分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2210910102

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　面向工业场景的防毒面具佩戴检测方法的研究

论 文 作 者　汪邦荣
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：2024 年 5 月 23 日

分类号：TP391　　单位代码：10363
密　级：公开　　学　　号：2210910102

题　　目　面向工业场景的防毒面具佩戴检测方法的研究

英文并列

题　　目　Research on Detection Method of Wearing Gas Masks for Industrial Scenarios

学生姓名：汪邦荣

指导教师：汪军 教授

专　　业：软件工程

研究方向：领域软件工程

论文答辩日期：2024 年 5 月 24 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：汤野荣

日期：2024年5月24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：汪郭荣　　指导教师签名：汪军

日期：2024 年 5 月 24 日　　日期：2024 年 5 月 24 日

面向工业场景的防毒面具佩戴检测方法的研究

摘 要

在工业生产中，防毒面具作为一种常用的防护用具，能够有效地过滤有害气体、粉尘和颗粒物，为工人的呼吸系统提供有力的保障。然而，由于部分工人安全意识缺失，未依照规定佩戴防毒面具，从而导致一系列诸如中毒和呼吸道损伤等安全事故。因此，构建一个能够实时检测工人防毒面具佩戴情况的系统，对预防安全事故至关重要。该系统能够对工人的面部区域进行实时监控，识别其是否正确佩戴了防毒面具。若发现有工人未按要求佩戴，系统将即刻发出警报，提醒工人采取必要的防护措施。该系统不仅能实时监控工人的防护状况，还能显著提升工人的安全意识，有效降低相关事故的发生率。

该文根据实际应用中的需求，深入分析了工业场景中防毒面具佩戴检测所面临的挑战，并采用基于深度学习的目标检测模型来构建防毒面具检测系统，同时针对这些挑战对模型进行了一系列优化。该文的主要工作和创新点如下：

(1) 通过对工业现场的实地调研，结合检测需求设计了合理的标注策略，进而构建了一个适用于工业场景的防毒面具数据集。根据数据集的统计信息，对工业场景中的检测难点进行了分析，其中包括目标尺度和类别不均衡等问题。同时，为了提升模型训练效果，增强模型的泛化能力，对数据集进行了离线数据增强操作。

(2) 为解决工业场景下防毒面具检测中存在的多尺度检测和类别不均衡等问题，以 Faster R-CNN 为基础进行了一系列的优化。首先，针对防毒面具检测场景中的多尺度问题，在 Faster R-CNN 中引入了特征金字塔网络。特征金字塔网络通过在不同层级提取特征并进行融合，有效地捕捉了不同大小目标的特征信息，使模型能够更加准确地检测小目标。其次，采用了在线难例挖掘算法来缓解数据集中的类不平衡问题。该算法通过自动选择难以分类的样本进行训练，使得模型更加关注那些容易被误分类的样本，以此提高模型的分类能力。最后，在训练过程中使用了 Mixup 和 Mosaic 对数据进行了扩充，使得模型能够更好地适应不同场景的变化。

(3) 为满足工业应用中的实时检测要求，解决复杂背景干扰导致的误检测、数据集缺乏多样性以及类别不均衡的问题，提出了一种基于 YOLOv5 的改进方法。首先，为了增强主干网络的提取能力，将高效的 ELAN 网络与 Efficient SE 模块相结合，使模型在保持轻量的同时提高对复杂背景的辨别能力。其次，提出了多阶段数据增强策略。该策略在训练过程中分阶段运用 Mosaic9 和 Mosaic4 生成更多数量的样本数据，从而提高了数据的多样性。最后，使用了 Varifocal loss 来缓解数据集中存在的类别不均衡问题。该损失函数能够更好地平衡不同类别的数据，使模型在训练过程中更关注数量较少但重要的类别。

关键词：防毒面具佩戴检测，深度学习，Faster R-CNN，YOLOv5

RESEARCH ON DETECTION METHOD OF WEARING GAS MASKS FOR INDUSTRIAL SCENARIOS

ABSTRACT

In industrial production, gas masks as a commonly used protective equipment, can effectively filter out harmful gases, dust and particulate matter, provide a strong guarantee for the respiratory system of workers. However, the lack of safety awareness among some workers, who do not wear gas masks as required, has resulted in a series of safety accidents, such as poisoning and respiratory injuries. Therefore, it is necessary to construct a system that can detect the wearing of workers' gas masks in real time to prevent safety accidents. The system can monitor the face area of workers in real time and recognize whether they are wearing gas masks correctly. If a worker is found to be not wearing the mask as required, the system will immediately issue an alarm to remind the worker to take the necessary protective measures. The system can not only monitor the protection status of workers in real time, but also significantly improve the safety awareness of workers and effectively reduce the incidence of related accidents.

According to the needs of practical applications, this dissertation deeply analyzes the challenges faced by the detection of gas mask wearing

in industrial scenes, and uses an object detection model based on deep learning to construct a gas mask detection system. At the same time, a series of optimizations have been made to the model for these challenges. The main work and innovations of this dissertation are as follows:

(1) Through on-site research of the industrial site and according to the detection requirements, a reasonable labeling strategy is formulated, enabling the construction of a data set of gas masks for industrial scenarios. Based on the dataset's statistical information, we analyzed the detection challenges in industrial scenarios, including issues such as target scale and category imbalance. Additionally, to enhance the model's training and generalization capabilities, offline data augmentation operations were performed on the dataset.

(2) In order to solve the problems of multi-scale detection and class imbalance in gas mask detection in industrial scenarios, a series of optimizations are carried out based on Faster R-CNN. Firstly, to address the multi-scale issue in gas mask detection scenarios, we incorporated a Feature Pyramid Network into Faster R-CNN. The FPN effectively captures the feature information of targets of different sizes by extracting and fusing features at different levels, enabling the model to accurately detect small targets. Secondly, the online hard example mining algorithm was employed to alleviate the class imbalance issue in the dataset. The algorithm improves the model's classification ability by automatically

selecting difficult-to-classify samples for training, making the model pay more attention to samples that are prone to misclassification. Finally, Mixup and Mosaic were used to augment the data during the training process, allowing the model to better adapt to changes in different scenarios.

(3) To address the issues of false detections caused by complex background, lack of diversity in datasets, and class imbalance in industrial applications, an improved method based on YOLOv5 is proposed. First, in order to enhance the extraction ability of the backbone network, the efficient ELAN network is combined with the Efficient SE module, so that the model can improve the ability to discriminate complex backgrounds while maintaining light weight. Second, a multi-stage data enhancement strategy is proposed. This strategy employs Mosaic9 and Mosaic4 to generate a larger number of sample data in stages during the training process, thus increasing the diversity of the data. Finally, Varifocal loss was used to alleviate the class imbalance issue existing in the dataset. This loss function can better balance the data of different categories, making the model more attentive to those categories with fewer samples but greater importance during training.

KEY WORDS: gas mask wearing detection, deep learning, faster r-cnn, yolov5

目录

第 1 章 绪论

1.1 研究背景及意义

近年来，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。我国”十四五”规划纲要明确提出未来五年要提高安全生产水平，并对完善落实安全生产责任制度、加强监测预警和执法、深入推进重点领域安全整治等做出了部署。多年来，我国安全生产水平稳步提高，在工业化、城镇化快速发展的同时保持了事故总量、较大事故、重特大事故数量持续下降。但从总体上判断，我国安全生产形势仍然复杂严峻，目前正处于事故持续下降的“多发平台期”，多数重点行业领域高风险特征没有根本改变，城市安全风险不断积累和显现，农村安全隐患日益增多，劳动者安全素质和全社会安全意识还有待提升，安全生产仍处于“爬坡过坎期”，防范生产安全事故仍然任重道远。

工业安全包括预防各种工业危害、职灾事故和职业病。从短期和长期来看，这些工业危害、职灾事故和职业病对工人，甚至对其家庭乃至全社会都会产生重大影响，包括人身伤亡、身体健康恶化、认知能力以及幸福感下降、社会和经济破坏、财产损失以及生存环境恶化。此外，这类危害还会降低企业的生产率和效率，且可能会削弱竞争力，并损害企业在供应链上的商誉，从而对经济和社会产生更广泛的影响。然而，工业活动永远不可能完全摆脱自然灾害和人为危害，因此必须尽可能全面地了解这些风险，才能告知监管机构，并根据现有的实践和技术采取适当的减灾措施[1]。

目前，许多工作已经尝试将成熟的管理模式与先进的人工智能技术相结合，以提高工人的职业安全意识，预防事故发生，并建立有效的预警系统。典型的应用如安全状态监测[2-3]、危险动作识别[4]等。但在这些工作中，对呼吸安全的关注并不多。特别是在工业生产中，由于防护措施不健全，或生产、运输、储存过程中发生意外，化学物质可污染环境，危害人体，引起疾病。有毒有害的物质以气体、蒸汽、雾、烟、粉尘等形式存在于空气中，主要通过呼吸道侵入人体，而鼻、咽、喉是最先受害者的器官，其受损程度与物质的化学特性、剂量和接触时间等因素有关。对于在这种环境中长时间工作的工人来说呼吸防护设备必不可少。防

毒面具作为一种重要的呼吸防护用品被普遍应用在有毒有害气体粉尘颗粒物等工业生产环境中。然而，由于工人和管理人员都容易忽视呼吸安全，一些因未佩戴防毒面具的呼吸系统疾病和事故时有发生。典型的疾病有矽肺、尘肺、石棉肺和职业性哮喘等。因此，为了提高工人的安全防护意识、预防安全事故发生，建立一个稳定且精确的防毒面具佩戴检测系统是十分有应用价值的。

1.2 国内外研究现状

在计算机视觉研究中，目标检测算法一直备受关注。与图像分类算法相比，目标检测的难度更大。图像分类算法只需对目标进行分类，而目标检测则需要结合目标定位和目标分类，运用图像处理技术和机器学习等多方面知识，在图像中确定感兴趣的对象。其中，目标分类主要用于判断图像中是否存在特定物体，而目标定位则是要确定这些物体的具体位置，并用矩形框准确标注。这意味着算法不仅要准确判断目标的类别，还要精确给出每个目标的位置。

传统的目标检测算法需要通过人工选取的特征来对物体进行检测并分类。人工提取的特征主要针对某些特定对象，需要手动设置一些用于提取特征的算子。然而现实场景往往复杂多变，例如光照、遮挡、尺度变化等，人工设置的算子难以在复杂的工业环境中保持较好的稳定性。近年来，随着深度学习的不断发展，以卷积神经网络为基础的目标检测技术在计算机视觉领域取得了显著的突破。目标检测算法的发展历程如图 1-1 所示。

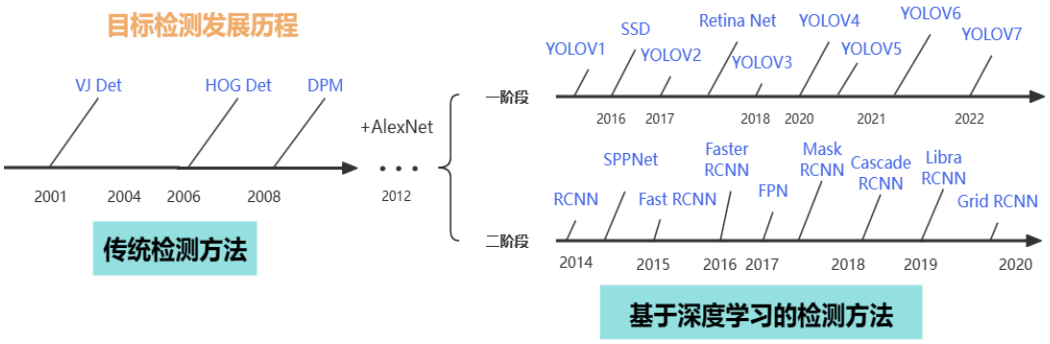


图 1-1 目标检测算法发展历程

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

传统目标检测的过程可以概括为以下三个步骤：首先，通过滑动窗口或图像金字塔方法生成候选区域；然后，利用手工设计的特征提取器从候选区域中提取特征；最后，使用分类器或回归器对提取的特征进行分类或定位，以实现目标检测。

在 2004 年，Lowe 等[5]提出了一种基于特征点的方法 SIFT (Scale Invariant Feature Transform)。该方法在图像特征提取和匹配方面具有较好的性能，但存在计算复杂度高，难以处理光滑和模型图像的缺点。因此，在 2005 年，Dalal 等[6]提出了梯度直方图算法 HOG (Histogram of Oriented Gradient)。该算法能够有效地提取图像中目标的形状和边缘信息，对光照变化和部分遮挡具有一定的鲁棒性，但对光照和背景变化敏感，尺度不变性较差。在 2008 年，Pedro 等[7]在 HOG 算法的基础上提出了 DPM (Deformable Parts Model)算法。该算法通过将目标分解为多个部件，并建立部件之间的关系模型，从而有效地进行目标检测。这种方法能够捕捉目标的多样性和变形，具有较好的检测性能。

传统的目标检测方法虽然在一些任务上取得了较好的性能，但也存在较多的缺点。首先，特征提取器的设计需要人工选择适当的特征，这限制了其对复杂场景和多种目标的适应能力。其次，分类器的性能高度依赖于训练数据的质量和数量，对于少量或者不平衡的数据集，分类器可能会表现不佳。最后，传统目标检测方法往往需要执行多个独立的步骤，导致计算复杂度较高，速度较慢。这些问题制约了它们在实际场景中的应用。

(2) 基于深度学习的目标检测算法研究现状

近年来，随着深度学习的蓬勃发展，基于深度学习的目标检测技术已广泛应用于各行各业。相较于传统的目标检测方法，基于深度学习的目标检测算法能够通过大规模数据集进行训练，并自动学习到目标的特征，避免了人工提取特征。同时，它还能保证良好的检测速度和准确性，并具备较强的普适性。

2012 年，Krizhevsky 等[8]首次在 Image Net LSVRC-2012 图像分类大赛中引入了卷积神经网络(AlexNet)，并以出色的表现脱颖而出。随后，卷积神经网络在图像分类领域得到广泛应用，并涌现出一系列具有里程碑意义的经典模型，如

VGG[9]、GoogLeNet[10]、ResNet[11]和 DenseNet[12]等。这些模型的不断发展不仅推动了图像分类方法的更新，也为目标检测领域带来了新的进步。

2014 年，Girshick 等[13]首次将卷积神经网络（Convolutional Neural Networks, CNN）应用于目标检测，并提出了两阶段检测模型——R-CNN（Region Convolutional Neural Network）。该算法以选择性搜索算法为基础，生成大量与目标相关的建议区域，然后借助 CNN 从这些区域中提取特征，最后使用支持向量机进行分类器训练。同年，He 等[14]为了解决输入图像大小固定的限制问题，提出了 SPPNet。该网络引入了空间金字塔池化，能够让网络适应不同大小的图像输入。2015 年，Girshick 等[15]针对 R-CNN 的检测速度过慢问题提出了 Fast R-CNN。与 R-CNN 相比，该算法的主要改进之处在于将输入图片送入卷积神经网络中生成特征图，并将选择性搜索算法生成的建议区域映射到特征图上进行分类与坐标回归，从而极大地加快了检测速度。2016 年，Girshick 等[16]通过对 R-CNN 和 Fast R-CNN 方法进行深入研究和总结提出了端到端检测模型——Faster R-CNN。该网络摒弃传统滑动窗口和选择性搜索方法，采用 Region Proposal Network[16]生成检测框，减少冗余计算，大幅度提高模型性能。在后续的研究中，大量学者对 Faster R-CNN 进行了不断的优化和完善，相继涌现出了 Mask R-CNN[17]、Cascade R-CNN[18]、Dynamic R-CNN[19]和 Sparse R-CNN[20]等改进模型。这些模型在保持 Faster R-CNN 的高效性能的同时，针对不同的应用场景和需求进行了优化，进一步扩展了 Faster R-CNN 的应用范围。

在 2016 年，Redmon 等[21]提出了一阶段检测算法——YOLO。YOLO 算法采用了全新的思路来进行目标检测。它将整个图像作为输入，并通过一个卷积神经网络直接输出目标的类别和位置信息。这意味着 YOLO 能够在一次前向传播中同时完成对多个目标的检测，而无需使用滑动窗口或者候选区域生成等复杂步骤。相比于 R-CNN 系列算法，YOLO 具有更快的速度和更高的实时性能，但精度不及两阶段算法。同年，Liu 等[22]提出了 SSD(Single Shot MultiBox Detector)算法。该算法通过在不同尺度的特征图上进行目标检测，可以同时检测不同大小的目标，并实现了对目标的有效检测和定位。2017 年，Redmon 等[23]研究人员对 YOLO 算法进行了改进，并推出了 YOLOv2 算法。YOLOv2 采用了多种方法优化 YOLOv1，例如锚框、多尺度训练和锚框聚类等。其中，锚框的使用可以帮助模

型更准确地识别目标的位置和大小，多尺度训练可以增强模型的鲁棒性和泛化能力，而锚框聚类则能进一步提高模型的准确性和效率。2018 年，Redmon 等[24]提出了 YOLOv3。YOLOv3 是 YOLO 系列的第三个版本，与之前的算法相比，其精度有了显著的提高。YOLOv3 提出了一种新的骨干网络 DrakNet53[24]，并引入了特征金字塔网络[28]，有效的提高了模型多尺度检测能力。2020 年，Alexey Bochkovskiy 及其团队发布了 YOLOv4[25]。YOLOv4 是对 YOLOv3 的整体改进，它使用了目标检测算法中的许多技巧，如使用跨阶段局部连接(CSPNet)[26]改进 DarkNet53，在多尺度融合模块中使用 SPP[27]和 PAFPN[29]，并提出了数据增强方法 Mosica。紧接着，ultralytics 公司发布了 YOLOv5 模型。这款新的模型是 YOLO 系列在工业应用的一个巨大进步。YOLOv5 相比于 YOLOv4，通过减小模型的复杂度并提高实时性，从而实现更快、更有效的目标检测。近几年，YOLOX[30]、YOLOv6[31]、YOLOv7[32]相继问世，YOLO 系列模型不断演进与完善，在工业生产监控、智能交通管理、人驾驶汽车等众多领域都取得了令人瞩目的成果。

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还比较少。因此，本研究的目的是填补这方面的研究空白。为了实现这一目标，本研究开展了广泛的文献研究，涉及人脸检测、工业生产中安全帽检测和公共卫生防护中口罩检测等相关领域。这些相关工作所提出的问题和解决思路，为本研究提供了珍贵的经验和参考价值。

在人脸检测领域，Tang 等[33]针对检测中不受控制条件下小的，模糊的，和部分遮挡的人脸，提出了一个上下文辅助的人脸检测器 PyramidBox。PyramidBox 设计了一种新的上下文锚框来监督学习小、模糊和部分遮挡人脸的上下文特征的监督信息。在特征融合上，该工作提出了一种低层特征金字塔网络整合高层的语义信息和低层的面部特征，它可以使模型检测不同尺度的人脸。Qi 等[34]提出了基于 YOLOv5 的人脸检测模型——YOLOv5-Face。该工作将人脸检测作为一个一般的目标检测任务来看待的，并在 YOLOv5 的基础上添加五个人脸关键点回归，并对关键点回归使用了 Wing loss 进行约束。Yu 等[35]针对现实场景中检测具有高度变化的人脸、姿势、遮挡、表情、外观和照明的挑战，提出了一种基于 YOLOv5 的实时人脸检测器 YOLO-FaceV2。该工作设计了一个名为 RFE 的感受

野增强模块来增强小人脸的感受野，并使用 NWD Loss 来弥补 IoU 对微小物体位置偏差的敏感性。针对人脸遮挡问题提出了一个名为 SEAM[35]的注意力模块并在训练中引入了排斥损失。王等[36]提出了一种融合多尺度特征的轻量级人脸检测算法。该算法首先利用现有的轻量级主干特征提取网络编码输入图像。然后，利用提出的颈部网络扩张特征图感受野，并将含有不同感受野的多尺度信息融合至单级特征图中。最后，利用提出的多任务敏感检测头对该单级特征图进行人脸分类、回归和关键点检测。

在工业安防领域中，基于深度学习的目标检测方法已经被广泛使用。徐等[37]针对工业安防中的安全帽佩戴检测算法对部分遮挡、尺寸不一和小目标存在检测难度大、准确率低的问题，提出了基于改进的 Faster R-CNN 和多部件结合的安全帽佩戴检测方法。相比于原始 Faster R-CNN，该方法检测准确率提高，对环境的适应性更强。李等[38]为解决复杂场景中安全帽检测的小目标和类别不均衡问题提出了一种基于 Faster R-CNN 改进方法。该方法通过增加锚点来提高对小目标的检测能力，同时使用 Focal loss 来替代原先的损失函数来缓解数据集中存在的类别不平衡问题。Li 等[39]针对开源头盔检测数据集 SHWD 的不足，构建了一个更为丰富的安全帽检测数据集。该数据集包含了之前数据集中所缺乏的各种工业场景。此外，该研究重新考虑了 YOLO 系列的样本选择方法，并在训练过程中提出了分层正样本选择机制，以提升 YOLOv5 的拟合能力。最后，受视频连续帧目标检测的启发，提出了一种基于盒密度的后处理算法，以有效抑制误检的出现。祁等[40]针对安全帽佩戴检测在复杂背景和密集小目标场景下的问题，提出了一种改进 YOLOv5 算法。该算法通过串联的混合池化和增加浅层输出来增强特征提取和目标位置信息的表达，同时利用坐标注意力机制扩大感受野，提高安全帽目标检测的准确率。

在公共健康防护领域，Mercaldo 等[41]提出了一种基于深度学习的自动检测人们是否佩戴口罩的方法。该方法利用迁移学习和 MobileNetV2 模型，能够准确识别图像或视频流中的口罩违规行为，并以相对概率定位人脸口罩检测相关区域。为了推进公共卫生事业，Sethi 等[42]设计了一种高精度、实时的技术，可以有效地监测公共场所未佩戴口罩的个体。该技术整合了一阶段和两阶段检测器，以实现较短的推理时间和更高的准确率，并应用迁移学习概念来融合多个特征图中具

有高层语义信息的 ResNet50[11]模型作为基准。此外，在口罩检测过程中还引入了边界框变换方法以提升定位性能。针对佩戴口罩等存在遮挡条件下的人脸检测问题，李等[43]提出了一种多尺度注意力学习 Faster R-CNN 人脸检测模型。该方法通过引入分组残差模块并结合空间-通道注意力结构改进模块来获取更细粒度特征，并使得模型能够自适应地学习不同目标特征。另外，李小波等[44]针对公共场所中目标尺度小且密集问题，提出了一种注重注意力机制并融合多尺度权重的口罩检测算法。该算法基于 YOLOv5s 骨干网络，在其内部引入 4 种不同类型注意力机制来抑制无关信息从而增强特征图表达能力，并显著提升对小尺寸目标物体进行有效检测之能力。

1.3 主要研究内容

从上述研究现状和发展趋势可以看出，基于深度学习的目标检测技术在安全防护领域具有良好的应用前景。使用深度学习技术构建一个端到端防毒面具佩戴检测模型，可以及时，准确的获取工人防毒面具佩戴情况，对提高工人职业安全意识、预防安全事故、提高企业管理能力意义深远。目前，工业安防检测逐渐朝着智能化的方向发展。虽然防毒面具检测是工业安防检测一个重要部分，但现有的研究很少关注这个方向。最主要的原因是工业生产环境十分复杂，难以获取大量的高质量的数据集。因此本文的主要工作和研究内容如下：

(1) 防毒面具数据集构建

深度学习是一项高度依赖数据的技术，只有高质量的数据才能训练出具有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，制作一个场景丰富、数据量充足的数据集至关重要。为构建一个高质量的防毒面具检测数据集，本工作首先实地调研工厂获取视频素材，并对其进行抽帧和数据清洗等操作以获得原始数据集。接着，根据实际检测需求设计合理标注策略，并使用标注工具对数据进行标注。之后，对已标注的数据进行统计分析以初步了解检测任务中存在哪些难点。最后，选择合适的离线增强方法对数据进行增强。

(2) 基于 Faster R-CNN 的防毒面具佩戴检测算法研究

本研究针对工业场景下中防毒面具检测所面临的多尺度检测和类别不均衡等问题，在 Faster R-CNN 的基础上进行了一系列优化。首先针对数据集中存在的多尺度检测问题，在 Faster R-CNN 中添加特征金字塔网络，以融合不同尺度

目标的特征，提高模型多尺度检测能力。其次，在训练过程中使用在线难例挖掘算法，提高模型对难样本的关注程度，缓解数据集中存在的类别不均衡问题。之后使用 Mosaic 与 Mixup 混合增强，提升样本的多样性，背景复杂度，缓解模型过拟合并提升模型在复杂场景下的鉴别能力。最后对改进后的算法进行了实验对比分析。

(3) 基于改进 YOLOv5 的防毒面具佩戴检测算法研究

针对工业应用中存在的复杂背景干扰、数据集缺乏多样性等问题，提出了一种基于 YOLOv5 的改进算法。首先，设计了 ELAN-ESE 模块，该模块有效的结合了 ELAN 与 Efficient SE 模块，让模型保持轻量的同时获得更丰富的梯度流信息，提高主干网络的特征抽取能力。其次，提出多阶段数据增强策略，该策略有效地生成了更大数量的样本数据，增强模型在实际应用场景中的泛化能力和鲁棒性。之后，使用 VariFocal Loss 缓解数据集中存在的类别不均衡问题。最后对改进后的算法进行了实验对比分析。

1.4 论文组织架构

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以及与本课题相似的研究。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。本章节对防毒面具佩戴检测的相关技术与理论基础进行了阐述。首先介绍了深度学习理论，包括其特点和应用领域。接着概述了卷积神经网络的基本概念和发展历程，并详细讲解了其结构，包括卷积层、池化层、激活函数、全连接层和损失函数等。此外，还探讨了解决训练过程中过拟合问题的方法。随后重点介绍了基于深度学习的目标检测方法，包括两阶段检测方法和一阶段检测方法，以及常用的目标检测损失函数。最后介绍了解决类别不均衡问题的常见方法。

第 3 章：防毒面具佩戴检测数据集。本章节对防毒面具佩戴检测数据集的相关信息进行了阐述。首先介绍了防毒面具的基本信息，包括其功能、结构和常见类型。接着介绍了数据集的标注策略与标注方式。之后介绍了数据集中使用的离线增强方法。最后展示了数据集的统计信息，包括数据采集的场景数量、数据集

中各个类别的标签数量和数据集中边界框尺寸分布。

第 4 章：基于 Faster R-CNN 的防毒面具佩戴检测研究。本章研究以 Faster R-CNN 模型为基础，针对防毒面具检测场景中的多尺度与类别不均衡问题进行了改进。首先介绍了数据增强策略，并解释了数据增强方法原理。之后介绍了本章提出的改进算法，包括 Faster R-CNN 网络结构、特征金字塔结构、在线难例挖掘算法原理。最后，通过消融实验和对比实验来验证改进方法的有效性，并分析了当前方法所存在的问题。

第 5 章：基于改进 YOLOv5 的防毒面具佩戴检测研究。本章主要研究了基于 YOLOv5 模型在实际工业应用场景中的问题，并提出相应的改进方法。首先介绍了多阶段数据增强策略。之后，介绍了改进 YOLOv5 模型的总体结构，包括 ELAN-ESE 模块的结构、模型具体的参数设置和 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 6 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。同时，还提出了未来的研究方向，包括数据集的完善、算法的优化以及实时性处理等。

第 2 章 相关技术与理论基础

2.1 深度学习理论

深度学习是一种模拟人类大脑神经网络结构和功能的机器学习方法。它通过构建多层次的神经网络模型，实现了端到端的自动学习，并能够提取高层次的抽象特征表示。这种方法赋予了机器对复杂模式和任务进行识别、分类、预测和生成等能力。在深度学习中，多层神经网络结构起到了关键作用。这种网络结构包括输入层、多个隐藏层和输出层。每个隐藏层的输出都作为下一层的输入，形成一种层级结构。这种层级结构使得模型能够自动学习和提取数据的多层次特征表示，无需人工设计特征提取器。

深度学习模型通常需要大量的数据集进行训练，以提高模型的泛化能力和准确性。在训练过程中，模型通过优化算法对数据集进行多次迭代学习，以适应各种不同的任务和场景。为了自动优化模型参数，深度学习采用了梯度下降算法。该算法根据损失函数对于每个参数的梯度来确定参数的更新方向。计算得到的梯度会用于更新网络中的权重和偏置，以使损失函数最小化。通过不断重复迭代过程，模型会持续优化自身参数，从而提高其对复杂模式和任务的识别准确率。梯度下降算法有多种变体，如小批量梯度下降、随机梯度下降和 Adam[45]等。

在深度学习领域中，不同的模型结构和算法被设计出来，以应对各种不同类型的任务。例如，卷积神经网络适用于图像处理任务，循环神经网络适用于处理序列数据，而 Transformer[46]则适用于自然语言处理任务。这些模型在计算机视觉、自然语言处理、语音识别和推荐系统等领域取得了显著的成就。随着技术的不断发展、大数据的不断积累以及计算资源的日益强大，深度学习技术将会不断创新和完善，其应用范围也将在更多领域得到拓展。

2.2 卷积神经网络

2.2.1 卷积神经网络结构

卷积神经网络（Convolutional Neural Network，CNN）是一种广泛应用于图像处理、语音识别和自然语言处理等领域的深度学习模型。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。

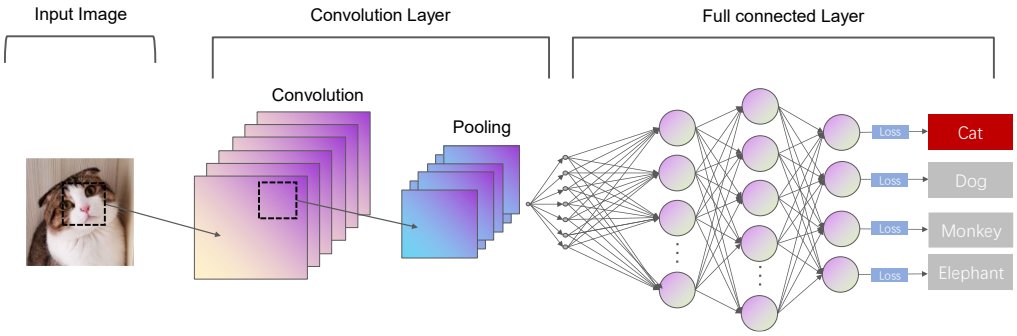


图 2-1 卷积神经网络的架构

Fig. 2-1 The architecture of convolutional neural networks

(1) 卷积层

卷积层是卷积神经网络的核心组件之一，它由多个卷积核组成。每个卷积核都可以对输入数据进行卷积操作，从而提取输入数据的局部特征。卷积操作是通过一个在输入数据上滑动的窗口来完成的，这个窗口通常被称为卷积窗口或卷积核。在卷积操作中，卷积核将与输入数据的对应位置进行点乘和求和运算。这一过程类似于卷积核在输入数据上滑动，并在每个位置对输入数据执行相同的运算。这种运算能够捕捉输入数据的局部特征，因为卷积核可被视为一个局部滤波器，它只关注输入数据在其周围的局部信息。卷积层的另一个重要特性是平移不变性，即无论输入数据的位置如何变化，卷积层的输出结果都保持不变。这是因为卷积核只关注局部信息，而忽略了其他位置的信息。这一特性使卷积层能够有效地处理平移不变性问题。图 2-2 展示了单个卷积核的计算过程。

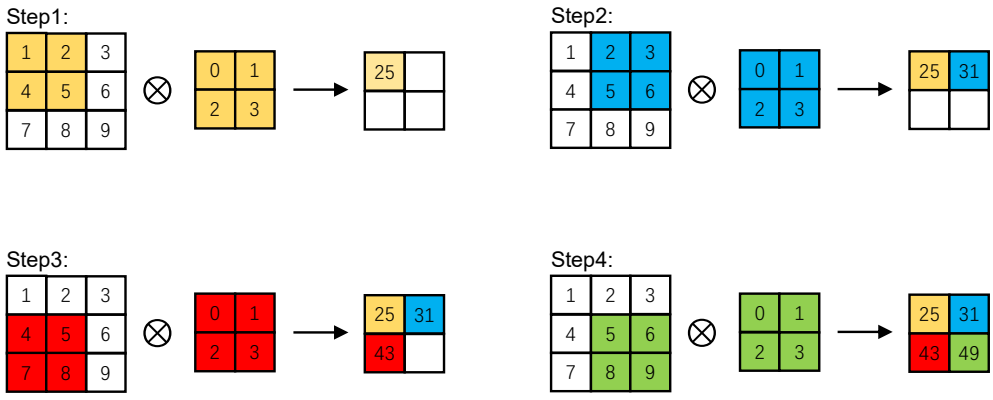


图 2-2 卷积计算示意图

Fig. 2-2 The diagram of convolution calculation

(2) 激活函数

激活函数通常应用于卷积层的输出，目的是为神经网络引入非线性变换。在卷积神经网络中，卷积层的作用是从输入数据中提取局部特征，但这些特征往往是线性的，难以准确表达复杂的非线性关系。因此，引入激活函数可以使神经网络更好地模拟各种非线性关系。以下将介绍几种常见的激活函数。

Sigmoid：该激活函数的输入是一个实数，输出被限制到 0 到 1。Sigmoid 函数可以解决梯度消失的问题，因为它的输出值在 0 到 1 之间，所以梯度不会趋于 0，但当输入值较大或较小时，会出现梯度消失或梯度爆炸的问题。Sigmoid 的数学表示如公式（2-1）所示：

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

ReLU：该函数可以将输入值映射到 0 或正数之间，可以解决梯度消失的问题，并且计算速度快[47]。但训练过程中可能会出现“死区”现象，即当输入值接近于 0 时，ReLU 函数的输出值也接近于 0，导致梯度几乎为 0，从而影响模型的训练效果。此外，ReLU 函数不太适用于处理一些需要负数输出的任务。ReLU 函数的数学表示如公式（2-2）所示:

$$f(x)_{ReLU} = max(0, x) \tag{2-2}$$

LeakyReLU：该函数是一种改进的 ReLU 激活函数，旨在解决 ReLU 在负数区间的“死区”问题。Leaky ReLU 的函数表达式为：

$$f(x)_{leakyReLU} = max(0, x) + \alpha \times min(0, x) \tag{2-3}$$

其中，α 是一个正数，通常取值为 0.01。在输入为负数时，该函数会输出一个较小的非零输出，这样在训练过程中就不会出现梯度几乎为 0 的情况，可以避免在训练过程中出现“死区”现象。

SiLU：该函数通过将输入与 sigmoid 函数相结合，使得输出具有非线性输出[48]。SiLU 的导数在输入大于 0 的部分为 sigmoid 函数，而在输入小于 0 的部分为 0，这种稀疏性有助于减少计算量和内存使用。可以避免 ReLU 可能出现的“死区”问题，适用于各种类型的神经网络任务。它的函数表达式为：

$$f(x)_{SiLU} = x \times sigmoid(\beta \times x) \tag{2-4}$$

(3) 池化层

池化层在 CNN 中起到减小特征图尺寸的作用，并保留主要的特征信息。常

见的池化操作包括最大池化和平均池化，通过划分输入数据的区域，并在每个区域内提取最大值或平均值。池化层能够减少数据的维度，提高网络的计算效率，并具有一定的平移不变性。图 2-3 展示了平均池化与最大池化的计算过程。

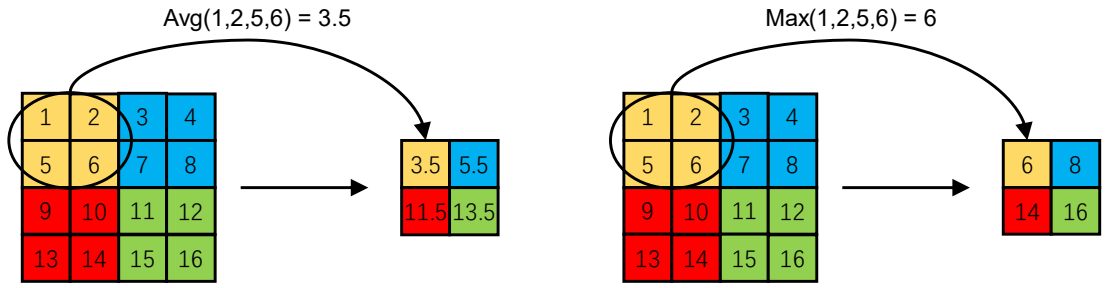


图 2-3 平均池化与最大池化的计算过程

Fig. 2-3 The calculation process of average pooling and max pooling

(4) 全连接层

全连接层将卷积层和池化层的输出连接起来，并应用于最终的分类或检测任务。如图 2-4 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，通过学习权重和偏置来进行信息的传递和转换。这样的连接方式使得网络能够学习到高级抽象特征，从而更好地完成复杂的分类或预测任务。

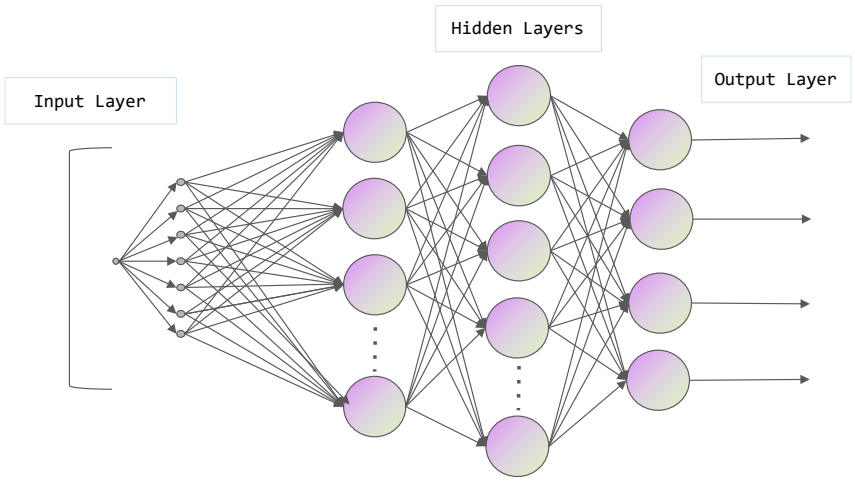


图 2-4 全连接层

Fig. 2-4 Fully connected layer

(5) 损失函数

在卷积神经网络中，损失函数是模型训练过程中的重要指标，用于衡量模型的预测输出与真实标签之间的差异，以指导模型的优化。常用的损失函数包括交叉熵损失函数、均方误差损失函数等。

交叉熵损失函数是最常用的损失函数之一，它常用于多分类问题。在分类问题中，通常使用 Softmax 函数将模型的输出转换为概率分布，然后使用交叉熵损失函数来衡量模型的输出与真实标签之间的差异。交叉熵损失函数的计算方式如下：

$$H(p,y) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(p_i) \tag{2-5}$$

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。

均方误差损失函数是另一种常用的损失函数，它主要用于回归问题。均方误差损失函数的计算方式为：

$$H(p,y) = \frac{1}{n}\sum_{i=1}^{n} (p_i - y_i)^2 \tag{2-6}$$

其中，n表示样本数量，y_i表示第i个样本的真实值，p_i 表示模型对第i个样本的预测值。均方误差损失函数的优点是可以有效地处理回归问题，并且不易受到异常值的影响。

2.2.2 卷积神经网络训练

在训练卷积神经网络时，可能会遇到过拟合或欠拟合的问题。过拟合指模型在训练数据上学习得过于复杂，以至于在测试数据上表现不佳。相反，欠拟合则是模型未能充分学习训练数据中的特征，导致在测试数据上表现不佳。为了避免这些问题，需要采取一些措施来提高模型的泛化能力。为了解决这些问题，研究人员提出了许多方法。

Data Augmentation：在训练网络时，避免过度拟合和者欠拟合最有效的方法之一是使用大量数据上进行训练。但在实际应用中，由于数据集的规模有限，很难获得足够大规模的数据。为了解决这个问题，可以使用数据增强技术来扩大训练数据集的规模，从而提高模型的泛化能力和鲁棒性。常用的数据增强方法有几何变换和色调变换等。在几何变换中，常用的有扭曲变形、旋转、缩放、反转、弹性变形[49]、随机裁剪和平移等。在色调变换中，常用的有调整图像曝光和饱和度等。此外还有一些如随机擦除[50]、CutOut[51]、MixUp[52]、CutMix[53]等增强方法。

Batch Normalization：该方法是一种在神经网络训练中广泛使用的技术，旨在解决内部协变量偏移问题，并加速训练过程[54]。在神经网络中，随着层数的增加，输入数据的分布可能会发生变化，导致模型难以训练和收敛。该方法通过对每一层的输入数据进行归一化处理，使其均值为 0，标准差为 1，从而使得输入数据的分布更加稳定，提高模型的训练速度和泛化性能，避免模型过拟合。

在模型训练过程中，参数更新问题与过拟合问题同样重要。卷积神经网络的训练通常会使用基于梯度优化的技术来更新参数。训练时，网络会不断更新参数，以优化网络性能，并寻找局部最优解，使误差最小化。学习率控制着每次参数更新的幅度，它定义了参数更新的步长。而训练迭代则是指包含完整训练数据集的一次完整的参数更新过程。在每个训练迭代中，网络参数会根据相应算法进行重复更新。以下是一些常用的参数更新方法。

Batch Gradient Descent(BGD)：该算法在执行过程中通过计算整个训练集的梯度来更新参数。对于规模较小的数据集，CNN 模型的收敛速度更快。然而，由于每个训练迭代中参数只更新一次，因此需要大量的计算资源。相比之下，对于较大的训练数据集，需要额外的时间才能收敛。

Stochastic Gradient Descent（SGD）：SGD 在训练过程中通过计算单个样本的梯度来更新模型参数，而不是使用整个数据集。这种随机采样的方法使 SGD 在训练大型数据集时更加高效，同时能够加速模型的收敛速度。然而，它使用单个样本来计算梯度，因此可能会在优化过程中产生较大的波动，导致训练不稳定。其次，SGD 需要设置学习率参数，如果设置不当，可能会影响模型的收敛效果。

Momentum：该方法对 SGD 进行了改进，通过引入动量项来加速梯度下降的收敛过程，并使参数更新路径更加平滑。然而，Momentum 算法仍存在一定的不稳定性。例如，如果动量系数设置不当，可能会导致参数更新波动较大。此外，该算法对 batch 大小的设置也较为敏感。

Adaptive Moment Estimation（Adam）：该算法作为一种自适应学习率的优化算法，融合了动量和自适应学习率的特性[45]。Adam 算法不仅考虑当前梯度，还兼顾过去梯度的累积以及历史信息。它借助维护动量和梯度的一阶矩与二阶矩，实现网络权重的更新，进而加快深度神经网络的训练和收敛速度。这使得 Adam 在多数情况下能够有效提升收敛速度，并在训练过程中自动调整学习率。此外，

Adam 算法还具有收敛迅速、适应不同参数和时间步长的能力，同时具备较高的计算效率和较小的内存需求。

Exponential Moving Average（EMA）：该方法用于平滑模型参数的更新和优化器的更新。EMA 通过对模型参数或梯度进行加权平均来减少更新的波动，使模型更加稳定，并提高模型的性能和泛化能力。

2.3 基于深度学习的目标检测算法

2.3.1 两阶段目标检测算法

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，它由两个阶段组成：建议区域生成和目标分类。在第一阶段，算法将产生众多关于目标的建议区域。在第二阶段，算法将检测建议区域，以获得精确的坐标和类别信息。两阶段目标检测模型的优点是准确性高，可以在不同尺度的目标上进行检测，并且可以处理重叠目标。一般来说，两阶段模型比一阶段模型更准确，但需要更多的计算资源。

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卷积神经网络，对目标检测任务进行了革新性的改进，具有里程碑式的意义。R-CNN 的算法原理如图 2-5 所示，其核心流程包括三个步骤：首先，使用 Selective Search 算法[13]生成两千个关于目标的候选区域；接着，将生成的候选区域进行裁剪，并送入卷积神经网络中进行特征提取；最后，将提取出的特征送入分类网络中进行分类。然而，由于在建议区域中存在大量的冗余计算，且不利于卷积神经网络的训练。

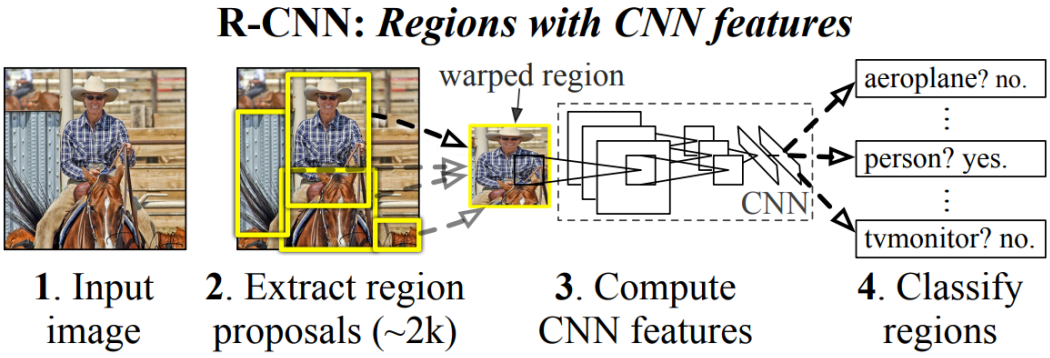


图 2-5 R-CNN 网络结构[13]

Fig. 2-5 The network architecture of R-CNN[13]

Fast R-CNN 是对 R-CNN 的改进。它沿用了 R-CNN 中的 Selective Search 算法，但与 R-CNN 不同的是，Fast R-CNN 先将图像送入卷积神经网络进行降维处理，再将候选区域映射到降维后的特征图上，最后将这些特征图送入分类模型中。这种方法降低了训练的难度，提高了训练效率，但仍然存在冗余计算的问题，未能实现端到端检测。

Faster R-CNN 是首个实现端到端目标检测的模型。如图 2-6 所示，Faster R-CNN 最大的特点是使用了 RPN（Region Proposal Network）来生成候选区域。通过神经网络自动学习生成建议区域，极大地缩减了回归任务的搜索范围，提升了检测效率。从图中可以看出，RPN 与分类层共享来自卷积层的特征图，这一设计使得 RPN 可以在共享的特征图上进行边界框回归，从而实现了目标检测任务的端到端训练。此外，Faster R-CNN 采用了 RoI Pooling 算法[16]，将不同大小的候选区域转换为固定长度的特征向量，使得模型可以更好地处理不同尺寸的目标。RoI Pooling 是一种局部特征提取方法，它将输入的特征图划分为多个区域，并对每个区域进行最大池化操作，以提取该区域的代表性特征。通过 RoI Pooling，Faster R-CNN 可以将不同尺寸的候选区域映射到一个固定长度的特征向量，从而便于后续的目标分类和边界框回归。

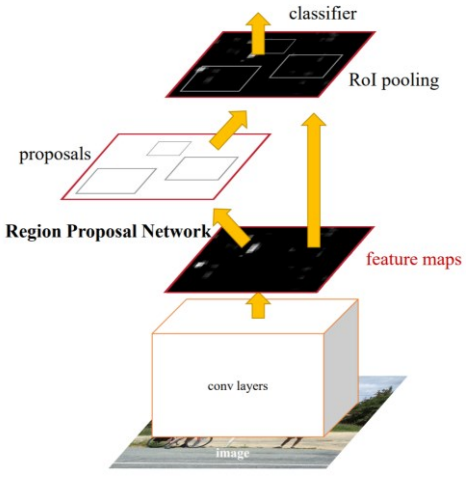


图 2-6 Faster R-CNN 网络结构[16]

Fig. 2-6 The network architecture of Faster R-CNN[16]

2.3.2 一阶段目标检测算法

一阶段目标检测算法是一种能够把目标检测问题转换为回归问题的方法，它的核心在于借助单一网络直接预测目标的边界框和类别。与两阶段算法相比，一

阶段物体检测算法具有以下优势：首先，结构简单。它仅需一个网络即可完成目标检测任务，相较于传统算法需要多个网络完成不同任务，其结构更为简洁，训练和调试也更为便捷。其次，计算资源需求较少。因为只需要一个网络来完成任务，所以计算资源的需求相对较低，可以在较低成本下实现高效的物体检测，适用于计算资源有限的场景。最后，实时性较强。由于计算资源需求少，能够在短时间内完成物体检测任务，实现实时物体检测，适用于对实时性要求较高的场景。

YOLOv1 是一种经典的一阶段目标检测算法。该算法利用单一神经网络来预测物体的类别和边界框的位置和大小。如图 2-7 所示，网络将输入图像划分为$S \times S$个区域（原文中使用7×7），每个区域负责预测一定数量的物体及其边界框。对于每个区域，YOLOv1 预测B个边界框以及它们的置信度，并预测C个条件概率代表该区域内存在的物体类别的概率。通过单次卷积神经网络的端到端训练，YOLOv1 算法省去了预测候选框的过程，极大地提高了检测效率。相较于 R-CNN 系列算法，YOLOv1 具有更快的检测速度，然而在精度方面稍显不足。

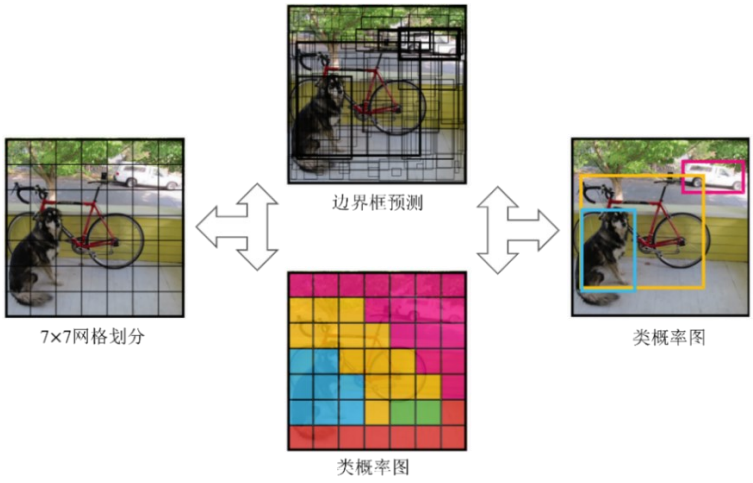


图 2-7 YOLO 算法的原理[21]

Fig. 2-7 YOLO algorithm principle[21]

YOLOv2 是 YOLO 系列的第二代算法。相比于第一代 YOLO，YOLOv2 在速度和准确性上都有了显著的提升。以下是 YOLOv2 的主要改进：首先使用 Darknet-19 网络代替 AlexNet。Darknet-19 是一种轻量级的卷积神经网络，它在保证较高精度的同时，减少了网络参数和计算量，从而提高了检测速度。其次引入 Anchor Boxes 来预测物体的形状和位置。Anchor Boxes 是一种先验框，它们在图像中不同位置和尺度上生成，用于检测不同大小和比例的物体。之后使用多

尺度训练和测试。YOLOv2 在训练和测试过程中使用多个不同尺度的图像来提高检测精度。在训练过程中，网络会在不同尺度的图像上进行训练，以便更好地适应不同大小和比例的物体。最后引入批量标准化操作。批量标准化操作可以在训练过程中对每一层的输入进行标准化处理，从而提高模型的泛化能力。

YOLOv3[24]是 YOLO 系列的第三个版本，由 Joseph Redmon 和 Ali Farhadi 于 2018 年提出。此算法运用了一系列创新设计，旨在提高目标检测的精确性和速度。

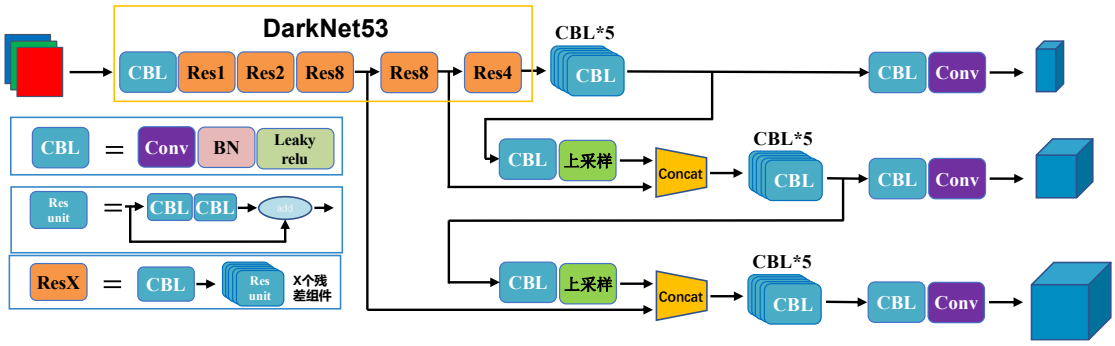


图 2-8 YOLOv3 结构图

Fig. 2-8 The structure of YOLOv3

如图 2-8 所示，YOLOv3 采用 Darknet-53 作为其主干网络。Darknet-53 网络参考了 ResNet 中的残差结构[11]，在部分层中使用了残差模块，以提升主干网络的特征提取能。其次，YOLOv3 引入了特征金字塔网络结构，以便在不同尺度上检测物体。通过在 Darknet-53 的基础上添加额外的卷积层来构建特征金字塔，YOLOv3 能够在不同尺度上检测物体，并且能够有效处理尺寸变化较大的物体。为了提高目标检测的精确度，YOLOv3 采用了多尺度预测。该算法借助三个不同尺度的输出层来检测不同尺寸的物体，每个尺度的输出层都会进行卷积操作，以预测不同大小的边界框和各个类别的概率。这种多尺度预测方法能够提高目标检测的准确性，并且可以有效处理不同尺寸的物体。此外，为了进一步提升目标检测的准确性，YOLOv3 采用了更为精细的边界框预测。具体而言，YOLOv3 沿用了 YOLOv2 中通过聚类获取先验框尺寸的方法，为每种下采样尺度设定了 3 种先验框，共计聚类出 9 种尺寸的先验框。

YOLOv4[25]是 YOLO 系列中第四个版本，它的主要贡献在于提出了一种实时、高精度的目标检测模型。这个模型可以使用如 1080Ti 或 2080Ti 等通用 GPU

来训练快速和准确的目标检测器。在模型结构方面，YOLOv4 进行了一些创新。首先，它将主干特征网络从 Darknet-53 改进为 CSPDarknet53[25]以更高效的提取特征。然后，YOLOv4 借鉴了 PANet[29]和特征金字塔的思想，通过构建路径聚合特征金字塔网络（PAFPN）来检测不同尺度的物体。最后，YOLOv4 还设计一些高效的卷积模块，如 CBM、CBL 等。除了模型结构优化外，YOLOv4 还在数据处理、网络训练、激活函数、损失函数等方面进行了优化。同时，YOLOv4 还验证了一些常用的 Bag-of-Freebies[56]和 Bag-of-Specials[25]方法的效果，对 SOTA 方法进行改进，使其效率更高，更适合单 GPU 训练。

YOLOv5 是 Ultralytics 公司提出的 YOLO 系列的第五个版本。如图 2-9 所示，YOLOv5 在重构骨干网络时采用了全新的设计理念，对原有结构进行了优化。该算法引入的 CBS 和 C3 模块，有效提高了网络性能和泛化能力。此外，这些改进使 YOLOv5 的骨干结构更具扩展性和可定制性，能够适应各种不同的应用场景和需求。除了骨干结构的优化，YOLOv5 还经历了多个版本的迭代，在实际应用中部署方便，能够快速适应不同的环境和任务需求。同时，它具有较高的预测速度和良好的检测精度，能够实现实时目标检测和定位。因此，YOLOv5 在工业领域得到了广泛的应用。

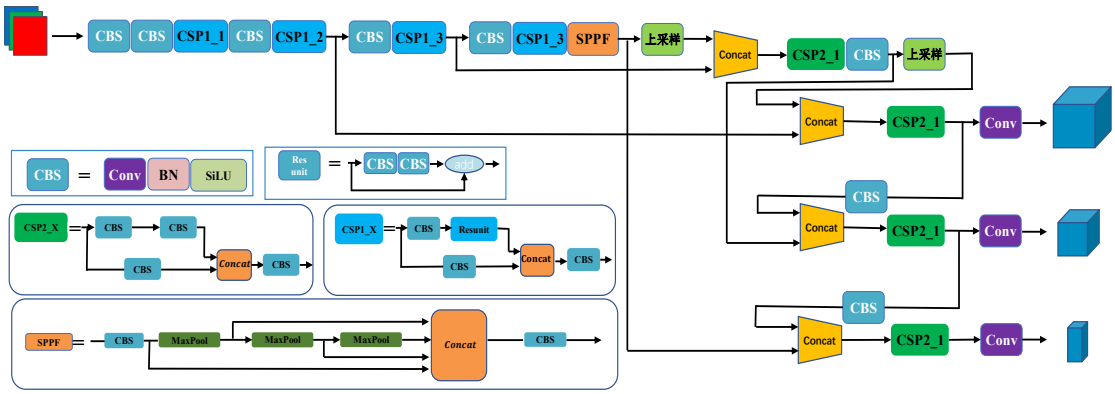


图 2-9 YOLOv5 结构图

Fig. 2-9 The structure of the YOLOv5

2.3.3 目标检测损失函数

在目标检测任务中，常见的损失函数包括用于边界框回归的损失函数和用于分类的损失函数。分类损失函数主要使用交叉熵损失函数，边界框损失函数包括平滑 L1 损失、均方误差损失和基于 IoU 的损失函数。边界框回归是目标检测中

的常用技术手段，它通过使用矩形框预测目标在图像中的位置，并对预测框进行精细调整。目前主流目标检测方法普遍采用 IoU (Intersection over Union)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。下面将介绍一些常用的 IoU 损失函数。

(1) IoU 损失函数

如图 2-10 所示，“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。

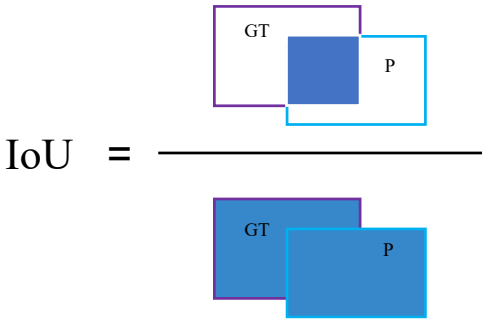


图 2-10 IoU 计算示意图

Fig. 2-10 IoU calculation diagram

IoU 损失[57]定义如公式（2-7）所示：

$$L_{IoU} = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{2-7}$$

(2) GIoU 损失函数

由于 IoU 损失无法处理目标未重叠和目标之间不同的对齐方式，Rezatofighi 等[58]提出了 GIoU 损失函数。公式（2-8）展示了 GIoU 损失的计算方式，其中A_c代表图 2-10 中“GT”与“P”的最小闭包区域面积，即同时包含了预测框和真实框的最小框的面积，U代表 “GT”与“P”的并集。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。

$$L_{GIoU} = 1 - IoU + \frac{A_c - U}{A_c} \tag{2-8}$$

(3) DIoU 损失函数

DIoU 考虑了目标与锚框之间的距离、重叠率和尺度，从而使目标框回归更加稳定[59]。相比之下，IoU 和 GIoU 在训练过程中可能会出现发散等问题。DIoU 的计算方法如公式（2-9）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闭包区域的对角线距离。

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{2-9}$$

(4) CIoU 损失函数

由于 DIoU 在计算中未考虑边界框回归中的长宽比，因此 CIoU 进一步在 DIoU 的基础上进行了改进[59]。其惩罚项如下面公式：

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2-10}$$

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：

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{2-11}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{2-12}$$

完整的 CIoU 损失函数的定义如公式（2-13）所示：

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2-13}$$

(5) EIoU 损失函数

EIoU[60]针对 CIoU 公式中的v未反应宽高分别与其置信度的真实差异，在 CIoU 的惩罚项基础上将纵横比的影响因子拆开分别计算目标框和锚框的长和宽。该损失函数包含三个部分：重叠损失，中心距离损失，宽高损失。前两部分延续 CIoU 中的方法，但是宽高损失直接使目标框与锚框的宽度和高度之差最小，使得收敛速度更快。EIoU 损失函数如公式（2-14）所示，其中C_w和C_h是覆盖两个 Box 的最小外接框的宽度和高度。

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{2-14}$$

2.3.4 评价指标

与分类任务不同，目标检测任务中的输出包含大量目标的位置、类别和置信度。因此，在评估模型性能时，需要事先设置交并比（Intersection over Union, IoU）阈值和置信度阈值，从而有效筛选出符合要求的预测结果。预测框和真实值之间的交并比用于计算$Precision$、$Recall$和AP。IoU 根据图 2-10 计算。$Precision$和$Recall$分别定义在公式（2-15）和公式（2-16）中。

$$Precision = \frac{TP}{TP + FP} \tag{2-15}$$

$$Recall = \frac{TP}{TP + FN} \tag{2-16}$$

TP（True Positive）表示 IoU 大于指定阈值的预测框的数量。相反，FP（False Positive）表示 IoU 小于或等于阈值的预测框的数量。最后，FN（False Negative）代表未检测到的实际框的数量。AP（Average Precision）是特定类别的准确率与召回率曲线下区域的面积，而mAP表示每个类别的AP的平均值。对于多类别任务，mAP（Mean Average Precision）可以使用公式（2-18）计算，其中AP_i表示给定类别的平均精度。

$$AP_i = \int_0^1 Precision(Recall) d(Recall) \tag{2-17}$$

$$mAP = \frac{\sum_{i=1}^{k} A\,P_i}{k} \tag{2-18}$$

$F_1\text{-}score$ 的定义如公式（2-19）所示，它是评价模型分类能力的综合评价指标。

$$F_1 - score = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{2-19}$$

目标检测中的常见评价指标主要有两种：一种是早期 VOC 数据集[61]提出的指标。另一种是广泛使用的 COCO 指标[62]。本文采用 COCO 指标对模型进行评价。

表 2-1 COCO 评价指标[62]

Table 2-1 COCO evaluation metric[62]

评价指标	含义
$mAP_{0.5:0.95}$	IoU介于 0.5~0.95 之间的 10 个mAP的平均值
$mAP_{0.5}$	IoU等于 0.5 时的mAP值
$mAP_{0.75}$	IoU等于 0.75 时的mAP值
mAP_{small}	目标尺寸小于32^2的mAP值
mAP_{medium}	目标尺寸介于32^2与96^2的mAP值
mAP_{large}	目标尺寸大于96^2的mAP值

表 2-1 详细展示了 COCO 评价指标。与传统的仅使用单一IoU阈值计算 mAP 的方法不同，COCO 指标采用了 10 个IoU阈值（范围在 0.50 到 0.95 之间，步长

为 0.05）来分别计算 mAP，然后对这 10 个阈值下的mAP取平均值，得到最终结果。而传统的mAP仅仅依据IoU为 0.5 的结果。因此，平均多级IoU能够更全面地评估检测效果。

2.4 类别不均衡

目标检测中存在多种多样的不均衡，这些不均衡会影响最终的检测精度。Oksuz 等[63]详细总结了目标检测中的不均衡问题，并将这些问题为四类：

(1) 类别不平衡：主要由样本数量上的差异引起。其中最典型的是前景和背景不平衡，即在训练过程中前景目标的数量远远少于背景目标的数量。

(2) 尺度不平衡：这种不平衡主要由目标的尺度差异引起，例如在 COCO 数据集中存在过多的小物体，或者在分配到特征金字塔时出现的不平衡情况。

(3) 空间不平衡：涉及到回归损失贡献的样本对不平衡、IoU 的不平衡以及物体分布位置的不平衡等问题。

(4) 目标不平衡：指多任务损失优化之间的不平衡，尤其是分类和回归损失之间的不平衡。

在防毒面具佩戴检测研究中，训练中存在的类别不平衡问题是值得关注的问题。在两阶段算法中，Girshick 等[64]提出了在线难例挖掘(Online Hard Example Mining, OHEM)来解决训练中前景目标数量与背景数量的不平衡问题。其主要思想根据样本的损失值对样本进行排序，选取损失值较高的部分样本送入网络再次训练。这种策略有效的避免了超参数对于训练的影响，使网络关注较难分类的样本，即少量前景目标与较难的分类背景目标。然而，OHEM 中定义的多任务损失函数在整个训练过程中权重相同，这种方式忽略了不同损失类型在训练过程中的影响。例如在训练期的后期，定位损失更为重要，因此 OHEM 对定位精度的关注可能不足。Li 等[65]提出了 S-OHEM，根据 loss 的分布抽样训练样本。相比 OHEM，S-OHEM 基于不同损失函数的分布来抽样选择困难样本，避免了仅使用高损失的样本来更新模型参数，但 S-OHEM 引入了额外的超参数 α、β，且没有给出普适性的超参数选择方式。

在一阶段算法中，大部分的研究都是通过改进损失函数来缓解类别不均衡问题。最常用的方法是从样本分布角度对损失函数添加权重因子，如 Balanced Cross Entropy。Lin 等[66]提出了 Focal Loss 损失函数来处理类别不均衡问题。Focal Loss

通过对不同样本的损失进行加权，使模型更加关注难分样本，从而提高模型在样本不平衡情况下的分类准确率。然而，Focal Loss 可能会使模型过度关注那些特别难分的样本，其中可能包含离群点，从而导致模型过拟合。因此，Liu 等[67]提出了 GHM(Gradient Harmonized Single-Stage Detector)来处理离群点问题。GHM 提出了梯度模长概念，通过梯度模长的大小来区分离群点和正常样本，能够有效的降低训练过程中模型对难以分类样本的关注度。此外，Zhang 等[68]提出了 Varifocal loss 作为 Focal loss 的改进版本。相较于 Focal loss，Varifocal loss 通过调整负例的贡献并保持正例的贡献，以实现对难易例平衡的控制。Li 等[69]提出了一种名为 Generalized Focal Loss 的方法。该方法通过引入边界框的不确定性统计量，来指导定位质量的估计。具体而言，该方法利用边界框的分布形状进行学习，从而提高目标检测器的性能。

2.5 本章小结

本章节对防毒面具佩戴检测的相关技术与理论基础进行了阐述。首先介绍了深度学习理论，包括其特点和应用领域。接着概述了卷积神经网络的基本概念和发展历程，并详细讲解了其结构，包括卷积层、池化层、激活函数、全连接层和损失函数等。此外，还探讨了解决训练过程中过拟合问题的方法。随后重点介绍了基于深度学习的目标检测方法，包括两阶段检测方法和一阶段检测方法，以及常用的目标检测损失函数。最后论述了类别不均衡问题以及常见的解决方法。

第 3 章 防毒面具数据集

3.1 防毒面具简介

防毒面具是工业生产中常用的劳动保护用品。根据工作原理的不同，防毒面具可分为过滤式和隔离式两种类型。过滤式防毒面具主要借助过滤盒或过滤棉来滤除有害气体、灰尘和颗粒物；而隔离式防毒面具则通过气体储存系统为使用者供氧，并在使用者的呼吸道、眼睛及面部形成保护屏障，避免其受到有害气体的侵害。这种防毒面具通常用于低氧且需要长时间作业的环境。在这两种防毒面具之中，过滤式防毒面具在工业生产中的使用频率相对更高，其使用范围也更为广泛。因此，防毒面具检测任务主要针对图 3-1 所示的 3M 过滤式防毒面具。该防毒面具配有 2 个滤毒盒，可用于过滤空气中的有害气体。滤毒盒内装有活性炭及其他吸附剂材料，能有效吸附和去除空气中的污染物。

图 3-1 3M 防毒面具

Fig. 3-1 3M gas mask

3.2 数据集标注策略

在本工作中，头部面向摄像头的方向以及工人是否配备防毒面具是划分标签的依据。头部朝向分为正面、侧面和背面三种类型。而根据受试者是否佩戴防毒面具，这三类又可进一步细化为七种标签。表 3-1 中详细描述了每个标签的名称和含义，图 3-2 则展示了标注的示例。在本研究的场景中，工人可能会佩戴不能

提供足够呼吸保护的口罩进入生产车间。这种情况与未佩戴防毒面具具有同样的风险，因为口罩无法有效阻隔有害气体或颗粒物。因此，数据集中设置了标签 5 和 6，以区分工人是否佩戴了有效的防毒面具。通过将头部朝向和是否佩戴防毒面具结合起来进行分类，能够使检测结果更好地反映工人的状态。

表 3-1 数据集的标签名称与描述

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

序号	标签名	标签描述
1	front_head_wear	佩戴防毒面具，头部正对着摄像机
2	side_head_wear	佩戴防毒面具，头部侧对着摄像机
3	front_head_no_wear	未佩戴防毒面具，头部正对着摄像机
4	side_head_no_wear	未佩戴防毒面具，头部侧对着摄像机
5	front_head_wear_wrong	佩戴错误的面具，头部正对着摄像机
6	side_head_wear_wrong	佩戴错误的面具，头部侧对着摄像机
7	back_head	头部背对摄像机

图 3-2 标签可视化

Fig. 3-2 Visualization of labels

3.3 离线数据增强

离线数据增强是一种常用的数据处理方法。其主要目的是在训练之前对现有数据进行一系列变换和扩充。通过这些变换和扩充，可以增加数据集的多样性和复杂性，从而提高模型的泛化能力和鲁棒性。离线数据增强将经过变换和扩充的数据存储在硬盘中，以备在进行模型训练时加载到内存或显存中使用。这样的处理方式可以有效地管理和利用数据，同时避免在训练过程中频繁地读取和处理原始数据，提高了训练的效率。同时，通过应用离线数据增强方法，可以模拟真实世界中的不同视角、尺度和形变情况，使得模型更好地适应各种实际应用场景。此外，离线数据增强还能够有效地利用已有数据，并扩充数据集的规模。通过对现有数据进行变换和扩充，可以生成更多样的训练样本，从而增加了训练数据的数量和多样性。这对于有限的数据集来说尤为重要，可以帮助模型降低模型过拟合的风险。

图 3-3 部分离线数据增强展示

Fig. 3-3 Partial offline data augmentation presentation

如图 3-3 所示，本工作使用了诸如 Cutout、Rotate、Flip 和 Shear 等离线数据增强方法对数据集进行了处理。在这些数据增强方法中，Cutout 数据增强方法

会随机遮挡图像的一部分，这有助于提高模型对存在随机遮挡物的图像的识别能力；Rotate 数据增强方法会旋转标注框内的图像，从而使模型能够更好地识别目标物体在旋转情况下的图像；Flip 数据增强方法会左右翻转标注框内的图像，这样模型能够更好地识别出发生左右翻转的目标物体；Shear 数据增强方法会剪切图像的水平边或垂直边，从而使模型能够更好地识别出发生水平或垂直方向形变的图像。

3.4 数据集统计信息

本工作从工厂中收集了大量的视频素材，并对这些视频素材进行视频抽帧，以生成原始的数据集。之后对原始的数据集进行数据清洗，筛选出满足检测要求的样本。图 3-4 展示了数据集中的部分场景。这些图像来源于 10 个不同的场景，涵盖了各种环境和状况。这种多样性有利于模型更好地适应各种实际应用情境。

图 3-4 数据集中部分场景展示

Fig. 3-4 Some scene demonstrations in the dataset

因可获取的工业场景样本较为有限，本工作从部分相似场景中采集样本以扩充数据集。此外，由于数据集中的负样本数量较少，本工作还借助网络爬虫技术，从必应图像库、百度图像库等图像网站获取负样本，作为补充数据。图 3-5 和图 3-6 展示了这些补充数据。这些数据可增加数据的多样性，有助于模型更好地适应各类实际应用场景，从而增强模型的鲁棒性和泛化能力。

图 3-5 相似场景中的样本

Fig. 3-5 The samples in similar scenes

图 3-6 补充的负样本数据

Fig. 3-6 Supplementary negative sample data

在对数据集进行数据补充之后，最终构建了一个包含 14 个场景、7998 张图像的数据集。这些场景涵盖了最为常见的情况，能够为模型提供充裕的数据以供训练，从而让模型学习到不同场景下的特征与规律，尽可能地提升检测的精准度。为了对模型展开训练，本项工作选取了 10 个场景中的 6084 个样本作为训练数据

集。其余 4 个场景里的 1934 张图片被当作测试数据集使用，用于评估模型在未知场景中的性能。

为了让数据集中各个类别的数量尽可能均衡，本工作提升了训练集中标注数量偏少类别的数据增强频率。图 3-7 展示了在数据增强前后数据集中各个类别的标注数量分布情况。从图中可以明显观察到，“front_head_wear_wrong”和“side_head_wear_wrong”这两个类别的标注数量经过数据增强后分别提升到了 3642 和 1889。通过增加对数量较少类别的数据增强频率，能够有效地平衡数据集中不同类别的样本数量。这对于目标检测模型的训练是至关重要的，因为不均衡的数据集会导致模型对数量较多类别的样本学习过多，而对数量较少类别的样本学习不足。这种情况下，模型可能会偏向于预测数量较多类别的目标，而在数量较少类别的目标上表现不佳。

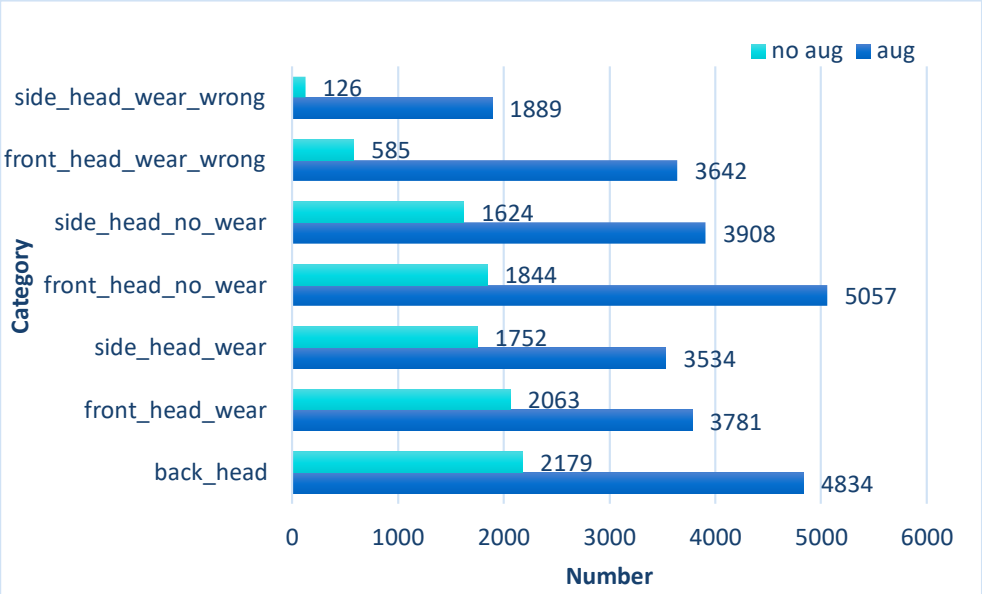


图 3-7 数据增强前后样本数量对比

Fig. 3-7 Comparison of the number of samples before and after data augmentation

图3-8展示了数据集中边界框的宽度和高度分布的散点图。从图中可以看出，大多数边界框的宽高小于300×300，并在尺度上表现出显著变化，表明数据集中有大量小目标。这种情况可能会造成前景和背景之间的不平衡。由于小目标在图像中所占面积较小，这种不平衡可能会导致模型将更多区域归类为背景，而忽略一些小目标。此外，小目标的特征信息较少，模型可能会将部分背景区域错误地分类为小目标，进而导致误检率上升。

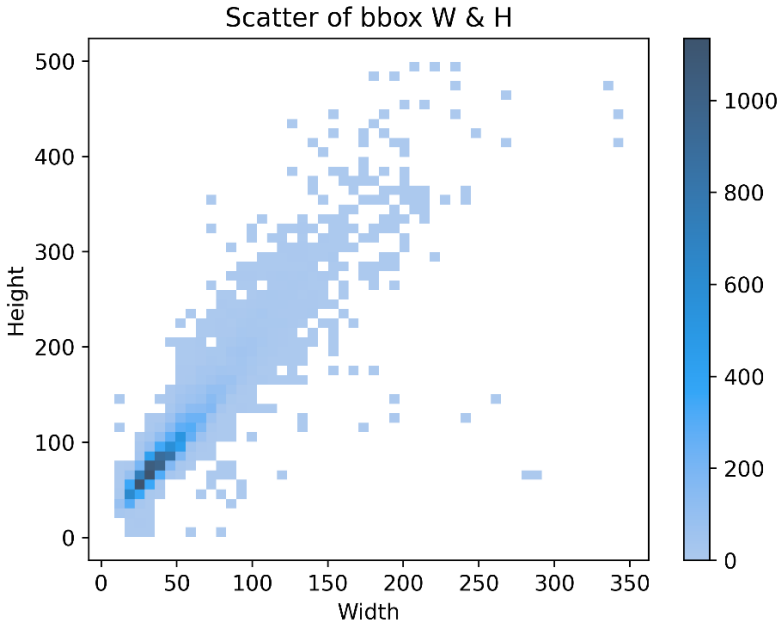


图 3-8 标注框宽高分布

Fig. 3-8 Bboxs width and height distribution

3.5 本章小结

本章节对防毒面具佩戴检测数据集的相关信息进行了阐述。首先介绍了防毒面具的基本信息，包括其功能、结构和常见类型。接着介绍了数据集的标注策略与标注方式。此外，介绍了数据集中使用的离线增强方法，包括 Cutout、Rotate、Shear 和 Flip 等。最后，展示了数据集的统计信息，包括数据采集的场景数量、数据集中各个类别的标签数量和数据集中边界框尺寸分布。

第 4 章 基于 Faster R-CNN 的防毒面具佩戴检测

本章以 Faster R-CNN 模型为研究基础，对防毒面具检测场景中存在的多尺度和类别不均衡等问题进行了改进。在训练过程中，为了提高数据集的多样性，采用了 Mixup 和 Mosaic 混合增强策略。Mixup 策略通过对训练样本进行线性插值，生成新的合成样本，使模型能够学习到更丰富的数据分布和特征表示。Mosaic 策略则通过将多个不同图像的局部区域拼接在一起，生成新的合成样本，进一步增加了数据集的多样性和复杂性。之后，为了解决多尺度问题，在 Faster R-CNN 模型中引入了特征金字塔模块。该模块允许模型有效地提取目标的多尺度信息，从而提高了模型在多尺度场景下的检测精度和鲁棒性。最后，为了缓解数据集中类别不均衡问题，在训练中采用了在线难例挖掘算法。该算法在训练过程中自动选择那些难以正确分类的样本进行重点训练。这种策略使模型能够更加关注难以分类的样本，提高对难样本的识别能力。

4.1 Faster R-CNN 模型研究及改进

4.1.1 数据增强策略

本研究在训练阶段应用了 Mixup 和 Mosaic 混合增强策略。具体而言，在数据预处理过程中以一定的概率对批量数据应用这两种数据增强方法。这种方式能够灵活地控制数据增强的强度，以达到最佳的数据增强效果。

Mixup 的原理是将两个不同样本的输入数据进行混合，从而生成新的训练样本。通过对多个样本进行混合，该方法提高了数据的多样性和复杂性，从而减少模型对单个样本的过度依赖。这样有效地限制了模型在单一样本范围内的过拟合现象，进而提升模型区分目标特征和背景特征的能力。其增强效果如图 4-1 所示。样本使用公式（4-1）[52]进行插值。公式（4-2）表示两个样本中的真值边界框的交集。

$$x = \lambda x_i + (1 - \lambda) x_j \tag{4-1}$$

$$y = y_i \cup y_j \tag{4-2}$$

其中x_i和x_j表示训练集中的图像，y_i和y_j表示它们对应的真值边界框。x是图像x_i和x_j的集合，y是从它们合并的对象的真实边界框的列表。λ是一个服从

$beta(\alpha,\beta)$的参数。为了得到混合样本的总损失，该方法使用两幅图像中目标检测损失的加权和。Mixup 的损失计算如公式（4-3）所示。其中α和β的值设置为1.5[56], L_i和L_j指的是x_i和x_j的损失。

$$L = \lambda L_i + (1-\lambda)L_j \tag{4-3}$$

图 4-1 Mixup 增强的效果展示

Fig. 4-1 Visualization of the effectiveness of Mixup

Mosaic 是一种在目标检测领域被广泛应用的数据增强方法，它能够有效提升训练图像的多样性和数量。图 4-2 展示了 Mosaic 的增强效果。该方法的原理是将四幅不同的图像拼接起来，形成一个全新的训练图像[25]。这种增强方式在保留目标物体位置的前提下，增添了许多小目标，并提升了背景的可变性与丰富性。此外，Mosaic 技术还可以生成多个包含相同目标物体但背景不同的训练图像，这有助于缓解训练数据不平衡的问题。

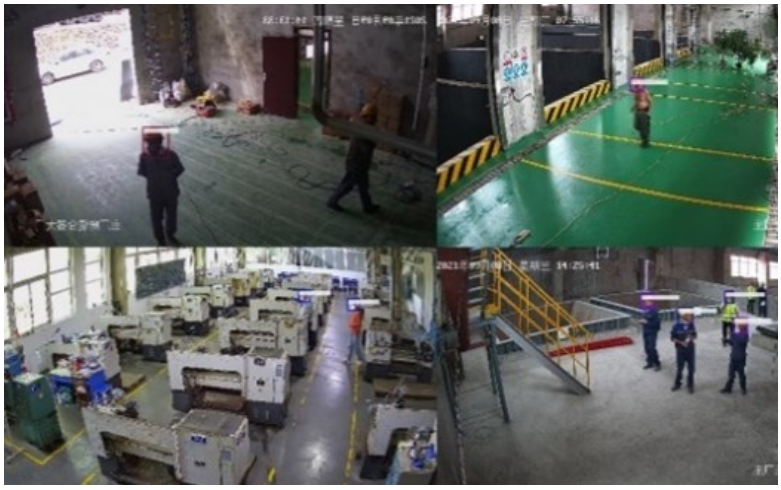


图 4-2 Mosaic 增强的效果展示

Fig. 4-2 Visualization of the Effectiveness of Mosaic

Mosaic 的实现过程如下：首先，从训练数据集中随机选择四张图像。然后，对每张图片进行几何变换和像素变换，如放大、缩小、平移、对比度与亮度调整等。接下来，从四张图片中各裁剪出一个区域，并将这个四个区域融合成一张图片。最后，对裁剪后的图片中的标注框进行相同的几何变换，同时按照算法设定的阈值过滤掉一些不符合要求的标注框。

4.1.2 Faster R-CNN 模型

Faster R-CNN 由三个部分构成：骨干网络、RPN(Region Proposal Network)和 ROI Head (Region of Interest Head)[16]。RPN 和 ROI Head 共享从骨干网络中提取的特征图。RPN 根据输入的特征图生成建议区域，随后 ROI Head 会对这些区域进行处理，执行坐标回归和分类操作。

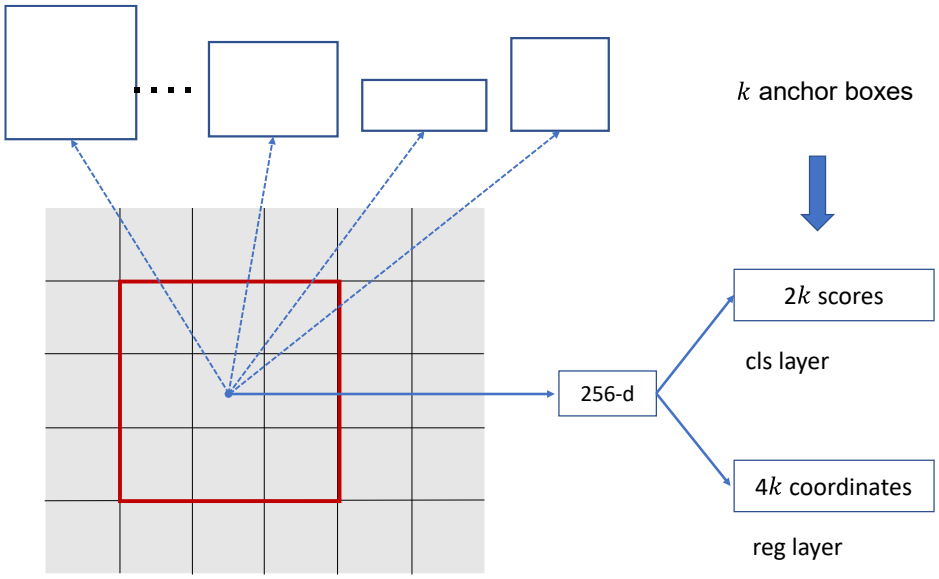


图 4-3 区域建议网络

Fig. 4-3 Region proposal network

如图 4-3 所示，RPN 的核心思想是生成大量关于目标的候选框[16]。RPN 的输入源于主干网络生成的特征图。接下来，它借助滑动窗口处理这些特征图，并同步预测k个候选框。每个滑动窗口都会生成一个 256 维的中间向量，将该向量输入到两个全连接层，以进行类别和坐标预测。类别分数代表一个候选框属于前景或背景的概率。在算法中，k个候选框被参数化为锚点。为了使网络能够更好地适应不同形状和大小的目标，RPN 在特征图的每个位置预置了k个具有不同长宽比的锚框，用于预测不同尺度图像的候选区域。因此，所有滑动窗口会同时预测 k 个锚框，每个滑动窗口的输出包括 $2k$个类分数和$4k$个坐标值。

ROI Head 用于检测和识别工人是否佩戴防毒面具。在训练过程中，图像首先进入主干网络进行特征提取，提取的特征图会被送入 RPN 以生成候选区域。接着，将候选区域中的特征图送入 ROI Pooling[16]，得到固定维度的特征向量。最后，将特征向量输入全连接层，进行目标类别的预测和坐标回归。

4.1.3 特征金字塔

特征金字塔网络（Feature Pyramid Network ,FPN）主要解决了目标检测中的多尺度问题，在不增加原始模型计算复杂度的情况下，通过简单改变网络连通性显著提升了多尺度目标的检测性能。在卷积神经网络的前向计算中，较低层包含更少的语义信息，但提供了更准确的目标位置；相反，较高层包含更多语义信息，但提供的目标位置较模糊。为了更好的融合底层与高层的特征信息，FPN 使用自底向上的路径、自顶向下的路径和水平连接结合了来自网络上层和下层的语义特征和位置信息。该方法大大提高了模型对多尺度目标特别是小目标的检测能力。

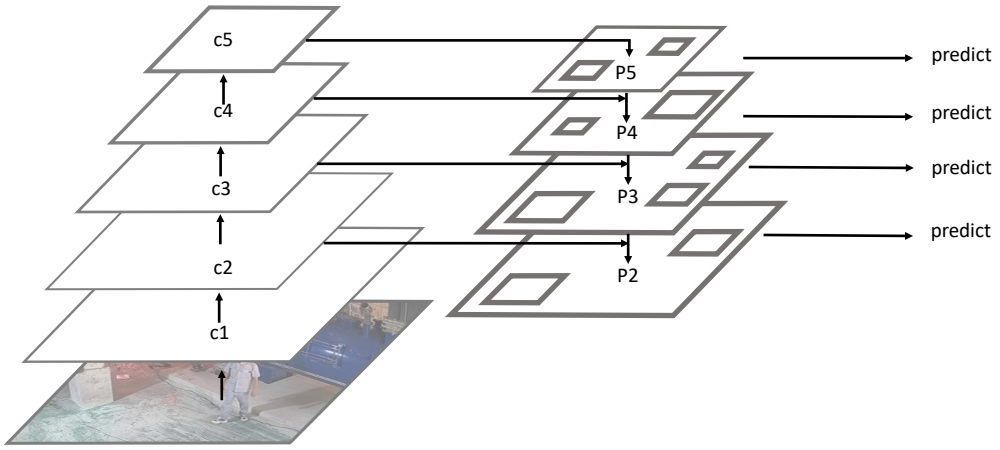


图 4-4 特征金字塔网络

Fig. 4-4 Feature pyramid network

FPN 结构如图 4-4 所示。本研究采用 ResNet5[11]作为骨干网络。它有五个卷积过程，第一个卷积过程由于计算量大，所以没有包含在特征金字塔中。最后四个卷积过程的输出记为{C2,C3,C4,C5}, FPN 的输出记为{P2,P3,P4,P5}。C5 通过1×1卷积进入自上而下的路径。在自顶向下的路径中，利用下层的位置信息和上层的语义信息，通过上采样将尺寸较小的特征图放大到与前一层特征图相同的大小。水平连接是上采样结果和自下而上生成的相同大小的特征图的融合。由于 P5 上采样后的分辨率与 C4 相同，可以直接将这两个特征图相加得到 P4。最后，通过3×3卷积输出每层的特征图。

4.1.4 在线难例挖掘

在训练过程中，RPN 会生成大量随机候选框。由于防毒面具目标在图像中所占比例较小，这将导致背景候选框数量过多，使得背景候选框与前景候选框的数量比例严重失衡，可能致使模型高度偏向背景预测。为解决正负样本不平衡的问题，本研究在 Faster R-CNN 的训练过程中采用了在线难例挖掘算法（Online Hard Example Mining，OHEM）。OHEM 通过增加一个额外的 ROI Head 来选择困难样本，然后利用这些样本训练原始的 ROI Head。OHEM 不仅能够提高模型精度、减少过拟合，还能提高训练效率。此外，该算法无需设定正负样本比例，大大降低了训练难度。如图 4-5 所示，ROI head 被扩展为两个共享参数的网络。其中一个 ROI Head 的参数是固定的，用于计算和排序所有候选区域的损失，进而选择一些损失较高的区域作为复杂样本。另一个 ROI Head 是可训练的，其输入是前一个 ROI Head 选择的复杂样本，输出则是预测的边界框坐标和分类结果。

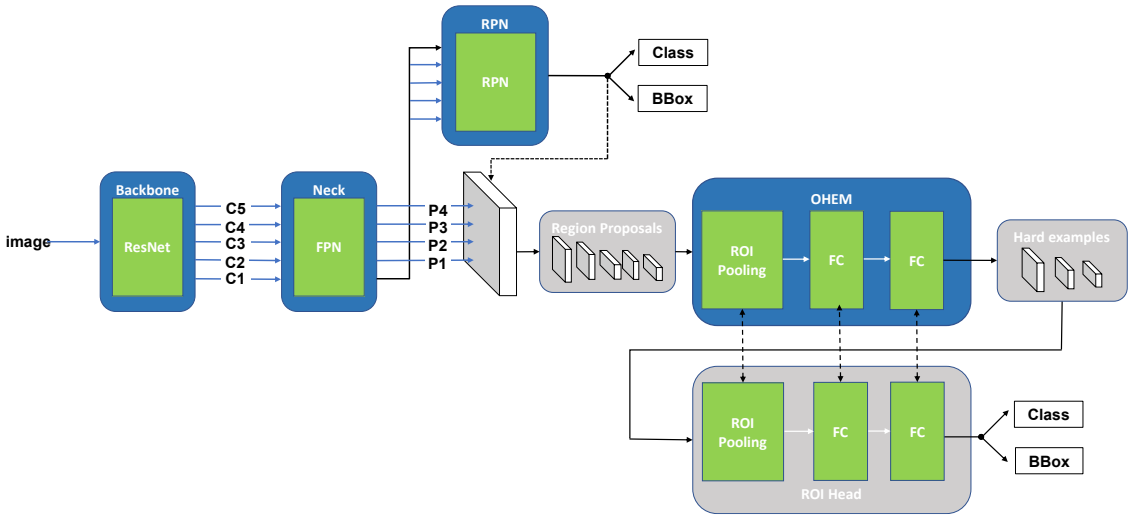


图 4-5 结合 FPN 和 OHEM 的 Faster R-CNN 结构

Fig. 4-5 The structure of Faster R-CNN with FPN and OHEM

4.2 实验结果与分析

4.2.1 实验环境

本章所有实验在 Ubuntu20.04 系统的主机上完成。硬件环境：Intel(R) Core(TM) i9-10940X CPU @3.30GHz，64G 内存，NVIDIA GeForce RTX 3090 GPU，24G 显存。软件环境：Python3.8，Pytorch1.7.1，Cuda11.2。

4.2.2 实验设置

本节的实验中使用了多种输入随机大小的图像，包括 1920×1080、1330×880 和 1024×512。在训练过程中，本研究采用了随机梯度下降优化器，并设置了动量为 0.9 和权重正则化参数为 0.0005。初始学习率被设定为 0.001，并采用多项式衰减策略进行学习率调整，其中权重衰减为 0.01。考虑到 GPU 性能的限制，选择了批量大小为 4。为了评估模型的有效性并发现数据集可能存在的问题，本研究使用了 COCO 指标并进行了一系列的测试。这些测试旨在深入分析模型在不同输入尺寸下的性能表现，并检查模型在识别目标中存在的问题。此外，实验还对模型的分类性能进行了分析，包括计算模型检测目标的精确度、召回率和F_1分数等。

4.2.3 消融实验

本节使用了 1920×1080 的测试输入大小和 0.1 的置信度阈值来比较 Faster R-CNN 中不同方法的精度。表 4-1 展示了实验结果。首先，在没有采用任何额外方法的情况下，直接训练 Faster R-CNN 得到的结果并不理想，$mAP_{0.5:0.95}$仅为 36.7%。接下来，本研究对加入 FPN 和 OHEM 的 Faster R-CNN 的精度表现进行了评估。在 Faster R-CNN 中引入 FPN 后，$mAP_{0.5:0.95}$显著提高了 19.1%，达到了 55.8%。这表明 FPN 的引入对于改善目标检测算法的性能是非常有效的。在使用 OHEM 训练模型后，$mAP_{0.5:0.95}$提高到了 57.9%。最后，应用 Mixup 和 Mosaic 混合数据增强方法后，模型的$mAP_{0.5:0.95}$得分提升了 2%。总体来看，通过运用 FPN、OHEM、Mixup 和 Mosaic 等技术手段，Faster R-CNN 在防毒面具检测任务中的性能得到了显著增强。

表 4-1 消融实验

Table 4-1 Ablation experiment

FPN	OHEM	Mixup	Mosaic	$mAP_{0.5}$(%)	$mAP_{0.75}$(%)	$mAP_{0.5:0.95}$(%)
				55.7	40.7	36.7
√				75.2	67.3	55.8
√	√			81.0	70.1	57.9
√	√	√	√	82.6	72.6	59.9

如图 4-6 所示，改进后的 Faster R-CNN 检测效果整体较好。从图中可以看出，改进后的模型在较小目标的检测上表现优异，具备较高的稳定性，没有出现漏检的情况。但是仍然存在背景误识别和错误分类的问题。这意味着模型在某些情况下可能会将背景区域错误地识别为目标，进而导致检测过程中的误报率上升。同时，错误的分类也会引发一些误报和漏报的情况，给系统的日常使用带来不良影响。因此，为了处理这些问题，本工作需要进一步扩大数据集的规模，并对模型的训练方法进行优化。

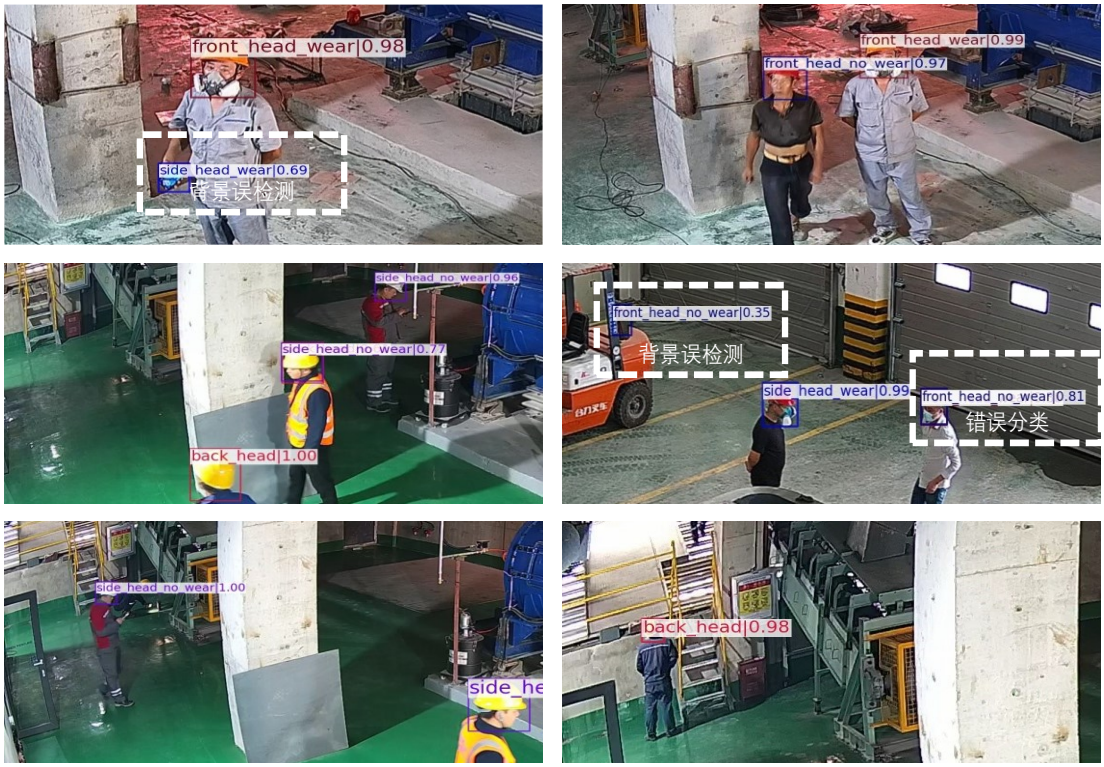


图 4-6 改进后的 Faster R-CNN 检测效果

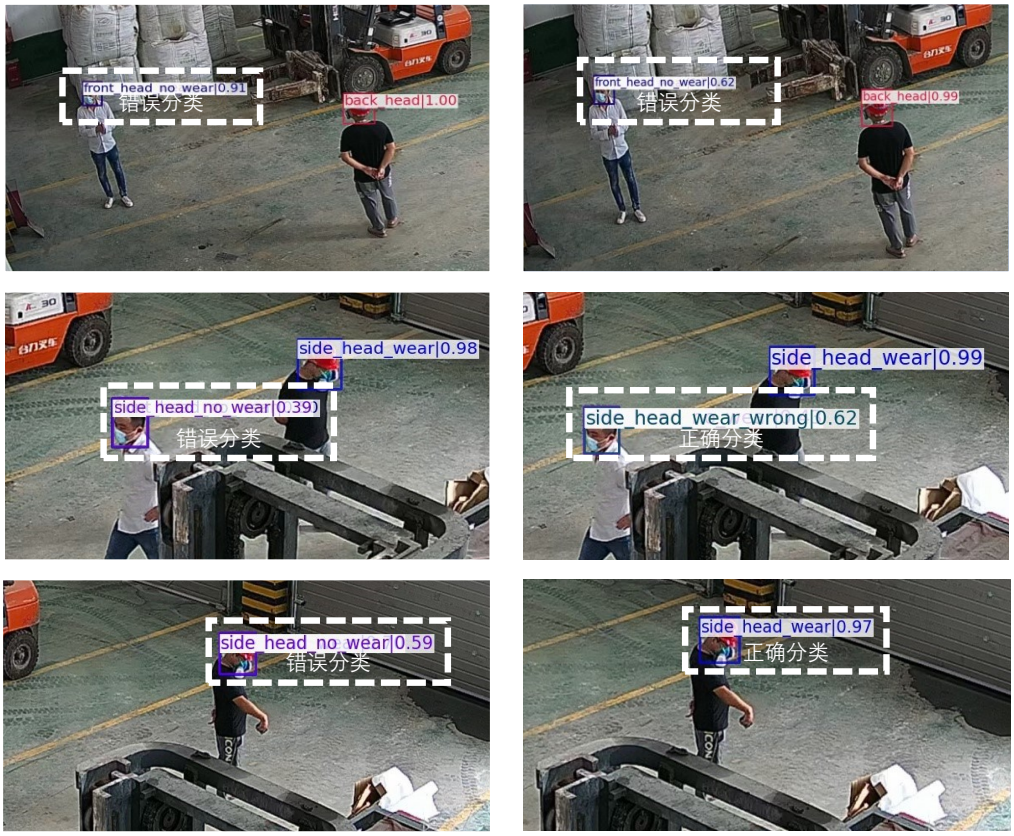
Fig. 4-6 Detection effect of improved Faster R-CNN

表 4-2 展示了各个类别的F_1得分，用于评估模型在分类能力上的性能。图 4-7 展示了改进前后的分类效果对比。总体而言，改进模型在目标分类方面的准确性有待提高。第 2、3 和 4 类别的F_1分数分别为 69.4%、72.0%和 62.6%。这些类别的F_1分数较低是由于检测器产生了 100 多个错误的阳性预测所致。同时，第 7 类别也存在问题，其产生了 88 个错误的阴性预测，从而降低了召回率和F_1分数。另外，由图 3-7 数据可知，在未增强的数据集中类别 6 和 7 数量较少，而且增强后仅提高采样率而未丰富样本特征，最终导致模型无法充分学习该两类别特征。

表 4-2 每个类别的 F1 分数

Table 4-2 The F1 score of each category

Index	Category	TP	FP	FN	P(%)	R(%)	F_1(%)
1	front_head_wear	232	51	71	82.0	76.6	79.2
2	side_head_wear	194	104	67	65.1	74.3	69.4
3	front_head_no_wear	134	63	41	68.0	76.6	72.0
4	side_head_no_wear	122	116	30	51.3	80.3	62.6
5	back_head	238	24	84	90.8	73.9	81.5
6	front_head_wear_wrong	73	18	54	80.2	57.5	67.0
7	side_head_wear_wrong	46	8	88	85.2	34.3	48.9



（a）改进前　　（b）改进后

图 4-7 改进前后的分类效果对比

Fig. 4-7 Comparison of classification effect before and after improvement

通过对表 4-1、表 4-2 及图 4-6 和图 4-7 所呈现信息的综合分析，可以得出：改进模型在目标定位上具有准确性，然而在背景过滤和分类能力上存在缺陷。这表明虽然 OHEM 和数据增强方法较好的提升模型泛化能力，但数据集中存在

的类别不平衡问题仍没有得到很好的缓解。此外，数据集中相同场景的样本重复率较高，会限制数据增强的效果。当存在大量相似或重复的样本时，数据增强可能会导致增加的样本之间相似度较高，缺乏多样性。

4.2.4 对比实验

本研究对比了 Faster R-CNN 与 SSD、YOLOv3、RetinaNet、YOLOv5 和 Cascade R-CNN 等经典模型，以及 YOLOX[30]和 Dynamic R-CNN[19]等最新模型的性能。为了验证这些模型在不同输入尺度下的检测能力，本研究采用了512×512、1024×1024和1280×1280像素的输入大小进行了测试。实验结果如表 4-3 所示。

表 4-3 不同尺度下的模型精度对比

Table 4-3 Comparison of model accuracy at different scales

Input	Method	Backbone	Neck	$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)	$mAP_{0.5:0.95}$ (%)
512×512	SSD	VGG16	-	**61.2**	51.4	**42.8**
	RetinaNet	ResNet50	FPN	41.5	34.9	27.5
	YOLOv5	CSPDarknet53	PAFPN	43.3	38.5	30.5
	YOLOX	Darknet53	PAFPN	55.7	**52.0**	40.7
	Casecade R-CNN	ResNet50	FPN	38.5	27.1	23.8
	Dynamic R-CNN	ResNet50	FPN	31.0	24.1	20.1
	Our	ResNet50	FPN	51.1	46.3	36.6
1024×1024	RetinaNet	ResNet50	FPN	67.4	61.9	48.2
	YOLOv3	Darknet53	-	56.0	48.4	38.7
	YOLOv5	CSPDarknet53	PAFPN	56.2	52.4	42.0
	YOLOX	Darknet53	PAFPN	71.3	**66.3**	52.0
	Casecade R-CNN	ResNet50	FPN	72.6	65.4	52.7
	Dynamic R-CNN	ResNet50	FPN	71.0	62.8	50.7
	Our	ResNet50	FPN	**75.4**	64.9	**52.8**
1280×1280	RetinaNet	ResNet50	FPN	75.4	65.7	53.4
	YOLOv3	Darknet53	-	71.4	63.9	50.3
	YOLOv5	CSPDarknet53	PAFPN	74.0	66.7	54.4
	YOLOX	Darknet53	PAFPN	70.7	67.0	51.5
	Casecade R-CNN	ResNet50	FPN	77.7	66.3	54.4
	Dynamic R-CNN	ResNet50	FPN	75.7	66.4	54.3
	Our	ResNet50	FPN	**82.1**	**69.0**	**56.7**

在最小输入尺寸512×512像素下，SSD 取得了最高的$mAP_{0.5}$分数，达到了61.2%。然而，由于其对较大输入尺寸的支持有限，无法对其进行进一步测试。在1024×1024像素的输入尺寸下，本研究改进的模型获得了最高的$mAP_{0.5}$分数，达到了 76.3%。而 YOLOX 在$mAP_{0.75}$方面表现最好，得分为 66.3%。然而，当输入大小增加到1280×1280像素时，改进的 Faster R-CNN 在所有指标上均超过了其他模型。改进模型在$mAP_{0.5}$取得了 82.1%的分数，在$mAP_{0.75}$取得了 69.0%的分数，在$mAP_{0.5:0.95}$取得了 56.7%的分数。与其他模型相比，改进的模型在检测更高输入尺寸的目标方面展现出了较好的准确性。这些结果表明，改进的模型特别适合于检测较大输入尺寸下的目标。相比于 YOLOv5、YOLOX 和 Cascade R-CNN 等模型，改进的模型在防毒面具检测精度上更为优越。

表 4-4 不同置信度阈值下的检测效果

Table 4-4 Detection effect under different confidence thresholds

Confidence thresholds	Model	$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)	$mAP_{0.5:0.95}$ (%)
0.1	YOLOv5	74.0	66.7	54.4
	Cascade R-CNN	79.7	65.9	56.1
	Dynamic R-CNN	73.8	64.1	53.5
	Faster R-CNN	75.2	67.3	55.8
	Our	**82.6**	**76.5**	**59.9**
0.2	YOLOv5	73.3	65.8	53.3
	Cascade R-CNN	76.5	63.7	54.0
	Dynamic R-CNN	70.8	62.5	51.7
	Faster R-CNN	71.4	63.4	52.9
	Our	**77.0**	**68.5**	**56.6**
0.3	YOLOv5	70.7	63.0	51.0
	Cascade R-CNN	73.9	61.7	52.4
	Dynamic R-CNN	66.8	59.6	49.3
	Faster R-CNN	68.7	61.5	51.1
	Our	**75.1**	**67.1**	**55.4**
0.4	YOLOv5	68.4	60.8	49.3
	Cascade R-CNN	70.9	59.5	50.4
	Dynamic R-CNN	62.6	56.9	46.8
	Faster R-CNN	67.0	61.0	50.3
	Our	**73.8**	**66.2**	**54.5**

在实际操作中，通常会采用较高的置信度阈值来过滤检测框，以筛除大量的错误检测。然而，置信度阈值的增大可能会导致模型的漏报率升高，从而降低模型的稳定性。因此，只有当模型具有较高的稳定性时，它对检测到的目标的置信度才会更高，并且在提高置信度阈值时，模型的漏报率才会降低。为了验证所研究模型的稳定性，本研究在原始1920×1080输入图像上评估了该模型在不同置信阈值（0.1、0.2、0.3 和 0.4）下的检测精度。评估结果如表 4-4 所示。

通过表中数据表明，通常情况下，置信度阈值的提升会导致所有模型的 mAP 分数下降。然而，值得注意的是，在置信度阈值增大的过程中，改进的 Faster R-CNN 展现出了较好的稳定性。具体来看，当置信度阈值设为 0.1 时，改进的 Faster R-CNN 在 IoU 比例为 0.5:0.95 时的 mAP 明显优于其他模型。当置信度阈值提高到 0.2 时，改进的 Faster R-CNN 的$mAP_{0.5:0.95}$为 56.6%，仍然高于其他模型。当置信度增加到 0.3 和 0.4 时，可以发现，改进模型的$mAP_{0.5:0.95}$、$mAP_{0.75}$和$mAP_{0.5}$这三个指标虽然均有明显下降，但与其他模型相比，仍然更具优势。综上所述，改进的模型能够稳定、有效地检测防毒面具的佩戴情况。

表 4-5 不同尺度目标检测精度比较

Table 4-5 Comparison of detection accuracy in object detection at different scales

Model	mAP_s(%)	mAP_m(%)	mAP_l(%)
Cascade R-CNN	13.2	51.3	64.8
Dynamic R-CNN	14.3	50.9	63.4
YOLOv3	10.2	45.9	61.6
YOLOv5	18.9	50.7	62.1
Faster R-CNN	5.8	36.1	51.1
Faster R-CNN + FPN	15.8	51.2	66.8
Faster R-CNN + (FPN, OHEM)	37.0	54.1	67.9
Faster R-CNN + (FPN, OHEM, Aug)	**40.3**	**55.9**	**70.1**

表 4-5 展示了下不同尺度类型目标的检测精度。根据表 4-5 的数据显示，在小尺度目标检测方面，Faster R-CNN 模型的平均精度仅为 5.8%，性能较低。然而，引入特征金字塔网络后的模型在小尺度上的平均精度提升至 15.8%。进一步，通过引入在线难例挖掘算法后，Faster R-CNN 在mAP_s达到了 37.0%。最后，在引

入数据增强方法后，改进的模型在小尺度目标检测方面的mAP_s达到了 40.3%。此外，在中尺度和大尺度目标检测方面，本研究改进的模型精度也较好，相比其他模型也有一定提升。这表明本改进的模型在多尺度目标检测方面具有较好的性能表现，能够满足不同场景下的需求。

4.3 本章小结

在本章中，一个改进的 Faster R-CNN 模型被应用于防毒面具佩戴检测任务中。针对该任务中存在的多尺度检测和类别不均衡问题，本研究采用了特征金字塔、在线难例挖掘、Mixup 和 Mosaic 混合增强等方法进行改进。本章首先介绍了 Mixup 和 Mosaic 数据增强的原理和作用，并说明了它们在训练中的具体使用方式。紧接着介绍了 RPN 网路和特征金字塔的结构，以及在线难例挖掘算法。然后进行了消融实验来验证改进方法的有效性，并列出了改进后的算法在数据集中每个类别的分类精度。通过实验结果可以得出，改进的方法能够有效提高模型的多尺度检测能力，但无法有效缓解数据集中存在的类别不均衡问题，并且在一些情况下可能会出现背景误检测。最后进行了对比实验，并在不同输入尺度下进行了测试，结果显示改进的模型具有较好的稳定性。此外，本章还对比了在相同输入尺度下不同尺度目标检测的精度，改进的模型展现出良好的小目标检测性能。

第 5 章 基于改进 YOLOv5 的防毒面具佩戴检测

本章研究主要基于 YOLOv5 模型，针对防毒面具检测在实际工业应用场景中存在的背景误检测、数据集缺乏多样性和分类性能较低等问题进行改进。为了提升模型对背景的学习能力，充分利用数据集并增加数据集的多样性，本章提出了一种多阶段数据增强策略。该策略以 Mosaic4 和 Mosaic9 为基础，在模型训练的各个阶段采用不同的数据增强方法，不仅有效提升了模型对背景的学习效果，还充分利用了数据集，增加了数据集的多样性。此外，本章还提出了 ELAN-ESE 模块以增强 YOLOv5 主干网络。该模块将 ELAN 结构与通道注意力机制相结合，使模型在保持轻量化的同时获取更丰富的梯度流信息，提高了模型的特征提取能力。为了解决类别不均衡问题，在训练中采用 Varifocal loss 作为分类损失函数。该损失函数能够提高模型对正例难样本的关注度，降低负样本的损失贡献，对复杂场景具有良好的适应性。通过这些改进，该模型在实际工业场景中展现出了较好的准确率和实时性。

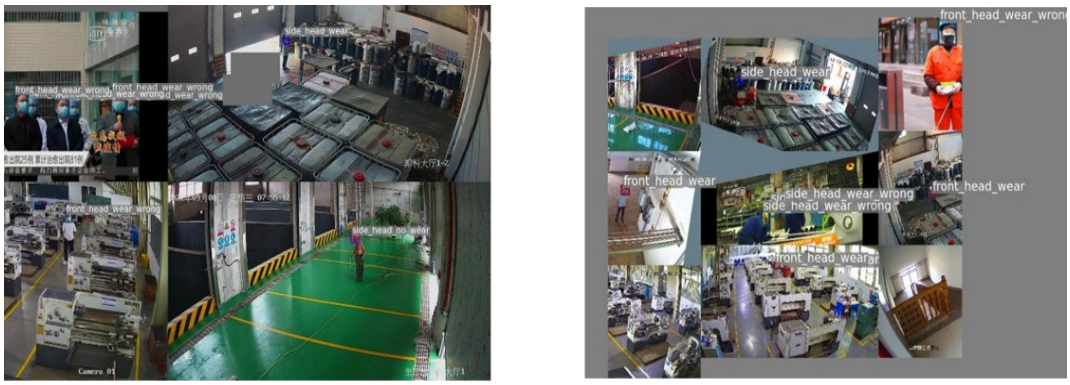
5.1 YOLOv5 模型研究与改进

5.1.1 多阶段数据增强策略

为了提升模型的泛化能力与准确性，深度学习模型通常需要使用大量的数据集来开展训练。然而，本工作在工业场景中获得的数据十分有限，存在样本量不足、重复度高和背景占比高等问题。这些问题可能会致使模型训练出现过拟合的情况，同时也会削弱对前景的学习效果，进而导致背景误识别。为了解决数据量不足引起的模型过拟合、背景误识别等问题，本研究在训练过程中提出了一种基于 Mosaic4 和 Mosaic9 的多阶段数据增强策略来扩充数据集。

如图 5-1 所示，Mosaic4 通过将四张图像合成一张的方式实现数据增强，而 Mosaic9 则是把九张图像组合成一张。与 Mosaic4 相比，Mosaic9 将 9 张图像融入到一个训练样本中，能够更好地融合不同目标和背景的信息，不仅扩大了训练集的规模，还增加了样本的多样性，有助于模型更好地学习不同场景下目标的特征。对于本工作所构建的防毒面具数据集，Mosaic9 在训练中引入了更多的变化和复杂性，从而为深度学习模型提供了更丰富的信息。此外，合成的图像中包含

了更多的场景和对象，模型需要学会同时处理和理解这些不同元素之间的关系和交互。这种综合性的训练使模型具有更强的适应能力，能够更好地应对未知的输入和复杂的场景。



（a）Mosaic4　　（b）Mosaic9

图 5-1 Mosaic4 与 Mosaic9 数据增强

Fig. 5-1 Mosaic4 and Mosaic9 data augmentation

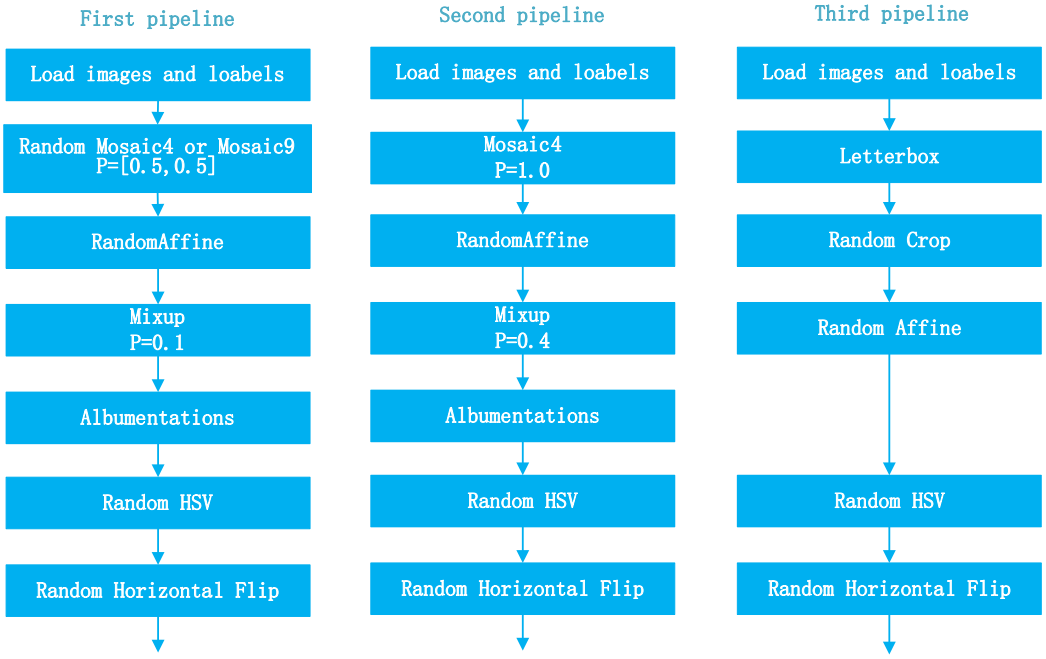


图 5-2 多阶段数据增强策略

Fig. 5-2 Multi-stage data augmentation strategy

然而，长时间使用 Mosaic9 或 Mosaic4 可能会导致模型学习到过多的冗余背景特征和一些无用的小目标信息，从而使模型难以收敛。为了解决这个问题，本工作从 YOLOX[30]的数据增强策略中得到启发，提出了一种多阶段数据集增强策

略。如图 5-2 所示，在初始训练阶段，该策略使用 Mosaic4、Mosaic9 和 Mixup 来促进模型对多样化背景特征的学习。在这个阶段，首先以 0.5 的概率对每个批次的数据应用 Mosaic9 和 Mosaic4 进行增强。然后，通过随机仿射变换将增强后的图像还原为正常状态。最后，以 0.1 的概率对该批次的样本应 Mixup 数据增强。这种增强方式可以有效提高样本的多样性，但生成的样本较为复杂，长时间使用可能会导致模型难以收敛。因此，在后续阶段，使用 Mosaic4 和 Mixup 来引导模型关注目标本身的特征，区分前景物体和背景，降低模型训练的难度，提高模型的收敛速度。最后，使用常规的增强技术进行微调，使模型更好地学习目标的特征信息。

5.1.2 改进 YOLOv5 模型

为了应对防毒面具佩戴检测应用中背景误检测、分类不准确等问题，且兼顾检测精度和速度，本节提出一种改进的 YOLOv5 模型。图 5-3 详细展示了本研究提出的模型结构。

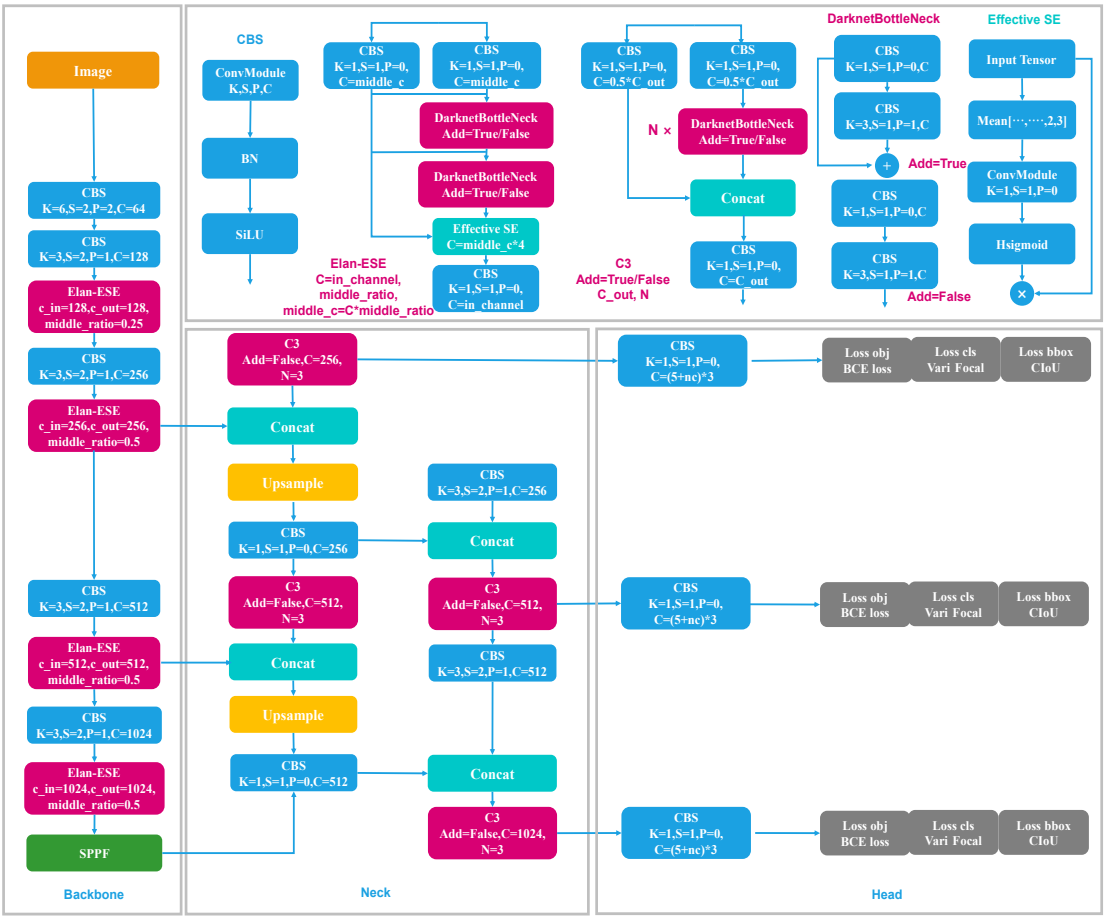


图 5-3 改进的 YOLOv5 模型

Fig. 5-3 The structure of improved YOLOv5 model

该改进算法引入了一个名为 ELAN-ESE 的新模块，它将高效的 ELAN[32]网络与通道注意力模块（Effective SE[71]）相结合。通过这种结合，使模型在保持轻量的同时，进一步提高了模型的稳定性，降低误检率。此外，为了缓解数据集中存在的类别不均衡问题，本研究采用了 Varifocal loss[68]来代替 BCE loss 损失函数。以下是对改进方法的描述：

(1) ELAN-ESE 模块

ELAN 网络是一种高效层聚合网络，它通过调控最短与最长梯度路径来构建深层次网络，从而有效促进模型提取复杂特征并提高收敛速度[32]。该设计理念不仅能使模型保持轻量特性，还能让其获得更全面的梯度流信息，进而提升模型整体性能。图 5-4 展示了 ELAN 网络的结构，其中“nb”代表特征提取块的数量，“nc”代表特征提取块中卷积模块的数量，“middle_ratio”代表输入通道的缩放系数，“index”代表每个特征提取块的序号。如图 5-4 所示，输入张量首先会经过两个卷积模块。其中，上方的卷积模块的输出会同时进入第一个特征提取块和 Concat 模块中，下方的卷积模块的输出会直接进入 Concat 模块。在 ELAN 块中，每个特征提取块的输出都会分别送入到后续的特征提取块和 Concat 模块中。最后，所有块的输出都会汇集到 Concat 模块中，并经过一个卷积块进行通道调整。由此可见，ELAN 网络可以通过调整“nc”、“nb”和“middle_ratio”可控制网络深度与输出宽度，从而有效进行特征提取。

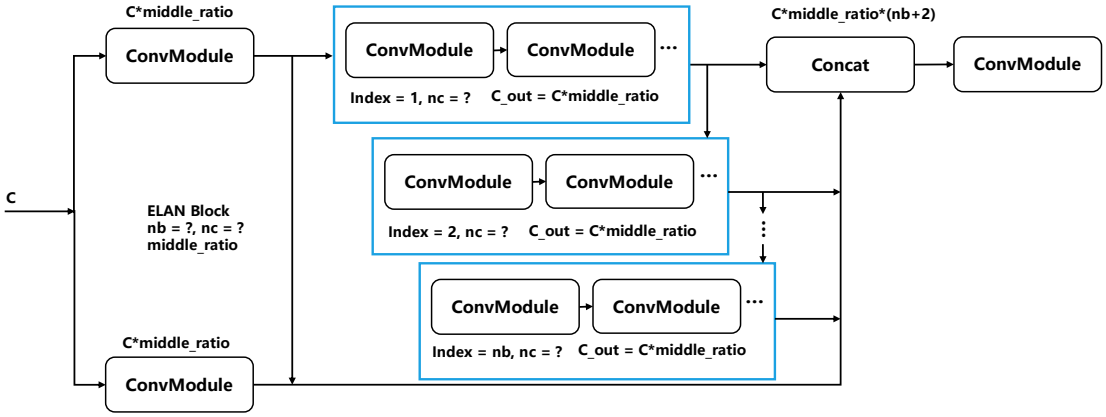


图 5-4 ELAN 网络的结构

Fig. 5-4 The structure of ELAN

本研究重新设计了 ELAN 模块，并引入了通道注意力机制，以确保在实现轻量化的同时增强其特征提取能力。如图 5-5 所示，本研究将 ELAN 块中的卷积特

征提取块替换为 DarkNet BottleNeck[24]，以进一步提升特征提取能力，并在 Concat 模块之后添加了通道注意力模块 Effective SE，以便更好地捕捉通道中的特征信息。在卷积特征提取块中，本研究使用了两个 DarkNet BottleNeck，每个 DarkNet BottleNeck 都包含两个卷积模块，即“nc=2，nb=2”。DarkNet BottleNeck 在卷积层中引入了额外的非线性激活，并通过残差连接有效提高了网络的特征表达能力。在通道方面，如图 5-4 所示，除了第一个 ELAN-ESE 模块之外，其余的 ELAN-ESE 模块使用的“middle ratio”为 0.5。这种设计使得具有C个输入通道数的特征图能够通过四条不同的提取路径进行处理，并通过 Concat 操作将它们合并成一个具有$2C$通道数的特征图。提高通道数可以让模型学习到更多的特征信息，从而更好地理解数据。每个通道都可以学习到不同维度的特征，因此增加通道数可以帮助网络捕捉更丰富的特征，以提高模型的鉴别能力。

然而，过高的通道数也会带来一些问题。首先，它会增加模型的复杂性和计算量，导致训练和推理速度变慢。其次，过多的通道数可能会导致冗余信息的出现，从而降低模型的性能。为了解决此问题，本研究在 Concat 操作后加入了通道注意力机制 Effective SE 模块。Effective SE 是 SE[70]的一个改进版本，它能够自适应地对通道特征的重要性进行调整，强化重要特征，抑制非重要特征，进而增强网络的表示能力。经过 Effective SE 模块处理后，输出的特征图将进入 CBS 模块。该模块的主要功能是将输出特征图的通道数还原至输入特征图的通道数。借助这两个模块，模型能够充分利用通道信息，同时确保模型的计算效率。

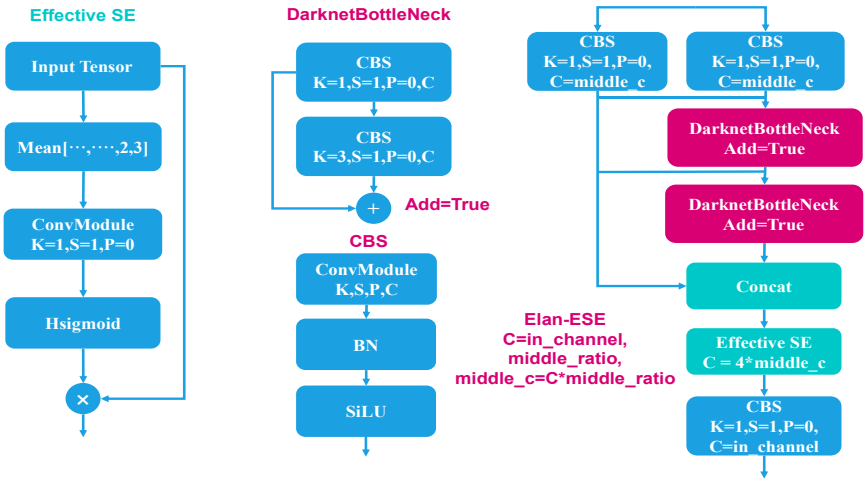


图 5-5 ELAN-ESE 模块的结构

Fig. 5-5 The structure of ELAN-ESE

(2) Varifocal loss

在 YOLOv5 中，损失函数包括三个部分：分类损失、边界框回归损失和置信度损失。分类损失和目标置信度得分损失通过二值交叉熵损失进行计算，边界框回归损失则通过 IoU 函数进行计算。整体损失函数如公式 5-1 所示。

$$L_{YOLOv5} = L_{cls} + L_{reg} + L_{obj} \tag{5-1}$$

然而，在多类分类问题中，不同类别的样本数量可能存在显著差异，从而导致类不平衡问题。二值交叉熵损失函数会倾向于将更多的注意力放在预测概率较高的类别上，也就是数量较多的类别。为了解决这个问题，本研究使用 Varifocal loss 代替二值交叉熵损失函数来缓解数据集中的类不平衡问题。Varifocal loss 的优点在于其能够通过对不同目标的重要性进行不对称处理，有效处理目标检测中的类别不平衡和难易样本问题，从而显著提升检测结果的质量。Varifocal loss 损失函数[68]定义如下:

$$VFL(p,q) = \begin{cases} -q(q\ \log(p) + (1-q)\ \log(1-p)) & q > 0 \\ -\alpha p^{\gamma} \log(1-p) & q = 0 \end{cases} \tag{5-2}$$

公式中，p表示模型预测正样本的概率，q表示样本的标签。当输入是一个正样本时，q表示预测边界框和真实边界框之间的交集和并集比。当输入为负样本时，$q = 0$。α是一个权重参数，用于平衡正负样本的数量。γ是一个可调参数，允许模型调整损失函数的难度。由公式可知，当q的值较大时，意味着当前预测框为正样本的概率较大，从而导致其 loss 值增大。这样，模型可以更加关注难以分类的正例。当q为 0 时，使用标准 Focal loss[66]来计算损失值。

5.2 实验结果与分析

5.2.1 实验环境

本章所有实验在 Ubuntu20.04 系统的主机上完成。硬件环境：Intel(R) Core(TM) i9-10940X CPU @3.30GHz，64G 内存，NVIDIA GeForce RTX 3090 GPU，24G 显存。软件环境：Python3.8，Pytorch1.7.1，Cuda11.2。

5.2.2 实验设置

本章实验中所采用的训练参数如表 5-1 所示。其中，输入图像大小设置为 1280×1280，批次大小为 8，训练周期为 60 轮。为了能得到更好的训练效果，本研究在训练过程中使用了指数滑动平均（Exponential Moving Average, EMA）

技术。由于显存限制，Faster R-CNN 等两阶段模型无法使用 EMA。为了评估模型的有效性并发现可能存在的问题，本研究使用了 COCO 指标并进行了一系列的测试。

表 5-1 训练参数设置

Table 5-1 Training parameter settings

参数	值
Input image size	1280×1280
Batch sizes	8
Epochs	60
Optimizer	SGD
Momentum	0.937
Weight decay	5e-4
EMA	True
MomentumEMA	2e-4

5.2.3 消融实验

为了全面评估不同改进策略对防毒面具检测模型的影响，本研究基于防毒面具数据集进行了一系列消融实验，并将结果汇总于表 5-2 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 YOLOv5s 作为基准模型，其检测结果$mAP_{0.5}$为 85.0%，$mAP_{0.75}$为 68.4%，$mAP_{0.5:0.95}$为 56.8%。在实验 2 中，将多阶段数据增强方法引入到模型训练中，基准模型的性能得到了较好的提升。在实验 3 中，将 ELAN-ESE 模块引入 YOLOv5s，并使得$mAP_{0.5}$达到 87.8%，$mAP_{0.75}$达到 69.0%，$mAP_{0.5:0.95}$达到 58.6%。此外，在实验 4 中评估了引入 Varifocal loss 对模型性能的影响。从表格第四行可以看出相较于基线模型，$mAP_{0.5}$提升 3.2%，这与实验 3 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。

实验 5 和实验 6 主要研究了不同方法组合对模型性能的影响。在实验 5 中测试了同时应用多阶段增强和 ELAN-ESE 的效果。结果显示，$mAP_{0.5:0.95}$与实验 2 一致，均为 58.1%; $mAP_{0.5}$和$mAP_{0.75}$略高于实验 2。但与实验 3 相比，其精度有所降低。实验 6 同时运用了 ELAN-ESE 和 Varifocal loss。结果显示，$mAP_{0.5:0.95}$

为 59.8%，相比实验 4 提高了 1.1%。基于实验 4、5 和 6 的结果，在训练中使用 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 7 中，本研究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。具体而言，与基线模型相比，所提方法在$mAP_{0.5}$指标上提高了 4.4%，$mAP_{0.75}$指标上提高了 4.7%，$mAP_{0.5:0.95}$指标上提高了 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 5-2 消融实验

Table 5-2 Ablation experiment

Index	Multi-pipeline	ELAN-ESE	Varifocal loss	$mAP_{0.5}$(%)	$mAP_{0.75}$(%)	$mAP_{0.5:0.95}$(%)
1				85.0	68.4	56.8
2	√			86.5	69.2	58.1
3		√		87.8	69.0	58.6
4			√	88.2	68.9	58.7
5	√	√		86.9	69.3	58.1
6		√	√	90.2	70.1	59.8
7		√	√	89.4	73.1	60.8

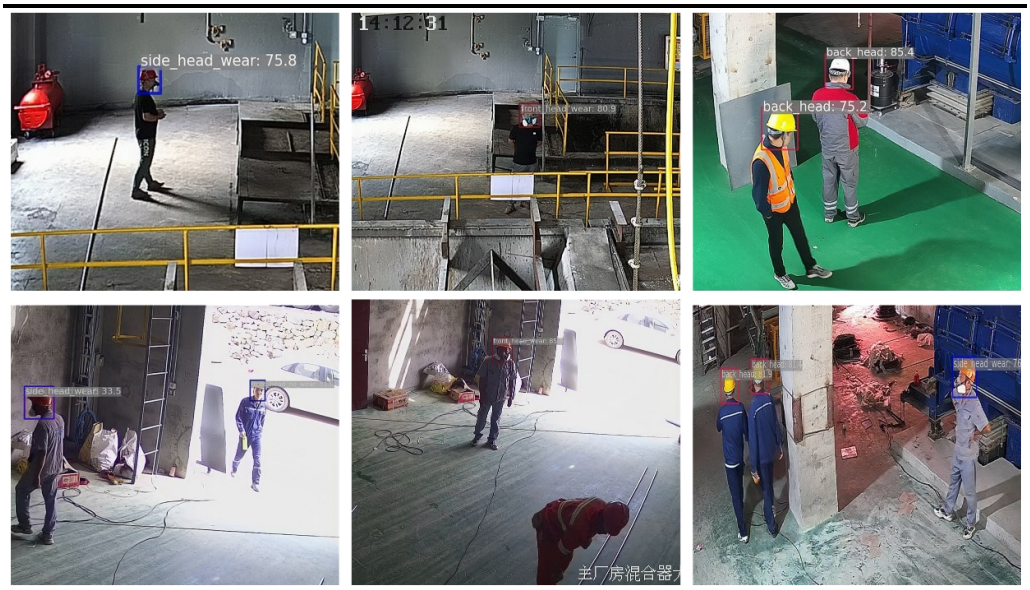


图 5-6 改进后的 YOLOv5 检测效果

Fig. 5-6 The detection effect of improved YOLOv5

图 5-6 展示了改进后的 YOLOv5 模型的检测效果。通过对图中检测结果的观察可知，改进后的模型在复杂背景下表现出了较好的稳定性，未出现误检测的情况，能够实现准确的定位和分类。但在一些背光条件下，模型检测的置信度会有所降低。

为了评估提出的多阶段数据增强方法在数据集上的增强效果，本研究使用 YOLOv5s、YOLOv6s 和 YOLOv7t 模型进行了三组对比实验，并与仅使用 Mosaic4 和 Mixup 混合增强进行了对比。实验结果如表 5-3 所示。与原始混合增强相比，多阶段增强的 YOLOv5s 模型性能有明显提升，$mAP_{0.5}$和$mAP_{0.5:0.95}$分别提升了 1.5%和 1.3%。对于 YOLOv6s 模型，使用多阶段增强将$mAP_{0.5}$提升 1.1 %，将$mAP_{0.5:0.95}$提升 1.4 %。对于 YOLOv7t 模型，采用多阶段增强方法将$mAP_{0.5}$和$mAP_{0.5:0.95}$分别提升 1.5 %和 1.4 %。实验结果表明，与仅使用 Mosaic4 和 Mixup 混合增强相比，多阶段数据增强方法能够提高模型的泛化性能。这反映了所提出的多阶段数据增强策略能够有效缓解数据集中样本重复率高的问题，从而提高样本的多样性。

表 5-3 不同数据增强方法的比较

Table 5-3 Comparison for different data augmentation methods

Model	Augmentation method	$mAP_{0.5}$(%)	$mAP_{0.5:0.95}$(%)
YOLOv5s	Mixup&Mosaic	85.0	56.8
YOLOv5s	Multi-pipeline	86.5(+1.5)	58.1(+1.3)
YOLOv6s	Mixup&Mosaic	86.6	56.6
YOLOv6s	Multi-pipeline	87.7(+1.1)	58.0(+1.4)
YOLOv7t	Mixup&Mosaic	80.9	52.8
YOLOv7t	Multi-pipeline	82.4(+1.5)	54.2(+1.4)

表 5-4 展示了改进前后模型检测各个类别的AP得分，用于评估模型在分类能力上的性能。从表 5-4 的测试结果可以明显看出，类别 5、6 和 7 的$AP_{0.5:0.95}$值均低于 50%，这表明模型在检测这三种类型的目标时性能不稳定。然而，在采用改进的方法后，模型的性能得到了显著提高。具体而言，第 6 类和第 7 类的$AP_{0.5}$分别提高了 7.2%和 19.3%，$AP_{0.5:0.95}$分别提高了 9.6%和 10.9%，第 5 类的$AP_{0.5}$和$AP_{0.5:0.95}$分别提高了 1.1%。因此，实验结果充分证明了改进的方法在缓解类别

不平衡问题和提高模型分类能力方面的有效性。但是，仍有部分类别的精度仍然较低，例如类别 5 的$AP_{0.5:0.95}$分数在改进后仍然低于 50%。这个问题需要进一步的对数据集和方法进行改进。

表 5-4 改进前后模型分类精度的对比

Table 5-4 Comparison of model classification accuracy before and after improvement

Index	Class	$AP_{0.5}$ (YOLOv5s,%)	$AP_{0.5:0.95}$ (YOLOv5s,%)	$AP_{0.5}$ (Our-s,%)	$AP_{0.5:0.95}$ (Our-s,%)
1	back_head	90.7	62.4	90.7	62.8
2	front_head_wear	94.3	66.1	95.1	67.8(+1.7)
3	front_head_no_wear	94.1	70.5	93.5	71.7(+1.2)
4	side_head_wear	88.6	61.2	92.1(+3.5)	64.0(+2.8)
5	side_head_no_wear	65.3	46.9	66.4(+1.1)	48.0(+1.1)
6	front_head_wear_wrong	89.5	45.6	96.7(+7.2)	55.2(+9.6)
7	side_head_wear_wrong	72.4	44.9	91.7(+19.3)	55.8(+10.9)

(a) 改进前

(b) 改进后

图 5-7 分类效果的可视化对比

Fig. 5-7 Visual comparison of classification effects

为了验证所提出方法的有效性，本研究在图 5-7 中展示了改进前后的检测结果。从结果中可以清楚地观察到，所提出的方法具有良好的分类能力和稳定性。

这表明该方法能够提取更复杂的特征信息，有效缓解类别不平衡问题，从而提高了检测的精度。此外，从图中还可以观察到在实际场景中所面临的挑战，例如工业环境中的照明变化和不同拍摄角度等问题。尽管本文的方法能够成功检测到昏暗灯光或高拍摄角度下的目标，但检测的置信度相对较低。解决这些挑战将成为未来改进的重要方向之一。

5.2.4 对比实验

本小节将改进方法与其他方法进行了比较。实验结果详见下表 5-5。首先，该方法与 RetinaNet、Cascade R-CNN 和第 4 章改进的 Faster R-CNN 等方法进行了比较。实验结果表明，改进的方法（“s”尺寸）在$mAP_{0.5:0.95}$指标上比 Faster R-CNN 提高了 4.7%，比 Cascade R-CNN 提高了 2.8%。随后，改进的方法与一些最新的方法进行比较，包括 YOLOX、YOLOv6 和 YOLOv7。由于本文改进的方法中尚未实现“l”和“x”尺寸的模型，因此仅比较了“s”和“m”尺寸的模型。

表 5-5 模型性能比较

Table 5-5 Performance Comparison of Models

Model	$mAP_{0.5}$(%)	$mAP_{0.5:0.95}$(%)	Param(M)	GFlOPs	FPS	Latency(ms)
Faster R-CNN	82.1	56.7	41.1	322.3	42.8	23.4
Cascade R-CNN	88.8	58.0	69.1	350.3	42.8	23.4
Retinanet	83.7	54.0	36.2	207.0	42.3	23.6
YOLOXs	86.6	56.1	8.9	53.3	72.7	13.8
YOLOv6s	87.7	58.0	9.6	49.3	74.6	13.6
YOLOv7t	82.4	55.2	6.0	20.3	100.0	10.0
YOLOv5s	86.5	58.1	7.0	31.8	99.1	10.1
Our-s	89.4	60.8	16.4	38.8	93.6	10.7
YOLOXm	88.2	59.5	25.2	147.0	45.9	21.8
YOLOv6m	87.9	59.2	34.2	162.0	45.5	22.0
YOLOv5m	88.3	60.9	20.8	96.2	71.5	14.0
Our-m	90.5	61.3	36.9	86.0	80.7	12.4

在”s”尺寸的模型中，本文所提出的改进模型表现优于多数其他模型。具体来说，当 IoU 阈值为 0.5 时，改进模型的精度达到了 89.4%，相较于 YOLOv5s 高

出 2.9%，相较于 YOLOv6s 高出 1.7%。在$mAP_{0.5:0.95}$指标上，改进模型的精度达到了 60.8%。此外，与 YOLOv5s 和 YOLOv6s 相比，改进的模型在$mAP_{0.5:0.95}$指标上提升了约 2% 对于”m”尺寸的模型，与 YOLOv6 和 YOLOX 相比，改进模型在$mAP_{0.5:0.95}$方面有 2%的提升，但与 YOLOv5 相比，仅提升了 0.4%。综合实验结果表明，本文改进的 YOLOv5 模型已具备较高的准确率，进一步增强了 YOLOv5 对防毒面具的检测能力。

在参数数量方面，本研究提出的方法虽然均高于其他模型，但在输入尺度为 1280×1280 的情况下，计算量并未成倍增加。这证明了 ElAN-ESE 模块的轻量性。在速度方面，本研究改进方法的”s”版本的 FPS 达到了 93.6，”m”版本的 FPS 达到了 80.7，可以满足实时检测的要求。

表 5-6 不同尺度目标的检测准确率对比

Table 5-6 Comparison of detection accuracy of objects at different scales

Model	mAP_s(%)	mAP_m(%)	mAP_l(%)
Faster R-CNN	40.4	51.7	66.9
YOLOv5s	18.9	50.7	62.1
YOLOv5m	23.6	56.3	71.1
YOLOXs	19.5	50.4	60.3
YOLOXm	24.2	56.6	66.5
YOLOv6s	11 .7	55.9	64.2
YOLOv6m	22.0	58.2	66.4
YOLOv7t	10.4	50.8	59.2
Our-s	17.2	57.3	69.9
Ours-m	19.8	57.7	70.4

表 5-6 对比了各个模型检测不同尺度目标的精度。其中，Faster R-CNN 的检测结果来自于第 4 章的改进方法。从表格数据可以看出，在小尺度目标检测方面，Faster R-CNN 的mAP_s达到了 40.3%，明显高于其他模型。在中尺度和大尺度上的精度与其他模型相差不大。综上所述，虽然本研究改进的 YOLOv5 模型在精度上普遍高于其他模型，但在小目标检测方面相较于其他模型仍存在一定的不足。因此，提高小目标检测的精度将是本研究未来需要努力的方向。

5.3 本章小结

在本章中，一种基于 YOLOv5 的改进模型被提出，并将其应用于防毒面具佩戴检测任务。针对数据集中存在的重复度高、背景复杂多变和类别不均衡等问题，采用了多阶段数据增强和 ELAN-ESE 模块等方法进行改进。首先，详细介绍了多阶段数据增强策略，并说明了每个阶段具体的增强方式。接着，介绍了 ELAN-ESE 模块的结构和 Varifocal loss。为了验证改进方法的有效性，本章进行了消融实验，对比了数据增强策略的效果以及改进算法在数据集每个类别上的分类精度。实验结果表明，改进的方法能够有效提高模型的分类能力，使模型能够更好地适应复杂场景，减少误检测。最后，进行了对比实验，结果显示改进的模型具有较好的精度，并且检测速度也满足应用需求。此外，该模型还与其他模型对比了不同尺度目标下的检测性能，结果显示改进的模型在小目标上的性能较低，这是未来需要改进的方向性。

第 6 章 总结与展望

6.1 工作总结

在工业生产中，防毒面具作为一种重要的呼吸保护用品，被广泛应用于包含有毒有害气体、粉尘、颗粒物等的工业生产场景中。然而，由于部分工人安全意识淡薄，未能按照规定正确佩戴防毒面具，从而导致呼吸道损伤和中毒等事故的发生。为解决此问题，构建一个实时监测工人防毒面具佩戴情况的系统至关重要。该系统利用摄像头和算法实时检测工人的面部状况，以判断工人是否正确佩戴了防毒面具。当发现工人未正确佩戴防毒面具时，系统会立即发出警报，提醒工人及时佩戴，从而避免事故的发生。此外，该检测系统还能提高工人的安全意识。通过实时监测和警报提醒，工人能够更加清楚地认识到正确佩戴防毒面具的重要性，从而更加自觉地遵守安全规定。本文采用基于深度学习的目标检测方法来解决工业场景中的防毒面具佩戴检测问题。针对检测中存在的类别不均衡、多尺度、复杂背景干扰等问题，对模型进行了改进。本文主要完成了以下工作：

(1) 为了训练基于深度学习的目标检测模型，足量的数据支撑是必要的。因此，本研究首先进行了实地调查，深入工厂内部收集了大量的视频素材。接下来，对这些视频素材进行了一系列的处理，包括视频抽帧和数据清洗，以生成原始的数据集。然后，根据实际检测需求，设计了一种合理的标注策略。该策略充分考虑了目标物体的各种属性和特征，以及它们在视频中出现的方式和位置。之后，运用该策略对原始数据集进行标注，为后续的训练和测试提供了基础。在标注完成后，还对数据进行了统计分析，初步揭示了检测任务中可能面临的挑战，如目标物体的尺寸分布、形状变化以及背景的复杂程度等。最后，选择了合适的离线增强策略来扩充数据，丰富了数据的多样性。

(2) 针对工业场景下防毒面具检测存在的多尺度检测和类别不均衡等问题，本文以 Faster R-CNN 为基础，进行了一系列改进。首先，在 Faster R-CNN 中引入了特征金字塔模块，该模块能够有效地融合不同尺度的目标特征，从而提高模型的多尺度检测能力。其次，在训练过程中采用了在线难例挖掘算法，该算法能够提升模型对难以分类样本的关注度，从而在一定程度上缓解了类别不均衡问题。

此外，为了提升样本的多样性，增强模型泛化能力，在训练中使用 Mixup 和 Mosaic 混合数据增强策略。通过实验表明，这些改进能够有效地提高模型的多尺度检测能力，尤其是小目标检测能力得到提升。同时，在不同的置信度下模型也表现较好的稳定性。然而，该模型在复杂场景中仍存在一些问题，容易导致误识别和目标丢失等问题。此外，模型的分类能力也未能够达到预期。

(3) 为满足工业应用对模型速度的要求，并有效处理复杂背景识别和类别不均衡问题，本文提出了一种基于 YOLOv5 的改进算法。首先，设计了 ELAN-ESE 模块。该模块将 ELAN 网络与通道注意力模块有效地结合起来，使模型在保持轻量级的同时，能够获得更丰富的梯度流信息，从而提高了模型的特征抽取能力。其次，为缓解数据集中样本重复度过高、类别不均衡问题，提出了多阶段数据增强策略，并在训练过程中使用了 VariFocal Loss。多阶段数据增强策略基于 Mosaic9、Mosaic4，在训练的不同阶段使用不同的混合数据增强，能够有效地生成更多的样本，让模型充分学习到背景特征与前景特征。最后，通过实验表明，改进后的模型能够较好地应对复杂缓解，降低误检测率。同时，分类能力得到提高，但小目标检测的能力还有待提升。这是本文未来研究的方向。

6.2 未来展望

在未来研究中，本工作将继续深入研究基于深度学习的目标检测算法，以进一步提高防毒面具佩戴检测的准确性和鲁棒性。具体而言，本工作将关注以下几个方面：

(1) 数据集完善：目前所使用的数据集可能存在标注错误或不均衡等问题，这会对模型的性能产生一定影响。未来将进一步完善数据集质量，提高模型的泛化能力。通过数据清洗、标注修正等方式来消除数据集中的噪声和偏差，使模型能够更好地从训练数据中学习到有用的特征和模式。

(2) 算法优化：尽管本文的方法在防毒面具佩戴检测中已取得较好的效果，但仍存在一些影响检测稳定性的问题。未来将继续研究更先进的深度学习算法和技术，如 Transformer、DETR 等，以进一步提高模型的检测性能。同时改进检测流程，尝试将目标检测与目标分类解耦。具体而言，让目标检测模型专注于检测工人的头部区域，随后将检测到的头部区域送入分类模型进行分类。

参考文献

[1] 王宇麒. 联合国工业发展组织《确保工业安全与防护手册》英汉翻译实践报告[D].东华大学,2023.
[2] 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,2020.
[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,2023.
[4] 罗国富,王源,李浩,等.基于改进 YOLOv5s 的智能车间工人不安全行为实时检测方法[J].计算机集成制造系统, 2024:1-15
[5] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.
[6] 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.
[7] 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.
[8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[16] 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.
[17] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE

international conference on computer vision. 2017: 2961-2969.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016, 9905():21-27.
[23] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
[24] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[25] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[26] 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.
[27] 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.
[28] 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.
[29] 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.
[30] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.
[31] 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.
[32] 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.
[33] 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.

[34] 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.
[35] Yu Z, Huang H, Chen W, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. arXiv preprint arXiv:2208.02019 (2022).
[36] 王建,宋晓宁.融合多尺度特征的轻量级人脸检测算法[J].模式识别与人工智能,2022,35(06):507-515.
[37] 徐守坤,王雅如,顾玉宛等.基于改进 Faster RCNN 的安全帽佩戴检测研究[J].计算机应用研究,2020,37(03):901-905.
[38] 李华,王岩彬,益朋等.基于深度学习的复杂作业场景下安全帽识别研究[J].中国安全生产科学技术,2021,17(01):175-181.
[39] 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.
[40] 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.
[41] Mercaldo F, Santone A. Transfer learning for mobile real-time face mask detection and localization[J]. Journal of the American Medical Informatics Association, 2021, 28(7): 1548-1554.
[42] Sethi S, Kathuria M, Kaushik T. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread[J]. Journal of biomedical informatics, 2021, 120: 103848.
[43] 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.
[44] 李小波,李阳贵,郭宁等.融合注意力机制的 YOLOv5 口罩检测算法[J].图学学报,2023,44(01):16-25.
[45] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
[46] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 5998-6008.
[47] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 807-814.
[48] Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural networks, 2018, 107: 3-11.
[49] Simard P Y , Steinkraus D , Platt J C . Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis[C]// 7th International Conference on Document Analysis and Recognition (ICDAR 2003), 2-Volume Set, 3-6.
[50] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]// Proceedings of the AAAI conference on artificial intelligence. 2017, 34(07): 13001-13008.
[51] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv:1708.04552, 2017.

[52] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
[53] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.
[54] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015: 448-456.
[55] Ruder S. An overview of gradient descent optimization algorithms[J]. arXiv preprint arXiv:1609.04747, 2016.
[56] Zhang Z, He T, Zhang H, et al. Bag of freebies for training object detection neural networks[J]. arXiv preprint arXiv:1902.04103, 2019.
[57] 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.
[58] 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.
[59] 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.
[60] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
[61] Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International journal of computer vision, 2015, 111: 98-136.
[62] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[J]. Computer Vision–ECCV 2014, 2014, 8693: 740-755.
[63] 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.
[64] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 761-769.
[65] Li M, Zhang Z, Yu H, et al. S-ohem: stratified online hard example mining for object detection[J]. arXiv:1705.02233, 2017.
[66] 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.
[67] 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.
[68] 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.

[69] 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.
[70] 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.
[71] Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 13906-13915.

攻读学位期间的成果

攻读学位期间发表的学术论文目录

[1] Wang Bangrong, Wang Jun, Xu Xiaofeng, et al. Gas Mask Wearing Detection Based on Faster R-CNN[J]. Journal of Ambient Intelligence and Smart Environments, 2024, 16(01):57 -71.

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(KZ32023034).

[2] 基于深度神经网络的零样本目标分类方法研究，安徽省自然科学基金项目(No. 2108085QF264, 2108085QF268).

致　谢

行文至此，我的硕士生涯即将画上句号。在这篇论文完成之际，我心中充满了无尽的感激与感慨。

首先，我要感谢我的导师汪军教授。从论文的选题、开题、实验设计到论文的撰写和修改，您都给予了我悉心的指导和无私的帮助。在我遇到困难和疑惑时，您总是耐心地倾听我的问题，并给予我建设性的意见和建议，让我能够顺利地克服一个又一个难关。

我还要向我的家人致以最诚挚的谢意，是你们给予了我最坚定的支持和毫无保留、无条件的爱。你们是我最坚强的后盾，最有力的支撑，在我遇到困难、陷入低谷时，是你们给予我鼓励，给予我安慰，给予我重新振作的力量。

此外，我要感谢我的朋友们，你们是我生活中最宝贵的财富。在这几年的学习生活中，我们一起度过了许多欢乐的时光，一起探讨学术问题，共同面对和完成项目。希望我们的友谊能够长存，无论将来身处何方，我们都能保持联系。

回顾这几年的硕士生涯，我收获的不仅仅是知识和能力，更重要的是成长和历练。这段经历让我变得更加坚强、自信和成熟。在未来的日子里，我将带着这份收获，勇敢地面对生活中的各种挑战。我将继续努力学习，不断提升自己的专业素养和综合能力，为实现自己的人生目标而奋斗。最后，再次感谢所有在我成长过程中给予我帮助和支持的人们。

尚德教学 唯实惟新