图文信息技术

基于深度学习的图像分类算法框架研究

罗雪阳，蔡锦达
（上海理工大学 出版印刷与艺术设计学院，上海 200000）

摘要：目的 提高图像分类精度是实现自动化生产的基础，提出一种更加准确的图像分类方法，使自动化包装和生产更加高效。**方法** 基于 ResNeSt 特征图组的思想，通过引入通道域和空间域注意力机制，并将自适应卷积核思想和 Gem 池化引入空间域注意力模块，从而使网络在空间域注意力机制中能够对不同图片使用不同的感受野使其关注更重要的部分，提出一种具有通道域和空间域注意力机制，且具有很好移植性的图像分类网络模型结构。**结果** 文中方法提高了图像分类准确度，在 ImageNet 数据集上，top-1 准确度为 81.39%。**结论** 文中提出的 ResNeSkt 算法框架优于目前的主流图像分类方法，同时网络整体结构具有很好的移植性，可以作为图像检测、语义分割等其他图像研究领域的主干网络。

关键词：ResNeSkt；图像分类与识别；包装和生产；图像检测；注意力机制
中图分类号：TP391.4 **文献标识码：**A **文章编号：**1001-3563(2021)21-0181-07
DOI：10.19554/j.cnki.1001-3563.2021.21.025

Framework of Image Classification Algorithm Based on Deep Learning

LUO Xue-yang, CAI Jin-da
(College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200000, China)

ABSTRACT: Improving the accuracy of image classification is the basis of automatic production. The work aims to proposes a more accurate image classification method to make automatic packaging and production more efficient. Based on the idea of ResNeSt feature graph group, by introducing the channel domain and spatial domain attention mechanism and introducing the idea of adaptive convolution kernel and gem pooling into the spatial domain attention module, the network could use different sensory fields for different pictures in the spatial domain attention mechanism to focus on more important parts. An image classification network model structure with channel domain and spatial domain attention mechanism and good portability was proposed. This method improved the accuracy of image classification. On ImageNet data set, the accuracy of top-1 was 81.39%. The ResNeSkt algorithm framework proposed in this paper is superior to the current mainstream image classification methods. At the same time, the overall network structure has good portability, and can be used as the backbone network in other image research fields such as image detection and semantic segmentation.
KEY WORDS: ResNeSkt; image classification and recognition; packaging and production; image detection; attention mechanism

图像分类在计算机视觉研究中是很多研究的基础以及关键，图像分类训练的深度学习网络通常可以用作其他相关研究的神经网络主体结构，例如图像检测[1—3]、语义分割[4—6]和姿势估计[7—8]。

自从 AlexNet[9]在 2012 年面世以来，深度卷积神经网络逐渐在图像分类领域占有一席之地，相关研究

收稿日期：2020-12-18
作者简介：罗雪阳（1996—），男，上海理工大学硕士生，主攻智能制造。
通信作者：蔡锦达（1963—），男，硕士，上海理工大学教授，主要研究方向为智能制造。

也已经从手工标注图像特征向深度学习特征所转换，这样使得人类的工作重心转移到如何设计更好的网络结构，以便更好地学习图像特征。随着深度学习领域的不断发展，深度学习的网络架构日益优化，目前使用的深度学习网络架构种类繁多，其中代表性的是 2015 年提出的 ResNet[10]引入残差思想，使深度神经网络中梯度消失的问题得到解决，并允许神经网络学习更为深层的特征成为目前最成功的卷积网络架构之一，同时其模块化的结构和可移植性使得目前主流下游应用程序也大多使用 ResNet 或其变体之一作为其网络主体结构以获得良好的效果。

将视觉注意力机制应用于深度学习的研究工作也在近几年逐步开展，现阶段主要的注意力机制常分为通道域、空间域以及混合域。通道域指在图像的通道层面使用注意力机制，主流方法是通过对各通道乘以所获得的权重系数以达到更关注某一个通道或忽略某一通道的目的，如 SENet[11]通过全局池化获得各通道的权重系数，学习通道之间的相关性，并给每个通道进行评价打分，通过将得到的分数系数与对应通道相乘，实现了将注意力机制应用于通道域的目的；SKNet[12]针对 SENet 进行改进，提出不同的图像应使用不同的感受野去获取图像特征，实现了不同的图像使用不同的卷积核权重，在不同的图像中能够动态地生成卷积核以获取不同的感受野。空间域指在图像的空间层面使用注意力机制，主流方法是将图像在空间上分割为不同的部分或直接使用各像素点乘以权重系数增强重要的特征，减弱不重要的特征，从而让提取的特征指向性更强，如 Spatial Transformer Networks（STN）通过将原图空间中的所有信息经过计算，变换到另一个空间中并将关键信息保留。混合域指在通道域与空间域上均使用注意力机制，具有代表性的 CBAM[13]使用平均池化与最大池化得到不同的特征描述的思想，先使用通道注意力模块学习各个通道间的相关性，之后在每一通道中使用空间注意力模块获得空间中的关键信息。将注意力机制与 ResNet 主体框架结合也得到了较好的效果，ResNest[14]在 ResNet 的基础上，并借鉴 ResNeXt[15]思想，在模块中首先使用 cardinality 为超参数 k 的单元进行组卷积，在每个单元中再次将特征图通过分为更细分化的子组以达到将特征图信息区分更细致的目的，其中，每组的特征表示是根据全局上下文信息选择系数权重，并将系数权重应用于每个特征子组通过加权决定的。

通过研究发现，文中发现 ResNeXt 与 ResNeSt 均只在通道域使用了注意力机制，并没有使用空间域注意力机制。现有的 CBAM 混合域机制与 ResNet 网络结构相结合时需要将其在每个 block 模块后加入，不利于网络整体结构的移植。受先前方法的启发，文中的网络在 ResNest 的网络基础上，提出 Spatial-Kernel Attention 模块，在不改变整体算法框架结构的基础上将空间域注意力机制引入。

1 算法框架设计

1.1 算法整体设计

ResNeSkt 算法整体结构与其他主流算法结构比较见图 1。

Output	ResNeXt-50(32×4d)	SENet-50	SKNet-50	ResNeSkt-50
112×112	7×7,64,stride 2			
56×56	3×3,max pool,stride 2			
56×56	[1×1,128 3×3,128，G=32 1×1,256] ×3	[1×1,128 3×3,128，G=32 1×1,256 fc,[16,256]] ×3	[1×1,128 SK[M=2,G=32,r=16],128 1×1,256] ×3	[1×1,128 3×3,128，G=32，r=2 1×1,256] ×3
28×28	[1×1,256 3×3,256，G=32 1×1,512] ×4	[1×1,256 3×3,256，G=32 1×1,512 fc,[32,512]] ×4	[1×1,256 SK[M=2,G=32,r=16],256 1×1,512] ×4	[1×1,256 3×3,256，G=32，r=2 1×1,512] ×4
14×14	[1×1,512 3×3,512，G=32 1×1,1024] ×6	[1×1,512 3×3,512，G=32 1×1,1024 fc,[64,1024]] ×6	[1×1,512 SK[M=2,G=32,r=16],512 1×1,1024] ×6	[1×1,512 3×3,512，G=32，r=2 1×1,1024] ×6
7×7	[1×1,1024 3×3,1024，G=32 1×1,2048] ×3	[1×1,1024 3×3,1024，G=32 1×1,2048 fc,[128,2048]] ×3	[1×1,1024 SK[M=2,G=32,r=16],1024 1×1,2048] ×3	[1×1,1024 3×3,1024，G=32，r=2 1×1,2048] ×3
1×1	7×7,global average pool,1000-d,fc,softmax			
#P	25.0M	27.7M	27.5M	27.6M
GFLOPs	4.24	4.25	4.47	4.32

图 1 ResNeSkt 与主流算法整体结构比较

Fig.1 Comparison of overall structure between ResNeSkt and mainstream algorithms

算法整体框架使用 ResNet 网络模型，其模块化的框架，使得 ResNeSkt 与 ResNet 一样具有很好的移植性。此处以 50 层网络结构为例，所有的网络都分成 5 部分：conv1，conv2_x，conv3_x，conv4_x，conv5_x，其中 conv1 为 1 层，conv2_x，conv3_x，conv4_x，conv5_x 中分别为 3，4，6，3 个 block，每个 block 为 3 层，最后为链接层，共计 50 层。相较于其他算法框架，ResNeSkt 由于在 block 中引入了空间注意力，使其参数量有了一定提升。

1.2　ResNeSkt block

SE-Net block，SK-Net block 与文中提出的 ResNeSkt block 算法流程见图 2。

由图 2 可知，ResNeSkt block 与 SE-Net block，SK-Net block 的共同点：三者的输入均是 h,w,c，其中 h,w 为输入的尺寸，c 为通道数；三者均是进行大小为 1×1 的卷积；三者为减少过拟合均使用了 ResNet 的残差链接机制。

ResNeSkt block 与 SE-Net block，SK-Net block 的不同点：SK-Net block 相较于 SE-Net block 在执行完尺寸为 1×1 的卷积后不再是直接使用 3×3 的卷积，而是分为 2 路，一路使用 3×3 卷积核卷积，一路使用 5×5 卷积核进行卷积，再把 2 路结果进行融合，使得 SK-Net 不仅能够和 SE-Net 考虑通道之间的权重，也考虑了 2 路卷积的权重。文中提出的 ResNeSkt block 相较于 SE-Net block，SK-Net block 引入了特征图组的思想。首先将输入分为 k 个特征图组，再将每个特征图组分为 R 个基数组，对每个基数组进行 1×1，3×3 的卷积后将每个基数组的结果进行融合后引入通道域和空间域的注意力机制，最后将每个特征图组的结果融合。

1.2.1　通道域注意力机制

文中沿用了 ResNeSt 思想，首先将图像特征分为若干特征图组，特征图组的数量由基数超参数 K 给出，通过基数超参数 R 可以将每个特征图组再次分割为 R 个基数组，因此基数组的总数为 $G=KR$。可以应用一系列变换 $\{F_1,F_2,F_3...F_G\}$ 到每个单独的基数组，每个基数组的中间表示为 $Ui=Fi(X),\ i\in\{1,2,...,G\}$。

由于每个特征图组可以分割为 R 个基数组，那么每个特征图组的组合表示可以通过跨多个基数组的逐元素求和来融合而获得，则第 k 个特征图组的表示为 $\hat{U}^k=\sum_{j=R(k-1)+1}^{Rk}U_j$，其中 $\hat{U}^k\in\Re^{H\times W\times C/K}$，$k\in1,2,...,K$，$\Re$ 表示原图像数组，H，W 和 C 是输出特征图的大小。Split Attention 模块见图 3。

由图 3 可知，Split Attention 模块的输入为 R 个基数组的卷积后结果，将所有结果进行融合，通过全局池化、归一化、ReLU 函数、Softmax 函数获得每个基数组相应的权重，即引入了注意力机制，将每个基数组乘以其相应权重并再次融合获得输出结果。

使用跨空间维度 $s^k\in\Re^{C/K}$（其中 $\Re$ 表示原图像数组）的全局平均池化来收集具有嵌入式通道统计信息的全局上下文信息[11—12]。第 c 个分量的计算见式（1）。

$$s_c^k=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\hat{U}_c^k(i,j) \tag{1}$$

式中：$\hat{U}_k$ 为第 k 个特征图组；H 为特征图组的高；W 为特征图组的宽。

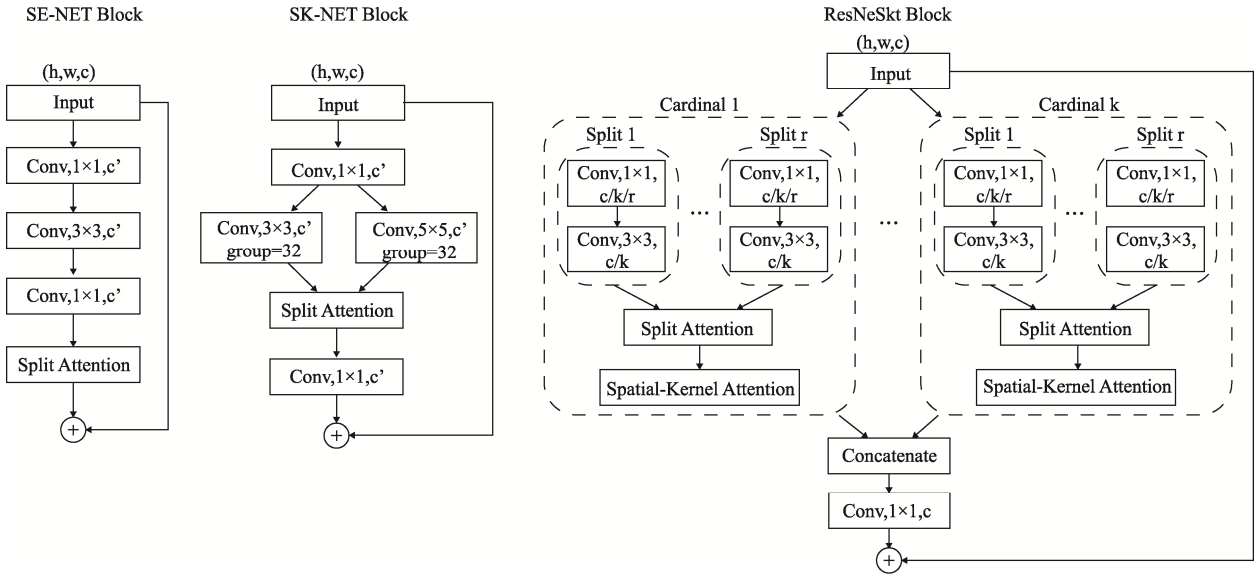

图 2　ResNeSkt block 与 SE-Net block、SK-Net block 算法流程

Fig.2 Algorithm flow chart of ResNeSkt block, SE-Net block and SK-Net block

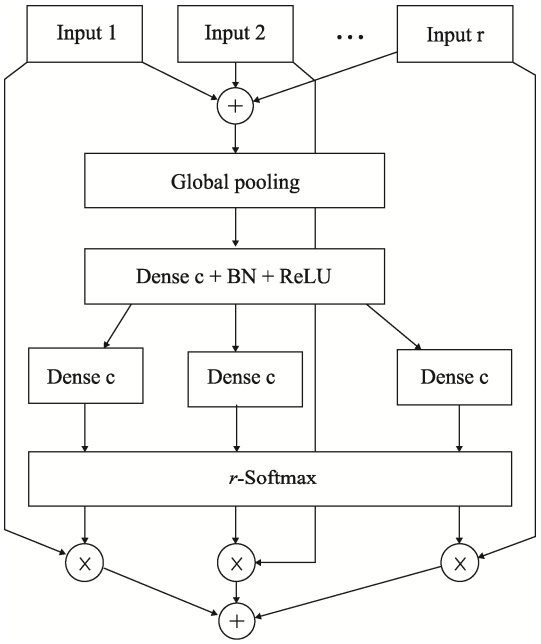

图 3 Split Attention 模块
Fig.3 Split Attention module

在基数组层面上使用注意力机制，特征图组表示的 $V^k \in \Re^{H\times W\times C/K}$ 是将得到的各个基数组的权重系数应用于相对应的基数组上后进行组合，即每个特征图组使用分割得到的 R 个基数组加权后组合而成。第 c 个特征图组的计算见式（2）。

$$V_c^k = \sum_{i=1}^{R} a_i^k(c) U_{R(k-1)+i} \tag{2}$$

式中：V_c^k 为各个基数组加权融合后得到的第 c 个特征图组，由于之前将一个特征图组分为 R 个基数组，所以在这里将 R 各基数组乘以其分配权重 $a_i^k(c)$ 后融合得到特征图组。分配权重 $a_i^k(c)$ 表达见式（3）。

$$a_i^k(c) = \begin{cases} \dfrac{\exp\left(\delta_i^c\left(s^k\right)\right)}{\sum_{j=0}^{R}\exp\left(\delta_j^c\left(s^k\right)\right)} & R>1 \\ \dfrac{1}{1+\exp\left(-\delta_i^c\left(s^k\right)\right)} & R=1 \end{cases} \tag{3}$$

式中：映射 δ_i^c 为 s^k 确定第 c 个分量的每个基数组的权重。

将每个基数组乘以分配权重的基数组 F_i，$i\in 1,2,...,G$，逐元素相加融合得到分割前对应每个的特征图组 $V^k, k\in 1,2,...,K$。

1.2.2 空间域注意力机制

先在通道层面上使用注意力机制，再在空间层面上使用注意力机制，可以获得更好的效果[13]。在 CBAM 的思想基础上引入 GeM 池化层[16]，GeM 池化层对全局最大池化和全局平均池化进行了融合，见式（4）。

$$x_k = \frac{\left(\sum_{i=0}^{H}\sum_{j=0}^{W}(x_{i,j,k})^p\right)^{\frac{1}{p}}}{H\times W} \tag{4}$$

式中：输入大小为 $H\times W\times C$，分别对应特征图的高、宽和通道数；$x_{i,j,k}$ 为特征图中的值；输出为 $x_k, k\in 1,2,...,C$；p 为 Gem 池化的关键参数，当 $p\to 1$ 时为全局平均池化，当 $p\to\infty$ 时为全局最大池化。

Sptial-Kernel Attention 模块见图 4。

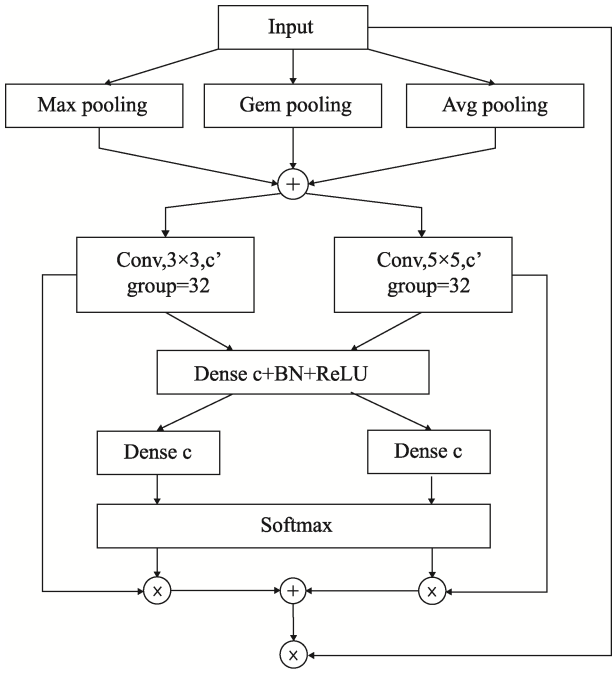

图 4 Sptial-Kernel Attention 模块
Fig.4 Spatial-Kernel Attention module

由图 4 可知，将由通道域注意力机制相加融合而成的特征图组 $V^k, k\in 1,2,...,K$ 同时进行全局平均池化、全局最大池化与 Gem 池化，使用 3 种不同的池化方法提取的图像高层次特征更加丰富，因此可以得到形状为 $H\times W\times 1/k$ 的 3 个特征图 $V_{\text{avg}}^k, V_{\max}^k, V_{\text{gem}}^k$，$k\in 1,2,...,K$。接着沿用 SKNet 的动态卷积核思想对不同图像使用的卷积核权重不同，即一种针对不同的图像动态生成卷积核的方法，将 3 个特征图 $V_{\text{avg}}^k, V_{\max}^k, V_{\text{gem}}^k, k\in 1,2,...,K$ 进行卷积核（文中以 3*3，5*5 卷积核举例，可以有多个分支）卷积分别得到 $V_{3*3}^k, V_{5*5}^k, k\in 1,2,...,K$ 的维度为 $H\times W\times 1/k$ 的 2 个特征图，将 2 个特征图按元素求和，因此 $U^k = V_{3*3}^k + V_{5*5}^k, k\in 1,2,...,K$。将 U^k 通过全局平均池化得到统计信息 T^k，见式（5）。

$$T^k = F_{\text{avg}}(U^k) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U^k(i,j) \tag{5}$$

式中：F_{avg} 为全局平均池化函数；H，W 分别为特征图的高与宽。

通过全连接层生成 compact 向量 $\boldsymbol{Z}=F_{fc}(T^k)=\delta(B(T^k))$，其中 δ 为 ReLu 激活函数，B 为批标准化(BN)。

通过 softmax 计算 2 个卷积核特征的 soft attention 向量 $\boldsymbol{A}_{3*3}$，$\boldsymbol{A}_{5*5}$，设 2 个卷积核特征的权重分别为 a_{3*3}，a_{5*5}，因此 $a_{3*3}+a_{5*5}=1$，且 $a_{3*3}=\dfrac{\exp(A_{3*3}Z)}{\exp(A_{3*3}Z)+\exp(A_{5*5}Z)}$，$a_{5*5}=\dfrac{\exp(A_{5*5}Z)}{\exp(A_{3*3}Z)+\exp(A_{5*5}Z)}$。将权重应用到对应的空间注意力特征图上，即 $U^k=a_{3*3}\cdot A_{3*3}+a_{5*5}\cdot A_{5*5}$，$a_{3*3}+a_{5*5}=1$。将得到的空间注意力权重应用到特征图 V^k 上，即 $V^k=V^k\times U^k, k\in 1,2,...,K$。

将得到的特征图组延通道维度链接起来得到与标准残差链接特征图相同维度的特征图 $V=Concat\left(V^1,V^2,...,V^k\right)$。假设每个 block 的输入为 Y，那么就有 $Y=V+T(X)$，其中 T 为跳跃链接映射。

2　实验设置与结果分析

2.1　实验设置

文中实验平台配置为：TX4000 8 GB 显存显卡、32 GB 内存和至强 W2255CPU 构建的硬件平台，软件系统为 Windows 10，神经网络框架使用 Python 3.7 和 Pytorch 1.4 搭建，采用 CUDA10.0 进行网络计算。

训练模型时实验参数为：训练过程中 1～50 个 epoch 的学习率为 1×10^{-3}，51～5100 个 epoch 的学习率为 1×10^{-4}，使用 SGD 方法；超参数 k，r 分别为 32，2；Gem 池化层参数 p 为 3；batch 大小为 8；网络输入的图像大小为 224 像素×224 像素。

2.2　数据集

文中使用 Imagenet-1K 数据集来评估算法框架的性能，Imagenet-1K 数据集包含 1000 种不同类型图片，依据计算机及 GPU 性能择选出 10 种类型图片（在 ImageNet 数据集中的编号分别为：n01514668，n01530575，n01728572，n01847000，n01871265，n02708093，n02814533，n02963159，n03991062，n11939491），其中用于训练的图片有 13 000 张，另选用于测试的图片有 500 张。这个数据集可以从 ImageNet 官网获得，数据集中的每张图片均进行中心裁剪至 224 像素×224 像素，考虑到测试集图片数量较少，故将训练集中 3000 张图片转移至测试集中，转移后得到训练集为 10 000 张图片，即每种类型图片为 1000 张作为训练输入，测试集为 3500 张。同时为考虑模型鲁棒性，对测试的图片进行随机旋转、随机亮度调整、随机对比度调整得到训练集图片 30 000 张，测试集图片 10 000 张。

2.3　实验结果与对比分析

为了验证所提方法的有效性，在 ImageNet-1K 上依据计算机及 GPU 性能选择出 10 类图片进行多组对比实验，并对实验结果进行分析。

2.3.1　准确度与损失值对比

直观地观察模型在整个训练过程中识别准确度和损失值的变化情况，可以更好地改进模型。当模型训练完成之后对其训练日志进行解析（以 ResNeSkt-50 举例），得到模型的识别精度和损失曲线分别见图 5 和图 6。实验设置训练总迭代次数为 100，得到训练集和测试集的损失率（loss）、准确率（accuracy）随迭代次数增加的变化曲线。由图 5 可知，准确率变化曲线随着迭代次数的增加，训练集的准确率在训练过程中虽然有所起伏，但总体处于一个不断上升的趋势。测试集准确率变化和训练集准确率变化相似；同时训练集和测试集的损失率随着迭代次数的增加也在不断降低，说明文中所提出的 ResNeSkt 在训练过程中没有发生过拟合和欠拟合的情况，模型收敛效果良好。当训练迭代次数到达 50 时，准确率和损失率基本趋于稳定，但仍然存在微小浮动，当 epoch 达到 51 时，训练学习率改变为 1×10^{-4}，使得 loss 值再次大幅度降低，使得模型在更小的区域里选择最优值。随着训练迭代次数的递增，浮动程度逐渐减弱，当 epoch 达到 100 时训练集和测试集最终得到的 accuracy 分别为 92.21%和 81.39%，训练集 loss 值趋近于 0.31，测试集 loss 值趋近于 0.64。这充分证明所提出的 ResNeSkt 网络结构用于图片分类上的有效性。

模型训练完成后，采用测试集数据进一步检测模型准确性，随机选取测试集数据输入模型检测模型准确性，ResNeSkt-50 在 ImageNet-1K 选择出来的 10 种类别数据集中表现出较高的精度，识别精度为 81.39%。

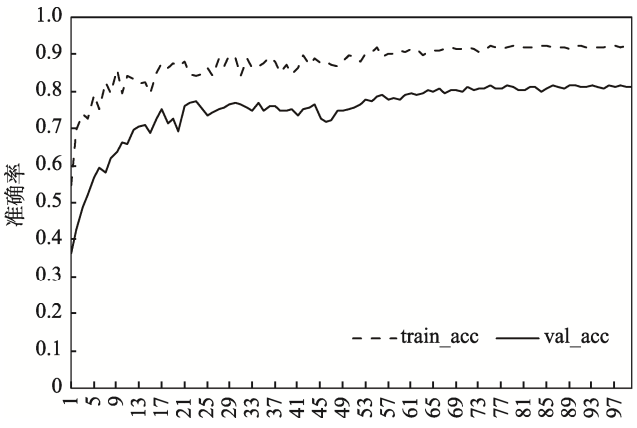

图 5　ResNeSkt 准确率变化曲线
Fig.5 ResNeSkt accuracy change curve

图 6 ResNeSkt 损失函数变化曲线
Fig.6 ResNeSkt loss function change curve

2.3.2 模型深度对于准确度的影响

文中进行模型深度对于图像分类准确度的影响实验，由于借鉴了 ResNet 的残差思想，使得文中网络在增加模型深度的同时不会发生梯度消失的问题，实验结果见表 1。其中模型名称后的 50，101，200 分别代表模型的层数；模型参数量$^{\#}P$ 表示模型的空间复杂度，参数量越大的模型其占用的空间越大，反之则越小；模型计算量 GFLOPs 代表模型的时间复杂度，计算量越大的模型其运算时间越长，反之则越短。模型参数量$^{\#}P$ 和模型计算量 GFLOPs 均可以通过 thop 工具箱中的 profile 函数获得。准确率由训练完成的模型对测试集图片逐一进行预测分类，将预测正确的图片张数与总体测试集图片张数进行比较得到，即预测正确的图片张数与总体测试集图片张数的比例即为准确率。

表 1 模型深度对于准确度的影响
Tab.1 Effect of model depth on accuracy

模型名称	参数量/MB	计算量	准确率/%
ResNeSkt-50	27.6	4.32	81.39
ResNeSkt-101	48.3	8.19	81.74
ResNeSkt-200	70.1	13.99	82.05

随着网络结构的加深，准确度得到明显的提高，但同时也伴随着参数量$^{\#}P$ 的增加和网络模型的计算量 GFLOPs 增大的问题。在提升网络深度的同时，虽然准确率也会有所提高，但随着模型层数的不断提高，准确率的提高相较于模型时间复杂度和空间复杂度的增大不再具有很好的成长性，也不利于实际研究中的使用。

2.3.3 实验结果与主流算法比较

为了比较文中所提出的 ResNeSkt 与现阶段主流算法的性能差异，给出了几种具有代表性的模型的对比结果，见表 2。

表 2 ResNeSkt 与现阶段主流算法的性能比较
Tab.2 Performance comparison between ResNeSkt and current mainstream algorithms

方法	参数量/MB	计算量	准确率/%
ResNet-50[10]	25.5	4.14	75.06
ResNeXt-50[15]	25.0	4.24	76.63
SENet-50[11]	27.7	4.25	77.76
SKNet-50[12]	27.5	4.47	78.14
ResNest-50[14]	27.5	4.25	80.06
ResNet-50+CBAM[13]	28.1	4.36	77.34
ResNeSkt-50(ours)	27.6	4.32	81.39
ResNet-101[10]	44.5	7.87	76.26
ResNeXt-101[15]	44.3	7.99	77.82
SENet-101[11]	49.2	8.00	78.35
SKNet-101[12]	48.9	8.46	78.78
ResNest-101[14]	48.3	8.07	80.53
ResNet-101+CBAM[13]	49.3	8.05	77.86
ResNeSkt-101(ours)	48.3	8.19	81.74

结果显示，相较于现有其他主流算法，文中提出的 ResNeSkt 图像分类准确度最高。究其原因，主要在于文中沿用特征图组的组卷积思想并加入了空间注意力机制，通过在空间域注意力中引入 Gem 池化与 SKNet 的不同 kernel size 思想，使其可以自适应地调整卷积核获得相对于特定图像的特定感受野，从而使得整个算法模型可以在通道域与空间域分别使用注意力机制，可以使模型提取到图像中更加有用的信息，从而使模型获得更好的效果。由于 ResNeSkt 引入了空间注意力机制，使得其相较于其他仅使用通道域注意力机制的算法参数量$^{\#}P$ 和计算量 GFLOPs 略有高，以此换来的是更高的准确率。相较于直接使用 CBAM 的 ResNet-50，由于 ResNeSkt 使用组卷积思想，使得其有着更少的计算量 GFLOPs 和模型参数量$^{\#}P$。

2.3.4 作为主干网络与目前主流算法比较

对比以 Faster-RCNN[15] 作为实验对象，Faster-RCNN 是何凯明等在 2015 年提出的目标检测算法，文中通过更换其主干网络进行实验，并分析实验结果，评价标准采用目标检测主流评价指标均值平均精度进行评价（mAP），实验结果见表 3。

通过将 ResNeSkt 与目前现有主流算法分别用作 Faster-RCNN 的主干网络，对比 Faster-RCNN 目标检测算法的 mAP 可知，使用文中提出的 ResNeSkt 作为主干网络的 Faster-RCNN，相较于使用其他现有主流算法拥有更高的均值平均精度，即 mAP。说明

表 3　更换主干网络的 Faster-RCNN 效果
Tab.3 Effect of replacing the Faster-RCNN of the backbone network

Method	Backbone	mAP/%
Faster-RCNN[15]	ResNet-50[10]	37.30
	ResNeXt-50[15]	40.10
	SENet-50[11]	41.90
	ResNest-50[14]	44.72
	ResNeSkt-50(ours)	45.53

ResNeSkt 可以用作其他算法的主干网络，体现其拥有很好的移植性，同时使用 ResNeSkt 作为主干网络的目标检测算法 Faster-RCNN 相较于使用其他主流方法作为主干网络有着更好的效果。

3　结语

以深度学习为基础，结合注意力机制，在前人的基础上，研究了具有通道域与空间域注意力机制且具有很好移植性的网络模型结构。通过实验证明，文中所提算法可以在图像分类方面达到较高的准确度，其准确度为 81.39%，文中提出的 ResNeSkt 的准确度要优于当前所有主流图像分类算法，并可以作为主干结构应用至其他图像研究领域，未来将会在此方法的广泛性上进一步展开研究。

参考文献：

[1] LIU Wei. Beach Sports Image Detection Based on Heterogeneous Multi-Processor and Convolutional Neural Network[J]. Microprocessors and Microsystems, 2021, 82: 1—6.

[2] JIANG Shui-qing, ZHENG Juan, XUE Wen-tao, et al. High Resolution Image Detection and Ultrasonic Evaluation of Hyperthyroidism Based on Hospital IoT System[J]. Microprocessors and Microsystems, 2021, 81: 1—6.

[3] 陈亮，张浩舟，燕浩. 基于深度学习算法的尿素泵体用铝型材表面瑕疵检测[J]. 流体机械，2020，48(8): 47—52.

[4] CHEN Liang, ZHANG Hao-zhou, YAN Hao. Surface Flaw Detection of Aluminum Profile for Urea Pump Body Based on Deep Learning Algorithm[J]. Fluid Machinery, 2020, 48(8): 47—52.

[5] FRIZZI S, BOUCHOUICHA M, GINOUX J, et al. Convolutional Neural Network for Smoke and Fire Semantic Segmentation[J]. Image Processing, IET, 2021(6): 1—14.

[6] KAMANN C, ROTHER C. Benchmarking the Robustness of Semantic Segmentation Models with Respect to Common Corruptions[J]. International Journal of Computer Vision, 2021, 129(2): 462—483.

[7] ZHU Y, SAPRA K, REDA F A, et al. Improving Semantic Segmentation Via Video Propagation and Label Relaxation[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020: 8848—8857.

[8] FANG H S, XIE S, TAI Y W, et al. Rmpe: Regional Multi-Person Pose Estimation[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334—2343.

[9] XIAO B, WU H, WEI Y. Simple Baselines for Human Pose Estimation and Tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 466—481.

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]// Advances in Neural Information Processing Systems, 2012: 1097—1105.

[11] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016: 770—778.

[12] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-Excitation Networks[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence(CVPR), 2019: 2011—2023.

[13] LI Xiang, WANG Wen-hai, HU Xiao-lin, et al. Selective Kernel Networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019: 510—519.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018: 3—19.

[15] ZHANG Hang, WU Chong-ruo, ZHANG Zhong-yue, et al. ResNeSt: Split-Attention Networks[J]. arXiv preprint arXiv, 2020, 4 : 1—14.

[16] REN Shao-qing, HE Kai-ming, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137—1149.

[17] BERMAN M, JÉGOU H, VEDALDI A, et al. Multi-Grain: a Unified Image Embedding for Classes and Instances[J]. ArXiv Preprint ArXiv, 2019, 5: 1—13.