

安防视频时间戳同步检测方法研究

彭祥云^{1,2}, 陈黎^{1,2}

(1. 湖北省智能信息处理与实时工业系统重点实验室, 湖北 武汉 430081;

2. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430081)

摘要:针对安防视频标注时间与标准时间存在偏差的问题,提出基于深度学习的时间戳同步检测方法。首先,针对目前的场景文本检测算法在检测视频时间戳区域时会出现时间戳区域检测不完全或检测框过大的现象,提出CBAP算法。该算法首先通过卷积神经网络得到字符级的检测结果与像素级的分割结果,再从字符级的检测结果和像素级的分割结果中提炼出最终检测结果,从而得到更为精确的文本实例包围框,能够有效地应对复杂多样的时间戳区域的检测任务。其次,提出基于容差匹配的时间戳同步判定方法用于解决视频图像在网络传输、编解码等各环节的延时导致时间同步判断出错的问题。最终,实验结果证明了提出的时间戳同步判定方法的合理性与有效性。

关键词:安防视频;深度学习;卷积神经网络;时间戳;文本检测;容差匹配

中图分类号:TP391.43

文献标识码:A

文章编号:1673-629X(2021)11-0195-07

doi:10.3969/j.issn.1673-629X.2021.11.032

Research on Synchronous Detection Method of Security Video Time Stamp

PENG Xiang-yun^{1,2}, CHEN Li^{1,2}

(1. Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, China;

2. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: Aiming at the problem that the marking time of security video deviates from the standard time, a time stamp synchronous detection method based on deep learning is proposed. First of all, for the current scene text detection algorithm, when detecting the time stamp region of the video, the detection of the time stamp area is not complete or the detection frame is too large. Therefore, CBAP algorithm is proposed. Character level detection results and pixel level segmentation results are obtained by convolution neural network firstly, and then the final detection results are extracted from the character level detection results and pixel level segmentation results, so as to obtain more accurate text instance envelope, which can effectively deal with the detection tasks of complex and diverse timestamp areas. Secondly, a time stamp synchronization decision method based on tolerance matching is proposed to solve the problem of time synchronization judgment error caused by the delay of image transmission, encoding and decoding. Finally, the experiment proves that the proposed method is reasonable and effective.

Key words: security video; deep learning; convolution neural network; time-stamp; text detection; tolerance matching

0 引言

随着平安城市、智能交通、智慧城市等重点行业基础设施和公共服务的持续建设,对庞大的监控视频进行统一的管理也显得尤为重要。2019年以来,公安部通过“公安视频图像信息联网应用运维管理平台”,陆续启动了对省级公安及下级业务部门的视频运行情况考核。发现部分摄像机存在时间信息与标准时间存在

偏差的现象。而时间信息作为安防视频图像必备的内容要素,一旦出现问题可能会使整个监控系统存在极大的风险。如时间信息标注错误的监控图像甚至可能因为不具备法律效力而无法作为证据,不仅增加了公安部门的破案难度,还会引发人们对监控系统存在意义与价值的质疑。

随着深度学习技术的快速发展,人工智能技术在

收稿日期:2020-11-18

修回日期:2021-03-18

基金项目:国家自然科学基金面上项目(61773297);智能信息处理与实时工业系统湖北省重点实验室开放基金资助项目(2016znss01A)

作者简介:彭祥云(1997-),男,硕士研究生,研究方向为文字检测、目标检测;陈黎,教授,博导,博士,研究方向为计算机视觉、图像处理。

安防领域的应用不断拓宽。因此,文中提出一种基于深度学习的视频图像时间戳同步检测方法,能够及时发现视频监控系统中与标准时间偏差程度超出容差阈值的监控摄像机,以便辅助人工进行校准工作。其中,时间戳同步检测的难点如下所述:

首先,时间戳区域检测作为后续工作的基础,时间戳区域检测结果的好坏将直接影响后续工作的好坏。因此,图像时间戳区域的检测要求输出的时间戳区域的边界框要完整且准确。既不能出现检测不完全的现象,也不能包含过多的背景信息。然而由于监控设备的多样,不同地区在标注时间戳信息时没有完全遵循《GA/T 751-2008 视频图像文字标注规范》标准。导致监控视频中的时间戳标注出现多样性与不规范的现象,如图 1 所示。此外,在自然场景下还会有不均匀的光照、相似的纹理特征等因素的影响。这些都给时间戳的检测任务带来了巨大的挑战。



图 1 时间戳标注多样性

其次,时间戳同步的判断标准或流程。如果简单采用某一时段设备时间与标准时间是否相同作为判断标准,可能存在如下问题:第一,获取数据是通过采集某一路监控点位监控设备的实时图像,并记录该点位图像采集发起时刻作为标准时间,而视频流采集图像帧在网络传输、设备响应等各环节存在不确定延时现象,导致一些时间戳同步的设备误报为不同步;第二,时间标注字符在背景透明的自然场景图像中,极易受到光线环境变化及关联因素的干扰,可能在不同时间段的识别结果不同,使得同一监控设备同时出现同步与不同步的矛盾情况。

针对目前时间戳同步检测的问题与难点,文中方法的主要贡献如下:

(1) 提出的 CBAP 方法融合了检测与分割的优点并在 fpn 网络中加入文字特征增强模块,提升模型对透明文字和文字与背景纹理特征相似等‘弱像素’文字区域的检测效果。

(2) 在模型的推理阶段提出候选框误差近邻传播连接的后处理方法,通过对检测分支和分割分支的输出做进一步优化,得到更加完整的检测结果。

(3) 提出一种基于图像的时间戳同步判定方法,能有效地根据图像判断出监控设备的时间与标准时间

是否同步。

1 国内外研究现状

针对时间戳的检测与识别,鲍复民等人^[1]提出一种基于模板匹配的方法,但对于字体大小变化十分敏感。姚文杰等人^[2]提出的方法需要对每个字符进行标注,需要耗费大量人力去标注训练样本。都难以应用到安防视频时间戳的检测与识别上。从大方向看,时间戳检测与标记牌文字检测^[3]、发票文字检测^[4]和车牌检测^[5]同属于场景文字检测范畴。而国内外最新基于深度学习的场景文字检测方法主要分为基于目标检测的方法和基于图像分割的方法两大类。

基于目标检测的方法一般是通过对 R-CNN^[6]、SSD^[7]或 yolo^[8]等算法进行改进,使之更符合场景文字检测。DeepText^[9]通过对 Faster R-CNN^[10]算法进行改进,在 RPN 模块后面加入 Inception^[11]模块并且加入歧义文字这一类别,因此能更好地获取文本区域。但对小区域的文本的检测效果不佳。杨宏志等人^[12]通过在 Faster R-CNN 算法中引入 LSTM,可以更好地对长文本进行定位。龙言等人^[13]采用分支结构的特征提取方法,使得算法能够感知到更加深层次的文本特征,从而提升算法检测精度。Liao 等人^[14]提出的基于 SSD 的 TextBoxes++ 算法不需要二次回归并且可以检测任意方向的文本。林泓等人^[15]提出基于多路精细化特征融合的方法可以针对难样本进行检测。

Zhang 等人^[16]首次将全卷积神经网络引入到场景文字检测,将文字视为一个需要分割的目标。杨剑锋等人^[17]通过融合多尺度文字特征图提出可以检测任意方向的文字检测算法。Yao 等人^[18]提出的整体嵌套边缘检测算法(holistically-nested edge detection, HED)通过将文字区域概率、字符概率和相邻字符连接概率融和到一个网络进行学习。基于实例分割的 PixelLink^[19]算法从实例分割结果中获取文本位置信息。该网络通过预测每个像素文本、非文本分类和每个像素的 8 个方向是否连接这 2 个任务,然后利用 Opencv 可以获得不同大小的文本连通域,接着使用噪声滤除操作,最后通过并查集合并出最终的文本框。PSENet^[20]主要通过预测多个分割结果 S_1, S_2, \dots, S_n , 其中 S_1 面积最小, S_n 面积最大,再通过扩张算法逐步扩张成正常文本大小。

2 文中方法

文中提出一种基于深度学习的视频图像时间戳同步检测方法,首先通过深度学习方法得到时间戳区域与时间信息,再通过设计同步匹配方法得出设备时间戳是否与标准时间同步的结论。

2.1 时间戳区域检测方法

由于安防视频图像时间戳标注的多样性,使得目前的文字检测算法在应用到安防视频的时间戳区域检测时往往无法取得令人满意的检测结果。主要存在以下问题:第一,时间戳区域检测不完全。第二,检测框过大,包含太多干扰信息。笔者认为基于目标检测的方法对物体级别的 Anchor 进行回归与分类,获得的信息多同时也会包含部分背景信息使得检测结果包含过

多背景。而基于分割的方法针对像素进行分类,在边缘处的分割效果往往不太理想,会出现检测结果缺少字符的现象。因此文中提出的 CBAP 算法,同时融合了检测与分割的优点,能更好地应对复杂的自然场景。CBAP 的网络结构如图 2 所示,主要分为两个分支:检测分支 (detection branch) 和分割分支 (segmentation branch)。

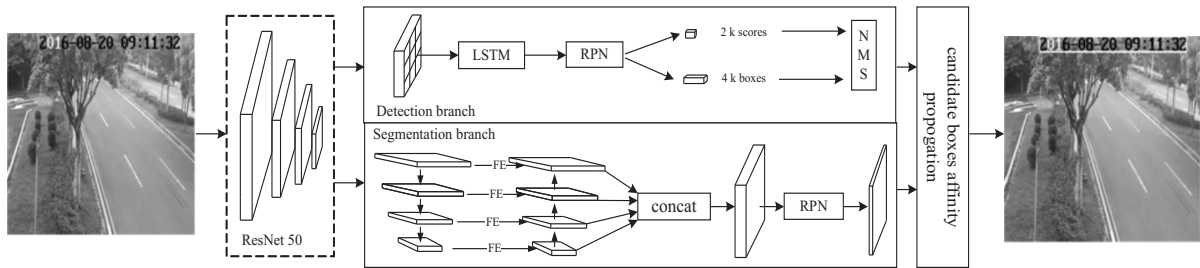


图 2 CBAP 网络结构

2.1.1 网络结构

检测分支的设计过程中主要借鉴了 CTPN^[21] 的网络结构。首先,将完整的时间戳区域划分为包含部分时间戳区域小块,将检测任务转变为对小时间块的检测,以此来解决长文字行的检测问题。其次,通过引入 LSTM 网络,获取时间戳的上下文联系,根据前后的时间块序列来提取相互之间的关系特征,能有效地提升检测的精度。

分割分支中的网络构建主要参考了 FPN (feature pyramid network) 的网络结构。FPN 最早由 Lin^[22] 等人提出用于解决目标检测中的多尺度问题。主要通过横向连接将低层的低级细粒度特征与高层的高级语义信息融合,在不增加计算量的情况下大幅提升了对小物体的检测性能。在近几年的场景文字检测的研究中,研究人员同样采用 FPN 结构在公共数据集上获得了不错的检测效果。但文中研究对象与公共数据集之间的差异较大,特别是对透明字体与文字背景混叠等困难样本的时间戳区域的检测,这些样本的文字与背景难以区分,目前的场景文字检测算法难以准确检测出时间戳区域。因此,文中在 FPN 的横向连接中加入了一组文字特征增强模块,提升模型对文字像素的感知。其中,文字特征增强模块主要参考了 Dai 等人^[23] 提出的可变形卷积的思想。Dai 等人认为固定几何结构的卷积对未知形变的建模存在固有的缺陷,这种缺陷来源于卷积单元对输入特征图的固定位置进行采样,为了解决这个缺陷,由此提出了可变形卷积,可以实现在当前位置附近随意采样。而文中以整个时间戳区域为研究对象,形状固定为长方形,水平方向的特征信息丰富。因此可以将卷积核固定为长方形,在获取文字特征信息的同时省去可变形卷积自适应探索的过

程。基于此,文中设计的文字特征增强模块具体结构如图 3 所示。首先,为了获得足够的感受野来应对不同尺度的时间戳区域的检测任务,引入扩张率为 2、卷积核为 3×3 的空洞卷积获得大感受野的特征图;然后在大感受野特征图上使用 1×5 卷积提取更多的横向纹理特征,得到更多关于文字特征的特征图;最后,将高层的特征图上采样后与之融合得到对文字特征加强后的金字塔特征图。

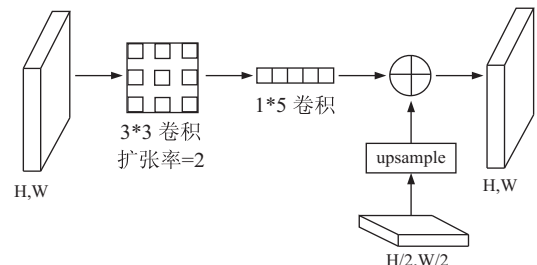


图 3 文字特征增强模块 (FE)

2.1.2 候选框误差近邻传播连接方法

在模型推理阶段,文中通过设计候选框误差近邻信息传播连接方法从网络检测分支与分割分支的结果中提炼出最终的检测结果。具体步骤为:

(1)划分集合:首先,对概率图进行二值化后得到文本与非文本二值图并通过获取连通域的最小外接矩形来得到分割分支的检测结果。然后,对检测分支的所有小检测框,通过计算其与每个分割分支的检测结果的水平距离与垂直重叠度确定小检测框所属于的文本实例。具体判断方法为按照垂直重叠度从大到小依次选取分割结果直到满足水平距离在 50 个像素内的条件。如果所有的分割结果都不满足条件,该小检测框直接舍弃。最终,属于同一个分割结果小检测框集合用来预测相应的文本行。其中,方框的水平距离计

算方法为两个方框的中心点的横坐标的差的绝对值减去两个方框的宽的一半;方框的垂直重合度的计算方法为两个方框在 y 轴投影的交集与两个方框在 y 轴投影并集的比值。

(2) 确定文本实例中心: 对于(1)中得到的每一个小检测框集合, 将其中与对应的分割分支的检测结果有交集的小检测框连接成基本的检测框 D , D 即为文本实例中心。将与对应的分割分支的检测结果没有交集的小检测框记为集合 C 。

(3) 计算吸引度 s 与可信度 a : 文本实例中心 D 对每个集合 C 中小检测框会产生一个吸引度 s , 用来表示文本实例中心 D 认为小检测框包含文本实例的可能性。集合 C 中的每个小检测框会拥有一个可信度 a 用来表示自身认为包含文本实例的可能性。其中, 可信度 a 的值为小检测框在分割分支输出的概率图上的平均值。吸引度 s 的计算公式为:

$$s = \max(0, 1 - e^{-r}) \quad (1)$$

其中, r 是根据 GA/T 751-2008《视频图像文字标注规范》标准和时间戳字符数量计算得到, 文中计算为 19; x 为文本实例中心 D 的长宽比。

(4) 获取最终结果: 一个小检测框是否与文本实例中心 D 连接取决于它的最终概率 P 是否大于设定阈值, 文中设定为 0.5。 P 的计算公式如下:

$$P = \lambda * s + (1 - \lambda) a \quad (2)$$

最终, 将所有符合条件的小检测框与文本实例中心 D 连接得到最终的检测结果。

2.2 时间戳同步匹配方法

首先, 针对同一监控设备在多次检测中可能出现同步与不同步并存的矛盾情况, 文中设计了‘ $N-1$ 命中’原则, 基本思想为, 在时间戳的多次识别中得到 N 个时间结果, 对这 N 个结果依次执行容差匹配, 只要有一个匹配成功, 就认为这路设备时间同步并停止对剩余结果的匹配过程。如果 N 个结果全部匹配失败, 则会提示该路设备时间异常, 需要人工修正。

其次, 由于视频流采集图像帧在网络传输、设备响应等各个环节存在不确定的时延现象, 文中使用了‘容差匹配’方法, 其主要思想是依据一次识别结果时间差值是否满足一定容差(Ω)进行条件判断, 以尽可能减小视频流采集图像帧在网络传输、设备响应等各环节存在不确定延现象对时间戳同步判断的影响。具体计算方法为: 从图片中识别出的时间信息记为 T_A , 应用任务在发出图片采集指令时的时间记为 T_B , 容差值为 Ω , 计算:

$$|T_A - T_B| < \Omega \quad (3)$$

如果上式成立, 表示视频监控设备中图像的时间信息符合标准时间的容差范围(Ω), 即为容差匹配。

否则即为不匹配。

根据上述的‘ $N-1$ 命中’原则与容差匹配方法, 时间戳同步匹配的具体流程如图 4 所示, 分为以下几个步骤:

步骤 1: 在某一时段, 请求视频监控系统中一路监控点位的摄像机的实时视频码流, 并通过解码得到视频图像。同时, 记录请求指令发出时的系统时间 T_0 ;

步骤 2: 将采集的视频图像输入基于深度学习的时间戳检测识别算法, 识别出时间值 T_i 并转换为标准时间戳格式输出, 对于无法转换为标准格式的统一输出为“1999-01-01 00:00:00”;

步骤 3: 计算步骤 1 中的系统时间 T_0 与步骤 2 中的识别时间 T_i 之间的差值 ΔT , 定义容差匹配的阈值 Ω 。如果 $\Delta T < \Omega$, 则该路摄像机时间同步, 判定结束。否则进入步骤 4;

步骤 4: 判断是否需要继续匹配, 如果继续匹配, 返回步骤 1。否则该路摄像机时间不同步。判定结束。

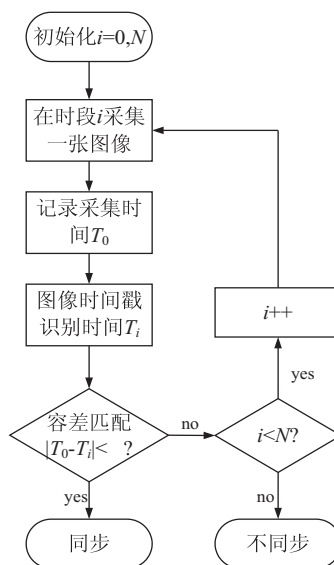


图 4 时间戳同步匹配流程

3 数据集及评价标准

3.1 数据集

文中实验中用于训练与测试的图像数据均来自于校企合作, 针对真实场景图片进行研究分析。数据集包含不同省份、不同时间段的市区、村镇和高速公路等场景采集的真实图像数据。根据不同任务, 文中将图分为时间戳区域检测数据集和时间戳同步数据集。

时间戳区域检测数据集包括三个集合, TR-4625 为训练集, 包含 4 625 张图片, 选取全天候各个时段全国各地随机抽取的图像, 包含透明字体、不同尺度文字和复杂背景下的文字; TS-2000 为测试集, 包含 2 000 张图片, 来源于湖北与广东, 选取的是白天时段且符合

GA/T 751-2008《视频图像文字标注规范》标准;TS-6060 为测试集,包含 6 060 张图片,来源于湖南、江西与江苏三个省份,时间段包括白天与夜晚,同时还出现透明字体、文字较小和复杂背景的现象。因此,相比于 TS-2000 集合,TS-6060 的图片复杂,检测难度更大。

时间戳同步检测包含一个测试集合 SYN-2648,包含 2 648 张图像,其中每张图像的标准时间(服务器发出采集指令的时刻)包含在图像名字中,然后通过人工对比设备时间(图像上标注时间)与标准时间的差值将集合 SYN-2648 分为差值在十秒内和差值在十秒以上两个文件。十秒为文中设置的容差匹配的容差值。

3.2 评价标准

目前文本检测的评价标准大多是基于目标检测评价标准的改变,没有考虑到文本检测任务的特殊性。文本检测的目的是为了识别,如果只是以交并比 IOU (intersection over union) 来判断文本是否被检测出来或检测框是否正确,无法体现出文本是否被检测完全。因此,从实际应用出发,文中提出一种以应用为驱动的文本检测的评价方法。该方法主要针对文本在实际应用中出现文本检测不完全但是交并比 IOU 却达到要求,特别是在水平方向上会缺少字符的问题。

假定对于一幅含特定目标区域的图像, $\{A_i \mid i = 1, 2, \dots, N\}$ 和 $\{B_j \mid j = 1, 2, \dots, M\}$ 分别为检测框 (prediction box) 集合与目标框 (ground-truth box) 集合。分别计算每个检测框 $\{A_i\}$ 和目标框 $\{B_i\}$ 的重叠比。计算方法如下:

$$\begin{cases} \text{IOU}_{AB}(i, j) = \frac{\iint (A_i \cap B_j)}{\iint (A_i \cup B_j)} \\ \text{IOU}_B(i, j) = \frac{\iint (A_i \cap B_j)}{\iint B_j} \end{cases} \quad (4)$$

其中, \cup 、 \cap 分别代表两个矩形框的并集与交集, $\iint(\cdot)$ 代表求面积运算。 $\text{IOU}_{AB}(i, j)$ 表示检测框 $\{A_i\}$ 与目标框 $\{B_i\}$ 在面积交集与并集的比值。 $\text{IOU}_B(i, j)$ 则以目标框 $\{B_i\}$ 为参考对象,侧重反映对于目标框的覆盖程度。此外,为了关注 box 在水平 x 方向的偏移程度,提出:

$$\begin{cases} \text{IOU}_{AB_x}(i, j) = \frac{\int (A_i \cap B_j)_x}{\int (A_i \cup B_j)_x} \\ \text{IOU}_{B_x}(i, j) = \frac{\int (A_i \cap B_j)_x}{\int (B_j)_x} \end{cases} \quad (5)$$

其中, $\int(\cdot)_x$ 代表求取矩形框在水平方向投影的长度。

$\text{IOU}_{AB_x}(i, j)$ 表示检测框 $\{A_i\}$ 与目标框 $\{B_i\}$ 在水平 x 方向(宽度方向)的交集与并集的比值, $\text{IOU}_{B_x}(i, j)$ 类似。检测框 $\{A_i\}$ 与目标框 $\{B_i\}$ 匹配可以依据阈值来确定。

$$\text{Match}_{AB}(i, j) = \begin{cases} 1, & \text{if } \begin{cases} \text{IOU}_{AB} > T_1 \\ \text{IOU}_B > T_1 + \lambda_1(1.0 - T_1) \\ \text{IOU}_{AB_x} > T_2 \\ \text{IOU}_{B_x} > T_2 + \lambda_2(1.0 - T_2) \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

其中, T_1, T_2 取 0.7, 0.9。系数 λ_1, λ_2 设为 0.5。

文中采用精准率 (Precision)、召回率 (Recall) 与 F 值对检测性能进行评价,计算方法如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$F = \frac{2\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

其中, TP、FP、FN、TN 的计算方式如下:

$$\begin{cases} \text{TP} = \text{TP} + 1, & \text{if } \forall A_i, \exists B_j, \text{Match}_{AB}(i, j) = 1 \\ \text{FP} = \text{FP} + 1, & \text{otherwise} \\ \text{FN} = \text{FN} + 1, & \text{if } \forall B_j, \exists A_i, \text{Match}_{AB}(i, j) = 1 \\ \text{TN} = \text{TN} + 1, & \text{otherwise} \end{cases} \quad (10)$$

时间戳同步检测采用的评价指标为准确率 (Accuracy, ACC), 计算公式为:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

其反映的是时间戳同步判断正确的样本占总样本的比例。TP 表示标签为同步,算法判定也为同步;TN 表示标签为不同步,算法判定也为不同步;FP 表示标签为不同步,算法判定为同步;FN 表示标签为不同步,算法判定为同步。

4 实验分析

4.1 时间戳区域检测

首先,为了分析提出的文字增强模块与候选框误差近邻传播连接方法对算法性能的影响,在相同训练集与测试集上的对比结果如表 1 所示。

其中, D_Net 为单独使用检测分支得到的检测结果, S_Net 为单独使用分割分支得到的检测结果, CBAP 为使用候选框误差近邻传播连接方法的检测结果, CBAP+TE 为使用候选框误差近邻传播连接方法

表 1 改进部分对性能的影响

| 方法 | TS-2000 | | | TS-6060 | | |
|---------|---------|-------|-------|---------|-------|-------|
| | P | R | F | P | R | F |
| D_Net | 0.899 | 0.908 | 0.903 | 0.687 | 0.747 | 0.711 |
| S_Net | 0.975 | 0.978 | 0.976 | 0.787 | 0.850 | 0.818 |
| CBAP | 0.974 | 0.977 | 0.975 | 0.794 | 0.852 | 0.822 |
| CBAP+TE | 0.982 | 0.984 | 0.983 | 0.820 | 0.869 | 0.844 |

与文字增强模块的检测结果。从表中可以看出,对于测试集 TS-2000,文中提出方法能在已经比较好的结果上更进一步。对于测试集 TS-6060,由于测试集中

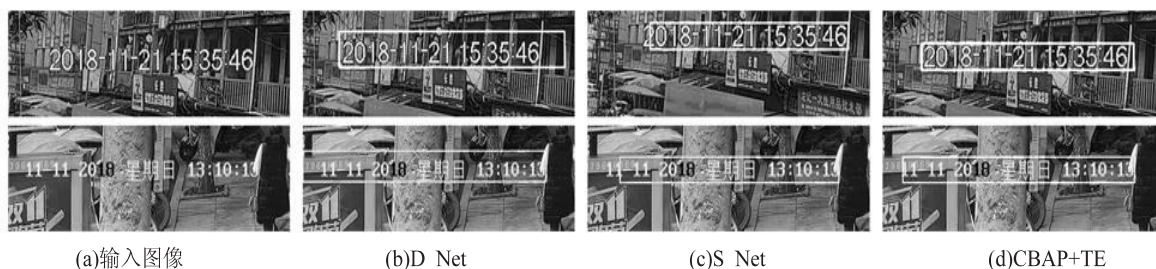


图 5 不同方法结果对比

其次,为了验证文中方法的有效性,选取了基于分割的 PSENet 与文中方法进行比较。得出的结果如表 2 所示。可以看出,文中方法在 TS-2000 测试集上能达到目前最新文本检测算法的效果。在 TS-6060 这个检测难度更大的测试集中,文中方法在精准率 (precision) 与 F 值上的表现更加优异。

表 2 与其他算法比较

| 方法 | TS-2000 | | | TS-6060 | | |
|---------|---------|-------|-------|---------|-------|-------|
| | P | R | F | P | R | F |
| PSENet | 0.907 | 0.937 | 0.922 | 0.749 | 0.873 | 0.806 |
| CBAP+TE | 0.982 | 0.984 | 0.983 | 0.820 | 0.869 | 0.844 |

4.2 时间戳同步检测

文中用 2 648 张图像测试了提出的时间戳同步判定方法,检测结果如表 3 所示,

表 3 时间戳同步检测结果

| 标签 | 同步 | 不同步 | 准确率 ACC |
|-----|-------|-------|---------|
| 同步 | 1 119 | 2 | 0.998 |
| 不同步 | 3 | 1 524 | |

从表中结果可以看出,时间戳同步判定的准确率达到了 0.998,证明了提出的基于容差匹配的时间戳同步判定方法的合理性。在失败的 5 组案例中,统计发现有 3 组是由于图像本身时间戳存在缺陷或干扰导致识别算法输出错误的时间信息,影响了时间戳同步的判定。如图 6 所示,a、b 组时间戳文字由于存在缺失、重叠的现象影响了识别结果的准确性,导致时间戳同步判定失败。而 c 组时间戳周围存在其他数字的干扰使得同步判定出错。

包含透明字体、复杂背景等各种不符合标注标准的样本,检测难度较大。单独使用基于检测或分割的方法很难在这种多模态测试集上取得较好的检测效果。而文中提出的方法融合了检测与分割的优点,能一定程度上提高模型对于这种多模态测试集的检测效果。同时,该方法在边缘细节部分的表现更加优秀,在文本区域的定位更加精确,能够为识别任务提供更好的样本。如图 5 所示,文中方法能得到更完整的文本区域且包含更少的非文本区域,而单独使用某一支得到的检测结果则出现缺少字符和检测框过大的现象。



图 6 同步判定失败图像

5 结束语

文中提出了一种针对安防视频监控图像的时间戳同步检测方法,包括时间戳区域检测与时间同步判定两个部分。首先,相比于目前的场景文本检测算法在检测时间戳区域时会出现检测框过大或检测不完全的现象,提出的时间戳区域检测算法针对安防视频监控图像时间戳区域的文字特点,能有效地检测出完整的时间戳区域。在 TS-2000 数据集上取得了 0.982 的精准率,在 TS-6060 数据集上取得 0.82 的精准率。其次,提出的时间同步判定方法,在测试集 SYN-2648 上取得了 0.998 的准确率,证明了该方法的合理性。

参考文献:

- [1] 鲍复民,李爱国,覃征.彩色照片时间戳识别[J].复旦学报:自然科学版,2004,43(5):914-917.
- [2] 姚文杰,艾旭升,栾咏红,等.建立时间戳识别模型的一种新方法[J].福建电脑,2019,35(11):84-86.
- [3] 于之靖,王嘉伟,郑建文,等.跨域标记牌文字检测算法研究[J].计算机应用与软件,2019,36(5):133-137.
- [4] 蒋冲宇,鲁统伟,闵峰,等.基于神经网络的发票文字检

- 测与识别方法[J]. 武汉工程大学学报, 2019, 41(6): 586-590.
- [5] 葛 艳, 陈 晨. 基于文字定位和颜色检测的多车牌定位方法研究[J]. 计算机与数字工程, 2018, 46(3): 575-579.
- [6] 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. Columbus, OH: IEEE, 2014: 580-587.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016: 21-37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV: IEEE, 2016: 779-788.
- [9] ZHONG Z, JIN L, ZHANG S, et al. Deeptext: a unified framework for text proposal generation and text detection in natural images[J]. arXiv:1605.07314, 2016.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [11] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, MA: IEEE, 2015: 1-9.
- [12] 杨宏志, 庞 宇, 王慧倩. 基于改进 Faster R-CNN 的自然场景文字检测算法[J]. 重庆邮电大学学报: 自然科学版, 2019, 31(6): 876-884.
- [13] 龙 言, 施水才, 肖诗斌. 基于分支 ResNet-50 的 EAST 场景文字检测算法[J]. 北京信息科技大学学报: 自然科学版, 2020, 35(3): 94-98.
- [14] 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.
- [15] 林 泓, 卢瑶瑶. 聚焦难样本的区分尺度的文字检测方法[J]. 浙江大学学报: 工学版, 2019, 53(8): 1506-1516.
- [16] ZHANG Z, ZHANG C, SHEN W, et al. Multi-oriented text detection with fully convolutional networks[C]//2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas, NV: IEEE, 2016: 4159-4167.
- [17] 杨剑锋, 王润民, 何 璇, 等. 基于 FCN 的多方向自然场景文字检测方法[J]. 计算机工程与应用, 2020, 56(2): 164-170.
- [18] YAO C, BAI X, SANG N, et al. Scene text detection via holistic, multi-channel prediction[J]. arXiv: 1606.09002, 2016.
- [19] DENG D, LIU H, LI X, et al. Pixellink: detecting scene text via instance segmentation[C]//Proceedings of the AAAI conference on artificial intelligence. New Orleans: AAAI, 2018.
- [20] WANG W, XIE E, LI X, et al. Shape robust text detection with progressive scale expansion network[J]. arXiv: 1903.12473, 2019.
- [21] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network[C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016: 56-72.
- [22] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI: IEEE, 2017: 936-944.
- [23] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]//2017 IEEE international conference on computer vision (ICCV). Venice: IEEE, 2017: 764-773.