

基于深度 CRF 模型的图像语义分割方法

胡 涛, 李卫华, 秦先祥[✉], 邱浪波, 李小春

(空军工程大学信息与导航学院, 西安, 710077)

摘要 从图像中提取多种特征向量堆叠为一个高维特征向量用于图像语义分割,会导致部分特征向量的分类能力减弱或丢失。针对此问题,提出了一种结合深度卷积神经网络 AlexNet 和条件随机场的图像语义分割方法。利用预训练好的 AlexNet 模型提取图像特征,再通过条件随机场对多特征及上下文信息的有效利用来实现图像的语义分割。与利用传统经典特征的方法进行对比,实验结果表明:在利用 AlexNet 模型提取特征进行图像语义分割时,Conv5 层为最有效的特征提取层,在 Stanford background 和 Weizmann horse 数据集下的识别准确率分别为 81.0% 和 91.7%,均高于其他 2 种对比方法,说明 AlexNet 可以提取更有效的特征,得到更高的语义分割精度。

关键词 语义分割;条件随机场;卷积神经网络;AlexNet 模型

DOI 10.3969/j.issn.1009-3516.2018.05.010

中图分类号 TP391 **文献标志码** A **文章编号** 1009-3516(2018)05-0052-06

An Image Semantic Segmentation Based on Deep CRF Model

HU Tao, LI Weihua, QIN Xianxiang[✉], QIU Langbo, LI Xiaochun

(Information and Navigation College, Air Force Engineering University, Xi'an 710077, China)

Abstract: Aimed at the problems that varieties of feature vectors extracted from the image is stacked into a high-dimensional feature vector for image semantic segmentation, and these lead to the weakening or loss of the classification ability of some feature vectors, an image semantic segmentation method based on deep convolution neural network AlexNet and conditional random fields is proposed. The pre-trained AlexNet model is utilized for extracting image features, and then the semantic segmentation of the image is achieved through the efficient use of conditional random fields for multiple features and context information. The experimental results compared with the methods using the traditional classical features show that Conv5 is the most effective feature extraction layer when AlexNet model is used to extract features for image semantic segmentation. The recognition accuracy in the Stanford background and Weizmann horse datasets is respectively 81.0% and 91.7%, and both the accuracy rates are higher than that of the two comparison methods, indicating that the deep convolution neural network can extract more effective features and obtain higher semantic segmentation accuracy.

Key words: semantic segmentation; conditional random field; convolution neural network; AlexNet model

收稿日期: 2017-12-27

基金项目: 国家自然科学基金(41601436;61403414;61703423);陕西省自然科学基金基础研究计划(2016JQ6070)

作者简介: 胡 涛(1994—),男,湖南浏阳人,硕士生,主要从事图像处理研究。E-mail:hu1862965@163.com

通信作者: 秦先祥(1986—),男,广西阳朔人,讲师,主要从事图像处理研究。E-mail:qinxianxiang@126.com

引用格式: 胡涛,李卫华,秦先祥,等.基于深度 CRF 模型的图像语义分割方法[J].空军工程大学学报(自然科学版),2018,19(5):52-57.
HU Tao, LI Weihua, QIN Xianxiang, et al. An Image Semantic Segmentation Based on Deep CRF Model[J]. Journal of Air Force Engineering University (Natural Science Edition), 2018, 19(5): 52-57.

图像语义分割(Semantic Segmentation)的目的 在于将图像划分成一系列具有特定语义信息的图像 区域^[1-2]。

目前国内外众多学者开展了图像语义分割研究,并提出了大量图像语义分割方法,其中典型的有 概率图模型方法(如条件随机场(Conditional Random Field, CRF)^[3]、马尔科夫随机场(Markov Random Field, MRF)^[4]和贝叶斯网络(Bayesian Network, BN)^[5])以及支持向量机(Support Vector Machine, SVM)^[6]等。与其他方法相比,基于 CRF 模型的方法由于具有很强的多特征融合及图像上下文 信息利用能力,已成为当前应用最为广泛的图像语义 分割方法之一^[3, 7-8]。

基于 CRF 模型的图像语义分割方法主要由两部分 组成:特征提取和像素语义分类。图像特征提取是 其中的重要前提和关键步骤。用于图像语义分割的 经典特征包括颜色直方图、方向梯度直方图(Histogram of Oriented Gradient, HOG)和纹理特征等^[6-8]。 然而,这些特征通常是针对具体问题设计,对先验 知识的依赖程度较高,在很多情况下其表征能力往 往不尽人意。解决该问题的一种常用思路是通过提 取大量不同类型的特征来构建图像的高维特征^[9], 但是这种方法容易导致部分特征向量的分类能力减 弱或丢失。因此,如何提取更具表达性的特征是当 前提高图像语义分割方法性能的关键所在。

近年来,深度学习技术在处理多种计算机视觉 任务上取得了巨大进展,多种深度神经网络模型相 继被提出^[10-12],其中卷积神经网络(Convolutional Neural Network, CNN)^[13]在图像处理中应用最为 广泛。CNN 无需设计显式的特征模型,而是通过对 大量图像数据进行学习来获取图像深层次信息,进 而用于完成具体任务。针对图像语义分割问题,一 些学者设计了直接完成图像语义分割的 CNN 模 型,并展现出优异的性能^[14-15]。然而,这类方法需 要巨大的图像样本数据量作支撑,而且需要耗费大 量时间进行网络训练。当实际中训练样本较少或对 实时性要求较高时,这类方法往往难以满足需求。 实际上,CNN 前端的不同网络层可以视为图像不同 层次特征的隐式提取过程。考虑到不同类型图像之 间往往存在共性,因此,一个经过大型数据量训练好 的 CNN 的前端网络可以作为一个特征提取的有效 模型^[16]。

基于此并考虑到 CRF 的多特征和上下文信息 利用优势,本文提出一种结合预训练 CNN 和 CRF 的图像语义分割方法。首先利用经典的 CNN 模 型——AlexNet 模型来提取图像深层次特征,再通 过 CRF 对多特征及上下文信息有效利用来完成图

像的语义分割。

1 CRF 模型

CRF 是 Lafferty 等^[10]提出的一种概率图模型, 可以表示为一个无向图模型 $G = \{V, E\}$,其中 V 表 示图中节点集合, E 表示节点间无向边的集合。

假设图像观测向量为 \mathbf{x} ,类别标签 y ,根据 Hammersley-Clifford^[17]定理,标记场的后验概率 $P(y|\mathbf{x})$ 服从 Gibbs 分布:

$$P(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\theta})} \prod_{c \in C} \phi_c(y_c|\mathbf{x}, \boldsymbol{\theta}) \quad (1)$$

式中:归一化函数 $Z(\mathbf{x}, \boldsymbol{\theta}) = \sum_y P(y|\mathbf{x}, \boldsymbol{\theta})$; ϕ_c 为 定义在集团 c 上的势函数。

势函数阶数的确定与实验需求紧密相关,阶数 越高,可以表征越大范围节点间的相关性,但模型复 杂度也会随之提升。目前常用的做法是仅定义单位 置和双位置势函数^[18],既可兼顾性能,模型复杂度 也不会过高。因此,式(1)可改写为:

$$P(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\theta})} \exp\left\{ \sum_{i \in \Omega} \phi_i(x_i, y, \boldsymbol{\theta}_A) + \sum_{i \in \Omega} \sum_{j \in N_i} \phi_{ij}(x_i, x_j, y, \boldsymbol{\theta}_I) \right\} \quad (2)$$

式中: Ω 为所有节点的集合; N_i 为节点 i 的邻域; $\boldsymbol{\theta} = \{\boldsymbol{\theta}_A, \boldsymbol{\theta}_I\}$ 为模型参数; ϕ_i 和 ϕ_{ij} 分别表示单位置和 双位置势函数。

目前,单位置与双位置势函数的形式很多,需根 据具体情况来选择或确定。当势函数形式给定后, 势函数定义的关键在于确定特征的形式。

2 AlexNet

近年来,随着数据集的快速发展和计算机性能 的大幅提升,出现了很多高性能的卷积神经网络深 层架构,如 AlexNet、ResNet 和 VGG-Net 等,其直 接将图像作为网络输入,每个卷积层都可以得到输 入图像不同的特征表达,避免了传统语义分割算法 中复杂的特征提取过程。

AlexNet 是一种经典的 CNN 模型^[13]。与实现 图像语义分割的 CNN 模型相比,其结构更为简单。 AlexNet 包含 5 个卷积层和 3 个全连接层。网络的 输入是待识别图像,输出是图像识别结果。

3 深度 CRF 模型

针对传统图像语义分割方法受限于人工特征表 征能力不强的问题,本文提出了一种基于深度 CRF

模型的图像语义分割方法,采用 AlexNet 提取图像深度特征,将提取到的特征用于训练 CRF 模型,实现图像语义分割。

3.1 深度特征提取

AlexNet 模型要求输入尺寸为 224×224 的图像,目前常用的图像语义分割数据集如 Stanford background 数据集的尺寸都是不固定的,因此需要对输入图像采用双线性插值^[19]上采样到 224×224 ,然后再输入到 AlexNet 中提取深度特征,其中,AlexNet 是在 ImageNet 数据集上已经完成预训练的网络。在 AlexNet 前 5 层提取的特征都是由多张特征图组成,故前 5 层都可作为特征提取层。AlexNet 后 3 层为全连接层,提取到的特征都是一维列向量,不适合作为训练本文 CRF 模型的特征。采用的 CRF 模型需要将提取到的特征与原图像每个像素点逐一对应,因此需将 AlexNet 模型中各层提取到的特征采用双线性插值上采样到原图像大小。

3.2 CRF 模型建立

CRF 模型中的单位置势函数 ϕ_i 用于对已知位置 i 属于某一类的概率进行决策,只与位置 i 的特征有关,邻域中相邻点的特征对语义类别标号不产生影响。本文采用广义线性模型定义单位置势函数^[20]:

$$\phi_i(x_i, y, \theta_A) = \exp(y_i \theta_A^T g_i(x)) \quad (3)$$

式中: θ_A 为特征向量 $g_i(x)$ 包含的特征所对应的权值,其值在参数估计中确定; $g_i(x)$ 表示第 i 个点的特征。

双位置势函数 ϕ_{ij} 不仅要考虑邻域中相邻位置 i 和 j 的特征对这 2 点的语义标签 y_i 和 y_j 的作用,还应考虑对应语义标签之间的相互关系。本文双位置势函数同样采用广义线性模型:

$$\phi_{ij}(y_i, y_j, \theta_i) = \exp(y_i y_j \theta_i^T u_{ij}(x)) \quad (4)$$

$$u_{ij}(x) = |g_i(x) - g_j(x)| \quad (5)$$

式中: $u_{ij}(x)$ 为联合特征向量,表示特征向量间的差异对标号的影响; θ_i 为向量 $u_{ij}(x)$ 的权重,其值在参数估计中确定; y_i 和 y_j 分别为相邻像素点的标号,当 y_i 和 y_j 相同时, $y_i y_j$ 的值为 1,当 y_i 和 y_j 不相同, $y_i y_j$ 的值为 -1。从式(4)~(5)可以看出,相邻像素点的特征越相似,则越可能被赋予相同的标号。

确定 CRF 模型后需在训练阶段估计模型参数 θ ,CRF 模型参数估计方法有很多,如梯度上升和树重置信传播算法(Tree-Reweighted Belief Propagation, TRW)等^[2]。其中文献[2]中 TRW 算法采用截断拟合策略,在模型训练阶段预设置确定的迭代次数,提高了模型效率,因此本文采用该方法估计模型参数。参数确定后,在固定的参数下搜索使得式(4)中的后验概率最大的最优解 y^* ,即 $y^* = \arg \max_y \log P(y|x, \theta)$,从而确定每个像素的类别

标签,实现图像语义分割。

4 实验设计与结果分析

为了验证本文方法的有效性,在 Stanford background 数据集和 Weizmann horse 数据集下比较了 CNN 特征与传统特征对分类性能的影响,分析了 AlexNet 模型中不同层特征与分类性能的关系,最后进行了效率分析。

4.1 实验数据集

目前用于图像语义分割的数据集有很多,其中,Stanford background 数据集^[21]和 Weizmann horse 数据集^[22]是比较常用的 2 种数据集,为便于与其他同类方法进行比较,本文方法采用这 2 种数据集进行实验。

4.2 不同特征对分类性能的影响分析

为验证本文采用 CNN 特征的方法相对采用传统特征方法对分类性能的提升效果,与 2 组方法进行对比实验:

方法 1:提取 HOG 特征,利用 CRF 分类。

方法 2:提取 HOG 和纹理特征,利用 CRF 分类。

实验分别在 Stanford background 数据集和 Weizmann horse 数据集下进行测试,随机选取数据集中 80% 的图像作为训练图像,其余作为测试图像。文献[2]是目前性能最好的基于 CRF 的图像语义分割方法之一,因此本文方法与 2 组对比方法都采用文献[2]中的 CRF 模型,并分别提取不同的特征进行对比。为了提高效率,本文进行测试的方法都将数据集当中的图像分辨率降为原来的 20%。通过实验确定迭代次数,作为对比的 2 种方法迭代次数设置为 100 次,本文方法迭代次数设置为 1 000 次,都达到最好性能。本文方法选择提取 AlexNet 模型 Conv5 层特征进行实验。有关不同卷积层特征和迭代次数对算法性能的影响将在 4.3 节和 4.4 节进行分析。本文的特征提取在 MatConvNet^[23]深度学习平台上完成。

图 1 是在 Stanford background 数据集下典型的语义分割效果对比图对比图,其中,第 1 行为测试图像,第 2 行为真值图,第 3 行为方法 1 的结果,第 4 行为方法 2 的结果,第 5 行为本文方法结果。

对比真值图可以看到,方法 1 和方法 2 存在较多区域错分的情况,比如第 1 列图像中的建筑区域有很大一部分被错分为其它类别,第 2 列的部分树木区域也被错分为其它类别。方法 2 由于增加了纹理信息,分类结果相对方法 1 有一定改善,但不明显,比如,第 3 列中的建筑区域在方法 1 中有 3 处被

错误识别为其他类别,而方法 2 中只有一处。本文提出的方法由于采用了 CNN 特征,与其他 2 组方法相比,可以减少错误识别,找到更加完整的物体区域,比如,第 1 列中原来被误认为前景对象的区域已正确识别为建筑,与此类似,第 2 列和第 3 列中,被错误识别为前景的区域均已正确识别。第 4 列图像左下角原来未被识别出来的车辆区域被正确识别。综上所述,基于 CNN 特征的方法可以有效纠正基于传统特征方法的错误识别。

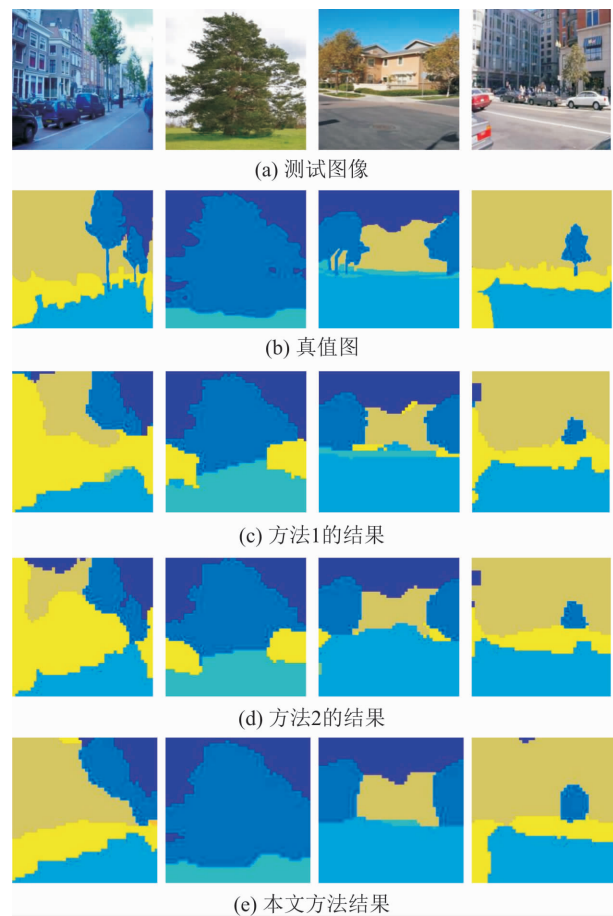


图 1 Stanford background 数据集下语义分割效果对比图
Fig. 1 Stanford background dataset semantic segmentation effect comparison chart

表 1 3 种方法在 Stanford background 数据集下的识别准确率

Tab. 1 The recognition accuracy of the three methods under the Stanford background dataset %

方法	天空	树木	山	道路	草地	水	建筑	前景	平均识别准确率
方法 1 ^[5]	90.1	69.6	87.0	78.0	87.8	67.4	75.4	50.8	71.6
方法 2 ^[5]	90.0	70.8	87.1	78.4	90.1	70.4	60.7	56.4	73.8
本文方法	90.5	72.4	88.0	78.5	90.5	73.9	60.8	62.9	81.0

表 2 3 种方法在 Weizmann horse 数据下的识别准确率

Tab. 2 The recognition accuracy of the three methods under the Weizmann horse dataset %

方法	背景	马	平均识别准确率
方法 1 ^[5]	88.5	58.8	86.4
方法 2 ^[5]	90.6	59.2	88.9
本文方法	93.0	69.2	91.7

为作量化对比,采用文献[2]提出的识别准确率 (Pixel Accuracy, PA) 作为评价标准,本文算法与 2 组对比方法在 Stanford background 和 Weizmann horse 数据集下的识别准确率比较见表 1 和表 2。与方法 1 和方法 2 相比,本文方法对表 1 中的 8 类物体和表 2 中的 2 类物体都取得了更高的准确率,性能更为优越。在 Weizmann horse 数据集下对背景和马的识别准确率都有所提高,与方法 1 相比,分别提高了 4.5% 和 10.4%。与方法 2 相比,分别提高了 2.4% 和 10%。在 Stanford background 数据集下对前景对象和水的识别准确率提高最为显著,较方法 1 分别提高了 12.1% 和 6.5%,较方法 2 分别提高了 6.5% 和 3.5%。

4.3 AlexNet 模型特征层选择

为了比较 AlexNet 模型中的哪一层特征更具表达力,以便选择合适的特征提取层,提取前 5 层特征进行实验并做精度评价,在 Stanford background 数据集下进行测试的结果见图 2。本文方法提取前 5 层特征进行测试,特征提取在 MatConvNet^[24] 深度学习平台上完成。CRF 模型训练阶段的迭代次数为最优迭代次数 1 000 次。AlexNet 前 5 层中,每层提取的特征都是由多张特征图组成,如 Conv1 层的特征为 96 张尺寸为 55×55 的特征图,插值到原图像大小后,得到 96 张尺寸为 224×224 的特征图,相当于对原图像的每一个像素点提取一个 96 维的特征向量。

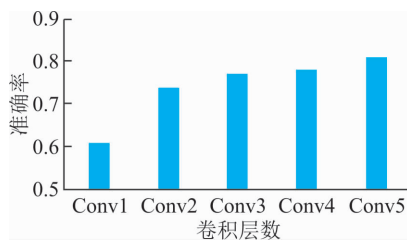


图 2 不同层特征分类结果精度对比图

Fig. 2 Accuracy comparison results of different layer classification results

从图 2 的分类结果精度对比图中可以看出:在 Stanford background 数据集下,随着卷积层数深度增加,分类精度呈上升趋势,在 Conv5 层达到最高。这是因为 AlexNet 模型中更深层特征更抽象,具有更高层次的语义信息。此外,Conv1 层特征对应的分类精度远低于其他几层特征对应的分类精

度,甚至低于很多利用传统特征的方法,这是因为第一层提取的特征都是些低级特征,如边缘、角点等。因此,在本文方法中,AlexNet模型特征提取层选择Conv5层。由于Weizmann horse数据集中的图像与Stanford background数据集中的图像相似性较强,因此不再针对Weizmann horse数据集进行不同层特征精度比较,同样选择Conv5层作为特征提取层。

4.4 迭代次数影响分析

为了分析迭代次数对方法性能的影响,在Stanford background数据集下比较本文方法和方法2在不同迭代次数下的分类精度。本文CRF模型利用文献[5]中的方法估计参数,其采用截断拟合策略,在模型训练阶段预设置确定的迭代次数,提高了模型效率。不同特征提取方法提取到的训练特征会影响使CRF模型达到最好性能的最优迭代次数,因此,针对不同的训练特征来找到最优迭代次数对语义分割精度的提高是非常重要的。在实验中通过一定的间隔设置不同的迭代次数来训练模型参数,观察其对语义分割精度的影响,从而确定最优迭代次数。

由4.2节可知,由于增加了纹理特征,方法2性能全面优于方法1,因此4.4节和4.5节中都仅选择方法2作为对比。

从图3可以看出,方法2在100次迭代时达到最优性能,本文方法在第1000次迭代时达到最优性能。在迭代次数小于400次时,方法2的性能较本文方法稍好,在迭代次数为100时,分类精度比本文方法高0.016,其他情况下性能差距很小,但当迭代次数高于400次时,本文方法性能更好,在迭代次数为1000时,比方法2性能高0.077。

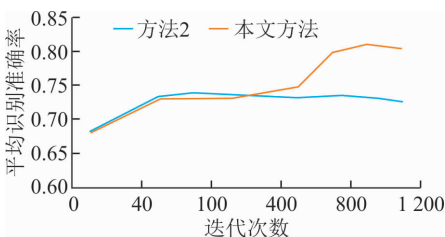


图3 Stanford background数据集下不同迭代次数影响比较
Fig. 3 Comparison of different iterations under Stanford background dataset

4.5 效率分析

为分析本文方法的效率,在Stanford background数据集下比较本文方法和方法2的特征提取时间以及训练时间。为便于对比,方法2训练阶段的迭代次数设置为100次,本文方法训练阶段的

迭代次数分别设置为100次和1000次。方法2的特征提取时间包含提取纹理和HOG特征的时间。本文的特征提取在MatConvNet深度学习平台上完成,提取Conv5特征并计算特征提取时间。

从表3可以看出,在特征提取阶段,利用AlexNet模型提取的深度特征耗时比人工特征(纹理和HOG)耗时多28.8s。在CRF模型训练阶段,当迭代次数都为100时,本文方法与方法2的训练时间相差很小。而本文方法在迭代次数为1000时,训练耗时为4045s。从总时间上来看,在相同迭代次数下,不同特征提取方法对总时间影响非常小,特征提取的时间仅占总时间的很少一部分。结合图3和表3可知,在迭代次数都为100时,本文方法与方法2准确率和效率相差无几,但此时方法2已达到最优性能,而本文方法在达到最优迭代次数之前,可以随着迭代次数的增加,在准确率上大幅超过方法2。因此,本文方法要优于方法2。

表3 Stanford background数据集下的效率比较
Tab. 3 Comparison of efficiency under the Stanford background dataset

方法	特征提取时间/s	CRF模型训练时间/s	总时间/s
方法2 ^[5]	2.2	611	613.2
本文方法(100)	31.0	648	679.0
本文方法(1000)	31.0	4045	4076.0

5 结语

本文提出一种基于深度CRF的方法,首先利用经典的CNN模型AlexNet来提取图像深度特征,再通过CRF对多特征及上下文信息有效利用来完成图像的语义分割。在Stanford background数据集和Weizmann数据集上的实验表明,本文算法由于采用CNN特征,与采用传统特征的方法相比,有效提升了识别准确率,但存在效率较低的问题。因此,如何在提升识别准确率的同时提高效率,是下一步的研究方向。

参考文献(References):

- [1] RUI C, BATISTA J, SMINCHISESCU C. Semantic Segmentation with Second-Order Pooling[C]// European Conference on Computer Vision. Florence, Italy: Springer-Verlag, 2012:430-443.
- [2] DOMKE J. Learning Graphical Model Parameters with Approximate Marginal Inference[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,

- 2013, 35(10):2454-2467.
- [3] KUMAR S, HEBERT M. Man-Made Structure Detection in Natural Images Using a Causal Multiscale Random Field[C]// IEEE International Conference on Computer Vision and Pattern Recognition. Tokyo, Japan: IEEE, 2003:119-126.
- [4] FENG X, WILLIAMS C K I, FELDERHOF S N. Combining Belief Networks and Neural Networks for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24 (4): 467-483.
- [5] FUREY T S, CRISTIANINI N, DUFFY N, et al. Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data[J]. Bioinformatics, 2000, 16(10):906-914.
- [6] FULKERSON B, VEDALDI A, SOATTO S. Class Segmentation and Object Localization with Superpixel Neighborhoods [C]// International Conference on Computer Vision. Kyoto:IEEE, 2009:670-677.
- [7] LUCCHI A, LI Y, SMITH K, et al. Structured Image Segmentation Using Kernelized Features [M]. [S. l.]: Springer Berlin Heidelberg, 2012.
- [8] LUCCHI A, LI Y, FUA P. Learning for Structured Prediction Using Approximate Subgradient Descent with Working Sets [C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, Oregon: IEEE, 2013:1987-1994.
- [9] 钟平. 面向图像标记的随机场模型研究[D]. 长沙:国防科学技术大学, 2008.
- ZHONG P. Research on Random Field Model for Image Marking [D]. Changsha: National University of Defense Technology, 2008. (in Chinese)
- [10] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [C]// Eighteenth International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann Publishers Inc, 2001:282-289.
- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. International Conference on Neural Information Processing Systems, 2012, 60 (2): 1097-1105.
- [12] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]//Proc of IEEE Conference on Computer Vision & Pattern Recognition. Miami, FL: IEEE, 2009:248-255.
- [13] ZHENG Y, LI Z, ZHANG C. A Hybrid Architecture Based on CNN for Image Semantic Annotation[M]. [S. l.]: Springer International Publishing, 2016.
- [14] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. Santiago, Chile: IEEE Computer Society, 2015:3431-3440.
- [15] CHEN L C, PAPANDEIOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40 (4): 834-848.
- [16] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
- LU H T, ZHANG Q C. A Review on the Application of Deep Convolutional Neural Networks in Computer Vision [J]. Data Acquisition and Processing, 2016, 31 (1): 1-17. (in Chinese)
- [17] LI S Z. Markov Random Field Modeling In Computer Vision[M]. [S. l.]: Springer-Verlag, 1995.
- [18] SHEN F, GAN R, YAN S, et al. Semantic Segmentation via Structured Patch Prediction, Context CRF and Guidance CRF[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society, 2017:5178-5186.
- [19] GRIBBON K T, BAILEY D G. A Novel Approach to Real-Time Bilinear Interpolation [C]// IEEE International Conference on Field-Programmable Technology. Bombay, India : IEEE, 2004:126-131.
- [20] 孙巍, 郭敏. 基于 SLIC 与条件随机场的图像分割算法[J]. 计算机应用研究, 2015, 32(12):3817-3820.
- SUN W, GUO M. Image Segmentation Algorithm Based on SLIC and Conditional Random Fields [J]. Application Research of Computers, 2015, 32 (12): 3817-3820. (in Chinese)
- [21] GOULD S, FULTON R, KOLLER D. Decomposing a Scene Into Geometric and Semantically Consistent Regions[J]. IEEE International Conference on Computer Vision, 2009, 30(2):1-8.
- [22] BORENSTEIN E, SHARON E, ULLMAN S. Combining Top-Down and Bottom-Up Segmentation[C]// Conference on Computer Vision and Pattern Recognition Workshop. Washington DC:IEEE, 2005:46-46.
- [23] VEDALDI A, LENC K. MatConvNet: Convolutional Neural Networks for MATLAB[C]// Proceedings of the 23rd ACM International Conference on Multimedia. New York, NY: ACM, 2014:689-692.