

文章编号: 2095-2163(2023)12-0093-05

中图分类号: TP391.41

文献标志码: A

# 基于改进 YOLOv5 的橘瓣检测方法

喻擎苍<sup>1</sup>, 周文博<sup>1</sup>, 邱锐<sup>1</sup>, 陶坚<sup>2</sup>

(1 浙江理工大学 计算机科学与技术学院, 杭州 310018; 2 浙江科技学院 信息与电子工程学院, 杭州 310023)

**摘要:** 针对柑橘罐头厂生产橘子罐头时人工分类橘瓣效率差的问题, 本文提出一种基于深度学习的橘瓣检测方法。以柑橘罐头厂生产线上的橘瓣为对象, 实地采集图像并制作了一个具有 2 500 张图像的橘瓣数据集; 在 YOLOv5s 模型的基础上针对橘瓣检测对象多、遮挡大的特点, 融合注意力机制并改进了损失函数, 得到改进的 YOLOv5s 模型。试验结果表明, 在该模型上橘瓣检测的平均精度达到 93.7%, 单张图像检测耗时 25 ms, 基于改进 YOLOv5s 模型的橘瓣检测方法能够满足工厂生产线的实际应用需求。本方法可以为橘瓣自动化分类设备提供高精度的视觉指导。

**关键词:** 深度学习; 橘瓣检测; YOLOv5;

## Orange flap detection method based on improved YOLOv5

YU Qingcang<sup>1</sup>, ZHOU Wenbo<sup>1</sup>, QIU Rui<sup>1</sup>, TAO Jian<sup>2</sup>

(1 School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China;

2 School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China)

**Abstract:** Aiming at the problem of poor efficiency of manual classification of orange flaps in orange canning factories, a method of orange flaps detection based on deep learning was proposed. In this paper, a data set with 2 500 images was made by taking the images from the production line of citrus canning plant as the object. Based on the YOLOv5s model, the attention mechanism is fused and the loss function is improved to achieve the improved YOLOv5s model, in which the average accuracy of the orange lobe detection task reaches 93.7% and the detection time of a single image is 25 ms. The experimental results show that the orange detection method based on the improved YOLOv5s model can meet the actual application requirements of the factory production line. This method can provide high-precision visual guidance for orange petal automatic sorting equipment.

**Key words:** deep learning; detection of orange petal; YOLOv5

## 0 引言

橘子罐头的生产工艺经过原料验收、选果清洗、分瓣、酸碱处理、清洗、橘瓣分类、橘瓣分级、选剔、装罐、称重、封罐头以及杀菌等工序<sup>[1]</sup>。其中, 酸碱处理是为了去掉橘瓣上的囊皮, 使橘瓣口感更佳, 但处理后仍有部分橘瓣上带有囊皮, 还有部分未分离橘瓣、有核橘瓣或破碎橘瓣等不良品, 需要分拣出来特殊处理。

目前, 罐头厂大多采取人工方法解决橘瓣的分类问题, 效率低、用工量非常大, 罐头生产的人工成本占总成本的 25% 左右; 人工的劳动强度大, 可持续性不足; 人工分类还存在误检、漏检情况, 造成食品安全隐患。实现橘瓣分类的自动化, 可以极大的提升罐头厂

的生产力, 减少人力成本, 还可以减少装罐后的坏果率, 提升食品安全等级, 具有重大现实意义。

采用机器视觉的图像处理技术可以实现高效的橘瓣自动检测, 消除人力检测橘瓣的弊端<sup>[2]</sup>。为了实现在流水线上快速准确的检测并分类密集的橘瓣, 本文提出一种基于改进 YOLOv5 的橘瓣检测方法, 可以为自动化橘瓣分类提供解决方案。

## 1 数据集制作

### 1.1 图像采集

本文的数据集是在柑橘罐头工厂实地采集, 自制的橘瓣数据集, 如图 1 所示。该数据集将橘瓣分为 5 种: 完好橘瓣、有核橘瓣、有囊皮橘瓣、未分离橘瓣和破碎橘瓣。为保证数据集的真实性和可靠性,

**作者简介:** 喻擎苍(1969-), 男, 博士, 教授, 硕士生导师, 主要研究方向: 计算机视觉、计算机控制; 周文博(1999-), 男, 硕士研究生, 主要研究方向: 机器学习、模式识别。

**通讯作者:** 陶坚(1966-), 男, 硕士, 副教授, 主要研究方向: 信息处理、自动控制。Email: 13355786986@qq.com

收稿日期: 2022-12-19

哈尔滨工业大学主办 ◆ 系统开发与应用

使用单反相机在流水线工人的工位前录制,获得时长 20 h 的录像;以录像为来源,考虑到不同密集程度、各个橘瓣的互相遮挡、橘瓣形状的差异和各类样本的数量等情况,截取了 4 032 像素 $\times$ 3 024 像素的图像,并对图像进行筛选、裁剪,最终得到满足实验条件的图像共 900 张,保存分辨率为 640 $\times$ 640,保存格式是.JPG。

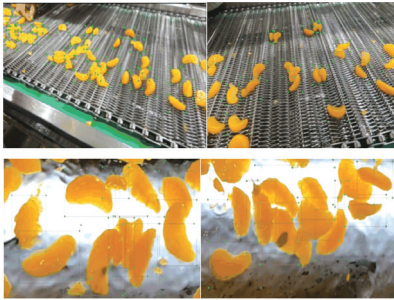


图1 橘瓣数据集

Fig. 1 Orange petal dataset

## 1.2 图像标注与划分

本文选用 Label 标注工具完成图像的标注,对数据集中每一张图片中的每一个橘瓣按照类别框选,得到 YOLO 格式的标注结果。

为了避免模型训练过程中的过拟合现象,对数据集进行了数据增广,即采用翻转变换、平移变换、尺度变换、噪声扰动、随机擦除和亮度调整等操作将数据集扩充到2 500张。将数据集按照 8 : 1 : 1 的比例划分为训练集、验证集以及测试集。

在完成标注工作后要配置 YOLO 的数据集文件,文件采用.yaml 格式,内容为数据集、验证集和测试集路径、分类的种类数以及各个类型的顺序。

## 2 YOLOv5 检测方法

### 2.1 改进 YOLOv5 网络模型

YOLOv5 目标检测模型是典型的 one-stage 检测器模型<sup>[3]</sup>。该模型能够将每一张输入图片直接回归出各个目标的位置和类别,具有精度高、速度快的特点,符合罐头工厂中高速流水线的应用场景。

YOLOv5 模型具有 4 种不同的结构:YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x。4 种结构网络深度和特征图的宽度都不相同,因此准确度和运行速度也不同,4 种结构具有各自适合的应用场景。本文基于 YOLOv5s 模型进行改进,改进的 YOLOv5s 模型结构如图 2 所示。

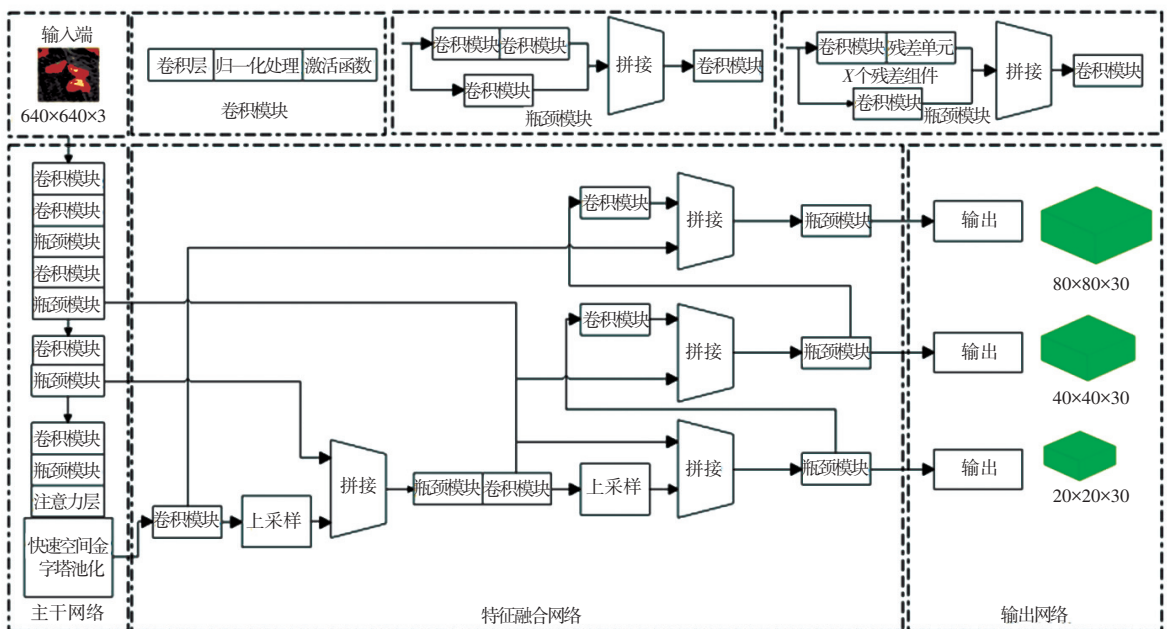


图2 改进的 YOLOv5s 模型结构

Fig. 2 Improved YOLOv5s model structure

### 2.2 输入端

输入端包括 3 个主要操作:Mosaic 数据增强、自适应锚框计算和自适应图片缩放。Mosaic 数据增强将图片随机裁剪拼接,增强数据集多样性;自适应锚框计算在训练开始前,采用 k-means 算法对数据集

进行分析,得出合适数据集的自定义大小的预设锚定框;自适应图片缩放将不同大小的输入图片缩放为固定大小,本文设定为 640 像素 $\times$ 640 像素。

### 2.3 改进主干网络

在主干网络中,改进 YOLOv5s 模型主要使用卷

积模块、瓶颈模块和快速空间金字塔池化模块。卷积模块包括卷积层、批标准化层和 Silu 激活函数, 保证下采样时信息不会丢失; 瓶颈模块将原输入分成两个分支, 将原输入划分为两个分支, 分别进行卷积操作, 最后融合这两个分支, 使模型提取出更丰富的特征, 精简了网络结构; 快速空间金字塔池化模块是空间金字塔池化模块的一种快速实现方式如图 3 所示, 包括卷积模块、最大值池化层和拼接, 先由卷积

模块提取特征输出, 再依序进入最大值池化层采样, 然后拼接融合各个尺度的特征, 最后经过卷积模块, 提高送入特征融合网络时所获得的信息量。

为了提升模型对重要特征的敏感程度, 在主干网络中融合了注意力机制, 新增注意力模块, 结构如图 4 所示, 对每一个特征图做全局池化处理, 通过加权方式强化有用特征, 抑制无用特征, 为特征融合阶段输入更好的特征。

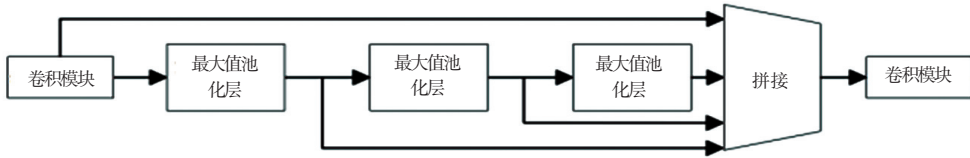


图 3 快速空间金字塔池化模块

Fig. 3 Spatial Pyramid Pooling-Fast component structure

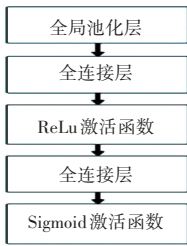


图 4 注意力模块结构

Fig. 4 Squeeze-and-Excitation module structure

### 2.4 特征融合

主干网络提取图像特征, 经过的卷积层数越多, 提取到目标的类别和颜色等语义信息越丰富, 但同时目标处于原始图像中的位置信息会越来越模糊。为了同时保留目标的语义信息以及位置信息, 该网络在特征融合部分使用了特征金字塔(FPN)加路径聚合网络(PAN)结构, 特征金字塔+路径聚合网络结构如图 5 所示。

度的特征信息, 80×80 维度上保留 40×40 维度上的信息, 这样做可以将高层的特征保留下来, 有利于小目标的检测。而 PAN 结构与 FPN 结构相反, 是一种自底向上的路线, 其在 FPN 融合出的 3 个维度的基础上从 80×80 维度开始提取信息给上层, 保证高层也拥有较为丰富的位置信息。

两种结构互相弥补缺点, 构成了一种语义信息与位置信息都较为丰富的多尺度、多层次特征金字塔。

### 2.5 输出端

输出端将特征融合网络中得到的各尺度特征图分别使用 1×1 卷积核扩展通道数, 扩展后通道数为  $(class + obj + x + y + w + h) * len(anchors)$ , 其中  $class$  表示类别总数,  $obj$  为目标的置信度,  $x$  为预测框的中心横坐标,  $y$  为预测框的中心纵坐标,  $w$  为预测框的宽度,  $h$  为预测框的高度,  $len(anchor)$  是每个检测层上锚框的数量, 带入本模型设定的类别与锚框数量, 最后输出通道数为 30 的特征图。经过损失函数计算与非极大值抑制后框出目标框。

### 2.6 改进损失函数

本文改进了 YOLOv5 中的默认边界框损失函数, 更换为考虑真实框与预测框之间匹配方向的 SIoU 损失函数。该损失函数由角度、距离、形状和交并比 4 种惩罚项组成, 角度的引入可以使预测框更快速的向  $x$  轴或  $y$  轴的一侧移动, 加快收敛速度<sup>[4]</sup>。SIoU 损失函数定义如式(1)所示:

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (1)$$

其中,  $IoU$  是真实框与预测框之间的交并比;  $\Delta$  为距离成本, 该距离成本考虑了两框之间的角度影响;  $\Omega$  为形状成本。

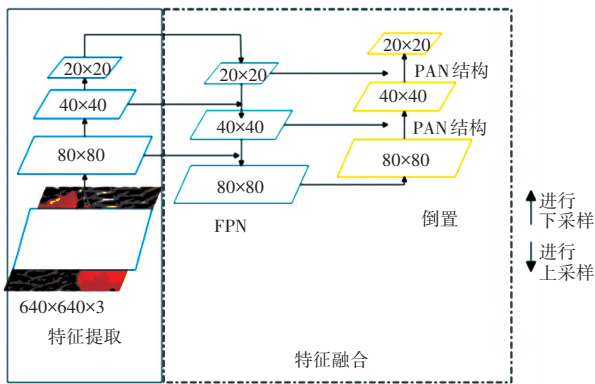


图 5 特征金字塔+路径聚合网络结构

Fig. 5 Feature pyramid network+path aggregation network structure

FPN 结构是一个自顶向下的特征金字塔, 其将提取到的不同维度的信息从高层到底层自上而下进行融合, 如在 40×40 这个维度上保留部分 20×20 维



### 3 实验

#### 3.1 实验环境

实验使用的 CPU 是英特尔®酷睿 i5-7300HQ, 2.50 GHZ, GPU 是英伟达 GTX1060, 显存 6.0 GB, 运行内存 14 GB, 操作系统为 Windows10-64 位, CUDA 版本 11.4, Python 版本为 3.7, 深度学习框架 Pytorch1.11。

#### 3.2 评价指标

选用平均精度均值 ( $mAP$ ) 和单张图片检测耗时作为评价指标。平均精度均值 ( $mAP$ ) 由各类别在给定交并比下的平均精度 ( $AP$ ) 平均得到, 平均精度均值、平均精度与模型的准确率 ( $P$ ) 和召回率 ( $R$ ) 的关系如式(2) ~ 式(5) 所示:

$$P = \frac{T_p}{T_p + F_p} \quad (2)$$

$$R = \frac{T_p}{T_p + F_N} \quad (3)$$

$$AP = \int_0^1 P(R) dr \quad (4)$$

$$mAP = \frac{1}{N} \sum_{i=0}^{N-1} AP_i \quad (5)$$

其中,  $N$  为橘瓣类别个数;  $T_p$  是被正确检测出的目标个数;  $F_p$  是被错误检测出的目标个数;  $F_N$  是被漏检的目标个数。

#### 3.3 模型训练

将训练轮次设定为 550 轮, 考虑到模型需要较好的泛化能力将迭代批次设置成为 16。训练过程中使用带有动量的随机梯度下降法对模型参数进行更新, 可以加快模型的训练。改进 YOLOv5s 训练参数见表 1, 改进 YOLOv5s 训练曲线如图 6 所示。

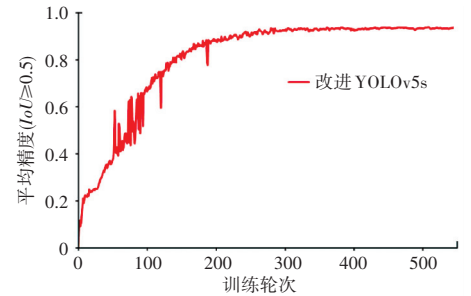
表 1 改进 YOLOv5s 训练参数

Table 1 Improved YOLOv5s training parameters

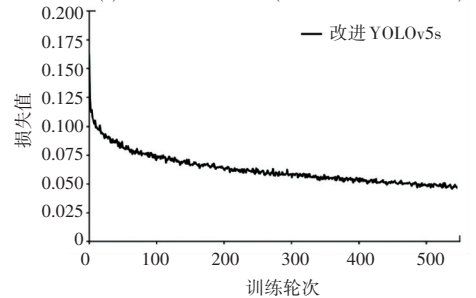
参数	数值
轮次 (Epochs)	550
批大小 (Batch)	16
初始学习率 (Lr0)	0.01
周期学习率 (Lrf)	0.2
动量 (Momentum)	0.9
权重衰减项 (Weight_decay)	0.000 5
置信度阈值 (Conf_thres)	0.5
交并比阈值 (Iou_thres)	0.5

整个训练过程中学习率的周期性调整比重不大, 因此损失值的下降以及平均精度均值的上升都没有出现断崖。在前 100 轮中, 损失值迅速下降至 0.075, 平均精度均值也迅速上升至 60%, 在第 100 至 300 轮中, 损失值的下降速度开始减缓, 接近

0.06, 而平均精度均值达到 90%, 迭代 300 轮往后, 损失值保持在 0.05 上下, 且平均精度均值达到最高值 93.7% 后不再上升, 模型已经达到收敛状态。



(a)  $mAP$  值变化曲线(交并比大于等于 0.5)



(b) 损失值变化曲线

图 6 改进 YOLOv5s 训练曲线

Fig. 6 Improved YOLOv5s training curve

#### 3.4 实验结果分析

本实验检测目标的 5 种类别橘瓣的测试结果见表 2。

表 2 各类橘瓣实验结果

Table 2 Results of various orange petal experiments

类别	准确率/%	召回率/%	单张图检测时间/ms
完好橘瓣	95.9	95.1	25
有核橘瓣	99.9	89.5	25
有囊皮橘瓣	90.0	83.7	25
未分离橘瓣	85.9	98.0	25
破碎橘瓣	91.7	78.2	25

有核橘瓣的精确度高达 99.9%, 同时召回率也有 89.5%, 为所有种类中最佳, 因为有核橘瓣的形状最为特殊, 一般在橘瓣的中心部分会有一个黑色的核, 特征明显, 因此最容易识别; 完好橘瓣的精确度也达到了 95.9%, 召回率有 95.1%, 因为数据量最大, 且完好橘瓣形状规则, 表面光滑, 因此也较容易识别; 未分离橘瓣在实际生产线上数量较少, 且两瓣橘瓣堆叠起来容易误识别为两个完好橘瓣, 因此识别率较低。该模型的平均精度均值 ( $mAP$ ) 为 93.7%, 并且运行速度达到了 25 ms/图, 可以快速且精准的完成分类任务。

为了验证改进 YOLOv5 模型的识别性能, 将改进 YOLOv5 模型与 YOLOv5s 模型、Faster-RCNN 模

型进行对比实验。对比实验过程中使用的数据集、实验环境保持一致, 各个模型的训练参数设置相同, 各模型的对比实验结果见表 3。

表 3 各模型对比实验结果

Table 3 The experimental results of each model are compared

模型	平均精度(交并比 阈值>0.5)/%	运行速度/ ms	模型大小/ MB	训练时长/ h
YOLOv5s	92.4	25.0	13.69	9.53
Faster-RCNN	90.54	67.5	113.5	17.64
改进 YOLOv5s	93.7	25.0	14.4	9.83

从表 3 可以看出, 改进 YOLOv5s 在模型大小和训练时长上有所增加, 但是平均精度(*mAP*) 增加了 0.3%, 相比 YOLOv5s 模型没有速度上的提升, 但是快于 Faster-RCNN, 改进 YOLOv5s 模型具有更好的性能。

将改进 YOLOv5s 模型与 YOLOv5s 模型在流水线上的检测效果对比与在传送带上的检测效果对比如图 7 和图 8 所示。在流水线应用场景中, 改进 YOLOv5s 模型的置信度整体优于 YOLOv5s; 在传送带应用场景中, 改进 YOLOv5s 模型虽然个别目标检测出的置信度低于 YOLOv5s, 但是没有出现误检情况, 具有更好的可靠性。

综上所述, 改进 YOLOv5s 在柑橘罐头厂加工橘子罐头时具有更好的适用性, 橘瓣的分类检测精确度提升, 同时在部分场景下的漏检率和误检率明显下降, 更适合工厂生产线上安装使用。

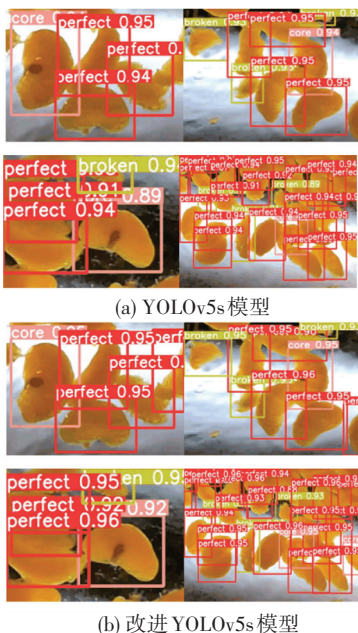
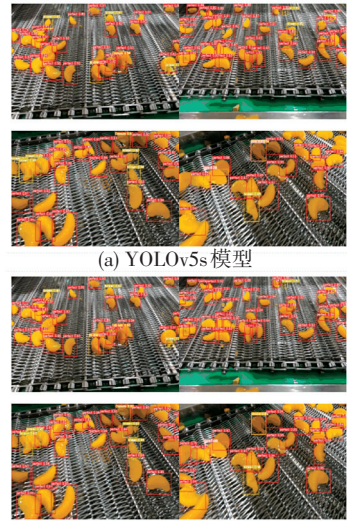


图 7 流水线上的检测效果对比

Fig. 7 Comparison of detection effects on the assembly line



(b) 改进 YOLOv5s 模型

图 8 传送带上的检测效果对比

Fig. 8 Comparison of inspection results on conveyor belts

### 4 结束语

为解决柑橘罐头工厂加工橘子罐头时的橘瓣实时检测问题, 本文使用基于深度学习的改进 YOLOv5s 模型对各类橘瓣进行检测。通过在罐头工厂内实地进行样本采集, 建立了完好橘瓣、有核橘瓣、有囊皮橘瓣、未分离橘瓣以及破碎橘瓣的数据集。通过改进 YOLOv5s 模型网络结构、引入注意力机制与 Siou 损失函数构建更适合橘瓣实时分类的网络模型, 使用该数据集训练网络并测试, 得到平均精确度 (*mAP*) 达 93.7%, 单张图推理速度达 25 ms 的网络模型。

通过对比实验的方式验证模型的性能。实验结果表明, 改进 YOLOv5s 模型具有更高的准确率, 并且在部分场景下误检率和漏检率更低。

综上所述, 本文基于改进 YOLOv5s 模型的橘瓣检测方法能够满足罐头工厂实际生产时的实时性、准确性要求, 能够为机器自动化分拣各类橘瓣提供视觉解决方法, 对罐头工厂的机器自动化发展具有一定的推进意义。

### 参考文献

- [1] 方修贵, 黄洪舸, 曹雪丹, 等. 柑橘加工常用技术(1)——柑橘果实的加工特性及加工产品[J]. 浙江柑橘, 2018, 35(1): 37-40.
- [2] 李琳芳, 王建军, 魏征, 等. 基于卷积神经网络的橘子分类识别研究[J]. 河南科技学院学报(自然科学版), 2020, 48(3): 68-73.
- [3] 李洪波, 廖详刚, 陈立. 基于机器学习 One-stage 目标检测算法的塑料自动识别系统[J]. 塑料技术, 2020, 48(12): 86-89.
- [4] GEVORGYAN Z. Siou Loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.