

图像语义分割问题研究综述

肖朝霞, 陈 胜

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘 要: 图像语义分割作为 AI 领域的重要分支, 是计算机视觉技术中的重要环节, 同时也是深度学习算法的重要应用。介绍深度学习应用于计算机视觉技术之前图像处理中语义分割传统算法, 并探讨语义分割问题中基于不同神经网络结构模型的深度学习算法及性能, 如 FCN、U-net 算法。最后针对当前图像语义分割算法应用, 总结展望未来研究方向。

关键词: 深度学习; 图像处理; 神经网络; U-net 网络

DOI: 10.11907/rjdk.173172

中图分类号: TP301

文献标识码: A

文章编号: 1672-7800(2018)008-0006-03

Review of Image Semantic Segmentation

XIAO Zhao-xia, CHEN Sheng

(School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: As an important branch of AI field, image semantic segmentation is an important part of computer vision technology and an important application of deep learning algorithm. This paper introduces the traditional algorithm of semantic segmentation applied in image processing prior to the application of deep learning in computer vision, and discusses the deep learning algorithm and its performance based on different neural network structure models in semantic segmentation, such as FCN and U-net algorithm. Finally, aiming at the application of the current image semantic segmentation algorithm, the paper summarizes the research direction.

Key Words: deep learning; image processing; neural network; U-net network

0 引言

图像语义分割是计算机视觉的关键技术, 与图像自动标注作为大规模图像处理和理解的两大重要任务^[1]。在机器视觉领域, 随机森林分类与纹理基元森林分类是图像语义分割的传统作法。最初的分割大都基于简单的像素级别的“阈值法”, 随着分割技术的不断改进, 基于“图划分”的分割法改善效果明显, 精度较高, 成为经典的传统语义分割方法之一^[2-3]。

在深度学习算法被创造性地引入机器视觉领域后, 语义分割问题有了突破性进展, 如全卷积神经网络 (fully convolutional networks, FCN)、基于多图学习与块对角约

束的全监督语义分割方法、基于模型评估的弱监督语义分割方法^[4]等许多基于神经网络训练的语义分割方法相继出现, 分割精度不断提高。

1 传统图像语义分割算法介绍

在传统的视觉算法中, 图像语义分割的经典作法有基于图划分的“Normalized cut”算法和“Grab cut”算法等。

1.1 Normalized Cut 算法

“Normalized cut”方法, 简称“N-cut”, 是基于图划分的语义分割方法之一^[5]。图划分是指将图像建模为图 (Graph) 的形式, 借助图理论 (Graph theory) 对图像进行语义分割。最小分割算法 (Min-cut algorithm) 就是其中较

收稿日期: 2017-12-12

基金项目: 国家自然科学基金项目 (201361235636); 上海市自然科学基金项目 (13ZR1410400); 沪江基金项目 (C14002)

作者简介: 肖朝霞 (1994-), 女, 上海理工大学光电信息与计算机工程学院硕士研究生, 研究方向为深度学习、生物医学图像处理; 陈胜 (1976-), 男, 博士, 上海理工大学光电信息与计算机工程学院副教授、硕士生导师, 研究方向为医学图像处理与分析、模式识别、计算机辅助诊断。

为经典的代表,如图 1 所示。

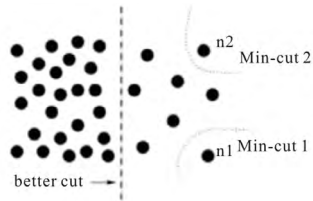


图 1 Min-cut 算法

Min-cut 算法不足之处在于只考虑局部信息,分割精度不高,为此 N-cut 提出一种考虑全局信息的方法进行图划分(Graph partitioning)^[6]。N-cut 算法将全图节点的权重($\text{assoc}(A, V)$ 和 $\text{assoc}(B, V)$)加在两个分割部分 A、B 中,使在划分离散点时,其中某一权重值接近于 1,从而达到考虑全局信息而摒弃划分离群点的目的。正是这样的算法改进,使得 N-cut 既能处理二类语义分割,还能将二分图扩展为 K 路图划分,即完成多语义的图像语义分割^[7],如图 2 所示。

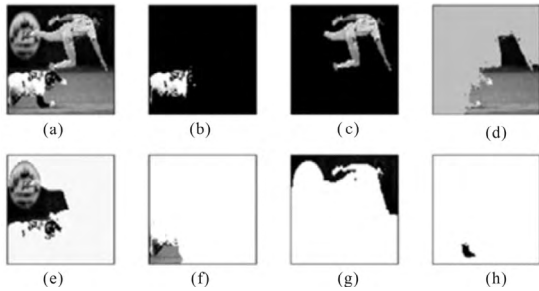


图 2 N-cut 算法多语义分割

1.2 Grab cut 算法

Grab cut 是一种经典的交互式图像语义分割方法^[8],是 N-cut 算法的改进,采用迭代式语法分割。在图像处理问题上,它利用传统的图像纹理信息及边界信息,借助用户少量交互操作获取前后背景的分割结果。Grab cut 算法可以很好地分割 RGB 图像,算法利用两个混合高斯模型(Gaussian Mixture Model, GMM)分别对前景和背景建模,还利用吉尔斯能量方程进行全局建模,而后不断迭代,将能量方程达到最优时的参数作为两个 GMM 的最优参数参与整个建模过程,从而使 GMM 准确描述某个像素属于前景或背景的概率。

用户交互式操作主要以包围框和涂写线条两种方式,为 Grab cut 算法提供辅助信息。以图 3 为例,用户在待分割的图像中用方框选中主要物体,Grab cut 就会默认方框中的像素包含主要物体、前景,此后经过图划分迭代求解,最终得到扣图结果。然而这样的分割效果并不是特别理想,它需要人为干预,用红色线条或点标明背景,用白色线条标明前景。只有在这样的基础上,才可能得到理想效果,不能做到完全自动化处理。



图 3 Grab cut 用户交互操作

2 基于深度学习的语义分割算法

自从深度学习引入机器视觉领域后,一系列基于不同神经网络训练的语义分割方法相继提出,使得分割精度不断提高。本文介绍在语义分割领域 3 种基于深度学习的代表性作法。

2.1 FCN 网络

全卷积神经网络(FCN)是目前语义分割最常用的网络,也是深度学习在图像语义分割任务上的开创性工作^[9]。深度卷积神经网络模型(CNN)是图像级语义理解的利器,而 FCN 则是基于 CNN 实现的像素级别的语义理解,适用于图像语义分割、边缘检测等应用场景^[10]。FCN 可以接受任意尺寸的输入图像,采用反卷积层对最后一个卷积层的特征图进行上采样,使它恢复到与输入图像相同的尺寸,从而可以对每个像素都产生一个预测,同时保留原始输入图像中的空间信息,最后在上采样的特征图上进行逐像素分类 FCN 的网络结构,如图 4 所示。

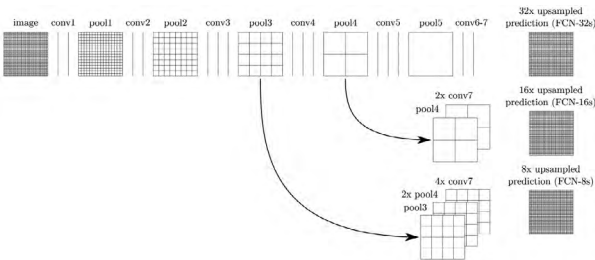


图 4 FCN 网络结构

在池化层下采样之后,原图的长和宽远大于响应张量的长和宽,就像素级别的直接训练而言,丢失了很多图像细节,增加了预测困难。针对这一问题,一方面 FCN 利用双线性插值采样到原图大小,另一方面 FCN 将 Pool4 和 Pool3 的响应结果,分别作为模型 FCN-16s 和 FCN-8s 的输出,与原来 FCN-32s 的输出结合在一起作最终的语义分割预测,如图 5 所示。

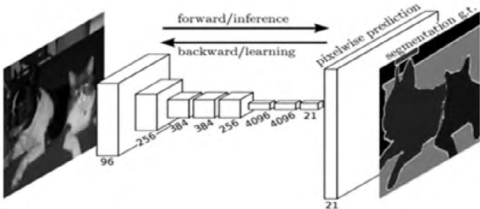


图 5 FCN 语义分割预测

图 6 是不同层作为输出语义分割结果的对比^[11],可以看出池化层下采样倍数的不同导致了不同语义分割精细程度。其中由于是最后一层卷积和池化的输出,该模型的 FCN-32s 下采样倍数最高,其对应的语义分割结果最为粗略;而 FCN-8s 则因下采样倍数较小可以取得较为精细的分割结果。

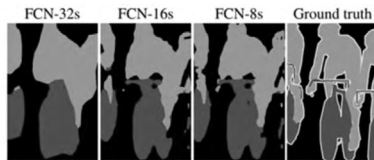


图 6 FCN 语义分割结果

FCN 训练比较复杂,需要训练 3 次才能得到 FCN-8s 分割效果,对图像的细节部分不够敏感。主要是因为上采样过程仅是一个简单的去卷积过程,在解码时,输入上采样层的标签图太稀疏;其次是对各图像像素在进行分类时,并未考虑像素间的关系,忽略了像素级别分类方法中的空间规整步骤,缺乏空间一致性。

2.2 U-net 网络

U-net 是一种基于 FCN 的语义分割网络,适用于医学图像分割^[12]。网络是一种编码器-解码器结构。编码器逐渐减少池化层的空间维度,解码器逐步修复物体的细节和空间维度。编码器和解码器之间通常存在快捷连接,因此能帮助解码器更好修复图像目标的细节。U-net 的结构如图 7 所示。

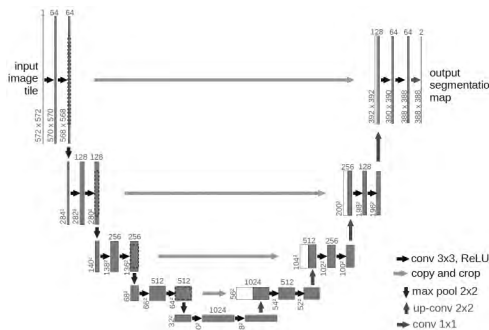


图 7 U-net 网络结构

U-net 作为一个全卷积网络,没有全连接层,输入、输出的内容都是图像。网络具有不同的分辨率层,其中较浅的高分辨率层决定像素定位,较深的则用于像素分类^[13]。20 个左右的卷积层经过 4 次上采样、4 次下采样。输入图像大于输出图像,对输入图像进行镜像操作。如图 8 所示。

和 FCN 相比,结构上比较大的改动在上采样阶段,上采样层包括多层特征。此外,U-net 只需要一次训练,而 FCN 需要 3 次训练。

2.3 SegNet 网络

SegNet 是一个编码器-解码器结构的卷积神经网络,是基于 FCN、修改后的 VGG-16 网络得到的语义分割网络^[14]。SegNet 的网络结构如图 9 所示。

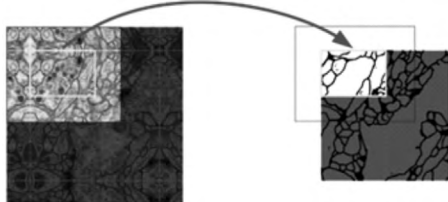


图 8 镜像操作

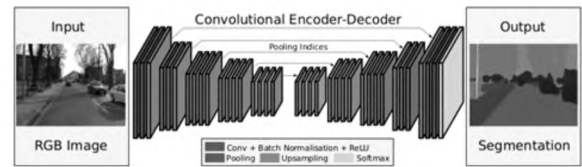


图 9 SegNet 网络结构

SegNet 和 FCN 最大区别是解码的上采样方法,在图 9 的结构中,编码器中每一个池化层都保存了池化指数,并且将其传递到后面对称的上采样层具体的上采样过程如图 10^[15]所示。

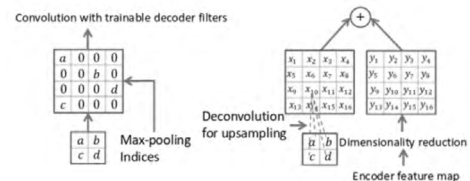


图 10 SegNet 与 FCN

文献^[16]中关于道路场景和室内场景的 SegNet 预测,如图 11 所示。



图 11 SegNet 预测

3 结语

基于深度学习的图像语义分割技术虽然与传统方法相比,取得了很好的分割效果,但是对数据标注的要求过高:不仅需要海量图像数据,而且这些图像还需提供精确到像素级别的标记信息。因此,越来越多的研究者开始将注意力转移到弱监督(weakly-supervised)条件下的图像语义分割上。弱监督条件下的图像语义分割可以作为深入研究的方向。

参考文献:

- [1] 曹攀,钱军浩,陈智,等. 基于图像分层树的图像语义分割方法[J/OL]. 计算机应用研究:1-11. [2017-07-21]. <http://www.aocmag.com/article/02-2018-08-039.html>.
- [2] ROTHER C, KOLMOGOROV V, BLAKE A. "GrabCut" - interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics, 2004.
- [3] 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.
- [4] 张珂. 图像语义分割算法研究[D]. 上海:复旦大学, 2014.

(下转第 12 页)

- Applied Mathematics, 1963, 11(2): 431-441.
- [16] NIELSEN H B. Damping parameter in marquardt's method [EB/OL]. http://www.imm.dtu.dk/documents/ftp/tr99/tr05_99.pdf.
- [17] PRESS W H, TEUKOSKY S A, VETTERLING W T, et al. Numerical recipes in C[M]. United Kingdom: Cambridge University Press, 1992.
- [18] TRANSTRUM M K, SETHNA J P. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization [EB/OL]. <https://arxiv.org/abs/1201.5885>.
- [19] MARK K T, BENJAMIN B M, JAMES P S. Why are nonlinear fits to data so challenging[J]. Physical Review Letters, 2010, 104(6): 1-4.
- [20] 陈德豪, 张琰. 常用阻尼最小二乘算法的改进[J]. 物测科技, 1994(1): 31-52.
- [21] 孙若昧, 郑斯华, 马林, 等. 阻尼最小二乘法联合测定震源位置和介质速度参数[J]. 地震, 1986(4): 29-37.
- [22] 何宗海. 用阻尼最小二乘法研究云南地区的 Q 值分布[J]. 西北大学学报: 自然科学版, 1995, 25(3): 237-240.
- [23] 宋林平. 阻尼加权最小二乘地震走时层析成像[J]. 计算物理, 1995, 12(4): 499-504.
- [24] KALACHAND, 戴成泰. 运用阻尼最小二乘法的广角地震反射时间反演[J]. 石油物探译丛, 1995(5): 27-37.
- [25] 阮百尧, 葛为中. 奇异值分解法与阻尼最小二乘法的对比[J]. 物探化探计算技术, 1997, 19(1): 47-49.
- [26] 赵龙, 刘淮, 陈哲. 递推阻尼最小二乘及其在 INS/双星组合中的应用[J]. 北京航空航天大学学报, 2003, 29(12): 1136-1138.
- [27] 王进, 李欣然, 苏盛. 用阻尼最小二乘优化算法实现电力系统综合静态负荷建模[J]. 长沙电力学院学报: 自然科学版, 2004, 19(1): 40-46.
- [28] 孙明玮, 张奇, 邵继法, 等. 鲁棒递推阻尼最小二乘算法[J]. 航空计算技术, 2003, 33(1): 22-25.
- [29] 张国芹, 朱长青, 王光霞. GIS 数据误差处理的阻尼最小二乘算法[J]. 测绘学院学报, 2004, 21(1): 8-10.
- [30] 马英杰, 虎胆·吐马尔拜, 沈冰. 利用阻尼最小二乘法求解 Van Genuchten 方程参数[J]. 农业工程学报, 2005, 21(8): 179-180.
- [31] JOSE P. The solution of nonlinear inverse problems and the Levenberg-Marquardt method[J]. Geophysics, 2007, 72(4): 1-16.
- [32] 孙秋超. 瞬变电磁数据改进阻尼最小二乘拟合算法研究[D]. 长春: 吉林大学, 2009.
- [33] 李培. 基于奇异值分解的瑞雷面波加权阻尼最小二乘反演[D]. 长沙: 中南大学, 2011.
- [34] 张皓, 刘建松, 黄金辉, 等. 最优化阻尼最小二乘法重力二维反演方法及其应用[J]. 中国西部科技, 2011, 11(31): 34-36.
- [35] 王晟, 胡志云, 张振荣, 等. 阻尼最小二乘算法 CARS 光谱温度拟合[J]. 强激光与粒子束, 2012, 12(11): 2565-2570.
- [36] 张亚兵, 陈同俊, 崔若飞, 等. 基于阻尼最小二乘法的煤层裂隙 P 波方位属性预测[J]. 煤田地质与勘探, 2012, 40(4): 79-81, 85.
- [37] 王亚强. 改进的阻尼最小二乘层析算法及影响因素分析[D]. 秦皇岛: 燕山大学, 2013.
- [38] 王园园, 刘斌, 王晨. 基于阻尼最小二乘法的瞬变电磁反演算法研究[J]. 电子测试, 2013(1): 1-7.
- [39] 王寅, 赵南京, 刘文清, 等. 阻尼最小二乘法用于激光诱导击穿光谱重叠特征谱线分离提取[J]. 光谱学与光谱分析, 2015, 35(2): 309-314.
- [40] 陈诚. 三轴天线座阻尼最小二乘跟踪策略研究[J]. 现代雷达, 2015, 37(7): 44-47.
- [41] 平晶晶. 基于阻尼最小二乘法的表面等离子激元定向耦合器的优化[D]. 南京: 南京航空航天大学, 2016.
- [42] 陈恒, 严良俊, 代小威, 等. 模拟退火法与阻尼最小二乘法联合反演岩石激电谱参数[J]. 工程地球物理学报, 2016, 13(2): 170-174.
- [43] 项伟, 白征东, 汤晓禹. 阻尼最小二乘法在任意欧拉角坐标转换中的应用[J]. 大地测量与地球动力学, 2016, 36(2): 167-170.
- [44] 林茂琼, 陈增强, 袁著祉. 基于阻尼最小二乘法的鲁棒自校正预测控制器[J]. 南开大学学报: 自然科学版, 1998, 31(2): 79-83.
- [45] 林茂琼, 陈增强, 贺江峰, 等. 基于阻尼最小二乘法的神经网络自校正一步预测控制器[J]. 控制与决策, 1999, 14(2): 165-168.
- [46] 王阿霞, 张文波. 阻尼最小二乘法在射线追踪中的应用[J]. 西安工程学院学报, 2001, 23(1): 43-45.

(责任编辑: 何 丽)

(上接第 8 页)

- [5] 程国建, 刘丽婷. 深度学习算法应用于岩石图像处理的可行性研究[J]. 软件导刊, 2017, 16(2): 54-56.
- [6] 王启冬. 基于数学规划的图划分模型研究[D]. 大连: 大连理工大学, 2009.
- [7] SHI J B, MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.
- [8] WU S Q. Automatic grab cut based lung extraction from endoscopic images with an initial boundary[A]. IEEE 13th International Conference on Signal Processing, 2017: 1374-1378.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Boston: IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [10] 相文波. 基于卷积神经网络的肺炎类型影像判别[D]. 哈尔滨: 哈尔滨理工大学, 2017.
- [11] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, et al. Conditional random fields as recurrent neural networks[C]. San-tiago: International Conference on Computer Vision, 2015.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]. Munich: International Conference on Medical Image Computing and Computer-assisted Intervention, 2015.
- [13] 楚恒. 像素级图像融合及其关键技术研究[D]. 成都: 电子科技大学, 2008.
- [14] 张明月. 基于深度学习的图像分割研究[D]. 长春: 吉林大学, 2017.
- [15] 翁健. 基于全卷积神经网络的全向场景分割研究与算法实现[D]. 济南: 山东大学, 2017.
- [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, 2015(99): 2481-2495.

(责任编辑: 江 艳)