第66卷 第6期
2020年12月

武汉大学学报(理学版)
J. Wuhan Univ. (Nat. Sci. Ed.)

Vol. 66 No. 6
Dec. 2020,519~531

DOI: 10.14188/j.1671-8836.2019.0002

图像分割方法综述

黄 鹏[1],郑 淇[1],梁 超[1,2†]

1.武汉大学 计算机学院,湖北 武汉 430072;

2.武汉大学 深圳研究院,广东 深圳 518000

收稿日期:2019-01-04 † 通信联系人 E-mail:cliang@whu.edu.cn

基金项目:国家自然科学基金(61876135,U1903214,61862015);湖北省自然科学基金(2019CFB472,2018AAA062,2018CFA024);
深圳市科技计划基础研究项目(JCYJ20170818143246278)

作者简介:黄 鹏,男,硕士生,现从事计算机视觉、模式识别与多媒体内容分析方面的研究。E-mail:2017282110180@whu.edu.cn

摘 要: 为了解图像分割领域的研究现状,对图像分割方法进行了系统性梳理,首先按照基于阈值、边缘、区域、聚类、图论及特定理论等6类方法介绍传统图像分割方法;然后介绍基于深度学习的分割方法,并探讨了几种常用的分割网络模型,包括全卷积网络(full convolutional network,FCN)、金字塔场景解析网络(pyramid scene parsing network,PSPNet)、DeepLab、Mask R-CNN;最后在图像分割的常用数据集上对同类方法进行了性能比较和分析。

关 键 词: 图像处理;图像分割;深度学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 1671-8836(2020)06-0519-13

Overview of Image Segmentation Methods

HUANG Peng[1], ZHENG Qi[1], LIANG Chao[1,2†]

1. School of Computer Science, Wuhan University, Wuhan 430072, Hubei, China;

2. Shenzhen Institute of Wuhan University, Shenzhen 518000, Guangzhou, China

Abstract: In order to understand the current research status in the field of image segmentation, the image segmentation methods are systematically sorted out. Firstly, traditional image segmentation methods are introduced according to 6 types of methods based on thresholds, edges, regions, clusters, graph theory, and specific theories. Then the segmentation methods based on deep learning are introduced, and several commonly used segmentation network models are discussed, including full convolutional network (FCN), pyramid scene parsing network (PSPNet), DeepLab, and Mask R-CNN. Finally, the performance comparison and analysis of similar methods are performed on the commonly used datasets for image segmentation.

Key words: image processing; image segmentation; deep learning

0 引 言

随着计算机技术的飞速发展,计算机视觉逐渐细化形成了自己的科学体系,其中图像分割作为图像处理领域的重要分支,起着越来越重要的作用。图像分割是指将图像划分成互不相交的、有意义的子区域,在同一个区域的像素点具有一定的相关性,不同区域的像素点存在一定的差异性,即是对图片中有相同性质的像素赋予相同标签的过程。

在智能安防、无人驾驶、卫星遥感、医学影像处理、生物特征识别等领域,图像分割可以提供精简且可靠的图像特征信息,进而有效地提高后续视觉任务的处理效率,具有重要意义。在实际应用过程中,根据应用场景的不同,需灵活地采用不同的图像分割方法,以满足不同分割任务的需求。

最早的图像分割方法应用在医学影像处理领域,

引用格式:黄鹏,郑淇,梁超. 图像分割方法综述[J]. 武汉大学学报(理学版),2020,66(6):519-531. DOI:10.14188/j.1671-8836.2019.0002.
HUANG Peng, ZHENG Qi, LIANG Chao. Overview of Image Segmentation Methods [J]. *J Wuhan Univ* (*Nat Sci Ed*), 2020, 66(6): 519-531. DOI:10.14188/j.1671-8836.2019.0002(Ch).

对影像中的特定目标分割后再进行医疗分析诊断。由于医学影像场景简单,背景和目标区别明显,在该领域中大多是通过简单的基于阈值的方法[1~3]进行粗糙的像素级别的分割。随着分割场景的复杂化,对分割技术的要求也愈加严格,陆续出现了基于边缘[4~9]、区域[10~12]、聚类[13~17]、图论[18~20]和特定理论[21~27]等的分割方法,分割的效果也因此得到了改善。特别是将深度学习引入到图像处理领域后,赋予了分割区域更准确的语义信息,图像分割问题也取得了突破性的进展,如:全卷积网络(full convolutional network, FCN)[28]、金字塔场景解析网络(pyramid scene parsing network, PSPNet)[29]、DeepLab[30~33]、Mask R-CNN[34]等基于深度学习的图像分割方法的出现,使得分割的准确度不断提高,分割的过程也更加智能化。

本文对图像分割方法进行综述,根据是否引入深度神经网络,将其分为传统图像分割方法和基于深度学习的图像分割方法,并分别对这些方法的性能、效果进行对比。

1 传统图像分割方法

传统图像分割方法是早期的分割手段,它们大多简单有效,经常作为图像处理的预处理步骤,用以获取图像的关键特征信息,提升图像分析的效率。本节将对传统分割方法进行阐述,主要介绍基于阈值、边缘、区域、聚类、图论及特定理论等常用且经典的分割方法,然后使用这些方法对相同的图像进行分割处理。

1.1 基于阈值的图像分割方法

基于阈值的图像分割方法[1~3]实质是通过设定不同的灰度阈值,对图像灰度直方图进行分类,灰度值在同一个灰度范围内的像素认为属于同一类并具有一定相似性,该类方法是一种常用的灰度图像分割方法。

用$f(i,j)$表示原始图像像素(i,j)的灰度值,通过设定阈值T,将图像中的像素分为目标和背景两类,实现输入图像$f(i,j)$到输出图像$g(i,j)$的变换:

$$g(i,j)=\begin{cases}1, & f(i,j)\geqslant T\\ 0, & f(i,j)<T\end{cases} \tag{1}$$

其中,$g(i,j)=1$表示属于目标类别的图像,$g(i,j)=0$表示属于背景类别的图像。

由此可见,基于阈值的图像分割方法的关键是选取合适的灰度阈值,以准确地将图像分割开来。如图1所示,本文针对同一灰度图像(图1(a)),设定不同的灰度阈值(T=80、120、160)分别进行阈值分割,得到不同效果的分割图,如图1(b)~图1(d)所示。由图1可知,阈值T越大,分为目标类别的像素点就越多,图像逐渐由浅变深。

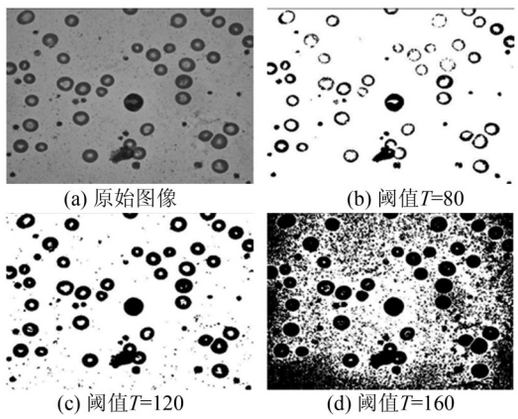


图1 不同阈值情况下的图像分割效果

Fig. 1 Image segmentation effect under different thresholds

对于基于阈值的图像分割方法,根据不同的准则有不同的分类,常见的分类为:基于点的全局阈值分割方法[1]、基于区域的全局阈值分割方法[2]、局部阈值分割方法[3]等。基于阈值的分割方法适用于目标灰度分布均匀、变化小,目标和背景灰度差异较明显的图像,简单易实现且效率高。然而,该类法通常只考虑像素自身的灰度值,未考虑图像的语义、空间等特征信息,且易受噪声影响,对于复杂的图像,阈值分割的效果并不理想。因此,在实际的分割操作中,基于阈值的分割方法通常作为预处理方法或与其他分割方法结合使用。

1.2 基于边缘的图像分割方法

在图像中若某个像素点与相邻像素点的灰度值差异较大,则认为该像素点可能处于边界处。若能检测出这些边界处的像素点,并将它们连接起来,就可形成边缘轮廓,从而将图像划分成不同的区域。

根据处理策略的不同,基于边缘的图像分割方法,可分为串行边缘检测法和并行边缘检测法[4]。串行边缘检测法需先检测出边缘起始点,从起始点出发通过相似性准则搜索并连接相邻边缘点,完成图像边缘的检测;并行边缘检测法则借助空域微分算子,用其模板与图像进行卷积,实现分割。

在实际应用中,并行边缘检测法直接借助微分算子进行卷积实现分割,过程简单快捷,性能相对优良,是最常用的边缘检测法。根据任务的不同,可灵活选择边缘检测算子,实现边缘检测完成分割。常用的边缘检测微分算子有:Roberts[5]、Sobel[6]、Prewitt[7]、LoG[8]、Canny[9]等。如图2所示,

本文分别使用不同的微分算子对相同的图像进行处理。从图2中可以看出,相较于图像背景,经边缘检测算子处理后,水果的边缘轮廓相对清晰,实现了图像分割的目的。

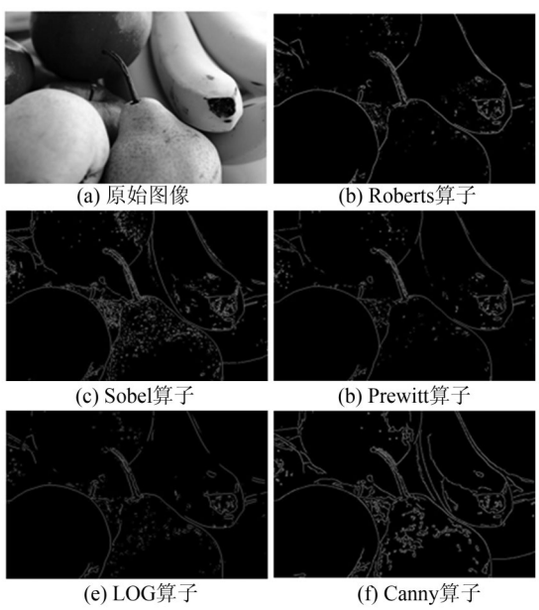


图2　采用不同微分算子时并行边缘检测法的图像分割效果

Fig. 2　Image segmentation effect of parallel edge detection with different differential operators

1.3　基于区域的图像分割方法

基于区域的图像分割方法是根据图像的空间信息进行分割,通过像素的相似性特征对像素点进行分类并构成区域。根据区域思想进行分割的方法有很多[10~12],其中较常用的有区域生长法和分裂合并法[12]。

区域生长法[12]指的是通过将具有相似性质的像素点集合起来,构成独立的区域,以实现分割。具体过程:先选择一组种子点(单个像素或小区域)作为生长起点,然后根据生长准则,将种子点附近与其具有相似特征的像素点归并到种子点所在的像素区域内,再将新像素作为种子点,反复迭代至所有区域停止生长。区域生长法中种子点和生长准则的选取至关重要,直接影响分割效果。种子点的选取除了人工选取法外,还可以用算法自动选取;生长准则可根据图像的颜色、纹理、空间等特征信息设定。

分裂合并法[12]的实质是通过不断地分裂合并,得到图像各子区域。具体步骤为:先将图像划分为规则的区域,然后根据相似性准则,分裂特性不同的区域,合并特性相同的邻近区域,直至没有分裂合并发生。该方法的难点在于初始划分和分裂合并相似性准则的设定。

图3为基于区域的图像分割方法分割效果图。首先对原始图像(图3(a))进行灰度化处理得到灰度图(图3(b)),然后分别用区域生长法和分裂合并法进行分割。区域生长法分割效果如图3(c)所示,该方法计算简单,但对噪声敏感,易导致区域空缺,图中头盔受背景颜色的干扰出现了残缺的现象;分裂合并法分割效果如图3(d)所示,它对复杂图像的分割有较好的效果,但其计算复杂,且分裂时边界可能被破坏,如图3(d)中,车轮的轮廓信息在合并过程中被破坏,导致车轮边缘出现了模糊现象。

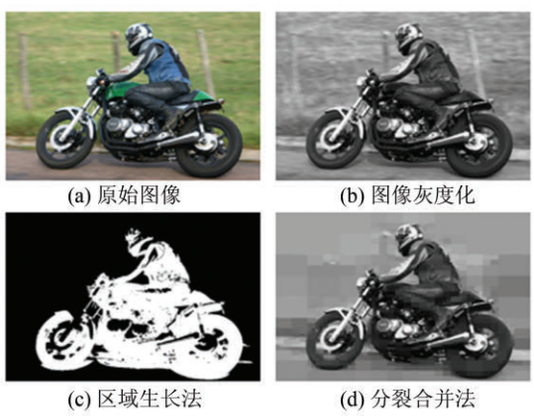


图3　基于区域的图像分割方法分割效果

Fig. 3　Segmentation effect of region-based image segmentation method

1.4　基于聚类的图像分割方法

基于聚类的图像分割方法将具有特征相似性的像素点聚集到同一区域,反复迭代聚类结果至收敛,最终将所有像素点聚集到几个不同的类别中,完成图像区域的划分,从而实现分割。

随着分割任务需求复杂化,聚类分割技术也在不断地发展。1995年,Cheng[13]在原始Mean Shift算法[14]的基础上定义了核函数和权值系数,使Mean Shift算法得到广泛的应用。2007年,Sheikh等[15]提出Medoidshift算法,与Mean shift算法类似,它能自动计算聚类数目,而且数据不必线性可分。2009年,Levinshtein等[16]提出基于几何流的超像素快速生成算法,称为TurboPixels。2012年,Achanta等[17]提出一种通过计算像素点距离和颜色相似度,聚类生成超像素的方法,称为简单线性迭代聚类(simple linear iterative clustering,SLIC)。SLIC适用于图像分割、姿势估计、目标跟踪及识别等计算机视觉应用,是经典的图像处理手段。下面以SLIC算法为例,进行详细介绍。

SLIC算法基于聚类思想,可以将图像中的像素划分为超像素块,因此也被称为超像素分割。该算法步骤如下:

1) 将RGB彩色图像通过映射转化到Lab颜色空间,Lab颜色空间由(L, a, b)三元素组成,其中,L代表亮度,a代表从洋红色至绿色的范围,b表示从黄色至蓝色的范围。相比于RGB空间,Lab空间能够保留更宽的色彩区域,提供更加丰富的色彩特征。

2) 将每个像素点颜色特征(L, a, b)及坐标(x,y)组合成向量(L, a, b, x, y)进行距离度量,包括像素点i和j之间的颜色距离d_c和空间距离d_s,具体公式如下

$$d_c=\sqrt{(l_j-l_i)^2+(a_j-a_i)^2+(b_j-b_i)^2} \tag{2}$$

$$d_s=\sqrt{(x_j-x_i)^2+(y_j-y_i)^2} \tag{3}$$

其中,$l_n(n=i,j)$表示在颜色空间中亮度的特征距离值;a_n、b_n分别表示在颜色空间中色阶品红、正黄系的特征距离值;x_n、y_n分别表示像素点的横、纵坐标值。

再通过D'对最终距离进行度量

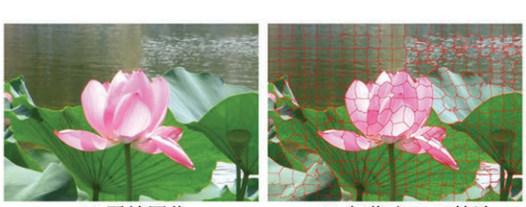
$$D'=\sqrt{\left(\frac{d_c}{N_c}\right)^2+\left(\frac{d_s}{N_s}\right)^2} \tag{4}$$

其中,N_c表示最大颜色距离,通常取常数m($m\in[1,40]$);N_s是类内最大空间距离,$N_s=S=\mathrm{sqrt}(N/K)$,N是图中像素点总数,K为预分割超像素块的总和,超像素块的大小为N/K,相邻种子点的距离为S。

综上,两个像素点之间的距离度量公式可表示为:

$$D'=\sqrt{\left(\frac{d_c}{m}\right)^2+\left(\frac{d_s}{S}\right)^2} \tag{5}$$

超像素SLIC算法中,像素间的相似性由对应(L, a, b, x, y)向量间的距离度量,两个向量的距离越小则对应像素点的性质越相似,反之,则对应像素点的性质相似性越低。根据这个相似性准则,可以对像素点进行聚类,实现图像的超像素分割。基于聚类的图像分割方法利用图像灰度、纹理等特征信息作为聚类准则,将图像分割转化成像素点聚类的问题,性能稳定且鲁棒性好。图4是使用超像素的SLIC算法得到的分割结果。由图4可知,超像素SLIC算法图像根据纹理特征,将图像划分为多个局部小区域,前景目标荷花和荷叶有明显的边缘轮廓信息。



(a) 原始图像 (b) 超像素SLIC算法

图4 超像素SLIC算法图像分割效果

Fig. 4 Image segmentation effect of super pixel SLIC algorithm

1.5 基于图论的图像分割方法

基于图论的图像分割方法将分割问题转换成图的划分,通过对目标函数的最优化求解,完成分割过程,包括:Graph Cut[18]、GrabCut[19]、One Cut[20]等常用算法。其中,Graph Cut算法基于图论的思想,将最小割(min cut)问题应用到图像分割问题中,可以将图像分割为前景和背景,是经典的基于图论的图像分割方法。下文以Graph Cut算法为例,对该类方法做具体介绍。

图5为原始图像映射成图结构后对应的S-T图。如图5所示,先将图像映射为带有权重的无向图$G=(V, E)$,其中,V是顶点的集合,E是边的集合,无向图中的节点对应原图中的像素点,对每个相邻的点进行连接形成边(实线),边的权重代表像素点之间的相似性。除此之外,每个节点还要和终端顶点S和T进行连接形成边(虚线),与S相连的边$R_p(1)$的权重由该节点(像素点)前景目标概率表示,与T相连的边$R_p(0)$的权重由该节点的背景概率表示。这样处理后,在无向图中就会形成两种顶点和边:一种是代表像素点的普通节点以及普通节点彼此相连形成的边;另一种是终端顶点S和T以及连接它和节点的边。

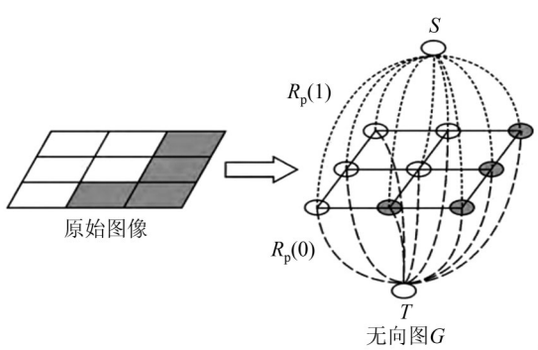


图5 S-T图

Fig. 5 S-T diagram

如果边集合E中的所有边都断开,将会导致S-T图的分开,称之为cut。若一种cut的过程中其对应边的所有权值之和最小,则称之为min cut,对应的能量损失函数最小。至此,将复杂的图像分割问题转化成了求解能量损失函数最小值的问题。通过寻找min cut过程的不断迭代,求得能量损失函数最小值,就可以实现前景目标与背景的分离,从而实现图像分割。如图6所示,使用Graph Cut算法对图片进行分割,可以获取前景目标大致的轮廓,实现目标与背景的分离。

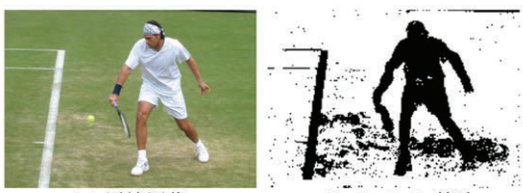


图6　Graph Cut算法图像分割效果

Fig. 6　Image segmentation effect of Graph Cut algorithm

基于图论的Graph Cut算法在利用图像灰度信息的同时使用了区域边界信息,通过最优化求解,得到最好的分割效果。然而,该算法计算量大,且更倾向于对具有相同类内相似度的图像进行分割。

1.6　基于特定理论的图像分割方法

随着分割任务要求及复杂度的提高,图像分割方法也在不断地改进,特别是在新理论和新方法的发展中,针对图像分割任务,出现了很多特定理论和方法。

Matheron[21]提出数学形态学的理论,后由Crespo等[22]将其应用在图像分割中,该方法能克服噪声影响获得清晰的边缘图像;Holland提出了遗传算法[23],并将其引入到多参数下的复杂分割任务中,模拟自然的优胜劣汰获得最优解,实现最优化的分割。除上述方法外,常用的分割理论方法还有:小波变换[24]、活动轮廓模型[25]、模糊理论[26]、粗糙集理论[27]等。由于该类方法涉及的图像分割技术较多,本文不做详细介绍和分割效果展示。

2　传统图像分割方法性能比较

针对不同的分割任务,可选取的分割技术种类繁多,每种方法都有各自的优劣,适用的情况也不尽相同。为此,本文在分割数据集PASCAL VOC[35]、Microsoft COCO[36]上对几种传统图像分割方法的性能进行了实验效果对比,用不同的分割方法对同一图像进行分割处理,图7左、右列图像分别为不同方法对来源于PASCAL VOC和Microsoft COCO数据集中图像的分割结果。其中,图7(a)为原始灰度图像;图7(b)是OTSU阈值法[1]分割效果,对于灰度区别较大的图像,能够明显地分割出前景目标(人和车);图7(c)为采用Canny微分算子[9]的并行边缘检测法分割效果,图像中的边缘轮廓信息比较明显,但也存在很多杂乱的噪声点;图7(d)是区域生长法[12]分割效果,能够分割出前景目标,但是目标区域的细节分割不够精细;图

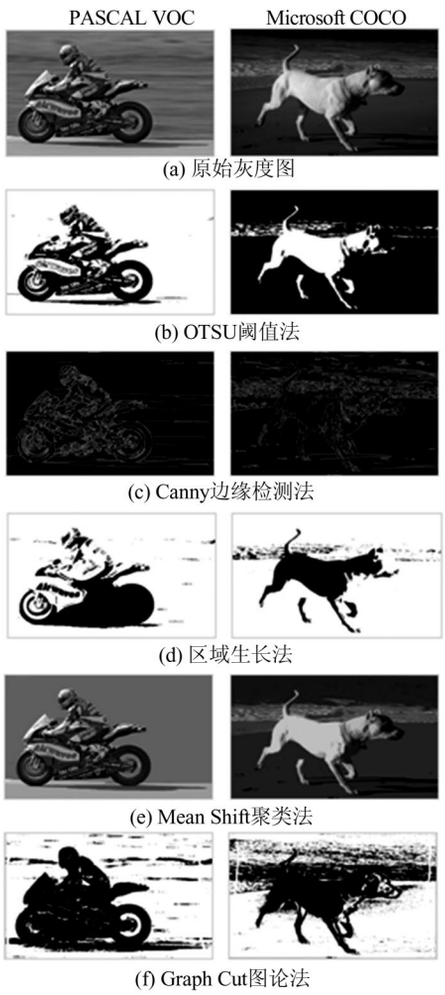


图7　传统分割方法效果

Fig. 7　Effect of traditional segmentation methods

7(e)为基于聚类的Mean Shift算法[14]分割效果,使用该算法可以对目标和背景区域进行各自的聚类,实现分割的效果;图7(f)是基于图论的Graph Cut[18]算法的分割效果,能够明显地区分出前景目标和背景,但也引入了噪声点。对比原始灰度图像(图7(a)),从分割效果图(图7(b)~7(f))中可以看出,传统分割方法在分割时会引入很多无关或者没有意义的阴影和区域,有的甚至会出现噪声点(图7(c)),对分割结果造成干扰,无法精准地区分出前景目标和背景。根据传统分割方法的实验效果,对不同分割方法的适用范围、算法难易程度、执行时间、内存占用大小等方面性能进行比较,结果如表1所示。

由图7和表1所示代表性方法的效果和性能,可归纳对应类别方法的性能。基于阈值的图像分割方法适用于目标与背景灰度差值大的灰度图,简单快捷,但对复杂图像的分割效果并不理想;基于边缘的图像分割方法对像素灰度值具有明显突变的图片处理效果较好,可以直接借助微分算子

表1 几种传统图像分割方法的性能对比

Table 1 Performance comparison of several traditional image segmentation methods

分割方法	适用图像	难易度	耗时	内存
OTSU阈值法[1]	目标与背景灰度差值大	易	短	小
Canny边缘检测法[9]	像素灰度值突变明显	一般	短	小
区域生长法[12]	大部分图像	难	长	大
Mean Shift算法[14]	目标类间性质区别明显	较难	较长	大
Graph Cut算法[18]	大部分图像	较难	较长	较大

获取图像中的轮廓信息,实现高效的分割,但该算法易受噪声影响;基于区域的图像分割方法对大部分图像都适用,但该方法法相对复杂,且时间复杂度高;基于聚类的图像分割方法则对目标类间性质区别明显的图片更适用,应用较广的是超像素分割方法,虽然能实现不错的分割效果,但也存在耗时较长的问题;基于图论的图像分割方法对大部分图像都能进行分割并且可以取得良好的效果,但该方法计算量大,一般需要通过交互实现分割。综上,传统图像分割方法各有利弊,在实际应用中,需要根据分割场景的需求灵活地选择分割算法。

3 基于深度学习的图像分割方法

传统图像分割方法大多利用图像的表层信息,对于需要大量语义信息的分割任务则不适用,无法应对实际的需求。随着深度学习的发展及引入,计算机视觉领域借此取得了突破性进展,卷积神经网络成为了图像处理的重要手段,将其引入到图像分割领域,可以充分利用图像的语义信息,实现图像的语义分割。为应对图像分割场景日益复杂化的挑战,一系列基于深度学习的图像语义分割方法被提出,实现了更加精准且高效的分割,使得图像分割的应用范围得到了进一步的推广。本节将重点介绍4种基于深度学习的经典分割方法,包括:FCN[28]、PSPNet[29]、DeepLab[30~33]、Mask R-CNN[34]。

3.1 FCN

全卷积网络(FCN)[28]是深度学习用于语义分割的开创之作,确立了图像语义分割(即对目标进行像素级别的分类)通用网络模型框架。通常,卷积神经网络(convolutional neural network,CNN)经过多层卷积之后接入若干个全连接层,将卷积层产生的特征图(feature map)映射成固定长度的特征向量进行分类。但FCN与CNN不同,如图8,FCN采用"全卷积"方式,在经过8层卷积处理后,对特征图进行上采样实现反卷积操作,然后通过SoftMax层进行分类,最后输出分割结果。

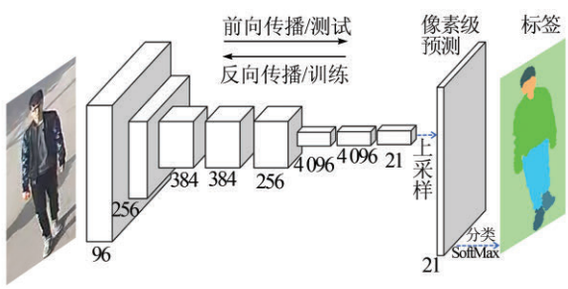


图8 语义分割模型FCN

Fig. 8 Semantic segmentation model FCN

在FCN模型中,由于经过多次卷积操作,特征图的尺寸远小于输入图,且丢失了很多底层的图像信息,如果直接进行分类,会影响分割精度。为此,FCN在上采样过程采用Skip策略。如图9所示,输入图像经过多次卷积(convolution)、池化(pooling),得到不同层级的特征图,将卷积7次后得到的Conv7层上采样后进行分类输出,得到FCN-32s模型的分割结果;将池化4次后得到的Pool4层,与双线性内插法处理后的Conv7层进行融合,上采样后进行分类得到FCN-16s模型的分割结果;将池化3次后得到的Pool3层,与双线性内插法处理后的Conv7层、Pool4层进行融合,上采样后进行分类得到FCN-8s模型的分割结果。通过把深层数据与浅层信息相结合,再恢复到原图的输出,得到更准确的分割结果,根据所利用的池化层的不同,分为FCN-32s、FCN-16s、FCN-8s。图10所示为使用不同FCN模型对同一图像进行分割得到的结果,其中标注图表示标准的分割结果(即真实值)。由图10可知,FCN-8s模型由于整合了更多层的特征信息,相比于FCN-32s和FCN-16s可以分割得到更加清晰的轮廓信息,分割效果相对较好。

FCN能对图像进行像素级别地分类,从而有效地解决了图像语义分割的难题,它可以输入任意尺寸的图像,且是首个端到端的分割网络模型,在分割领域具有重要意义。图11所示为FCN-8s模型

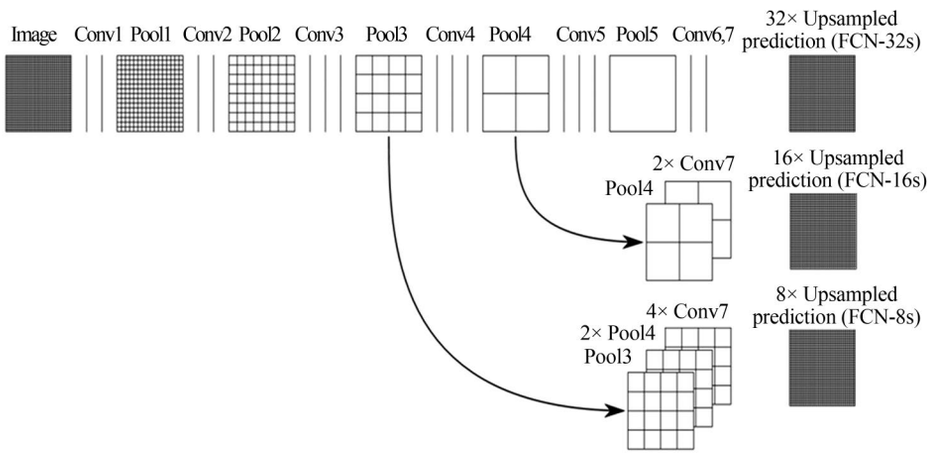


图 9　FCN 结构图[28]

Fig. 9　Structure diagram of FCN[28]

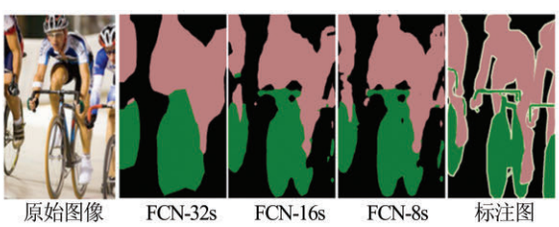


图 10　不同 FCN 模型分割效果[28]

Fig. 10　Segmentation effects of different FCN models[28]

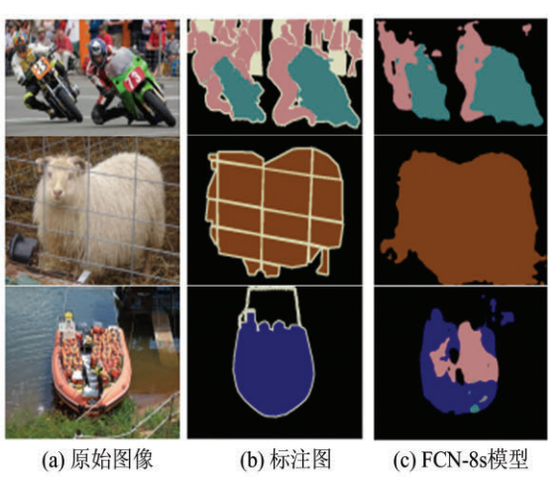


图 11　FCN-8s 模型图像分割效果[28]

Fig. 11　Image segmentation effect of FCN-8s model[28]

对不同类别(人、车、羊、船等)目标的分割效果。实验表明,FCN 的网络相对较大,对图像的细节信息不够敏感,且由于像素点之间的关联性较低,导致目标边界模糊,如图 11(c)所示,前景目标的轮廓分割得不够细致。

3.2　PSPNet

金字塔场景解析网络(PSPNet)[29]整合上下文信息,充分利用全局特征先验知识,对不同场景进行解析,实现对场景目标的语义分割。如图 12 所示,给定输入图像,首先使用 CNN 得到最后一个卷积层的特征图,再用金字塔池化模块(pyramid pooling module)收集不同的子区域特征,并进行上采样(upsample),然后串联(concat)融合各子区域特征以形成包含局部和全局上下文信息的特征表征,最后将得到的特征表征进行卷积和 SoftMax 分类,获得最终的对每个像素的预测结果。

PSPNet 针对场景解析和语义分割任务,能够提取合适的全局特征,利用金字塔池化模块将局部和全局信息融合在一起,并提出了一个适度监督损失的优化策略,在多个数据集上的分割精度都超越了 FCN[28]、DeepLab-v2[31]、DPN[37]、CRF-RNN[38]

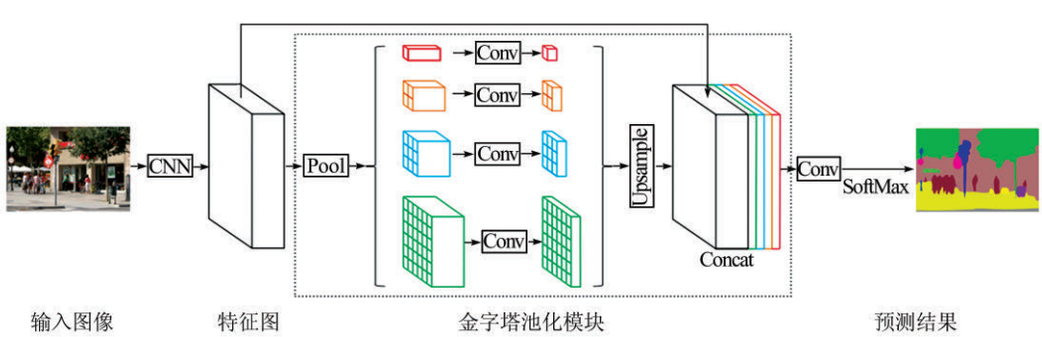


图 12　PSPNet 框架[29]

Fig. 12　PSPNet framework[29]

等模型,性能良好。图 13 为 PSPNet 模型对不同目标(牛、飞机、人等)的分割效果。如图 13,前景目标分割精细,但对目标间有遮挡的情况处理得不够理想,如图中(第 3 行)桌子受遮挡影响,边缘分割得不够精准。

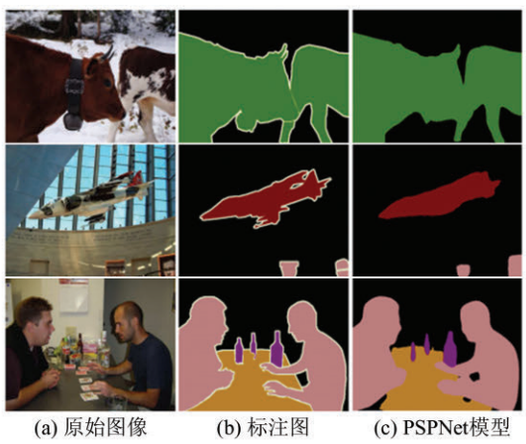


图 13 PSPNet 模型图像分割效果[29]

Fig. 13 Image segmentation effect of PSPNet model[29]

3.3 DeepLab

DeepLab 系列模型[30~33]是 Chen 等提出的深度卷积神经网络(deep convolutional neural network, DCNN)模型,其核心是使用 atrous 卷积[30],即采用在卷积核里插孔的方式,不仅能在计算特征响应时明确地控制响应的分辨率,而且还能扩大卷积核的感受野,在不增加参数量和计算量的同时,能够整合更多的特征信息。

最早的 DeepLab 模型[30]如图 14 所示,输入图像经过带有多孔(atrous)卷积层的 DCNN 处理后,得到粗略的评分图,双线性内插值上采样后引入全连接条件随机场(conditional random fields,CRF)[30]作为后处理,充分考虑全局信息,对目标边缘像素点进行更准确地分类,排除噪声干扰,从而提升分割精度。

DeepLab-v2[31]在 DeepLab 模型基础上将 atrous

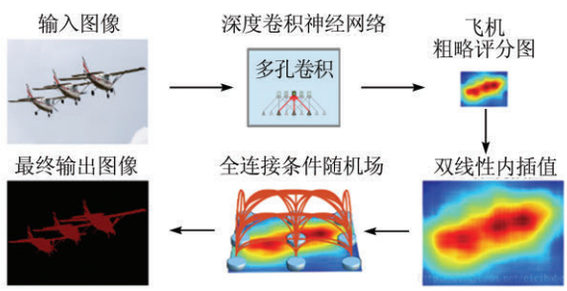


图 14 原始 DeepLab 模型[30]

Fig. 14 The original DeepLab model [30]

卷积层扩展为多孔空间金字塔池化(atrous spatial pyramid pooling,ASPP)[31]模块,级联多尺度 atrous 卷积层并进行特征图融合,保留全连接 CRF 作为后处理。

DeepLab-v3[32]如图 15,输入图像经过卷积池化后,图像尺寸缩小了 4 倍,再依次经过 3 个 Block 模块[39](Block1~ Block3)进行卷积、线性整流函数(rectified linear unit, ReLU)、池化处理,图像依次缩小 8、16、16 倍,然后经过 Block4 处理后进入 ASPP 模块,ASPP 通过融合不同多孔卷积(插孔数 rate=6、12、18)处理后,与 1×1 卷积层、全局池化层进行整合,得到缩小 16 倍的特征图,再进行分类预测得到分割图。DeepLab-v3+模型[33]采取编解码结构,如图 16,将 DeepLab-v3 模型作为编码部分,对图像进行处理后输出 DCNN 中浅层特征图和经过 ASPP 融合卷积后的特征图,并将两者作为解码部分的输入。进入解码模块,先对输入的浅层特征图卷积,再与经过上采样的 ASPP 特征图进行融合,然后经过卷积、上采样操作输出原始尺寸大小的分割图,实现端到端的语义分割。DeepLab-v3+模型对不同目标(人、马、狗等)的分割效果如图 17 所示,其中第 1、3 列表示输入图像,2、4 列表示对应的分割结果。由图 17 可知,分割后的图像中能够明显区分出前景目标和背景,目标边缘轮廓清晰,说明该模型能够实现细粒度的分割。

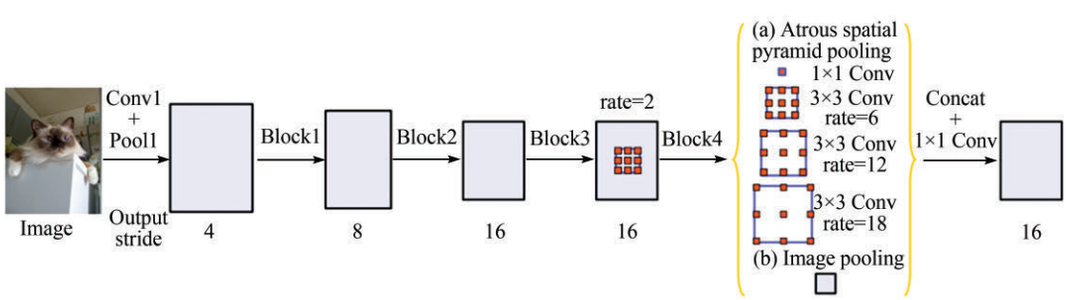


图 15 DeepLab-v3 模型结构[32]

Fig. 15 DeepLab-v3 model structure [32]

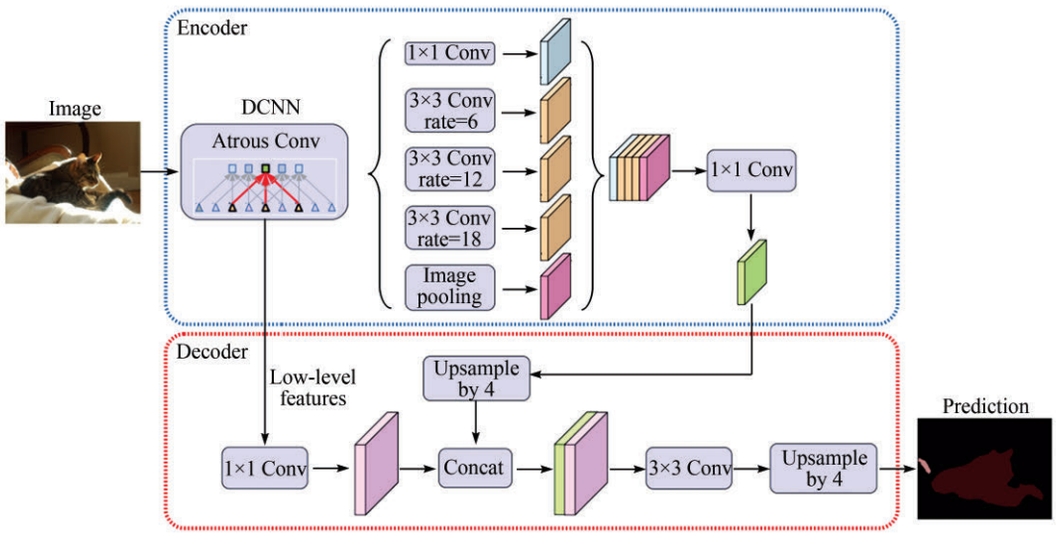


图16　DeepLab-v3+模型结构[33]

Fig. 16　DeepLab-v3+ model structure[33]

图17　DeepLab-v3+模型图像分割效果[33]

Fig. 17　DeepLab-v3+ model segmentation effect[33]

3.4　Mask R-CNN

Mask R-CNN[34]是He等基于Faster R-CNN[40]提出的用于图像分割的深度卷积网络,在进行目标检测的同时实现高质量的分割。Mask R-CNN框架如图18,第一阶段,首先用区域建议网络(region proposal networks,RPN)[40]提取出候选目标的边界框,然后对边界框里面的内容(regions of interest,RoI)进行RoIAlign[34]处理,将RoI划分为$m\times m$的子区域;第二阶段,与预测类和边界框回归任务并行,增加了为每个RoI输出二分类掩码的分支,可理解为用FCN对每个RoI进行分割,以像素到像素的方式预测分割掩码。

区别于FCN[28]、PSPSet[29]、DeepLab[30]等模型实现的语义分割,Mask R-CNN在语义分割的基础上实现了实例分割。语义分割用来识别图像中存在的内容以及位置,而实例分割是在语义分割的基础上区分同一类别下的不同个体,可以得到更精确的目标信息。与现有的实例分割模型FCIS[41]、MNC[42]等相比,Mask R-CNN模型不仅分割精度更高,而且模型更加灵活,可以用来完成多种计算机视觉任务,包括目标分类、目标检测、实例分割、人体姿态识别等。在训练阶段,Mask R-CNN模型使用多任务损失约束L,其表达式如下:

$$L=L_{\mathrm{cls}}+L_{\mathrm{box}}+L_{\mathrm{mask}} \tag{6}$$

其中,L_{cls}表示目标分类的损失,L_{box}为检测任务的损失,L_{mask}是实例分割损失。

Mask R-CNN模型复杂场景的分割效果如图19所示,图中的前景目标在实现精准检测定位的同时,实现了实例分割,对同类目标不同个体进行了区分。

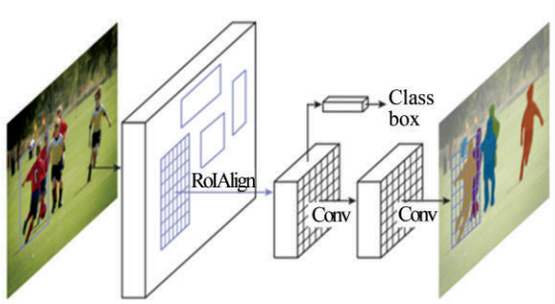


图18　实例分割的Mask R-CNN框架[34]

Fig. 18　Mask R-CNN framework for instance segmentation[34]

图19　Mask R-CNN模型图像分割效果[34]

Fig. 19　Image segmentation effect of Mask R-CNN model[34]

4 基于深度学习的图像分割方法性能比较

为了公正、客观地对各种基于深度学习的图像分割方法的性能进行科学的比较,需要使用分割领域标准的数据集进行评测,本文采用常用的3个深度学习分割数据集:PASCAL VOC[35]、Microsoft COCO[36]和Cityscapes[43]。在此基础上,对上文提到的4种经典分割方法进行了定性和定量的比较。

4.1 定性分析

如图20所示,使用以上方法,对相同的图片(取自Microsoft COCO数据集)进行分割处理。分割效果如图20所示,对比语义分割真实值(ground truth)(图20(b)),从使用FCN-8s进行分割所得效果(图20(d))来看,该算法对于大部分目标的类别有较准确的判断,但对于全局特征信息应用得不够充分且对图像细节不够敏感;PSPNet(图20(e))对大部分目标能实现准确地分类,特别是对于一些复杂交通场景的分割也能取得不错的效果,但可能会丢失目标的边缘信息,导致轮廓分割得不够精准;DeepLab-v3+(图20(f))对绝大部分目标能进行精准分类并且对边缘细节处理的效果较好,整体的分割效果比较好。Mask R-CNN(图20(g))属于实例分割的方法,对比实例分割ground truth(图20(c)),该方法在对目标进行语义分割的基础上,能对同一类别的不同个体进行区别,获得较高的分类准确率,但是当目标之间有遮挡时分割得不够精准。综合比较,FCN、PSPNet、DeepLab-v3+对目标进行语义分割都能取得不错的效果,Mask R-CNN适用于实例分割,也能对目标进行精准地分类。

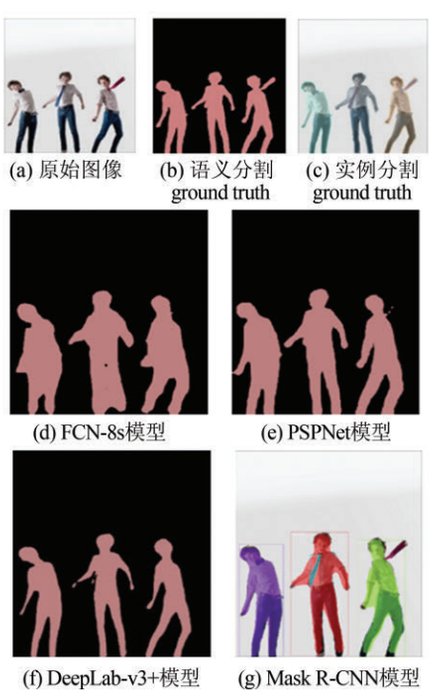


图20 基于深度学习的图像分割方法效果

Fig. 20 The effect of deep learning segmentation methods

4.2 定量分析

在上述深度学习分割数据集上,基于现有实验环境:CPU-Intel i7-6900K、GPU-Nvidia Titan Xp,将本文介绍的语义分割方法(FCN-8s[28]、PSPNet[29]、DeepLab[31~33])和实例分割方法(Mask R-CNN[34])分别与其他基于深度学习的图像分割方法的性能指标进行定量比较,结果如表2~4。

1)针对语义分割方法,采用平均交并比(mean intersection over union,mIoU)作为精度衡量指标,该值表示两个集合的交集和并集之比,在语义分割的问题中,这两个集合分别为真实值和预测值的集合。

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}\times 100\% \quad (7)$$

其中,k表示前景目标的类别数,共有$k+1$个类别(含目标和背景);i,j均表示类别号;p_{ii}表示分类正

表2 不同语义分割方法在PASCAL VOC数据集上的性能

Table 2 Performance of different semantic segmentation methods on PASCAL VOC dataset

排序	分割方法	mIoU/%
1	DeepLab-v3+[33]	89.0
2	DeepLab-v3[32]	85.7
3	DeepLab-v2[31]	79.7
4	PSPNet[29]	85.4
5	FCN-8s[28]	67.2
6	CRF-RNN[38]	74.7
7	DPN[37]	77.5

表3 不同语义分割方法在Cityscapes数据集上的性能

Table 3 Performance of different semantic segmentation methods on Cityscapes dataset

排序	分割方法	mIoU/%
1	DeepLab-v3+[33]	82.1
2	DeepLab-v3[32]	81.3
3	DeepLab-v2[31]	70.4
4	PSPNet[29]	81.2
5	FCN-8s[28]	65.3
6	CRF-RNN[38]	62.5
7	DPN[37]	66.8

确的像素;p_{ij}、p_{ji}均表示分类错误的像素。

从表2和表3中可以看出,相比于其他分割模型,DeepLab-v3+在PASCAL VOC、Cityscapes等数据集上均能够取得最高的准确性,分别是89.0%和82.1%。

2)针对实例分割方法,采用像素精度(pixel accuracy,PA)作为衡量指标,表示分类正确的像素占总像素的比例。具体计算公式

$$\mathrm{PA}=\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\times 100\% \tag{8}$$

将Mask R-CNN[34]方法与FCIS[41]、MNC[42]方法在Microsoft COCO数据集上的性能结果进行比较,结果如表4所示。由表4可知,与其他两种方法相比,Mask R-CNN在Microsoft COCO数据集上性能优异,像素精度达37.10%。

表4　不同实例分割方法在Microsoft COCO数据集上的性能

Table 4　Performance of different instance segmentation methods on Microsoft COCO dataset

排序	分割方法	PA/%
1	Mask R-CNN[34]	37.10
2	FCIS[41]	33.60
3	MNC[42]	24.60

5　结　语

本文在阐述图像分割概念的基础上,着重介绍了几类传统图像分割方法和基于深度学习的图像分割方法,选取每一类方法中的代表性算法进行研究和分析,并在图像分割的常用数据集上对同类算法进行了性能比较。

综观图像分割领域近几年的发展,在实际的分割任务中,需根据应用场景的不同,灵活地选择分割方法,有的甚至需要将多种分割方法结合使用,获得最佳的分割效果。随着分割技术的不断发展,图像分割在计算机视觉任务中的应用越来越广泛,分割的准确性和速度也有了明显的提升,但仍然存在一些难题:1)分割数据集匮乏,标注工作繁重;2)小尺寸目标的分割不够精准;3)分割算法计算复杂;4)无法实现实时交互式分割,阻碍了分割技术的落地、应用和推广。这些问题将成为未来的研究热点,具有极其重要的研究价值和意义。

参考文献:

[1] OTSU N. A threshold selection method from gray-level histograms [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **9**(1): 62-66. DOI: 10.1109/TSMC.1979.4310076.

[2] PUN T. A new method for grey-level picture thresholding using the entropy of the histogram [J]. *Signal Processing*, 1980, **2**(3): 223-237. DOI: 10.1016/0165-1684(80)90020-1.

[3] YEN J C, CHANG F J, CHANG S. A new criterion for automatic multilevel thresholding [J]. *IEEE Transactions on Image Processing*, 1995, **4**(3): 370-378.

[4] KHAN J F, BHUIYAN S M A, ADHAMI R R. Image segmentation and shape analysis for road-sign detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2011, **12**(1): 83-96. DOI: 10.1109/ TITS.2010.2073466.

[5] ROSENFELD A. The max Roberts operator is a Hueckel-type edge detector [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1981, **3**(1): 101-103. DOI: 10.1109/TPAMI.1981.4767056.

[6] LANG Y, ZHENG D. An improved Sobel edge detection operator [C]// *IEEE International Conference on Computer Science & Information Technology*. New York: IEEE Press, 2010: 67-71. DOI: 10.1109/ ICCSIT.2010.5563693.

[7] YANG L, WU X Y, ZHAO D W, *et al*. An improved Prewitt algorithm for edge detection based on noised image [C]// 2011 4*th International Congress on Image and Signal Processing*. New York: IEEE Press, 2011: 1197-1200.

[8] ULUPINAR F, MEDIONI G. Refining edges detected by a LoG operator [J]. *Computer Vision Graphics & Image Processing*, 1990, **51**(3): 275-298. DOI: 10.1016/0734-189X(90)90004-F.

[9] LI E S, ZHU S L, ZHU B S, *et al*. An adaptive edge-detection method based on the Canny operator [C]//2009 *International Conference on Environmental Science and Information Application Technology*. New York: IEEE Press, 2009: 465-469. DOI: 10.1109/ESIAT.2009.49.

[10] ZHANG Y J. An Overview of Image and Video Segmentation in the Last 40 Years [EB/OL]. [2018-02-10]. *https://www.irma-international.org/viewtitle/*4834/? *isxn*=97815914 07539.

[11] PHAM D L, XU C Y, PRINCE J L. A survey of current methods in medical image segmentation [J]. *Annual Review of Biomedical Engineering*, 2000, **2**(1): 315-337.

[12] TREMEAU A , BOREL N. A region growing and merging algorithm to color segmentation [J]. *Pattern Recognition*, 1997, **30**(7): 1191-1203. DOI: 10.1016/S0031-3203(96)00147-1.

[13] CHENG Y. Mean shift, mode seeking, and clustering [J].

IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, **17**(8): 790-799. DOI: 10.1109/34.400568.

[14] FUKUNAGA K, HOSTETLER L D. The estimation of the gradient of a density function, with applications in pattern recognition [J]. *IEEE Transactions on Information Theory*, 1975, **21**(1): 32-40. DOI: 10.1109/TIT.1975.1055330.

[15] SHEIKH Y A, KHAN E A, KANADE T. Mode-seeking by Medoidshifts [C]//2007 *IEEE* 11*th International Conference on Computer Vision*. New York: IEEE Press, 2007: 1-8. DOI: 10.1109/ ICCV.2007.4408978.

[16] LEVINSHTEIN A, STERE A, KUTULAKOS K N, *et al*. TurboPixels: Fast superpixels using geometric flows [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(12): 2290-2297. DOI: 10.1109/ TPAMI.2009.96.

[17] ACHANTA R, SHAJI A, SMITH K, *et al*. SLIC superpixels compared to state-of-the-art superpixel methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2274-2282. DOI: 10.1109/ TPAMI.2012.120.

[18] BOYKOV Y Y, JOLLY M P. Interactive graph cuts for optimal boundary region segmentation of objects in N-D images [C]//*Proceedings Eighth IEEE International Conference on Computer Vision*. New York: IEEE, 2001: 105-112. DOI: 10.1109/ICCV.2001.937505.

[19] ROTHER C, KOLMOGOROV V, BLAKE A, *et al*. "GrabCut": Interactive foreground extraction using iterated graph cuts [J]. *International Conference on Computer Graphics and Interactive Techniques*, 2004, **23**(3): 309-314. DOI: 10.1145/1015706.1015720.

[20] TANG M, GORELICK L, VEKSLER O, *et al*. GrabCut in One Cut [C]//*Proceedings of the* 2013 *IEEE International Conference on Computer Vision*. New York: IEEE Press, 2013: 1769-1776.

[21] MATHERON G. *Random Sets and Integral Geometry* [M]. New York: Wiley Press, 1975.

[22] CRESPO J, W.SCHAFER R, SERRA J, *et al*. The flat zone approach: A general low-level region merging segmentation method [J]. *Signal Processing*, 1997, **62**(1): 37-60.

[23] HOLLAND J H. Genetic algorithms and the optimal allocation of trials [J]. *SIAM Journal on Computing*, 1973, **2**(2): 88-105. DOI: 10.1137/0202009.

[24] LIU H H, CHEN Z H, CHEN X H, *et al*. Multiresolution medical image segmentation based on wavelet transform [C]// 2005 *IEEE Engineering in Medicine and Biology* 27th *Annual Conference*. New York: IEEE Press, 2006: 3418-3421. DOI: 10.1109/IEMBS. 2005. 1617212.

[25] YANG X, CHUNG A C S, JIAN Y. An active contour model for image segmentation based on elastic interaction [J]. *Journal of Computational Physics*, 2006, **219**(1): 455-476. DOI: 10.1016/j.jcp.2006.03.026.

[26] SAHA P K, UDUPA J K. Relative fuzzy connectedness among multiple objects: Theory, algorithms, and applications in image segmentation [J]. *Computer Vision and Image Understanding*, 2001, **82**(1): 42-56. DOI: 10.1006/cviu. 2000.0902.

[27] ROY P, GOSWAMI S, CHAKRABORTY S, *et al*. Image segmentation using rough set theory: A review [J]. *International Journal of Rough Sets & Data Analysis*, 2014, **1**(2): 62-74. DOI: 10.1006/cviu.2000.0902.

[28] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the* 2015 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2015: 3431-3440. DOI: 10.1109/CVPR.2015.7298965.

[29] ZHAO H S, SHI J P, QI X J, *et al*. Pyramid scene parsing network [C]//*Proceedings of the* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2017: 2881-2890. DOI: 10.1109/CVPR.2017.660.

[30] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al*. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs [EB/OL]. [2018-08-09]. *https://ui.adsabs.harvard.edu/abs/*2014*arXiv*1412.7062*C/abstract*.

[31] 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. DOI: 10.1109/TPAMI. 2017.2699184.

[32] CHEN L C, PAPANDREOU G, SCHROFF F, *et al*. Rethinking Atrous Convolution for Semantic Image Segmentation [EB/OL]. [2018-05-09]. *https://arxiv.org/pdf/*1706.05587*.pdf*.

[33] CHEN L C, ZHU Y, PAPANDREOU G, *et al*. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [EB/OL]. [2018-08-09]. *https://arxiv.org/pdf/*1802.02611*v1.pdf*.

[34] HE K M, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN [C]//2017 *IEEE International Conference on Computer Vision* (*ICCV*). New York: IEEE Press, 2017: 2980-2988. DOI: 10.1109/ICCV.2017.322.

[35] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, *et al*. The PASCAL visual object classes challenge: A retrospective [J]. *International Journal of Computer Vision*, 2015, **111**(1): 98-136. DOI: 10.1007/s11263-014-0733-5.

[36] LIN T Y, MAIRE M, BELONGIE S, *et al*. Microsoft COCO: Common objects in context [C]//*European Conference on Computer Vision* (*LNCS* 8639). Berlin: Springer, 2014: 740-755. DOI: 10.1007/978-3-319-10602-1_48.

[37] CHEN Y P, LI J N, XIAO H X, *et al*. Dual path networks [C]//*Proceedings of the* 31*st International Conference on Neural Information Processing Systems*. New York: ACM, 2017: 4467-4475.

[38] CHENG J, SUN Y, MENG Q H. A dense semantic mapping system based on CRF-RNN network [C]//2017 18*th International Conference on Advanced Robotics* (*ICAR*). New York: IEEE Press, 2017: 589-594. DOI: 10.1109/ICAR.2017.8023671.

[39] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition [C]// 2016 *IEEE Conference on Computer Vision & Pattern Recognition*. Washington D C: IEEE Computer Society, 2016:770-778.

[40] REN S Q, HE K M, GIRSHICK R, *et al*. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [EB/OL]. [2018-08-08]. *https://arxiv.org/pdf/*1506.01497*.pdf*.

[41] LI Y, QI H Z, DAI J, *et al*. Fully convolutional instance-aware semantic segmentation [C]//*Proceedings of the* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. New York : IEEE Press, 2017: 2359-2367.

[42] DAI J, HE K, SUN J. Instance-aware semantic segmentation via multi-task network cascades [C]//*Proceedings of the* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2016: 3150-3158.

[43] CORDTS M, OMRAN M, RAMOS S, *et al*. The Cityscapes dataset for semantic urban scene understanding [C]//*Proceedings of the* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. New York : IEEE Press, 2016: 3213-3223.

□