基于改进 YOLOv5s 的药盒钢印日期识别方法

黄杨乐天，刘宜胜，王俊茹*
（浙江理工大学 机械工程学院，杭州 310018）

摘要：目的 药盒的钢印日期与背景对比度低，字符轮廓不明显，识别易受环境光线干扰，对此提出一种基于机器视觉的识别方法。**方法** 使用改进 YOLOv5s 模型，首先对采集的药盒数据集进行透视变换校正，并进行数据增强。通过在模型的骨干网络中融合位置注意力机制（CA），减少冗余信息的干扰；颈部网络根据加权双向特征金字塔网络（BiFPN）引入权重，更好地平衡不同尺寸图层的特征信息；引入动态聚焦损失函数（WIoU），降低高质量样本对训练的干预，提高模型的泛化能力。**结果** 在自建钢印字符数据集上的实验结果表明，改进网络对药盒钢印日期识别的平均精度值达到了 99.41%，比原始模型提升了 2.38%，帧率为 80.01 帧/s。**结论** 改进后的 YOLOv5 模型对药盒钢印日期的检测精度优于原有网络，对可以满足药盒生产线的实时性要求。

关键词：钢印日期；透视变换；目标检测；加权特征图；注意力机制
中图分类号：TP391 **文献标志码：**A **文章编号：**1001-3563(2024)07-0189-08
DOI：10.19554/j.cnki.1001-3563.2024.07.024

Improved YOLOv5s-based Date Recognition Method for Steel Stamps on Pill Boxes

HUANG Yangletian, *LIU Yisheng*, *WANG Junru**
(School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China)

ABSTRACT: The work aims to propose a machine vision-based recognition method for pill boxes with low contrast between the steel-stamped date and the background, inconspicuous character outlines, and recognition susceptible to interference by ambient light. An improved YOLOv5s model was used to correct the collected pill box dataset by perspective transformation and data enhancement. By fusing the Coordinate Attention (CA) in the backbone network of the model, the interference of redundant information was reduced. The neck network introduced weights according to the Bi-directional Feature Pyramid Network (BiFPN) to better balance the feature information of the layers of different sizes. The Wise-IoU (WIoU) was introduced to reduce the intervention of high-quality samples in the training and to improve the model's generalization ability. The experimental results on the self-constructed steel-stamped character dataset showed that the average accuracy of the improved network for recognizing the steel-stamped date of the pill box reached 99.41 %, which was 2.38 % higher than that of the original model, and the frame rate was 80.01 f/s. The improved YOLOv5 model can detect the steel-stamped date of the pill box with a better accuracy than that of the original network, and it can meet the real-time requirement of the production line of the pill box.

KEY WORDS: steel-stamped date; perspective transformation; target detection; weighted feature maps; coordinate attention

收稿日期：2023-07-11
基金项目：浙江省“尖兵”“领雁”研发攻关计划项目（2023C01158）
*通信作者

生产日期是药盒上的重要信息，很多厂家选择使用钢印字符。因为相较于喷墨、印刷等方式，钢印的过程相对简单，不易受物理磨损，不会随时间而脱落。起初，厂家雇佣劳动力来检测产品日期。这种检测方式不仅速度较慢，而且易导致视觉疲劳，漏检、错检率较高。随着科技进步，人们利用传统的图像处理来识别字符。比较常用的有：OCR 光学字符识别；基于文字特征提取，对字符特征进行识别；利用建立好的字符库进行模板匹配。而钢印字符由药盒表面施压生成，这种字符和背景颜色相近，轮廓并不明显[1]。此外，药盒钢印字符的样式和尺寸不统一，模板匹配的泛用性较差，故传统方法的效果并不理想。

近年来，基于深度学习的神经网络在识别领域得到了广泛应用，大致分为 2 种：以 R-CNN[2-4]系列为代表，首先从图像中提取候选区域，再对候选区域进行分类和回归的两阶段目标检测算法；以 YOLO(You Only Look Once) [5]系列和 SSD[6]等为代表，将整个图像输入卷积神经网络，提取图像特征进行回归和预测的单阶段目标检测算法。相比双阶段算法，单阶段算法相对牺牲了一些精度，但提升了识别速度，更好地满足生产线的实时性要求。YOLOv5 是兼顾精度和速度的模型，非常适合部署在工厂设备上[7]。Santoso 等[8]提出基于 YOLO 的 Kawi 铜质铭文识别方法；宫鹏涵[1]提出一种基于 YOLOv5 的枪械钢印字符识别方法；Zhang 等[9]针对铸造压印字符使用 YOLOv5 进行识别。综上所述，本文提出一种基于改进 YOLOv5s 的药盒钢印日期识别方法，主要改进为：在骨干网络中融合位置注意力机制（CA），以减少背景信息的干扰；引入双向特征金字塔结构（BiFPN），更好地融合多尺度特征和上下文信息；采用动态聚焦损失函数（WIoU）提高模型对不平衡样本的学习能力。

1　检测模型

1.1　模型结构

该模型由骨干网络（Backbone）、颈部网络（Neck）、头部网络（Head）3 个部分组成。Backbone 网络用于提取特征，使用连续的卷积模块来提取各类尺寸的特征图，以保留相应的特征信息。Neck 网络进一步提取特征，使大尺度的语义信息和小尺度的细节信息可以更好的融合，增强多尺度的目标定位能力。Head 网络有 3 个检测头，用于目标的预测和回归。图 1 为改进 YOLOv5s 模型结构。

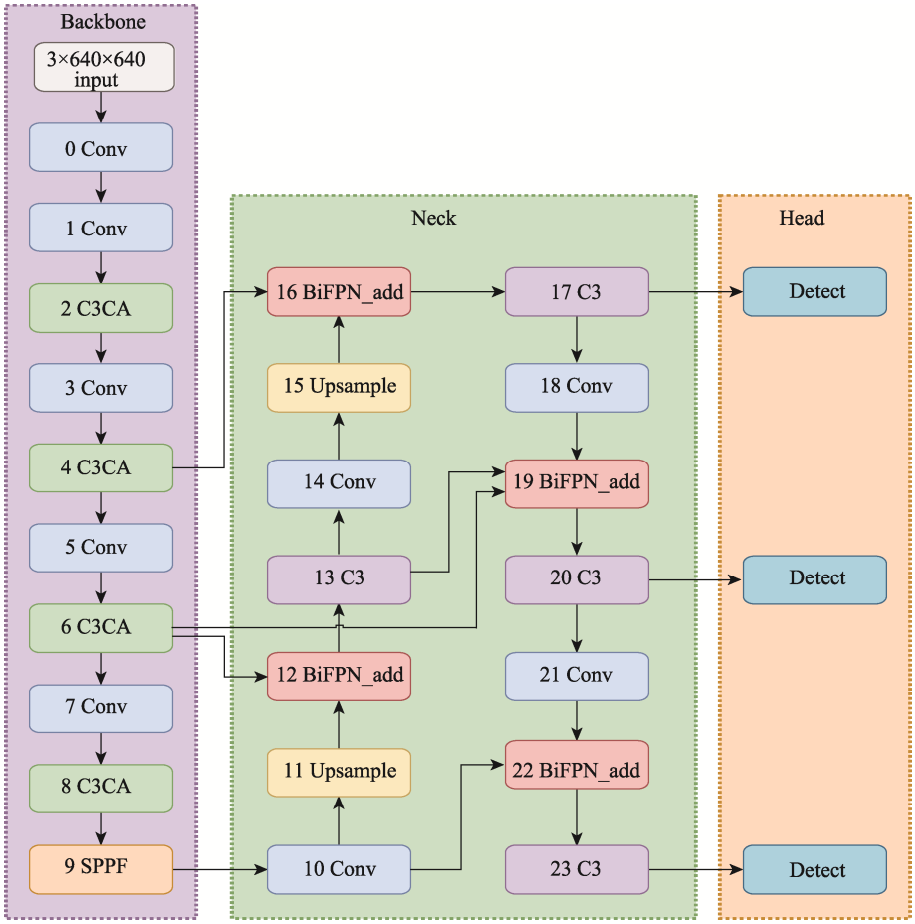


图 1　改进 YOLOv5s 模型结构
Fig.1 Improved YOLOv5s structure

1.2　改进模型

1.2.1　骨干网络改进

注意力机制可以使有限的计算资源更加集中地应用于关键的识别区域。常用的注意力机制有 SE[10]、ECA[11]、CBAM[12]和 CA[13]等。前两者属于通道注意力机制，CBAM 模块在此基础后加入了空间注意力机制，以使模型更关注重要区域。CA 模块将位置信息嵌入通道注意力机制中，以更高效地获取 2 种信息，而不增加过多计算量。如图 2 所示，CA 模块先将输入特征图（$C\times H\times W$）从宽度和高度 2 个方向使用池化核（H，1）和（1，W）进行全局平均池化，分别获得其特征图。再分别沿 2 个空间方向聚合特征，这使得注意力机制保留位置信息的同时捕获通道信息，有助于模型更精确地定位感兴趣的像素区域。

1.2.2　特征金字塔改进

BiFPN[14]旨在提高模型的感受野和特征表征能力，能够有效地提高模型的性能和效率。如图 3 所示，相较于左侧的 FPN 结构，BiFPN 删除了只有一条输入的节点（P3、P6）。这样的节点并没有很好地融合不同的特征信息，反而会增加模型参数量。同时，BiFPN 在同一层的输入和输出之间添加了额外的融合通路（虚线箭头通路），以较小的计算成本进一步增强特征融合能力。此外，原有融合方式只是简单地将特征图叠加在一起。然而，不同输入特征具有不同的分辨率和贡献度，简单的加和并不是最佳选择。BiFPN 提出了一种简单高效的加权特征融合机制，能够快速对融合后的特征进行归一化处理，从而减少模型计算量。因此，将网络的第 6 层和第 19 层连接。融合计算见式（1）。

$$P_{19}^{\text{out}} = Conv\left(\frac{w_1 P_6^{\text{in}} + w_2 P_{13}^{\text{in}} + w_3 P_{18}^{\text{in}}}{w_1 + w_2 + w_3 + \varepsilon}\right) \tag{1}$$

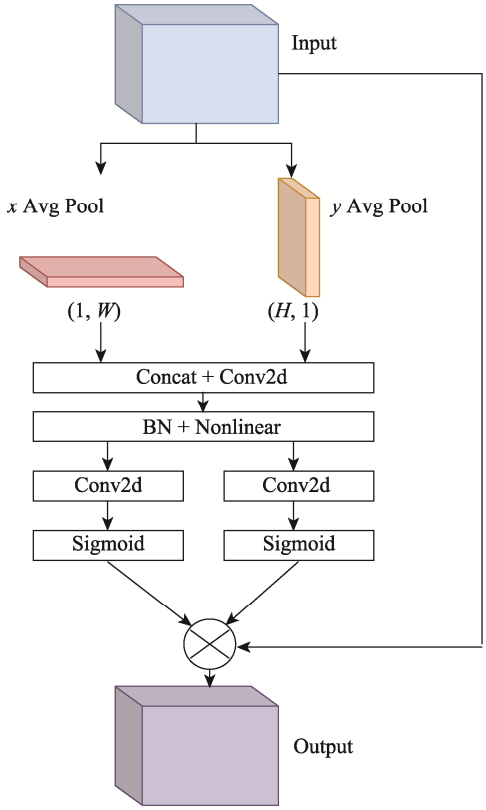


图 2　CA 注意力机制模块结构
Fig.2 CA module structure

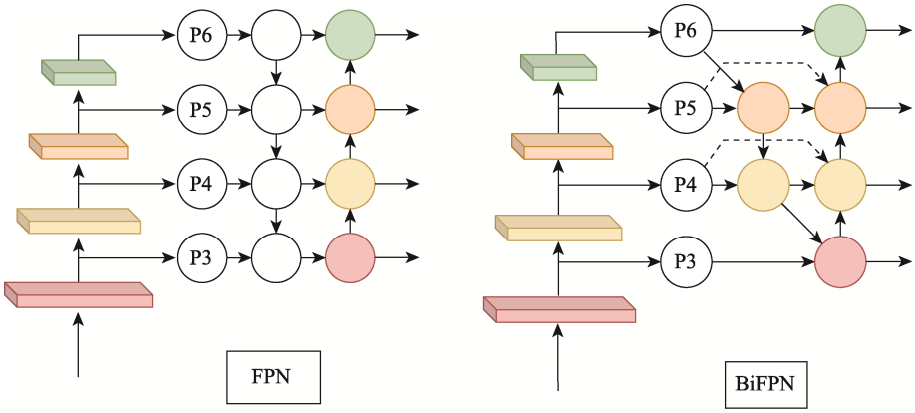
式中：w_i 为一个可学习权重，通过 ReLU 激活函数来保证每一个 $w_i \geqslant 0$；ε 为一个避免数值不稳定的系数，其值为 0.000 1。

1.2.3　损失函数改进

损失函数用于衡量预测值与实际值之间差异的函数。在目标检测领域经常使用 IoU 损失函数。其计算式见式（2）。

$$L_{\text{IoU}} = 1 - I_{\text{IoU}} \tag{2}$$

式中：I_{IoU} 为预测边界框与真实边界框之间的交集与并集之比。



a FPN结构　　b BiFPN结构

图 3　特征金字塔结构
Fig.3 Characteristic pyramid structure

为了防止损失函数梯度过小，影响模型的学习，研究者提出许多改进 IoU 计算方法[15-18]。如图 4 所示，由于生产日期的性质，样本不平衡无法避免。而且数据集难免包含低质量样本，导致其对数量较少的类别影响更大，降低模型的泛化能力。故引入 WIoU 损失函数，以不同的关注程度对待高质量样本和中低质量样本。当预测框和真实框相匹配时，WIoU 可以降低此类高质量样本对模型训练的干预，并且在防止低质量样本产生有害梯度的前提下更加关注于中低质量样本，提高模型泛化能力。WIoU 计算式见式（3）。

$$L_{\mathrm{Loss}} = R_{\mathrm{WIoU}} L_{\mathrm{IoU}} \tag{3}$$

$$R_{\mathrm{WIoU}} = \exp\left(\frac{\left(x-x_{\mathrm{gt}}\right)^2+\left(y-y_{\mathrm{gt}}\right)^2}{\left(W_{\mathrm{g}}^2+H_{\mathrm{g}}^2\right)}\right) \tag{4}$$

式中：x_{gt} 和 y_{gt} 为真实框坐标；W_{g} 和 H_{g} 分别为最小包围框的宽和高。

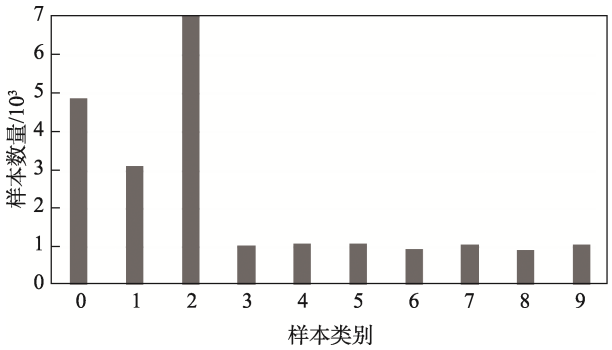


图 4　数据集样本数量分布
Fig.4 Dataset sample distribution

2　实验与结果分析

2.1　实验准备

2.1.1　数据集建立

本实验数据集为志愿者在不同条件下采集的图像。如图 5 所示，包括国内外不同种类的药盒，其字符形状、尺寸、分布密集程度各有差异，以提高模型的泛用性。

2.1.2　透视变换

相机采集药盒时可能会出现倾斜现象，而且会包含无效的背景信息，故利用透视变换对其进行校正。如图 6 所示，不仅使目标区域处于水平状态，且四周的背景被去除，利于模型关注目标区域处理更少的像素。透视变换公式为：

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{5}$$

$$\begin{cases} \boldsymbol{T}_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ \boldsymbol{T}_2 = \begin{bmatrix} a_{31} & a_{32} \end{bmatrix} \\ \boldsymbol{T}_3 = \begin{bmatrix} a_{13} & a_{23} \end{bmatrix}^{\mathrm{T}} \end{cases} \tag{6}$$

$$\begin{cases} u = \dfrac{x'}{z'} = \dfrac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}} \\ v = \dfrac{y'}{z'} = \dfrac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}} \\ w = \dfrac{z'}{z'} = 1 \end{cases} \tag{7}$$

原始图像像素点的齐次坐标（x，y，z）对应变换之后的归一化齐次坐标（u，v，w）。矩阵 $\boldsymbol{T}$ 为透视变换矩阵。$\boldsymbol{T}_1$ 为线性变换，如放缩、旋转，$\boldsymbol{T}_2$ 为平移操作，$\boldsymbol{T}_3$ 产生透视变换。

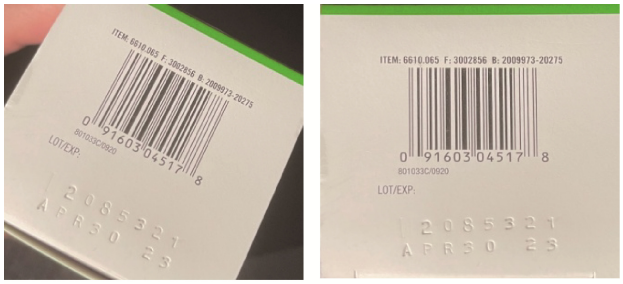
2.1.3　数据增强

数据集以 8∶1∶1 的比例划分为训练集、测试集、验证集，供训练网络模型使用。为了增加训练样本并提高网络模型的泛化能力，对训练集进行扩充。采用图像模糊、加噪、缩放、翻转、剪裁和调整亮度等处理方式随机组合，得到最终数据集图像共 1 620 张。

a 清晰字符　　b 失焦字符　　c 弱光字符

d 曝光字符　　e 反光字符　　f 分散字符

图 5　不同采集条件的图像
Fig.5 Images under different collection conditions



a 原图像　　b 透视变换后图像

图 6　透视变换结果对比
Fig.6 Comparison of perspective transformation results

2.2　实验环境

实验平台操作系统为 Windows10，显卡为 NVIDIA GeForce RTX3090，24 G 显存，使用 PyTorch 深度学习框架，版本为 1.12.1，CUDA 版本为 11.3，编译语言为 Python3.9.15；图片输入尺寸为 640×640；批量大小设为 16，迭代次数设为 100 轮，学习率设为 0.001。

2.3　评价指标

为衡量识别药盒钢印字符的效果以及模型性能，选取神经网络性能评估指标：精确率（Precision，P）、召回率（Recall，R）、平均精度（Average Precision，A_P）、平均精度均值（mean Average Precision，M_{AP}）、每秒浮点运算次数（Flops）、平均帧率（Frame Per Second，FPS）和参数量。相关计算公式如下：

$$P=\frac{T_P}{T_P+F_P} \tag{8}$$

$$R=\frac{T_P}{T_P+F_N} \tag{9}$$

$$A_P=\int_0^1 P(r)\mathrm{d}r \tag{10}$$

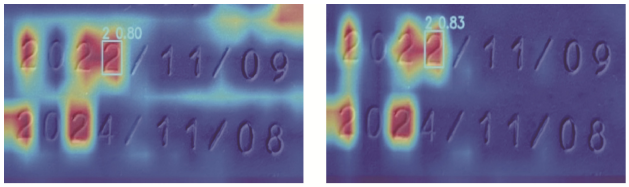
$$M_{AP}=\frac{1}{n}\sum_{i=1}^{n}A_{Pi} \tag{11}$$

式中：T_P 为被模型预测为正类的正样本；F_P 为被模型预测为正类的负样本；F_N 为被模型预测为负类的正样本；A_P 为 PR 曲线所围成的面积；M_{AP} 为所有类别的 A_P 的平均值。本实验 M_{AP} 指标选用 $M_{AP_0.5}$，即阈值为 0.5。

2.4　结果对比

2.4.1　骨干网络对比

为评估注意力机制的效果，对特征图进行热力图可视化，并对比了不同的插入位置，结果分别如图 7 和表 1 所示。红色表示模型的关注程度高，可以看出原模型在识别字符 2 时，同时也关注了右上角的背景信息。相比之下，添加 CA 的模型，集中于目标区域，且在字符识别方面的表现更加优秀，置信度更高。这表明注意力机制对提高模型性能具有显著的效果。此外，相比直接插入，融合跨阶段模块（C3CA）的参数量更小且 M_{AP} 值更高。说明添加过多的注意力机制可能使图像中的目标信息被当成了背景信息，导致模型的精度下降。



a 未添加CA模块　　b 添加CA模块

图 7　模型热力图可视化
Fig.7 Visualization of model thermal diagram

表 1　CA 插入位置的结果对比
Tab.1 Comparison results of CA in different layers

模块	参数量	M_{AP}/%
CA	7 072 247	97.2
C3CA	7 067 567	97.5
CA+C3CA	7 093 215	96.3

2.4.2　损失函数对比

各类损失函数在不同应用场合的性能不同，需要根据任务的要求具体情况具体分析，故在本钢印字符检测任务中进行了对比，总损失越小，代表预测值与期望值更接近，结果如图 8 所示。可以看出，WIoU 损失函数在本实验数据集上取得较好的效果，损失收敛速度快且总损失较小，GIoU 的效果相对最差。

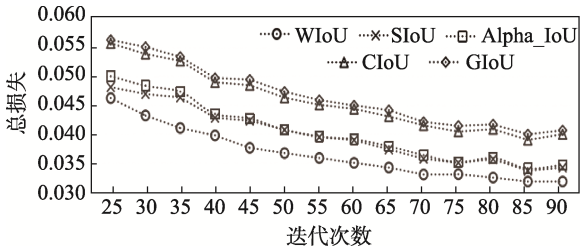
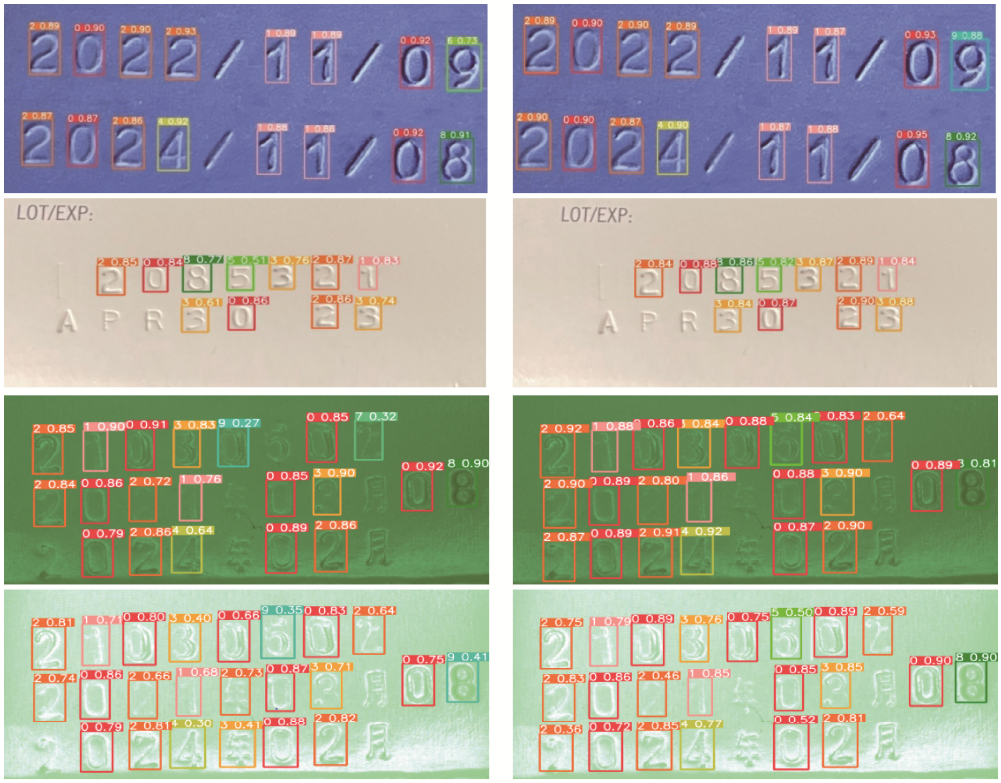


图 8　不同损失函数对比
Fig.8 Comparison of different loss functions

2.4.3　识别结果

图 9 和表 2 为原网络和改进网络的部分识别结果对比。无论对于正常或曝光、弱光图像，还是改进后的漏检、误检现象明显少于原网络，尤其是改进后的模型对几何特征相似的字符 6 和 9 的检测效果更好。证明改进网络对目标的特征提取能力更高，检测效果更好。



a YOLOv5s　　b 改进YOLOv5s

图 9　识别结果对比

Fig.9 Comparison of recognition results

表 2　改进前后对字符 6 和 9 的检测结果对比

Tab.2 Comparison of detection results for characters 6 and 9 after improvement

类别	YOLOv5s		改进 YOLOv5s	
	精确率 P/%	召回率 R/%	精确率 P/%	召回率 R/%
字符 6	95.0	84.4	98.1	96.4
字符 9	97.8	97.6	99.0	100.0

2.4.4　消融实验

为了验证各种改进模块对模型性能的有效性，进行了消融实验，结果见表 3。根据模型 A、B、C 的表现，可以发现，虽然 C3CA 和 BiFPN 模块增加了模型的参数量，但对帧率的影响不大，而 WIoU 模块对参数量没有影响。3 个模块单独应用都能提升检测效果，其中 BiFPN 模块的效果最为显著，提升了 1.08%。由模型 D、F、G 可以发现组合不同的模块能够进一步提升模型性能，M_{AP} 提升分别为 2.03%、1.79%、1.34%。当 3 种模块共同作用时，模型的识别效果最佳，M_{AP} 提升了 2.38 %，具有更好的性能。

2.4.5　模型对比

为进一步验证本文改进模型的性能，通过参数量、M_{AP} 和 GFLOPs 这 3 个指标与常用的检测模型进行对比，对比结果见图 10。结果显示，改进的 YOLOv5

表 3　消融实验结果

Tab.3 Results of ablation experiment

模型	改进			平均精度均值/%	参数量	浮点运算次数	帧率/(帧·s^{-1})
	C3CA	BiFPN	WIOU				
YOLOv5s				97.03	7 046 599	15.9	86.96
YOLOv5s-A	√			97.52	7 067 567	16.1	80.65
YOLOv5s-B		√		98.11	7 185 584	16.3	80.00
YOLOv5s-C			√	97.34	7 046 559	15.9	86.21
YOLOv5s-D	√	√		99.06	7 211 232	16.4	81.97
YOLOv5s-F		√	√	98.82	7 185 584	16.3	78.13
YOLOv5s-G	√		√	98.37	7 067 567	16.1	79.37
改进 YOLOv5s	√	√	√	99.41	7 211 232	16.4	80.01

注：√表示本次实验使用了该改进模块。

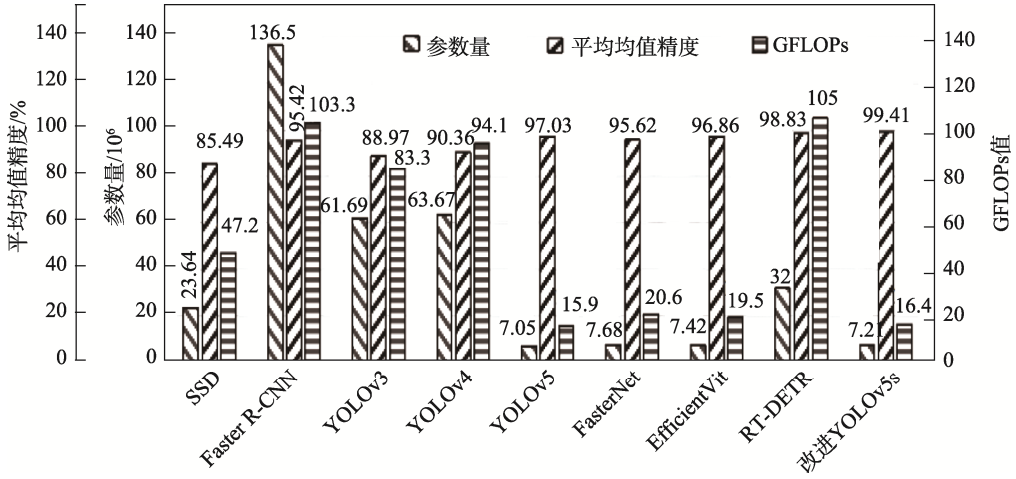


图 10　各类模型性能对比

Fig.10 Performance comparison of various models

相较于 SSD、Faster R-CNN、YOLOv3 和 YOLOv4 算法，在 M_{AP} 方面分别提高了 13.92%、3.99%、10.44%、9.05%。同时，改进算法拥有更小的参数量，适用于资源有限的部署条件。与 FasterNet 等参数量基本持平的算法相比，改进的 YOLOv5 精度最高，达到了 99.41%。此外，虽然 RT-DETR[19]的 M_{AP} 值相对接近，但其 GFLOPs 值高达 105，表明无法在较低的计算资源下实时运行。因此，改进 YOLOv5s 在高精度检测和高效率计算之间取得了良好的平衡，非常适用于对性能和效率要求都较高的应用场景。因此，改进的 YOLOv5 能够满足工业钢印字符检测的要求。

3　结语

本文针对药盒钢印字符识别任务中的多种复杂情况，如目标与背景对比度低、目标图像模糊和光线方向等情况，提出了一种基于改进 YOLOv5 的药盒钢印字符识别方法。首先对采集的图像进行透视变换，图像增强等预处理，通过在骨干网络中添加 CA 注意力机制，使其更集中关注识别对象的位置信息；颈部网络采用了 BiFPN 的结构，更好地融合不同尺度的特征信息；使用 WIoU 损失函数，注重中低质量样本的贡献，减少样本不平衡带来的影响。实验结果表明，本文提出的模型相较于常用的目标检测模型在参数量方面均有下降，与 YOLOv5s 模型基本持平，而在精度方面有显著提升，M_{AP} 达到了 99.41%。改进后的模型具有良好的泛用性，可以满足药盒生产流水线的实时精度要求，并可为其他钢印字符识别任务提供参考。

参考文献：

[1] 宫鹏涵. 基于 YOLOv5 算法的钢印字符识别方法[J]. 兵器装备工程学报, 2022, 43(8): 101-105.

GONG P H. Character Recognition Research of Steel Embossing Based on YOLOv5[J]. Journal of Ordnance Equipment Engineering, 2022, 43(8): 101-105.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Conference on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE, 2014: 580-587.

[3] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448.

[4] REN S Q, HE K M, 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.

[5] 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 (CVPR). Las Vegas: IEEE, 2016: 779-788.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]// European Conference on Computer Vision (ECCV). Amsterdam: Springer, 2016: 21-37.

[7] JI Z, WU Y, ZENG X, et al. Lung Nodule Detection in Medical Images Based on Improved YOLOv5s[J]. IEEE Access, 2023, 11: 76371-76387.

[8] SANTOSO R, SUPRAPTO Y K, YUNIARNO E M. Kawi Character Recognition on Copper Inscription Using YOLO Object Detection[C]// Proceedings of 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). Sura-

baya: IEEE, 2020: 343-348.

[9] ZHANG Z, YANG G, WANG C, et al. Recognition of Casting Embossed Convex and Concave Characters Based on YOLO v5 for Different Distribution Conditions[C]// Proceedings of 2021 International Wireless Communications and Mobile Computing (IWCMC). New York: IEEE, 2021: 553-557.

[10] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 7132-7141.

[11] WANG Q, WU B, ZHU P, *et al.* ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE/CVF, 2020: 11534-11542.

[12] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional Block Attention Module[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3-19.

[13] HOU Q, ZHOU D, FENG J. Coordinate Attention for Efficient Mobile Network Design[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE/CVF, 2021: 13713-13722.

[14] TAN M, PANG R, LE Q V. Efficientdet: Scalable and Efficient Object Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE/CVF, 2020: 10781-10790.

[15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[16] HE J, ERFANI S, MA X, et al. Alpha-IOU: A Family of Power Intersection over Union Losses for Bounding Box Regression[J]. Advances in Neural Information Processing Systems, 2021, 34: 20230-20242.

[17] GEVORGYAN Z. SIoU loss: More Powerful Learning for Bounding Box Regression[J/OL]. arXiv preprint, 2022: 1-12. https://arxiv.org/abs/2205.12740.

[18] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J/OL]. arXiv preprint, 2023: 1-8. https://arxiv.org/abs/2301.10051.

[19] LYU W, XU S, ZHAO Y, et al. Detrs Beat Yolos on Real-Time Object Detection[J/OL]. arXiv preprint, 2023: 1-11. https://arxiv.org/abs/2304.08069.