

# 图像分割方法综述研究\*

周莉莉, 姜枫

(南京理工大学泰州科技学院 移动互联网学院, 江苏 泰州 225300)

**摘要:** 图像分割是计算机视觉领域重要和基础性的问题,也是颇具挑战性的任务。为了解该问题的研究现状、存在问题及发展前景,在广泛调研现有文献和最新成果的基础上,针对2000年之后主流的图像分割方法进行了研究,将之分为四类:基于图论的方法、基于聚类的方法、基于分类的方法以及结合聚类和分类的方法,对每类方法所包含的典型算法,尤其是该领域最近几年发表的最新文章的基本思想、优缺点进行介绍和分析。最后介绍了图像分割常用的基准数据集和算法评价指标,对比各种算法并总结全文,对未来可能的发展趋势进行了展望。

**关键词:** 图像分割; 图论; 聚类; 分类

**中图分类号:** TP391.4    **文献标志码:** A    **文章编号:** 1001-3695(2017)07-0-0

**doi:**10.3969/j.issn.1001-3695.2017.07.001

## Survey on image segmentation methods

Zhou Lili, Jiang Feng

(College of Mobile Internet, Taizhou Institute of Sci. & Tech., NUST, Taizhou Jiangsu 225300, China)

**Abstract:** Image segmentation is an important and fundamental problem in computer vision, meanwhile it's a challenging task. In order to find out the state-of-the-art, main problems and future trends of image segmentation, this paper introduced the mainstream image segmentation methods after 2000 on the basis of extensive research on the existing literatures and the latest achievements. These methods are categorized into four classes: graph theory based methods, clustering based methods, classification based methods, and hybrid methods of clustering and classification. The basic ideas, advantage and disadvantage of typical algorithms belong to each category, especially the most recently published papers were introduced and analyzed. Finally, this paper introduced the datasets which were commonly used as benchmark and evaluation metrics, compared all the algorithms, summarized the work and forecasts some potential future research work.

**Key words:** image segmentation; graph theory; clustering; classification

## 0 引言

图像分割是指将图像分成若干互不重叠的子区域,使得同一个子区域内的特征具有一定相似性,不同子区域间特征呈现较为明显的差异。图像分割是图像识别、场景解析、对象检测等任务的预处理,是计算机视觉中一项基础的任务。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>[3,4]</sup>等。2000年左右,开始出现基于超像素的图像分割方法<sup>[5]</sup>,这种处理方式将具有相似特征的像素分组,使图像块包含单个像素所不具备的图像内容信息,并提高了后续处理任务的效率。根据算法实现原理不同,超像素方法可以分为基于图论的方法和基于聚类的方法。前者的主要思想是将图像映射为带权无向图,图像的像素对应图的顶点,像素信息对应顶点属性,像素之间的相似性(或差异性)对应边的权值,将图像分割问题转换为图的顶点标注问题。后者主要思想是根据图像中的单个像素及像素之间的相互信息,如颜色、亮度、纹理等,利用数据挖掘中聚类算法,将具有相近特征的相邻像素聚到同一个图像块。

近年来,分类技术在物体检测、图像分类、目标识别等计算

机视觉领域得到了广泛的应用,并取得了巨大的成功,在图像分割方面也进行了一些有益的尝试。基于分类的方法将图像分割问题看作是图像单个像素的分类问题,先以有标注的图像作为训练样本,训练支持向量机(support vector machine, SVM)<sup>[6]</sup>、逻辑回归(logistic regression, LR)、神经网络(neural network, NN)等分类器,再以训练好的分类器对输入图像进行逐像素分类,根据像素分类情况得到图像分割结果。

分类方法是一种有监督的学习算法,需要大量的标注图像作为样本训练分类器,但是对图像进行像素级的标注是一件费时费力的工作,因此这样的样本往往数量不多。而基于聚类的图像分割方法是一种无监督的算法,近几年有人提出了通过结合聚类和分类算法进行图像分割的方法。其算法思路为,首先使用聚类算法从图像中生成初始化目标候选区域集,接着对区域内容进行描述和分类,最后根据区域分类结果构建全图标注,完成图像分割。

相比于其他图像分割的综述文章,本文的主要贡献如下:

a) 主要介绍2000年以后主流的图像分割方法,将之归纳为四种类型,分别是基于图论的算法,基于聚类的算法,基于分

**收稿日期:** 2016-07-30; **修回日期:** 2016-11-28    **基金项目:** 国家自然科学基金资助项目(61373012); 江苏省高校自然科学基金项目(15KJB520016)

**作者简介:** 周莉莉(1982-),女,江苏泰州人,讲师,主要研究方向为图像处理(18749813@qq.com);姜枫(1980-),男,副教授,博士研究生,主要研究方向为计算机视觉、机器学习。

类的算法,结合聚类和分类的算法;

b)对图像分割常用的公测数据集以及衡量算法性能指标进行了归纳和总结;

c)对近年来流行的卷积神经网络在图像分割的应用作了重点介绍。

### 1 基于图论的图像分割方法

基于图论的方法是一种自顶向下的全局分割方法,其主要思想是将整幅图像映射为一幅带权无向图  $G = (V, E)$ ,其中  $V$  是顶点的集合,  $E$  是边的集合,图像每个像素对应图中一个顶点,像素之间的相邻关系对应图的边,像素特征之间的相似性或差异性表示为边的权值。将图像分割问题转换成图的划分问题,通过对目标函数的最优化求解,完成图像分割过程。

#### 1.1 Normalized Cuts 算法

Wu 等人<sup>[7]</sup>根据图论中最小割 (min-cut) 的定义,将图像的最优分割问题转换为求解对应图的最小割问题。其形式化定义为: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  $A \cup B = V, A \cap B = \Phi$ 。定义最小割的目标函数为

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

根据式(1)定义,这种最小割方法仅考虑子图间耦合度最低,忽略了子图内部结点的耦合情况,倾向于分离单个节点的情况。因此,Shi 等人<sup>[8]</sup>在 2000 年提出将 normalized cuts (简称 NCuts)算法进行归一化,既考虑子图间的差异,同时考虑子图内部的相似性,定义全局目标函数:

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

其中:  $\text{assoc}(A, V) = \sum_{u \in A, t \in V} w(u, t)$ ,表示子集  $A$  中所有节点到图中所有节点的边权值之和。用这种方式定义两个区域的不相关性,使单个节点分割的结果不再满足  $\text{Ncut}(A, B)$  最小,避免分割单个孤点的情况。

NCuts 算法通过图像的轮廓特征和纹理特征来全局最小化目标函数,能生成规则的超像素。不足之处在于该算法的图像边界的贴合性一般,且该问题是 NP-hard 问题,对大图像问题的求解非常困难。

为此,Ren 等人<sup>[5]</sup>提出先用 NCuts 算法将图像分成数个较大的子图,接着对每个子图使用 K-means 算法进一步划分,降低了运算复杂度。2015 年, Li 等人<sup>[9]</sup>提出线性谱聚类 (linear spectral clustering, LSC) 算法。LSC 算法基于 K 路 NCuts (K-way Ncuts) 算法<sup>[8]</sup>的目标函数,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 NCuts 算法的目标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 NCuts 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 FH 算法

FH 算法<sup>[10]</sup>由 Felzenswalb 和 Huttenlocher 于 2004 年提出,是一种基于图像最小生成树的算法。该算法将图像映射成无向图  $G = (V, E)$ ,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,记为  $w(e)$ 。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  $E'$  是  $E$  的非空子集。

定义区域  $C$  的内部差异 (internal difference) 为:  $\text{Int}(C) =$

$\max_{e \in \text{MST}(C, E)} w(e)$ ,即区域  $C$  的所有生成树中的最大权值之和。定义区域  $C_1$  和  $C_2$  之间的差异 (difference) 为:  $\text{Dif}(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2, (v_i, v_j) \in E} w(v_i, v_j)$ ,也就是在所有能够连接区域  $C_1$  和  $C_2$  的边中权值最小的边的权值。

对于两个区域  $C_1$  和  $C_2$ ,如果  $\text{Dif}(C_1, C_2)$  大于  $\text{Int}(C_1)$  和  $\text{Int}(C_2)$  中较小的一个,则可以认为  $C_1$  和  $C_2$  之间有边界,定义为

$$D(C_1, C_2) = \begin{cases} \text{true} & \text{if } \text{Dif}(C_1, C_2) > \text{MInt}(C_1, C_2) \\ \text{false} & \text{otherwise} \end{cases}$$

$$\text{MInt}(C_1, C_2) = \min(\text{Int}(C_1) + \tau(C_1), \text{Int}(C_2) + \tau(C_2)) \tag{3}$$

其中:  $\tau$  是一个阈值函数,用于控制两个区域之间的差异必须大于其内部差异。

FH 方法通过对图中的节点进行聚类实现分割,其生成的超像素就是像素集合的最小生成树,能较好地保持图像边界。该算法运行速度很快,但无法控制图像块的数量和紧凑程度。

#### 1.3 Graph Cuts 算法

Graph Cuts 算法<sup>[11]</sup>由 Boykov 等人于 2006 年提出,该算法将图像映射成无向图  $G = (V, E)$ ,但该图比普通图多了两个顶点,分别为  $S$  和  $T$ ,图中的顶点和边分为两种类型:

第一种为普通顶点,对应于图像中的每个像素,每两个相邻像素的连接叫 n-links。

第二种为  $S$  (source:表示前景)和  $T$  (sink:表示背景),每个普通顶点和这两个终端顶点之间都有连接,称为 t-links。

图中每条边都有一个非负的权值  $w_e$ ,一个分割就是图中边集合  $E$  的一个子集  $C$ ,这个分割的代价 (表示为  $|C|$ ) 就是  $C$  中所有边的权值的总和。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,所以称为割。如果一个割,它的边的所有权值之和最小,就称为最小割。这个最小割把图的顶点划分为两个不相交的子集  $S$  和  $T$ ,其中  $s \in S, t \in T$  和  $S \cup T = V$ 。这两个子集分别对应于图像的前景像素集和背景像素集,这样就相当于完成了图像分割。

算法相关定义如下:  $P$  是所有像素,  $N$  是  $P$  中所有相邻的点对,  $A = (A_1, \dots, A_p, \dots, A_{|P|})$  是一个二进制向量,  $A_p$  指示像素  $p$  属于哪个区域,可以是前景 (obj) 或者 (bkg)。

$$E(A) = \lambda \cdot R(A) + B(A)$$

$$R(A) = \sum_{p \in P} R_p(A_p)$$

$$B(A) = \sum_{p, q \in N} B_{p, q} \cdot \delta_{A_p \neq A_q} \tag{4}$$

其中:  $R_p(A_p)$  表示给像素  $p$  分配标签  $A_p$  所产生的代价,该项的值可以通过比较像素  $p$  的灰度值和给定的目标和前景的灰度直方图来获得。  $B_{p, q}$  理解为  $p$  和  $q$  之间不连续的代价,  $p$  和  $q$  越相似,  $B_{p, q}$  值越大,反之趋于 0。  $E(A)$  是目标函数,它由区域项  $R(A)$  和边界项  $B(A)$  构成,求解最终目标是 minimized 该目标函数。

Graph Cuts 算法同时利用了图像的像素灰度信息和区域边界信息,且目标函数是构建在全局最优的框架下,保证了分割效果。但其也有一定缺点,如需要大量的矩阵广义特征向量运算,且其分割结果更倾向于具有相同的类内相似度。

#### 1.4 Superpixel Lattice 算法

Moorer 等人<sup>[12]</sup>于 2008 年提出了一种无监督的分割算法,称为 Superpixel Lattice。

Supixel Lattice 算法以二维图像边界图作为输入,该图保存了两个像素之间存在边界的概率,称为边界代价图(boundary cost map)。算法的原理是在图中寻找最小带权路径,使得边界代价图最小。该算法是一个迭代的过程,先将图在水平和垂直方向分别进行二分,形成四个区域,在预先设定的 strip 中搜索最优路径;接着在图的水平和垂直方向各增加一条路径,使图像被分成九个区域。如此反复进行,最终将图像分割成单独的图像块。该方法强制性地要求图像块形成网格结构,尽管增加了这一约束条件,但其在分割准确度和算法运行效率上保持了良好的性能。

为了克服 Supixel Lattice 算法过分依赖边界代价图的缺点,2009 年,Moore 等人<sup>[13]</sup>在该算法的基础上加以改进,在超像素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近乎均匀的超像素。

### 1.5 Seeds 算法

基于图论的图像分割方法通过构造目标函数并求解,为了取得更好的分割效果,目标函数通常较为复杂,造成算法时间复杂性高、不能满足实时应用的要求。针对这一情况,Bergh 等人<sup>[14]</sup>提出了 SEEDS(superpixels extracted via energy-driven sampling)算法。

SEEDS 预生成一个超像素分割,通过不断地修正边界求精获得最优分割效果,目标函数定义为

$$E(s) = H(s) + \gamma G(s) \quad (5)$$

该目标函数  $E(s)$  由两项组成,其中  $H(s)$  表示图像块颜色分布,通过图像块的颜色密度分布进行计算。文中假设不同图像块的颜色分布彼此独立,而同一个图像块的颜色分布应尽可能均匀。 $G(s)$  表示图像块边界形状的先验知识,该项对分割边界的局部不规则性进行惩罚,有助于生成紧凑、平滑的边界。 $\gamma$  是用于调节这两项的权值因子。SEEDS 使用爬山法(hill-climbing)最优化目标函数,以迭代的方式、通过寻找最小局部变化更新解。

## 2 基于聚类的图像分割方法

聚类方法是将对象的集合分成由类似的对象组成的多个类的过程。聚类的思想可以应用到图像分割中,将图像中具有相似性质的像素聚类到同一个区域或图像块,并不断迭代修正聚类结果,直至收敛,从而形成图像分割结果。

### 2.1 Meanshift 算法

Meanshift 算法最初由 Fukunaga 等人<sup>[15]</sup>提出。1995 年,Cheng 等人<sup>[16]</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。2002 年,Comaniciu 等人<sup>[17]</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,将具有相同模点的像素聚类在一起形成超像素区域。

Meanshift 算法的基本思想如下:在  $d$  维空间中,任选一点作为圆心,以  $h$  为半径作高维球。圆心和每个落在球内的点都会产生一个以圆心为起点、该点为终点的向量,将这些向量相加,其结果就是 Meanshift 向量。继续以 Meanshift 向量的终点为圆心作高维球,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 作图像分割,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

$$y_{k+1}^{\text{mean}} = \arg \min_z \sum_i \|x_i - z\|^2 \varphi\left(\left\|\frac{x_i - y_k}{h}\right\|^2\right) \quad (6)$$

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。

该算法稳定性、鲁棒性较好,有着广泛的应用。但是其速度较慢,分割时所包含的语义信息较少,所以分割效果不够理想,无法有效控制图像块数量。

### 2.2 Medoidshift 算法

Sheikh 等人<sup>[18]</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 算法相类似,Medoidshift 算法也能够自动计算聚类的数目,并且数据不必线性可分。

Medoidshift 算法比 Meanshift 算法的优势体现在三处。

a) Medoidshift 算法是一种增量聚类算法,前次迭代的计算结果可以在以后的迭代中重复利用。

b) Meanshift 算法需要定义均值(mean)的概念,而 Medoidshift 算法则不需要,只需要定义两个点之间距离(distance)的概念即可,因此可以直接利用距离矩阵运算。

c) Meanshift 算法需要定义算法终止条件,而 Medoidshift 算法则不需要。

在 Medoidshift 算法中,每次迭代并非计算新位置  $y_{k+1}$ ,而是计算新的中心(medoid)。一个中心点  $y \in \{x_i\}$  的定义如下:

$$y_{k+1}^{\text{medoid}} = \arg \min_{z \in \{x_i\}} \sum_i \|x_i - z\|^2 \varphi\left(\left\|\frac{x_i - y_k}{h}\right\|^2\right) \quad (7)$$

其参数的含义同式(6)相同,对比式(6)和式(7)可以发现,Meanshift 算法选择的点使目标函数达到最小值,而 Medoidshift 算法选择的点是从所有的  $\{x_i\}$  中能够使目标函数达到最小值的点。

Medoidshift 算法的缺陷在于其时间复杂度较高,可以证明,其时间复杂度为  $O(N^3)$ ,通过改进可以降低到  $O(N^{2.38})$ ,而 Meanshift 算法时间复杂度为  $O(dN^2 T)$ ,其中  $d$  是数据维度, $T$  为算法迭代次数,显然  $dT \gg N$ 。Vedaldi 等人<sup>[19]</sup>对 Medoidshift 算法中点对的距离限制为欧氏距离,并不断促使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的像素移动来实现图像的分割。该算法称为 Quickshift,其时间复杂度为  $O(dN^2)$ (其中  $d$  是一个小常数),低于 Meanshift 算法的时间复杂度。但 Quickshift 算法是非迭代的,无法有效地控制图像块的大小和数量。

### 2.3 Turbopixels 算法

2009 年,Levinshtein 等人<sup>[20]</sup>提出了一种几何流(geometric flows)的超像素快速生成算法,称为 TurboPixels,该算法将图像分割成近似网格结构的图像块,算法流程描述如算法 1 所示。

TurboPixels 算法生成的图像块满足以下五个条件:a)各图像块尺寸均匀;b)图像块内保持连通;c)图像块比较紧凑;d)图像块边界光滑;e)各图像块彼此不重叠。通过 TurboPixels 算法生成的图像块较好地保持了图像的局部边界,并限制了欠分割(under segmentation)的发生。该算法时间复杂度近似为  $O(N)$ ,尤其适用于百万像素级的大图像。

算法 1 TurboPixels 算法

- 1 放置初始化种子
- 2 repeat
- 3 第  $T$  次进化边界
- 4 估算未分配区域的骨架
- 5 更新边界像素的速度以及边界附近未分配像素的速度
- 6 until 没有进一步的进化

### 2.4 SLIC 算法

Achanta 等人<sup>[21]</sup>在 2012 年提出一种简单的超像素方法,称为 SLIC(simple linear iterative clustering)。该算法通过计算图像中像素的颜色相似度以及距离进行聚类生成超像素。

图像中的每个像素被分解成一个五维向量  $\{l, a, b, x, y\}$ , 其中  $l, a, b$  是 CIELab 颜色空间中的分量,  $x, y$  是像素的坐标。定义两个像素点之间的距离为

$$D_S = d_{lab} + \frac{m}{S}d_{xy}$$

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2}$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \quad (8)$$

其中:  $m$  是用来控制超像素紧凑度的参数,一般在 1 ~ 20;  $S = \sqrt{\frac{N}{K}}$ ,  $K$  为设置的超像素的块数,  $N$  为图像的总像素数, SLIC 算法描述见算法 2。

#### 算法 2 SLIC 算法

按照步长  $S$  对图像进行采样,初始化聚类中心

$$C_k = [l_k, a_k, b_k, x_k, y_k]^T$$

- 1 修正每个聚类中心到梯度最小的位置
- 2 repeat
- 3 for 每个聚类中心  $C_k$  do
- 4 将聚类中心周边  $2S \times 2S$  区域内像素分配到最近聚类
- 5 end for
- 6 重新计算聚类中心以及残差  $E$
- 7 until  $E \leq \epsilon$
- 8 强制连通

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,所以 SLIC 算法时间复杂度为  $O(N)$ , 并且 SLIC 能够生成紧凑、近似均匀的超像素。

## 3 基于分类的图像分割方法

### 3.1 QEM 方法

SVM<sup>[6]</sup>是一种性能良好的分类器,最近被用来解决图像分割问题<sup>[22-24]</sup>, 通过从图像中提取像素级特征,使用 SVM 进行逐像素分类,从而实现图像分割。

然而,图像的像素级特征是从图像亮度或颜色不同通道(如 RGB、HSV 等)提取的,未考虑各通道之间的关系,其次 SVM 的训练速度比较缓慢。因此,Wang 等人<sup>[25]</sup>提出四元指数矩(quaternion exponent moments, QEM)方法。使用四元指数矩,考虑包括图像各颜色通道之间的关系在内的像素级特征。将特征作为孪生支持向量机(twin support vector machine, TSVM)<sup>[26]</sup>的输入,TSVM 事先使用 Arimoto 熵阈值选择训练样本进行训练。最后利用训练好的 TSVM 模型对图像逐像素分类,根据分类结果实现彩色图像分割。

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。

### 3.2 FCN 方法

Long 等人<sup>[27]</sup>于 2015 年提出的全卷积网络(fully convolutional networks, FCN)方法,提出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,解决图像语义分割问题。

FCN 方法利用了 VGG 16 网络<sup>[28]</sup>, 该网络具有 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,并在 pool3 和 pool4 层后加上反卷积层,采用双线性上采样的方法将粗糙(coarse)输出转换为密集(dense)输出。

FCN 方法的主要优势如下:

a) 实现像素级的预测。传统的卷积网络需要做下采样(subsampling), 因此其输出图像大小会降低。在 FCN 中,将 AlexNet<sup>[29]</sup>、VGG<sup>[28]</sup> 和 GoogLeNet<sup>[30]</sup> 等经典卷积网络的全连接层全部转换为卷积层。这样做能够充分利用预训练的网络,只需调优(fine-tuning)即可,训练非常高效;同时在卷积层得到的特征图(feature map)上采用双线性插值上采样,使输出的分割图像和输入图像尺寸相同。

b) 综合利用图像全局信息和局部信息。图像的全局信息包含语义信息,局部信息包含位置信息,FCN 采取采用 skip layer 的方法,在浅层处减小上采样的步长,得到的精细层(fine layer)和高层得到的粗糙层(coarse layer)做融合,然后上采样得到输出,以此兼顾全局信息和局部信息,取得了良好的分割效果提升。

FCN 方法在实际使用中存在如下弊端:(a) 由于使用了固定尺寸的感知野(滤波器),所以只能检测和处理单一尺度的语义目标;(b) 物体的细节结构可能会丢失或边界模糊。

Krähenbüh 等人<sup>[31]</sup>采用条件随机场(conditional random fields, CRF)的方法进行边界优化。Noh 等人<sup>[32]</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 只能处理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物体细节,得到高质量分割效果。

### 3.3 Zoom-out 方法

QEM 方法和 FCN 方法的原理是利用图像局部特征信息指导像素分类,Mostajabi 等人<sup>[33]</sup>认为仅使用图像局部特征只能部分反映图像内容信息,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(feedforward semantic segmentation with zoom-out features)算法。

Zoom-out 算法的核心思想在于利用巧妙设计的 Zoom-out 结构从图像提取多个级别的特征用于像素分类,特征从低到高分为四个级别:

a) 局部(local)。最低级别特征称为局部特征,包含颜色、纹理、密度/梯度模式,以及其他可以在较小连续区域内计算的属性。相邻区域的局部特征有可能差异较大,这在物体的边界处尤其明显。

b) 近端(proximal)。近端所指的范围比局部特征空间区域稍大,在这个级别能够获取一些视觉线索,近端特征包含局

部范围无法获取的特征。对于特征分布不均匀的物体,近端特征比局部特征能更好地反映特征分布。

c)远端(distant)。远端特征指能够包含整个物体或者物体一部分大小级别的特征,这个级别有足够的信息可用于理解形状、颜色和梯度的复杂模型、空间布局等。远端特征能够跨越物体的真实边界,因此可以提取物体和相邻物体的高层特征。

d)场景(scene)。场景特征是最高级别特征,可以描述整幅图像的全局信息,能够用于图像级别的分类。通过对场景特征的分类可以辅助识别场景中的物体种类,提升高图像分割的准确率。

Zoom-out方法利用CNN不同层的输出提取上述多个级别的图像特征,组合成特征向量 $\varphi(s, I) = [\varphi_1(s, I), \dots, \varphi_L(s, I)]$ ,其中 $I$ 为输入图像, $s$ 是图像中某个区域, $L$ 是图像级别的数目。训练网络时,需要为每个区域 $s$ 做标注,记为 $y_s$ ,所选用的目标函数为

$$-\frac{1}{N} \sum_{i=1}^N \frac{1}{f_c} \log \hat{p}(y_i | \varphi(s_i, I_i)) \quad (9)$$

其中: $N$ 是训练样本总数; $f_c$ 表示类 $c$ 中包含的训练样本数; $\log \hat{p}(y_i | \varphi(s_i, I_i))$ 表示图像 $I_i$ 中区域 $s_i$ 能够正确标注的估算概率。

Zoom-out方法在Pascal VOC 2010、VOC 2011、VOC 2012等公测集上均取得了当年的最好成绩。

#### 4 结合聚类和分类的图像分割方法

分类方法是有监督的机器学习算法,需要大量的标注数据作为训练样本,事实上,这样的像素级(pixel-wise)的标注图像样本非常稀少,难以胜任分类器的训练任务。而聚类方法是一种无监督的学习算法,无须标注图像作为训练样本。因此结合无监督的聚类算法和有监督的分类算法各自优势,研究图像分割算法,也是近年的热点之一。这类方法的思路通常分为三个步骤:(a)使用聚类算法生成目标候选区域集;(b)使用分类算法对各区域分类;(c)根据区域分类结果构建全图标注,完成图像分割。

##### 4.1 O2P方法

Carreira等人<sup>[34]</sup>于2012年提出了二阶池化(second-order pooling, O2P)方法。该算法包含三个步骤,首先采用CPMC算法<sup>[35]</sup>从原始图像中提取出一系列候选区域集,然后对每一个候选区域进行特征描述并分类,最后根据分类结果完成语义分割。

O2P算法的主要贡献在于第二个步骤,即对候选区域进行特征描述和分类采取了不同的方法。常用的特征描述方法一般采用词袋模型(bag of words, BoW)和方向梯度直方图(histogram of oriented gradient, HOG),这样做的缺点是需要分类器采取非线性的核函数,并且需要在整幅图像中采用滑动窗口探测技术,这些都是非常费时的操作。

$$G_{\text{avg}}(R_j) = \frac{1}{|F_{R_j}|} \sum_{i: (i \in R_j)} x_i \cdot x_i^T \quad (10)$$

$$G_{\text{max}}(R_j) = \max_{i: (i \in R_j)} x_i \cdot x_i^T \quad (11)$$

在O2P中,采取了先特征提取,后二阶池化的方法,可以采取二阶均值池化(2AvgP)或二阶最大池化(2MaxP)的方法,其公式分别如式(10)和(11)所示。从式中可以看出,二阶均值池化会得到对称正定矩阵(symmetric positive definite, SPD)。SPD具有很好的几何性质,构成黎曼流形,利用对数操作将其投影到正切空间,则可以通过简单的线性分类器对特征进行分

类,从而避免了复杂的非线性运算。相比于一阶池化方法,该算法取得了更好的分割效果。

##### 4.2 SDS方法

同步检测及分割(simultaneous detection and segmentation, SDS)方法由Hariharan等人于2014年提出<sup>[36]</sup>,该方法可用于图像中对象检测和语义分割两个任务。其算法分为四个步骤。

a)候选区生成。使用MCG算法<sup>[37]</sup>从每幅图像中生成2000个左右的区域候选集,每个区域是一个物体的最小外接矩形。

b)特征提取。使用卷积神经网络(convolutional neural network, CNN)从每个区域提取特征,除了提取最小外接矩形的特征外,同时提取区域前景特征,将两部分特征联合训练CNN特征。

c)区域分类。利用CNN提取的特征,使用训练好的SVM对每个候选区进行分类。

d)区域增强。对候选区评分使用非最大约束 non-maximum suppression, NMS,接着使用CNN提取的特征生成掩码(mask),将掩码和原始区域候选集融合提升分割效果。

该算法的优势在于使用CNN提取区域特征,利用候选区最小外接矩形(规则图像)和前景(不规则图像)两部分特征联合训练CNN,这样做取得了比单独使用原始区域更精确的分割效果。但是,每幅图像2000个左右的候选区域会带来非常大的计算量,因此还不适用于实时场景。

##### 4.3 R-CNN方法

Girshick等人<sup>[38]</sup>利用近年来在深度学习方面取得的成果,于2014年提出了区域卷积神经网络(regions with CNN, RCNN)方法,该方法可用于图像目标检测和语义分割,算法分为三个步骤:

a)从原始图像中使用selective search方法<sup>[39]</sup>抽取约2000个区域建议,这些区域建议均是具体对象类别无关的。

b)将每个区域变换为固定大小(227×227)的RGB图像,作为输入,利用卷积神经网络(含5个卷积层和2个全连接层)计算每个区域的特征。

c)利用步骤b)提取的特征以及训练标注,为每一个对象类构造SVM分类器。由于训练样本非常庞大,所以使用了Hard negative mining方法<sup>[40]</sup>,该方法收敛速度很快,且能有效提高平均准确率。

R-CNN的主要贡献在于率先将CNN作用于区域建议以定位并分割物体,其次是提出了一种在标注样本稀少的情形下有效训练大型卷积神经网络的方法。但其也有一定的局限性,(a)该算法依赖于步骤a)中区域建议生成的质量和数量;(b)该算法要将每个区域变换为固定大小图像作为CNN输入,这种操作会产生图像形变从而影响最终的效果。

## 5 算法比较与分析

### 5.1 图像分割常用数据集

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包括:

1)PASCAL VOC<sup>[41]</sup>。PASCAL VOC(pattern analysis, statistical modeling and computational learning visual object classes)

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。最初该集合中只包含 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 9993 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>[42]</sup>。SBD (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>[43]</sup>。该数据集中包括 101 个类别的物体, 每种物体包含 40 ~ 800 张图像不等, 大部分物体有 50 张左右图像, 每张图像的尺寸大约为 300 × 240 像素。该数据集后来发展为 Caltech 256, 包含 256 个类别, 共 30 607 张图像。

4) BSDS<sup>[44]</sup>。BSDS (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查找算法的性能。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)。数据集由彼此没有交叉的训练集、验证集和测试集三部分组成。

5) MSRC<sup>[45]</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, 包含 23 类物体, 共 591 张图像, 其中 21 个类是常用的, 每张图像有像素级的类别标注。

6) SIFT Flow<sup>[46]</sup>。包含 33 个语义类别目标, 以及 3 个地理类别目标, 共 2 688 张像素级标注的图像, 其中大部分为户外场景, 如街道、海滩、山脉、建筑等。该数据集由 2 488 张训练图像和 200 张测试图像构成。

### 5.2 算法性能衡量指标

为了科学地评价图像分割算法性能的优劣, 往往需要使用统一的指标进行定量比较。根据前述图像分割方法的不同, 这些指标分为两类, 具体如下:

第一类适用于超像素方法, 包括基于图论的算法和基于聚类的算法, 这类指标包括:

#### a) 边界召回率

边界召回率 (boundary recall, BR) 是指边界真值出现在算法得到的分割边界中的比率, 其数学定义为<sup>[47]</sup>

$$BR = \frac{TP}{TP + FN} \tag{12}$$

其中: TP (true positives) 为在边界真值中像素同时出现在算法得到的边界中的像素数目; FN (false negatives) 为在边界真值中的像素但没有出现在算法得到的边界像素中的数目。

#### b) 欠分割错误率

欠分割错误率 (undersegmentation error, UE) 用来度量超像素溢出真值的程度, 其计算方法有多种, 其中一种定义为<sup>[47]</sup>

$$UE = \frac{\sum_i \sum_{k: S_k \cap G_i \neq \emptyset} |S_k - G_i|}{\sum_i |G_i|} \tag{13}$$

其中:  $G$  为真值分割;  $S$  为算法得到的分割;  $| \cdot |$  为超像素中包含的像素数目。

#### c) 可达分割准确率

可达分割准确率 (achievable segmentation accuracy, ASA) 是一种算法性能上界的度量, 它给出了使用该算法得到的超像素分割作为输入, 能得到的对象分割的最高准确度, 其定义

为<sup>[48]</sup>

$$ASA = \frac{\sum_i \max_k |S_k \cap G_i|}{\sum_i |G_i|} \tag{14}$$

除以上三个指标外, 紧密度 (compactness)、面积方差、圆度等也是常用于衡量超像素算法性能的指标。

第二类指标适用于语义分割方法性能的评价, 基于分类的算法和结合聚类与分类的算法都属于语义分割的范畴, 这类指标包括:

#### a) 像素准确率

像素准确率 (pixel accuracy, PA) 用于计算正确分割的像素数目与图像像素总数的比例, 其定义为<sup>[27]</sup>

$$\frac{\sum_{i=1}^N n_{ii}}{\sum_{i=1}^N t_i} \tag{15}$$

其中:  $N_{cl}$  为图像中对象类别总数;  $n_{ij}$  表示实际类别为  $i$ ; 预测类别为  $j$  的像素数目;  $t_i$  为属于类别  $i$  的像素数目。

#### b) 平均准确率

平均准确率 (mean accuracy, MA) 是指各种类别对象的准确率平均值, 其定义为<sup>[27]</sup>

$$\frac{1}{N_{cl}} \sum_{i=1}^N \frac{n_{ii}}{t_i} \tag{16}$$

#### c) 平均 IoU

平均 IoU (mean intersection over union, mean IoU) 用于衡量分割结果与真值的交并集比例, 其定义为<sup>[27]</sup>

$$\frac{1}{N_{cl}} \sum_{i=1}^N \frac{n_{ii}}{t_i + \sum_{j=1}^N n_{ji} - n_{ii}} \tag{17}$$

### 5.3 算法分析与比较

如上所述, 基于内容的图像分割算法主要包含四种类型, 其中基于图论和基于聚类的方法采取无监督学习的方式, 都属于超像素方法, 以将具有相同或相近属性的像素归类到同一区域为分割目标; 基于分类的方法属于有监督学习方法, 结合聚类和分类的方法属于弱监督学习方法, 这两种方法都能够为图像进行逐像素的类别标注。在分析和比较算法时, 将前两种方法作为一类, 后两种方法作为另一类分别进行。

基于聚类和基于图论的图像分割研究工作分析与比较结果如表 1 所示, 主要比较因素包括: 发表年份、算法计算复杂度、图像块数量可控性、图像块紧凑度可控性等, 其中  $N$  表示图像中像素数量,  $N/A$  表示相关论文中未提及该项数据。

基于分类和结合聚类与分类的图像分割研究工作分析与比较结果如表 2 所示, 主要比较因素包括: 发表年份、是否需要标注图像、使用分类器类别、算法应用的数据集、算法特点等。

表 1 基于聚类和基于图论的图像分割算法

分类	算法名称	作者	发表年份	计算复杂度	图像块数量可控	图像块紧凑度可控
基于图论的方法	Neuts	Shi 等人 <sup>[8]</sup>	2000	NP-hard	是	是
	FH	Felzenswalb 等人 <sup>[10]</sup>	2004	$O(N \log N)$	否	否
	Graph Cuts	Boykov 等人 <sup>[11]</sup>	2006	NP-hard	N/A	N/A
	Superpixel Lattice	Moorer 等人 <sup>[13]</sup>	2008	$O(N^{3/2} \log N)$	是	是
	SEEDS	Bergh 等人 <sup>[14]</sup>	2012	$O(N)$	否	否
基于聚类的方法	LSC	Li 等人 <sup>[9]</sup>	2015	$O(N)$	是	是
	Meanshift	Comaniciu 等人 <sup>[17]</sup>	2002	$O(N^2)$	否	否
	Medoidshift	Sheikh 等人 <sup>[18]</sup>	2007	$O(N^3)$	否	否
	Quickshift	Vedaldi 等人 <sup>[19]</sup>	2008	$O(N^2)$	否	否
	Turbopixels	Levinshtein 等人 <sup>[20]</sup>	2009	$O(N)$	是	是
	SLIC	Achanta 等人 <sup>[21]</sup>	2012	$O(N)$	是	是

表2 基于分类和结合聚类与分类的图像分割算法

分类	算法名称	作者	发表年份	需要标注图像	分类器类别	应用数据集	特点
基于分类的方法	QEM	Wang 等人 <sup>[25]</sup>	2016	像素级	TSVM	BSDS <sup>[44]</sup> MSRC <sup>[45]</sup>	利用了图像各颜色通道之间的关联,具有较好的鲁棒性,算法运行速度较快。
	FCN	Long 等人 <sup>[27]</sup>	2015	像素级	CNN	PASCAL VOC <sup>[41]</sup> SIFT Flow <sup>[46]</sup>	融合图像局部和全局信息,实现端到端的像素级图像分割。
	Zoom-out	Mostajabi 等人 <sup>[33]</sup>	2015	实例级	CNN + LR	PASCAL VOC <sup>[41]</sup> SBD <sup>[42]</sup>	利用 CNN 提取图像多级特征,使用融合特征进行像素分类,实现图像分割。
结合聚类和分类的方法	O2P	Carreira 等人 <sup>[34]</sup>	2012	图片级	SVM	PASCAL VOC <sup>[41]</sup> Caltech101 <sup>[43]</sup>	使用二阶池化方法描述特征,使用线性 SVM 分类,分割效果好,速度较快。
	SDS	Hariharan 等人 <sup>[36]</sup>	2014	图片级	CNN + SVM	SBD <sup>[42]</sup> PASCAL VOC <sup>[41]</sup>	融合区域前景和最小外接矩形提取特征,使用 CNN 提取图像特征,利用 NMS 提升分割效果。
	R-CNN	Arbelaez 等人 <sup>[37]</sup>	2014	图片级	CNN + SVM	PASCAL VOC <sup>[41]</sup>	使用区域 CNN 提取图像特征,定位并分割物体;提出在标注样本稀少的情况下使用深度网络的方法。

## 6 未来发展方向和趋势

图像分割问题一直是计算机视觉、图像处理领域的研究热点,近年来针对图像分割的研究取得了大量的成果,但在该领域仍存在不少问题需要解决,值得研究人员进一步关注。

1) 性能良好,算法复杂度低的超像素算法。超像素作为一种重要的图像分割方法,目前存在的主要问题是边界贴合度高和算法时间复杂度低是一对矛盾。提高边界贴合度的常用做法是在超像素方法中引入图像全局信息辅助局部信息,但这通常意味着构造更加复杂的目标函数,增加算法运行时间。因此,如何平衡两者之间的关系,一直是学者们致力解决的问题。

2) 基于弱标注信息的语义分割方法。本文所介绍的主流图像公测数据集大部分标注样本均是图像级(image-level)或实例级(instance-level)的弱标注,而现在主要的语义分割方法均要借助像素级(pixel-wise)的强标注样本。在这方面,文献[49,50]进行了一些有益的尝试,但分割效果有进一步提升的空间。

3) 针对特定应用场景的图像分割问题。本文所介绍的算法主要针对通用场景图片,而在不同应用领域对图像分割有不同的标准和要求。文献[51,52]介绍了图像分割在医学图像领域的应用,文献[53,54]介绍了地质领域的图像分割,文献[55]介绍了遥感图像的分割。如何针对各种不同的应用的特点和需求,开发具有针对性的图像分割算法也颇具挑战。

4) 交互式图像分割方法。全自动的图像分割方法是一项非常困难的任务,在分割过程中利用少量的用户交互信息,可以有效提升图像分割的效果。在这方面,图割法(graph cut)、随机游走(random walks)等方法及其扩展<sup>[56,57]</sup>是研究的趋势之一。

5) 针对多维图像的图像分割方法。本文所述的研究大部分面向传统二维图像,近年来,针对多张图像提取相同前景目标的协同分割(CoSegmentation),针对RGB-D图像的分割算法,以及针对视频的分割方法也逐渐增多,成为发展的方向之一。

## 7 结束语

本文对2000年以来的针对图像分割问题的相关文献进行了较为细致的梳理,并在此基础上根据算法原理对其进行分类,选取每一类方法中的代表性算法进行研究和分析,指出其优缺点,并就同类算法进行对比。此外,本文对图像分割的常用公测数据集以及算法评价指标进行了介绍。最后就图像分

割问题未来发展的方向和趋势进行了论述。

(见电子版)

### 参考文献:

- [1] Otsu N. A threshold selection method from gray-level histograms[J]. *Automatica*, 1975, 11(285-296): 23-27.
- [2] Davis L S. A survey of edge detection techniques[J]. *Computer Graphics and Image Processing*, 1975, 4(3): 248-270.
- [3] Adams R, Bischof L. Seeded region growing[J]. *Pattern IEEE Trans on Analysis and Machine Intelligence*, 1994, 16(6): 641-647.
- [4] Meyer F. Skeletons and watershed lines in digital spaces[C]// *International Society for Optics and Photonics*. 1990: 85-102.
- [5] Ren Xiaofeng, Malik J. Learning a classification model for segmentation[C]// *Proc of the 9th IEEE International Conference on Computer Vision*. [S. l.]: IEEE Press, 2003: 10-17.
- [6] Vapnik V N, Vapnik V. *Statistical learning theory*[M]. New York: Wiley, 1998.
- [7] Wu Zhenyu, Leahy R. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1993, 15(11): 1101-1113.
- [8] Shi Jianbo, Malik J. Normalized cuts and image segmentation[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [9] Li Zhengqin, Chen Jiansheng. Superpixel segmentation using linear spectral clustering[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. [S. l.]: IEEE Press, 2015: 1356-1363.
- [10] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation[J]. *International Journal of Computer Vision*, 2004, 59(2): 167-181.
- [11] Boykov Y, Funka-Lea G. Graph cuts and efficient ND image segmentation[J]. *International Journal of Computer Vision*, 2006, 70(2): 109-131.
- [12] Moore A P, Prince J D, Warrell J, et al. Superpixel lattices[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. [S. l.]: IEEE Press, 2008: 1-8.
- [13] Moore A P, Prince S J D, Warrell J, et al. Scene shape priors for superpixel segmentation[C]// *European Conference on Computer Vision*. Berlin Heidelberg: Springer, 2009: 771-778.
- [14] Van den Bergh M, Boix X, Roig G, et al. Seeds: superpixels extracted via energy-driven sampling[M]// *Computer Vision-ECCV*. Berlin Heidelberg: Springer, 2012: 13-26.
- [15] Fukunaga K, Hostetler L D. The estimation of the gradient of a density function, with applications in pattern recognition[J]. *IEEE Trans on Information Theory*, 1975, 21(1): 32-40.
- [16] Cheng Yizong. Mean shift, mode seeking, and clustering[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1995, 17(8): 790-799.

- [17] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619.
- [18] Sheikh Y A, Khan E A, Kanade T. Mode-seeking by medoidshifts [C]//Proc of the 11th IEEE International Conference on Computer Vision. [S. l.]:IEEE, 2007: 1-8.
- [19] Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking[C]//European Conference on Computer Vision. Berlin Heidelberg:Springer, 2008: 705-718.
- [20] Levinshtein A, Stere A, Kutulakos K N, *et al.* Turbopixels: Fast superpixels using geometric flows[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2009, 31(12): 2290-2297.
- [21] Achanta R, Shaji A, Smith K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [22] Bai Xuefei, Wang Wenjian. Saliency-SVM: an automatic approach for image segmentation[J]. *Neurocomputing*, 2014, 136: 243-255.
- [23] Yu Zhiwen, Wong H S, Wen Guihua. A modified support vector machine and its application to image segmentation[J]. *Image and Vision Computing*, 2011, 29(1): 29-40.
- [24] Senyukova O V. Segmentation of blurred objects by classification of isolabel contours[J]. *Pattern Recognition*, 2014, 47(12): 3881-3889.
- [25] Wang Xiangyang, Wu Zhifang, Chen Liang, *et al.* Pixel classification based color image segmentation using quaternion exponent moments[J]. *Neural Networks*, 2016, 74: 1-13.
- [26] Kr? hnb?uhl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]//Advances in Neural Information Processing Systems. 2011:1-9.
- [27] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE Press, 2015: 3431-3439.
- [28] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv*: 1409.1556, 2014.
- [29] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [30] Szegedy C, Liu Wei, Jia Yangqing, *et al.* Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE Press, 2015: 1-9.
- [31] Khemchandani R, Chandra S. Twin support vector machines for pattern classification[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2007, 29(5): 905-910.
- [32] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation [C]//Proc of the IEEE International Conference on Computer Vision. 2015: 1520-1528.
- [33] Mostajabi M, Yadollahpour P, Shakhnarovich G. Feedforward semantic segmentation with zoom-out features [C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2015: 3376-3384.
- [34] Carreira J, Sminchisescu C. CPMC: automatic object segmentation using constrained parametric min-cuts[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1312-1328.
- [35] Carreira J, Caseiro R, Batista J, *et al.* Semantic segmentation with second-order pooling [C]//European Conference on Computer Vision. Berlin Heidelberg:Springer, 2012: 430-443.
- [36] Hariharan B, Arbel?ez P, Girshick R, *et al.* Simultaneous detection and segmentation[M]//Computer Vision-ECCV. [S. l.]:Springer International Publishing, 2014: 297-312.
- [37] Arbelaez P, Pont-Tuset J, Barron J, *et al.* Multiscale combinatorial grouping[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE Press, 2014:328-335.
- [38] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2014: 580-587.
- [39] Uijlings J R, Van de Sande K E A, Gevers T, *et al.* Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [40] Felzenszwalb P F, Girshick R B, McAllester D, *et al.* Object detection with discriminatively trained part-based models[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [41] Everingham M, Van Gool L, Williams C K I, *et al.* The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [42] Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions[C]//The IEEE 12th International Conference on Computer Vision. [S. l.]:IEEE, 2009: 1-8.
- [43] Li Feifei, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[J]. *Computer Vision and Image Understanding*, 2007, 106(1): 59-70.
- [44] Martin D, Fowlkes C, Tal D, *et al.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//Proc of the 8th IEEE International Conference on Computer Vision. [S. l.]:IEEE, 2001.
- [45] Wang Xiaoru, Du Junping, Wu Shuzhe, *et al.* Cluster ensemble-based image segmentation[J]. *International Journal of Advanced Robotic Systems*, 2013, 10: 297-308.
- [46] Liu Ce, Yuen J, Torralba A. Nonparametric scene parsing via label transfer[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2368-2382.
- [47] Neubert P, Protzel P. Superpixel benchmark and comparison [C]//Proc Forum Bildverarbeitung. 2012: 1-12.
- [48] Liu Mingyu, Tuzel O, Ramalingam S, *et al.* Entropy rate superpixel segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE Press, 2011: 2097-2104.
- [49] Pinheiro P O, Collobert R. Weakly supervised semantic segmentation with convolutional networks[J]. *arXiv*:1411.6228, 2014.
- [50] Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation[C]//Advances in Neural Information Processing Systems. 2015: 1495-1503.
- [51] Yang Xin, Cheng K T T, Chien A. Accurate vessel segmentation with progressive contrast enhancement and Canny refinement[C]//Asian Conference on Computer Vision. [S. l.]:Springer International Publishing, 2014: 1-16.
- [52] Cire?an D, Giusti A, Gambardella L M, *et al.* Deep neural networks segment neuronal membranes in electron microscopy images [C]//Advances in Neural Information Processing Systems. 2012: 2843-2851.
- [53] Jungmann M, Pape H, Wi?bkirchen P, *et al.* Segmentation of thin section images for grain size analysis using region competition and edge-weighted region merging [J]. *Computers & Geosciences*, 2014, 72:33-48.
- [54] Filho I M. Segmentation of sandstone thin section images with separation of touching grains using optimum path forest operators[J]. *Computers & Geosciences*, 2013, 57(4):146-157.
- [55] Li Zhao, Hao Fang. Multi-scale and multi-feature segmentation of high resolution remote sensing image [J]. *Journal of Multimedia*, 2014, 9(7):948-954.
- [56] Diebold J, Demmel N, Haz? rba? C, *et al.* Interactive multi-label segmentation of RGB-D images [C]//International Conference on Scale Space and Variational Methods in Computer Vision. Springer International Publishing, 2015: 294-306.
- [57] Liang Xianpeng, Zhang Xiaoping, Shang Li, *et al.* Locally biased discriminative clustering method for interactive image segmentation [C]//International Conference on Intelligent Computing. [S. l.]:Springer International Publishing, 2016: 514-522.