基于卷积神经网络的图像分类研究进展*

覃　晓[1],黄呈铖[1],施　宇[1],廖兆琪[1],梁新艳[1],元昌安[2**]
(1.南宁师范大学计算机与信息工程学院,八桂学者创新团队实验室,广西南宁　530000;2.广西科学院,广西南宁　530007)

摘要:基于卷积神经网络的图像分类算法的优势是传统方法无法比拟的。卷积神经网络利用其设计好的网络结构和权值共享的特点,能够从数量庞大的训练数据中学习图像底层到高级语义的抽象特征,而且端到端的学习省去了在每一个独立学习任务执行之前所做的数据标注。多年来,卷积神经网络经过科研人员的探索和尝试,从最开始的多层神经网络模型,演变出多种优化结构,性能不断提高。本文介绍了基于卷积神经网络图像分类算法的研究进展,叙述了卷积神经网络在图像分类中的经典模型和近年来的改进方法,并对各个模型进行分析,展示各种方法在ImageNet公共数据集上的性能表现,最后对基于卷积神经网络的图像分类算法的研究进行总结和展望。

关键词:卷积神经网络　图像分类　经典模型　改进方法　性能对比

中图分类号:TP31　　**文献标识码**:A　　**文章编号**:1005-9164(2020)06-0587-13
DOI:10.13656/j.cnki.gxkx.20210119.001

0　引言

图像分类是图像分割、目标检测[1-7]、人脸识别[8-10]、行为识别和姿态估计[11-13]等视觉任务的基础任务。近年来,深度学习模型已被证明是具有卓越分类能力的机器学习算法。深度学习强调模型结构的深度和特征学习的重要性,采用有监督或无监督的方式对图像从底层到高级的语义特征进行学习。典型的深度学习模型有深度信念网络(Deep Belief Network,DBN)[14]、受限玻尔兹曼机(Restricted Boltzmann Machine,RBM)[15]、卷积神经网络(Convolutional Neural Network,CNN)[16,17]等。卷积神经网络已被证明是解决各种视觉任务的有效模型[18-21]。

卷积神经网络是受到动物视觉神经系统启发被提出的。加拿大科学家David H. Hubel[22]和瑞典科学家Torsten N. Wiesel[23]从1958年开始对猫视觉皮层进行研究,他们在V1皮层里发现两种细胞,简单细胞(Simple Cells)和复杂细胞(Complex Cells),这两种细胞的共同特点就是每个细胞只对特定方向的条形图样刺激有反应,两者的主要区别是简单细胞对应视网膜上的光感受细胞所在的区域很小,而复杂细胞则对应更大的区域,这个区域被称作感受野(Receptive Field)。早在1980年,日本的Kunihiko Fukushima[24]提出的Neocognitron模型,即为模拟

*国家自然科学基金项目(61962006),广西创新驱动重大项目(AA18118047)和广西研究生教育创新计划项目(YCSW2019182)资助。
【作者简介】
覃　晓(1973—),女,副教授,主要从事人工智能和图像处理研究。
【**通信作者】
元昌安(1964—),男,博士,教授,主要从事人工智能与数据挖掘研究,E-mail:68852917@qq.com。
【引用本文】
覃晓,黄呈铖,施宇,等.基于卷积神经网络的图像分类研究进展[J].广西科学,2020,27(6):587-599.
QIN X,HUANG C C,SHI Y,et al. Research Progress of Image Classification based on Convolutional Neural Network [J]. Guangxi Sciences,2020,27(6):587-599.

这种作用的结构。随后 LeCun 等[25]基于 Fukushima 的研究工作,使用误差梯度回传策略设计和训练 CNN(被称为 LeNet-5),这就是现今计算机视觉领域第一个卷积神经网络。

基于卷积神经网络的图像分类方法,可以从海量且有噪声的图像中学习到目标的高层特征,且这种特征对于目标某种程度的形变有很好的鲁棒性。卷积神经网络是一种带有卷积结构的深度神经网络,其将原始图像输入网络,经过数据预处理后,网络各节点传递图像数据并经过逐层的权重迭代更新和计算,最终输出类别标签上的概率分布,图像的特征由卷积神经网络自动学习,不需要手动设计。此外,卷积神经网络卷积层中的网络稀疏连接和卷积核权值共享两大特性,有利于神经网络的快速学习并避免过度拟合[26],使得网络结构变得简单,适应性更强,泛化能力提高[27]。

CIFAR-10[28]、CIFAR-100[28],特别是 ImageNet 数据集[29]是目前深度学习图像领域应用最广泛的数据集。2012 年举办的图像分类竞赛(ImageNet Large Scale Visual Recognition Challenge)中,由 Alex Krizhevsky 等[30]实现的深层 CNN 结构系统获得冠军,当时取得了最佳分类效果,也是在此后,更多的更深更宽的神经网络被提出,如 VGGNet[31]、GoogleNet[32]、ResNet[33]、DenseNet[34] 等。近几年来,有研究者开始对神经网络的特征图通道、空间、卷积核等方面进行结构上的改进,提出 SE block[35]、SK block[36]等注意力机制,它们可以直接放入现有的神经网络中进行训练。也有研究者使用现有的网络,如 ResNet 用作基线网络进行模型改进研究,如 Res2Net[37]、ResNeSt[38],提出的这些模型在广泛使用的图像分类 ImageNet 数据集上,均得到优于基线网络的准确率,网络性能得到一致的提升。下面本文将对上述卷积神经网络的研究过程进行详细介绍。

1 卷积神经网络经典模型

1.1 VGGNet 网络

VGGNet 是由 Simonyan 等[31]提出的非常经典的 CNN 模型,在 2014 年的 ImageNet 挑战赛中以 92.3%的准确率获得亚军。VGGNet 整体结构十分简洁,主体采用多个 3×3 的小卷积核,每个卷积后都伴有非线性激活函数。3 个 3×3 的卷积可等价于 1 个 5×5 卷积,但小卷积核替代大卷积核,却能增加网络的非线性表达能力,因而 VGGNet 在图像复杂特征表达上具有天然的优势。每隔 2 个或者 3 个 3×3 的卷积层就连接一个 2×2 的最大池化层,将卷积层和最大池化层进行反复堆叠即构成 VGGNet 网络主体。部分结构采用 1×1 的卷积核,后伴有激活函数,在不影响输入输出维度的情况下,增加模型的非线性。如果卷积层和全连接层堆叠的层数总共为 16 层,那么就称这个网络为 VGG16。研究表明,当 VGG 模型的层数过深时,会因为参数的误差变大导致网络的退化,因此一般 VGG 的模型多数为 16—19 效果最佳。因为结构简单和性能稳定的优势,现在 VGGNet 仍在图像处理领域被广泛研究和应用。图 1 展示了 VGGNet 的网络架构。

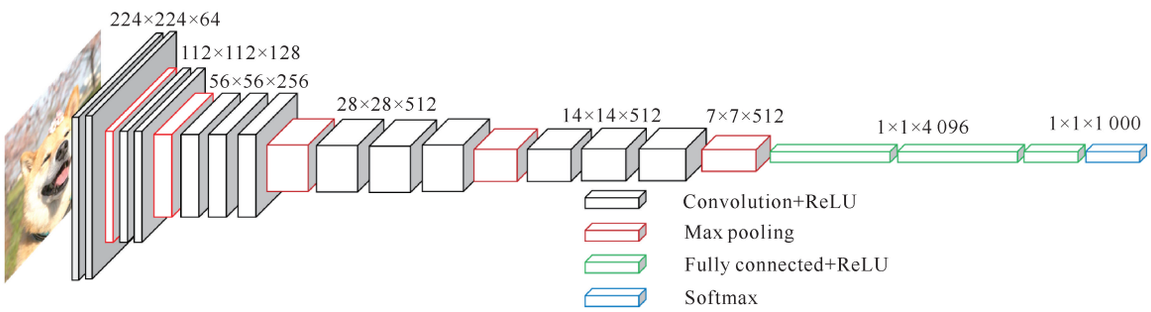

图 1 VGGNet 的网络架构

Fig. 1 Network architecture of VGGNet

从 VGGNet 提出到现在,仍然有许多研究者将其使用到自然应用场景中。孙新立[39]通过构建 VGG16 模型,以颜色和纹理作为煤和矸石图像的类别特征,结合迁移学习方法,解决了目前人工捡矸法、机械湿选法操作过程中无法解决的效率问题,其识别准确率达到 99.18%,能够有效地识别出煤和矸石。田佳鹭等[40]提出改进的 VGG16 模型来对猴子图像数据进行分类,模型的优化包括利用 Swish 作为激活函数,将 softmax loss 与 center loss 相结合作为损失函数以实现更好的聚类效果,采用性能完善的 Adam

优化器;同时用训练集训练模型以确定微调参数信息,再用测试集检验模型准确性;该方法对猴子图像分类的准确度可达到 98.875%,分类速度也得到显著提升。

1.2 Inception 网络

Szegedy 等[32]提出的 Inception-v1 模型使用 22 层 Inception 网络结构模块,Inception-v1 模型与以往的卷积神经网络模型相比,其特点在于最后一层使用全局平均池化层代替全连接层,降低空间参数,让模型更加健壮,抗过拟合效果更佳。该模型在 ILSVRC 2015 比赛分类任务上以 93.3%的正确率获得该届比赛的冠军。Inception 模块的主要思路是使用一个密集成分来近似或者代替最优的局部稀疏结构,使其能够降低网络复杂度,达到提高神经网络的计算资源利用率,提升其分类准确率并且减少过拟合的情况。经典的 Inception 模块如图 2 所示。

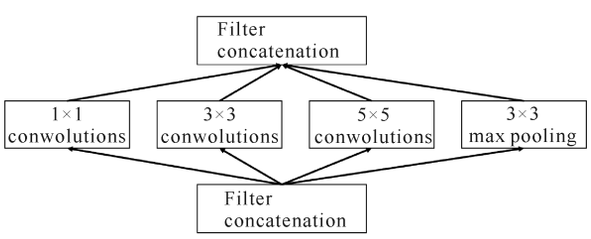

图 2 经典的 Inception 模块

Fig. 2 Classical Inception module

模块中包含几种不同大小的卷积:1×1 卷积、3×3 卷积、5×5 卷积以及一个 3×3 的池化层,让网络增加了对不同尺寸度的适应性,这些卷积层和池化层会把它们得到的特征组融合到一起,输入给下一层 Inception 模块。

一般而言,提升网络性能最直接的方法是增加网络的深度和宽度。其中,网络的深度指的是网络的层数,宽度指的是每层的通道数。当深度和宽度不断增加时,需要学习的参数也不断增加,巨大的参数容易发生过拟合,并且会导致计算量加大。Incaption-v1 模型通过引入稀疏特性,将全连接层转换成稀疏连接的方法解决上述问题。Incaption-v1 模型取得成功之后,研究人员在其基础上进行诸多应用和改良。在应用方面,Shichijo 等[41]构建了一种能自行诊断幽门螺杆菌感染的 Inception-v1 模型,经过测试,模型的准确性为 87.7%,诊断时间 194 s;Zhuo 等[42]提出一种使用卷积神经网络的车辆分类新方法,先对 Inception-v1 模型进行 ILSVRC-2012 数据集预训练,分类中汽车共分为公共汽车、汽车、摩托车、小巴、卡车和货车 6 类,经过试验平均准确率为 98.26%。在改良方面,针对 Inception-v1 模型收敛速度太慢的问题,Ioffe 等[43]将 BN 层引入 Inception-v1 模型,提出 Inception-v2 模型;Szegedy 等[44]为提高网络学习效率,进一步引入卷积因子化的思想,把较大的卷积分解成较小的卷积,通过级联的方式来减少参数量,进而提出 Inception-v3 模型。之后 Szegedy 等[45]又对 Inception-v3 模型进行优化,提出 Inception-v4 和 Inception-ResNet 模型。

1.3 ResNet 网络和 DenseNet 网络

理论上,网络深度越深,模型的非线性表达能力越好,分类性能应该更好。但研究表明,神经网络因为网络层数太深,会引发梯度反向传播中的连乘效应,从而导致梯度爆炸[46]、梯度消失[47]等问题,研究人员通过在网络中加入 BN[48]、Dropout[49]的方式,解决上述问题。He 等[33]创新性提出的 ResNet 网络,从改变网络结构的角度解决深层网络梯度消失问题,在训练集和验证集上,都证明越深的 ResNet 网络模型,其错误率越小。

ResNet 使用的残差学习单元如图 3 所示,整个残差学习单元除了正常的权重层输出外,还有一个分支把输入直接相连到输出上,该输出和权重层输出做相加运算得到最终的输出。若输入 x 学习,最终得到的输出为 $H(x)$,$F(x)$是权重层的输出,则整个残差学习过程中可以表达为 $H(x)=F(x)+x$。如果权重层没有学习到特征,即 $F(x)$为 0,则 $H(x)$为一个恒等映射(Identity Mapping)。多个残差学习单元产生多个旁路的支线将输入信息绕道传到输出,保护信息的完整性,确保最终的错误率不会因为深度的变大而越来越差。

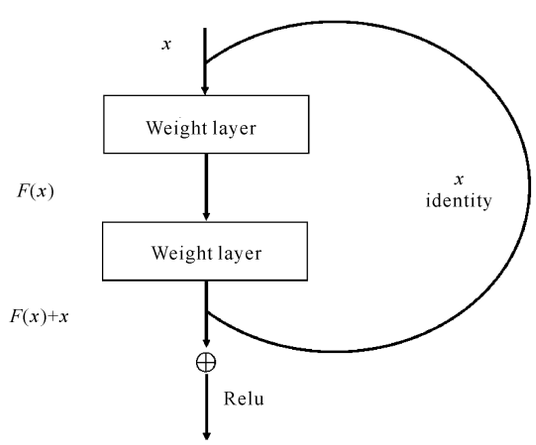

图 3 残差学习单元

Fig. 3 Residual learning unit

残差学习单元使用恒等映射,把当前输出直接传输给下一层网络,相当于一个捷径连接(Shortcut Connection),也称为跳跃连接(Skip Connection),同时在后向传播过程中,将下一层网络的梯度直接传递给上一层网络,这样就解决了深层网络的梯度消失问题。捷径连接并不会产生额外的计算量。ResNet 的残差学习单元解决了深度网络的退化问题,可以训练出更深的网络,是深度网络的一个历史性突破。

在模型表征方面,虽然 ResNet 不能更好地表征图像某一方面的特征,但是其允许逐层深入地表征更多模型,利用网络深度去理解图像更多有意义的语义特征。ResNet 的设计使得前向传播和反向传播算法可以顺利进行,因此,在极大程度上,ResNet 使得优化较深层模型更为简单。此外,ResNet 的捷径连接既不产生额外的参数,也不会增加计算的复杂度。捷径连接通过简单的执行函数映射,并将它们的输出添加到叠加层的输出,再通过反向传播。所以,整个 ResNet 网络可以被看作是一种端到端的训练模式。

与 ResNet 的原理一致,DenseNet[34] 也是建立层与层之间的连接来达到特征重用的目的,不同的是 DenseNet 层与层之间的连接具有密集连接特性,进一步增强了特征重用的效果。DensNet 的基本架构如图 4 所示,深度 DenseNet 具有 3 个稠密块(Dense Blocks)。DensNet 与 ResNet 模型的区别在于,在网络中任何两层之间都会有直接的连接,也可以理解为网络每一层的输入都是前面所有层输出的集合。从结构上看,感觉是极大地增加了网络的参数和计算量,但实际上 DenseNet 模型比其他网络有着更高的效率。由于每一层都包含之前所有层的输出信息,通过网络每层的特征重复利用,当前层只需要较少的特征图就可以把 DenseNet 模型的每一层设计得特别窄,从而解决网络的冗余现象,降低参数量。在稠密块中的每一个单元实际上都是一个瓶颈层(Bottleneck Layer),其中包括一个 1×1 卷积核和一个 3×3 卷积核。相邻块之间的层称为过渡层(Transition Layer),具体包括一个 BN 层、一个 1×1 卷积和一个池化层,通过卷积和池化来改变特征图的大小,同样能降低冗余。一个块中有 N 个特征图,通过一个 0—1 的参数来限制输出的特征图数量。

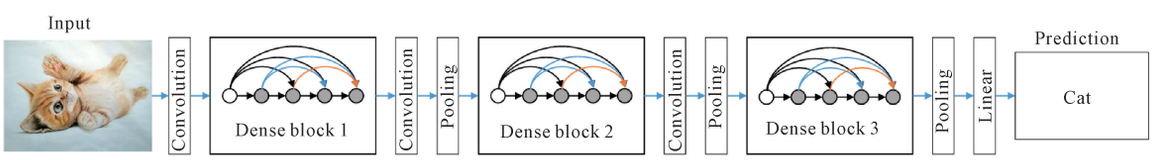

图 4　DensNet 基本架构

Fig. 4　Basic architecture of DensNet

DensNet 网络通过学习比较少的特征图来降低特征学习的冗余,鼓励特征重用,一定程度上缓解了梯度消失问题。因为不需要重新学习冗余特征图,这种密集连接模式相对于传统的卷积网络,大大减少了参数的数量,在比 ResNet 网络使用参数更少的情况下,达到 ResNet 网络的准确率。

2　卷积神经网络改进方法

尽管卷积神经网络的经典模型在视觉大赛中表现出色,但若识别的局部区域大小多样化且图像包含较多的噪声,则不能充分提取目标特征的信息,影响网络的分类与识别性能。于是人们开始考虑是否能够通过改进特征图通道、卷积核等来进一步提高网络提取特征能力。近 3 年来,通过引入注意力机制方法改进 ResNet 网络,从而提高 CNN 图像分类算法性能的研究十分广泛。其中,注意力机制模块的作用是使卷积神经网络在训练过程中更加精准地判断输入图像的哪个部分需要更加关注,从而从关键部分进行特征提取,得到重要信息,提升网络性能。而 ResNet 由于网络残差块的特殊效果,对 ResNet 网络的改进已经成为一个专门的、独特的卷积神经网络研究,以下将对用于图像分类的 ResNet 网络的改进方法进行介绍。

2.1　SENet

卷积核作为卷积神经网络的核心部件,其本质上只建模图像的空间信息,并没有建模通道之间的信息。面对现实中复杂的图像,深层网络虽然可以获取图像更高层的语义信息,但是增加网络层数难免会让训练过程出现过拟合现象,效果有时甚至比浅层网络更差。因此,Hu 等[35] 提出一种利用通道信息理解图像语义的注意力机制模型 SENet,即关注特征在通道之间的关系,它是 ImageNet 2017 收官赛的冠军模型,在很大程度上减小了之前模型的错误率,并且复杂度低,新增参数量和计算量都很小,属于轻量级

模型。

SE 块的构成如图 5 所示,SENet 模块主要由挤压(Squeeze)和激励(Excitation)两部分组成。在 Squeeze 部分中,C 个大小为 $H\times W$ 的特征图通过全局平均池化,特征图 $H\times W\times C$ 被压缩成一维的 $1\times 1\times C$,即相当于一维的参数获得了 $H\times W$ 大小的全局信息,具有全局的感受野。在 Excitation 部分中,将 Squeeze 得到的 $1\times 1\times C$ 加入全连接层(Fully Connected),通过参数 W 为每个特征通道生成权重,获取每个通道的重要性,以此来显示建模通道之间的相关性。得到不同通道的重要性大小后,通过乘法主通道加权到先前的特征,实现在通道上对原始特征的重新校准。

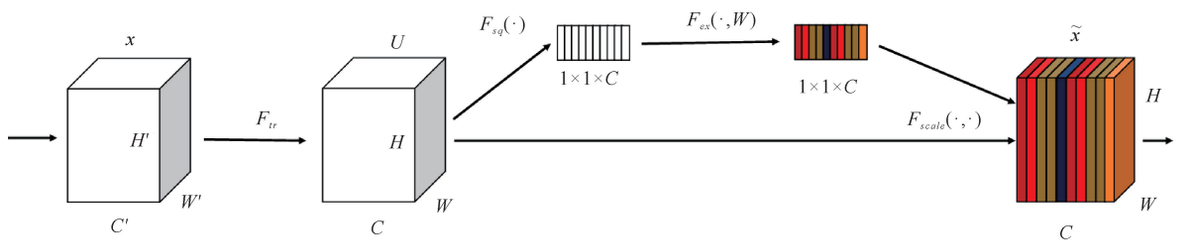

图 5 SENet 模块

Fig. 5 SENet module

其他新网络架构的提出需要进行大量的调参工作才能达到理想效果,SENet 模块则可以根据实际研究需求,直接嵌入在其他卷积神经网络架构中(图 6),组成 SE-Inception、SE-ResNet 等网络。

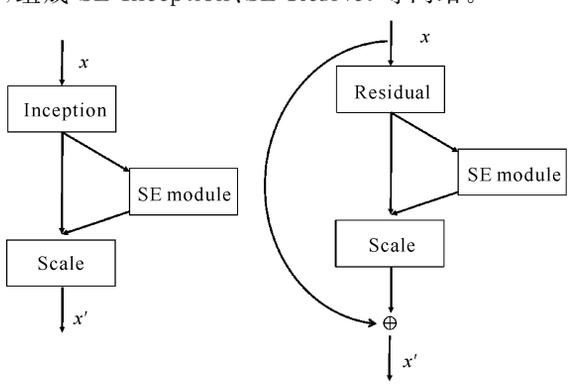

图 6 SE-Inception (左)和 SE-ResNet (右)

Fig. 6 SE-Inception (left) and SE-ResNet (right)

注意力机制的提出并应用于 CNN 模型结构是近年来改进 CNN 结构的一大亮点,已经有大量研究受到 SENet 的启发,进行改进并应用到图像识别领域,并取得一定进展[50,51]。注意力机制的核心目标是从全局众多复杂信息中选择出对当前视觉任务目标更重要的信息。在图像分类中,当神经网络获取对目标任务更关键的信息,也就相当于找到了对分类目标更具有判别性的区域,从而提高网络的分类精度。传统的图像分类模型只是单纯地利用不同大小的卷积核在空间维度和特征维度提取特征,然后将信息进行聚合从而获取全局信息。这种提取特征的过程容易因为噪声的干扰不能准确定位到目标任务的可区分区域,反而容易提取到不相关的识别区域。SENet 能够从通道的角度,通过全面捕获通道依赖性,自动获取每个特征通道的重要程度,然后依照这个重要程度提升有用的特征并抑制对当前任务用处不大的特征,从而提高网络识别性能。

2.2 SKNet

Li 等[36]提出的 SKNet 是一种动态选择机制,允许每个神经元根据输入信息的多个尺度自适应地调整其感受野大小。SKNet 在计算上属于轻量级,参数量和计算成本只是轻微增加。SKNet 在 ImageNet 2012 数据集上,相比其他常见的模型,如 Inception-v4[45]、SENet-101[35]等均有提升,Top-1 错误率仅为 18.40%。

SENet 是针对特征图的通道注意力机制的研究,SKNet 则是针对卷积核的注意力机制的研究,着重突出卷积核的重要性。常见的卷积神经网络中,对于特定任务的特定模型,卷积核大小是确定的,而在 SKNet 模型中,卷积核的大小可以是不确定的,其允许网络可以根据输入信息的多个尺度自适应地调节接受域大小,其灵感来自于人的双眼在看不同尺寸、不同远近的物体时,视觉皮层神经元接受域大小会根据外界刺激来进行调节。

如图 7 所示,SKNet 由 Split、Fuse 和 Select 3 个运算符组成。Split 运算符使用不同大小的卷积核(3×3 和 5×5)对输入图像进行卷积,生成具有各种内核大小的多个路径。Fuse 运算符聚合来自多个路径的信息,类似于 SENet 模块[35]的挤压-激励处理,两个特征图相加后,进行全局平均池化和全连接操作,输出两个矩阵 a 和 b,其中矩阵 b 为冗余矩阵,

$b=1-a$。Select 运算符实际上对应于 SENet 模块中的 Scale 运算,将两个权重矩阵 a 和 b 与对应经过 Split 运算得到的两个特征图进行加权操作。图 7 仅展现了双分支的情况,实际应用中可以扩展为多分支。

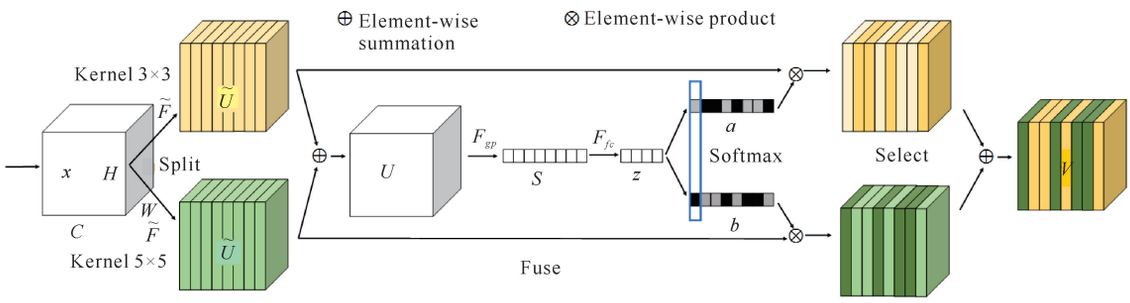

图 7　SKNet 模块

Fig. 7　SKNet module

与 Inception 网络[44]中的多尺度不同,SKNet 是让网络自己选择合适的尺度。与 SENet[35]仅考虑通道之间的权重不同,SKNet 不仅考虑通道之间的权重,还考虑分支中卷积的权重,充分利用组卷积和深度带来的较小理论参数量和计算量的优势,使得模块在增加多路与动态选择的设计中不会带来很大的计算开支。总的来说 SKNet 相当于给网络融入软注意力机制,是一种泛化能力更好的网络结构。虽然 Inception 网络的多尺度设计精妙,效果也不错,但实际上是通过人工设计卷积核的大小进行特征提取。SKNet 这种为获取不同特征信息可以自适应卷积核大小的结构,是卷积神经网络模型极限的一次重要突破。目前已经有相关研究引入 SKNet 的思想改进网络[52,53]。

SKNet 是对特征图进行权重调整的网络结构,它并不会改变特征图的大小,因此,它可以嵌入到现有网络模型中的任意地方来对特征进行提取操作,如 SK-ResNeXt50、SK-ResNet50 等,进而提升网络的实验效果。

2.3　Res2Net

尽管 ResNet 有效地解决了网络深度加深造成的梯度爆炸和梯度消失的问题,使得特征提取能力更强大,但在 ResNet 网络的残差块中,最具有特征提取能力的卷积核只有通用的单个 3×3 卷积核。然而,自然场景下的物体可能以不同的尺寸出现在一个图像中,物体的基本上下文信息可能占据比物体本身大得多的区域,ResNet 的单尺度卷积核获得的单一尺度感知信息不能很好地理解自然场景下的图像、对象及其周围环境,导致网络不能准确识别目标对象。为解决该问题,有研究通过设计多尺度特征[54-57]提取模型来提升网络性能。Gao 等[37]在 ResNet 的基础上,提出 Res2Net 网络,通过在单一残差块中对残差连接进行分级,进而可以提取到细粒度层级的多尺度表征,同时增加每一层的感受野大小,以此方式进一步提高网络模型的分类与识别性能。图 8 左图给出 ResNet 的瓶颈块(Bottleneck Block)和 Res2Net 模块的结构对比图。

ResNet 网络中瓶颈块的结构采用 1×1、3×3、1×1 三层卷积层结构。Res2Net 模块只是修改了瓶颈块中的 3×3 卷积层,在单个残差块内构造类似于分层的残差连接,并引入一个新的超参数 Scale,它将输入通道数平均分成多个特征通道,Scale 越大表明多尺度提取能力越强。图 8 右图展示了 Scale=4 时 Res2Net 模块示意图,即将第一层 1×1 卷积层的输出特征分为 4 组,为减少参数量并将特征重利用,第一组无卷积操作,随后的第二、第三组会将卷积后得到的特征图输出的同时传入下一组,第三、第四组每次卷积操作都会接收到前面每一组特征的信息。如此,网络将得到不同数量以及不同感受野大小的输出,如 Y_2 得到 3×3 的感受野,那么 Y_3 得到 5×5 的感受野,Y_4 会得到更大尺寸 7×7 的感受野,最后将这 4 个输出进行融合并经过一个 1×1 卷积。也就是说,每个 3×3 卷积核可以接受来自该层前面的所有特征,每次分类特征经过 3×3 的卷积处理后,其输出的感受野要比输入更大。这样的操作,将使每一组卷积最终的输出信息更加丰富、更具有多尺度特征。相较于 ResNet 的模块单一尺度的特征提取,Res2Net 的模块设计提升整个网络的多尺度特征提取能力,更适应于自然场景下的图像分类与识别。

Res2Net 的通用性和便利性与其他多尺度模型

相比要更强，可以与现有的最新模块或网络相整合，例如 ResNeXt[58]、SE 模块[35]等。

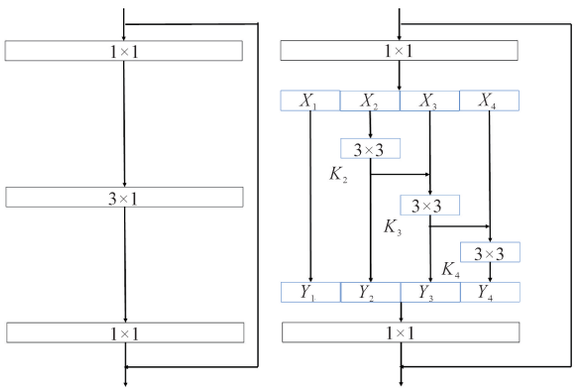

图 8　ResNet 的瓶颈块（左图）和 Res2Net 模块（右图）

Fig. 8　Bottlecheck block in ResNet (left figure) and Res2Net module (right figure)

2.4　ResNeSt

ResNet 网络的设计初衷很大程度上缓解了网络的退化问题，使得网络可以学习更深层次的特征，但其感受野大小是固定且单一的，无法融合不同尺度的特征，也未能充分利用跨通道特征之间的相互作用，ResNeSt[38]的设计弥补了这些缺点。第一，ResNeSt 网络借鉴 GoogleNet[32]采用多路径机制，每个模块由不同大小的卷积核组成，在网络层数足够深的情况下可以提取到不同尺度特征同时减小计算量。第二，为了进一步提取多样性的目标特征，ResNeSt 网络借鉴 ResNeXt[58]的设计思想，在残差块中采用组卷积、多分支的架构。不同组之间形成的不同子空间可以让网络学到更丰富的多样性特征。第三，设计的网络架构在保证良好的特征多样性提取能力情况下，让网络聚焦于局部信息，避免噪声干扰，有助于整个网络实现更精准的图像识别任务。因此，ResNeSt 网络借鉴 SENet[35]和 SKNet[36]的思想，将注意力机制的思想引入分组卷积中，不仅建模通道之间的重要程度，建立通道注意力，同时用非线性方法聚合来自多个卷积核的信息，建立特征图注意力。最终的 ResNeSt 网络是基于 ResNet[33]进行修改的，在单个网络内合并特征图的拆分注意力，将通道维度的注意力机制扩展到特征图组表示，形成模块化。

图 9 展示了 ResNeSt 模块，其中包含特征图组（Feature-map Group）和拆分注意力（Split Attention）操作。特征图组中，特征图被分为多个组，每个

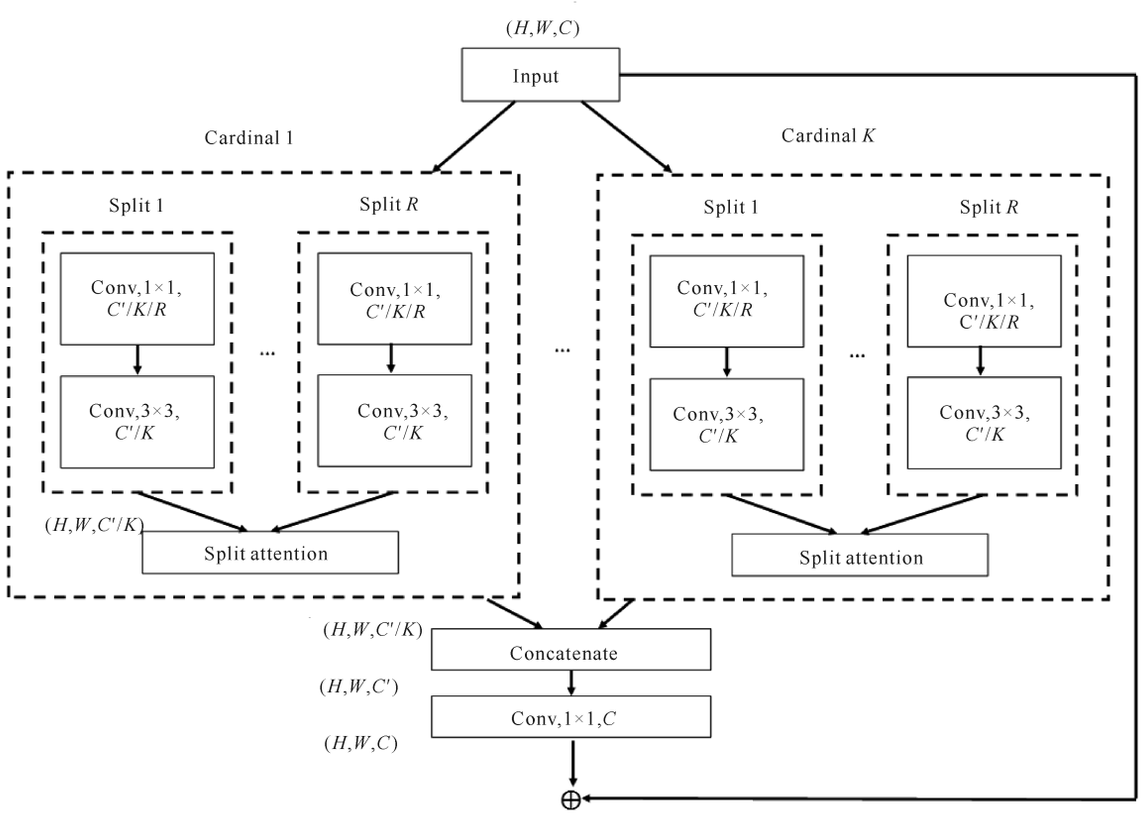

图 9　ResNeSt 模块

Fig. 9　ResNeSt module

组内又进行分组。超参数 K 和 R 分别表示特征图组的数量和基数组内的分组数,基数组内的每个分组称为 Splits,总的特征图分组数可以表示为 $G=K\times R$。在基数组中的每个 Split 进行 1×1 和 3×3 的卷积,得到 R 个特征图后进行拆分注意力操作。每个基数组得到的输出进行 Concat 操作,再与 Shortcut 路径中 1×1 卷积配合。多个 ResNeSt 模块堆叠最终组成 ResNeSt 网络。

ResNeSt 网络与现有的 ResNet 变体相比,不需要增加额外的计算量,且可以作为其他任务的骨架。ResNeSt 在 ImageNet 图像分类数据集上的准确率超越了 ResNet[33]、ResNeXt[58]、SENet[35] 和 EfficientNet[59],是当前 ResNet 网络的最强改进版本。

3 ImageNet 数据集上各类网络模型性能对比

本文在前面章节中介绍并分析了在卷积神经网络中图像分类领域网络的发展现状。以下是在相关参考文献的实验部分使用 ImageNet 数据集进行分类得到的 Top-1 性能对比。表 1 展示了在 ImageNet 数据集上不同 CNN 模型的性能对比,仅选取最优结果。GFLOPs 表示每秒 10 亿次的浮点运算数,Params 表示参数数量。

从表 1 可以看出,第一,网络的深度越深,准确率通常越高,参数量和计算量也会相应增加。ResNet 网络从 50 层增加到 152 层、ResNeXt 从 50 层增加到 101 层、DenseNet 从 121 层增加到 264 层、SENet 从 50 层增加到 154 层、SKNet 从 50 层增加到 101 层、ResNeSt 从 50 层增加到 101 层时,网络性能均有所提升,证明增加网络的深度是提高分类效果的重要因素,卷积网络深度越深,能获取的信息越多,得到的特征也越丰富,准确率就越高。但此时不可避免的是网络越深,神经元的数量也就越多,必然导致参数量和运算量的增加。第二,当网络深度到达一定程度时,网络会出现退化的现象,准确率下降,但增大输入图像的大小进行训练,在一定程度上可以缓解网络的退化,网络能识别到的图像信息越多。ResNet 从 101 层增加到 152 层时,输入 224×224 的图像进行训练时准确率出现下降。尽管 ResNet 能够训练很深的网络,但也存在一定瓶颈,在网络深度更深的情况下,网络无法更好收敛。当输入图像大于 224×224 时,ResNet 网络从 101 层增加到 152 层时准确率反而上升,说明增大输入图像的大小可以改善网络的分类性能。第三,在更深的网络模型中,DenseNet 具有较强的竞争力,当网络深度增加到 264 层时,分类准确率仍然有所提升。DenseNet 脱离加深网络层数和加宽网络结构来提升网络性能的定式思维,从特征的角度考虑,通过特征重用、旁路(Bypass)设置以及密集连接的特性很大程度上缓解了过拟合问题的产生,能够训练比 ResNet 更深的网络且保证准确率不受深度影响而下降。第四,多尺度和多分支的网络相较于同深度的其他网络获得的准确率更高,如 Inception 网络、SKNet 网络和 ResNeSt 网络,这类网络能够对不同粒度的特征进行采样,获取更加强大的特征表达,从而提高准确率。第五,在神经网络中,注意力模块通常是一个额外的神经网络,能够给输入的不同部分分配不同的权重。引入 SE、SK、CBAM 等注意力模块的网络,相较于基线网络的准确率高,这类网络模型能够忽略无关信息,不断聚焦到最具辨别性的区域,进一步提高网络的分类准确率,且增加的参数量很小。第六,通过将 ResNeSt 与其他 50 层和 101 层配置、类似复杂度的 ResNet 变体作比较,ResNeSt 的 Top-1 准确率达到最高,且参数量没有大幅度增加,性能得到显著提升,网络的泛化能力极强。说明在卷积神经网络中,分组卷积、多尺度方法、注意力机制等思想的共同引入可以从卷积、通道、特征图等方面全面理解图像信息,进一步提升网络的分类效果。

表 1 ImageNet 数据集上各类网络模型性能对比

Table 1 Performance comparison of various network models on ImageNet dataset

卷积神经网络模型 CNN model	Top-1 准确率 Top-1 accuracy		GFLOPs	Params (M)
	224×224	320×320/299×299		
VGG16[31]	72.98	—	15.47	—
VGG16-SE[35]	74.78	—	15.48	—
Inception-v3[44]	—	78.80	5.73	27.1
Inception-v4[45]	—	80.00	12.31	42.0
Inception-ResNet-v2[45]	—	80.10	13.22	55.0
ResNet-50[33]	76.15	76.86	4.14	25.5
ResNet-101[33]	77.37	78.17	7.87	44.5
ResNet-152[33]	77.00	78.70	—	—
ResNeXt-50[58]	77.77	78.95	4.24	25.0
ResNeXt-101[58]	78.89	80.14	7.99	44.3
DenseNet-121[34]	76.39	—	—	20.0

续表 1

Continued table 1

卷积神经网络模型 CNN model	Top-1 准确率 Top-1 accuracy		GFLOPs	Params (M)
	224×224	320×320/299×299		
DenseNet-169[34]	77.92	—	—	—
DenseNet-201[34]	78.54	—	—	—
DenseNet-264[34]	79.20	—	—	—
ResNeXt-50+CBAM[60]	78.60	79.62	4.25	27.7
ResNeXt-101+CBAM[60]	79.40	80.58	8.00	49.2
SENet-50[35]	78.88	80.29	4.25	27.7
SENet-101[35]	79.42	81.39	8.00	49.2
SENet-154[35]	81.32	82.72	—	—
Res2Net-50[37]	77.99	78.59	4.20	25.0
Res2NeXt-50[37]	78.24	—	4.20	25.0
SE-Res2Net-50[37]	78.44	—	4.20	25.0
SKNet-50[36]	79.21	80.68	4.47	27.5
SKNet-101[36]	79.81	81.60	8.46	48.9
ResNeSt-50-fast[38]	80.64	81.43	4.34	27.5
ResNeSt-50[38]	81.13	81.82	5.39	27.5
ResNeSt-101-fast[38]	81.97	82.76	8.07	48.2
ResNeSt-101[38]	82.27	83.00	10.20	48.3

4 展望

本文介绍了卷积神经网络在图像分类中的优点以及发展趋势,分析了卷积神经网络在图像分类中的经典模型、近年来的改进方法及其在 ImageNet 公共数据集上的性能表现。尽管卷积神经网络在图像分类领域取得丰硕的成果,但在实际应用中仍存在一些具有挑战性的研究问题:

(1)在深度学习中进行的图像分类研究难免需要针对特定任务数量庞大的图像数据集,但这些数据集不易获取和采集,导致目前使用监督学习方法的图像分类研究进展缓慢,大部分提出的创新性网络均在仅有的几个大型公共数据集上进行评估,针对特定分类任务的研究十分零散且数据集样本较少。对于一个新的应用领域,在数据集较少的情况下,有相关的研究[61,62]采用迁移学习方法处理小样本数据集进行图像分类,该方法使用在 ImageNet 大型数据集上训练得到的模型和参数,通过迁移训练方法进一步优化使用小样本目标数据集训练的模型,从而有效地利用小型数据并保证网络具有良好的鲁棒性和泛化能力。因此迁移学习方法应用在特定的图像分类任务中可以取得较好的分类效果。

(2)自然场景下的图像分类具有巨大的挑战性,这类数据集的图像往往含有较大的噪声,背景复杂,导致网络无法准确分辨出目标物体的类别,使得分类准确率较低。此外细粒度的图像分类由于子类别间细微的类间差异以及较大的类内差异,在某些类别上甚至连专家都难以区分,较之普通的图像分类任务,具有较大难度。自然场景下的图像分类任务无疑在现实世界中更具有意义,且此类数据集的收集和处理较容易,有利于切实推进现实的生产工作。细粒度图像分类更是对卷积神经网络挖掘细粒度特征发起的一大挑战,利于图像分类领域的研究发展。

(3)目前有相当一部分图像分类网络的深度较深、模型结构设计复杂甚至存在冗余结构,其参数量增大导致计算成本增加。DenseNet 网络虽然从特征的角度考虑,通过特征重用加强特征的传递,但是网络的连接十分冗余,对其进行剪枝将成为未来的研究方向。虽然 Res2Net 的计算复杂度与等效的 ResNet 相似,但它的运行速度仍然比对应的 ResNet 慢,不能高效地处理图像识别任务。ResNeSt 网络虽然集合多种优秀模型的优点,在 ImageNet 数据集上的实验效果也很优秀,但是网络模型结构太过臃肿,并且训练模型需要大量的调参技巧,使用多种模型堆叠设计的痕迹较重。另外在当前移动设备普及的情况下,这种复杂度高、深度太深的网络并不能在移动端很好地应用。因此,未来的研究热点必将趋向轻量级、高效的网络发展。

(4)深度学习方法在图像分类中的可解释性差,不同于传统图像分类方法有严谨的数学理论依据作为支撑,目前使用深度学习的方法更偏向于将网络进行可视化解释网络观察到所分类图像的某个位置。如何从更科学的角度诠释深度学习方法将是未来一个十分重要的研究课题。

(5)虽然注意力机制模块可以一定程度上提升分类模型的性能,但 SENet 实际上只是从通道的角度关注特征图,没有从空间、非局部、全局等角度关注特征图,未来的研究工作可以将这些角度很好地整合在一起,形成一种关注能力更强的注意力机制。SKNet 使用时涉及到分组数量、卷积核大小的选择问题,在实际使用中仍需要经过大量的实验调整参数才能找

到合适的参数设置,不能真正做到脱离人工设定分组数量和卷积核大小来适用于不同尺度的目标对象,使网络达到最优。此外,全局池化的使用容易混淆多个物体的尺度信息,使分类精度降低。注意力机制模块在网络中的放置位置也会影响到神经网络的整体效果,同样需要经过大量实验找到适合的添加位置才能达到最佳的实验效果,灵活性较低。因此,如何设计出能够真正自动适应神经网络训练的注意力机制是一个十分有趣的研究方向。

总的来说,相对于传统图像分类方法,基于深度学习的卷积神经网络在特征表示上具有极大的优越性,且应用广泛,性能优异。随着研究的深入、数据集数量的增加和实际应用场景的增多,卷积神经网络的模型必然更复杂,更具有挑战性。因此,在未来的深度学习研究中,使用卷积神经网络进行改进和创新依旧是最终实现人工智能的最佳途径和方法。

参考文献

[1] FAN Q,ZHUO W,TANG C K,et al. Few-shot object detection with attention-RPN and multi-relation detector [C]. 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle,WA,USA:IEEE,2020:4013-4022.

[2] ZHANG S,CHI C,YAO Y,et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]. 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 9759-9768.

[3] IBRAHIM M S,VAHDAT A,RANJBAR M,et al. Semi-supervised semantic image segmentation with self-correcting networks [C]. 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle,WA,USA:IEEE,2020:12715-12725.

[4] PENG S,JIANG W,PI H,et al. Deep snake for real-time instance segmentation [EB/OL]. [2020-07-01]. https://arxiv.org/abs/2001.01629v1.

[5] YANG L,ZHUANG J,FU H,et al. SketchGCN:Semantic sketch segmentation with graph convolutional networks [EB/OL]. [2020-07-02]. https://arxiv.org/abs/2003.00678.

[6] XIE E,SUN P,SONG X,et al. PolarMask:Single shot instance segmentation with polar representation [EB/OL]. [2020-07-26]. https://arxiv.org/abs/1909.13226.

[7] CHEN H,SUN K,TIAN Z,et al. BlendMask:Top-Down meets bottom-up for instance segmentation [EB/OL]. [2020-07-26]. https://arxiv.org/abs/2001.00309.

[8] SHI Y,YU X,SOHN K,et al. Towards universal representation learning for deep face recognition [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA:IEEE,2020.

[9] WANG K,PENG X,YANG J,et al. Suppressing uncertainties for large-scale facial expression recognition [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA:IEEE,2020.

[10] LI L,BAO J,ZHANG T,et al. Face X-ray for more general face forgery detection [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA:IEEE,2020.

[11] MUNRO J,DAMEN D. Multi-modal domain adaptation for fine-grained action recognition [C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul,Korea:IEEE,2019.

[12] ZHANG F,ZHU X,DAI H,et al. Distribution-aware coordinate representation for human pose estimation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA:IEEE,2020.

[13] HUANG J,ZHU Z,GUO F,et al. The devil is in the details:Delving into unbiased data processing for human pose estimation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle,WA,USA:IEEE,2020.

[14] HINTON G E,OSINDERO S,THE Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation,2006,18(7):1527-1554.

[15] SALAKHUTDINOV R,HINTON G. An efficient learning procedure for deep boltzmann machines [J]. Neural Computation,2012,24(8):1967-2006.

[16] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-basedlearning applied to document recognition [J]. Proceedings of the IEEE,1998,86(11):2278-2324.

[17] LECUN Y,BOSER B,DENKER J S,et al. Backpropagation appliedto handwritten zip code recognition [J]. Neural Computation,1989,1(4):541-551.

[18] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks [C]. Conference and Workshop on Neural Information Processing Systems,2012.

[19] LONG J,SHELHAMER E,DARRELL T. Fully conv-

olutional networks for semantic segmentation [C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015.

[20] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, 1: 91-99.

[21] TOSHEV A, SZEGEDY C. DeepPose: Human pose estimation via deep neural networks [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014.

[22] HUBEL D H. Evolution of ideas on the primary visual cortex, 1955 – 1978: A biased historical account [J]. 1982, 2(7): 435-469.

[23] WIESEL T N. Postnatal development of the visual cortex and the influence of environment [J]. Nature, 1982, 299: 583-591.

[24] FUKUSHIMA K, MIYAKE S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition [M]// AMARI S, ARBIB M A. Competition and cooperation in neural nets. Berlin, Heidelberg: Springer, 1982: 267-285.

[25] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[26] COOK J A, RANSTAM J. Overfitting [J]. British Journal of Surgery, 2016, 103(13): 1814.

[27] ANTOL S, AGRAWAL A, LU J, et al. VQA: Visual question answering [C]. The 2015 IEEE International Conference on Computer Vision. Santiago, Chile. Piscataway: IEEE, 2015: 2425-2433.

[28] LI H, LIU H, JI X, et al. CIFAR10-DVS: An event-stream dataset for object classification [J]. Frontiers in Neuroence, 2017, 11: 309. DOI: 10. 3389/fnins. 2017. 00309.

[29] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009). Miami, Florida, USA: IEEE, 2009.

[30] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]. Proceedings of Advances in Neural Information Processing Systems, Lake Tahoe, USA, 2012, 1: 1097-1105.

[31] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-06-20]. http://citeseerx. ist. psu. edu/viewdoc/download? doi = 10. 1. 1. 740. 6937&rep = rep1&type = pdf.

[32] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.

[33] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 770-778.

[34] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks [EB/OL]. [2020-06-23]. https://arxiv. org/abs/1608. 06993.

[35] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[36] LI X, WANG W, HU X, et al. Selective kernel networks [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 510-519.

[37] GAO S H, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access), 2019, 32(2): 652-662. DOI: 10. 1109/TPAMI. 2019. 2938758.

[38] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: Split-attention networks [EB/OL]. [2020-06-30]. https://arxiv. org/abs/2004. 08955? context = cs. CV.

[39] 孙新立. 基于卷积神经网络的煤矸石图像识别[J]. 电脑知识与技术, 2020, 16(21): 16-22.

[40] 田佳鹭, 邓立国. 基于改进 VGG16 的猴子图像分类方法[J]. 信息技术与网络安全, 2020, 39(5): 6-11.

[41] SHICHIJO S, NOMURA S, AOYAMA K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images [J]. EBioMedicine, 2017, 25: 106-111.

[42] ZHUO L, JIANG L, ZHU Z, et al. Vehicle classification for large-scale traffic surveillance videos using Convolutional Neural Networks [J]. Machine Vision and Applications, 2017, 28: 793-802.

[43] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-02-11) [2020-06-20]. https://arxiv. org/abs/1502. 03167.

[44] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 2818-2826.

[45] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [EB/OL]. [2020-06-20]. https://arxiv.org/abs/1602.07261.

[46] PASCANU R, MIKOLOV T, BENGIO Y. Understanding the exploding gradient problem [EB/OL]. [2020-06-23]. https://arxiv.org/abs/1211.5063v1.

[47] SQUARTINI S, PAOLINELLI S, PIAZZA F. Comparing different recurrent neural architectures on a specific task from vanishing gradient effect perspective [C]. 2006 IEEE International Conference on Networking, Sensing and Control. Ft. Lauderdale, FL, USA: IEEE, 2006: 380-385.

[48] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. Proceedings of the 32nd International Conference on International Conference on Machine Learning. 2015, 37: 448-456.

[49] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3(4): 212-223.

[50] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks [C]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2018: 421-429.

[51] QIN X, JIANG J, FAN W, et al. Chinese cursive character detection method [J]. The Journal of Engineering, 2020(13): 626-629.

[52] ZHAI M, XIANG X, LV N, et al. SKFlow: Optical flow estimation using selective kernel networks [J]. IEEE Access, 2019, 7: 98854-98865.

[53] JI H, LIU Z, YAN W Q, et al. Early diagnosis of alzheimer's disease based on selective kernel network with spatial attention [J]. Asian Conference on Pattern Recognition, 2019: 503-515.

[54] CHEN C F, FAN Q, MALLINAR N, et al. Big-little net: An efficient multi-scale feature representation for visual and speech recognition [C]. International Conference on Learning Representations, 2019.

[55] CHEN Y, FAN H, XU B, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019.

[56] CHENG B, XIAO R, WANG J, et al. High frequency residual learning for multi-scale image classification [C]. British Machine Vision Conference (BMVC), 2019.

[57] SUN K, ZHAO Y, JIANG B, et al. High-resolution representations for labeling pixels and regions [EB/OL]. [2020-06-23]. https://arxiv.org/abs/1904.04514.

[58] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks [C]. Computer Vision and Pattern Recognition, 2017: 5987-5995.

[59] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, 2019: 6105-6114.

[60] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. European Conference on Computer Vision, 2018: 3-19.

[61] WU Y, QIN X, PAN Y, et al. Convolution neural network based transfer learning for classification of flowers [C]. 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP). Shenzhen, China: IEEE, 2018: 562-566.

[62] SEO Y, SHIN K S. Image classification of fine-grained fashion image based on style using pre-trained convolutional neural network [C]. 2018 IEEE 3rd International Conference on Big Data Analysis. Shanghai, China: IEEE, 2018: 387-390.

Research Progress of Image Classification based on Convolutional Neural Network

QIN Xiao[1], HUANG Chengcheng[1], SHI Yu[1], LIAO Zhaoqi[1], LIANG Xinyan[1], YUAN Chang'an[2]

(1. BAGUI Scholar Innovation Team Laboratory, School of Computer & Information Engineering, Nanning Normal University, Nanning, Guangxi, 530000, China; 2. Guangxi Academy of Scienses, Nanning, Guangxi, 530007, China)

Abstract: The advantages of image classification algorithms based on convolutional neural network are unmatched by traditional methods. Convolutional neural network uses its designed network structure and weight sharing characteristics to learn abstract features from the bottom of the image to the high-level semantics from a huge amount of training data. End-to-end learning eliminates the need for data labeling before the execution of each independent learning task. Over the years, after research and experimentation by researchers, the convolutional neural network has evolved a variety of optimized structures from the first multilayer neural network model, and its performance has been continuously improved. This article introduces the research progress of image classification algorithm based on convolutional neural network, describes the classic model of convolutional neural network in image classification and the improved methods in recent years. Each model is been analyzed, and the performance of various methods on ImageNet public dataset are shown. Finally, the research of image classification algorithm based on convolutional neural network is summarized and prospected.

Key words: convolutional neural network, image classification, classic model, improved methods, performance comparison

责任编辑:陆　雁