

基于深度学习的语义分割综述

杨洁洁, 杨 顶

(三峡大学计算机与信息学院, 湖北 宜昌 443002)

摘要: 语义分割是深度学习计算机视觉方面的核心领域, 有着很深的研究价值。语义分割技术的发展在近几年趋于成熟, 从传统的方法到基于卷积神经网络方法的突破, 构建了端到端的语义分割深度学习神经网络算法。这些方法被用于人工智能当中, 应用在无人驾驶, 遥感影像检测, 医疗影像研究等方面。基于对经典语义分割算法进行学习, 每个经典算法都有自己的特点, 值得在此一一总结阐述。文章将针对语义分割的发展, 优秀算法的网络架构特色, 应用的场景进行介绍, 最后将对语义分割算法作小结和展望。

关键词: 深度学习; 计算机视觉; 传统算法; 语义分割; 卷积网络

中图分类号: TP391

文献标识码: A

文章编号: 2096-9759(2022)02-0069-04

Overview of semantic segmentation based on deep learning

YANG Jiejie, YANG Ding

(College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443002, China)

Abstract: Semantic segmentation is the core field of deep learning computer vision, which has deep research value. The development of semantic segmentation technology has matured in recent years. From traditional methods to methods based on convolutional neural networks, an end-to-end deep learning neural network semantic segmentation algorithm is realized. These methods are used in artificial intelligence, applied in unmanned driving, remote sensing image detection, medical image research and so on. Based on the learning of classical semantic segmentation algorithms, each classical algorithm has its own characteristics, which is worthy of summarizing and elaborating here. This article will focus on the development of semantic segmentation, the network architecture features of excellent algorithms, and the application scenarios. Finally, it will summarize and look forward to semantic segmentation.

Key words: Deep Learning; Computer Vision; Traditional Algorithm; Semantic Segmentation; Convolutional Network

0 引言

人工智能在生活中的多方使用, 推进着深度学习计算机视觉领域在各方进行技术创新, 技术融合。在图像分割当中, 许多研究人员都将目标转向语义分割算法研究当中, 近些年已经研究出了很多经典的语义分割算法, 这些算法相较于传统算法的不同之处是使用了神经网络进行分割, 并不断改进, 进而应用在实际生活图像处理方面。

从传统图像分割算法到基于深度学习语义分割算法, 形成了图像分割算法史。传统方法中, 研究者提取了图像的颜色值、灰度值和几何形状等特征将图像分割成多个没有相交的区域, 这些区域内具有相似性特征, 不同区域具有不同的特征, 最后对这些分割的区域进行语义标注得到图像分割。此类方法分割的精度不够高, 为了解决这个问题, 研究者开始探索将神经网络应用在语义分割算法上, 神经网络可以自主学习图像特征, 提升了分割精度, 得到了端到端, 像素到像素的语义分割算法。

从 2015 年 FCN 的提出开始, 经典的基于神经网络的语义分割算法开始出现, 这些文章值得研究学习。本文将在第一章介绍 4 个简单的传统图像分割算法, 介绍 4 个现阶段经典的基于深度学习的语义分割算法。第二章介绍 3 个语义分割算法实验常用数据集, 第三章介绍语义分割的实际应用场景, 第四节进行总结与展望。

1 语义分割算法

1.1 传统图像分割算法

研究传统语义分割算法, 可以在后期与基于神经网络的语义分割算法进行对比理解, 传统图像分割算法不可小觑, 它为后期语义分割的发展奠定了基础。以下介绍 4 个传统图像分割方法: 阈值分割法、马尔可夫随机场模型、边缘检测法、区域法。

1.1.1 阈值分割法

传统分割算法中最为基本的, 使用最广泛的是阈值分割法^[1]。阈值分割法会设定不同的阈值, 这些阈值在图像上代表像素点, 会根据不同灰度级的像素点将图像的不同区域分割出来, 从而分离目标和背景。阈值分割法的关键点是最佳阈值的选取, 找出最佳阈值可以使得图像分割更为合理, 得到效果好的分割图。

阈值分割法中常用的方法有: 灰度直方图双峰法、最大类间方差(Otsu)法、自适应法等。传统的阈值分割法只专注于研究像素本身的灰度值, 而像素点在空间上的分布却没有关注到, 这样就会受到噪声的影响, 分割就会不精确。

1.1.2 边缘检测法

边缘检测算法^[2]定位图像边缘准确, 且运算速度快。边缘是目标与背景之间灰度值突变的边界线, 边缘检测算法对边缘存在灰度差异较大的点定为边缘点, 得到全部边缘点后, 将边缘点相连成有连贯性的轮廓, 对不同的轮廓区域进行图像分割。

边缘检测分割方法因检测边缘点的导数算子不同, 可分为一阶和二阶导数算子。常见的一阶导数算子包括梯度算子、

收稿日期: 2022-01-05

作者简介: 杨洁洁(1995-), 女, 陕西渭南人, 硕士, 主要研究方向: 计算机视觉。

Roberts 算子、Sobel 算子、Canny 算子等。常见的二阶导数算子包括 Laplacian 算子、Wallis 算子、Kirsch 算子等。边缘检测的两个缺点分别为,其一检测的边缘点如果不够密集,边缘点连接构成的边缘可能不够连续不够封闭。其二图像的高细节区域会有很多碎小边缘,不能形成大区域。在图像分割当中,边缘检测法可与其他算法相结合进行分割。

1.1.3 区域法

基于区域的分割方法^[3]其原理是图像各个区域,如果有相同性或相似性特征,就会被组为同一个区域。区域法被分为两种方法,其一区域生长,它是串行区域技术,基本思想是将像素点具有相同或相似的邻域像素进行合并构成更大的区域,直至领域像素不再与区域的像素值相似或相同时,结束合并,得到区域分割结果。其中区域生长的关键点是找到一个像素点即种子点。其二区域分割合并,算法流程是整个图像会先一直分裂成各个子区域,子区域是具有相似性或相同性的像素,然后将各个子区域之间具有相似或相同性地进行迭代合并。

在图像分割当中,会将区域生长和区域分割合并算法进行结合使用,由于其分割特性,在实际应用中,对复杂的分割场景分割效果会较为理想。

1.1.4 马尔可夫随机场模型算法

基于马尔可夫随机场模型^[4]是一种关于统计学习的图像分割算法,马尔可夫是对图像局部区域分割,可以根据图像中像素之间的相关性来确定先验概率,是建立在马尔可夫模型(MRF)和贝叶斯理论(Bayes)基础上的,由[4]知,图像分割问题的目标函数由统计决策和估计理论中的最优化准则来确定,利用优化算法求解满足这些约束条件下的最大可能分布,从而将分割问题转换为 MRF 的最优化问题。其中最常用最优化准则是最大后验概率(MAP)。

1.2 卷积神经网络语义分割算法

传统图像分割算法提取了图像的低级语义(图像像素的颜色、纹理和形状信息),没有考虑图像中的高级语义信息,使得图像的分割效果不够好,基于深度学习的卷积神经网络语义分割算法提取图像的中高层语义信息(图像或图像区域所包含的对象和实体进行分割),会比传统算法得到更高精度的语义分割结果。其经典的卷积神经网络语义分割算法有 FCN, PSPNet, U-Net, Deeplab 系列等。

1.2.1 FCN

FCN 网络^[5]是 CVPR2015 的最佳论文,是基于深度学习的语义分割技术的开山之作。FCN 将 CNN 网络(AlexNet, VGGNet 等)中的全连接层换成全卷积层,构建了端到端,像素到像素的语义分割网络。FCN 的网络特点有:卷积化、上采样和跳跃结构。

卷积化(Convolutional),例如 CNN 网络一共有 8 层,前 5 层是卷积层, CNN 的后 3 层的全连接层全部转换为卷积层。全连接层是一维向量,它会将二维空间的图像压缩为一维向量,这样图像就失去了空间信息,后三层转换为全卷积层,图像就会一直保持二维性,空间信息就得到保留。

上采样(Upsample),上采样也叫反卷积操作,上采样操作在网络最后一层。当多层卷积使得图像特征图变小后,加入上采样就使得图像最后的输出分割图尺寸与原图的尺寸保持一致。

跳跃结构(Skip Layer),FCN 网络中采用跳跃结构,将不同池化的结果进行上采样之后结合(结合方式为 Element-wise add)在一起,得到最后的输出,分别有 8 倍的上采样,16 倍的上采样和 32 倍的上采样。

FCN 网络效率高,复杂度低,语义分割结果还不够精细,此网络没有考虑像素之间的关系,缺乏空间的一致性,会造成分割时类别的误判。

1.2.2 PSPNet

2017 年 PSPNet^[6]文章对 FCN 网络存在的分割问题进行了改进,其主要创新点是提出了空间金字塔模块,图像的上下文信息和多尺度信息会被提取,分割层就有了更多的全局信息,降低了 FCN 网络中会使得图像类别误分割的概率。PSPNet 在网络层添加空洞卷积(dilated convolution),全局均池化操作将图像感受野增大(Receptive Field),包含了图像的浅层深层特征。

空间金字塔模块如图 1(Pyramid Pooling Module),其包括全局池化操作,1*1 卷积操作,以及上采样和 concat 操作。网络操作流程为:输入图像经过 CNN 操作得到特征图,特征图经过四种不同尺度的池化操作,得到四个不同尺度的特征,再经过 conv 1*1 得到通道数为输入通道数的 1/4 的四个特征,再行双线性插值上采样使得四个尺度特征尺寸与特征图大小相同,然后将所有特征与原始特征图进行 concat 用于后续分割预测。空间金字塔模块融合了不同尺度的上下文信息,增大了感受野区域,1*1 卷积融合通道信息。

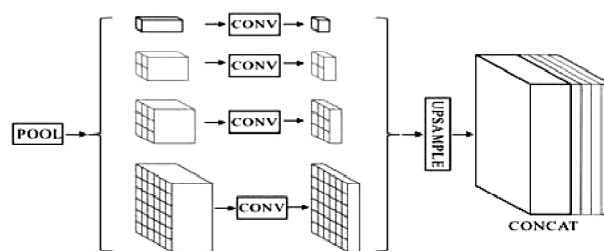


图 1 空间金字塔模块图源^[6]

1.2.3 U-Net

U-Net 网络^[7]是 2015 年发表在 MICCAI 上的文章,是基于 FCN 架构的改进,此网络结构因其采用了 Encoder 和 Decoder 的 U 型结构及 skip-connection,结合了图像的底层和高层信息,广泛应用在解决医学领域的图像处理问题中。

U-Net 网络为两个部分,其 Encoder 为特征提取部分,下采样 4 次,总共在 Encoder 部分下采样了 16 倍,提取了图像的底层信息,即上下文语义信息,有利于类别分类。其 Decoder 为上采样部分,每次上采样之后就与下采样对应的通道数相同尺度相融合,即 skip-connection。Decoder 部分提取的是图像高层信息,对图像分割提供精细的特征定位。

在特征融合方面,U-Net 新颖的融合方式是将特征在 channel 维度进行拼接融合,而 FCN 的融合是逐点相加。U-Net 适用医学影像分割的原因是医学影像的语义简单,人体器官结构固定,U-Net 能运用低分辨率的信息,为目标类别分类;医学影像的边界模糊,U-Net 能提供高分辨率的信息,提供精确定位。

1.2.4 DeepLab 系列

DeepLab^[8-11]是谷歌团队基于 CNN 开发的语义分割模型,

目前有四个算法,最新算法是 DeepLabv3+。

DeepLabV1 是基于深度卷积神经网络 DCNN 的改进,解决了 DCNN 在分割图像上存在的两个问题。其一重复池化和下采样导致分辨率降低,使得图像位置信息无法恢复,DeepLabV1 运用 Atrous 卷积解决。其二由于空间不变性导致的细节信息的丢失,DeepLabV1 运用全连接的 CRF 提高了模型精细细节的分割。DeepLabV1 网络架构使用 backbone 为 VGG-16,使用了 Atrous 卷积,以及全连接 CRF。Atrous 卷积的一维算法操作如图 2,图中 Input stride 为空洞因子,算法在卷积核大小相同时,通过增加空洞因子来增大感受野。全连接 CRF 主要是在分割时确定一个位置的像素值时,会考虑这个像素周围的像素点值,用图像的全局信息来恢复图像的局部信息,精确图像的轮廓。

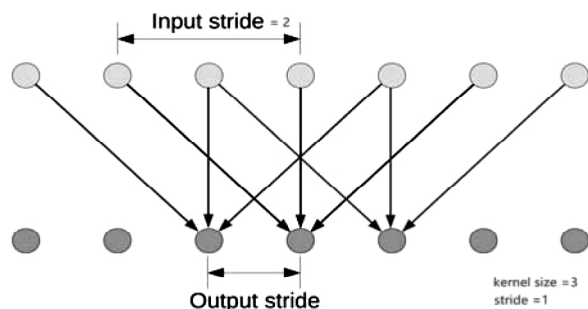


图 2 Atrous 卷积图源^[8]

DeepLabV2 在 V1 的基础上,改进网络架构,比 V1 多解决的问题是:图像中存在多尺度物体。其改进的网络架构有:丢弃一些池化操作,使用空洞卷积;使用全连接 CRF 细化细节信息;提出空间金字塔池化 ASPP,解决图像存在多尺度物体的问题。ASPP 模块使用多条平行的有不同采样率的空洞卷积在给定的特征层上进行重采样,再将特征进行融合。

DeepLabV3 在 V2 的基础上,不再使用全连接 CRF,研究更有深度结构的 Atrous 卷积并且改进 ASPP 模块。串联不同膨胀率的 Atrous 卷积得到更深结构,并行不同膨胀率的空洞卷积优化 ASPP。

DeepLabV3+也不使用全连接 CRF,提出了深度可分离卷积,把 DeepLabV3 看作 Encoder,添加 Decoder 得到 DeepLabV3+模型。作者将 SPP 与 Encoder-Decoder 模型相结合,把 Xception 和深度可分离卷积应用到 ASPP 和 Decoder 当中。其中 SPP 模块可以提取更多的尺度信息,Encoder-Decoder 模型可恢复物体的边缘信息,提出的深度可分离卷积在保持性能的同时可大大减少计算量。

2 数据集

语义分割的数据集从不同维度来划分,有相当多的数据集。基于二维图像方面热门的数据集有: PASCAL VOC2012 系列、Microsoft COCO 系列、Cityscapes、SYNTHIA、CamVid、KITTI 等数据集。下面介绍三个比较常用的数据集。

2.1 PASCAL VOC2012 数据集

PASCAL VOC2012 是 Pascal 系列语义分割方向最常用的数据集,其训练集,验证集,测试集分别为,1464,1449,1456 张,经过数据增强后的数据集其训练集达到了 10582 张。VOC2012 有 20 个类和 1 个背景类,从人、动物、交通工具、室内四个类型出发分为 20 个类别,加上背景为 21 类。VOC2012 中用于分割比赛的图片包括原图 jpg 格式,图像分类分割 png 格式和图像

物体分割 png 格式。图像分类分割中,物体分割颜色是特定,而图像物体分类分割中,图像生成由不同物体生成不同轮廓,颜色随机填充。VOC2012 数据集文件夹包含 5 个文件夹,Annotations 文件夹对应其图像的 xml 信息,ImageSets 文件夹中的 Segmentation 的三个 txt 文件为标记的用于图像分割的图像,JPEGImages 文件夹为原图,SegmentationClass 文件夹包含 png 图像用于图像分割分类,SegmentationObject 文件夹的 png 图为不同物体的分割。

2.2 Microsoft COCO 数据集

Microsoft COCO 数据集用于场景理解的任务中,图像来源于复杂背景下的生活场景。COCO 数据集有 91 个物体类别,它含有 32.8 万张图片,标注实例有 250 万个,是目前为止最大的语义分割数据集。这个大规模数据集专注于解决图像场景理解中的三个关键问题:目标分类,目标检测和场景语义标注。COCO 数据集其特点是每一张图片平均由 3.5 个类别和 7.7 个实例组成的,评估标准比 PASCAL VOC 严格,大家乐意用它来测评模型质量。

2.3 Cityscapes 数据集

Cityscapes 数据集是对城市街道场景的语义理解,从 50 多个不同的城市中根据不同的季节,良好的天气情况手动选出视频中的帧,由大量动态对象,各种场景布局和变化的背景所记录产生的大规模数据集,有 5000 张高质量标注的图片,2 万多张粗糙的标记图片。它提供了 8 种类型 30 个类别的语义化,实例化和密集像素的标记,8 种类型为平面、人、车辆、建筑、物体、自然、天空、虚空,用于训练深度神经网络。

3 实际应用领域

语义分割应用于地理信息系统,对于一张卫星遥感影像,通过对语义分割模型训练,训练好的模型就可以识别图像中的道路、森林、庄稼、建筑物等类别。这可以直接应用在森林资源,环境的监测管理,农田生物植物量的检测与预估及城市道路规划等方面,提高了在这些研究区域的效率并且降低了人工成本。

语义分割应用于自动驾驶,自动驾驶的车辆需要感知周围环境,自动进行规划路线,使得驾驶车辆安全前行,其安全性使得语义分割技术需要高精度的道路分割,其分割类别包括道路,人和车。自动驾驶车辆可以通过车载摄像头或者雷达探测器中拍摄的图像进行神经网络,后台设备就会将图像分割成相应类别,识别出道路哪些区域可绕过行人和其他车辆安全在驾驶。

语义分割应用于医学影像中,涉及在医学治疗中肿瘤的分割识别,头部 CT 扫描分割及扫描估计器官的体积等。神经网络与医疗诊断的结合,将使智能医疗领域快速发展。

除外语义分割还应用于机器人领域,场景理解,时尚产业,卫星图像处理等领域,在这些领域起着重要的作用。

4 总结与展望

本文从传统图像分割算法出发,首先介绍了 4 种传统图像分割算法,又介绍了 FCN、PSPNet、U-Net 和 Deeplab 四个经典的基于深度学习的语义分割网络,非常值得语义分割的研究者进行学习探讨。接着介绍了语义分割的数据集,进行实验的研究者可根据数据集特色选取相应数据集进行语义分割实验。最后介绍了三个语义分割实际应用场景,对人类社会

的改善。

根据对语义分割传统和基于深度学习的语义分割模型进行总结, 未来的语义分割具有以下挑战:

(1) 数据集处理的工作较费时间, 若需要创建一个新的数据集需要大量的时间去标注数据, 对类别的定义也多加衡量, 且还要中衡数据集类别的数量分布, 耗时耗力。

(2) 语义分割需要在实际应用中结合, 就需要考虑实际环境的参数量, 计算量方面是否达到语义分割实验的要求, 语义分割需要设计出轻量级的, 计算量低的模型, 才能在实际场景中广泛使用。

(3) 语义分割在其视频语义分割, 实时语义分割方面, 还具有很大研究价值, 不仅需要提升分割精度, 还需要在速度上进行提升。

参考文献:

- [1] 刘硕. 阈值分割技术发展现状综述 [J]. 科技与应用, 2020 (24):129-130.
- [2] 黄旭, 张世义, 李军. 图像分割技术研究综述 [J]. 装备机械, 2021(02):6-9.
- [3] 王秋萍, 张志祥, 朱旭芳. 图像分割方法综述 [J]. 信息记录材料, 2019, 20 (07): 12-14. DOI: 10.16009/j.cnki.cn13-1295/tq. 2019.07.005.
- [4] 徐胜军, 韩九强, 刘光辉. 基于马尔可夫随机场的图像分割方法综述 [J]. 计算机应用研究, 2013, 30(09):2576-2582.

- [5] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39 (4): 640-651.
- [6] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network [C]// IEEE Computer Society. IEEE Computer Society, 2016.
- [7] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
- [8] Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs [J]. arXiv, 2014.
- [9] 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, 2018, 40(4):834-848.
- [10] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [J]. arXiv preprint arXiv:1706.05587, 2017.
- [11] 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.

(上接第 68 页)

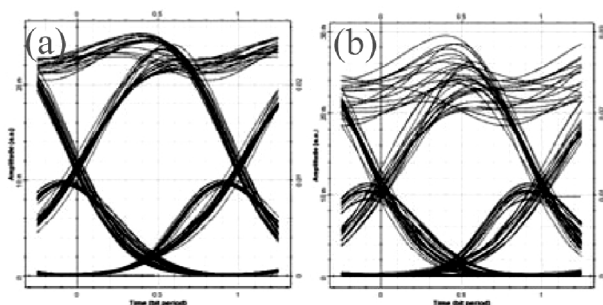


图 5 通道 5 两种补偿方式眼图对比图

在图 4 和图 5 中, (a) 为干路 DCF+支路 CFBG 的前置补偿方案的眼图; (b) 为传统前置色散补偿方案的眼图。从图中分别可以看出 (a) 的眼图比 (b) 的眼图张开度大, 效果好。两种色散补偿方式相结合的方案中, 支路 CFBG 精确地针对各路光波色散斜率造成的影响进行了补偿, 大幅度减小了各路信号由于波长不同造成的色散补偿差异, 因此得到了较好的信号质量。

4 结语

本文着重分析了采用 80Gb/s 的 DCF 前置色散补偿技术的 WDM 系统。使用 Optisystem 软件, 采用传统 DCF 前置色散补偿和 DCF 前置色散补偿加分波后 CFBG 的混合补偿方案, 通过眼图、误码率和 Q 值对系统的性能进行了比较评估。仿真实验结果表明, 干路 DCF 支路 CFBG 的混合补偿方案的 Q 值和眼图效果都优于传统方案, WDM 系统的性能得到了进一步的提高。

参考文献:

- [1] Neheeda P, Pradeep M, Shaija P J. Analysis of WDM System

With Dispersion Compensation Schemes [J]. Procedia Computer Science, 2016, 93 (2016):647- 654.

- [2] SAYEEDA Khatoon, JAISWAL A K, ADITI Agrawal. Performance Evaluation of Post and Symmetrical DCF Technique with EDFA in $32 \times 10, 32 \times 20$ and 32×40 Gbps WDM Systems [J]. International Journal of Current Engineering and Technology, 2017, 7(4):1416-1421.
- [3] Permadi Aji Insan and Isnawati Anggun Fitriani and Zulherman and Dodi. Comparative analysis of the dispersion compensating fiber (DCF) scheme in long-haul dense wavelength division multiplexing (DWDM) [J]. Journal of Physics: Conference Series, 2019, 2019(1367):012063-012063.
- [4] M.L.Meena, Raj Kumar Gupta. Design and comparative performance evaluation of chirped FBG dispersion compensation with DCF technique for DWDM optical transmission systems [J]. International Journal for Light and Electron Optics, 2019, 188 (2019):212-224.
- [5] Kamaldeep Kaur and Beant Kaur. Dispersion Compensation Techniques: A Review [J]. Research Cell: An International Journal of Engineering Sciences, 2016, 20(0) : 58-61
- [6] 申静, 潘建. 色散补偿光纤的分波长色散补偿 [J]. 光通信技术, 2019, 43(2):27-29.
- [7] 曹雪. 以啁啾光栅对波分复用系统色散补偿的研究 [J]. 光电子·激光, 2021, 32(02):217-222.
- [8] 曹雪. WDM 系统色散分波补偿技术研究 [J]. 光通信技术, 2013, 37(12):18-20.