分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220320119

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于改进深度学习模型的安全帽佩戴检测技术研究

论文作者　方楠
校内导师　王凤随 教授
校外导师　曾红军 工程师

专业学位类别　电子信息
研究方向（领域）　目标检测

论文提交日期：2025 年 5 月 15 日

分类号：TP391　　单位代码：10363

密 级：公 开　　学 号:2220320119

题　目 基于改进深度学习模型的安全帽佩戴检测技术研究

题　目 Research on Safety Helmet Wearing Detection Technology Based on Improved Deep Learning Model

学　生　姓　名：方楠

校　内　导　师：王凤随（教授）

校　外　导　师：曾红军（工程师）

专　　业：电子信息

研　究　方　向：目标检测

论文答辩日期：2025 年 5 月 12 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，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：方楠

日期：2025 年 5 月 12 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 ☑。

学位论文作者签名：方楠　　指导教师签名：王凤随

日期：2025 年 5 月 12 日　　日期：2025 年 5 月 12 日

基于改进深度学习模型的安全帽佩戴检测技术研究

摘 要

近年来，施工现场安全事故频发，很大一部分原因是工人未佩戴安全帽。因此，研究施工现场安全帽佩戴检测的方法，对降低事故发生率、保障工人生命安全具有重要意义。传统的人工监管方法成本较高且精度较低，而基于深度学习的目标检测技术则为安全帽佩戴检测提供了新途径。然而，由于施工现场环境的复杂，图像中安全帽的纹理和颜色特征易产生非线性失真，模型难以准确提取头部区域的有效信息。此外，安全帽在图像中占比较小，卷积神经网络通过逐层下采样扩大感受野的同时，微小目标的空间信息在高层特征中被稀释，其边缘纹理等关键细节难以被有效表征。因此，为解决安全帽佩戴检测技术中存在的检测精度低、漏检误检以及对小目标检测效果不理想的问题，论文提出了基于改进深度学习模型的安全帽佩戴检测技术研究。论文以无锚框网络 CenterNet、一阶段 YOLO 系列网络作为基线模型，通过引入注意力机制和替换损失函数等策略提高对关键特征的关注、优化特征信息的传播，实现模型检测性能的提升。其主要研究内容如下：

（1）针对安全帽佩戴检测算法缺乏对目标上下文信息的关注，存在模型检测精度低且参数量过多的问题，提出一种基于改进 CenterNet 模型的检测算法。首先，设计一种采用部分卷积和级联融

合策略的轻量化主干网络(Partial Convolution and Cascade Fusion Lightweight Backbone Network, PCCFNet)，实现更高的检测精度和更少的参数量；其次，在主干网络中引入 CBAM(Convolutional Block Attention Module)，提高模型的特征提取能力，抑制冗余信息的传播；最后，对热力图损失函数进行改进，提高对正样本的关注，降低对负样本的关注。实验结果表明，改进后算法的参数量为 15.89M，在三个公开的安全帽佩戴检测数据集 Safety Helmet Wearing Dataset(SHWD)、Safety Helmet Detection Dataset(SHDD)和 GDUT-Hardhat Wearing Detection(GDUT-HWD)上的平均精度均值(mean Average Precision, mAP)分别达到了 93.05%、92.72%和 90.65%。相较于基线算法，改进后算法的参数量减少了 16.77M，mAP 分别提高了 3.37%、2.75%和 7.03%，证明了所提算法的有效性。

（2）针对安全帽佩戴检测算法在复杂施工环境下易出现漏检误检的问题，提出一种基于改进 YOLOv5s 模型的检测算法。首先，设计一种引入注意力机制的特征融合模块 Star-CAA Block 代替原算法主干网络中的部分 C3 模块，自适应聚焦不同层级的语义信息；其次，将损失函数替换为 MPDIoU Loss，在处理安全帽倾斜、侧视等非轴向对齐情况时，通过顶点距离的对称性减少方向误判。同时，对公开数据集 SHWD 进行预处理，筛选和补充更多复杂施工现场的图像素材，确保重构后的数据集 R-SHWD 能够更准确地模拟实际施工现场的复杂性和挑战性。实验结果表明，改进后算法的参数量为

8.08M，在数据集 R-SHWD 上的 mAP 达到了 87.00%。相较于基线算法，mAP 提高了 2.30%，同时对识别结果进行可视化分析，证明了所提算法的有效性。

（3）针对安全帽佩戴检测算法对小目标检测效果不理想的问题，提出一种基于改进 YOLOv7-tiny 模型的检测算法。首先，在主干网络中引入近乎无参的 Triplet Attention，在不显著增加计算开销的前提下，精准捕获小目标的细粒度特征，强化模型对小目标的定位能力；其次，利用 K-means 聚类算法通过无监督的方式自动分析数据集中安全帽的尺寸分布，生成与真实目标尺度匹配的锚框，显著提升小目标先验框的适配性，增强模型对小目标的初始定位能力；最后，将损失函数替换为 SIoU Loss，通过引入角度损失项，将边界框回归的方向作为优化目标，有效解决了小目标因位置偏移导致的定位模糊问题。实验结果表明，改进后算法的参数量为 6.01M，在数据集 R-SHWD 上的 mAP 达到了 89.40%。相较于基线算法，在不额外增加参数量的情况下提高了 3.30%，并通过分析图像的可视化结果，证明了所提算法的有效性。

关键词：目标检测，深度学习，安全帽佩戴检测，特征融合，注意力机制，损失函数

RESEARCH ON SAFETY HELMET WEARING DETECTION TECHNOLOGY BASED ON IMPROVED DEEP LEARNING MODEL

ABSTRACT

In recent years, construction site safety accidents occur frequently, and a large part of the reason is that workers do not wear safety helmets. Therefore, it is of great significance to study the method of safety helmet wearing detection at construction sites to reduce the accident rate and protect the life safety of workers. The traditional manual supervision methods have high cost and low accuracy, while deep learning-based target detection technology provides a new way for safety helmet wearing detection. However, due to the complexity of the construction site environment, the texture and color features of the safety helmet in the image are prone to nonlinear distortion, and it is difficult for the model to accurately extract the effective information of the head region. In addition, the safety helmet occupies a relatively small area in the image, and while the convolutional neural network expands the receptive field through layer-by-layer downsampling, the spatial information of the tiny target is diluted in the high-level features, and the key details such as its edge texture are difficult to be effectively characterized. Therefore, in order to solve the problems of low detection accuracy, missed detection and misdetection, as well as unsatisfactory detection effect on small targets in safety helmet wearing detection technology, the paper proposes a research on safety helmet wearing detection technology based on improved deep learning

model. The paper using the anchorless frame network CenterNet and one-stage YOLO series network as the baseline model, and improving the attention to key features and optimizing the propagation of feature information by introducing the attention mechanism and replacing the loss function and other strategies, so as to realize the enhancement of the model's detection performance. Its main research:

(1) Aiming at the safety helmet wearing detection algorithm's lack of attention to the target's context information and the problems of low model detection accuracy and excessive number of parameters, we propose a detection algorithm based on the improved CenterNet model. Firstly, a Partial Convolution and Cascade Fusion Lightweight Backbone Network (PCCFNet) is designed to achieve higher detection accuracy and less number of parameters; secondly, a CBAM (Convolutional Block Attention Module) in the backbone network to improve the feature extraction ability of the model and inhibit the propagation of redundant information; finally, the loss function of the heat map is improved to increase the attention to positive samples and decrease the attention to negative samples. The experimental results show that the improved algorithm has a parameter count of 15.89M on three publicly available safety helmet wearing detection datasets SHWD (Safety Helmet Wearing Dataset), SHDD (Safety Helmet Detection Dataset) and GDUT-HWD (GDUT-Hardhat Wearing Detection) on which the mean Average Precision (mAP) reaches 93.05%, 92.72% and 90.65%, respectively. Compared to the baseline algorithm, the improved algorithm has 16.77M fewer parameters and 3.37%, 2.75% and 7.03% higher mAP, respectively, which proves the effectiveness of the proposed algorithm.

(2) Aiming at the problem that safety helmet wearing detection algorithm is prone to miss detection and misdetection in complex

construction environments, a detection algorithm based on the improved YOLOv5s model is proposed. Firstly, a feature fusion module Star-CAA Block, which introduces an attention mechanism, is designed to replace part of the C3 module in the backbone network of the original algorithm, to adaptively focus on the semantic information at different levels; secondly, the loss function is replaced with the MPDIoU Loss, which reduces the directional misjudgement by the symmetry of the vertex distances when dealing with the non-axial alignment cases such as tilted safety helmets and side-views safety helmets. Meanwhile, the publicly available dataset SHWD is preprocessed to screen and supplement more image materials of complex construction sites to ensure that the reconstructed dataset R-SHWD can more accurately simulate the complexity and challenge of actual construction sites. The experimental results show that the improved algorithm has a parameter count of 8.08M and a mAP of 87.00% on the dataset R-SHWD. Compared to the baseline algorithm, the mAP is increased by 2.30%, while visualization and analysis of recognition results, proving the effectiveness of the proposed algorithm.

(3) Aiming at the problem that the safety helmet wearing detection algorithm is not ideal for small target detection, a detection algorithm based on the improved YOLOv7-tiny model is proposed. Firstly, the near-parameterless Triplet Attention is introduced into the backbone network to accurately capture the fine-grained features of small targets and strengthen the model's ability to locate small targets without significantly increasing the computational overhead; secondly, the K-means clustering algorithm is used to automatically analyze the size distribution of the safety helmets in the dataset by unsupervised means to generate an anchor that matches the scale of the real target frame, which significantly improves the fitness of the a priori frame of the small target and enhances the initial localization

ability of the model to the small target; finally, the loss function is replaced with SIoU Loss, and the direction of the bounding box regression is taken as the optimization objective by introducing the angular loss term, which effectively solves the localization ambiguity problem of the small target due to the positional offset. The experimental results show that the number of parameters of the improved algorithm is 6.01M, and the mAP on the dataset R-SHWD reaches 89.40%. Compared with the baseline algorithm, it is improved by 3.30% without additional increase in the number of parameters, and the effectiveness of the proposed algorithm is demonstrated by analyzing the visualization results of the images.

KEY WORDS: object detection, deep learning, safety helmet wearing detection, feature fusion, attention mechanism, loss function

目录

第 1 章 绪论

1.1 研究背景与意义

随着国民经济的持续增长，我国的城市化进程显著提速，居民生活质量获得极大改善。在此进程中，建筑产业规模呈现持续扩张态势。根据国家统计局 2023 年统计公报数据，我国建筑业全年完成总产值 315911.85 亿元，占全国 GDP 的 24.41%，建筑业从业人员达到 5253.75 万人，占全国劳动力的 6.81%。从 2014 年至 2023 年，十年间建筑业生产总值从 176713.40 亿元增长到 315911.85 亿元，建筑业从业人员从 4960.58 万人增长到 5253.75 万人。然而，因施工现场环境相对复杂、人员频繁流动，常给建筑业工作者带来众多安全威胁，使得建筑业成为事故多发且安全形势严峻的行业之一[1]。这些安全事故不仅可能导致工程进度受阻，更值得关注的是其对施工人员生命安全构成的重大威胁。鉴于此，该问题已成为工程管理领域亟待解决的关键难题。据数据显示，全球建筑业年均发生安全生产事故约 6 万起，而我国近五年年均发生施工安全事故约 2000 余起[2]。高处坠落和物体打击是引发安全事故的主要诱因，其中颅脑损伤占比达 80%[3]，因未佩戴安全帽而导致的致死案例占比则超过 50%[4]。此类事故的发生往往伴随着不可逆的严重后果，而安全帽作为重要的头部防护装备，其设计原理在于通过高密度缓冲材料吸收并分散冲击能量，能有效保护人体颅腔等脆弱部位。

传统的安全帽佩戴检测方法通常采用人工巡检的方式。然而，由于施工区域范围较广、作业环境多样、天气条件不可控以及人员流动性强等因素，人工巡检的可靠性受到挑战。此外，人类视觉系统存在生理极限，长时间的连续监测会导致注意力分散和视觉疲劳，难以实现对施工人员安全帽佩戴状态的持续有效监管。由此可见，传统的人工监督方法不仅效率低下且资源消耗较大，更存在显著的安全隐患。目前，人工监督方法在复杂施工环境下存在显著不足，难以满足智能化施工对实时性和精确性的要求。因此，应用计算机视觉技术替代传统的人工监督方法，实现对施工现场作业人员安全帽佩戴状态的实时检测，在工程管理领域具有重要的现实意义和应用价值。

近年来，人工智能技术持续迭代演进，计算机视觉技术作为分支之一，在行人检测[5]、手势识别[6]、自动驾驶[7]等多元化应用场景中发挥着关键作用。目标检测技术则凭借其在复杂场景下的对象识别能力，成为当前智能视觉领域的研究焦点。因此，采用目标检测算法来为安全帽佩戴的实时自动检测提供新的解决方案。当前主流的目标检测算法普遍采用深度学习框架，和传统的人工监督相比，基于深度学习的安全帽佩戴检测算法具有成本低、响应速度快、全天候不间断工作等优势，能够更好地保障施工人员的安全，有效降低施工现场事故的发生率。然而，基于深度学习的安全帽佩戴检测算法仍存在明显的不足之处：首先，施工现场环境相对复杂，多源干扰信息并存的情况下，易引发漏检和误判问题，导致检测性能下降；其次，安全帽这类小尺寸目标由于其物理尺寸和图像分辨率的比例失衡问题，在图像中占据较小的像素区域，导致算法在特征提取过程中面临有效信息不足的挑战，进而导致检测精度受限。因此，论文提出了三种基于改进深度学习模型的安全帽佩戴检测算法，不仅能提高检测效率并降低人力成本，还能有效降低事故的发生率，为施工人员的生命安全提供可靠保障。此外，改进后算法还将为智慧工地建设的全面推进提供关键技术支撑。综上所述，论文提出的基于改进深度学习模型的安全帽佩戴检测算法，不仅为后续的学术探索提供方法论支撑，还在实践领域展现出潜在的应用价值。

1.2 国内外研究现状

1.2.1 目标检测算法研究现状

作为计算机视觉领域的核心研究方向，目标检测技术旨在实现对静态图像及动态视频序列中感兴趣目标的精准定位和分类，目前在人脸识别[8]、医学影像[9]等交叉学科领域展现出显著的应用成效。根据底层技术框架的差异，目标检测算法可分为两类：一类是依赖手工设计特征的传统检测算法，另一类是基于卷积神经网络的检测算法。

在传统检测算法领域内，研究主要聚焦于手工设计的特征提取方法和图像分析技术，致力于实现目标特征的精准定位和识别。预处理作为数据增强的重要环节，通过滤波算法和噪声抑制等技术手段，对原始图像数据进行质量优化，

为后续图像处理提供更优质的输入条件。区域建议是采用滑动窗口对图像进行多尺度扫描，从而生成候选目标区域。特征提取主要借助手工设计的方法，识别并提取和任务相关的特征。区域分类的核心在于构建分类模型，通过多轮迭代动态调整参数配置，从而完成目标识别任务。传统目标检测算法的检测流程如图 1-1 所示。

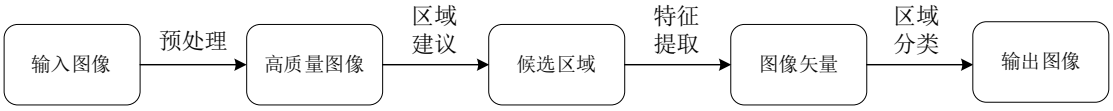


图 1-1 传统目标检测算法检测流程

区域建议用来生成候选区域。传统目标检测算法中，候选区域的生成方法有滑动窗口法[10]、选择性搜索法[11]等。滑动窗口法是最简单直接的候选区域生成方法，即窗口以预设步长在图像平面移动，每次操作均会生成一个特定的特征区域，再确定目标在该区域内的存在概率。虽然滑动窗口法在图像处理领域内具有操作简便的优势，但也存在一个关键局限：窗口的尺寸参数需要研究者根据具体任务自主设定。此外，滑动窗口法需要遍历整个数据窗口，这显著增加了计算复杂度并导致计算效率降低。因此，在有严格实时性要求的应用场景下，滑动窗口法难以满足需求。相较于滑动窗口法的全部穷举方案，选择性搜索法通过优化候选区域筛选机制，可有效排除识别效果不佳的区域。在多尺度目标检测领域，选择性搜索法通过将图像划分为若干个子区域，基于相邻区域的特征相似性计算，采用自底向上的区域合并策略，逐步融合具有相似纹理、颜色或形状特征的相邻区域，实现对图像中多尺度目标的全面遍历。在此基础上，选择性搜索法通过整合色彩空间分析、相似性评估及阈值动态设定等多维策略，有效扩展了候选区域的覆盖范围，确保了检索结果的准确性。

特征提取是提取每个候选区域中的特征。传统的目标检测算法中，经常使用 HOG[12](Histogram of Oriented Gradients)、DPM[13](Deformable Parts Model)等算法作为特征提取器。2005 年，Dalal 等人首次提出 HOG 算法。改进后算法在行人检测任务中展现出显著优势，通过对输入图像采用多尺度缩放来实现多分辨率分析，确保对不同尺寸目标精细化特征的提取。算法通过统计梯度方向分布对目标的边缘信息进行编码，但其特征设计依赖人工经验，导致在复杂场景下的泛化能力不足，难以满足实际工程应用的需求。2008 年，Felzenszwalb 等

人提出了 DPM 算法。针对具有复杂形变特征的目标，该算法展现出显著的处理优势，但存在稳定性不足的挑战，且在复杂场景下的实际应用也受到制约。

分类器的核心功能包括对目标区域和背景的划分及识别，同时完成对潜在候选区域的分类。传统的目标检测算法中，经典的分类器有 SVM(Support Vector Machines)、AdaBoost[14]和 Random Forest[15]等。1995 年，Cortes 等人提出了 SVM 算法。该算法的核心在于构建最优决策边界实现数据分类和借助核函数进行特征空间的非线性变换，从而完成分类回归任务。同年，Freund 等人提出了 AdaBoost 算法。该算法通过集成多组弱分类器构造强分类器模型，弱分类器的权重参数通过动态更新机制实现迭代优化，能够引导后续分类器将判别焦点集中于误判样本子集。值得注意的是，AdaBoost 算法对噪声数据及异常样本具有较高的敏感性。此外，当分类器数量不断增加时，模型面临过拟合风险。2001 年，Breiman 等人提出了 Random Forest 算法。该算法集成多个决策树模型，通过投票机制对预测结果进行加权融合，最终输出决策结果。

传统的目标检测算法虽已取得显著进展，但仍存在许多待解决的关键挑战。例如，区域选择器和特征提取器存在较高程度的冗余。当目标的形态发生变化时，模型所提取的特征和预设特征之间将产生显著差异，这种特征偏差会直接导致模型泛化能力受限。而基于深度卷积神经网络的目标检测算法，通过网络架构实现了底层特征和高层语义信息的有效融合，在特征提取过程中展现出更强的抗干扰能力，能够显著提升模型在复杂场景下的鲁棒性。所以，基于深度学习的目标检测算法成为当前领域内的研究焦点。

2012 年，AlexNet 算法[16]在 ImageNet 视觉识别挑战赛上取得突破性成果，这一里程碑事件推动了深度学习技术的快速发展。该算法通过自动特征提取机制，有效克服了传统方法过度依赖人工经验的局限性，逐渐确立了其在目标检测领域的主导地位。当前基于深度学习框架的目标检测算法可分为两种经典范式：其一是采用级联架构的两阶段检测算法，这类算法通过独立的候选区域生成和特征分类模块完成目标检测任务。其二是端到端的单阶段检测算法，这类算法通过构建统一的特征学习框架，将目标定位和识别任务进行联合优化。目标检测算法的发展脉络如图 1-2 所示。

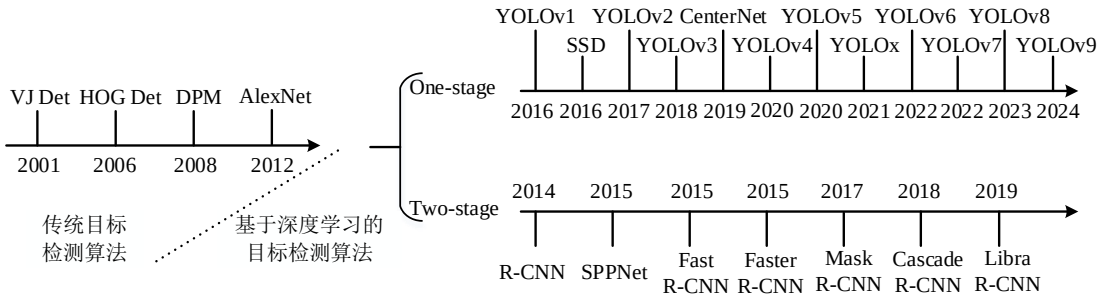


图 1-2 目标检测算法发展脉络

两阶段目标检测算法通过区域生成网络，从输入图像中提取一系列潜在目标区域，再针对这些候选区域进行精细化处理，通过卷积神经网络完成目标类别的语义分类，并同步实现边界框坐标的精确回归。2014 年，Girshick 等人[17]首次提出 R-CNN 算法。该算法采用选择性搜索法生成候选框集合，再通过卷积神经网络对各候选区域实施特征提取，最后借助 SVM 实现目标分类。由于算法生成的候选区域数量可达数千级别，这种高密度的锚框虽然能有效覆盖目标，但同时也会导致计算复杂度的显著增加。2015 年，He 等人[18]则提出了 SPPNet 算法。该算法在 R-CNN 框架上进行优化升级，通过将空间金字塔池化(Spatial Pyramid Pooling，SPP)引入到特征提取阶段，减少了缩放图像对结果的影响。但 SPPNet 算法只对全连接层进行微调，容易导致检测效率降低。同年，Girshick 等人[19]又提出了 Fast R-CNN 算法。该算法以经典的 VGG16[20]作为骨干网络，通过在深层特征图上引入候选区域生成机制，有效共享了卷积层的计算资源，算法性能得到极大改善。2017 年，Ren 等人[21]对 Fast R-CNN 算法进一步改进，提出了 Faster R-CNN 算法。该算法创新性地引入基于区域候选网络的目标候选框生成机制，实现了从图像输入到目标定位的端到端训练过程，使模型的推理速度显著提升。在 Faster R-CNN 基础上，研究人员又推出 Cascade R-CNN[22]、Mask R-CNN[23]、Libra R-CNN[24]等高效算法，不断推动检测精度和速度的提升。尽管技术不断完善，两阶段目标检测算法的计算资源消耗仍相对较大，速度在特定场合需求下依然有待提高。

与两阶段目标检测算法相比，一阶段检测算法采用端到端的直接预测机制，通过优化特征提取和目标定位的联合学习过程，在前向推理过程中同步输出目标类别置信度和边界框坐标参数，有效避免了候选区域生成和后续分类回归的分步处理流程，在保证检测精度的前提下显著提升了推理速度。2016 年，

Redmon 等人[25]提出了 YOLOv1 算法。该算法将特征图离散化为等间距的网格单元，通过回归预测机制估计目标中心点在网格单元内的空间位置分布。同年，Liu 等人[26]在研究一阶段检测算法时提出了 SSD 算法。该算法创新性地提出了一种基于卷积神经网络的多尺度特征处理框架，通过在不同空间分辨率的特征图上实施并行化特征提取操作，有效完成了目标定位和分类任务。随后，为改进 YOLOv1 的性能，Redmon 团队又提出了 YOLOv2 算法[27]和 YOLOv3 算法[28]。其中 YOLOv2 通过引入含有锚框的卷积结构、批归一化等优化方案，有效解决了单阶段检测算法在小目标识别和定位精度方面的局限性；团队提出的 YOLOv3 学习了特征金字塔网络[29](Feature Pyramid Network，FPN)的设计思想，通过三条分支去处理不同尺度的目标，同时将主干替换为深度残差网络 DarkNet53，并引入残差连接解决了梯度消失问题，进一步提升了算法的检测精度。同年，Law 等人[30]提出了具有创新意义的 CornerNet 算法，该算法摒弃了传统的锚框预测方法，转而独立识别物体的左上角和右下角关键点，并使用嵌入式向量进行角点对匹配，预测物体的边界框坐标，将目标检测任务转换为关键点检测任务，为后续研究提供了新的视角。尽管 CornerNet 摆脱了锚框的依赖，但并未考虑物体的关键内部特征，提高了误判风险。此外，它也在处理物体关键点被遮挡时显现出不足。为了克服这些问题，Zhou 等人[31]提出的 CenterNet 算法则关注预测物体的中心点，利用深层特征，通过识别热力图上的局部最大值点简化处理流程，可适用于 2D 目标检测、3D 物体检测和人体姿势估计等任务。然而，CenterNet 在实操中存在下采样导致的中心点重合问题。2020 年，Bochkovskiy 等人[32]通过对 YOLOv3 进行改进提出了 YOLOv4 算法。该算法在模型颈部融合了空间金字塔池化结构 SPP 和路径聚合网络[34](Path Aggregation Network，PANet)。其中 CSP 结构[33]通过特征映射的分层拆分和跨层级融合策略，在显著降低计算复杂度的同时，仍能保持优秀的精度表现。同年 6 月，Glenn 等人[35]提出了 YOLOv5 算法，YOLOv5 是 YOLO 系列的一个延伸。该算法在改进方案中采用了 Mosaic 数据增强和动态图像尺寸调整技术，能够依据目标检测任务的复杂程度对模型规模进行动态优化。2021 年，Ge 等人[36]提出了 YOLOx 算法。通过移除传统锚框机制，构建了无锚框检测体系。2022 年，美团提出了 YOLOv6 算法[37]。该算法的设计相较于理论层面的创新，更着重考虑工业场景

的实际需求。同年，Wang等人[38]提出YOLOv7算法，通过采用基于级联的模型缩放策略提高了算法的鲁棒性和模型的泛化能力，但该算法还存在着检测结果不稳定的缺陷。2023年，Ultralytics公司发布了YOLOv8算法。该算法采用更全面的数据增强技术、动态学习率调整策略以及改进后的CIoU Loss等。2024年2月，Wang等人[39]提出YOLOv9算法，该算法利用信息瓶颈原理，通过实施可编程梯度信息来保留深层网络中的重要数据，确保更可靠的梯度生成，提高了目标检测模型的收敛性和性能。

1.2.2 安全帽佩戴检测算法研究现状

基于计算机视觉的安全帽佩戴检测技术旨在通过图像分析定位施工人员的头部区域，进而识别是否存在未佩戴安全帽的情况。国内外研究人员围绕此技术提出了两类算法：传统检测算法和基于深度学习的检测算法，下文将对相关算法进行介绍。

传统的检测算法主要依赖人工设计的特征提取模块来实现目标的特征表示。其中 Wen 等人[40]基于 ATM 监控场景展开研究，提出了一种基于改进 Hough 变换的安全帽佩戴检测算法。该算法以安全帽的圆弧特征为核心分析要素，但依赖近距离获取的高分辨率图像，导致其在复杂施工环境下的适用性受到限制。Felzenszwalb 等人[41]将图像频域信息和方向梯度直方图进行结合。通过对目标进行身份识别，进而结合圆环霍夫变换和目标色彩特征实现安全帽的佩戴检测。虽然该算法可以完成检测，但其颜色识别范围受限于预设类别，对非标定色系的安全帽缺乏有效检测能力。Du 等人[42]从感知上下文信息出发，利用经典的 Haar-Like 特征提取算法对人脸进行特征提取并定位，再根据人脸上方区域的颜色信息判断人员是否佩戴安全帽，但在光线较差区域或阴雨天气环境下，该算法的检测效果并不理想。Kelm 等人[43]设计了一种移动射频门禁系统，该系统可在特定场景下识别安全帽的存在状态，但无法识别是否正确佩戴安全帽。刘晓慧等人[44]提出肤色检测和Hu矩在安全帽佩戴检测中的应用。虽然取得了较好的检测效果，但只能对有颜色的安全帽进行检测，且当图像中出现遮挡、光照强度变化剧烈的情况时，算法鲁棒性不强，有很大的局限性。贾俊苏等人[45]运用可变性部件模型进行安全帽特征信息的处理，利用基于块的局部二值模式处理

安全帽的有效信息，再通过支持向量机对安全帽进行训练和检测。该算法虽然可以检测出安全帽，但检测精度较低，还有很大的提升空间。李琪瑞等人[46]利用凸字算法检测人体头部。该研究采用 Vibe 算法对人体目标进行定位，再结合 HOG 特征提取器和支持向量机，构建安全帽佩戴检测模型。但由于环境的复杂，该算法不能满足实际施工现场实时检测的要求。Marin-Jimenez 等人[47]则提出了一种基于背景差分法的安全帽佩戴检测算法。首先对图像的前景区域进行目标轮廓的特征提取，再通过构建分类模型，针对人员的头部区域和安全帽区域开展联合检测。但在目标被遮挡场景下，该算法易出现误检等问题。

基于深度学习的安全帽佩戴检测算法在训练过程中引入特征自学习机制，有效克服了手工设计特征存在的不足。下面将系统梳理两阶段安全帽佩戴检测算法和一阶段安全帽佩戴检测算法的研究现状。

在两阶段安全帽佩戴检测算法的研究过程中，Fang 等人[48]提出一种基于 Faster R-CNN 的安全帽佩戴检测算法，该算法收集了 81000 张图像作为训练数据集，涵盖真实安全帽佩戴检测场景中存在的多种问题，包含不同比例、遮挡和小物体检测等，该算法也开启基于深度学习的安全帽佩戴检测算法研究。Sun 等人[49]则提出了一种融合注意力机制的改进 Faster R-CNN 安全帽佩戴检测算法。通过引入以不同规模的全局信息为重点的注意力机制，改进后算法的检测精度得到明显提高，但由于大幅增加了模型复杂度，并不适合部署在嵌入式设备上。Saravanan 等人[50]又提出一种基于 Mask R-CNN 的多阶段头盔佩戴检测与车牌提取系统。通过任务分解实现精准信息获取，避免传统单阶段算法的过载问题。然而，恶劣天气环境可能导致算法的检测精度下降，误检率上升。

在一阶段安全帽佩戴检测算法的研究过程中，Li 等人[51]通过在 SSD 算法中引入特征融合模块，充分提取到安全帽的特征信息，提高了算法的检测精度，但改进后算法对于远距离目标的检测能力较弱。王海瑞等人[52]则针对 CenterNet 算法的损失函数进行改进，并采用 ASFF 自适应模块整合主干网络提取到的特征，尽管检测精度得到一定程度的提升，但网络参数量过多。Li 等人[53]又提出一种基于 YOLOv4 的轻量级安全帽佩戴检测算法。通过融合高层低噪声特征和底层细节特征，提升对小目标的检测能力。然而，因未针对性学习安全帽的判别特征，改进后算法存在漏检和误判风险。Lee 等人[54]则又提出一种基于

YOLOv5 的安全帽佩戴检测算法。将安全帽佩戴检测任务扩展为三类，避免传统算法将帽子误判为安全帽的隐患。然而，改进后算法的训练数据并非来自实际施工现场，这导致模型的适应性不足。Han 等人[55]又提出一种基于改进 YOLOx 的夜间安全帽佩戴检测算法。通过构建残差模块，算法的检测性能显著提升，但将算法应用于日间场景时，检测精度出现下降。随后 Li 等人[56]提出一种基于 YOLOv8 的安全帽佩戴检测算法。通过在 C2f 模块中引入坐标注意力机制，提升了对小目标和复杂尺寸目标的特征提取效果。然而，改进后算法仅使用 P3 和 P4 层作为检测头，易损失高层语义信息。Zhang 等人[57]接着提出一种基于 YOLOv9 的安全帽佩戴检测算法。利用姿态估计模型 YOLO-Pose 提取人体关键点，间接推断被遮挡头部的可能位置。但在密集目标场景下，安全帽检测框易重叠，提取到的关键点可能交叉干扰。

研究表明，与传统检测算法相比，基于深度学习模型的安全帽佩戴检测算法在检测性能方面展现出显著优势。然而，现有算法仍存在以下局限性：首先，该类算法假设的检测场景较为理想，未能全面涵盖施工现场复杂多变的环境特征；其次，小尺寸目标检测精度仍有待提升，在遮挡、低分辨率等场景下的检测效果未达到预期。此外，模型的轻量化设计仍存在不足，推理速度和计算消耗之间的平衡问题仍需优化，难以满足实时检测的应用需求。

1.3 论文的主要研究内容

论文系统梳理和综合分析了国内外安全帽佩戴检测领域的研究进展，整理出当前检测技术面临的核心挑战。在此基础上，重点聚焦基于深度学习的检测算法，深入探讨了经典模型的检测机制和核心原理，为后续研究方向的拓展奠定了理论基础并提供了方法论指导。论文主要通过对 CenterNet、YOLOv5s 以及 YOLOv7-tiny 模型进行改进，旨在提高检测的精度、减少模型的参数量、缓解漏检误检以及对小目标检测效果不理想的问题，进一步提升算法的检测性能，主要研究内容如下：

（1）基于改进 CenterNet 模型的安全帽佩戴检测算法

从提高检测精度和减少模型参数量的角度出发，提出一种基于改进 CenterNet 模型的安全帽佩戴检测算法。首先，针对检测精度低的问题，设计了

轻量化主干网络 PCCFNet 替代算法原主干网络 ResNet50，通过参数重分配显著增强了特征交互模块的权重占比，提升了多尺度特征融合的表征能力；其次，针对改进后模型在深层卷积操作中易受梯度弥散影响，导致细粒度特征表征能力下降的问题，在主干网络中引入 CBAM，轻量化的模块设计可无缝嵌入主干网络架构中，显著提升模型的鲁棒性；最后，针对安全帽的高频细节特征易与背景纹理产生特征混淆的问题，对热力图损失函数进行改进，提高对安全帽正样本的关注以再次提升检测精度。

（2）基于改进 YOLOv5s 模型的安全帽佩戴检测算法

从缓解复杂施工环境下安全帽漏检误检问题的角度出发，提出一种基于改进 YOLOv5s 模型的安全帽佩戴检测算法。首先，为解决信息丢失的问题，设计了特征融合模块 Star-CAA Block 代替原算法主干网络中的部分 C3 模块，通过动态调整不同场景下特征的贡献度，提升模型对遮挡、倾斜等复杂情况的鲁棒性，减少因局部特征丢失导致的误判；其次，针对损失函数 CIoU Loss 过度优化长宽比损失项，导致密集目标漏检或方向异常目标误检的问题，将损失函数替换为 MPDIoU Loss，通过最小化预测框和真实框的最小点距离，避免了因长宽比线性假设和安全帽实际比例不匹配而导致的优化偏差，更稳定地约束边界框的回归方向。同时，针对数据集 SHWD 中的部分图像来源于教室监控下画面，不符合实际施工现场环境，不适用于模型训练的问题。通过优化并清理数据集，系统筛选和增补高质量图像，增强了数据代表性和场景覆盖度，从而构建更具真实性和泛化能力的数据集 R-SHWD。

（3）基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法

从提高对安全帽小目标检测效果的角度出发，提出一种基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法。首先，针对模型在特征学习过程中对小目标的语义信息关注度不足，导致检测效果不佳的问题，在主干网络中引入近乎无参的 Triplet Attention，保持模型轻量化特性的同时，有效克服了复杂背景干扰、特征响应弱化等问题，实现复杂环境下小目标的精准定位；其次，针对基线模型的三组锚框是针对 coco 数据集中的目标尺度来设计，不适用于数据集 R-SHWD 的问题，采用无监督学习方法，基于 K-means 算法对安全帽目标尺寸进行自主聚类分析，通过挖掘数据内在的分布特征，构建与真实目标几何特性高

度契合的锚框参数，有效降低小目标在特征图上的语义丢失概率，再次提升模型对小目标的定位精度；最后，针对小目标的特征信息较少，损失函数 CIoU Loss 依赖不准确的长宽比信息进行回归，导致预测框偏离真实目标的问题，将损失函数替换为 SIoU Loss，在边界框回归任务中引入角度损失项，将预测框的方向一致性纳入损失约束，有效缓解了小目标因坐标偏移引起的边界模糊问题。

1.4 论文的组织结构

全文组织结构共分为六章，具体如下：

第一章是绪论部分。首先，系统阐述了安全帽佩戴检测技术的研究背景、理论价值及现存挑战；其次，全面梳理了目标检测及安全帽佩戴检测技术的国内外研究进展；最后，详细说明了本研究的核心内容和技术路线。

第二章对基于深度学习的安全帽佩戴检测相关技术展开介绍。包括卷积神经网络的重要组成、经典的安全帽佩戴检测模型以及该技术的评价指标。

第三章是基于改进 CenterNet 模型的安全帽佩戴检测算法研究。首先，概述了算法的研究思路和整体框架；其次，阐述了算法中的各个模块，包括轻量化主干网络 PCCFNet、CBAM 以及损失函数；然后，从与其他算法对比、与基线算法对比的实验中说明该算法的先进性，从整体消融实验以及可视化实验中分析该算法的有效性；最后，对该章内容进行总结。

第四章是基于改进 YOLOv5s 模型的安全帽佩戴检测算法研究。首先，概述了算法的研究思路和整体框架；其次，阐述了算法中的各个模块，包括引入注意力机制的特征融合模块 Star-CAA Block、损失函数以及重构数据集 R-SHWD；然后，开展了一系列消融实验和对照实验，以验证所提算法的有效性；最后，对该章内容进行总结。

第五章是基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法研究。首先，概述了算法的研究思路和整体框架；其次，阐述了算法中的各个模块，包括 Triplet Attention、K-means 聚类算法以及损失函数；然后，在数据集 R-SHWD 上设置相关消融实验和可视化分析；最后，对该章内容进行总结。

第六章是总结展望，概述论文研究内容和工作，同时分析当前研究存在的不足，并指出今后的工作方向。

第 2 章 基于深度学习的安全帽佩戴检测相关技术分析

2.1 引言

随着计算机硬件性能的不断提升，深度学习技术迅速发展。在安全帽佩戴检测领域，过去依赖手工设计的特征提取方法已被基于卷积神经网络的特征提取方法取代。基于卷积神经网络在特征学习方面的显著优势和智慧工地场景需求的增加，以深度学习为基础的安全帽佩戴检测技术被普及并得到广泛认可。

本章是基于深度学习的安全帽佩戴检测技术分析。为更好地理解本论文的研究内容，本章从卷积神经网络的四个组成部分、经典的安全帽佩戴检测模型，以及该技术的评价指标展开介绍。

2.2 卷积神经网络

卷积神经网络[58](Convolutional Neural Network, CNN)是一种深度学习模型，能够自主学习视频、图像中的有用信息，主要通过权值共享、局部感受野和池化操作实现分类、检测等任务，是现代深度学习领域的重要组成部分。如图 2-1 所示，CNN 网络主要由输入层、卷积层、池化层、全连接层和激活函数五个部分组成。

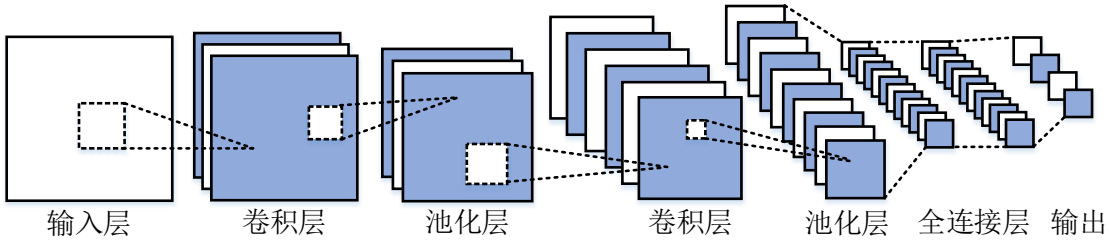


图 2-1 CNN 结构图

2.2.1 卷积层

卷积层作为 CNN 中的核心单元，通过限制神经元仅连接到输入数据的局部邻域，并引入参数共享机制，能有效提取图像中的局部特征，同时显著减少模型参数量。卷积层包括三部分：待处理的图像数据、串联卷积核构建的滤波单

元，以及输出的特征映射集合。特征映射作为 CNN 中卷积操作后的中间结果，其作用是对图像中的局部纹理特征和空间位置关系进行编码。通常为了增强 CNN 的特征表达能力，可通过在各卷积层中引入多组差异化特征映射，实现图像特征的精细化表达；也可通过增加网络深度，促进对深层特征的提取，提升模型的表达能力。

卷积层操作的实质是对输入数据按元素相乘并累加得到卷积结果，其计算过程如图 2-2 所示。以 1 通道 5×5 的输入图像为例，使用 3×3 的卷积核，按照规定步长 1，从左到右、从上到下进行滑动计算，输出 3×3 的特征图。在实际的卷积神经网络架构中，每个卷积层通常由多个并行卷积核构成，通过协同执行卷积操作，共同提取输入特征的不同维度信息，公式(2-1)可表示卷积操作。其中，a^{l-1}表示输入特征图，k^l 表示卷积核，l 表示第 l 个卷积核，a^l 表示卷积操作后的特征图。对输入为三通道的可见光图像进行卷积操作，则需要将每个通道与卷积核进行计算，然后对应位置求和，输出特征映射的矩阵表示。

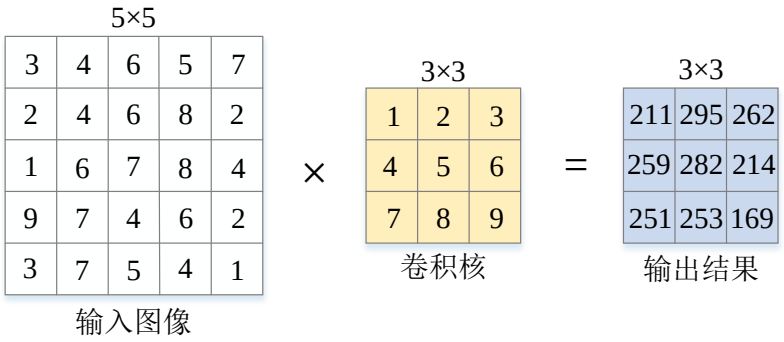


图 2-2 卷积计算过程

$$a_j^l = f\left(b_j^l + \sum_{i \in M_j^l} a_i^{l-1} * k_{ij}^l\right) \tag{2-1}$$

相较于人工神经网络(ANN)，CNN 借助局部感知域机制和参数共享策略，实现图像信息的高效抽取和层次化表达，不仅降低了模型参数量，更通过特征层级化处理提升了视觉信息的抽象表达能力。权重共享即利用权内系数相同的滤波器，对输入图像进行卷积处理。卷积层由许多可学习的卷积核集合组成，当处理复杂的图像数据时，卷积核内的神经元无法与上层的所有神经元进行连接，故将神经元和输入数据的局部区域连接，被覆盖的区域称为感受野。感受野大小决定了对数据信息的理解程度，较大的感受野能够增强对上下文信息的

感知能力，较小感受野则更关注细节信息。如公式(2-2)所示，每一层的抽象层次可以由感受野的值来推测。

$$l_k = l_{k-1} + \left((f_k - 1) * \prod_{i=1}^{k-1} s_i \right) \tag{2-2}$$

其中，l_{k-1}表示前一层的感受野，f_k表示第 n 层卷积核的尺寸，s_i表示前一层的步幅，l_k表示该层感受野的尺寸大小。

此外，CNN 通过自动学习特征表示，无需手工设计特征，提升了其在计算机视觉任务中的处理性能和泛化能力。而 ANN 在处理高维图像数据时，其全连接结构往往导致模型参数量呈指数级增长，进而面临过拟合风险。

2.2.2 池化层

池化层作为 CNN 的重要组成部分，常和卷积层交替堆叠构建特征提取模块，通过在特征图上采用滑动窗口逐区域遍历，对特征信息实施降维聚合操作。池化层通过下采样处理，能显著降低特征图的空间维度，有效减少模型参数量并降低计算复杂度。通过特征聚合策略，能有效保留输入特征图的核心语义信息，同时降低噪声干扰和冗余细节的影响，显著提升模型的鲁棒性及泛化能力。常见的池化有最大池化(Max Pooling, MP)和平均池化(Average Pooling, AP)。最大池化和平均池化操作分别如图 2-3 和图 2-4 所示。

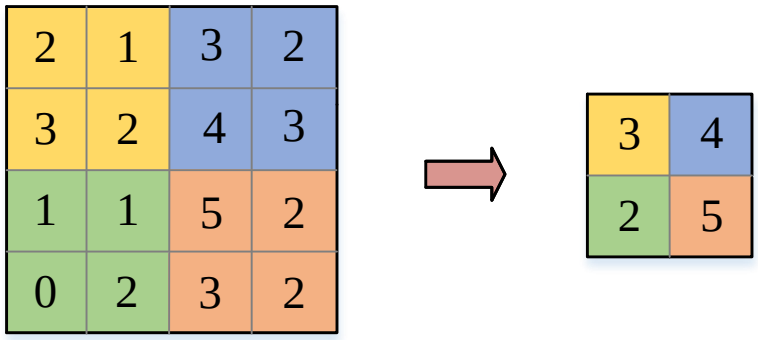


图 2-3 最大池化操作

最大池化操作的核心是提取局部感受野内的最大值作为输出特征，通过调节池化窗口尺寸和滑动步长参数，能有效控制特征图的空间分辨率。

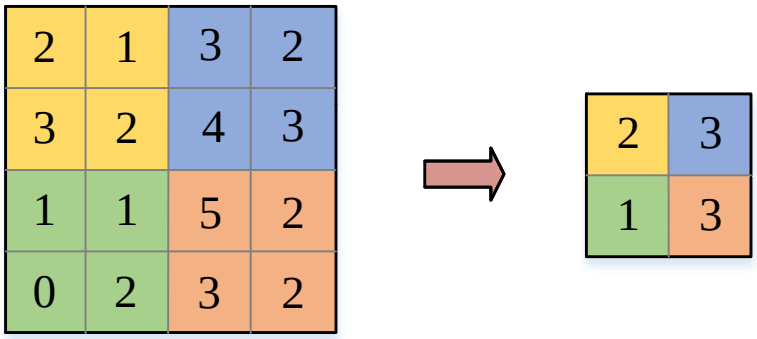


图 2-4 平均池化操作

平均池化操作通过计算滑动窗口覆盖区域内的元素均值生成输出结果。与最大池化相比，平均池化通过邻域信息的整合机制，有效降低了噪声干扰，并对图像特征进行平滑处理，这一特性使其在对特征平滑度要求较高的视觉任务中具有显著优势。

2.2.3 全连接层

在 CNN 中，全连接层(Fully Connected Layer)承担着分类决策的功能，通常位于神经网络的末端，表现为当前层神经元和前一层所有神经元的完全连接。通过特征提取过程学习到权重参数和偏置项，实现从分布式特征表示到样本标签空间的非线性映射。全连接层操作如图 2-5 所示。

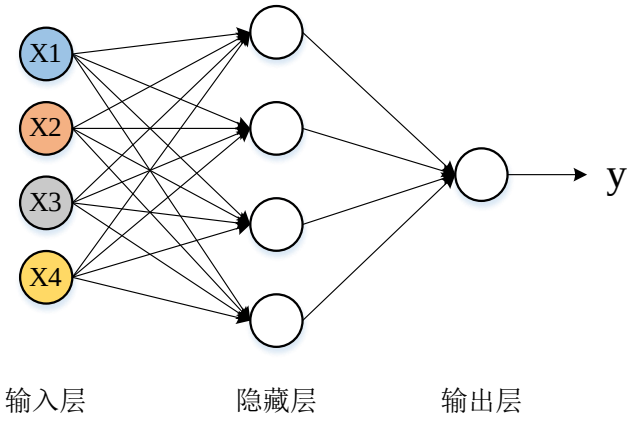


图 2-5 全连接层操作

全连接层由于其较强的灵活性和非线性表达被广泛应用，然而该层的引入会导致网络参数量呈指数级增长，可能造成参数冗余问题，显著增加过拟合风险，导致模型泛化性能下降。对此，研究人员通常采用 Dropout 正则化技术进行优化，该方法以 1-p 的概率随机失活神经元，并将其输出置零处理，从而阻断反

向传播路径。

2.2.4 激活函数

在卷积神经网络的复杂架构中，激活函数发挥着举足轻重的作用，它负责将卷积层提取到的特征引入非线性变换过程。鉴于卷积操作的线性本质是输入数据和卷积核的线性组合，为了能识别出更为复杂的特征，网络需借助激活函数，来加入所需的非线性映射功能。由此，激活层利用经过挑选的激活函数，为经过卷积处理的数据赋予新的活性，使得特征图中的信息演化为非线性特质的形式。在众多激活函数中，常用的激活函数有 Sigmoid[59]和 ReLU[60]等。

（1）Sigmoid 激活函数：Sigmoid 函数形状似 S 型曲线，是一种在两端进入饱和状态的函数。当处理各种定义域内的输入时，Sigmoid 函数能够将输入转化为介于(0,1)这一有界区间内的输出结果，赋予网络输出映射渐变的特性，Sigmoid 函数公式如式(2-3)所示。

$$f(x)=\frac{1}{1+e^{-x}} \tag{2-3}$$

其导数为：

$$f'(x)=f(x)(1-f(x)) \tag{2-4}$$

Sigmoid 函数及其微分图像如图 2-6 所示。可以看出在输入为零时，Sigmoid 函数梯度达到了其最大峰值，约为 0.25，输出值变化较为灵敏。当输入开始在任何方向偏离这个平衡点时，梯度逐步降低，渐渐向零逼近，会出现梯度消失问题，这给模型训练带来了阻碍。

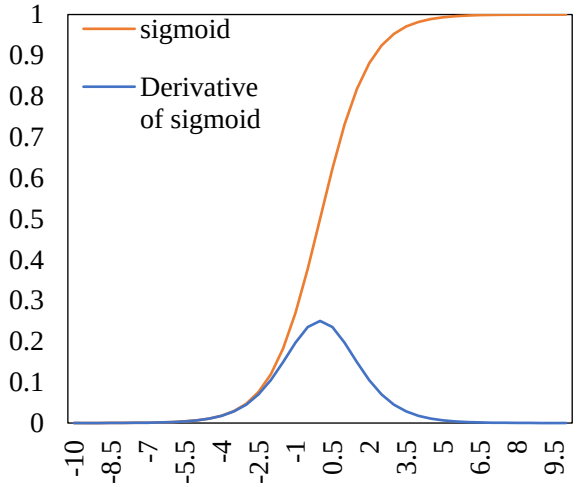


图 2-6 Sigmoid 函数及其导数图像

（2）ReLU激活函数：ReLU函数又被称为修正线性单元，是一种常用的激活函数，因其结构的简约性和高效的计算性，使神经网络达成快速收敛成为可能。ReLU 函数不仅优化了训练过程中的计算速度，而且在很大程度上缓解了训练中梯度消失这一通病，其计算过程如公式(2-5)所示。

$$f(x)=\begin{cases}0 & x\le 0\\ x & x>0\end{cases} \tag{2-5}$$

其导数为：

$$f(x)=\begin{cases}0 & x\le 0\\ 1 & x>0\end{cases} \tag{2-6}$$

ReLU 函数图像及微分图像如图 2-7 所示。可以看出，在输入值跨越零点进入正数领域时，梯度值恒为 1，能够避免梯度消失的问题，从而维持信息流的传递。当输入值位于负数范围时，梯度值将保持恒定为零的状态，这一特性使得模型训练过程中，部分神经元可能会陷入未被激活的状态。这些未被激活的神经元在模型表达过程中将失去作用，无法对模型的输出结果产生影响。

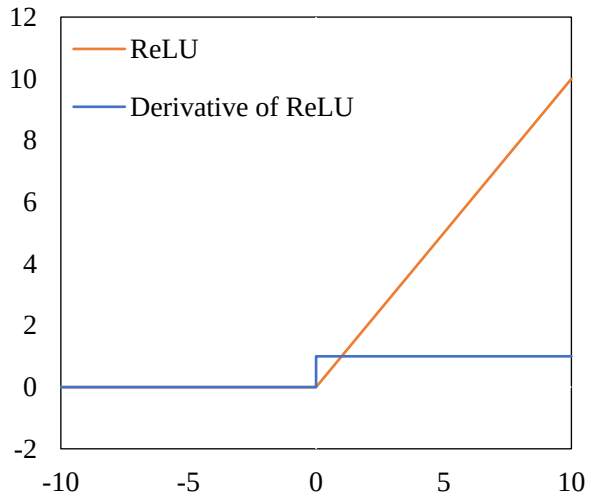


图 2-7 ReLU 激活函数及其导数图像

2.3 经典模型架构

2.3.1 CenterNet 模型

CenterNet 目标检测算法提出了以点代替框进行目标检测的新思想。具体而言，该算法分为两个核心部分：目标中心定位和尺寸回归。在目标中心定位部

分，通过高斯核生成高置信度热力图。在尺寸回归环节，模型将每个物体的中心像素作为独立的训练样本点，从而能够精确地预测出该点所对应物体的宽度和高度。此外，模型还通过预测中心偏移量的方式，精细地校正了输出所带来的离散误差，以实现更精准的目标定位。其网络框架如图 2-8 所示。

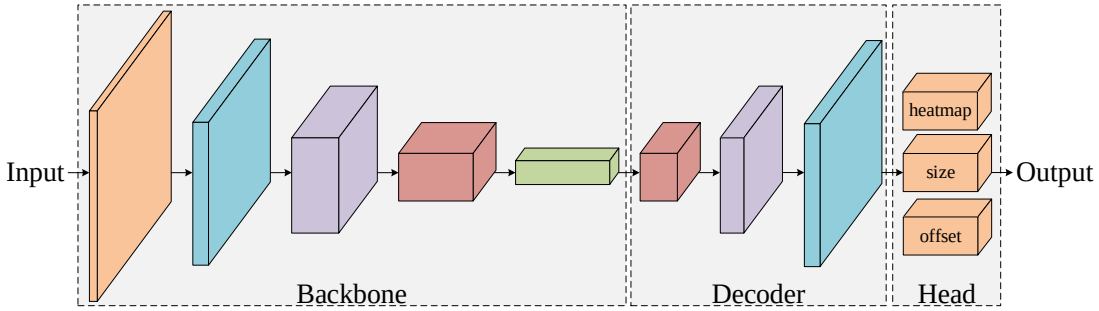


图 2-8 CenterNet 网络框架

CenterNet 模型采用 ResNet[61]作为主干特征提取网络，来完成视觉特征提取任务。ResNet 这一结构的提出，是为了解决卷积神经网络深度加深引起的性能下降问题，其核心在于设计了跨越数层的直连通道，通过将前层的原始输入直联至当前层的输出端，构建了残差架构。优势之处在于，模型引入残差连接的设计，使得网络在增加深度的同时，保持各层的输出基于前一层的输出进行递进式累加，从而显著降低了信息损耗的风险。这种设计策略有效提升了模型的稳定性和训练效果。假设将输入变量 x 传递至网络的某一层操作，那么该层处理后的输出可记为 $h(x)$，则 ResNet 网络的具体运算过程如式(2-7)所示。

$$f(x)=h(x)-x \tag{2-7}$$

其中，$f(x)$代表当前网络层输出和输入之间的差异，即残差。当某一层网络的输出 $h(x)$和该层接收的输入 x 相匹配时，$f(x)$为零，表示模型在此阶段没有提取到新信息。换言之，当网络输出 $h(x)$比输入 x 更接近理想输出时，$f(x)$的值通常较小。该机制保证网络的每一层通过残差学习提取到独特的特征表示，从而有效提升整体网络的检测精度。

ResNet 网络由多个残差结构组合而成，每个残差结构内部不仅嵌入了多个卷积层，还引入了跨层连接结构。这一连接结构确保了特征信息在层间得以直接高效的传递。在模型的训练过程中，这些残差单元的输出会和其自身输入融合，融合后的信息输入到后续的网络层去执行进一步的处理。CenterNet 模型的核心步骤可分为以下四个部分。

（1）首先，输入尺寸为 1×3×512×512 的图像至卷积网络，采用 ResNet 骨干网络进行多阶段下采样处理，逐层提取图像的语义特征，生成尺寸为 1×2048×16×16 的特征映射。

（2）其次，采用三层反卷积结构对特征图进行上采样操作，旨在恢复更丰富的图像细节信息，生成维度为 1×64×128×128 的高分辨率特征表达。随后，由各独立检测分支分别输出目标中心点偏移量、空间尺寸参数以及热力图等多维度特征信息。

（3）然后，再对生成的热力图进行归一化处理，通过 Sigmoid 函数将其映射到 0 至 1 之间，并利用最大池化操作提取出热力图中的极大值点，同时将非极大值点置为零。再根据预设的置信阈值，筛选出前 100 个具有最高置信度的候选样本，并计算出目标的中心点位置、宽度和高度信息，以及中心点偏移量。

（4）最后，在得到包含中心点偏移和目标尺寸信息的候选样本后，采用解码方式结合这些信息，构建出目标的检测框，最终输出整体的目标检测结果。

2.3.2 YOLOv5 模型

2020 年，Ultralytics 公司发布单阶段目标检测算法 YOLOv5。依据模型的规模，YOLOv5 被分成四种型号：小型的 s、中型的 m、大型的 l 以及超大型的 x，其在 coco 数据集[62]上的测试结果表明，随着网络规模的增大，检测性能同步提升。其网络结构包括输入端、主干网络(Backbone)、特征融合网络(Neck)和输出端(Head)四部分，如图 2-9 所示。下文将对 YOLOv5 中的重要组成部分展开详细介绍。

（1）输入端

YOLOv5 作为一种强大的实时目标检测模型，在对输入图像进行处理的过程中，整合了一系列创新的技术手段，以增强模型对于各类目标的识别和检测精度。首先，在预处理阶段，采纳了马赛克数据扩增方法，通过将多个图像进行随机的缩放、剪切以及拼接处理，形成新的训练样本。经实验验证，该方法对于小目标检测的性能提升具有显著效果；其次，在锚框尺寸的确定上，引入了一种基于目标检测任务具体数据集标注情况的自适应锚框尺寸计算方法。通过这种方法，能够在模型训练初期确立最适宜的锚框大小，与标注框的差异作为反向误差传播的基础，持续对模型参数进行迭代优化。此外，还引入了自适

应图像缩放技术，通过专门的计算方法减少了对图像的过度填充，这不仅在训练过程中有助于提取更多有效特征，同时在推理过程中可以大幅减少计算量，有效提升模型的处理速度。

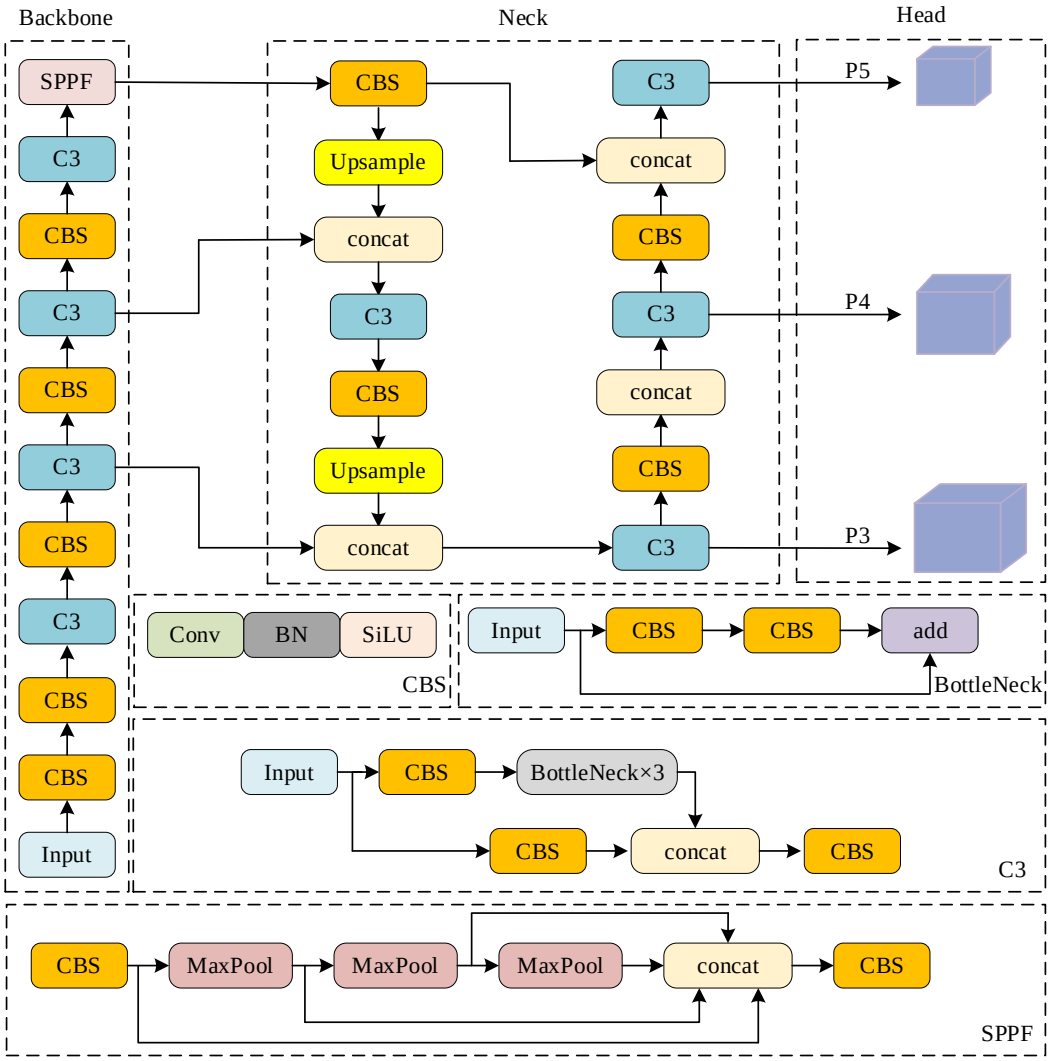


图 2-9 YOLOv5 网络框架

（2）主干网络

在 YOLOv5 目标检测算法的的框架当中，主干特征提取网络采用了 Wang 等人提出的 CSPNet 结构，并与经典的 ResNet 残差结构结合，形成了一种高效的特征提取网络。CSP 结构如图 2-10 所示，假设输入通道数为 c1，模型经过两次 1×1 的卷积处理，输出具有 c_通道数的中间特征图。随后，该特征图被划分为两个分支，一个继续按照残差结构传递，另一个保持不变。最终，两个分支的特征图在融合后，再次经过 1×1 卷积生成 c2 通道数的输出特征图。这种结构和残差网络的残差块相结合形成了特征提取网络的主要部分 C3，即包含三次卷

积操作的 CSP 结构。这种创新架构显著提升了卷积神经网络的学习效率，在降低计算成本和内存消耗的同时，确保了模型的精度。

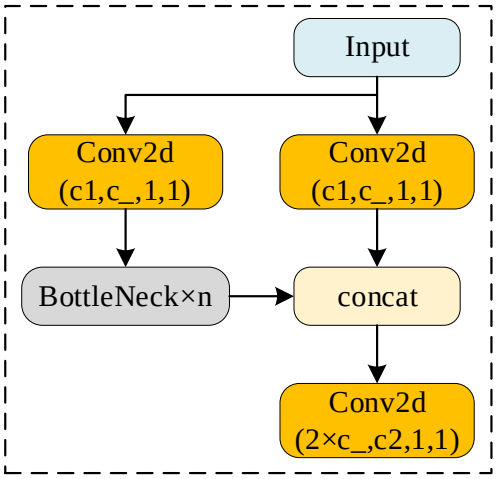


图 2-10 CSP 网络结构

（3）特征融合网络

在 YOLOv5 目标检测算法的框架当中，特征融合网络承担着串联主干网络和输出层的关键角色，通常被称为 Neck。该网络的核心作用是对各个层级的特征图进行有效整合，以增强模型对于各种尺寸目标的识别精度。借鉴 YOLOv4 的先进设计，YOLOv5 也采用特征金字塔网络代替像素聚合网络[63]，完成特征图的高效融合。在其主干网络部分，多级卷积层的堆叠对输入数据执行了连贯的下采样处理，从而生成了一系列具有层次性的特征图。其中，高层特征图承载着重要的语义信息，而底层特征图则包含了精确的定位以及清晰的轮廓信息。FPN 结构对高层特征图进行上采样，且与前一级底层特征图通过 Concat 操作进行通道方面的融合。结合 C3 模块和其他卷积操作后，FPN 在传递深层的高层语义信息到底层特征图的过程中，显著增强了模型对小尺寸目标的检测性能。另外，FPN 自上而下的信息流动策略，旨在补偿特征提取过程中可能损失的定位细节信息。

（4）输出端

在 YOLOv5 目标检测算法的框架当中，输出层扮演着不可或缺的作用，其根本任务在于对 Neck 阶段得到的多尺度特征图进行整合性分析，并且负责完成对物体的精准定位和类别鉴定。该过程将输入图像逐一细致划分为数个 S×S 的格点矩阵，并在每一单元格内生成诸多预选框作为潜在的检测候选。随后，算法实行非最大抑制操作，目的是为了消除那些信任度不足的候选框，从而保障

检测成果的高精度和稳定性。

2.3.3 YOLOv7-tiny 模型

YOLOv7-Tiny 作为一种轻量级的实时目标检测算法，其网络结构由输入端、主干网络(BackBone)、特征提取网络(Neck)和输出端(Head)四部分组成，如图 2-11 所示。下面将对 YOLOv7-Tiny 中的重要组成部分展开详细介绍。

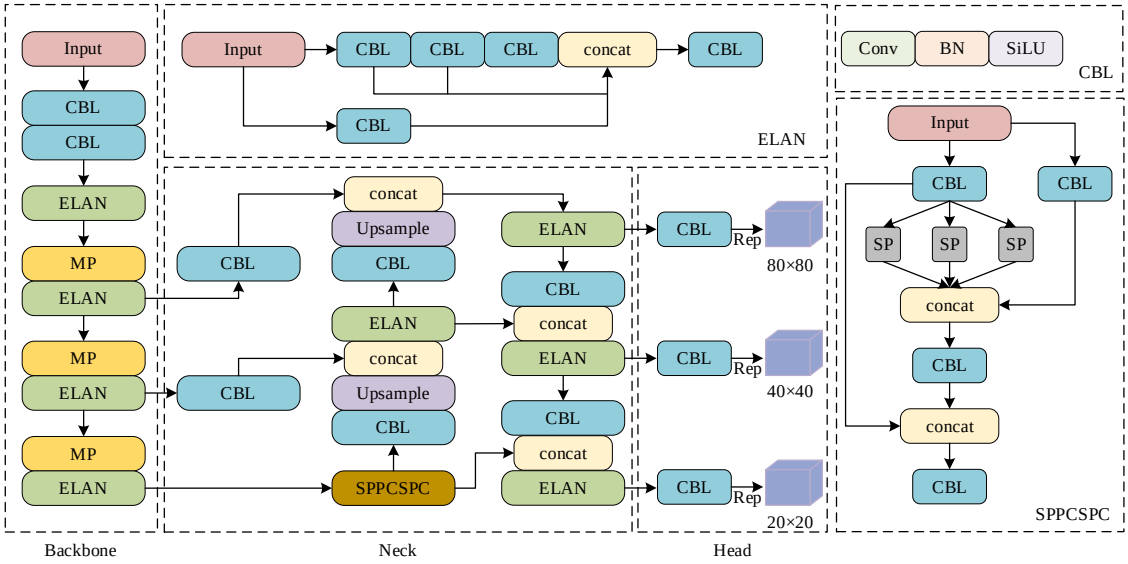


图 2-11 YOLOv7-tiny 网络框架

（1）输入端

模型在输入端采用多模态数据增强技术，主要包括 Mosaic 数据增强策略和自适应图像尺寸调整。其中，Mosaic 数据增强通过多源图像采样，对四张训练图像进行随机比例缩放、几何变换和拼接组合，构建复合训练样本，有效扩大了数据集的覆盖范围，从而提升模型对复杂场景的适应能力。自适应图像尺寸调整通过分析目标的尺寸和比例特征，动态调整输入图像的空间维度，能够有效增强模型对小尺寸目标的检测性能。

（2）网络架构

如图 2-11 所示，不同颜色的模块对应不同的操作类型。其中，CBL 模块作为模型的基础构成单元，采用可变步长和多尺度卷积核的标准卷积进行操作。MP 模块代表最大池化操作，步长和卷积核均为 2，能够实现两倍空间维度的下采样。ELAN 模块作为基础的特征提取单元，借鉴了残差结构的优势，把输入特征均分到两条通道开展特征提取工作。SPPCSPC 模块是特征提取网络中的最

后一个单元，同样借助双分支结构开展特征提取工作，一条进行全局池化操作，另一条仅进行 CBL 操作。Upsample 模块主要承担图像尺寸的上采样操作。Rep 模块代表重参数化卷积操作。

模型经过五次下采样和两次特征融合，由三个多尺度检测分支分别输出尺寸为 20×20、40×40 和 80×80 的锚框，以适应多尺度目标的任务需求。五次下采样操作包括两次步长为 2 的 CBL 模块以及三次 2 倍最大池化，特征融合操作则由两次上采样完成。

（3）损失函数

YOLOv7-Tiny 模型的损失函数包括目标检测损失、分类损失和边界框坐标损失。目标检测损失用于评估模型的检测精度，其由置信度损失和标定框位置损失组成。同时，使用交叉熵损失函数衡量模型对类别标签的预测性能，使用平方损失函数量化模型对边界框回归的预测精度。

2.4 评价指标

常见的指标包括精确率(Precision, P)、召回率(Recall, R)、平均精度(Average Precision，AP)、平均精度均值 mAP 以及每秒帧率(Frame Per Second, FPS)。

精确率(P)：在判定为正类的样本集合中，真实标签为正类的样本所占比例，计算公式如式(2-8)所示。

$$P = \frac{T_P}{T_P + F_P} \tag{2-8}$$

其中，T_P 表示判定为正类的正样本数量，代表的是被正确分类的正类。F_P 表示判定为正类的负样本数量，代表的是被错误分类的负类。

召回率(R)：在实际正类的样本集合中，被识别为正类的样本所占比例，计算公式如式(2-9)所示。

$$R = \frac{T_P}{T_P + F_N} \tag{2-9}$$

其中，T_P 与式(2-8)中的含义相同，F_N 表示判定为负类的正样本数量，代表的是被错误分类的正类。

平均精度(AP)：检测样本的准确率均值。利用不同精确率和召回率的组合

点绘制的曲线与坐标轴所围成的区域面积，即 P-R 曲线下的面积。平均精度的计算如式(2-10)所示。

$$AP = \frac{\sum P}{N} \tag{2-10}$$

其中，N表示图像数量。

平均准确率均值(mAP)：评估模型检测性能的核心指标，在完成多类别检测任务后，首先计算各类别的 AP，随后对所有 AP 进行算术平均运算。即：

$$mAP = \frac{\sum AP}{NC} \tag{2-11}$$

其中，NC表示目标的类别总数。

每秒帧率(FPS)：检测器每秒可以处理的图像帧数，值越大检测速度就越快。

2.5 本章小结

本章从以下几个方面对基于深度学习的安全帽佩戴检测技术进行简要阐述。首先，详细概述了卷积神经网络的主要组成部分，分别为卷积层、池化层、全连接层和激活函数；其次，对经典的安全帽佩戴检测模型进行了分析；最后，对安全帽佩戴检测领域的常用评价指标进行了介绍。

第 3 章 基于改进 CenterNet 模型的安全帽佩戴检测算法

3.1 引言

安全帽佩戴检测作为计算机视觉领域的重要应用之一，在电力与能源设施巡检场景下发挥着至关重要的作用。准确高效地检测安全帽佩戴情况，可以有效预防事故的发生，保障人员生命安全。近年来，基于深度学习的目标检测算法在安全帽佩戴检测领域取得了显著进展。然而，现有算法仍存在一些挑战。首先，在实际检测场景下，现有算法往往只关注目标本身的特征，忽略了其与周围环境的上下文信息，导致检测精度受限；此外，经典的目标检测算法，例如 YOLO 系列等，网络结构相对复杂，参数量庞大，难以满足实时性要求较高的场景需求。

为了提升算法的性能，学者们做出大量研究。Wang 等人[64]提出一种基于 YOLOv4-tiny 的安全帽佩戴检测算法。通过简化融合路径和减少冗余层，降低了模型参数量和计算复杂度，但也使得特征提取能力下降，导致精度下降。Han 等人[65]在 YOLOv5 检测模块之前设计了一种新颖的超分辨率重建模块。通过提高图像的分辨率来提升检测精度，但设计的新模块带来了参数冗余，导致检测速度不够理想。Jiao 等人[66]采用迁移学习技术对样本进行联合标注和训练，减少数据需求并提升模型的泛化能力。虽然改进后算法加快了模型收敛速度，但并未考虑轻量化，模型体积方面还有很大的提升空间。

针对以上问题，本章从提高检测精度和减少模型参数量的角度出发，提出一种基于改进 CenterNet 模型的安全帽佩戴检测算法。首先，设计了轻量化主干网络 PCCFNet，引入部分卷积模块替代标准卷积操作，有效降低了模型参数量和计算复杂度，同时通过级联融合策略实现底层细节特征和高层语义信息的渐进式融合，显著提高了目标的定位精度；其次，在主干网络中引入 CBAM，通过通道和空间维度的协同自适应校准，实现对目标特征的全局-局部联合增强，有效捕获安全帽和人体头部的关联性；最后，针对热力图损失函数进行改进，通过引入动态权重增强对正样本的关注，再次提升检测精度。

3.2 算法流程

本章基于 ResNet50 骨干网络提出适用于安全帽佩戴检测的算法。通过设计轻量化主干网络 PCCFNet、引入 CBAM 和修改热力图损失函数对原始的 CenterNet 模型进行改进，在大幅降低参数量的同时，提高模型对安全帽佩戴检测的精度。

改进后的网络结构如图 3-1 所示，由主干特征提取模块(Backbone)、特征融合模块(Neck)和检测头(Head)三部分组成。首先，主干特征提取模块通过提取输入图像的多尺度特征，整合不同感受野的语义信息，输出层次化特征表示。其次，特征融合模块将主干模块输出的深层特征进行融合。最后，由引入的 CBAM 实现特征权重的动态调整，更好适应不同背景下安全帽的颜色变化，以及不同角度拍摄时的形状变化，从而提高检测的稳定性和可靠性。同时，对热力图损失函数进行改进，通过重点关注安全帽正样本，模型能更好理解目标的本质特征，提高泛化能力，使其在实际应用中具有更好的性能表现。

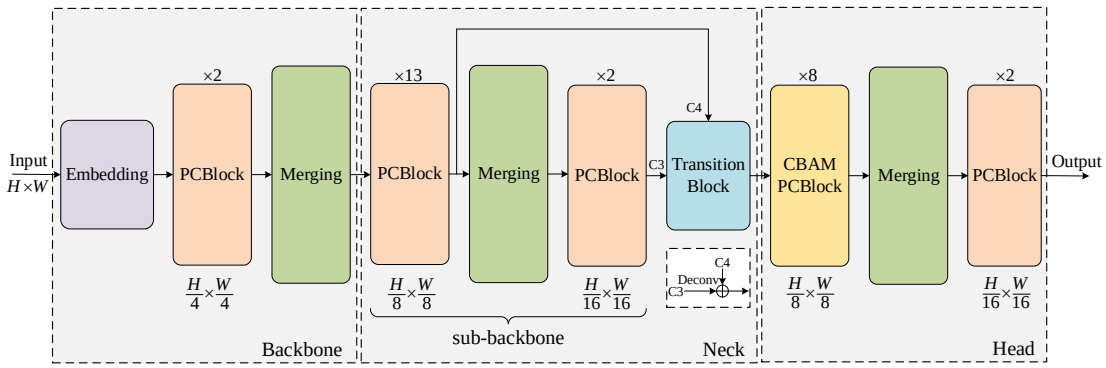


图 3-1 整体网络结构

3.3 算法原理

3.3.1 部分卷积模块

为了实现网络轻量化以减少开销，采用了部分卷积(Partial Convolution, PConv)来提取特征。改进的轻量化主干网络主要由 PCBlock 组成，其结构如图 3-2(a)所示，包括一个 PConv 层和两个 PWConv(Pointwise Convolution)层。该模块采用倒置残差结构，中间层执行通道维度扩展操作，同时通过残差连接实现输入特征的复用。图 3-2(b)解释了 PConv 是如何工作的，其中 w、h 为输入特征

图的宽、高，c_p 为应用常规卷积的通道数，c 为剩余通道数，k 为卷积核大小。它只对部分输入通道应用标准卷积以实现空间特征编码，其余通道保留恒定映射。经实验验证，部分卷积比 $r = c_p / c = 1/3$ 时取得最佳实验效果。

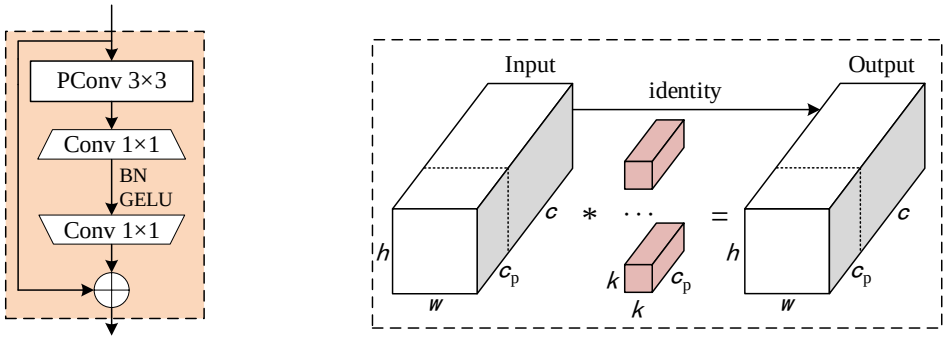


图 3-2 PCBlock 结构

对于神经网络来说，归一化层和激活层也必不可少。然而，许多工作[67-68]在整个网络中过度使用这些层，这可能会限制特征多样性，从而降低性能。相比之下，本章只将它们放在每个中间 PWConv 层之后，以保持特征多样性。此外，实验中使用了批归一化[69](Batch Normalization, BN)，BN 的好处是它可以合并到相邻的 Conv 层中，从而高效快速地进行推理。至于激活层，考虑到运行时间和有效性，选择 GELU 激活函数。

3.3.2 级联融合策略

多尺度特征对于目标检测任务至关重要，对于许多密集的检测任务，需要多尺度特征来处理不同尺度的对象。特征金字塔网络及其变体就被广大学者用于多尺度特征提取和融合。值得注意的是，现有网络在多尺度特征融合方面存在局限性，这主要是因为特征融合模块的参数量相对较少，适当增加模块的参数量占比可有效提升模型性能。

FPN 的特征融合模块可以看作是由对相邻层特征进行整合的整合操作和对整合特征进行转换的转换操作构成的。本章采用了一种新的策略，称为级联融合策略，用于安全帽佩戴检测。除了用于提取初始高分辨率特征的主干以外，还设计了级联阶段以生成多尺度特征，该阶段由特征提取子主干和用于特征整合的过渡块构成，通过将特征整合操作插入到主干中，确保整合后的信息在后续阶段能有效转换。这意味着，过渡块之后的所有参数均参与特征融合过程，

这样的设计显著增加了主干网络中用于特征融合的参数比例，进而提升了特征融合的深度和有效性。

采用部分卷积和级联融合的主干网络如图 3-3 所示，输入一张尺寸为$H\times W$的 RGB 图像，经过一个 Embedding（步幅为 4 的 4×4 卷积层）和多个连续的 PCBlock 处理，提取到$H/4\times W/4$的高分辨率特征。在主干网络的多级结构中，高分辨率的特征经过一个 Merging（步幅为 2 的 2×2 卷积层）下采样和通道数扩展后，被送入 Neck 中，图中 Transition Block 将不同尺度的特征逐步上采样和组合。最后，由 Head 输出的特征用于目标检测任务。

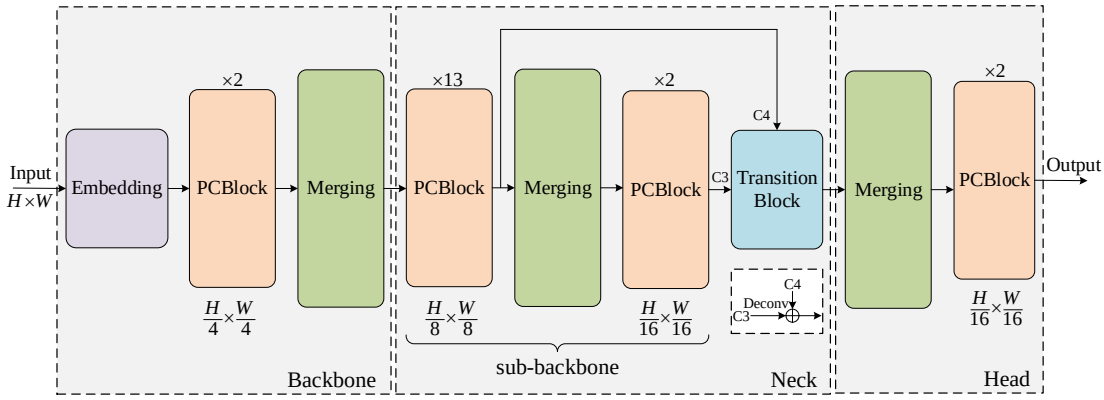


图 3-3 主干网络结构

3.3.3 注意力机制模块

安全帽图像中存在一定数量的目标，占整幅图像比例较小，易受背景等因素的影响。改进后的主干网络在进行深层次卷积时易丢失部分目标的特征信息，不利于检测。为了进一步提高物体分类能力和目标检测精度，在主干网络中引入 CBAM，其结构如图 3-4 所示。

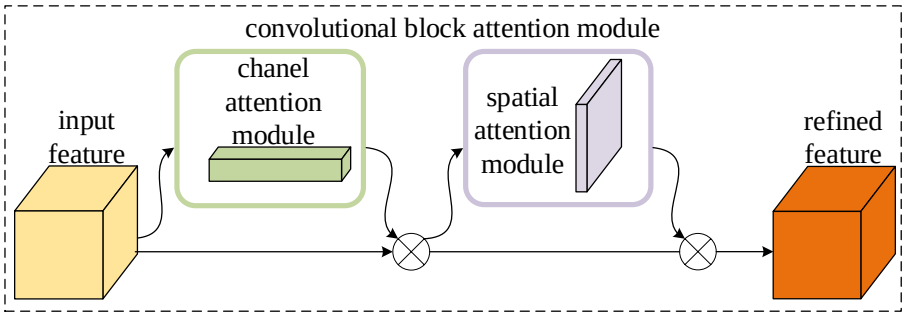


图 3-4 CBAM 结构

图 3-5 为通道注意力模块。首先，对输入特征图 F 做全局最大池化(Global Max Pooling, GMP)和全局平均池化(Global Average Pooling, GAP)。其次，将池

化后结果送入多层感知器并进行加和操作。最后，经过 Sigmoid 函数得到输出。

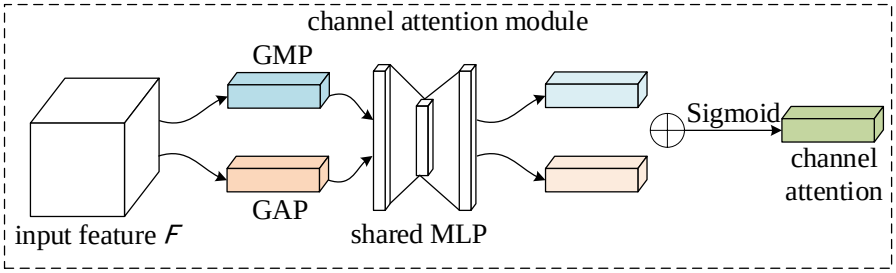


图 3-5 通道注意力模块

图 3-6 为空间注意力模块。首先，将通道注意力模块的输出作为输入，经过 GMP 和 GAP 维度压缩为 1×1。其次，将特征图进行通道拼接，经过一个 7×7 卷积，降维成一个通道。最后，经过 Sigmoid 函数得到输出。

CBAM 的具体引入位置见图 3-7，能在几乎不影响模型参数量的情况下，使模型对重要内容和重要位置进行关注，从而提高检测精度。

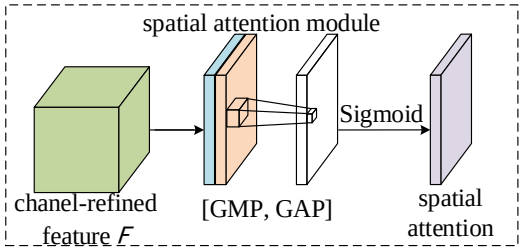


图 3-6 空间注意力模块

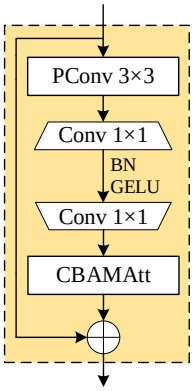


图 3-7 CBAM 引入位置

3.3.4 损失函数设计

针对安全帽数量有限且与背景占比失衡的问题，本章对热力图损失函数进行改进。采用 $\left(Y_{xyc}-\hat{Y}_{xyc}\right)^2$ 来增加 Y_{xyc} 为其他情况下的损失，在正样本损失项中加入动态权重系数 ω，强化对正样本特征的学习，在负样本损失项加入 1-ω，抑制其在总损失中的贡献度。改进后的热力图损失函数 $L_{\omega k}$ 见公式(3-1)所示。

$$L_{\omega k}=\frac{-1}{N}\sum_{xyc}\begin{cases}\omega\left(1-\hat{Y}_{xyc}\right)^{\alpha}\log_e\left(\hat{Y}_{xyc}\right) & Y_{xyc}=1\\ (1-\omega)\left(Y_{xyc}-\hat{Y}_{xyc}\right)^{2}\left(1-Y_{xyc}\right)^{\beta}\left(\hat{Y}_{xyc}\right)^{\alpha}\log_e\left(1-\hat{Y}_{xyc}\right) & \text{otherwise}\end{cases} \tag{3-1}$$

式中：α、β 表示超参数，本章取 $\alpha=2$、$\beta=4$；ω 也表示超参数，是对正

负样本的关注权重，经实验验证 $\omega=0.7$ 时取得最佳实验效果。N 表示一张图像中安全帽的数量，xyc 表示热力图上的所有坐标点，$\hat{Y}_{xyc}$ 表示预测值，Y_{xyc} 表示标注的真实值。

采用 L_1 Loss 计算的宽高损失 L_{size} 和中心点偏移损失 L_{off} 见公式(3-2)和公式(3-3)所示。

$$L_{\text{size}}=\frac{1}{N}\sum_{k=1}^{N}\left|\hat{S}_{\text{p}k}-S_k\right| \tag{3-2}$$

$$L_{\text{off}}=\frac{1}{N}\sum_{p}\left|\hat{O}_{\tilde{\text{p}}}-\left(\frac{p}{R}-\tilde{p}\right)\right| \tag{3-3}$$

式中：$\hat{S}_{\text{p}k}$ 表示安全帽预测尺寸，S_k 表示安全帽真实尺寸，$\hat{O}_{\tilde{\text{p}}}$ 表示预测的偏移量，P 表示图片中目标的中心点坐标，R 为缩放比例，$\tilde{P}$ 表示缩放后中心点的近似整数坐标。

综上所述，训练过程中损失函数 L_{ω} 的具体计算见公式(3-4)所示，设置 $L_{\text{size}}=0.1$，$L_{\text{off}}=1$。

$$L_{\omega}=L_{\omega\text{k}}+\lambda_{\text{size}}L_{\text{size}}+\lambda_{\text{off}}L_{\text{off}} \tag{3-4}$$

3.4 实验结果与分析

3.4.1 实验环境与参数设置

如表 3-1 所示，本章训练模型使用的是 Windows 操作系统，CPU 为 Intel Xeon Platinum 8255C @2.50GHz，GPU 为 NAVIDIA GeForce RTX 3090，深度学习框架为 Pytorch 1.11.0，编程语言为 Python 3.8。

表 3-1 实验环境配置

Configuration name	Configuration information
Operating system	Windows
CPU	Intel Xeon Platinum 8255C @2.50GHz
GPU	NAVIDIA GeForce RTX 3090
Deep learning framework	Pytorch 1.11.0
Programming language	Python 3.8

实验设置输入图像大小为 520×520，应用 Adam 优化器，并将动量因子设为 0.9，最大学习率设为 5e-4，Batch_size 设为 16，一共训练 300 个 Epoch。参数设置如表 3-2 所示。

表 3-2 实验参数设置

Experimental parameters	Value
Input_shape	520×520
Optimizer	Adam
Momentum	0.9
Max_lr	5e-4
Min_lr	Max_lr × 0.01
Batch_size	16
Epoch	300

3.4.2 实验数据集与评价指标

本章所用数据集为公开的安全帽佩戴检测数据集 SHWD、SHDD 和 GDUT-HWD，各数据集所含图像数量及标签类别见表 3-3 所示。设置训练集加验证集与测试集之比为 9:1，其中训练集与测试集之比同样为 9:1。

表 3-3 数据集所含图片数量及标签类别

Dataset	Images	Label				
SHWD	7581	Hat			Person	
SHDD	5000	Helmet			Head	
GDUT-HWD	3174	Blue	Red	White	Yellow	None

本章从客观评价和主观评价两方面来评估改进后 CenterNet 模型的性能。以每一类别的 AP 和 mAP 为衡量标准用于客观评价；利用改进前和改进后算法对图像进行检测，将预测到的目标框作为主观评价的依据。

3.4.3 PCCFNet 部分卷积比选择

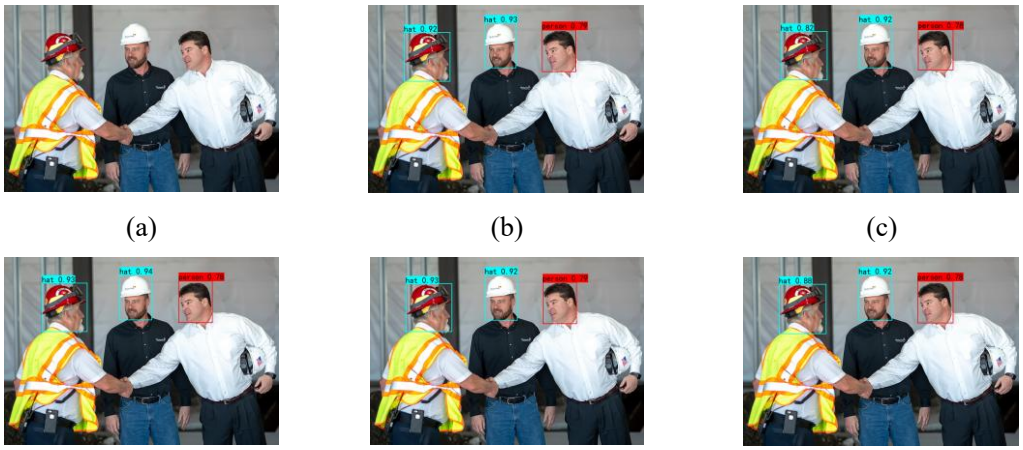
为了确定部分卷积比 r 的取值，单独对 r 进行实验。经实验验证，r 取 1/3 综合效果最佳。表 3-4 为具体的实验结果。首先，表 3-4 中“-”表示在 PCCFNet 中对全部输入通道应用常规卷积，其 mAP 和 FPS 均低于 r 取 1/3，同时参数量远多于 r 取 1/3。其次，r 取 1、1/2、1/4，mAP 和 FPS 也均低于 r 取 1/3；除 r 取 1/4，模型参数量略少于 r 取 1/3 以外，其余取值情况下的模型参数量均多于 r 取 1/3。由此得出 r 取 1/3，总体评价指标较优。

表 3-4 r 取值对比实验

r	Hat@0.5/%	Person@0.5/%	mAP@0.5/%	Parameters/ 10^6	FPS/frames
1	93.54	90.32	91.93	19.97	57.80
1/2	93.65	90.15	91.90	16.88	54.34
1/3	93.43	90.46	91.94	15.82	61.18

1/4	93.13	90.45	91.79	15.32	61.12
-	93.71	90.13	91.92	36.56	57.94

图 3-8 为 r 取不同值时，模型输出的可视化结果。其中，(a)为输入图像，(b)为 r 取 1 时的可视化结果，(c)为 r 取 1/2 时的可视化结果，(d)为 r 取 1/3 时的可视化结果，(e)为 r 取 1/4 时的可视化结果，(f)为全部输入通道应用常规卷积时的可视化结果。对比(b)(c)(d)(e)(f)可以发现，当 r 取 1/3，相较其他取值情况，安全帽佩戴检测的效果最佳，同时结合表 3-4 中的实验结果可知，r 取 1/3，mAP 和 FPS 最高，参数量较低，综合效果最佳。



(a) (b) (c)

(d) (e) (f)

图 3-8 部分卷积比 r 取值不同时的检测结果

3.4.4 损失函数超参数选择

为了确定损失函数中 ω 的取值，单独对 ω 进行实验。初始值设为 0.6，以 0.1 为步长进行递增，表 3-5 为具体的实验结果。

表 3-5 ω 取值对比实验

ω	Hat@0.5/%	Person@0.5/%	mAP@0.5/%
0.6	91.88	88.23	90.06
0.7	92.57	88.44	90.50
0.8	91.89	88.49	90.19
0.9	92.54	87.78	90.16
-	91.26	88.10	89.68

由表 3-5 可知，ω 取值由 0.6 递增至 0.7 时，检测精度呈现正向增长，而当 ω 进一步提升至 0.9 时，检测精度呈现明显的负向变化。表 3-5 中“-”表示原损失函数，其检测精度低于 ω 为 0.7 时。由此得出当 $\omega=0.7$ 时，总体评价指标较优。

3.4.5 消融实验

为了证实本章所提算法的可靠性，在 SHWD 数据集上对各个改进模块进行了消融实验。在 CenterNet 模型的基础上，依次进行改进：设计轻量化主干网络 PCCFNet、引入 CBAM、改进热力图损失函数。消融实验结果如表 3-6 所示。

表 3-6 消融实验结果

Method	PCCFNet	CBAM	L_{ω}	mAP@0.5/%	Parameters/ 10^6	FPS/frames	Model Size/MB
CenterNet	×	×	×	89.68	32.66	85.57	124.93
Experiment 1	√	×	×	91.94	15.82	61.18	60.58
Experiment 2	×	√	×	90.78	35.18	50.52	134.56
Experiment 3	×	×	√	90.50	32.66	89.33	124.93
Experiment 4	√	√	×	92.63	15.89	45.60	61.83
Experiment 5	√	×	√	92.35	15.82	59.20	61.83
Experiment 6	×	√	√	90.77	35.18	48.44	134.56
Experiment 7	√	√	√	93.05	15.89	47.31	61.83

由表 3-6 可知，本章设计的三种改进均可提升基于 CenterNet 模型的安全帽佩戴检测算法的性能。其中，PCCFNet 表示设计一种轻量化主干网络代替原算法主干网络 ResNet50，CBAM 表示在改进后的主干网络中引入 CBAM，L_{ω}表示对热力图损失函数进行改进。原 CenterNet 模型的 mAP 为 89.68%，参数量为 32.66M，FPS 为 85.57，模型大小为 124.93MB。Experiment 1 是将主干网络替换为 PCCFNet，模型大小为 60.58MB，参数量为 15.82M，FPS 为 61.18，mAP 为 91.94%，实验结果表明采用 PCCFNet 有效控制模型参数量的同时，显著提升了检测精度，并满足实时性需求；Experiment 4 引入 CBAM 后，模型大小为 61.83MB，参数量为 15.89M，FPS 为 45.60，mAP 为 92.63%，可以看出 CBAM 能够优化特征传播，再次提高检测精度；Experiment 7 改进热力图损失函数后，mAP 为 93.05%，检测精度得到进一步提高。

图 3-9 是部分消融实验每 5 轮评估一次的 mAP 变化曲线和 300 轮 Train loss 变化曲线，每个消融实验的 mAP 和 Train loss 通过不同颜色和线型的组合在图中加以区分。从图中可以看出，与 CenterNet 基线相比，Experiment 7 即改进后算法的 mAP 达到了最高，并且训练过程中的总损失均维持在较低区间，相较于基线表现出更强的损失控制能力，不仅实现了更快的收敛速度，还表现出更优的平滑度和稳定性。总体而言，算法的改进效果明显。

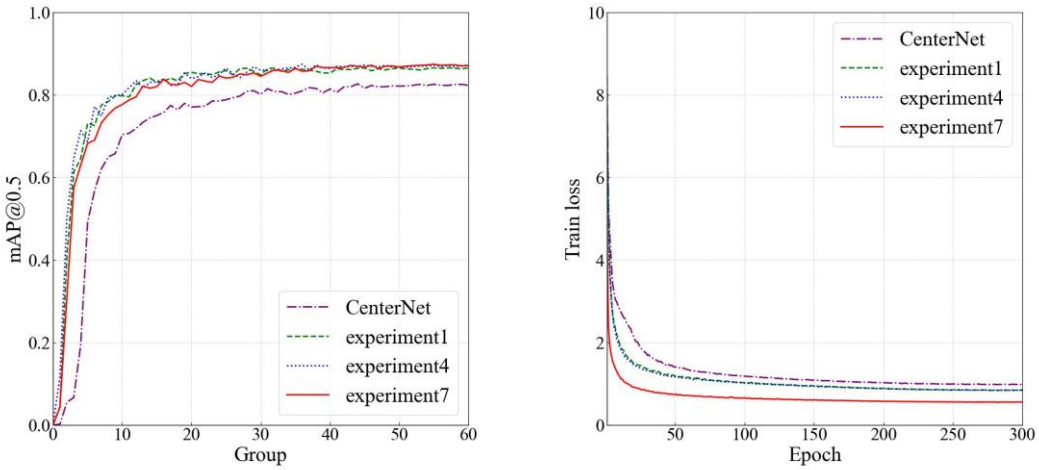


图 3-9 部分消融实验 mAP 和 Train loss 结果

3.4.6 与其他方法进行比较

为了更加直观地评价本章所提改进算法的性能，对以下算法进行了对比试验，表 3-7、表 3-8 和表 3-9 显示了本章改进算法与其他算法在 SHWD、SHDD 和 GDUT-HWD 数据集上的比较。

首先，由表 3-7 可知，改进后的 CenterNet 算法在 SHWD 数据集上的 mAP 达到了 93.05%，参数量为 15.89M，FPS 为 47.31，模型大小为 61.83MB，相较于其他算法综合效果最佳，具有精度高、参数量少的特点。与主流算法相比，虽然改进后的算法 mAP 略低于 YOLOv3，但参数量仅为 YOLOv3 的 1/4，模型大小仅为 YOLOv3 的 1/2，模型结构更加简洁，能有效适配工业生产设备的部署需求。与 YOLOv4、YOLOv5s、YOLOv6s 和 YOLOv7 相比，mAP 分别提升了 6.11%、1.35%、5.65%和 0.35%，同时参数量明显低于 YOLOv4、YOLOv6s 和 YOLOv7，FPS 略高于 YOLOv4。与近年来的安全帽佩戴检测算法[70]、[71]相比，mAP 分别获得 3.45%、1.65%的明显提升。与基于 ResNet50 的基线算法相比，Hat 和 Person 两类别的检测精度分别提升了 3.88%和 2.86%，mAP 提升了 3.37%，同时参数量不到基线算法的 1/2，模型大小仅为基线算法的 1/2，虽然 FPS 下降了 38.26，但能够保障大多数检测场景的实时处理要求。

表 3-7 不同算法在 SHWD 数据集上的实验结果对比

Method	Hat@0.5/%	Person@0.5/%	mAP@0.5/%	Parameters/ 10^6	FPS/frames	Model Size/MB
YOLOv3	92.10	94.30	93.20	61.50	125.00	117.77
CenterNet	91.26	88.10	89.68	32.66	85.57	124.93

YOLOv4	89.47	84.42	86.94	64.36	42.21	244.41
YOLOv5s	91.30	92.10	91.70	7.02	454.55	13.71
[71]	91.90	90.90	91.40	-	137.00	23.00
YOLOv6s	-	-	87.40	18.50	333.33	38.73
YOLOv7	91.50	93.90	92.70	36.49	128.21	71.33
[70]	-	-	89.60	-	32.00	-
Ours	95.14	90.96	93.05	15.89	47.31	61.83

其次，由表 3-8 可知，改进后的 CenterNet 算法在 SHDD 数据集上的 mAP 达到了 92.72%，除略低于 YOLOv3 以外，相较于 YOLOv4、YOLOv5s、YOLOv6s 和 YOLOv7，分别提升了 6.63%、0.42%、3.52%和 3.72%。与近年来的安全帽佩戴检测算法[70]相比，mAP 提升了 1.02%。与基于 ResNet50 的基线算法相比，Helmet 和 Head 两类别的检测精度分别提升了 1.11%和 4.40%，mAP 提升了 2.75%。

表 3-8 不同算法在 SHDD 数据集上的实验结果对比

Method	Helmet@0.5/%	Head@0.5/%	mAP@0.5/%	Parameters/10^6	FPS/frames	Model Size/MB
YOLOv3	94.50	91.70	93.10	61.50	65.36	117.77
CenterNet	92.44	87.50	89.97	32.66	90.00	124.93
YOLOv4	87.14	85.04	86.09	64.36	45.68	244.41
YOLOv5s	93.00	91.70	92.30	7.02	129.87	13.71
YOLOv6s	-	-	89.20	18.50	255.75	38.73
YOLOv7	89.10	88.80	89.00	36.49	91.74	71.33
[70]	-	-	91.70	-	32.00	-
Ours	93.55	91.90	92.72	15.89	45.88	61.83

最后，由表 3-9 可知，改进后的 CenterNet 算法在 GDUT-HWD 数据集上的 mAP 达到了 90.65%，相较于 YOLOv3、YOLOv5s 和 YOLOv6s，分别提升了 1.45%、2.95%和 6.95%。与近年来的安全帽佩戴检测算法[72]相比，mAP 虽略有降低，但参数量不到其 1/5，FPS 提高了 33.23。与基于 ResNet50 的基线算法相比，mAP 提升了 7.03%。

表 3-9 不同算法在 GDUT-HWD 数据集上的实验结果对比

Method	Label@0.5/%					mAP@0.5/%	Parameters/ 10^6	FPS/frames	Model Size/MB
	Blue	Red	White	Yellow	None				
YOLOv3	92.60	87.50	90.00	88.30	87.40	89.20	61.50	133.33	117.77
CenterNet	87.68	82.18	78.42	89.55	80.29	83.62	32.66	87.40	124.93
YOLOv5s	91.60	84.70	89.00	87.70	85.40	87.70	7.02	344.83	13.71
YOLOv6s	-	-	-	-	-	83.70	18.50	328.95	38.73
[72]	91.81	90.89	95.86	92.69	93.09	92.87	86.70	12.88	-
Ours	95.31	90.72	87.27	92.88	87.05	90.65	15.89	46.11	61.83

综合以上实验结果可知，本章所提改进 CenterNet 模型具有较低的参数量，

在三个公开的安全帽佩戴检测数据集上均达到较高的 mAP，同时能够满足实时性需求，易于在生产实践设备上进行部署。

3.4.7 可视化结果分析

为了验证所提算法的有效性，如图 3-10 所示，对基线算法和改进后算法在数据集 SHWD 上获得的可视化结果进行分析。其中，第一行为基线算法的可视化结果，第二行为改进后算法的可视化结果。

(a) 基线算法可视化结果

(b) 改进后算法可视化结果

图 3-10 部分可视化结果对比

结果显示，本章改进后算法在安全帽佩戴检测场景中的表现优于基线算法。例如，从前两列的对比图可以看出，改进后算法能显著提高单目标和多目标环境下安全帽佩戴检测的精度；进一步观察第三列图像可以发现，在光照剧烈的环境下，原算法对最左侧目标的置信度得分较低，但改进后算法不仅提高了置信度得分，还使边界框更精准；最后观察第四列可以发现，在光照昏暗的环境下，改进后算法明显提高了目标的定位精度。

3.5 本章小结

本章提出了一种基于改进 CenterNet 模型的安全帽佩戴检测算法。首先，通过部分卷积模块和级联融合策略的协同设计，实现了参数量和检测精度的平衡，减少冗余参数的同时保留了关键通道的多尺度信息；其次，在主干网络中引入 CBAM，一方面通过通道注意力模块评估特征通道的语义重要性，另一方面通过空间注意力模块聚焦头部区域的空间分布特性；最后，对热力图损失函数进

行改进，动态调整正负样本的权重分配，缓解了训练过程中样本分布不均衡的问题。本章算法的参数量为 15.89M，在三个公开的安全帽佩戴检测数据集 SHWD、SHDD 和 GDUT-HWD 上的 mAP 分别达到了 93.05%、92.72%和 90.65%。相较于基线算法，改进后算法的参数量减少了 16.77M，mAP 分别提高了 3.37%、2.75%和 7.03%。实验结果显示，模型参数量大幅下降、检测精度显著提升，充分验证了本章算法的可靠性，能够在电力与能源设施巡检场景下辅助监控系统实施安全帽佩戴检测，完成智能化的远程安全督导任务。

第 4 章 基于改进 YOLOv5s 模型的安全帽佩戴检测算法

4.1 引言

在狭窄空间及夜间维护作业场景下，安全帽佩戴检测发挥着极其关键的作用，建立智能化的安全帽佩戴检测系统，是防范意外伤害的重要举措。然而，经典的安全帽佩戴检测算法在特征学习过程中不能充分实现多尺度特征对齐，底层细节特征和高层语义特征会因分辨率或语义差异产生冲突，导致局部关键信息被抑制，且在遮挡严重时，模型难以精准定位安全帽轮廓，进而引发漏检误检问题。

论文第三章算法采用部分卷积模块和级联融合策略，在主干网络中引入 CBAM，并对热力图损失函数进行改进，显著提高了模型的检测精度，轻量化设计也大幅减少了参数量。然而，深层网络的特征提取受限于感受野扩张，高层语义特征虽能捕获全局上下文信息，但会弱化局部细节，这种全局和局部特征的失衡易使模型对非标准佩戴姿态的安全帽适应性下降，造成目标误检。同时，公开数据集缺乏对安全帽多样性及环境复杂性的充分覆盖，模型在训练阶段易受局部干扰和特征混淆影响，导致对边缘目标的漏检。

针对以上问题，本章从缓解安全帽佩戴检测易出现漏检误检问题的角度出发，提出一种基于改进 YOLOv5s 模型的安全帽佩戴检测算法。首先，设计了特征融合模块 Star-CAA Block 代替原算法主干网络中的部分 C3 模块，通过整合不同层次和类型的特征信息，动态调整各特征的权重，使底层细节特征和高层语义特征更好地互补，避免因特征信息缺失而导致的漏检误检；其次，采用 MPDIoU Loss 优化模型，根据损失函数的反馈不断调整预测结果，提高安全帽的识别精度，降低误检率。同时，对数据集 SHWD 进行优化，剔除冗余和无效数据，挑选并补充高质量图像，提升数据代表性和场景覆盖度。

4.2 算法流程

本章基于 CSPDarkNet53 骨干网络提出适用于安全帽佩戴检测的算法。通过设计特征融合模块 Star-CAA Block 和修改 CIoU Loss 对原始的 YOLOv5s 模型进

行改进，以缓解漏检误检的问题。

改进后的网络结构如图 4-1 所示，由主干特征提取模块(Backbone)、特征融合模块(Neck)和检测头(Head)三部分组成。首先，主干特征提取模块通过卷积提取输入图像的多尺度特征，并输出不同抽象层次的特征信息。其次，特征融合模块将主干模块输出的深层特征进行融合，以学习图像不同尺度的信息。最后，采用 MPDIoU Loss 替换 CIoU Loss，使网络学习更准确的边界框回归参数，避免因边界框过大或过小而遗漏安全帽的部分区域，进而降低漏检误检的可能。

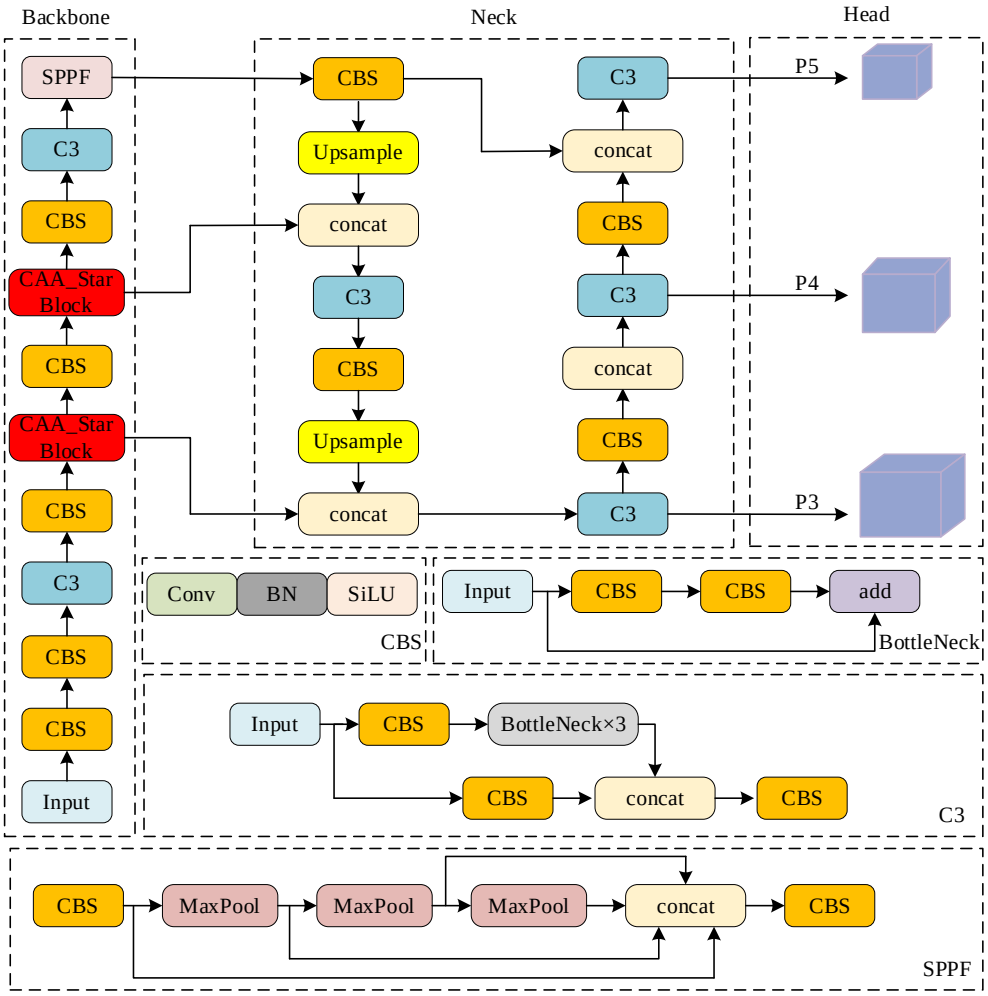


图 4-1 整体网络结构

4.3 算法原理

4.3.1 特征融合模块

为了提高模型在安全帽佩戴检测任务中的表现，增强目标特征在多维度空

间中的语义描述能力，减少漏检误检的可能，本章使用 Star Block 代替原算法主干网络中的部分 C3 模块。Star Block 通过引入星型运算，在低维输入数据空间可有效抽取复杂的高维非线性特征。在单层神经网络架构下，Star Block 通过整合权重矩阵和偏差参数，形成了统一的表示形式，表示为 $w=\begin{bmatrix} W \\ B \end{bmatrix}$，其中 W 表示权重部分，B 表示偏置项。同样地，输入向量 X 通过添加固定值元素扩展为包含常数项的矩阵，$x=\begin{bmatrix} X \\ 1 \end{bmatrix}$。通过这种方法，Star Block 实现了星形运算：$(w_1^{\mathrm{T}}x)\times(w_2^{\mathrm{T}}x)$。鉴于单输入单输出模型结构的简洁性，本章优先对其进行分析。具体来说，定义 w_1，w_2 和 $x\in R^{(d+1)\times 1}$，其中 d 表示输入通道数。这很容易扩展到多个输出通道，其中 w_1，$w_2\in R^{(d+1)\times n}$。星型运算公式如下所示。

$$\begin{aligned}&(w_1^{\mathrm{T}}x)\times(w_2^{\mathrm{T}}x)=(\sum_{i=1}^{d+1}w_1^i x^i)\times(\sum_{j=1}^{d+1}w_2^j x^j)= \\ &\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}w_1^i w_2^j x^i x^j=\alpha_{(1,1)}x^1x^1+\ldots+\alpha_{(d+1,d+1)}x^{d+1}x^{d+1}\end{aligned} \tag{4-1}$$

$$\alpha_{(i,j)}=\begin{cases} w_1^i w_2^j & i=j \\ w_1^i w_2^j+w_1^j w_2^i & i\neq j \end{cases} \tag{4-2}$$

其中，i 和 j 对通道进行索引，α 表示每个项目的系数。

星型运算最终可以展开为 $\frac{(d+2)(d+1)}{2}$ 种组合，如式(4-1)所示。在多维层级结构的叠加下，潜在维度可通过递归机制实现接近无限的扩展。设 S_n 为第 n 次迭代的星型运算输出:

$$\begin{aligned}&S_1=\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}w_{(1,1)}^i w_{(1,2)}^j x^i x^j \\ &S_2=\mathrm{W}_{2,1}^{\mathrm{T}}S_1\times\mathrm{W}_{2,2}^{\mathrm{T}}S_1 \\ &S_3=\mathrm{W}_{3,1}^{\mathrm{T}}S_2\times\mathrm{W}_{3,2}^{\mathrm{T}}S_2 \\ &\ldots \\ &S_n=\mathrm{W}_{n,1}^{\mathrm{T}}S_{n-1}\times\mathrm{W}_{l,2}^{\mathrm{T}}S_{n-1}\end{aligned} \tag{4-3}$$

通过多层星型运算的级联叠加，能实现潜在特征维度的指数级扩展。该模块具有结构简约和功能高效的双重优势，其紧凑的网络架构在抑制冗余信息的

同时降低了计算开销，从而为本章提出的研究问题提供了理想的解决方案。具体而言，本章将 YOLOv5s 主干网络中的部分 C3 模块替换为 Star Block 中第 1 次迭代的星型运算输出。

同时，为了更好地提取特征，本章将 CAA 注意力机制引入 Star Block 中，形成 Stat-CAA Block，模块结构如图 4-2 和图 4-3 所示。

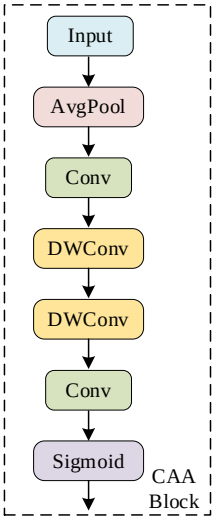


图 4-2 CAA Block 结构

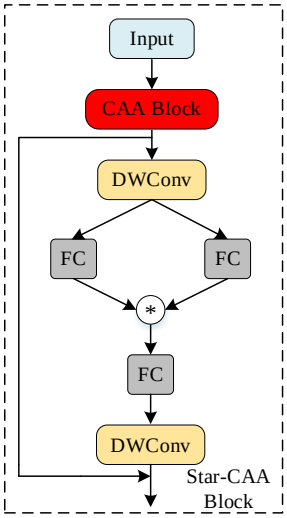


图 4-3 Star-CAA Block 结构

CAA 注意力机制先借助全局平均池化操作提取特征图上的统计信息，再利用 1×1 卷积对局部区域内的关键语义特征进行挖掘，具体表达式如式(4-4)所示。

$$F_{\text{pool}} = \text{Conv}(P_{avg}(X)) \tag{4-4}$$

式中，P_{avg} 代表全局平均池化操作。

然后，利用两个具有填充功能的一维卷积替代 $S \times S$ 的二维深度可分离卷积，保持特征提取能力的同时，有效提升了检测速度。鉴于安全帽的几何形态呈现近似矩形，针对此类目标的图像处理任务，一维卷积能够实现更好的特征提取效果，并且不会显著增加计算复杂度，该过程如下所示。

$$F_{\text{W}} = \text{DWConv}_{1\times s}(F_{\text{pool}}) \tag{4-5}$$

$$F_{\text{H}} = \text{DWConv}_{s\times 1}(F_{\text{W}}) \tag{4-6}$$

式中，$\text{DWConv}_{1\times s}$，$\text{DWConv}_{s\times 1}$ 代表卷积核大小为 $1\times S$ 和 $S\times 1$ 的一维卷积。

最后，特征选择矩阵经过 Sigmoid 函数映射生成 [0,1] 的注意力权重矩阵 $A \in R^{C\times H\times W}$（$C$ 代表通道数，H 代表高度，W 代表宽度）。过程如式(4-7)所示。

$$A=\mathrm{Sigmoid}(\mathrm{Conv}_{1\times1}(F_{\mathrm{H}})) \tag{4-7}$$

4.3.2 损失函数设计

YOLOv5s 的损失函数由分类损失、定位损失和置信度损失组成，其中，通常采用 CIoU Loss[73]作为定位损失来测评真实框(Ground Truth, GT)和预测框(Predicted Truth, PT)之间的距离。CIoU Loss 同时衡量了真实框和预测框二者间的重叠面积、质心距离和宽高比，公式如式(4-8)、(4-9)、(4-10)、(4-11)所示。

$$L_{\mathrm{CIoU}}=1-\mathrm{IoU}+\frac{\rho^2(b,b^{gt})}{\mathrm{c}^2}+\alpha v \tag{4-8}$$

$$v=\frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h})^2 \tag{4-9}$$

$$\alpha=\frac{v}{(1-\mathrm{IoU})+v} \tag{4-10}$$

$$\mathrm{IoU}=\frac{\mathrm{PT}\cap\mathrm{GT}}{\mathrm{PT}\cup\mathrm{GT}} \tag{4-11}$$

其中，b 和 b^{gt} 表示 PT 和 GT 的中心点，$\rho^2(b,b^{gt})$ 表示 PT 和 GT 中心点之间的欧氏距离，c 表示包含 PT 和 GT 的最小长方形的对角线长度，v 表示 PT 和 GT 的长宽比的相似度，w^{gt},h^{gt} 和 w，h 分别表示 PT 和 GT 的宽度和高度，α 为平衡系数。

在多数情况下，CIoU Loss 能很好地计算损失，但在安全帽佩戴检测任务中，仍存在一定的局限性。首先，当预测框和真实框无重叠时，梯度消失问题会导致模型难以有效优化预测框位置；其次，宽高比损失项对极端比例差异的敏感性不足，易因角度偏差引发漏检或置信度下降；最后，中心点距离损失计算受目标尺寸影响，加剧了密集场景下的定位偏差和邻近误检。

为解决安全帽佩戴检测中因背景复杂、遮挡频繁及密集分布导致的漏检误检问题，本章采用 MPDIoU Loss 替换损失函数 CIoU Loss，通过最小化预测框和真实框的关键点距离平方和来优化定位精度，其核心公式如式(4-12)所示。

$$L_{\mathrm{MPDIoU}}=1-IoU+\frac{d_1^2}{h^2+w^2}+\frac{d_2^2}{h^2+w^2} \tag{4-12}$$

其中，d_1 和 d_2 分别表示预测框和真实框左上角点 $(x_1^{pred}+y_1^{pred})$ 和 $(x_1^{gt}+y_1^{gt})$ 的欧氏距离，以及右下角点 $(x_2^{pred}+y_2^{pred})$ 和 $(x_2^{gt}+y_2^{gt})$ 的欧式距离，具体计算公

式如式(4-13)、(4-14)所示。

$$d_1^2 = (x_1^{pred} + x_1^{gt})^2 + (y_1^{pred} + y_1^{gt})^2 \tag{4-13}$$

$$d_2^2 = (x_2^{pred} + x_2^{gt})^2 + (y_2^{pred} + y_2^{gt})^2 \tag{4-14}$$

相较于 CIoU Loss 依赖中心点距离和宽高比损失项的复杂设计，MPDIoU Loss 通过关键点距离优化，直接约束预测框和真实框的几何形状对齐。首先，即使预测框和真实框无重叠，d_1 和 d_2 仍能提供明确的方向梯度，引导模型逐步调整预测框位置，有效缓解遮挡场景下的漏检；其次，通过同步优化左上角和右下角距离，MPDIoU Loss 能自动平衡宽度 $x_2 - x_1$ 和高度 $y_2 - y_1$ 的差异，避免 CIoU Loss 因角度损失项对倾斜目标的过度敏感，减少误检；最后，关键点距离的对称计算使预测框更贴合真实框边缘，降低密集排列时因中心点轻微偏移导致的框重叠或错位。

4.4 实验结果与分析

4.4.1 实验环境与参数设置

如表 4-1 所示，本章训练模型使用的是 Windows 操作系统，CPU 为 Intel Xeon Platinum 8255C @2.50GHz，GPU 为 NAVIDIA GeForce RTX 3090，深度学习框架为 Pytorch 1.11.0，编程语言为 Python 3.8。

表 4-1 实验环境配置

Configuration name	Configuration information
Operating system	Windows
CPU	Intel Xeon Platinum 8255C @2.50GHz
GPU	NAVIDIA GeForce RTX 3090
Deep learning framework	Pytorch 1.11.0
Programming language	Python 3.8

实验设置输入图像大小为 640×640，应用 SGD 优化器，并将动量因子设为 0.937，最大学习率设为 0.01，设定的权重衰减为 0.0005，Batch_size 设为 16，一共训练 300 个 Epoch。参数设置如表 4-2 所示。

表 4-2 实验参数设置

Experimental parameters	Value
Input_shape	640×640
Optimizer	SGD
Momentum	0.937

Max_lr	0.01
Weight decay	0.0005
Batch_size	16
Epoch	300

4.4.2 实验数据集与评价指标

由于公开数据集 SHWD 涵盖的场景类型多样性不足且非施工场景占比较高，同时未佩戴安全帽的负样本主要采集于教室监控场景，难以为复杂环境下的安全帽佩戴检测任务提供充分的训练数据。

本章首先从 SHWD 数据集中筛选出符合需求的图像样本，再通过实地拍摄、网络爬取和视频帧提取等多源数据采集方式构建了 R-SHWD 数据集。在数据标注环节，采用 LabelImg 工具对安全帽目标进行人工标记，并将标注结果分别以 xml、txt 和 json 三种格式存储。标注界面如图 4-4 所示。

图 4-4 LabelImg 标注界面

本章采用适用于 YOLO 模型训练的 txt 文件进行数据存储，将标注完成的标签放置在对应文件夹内，并且确保标签名和数据集中的图像名一一对应。图 4-5 所示的数据标注文件采用了特定格式，每行记录对应一个独立标注对象。具体而言，每条记录包含五个数值特征：首项数值对应标注目标的类别（0表示佩戴安全帽，1 表示未佩戴安全帽），后续四个数值依次对应目标框中心点的横纵坐标值及目标框的宽度和高度，所有空间参数均采用归一化方式处理。

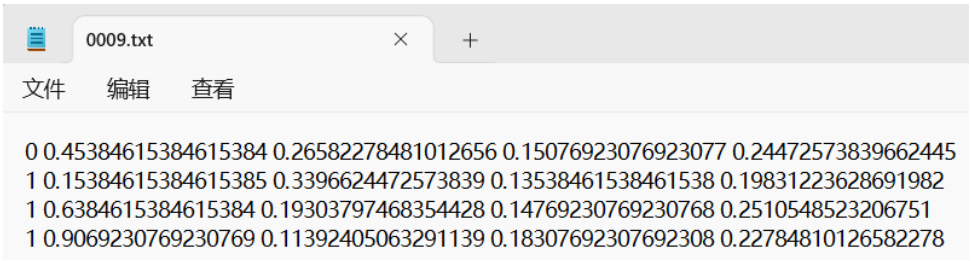


图 4-5 数据标签文件

数据集 R-SHWD 包含 5200 张图像，其中佩戴安全帽的正样本，标记为“Helmet”，共计 16634 个；未佩戴安全帽的负样本，标记为“Head”，共计 12414 个。部分数据如图 4-6 所示。本章设置训练集加验证集与测试集之比为 9:1，其中训练集与测试集之比同样为 9:1。

图 4-6 部分数据

在测评时，本章从客观评价和主观评价两方面来评估改进后 YOLOv5s 模型的性能。以每一类别的 AP 和 mAP 为衡量标准用于客观评价；利用改进前和改进后算法对图像进行检测，将预测到的目标框作为主观评价的依据。

4.4.3 消融实验

为了验证本章算法的有效性，在数据集 R-SHWD 下进行消融实验，实验结果如表 4-3 所示。

表 4-3 消融实验结果

Method	Star-CAA	MPDIoU	mAP@0.5/%	Parameters/10^6	FPS/frames	Model Size/MB
YOLOv5s	×	×	84.70	7.02	294.12	13.71
Experiment 1	√	×	86.40	8.08	200.00	15.82
Experiment 2	×	√	85.60	7.02	285.71	13.71
Experiment 3	√	√	87.00	8.08	196.08	15.82

由表 4-3 可知，本章设计的两种改进均可提升基于 YOLOv5s 模型的安全帽佩戴检测算法的性能。其中，Star-CAA 表示通过引入 Star-CAA Block 代替原算

法主干网络中的部分 C3 模块，MPDIoU 表示采用 MPDIoU Loss 替换损失函数 CIoU Loss。原 YOLOv5s 模型的 mAP 为 84.70%，参数量为 7.02M，FPS 为 294.12，模型大小为 13.71MB。Experiment 1 表示在基线模型的基础上引入 Star-CAA Block。由实验结果可知，mAP 增加了 1.70%，说明 Star-CAA Block 通过引导模型重点关注安全帽相关区域，并赋予更高权重，精确识别出了安全帽，有效降低了漏检率。Experiment 3 表示在 Experiment 1 的基础上再采用 MPDIoU Loss 对网络进行监督学习。由实验结果可知，与基线算法相比，mAP 获得 2.30%的提升，说明 MPDIoU Loss 通过优化预测框和真实框的左上角和右下角关键点坐标，更精准地衡量了边界框的匹配程度，通过调整预测框的位置和尺寸，减少了因背景复杂或密集遮挡导致的漏检误检。

4.4.4 与其他方法进行比较

为了更加直观地评价本章所提算法的性能，对以下算法进行了对比试验，表 4-4 显示了本章算法和其他算法在 R-SHWD 数据集上的比较。

表 4-4 不同算法在 R-SHWD 数据集上的实验结果对比

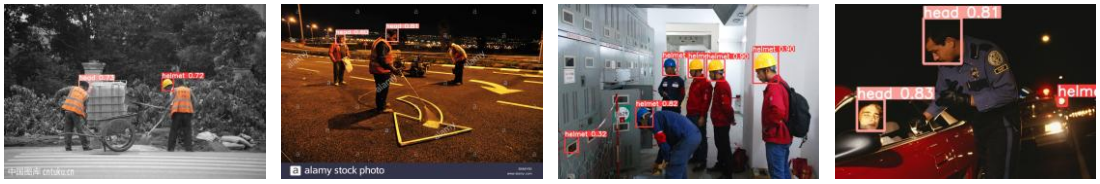
Method	Helmet@0.5/%	Head@0.5/%	mAP@0.5/%	Parameters/ 10^6	FPS/frames	Model Size/MB
YOLOv3-tiny	86.40	82.10	84.20	8.67	476.20	16.62
CenterNet	88.00	81.00	84.69	32.66	87.41	124.93
YOLOv4	89.00	81.00	84.87	64.36	35.84	244.41
YOLOv5s	85.60	83.80	84.70	7.02	294.12	13.71
YOLOv6s	-	-	82.80	18.50	303.03	38.73
YOLOv7-tiny	87.50	84.80	86.10	6.01	212.77	11.70
Ours	88.40	85.60	87.00	8.08	196.08	15.82

由表 4-4 可知，改进后算法在 R-SHWD 数据集上的 mAP 达到了 87.00%，参数量为 8.08M，FPS 为 196.08，模型大小为 15.82MB，相较于其他算法综合效果最佳。与 YOLOv3-tiny 算法相比，改进后算法 mAP 提高了 2.80%，参数量、模型大小均少于 YOLOv3-tiny，模型具备更低的复杂度，在工业生产设备上的适用性更为突出。与 CenterNet、YOLOv4、YOLOv6s 和 YOLOv7-tiny 相比，mAP 分别提高了 2.31%、2.13%、4.20%和 0.90%，同时参数量明显低于 CenterNet、YOLOv4 和 YOLOv6s，FPS 略高于 CenterNet 和 YOLOv4。与基线算法相比，Helmet 和 Head 两类别的检测精度分别提高了 2.80%和 1.80%，mAP 提高了 2.30%，虽然改进后算法的参数量和模型大小略多于基线算法，FPS 也出

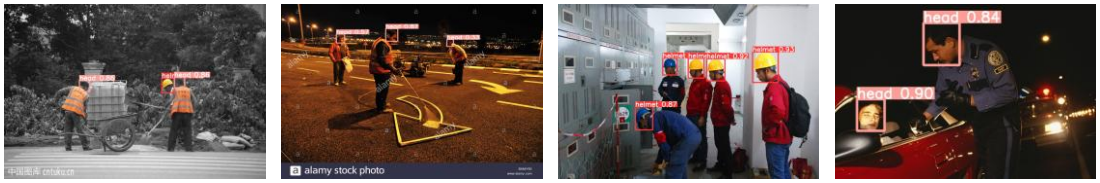
现下降，但在多样化检测场景中均能满足实时性处理要求。

4.4.5 可视化结果分析

为了验证所提算法的可靠性，如图4-7所示，对基线算法和改进后算法在数据集 R-SHWD 上获得的可视化结果进行分析。其中，第一行为基线算法的可视化结果，第二行为改进后算法的可视化结果。



(a) 基线算法可视化结果



(b) 改进后算法可视化结果

图 4-7 部分可视化结果对比

结果显示，本章改进后算法在安全帽佩戴检测场景中的表现优于基线算法。从前两列的可视化结果可以看出，在不同的施工场景下，基线算法都出现了漏检问题，而本章算法能够准确检测出远处未戴安全帽的工人。从第三列、第四列的可视化结果可以看出，在多目标和低光照环境下，基线算法都出现了误检问题，而本章算法通过加强模型的多尺度特征提取能力，有效降低了误检率。

4.5 本章小结

本章提出了一种基于改进 YOLOv5s 模型的安全帽佩戴检测算法。首先，设计了一种引入注意力机制的特征融合模块 Star-CAA Block 代替原算法主干网络中的部分 C3 模块，在多尺度特征融合中动态分配权重，优先强化底层高分辨率特征的细节信息，并结合局部注意力机制捕捉长距离依赖关系，关联被遮挡目标和上下文信息，提高了检测精度，有效缓解了漏检误检问题。其次，通过采用 MPDIoU Loss 替换 CIoU Loss，不仅计算了预测框和真实框的交并比，还额外引入了两框中心点坐标及四个角点坐标的最小化距离损失项，更精准地计算

了两框的空间位置偏差。实验结果显示，在 R-SHWD 数据集上的平均精度均值达到了 87.00%，与基线算法相比，mAP 提高 2.30%。可视化结果表明，本章算法能在充分提取安全帽特征的同时减少背景干扰，提高安全帽佩戴检测的精度，同时有效缓解了狭窄空间及夜间维护作业场景下安全帽漏检误检的问题。

第 5 章 基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法

5.1 引言

在计算机视觉领域内，小目标检测一直是一项颇具挑战的任务，主要由于小目标通常具有低分辨率、复杂背景等难以处理的特性，传统方法往往难以准确识别和定位。在密集钢筋捆扎作业场景下，安全帽佩戴检测是一项重要的应用，对于工人而言，安全帽可以保护头部免受外部伤害。因此，监管安全帽佩戴行为对于减少意外事故的发生具有重要的现实意义。然而，摄像机视角下拍摄到的安全帽通常占据较小尺寸和展现出较低对比度，这使得目前的安全帽佩戴检测模型在提取小目标特征以及精确定位其位置方面面临重大挑战。

论文第三章和第四章算法都利用特征融合策略学习更多的目标特征来完成特征提取，并通过替换损失函数来优化网络参数，这样的设计方式明显提升了基线算法的性能。然而，在复杂多变且具有挑战性的应用场景下，当出现分辨率较低的小目标时，这种特征融合方式没有关注到来自网络更底层的丰富的细节特征，导致针对小目标表征的判别力不足，使得第三章和第四章算法对于小目标的检测效果不够理想。

针对以上问题，本章从挖掘底层特征和优化损失函数的角度出发，同时考虑到监控场景下对检测速度的需求以及实际应用中轻量化模型更易于部署的优势，提出一种基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法。首先，在主干网络中引入 Triplet Attention，通过其独特的跨维度交互机制，显著提升了小目标的检测性能。其次，联合 K-means 聚类算法重新设计更精细的锚框参数，提高先验包围框和不同尺度目标的匹配度。最后，采用 SIoU Loss 优化损失函数的计算方式，以弱化 CIoU Loss 对小目标位置偏差的敏感度，从而提高算法的定位能力。

5.2 算法流程

本章基于 ELAN 骨干网络提出适用于安全帽佩戴检测的算法。通过引入 Triplet Attention、应用 K-means 聚类算法和修改 CIoU Loss 对原始的 YOLOv7-

tiny 模型进行改进，以提高对安全帽小目标的检测精度。

改进后的网络结构如图 5-1 所示，由主干特征提取模块(Backbone)、特征融合模块(Neck)和检测头(Head)三部分组成。首先，主干特征提取模块负责学习输入图像的特征，再输出不同层次的信息。其次，特征融合模块将主干模块输出的深层特征进行融合，最终生成三个检测头，以学习图像三种不同尺度的特征。同时，采用 K-means 聚类算法重新设计预设锚框的大小和长宽比例，实现先验包围框和目标匹配度的提高，并采用 SIoU Loss 替换 CIoU Loss，对网络捕获的特征进行监督，以弱化对小目标位置偏差的敏感度，从而实现安全帽小目标的精准检测。

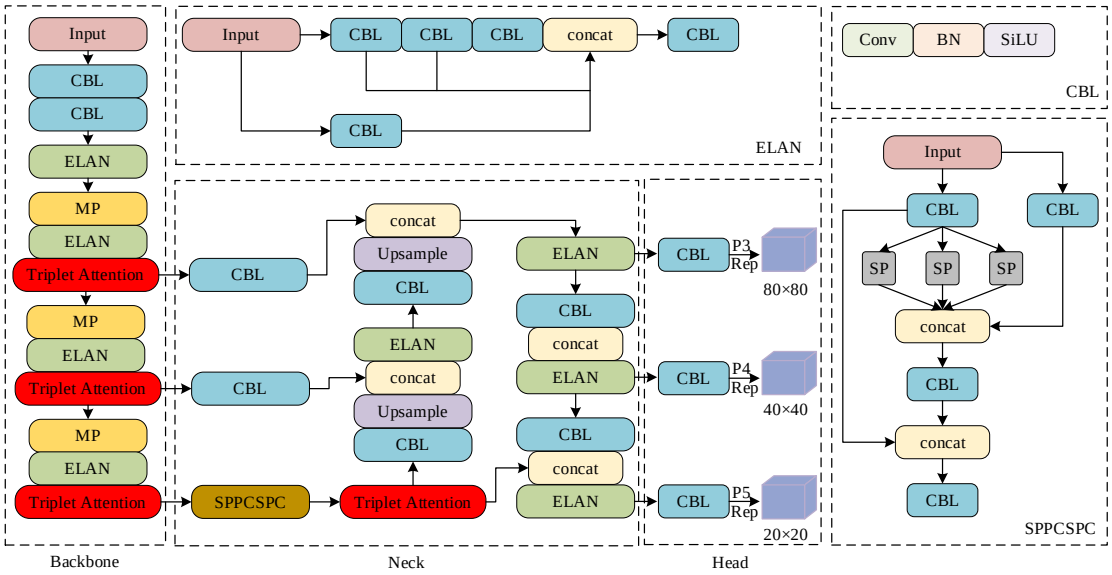


图 5-1 整体网络结构

5.3 算法原理

5.3.1 注意力机制模块

为了提高模型对关键特征的关注，更聚焦训练安全帽小目标的特征。在 Backbone 中引入 Triplet Attention，以可忽略的计算开销，使模型对重要内容和重要位置进行关注，从而提高检测精度，其结构如图 5-2 所示。

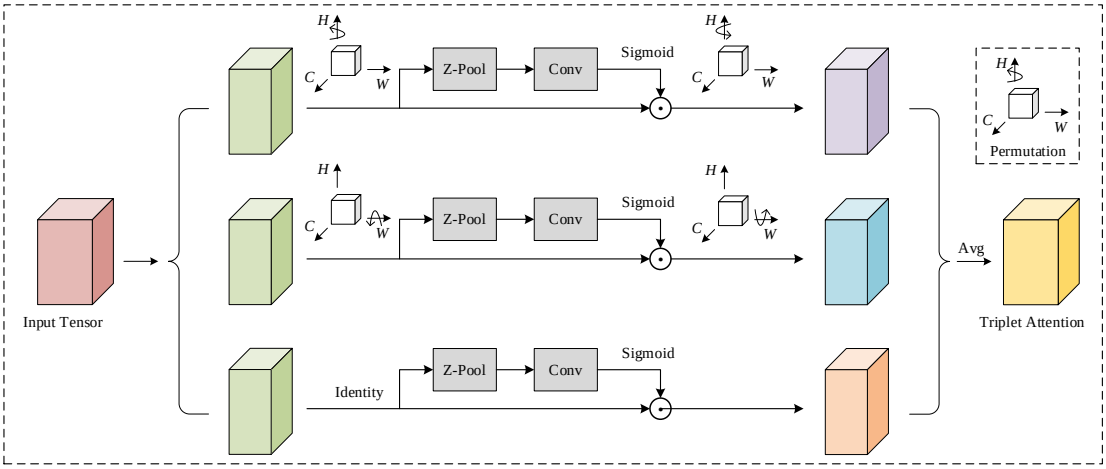


图 5-2 Triplet Attention 结构

第一分支通过分析通道维度 C 和空间维度 H 间的相互作用关系，生成相应的注意力权重矩阵。首先，输入特征经过 Permutation 旋转为 W×H×C；接着，执行 Z 池化操作，再经过 7×7 卷积和 Sigmoid 激活函数；最后，旋转回 C×H×W。第二分支通过分析通道维度 C 和空间维度 W 间的相互作用关系，生成相应的注意力权重矩阵。首先，输入特征旋转为H×C×W；接着，执行与第一分支相同的操作。第三分支负责捕获空间维度之间的依赖关系。首先，输入特征执行 Z 池化操作，再经过 7×7 卷积；接着，通过 Sigmoid 函数得到空间注意力权重。

最后，将三个分支的输出加权融合后，与原始特征相乘完成动态增强。整个过程通过参数共享和轻量化设计降低计算量，始终保留原始特征维度，避免因降维导致小目标细节的丢失。这种多维度协同机制使模型能从通道、空间方向全面筛选细小特征，适用于复杂背景下对安全帽小目标稀疏特征的捕捉。

相较于传统注意力机制，Triplet Attention 的核心竞争力源于其跨维度的全面性和轻量化设计。Triplet Attention 通过三个分支同步建模通道-高度、通道-宽度、高度-宽度的交互，形成多维度协同感知，既能强化小目标的边缘纹理特征，又能抑制背景噪声。其轻量化设计则体现在无降维操作和参数共享，计算复杂度仅为线性增长，显著低于 CBAM 的二次复杂度。此外，空间分支采用分解式方向池化，比传统二维卷积更能精准定位离散小目标的空间分布。

5.3.2 K-means 聚类算法

YOLOv7-tiny 的原始模型有三个检测头，如图 5-1 所示，即 P3、P4 和 P5，三个检测头分别对应三组预设锚框。在基于 YOLOv7-tiny 的目标检测中，三组

锚框是针对 coco 数据集中的目标尺度来设计，并不适用于数据集 R-SHWD。为了更好地适应数据集，本章重新设计了三组锚框，以提高先验包围框和目标的匹配度，采用 K-means 聚类算法对训练数据集中的目标边界框实施聚类分析，以确定最优的先验框尺寸。在对边界框聚类时，以边界框的宽度和高度作为特征，进行归一化处理，公式如(5-1)、(5-2)所示。

$$w = \frac{w_{\text{box}}}{w_{\text{img}}} \tag{5-1}$$

$$h = \frac{h_{\text{box}}}{h_{\text{img}}} \tag{5-2}$$

式中，w_{box}，h_{box} 分别表示边界框的宽度和高度，w_{img}，h_{img} 分别表示输入图像的宽度和高度。对边界框进行 K-means 聚类的步骤是：

步骤 1：选择 K 个边界框作为预设锚框；

步骤 2：利用 IOU 度量将各个方框分配给最近的锚框；

步骤 3：计算各个聚类中全部框的高度和宽度的平均值且更新锚点；

步骤 4：重复以上步骤，直至达到最大迭代次数，或锚框的尺寸不再改变。

K-means 聚类后锚框的大小如表 5-1 所示。

表 5-1 调整后的锚框

检测层	聚类后
80×80	[8,13, 19,28, 31,48]
40×40	[51,70, 67,114, 94,188]
20×20	[146,107, 159,252, 277,350]

5.3.3 损失函数设计

YOLOv7-tiny 的损失函数由三部分组成，分别为置信度损失 L_{obj}、分类损失 L_{cls} 和边界框损失 L_{box}。其中，L_{box} 采用的是 CIoU Loss。相较于 DIoU Loss，CIoU Loss 的优点是引入了预测框和真实框的长宽比，使损失函数更加关注边界框的形状。缺点是由于小目标边界框的绝对尺寸极小，其长宽比的轻微扰动会引发不成比例的高额惩罚，导致模型过度优化长宽比而忽视更核心的中心点偏移或重叠区域不足问题。同时，CIoU Loss 对中心点距离的度量也存在尺度盲区，使得模型难以捕捉细微的位置偏差。

为了解决上述问题，本章将损失函数替换为 SIoU Loss。进一步考虑了预测

框逼近真实框的损失问题，使损失函数从四个层面进行处理，分别为角度损失、距离损失、形状损失和 IOU 损失，显著优化了边界框回归的效率和精度，在小目标检测场景下表现出更强的适应性和鲁棒性。

首先，由于安全帽小目标的边界框面积较小，因此中心点偏移对 IoU 的影响更为显著。而角度损失能够有效引导模型优先优化中心点对齐，减少因方向偏差导致的定位误差。角度损失的计算见公式(5-3)所示，σ 表示真实框和预测框中心点的距离，c_{h} 表示真实框和预测框中心点的高度差。

$$\varphi = 1 - 2 * \sin^2\left(\arcsin\left(\frac{c_{\mathrm{h}}}{\sigma}\right) - \frac{\pi}{4}\right) \tag{5-3}$$

其次，对于安全帽小目标而言，即使中心点偏移只有几个像素，也可能占据目标尺寸的较大比例。而 SIoU Loss 重新设计的距离损失，采用了一种更符合几何直觉的计算方式，通过结合角度信息动态调整距离损失的权重，使得小目标的中心点偏移能够更准确地反映在损失值中，更敏感地捕捉细微偏差，从而提升定位精度。距离损失的计算见公式(5-4)、(5-5)、(5-6)所示，$b_{\mathrm{c_x}}^{\mathrm{gt}}$ 、$b_{\mathrm{c_y}}^{\mathrm{gt}}$ 分别表示真实框的中心点坐标，$b_{\mathrm{c_x}}$ 、$b_{\mathrm{c_y}}$ 分别表示预测框的中心点坐标，c_{w} 、c_{h} 分别表示真实框和预测框最小外接矩形的宽高。

$$\Delta = 2 - \mathrm{e}^{-(2-\varphi)\rho_{\mathrm{x}}} - \mathrm{e}^{-(2-\varphi)\rho_{\mathrm{y}}} \tag{5-4}$$

$$\rho_{\mathrm{x}} = \left(\frac{b_{\mathrm{c_x}}^{\mathrm{gt}} - b_{\mathrm{c_x}}}{c_{\mathrm{w}}}\right)^2 \tag{5-5}$$

$$\rho_{\mathrm{y}} = \left(\frac{b_{\mathrm{c_y}}^{\mathrm{gt}} - b_{\mathrm{c_y}}}{c_{\mathrm{h}}}\right)^2 \tag{5-6}$$

最后，SIoU Loss 的形状损失通过动态调整长宽比的权重，能够解决对小目标几何变化过度敏感的问题。形状损失的计算见公式(5-7)、(5-8)、(5-9)所示，w^{gt} 、h^{gt} 、w 、h 分别表示真实框和预测框的宽高，θ 为形状损失的关注权重。

$$\Omega = \left(1 - e^{-w_{\mathrm{w}}}\right)^{\theta} + \left(1 - e^{-w_{\mathrm{h}}}\right)^{\theta} \tag{5-7}$$

$$w_{\mathrm{w}} = \frac{\left|w - w^{\mathrm{gt}}\right|}{\max\left(w, w^{\mathrm{gt}}\right)} \tag{5-8}$$

$$w_{\mathrm{h}}=\frac{\left|h-h^{\mathrm{gt}}\right|}{\max\left(h,h^{\mathrm{gt}}\right)} \tag{5-9}$$

SIoU Loss 的计算见公式(5-10)所示。

$$L_{\mathrm{SIoU}}=1-\mathrm{IoU}+\frac{\Delta+\Omega}{2} \tag{5-10}$$

5.4 实验结果与分析

5.4.1 实验环境与参数设置

如表 5-2 所示，本章训练模型使用的是 Windows 操作系统，CPU 为 Intel Xeon Platinum 8255C @2.50GHz，GPU 为 NAVIDIA GeForce RTX 3090，深度学习框架为 Pytorch 1.11.0，编程语言为 Python 3.8。

表 5-2 实验环境配置

Configuration name	Configuration information
Operating system	Windows
CPU	Intel Xeon Platinum 8255C @2.50GHz
GPU	NAVIDIA GeForce RTX 3090
Deep learning framework	Pytorch 1.11.0
Programming language	Python 3.8

实验设置输入图像大小为 640×640，应用 Adam 优化器，并将动量因子设为 0.937，最大学习率设为 0.01，设定的权重衰减为 0.1，Batch_size 设为 16，一共训练 300 个 Epoch。参数设置如表 5-3 所示。

表 5-3 实验参数设置

Experimental parameters	Value
Input_shape	640×640
Optimizer	Adam
Momentum	0.937
Max_lr	0.01
Weight decay	0.1
Batch_size	16
Epoch	300

5.4.2 实验数据集与评价指标

本章采用第四章数据集 R-SHWD 进行验证。测评时，本章从客观评价和主观评价两方面来评估改进后 YOLOv7-tiny 模型的性能。以每一类别的 AP 和 mAP

为衡量标准用于客观评价；利用改进前和改进后算法对图像进行检测，将预测到的目标框作为主观评价的依据。

5.4.3 消融实验

本章在数据集 R-SHWD 下进行消融实验，评估所提模型和模块的有效性，消融实验结果如表 5-4 所示。

表 5-4 消融实验结果

Method	Triplet Att	K-means	SIoU	mAP@0.5/%	Parameters/10^6	FPS/frames	Model Size/MB
YOLOv7-tiny	×	×	×	86.10	6.01	212.77	11.70
Experiment 1	√	×	×	88.30	6.01	41.32	11.72
Experiment 2	×	√	×	87.50	6.01	204.08	11.70
Experiment 3	×	×	√	88.00	6.01	208.33	11.70
Experiment 4	√	√	×	89.00	6.01	40.32	11.72
Experiment 5	√	×	√	89.30	6.01	40.32	11.72
Experiment 6	×	√	√	88.80	6.01	217.39	11.70
Experiment 7	√	√	√	89.40	6.01	42.55	11.72

由表 5-4 可知，本章设计的三种改进均可提升基于 YOLOv7-tiny 模型的安全帽佩戴检测算法的性能。其中，Triplet Att 表示通过引入 Triplet Attention 对模型的底层特征进行学习，K-means 表示对模型使用聚类算法进行锚框参数的调整，SIoU 表示将损失函数 CIoU Loss 替换为 SIoU Loss 进行监督学习。原 YOLOv7-tiny 模型的 mAP 为 86.10%，参数量为 6.01M，FPS 为 212.77，模型大小为 11.70MB。Experiment 1 表示在基线模型的基础上引入 Triplet Attention。由实验结果可知，检测精度得到明显提升，mAP 增加了 2.10%，说明 Triplet Attention 有捕捉小目标的空间细节和上下文信息的能力，能有效学习底层特征，同时低计算开销也使其易于集成到实时检测框架中。Experiment 4 表示在 Experiment 1 的基础上再使用 K-means 聚类算法对模型的锚框参数进行优化。由实验结果可知，mAP 增加了 0.70%，说明 K-means 聚类算法有提高先验锚框和目标匹配度的作用，能有效适应数据集中目标的真实分布。Experiment 7 表示在 Experiment 4 的基础上将损失函数 CIoU Loss 替换为 SIoU Loss。由实验结果可知，与基线算法相比，mAP 得到了 3.30%的提升，说明 SIoU Loss 有提升模型检测能力的作用，能有效弱化对小尺度目标位置偏差的敏感度。

5.4.4 与其他方法进行比较

为了更加直观地评价本章所提改进算法的性能，对以下算法进行了对比试验，表 5-5 显示了本章算法和其他算法在 R-SHWD 数据集上的比较。

表 5-5 不同算法在 R-SHWD 数据集上的实验结果对比

Method	Helmet@0.5/%	Head@0.5/%	mAP@0.5/%	Parameters/ 10^6	FPS/frames	Model Size/MB
YOLOv3-tiny	86.40	82.10	84.20	8.67	476.20	16.62
CenterNet	88.00	81.00	84.69	32.66	87.41	124.93
YOLOv4	89.00	81.00	84.87	64.36	35.84	244.41
YOLOv5s	85.60	83.80	84.70	7.02	294.12	13.71
YOLOv6s	-	-	82.80	18.50	303.03	38.73
YOLOv7-tiny	87.50	84.80	86.10	6.01	212.77	11.70
第 4 章算法	88.40	85.60	87.00	8.08	196.08	15.82
Ours	91.10	87.70	89.40	6.01	42.55	11.72

由表 5-5 可知，改进后的 YOLOv7-tiny 模型在 R-SHWD 数据集上的 mAP 达到了 89.40%，参数量为 6.01M，FPS 为 42.55，模型大小为 11.72MB，相较于其他算法综合效果最佳。与 YOLOv3-tiny 相比，改进后算法 mAP 提高了 5.20%，参数量不到 YOLOv3-tiny 的 3/4，模型大小仅为 YOLOv3-tiny 的 2/3，模型相对更简单，更适合布置在生产实践设备上。与 CenterNet、YOLOv4、YOLOv5s、YOLOv6s 和第四章算法相比，mAP 分别提升了 4.71%、4.53%、4.70%、6.60%和 2.40%，同时参数量明显低于 CenterNet、YOLOv4、YOLOv5s、YOLOv6s 和第四章算法，FPS 略高于 YOLOv4。与基线算法相比，Helmet 和 Head 两类别的检测精度分别提升了 3.60%和 2.90%，mAP 提升了 3.30%，改进后算法的参数量和模型大小与基线算法基本一致，虽然 FPS 稍有下降，但能够满足大多数安全帽佩戴检测场景的实时性需求。

5.4.5 可视化结果分析

为了验证所提算法的有效性，如图 5-3 所示，对基线算法和改进后算法在数据集 R-SHWD 上获得的可视化结果进行分析。其中，第一行为基线算法的可视化结果，第二行为改进后算法的可视化结果。

(a) 基线算法可视化结果

(b) 改进后算法可视化结果

图 5-3 部分可视化结果对比

结果显示，本章改进后算法在安全帽小目标佩戴检测场景中的表现优于基线算法。例如，从前三列的对比图可以看出，图像中单一小尺度目标的置信度得分有了不同幅度的提升，锚框的定位也更准确，得益于本章算法将损失函数 CIOU Loss 替换为 SIoU Loss，有弱化小目标位置偏移敏感度的作用，从而加强了模型对小目标位置的度量能力。从第四列的对比图可以看出，图像中多个小尺度目标的置信度得分呈现出差异化增加，同时实现锚点回归精度的同步优化。

5.5 本章小结

本章提出了一种基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法。首先，在主干网络中引入 Triplet Attention，通过多向上下文感知构建目标和周围环境的动态关联，有效缓解了因下采样导致的小目标空间信息丢失问题；其次，联合 K-means 聚类算法重新设计了多尺度分布锚框，增强预设候选框和多尺度目标的统计分布对齐；最后，采用 SIoU Loss 来评估不同尺度目标的分布相似度，减小模型对目标位置偏差的敏感度，提高模型对多尺度目标的定位能力。通过在 R-SHWD 数据集上的实验结果，验证了改进后模型的有效性，相比于基线模型，平均精度均值提升了 3.3%。部署于密集钢筋捆扎作业场景时，可以更好地适应复杂多变的真实环境，在一定程度上缓解了安全帽小目标信息丢失和尺度敏感度差异大引起的检测效果不佳的问题。

第 6 章 总结与展望

6.1 工作总结

安全帽佩戴检测作为目标检测领域的关键研究方向，在工业安防体系具有重要的工程价值。随着卷积神经网络技术的突破，基于深度学习的安全帽佩戴检测技术取得了显著进展，但实际应用仍面临着多重技术挑战。首先，复杂场景下的多模态干扰，包括设备遮挡、人员密集流动及非结构化环境下的视角变化等，对目标特征提取提出了更高要求。其次，安全帽佩戴检测模型的结构优化常涉及多模块集成，这种改进方式在显著增加参数量的同时，也不可避免地引发了推理速度的下降，使得算法在实际场景部署时难以符合实时性需求。因此，论文从这两个角度出发，基于深度信息融合和注意力机制等方面展开研究，具体的研究内容总结如下：

（1）提出了基于改进 CenterNet 模型的安全帽佩戴检测算法。利用特征融合跨层级的聚合方式，将底层高分辨率细节信息和高层抽象语义信息进行加权融合，弥补了单一尺度特征的表征局限性，再通过校准通道维度的特征响应，有效强化了和目标相关的语义信息，同时整体计算开销较低且无需引入额外参数，可无缝嵌入主干网络中而不显著增加模型复杂度。此外，通过引入正样本加权因子，抑制了冗余负样本的响应。该算法有效解决了电力与能源设施巡检场景下安全帽佩戴检测模型检测精度低且参数量过多的问题。

（2）提出了基于改进 YOLOv5s 模型的安全帽佩戴检测算法。设计了一种引入注意力机制的特征融合模块，通过特征加权增强底层高分辨率特征的细节信息，并结合高层语义特征的空间上下文关联，提高对遮挡或低对比度安全帽的定位能力，再在传统 IoU 度量的基础上，引入多维几何约束因子，同步优化预测框和真实框的中心点距离、宽高差异及重叠区域匹配度，有效缓解因定位偏差导致的漏检问题。此外，对数据集进行优化，剔除冗余和无效数据，挑选并补充高质量图像，丰富了数据代表性和场景覆盖度。该算法缓解了狭窄空间及夜间维护作业场景下安全帽佩戴检测模型漏检误检的问题。

（3）提出了基于改进 YOLOv7-tiny 模型的安全帽佩戴检测算法。通过在底

层高分辨率特征中加强局部细节响应，在高层语义特征中增强对小目标的上下文关联，缓解了因下采样导致的特征弱化问题，实现对小目标的精准定位，再采用 K-means 算法对目标宽高分布进行无监督聚类，生成贴合实际小目标尺寸的候选锚框集合，更精准地适配不同场景下安全帽的尺度多样性，减少因锚框和目标尺寸失配导致的漏检问题。同时，通过计算预测框和真实框中心点连线的方向偏差，约束了边界框回归的角度对齐特性，有效缓解预测框在迭代过程中的滑动现象，并重新定义了中心点距离的度量方式，使得中心点偏移量的损失和角度偏差呈负相关，解决了小目标因特征提取不充分导致的定位抖动问题。此外，聚焦于预测框和真实框的宽高比匹配度，通过引入可调控的超参数对宽度和高度的相对差异进行非对称惩罚，使模型在边界框回归过程中更关注目标本身的形态特征。该算法缓解了密集钢筋捆扎作业场景下安全帽佩戴检测模型对小目标检测效果不佳的问题。

6.2 工作展望

目前安全帽佩戴检测技术仍处于发展阶段，在复杂的施工环境和安防领域严苛的检测精度要求下，安全帽佩戴检测将面临更多挑战。虽然论文所提算法在一定程度上改进了安全帽佩戴检测模型的性能，但仍有很多不足之处，有待深入探究和突破。

（1）构建规模更大的安全帽佩戴检测公共数据集。论文算法采用的数据集为公开数据集 SHWD、SHDD、GDUT-HWD 以及重构数据集 R-SHWD，由于数据集规模不大，存在一定局限性。在安全帽佩戴检测研究领域内，训练样本的不充足也限制了安全帽佩戴检测技术的发展。因此，迫切需要新的大规模高质量数据集，提高模型的泛化能力，缩小和现实监控系统数据量之间的差异。

（2）基于无监督的安全帽佩戴检测技术研究。论文研究基于有监督的安全帽佩戴检测算法，所使用数据集中的训练数据均包含标签信息。而在实际监控环境下，检索的安全帽佩戴图像没有数据标注，考虑到安全帽佩戴检测的实用性和应用前景，基于无监督的安全帽佩戴检测技术研究更贴近实际应用。

参考文献

[1] Zhang Z T, Li H, Guo H L, et al. Causal inference of construction safety management measures towards workers' safety behaviors: a multidimensional perspective[J]. Safety Science, 2024, 172: 106432-106444.

[2] Gu J L, Guo F. How fatigue affects the safety behaviour intentions of construction workers an empirical study in Hunan, China[J]. Safety Science, 2022, 149: 105684-105694.

[3] 徐桂芹, 黄晓琼. 建筑业安全生产事故特点与研究热点分析[J]. 建筑安全, 2021, 36(11): 6-9.

[4] Liu S X, Wei Y Y, Yang D J, et al. A systematic review of antecedents of workers' safety behavior: a grounded theory analysis[J]. Safety Science, 2025, 185: 106778-106799.

[5] Huang S C, Zhang S B, Jiao Y F. Ms-VLPD: a multi-scale VLPD based method for pedestrian detection[J]. Expert Systems with Applications, 2025, 271: 126613-126625.

[6] Wu X, Ma Y H, Zhang S J, et al. Yo3RL-Net: a fusion of two-phase end-to-end deep net framework for hand detection and gesture recognition[J]. Alexandria Engineering Journal, 2025, 121: 77-89.

[7] Cui J P, Yuan L, Xiao W D, et al. SEAE: stable end-to-end autonomous driving us -ing event-triggered attention and exploration-driven deep reinforcement learning [J]. Displays, 2025, 87: 102946-102956.

[8] Li Y B, Hu C, Wang R, et al. Hierarchical feature transformation attack: generate transferable adversarial examples for face recognition[J]. Applied Soft Computing, 2025, 172: 112823-112836.

[9] Zhang X H, Zhao J Y, Zhang F, et al. A deep learning method for medical image quality assessment based on phase congruency and radiomics features[J]. Optics and Lasers in Engineering, 2025, 186: 108772-108783.

[10] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with

discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.

[11] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104: 154-171.

[12] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005: 886-893.

[13] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008: 1-8.

[14] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[15] Breiman L. Random forests[J]. Machine Learning, 2001, 45: 5-32.

[16] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[C]. Proceedings of the International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA, 2012: 1097-1105.

[17] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014: 580-587.

[18] He K M, Zhang X Y, Ren S Q, 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.

[19] Girshick R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440-1448.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 2015: 1446-1466.

[21] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object

detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[22] Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 6154-6162.

[23] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Venice, Italy, 2017: 2961-2969.

[24] Pang J, Chen K, Shi J, et al. Libra R-CNN: towards balanced learning for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 821-830.

[25] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 779-788.

[26] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21-37.

[27] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 7263-7271.

[28] Farhadi A, Redmon J. YOLOv3: an incremental improvement[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 1-6.

[29] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 2117-2125.

[30] Law H, Deng J. Cornernet: detecting objects as paired keypoints[C]. Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 734-750.

[31] Duan K, Bai S, Xie L, et al. Centernet: keypoint triplets for object detection[C].

Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Kore, 2019: 6569-6578.

[32] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 321-333.

[33] Wang C Y, Mark Liao H Y, 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, Seattle, WA, USA, 2020: 1571-1580.

[34] Yang J, Fu X, Hu Y, et al. PanNet: a deep network architecture for pan-sharpening[C]. Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 5449-5457.

[35] Zhang J C, Tian M M, Yang Z R, et al. An improved target detection method based on YOLOv5 in natural orchard environments[J]. Computers and Electronics in Agriculture, 2024, 219: 108780-108795.

[36] Ge Z, Liu S, Wang F, et al. Yolox: exceeding yolo series in 2021[J]. arXiv preprint arXiv: 2107.08430v2, 2021.

[37] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv: 2209.02976v1, 2022.

[38] 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/CV F Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 7464-7475.

[39] Wang C Y, Yeh I, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information[J]. arXiv preprint arXiv: 2402.13616v2, 2024.

[40] Wen C Y, Chiu S H, Liaw J J, et al. The safety helmet detection for ATM's surveillance system via the modified hough transform[C]. Proceedings of the IEEE International Carnahan Conference on Security Technology, Taipei China, 2003: 364-369.

[41] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with

discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.

[42] Du S, Shehata M, Badawy W. Hard hat detection in video sequences based on face features, motion and color information[C]. Proceedings of the International Conference on Computer Research and Development, Shanghai, China, 2011, 4: 25-29.

[43] Kelm A, Lauβat L, Meins-becker A, et al. Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites[J]. Automation in Construction, 2013, 36: 38-52.

[44] 刘晓慧, 叶西宁. 肤色检测和 Hu 矩在安全帽识别中的应用[J]. 华东理工大学学报（自然科学版）, 2014, 40(3): 365-370.

[45] 贾峻苏, 鲍庆洁, 唐慧明. 基于可变形部件模型的安全头盔佩戴检测[J]. 计算机应用研究, 2016, 33(3): 953-956.

[46] 李琪瑞. 基于人体识别的安全帽视频检测系统研究与实现[D]. 电子科技大学, 2017: 1-6, 34-59.

[47] Marin-Jimenez M J, Kalogeiton V, Medina-Suarez P, et al. LAEO-net: revisiting people looking at each other in videos[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 3477-3485.

[48] Fang Q, Li H, Luo X, et al. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos[J]. Automation in Construction, 2018, 85: 1-9.

[49] Sun G D, Li C, Zhang H. Safety helmet wearing detection method fused with self-attention mechanism[J]. Computer Engineering and Applications, 2022, 58(20): 300-304.

[50] Saravanan M, Rajini G K. Comprehensive study on the development of an automatic helmet violator detection system (AHVDS) using advanced machine learning techniques[J]. Computers and Electrical Engineering, 2024, 118(A): 109289-109322.

[51] Li M S, Han Q P, Zhang T Y, et al. Safety helmet detection method of improved

SSD[J]. Computer Engineering and Applications, 2021, 57(8): 192-197.

[52] 王海瑞, 赵江河, 吴蕾等. 针对 CenterNet 缺点的安全帽检测算法改进[J]. 湖南大学学报（自然科学版）, 2023, 50(8): 125-133.

[53] Li H B, Wu D C, Zhang W M, et al. YOLO-PL: helmet wearing detection algorithm based on improved YOLOv4[J]. Digital Signal Processing, 2024, 144: 104283-104293.

[54] Lee J Y, Choi W S, Choi S H. Verification and performance comparison of CNN-based algorithms for two-step helmet-wearing detection[J]. Expert Systems with Applications, 2023, 225: 120096-120109.

[55] Han G J, Wang R X, Xie W Y, et al. Night construction site detection based on ghost-YOLOx[J]. Connection Science, 2024, 36(1): 2316015-2316042.

[56] Li Y, Xu H Y, Zhu X Z, et al. THDet: a lightweight and efficient traffic helmet object detector based on YOLOv8[J]. Digital Signal Processing, 2024, 155: 104765-104778.

[57] Zhang Y, Huang S Z, Qin J Z, et al. Detection of helmet use among construction workers via helmet-head region matching and state tracking[J]. Automation in Construction, 2025, 171: 105987-106003.

[58] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.

[59] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning[C]. Proceedings of the International Workshop on Artificial Neural Networks, Malaga-Torremolinos, Spain, 1995: 195-201.

[60] Glorot X, Bordes A, Bengio Y, et al. Deep sparse rectifier neural network[C]. Proceedings of the International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 2011: 315-323.

[61] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 770-778.

[62] Tsung Y L, Michael M, Serge B, et al. Microsoft coco: common objects in context[J]. Lecture Notes in Computer Science, 2014, 8693(1): 740-755.

[63] Wang W, Xie E, Song X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 8440-8449.

[64] Wang J B, Wu Y X. Safety helmet wearing detection algorithm of improved YOLOv4-tiny[J]. Computer Engineering and Applications, 2023, 59(4): 183-190.

[65] Han J, Liu Y C, Li Z P, et al. Safety helmet detection based on YOLOv5 driven by super-resolution reconstruction[J]. Sensors, 2023, 23(4): 1822-1835.

[66] Jiao X, Li C, Zhang X, et al. Detection method for safety helmet wearing on construction sites based on UAV images and YOLOv8[J]. Buildings, 2025, 15(3): 354-366.

[67] Sandler M, Howard A, Zhu M L, et al. Mobilenetv2: inverted Residuals and Linear Bottlenecks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 4510-4520.

[68] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020. 1580-1589.

[69] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]. Proceedings of the International Conference on Machine Learning, Lille France, 2015: 448-456.

[70] Song R J, Wang Z M. RBFPDet: an anchor-free helmet wearing detection method[J]. Applied Intelligence, 2023, 53(5): 5013-5028.

[71] 张玉涛, 张梦凡, 史学强等. 人员安全帽佩戴轻量化检测方法研究[J]. 安全与环境学报, 2023, 23(2): 474-480.

[72] Liang H, Seo S Y. UAV low-altitude remote sensing inspection system using a small target detection network for helmet wear detection[J]. Remote Sensing, 2023, 15(1): 196-215.

[73] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]. Proceedings of the AAAI 2020-34th AAAI Conference on Artificial Intelligence, New York, USA, 2020: 12993-13000.

致谢

行文至此落笔为终，二十余载求学路，或顺利或坎坷。回首这漫长而短暂的研究生旅程，曾笑过、哭过、憧憬过也迷茫过，目光所及，皆是回忆。不知不觉已走过三年，在安徽工程大学的学习生活使我受益匪浅。纵有万般不舍，仍心存感激。借此，向我的家人、老师、朋友和同学们致以最真诚的感谢。

桃李不言，下自成蹊。首先我要感谢我的导师王凤随教授，研究生生涯很幸运成为王老师的学生。老师求真务实的科研作风、严以律己的人生态度以及温和谦虚的处事风格深深影响着我，使我受益颇多。言传不如身教，感谢您三年以来对我的栽培，教我专业能力、受以学术素养。从论文的选题、构思、撰写、修改和定稿，离不开老师的悉心指导，一遍遍地审阅、以及逐字逐句地斟酌。每每遇到困难、迷茫无所时，与老师交流后总能扫清我心中忧虑，豁然开朗。遇到一位认真负责的好老师不易，再次感谢王老师三年的淳淳教诲与帮助。祝愿老师万事顺遂，喜乐平安。

父母之爱子，则为之计深远。感谢我的爸爸妈妈。感谢您们不求回报的付出，尊重我成长道路上的每一次选择，做我最坚实的后盾，让我有义无反顾向前地勇气。感谢家人永远尊重我，支持我，让我心无旁骛地学习，给我幸福温暖的家庭环境，是我永远坚定的避风港。希望时间走的慢一些，让我陪伴你们的时间长一点，以后努力成为你们的依靠。祝愿我的家人永远身体健康，幸福平安。

岁月虽轻浅，时光亦潋滟。特别感谢 404 课题组的全体同学。感谢师兄师姐对我的照顾，在课题研究上提供细心细致的帮助与支持，在学习中给我建议，在撰写论文给我指导。感谢同门们互帮互助一起解决实验困难，共同奋斗，为我的实验生活增添了许多回忆。感谢实验室所有小伙伴三年的陪伴，使我的研究生学习生活过得非常充实快乐。愿未来万里鹏程，一切顺利！

浅喜似苍狗，深爱如长风。感谢我的女友张娜，从本科到研究生，这一路走来你给予了我无数的包容与关怀。很幸运与你相遇，很庆幸这一路有你的陪伴，互相见证彼此从青涩懵懂到成熟的蜕变。感谢你一直的支持与鼓励，感谢在读研三年的起起落落中有你疏导我的坏情绪，做我的精神支柱。感谢你坚定

不移的选择，给予我无限的爱与包容。人生路漫漫，有你陪伴的日子平凡且浪漫，愿未来我们携手共进退，一起奔赴美好的未来。

前程可奔赴，岁月可回首。感谢一路走来从未放弃，努力向前的自己。人生的路走的每一步都算数，二十余载求学路中每个刻苦拼搏的日夜，自我治愈、坚持的瞬间，都在不经息间化为我前行的力量，使我收获坚韧、成熟与责任。愿自己能一直保持热枕，山高路远，继续前行！

最后，感谢所有参加本论文评审和答辩的老师们，谢谢你们的辛苦付出！

攻读学位期间取得的学术成果情况

1、录用学术论文情况：

[1] 方楠, 王凤随. 基于改进 CenterNet 网络的安全帽佩戴检测算法. 传感技术学报.

尚德教学 唯实惟新