



山西大学
SHANXI UNIVERSITY

2024 届硕士学位论文

基于深度学习的安全帽与反光衣检测研究

作者姓名	张学立
指导教师	贾新春 教授
学科专业	电子信息
研究方向	目标检测、深度学习
培养单位	自动化与软件学院
学习年限	2021 年 9 月至 2024 年 6 月

二〇二四年六月

山西大学

2024 届硕士学位论文

基于深度学习的安全帽与反光衣检测研究

作者姓名	张学立
指导教师	贾新春 教授
学科专业	电子信息
研究方向	目标检测、深度学习
培养单位	自动化与软件学院
学习年限	2021 年 9 月至 2024 年 6 月

二〇二四年六月

Thesis for Master's degree, Shanxi University, 2024

**Research on Intelligent Detection of Safety Helmet and
Reflective Clothing Based on Deep Learning**

Student Name	Xue-li Zhang
Supervisor	Prof. Xin-chun Jia
Major	Electronic Information
Specialty	Object Detection, Deep Learning
Department	School of Automation and Soft- ware Engineering
Research Duration	2021.09-2024.06

June, 2024

摘 要

安全帽和反光衣在诸多行业已经成为基本的安全装备，这些装备能够保护工作人员的安全。在建筑工地、工厂、交通等危险作业环境中，工人佩戴安全帽可以减少头部伤害。反光衣在光线较暗或者夜间工作时，能够反射光线，增加工人的可见性，减少被车辆或其他移动机械碰撞的风险。通过智能监控系统自动检测工人是否佩戴了安全帽和反光衣，可以帮助管理者及时发现并纠正不安全行为，从而预防事故的发生。在智能监控中，能够准确识别和追踪佩戴安全帽和反光衣的人员具有重要的意义。

为此制作安全帽与反光衣数据集，而且分成了四类佩戴情况，以适应实际的检测环境。并且以 YOLOv5s 为基准，设计了两种改进的安全帽与反光衣检测算法。

其一，将 YOLOv5s 的网络中的普通卷积替换为 GhostConv，并且将 C3Ghost 模块代替部分 C3 模块；并且在网络中加入了 CA 注意力机制，提高对目标的关注；使用 WIoU 损失函数。通过这些改进方式，降低了模型的大小，减少了的计算资源。使其更容易部署在资源受限的环境中，且集成到各种平台和设备上。

其二，将 YOLOv5s 的网络模型中颈部网络中 C3 模块替换为 C3SAC 模块，提高其特征提取能力；引入注意力机制模块；替换损失函数为 SIoU；使用轻量级上采样算子 CARAFE，更好地保留特征信息。通过以上方式，提高了检测的平均精度。在数据集上表现良好。

本文提出的两种改进方式，侧重点不同，以此来适应复杂多变的使用环境。不仅提高了工作人员的个人安全，而且对于构建智能、高效、安全的作业环境具有重要意义。同时也帮助管理人员更好地了解工人的分布和活动，优化资源分配，提高工作效率。

关键词： 安全帽与反光衣；安全检测；深度学习；轻量化算法

ABSTRACT

Helmet and reflective clothing have become basic safety equipments in many industries, and these equipments can protect the safety of workers. In dangerous working environments such as construction sites, factories, and traffic, workers wearing helmets can reduce head injuries. Reflective clothing can reflect light when dim or working at night, increasing the visibility of workers and reducing the risk of collision by vehicles or other moving machines. Through the intelligent monitoring system to automatically detect whether the workers wear hard hats and reflective clothing, it can help the managers to find and correct the unsafe behaviors in time, so as to prevent accidents. In the intelligent monitoring, it is of great significance to accurately identify and track the people wearing hard hats and reflective clothing.

A dataset of safety helmets and reflective clothing has been created for this purpose, and it has been categorized into four types of wearing conditions to adapt to an actual detection environment. Based on YOLOv5s, two improved target detection algorithms are designed in this thesis.

First, replace the convolution in the YOLOv5s network model with GhostConv, and the Ghost module replaces the original partial convolution with the C3 module; and add the CA attention mechanism in the network to increase the attention to the target table; and use the WIoU loss function. Through these improvements, the size of the model is reduced and the computing resources are reduced. Make it easier to deploy in resource-constrained environments and integrate into a variety of platforms and devices.

Second, replace C3 module in the neck network of YOLOv5s network model with C3SAC module to improve its feature extraction ability; introduce attention mechanism module; replace loss function as SIOU; and the lightweight upsampling operator CARAFE is used to better retain feature information. Through the above methods, the average accuracy of the detection is improved. Performs well on the data set..

This thesis proposes two improvement methods with different emphasis points, in order to adapt to the complex and changeable applied environment. It not only improves the personal safety of workers, but also is of great significance to the construction of an

intelligent, efficient and safe working environment. At the same time, it also helps managers to better understand the distribution and activities of workers, optimize resource allocation, and improve work efficiency.

Key words: Safety helmet and reflective clothing; Safety detection; Deep learning; Lightweight algorithm

目 录

1 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 目标检测技术发展现状.....	2
1.2.2 安全帽与反光衣检测研究.....	4
1.3 主要研究内容.....	6
1.4 论文框架.....	6
2 相关知识介绍	8
2.1 深度学习简述.....	8
2.2 卷积神经网络.....	9
2.3 YOLO 检测算法.....	11
2.3.1 YOLOv1	11
2.3.2 YOLOv2	12
2.3.3 YOLOv3	13
2.3.4 YOLOv4	15
2.4 本章小结.....	15
3 基于 Ghost 改进的 YOLOv5 检测算法.....	16
3.1 引言.....	16
3.2 YOLOv5s 算法	16
3.3 改进的 YOLOv5s 算法	18
3.3.1 Ghost 模块的引入	18
3.3.2 CA 注意力机制的引入	19
3.3.3 WIOU 损失函数的替换	21
3.3.4 轻量化 YOLOv5s 网络模型	22
3.4 实验结果与分析.....	22
3.4.1 实验环境.....	22
3.4.2 数据集.....	23
3.4.3 评估指标	25
3.4.4 实验结果.....	26
3.4.5 消融实验.....	27
3.4.6 对比实验.....	28
3.5 本章小结.....	30
4 基于 SACConv 改进的 YOLOv5s 的检测算法.....	31
4.1 引言.....	31

4.2 改进的 YOLOv5 算法	31
4.2.1 SAConv 模块	31
4.2.2 SIoU 损失函数	32
4.2.3 CARAFE 轻量级上采样算子	33
4.2.4 改进后的 YOLOv5s 网络结构	33
4.3 实验结果与分析	35
4.3.1 实验结果	35
4.3.2 消融实验	36
4.3.3 对比实验	36
4.4 本章小结	39
5 总结与展望	40
5.1 总结	40
5.2 展望	40
参考文献	42

1 绪论

1.1 研究背景及意义

随着时代的发展，人们安全意识已不断提高，以人为本的理念逐步地深入人心，对工作人员的安全防护管理逐步成为各施工单位、交通作业等部门的工作重点。

根据欧盟统计局和 IBS 于 2020 年的统计，建筑工地的头部伤害数量占致命事故的 30% 以上，可见佩戴安全帽对建筑工人的生命安全至关重要^[1]。安全帽可以防护头部，在发生撞击、碰撞或高处坠物时，为头部提供保护，减少或避免头部受到伤害。在建筑业、制造业、交通运输等行业中，安全帽可以有效地减少头部伤害。目前在许多国家和地区，法律规定在某些危险工作环境中必须佩戴安全帽，以保障工人的安全。反光衣通常由特殊的反光材料制成，能够在光线照射下反射出明亮的颜色，从而在夜间或低光照条件下提高穿着者的可见性。在道路施工、夜间作业等环境中，反光衣可以帮助工人或行人被司机及时发现，减少被车辆撞击的风险，预防事故发生。反光衣不仅适用于工地，还广泛用于交通指挥、环卫清洁、紧急救援等多种场合。安全帽与反光衣的结合使用，可以为工作者提供全面的个人防护，减少事故发生的可能性，保护生命安全。而通过采用先进的安全管理技术，可以提升企业的安全管理形象，增强公众对企业的信任和认可。

通过安全帽与反光衣的检测技术，不仅可以确保工人在施工现场始终佩戴适当的个人防护装备，从而降低事故发生的风险，提高安全管理水平。通过实时监测和预警，可以及时纠正未佩戴的行为，预防潜在的事故。

传统的安全管理依赖于人工巡查，不仅耗时耗力，而且难以实现全天候、全方位的监控。依托于深度学习的目标检测技术，能够实现对监控视频的智能分析处理，自动识别不合规的行为，并实时反馈。这种智能化的管理方式有助于构建更加科学、高效的安全生产管理体系。目标检测技术则也可以在各种环境条件下稳定工作，减少因人为因素导致的误检和漏检。通过目标检测，可以减少对人工巡查的依赖，节省人力资源，降低管理成本。安全帽与反光衣检测研究也可以推动了计算机视觉、图像识别等技术的发展，这些技术在其他领域也有广阔的使用。随着人工智的发展，目标检测在准确度、速度和适用性上都有了显著提升，其在安全帽与反光衣检测上的应用，体现了科技进步对于提高生产安全水平的贡献。

1.2 国内外研究现状

1.2.1 目标检测技术发展现状

目标检测技术是计算机视觉的一项主要任务，主要目标是确定图像或视频中特定物体的位置和类别。自 20 世纪 90 年代以来，目标检测技术经历了从传统方法到深度学习方法的重要转变。

传统的目标检测框架由三部分构成，分别为区域选择器、特征提取器和分类器如图 1-1 所示。2001 年，Viola 等人提出了一种革命性的人脸检测算法 Viola-Jones^[2]，能够在没有肤色等先验信息的情况下实现实时检测。这种检测器采用了“滑动窗口”技术，通过积分图像、特征选择和检测级联技术，显著提高了检测速度。2005 年提出的 HOG^[3] (Histogram of Oriented Gradients) 是一种在灰度图像中检测物体的方法，该方法通过构建图像中每个单元格的梯度直方图来形成特征。与 SVM^[4] 分类器相结合，HOG 特征在行人检测等众多领域中得到了广泛应用。但是它采用滑动窗口方法来提取特征，因此需要很长时间来识别对象。Felzenszwalb 等人在 2008 年提出了 DPM^[5] (Deformable Part-based Model) 这一物体检测框架。它是基于 HOG 的，但加入了可变形部件的概念，以更好地处理物体姿态的变化。D.G.Lowe 等人对 SIFT 算法^[6] 的提出对于计算机视觉领域来说是一个重要的里程碑，因为它提供了一种强大的方法来检测和描述图像中的局部特征，这些特征对图像的缩放和旋转具有不变性，因此在目标识别和图像匹配等任务中表现出色。但它对于图像纹理要求较多，如果遇到模糊图像或平滑处理图像则无法发挥作用。由 J.R.R. Uijlings 等人提出的 Selective Search 算法^[7]，使用图像分割来生成一系列的区域，然后通过计算这些区域之间的相似度来选择性地合并它们，以形成更大的区域。减少了需要评估的候选区域的数量，从而提高检测和分割的效率。传统目标检测虽然在特定场景下有一定的效果，但普遍存在着一些不足，基于滑动窗口的区域选择策略具有高计算复杂度和高窗口冗余度，在复杂的环境中，传统算法往往难以区分前景目标和背景噪声，导致许多误检。传统目标检测方法对目标尺度的变化适应性较差，通常需要通过图像金字塔技术来处理不同尺度下的目标，这不仅增加了计算复杂度。还有特征提取器不够鲁棒、分类器准确性不足以及处理速度慢等问题。

伴随着深度学习技术的迅猛进步，尤其是在卷积神经网络 (CNN) 的广泛运用下，目标检测技术步入了快速发展的轨道。在 2012 年的 ImageNet 大规模视觉识别挑战赛 (ILSVRC) 上，AlexNet^[8] 实现了里程碑式的突破。将图像识别的错误率从

26.2%降低到了 15.3%，大幅度超越了之前的传统机器学习方法，从此开启了深度学习在计算机视觉领域的新篇章。

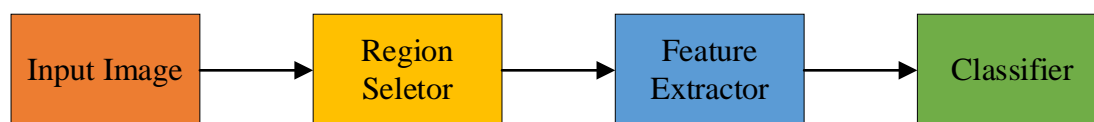


图 1-1 传统目标检测算法的基本架构

Fig. 1-1 The basic architecture of traditional target detection algorithm

目前，基于深度学习的目标检测算法主要被划分为两个主要类别：即两阶段检测算法和一阶段检测算法。目标检测发展如图 1-2 所示。

两阶段检测器首先使用区域选择网络 RPN^[9] (Region Proposal Network) 来生成候选的目标区域，然后对这些区域进行分类和边界框回归。典型的两阶段检测器包括 R-CNN (Regions with CNN features) 系列，如 Faster R-CNN 和 Mask R-CNN 等。这些方法通过 RPN 来提高候选区域的生成效率，并通过后续的网络对目标进行精确的分类和定位。

在 2014 年，R. Girshick 等研究者在目标检测领域提出了 R-CNN^[10] 算法，通过结合区域提议方法和 CNN 特征提取，显著提高了目标检测的准确率。R-CNN 使用选择性搜索来生成候选区域，然后利用 CNN 提取特征，最后使用 SVM 分类器进行分类。为了解决 R-CNN 中候选区域尺寸不一导致的特征图尺寸不匹配问题，SPPnet^[11] (Spatial Pyramid Pooling Net) 引入了空间金字塔池化层，可以接受不同尺寸的特征图并输出固定尺寸的特征向量。Fast R-CNN^[12] 通过 RoI (Region of Interest) Pooling 层改进了 SPPnet，它可以接受任意尺寸的候选区域并输出固定尺寸的特征向量，同时引入了 RoI 概念。Faster R-CNN^[13] 引入了区域建议网络，它可以共享卷积特征并实时生成候选区域，极大地提高了检测速度和效率。Faster R-CNN 标志着首个实质性的两阶段目标检测网络的出现。在 Faster R-CNN 的基础上，Mask R-CNN^[14] 添加了一个分支来预测目标的分割掩码，使其能够进行实例分割。

一阶段检测器直接在图像上预测目标类别和位置，不需要生成候选区域。典型的一阶段检测器包括 YOLO^[15] 和 SSD^[16] 等。这些方法通常具有更快的处理速度，但可能在精度上略低于两阶段检测器。YOLO 是一个革命性的网络，它将目标检测作为回归问题处理，直接在图像中预测目标的类别和位置。YOLO 的出现使得实时目标检测成为可能。SSD 使用了一个类似 YOLO 的网络结构，但在特征图上使用不同尺寸的锚框来预测目标位置，这种方法在精确度上有所提升。YOLOv2^[17] 在 YOLOv1

的基础上进行了多方面的改进，引入了新的预测机制，预测边界框的置信度和类别概率，改进了网络结构。YOLOv3^[18]引入了新的网络架构 Darknet-53，以及一个用于提高对小目标检测能力的特性 PANet（Path Aggregation Network）。YOLOv3 还引入了锚框的概念，以改善边界框的预测。YOLOv4^[19]进一步改进了网络结构，引入了 CSPDarknet-53+模型，该模型使用了一种新的卷积块 CSPNet（Cross Stage Partial networks）。YOLOv5 使用了 SiLU 激活函数，引入了 FocusNet 注意力机等，进一步提高了检测精度。

近年来，目标检测技术还引入了 Transformer 架构，这在很大程度上推动了目标检测技术的发展。基于 Transformer 的方法，如 DETR（Detection Transformer），通过将 Transformer 的结构应用于目标检测任务，实现了端到端的目标检测，在众多标准数据集上，该模型展现出了卓越的表现。

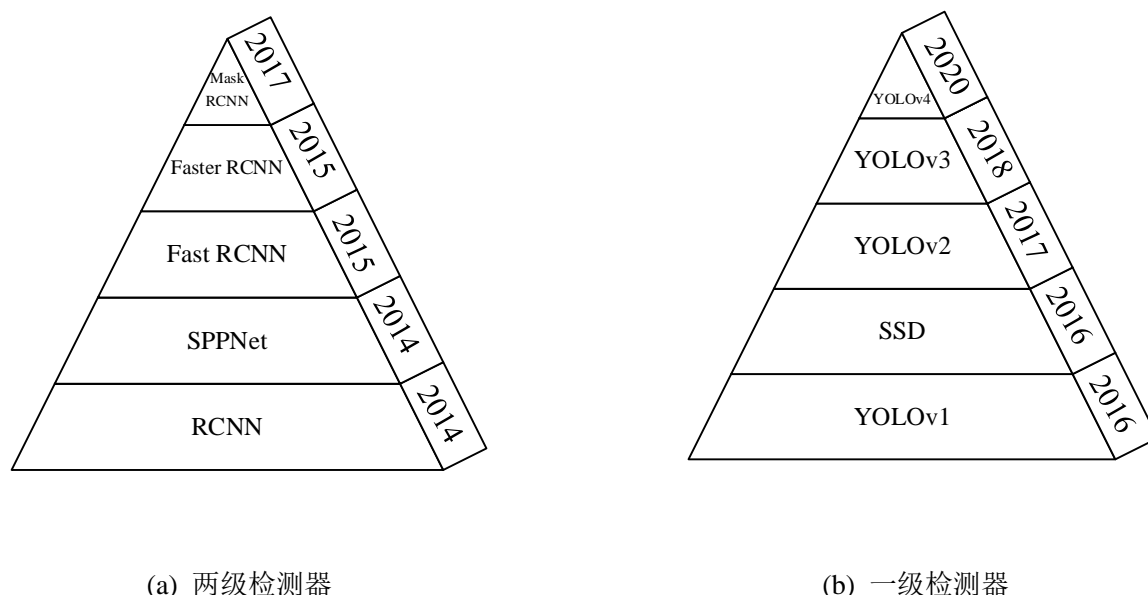


图 1-2 目标检测模型的分类

Fig. 1-2 Classification of target detection models

1.2.2 安全帽与反光衣检测研究

Park 等人^[20]通过 HOG 定向梯度直方图和 HSV 颜色直方图来检测安全帽。Sun X^[21]等人提出了一种结合多特征融合和支持向量机的安全帽检测方法，以提高安全帽识别率。Wu H 等人^[22]通过局部二进制模式（LBP），几何不变矩（HMI）和颜色直方图（CH）组成的混合描述符，以提取具有不同颜色（红色，黄色和蓝色）的头盔特征。冯国臣等人^[23]利用混合高斯模型进行前景检测，确定为人体后头部区域的安

全帽进行识别。这些传统的检测方实时性较弱小，且对于背景复杂的环境检测不够准确。

随着深度学习的发展与计算机算力的迅速提升，安全帽与反光衣的检测技术正处于快速发展阶段，研究者们致力于提高算法的实时性能与检测精度，以确保在实际应用中能够起到更好的效果。Cheng 等人^[24]通过构造基于深度可分离卷积和通道注意机制模块取代原始的卷积层，改进的空间金字塔池（SPP）模块等方法来改进 YOLOv3-tiny，进而检测安全帽，不仅实现了模型的轻量化还改善了检测指标。Zhang 等人^[25]使用改进的 YOLOv5s 算法检测六种安全帽佩戴方式，通过 K-means 方法重新设计了先验框的大小，并且在算法中增加了一个额外的层，以提高模型对小目标的识别能力，将 YOLOv5 算法中的 CIoU 损失函数替换为 EIoU 损失函数，提高了 YOLOv5s 算法对安全帽检测的准确性。Chen 等人^[26]使用轻量级网络 PP-LCNet 作为 YOLOv4 主干网络，使用坐标注意机制模块以增强特征信息，并且使用了 SIoU 损失函数，提高了检测精度，降低了模型大小，实现了实时高效的安全帽佩戴检测。Song 等人^[27]结合了多目标跟踪算法 DeepSort 和 YOLOv5 算法，提高了在复杂多变的施工环境中安全帽的检测精度。Hayat 等人^[28]提出了一种改进 YOLOv5x 算法，检测弱光图像中的安全帽。Xu 等人^[29]提出改进的 YOLOv5 算法，通过坐标空间注意力模块和多尺度非对称卷积增强模型关注相关特征的能力，提高了仓储场景中小目标安全头盔的检测。Zhang 等人^[30]提出了一种新的安全帽检测算法 FEFD-YOLOV5，其通过添加专门针对小目标检测的浅层检测头，并结合 SENet 通道注意力模块来压缩全局空间信息来增强检测性能，从而提高模型在相应场景下的平均精度。

Ma 等人^[31]通过 K-means++ 聚类方法，引入 SPPF 模块，使不同尺度的特征信息充分融合，获得更丰富的特征表示，提高了 YOLOV4-tiny 在自制反光衣数据集上的检测精度。谢国波等人^[32]通过改进的 YOLOX 算法对反光衣与安全帽进行检测，引入 CAM 模块，增强感受野，采用 Varifocal Loss 损失函数解决样本分类不均匀问题，最终提升了 YOLOX 算法检测速度与平均精度。骆杰豪^[33]基于深度学习算法，提出了一种行人安全穿戴级联检测算法。程换新等人^[34]提出了改进的 YOLOX-S 检测方法，引入 BiFPN 模块，提高网络特征提取的能力，使用了 GIoU 损失函数，并且使用 Mosaic 提高了网络模型的识别准确率，提高安全帽反光衣检测。韩泽佳等人^[35]改进了 SSD 算法，通过将 ResNet50 网络替换 VGG-16 作为特征提取网络，加入可形变卷积模块方法，提高了对安全帽和反光衣的检测精确度和速度。

1.3 主要研究内容

本文主要研究的是的安全帽与反光衣的佩戴检测，分析了现有的一些算法的不足，结合数据集进行了算法的改进。并进行了大量的实验与分析。通过查阅大量文献与学位论文，分析了安全帽与反光衣的佩戴检测研究，了解各种目标检测算法的优势，经过筛选，决定采用 YOLOv5s 作为检测算法，并对其进行了优化升级。

本文首先使用 Labelimg 对数据集进行了标注，并且分成了四类，helmet（安全帽），nohelmet（无安全帽），reflective_clothes（反光衣）和 other_clothes（其他衣服）。避免了单一场景，更加符合实际情况。并使用改进算法进行训练，在数据集上表现良好。证明了本文所提出算法的有效性。然后，对 YOLOv5s 检测算法进行了轻量化的改进，不仅保证了检测精度与检测速度，而且尽可能的降低了其参数量与计算量，使其更适合在移动设备与嵌入式设备中使用。最后，通过改进 YOLOv5s 检测算法，提高了网络的特征提取的能力。修改损失函数提高了模型的收敛速度和预测框的定位准度。添加注意力机制模块，更能关注到检测的区域。替换上采样算子，来聚合上下文信息。最终提高了算法的检测精度，使其更匹配对安全帽与反光衣的检测。

1.4 论文框架

本篇论文共有五章，具体内容如下：

第一章，绪论。本章首先阐述了安全帽与反光衣检测研究的背景与意义，分析了研究的重要性。接下来通过众多文献研究了国内外目标检测发展现状，并且分析了现有针对安全帽与反光衣的目标检算法，确定了基于改进 YOLOv5s 算法的研究方向。最后介绍了本文结构。

第二章，相关知识介绍。本章介绍了深度学习基本理论。介绍了卷积神经网络的结构。最后，详细的叙述了一阶段检测 YOLO 系列的代表算法，介绍了其网络结构，优点和不足。还有每一代 YOLO 相对于上一代的改进。

第三章，基于 Ghost 改进的 YOLOv5 检测算法。首先，介绍了 YOLOv5s 检测算法的结构，各个功能模块。介绍了基于 YOLOv5s 的轻量化改进方式，并且详细的阐述了 Ghost 模块，CA 注意力机制和 WIoU 损失函数的结构，功能和优点。然后，介绍了 YOLO 算法常用的几种评估指标，平均精度，精确率，召回率，参数量和计算量，说明其实际含义和计算方法。接着介绍了本篇文章所用的数据集的构成，标注

方法和数据集的数量。最后，最后介绍了实验环境的搭建和参数配置。介绍了实验的方法。通过不同的消融实验和对比实验，更进一步展示了所提出算法的优异性。

第四章，基于 SAConv 改进的 YOLOv5 检测算法。首先，介绍了 SAConv 卷积结构以及和 C3 模块结合的方式。然后介绍了 SIoU 损失函数的计算公式和 CARAFE 轻量级上采样算子的特点。最后，通过实验对比，验证所改进网络的优越性。

第五章，总结与期望。本章对研究成果进行了总结，并且对未来的研究方向进行了展望。

2 相关知识介绍

2.1 深度学习简述

深度学习通过模拟人脑的运行方式，自动地从大量数据中学习并提取特征，进而实现对数据的分类、识别、生成等任务。深度学习的发展，得益于计算能力的提升、大数据的积累以及算法的创新，其应用范围已涵盖计算机视觉、人脸识别、自然语言处理等多个领域。

深度学习起源于人工神经网络（ANN）的研究，并在长达数十年的持续发展中，超越了其他许多高性能的机器学习技术，目前被广泛认为是效率最高的方法之一。深度学习被称为深度神经网络，传统的神经网络仅有 2 到 3 层隐藏层，而深度网络由大量的神经元组成，这些神经元之间通过连接相互关联，形成一个层次结构的网络。简单神经网络结构如图 2-1 所示，神经网络中的每一个神经元都代表一个特征，该层接收来自上一层神经元的信号输入，经过特定的处理转换，再将结果输出至下一层的神经元。通过这种方式，神经网络可以在多层神经元之间传递和加工特征，从而学习到数据中的复杂模式和规律。

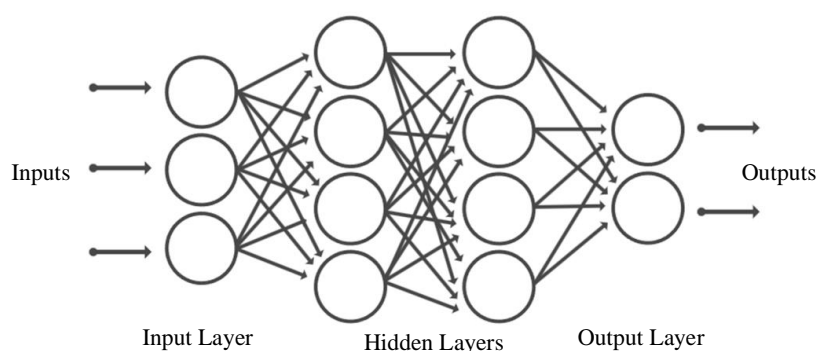


图 2-1 神经网络结构

Fig. 2-1 Structure of neural network

深度学习的一个关键特点是层次结构，它包括输入层、隐藏层和输出层。输入层负责接收原始数据，隐藏层负责提取和转换数据中的特征，输出层则负责生成最终的预测结果。在深度学习中，神经网络通常包含多层隐藏层，这使得网络能够学习到更加抽象和复杂的特征。

在计算机视觉领域，深度学习技术已经实现了突破性的成果。卷积神经网络

(CNN) 是一种专注于图像处理的深度学习模型，它在图像分割、目标检测等应用中表现出色。卷积神经网络通过引入卷积操作，能够有效地提取图像中的局部特征，从而实现对图像的高效识别。除了计算机视觉，深度学习在语音识别领域也取得了重要的进展。循环神经网络(RNN)是一种适用于处理序列数据的深度学习模型，它在语音识别任务中具有优势。循环神经网络能够捕捉语音信号中的时序信息，从而准确地识别出语音中的单词和句子。自然语言处理(NLP)是深度学习的另一个重要应用领域。在NLP中，深度学习模型能够处理和理解文本数据，实现诸如文本分类、情感分析、机器翻译等任务。通过训练神经网络模型，可以让计算机理解文本中的语义信息，进而实现对文本的高效处理和分析。

2.2 卷积神经网络

卷积神经网络是一种具有深层架构的前馈型网络，它是深度学习领域的一个标志性算法，可以进行监督学习和非监督学习，包含卷积计算，能够有效地处理图像数据。卷积神经网络模仿生物的视知觉机制构建，具有较强的鲁棒性和泛化能力，在计算机视觉领域取得显著成果。卷积神经网络的网络结构包括输入层、卷积层、池化层和全连接层。

(1) 输入层

输入层的主要任务是对原始图像数据进行预处理。去均值：减去图像的平均像素值，以消除图像亮度 and 对比度的偏差。归一化：将像素值缩放到一个较小的范围，如 0 到 1 之间，以加速计算并提高模型的稳定性。数据增强：通过旋转、翻转、缩放、裁剪等手段增加训练样本的多样性，以提高模型的泛化能力。输入层接收的数据通常是三维的，即具有宽度、高度和颜色通道的数量。对于灰度图像，数据维度是二维的。输入层对图像的每个像素执行操作，在灰度图像中，将图像转换为一维数组或二维数组，以便后续层能够处理。

(2) 卷积层

卷积层通过卷积核在输入数据上滑动，以提取局部特征。卷积层能够学习到数据的层次表征，将原始数据映射到更抽象的层面。这种表征有助于后续更好地理解数据的结构。卷积核的滑动，减少了数据的维度。卷积层中的每个卷积核是一个小的权重矩阵，其大小取决于卷积的类型(如大小为 3×3 或 5×5)。卷积核在输入数据上滑动，与每个局部区域相乘并求和，形成卷积层的输出。为了控制输出特征图的

大小，可以在输入数据周围填充 0（padding）或在滑动过程中设置步幅（stride），卷积计算过程如图 2-2 所示。

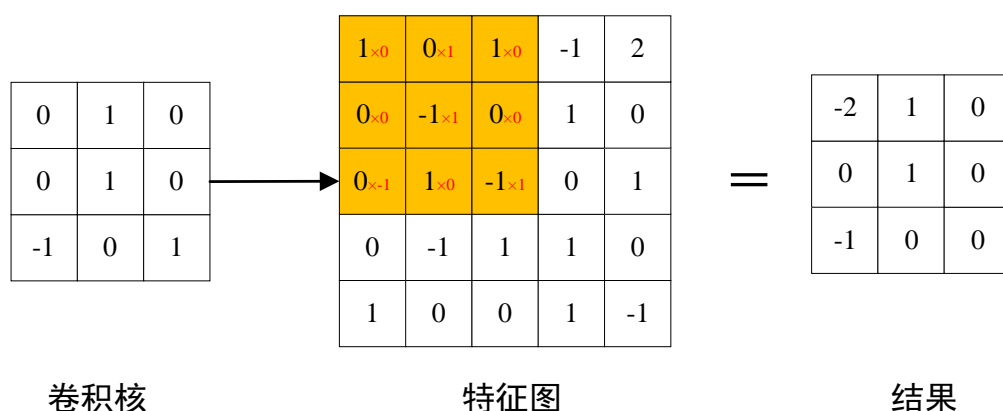


图 2-2 卷积计算示意图

Fig. 2-2 Convolution computation diagram

特征图长宽计算公式：

$$H_2 = \frac{H_1 - F_H + 2P}{S} + 1 \quad (2.1)$$

$$W_2 = \frac{W_1 - F_W + 2P}{S} + 1 \quad (2.2)$$

其中， H_1 ， W_1 表示输入特征图的长度和宽度； H_2 ， W_2 表示输出特征图的长度和宽度； F_H ， F_W 表示卷积核大小； S 表示滑动窗口的额步长； P 表示边界填充几圈 0。

（3）池化层

池化层通过减小特征图的尺寸来降低数据的分辨率。这可以减少模型所需的参数量和计算量，同时保持重要特征。池化层在卷积层获得的特征图上滑动，对每个局部区域执行池化操作，以提取该区域的特征。这有助于网络关注更具区分性的特征，忽略噪声和不重要的细节。常见的池化方式包括最大池化（Max Pooling）、均值池化（Average Pooling）和全局平均池化（Global Average Pooling）等。其中最大池化过程如图 2-3 所示。

（4）全连接层

全连接层将前面卷积层和池化层输出的特征图（feature map）转换为一维向量。这一过程可以看作是将空间维度的特征整合到数值维度上，使得模型能够处理一维数据，进行后续的分类或回归任务。在回归任务中，全连接层输出一个或多个数值，直接对应于预测的目标值。

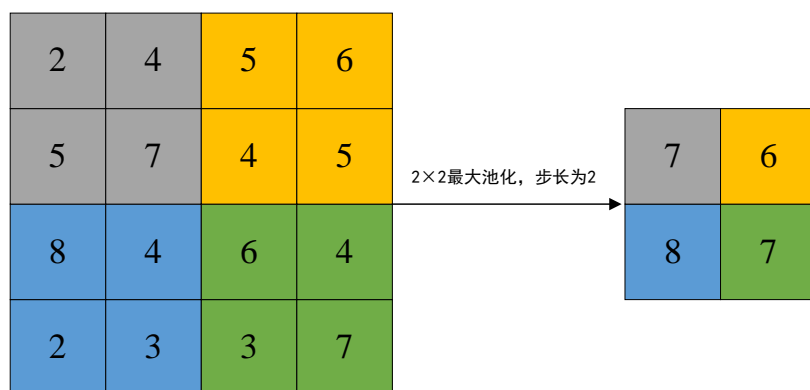


图 2-3 池化计算示意图

Fig. 2-3 Pooling calculation schematic diagram

2.3 YOLO 检测算法

2.3.1 YOLOv1

YOLOv1 网络设计是基于 GoogLeNet 模型，并进行了修改以适应目标检测任务，YOLOv1 网络结构如图 2-4 所示。YOLOv1 网络包含 24 个卷积层，这些层用于图像的提取特征。卷积层使用了不同的卷积核尺寸，如 3×3 和 1×1 ，以及步长为 1 或 2 的卷积操作。网络中有 4 个最大池化层，用于减小特征图的尺寸，并增加特征的感受野。在网络的最后部分，有两个全连接层，用于将卷积层提取的特征映射到预测的边界框和类别概率。受到全连接层的限制，输入图像为固定尺寸，一般为 448×448 像素。输出层中 YOLOv1 的输出是一个 $7 \times 7 \times 30$ 的张量，其中 7×7 对应于将图像分割成的网格单元，每个网格单元预测 2 个边界框，因此每个单元有 30 个输出值。这 30 个值包括 2 个边界框的 5 个坐标值（x, y, w, h, 置信度）和 20 个类别概率。

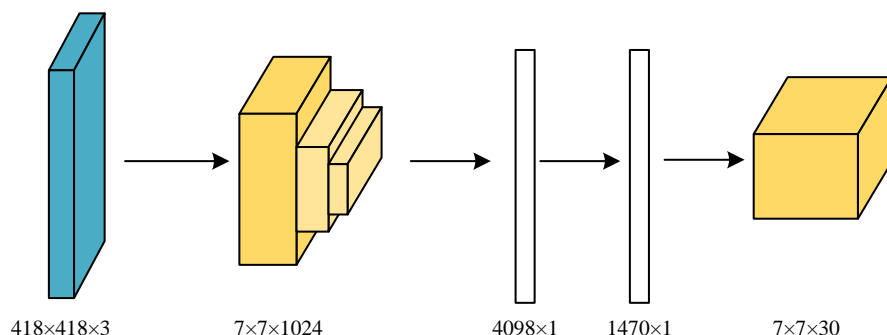


图 2-4 YOLOv1 网络结构

Fig. 2-4 YOLOv1 network structure

YOLOv1 中的位置损失，用于计算每个边界框的预测坐标与真实坐标之间的误差；置信度损失，计算与真实边界框的 IoU（交并比）的差的平方作为损失，置信度表示预测边界框中包含目标的可信度以及该边界框的准确度。置信度定义为预测边界框与真实边界框的 IoU 与目标存在的概率的乘积。分类损失，当网格单元中存在目标时，计算预测与真实类别之间的交叉熵损失。

由于多个网格单元可能会预测到同一个目标，因此需要进行非极大值抑制来去除冗余的边界框。具体操作是：对于每个类别，选择具有最高置信度的边界框，然后删除与其 IoU 大于设定阈值的其它边界框。NMS 处理后，剩下的边界框及其对应的类别和置信度即为检测结果。

YOLOv1 将目标检测问题转化的回归问题，这是一个开创性意义。通过单个卷积神经网络实现了端到端的目标检测，极大地提高了检测速度，使得实时目标检测成为可能，从而实现了显著的检测速度和相对较好的性能。而且简化了检测流程，将目标检测的定位、分类和检测整合到一个统一的框架中。在训练时考虑到了背景信息，减少了将背景错误分类为目标概率。而且具有一定的泛化能力，能够处理多种不同的目标检测任务。当然 YOLOv1 也存在着一些不足，在处理小目标时性能较差，而且在边界框的定位精度上可能不够高。

2.3.2 YOLOv2

YOLOv2 的模型架构是在其前身 YOLOv1 的基础上进行改进的，以提高物体检测的准确性和速度。YOLOv2 的模型拥有一个新的卷积神经网络架构 Darknet-19。Darknet-19 设计不仅提高了计算速度，而且也保持较高的识别准确性。

Darknet-19 网络包含 19 个卷积层和 5 个最大池化层，Darknet-19 网络结构如图 2-5 所示。它的卷积层主要用于特征提取，而最大池化层则用于减小特征图的尺寸。Darknet-19 还使用了批量归一化（Batch Normalization）来提高训练的稳定性和收敛速度。Darknet-19 的卷积层中包含 3x3 和 1x1 卷积核。Darknet-19 使用最大池化层用于减小特征图的尺寸，同时保留最重要的信息。在每个卷积层之后，YOLOv2 使用了批量归一化来规范化激活函数的输出，有助于提高网络的训练速度和稳定性。

为了提高对小物体的检测能力，YOLOv2 引入了 passthrough 层，它可以将高分辨率的特征图与低分辨率的特征图相结合，使得网络能够利用更细粒度的特征。而且从 Faster R-CNN 中借鉴概念，使用了锚框来预测不同形状的边界框。为了使网络能够检测不同大小的物体，YOLOv2 在不同尺寸上检测物体，在训练过程中会改变

输入图像的尺寸，这样网络就能学习到在不同分辨率下检测物体的能力。

YOLOv2 模型通过这些改进，使其在保持实时性能的同时，提高了物体检测的准确性。虽然 YOLOv2 在检测速度和检测性能之间取得了良好的平衡，但模型复杂度仍然较高，对于一些计算资源有限的设备来说，部署存在困难。YOLOv2 在处理遮挡物体与小物体时，检测可能不够准确。

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		2×2/2	112×112
Convolutional	64	3×3	112×112
Maxpool		2×2/2	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		2×2/2	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		2×2/2	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		2×2/2	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

图 2-5 Darknet-19 网络结构

Fig. 2-5 Darknet-19 network structure

2.3.3 YOLOv3

YOLOv3 是继 YOLOv2 之后的下一代实时物体检测算法。YOLOv3 在保持检测

速度的同时，进一步提高了物体检测的准确性和鲁棒性。YOLOv3 采用了新的基础网络架构 Darknet-53，包含了 53 个卷积层，网络比 YOLOv2 中的 Darknet-19 更深。Darknet-53 提高网络的感受野，同时减少计算量，使其在图像检测任务中更加有效。YOLOv3 网络结构如图 2-6 所示。

YOLOv3 在三个不同尺度的特征图上进行预测，这些特征图分别对应于不同的分辨率和感受野。这种多尺度预测机制使得 YOLOv3 能够有效地检测到不同大小的物体。YOLOv3 通过将低层特征图上采样并与高层特征图相加，实现了细粒度特征的融合。这种融合有助于网络在检测小物体时能够利用到更丰富的细节信息。

YOLOv3 使用了与 Faster R-CNN 类似的锚框机制，但 YOLOv3 为每个尺度使用三个锚框进行预测，总共使用了 9 个锚框。这些锚框的尺寸和比例是通过聚类算法在训练数据集上自动学习得到的。在 YOLOv2 中，每个锚框使用了 softmax 层来预测物体的类别，而在 YOLOv3 使用逻辑回归来预测每个锚框的类别。YOLOv3 不是使用 softmax，而是使用独立的逻辑回归分类器来预测每个类别的概率，这使得每个类别预测是独立的，可以处理多标签分类问题。YOLOv3 为每个锚框预测了 4 个边界框坐标、1 个对象得分和多个类别概率。对象得分表示网络对当前锚框是否包含一个物体的置信度。

YOLOv3 在训练时使用了多种数据增强技术，如随机裁剪、颜色变换和水平翻转，以提高模型的泛化能力。此外，还使用了多尺度训练，以使模型能够适应不同大小的输入图像。

YOLOv3 在保持实时性能的同时，显著提高了物体检测的准确性和鲁棒性，同时在速度上具有明显优势。这使得 YOLOv3 在需要实时物体检测的应用场景中非常受欢迎。

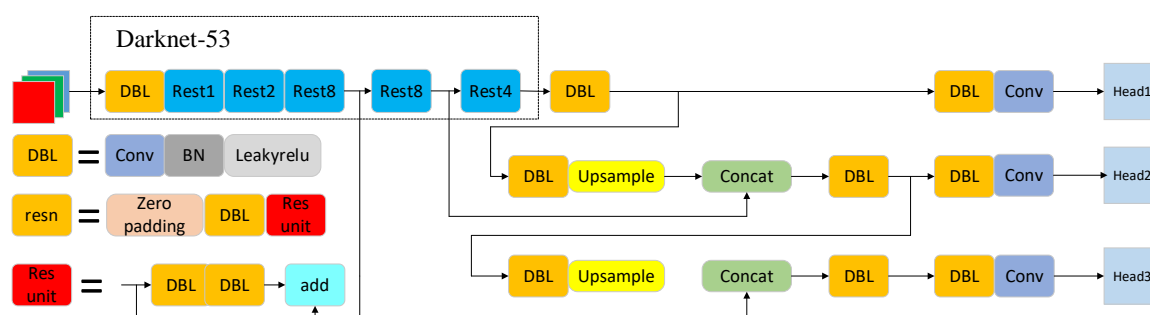


图 2-6 YOLOv3s 网络结构

Fig. 2-6 YOLOv3s network structure

2.3.4 YOLOv4

YOLOv4 作为第 4 代改进版本，它在 YOLOv3 的基础上引入了多种最新的计算机视觉技术，以提高物体检测的准确性和鲁棒性^[36]。YOLOv4 使用了 CSPDarknet53（Cross Stage Partial networks, CSP）作为其主干网络。CSPDarknet53 这是一种基于 Darknet 的网络架构，通过引入跨阶段层次结构来减少计算量并提高效率，具有更好的性能和更少的参数^[37]。

YOLOv4 的颈部网络采用了 SPP（Spatial Pyramid Pooling）模块和 PANet 结构作为颈部网络。SPP 模块可以增加网络的感受野，而 PANet 则有助于更好地融合不同层次的特征图。

YOLOv4 的检测头与 YOLOv3 类似，使用多个尺度进行预测，并采用了锚框来预测物体的位置和大小。每个锚框都会预测一个边界框、一个置信度分数和多个类别概率。

YOLOv4 使用了多种数据增强技术，这些技术可以增加训练数据的多样性，提高模型的泛化能力。使用了非单调 Mish 激活函数，以提供更好的梯度流和性能。为了防止过拟合。还采用了 DropBlock 正则化技术。YOLOv4 使用了 CIoU 损失函数来优化边界框的预测，以更好地处理边界框的回归问题。YOLOv4 在训练过程中使用了多尺度训练，以使模型能够适应不同大小的输入图像。

YOLOv4 通过结合上述多种技术和策略，显著提高了物体检测的准确性和鲁棒性。在 COCO 数据集上的实验结果表明，YOLOv4 在保持实时性能的同时，取得了与当时最先进的物体检测算法相媲美的性能。

2.4 本章小结

本章主要介绍基于深度学习的目标检测相关知识。首先，介绍了深度学习的基本理论。然后介绍了卷积神经网络的结构，卷积神经网络是适合于图像处理的深度学习模型。最后，详细介绍目标检测的一阶段检测算法，对 YOLO 系列第一代到第四代的模型结构进行了说明，同时也介绍了 YOLO 存在的不足和迭代升级后的创新点。

3 基于 Ghost 改进的 YOLOv5 检测算法

3.1 引言

YOLO 检测算法因为其检测速度快而准，所以被广泛应用于需要实时处理的场景。然而，原始的 YOLO 模型可能对于资源受限的环境（如移动设备或嵌入式系统）来说仍然太大或太复杂。因此，本节基于 YOLOv5s 检测算法进行了改进，设计了一种轻量化的网络模型，并且提高了一定检测精度。在自制数据集上表现良好。

3.2 YOLOv5s 算法

YOLOv5 是 YOLO 系列第 5 代目标检测模型，由 ultralytics 团队开发，其中根据网络深度和宽度的不同，可以分为 s、m、l 和 x 版本。YOLOv5s 模型^[38]的网络深度和宽度都比较小。它以其轻量级的设计、出色的性能和易用性而受到广泛应用。是本文所改进模型的基础。

YOLOv5s 网络主要由 3 部分构成，Backbone（主干网）、Neck（颈部）和 Head（检测头）构成。YOLOv5s 网络结构如图 3-1 所示。主干网络主要负责从输入图像中提取多尺度特征。由 CBS（Conv+BN+SiLU）、C3 模块和 SPPF（Spatial Pyramid Pooling - Fast）快速空间金字塔池化构成^[39]。SiLU 激活函数是 ReLU 激活函数的一种变种，它将 Sigmoid 函数和 ReLU 函数结合起来。SiLU 在大于等于 0 时，提供了正的梯度，这有助于缓解梯度消失问题，尤其是在深层网络中；在小于等于 0 时输出均值为 0，这有助于保持模型的稳定性，并减少了对输入数据分布的偏移。其中 C3 模块作为主要结构，这是 YOLOv5s 的核心特征提取部分，C3 模块中的 BottleNeck 结构减少计算复杂度，同时保持或提高模型性能^[40]。SPPF 通过使用多个最大池化操作，在不同的尺度上对特征图进行池化，然后拼接特征图，从而增加网络对不同尺度特征的感知能力。这使得网络能够更有效地捕捉到图像中的目标信息和上下文关系^[41]。

颈部网络负责将主干网络提取的特征进行级联和融合，以形成更丰富的特征表示。YOLOv5s 的颈部网络部分主要包括 PANet 结构，它能够有效地将不同层次的特征图融合，以提供更好的多尺度信息。PANet 结构如图 3-2 所示。FPN 是一个自顶向下的特征融合框架，它通过上采样操作将高层的特征图映射到与原始图像相同的尺

寸，同时通过横向连将高层特征与低层特征进行融合，确保了语义信息与位置信息的有效结合。在 FPN 的基础上，PANet 增加了一个自底向上的路径，这个路径通过一系列卷积操作增强低层特征，并将其高效地传递到高层，使得低层的位置信息能够更好地辅助高层的语义信息。

检测头包含预测层，负责输出每个目标的类别概率、位置坐标和置信度。输出层使用了 YOLOv5s 独特的 Detect 层，该层会处理特征图并输出检测结果。

YOLOv5s 在训练和测试时，采用了 Mosaic 数据增强策略，以此丰富训练数据并提升模型的泛化性能。此外，YOLOv5s 还引入了自适应锚框的概念，它允许模型自动学习和调整锚框的大小和比例，以更好地适应不同尺寸和比例的目标。

在输入输出方面，YOLOv5s 接受尺寸为 640x640 的图像，并输出一个张量，其中包含检测框的位置、大小、类别概率和目标置信度。这些输出可用于后续的目标识别和跟踪任务。

YOLOv5s 模型的大小和计算复杂度相对较小，这使得它非常适合在资源受限的环境中部署，如嵌入式设备或无人机。其高性能和准确性也使其成安全帽与反光衣检测任务的首选模型。

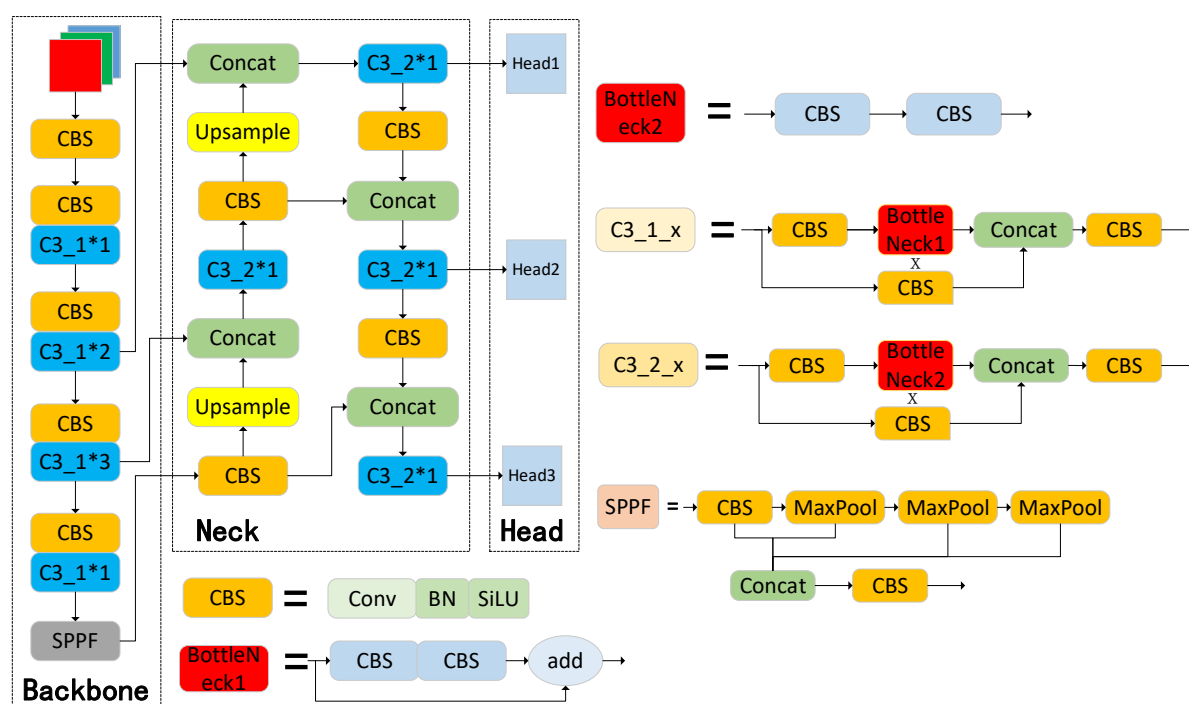


图 3-1 YOLOv5s 网络结构

Fig. 3-1 YOLOv5s network structure

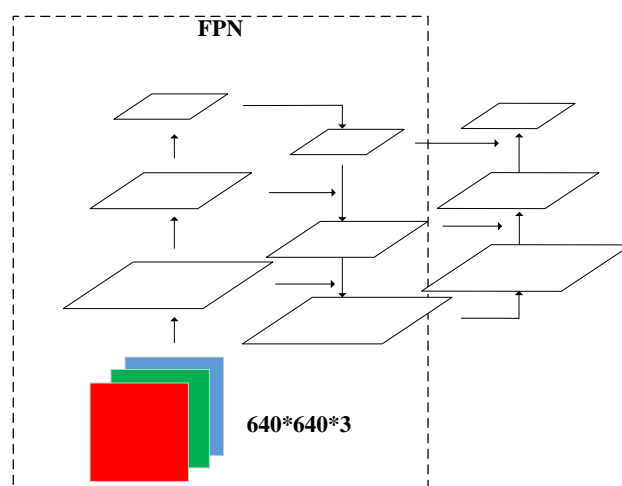


图 3-2 PANet 网络结构

Fig. 3-2 PANet network structure

3.3 改进的 YOLOv5s 算法

3.3.1 Ghost 模块的引入

在实际的应用环境中，YOLOv5s 网络模型性能容易受到硬件内存与计算量的影响。为了适应移动与嵌入式设备，引入了 GhostNet^[42]网络结构，以降低模型复杂度。GhostNet 是一种轻量化的网络，与传统的卷积神经网络相比，GhostNet 能够通过较少的计算成本生成更多的特征图。GhostNet 网络主要由 Ghost 模块构成^[43]。

Ghost 模块将传统的卷积与简单的线性运算相结合，利用了特征提取与特征图冗余的关系^[44]。Ghost 模块结构如图 3-3。在 YOLOv5s 网络中，Ghost 模块代替了传统了卷积过程，先对特征图进行普通卷积操作，进行通道压缩，然后进行线性变换获得更多的特征图，然后将得到的特征图进行拼接操作，得到最终的特征图，其中 Φ 表示线性变换^[39]。普通卷积浮点计算量 FLOPs 的公式如下：

$$F_1 = n \times h' \times w' \times c \times g \times g \quad (3.1)$$

Ghost 模块的浮点计算量 FLOPs 公式如下：

$$F_2 = \frac{n}{s} \times h' \times w' \times c \times g \times g + (s-1) \times h' \times w' \times \frac{n}{s} \times d \times d \quad (3.2)$$

其中， n 为输出通道的数量； h' 为输出特征的高度； w' 为输出特征的宽度； c 为输入通道的数量； g 为卷积核大小； s 为 Ghost 模块中生成的特征图数量； d 为线性操作卷

积核的大小。

$$\frac{F_1}{F_2} = \frac{c \times g \times g}{(1/s) \times c \times g \times g + d \times d \times (s-1)/s} \quad (3.3)$$

$$\approx \frac{s \times c}{s + c - 1} \approx s$$

其中 $s \ll c$ ，由计算可得知，普通卷积的计算量大约为 Ghost 模块计算量的 s 倍。GhostBottleNeck 中初始的 Ghost 模块旨在扩充通道数量，而随后的 Ghost 模块则负责将通道数量降低，确保与初始输入的通道数匹配。GhostBottleNeck 模块如图 3-4 所示。将 C3 模块与 GhostBottleNeck 结合，得到 C3Ghost 模块，降低了模型的参数量与计算量。C3Ghost 模块如图 3-5 所示。

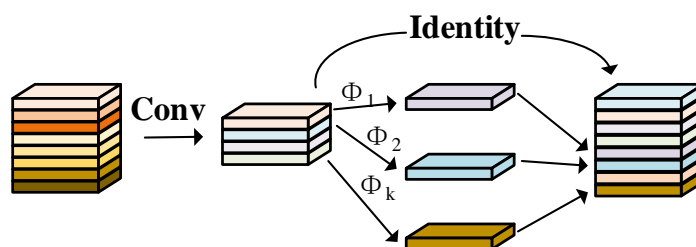


图 3-3 Ghost 模块

Fig. 3-3 Ghost module

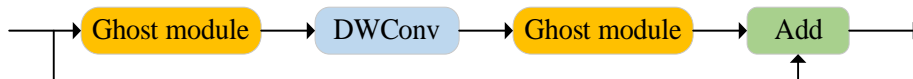


图 3-4 GhostBottleNeck 模块

Fig.3-4. GhostBottleNeck Module

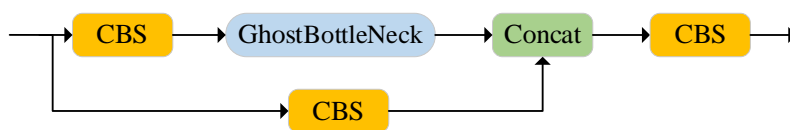


图 3-5 C3Ghost 模块

Fig. 3-5 C3Ghost Module

3.3.2 CA 注意力机制的引入

注意力机制在复杂背景下可以帮助模型忽略与目标物体无关的噪声信息，减少干扰，提高检测的准确性。可以帮助模型聚焦于图像中与目标物体最相关的区域，

从而增强这些区域特征的表达能力，更好地识别和定位目标物体。在工业生产和交通作业，周围环境复杂多变，为了使网络更能关注到安全帽与反光衣，忽略背景信息，引入了注意力机制。

CA (Coordinate Attention) 注意力机制是一种轻量化的注意力机制模块，目的是提高卷积神经网络的性能和效率。这种机制不仅考虑了通道间的关系，而且注意了位置信息，这有助于模型更好地定位和识别目标。

CA 注意力机制的显著优点在于它能够整合通道和空间位置信息^[45]。SE 注意力忽略了位置信息的关键性，这一信息这对于图像任务至关重要。CA 既能提取通道间的信息，还考虑了横向和纵向的信息，有助于模型关注到目标^[46]。CA 注意力机制因其灵活性和轻量级特性，能够轻松地被整合到卷积神经网络的核心位置中，而且能带来性能上的提升^[47]。

CA 注意力机制的实现包括两个主要步骤：在通道特征提取中，通过全局平均池化操作，使用两个 1 维向量的编码操作来捕捉单方向上的长距离关系，同时保留另一个方向上的空间信息。在通道注意力计算中，利用提取的位置信息，通过卷积和激活函数处理，生成最终的注意力权重。CA 注意力机制网络结构如图 3-6 所示

具体过程为，输入特征图 X ，使用 $(H,1)$ 和 $(1,W)$ 两个池化核，沿着宽度和高度两个方向进行平均池化。表达式如下：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (3.4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w) \quad (3.5)$$

其中， c 表示通道， h 表示高度， w 表示宽度。接下来将两个特征图拼接并且进行卷积变换，表达式如下：

$$f = \partial(F_1([z^h, z^w])) \quad (3.6)$$

$$g^h = \sigma(F_h(f^h)) \quad (3.7)$$

$$g^w = \sigma(F_w(f^w)) \quad (3.8)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (3.9)$$

其中， $[, \cdot]$ 表示沿空间维度的拼接， F_1 表示 1×1 卷积函数， ∂ 表示非线性激活函数， f 表示中间特征图， σ 为 sigmoid 函数， g 表示注意力权重。将 f 沿着为垂直和水平方向的分解为 f^h 和 f^w ，然后经过卷积变换，最终得到输出 y 。

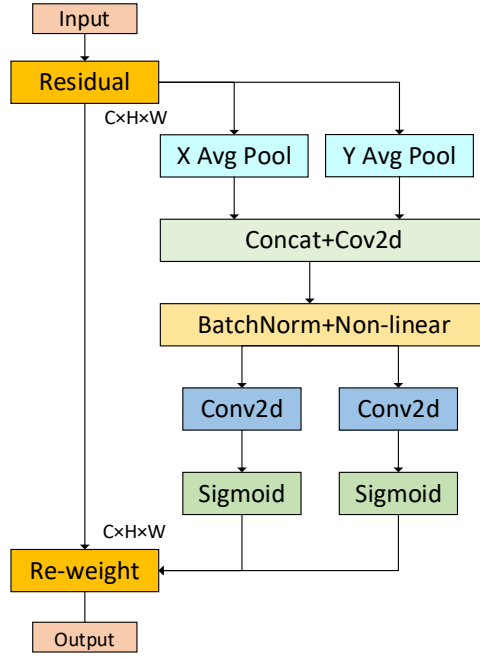


图 3-6 CA 注意力机制网络结构

Fig. 3-6 CA Attention mechanism network structure

3.3.3 WIOU 损失函数的替换

原始的 YOLOv5s 模型采用了 CIoU 损失函数作为其边界框的损失函数^[48]。CIoU 损失函数在 IoU 的基础上考虑了预测框与真实框之间的中心点距离以及宽高比，从而能够更全面地评估两个边界框之间的差异。CIoU 损失函数的计算公式如下：

$$CIoU = IoU - \frac{p^2(b, b^{gt})}{c^2} - \alpha v \quad (3.10)$$

其中， IoU 是预测框与真实框的交并比； b 和 b^{gt} 是预测框和真实框的中心点坐标； p 是计算两点之间距离函数； c 为最小包围框对角线长度。

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (3.11)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \quad (3.12)$$

其中， w^{gt} 和 h^{gt} 分别表示真实框的宽和高； w 和 h 分别表示预测框的宽和高。CIoU 损失函数考虑了预测框与真实框的重叠，位置关系和形状差异，这使得在训练目标检测模型时能够更有效地计算边界框损失函数。但是由于数据集中存在这一些低质量图片，使用 CIoU 损失函数会加剧对这些图片的惩罚力度，导致算法的泛化能力降低。因此本章使用 WIOU 损失函数来替换 CIoU 损失函数^[49]。WIoUv1 计算公式如下：

$$Loss_{WIoU} = R_{WIoU} Loss_{IoU} \quad (3.13)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)}\right) \quad (3.14)$$

其中, $R_{WIoU} \in [1, e]$, 会放大正常质量锚框的 $Loss_{IoU}$; x 和 y 为预测框的中心点坐标; x_{gt} 和 y_{gt} 为真实框中心点坐标; H_g 和 W_g 为最小包围框的高度和宽度如图 3-7 所示。

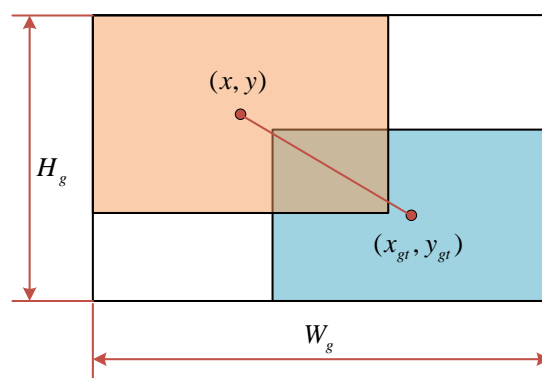


图 3-7 WIoU 示意图

Fig. 3-7 Schematic of WIoU

3.3.4 轻量化 YOLOv5s 网络模型

改进后的轻量化 YOLOv5s 网络模型如图 3-8 所示。其中, 使用 GhostConv 替换了网络中原来所有的卷积操作, 使用 C3Ghost 模块替换了原本的 C3 模块, 极大的降低了模型复杂度, 减小了计算量。但是轻量化的网络结构也带来了检测精度的损失。在整个网络中加入了 6 个 CA 注意力机制模块, 在不增加过多参数量的情况下, 增强了对感兴趣特征图区域的定位与识别^[39]。最后, 将 CIoU 损失函数替换为 WIoU 损失函数, 提高了模型的泛化能力。

3.4 实验结果与分析

3.4.1 实验环境

实验所使用服务器的操作系统为 Windows10, GPU 为 NVIDIA Quadro RTX3090, CPU 为 Intel(R) Xeon(R) Silver 4310 CPU @ 2.10GHz。搭建的环境为 python3.11.5, pytorch2.1.0, CUDA12.3。其中训练图片大小为 640×640 , 每次训练的批量大小设置为 32, 训练迭代设置 300 个 epoch, 学习率设置为 0.01, 其他的参数设置如表 3-1 所示。

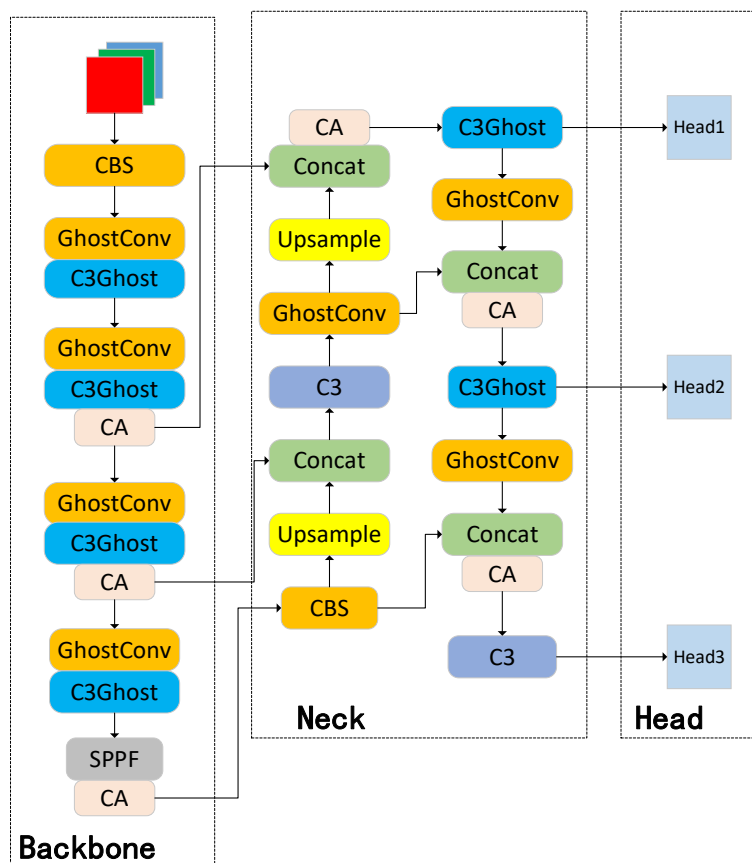


图 3-8 改进的 YOLOv5s 网络结构

Fig. 3-8 The improved YOLOv5s network structure

表 3-1 YOLOv5s 参数设置

Tab. 3-1 YOLOv5s parameter setting

参数	数值
图片大小	640×640
迭代次数	300
批次大小	32
学习率	0.01
动量	0.937
权重衰减	0.0005

3.4.2 数据集

本文使用自制数据集，用 LabelImg 工具对其进行了数据标注与分类，并且剔除了其中不相关的图片。总共分为 4 种类别，分别为 helmet（安全帽），nohelmet（无

安全帽), reflective_clothes (反光衣) 和 other_clothes (其他衣服)。数据集图片如图 3-9 所示, 其中包含了四种检测类别与数据增强后的图片。安全帽与反光衣初始图片 1083 张, 为了增加样本数量, 对图像进行了随机旋转, 镜像和加噪等操作将图片扩充为 5415 张。将图片按照 8: 2 的比例随机划分为训练集与测试集, 其中训练集 4332 张图片, 测试集 1083 张图片。

由于单张图片中存在着多个标签, 为了解各类标签数量, 制作了标签统计图如图 3-10 所示。其中反光衣标签数量最少, 数量约 4 千; 其他衣服标签最多, 数量超过 8 千, 总计超过 24500 个标签。



图 3-9 数据集图片

Fig. 3-9 Dataset picture

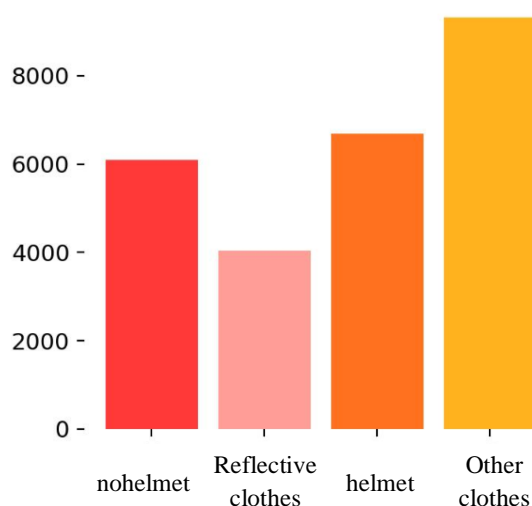


图 3-10 标签数量统计图

Fig. 3-10 Label quantity statistics

为了掌握标注框的分布情况, 制作了标注框分布图, 如图 3-11 所示。所有标注

框中心点相对于整幅图的位置如图 3-11 (a) 所示, 可见绝大部分目标都集中在图片的中间位置。所有标注框相对于整幅图的高宽比例如图 3-11 (b) 所示, 样本集中分布于左下角, 说明数据集中小目标居多。

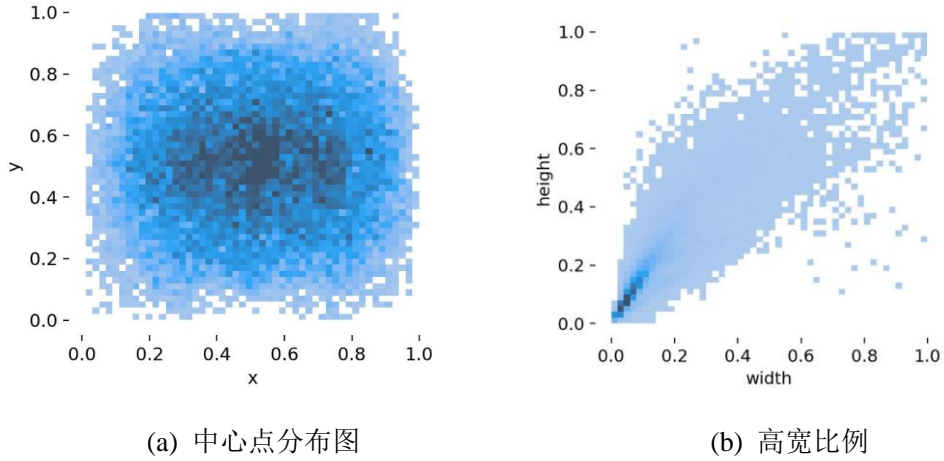


图 3-11 标注框分布图

Fig. 3-11 Label box distribution diagram

3.4.3 评估指标

mAP (mean Average Precision) 是衡量计算机视觉任务与目标检测任务中的一种常见指标。它是衡量目标类别检测平均精度的重要标准。其中 mAP 的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (3.15)$$

$$R = \frac{TP}{TP + FN} \quad (3.16)$$

$$AP = \int_0^1 P(R) dR \quad (3.17)$$

$$mAP = \frac{\sum_{i=0}^n AP_i}{N} \quad (3.18)$$

其中, P (Precision) 表示精确率, R (Recall) 表示召回率, i 表示类别数。TP 表示正样本被预测为正类, FP 表示负样本被预测为正类, FN 表示正样本被预测为负类。mAP@0.5 表示 IOU 阈值为 0.5 的 mAP 值, mAP@0.5:0.95 表示 IOU 阈值为 0.5~0.95, 且步长为 0.05 的平均 mAP 值。mAP 值越高, 说明模型在检测不同类别对象时的性能越好。这个指标在评估和比较目标检测算法的性能时非常重要。

在目标检测任务中, 参数量 (Params) 和计算量 (FLOPs) 是评估模型复杂度的重要指标。其中, 参数量用来衡量算法的复杂程度。在 YOLO 模型中, 参数量取决

于模型的架构。每个卷积层的参数量是由卷积核的数量、大小以及输入和输出的特征图的通道数决定的。卷积的参数量计算公式如下：

$$Params = k^2 \times C_{in} \times C_{out} \quad (3.19)$$

计算量通常用来衡量模型在进行前向传播时所需的浮点运算次数。FLOPs 是“billion floating-point operations”的缩写，即每秒十亿次浮点运算。计算量取决于模型特征图的大小，输入和输出通道数的多少。

在 YOLO 模型中，计算量主要来自于卷积层，因为卷积操作是计算密集型的。每个卷积层的计算量可以通过以下公式近似计算：

$$FLOPs = n \times C_{out} \times H_{out} \times W_{out} \times C_{in} \times k^2 \quad (3.20)$$

其中， C_{in} 和 C_{out} 分别表示输入与输出通道数， H_{out} 和 W_{out} 分别表示输出特征图的高和宽， k 表示卷积核大小，将所有层的计算量相加，得到整个模型的计算量。 $1GFLOPs=10^9FLOPs$ 。

对于具体的 YOLO 的参数量和计算量取决于它们的架构设计，如卷积层的数量、深度、宽度以及输入图像的分辨率。而且，YOLOv5 有不同大小的版本，它们的参数量和计算量随着模型大小的增加而增加。

为了获取特定 YOLO 模型的参数量和计算量，可以使用 PyTorch 的 `torchinfo` 或 `thop`，这些工具可以分析模型的结构并计算其参数量和计算量。需要注意的是，参数量和计算量并不总是直接关联的。一些模型可能有大量的参数，但计算量并不高，这可能是因为模型中有大量的稀疏连接或参数共享。同样，一些参数量少的模型，其计算量可能很高。

3.4.4 实验结果

改进的 YOLOv5s 不同类别 AP 如表 3-2 所示，所有类别的平均精确率为 96.1%，召回率为 90.2%，mAP@0.5 值为 94.8%。可以看出无安全帽的 mAP@0.5 值最小，达到 90.8%，反光衣的 mAP@0.5 值最高，达到 97.5%。反光衣的检测效果最好。改进 YOLOv5 的 mAP@0.5 曲线随着迭代次数增长曲线如图 3-12 所示。50 个 Epoch 时 mAP@0.5 值到达 80%，110 个 Epoch 时 mAP@0.5 值到达 90%，之后开始收敛，最终的 mAP@0.5 值为 94.8%。

P-R 曲线代表精确率与召回率之间的关系。横坐标为召回率，纵坐标为精确率。P-R 曲线与坐标轴围成的面积为此种类型的 AP，所以曲线越靠近右上方，AP 越大，算法性能越好^[39]。如图 3-13 所示四条细线分别代表安全帽、无安全帽、反光衣和其

他类衣服，蓝色实线为三种类别的平均精度 $mAP@0.5$ 。可以看出，所有曲线都非常靠近右上角，围成面积占到 90% 以上。

表 3-2 改进的 YOLOv5s 不同类别 AP

Tab. 3-2 Improved YOLOv5s APs of different classes

类别	精确率/%	召回率/%	$mAP@0.5$
All	96.1	90.2	94.8
Nohelmet	96.4	88.5	90.8
Reflective_clothes	97.2	93.5	97.5
Helmet	97.2	91.6	95.7
Other_clothes	95.8	90.6	95.1

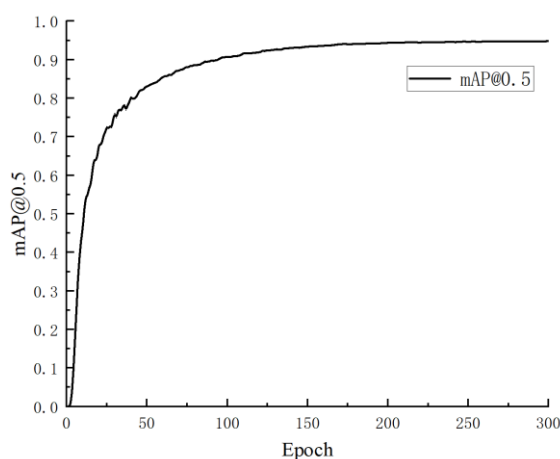


图 3-12 改进 YOLOv5 mAP 曲线

Fig. 3-12 Improved YOLOv5s mAP Curve

3.4.5 消融实验

为了验证本文所提出模型算法有效性，在自制数据集上进行了消融实验如表 3-3 所示。Ghost 模块替换了原有的普通卷积与 C3 模块，极大降低了模型的参数量，计算量和模型大小。为模型轻量化的主要功能模块。但是模型的轻量化导致了模型检测精度的精度，使模型的检测精度降低到了 93.6%。CA 注意力模块，在不增加模型复杂度的情况下极大的提高了模型的平均精度，达到 95%。与原始模型相比较， $mAP@0.5$ 提高了 0.9%。与替换了 Ghost 的 YOLOv5s 相比较，模型的 $mAP@0.5$ 值提高了 1.4%。WIoU 损失函数的引入也给模型带来了检测精度的提高，与原始模型相比较， $mAP@0.5$ 提高了 0.3%。将各个模块都加入到原始的 YOLOv5s 模型中，不

仅模型的 mAP@0.5 提高了 0.7%，而且模型的参数量降低 2.5×10^6 ，达到 4.5×10^6 。模型的计算量和模型大小分别降低了 6.6 GFLOPs 和 4.6 MB，分别达到了 9.2 GFLOPs 和 9.1MB。

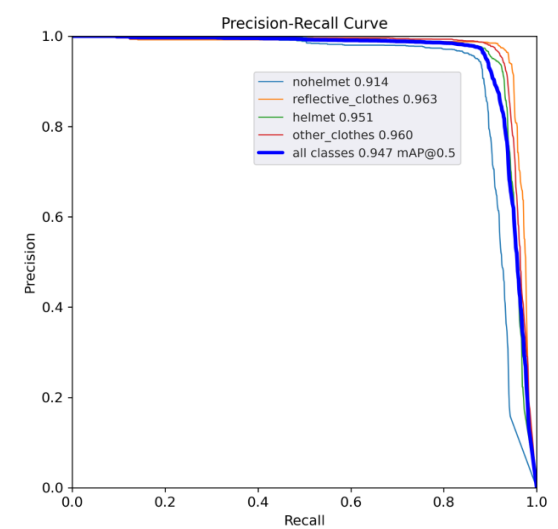


图 3-13 改进 YOLOv5 mAP 曲线

Fig.3-13 Improved YOLOv5s mAP Curve

表 3-3 消融实验

Tab. 3-3 Ablation experiment

模型	mAP@5/%	参数量 /10 ⁶	计算量 /GFLOPs	模型大小/MB
YOLOv5s	94.1%	7.02	15.8	13.7MB
+Ghost	93.6%	4.5	9.1	9.0MB
+CA	95%	7.09	15.9	13.8MB
+WIOU	94.4%	7.02	15.8	13.7MB
+Ghost+CA	94%	4.5	9.2	9.1MB
+CA+ WIOU	95.4%	7.0	15.9	13.8MB
+Ghost +WIOU	93.9%	4.5	9.1	9.0MB
+Ghost+ CA+ WIOU	94.8%	4.5	9.2	9.1MB

3.4.6 对比实验

为了直观地展示检测效果，制作了图 3-14 的检测结果对比。上图为原始的 YOLOv5s 检测结果，下图为改进的 YOLOv5s 检测结果。从图中可以清晰地看出，

四种类别安全帽, 无安全帽, 反光衣和其他衣服, 改进后 YOLOv5s 识别的置信度水平都要高于原始的 YOLOv5s 检测算法。而且, 在第二幅图中, 原始的 YOLOv5s 存在误检, 将背景识别为了其他衣服, 而改进后 YOLOv5s 则不存在这种问题。



图 3-14 检测结果对比

Fig. 3-14 Comparison of test results

为了展示改进模型与其他同类型模型的优势, 进行了以下表 3-4 的对比实验。可以看出 YOLOv7-Tiny 与原始的 YOLOv5s 算法相比较, 模型的复杂程度相差不大, 但是 YOLOv7-Tiny 的 $mAP@0.5$ 为 92.7%, 比原始 YOLOv5 低 1.4%。改进的 YOLOv5s 的模型复杂度要远低于 YOLOv7-Tiny, 而且 $mAP@0.5$ 高了 YOLOv7-Tiny 算法 2.1%。YOLOv7 的检测精度要比改进的 YOLOv5s 高, 但其模型复杂度过高, 参数量达到 36.5×10^6 , 计算量达到 103.2 GFLOPs, 模型大小达到 71.3 MB。YOLOv3-Tiny 为轻量型的算法, 其模型的复杂度比原始 YOLOv5s 略高, 且检测精度比原始 YOLOv5s 低 7.7%。YOLOv3 的 $mAP@0.5$ 低于改进的 YOLOv5s 模型 0.2%, 且模型参数, 计算量远远高于改进后的 YOLOv5s。

表 3-4 对比实验

Tab. 3-4 Contrast experiment

模型	$mAP@0.5/\%$	参数量/ 10^6	计算量 /GFLOPs	模型大小/MB
YOLOv5s	94.1	7.02	15.8	13.7
YOLOv3-Tiny	86.4	8.67	13.0	16.6
YOLOv7-Tiny	92.7	6.01	13.0	11.7
YOLOv3	94.6	61.5	154.6	117
YOLOv7	95.5	36.5	103.2	71.3
改进的 YOLOv5s	94.8	4.5	9.2	9.1

3.5 本章小结

本章首先介绍了 YOLOv5s 算法网络模型的构成与功能。然后对 YOLOv5s 检测算法进行了轻量化的改进，介绍了各个改进模块的原理和主要功能，通过加入 Ghost 模块来降低模型的复杂度，注意力机制模块的添加提高了网络对感兴趣区域的关注，更优秀的损失函数计算方法提高了检测精度。最后，介绍了实验的环境和超参数，实验环境和评估指标。并且在数据集上进行了实验，结果表明改进后的 YOLOv5s 参数量降低了 35.7%，计算量减低了 41.7%，模型大小降低了 33.6%，并且模型的精度提高了 0.7%，达到 94.8%。实验效果良好。

4 基于 SAConv 改进的 YOLOv5s 的检测算法

4.1 引言

为了进一步提高安全帽与反光衣的检测精度，降低在实际环境中由于误检漏检而造成的损失，本章节提出了一高精度 YOLOv5s 检测算法。首先，在颈部网络中，将原有的 C3 模块替换为 C3SAC 提高检测精度。其次，在网络中添加了 CA 注意力机制，提高对检测物体的注意力。最后，将 CIoU 损失函数替换为 SIoU 损失函数，改善了对位置损失的计算，使用 CARAFE 轻量级上采样算子，提供更丰富的语义信息。经过一系列实验验证，确认了所提算法的有效性。

4.2 改进的 YOLOv5 算法

4.2.1 SAConv 模块

在传统的卷积神经网络中，卷积操作的感受野是固定的，这对于捕捉图像中不同尺度的特征是有局限性的^[50]。SAConv (Switchable Atrous Convolution) 通过采用可切换的空洞卷积，可以动态地调整卷积核的感受野大小，从而能够有效地捕捉多尺度上下文信息^[51]。

具体来说，SAConv 通过网络中引入多个并行的分支，每个分支使用不同扩张率的空洞卷积。这些分支可以捕获不同尺度的上下文信息。每个分支独立地应用空洞卷积来提取特征图。通过一个卷积层将所有分支的特征图融合在一起。这个卷积层的权重是通过训练过程学习的，它决定了不同尺度特征的相对重要性。融合后的特征图被送入后续的网络层。SAConv 结构如图 5-1 所示。SAConv 模块由三个部分构成，分别为 SAConv 组件和前后的两个全局上下文模块，SAConv 的结构如图 4-1 所示。SAConv 的输出公式为如下：

$$y = S(x) \cdot \text{Conv}(x, w, 1) + (1 - S(x)) \cdot \text{Conv}(x, w + \Delta w, r) \quad (4.1)$$

其中， x 和 y 表示输入和输出。 r 表示超参数； Δw 表示可训练权重； $S(x)$ 为开关函数，以适应不同尺度。

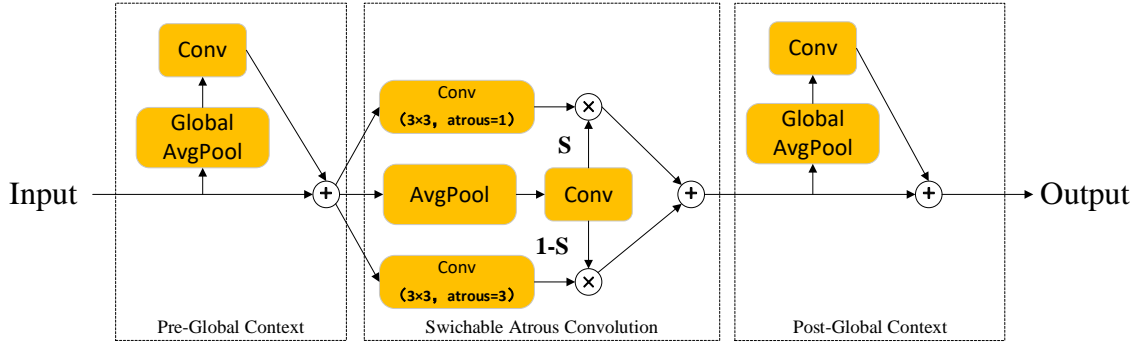


图 4-1 SACConv 结构图

Fig. 4-1 SACConv structure diagram

4.2.2 SIOU 损失函数

SIOU 损失函数^[52]考虑了真实框和预测框之间方向的匹配问题，考虑到所需回归之间的向量角度，重新定义了惩罚指标，加快了收敛速度。SIOU 损失函数由四种成本函数组成，角度成本 (Angle cost)，距离成本 (Distance cost)，形状成本 (Shape cost) 和 IoU 成本 (IoU cost)^[53]。角度成本计算公式子如下：

$$\begin{cases} c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \\ \sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \\ x = \frac{c_h}{\sigma} = \sin(\alpha) \\ \Lambda = 1 - 2 * \sin^2(\arcsin(x) - \frac{\pi}{4}) \end{cases} \quad (4.2)$$

其中， $b_{c_x}^{gt}$ 和 $b_{c_y}^{gt}$ 表示真实框中心点坐标； b_{c_x} 和 b_{c_y} 表示预测框中心点坐标； c_h 表示两框中心点的高度差； σ 表示两框中心点的距离。

距离成本计算公式子如下：

$$\begin{cases} \rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{v_w} \right)^2 \\ \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{v_h} \right)^2 \\ r = 1 - \Lambda \\ \Delta = \sum_{t=x,y} (1 - e^{-r\rho_t}) \end{cases} \quad (4.3)$$

其中， v_w 和 v_h 表示最小包围框的宽度和高度； ρ_x 和 ρ_y 表示真实框和预测框的偏离程度。

形状成本算公式子如下：

$$\begin{cases} w_h = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ h_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \\ \Omega = \sum_{t=w,h} (1 - e^{-wt})^\theta \end{cases} \quad (4.4)$$

其中, w 和 h 表示预测框的宽和高; w^{gt} 和 h^{gt} 表示真实框的宽和高。

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4.5)$$

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (4.6)$$

其中, A 和 B 分别表示真实框和预测框。 L_{SIoU} 表示 SIoU 损失函数。SIoU 损失函数有助于模型更准确地预测边界框的位置和大小, 从而提高检测的整体性能。

4.2.3 CARAFE 轻量级上采样算子

YOLOv5s 使用最近邻插值作为上采样方式, 将输入图像中的每个像素复制到输出图像的最近邻像素位置来实现上采样, 无法提取丰富的语义信息。轻量级上采样算子 CARAFE (Content-Aware ReAssembly of FEatures)^[54] 是一种用于高效图像处理的技术, 能够保持特征信息的完整, 在更大感受野上聚合上下文信息。CARAFE 分别由上采样核预测模块 (Kernel Prediction Module) 和特征重组模块 (Content-aware Reassembly Module) 两个模块构成^[55]。上采样核预测模块能够通过通道压缩减小特征图的通道数量, 进而降低了计算量; 通过特征编码扩大感受野, 提高对特征信息的利用; 最后利用 softmax 函数进行归一化处理, 使上采样核权重参数和为 1。特征重组是将输出特征图的每个位置映射回输入特征, 取出以其为中心的 $k_{up} \times k_{up}$ 区域, 与这点预测的上采样核作点积, 得到最终的输出值。所获得的特征图更能获得局部区域相关点的信息。上采样算子 CARAFE 在网络结构中的位置如下图 4-2 所示。

4.2.4 改进后的 YOLOv5s 网络结构

改进后的 YOLOv5s 网络模型如图 4-3 所示, 首先, 在颈部网络的 C3 模块中引入了 SAConv 卷积, 组合成了新的模块 C3SAC, 增强了颈部网络的特征提取能力。然后, 在主干网络和颈部网络分别添加了 CA 注意力机制模块, 以增强对目标的关注度。最后, 将 YOLOv5s 中的损失函数替换为 SIoU 损失函数。并且将最近邻插值上采样替换为 CARAFE 轻量级上采样, 减少特征的损失。

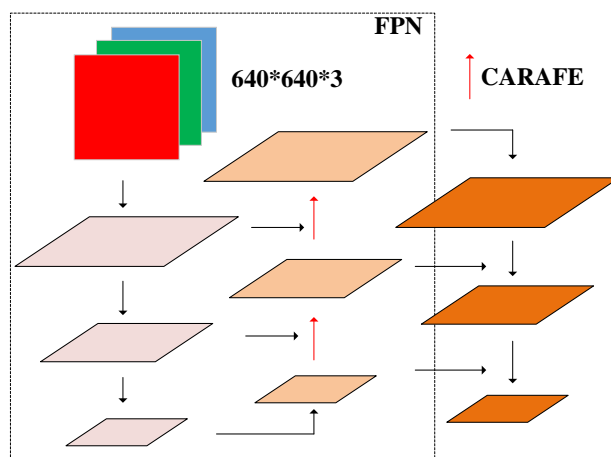


图 4-2 PANet 网络结构中的 CARAFE

Fig. 4-2 CARAFE in PANet network architecture

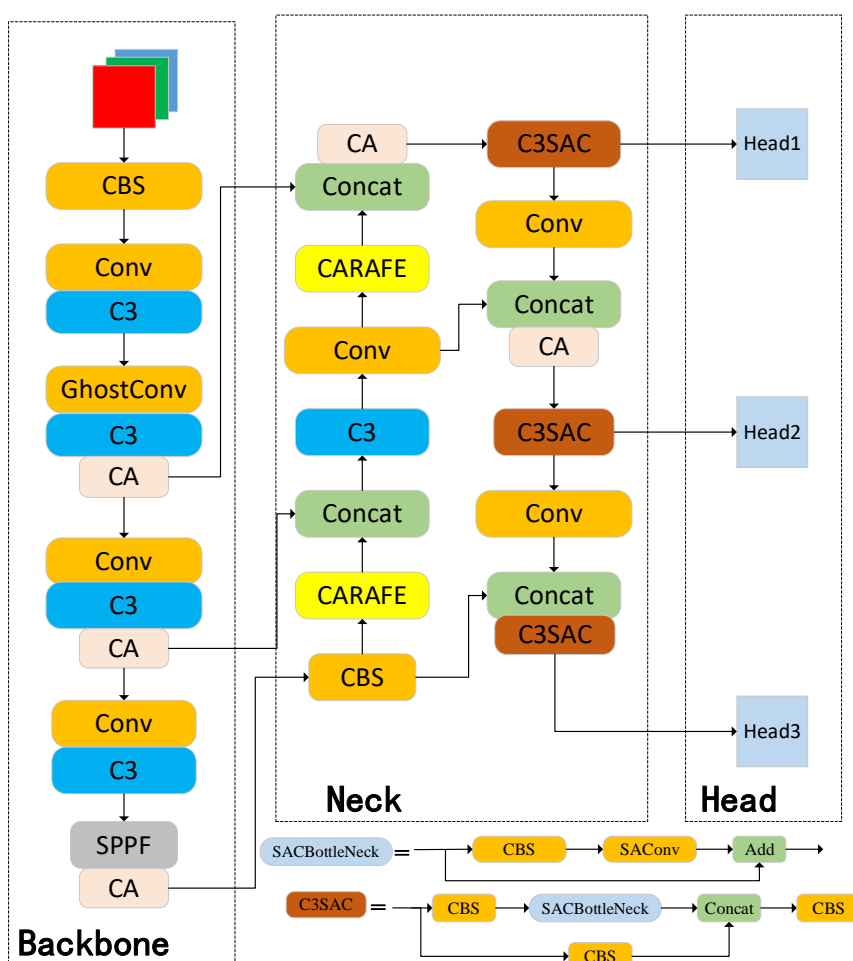


图 4-3 改进的 YOLOv5s 网络结构

Fig. 4-3 The improved YOLOv5s network structure

4.3 实验结果与分析

4.3.1 实验结果

本章使用的实验环境与第三章相同，也使用了相同的数据集。改进的 YOLOv5s 的 mAP@0.5 增长曲线如图 4-4 所示，第 65 个 epoch 时，mAP@0.5 达到了 90% 以上，并且开始逐渐收敛，最终达到 96.1%。YOLOv5s 算法和改进的 YOLOv5s 的 mAP 值和不同类别 AP 分别如表 4-1 和表 4-2 所示。改进的 YOLOv5s 算法中，反光衣的精确率，召回率，mAP@0.5 和 mAP@0.5:0.95 最高，分别达到 97.8%，94.7%，98.5% 和 76.8%。改进的 YOLOv5s 的总平均精度 mAP@0.5 提高了 2.0%，无安全帽的 mAP@0.5 提高了 1.9%，反光衣提高了 1.8%，安全帽提高了 1.9%，其他衣服提高了 2.2%。

表 4-1 改进的 YOLOv5s 不同类别 AP

Tab. 4-1 Improved YOLOv5s APs of different classes

类别	精确率/%	召回率/%	mAP@0.5	mAP@0.5:0.95
All	97.1	91.6	96.1	71.9
nohelmet	96.6	86.6	92.5	67.7
Reflective_clothes	97.8	94.7	98.5	76.8
helmet	96.3	92.7	96.5	73.3
Other_clothes	97.6	92.3	96.8	69.6

表 4-2 YOLOv5s 不同类别 AP

Tab. 4-2 YOLOv5s APs of different classes

类别	精确率/%	召回率/%	mAP@0.5	mAP@0.5:0.95
All	95.2	89	94.1	67.8
nohelmet	95.5	84.1	90.6	64.1
Reflective_clothes	95.9	92.2	96.7	72.1
helmet	93.7	90.2	94.6	70.2
Other_clothes	95.5	89.5	94.6	64.8

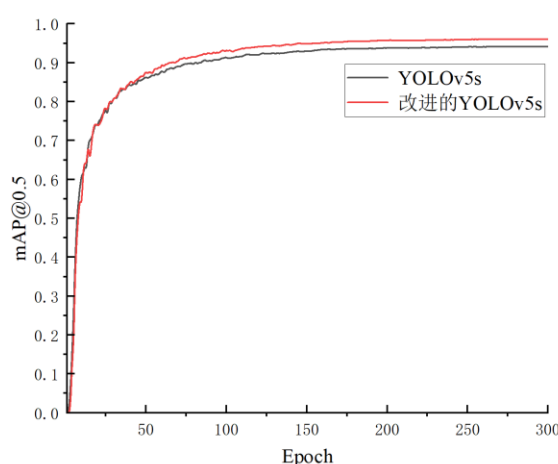


图 4-4 改进 YOLOv5s mAP 曲线

Fig. 4-4 Improved YOLOv5s mAP Curve

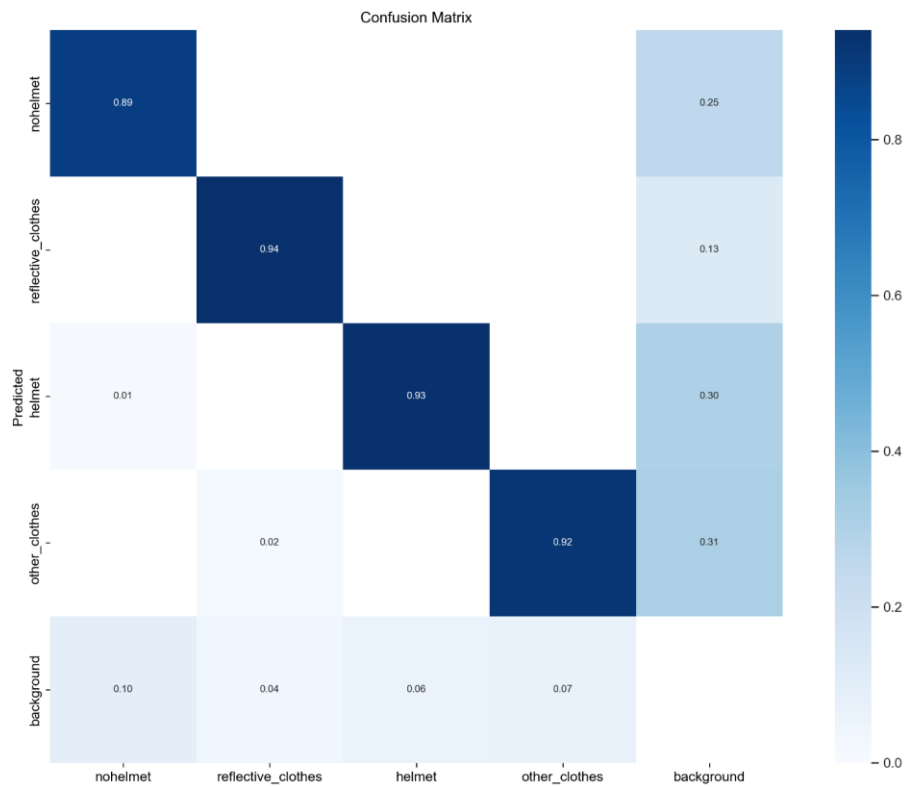
4.3.2 消融实验

为了验证本章提出的改进策略的有效性,进行了以下消融实验。从表 4-3 中可知,将 C3 模块替换为 C3SAC 后,模型的平均精度 $mAP@0.5$ 提高了 0.6%;此外,在引入了 CA 之后,模型的 $mAP@0.5$ 提升了 0.9%;CIoU 损失函数替换为 SIoU 损失函数后,模型的 $mAP@0.5$ 提高了 0.5%;引入 CARAFE 上采样算子后, $mAP@0.5$ 提高了 0.4%。其中对 $mAP@0.5:0.95$ 提升最明显的为 CA 注意力机制,达到 69.5%,提高了 1.7%。最后改进的 YOLOv5s 检测算法的识别精度 $mAP@0.5$ 达到 96.1%,提高了 2.0%。 $mAP@0.5:0.95$ 达到 71.9%,提高了 4.1%。

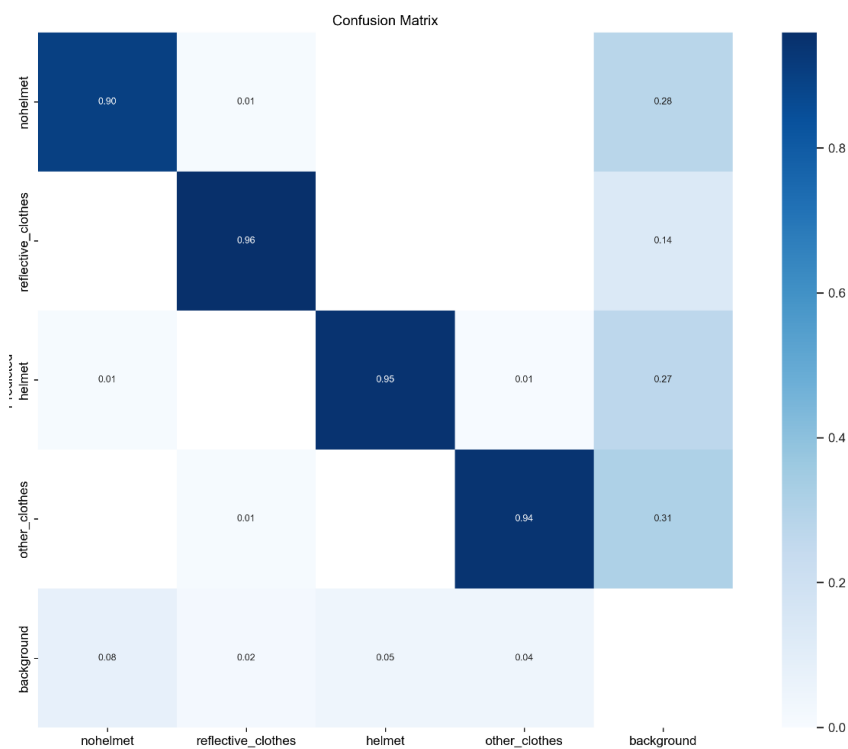
4.3.3 对比实验

混淆矩阵是对分类问题预测结果的总结。有助于直观的了解到目标检测模型的分类错误。原始算法的混淆矩阵如图 4-5 (a) 所示,改进算法的混淆矩阵如图 4-5 (b) 所示。对比可以明显的看出,改进算法各类别正确识别的概率要高于原始算法。

为了清晰地呈现检测成果以及改进算法的优越性,制作了图 4-6 的检测结果对比。在不同环境下,检测安全帽与反光衣的佩戴情况。从检测结果可以看出,改进后的 YOLOv5 算法的对于四种类别的检测置信度更高,且预测框更贴合被检测物体。



(a) 原始算法混淆矩阵



(b) 改进算法混淆矩阵

图 4-5 混淆矩阵

Fig. 4-5 confusion matrix

表 4-3 消融实验

Tab. 4-3 Ablation experiment

模型	精确率/%	召回率/%	mAP@0.5	mAP@0.5:0.95
YOLOv5s	95.2	89	94.1	67.8
+C3SAC	94.5	90.5	94.7	68.8
+CA	96	85.6	95	69.5
+SIoU	95.6	89.9	94.6	68
+CARAFE	95.6	89.8	94.5	68.6
+CA+SIoU	96.3	90.7	95.3	69.9
+CA+ SIoU+C3SAC	96.7	91.6	95.7	71.2
+CA+ SIoU+C3SAC+CARAFE	97.1	91.6	96.1	71.9



图 4-6 检测结果对比

Fig. 4-6 Comparison of test results

为了展示改进算法与其他同类型算法的优势，将改进后的 YOLOv5s 检测算法与 YOLOv3，YOLOv3-Tiny，YOLOv7 和 YOLOv7-Tiny 进行了以下对比，如表 4-4 所示。可以看改进 YOLOv5s 的 mAP@0.5 和 mAP@0.5:0.95 均高于其他检测算法。YOLOv3-Tiny 与 YOLOv7-Tiny 这种相较轻量化的算法的检测精度均低于原始的 YOLOv5s。其中 YOLOv3-Tiny 的检测精度最低，mAP@0.5 和 mAP@0.5:0.95 分别达到了 86.4%和 53.5%。YOLOv3 和 YOLOv7 的检测精度均高于原始的 YOLOv5s 检测算法，但是均低于改进后的 YOLOv5s。

表 4-4 对比实验

Tab. 4-4 Contrast experiment

模型	精确率/%	召回率/%	mAP@0.5	mAP@0.5:0.95
YOLOv5s	95.2	89	94.1	67.8
YOLOv3-Tiny	89.5	77.5	86.4	53.5
YOLOv7-Tiny	92.5	86.1	92.7	62.4
YOLOv3	95.9	90.4	94.6	70.4
YOLOv7	95.7	90.7	95.5	71
改进的 YOLOv5s	97.1	91.6	96.1	71.9

4.4 本章小结

本章提出了一种改进的 YOLOv5s 高精度检测算法。首先,介绍 SAConv 卷积的结构, C3SAC 模块的组成。其次,介绍了 SIoU 损失函数的计算公式及其优点,还介绍了 CARAFE 轻量级上采样算子的结构。最后,展示了实验结果,通过消融实验与对比实验验证了所改进算法的有效性,展示了其在检测算法中的优势,实验结果表明改进的 YOLOv5s 检测算法 mAP@0.5 达到 96.1%,提高了 2.0%; mAP@0.5:0.95 达到 71.9%,提高了 4.1%。

5 总结与展望

5.1 总结

在施工环境与交通作业中，安全帽与反光衣的配带起到了至关重要的作用。为了适应实际环境中复杂多变的场景，本文提出两种改进的 YOLOv5s 检测算法。第一种为轻量化的 YOLOv5s 检测算法，该检测算法是为了解决在实际的安全帽与反光衣检测过程中存在的硬件内存，算力不足的问题。并且在网络模型轻量化的同时，提高了一定的检测精度。第二种为更高精度的 YOLOv5s 检测算法，该检测算法进一步的提高了算法的平均精度，并且没有过多的增加 YOLOv5s 检测算法原有的模型复杂度。本文也制作了适用于安全帽与反光衣的数据集，使用 LabelImg 进行了标注。主要标注为 4 类佩戴情况，分别安全帽，无安全帽，反光衣和其他衣服。这四种情况相较于单一的安全帽或反光衣检测，更加的符合实际场景和工作人员的佩戴情况。

其中轻量化的 YOLOv5s 检测算法，是将 GhostNet 中的轻量型的 Ghost 模块融合到了 YOLOv5s 的 C3 模块中，并且 GhostConv 卷积替换了普通卷积，在尽可能的减少模型复杂度同时，保持了一定的检测精度。同时在模型的主干网络和特征融合层引入了注意力机制模块，忽略实际环境中的复杂背景，而关注安全帽与反光衣区域，进一步提高了模型的特征提取能力。最后引入了 WIoU 损失函数，使模型更能关注普通质量的锚框，提高了算法检测精度。

其中更高性能的 YOLOv5s 检测算法，是将 SAConv 卷积模块引入到了颈部网络的 C3 模块中，该卷积可以提供更大的感受野，从而提高了 C3 的特征提取能力。注意力机制模块提高了对安全帽与反光衣的检测。引入了 SIoU 损失函数使模型更好的计算位置损失，从而提高检测精度，使用 CARAFE 算子来丰富上采样过程中的语义信息。

5.2 展望

1.数据集。本文采用的数据集为网络采集的自制数据集，以后可以在实地采集更有针对性的数据集。并且可以增加标注类别，比如未正确佩戴安全帽与反光衣的情况。也可以增加其他防护类别，如安全绳等。也可以考虑天气因素，雨天，雾天和雪天等天气对检测精度的影响。考虑在夜晚等光线不足时的情况。

2.进一步提高模型轻量化与精度。本文已经提出了两种改进的 YOLOv5s 检测算法，在数据集上表现良好，且通过实验验证了其有效性。但是在实现了模型的轻量化与精度的提高方面，还存在着一些进步的空间。

3.实地检验。由于实验设备与场地的影响，本文主要使用实验室计算机对安全帽与反光衣的检测，未实际在嵌入式与移动设备上进行检测。之后可以在现场的检测环境，使用当地的设备，测试实际的工作表现。

4.系统设计。本文目前使用 pytorch 深度学习框架实现了安全帽与反光衣的检测，没有检测界面，不适合在现场使用。以后可以设计检测平台，在现场供非专业人员更方便使用检测算法。

参考文献

- [1]WÓJCIK B, ŻARSKI M, KSIAŻEK K, et al. Hard hat wearing detection based on head keypoint localization[J]. arXiv:2106.10944, 2021.
- [2]VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [C]. Conference on Computer Vision and Pattern Recognition, 2001, 1: 511-518.
- [3]DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 1: 886-893.
- [4]HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and their Applications, 1998, 13(4): 18-28.
- [5]FELZENSZWALB P, GIRSHICK R, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627–1645.
- [6]LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [7]UIJLINGS J, VAN DE SANDE K, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision 2013, 104(2):154–171.
- [8]KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [9]HOSANG J, BENENSON R, DOLLÁR P, et al. What makes for effective detection proposals[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(4): 814-830.
- [10]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, 2014: 580-58
- [11]HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [12]SHICK R. Fast r-cnn[C]. IEEE International Conference on Computer Vision, 2015:

- 1440-1448.
- [13]REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2017, 39(6): 1137-1149.
- [14]HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2):386-397.
- [15]REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [16]LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. European Conference on Computer Vision, 2016: 21-37.
- [17]REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271
- [18]REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804.02767, 2018.
- [19]BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv:2004.10934, 2020.
- [20]PARK M W, ELSAFTY N, ZHU Z. Hardhat-wearing Detection for Enhancing On-Site Safety of Construction Workers [J]. Journal of Construction Engineering Management, 2015,141(9):1-16.
- [21]SUN X, XU K, WANG S, et al. Detection and tracking of safety helmet in factory environment[J]. Measurement Science and Technology, 2021, 32(10): 105406
- [22]WU H, ZHAO J. An intelligent vision-based approach for helmet identification for work safety[J]. Computers in Industry, 2018, 100: 267-277.
- [23]冯国臣, 陈艳艳, 陈宁等.机器视觉的安全帽自动识别技术研究[J]. 机械设计与制造工程, 2015, 44(10): 39-42.
- [24]CHENG R, HE X, ZHENG, Z, et al. Multi-Scale Safety Helmet Detection Based on SAS-YOLOv3-Tiny[J]. Applied Sciences, 2021, 11(8): 3652-3652.
- [25]ZHANG Y J, XIAO F S, LU Z M. Helmet Wearing State Detection Based on Improved Yolov5s[J]. Sensors. 2022; 22(24): 9843-9843.

- [26]CHEN J, DENG S, WANG P, et al. Lightweight Helmet Detection Algorithm Using an Improved YOLOv4[J]. Sensors. 2023, 23(3): 1256-1256.
- [27]SONG H, ZHANG X, SONG J, et al. Detection and tracking of safety helmet based on DeepSort and YOLOv5[J]. Multimedia Tools and Applications, 2023, 82(7): 10781-10794.
- [28]HAYAT A, MORGADO-DIAS F. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety[J]. Applied Sciences, 2022, 12(16): 8268-8268.
- [29]XU H, WU Z. MCX-YOLOv5: efficient helmet detection in complex power warehouse scenarios[J]. Journal of Real-time Image Processing, 2024, 21(2): 27.
- [30]ZHANG Y, QIU Y, BAI H. FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising[J]. Electronics (Basel), 2023, 12(13):2902.
- [31]Ma W, Guan Z, Wang X, et al. YOLO-FL: A Target Detection Algorithm for Reflective Clothing Wearing Inspection[J]. Displays, 2023, 80: 102561.
- [32]谢国波, 谢建辉, 林志毅等. 基于 CT-YOLOX 的反光衣与安全帽检测算法[J]. 国外电子测量技术, 2023, 42(10): 51-58.
- [33]骆杰豪. 施工人员安全穿戴检测关键技术研究[D]. 华南理工大学, 2022.
- [34]程换新, 蒋泽芹, 程力等. 基于改进 YOLOX-S 的安全帽反光衣检测算法[J]. 电子测量技术, 2022, 45(06): 130-135.
- [35]韩泽佳, 肖秦琨, 张立旗. 基于改进 SSD 安全头盔反光衣检测算法[J]. 自动化与仪表, 2021, 36(09): 63-68.
- [36]YU H, TAO Y, CUI W, et al. Research on application of helmet wearing detection improved by YOLOv4 algorithm[J]. Mathematical Biosciences and Engineering, 2023, 20(5): 8685-8707.
- [37]LI H, WU D, ZHANG W, et al. (2024). YOLO-PL:Helmet wearing detection algorithm based on improved YOLOv4[J]. Digital Signal Processing, 2024, 144: 104283.
- [38]TAI W, WANG Z, LI, W, et al. DAAM-YOLOV5: A Helmet Detection Algorithm Combined with Dynamic Anchor Box and Attention Mechanism[J]. Electronics

- (Basel), 2023, 12(9): 2094.
- [39]张学立, 贾新春, 王美刚等. 安全与反光衣的轻量化检测: 改进 YOLOv5s 的算法[J]. 计算机工程与应用, 2024, 60(01): 104-109.
- [40]SHAN C, LIU H, YU Y. Research on improved algorithm for helmet detection based on YOLOv5[J]. Scientific Reports, 2023, 13(1): 18056.
- [41]ELESAWY A, MOHAMMED ABDELKADER E, OSMAN H. A Detailed Comparative Analysis of You Only Look Once-Based Architectures for the Detection of Personal Protective Equipment on Construction Sites[J]. Eng. 2024; 5(1): 347-366.
- [42]HAN K, WANG, Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]. IEEE Conference on Computer Vision and Pattern Recognition Seattle, 2020: 1577-1586.
- [43]李昂, 孙士杰, 张朝阳等. 改进 YOLOv5s 的轨道障碍物检测模型轻量化研究[J]. 计算机工程与应用, 2023, 59(04): 197-207.
- [44]邱天衡, 王玲, 王鹏, 白燕娥. 基于改进 YOLOv5 的目标检测算法研究[J]. 计算机工程与应用, 2022, 58(13): 63-73.
- [45]HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, 2020: 13708-13717.
- [46]PANG L, LI B, ZHANG F, et al. A lightweight YOLOv5-MNE algorithm for SAR Ship detection[J]. Sensors, 2022, 22(18): 7088.
- [47]宋晓凤, 吴云军, 刘冰冰等. 改进 YOLOv5s 算法的安全帽佩戴检测[J]. 计算机工程与应用, 2023, 59(02): 194-201.
- [48]郑陆石, 胡晓锋, 于伟国等. 基于改进 YOLOv7-tiny 的坦克车辆检测方法[J]. 兵器装备工程学报, 2023, 44(12): 285-292.
- [49]TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J], arXiv:2301.10051, 2023.
- [50]QIAO S, CHEN L, YUILLE A L. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021:10208-10219.
- [51]梁秀满, 李然, 于海峰等. 改进 YOLOv7 的水下目标检测算法[J]. 计算机工程与

应用: 1-14[2024-02-26].

- [52]GEVORGYAN Z. SIOU Loss: More Powerful Learning for Bounding Box Regression, arXiv:2205.12740,2022.
- [53]孙迟, 刘晓文. 基于 YOLOv7-tiny 改进的矿工安全帽检测[J]. 中国科技论文, 2023, 18(11): 1250-1256+1274.
- [54]WANG J, CHEN, K XU R, et al. CARAFE: Content-Aware ReAssembly of Features[C]. IEEE International Conference on Computer Vision, 2019: 3007-3016.
- [55]WANG H, ZHANG S, ZHAO S, et al. Fast detection of cannibalism behavior of juvenile fish based on deep learning[J]. Computers and Electronics in Agriculture, 2022, 198: 107033.