

改进 YOLOv5 算法的钢筋端面检测^①



王超¹, 张运楚^{1,2}, 孙绍涵¹, 张汉元^{1,2}

¹(山东建筑大学信息与电气工程学院, 济南 250101)

²(山东省智能建筑技术重点实验室, 济南 250101)

通信作者: 张运楚, E-mail: yczhang@sdjzu.edu.cn

摘要: 钢筋是基建行业不可或缺的结构材料, 无论是钢筋生产过程、还是施工现场, 对钢筋进行准确计数是必不可少的环节。成捆钢筋存在端面密集、直径尺度不统一、端面边界粘连、端面与背景融合、端面之间存在遮挡等问题。针对上述问题, 提出了一种改进的 YOLOv5 模型框架, 以降低密集小目标漏检率、误检率。针对钢筋端面数据集稀缺、没有公开的大型数据集并且钢筋端面特征较弱的问题, 自建了钢筋端面数据集, 使用半自动标注法对数据集进行标注, 并采用数据增强算法对钢筋端面数据集进行扩充。修改了 YOLOv5 中的主干网络, 增加空间金字塔池 (spatial pyramid pooling, SPP) 和小目标检测层, 以获取更大的特征图; 使用特征金字塔模型 (feature pyramid network, FPN) 和路径聚合网络 (path aggregation network, PAN) 对多尺度特征图融合, 提高密集小目标检测精度。在 Data Fountain 钢筋盘点竞赛数据集和自建钢筋数据集上设计了多组对照试验。实验结果表明, 提出的改进算法 YOLOv5-P2 模型对钢筋端面的检测效果最佳, 钢筋端面平均精度均值 (mean average precision, *mAP*) 达到了 99.9%, 相比于 YOLOv3、YOLOv4、ScaledYOLOv4 以及 YOLOv5 主流算法, 模型的 *mAP* 分别提升了 9.6%、7.9%、7.0%、1.1%, 在工厂真实环境条件下进行测试时都有较稳定的表现, 在测试集上相对于原始模型检测精度提升了 2.1%。通过修改 YOLOv5 的主干网络中 SPP 模块位置和增加检测层都能够显著提升密集小目标检测精度, 更好的提取到钢筋端面的边缘特征, 取得 99.9% 的平均精度均值。

关键词: 深度学习; YOLOv5; 数据增强; 钢筋计数; 密集小目标检测; 检测方法; 卷积神经网络

引用格式: 王超, 张运楚, 孙绍涵, 张汉元. 改进 YOLOv5 算法的钢筋端面检测. 计算机系统应用, 2022, 31(4): 68-80. <http://www.c-s-a.org.cn/1003-3254/8433.html>

Steel-bar End Face Detection Based on Improved YOLOv5 Algorithm

WANG Chao¹, ZHANG Yun-Chu^{1,2}, SUN Shao-Han¹, ZHANG Han-Yuan^{1,2}

¹(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

²(Shandong Key Laboratory of Intelligent Buildings Technology, Jinan 250101, China)

Abstract: A steel bar is an indispensable structural material in the infrastructure industry, and accurate counting of steel bars is an essential link in both the steel-bar production process and the construction site. There are some problems in steel-bar bundles, such as dense end faces, non-uniform diameter scale, end-face boundary adhesion, fusion of end face and background, and end-face occlusion. To solve the above problems, this study proposes an improved YOLOv5 model framework to reduce the missed detection rate and the false detection rate of dense small targets. Considering the scarcity of the steel-bar end face dataset, the absence of a large public dataset in this field, and the weak feature of the steel-bar end face, we built a steel-bar end face dataset with the semi-automatic labeling method for dataset labeling and the data enhancement algorithm for dataset expansion. Moreover, the backbone network in YOLOv5 was modified, and the spatial pyramid pooling (SPP) and the small target detection layer were added to obtain larger feature maps. The feature pyramid

① 基金项目: 国家自然科学基金 (62003191)

收稿时间: 2021-06-27; 修改时间: 2021-07-02; 采用时间: 2021-08-13; csa 在线出版时间: 2022-03-22

network (FPN) and path aggregation network (PAN) were used to fuse multi-scale feature images to improve the accuracy of dense small target detection. Several groups of control tests were designed based on the Data Fountain steel-bar stocktaking competition dataset and the self-built steel bar dataset. The experimental results show that the improved algorithm YOLOv5-P2 model has the best performance on the steel-bar end face detection, and the mean average precision (*mAP*) of the steel-bar end face reaches 99.9%. Compared with the mainstream algorithms of YOLOv3, YOLOv4, ScaledYOLOv4, and YOLOv5, the proposed model has its *mAP* increased by 9.6%, 7.9%, 7.0%, and 1.1%, respectively. When tested in the real environment of factories, the model has stable performance, and its detection accuracy is improved by 2.1% compared with the original model on the test dataset. The position modification of the SPP module in the backbone network of YOLOv5 and the adding of detection layers can all significantly improve the detection accuracy of dense small targets with better edge feature extraction of the steel-bar end face and an *mAP* of 99.9%.

Key words: deep learning; YOLOv5; data enhancement; rebar counting; dense small target detection; detection method; convolutional neural network

钢筋作为不可或缺的结构材料,广泛应用于建筑、桥梁、交通等基建行业。无论是钢筋生产过程、还是施工现场,对钢筋进行准确计数是必不可少的环节。

在钢筋生产过程中,经轧制、冷却、剪切生产的散状钢筋,在传送链床上进行计数成捆、收集打捆、卸料,然后对每捆钢筋数量进行复核计数、称重,最后挂牌入库。散状钢筋常采用视频动态计数,通过融合多帧视频图像中钢筋端面检测和跟踪结果来实现。成捆钢筋的复核计数,则通过检测单帧图像中钢筋端面的数量来完成,准确度要求高,以捆为单位不低于98%,同时检测时间应小于100 ms/捆。在施工现场,需对进场成捆钢筋的数量进行复验,目前多采用人工计数,劳动强度大、工作效率低、出错率高,无法满足智慧工地的管理要求。

视觉成捆钢筋计数技术的关键,在于对钢筋端面的检测与分割,主要有基于传统图像处理的钢筋计数方法和基于深度学习的钢筋计数方法。

基于传统图像处理的钢筋计数方法主要利用图像增强、滤波、阈值分割进行钢筋端面检测和计数^[1],或者对预处理后的图像进行模板覆盖以及空间相关性计算对钢筋端面进行检测计数^[2]。罗山等人^[3]利用二值化后图像的面积法对成捆钢筋的根数进行统计,臧晶等人^[4]首先利用双边滤波算法去噪声。然后采用最大类间方差(Otsu)算法对钢筋端面分割,再使用连通区域法对钢筋进行标记。袁中锦等人^[5]运用直方图均衡化增强对比度后,再用高斯模板计算空间相关性,使钢筋截面特征更加显著,通过统计相关矩阵中的二维局部极大值点个数来得到最终计数。

近年来,基于深度学习的卷积神经网络(convolu-

tional neural networks, CNN)在目标检测领域得到了快速发展,如YOLOv3^[6]、YOLOv4^[7]、YOLOv5、SSD^[8]等单步目标检测算法和Fast R-CNN^[9]、Faster R-CNN^[10]等两步目标检测算法。

石京磊^[11]利用卷积神经网络与红外相机对钢筋进行静态与动态检测,通过改进ResNeXt101网络获得了一种级联式目标候选头网络对成捆钢筋端面进行静态检测和计数。Wu等人^[12]通过深度学习语义分割来实现钢筋计数,对弱监督标注的数据集训练语义分割,主干网络采用改进的U-net预测3层不同的mask,推断得到语义分割二值图像,再使用轮廓提取算法找到各个钢筋端面。谢海桢^[13]提出了基于全卷积网络的密集多目标识别算法MSFCN,在自建数据集上准确率和召回率均达到95%以上,相较传统算法精准度高出10%。唐楚柳^[14]提出了一种由Inception-RFB-FPN和改进ResNet34-FCN组成的SWDA-CNN网络用于钢筋计数,实验表明该网络提高了定位精度并节省推理时间。明洪宇等人^[15]提出了一种基于RetinaNet目标检测框架,采用基于EM算法的高斯混合聚类方法解决检测歧义,改进后的模型*mAP*值为93%。王志丹^[16]基于Faster R-CNN算法利用微调后的预训练模型对增强后的数据集进行训练,得到98.3%的*mAP*值,较原始Faster R-CNN提高了2.1%,但该模型的计算复杂度较高,在实际应用中难以达到实时检测。

由于剪切的钢筋端面颜色接近,打捆后挤在一起,彼此之间界线不显著,且端面参差不齐、面积大小不一、形状不很规则等原因,基于传统图像处理的钢筋计数方法准确度和鲁棒性难以提高。基于深度学习的单

步目标检测算法对密集小目标的漏检率较高,两步目标检测算法的实时性差。

本研究主要应用于两大场景,第一是钢筋生产企业入库复检,仓库的环境通常是一个整体偏暗的环境,需要灯带进行打光,除了需要打光之外在某些比较苛刻的工况下,需要定制固定的工位来进行检测,真实场景如自建数据集所示;第二是建筑工地计数,卡车运载着大量的钢筋进场,现场的验收人员需要对其进行盘点,该场景中拍照角度存在偏移,现场光照的不均,拍摄距离的远近不同导致钢筋面存在尺度不一致,真实场景如公共数据集所示。

本文提出一种基于改进YOLOv5的钢筋端面识别计数算法,主要贡献如下:

(1) 提出一种半自动数据标注方法,在公开数据集DataFountain钢筋盘点竞赛数据集上训练YOLOv5初始模型,对工厂环境下的自建数据集进行检测标注,自动生成标注文件,然后进行手工校正,大大提升数据集的标注效率。

(2) 通过增加采样融合扩大特征图的尺度、增加模型检测层个数、改变SPP模块在主干网络中的位置,获取更充足的端面纹理特征,丰富端面间轮廓信息,提升端面识别率。

(3) 在公开数据集与自建数据集上与多个主流检测算法进行消融实验(ablation experiment),改进后的算法平均精度均值*mAP*达99.9%,高于所有其他模型,比原始YOLOv5模型提升了1.1%,精准度*Precision*达95%,召回率*Recall*达99.8%。

1 YOLOv5的改进及钢筋端面检测

本节对成捆钢筋端面检测的难点进行了总结并简要阐述了单阶段YOLO系列检测算法框架,对YOLOv5初步实验后存在的问题进行分析,并提出改进思路。

1.1 成捆钢筋端面的特点及检测难点

成捆钢筋的端面一种密集型小目标,在检测时具有目标类别单一、目标分布较为集中、彼此区分度低等特点。采用深度学习模型对成捆钢筋端面进行检测计数会面临以下难点:

(1) 钢筋端面尺寸不一。钢筋端面直径视规格型号差距较大,本文涉及钢筋端面直径在12–32 mm之间。

(2) 每捆钢筋数量差异较大,视钢筋规格型号有95、290、380根。

(3) 应用场景复杂。拍摄的距离、角度和光线不完

全受控,切割端面形状差异较大、端面锈蚀颜色和亮度差异较大且存在边界粘连,部分钢筋端面窜出或缩进造成遮挡等问题。

(4) 计数精度和准确度要求高。钢筋产量和使用基数很大,如果检测结果中出现了误检和漏检,需要人工从大量的标记点中找出。只有达到极高的计数精度和准确度,才能保证验收人员的使用体验。

1.2 YOLO系列目标检测框架

YOLO系列为单步目标检测框架,对输入图像直接进行分类概率回归和包围框坐标回归实现目标检测。从YOLOv3开始摒弃了池化层和全连接层,使用Darknet53代替Softmax对目标特征进行多尺度预测。YOLOv4是对YOLOv3的改进,网络结构主要包括主干特征提取网络CSPDarknet53、空间金字塔池化(SPP^[17])、路径聚合网络(PANet^[18]),同时使用了多种数据增强技术来提升检测性能,相较于YOLOv3模型*mAP*值提升10%。2020年Ultralytics发布了YOLOv5,其性能与YOLOv4相当但推理速度更快,且模型框架更便于工程部署。YOLOv5使用C3Darknet作为主干网络从输入图像中提取丰富的信息特征,使用PANet作为Neck来聚合特征,模型检测层与YOLOv3相同,Conv卷积模块的激活函数采用SiLU^[19]函数。此外,YOLOv5通过设置depth_multiple和width_multiple两个参数调节主干网络的深度和宽度并划分出4个量级的模型:YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。

1.3 YOLOv5算法与改进方式

由于钢筋数据集中钢筋端面分布密集、类内差异较大、钢筋端面半径小且不统一,因此采用YOLOv5x这个较深层次的网络作为原始检测模型。

1.3.1 K-means聚类锚点框

YOLOv5算法中,预测部分分为3个特征层,每个特征层含有3个大小不同的先验框,先验框是通过K-means聚类算法对数据集中标注检测目标聚类得到的。在聚类过程中,通过随机化初始*K*个聚类中心位置、计算目标标注框与聚类中心点的交并比、分配交并比最大的聚类中心、重新计算聚类中心、直到聚类中心不再发生改变。最终聚类好*K*个先验锚点框(anchor)的宽、高。具体的目标函数*D*为:

$$D = \min \sum_{i=1}^n \sum_{j=1}^k \left[1 - \frac{box_i \cap cen_j}{box_i \cup cen_j} \right] \quad (1)$$

式中, box_i 为检测目标中的第*i*个标注框的区域面积;

cen_j 是第 j 个聚类中心的区域面积; $box_i \cap cen_j$ 为标注框与聚类中心区域交集的面积; n 为检测目标的数量, k 为聚类中心个数.

聚类时以准确率 (*Accuracy*) 作为聚类结果的评价指标, 计算公式为:

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \max \left(\frac{box_i \cap cen_1}{box_i \cup cen_1}, \frac{box_i \cap cen_2}{box_i \cup cen_2}, \dots, \frac{box_i \cap cen_k}{box_i \cup cen_k} \right) \quad (2)$$

使用 K-means 算法对自建钢筋数据集的真实包围框进行聚类来确定性能最佳的 K 值, 结果由图 1 所示. 当 $K=12$ 时, $Accuracy=91.57$, 指标处于较好性能. 图 2 给出了 K 值分别为 6、9、12、15 时数据集的真实框聚类结果图.

1.3.2 改进 YOLOv5 算法

在自建数据集上对原始 YOLOv5 模型进行初步实验, 自建数据集的构建将在第 2 节介绍, 训练参数设置将在第 3.1 节给出, 模型评价指标详见第 1.4 节.

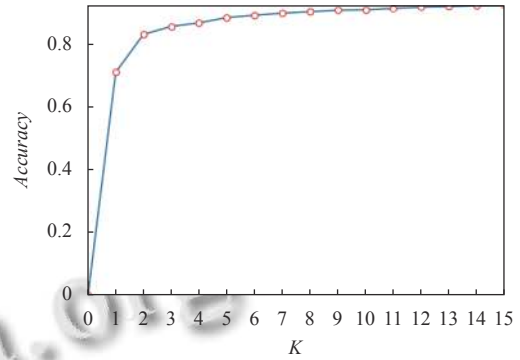


图 1 *Accuracy-K* 折线图

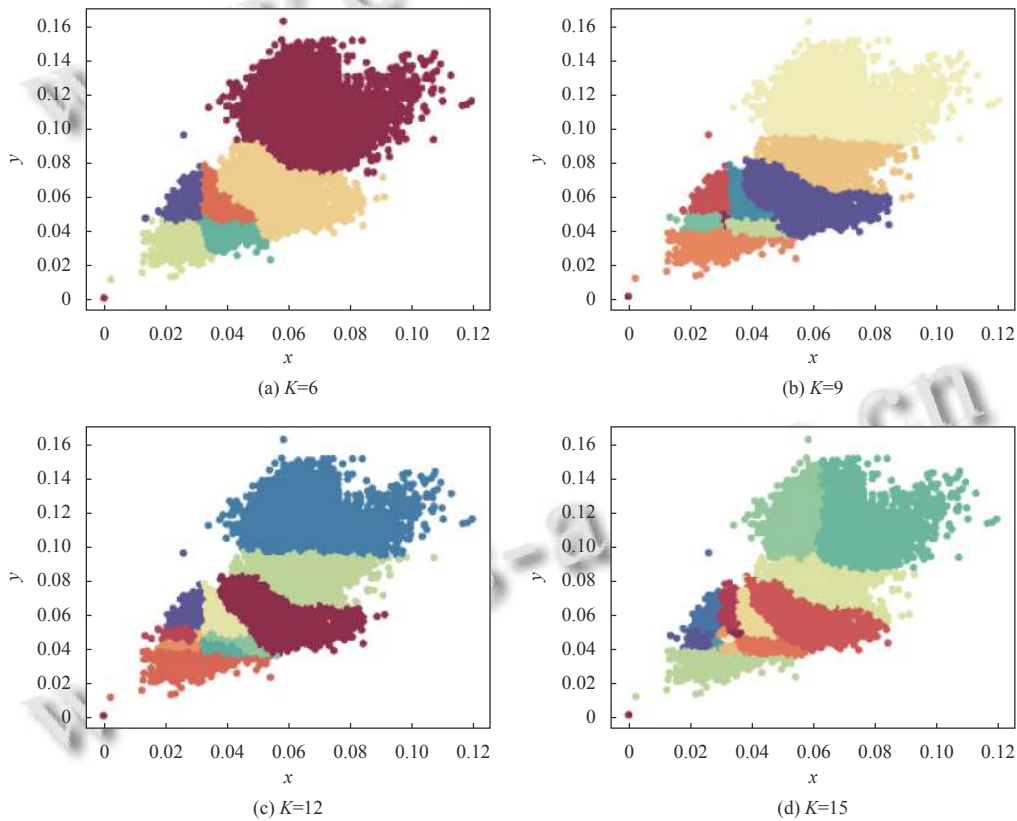


图 2 样本包围框 K-means 聚类可视化图

训练结果如图 3 所示, 由曲线图可以看到精准度只有 80% 左右, 当 *IOU* 值为 0.5 时的平均精度均值 (*mAP*) 为 98.8%. 对测试集进行检测的结果如图 4 和图 5 所示, 图中部分遮挡目标存在漏检和误检情况, 原因是 YOLOv5 模型只有 3 个检测层各对应 3 组初始化 anchor, 当输入图像尺寸为 640×640 时, 主干网络

C3Darknet 第 3 层对应的检测层大小为 80×80 , 此时该层对应的感受野最小, 只能用来检测大小在 8×8 以上的目标. 主干网络第 5 层对应的检测层的大小为 40×40 , 只能用来检测尺度在 16×16 以上的目标. 主干网络第 7 层对应的检测层的大小为 20×20 , 此时对应的感受野最大, 只能用来检测 32×32 以上的目标.

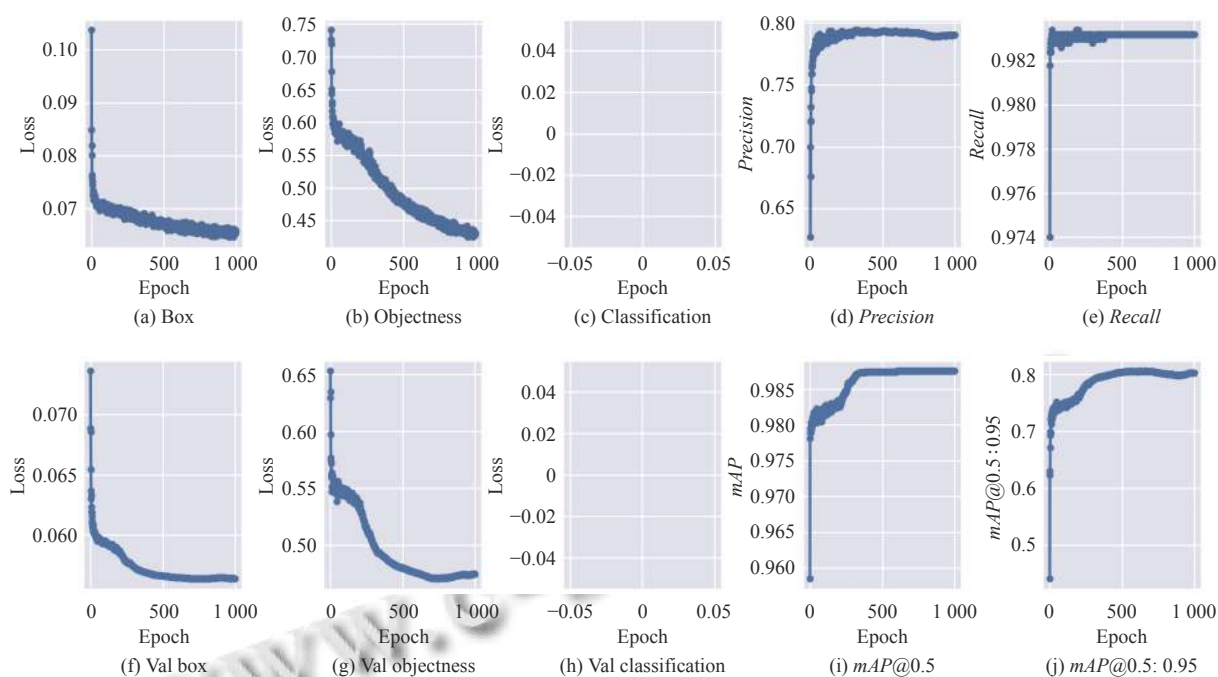


图3 原始YOLOv5模型训练曲线图

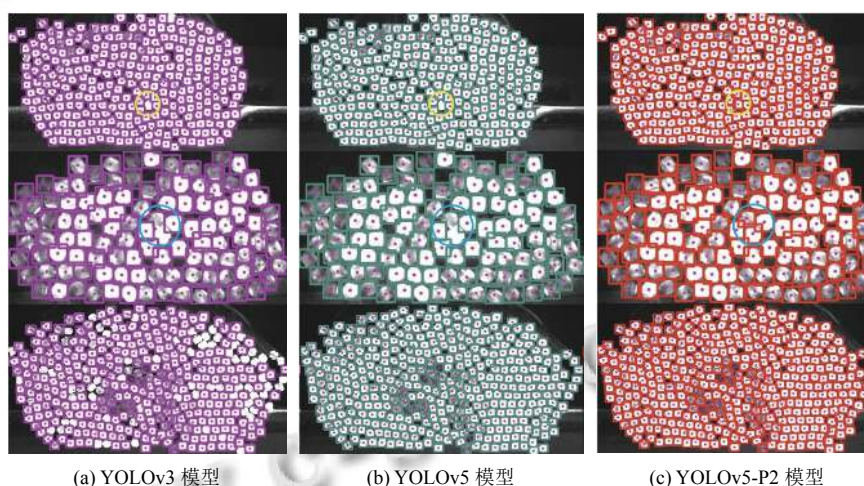


图4 模型自建数据集检测效果图

因此,当被检测目标尺度小于 8×8 时,无法被检测器准确检测到,为了避免这种情况,在Neck结构的第17层后,继续对特征图进行上采样等处理,使特征图继续扩大,同时在Neck第20层,将获取到的大小为 160×160 的特征图与主干网络第2层特征图进行Concat融合,以此获取更大的特征图进行小目标检测.改进后的YOLOv5-P2模型结构如图6所示,其中主干Backbone红色箭头加粗虚线是新增采样对应于Head层 160×160 预测头.由K-means算法对数据集聚类得到效果最佳的K值为12,当anchor的数量为12

时能够得到更精确的anchor的宽高值.所以网络模型改进思路的正确性得到初步验证.

为了找到适配数据的最佳感受野大小,另一种改进方法是把SPP模块当作变量,添加在主干网络的不同位置.本文在主干网络第6层与第7层之间加入SPP模块,增大感受野提取重要的特征,来提升钢筋端面小目标检测精度.SPP模块置于位置如图6加粗红色虚线处所示,得到改进的YOLOv5-SPP模型.

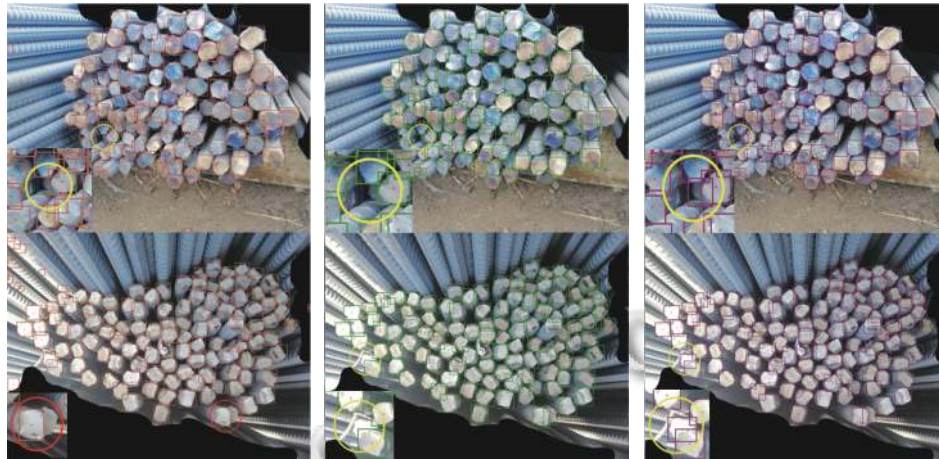
1.3.3 YOLOv5改进模型损失函数

损失函数反映了预测框与真实框之间的差值,是

衡量检测模型性能的重要指标。YOLOv5 模型中损失函数 $Loss$ 包括预边界损失函数 L_{box} 、置信度损失函数 L_{conf} 、分类损失函数 L_{cls} , 因此损失函数可以表示为:

$$Loss = \lambda_{box}L_{box} + \lambda_{conf}L_{conf} + \lambda_{cls}L_{cls} \quad (3)$$

式中, λ_{box} 、 λ_{conf} 、 λ_{cls} 分别是坐标损失权重, 置信度损失权重和分类损失权重。



(a) YOLOv3 模型 (b) YOLOv5 模型 (c) YOLOv5-P2 模型

图 5 Data Fountain 数据集检测效果图

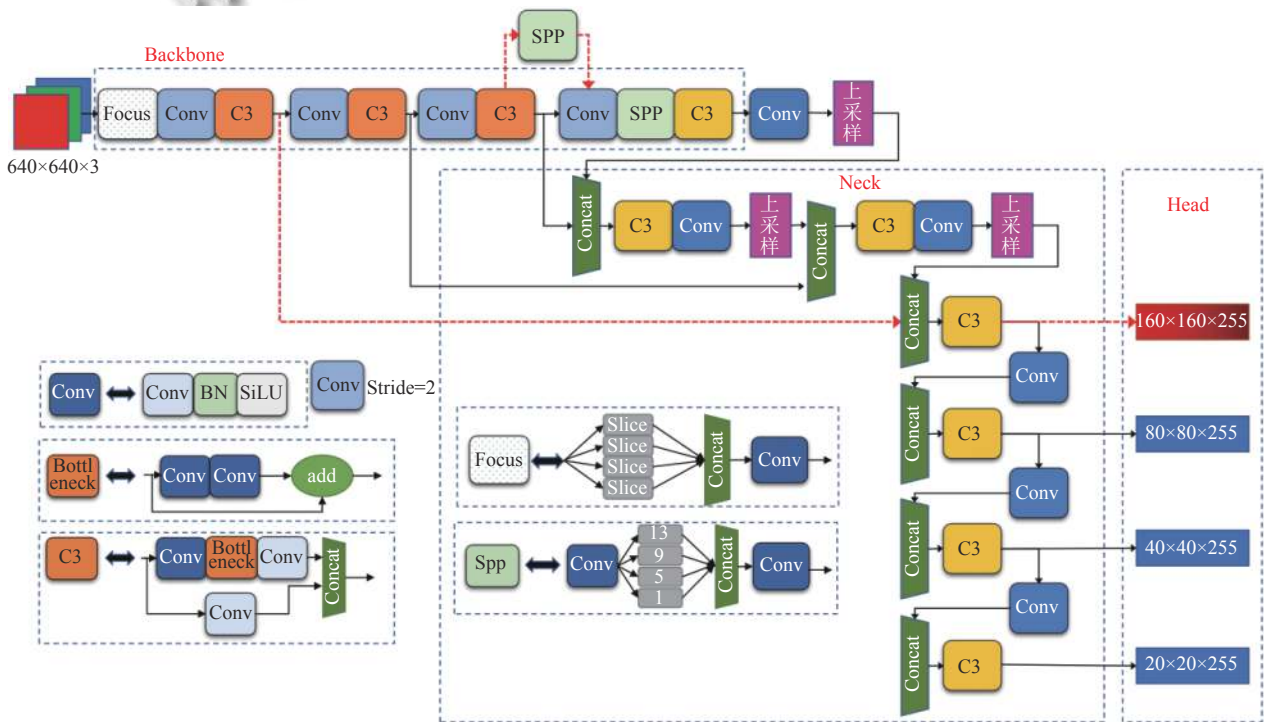


图 6 改进后的 YOLOv5-P2 模型网络结构图

$$L_{box} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[1 - IOU(b_i, \hat{b}_i) + \frac{\rho^2(b_i, \hat{b}_i)}{c^2} \right] \quad (4)$$

式中, I_{ij}^{obj} 表示包含目标的第 i 个单元的第 j 个锚点; b_i 和 \hat{b}_i 分别代表预测框和真实框; IOU ^[20] 表示预测框与真实框的交并比; ρ 是真实框与预测框中心点之间的欧

氏距离^[21]; c 是能包围预测框与真实框的最小框的对角线长度。

$$L_{conf} = \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C_i - \hat{C}_i)^2 \quad (5)$$

式中, C_i 表示预测框的置信度得分, \hat{C}_i 表示真实框的置信度得分; I_{ij}^{obj} , I_{ij}^{noobj} 分别代表第 i 个单元的第 j 个锚点。

分类损失函数采用二值交叉熵损失, 损失函数公式为:

$$L_{\text{cls}} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c))] \quad (6)$$

其中, c 是当前检测到的目标类别, classes 是所有目标类别; $\hat{p}_i(c)$ 和 $p_i(c)$ 分别是预测类别和真实类别。

1.4 检测模型评价指标

mAP (mean average precision) 是多个验证集的平均精度均值, 在目标检测任务中作为衡量检测精度的重要指标。 AP (average precision) 为平均精度, 是 P-R (precision-recall) 曲线与坐标轴围成的面积值。 P-R 曲线是以 $Recall$ 和 $Precision$ 作为横纵坐标的二维曲线。 $Precision$ 和 $Recall$ 的定义如式 (7) 和式 (8):

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

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

式中, TP 为正确划分为钢筋的数量, FP 为错误划分为钢筋的数量, FN 为被错误地划分为非钢筋的数量。

绘制 P-R 曲线来计算单个类的 AP 值, 再计算各类 AP 值的平均值, 得到整个模型的 mAP 值:

$$AP = \int_0^1 p(r) dr \quad (9)$$

$$mAP = \sum_{n=1}^C AP_n \quad (10)$$

式中, p 为准确率, r 为召回率, C 为类别总数。 本文是对钢筋的单类目标检测, 所以计算所得 AP 值就是 mAP 。

除了 mAP 外, 目标检测算法的另一个重要性能指标是速度。 衡量速度的一个常用指标是每秒帧数 (FPS), 即每秒可以处理的图片数量。

2 数据集的构造及增强

2.1 数据集的获取

本文使用的数据集由两部分组成, 一部分是由 Data Fountain 的智能盘点钢筋数量 AI 识别比赛标注好的数据集, 共有标注图片 250 张, 图片规格是 2666×2000; 另一部分是自建的企业私有数据集, 使用 HIKVISION 摄像机拍摄钢厂链条上成捆钢筋所得, 共有未标注的图片 958 张, 图片规格为 1600×1200。

2.2 数据集的半自动标注方式

钢筋数据集手工标注耗时费工, 为提高工作效率, 本文利用已经标注好的智能盘点钢筋数据集, 训练得

到初步模型, 然后使用该模型检测未标注的自建数据集, 生成 XML 标注文件, 再使用 LabelImg 数据集标注软件进行手工修正标注框, 以防止漏标和误标, 如图 7。

2.3 数据集增强

原始数据集共有 1208 张图像, 为避免模型训练产生过拟合, 本文采用数据增强扩充数据集。 通过水平/垂直镜像、亮度/色彩调节、随机裁剪等方式对数据集进行扩充, 并自动生成扩充后的 XML 标注文件, 最终的数据集被扩充为 3295 张图像。 图 8 给出了数据增强得到的样本及其自动标注示例。

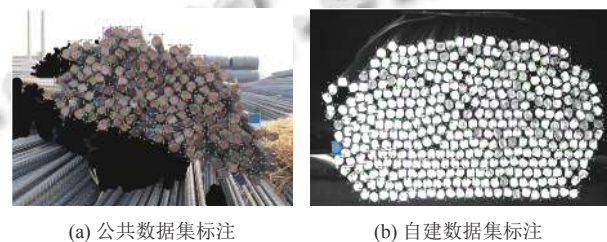


图 7 图像半自动标注示例

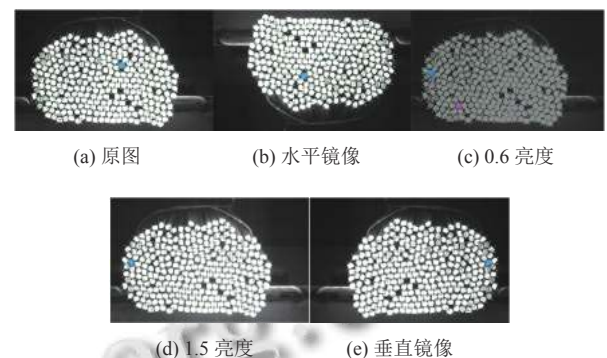


图 8 数据增强得到的样本及其自动标注

2.4 数据集划分

本文使用数据集中包含了日间室外逆光、顺光和夜晚室内灯带打光等场景。

将数据集以 8:1:1 比例划分训练集、验证集和测试集。 选取数据集中 2636 幅图像作为训练数据集, 选取未经过数据增强的图片 330 张作为测试集。 对划分的训练集真实目标分布情况进行可视化, 结果如图 9 所示。

3 实验及结果分析

本实验硬件配置使用 Intel Core i7-9800X CPU @3.80 GHz 处理器、NVIDIA GTX 2080 Ti 显卡、32 GB RAM、3 TB 机械硬盘, 操作系统是 Windows 10 64 位系统。 编程语言为 Python, 深度学习框架为 PyTorch, GPU 加速库为 CUDA10.2 和 CUDNN8.0。

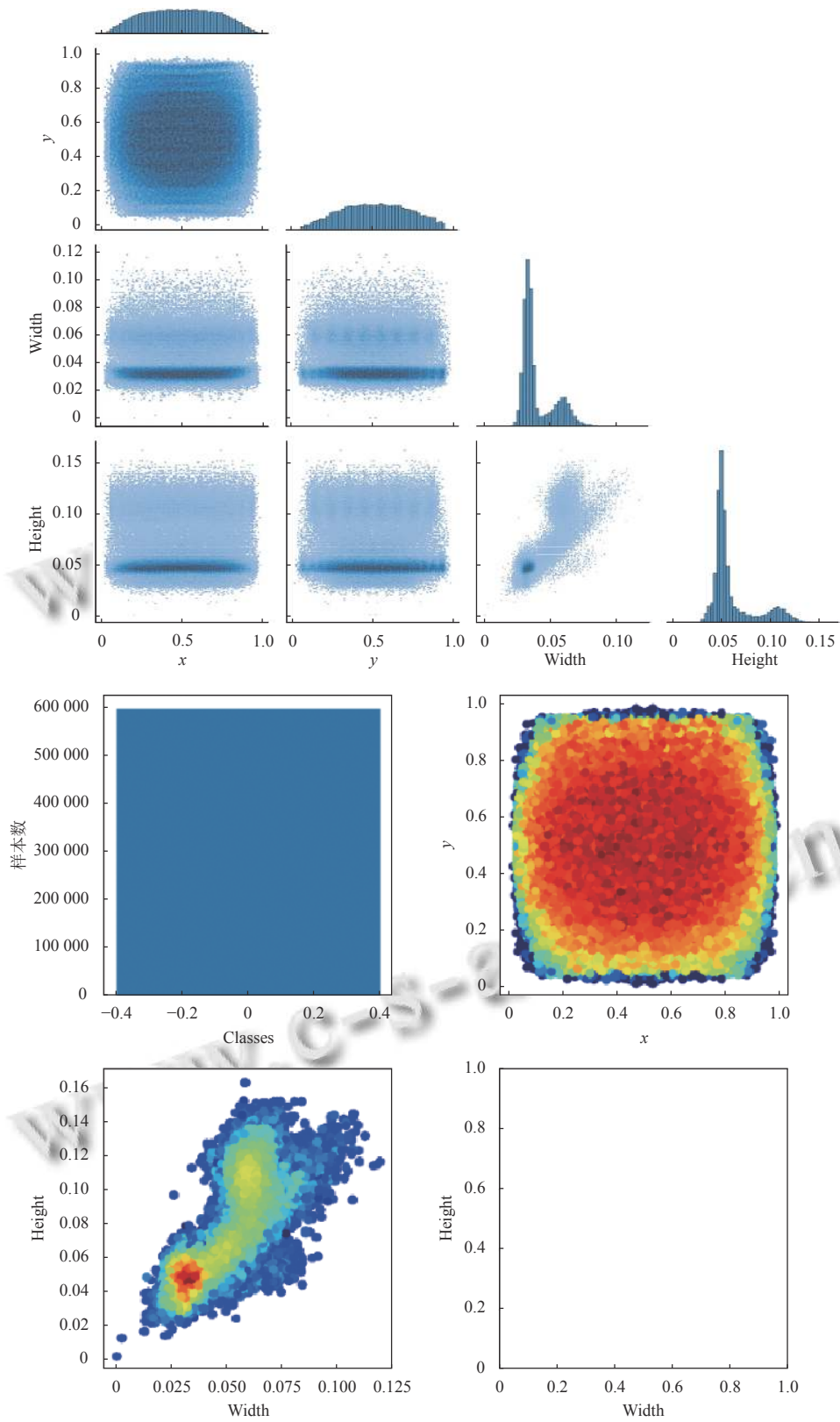


图9 训练集的目标分布统计图

3.1 模型训练

YOLOv5 钢筋端面检测模型使用的深度学习框架为 PyTorch. YOLOv5 算法训练的初始参数设定为: 优化器使用随机梯度下降 SGD^[22] 和余弦学习速率衰减策略来训练网络, 初始学习率 (learning rate) 为 0.01, 最终衰减速率为 0.001, 迭代次数 (epochs) 为 1000. 批处理大小 (batch size) 为 8, 动量因子为 0.937, 输入图像分辨率为 640×640, 设置损失权重 $\lambda_{\text{box}} = 0.05$ 、 $\lambda_{\text{conf}} = 1$ 、 $\lambda_{\text{cls}} = 0.5$ 、 $\lambda_{\text{obj}} = 1$ 、 $\lambda_{\text{noobj}} = 0.5$.

批量处理多张训练图片时, 同 YOLOv4 算法类似采用了 Mosaic^[23]、Cutout^[24] 等方法, 此外还使用了图像扰动, 改变亮度、对比度、饱和度、色调, 加噪声、随机缩放、随机裁剪 (random crop)、翻转、旋转、随机擦除等数据增强方式.

3.2 实验结果分析

本文通过在模型 Head 层修改检测层个数、增加采样融合扩大特征图的尺寸提升模型的性能. 首先, 为了改善钢筋端面密集小目标的检测效果, 在原始主干网络上增加 P2 采样层, 并在 head 层增加第 4 个输出层. 与原始模型相比, 改进后的 YOLOv5-P2 模型的 *mAP* 和 FPS 如表 1 所示.

表 1 模型添加小目标检测层的测试结果

钢筋检测模型	Aug.	K-means	输出层数	FPS	<i>mAP</i> (%)
YOLOv5x	√	√	3	35	98.8
YOLOv5-P2	√	√	4	28	99.9

注: 加粗字体为每列最优值, “■”表示未使用, “√”表示使用

由表 1 可知, 改进后模型在主干网络增加 P2 采样层、Head 输出增加为 4 层时, 模型的输出特征图增大, 在 Neck 层提取到更多轮廓特征, 提升了小目标检测精度, *mAP* 值由 98.8% 提升到 99.9%, *Precision* 由原始模型的 80% 改进后提升到 95%. 尽管检测速度有所降低, 仍能满足实际需要.

为了找到适配数据的最佳感受野大小, 本文在主干网络的第 6 层与第 7 层之间加入 SPP 模块, 得到 YOLOv5-SPP 模型, 增大感受野提取重要的特征, 来提升钢筋端面小目标检测精度. 对比结果如表 2 所示.

表 2 模型改变 SPP 模块位置的测试结果

钢筋检测模型	K-means	Backbone 6~7 SPP	FPS	<i>mAP</i> (%)
YOLOv5x	√	■	35	98.8
YOLOv5-SPP	√	√	33	99.8

注: 加粗字体为每列最优值, “■”表示未使用, “√”表示使用

本文的检测目标是钢筋端面. 在增强后的训练数据集上, 将检测的钢筋端面预测框的平均精度均值 *mAP* 作为检测效果的衡量标准.

为了能够更好的观察各个检测层的输出以及模型各检测层的特征提取效果, 对改进前后模型的检测层的特征图 (feature map) 进行了可视化. 如图 10 所示, 其中图 10(a) 为原图, 图 10(b)–图 10(d) 为原始模型的检测层可视化结果, 图 10(e)–图 10(h) 为改进模型检测层可视效果. 由改进前后的图 10(b)、图 10(e), 即小目标特征图可以看出改进后的图 10(e) 的可视化效果更佳, 与原始图像具有空间上的对应关系, 目标间的轮廓信息更丰富, 此时的感受野最小, 能够提升对小目标的检测效果.

本文除了将改进模型与原始 YOLOv5 模型进行对比外, 还与 YOLOv3、YOLOv3-SPP 以及 YOLOv4 和 ScaledYOLOv4^[25] 等主流实时目标检测算法进行消融实验对比, 结果如表 3 所示. 表中, Augmentation 代表是否进行数据增强, K-means 表示模型是否对锚点进行聚类, epoch 表示训练迭代的次数, 通过控制不同的变量来验证改进算法的效果.

未进行 K-means 聚类之前默认的锚点框为 [10, 13], [16, 30], [33, 23], [30, 61], [62, 45], [59, 119], [116, 90], [156, 198], [373, 326], 进行 K-means 聚类之后锚点框为 [18, 30], [18, 34], [19, 23], [19, 27], [20, 31], [22, 29], [22, 33], [32, 42], [39, 67], YOLOv5-P2 模型使用的锚点框为 [17, 31], [17, 22], [18, 27], [19, 32], [19, 30], [20, 24], [20, 36], [21, 28], [21, 31], [24, 33], [33, 43], [39, 67].

与目前主流的目标检测算法相对比, 本文改进的算法在 *Precision* 以及 *mAP* 均取得最优. 其原因如下: (1) 平滑的激活函数允许更好的信息深入神经网络, 从而得到更好的准确性和泛化. 相较于 YOLOv3, YOLOv4 模型, 本文模型使用了 SiLU 激活函数, 与 Lekey-ReLU 相比, SiLU 梯度更平滑, 保持准确性更好且更准取得地传播信息. (2) Neck 网络与 YOLOv4 模型相似本模型使用 SPP 与 PANet 结构. SPP 利用 4 个池化核大小分别为 13×13、9×9、5×5、1×1 进行处理, 该结构能分离出最显著的上下文特征, 是强有力的特征提取, 池化后再进行堆叠; PANet 结构是一种反复提取特征的实例分割算法, 过程包括上采样、再堆叠卷积重复, 之后再下采样、堆叠. 此外本文算法还采用 Mosaic 数据增强, 扩充了数据集, 均衡大中小目标

数量: 随机使用多张图片随机拼接, 且通过随机缩放可以获得很多小目标, 让网络的鲁棒性更好. 通过以上几

点对密集小目标检测效果有了很大提升, 这也证明了本文模型的训练效果.

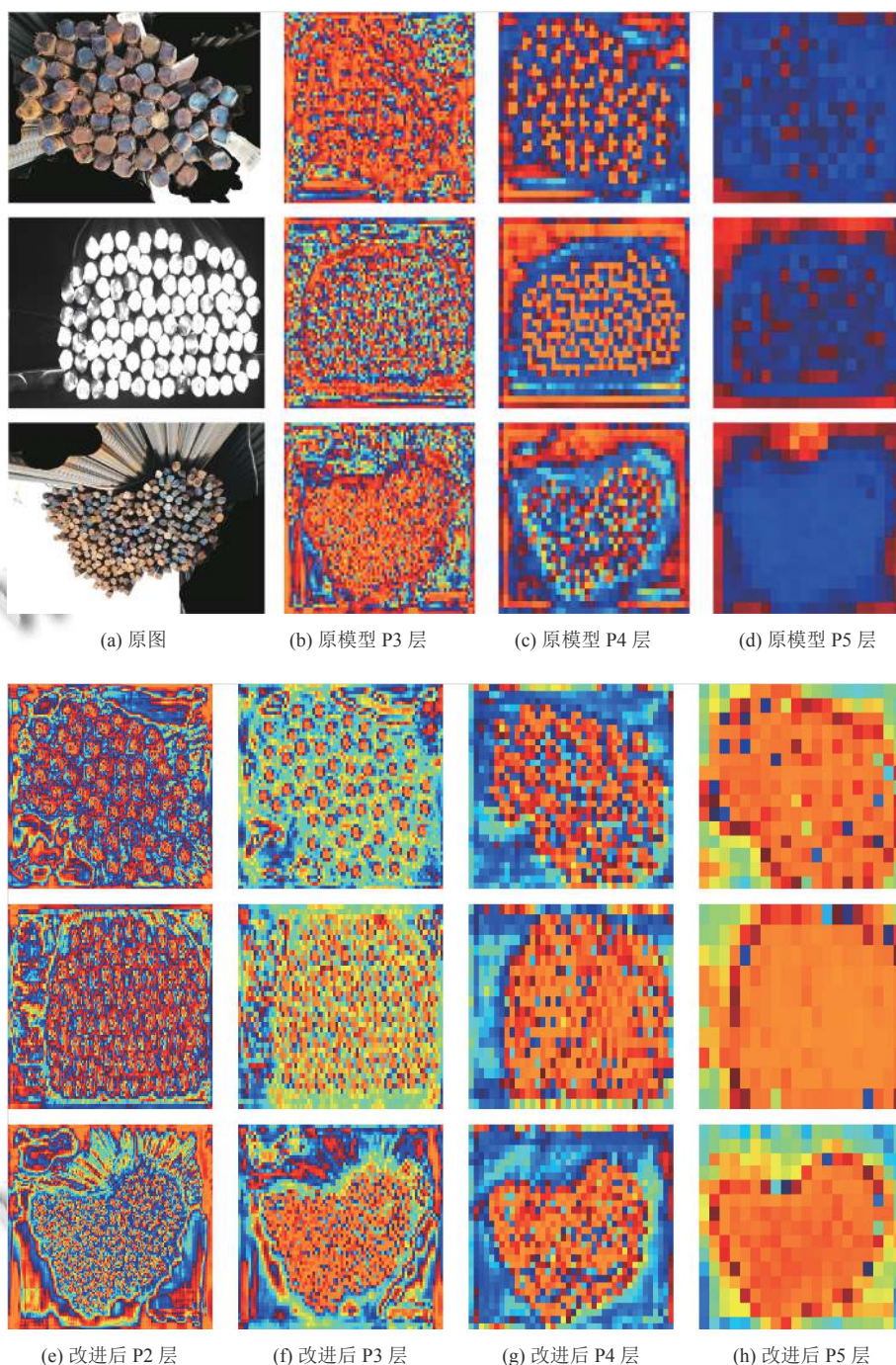


图 10 特征图可视化

如表 3 中所示, 分别对几大主流目标检测模型采用相同的数据集进行训练, 其中 YOLOv3 模型、YOLOv3-SPP 模型均在 Windows 10 环境下进行的, 输入图片尺度为 640×640, batch size 为 8 训练的迭代次

数 epoch 为 200. YOLOv4 模型与 YOLOv4-CSP 模型均在 Linux 环境下训练的, 其中 YOLOv4 模型的输入图片尺度为 608×608, YOLOv4-CSP 模型输入图片尺度为 640×640, batch size 为 4, 训练的迭代次数 epoch

为 500. 训练得到的模型分别在 330 张测试集上设置置信度参数 $\text{conf}=0.15$, 非极大值抑制阈值 $\text{IOU}=0.45$ 进行检测得到 mAP 值以及 FPS.

图 11 给出了 3 个钢筋端面检测模型的 mAP 值以

及 P-R 曲线图, 其中蓝色曲线 1 代表 YOLOv5-P2 模型, 绿色曲线 2 代表 YOLOv5 模型, 红色曲线 3 代表 YOLOv3 模型. 由曲线图可知改进后的 YOLOv5-P2 模型检测效果优于另外两个模型.

表 3 不同检测算法消融实验检测精度对比

钢筋检测模型	image size	batch size	Augmentation	Epoch	K-means	FPS	mAP (%)	$mAP@0.5:0.9$ (%)
YOLOv3	640	8	√	200	■	62.5	90.3	69.4
YOLOv3	640	8	■	200	√	62.5	86.9	66.6
YOLOv3	640	8	√	200	√	62.5	93.5	72.7
YOLOv3-SPP	640	8	√	200	■	58.8	93.1	73.6
YOLOv3-SPP	640	8	■	200	√	58.8	86.6	66.6
YOLOv3-SPP	640	8	√	200	√	58.8	93.5	72.1
YOLOv4*	608	4	√	500	■	51.7	92.0	-
YOLOv4-CSP*	640	4	√	500	√	56.2	92.9	73.4
YOLOv5s	640	8	√	500	√	71	85.2	57.9
YOLOv5x	640	4	√	500	√	35	98.8	80.2
YOLOv5-P2(ours)	640	4	√	500	√	28	99.9	85.8

注: 加粗字体为每列最优值, “■”表示未使用, “√”表示使用; “-”表示未进行实验; “*”指在Linux系统训练模型

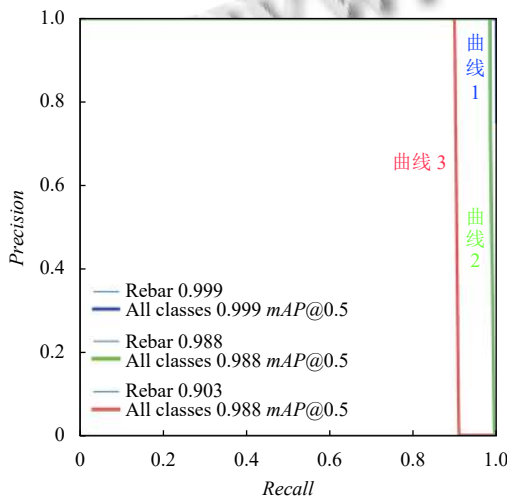


图 11 各个模型 P-R 曲线与 mAP 值

图 12 给出了为本文改进后的 YOLOv5-P2 模型的训练曲线图, 主要指标为 $mAP@0.5$ 、精准率 $Precision$ 、召回率 $Recall$ 以及训练集、验证集的目标与包围框损失曲线, 由于本文设计的是单类目标检测器因此分类损失函数恒为零.

使用 TensorBoard 对训练过程进行可视化, 可视化结果由图 13 所示, 模型输入图片尺度为 640×640 , $\text{batch-size}=8$ 时的训练曲线图. 图 13(a) 是 $mAP@0.5$ ($\text{IOU}=0.5$) 曲线图, 即 $\text{IOU}=0.5$ 时, 计算所有图片每一类的 AP 值; 图 13(b) 是 $mAP@0.5:0.9$ 曲线图: 表示在 IOU 不同阈值 (从 0.5 到 0.9, 步长 0.05) 的 mAP 值; 图 13(c) 表示训练集预测框损失曲线图; 图 13(d) 表示验证集目标损失曲线图; 图 13(e) 表示验证集预测框损失曲线图;

图 13(f) 表示训练集目标损失曲线图.

从图 13 可以看出, 改进后的两个模型在 $mAP@0.5$ 曲线 (曲线 1、曲线 2) 以及 $mAP@0.5:0.95$ 曲线图中表现最优, 这是衡量检测器的首要标准. 图 13 中曲线 1 代表 YOLOv5-P2 模型, 曲线 2 代表 YOLOv5-SPP 模型, 曲线 3 代表 YOLOv5x 模型, 曲线 4 代表 YOLOv3 模型, 曲线 5 代表 YOLOv5s 模型. 在训练损失以及验证损失曲线图中 YOLOv5-P2 模型 (曲线 1) 比 YOLOv5-SPP 模型 (曲线 2) 损失更低, 说明模型预测框与真实框更接近.

图 12 给出了置信度阈值一致时训练的 3 种模型, 在自建数据测试集设置置信度阈值 $\text{conf}=0.15$, NMS 阈值 $\text{IOU}=0.25$ 时钢筋端面检测的效果. 从图中可以看到, 对于目标边界存在粘连的情况, YOLOv3、YOLOv5 出现漏检, 改进后的 YOLOv5-P2 则能够准确检测出, 说明改进后的模型对小目标的检测效果有很大提升.

图 13 给出了当超参数设置一致时训练的 3 种模型, 在 Data Fountain 测试数据集上设置置信度阈值 $\text{conf}=0.15$, NMS 阈值 $\text{IOU}=0.25$ 时钢筋端面检测的效果, 从圈出的细节图可以看出, YOLOv3 模型和 YOLOv5 对存在遮挡、光线过强的目标有漏检、误检情况, 而 YOLOv5-P2 模型能够准确的检测到钢筋目标, 说明改进后的模型对两个数据集同时具有很强的自适应性和健壮性.

改进后 YOLOv5-P2 模型相较于改进前 YOLOv5 模型, 增加了网络上增加 P2 采样层, 并在 head 层增加第 4 个输出层, 目标间的轮廓信息更丰富, 锚点框由

9个增加到了12个,感受野减小特征图增大能够提升对小目标的检测效果.与YOLOv3模型相比优化了激

活函数和Mosaic增强以及网络结构改良,因此对于实例中目标检测效果有较大提升.

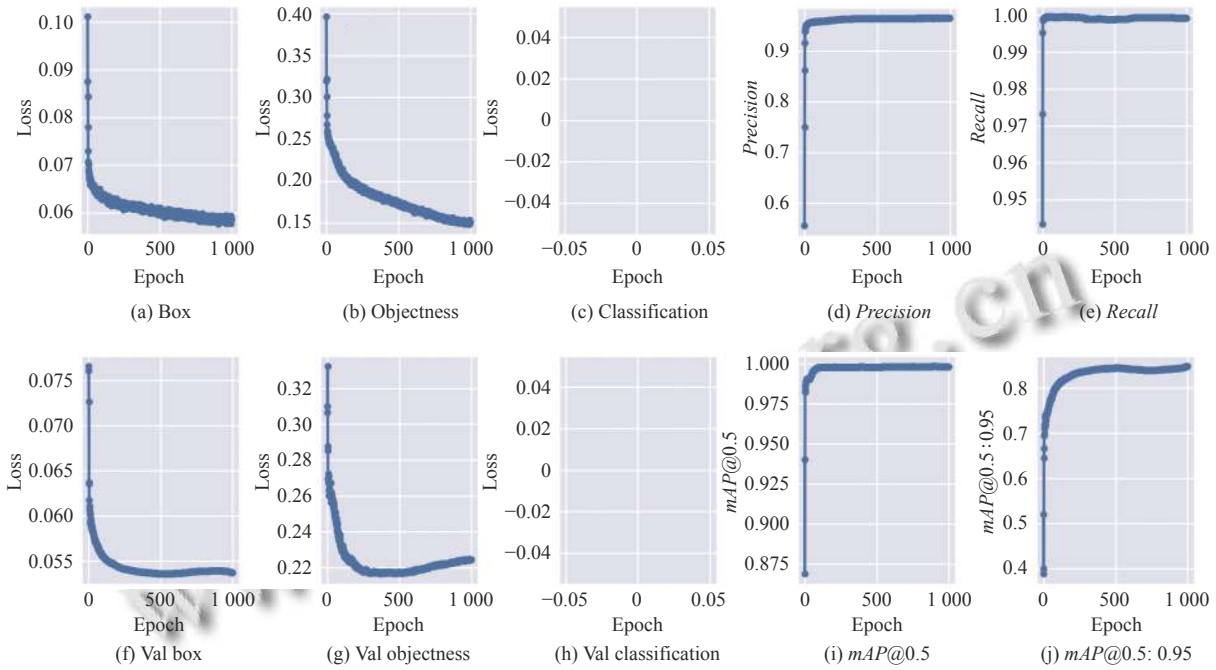


图12 YOLOv5-P2模型训练曲线图

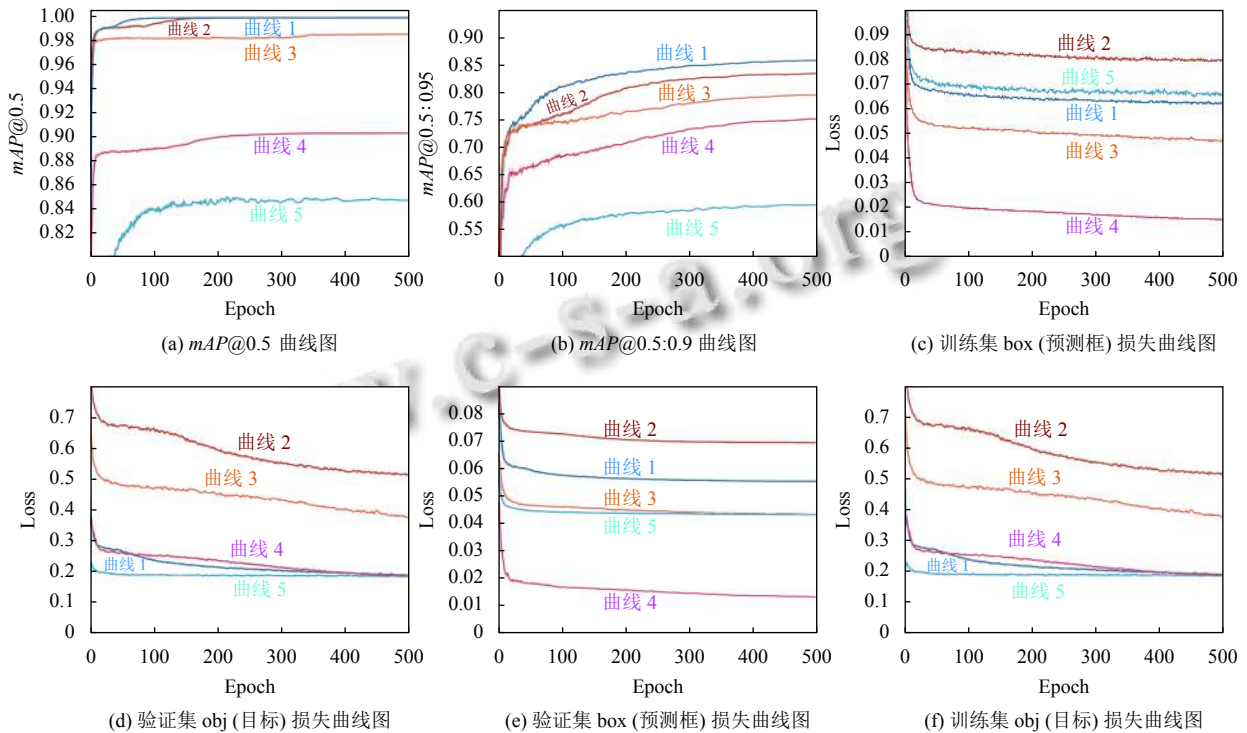


图13 各模型训练曲线对比图

为了进一步说明改进后的模型在应用于中的优越性,在330张训练集上设置相同的置信度阈值和非极

大值抑制阈值对模型进行测试得到的结果如表4所示,改进后的算法模型在测试集上检测的误检率和漏检率

明显降低,正确率相对于改进前提升了2.1%。

4 结论与展望

钢筋是建筑施工领域一种重要的材料,精准、快速的盘点出钢筋的数量代替人工盘点是一项具有重要意义的工作。本文根据钢筋端面的特点,基于改进的YOLOv5模型框架,设计了一种钢筋端面检测计数算法。本文的主要贡献为:(1)针对目前钢筋数据集稀缺且实际应用场景复杂的问题,创建了一个新的钢筋端面数据集,并使用水平、垂直镜像、亮度色彩调节、随机裁剪等方法进行数据增强;使用预训练模型对数据进行半自动标注。(2)通过改进YOLOv5网络结构,使输出特征图感受野扩大,路径聚合网络(PANet)能够提取更多的目标边缘特征,从而提升了密集小目标的检测精度。实验结果表明,本文方法在钢筋端面检测中的*mAP*和*Precision*等指标均有所提高,能满足实时检测的基础上显著提升密集成捆钢筋的检测精度。未来工作将考虑进一步改进网络结构对深层网络进行剪枝,在保证精度的基础上提升检测速度。

表4 在测试集多个模型检测结果对比

模型	误检 (张)	漏检 (张)	正确 (张)	误检率 (%)	漏检率 (%)	正确率 (%)
YOLOv3	20	99	224	6.06	30	63.94
YOLOv5	8	6	316	2.42	1.81	95.77
YOLOv5-P2	6	1	323	1.81	0.3	97.89

参考文献

- 皮琳琳. 图像识别技术在棒材计数中的应用. 城市建设理论研究: 电子版, 2013, (1): 1-4.
- 周群, 胡光锐. 基于模板覆盖法的钢筋在线计数方法. 上海交通大学学报, 1999, 33(11): 1462-1464. [doi: 10.3321/j.issn.1006-2467.1999.11.034]
- 罗山, 黄欢, 刘继红. 一种基于图像处理的成捆钢筋计数方法. 微计算机应用, 2008, 29(6): 94-97. [doi: 10.3969/j.issn.2095-347X.2008.06.019]
- 臧晶, 刘冠楠, 杨慧英. Otsu算法的打捆钢筋计数方法研究. 沈阳理工大学学报, 2017, 36(6): 47-50, 60. [doi: 10.3969/j.issn.1003-1251.2017.06.010]
- 袁中锦, 刘宇, 时娟. 基于空间相关性的成捆钢筋图像计数算法研究. 电脑与信息技术, 2019, 27(4): 10-13. [doi: 10.3969/j.issn.1005-1228.2019.04.004]
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934, 2020.
- Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector. Proceeding of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21-37. [doi: 10.1007/978-3-319-46448-0_2]
- Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440-1448.
- Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149. [doi: 10.1109/TPAMI.2016.2577031]
- 石京磊. 基于卷积神经网络的钢筋计数算法研究 [硕士学位论文]. 武汉: 华中科技大学, 2019.
- Cheng ZH, Wu YX, Xu ZH, et al. Segmentation is all you need. arXiv: 1904.13300, 2019.
- 谢海桢. 复杂场景下的棒材计数算法研究与应用 [硕士学位论文]. 成都: 电子科技大学, 2020.
- 唐楚柳. 基于卷积神经网络的钢筋端面定位与分割 [硕士学位论文]. 桂林: 广西师范大学, 2020.
- 明洪宇, 陈春梅, 刘桂华, 等. 基于RetinaNet的密集型钢筋计数改进算法. 传感器与微系统, 2020, 39(12): 115-118.
- 王志丹. 基于小数据的钢筋检测方法的研究. 邮电设计技术, 2020, (2): 39-44. [doi: 10.12045/j.issn.1007-3043.2020.02.009]
- He KM, Zhang XY, Ren SQ, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916. [doi: 10.1109/TPAMI.2015.2389824]
- Liu S, Qi L, Qin HF, et al. Path aggregation network for instance segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759-8768. [doi: 10.1109/CVPR.2018.00913]
- Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks, 2018, 107: 3-11. [doi: 10.1016/j.neunet.2017.12.012]
- Yu JH, Jiang YN, Wang ZY, et al. UnitBox: An advanced object detection network. Proceedings of the 24th ACM international conference on Multimedia. Amsterdam: ACM, 2016. 516-520.
- Ratner B. Pythagoras: Everyone knows his famous theorem, but not who discovered it 1000 years before him. Journal of Targeting, Measurement and Analysis for Marketing, 2009, 17(3): 229-242. [doi: 10.1057/jt.2009.16]
- Paras. Stochastic gradient descent. Optimization, 2014. [doi: 10.1007/978-1-4842-2766-4_8]
- Zhang HY, Cisse M, Dauphin YN, et al. mixup: Beyond empirical risk minimization. arXiv: 1710.09412, 2017.
- DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv: 1708.04552, 2017.
- Wang CY, Bochkovskiy A, Liao HYM. Scaled-YOLOv4: Scaling cross stage partial network. arXiv: 2011.08036, 2020.