ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 32,No. 6,Nov. 2017,pp. 1097－1106
DOI:10.16337/j.1004-9037.2017.06.004

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

适用于文字检测的候选框提取算法

朱盈盈[1]　张拯[1]　章成全[1]　张兆翔[2]　白　翔[1]　刘文予[1]

(1.华中科技大学电子信息与通信学院,武汉,430074;2.中国科学院自动化研究所类脑智能研究中心,北京,100080)

摘　要: 在文字检测的相关研究中,针对文字的候选框提取方法并未得到广泛关注与深入挖掘。一方面由于文字本身结构和一般物体具有较强的差异性,另一方面由于文字对检测的精度要求高。本文提出了一种针对文字的候选框提取算法,该算法首先利用全卷积网络进行快速预测文字区域,有效地减少了候选框的搜索范围,然后针对文字特性对 EdgeBox 算法进行改进,使之适用于自然场景文字候选框的提取。此外,本文在两个自然场景文字检测的标准数据集上对该算法进行了评测,并与其他已有的候选框提取方法进行了比较。实验结果表明本文方法相较其他算法,具有更好的性能和鲁棒性。

关键词: 物体候选框;自然场景文字检测;全卷积网络;EdgeBox

中图分类号: TP391　　**文献标志码:** A

Proposal Extraction Method for Text Detection

Zhu Yingying[1], Zhang Zheng[1], Zhang Chengquan[1], Zhang Zhaoxiang[2], Bai Xiang[1], Liu Wenyu[1]

(1. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China;2. Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Science, Beijing,100080, China)

Abstract: In the study of text detection, the proposal extraction method is not widely concerned and deeply studied, due to the structure of the text and otherness of the general object and the high precision requirement of text detection. In this paper, we propose a proposal extraction method for text detection. The proposed method firstly utilize the fully convolutional network to predict the text regions, which can effectively reduce the search range of the proposal extraction. Then, the EdgeBox algorithm is improved to make it suitable for the text proposal extraction in natural scenes. In addition, the proposed method is evaluated on two standard natural scene text detection benchmarks, and compared with other existing methods. Results show that the proposed method has better performance and robustness than other methods.

Key words: object proposal; scene text detection; fully convolutional network; EdgeBox

---

基金项目: 国家自然科学基金优秀青年基金(61222308)资助项目;国家自然科学基金重点(61733007)资助项目;国家自然科学基金(61573160,61572207)资助项目;教育部新世纪优秀人才支持计划(NCET-12-0217)资助项目;华中科技大学自主创新基金资助项目。

收稿日期: 2016-04-16;**修订日期:** 2016-04-19

引 言

随着智能手机、可穿戴设备等移动终端的广泛普及，越来越多的新型应用场景需要利用自然场景中丰富的文字信息，例如图像搜索、场景理解、人机互动、目标定位和自动巡航等[1,2]。因此，自然场景中的文字自动阅读(包含文字检测、文字识别和文字语种识别等环节[3,4])成为了近几年计算机视觉和文档分析领域的热门研究课题。文字检测作为文字自动阅读的第1步，受到了越来越多的国内外学者的关注。然而自然场景下文字本身的多样性和背景的复杂性使得自然场景文字的自动检测变得极具挑战性。自然场景文字的多样性体现在文字的字体、大小、方向和颜色等存在多种可能。背景复杂性，体现在自然场景中某些特殊纹理和形状如：符号、窗户、树叶和杂草等与文字较为相似很难被区分开。再加上光照不均、遮挡、模糊、噪声和低分辨率等因素，使得自然场景中的文字检测变得极为困难。为了解决上述困难，近几年已经有很多学者提出了许多文字检测的方法。这些方法主要可以分为3类：基于纹理的方法、基于连通域的方法和混合的方法。基于纹理的方法[5-7]将文字当作一种特殊的纹理，结合滑动窗方法取得文字候选框，并利用文字的纹理特征如边缘[8]等信息来区分文字和非文字部分。由于该类方法需要扫描较多的尺度，因此通常基于纹理的方法计算复杂度很高，对于尺度变化非常敏感。基于连通域的方法[9-12]主要通过文字笔画、颜色分割和提取极值区域等方法提取出候选的联通区域，然后通过人为设计的规则或者分类器的方法滤除非文字区域，最后结合一系列后续的部件组合策略和分词手段来获得最终的文字候选框。一般来说，这种方法效率更高，而且对尺度变化不敏感，因此成为近几年文字检测领域的主流方法。混合的方法[13-14]结合了基于纹理的方法和基于连通域的方法，利用了这两类方法的优势，通过连通域的方法获取候选字符，结合纹理的特征进行过滤，从而获得较好的性能。总结分析，可以发现这3类方法其实都是自底向上的流程：先取得候选字符，然后结合一系列后续的组合、过滤和分词手段，得到最终的文字候选框，尤其是英文文字检测，往往最终需要得到的候选框是单词级别的包围盒。显然，这样的文字检测流程相当繁琐，整个算法对于中间每个环节的性能要求都是相当高。而本文的出发点，则是直接寻求单词级别候选框，从而可以舍弃掉繁琐的后续处理步骤。

近些年，物体候选框检测算法[15-18]在计算机视觉领域中受到了越来越广泛的关注。作为目标检测算法的前处理，这类算法可以快速地提供物体的潜在位置，缩小目标检测的搜索范围，有效地降低了目标检测算法的资源开销以及提高检测速度。物体候选框的提取方法大致可以归纳为3大类别：基于区域的方法、基于边缘的方法以及两者结合的方法。基于区域的方法，通常先将图片划分成多个子块，然后结合多种区域融合的手段将不同子块进行多层次组合，通过计算组合结果的包围盒信息来获得物体候选框。例如选择性搜索方法[19]对图像首先进行超像素分割，然后通过多种策略来逐层组合超像素图像生成物体候选框。Yanulevskay等[20]则通过一个可学习模型来融合超像素，从而进一步改善选择性搜索方法。此外，也有一些方法通过直接分割出物体来生成候选框，如多尺度组合分组(Multiscale combinatorial grouping, MCG)文献[21]采取了多层次和多尺度图像区域分割方法，直接生成物体的候选框。基于边缘的做法相对少一些，典型的代表工作有程明明等[16]提出的BING方法，通过提取滑动窗口内图像区域的梯度特征，并对特征值进行二值化处理，利用该二值化特征训练一个分类器，能够快速地为每个滑动窗赋予得分，该方法效率高、速度快，但得到的框不是很紧凑。Piotr等[22]提出了Edge Boxes方法，该方法建立在一般物体具有闭合的轮廓假设基础上，根据滑动窗所包含的边缘片段是否构成闭合轮廓进行得分评估，该方法得到的物体候选框在性能和速度上均达到较高水准，是目前综合性能最好的方法之一。两者结合的最新工作也有一些，比如Alexe等[23]通过结合多种区域的特征信息来对物体候选框进行排序，从而有较为显著的性能提升。而Kuo等[24]通过学习一个卷积神经网络模型来给物体候选框进行快速有效的重排序。在最近的一些物体检测工作中，R-CNN[25]和Fast R-CNN[26]都是结合候选框的方法实现物体检测。该系列方法首先提取目标物候选框，然后训练一个卷积神经网络模型[27]用

来提取每个候选框对应图像区域的深度学习特征,最后用深度学习特征训练支持向量机分类器进行最终的判决,该系列方法在物体检测公开数据集上取得了领先的结果。在文字检测领域,也有少量学者将物体候选框检测算法应用于文字检测这一问题上。文献[28]提出一种基于物体候选框的端到端文字检测与识别算法,该算法使用 EdgeBox 算法与 ACF 算法产生单词级别的文字候选框。文献[29]则在最大稳定极值区域(Maximally stable extremal regions,MSER)的基础上,设计了针对文字的候选框检测方法,并取得了良好的效果。受到上述工作的启发,本文提出了一种基于全卷积网络和改进的 EdgeBox[15] 的文字候选框提取算法,该方法能够快速准确地直接找到文字候选框。首先利用全卷积网络预测文字的大概区域,有效减少了候选框的搜索范围。另外,针对文字的特性,本文改进了 EdgeBox 算法,使之能够更准确地提取文字的候选框。

1 文字候选框提取算法整体流程

本文采用区域分割与目标候选框提取相结合的方法对自然场景中的文字候选框进行提取。文字候选框提取算法大致由两个步骤组成:文字区域分割和文字候选框提取,具体流程如图 1 所示。自然场景图片中的文字往往较小,且可能出现在图片的任意位置。文字区域分割可以提取出自然场景图片中可能出现文字的位置,缩小了文字候选框的搜索范围,降低了算法的时间复杂度,同时提高了候选框定位的准确性。在分割出文字区域后,对于每一块分割出来的图像区域,本文采用 EdgeBox 算法[15] 进行文字候选框的提取。由于传统的 EdgeBox 算法所使用的边缘提取算法对文字这一特殊物体的边缘提取效果较差,因此,本文使用全卷积网络改进了文字边缘的提取效果(图 1),并在 EdgeBox 算法提取候选框的过程中,将其与 EdgeBox 算法所采用的边缘相结合,从而提高了算法性能。

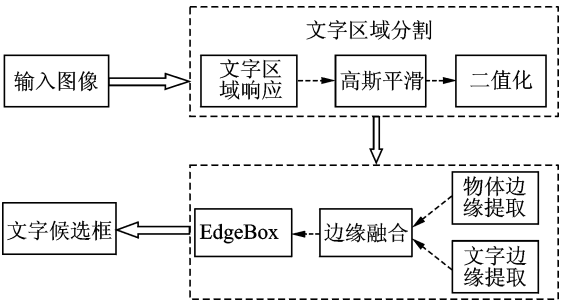


图 1 文字候选框提取流程图

Fig. 1 Pipeline of the proposed algorithm

2 基于全卷积网络的文字区域分割

2.1 全卷积网络

全卷积网络不同于传统的卷积神经网络,网络中不包含全连接层,而仅由卷积层、采样层和激励层等组成,摆脱了网络输入图像的尺寸限制,能够做到端到端的像素级别识别,如目标分割[30] 和边沿检测[31] 等工作。由于网络中不同卷积层对应的感知域的大小不同,因此不同的卷积层捕捉到的特征往往蕴含着来自不同层次和不同视野的上下文信息。浅层的卷积特征代表着低层次的、窄视野的上下文信息,高层的卷积特征代表着高层次的、宽视野的上下文信息。而将不同卷积层捕捉到的特征进行有效结合,则能够生成更具丰富性、层次性的特征表示。在全卷积网络中不同卷积层产生的特征图大小并不一致,为了融合不同尺度的特征图,要对尺度较小的特征图进行上采样。如图 2 所示,由不同卷积层产生的特征图,通过上采样层,被还原至原图尺寸,再输入到由卷积核大小为 1×1 的卷积层替代的线性判别器,生成最终的响应图。用数学公式表示为

$$F^k=\alpha_k f^k \tag{1}$$

$$M_{ij}=\sum_k w_k F_{ij}^k \tag{2}$$

式中:f^k 为第 k 层的特征图,α_k 为上采样的比例,M_{ij} 对应原图位置(i,j)的响应。w_k 为最终链接的 $1\times$

1 卷积层的第 k 组权重。

2.2 文字区域响应与分割

文字是一种具有独特纹理特性的物体，通过刻画这种纹理特性，可以较为容易地将文字与背景区分开。本文把落在文字条包围盒内的像素点当作是正样本，把落在文字条包围盒外的像素点视作负样本，利用全卷积网络训练了一个预测每个像素点是否属于文字区域的模型。图 3(a)为输入的自然场景图片，利用训练好的全卷积网络模型产生文字区域响应图，如图 3(b)所示。利用窗口大小为的 k 高斯滤波器对响应图进行平滑处理，然后设定判别阈值 T 进行二值化处理，如图 3(c)。对于二值化后的响应图，按照 8 邻域联通的方法进行连通域的求解，并计算每个连通域的最小包围盒。最后根据这些最小包围盒，将对应的文字候选区域提取出来。

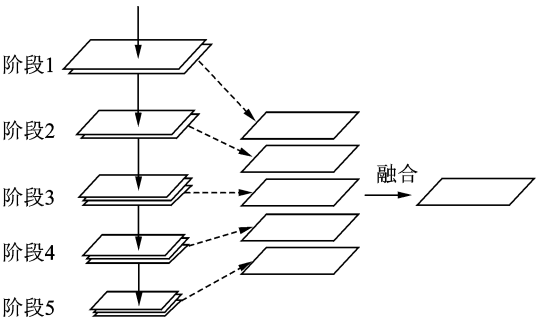
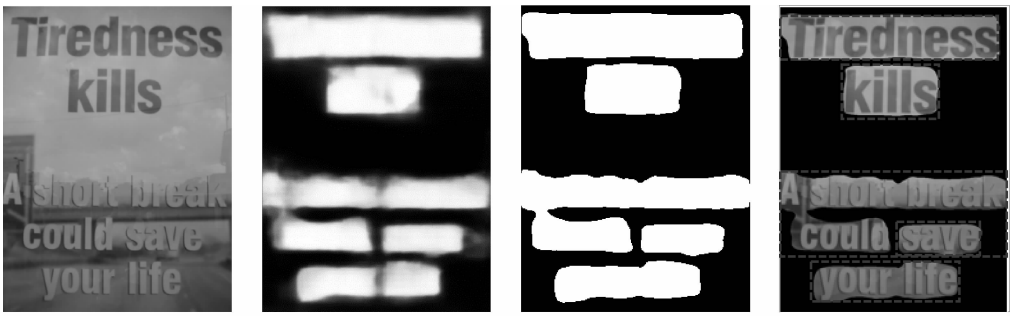


图 2 全卷积网络架构

Fig. 2 Architecture of fully convolutional network



(a) 原始图像 (a) Original image　(b) 文字区域响应图 (b) Text region map　(c) 二值化响应图 (c) Binary image　(d) 文字候选区域 (d) Text region candidates

图 3 文字区域分割

Fig. 3 Text region segmentation

3 基于 EdgeBox 的文字候选框提取

3.1 EdgeBox 算法

该算法由 Piotr 等[15]提出，其主要思想是基于物体边缘信息，提取一系列物体候选框，用于加速现有的目标物体检测算法。该算法首先利用文献[32]的结构化边缘检测算法求得图像的边缘响应图，接着根据边缘点的空间和方向约束进行组合得到多组边缘片段，然后根据边缘片段与候选框的空间几何关系给每个边缘片段赋予权值，并根据这些边缘片段的权值给候选框赋予得分。算法的具体步骤归纳如下：

步骤 1 输入图像，利用文献[32]的方法对图像进行边缘检测，通过非最大抑制处理得到最终的边缘响应图。

步骤 2 根据边缘点的空间相邻关系和方向的约束对边缘点进行组合，得到多个边缘片段。

步骤 3 计算任意两个边缘片段的相似度，其计算公式为

$$a(s_i, s_j) = \left| \cos(\theta_i - \theta_{ij}) \cos(\theta_j - \theta_{ij}) \right| \tag{3}$$

式中：s_i 和 s_j 为两个边缘片段，θ_i，θ_j 和 θ_{ij} 分别为两组边缘点的平均方向角和两组边缘之间的方向夹角。

步骤 4 给定一个候选框,计算每个边缘片段在该候选框内的权值。具体可以分为 4 类:位于候选框内的边缘片段权值为 1,和候选框边界相交的边缘片段权值为 0,边缘点和边缘点几何中心均在候选框之外的边缘片段权值也为 0,而剩下的边缘点全在候选框外部而其几何中心位于候选框内部的边缘片段其权值计算方式为

$$w_b(s_i)=1-\max_{T}\prod_{j}^{|T|-1}a(t_j,t_{j+1}) \tag{4}$$

式中:T 为该边缘片段往候选框内部延伸,和候选框边界相交的边缘片段组成的路径。如果该路径不存在,则将权值赋为 1。

步骤 5 统计和候选框相关联的所有边缘片段并计算候选框的最终得分,其得分计算方式为

$$h_b=\frac{\sum_i w_b(s_i)m_i}{2(b_w+b_h)^{\kappa}} \tag{5}$$

$$h_b^{\text{in}}=h_b-\frac{\sum_{p\in b^{\text{in}}}m_p}{2(b_w+b_h)^{\kappa}} \tag{6}$$

式中:m_i 和 b_w 均为边缘片段的模长;b_h 和 b^{in} 分别为候选框的长和宽;b^{in} 为落在候选框内的边缘片段的集合;κ 值取 1.5,用于抵消大的矩形框包含较多的边缘片段所带来的影响。文献[15]在一般物体候选框提取的实验中发现,位于物体内部的边缘片段往往没有物体轮廓上的边缘重要,所以最终候选框的得分由式(6)计算的 h_b^{in} 来表示。

3.2 EdgeBox 提取文字候选框

3.2.1 文字边沿提取

实验中发现,尽管 EdgeBox 算法所使用的边缘提取算法适用于一般物体候选框的提取,然而对于文字而言并不十分合适。受到全卷积网络在一般物体边缘检测[31]上取得的突破性成果的启发,以文字的边缘所在像素点为正样本,其他像素点为负样本,训练了一个全卷积网络,针对自然场景图片中的文字边缘进行提取。图 4(a)为输入自然场景图片,图 4(b)为 EdgeBox 所使用的边缘图片,图 4(c)为针对文字边缘训练的全卷积网络所产生的文字边缘响应图。可以看到,改进后的边缘检测模型对文字边缘的响应更强,同时也能抑制非文字边缘的响应。实验中,将两种边缘融合在一起,可以取得更好的效果,融合的边缘图如图 4(d)所示。

(a) 输入自然场景图片 (a) Input image

(b) 边缘图[27] (b) Edge map[27]

(c) 本文边缘图 (c) Proposed edge map

(d) 两种边缘图混合边缘图 (d) Fused edge map

图 4 文字边缘响应图

Fig. 4 Text edge response map

3.2.2 候选框得分计算

相对于一般物体而言,文字由许多笔画组成,不仅具备有较为规律的外轮廓边缘特性,同时其内部的边缘结构也呈现特殊的规律性,这也是文字与窗户、树叶和杂草等复杂易混淆背景的显著区别之处。

因此,在计算候选框得分时,不应像一般物体一样,忽略内部的边缘信息,而应当将文字的外轮廓边缘信息和文字内部的边缘信息等同对待。因此,在计算每个候选框的得分时,本算法加上了选框内部的边缘得分,即用式(5)计算的 h_b 作为最终的候选框得分。与其他文字候选框提取方法或一般物体的候选框提取方法相同,本算法也使用得分来对所有候选框进行排序:得分越高的候选框,其包含文字的概率越大。

4 实验结果与分析

为了与其他最新的文字候选框提取算法进行全面的比较,在 ICDAR2013 和 SVT 两个标准数据集对本文的方法进行了评估。

4.1 数据集与评价标准

为了衡量本文所提出算法的性能,在两个标准数据集上将该方法与其他算法进行了比较。ICDAR2013:该数据集是 ICDAR2013 文字检测竞赛采用的标准数据集,共包含 229 张训练图片与 233 张测试图片。该数据集以水平文字为主,涵盖了复杂光照、模糊和低分辨率等各种极端情况下的文字图片。SVT:该数据集共包含 349 张(100 张训练图片和 249 张测试图片)从 Google 街景地图中采集的自然场景图片。相较于 ICDAR2013 数据集,SVT 数据集中文字尺度更小,背景更复杂,更接近自然场景。评价标准:本文采用与 TextProposal[29] 相同的评价标准衡量算法性能,即召回率-候选框数量曲线。该评价标准包含召回率、候选框数量和交并比(Intersection over union, IoU)3 个指标。其中,召回率被定义为检测到真实目标的数量占总数量的比例,若真实目标与某个候选框的 IoU 大于某个阈值,则认为该真实目标被检测到了。召回率-候选框数量曲线描述了候选框数量与召回率之间的关系,该曲线被广泛应用于一般物体候选框提取算法的评价。

4.2 实现细节

本文使用两个结构相同的全卷积网络分别用来提取文字区域与文字边缘。网络结构如图 2 所示。网络中的 5 个卷积阶段(共包含 13 层卷积层)使用 VGG-16 模型[33] 的参数进行初始化。对于每个卷积阶段都添加了一个预测分支,该分支包含一个核大小为 1×1 的卷积层和一个上采样层。为了结合不同尺度和不同阶段的信息,进一步将不同预测分支的响应结果进行线性融合,从而得到最终的响应图。对于文字区域的全卷积网络,扩增了 ICDAR2013 和 SVT 的训练集,并随机采样了 30 K 张大小为 500×500 的图片作为训练样本。在测试阶段,每张图片被按比例缩放到 3 个高度(高度为 200,500 和 1 000 像素)分别进行预测,不同尺度的响应图被缩放至原图大小并取均值,得到最终的响应图。对于文字边缘的全卷积网络,使用 ICDAR2013 的训练集提供的字符级分割数据作为边缘图的训练集:在字符分割的图片上使用 Canny 算子提取边缘,并以此作为文字边缘图进行训练。与文字区域网络的训练策略一样,也对 ICDAR2013 的数据进行了扩增,并随机采样了 30 K 张大小 500×500 的图片作为训练样本。测试阶段,每张图片均被按比例缩放至 800 像素的高度进行预测,再还原至原图大小,得到原图的边缘响应图。对于所有的实验,参数均设置为 $k=20$, $T=0.7$, EdgeBox 最大保留框个数为 10^4 个,最小面积为 100,最大长宽比为 10,其他参数均使用默认参数。实验环境为 2.0 GHz 8-core CPU, 64 GB RAM, GTX TitanX,所有程序都使用 Torch7、Matlab 与 C++实现。

4.3 各改进部分对性能的影响

本文针对文字的特性,在 EdgeBox 的基础上提出了 3 个部分的改进,为了分析这些改进对性能的影响,在 ICDAR2013 数据集上对这些改进进行分析。图 5 为使用不同策略时在 ICDAR2013 数据集上性能的比较情况。其中,EB 表示原始的 EdgeBox 算法,EB+C 表示结合了文字区域分割的算法,EB+CE 表示同时结合了文字区域分割与边缘融合的算法,EB+CEM 表示在 EB+CE 的基础上对得分计算方法进行了修改的算法。可以看到,相较于原始的 EdgeBox,3 种改进策略均在不同程度提高了算法的性

能。表 1 为在候选框保留个数为 1 000 时,不同部分的召回率结果:相较于原始的 Edge Box 算法,分别提高了 5%,9%和 11%的召回率。除召回率的提升之外,改进也降低了 EdgeBox 的时间复杂度。表 2 为原始的 EdgeBox 算法与本算法在 ICDAR2013 数据集上运行时间的比较与分析。EB 表示原始的 EdgeBox 算法的运行时间,Edge 和 Region 分别表示使用全连接网络提取文字边缘与文字区域的运行时间,本文提出的算法的最终运行时间已包含边缘提取和文字区域提取所消耗的时间。可以看到,尽管计算边缘和字符区域增加了额外的时间开销,但是由于字符区域大大降低了 EdgeBox 的搜索范围,因此算法的运行时间有了显著的降低。

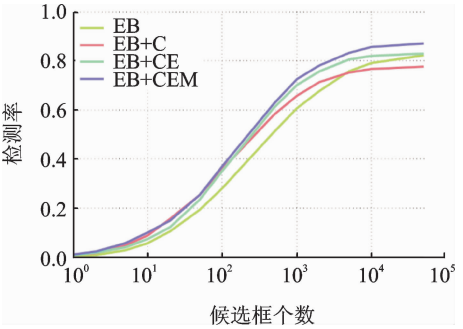


图 5 不同改进部分对性能的影响

Fig. 5 Effect of different improvements on performance

表 1 不同改进部分在保留候选框数量为 1 K 时的召回率

Tab. 1 Recall of different contributions on 1 K proposals

Contribution	EB	EB+C	EB+CE	EB+CEM
Recall	0.61	0.66	0.70	0.72

表 2 算法运行时间的比较与分析

Tab. 2 Comparison and analysis of the running time

Part	EB/s	Edge/s	Region/s	本文方法/s
t	3.7	0.34	0.7	2.19

4.4 与其他算法的比较

将本文提出的算法与其他文字候选框提取算法以及一般物体候选框提取算法进行比较。这些比较算法包括:TextProposal[29], BING[16],随机约束方法(Randomized Prim's, RP)[17],和可测量物体候选框(Geodesic Object Proposals,GOP)[18]。实验中,所有参与比较的算法均使用作者提供的公开代码或论文中的结果,并采用默认参数进行比较。ICDAR2013 上的比较结果:图 6 为不同算法在 ICDAR2013 上的比较结果。其中,TextProposal 有标准和快速两种算法,分别记为 TP_FULL 和 TP_FAST。可以看到,本文的算法在保留的候选框数量为 1 000 时的召回率均高于其他算法:在 IoU 为 0.5 和 0.7 时,比之前最好的文字候选框提取方法 TextProposal 分别提高 1%和 3%。在保留的候选框数量更少时,本文的方法在召回率上相较于其他算法的提升更加明显。在保留的候选框数量为 100 时,本文的方法比其他方法分别提升 11%(IoU=0.5)和 9%(IoU=0.7)。而在候选框数量更多时,本文的方法则在召回率上低于 TextProposal。SVT 上的比较结果:图 7 为不同算法在更接近于真实场景的 SVT 数据集上的比较结果。结果显示,本文提出的算法在 SVT 数据集上的性能要远好于其他算法,当候选框个数为 1 000 时,相比于之前最好的方法 TextProposal 在 IoU=0.5 和 IoU=0.7 时分别提升 14%和 33%。这体现出本文的算法在复杂场景下的有效性和鲁棒性。

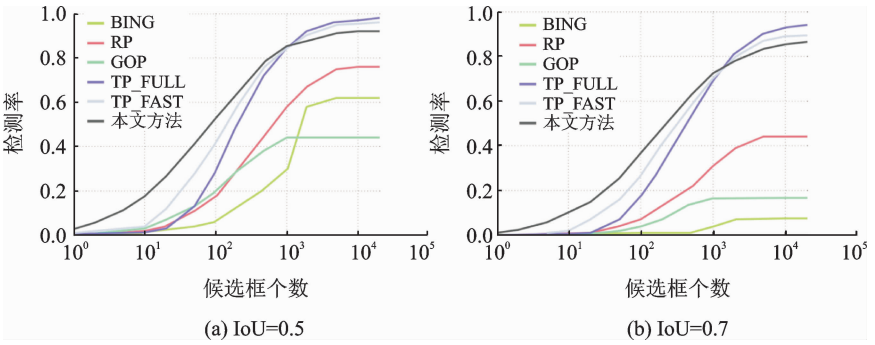


图 6 不同算法在 ICDAR2013 数据集下的比较

Fig. 6 Performance of different algorithms on ICDAR2013 Dataset

4.5 该算法存在的不足

本文提出的算法在两个数据集上均取得了较好的效果，图 8 展示了一些检测结果，其中红色矩形为目标候选框，绿色矩形为与目标候选框的 IoU 最大的文字候选框。然而，本文的方法依然存在不足，在一些复杂自然场景下候选框并不能很好捕捉住文字区域，如图 9 展示了一些失败的情形。分析这些错误的情形大致可以归结如下：对于强光照、低对比度或遮挡的情况下易造成区域响应的不连续，从而使得最终生成的候选框不能框住完整的文本条；对于尺度较小而且长宽较大的文本条，文字的外在轮廓信息不容易捕捉住，通过采取适当的多尺度缩放策略一定程度上可以解决，但也会造成多余的计算开销。此外，本文提出的文字候选框方法主要针对于水平或近似水平的文本设计，暂时没有考虑到多方向文本的情况。

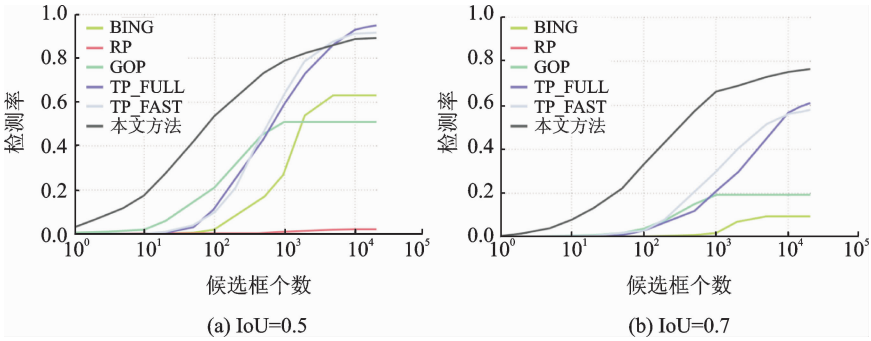


图 7 不同算法在 SVT 数据集下的比较

Fig. 7 Performance of different algorithms on SVT Dataset

图 8 本算法在 ICDAR2013 和 SVT 数据集上的结果

Fig. 8 Several cases in ICDAR2013 and SVT of our method

图 9 一些失败的结果

Fig. 9 Several failure cases of our method

5 结束语

本文提出了一种针对文字的候选框提取算法。该算法能够提供一系列规模较小、精度较准的文字(单词级别)候选框，大大减少了文本检测的搜索范围，能够有效地运用于文字检测。本文的主要创新工

作在于:(1)针对文字特性出发,将针对物体边缘和文字边缘的两种响应图有效融合,从而使得最终的边缘响应图能够同时具备物体的完整封闭特性和文字内部规律纹理特性。(2)针对性地改进 EdgeBox 对于文字候选框的得分计算方法,使得产生的候选框能更好地按照文字特性排序。在 ICDAR2013 与 SVT 两个标准数据集上进行评测,本文提出的文字候选框的性能均优于其他算法,在绝大多数场景本文算法所产生的候选框均能很好地覆盖住文本区域,从而证明了该算法的有效性和鲁棒性。然而,本文的工作依旧存在着较多改善空间。首先本文提出的候选框提取算法在较为复杂的场景下对文本条的捕捉能力还有较大的提升,尤其是在 SVT 数据集上,因此将区域特征和边缘信息有效结合从而设计更加合理的得分计算公式可能是一个较为不错的解决方向。此外,本文的算法是针对水平文字出发的,并不能解决多方向的文本检测,因此下一步研究多方向文本候选框提取方法将极具较大的现实意义。最后,本文将考虑结合文本候选框提取方法来实现完整的文字检测和识别系统。

参考文献:

[1] 罗斌, 部伟, 汤进, 等. 复杂环境下基于角点回归的全卷积神经网络的车牌定位[J]. 数据采集与处理, 2016, 31(1): 65-72.
Luo Bin, Gao Wei, Tang Jin, et al. Learning corner regression-based fully convolutional neural network for license plate localization in complex scene[J]. Journal of Data Acquisition & Processing, 2016, 31(1): 65-72.

[2] Chen Chunmin, Chen Lingwei. A novel approach for semantic event extraction from sports webcast text[J]. Multimedia Tools and Applications, 2014, 71(3): 1937-1952.

[3] Cong Yao, Bai Xiang, Liu Wenyu. A unified framework for multioriented text detection and recognition[J]. IEEE Transactions on Image Processing, 2014, 23(11): 4737-4749.

[4] Zhu Yingying, Yao Cong, Bai Xiang. Scene text detection and recognition: Recent advances and future trends[J]. Frontiers of Computer Science, 2016, 10(1): 19-36.

[5] Zhong Y, Karu K, Jain A K. Locating text in complex color images[C]//Document Analysis and Recognition, Proceedings of the Third International Conference on. [S. l.]: IEEE, 1995, 1: 146-149.

[6] Kim K I, Jung K, Kim J H. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003, 25(12): 1631-1639.

[7] Gllavata J, Ewerth R, Freisleben B. Text detection in images based on unsupervised classification of high-frequency wavelet coefficients[C]//Pattern Recognition, 2004, Proceedings of the 17th International Conference on. [S. l.]: IEEE, 2004, 1: 425-428.

[8] 邓彩霞, 侯杰, 张晓卫. 改进的自适应阈值方法用于文字图像边缘检测[J]. 数据采集与处理, 2006, 21(S): 63-66.
Deng Caixia, Hou Jie, Zhang Xiaowei. Improved adaptive threshold method and its application in character image edge detection[J]. Journal of Data Acquisition and Processing, 2006, 21(S): 63-66.

[9] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. [S. l.]: IEEE, 2010: 2963-2970.

[10] Yao C, Bai X, Liu W, et al. Detecting texts of arbitrary orientations in natural images[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. [S. l.]: IEEE, 2012: 1083-1090.

[11] Yi C, Tian Y L. Text string detection from natural scenes by structure-based partition and grouping[J]. Image Processing, IEEE Transactions on, 2011, 20(9): 2594-2605.

[12] Huang W, Lin Z, Yang J, et al. Text localization in natural images using stroke feature transform and text covariance descriptors[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2013: 1241-1248.

[13] Pan Y F, Hou X, Liu C L. A hybrid approach to detect and localize texts in natural scene images[J]. Image Processing, IEEE Transactions on, 2011, 20(3): 800-813.

[14] Yangxing Liu, Ikenaga T. A contour-based robust algorithm for text detection in color images[J]. IEICE transactions on information and systems, 2006, 89(3): 1221-1230.

[15] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[M]//Computer Vision—ECCV 2014. [S. l.]: Springer International Publishing, 2014: 391-405.

[16] Cheng M M, Zhang Z, Lin W Y, et al. BING: Binarized normed gradients for objectness estimation at 300fps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: 2014: 3286-3293.

[17] Manen S, Guillaumin M, Gool L. Prime object proposals with randomized prim's algorithm[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE,2013: 2536-2543.

[18] Krhenbühl P, Koltun V. Geodesic object proposals[M]. Computer Vision—ECCV 2014. [S. l.]:Springer International Publishing, 2014: 725-739.

[19] Van de Sande K E A, Uijlings J R R, Gevers T, et al. Segmentation as selective search for object recognition[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. [S. l.]:IEEE, 2011: 1879-1886.

[20] Yanulevskaya V, Uijlings J, Sebe N. Learning to group objects[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2014: 3134-3141.

[21] Arbeláez P, Pont-Tuset J, Barron J, et al. Multiscale combinatorial grouping[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2014: 328-335.

[22] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[M]//Computer Vision—ECCV 2014. [S. l.]: Springer International Publishing, 2014:391-405.

[23] Alexe B, Deselaers T, Ferrari V. What is an object[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. [S. l.]: IEEE, 2010: 73-80.

[24] Kuo W, Hariharan B, Malik J. Deepbox: Learning objectness with convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE, 2015: 2479-2487.

[25] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]:IEEE, 2014: 580-587.

[26] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE, 2015: 1440-1448.

[27] 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 Hongtao, Zhang Qinchuan. Applications of deep convolutional neural network in computer vision[J]. Journal of Data Acquisition & Processing, 2016, 31(1): 1-17.

[28] Jaderberg M, Simonyan K, Vedaldi A, et al. Reading text in the wild with convolutional neural networks[J]. International Journal of Computer Vision, 2016, 116(1): 1-20.

[29] Gomez L, Karatzas D. Object proposals for text extraction in the wild[C]//Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. [S. l.]:IEEE, 2015: 206-210.

[30] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2015: 3431-3440.

[31] Xie S, Tu Z. Holistically-nested edge detection[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE,2015: 1395-1403.

[32] Dollár P, Zitnick C. Structured forests for fast edge detection[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE,2013: 1841-1848.

[33] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the International Conference on Learning Representations. 2015.

作者简介:

朱盈盈(1988-),女,博士研究生,研究方向:自然场景中文字、交通标志的检测与识别,E-mail:yyzhu@hust.edu.cn。

张拯(1990-),男,硕士研究生,研究方向:自然场景文字检测与识别。

章成全(1990-),男,硕士研究生,研究方向:图像分类、场景文字检测。

张兆翔(1983-),男,研究员,研究方向:模式识别、计算机视觉、类脑智能。

白翔(1981-),男,教授,研究方向:物体识别、形状分析、场景文字识别、智能信息系统。

刘文予(1963-),男,教授,研究方向:无线通信、多媒体信息处理、计算机视觉,E-mail: wyliu@mail.hust.edu.cn。