分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220920102

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于YOLO的防毒面具佩戴检测方法的研究

论 文 作 者　朱毅
指 导 教 师　汪军(教授)
学科（专业）　软件工程
研 究 方 向　领域软件工程

论文提交日期：2025 年 5 月 9 日

分类号：TP391　　单位代码：10363
密　级：公开　　学　　号：2220920102

题　　目 基于 YOLO 的防毒面具佩戴检测方法的研究

英文并列

题　　目 Reseach on Gas Mask Wearing Detection Based on YOLO

学　生　姓　名：朱毅
校　内　教　师：汪军(教授)

专业 学位 类别：软件工程

研究方向（领域）：领域软件工程

论文答辩日期：2025 年 5 月 10 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风，所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果，除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全认识到本声明的法律后果由本人承担。

学位论文作者签名：朱毅

日期：2015 年 5 月 9 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

保密 □，在_____年解密后适用本版权书。

本学位论文属于

不保密 ☑。

学位论文作者签字：朱毅　　指导老师签字：

日期：2025 年 5 月 9 日　　日期：2025 年 5 月 9 日

基于 YOLO 的防毒面具佩戴检测方法的研究

摘 要

在化工生产环境中，防毒面具作为关键的个人防护装备，能够高效过滤有害气体、粉尘及悬浮颗粒物。但是因未正确佩戴防毒面具而导致的呼吸道损伤及中毒事故仍频繁发生。为解决这一问题，需要实现对防毒面具佩戴情况进行检测，通过实时采集并分析作业人员面部区域的图像数据，自动识别其是否规范佩戴防毒面具。本文基于工业场景中的实际需求，针对该任务面临的复杂背景干扰、目标多尺度变化及类别分布不均等技术难点，本文采用基于深度学习的目标检测框架，构建了一套高效的防毒面具检测系统，并针对上述挑战提出了一系列模型优化策略。本文的主要贡献与创新点如下：

（1）基于对工业现场的实地考察，本研究结合防毒面具佩戴检测任务的实际需求，构建了一个面向工业场景的专用防毒面具佩戴检测数据集。通过对该数据集的特征进行分析揭示了任务中存在的关键难点，为模型训练提供高质量的数据支持。

（2）为解决复杂背景干扰和样本不均衡问题，以 YOLOv5 为基础模型，提出了一种改进模型。首先，在原有三个检测层的基础上，引入了一个小目标检测头 SODH，以增强模型对微小目标的捕捉能力；其次，在 Neck 层中嵌入 CBAM 模块，通过增强待检测目标的位置特征来有效区分目标与背景信息，提升检测精度，同时避免了计算复杂度的显著增加；最后，将原模型的二元交叉熵损失函数替换为基于样

本质量评估的 Varifocal Loss，通过动态权重调节机制使模型更加关注正样本，从而改善检测准确率。

（3）为解决特征层信息融合差和多尺度目标检测问题，以 YOLOv8 为基准框架，提出了一种改进模型。首先，在空间金字塔池化融合模块中引入了大核分离卷积注意力模块 LSKA，以增强不同特征层之间的语义信息融合能力；其次，将主干网络中 P2 层的输出集成到颈部结构，并额外增加了一个小目标检测头 SODH，从而提升模型对小尺度目标的检测性能；最后，在检测头部引入了高效多尺度注意力模块 EMA，能够有效学习待检测目标的多尺度语义信息。

（4）为解决小目标检测困难、特征信息提取能力差的问题，提出基于 YOLOv10 的改进模型，首先设计了小目标语义信息增强金字塔 STSIEP，以增强对小目标的语义信息特征提取效果，其次在 Neck 中使用 RCS-OSA 替换原模型的 C2f，增强网络对不同特征层的信息提取能力，最后使用指数滑动样本加权函数替换原模型的 BCE Loss，解决样本不平衡的问题，增强了模型的鲁棒性。

关键词：防毒面具佩戴检测，深度学习，YOLOv5，YOLOv8，YOLOv10

RESEACH ON GAS MASK WEARING DETECTION BASED ON YOLO

ABSTRACT

In the chemical production environment, gas masks, as key personal protective equipment, can efficiently filter harmful gases, dust and suspended particles. However, respiratory injuries and poisoning accidents caused by not wearing gas masks correctly still occur frequently. In order to solve this problem, it is necessary to detect the wearing of gas masks, and automatically identify whether the operator is wearing a gas mask by collecting and analysing the image data of theirs facial area in real time. Based on the actual needs in industrial scenarios and the technical difficulties faced by this task, such as complex background interference, multi-scale changes in targets and uneven distribution of categories, this paper adopts a target detection framework based on deep learning, constructs a set of highly efficient gas mask detection system, and proposes a series of model optimisation strategies for the above challenges. The main contributions and innovations of this paper are as follows:

(1)The present study is based on field visits to industrial sites, the results of which are used to construct a specialised gas mask-wearing detection dataset for industrial scenarios. This is achieved by combining the actual needs of the gas mask-wearing detection task. The features of

this dataset are analysed to reveal the key difficulties in the task and provide high-quality data support for model training.

(2) To address the issues of complex background interference and class imbalance, an improved model based on YOLOv5 is proposed for the task of detecting mask-wearing. First, to tackle the problem of missed detection of small-scale targets, a SODH is added to enhance the model's ability to capture tiny targets. Secondly, a CBAM module is embedded in the Neck layer to effectively differentiate target and background information by enhancing the positional features of the objects being detected, improving detection accuracy while avoiding significant increases in computational complexity. Finally, the original BCE Loss is replaced with Varifocal Loss, which is based on sample quality assessment. This dynamic weight adjustment mechanism makes the model focus more on positive samples, thereby improving detection accuracy.

(3) To address the issues of poor feature layer information fusion and multi-scale object detection, an improved model is proposed based on the YOLOv8 framework. First, a Large Kernel Separable Convolution Attention module is introduced in the Spatial Pyramid Pooling fusion module to enhance the semantic information fusion ability between different feature layers. Secondly, the output of the P2 layer in the backbone network is integrated into the neck structure, and an additional

small object detection head is added to improve the model's performance in detecting small-scale objects. Finally, an efficient multi-scale attention module is introduced at the detection head to effectively learn multi-scale semantic information of the objects to be detected.

(4) To address the difficulties in small object detection, poor feature information extraction, an improved model based on YOLOv10 is proposed. First, a Small Target Semantic Information Enhancement Pyramid is designed to improve the semantic information feature fusion for small objects. Next, RCS-OSA is used in the Neck to replace the original model's C2f, enhancing the network's ability to extract information from different feature layers. Finally, an exponential sliding sample weighting function is used to replace the original model's BCE Loss, addressing the sample imbalance issue and improving the model's robustness.

KEY WORDS: Gas Mask Wearing Detection, Deep Learning, YOLOv5, YOLOv8, YOLOv10

目 录

第 1 章 绪论

1.1 研究背景与意义

随着全球工业化进程的持续推进，政府、企业、地区和国际层面对工业安全，特别是职业安全和健康的认识都在提高。尤其是在化工、矿业、冶金、建筑等高风险行业，工人们常常面临各种潜在的安全隐患，而其中危险性较大的便是有毒有害气体的泄漏。为了有效避免气体中毒事故的发生，防毒面具作为一种重要的个人防护装备，发挥着至关重要的作用。然而，防毒面具的佩戴是否规范、是否到位，直接关系到防护效果，因此确保工人在工作过程中佩戴防毒面具并且佩戴合规，成为提高工业安全、降低事故风险的重要环节。尽管防毒面具的使用已经被广泛普及，并且相关标准和法规也有明确规定，但由于工作环境复杂、工人素质参差不齐、工作压力大等原因，传统的人工检查方法常常面临着执行效率低、判断不准确、漏检和错检等问题，无法做到实时、有效的监控和管理。特别是在一些大型企业或特殊环境中，工人数量庞大且环境复杂，人工检查的可行性和准确性大大降低。

近年来，随着深度学习算法创新和高性能计算平台的迭代升级，基于图像识别的自动化检测技术在工业安全领域取得了显著进展。各种工业检测应用如安全帽佩戴检测[1-3]、危险动作检测[4]、抽烟检测[5]等，已经在多种生产环境中得到了初步应用，并取得了良好的效果。这些技术不仅有效提高了安全管理的自动化水平，还显著降低了人工检查的负担和错误率。特别是在一些人员密集、环境复杂的工作场景中，自动化监控系统通过图像识别技术能够实时判断工人是否佩戴了安全帽、是否存在危险动作等，从而及时发出预警，防止安全事故的发生。例如，安全帽佩戴检测系统能够通过摄像头监控工人是否戴上了合格的安全帽，而危险动作检测技术则可以自动识别工人是否存在高空作业、未系安全带等行为，以保障工人的生命安全。在这一背景下，防毒面具佩戴检测作为一种新兴的工业安全检测应用，具有重要的研究价值。尽管现有的工业检测系统已经在一定程度上取得了成功，但与安全帽检测和危险动作检测等领域相比，防毒面具佩戴检测仍处于较为初步的研究阶段。由于工人和管理人员都容易忽视呼吸安全，工人持续接触有毒有害气体、粉尘会显著提高患职业病的可能性[6]，典型的职业病有矽肺、尘肺、石棉肺、哮喘等。因此，为了提高工作人员正确佩戴防毒面具意识和企业

安全检查水平，推动整个行业的技术革新和数字化转型，建立一个稳定且精确的防毒面具佩戴检测系统是具有实际应用价值的。

1.2 国内外研究现状及其发展趋势

在计算机视觉领域，目标检测算法的研究始终占据重要地位。相较于图像分类任务，目标检测面临更大的技术挑战。图像分类仅需实现目标的类别识别，而目标检测则需同时完成目标定位与分类双重任务，这要求综合运用图像处理技术、机器学习等多学科知识，以实现对图像中特定目标的精确识别与定位。目标分类旨在判定图像中是否存在特定目标物体，目标定位则需确定目标的空间位置信息，通常以边界框的形式进行标注。因此，目标检测算法不仅需要实现目标的准确分类，还必须提供目标在图像中的精确位置信息。图 1-1 展示了目标检测算法的演进历程。

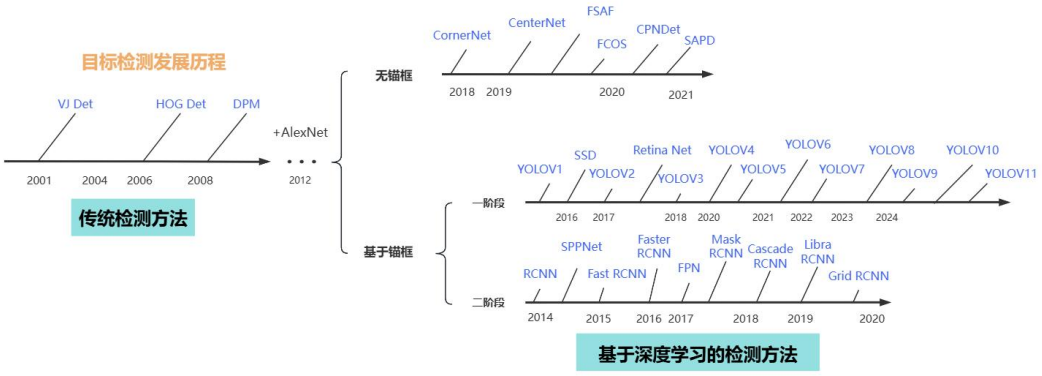


图 1-1 目标检测算法发展历程

Fig. 1-1 Development of object detection algorithms

1.2.1 目标检测算法研究现状

(1) 传统目标检测研究现状

传统目标检测方法主要依托图像处理技术实现目标的分类与定位任务。其技术框架通常包含以下核心步骤：首先，通过滑动窗口或选择性搜索等区域生成策略，在输入图像中提取潜在目标区域；其次，利用人工设计的特征提取器从候选区域中获取特征表示；最后，基于预定义的底层特征完成目标的检测与分类任务。这一方法依赖于人工特征工程，其性能受限于特征提取器的设计质量。

2004 年，Lowe 等人提出了基于特征点的尺度不变特征变换（Scale-Invariant Feature Transform, SIFT）方法[7]。该方法在图像特征提取与匹配任务中表现出色，

但其计算复杂度较高，且在处理光滑区域和纹理缺失图像时存在局限性。为克服上述问题，Dalal 等人于 2005 年提出方向梯度直方图（Histogram of Oriented Gradients, HOG）算法[8]。该算法能够有效提取目标的形状与边缘特征，对光照变化和局部遮挡具有一定的鲁棒性，但其对光照与背景变化的敏感性较高，且尺度不变性表现欠佳。

为进一步提升检测性能，Pedro 等人于 2008 年在 HOG 算法的基础上提出了可变形部件模型（Deformable Part Model, DPM）算法[9]，该算法实现了对目标多样性与形变的有效识别。DPM 方法在目标检测任务中取得了显著的性能提升，成为传统目标检测领域的重要里程碑。

传统目标检测方法尽管在某些特定任务中表现出一定的性能优势，但其固有局限性显著制约了其实际应用范围。首先，这类方法依赖于人工设计的特征提取器，需要针对特定任务手动选择特征，这种设计方式难以适应复杂场景下的多样化目标检测需求。其次，分类器的性能与训练数据的规模和质量密切相关，在数据量有限或类别分布不均衡的情况下，分类器的泛化能力往往显著下降。此外，传统方法通常采用多阶段处理流程，包括候选区域生成、特征提取和分类等独立步骤，这种串行处理机制导致计算效率低下，难以满足实时性要求。上述局限性严重限制了传统目标检测方法在实际应用场景中的推广与部署。

(2) 基于深度学习的目标检测算法研究现状

当前，人工智能技术蓬勃发展，基于深度学习的目标检测方法正在逐步取代传统方法。首先，采用数据驱动，自动提取目标的深层语义特征，从而避免了人工特征设计的局限性；其次，这类方法在检测速度与精度之间实现了更好的平衡，能够满足工业场景下的实时处理需求；此外，具有较强的泛化能力，能够适应多样化的场景与目标类别，展现出更广泛的适用性。这些优势使其成为当前目标检测领域的主流研究方向。

2012 年，Krizhevsky 等学者在 ImageNet LSVRC-2012 图像分类竞赛中首次引入深度卷积神经网络（AlexNet）[10]，并以显著优势取得了突破性成果。此后，卷积神经网络（Convolutional Neural Network, CNN）在图像分类任务中得到广泛应用，并衍生出一系列具有重要影响力的经典架构，例如 VGG[11]、GoogLeNet[12]、ResNet[13]以及 DenseNet[14]等。这些模型的持续演进不仅推动了图像分类技术的

革新，也为目标检测领域的发展提供了新的理论基础与技术支撑，促进了检测性能的显著提升。

2014 年，Girshick 等人开创性地将卷积神经网络引入目标检测领域，提出两阶段检测框架——区域卷积神经网络（Region-based Convolutional Neural Networks, R-CNN）[15]。同年，He 等人针对输入图像尺寸固定的限制，提出了空间金字塔池化网络（Spatial Pyramid Pooling Network, SPPNet），实现了对任意尺寸输入图像的处理[16]。2015 年，Girshick 等人针对 R-CNN 检测效率低下的问题，提出了 Fast R-CNN[17]。该框架的核心改进在于将整张图像输入卷积神经网络生成特征图，并将候选区域映射到特征图上进行分类与边界框回归，显著提升了检测速度。2016 年，Girshick 等人进一步优化框架，提出了 Faster R-CNN[18]。该模型摒弃了传统的滑动窗口和选择性搜索方法，引入了区域建议网络（Region Proposal Network, RPN），实现了端到端的检测，大幅减少了计算冗余，进一步提高了检测效率与精度[18]。此后，研究者们基于 Faster R-CNN 框架进行了大量改进与优化，相继提出了 Mask R-CNN[19]、Cascade R-CNN[20]、Dynamic R-CNN[21]以及 Sparse R-CNN[22]等衍生模型。这些改进模型在保持 Faster R-CNN 高效性能的同时，针对不同应用场景与需求进行了针对性优化，进一步拓展了其应用范围与性能边界。

2016 年，Redmon 等人提出了一种革新性的一阶段目标检测算法——YOLO（You Only Look Once）[23]。YOLO 采用了一种全新的检测范式，采用全局图像输入策略，通过一个卷积神经网络实现目标类别识别与边界框回归的联合预测，这种一体化设计使得系统能够并行处理多个目标的检测任务，避免了滑动窗口或候选区域生成等复杂步骤。与 R-CNN 系列算法相比，YOLO 在检测速度与实时性方面具有显著优势，但其检测准确率要略低于 R-CNN 系列算法。Liu 等人提出了单次多框检测器（Single Shot MultiBox Detector, SSD）[24]，该算法通过在不同尺度的特征图上进行检测，实现了对多尺度目标的有效定位与识别。2017 年，Redmon 等人对 YOLO 进行了改进，提出了 YOLOv2[25]。YOLOv2 引入了锚框、多尺度训练以及锚框聚类等优化策略。其中，锚框机制提升了模型对目标位置与尺寸的预测能力，多尺度训练增强了模型的鲁棒性与泛化性能，而锚框聚类则进一步提高了检测精度与效率。2018 年，采用了一种全新的骨干网络——DarkNet53

的 YOLOv3[26]被提出，并引入了特征金字塔网络（Feature Pyramid Network, FPN）[27]，显著提升了模型的多尺度检测能力。2020 年，Alexey Bochkovskiy 团队发布了 YOLOv4[28]，采用跨阶段局部连接网络（Cross Stage Partial Network, CSPNet）[29]改进 DarkNet53，在多尺度融合模块中结合了空间金字塔池化（Spatial Pyramid Pooling, SPP）[30]和路径聚合特征金字塔网络（Path Aggregation FPN, PAFPN）[31]。随后，Ultralytics 公司发布了 YOLOv5[32]，该模型在工业应用中取得了显著进展。YOLOv5 通过降低模型复杂度并提升实时性，实现了更高效的目标检测。近年来，YOLO 系列模型持续演进，相继推出了 YOLOX[33]、YOLOv6[34]、YOLOv7[35]、YOLOv8[36]、YOLOv9[37]以及 YOLOv10[38]等版本。这些改进模型在工业生产监控、智能交通管理、自动驾驶等多个领域取得了显著成果，推动了目标检测技术的不断发展与应用普及。

1.2.2 相关领域研究现状

目前，针对工业环境中防毒面具佩戴检测方法的研究尚处于初步探索阶段，相关文献较为匮乏。因此，本研究旨在填补这一领域的研究空白，为工业安全防护提供技术支持。为实现这一目标，本研究系统梳理了相关领域的研究成果，包括人脸检测技术、工业生产场景中的安全帽检测方法以及公共卫生领域的口罩佩戴检测算法等。这些相关研究不仅揭示了目标检测任务中的共性挑战，如遮挡、光照变化和复杂背景等问题，同时也提供了多样化的解决思路与技术框架，为本研究提供了支持。通过借鉴与创新，本研究致力于构建一种高效、鲁棒的防毒面具佩戴检测方法，以应对工业环境中的实际需求。

在人脸检测研究中，针对复杂环境下小尺度、模糊及部分遮挡人脸的检测问题，Tang 等人提出了一种基于上下文辅助的检测框架 PyramidBox[39]。该框架创新性地引入了上下文锚点机制和一种多层次特征金字塔网络，将高层语义信息与低层面部特征进行有效融合，从而提升了模型在多尺度人脸检测中的性能。Qi 等人则基于 YOLOv5 架构，提出了一种改进的人脸检测模型 YOLOv5-Face[40]。该模型将人脸检测视为通用目标检测任务，并在 YOLOv5 的基础上增加了人脸关键点回归模块，同时采用 Wing loss 对关键点回归进行约束，进一步提升了检测精度。针对现实场景中人脸尺度变化大、姿态多样、遮挡严重以及光照复杂等挑战，Yu 等人提出了 YOLO-FaceV2 模型[41]。该模型在 YOLOv5 的基础上，设

计了感受野增强 RFE 模块，以增强对小尺度人脸的检测能力，并采用 NWD Loss 来缓解 IoU 对微小目标位置偏差的敏感性。此外，作者还引入了空间注意力机制模块和排斥损失函数，以应对人脸遮挡问题。张等人提出一种改进型级联神经网络检测算法来解决人脸密集样本的漏检问题[42]，采用 Soft-NMS 对候选框进行筛选，通过连续函数调整重叠框的置信度评分，有效提升了模型在目标密集场景下的检测性能。

在工业安防领域，基于深度学习的目标检测技术已得到广泛应用。针对施工场景下由于人员密集、目标相互遮挡及环境背景复杂等因素造成小尺度样本出现漏检和误检的问题，陈等人提出了一种改进型 YOLOv8n 检测方法[43]。该方法通过 S-GFPN 结构优化 YOLOv8n 特征融合模块，抑制背景信息的干扰，增强了对复杂环境的适应能力，同时采用 MPDIOU 损失函数来应对待检测目标尺度变化不敏感的问题。为进一步提升安全帽检测的性能，Li 等人针对开源头盔检测数据集 SHWD 的局限性，构建了一个涵盖多种工业场景的综合性安全帽检测数据集[44]，同时，作者重新设计了 YOLO 的正样本选择策略，提出了分层正样本选择机制，以增强 YOLOv5 模型的拟合能力。此外，受视频连续帧目标检测的启发，作者提出了一种基于盒密度的后处理算法，有效减少了误检率。祁等人针对复杂背景和密集小目标场景下的安全帽检测问题，提出了一种改进的 YOLOv5 算法[45]，该算法通过引入串联混合池化模块和增强浅层特征输出，优化了特征提取与目标位置表达能力，同时结合坐标注意力机制扩大感受野，进一步提高了检测精度。

在公共健康防护领域，李等人[46]设计了一种高精度、实时的方法，可以有效地监测公共场所未佩戴口罩的个体，该方法通过在主干特征提取深层网络中添加 CA 注意力模块，使模型能够获取更多的区域信息和通道特征，同时添加 BiFPN 跨阶段多尺度特征融合设计，实现不同尺度目标空间分布特征与高层语义信息的融合，提升了对小尺度目标的识别准确率。

1.3 论文研究内容

从上述的研究现状与发展趋势来看，基于深度学习的目标检测技术在安全防护领域展现出广阔的应用前景。利用深度学习技术构建端到端的防毒面具佩戴检测模型，能够实时、精准地获取工人的防毒面具佩戴状态，这对于提升工人职业

安全意识、预防安全事故发生以及增强企业安全管理效能具有重要的现实意义。目前，工业安防检测正朝着智能化方向快速发展。然而，尽管防毒面具检测是工业安防检测的重要组成部分，但现有研究对此关注不足，相关成果较为稀缺。

基于此，本文在构建防毒面具佩戴检测数据集后选取 YOLOv5、YOLOv8 和 YOLOv10 作为研究对象，是因为其技术演进特征与差异化优势具备研究价值：YOLOv5 作为 YOLO 系列中应用最广泛的版本，凭借其高度优化的工程架构和丰富的文档资源，在开源社区中形成了完善的生态体系，成为工业界快速落地的首选方案。YOLOv8 其无锚框设计摒弃了传统基于预定义锚框的检测机制，转而采用更简洁的中心点预测方法，配合动态标签分配策略，有效降低了模型复杂度，在多个数据集上展现出优于前代模型的精度-速度平衡特性。YOLOv10 则通过双分支一致性匹配策略实现了 NMS-Free 的端到端检测框架，彻底消除非极大值抑制后处理环节带来的计算冗余，相较于 YOLOv8 有更低的推理延迟，为实时高密度目标检测任务提供了新的技术路径。鉴于上述技术演进特征，本研究选取 YOLOv5、YOLOv8 和 YOLOv10 作为研究对象进行系统性改进与对比实验。主要工作如下：

（1）防毒面具佩戴检测数据集构建

数据驱动的深度学习的模型性能与训练数据的质量与规模存在很大关系，只有通过高质量的数据才能训练出具备良好泛化能力的模型。因此，在研究基于深度学习的防毒面具佩戴检测算法时，构建一个场景多样、数据充足的高质量数据集是至关重要的。为实现这一目标，本研究首先通过实地调研工厂环境，采集实际场景下的视频素材，并对其进行抽帧和数据清洗等预处理操作，以获取原始数据集。随后，结合任务特点选择离线数据增强方法，对数据集进行扩充与优化，以进一步提升数据的多样性与模型的鲁棒性。

（2）针对复杂背景干扰和样本不均衡的改进方法

首先，在 YOLOv5 原有的 P3、P4、P5 基础上，引入了一个小目标检测头 SODH，辅助进行目标检测任务，然后在 Neck 层添加了 CBAM 模块，通过有效区分目标与背景信息，提升检测精度，最后选择 VariFocal Loss 作为损失函数替换原算法的二元交叉熵损失函数，使算法更侧重于处理高质量的正样本，提升了对小尺度的检测准确率，最后设计对照实验对改进方案的有效性进行了系统性评估。

（3）针对特征层信息融合差和多尺度检测的改进方法

本文将大核分离卷积注意力模块添加到空间金字塔池化融合模块中，提升不同特征层间的语义融合；然后将主干的 P2 层的输出集成到颈部，增加了一个小目标检测头，辅助进行目标检测任务；最后在头部添加了高效多尺度注意力模块，使网络更高效地学习待检测目标语义信息，最后对改进后的算法进行了实验对比分析。

（4）针对小目标检测困难和特征信息提取差的改进方法

本文设计小目标语义信息增强金字塔模块进行特征融合，以增强模型对防毒面具佩戴的检测效果。其次，为了提高模型的检测速度和特征提取能力，在 RepVGG 中引入了 RepConv ShuffleNet 和 RCS 的 OneShot 聚合替换模型颈部浅层网络中的 C2f 模块。最后，采用 EMA-slide Loss 替换原模型损失函数，以增强类别分类能力，提高训练的稳定性，最后对改进后的算法进行了实验对比分析。

1.4 论文结构安排

本文具体组织架构如下：

第 1 章：绪论。首先介绍工业场景下防毒面具佩戴检测的研究背景与意义。之后介绍了国内外相关技术的发展与研究。最后介绍了本文的主要研究内容。

第 2 章：相关技术与理论基础。本章围绕防毒面具佩戴检测任务，系统性地梳理了相关技术体系与理论基础。首先阐述了深度学习的基本原理与核心特征，随后介绍了卷积神经网络的发展脉络，并详细解析了其核心组件。在目标检测方法部分，对比分析了两阶段与单阶段检测框架的算法原理及性能特点，并详细讨论了检测任务中常用的损失函数设计。最后，针对样本分布不均衡这一普遍性问题，介绍了当前主流的处理方案。

第 3 章：本章节系统阐述了防毒面具佩戴检测数据集的信息及实验环境。首先，对防毒面具的基本属性进行了概述，涵盖其核心功能、结构组成以及主要分类，为检测任务提供了必要的背景知识。其次，详细描述了数据集的统计特征进行了全面分析。

第 4 章：针对复杂背景干扰和样本不均衡的改进方法。本章主要针对模型检测过程中的复杂背景干扰和样本不均衡问题，提出基于 YOLOv5 的改进方法。首先介绍了改进 YOLOv5 模型的总体结构，包括小目标检测头、CBAM 模块结

构和 VariFocal Loss 原理。最后对各项实验的结果进行分析和讨论，并通过可视化检测效果对比分析模型存在的问题。

第 5 章：针对特征层信息融合差和多尺度检测的改进方法。本章主要研究了在实际工业应用场景中的检测模型特征层信息融合差和多尺度目标检测问题，并提出基于 YOLOv8 的改进方法。首先介绍了改进 YOLOv8 模型的总体结构，包括 LSKA 模块结构、高效多尺度注意力模块结构和小目标检测头。最后对各项实验的结果进行分析和讨论，并通过可视化检测效果对比分析模型存在的问题。

第 6 章：针对小目标检测困难和特征信息提取差的改进方法。本章主要研究了针对小目标检测困难、特征信息提取能力差以及样本不均衡的问题，提出相应的改进方法。首先介绍了改进 YOLOv10 模型的总体结构，包括 STSIEP 模块结构、RCSOSA 模块结构和 EMA-Slide Loss 原理。最后对各项实验的结果进行分析和讨论，并通过可视化检测效果对比分析模型存在的问题。

第 7 章：总结与展望。本章首先对本研究所做的具体工作进行了总结和分析。然后提出对不足之处的改进方向，包括数据集的完善、算法的优化以及实时性处理等。

第 2 章 防毒面具佩戴检测相关理论基础

2.1 深度学习相关理论

深度学习（Deep Learning, DL）以多层人工神经网络结构为核心，能够从海量数据中自动提取特征并执行复杂的模式识别任务。相较于传统机器学习依赖人工特征设计的方法，深度学习通过深度网络的多层结构直接从原始数据中学习特征信息，建立端到端的学习模式。这种数据驱动的特征构建方式不仅突破了传统方法的表征瓶颈，更通过多层次的特征组合优化，实现了模型建模能力与泛化能力的协同提升。与传统机器学习方法不同，深度学习不再依赖人工设计特征，而是通过多层神经网络结构直接从原始数据中学习潜在的规律，完成端到端的建模过程。其核心思想是通过多层神经元和非线性激活函数，逐步提取和抽象数据中的高阶特征，从而显著增强模型的表达能力和泛化能力。

深度学习以神经网络为基本架构，其结构通常由输入层、多个隐藏层以及输出层构成。每一层的神经元通过可调节的权重与前一层相连接，隐藏层的输出作为下一层的输入，形成一种层次化的特征提取机制，这种层级化设计使得模型能够自动从数据中学习到不同特征层的信息。为了优化参数，深度学习普遍采用反向传播算法。该算法该算法基于链式求导法则，计算目标函数相对于各层网络参数的偏导数确定参数的更新方向，并采用随机梯度下降等优化方法对网络权重矩阵和偏置向量进行迭代更新，以逐步降低损失函数的值。通过这一过程的反复执行，模型能够不断优化其内部参数，从而提升对复杂数据模式的识别能力。

深度学习的模型架构根据任务类型的不同而有所区分。例如，卷积神经网络因其在局部特征提取方面的优势，广泛应用于图像处理任务；循环神经网络则因其对序列数据的处理能力的特点，常用于时间序列分析任务；而 Transformer[48] 模型凭借其自注意力机制，突破了传统神经网络在并行计算与长程依赖捕获方面的局限性，在自然语言处理领域展现了显著优势。这些模型在计算机视觉、自然语言处理、语音识别以及推荐系统等领域展现了卓越的性能。随着大规模数据集的不断积累和计算资源的持续提升，深度学习的应用范围也在不断扩展和优化。其在各个领域的成功应用，不仅推动了技术的进步，也为解决实际问题提供了新的思路和方法。

2.2 卷积神经网络

作为深度学习体系中的核心架构之一，卷积神经网络在计算机视觉领域发挥着关键作用。受生物视觉感知机制的启发，卷积神经网络采用分层特征学习策略，通过局部感受野和权值共享机制，实现了对图像空间特征的自动学习。其核心思想是通过卷积操作和池化操作进行特征提取，并通过全连接层进行分类或预测。图 2-1 展示了一个用于图像分类的 CNN 架构。

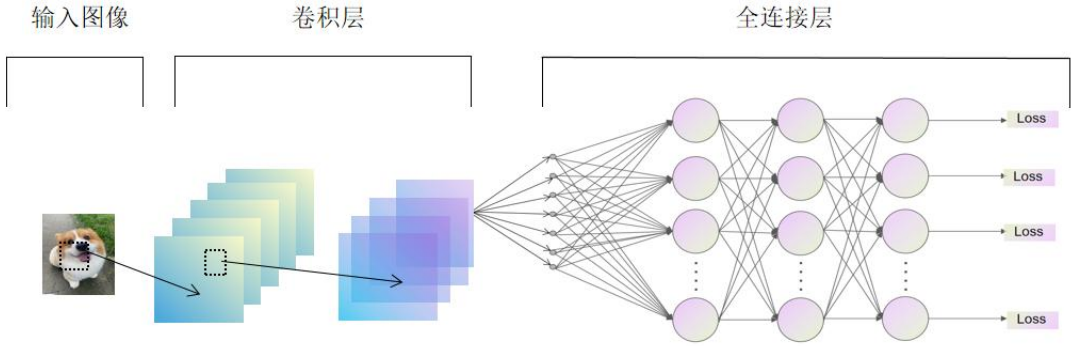


图 2-1 CNN 架构

Fig.2-1 CNN architecture

（1）卷积层

卷积层作为卷积神经网络的核心组件，其主要功能是通过卷积操作从输入数据中提取局部特征，并逐步学习到更高层次的抽象特征表示。卷积操作通过使用可训练的卷积核在输入数据上进行滑动，并与局部区域进行加权求和，生成特征图[49]。卷积核的尺寸、步长以及填充方式共同决定了特征图的输出维度。多个卷积核可以并行工作，每个卷积核能够学习到不同的特征模式，例如边缘、纹理或颜色梯度等。通过多层卷积层的堆叠，网络能够从低级的局部特征逐步提取到更高级的语义特征，从而实现对复杂数据的有效建模。此外，卷积层还具有多项显著优势：首先，权重共享机制显著减少了模型的参数量，提升了计算效率；其次，局部感受野的设计使得卷积核仅关注输入数据的局部区域，从而实现对局部特征的精确提取；最后，卷积层具备一定的平移不变性，即当输入数据中的目标发生位置变化时，网络仍能保持稳定的识别性能。因为这种强大的特征学习能力，卷积结构不仅在计算机视觉领域得到广泛的应用，还在时序信号处理、语音识别、自然语言处理等多个方向展现出卓越的性能，成为现代深度学习框架中的基础构建模块。图 2-2 展示了一个基本的卷积过程。

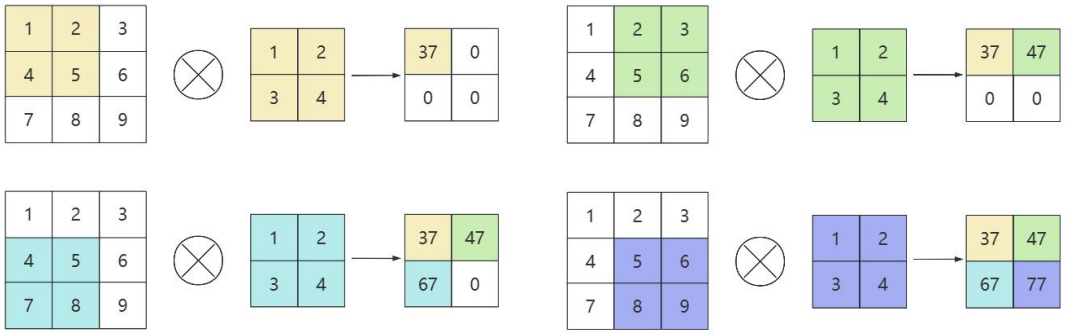


图 2-2 卷积计算示意图

Fig.2-2 The diagram of convolution calculation

（2）激活函数

激活函数的主要作用是引入非线性变换，能够帮助网络捕捉到输入数据中的复杂模式和特征，提升模型的表达能力和学习能力[50]。没有激活函数，神经网络将只能进行线性变换，无法处理复杂的任务。常见的激活函数有 Sigmoid[51]、Tanh[52]、Mish[53]、ReLU[54]、Leaky_ReLU[55]和 SiLU[56]，如图 2-3 所示。

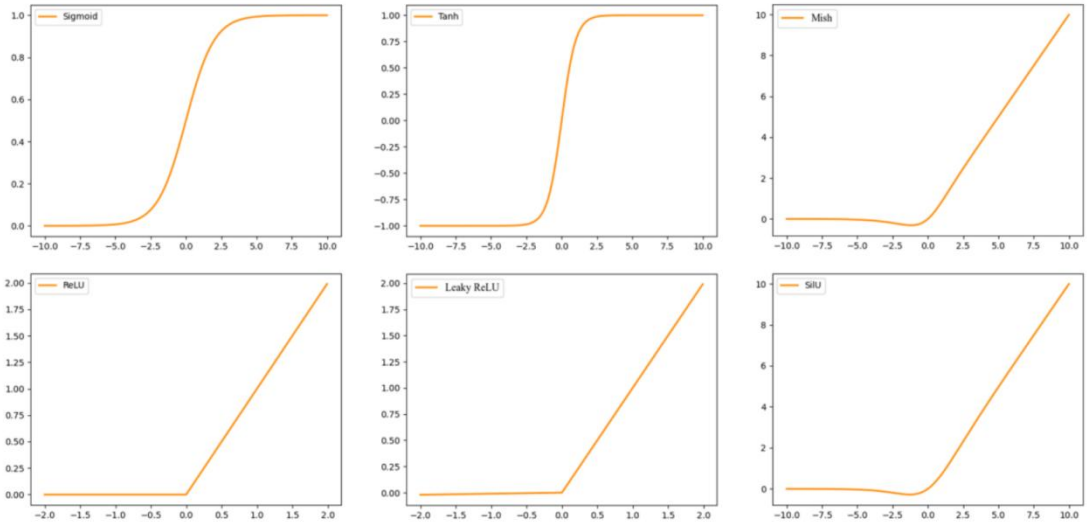


图 2-3 常见激活函数图像

Fig.2-3 The graphs of common activation functions

Sigmoid 激活函数图像呈 S 形，这意味着其在接近 0 时变化较快，而在输入值远离 0 时接近平稳。Sigmoid 函数具有将任意实数输入映射到（0,1）区间的特性，常用于神经网络的隐藏层或输出层。其公式为：

$$\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{2-1}$$

Tanh 函数的图像呈 S 形曲线，与 Sigmoid 函数类似，但它的值在输入趋向正无穷和负无穷时分别接近 1 和-1，常用于 RNN 和 LSTM 等序列建模任务的激活函数。

$$\mathrm{Tanh}(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2-2}$$

Mish 函数的图像是一种平滑、无间断的曲线，正值可以达到任何高度，能够有效避免因值限制引起的梯度饱和问题。但是由于相较于 Sigmoid、ReLU 等计算复杂度较高，因此应用并不广泛。

$$\mathrm{Mish}(x)=x\times\mathrm{Tanh}\left(\ln\left(1+e^{x}\right)\right) \tag{2-3}$$

ReLU 函数被广泛应用于神经网络的每一层，尤其是在卷积神经网络和全连接神经网络中。ReLU 的优点在于计算简单、收敛速度快，并且通过保持正区间的线性特性有效缓解了反向传播中的梯度衰减现象。

$$\mathrm{ReLU}(x)=\max(0,x) \tag{2-4}$$

Leaky_ReLU 是 ReLU 的一种变种。它通过对负输入值引入一个非常小的斜率来解决 ReLU 中可能出现的梯度消失问题。当输入值为负数时，Leaky_ReLU 不会将输出截断为 0，而是会乘以一个非常小的常数即 negativa_slope。

$$\text{Leaky\_ReLU}\ (x)=\begin{cases}x & x\geq 0\\ \text{negativa\_slope}*x & x<0\end{cases} \tag{2-5}$$

SiLU 函数是一种结合了 Sigmoid 函数和线性函数特性的激活函数，在负值区域不会完全丢失信息，而是平滑地接近于 0，避免了训练过程中的过度梯度消失。

$$\mathrm{SiLU}(x)=x*\mathrm{Sigmoid}(x) \tag{2-6}$$

（3）池化层

池化层作为特征降维的关键组件，通过下采样操作降低特征图的空间维度，在保持关键特征的同时实现计算效率的提升。该层通常部署在卷积层之后，通过局部区域提取特征信息。池化层的工作原理是对输入的特征图进行子区域的降维操作，通常包括最大池化和平均池化。图 2-4 展示了最大池化和平均池化计算过程。最大池化最常用的一种池化方法，在每个小区域中选择最大值作为该区域的输出，能够有效保留边缘和纹理等重要信息。平均池化是在每个小区域内计算该

区域所有值的平均值，相比最大池化更平滑，能够保留更平稳的信息。

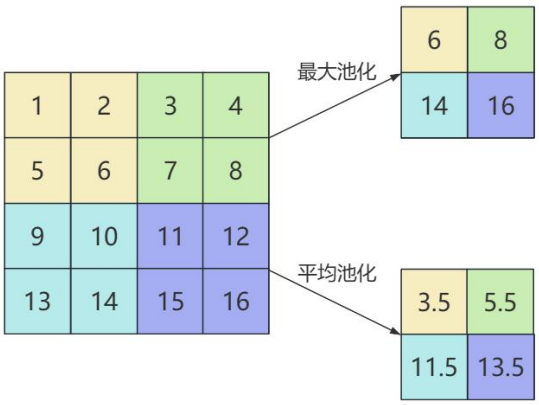


图 2-4 最大池化和平均池化示意图

Fig.2-4 Schematic diagram of max pooling and average pooling

（4）全连接层

全连接层是神经网络中一种常见的层类型，通常用于卷积神经网络中的最后部分，卷积层提取了图像的局部特征，池化层进一步降维，而全连接层则将这些局部特征融合成一个全局的特征表示，用于最终的分类或回归任务。结构如图 2-5 所示，全连接层中的每个神经元与前一层的所有神经元相连，每个输入节点都与输出层的每个神经元都有权重连接[57]，从而使得模型可以学习到数据中的复杂关系和模式。

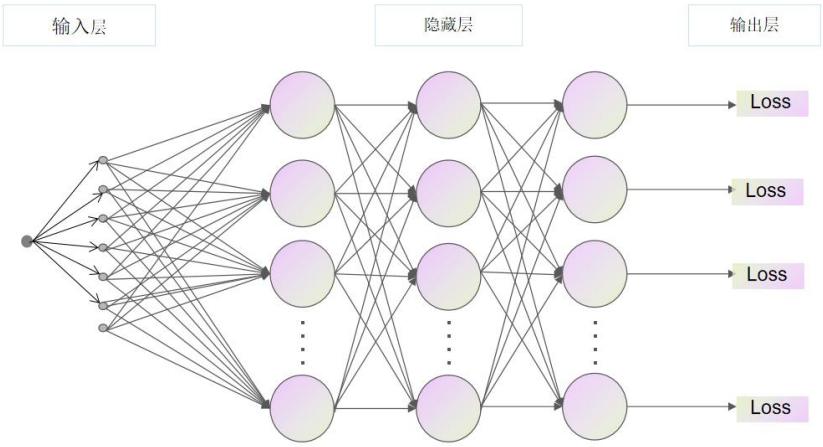


图 2-5 全连接层结构图

Fig.2-5 Diagram of fully connected layer structure

（5）损失函数

在训练卷积神经网络时，损失函数主要用于度量模型的预测与真实标签之间的差距。通过最小化损失函数，模型学习到最优的参数，使得其预测结果更接近真实标签[58]。通过任务类型进行分类可以分为分类任务中的交叉熵损失和回归任务中的均方误差损失函数等。

交叉熵损失函数作为监督学习中的核心评价指标，通过最小化预测概率分布与真实分布之间的差异来驱动模型输出逼近真实标签分布，使其成为分类任务中最常用的优化目标函数之一，其公式如公式（2-7）所示。

$$H(p,y) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(p_i) \tag{2-7}$$

其中，n表示样本数量，y_i表示第i个样本的真实标签，p_i表示模型对第i个样本的预测概率。交叉熵损失函数的优点是可以处理多分类问题，并且可以有效地避免模型出现过拟合现象。

在回归任务中，均方误差作为常用的评估指标，其核心机制是通过计算预测输出与实际结果之间的平方差均值，持续优化模型参数以缩小预测偏差。

$$H(p,y) = \frac{1}{n}\sum_{i=1}^{n} (p_i - y_i)^2 \tag{2-8}$$

其中，n表示样本数量，y_i表示第i个样本的真实标签，p_i表示模型对第i个样本的预测概率。均方误差损失的优势在于是一个平滑的函数，有效避免了梯度消失问题。

2.3 目标检测方法

2.3.1 双阶段目标检测方法

两阶段目标检测模型是目标检测领域中最常用的模型之一，通常分为两个阶段：区域提议阶段和目标分类与回归阶段。在一阶段，算法首先需要从输入图像中生成一组大致覆盖了图像中可能存在物体的位置的候选区域。在二阶段，算法对这些区域进行精确的分类和回归。

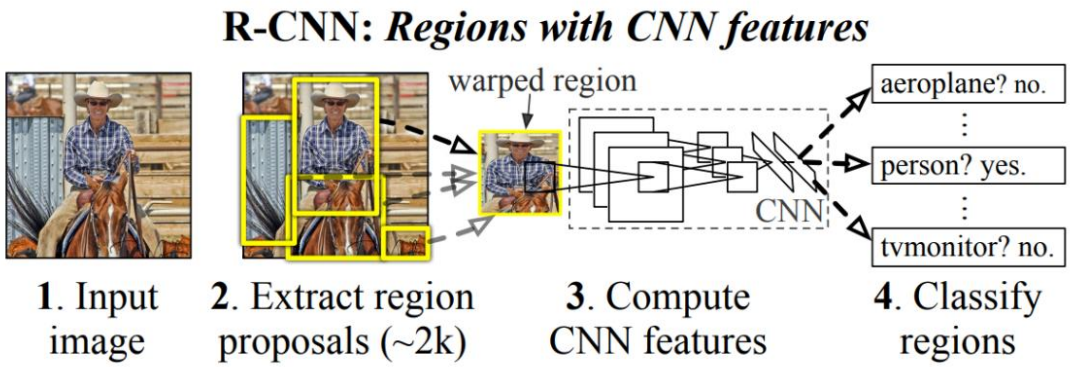


图 2-6 R-CNN 网络结构图[15]

Fig.2-6 Diagram of R-CNN network architecture[15]

R-CNN（Region-based Convolutional Networks，R-CNN）是目标检测领域两阶段模型的开山之作，其首次在目标检测模型中引入 CNN，使得目标检测精度有了显著提升，是深度学习在计算机视觉中应用的一个重要里程碑。R-CNN 的算法原理如图2-6所示，首先通过选择性搜索策略生成可能包含目标的区域建议；然后借助预训练的深度卷积网络对每个候选区域进行特征编码，获得具有判别性的特征表示；最后构建专门的分类器网络，对提取的区域特征进行类别判定与边界框回归。

由于 R-CNN 需要对每个候选区域单独进行卷积特征提取，导致计算量极大，因此后续有了一系列改进，如 Fast R-CNN，Faster R-CNN 等等。但由于两阶段目标检测算法需要分为区域提议阶段和目标分类与回归阶段两个阶段，导致计算开销大，整体推理速度慢，不适用于实时检测。

2.3.2 单阶段目标检测方法

单阶段检测模型通过端到端的设计实现了效率突破。与两阶段方法需依赖区域建议网络预生成候选框不同，该模型直接在卷积神经网络的特征图上进行空间网格划分——每个网格单元不仅负责回归预设锚框的坐标偏移量，还同步计算多类别概率分布。相较于传统的两阶段算法，由于一阶段检测算法省去了候选区域生成和后续的区域分类步骤，因此具有显著的速度优势，能够实现实时目标检测，如较多应用于自动驾驶、安防等领域。

在单阶段检测器发展历程中，YOLO 系列算法开创了全新的检测模式。该算法创新性地将目标定位与识别任务统一为端到端的回归问题，通过网格化策略实现了检测效率的突破性提升。如图 2-7 所示，YOLOv1 网络首先将输入图像划分为 s×s 个网格，然后每个网格单元预测 B 个边界框，并预测 C 个类别的概率，表示该网格单元中包含目标的类别。这种一体化设计摒弃了传统的区域建议机制，显著提高了检测速度，为实时目标检测提供了可行的解决方案。

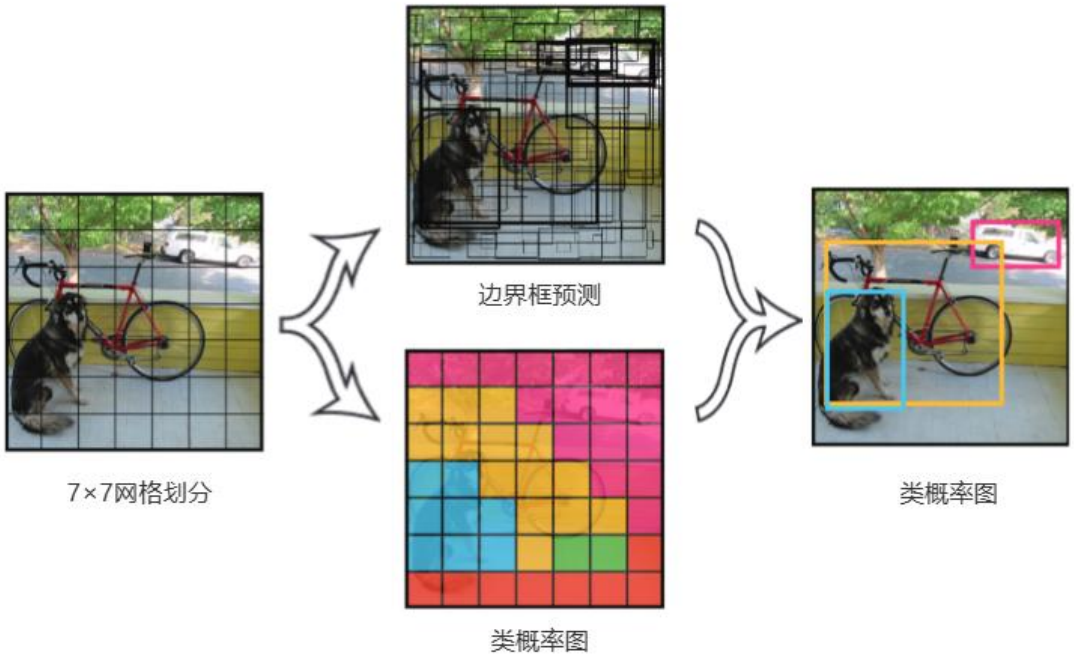


图 2-7　YOLOv1 算法原理图[23]

Fig.2-7 YOLOv1 algorithm principle diagram[23]

但是由于 YOLOv1 是将输入图像分割成一个固定的网格，导致没有足够的分辨率对小目标进行精确检测，因此对小物体的检测精度较低，并且还存着这明显的类别不均衡问题等。而后续版本通过采用先进的模块化策略继续创新，从而做到更轻量，检测精度更高，综合性能更好。

YOLOv5 是 Ultralytics 公司提出的 YOLO 系列的第五个版本，拥有更简单的网络结构和更强的检测效果。其结构图如图 2-8 所示，主要改进了 CBS 模块和 SPPF 模块，CBS 模块主要由 Conv、Batch Normalization 和 SiLU 激活函数构成，帮助网络在各个阶段提取和处理图像特征，而 SPPF 模块是对之前版本的 SPP 模快的改进，主要是用来增强网络对多尺度特征的提取能力，使得模型可以更好地检测不同大小的物体。

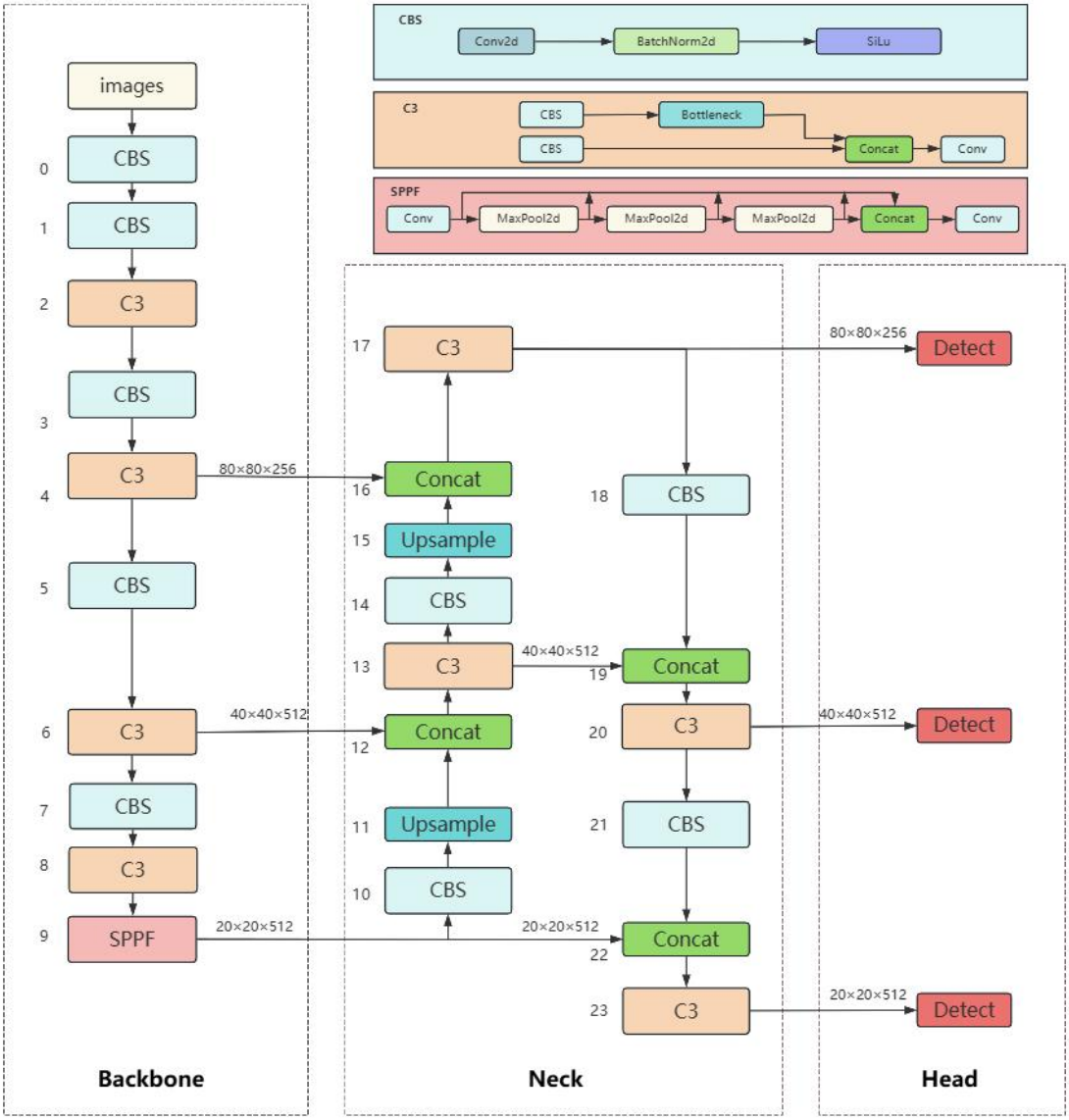


图 2-8 YOLOv5 模型结构图

Fig.2-8 The architectural diagram of the YOLOv5 model.

YOLOv7 是由开发 YOLOv4 的 Wang, Chien-Yao 等人提出的，是在 5 FPS 到 160 FPS 范围内的速度和准确度最先进的目标检测算法之一。其结构如图 2-9 所示，相较于 YOLOv5 主要提出了 ELAN 和模型重参数化的创新。ELAN 作为 backbone 的一部分，与 MP 结构结合使用，显著提升了网络的学习能力和效率，而模型重参数化在不增加计算复杂度的情况下，通过重新组织网络结构和优化参数来提升模型的表达能力。二者为 YOLOv7 在目标检测任务中的出色表现奠定了坚实的基础。

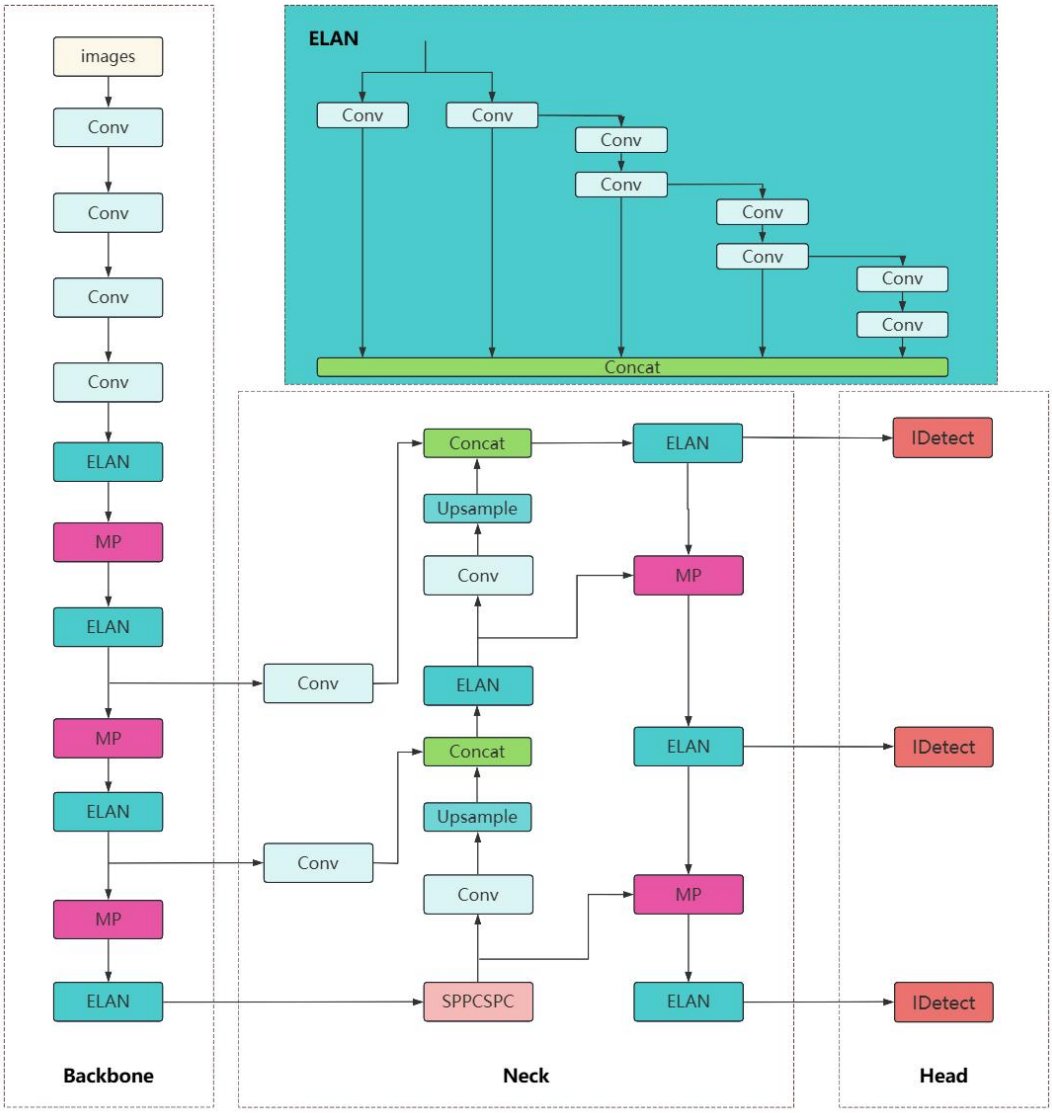


图 2-9 YOLOv7 结构图

Fig.2-9 The architectural diagram of the YOLOv7 model.

YOLOv8 是 Ultralytics 公司在 YOLOv5 基础上进一步优化和改进的目标检测算法，相较于 YOLOv5 有较大创新。YOLOv8 参考 C3 模块的残差结构以及 YOLOv7 的 ELAN 思想提出 C2f 模块，将耦合头替换为分别提取类别特征和位置特征解耦器检测头并优化损失函数，同时采用 Anchor-Free 思想，显著提高了模型对目标检测的准确性、效率和泛化能力，其结构如图 2-10 所示。

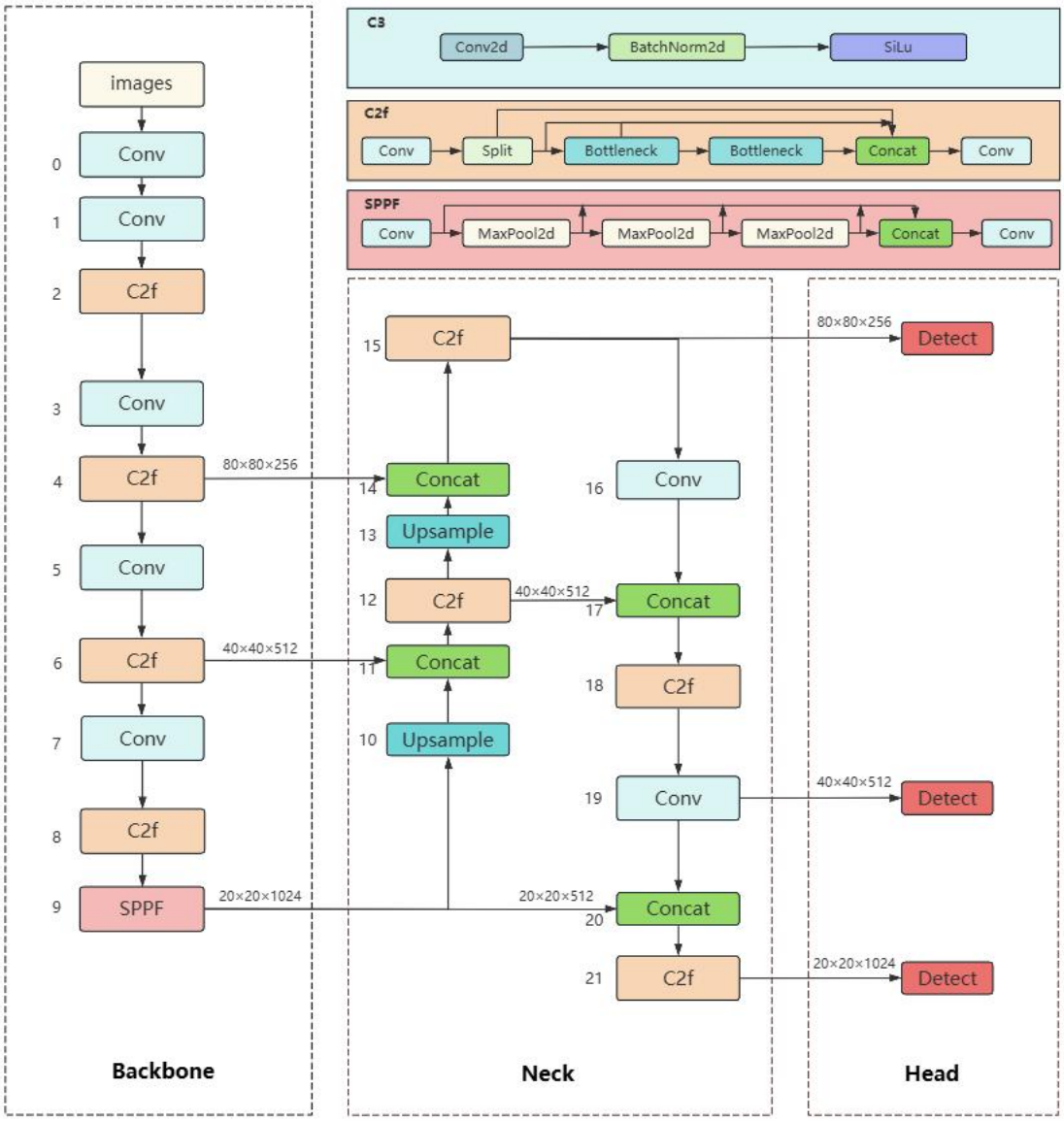


图 2-10 YOLOv8 结构图

Fig.2-10 The architectural diagram of the YOLOv8 model.

YOLOv10 由清华大学的研究团队于 2024 年提出。其基本框架源自 YOLOv8，但创新性提出了 PSA 模块和支持无 NMS 训练。其结构如图 2-11 所示，PSA 模块结合了通道和空间注意力，被巧妙地融入模型架构中，通过极化过滤保持高分辨率，提升了模型学习全局表示的能力。同时，模型在训练期间利用一对多分配的丰富监督信号，而在推理期间则使用一对一分配的预测结果，从而实现无 NMS 的高效推理，YOLOv10 由于 NMS-free 方法在后处理速度上优于 yolov8，更适合 real-time 检测。

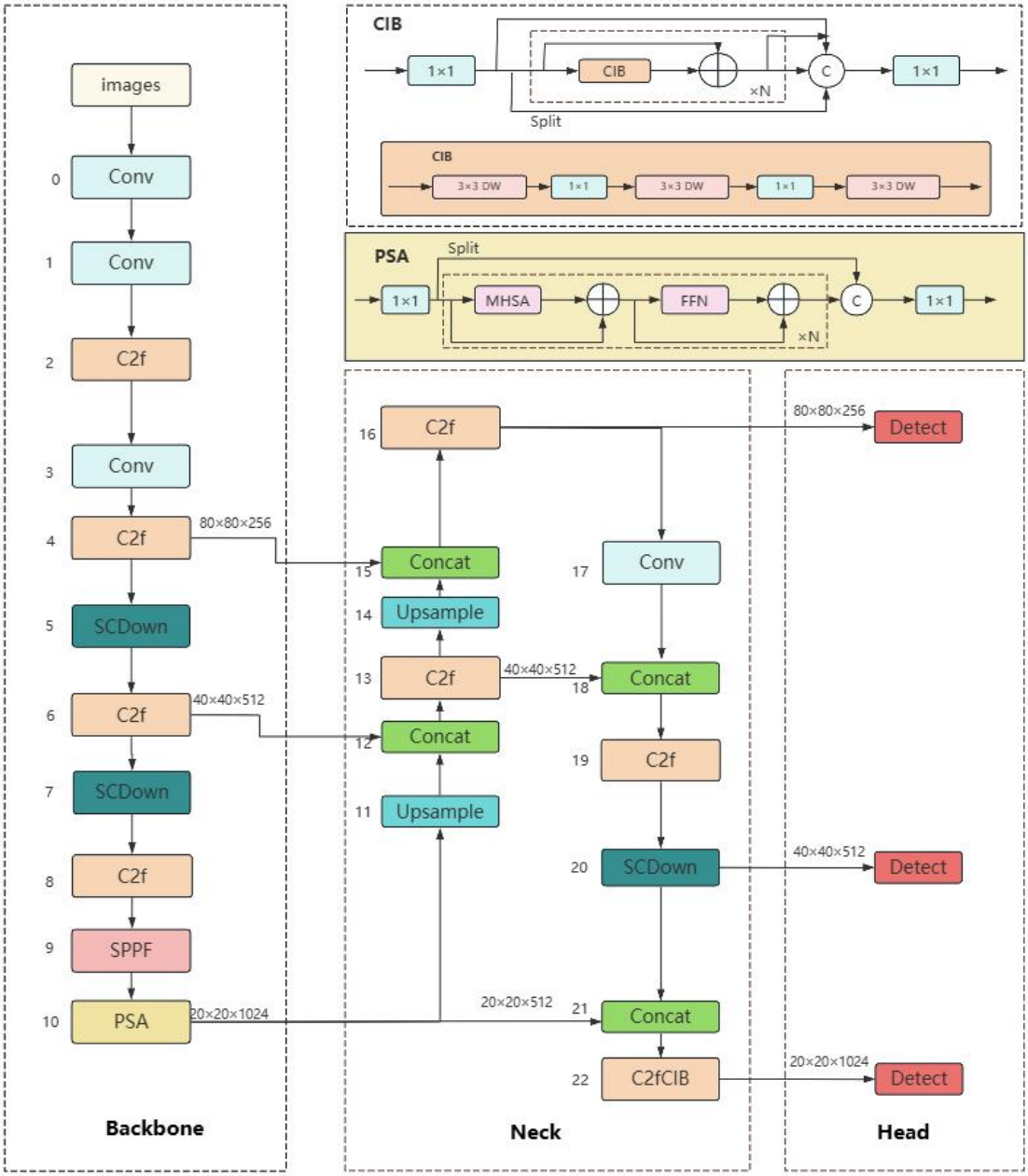


图 2-11 YOLOv10 结构图

Fig2-11 The architectural diagram of the YOLOv10 model.

表 2-1 是 YOLO 系列各个版本的对比。本研究选取 YOLOv5、YOLOv8、YOLOv10 进行系统性对比，因其技术框架具有以下优点：YOLOv5 凭借工程架构优化与完善生态成为工业工业界快速落地的首选方案；YOLOv8 采用无锚框设计和动态标签分配策略，实现精度与速度的平衡，有效降低了模型复杂度；YOLOv10 首创双分支一致性匹配策略，构建 NMS-Free 的端到端检测框架，相

较于 YOLOv8 有更低的推理延迟。三版本技术迭代路径清晰，具备典型研究价值。

表 2-1 YOLO 系列对比表

Table 2-1 YOLO series comparison

模型	提出时间	主要改进结构	特点
YOLOv1	2015 年 6 月	GoogLeNet	首个实时检测模型
YOLOv2	2016 年 12 月	Darknet-19	引入 Anchor Boxes
YOLOv3	2018 年 4 月	Darknet-53	引入残差网络
YOLOv4	2020 年 4 月	CSPDarknet-53	改进特征融合和路径聚合
YOLOv5	2020 年 6 月	C3、SPPF	网络结构优化易于部署
YOLOv6	2022 年 6 月	EfficientRep	专注效率支持多平台部署
YOLOv7	2022 年 7 月	E-ELAN、MP	模型重参化、优化标签分配
YOLOv8	2023 年 1 月	C2f、Anchor-Free	支持分类、分割、检测等任务
YOLOv9	2024 年 1 月	PGI、GELAN	弥补前馈过程信息损失
YOLOv10	2024 年 5 月	PSA	支持无 NMS 训练

2.3.3 目标检测损失函数

在目标检测中，除了上文介绍的深度学习中常用的交叉熵损失函数和均方误差损失函数之外，还有一类边界框损失函数，用于衡量预测边界框（Predicted Bounding Box，P）与真实边界框（Ground Truth Bounding Box，GT）之间的差异，通过优化边界框损失函数能够直接提高目标检测模型的定位精度[59]。目前主流目标检测方法普遍采用交并比 (Intersection over Union，IoU)损失函数来度量预测框和真实框之间的重合区域，从而引导模型训练。下面介绍一些常见的 IoU 损失函数。

（1）IoU 损失函数

IoU 计算过程如图 2-12 所示，其中，“GT”代表真实框，“P”代表预测框，IoU 计算的是预测边界框与真实边界框的交集与并集之比。

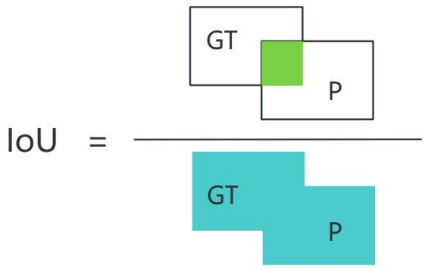


图 2-12 IoU 计算过程图

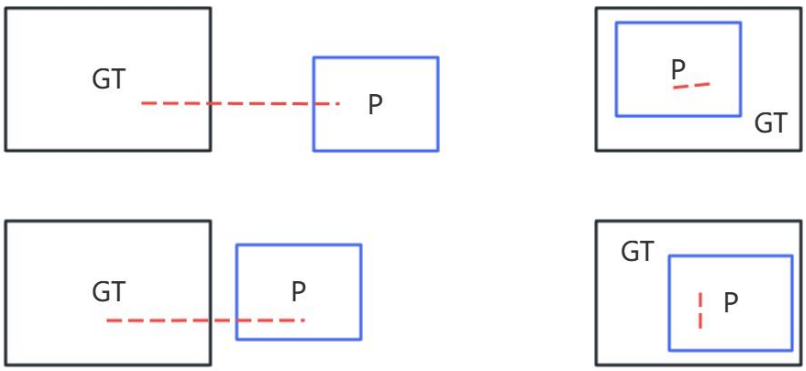
Fig.2-12 Diagram of the IoU calculation process

IoU 损失函数[60]的定义如公式 2-9 所示。

$$L_{\mathrm{IoU}} = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{2-9}$$

但是，该计算公式存在两个问题，如图 2-13 所示。第一、非交集情况：当预测边界框与真实边界框没有交集时，IoU 值恒为零，且定位损失恒为 1。在这种情况下，IoU 无法提供有效的梯度信息，也无法反映预测框与真实框之间的空间距离。这种缺乏区分能力的特点阻碍了模型训练中的优化过程。

第二、固定尺寸包含情况：当真实框完全包含预测框且两者的尺寸固定时，IoU 值仅由预测框与真实框的面积比决定。因此，无论预测框在真实框内的具体位置如何，定位损失都保持不变。这一局限性削弱了 IoU 在指导模型实现精确定位方面的能力，因为它未能考虑边界框的空间对齐情况。



(a)问题一　　(b)问题二

图 2-13 IoU 存在的两个问题

Fig.2-13 Two issues with IoU

这些不足表明，需要设计更复杂的损失函数来解决 IoU 的局限性，从而提高目标检测任务中边界框定位的精度。

针对第一个问题，Rezatofighi 等人提出 GIoU[61]，GIoU 在传统 IoU 的基础上进行了扩展，引入了最小包含框（Minimum Bounding Box, MBB）的概念，该矩形是能够同时包含真实框和预测边界框的最小矩形区域。这一改进解决了在预测框与真实框无交集情况下 IoU 的局限性，从而为边界框回归提供了更鲁棒且信息更丰富的度量方法。如图 2-14 所示。

图 2-14 GIoU 示例图

Fig.2-14 Diagram of GIoU

其中，绿色框代表最小包含框，是能够同时包含真实框（以黑色表示）和预测框（以蓝色表示）的最小矩形区域。通过从绿色框的面积中减去黑色框和蓝色框的联合面积，并将该差值相对于绿色框的面积进行归一化，可以量化真实框与预测框之间的空间差异。

公式（2-10）展示了 GIoU 损失的计算方式，其中A_c代表图 2-14 中最小包含框面积，即同时包含了预测框和真实框的最小框的面积。真实框与预测框之间最小横轴距离为x_1，最大横轴距离为x_2；最小纵轴距离y_1，最大纵轴距离为y_2。U代表真实框与预测框的并集[62]。通过引入A_c，GIoU 不仅关注重叠区域，还关注其他的非重合区域，能够准确地反映预测框和真实框的重叠方式。

$$L_{\mathrm{GIoU}} = 1 - \mathrm{IoU} + \frac{A_c - U}{A_c} \tag{2-10}$$

$$A_c = (x_2 - x_1)(y_2 - y_1) \tag{2-11}$$

尽管 GIoU 有效解决了传统 IoU 在预测框与真实框无重叠情况下失效的问题，但并未解决第二个问题。2020 年天津大学研究团队提出的 DIoU 技术提供相应的解决方案[63]。DIoU 在 IoU 的基础上引入了中心点距离的惩罚项，通过最小化预

测框与真实框中心点之间的欧氏距离，进一步优化了边界框的回归精度，如图2-15所示。

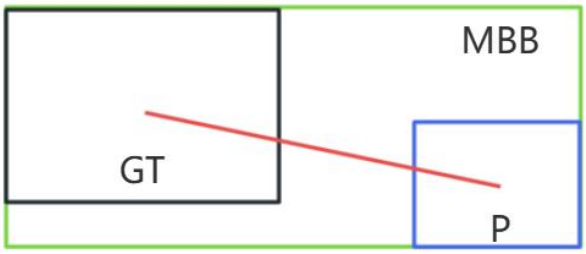


图2-15 DIoU示例图

Fig.2-15 Diagram of DIoU

DIoU的计算方法如公式（2-12）所示，其中b和b^{gt}分别代表预测框和真实框的中心点，ρ表示两个中心点之间的欧式距离。此外，c代表能够同时包含预测框和真实框的最小包含框的对角线距离。

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{2-12}$$

由于DIoU在计算中未考虑边界框回归中的长宽比，因此天津大学研究团队在DIoU的基础上进行改进提出了CIoU[63]，这是目前YOLO系列正在使用的IoU损失函数。其惩罚项如下：

$$R_{\mathrm{CIoU}} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2-13}$$

其中α是权重函数，而v用来度量长宽比的相似性，定义如下：

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{2-14}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{2-15}$$

完整的 CIoU 损失函数的定义如公式所示：

$$L_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2-16}$$

2.3.4 评价指标

在目标检测任务中，模型的输出通常包括目标类别的置信度分数以及检测目标的边界框坐标信息。为了评估模型的性能，通常需要设置两个关键阈值：置信

度阈值和 IoU 阈值。首先，置信度分数低于预设阈值的预测框将被过滤掉，以确保检测结果的可靠性。其次，通过计算预测框与真实框之间的 IoU 值，可以进一步衡量检测的准确性，并基于此计算精确率（Precision，P）、召回率（Recall，R）以及平均精度均值（mean Average Precision, mAP）等核心评价指标。此外，还有参数量（Params）、计算量（GFLOPs）、帧率（FPS）这些指标共同构成了目标检测模型性能评估的综合体系。

图 2-16 为精确率与召回率示意图。True Positives（TP）是正样本被正确分配的数量；True Negatives （TN）是负样本被正确分配的数量；False Positives（FP）是被错误分配为正样本的数量；False Negatives（FN）是被错误分配为负样本的数量。

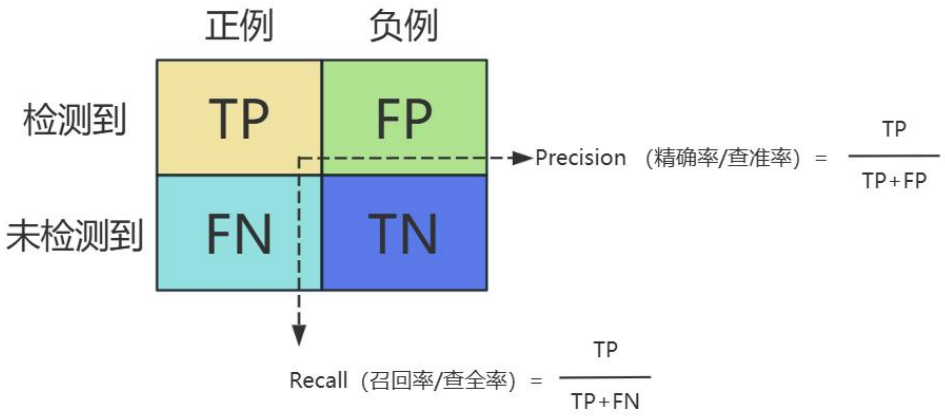


图 2-16 精确率与召回率示意图

Fig.2-16 Diagram of Precision and Recall

精确率和召回率分别定义在公式（2-17）和公式（2-18）中。精确率P是指被正确分配的正样本数占总分配的正样本。召回率R是指被正确分类的正样本数与总正样本数的比例。

$$P = \frac{TP}{(TP + FP)} \tag{2-17}$$

$$R = \frac{TP}{(TP + FN)} \tag{2-18}$$

AP是特定类别的准确率-召回率曲线下的区域，而mAP表示每个类别的AP的平均值。对于多类别任务，mAP可以使用公式（2-20）计算，其中AP_i表示给定类别的平均精度。

$$AP_i = \int_0^1 \mathrm{P}(\mathrm{R})d(\mathrm{R}) \tag{2-19}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{2-20}$$

F_1评价模型分类能力的综合评价指标，其定义如公式（2-21）所示。本文使用F_1指标来评估模型对数据集中每个类别的分类效果。

$$F_1 = 2\frac{P * R}{P + R} \tag{2-21}$$

2.4 类别不均衡

在目标检测任务中，存在多种形式的不均衡问题，这些不均衡现象会显著影响模型的检测效果。Oksuz 等人对此进行了系统性归纳[64]，将目标检测中的不均衡问题划分为以下四类：

（1）类别不均衡（Class Imbalance）：主要由不同类别样本数量分布不均引起，其中最典型的是前景目标与背景目标之间的数量差异。在训练过程中，背景样本的数量通常远多于前景样本，导致模型对背景的过拟合，从而降低了对前景目标的检测能力。

（2）尺度不均衡（Scale Imbalance）：这种不均衡主要源于目标物体在尺度上的多样性。例如，在 COCO 数据集中，小尺度目标的占比显著高于大尺度目标。此外，在特征金字塔网络中，不同尺度目标的特征分配也可能存在不均衡现象。

（3）空间不均衡（Spatial Imbalance）：涉及多个方面，包括回归损失贡献的样本对不均衡、IoU 分布的不均衡以及目标物体在图像空间位置分布的不均衡。这些问题可能导致模型在某些区域或特定 IoU 范围内的检测性能下降。

（4）目标不均衡（Objective Imbalance）：主要体现在多任务损失优化过程中，分类损失与回归损失之间的权重分配不均衡。这种不均衡可能导致模型在分类精度和定位精度之间无法取得最优平衡，从而影响整体检测性能。

在一阶段目标检测算法中，研究者们主要通过优化损失函数来缓解类别不均衡问题。Lin 等人提出的 Focal Loss 是一种典型的解决方案[65]，其通过对不同样本的损失进行动态加权，使模型更加聚焦于难以分类的样本，从而显著提升了类

别不均衡情况下的分类精度。此外，Focal Loss 不仅有效缓解了样本不均衡问题，还对模型的整体性能提升具有积极作用。然而，Focal Loss 可能导致模型过度关注极端难分的样本，其中可能包含离群点，从而引发过拟合问题。为了解决这一问题，Liu 等人提出了梯度调和单阶段检测器（Gradient Harmonized Single-Stage Detector, GHM）[66]。GHM 引入了梯度模长的概念，通过梯度模长的大小来区分离群点与正常样本，从而有效降低了模型对难分样本的过度关注，进一步提升了模型的鲁棒性和性能。此外，Zhang 等人提出了 Varifocal Loss[67]，作为 Focal Loss 的改进版本。Varifocal Loss 通过调整负样本的损失贡献，同时保持正样本的损失贡献，实现了对难易样本的平衡控制。Li 等人提出了一种名为 Generalized Focal Loss 的方法[68]，该方法通过引入边界框的不确定性统计量，指导定位质量的学习，这一方法被 YOLOv8 使用与 CIoU 一起构成了回归损失。这些方法从不同角度解决了类别不均衡问题，为目标检测算法的优化提供了多样化的思路。

2.5 本章小结

本章围绕防毒面具佩戴检测任务，系统性地梳理了相关技术体系与理论基础。首先阐述了深度学习的基本原理与核心特征，随后介绍了卷积神经网络的发展脉络，并详细解析了其核心组件。在目标检测方法部分，对比分析了两阶段与单阶段检测框架的算法原理及性能特点，并详细讨论了检测任务中常用的损失函数设计，包括交并比及其衍生形式的应用。最后，本章还深入探讨了目标检测中的类别不均衡问题，并总结了当前一阶段目标检测算法中主流的解决方案。这些内容为后续防毒面具佩戴检测模型的构建与优化奠定了坚实的理论基础。

第 3 章 数据集及实验环境

3.1 数据集

3.1.1 防毒面具简介

防毒面具是一种用于保护佩戴者呼吸系统的个人防护装备，主要用于过滤空气中的有害物质，如有毒气体、化学污染物、生物病原体以及颗粒物等。它在军事、工业、医疗等领域具有广泛的应用。根据工作原理的不同，主要可分为过滤式和隔离式两种类型。过滤式防毒面具因结构简单、成本较低等优点，在工业场景下应用更加广泛。防毒面具佩戴检测数据集的研究对象为图 3-1 所示的 3M 半面罩型防毒面具。该面具配备双滤毒盒设计，能够高效过滤空气中的有害气体。滤毒盒内部填充了活性炭及其他高效吸附材料，通过物理吸附和化学反应的协同作用，显著提升了对空气中污染物的去除效率。

图 3-1 3M 防毒面具

Fig. 3-1 3M gas mask

3.1.2 数据集统计信息

本文中使用的数据集包括来自真实的工业场景和模拟场景的数据。该数据集并经过筛选整理和离线数据增强后，最终构建了一个包含 7998 张图像的数据集。本数据集根据头部对摄像头的方向将其分为三个不同的类别：正对摄像头、侧对摄像头和背对摄像头。接下来，考虑到实际场景下使用防毒面具的情况，将这三个类别进一步细分为 7 个标签。表 3-1 和图 3-2 提供了标签名称和含义。例如，

“佩戴错误”这样的标签指的是戴口罩。在工业环境中，只佩戴口罩的工人进入生产车间较常见，不符合安全防护要求，也应列为被检测的目标。

表 3-1 数据集标签表

Table 3-1 The label names and descriptions of the dataset

序号	标签	标签含义
1	Front_head_wear	正确佩戴防毒面具正对镜头
2	Side_head_wear	正确佩戴防毒面具侧对镜头
3	Front_head_no_wear	未佩戴防毒面具正对镜头
4	side_head_no_wear	未佩戴防毒面具侧对镜头
5	Front_head_wear_wrong	错误佩戴防毒面具正对镜头
6	side_head_wear_wrong	错误佩戴防毒面具侧对镜头
7	Back_head	背对镜头

图 3-2 标签可视化

Fig. 3-2 Visualization of labels

在本研究中，为了构建训练模型，从 11 个典型场景中选取了 6084 个样本作为训练数据集。这些场景涵盖了最常见的环境与情境，能够为模型提供多样化的

数据支持，从而增强其泛化能力。通过利用这些样本进行训练，模型能够有效学习不同场景下的特征与规律。此外，剩余的 4 个场景中的 1934 个样本被划分为测试数据集，用于评估模型在未知场景下的性能表现。

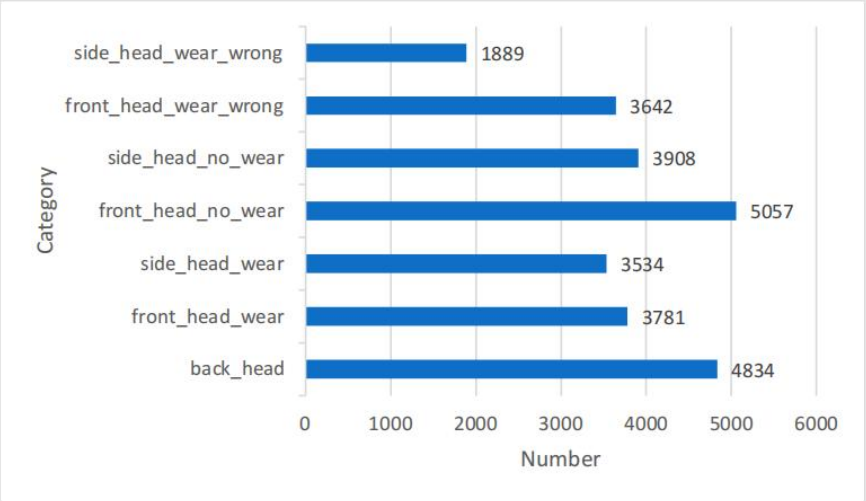


图 3-3 数据集类别的统计信息

Fig. 3-3 Statistics of each category in the dataset.

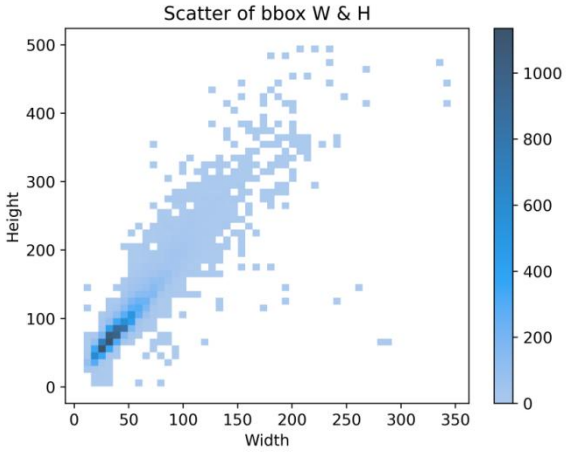


图 3-4 标注框宽高分布

Fig. 3-4 Bboxs width and height distribution

图 3-3 和图 3-4 分别展示了数据集类别的统计信息以及标注框宽高比的分布情况。图 3-4 进一步以散点图的形式呈现了数据集中边界框长度与宽度的分布。从图中可以看出，大多数边界框的尺寸小于 300×300 像素，并且在尺度上呈现出显著的多样性，这表明数据集中包含大量小尺度目标。小目标的存在可能导致前景与背景之间的不均衡问题。由于小目标在图像中占据的区域较小，这种不均衡可能使模型倾向于将更多区域分类为背景，从而忽略部分小目标。此外，小目标

的显著特征数量有限，可能导致模型将某些背景区域误分类为目标，进而增加误检率。这些问题需要在模型设计与训练过程中予以重点关注和优化。

3.2 实验环境

本文的所有实验环境一致，环境参数及训练参数如表 3-2、表 3-3 所示。

表 3-2 环境参数

Table 4-1 Environment parameter settings

硬件	型号
CPU	Intel® Core™ i9-10980XE CPU @ 3.00GHz
GPU	NVIDIA GeForce RTX3090
操作系统	Ubuntu20.4
显存大小	24GB
运行内存	64GB

表 3-3 训练参数设置

Table 4-2 Training parameter settings

参数	设定值
训练轮数	200
批量大小	16
图片尺寸	1280×1280
初始学习率	0.01
学习率动量	0.937
优化器	SGD

3.3 本章小结

本章阐述了防毒面具佩戴检测数据集的相关信息及本文主要实验环境。首先，对防毒面具的基本特性进行了介绍，涵盖其核心功能、结构组成以及常见分类。随后，对数据集的统计特征进行了分析，包括各类别标签的数量分布以及边界框尺寸的统计特性。这些信息为后续模型的训练与评估提供了重要的数据基础。

第 4 章 针对复杂背景干扰和样本不均衡的改进方法

本章针对防毒面具检测场景中存在的复杂背景干扰和样本不均衡的问题，以YOLOv5 模型为基础模型进行针对性改进，用于防毒面具佩戴检测任务。具体改进包括以下三个方面：首先，在原有三检测层的基础上，引入了一个小目标检测头（Small Object Detection Head，SODH），以增强模型对小尺度目标的检测能力；其次，在模型中嵌入了 CBAM（Convolutional Block Attention Module，CBAM）注意力机制模块[69]，通过有效区分目标与背景信息，提升检测精度，同时避免了计算复杂度的显著增加；最后，将原模型的二元交叉熵损失函数替换为 Varifocal Loss[70]，以确保模型在训练过程中更加关注高质量的正样本，从而优化检测性能。

4.1 YOLOv5 模型与改进

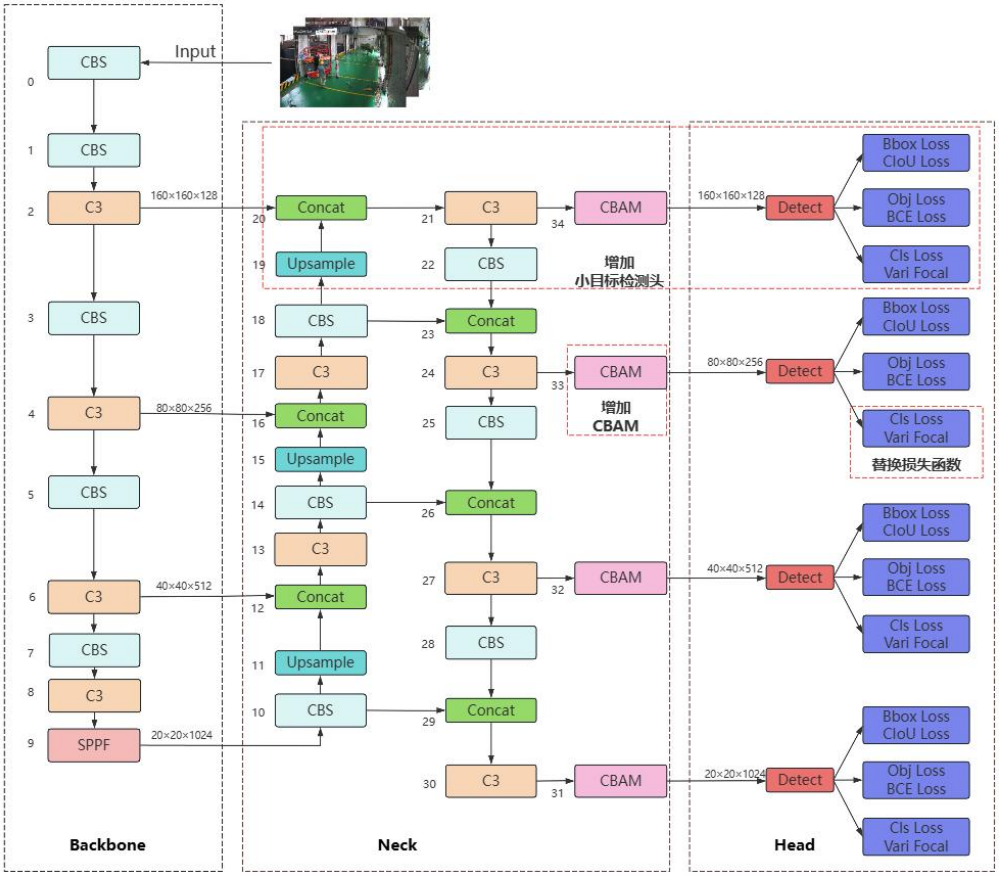


图 4-1 改进 YOLOv5 模型结构图

Fig.4-1 Diagram of improved YOLOv5 model architecture

图 4-1 展示了改进后的 YOLOv5-SCV 整体结构图，图像输入后在主干网络部分进行特征提取，在颈部网络中经过新添加的 SODH 和 CBAM 模块进行特征提取，最后在检测头进行分类和定位。

4.1.1 新增小目标检测头

由于 YOLOv5 目标检测网络的默认采用三个检测头，分别对应中等、较大和最大尺度的目标检测，但由于深层特征图经过多次下采样，空间分辨率较低，容易丢失小目标的细节信息，导致小目标检测性能不足。因此，本节通过添加一个小目标检测头来缓解检测防毒面具佩戴情况时的小目标丢失问题。

如图 4-2 所示，新增的红色箭头指示了额外的特征传递路径，而黄色模块则代表了引入的新增结构。通过将特征提取网络中仅经过两次下采样操作的特征图（即图中维度为 160×160×128 的浅蓝色模块）与新增模块相结合，采用路径聚合策略进行多层次特征融合，最终输出一个具有更高维度的检测头。这一改进有效提升了模型对小尺度目标的检测能力，增强了对细微特征的感知与提取，从而显著优化了整体目标检测性能。

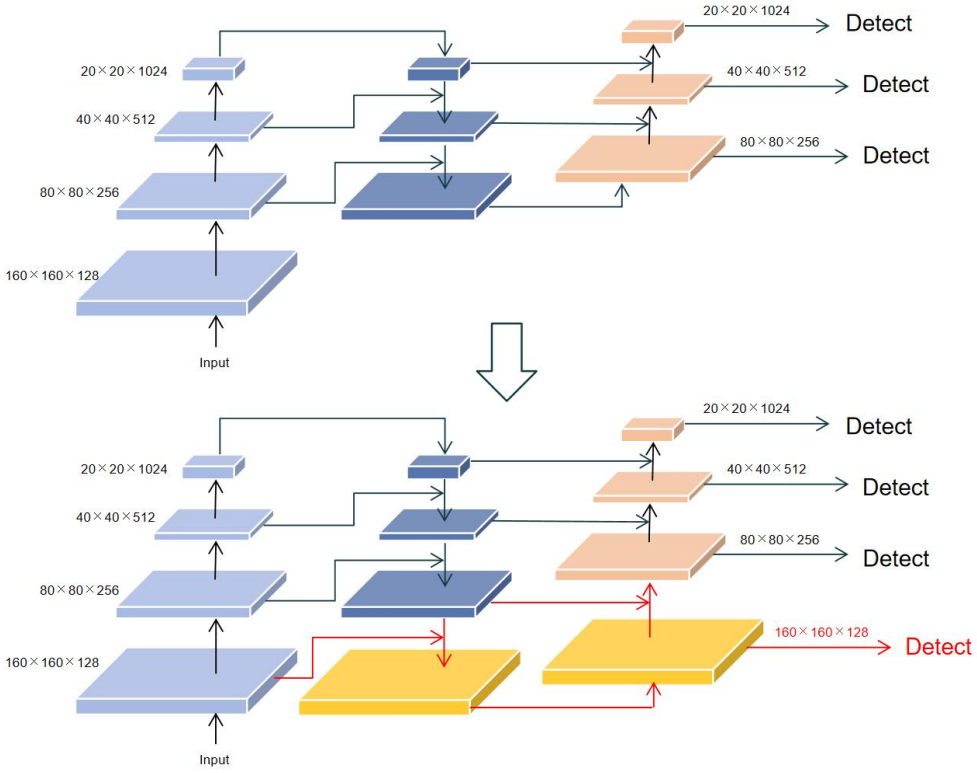


图 4-2 小目标检测头示意图

Fig.4-2 Diagram of small object detection head

4.1.2 新增 CBAM 注意力机制

在计算机视觉领域，传统卷积神经网络往往难以有效捕捉复杂场景中目标的多样性特征，这在一定程度上制约了其在实际应用中的性能表现。为了解决这一问题，Sanghyun Woo 等人在 2018 年提出的一种轻量级注意力机制 CBAM，旨在克服传统卷积神经网络在处理多尺度、多形状以及多方向信息时的固有局限性。其设计灵感源于人类视觉系统中注意力机制的工作方式，即通过选择性聚焦于重要信息来提升感知效率。如图 4-3 所示的 CBAM 通过引入双重注意力机制——通道注意力模块（Channel Attention Module，CAM）和空间注意力模块（Spatial Attention Module，SAM），显著增强了网络的特征提取能力。通道注意力模块专注于学习特征图中各通道的重要性，而空间注意力模块则聚焦于特征图中各空间位置的关键性。

如图 4-4 所示的通道注意力机制的目标是学习每个通道的重要性，并为重要的通道分配更高的权重。其核心思想是通过全局信息聚合和通道间关系建模，动态调整特征图中各通道的权重。

如图 4-5 所示的空间注意力机制的目标是学习特征图中每个空间位置的重要性，并为重要的位置分配更高的权重。其核心思想是通过空间信息聚合和位置关系建模，动态调整特征图中各位置的权重。

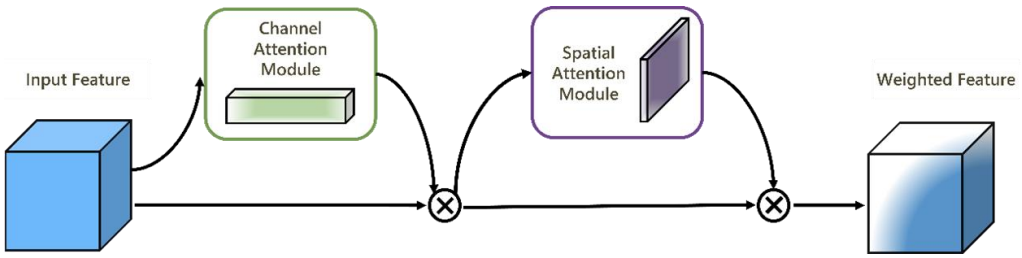


图 4-3 CBAM 注意力机制模块

Fig.4-3 CBAM attention mechanism module

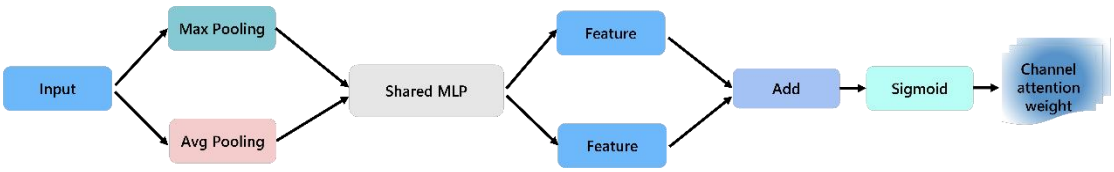


图 4-4 通道注意力模块

Fig.4-4 Channel attention module

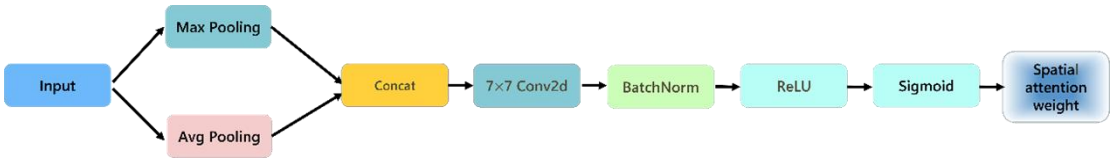


图 4-5 空间注意力模块

Fig4-5 Spatial attention module

4.1.3 解决样本不均衡问题

在 YOLOv5 模型中，损失函数由三个核心组件构成：分类损失、边界框回归损失以及置信度损失。其中，分类损失和置信度损失均采用二元交叉熵损失函数（Binary Cross-Entropy Loss, BCE Loss）进行计算，以优化目标类别的预测精度和目标存在的置信度评分。边界框回归损失则基于 IoU 函数进行设计，旨在精确度量预测边界框与真实边界框之间的空间匹配程度，从而提升定位精度。其整体损失函数如公式（4-1）所示。

$$L_{YOLOv5} = L_{cls} + L_{box} + L_{obj} \tag{4-1}$$

然而，二元交叉熵损失函数倾向于将更多的注意力分配给预测概率较高的类别，即样本数量较多的类别。这种特性导致模型对样本数量较少的类别关注不足，而本文所研究的防毒面具佩戴检测数据集中存在部分类别样本分布不均衡的问题，进一步加剧了这一问题。为了解决类别不均衡问题，本章采用 Varifocal Loss 替代 YOLOv5 模型中原有的二元交叉熵损失函数。Varifocal Loss 的核心优势在于其能够通过对正负样本进行非对称加权处理，动态调整不同目标的重要性，从而有效缓解类别不均衡以及难易样本分布不均的问题。这种改进显著提升了模型对少数类别的检测能力，同时优化了整体检测结果的质量。Varifocal Loss 的公式如下：

$$Q_{VFL}(p,q) = \begin{cases} -q[q\log p + (1-q)\log(1-p)], q > 0 \\ -\alpha p^{\beta}\log(1-p), q = 0 \end{cases} \tag{4-2}$$

在公式(4-2)中，p表示模型预测为正样本的概率，q表示样本的标签值。当输入为正样本时，q表示预测边界框与真实边界框之间的交并比 IoU；当输入为负样本时，q 的值为 0。参数α用于调节正负样本之间的权重平衡，而参数β 则用于控制损失函数的难度，使模型能够动态调整对难易样本的关注程度。从公式中可以看出，当q 的值较大时，表明当前预测框为正样本的概率较高，从而导致其损

失值增大，这使得模型能够更加聚焦于难以分类的正样本。当 q 为 0 时，损失值则通过标准的 Focal Loss 进行计算。

4.2 实验结果与分析

4.2.1 消融实验

为了全面评估不同改进策略对防毒面具佩戴检测模型的影响，本研究基于防毒面具佩戴检测数据集进行了一系列消融实验，并将结果汇总于表 4-1 中。实验 1 至实验 4 分别探讨了各个模块对模型性能的影响。在实验 1 中，采用 YOLOv5n 作为基准模型，其检测结果F_1为 69.0%，$map_{0.5}$为 74.8%，$map_{0.5:0.95}$为 52.1%。在实验 2 中，将小目标检测头增加到模型中，基准模型的性能得到了较好的提升。F_1提升 7%，$map_{0.5}$提升 8.5%，$map_{0.5:0.95}$提升 4.5%。在实验 3 中，将 CBAM 模块引入到 YOLOv5n，使得模型F_1达到 71.0%，$map_{0.5}$达到 81.2%，$map_{0.5:0.95}$达到 57.3%。实验 4 评估了引入 Varifocal loss 对模型性能的影响，从表格第四行可以看出相较于基准模型，F_1提升 6.8%，这与实验 2 的结果相近。综上所述，本文介绍的三种方法都有效提高了 YOLOv5 模型对于防毒面具佩戴检测效果。

表 4-1 消融实验

Table 4-1 Ablation experiment

序号	SODH	CBAM	VariFocal Loss	F_1(%)	$map_{0.5}$(%)	$map_{0.5:0.95}$(%)
1				69.0	74.8	52.1
2	√			76.0	83.3	56.6
3		√		71.0	81.2	57.3
4			√	75.8	81.8	57.7
5	√	√		80.0	85.4	57.5
6	√		√	79.2	85.1	58.7
7		√	√	79.8	85.4	59.1
8	√	√	√	80.2	86.7	60.0

实验 5、实验 6 和实验 7 主要研究了不同方法组合对模型性能的影响。在实验 5 中，本研究测试了同时增加小目标检测头和 CBAM 模块的效果。实验结果F_1达到了 80.0% ，$map_{0.5}$达到了 85.4%，$map_{0.5:0.95}$达到了 57.5%。实验 6 将小目标

检测头和 Varifocal loss 同时使用，结果显示F_1为 79.2%，$map_{0.5}$为 85.1%，$map_{0.5:0.95}$为 58.7%。实验 7 将 CBAM 模块和 Varifocal loss 同时使用，F_1、$map_{0.5}$和$map_{0.5:0.95}$相较于实验 6 有稍微提升。基于实验 4、6、7 的结果，将基准模型的二元交叉熵损失函数替换为 Varifocal Loss 能够有效提高模型的精度，这主要得益于其能够有效地缓解类别不平衡问题。

最后，在实验 8 中，本章对所提出的方法进行了全面测试。通过融合所有改进策略，模型的精度得到了显著提升。与基准模型相比，所提方法在F_1指标上提高了 11.2%，$map_{0.5}$指标上提高了 11.9%，$map_{0.5:0.95}$指标上提高了 7.9%。这些实验结果充分验证了所提方法在提高防毒面具检测性能方面的有效性。

4.2.2 对比实验

在本节对比实验中，本文提出的 YOLOv5-SCV 模型与当前主流目标检测模型进行了全面的性能对比，实验结果如表 4-2 所示。首先对比经典的两阶段算法 Faster R-CNN、Cascade R-CNN，之后对比单阶段目标检测算法 RetinaNet 以及 YOLO 系列的不同版本，如 YOLOv3、YOLOv5s/m/l、YOLOv7tiny。实验从检测精度、计算效率、推理速度和模型复杂度四个维度进行了详细分析，以验证 YOLOv5-SCE 的综合性能。

在检测精度方面，YOLOv5-SCV 在$map_{0.5}$和$map_{0.5:0.95}$两个关键指标上均表现出色。$map_{0.5}$达到了 86.7%，显著高于 Faster R-CNN、RetinaNet 和 YOLOv5s，与 YOLOv5l 和 YOLOv8s 相当，但略低于 Cascade R-CNN。然而，Cascade R-CNN 的高精度是以更高的计算成本和参数量为代价的。在$map_{0.5:0.95}$指标上，YOLOv5-SCV 以 60.0%的成绩超越了所有对比模型，包括两阶段检测算法的 Cascade R-CNN 以及 YOLOv5 系列的其他版本和 YOLOv7tiny，展现了其在复杂场景下的鲁棒检测能力。

在计算量方面，YOLOv5-SCV 的 GFLOPs 仅为 19.2，显著低于 Faster R-CNN、Cascade R-CNN 和 RetinaNet，同时也低于 YOLOv5m 和 YOLOv5l。尽管 YOLOv5s 和 YOLOv7tiny 的计算量更低，但它们的检测精度均低于 YOLOv5-SCV。这表明 YOLOv5-SCV 在计算效率与检测精度之间实现了更好的平衡。

在推理速度方面，YOLOv5-SCV 的 FPS 达到了 196，在对比模型中处于较高水平，显著高于 Faster R-CNN、Cascade R-CNN、RetinaNet 以及 YOLOv3。这一

结果表明，YOLOv5-SCV 能够满足实时检测的需求，尤其适用于对速度要求较高的应用场景。

在模型复杂度方面，YOLOv5-SCV 的参数量为 7.54M，相较于 YOLOv5s 和 YOLOv7tiny 略有增加。与 Faster R-CNN 和 Cascade R-CNN 相比，YOLOv5-SCV 的模型复杂度显著降低，更适合资源受限的嵌入式设备或移动端应用。

表 4-2 模型性能比较

Table4-2 Performance Comparison of Models

模型	主干	$map_{0.5}$ (%)	$map_{0.5:0.95}$ (%)	GFLOPs	FPS	Params (M)
Faster R-CNN	ResNet-50-FPN	83.7	55.3	322.34	20	41.15
Cascade R-CNN	ResNet-50-FPN	88.0	58.0	350.35	17	69.17
RetinaNet	ResNet-50-FPN	83.7	54.0	331.00	21	36.23
YOLOv3	Darknet-53	85.5	56.0	154.60	56	61.55
YOLOv5s	CSP-DarkNet-53	74.8	52.1	15.80	208	7.03
YOLOv5m	CSP-DarkNet-53	81.5	56.9	47.90	110	20.89
YOLOv5l	CSP-DarkNet-53	85.1	58.3	107.70	70	46.16
YOLOv7tiny	CSP-DarkNet-53	83.5	55.3	13.20	526	6.03
YOLOv5-SCV	CSP-DarkNet-53	86.7	60.0	19.2	196	7.54

综合以上实验结果，YOLOv5-SCV 在检测精度、计算效率、推理速度和模型复杂度之间实现了良好的平衡。其性能提升主要归功于本文提出的改进策略，包括网络结构优化、特征提取增强以及损失函数改进等。这些改进使得模型在保持较低计算成本和参数量的同时，显著提升了检测精度和推理速度。实验结果表明，YOLOv5-SCV 在目标检测任务中具有显著优势，尤其在精度与效率的平衡上表现突出，为实际应用提供了更优的解决方案。

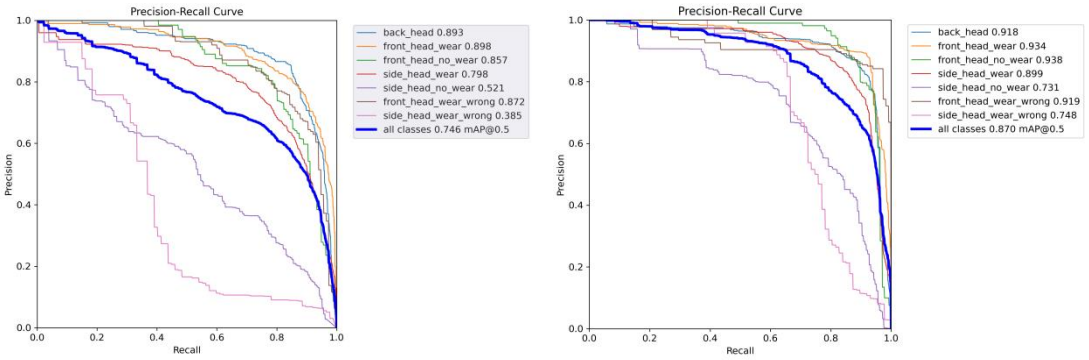
4.2.3 实验结果可视化

最后，对部分实验结果进行可视化展示。图 4-6 展示了改进后的 YOLOv5 模型的检测效果。实验结果表明，改进后的模型在复杂背景条件下展现出优异的检测稳定性，能够有效避免误检现象，并实现精准的目标定位与分类。然而，在背

光场景下，模型的检测置信度出现了一定程度的下降，这可能会对实际应用中的检测可靠性造成影响。尽管模型在多数场景下表现稳健，但背光条件下的性能衰减仍需进一步优化，以提升其在多样化环境中的适应性和鲁棒性。未来研究将针对这一问题展开深入探索，以增强模型在极端光照条件下的检测能力。

图 4-6 改进后的 YOLOv5 模型检测效果图

Fig.4-6 Detection results diagram of the improved YOLOv5 model



（a）YOLOv5 模型 P-R 图　　（b）改进 YOLOv5 模型 P-R 图

图 4-7 P-R 图对比

Fig4-7 P-R curve comparison

如图 4-7 所示，改进后的 YOLOv5 模型在检测性能上相较于基准模型提升显著，$map_{0.5}$ 提高了 12.4%。所有目标类别的检测精度均有所改善，其中

“side_head_wear_wrong”类别的提升幅度最为显著，从基准模型的 38.5%，达到了 74.8%。这是由于该类别样本数量相对较少，而改进后的 Varifocal Loss 能够有效分配更多的注意力资源至稀疏样本，从而增强模型对少数类别的检测能力。这一优化策略不仅提升了模型对“side_head_wear_wrong”类别的识别效果，同时也进一步增强了整体检测性能，体现了改进算法在类别不平衡问题上的有效性。

4.3 本章小结

本章针对防毒面具检测场景中存在的复杂背景干扰和样本不均衡的问题，提出了一种基于 YOLOv5 框架的改进模型，并将其应用于防毒面具佩戴检测任务。首先，详细阐述了小目标检测头的设计原理及其在提升小目标检测性能中的作用。其次，介绍了 CBAM 注意力模块的多层次结构及其在增强特征表达能力方面的优势。此外，采用 Varifocal Loss 替代传统损失函数，以解决类别不均衡问题并提高模型对难样本的识别能力。为验证改进策略的有效性，本章设计并实施了消融实验，实验结果证实了三种改进方法在提升模型性能方面的显著贡献。实验结果表明，改进后的模型在分类能力上取得了显著提升，能够更好地适应复杂多变的检测场景。最后，通过与其他主流模型的对比实验，进一步验证了改进模型在检测精度和推理速度上的优越性，充分满足了实际应用的需求。

第 5 章 针对特征层信息融合差和多尺度检测的改进方法

上章提出的 YOLOv5-SCV 模型虽然在存在的复杂背景干扰和类别不均衡等问题上提出了解决方案，但并未解决网络对提取到的不同特征层信息融合效果差和存在多尺度目标检测的问题，因此本章提出了 YOLOv8-LSE 模型解决上述问题。

5.1 YOLOv8 模型与改进

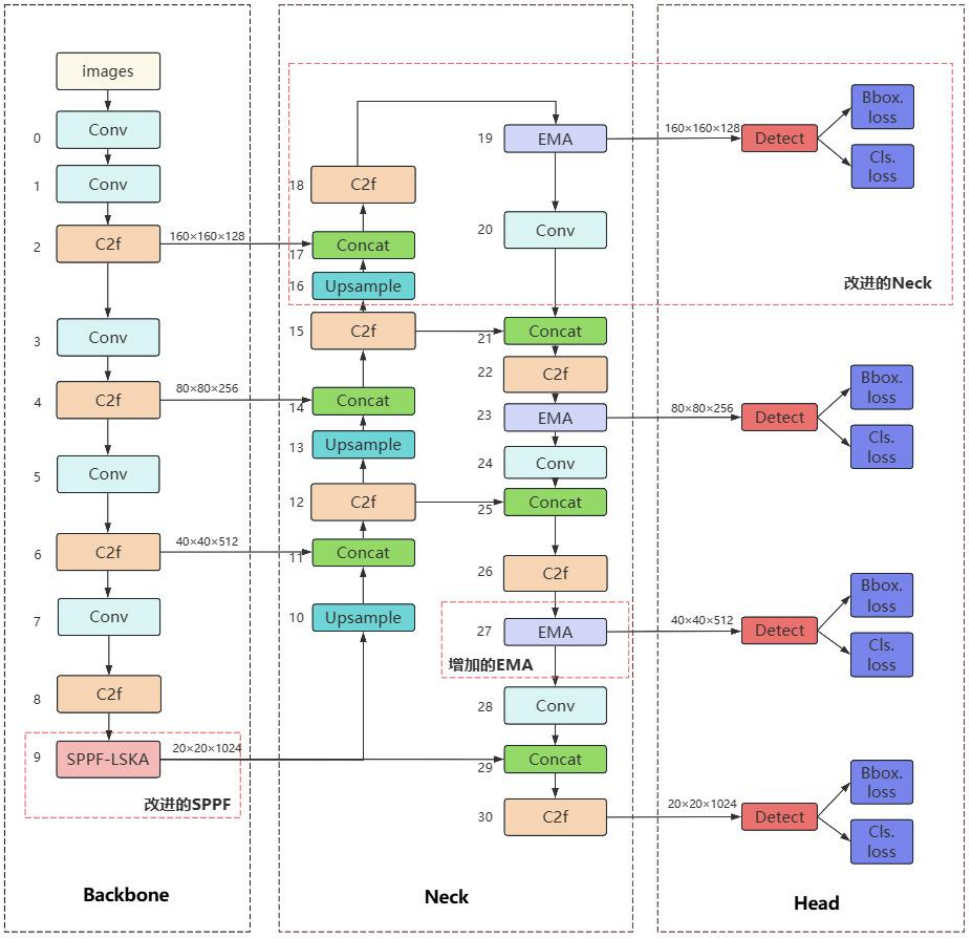


图 5-1 改进的 YOLOv8 模型结构图

Fig.5-1 Diagram of improved YOLOv8 model architecture

本章针对网络提取到的不同特征层信息融合效果差和存在多尺度目标检测的问题，以 YOLOv8 为基准框架，提出了一种改进模型——YOLOv8-LSE，并将其应用于防毒面具佩戴检测任务。改进主要包括以下三个方面，如图 5-1 所示：

首先，在空间金字塔池化融合模块（Spatial Pyramid Pooling Fusion, SPPF）[32]中引入了大核分离卷积注意力模块（Large Separable Kernel Attention, LSKA）[74]，以增强不同特征层之间的语义信息融合能力；其次，将主干网络中 P2 层的输出集成到颈部结构，并额外增加了一个小目标检测头，从而提升模型对小尺度目标的检测性能；最后，在检测头部引入了高效多尺度注意力模块（Efficient Multi-scale Attention, EMA）[75]，该模块能够有效学习待检测目标的多尺度语义信息。

5.1.1 改进的 SPPF 模块

在传统的卷积神经网络中，卷积操作的大小通常较小。这些小卷积核能够捕捉局部特征，但在捕捉全局依赖或长距离依赖时会遇到限制。为了克服这一问题，Kin 等人于 2023 年提出大核可分离核注意力机制 LSKA。LSKA 结合了大核卷积和可分离卷积两种思想，通过大尺寸卷积核扩大感受野以捕捉长距离依赖关系，同时利用可分离卷积将标准卷积分解为深度卷积和逐点卷积，从而显著减少参数量和计算量，提升模型的表达能力和效率。此外，LSKA 还引入了注意力机制，动态调整特征图中不同位置的重要性，以增强模型对关键信息的关注。

LSKA 注意力机制的结构主要由以下几个关键组件构成：

1.深度卷积：通过小尺寸卷积核对输入数据进行深度卷积操作，以提取局部特征并增强特征的细节表达能力；

2.展平操作：对深度卷积输出的特征图进行展平处理，以便适应后续的大核卷积操作；

3.可分离卷积：采用大尺寸可分离卷积核进行卷积计算，该卷积核由多个小卷积核级联组成，从而在减少参数数量和计算复杂度的同时，保留大感受野的优势；

4.注意力权重计算：基于可分离卷积的输出，计算注意力权重矩阵，该矩阵表征了特征图中不同位置之间的相关性；

5.权重应用：将注意力权重与输入特征图进行逐元素相乘，生成加权后的特征表示，以突出关键信息；

6.聚合特征：对加权特征进行聚合操作，生成统一的特征表示；

7.输出：将聚合后的特征传递至后续网络层，用于模型的训练或推理过程。这一结构设计使得 LSKA 能够在高效计算的同时，显著增强模型对全局上下文信息的捕捉能力。

如图 5-2 所示，将 LSKA 模块融合进空间金字塔池化融合模块 SPPF 中，能够进一步增强多尺度特征的语义融合能力。SPPF 通过多尺度池化操作提取不同感受野的特征，但其在特征融合过程中对全局上下文信息的捕捉能力有限。引入 LSKA 后，SPPF 能够利用大核可分离卷积增强特征层之间的长距离依赖关系，同时通过注意力机制动态调整多尺度特征的权重，从而提升特征融合的效果。这种改进使得模型在复杂背景和多尺度目标检测任务中表现更加鲁棒，尤其适用于小目标检测场景。实验结果表明，将 LSKA 融入 SPPF 后，模型的检测精度显著提升，同时保持了较高的推理效率，验证了该改进策略的有效性。

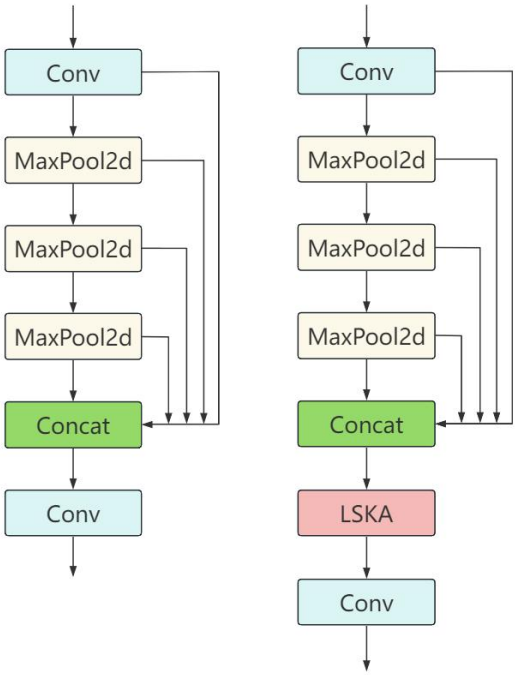


图 5-2 SPPF 和 SPPF-LSKA 模块示意图

Fig.5-2 Diagram of SPPF and SPPF-LSKA modules

5.1.2 改进的 Neck

由于 YOLOv8 网络模型的 Neck 结构与 YOLOv5 相同，都默认采用三个检测头，分别对应中等、较大和最大尺度的目标检测。鉴于上一章新增小目标检测头的分类性能和检测效果均有较大提升。因此，在本节中，对 Neck 结构的改进延

续了上一章中增加小目标检测头的设计思路，并在此基础上进一步优化了特征融合策略。本节的改进与上一章新增小目标检测头的工作保持一致，均在 Neck 部分引入了额外的小目标检测头，以增强模型对小尺度目标的检测能力。这一设计通过充分利用浅层特征的高分辨率信息，有效提升了模型在复杂场景下对小目标的定位与识别精度。本节的主要贡献在于验证了上一章改进策略的通用性和可扩展性，同时进一步分析了小目标检测头在不同任务场景下的性能表现。

5.1.3 新增 EMA 模块

在当前主流的注意力机制研究中，SE（Squeeze and Excitation，SE）[75]、CBAM 等模块在提升模型性能方面展现了显著的效果。但是，这些机制仍存在若干局限性。SE 模块通过 2D 全局平均池化计算通道注意力，能够在较低的空间成本下提升模型表现[76]，但其设计未充分考虑位置信息的建模，导致空间特征的利用不足。CBAM 模块则通过减少输入张量的通道维度并结合卷积神经网络捕捉位置信息，然而，由于卷积操作的局部特性，其在处理长距离依赖关系时表现受限。此外，通道降维策略在跨通道关系建模中可能对深层视觉特征的提取产生负面影响。以 CA（Coordinate Attention，CA）[77]为例，尽管其在精度上有所提升，但其全局注意力权重的计算需要消耗额外的计算资源，且难以有效捕捉通道间的长距离依赖关系。同时，CBAM 模块因引入了额外的注意力分支，导致模型复杂度和计算量增加，且其性能可能随着网络深度的增加而逐渐衰减。这些局限性表明，现有注意力机制在效率、长距离依赖建模以及计算资源消耗方面仍有优化空间。

针对上述问题，深圳航天科工于 2023 年 6 月提出了一种创新的高效多尺度注意力机制 EMA 模块，其模块结构如图 5-3 所示。该机制在处理输入特征图时，首先将通道划分为若干组，每组包含特定数量的通道；随后，将每组通道重塑为批量维度，并通过 1×1 卷积核实现降维操作；接着，利用 2D 全局平均池化对 3×3 分支中的全局空间信息进行编码，同时将 1×1 分支的特征转换为与联合激活机制兼容的维度[78]；最后，将降维后的特征图与原始特征图相加，并通过 Sigmoid 函数生成注意力权重系数。EMA 机制的核心优势在于其能够在多尺度上有效捕捉跨通道关系，通过分组和重塑操作在同一组内及不同组之间建立通道间的关联，从而显著提升深度视觉表示的建模能力。此外，EMA 机制在计算效率方面表现

突出，其采用的分组和降维策略避免了传统多尺度注意力模块中重复卷积操作带来的计算开销，在维持强大特征提取能力的同时显著降低了计算成本。

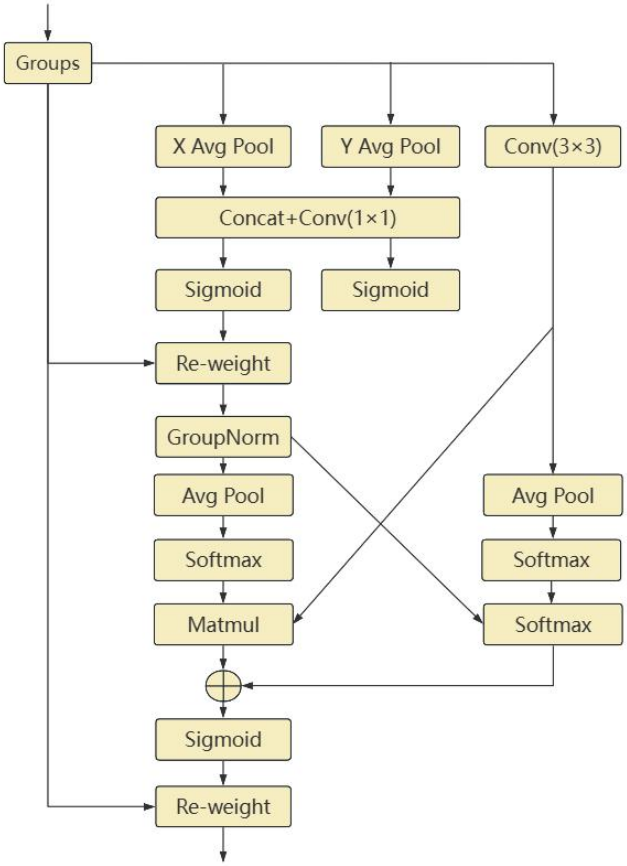


图 5-3 EMA 模块示意图

Fig.5-3 Diagram of the EMA module.

将 EMA 模块集成到 YOLOv8 的网络结构中，可以显著提升防毒面具佩戴检测模型的性能。首先，EMA 模块通过多尺度注意力机制增强了特征金字塔网络的多层次特征融合能力。在防毒面具检测任务中，目标尺度可能因距离差异而变化，且背景复杂多变，传统 Neck 层在特征融合时难以充分捕捉全局上下文信息。EMA 模块通过分组和重塑操作，在同一组内及不同组之间建立通道关联，从而更好地建模全局上下文信息，显著提升了模型对不同尺度目标的检测能力。其次，EMA 模块通过分组和降维策略，避免了传统注意力机制中重复卷积操作带来的计算开销，在保持强大特征提取能力的同时显著降低了计算成本，确保了模型在实时检测任务中的高效性。此外，EMA 模块通过动态生成注意力权重，能够自适应地增强关键特征并抑制冗余信息，从而提升模型在复杂场景下的检测精度和鲁棒性。

5.2 实验结果与分析

5.2.1 消融实验

为了全面评估不同改进策略对防毒面具佩戴检测模型的影响，本研究基于防毒面具佩戴检测数据集进行了一系列消融实验，并将结果汇总于表 5-1 中。实验 1 至实验 4 分别探讨了各个模块对模型性能的影响。在实验 1 中，采用 YOLOv8 作为基准模型，其检测结果F_1为 76.5%，$map_{0.5}$为 80.6%，$map_{0.5:0.95}$为 57.0%。在实验 2 中，将 LSKA 增加到模型中，基准模型的性能得到了较好的提升。F_1提升 1.3%，$map_{0.5}$提升 4.3%，$map_{0.5:0.95}$提升 1.3%。在实验 3 中，将小目标检测头模块引入到 YOLOv8 中，使得模型F_1达到 75.3%，$map_{0.5}$达到 82.3%，$map_{0.5:0.95}$达到 57.9%。实验 4 评估了增加 EMA 注意力机制模块对模型性能的影响，从表格第四行可以看出相较于基准模型，F_1提升 1.3％，这与实验 2 的结果相同。综上所述，本文介绍的三种方法都有效提高了 YOLOv8 模型对于防毒面具佩戴检测效果。

实验 5、实验 6 和实验 7 主要研究了不同方法组合对模型性能的影响。在实验 5 中，本研究测试了同时增加 LSKA 和小目标检测头模块的效果。实验结果F_1达到了 76.2% ，$map_{0.5}$达到了 84.5%，$map_{0.5:0.95}$达到了 59.3%。实验 6 将 LSKA 和 EMA 同时使用，结果显示F_1为 76.6%，$map_{0.5}$为 84.3%，$map_{0.5:0.95}$为 58.5%。实验 7 将小目标检测头模块和 EMA 同时使用，F_1相较于实验 6 有稍微提升。

表 5-1 消融实验表

Table 5-1 Ablation experiment

序号	LSKA	SODH	EMA	F_1	$map_{0.5}$（%）	$map_{0.5:0.95}$（%）
1				76.5	80.6	57.0
2	√			77.8	84.9	58.3
3		√		75.3	82.3	57.9
4			√	77.8	83.6	57.5
5	√	√		76.2	84.5	59.3
6	√		√	76.6	84.3	58.5
7		√	√	77.5	83.1	58.0
8	√	√	√	79.0	85.6	60.5

最后，在实验 8 中，本章对所提出的方法进行了全面测试。通过融合所有改进策略，模型的精度得到了显著提升。具体的说，与基准模型相比，所提方法在F_1指标上提高了 2.5%，$map_{0.5}$指标上提高了 5.0%，$map_{0.5:0.95}$指标上提高了 3.5%。这些实验结果充分验证了所提方法在提高防毒面具检测性能方面的有效性。

5.2.2 对比实验

在本节对比实验中，本文提出的 YOLOv8-LSE 模型与当前主流目标检测模型进行了全面的性能对比，实验结果如表 5-2 所示。首先对比经典的两阶段算法 Faster R-CNN、Cascade R-CNN，之后对比单阶段目标检测算法 RetinaNet 以及 YOLO 系列的不同版本（如 YOLOv8n/s 和 YOLOv10n）。实验从检测精度、计算效率、推理速度和模型复杂度四个维度进行了详细分析，以验证 YOLOv8-LSE 的综合性能。

在检测精度方面，YOLOv8-LSE 在 mAP@0.5 指标上达到 85.6%，在 mAP@0.95 指标上达到 60.5%，均优于大多数对比模型。例如，Faster R-CNN 的 mAP@0.5 为 83.7%，Cascade R-CNN 为 88.0%，而 YOLOv5-SCV 为 86.7%。尽管 Cascade R-CNN 在 mAP@0.5 上略高于 YOLOv8-LSE，但其计算成本和参数量显著增加，而 YOLOv8-LSE 在保持高精度的同时实现了更高的效率。

在计算量方面，YOLOv8-LSE 的计算成本仅为 12.5 GFLOPs，显著低于 Faster R-CNN 和 Cascade R-CNN，同时也优于 YOLOv5-SCV 和 YOLOv8s。与轻量级模型 YOLOv8n 和 YOLOv10n 相比，YOLOv8-LSE 在计算效率上略有增加，但其精度提升更为显著。

在推理速度方面，YOLOv8-LSE 的推理速度达到 170 FPS，能够满足工业场景中的实时检测需求。相比之下，Faster R-CNN 和 Cascade R-CNN 的推理速度分别仅为 20 FPS 和 17 FPS，远低于 YOLOv8-LSE。尽管 YOLOv8n 和 YOLOv10n 在速度上更快，但其检测精度显著低于 YOLOv8-LSE。

在模型复杂度方面，YOLOv8-LSE 的参数量优化至 3.20M，远低于 YOLOv5-SCV 和 YOLOv8s，同时也优于 Faster R-CNN 和 Cascade R-CNN。与 YOLOv8n 和 YOLOv10n 相比，YOLOv8-LSE 在参数量上略有增加，但其精度和效率的提升更为显著。

表 5-2 模型性能比较

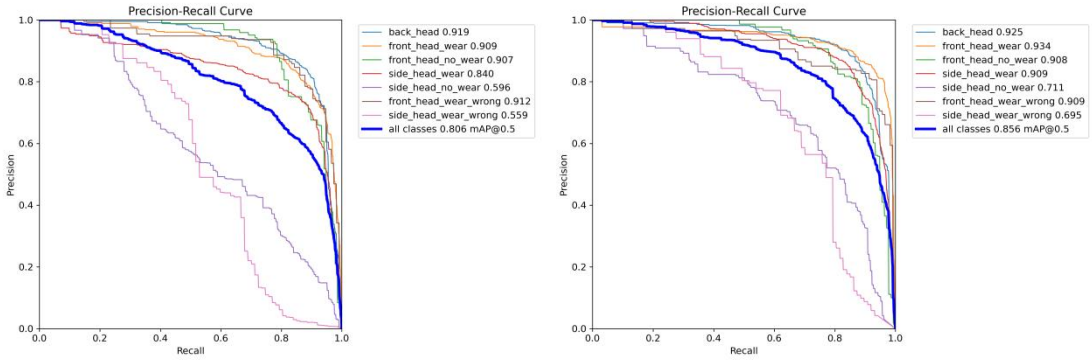
Table 5-2 Performance Comparison of Models

模型	主干	$map_{0.5}$ (%)	$map_{0.5:0.95}$ (%)	GFLOPs	FPS	Params (M)
Faster R-CNN	ResNet-50-FPN	83.7	55.3	322.34	20	41.15
Cascade R-CNN	ResNet-50-FPN	88.0	58.0	350.35	17	69.17
RetinaNet	ResNet-50-FPN	83.7	54.0	331.00	21	36.23
YOLOv5-SCV	CSP-DarkNet-53	86.7	60.0	19.2	196	7.54
YOLOv8n	CSP-DarkNet-53	80.6	57.0	8.9	357	3.16
YOLOv8s	CSP-DarkNet-53	85.0	59.0	28.8	154	11.2
YOLOv10n	CSP-DarkNet-53	83.2	58.1	6.5	413	2.26
YOLOv8-LSE	CSP-DarkNet-53	85.6	60.5	12.5	170	3.20

综合以上实验结果，YOLOv8-LSE 在检测精度、计算效率、推理速度和模型复杂度之间实现了良好的平衡。其性能提升主要归功于本文提出的改进策略，包括网络结构优化、特征提取增强等方法。这些改进使得模型在保持较低计算成本和参数量的同时，显著提升了检测精度和推理速度。实验结果表明，YOLOv8-LSE 在目标检测任务中具有显著优势，尤其在精度与效率的平衡上表现突出，为实际应用提供了更优的解决方案。

5.2.3 实验结果可视化

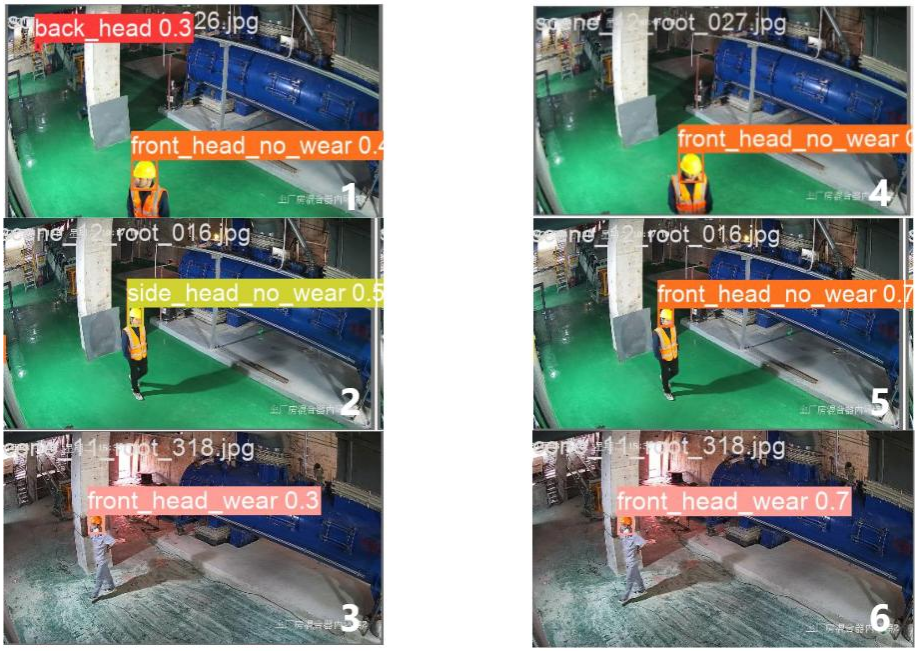
最后，通过可视化分析对部分实验结果进行了展示。如图 5-4 所示，改进后的 YOLOv8 模型在检测性能上相较于基准模型实现了显著提升，整体精度提高了 5%。实验结果表明，YOLOv8 基准模型在样本数量较多的类别上已经表现出较高的识别精度，而改进后的模型在所有目标类别的检测精度上均有所提升。其中，“side_head_wear_wrong”类别的提升幅度最为显著，其精度从基准模型的 55.9%提升至 69.5%。这一显著提升主要归因于该类别在基准模型中的检测精度相对较低，改进后的模型通过优化特征提取和融合策略，进一步增强了对该类别的识别能力。改进后的模型在整体检测性能上也得到了优化，为模型在复杂场景下的整体性能提升提供了可靠支持。



（a）YOLOv8 模型 P-R 图　　（b）YOLOv8-LSE 模型 P-R 图

图 5-4 P-R 图对比

Fig5-4 P-R curve comparison



（a）YOLOv8 检测效果图　　（b）YOLOv8-LSE 检测效果图

图5-5 改进前后的检测效果对比

Fig.5-5 Comparison of detection performance and after improvement

图 5-5 展示了 YOLOv8 基准模型与改进后的 YOLOv8 模型的检测效果对比。从图中左侧的基准模型检测结果可以看出，原模型存在明显的误检、错检以及检测精度不足的问题。分图 1 中，模型错误地将背景区域误检为“back_head”；分图 2 中，模型将“front_head_no_wear”类别错误地识别为“side_head_no_wear”；分图 3 中，模型对“front_head_wear”类别的检测置信度仅为 30%左右，表明其定位与分类能力存在显著不足。相比之下，改进后的 YOLOv8 模型不仅能够准确识

别目标类别，还显著提升了检测精度，有效解决了基准模型中的误检和错检问题。实验结果可视化表明，改进策略在提升模型检测性能方面具有显著效果。

5.3 本章小结

本章针对网络提取到的不同特征层信息融合效果差和存在多尺度目标检测的问题，提出了一种基于 YOLOv8 框架的改进模型。首先，详细阐述了 LSKA 的设计原理及其对增强多尺度特征的语义融合能力；其次，介绍了小目标检测头的优势。此外，采用 EMA 注意力模块以增强模型对目标特征信息的学习能力，提高对小目标的检测精度。为验证改进策略的有效性，本章设计并实施了消融实验，实验结果证实了三种改进方法在提升模型性能方面的显著贡献。实验结果表明，改进后的模型在分类能力上取得了显著提升，能够更好地适应复杂多变的检测场景；最后，通过与其他主流模型的对比实验，进一步验证了改进模型在检测精度和推理速度上的优越性，充分满足了实际应用的需求。

第 6 章 针对小目标检测困难和特征信息提取差的改进方法

上章提出的 YOLOv8-LSE 模型虽然在不同特征层信息融合效果差和多尺度目标检测的问题上提出了解决方案，但小目标检测头 SODH 造成了过多的冗余计算，导致模型复杂度变大，且并未对小目标检测困难、特征信息提取能力差以及样本不均衡的问题进行解决，因此本章提出了基于 YOLOv10 的改进方法解决上述问题。

6.1 YOLOv10 模型与改进

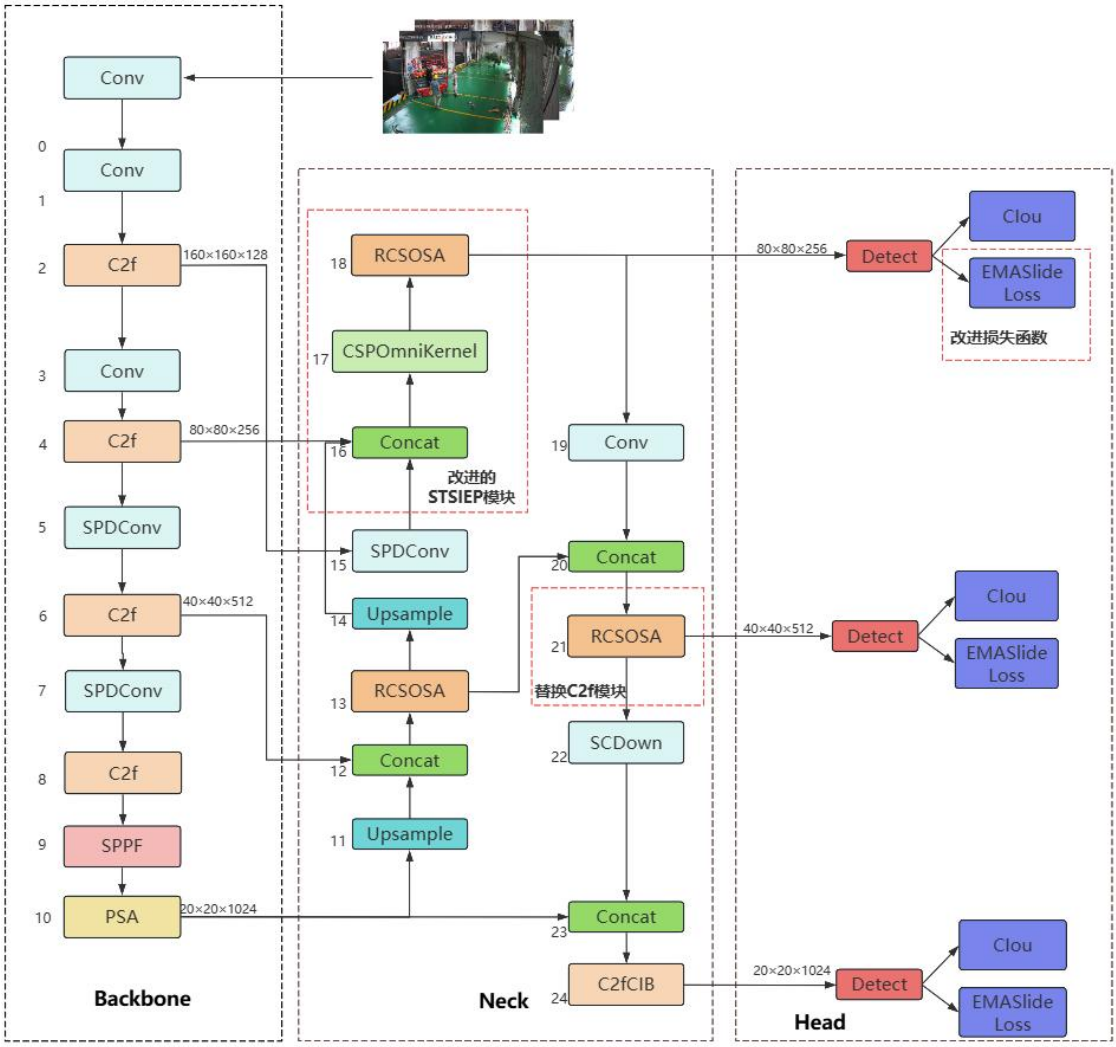


图 6-1 改进的 YOLOv10 模型结构图

Fig.6-1 Diagram of improved YOLOv10 model architecture

本章针对模型对小目标检测困难、特征信息提取能力差的问题以 YOLOv10 为基准模型进行改进，首先设计了小目标语义信息增强金字塔 STSIEP，以增强对小目标的语义信息特征融合效果，其次在 Neck 中使用 RCS-OSA 替换原模型的 C2f，对提取的特征信息进行增强及融合，最后使用指数滑动样本加权函数 EMA-Slide Loss 替换原模型的交叉熵损失函数，解决样本不平衡的问题，增强了模型的鲁棒性。SRE-YOLO 目标检测平均精度 mAP@0.5:0.95 相较于基线模型提高了 3.3%，达到了 61.4%，表明该模型具有较高的检测精度。改进后 SRE-YOLO 模型结构图如图 6-1 所示。

6.1.1 改进的特征融合模块

原有的 YOLOv10 模型只有 P3、P4、P5 三种不同尺度的检测层，P3 层的特征图已经下采样到 80x80，对于小于 32x32 的目标来说特征信息较少，因此，检测小目标是比较困难的。而传统的方法通常是在 P3 层之前增加一个额外的小目标检测头，以增强对小目标的检测。然而，这可能会导致更高的计算复杂度和更长的后处理时间。为了避免这些问题，本章使用 CSP 思想[23]和 OmniKernel[78]进行改进，设计了 CSPOmniKernel 进行特征融合构建出小目标语义信息增强金字塔 STSIEP，CSPOmniKernel 模块结构如图 6-2 所示。该模块应用核大小为 K×K 的深度卷积来追求大的感受野，并使用 1×K 和 K×1 深度卷积来获取上下文信息，通过 SPDConv 处理语义信息，提取富含小目标信息的特征。然后将这些特征集成到 P3 层中。这种方法有效地学习了从全局到局部尺度的特征表示，最终提高了小目标的检测性能。

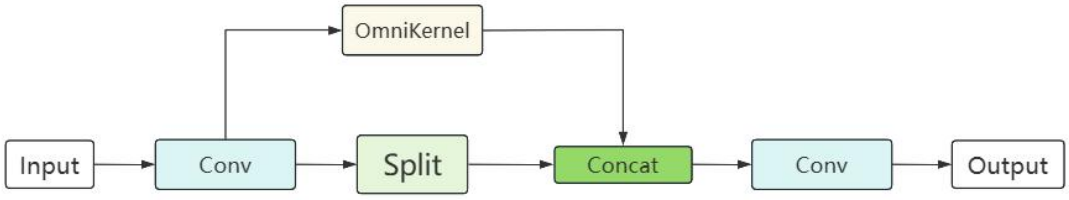


图 6-2 CSPOmniKernel 模块结构图

Fig.6-2 Diagram of the CSPOmniKernel module

CSP 思想是在 YOLOv4 和 YOLOv5 中首次引入，并成为后续 YOLO 系列模型的重要优化策略，其核心在于通过部分 Partial Cross Stage Connect 来减少计算冗余、增强特征复用能力，并缓解深层网络中的梯度消失问题。CSP 模块通过将

输入特征图沿通道维度分为两部分来实现部分跨阶段连接。其中一部分特征图直接传递到下一阶段，保留原始信息；另一部分则经过密集的卷积操作提取更丰富的特征表示。随后，这两部分特征图通过拼接或逐元素相加的方式进行融合，从而整合不同层次的信息。这种设计不仅减少了计算量，还增强了特征的多样性和复用能力。

OmniKernel 模块是 Cui 等人于 2024 年提出的一种卷积神经网络模块，专为图像恢复任务设计，旨在通过多尺度特征表示学习提升模型性能。其核心思想在于结合全局、大尺度和局部特征表示，以增强模型对多尺度退化图像的重建能力。OmniKernel 模块的设计灵感来源于 Transformer 模型的全局感受野优势，但通过卷积神经网络实现，避免了 Transformer 因输入尺寸二次增长带来的计算复杂度问题。

OmniKernel 模块如图 6-3 所示，由三个并行分支组成，分别用于捕捉不同尺度的特征信息。全局分支通过双域通道注意模块（Dual-domain Channel Attention Module, DCAM）和基于频率的空间注意模块（Frequency-based Spatial Attention Module, FSAM）实现全局感知场。DCAM 在通道维度上动态调整特征权重，而 FSAM 利用傅里叶变换在频域中捕捉全局信息，从而增强模型对整体图像内容的理解。大分支采用异常大的卷积核进行深度卷积，以扩大感受野，并通过 1×K 和 K×1 的条状卷积捕捉不同形状的上下文信息，进一步提升特征表示能力。局部分支则使用 1×1 深度卷积补充局部信息，确保模型能够捕捉小尺度退化细节。通过将这三个分支的特征进行融合，OmniKernel 模块能够有效整合从全局到局部的多粒度特征，显著提升图像恢复的精度。

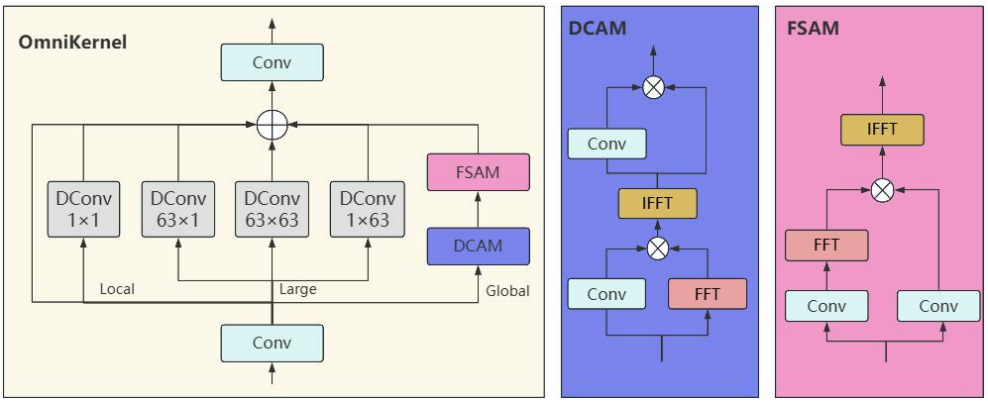


图 6-3 OmniKernel 模块结构图

Fig.6-3 Diagram of the OmniKernel module

6.1.2 改进信息提取模块

RCS-OSA 模块是 Ming Kang 等人于 2023 年提出的一种目标检测模块[79]，旨在通过优化特征提取过程提升模型的计算效率和检测精度。其核心思想结合了 RCS（Reparameterized Convolution based on channel Shuffle，RCS）和 OSA（One-Shot Aggregation，OSA）技术，通过通道混洗、重参数化卷积和一次性聚合，显著增强了模型的特征表示能力和计算效率。

RCS-OSA 模块的主要目的是有效地减小通道尺寸，增强空间对象注意力。RCS 模型结构如图 6-4 所示，使用具有 3×3 卷积、1×1 卷积和 Identity 分支的 RepVGG 结构进行训练，多分支拓扑允许在训练过程中提供更丰富的特征信息。在推理阶段，使用 3×3 卷积的 RepConv 进行结构重参数化，简单的单分支结构减少了推理阶段的内存和计算量，提高了推理速度。同时在通道维度上由 RepVGG 和 RepConv 进行拼接，然后用张量的通道分割部分拼接。最后，通过 Channel Shuffle 对两个分支特征图通道进行重组，促进不同通道之间的信息交换，提取更丰富的特征信息。为了减轻与 RCS 模块相关的计算负担，通过结合 RCS 和 OSA 提出了 RCS-OSA 模块，如图 6-5 所示。该模块通过反复堆叠 RCS 模块来实现功能重用。OSA 路径上保留了三个特征级联，以减少计算工作量，实现高精度的快速推理。

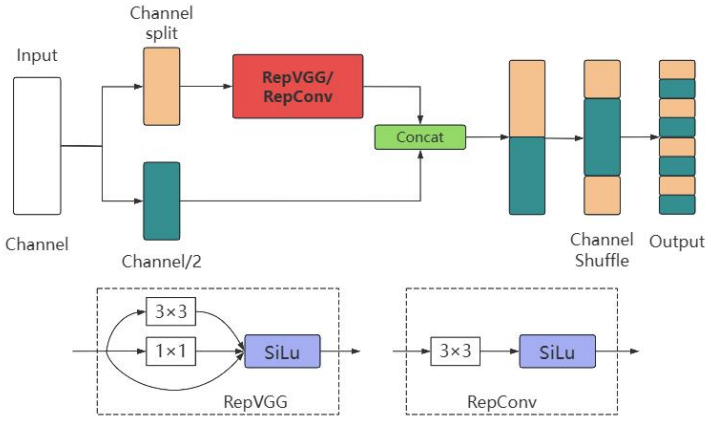


图 6-4 RCS 模块结构图

Fig.6-4 Diagram of the RCS module

将 RCS-OSA 模块集成到 YOLOv10 模型中，对防毒面具佩戴检测任务具有显著的性能提升。首先，RCS-OSA 模块通过增强特征提取能力，提高了模型对小尺度目标的检测精度。其次，其高效的计算特性使得模型在保持高精度的同时，

可以实现工业场景下实时检测，特别适用于复杂场景下的防毒面具佩戴检测任务。此外，RCS-OSA 模块通过一次性聚合和通道混洗操作，增强了模型对多尺度目标的检测能力，有效减少了误检和漏检现象。

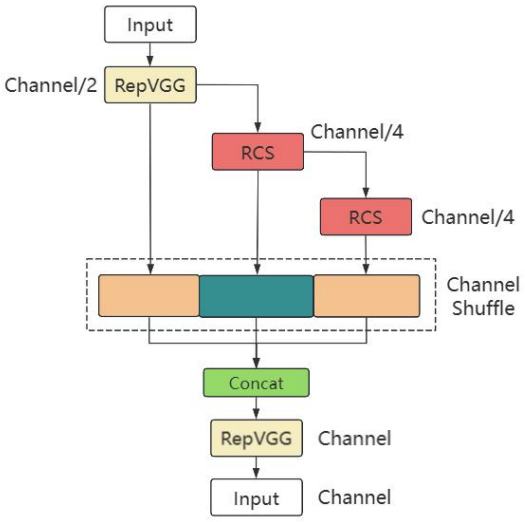


图 6-5 RCS-OSA 模块结构图

Fig.6-5 Diagram of the RCS-OSA module

6.1.3 改进损失函数

YOLOv10 目前的类别分类损失采用二元交叉熵损失。交叉熵损失在训练过程中需要模型具有较高的稳定性。然而，在模型训练的初始阶段学习率必须较小，否则可能导致不稳定、振荡或梯度爆炸。同时，交叉熵损失对类别不平衡问题非常敏感，这意味着训练集中样本较少的类别可能会被忽略[79]。相比之下，Yu 等人于 2022 年提出的样本加权函数 Slide-Loss 在缓解类别不平衡问题方面表现出显著效果。为了进一步增强训练过程中的模型稳定性，刘等人基于 Slide-Loss 提出了改进的 EMA-SlideLoss。SlideLoss 的公式如公式（6-1）所示，其中μ表示所有边界框的平均交并比值即 IoU 值。IoU 值低于μ的样本被归类为负样本，高于μ的样本被归类为正样本。然而，由于分类边界不明确，边界附近的样本往往会产生较大的损失。为了更好地利用这些边界样本来训练网络，本章首先根据参数μ将样本分为正样本和负样本。随后，应用样本加权函数 SlideLoss 以增强模型训练时对边界样本的关注度。

$$f(x)=\begin{cases}1 & x\leqslant\mu-0.1\\ e^{1-\mu} & \mu<x<\mu-0.1\\ e^{1-x} & x\geqslant\mu\end{cases} \tag{6-1}$$

$$EMA_t = d \times \left[1 - \exp\left(-\frac{t}{\text{tau}}\right)\right] \times EMA_{t-1} + (1-d) \times \text{ auto\_iou} \quad (6\text{-}2)$$

$$f(x) = \begin{cases} 1 & x \leqslant EMA_t - 0.1 \\ e^{1-EMA_t} & EMA_t < x < EMA_t - 0.1 \\ e^{1-x} & x \geqslant EMA_t \end{cases} \quad (6\text{-}3)$$

但是，样本加权函数 Slide Loss 并未实现真正的滑动机制，因此本章引入指数移动平均（Exponential Moving Average, EMA）的思想，以优化均值μ的计算过程。通过动态调整指数移动平均值，并对损失函数进行自适应优化。如式（6-2）所示，其中 auto_iou 表示时间步的观测值，d 为衰减因子，t 为当前迭代次数，tau 为训练总迭代次数。当 t=tau 时，衰减因子 d 逐渐增加至e^{e-1}。这种设计使得在训练初期，模型更倾向于依赖历史数据（即 EMA），而在训练后期，模型则更加关注当前观测值[79]。最终的损失函数计算如式（6-3）所示。

尽管本研究所使用的数据集已具备较高的广泛性和丰富性，但由于负样本数量相对较少，仅依靠这些数据训练模型仍存在不足。为此，本章引入了一种改进的损失函数——EMA-Slide Loss，有效解决了样本不平衡问题，并显著增强了模型的鲁棒性。实验结果表明，EMA-Slide Loss 在提升模型性能的同时，能够更好地适应复杂场景下的训练需求。

6.2 实验结果与分析

6.2.1 消融实验

为了系统评估 STSIEP、RCS-OSA 和 EMA-Slide Loss 模块对模型性能的贡献，我们进行了全面的消融实验。实验设计旨在分析这些模块对$map_{0.5:0.95}$、计算复杂度和参数量的影响，并将结果汇总于表 6-1 中。实验 1 至实验 4 分别探讨了各个模块对模型性能的影响。在实验 1 中，采用 YOLOv10 作为基准模型，模型的$map_{0.5:0.95}$为 58.1%，计算复杂度为 6.5 Gflops，参数量为 2.3M。在实验 2 中，引入 STSIEP 模块后，模型精度显著提升至 60.4%，但计算复杂度和参数量分别增加至 12.4 Gflops 和 3.1M，表明该模块在提升性能的同时增加了计算负担。在实验 3 中，仅使用 RCS-OSA 模块时，模型的$map_{0.5:0.95}$为 59.3%，计算复杂度和参数量分别增加至 15.3 Gflops 和 4.1M。实验 4 评估了将基准模型的损失函数替换

为 EMA-Slide Loss 对模型性能的影响，从表格第四行可以看出相较于基准模型，$map_{0.5:0.95}$提升至 58.6%。综上所述，本文介绍的三种方法都有效提高了 YOLOv8 模型对于防毒面具佩戴检测效果。

实验 5、实验 6 主要研究了不同方法组合对模型性能的影响。在实验 5 中，本研究测试了同时使用 STSIEP 和 RCS-OSA 模块时，$map_{0.5:0.95}$进一步提升至 60.7%，计算复杂度和参数量分别为 15.4 Gflops 和 4.2M，表明这两个模块具有协同效应。实验 6 同时使用 RCS-OSA 和 EMA-Slide Loss 模块时，模型的$map_{0.5:0.95}$为 59.8%。

最后，在实验 7 中，本章对所提出的方法进行了全面测试。通过融合所有改进策略，模型的精度得到了显著提升。与基准模型相比，$map_{0.5:0.95}$指标上提高了 3.3%，达到了 61.4%。这些实验结果充分验证了所提方法在提高防毒面具检测性能方面的有效性。

表 6-1 消融实验表

Table 6-1 Ablation experiment

序号	STSIEP	RCS-OSA	EMA-Slide Loss	$map_{0.5:0.95}(\%)$	Gflops	Params(M)
1				58.1	6.5	2.3
2	√			60.4	12.4	3.1
3		√		59.3	15.3	4.1
4			√	58.6	6.5	2.3
5	√	√		60.7	15.4	4.2
6		√	√	59.8	15.3	4.1
7	√	√	√	61.4	15.4	4.2

6.2.2 对比实验

为验证 STSIEP 改进特征提取模块的有效性，现将其他改进 FPN 的模块引入到 YOLOv10 基线模型中，进行对比实验，实验结果如表 6-2 所示。传统的增强检测小目标效果的方法是在 P3 层之前增加一个额外的小目标检测头 SODH，如第四章和第五章本工作一样，但是从实验 1 和实验 2 的对比中可以看出，本章提出的 STSIEP 仅仅比 SODH 高出了 0.3 的参数量，但是$map_{0.5:0.95}$提升了 1.8%，

且计算量也降低了2.7G，因此本章提出的STSIEP具有对小目标更强的检测效果。相较于CGRFPN[80]，从实验3中可以看出，虽然STSIEP的计算量和参数量都更大，但明显$map_{0.5:0.95}$高于CGRFPN2.0%。

表6-2 STSIEP模块对比实验表

Table 6-2 Comparative Experiment Table of the STSIEP Module

序号	STSIEP	SODH	CGRFPN	$map_{0.5:0.95}$（%）	Gflops	Params(M)
1	√			60.4	12.4	3.1
2		√		58.6	15.7	2.8
3			√	58.4	6.8	2.7

为验证RCS-OSA替换C2f的有效性，现将其他替换C2f的模块引入到YOLOv10基线模型中，进行对比实验，实验结果如表6-3所示。RCS-OSA相较于MogaBlock[81]和WTConv[82]，$map_{0.5:0.95}$都有显著优势，分别高出0.9%和1.4%，且参数量提升不大。

表6-3 RCS-OSA模块对比实验表

Table 6-3 Comparative Experiment Table of the RCS-OSA Module

序号	RCS-OSA	MogaBlock	WTConv	$map_{0.5:0.95}$（%）	Gflops	Params(M)
1	√			59.3	15.3	4.1
2		√		58.4	7.6	5.4
3			√	57.9	5.8	2.1

接下来将本文提出的YOLOv10-SRE模型与当前主流目标检测模型进行了全面的性能对比，实验结果如表6-4所示。首先对比经典的两阶段算法Faster R-CNN、Cascade R-CNN，之后对比单阶段目标检测算法RetinaNet，还有前两章本文提出的YOLOv5-SCV和YOLOv8-LSE。实验从检测精度、计算效率、推理速度和模型复杂度四个维度进行了详细分析，以验证YOLOv10-SRE的综合性能。

YOLOv10-SRE相较于常见的两阶段检测算法Faster R-CNN、Cascade R-CNN和单阶段检测算法RetinaNet、和YOLOv11和HYPER-YOLO[83]，$map_{0.5:0.95}$是最高的，达到了61.4%，$map_{0.5}$达到了88.0%，同Cascade R-CNN并列最高，但是计算量和参数量都显著优于Cascade R-CNN。

相较于上一章的改进 YOLOv8 算法，本章改进的 YOLOv10 算法，由于集成的模块的参数量和计算量稍大，但是检测效果也是最高的，$map_{0.5}$分别高于 YOLOv5-SCV 和 YOLOv8-LSE1.3%和 2.4%，$map_{0.5:0.95}$分别高出 1.4%和 0.9%，并且因为 YOLOv10 模型的 NMS-free 创新，本章提出的 YOLOv10-SRE 的 FPS 是最高的。

综合以上实验结果，YOLOv10-SRE 在检测精度、计算效率、推理速度和模型复杂度之间实现了良好的平衡。其性能提升主要是因为于本章提出的改进策略，包括优化特征提取模块、损失函数改进等方法。这些改进使得模型在保持较低计算成本和参数量的同时，显著提升了检测精度和推理速度。

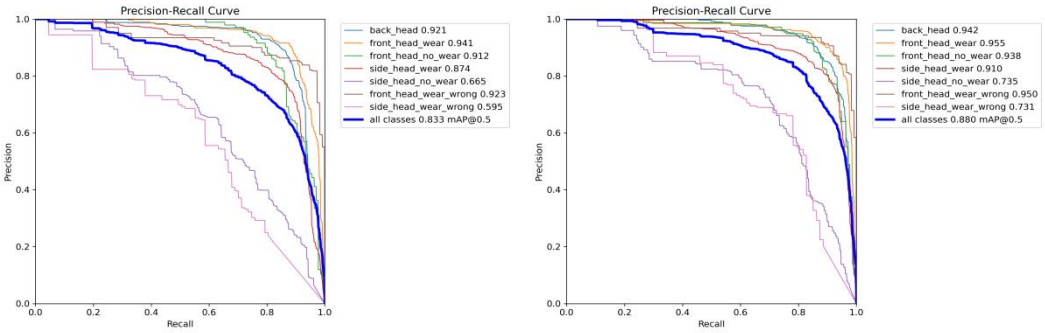
表 6-4 模型性能比较

Table 6-4 Performance Comparison of Models

模型	主干	$map_{0.5}$ (%)	$map_{0.5:0.95}$ (%)	GFLOPs	FPS	Params (M)
Faster R-CNN	ResNet-50-FPN	83.7	55.3	322.34	20	41.15
Cascade R-CNN	ResNet-50-FPN	88.0	58.0	350.35	17	69.17
RetinaNet	ResNet-50-FPN	83.7	54.0	331.00	21	36.23
YOLOv10n	CSP-DarkNet-53	83.2	58.1	6.5	413	2.26
YOLOv11n	CSP-DarkNet-53	84.0	58.5	6.6	384	2.6
HYPER-YOLO	CSP-DarkNet-53	85.6	58.7	13.7	161	3.9
YOLOv5-SCV	CSP-DarkNet-53	86.7	60.0	19.2	196	7.54
YOLOv8-LSE	CSP-DarkNet-53	85.6	60.5	12.5	170	3.20
YOLOv10-SRE	CSP-DarkNet-53	88.0	61.4	15.4	232	4.20

6.2.3 实验结果可视化

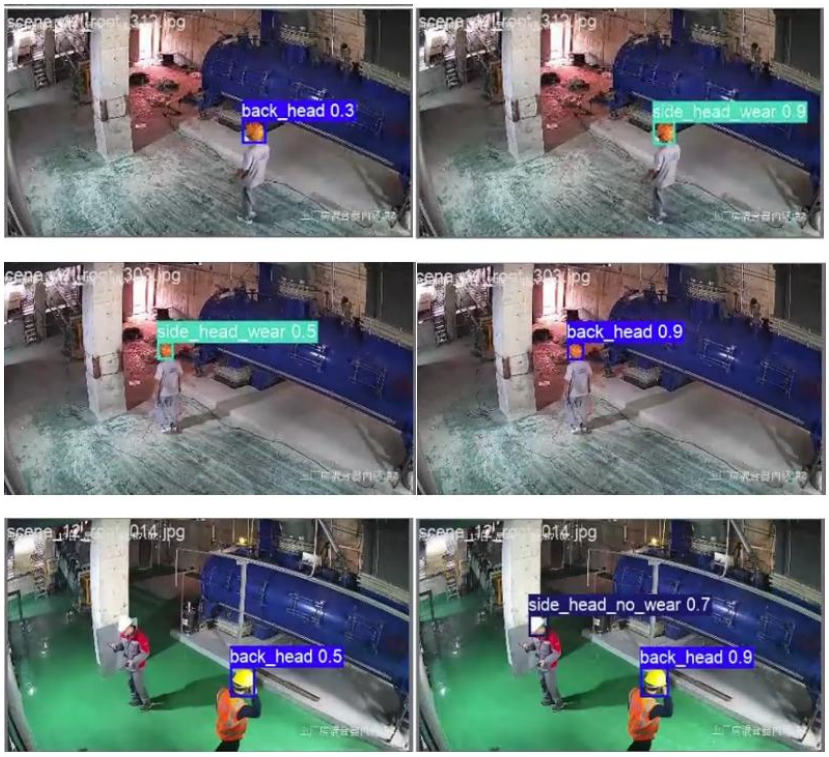
最后，对部分测试的结果进行可视化展示，从图 6-5 中可以看出本文改进的 YOLOv10-SRE 相较于基准模型的 mAP@0.5 提升了 4.7%，所有类别的检测精度都有所提升，其中提升最大的类别为“side_head_wear_wrong”，提升了 13.6%，这是由于更换损失函数为 EMA-SlideLoss，提升了模型对少样本类别的关注度，从而增强了检测效果。



（a）YOLOv10 模型 P-R 图　　（b）YOLOv10-SRE 模型 P-R 图

图 6-5 P-R 图对比

Fig6.5 P-R curve comparison



（a）原模型检测效果图　　（b）SRE-YOLO 检测效果图

图6-6 改进前后的检测效果对比

Fig.6-6 Comparison of detection performance and after improvement

在图 6-6 中，对比了原模型 YOLOv10n（a）和本研究改进的 YOLOv10-SRE 模型（b）的检测效果，第一行到第三行分别是正确佩戴防毒面具侧对镜头、背对镜头、未佩戴防毒面具侧对镜头及背对镜头，可以看出相比较于原 YOLOv10n 模型，SRE-YOLO 对小于 32×32 的待检测小目标的更加敏感，能够更好的兼顾

深层特征的语义信息和浅层特征的细节信息。明显提升了检测精度和检测正确率，同时改善了原模型的漏检问题。

6.3 本章小结

本章针对模型对小目标检测困难、特征信息提取能力差的问题，提出了一种基于 YOLOv10 框架的改进模型。首先设计了小目标语义信息增强金字塔 STSIEP，以增强对小目标的语义信息特征融合效果，其次在 Neck 中使用 RCS-OSA 替换原模型的 C2f，对提取的特征信息进行增强及融合，最后使用指数滑动样本加权函数 EMA-Slide Loss 替换原模型的交叉熵损失函数，解决样本不平衡的问题，增强了模型的鲁棒性。首先，详细阐述了 STSIEP 的设计原理及其对增强多尺度特征的语义融合能力。其次，介绍了 RCS-OSA 替换原模型的 C2f 的优势。此外，采用 EMA-Slide Loss 替代传统损失函数，以解决类别不均衡问题并提高模型对难样本的识别能力。为验证改进策略的有效性，本章设计并实施了消融实验，实验结果证实了三种改进方法在提升模型性能方面的显著贡献。实验结果表明，改进后的模型在检测能力上取得了显著提升，能够更好地适应复杂多变的检测场景。最后，通过与其他主流模型的对比实验，进一步验证了改进模型在检测精度和推理速度上的优越性，充分满足了实际应用的需求。

第 7 章 总结与展望

7.1 工作总结

在化工生产环境中，防毒面具作为关键的个人防护装备，能够有效滤除有害气体、粉尘及颗粒物，从而保障作业人员的呼吸系统安全。然而，由于操作规范执行不力或其他因素，部分工人未能严格按照要求佩戴防毒面具，导致一系列呼吸道损伤及中毒等安全事故的发生。为此，开发一套实时检测防毒面具佩戴状态的系统显得尤为重要。该系统依托摄像头采集图像数据，并结合先进的算法对作业人员面部区域进行实时分析，以判定其是否规范佩戴防毒面具。当检测到未按规定佩戴的情况时，系统将即时触发警报机制，提醒相关人员及时纠正行为，从而有效预防潜在的安全隐患。此外，该监测系统还具有提升作业人员安全意识的辅助功能。通过实时监控与预警反馈，工人能够更加深刻地认识到规范佩戴防毒面具的必要性，进而增强遵守安全操作规程的自觉性。本研究采用基于深度学习的目标检测技术，针对工业场景中防毒面具佩戴检测任务所面临的类别分布不均衡、目标多尺度变化以及复杂背景干扰等挑战，对模型进行了针对性优化与改进。本文的主要贡献包括以下几个方面：

（1）为构建高性能的深度学习目标检测模型，本研究深入工业现场采集了大量作业场景的视频素材。随后，对获取的视频数据进行了预处理，以构建初始数据集。对原始数据集进行了精细化标注，后进一步对数据集进行了统计分析，初步揭示了任务中潜在的技术难点，例如目标尺度的显著差异、形态的多样性以及背景环境的复杂干扰等。

（2）针对防毒面具检测场景中存在的复杂背景干扰和样本不均衡的问题，提出了一种基于 YOLOv5 的改进模型。首先，在原有三检测层的基础上，引入了一个小目标检测头 SODH，以增强模型对小尺度目标的检测能力；其次，在模型中嵌入了 CBAM 注意力机制模块，通过有效区分目标与背景信息，提升检测精度，同时避免了计算复杂度的显著增加；最后，将原模型的二元交叉熵损失函数替换为 Varifocal Loss，以确保模型在训练过程中更加关注高质量的正样本，从而优化检测性能。相较于基准模型，改进后的 YOLOv5 模型$map_{0.5:0.95}$指标上提高了 7.9%。

（3）针对网络提取到的不同特征层信息融合效果差和存在多尺度目标检测的问题，提出一种基于 YOLOv8 的改进方法。首先，在空间金字塔池化融合模块 SPPF 中引入了大核分离卷积注意力模块 LSKA，以增强不同特征层之间的语义信息融合能力；其次，将主干网络中 P2 层的输出集成到颈部结构，并额外增加了一个小目标检测头 SODH，从而提升模型对小尺度目标的检测性能；最后，在检测头部引入了高效多尺度注意力模块 EMA，能够有效学习待检测目标的多尺度语义信息。相较于基准模型，改进后的 YOLOv8 模型$map_{0.5:0.95}$指标上提高了 3.5%。

（4）针对模型对小目标检测困难、特征信息提取能力差的问题，提出基于 YOLOv10 的改进模型 SRE-YOLO，首先设计了小目标语义信息增强金字塔 STSIEP，以增强对小目标的语义信息特征融合效果，其次在 Neck 中使用 RCS-OSA 替换原模型的 C2f，对提取的特征信息进行增强及融合，最后，采用 EMA-slide Loss 替换原模型损失函数，以增强类别分类能力，提高训练的稳定性。相较于基准模型，改进后的 YOLOv10 模型$map_{0.5:0.95}$指标上提高了 3.3%。

7.2 未来展望

在后续研究中，本研究计划进一步探索基于深度学习的目标检测算法，旨在显著提升防毒面具佩戴检测任务的精度与鲁棒性。具体研究方向包括但不限于以下几个方面：

（1）数据集优化：当前研究中所采用的数据集可能存在标注误差或类别分布不均衡等问题，这些问题将对模型的性能产生显著影响。在后续工作中，本研究将致力于提升数据集的质量，以增强模型的泛化能力。具体措施包括实施数据清洗流程以剔除噪声样本，以及通过标注修正机制消除标注偏差。此外，还将引入数据平衡技术以缓解类别不均衡问题。通过这些优化手段，确保模型能够从训练数据中更有效地提取鲁棒的特征表示和学习泛化模式，从而提升其在真实场景中的检测性能。

（2）尽管本研究在防毒面具佩戴检测任务中采用了 YOLO 系列，并对其中的 YOLOv5、YOLOv8 和 YOLOv10 进行改进，但现有方法仍存在一定的优化潜力。在后续研究中，将重点探索更先进的深度学习框架与技术，例如更先进的 YOLO 版本和研究人员提出的更高效的目标检测模块。通过引入这些前沿技术，

旨在进一步提升模型的检测精度、鲁棒性及泛化能力，以应对复杂工业场景中的多样化挑战。

（3）实现目标追踪：在工业场景的实际应用中，由于工作人员进行走动会背对摄像头，无法检测是否规范佩戴防毒面具，因此只进行工作人员的防毒面具佩戴检测任务是不够的，后续希望可以对工作人员进行目标追踪，通过改进追踪算法对背对摄像头的工作人员进行追踪检测，满足更多的任务需求，具有更高的应用价值。

参考文献

[1] 吴迪.基于计算机视觉的施工人员安全状态监测技术研究[D].哈尔滨工业大学,2019.DOI:10.27061/d.cnki.ghgdu.2019.004172.

[2] 赵云.基于改进 YOLOv5 的安全帽佩戴检测系统的设计与实现[D].成都理工大学,2023.

[3] 王晨.面向作业安防的智能检测技术研究与实现[D].西安工业大学,2022.DOI:10.27391/d.cnki.gxagu.2022.000454.

[4] 赵月爱,郝慧琦,王玲.融合自注意机制的矿工行为识别算法[J/OL].计算机技术与发展,1-9[2024-12-31].https://doi.org/10.20165/j.cnki.ISSN1673-629X.2024.0364.

[5] 肖赣涛.基于深度学习的复杂场景抽烟行为检测研究与实现[D].西南交通大学,2021.DOI:10.27414/d.cnki.gxnju.2021.002494.

[6] Barling J E, Frone M R. The psychology of workplace safety[M].Washington: American Psychological Association, 2004.

[7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.

[8] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[9] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 IEEE conference on computer vision and pattern recognition. Ieee, 2008: 1-8.

[10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[12] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[14] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[15] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[16] 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.

[17] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[18] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(06): 1137-1149.

[19] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[20] Cai Z, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(5): 1483-1498.

[21] Zhang H, Chang H, Ma B, et al. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training[C]//European Conference on Computer Vision. 2020: 260-275.

[22] Sun P, Zhang R, Jiang Y, et al. Sparse r-cnn: End-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14454-14463.

[23] 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. 2016: 779-788.

[24] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016, 9905:21-27.

[25] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[26] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[27] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[28] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[29] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.

[30] 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.

[31] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[32] Jocher G, Chaurasia A, Stoken A, et al. ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations[J]. Zenodo, 2022.

[33] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[34] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework

for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[35] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.

[36] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

[37] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision. Springer, Cham, 2025: 1-21.

[38] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.

[39] Tang X, Du D K, He Z, et al. Pyramidbox: A context-assisted single shot face detector[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 797-813.

[40] Qi D, Tan W, Yao Q, et al. YOLO5Face: why reinventing a face detector[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 228-244.

[41] Yu Z, Huang H, Chen W, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. arXiv preprint arXiv:2208.02019 (2022).

[42] 张子振,南钢洋,孟凡超,等.一种改进型级联神经网络检测算法及加速处理[J].计算机仿真,2024,41(02):255-260+316.

[43] 陈亮,王璇,雷坤.复杂场景下跨层多尺度特征融合安全帽佩戴检测算法[J/OL].计算机应用,1-11[2024-12-31].

[44] Li Z, Xie W, Zhang L, et al. Toward efficient safety helmet detection based on YoloV5 with hierarchical positive sample selection and box density filtering[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.

[45] 祁泽政,徐银霞.改进 YOLOv5s 算法的安全帽佩戴检测研究[J].计算机工程与应用,2023,59(14):176-183.

[46]李毅,徐慧英,朱信忠,等.基于 YOLOv5n 模型改进的口罩检测算法：Mask-YOLO[J/OL].计算机工程,1-15[2024-12-31].

[47]李泽琛,李恒超,胡文帅等.多尺度注意力学习的 Faster R-CNN 口罩人脸检测模型[J].西南交通大学学报,2021,56(05):1002-1010.

[48]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30(): 5998-6008.

[49]易琴.基于分布式学习的 MIMO 系统研究[D].电子科技大学,2023.DOI:10.27005/d.cnki.gdzku.2023.000956.

[50]史墨可.激光立体成形裂纹缺陷预测及工艺参数优化[D].河南科技大学,2023.DOI:10.27115/d.cnki.glygc.2023.000254.

[51]Yin X, Goudriaan J, Lantinga E A, et al. A flexible sigmoid function of determinate growth[J]. Annals of botany, 2003, 91(3): 361-371.

[52]Fan E. Extended tanh-function method and its applications to nonlinear equations[J]. Physics Letters A, 2000, 277(4-5): 212-218.

[53]Misra D. Mish: A self regularized non-monotonic activation function[J]. arXiv preprint arXiv:1908.08681, 2019: 1-14.

[54]He J, Li L, Xu J, et al. ReLU deep neural networks and linear finite elements[J]. arXiv preprint arXiv:1807.03973, 2018: 1-50.

[55]Xu J, Li Z, Du B, et al. Reluplex made more practical: Leaky ReLU[C]. 2020 IEEE Symposium on Computers and communications (ISCC), 2020: 1-7.

[56]Jocher G, Stoken A, Borovec J, et al. ultralytics/yolov5: v4. 0-nn. SiLU () activations, Weights & Biases logging, PyTorch Hub integration[J]. Zenodo, 2021.

[57]付琪.基于深度学习的非接触式掌纹识别方法研究[D].沈阳工业大学,2023.DOI:10.27322/d.cnki.gsgyu.2023.000662.

[58]王文浩.基于改进 YOLOv8 的水面漂浮物检测模型[D].南昌大学,2024.DOI:10.27232/d.cnki.gnchu.2024.002316.

[59]马冬梅,郭智浩,罗晓芸.改进 YOLOv5s-Seg 的高效实时实例分割模型[J].计算机工程与应用,2024,60(16):258-268.

[60]Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection

network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.

[61] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 658-666.

[62] 陈好男.基于YOLO的密集小目标检测算法研究[D].齐鲁工业大学,2024.DOI:10.27278/d.cnki.gsdqc.2024.000042.

[63] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[64] Oksuz K, Cam B C, Kalkan S, et al. Imbalance problems in object detection: a review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3388–3415.

[65] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[66] Li B, Liu Y, Wang X. Gradient harmonized single-stage detector[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 8577-8584.

[67] Zhang H, Wang Y, Dayoub F, et al. Varifocalnet: An iou-aware dense object detector[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8514-8523.

[68] Li X, Wang W, Wu L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020, 33: 21002-21012.

[69] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[70] Zhang H, Wang Y, Dayoub F, et al. Varifocalnet: An iou-aware dense object detector[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8514-8523.

[71] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: Rethinking the large kernel attention design in cnn[J]. Expert Systems with Applications, 2024, 236: 121352

[72] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, 2023: 1-5.

[73] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[74] 陈洋,张欣,陈孝玉龙,等.CA-MobileNet V2:轻量化的作物病害识别模型[J].计算机工程与设计,2024,45(02):484-490.DOI:10.16208/j.issn1000-7024.2024.02.021.

[75] 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. 2021: 13713-13722.

[76] 王召龙,张洁.基于改进的YOLOv8轻量级火灾检测算法研究[J].计算机技术与发展,2024,34(10):61-68.DOI:10.20165/j.cnki.ISSN1673-629X.2024.0206.

[77] Cui, Y.; Ren, W.; Knoll, A. Omni-Kernel Network for Image Restoration. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024 ; Volume 38, pp. 1426–1434.

[78] Kang M, Ting C M, Ting F F, et al. RCS-YOLO: A fast and high-accuracy object detector for brain tumor detection[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 600-610.

[79] 刘伟宏,李敏,朱萍,等.基于 YOLOv8n 改进的织物疵点检测算法[J].棉纺织技术,2024,52(10):19-25.

[80] Ni Z, Chen X, Zhai Y, et al. Context-Guided Spatial Feature Reconstruction for Efficient Semantic Segmentation[J]. arXiv preprint arXiv:2405.06228, 2024.

[81] Li S, Wang Z, Liu Z, et al. Moganet: Multi-order gated aggregation network[C]//The Twelfth International Conference on Learning Representations. 2023.

[82] Finder S E, Amoyal R, Treister E, et al. Wavelet convolutions for large receptive fields[C]//European Conference on Computer Vision. Springer, Cham, 2025: 363-380.

[83] Feng Y, Huang J, Du S, et al. Hyper-yolo: When visual object detection meets hypergraph computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

攻读学位期间发表的学术论文目录

[1] YOLOv8-LPE——面向工业场景的防毒面具佩戴检测方法[J/OL].青岛理工大学学报(已录用，未发表)

[2] YOLOv5s-PCV——防毒面具佩戴检测方法[J/OL].佳木斯大学学报(已录用，未发表)

攻读学位期间参与的科研项目

[1]安徽未来技术研究院企业合作项目，面向工业安全的工业互联网个性化 AI 研究与应用(2023gyh210).

[2]安徽高科电子股份有限公司企业合作项目，基于机器视觉的焊缝坡口自动识别定位跟踪与自动焊接技术的研究与开发(HX-2024-08-003).

攻读学位期间参与的学科竞赛

[1]2023 年安徽省大学生网络与分布式系统创新设计大赛，研究生组，二等奖.

[2]第九届安徽省“互联网+”大学生创新创业大赛《去危就安——面向工业安全生产的防护设备佩戴 AI 检测》，铜奖.

[3]第九届安徽省“互联网+”大学生创新创业大赛《基于深度学习的汽车轮胎参数云边协同检测系统》，铜奖.

[4]第九届安徽省“互联网+”大学生创新创业大赛《MVIC——面向医疗产业的 3D 人脸重建及变形技术》，铜奖.

致谢

行文至此，落笔为终。回望求学之路，感慨万千。在这段充满挑战与成长的旅程中，无数人曾予我温暖与力量，谨以此篇致谢，向所有给予我支持的人致以最真挚的谢意。

首先，我要特别感谢我的研究生导师，感谢他在过去几年中对我的悉心指导与关怀。在整段研究生生涯里，汪教授不仅为我提供了宝贵的学术资源和研究方向，而且在我遇到困难时，总是耐心地为我解答问题，指引我走出困境。同时，汪教授严谨的治学态度、务实的工作作风和宽厚的人格魅力，都深深地影响着我，让我受益终身。

其次，我要感谢我的同学们和实验室的伙伴们。在研究过程中，大家的讨论和分享让我对问题有了更为广泛的思考视角。在团队合作中，我学到了很多，不仅是学术上的帮助，更是在人际交往和团队协作方面的成长。大家互相支持、互相鼓励，成为我在研究生涯中不可或缺的一部分。

最后，感谢我的家人。在我求学的道路上，他们一直给予我无尽的支持与理解，特别是在我遇到困难时，他们总是我的坚强后盾。是你们的默默付出和无条件的爱，成就了我今天的每一步。

在此，我再次向所有帮助过我的人表示衷心的感谢。未来的日子里，我将继续努力，不负众望。