

DOI:10.19344/j.cnki.issn1671-5276.2026.02.022

融入反光特征的金屬壓印字符識別深度學習模型

趙睿棟¹, 李立¹, 歐陽², 崔大賓¹

(1. 西南交通大學 機械工程學院, 四川 成都 610031; 2. 成都大學 機械工程學院, 四川 成都 610106)

摘 要: 針對因金屬材質反光而導致的火車大線銘牌壓印字符識別精度低的問題, 提出一種融入反光特征的金屬壓印字符識別深度學習模型, 記為 RFF-YOLOv5-C3CA。該模型以 YOLOv5 為基礎, 引入反光特征融合模塊(RFF), 增強圖像中壓印字符區域的對比度, 以便從中提取有效的圖像特征; 採用具有方向信息和位置信息的目标注意力模塊 C3CA, 增強模型對壓印字符的感知能力。實驗結果表明: 與 YOLOv5 模型相比, 所提模型對火車大線銘牌壓印字符數據集的識別精確度、召回率、以及 mAP 值分別提高了 3.46 個百分點、1.41 個百分點和 1.72 個百分點, 檢測速度 FPS 為 41 f/s, 由此說明所提算法的有效性。

关键词: 壓印字符; 反光特征; YOLOv5; 字符識別; 注意力機制

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1671-5276(2026)02-0114-07

Deep Learning Model for Metal Embossed Character Recognition Incorporating Reflective Features

ZHAO Ruidong¹, LI Li¹, OU Yang², CUI Dabin¹

(1. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China;

2. School of Mechanical Engineering, Chengdu University, Chengdu 610106, China)

Abstract: To improve the low recognition accuracy of characters embossed on train line nameplates due to the reflection of metal materials, a deep learning model for recognition of metal embossed characters incorporating reflective features is proposed, namely RFF-YOLOv5-C3CA. The proposed model, based on YOLOv5, introduces the reflective feature fusion (RFF) module to enhance the contrast of the imprinted character area in the image in order to extract effective image features, uses target attention module C3CA with direction information and position information reinforcing the model's perception of embossed characters. The experiment results show that compared with the YOLOv5 model, the recognition accuracy, recall rate, and mAP value of the proposed model on the train nameplate embossed character data set increases soon by 3.46 percentage points, 1.41 percentage points, and 1.72 percentage points respectively, and the detection speed FPS reaches 41 f/s, which demonstrates the effectiveness of the proposed algorithm.

Keywords: embossed characters; reflective features; YOLOv5; character recognition; attention mechanism

0 引言

随着经济贸易的快速发展铁路货车所承载的运载工作量日益增加, 对铁路货车进行必要的维护维修以保证铁路货车的正常运行就显得非常重要。对火车大线(如图 1 所示)的检测是其中重要的任务之一, 即对承载线路有效时间的压印字符铭牌进行检测, 压印字符铭牌如图 2 所示。压印字符铭牌是嵌入于火车大线的两端接头处。本文所要研究的对象是金属铭牌上的压印字符, 每一张铭牌包含了该火车大线的出厂编号以及检修时间的压印字符信息, 其中出厂编号是由该火车大线生产并检修合格出厂时的后两位年份数字、

月份以及该火车大线的 3 位数字身份码组成; 检修时间表示该火车大线生产并检修合格出厂时的年份以及月份。为保证铁路的正常运行, 需淘汰压印字符中的检修时间超过两年的火车大线。当下该工作是依靠人工识别金属铭牌上面的压印字符, 而传统的人工检查存在主观性强, 检测结果易受人眼疲劳的影响, 因此迫切需要一种自动化的检测手段。检测的难点包括: 1) 由金属制造的铭牌易反光, 导致字符信息不完整; 2) 由金属制造的铭牌表面易被氧化、油垢污染, 从而导致压印字符不清晰。这些难点导致现有的目标检测算法对该类压印字符检测效果不是很好。针对该问题, 本文提出融入反光特征的金屬壓印字符識別深度學習

基金项目: 四川省重大科技专项项目(2022ZDZX0007)

第一作者简介: 趙睿棟(1999—), 男, 山西运城人, 硕士研究生, 研究方向为机器视觉、图像处理, 969040157@qq.com。

模型。



图1 火车大线实物图



图2 压印字符铭牌图像

近年来,基于图像处理的字符检测技术在文本识别^[1]、喷点字符识别^[2]和身份证识别^[3]等领域取得了广泛的应用,其主要可分为传统图像处理和基于深度学习的图像处理方法。在传统图像处理方法中,字符分割是字符识别的前提,很大程度上影响着字符识别的准确性。针对压印字符分割,张传果等^[4]提出了一种先筛选字符区域后进行形态学优化的字符分割方法,克服了传统方法直接用形态学优化造成的干扰区域粘连字符的弊病。高潮等^[5]使用基于HOG的局部动态阈值算法和OTSU方法提取压在机械零件上的字符的两个二进制图像,然后将这两个结果进行融合以获得更优化的二进制图像。在基于深度学习的图像处理方法中,针对压印字符识别,汪志成等^[6]在

网络各卷积层中采用小尺寸卷积核,使用InceptionV2卷积模块取代C5全连接层,以提高LeNet-5模型对压印字符的识别精度。李建美等^[7]首先使用Gabor滤波器直接提取凸字符图像在水平和垂直方向上的局部笔画特征,以提高字符识别的准确率。此外,一些研究提出通过获取字符的深度信息来突出字符和背景之间的对比。向忠等^[8]通过利用在4个方向照明下的图像融合,增强了压印字符与背景的对比度。

传统图像处理方法大多依次进行字符定位、字符分割、字符识别,过程繁琐且最终的字符识别效果特别依赖于分割的好坏。现有的基于深度学习的图像处理算法用于压印字符时,大多针对高质量的图像,不适用于反光严重的火车大线压印字符铭牌的检测。因此,本文提出在YOLOv5s模型^[9]的基础上,引入反光特征融合模块(reflective feature fusion, RFF)和注意力机制模块C3CA,记为RFF-YOLOv5-C3CA,如图3所示。本文方法的主要贡献如下:

- 1) 针对金属铭牌字符图像反光问题,将反光问题转化为可提取的反光特征,增强了字符特征,降低了背景影响,提出字符反光特征融合网络模块;
- 2) 将所提字符反光特征融合网络模块与目标检测算法相结合,同时对网络模型融入了注意力机制,提出RFF-YOLOv5压印字符识别模型;
- 3) 为进一步提高模型的识别效果,提出一种简单有效的C3CA混合模块,增强了特征的表达能力,提高了识别准确率。

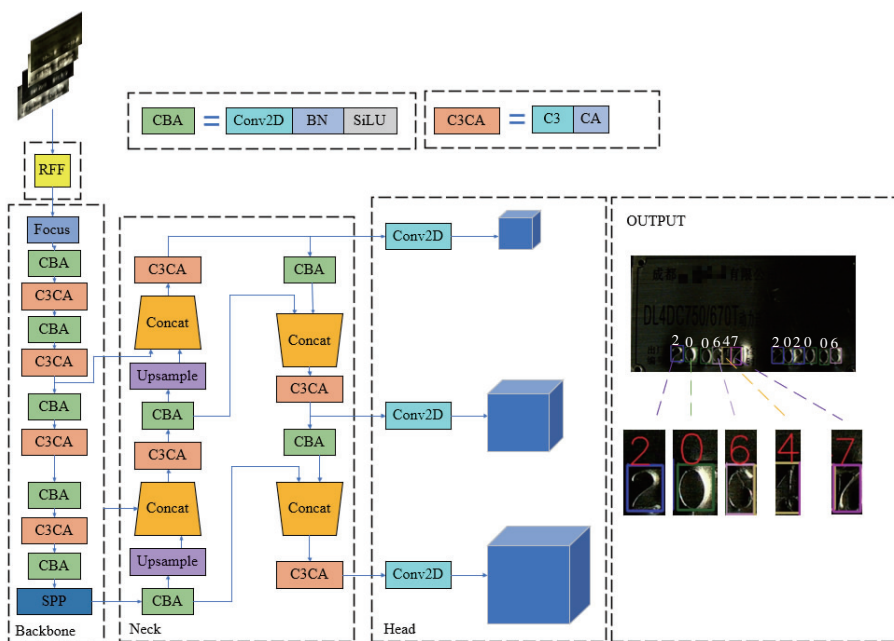


图3 所提模型网络结构

1 所提模型

1.1 网络结构

本文提出的融入反光特征的金属压印字符识别深度学习模型 RFF-YOLOv5-C3CA 主要由 3 部分组成:YOLOv5 负责基本的算法框架,反光特征融合模块 RFF 进行提取、融合、增强字符反光特征,C3CA 注意力机制混合模块用于提高特征的表达能力。YOLOv5 的基本网络结构主要由主干网络、多尺度特征融合网络、以及检测头 3 部分组成。YOLOv5 使用主干网络 Backbone 进行特征提取,该部分主要由 Focus、CSP 和 SPP 组成,Focus 对特征图进行切片操作,CSP 由残差网络结构组成,解决梯度消失和过拟合现象,SPP 解决了候选锚框与特征图各层之间的对齐问题。多尺度特征融合网络 Neck 采用特征金字塔从上向下传递语义信息和路径聚合结构来传递定位信息,提高特征层之间的交互性,对提取到的有效特征层信息进行交互融合,从而进一步提高模型的检测性能。检测头 Head 输出 3 个不同尺度大小的预测特征图并使用非最大抑制算法(NMS)筛选预测框,得到最终识别目标。

1.2 反光特征融合模块 RFF

在采集金属铭牌压印字符图像的过程中,由于现场环境光照条件差、存在把手遮挡、铭牌摆放位置与光线角度变化等原因,迫使在采集压印字符铭牌数据时大多会受到较为明显的反光噪声影响。图 4 展示了在成都铁路局某段火车大线检修处真实场景下采集到的图像,可以看到想要获得一张完整的铭牌图像,只能从两个角度分别采集,然后通过仿射变换等一系列操作拼接成一张图像,同时图像中的字符与背景过于相近不易辨认。这些难点导致现有的检测模型对该类压印字符的检测效果较差,因此本节参考文献[8],期望通过反光特征融合模块 RFF,尽可能的减少因金属铭牌压印字符对比度低以及反光噪声带来的影响,最终提高目标检测模型的检测效果。



图 4 火车大线检修处采集的压印字符铭牌图像

为解决上述干扰信息对金属铭牌压印字符识别的影响,采用反光特征融合模块使反光噪声转换为携带着字符特征信息的反光特征并加以利用,因此,本文按照如图 5 所示的采集方法进行图像采集,相机位于铭牌的正上方,铭牌放置于 O 点;在保持高度 h 一致的情况下,拍摄光源在同一圆周上的 $\theta = 0^\circ$ 、 $\theta = 90^\circ$ 、 $\theta = 180^\circ$ 、 $\theta = 270^\circ$ 4 个位置的图像。图 6 为采集到的图像示例。反光特征的提取如图 7 所示。

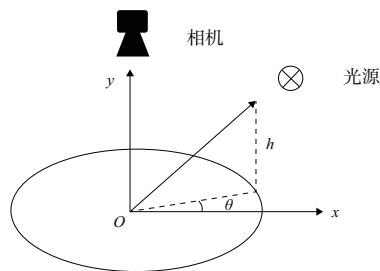


图 5 图像采集示意图

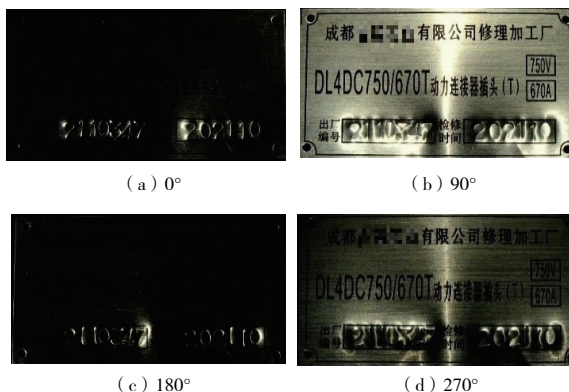


图 6 不同角度采集的图像

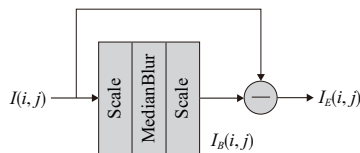


图 7 反光特征的提取

反光特征提取之后,对采集到的图像数据集通过以下步骤提取和消除背景信息:1)使用线性插值算法来减小原始图像的大小;2)使用中值滤波算法对缩小后的图像进行模糊处理,以消除表面上的压印字符,从而获得近似的背景图像,其中保留了图像的整体亮度分布;3)使用线性插值算法将模糊图像放大回原始图像大小,并用作原始图像的背景;4)设置图像的参考灰度值 128,通过从背景图像的灰度值中减去原始图像的灰度值来消除

背景。可以通过以下两个公式来描述上述过程:

$$I_B(i,j) = \text{Scale}(\text{MedianBlur}(\text{Scale}(I(i,j), 1/n), n) \quad (1)$$

$$I_E(i,j) = 128 + (I(i,j) - I_B(i,j)) \quad (2)$$

式中: $\text{Scale}(\cdot)$ 表示对图像进行线性插值运算; n 表示放大因子; $1/n$ 表示缩小因子, 根据样本的压印字符宽度, 将 n 设置为 10; $\text{MedianBlur}(\cdot)$ 表示对图像进行中值滤波; $I(i,j)$ 表示原始灰度图像; $I_B(i,j)$ 表示灰度背景图像; $I_E(i,j)$ 表示提取到的反光特征图像; i 和 j 表示图像的像素坐标。图 8(a) 是光源在 $\theta=0^\circ$ 位置处采集到的原图, 图 8(b) 经过灰度化和按上式提取到的特征图。

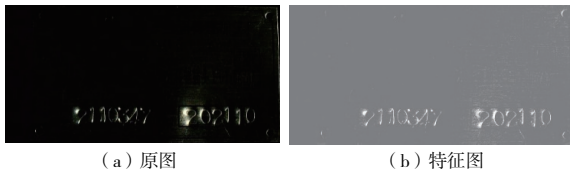


图 8 反光特征图像提取

提取到 4 张反光特征图像后, 采用 RFF 模块进行反光特征融合, 如图 9 所示。

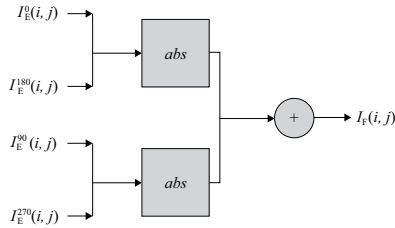


图 9 反光特征的融合

具体实现如下式:

$$V_1(i,j) = \text{abs}(I_E^0(i,j) - I_E^{180}(i,j)) \quad (3)$$

$$V_2(i,j) = \text{abs}(I_E^{90}(i,j) - I_E^{270}(i,j)) \quad (4)$$

$$I_F(i,j) = V_1(i,j) + V_2(i,j) \quad (5)$$

式中: $I_E^0(i,j)$ 、 $I_E^{90}(i,j)$ 、 $I_E^{180}(i,j)$ 、 $I_E^{270}(i,j)$ 分别表示 4 个方向上的反光特征图像; $V_1(i,j)$ 表示图像 $I_E^0(i,j)$ 与图像 $I_E^{180}(i,j)$ 的差异, 即融合了两个方向上的反光特征的图像, $V_2(i,j)$ 同理; $I_F(i,j)$ 表示最终反光特征融合图像, 如图 10 所示。



图 10 反光特征融合图像

1.3 C3CA 注意力机制混合模块

在网络结构中添加了反光特征融合模块后, 本文考虑在网络结构中添加注意力机制, 过去几年的轻量化网络模型研究中, 很多研究体现出来注意力机制会给模型性能带来显著性的提高。注意力机制能够提高深度学习算法的学习能力、泛化能力以及鲁棒性, 如 SE^[10]、CA^[11]、CBAM 等^[12]。SE 网络通过给各特征图分配不同的权重, 关注更有用的特征, 但只关注了通道信息而忽略了空间信息。CBAM 则是在通道注意力的基础上增添了空间注意力通道, 但只考虑了局部区域信息。而 CA 注意力是通过水平方向和垂直方向上分别进行平均池化, 再使用转换器对空间信息进行编码, 最后把空间信息通过加权的方式融合进通道中, 这样就实现了 CA 注意力机制对空间信息和通道信息的全面考虑。

因此, 本文选择将 CA 注意力模块 (如图 11 所示) 与 CSP 模块相结合得到 C3CA 注意力机制混合模块, 提高了网络的检测精度和定位精度。

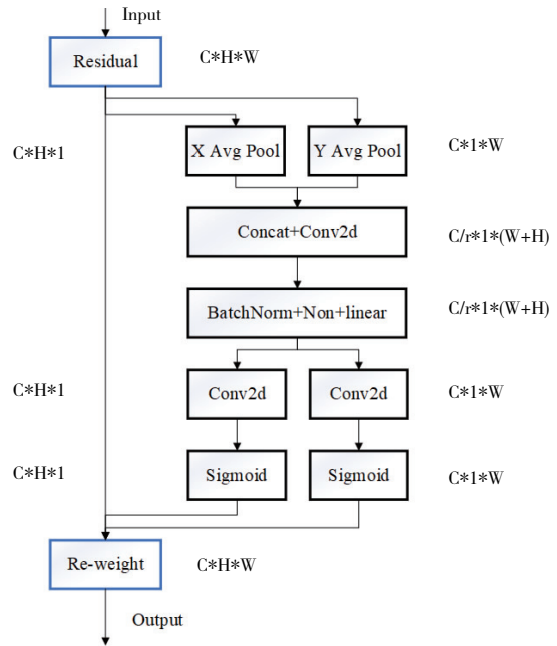


图 11 CA 注意力机制网络结构

CA 模块分两步编码通道关系和空间关系。一个步骤是坐标信息嵌入, 另一个是坐标注意力生成。具体过程如下。

1) 首先对宽度和高度特征图进行编码, 即在输入特征图的宽度和高度方向上分别进行全局平均池化处理。高度和宽度特征映射计算公式为

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (6)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (7)$$

式中: H 表示该层特征图的高度; W 表示该层特征图的宽度; $Z_c^h(h)$ 表示沿 H 方向的第 c 个通道的输出结果; $Z_c^w(w)$ 表示沿 W 方向的第 c 个通道的输出结果; $x_c(h, i)$ 表示输入特征图沿 H 方向的输入; $x_c(j, w)$ 表示输入特征图沿 W 方向的输入。

2) 将具有嵌入的特定方向信息的这两个特征图分别编码为 2 个注意力图, 该过程为坐标注意力生成。对应产生的位置信息会被保存起来, 存放在注意力图内。位置信息是指特征图沿 H 方向提取的信息和沿 W 方向提取的信息。

$$f = \delta(F_1([Z^h, Z^w])) \quad (8)$$

式中: $[,]$ 表示沿空间维度的拼接操作; F_1 表示卷积操作; δ 表示 Sigmoid 激活函数; $f \in R^{C/r * (H+W)}$ 表示在水平方向和垂直方向编码空间信息的中间特征图; r 表示控制注意力模块大小的超参。

3) 对原始特征在宽度和高度方向分别进行 1×1 卷积运算 (F_h, F_w) 和激活函数运算 σ 处理:

$$g^h = \sigma(F_h([f^h])) \quad (9)$$

$$g^w = \sigma(F_w([f^w])) \quad (10)$$

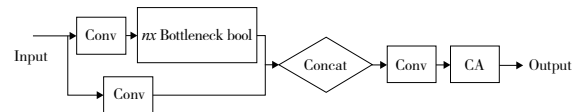
式中: f^h 和 f^w 分别表示 f 沿 H 和 W 两个方向的特征图; g^h 和 g^w 表示与输入特征具有相同通道数的注意力权重。

4) 最后将 CA 模块的输入特征图与具有宽度和高度的注意权值的特征图相乘, 得到最终特征图。计算公式为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (11)$$

式中: $y_c(i, j)$ 表示第 c 个通道的输出; $x_c(i, j)$ 表示第 c 个通道的输入; $g_c^h(i)$ 表示第 c 个通道上沿 H 方向的注意力权重; $g_c^w(j)$ 表示第 c 个通道上沿 W 方向的注意力权重。

本文将 CA 注意力机制融入 C3 模块中, 混合模块结构如图 12 所示, 首先, 输入的特征图信息进入 C3 结构中, 经过 C3 的 Conv 混合层处理后, 残差网络会汇总卷积层 (1×1) 和卷积层 (3×3) 的结果, 然后将上述混合层的结果与残差网络的结果相连接并再一次经过混合层处理, 最终送入 CA 注意模块。



C3CA 混合模块

图 12 C3CA 模块结构

通过在 YOLOv5 网络中的每一个 C3 模块后添加 CA 注意模块, 实现在每一层的卷积特征图中都能够将字符位置信息嵌入到通道注意中, 保留图像中压印字符的位置信息, 并将其纳入接下来的卷积操作中, 使所需检测目标位置信息贯穿整个网络模型之中, 增强模型对压印字符的关注能力, 使其能够将大部分计算资源分配到目标位置, 最终提高因图像采集条件差所导致的模糊字符图像的检测效果。

2 实验结果与分析

2.1 实验配置

本文使用的数据集是采用大恒工业相机, 在实验室中采集而来, 相机型号为 MER2-302-37GM/C。为验证 RFF-YOLOv5 算法的检测性能, 本研究使用同一样本数据集对 SSD、Faster R-CNN、YOLOv7 和 YOLOv5 算法进行了训练。

网络超参数配置如下: 在模型中设置目标的类别置信度阈值为 0.5, 初始学习率为 0.01, 权重衰减系数设置为 0.000 5, batch size 设置为 16, 1 000 个 epoch。具体测试环境: GPU 为 NVIDIA GeForce GTX 1060, 显存为 6 GB, CUDA 版本 11.0, 集成开发环境是 Pycharm, 编程语言是 Anaconda Python 3.7。

通过采集 232 张火车大线铭牌在 4 个方向照明下的图像, 总共有 928 张图像。在训练时按照 8:2 的比例, 将数据集随机划分为包含 744 张图片的训练集和 184 张图片的测试集。

2.2 评价指标

为更好地评估模型的综合性能, 采用精确度 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP) 作为字符识别结果的评价指标。如表 1 所示, 模型检测结果的定义主要分为真阳性 (true positive, TP)、假阳性 (false positive, FP)、真阴性 (true negative, TN)、假阴性 (false negative, FN) 4 种情况。

表 1 模型对不同检测结果的定义

	正类	非正类
预测为正类	真阳性 (TP)	假阳性 (FP)
预测为非正类	真阴性 (FN)	假阴性 (TN)

精确度 P 和召回率 R 的计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \quad (12)$$

$$R=\frac{T_p}{T_p+F_N} \quad (13)$$

式中: T_p 表示正类被正确预测的部分; F_p 表示非正类被预测为正类的部分; F_N 表示正类被预测错误的部分,即为漏检的部分。

mAP 的计算公式如下:

$$\text{mAP}=\frac{\sum_{i=1}^k A_{p_i}}{k} \quad (14)$$

式中: k 表示数据集中检测的目标类别个数; A_p 表示不同召回率下精确度的线下面积;mAP 即为各类 A_p 值的平均值。

2.3 对比实验

为验证本文模型在压印字符目标检测的性能,本文采用主流的目标检测模型进行对比,包括 SSD^[13]、Faster R-CNN^[14]、YOLOv7^[15] 和 YOLOv5 模型,实验结果如表 2 所示。由表 2 可知,本文提出的 RFF-YOLOv5-C3CA 模型表现良好, P 值为 95.23%, R 值为 81.73%,mAP 值为 89.31%,检测速度 FPS 为 41。改进后的模型优于 SSD、Faster R-CNN、YOLOv7、以及 YOLOv5。此外,RFF-YOLOv5-C3CA 的检测速度 FPS 是 YOLOv7 的两倍多,满足实时检测的要求。

表 2 不同模型的实验结果对比

模型	P/%	R/%	mAP/%	FPS/(f/s)
SSD	73.22	74.80	74.12	28
Faster R-CNN	70.37	79.65	77.10	8
YOLOv7	93.35	83.28	89.16	16
YOLOv5	91.77	80.32	87.59	52
RFF-YOLOv5-C3CA	95.23	81.73	89.31	41

为了定性比较改进模型的识别效果,在图 13 中展示了部分数据的识别结果,其中第 1 行为原图,第 2 行为 YOLOv5 的检测结果,第 3 行为 RFF-YOLOv5-C3CA 的检测结果。可以发现,所提出的 RFF-YOLOv5-C3CA 可以检测不同光照角度情况下的压印字符(反光现象较为明显的目标),相较于原始的 YOLOv5 模型,本文模型能够更好的检测到位于反光区域的压印字符。特别地,本文提出的 RFF-YOLOv5-C3CA 算法能够检测出不易区分的“0”、“6”、“8”、“9”等字符,优于 YOLOv5 原始算法。主要原因是 C3CA 模块和 RFF 模块增强了其字符特征,从而提高了检测能力。



图 13 多目标检测结果比较

2.4 消融实验

本文提出采用 RFF 模块和 C3CA 模块来提高 YOLOv5 的检测性能。为测试所提不同模块对模型检测性能的影响,通过控制变量去掉各项改进,设置消融实验,实验结果如表 3 所示。

表 3 不同模块对网络性能的影响

RFF 模块	C3CA 混合模块	mAP/%	时间/ms
×	×	87.59	19.3
✓	×	88.60	23.1
×	✓	88.20	22.0
✓	✓	89.31	24.6

从表 3 的结果中可以看出,在基础网络 YOLOv5 中加入反光特征融合模块 RFF, mAP 值提升了 1.01 个百分点,表明 RFF 模块通过融合不同角度照明下的字符特征,增强了反光情况下压印字符的特征表达能力,能有效的减少复杂环境下对识别压印字符的影响,其有助于提高模型学习能力并提高了检测性能。然后加入 C3CA 混合模块, mAP 值进一步提升了 0.6 个百分点,这表明 C3CA 混合模块通过利用和保留有效的通道信息和位置信息,能对模型产生积极的影响。

此外,为进一步展示 RFF 模块的有效性,通过图 14 展示了训练和验证的结果。从 RFF-YOLOv5-C3CA 模型的压印字符识别结果可以看到,字符“6”与字符“9”的 AP 值最低,分别为 0.693 和 0.802,字符“5”的检测率为 0.839,字符“0”识别率高达 0.995,其余字符识别率均超过 85%。

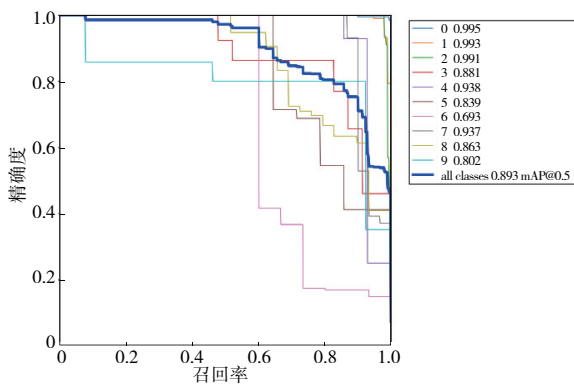


图 14 RFF-YOLOv5-C3CA 模型 PR 曲线

3 结语

本文提出了一种融入反光特征的金属压印字符识别深度学习模型,通过添加反光特征融合模块 RFF 增强压印字符特征信息,将反光噪声转变为可提取的反光特征,改善了因光照对金属铭牌压印字符识别结果的影响;通过引入 C3CA 注意力机制混合模块,来提高网络的检测精度和定位精度。实验结果表明:与原始的 YOLOv5 网络模型相比,本文提出的 RFF-YOLOv5-C3CA 模型对金属铭牌压印字符识别的 mAP 值提高了 1.72 个百分点,具有更强的特征提取能力。此外,与多种目标检测模型相比,本文模型对金属铭牌压印字符识别效果均有一定程度的优势。实验结果证明了本文模型的有效性。在未来,拟在本文提出的网络结构上作进一步优化,针对不易识别到的字

符和识别错误的字符问题进行探索,不断提高压印字符识别精度。

参考文献:

- [1] ZHANG J, BI H L, CHEN Y J, et al. SmartHandwriting: handwritten Chinese character recognition with smartwatch[J]. IEEE Internet of Things Journal, 2019, 7(2): 960-970.
- [2] 苗相彬, 罗建桥, 李柏林. 基于秩正则几何变换的无监督字符缺陷检测[J]. 机械制造与自动化, 2023, 52(5): 127-133.
- [3] RYAN M, HANAFIAH N. An examination of character recognition on ID card using template matching approach[J]. Procedia Computer Science, 2015, 59: 520-529.
- [4] 张传果, 刘建群. 柱面压印字符识别算法研究和实现[J]. 机床与液压, 2018, 46(13): 37-41.
- [5] 高潮, 常永鑫, 郭永彩. 机械工件数字识别的二值化算法研究[J]. 光电工程, 2010, 37(6): 1-5.
- [6] 汪志成, 何坚强, 翁嘉鑫, 等. 基于改进 LeNet-5 的压印字符识别[J]. 计算机仿真, 2022, 39(2): 441-446.
- [7] 李建美, 路长厚, 李国平. 基于 Gabor 变换的凹凸字符图像特征抽取新方法[J]. 系统仿真学报, 2008, 20(8): 2133-2136.
- [8] 向忠, 张建峰, 钱森, 等. 多方向照明下金属冲压字符图像融合算法[J]. 计算机辅助设计与图形学学报, 2018, 30(9): 1637-1642.
- [9] WEN G, LI S, LIU F, et al. YOLOv5s-CA: a modified YOLOv5s network with coordinate attention for underwater target detection[J]. Sensors (Basel), 2023, 23(7): 3367.
- [10] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.
- [11] TANG H Y, LIANG S, et al. A visual defect detection for optics lens based on the YOLOv5-C3CA-SPFP network model [J]. Optics Express, 2023, 31(2): 2628.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]// Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [14] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149.
- [15] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.[J]. 2022, DOI:10.48550/arXiv.2207.02696.

收稿日期: 2024-01-25