

王景浩. 基于 YOLOv5s 的单检测框多标签检测方法研究[J]. 智能计算机与应用, 2025, 15(3): 219–222. DOI: 10.20169/j.issn.2095-2163.250333

# 基于 YOLOv5s 的单检测框多标签检测方法研究

王景浩

(西安石油大学 计算机学院, 西安 710065)

**摘要:** 当前的目标检测方法通常关注的是检测对象的单一类别标签,然而在实际场景中,经常需要识别对象的多个相关标签,从而为对象提供更全面的特征描述。本研究基于 YOLOv5s 框架开发一种单检测框多标签检测方法,实现在单一检测框内同时识别和标记多个相关标签。为提高多标签分类准确性和效率,提出联合预测策略。使用引入层次权重和层级联合性的损失函数,通过权衡不同任务的权重、考虑不同标签的层级关系,为相关标签建立联系。使用联合预测策略相比使用前  $mAP@0.5$  提升了 4.9%。该研究为基于 YOLOv5s 实现多标签检测方法,为多标签检测提供了新的视角和改进方向,对实际对象识别和分类具有重要的实用价值。

**关键词:** 单检测框多标签检测; YOLOv5s; 联合预测; 层次权重

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2025)03-0219-05

## Research on single detection box multi-label detection method based on YOLOv5s

WANG Jinghao

(School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

**Abstract:** Current object detection methods typically focus on detecting single class labels for objects. However, in real-world scenarios, it is often necessary to identify multiple relevant labels for an object in order to provide a more comprehensive feature description. The paper develops a single detection box multi-label detection method based on the YOLOv5s framework, which enables simultaneous recognition and labeling of multiple relevant labels within a single detection box. To improve the accuracy and efficiency of multi-label classification, a joint prediction strategy is proposed. The paper uses a loss function that incorporates hierarchical weights and hierarchical associations, balancing the weights of different tasks and considering the hierarchical relationships between different labels to establish connections between relevant labels. Using the joint prediction strategy, the paper achieves a 4.9% improvement in  $mAP@0.5$  compared to previous methods. This research provides a new perspective and direction for implementing multi-label detection based on YOLOv5s and holds significant practical value for real-world object recognition and classification.

**Key words:** single detection box multi-label detection; YOLOv5s; joint prediction; hierarchical weight

## 0 引言

随着机器视觉和深度学习的快速发展,基于深度学习的单标签目标检测技术(如 YOLO、Faster R-CNN、SSD 等)取得了长足进步<sup>[1]</sup>。现今这些技术已成功应用于多个领域。然而在实际应用中,传统的单标签目标检测技术在处理具有多个标签的物体时存在局限性。实际上一个物体往往具有多个相关标签<sup>[2]</sup>,但传统检测技术往往只能预测一个主要标签,为此就限制了这些技术在真实世界中的应用。

当涉及到实际应用时,单检测框多标签检测展现出其不可或缺的重要性和研发需求。在肺部影像分析中,可以标记肺部病变类型的同时,判断病变大小、形状以及是否为结节、肿块等。帮助医生评估肺部疾病的性质和进展。在交通监控系统中,判断车辆品牌的同时,检测车型、颜色和车辆状态等信息。通过捕捉车型和车辆状态的关联性,提供对交通拥堵更加准确的分析预测。

研究提出的基于 YOLOv5 的单检测框多标签方法,在实现多分类的同时,单检测框满足了实际应用

基金项目: 陕西省重点研发计划(2022GY-035)。

作者简介: 王景浩(1999—),男,硕士研究生,主要研究方向:计算机视觉。Email: wangjh026@163.com。

收稿日期: 2023-09-09

中需要定位的需求。考虑多个标签之间相关性问题,提出联合预测策略,增强神经网络对标签之间语义相关性学习。并在足球机器人系统上进行了相关实验验证。

## 1 基于 YOLOv5s 实现单检测框多标签检测

目标检测的核心任务是定位并分类图像中的物体。基于深度学习的目标检测算法取代了传统的手动特征选择,分为 Two-stage 目标检测和 One-stage 目标检测两种方法。其中,Two-stage 方法首先生成区域建议(region proposal),随后使用卷积神经网络对这些区域进行分类。一些著名的 Two-stage 目标检测算法包括 RCNN<sup>[3]</sup>、Fast R-CNN<sup>[4]</sup>、Faster R-CNN<sup>[4]</sup> 和 Mask R-CNN<sup>[5]</sup> 等。而 One-stage 方法则直接通过深度卷积神经网络进行目标的定位和分类,省略了生成候选区域的过程。常见的 One-stage 目标检测算法包括 YOLO 系列<sup>[6-8]</sup>、SDD<sup>[9]</sup> 和 FSSD<sup>[10]</sup> 等。

在处理多标签任务时,Two-stage 目标检测方法在候选框区域生成和再定位过程中较为细致,因此更适应多标签问题的定位和分类需求。然而,由于这种精细的设计,使其计算复杂度较高,训练和推理时间相对较长。相比之下,One-stage 方法能够在单一阶段完成分类和定位,具备较高的计算效率,但其设计更偏向于单标签目标,对于多标签任务的适应性相对较低。

为了保留目标检测实时性,本研究基于 YOLOv5s 框架进行实现,通过修改数据标注、预处理阶段、非极大值抑制方法等,有效解决了多标签任务适应性的问题。

### 1.1 数据标注

标注数据是使用 YOLOv5 时的关键组成部分,应当满足一定的格式要求,每个标注行主要包含以下内容:类别索引、边界框中心坐标相对于图像宽度和高度的比例、以及边界框的宽度和高度相对于图像宽度和高度的比例。

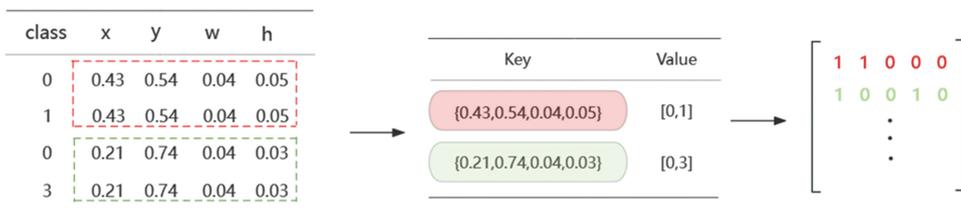


图1 改进预处理模块

Fig. 1 Improved preprocessing module

处理多标签任务时,在特定标注格式的基础上进行修改,以适应多标签任务。如果某个目标具有多个标签,只需修改其类别标签,而保持边界框坐标、中心点和尺寸不变。

在进行标注时,要确保每个标注与相应的图像匹配,以便让 YOLOv5 正确地学习和预测图像中的多标签信息。通过合理标注数据,YOLOv5 可以在多标签任务中准确地识别和定位图像中的目标,为多个属性或特征提供全面的描述。例如单个目标属于类别 0、1、2,则该目标标注格式见表 1。

表1 数据标注格式

Table 1 Data annotation format

Class	x	y	w	h
0	0.429 688	0.538 542	0.037 500	0.047 917
1	0.429 688	0.538 542	0.037 500	0.047 917
2	0.429 688	0.538 542	0.037 500	0.047 917

### 1.2 预处理阶段

修改数据标注后,如果直接输入网络进行预测,则和单标签任务没有区别。为了适应多标签任务,本研究对 YOLOv5 的数据预处理部分进行了修改。

首先,将标注信息整合到有序字典中。该有序字典的键包括了边界框坐标以及图像的宽度和高度,而类别标签则作为有序字典的值。这样的数据结构确保了目标和其对应类别标签之间的一一对应关系。同时,还在构建有序字典的过程中生成了一个列表,用于存储与每个目标相关的类别标签。这一步骤保证了类别标签与其对应的目标的一致性。

其次,在数据预处理完成后,为每个检测框的类别索引所对应的位置分配了正样本权重,其他位置则被赋予了负样本权重。

预处理阶段修改后,确保了在进行回归计算时,每个检测框内的多个标签都能够被同时送入损失函数。使得 YOLOv5 能够更好地适应多标签任务,为神经网络学习单个检测框内多个标签的需求提供了有效的途径,减少了标签关联性的损失。若类别数量为 5,且正样本权重为 1,负样本权重为 0,则预处理过程如图 1 所示。

### 1.3 非极大值抑制方法

在单标签的目标检测中,每个候选框在经过非极大值抑制(NMS)之前,通常会选择其类别预测概率中的最大值作为候选框的类别标签。

本研究采用了一种不同的方法以适应多标签情况。得到检测框的类别预测信息后,使用置信度阈值来筛选出一些候选框,将其中大于置信度阈值的预测类别标签与对应的检测框信息取出,例如一物体的预测类别标签概率值,有 2 类大于阈值,则类别与检测信息分别组合成 2 个单标签检测框。

这种分离方法将检测框的格式重新还原成了 YOLOv5 所需的标签信息格式。由于在分离时将同一样本的类别标签与检测框信息一同分离出来,因此确保了同一样本的每个类别都有相同的检测框信息。这一创新的方法不仅使得多标签目标检测变得可行,还在保持非极大值抑制效果的前提下,实现了更精确的多标签检测结果。

## 2 完善单检测框多标签检测方法

在将单检测框多标签任务引入 One-stage 目标检测算法中,面临着许多挑战和限制。其中,标签关联问题即是目前备受关注的研究热点之一。因为单一检测框需要同时预测多个标签,而这些不同标签之间可能存在着复杂的关联性。同时,标签不平衡问题也是目前亟待探索的研究内容之一。

针对这些问题,本研究提出了一种创新的解决

方案,使用引入层级权重和层级关联性的损失函数。

### 2.1 标签层级关系

在多标签任务中,不同标签之间可能具有一定的层次结构或相关性。父子类层级关系是其中的一种关系,指的是某个标签被视为另一个标签的“父类”或“子类”。这种层级关系用于描述标签之间的从属关系或上下级联系,通常是在语义或逻辑上存在联系的标签之间建立的。

例如本实验中,足球机器人色标设计如图 2 所示,正方向朝向情况如图 3 所示。

在车号标签判断中,中间矩形块 A 部分为队标部分,颜色分为黄色或蓝色两队;矩形块 B 和矩形块 C 为色标辅助色块,根据其颜色判断黄队或蓝队中的具体车号。

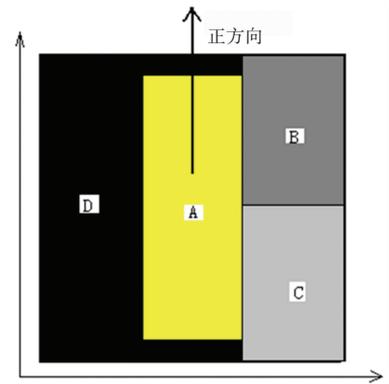


图 2 足球机器人色标设计  
Fig. 2 Soccer robot color code design

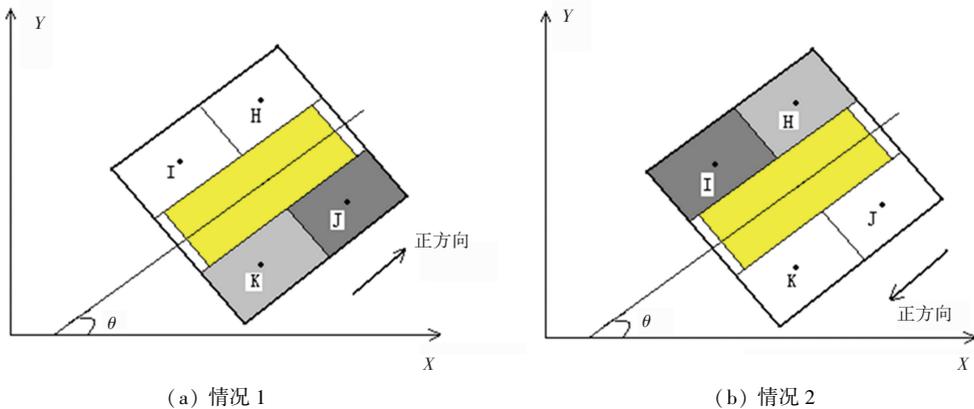


图 3 机器人正反向与色标关系

Fig. 3 The relationship between forward and reverse direction of robot and color code

在本研究中,机器人的正方向朝向是根据队标的长轴与 B 部分及 C 部分的位置关系来判定的。具体而言,如果长轴的右侧带有辅助色块,则确定该方向为正方向。在本文的检测任务中,首先根据色标判断出车号,接着根据所在位置来判断机器人的

正方向朝向。

为了实现这一判定过程,将车号标签设置为父标签,而正方向朝向设置为子类标签。这种层级关系体现了父标签包含了子类标签所具有的特征和属性,因此子类标签的预测结果对于父类标签的预测

也具有一定的影响。为了保持标签之间的关联性,文中对损失函数进行了优化,引入了层次权重和层级关联性的概念。

## 2.2 改进损失函数

在本次检测研究中,为适应多标签任务,分类损失采用引入层级权重和层级关联性的交叉熵损失函数。根据父标签  $P$  和子标签  $C$  使用以下损失函数来融入层级信息。父标签交叉熵损失如下所示:

$$L_P = -(y_P \times \log(p_P) + (1 - y_P) \times \log(1 - p_P)) \quad (1)$$

子标签交叉熵损失可由下式来表示:

$$L_C = -(y_C \times \log(p_C) + (1 - y_C) \times \log(1 - p_C)) \quad (2)$$

在父子标签之间预测差异引入关联性损失项。层级关联性损失可使用如下公式来计算:

$$L_{rel} = (p_P - p_C)^2 \quad (3)$$

联合损失函数的计算公式如下:

$$L_{joint} = w_P \times L_P + w_C \times L_C + \lambda \times L_{rel} \quad (4)$$

层级联合性损失 ( $L_{rel}$ ), 通过计算父类概率 ( $p_P$ ) 和子类概率 ( $p_C$ ) 之间的平方差来衡量层级关联性。通过最小化层级关联性损失, 模型被引导去学习标签之间的关联性, 有助于模型在训练时考虑到不同标签之间的语义和层级关系, 提高多标签检测任务的准确性和效果。

联合损失函数 ( $L_{joint}$ ), 可以通过增加子类标签权重的方法, 在准确预测子类标签的同时, 分析可知子类标签包含父类标签的一部分特征和属性, 即可将通过层级关系传递给父类标签的预测, 因此提升了父类标签的准确性。

对于类别数量不平衡问题, 联合损失函数可以针对样本数量较少的类别增加类别权重, 使得训练时关注样本较少的类别, 从而解决了类别不平衡问题。

## 3 实验结果与分析

### 3.1 实验环境和实验数据

本实验采用 Intel Core i5-8300 处理器, 内存为 16 G, 显卡为 NVIDIA GeForce GTX 1050Ti Laptop GPU, 其显示内存大小为 4 GB。程序编程采用 Python 语言和 Pytorch 深度学习框架, 并使用 GPU 加速软件 CUDA10.1、Python3.8、Pytorch1.8。在该实验中, 采用的数据增强方法有缩放、平移以及 Mosaic 数据增强等。

本实验使用数据集共包含 1 500 张 5V5 足球机

器人比赛图像, 以人工方式标注定义机器人车号类别以及车头正方向朝向类别。共 14 个类别, 其中 1~10 作为足球机器人的车号类别标签, 定义机器人车头朝向与  $X$  轴夹角小于  $180^\circ$  为正方向。其余 4 个类别分别为: ‘Positive’ 为正反向标签, ‘Negative’ 为反方向, ‘Right’ 为车头朝向与  $X$  轴正方向相同, ‘Left’ 为朝向与  $X$  负方向相同。

### 3.2 模型训练

在模型训练阶段, 采用了一系列参数设置以达到最佳的性能。训练集和验证集比例为 7:3, 迭代批量大小设置为 8, 并使用初始学习率为 0.01 的 SGD 优化算法, 其中动量为 0.937, 衰减系数为 0.0005, 设定最大迭代 Epoch 为 120。以确保模型充分学习数据特征。

在引入层次权重的损失函数参数方面, 经过多次实验和反复验证, 发现 0.5 和 2.0 的权重设置在训练过程中表现出色, 因此将 0.5 作为父类的权重, 以减少父类标签在训练中的影响, 因为父类标签通常具有更广泛的语义涵义, 允许在训练过程中有更多的误差容忍度。相反, 将子类权重设为 2.0, 以强调对子类标签的学习, 更好地捕捉子类的特征。这是因为子类标签包含父类标签的一部分特征和属性, 通常具有更具体的语义信息, 对模型的预测更为敏感。

### 3.3 实验结果

在损失函数加入层级权重和层级关联性前与加入后的 YOLOv5s 在多标签检测应用场景下, MAP 数据对比如图 4 所示。图 4 中, YOLOv5s 的 MAP@0.5 达到了 0.932, 加入后的 YOLOv5s 的 MAP@0.5 达到了 0.981, 提升了 4.9%。并且算法收敛速度得到提升。

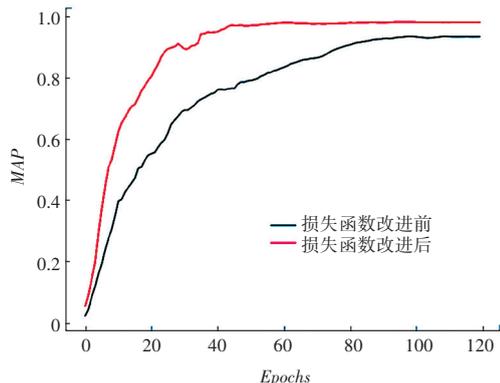


图4 损失函数改进前后算法 MAP@0.5 值对比

Fig. 4 Comparison of MAP@0.5 values before and after loss function improvement

为验证本研究提出的在多标签检测下的改进方案,在相同的实验环境与数据集下,进行消融实验。实验结果见表2。由表2可以看出,在改进损失函数后,准确率 (*Precision*) 达到 0.98。召回率 (*Recall*) 提升到 0.97。可见改进的方案相比原 YOLOv5s 损失函数在多标签检测效果上具有明显优势。

表 2 算法检测性能对比

Table 2 Algorithm detection performance comparison

模型	<i>Precision</i> /%	<i>Recall</i> /%	<i>MAP</i> /%	<i>F1</i>
YOLOv5s	0.90	0.94	0.93	0.92
完善后 YOLOv5s	0.98	0.97	0.98	0.98

为了展示改进的 YOLOv5s 网络模型的检测效果,从测试集中随机选取图片进行测试,原 YOLOv5s 网络模型与损失函数改进后的 YOLOv5s 网络模型在多标签任务下对比结果如图 5、图 6 所示。



图 5 改进前检测结果

Fig. 5 Test results before the improvement



图 6 改进后检测结果

Fig. 6 Test results after the improvement

完善前 YOLOv5s 模型在多标签检测下检测小车的车号和车头正方向状态时,平均检测精度分别为 0.876 和 0.836。然而,在改进损失函数后,平均检测精度得到了提升,分别达到了 0.886 和 0.883。这表明在单一检测框多标签检测任务中,经过改进的 YOLOv5s 网络模型在多标签检测下展现出更高

的检测精度。

值得注意的是,改进的方法不仅在检测精度方面取得了提升,同时标签之间的关联性也得以增强。这意味着改进后的模型更能适应多标签任务,可以捕捉标签之间更为合理的组合,从而对图像中的实际情况能进行更准确的组合检测。

## 4 结束语

本文研究提出的基于 YOLOv5s 框架的单检测框多标签检测方法,在经过损失函数的改进后,解决了 One-stage 目标检测在多标签任务上不适用的问题,进一步实现了在单一检测框的基础上同时识别和定位多个标签信息,并且为对象检测任务的多属性特征提供了更全面、准确的描述。本文成果有助于在对象检测领域中深入探索多标签检测的方法和技术,而且也将对后续多标签检测方法的研究和应用提供有益的借鉴参考价值。

## 参考文献

- [1] 张阳婷,黄德启,王东伟,等. 基于深度学习的目标检测算法研究与应用综述[J]. 计算机工程与应用,2023,59(18):1-13.
- [2] TSOU MAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data [M]// Data Mining and Knowledge Discovery Handbook. Cham: Springer, 2009: 667-685.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2014:580-587.
- [4] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2015:1440-1448.
- [5] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems. Montreal, Canada: NIPS Foundation, 2015:91-99.
- [6] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ:IEEE, 2017:2980-2988.
- [7] HAN Xiaohong, CHANG Jun, WANG Kaiyuan. Real-time object detection based on YOLO-v2 for tiny vehicle object[J]. Procedia Computer Science,2021, 183:61-72.
- [8] ZHAO Liqun, LI Shuaiyang. Object detection algorithm based on improved YOLOv3[J]. Electronics,2020,9(3):537.
- [9] YU Jimin, ZHANG Wei. Face mask wearing detection algorithm based on improved YOLOv4[J]. Sensors,2021,21(9):3263.
- [10] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [M]//European Conference on Computer Vision. Cham: Springer,2016:21-37.
- [11] LI Zuoxin, LU Yang, ZHOU Fuqiang. FSSD: Feature fusion single shot multibox detector[J]. arXiv preprint arXiv,1712.00960, 2017.