

基于深度神经网络的图像语义分割综述

郑 凯, 李建胜

(信息工程大学, 河南 郑州 450001)

摘 要: 近年来, 基于深度学习的方法在计算机视觉、模式识别等领域取得了令人鼓舞的成就, 并且在图像分类、目标检测等方面的精度已经超过人类手工操作。图像语义分割作为场景理解的基础, 分割的精度直接决定着诸如自动驾驶、三维重建、无人机控制与目标识别等应用的质量。本文首先简要介绍了图像分割问题的历史发展脉络, 而后重点针对采用深度学习方法的图像分割进行阐述, 包括基于全卷积网络以及基于整合图像上下文知识的方法, 并分析其优缺点, 最后, 对未来可能的发展趋势与研究热点进行了展望。

关键词: 图像分割; 语义分割; 实例分割; 全卷积网络; 上下文知识

中图分类号: P208 **文献标识码:** A **文章编号:** 1672-5867(2020)10-0119-07

An Overview of Image Semantic Segmentation Based on Deep Learning

ZHENG Kai, LI Jiansheng

(Information Engineering University, Zhengzhou 450001, China)

Abstract: Recently, methods based on deep learning earned encouraging achievements in computer vision, pattern recognition and its accuracy on image classification and object detection has exceeded human. As the base of scene parsing, the accuracy of image semantic segmentation directly determines the quality of autopilot, three-dimension reconstruction, drone control and target recognition and so on. Starting from the historical development thread of image segmentation problem, methods for image semantic segmentation based on deep learning are expounded and the advantages and disadvantages of methods based on fully convolutional network and those based on integrating image context knowledge are analyzed. Finally, developing trends and researching hotspots are forecasted.

Key words: image segmentation; semantic segmentation; instant segmentation; fully convolutional network; context knowledge

0 引 言

图像分割是将图像细分为其组成部分或对象的过程^[1], 这是图像处理到图像分析的关键步骤。从20世纪60年代到深度学习广泛应用于图像领域之前, 研究人员已经提出了上千种方法用于解决图像分割问题, 如基于阈值的方法^[2]、基于边缘的方法^[3]、基于区域的方法^[4]、基于图论的方法^[5]、基于聚类的方法^[6]、基于模糊理论的方法^[7]等结合特定理论和工具的方法。但是这些方法大多是借助图像像素自身的中、低阶视觉信息, 如颜色、纹理、形状、粗糙度、方向度与紧凑度的相关性来进行图像分割, 针对特定背景的分割能力较强, 但在较复杂的任务上表现差强人意。所以, 一直以来还没有一种普遍适用的分割解决方案, 往往需要结合相关领域的先验知识才

能进行有效的分割。

图像语义是对像素信息的进一步组织和抽象, 使其符合人类活动的语义和逻辑^[8]。具体来说, 是图像或图像区域中包含的对象或实体的类别等高级信息。1978年, Ohta等人提出, 图像语义是指在图像中为每个像素分配一个预先定义的表示其语义知识的类别标签。当每个像素有了计算机能识别的标签, 就能应用于目标识别乃至场景理解等应用场景。当前, 越来越多的应用场景需要利用图像中的语义知识, 包括自动驾驶中的场景识别与理解^[9]、无人机控制与应用^[10]、医学影像辅助分析^[11]、图像搜索^[12]、增强现实^[13]等。近年来, 深度学习引爆了图像处理领域的革命, 图像分类、语义分割和目标识别等许多计算机视觉问题都开始使用深度神经网络来解决, 最早出现且最常用的卷积神经网络(CNN), 其在准确率和

收稿日期: 2019-07-16

作者简介: 郑 凯(1990-), 男, 福建三明人, 控制工程专业硕士研究生, 主要研究方向为计算机视觉、定位理论与方法。

通讯作者: 李建胜(1975-), 男, 湖北钟祥人, 教授, 博士, 2011年毕业于信息工程大学摄影测量与遥感专业, 主要从事计算机视觉、视觉导航、SLAM等方面的教学与科研工作。

效率方面都远远超过传统方法。

国内外一些图像语义分割的综述,比如文献[1]等,这些文献在总结现有方法,讨论数据集和算法评价指标以及展望未来研究方向等方面给出了详细阐述。基于深度学习的图像分割是当前一个十分热门且发展迅速的研究领域,故提供基于深度学习的图像分割相关介绍,并按主要原理梳理典型方法的发展历程对于把握当前研究方向、加深对问题的认识是有意义的。

1 目前的主要方法

深度学习图像分割算法的发展经历了从粗糙到精细、从局部到全局的过程。本节通过思维导图归类整理常用的方法,并重点分析典型方法的实现过程。总的来说,主要方法围绕全卷积网络与图像上下文知识展开。基于全卷积方法的基本原理是采用级联的卷积层与池化层,不断抽象图像中的特征,得到特征图(feature map),最后通过转置卷积(transposed convolution)插值得到还原为原始大小的特征图,完成图像逐像素的语义分割。基于图像上下文知识的基本原理是在 CNN 的处理过程中加入图像特征的全局信息,有 3 种不同的方式:一种是在 CNN 中通过不同层特征的跳跃连接或多尺度网络的特征提取,得到融特征合,或通过扩张卷积(dilated convolution)得到更大的图像感受野,从而得到更多的图像语义知识辅助最后的语义分割;一种是在 CNN 处理结束后引入条件随机场(CRF)作为网络的后续模块对分割结果进行优化;还有一种是借助于循环神经网络(RNN)的记忆能力,将图像特征作为序列输入,用以对全局上下文信息建模,改善语义分割结果。

1.1 基于全卷积网络的分割方法

深度学习在高级视觉任务中取得了巨大的成功,特别是有监督学习的方法如 CNN 等用于图像分类或物体检测,这些成果激励并启发研究者探索这些方法的语义分割能力。2014 年,第一个用于图像分割的全卷积网络(FCN)^[15]被提出,其中全连接层被卷积层取代,用于分类的 CNN 网络被转化为生成空间热图的分割网络。同时,添加转置卷积层以实现上采样,使得网络可以执行密集推理并学习到图像中每个像素点的语义标签。基于全卷积网络的分割方法如图 1 所示。



图 1 基于全卷积网络的分割方法导图

Fig.1 mapping of segmentation method based on total convolution network

1.1.1 编码器-解码器

编码器网络通常使用 VGG-16 等基础分类网络,并

移除其全连接层,这个部分通常被称为编码器以产生低分辨率的图像或特征映射。解码器用于学习如何将这些低分辨率的图像转换为图像分割映射到像素级别上去,这部分被称为解码器。

SegNet^[16]是一个典型的例子。其网络分为两部分,左边是连续卷积层提取特征,通过池化增大感受野,同时图像变小,该过程称为编码;右边是转置卷积与上采样,通过转置卷积使得图像分类后特征得以重现,上采样还原到图像原始尺寸,该过程称为解码,最后通过 Softmax 分类器,输出不同分类的最大值,得到最终分割图,网络结构如图 2 所示。

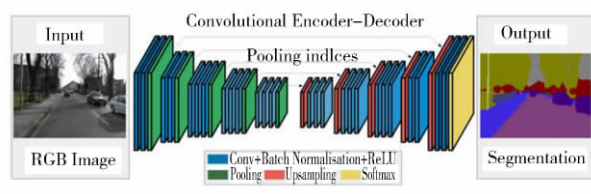


图 2 SegNet 架构示意图

Fig.2 Schematic diagram of segNet architecture

在此基础上,发展出了诸如贝叶斯 SegNet、Unet、RefineNet 以及带有大维度卷积核的全局卷积网络 GCN 等网络。

1.1.2 实例分割

实例分割通常被认为是语义分割的后续工作。如果输入图像中有 2 个来自相同类别的对象,普通语义分割不能将其区分为单独的对象,而实例分割可以用不同颜色区分同一类物体的各个对象。

Hariharan 等人^[17]于 2014 年提出了一种同时检测和分割的方法,该方法采取一种自下而上的层次化图像分割方法——多尺度可结合组(MCG),以此得到网络建议的分割区域。对于每个区域,使用 R-CNN 网络(Regions with CNN)提取特征,接着,对每个建议的区域使用线性支持向量机在 CNN 顶层特征上进行分类,最后通过非极大值抑制输出结果。

Pinheiro 等人^[18]提出了深度掩膜(DeepMask)模型,它使用单个卷积网络对输入图像给出分割提议。对于 1 个输入的图像块预测出 1 个分割的覆盖区域,同时生成图像块中包含物体的概率,网络同时学习上述 2 项任务,除了网络的最后 1 层外,多数的网络层是由 2 个任务共享的。

He 等人^[19]于 2018 年提出了 Mask R-CNN 模型,统一了图像目标检测和语义分割。它的贡献主要是以下几点:Mask R-CNN 使用类似 Faster R-CNN 的框架,Faster R-CNN 的输出是物体的预测框(bounding box)和类别,而 Mask R-CNN 多了 1 个分支,用来预测物体的语义分割图。同样是让网络同时学习 2 项任务,并互相促进。其次,在语义分割中使用二值预测(Binary Mask)。之前的语义分割预测类别需要使用 0、1、2、3 等数字代表各个类

别,而在 Mask R-CNN 中,目标检测分支会预测类别,分割分支只需要用 0 和 1 预测该物体的形状掩膜。第三,Mask R-CNN 提出了 RoiAlign 用来替换 Faster R-CNN 中的 RoiPooling。RoiPooling 的作用是将输入图像中任意一块区域对应到特征图中的对应区域,由于使用了近似方法来寻找对应区域,导致对应关系与实际情况有偏差。在分类任务中可以忽略此近似带来的偏差,但对于分割来说影响不能忽视。为解决该问题,RoiAlign 使用线性插值改进了近似操作,以找到更精准的对区域。

1.1.3 三维数据分割

三维几何数据如点云或网格数据等因其提供附加的维度使算法可以利用更丰富的空间信息,从而有助于分割。但是,绝大多数基于 CNN 的深度学习分割技术,原本并不是被设计为解决非结构化或者不规则输入问题。为了使用卷积网络中参数共享以及其他优化形式,多数研究者借助三维立体网格将非结构化的不规则点云或网格转化为普通表示形式,然后再将其输入网络中。例如,Huang 等人^[20]将点云数据通过一个密集的立体网格对其进行分析,生成了一系列的占位体元,将其作为三维 CNN 的输入并为每个体元产生 1 个标签,最后,将标签映射回点云。虽然该方法已被成功应用,其仍存在难以量化、丢失空间信息以及不必要的大规模表示等。因此,研究者们致力于构造出一种可以直接处理非结构化三维点云或网格的分割模型。

PointNet^[21]给出了一个同时进行分类和分割的端到端模型。图 3 展示了它处理无序三维点集的双模块结构。

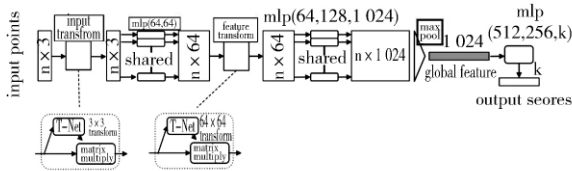


图 3 PointNet 网络模型架构^[21]

Fig.3 PointNet network model architecture^[21]

PointNet 网络分为两部分,分别负责分类和分割。分类子网络以点云为输入,采用一系列变换以及多层感知机(MLP)来生成特征,然后使用最大池化来生成全局特征以描述原输入的点云,并对这些全局特征进行分类,然后为每一个类得出分数。分割子网络将全局特征与分类网络生成的每个点的特征串联起来,然后应用另外的两个 MLP 来生成特征,并计算出每个点属于每一类的得分。

1.1.4 视频序列

视频作为图像序列,其分割处理与单幅图像分割相似又不完全相同。当处理图像序列时,许多方法简单地使用逐帧处理的方法,虽然可以得到不错的结果,但因其处理非常耗时,因此通常不可行。另外,逐帧处理完全忽略了图像间的短期连续性以及一致性信息,而这些信息极可能会提高系统准确率并减少运行时间。

针对图像序列,最具标志性的工作是 Shelhamer 等

人^[22]提出的时钟 FCN 网络。该网络源于 FCN,使用视频中的短期线索来减少推理时间,同时保证正确率。这种时钟方法的原理如下:网络中短期的特征变化率在各层中经过某帧时是变化的,因此来自浅层的特征变化的比来自深层的特征要快。基于此假设,各层可以被分为不同的阶段,从而可以按照其深度指定不同的更新率来进行处理。由于图像序列语义上的稳定性,深度特征可以在各帧被保持,提取出的特征在静止的帧时将会被保留,而在动态的帧时将会被重新计算,从而节省了推理时间。图 4 展示了这种时钟 FCN 的网络结构。

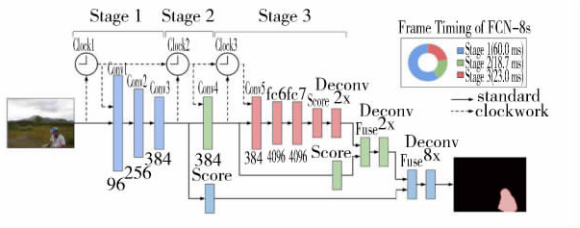


图 4 三阶段的时钟 FCN 模型以及其对应的时钟速率^[22]

Fig.4 Three phase clock FCN model and its corresponding clock rate^[22]

在视频序列上应用三维卷积也是一种重要的思想,Tran 等人于 2015 年提出了一种深度的端到端的、体元到体元的预测模型。该模型将输入的视频片段分为包含 16 个帧的小片段,对每个片段单独进行预测。其主要的贡献在于使用了三维卷积技术,通过三维的滤波器,适应了从多通道数据中学习出的时空联合特征。

1.2 基于上下文知识的分割方法

语义分割需要对多种空间尺度的信息予以整合,也需要对局部与全局信息进行平衡。一方面,低层的或局部的信息对于提高像素级别预测是有帮助的;另一方面,整合图像全局的上下文信息对于解决局部模糊性问题来说也是重要的。一般的 CNN 模型对于处理这种平衡不是很擅长。池化层可以使网络取得某种程度的空间不变性并保持同样的计算效率,却丢失了全局的上下文信息。即便是没有带池化层的 CNN 网络,功能也同样受限,因其神经元的感受野只能随着层数线性增长。

目前,常采用多种方法来使 CNN 掌握全局信息:用条件随机场(CRF)作为后处理过程来调优结果、特征多尺度聚合,以及对上下文的建模放到另一种深度网络中,如 RNN 等(如图 5 所示)。

1.2.1 条件随机场(CRF)

CNN 结构内在的空间不变性限制了其在图像分割上的准确率。优化分割的输出并强化其捕捉细粒度信息的一个通用的办法是引入 CRF 作为网络的后处理模块。CRF 方法是一种基于底层图像像素强度进行“平滑”分割的图模型,它能将像素强度相似的点标记为同一类别,促成了底层图像信息相互关系与产生像素级类别标签的多类别推理输出的结合,这种结合对于捕捉长期依赖性质尤其重要,弥补了 CNN 的不足。

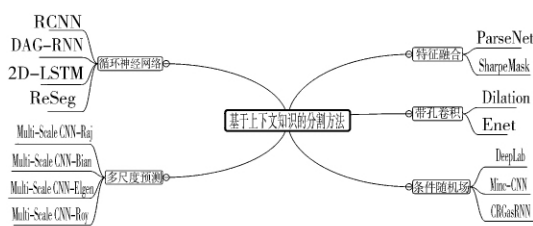


Fig.5 Segmentation method based on context knowledge

DeepLab 模型^[23]使用了全连接的 CRF 模型作为其流程中的一个独立的后处理步骤,以此优化分割结果。该模型将每个像素建模为某区域内的一个节点,无论两个像素距离多远,其两两之间的关系都会被考虑,因此,该模型也被称为全连接因子图。使用此模型后,分割过程中的短期、长期的像素相互关系都将被考虑,图像的细节信息将被保留,而由于 CNN 的空间不变性,这些信息是 CNN 结构所未能考虑的。

Zheng 等人提出的 CRFsRNN 模型^[24]则用 CRF 来优化 FCN 网络的分割结果。其主要贡献是将密集 CRF 转换为带有成对势能的形式,并作为网络的组成部分加入了 FCN 网络,通过平均场近似,将其视为 RNN 结构,该工作成功地将 CRF 与 RNN 整合在一起成为一个完整的端对端的网络。这篇论文说明可以将 CRF 转换为 RNN,并将其集成到 FCN 模型中,实现对全局信息的建模。

1.2.2 带孔卷积

带孔卷积又称扩张卷积,是对采用 Kronecker 积的卷积核的拓展,它可以指数级地扩大感受野而不增加卷积核参数数量。带孔卷积是常规的利用上采样滤波的方法,它的扩张率控制着上采样因子,扩张率分别为 0、1、2。这意味着扩张卷积可以在任意分辨率图片上高效地提取密集特征。普通的卷积其实是扩张率为 1 时的带孔卷积。

使用扩张卷积的典型成果是 Yu 等人提出的多尺度上下文聚合模型、上文提及的 DeepLab 系列模型,以及实时分割网络 ENet,增加了金字塔池化模块的 PSPNet。扩张卷积使得模型具有更大的感受野,不会过度地对图像特征进行下采样,同时不增加额外的系统开销,使其得到了广泛应用。

1.2.3 多尺度预测

整合上下文知识的另一种做法是使用多尺度预测。传统 CNN 中几乎每个独立参数都会影响到得到的特征映射的大小,换句话说,即便是相似的架构也会对输入图像的像素数量产生较大的影响,影响最后得到的特征映射。这意味着滤波器将会检测特定尺度的特征。另外,深度网络的参数一般都与要解决的问题紧密相关,这也使得模型向不同尺度的扩展变得更难。使用多尺度的网络能解决多尺度特征学习的问题,这种网络一般都是选用多个处理不同尺度的网络,最后将它们的预测结果结合,产生一个唯一的输出。

Raj 等人^[25]提出了全卷积 VGG-16 基础网络的一种多尺度版本,有 2 个路径分支,一个是使用一个浅层的卷积网络在原始分辨率上处理输入,另一个使用全卷积 VGG-16 和一个额外的卷积层是在 2 倍分辨率上处理。第二个分支得到的结果经过上采样后与第一个分支的结果相结合,这个综合结果再经过级联的卷积层,得到最终的输出。最终,这个网络对尺度变换具有更好的鲁棒性。Eigen 等人采取了相似的思路解决这个问题,他们选用了包含 4 个有着相同架构的多尺度 CNN 的网络,以一个从粗糙到精细的尺度序列来逐步地提取特征,优化输出结果。

1.2.4 特征融合

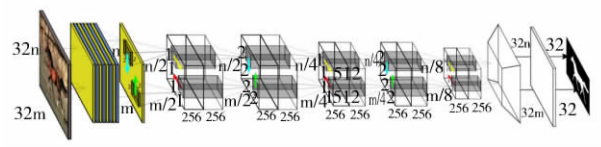
特征融合是向全卷积神经网络架构中加入上下文信息的一种方式,它是将一个较前面的网络层提取到的全局特征与一个相对后边的层提取的局部特征相结合。常见的方法有早融合和晚融合两种。

早融合即先融合多层的特征,然后在融合后的特征上训练预测器。这一方法来自 ParseNet^[26]中的上下文模块。首先,全局特征被上池化为与局部特征相同的尺寸,然后将这两种特征进行串联后得到融合特征,以此得到上下文信息,而后输入到下一层或者直接用于分类器的学习。晚融合是采用跳跃连接的方式对 FCN 网络利用进行延迟特征融合,也是通过将不同层产生的特征映射相结合,类似于在每层上做出独立的预测后再对结果进行融合。

1.2.5 循环神经网络

CNN 网络在非一维数据如图像、视频等处理上取得了成功,然而,这些网络由于依赖于人为设计的卷积核,只能学习到局部上下文信息。而 RNN 得益于其拓扑结构,在对序列的建模上取得了成功。通过将图像特征作为序列输入,可以较好地全局上下文信息建模并改善语义分割结果。

Visin 等人^[27]提出了 ReSeg 模型用于语义分割,如图 6 所示。输入图像在第一层 VGG-16 层中被处理,特征映射结果输入一个或更多的后续层中进行微调。最终,使用的是基于转置卷积的上采样层调整特征映射的尺寸。在本方法中,采用门控循环单元(GRU) 来平衡占用空间与计算复杂度。一般的 RNN 在建模长期依赖关系时表现不好,主要因为梯度消失问题的存在。而长短期记忆网络(LSTM) 和 GRU 是该领域目前较好的两种方法,可以避免梯度消失问题。

图 6 ReSeg 网络示意图^[27]Fig.6 Schematic diagram of ReSeg network^[27]

另一种捕获全局信息的方法是更大的输入窗口的使用,它可以建模更大范围内的上下文信息,但同时也降低了图像的分辨率,而且带来了其他类似于窗口重叠等的问题。然而,Pinheiro 等人^[28]提出了循环卷积神经网络(RCNN)来使用不同的窗口大小循环地训练,相当于考虑了之前网络层中的预测信息。通过这种方法预测出的标签将自动地平滑,从而使网络表现更好。

表 1 典型方法在不同数据集上的表现
Tab.1 Performance of typical methods on different data sets

模型	数据集						
	PASCAL VOC 2012 (mIOU)	PASCAL-cOTENT (mIOU)	COCO2017 (AP)	CamVid	Citys cape (mIOU)	SUN-3D	Stanford-2D-3D-S
FCN	62.2	39.1			65.3		
Bayesian SegNet	60.5			63.1			
Multi-scale CNN-Eigen	62.6						
CRFasRNN	74.7	39.3			62.5		
Deeplab v1	79.7	45.7			70.4		
Deeplab v3	86.9				80.3		
Deeplab v3+	89.0				82.1		
Mask-RCNN			41.8				
PSPNet	85.4				80.2		
PointNet							47.71
LSTM-CF						58.5	
Clockwork Convnet					64.4		

1.3.2 结果分析

结果可以归纳为两点:一是由于有些论文中的方法没有描述其实验的配置,使得完全复现存在困难;二是很多方法缺少对时间开销与占用内存的评价。多数方法关注准确率而忽视了时间与空间效率。然而在实际应用时,这些问题是很重要的。多数算法将被应用于移动设备上,如自动驾驶汽车、无人机、微型机器人等,这些应用都受限于计算消耗以及内存空间。

从分割精度角度看,目前 DeepLab 系列方法精度较高,几乎在每个二维图像数据集上超过了其他方法。2.5 维数据集上,循环的网络如 LSTM-CF 精度较高。三维数据的分割问题刚刚起步,PointNet 为解决无序点云的分割问题开辟了道路。最后,视频序列同样是一个待深入研究的领域。

2 未来的发展趋势

作为多学科交叉的领域,特别是人工智能技术的引入,图像分割得到了由简单粗糙到精细复杂、由人工设计到自主感知的跨越式发展,总的目标是更快速、更精确、更便携。由于基于深度学习的图像分割主要依赖数据样本以及神经网络,所以未来的工作主要在这两个方面展开。

1.3 不同方法的对比与分析

1.3.1 典型方法在不同数据集上的对比

表 1 给出了提供该数据集上结果的方法在其测试集上的准确率,从原始的 FCN 开始,到复杂的模型如 CRFas-RNN 以及目前最优的 DeepLab v3+,图像语义分割的准确率呈现不断上升的清晰趋势。

2.1 数据集方面

1) 三维数据集。总体呈现出向 2.5 维和三维方向发展的趋势。虽然利用三维数据的一些方法如 PointNet 等已经出现,但真实生活场景中的应用需求需要更多更大更通用的数据集出现。

2) 序列数据集。目前仅有少数数据集是基于图像序列的,这些数据集对于利用时间序列信息的方法的发展很有利,二维和三维数据联系起来必将产生新的研究热点。另外,在实际应用场景下,能够达到常见的摄像机 25 帧帧率每秒的处理速度与算法精度同等重要。多数方法达不到这个帧率,比如,FCN-8 处理 1 张低分辨率的图像就需要 100 ms。因此,未来必然更加关注实时分割的质量,新的算法将必须在准确率与时间开销之间寻求平衡。

3) 数据增强与迁移学习。数据增强技术已被证明有利于加速训练收敛,避免过拟合并增强模型的泛化能力。比如,针对医学影像分割提出 U-net 方法的 Ronneberger 等人发现给训练样本随机的弹性变形是小数据集训练成功的关键之一。从头训练一个神经网络通常是不可行的,因为训练需要庞大的数据集,且深度网络收敛也需要很长的时间。Yosinski 等人^[29]证明了从较不相关的任务中迁移学习来的特征也要比直接从随机初始化学习的特征要好。由于收集和创建像素级别的图像分割数据集

比较困难,所以这些数据集的规模一般不如图像分类数据集大。而在 RGB-D 或三维数据分割上,这类数据集规模更小。因此,迁移学习将会是分割领域的大势所趋。

4) 无监督学习。大数据时代数据来源广泛,数据量也呈指数级增长,但事实证明许多数据是无效的:某些场景的数据量不足,或者数据质量不高,还可能只有部分被贴上标签或者标签贴得不合适,这些数据无法满足特定的需求。另外,由于使用小数据集训练复杂深度神经网络时,可能会导致模型过拟合,而人工标注数据又是代价高昂的,因此,使用无标注的无监督学习方法是深度学习的未来。

2.2 深度神经网络方面

1) 图卷积网络(GCN)。卷积神经网络处理的数据通常是二维有规则结构的数据,但现实生活中有很多数据不具备规则的空间结构,比如推荐系统、电子交易、计算几何、脑信号、分子结构等抽象出的图谱。这些图谱结构每个节点连接都不尽相同,不同的节点有不同的连接,是不规则的数据结构。如之前所述,处理三维数据如点云等目前尚未解决,由于点云内在的无序性及非结构性,传统的神经网络如 CNN 等不能直接应用,除非使用某种离散化手段使其结构化。一个研究方向是将点云处理为图,使图模型网络化,融入卷积神经网络进行训练。

2) 融合上下文知识。虽然 FCN 是语义分割的一种基础方法,但是 FCN 网络缺乏对于上下文等特征的建模,而上下文知识对分割结果的提高有很大影响。将 CRF 转换为 RNN 模块融入 FCN 网络等方法建立了一个端对端的解决方案,可以提高算法在真实生活场景下的性能,另外多尺度及特征融合方法也取得了较大的进展。总而言之,这些方法已经取得了不小的进步,但是仍存在进一步研究的空间。

3) 模型移动化与轻量化。用于图像分割的深度神经网络一般需要较大的存储空间,而某些小型平台如手机、微型机器人等受限于其平台空间,为了适应这些应用场景,深度网络规模不能过大。一方面,可以通过研究适合移动设备和相应 APP 的小型神经网络实现,还可以通过基于现有模型的简化来实现。在模型简化方面,剪枝(Pruning)是热门的研究方向,它的作用是简化网络,使得网络在保留多数信息的同时减少参数,同时保留了原网络的准确率。相信未来随着 5 G 网络的普及,深度神经网络将会成为手机等移动载体的标配。

3 结束语

基于深度学习方法的图像语义分割方法使人们能够高效、准确地获取到所需要的图像语义信息,并且实现计算机对图像的自主理解,具有巨大的研究意义和实用价值。在当下信息爆炸的时代,信息的获取越来越容易,而利用深度学习方法实现对图像、视频等视觉大数据的语义分析研究正处于初级阶段,仍然存在很多挑战和难题需要解决,本文提供的基于深度学习的图像语义分割方

法综述将为该领域学习者和研究人员提供解决思路和参考价值。

参考文献:

- [1] ZAITOUN N M, AQEL M J. Survey on image segmentation techniques [J]. Procedia Computer Science, 2015, 65: 797-806.
- [2] OTSU N. A Threshold selection method from Gray-Level Histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [3] SANDHU P S, JUNEJA M, WALIA E. Comparative analysis of edge detection techniques for extracting refined boundaries [C]// International Conference on Machine Learning and Computing(icmlc. 2009.
- [4] ADAMS R, BISCHOF L. Seeded region growing [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 16(6): 641-647.
- [5] REN X, MALIK J. Learning a classification model for segmentation [J]. Iccv, 2003, 1: 10-17 vol.1.
- [6] GOMEZ D, YANEZ J, GUADA C, et al. Fuzzy image segmentation based upon Hierarchical Clustering [J]. Knowledge Based Systems, 2015, 87(10): 26-37.
- [7] 赵静, 黄国满. 基于模糊集理论的彩色图像融合分割 [J]. 测绘与空间地理信息, 2018, 41(8): 107-110, 117.
- [8] 邓莉琼, 张贵新, 郝向宁. 基于知识图谱的图像语义分析技术及应用研究 [J]. 计算机科学与应用, 2018, 8(9): 1364-1371.
- [9] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The kitti vision benchmark suite [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012.
- [10] 侯永宏, 叶秀峰, 张亮, 等. 基于深度学习的无人机人机交互系统 [J]. 天津大学学报: 自然科学与工程技术版, 2017, 50(9): 967-974.
- [11] 马乐乐, 李照洋, 董嘉蓉, 等. 基于计算机视觉及深度学习的无人机手势控制系统 [J]. 计算机工程与科学, 2018, 40(05): 872-879.
- [12] 王弈, 李传富. 人工智能方法在医学图像处理中的研究新进展 [J]. 中国医学物理学杂志, 2013, 30(3): 4138-4143.
- [13] WAN J, WANG D, HOI S C H, et al. Deep learning for content-based image retrieval: a comprehensive study [C]// Proceedings of the ACM International Conference on Multimedia. ACM, 2014.
- [14] ABDI L, MEDDEB A. Driver information system: a combination of augmented reality and deep learning [C]// Symposium on Applied Computing. 2017.
- [15] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,

- 2014, 39(4): 640–651.
- [16] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder – decoder architecture for scene segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017: 1–1.
- [17] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Simultaneous detection and segmentation [C]//European Conference on Computer Vision. Springer, Cham, 2014: 297–312.
- [18] PINHEIRO P O O, COLLOBERT R, Dollár P. Learning to segment object candidates [C]//Advances in Neural Information Processing Systems. 2015: 1990–1998.
- [19] KAIMING H, GEORGIA G, PIOTR D, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018: 1–1.
- [20] HUANG J, YOU S. Point cloud labeling using 3D Convolutional neural network [C]// International Conference on Pattern Recognition. IEEE, 2016.
- [21] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652–660.
- [22] SHELHAMER E. Clockwork convnets for video semantic segmentation [M]. European Conference on Computer Vision. Springer, Cham, 2016.
- [23] 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 & Machine Intelligence, 2018, 40(4): 834–848.
- [24] ZHENG S, JAYASUMANA S, ROMERA P B, et al. Conditional random fields as recurrent neural networks [C]// Proceedings of the IEEE international conference on computer vision. 2015: 1 529–1 537.
- [25] RAJ A, MATURANA D, SCHERER S. Multi-scale convolutional architecture for semantic segmentation [J]. Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RITR-15-21, 2015.
- [26] LIU W, RABINOVICH A, BERG A C. Parsenet: looking wider to see better [J]. arXiv preprint arXiv: 1506.04579, 2015.
- [27] VISIN F, CICCONE M, ROMERO A, et al. Reseg: a recurrent neural network-based model for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 41–48.
- [28] PINHEIRO P H O, COLLOBERT R. Recurrent convolutional neural networks for scene labeling [C]//31st International Conference on Machine Learning (ICML). 2014 (CONF).
- [29] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [C]// Advances in neural information processing systems. 2014: 3 320–3 328.

[编辑: 张 曦]

(上接第 118 页)

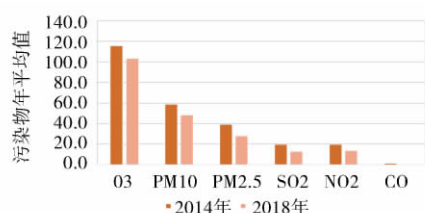


图 6 2014、2018 年污染物浓度年均值

Fig.6 Annual average concentration of pollutants in 2014 and 2018

5 结束语

1) 工业用地面积增长最多,非城市建设用地面积减少最多。

在搬迁改造实施过程中,老工业区炼油厂新建炼油装置,工业用地面积有所增长。与 2014 年相比,工业用地面积增长了 2.5%,非城市建设用地面积减少了 3.2%。

2) 加大重点企业整治力度,环境污染逐步改善。

随着当地政府对污染以及重点皮革、石化企业的整治,2014—2018 年,老工业区空气质量逐年改善,2018 年

空气质量一级和二级的比例较 2014 年上升了 8 个百分点,污染物浓度逐年下降。

参考文献:

- [1] 国家改革和发展委员会振兴司.茂名市某老工业区搬迁改造工作纪实 [Z].北京: 国家改革和发展委员会振兴司,2016.
- [2] 自然资源部.基础性地理国情监测内容与指标: GQJC 03—2018 [S].北京: 自然资源部,2018.
- [3] 国家测绘地理信息局.基础性地理国情监测数据技术规范: GQJC 01—2018 [S].北京: 国家测绘地理信息局,2018.
- [4] 国家基础地理信息中心.地理国情普查内业编辑与整理技术规定: GDPJ 12—2013 [S].北京: 国务院第一次全国地理国情普查领导小组办公室,2013.
- [5] 中国城市规划设计研究院.城市用地分类与规划建设用地标准: GB 50137—2011 [S].北京: 中国计划出版社,2011.
- [6] 欧哲坛,李超群.基于地理国情普查数据的昌江县沿海地表覆盖变化监测 [J].测绘与空间地理信息,2018,41(7),141–144,147.

[编辑: 任亚茹]