

改进 EAST 算法的游戏场景文本检测

池 凯 赵逢禹

(上海理工大学 光电信息与计算机工程学院 上海 200093)

E-mail: ck384353159@163.com

摘 要: 文本识别是模式识别应用的一个重要领域,其中文本检测是文本识别的关键所在。EAST 是目前较为新颖的自然场景文本检测算法,能够高效准确地实现文本定位,但是仍存在长文本检测不佳、感受野不够大的问题。本文针对复杂游戏场景及传统算法定位问题,对 EAST 算法进行了改进优化,在 EAST 网络结构中加入 BLSTM 神经网络,并优化了 shrink_poly 权重策略以提高感受野,提高文本定位效果。实验结果表明,该方案能高效地对游戏场景文本进行检测,极大地提高了检测准确率。

关键词: 游戏场景; 文本检测; EAST; 感受野

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2020)10-2189-05

Game Scene Text Detection Based on Improved EAST Algorithm

CHI Kai ZHAO Feng-yu

(School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: Text recognition is an important field of pattern recognition applications in which text detection is the key to text recognition. Though EAST is a relatively new natural scene text detection algorithm that can efficiently and accurately implement text positioning, there are still problems with poor long text detection and insufficient receptive field. This article focused on the complex game scenarios and traditional algorithm positioning problems and the EAST algorithm is improved by adding a BLSTM neural network to the EAST network structure and the shrink_poly weighting strategy to improve the receptive field and improve the text positioning effect. The experimental results show that this scheme can efficiently detect the text of the game scene and greatly improve the detection accuracy.

Key words: game scene; text detection; EAST; receptive field

1 引言

视频游戏中蕴含着丰富的文本信息,而这些文本信息往往渗透着对游戏制作及全局规划的理解;因此从游戏场景中提取所包含的文本信息,对视频游戏的分析与处理都具有重要的意义。

然而游戏场景中图像变换速率较快,且具有复杂的背景和多样的特殊字符;相较于自然场景,游戏场景的色彩和光影效果较强,部分场景物体具有虚化现象,容易产生视觉深度;与此同时,游戏场景中长文本较多且文本之间具有联系性,对检测算法的感受野要求较高;这一系列的特殊性质给游戏场景文本的检测造成了较大的困难,使得从游戏场景中提取文本信息具有一定挑战性。目前众多学者在场景文本检测方面做了大量的研究并取得了相应的进展。尹芳等人利用迭代算法(Adaboost)将场景中文本定位问题转化为二分类问题^[1],将多个弱分类器组合成一个强分类器,实现了对场景中文本的定位检测;王梦迪等人提出了一种基于多方向边缘检测和自适应特征融合的自然场景文本定位方法,利用三通道八方向的边缘检测通道提取出备选文本域进行特征提取与融合从而定位文本^[2];Xiang Bai 等人提出由粗到精的方法对自然场景中文本进行检测,使用全卷积的神经网络来训练和预测不

同方向的语言和字体,将局部信息和全局信息结合起来^[3];而 Zhou X 等人提出的 EAST^[4] 文本检测算法在全卷积网络的基础上采用了端到端的文本定位方法,使用单个轻量级的神经网络直接对整个图片进行高效快速地预测,消除了复杂步骤;并且 EAST 算法允许端到端的训练和优化,优化了中间组件也省去了复杂步骤,其中高效的算法和精简的网络结构提高了检测精度与效率。

在上述研究中为了达到较高水准的文本检测水平,采用了复杂文本检测通道与文本提取算法,导致文本定位速率较慢、检测效果不佳。经典的 EAST 文本检测算法由于网络结构的限制导致文本感受野较小,存在对长文本以及复杂背景文本检测不佳的问题。因此本文针对游戏场景的文本定位识别问题选取 EAST 作为基础文本检测算法,对 EAST 算法网络结构提出了相应的改进,加强了特征筛选并且在网络结构中加入 BLSTM 层,使得 EAST 文本感受野长增加;同时针对 shrink_poly(收缩文本框策略)提出比例权重更改,从而改善了 EAST 的文本检测效果,提高了算法性能。

2 EAST 文本检测算法

EAST 检测算法采用二阶段(two-stage)文本检测的方法

收稿日期: 2019-12-10 收修改稿日期: 2020-01-03 基金项目: 国家自然科学基金项目(61202376 61003031)资助。作者简介: 池 凯,男,1995 年生,硕士研究生,研究方向为图像识别、机器学习;赵逢禹,男,1963 年生,博士,教授,CCF 会员,研究方向为 web 服务与安全、软件工程和软件质量控制、图像识别、机器学习等。

并基于全卷积网络(FCN)直接对文本内容进行检测,该算法检测通道结合了密集检测框(DenseBox)和分割网络(Unet)中的特性,利用单个神经网络进行文本特征提取,并从整幅图片中预测任意方向的四边形文本行^[4].

2.1 网络结构

EAST 神经网络层采用 U-shape 的思想逐步合并特征图,并且每层网络保持对上层的较少采样;既利用不同级别的特征值,又节约了一定的计算成本,减少了网络层中参数的数量^[4]. EAST 整体网络结构分为 3 个部分,如图 1 所示,分别为特征提取层、特征融合层以及输出层.

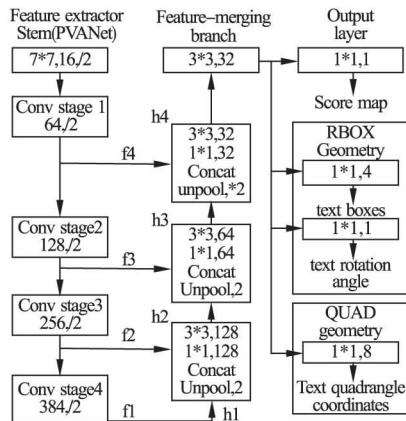


图 1 EAST 网络层的结构

Fig. 1 Structure of the EAST network layer

特征提取层:采用 VGG16 模型中 pooling-2 -pooling-5 阶段作为特征提取阶段的基础网络结构模型^[4];首先输入一张图片,然后经过四个阶段的卷积层处理,并分别从四个阶段(stage1 stage2 stage3 stage4)得到 f4、f3、f2、f1,每个特征图大小分别为输入图像的 1/32、1/16、1/8 和 1/4.

特征融合层:采取逐层合并的方式从后往前做上采样,将第一步抽出的特征图输入到池化层(unpooling)进行扩展,再使用连接函数(concat)对上下层特征图(记为 h_i)进行连接;并通过卷积运算(conv1×1 bottleneck)减少通道数量和计算量,最后将得到的 h4 特征图通过三层卷积核运算(conv3×3)生成最终的特征图.

输出层:输出得分图(score map)和 4 个旋转文本框(RBOX)以及文本框角度信息,表示从像素位置到矩形的顶部、右侧、底部、左侧边界的 4 个距离值;或者输出得分图和四边形(QUAD)中 8 个文本坐标信息,分别表示从矩形的四个顶点到像素位置的坐标偏移.

2.2 损失函数

EAST 总体损失函数(公式(1))由两部分组成:分类损失函数 L_s 和几何损失函数.公式(2)为分类损失函数,其中 \hat{Y} 是预测值, Y^* 是正确预测值, β 是平衡因子;公式(3)为几何损失函数, λ_g 表示两个损失函数之间的重要性;公式(4)和公式(5)为旋转回归损失函数, L_{AABB} 部分使用 IoU (Intersection over Union) 损失函数使计算更加平均化, R_a 表示预测的 AABB 几何形状, R_b 是其相应的标注框; L_θ 损失函数主要计算边框与水平方向的夹角, θ_a 是对旋转角度的预测, θ_b 表示标注值.最后,总体几何损失是 L_{AABB} 损失与 L_θ 损失的

加权和.

$$L = L_s + \lambda_g L_g \quad (1)$$

$$L_s = -\beta Y^* \log \hat{Y} - (1 - \beta) (1 - Y^*) \log (1 - \hat{Y}) \quad (2)$$

$$L_g = L_{AABB} + \lambda_\theta L_\theta \quad (3)$$

$$L_{AABB} = -\log \text{IoU}(R_a, R_b) = -\log \frac{|R_a \cap R_b|}{|R_a \cup R_b|} \quad (4)$$

$$L_\theta(\theta_a, \theta_b) = 1 - \cos(\theta_a - \theta_b) \quad (5)$$

在 L_s 与 L_g 的计算中,以平衡交叉熵函数(cross_entropy)作为 score map 损失函数来解决样本不平衡分布的问题,既避免了复杂参数也对目标数据进行了分布处理,提高了网络性能.通过交叉熵函数尽可能地使模型输出的分布与训练样本的分布一致;交叉熵(cross_entropy)主要呈现出的是实际输出(概率)与期望输出(概率)的差距,两个概率分布越接近则交叉熵的值就越小.假设 $H(p, q)$ 为交叉熵, p 为期望输出概率分布,实际模型输出概率分布为 q ,则:

$$H(p, q) = -\sum_x p(x) \log q(x) \quad (6)$$

3 改进 EAST 检测算法

3.1 网络结构优化

EAST 原始网络模型为了扩大输出单元的感受野在池化层阶段采用了大量计算复杂度较小的下采样,然而较多池化层提取使得特征样本尺寸减少,使得上采样阶段回复分辨率的难度增加,导致部分输出特征映射感受野较小,不能编码更多尺度信息;因而在实际检测中段落边缘检测效果略差的问题.因此本文针对 EAST 段落边界检测不佳的问题,简化卷积层结构并扩大了卷积核通道数从而加强了对特征值的提取筛选.在特征融合的基础上加强了对特征样本的序列化处理.在 EAST 原先网络模型中加入 BLSTM^[5] (Bi-directional Long Short-Term Memory) 神经网络层,改进过的网络结构如图 2 所示.

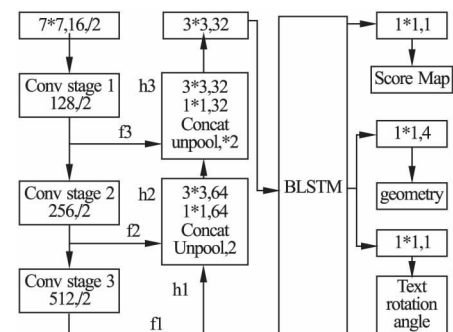


图 2 改进 EAST 网络层的结构

Fig. 2 Improved structure of the EAST network layer

EAST 在基于 VGG16 的基础上提取了四个级别的特征图,分别为输入图像的 1/32、1/16、1/8、1/4;考虑到原先网络感受野较小的问题以及游戏场景的特殊性(无旋转字体以及较小文本),在本文中修改为提取三个阶段的特征图,在保持参数数量的基础上增加卷积核的通道数从而大量有效地筛选的更多特征,使得特征层输出更大尺度的有效感受野.

本文加入的 BLSTM 是一种特殊的循环长短期记忆神经网络,由双向 LSTM^[6] 神经网络组成;BLSTM 具有学习长期

依赖的能力, 能将每个特征的前后序列呈现为两个单独的隐藏状态, 以分别捕获序列过去和未来的信息, 然后再将两个隐藏的特征序列连接起来形成一个新的特征样本进行最终输出^[7] (如图 3 所示). 本文将经过特征融合所得到的序列样本 (feature map) 放入 BLSTM 中进行循环特征连接, 使得到新的序列样本更加合理均匀且具有联系性; 然后再经过 1×1 的卷积输出 ROI (region of interest-proposal) 所对应的 score map 和相应的几何信息.

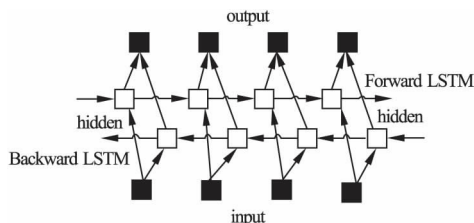


图 3 BLSTM 网络传输结构

Fig. 3 BLSTM network transmission structure

3.2 shrink_poly (收缩矩阵) 改进

EAST 在制作预测四边形文本框 (polygon) 时采用了收缩矩阵 (shrink_poly) 的策略, 而收缩矩阵在一定程度上能够减少人工标注带来的误差.

$$r_i = \min(D(p_i, p((i \bmod 4) + 1)), D(p_i, p((i + 2) \bmod 4) + 1))$$

其中 p_i 表示 4 个顶点, r_i 表示矩形 4 条边的长度, $i = 0, 1, 2, 3$; 并直接计算出矩形相邻两点之间的 L2 距离, 取其中较小的一个. 实际计算矩形边长时, 通过求取向向量范数计算相邻两点距离 (L2 norm).

考虑到使得收缩更加优化, 在计算过程中尝试增加对于特殊边长的权重收缩, 并赋予相应的数值权重. 其中对于较短的矩形边可以取正常固定值的距离收缩 (shrink); 而对于较长的矩形边则要采取减少收缩的策略. 当矩形长边和短边在相同倍数收缩时会导致训练完成后的四边形矩形框整体收缩过度, 出现长边边界预测不准的情况, 导致边界效果检测较差. 这里对于矩形长边尝试采取 0.1 倍固定值的收缩, 从而改善边界的检测效果, 如算法 1 所示.

算法 1. shrink_poly 策略改进

输入: Four vertices p_1, p_2, p_3, p_4

输出: Shrunk vector

```

1. function shrink_poly( p1, p2, p3, p4)
2. if np. l. n( ( p1 - p2) + ( p2 - p3) ) > np. l. n( ( p1 - p4) + ( p2 - p3) )
3.     np. arctan2( ( p1, p2) ), ( p3, p4) )
4.     [ p1, p2 ] ← 0.1 shrink
5.     [ p3, p4 ] ← 0.1 shrink
6.     np. arctan2( ( p1, p4) ), ( p2, p3) )
7.     [ p1, p4 ] ← 0.3 shrink
8.     [ p2, p3 ] ← 0.3 shrink
9.     return shrink_poly
10. Else np. arctan2( ( p1, p2) ), ( p3, p4) )
11.     [ p1, p4 ] ← 0.1 shrink
12.     [ p2, p3 ] ← 0.1 shrink
13.     np. arctan2( ( p1, p4) ), ( p2, p3) )
14.     [ p1, p2 ] ← 0.3 shrink
15.     [ p3, p4 ] ← 0.3 shrink

```

```

16.     return shrink_poly
17. End if
18. end function

```

4 实验及分析

为了更好地验证本次算法的性能, 并对实验数据有较快地分析, 本次实验平台选择为 Matlab 2012b, PyCharm, 计算机配置为 windows 与 Linux 双系统, 处理器 Intel core i7, 主频 3.10GHz, 内存 16GB, 显卡为 GTX1070.

4.1 实验步骤

本次实验主要分为两个步骤: 构建图像数据集、文本定位训练与验证, 流程如图 4 所示.

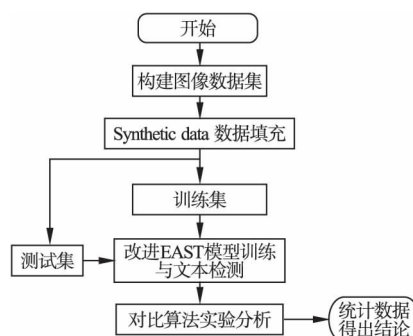


图 4 模型训练与验证实验流程图

Fig. 4 Model training and verification experiment flowchart

4.2 构建图像数据集

本次实验首先从游戏视频中通过人工逐帧截取的方式获取大量场景图像^[8], 由于游戏场景中的文本信息较少, 不利于制作游戏场景训练集, 因此需要对游戏场景进行人工填充数据, 而人工合成数据可以准确知道文本的标签 (label) 信息, 省去了人工标注数据的时间; 采用 VGG 提出的 Synthetic Data^[9] 技术将数据自然排列进入游戏场景空白处, 从而快速生成大量符合有利于实验要求的标签化数据, 制作出游戏场景专属的训练集.

游戏场景 (pubg) 中的大部分文本是集中在区域内出现的, 因此要把背景图像分割成连续的区域, 再将要填入的游戏文本字体嵌入其中相应区域. 当放入文本的位置、方向确定后, 再对文本赋予一些相应的颜色. Synthetic Data 会把文字和背景图像的像素分为两个集合, 其中一个集合是文字的颜色, 另外一个则是背景的颜色, 因此数据填充会根据游戏场景的设定, 设置相匹配的颜色. 由于游戏场景中的文字浮于图像的表面, 所以无需进行太多的深度测试和筛选, 因此在这里将深度筛选部分设为固定值, 将文字更符合地排入游戏场景中, 从而满足游戏场景的特殊性质. 数据填充流程如表 1 所示.

经过人工合成数据后, 游戏场景图像中的文本得到了有效地填充, 其中效果如图 5 所示.

4.3 文本定位训练与验证

本次实验使用经过 resnet-50 网络模型^[10] 训练后的模型作为实验预训练模型, 将游戏场景图像经过降噪、平移等处理后, 将每个 batch size 设置为 50 进行训练; 其中模型的初始学习率设为 0.001, 衰减率为 1000 倍, 并经过对比训练筛选取得

较优模型.

表 1 数据填充流程

Table 1 Data population process

人工合成数据: Synthetic Data for text
输入: 30000 张游戏场景图像,eko 文本字体, 新闻报纸数据集.
输出: 经过人工合成后的场景图像.
1. 首先确定每个场景生成多少张合成图片及每张合成图片的合成文本区域,对于每次文本合成,随机选取 1 个文本域进行填充,再确定填入字体和字体的弯曲程度;
2. 根据文本域的平均宽高,设置字体的大小,从新闻报纸数据集中(Newsgroup 20 dataset) 选择要放入的文本内容;
3. 根据要填写的文本创建矩形文本框,并将文本嵌入到文本框内(先中间,后两边),并将文本框进行截取;
4. 将文本和文本域进行碰撞检测,如果存在碰撞,则此区域不合适;将会继续寻找合适的配对;
5. 通过矩阵变换将得到的单位矩阵变换成符合场景的角度的文本区域,然后再使用 opencv ^[11] 进行高斯模糊处理;
6. 最后为文本选择一种颜色(根据设定),将涂色好的文本区域放到原图片中得到输出图片.



图 5 场景数据填充图

Fig.5 Scene data fill map

实验中改进 EAST 算法网络模型是基于 VGG-16 框架进行构建的,网络模型主要由输入层,特征提取层,融合层, BLSTM 层以及输出层组成.配置如表 2 所示, f 表示卷积核的

表 2 网络参数表

Table 2 Network parameter table

网络层类型	参数配置
输入层	512* 512 场景图片
特征提取层	
卷积层	$f: 128 \quad k: 3*3 \quad s: 1 \quad p: 0$
卷积层	$f: 256 \quad k: 3*3 \quad s: 1 \quad p: 0$
卷积层	$f: 512 \quad k: 3*3 \quad s: 1 \quad p: 0$
特征融合层	
池化层	窗口 $2*2 \quad s: 2$
卷积层	$f: 32 \quad k: 3*3 \quad f: 32 \quad k: 1*1$
卷积层	$f: 64 \quad k: 3*3 \quad f: 64 \quad k: 1*1$
BLSTM 层	#隐藏单元: 256
BLSTM 层	#隐藏单元: 256
输出层	Score map

数量 k 表示卷积核尺寸, s 表示步长, p 表示填充尺寸,其中池化层阶段采用了 $2*2$ 的池化窗口来扩增感受野,而 BLSTM 层中加入了 256 个隐藏单元.

为了进一步验证改进 EAST 算法的性能,本次的实验利用 nvidia gtx1080ti 显卡在游戏场景的专属训练集上对原 EAST 算法与改进 EAST 算法的文本定位效果进行了检测和比较.如表 3 所示,表中数据显示本文改进算法平均召回率为 77.89%,准确率为 85.17%,平均 F-score 为 82.76%;从表 2

中可以看出,本文算法的召回率要优于原 EAST 算法,比 EAST 检测算法准确率高;而改进 EAST 算法的 F-score 也有所提升,在两个算法执行速度上相差无几的情况下,改进 EAST 算法在游戏场景上的文本检测性能要优于原 EAST 算

表 3 改进 EAST 性能表现

Table 3 Improved EAST performance

模型	Precision(%)	Recall(%)	F-score(%)	FPS
EAST	81.13	77.48	79.98	13.2
Ours	85.17	77.89	82.76	13.4

法.其中图 6 为 EAST 算法和改进 EAST 算法在实际游戏场景中的检测效果对比,从图中对比可以看出, EAST 算法整体检测效果较好,但是对于长文本的检测有略微缺陷,对长文本检测不够完整;而改进 EAST 算法对于长文本的检测效果较好,能够较为完整地检测长文本.



图 6 原 EAST 算法与改进 EAST 检测效果对比图

Fig.6 Comparison of original EAST algorithm and improved EAST detection effect

在上述实验的基础上,本文选取 CTPN 等一些优秀的文本定位算法与改进 EAST 算法在游戏场景数据集上进行比较验证. CTPN 算法结合 RNN 与 CNN 的特点直接利用卷积在特征图上生成文本区域并进行快速检测^[12]; PixelLink 算法利用像素分割出文本实例并提取文本预测框,处理文字特征上效果明显,具有较好的定位精度^[13]; MSER 算法具有良好的抗噪性和实用性被广泛运用于自然场景的文本检测^[14,15]; EAST + PVANET 是原论文作者对 EAST 的强化改进,并在 ESAT 算法中引入了改进的 NMS^[4]; SegLink 算法能够快速地对多方向任意长度文本进行检测^[7]; Wordsup 算法能够在文本行和单词标注的数据集上训练出检测模型^[5];而 TextBoxes 算法具有较好的文本感受野,检测速度较快^[16].其中在游戏场景数据集上对这几种算法的比较如表 4 所示.

由表 4 可以看出,改进 EAST 算法文本检测的准确率和检测速度要优于其他的对比算法,而在召回率上改进 EAST 算法略低于处理文字特征较强的 PixelLink 算法和 EAST + PVA 算法,同时在 F-score 上改进 EAST 算法与 pixelLink 算法领先于其他算法.可以很直观地感受到改进 EAST 算法在游戏场景的各项数据中性能表现都是较为出色的,对游戏场

景文本检测较快,综合性能与其他算法相比具有一定的优势.这一结果证明了改进 EAST 算法能够有效地提取游戏场中的景特征值且具有出色的文本感受野,对背景较为复杂的游戏场景图像检测效果相较于其他算法更好^[17-19].

表 4 改进 EAST 与其他方法在游戏场景性能对比

Table 4 Improved EAST and other methods in game scene performance comparison

模型	Precision(%)	Recall(%)	F-score(%)	FPS
CTPN	74.8	69.2	70.92	11.4
PixelLink	82.93	82.38	82.98	7.6
EAST + PN	83.17	78.13	80.16	12.7
MSER	69.95	65.86	68.92	5.2
SegLink	80.50	70.46	81.92	8.7
WordSup	80.72	76.32	80.12	9.4
TextBoxes	79.45	74.30	78.58	6.7
Ours	85.17	77.89	82.76	13.4

为了进一步验证改进 EAST 算法的性能,本文选取在上述游戏场景数据集表现较好的算法在 ICDAR 2015 数据集上进行了对比实验.ICDAR 2015 数据集包含 1000 张训练图和 500 张测试图,其中文本具有旋转性质和噪声干扰,给检测带来了和很大难度.实验中随机选取了 ICDAR 2015 的 200 张图像,其中结果如表 5 所示.

表 5 改进 EAST 与其他方法在 ICDAR 2015 上性能对比

Table 5 Performance comparison between improved EAST and other methods at ICDAR 2015

模型	Precision(%)	Recall(%)	F-score(%)	FPS
CTPN	73.2	55.30	60.92	7.1
PixelLink	80.93	82.78	81.98	3.2
Seglink	74.17	76.80	75.36	4.6
WordSup	79.95	77.46	78.92	5.2
Ours	81.17	81.89	81.75	6.4

从表 5 中可以看出,改进 EAST 算法在 ICDAR 2015 数据上的表现也是较为出色的.对比在游戏场景数据集上表现,改进 EAST 算法在 ICDAR 2015 数据上检测准确率有所下降,但是仍是最为优秀的;在算法执行速度上领先于检测精准的 PixelLink 算法.综合而言,改进 EAST 算法展现出了良好的检测性能,准确率较高.

5 结 语

本文在 EAST 算法的基础上提出了改进,加入了 BLSTM 网络,优化了 shrink_poly 权重策略,提升了 EAST 算法的性能.在游戏场景和 ICDAR2015 数据集上做了大量的对比实验,从而验证了方法的有效性.该方案还仍有提高的可行性,在实验流程上还能对图像预处理的部分再进一步提高完善,加强对图像二值化以及降噪方面的细致处理.同时还要加强数据样本收集,增加数据训练,以达到更好的训练效果.针对以上两个问题,在后续实验研究中将会加大数据预处理的力度,努力搜集样本数据,从而再次提高检测准确率.

References:

[1] Yin Fang ,Zheng Liang ,Chen Tian-tian. Chinese text localization based on Adaboost algorithm [J]. Computer Engineering and Ap-

plications 2017 ,53(4):200-204.

- [2] Wang Meng-di ,Zhang You-mei ,Chang Fa-liang. Natural scene text localization based on edge detection and feature fusion [J]. Computer Science 2017 ,44(9):300-303.
- [3] Zhang Z ,Zhang C ,Shen W ,et al. Multi-oriented text detection with fully convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR) , 2016: 2642-2651.
- [4] Zhou Xin-yu ,Yao Cong ,Wen He ,et al. EAST: an efficient and accurate scene text detector [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2017: 2642-2651.
- [5] Hu H ,Zhang C ,Luo Y ,et al. WordSup: exploiting word annotations for character based text detection [C]//2017 IEEE International Conference on Computer Vision(ICCV) 2017: 4950-4959.
- [6] Shi X ,Chen Z ,Wang H ,et al. Convolutional LSTM network a machine learning approach for precipitation nowcasting [C]//Advances in Neural Information Processing Systems(NIPS) 2015: 802-810.
- [7] Shi B ,Bai X ,Belongie S. Detecting oriented text in natural images by linking segments [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2017: 3482-3490.
- [8] Jiang Meng-di ,Cheng Jiang-hua ,Chen Ming-hui ,et al. A review of video and image text extraction methods [J]. Computer Science , 2017 44(B11): 8-18.
- [9] Gupta A ,Vedaldi A ,Zisserman A. Synthetic data for text localisation in natural images [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2016: 2315-2324.
- [10] He K ,Zhang X ,Ren S ,et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2016: 770-778.
- [11] Yang P ,Rong G ,Peng G ,et al. Research on lip detection based on OpenCV [C]//2011 International Conference on Transportation Mechanical and Electrical Engineering(TMEE) 2011: 1465-1468.
- [12] Zhi T ,Huang W ,Tong H ,et al. Detecting text in natural image with connectionist text proposal network [C]//European Conference on Computer Vision(ECCV) , Springer ,Cham 2016: 56-72.
- [13] Deng D ,Liu H ,Li X ,et al. PixelLink: detecting scene text via instance segmentation [C]//2018 the Association for the Advance of Artificial Intelligence(AAAI) 2018: 6773-6780.
- [14] He Si-nan ,Guo Yong-jin ,Zhang Li. Multi-directional natural scene text detection [J]. Application Research of Computers ,2018 ,35(7): 2193-2196.
- [15] Tian Qing-yue ,Gao Zhi-rong ,Xiong Cheng-yi ,et al. MSER natural scene text detection with joint edge enhancement [J]. Journal of Chinese Computer Systems 2017 ,38(11): 2604-2609.
- [16] Liao M ,Shi B ,Bai X ,et al. TextBoxes: a fast text detector with a single deep neuralnetwork [C]//Association for the Advance of Artificial Intelligence(AAAI) 2017: 4161-4167.
- [17] Frinken V ,Uchida S. Deep BLSTM neural networks for unconstrained continuous handwritten text recognition [C]//2015 13th International Conference on Document Analysis and Recognition (ICDAR) 2015: 911-915.
- [18] Arbelaez P ,Maire M ,Fowlkes C ,et al. Contour detection and hierarchical image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI) 2011 ,33(5): 898-916.
- [19] Simonyan K ,Zisserman A. Very deep convolutional networks for large-scale image recognition [C]//International Conference on Learning Representations(ICLR) 2014.

附中文参考文献:

- [1] 尹 芳 ,郑 亮 ,陈田田. 基于 Adaboost 算法的场景中文文本定位 [J]. 计算机工程与应用 2017 ,53(4): 200-204.
- [2] 王梦迪 ,张友梅 ,常发亮. 基于边缘检测和特征融合的自然场景文本定位 [J]. 计算机科学 2017 ,44(9): 300-303.
- [8] 蒋梦迪 ,程江华 ,陈明辉 ,等. 视频和图像文本提取方法综述 [J]. 计算机科学 2017 ,44(B11): 8-18.
- [14] 何思楠 ,郭永金 ,张 利. 多方向自然场景文本检测 [J]. 计算机应用研究 2018 ,35(7): 2193-2196.
- [15] 田清越 ,高志荣 ,熊承义 ,等. 联合边缘增强的 MSER 自然场景文本检测 [J]. 小型微型计算机系 2017 ,38(11): 2604-2609.