

卷积神经网络模型在图像识别中的应用综述

王煜瑄¹, 李振一², 曲亚川¹

(1. 青岛理工大学机械与汽车工程学院, 山东 青岛 266520;

2. 青岛理工大学信息与控制工程学院, 山东 青岛 266520)

摘要 传统的图像识别技术以浅层结构网络为主, 通过计算机对捕获的图片进行处理, 比如人脸识别、商品条形码识别及手写识别等。卷积神经网络模型依据其分类的准确度高、自适应能力强等特点被广泛应用于图像识别领域, 但是卷积神经网络模型也有数据采集量需求大、计算量大等问题, 这些问题的存在使得卷积神经网络无法满足高智能的自动化领域的发展, 如医疗研究、航空航天等领域。本文针对卷积神经网络的典型模型进行优缺点分析并阐明了各种模型之间的关联性。同时, 整理了卷积神经网络在图像识别领域的理论研究案例。最后, 总结卷积神经网络在图像处理领域的发展情况以及当前CNN模型发展面临的难题以及解决方案。

关键词 卷积神经网络模型; 图像识别; 自适应能力

中图分类号: TP317.4

文献标识码: A

文章编号: 1007-0745(2023)08-0008-04

人工智能是能够自主感知周围环境并做出响应行为来达到某一目标的智能体, 其中深度学习作为人工智能的一个分支, 以人工神经网络为架构, 对数据中的高级抽象进行表征学习。随着深度学习技术的发展, 在自然语言处理、计算机视觉、机器学习、卷积神经网络、深度神经网络、蒙特卡洛、语言处理、自动驾驶等领域都得到了广泛应用。

人类对世界的认知多数来自视觉。在计算机视觉训练的过程中用到的一种关键技术是卷积神经网络模型(Convolution Neural Networks, CNN)。

深度学习在图像识别方面的出色表现通常是通过神经网络与卷积神经网络模型来实现的, 神经网络虽然对不同的数据都具有很轻的非线性拟合能力, 但是在语音、图像上迟迟得不到突破。为了发展高精度、高可靠度的图像识别技术, 需要对现有神经网络模型进行分类总结与改善。本文主要对基于卷积神经网络的图像识别算法进行回顾和总结: 第1节, 主要针对图像识别模型的定义和研究必要性进行讨论, 介绍图像识别技术的必要性。第2节, 对卷积神经网络基本概念进行阐述。第3节, 介绍目前较为常见的卷积神经网络模型并分析其优缺点。第4节, 通过引用论述说明图像识别算法的研究方向、实际应用和当前面临的挑战。第5节, 对全文进行总结, 对神经网络图像识别算法模型当前面临的问题进行简述, 针对这些挑战给出可能的解决方向。

1 图像识别模型的定义和研究必要性

图像识别是指利用计算机分辨图片中的人物、物体位置等, 并在此基础上进行分析与理解, 最终实现不同模式目标和对象的识别。图像识别的结果取决于神经网络模型的选择以及亟待解决的问题。目前, 神经网络模型被使用在多种场合, 其不透明性、非线性拟合始终被认为是神经网络的主要缺陷之一。这种不透明模型会产生没有根据的结果与识别, 这种无根据结果会导致潜在的信任危机和安全危机, 尤其在必须保证模型高度可靠的领域, 例如医学诊断、自动驾驶、刑事司法、建筑施工等。

2 卷积神经网络概述

卷积神经网络的基础是BP神经网络, 是一种按误差反向传播训练的多层前馈网络, 它们均采用前向传播方式来计算输出值, 采用反向传播方式调整权值和偏置。CNN与BP神经网络最大的不同点是CNN模型相邻两层之间的神经元是部分连接结构, 而不是全连接结构, 即下一层的某个神经元的感知区域只能覆盖上一层的一部分神经元。目前卷积网络模型的发展趋向于多元化, 但是整体结构还是1998年Le Cun和Bottou等人设计出的Le Net-5卷积神经网络。

卷积神经网络主要包括输入层、卷积层、激活函数、池化层、全连接层。每一层均有多个特征图。其结构包括以下几个方面特点:

1. 上一层的局部区域输入得到特征提取过程中每

个节点的特征值。

2. 多个特征图组成特征层且每个特征映射都是二维平面映射, 特征图中所有节点共享相同的卷积核。

3. 卷积层后是局部下采样层, 经处理后的特征图分辨率和变形敏感度降低。

4. 多层神经网络作为最后的分类器。

假设输入的原始图像用 A_0 表示, 用 F_A 表示第 A 层的特征图, 则由如下公式得到下一层的特征图:

$$F_A = f(F_{(A-1)} \otimes W_A + b_A) \quad [1]$$

其中 $f(x)$ 代表非线性激活函数, W_A 为卷积核, \otimes 代表卷积运算, b_A 代表偏置项。

卷积层后一般接入池化层, 池化层是对特征进一步的抽象与降维, 其结果一般是池化区域最大值或者平均值。平均池化是对卷积核内的元素取平均值, 最大池化是对卷积核内的元素取最大值, 通过此操作可以在避免噪声的同时减小卷积引起的均值偏差, 更精确地反映图像的边缘纹理特征。

全连接层在卷积神经网络中起到“分离器”的作用, 通常为模型的最后一层, 将池化层的所有特征矩阵转化成一维的特征大向量, 对提取的特征进行分类。

3 卷积神经网络用于图像识别的常用网络模型

LeNet 模型的提出使得基于卷积神经网络的图像识别模型成为研究重点, 当前利用卷积神经网络模型进行图像识别成果众多, 本节将对图像识别领域具有代表性的卷积神经网络模型进行整理, 将模型的使用条件、优点、缺点等做了详细的描述。

3.1 AlexNet

由于 LeNet 模型架构简单并使用全连接层, 导致其只能使用在内容简单的场景也需要消耗大量计算资源, 所以该模型很少用在实际任务中。与 LeNet 不同的是, AlexNet 模型进行了更深的网络设计并借助参数优化来获取更高维的特征, 以应对特征复杂的图像。

该模型不需要将输入的图像转化为灰度图, 并且首次加入了 CPU 并行运算部署更深、更宽的网络, 以更快的速度进行训练。在模型结构方面利用 ReLU 作为激活函数, 收敛速度较 Sigmoid 和 tanh 更快且不需要进行指数运算, 解决了神经网络中的梯度消失与梯度饱和问题。为了进一步解决收敛问题, AlexNet 模型引入 LRN 局部响应归一化的概念, 实现了“侧抑制”, 即对局部神经元的活动创建竞争机制, 使得其中响应比较大的值变得相对更大, 并抑制其它反馈较小的神经元, 增强了模型泛化能力。同时, Dropout 随机失活

神经元的加入能够随机删除神经元, 使神经网络模型避免过拟合。由于 AlexNet 模型设计池化步长小于池化尺寸, 需要进行重叠的最大池化来避免平均池化产生的模糊效果。

3.2 VGGNet

VGGNet 是在 AlexNet 的基础上创造更深的网络, 在网络深度上达到 19 层。此外, 该模型的适应能力较强, 可根据需求, 将网络深度与网络能力之间的关系展开全面且严格化的模拟。VGGNet 模型中大多使用 3×3 卷积操作和 2×2 的最大池化操作, 使得特征图经过很深的网络之后尺寸减小, 进而增加网络层数。VGGNet 模型为了减少网络参数, 通过多次重复引入包含了小尺寸卷积核的卷积层串联的方式。与具有较大核的卷积层相比, 此方式能够保证相同感受野的同时减少网络的参数量, 提高了网络的非线性表达能力, 获得更好的特征提取与特征学习能力。VGGNet 网络具有模型简洁, 收敛和迭代速度快的特点, 并且由于使用的卷积核较小, 所以计算量速度快。

3.3 Faster R-CNN

Faster R-CNN 算法包括 4 个部分, 分别是对于图像中特征的获取、生成 anchor 区域、检测目标分类和获取目标位置, 然后利用一个神经网络将这 4 个模块结合起来, 训练了一个端到端的网络。接下来叙述 Faster R-CNN 模型的大致流程, 首先将图片输入到 Faster R-CNN 网络, 经过共享卷积层计算后等到公共特征图, 这一步对于输入的图像尺寸并没有要求; 特征图经过 RPN 网络的加工处理得到 anchor 区域, 判断 anchor 区域中是否存在检测物体以及物体类别; 如果经过判断得到的是需要的信息, 则输入到 ROI 池化层得到相同维度的特征, 然后将相同维度的特征进行 softmax 分类以及回归。如果检测的不是前景特征, 则把这些生成 anchor 区域舍去。然后对 anchor 区域进行位置修正, 从而形成较为准确的候选区域。经过不断地技术升级, 该模型能够实现精度较高地物体检测性能。Faster R-CNN 模型首次提出了 RPN 网络, 在提高精度的同时提高了训练速度, 解决了 SS 网络选择性搜索的弊端。模型在算法中使用了 ROI Pooling 技术, 根据预选框的位置坐标在特征图中将相应区域池化为固定尺寸的特征图, 以便进行后续的分类和候选框回归操作。此外, 为了提高检测模型对于全局特征地检测能力以及加大局部特征之间关联性和相似性特征的鉴别性, Faster R-CNN 模型引入了注意力机制。

3.4 残差网络 ResNet

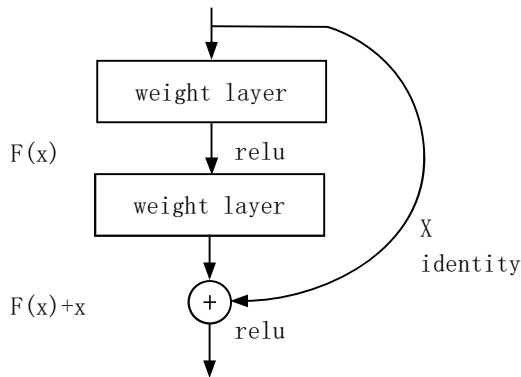


图1 残差单元

为了探索更深层的卷积神经网络同时避免非常深的网络导致精度上的过度拟合和饱和问题，微软研究院何凯明团队搭建了融入残差块 ResNet。ResNet 模型的基本思想是引入了能够跳过这一层或多层的跳接。即图 1 中的由输入 x 连接到 $F(x)+x$ 的弧线。这消除了神经网络由于深度过大而造成训练困难的问题。这使得 ResNet 在图像识别、检测、定位等任务中取得了很好的效果。

3.5 全卷积神经网络模型 (FCN)

全卷积神经网络将卷积神经网络的全连接层转化成了卷积层。与早期的卷积神经网络不同，FCN 可以输入任意尺寸的图像，利用反卷积层进行上采样操作，以此对最深的卷积层产生的特征图进行处理，从而得到与原始图像相同尺寸的特征图，这样就会对每一个像素都产生一个预测。由于 FCN 模型去掉了全连接层，故该模型能够保留原始图像中的空间信息进而解决语义分割问题。

3.6 Mask R-CNN

Mask R-CNN 模型是在 Faster R-CNN 的基础上融合 FCN 的像素识别方法，将原有 ROI pooling 改进为更加精确的 ROI Align。该算法模型由 ResNet101 特征提取网络、FPN 特征融合网络、RPN 区域候选网络、Mask R-CNN Head 输出层 4 部分组成^[2]。

首先 Mask R-CNN 模型将 ResNet 残差网络结构与特征金字塔相结合的方式，对原有的特征图提取的卷积层和池化层结构做出了优化，生成多尺度特征映射图的同时兼顾小目标和细节特征。然后通过 ROI Align 在特征图的各个像素点上产生一系列的预选框，最后通过全卷积神经网络完成对标注图像的提取，完成精确的图像识别。

4 卷积网络在图像识别中的应用以及面临的挑战

卷积神经网络经过了较长时间的发展，已经在很多领域都取得了较好的实验结果，并在一些领域已经有了商业化应用。基于卷积神经网络的图像识别技术具有提取特征能力强、识别精度高以及实施性强等优点，在人脸识别、医疗图像处理、遥感图像识别、农业病虫害识别、公安刑侦等方面运用广泛。

4.1 人脸识别

随着模式识别技术的发展，人脸识别技术逐渐运用到视频监控系统、智能支付、刑侦系统等应用中。但是由于人脸识别因光照、遮挡物等引起的不确定性，使得人脸识别技术依旧面临着许多挑战。在某些特殊的应用中，如皮肤检测，需要采用特殊的光照技术进行人脸拍摄并识别。孙金龙^[3]等人利用 RetinaFace 算法进行网络增强训练，进一步优化不同照明条件的人脸识别技术的鲁棒性。在现实情况中，人脸容易出现不同程度的遮挡，从而造成人脸识别困难。陈灿林^[4]提出基于 Wasserstein 距离的局部遮挡人脸修复算法来修复人脸被遮挡的部分，上述研究的测试结果均表明系统对遮挡人脸识别具有很好的识别效果。

随着科技的发展，虽然人脸识别技术在许多领域均有突破，但是此技术依旧存在潜在的缺陷，比如侵犯权力、潜在的数据盗窃、3D 掩码攻击等问题。

4.2 医疗图像处理

深度学习在医疗图像辅助诊断领域取得了重大的成功。在临床方面，医生可以通过医学成像来诊断患者病情。但是目前对于影像判断病情仍需要人工阅片，在这过程中医生在诊断病情时的主观性不可忽视。而卷积神经网络凭借其能够解决图像级识别问题中的高分辨率小目标识别问题、区域级目标识别问题中的不完整目标标注问题、像素级目标识别的识别精度问题等，成为图像分割、图像分类等领域的重要研究热点之一。在医学领域中，医学图像分割能够对器官形态、病区面积、细胞轮廓等进行分割，在辅助治疗中，医学图像分割可以使医生做出更准确、更权威的治疗方案。张光华^[5]等人通过 LiTS 肝影像数据集和自有数据集，提出的肝分割和肝病灶分割算法提高了肝病灶分割的精确性。黄凯达^[6]等人提出了一种基于空间和密度转换模型来对肺部 CT 影像进行数据增强的算法，为肺部分割提供了很好的技术支持。

4.3 遥感图像识别

随着高分辨率的遥感图像需求不断增长，卫星感

受器等器件受有效载荷、成本和宽带技术的限制,已经无法满足遥感图像分类、图像分辨率高、包含丰富的空间细节信息等技术要求。朱萌^[7]等人研究了一种基于多形态稀疏分解的遥感图像融合方法,能够向卷积神经网络输入稀疏分解的不同形态成分,然后经过关键特征提取能够得到高分辨率并含有丰富信息的图像。王乐萱^[8]等人提出了基于注意力机制的多尺度残差卷积神经,解决了沙漠地区遥感图像地物目标尺度大、难以准确地自动化分类的问题。

卷积神经网络的快速发展使得遥感影像技术得到发展,但是对于某些应用,比如卷积神经网络在云阴影检测、复杂地形勘测等领域还处在初级阶段,待改善的地方较多。

4.4 农业病虫害识别

近年来,由于病虫害种类繁多、植物轮廓和纹理脉络的变化,使得农作物特征提取困难。为了解决此问题,研究人员将目标聚焦于卷积神经网络,利用其擅长提取轮廓、纹理等特征来进行病虫害的检测与识别。

在农作物病害的特征提取过程中,由于数据集分布不均匀、光照等背景因素原因,造成识别准确度低。方桃^[9]等人在 ResNet50 模型用 Focal loss 函数替换了标准交叉熵函数,并采用 Adam 优化方法提高精度,在数据集 Plant Village 上进行实验验证,证明经过改进的算法具有一定的优越性。李进^[10]等人提出了基于多尺度残差网络 MSA-ResNet 的农作物病害识别,将 AI Challenger 2018 数据集分布不均匀的问题进行数据预处理,删除样本数量极少的类别后对数据集标签进行重采样操作,使数据集各类分布较为均匀。在残差块中采用多尺度卷积核替换单一尺度的卷积核来解决农作物病害病斑症状的大小、颜色和形状在不同程度和不同时期不同的问题,通过实验得农作物病害识别准确率达到 89.64%,具有较优的性能表现。

随着卷积神经网络技术的发展,其对农作物病害识别具有准确性、快速性等特点,能够实时有效地检测出病虫害的类别和严重程度,但是由于实际农田的不可靠因素过多,如大气质量、光照因素、种植疏密等问题,其图像采集依旧是一个重要问题。

5 总结与展望

卷积神经网络能通过对数据不断地学习然后提取特征,优化自身性能,所以具有很强的泛化性,其克服了传统模型不能预测复杂的非线性模型的缺点。但是利用卷积神经网络进行图像识别仍然存在许多亟待解决的问题,这些问题是图像识别基础进一步发展的阻碍。

1. 卷积神经网络模型的核心任务之一是图像语义分割,虽然已经对 CNN 网络进行多种多样的改进,使模型既能保证模型简化还能保持良好的分割性、鲁棒性,但还是没有一个基本模型来同时解决模型简化以及良好分割问题。

2. 在生活中图数据具有多样性,也会出现环境中存在与识别物体形状相似的物体,这使得图卷积神经网络的设计需要更加复杂,在卷积神经网络的算法优化等方面带来了诸多挑战。

3. 目前的一些模型是根据某一数据集进行设计的,无论是网络层次结构还是具体的实验研究都是基于固定模型进行设定的,如果能够实现模型对于不同数据集进行自适应设计网络深度与识别训练,可以提高模型的通用性与智能化。

综上所述,虽然卷积神经网络依旧存在许多需要突破的问题,如对卷积神经网络的初始状态参数,寻优算法的选取等问题,但是目前在图像识别领域也出现了新的研究方法和思路来推进图像的高分辨率、高可靠性、实时性等特点,比如使用 Transformer 模型结构代替卷积块等,所以在图像分类、检测、识别等领域应用卷积神经网络依旧是未来的重要研究方向。

参考文献:

- [1] 吕智愚. 基于多尺度几何分解与神经网络的图像去噪研究 [D]. 大连: 大连理工大学, 2021.
- [2] 宋玲, 夏智敏. 人体关键点检测的 Mask R-CNN 网络模型改进研究 [J]. 计算机工程与应用, 2021, 57(01): 150-160.
- [3] 孙金龙, 吴振宁, 肖仲喆, 等. 关于人脸识别算法对不同照明技术的鲁棒性研究 [J]. 电子器件, 2022, 45(05): 1123-1128.
- [4] 陈灿林. 面向部分遮挡人脸识别的研究与实现 [D]. 成都: 电子科技大学, 2019.
- [5] 张光华. 基于卷积神经网络的肝分割技术研究 [D]. 西安: 西安电子科技大学, 2019.
- [6] 黄凯达. 基于模型学习空间与密度转换的肺部图像分割数据增强算法研究与实现 [D]. 重庆: 重庆大学, 2020.
- [7] 朱萌. 基于多形态卷积神经网络的遥感图像融合 [D]. 烟台: 烟台大学, 2022.
- [8] 王乐萱. 基于多分辨率特征融合的沙漠遥感图像识别 [D]. 南京: 南京信息工程大学, 2021.
- [9] 方桃. 基于卷积神经网络的农作物叶片病害分类研究 [D]. 合肥: 安徽大学, 2020.
- [10] 李进. 基于卷积神经网络的多种作物病害识别研究 [D]. 镇江: 江苏大学, 2022.