



## 基于深度学习的工业视觉箱体字符识别与判断

葛永杰<sup>1,3</sup>, 王丽丹<sup>1,2,3,4</sup>, 陈定喜<sup>5</sup>, 段书凯<sup>1,2,3,4,6</sup>, 干秀灵<sup>1,3</sup>

(1. 西南大学 电子信息工程学院, 重庆 400715; 2. 智能传动和控制技术国家地方联合工程实验室, 重庆 400715;  
3. 类脑计算与智能控制重庆市重点实验室, 重庆 400715; 4. 重庆市脑科学协同创新中心, 重庆 400715;  
5. 美的集团, 广东 佛山 528311; 6. 西南大学 人工智能学院, 重庆 400715)

**摘要:** 工厂生产线上的商品包装外箱文本印刷存在残缺, 无法及时检出会影响流通销售。制作工业商品外观信息数据集, 提出基于深度学习的工业视觉箱体字符识别与匹配判断方法。合并YOLOv3中的卷积层和批量归一化层, 引入GIoU作为边界框损失函数并设计自适应调整定位坐标的方法, 优化在原始图像上进行文本检测定位的速度与精度。同时, 训练并对比CRNN和Tesseract两种识别引擎在已裁剪文本图片上的识别性能, 设计字符匹配方法判断字符识别正确与否并输出结果, 从而减少误判。对基于该方法的系统进行生产线实测, 实验结果表明, 其识别准确率可达99.5%, 单件商品的外观拍照、检测识别、输出结果耗时仅3 s左右, 表明所提方法能够实现实时监测。

**关键词:** 深度学习; YOLOv3算法; 卷积递归神经网络; 字符识别; 外观信息; 实时监测

开放科学(资源服务)标志码(OSID):



中文引用格式: 葛永杰, 王丽丹, 陈定喜, 等. 基于深度学习的工业视觉箱体字符识别与判断[J]. 计算机工程, 2022, 48(1): 296-304.

英文引用格式: GE Y J, WANG L D, CHEN D X, et al. Character recognition and judgment of industrial vision box based on deep learning[J]. Computer Engineering, 2022, 48(1): 296-304.

## Character Recognition and Judgment of Industrial Vision Box Based on Deep Learning

GE Yongjie<sup>1,3</sup>, WANG Lidan<sup>1,2,3,4</sup>, CHEN Dingxi<sup>5</sup>, DUAN Shukai<sup>1,2,3,4,6</sup>, GAN Xiuling<sup>1,3</sup>

(1. College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China;  
2. National and Local Joint Engineering Laboratory of Intelligent Transmission and Control Technology, Chongqing 400715, China;  
3. Chongqing Key Laboratory of Brain-Inspired Computing and Intelligent Control, Chongqing 400715, China;  
4. Chongqing Brain Science Collaborative Innovation Center, Chongqing 400715, China; 5. Midea Group, Foshan, Guangdong 528311, China;  
6. School of Artificial Intelligence, Southwest University, Chongqing 400715, China)

**[Abstract]** If the incomplete text printing on commodity packaging boxes produced by factory production lines cannot be detected in time, the sales and circulation of the commodities will be affected. This paper presents a deep learning-based box character recognition and matching method for industrial vision, and also makes a data set of industrial commodity appearance information for the method. By merging the convolutional layer and the batch normalization layer of YOLOv3, and introducing GIoU as the loss function of the boundary box, a method for adaptive positioning coordinate adjustment is designed, which improves the speed and accuracy of text detection and location on the original image. Then the recognition performance of the trained CRNN and Tesseract engines on cropped text images is compared. The designed character matching method is used to judge whether the character recognition result is correct, and the result is output, which reduces the misjudgment. The system based on this method is tested on a production line, and the experimental results show that the system displays an accuracy of 99.5%. It takes about 3 s to take a photo of the appearance, detect and recognize the characters, and output the result of a single product, which demonstrates that the proposed method enables real-time monitoring.

**[Key words]** deep learning; YOLOv3 algorithm; Convolutional Recurrent Neural Network (CRNN); character recognition; appearance information; real-time monitoring

DOI: 10.19678/j.issn.1000-3428.0059680

基金项目: 国家重点研发计划(2018YFB1306600); 国家自然科学基金(62076207, 62076208, U20A20227, 61672436); 重庆市基础科学与前沿技术研究专项重点项目(cstc2017jcyjBX0050)。

作者简介: 葛永杰(1993—), 男, 硕士研究生, 主研方向为机器学习、计算机视觉; 王丽丹(通信作者), 教授、博士、博士生导师; 陈定喜, 硕士; 段书凯, 教授、博士、博士生导师; 干秀灵, 硕士研究生。

收稿日期: 2020-10-10 修回日期: 2020-12-24 E-mail: ldwang@swu.edu.cn

## 0 概述

工厂商品包装外箱制作是工业生产流程中的重要环节,外箱的产品批号和生产日期信息具有重要意义,出厂前须确保字迹清晰完整,不能出现漏印、错印、空白箱等现象,否则会影响商品的流通与销售。目前,工厂中商品包装外箱上字符检测的相关方法和应用产品较少,且实时性与效果不佳,主要处于人工检查阶段,未实现实时自动化全覆盖检测。商品包装外箱上的字符检查,可以利用光学字符识别(Optical Character Recognition, OCR)方法实现生产线智能化字符识别判断与商品全覆盖。

OCR<sup>[1]</sup>一般包含文本检测和文本识别两个阶段。传统OCR的文本检测以图像处理方法为主进行文字定位,处理的对象往往局限于成像清晰、排列规整的文档图像,无法很好地处理背景复杂的图像。随着深度学习技术的发展,在OCR的文本检测领域中相继出现了一系列基于深度学习的文本检测算法,如CTPN算法<sup>[2]</sup>借助anchor回归机制并使用滑动窗口来检测目标所在区域,SegLink算法<sup>[3]</sup>通过引入Segment和Linking两个概念实现了对具有一定旋转角度文本的检测,EAST算法<sup>[4]</sup>实现了对各个方向文本的检测。同时,由于文本检测可以被视为一种目标检测,因此也可以利用主流的目标检测算法进行文本检测与定位。文献[5]基于R-CNN(Region-Convolutional Neural Network)和Fast RCNN<sup>[6]</sup>目标检测算法提出Faster R-CNN算法,实现了区域建议网络和检测网络共享卷积特征,使运行速度获得大幅提升。文献[7-9]提出的YOLO系列算法,将候选框提取、特征提取、目标分类与定位统一到一个神经网络中,具有运行速度快和检测准确率高的优势。OCR文字识别阶段以长短期记忆网络(Long Short-Term Memory, LSTM)<sup>[10]</sup>和CTC<sup>[11]</sup>(Connectionist Temporal Classification)的组合为主,同时也有实现端到端文字检测与识别的方法,如文献[12]通过共享训练特征、监督的方法完成文本检测和识别。目前OCR在车牌、票据、手写体等方面有很多方法和应用产品。文献[13]针对手写体数字识别提出一种基于融合卷积神经网络的字符识别模型。文献[14]对票据文本字符识别进行了研究,从光学系统、硬件结构和深度学习算法等方面进行优化,设计一个基于CCD的票据图像实时文本检测机器视觉系统。文献[15]针对车牌识别速度慢的问题,提出一种自然交通场景下车牌定位和识别的端到端深度学习模型。此外,文献[16]基于卷积神经网络对汽车零件编码字符进行识别,文献[17]基于几何特征对IC芯片字符分割识别,文献[18]基于改进卷积神经网络对铝轮毂背腔字符识别,文献[19]针对字符质量低的文本提出一种图像增强算法和卷积循环神经网络字符识别模型。目前,许多企业都在开展各种场景下字符识别的研究,并逐步对外提供各场景下的商用服务。

在计算机技术和人工智能应用不断发展的背景下,研究基于深度学习的工业视觉方法具有重要意义。本文

结合工厂生产需求和产业升级改造要求,提出基于深度学习的工业视觉箱体字符识别与判断方法。通过使用具体的工业商品外观信息数据集,利用深度学习方法实现箱体字符的检测与识别,从而提升生产线自动化和智能化水平。

## 1 工业场景商品外观信息数据集的制作与标注

本文检测识别的目标文本即商品外观信息,具体指箱体的产品批号和生产日期。本文使用工业相机在生产线上共采集20 000幅商品外观信息图像,制作工业场景商品外观信息数据集,其中图像分辨率为1 600像素×1 200像素。

OCR的首要任务便是文本检测,文本字符的变化和应用场景的复杂性,对检测方法的性能有很大的影响。因此,本文从文本、图像和应用场景等多方面考虑,制作了工业场景商品外观信息数据集,采集图像的特点和应用要求如下:

1) 字符文本长度存在变化,如图1(a)和图1(b)所示,工厂生产线的产品批号种类不一,导致图像上的产品批号字符文本长度存在变化。目前需要识别的产品批号种类近30种,而且每年每季度会有相应的更新,其为数字、字母、符号组合文本,所包含字符数量从13到25不等。

2) 字符文本内容存在变化,如图1(a)和图1(b)所示,工厂生产线生产商品的批号具有随机性,并且产品批号在不断变化更新,同一条生产线,一天之内可能生产多个批号商品。

3) 字符文本字体存在变化,如图1(c)和图1(d)所示,工厂各生产线生产的商品日期文本的形式相对固定,但日期文本字符粗细存在明显变化。

4) 图像表面光线分布不均匀,如图1(e)~图1(g)所示,工厂生产线光线环境较差,虽然利用补光灯进行光照增强,但相机拍摄到的图像成像质量依然不佳,同一幅图像上光线不均的情况十分普遍。同时通过图1(h)的直方图可知,图像的灰度值集中分布在75以内,成像整体偏暗。

5) 图像存在远近变化,如图1(e)和图1(f)所示,前者为近景,后者为远景。工厂生产线生产商品是一个运动的过程,且相机对同一商品箱体两侧拍照。图像的远近变化影响了图像表面光线强度和箱体字符大小的变化。

6) 基于该方法的系统应用环境存在变化,如图1(e)~图1(g)所示,其中,图1(e)、图1(f)为同一生产线,图1(g)为另一条生产线。工厂环境复杂,不同生产线物理环境变化较大,例如箱体大小、光线强度、拍摄角度、生产线运行速度等,导致成像质量不一,同时字符格式也存在相应变化。

7) 基于该方法的系统需要满足不同生产线可以使用同一个网络模型的要求,网络要具有优良的鲁棒性,能够适应生产线环境变化。

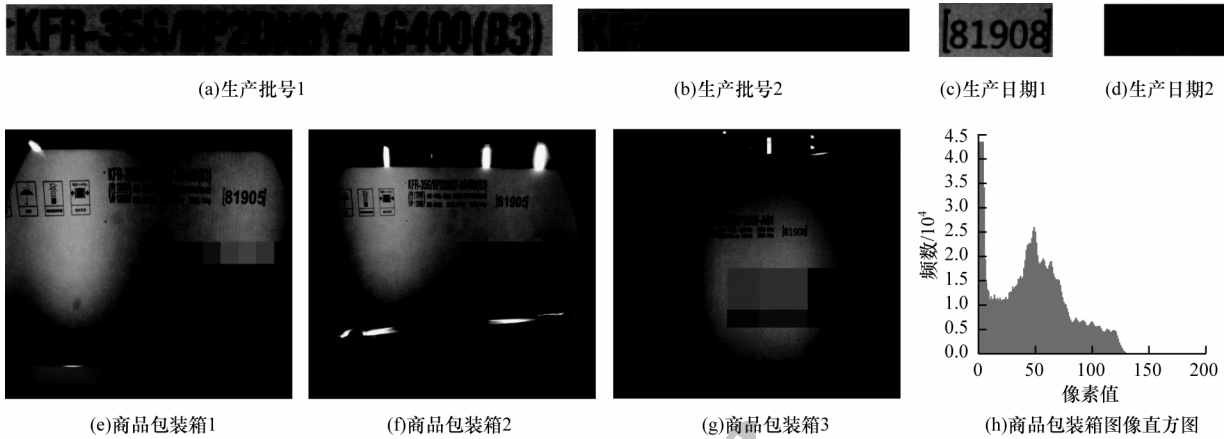


图1 生产线实拍图、切割图及直方图

Fig.1 Real pictures ,cutting diagrams and histogram in production line

利用图像标注工具手动地对工业商品外观信息数据集进行标注,设置产品批号(model)和生产日期(date)两个类别。

### 2 字符识别与匹配判断

字符识别与匹配判断包含3个模块,即目标检测模块、字符识别模块和匹配判断模块。目标检

测模块利用CNN对图像中的产品批号和生产日期文本区域进行检测定位,然后依据定位文本框裁剪图像并送入字符识别模块进行字符识别,将字符识别结果输送到匹配判断模块,利用模板组合字符文本对字符识别结果进行分段全匹配,输出相应的结果提示。字符识别与判断流程如图2所示。

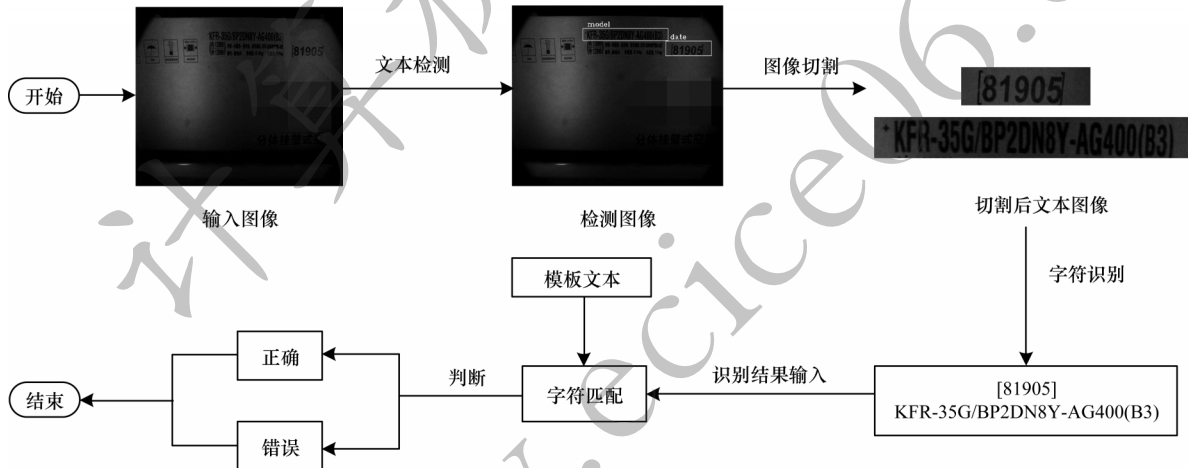


图2 字符识别与匹配判断流程

Fig.2 Procedure of character recognition and matching judgment

#### 2.1 目标检测模块

##### 2.1.1 目标检测算法

商品外观信息的文本格式较为固定,不存在较大的角度旋转问题,因此,本文利用目标检测算法进行文本检测。考虑对速度和精度的高要求,选取YOLO系列的算法,在源码YOLOv3基础上将卷积层与批量归一化层合并,并将GIoU<sup>[20]</sup>(Generalized Intersection over Union)作为回归目标框的损失函数,称为G-YOLOv3。将工业商品外观信息数据集以6:2:2的比例分为训练集、验证集和测试集,对YOLOv2、YOLOv3和G-YOLOv3分别进行训练,网络共迭代10 000次,每1 000次保存1个模型,初始学习率为0.001,权值衰减率和动量分别为0.000 5

和0.9。在CPU为i7-4720HQ、GPU为GTX 950M、内存为8 GB的计算机上进行测试,设置IoU<sup>[21]</sup>(Intersection over Union)阈值为0.75,测试结果如表1所示。可以看出,G-YOLOv3对字符文本的检测性能要优于YOLOv2和YOLOv3,在运行速度方面,YOLOv2性能出色,但其综合性能不如G-YOLOv3。

表1 目标检测性能测试结果

Table 1 Performance test results of object detection

检测算法	mAP	耗时/ms
YOLOv2	0.791	35
YOLOv3	0.884	50
G-YOLOv3	0.951	39

利用工业商品外观信息数据集训练,并绘制损失函数曲线,图3(a)为YOLOv3的损失函数曲线,图3(b)为G-YOLOv3的损失函数曲线。可以看出,G-YOLOv3的训练效果比较理想。选取G-YOLOv3的损失函数曲线趋于平缓的区间内保存的模型进行测试,即在函数

值8 000~10 000内保存的模型,检测定位的效果如图4所示。可以看出,模型能够准确定位同一商品箱体两侧的产品批号和生产日期字符文本区域,也能够适应不同生产线的物理环境,处理不同类型的文本字符以及不同的字符字体。

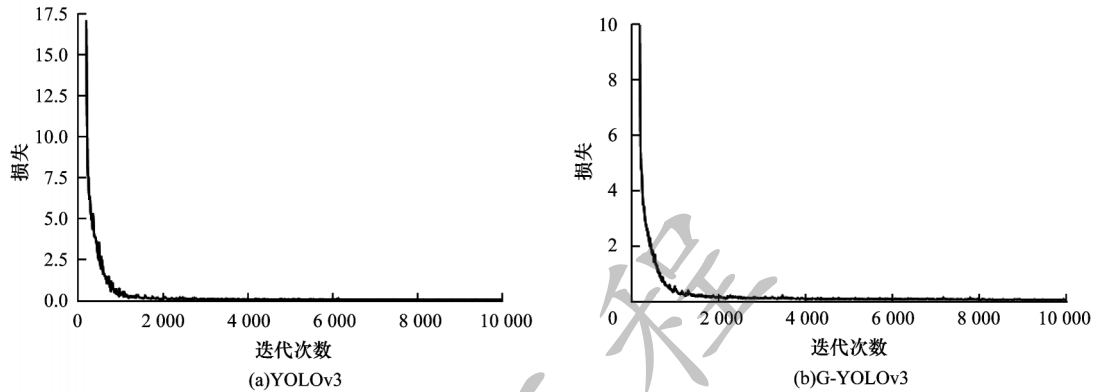


图3 损失函数曲线

Fig.3 Loss function curve

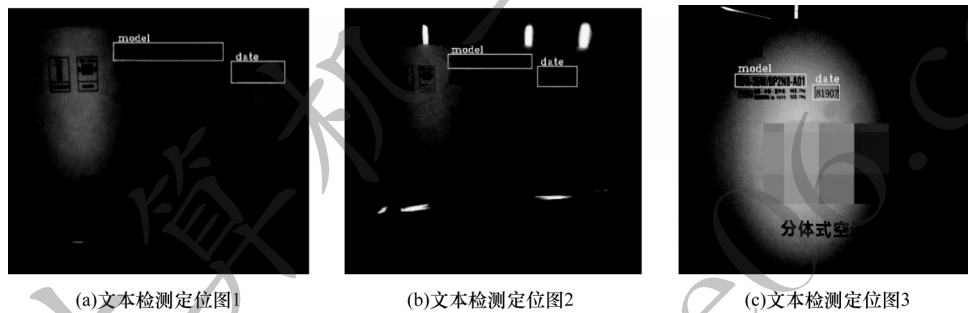


图4 实时检测定位图

Fig.4 Real time detection and positioning map

2.1.2 边界框优化

G-YOLOv3 经过工业商品外观信息数据集的训练,对多数图像的目标文本区域能够实现精准检测定位,但出现了部分图像文本定位存在少量偏差的问题。由于对目标区域的切割是依据边界框来进行的,少量的偏差会导致切割字符不准确,致使识别时出现错误,因此必须对目标文本实现精准定位,即边

界框包含完整的字符文本。在未进行边界框优化的情况下,G-YOLOv3 对于大部分图像能够实现精确定位和准确切割,最终正确识别,如表2中第1行所示;但在部分图像中,切割后的图像未能实现包含完整字符文本,一侧或者两侧会出现切割文本字符的情况,导致字符残缺或者整个字符缺失,造成误识别,如表2中第2行和第3行所示。

表2 边界框的影响

Table 2 Influence of bounding box

定位图像	切割后图像	识别结果	结果
		KFR-35G/BP2DN8Y-PH400(B3)	正确
		KFR-35G/BP2DN8Y-PH400(B:	字符残缺误识别
		(FR-35G/BP2DN8Y-PH400(B	字符残缺、缺失误识别

从生产线实时获取箱体两侧的图像存在远近的变化,即目标文本存在大小的变化,但定位边界框在高度上总能很好地适应这一变化。因此,可以借助边界框的高度变化来调整起始位置及其宽度,实现对G-YOLOv3边界框的改进优化。

边界框有4个参数,即 $x, y, w, h$ ,其中 $x, y$ 为边界框的左上顶点坐标, $w$ 为边界框的宽度, $h$ 为边界框的高度。任意时刻 $t$ 的边界框参数为 $(x_t, y_t, w_t, h_t)$ ,上一时刻边界框参数为 $(x_{t-1}, y_{t-1}, w_{t-1}, h_{t-1})$ ,通过参数 $h$ 的变化来调整边界框的定位区域,使其完全包含目标文本

区域。对于起始位置,即边界框的左上顶点,只需调整横轴坐标  $x$  的值即可,计算方法如式(1)所示:

$$x_i = \begin{cases} x_i \cdot \frac{h_i}{h_{i-1}}, & h_i < h_{i-1} \\ x_i \cdot \frac{h_{i-1}}{h_i}, & h_i \geq h_{i-1} \end{cases} \quad (1)$$

同时,对边界框的宽度  $w$  进行自适应变化并消除因对起始位置的调整而给边界框宽度带来的影响,其计算方法如式(2)所示:

$$w_i = \begin{cases} w_i \cdot \frac{h_{i-1}}{h_i} + \left(x_i - x_i \frac{h_i}{h_{i-1}}\right), & h_i < h_{i-1} \\ w_i \cdot \frac{h_i}{h_{i-1}} + \left(x_i - x_i \frac{h_{i-1}}{h_i}\right), & h_i \geq h_{i-1} \end{cases} \quad (2)$$

通过对 G-YOLOv3 边界框的改进优化,使边界框能够覆盖各种长度的字符文本,准确定位目标区域,包含目标文本的全部字符,为之后的字符识别奠定基础,改进优化后效果如图 5 所示。

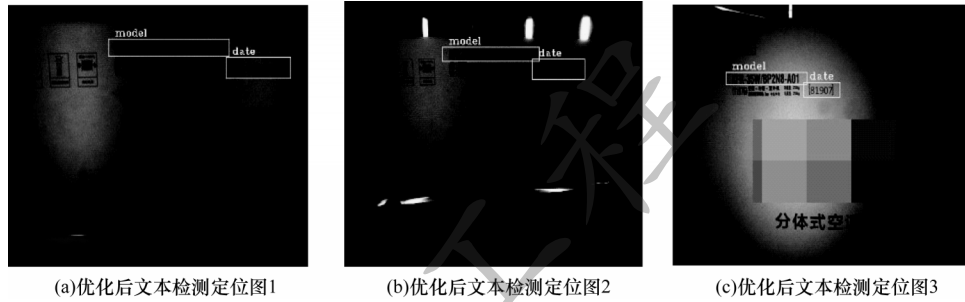


图 5 边界框优化后测试结果图

Fig.5 Test results after optimization of bounding box

## 2.2 字符识别模块

字符识别模块利用 Tesseract<sup>[22]</sup>进行识别,同时端到端识别的卷积递归神经网络(Convolutional Recurrent Neural Network, CRNN)<sup>[23]</sup>对该场景下的文本进行调优训练,并对比性能,该部分框架如图 6 所示。

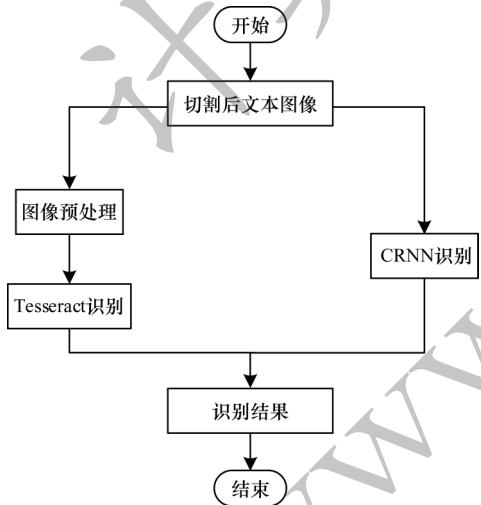


图 6 字符识别模块框架

Fig.6 Framework of character recognition module

### 2.2.1 Tesseract 识别引擎

利用 Tesseract 自带的权重,出现混淆字符的种类较多,频率较高,后续易混处理繁琐,为满足工厂生产线对实时性的高要求,必须进行调优训练。

#### 1) 调优训练

依据项目字符文本制作待训练文本数据集,在 ubuntu 16.04 版本下对 Tesseract 进行调优训练,经过近百万次迭代训练形成项目权重。

利用 10 000 张图像对自带权重和项目权重进行测试,测试结果如表 3 所示。可以看出,在自带权重的基础上调优训练形成的项目权重性能更加优越,其减少了混淆字符出现的种类,大部分易混字符的出现频率显著下降,下降率集中在 50% 以上,但个别字符也存在其他情况,例如:字母“B”易混频率虽然取得显著下降,但出现易混的次数不容忽视,仍然维持在较高值;数字“0”的易混频率虽然出现了上升,但出现次数不足 1%。对于这些情况,后期的易混处理要求较高。

表 3 权重测试统计表

Table 3 Statistical table of weight test

字符类	易混字符类	自带权重	项目权重	下降率/%
K	A	129	40	68.99
	Y	15	0	100.00
	X	388	167	56.96
G	0	374	212	43.32
	6	2 422	87	96.41
B	8	2 126	1205	43.32
D	0	1 368	434	68.27
	O	20	4	80.00
8	B	976	782	19.88
	S	1 515	22	98.55
G		70	13	81.42
0	D	18	69	-73.91

利用部分产品批号文本对自带权重和项目权重进行测试,并统计各个产品批号字符文本的置信度,测试结果如表 4 所示。可以看出,自带权重中的字符置信度集中分布在 0.75~0.8 之间,项目权重中的字符置信度分布在 0.9 左右,实现了至少 10% 的性能提升。

表 4 部分批号置信度

Table 4 Confidences of some batch numbers

部分产品批号字符文本	自带权重	项目权重	提升率/%
KF-23G/G01-D3	0.76	0.89	17.11
KF-26G/Y-DH400(D3)	0.77	0.87	12.99
KFR-35W/BP2N8-A02	0.48	0.90	87.5
KFR-35W/BP3N8-B23	0.51	0.87	70.59
KFR-32G/DY-DH400(D3)	0.78	0.91	16.67
KFR-26G/BP2DN8Y-PH400(B3)	0.78	0.92	17.95
KFR-35G/BP2DN8Y-AG400(B3)	0.57	0.91	59.65

2) 图像预处理

通过图 4 和表 2 的直观分析,可以得知工厂生产线拍摄的图像质量较差。为了提高 Tesseract 的识别正确率,需要对切割之后的图像进行优化增强。切割后得到的文本区域图像亮暗不均,总体偏暗,需要对其进行二值化处理,但不宜采用固定阈值的二值化方法,否则会出现大量噪音,如图 7 所示。本文采用自适应阈值的二值化处理,即通过图像的局部特征自适应地设定阈值进行二值化处理来避免出现这种情况。将二值图以白底黑字的形式呈现,然后使用中值滤波算法去除二值图中的噪声点。经过图片测试并且二值化处理后的图像质量相接近,此处的中值滤波器的尺寸选用 5×5。



(a) 固定阈值二值化图1



(b) 固定阈值二值化图2

图 7 二值化图

Fig.7 Binary graph

表 5 BlockSize 值对识别结果的影响

Table 5 Influence of Blocksize value to recognition result

切割后文本图像	自适应阈值二值化图	BlockSize 值	识别结果
		35	识别: KFR-35G/BP2DN8Y-AG400(83) 真值: KFR-35G/BP2DN8Y-AG400(B3)
		55	识别: KFR-35G/BP2DN8Y-AG400(B3) 真值: KFR-35G/BP2DN8Y-AG400(B3)
		105	识别: KFR-35G/BP2DN8Y-AG400(83) 真值: KFR-35G/BP2DN8Y-AG400(B3)
		35	识别: KFR-35G/BP2DN8Y-AG400(83) 真值: KFR-35G/BP2DN8Y-AG400(B3)
		45	识别: KFR-35G/DY-DH400(D3) 真值: KFR-35G/DY-DH400(D3)
		25	识别: KF-23G/Y-DH400(D3) 真值: KF-23G/Y-DH400(D3)
		15	识别: KF-236/V-DHA00(D3) 真值: KF-23G/Y-DH400(D3)

由于切割后得到的文本区域图像质量存在差异,因此在进行自适应阈值二值化时,需要对计算区域阈值中的区域大小(BlockSize)进行动态调整,其数值只能为奇数。BlockSize 值动态调整方式如图 8 所示。

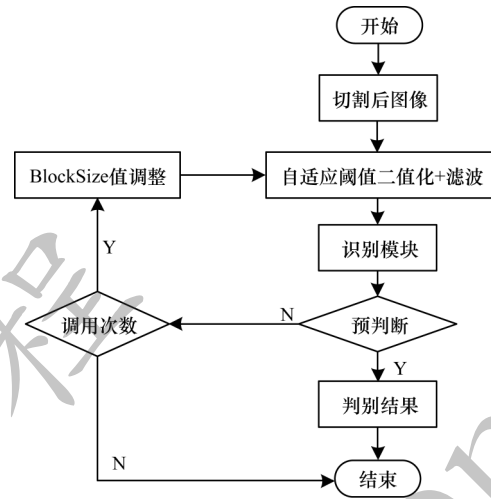


图 8 BlockSize 值调整流程

Fig.8 Procedure of adjusting Blocksize value

在同一条生产线采集图像对不同的 BlockSize 值进行测试,测试结果如表 5 所示。可以看出,同一图像在不同 BlockSize 值条件下,二值化图像会有变化,识别结果也有不同。同一图像,若 BlockSize 值过小,二值化图像中的部分字符信息损失严重,导致出现误识别;若 BlockSize 值过大,二值化图像中的字符出现粘连现象,出现误识别的可能性会提升。同一生产线含有相同产品批号的不同图像在相同 BlockSize 值条件下,二值化图像中字符信息保存完整程度不同,导致识别结果不同;同一生产线含有不同产品批号的不同图像对 BlockSize 值的要求也不同。

通过大批量图像测试,本文将BlockSize的初始值设为35值。BlockSize必须为奇数,但动态调整要适中,避免出现二值化图像中字符信息损失严重和字符间粘连的现象,因此,在算法中对BlockSize值的调整,设置了5次调整限制,并依据式(3)来改变BlockSize的数值,实现动态调整:

$$\text{BlockSize} = \text{BlockSize} + 10 \quad (3)$$

### 2.2.2 CRNN 识别引擎

利用经过数据集训练和边界框优化的G-YOLOv3模型,对从生产线采集的图像进行批量处理切割,形成文本区域图像,并选取产品批号图像120 000张,生产日期图像30 000张组成训练集,制作相应文本进行训练形成相应模型。选取8 000张产品批号图像和2 000张生产日期图像对CRNN进行测试,生产日期字符文本识别可达到100%的正确率,但产品批号存在部分易混字符,统计结果如表6所示。可以看出,在CRNN的测试中出现易混字符的频率集中出现在5%以下,并且易混种类较少。同时结合表3~表5可知,训练之后的CRNN字符识别性能要优于Tesseract。CRNN识别字符无需对图像进行预处理,减少了时间损耗,同时可以使用GPU加速,但Tesseract识别字符需要预处理图像进行增强,无法使用GPU加速,因此,在字符识别速度上Tesseract

要弱于CRNN。在内存为16 GB和显卡为GTX 1060的计算机上进行测试,结果表明,在字符识别模块中,CRNN每次耗时45 ms左右,Tesseract每次耗时150 ms左右。综合运行速度和识别字符性能,字符识别模块中选用CRNN进行字符识别。

表6 CRNN 测试结果

Table 6 CRNN test result

字符类	易混字符类	出现次数	出现频率/%
R	P	363	4.54
G	6	208	2.60
D	0	37	0.46
B	8	120	1.50
8	B	354	1.86
0	D	295	4.43
4	A	133	1.66

### 2.3 字符匹配模块

字符识别模块完成对目标文本区域的字符识别,其将识别结果输入字符匹配模块综合判断,然后输出判别结果。字符匹配模块流程如图9所示。

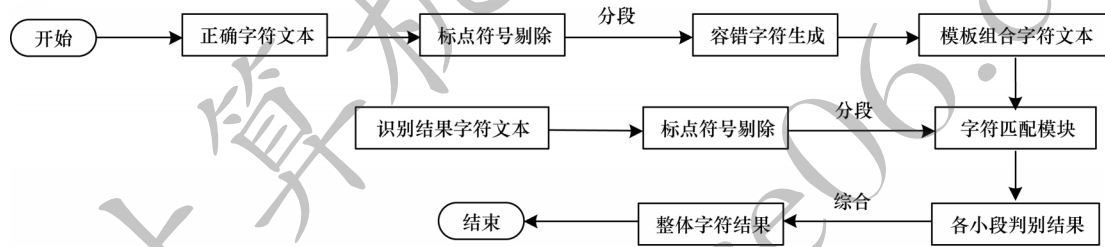


图9 字符匹配模块流程图

Fig.9 Flow chart of character matching module

生产日期为纯数字文本,且长度较短(一般为7个字符),形式固定,无需对其进行分段,可对识别文本直接匹配判断输出相应判别结果。产品批号是复杂的数字、字母、标点符号组合字符文本,文本长度存在变化且过长,文本内容也在变化更新,存在容易出现易混字符的文本,对其进行分段全匹配。本文针对产品批号字符文本,依据表6设计容错字符生成算法,进而提升系统的整体性能。

1)扫描商品包装外箱上的一维条形码从生产管理系统获取正确的产品批号字符文本,将文本中的标点符号剔除,统计剔除后字符文本的长度为L,设置分段数为5,则前4小段各为 $\lfloor L/5 \rfloor$ 个字符,第5小段为 $L - 4 \times \lfloor L/5 \rfloor$ 个字符。对每小段进行容错字符生成,产生模板组合字符文本。例如,小段“KFR”,经过容错字符生成算法处理后生成“KFRKFP”模板组合字符文本。若为相同条形码,上述容错处理算法只执行一次。

2)对字符识别模块输入的识别结果进行标点符号剔除,和剩余字符相应的分段处理,形成小段文本。

3)将识别结果的小段文本与其对应的模板组合字符文本进行全字符匹配,输出相应小段判别结果。

4)综合各个小段判别结果,输出整个识别文本判别结果。

字符匹配模块中对易混字符的处理尤为重要,在

保证系统满足生产线实时性要求的条件下,依据表6设计的容错字符生成算法既能降低系统出现误判的可能性,又能对外观信息文本存在印刷质量问题的商品进行及时报警提醒,保障了工厂生产线的正常运行。

## 3 生产线实测

将基于本文方法的系统在工厂生产线上进行实测,实现商品全覆盖、实时监测、自动统计、异常处理,生产线系统装置示意图如图10所示。

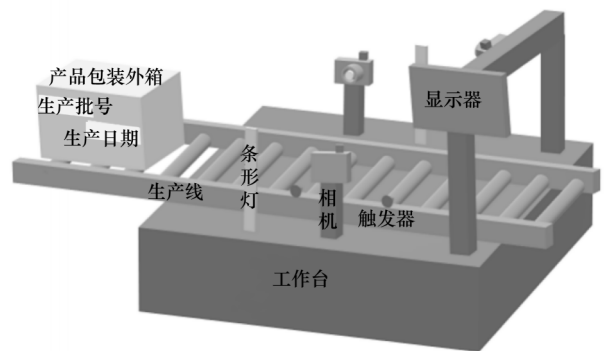


图10 生产线系统装置示意图

Fig.10 Schematic diagram of production line system

生产线计算机配置为内存16 GB、显卡GTX 1060,系统从相机拍照到输出判别结果耗时3 s左右,其中算法部分运行时间在200 ms以内,准确率可到达99.5%,

对存在字符印刷异常的商品实时报警提示,检测出的存在印刷质量问题的产品批号和生产日期字符文本部分图像如表7所示。

表7 存在印刷质量问题的部分图像

Table 7 Some images with printing quality problems

印刷质量问题文本图像	正确印刷文本	印刷质量问题
	81908	漏墨
	81908	错印
	KFR-35W/BP3N8-B12	漏印
	KFR-35G/DY-DH400(D3)	水平漏墨
	KFR-35G/BP3DN8Y-PG100(B1)	水平拖墨
	产品批号和生产日期	空白箱

#### 4 结束语

针对工厂商品包装外箱字符检测问题,本文制作数据集并提出一种基于深度学习的字符识别与判断方法。该方法合并YOLOv3网络的卷积层与批量归一化层,引入边框损失函数GloU,并利用边框调整优化方法,使边界框完全包含文本区域,降低误识别率。同时,将经过场景文本训练的CRNN应用于识别捕获的文本图片内容,保证快速和高准确识别,并设计字符模板匹配方法减少误判,保证生产线的良好运行。实验与生产线实测结果表明,本文方法能够达到全覆盖、实时监测、高准确率的应用目标。后续将研究深度卷积网络的优化方法和边界框损失函数的调整方法,减少图像处理过程,进一步提升系统的运行速度与识别精度。

#### 参考文献

- [1] LONG S, HE X, YAO C. Scene text detection and recognition; the deep learning era[J]. International Journal of Computer Vision, 2021, 129: 161-184.
- [2] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network [C]// Proceedings of 2016 IEEE European Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2016: 56-72.
- [3] SHI B, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 3482-3490.

- [4] ZHOU X, YAO C, WEN H, et al. EAST: an efficient and accurate scene text detector [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2642-2651.
- [5] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [6] GIRSHICK R. Fast R-CNN [C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1440-1448.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6517-6525.
- [9] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08)[2020-10-05]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [11] GRAVES A, FERNANDEZ S, GOMEZ F, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks [C]// Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, USA: [s. n.], 2006: 369-376.
- [12] LIU X D, LIANG D, YAN S. FOTS: fast oriented text spotting with a unified network [C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern

- Recognition. Washington D. C. , USA :IEEE Press, 2018; 5676-5685.
- [13] 陈玄,朱荣,王中元. 基于融合卷积神经网络模型的手写数字识别[J]. 计算机工程, 2017, 43(11): 187-192.  
CHEN X, ZHU R, WANG Z Y. Handwritten digit recognition based on fused convolution neural network model[J]. Computer Engineering, 2017, 43(11): 187-192. (in Chinese)
- [14] ZHAO S, SUN L, LI G, et al. A CCD based machine vision system for real-time text detection [J]. Frontiers of Optoelectronics, 2019(7): 1-7.
- [15] 史建伟,章韵. 基于改进YOLOv3和BGRU的车牌识别系统[J]. 计算机工程与设计, 2020, 41(8): 2345-2351. (in Chinese)  
SHI J W, ZHANG Y. License plate recognition system based on improved YOLOv3 and BGRU [J]. Computer Engineering and Design, 2020, 41(8): 2345-2351. (in Chinese)
- [16] LI D Y, TANG Q, ZHOU H, et al. Character recognition for automotive parts coding based on convolutional neural network[J]. Journal of Physics: Conference Series, 2020, 1518: 1-9.
- [17] 郭晓峰,王耀南,毛建旭. 基于几何特征的IC芯片字符分割与识别方法[J]. 智能系统学报, 2020, 15(1): 144-151.  
GUO X F, WANG Y N, MAO J X. IC chip character segmentation and recognition method based on geometric features[J]. CAAI Transactions on Intelligent Systems, 2020, 15(1): 144-151. (in Chinese)
- [18] 程淑红,周斌. 基于改进CNN的铝轮毂背腔字符识别[J]. 计算机工程, 2019, 45(5): 182-186.  
CHENG S H, ZHOU B. Character recognition of aluminum wheel hub back cavity based on improved CNN [J]. Computer Engineering, 2019, 45(5): 182-186. (in Chinese)
- [19] 何鑑一,杨国为. 基于深度学习的光照不均匀文本图像的识别系统[J]. 计算机应用与软件, 2020, 37(6): 184-190, 217.  
HE L Y, YANG G W. Recognition system of uneven illumination text image based on deep learning [J]. Computer Applications and Software, 2020, 37(6): 184-190, 217. (in Chinese)
- [20] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA :IEEE Press, 2020; 658-666.
- [21] YU J, JIANG Y, WANG Z, et al. UnitBox: an advanced object detection network[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York, USA: ACM Press, 2016: 1-5.
- [22] SPORICI D, BOIANGIU C A. Improving the accuracy of Tesseract4.0 OCR engine using convolution-based preprocessing[J]. Symmetry, 2020, 12(5): 715.
- [23] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(11): 2298-2304.

编辑 金胡考