

## 基于内容的图像分割方法综述<sup>\*</sup>

姜 枫<sup>1,2,3</sup>, 顾 庆<sup>1,2</sup>, 郝慧珍<sup>1,2,4</sup>, 李 娜<sup>1,2</sup>, 郭延文<sup>1,2</sup>, 陈道蓄<sup>1,2</sup>



<sup>1</sup>(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210023)

<sup>2</sup>(南京大学 计算机科学与技术系, 江苏 南京 210023)

<sup>3</sup>(南京理工大学泰州科技学院 移动互联网学院, 江苏 泰州 225300)

<sup>4</sup>(南京工程学院 通信工程系, 江苏 南京 211167)

通讯作者: 顾庆, E-mail: guq@nju.edu.cn

**摘 要:** 图像分割是指将图像分成若干具有相似性质的区域的过程, 是许多图像处理任务的预处理步骤. 近年来, 国内外学者主要研究基于图像内容的分割算法. 在广泛调研大量文献和最新成果的基础上, 将图像分割算法分为基于图论的方法、基于像素聚类的方法和语义分割方法这 3 种类型并分别加以介绍. 对每类方法所包含的典型算法, 尤其是最近几年利用深度网络技术的语义图像分割方法的基本思想、优缺点进行了分析、对比和总结. 介绍了图像分割常用的基准数据集和算法评价标准, 并用实验对各种图像分割算法进行对比. 最后进行总结, 并对未来可能的发展趋势加以展望.

**关键词:** 图像分割; 图论; 聚类; 语义分割; 深度神经网络

**中图法分类号:** TP391

中文引用格式: 姜枫, 顾庆, 郝慧珍, 李娜, 郭延文, 陈道蓄. 基于内容的图像分割方法综述. 软件学报, 2017, 28(1): 160-183. <http://www.jos.org.cn/1000-9825/5136.htm>

英文引用格式: Jiang F, Gu Q, Hao HZ, Li N, Guo YW, Chen DX. Survey on content-based image segmentation methods. Ruan Jian Xue Bao/Journal of Software, 2017, 28(1): 160-183 (in Chinese). <http://www.jos.org.cn/1000-9825/5136.htm>

## Survey on Content-Based Image Segmentation Methods

JIANG Feng<sup>1,2,3</sup>, GU Qing<sup>1,2</sup>, HAO Hui-Zhen<sup>1,2,4</sup>, LI Na<sup>1,2</sup>, GUO Yan-Wen<sup>1,2</sup>, CHEN Dao-Xu<sup>1,2</sup>

<sup>1</sup>(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

<sup>2</sup>(Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China)

<sup>3</sup>(College of Mobile Internet, Taizhou Institute of Science & Technology, Nanjing University of Science and Technology, Taizhou 225300, China)

<sup>4</sup>(Department of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

**Abstract:** Image segmentation is the process of dividing the image into a number of regions with similar properties, and it's the preprocessing step for many image processing tasks. In recent years, domestic and foreign scholars mainly focus on the content-based image segmentation algorithms. Based on extensive research on the existing literatures and the latest achievements, this paper categorizes image segmentation algorithms into three types: graph theory based method, pixel clustering based method and semantic segmentation method. The basic ideas, advantage and disadvantage of typical algorithms belong to each category, especially the most recent image

\* 基金项目: 国家自然科学基金(61373012, 91218302, 61321491, 61373059); 江苏省高校自然科学研究项目(15KJB520016); 江苏省自然科学基金(BK20150016)

Foundation item: National Natural Science Foundation of China (61373012, 91218302, 61321491, 61373059); Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China (15KJB520016); Natural Science Foundation of Jiangsu Province (BK20150016)

收稿时间: 2016-06-16; 修改时间: 2016-09-07; 采用时间: 2016-09-29; jos 在线出版时间: 2016-11-04

CNKI 网络优先出版: 2016-11-04 14:29:23, <http://www.cnki.net/kcms/detail/11.2560.TP.20161104.1429.001.html>

semantic segmentation algorithms based on deep neural network are analyzed, compared and summarized. Furthermore, the paper introduces the datasets which are commonly used as benchmark in image segmentation and evaluation criteria for algorithms, and compares several image segmentation algorithms with experiments as well. Finally, some potential future research work is discussed.

**Key words:** image segmentation; graph theory; clustering; semantic segmentation; deep neural network

图像分割是指根据一定的相似性准则将图像划分成不同区域的过程,是计算机视觉、图像处理等领域的基础性问题之一,是图像分类、场景解析、物体检测、图像 3D 重构等任务的预处理.其研究从 20 世纪 60 年代开始,至今仍然是研究的热点之一,并且被广泛应用于医学影像分析、交通控制、气象预测、地质勘探、人脸与指纹识别等诸多领域.

图像分割的任务是将一幅图像划分为互不相交的连通区域,其数学定义为:令集合  $R$  表示整个图像区域,图像分割就是将  $R$  分成一组连通的非空子集(子区域) $\{R_1, R_2, \dots, R_N\}$ ,使  $\bigcup_{i=1}^N R_i = R, R_i \cap R_j = \emptyset (i \neq j)$ , 并存在判断区域一致性的逻辑谓词  $P(\cdot)$ ,使得  $P(R_i) = \text{True}, P(R_i \cap R_j) = \text{False}, i \neq j$ .

传统的图像分割方法主要包括阈值法<sup>[1]</sup>、边界检测法<sup>[2]</sup>、区域法<sup>[3]</sup>等,这些方法的实现原理有所不同,但基本都是利用图像的低级语义,包括图像像素的颜色、纹理和形状等信息<sup>[4]</sup>,遇到复杂场景时实际分割效果不尽理想.

如何有效地利用图像自身包含的内容信息,结合图像中级、高级语义提升图像分割效果,成为近年来研究的热点.这类方法可以认为是基于内容的图像分割,亦即:不是简单地将图像分割归为边界检测或区域查找,而是利用图像或区域的内容信息,例如特征相似性、图像语义等,来确定图像中的区域划分.

图像的中级语义是指将具有相似特征的相邻像素构成的图像块所具有的像素以外的附加信息,如图像块的粗糙度、对比度、方向度、紧凑度等,以此辅助图像分割并提升效果.2000 年左右,研究人员将图论的理论和方法引入图像分割问题.其原理是将待分割图像映射为带权无向图,根据图的顶点以及边的信息构造代价函数并加以优化,将图像分割问题转换为图的顶点标注问题,标号相同的顶点所对应的像素属于同一个图像块,这些图像块称为超像素(superpixel)<sup>[5]</sup>.另一种思路是借鉴无监督(unsupervised)学习的思想,根据图像中单个像素的信息以及像素之间的相互关系,利用数据挖掘中的聚类算法,将具有相似特征的相邻像素归并到同一个超像素.

图像的高级语义是指图像或图像区域所包含的对象或实体的类别等语义信息,高级语义下的图像分割称为语义分割(semantic segmentation).图像语义分割是指能够将图像中前景和背景分离,并识别出每个前景目标的类别,其数学定义为:对于图像  $R = \{p | i = 1, 2, \dots, M\}$ ,语义分割就是要指定每个区域(或像素)的类别,即,确定  $F: R \rightarrow CL$ ,其中,  $CL = \{c_1, c_2, \dots, c_K\}$  为预设的  $K$  个类别,相当于给每个像素赋予一个语义标号.例如:针对室内场景的语义分割,其目标是给定一个室内场景,例如家居场景的图像,将其分割为地板、桌子、椅子、画像、柜橱等区域.如图 1 所示<sup>[6]</sup>,左图为原始图像,右图为分割效果图,用不同颜色表示不同语义的对象.

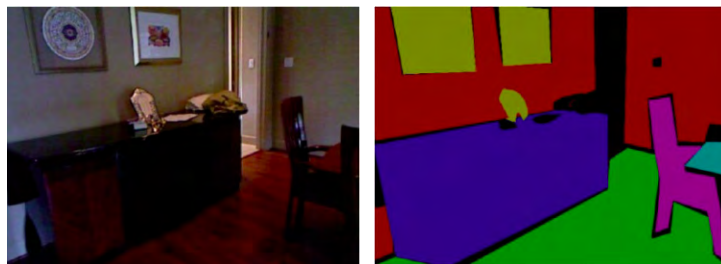


Fig.1 Image semantic segmentation

图 1 图像语义分割

语义分割算法利用有监督(supervised)机器学习技术,实现时需要以大量标注图像作为样本进行训练.然而,对图像进行像素级(pixel-wise)标注是一件相当费时费力的工作,因此,这样的强标注样本比较难以获得.最近出现了一类可以利用具有弱标注图像样本,如图像级(image-level)标注或边界框(bounding box)标注样本进行弱监

督(weakly-supervised)学习,或是结合少量强标注样本和大量弱标签样本进行半监督(semi-supervised)学习的方法,作为有监督学习方法的辅助和拓展,构建图像语义分割系统。

图像分割问题一直都是计算机视觉领域的热门话题之一,每年都有大量的新方法呈现.为了对该问题进行系统的研究与阐述,我们通过 Google scholar,IEEE,ACM,Elsevier 以及 CNKI 等搜索近 5 年的图像分割综述类文章,并选择相关程度较高、质量较好的文献作为参考.通过整理、阅读和研究,部分文献是比较传统分割技术的综述,如阈值法、边界法、区域法等<sup>[7,8]</sup>.另一部分文献是对某一类图像分割方法的综述,如文献[9-11]是基于图论方法的综述,文献[12,13]是基于聚类方法的综述.而据我们查询的结果来看,近几年非常热门的语义分割技术,尤其是深度网络在这方面的应用尚未有综述文章提及.鉴于此,在研究内容上,本着系统、科学、前沿的原则,本文不再介绍 2000 年之前的图像分割技术,主要阐述和比较 2000 年之后的以图像中、高层特征为依据、基于内容的图像分割方法,如图 2 所示,将其分为基于图论的方法、基于像素聚类的方法以及图像语义分割方法,一些文献中将前两种方法都归类为超像素方法<sup>[14,15]</sup>,每种类型的方法中又包含若干具体算法.与其他同类文献相比,本文的主要贡献如下:

- (1) 针对 2000 年以来基于内容的图像分割方法进行了较为系统、全面的研究和分类,对每类方法中的典型算法进行了阐述和分析;
- (2) 对 2012 年以来,以卷积神经网络(convolutional neural network,简称 CNN)为代表的深度学习技术在图像分割中应用的情况和趋势进行了研究和阐述;
- (3) 对图像分割算法常用的实验数据库、算法性能评价指标进行了研究和介绍;对典型算法的主要指标和特点进行了分析、总结和对比;
- (4) 讨论了图像分割中目前存在的、亟待解决的主要问题以及未来发展的趋势.

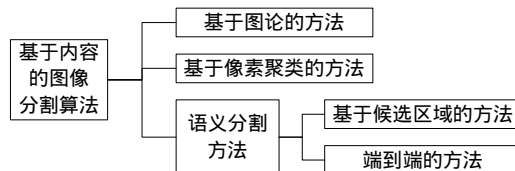


Fig.2 Classification chart of content based image segmentation algorithms

图 2 基于内容的图像分割算法分类图

本文第 1 节介绍基于图论的图像分割方法.第 2 节介绍基于像素聚类的图像分割方法.第 3 节介绍图像语义分割方法,包括基于候选区域的方法和端到端的方法.第 4 节介绍图像分割常用数据集、算法性能衡量指标以及通过实验对比各种算法.最后总结全文,并对未来可能的研究方向和发展趋势加以展望.

## 1 基于图论的分割方法

基于图论的方法利用图论领域的理论和方法,将图像分割问题看作是图的顶点划分问题.通用的方法是将待分割的图像映射为一幅带权无向图  $G=(V,E)$ ,其中,  $V=\{v_1, \dots, v_n\}$  是顶点的集合,  $E$  为边的集合.如图 3 所示.

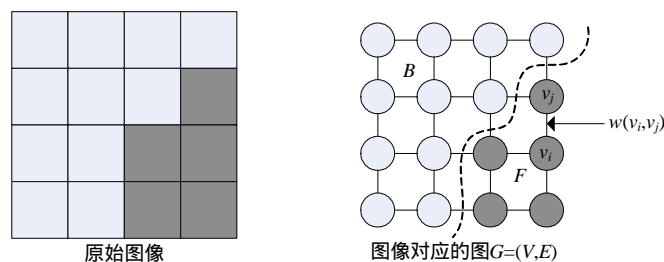


Fig.3 Transform from image to graph

图 3 图像到图的转换

由图 3 可知,原始图像中每个像素映射为无向图的一个顶点,像素自身的特征信息对应图中顶点属性;像素之间的相邻关系映射为图的边,像素特征之间的相似性或差异性对应边的权值,表示为  $w(v_i, v_j)$ , 其中,  $(v_i, v_j) \in E$ .

通过构造关于图信息的代价函数并求解,将图的顶点集合二分为前景集  $F$  和背景集  $B$ , 并且  $F \cap B = \emptyset, F \cup B = V$ . 迭代地使用上述二分方法,可以将图像分为多个图像块.

### 1.1 基于谱聚类的方法

谱聚类方法建立在谱图理论的基础之上,通过构造关于原图的拉普拉斯矩阵并求解特征值和特征向量,对图中顶点进行前景背景分离,以解决图像分割问题.

Wu 等人<sup>[16]</sup>根据网络中最大流和最小割等价的原理,将图像的最优分割问题转化为求解对应图的最小割问题.其形式化描述为:将图像对应的图  $G=(V, E)$  分割为两个不相交的集合  $A$  和  $B$ , 且有  $A \cup B = V, A \cap B = \emptyset$ . 定义最小割(min cut)的代价函数为

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \quad (1)$$

根据公式(1)的定义,最小割方法考虑了最大子图间的差异性,但忽略了子图内部差异性应尽可能小的原则,倾向于分离图中的单个节点.

Shi 等人<sup>[17]</sup>在 2000 年提出规范割(normalized cut, 简称 NCut)算法,对公式(1)进行归一化,在考虑了子图间的差异性的基础上,同时考虑了子图内部的相似性,定义全局代价函数为

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(B, A)}{assoc(B, V)} \quad (2)$$

其中,  $assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$ , 表示子集  $A$  中所有节点到图中所有节点的边权值之和.由于进行了归一化处理,单个节点分割的结果不再满足  $Ncut(A, B)$  最小,从而避免了分割单个孤点的情况.

NCut 算法通过图像的轮廓特征以及纹理特征来全局最小化代价函数,可生成较规则的图像块.最优化代价函数(2)被转换成求解拉普拉斯矩阵的特征值问题,属于 NP 完全问题.为此,Ren 等人<sup>[5]</sup>提出先用 NCut 算法将图像分成几个较大的子图,对每个子图使用  $k$ -means 算法进一步细化初始划分,从而降低算法的时间复杂度.徐海霞等人<sup>[18]</sup>先通过顶点凝聚,使用凝聚图序列逼近原图,再利用 NCut 方法生成粗尺度分割,最后依据图像统计特征,使用混合模型的最大后验概率优化分割结果得到细尺度分割结果.

Sarkar 等人<sup>[19]</sup>改进了 NCut 算法,提出了平均割(average cut, 简称 ACut)算法,定义代价函数为

$$Avcut(A, B) = \frac{cut(A, B)}{|A|} + \frac{cut(B, A)}{|B|} \quad (3)$$

公式(3)中,  $|A|$  和  $|B|$  分别表示子集  $A$  和  $B$  中的顶点数.该代价函数表示图中边界损失与区域面积的比值之和,最小化该代价函数能够产生准确的划分.ACut 算法的不足与 NCut 算法类似,容易分割出只包含几个顶点的较小子图.

Ding 等人<sup>[20]</sup>针对上述不足,提出了最小最大割(min-max cut, 简称 MCut)算法,定义代价函数为

$$Mcut(A, B) = \frac{cut(A, B)}{W(A)} + \frac{cut(B, A)}{W(B)} \quad (4)$$

公式(4)中,  $W(A)$  定义为子集  $A$  中所有边的权值之和.根据类间相似度最小和类内相似度最大原则,需要同时最小化  $cut(A, B)$  和最大化  $W(A)$  及  $W(B)$ , 最小化公式(4)能够分割出较为平衡的割集.

Li 等人<sup>[21]</sup>提出了线性谱聚类(linear spectral clustering, 简称 LSC)算法,该算法基于  $K$  路 NCut( $K$ -way NCut)算法<sup>[17]</sup>的代价函数,使用核函数将像素值和坐标映射到高维特征空间,通过证明带权  $k$ -means(weighted  $k$ -means)算法和  $K$ -way NCut 算法的代价函数共享相同的最优点,迭代地使用  $k$ -means 算法在高维特征空间聚类代替 NCut 算法中特征值和特征向量的求解,将算法复杂度降低到  $O(N)$ .

### 1.2 基于图割理论的方法

在计算机视觉领域,马尔可夫随机场(Markov random field, 简称 MRF)<sup>[22]</sup>经常用来对图像像素级特征进行

建模.由于MRF和吉布斯分布(Gibbs distributions)的等价关系,可将MRF的联合分布转换成简单的形式,通过贝叶斯最大后验(maximum a posterior,简称MAP)估计求解代价函数,将图像分割问题转化为代价函数优化问题.Greig等人<sup>[23]</sup>首先提出,min-cut/max-flow算法可用于代价函数的优化,称为图割,并将其应用在图像恢复技术中,随后也被用于图像分割问题.

2006年,Boykov等人基于图割理论提出了Graph Cuts算法<sup>[24]</sup>,该算法将图像映射成无向图 $G=(V,E)$ , $V$ 中顶点包含两类:第1类是原始图像每个像素点对应的顶点,称为普通顶点;第2类是增加的两个顶点, $S$ (source)和 $T$ (sink),分别表示前景和背景.相应地, $E$ 中的边也分为两种类型:普通顶点间的连接称为 $n$ -links,普通顶点和 $S$ 或 $T$ 之间的连接称为 $t$ -links.Graph Cuts中的Cuts是指这样一个边的集合,该集合包括了上述两种类型的边,该集合中所有边的断开会导致残留 $S$ 和 $T$ 图的分开,称为割.边的权值之和最小的割称为最小割.最小割将图的顶点划分为两个不相交的子集,分别对应于图像的前景像素集和背景像素集.

Graph Cuts算法利用了图像的像素灰度信息和区域边界信息,代价函数构建在全局最优的框架下,保证了分割效果.但Graph Cuts是NP难问题,且分割结果更倾向于具有相同的类内相似度.Rother等人<sup>[25]</sup>在文献[24]的基础上提出了基于迭代的图割方法,称为Grab Cut算法.该算法使用高斯混合模型(Gaussian mixture model,简称GMM)对目标和背景建模<sup>[26]</sup>,利用了图像的RGB色彩信息和边界信息,通过少量的用户交互操作得到非常好的分割效果.文献[27]在代价函数中引入高阶势能项,用以同时描述单个像素的细节信息以及像素之间的关联信息,提高了模型的表达精度.通过实验对比,该算法性能优于Grab Cut,并可应用于视频的分割.

Moorer等人<sup>[28]</sup>于2008年基于最小割的概念提出Superpixel Lattice算法,该算法以二维图像边界代价图(boundary cost map)作为输入,该图保存了两个像素之间存在边界的概率.以迭代的方式在图像水平和垂直方向寻找最有可能通过分割边界的路径,使图像不断被划分成小的块.该方法生成的图像块形状比较规则,接近网格结构,分割准确度也较高.

为了克服Superpixel Lattice算法<sup>[28]</sup>过分依赖边界代价图的缺点,2009年,Moore等人<sup>[29]</sup>对文献[28]加以改进,在分割中加入先验信息,通过学习一个描述物体边界的空间密度概率模型,采用过分割算法将图像划分成密度基本均匀的图像块.

图割方法中,为了取得更好的分割效果,通常需要构造复杂的代价函数,算法时间复杂性高、难以满足实时应用的要求.针对这一情况,Bergh等人<sup>[30]</sup>提出了SEEDS(superpixels extracted via energy-driven sampling)算法.SEEDS算法首先预生成一个初始分割;接着,通过移动相邻图像块之间的边界或者交换相邻图像块中的像素点,不断地修正分割边界以获得最优效果.其代价函数包含两项,分别代表图像块颜色密度分布项和分割边界局部不规则性惩罚项,利用爬山法(hill-climbing)最优化代价函数,以迭代的方式,通过寻找最小局部变化更新求解.实验结果表明:该算法运行速度快,并且算法性能优秀.

### 1.3 基于最小生成树的方法

最小生成树(minimal spanning tree,简称MST)是图论中的一个概念,是指边的权值之和最小的生成树.最小生成树的求解算法主要有Kruskal<sup>[31]</sup>,Prim<sup>[32]</sup>等.Zahn等人<sup>[33]</sup>利用最小生成树和聚类结构之间的内在联系,最先将最小生成树算法应用于图像分割问题.对于真实场景的图像,如果目标内部差异较大,这种简单的权值计算方法则易导致分割效果不佳.Morris等人<sup>[34]</sup>提出了一种层次化的图像分割方法,该方法根据相似像素聚集、相异像素分离的原则,能够获得不同尺度的分割,在最小生成树最大权值处进行划分,得到图像分割结果.Kwok等人<sup>[35]</sup>对文献[34]中的算法进行优化,提出快速递归最小生成树法,提升了算法运行速度.

2004年,Felzenswalb和Huttenlocher<sup>[36]</sup>提出FH算法,该算法将图像映射成无向图 $G=(V,E)$ ,其边的权值是一个用以衡量两个像素间非相似程度的非负值.一个分割 $S$ 就是将 $V$ 分成不同区域,其中每个子区域 $C \in S$ ,对应于图 $G'=(V,E')$ 的一个连通子图,其中 $E'$ 为 $E$ 的非空子集.区域的内部差异以其内部最小生成树的最大权值表示,两个区域之间的差异表示为在所有能连接两个区域的边中权值最小的边的权值.如果两个区域之间的差异大于两个区域中任意一个区域的内部差异,则认为两个区域之间存在边界.

FH方法综合考虑了区域内部相似度和区域之间的差异性,通过对图中的顶点进行聚类实现分割,能够较好

地保持图像边界.该算法的时间复杂度为  $O(M \log N)$ ,无法控制图像块的数量和紧凑程度,并且分割效果对选定的阈值依赖性较强.

孟庆涛等人<sup>[37]</sup>针对上述不足,先将图像分成指定大小的规则块,将块映射为图的顶点,根据块内颜色纹理特征统计量和小波能量直方图计算块间距离,作为图中边的权值,利用最小生成树进行顶点聚类,解决了 FH 算法对纹理图像和边缘突变图像分割效果不理想的问题.叶伟等人<sup>[38]</sup>根据 Mumford-Shah 理论计算区域间基于结合和几何性质的权值,加入到最小生成树的区域合并公式中,能够生成更加简洁、平滑的分割边界.

## 2 基于像素聚类的分割方法

机器学习中的聚类方法也可以用于解决图像分割问题,其一般步骤是:首先初始化一个粗糙的聚类,然后使用迭代的方式将空间距离接近、颜色、亮度、纹理等特征相似的像素点聚类到同一超像素,直至收敛,从而得到最终的图像分割结果.

### 2.1 基于Meanshift的方法

Meanshift 算法的原理如下:在  $d$  维空间中,任选一点作为圆心,以  $h$  为半径做圆.圆心和圆内的每个点都构成一个向量.将这些向量进行矢量加法操作,得到的结果就是 Meanshift 向量.继续以 Meanshift 向量的终点为圆心做圆,得到下一个 Meanshift 向量.通过有限次迭代计算,Meanshift 算法一定可以收敛到图中概率密度最大的位置,即数据分布的稳定点,称为模点.利用 Meanshift 做图像分割,就是把具有相同模点的像素聚类到同一区域的过程,其形式化定义为

$$y_{k+1}^{mean} = \arg \min_z \sum_i \|x_i - z\|^2 \varphi\left(\left\|\frac{x_i - y_k}{h}\right\|\right) \quad (5)$$

其中,  $x_i$  表示待聚类的样本点,  $y_k$  代表点的当前位置,  $y_{k+1}$  代表点的下一个位置,  $h$  表示带宽.

Meanshift 算法最初由 Fukunaga 等人<sup>[39]</sup>提出.1995 年,Cheng 等人<sup>[40]</sup>发表的文献定义了核函数和权值系数,使 Meanshift 算法得到了广泛应用.李乡儒<sup>[41]</sup>和文志强<sup>[42]</sup>讨论了 Meanshift 算法的收敛性.2002 年,Comaniciu 等人<sup>[43]</sup>提出了基于核密度梯度估计的迭代式搜索算法,其基本思想是:通过定位密度函数的局部最大位置,将具有相同模点的像素聚类在一起,形成超像素区域.

Meanshift 算法的稳定性、鲁棒性较好,有着广泛的应用.但是分割时所包含的语义信息较少,分割效果不够理想,无法有效地控制超像素的数量,且运行速度较慢,不适用于实时处理任务.

针对 Meanshift 算法运行速度较慢的问题,李宏益等人<sup>[44]</sup>利用近年来计算机硬件技术取得的进步,使用 TBB(threading building block)和 CUDA(compute unified device architecture)对算法进行多核改造和 GPU(graphic processing unit)并行化改造,大幅缩短算法运行时间.

Sheikh 等人<sup>[45]</sup>在 Meanshift 算法的基础上提出了一种模式搜索算法,称为 Medoidshift 算法.与 Meanshift 算法类似,Medoidshift 算法也能自动计算聚类的数目.在 Medoidshift 算法中,每次迭代需要计算新的中心(medoid),该中心点是指在所有的  $\{x_i\}$  中能够使公式(5)目标函数达到最小值的点,这是一种增量聚类算法.

Medoidshift 算法的缺陷在于其时间复杂度较高,为  $O(N^3)$ ,通过改进可以降低到  $O(N^{2.38})$ .而 Meanshift 算法的时间复杂度为  $O(dN^2T)$ ,其中,  $d$  是数据维度,  $T$  为算法迭代次数.显然,  $dT \ll N$ .Vedaldi 等人<sup>[46]</sup>对 Medoidshift 算法中点对的距离限制为欧氏距离,引导像素特征空间中的每一个数据点向着能使 Parzen 密度估计增大的最近像素移动来实现图像分割.该算法称为 Quickshift,时间复杂度为  $O(N^2)$ ,低于 Meanshift 算法的时间复杂度.

### 2.2 基于Turbopixels的方法

基于 Meanshift 的图像分割方法无法控制生成超像素的紧凑度(compactness)<sup>[47]</sup>,并且在图像对比度较差或包含阴影时,非常容易产生欠分割(undersegmentation)现象.

2009 年,Levinshtein 等人<sup>[48]</sup>在算法中引入紧凑度约束,提出了一种基于几何流(geometric flow)的超像素生成算法,称为 TurboPixels,该算法能够生成尺寸均匀、紧凑、边界光滑的超像素.Turbopixels 算法的原理是:利用

几河流水平集方法,通过对初始种子点的膨胀,在已分配区域和未分配区域之间通过曲率演化获得超像素边界. TurboPixels 算法生成的超像素较好地保持了图像的局部边界,有效地限制了欠分割的发生.算法时间复杂度近似为  $O(N)$ ,适用于百万像素级的大图像.

在 TurboPixels 的基础上,Xiang 等人<sup>[49]</sup>提出一种方法,从待分割图像中学习特征图(eigen-image),这些特征图用以生成多维梯度算子得到图像分割,并利用图像金字塔加速运算.Cigla 等人<sup>[50]</sup>使用带凸约束的  $k$ -means 算法进行像素初始聚类,再迭代地使用 NCut 算法将图像不断进行二分,生成与 TurboPixels 效果类似、近似均匀、边界贴合度高的超像素.

### 2.3 基于SLIC的方法

Achanta 等人在 2010 年提出了 SLIC(simple linear iterative clustering)算法.将图像中的每个像素分解成一个五维向量  $\{l, a, b, x, y\}$ ,其中,  $\{l, a, b\}$  是 CIE Lab 颜色空间中的分量,  $\{x, y\}$  是像素的坐标.定义像素  $i$  和聚类中心  $k$  之间的距离为

$$D_s = d_{lab} + \frac{m}{S} d_{xy} \quad (6)$$

其中,  $d_{lab}$  是两个像素 lab 空间的欧氏距离;  $d_{xy}$  是两个像素坐标的欧式距离;  $m$  是用来控制超像素紧凑度的参数,一般在 1~20 之间;  $S = \sqrt{N/K}$ ,  $K$  为设置的超像素的块数,  $N$  为图像的总像素数. SLIC 算法首先将图像分成规则的块,每个块为一个初始聚类.计算每个聚类的中心,然后以迭代的方法从每个聚类中心出发,搜索周边  $2S \times 2S$  区域内的像素,并将其分配到距离最近的聚类,直至算法收敛.

SLIC 算法的实质是将  $k$ -means 算法用于超像素聚类,众所周知,  $k$ -means 算法的时间复杂度为  $O(NKI)$ ,其中,  $N$  是图像的像素数,  $K$  是聚类数,  $I$  是迭代次数.由于在 SLIC 中,每个点仅需要和周边最多 8 个点进行运算,且迭代次数是常数,因此, SLIC 算法的时间复杂度为  $O(N)$ ,且能够生成紧凑、近似均匀的超像素.

2012 年, Achanta 等人<sup>[14]</sup>对 SLIC 算法进行了改进,采用更复杂的距离计算公式 ASLIC(adaptively SLIC)以及 GSLIC(geodesic SLIC).在 ASLIC 中,根据每个聚类中观察到的最大空间距离和最大颜色距离( $m_s, m_c$ ),对  $d_{lab}$  和  $d_{xy}$  进行归一化,改写距离公式如下:

$$D_s = \sqrt{\left(\frac{d_{lab}}{m_c}\right)^2 + \left(\frac{d_{xy}}{m_s}\right)^2} \quad (7)$$

而在 GSLIC 中,使用测地线距离(geodesic distance)代替  $D_s$ .

Ren 等人<sup>[51]</sup>提出 gSLIC,是使用 NVIDIA CUDA 框架的 GPU 并行加速版本 SLIC,通过使用单块显卡,使算法实现提速 10~20 倍,从而使 SLIC 算法可应用于实时场景.2012 年, Schick 等人<sup>[47]</sup>对 SLIC 算法进行改进,对位于分割边界的像素重新分配,保证分割后超像素的完整性,并通过引入参数控制超像素生成的紧凑度.孙巍等人<sup>[52]</sup>提出了一种基于 SLIC 的条件随机场(conditional random fields,简称 CRF)图像分割算法,首先,利用 SLIC 将图像预分割成超像素;然后,以超像素作为图的顶点构建 CRF 模型;最后,通过极大似然估计及模型推断得到分割结果,提升了算法效率.

## 3 图像语义分割方法

聚类方法可以将图像分割成大小均匀、紧凑度合适的超像素块,为后续的处理任务提供基础.但在实际场景的图片中,一些物体的结构比较复杂,内部差异性较大,仅利用像素点的颜色、亮度、纹理等较低层次的内容信息不足以生成好的分割效果,容易产生错误的分割.因此,需要更多地结合图像提供的中高层内容信息辅助图像分割,称为图像语义分割.语义分割的概念最早由 Ohta 等人于 1978 年提出,其定义为:为图像中每个像素分配一个预先定义的表示其语义目标类别的标签<sup>[53]</sup>.

语义分割的方法分为两种:一种是自顶向下(top-down)的方法<sup>[54]</sup>,即,使用物体形状模型在待分割图像中做匹配搜索,由于每种物体的形状可变性非常大,这种方法的效果和适应性不佳;另一种是自底向上(bottom-up)的



方法<sup>[55]</sup>,这类方法无需物体形状的先验知识,首先从图中生成区域假设(region hypotheses)或候选区域(region candidates),然后对候选区域进行分类打分并预测.这种方法的主要问题是缺乏能够生成高质量候选区域的算法,因此最初采用在物体检测中的滑动窗口方法生成候选区域,这种方法相当费时费力,而且无法描述不规则形状的物体.2010年,Carreira等人提出的CPMC(constrained parametric min-cuts)算法<sup>[56]</sup>能够在图像中生成高质量的候选区域,使自底向上的方法得到了长足的发展,第3.1节介绍的基于候选区域的方法其思想即来源于此类方法.需要说明的是:在自底向上的方法中,有一类方法直接以图像像素<sup>[26,57,58]</sup>或超像素<sup>[59,60]</sup>为处理单位,提取其自身及邻域特征用于语义分割,第3.2节介绍的端到端的方法即是这种方法的发展和延伸.

### 3.1 基于候选区域的方法

图像语义分割方法中最典型的是基于候选区域的方法,这种方法的处理步骤如图4所示<sup>[61]</sup>.第一,图像底层特征提取,包括颜色、纹理、形状等基本特征;第二,图像分割,根据提取的底层特征将图像分割成不同区域;第三,结合先验知识,对每个区域进行语义分类,并根据分类结果对原图像进行标注.最终得到图像语义分割结果.

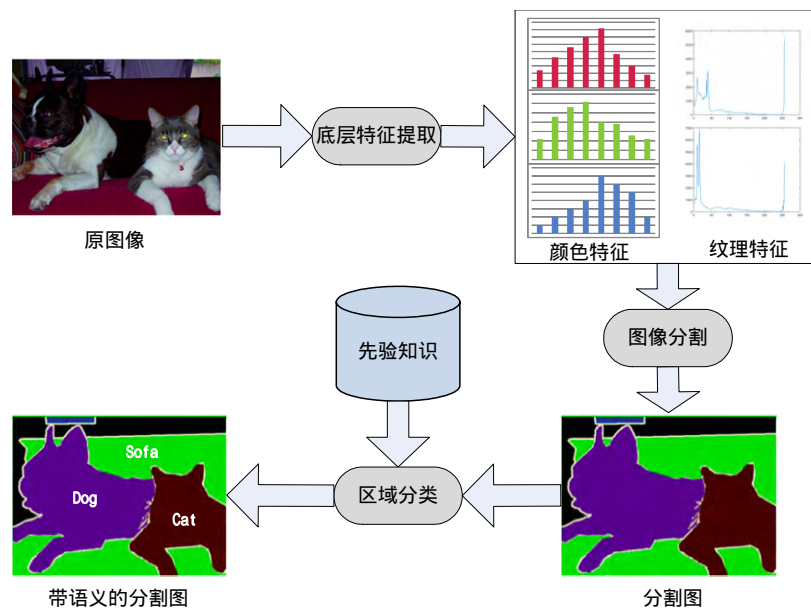


Fig.4 Pipeline for image semantic segmentation

图4 图像语义分割流程

2012年,Carreira等人<sup>[62]</sup>首先采用CPMC算法<sup>[56]</sup>从原始图像中提取出一系列候选区域;然后,将候选区域与图像中层特征(物体形状特征)匹配并打分排序,从中选出一定数量的有效区域;计算每个有效候选区域属于特定类别的概率并排序,最终获得分割结果.该算法在PASCAL VOC 2010竞赛<sup>[63]</sup>中显示的分割准确率为34.30%.在此基础上,Carreira等人<sup>[64]</sup>提出了O2P(second-order pooling)方法.该方法在特征描述阶段采取二阶均值池化或二阶最大池化的方法,并且利用黎曼流形的几何特性,可以描述任意形状区域的特征.使用SIFT算子对局部特征进行加强,最终在PASCAL VOC 2011竞赛中显示出的分割准确率为47.6%.

Girshick等人<sup>[65]</sup>于2014年提出了R-CNN(regions with CNN)方法,用于图像目标检测和语义分割.该算法分为3个步骤:首先,从原始图像中使用selective search方法<sup>[66]</sup>抽取候选区域集;第2步,将每个区域变换为固定大小(227×227)的RGB图像,利用包含5个卷积层和2个全连接层的CNN提取每个区域的特征;最后,利用提取的特征以及训练标注,为每个对象类构造支持向量机(support vector machine,简称SVM)分类器.该方法借鉴了迁移学习的思想,首先利用ImageNet<sup>[67]</sup>中大量的图像进行预训练,再使用PASCAL VOC中的图像调优



(fine-tuning),最终在 PASCAL VOC 2011 竞赛中显示出的分割准确率为 47.9%.

2014 年,Hariharan 等人<sup>[68]</sup>提出了 SDS(simultaneous detection and segmentation)方法,该方法可用于图像中对象检测和语义分割两个任务.首先使用 MCG 算法<sup>[69]</sup>从每幅图像中生成候选区域集,每个区域是一个物体的边界框;之后,使用 CNN 从边界框和区域前景提取特征,将两部分特征联合训练 CNN;然后,利用这些特征,使用支持向量机对每个候选区分类;最后,使用非最大约束(non-maximum suppression,简称 NMS)将掩码和原始区域候选集融合提升分割效果.该算法在 PASCAL VOC 2011 竞赛中显示出的分割准确率为 52.6%.

在 SDS 算法的基础上,Liu 等人<sup>[70]</sup>提出了 MPA(multi-scale patch aggregation)算法,不使用区域生成算法,采用不同大小的滑动窗口对原图进行卷积,将图像重构为有重叠的多尺度特征图,并通过尺度对齐(scale alignment)将多尺度特征图映射到相同尺度,用来并行完成图像分类和分割任务.在 PASCAL VOC 2012 竞赛中,表现出的分割准确率为 62.1%.更为重要的是,由于未使用候选区生成算法,在相同的机器配置运行 SDS 算法和 MPA 算法分割一幅图像的时间分别为 60.1s 和 8.7s,速度取得了显著的提升.

R-CNN 中提取的候选区域之间相互重叠,重叠部分的特征被重复提取,非常耗时.He 等人<sup>[71]</sup>在 R-CNN 卷积层之后加入空间金字塔池化层(spatial pyramid pooling,简称 SPP),该层接收任意大小的输入图像并产生固定长度的特征向量,克服了文献[65]中区域变换的弊端.并且,将从原图中提取特征转换为从卷积图中提取特征,无需重复性地对重叠区域进行耗时的卷积操作,提升了算法效率.Girshick 等人<sup>[72]</sup>借鉴文献[71]的思想,提出了 Fast R-CNN,将候选区域映射到 CNN 最后一个卷积层的特征图上,确保每张图片只提取一次特征,提高了运行速度.

Fast R-CNN 改善了特征提取的速度,但是候选区域的生成速度仍然是算法性能的瓶颈.因此,Ren 等人<sup>[73]</sup>又提出了 Faster R-CNN,在 Fast R-CNN 的基础上增加一个区域建议网络(region proposal network,简称 RPN),共享全图卷积特征检测图中的目标,提高了候选区域的生成速度.实验结果表明:利用 K40 GPU,使用 selective search 从一幅图像生成候选区域时间为 1 510ms;而使用 RPN 方法仅需 10ms,极大地提升了图像分割算法的效率.

### 3.2 端到端的方法

基于区域分类的图像语义分割方法需要首先从图像中生成目标候选区域集,该区域生成质量的优劣直接影响到算法的效果,并且生成候选区域集合耗时较长,影响了整个算法的性能.因此近年来,一些学者提出一种更为直接的以像素直接分类为基础的语义分割方法.其基本思想是:以大量带有像素级标注的图像为样本,训练支持向量机(support vector machines,简称 SVM)<sup>[74]</sup>、神经网络等分类器,对图像中每个像素分类,其过程如图 5 所示.输入是原图像,利用真值(ground truth)作为监督信息训练分割系统,输出是源图像对应的语义分割图,中间过程对用户来说可以看作是黑盒,这是一种端到端的方法.

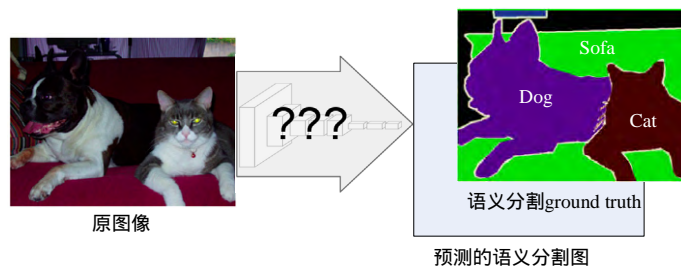


Fig.5 Demonstration for end to end image semantic segmentation

图 5 端到端的图像语义分割方法示意图

在端到端的方法中,分类器(如 SVM)被用于解决图像分割问题<sup>[75-77]</sup>,从图像中提取像素级特征进行逐像素分类,从而实现图像分割.然而,上述方法使用的特征是从图像不同通道(如 RGB,HSV 等)提取的,未考虑各通道之间的关系.因此,Wang 等人<sup>[78]</sup>提出了 QEM(quaternion exponent moments)方法.首先,使用 QEM 描述包括图像各颜色通道之间关系在内的像素特征;然后,使用孪生支持向量机(twin support vector machines,简称 TSVM)<sup>[79]</sup>

进行分类.QEM 算法的优势是对噪声、几何形变、颜色变化有很好的鲁棒性,并且 TSVM 分类器计算效率高,分类效果好.

Long 等人<sup>[80]</sup>于 2014 年提出了 FCN(fully convolutional network)方法,设计了一种针对任意大小的输入图像,训练端到端的全卷积网络的框架,实现逐像素分类,奠定了使用深度网络解决图像语义分割问题的基础框架.FCN 方法利用了 VGG 16 网络,VGG 16 网络具有 16 个卷积层、5 个最大池化层、3 个全连接层以及 1 个 softmax 层.FCN 将 3 个全连接层转换为卷积层,并移除 softmax 层,将原本用于全图分类的网络转换为生成图像分割的网络.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 62.2%.

为了克服卷积网络最后输出层缺少空间位置信息这一不足,Long 等人<sup>[80]</sup>通过双线性插值上采样和组合中间层输出的特征图,将粗糙(coarse)分割结果转换为密集(dense)分割结果.Eigen 等人<sup>[81]</sup>提出了一个包含两个卷积网络的系统模型,其中一个卷积网络专门用于对预测结果优化获取高精度的分割效果,将分割准确率提升到 62.6%.

Hariharan 等人<sup>[82]</sup>借用图像金字塔(image pyramid)的思想,通过将原图中指定位置对应 CNN 各个层次的输出特征图,组合成 hypercolumn 向量描述图像中统计结构特征,提升了图像分割的精细度,该方法也可用于物体检测和特征点预测等任务.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 62.6%.

Mostajabi 等人<sup>[83]</sup>认为:FCN 方法中使用固定大小的滤波器捕获的图像局部特征只能部分地反映图像内容信息,融合图像多个级别特征有助于提升分割效果.于 2015 年提出 Zoom-out 方法,其核心思想在于,从图像提取局部(local)、近端(proximal)、远端(distant)、场景(scene)这 4 个级别的特征组合成特征向量,克服了单一尺度感知野无法检测不同尺度目标的不足.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 64.4%.

Chen 等人<sup>[84]</sup>在 FCN 框架的末端增加全连接 CRF<sup>[85]</sup>,提出了 DeepLab 模型,首先使用双线性插值法对 FCN 的输出结果上采样得到粗糙分割结果,以该结果图中每个像素为一个节点构造 CRF 模型并求解,对分割结果求精.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 71.6%.

Zheng 等人<sup>[86]</sup>在文献[84]工作的基础上,将 CRF 建模为循环神经网络(recurrent neural network,简称 RNN).该网络可以直接使用 BP(back propagation)算法进行端到端的训练,不需要离线地分别训练 CNN 和 CRF 模型.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 72.0%.

Noh 等人<sup>[87]</sup>对 FCN 网络架构进行了改进,通过学习一个和 FCN 网络完全对称的解卷积网络,一方面可以检测到图像中不同级别尺度的目标实例,从而避免了 FCN 只能检测单一尺度语义目标的弊端;另一方面,通过解卷积层(deconvolution)和反池化层(unpooling)的结合,在输出的像素分类图中更好地反映了物体细节,得到了高质量的分割效果.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 72.5%.

Yu 等人<sup>[88]</sup>设计了一种专用于图像像素预测的网络,该网络的特殊之处在于其中不包含池化层,卷积层采用扩张卷积(dilated convolution)操作.扩张卷积是小波分解中常用的方法<sup>[89]</sup>,在普通卷积操作中引入不同的扩张因子,使感知野大小呈指数级增长,从而捕获图像中多尺度上下文信息并加以聚合,以提升像素预测的准确率.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 75.3%.

Chen 等人<sup>[90]</sup>以 FCN 为基础架构并引入注意力模型(attention model),将注意力模型应用于特征的尺度维度,即:对输入图像首先进行尺度变换,将变化后的图像并行作为 FCN 的输入,并针对每个尺度的图像学习对应的像素级权值图(weight map),这些权值图加权后用于 FCN 逐像素分类标注.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确率为 75.7%.

在文献[91]中,Chen 等人将 DeepLab 中的 3 个特征图层上采样为与原图具有相同的分辨率并进行级联,构建边界预测网络 EdgeNet,得到边界图,区域变换(domain transform)网络将 DeepLab 得到的粗糙分割结果结合边界图进行二维边界保留滤波操作,优化物体边缘位置.由于区域变换操作仅在 2 个维度滤波,而全连接 CRF 在 5 个维度双向滤波(位置空间 2 维,颜色空间 3 维),因此在速度上数倍于 DeepLab,并且在 PASCAL VOC 2012 竞赛中显示出的分割准确度达到 76.3%.

Ghiasi 等人<sup>[92]</sup>针对语义分割结果丢失空间位置信息的问题,基于拉普拉斯金字塔的思想,将每层得到的低

分辨率特征图通过重构的方式,即将特征图表示为一组各种类别物体基函数的线性组合,得到低频粗糙分割图.进而通过跨层方式(skip layer)<sup>[80]</sup>加入高频边界图的信息对分割图求精.该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确度为 76.8%.

Lin 等人<sup>[93]</sup>在图像语义分割中引入上下文信息,通过 CNN 建模 CRF 点对势能(pairwise potential)捕获相邻图像块之间的语义关系<sup>[94]</sup>,称为“区域-区域”上下文(patch-patch context),以此提升粗糙预测结果;以多尺度图像作为输入,通过卷积层得到多尺度特征图,使用滑动金字塔池化技术并连接各个尺度特征图,以此建模“区域-背景”上下文(patch-background context).该方法在 PASCAL VOC 2012 竞赛中显示出的分割准确度为 78.0%.

端到端的图像语义分割方法建立在有监督的机器学习算法基础之上,需要大量的像素级标注样本,根据文献[95]的研究,对图像进行像素级标注的耗时是标定图像中物体位置耗时的 15 倍,因此,像素级标注的图像难以大量获得,这就限制了端到端方法的应用范围.另一方面,ImageNet<sup>[67]</sup>,Microsoft COCO<sup>[95]</sup>以及 Pascal VOC<sup>[63]</sup>中提供了大量图像级标注和边界框标注的图像,这些弱标注信息可以用来辅助语义分割.

Kumar 等人<sup>[96]</sup>提出了一种从弱标签数据学习的算法,将边界框标注和图像级标注图像中缺失的图像分割信息看作是隐含变量,将参数学习问题用 LSVM(latent structural SVM)建模.首先利用原图像和边界框标注,通过极小化代价函数预测图像分割;然后,使用该分割结果定义图像类标注;最后,利用该标注学习模型参数.

Dai 等人<sup>[97]</sup>提出了 BoxSup 方法,使用边界框标注样本图像,以弱监督学习方式,迭代地自动生成候选区域并训练卷积网络.首先,使用文献[56,66,69]中的方法生成候选分割掩码,使用这些掩码进行有监督的网络训练,网络参数更新后,再利用新网络提升掩码精度,如此反复迭代.起初的掩码比较粗糙,随着迭代的进行逐渐变得准确,并能提供有用的信息用于网络训练.该弱监督学习算法在 PASCAL VOC 2012 竞赛中显示出的分割准确度为 66.2%,接近于有监督学习方法的效果.

Papandreou 等人<sup>[98]</sup>建立在 DeepLab 模型<sup>[84]</sup>之上,提出了一种针对图像级标注的弱监督学习算法.以原图像  $x$  以及图像级标注  $z$  作为已知条件,将图像的像素级分割信息  $y$  看作隐含(latent)变量,构建概率图模型并推导 EM(expectation maximization)算法学习模型参数  $\theta$ .在 E 步中,利用后验概率最大估算隐含变量  $y$ ;在 M 步中,使用随机梯度下降算法求解参数  $\theta$ .在此基础上,提出了 3 种利用边界框标注进行模型学习的算法.该弱监督学习算法在 PASCAL VOC 2012 竞赛中显示出的分割准确度为 62.2%,而以 10%强标注数据和 90%弱标注数据相结合的半监督学习将准确率提高到 73.9%,超过了当时最优的有监督学习的效果.

## 4 图像分割算法实验

本节介绍目前常用的检验图像分割算法性能的公用数据集,以及定量地衡量算法性能的各种评价指标,并对前述各种类型的典型图像分割算法进行了分析、总结和比较,最后,通过实验展示各种算法的分割效果.

### 4.1 实验用数据集

为了科学、一致地评价各类图像分割算法的性能,需要使用标准的图像数据集进行测试和对比,目前,常用的图像数据集见表 1,包括:

(1) PASCAL VOC<sup>[63]</sup>.PASCAL VOC 的全称为 Pattern Analysis,Statistical Modeling and Computational Learning Visual Object Classes,它提供了视觉对象分类和识别、图像分割、动作识别的标准图像标注数据集和平台.最初,该图像库只包含 4 个类别的图像,2006 年增加到了 10 个类,2007 年又扩充为 20 个类.最新的 PASCAL VOC 2012 包含 20 个类别,其中,用于图像分割任务的图像有 9 993 张.如今, PASCAL VOC 图像集已经成为了计算机视觉各领域最为常用的基准数据集;

(2) SBD<sup>[99]</sup>.SBD 的全称是 Stanford Background Dataset,是用于衡量语义场景理解方法性能的图像数据集.其图像是从 LabelMe,MSRC,PASCAL VOC 等公测数据集中抽取的 715 张图像,这些图像都是一些户外场景,图像尺寸接近 320×240 像素,每幅图像中至少包含一个前景目标;

(3) Caltech101<sup>[100]</sup>.该数据集中包括 101 个类别的物体,每种类别包含 40~800 张图像不等,大部分类别包含 50 张左右图像,该数据集后来发展为 Caltech 256,包含 256 个类别,共 30 608 张图像;

(4) BSDS<sup>[101]</sup>.BSDS 的全称为 Berkeley Segmentation Dataset,是一个自然图像数据集,用于比较不同分割算法和边界检测算法的性能.该数据集包含 500 张自然图像,每张图像都有人工标注的分割真值(ground truth).数据集由彼此不交叉的训练集、验证集和测试集这 3 部分组成;

(5) MSRC<sup>[102]</sup>.MSRC 是由微软剑桥研究院建立的用于图像场景理解、物体分割的数据集,其 V2 版本包含 23 类物体,共 591 张图像,每张图像有像素级的类别标注;

(6) SIFT Flow<sup>[103]</sup>.包含 8 种户外场景,其中大部分为户外场景,如街道、海滩、城市、山脉、建筑等,33 种语义类别目标以及 3 个地理类别目标,共 2 688 张图像,每张图像有像素级的类别标注.

Table 1 Common datasets for image segmentation

表 1 图像分割常用数据集

图像数据库	图像数	类别数	每类样本数	图像分辨率(宽度×高度)	图像标注级别
PASCAL VOC	9 993	20	N/A	(142~500)×(71~500)	像素级/边界框
SBD	715	8	N/A	320×240	像素级
Caltech101	9 144	101	31~800	不规则	边界框
Caltech256	30 608	256	80~827	不规则	边界框
BSDS	500	N/A	N/A	481×321 或 321×481	分割边界
MSRC V1	240	8	30	320×213 或 213×320	像素级
MSRC V2	591	23	21~30	320×213 或 213×320	像素级
SIFT Flow	2 688	33	N/A	256×256	像素级

表 1 列出了用于图像分割的常用图像数据库集,比较因素包括数据库名、图像数、类别数、每类样本数、图像分辨率、图像标注级别.其中,N/A 表示无法准确提供相关信息.图像分辨率中的“不规则”表示该数据集中图像的分辨率没有统一规格.

#### 4.2 图像分割算法的评价指标

通过对现有文献的整理研究,将衡量图像分割的指标分为基于分割边界的评价指标、基于分割区域的评价指标和基于像素的评价指标.

##### (1) 基于分割边界的评价指标

基于分割边界的评价指标主要是边界召回率(boundary recall,简称 BR),是衡量真值边界和算法生成的边界之间的一致性程度的指标<sup>[30]</sup>.算法的边界召回率越高,表示其生成的分割边界与真实边界越接近.分割  $s$  的边界召回率定义如下<sup>[38]</sup>:

$$BR(s) = \frac{\sum_{p \in B(g)} I(\min_{q \in B(s)} \|p - q\| < \varepsilon)}{Area(B(g))} \quad (8)$$

其中, $B(g)$ 和  $B(s)$ 分别为超像素边界真值和算法生成的超像素边界的集合. $I(\cdot)$ 为指示函数,如果算法生成的超像素中的边界像素位于超像素真值中的边界像素  $\varepsilon$ 范围之内,则返回 1;否则,返回 0. $Area(s)$ 表示集合  $s$  的面积.

##### (2) 基于分割区域的评价指标

基于分割区域的评价指标主要包括欠分割错误率、可达分割准确率、紧凑度等.欠分割错误率(undersegmentation error,简称 UE)用以衡量生成的超像素“溢出”到超像素真值区域的程度,算法生成的分割  $s$  的欠分割错误率定义为<sup>[30]</sup>

$$UE(s) = \frac{\sum_i \sum_{j: s_j \cap g_i \neq \emptyset} (Area(s_j) - Area(g_i))}{\sum_i Area(g_i)} \quad (9)$$

其中, $Area(s)$ 表示超像素  $s$  的面积, $g_i$ 为超像素真值, $s_k$ 为算法生成的超像素.

可达分割准确率(achievable segmentation accuracy,简称 ASA)给出了使用超像素方法进行预处理所能获得的最好的图像分割效果的性能,是用最终分割结果间接衡量超像素方法性能的指标,其定义为<sup>[30]</sup>

$$ASA(s) = \frac{\sum_j \max_i |s_j \cap g_i|}{\sum_i Area(g_i)} \quad (10)$$

紧凑度(compactness)是指每个超像素形状的规则程度、面积大小以及分割边界光滑程度<sup>[47]</sup>.在实际使用时,常用的做法是计算超像素形状和圆形的接近程度,令  $A_C$  和  $A_S$  分别表示周长相等的圆形和任意形状的超像素的面积,则超像素和圆的面积之比定义为

$$Q_S = \frac{A_S}{A_C} \quad (11)$$

对于一幅给定的图像,其紧凑度定义为

$$CO = \sum_{S \in O} Q_S \cdot \frac{Area(S)}{Area(I)} \quad (12)$$

其中,  $O$  代表图像  $I$  生成的超像素集合,  $Area(S)$  和  $Area(I)$  分别代表超像素  $S$  的面积和图像  $I$  的面积.

### (3) 基于像素的评价指标

基于像素的评价指标包括像素准确率、平均准确率、平均 IU、带权 IU 等<sup>[80]</sup>.令图像中包含物体的类别数为  $N$ ,  $n_{ij}$  表示实际类别为  $i$ 、预测类别为  $j$  的像素数量,  $t_i = \sum_j n_{ij}$  表示类别  $i$  中包含的像素总数.

像素准确率(pixel accuracy,简称 PA)是指图像中分割正确的像素数和图像中总像素数的比例,其定义为<sup>[80]</sup>

$$\frac{\sum_{i=1}^N n_{ii}}{\sum_{i=1}^N t_i} \quad (13)$$

平均准确率(mean accuracy,简称 MA)是指所有类别物体像素准确率的平均值,其定义为<sup>[80]</sup>

$$\frac{1}{N} \sum_{i=1}^N \frac{n_{ii}}{t_i} \quad (14)$$

平均 IU(mean intersection over union,简称 Mean IU)是指分割结果与真值的交集与其并集之比,定义为<sup>[80]</sup>

$$\frac{1}{N} \sum_{i=1}^N \frac{n_{ii}}{t_i + \sum_{j=1}^N n_{ji} - n_{ii}} \quad (15)$$

带权 IU(frequency weighted intersection over union,简称 FW IU)是在平均 IU 的基础上,考虑每个类在全图的权重进行加权平均的结果,其定义为<sup>[80]</sup>

$$\left( \sum_{k=1}^N t_k \right)^{-1} \sum_{i=1}^N \frac{t_i n_{ii}}{t_i + \sum_{j=1}^N n_{ji} - n_{ii}} \quad (16)$$

## 4.3 图像分割算法的分析与比较

基于内容的图像分割典型算法研究工作的分析与比较结果见表 2,主要比较因素包括算法分类、算法名称、发表年份、时间复杂度、算法测试用数据集、算法评价指标、算法特点以及算法是否需要标注的额外代价等.时间复杂度中的  $N$  表示图像中的像素数量,  $N/A$  表示相关论文中未提及该数据或该项无需此数据.

在基于图论的方法和基于像素聚类的方法中, NCut<sup>[17]</sup>, FH<sup>[36]</sup>, Graph Cuts<sup>[24]</sup>, Meanshift<sup>[43]</sup> 这几种经典算法已经实用化,且有标准实现.这些算法的时间复杂度普遍偏高.在此基础上的改进算法,如 SEEDS<sup>[30]</sup>, LSC<sup>[21]</sup>, SLIC<sup>[14]</sup> 等,无论是生成图像块的质量和算法的时间复杂度上都有更好的表现,实验结果表明<sup>[21]</sup>:在 Intel 3.4GHz 双核 CPU, 2GB 内存的计算机上,这 3 种算法分割一张分辨率为 481×321 像素的图像所需时间分别为 0.213s、0.314s 和 0.919s,基本能够满足实时性要求.

在语义分割方法中,大多数方法都放弃使用人工设计(hand-craft)特征算子,转而使用卷积网络为代表的自动特征提取技术.实验结果表明:将卷积网络用于图像分割,相比于常用经典算子,如 SIFT<sup>[104]</sup>, HOG<sup>[105]</sup> 等,效果有了质的提升,在 PASCAL VOC 2011 图像分割竞赛中,将准确率从 47.6%<sup>[68]</sup> 提升到 71.6%<sup>[84]</sup>.在算法的运行时间方面,基于候选区域的方法在生成候选区域及区域描述时耗时较多,所以实时性不强.而端到端的方法通过改造 AlexNet<sup>[106]</sup> 和 GoogLeNet<sup>[107]</sup> 等经典卷积网络,将图像分类的网络转换为逐像素预测的网络.虽然网络训练耗时较长,但在网络训练完成后,预测每张图像所需的时间为毫秒级.文献[80]的实验结果表明:SDS 方法分割一张图片耗时约为 50s,而 FCN 方法仅为 175ms.

对于像素级标注样本不足的情景,文献[97,98]提出了基于弱标注样本的处理方法和技术,使用改进的超像素生成方法从图像中提取候选区域,结合 Microsoft COCO<sup>[95]</sup>等图像数据集中提供的大量边界框标注区域进行半监督或弱监督学习,作为强标注样本学习的补充,在 PASCAL VOC 2012 中将分割准确率提升到了 73.9%<sup>[98]</sup>。

Table 2 Comparison of image segmentation algorithms

表 2 图像分割方法对比

分类	名称	发表年份	时间复杂度	数据集	评价指标	特点	额外代价
基于图论的分割方法	NCut <sup>[17]</sup>	2000	NP-hard	BSDS	UE/BR/ASA	利用谱聚类算法求解图像分割问题,生成的图像块数量不定,紧凑度不可控,算法时间复杂度高	无
	FH <sup>[36]</sup>	2004	$O(M\log N)$	MSRC	UE/BR	利用最小生成树算法对图像进行分割,生成的图像块边界贴合度较高,形状和大小不规则,算法稳定性不好	无
	Graph cuts <sup>[24]</sup>	2006	NP-hard	N/A	N/A	利用图论中最大流和最小割定理解决图像分割问题,但图像块数量和紧凑度不可控,算法时间复杂度高	无
	Superpixel lattice <sup>[28]</sup>	2008	$O(N^{3/2}\log N)$	BSDS	BR	迭代地对图像进行网格划分,生成的图像块形状规则、大小均匀,但算法过分依赖输入的边界代价图,时间复杂度较高	无
	SEEDS <sup>[30]</sup>	2012	$O(N)$	BSDS	UE/BR/ASA	对初始分割使用爬山法迭代求精,获得边界贴合度好的最终分割效果,算法性能优秀,时间复杂度低	无
	LSC <sup>[13]</sup>	2015	$O(N)$	BSDS	UE/BR/ASA	使用核函数将特征映射到高维空间,利用 $k$ -means 算法求解 $K$ 路 NCut 问题,算法性能优秀,时间复杂度低	无
基于像素聚类的方法	Meanshift <sup>[43]</sup>	2002	$O(N^2)$	N/A	N/A	利用模式搜索算法,通过将像素点聚类到不同模点实现图像分割,生成的图像块数量不定,紧凑度不可控,算法时间复杂度较高	无
	Medoidshift <sup>[45]</sup>	2007	$O(N^3)$	N/A	N/A		
	Quickshift <sup>[46]</sup>	2008	$O(N^2)$	Caltech101	UE/BR	利用几河流水平集方法,通过对初始区域的膨胀与演化生成大小均匀、紧凑度好的图像块	无
	Turbopixels <sup>[48]</sup>	2009	$O(N)$	BSDS	UE/BR	利用改进的 $k$ -means 算法对 5 维特征向量聚类,生成大小均匀、紧凑度好的图像块,算法时间复杂度低	无
基于候选区域的语义分割方法	SFGR <sup>[62]</sup>	2012	N/A	PASCAL VOC	Mean IU	提出此类方法的基本框架,使用 CPMC 生成候选区域,利用中层特征去除无效候选区域	图像级标注
	O2P <sup>[64]</sup>	2012	N/A	PASCAL VOC/Caltech101	Mean IU	使用二阶池化方法描述任意形状区域的特征,使用线性 SVM 分类区域,实现图像分割	图像级标注
	SDS <sup>[68]</sup>	2014	N/A	PASCAL VOC/SBD	Mean IU	使用MCG生成候选区域,利用CNN融合区域前景和边界框提取特征,利用NMS提升分割效果	图像级标注
	R-CNN <sup>[60]</sup>	2014	N/A	PASCAL VOC	Mean IU	使用selective search生成候选区域,利用迁移学习思想预训练网络,使用CNN提取图像区域特征	图像级标注
	Fast/Faster R-CNN <sup>[72,73]</sup>	2015	N/A	PASCAL VOC	Mean IU	使用 RPN 提升候选区域生成速度;通过共享卷积层特征图提升区域特征提取速度	图像级标注
	SPP <sup>[71]</sup>	2015	N/A	PASCAL VOC	Mean IU	使用 SPP 将任意大小的输入图像生成固定长度的特征向量,共享卷积层输出的特征图,提升图像分割速度	图像级标注

Table 2 Comparison of image segmentation algorithms (Continued)

表 2 图像分割方法对比(续)

分类	名称	发表年份	时间复杂度	数据集	评价指标	特点	额外代价
端到端的语义分割方法	FCN <sup>[80]</sup>	2014	N/A	PASCAL VOC/ SIFT Flow	PA/MA/ Mean IU/ FW IU	以 CNN 为基础,率先提出端到端的图像分割方法框架,融合图像局部和全局信息,实现像素级的语义分割	像素级标注
	DeepLab <sup>[84]</sup>	2014	N/A	PASCAL VOC	Mean IU	使用全连接 CRF 辅助 CNN 定位图中物体位置,优化图像分割效果	像素级标注
	RNN <sup>[86]</sup>	2015	N/A	PASCAL VOC	Mean IU	以 DeepLab 为基础,将 CRF 建模为 RNN,实现端到端的训练,提升图像分割精度	像素级标注
	Zoom-out <sup>[83]</sup>	2015	N/A	PASCAL VOC/ SBD	PA/MA/ Mean IU	组合图像中从小到大的多个视觉级别的特征,挖掘图像空间结构信息,辅助提升图像分割效果	像素级标注
	Hypercolumn <sup>[82]</sup>	2015	N/A	PASCAL VOC	Mean IU	将原图中指定位置多个层次的输出特征图组合成 hypercolumn 向量,提升图像分割的效果	像素级标注
	BoxSup <sup>[97]</sup>	2015	N/A	PASCAL VOC	Mean IU	利用边界框标注,使用弱监督学习方法实现图像语义分割	边界框级标注
	DeepLab-EM <sup>[98]</sup>	2015	N/A	PASCAL VOC	Mean IU	利用边界框和图像级标注,结合 EM 算法,使用弱监督和半监督方法实现图像语义分割	边界框级/图像级标注
	CRF-RNN <sup>[88]</sup>	2016	N/A	PASCAL VOC	Mean IU	采用扩张卷积操作,使感知野大小呈指数级增长得到多尺度信息并融合提升分割效果	像素级标注
	DeepLab-DT <sup>[91]</sup>	2016	N/A	PASCAL VOC	Mean IU	在 DeepLab 中使用区域变换代替全连接 CRF,提升算法精度与速度	像素级标注
	CRF-Attention <sup>[90]</sup>	2016	N/A	PASCAL VOC	Mean IU	引入注意力模型,计算图像各尺度特征权重,优化分割效果	像素级标注
	LRR <sup>[92]</sup>	2016	N/A	PASCAL VOC	Mean IU	采用拉普拉斯金字塔重构低分辨率分割图,提升语义分割精度	像素级标注
	Contextual infomation <sup>[93]</sup>	2016	N/A	PASCAL VOC/ SIFT Flow	PA/MA/ Mean IU	利用 CRF 对图像上下文信息建模,提升语义分割效果	像素级标注

#### 4.4 图像分割算法实验

为了进一步加深对算法的理解和认识,本节使用前述的一些典型算法进行图像分割对比实验.实验选取公测数据集中的图像作为对象,实验分为两部分:第 1 部分针对基于图论的方法和基于像素聚类的方法,第 2 部分针对语义分割方法.

实验 1 对比 7 种常用的超像素方法,包括 FH<sup>[36]</sup>,Meanshift<sup>[43]</sup>,Quickshift<sup>[46]</sup>,Turbopixels<sup>[48]</sup>,SEEDS<sup>[30]</sup>,SLIC<sup>[14]</sup>,LSC<sup>[21]</sup>,实验所用代码均来自于相关文献的公开代码库,实验效果如图 6 所示.图 6(h)所示 LSC 算法得到的分割边界贴合度最高,并且图像块大小和形状都比较规则.图 6(g)中,SLIC 算法分割的边界贴合度比 LSC 略逊,主要原因是 SLIC 是在局部空间搜索优化,而 LSC 是建立在全局最优函数的基础上.图 6(f)所示的 SEEDS 算法得到的图像边界贴合度也很好,但是超像素不够规则.Turbopixels 算法采用了水平集方法,生成的图像块规则性非常好,边界也比较贴近于真实分割边界.图 6(c)和图 6(d)所示 Meanshift 和 Quickshift 同属模式搜索算法,生成的图像块数量不可控,形状不规则.图 6(b)所示 FH 算法生成的分割图边界贴合度也很高,但是图像块的大小和规则性差异都很大.

实验 2 对比 8 种语义分割方法的分割效果,包括 SDS<sup>[68]</sup>,FCN<sup>[80]</sup>,DeepLab<sup>[84]</sup>,BoxSup<sup>[97]</sup>,Hypercolumn<sup>[82]</sup>,RNN<sup>[86]</sup>,DeepLab+EM<sup>[98]</sup>,DeepLab+DT<sup>[91]</sup>.图 7 显示了各种算法实验的实验效果,从中可以发现:SDS 是基于候选区域的方法,对于小区域分类易出现错误,因此易丢失图像中的较小区域,如图 7(c)中马腿部分所示;BoxSup 是一种弱监督方法,仅以边界框级标注训练网络,因此分割效果较差,如图 7(f)所示;其余方法均是 FCN 及其衍生方



法,以图 7(d)所示的 FCN 方法分割效果为基准,图 7(g)所示 Hypercolumn 方法结合多层次特征图,分割效果较 FCN 有所提升;图 7(e)所示使用 CRF 对输出进行二次优化,因此分割边缘更加精准;图 7(i)是用 DeepLab 结合半监督学习方法得到的分割效果;图 7(j)所示使用区域变化代替全连接 CRF 计算,虽然分割效果略差于图 7(e),算法运行速度却有大幅度提升;RNN 方法将 CRF 建模为 RNN,能够较好地分割出图中较小物体,如图 7(h)所示,灰色部分为远处的汽车。

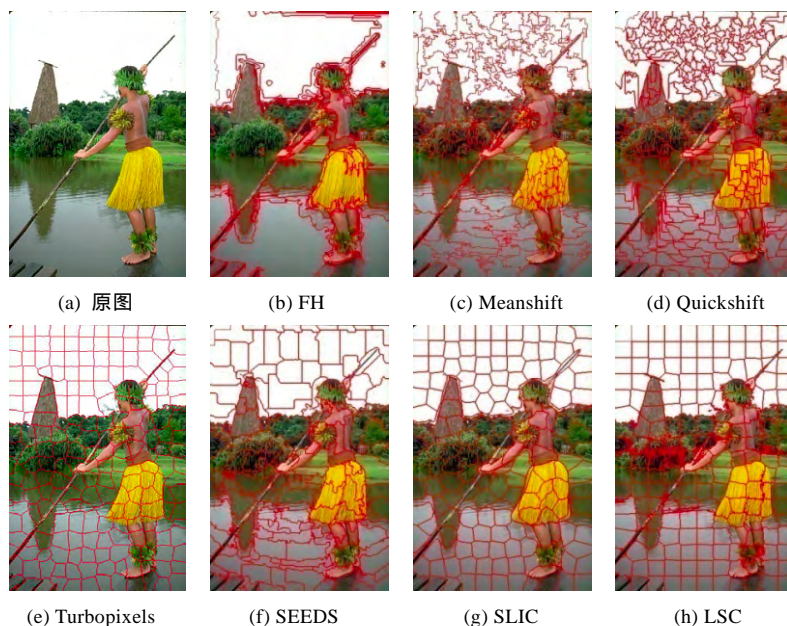


Fig.6 Comparison for superpixel segmentation experiments

图 6 超像素分割实验对比

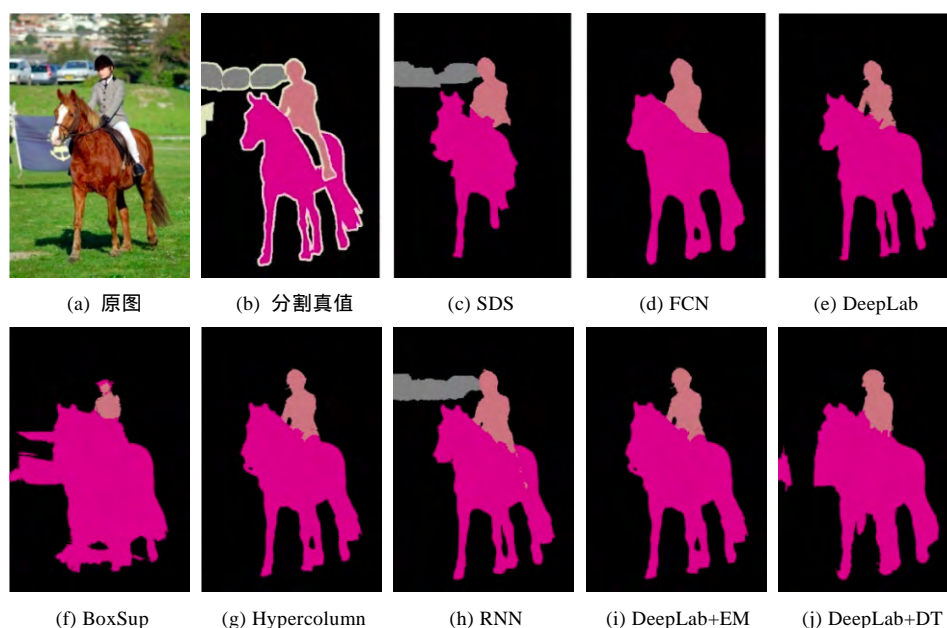


Fig.7 Comparison of semantic segmentation experiments

图 7 语义分割实验对比

## 5 总结和展望

如今,基于内容信息的图像分割方法已成为图像分割方法的主流,尤其是随着 2010 年以来,深度网络在计算机视觉领域的巨大成功,以卷积神经网络为基础的图像语义分割方法取得了突破性的进展,将图像分割带到了一个新的高度.本文对基于内容的图像分割方法进行了较为细致的梳理和分类,对基于图论的方法、基于像素聚类的方法和语义分割方法分别进行介绍,对每类方法中的代表性算法进行了研究、分析和对比,指出其特点.在现有研究成果的基础上,我们认为,在基于内容的图像分割领域,还存在以下的研究点.

### (1) 针对弱标注图像的语义分割方法.

在目前已有的图像公测数据库中,大部分样本均带有边界框、图像级或实例级(instance-level)的弱标注,具有像素级强标注的样本较少.另一方面,很多网站,如 Flickr 等,可由用户自己为图片添加标签,这些标签也是一种弱标注数据.如何利用好这些弱标注样本数据来提升图像分割效果,是当下的研究热点之一.尽管本文已经列举了一些弱监督或半监督的学习方法<sup>[87,96-98,108,109]</sup>,文献[110]仍然给出使用图像级标注结合弱监督学习算法训练语义来分割网络.文献[111]设计了由分类网络和分割网络组成的系统,利用集合的幂集运算扩充标注训练样本,实现弱监督方式的语义分割,但尚有进一步提升的空间.

### (2) 边界贴合度高,计算复杂度低的超像素生成算法.

研究表明,图像全局信息在人类认知中起到至关重要的作用.因此在基于超像素的图像分割算法中,应考虑以图像全局信息辅助局部信息,得到与物体真实边界贴合度高的分割结果.然而,全局信息的引入通常意味着构造更加复杂的代价函数,从而增加算法的复杂度.因此,如何平衡算法效果和算法复杂性之间的关系,具有一定的研究意义.

### (3) 在语义分割算法中实现实例级的分割.

本文介绍的语义分割方法可以将图像中不同类型的目标分开,但无法区分相同类型目标的不同实例.文献[112]借鉴了文献[113,114]中的思想,使用层次聚类方法构建分割树,通过求树的最优割,搜索图中的不同实例并加以标注.文献[115]在 FCN<sup>[80]</sup>的基础上,以弱监督学习方式提出了多实例学习(multiple instance learning,简称 MIL)框架.文献[116]构建了共享卷积特征由实例区分、掩码预测和目标分类这 3 部分组成的端到端级联网络,一定程度上解决了该问题.

### (4) 在特定应用场景的图像分割问题.

目前研究的图像分割算法主要针对户外场景的公测数据集,这些图像大多经过人工筛选和处理,与实际应用场景中的真实图像存在较大差异.文献[117,118]介绍了图像分割在医学图像领域的应用,文献[119,120]介绍了地质领域岩石薄片显微图像的分割,文献[121]介绍了遥感图像的分割.如何针对各种不同的应用开发行之有效的图像分割算法,是一项颇具挑战性的任务.

### (5) 应用于实时视频的图像分割算法.

目前,对图像分割算法的研究主要集中于静态图像,针对视频的研究很少.文献[122]使用 GPU 加速,提出了一种同步二维图像分割算法,并成功应用于视频目标跟踪,文献[123]提出一种针对移动摄像机的语义映射学习方法,用于 RGB-D 视频的实时语义分割,这是一些有益的尝试.但是这些方法仍是以二维静态图像的方法为基础,并未很好地利用视频中各帧之间的时序关系,因此在视频图像分割方面,亟需一些新的、开创性的工作引领未来的研究方向.

### (6) 基于多张图像的协同分割(CoSegmentation)算法<sup>[124,125]</sup>.

协同分割最早由 Rother 等人<sup>[126]</sup>提出,是指同时在一组图像中提取相同的前景目标,协同分割普遍比单张图片的分割效果要好.文献[127]提出一种内容敏感的图像集分割方法,通过少量用户交互实现协同分割,并迭代地对分割结果求精.文献[128]使用随机游走(random walker)法实现协同分割,并在优化过程中使用 GPU 加速.文献[129]提出一种样本驱动的半自动图像集协同分割方法,首先选取样本,以手工交互分割方式训练 SVM,对剩余样本计算超像素前景置信度,并使用图割算法自动分割.在协同分割算法中如何尽量减少甚至不需要用户交互、如何对包含两个以上前景目标的图像进行分割,都具有一定的研究意义.

## References:

- [1] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems Man and Cybernetics*, 1979,9(1):62–66. [doi: 10.1109/tsmc.1979.4310076]
- [2] Ziou D, Tabbone S. A study of edge detection techniques for segmentation computing approaches. *Int'l Journal of Computer Applications*, 2010,CASCT(1):35–41. [doi: 10.5120/993-25]
- [3] Adams R, Bischof L. Seeded region growing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994,16(6):641–647. [doi: 10.1109/34.295913]
- [4] Smeulders AW, Worring M, Santini S, Gupta S, Jain R. Content-Based image retrieval at the end of the early years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(12):1349–1380. [doi: 10.1109/34.895972]
- [5] Ren X, Malik J. Learning a classification model for segmentation. In: *Proc. of the 10th Int'l Conf. on Computer Vision*. IEEE, 2003. 10–17. [doi: 10.1109/iccv.2003.1238308]
- [6] Cheng MM, Zheng S, Lin WY, Vineet V, Sturges P, Crook N, Mitra J, Torr P. ImageSpirit: Verbal guided image parsing. *ACM Trans. on Graphics*, 2014,34(1):1–11. [doi: 10.1145/2682628]
- [7] Khan W. A survey: Image segmentation techniques. *Int'l Journal of Future Computer and Communication*, 2014,3(2):89–93. [doi: 10.7763/ijfcc.2014.V3.274]
- [8] Zuva T, Olugbara OO, Ojo SO, Ngwira SM. Image segmentation, available techniques, developments and open issues. *Canadian Journal on Image Processing and Computer Vision*, 2011,2(3):20–29.
- [9] Peng B, Zhang L, Zhang D. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 2013,46(3):1020–1038. [doi: 10.1016/j.patcog.2012.09.015]
- [10] Yi F, Moon I. Image segmentation: A survey of graph-cut methods. In: *Proc. of the 2012 Int'l Conf. on Systems and Informatics*. Washington: IEEE, 2012. 1936–1941. [doi: 10.1109/icsai.2012.6223428]
- [11] Liu ST, Yin FL. The basic principle and its new advances of image segmentation methods based on graph cuts. *ACTA AUTOMATICA SINICA*, 2012,38(6):911–922 (in Chinese with English abstract). [doi: 10.3724/sp.j.1004.2012.00911]
- [12] Gómez D, Yáñez J, Guada C, Rodríguez T, Javier M, Edwin Z. Fuzzy image segmentation based upon hierarchical clustering. *Knowledge-Based Systems*, 2015,87:26–37. [doi: 10.1016/j.knosys.2015.07.017]
- [13] Naz S, Majeed H, Irshad H. Image segmentation using fuzzy clustering: A survey. In: *Proc. of the 6th Int'l Conf. on Emerging Technologies*. IEEE, 2010. 181–186. [doi: 10.1109/icet.2010.5638492]
- [14] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012,34(11):2274–2282. [doi: 10.1109/tpami.2012.120]
- [15] Song XY, Zhou LL, Li ZG, Chen J, Zeng L, Yan B. Review on superpixel methods in image segmentation. *Journal of Image and Graphics*, 2015,20(5):599–608 (in Chinese with English abstract). [doi: 10.11834/jig.20150502]
- [16] Wu Z, Leahy R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1993,15(11):1101–1113. [doi: 10.1109/34.244673]
- [17] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(8):888–905. [doi: 10.1109/34.868688]
- [18] Xu HX, Tian Z, Ding MT. Multiscale segmentation for SAR image based on spectral clustering and mixture model. *Journal of Image and Graphics*, 2010,15(3):450–454 (in Chinese with English abstract). [doi: 10.11834/jig.20100315]
- [19] Sarkar S, Soundararajan P. Supervised learning of large perceptual organization: Graph spectral partitioning and learning automata. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(5):504–525. [doi: 10.1109/34.857006]
- [20] Ding CHQ, He X, Zha H, Gu M, Simon HD. A min-max cut algorithm for graph partitioning and data clustering. In: *Proc. of the 1st Int'l Conf. on Data Mining*. IEEE, 2001. 107–114. [doi: 10.1109/icdm.2001.989507]
- [21] Li Z, Chen J. Superpixel segmentation using linear spectral clustering. In: *Proc. of the 28th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2015. 1356–1363. [doi: 10.1109/cvpr.2015.7298741]
- [22] Geman S, Graffigne C. Markov random field image models and their applications to computer vision. In: *Proc. of the Int'l Congress of Mathematicians*. American Mathematical Society, 1986. 1496–1517. <http://www.mathunion.org/ICM/ICM1986.2/Main/icm1986.2.1496.1517.ocr.pdf>

- [23] Greig DM, Porteous BT, Seheult AH. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, 1989,51(2):271–279.
- [24] Boykov Y, Funka-Lea G. Graph cuts and efficient n-d image segmentation. *Int'l Journal of Computer Vision*, 2006,70(2):109–131. [doi: 10.1007/s11263-006-7934-5]
- [25] Rother C, Kolmogorov V, Blake A. GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics*, 2004,23(3):309–314. [doi: 10.1145/1186562.1015720]
- [26] Blake A, Rother C, Brown M, Perez P, Torr P. Interactive image segmentation using an adaptive gmmrf model. In: *Proc. of the 8th European Conf. on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2004. 428–441. [doi: 10.1007/978-3-540-24670-1\_33]
- [27] Liu L, Shi ZG, Su HR, Li H. Image segmentation based on higher order Markov random field. *Journal of Computer Research and Development*, 2013,50(9):1933–1942 (in Chinese with English abstract).
- [28] Moore AP, Prince SJD, Warrell J, Mohammed U, Jones G. Superpixel lattices. In: *Proc. of the 2008 IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2008. 1–8. [doi: 10.1109/cvpr.2008.4587471]
- [29] Moore AP, Prince SJD, Warrell J, Mohammed U, Jones G. Scene shape priors for superpixel segmentation. In: *Proc. of the 12th Int'l Conf. on Computer Vision*. IEEE, 2009. 771–778. [doi: 10.1109/ICCV.2009.5459246]
- [30] Van Den Bergh M, Boix X, Roig G, Capitanib D, Gool LV. SEEDS: Superpixels extracted via energy-driven sampling. In: *Proc. of the 12th European Conf. on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2012. 13–26. [doi: 10.1007/978-3-642-33786-4\_2]
- [31] Kruskal JB. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. of the American Mathematical Society*, 1956,7(1):48–50. [doi: 10.2307/2033241]
- [32] Prim RC. Shortest connection networks and some generalizations. *Bell Labs Technical Journal*, 2010,36(6):1389–1401. [doi: 10.1002/j.1538-7305.1957.tb01515.x]
- [33] Zahn CT. Graph-Theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. on Computers*, 1971,20(1):68–86. [doi: 10.1109/t-c.1971.223083]
- [34] Morris OJ, Lee MDJ, Constantinides AG. Graph theory for image analysis: An approach based on the shortest spanning tree. *Communications, Radar and Signal Processing, IEE Proc. F*, 1986,133(2):146–152. [doi: 10.1049/ip-f-1.1986.0025]
- [35] Kwok SH, Constantinides AG. A fast recursive shortest spanning tree for image segmentation and edge detection. *IEEE Trans. on Image Processing*, 1997,6(2):328–332. [doi: 10.1109/83.551705]
- [36] Felzenszwalb PF, Huttenlocher DP. Efficient graph-based image segmentation. *Int'l Journal of Computer Vision*, 2004,59(2):167–181. [doi: 10.1023/b:visi.0000022288.19776.77]
- [37] Meng QT, Gong SR, Liu CP, Wang ZH. Graph-Based color-texture region segmentation. *Journal of Image and Graphics*, 2009,14(10):2092–2096 (in Chinese with English abstract). [doi: 10.11834/jig.20091027]
- [38] Ye W, Wang YJ. MST image segmentation based on Mumford-Shah theory. *Journal of Computer-Aided Design & Computer Graphics*, 2009,21(8):1127–1133 (in Chinese with English abstract).
- [39] Fukunaga K, Hostetler L. The estimation of the gradient of a density function. *IEEE Trans. on Information Theory*, 1975,21(1):32–40. [doi: 10.1109/tit.1975.1055330]
- [40] Cheng Y. Mean shift, mode seeking, and clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1995,17(8):790–799. [doi: 10.1109/34.400568]
- [41] Li XR, Wu FC, Hu ZY. Convergence of a Mean Shift algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2005,16(3):365–374 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/20050305.htm>
- [42] Wen ZQ, Cai ZX. Convergence analysis of Mean Shift algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2007,18(2):205–212 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/20070204.htm>
- [43] Comaniciu D, Meer P. Mean Shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002,24(5):603–619. [doi: 10.1109/34.1000236]
- [44] Li HY, Wu SP. Parallelization of Mean Shift image segmentation algorithm. *Journal of Image and Graphics*, 2013,18(12):1610–1619 (in Chinese with English abstract). [doi: 10.11834/jig.20131209]

- [45] Sheikh YA, Khan EA, Kanade T. Mode-Seeking by medoidshifts. In: Proc. of the 12th IEEE Int'l Conf. on Computer Vision. IEEE, 2007. 1–8. [doi: 10.1109/icc.2007.4408978]
- [46] Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. In: Proc. of the 10th European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2008. 705–718. [doi: 10.1007/978-3-540-88693-8\_52]
- [47] Schick A, Fischer M, Stiefelhagen R. Measuring and evaluating the compactness of superpixels. In: Proc. of the 21st Int'l Conf. on Pattern Recognition. IEEE, 2012. 930–934. <http://ieeexplore.ieee.org/document/6460287/>
- [48] Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ, Siddiqi K. Turbopixels: Fast superpixels using geometric flows. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(12):2290–2297. [doi: 10.1109/tpami.2009.96]
- [49] Xiang S, Pan C, Nie F, Zhang CS. Turbopixel segmentation using eigen-images. IEEE Trans. on Image Processing, 2010,19(11):3024–3034. [doi: 10.1109/tip.2010.2052268]
- [50] Cigla C, Alatan AA. Efficient graph-based image segmentation via speeded-up turbo pixels. In: Proc. of the 17th Int'l Conf. on Image Processing. IEEE, 2010. 3013–3016. [doi: 10.1109/icip.2010.5653963]
- [51] Ren CY, Reid I. gSLIC: A real-time implementation of SLIC superpixel segmentation [Ph.D. Thesis]. Oxford: Department of Engineering, University of Oxford, 2011.
- [52] Sun W, Guo M. Image segmentation based on SLIC and conditional random field. Application Research of Computers, 2015,32(12):3817–3820 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2015.12.069]
- [53] Csurka G, Perronnin F. An efficient approach to semantic segmentation. Int'l Journal of Computer Vision, 2011,95(2):198–212. [doi: 10.1007/s11263-010-0344-8]
- [54] Mori G, Ren X, Efros AA, Malik J. Recovering human body configurations: Combining segmentation and recognition. In: Proc. of the 2004 IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2004. 326–333. [doi: 10.1109/cvpr.2004.1315182]
- [55] Todorovic S, Ahuja N. Learning subcategory relevances for category recognition. In: Proc. of the IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2008. 1–8. [doi: 10.1109/cvpr.2008.4587366]
- [56] Carreira JO, Sminchisescu C. CPMC: Automatic object segmentation using constrained parametric min-cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(7):1312–28. [doi: 10.1109/tpami.2011.231]
- [57] Shotton J, Winn J, Rother C, Criminisi A. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proc. of the 9th European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2006. 1–15. [doi: 10.1007/11744023\_1]
- [58] He X, Zemel RS, Carreira-Perpi MA. Multiscale conditional random fields for image labeling. In: Proc. of the 22nd IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2004. 695–702. [doi: 10.1109/cvpr.2004.1315232]
- [59] Fulkerson B, Vedaldi A, Soatto S. Class segmentation and object localization with superpixel neighborhoods. In: Proc. of the 12th Int'l Conf. on Computer Vision. IEEE, 2009. 670–677. [doi: 10.1109/icc.2009.5459175]
- [60] Gonfaus JM, Boix X, Van DW, Bagdanov AD, Serrat J, Gonzalez J. Harmony potentials for joint classification and segmentation. In: Proc. of the 22th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2010. 3280–3287. [doi: 10.1109/cvpr.2010.5540048]
- [61] Zhang JH. Semantic segmentation and annotation of image based on self-organizing feature map (SOFM) [MS. Thesis]. Shanghai: Shanghai Jiaotong University, 2010 (in Chinese with English abstract).
- [62] Carreira J, Li F, Sminchisescu C. Object recognition by sequential figure-ground ranking. Int'l Journal of Computer Vision, 2012, 98(3):243–262. [doi: 10.1007/s11263-011-0507-2]
- [63] Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VoC) challenge. Int'l Journal of Computer Vision, 2010,88(2):303–338. [doi: 10.1007/s11263-009-0275-4]
- [64] Carreira J, Rui C, Batista J, Sminchisescu C. Semantic segmentation with second-order pooling. In: Proc. of the 12th European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2012. 430–443. [doi: 10.1007/978-3-642-33786-4\_32]
- [65] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc. of the 27th IEEE Computer Society Conf. on CVPR. Washington: IEEE Computer Society, 2014. 580–587. [doi: 10.1109/cvpr.2014.81]

- [66] Uijlings JRR, Sande KEAVD, Gevers T, Smeulders AWM. Selective search for object recognition. *Int'l Journal of Computer Vision*, 2013,104(2):154–171. [doi: 10.1007/s11263-013-0620-5]
- [67] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Li FF. Imagenet large scale visual recognition challenge. *Int'l Journal of Computer Vision*, 2015,115(3):211–252. [doi: 10.1007/s11263-015-0816-y]
- [68] Hariharan B, Arbel EZP, Girshick R, Malik J. Simultaneous detection and segmentation. In: *Proc. of the 13th European Conf. on Computer Vision*. Springer Int'l Publishing, 2014. 297–312. [doi: 10.1007/978-3-319-10584-0\_20]
- [69] Arbelaez P, Ponttuset J, Barron J, Marques F, Malik J. Multiscale combinatorial grouping. In: *Proc. of the 27th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2014. 328–335. [doi: 10.1109/cvpr.2014.49]
- [70] Liu S, Qi X, Shi J, Zhang H, Jia J. Multi-Scale patch aggregation (MPA) for simultaneous detection and segmentation. In: *Proc. of the 29th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2016. 3141–3149.
- [71] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Proc. of the 13th European Conf. on Computer Vision*. Springer Int'l Publishing, 2014. 346–361. [doi: 10.1007/978-3-319-10578-9\_23]
- [72] Girshick R. Fast r-CNN. In: *Proc. of the 16th Int'l Conf. on Computer Vision*. IEEE, 2015. 1440–1448. [doi: 10.1109/iccv.2015.169]
- [73] Ren S, He K, Girshick R, Sun J. Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2015. 1–9. [doi: 10.1109/tpami.2016.2577031]
- [74] Vapnik VN. *Statistical Learning Theory*. New York: Wiley, 1998.
- [75] Bai X, Wang W. Saliency-SVM: An automatic approach for image segmentation. *Neurocomputing*, 2014,136(8):243–255. [doi: 10.1016/j.neucom.2014.01.008]
- [76] Yu Z, Wong HS, Wen G. A modified support vector machine and its application to image segmentation. *Image and Vision Computing*, 2011,29(1):29–40. [doi: 10.1016/j.imavis.2010.08.003]
- [77] Senyukova OV. Segmentation of blurred objects by classification of isolabel contours. *Pattern Recognition*, 2014,47(12):3881–3889. [doi: 10.1016/j.patcog.2014.06.007]
- [78] Wang XY, Wu ZF, Chen L, Zheng HL, Yang HY. Pixel classification based color image segmentation using quaternion exponent moments. *Neural Networks*, 2015,74:1–13. [doi: 10.1016/j.neunet.2015.10.012]
- [79] Khemchandani R, Jayadeva, Chandra S. Twin support vector machines for pattern classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007,29(5):905–910. [doi: 10.1109/tpami.2007.1068]
- [80] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proc. of the 28th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2015. 1337–1342. [doi: 10.1109/cvpr.2015.7298965]
- [81] Eigen D, Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *Proc. of the 16th Int'l Conf. on Computer Vision*. IEEE, 2015. 2650–2658. [doi: 10.1109/iccv.2015.304]
- [82] Hariharan B, Arbelaez P, Girshick R, Malik J. Hypercolumns for object segmentation and fine-grained localization. In: *Proc. of the 28th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2015. 447–456. [doi: 10.1109/cvpr.2015.7298642]
- [83] Mostajabi M, Yadollahpour P, Shakhnarovich G. Feedforward semantic segmentation with zoom-out features. In: *Proc. of the 28th IEEE Conf. on CVPR*. Washington: IEEE Computer Society, 2015. 3376–3385. [doi: 10.1109/cvpr.2015.7298959]
- [84] Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A. Semantic image segmentation with deep convolutional nets and fully connected crfs. In: *Proc. of the 3rd Int'l Conf. on Learning Representations*. 2015. <https://arxiv.org/abs/1412.7062>
- [85] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFS with gaussian edge potentials. In: *Proc. of the Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2012. 109–117. <http://papers.nips.cc/paper/4296-efficient-inference-in-fully-connected-crfs-with-gaussian-edge-potentials>
- [86] Zheng S, Jayasumana S, Romeraparedes B, Vineet V, Su Z, Du D, Huang C, Torr PHS. Conditional random fields as recurrent neural networks. In: *Proc. of the 16th Int'l Conf. on Computer Vision*. IEEE, 2015. 1529–1537. [doi: 10.1109/iccv.2015.179]
- [87] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. In: *Proc. of the 16th Int'l Conf. on Computer Vision*. IEEE, 2015. 1520–1528. [doi: 10.1109/iccv.2015.178]

- [88] Yu F, Koltun V. Multi-Scale context aggregation by dilated convolutions. In: Proc. of the 4th Int'l Conf. on Learning Representations. 2016. <https://arxiv.org/abs/1511.07122>
- [89] Shensa MJ. The discrete wavelet transform: Wedding the a trous and mallat algorithms. IEEE Trans. on Signal Processing, 1992, 40(10):2464–2482. [doi: 10.1109/78.157290]
- [90] Chen LC, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: Scale-Aware semantic image segmentation. In: Proc. of the 29th IEEE Computer Society Conf. on CVPR. Washington: IEEE Computer Society, 2016. 3640–3649. [doi: 10.1109/CVPR.2016.396]
- [91] Chen L, Barron JT, Papandreou G, Murphy K, Yuille AL. Semantic image segmentation with task-specific edge detection using cnns and a discriminatively trained domain transform. In: Proc. of the 29th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2016. 4545–4554. [doi: 10.1109/CVPR.2016.492]
- [92] Ghiasi G, Fowlkes C. Laplacian pyramid reconstruction and refinement for semantic segmentation. In: Proc. of the 14th European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2016. 519–534. [doi: 10.1007/978-3-319-46487-9\_32]
- [93] Lin G, Shen C, Reid I. Efficient piecewise training of deep structured models for semantic segmentation. In: Proc. of the 29th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2016. 3194–3203. [doi: 10.1109/CVPR.2016.348]
- [94] Kolesnikov A, Guillaumin M, Ferrari V, Lampert CH. Closed-Form approximate CRF training for scalable image segmentation. In: Proc. of the 13th European Conf. on Computer Vision. Springer Int'l Publishing, 2014. 550–565. [doi: 10.1007/978-3-319-10578-9\_36]
- [95] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, Zitnick L. Microsoft COCO: Common objects in context. In: Proc. of the 13th European Conf. on Computer Vision. Springer Int'l Publishing, 2014. 740–755. [doi: 10.1007/978-3-319-10602-1\_48]
- [96] Kumar MP, Turki H, Preston D, Koller D. Learning specific-class segmentation from diverse data. In: Proc. of the 14th Int'l Conf. on Computer Vision. IEEE, 2011. 1800–1807. [doi: 10.1109/iccv.2011.6126446]
- [97] Dai J, He K, Sun J. BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proc. of the 16th Int'l Conf. on Computer Vision. IEEE, 2015. 1635–1643. [doi: 10.1109/iccv.2015.191]
- [98] Papandreou G, Chen LC, Murphy KP, Yuille AL. Weakly-And semi-supervised learning of a deep convolutional network for semantic image segmentation. In: Proc. of the 16th Int'l Conf. on Computer Vision. IEEE, 2015. 1742–1750. [doi: 10.1109/iccv.2015.203]
- [99] Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions. In: Proc. of the 13th Int'l Conf. on Computer Vision. IEEE, 2009. 1–8. [doi: 10.1109/iccv.2009.5459211]
- [100] Li FF, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding, 2007,106(1):59–70. [doi: 10.1016/j.cviu.2005.09.012]
- [101] Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application toevaluating segmentation algorithms and measuring ecological statistics. In: Proc. of the 9th Int'l Conf. on Computer Vision. 2001. 416–423. [doi: 10.1109/iccv.2001.937655]
- [102] Wang X, Du J, Wu S, Li X, Li F. Cluster ensemble-based image segmentation. Int'l Journal of Advanced Robotic Systems, 2013, 10(4):261–270. [doi: 10.5772/56769]
- [103] Liu C, Yuen J, Torralba A. Nonparametric scene parsing via label transfer. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,33(12):2368–2382. [doi: 10.1109/cvprw.2009.5206536]
- [104] Lowe DG. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004,60(2):91–110. [doi: 10.1023/b:visi.0000029664.99615.94]
- [105] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the 2005 IEEE Int'l Conf. on Computer Vision. Washington: IEEE Computer Society, 2005. 886–893. [doi: 10.1109/cvpr.2005.177]
- [106] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2012. 1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>



- [107] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proc. of the 28th IEEE Computer Society Conf. on CVPR. Washington: IEEE Computer Society, 2015. 1–9. [doi: 10.1109/cvpr.2015.7298594]
- [108] Xia W, Domokos C, Dong J, Cheong LF, Yan S. Semantic segmentation without annotating segments. In: Proc. of the 15th IEEE Int'l Conf. on Computer Vision. IEEE, 2013. 2176–2183. [doi: 10.1109/icc.2013.271]
- [109] Zhu J, Mao J, Yuille A. Learning from weakly supervised data by the expectation loss SVM (e-SVM) algorithm. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2014. 1125–1133. <http://papers.nips.cc/paper/5287-learning-from-weakly-supervised-data-by-the-expectation-loss-svm-e-svm-algorithm>
- [110] Pinheiro PO, Collobert R. From image-level to pixel-level labeling with convolutional networks. In: Proc. of the 28th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2015. 1713–1721. [doi: 10.1109/cvpr.2015.7298780]
- [111] Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2015. 1495–1503. <http://papers.nips.cc/paper/5858-decoupled-deep-neural-network-for-semi-supervised-semantic-segmentation>
- [112] Silberman N, Sontag D, Fergus R. Instance segmentation of indoor scenes using a coverage loss. In: Proc. of the 13th European Conf. on Computer Vision. Springer Int'l Publishing, 2014. 616–631. [doi: 10.1007/978-3-319-10590-1\_40]
- [113] Ladick LU, Russell C, Kohli P, Torr PHS. Associative hierarchical CRFS for object class image segmentation. In: Proc. of the 13th Int'l Conf. on Computer Vision. IEEE, 2009. 739–746. [doi: 10.1109/icc.2009.5459248]
- [114] Lempitsky V, Vedaldi A, Zisserman A. A pylon model for semantic segmentation. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2011. 1485–1493. <http://papers.nips.cc/paper/4201-ylon-model-for-semantic-segmentation>
- [115] Chen YT, Liu X, Yang MH. Multi-Instance object segmentation with occlusion handling. In: Proc. of the 29th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2016. 3470–3478. [doi: 10.1109/CVPR.2015.7298969]
- [116] Dai J, He K, Sun J. Instance-Aware semantic segmentation via multi-task network cascades. In: Proc. of the 29th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2016. 3150–3158. [doi: 10.1109/CVPR.2016.343]
- [117] Yang X, Cheng KTT, Chien A. Accurate vessel segmentation with progressive contrast enhancement and canny refinement. In: Proc. of the 2015 Asian Conf. on Computer Vision. Springer Int'l Publishing, 2014. 1–16. [doi: 10.1007/978-3-319-16811-1\_1]
- [118] Dan CC, Giusti A, Gambardella LM, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2012. 2852–2860. <http://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images>
- [119] Jungmann M, Pape H, Wißkirchen P, Clauser C, Berlage T. Segmentation of thin section images for grain size analysis using region competition and edge-weighted region merging. Computers & Geosciences, 2014,72:33–48. [doi: 10.1016/j.cageo.2014.07.002]
- [120] Filho IM. Segmentation of sandstone thin section images with separation of touching grains using optimum path forest operators. Computers & Geosciences, 2013,57(4):146–157. [doi: 10.1016/j.cageo.2013.04.011]
- [121] Li Z, Fang H. Multi-Scale and multi-feature segmentation of high resolution remote sensing Image. Journal of Multimedia, 2014, 9(7):948–954. [doi: 10.4304/jmm.9.7.948-954]
- [122] Prisacariu VA, Reid ID. PWP3D: Real-Time segmentation and tracking of 3D objects. Int'l Journal of Computer Vision, 2012, 98(3):335–354. [doi: 10.1007/s11263-011-0514-3]
- [123] STückler J, Waldvogel B, Schulz H, Behnke S. Dense real-time mapping of object-class semantics from RGB-D video. Journal of Real-Time Image Processing, 2014,10(4):599–609. [doi: 10.1007/s11554-013-0379-5]
- [124] Wang C, Guo Y, Zhu J, Wang L. Video object co-segmentation via subspace clustering and quadratic pseudo-boolean optimization in an mrf framework. IEEE Trans. on Multimedia, 2014,16(4):903–916. [doi: 10.1109/tmm.2014.2306393]
- [125] Wang L, Xia T, Guo Y, Liu L, Wang J. Confidence-Driven image co-matting. Computers & Graphics, 2014,38:131–139. [doi: 10.1016/j.cag.2013.10.014]

- [126] Rother C, Minka T, Blake A, Kolmogorov V. Cosegmentation of image pairs by histogram matching—Incorporating a global constraint into MRFS. In: Proc. of the 2006 IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2006. 993–1000. [doi: 10.1109/cvpr.2006.91]
- [127] Fu Y, Guo Y. Content-Sensitive collection snapping. In: Proc. of the 2011 Int'l Conf. on Multimedia and Expo. IEEE, 2011. 1–6. [doi: 10.1109/icme.2011.6011876]
- [128] Grady L. Random walks based multi-image segmentation: Quasiconvexity results and GPU-based solutions. In: Proc. of the 25th IEEE Conf. on CVPR. Washington: IEEE Computer Society, 2012. 1656–1663. [doi: 10.1109/cvpr.2012.6247859]
- [129] Wang LB, Guo YW, Xia TC, Jin GP. Example-Driven semi-automatic image collection segmentation. Journal of Computer-Aided Design & Computer Graphics, 2013,25(6):794–801 (in Chinese with English abstract).

#### 附中文参考文献:

- [11] 刘松涛,殷福亮.基于图割的图像分割方法及其新进展.自动化学报,2012,38(6):911–922. [doi: 10.3742/sp.j.1004.2012.00911]
- [15] 宋熙煜,周利莉,李中国,陈健,曾磊,闫钺.图像分割中的超像素方法研究综述.中国图像图形学报,2015,20(5):599–608. [doi: 10.11834/jig.20150502]
- [18] 徐海霞,田铮,丁明涛.基于谱聚类与混合模型的 SAR 图像多尺度分割.中国图像图形学报,2010,15(3):450–454. [doi: 10.11834/jig.20100315]
- [27] 刘磊,石志国,宿浩茹,李红.基于高阶马尔可夫随机场的图像分割.计算机研究与发展,2013,50(9):1933–1942.
- [37] 孟庆涛,龚声蓉,刘纯平,王朝晖.一种基于图的颜色纹理区域分割方法.中国图像图形学报,2009,14(10):2092–2096. [doi: 10.11834/jig.20091027]
- [38] 叶伟,王远军.基于 Mumford-Shah 理论的最小生成树图像分割方法.计算机辅助设计与图形学学报,2009,21(8):1127–1133.
- [41] 李乡儒,吴福朝,胡占义.均值漂移算法的收敛性.软件学报,2005,16(3):365–374. <http://www.jos.org.cn/1000-9825/20050305.htm>
- [42] 文志强,蔡自兴. Mean Shift 算法的收敛性分析.软件学报,2007,18(2):205–212. <http://www.jos.org.cn/1000-9825/20070204.htm>
- [44] 李宏益,吴素萍. Mean Shift 图像分割算法的并行化.中国图像图形学报,2013,18(12):1610–1619. [doi: 10.11834/jig.20131209]
- [52] 孙巍,郭敏.基于 SLIC 与条件随机场的图像分割算法.计算机应用研究,2015,32(12):3817–3820. [doi: 10.3969/j.issn.1001-3695.2015.12.069]
- [61] 张继昊.基于 SOFM 自组织特征映射网络的图像语义分割与标识[硕士学位论文].上海:上海交通大学,2010.
- [129] 汪戴波,郭延文,夏天辰,金国平.样本驱动的非自动图像集前背景分割.计算机辅助设计与图形学学报,2013,25(6):794–801.



姜枫(1980 - ),男,江苏泰州人,博士生,副教授,CCF 专业会员,主要研究领域为机器学习,计算机视觉.



李娜(1991 - ),女,硕士生,主要研究领域为机器学习,数据挖掘,图像处理.



顾庆(1972 - ),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为分布计算与并行处理,图像处理.



郭延文(1980 - ),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为计算机图形学,图像视频等多媒体信息处理,计算机视觉.



郝慧珍(1974 - ),女,讲师,CCF 学生会员,主要研究领域为人工智能,计算机视觉.



陈道蓄(1947 - ),男,教授,博士生导师,CCF 会士,主要研究领域为分布计算与并行处理.