文章编号：2095-2163(2024)02-0048-07 中图分类号：TP391 文献标志码：A

基于深度学习的场景文本检测方法研究综述

张 静，孙巧榆，刘珍兵
（江苏海洋大学 电子工程学院，江苏 连云港 222005）

摘 要：文本检测技术在社会中有着广泛的应用，随着深度学习的加入，文本检测技术得到了进一步的提升。近年来基于深度学习的检测算法逐渐增多，针对场景文本检测的各种问题提出了相应的解决方法，提升了场景文本检测算法的性能。本文对这些算法进行了归纳、分析和总结，将这些算法大致分为基于回归和基于分割两种类型，并对其性能进行了对比，最后基于这些算法的研究内容为文本检测领域未来的发展提出了新的研究方向。

关键词：深度学习；文本检测；场景文本

Review of scene text detection methods based on deep learning

ZHANG Jing, SUN Qiaoyu, LIU Zhenbing
(School of Electronic Engineering, Jiangsu Ocean University, Lianyungang Jiangsu 222005, China)

Abstract: Text detection technology has a wide range of applications in society, and with the integration of deep learning, it has been further enhanced. In recent years, the number of detection algorithms based on deep learning has gradually increased, and corresponding solutions have been proposed for various problems in scene text detection, improving the performance of these algorithms. This paper summarizes, analyzes, and concludes these algorithms, categorizing them into two main types: regression-based and segmentation-based. Their performances are compared, and based on the research on these algorithms, new research directions are proposed for the future development of the text detection field.

Key words: deep learning; text detection; scene text

0 引 言

自古以来，文字作为文化传承的一种载体，记载了中国发展的重要历史，使得华夏文明得以保存和流传。随着生活方式的改变，文字的保存方式逐渐由纸质稿转向电子稿，更利于文化的传播和信息的保存。如何识别文本图像中的文本成为广泛研究的课题，研究的对象也由文本文档图像转向自然场景文本图像。一方面，在某些固定场景下，提取图像中的文本在信息识别应用上发挥着重要的作用，如发票、银行卡等信息整理；另一方面，计算机视觉领域对于图像的研究也有重要的意义，图像中的信息可以帮助计算机“理解”图像，其中也包含图像中的文本信息，如自动驾驶技术和道路违法车辆拍摄的应用。但是自然场景下获取的图像是杂乱无章的，光照、模糊、复杂背景、文本方向和形状的随机性等都影响着文本图像信息的提取。文本图像的检测从传统方法转向基于深度学习方法，从单语种文本的研究到多语种文本的研究，在这个领域出现了大量的自然场景文本检测算法。传统的文本检测算法需要手动设计特征框，定文本位置，而基于深度学习的文本检测算法可以自主的学习文本特征，具有更强的泛化性。

本文对近几年对文本检测产生影响的基于深度学习的自然场景文本检测算法进行了介绍和分析。文本检测的深度学习算法大致分为基于回归与基于分割两种方法，也有将两者结合的其他算法。

1 基于回归的场景文本检测算法

文本检测可以看作一般目标检测的延伸，基于回归的方法采用目标检测框架检测文本，通过卷积得到文本区域提议，对文本区域提议进行分类和回

作者简介：张 静(1997-)，女，硕士研究生，主要研究方向：图像分析与智能系统；刘珍兵(1999-)，男，硕士研究生，主要研究方向：图像分析与智能系统。
通讯作者：孙巧榆(1973-)，女，博士，副教授，主要研究方向：机器智能与图像分析。Email：sunqy@jou.edu.cn

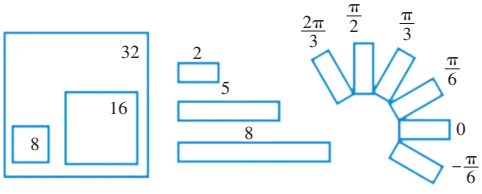
收稿日期：2023-03-03

归得到最终文本框，通过预设锚框得到文本框提议的称为间接回归，不预设锚框的称为直接回归。

1.1　间接回归法

对水平文本的研究，CTPN（Connectionist Text Proposal Network）采用垂直的锚框检测部分文本的位置，再根据文本的顺序特征进行连接细化，得到文本区域[1]；Inception RPN（Inception Region Proposal Network）将 SSD（Single Shot multibox Detector）的全连接层换成卷积层，并重新设计了锚框的比例，使之更适用于水平文本的检测[2]。

对多方向文本的研究，TextBoxes++使用长卷积核提取文本特征，对矩形框进行回归，得到多角度的文本框回归信息[3]；RRPN（Rotation Region Proposal Networks）将角度信息加入了 RPN 中，使用的锚框策略如图 1 所示，TextBoxes++和 RRPN 两种方式都会产生大量的候选框[4]；SBD（Sequential-free Box Discretization）将区域提议的边离散成 8 个关键边，通过学习序列标记匹配类型重建文本边框，避免了角度引起的不稳定[5]。



(a) 尺寸　　(b) 比例　　(c) 角度

图 1　RRPN 的锚点策略

Fig. 1　Anchor strategy of RRPN

对于弯曲文本和任意形状文本的研究，CTD（Curve Text Detector）分开回归外接矩形的高和宽，预测粗略提议边框上 14 个点高和宽的偏移量来修正文本框，采用 14 点坐标表示弯曲文本，如图 2 所示[6]；SLPR（Sliding Line Point Regression）对输出的候选矩形框用线进行横向和纵向的滑动，连接与文本边缘相交的点，形成多边形文本框，可以表示任意形状的文本[7]；ATRR（Adaptive Text Region Representation）使用 LSTM（Long Short-Term Memory）对候选提议边界框上的点坐标进行迭代回归，得到能够完整表示文本框的原图像提取轮廓线[8]；CSE（Conditional Spatial Expansion）对区域提议的文本区域节点重新扩展，再映射回采用 IoU（Intersection over Union）损失优化 RPN 的回归损失，分别对提议区域的垂直和水平方向的局部纹理进行建模，得到轮廓点，对轮廓点进行非极大值抑制等处理，减少文本的误报情况[9]。

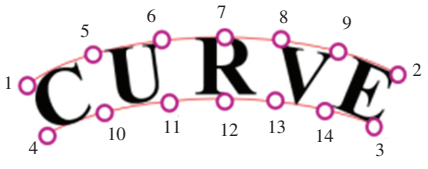


图 2　弯曲文本坐标表示

Fig. 2　Coordinate representation of curved text

1.2　直接回归法

对水平文本的研究，RRPN++将特征图分割成网格，对网格中坐标进行回归，减少推理时间[10]；AF-RPN（Anchor-Free Region Proposal Network）结合 FPN（Feature Pyramid Network）采用 3 个不同尺寸的特征图，分别检测大、中、小 3 种尺寸的文本，生成高质量的文本建议区域[11]。

对多方向文本的研究，EAST（Efficient and Accuracy Scene Text detection pipeline）直接由多通道全卷积网络进行特征提取，得到文本概率图、文本框坐标和角度，去除了候选建议和文本行生成等步骤，提高了文本检测的速度，但是对于较长的文本检测效果不理想[12]；RRD（Rotation sensitive Regression Detector）用对文本方向敏感和不敏感的两种特征图分别对文本框进行回归和分类，提高了对长文本的检测性能，但是对于过于分散的文本、并排的垂直文本可能会出现错误的文本框分割[13]；GNM（Geometry Normalization Module）生成多个几何感知图，增强网络对较大文本几何方差的学习能力[14]；ITN（Instance Transformation Network）采用几何感知的自适应感受野表征文本区域，直接检测得到多方向文本行[15]。针对检测文本容易断开、长文本定位不准确，数据集过少的问题，文献[16]对长文本进行分段，增加训练集数据，然后将文本的 4 条边看作 4 个类型对象进行训练，可以有效地检测过于靠近的文本行。IncepText 在特征提取阶段和 PSROI Pooling（Position Sensitive ROI Pooling）阶段中加入可变形卷积，调整网络的感受野和增加区域的偏移量以覆盖任意方向的文本[17]。

对任意形状文本的研究，ABCNet（Adaptive Bezier-Curve Network）采用 Bezier 曲线拟合任意形状的文本，用点固定曲线对曲线进行回归，转换成多边形文本框坐标，极大地提高了任意文本的检测速度[18]；TextBPN（Text Boundary Proposal Network）将边界建议转换成具有拓扑结构和序列上下文的 N 个点的表示方式，然后使用由图形卷积网络（Graph Convolutional Network，GCN）和 RNN（Recurrent

Neural Network)迭代细化点的结构,并使用先验信息加以引导,得到文本框[19]。

1.3 基于回归方法的性能对比

在公开数据集 ICDAR2013、ICDAR15 和MSRA-TD500 上,部分基于回归方法的性能对比见表 1,数据集 ICDAR2013 主要由水平文本组成,数据集 ICDAR15 和 MSRA-TD500 包含倾斜文本;在公开数据集 CTW1500 和 Total-text 上,部分基于回归方法的性能对比见表 2,数据集 CTW1500 和 Total-text 包含弯曲文本。

表 1 和表 2 中 P 表示准确率,R 表示召回率,F 表示综合评价指标。

表 1 部分基于回归的方法在 ICDAR2013、ICDAR15 和 MSRA-TD500 上的性能对比

Table 1 Performance comparison of regression-based methods on ICDAR2013, ICDAR15 and MSRA-TD500

方法	ICDAR2013			ICDAR2015			MSRA-TD500		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
CTPN[1]	0.93	0.83	0.88	0.74	0.52	0.61	-	-	-
Inception RPN[2]	0.87	0.83	0.85	-	-	-	-	-	-
TextBoxes++[3]	0.74	0.86	0.80	0.76	0.87	0.81	-	-	-
RRPN[4]	0.95	0.88	0.91	0.84	0.77	0.80	0.82	0.69	0.75
SBD[5]	-	-	-	0.88	0.92	0.90	-	-	-
CTD[6]	-	-	-	-	-	-	0.77	0.84	0.80
SLPR[7]	-	-	-	0.85	0.83	0.84	-	-	-
ATRR[8]	0.89	0.93	0.91	0.86	0.89	0.87	0.82	0.85	0.83
RRPN++[10]	-	-	-	0.86	0.87	0.86	-	-	-
AF-PPN[11]	-	-	0.92	-	-	0.89	-	-	-
EAST[12]	0.90	0.94	0.92	0.83	0.89	0.86	-	-	-
RRD[13]				0.78	0.83	0.80	0.67	0.87	0.76
GNM[14]	0.75	0.88	0.81	0.79	0.85	0.82	0.73	0.87	0.79
ITN[15]	-	-	-	0.86	0.90	0.88	-	-	-
文献[16]	-	-	-	0.85	0.74	0.79	0.90	0.72	0.80
IncepText[17]	0.87	0.91	0.89	-	-	-	0.77	0.83	0.80
TextBPN[19]	-	-	-	0.80	0.90	0.85	0.79	0.87	0.83
CTPN[1]	-	-	-	-	-	-	0.84	0.86	0.85

表 2 部分基于回归的方法在 CTW1500 和 Total-text 上的性能对比

Table 2 Performance comparison of regression-based methods on CTW1500 and Total-text

方法	CTW1500			Total-text		
	P	*R*	*F*	*P*	*R*	*F*
CTD[6]	0.69	0.77	0.73	0.71	0.74	0.73
SLPR[7]	0.80	0.70	0.74	-	-	-
ATRR[8]	0.80	0.80	0.80	0.76	0.80	0.78
CSE[9]	0.81	0.76	0.78	0.81	0.79	0.80
ABCNet[18]	0.84	0.83	0.83	0.83	0.86	0.85
TextBPN[19]	-	-	0.74	-	-	0.75
CTD[6]	0.83	0.86	0.85	0.85	0.90	0.87

1.4 小结

基于回归方法的难点主要在于极端纵横比文本的检测和任意形状文本框的拟合,基于间接回归方法的改进方向主要在于设计锚框的比例和文本轮廓的修正,通过几何感知增强对大文本的检测,采用多尺度特征融合,增强特征图的特征提取,以及对文本边框进行训练生成文本轮廓。

2 基于分割的场景文本检测算法

基于分割的方法多数受到 FCN(Fully Convolutional Networks)的影响,通常是对图像进行像素级的预测,然后将预测为文本的像素聚合成单个字符,再进行连接或者直接聚合为文本行[20]。

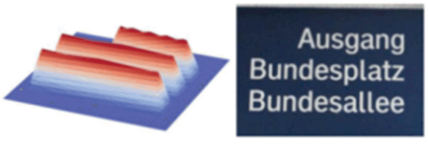
2.1 基于分割的场景文本检测算法

2.1.1 基于像素关系的方法

CCTN(Cascaded Convolutional Text Network)在

特征提取过程中加入并联的多尺度的卷积核层，增强特征图对多尺度文本的捕获能力，并根据文本几何特征提取文本行，用来检测水平文本[21]；MCN（Markov Clustering Network），将特征图根据文本的局部相关性和语义信息进行编码，生成随机流图（Stochastic Flow Graph，SFG），在 SFG 上采用马尔洛夫聚类（Markov Clustering，MC）将属于同一文本行的像素进行聚类，可以检测尺寸变化大和旋转的文本[22]。PMTD（Pyramid Mask Text Detector）将像素点与文本中心的距离软映射到[0，1]区间内作为像素点的值，将像素点的值作为高度得到金字塔，然后使用平面聚类算法对金字塔的底部进行迭代，得到最适合的文本框，避免了大量非字符区域被划分到正样本的集合中进行训练的问题[23]；DBNet（Differentiable Binarization Net）提出可微分二值化模块作为二值化的近似函数，进一步区分文本与非文本，提高检测效果，后来增加了 ASF（Adaptive Scale Fusion）模块，用空间注意力学习不同尺度的特征图的权重，增加尺度融合的鲁棒性[24]。

针对相近文本粘连和长文本检测问题，文献[25]引入形状感知嵌入（Learning Shape－Aware Embedding）检测文本，一个分支在特征合并模块中加入位置信息得到嵌入特征，一个分支生成文本中心图和文本全图进行文本分割，根据像素的嵌入距离对像素进行聚类得到文本区域；TextMountain 首先计算得到文本中心边界概率和文本中心方向，将像素的概率值升高得到中心边界概率，如图 3 所示；根据边界像素概率增长的方向和文本中心方向对文本进行分组，避免了因文本方向和感受野过小带来的问题[26]。



中心边界概率　　原始图像

图 3　TextMountain 图示例

Fig. 3　Example of TextMountain

2.1.2　基于组件连接的方法

SegLink（Segment Linking）检测带角度的单个字符，根据不同尺度特征上字符之间的关系将字符连接成文本行[27]；TextSnake 根据文本区域找出文本中心线，在中心线上画一系列的圆，连接包裹圆的最小多边形得到文本框[28]；CRAFT（Character Region Awareness For Text detection）通过高斯热图编码像素的区域得分和亲和度得分，分别得到字符框和相邻字符亲和度框，然后连接字符得到文本行[29]；文献[30]将文本组件之间的关系转成非欧距离，修改 IPS（Instance Pivot Subgraph），将单个文本组件转换成局部图作为一个节点，然后根据节点的 RROI（Rotated Region of Interest）特征和几何属性使用 GCN 对节点进行深度关系推理并连接生成文本行。

2.1.3　基于文本中心扩展的方法

为了缓解过于靠近的文本的误检问题，PSENet（Progressive Scale Expansion Network）在最小核尺寸的分割图上得到彼此分离的文本区域，然后利用较大的尺寸文本核分割图对文本区域进行渐进膨胀扩展，如图 4 所示[31]。PAN（Pixel Aggre－gation Network）采用轻量级的主干网络，结合 FPN 增强特征图表达能力，并改变特征融合的方式，减少运算量；根据文本中心与文本像素的距离对文本像素进行聚类，从而在不降低精度的条件下提升检测速度[32]。

也有对引导掩码进行增强的方法，Guided CNN 在训练中加入随机合成文本扩展引导掩码，从而过滤更多背景区域，提高检测速度[33]。

2.2　部分基于分割算法的性能对比

在公开数据集 CTW1500 和 Total－text 上，部分基于分割算法的性能对比见表 3；在公开数据集 ICDAR2013、ICDAR15 和 MSRA－TD500 上，部分基于分割算法的性能对比见表 4。

表 3　部分基于分割的方法在 CTW1500 和 Total－text 上的性能对比

Table 3　Performance comparison of segmentation－based methods on CTW1500 and Total－text

方法	CTW1500			Total－text		
	P	*R*	*F*	*P*	*R*	*F*
DBNet[24]	0.86	0.80	0.83	0.87	0.82	0.84
文献[25]	0.87	0.82	0.85	0.88	0.83	0.86
TextMoutain[26]	0.77	0.82	0.80	－	－	－
TextSnake[28]	0.82	0.83	0.83	－	－	－
CRAFT[29]	0.67	0.85	0.75	0.82	0.74	0.78
文献[30]	0.81	0.86	0.83	0.79	0.87	0.83
PSENET[31]	0.83	0.85	0.84	0.84	0.86	0.85
PAN[32]	0.84	0.79	0.82	0.84	0.77	0.80
DBNet[24]	0.86	0.81	0.83	0.89	0.81	0.85

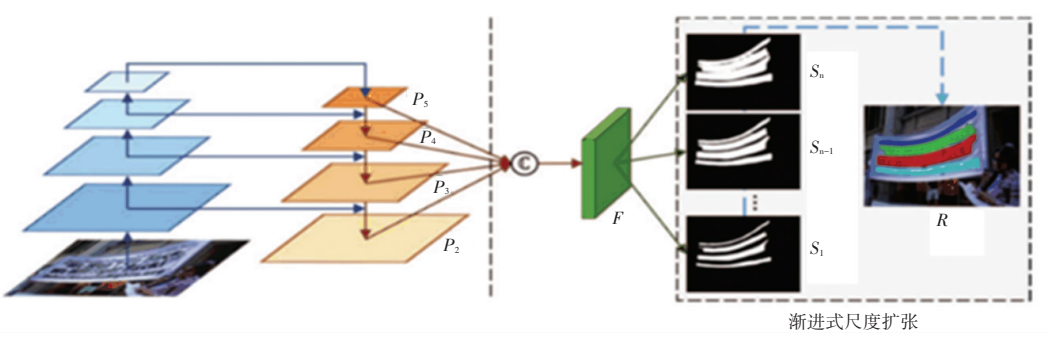


图 4 PSENet 网络结构

Fig. 4 Architecture of PSENet network

表 4 部分基于分割的方法在 ICDAR2013、ICDAR15 和 MSRA-TD500 上的性能对比

Table 4 Performance comparison of segmentation-based methods on ICDAR2013, ICDAR15 and MSRA-TD500

方法	ICDAR2013			ICDAR2015			MSRA-TD500		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
CCTN[21]	0.90	0.83	0.86	–	–	–	0.65	0.79	0.71
MCN[22]	0.88	0.87	0.88	0.72	0.80	0.76	0.88	0.79	0.83
PMTD[23]	–	–		0.85	0.72	0.78	0.91	0.87	0.89
DBNet[24]	–	–	–	0.91	0.83	0.87	0.91	0.79	0.84
文献[25]	–	–	–	0.90	0.83	0.87	0.91	0.83	0.87
TextMoutain[26]	–	–	–	0.85	0.88	0.86	0.81	0.84	0.82
SegLink[27]	–	–	–	0.88	0.84	0.86	–	–	–
TextSnake[28]	0.87	0.83	0.85	0.73	0.76	0.75	0.86	0.70	0.77
CRAFT[29]	–	–	–	0.84	0.80	0.82	0.83	0.73	0.78
文献[30]	0.93	0.97	0.95	0.84	0.89	0.86	0.78	0.88	0.82
PSENET[31]	–	–	–	0.84	0.88	0.86	0.82	0.88	0.85
PAN[32]	–	–	–	0.86	0.84	0.85	–	–	–
文献[30]	–	–	–	0.84	0.81	0.82	0.84	0.83	0.84

2.3 小结

基于分割方法的难点主要在于相邻文本粘连，改进的方向主要在于将文本像素或字符映射到其他领域进行分类或聚合，或者采用半监督字符级别的训练。

3 其他文本检测算法及性能对比

3.1 其他文本检测算法

基于回归方法不能很好地处理极端尺寸的文本，基于分割的方法存在复杂的后处理步骤，针对此问题，文献[34]结合目标检测和分割方法的思想，提出了采用角点检测和相对位置分割的方法。针对一些算法在对齐文本与文本框时运算量过大和感受野不足的问题，MOST(Multi-Oriented Scene Text detector)在对齐文本特征阶段，使用可变形卷积对粗略检测进行采样，预测采样点偏移，然后将采样点均匀分布在粗略检测框中，产生自适应的感受野，改善文本极端尺寸比例问题，提高了检测速度[35]。为了优化对长文本和任意形状文本的检测，LOMO(Look More than Once)引入角点注意力回归四边形文本框角点坐标的偏移，多次迭代细化，得到文本区域，再根据文本区域、文本中心线和边界偏移重建多方向文本行边界得到文本框，改善了对长文本的检测效果[36]；PCR(Progressive Contour Regression)对区域提议的轮廓点进行均匀采样，根据点的位置信息和语义信息预测点的偏移得到新的轮廓框，对新轮廓框预测角点生成旋转的矩形框，再重新进行轮廓检测，采用轮廓得分进行指导，多次迭代回归轮廓点得到最终的文本轮廓，避免了冗余点或噪声点对文本轮廓的影响[37]；FCE(Fourier Contour Embedding)采用傅里叶变换将区域提议轮廓的采样点变换到傅里叶领域，计算该文本区域像素的傅里

叶特征向量，对采样点根据傅里叶特征向量逆变换（Inverse Fourier Transformation，IFT）到空间域，得到密集的文本轮廓点序列，使得文本框更为紧凑[38]。

3.2　性能对比

在公开数据集 CTW1500 和 Total-text 上，3.1 节所述算法的性能对比见表 5，在公开数据集 ICDAR2013、ICDAR15 和 MSRA-TD500 上，3.1 节所述算法的性能对比见表 6。

表 5　其他文本检测算法在 CTW1500 和 Total-text 上的性能对比

Table 5　Performance comparison of other text detection algorithms on CTW1500 and Total-text

方法	CTW1500			Total-text		
	P	*R*	*F*	*P*	*R*	*F*
LOMO[36]	0.69	0.89	0.78	0.75	0.88	0.81
PCR[37]	0.82	0.87	0.84	0.82	0.88	0.85
FCENet[38]	0.83	0.87	0.85	0.82	0.89	0.85

表 6　其他文本检测算法在 ICDAR2013、ICDAR15 和 MSRA-TD500 上的性能对比

Table 6　Performance comparison of other text detection algorithms on ICDAR2013, ICDAR15 and MSRA-TD500

方法	ICDAR2013			ICDAR2015			MSRA-TD500		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
文献[34]	0.93	0.79	0.85	0.94	0.70	0.80	0.87	0.76	0.81
MOST[35]	–	–	–	0.89	0.87	0.88	0.90	0.82	0.86
LOMO[36]	–	–	–	0.83	0.91	0.87	–	–	–
PCR[37]	–	–	–	–	–	–	0.83	0.90	0.87
FCENet[38]	–	–	–	0.82	0.90	0.86	–	–	–

4　结束语

本文对近几年基于深度学习的文本检测方法进行了分析、归纳和总结，并根据这些信息对文本检测的未来发展趋势进行了展望：

（1）文本检测的速度。一方面实际应用的需求推动着检测速度的研究，一方面网络优化也逐渐得到关注，在检测效果相同时，模型处理文本图像的速度更能体现网络的性能；

（2）多语言文本的检测。随着国家之间交流的增加，多语言文本检测的需求增加，而现存的算法多数关注于单一文本，因此，多语言文本的检测以及分类将逐渐增加；

（3）残缺文本的检测。现存的文本检测算法没有针对存在遮挡、过度曝光或者磨损的文本进行研究，但这些文本在场景文本图像中广泛存在，提升对这类文本的检测可以进一步提高文本检测算法的效果，因此这类文本也逐渐得到了文本检测研究学者的关注。

参考文献

[1] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 56-72.

[2] ZHONG Z, JIN L, HUANG S. Deeptext: A new approach for text proposal generation and text detection in natural images[C]// Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017: 1208-1212.

[3] LIAO M, SHI B, BAI X. Textboxes++: A single-shot oriented scene text detector[J]. IEEE Transactions on Image Processing, 2018, 27(8): 3676-3690.

[4] MA J, SHAO W, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.

[5] LIU Y, HE T, CHEN H, et al. Exploring the capacity of sequential-free box discretization network for omnidirectional scene text detection[J]. arXiv preprint arXiv:1912.09629, 2019.

[6] LIU Y, JIN L, ZHANG S, et al. Curved scene text detection via transverse and longitudinal sequence connection[J]. Pattern Recognition, 2019, 90: 337-345.

[7] ZHU Y, DU J. Sliding line point regression for shape robust scene text detection[C]//Proceedings of the 2018 24th international conference on pattern recognition (ICPR). IEEE, 2018: 3735-3740.

[8] WANG X, JIANG Y, LUO Z, et al. Arbitrary shape scene text detection with adaptive text region representation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 6449-6458.

[9] LIU Z, LIN G, YANG S, et al. Towards robust curve text detection with conditional spatial expansion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7269-7278.

[10] MA J. Rrpn++: Guidance towards more accurate scene text detection[J]. arXiv preprint arXiv:2009.13118, 2020.

[11] ZHONG Z, SUN L, HUO Q. An anchor-free region proposal network for Faster R-CNN-based text detection approaches[J]. International Journal on Document Analysis and Recognition (IJDAR), 2019, 22(3): 315-327.

[12] ZHOU X, YAO C, WEN H, et al. East: An efficient and

accurate scene text detector [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5551-5560.
[13] LIAO M, ZHU Z, SHI B, et al. Rotation-sensitive regression for oriented scene text detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5909-5918.
[14] XU Y, DUAN J, KUANG Z, et al. Geometry normalization networks for accurate scene text detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9137-9146.
[15] WANG F, ZHAO L, LI X, et al. Geometry-aware scene text detection with instance transformation network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1381-1389.
[16] XUE C, LU S, ZHAN F. Accurate scene text detection through border semantics awareness and bootstrapping [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 355-372.
[17] YANG Q, CHENG M, ZHOU W, et al. Inceptext: A new inception-text module with deformable psroi pooling for multi-oriented scene text detection [J]. arXiv preprint arXiv: 1805.01167, 2018.
[18] LIU Y, CHEN H, SHEN C, et al. Abcnet: Real-time scene text spotting with adaptive bezier-curve network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9809-9818.
[19] ZHANG S X, ZHU X, YANG C, et al. Adaptive boundary proposal network for arbitrary shape text detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1305-1314.
[20] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.
[21] HE T, HUANG W, QIAO Y, et al. Accurate text localization in natural image with cascaded convolutional text network [J]. arXiv preprint arXiv: 1603.09423, 2016.
[22] LIU Z, LIN G, YANG S, et al. Learning markov clustering networks for scene text detection [J]. arXiv preprint arXiv: 1805.08365, 2018.
[23] LIU J, LIU X, SHENG J, et al. Pyramid mask text detector [J]. arXiv preprint arXiv: 1903.11800, 2019.
[24] LIAO M, WAN Z, YAO C, et al. Real-time scene text detection with differentiable binarization [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 11474-11481.
[25] TIAN Z, SHU M, LYU P, et al. Learning shape-aware embedding for scene text detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4234-4243.
[26] ZHU Y, DU J. Textmountain: Accurate scene text detection via instance segmentation [J]. Pattern Recognition, 2021, 110: 107336.
[27] SHI B, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2550-2558.
[28] LONG S, RUAN J, ZHANG W, et al. Textsnake: A flexible representation for detecting text of arbitrary shapes [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 20-36.
[29] BAEK Y, LEE B, HAN D, et al. Character region awareness for text detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9365-9374.
[30] ZHANG S X, ZHU X, HOU J B, et al. Deep relational reasoning graph network for arbitrary shape text detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9699-9708.
[31] WANG W, XIE E, LI X, et al. Shape robust text detection with progressive scale expansion network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9336-9345.
[32] WANG W, XIE E, SONG X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8440-8449.
[33] YUE X, KUANG Z, ZHANG Z, et al. Boosting up scene text detectors with guided CNN [J]. arXiv preprint arXiv: 1805.04132, 2018.
[34] LYU P, YAO C, WU W, et al. Multi-oriented scene text detection via corner localization and region segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7553-7563.
[35] HE M, LIAO M, YANG Z, et al. MOST: A multi-oriented scene text detector with localization refinement [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8813-8822.
[36] ZHANG C, LIANG B, HUANG Z, et al. Look more than once: An accurate detector for text of arbitrary shapes [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10552-10561.
[37] DAI P, ZHANG S, ZHANG H, et al. Progressive contour regression for arbitrary-shape scene text detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7393-7402.
[38] ZHU Y, CHEN J, LIANG L, et al. Fourier contour embedding for arbitrary-shaped text detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3123-3131.