自然场景文本检测技术研究

周贤望

（广东技术师范大学，广东 广州 510665）

摘　要　自然场景文本检测作为图像处理领域中的重要分支之一，其在信息检索、智能办公、智慧城市等领域存在着广泛应用。在此背景下，国内外诸多学者针对一些具有挑战性的场景文本检测任务进行了深入研究。本文对文本检测所解决问题的研究趋势进行分类，从水平方向文本检测方法、任意方向文本检测方法、任意形状文本检测方法三个方面阐述了现有场景文本检技术的研究现状及发展前景，以期为相关人员提供参考。

关键词　深度学习；文本检测；自然场景；图像处理

中图分类号：TP317.4　　**文献标识码**：A　　**文章编号**：2097-3365(2023)10-0001-03

1 前言

在现今数字化高度发展的时代，由于移动设备普及和人们对图片信息的需求增加，自然场景图片中的文本信息越来越多。自然场景图片中的文本信息所包含的语义对于人们理解世界和交流思想具有非常重要的作用。然而，自然场景复杂的背景极大地增加了图像中文本的检测难度。因此，场景文本检测与识别成为计算机视觉领域的研究热点。文本检测与识别技术已经广泛应用于拍照翻译软件、商品识别、信息检索、智能办公、智慧城市等领域[1]。因此，对于场景文本检测技术的研究具有重要的意义和价值。文本检测的主要目标是在数字图像或视频中自动识别并定位文本区域。这是文本识别技术的必要步骤，文本检测的准确率将直接影响最终的识别结果，在整个文字识别任务中占据着重要的位置。

文本检测是光学字符识别(Optical Character Recognition，OCR）技术中的一个非常关键的环节，它是OCR技术的第一步，能够从图像中准确地定位出文本区域，为后续的文本识别和分析提供必要的信息。文本检测的准确性直接影响了整个OCR系统的性能，因此，重视和优化文本检测技术非常重要。最初的研究主要聚焦于水平文本检测，随着深度学习的发展，文本检测的研究方向也慢慢从任意方向的文本检测到当前的任意形状文本检测，在场景文本检测领域，诸多学者进行了大量的研究工作并提出了许多文本检测方法，但是仍然存在一些问题，例如小目标文本不容易检测、文本角度多样以及任意形状文本难以准确检测。因此，进一步的研究仍有必要，以提高场景文本检测的准确性和实用性。

2 场景文本检测研究现状

当前，文本检测的方法可以按照所研究的问题分类为不同类型：水平方向文本检测方法、任意方向文本检测方法、任意形状文本检测方法。

2.1 水平方向文本检测方法

在水平文本检测方法中，很多方法采用边缘检测技术来提取自然场景中的文本候选区域，因为这些文本通常具有丰富的边缘和角点信息[2]。其中，最大稳定极值区域[3]是最为经典的文本检测算法之一，该方法的核心思想来自分水岭算法。它利用文本区域稳定的不相连“极值点”来定位和分割字符笔画边缘信息。具体而言，MSER首先将灰度图像进行二值化处理，逐渐提高阈值。这类似于分水岭算法中水平面上升的过程。在这个过程中，一些“山谷”和“较矮的丘陵”将被淹没。如果从空中俯视，则图像将被分为陆地和水域两部分，对应于切分字符和背景的二值图像。每个阈值都将生成一个二值图像，通过对灰度图像进行二值化处理，并逐步提高阈值，可以获得字符和背景的二值图像。据此，可以采用规则或分类器来定位和预测文本候选区域。另外，笔画宽度变换算法是一种针对笔画两侧边缘平行的特点的文本检测方法。该方法通过对高对比度边缘进行逐像素分析，从垂直于边缘的方向上找到与之平行的边缘上的一点，由这两点构成一个笔画横截面并将许多宽度相似的笔画横截面连接起来，能够有效地定位文本位置[4]。最后，CTPN模型将文本区域视为文本组件序列，结合目标检测方法能够克服任意长度文本的检测难点[5]。然而，该方法只能检测水平的文本区域。综上所述，基于边缘的文本检测方法和MSER算法是常用的文本检测方法，笔

画宽度变换算法是一种改进的文本检测方法，CTPN模型可以克服任意长度文本的检测难点，但只能检测水平的文本区域。

2.2 任意方向文本检测方法

近年来，计算机视觉领域中的目标检测得到了迅猛发展，作为目标检测研究内容的特定领域中的文本检测也得到了极大发展，该领域目前已经涌现出一大批任意方向文本检测方法。

TextBoxes网络结构使用不同卷积层的多尺度特征来检测文本，可以有效地识别不同尺寸文本。此外，该网络还可以根据文本区域的纵横比，设置不同的纵横比来检测不同大小和不同方向的文本。然而，TextBoxes网络的低层特征表达能力相对较弱，这会导致它在预测小尺寸文本方面的准确率不是很高。此外，非极大值抑制算法处理候选文本框的结果不理想。余峥[6]通过将TextBoxes网络中不同的特征层相融合并利用邻域候选文本框的位置关系构建了一个新的神经网络，该网络可以提高任意方向文本检测的性能。基于CTPN模型提出的任意方向的文本检测模型SegLink，为了克服CTPN模型无法检测倾斜文本的缺点，通过预测分段八个方向是否有与其他分段连接，使预测分段可以链接生成任意方向的文本框。算法要点如下：先检测文本或者文本行的局部区域，再将这些局部区域连接起来形成一个完整的单词或者文本行。它将文字检测任务分解成两个子任务：检测文字片段和预测片段之间的连接关系。方承志等[7]提出了一种基于残差网络及笔画宽度变换的自然场景文本检测算法。该算法引入了残差结构来加深网络深度，扩大感受野并避免梯度消失问题，从而提升了网络的学习能力。此外，该算法还将预测框和真实文本框之间的中心点距离作为惩罚项加入损失函数，有效区分了不同重叠方式的检测框，进一步提高了检测精度。

2.3 任意形状文本检测方法

任意形状文本检测的发展要从两个重要的数据集说起：CTW-1500和Total-text。自2017年提出这两个数据集之后，大量学术界和工业界关于任意形状文本的研究纷至沓来。李伟冲[8]在现有的任意方向场景文字检测算法TextBoxes的基础上，提出了一种端到端可训练的任意形状文本检测和识别方法，从而实现文字的同时检测和识别。该方法利用带有倾斜角度的文本框，能够实现对不规则形状的文本检测，并通过特征金字塔网络结构和全卷积层来提高检测精度。为了能够适应文本的检测和识别，他在TextBoxes的文本检测分支中添加了对四边形文字框角度的预测，并且通过添加文本识别分支扩展了TextBoxes的网络结构。此外，他引入了特征金字塔网络结构和全卷积层来提高检测精度，使得模型能够有效地检测不同尺寸的文本。通过利用四边形文本框或者包含倾斜角度的文本框实现不规则形状的文本检测。同时，SegLink++模型也是一个很好的解决方案，它引入线段和点两种类型的链接来定义文本区域，并利用深度网络的多级合并细节的能力来处理各种尺度和形状的文本信息，可以检测任意形状的文本。这些模型的引入，为任意形状文本的检测和识别提供了更加有效的工具和技术，从而实现对各种形状的文本进行更好的检测和理解。

Long等[9]提出了一种名为Textsanke的非常灵活的文本实例表征方法。该方法利用一系列连接且重叠的圆盘来表示文本区域，每个圆盘的圆心在文本区域中心线上。这种方法能够实现对线性文本和不规则文本的检测。唐秦[10]将自然场景下的文本检测与识别分为两个任务进行研究，并提出了一种特征聚合与感受野增强的场景文本检测算法，该算法能够获得更加稳定且精确的任意形状文本检测器。这种方法是在PSENet（Progressive Scale Expansion Network）的基础上进行改进的，通过加入特征聚合与感受野增强模块，实现了不同尺度特征信息的提取与融合，并增强了网络低层特征的感受野。白鹤翔等[11]在PSENet模型基础上，加入了三个用以增强边缘特征的网络模块。其中，浅层特征增强模块可有效增强包含更多边缘特征的浅层特征；边缘区域检测分支将普通特征和边缘特征进行区分以对目标的边缘特征进行显式建模；分支特征融合模块可将两种特征在识别过程进行更好的融合，提升了任意形状的文本检测准确率和召回率。这些方法为任意形状的文本检测提供了更加灵活、准确的解决方案，并同时提高了文本检测的准确率和召回率。这些方法可能会在图像处理领域得到广泛的应用。

通常在进行文本检测任务时，采用分割网络来处理预测的概率图并转化为二值图以优化模型训练和计算。然而，传统的二值化过程不可微分，需要进行繁琐的后处理，这会严重影响网络性能和收敛速度。为了解决这个问题，蔡鑫鑫[12]提出了一种基于分割的方法，该方法使用低成本的分割头和高效的后处理，分割头由特征增强和特征融合模块组成，前者提供多层次信息指导分类，后者将深度特征集合成最终特征进行分割。并利用可微分二值化模块（DB）将概率图转换为文本区域，从而提高了文本检测的准确性。Liao

等[13]在DBNet的基础上提出了DBNet++并引入了自适应尺度融合（ASF）模块，该模块可自适应地融合不同尺度的特征以提高尺度的鲁棒性。但两者的不足之处都在于难以检测重叠文本。

3 发展与挑战

目前，深度学习已经成为自然场景文本检测领域的重要研究工具，但该领域的研究方法仍有一定的局限性。下面将介绍该领域存在的一些问题以及未来的主要研究方向。

当前主流的文本检测方法中，都是以矩形或者四边形作为文本区域检测框，这种线性文本区域检测框的设定方式导致了这类方法无法很好地适应任意形状的文本。因此，可以通过提高模型对任意形状文本的检测性能。这种描述方式不仅需要保证检测结果的准确性和鲁棒性，还需要考虑到计算效率的问题。基于这样的要求，近年来出现了各种各样的文本框描述方式，例如基于分割的方法、锚点定位的方法、密集预测的方法等。这些方法在提高文本检测性能方面都有着各自的优缺点，具体选择哪种方法需要根据实际应用场景和需求进行权衡。总之，设计合适的文本区域描述方式是提高文本检测性能至关重要的研究思路。

此外，以目标检测模型为基础改进的文本检测方法往往忽略了文本特征与其他目标物体特征的独特性，导致在一些场景下检测效果并不理想。针对这个问题，可以从文本组件笔画特征进行考虑，可以先设计微文本框去检测文本组件，再利用微分的思想将这些微文本框进行拼接组合成任意形状的文本区域框。

与传统的文本检测方式相比，微文本框的设计可以进一步增加文本检测的灵活性，提高任意形状文本检测的性能和准确率。此外，使用微文本框可以有效地解决文本形状和大小的差异问题，对检测尺寸差异性大的文本场景非常有效。因此，将微文本框引入文本检测技术是一种非常有前途的方法，可以为今后的相关工作提供重要参考，也有很好的应用前景。

除了通过以上思路来提高文本检测的准确率之外，基于直接边框回归的思路也是提高文本检测速度的一个重要思路，基于此思想的方法可以直接预测任意形状的文本区域。这种方法可以有效避免传统方法需要先生成大量的候选框以及复杂的后处理过程，从而提供更快速、更精准的文本检测能力。另外，由于移动设备终端的处理能力有限，构建更轻量化的文本检测网络也将成为未来的重要需求。如何权衡好文本检测模型的检测速度和精度是未来研究中的重要方向，需要不断探索新的文本区域描述方式、模型结构以及优化算法，以提高文本检测的性能，并在实现高效的同时确保准确性，满足各种场景下的需求。

4 结论

自然场景文本检测目前是计算机视觉和模式识别领域的研究热点之一，其方法已逐步从经典方法转向基于深度学习的方法，并且研究对象涵盖了水平方向文本到任意方向和任意形状的文本。本文主要整理了近年来基于深度学习的文本检测方法，并根据文本检测技术要解决的问题对研究者们所提出的思想、方法进行分类，并阐述了其待解决问题和发展趋势。

参考文献：

[1] 李益红,陈袁宇.深度学习场景文本检测方法综述[J].计算机工程与应用,2021,57(06):42-48.
[2] 余波,吴静,周琦宾.一种基于改进Canny算子的边缘检测算法[J].制造业自动化,2022,44(08):24-26,43.
[3] Matas J,Chum O,Urban M, et al.Robust wide-baseline stereo from maximally sTab extremal regions[J].Image Vision Computing,2004,22(10):761.
[4] 周鹏飞.自然场景图像中的文本检测与识别技术研究[D].西安:西安理工大学,2019.
[5] Tian Z, Huang W, He T, et al. Detecting text in natural image with connectionist text proposal network[C]// Computer Vision – ECCV,2016.
[6] 余峥.基于改进TextBoxes的自然场景文本检测算法[D].上海:华东师范大学,2018.
[7] 方承志,倪梦媛,唐亮.基于残差网络及笔画宽度变换的场景文本检测[J].计算机技术与发展,2023,33(01):49-55.
[8] 李伟冲.基于改进TextBoxes++的多方向场景文字识别算法的研究[J].现代计算机(专业版),2018,636(36):67-72.
[9] 王明宇.基于深度学习的自然场景多方向文本检测与识别[J].电子技术与软件工程,2021,218(24):93-96.
[10] 唐秦.自然场景下任意形状文本检测与识别算法研究[D].南昌:南昌航空大学,2021.
[11] 白鹤翔,王浩然.基于边缘特征增强的任意形状文本检测网络[J].自动化学报,2023,49(05):1019-1030.
[12] 蔡鑫鑫,王敏.基于分割的任意形状场景文本检测[J].计算机系统应用,2020,29(12):257-262.
[13] Liao M,Zou Z,Wan Z,et al.Real-time scene text detection with differentiable binarization and adaptive scale fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,45(01):919-931.