



计算机科学与探索

Journal of Frontiers of Computer Science and Technology

ISSN 1673-9418, CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目: 卷积神经网络模型发展及应用
作者: 严春满, 王铖
网络首发日期: 2020-09-29
引用格式: 严春满, 王铖. 卷积神经网络模型发展及应用. 计算机科学与探索.
<https://kns.cnki.net/kcms/detail/11.5602.TP.20200929.1012.007.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

卷积神经网络模型发展及应用

严春满, 王 铖⁺

西北师范大学 物理与电子工程学院, 兰州 730070

+ 通信作者 E-mail: 610487244@qq.com

摘 要: 深度学习是机器学习和人工智能研究的最新趋势, 作为一个十余年来快速发展的崭新领域, 越来越受到研究者的关注。卷积神经网络模型是深度学习模型中最重要的一种经典结构, 其性能在近年来深度学习任务上逐步提高。由于可以自动学习样本数据的特征表示, 卷积神经网络已经广泛应用于图像分类、目标检测、语义分割以及自然语言处理等领域。本文首先分析了典型卷积神经网络模型为提高其性能增加网络深度以及宽度的模型结构, 分析了采用注意力机制进一步提升模型性能的网络结构, 然后归纳分析了目前的特殊模型结构, 最后总结并讨论了卷积神经网络在相关领域的应用, 并对未来的研究方向进行展望。

关键词: 卷积神经网络模型; 特征提取; 计算机视觉; 自然语言处理

文献标志码: A **中图分类号:** TP391.4

严春满, 王铖. 卷积神经网络模型发展及应用[J]. 计算机科学与探索

YAN C M, WANG C. Development and Application of the Convolutional Neural Network Model[J]. Journal of Frontiers of Computer Science and Technology

Development and Application of the Convolutional Neural Network Model

YAN Chunman, WANG Cheng⁺

College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China

Abstract: Deep learning is the latest trend of machine learning and artificial intelligence research. As a new field of rapid development in the past ten years, more and more researchers pay attention to it. Convolutional neural network model is one of the most important classical structures in deep learning model, and its performance has been gradually improved in recent years. Since it can automatically learn the feature representation of sample data, convolutional neural network has been widely used in image classification, object detection, semantic segmentation and natural language processing. This paper first analyzes the model structure of typical convolutional neural network model, which increases the depth and width of the network in order to improve its performance, analyzes the network structure that uses attention mechanism to further improve the performance of the model, and then summarizes and analyzes the current special model structure. Finally, it summarizes and discusses the application of convolutional neural network in related fields, and develops the future research direction Hope.

Key words: Convolution neural network model; feature extraction; computer vision; natural language processing

* The Natural Science Foundation of China under Grant No. 61861041 (国家自然科学基金).

1 引言

卷积神经网络 (Convolution Neural Network, CNN) 在计算机视觉^[1-6]、自然语言处理^[7-9]等领域已被广泛应用。在卷积神经网络兴起之前, 主要依靠人工针对特定的问题设计算法, 比如采用 Sobel、LOG、Canny、Prewitt 等^[10-13]算子进行边缘检测, 采用 Harris、DOG、FAST、SIFT 等^[14-17]用于角点等特征检测, 并且采用传统分类器如 K 近域、支持向量机、稀疏分类器等^[18-20]进行分类。特征提取和分类器的设计是图片分类等任务的关键, 对分类结果的好坏有着最为直接的影响。卷积神经网络可以自动地从训练样本中学习特征并分类, 解决了人工特征设计的局限性。神经网络的思想起源于 1943 年 McCulloch 和 Pitts 提出的神经元模型^[21], 简称 MCP 神经元模型。它是利用计算机来模拟人的神经元反应的过程, 具有开创性意义。此模型将神经元反应简化为三个过程: 输入信号线性加权、求和、非线性激活。1958 年到 1969 年为神经网络模型发展的第一阶段, 称为第一代神经网络模型。在 1958 年 Rosenblatt 第一次在 MCP 模型上增加学习功能并应用于机器学习, 发明了感知器算法^[22], 该算法使用 MCP 模型能够采用梯度下降法从训练样本中自动学习并更新权值, 并能对输入的多维数据进行二分类, 其理论与实践的效果引起了神经网络研究的第一次浪潮。1969 年美国数学家及人工智能先驱 Minsky 在其著作中证明感知器本质上是一种线性模型^[23], 只能处理线性分类问题, 最简单的异或问题都无法正确分类, 因此神经网络的研究也陷入了近 20 年的停滞。1986 年到 1988 年是神经网络模型发展的第二阶段, 称为第二代神经网络模型。1986

年 Rumelhart 等人提出了误差反向传播算法(Error Back Propagation Algorithm, 简称 BP 算法)^[24]。BP 算法采用 Sigmoid 进行非线性映射, 有效解决了非线性分类和学习的问题, 掀起了神经网络第二次研究高潮。BP 网络是迄今为止最常用的神经网络, 目前大多神经网络模型都是采用 BP 网络或者其变化形式。早期神经网络缺少严格数学理论的支撑, 并且在此后的近十年时间, 由于其容易过拟合以及训练速度慢, 并且在 1991 年反向传播算法被指出在后向传播的过程中存在梯度消失的问题^[25], 神经网络再次慢慢淡出人们的视线。

1998 年 LeCun 发明了 LeNet-5, 并在 Mnist 数据集达到 98% 以上的识别准确率, 形成影响深远的卷积神经网络结构, 但此时神经网络的发展正处于下坡时期, 没有引起足够的重视。

从感知机提出到 2006 年以前, 此阶段称为浅层学习, 2006 年至今是神经网络的第三阶段, 称为深度学习。深度学习分为快速发展期 (2006 年-2012 年) 和爆发期 (2012 年-至今), 2006 年 Hinton 提出无监督的“逐层初始化”策略以降低训练难度, 并提出具有多隐层的深度信念网络 (Deep Belief Network, DBN)^[26], 从此拉开了深度学习大幕。

随着深度学习理论的研究和发展, 研究人员提出了一系列卷积神经网络模型。为了比较不同模型的质量, 我们收集并整理了文献中模型在分类任务上的识别率, 如图 1 所示。由于部分模型并未在 ImageNet 数据集测试识别率, 我们给出了其在 Cifar-100 或 mnist 数据集上的识别率。其中, Top-1 识别率指的是 CNN 模型预测出最大概率的分类为正确类别的概率。Top-5 识别率指的是 CNN 模型预测出最大概率的前 5 个分类里有正确类别的概率。

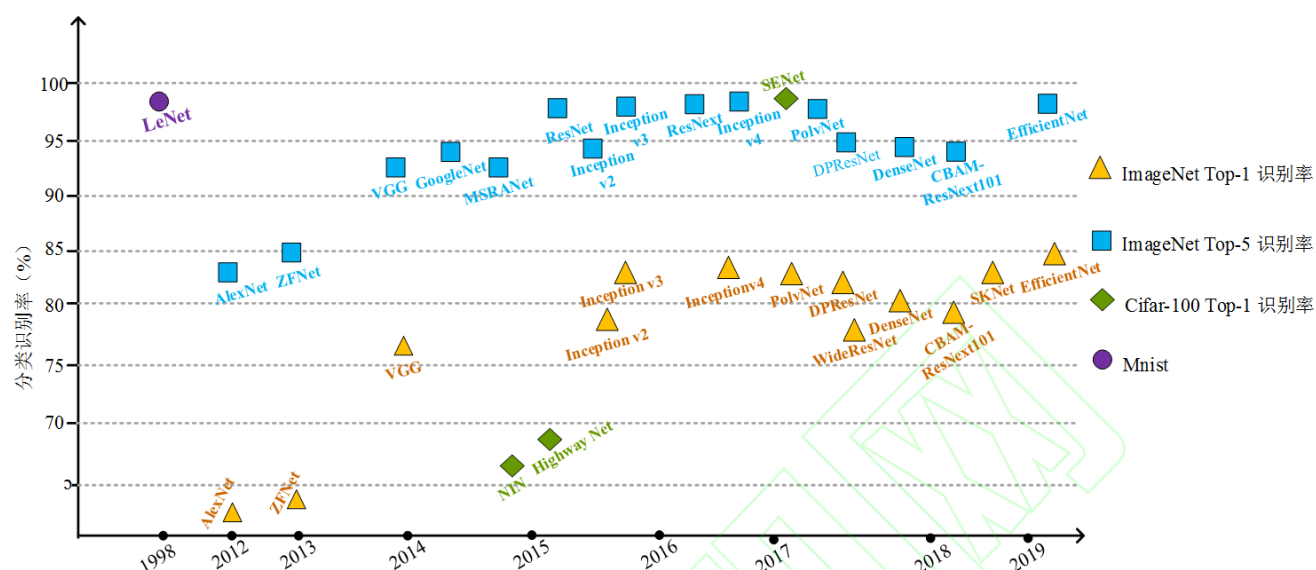


Fig.1 Recognition accuracy of typical CNN models in classification tasks

图 1 典型 CNN 模型在分类任务中的识别准确率

2012 年, 由 Alex Krizhevshy 提出的 AlexNet 给卷积神经网络迎来了历史性的突破。AlexNet 在百万量级的 ImageNet 数据集上对图像分类的精度大幅度超过传统方法, 一举摘下了视觉领域竞赛 ILSVRC2012 的桂冠。自 AlexNet 之后, 研究者从卷积神经网络的结构出发进行创新, 主要有简单的堆叠结构模型, 比如 ZFNet、VGGNet、MSRNet。堆叠结构模型通过改进卷积神经的基本单元并将其堆叠以增加网络的深度提升模型性能, 但仅在深度这单一维度提升模型性能具有瓶颈; 后来在 NIN 模型提出使用多个分支进行计算的网中网结构模型, 使宽度和深度都可增加, 具有代表性的模型有 Inception 系列模型等; 随着模型深度以及宽度的增加, 网络模型出现参数量过多、过拟合以及难以训练等诸多问题。ResNet 提出残差结构后, 为更深层网络构建提出解决方案, 随即涌现出很多残差结构模型, 比如基于 ResNet 改进后的 ResNeXt、DenseNet、PolyNet、WideResNet, 并且 Inception 也引入残差结构形成了 Inception-ResNet-block, 以及基于残差结构并改进其特征通道数量增加方式的 DPRNet; 与之前空间维度上提升模型性能的方法相比, 注意力机制模型通过通道注意力和空间注意力机制可以根据特征通道重要程度进一步提升模型性能, 典型的模型为 SENet、SKNet 以及 CBAM。

传统的卷积神经网络模型性能十分优秀, 已经应用到各个领域, 具有举足轻重的地位。由于卷积神经网络的模型十分丰富, 有些模型的结构或用途比较特殊, 在本文中统称为特殊模型, 包括具有简单的结构和很少参数量的挤压网络模型 SqueezeNet, 采用无监督学习的生成对抗网络模型 (GAN), 具有完全相同的两路网络结构以及权值的孪生神经网络模型 SiameseNet, 以及通过线性运算生成其他冗余特征图的幽灵网络 GhostNet。

由于卷积神经网络的一系列突破性研究成果, 并根据不同的任务需求不断改进, 使其在目标检测、语义分割、自然语言处理等不同的任务中均获得了成功的应用。

基于以上认识, 本文首先概括性地介绍了卷积神经网络的发展历史, 然后分析了典型的卷积神经网络模型通过堆叠结构、网中网结构、残差结构以及注意力机制提升模型性能的方法, 并进一步介绍了特殊卷积神经网络模型的以及模型结构, 最后讨论了卷积神经网络在目标检测、语义分割以及自然语言处理领域的典型应用, 并对当前深度卷积神经网络存在的问题以及未来发展方向进行探讨。

2 基本卷积神经网络及运算单元

LeNet-5^[27]是最早的基本卷积神经网络, 其结构

如图 2 所示, 包含了 CNN 的基本运算单元: 卷积层、池化层、非线性单元、全连接层。典型的 CNN 体系结构通常为卷积层与池化层交替, 最后通过一

个或者多个全连接层输出结果。在某些情况下采用全局平均池化层替换全连接层, 并加入了批归一化等操作进一步优化 CNN 的性能。

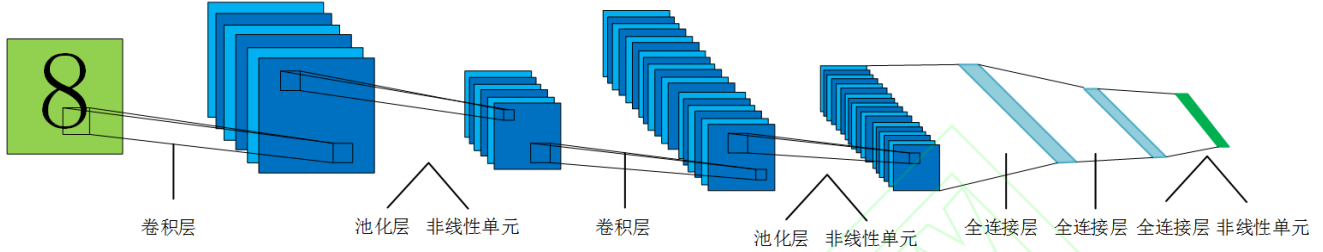


Fig.2 Structure of LeNet-5

图 2 LeNet-5 结构

1) 卷积层 (Convolution Layer, Conv) 又称之为特征提取层, 主要用于提取图像的特征。其由一组卷积核组成, 并且卷积核的权重值可以根据目标函数自动学习更新。卷积层是卷积神经网络的核心。

2) 池化层 (Pooling Layer) 又称为下采样层, 一般在连续的两层卷积层之间进行降维操作, 能够有效减少模型参数量以及减轻网络的过拟合现象。一般有最大池化层 (MaxPooling) 和平均池化层 (Average Pooling) 等。

3) 非线性单元由非线性激活函数构成, 根据其特性分为饱和和非线性激活函数, 如 Sigmoid 函数、Tanh 函数等, 以及不饱和非线性激活函数, 如 Relu 函数, Leaky Relu 函数等。非线性单元是对卷积层输出结果做非线性映射, 使得神经网络可以任意逼近任何非线性函数, 提升了模型的特征表达能力。

4) 批归一化 (Batch Normalization, BN) 就是将输入数据变换到标准正态分布中, 使非线性单元

输入值落入梯度较大的取值区间, 避免梯度消失问题, 并且加快模型训练的收敛速度。

5) 全连接层 (Fully Connected Layers, FC) 通常在神经网络的尾部用于分类任务, 会把卷积层输出的二维向量转化为一维向量。由于全连接层连接了上层所有的神经元权重, 所以存在众多冗余参数并占据了大量的硬件资源, 且破坏了图像的空间结构。目前普遍使用卷积层或全局平均池化层 (Global Average Pooling Layers, GAP) 代替全连接层。

3 典型卷积神经网络模型

本节首先对四种结构的网络模型从模型机制、优缺点、应用建议三个方面进行对比分析如表 1 所示。分析表 1 可见, 由于网络模型的机制不同, 在具体应用网络时需要根据网络特点选用, 并做出相应的优化。

Table 1 Comparative analysis of four model structures

表 1 四种模型结构对比分析

模型名称。	模型机制。	优缺点。	应用建议。
堆叠结构模型。	仅有一条主干网络。	网络结构简洁, 易于软硬件构建, 但深层网络难以训练。	应用于数据集较小的任务中。
网中网结构模型。	多条网络分支在多个尺度并行计算再聚合。	能提取更丰富的特征, 且能提高非线性表达能力, 但参数量大, 计算成本高。	尽可能避免应用于资源极度紧缺的设备中。
残差结构模型。	短路机制网络。	解决了深层网络模型构建问题以及退化问题, 但有随机深度现象。	应用于构建深层次网络。
注意力机制模型。	通道注意力与空间注意力。	增强模型特征自主提取能力, 但增加了模型复杂度以及计算量。	需进一步增强原有模型性能以及泛化能力时采用。

3.1 堆叠结构模型

堆叠结构模型指的是仅仅通过网络层堆叠而无其他拓扑结构形成的网络模型。早期的神经网络模型通过不断改进神经网络的基本运算单元，并将其堆叠形成网络模型，如 LeNet、AlexNet、ZFNet、VGGNet、MSRANet 等。LeNet 奠定了 CNN 的基础，而 AlexNet 网络在 LeNet 基础上将 Relu 激活函数作为非线性单元，并且添加 Dropout 以及局部响应归

一化层 (Local Response Normalization, LRN) 防止网络过拟合。由于早期 GPU 显存的限制, AlexNet^[28] 将一个堆叠结构模型拆分后通过两个 GPU 进行协同训练, 随着硬件平台发展, 可以单独使用一个 GPU 平台进行网络训练, 不再拆分模型结构。ZFNet 和 VGGNet 是典型的堆叠结构模型, 其模型结构如图 3 所示。

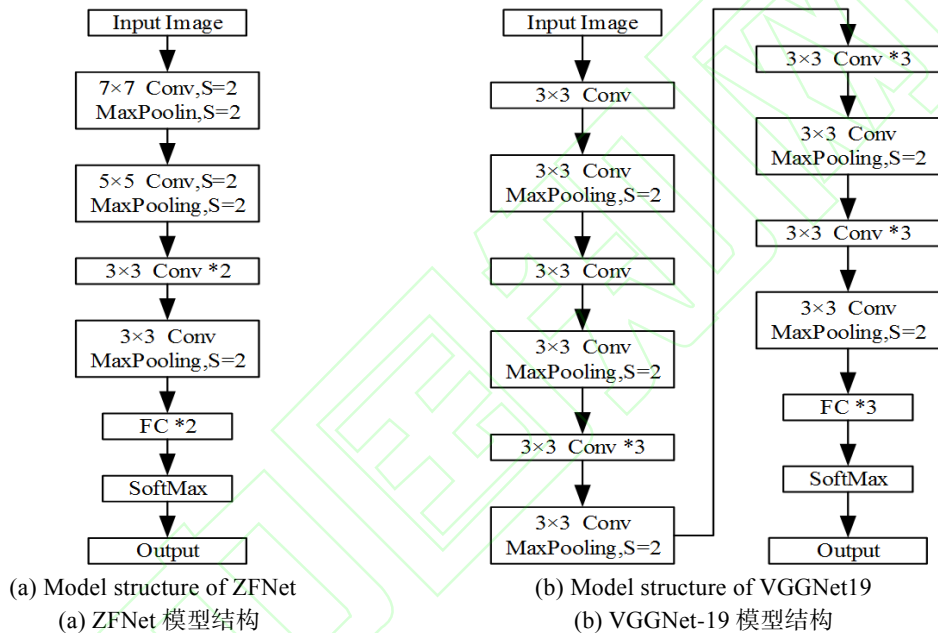


Fig.3 Two structural models of stacking
图 3 两种堆叠结构模型

在 2013 年之前, 提升 CNN 性能主要依靠反复实验, 并不明白改进的原因。ZFNet^[29] 网络结构与 AlexNet 类似, ZFNet 作者提出了多层反卷积可视化技术, 可监视网络中隐藏层的特征。作者在实验中发现 AlexNet 结构学习到的第一层特征只对于高频和低频信息有了收敛, 对于中层信息却没有收敛, 并且第二层特征出现了混叠失真。为了解决这些问题, 在 ZFNet 网络中减小了卷积核尺寸以及步长, 形成 ZFNet 模型结构如图 3 (a) 所示, 从而最大限度地提高了 CNN 的学习能力。通过可视化技术对 CNN 结构重新调整有助于分类性能的提升, 并且在以后卷积网络模型的发展中可以看到, 普遍采用了更小的 3×3 卷积核。

相比之下, VGGNet^[30] 则是探索 CNN 的深度, 并发现网络的深度是模型优良性能的关键部分。VGGNet 结构与 AlexNet 类似, 但在卷积中使用更小的 3×3 卷积核, 通过反复堆叠 3×3 的小型卷积核和 2×2 的最大池化层, 并且去掉了耗费计算资源的 LRN, 形成了更深的 VGGNet, 如图 3 (b) 所示为 19 层 VGGNet 结构。之所以使用 3×3 的卷积核, 是因为同时堆叠 3×3 小卷积核, 可以达到大尺寸卷积核的效果, 并且小尺寸卷积核可以减少参数的数量, 提供了较低的计算复杂性, 且有更多的非线性, 增强了模型的泛化能力。VGGNet 加深了模型的网络结构模型的深度, 相比 AlexNet 在 Imagenet 数据集上准确率得到大幅提升, 赢得了

ILSVRC2014 定位任务的冠军以及分类任务的亚军。

MSRANet^[31]与 VGGNet 结构类似,但是其针对非线性单元中的 Relu 函数在输入为负时,会导致神经元输出为 0 的“神经元死亡问题”提出了 PReLU 激活函数,合理的保留负向信息。

堆叠结构模型由于其仅仅在网络的深度单维度上进行提升,导致其性能并不突出。EfficientNet^[32]探索出在网络深度、网络宽度、图像分辨率三个维度上共同提升模型性能的思想,进一步提高堆叠结构模型的性能。堆叠结构模型由于其结构简单,增加模型深度容易,并且便于硬件以及软件实现,所以其应用十分广泛。

3.2 网中网结构模型

网中网结构模型是使用多个神经网络分支进行运算,再将各分支运算结果连接形成的网络模型。网中网结构模型在 NIN^[33]中提出,由于其采用较少的参数量就取得了 AlexNet 效果而产生深远影响。在各分类任务中输入特征通常是高度非线性的, NIN 网络在每个卷积层内引入一个微型网络,相比堆叠结构能更好的抽象每个局部块的特征,其网络结构如图 4 所示。

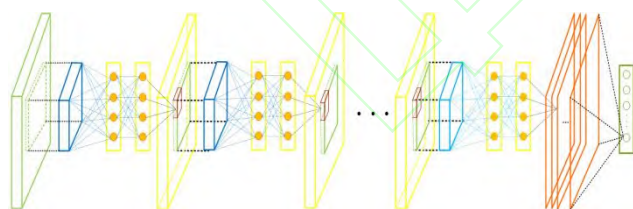


Fig.4 Structure of NIN model

图 4 NIN 模型结构

NIN 将如图 5(a)所示的广义线性卷积层 GLM 替换为如图 5(b)所示多层感知机 MLP,即在线性卷积层后添加 1×1 卷积层,作为 NIN 网络构建的基本单元,提高了特征的抽象表达能力,作者是第一个使用 1×1 卷积层构建网络,具有划时代的意义。在现在看来 1×1 卷积层还可以实现跨通道特征融合和通道升降维。

网中网结构模型以 Inception 系列模型为代表,其是由不同的 Inception block 构建而成。GoogLeNet

(Inception V1)^[34]是典型的网中网结构模型,与 VGGNet 和 AlexNet 相比,在网络结构做出巨大改变,通过网中网结构增加了模型深度以及宽度,相比 VGGNet 也有了更高的准确率的提升。

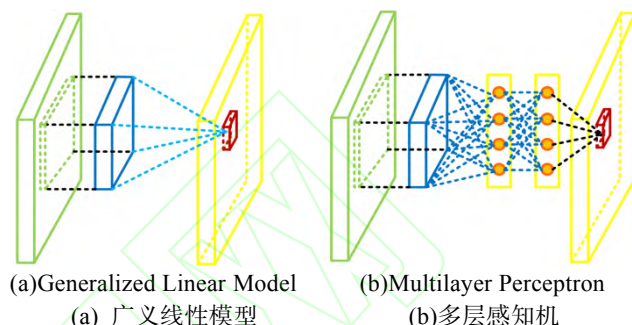
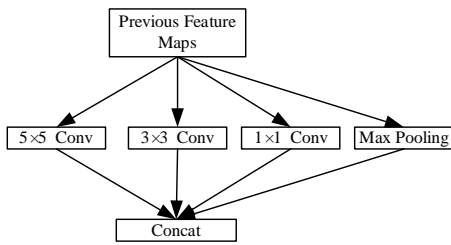


Fig.5 GLM and MLP layer

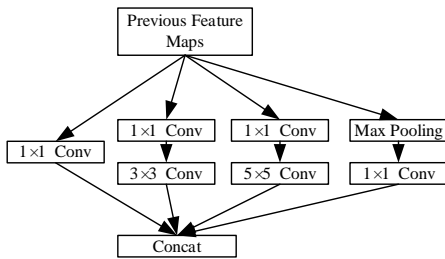
图 5 GLM 与 MLP 层

GoogLeNet 是由原始 Inception block 构建,其 block 结构如图 6(a)所示,通过拆分、变换和合并思想整合了多尺度卷积变换,包括 1×1 、 3×3 、 5×5 卷积以及一个池化操作并联而成,使模型具有不同大小的感受野,也增加了网络对多尺度的适应性,而池化操作,能防止模型过拟合同时也增强了网络的平移不变性,提高模型的泛化能力。但是,将多个 Google 模块不断叠加后,每个模块输入都是先前模块的输出特征图,导致深层的网络特征图通道数过多,且 5×5 的卷积核所需的计算量巨大。为了避免这种情况,Google 参考 NIN 网络,又提出了既能保持网络结构的稀疏性,又能利用硬件对密集矩阵高计算性能的 Inception V1 模块,如图 6(b)所示。模型在 3×3 和 5×5 卷积之前以及最大池化操作后加入 1×1 卷积构成瓶颈层,一方面,削减每一层的特性通道数量,大大减少了参数量节约计算成本;另一方面,将同一位置不同通道的特征连接在一起,加强了不同通道间的信息交流与特征融合。但瓶颈层会导致有用信息丢失,并减小了下一层的空间特征。为此,Inception 模块单独并联一路 1×1 卷积,这样既能保留上层原始特征,又能跨通道组织信息,进行信息整合。



(a) Original Inception block

(a) 原始 Inception 模块

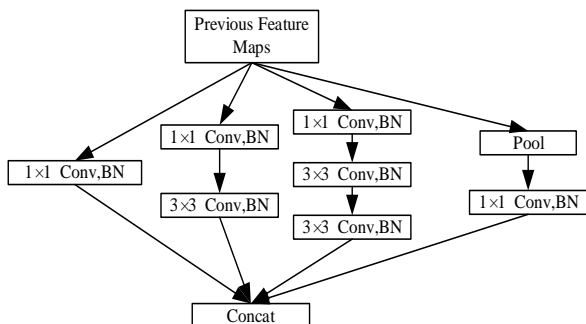


(b) Inception V1 block

(b) Inception V1 模块

Fig.6 Two kinds of concept modules**图 6 两种 Inception 模块**

Google 在 2015 年提出 Inception V2 模型^[35], 其 block 结构如图 7 所示。相对于 Inception V1 block 结构主要在两方面进行改进。一方面, 将 V1 模型中 5×5 卷积替换为两个 3×3 卷积, 加深了神经网络模型的深度, 降低了模型参数, 同时建立了更多的非线性变换, 使得 CNN 对特征的学习能力更强; 另一方面, 在每个卷积层后加入 BN 层, 避免梯度消失问题, 而且梯度变大意味着学习收敛速度快, 能大大加快训练速度提高学习效率。BN 层的加入对模型性能产生很大影响, 在此后的神经网络模型结构中成为“标配”。

**Fig.7 Inception V2 block****图 7 Inception V2 模块**

网中网结构模型在每个卷积层内引入一个微型网络后加深了网络的深度以及宽度, 增强了网络特征表达能力。

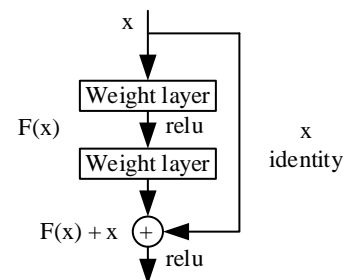
3.3 残差结构模型

残差结构模型是在结构中引入短路机制, 使模型的输出表述为输入和输入的非线性变换进行线性叠加的模型结构。卷积神经网络模型从 AlexNet 的 8 层, 到 VGGNet 的 19 层, 再到 GoogleNet 的 22 层, 模型的层数逐渐增多, 并且模型的性能越来越好, 更深的网络模型意味着更好的非线性表达能力, 能够更好的拟合复杂的特征。

但何凯明团队发现神经网络深度提升不能简单的通过层堆叠实现, 网络结构的过深后, 容易出现“退化现象”, 即随着网络深度加深, 网络训练误差却增大。为了解决退化现象, 残差结构引入恒等快捷链接, 其设计启发于 Schmidhuber 教授在 1997 年根据长短期记忆网络 (Long Short Term Memory Network, LSTM) 中门机制原理设计的 Highway Network^[36], 构造了如图 8 所示残差结构。残差结构直接把恒等映射作为网络的一部分, 使学习目标变为学习一个残差函数:

$$F(x) = H(x) - x \quad (1)$$

当 $F(x) = 0$ 时, 即堆积网络层仅仅做了恒等映射, 残差结构会保证网络加深后模型性能不会下降, 实际上 $F(x) \neq 0$, $F(x)$ 也会输入特征基础上学习到新的特征, 并通过残差结构与 $F(x)$ 叠加, 加强特征重用, 从而拥有更好的性能。

**Fig.8 Residual structure****图 8 残差结构**

残差结构中 $F(x)$ 的形式是灵活的, 通过构造不

同的 $F(x)$ 形成了两种构成 ResNet^[37] 的 block, 如图 10 所示。常规残差模型如图 9 (a) 所示, 当残差块加入 1×1 卷积层形成如图 9 (b) 所示结构, 思路与 Inception 一样, 通过第一层 1×1 卷积层形成瓶颈结构, 可以减少 3×3 卷积层计算量, 然后 3×3 卷积层进行特征提取, 最后 1×1 卷积层将通道数恢复, 整个过程是“压缩—卷积特征提取—扩张”。

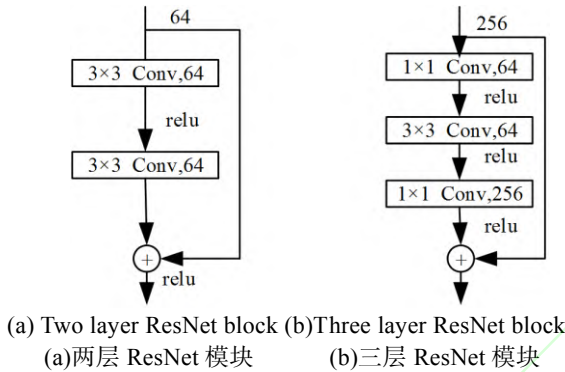


Fig.9 Two types ResNet block
图 9 两种 ResNet 模块

ResNet 在 VGGNet 基础上进行了修改, 通过残差结构使上一个残差块的信息没有阻碍的流入到

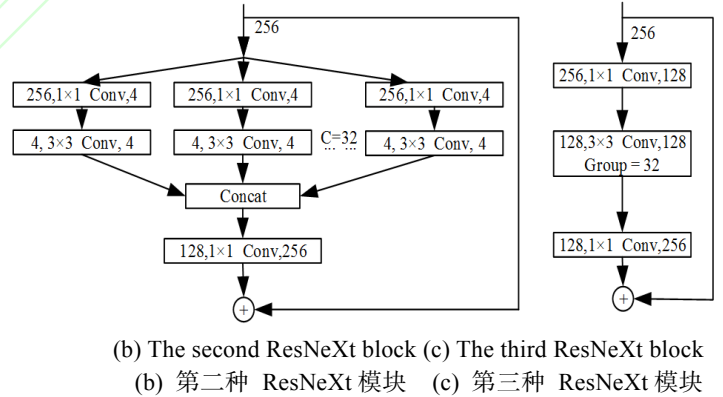
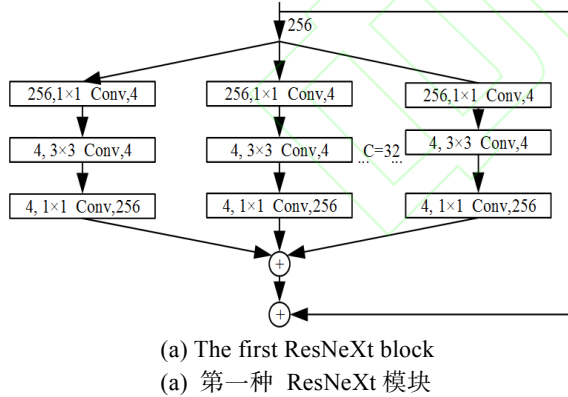


Fig.10 Three ResNeXt blocks
图 10 三种 ResNeXt 模块

DenseNet^[39] 结构将单层特征重用扩展到多层, 采用残差结构将每个层与其它所有层连接地密集卷积网络。Dense Block 内部均采用了密集连接, 在第 i 层与之前所有层特征映射, 即:

$$x_i = H_i([x_0, x_1, x_2, \dots, x_{i-1}]) \quad (3)$$

其中, $[x_0, x_1, x_2, \dots, x_{i-1}]$ 表示特征映射的级联, x_i 为第 i 层的特征输出, $H_i(\cdot)$ 为非线性变换函数组合。

下一个残差块, 增强了信息流通, 并且也避免了网络过深所引起退化问题。ResNet 通过残差结构达到了 152 层的深度, 为更深层的卷积网络提供了思路。

ResNeXt^[38] 是对 ResNet 的改进, ResNeXt 使用 block 相同的平行拓扑结构代替 ResNet 的 block, 即:

$$H(x) = x + \sum_{i=1}^C F_i(x) \quad (2)$$

在公式(2)中, C (cardinality) 是需要聚合的变换的组数, $F_i(x)$ 可以是任意函数, 在神经网络中为任意的网络结构。在 ResNeXt 中的 $F_i(x)$ 拥有同样的 block 结构, 聚合 C 个同样的 block 形成不同的 ResNeXtBlocks, 如图 10 所示为 $C=32$ 时不同的 ResNeXtBlocks, (a) 结构参数数量与单个 ResNet Block 差别很小, (b) 结构在各个分支经两层卷积后将各个分支的特征通道合并再进行卷积, 而 (c) 结构采用比较简洁而且速度更快的分组卷积实现, 作者在文中明确说明这三种结构是等价的。ResNeXt 的每个分支的拓扑结构是相同的, 不需要人工设计每个分支, 改进后的网络模型泛化能力相比 ResNet 较强。

一种典型的 Dense Block 如图 11 所示。从表面上看, DenseNet 与 ResNet 区别仅在于使用残差结构连接之前所有层。然而, 这个看似很小的修改导致了两个网络实质上不同。DenseNet 不需要重新学习冗余特征图, 这种密集连接模式相对于其他结构模型需要更少的参数, 并且改进了整个网络的信息流和梯度, 并且每个层直接访问来自损失函数和原始输入信号的梯度, 这使得训练深层网络变得更简单。此

外，密集连接具有正则化效果，减少了对训练集较小的任务的过拟合。

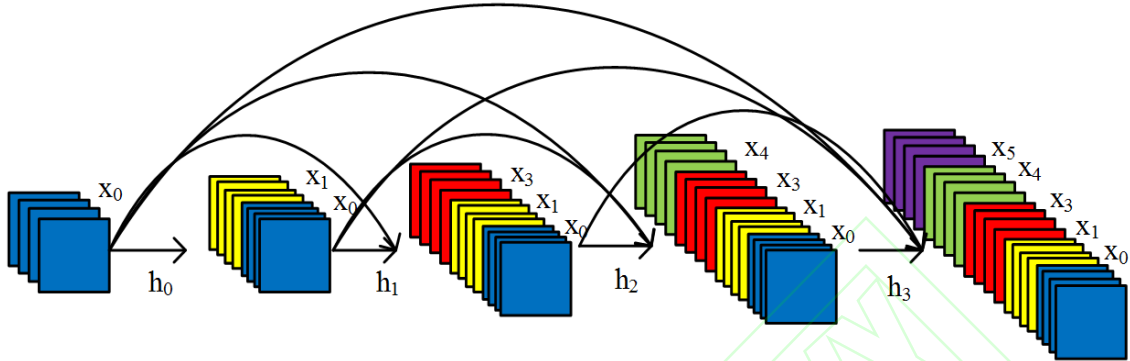


Fig.11 DenseNet block
图 11 DenseNet 模块

当残差结构提出后，Google 开始研究 Inception 网络和残差网络的性能差异以及结合的可能性，提出了 Inception-ResNet-V2^[40]，其模块结构如图 12 所示。Inception V4 模型中使用的 Inception 模块与 V3 基本一致，但是加入残差结构，可以加快训练的速度。但在实验中发现如果滤波器数量超过 1000，残差网络开始出现不稳定并且难以训练，于是在激活层之前加入残差放缩因子即形成 Activation Scaling 结构。通过控制残差放缩因子数值(0.1~0.3)进行放缩，便能够稳定训练。

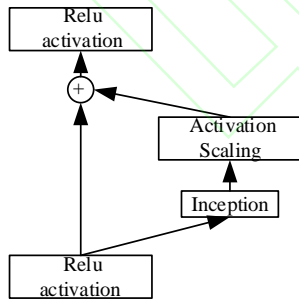


Fig.12 Inception-resnet-v2block
图 12 Inception-ResNet-V2 模块

Dongyoon Han 等人在研究中发现残差单元内下采样层对整体模型性能影响具有不确定性，为了避免下采样层对模型性能的影响，提出了 DPResNet^[41]。在整个模型结构中将特征通道数增加方式采用线性或者非线性增加，最终模型内特征通道数量形成了“金字塔”结构，如图 13 所示。通过这种结构使受下采样层影响较大的残差单元的负担均匀分布在所有单元上。

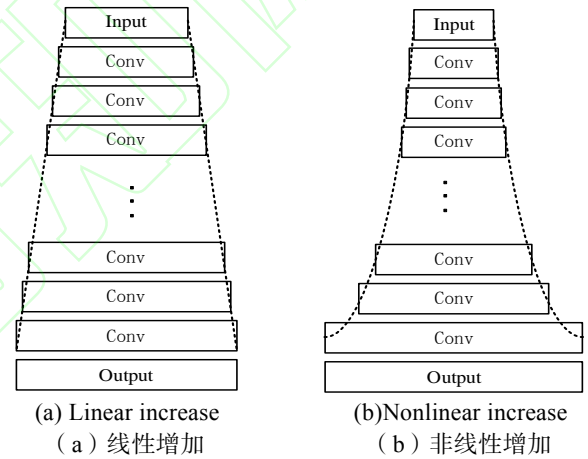


Fig.13 Channel pyramid structure generation
图 13 通道金字塔结构生成方式

由于残差单元通道数量输入和输出不一致，所以无法直接采用恒等映射，但是如果采用 1×1 卷积层对输入通道数目升维，会阻碍通道间的信息传递。DPResNet 采用如图 14 所示的残差单元结构，在残差路径上将输入特征通道后填充零矩阵通道与输出通道直接相加。

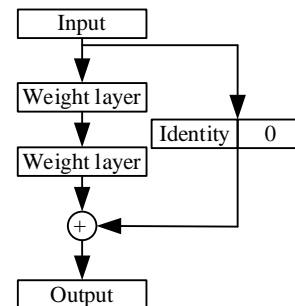


Fig.14 DPResNetblock
图 14 DPResNet 模块

除此之外,由于 ResNet 的短路连接,会出现“随机深度”现象,即只有部分残差块学到了有用信息。为此 Sergey 等人提出了深度较浅但宽度更宽的模型 WideResNet^[42]。由于宽度增加会增加计算成本,为此 Zhang 等人提出了 PloyNet^[43],从多项式角度推导以构建更复杂的 block 结构,获得比仅增大宽度与深度更大的效益。

残差结构模型通过引入短路机制形成恒等映射,有效的解决了模型的退化问题,为更深层网络模型的构建提供了解决方案。

3.4 注意力机制模型

注意力机制模型通过自动学习的方式获取到需

要重点关注的特征,抑制其他无用特征的模型结构。

通过之前介绍的模型发现,研究者的很多工作是在空间维度上提升模型的性能,而 SENet^[44]能自动获取到每个通道特征的重要程度,提升有用特征通道权重并抑制其他无用特征通道,是一种通道注意力模型。

SE 模块结构如图 15 所示。首先给定一个输入 x ,其特征通道数为 c_1 ,通过一系列变换后得到特征通道数为 c_2 的特征。与传统的 CNN 不同的是,将通过三个操作来重标定前面得到的特征。

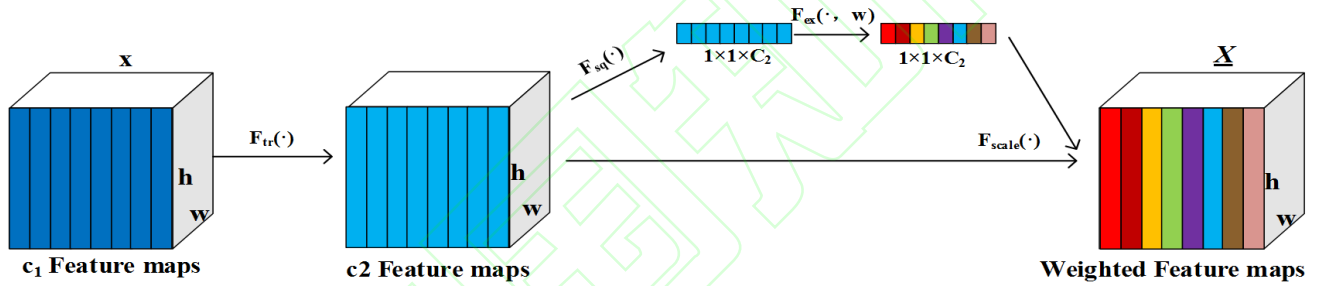


Fig.15 Squeeze and Excitation Module
图 15 Squeeze-and-Excitation 模块

首先是 Squeeze 操作,将每个二维的特征通道压缩为一个实数,这个实数某种程度上具有全局的感受野,并且输出的维度和输入的特征通道数相匹配。其次是 Excitation 操作,为每个特征通道生成权重 w ,显式地建模特征通道间的相关性。最后进行 Reweight 的操作,将权重 w 通过乘法逐通道加权到先前的特征上,完成对原始特征的重标定。

SE 模块可以嵌入到现有的网络结构中,从而提升模型对通道特征的敏感性。图 16 是将 SE 嵌入到 Inception 以及 ResNet 模块中的结构。通过在原始网络结构单元中嵌入 SE 模块,我们可以获得不同种类的 SENet,如:SE-Inception、SE-ResNet、SE-ReNeXt、SE-Inception-ResNet-v2 等,只需要增加较少的计算量便可以提升模型的性能。

标准卷积网络中,卷积层的感受野为固定大小。而 SKNet^[45]采用一种非线性的方法融合不同卷积核提取的特征,实现感受野尺寸的自动调整以便获得不同尺寸的空间信息,是一种空间注意力机制模型。

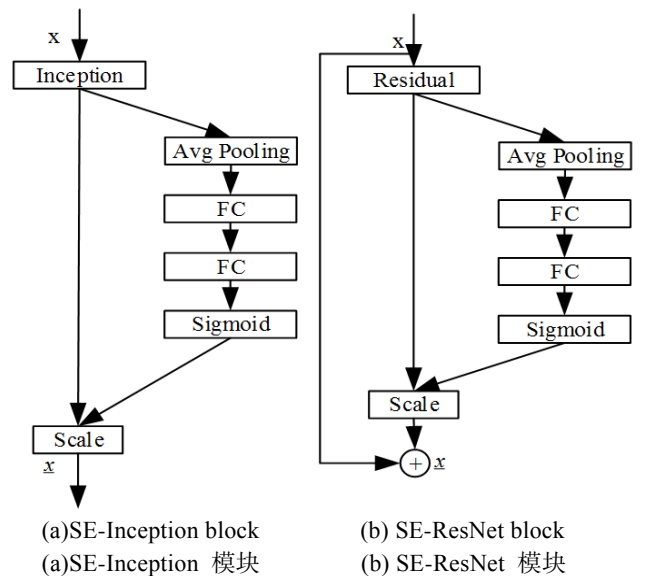


Fig.16 Two kinds of network structure unit embedded in SE module

图 16 两种嵌入 SE 模块的网络结构单元

SKNet 是由多个 SK 单元堆叠而成,SK 单元是由 1×1 卷积层,SK 卷积组成。SKNet 能够实现感

受野的自调整, 关键就在于 SK 卷积, 该卷积包含 Split, Fuse, Select 三个操作, 其结构如图 17 所示。其包含两个分支, 对于输入 $h \times w \times c$ 大小的特征图, 首先进行 Split 操作, 输入特征图分别经过两个卷积核尺寸分别为 3×3 和 5×5 的组卷积, 输出两个特征图分别为 $U_1 = h \times w \times c$, $U_2 = h' \times w' \times c'$, 并且为了进一步提升网络性能, 将 5×5 卷积分解为两个 3×3 卷积堆叠。然后进行 Fuse 操作, 形成一个门机

制控制流入下一个卷积层中不同分支的信息流。首先进行通道融合 $U = U_1 + U_2$, 然后使用 GAP 来编码全局信息 s , 然后通过 FC 进行降维并对特征权重进行调整得到紧凑特征信息 z , 最后进行 Select 操作, 通过 softmax 操作得到注意力权重 a, b , 将权重通过乘法逐通道加权到先前的特征上, 将输出特征相加得到最终的输出特征 V 。

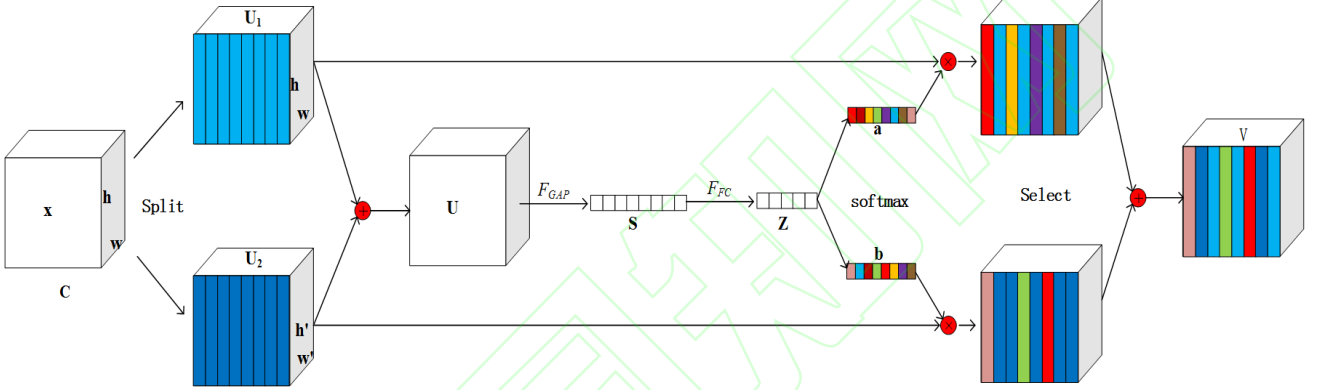


Fig.17 Selective kernel containment module
图 17 Selective Kernel Convolution 模块

SK 卷积可以嵌入到现有的网络结构中, 只需要将原网络中所有具有较大尺寸的卷积核都替换为 SK 卷积, 使网络自动选择合适的感受野大小, 成为一种泛化能力更好的注意力机制模型。

CBAM 模型^[46]结合了通道注意力机制与空间注意力机制, 相比单注意力机制模型具有更好的特征表达能力, 并且作为一种轻量级的模型, 可以无缝的集成到现在的任何 CNN 模型架构中。

注意力机制可以使得神经网络具备专注于其输入特征子集的能力, 解决了信息超载问题。

4 特殊卷积神经网络模型

4.1 挤压网络模型

SqueezeNet^[47]是最开始关注模型大小的问题, 探索模型体量和准确率之间的平衡, 而不是仅仅提高准确率。

卷积层的浮点数计算量 (floating point of operations, FLOPs) 主要是由卷积核大小 k 以及输入特征图通道数量 C_{in} 决定, 若输出特征图尺寸为

$H \times W \times C_{out}$, 那么其 $FLOPs = H \times W \times C_{in} \times k \times k \times C_{out}$ 。 C_{out} 过小即输出通道数量过少则会对特征提取不完全, 因此通过减小 C_{in} 可以保证准确率的同时有效的降低计算量。

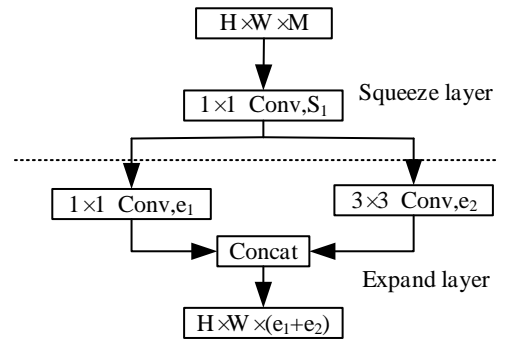


Fig.18 Fire Module
图 18 Fire 模块

SqueezeNet 是通过减少输入通道的数量形成轻量化模型的代表结构。其整个网络是通过一个个的 Fire Module 连接组合而成。Fire Module 由一个 squeeze 卷积层和一个 expand 卷积层组成, squeeze 卷积层仅仅包含 1×1 卷积核, expand 卷积层仅包

含 1×1 和 3×3 卷积核，其具体操作方式如图 18 所示。首先上一层大小为 $H \times W \times M$ 特征图会经过 Squeeze layer 通过 1×1 卷积极大减少通道数量，输出得到 S_1 个特征图 ($S_1 < M$)； $H \times W \times S_1$ 特征图到 Expand layer 分别经过 1×1 卷积层和 3×3 卷积层进行卷积，再将输出结果进行拼接，输出为 $H \times W \times (e_1 + e_2)$ 大小特征图。Fire Module 内全部采用标准卷积，所以它的模型体量小就是减少了输入通道数量，最终形成为 AlexNet 参数量 1/50 的同性能 SqueezeNet 模型。

SqueezeNet 提出后开拓了轻量级卷积神经网络模型设计这一方向，这是 SqueezeNet 的最大贡献。

4.2 生成对抗神经网络模型

生成对抗神经网络 (Generative Adversarial Network, GAN)，是一类在无监督学习中使用的神经网络。由 Ian Goodfellow^[48] 首先提出，与传统的卷积神经网络模型相比，对抗神经网络模型可以自动学习原始真实样本集的数据分布，并将输入噪声生成与真实样本类似的数据，其网络模型结构如图 19 所示。

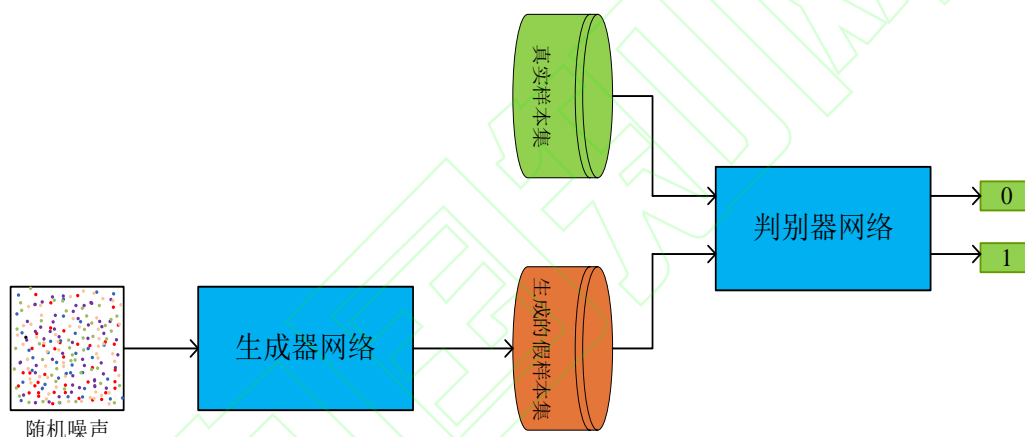


Fig.19 Model structure of Gan network
图 19 GAN 网络模型结构

生成器网络与判别器网络是两个完全独立的神经网络。对于生成器模型，输入是随机数组，通过生成网络模型生成与真实样本相似的假样本集，然后将真样本集以及其标签 1、假样本集及其标签 0 输入到判别器网络模型中。对于判别网络来说，这个问题便变为有监督的二分类问题，假如输入为真样本，网络输出就接近 1，输入为假样本，网络输出接近 0。对判别网络它希望尽量去区分样本是来自真样本集还是假样本集，生成网络与判别网络的目的正好是相反的。对于生成网络希望生成以假乱真的假样本，使判别网络无法区分真假样。这是一个生成器网络与判别器网络相互博弈的过程，所以叫做对抗网络模型。

生成对抗神经网络模型已应用于许多领域，如文本生成图像、提高图片分辨率、药物匹配、检索特定模式图片等任务。并且使得现有问题从有监督

学习过渡到无监督学习，而无监督学习才是自然界中普遍存在的，因为很多时候无法拿到监督信息。

4.3 孪生神经网络模型

孪生神经网络模型与传统的卷积神经网络模型相比主要用于判断两个输入样本的相似度。它在不需知道每个样本标签属于哪个类的情况下，能够利用给定的数据学习一个相似性的度量，并去对比与新样本的相似程度。如图 20 所示为 SiameseNet^[49] 的网络结构，它是由两个结构完全相同，并且共享权值的双胞胎子网络以及某种向量距离的度量方式构成，对于其他孪生神经网络模型中的双胞胎子网络甚至可以是一个网络。孪生网络有两个输入 X_1 与 X_2 ，它们分别通过双胞胎子网络将其转换为低维向量 $G_w(X_1)$ 与 $G_w(X_2)$ 。由于子网络具有相同的计算函数，从而能维持在原有空间的

“语义”距离，然后再通过某种距离度量方式计算两个输入向量的距离，进而判断出两个输入的相似度 E_W 。

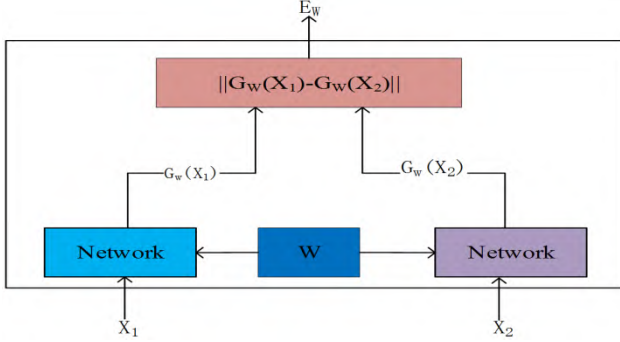


Fig.20 Structure of siamesenet network
图 20 SiameseNet 网络结构

孪生神经网络已经在人脸识别、目标追踪等任务中获得了非常成功的应用，也是一种成功的网络模型结构。

4.4 幽灵网络模型

幽灵网络模型与传统的神经网络模型相比最大的不同在于特征图的生成方式。传统的神经网络模型的特征图是通过卷积层生成所有的特征图，而幽灵网络模型中相似特征图是通过其中一张特征图线性操作得到的。研究者在 ResNet-50 中使用可视化技术发现在一个训练良好的卷积神经网络模型的特征图中，通常会包含丰富甚至冗余的信息。华为公司的 Ghost Net^[50]采用 Ghost 模块将一个特征图通过线性运算后生成其他冗余特征图，可以用更少的参数生成更多更丰富的特征。

Ghost Module 由两部分组成，如图 21 所示。第一部分涉及普通卷积，Ghost 模块并没有使用深度可分离卷积等操作，因为逐点卷积层占用大量内存和 FLOPs 的问题没有有效的解决方式。虽然采用了普通卷积但卷积核总数受到严格控制。输入的特征图首先通过普通卷积操作生成小数量的 m 个固有特征图，此时的空间大小信息已经固定。第二部分分为特征映射， m 个固有特征图经线性运算来生成更多的特征映射为 n ，具体公式为：

$$y_{ij} = \phi_{i,j}(y'_i), \forall i = 1, 2, \dots, s(4)$$

其中， y'_i 表示固有特征图中的第 i 个特征图，

$\phi_{i,j}$ 表示对第 i 个特征图进行的第 j 个线性变换。某一层特征图可能对应多个线性变换，得到多个变换结果， y_{ij} 为输出特征图。在变换的最后，为保留固有特征中的信息增加恒等映射，也就是将固有的特征图拼接到变换后的特征图，便生成 n 维特征图。

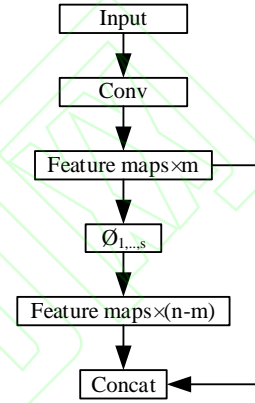


Fig.21 Ghost Module
图 21 Ghost 模块

GhostNet 使用 MobileNet V3 的基本架构，利用 Ghost block 代替了 MobileNet V3 中的 block，并且加入轻量级注意力模块 SE 模块，最终 GhostNet 通过冗余特征重生技术改进的模型，在各种计算复杂度级别上准确率以及实际推理速度均优于目前其他轻量级网络模型。

5 卷积神经网络的典型应用

卷积神经网络在图像分类领域取得一系列突破，各个应用领域将不同 CNN 作为特征提取的主干网络并添加不同的功能单元结构形成了新的应用模型，将 CNN 融入到了不同的深度学习任务中，由于其出色检测效果及性能逐渐取代传统方法目前已成为目标检测、语义分割、自然语言处理等领域的研究热点。

5.1 目标检测

目标检测（Object Detection）的主要任务是从图像中准确定位和分类特定对象，并给出每个目标的位置和大小，是计算机视觉领域的核心问题之一。

早在 1994 年 CNN 就成功应用在目标检测中，但是与传统目标检测方法相比在性能上没有太大

优势。直到进入到深度学习时代,将 CNN 作为特征提取网络较好应用到目标检测任务上提出了 R-CNN 模型,其性能远超传统目标检测方法。此后提出了一系列优秀的目标检测模型,基于 CNN 的目标检测方法成为主流。基于 CNN 的目标检测方法与传统检测算法步骤类似,分为三个阶段,首先生成候选区域,然后提取待分类候选区域特征,最后对提取到的特征进行分类。其主要包括两种类型,一种是基于锚点的目标检测,另一种是基于 Anchor-free 的目标检测。其中基于锚点的目标检测又分基于候选区域网络的检测模型和基于回归方法的检测模型。

R-CNN^[51]是最初基于候选区域的检测模型,其在每个阶段都采用了不同的方法,在模型训练和应用上都有很多弊端。在 Girshick 等人^[52]提出改进的 Fast R-CNN,将分类器与边界回归器加入到 CNN 网络中,从而整个过程只对图像进行了一次卷积操作,利用卷积网络实现了后两个阶段的统一;Ren 等人^[53]提出的 Faster R-CNN,在 Fast R-CNN 基础上添加 CNN 作为区域生成网络(RPN)来生成检测框以取代选择性搜索方法,使得区域生成、对象分类、边界回归共享卷积特征,加快了目标检测的速度,实现目标检测三个阶段的统一;Jifeng Dai 等人^[54]提出基于区域的全卷积目标检测网络(R-FCN),在 CNN 特征层后加入位置敏感的卷积层替换全连接层,解决了全连接层丢失目标精确位置信息的问题;Spyros Gidaris 等人^[55]提出基于多区域的 CNN 识别模型 MR-CNN,通过提取多区域的特征来丰富目标的特征,达到提高目标检测效果。基于候选区域的检测模型目标检测准确率,非常依赖候选区提取方法的准确度,为此提出了基于回归方法的检测模型。

另一种基于回归方法的检测模型,最初是由 Pierre Sermanet 等人提出 OverFeat 模型^[56],将 CNN 作为回归器,把整副图像看成一个候选区域输入到 CNN 中实现目标检测。其结构相比基于候选区域的检测模型网络结构更加简单、实时性更高。随后 Yolo 系列和 SSD 网络都是典型的基于回归方法的

检测模型。Yolo 最初的提出就是针对基于候选区域网络模型检测速度问题提出,Yolo^[57]仅采用一个神经网络在输出层回归边界框位置以及分类,虽然目标检测精度略有下降,但是提高了目标检测的实时性,Fast Yolo 更是达到了 155 帧/s。随后提出的 SSD^[58]通过融合六个不同尺度的特征层可以对不同尺度的物体进行检测,并且采用非极大抑制方法以及难分样本挖掘技术处理大量产生的负样本,SSD 采用单神经网络多边界框检测的方式,直接完成了候选区域选取和分类两个步骤,其达到了较高的检测精度以及速度,目前 SSD 算法已经被移植到手机移动端并流畅运行。Yolo v2^[59]针对 Yolo 回归方法召回率以及检测率不高的问题,采用批正则化、维度聚类、直接位置预测、多尺度训练等方法进行改进,相较于 SSD 以及 Faster R-CNN 性能更胜一筹。与此同时,提出的 Yolo 9000 在 ImageNet 和 COCO 数据集采用联合训练和检测的方法,使 Yolo9000 能对超过 9000 类目标进行实时检测。DSSD^[60]采用 Top-Down 结构,将浅层特征图输入到反卷积模型中和深层的特征图在对应的通道上做乘法运算,将高层的语义信息融入到低层的特征信息中,以此来提高特征精度。YOLO v3^[61]为了实现多尺度预测参考特征金字塔网络 FPN 在 3 个尺度上进行检测,提高了小物体检测效果。以前的目标检测算法大都是在 ImageNet 数据集进行预训练,然后针对特定任务进行微调,导致模型结构灵活性差,且难以改变网络结构,并且计算量也大,限制其应用场合。Zhiqiang Shen 等人^[62]首次提出了一种能从零开始训练的检测器 DSOD,并且采用了 Densenet 结构思想,相比 SSD 模型准确率以及模型大小方面更胜一筹。Yolo v4 通过改进主干网络,并且添加 SPP 模块提升模型的感受野,采用 PANet^[63]特征融合方法代替 FPN,Yolo v4^[64]总得来说对近年来所有比较好的目标检测方法总结的并改进基础上取各家之所长形成,其目前与其他 SOTA 目标检测器相比在速度和准确性上都优于其他检测器。

虽然基于 Anchor 的目标检测模型在性能上有了很大的提升,但是由于预设的锚点大多是负样本,

在训练过程中会导致正样本和负样本的失衡, 锚框的超参数巨大并且锚框的宽高比对数据集比较敏感, 会影响其检测性能。基于 Anchor-free 的目标检测模型消除了对锚框的需求。Hei Law 等人^[65]提出了 CornerNet, 使用 CNN 生成两组热图分别来检测物体的左上角点和右下角点, 并判断两个角点是否属于同一目标, 并对角点位置偏差进行修正, CornerNet 是 Anchor-free 目标检测的开创性工作。ExtremeNet^[66]改进 CornerNet 将目标检测改为检测目标边框的 4 个极值点; Duan 等人^[67]提出了基于中心点的 CenterNet, 将目标检测改变为关键点估计问题, 在中心点基础上回归出目标的其他属性, 如尺寸、方向、姿态等。FSAF 模型^[68]在 FPN 每层后添加 Anchor-free 分支去计算该目标在不同的特征分支上的得分, 将该目标分配给得分最高的分支上进行检测。FCOS 模型^[69]是一个基于全卷积的检测网络, 通过“center-ness”来预测在特征图中每个像素点相对框中心点的偏差, 在后处理中仅使用 NMS 使其检测效果比肩基于 Anchor 的目标检测算法。

基于 CNN 的目标检测在各个领域均得到了广泛应用。Vailant 等人^[70]首次提出将 CNN 应用于人脸检测。Sermanet 等人^[71]提出了将 CNN 应用于行人检测。Li 等人^[72]采用传统领域知识和深度学习优势的融合实现了车辆车色识别任务。P. Sermanet, Y. LeCun 使用多尺度卷积神经网络识别交通标志^[73]。Linda 等人^[74]实现基于 CNN 的新冠肺炎 (COVID-19) 检测。Huang 等人^[75]改进 R-FCN 实现船舶识别。Luo 等人^[76]改进 YOLO 实现火灾检测报警。Ye 等人^[77]改进 SSD 适用于自动驾驶场景。

5.2 语义分割

语义分割是从像素级别上理解图像, 通过对图像中的每个像素点进行分类, 将图像分割成若干个具有特定语义类别的区域, 也是计算机视觉领域的核心问题之一。深度学习兴起后, 基于 CNN 的语义分割方法由于其效率以及准确率远超传统方法而成为主流, 主要有基于解码器方法和基于多特征融合方法。

语义分割取得跨越式进步是 2014 年 Jonathan

Long 等人^[78]提出的基于全卷积神经网络 (FCN), 真正开启像素级语义分割网络。FCN 是在 CNN 分类网络的基础上改进, 将 CNN 中的全连接层替换为卷积层, 从而建立全卷积网络。输入任意尺寸的图像后, 从而产生对每个像素的分类结果, 这部分称为编码器。将分类结果再通过反卷积操作映射到原图像大小, 产生像素级别的标签, 从而获得语义分割的结果, 这部分称为解码器。FCN 提出了的基于 CNN 的编码器-解码器体系结构, 证明了在可变大小的图像上以端到端的方式训练网络完成语义分割, 并在其基础上不断改进得到了性能更好的语义分割网络。Badrinarayanan 等人^[79]提出 SegNet, 该网络在编码器池化时保留池化层索引, 在解码器中根据索引准确将该值映射到初始位置, 该改进不需要额外学习就改善了图像分割的效果。为了在浅层捕捉充足的空间信息帮助模型恢复目标细节, DeepLabv3+^[80]在网络中添加一个简单有效的解码模块。Ronnebergeretal 等人^[81]提出了用于分割生物显微镜图像的 U-Net; Milletari 等人^[82]提出用于三维医学图像分割模型 V-Net。

为了充分利用目标的上下文信息, 进一步优化分割效果, 需要对 CNN 中不同层次的信息进行融合。Liu 等人^[83]提出了 ParseNet 模型, 通过 GAP 提取图像的全局特征与局部特征融合, 将全局信息融合到 FCN 中。Li 等人^[84]提出 RefineNet, 通过残差连接利用了下采样过程中的信息, 使前层的特征可以直接强化高层的语义特征。DeepLabv2^[85]和 PSPNet^[86]采用空洞空间金字塔池化模块有效的捕获上下文信息, 提高模型获取全局信息的能力。OCNet^[87]在每个尺度上使用金字塔语义模块捕获目标像素的上下文信息。SPNet^[88]通过引入一个新的条带池模块, 使骨干网络能有效的利用远程的上下文信息; CPNet^[89]设计上下文优先的网络; 此外, 在 GCN (global convolutional network)^[90]、PAN (pyramid attention network)^[91]、DFN (discriminative feature network)^[92]等模型中应用可以逐步融合骨干网络中不同层次的特征图的 U 型结构。

随着注意力机制在 CNN 分类中的成功应用,

其在帮助语义分割中生成高质量的分割结果越来越重要。Wu 等人^[93]提出 DANet 利用注意力机制对全局上下文信息计算权重向量,利用权重向量细化局部及全局上下文信息。RANet^[94]将注意力机制融合到前向过程中,使局部信息得到增强。Yude Wang 等人^[95]提出一种自监督等变注意力机制模型 (SEAM)。Sungha Choi 等人^[96]提出高度驱动的注意力网络 (HANet) 改善城市场景的语义分割。

基于 CNN 的语义分割在各个场景得到了广泛应用。Lee 等人^[97]对多种场景下的车道线分割。Diakogiannis 等人^[98]将提出基于 CNN 的遥感图像分割框架 ResUNet-a。Tran 等人^[99]SegNet 结构对显微血液涂片图像中的红细胞和白细胞进行分割。Edupuganti 等人^[100]利用 FCN 对眼底图像中视盘和杯状区域进行分割,用以辅助眼部疾病的诊断。Liu 等人^[101]改进 UNet 对脑海马体 3D 分割。Wang 等人^[102]改进 DeepLab 对肠息肉完成分割诊断。Wang 等人^[103]对 FCN 改进完成玉米叶片病斑的分割。“腾讯觅影”在新冠肺炎 (COVID-19) 能 1 分钟内对 CT 图像进行自动分割,帮助医生准确评判患者肺炎的严重程度。

5.3 自然语言处理

CNN 最开始是用于计算机视觉领域中,由于其优越的特征提取能力现在也被广泛用于自然语言处理 (Natural Language Processing, NLP) 中。研究者基于 CNN 设计各种模型,在 NLP 领域的研究取得了一个又一个突破。

CNN 应用到 NLP 中最早可以追溯到 2008 年 Collobert 等人^[104]将词向量方法与 CNN 相结合,为目前许多方法的改进和提升奠定了基础。在 2011 年 Collobert 等人^[105]进一步扩展了他们的研究工作,提出用于各种 NLP 任务的通用框架。2014 年 Kim 等人^[106]提出基于预训练 Word2vec 的 TextCNN 模型用于句子的分类任务。2014 年 Kalchbrenner 等人^[107]提出用于语句建模的动态卷积网络 DCNN,其采用动态挑选最大特征的池化策略,在保留了特征顺序的同时也有利于累计句子中的关键信息,解决 CNN 中文本的长距离依赖关系的问题,完成了情感分类

任务。DPCNN^[108]通过增加网络的深度来捕捉文本的长距离依赖关系,提高了用于文本分类的准确度。Siwei Lai 等人^[109]提出了同时利用了 RNN 和 CNN 的优点的 RCNN 模型用于文本分类。

除此之外,ByteNet^[110]应用于具有挑战性的机器翻译任务;SeqGan^[111]完成文本生成任务。Yan 等人^[112]提出深度文本生成模型 GradeGAN 对音乐生成评价信息。Li 等人^[113]对评论数据提出标签方面情感联合模型。Tu 等人^[114]无池化层的中文分词模型 PCNN。并且 Facebook 的 Pytext 以及百度的 PaddleNLP 的自然语言处理框架均已成熟应用。

6 总结与展望

随着卷积神经网络相关研究的不断推进,其性能以及模型复杂度也随之提高。本文对目前性能优秀的典型卷积神经网络模型结构进行了分析。典型的卷积神经网络模型已经取得了显著的成果,在图像分类以及识别中已经有很高的准确率,主要有增加网络结构的宽度以及深度、融合通道域和空间域注意力机制等关键技术。虽然传统卷积神经网络模型识别率已经很高,但是通过向原图片中加入一些特定的噪点,这种人工扰动可以轻易使神经网络模型对图像进行错误分类,如何解决这一问题并提高模型的泛化能力,是亟待解决的问题。并且随着神经网络深度以及宽度的加深,训练成本逐渐增加,若能够将针对特定问题的先验知识加入到模型的构建中,将能够大大加快模型训练速度。除此之外,卷积神经网络的结构研究还具有很大空间,模型性能的提升需要更加合理网络结构设计,并且网络模型的超参数的设置依赖实验和经验,对参数的量化分析是卷积神经网络一个有待解决的问题。虽然卷积神经网络正处于研究热度非常高的阶段,但是关于其完备的数学解释与证明目前仍然较为匮乏,对其相关理论研究对卷积神经网络进一步发展,以及解决当前网络结构的缺陷问题都具有重要意义。

本文列举的特殊模型也为传统卷积神经网络模型提出了更多设计思路:

(1) 典型卷积神经网络模型需轻量化设计。

在过去卷积神经网络研究重点在于算法设计, 对于其模型应用的具体部署平台欠缺考虑, 设计硬件友好型模型结构将有助于进一步提高模型性能, 也是模型结构设计的重要研究方向。

(2) 加强弱监督或无监督学习的卷积模型结构研究; 自然界中无监督学习是更加普遍的存在, 并且符合人脑的思维模式, 虽然无监督与弱监督在图像识别方面已取得了一定的进展, 但是, 与监督学习相比, 非监督学习与半监督学习在图像识别精确度方面依然差距较大。

(3) 构建多输入的卷积神经网络模型结构。多信息输入能更充分利用原始数据中隐含的特征表达, 以更少的训练代价获得更好的识别效果。并且在识别过程中共享网络结构, 能进一步加快识别效率。

(4) 研究效率更高的特征生成方式。高效的冗余特征生成方式能减少模型的参数量, 生成更丰富的特征图, 也是值得研究的问题。

目前深度神经网络模型研究仍处于起步阶段, 许多问题尚未得到解决。然而这仍然是一个极具研究价值的领域。无论是典型卷积神经网络还是特殊卷积神经网络模型, 目前在智能安防、虚拟现实、智能医疗、自动驾驶、可穿戴设备以及移动支付等领域中均有应用。深度神经网络模型发展对引领未来科技发展, 人工智能产业的发展具有关键作用。

References:

- [1] Sun B, Ju Q Q, Sang Q B. Image dehazing algorithm based on FC-DenseNet and WGAN[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(8): 1380-1388.
- [2] Sampedro C, RodriguezVazquez J, RodriguezRamos A, et al. Deep learning-based system for automatic recognition and diagnosis of electrical insulator strings[J]. IEEE Access, 2019, 7: 101283-101308.
- [3] Harikrishnan J, Sudarsan A, Sadashiv A, et al. Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision[C]//2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN). IEEE, 2019: 1-5.
- [4] Li D J, Li D G, Yang L. Application of Convolutional Neural Network in Dynamic Gesture Tracking[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(5): 841-847.
- [5] Lin K Z, Bai J X, Li H T, et al. Facial Expression Recognition with Small Samples Fused with Different Models Under Deep Learning[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(3): 482-492.
- [6] Wang ZY, Yuan C, Li JC. Instance Segmentation with Separable Convolutions and Multi-level Features[J]. Journal of Software, 2019, 30(4): 954-961.
- [7] Xiao L, Chen BL, Huang X, Liu HF, Jing LP, Yu J. Multi-label Text Classification Method Based on Label Semantic Information[J]. Journal of Software, 2020, 31(4): 1079-1089.
- [8] Li D M, Tan W. Research on Named Entity Recognition Method in Plant Attribute Text[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(12): 2085-2093.
- [9] Jaf S, Calder C. Deep learning for natural language parsing[J]. IEEE Access, 2019, 7: 131363-131373.
- [10] Roberts L G. Machine perception of three-dimensional solids[D]. Massachusetts Institute of Technology, 1963.
- [11] Pellegrino F A, Vanzella W, Torre V. Edge detection revisited[J]. IEEE Trans Syst Man Cybern B Cybern, 2004, 34(3): 1500-1518.
- [12] Marr D, Hildreth E. Theory of Edge Detection[J]. Proceedings of the Royal Society B: Biological Sciences, 1980, 207(1167): 187-217.
- [13] Canny J. A computational approach to edge detection[J]. IEEE Transactions on pattern analysis and machine intelligence, 1986 (6): 679-698.
- [14] Harris C G, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15(50): 10-5244.
- [15] Lowe D G. Object recognition from local scaleinvariant features[C]//Proceedings of the seventh IEEE international conference on computer vision. IEEE, 1999, 2: 1150-1157.
- [16] Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 32(1): 105-119.
- [17] Lowe D G. Object recognition from local scale-invariant features[C]//Proceedings of the seventh IEEE international conference on computer vision. Ieee, 1999, 2: 1150-1157.
- [18] Cover T, Hart P. Nearest neighbor pattern classification[J]. IEEE transactions on information theory, 1967, 13(1): 21-27.
- [19] Zhao Q, Principe J C. Support vector machines for SAR automatic target recognition[J]. IEEE Transactions on Aerospace and Electronic Systems, 2001, 37(2): 1-11.

- 643-654.
- [20] Thiagarajan J J, Ramamurthy K N, Knee P, et al. Sparse representations for automatic target classification in SAR images[C]//International Symposium on Communications. IEEE, 2010: 1-4.
 - [21] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.
 - [22] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological Review, 65(6):386-408.
 - [23] Jan M. Book Review: Perceptrons, An Introduction to Computational Geometry[J]. Bull. amer. math. soc, 1972, 78(1): 12-16.
 - [24] Rumelhart D E , Hinton G E , Williams R J . Learning Representations by Back Propagating Errors[J]. Nature, 1986, 323(6088):533-536.
 - [25] Hochreiter S. Untersuchungen zu dynamischen neuronalen Netzen[J]. Diploma, Technische Universität München, 1991, 91(1).
 - [26] Hinton G E , Osindero S , Teh Y W . A fast learning algorithm for deep belief net[J]. Neural Computxtion, 2006, 18(7):1527-1554.
 - [27] 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.
 - [28] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
 - [29] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Springer, Cham, 2014: 818-833.
 - [30] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large Scale Image Recognition[J]. Computer ence, 2014.
 - [31] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.
 - [32] Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[J]. 2019: 6105-6114.
 - [33] Lin M , Chen Q, Yan S . Network in Network[C]// International Conference on Learning Representations, 2014:1-10.
 - [34] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1-9.
 - [35] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
 - [36] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural computation, 1997, 9(8):1735-1780.
 - [37] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
 - [38] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
 - [39] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
 - [40] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-first AAAI conference on artificial intelligence. 2017.
 - [41] Yamada Y, Iwamura M, Kise K. Deep pyramidal residual networks with separated stochastic depth[J]. arXiv preprint arXiv:1612.01230, 2016.
 - [42] Zagoruyko S, Komodakis N. Wide residual networks[J]. arXiv preprint arXiv:1605.07146, 2016.
 - [43] Zhang X, Li Z, Change Loy C, et al. Polynet: A pursuit of structural diversity in very deep networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 718-726.
 - [44] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 7132-7141.
 - [45] Li X, Wang W, Hu X, et al. Selective kernel networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2019: 510-519.
 - [46] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV), 2018: 3-19.
 - [47] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.
 - [48] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 2014 Advances in Neural Information Processing Systems. BerlinSpringer, 2014: 2672-2680.
 - [49] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks[J]. computer vision and pattern recognition, 2015: 4353-4361.
 - [50] Han K, Wang Y, Tian Q, et al. GhostNet: More features

- from cheapoperations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.
- [51] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [52] Girshick R. Fast R-CNN[C]//Proc of IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [53] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [54] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems, 2016: 379-387.
- [55] Gidaris S, Komodakis N. Object detection via a multi-region and semantic segmentation aware cnn model[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1134-1142.
- [56] Sermanet P, Eigen D, Zhang X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. *Eprint Arxiv*, 2013.
- [57] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [58] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [59] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7263-7271.
- [60] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector[J]. *arXiv preprint arXiv:1701.06659*, 2017.
- [61] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. *arXiv preprint arXiv:1804.02767*, 2018.
- [62] Shen Z, Liu Z, Li J, et al. DSOD: Learning Deeply Supervised Object Detectors from Scratch[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Computer Society, 2017: 1919-1927.
- [63] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
- [64] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. *arXiv preprint arXiv:2004.10934*, 2020.
- [65] Law H, Heng J. Cornernet: Detecting objects as paired keypoints [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734-750.
- [66] Zhou X, Zhou J, Krahenbuhl P. Bottom up object detection by grouping extreme and center points[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 850-859.
- [67] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 6569-6578.
- [68] Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 840-849.
- [69] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE international conference on computer vision, 2019: 9627-9636.
- [70] Vaillant R, Monroq C, Le Cun Y. Original approach for the localisation of objects in images[J]. *Vision, Image and Signal Processing*, 1994, 141 (4) : 245-250.
- [71] Sermanet P, Kavukcuoglu K, Chintala S, et al. Pedestrian detection with unsupervised multi-stage feature learning[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3626-3633.
- [72] Hu C, Bai X, Qi L, et al. Vehicle color recognition with spatial pyramid deep learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16 (5): 2925-2934.
- [73] Sermanet P, Lecun Y. Traffic sign recognition with multiscale convolutional networks[C]//International Joint Conference on Neural Networks, 2011: 2809-2813.
- [74] Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images[J]. *arXiv preprint arXiv:2003.09871*, 2020.
- [75] Hang Z J, Sang Q B. Ship Detection Based on Improved R-FCN[J]. *Journal of Frontiers of Computer Science and Technology*, 2020, 14(6): 1045-1053.
- [76] Luo X Q, Pan S L. Improved YOLOV3 fire detection method[J]. *Computer Engineering and Applications*, 2020, 56(17): 187-196.
- [77] Ye Z Y, Zheng J L. Object Detection Algorithm DFSSD Based on Automatic Driving Scene[J]. *Computer Engineering and Applications*, 2020, 56(16): 139-147.
- [78] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [79] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A

- Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [80] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Proceedings of the European conference on computer vision (ECCV)*. 2018: 801-818.
- [81] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]// *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015: 234-241.
- [82] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//*2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016: 565-571.
- [83] Liu W, Rabinovich A, Berg A C. Parsenet: Looking wider to see better[J]. *arXiv preprint arXiv:1506.04579*, 2015.
- [84] Lin G, Milan A, Shen C, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017: 1925-1934.
- [85] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 40(4): 834-848.
- [86] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 2881-2890.
- [87] Yuan Y, Wang J. Ocnet: Object context network for scene parsing[J]. *arXiv preprint arXiv:1809.00916*, 2018.
- [88] Al Arif S M M R, Knapp K, Slabaugh G. SPNet: Shape prediction using a fully convolutional neural network[C]//*International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer, Cham, 2018: 430-439.
- [89] Yu C, Wang J, Gao C, et al. Context Prior for Scene Segmentation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 12416-12425.
- [90] Peng C, Zhang X, Yu G, et al. Large kernel matters--improve semantic segmentation by global convolutional network[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017: 4353-4361.
- [91] Li H, Xiong P, An J, et al. Pyramid attention network for semantic segmentation[J]. *arXiv preprint arXiv:1805.10180*, 2018.
- [92] Yu C, Wang J, Peng C, et al. Learning a discriminative feature network for semantic segmentation[C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 1857-1866.
- [93] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019: 3146-3154.
- [94] Yang L, Han Y, Chen X, et al. Resolution Adaptive Networks for Efficient Inference[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 2369-2378.
- [95] Wang Y, Zhang J, Kan M, et al. Self-supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation[C]//*Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition*. 2020: 12275-12284.
- [96] Choi S, Kim J T, Choo J. Cars Can't Fly Up in the Sky: Improving Urban-Scene Segmentation via Height Driven Attention Networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9373-9383.
- [97] Lee S, Kim J, Shin Yoon J, et al. Vpgnet: Vanishing point guided network for lane and road marking detection and recognition[C]//*Proceedings of the IEEE international conference on computer vision*, 2017: 1947-1955.
- [98] Diakogiannis F I, Waldner F, Caccetta P, et al. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114.
- [99] Tran T, Kwon O H, Kwon K R, et al. Blood cell images segmentation using deep learning semantic segmentation[C]//*2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)*. IEEE, 2018: 13-16.
- [100] Edupuganti V G, Chawla A, Kale A. Automatic optic disk and cup segmentation of fundus images using deep learning[C]//*2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018: 2227-2231.
- [101] Liu C, Xiao Z Y, Wu X X. Application of ThreeDimensional Convolution Network in Brain Hippocampus Segmentation[J]. *Journal of Frontiers of Computer Science and Technology*, 2020, 14(3): 493-501.
- [102] Wang Y G, Xi Y Y, Pan X Y. Method for Intestinal Polyp Segmentation by Improving DeepLabv3+ Network[J]. *Journal of Frontiers of Computer Science and Technology*, 2020, 14(7): 1243-1250.
- [103] Wang Z, Shi Y, Li Y B. Segmentation of Corn Leaf Diseases Based on Improved Fully Convolutional Neu-

- ral Network[J]. Computer Engineering and Applications, 2019, 55(22): 127-132.
- [104] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multi-task learning[C]//Proceedings of the 25th international conference on Machine learning. 2008: 160-167.
- [105] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. Journal of machine learning research, 2011, 12(Aug): 2493-2537.
- [106] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. ACM, 2014:1746-1751.
- [107] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [108] Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 562-570.
- [109] Wang R, Li Z, Cao J, et al. Convolutional recurrent neural networks for text classification[C]//2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-6.
- [110] Kalchbrenner N, Espeholt L, Simonyan K, et al. Neural machine translation in linear time[J]. arXiv preprint arXiv:1610.10099, 2016.
- [111] Yu L, Zhang W, Wang J, et al. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient[C]//Thirty First AAAI Conference on Artificial Intelligence, 2017.
- [112] Yan D, He J, Liu H Y. Considering Grade Information for Music Comment Text Automatic Generation[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(8): 1389-1396
- [113] Cao Y, Li T R, Jia Z, et al. BGRU: New Method of Chinese Text Sentiment Analysis[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(6): 973-981.
- [114] Tu W B, Yuan Z M, Yu K. Convolutional Neural Networks Without Pooling Layer for Chinese Word Segmentation[J]. Computer Engineering and Applications, 2020, 56(2): 120-126.

附中文参考文献:

- [1] 孙斌, 睢青青, 桑庆兵. 结合 FC-DenseNet 和 WGAN 的图像去雾算法[J]. 计算机科学与探索, 2020, 14(8): 1380-1388.
- [4] 李东洁, 李东阁, 杨柳. 卷积神经网络在动态手势跟踪中的应用[J]. 计算机科学与探索, 2020, 14(5): 841-847.
- [5] 林克正, 白婧轩, 李昊天, 等. 深度学习下融合不同模型的小样本表情识别[J]. 计算机科学与探索, 2020, 14(3): 482-492.
- [6] 王子愉, 袁春, 黎健成. 利用可分离卷积和多级特征的实例分割[J]. 软件学报, 2019, 30(4): 954-961.
- [7] 肖琳, 陈博理, 黄鑫, 等. 基于标签语义注意力的多标签文本分类[J]. 软件学报, 2020, 31(4): 1079-1089.
- [8] 李冬梅, 檀稳. 植物属性文本的命名实体识别方法研究[J]. 计算机科学与探索, 2019, 13(12): 2085-2093.
- [75] 黄致君, 桑庆兵. 改进 R-FCN 的船舶识别方法[J]. 计算机科学与探索, 2020, 14(6): 1045-1053.
- [76] 罗小权, 潘善亮. 改进 YOLOV3 的火灾检测方法[J]. 计算机工程与应用, 2020, 56(17): 187-196.
- [77] 叶召元, 郑建立. 基于自动驾驶场景的目标检测算法 DFSSD[J]. 计算机工程与应用, 2020, 56(16): 139-147.
- [101] 刘辰, 肖志勇, 吴鑫鑫. 三维卷积网络在脑海马体分割中的应用[J]. 计算机科学与探索, 2020, 14(3): 493-501.
- [102] 王亚刚, 郝怡媛, 潘晓英. 改进 DeepLabv3+ 网络的肠道息肉分割方法[J]. 计算机科学与探索, 2020, 14(7): 1243-1250.
- [103] 王振, 师韵, 李玉彬. 基于改进全卷积神经网络的玉米叶片病斑分割[J]. 计算机工程与应用, 2019, 55(22): 127-132.
- [112] 严丹, 何军, 刘红岩, 等. 考虑评级信息的音乐评论文本自动生成[J]. 计算机科学与探索, 2020, 14(8): 1389-1396.
- [113] 曹宇, 李天瑞, 贾真, 等. BGRU: 中文文本情感分析的新方法[J]. 计算机科学与探索, 2019, 13(6): 973-981.
- [114] 涂文博, 袁贞明, 俞凯. 无池化层卷积神经网络的中文分词方法[J]. 计算机工程与应用, 2020, 56(2): 120-126.



YAN Chunman was born in Huating, Gansu Province, China. He received the Ph. D. degree in Circuits and System, with the Institute of Intelligent Control & Image Engineering, Xidian University, Xi'an, China. He is an associate professor and master's supervisor at Northwest Normal University. His research interests include pattern recognition, artificial intelligence, etc.

严春满 (1970-), 男, 甘肃省华亭市人, 2012 年于西安电子科技大学获博士学位, 目前为西北师范大学副教授, 硕士生导师, 主要研究领域为模式识别、人工智能。发表 EI 检索论文 10 余篇, 主持国家自然科学基金项目 1 项, 参与国家自然科学基金项目 2 项。



WANG Cheng was born in 1997. He is a M.S. candidate at Northwest Normal University. His research interests include pattern recognition, artificial intelligence, Embedded system, etc.

王铖 (1997-), 男, 山东省淄博市人, 西北师范大学硕士研究生, 主要研究领域为模式识别, 人工智能, 嵌入式系统等。