]\right)\tag{2}$$

式中，$\boldsymbol{F}_2$ 表示特征通道组间重排序的集合。随机打乱通道可以减少过拟合的风险，增加模型的稳定性。

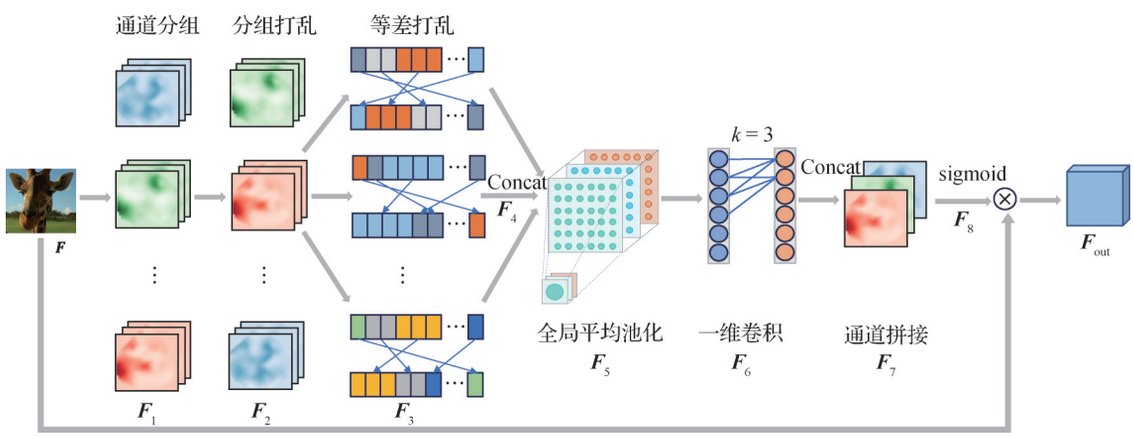


图 3　FARM 模块结构图

Fig. 3　FARM module structure diagram

然后，对 $\boldsymbol{F}_2$ 集合中每个组内部进行等差排列后再分组，根据等差数列求和公式，具体为

$$C_i=ma_1+\frac{\left[m(m-1)d\right]}{2}\tag{3}$$

式中，a_1 表示数列的首项，d 表示步长，C_i 为每组内的通道数，$C_i=\frac{C}{n}$，m 为等差数列的项数，所以每组中的通道特征 $\boldsymbol{f}_{ij}$ 可以表示为 $\boldsymbol{f}_{ij}=\left[\boldsymbol{f}_{i1},\boldsymbol{f}_{i2},\cdots,\boldsymbol{f}_{im}\right]$。其中，$j$ 表示 $\boldsymbol{f}_i$ 中某个子组通道特征在集合中的位置，由于等差数列采用首项 $a_1=1$、步长 $d=1$，所以根据式(3)可求出等差排列的项数 m，表示为

$$m=\frac{-1+\sqrt{\left(-1+8C_i\right)}}{2}\tag{4}$$

因为计算出的结果可能不为整数，所以将 m 向下取整，求得项数 M，再根据式(4)重新得到通道数 C_k，所以数列的最后一项 P 可以表示为 $P=C_i-C_k$。由于数列采用首项为 1、步长为 1 的递增方式，所以将每组的特征通道集合 $\boldsymbol{f}_{ij}$ 以 $[1,2,3,\cdots,P]$ 所表示的数列进一步分组，每个子组中的通道数取决于数列中的数字，进而得到子组的集合 $\boldsymbol{F}_{ij}$，然后对其进行随机打乱，得到子组重排列特征通道的集合 $\boldsymbol{F}_3$，可表示为 $\boldsymbol{F}_3=Random\left(\boldsymbol{F}_{ij}\right)=Random\left(\left[\boldsymbol{F}_{i1},\boldsymbol{F}_{i2},\boldsymbol{F}_{i3},\cdots,\boldsymbol{F}_{ip}\right]\right)$。这样可以使相邻的通道不再相连，将通道按照步长为 1 的等差数列分组，既可以打乱部分原始通道间的前后顺序，同时又可以保留部分通道间的前后关系，引入一定的随机性，使得模型能够全面地考虑不同通道之间的相互作用，避免过度依赖特定通道。这种操作使通道之间具有更好的异质性，从而增加模型对通道间相互作用的探索能力，增加信息交流，提取更加丰富的特征。再对 $\boldsymbol{F}_3$ 进行通道维度上的拼接得到 $\boldsymbol{F}_4\in\mathbf{R}^{\frac{C}{n}\times H\times W}$，可表示为 $\boldsymbol{F}_4=Concat\left(\boldsymbol{F}_3\right)=\left[\boldsymbol{K}_1,\boldsymbol{K}_2,\boldsymbol{K}_3,\cdots,\boldsymbol{K}_n\right]$。其中，$[\boldsymbol{K}_1,\boldsymbol{K}_2,\boldsymbol{K}_3,\cdots,\boldsymbol{K}_n]$ 表示 $\boldsymbol{F}_3$ 中每个子组集合的通道融合成为一个更为丰富的特征表示，更好地展示输入图像的复杂性和多样性。

接下来，对所有通道进行全局平均池化，将通道的空间维度缩减为 1×1，得到 $\boldsymbol{F}_5$，具体为

$$\boldsymbol{F}_5=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{F}_4(a,b)\tag{5}$$

式中，$H\times W$ 表示输入特征图的大小，$\boldsymbol{F}_4(a,b)$ 表示特征图在位置 (a,b) 上的值，$\boldsymbol{F}_5\in\mathbf{R}^{\frac{C}{n}\times1\times1}$ 表示全局平局

池化结果，此操作降低了训练的参数量，有效地聚合特征图的信息，然后使用一维卷积（1D convolution）对所得的全局平局池化结果进行特征提取，可表示为

$$\boldsymbol{F}_6 = Conv1d(\boldsymbol{F}_5) = Conv1d\left(\sum_{x=1}^{H}\sum_{y=1}^{W}\frac{[\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3, \cdots, \boldsymbol{K}_n]}{H \times W}\right) \tag{6}$$

式中，$Conv1d(\cdot)$表示一维卷积，与二维卷积（2D convolution）在图像处理上不同，如图4所示，K表示卷积核大小，P表示步长，一维卷积只在一个维度上滑动卷积核对输入数据进行局部特征提取，可以促进通道之间的特征交互，常用的通道注意力机制如SENet，其使用了两个非线性的全连接层来提取通道特征，这会导致推理阶段效率低并引起不必要的复杂性，而一维卷积具有局部感受野、参数共享和空间平移不变性等优点使得一维卷积在提取通道特征时更加高效。

对于给定输入特征序列$\boldsymbol{x}$和长度为E的卷积核q，一维卷积的输出特征序列$\boldsymbol{y}$计算为

$$\boldsymbol{y}[t] = \sum_{q=0}^{E-1}\left(\boldsymbol{x}[t+q] \times \boldsymbol{h}[q]\right) \tag{7}$$

式中，t为输出序列$\boldsymbol{y}$的索引，q为卷积核内的索引。$\boldsymbol{y}[t]$表示输出特征序列的第t个元素，$\boldsymbol{x}[t+q]$表示输入特征序列的第$t+q$个元素，$\boldsymbol{h}[q]$表示卷积核的第q个元素。经过n次的全局平均池化和一维卷积操作得到了n组特征输出$\boldsymbol{F}_6 \in \mathbf{R}^{\frac{C}{n}\times 1\times 1}$，将这些特征输出在通道维度上进行拼接，可以将不同特征的信息融合在一起得到$\boldsymbol{F}_7 \in \mathbf{R}^{C\times 1\times 1}$，可表示为$\boldsymbol{F}_7 = Concat(\boldsymbol{F}_6)$，使$\boldsymbol{F}_7$具有与输入特征$\boldsymbol{F}$相同的通道维度$C$。通过拼接多个特征输出，可以获得更全面、更丰富的特征表示，增强特征间的融合和交互，提高模型的表达能力。最后，通过sigmoid激活函数得到通道特征权重矩阵$\boldsymbol{F}_8$，可表示为$\boldsymbol{F}_8 = \text{sigmoid}(\boldsymbol{F}_7)$。sigmoid具体为

$$R(x) = \frac{1}{1+\exp(-x)} \tag{8}$$

式中，$\exp(-x)$表示以自然常数e为底的指数函数，将注意力映射到0到1之间的概率值，用于确定不同通道的重要性，较大的概率值反映了对应通道更加重要，较小的概率表示对应通道相对不重要，这种方式能更有效地捕捉通道中的重要部分，因为sigmoicl也是一种非线性函数，有助于引入非线性关系，学习更为复杂的特征。

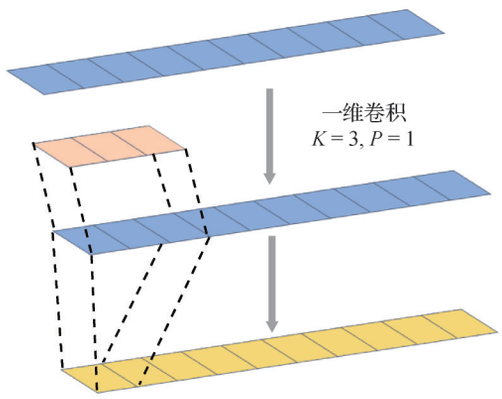


图4　一维卷积示意图

Fig. 4　Schematic diagram of one-dimensional convolution

最后，将通道特征权重矩阵$\boldsymbol{F}_8$与输入图像$\boldsymbol{F}$进行逐元素相乘得到特征重排列注意力机制的特征图$\boldsymbol{F}_{\text{out}}$，可表示为

$$\boldsymbol{F}_{\text{out}} = \boldsymbol{F} \times \boldsymbol{F}_8 = \boldsymbol{F} \times \text{sigmoid}\left(Concat\left(Conv1d\left(\sum_{x=1}^{H}\sum_{y=1}^{W}\frac{[\boldsymbol{K}_1, \boldsymbol{K}_2, \cdots, \boldsymbol{K}_n]}{H \times W}\right)\right)\right) \tag{9}$$

1.4　双池化残差网络模块

为了解决深层网络中的网络退化问题，不同于AlexNet和VGG的简单堆叠卷积层的方式，ResNet在网络中创新地引入了残差连接，通过跳跃连接将输入特征直接添加到输出特征上，有效地解决了深层网络训练过程中梯度爆炸和梯度消失的问题。

假设输入的图像特征为$\boldsymbol{x}$，残差块的输出特征为$\boldsymbol{F}(\boldsymbol{x})$，残差连接是将二者相加得到最后的输出特征$\boldsymbol{H}(\boldsymbol{x})$，可以表示为$\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{x}$，输入特征$\boldsymbol{x}$不经过复杂的变换和卷积操作，能够保留更加完整的特征信息，通过这种方式，网络可以更容易地学习到输入和输出之间的差异。

但是，ResNet中的残差连接仍然存在局限性，在通道维度发生变化的时候使用跳跃连接，为了匹配输入和输出的维度，通常使用步长为2、大小为1×1的卷积核进行卷积映射操作，如图5所示，为了使输出特征图的宽度和高度减半，步长设置为2。但是这意味着每次卷积都会跳过特征矩阵上的部分像素点，造成大量的特征信息丢失，丢失的信息包括纹理、形状及类别等重要属性的细节信息，使模型在后续网络中处理特征时产生错误的推理和预测，从而

降低模型的泛化能力和鲁棒性。如图5所示，以输入特征$\boldsymbol{A}_1$为例，$\boldsymbol{A}_1 \in \mathbf{R}^{3\times5\times5}$，经过升维卷积操作，提取图像特征，根据式(1)可以得出输出图像$\boldsymbol{A}_2 \in \mathbf{R}^{6\times3\times3}$，输出图像的空间分辨率仅为输入图像的1/3，意味着输入图像的通道矩阵将会有大量的像素点被跳过，如图5中浅色的方块所示。图像中的细节被模糊化，从而影响后续的网络层导致无法有效地提取特征。本文将所提出的双池化残差(DPR)模块放置在ResNet跳跃连接上，替换原有的1×1升维卷积映射，有效解决了上述残差模块中特征矩阵的大量像素点被跳过所造成的特征信息丢失问题。

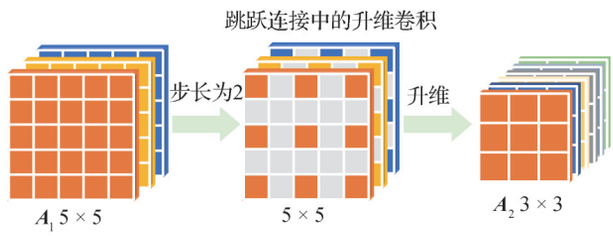


图5　原始残差卷积示意图

Fig. 5　Schematic diagram of original residual convolution

双池化残差模块DPR的结构如图6所示，输入特征$\boldsymbol{M} \in \mathbf{R}^{C\times H\times W}$，其中$(H, W)$为特征图的空间维度，$C$为通道数。DPR同时推导2个分支的二维特征矩阵，对输入图像提取不同层次的特征，提高特征的多样性，使得网络能够更好地理解图像的不同方面。

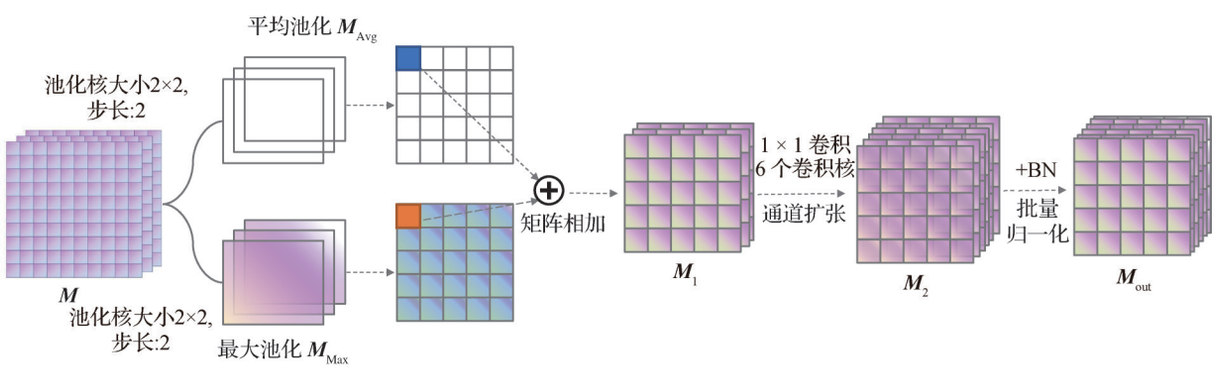


图6　双池化残差结构

Fig. 6　Dual branch residual structure

DPR模块首先对输入特征M进行卷积核大小为2×2、步长为2的池化操作，对输入特征图的每个窗口位置，分别取其通道上的平均值和最大值，得到$\boldsymbol{M}_{\text{Avg}}$和$\boldsymbol{M}_{\text{Max}}$，如图6所示。根据式(1)可以计算出输出特征$\boldsymbol{M}_{\text{Avg}}$、$\boldsymbol{M}_{\text{Max}} \in \mathbf{R}^{C\times\frac{H\times W}{2}}$。

平均池化操作为

$$\boldsymbol{M}_{\text{Avg}}(i,j)=\frac{1}{N}\sum_{m=i}^{i+K-1}\sum_{n=j}^{j+K-1}\boldsymbol{M}(m,n) \tag{10}$$

式中，$\boldsymbol{M}_{\text{Avg}}(i,j)$是平均池化后特征图上的元素，$\boldsymbol{M}(m,n)$是输入特征图上的元素，$K$是池化窗口的大小，$N$是窗口内的元素数，$i,j$是池化后特征图上的索引，$m,n$是输入特征图上对应窗口内的索引。

使用平均池化操作，如图7(a)所示，k表示卷积核大小，s表示步长。计算输入特征图上每个窗口内元素的平均值来减小特征图的尺寸，同时保留关键信息，减少局部变化和噪声的影响，并提取更加平滑的特征表示，提高模型的鲁棒性。步长为2的池化窗口对图像全局像素点进行扫描而不会对邻近的特征值进行重复计算。

最大池化操作可表示为

$$\boldsymbol{M}_{\text{Max}}(i,j)=\max_{m=i}^{i+K-1}\max_{n=j}^{j+K-1}\boldsymbol{M}(m,n) \tag{11}$$

式中，$\boldsymbol{M}_{\text{Max}}(i,j)$是最大池化后特征图上的元素，$\boldsymbol{M}(m,n)$是输入特征图上的元素，$K$是池化窗口的大小，$i,j$是池化后特征图上的索引，$m,n$是输入特征图上对应窗口内的索引。

使用最大池化操作遍历整个特征图，如图7(b)所示，k表示卷积核大小，s表示步长。每个窗口的最大值被选取为对应位置上的输出值，有助于保留输入数据中最显著的特征，在降低特征图尺寸的同时提供更具代表的特征表示。通过保留最为显著的特征并丢弃其他不重要的特征，最大池化大大减少提取特征的计算量，从而提高模型的效率。

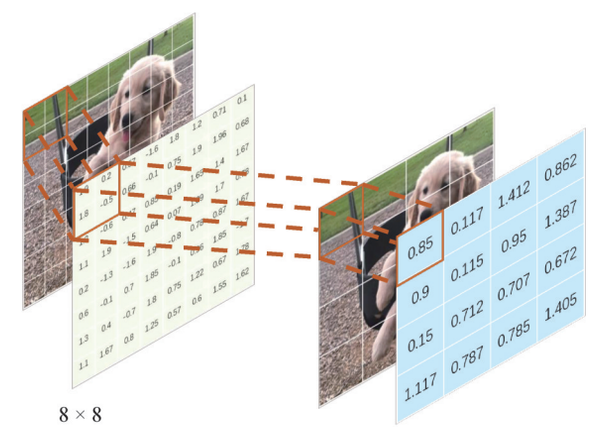
从两个分支上可以分别得到输入特征中每个

2×2窗口内的平均值和最大值集合。特征平均值是一组特征的算数平均值，用于描述当前特征的典型值，代表了特征的平均水平或集中趋势。特征最大值是每个2×2窗口内最显著的特征，可以用于检测图像中的边缘、亮点或者显著的目标物体。将两个不同分支上的特征矩阵进行逐元素相加得到$\boldsymbol{M}_1$，可以表示为$\boldsymbol{M}_1=\boldsymbol{M}_{\text{Avg}}+\boldsymbol{M}_{\text{Max}}$。

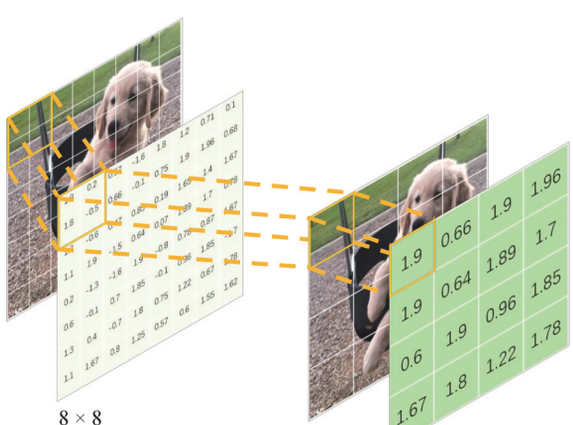
此操作既得到了输入图像的显著特征又得到了典型特征，可以增强特征的表达能力，有助于网络更好地捕捉图像中的重要信息，并提高模型性能。再将所得到的$\boldsymbol{M}_1$，通过步长为1、大小为1×1的卷积核进行升维卷积操作得到$\boldsymbol{M}_2$，可表示为

$$\boldsymbol{M}_2=Conv2d\left(\boldsymbol{M}_1\right)=Conv2d\left(\boldsymbol{M}_{\text{Avg}}+\boldsymbol{M}_{\text{Max}}\right) \tag{12}$$

式中，$Conv2d(\cdot)$表示二维卷积，$\boldsymbol{M}_2\in\mathbf{R}^{2C\times\frac{H\times W}{2}}$，在不改变特征图空间维度的情况下引入更多特征通道，学习更为丰富的特征表示，并且与主分支上的输出特征图维度保持一致，确保在残差连接中通道矩阵能够进行元素级相加操作。



(a) 平均池化



(b) 最大池化

图7　双池化示意图

Fig. 7　Schematic diagram of dual pooling ((a) average pooling; (b) maximum pooling)

随着神经网络加深，位于网络前端的参数经历多次线性和非线性变换，这可能导致它们的微小调整在传递过程中被逐渐扩大，因此，在反向传播过程中，这些浅层参数的梯度往往变得非常微小，从而减慢了参数的更新速度。这种现象可能会使得深层网络的训练过程变得更加困难，甚至难以达到收敛状态。针对此问题，对$\boldsymbol{M}_2$进行批量归一化操作(batch normalization，BN)得到$\boldsymbol{M}_{\text{out}}$，可表示为

$$\boldsymbol{M}_{\text{out}}=BN\left(\boldsymbol{M}_2\right)=BN\left(Conv2d\left(\boldsymbol{M}_{\text{Avg}}+\boldsymbol{M}_{\text{Max}}\right)\right) \tag{13}$$

通过对特征图进行归一化，可以减轻梯度消失和过拟合的影响，同时减少网络对权重初始化的依赖。

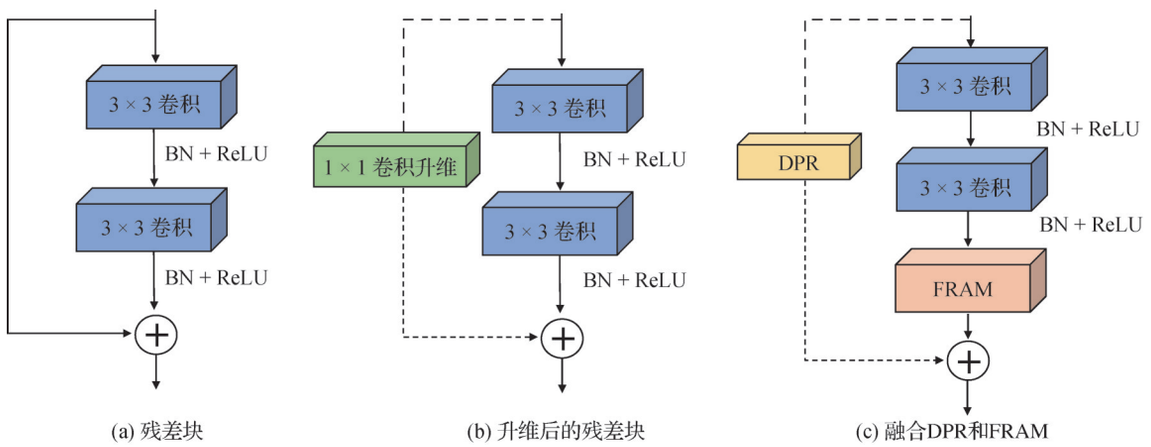
1.5　融合特征重排列注意力机制和双池化残差模块

在ResNet的残差块中，原始输入会通过一个跳跃连接绕过一部分层直接与经过卷积层的输出相加，从而在网络中引入了一个直通路径，以解决深层网络中信息丢失的问题，有利于保留原始的输入信息。原始残差块(简记为Block)结构如图8(a)(b)所示，通常由两个卷积层、批标准化层和激活函数组成。这些残差块中的卷积通常采用3×3大小的卷积核来提取特征。激活函数通常使用ReLU来引入非线性关系，在遇到输入输出维度发生变化时，跳跃连接使用1×1大小的卷积核进行升维后再与输出进行元素间的相加。

图8(c)为本文融合FRAM和DPR的残差块部分(简记为D-Block)，由2个3×3的卷积和FRAM以及残差分支(DPR)组成。FRAM模块通过学习不同顺序通道组合的权重，实现了特征图内部的通道交叉和信息交流，增强图像特征间的相互关系，自适应地调整对不同通道的注意力分配。跳跃连接中的DPR模块通过对输入图像的所有特征矩阵像素点进行更为具体的特征提取，取代原来造成大量特征信息丢失的跳跃连接。

综上所述，FDPRNet针对ResNet-34的改进如下：1)针对尺寸较小的输入图像，将ResNet-34首层卷积核大小为7×7的卷积替换为卷积核大小为3×3的卷积，保留更多特征信息，增加非线性能力，删除

最大池化层，获得更好的局部细节捕捉能力。2)为了实现特征图内部的信息交流和通道交叉，在残差块中添加FRAM模块，可以增强特征之间的相互作用，模型能够更好地捕捉上下文中的信息，并提高特征之间的交互性。3)为解决跳跃连接中特征矩阵的大量像素点被跳过所造成的特征信息丢失问题，通过引入DPR模块来替换掉原来简单的1×1升维卷积，对输入图像提取不同层次的特征，增加特征的多样性，从而使网络更全面地理解图像的内容。



(a) 残差块　(b) 升维后的残差块　(c) 融合DPR和FRAM

图8　3种残差块

Fig. 8　Three types of residual blocks ((a) residual block; (b) upgraded residual block; (c) fusion of DPR and FRAM)

2　实验及结果分析

2.1　实验环境

实验采用的操作系统为Linux5.15.133。训练所需的硬件环境参数为NVIDIA Tesla P100显卡、60 GB内存。软件环境为PyTorch2.0.0、CUDA11.4、CUDNN8.09。编程语言为Python3.10.12。模型的训练效果与CNN神经网络参数的设置密不可分。本模型使用Batch_size为128，训练的总周期为200(epochs)，初始学习率设置为0.1，分别在60、120、160(epochs)时将学习率缩减为原来的0.2倍，采用随机梯度下降(stochastic gradient descent，SGD)优化器。

在CIFAR-10(Canadian Institute for Advanced Research，10 classes)、CIFAR-100、SVHN(street view house numbers)、Flowers-102、NWPU-RESISC45(Northwestern Polytechnical University remote sensing image scene classification dataset)和STL-10等6个公共数据集上进行实验。

CIFAR-10是一个10种类别的数据集，包含常见的物体类别，如：飞机、汽车和猫等，其中每个类别有6 000幅彩色图像。CIFAR-100是一个100种类别的数据集，包含更具体的类别，如：不同的花朵、蔬菜等，其中每个类别有600幅彩色图像。SVHN是一个街景房屋号码的数据集，包含了从Google街景图像中提取的数字图像，用来识别房屋号码牌。Flowers-102包含了102个不同的花卉种类，每个种类大约有40～258幅图像，由于拍摄条件和角度不同，因此数据集中的图像具有不同的尺寸和分辨率。NWPU-RESISC45是一个45分类的数据集，包含了不同的地理环境，如：机场、体育场和高速公路等，每个类别包含700幅图像。STL-10数据集来自于ImageNet，经过裁剪和调整大小处理，最终形成10个类别的图像样本，涵盖了不同背景、光照条件、视角和姿势等。

具体数据如表1所示。本文在基本数据增强后的数据集进行实验，基本的数据增强包括填充随机裁剪、随机水平翻转、转换为张量和归一化。实验的主要评价指标为验证集的分类准确率。

2.2　首层特征提取模块实验

实验通过替换ResNet-34网络的第1层卷积核尺寸和删除最大池化层，探索对小尺寸数据集进行图像分类时模型性能的影响，首先将原始的7×7大小的卷积核替换为3×3大小的卷积核，并删除最大池化层。通过比对基线模型(ResNet-34)和改进的模型(缩小卷积核并删除最大池化层)在CIFAR-100数据集上的性能表现，可以评估模型改进的有效性。

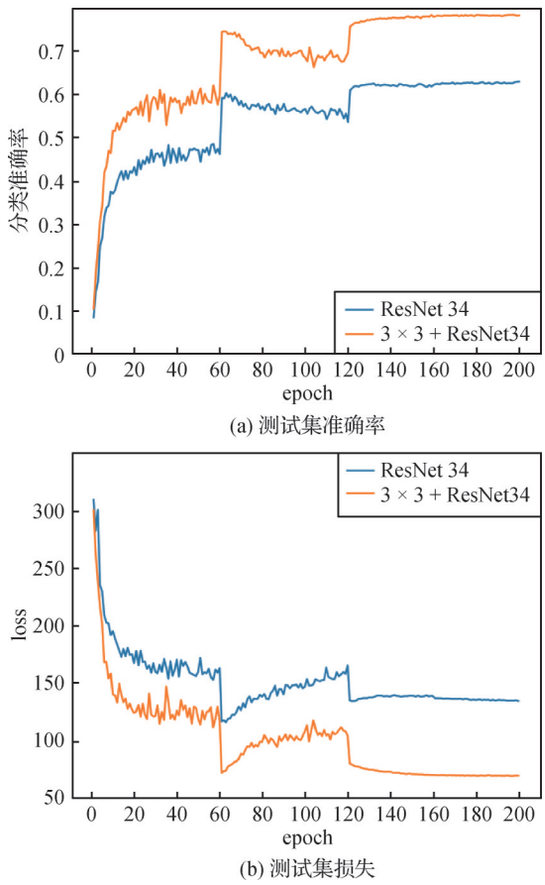
表1　实验数据集

Table 1　Experimental dataset

名称	图像尺寸/像素	分类数	训练集数量	测试集数量
CIFAR-10	32 × 32	10	50 000	10 000
CIFAR-100	32 × 32	100	50 000	10 000
SVHN	32 × 32	10	73 257	26 032
Flowers-102	96 × 96	102	14 740	15 153
NWPU-RESISC45	256 × 256	45	27 000	4 500
STL-10	96 × 96	10	5 000	8 000

本次实验使用上述相同的实验环境来训练两个模型，实验结果如图9所示。

从图9可以看出，改进的模型相较于基线模型有着较高的准确率，如图9(a)所示，改进模型的准确率达到了78.47%，而基线模型的准确率为63.01%，相比之下改进模型的准确率提升了15.46%。对于损失值来说同样有着显著的下降，如图9(b)所示，改进的模型下降了46.73。通过对比实验结果，发现针对小尺寸数据集较小的卷积核提供了更细微的感受野，使模型能够捕捉图像中的细节特征，通过删除最大池化层为模型保留了更多的空间信息。所以为了模型的性能能够达到最高，使用改进的模型作为后续实验的基线模型。



(a) 测试集准确率

(b) 测试集损失

图9　修改网络首层卷积对分类准确率的影响

Fig. 9　The impact of modifying the first layer convolution of the network on classification accuracy ((a) test set accuracy; (b) test set loss)

2.3　FRAM模块实验

2.3.1　FRAM插入位置实验

注意力机制插入不同的位置会对CNN神经网络的识别率产生不同影响，不恰当的插入位置可能会引入过多的模型复杂度和参数，导致模型过拟合，在新数据上的泛化能力变差，所以选择合适的插入位置能够提高网络的上下文理解能力和预测性能，帮助网络更好地提取输入图像的关键信息，从而提高准确率。为了比较插入FRAM注意力机制的位置和数量对准确率的影响，本文设计了6种不同的插入方案，分别是底层插入、中层插入、高层插入和多层插入，图10展示了不同Block的结构。

在确保训练参数一致的基础下，使用CIFAR-100、CIFAR-10和SVHN数据集在ResNet-34网络模

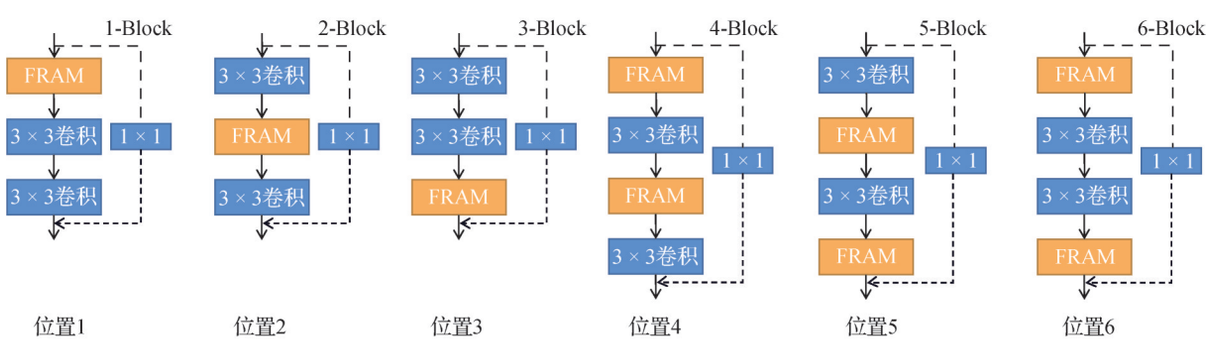


图10　FARM-Block位置和数量的排列结构图

Fig. 10　Arrangement structure diagram of FARM-Block position and quantity

型中训练和验证，实验结果分别如表2和图11所示。从实验结果可以看出，将FRAM模块插入在位置3中，在CIFAR-100、CIFAR-10和SVHN数据集上分别取得79.6%、95.59%和96.92%的分类准确率，均为最高值，说明将FRAM注意力机制插入在高层网络中效果最好。首先通过高层网络提取重要的语义特征，扩大感受野，从局部特征转化为全局特征，再利用FRAM注意力机制确定高层网络中最具有影响力的特征，以便网络可以自适应地关注输入图像的重点区域，忽略无关信息，提高了网络对关键特征的感知能力。因此，将FRAM模块放在位置3上能够有效地提升模型验证集的准确率，以最大程度发挥网络的优势。

表2　不同位置对准确率的影响

Table 2　The impact of accuracy at different positions

/%

位置	CIFAR-100	CIFAR-10	SVHN
1	77.98	95.30	96.79
2	77.94	95.43	96.71
3	**79.60**	**95.59**	**96.92**
4	77.63	95.37	96.71
5	78.98	95.55	96.59
6	78.37	95.46	96.79

注：加粗字体表示各列最优结果。

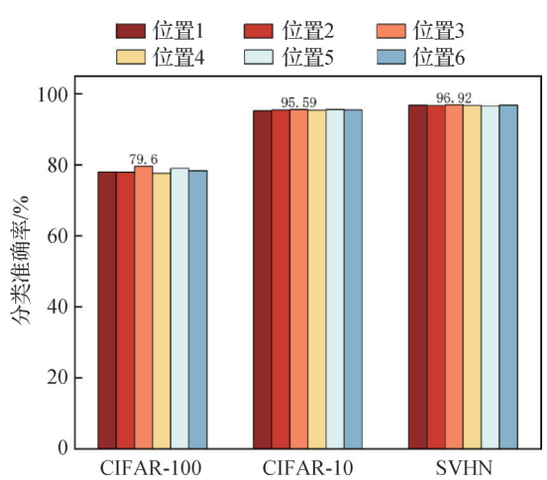


图11　注意力机制在不同位置对分类准确率对比

Fig. 11　Comparison of classification accuracy of attention mechanism at different positions

2.3.2　FRAM不同分组参数实验

超参数（hyperparameters）是深度学习和机器学习模型中的一种设置参数，它们决定了模型的结构和学习过程，选择一组合适的超参数会对网络有着巨大的提升效果。FRAM模块中的超参数为特征通道的分组数，较小的分组数目可以减少参数量和计算量，但也可能会导致特征通道间的信息交叉不充分，较大的分组数可以增加通道间的信息交流，但也会增加系统的负担。

为了深入探究FRAM的分组参数对网络模型准确率的影响，在2.3.1节中位置3的基础上设置5组不同的分组参数，分别为方案1（1-1-1-1）不进行分组、方案2（2-4-8-16）每组32通道、方案3（4-8-16-32）每组16通道、方案4（8-16-32-64）每组8通道、方案5（16-32-64-128）每组4通道。网络中一共有4个Layer层，如图1所示。其中Layer1有64通道数，Layer2有128通道数，Layer3有256通道数，Layer4有512通道数。以方案3为例，（4-8-16-32）分别表示每个Layer层中的分组数，将Layer1、Layer2、Layer3和Layer4各分成每组16通道，使不同组内的卷积核可以专注不同的特征子空间，有助于强化特征提取。用上述不同的5种方案分别在CIFAR-100、CIFAR-10和SVHN数据集进行训练和测试，并设置200个迭代轮次，使用基本的数据增强，根据最高准确率指标来选择最佳的分组参数取值。

表3和图12为不同分组参数对FRAMNet网络模型在3个数据集上的实验结果（分类准确率）。可以看出，分组参数并不是越大或者越小效果越好。较大的分组参数会导致计算资源消耗和内存占用增加，每个分组内通道数过少，缺乏多样性和差异性，从而限制了模型的表达能力；较小的分组参数意味着较多的通道被分到每个组中，无法充分打破原有通道之间的联系，缺乏通道之间的信息交互，导致无

表3　不同分组参数的实验结果（分类准确率）

Table 3　Experimental results of different grouping parameters（classification accuracy）

/%

方案	CIFAR-100	CIFAR-10	SVHN
1	78.63	95.45	96.80
2	79.24	95.55	96.70
3	**79.60**	**95.59**	**96.92**
4	79.23	95.30	96.56
5	79.34	95.59	96.70

注：加粗字体表示各列最优结果。

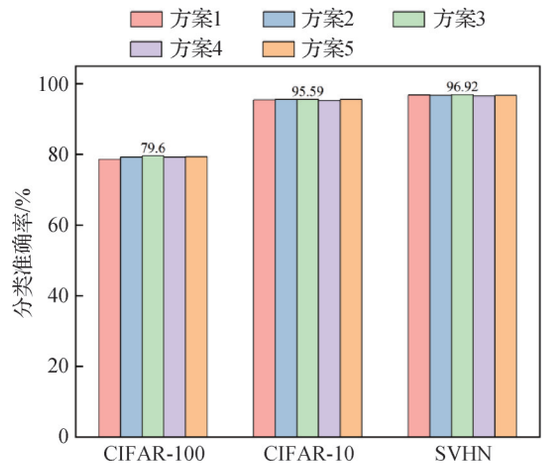


图12　不同分组参数对分类准确率的影响

Fig. 12　The impact of different grouping parameters on classification accuracy

法有效地学习到通道间不同组合的特征。

方案3(4-8-16-32)每组16通道,在CIFAR-100、CIFAR-10、SVHN数据集上分别取得79.60%、95.59%、96.92%的分类准确率,均为最高值。综上所得,16为最佳的分组参数,每个子组内再根据等差数列(1, 2, 3, 4, 5, 1)进行分组,然后随机打乱,这样能够减少模型对通道顺序的敏感性,使得不同组之间的通道产生交互,又不会让所有的通道之间失去前后联系,从而提高模型的泛化能力和鲁棒性,使其能够更好地适应新数据。

2.3.3　FRAM注意力机制对比实验

为了进一步验证FRAM模块对ResNet-34网络模型性能提升的有效性,选取不同类型注意力机制模块与本文FRAM进行对比分析。选取4种注意力机制作为对比分析模块,4种对比注意力机制分别是SENet(squeezeand excitation networks)、ECA(efficient channel attention for deep convolutional neural networks)(Wang等,2020)、CA(coordinate attention for efficient mobile network design)(Hou等,2021)、CBAM(convolutional block attention module)(Woo等,2018),各模块简介如表4所示。实验中具体操作为将4种注意力机制分别插入到ResNet-34网络中,在CIFAR-100、Flowers-102和STL-10数据集上进行训练,并对比分析其实验结果。

表4　注意力机制简介

Table 4　Introduction to attention mechanisms

注意力机制	简介
SE	SE引入通道注意力机制,动态地重新校准通道权重增强特征表示,使网络关注重要的特征通道,减少无关通道的影响。
ECA	ECA是对SE的改进,通过局部跨通道的交互直接从原始特征图学习通道权重,有效保留了通道间的重要信息,同时减少了模型复杂度,实现了更高效的通道注意力机制。
CA	CA同时考虑通道关系和位置信息的注意力模块,通过精确的位置信息对通道关系和长程依赖进行编码,引入纵向和横向的坐标信息增强模型对空间位置的感知能力,使得网络能够捕获长距离的依赖关系。
CBAM	CBAM是一种集成了空间和通道注意力的模块,能够在通道和空间两个维度上产生注意力特征图信息,用于增强卷积神经网络的特征表示能力。

依次在CIFAR-100、Flowers-102和STL-10数据集上测试不同注意力机制的分类准确率,由于Flowers-102和STL-10数据集的尺寸要比CIFAR-100更大,所以将ResNet-34的首层网络特征提取模块修改回7×7大小的卷积,并添加最大池化以应对大尺寸图像的特征提取。实验结果(分类准确率)如表5和图13所示,由实验结果可知,FRAM在3个数据集上的分类准确率均高于其他4种注意力机制模块。

由实验结果可知,SE和ECA的加入,使得网络模型的准确率有明显提升,通过引入通道维度注意力,学习通道重要性来提升模型表达能力,但是分类效果弱于FRAM,这是因为特征之间的关系不仅仅局限于单个通道特征,而是不同通道之间的特征相互影响和交互,这种跨通道的交互涉及非线性、高阶的关系,无法简单地通过ECA和SE来捕捉;加入CA的网络在Flowers-102数据集上分类准确率仅次于FRAM,其利用空间坐标信息,帮助模型更好地理解和利用图像中不同位置的特征,有助于提高模型对图像中重要区域的关注能力,但是相比于更简单的通道注意力机制,CA的实现相对复杂,需要处理横纵坐标,引入了额外的计算,增加了过拟合风险;CBAM虽然引入空间和通道注意力,但是相较于

表 5　引入不同注意力机制的实验结果(分类准确率)

Table 5　Experimental results of introducing different attention mechanisms(classification accuracy)

/%

网络	CIFAR-100	Flowers-102	STL-10
ResNet-34	77.94	82.88	69.97
+SE	79.26	86.43	70.62
+ECA	78.87	85.08	73.62
+CA	78.65	86.55	71.55
+CBAM	78.73	85.45	73.75
+FRAM	**79.60**	**87.16**	**73.85**

注:加粗字体表示各列最优结果;“+”代表引入对应注意力机制。

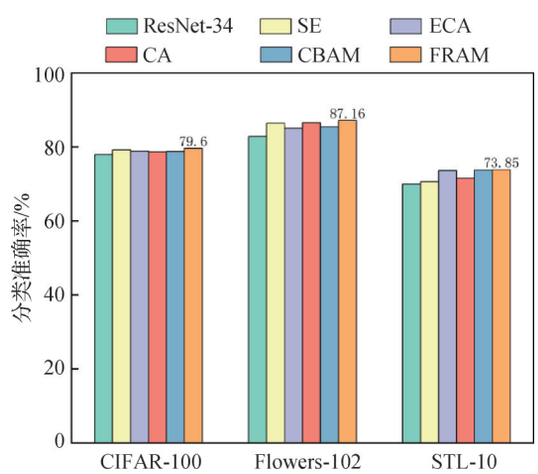


图 13　不同注意力机制对分类准确率的影响

Fig. 13　The impact of different attention mechanisms on classification accuracy

FRAM 在 3 个数据集上的准确率分别低 0.87%、1.71%和 0.1%,这是因为空间和通道注意力是同时引入的,这意味着它们会相互影响,当空间注意力和通道注意力在特定任务上的贡献不一致时,两种注意力机制之间的相互干扰可能会导致性能下降。

与其他 4 种注意力机制相比,FRAM 在 3 个数据集上分别平均提升 0.72%、1.28%和 1.46%。说明本文所提出的特征重排列注意力模块 FRAM 能够在不同数据集上提供更好的性能,FRAM 实现了跨通道的重新交互,在特征图内部之间进行通道交叉和信息交流,更好地提取前后通道间的信息,同时增强不同通道特征间的交互能力。

2.4　DPR 模块实验

通过修改残差结构,改进图像分类模型的性能,如图 8(b)(c)所示,将 ResNet 中的跳跃连接替换为 DPR 模块。为了验证本文提出的双池化残差模块(DPR)的有效性,选取 ResNet-34 作为基线模型进行比较分析,将网络首层替换为卷积核大小为 3×3 的卷积,并删掉最大池化层,使模型性能达到最佳。使用 CIFAR-100、CIFAR-10 和 SVHN 数据集进行测试,其他参数条件同上述实验一致。

3 个数据集上的实验结果如图 14 所示,DPRNet

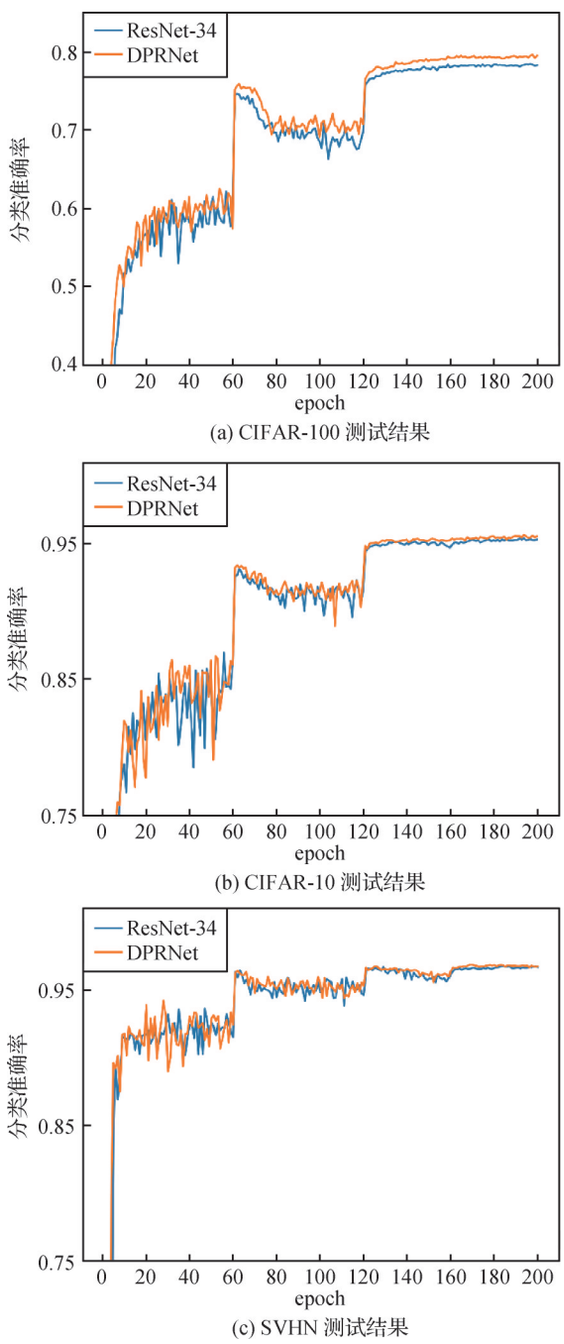


图 14　双分支残差(DPR)模块测试结果

Fig. 14　Test results of dual branch residual (DPR) module ((a) CIFAR-100 test results; (b) CIFAR-10 test results; (c) SVHN test results)

在CIFAR-100、CIFAR-10和SVHN数据集上的分类准确率都要高于ResNet-34，分别为79.64%、95.66%和96.91%，对比ResNet-34分别提升了1.7%、0.26%和0.12%。综合以上数据可以表明，DPR模块有效地解决了ResNet跳跃连接特征丢失的问题，改变了特征的传递方式，保持跳跃连接中特征的完整性。通过同时获得输入图像的显著特征和典型特征，可以增强特征的表达能力，从而帮助网络更有效地捕捉图像中的重要信息，并提升模型的性能。

2.5 消融实验

为了验证DPR模块和FRAM模块对FDPRNet网络的性能的影响，本文设计了FDPRNet消融实验，将DPR模块、FRAM模块分别加入和共同加入到ResNet-34网络中，并与ResNet-34网络进行对比分析。为了保证对比公平性，使用5种不同的数据集，其中CIFAR-100、CIFAR-10、SVHN为小尺寸数据集，Flowers-102为多种尺寸并存的数据集，NWPU-RESISC45为大尺寸数据集。在图像预处理中，将Flowers-102数据集裁剪为96×96像素，由于Flowers-102和NWPU-RESISC45数据集的尺寸都要比CIFAR-100、CIFAR-10、SVHN更大，所以在ResNet-34网络模型中将首层网络特征提取模块修改回7×7大小的卷积并添加最大池化以应对大尺寸图像的特征提取。此次消融实验的衡量标准为验证集的最高分类准确率，实验结果如表6所示。

由表6可知，与原模型ResNet-34相比，在CIFAR-100、CIFAR-10、SVHN、Flowers-102和NWPU-RESISC45数据集中，ResNet-34+FRAM的准确率分别提升了1.66%、0.19%、0.13%、4.28%和2.00%；ResNet-34+DPR的准确率分别提升了1.70%、0.26%、0.22%、3.18%和1.31%；ResNet-34+FRAM+DPR的准确率分别提升了2.07%、0.30%、0.17%、8.31%和2.47%。与原模型ResNet-34相比，ResNet-34+FRAM+DPR的准确率提升幅度最大，由于模型在分类更多的数据集上能够学习到更丰富的特征表示，所以在CIFAR-100、Flowers-102、NWPU-RESISC45上提升效果最明显，CIFAR-10和SVHN数据集相对于CIFAR-100数据集样本量大一些，且种类较少。因此，几种模型的整体分类精度都很高，导致准确率提升不大。综上可得，无论是在小尺寸还是大尺寸、分类较少还是分类较多的数据集上，FRAM和DPR模块均对ResNet-34网络的识别率起到提升作用，并且两种模块的组合（即FDPR）对网络识别率的提升效果更佳。

表6 引入不同模块的实验结果（分类准确率）

Table 6 Experimental results of introducing different modules (classification accuracy)

/%

模型	CIFAR-100	CIFAR-10	SVHN	Flowers-102	NWPU-RESISC45
ResNet-34	77.94	95.40	96.79	82.88	90.04
ResNet-34+FRAM	79.60	95.59	96.92	87.16	92.04
ResNet-34+DPR	79.64	95.66	96.91	86.06	91.35
ResNet34+FRAM+DPR	**80.01**	**95.7**	**96.96**	**91.19**	**92.51**

注：加粗字体表示各列最优结果；“+”表示添加对应模块。

2.6 不同网络层数对实验结果的影响

为了适应图像数据的复杂性，考虑选择主流的残差卷积神经网络模型，如ResNet-18、ResNet-34、ResNet-50和ResNet-101作为基线模型，将FRAM和DPR模块共同加入到这4种网络中，分别采用相同的实验环境和实验参数进行测试，在CIFAR-100数据集上进行训练，得到的分类结果如图15所示。

由图15可知，嵌入了两个模块的FDPRNet-34在CIFAR-100数据集上的分类准确率最高，为80.01%，分别比FDPRNet-18、FDPRNet-50和FDPRNet-101提高1.13%、6.39%和38.82%。这意味着增加网络层不一定能够带来更好的分类性能，相较于更深层的网络，FDPRNet-34能够更好地捕捉图像中的特征，从而取得更高的准确率。因此选择ResNet-34作为本文的骨干网络模型。

2.7 不同数据增强对实验结果的影响

数据增强主要通过对原始图像进行多种变换，以增加数据多样性和数量，从而扩充原始数据集的

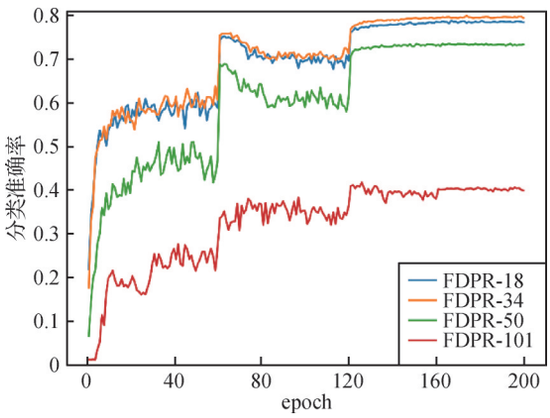


图 15　不同网络层数对分类准确率的影响

Fig. 15　The impact of different network layers on classification accuracy

规模，有助于增强模型泛化能力。为探究不同数据增强方法对网络模型产生的影响，选择以下 2 种数据增强方法。TrivialAugmentWide（Müller 和 Hutter，2021）从一组数据增强函数的集合中随机选取一个数据增强函数，使模型能够更好地学习到不同变体的图像特征；Mixup（Zhang 等，2018）将两幅不同图像的标签和特征进行混合，生成新的训练样本。将 TrivialAugmentWide 和 Mixup 分别加入到训练过程中，实验结果（分类准确率）如表 7 所示。

表 7　引入不同数据增强的实验结果（分类准确率）

Table 7　Experimental results of introducing different data augmentations（classification accuracy）

/%

数据增强方法	CIFAR-100	CIFAR-10	SVHN
FDPRNet	80.01	95.7	96.96
FDPRNet+TrivialAugmentWide	82.03	96.94	97.46
FDPRNet+Mixup	81.73	96.17	97.09
FDPRNet+TrivialAugmentWide+Mixup	**82.76**	**97.13**	**97.65**

注：加粗字体表示各列最优结果；“+”表示引入对应方法。

由表 7 可知，通过引入 TrivialAugmentWide 和 Mixup 这两种数据增强技术，FDPRNet 模型的准确率得到了显著提升。与原模型 FDPRNet 相比，在 CIFAR-100、CIFAR-10 和 SVHN 数据集中，FDPRNet + TrivialAugmentWide + Mixup 的准确率提升幅度最大，分别提升了 2. 75%、1. 43% 和 0. 69%。它们的组合可以增强模型特征学习能力，帮助模型更好地应对输入图像变化和噪声干扰，同时避免过度数据增强影响网络的实验结果。综上所述，两种数据增强技术在相互配合下能够带来更好的性能提升。

2. 8　对比实验

为了验证 FDPRNet 的有效性，与常见的图像分类网络模型 ResNet-34、EfficientNets（Tan 和 Le，2020）、CAPR-DenseNet（classification and attribute prediction with residual dense networks）（Zhang 等，2019）、GhostNet（Han 等，2020）、RTSA Net-101（improved resNet image classification model based on tensor synthesis attention）（邱云飞 等，2023）、GLS-GAN（generative least squares generative adversarial networks）（Qi，2020）、NNCLR（nearest neighbor contrastive learning of visual representations）（Dwibedi 等，2021）、CCT-6/3x1（Hassani 等，2022）、Couplformer（Lan 等，2021）、MMA-CCT-7/3×2（multi-column convolutional neural networks with 7 layers and 3 × 2 convolution kernels）（Konstantinidis 等，2023）和 DAMS-Net（double-branch multi-attention mechanism based sharpness-aware classification network）（姜文涛 等，2023）在 CIFAR-100、CIFAR-10 和 SVHN 数据集上进行对比实验。本文中对分类准确率的比较实验数据来源如下：对于未开源的代码，优先采用对比网络所对应论文提供的实验结果；对于开源代码通过使用对比网络所对应的论文提供的开源代码进行复现实验。各网络在 3 个数据集上的分类准确率如表 8 所示。

通过对比表 8 中所有模型的分类准确率可知，FDPRNet 在不同的分类数据集上均优于其他 11 种网络模型。其中 FDPRNet 在 3 个数据集上具有最佳的识别效果，其最高准确率分别为 82. 76%、97. 13% 和 97. 65%。对比其他 11 种网络，准确率平均提升了 4. 78%、2. 57% 和 2. 47%。

准确率次优的网络是 RTSA Net-101 和 DAMS-Net，它们在 3 个数据集的准确率分别为 81. 6%、96. 12%、96. 67% 和 80. 50%、96. 51%、94. 26%。这两个网络和 FDPRNet 都以 ResNet 残差网络为基础加以改进。RTSA Net-101 在网络中嵌入了张量合成注意力模块，能够聚焦于贡献度高的特征，提高分类准确率，DAMSNet 通过双分支多注意力机制模块，减少非关键特征所造成的影响，使网络高效地定位到目标区域。但是上述网络均没能解决残差网络中部分特征信息丢失的问题，导致分类效果低于

表8　不同网络在3种数据集上的分类准确率

Table 8　Classification accuracy of different network on three datasets

/%

网络	CIFAR-100	CIFAR-10	SVHN
ResNet-34	77.94	95.4	96.79
EfficientNets	75.96	94.01	93.32
CAPR-DenseNet	78.84	94.24	94.95
GhostNet	77.15	94.92	93.86
RTSA Net-101	81.60	96.12	96.67
GLS-GAN	–	91.70	94.02
NNCLR	79.00	93.70	–
CCT-6/3x1	77.31	95.29	–
Couplformer	73.92	93.54	94.26
MMA-CCT-7/3×2	77.50	94.74	–
DAMSNet	80.50	96.51	97.60
FDPRNet	**82.76**	**97.13**	**97.65**

注：加粗字体表示各列最优结果；“–”表示未知结果。

FDPRNet。

EfficientNets在3个数据集上的准确率分别为75.96%、94.01%和93.32%，其通过在不同维度上对模型进行统一的缩放，在准确性和计算效率之间找到了一个最佳的平衡点，但是由于受到复合系数的影响效果不理想。

MMA-CCT-7/3×2和Couplformer均为采用Transformer自注意力机制的网络模型。尽管在某些数据集上较GLS-GAN有所提升，但是自注意力机制缺乏位置信息并且对噪声和异常值敏感，导致网络泛化能力较差。

其他网络虽然在3个数据集上的表现效果不错，但通过对比发现，对于通道特征信息交互以及残差块中通道矩阵上特征信息丢失的问题，FDPRNet效果最佳，这证明了本文所提的FDPRNet模型在图像分类方向上具有一定的实用性和有效性。

2.9　特征可视化分析

为了进一步说明FDPRNet网络特征提取的变化，选择MobileNet_V2（Sandler等，2018）、ShuffleNet（Zhang等，2017）和ResNet-34网络进行对比。随机选择5幅种类不同的图像，分别对各网络的最后一个卷积层进行Grad-CAM（gradient-weighted class activation mapping）（Selvaraju等，2017）特征可视化，如图16所示。网络模型可以提取出不同种类图像的主要特征，并通过不同种类动物或物品独有的特征进行识别，蓝色一般来说表示低权重而红色表示高权重，对于分类任务，红色区域中的特征对模型的预测起着关键作用，而蓝色区域中的特征对模型的预测贡献度较低。

由图16（b）（c）可以清晰地观察到，MobileNet_V2和ShuffleNet的热力图显示效果不佳，由于它们采用了轻量级的网络结构，无法捕捉到细微和更复杂的特征，导致热力图中红色区域只将部分关键特征所覆盖。相较于MobileNet_V2和ShuffleNet，ResNet-34的网络层更深，如图16（d）所示，能够更好地捕捉与分类任务相关的关键特征，红色区域能够将图像中的大部分特征所覆盖，微小的细节被忽略。

由特征可视化分析可以看到，FDPRNet通过使用FRAM注意力机制实现了特征图内部的信息交流和通道交叉，能够更加高效地利用关键特征，DPR模块改善了ResNet-34在跳跃连接时特征丢失的情况。相比于其他的网络，如图16（e）所示，FDPRNet能够注意到马的腿，鹿的耳朵等细节特征，减少周边冗余特征信息。证明FDPRNet能够更有效地提取不同种类的图像特征。

3　结　论

本文提出了一种高效图像分类网络—特征重排列注意力机制的双池化残差分类网络（FDPRNet）。首先修改网络首层卷积核尺寸，删除最大池化层，减少特征丢失，极大地保留原始图像的特征信息。然后通过特征重排列注意力机制（FRAM）模块，实现了特征图内部的信息交流和通道交叉，有效利用通道特征，增强模型对通道排列关系的关注程度，解决了图像通道之间固定的通道顺序可能导致特征偏重的问题。最后将双分支残差（DPR）模块放置在残差块的跳跃连接上，分别对输入图像中特征矩阵的所有像素点提取不同层次的特征，提高特征的多样性，解决了残差块中局部特征信息丢失问题。FRAM和DPR模块使网络具备更强的表达能力和鲁棒性。

FDPRNet与原ResNet-34相比，在CIFAR-100、CIFAR-10、SVHN、Flowers-102和NWPU-RESISC45数据集上进行对比实验，准确率分别提升2.07%、

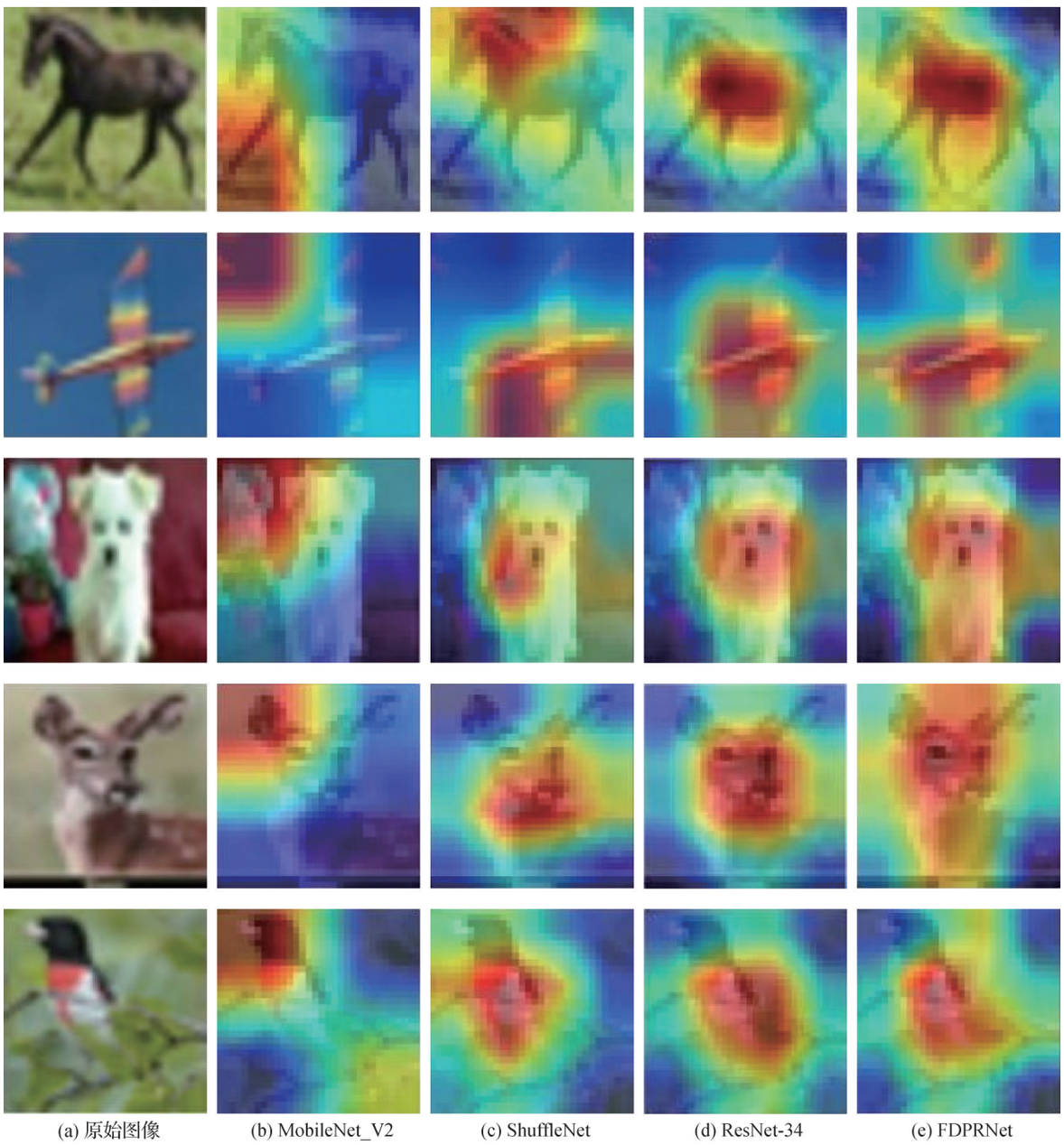


图 16　不同网络的可视化图像

Fig. 16　Visualization images of different networks

((a) original images; (b) MobileNet_V2; (c) ShuffleNet; (d) ResNet-34; (e) FDPRNet)

0.30%、0.17%、8.31% 和 2.47%。实验结果表明，在使用 FRAM 和 DPR 模块后，网络不但能够更好地捕捉关键特征，而且还能够关注到特征图中的细节特征，进一步提升了网络模型的鲁棒性。

然而，本文网络也存在不足之处，在增加 FDPRNet 的网络深度时会导致网络的性能下降，准确率降低。在接下来的工作中，一方面将使用 FRAM 和 DPR 模块提升网络的图像分类能力；另一方面，将改进 FDPRNet 网络，在加深网络的同时提高准确率，进一步优化模块结构，拓展其应用范围，将模型应用到更多的计算机视觉任务上。

参考文献(References)

Dwibedi D, Aytar Y, Tompson J, Sermanet P and Zisserman A. 2021. With a little help from my friends: nearest-neighbor contrastive learning of visual representations [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/2104.14548.pdf

Han K, Wang Y H, Tian Q, Guo J Y, Xu C J and Xu C. 2020. Ghost-

Net: more features from cheap operations//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1577-1586 [DOI: 10.1109/cvpr42600.2020.00165]

Hassani A, Walton S, Shah N, Abuduweili A, Li J C and Shi H. 2022. Escaping the big data paradigm with compact transformers [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/2104.05704.pdf

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hou Q B, Zhou D Q and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang Z L, Wang X G, Huang L C, Huang C, Wei Y C and Liu W Y. 2019. CCNet: criss-cross attention for semantic segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 603-612 [DOI: 10.1109/iccv.2019.00069]

Jiang W T, Zhao L L and Tu C. 2023. Double-branch multi-attention mechanism based sharpness-aware classification network. Pattern Recognition and Artificial Intelligence, 36(3): 252-267 (姜文涛, 赵琳琳, 涂潮. 2023. 双分支多注意力机制的锐度感知分类网络. 模式识别与人工智能, 36(3): 252-267) [DOI: 10.16451/j.cnki.issn1003-6059.202303005]

Konstantinidis D, Papastratis I, Dimitropoulos K and Daras P. 2023. Multi-manifold attention for vision transformers [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/2207.08569.pdf

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]

Lan H, Wang X H and Wei X. 2021. Couplformer: rethinking vision transformer with coupling attention map [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/2112.05425.pdf

Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Müller S G and Hutter F. 2021. TrivialAugment: tuning-free yet state-of-the-art data augmentation [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/2103.10158.pdf

Qi G J. 2020. Loss-sensitive generative adversarial networks on lipschitz densities. International Journal of Computer Vision, 128(5): 1118-1140 [DOI: 10.1007/s11263-019-01265-2]

Qiu Y F, Zhang J X, Lan H and Zong J X. 2023. Improved resnet image classification model based on tensor synthesis attention. Laser and Optoelectronics Progress, 60(6): #0610008 (邱云飞, 张家欣, 兰海, 宗佳旭. 2023. 融合张量合成注意力的改进ResNet图像分类模型. 激光与光电子学进展, 60(6): #0610008) [DOI: 10.3788/LOP212836]

Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]

Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/1409.1556.pdf

Song W, Cai W Y, He S Q and Li W J. 2021. Dynamic graph convolution with spatial attention for point cloud classification and segmentation. Journal of Image and Graphics, 26(11): 2691-2702 (宋巍, 蔡万源, 何盛琪, 李文俊. 2021. 结合动态图卷积和空间注意力的点云分类与分割. 中国图象图形学报, 26(11): 2691-2702) [DOI: 10.11834/jig.200550]

Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]

Tan M X and Le Q V. 2020. EfficientNet: rethinking model scaling for convolutional neural networks [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/1905.11946.pdf

Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]

Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/1711.07971.pdf

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module [EB/OL]. [2024-02-04]. http://arxiv.org/pdf/1807.06521.pdf

Wu B, Liu Y A and Zhao J. 2024. Classification network for 3D point cloud based on spatial structure convolution and attention mechanism. Journal of Image and Graphics, 29(2): 520-532 (武斌, 刘溢安, 赵洁. 2024. 结合空间结构卷积和注意力机制的三维点云分类网络. 中国图象图形学报, 29(2): 520-532) [DOI: 10.

11834/jig.230137］

Xie S N，Girshick R，Dollár P，Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Honolulu，USA：IEEE：5987-5995［DOI：10.1109/CVPR.2017.634］

Zagoruyko S and Komodakis N. 2017. Wide residual networks［EB/OL］.［2024-02-04］. http://arxiv.org/pdf/1605.07146.pdf

Zhang H Y，Cisse M，Dauphin Y N and Lopez-Paz D. 2018. mixup：beyond empirical risk minimization［EB/OL］.［2024-02-04］. http://arxiv.org/pdf/1710.09412.pdf

Zhang K，Guo Y R，Wang X S，Yuan J S，Ma Z Y and Zhao Z B. 2019. Channel-wise and feature-points reweights densenet for image classification//Proceedings of 2019 IEEE International Conference on Image Processing（ICIP）. Taipei，China：IEEE：410-414［DOI：10.1109/ICIP.2019.8802982］

Zhang X Y，Zhou X Y，Lin M X and Sun J. 2017. ShuffleNet：an extremely efficient convolutional neural network for mobile devices［EB/OL］.［2024-02-04］. http://arxiv.org/pdf/1707.01083.pdf

Zhou B L，Khosla A，Lapedriza A，Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas，USA：IEEE：2921-2929［DOI：10.1109/CVPR.2016.319］

作者简介

袁姮，女，副教授，主要研究方向为图像与视觉信息计算、模式识别与人工智能。E-mail：yuanheng@lntu.edu.cn

刘杰，通信作者，男，硕士研究生，主要研究方向为图像处理、模式识别和人工智能。E-mail：2269224774@qq.com

姜文涛，男，副教授，主要研究方向为模式识别与人工智能、图像与视觉信息计算。E-mail：jiangwentao@lntu.edu.cn

刘万军，男，教授，主要研究方向为软件工程理论、图像与视觉信息计算、模式识别与人工智能。

E-mail：liuwanjun@lntu.edu.cn