分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220110135

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目 基于深度学习的车间人员安全监测与预警

论 文 作 者　刘 懿 平
校 内 导 师　江本赤 教授
校 外 导 师　赵立军 高级工程师
专业学位类别　机 械
研究方向（领域）　计算机视觉

论文提交日期：2025 年 5 月 20 日

分类号：TP391　　学校代码：10363

密　级：公开　　学　　号：2220110135

基于深度学习的车间人员安全监测与预警

DEEP LEARNING-BASED WORKSHOP PERSONNEL SAFETY MONITORING AND EARLY WARNING

学生姓名：刘懿平

校内导师：江本赤 教授

校外导师：赵立军　高级工程师

专　　业：机 械

研究方向：计算机视觉

论文答辩日期：2025 年 5 月 9 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： 

日期： 2025 年 5 月 16 日

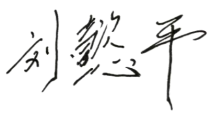
安徽工程大学学位论文版权使用授权书

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：  　　指导教师签名： 

日期：2025 年 5 月 16 日　　日期：2025 年 5 月 16 日

基于深度学习的车间人员安全监测与预警

摘 要

随着智能制造技术快速发展，生产车间的自动化与智能化水平不断提升，与此同时，车间内部的安全隐患也愈加突出。面对安全情况较复杂的车间环境，传统的安全监控系统在动态目标监测以及人机交互安全预警等方面存在诸多不足。本文针对生产车间内的安全监测问题，基于深度学习技术开展相关研究，力求解决行人状态检测、安全帽佩戴监测以及无人叉车与行人碰撞预警三个方面的问题，主要完成了以下工作。

首先，设计行人状态检测和安全帽佩戴监测算法框架。在传统YOLO目标检测算法中，引入轻量化网络结构FasterNet与具有动态聚焦机制的损失函数，并添加CBAM注意力机制。增强复杂车间环境下，行人状态及安全帽佩戴情况监测准确性和实时性。

其次，研究三维距离恢复方法，建立人机碰撞分级预警机制。在端到端的行人轨迹预测基础之上，实时提取监控系统的图像像素点，依据相机成像几何关系，在线计算人机欧氏距离，实现单目相机的人机三维距离恢复，以预警人机碰撞风险，提升车间内人机交互的安全性。

最后，设计实验方案验证本文所提方法。首先，采集车间监控图像信息，制作包含行人与叉车的数据集，并按一定比例划分训练

集与测试集；然后，将本文算法与传统算法分别在测试集中实验，验证改进算法有效性。此外，为验证本算法的泛化性能，在不同的车间光照条件和不同视角下分别采集图像进行测试；最后，在公开数据集中，将本文行人轨迹预测模型与其他模型对比；同时验证了本文方法对人机碰撞预警的有效性。

关键词：深度学习；目标检测；轨迹预测；车间安全

DEEP LEARNING-BASED WORKSHOP PERSONNEL SAFETY MONITORING AND EARLY WARNING

ABSTRACT

With the rapid development of intelligent manufacturing technology, the automation and intelligence levels of production workshops are constantly improving. At the same time, the safety hazards inside the workshops are becoming more prominent. In the face of the complex workshop environment, the traditional safety monitoring system has many deficiencies in dynamic target monitoring and human-computer interaction safety warning. In this paper, for the safety monitoring problem in the production workshop, based on deep learning technology to carry out related research, and strive to solve the pedestrian state detection, helmet wear monitoring and unmanned forklift and pedestrian collision warning three aspects of the problem, mainly completed the following work.

First, the algorithm framework for pedestrian state detection and helmet wearing monitoring is designed. In the traditional YOLO target detection algorithm, introduce the lightweight network structure FasterNet with the loss function with dynamic focusing mechanism, and

add the CBAM attention mechanism. Enhance the accuracy and real-time monitoring of pedestrian status and helmet wearing under complex workshop environment.

Secondly, the three-dimensional distance recovery method is studied to establish a graded warning mechanism for human-machine collision. On the basis of end-to-end pedestrian trajectory prediction, the image pixels of the monitoring system are extracted in real time, and based on the geometric relationship of the camera imaging, the Euclidean distance between man and machine is calculated online, so as to realize the three-dimensional distance recovery of man-machine with monocular camera, in order to warn the risk of man-machine collision and enhance the safety of human-computer interaction in the workshop.

Finally, an experimental program is designed to verify the method proposed in this paper. Firstly, the workshop surveillance image information is collected, the data set containing pedestrians and forklifts is produced, and the training set and test set are divided according to a certain ratio; then, the algorithm of this paper and the traditional algorithm are experimented in the test set respectively to verify the effectiveness of the improved algorithm. In addition, in order to verify the generalization performance of this algorithm, the images are collected under different workshop lighting conditions and different viewing angles for testing; finally, the pedestrian trajectory prediction model of this paper

is compared with other models in the public datasets; at the same time, the validity of this paper's method is verified for the warning of human-machine collision.

KEY WORDS: deep learning; target detection; trajectory prediction; shop floor safety

目录

插图清单

插表清单

第 1 章 绪论

1.1 研究背景与意义

随着工业 4.0 和智能制造的发展，现代生产车间逐渐引入自动化和智能化设备，例如，采用工业机器人和无人驾驶叉车等技术，能够有效提升生产效率并减少运营成本，如图 1-1 所示。然而，随着这些新技术的应用，车间的复杂度大幅提升，尤其是在车间内的动态环境中，行人与自动设备交互愈加频繁，车间内的安全问题日益凸显。传统的安全监控系统依赖于人工监控或基于规则的简单算法，无法有效应对当前车间内密集行人、动态目标和复杂环境带来的挑战，这导致检测准确性与实时性能难以满足实际应用的要求。

在此背景下，深度学习技术因其在图像处理、目标检测和行为预测等领域的出色表现，逐渐被应用于工业安全监控中。凭借其在复杂场景中能够更有效地提取特征。特别是卷积神经网络（Convolutional Neural Network，CNN）在目标检测领域的广泛应用，显著提升了图像识别与检测的精度和效率。同时，随着无人化设备的普及，如何避免人机交互过程中的安全事故，成为智能车间安全管理的重要课题。

图 1-1 智能化车间

本研究基于深度学习技术，旨在解决生产车间内的三大关键安全问题：密集行人检测、安全帽佩戴识别以及无人叉车与行人碰撞预警研究，不仅具备重要的理论意义，还展现出突出的实用价值。

首先，车间环境通常人员密集，行人穿行频繁，常规的目标检测算法难以在这种复杂场景下准确识别每个目标。因此，本文提出了一种优化 YOLOv5 算法，通过改进网络架构和损失函数设计，有效提升了密集行人检测的精度与实时性能。这对于提升车间安全监控的智能化水平，减少行人相关事故发生具有重要意义。

此外，车间工人正确佩戴安全帽是确保车间安全运行的关键因素之一。目前的安全帽检测技术在复杂场景下的识别精度较低，且实时性能不足。通过结合 FasterNet 与 YOLOv5 的优势，本文设计了一种轻量化且高效的安全帽佩戴检测算法，能够在不同环境中实现高效的安全检测，这有助于提高工厂安全生产管理的规范化程度。

最后，无人叉车等自动化设备的广泛应用带来了行人和叉车之间碰撞的潜在风险。通过设计基于行人轨迹预测和三维距离恢复的碰撞预警系统，本文为车间内复杂的动态交互提供了有效的安全保障机制。该系统的应用将显著降低车间内人机碰撞事故的发生率，提升智能车间的安全性与自动化水平。

本文所研究的车间人员安全检测与预警方法，将在理论与实践两方面对生产车间的安全管理起到重要的推动作用，为智能制造环境下的安全检测技术带来了新的思路和解决途径。

1.2 国内外研究现状

随着工业自动化与智能化的快速发展，生产车间的安全监测问题逐渐成为学术界和工业界关注的重点。近年来，基于深度学习的目标检测、行人轨迹预测以及安全监控技术在多个领域中得到了广泛的研究与应用，国内外研究人员在这些方向上取得了显著的成就。本文旨在从人机碰撞预警、行人检测方法以及轨迹预测研究三个视角出发，对国内外相关领域的最新研究进展进行系统梳理与综述。

1.2.1 国内外人机碰撞预警研究现状

（1）国外人机碰撞预警研究现状

国外在人机碰撞预警的研究起步较早，美国、欧洲和日本等地区在人机避障领域积累了丰富经验，研究重点涵盖感知技术、决策算法及人机交互优化。

美国谷歌通过激光雷达（LiDAR）、摄像头和雷达的多传感器融合技术，构建了高精度的动态环境模型，用于实时检测行人、车辆等障碍物[1]。Katrakazas 等人[2]提出了一种基于概率 occupancy grid 的碰撞预警方法，能有效识别动态障碍物的运动轨迹。此外，欧洲的 CityMobil2 项目[3]针对城市低速场景，开发了基于立体视觉的避障系统，显著提高了对近距离行人的检测精度。

国外研究广泛采用强化学习和深度学习优化碰撞预警策略。Schwarting 等人[4]提出了 Social Value Orientation 模型，通过预测人类行为意图实现动态碰撞预警。同时，日本丰田研究所开发了一种基于博弈论的路径规划算法，针对行人密集区域的人机交互场景，能够平衡安全性和通行效率。

国外研究还关注碰撞预警过程中与人类驾驶员或行人的交互。例如，德国的“Automated Driving HMI”项目设计了直观的提示系统，通过声音和灯光信号提醒外部行人避让。此外，特斯拉通过 OTA（Over-the-Air）更新不断优化其 Autopilot 系统的避障性能，路测数据显示其在复杂场景下的误判率逐步降低。

（2）我国人机碰撞预警研究现状

我国在人机碰撞预警领域的研发近年来加速推进，依托丰富的应用场景和政策支持，此研究逐渐形成特色，聚焦于本土化场景适应、技术融合与产业化落地。国内高校和企业如清华大学、百度 Apollo 在多传感器融合方面取得进展。清华大学魏凌涛[5]团队提出了一种基于 LiDAR 与视觉融合的碰撞预警算法，针对中国城市道路的行人密集特性优化了障碍物检测精度。哈尔滨工业大学于赫年[6]改进静态算法以适应动态环境并提出多步前瞻性主动避障算法。百度 Apollo 平台则通过开源数据集验证了其感知系统在雨雾天气下的鲁棒性，显著降低了漏检率。

1.2.2 行人检测技术研究现状

行人检测技术作为目标检测领域中备受关注的研究热点之一，对安防、辅助驾驶等多个应用场景具有重要的学术和实用价值。当前，围绕行人检测的学术研究呈现快速增长的趋势，其研究领域不断扩展，研究层次也逐步加深。例如，在该方向的研究中，既包含基础的行人检测技术探索，也涵盖了行人重识别[7][8][9]、行人姿态估计[10][11][12]的研究等。

近年来，行人检测技术取得了显著的进步[13][14][15][16]，然而该领域依然面临

诸多难题。以下从三个方面分析其主要挑战：首先，环境条件的复杂性与多样性对检测效果构成显著威胁。光照强弱、天气变化以及拍摄角度的不同均可能干扰图像质量，进而降低检测的准确性。例如，阴雨天气或逆光条件下，行人特征可能变得模糊。其次，行人姿态的多样化增加了检测难度。当行人做出蹲下、弯腰等动作时，其轮廓会发生明显变形，导致传统检测模型难以正确识别，从而影响精度。行人与环境的交互进一步复杂化了检测任务。在光线不足的环境中，身着深色衣物的行人往往难以被辨识；而当行人由于被树木、车辆等物体部分遮挡，从而导致图像信息部分缺失，使得视觉检测系统捕捉的目标特征丢失。目前，行人检测技术主要依赖以下三类方法：基于背景建模的目标检测算法，通过分离前景与背景实现行人识别；基于传统特征提取的算法，利用手工设计的特征进行目标检测；基于深度学习的目标检测算法，借助神经网络自动提取特征，近年来表现出较强的适应性和性能。综上，尽管行人检测技术在发展中取得了长足进步，但环境、姿态及交互等因素带来的挑战仍需进一步研究与突破，以提升检测系统的鲁棒性和精确性。

背景建模的目标检测算法利用目标的运动属性进行识别，通过分析图像中发生变化的区域来完成目标的检测。该类方法主要有帧差法[17]、光流法[18]、背景减除法[19]和图像匹配法[20]等。Lipton 等[21]提出一种通过判断连续两帧图像像素差值是否发生变化的检测方法。该方法具有较强的实时性，能够快速处理图像数据，但在面对运动速度过快或过慢的目标时，其检测性能会显著下降。盖绍彦[22]等人开发了一种基于 ORB 特征的图像匹配算法。该算法检测精度较高，然而在匹配时间上仍有进一步提升空间。

在行人检测算法的研究领域中，世界范围内研究人员针对特征提取方法深入地进行探索。Wang[23]等人结合 HOG 与 LBP（局部二值模式直方图）两种特征提取方法，并运用 SVM（支持向量机）对提取的特征进行分类处理。尽管这一方法在实际图像测试中取得了良好的检测效果，但由于其实时性能不足，难以广泛应用于实际场景。此外，近几年的研究中，Felenszwalb 等人[24]通过对目标内部构造进行分析，开发出 DPM（局部形变模型），该模型在检测不同姿态的行人时表现出较强的能力。

基于深度学习的行人检测技术也成为学术研究领域热点。与传统特征提取方法相比，神经网络模型在行人检测任务中的性能更加优良。现如今，许多基于深度学习的目标检测算法已相继问世，例如 Faster-RCNN[25]、SSD[26]、FPN[27]、YOLO[28][29][30][31]等。

尽管如此，这些现有方法离做到成熟应用仍存在一定差距具有诸多挑战亟待攻克。例如，当面临行人拥挤场景或者行人部分躯干被遮挡时，目标检测鲁棒性严重下降；在一些灯光昏暗场景下或有雾气情况下，会出现行人检测漏检、误检发生。在无人驾驶等视角高动态场景下，会因视场运动而造成成像模糊，从而影响算法识别准确性。因此行人检测技术仍有诸多问题亟待解决。

1.2.3 轨迹预测研究现状

随着无人驾驶技术走向成熟并逐步应用化商业化，由于智能汽车在工作过程中需通过视觉技术感知周围环境从而对驾驶路线进行合理规划，因此针对行人实时轨迹预测的技术研究成为关注的焦点领域，尤其是行人动态行为进行精确预判，以便在潜在风险出现时及时发出警报，从而有效预防交通事故的发生。当前主流的行人轨迹预测解决方案可分为两大类（1）基于传统方法：通过社会力模型或卡尔曼滤波的方法，这类方法适用于简单场景，在实际场景下效果差（2）基于深度学习方案：通过长短期记忆网络（LSTM）或生成对抗网络（GAN）等进行行人轨迹预测，此类方法对硬件算力要求高，但具有良好预测效果。

（1）传统方法轨迹预测

基于传统方法的行人轨迹预测解决方案主要依赖物理模型与统计技术，在深度学习兴起前占据主导地位并奠定了研究基础。例如，Helbing 等人[32]研究扩展社会力模型，专注于紧急情况下行人轨迹预测，通过仿真分析了恐慌状态下人群的动态特征，验证模型在异常行为预测中的适用性。Johansson[33]等人引入视频跟踪数据和进化算法优化社会力模型的参数，在真实场景中预测行人轨迹中具有良好精度。除此之外，传统方法还包括基于卡尔曼滤波[34]的预测技术以及动态贝叶斯网络[35]的预测方案。Schneider[36]研究比较了卡尔曼滤波和粒子滤波在行人轨迹预测中的性能，基于递归贝叶斯框架分析传统方法在实时性和精度上的表现，实验得证传统方法在简单场景下的效果。虽然在某些特定场景下

传统轨迹预测算法具有良好预测效果但由于行人运动的随机性、多样性、复杂性在实际场景中处理行人轨迹预测任务时仍存明显问题。

（2）基于 LSTM 的轨迹预测算法

神经具有强大的自学习能力，适合处理轨迹预测这一类的随机性强复杂程度高的任务。因此近年来，各大科研机构的研究者更偏向于用深度学习的方法解决行人轨迹预测问题。长短期记忆网络（Long Short Term Memory network，LSTM）[37]能有效捕捉时间序列中的长期依赖，在轨迹预测中，LSTM 常用于分析行人历史位置序列从而预测行人轨迹。Liu 等人[38]率先利用 LSTM 算法处理行人轨迹预测任务，但其研究具有局限性，未将行人自身运动的随机性考虑到模型构建中。随后，Alahi 等人[39]通过结合 LSTM 和社会池化层（Social Pooling）预测拥挤场景下行人轨迹，首次在深度学习框架中将行人间空间交互加入，有效提升多人轨迹预测准确性。Gupta 等人[40]在 Social LSTM 基础上引入生成对抗网络（GAN），提出 Social GAN，通过生成器预测多样化的社会可接受轨迹并由判别器评估合理性，大幅提高轨迹预测的多模态性和现实性。

现有的人机碰撞预警技术多针对城市道路或无人驾驶场景，缺乏对车间内设备遮挡、光照不足等复杂环境的适应性，且实时性与误判率难以平衡；行人检测方法在遮挡严重、姿态多样及特征模糊的车间场景中鲁棒性不足，未能充分优化工业人员行为模式；轨迹预测研究虽在简单或社会交互场景下有效，但在车间内人员与设备交互、路径随机性强的环境下预测精度下降，且深度学习方案对算力要求高，未兼顾车间实时性与低成本需求。因此，利用行人轨迹预测模型避免人机冲突的研究仍具有亟待突破的技术瓶颈。

1.3 论文主要工作与章节安排

1.3.1 研究内容

本文围绕生产车间中安全监测问题，针对行人检测、安全帽佩戴检测以及无人叉车与行人碰撞预警三大场景，基于深度学习，提出一系列算法改进与方案。本文的主要工作如下：

（1）采集真实车间工作环境下中行人及行人佩戴安全帽图像，制作行人检测和安全帽佩戴检测数据集。

（2）利用目标检测算法针对行人及其安全帽佩戴情况进行识别检测，根据车间安全生产实时性需求对算法优化。

（3）通过目标检测算法获取行人及叉车矩形区域，求解出两类候选框的中心坐标。获取相机参数后利用单目三维距离恢复方法，求解行人及叉车的欧氏距离。

（4）根据无人叉车工作场景下行人轨迹预测需求，研究轨迹预测算法，设计一种基于自动编码器的行人轨迹预测模型，利用车间监控实现端到端的行人轨迹预测。

（5）提出车间场景下人机碰撞分级预警模型，量化危险程度。

1.3.2 技术路线

本文所采取的技术路线如图 1-2 所示。

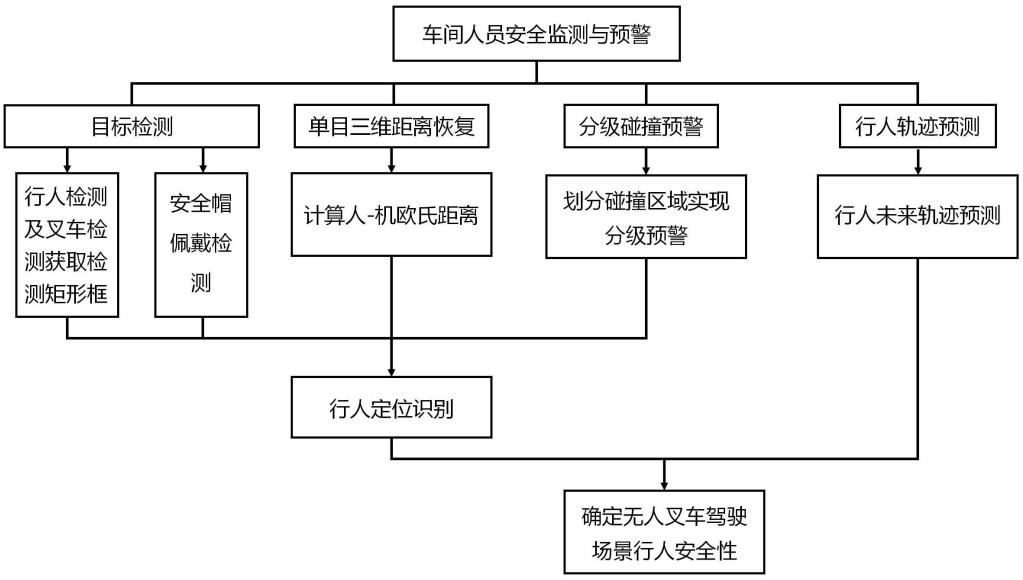


图 1-2 技术路线

1.3.3 论文章节安排

本文共分为六章，具体结构安排如下：第一章绪论；第二章介绍行人轨迹预测相关理论基础；第三章基于深度学习研究目标检测算法；第四章研究三维距离恢复方法，依据成像几何关系，计算人机欧氏距离，并建立人机碰撞分级预警机制；第五章针对行人轨迹预测环节展开研究，根据叉车行人轨迹预测的特点和需求，在厂区无人叉车工作环境下进行实验；第六章总结与展望。

第 2 章 行人轨迹预测的相关理论基础

2.1 引言

作为机器学习的重要分支，深度学习以人工神经网络为理论基础，其网络结构通过多层级联节类似生物神经系统处理信息工作原理。相较于常规启发式算法，人工神经网络具有自主学习能力。近年来学术界针对各类深度网络架构优化研究不断深入，其中卷积神经网络（CNN）与循环神经网络（RNN）分别在不同实际应用任务中展现出独特的优势。作为深度神经网络典型架构之一，卷积神经网络在目标检测与行为识别领域效果出众。其通过逐层特征变换机制，实现从原始输入数据到高阶语义特征的自动化特征学习，这种分层获取信息的能力使其能够有效捕捉数据中的有用信息。在处理时序信号方面，循环神经网络通过构建记忆单元保留历史状态信息，因此其在轨迹预测等序列分析任务中持续发挥重要作用。本章将系统阐述人工神经网络的基础理论，深入解析卷积神经网络和长短期记忆网络的机理特性，并探讨行人检测与轨迹预测领域的主流算法框架，为后续研究提供理论基础。

2.2 人工神经网络

人工神经网络理论可溯源至 20 世纪 40 年代，其发展轨迹经历了三个关键阶段：感知器时期（1940s-1980s）、BP 算法时代（1980s-2010s）以及深度学习时代（2010s 至今）。上世纪八十年代中期误差反向传播算法（BP 算法）的提出解决了多层网络参数优化难题，成为推动该领域跨越式发展的里程碑。反向传播关键作用在于，它能在预测偏差出现时，自主调整神经元的参数。通过迭代训练，神经网络能够逐步优化，以更好地解决实际问题。当前，基于深度学习的神经网络已在智能计算领域确立主导地位，广泛应用在处理复杂任务的特征学习与决策推理[41]。

图 2-1 展示了典型类脑计算模型的结构组成，其由多层级联的处理单元构成。从信号传递方向分析，该架构包含三个功能模块：信号接收接口层（输入端）、特征抽象转换层（中间处理域）以及决策输出层（结果生成端）。

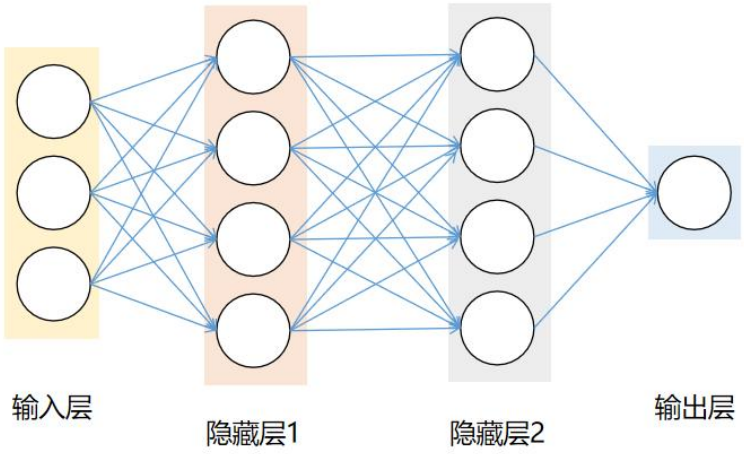


图 2-1 人工神经网络结构图

2.3 卷积神经网络

20 世纪 80 年代末期，Lecun 团队[42]在字符识别领域首次实现卷积运算的工程化应用，标志着特征提取新方法的诞生。1998 年提出的 LeNet-5 架构[43]，如图 2-2 所示。确立了现代卷积网络的基础框架，其分层特征学习机制为后续研究提供了重要启示。值得注意的是，早期研究者在尝试扩展网络深度时遭遇了模型泛化性能下降的困境，主要表现为梯度弥散现象和参数优化困境。针对深度模型训练难题，Hinton 研究组[44]于 2006 年提出分层预训练策略，揭示了深层网络在特征抽象能力方面的理论优势。2012 年 ImageNet 竞赛中突破性成果的取得[45]，不仅验证了深度卷积架构的有效性，更引发了计算机视觉研究模式的转变。随着模型鲁棒性的持续提升，该技术在工业检测、环境感知、医学影像分析等领域的应用边界不断拓展，现已成为智能计算系统的核心组件。

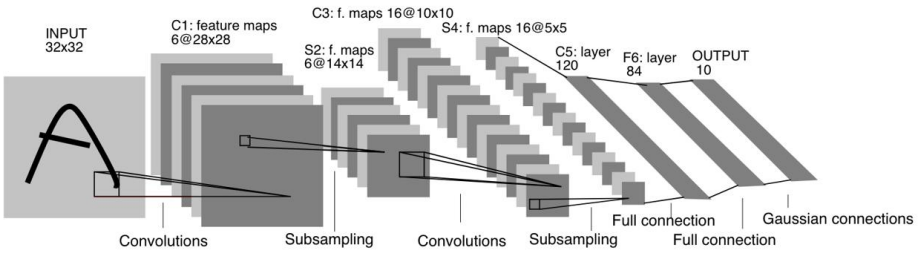


图 2-2 LeNet-5 结构图

经典 CNN 由卷积层与池化层构成核心处理链，如图 2-3 所示。卷积层生成特征图时采用零填充防止边缘信息丢失，通过步长控制采样密度；池化层则对卷积输出的空间冗余进行压缩，有效抑制过拟合。两者协同实现特征精炼与维

度控制。

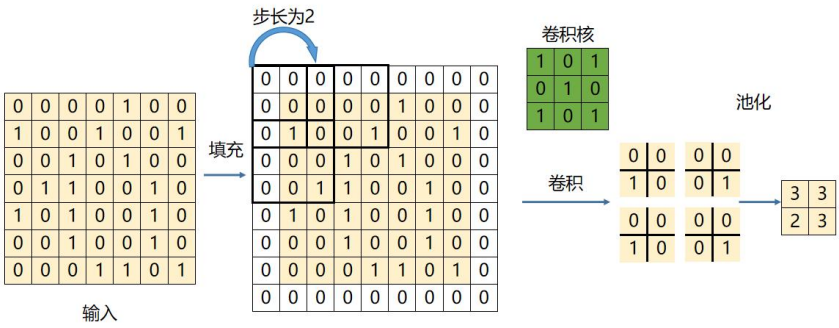


图 2-3 卷积网络操作示意图

图 2-4 展示了常用的非线性激活函数的数学特性及其梯度传播曲线，不同函数类型在梯度稳定性、计算效率等方面展现出差异化特征，使得在为满足不同任务要求需选择合适的非线性激活函数。

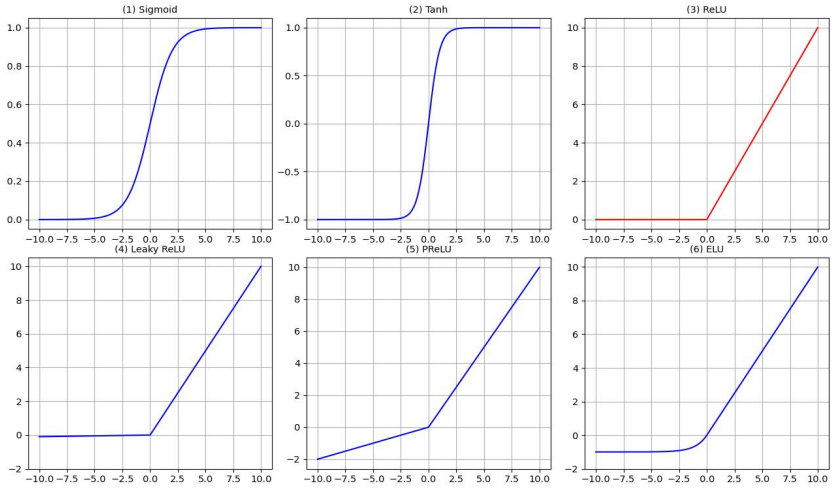


图 2-4 主要的激活函数及其图像

卷积神经网络利用损失函数评估模型输出结果与真实值之间的偏差。传统的损失函数有许多种类，其中包括平均绝对误差（MAE）、均方误差（MSE），以及用于分类任务的交叉熵损失（Cross-Entropy Loss）。

2.4 长短期记忆网络

长短期记忆网络（LSTM）首次采用门控机制重新定义了循环神经网络对时序信息的处理规则。该网络采用动态记忆管理机制，根据上下文特征自发调节历史信息存储强度与更新权重。这种自主信息筛选机制，使其在处理时间序列等具有动态特性的数据时展现出卓越的运算能力。长短期记忆网络的架构创

新体现在其三重门控协同工作机制中，包含输入门、输出门和遗忘门。其计算过程如图 2-5 所示。信息处理流程涉及神经元与门控单元的协同运算，通过多阶段非线性变换实现时序特征编码。输入门作为信息筛选模块，调控外部信息向记忆单元的输入强度；遗忘门则通过对历史记忆进行动态衰减控制，实现记忆容量的自适应优化。实证研究表明，当参数收敛时，当输入门激活值趋近零而遗忘门趋近一，系统将自动抑制当前输入信号的传播路径，同时保持上一时刻状态稳定。这种操作模式对应于模型对稳态环境特征的记忆保持机制。相反，当输入门接近于一且遗忘门接近与零，网络会优先更新记忆单元中的信息存储，此时系统的短期记忆窗口显著收缩以适应动态环境变化。这种基于门控参数动态平衡的调控方法，解决了传统递归网络在长程依赖算法中的梯度衰减效应，为复杂时序模式挖掘提供了可解释的计算框架。

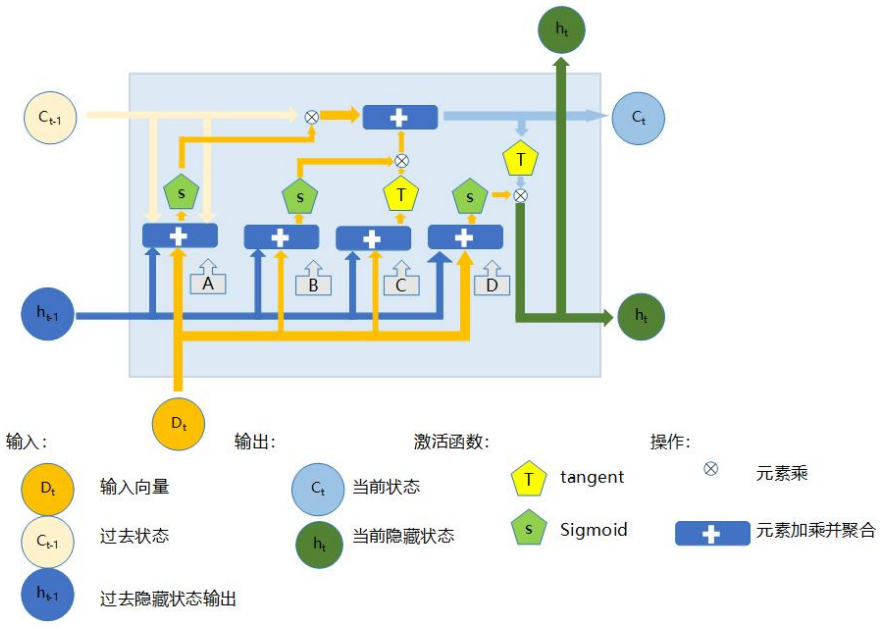


图 2-5 LSTM 计算过程示意图

2.4.1 基于人工特征的行人检测算法

在传统机器学习框架中，行人检测任务包括训练与检测两个阶段。首先通过特征处理阶段完成判别性特征构建与分类边界学习，随后在推理阶段执行多尺度滑动窗口的特征匹配与目标定位。研究人员通常基于目标在二维图像空间的形态学特征进行提取，包括边缘梯度分布、区域纹理模式以及形状轮廓特征等几何不变性特征。例如，基于矩形区域差分运算的 Haar 特征描述子、刻画局部梯度方向直方图（HOG）的统计分布特征，以及反映邻域灰度对比关系的

LBP 纹理算子等经典特征提取方案。这些人工设计的特征提取器与分类器的组合构成了早期检测系统的技术框架。LBP+AdaBoost 行人检测算法：局部二值模式（Local Binary Patterns，LBP）特征最初被用于纹理分析。Enzweiler 等人[46]首次将 LBP 特征应用于行人检测。LBP 特征能够有效地捕捉行人外观的局部纹理信息，并结合 AdaBoost 算法构建鲁棒的行人检测器。HOG+SVM 行人检测算法：2005 年，Dalal 等人提出一种行人检测方案，其核心是结合使用 HOG 特征与 SVM 分类器。HOG 特征侧重于捕捉图像的边缘信息，例如边缘的方向和强度。与其它特征相比，HOG 的优势在于能够更精准地描述局部细节，并且对光照变化等因素具有较强的鲁棒性。因此，HOG 特征非常适合用于表达行人的特征信息。

由于 HOG 特征具有严格的空间约束性。当检测目标呈现非典型姿态（如肢体大幅摆动）或存在多向运动时，其梯度方向分布的统计规律发生畸变，导致特征空间可分性显著降低。这种几何敏感特性使其难以适应工业场景等行人动作复杂环境下的检测需求。

2.4.2 基于深度学习的行人检测算法

基于深度学习的视觉感知技术革新了传统目标识别方式，其通过自主特征学习机制突破了传统特征提取的局限性。当前深度学习检测模型根据架构差异主要分为两大类：单次阶段检测框架与双阶段检测框架。本研究系统梳理 2012 年至今的方法演进路径，如图 2-6 所示。重点解析其在行人检测领域的架构创新与技术迁移规律，探讨当前主流算法的实现机理与优化方向。

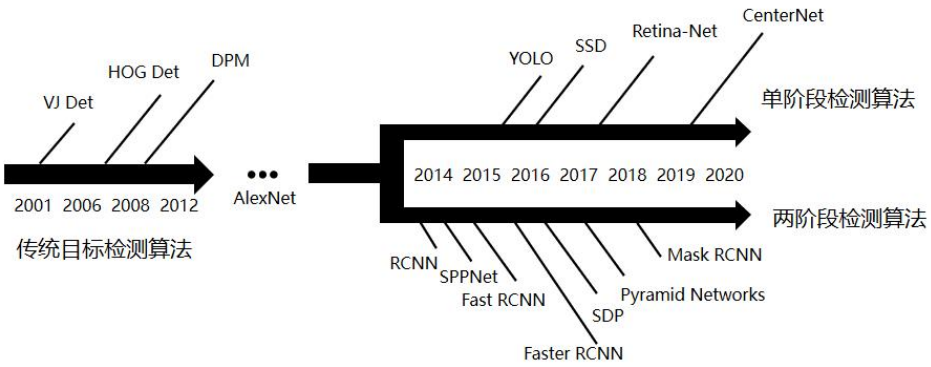


图 2-6 目标检测算法主要发展历程

2015 年，Girshick 等人[47]Faster R-CNN 作为两阶段检测里程碑式改进，其创新性地引入区域生成网络。相较于 Fast R-CNN 仍依赖选择性搜索（Selective

Search）进行候选框生成导致的效率约束，该模型通过端到端区域建议网络（Region Proposal Network, RPN）与主干网络的参数复用，实现了特征计算图的全局优化。如图 2-7 所示，其架构演进为四阶段特征处理流程：基于深度卷积网络的多尺度特征提取、RPN 层的锚框建议生成、ROI Pooling 层的区域特征对齐以及后处理阶段的非极大值抑制（NMS）优化。

图 2-7 Faster R-CNN 流程图

区域建议网络（RPN）作为 Faster R-CNN 的核心创新模块，通过全卷积滑动机制实现候选区域密集预测。如图 2-8 所示，该网络作用于骨干网络输出的特征张量，采用多尺度锚点模板与多纵横比配置（的参数化生成策略。每个空间坐标点对应 k 个几何先验框（k=9），通过平移不变性特性实现全图覆盖。其中预设的锚点模板作为基准参考系，通过可学习的缩放和平移参数动态调整建议框的几何形态。

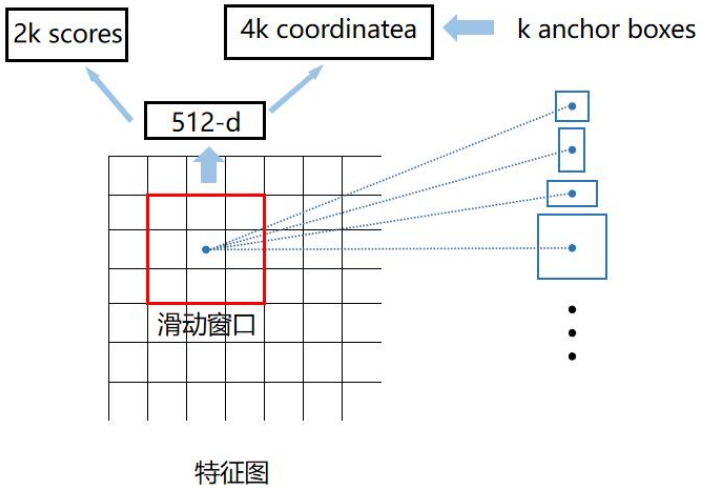


图 2-8 RPN 网络原理图

在车间生产场景下，行人检测算法要求兼顾检测精度和实时性。因此两阶段目标检测算法难以满足实际应用的要求。

作为深度学习目标检测领域的开创性框架，YOLO（You Only Look Once）通过单阶段检测突破了传统两阶段方法的性能瓶颈。该框架在确保目标识别准

确率的同时，显著提升了计算效率，为实时检测系统提供了新的解决方案。在迭代发展的 YOLO 系列算法中，YOLOv5 凭借其平衡的精度，速度优势成为工业界主流实施方案。本论文将基于该版本的技术实现，深入剖析其网络架构及创新特性，具体包括：采用 CSPDarknet53 的主干网络结构、融合 PANet 的特征金字塔机制，以及自适应锚框计算等核心模块。YOLOv5 的网络结构如图 2-9 所示。YOLOv5 的主干网络由 CSP、SPP 和 Focus 模块构成，主要用于提取输入图像的特征。作为 YOLOv5 网络架构的核心处理流程，特征提取阶段首先对预处理后的输入数据进行多级特征变换。

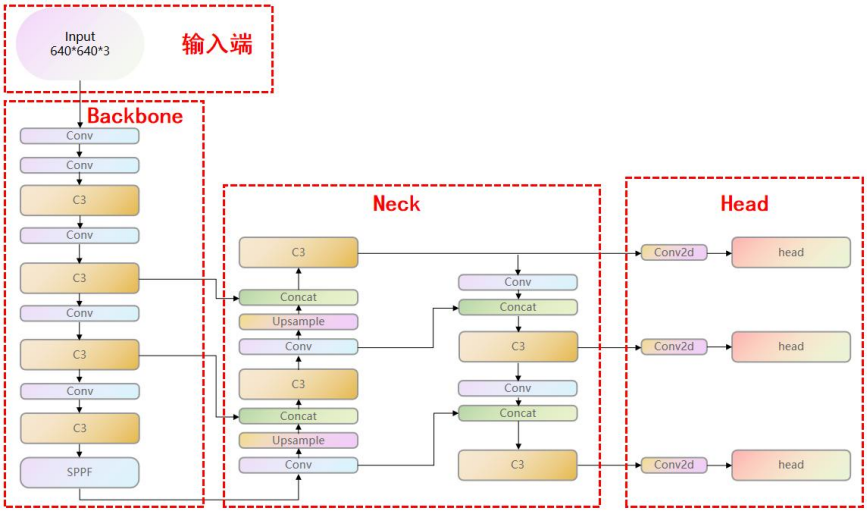


图 2-9 YOLOv5 网络结构图

2.5 行人轨迹预测算法

行人运动轨迹预测研究领域的重要突破源自斯坦福团队提出的递归神经网络架构。该研究构建了基于时序建模的递归推理框架，通过编码行人历史位移序列的时空相关性，实现高密度人群场景下的多智能体运动轨迹推断。其创新性体现在社交特征聚合机制的设计——在目标个体的 LSTM 状态更新过程中，动态构建局部交互图进行分布式学习。与传统方法相比，该方法实现了三个维度的突破：首先采用端到端学习策略替代人工特征工程，通过反向传播自动捕获复杂环境约束；其次引入邻域状态融合算子，使各节点的隐藏状态在欧氏空间内形成信息场，实现群体运动影响的概率建模；最后建立位移残差预测机制，将笛卡尔坐标系下的绝对位置预测转化为相对运动矢量估计。

Social-LSTM 模型的架构如图 2-10 所示。该框架为场景中的每个行人分配独立的 LSTM 单元，这些单元共享统一的参数配置。每个 LSTM 单元通过学习个体在时序上的状态变化来实现轨迹预测功能。特别地，模型引入了社交池化机制，该机制通过整合预测目标周围行人的隐藏状态信息，同时保持空间关系的完整性。这种参数共享的设计使得各个 LSTM 单元能够获取相邻单元的联合隐藏特征，从而实现群体行为的协同建模。

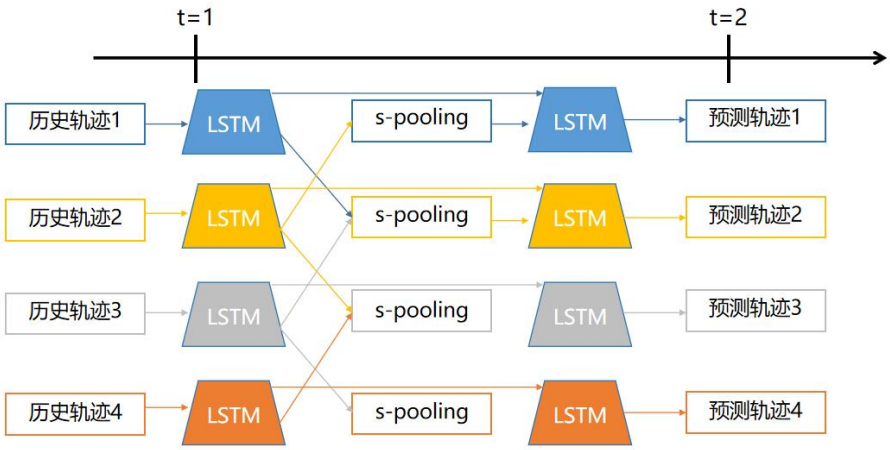


图 2-10 Social-LSTM 模型结构

Social-LSTM 便是通过处理不同时刻下的行人位置坐标来进行轨迹预测的。首先，将不同时刻下的坐标 $X_i^t\left(x_i^t, y_i^t\right)$进行式（2-1）的嵌入操作后，使得原始的位置坐标信息转化成了 LSTM 可以接收和处理的数据 e_i^t。然后，将池化向量 H_t^i 进行式（2-2）处理后得到嵌入向量 a_i^t。最后，将处理后得到的数据 e_i^t、a_i^t 和上一时刻的隐藏状态 h_i^{t-1} 输入到 LSTM 单元中得到当前时刻的隐藏状态，计算过程见式（2-3）。

$$e_i^t = \phi(X_i^t; W_e) \tag{2-1}$$

$$a_i^t = \phi(H_t^i; W_a) \tag{2-2}$$

$$h_i^t = LSTM\left(h_i^{t-1}, e_i^t, a_i^t; W\right) \tag{2-3}$$

其中，$\phi(\ )$是一个带有 ReLU 单元非线性的嵌入函数，W_e 和 W_a 为嵌入权重，LSTM 预测单元的权重表示为 W。

2.6 本章小结

本章系统性地概述了人工神经网络（ANN）、卷积神经网络（CNN）以及长短期记忆网络（LSTM）的基本原理。在此基础上，深入探讨了当前主流行人检测算法的技术框架与工作机制。根据方法论差异，将行人检测技术划分为传统特征提取与深度学习特征提取两大类别，并结合工业场景中叉车行人检测的具体需求，分析各类算法优缺点。此外，针对行人轨迹预测领域，本章从传统先验知识驱动与循环神经网络驱动两个维度展开对比研究，详细阐述了不同预测方法的技术特点及其局限性。

第 3 章　基于深度学习的目标检测算法研究

3.1 引言

随着工业自动化进程的发展，生产车间中的安全管理面临新的挑战。一方面，车间内部人员行动复杂，且人与移动设备（如无人叉车、自动引导车 AGV 等）的频繁交互，安全事故发生的风险大大增加。对行人进行精确检测和实时监控，是确保车间安全的关键。另一方面，在复杂的车间生产环境中，安全帽佩戴是确保人员安全的必要措施。然而，传统的人工监督方法难以实现全覆盖、全天候的监测，而且增加了管理成本，如图 3-1 所示。

图 3-1 人工安全监测

虽然当前智能安全监控系统中目标检测算法在行人检测与识别方面取得了一定成果，但因生产车间环境的复杂性，传统目标检测算法在应对复杂背景和遮挡等问题时，仍难以准确检测出行人并判断其安全帽佩戴情况。因此，需一种具有高精度、实时性和鲁棒性的新型检测算法，以解决以下问题：

（1）准确检测人员并识别安全帽佩戴状态：车间内行人移动频繁，加之设备复杂、背景干扰和遮挡现象普遍，传统检测方法难以在此类复杂环境下精确识别行人及其安全帽佩戴状态。

（2）实时响应与智能预警：生产车间中的安全事件具有突发性，要求安全

帽佩戴检测系统具备高度的实时性，以便在发现异常时及时触发预警并提醒相关人员。这不仅要求算法具备高检测精度，还要满足实时性的要求，实现对安全隐患的快速响应。

（3）适应车间环境的复杂性和多变性：生产车间的环境复杂多变，不同区域的光照强度、背景差异显著，人员佩戴的安全帽颜色、形状也各不相同。这要求检测系统具有良好的鲁棒性，能够在各种工况、背景和光照条件下保持稳定检测性能。

本研究旨在设计一种高精度、实时检测算法，以克服现有技术在复杂车间环境下行人检测和安全帽佩戴识别方面的局限性。该算法将有助于提升工业生产车间的安全管理水平，减少安全事故的发生，为构建更安全、高效的生产环境提供技术支持。

3.2 YOLO 系列算法

2015 年，Redmon 等人[49]提出了 YOLO（You Only Look Once）目标检测算法。该算法以整张图像为输入，通过卷积神经网络实现目标定位与分类，从而达到快速精准的检测效果。

YOLO 是一种单阶段的目标检测技术，将复杂的检测任务简化为直接回归问题，省去了候选框预测的步骤，从而实现了端到端实时检测。与传统的两阶段算法相比，YOLO 在检测效率上具有明显优势，能够很好地满足实时检测的应用需求。如图 3-2 所示，该图表对比了 YOLO 与其他算法在帧率（FPS）和平均精度均值（mAP）方面的性能表现。

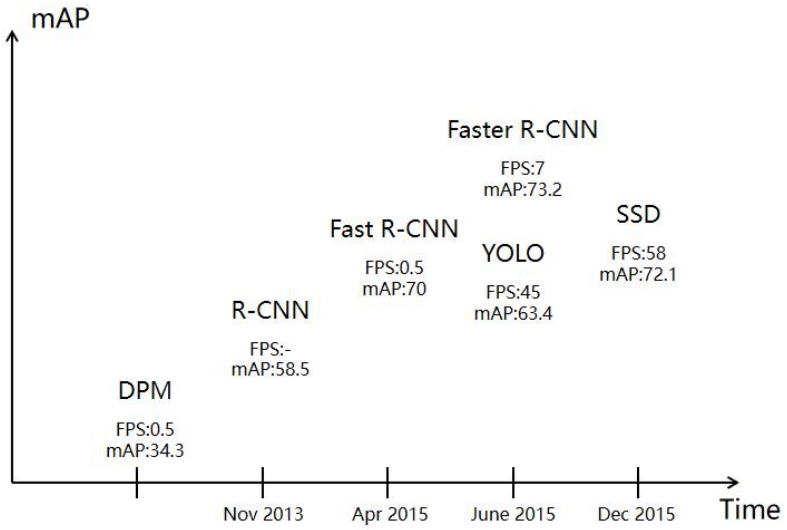


图 3-2 YOLO 与其他算法 FPS 和 mAP 的比较[50]

YOLO 系列算法核心在于直接预测目标边界框位置坐标和类别概率，从而实现对整幅图像的准确识别。虽然 YOLO 算法在特征提取方面表现出较高的运行速度和精准度，但其在目标定位上的误差相对较大，针对特殊场景下的目标检测任务仍需对其进行改进。

YOLOv1 通过将输入图像分割成 S×S 个网格单元来实现目标检测。当目标的中心位于某个网格单元内时，该单元负责对该目标进行预测。每个网格单元会生成 B 个边界框，每个边界框包含位置坐标以及表示目标存在的置信度信息。置信度的定义如下所述：

$$p = pr(object) \times \mathrm{IoU}_{pred}^{truth} \tag{3-1}$$

当目标未被边界框框选出时，其置信度将被打为零分。交并比（IoU）用于衡量真实边界框与预测边界框重叠程度，以此评估二者之间的相关性。

2016 年，YOLOv2 算法[51]被开发出来，通过优化实现了目标检测框架的突破性进展，其网络结构如图 3-3 所示。使模型在检测精度、运算效率及多尺度目标适应性三个维度均取得显著提升，特别针对前代模型存在的坐标回归误差问题，重点重构了定位机制。

2018 年，Redmon[52]在 YOLOv2 基础上踢出 YOLOv3 算法。其通过将网络架构从 Darknet-19 升级为 Darknet-53，并融入特征金字塔网络（FPN），实现了对多尺度目标的有效检测。并且 YOLOv3 采用逻辑回归取代 Softmax，从而在保持实时性的前提下，兼顾了检测精度。

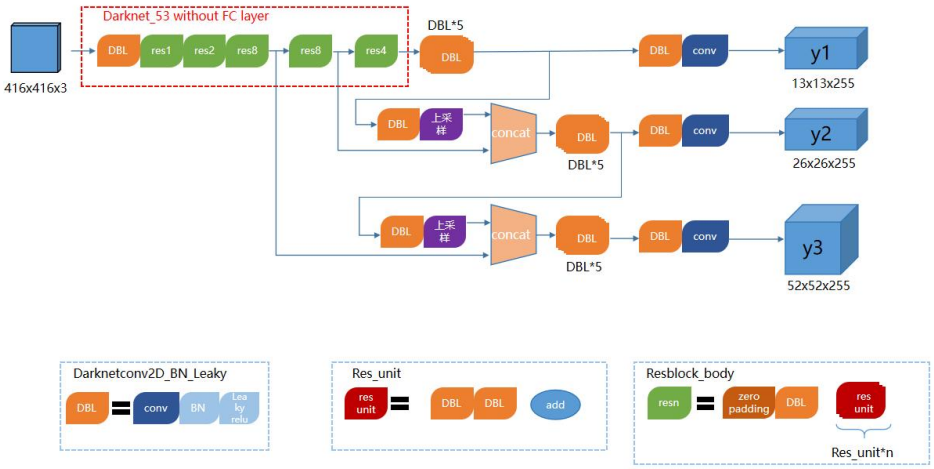


图 3-3 YOLOv3 网络结构图

YOLOv3 在网络架构层面实现了突破性改进，其核心技术提升体现在以下方面：采用残差结构构建深度网络体系，将特征提取主干升级为 Darknet-53 复合架构，如图 3-4 所示。

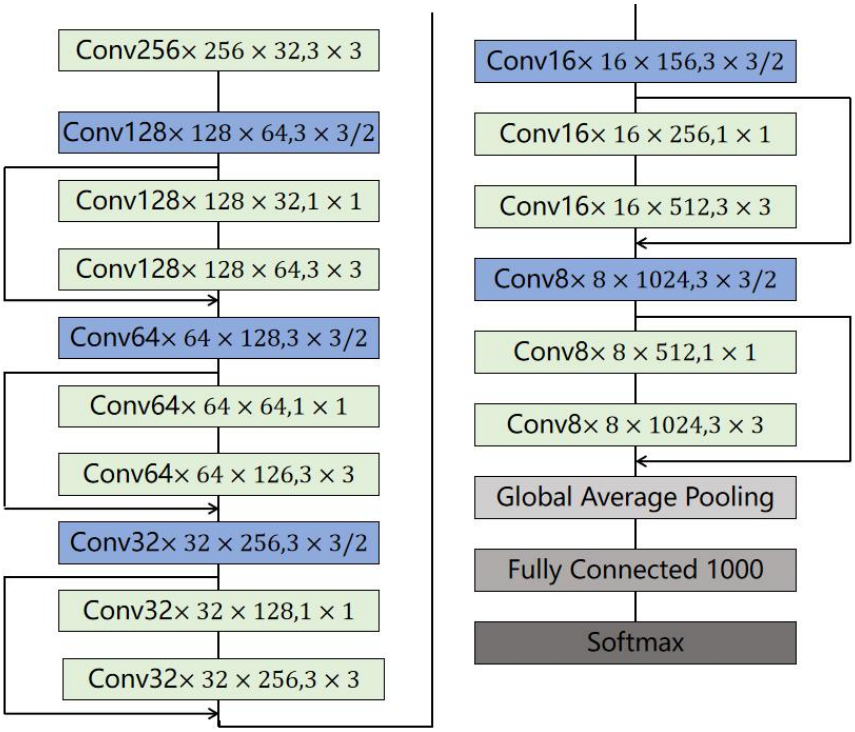


图 3-4 Darknet-53 特征提取网络结构

随着研究推进，后续又推出了一系列新的 YOLO 系列目标检测算法。本文选取 YOLOv5 作为目标检测算法，主要由以下几点原因：（1）高效性与实时性兼顾：YOLOv5 在设计上优化了计算效率，保持较高精度的同时实现快速推理，可满足车间目标监测的实时性；（2）对小目标检测的优化：通过自适应 Anchor

优化和 Focus 结构增强了对小目标的特征提取能力，适合本文监控视角下的目标检测任务；（3）模型规模适中：YOLOv5 模型在能够完成复杂检测任务的同时，避免过高的算力消耗，满足本文端到端的检测任务需求。因此，本文选用 YOLOv5 作为基础视觉检测模型。

3.3 基于 FasterNet 改进的 YOLOv5 检测算法

虽然 YOLOv5 在精度上更具优势，并且其模型结构相对简洁，更易满足实时性的需求。然而，本文提出的基于行人运动轨迹预测的人机碰撞预警方法，要求对行人图像进行两次网络计算。为确保叉车行人检测的实时性，需尽量减少每个阶段算法的计算时间。因此，本研究对 YOLOv5 的网络结构进行了优化和简化。同时，受轻量化网络结构 FasterNet[53]网络启发，借鉴了 FasterNet 卷积块的设计，替换了 YOLOv5 算法中的主干网络。降低模型参数、提高运算速度的同时，保证了检测精度不受影响。

3.3.1 FasterNet 网络

FasterNet 网络是一种通过减少 FLOPS(每秒浮点运算)而实现快速计算的轻量化神经网络。FasterNet 的基本组成单元是 Partial Convolution(PConv)，其结构如图 3-5 所示，

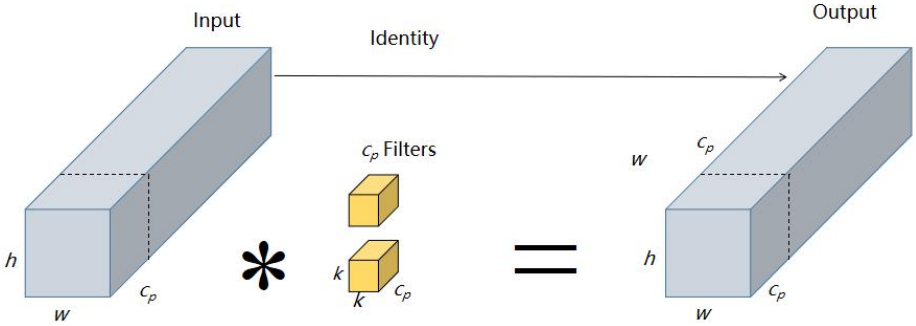


图 3-5 Partial Convolution 结构图

该方法仅对部分输入通道应用标准卷积（Conv）以提取空间特征，同时保持其他通道的数据不变，因此部分卷积（PConv）的浮点运算次数（FLOPs）仅为 $h\times w\times k^2\times c_p^2$。为了更全面、高效地利用所有通道的信息，本研究进一步将点态卷积（PWConv）融入部分卷积（PConv）之中。这种组合在输入特征图上的有效感受野呈现出类似 T 形的卷积特性，相较于传统的卷积运算，它更倾向于关注中心区域，如图 3-6(b)所示。

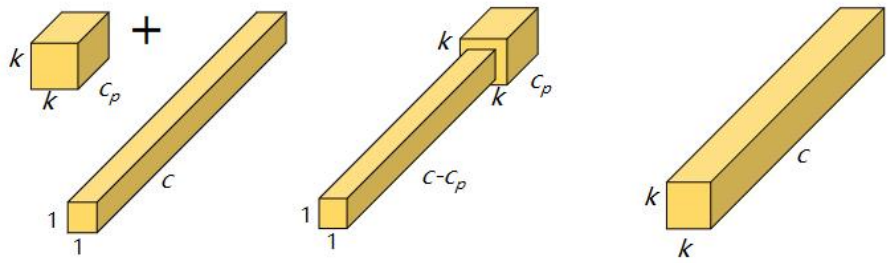


图 3-6 (a)PConv+PWConv(b)T-shaped Conv(c)regular Conv

FasterNet 网络的整体结构如图 3-7 所示。它包含四个分级阶段，每个阶段前均设置了一个嵌入层（采用 4×4 常规卷积，步幅为 4）或合并层（采用 2×2 常规卷积，步幅为 2），以实现空间下采样和通道数的增加。每个阶段由若干 FasterNet 块组成，其中最后两个阶段被分配了更多的计算资源以增强性能。每个 FasterNet 块包括一个部分卷积（PConv）层，随后连接两个点态卷积（PWConv，或称 1×1 卷积）层。这些层共同构成一个反向残差块结构，其中中间层的通道数经过扩展，并通过快捷连接重用输入特征。

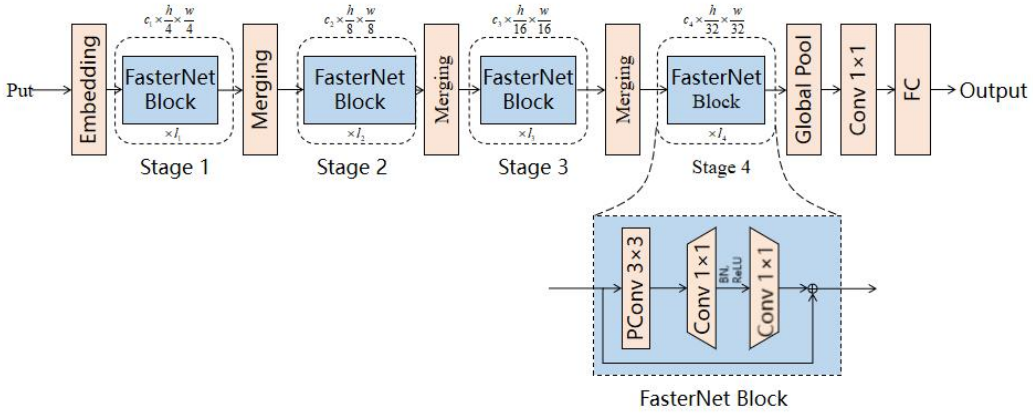


图 3-7 FasterNet 网络结构

FasterNet 采用四阶段层级架构，各阶段由若干 FasterNet 模块堆叠构建，起始部分配置嵌入层或特征整合层以实现特征重组。该网络的特征分类功能主要通过后续三个处理阶段完成。在模块设计层面，每个基础单元采用部分卷积层（PConv）与两个逐点卷积层（PWConv）顺序连接的拓扑结构。值得注意的是，本架构在保持特征表示多样性的同时，通过优化层间操作显著降低计算时延，仅在中间层级后引入归一化处理及非线性激活层，该精简设计有效平衡了模型性能与计算效率。

3.3.2 损失函数替换

目标检测任务中，场景多目标共存特性导致算法生成大量候选检测框。为此需采用非极大值抑制算法（Non-Maximum Suppression, NMS）进行后处理优化，该机制基于检测框置信度分数及交叠区域分析，通过保留置信度峰值框体并剔除其邻域内高重叠框体，实现检测结果的去冗余化处理，从而显著提升检测精度与运算效率[54]。原始 YOLOv5 模型采用 GIoU 损失函数，其原理参见公式（3-2）[55]。其计算过程包含三个关键参数：预测框与真实框的重叠区域面积 R_{IoU} 、包含两者的最小闭包区域 C 以及两者的联合区域 $A \cup B$ 。该函数通过构建空间约束优化策略，有效改善了传统 IoU 在无交叠情况下的梯度消失问题。然而当预测框与真实框完全重叠或其宽高比完全匹配时，闭包区域 C 与联合区域 $A \cup B$ 将呈现等值特性，此时 GIoU 退化为标准 IoU 度量，导致空间位置信息表达能力下降，进而引发模型收敛效率降低与检测时延增加等问题[56]。

$$\mathcal{R}_{GIoU} = \mathcal{R}_{IoU} - \frac{\left|C-(A \cup B)\right|}{\left|C\right|} \tag{3-2}$$

训练数据中普遍存在的低质量样本易受几何畸变（如目标距离偏差和长宽比例失调）干扰，此类噪声样本会导致模型表达能力退化[57]。针对该问题，本研究提出改进方案，设计融合动态距离感知机制的加权交并比（Weighted IoU, WIoU）损失函数，其具体表达式见公式（3-3）和公式（3-4）。该创新性方法通过构建可学习的空间权重矩阵，在传统 IoU 计算框架中嵌入几何敏感因子，有效克服了常规损失函数对异常尺度样本的敏感性。

$$\mathcal{L}_{WIoUv1} = \mathcal{R}_{WIoU}\mathcal{L}_{IoU} \tag{3-3}$$

$$\mathcal{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{3-4}$$

在此公式中，x 和 y 表示预测框中心点的位置坐标，而 x_{gt} 和 y_{gt} 则对应真实框中心点的坐标值。W_g 和 H_g 分别定义为能够完全包含预测框与真实框的最小矩形框的宽度和高度。其中，$\mathcal{R}_{WIoU} \in [1, e)$ 的功能是将普通质量的先验框进行放大，而 $\mathcal{L}_{IoU} \in [0,1]$ 则起到缩小高质量先验框的作用。当先验框与目标框完全重叠时，算法会特别关注两者中心点之间的距离，这一过程如图 3-8 所示。

为避免 $\mathcal{L}_{IoU} \in [0,1]$ 发生梯度消失的现象，此处对 W_g 和 H_g 的处理方式进行了调整，具体是将它们从梯度计算中剥离出来。即，通过将 W_g 和 H_g 的值设定为固定常量，使其在反向传播过程中不参与梯度更新，从而防止因数值过大而减缓模型的收敛速度（公式中用*符号标记此操作）。

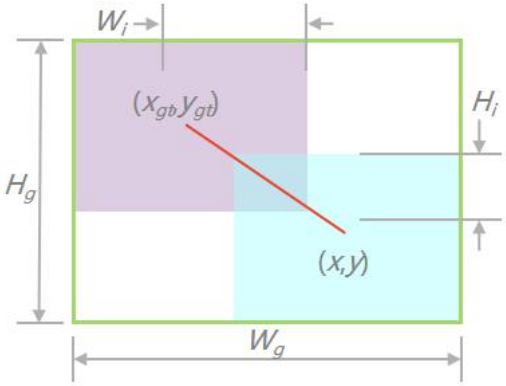


图 3-8 最小封闭框及其中心距

3.4 数据集制作

本文利用车间监控相机视频流，将视频转化为单帧图像制作数据集，数据集分为行人、佩戴安全帽、叉车。为确保模型泛化能力，对标注后的数据通过添加噪声、改变亮度、剪裁旋转等方法进行数据增强。最终形成 5000 张图片的数据集，并按照 9:1 的比例将数据集划分为训练集和验证集，如图 3-9 所示。

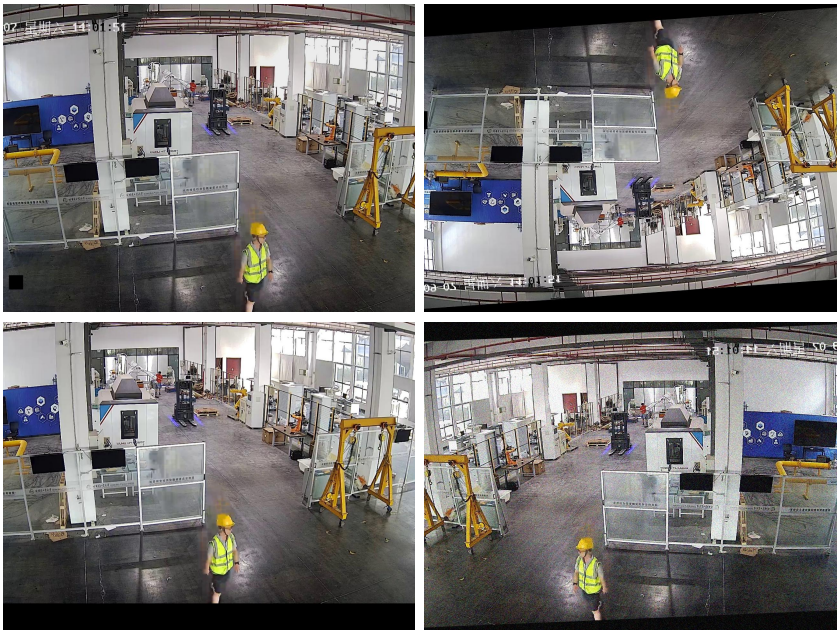


图 3-9 人机共存的车间现场数据集示例

3.3.3 引入注意力机制

车间监控图像中存在着一定数量的小目标，占整幅图像比例较小，易受环境等因素的影响。原版的 YOLOv5s 网络在进行深层次卷积时常常会丢失小目标的安全帽的信息。为了进一步提升图像中位置信息与安全帽特征之间的关联性，同时增强网络对关键信息的特征表达能力，本文采用了 CBAM[58]注意力机制。该机制的网络架构如图 3-10 所示。图中所示的注意力机制架构中，CBAM（Convolutional Block Attention Module）采用通道-空间双域协同的设计方法，实现了特征图谱的多维优化。该模块通过级联式特征校准机制，首先对特征张量实施通道域权值重标定，继而完成空间维度的聚焦增强。具体而言，通道注意力子模块通过建立跨通道关联性度量矩阵，对特征通道进行动态重要性分配；而空间注意力子模块则通过构建二维空间权重映射，强化目标区域的响应强度。这种双阶段注意力递进机制，通过建立通道敏感性与空间显著性的联合优化模型，有效提升了特征表达的判别能力。

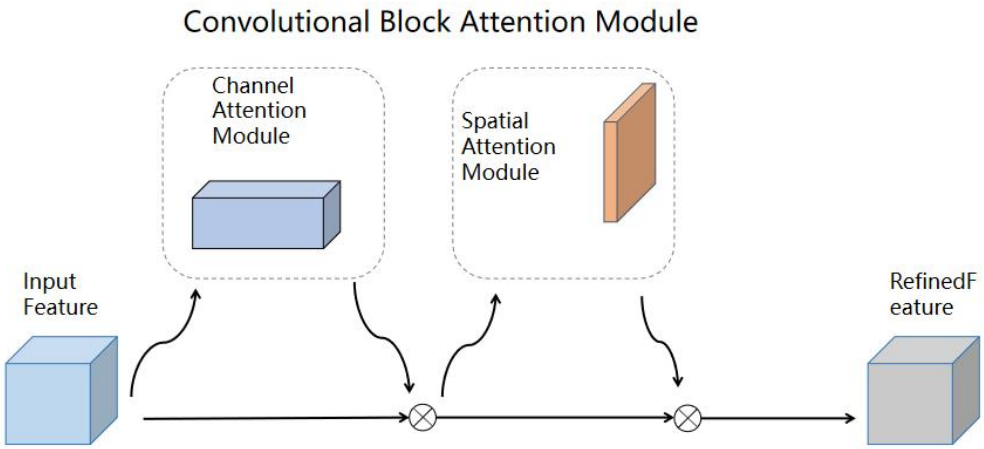


图 3-10 CBAM(Convolutional Block Attention Module)模块结构

通道注意力模块结构如图 3-11 所示，其通过双路空间特征压缩机制实现信息聚合：全局平均池化（Average Pooling）与全局最大池化（Max Pooling）分别提取空间统计特征，生成特征向量 F_{avg}^{C} 和 F_{max}^{C} 。这两个特征描述符随后被馈入权值共享的多层感知器（MLP）网络，该网络采用通道维度缩减率 r（隐藏层维度设定为 $R^{C/r\times 1\times 1}$ ）的参数优化策略以控制计算开销，最终通过非线性映射合成通道注意力权重矩阵 $M_c \in R^{C\times 1\times 1}$,如公式(3-5)所示。其中， $W_0 \in R^{C\times C/r}$ 和 $W_1 \in R^{C\times C/r}$ 表

示全连接层的权重参数，W_0 与 W_1 实施跨路径权值共享机制，特征变换过程经 ReLU 非线性激活处理后，由 Sigmoid 函数完成概率空间规范化。

$$\begin{aligned} M_C(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0 F_{avg}^c)) + W_1(W_0(F_{max}^c))) \end{aligned} \tag{3-5}$$

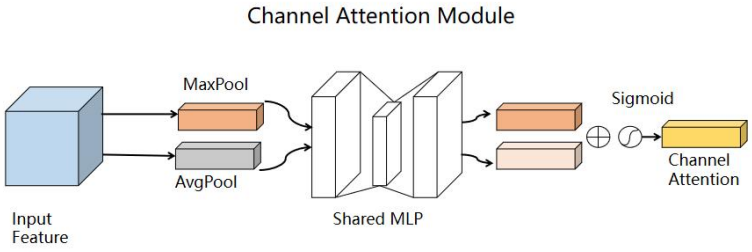


图 3-11 通道注意力模块结构

空间注意力模块与通道注意力模块的主要区别在于，其核心目标是识别输入图像中哪些区域的信息更具重要性，其结构设计如图 3-12 所示。该模块首先利用两种池化操作对特征图的通道信息进行聚合，从而生成两个特征映射：$F_{avg}^s \in \boldsymbol{R}^{1\times H\times W}$ 和 $F_{max}^s \in \boldsymbol{R}^{1\times H\times W}$。其中，$F_{avg}^s \in \boldsymbol{R}^{1\times H\times W}$ 表示跨通道的平均池化特征，而 $F_{max}^s \in \boldsymbol{R}^{1\times H\times W}$ 则代表跨通道的最大池化特征。接下来，这两个特征映射被拼接在一起，并通过一个标准卷积层进行处理，生成一个二维的空间注意力图。随后，通过 Sigmoid 函数对该注意力图进行归一化处理，得到最终的空间注意力分布。整个过程的详细计算公式可参见公式（3-6），式中 σ 表示 Sigmoid 函数，$f^{7\times7}$ 表示 7×7 大小的卷积核。

$$\begin{aligned} M_S(F) &= \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)]) \\ &= \sigma(f^{7\times7}([F_{avg}^s; F_{max}^s])) \end{aligned} \tag{3-6}$$

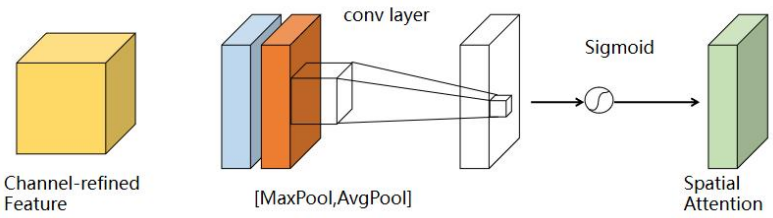


图 3-12 空间注意力模块结构

3.3.4 改进后的 YOLOv5

面向工业车间检测场景的工程化需求，本研究提出基于 YOLOv5s 架构的多维度优化方案，在保持基础检测精度的前提下，实现算法轻量化与部署适应性提升，满足嵌入式平台资源约束及工业现场实时检测需求。改进方案包含三个核心模块：（1）主干网络轻量化重构：针对原始模型存在的计算冗余问题，移植 FasterNet 作为特征提取主干，通过结构重参数化技术压缩网络深度，降低运行时内存占用；（2）动态边界回归优化策略：采用融合注意力引导的 Wise-IoU 损失函数，该函数通过构建动态评价机制实现异常样本抑制，其分层权重分配策略显著提升边界框回归的鲁棒性；（3）级联式注意力架构：在特征金字塔与检测头连接处嵌入 CBAM 双路注意力模块，形成空间-通道联合增强机制，强化网络对工件细微特征的捕获能力。改进后的网络架构如图 3-13 所示，其通过算法精度、计算效率与工程适用性的协同优化，达成工业场景下的最优平衡。

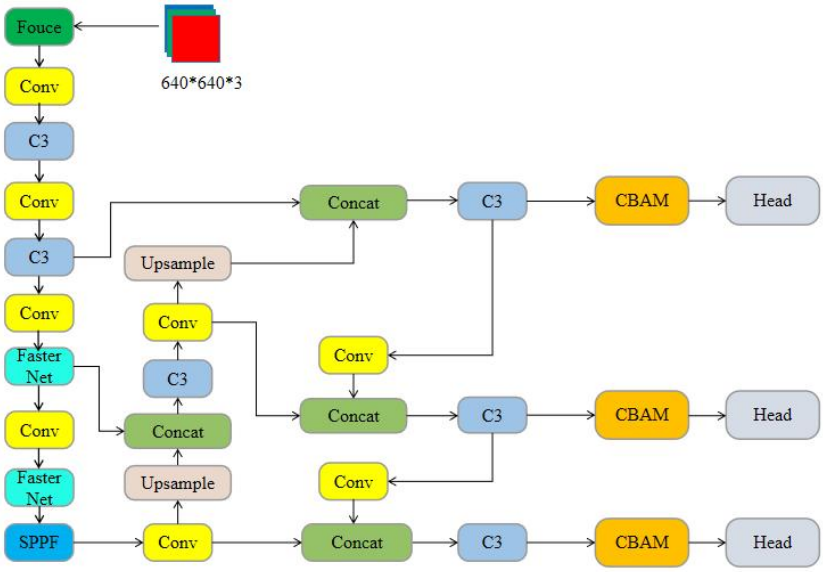


图 3-13 改进后的 YOLOv5s 模型

3.5 实验软硬件环境与评估指标

3.5.1 实验软硬件环境

训练平台和硬件采用 Windows 10 系统、NVIDIA GeForce RTX3060 图像处理器。在软件部署方面采用 Pytorch1.7 为深度学习框架，利用 CUDA 计算平台和 GPU 加速库 cuDNN 来降低内存占用并大幅提高 GPU 计算性能。详细参数设

置如表 3-1 所示。

表 3-1 训练参数设置

参数名称	参数设置
输入图像大小	640×640
训练轮数	500
步长大小	32
优化器	SGD
动量	0.937
初始学习率	0.01
学习率因子	0.01

3.5.2 评价指标

为验证算法改进的有效性，本文利用检测精度（Precision）、召回率（Recall）、均值平均精度 *mAP*（Mean Average Precision）作为标准衡量算法性能[59]，$mAP^{0.5:0.95}$ 表示当阈值为0.5到0.95之间时总体平均 *mAF*值，计算公式如下所示。

$$Precision=\frac{TP}{TP+FP} \tag{3-7}$$

$$Recall=\frac{TP}{TP+FN} \tag{3-8}$$

$$mAP=\frac{1}{C}\sum_{i=1}^{c}AP_i \tag{3-9}$$

上述公式中 *TP* 代表正样本被正确识别出来；*FN* 代表正样本被错误识别为负样本；*FP* 代表负样本被错误识别为正样本；*AP* 代表单个类别上的平均精度；*C* 代表类别个数。

3.6 目标检测实验结果

为验证本文改进算法的有效性，并使用目前比较流行的几种目标检测网络进行对比试验：Faster RCNN、SSD、YOLOv4 Tiny、YOLOv7。实验过程遵循控制变量原则，实验软硬件环境保持不变。评价指标采用 mAP、参数量、GFLOPS、Precision、推理时间。实验结果如表 2 所示。

由表 3-2 可见，模型在构建的数据集上，均值平均精度比 Faster RCNN、SSD、YOLOv4 Tiny、YOLOv7、YOLOv8 算法分别提高了 18.5%、34.4%、28.6%、2.1%、0.5%，具有更优秀的检测性能。模型计算量大小排序为

YOLOv7<Ours<YOLOv4 Tiny<YOLOv8<SSD<Faster RCNN,推理时间的排序为Ours<YOLOv7<YOLOv8<YOLOv4 Tiny<SSD<Faster RCNN,实验证明本文算法符合实时检测的需求。综合来看改进算法相较于最新的 YOLOv8 有着更低的参数量和计算量，在推理时间减少 22ms 的前提下，均值平均精度和精确度仍提高 0.5%、0.7%，与传统的 Faster RCNN、SSD 模型相比，改进算法不仅在参数量、计算量、推理时间等方面均大大降低，而且在均值平均精度和精确度等方面具有较大提升。与 YOLOv7 相比，尽管改进后的算法模型参数规模有所增加，但由于计算复杂度显著降低，其检测延迟依然缩短了 3 毫秒。同时，均值平均精度（mAP）和精确度分别提高了 2.1%和 3.1%。相较于轻量级网络 YOLOv4 Tiny，该算法在 mAP 和精确度等关键指标上均实现了显著增强。综合上述分析，本研究所提出的 YOLOv5 改进模型展现出较高的 mAP，足以满足实际应用中的精度需求；与此同时，其计算复杂度和模型参数规模均维持在较低水平，便于在边缘设备上高效部署。

表 3-2 对比实验结果

Model	mAP@.5(%)	Parameters(M)	GFLOPS	Precision(%)	Inference Time(ms)
Faster RCNN	73.8	82.90	216.2	59.4	126
SSD	58.0	26.73	23.6	85.9	89
YOLOv4 Tiny	63.7	7.23	9.4	74.1	45
YOLOv7	90.2	3.67	103.2	87.8	21
Ours	92.3	6.13	14.1	90.9	5

为进一步验证本文针对车间环境下行人状态检测及安全帽佩戴监测算法改进的有效性和泛化性能，本文在车间采集两组不同行人位姿，并通过调整图像灰度模拟暗光环境下应用场景，如图 3-14 所示。

图 3-14 不同光照条件下两组行人位姿

粉色检测框表示未佩戴安全帽，红色检测框表示佩戴安全帽。YOLOv5 检测结果如图 3-15 所示，当车间内行人出现遮挡时，会出现漏检情况。本文所提出的方法如图 3-16 所示，可以看出相较于传统 YOLOv5 算法利用本文提出的改进算法针对不同光照环境中在车间行人，部分躯体被遮挡的情况下，可以精准做出检测，同时证明模型具有良好的泛化性能。

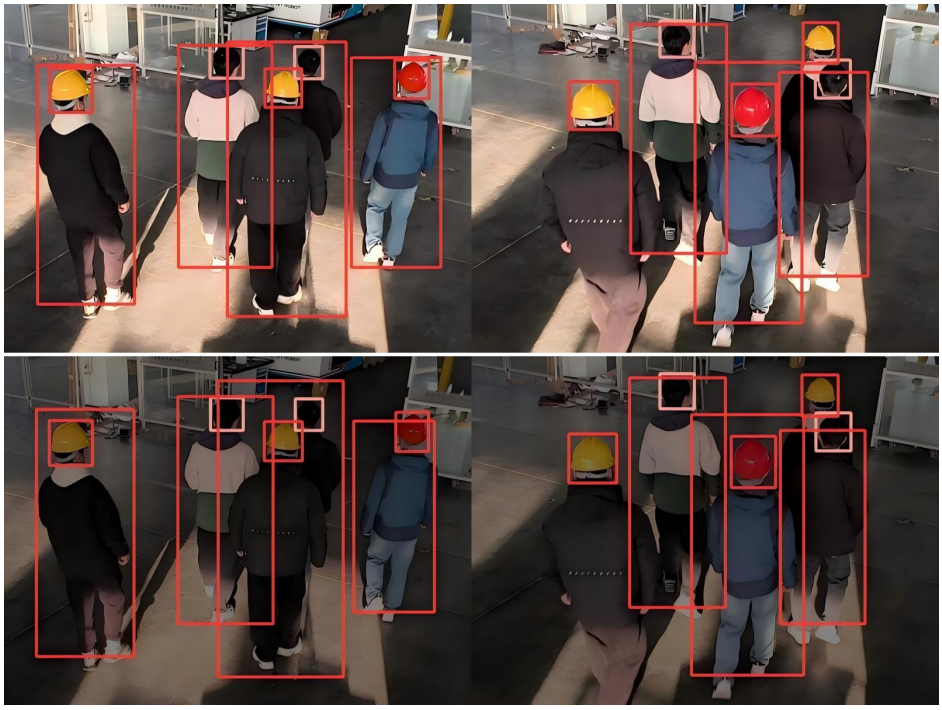


图 3-15 YOLOv5 检测结果

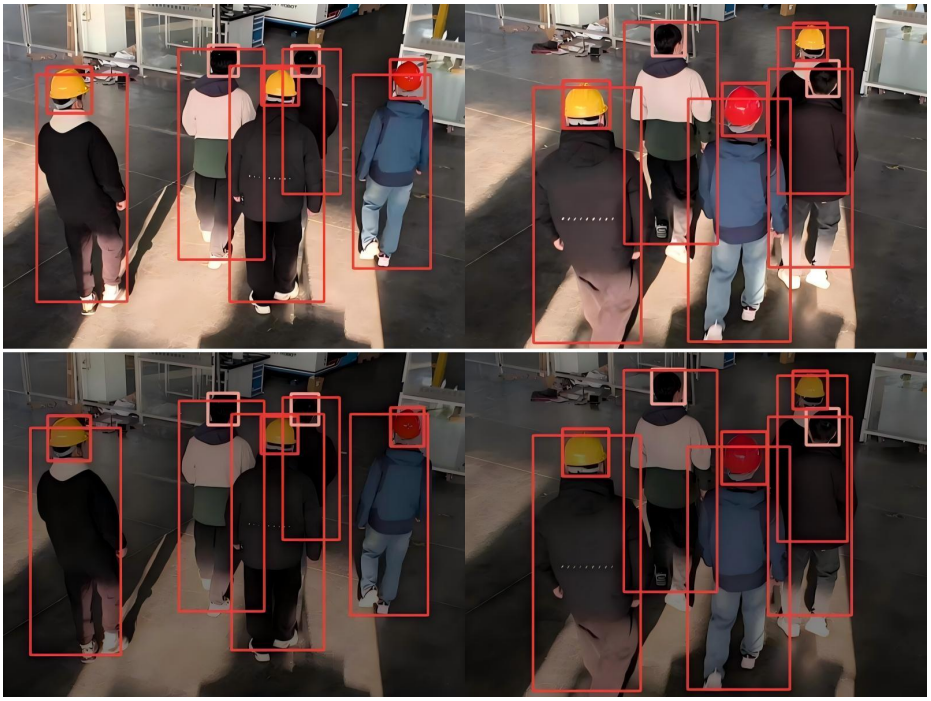


图 3-16 本文方法检测结果

3.7 本章小结

针对叉车操作场景中对行人检测及安全帽佩戴情况监测的需求，本章提出了一种基于 FasterNet 优化的 YOLOv5 算法改进方案。该方案通过将 FasterNet 集成到 YOLOv5 的主干网络中，显著减少了模型的参数数量和计算负荷，将原损失函数替换为具有动态聚焦机制的 WIoU 损失函数，并添加 CBAM 注意力机制。从而有效降低了算法的运算成本，并提升了检测的运行速度。考虑到叉车作业环境中光照条件的不一致性，本文通过图像增强技术对输入图像进行了预处理，以丰富数据集的多样性和复杂度，从而增强模型的适应能力。为了评估所提算法的实际应用性能，利用某车间内部监控拍摄画面，完成数据集的采集与加工，共生成了 5000 张基于监控视角的图像数据。实验结果显示，通过真实环境测试，基于 FasterNet 改进的 YOLOv5 算法在实时处理能力和检测精确度上均取得了显著提升，充分满足了叉车场景下行人检测的实际要求。

第 4 章 基于三维距离恢复的人机碰撞预警研究

4.1 引言

目前的视觉测距方法主要分为单目测距、双目立体视觉测距、结构光测距等。考虑到车间实际环境中大多使用单目监控相机，因此本文基于单目三维距离恢复利用单目相机提出一种车间场景下人机间距离检测方法。在目标检测的基础上，提取目标框中心点图像坐标。由于车间中无人叉车尺寸确定，利用行人平均身高可以计算出叉车和监控以及行人和监控的距离。之后在已知相机视场角的情况下通过计算视场中两物体夹角和距离信息通过余弦定理即可计算出人机距离，具体计算流程见图 4-1 所示。

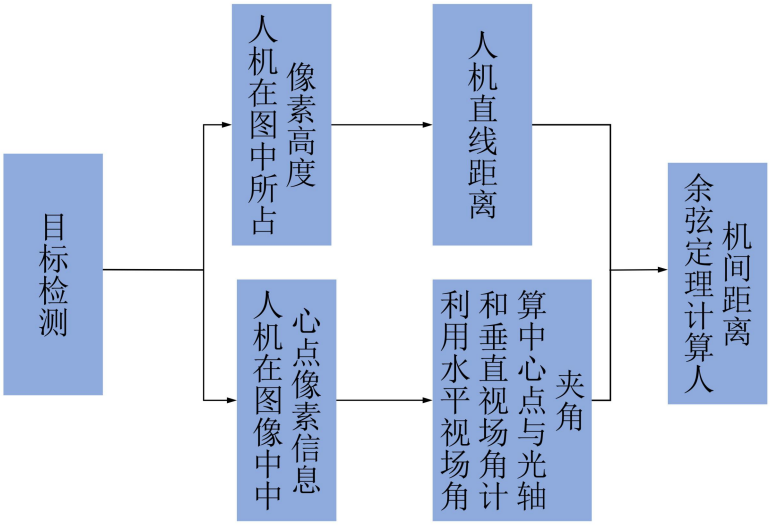


图 4-1 基于单目三维距离恢复人机距离检测原理示意图

4.2 单目视觉定位原理

4.2.1 参考坐标系设定及坐标转换

单目视觉定位技术的关键在于构建图像坐标系、相机坐标系与世界坐标系间的几何映射模型。在计算机视觉应用中，目标物体的空间定位需通过多层次坐标转换实现：首先将像素坐标系中的二维观测数据转换至相机坐标系的三维特征，进而通过刚体变换建立其与世界坐标系的关联关系。以典型应用场景为例，行人检测算法输出的原始位置信息通常基于像素坐标系，而要解算目标在

三维物理空间中。的精确坐标，需要通过多步骤的坐标转换实现，该过程涉及多个空间基准系的协同映射

（1）像素坐标系

像素坐标系定义为以图像左上顶点为原点的二维笛卡尔坐标系，其坐标定义基于图像宽度（w）与高度（h）的正交轴向，具体几何关系如图 4-2 所示。在此模型下，数字图像可被解构为 h×w 维的离散采样阵列，每个像素单元的位置参数由二维离散网格结构唯一确定：横向坐标u_o表示像素所在列的索引值，纵向坐标v_o对应其所在行的序列号，二者共同构成像素点在阵列中的定位基准。

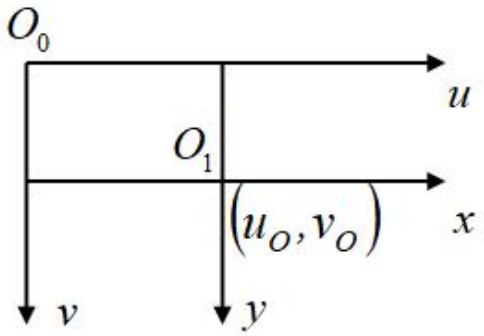


图 4-2 像素坐标系

（2）图像坐标系

像素坐标系向物理空间的映射需通过建立具有物理量纲的规范化图像坐标系实现，该过程包含两个关键步骤：首先对像素坐标进行空间离散性消除的归一化处理，其次通过式（4-1）所示的线性变换建立坐标基准O_{lxy}。在此转换模型中，参数 dx 表示传感器横向离散间距参数（单位：毫米/像素），dy 则定义纵向离散间距参数，二者共同构成图像平面内像素阵列的物理尺度基准。

$$\begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4-1}$$

（3）相机坐标系

相机光心处设定成相机坐标系原点，其X_C轴和Y_C轴分别与图像坐标系的 x 轴和 y 轴保持平行，而Z_C轴则与相机的光轴一致，如图 4-3 所示。基于图 4-3 所展示的成像原理，通过式（4-2）可以导出坐标转换的数学表达式，其中 f 表

示相机的焦距。

$$Z_c\begin{bmatrix}x\\y\\1\end{bmatrix}=\begin{bmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}X_C\\Y_C\\Z_C\\1\end{bmatrix} \tag{4-2}$$

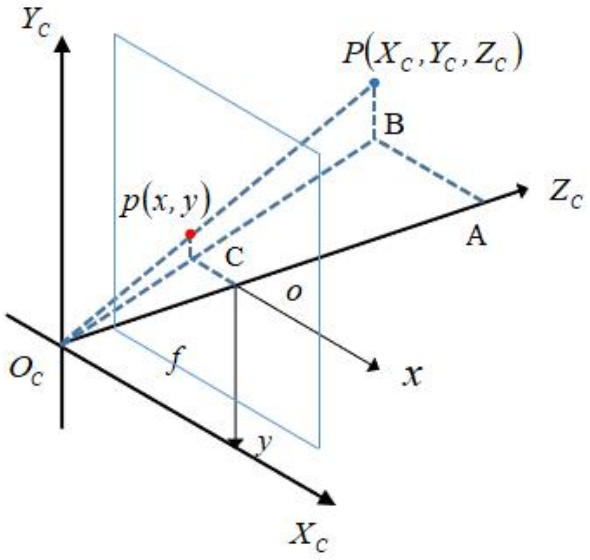


图 4-3 坐标系转换示意图

（4）世界坐标系

为建立相机观测数据与真实物理空间的几何关联，需引入具有绝对物理基准的世界坐标系。在工程实践中，该坐标系的基准框架可根据应用场景自由选定正交基底。相机坐标系至世界坐标系的坐标映射可通过式（4-3）所示的刚体变换模型实现，其数学表达式包含两个核心参数：平移向量T描述两坐标系原点间的欧氏空间位移量，旋转矩阵R则表示坐标系轴向的旋转变换关系，二者共同构成摄像机在世界坐标系中的六自由度位姿参数。

$$\begin{bmatrix}X_C\\Y_C\\Z_C\end{bmatrix}=R\begin{bmatrix}X_W\\Y_W\\Z_W\end{bmatrix}+T \tag{4-3}$$

根据坐标系转换关系即可推导出，若图像里存在像素坐标(x,y)即可转换为其在世界坐标系下的坐标(X_W,Y_W,Z_W)，其计算如式（4-4）所示。

$$Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x&0&u_0&0\\0&f_x&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&T\\0^T&1\end{bmatrix}\begin{bmatrix}X_W\\Y_W\\Z_W\\1\end{bmatrix} \tag{4-4}$$

式中$\begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_x & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$为相机的内参，$\begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix}$为相机的外参。

4.2.2 成像过程中的坐标系转换

深度信息重构需依托相机坐标系与物体坐标系间的空间几何关联模型。该过程基于投影几何理论构建：首先通过齐次变换矩阵将三维物点映射至相机坐标系，继而通过刚体变换参数实现与物体坐标系的逆向映射。为实现精确深度解算，通常采用基于已知几何特征的参照物（如标定板）建立共线方程，通过对齐次坐标系的参数解算完成空间位姿参数辨识，最终获取目标物体在欧氏空间中的三维坐标信息。

（1）图像坐标系和相平面坐标系的转换

坐标系映射参数的精确解算需综合考量摄像机内参（焦距、主点等）、外参（位置姿态）及传感器几何成像特性。在数字摄影测量学框架下，主点坐标（图像平面中心基准）作为关键参考基准，其作用在于建立像素分辨率与物理成像平面之间的尺度归一化关系，该映射关系可通过公式（4-5）所示的仿射变换模型进行数学表示。

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \tag{4-5}$$

基于成像几何参数体系，空间尺度参数 d_x 与 d_y 分别表示图像传感器在水平与垂直轴向的物理分辨率参数 $lpixel = dxmm$。该模型构建需满足两个基本约束条件：其一，需严格遵循针孔相机模型的理想透视投影假设；其二，要求成像系统满足无畸变光学条件。通过参数代换与齐次坐标变换的代数运算，可推导出如式（4-6）所示的参数化投影方程，其完整显示了像素坐标到物理空间的确定性映射关系。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4-6}$$

（2）图像坐标系与相机坐标系的转换

图像平面坐标系与摄像机投影中心的三维空间几何投影关系见图 4-4 所示。

为简化代码计算，以 $O_C - X_C Y_C$ 平面作为成像平面的对称处理基准。在平面中 $\Delta MNO_C \sim \Delta oCQ$，$\Delta PNQ_C \sim \Delta pCO_C$，因此得到几何关系表达式如公式（4-7）

$$\frac{MN}{oC} = \frac{MO_C}{oO_C} = \frac{PN}{pC} = \frac{x_C}{x} = \frac{x_C}{y} = \frac{z_C}{f} \tag{4-7}$$

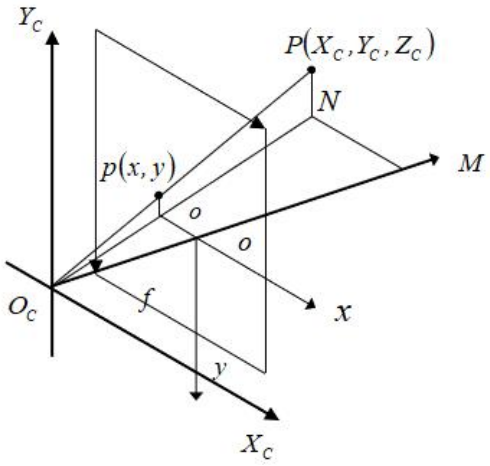


图 4-4 相机坐标系与图像坐标系

根据以上公式，并结合图 4-4 中的几何关系，可以得出应对的几何关系表达式如下，即公式（4-8）所示：

$$Z_C \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \tag{4-8}$$

（3）相机坐标系与世界坐标系的转换

坐标系转换时，将坐标值代入公式即可得到两个坐标系之间的转换关系。示意图如图 4-5 所示。

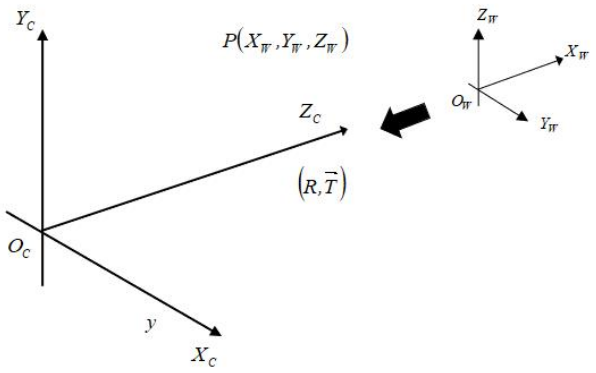


图 4-5 世界坐标系和相机坐标系

通过带入旋转矩阵$(\boldsymbol{R})$和平移矩阵$(\boldsymbol{T})$，可以得到如下表达式，即公式（4-9）所示：

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + T \tag{4-9}$$

将上述公式（4-9）进行方程整理，可以得到如下方程式：

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{4-10}$$

在公式（4-10）中，$\boldsymbol{R}$是一个3×3的矩阵，$\boldsymbol{T}$是一个3×1的矩阵。为进一步简化上述公式，本文将坐标$O_C - X_C Y_C$变换为$o - xy$，并重新进行转换分析。旋转示意图如图4-6所示。

在三维空间变换中，当参考坐标系经历旋转变换时，原坐标系下点的坐标参数将在新的正交基底体系中产生规律性改变。这种空间坐标转换关系可通过线性代数中的正交变换矩阵进行数学描述，具体表示形式如式（4-11）所示：

$$\begin{cases} x = x'\cos\theta - y'\sin\theta \\ y = y'\sin\theta - x'\cos\theta \\ z = z' \end{cases} \tag{4-11}$$

整理公式（4-11），得到其矩阵方程如（4-12）：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_1 \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4-12}$$

同理得到：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\varphi & sin\varphi \\ 0 & -sin\varphi & cos\varphi \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_2 \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4-13}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} cos\omega & 0 & -sin\omega \\ 0 & 1 & 0 \\ sin\omega & 0 & cos\omega \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_3 \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4-14}$$

因而，整合以上公式处理得到旋转矩阵方程：$\boldsymbol{R} = \boldsymbol{R}_1\boldsymbol{R}_2\boldsymbol{R}_3$。

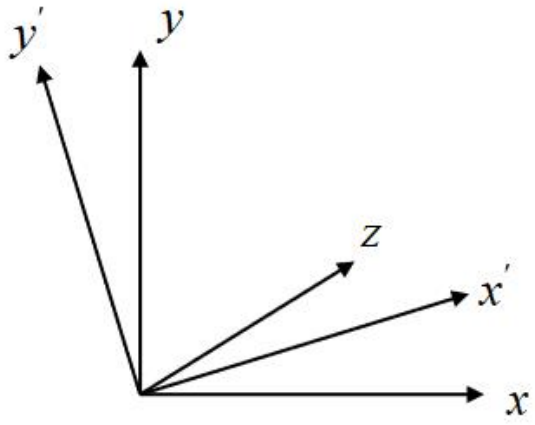


图 4-6 绕 Z 轴旋转示意图

4.2.3 测距模型

在空间位姿传递过程中，复合旋转矩阵的乘法运算可实现多坐标系间的参数映射。基于相邻坐标系间的线性变换关系，运用公式（4-15）对空间位姿参数进行矩阵运算，可推导出目标坐标系下的参数特征。

$$\begin{aligned} Z_C\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \end{aligned} \tag{4-15}$$

在该公式中，相机的内部特性，例如焦距等参数，被封装在相机内参矩阵 $\begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ 内，而相机的外部属性，如位置及朝向，则由相机外参矩阵 $\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \vec{0} & 1 \end{bmatrix}$ 反映。因而，为了利用上述公式实现坐标系之间的转换，就必须掌握相机的内、外参数矩阵的具体取值。

在光学系统建模过程中，所构建的坐标变换方法基于理想化假设体系，未纳入镜头畸变、传感器误差及机械装配偏差等非理想因素。然而实际工程场景中普遍存在的相机光轴偏移与安装位姿扰动，使得真实成像系统的几何非线性特征显著增强，导致其数学模型复杂度远超理想化针孔成像原理的理论承载范围。

针对透镜径向畸变校正问题，可基于泰勒多项式展开理论，围绕原点（r=0）对畸变函数进行低阶近似展开。该方法可建立理想坐标(X_{rcoor},Y_{rcoor})与畸变坐标(x_p,y_p)之间的映射关系，其数学表达如式（4-16）所示。具体而言，通过将畸变函数在光学中心邻域内展开为泰勒级数，并截取前若干项构建近似模型，能够有效实现非线性畸变的参数化表示与坐标变换计算。

$$\begin{cases} X_{\text{rcoor}} = x_p(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ Y_{\text{rcoor}} = y_p(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \tag{4-16}$$

由式（4-16）得，如果三个参数$k_1>0$、$k_2>0$、$k_3>0$，会导致枕型畸变；但是如果都小于 0，则为桶型畸变，如图 4-7 所示。

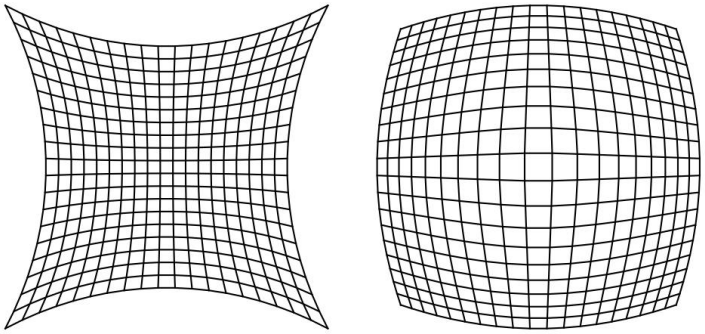


图 4-7 畸变示意图

针对光学系统中因机械装配偏差引发的非对称畸变问题，其本质可归结为切向畸变效应。该畸变类型主要源于透镜组件与成像平面间的非理想位姿关系，具体表现为投影坐标系下的几何形变特征，其参数化解析模型如式（4-17）所示，该方程精确反映了物像坐标在切向维度上的非线性映射关系。

$$\begin{cases} X_{\text{tcoor}} = x_p + \left[2p_1 x_p y_p + p_x\left(r^2 + 2x_p^2\right)\right] \\ Y_{\text{tcoor}} = y_p + \left[2p_1 x_p y_p + p_x\left(r^2 + 2x_p^2\right)\right] \end{cases} \tag{4-17}$$

其中，(X_{tcoor},Y_{tcoor})表示矫正后坐标，(x_p,y_p)为矫正前坐标。通过 p1 与 p2 来校正畸变。相机畸变模型包含(k_1,k_2,k_3,p_1,p_2)五个参数，这些参数用于刻画相机成像时的畸变特征。为了校正这些内在的畸变，需要将相机进行标定。

4.3 单目相机标定及测距

4.3.1 张正友标定方法简介

针对相机内参矩阵的精确解算需求，需通过严格的标定流程实现参数估计。在现有标定方法中，张氏平面标定法因其鲁棒性与工程实用性成为计算机视觉领域的主流方法。相较于需三维标定物的传统方法，该技术通过引入基于二维平面模板的灵活成像约束，显著降低了对实验配置的严苛要求，同时提升了径向/切向畸变系数的收敛精度。图 4-8 展示了平面标定板特征点的空间约束关系图解，而图 4-9 则系统描述了从几何特征提取到非线性优化迭代的完整标定算法执行逻辑拓扑图。

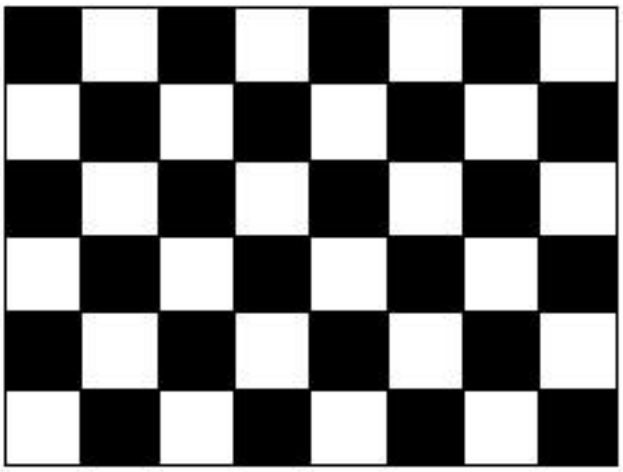


图 4-8 标定靶标示意图

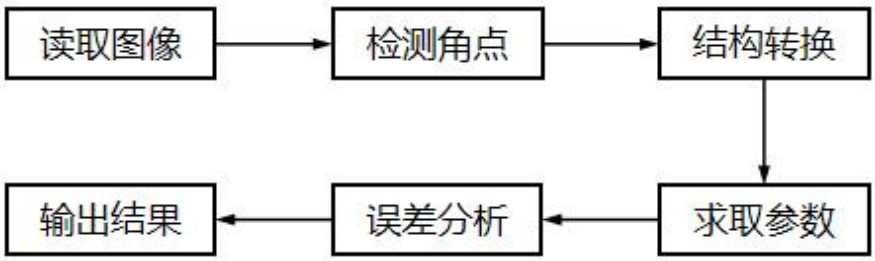


图 4-9 标定流程图

为了确定相机的内部参数，初始步骤是使用已知的角点坐标来计算相机的参数矩阵。具体而言，当棋盘格上的任意角点 z_w 等于零时，可以推导出公式（4-18）。

针对相机内参的精确解算需求，需基于二维标定板特征点的空间坐标观测

值构建投影映射关系的参数化模型。在标定算法的关键预处理阶段，当特征点齐次坐标满足 z_W 为零的归一化条件时（对应棋盘格角点位于基准坐标系原点），可依据透视投影方程推导出表示内参矩阵正交投影约束的关键方程式（4-18），该约束条件为后续非线性优化提供了初始估计的理论基础。

$$Z_C=\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x & 0 & u_0 & 0\\0 & f_y & v_0 & 0\\0 & 0 & 0 & 0\end{bmatrix}\begin{pmatrix}R_1 & R_2 & T\end{pmatrix}\begin{bmatrix}X_W\\Y_W\\1\end{bmatrix}=A\begin{pmatrix}R_1 & R_2 & T\end{pmatrix}\begin{bmatrix}X_W\\Y_W\\1\end{bmatrix} \quad (4\text{-}18)$$

因为 $Z_W=0$，$\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 是旋转矩阵前两列，$\boldsymbol{A}$ 代表内参矩阵。参数代入后，得到式（4-19）。

$$A(R_1R_2R_3)=H \quad (4\text{-}19)$$

而后通过代换得到式（4-20）：

$$\begin{bmatrix}u\\v\\1\end{bmatrix}=\frac{1}{z_c}H\begin{bmatrix}X_W\\Y_W\\1\end{bmatrix} \quad (4\text{-}20)$$

其中，$\boldsymbol{H}$ 代表内外参数矩阵的积，意味着内外参数矩阵相乘的结果，已知 $\boldsymbol{R}$ 被定义为一个正交矩阵。则有：

$$R_1^TR_2=0 \quad (4\text{-}21)$$

$$R_1^TR_1=R_2^TR_2=1 \quad (4\text{-}22)$$

在公式（4.21）和（4.22）中带入 $\boldsymbol{H}$ 可得到：

$$H_1^TA^{-T}A^{-1}H_2=0 \quad (4\text{-}23)$$

$$H_1^TA^{-T}A^{-1}H_1=H_2^TA^{-T}A^{-1}H_2=0 \quad (4\text{-}24)$$

基于最小二乘估计准则，利用不少于四个特征角点的坐标数据，可构建超定方程组并迭代求解出单应性矩阵 $\boldsymbol{H}$ 的最优估计值。通过采集多组非共线特征点数据，并依据其空间对应关系构建线性方程组，代入公式（4-23）和（4-24）中，可在满足投影几何约束的条件下，通过矩阵运算求得单应性变换矩阵的解析解。

基于张正友标定框架的畸变校正模型，主要针对相机成像系统中的径向畸变成因进行参数辨识。该算法着重建立二阶非线性畸变补偿的数学模型，其核心校正方程可表述为式(4-25)所示:

$$\begin{cases} x_r = x_p(1 + k_1r^2 + k_2r^4 + k_3r^6) \\ y_r = y_p(1 + k_1r^2 + k_2r^4 + k_3r^6) \end{cases} \quad (4\text{-}25)$$

依据公式（4-26），在进行坐标转换时考虑畸变效应，其中(u_r, v_r)代表畸变后的像素坐标值。

$$\begin{cases} u_r = \frac{x_r}{dx} + u_0 \\ v_r = \frac{y_r}{dy} + v_0 \end{cases} \quad (4\text{-}26)$$

通过将畸变公式代入（4-26）中，可得表达式（4-27）。

$$\begin{cases} u_r - u_0 = (u - u_0)(1 + k_1r^2 + k_2r^4) \\ v_r - v_0 = (v - v_0)(1 + k_1r^2 + k_2r^4) \end{cases} \quad (4\text{-}27)$$

对公式（4-27）进行代换可得表达式（4-28）。

$$\begin{cases} u_r = u + (u - u_0)(k_1r^2 + k_2r^4) \\ v_r = v + (v - v_0)(k_1r^2 + k_2r^4) \end{cases} \quad (4\text{-}28)$$

将公式转化为矩阵的形式可得公式（4-29）。

$$\begin{bmatrix} (u - u_o)r^2 & (u - u_o)r^4 \\ (v - v_o)r^2 & (v - v_o)r^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u_r - u \\ v_r - v \end{bmatrix} \quad (4\text{-}29)$$

基于式(4-29)建立的数学模型，每个特征点可构建两个线性约束条件，其约束参数包含理论投影坐标与畸变校正后的实际成像坐标两组已知量。通过矩阵分解方法对该方程组实施代数变换，最终可推导出如式(4-30)所示的简化求解形式。

$$Dk = d \quad (4\text{-}30)$$

在此阶段，再次采用最小二乘法来计算畸变参数矩阵$\boldsymbol{k}$，其具体表达式如式（4-31）所示，而这个$\boldsymbol{k}$矩阵就是的畸变参数。

$$k = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = (D^T D)^{-1} D^T d \quad (4\text{-}31)$$

4.3.2 监控相机标定及结果

本研究采用经典的张氏标定法对其进行了相机标定实验。标定目标为标准棋盘格，格子大小为 30×30，在不同视角和距离下采集 20 张图像，环境为室内自然光照，标定工具为 OpenCV 4.5.0 库，通过最小化重投影误差优化相机参数。

标定结果表明，相机内参矩阵为$\mathrm{M}=\begin{bmatrix}775.20 & 0 & 960.50\\ 0 & 775.30 & 540.75\\ 0 & 0 & 1\end{bmatrix}$，其中$\mathrm{f_x}=775.20$像素、$\mathrm{f_y}=775.30$像素、$\mathrm{c_x}=960.50$像素、$\mathrm{c_y}=540.75$像素，主点坐标接近图像中心（1920/2,1080/2），表明镜头中心与图像中心对齐良好。该相机焦距与分辨率匹配，畸变幅度较小，重投影误差低，为后续视觉任务提供了可靠的几何基础。

4.3.3 相机校准及直线距离获取

在利用 YOLOv5 检测跟踪目标后，获取检测框对角点的纵坐标，y_1、y_2。则目标在图像中的高度为：

$$h_{\text{image}}=y_1-y_2 \tag{4-32}$$

由于相机是倾斜的所以物体成像高度h_{image}会受到透视缩短的影响，利用相机安装倾角ε对成像高度校准：

$$h_{\text{corrected}}=h_{\text{image}}\cdot\cos\varepsilon \tag{4-33}$$

已知感光元件大小为 1/2.7"，成像比例为 16：9，图像分辨率为 1920×1080。则通过公式（4-34）,（4-35）可以计算出焦距f在像素单位下的表示f_{y}

$$y_{\text{px\_size\_y}}=\frac{S_{\text{y}}}{W_{\text{y}}} \tag{4-34}$$

$$f_{\text{y}}=\frac{f}{y_{\text{px\_size\_y}}} \tag{4-35}$$

$y_{\text{px_size_y}}$表示在y轴分量上单位像素在感光元件上的尺寸，S_{y}表示感光元件竖直方向尺寸，W_{y}表示图像在竖直方向上像素。

相机到物体的直线距离为：

$$d=\frac{H_{\text{object}}\cdot f_{\text{y}}}{h_{\text{corrected}}} \tag{4-36)}$$

H_{object}为物体实际高度。本文使用无人叉车尺寸为尺寸为 2200×950×1900，中国成年男性平均身高约 1.726m[60]，即$H_{\text{forklift}}=1.900\text{m}$，$H_{\text{worker}}=1.726\text{m}$。

4.3.4 基于三维距离恢复的人机距离计算

由于相机视场角θ固定，利用图中像素点间坐标通过几何关系即可解出行人及无人叉车同相机间夹角，利用余弦定理便可获得人机间相对距离。

通过 YOLOv5 算法对行人和无人叉车检测跟踪获取检测框中心点分别为 $P\mathrm{p}(x_\mathrm{p},y_\mathrm{p})$、$Pf(x_\mathrm{f},y_\mathrm{f})$行人和叉车相对于相机光轴的水平角度分别为：

$$\alpha_\mathrm{p}=\left(\frac{x_\mathrm{p}-\frac{W}{2}}{W}\right)\cdot\theta_h \tag{4-37}$$

$$\alpha_\mathrm{f}=\left(\frac{x_\mathrm{f}-\frac{W}{2}}{W}\right)\cdot\theta_h \tag{4-38}$$

相对于相机光轴的垂直角度分别为：

$$\beta_\mathrm{p}=\left(\frac{y_\mathrm{p}-\frac{H}{2}}{H}\right)\cdot\theta_\mathrm{v} \tag{4-39}$$

$$\beta_\mathrm{f}=\left(\frac{y_\mathrm{f}-\frac{H}{2}}{H}\right)\cdot\theta_\mathrm{v} \tag{4-40}$$

将 $P\mathrm{p}$、$P\mathrm{f}$ 的方向分别用两个单位向量v_p、v_f表示：

$$v_\mathrm{p}=\left(\cos\left(\alpha_\mathrm{p}\right)\cos\left(\beta_\mathrm{p}\right),\ \sin\left(\alpha_\mathrm{p}\right)\cos\left(\beta_\mathrm{p}\right),\ \sin\left(\beta_\mathrm{p}\right)\right) \tag{4-41}$$

$$v_\mathrm{f}=\left(\cos\left(\alpha_\mathrm{f}\right)\cos\left(\beta_\mathrm{f}\right),\ \sin\left(\alpha_\mathrm{f}\right)\cos\left(\beta_\mathrm{f}\right),\ \sin\left(\beta_\mathrm{f}\right)\right) \tag{4-42}$$

v_p、v_f之间夹角$\Delta\eta$可以通过公式（4-52）计算，其中$v_\mathrm{p}\cdot v_\mathrm{f}$是两个向量点积，$|\mathrm{v_p}|$和$|\mathrm{v_f}|$是向量的模长。

$$\cos\left(\Delta\ \eta\right)=\frac{v_\mathrm{p}\cdot v_\mathrm{f}}{|v_\mathrm{p}||v_\mathrm{f}|} \tag{4-43}$$

由于v_p和v_f是单位向量，即$|v_\mathrm{p}|$=$|v_\mathrm{f}|$=1，则有：

$$\Delta\eta=\arccos\left(\cos\left(\alpha_\mathrm{p}\right)\cos\left(\beta_\mathrm{p}\right)\cos\left(\alpha_\mathrm{f}\right)\cos\left(\beta_\mathrm{f}\right)+\right.$$

$$\left.\sin\left(\alpha_\mathrm{p}\right)\cos\left(\beta_\mathrm{p}\right)\sin\left(\alpha_\mathrm{f}\right)\cos\left(\beta_\mathrm{f}\right)+\sin\left(\beta_\mathrm{p}\right)\sin\left(\beta_\mathrm{f}\right)\right) \tag{4-44}$$

为化简计算，本文通过近似公式计算夹角 $\Delta\eta$

$$\Delta\eta=\sqrt{\left(\alpha_\mathrm{f}-\alpha_\mathrm{p}\right)^2+\left(\beta_\mathrm{f}-\beta_\mathrm{p}\right)^2} \tag{4-45}$$

因此，由余弦定理可知人机间距离d_pf：

$$d_\mathrm{pf}=\sqrt{d_\mathrm{p}^2+d_\mathrm{f}^2-2\times d_\mathrm{p}\times d_\mathrm{f}\times\cos\left(\Delta\eta\right)} \tag{4-46}$$

4.4 叉车及监控安全范围划分

4.4.1 叉车行驶坐标建立

在工业智能运输设备控制领域，构建基于空间关系的动态风险评估模型是实现人机协同作业的关键。本研究提出以载具本体坐标系为基准建立全局参考系：设定载具行进方向为 Y 轴正向，横向位移对应 X 轴，垂直维度表示为 Z 轴，从而形成符合运动学特征的笛卡尔空间坐标系。通过解算行人目标在该坐标系中的欧氏距离参数，建立三级渐进式安全防护机制。该模型将人机交互空间划分为临界预警区、动态避障区及紧急制动区，如附图 4-10 所示。

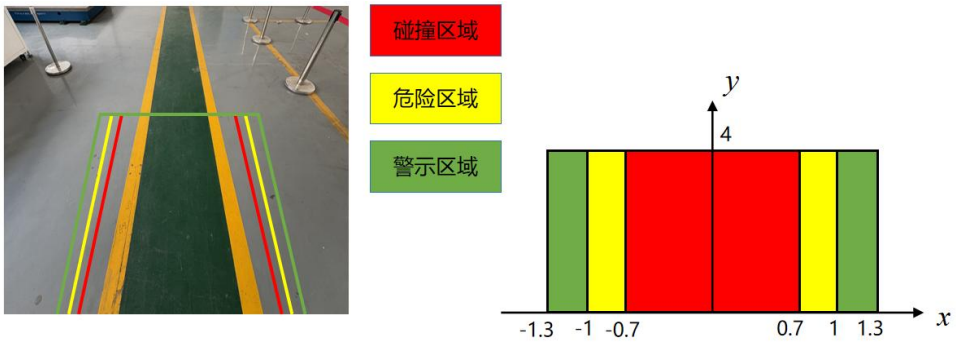


图 4-10 叉车行驶坐标系示意图

在工业车辆运动安全控制领域，本研究构建基于运动学特征与安全防护需求的分层预警机制。根据载具动力学模型，将本体坐标系 Y 轴正向 4 米范围设定为一级安全域场（碰撞区域，对应红色示警区），该阈值边界由典型托盘尺寸（1.5 米）与人机交互动力学参数共同确定。当行人目标进入该临界域时，系统触发急停以规避碰撞风险。基于此阈值，沿 X-Y 平面构建 0.3 米缓冲区形成二级安全区域（危险区域，黄色示警区），此时载具需执行减速步骤并触发声光告警装置。进一步通过 0.3 米缓冲区扩展生成三级安全域场（警示区域，绿色示警区），系统在该区域仅需激活声学报警装置。

4.4.2 人机防碰撞分级预警

综合人机目标检测与空间距离测量结果,依据无人叉车运动特性设计了人机碰撞三级安全预警范围并基于几何位置等效方法对碰撞预警进行可视化模拟。

因此本文以无人叉车车身中心位置为圆心，划分三级预警状态，分别为一级（接近状态）、二级（预警状态）、三级（危险状态），如图 4-11 所示。本文所使用的无人叉车工作速度为 1.5m/s，以行人正常行进速度 1m/s 为标准，即

人机相对速度为 2.5m/s。在保证行人具有 1s 的安全反应时间的前提下，可以推算出危险反应距离 R2-R1=2.5m。进一步设定接近状态与危险状态距离相同（R3-R2=R2-R1）,即 R1=2.5m、R2=5m、R3=7.5m，出于对安全性考量保证距离冗余，本文将 R1、R2、R3 的值取整，即 R1=3m、R2=5m、R3=8m。依据人机空间距离，由远到近将行人状态分为接近状态、预警状态和危险状态。当行人出现在 R3~R2 范围内时将对其未来 3 秒轨迹进行预测，当预测的行人轨迹端点在预警范围内时将出发预警系统，为行人提供足够反应时间，从而有效避免人机碰撞事故的发生。

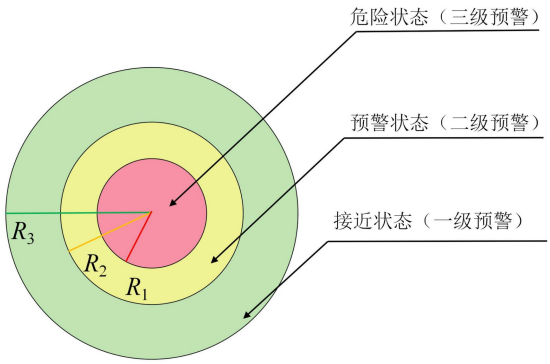
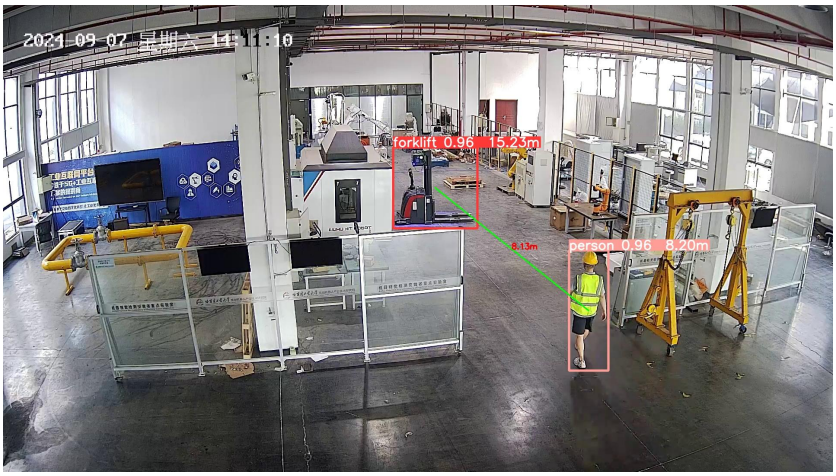


图 4-11 碰撞风险分级模型

4.5 单目测距实验及结果比较

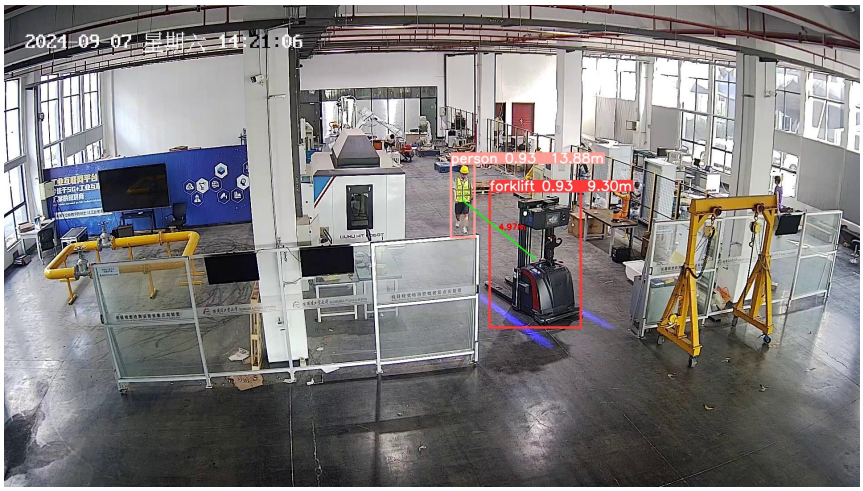
为更好的验证本文所提出方法的可行性，通过截取一段车间内监控视频进行验证。分别从单目距离测量和人机碰撞预警两个方面进行验证。

为方便测量固定人机间距离，本实验只考虑无人叉车静止行人移动状态。本文选取真实厂区监控下三个不同叉车行人位置进行实验，并在实验前测量出准确的人机距离，如图 4-12 所示。



实际距离 8m，检测距离 8.13m

实际距离 6m，检测距离 6.05m



实际距离 5m，检测距离 4.97m

图 4-12 三种不同人机位置下的单目测距结果

将本文提出的单目测距算法与实际人机距离对比，结果如表 4-1 所示。可

以看出本文所提出的基于几何关系的单目测距人机距离测算具有良好效果。

表 4-1 三种不同人机位置下测距结果与实际距离比较

位置	单目算法测量结果（m）	实际人机距离(m)	相对误差(%)
(a)	8.13	8	1.63
(b)	6.05	6	0.83
(c)	4.97	5	0.6

4.6 本章小结

本章基于单目相机测距原理开展了相机的标定工作，成功获取了相机的相关参数，并推导出了从图像坐标系到世界坐标系的坐标转换关系，从而实现了对行人位置的精准定位。此外，本研究还构建了叉车在行驶过程中的坐标系，并结合叉车的运行特性，对坐标系进行了安全等级的划分。通过这一系列工作，成功搭建了从目标检测到目标定位的转换机制。

在单目视觉测距理论框架下，本章完成了相机参数标定与几何建模，通过分析透视投影模型中的内参矩阵与外参矩阵，构建了像素坐标系至全局参考系的映射关系，实现了行人目标三维位姿的精确定位。针对工业载具运动特性，以载具行进方向为基准轴建立笛卡尔空间基准，并基于运动约束条件与人机交互安全阈值建立了动态风险场域分层模型。

第 5 章 基于 Y-Net 网络的行人轨迹预测算法研究

5.1 引言

针对无人驾驶工业车辆的路径安全规划问题，需重点识别两种行人动态：路径正向干扰型与轨迹交叉冲突型。在智能物流系统中，视觉感知系统需在实时检测行人的基础上，同步解析其运动学特征。通过构建行人运动学模型预测其短时轨迹，可实现动态风险场的梯度建模，从而量化评估不同空间位置的碰撞概率。基于此概率分布，系统可构建自适应避障机制和多级预警系统，通过轨迹优化算法生成满足安全阈值的行驶策略。该技术路径不仅有效降低了人机交互场景下的运动干涉风险，更通过前瞻性决策机制显著提升了工业车辆在复杂工况下的运行鲁棒性。

工业场景下的运动轨迹预测研究面临经典算法适用性不足的双重挑战。传统运动学模型基于匀速直线假设构建预测函数，其有效性局限于理想状态下的刚体运动。实际行人运动具有显著的非线性特征，受多体交互与环境约束耦合影响，导致常规线性外推方法产生显著预测偏差。虽然后续提出的社交长短期记忆网络（Social LSTM）通过构建空间交互矩阵实现了邻近行人运动关联建模，但其环境耦合机制仍存在建模缺失问题。更深层次的矛盾体现在评估体系维度：现有研究普遍采用 5-8 秒的长时预测评价标准，这与工业车辆 3 秒以内的决策响应周期产生时域失配，不仅导致预测置信度呈指数衰减，更引发不必要的计算冗余。本工作创新性地建立面向工业车辆的轨迹预测方法，基于特征解耦的自动编码架构开发短时域高精度预测模型，其技术框架包含多尺度时空注意力机制与环境语义融合模块。

5.2 基于自动编码器的行人轨迹预测模型

5.2.1 自动编码器概述

工业级特征学习领域，自动编码器（Autoencoder，AE）作为典型深度生成架构，其核心机理建立在信息瓶颈原理与特征解耦方式之上。该模型采用双通道编解码结构：编码器通过非线性变换将高维输入映射至低维潜变量空间，实

现特征压缩与信息蒸馏；解码器则基于潜变量进行信号重构，形成闭环特征学习系统。训练过程中，模型通过反向传播算法优化重构误差函数，驱动网络提取具有最大互信息的本质特征，同时抑制观测噪声与数据冗余。

从功能维度分析，AE 在特征工程领域展现出多重应用价值：（1）通过建立输入-潜变量-输出的双向映射关系，实现高维数据的流形学习；（2）基于潜变量空间的可解释性特征，构建面向异常检测的判别基准；（3）结合变分推理框架衍生生成式建模能力，实现数据分布的概率建模。特别在时序数据处理方面，AE 通过时间卷积或循环神经网络构建时序编码器，可有效捕获动态系统的相位特征与状态转移规律。

潜变量空间作为深度学习特征学习的核心载体，其数学本质是数据流形的低维嵌入空间。与传统欧氏空间相比，该空间具有以下特性：（1）拓扑保持性：保留原始数据的邻域结构；（2）特征解耦性：各维度对应独立语义因子；（3）操作可行性：支持向量插值与扰动分析。针对工业时序预测任务，本文提出潜变量空间转换策略：首先通过时序 AE 提取设备运行状态的紧凑特征，进而构建基于潜变量的预测模型。该方法突破原始信号空间的维度灾难问题，利用潜变量间的动力学耦合关系提升预测精度。

5.2.2 Y-Net 网络结构

Y-Net 网络是以 U-Net 分割网络为基础进一步发展的一种新型神经网络架构，它巧妙地整合了图像分割与分类功能，在乳腺癌的分割和诊断应用中表现出卓越的性能。Y-Net 的整体结构与 U-Net 有诸多相似之处，其上半部分同样采用了经典的“编码-解码”设计。整个网络主要由卷积层和池化层构成，未引入全连接层，具体架构可参考图 5-1 所示。在 Y-Net 中，编码网络可以被看作是由编码模块和下采样模块逐层堆叠而成的复合结构。编码模块的核心任务是从输入数据中提取和学习特征表示，而下采样操作则通过逐步缩小特征图的空间尺寸，帮助网络捕捉不同尺度下的不变性特征。然而，这一过程不可避免地会导致空间信息的部分丢失。为了弥补这一缺陷，解码器部分通过上采样模块和解码模块的组合进行设计。上采样模块能够有效恢复因下采样而降低的空间分辨率，而解码模块则通过整合信息，逐步重建编码阶段丢失的空间细节。与 U-Net 类似，Y-Net 也在编码器和解码器之间引入了跳跃连接，这一设计使得低层次的特

征信息能够直接传递到解码器，与高层次语义信息融合，从而提升分割精度。然而，Y-Net 在细节上进行了创新改进，其编码部分采用了 ESP block（高效空间金字塔模块）替代传统的卷积层。这一替换显著拓宽并加深了网络结构，使其具备更强的特征提取能力。研究表明，相比于较小规模的卷积神经网络（CNN），更大、更复杂的 CNN 架构通常在性能上具有明显优势。此外，Y-Net 还在 U-Net 的编码框架基础上进行了扩展，新增了一个分支结构。这一额外分支进一步增强了网络的多任务处理能力，使其在分割和分类任务中均表现出色。与传统方法相比，Y-Net 凭借其优化的设计和更强大的表达能力，在乳腺癌相关的分割与诊断任务中取得了显著优于传统方法的性能表现。

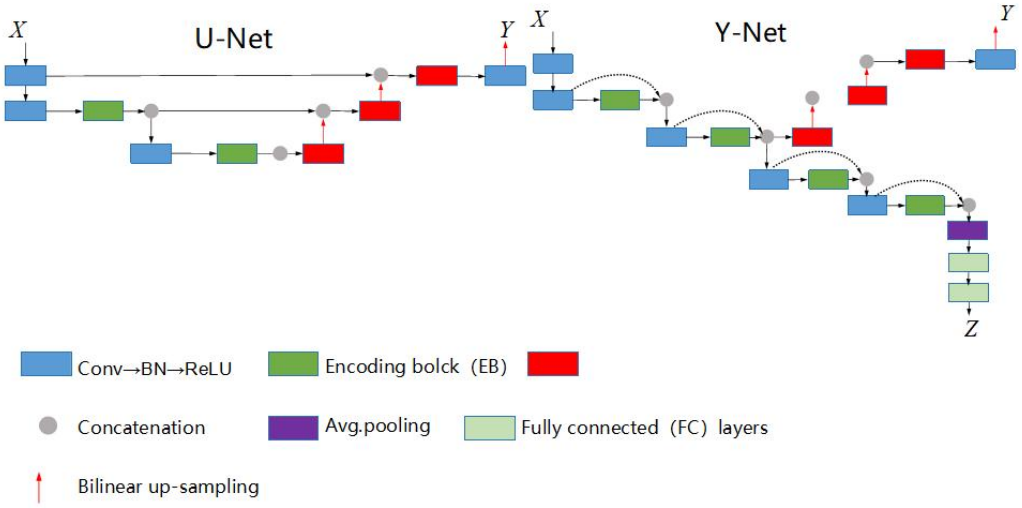


图 5-1 U-Net 与 Y-Net 的结构

5.2.3 轨迹预测算法设计

为实现行人轨迹实时预测本文结合座标卷积与自动编码器设计一种新型编码器，结构如图 5-2 所示。首先将 RGB 图像$\mathcal{I}$通过语义分割网络 Y-Net 处理，生成图像$\mathcal{I}$对应语义分割图 $\boldsymbol{S}$ 将行人过去运动$\{\mathbf{u}_n\}_{n=1}^{n_p}$转化为存在$n_p$个通道的轨迹热力图 $\boldsymbol{H}$，其中每帧所对应通道如公式（5-1）所示。式中 $\mathbf{H}(n,i,j)$表示在轨迹热力图中，对应过去第 n 帧的通道，在图像坐标(i,j)位置上的值，$\mathbf{u}_n$表示行人在过去第 n 帧时的真实位置坐标，利用$\|(i,j)-u_n\|$计算像素(i,j)与行人在第 n 帧时位置 $\mathbf{u}_n$ 的欧氏距离，并将归一化处理后的距离乘以 2，将值放缩到 0 到 2 之间，使热力图范围值方便后续处理。

$$H(n,i,j)=2\frac{\|(i,j)-u_n\|}{max_{(x,y)\in\mathcal{I}}\|(x,y)-u_n\|} \tag{5-1}$$

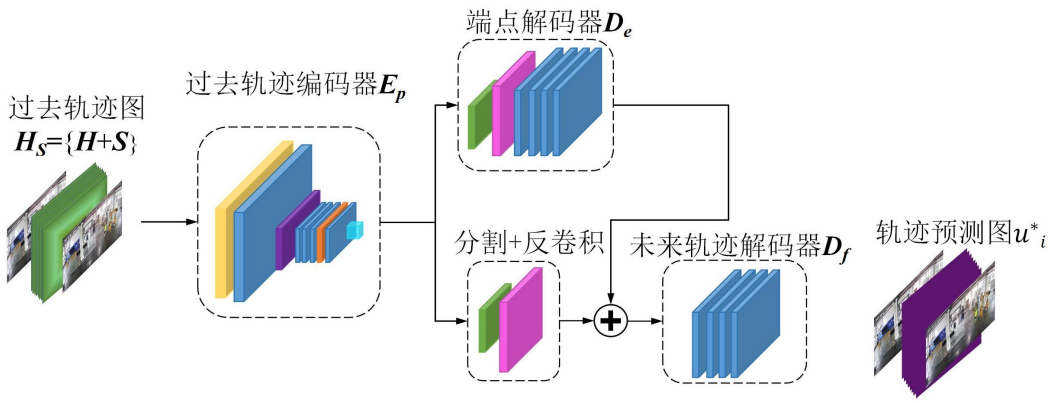


图 5-2 基于自动编码器的轨迹预测算法

将语义分割图 $\boldsymbol{S}$（维数为$H \times W \times n_p$）和轨迹热力图 $\boldsymbol{H}$（维数为$H \times W \times N_c$）沿深度方向拼接得到 $\boldsymbol{H_S}$（维数为$H \times W \times (n_p + N_c)$）。之后将 $\boldsymbol{H_S}$送入编码器 $\boldsymbol{E_p}$中，$\boldsymbol{E_p}$由若干个块组成，每个块后都有一个最大池化操作，这使得空间尺寸逐级减半，池化后经过卷积和激活函数将通道维度逐级递增。编码器 $\boldsymbol{E_p}$最终输出 $\boldsymbol{H_M}$与其中间输出 $\mathbf{H}_m(1 \le m \le M)$将被送入端点解码器 $\boldsymbol{D_e}$并在上采样后通过加权输入未来轨迹解码器 $\boldsymbol{D_f}$中。

$\boldsymbol{D_e}$的输出层由一个卷积后跟一个 Sigmoid 函数组成，对于每个 N^w选定的航路点 $\mathbf{u}_{wi}$和目的地$\mathbf{u}_{n_p+n_f}$在归一化后产生一个确定的$\mathbb{P}(\mathbf{u}_{wi})$、$\mathbb{P}\left(\mathbf{u}_{n_p+n_f}\right)$。$\boldsymbol{D_e}$的输出维度为$H \times M \times (N^w + 1)$,因此对于每个航路点和目的地都将产生一个$H \times M$矩阵，矩阵中第$(i,j)$个元素表示行人在特定帧下$(i,j)$处的估计概率值。

根据 $\boldsymbol{D_e}$中的概率分布得到航路点和目的地后产生 K_a个航路点和 K_e个端点，那么通往一个相同的端点则有 K_a条路经。其次将获得的端点 $\hat{\mathbf{u}}_{n_p+n_f}$和中间航路点$\{\hat{\mathbf{u}}_{wi}\}_{i=1}^{N^w}$的坐标转化为相应的热力图，用 $\mathbf{H}_g$表示。之后对向量 $\mathbf{H}_g$下采样以匹配每个块的空间尺寸。$\boldsymbol{D_f}$结构与 $\boldsymbol{D_e}$相似，最终的轨迹分布通过独立于通道每个像素 sigmoid 得到，为未来每个时间步长产生独立的空间大小为$H \times M$的热图。由于预测结果是每个时间步长上的显示概率分布，因此直接将损失加在估计分布$\mathbb{P}$上。$\boldsymbol{E_P}$、$\boldsymbol{D_e}$、$\boldsymbol{D_f}$这三个网络都是端到端联合训练的，在预测轨迹端点 end、航路点 waypoints 以及轨迹分布中使用的都是二进制交叉熵的加权组合。使用 λ_1=λ_2=1 对二元交叉熵损失进行加权。具体公式如下：

$$\mathcal{L}_{\text{end}} = \mathbf{BCE}\left(\mathbb{P}\left(\mathbf{u}_{n_p+n_f}\right), \widehat{\mathbb{P}}\left(\mathbf{u}_{n_p+n_f}\right)\right) \tag{5-2}$$

$$\mathcal{L}_{\text{waypoints}} = \sum_{i=1}^{N^w} \mathbf{BCE}\left(\mathbb{P}(\mathbf{u}_{wi}), \widehat{\mathbb{P}}(\mathbf{u}_{wi})\right) \quad (5\text{-}3)$$

$$\mathcal{L}_{\text{future trajectory}} = \sum_{i=n_p}^{n_p+n_f} \mathbf{BCE}\left(\mathbb{P}(\mathbf{u}_i), \widehat{\mathbb{P}}(\mathbf{u}_i)\right) \quad (5\text{-}4)$$

$$\mathcal{L} = \mathcal{L}_{\text{end}} + \lambda_1 \mathcal{L}_{\text{waypoints}} + \lambda_2 \mathcal{L}_{\text{future trajectory}} \quad (5\text{-}5)$$

本文对最终位置的估计非参数概率分布进行最大值索引选取操作，最终得到一个最可能的目标位置点$\mathbf{u}^*{}_{n_p+n_f}$，如公式（5-6）所示。

$$u^*{}_{n_p+n_f} = softargmax\left(\widehat{\mathbb{P}}\left(\mathbf{u}_{n_p+n_f}\right)\right) \quad (5\text{-}6)$$

对未来每个时间步长$i \in \{n_p, \ldots, n_p + n_f\}$，计算轨迹位置$u^*{}_i = softargmax\left(\widehat{\mathbb{P}}(\mathbf{u}_i)\right)$，同时确保$K_a$=1，其中$\widehat{\mathbb{P}}(\mathbf{u}_i)$是以最可能目标为条件预测的概率分布，$u^*{}_i$为输出的预测轨迹。

本文利用过去 6 秒轨迹预测未来 3 秒轨迹，由于$n_p = t_p \times FPS$，$n_f = t_f \times FPS$，且本文所使用数据集帧率在预处理后为 2.5FPS，所以本文$n_p = 15$，$n_f = 7.5$。

5.2.4 损失函数选择

为了定量分析算法在训练阶段的性能表现，模型依赖损失函数来衡量每次输出结果与真实值之间的偏差，并以此为基础计算反向传播所需的梯度值。损失值越小，说明模型的训练效果越优，其输出结果与真实的行人轨迹数据越贴近。

在本研究中，选用二维高斯密度函数作为损失函数，其具体的数学表达式如下所述：

$$p(x,y) = -\log \frac{1}{2\pi S_x S_y \left(\sqrt{1-tan^2\theta}\right)} exp\left(\frac{\left(\frac{x}{S_x}\right)^2 + \left(\frac{y}{S_y}\right)^2 - 2\left(\frac{tan\theta xy}{S_x S_y}\right)}{2\left(1-1-tan^2\theta\right)}\right) \quad (5\text{-}7)$$

式中x，y分别为预测坐标点的横纵坐标，S_x，S_y代表的是x，y分别对应的协方差矩阵。

5.3 行人轨迹预测及碰撞预警实验结果

5.3.1 实验环境和参数

训练平台和硬件采用 Windows 10 系统、NVIDIA GeForce RTX3060 图像处理器。在软件部署方面采用 Pytorch1.7 为深度学习框架，利用 CUDA 计算平台和 GPU 加速库 cuDNN 来降低内存占用并大幅提高 GPU 计算性能。详细参数设置如表 5-1 所示。

表 5-1 训练参数设置

参数名称	参数设置
输入图像大小	640×640
训练轮数	500
步长大小	32
优化器	SGD
动量	0.937
初始学习率	0.01
学习率因子	0.01

本文使用海康威视监控相机型号为 DS-IPC-B12H-LFT(PoE)，相机及感光元件参数如表 5-2 所示。本文使用 CDD20 型无人叉车，尺寸为 2200×950×1900，实验环境如图 5-3 所示。

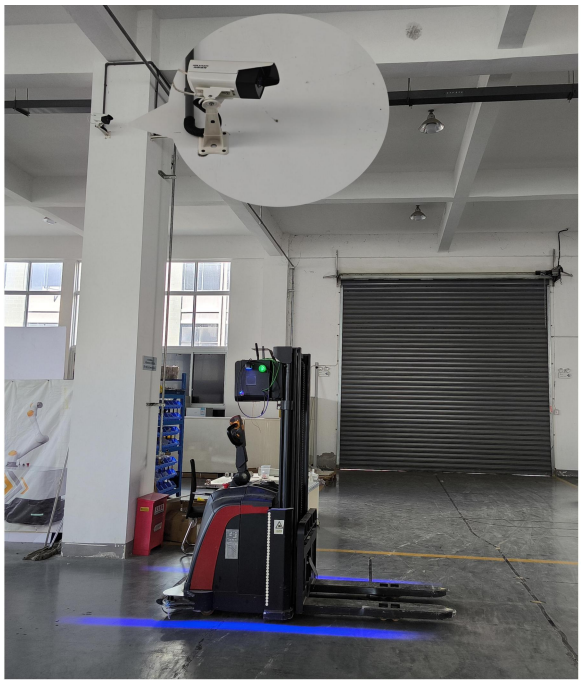


图 5-3 实验环境

表 5-2 相机及感光原件参数

图像分辨率	焦距	水平视场角θ_h	垂直视场角θ_v	传感器类型
1920×1080	4mm	91°	48°	1/2.7" Progressive Scan CMOS

5.3.2 性能评估指标

运动预测模型的性能评估体系中，两个核心评估参数分别从全局一致性和终点准确性两个维度衡量模型性能：平均位移误差（Average Displacement Error, ADE）与最终位移误差（Final Displacement Error, FDE）。其中，ADE 参数通过计算预测轨迹与真实轨迹各时间步对应点的欧氏距离均值，定量表示模型在整个轨迹序列上的综合预测精度，其数学表达式如下：

$$ADE(i) = \frac{1}{len}\sum_{k=1,2,\ldots,len}\sqrt{\left(x_{pred}^{k} - x_{true}^{k}\right)^2 + \left(y_{pred}^{k} - y_{pred}^{k}\right)^2} \quad (5\text{-}8)$$

$$ADE(i) = \frac{1}{N}\sum_{i=1,2,\ldots,N} A\,DE(i) \quad (5\text{-}9)$$

式中len表示每一条轨迹预测的轨迹点数目，x_{pred}^{k}，y_{pred}^{k}分别表示预测的第k个轨迹点的横纵坐标，x_{true}^{k}，y_{true}^{k}分别表示第k个轨迹点的实际横纵坐标。N表示轨迹的总数目。

最终位移误差（Final Displacement Error, FDE）：该指标聚焦于轨迹终点预测的空间偏移量，定义为所有预测轨迹终点与对应真实终点间欧氏距离的统计期望值。在轨迹预测模型的评估体系中，FDE 通过量化终点位置偏差的分布特征，重点评估模型对运动趋势终态的空间定位能力。其数学表达式如下所示：

$$FDE(i) = \sqrt{\left(x_{pred}^{final} - x_{true}^{final}\right)^2 + \left(y_{pred}^{final} - y_{true}^{final}\right)^2} \quad (5\text{-}10)$$

$$FDE = \frac{1}{N}\sum_{i=1,2,\ldots,N} FDE(i) \quad (5\text{-}11)$$

式中，x_{pred}^{final}，y_{pred}^{final}分别表示预测轨迹终点的横纵坐标，x_{true}^{final}，y_{true}^{final}分别表示实际轨迹终点的横纵坐标。

5.3.3 轨迹预测数据集引用

在深度神经网络训练阶段，数据分布的多样性与场景覆盖度直接影响模型的泛化性能。针对单一场景采集的行人运动数据存在显著的数据偏差问题，本

研究采用多源跨场景轨迹数据集联合训练策略，通过增强训练样本的时空特征完备性，有效提升模型对复杂运动模式的学习能力。

当前，自动驾驶领域已涌现出多个公开数据集，例如 ApolloScape 和 BDD100k[61]等。在轨迹预测领域的数据生态中，当前研究存在显著的模态适配性失衡问题，多数开源数据集以计算机视觉任务为导向构建数据标注体系，而具备完整运动学特征描述的轨迹预测专用数据资源存在显著缺口。本研究针对该问题建立跨模态数据适用性评估框架，对主流数据集进行多维特征解构与预测任务适配度分析，详细结果可参见表 5-3。

表 5-3 轨迹预测数据集

数据集	场景	描述
UCY[62]	中等人群密度行人轨迹数据集	三段视频
ETH[63]	拥挤环境下的小型行人轨迹数据集	一段视频
Stanford[64]	大型行人轨迹数据集	在斯坦福校园内的 8 个场景下的行人轨迹视频
HGSIM	车辆轨迹数据集	加州高速公路下拍摄的监控数据，标注较为粗糙
HighD	车辆轨迹数据集	轨迹数据集，包括车辆行为指标数据
TraPHic	驾驶人视野下的轨迹数据集	于亚洲收集的 48 个视频数据，数据较为粗糙
HEV-I	本田专业轨迹预测数据集	来源于美国的轨迹数据集，仅供美国科研机构使用

在轨迹预测研究领域中，ETH 和 UCY 数据集因其广泛适用性而备受青睐。ETH 数据集由苏黎世联邦理工学院开发与发布，主要涵盖两种典型场景：ETH 场景和 HOTEL 场景。其中，ETH 场景的数据来源于某酒店入口处，通过俯视角度的摄像头采集；HOTEL 场景的数据则是在某火车站附近，利用相似的俯视视角进行记录。ETH 数据集提供了包括轨迹帧数、行人标识（ID）以及行人在实际空间中的横纵坐标等信息。另一方面，UCY 数据集则记录了两种不同环境下的行人运动轨迹，具体包括大学校园内的 Univ 数据集，以及购物街区域的 Zara1 和 Zara2 数据集。

在行人运动预测研究领域，ETH 与 UCY 数据集因其基准地位被广泛采用为轨迹预测算法的基准测试标准。苏黎世联邦理工学院发布的 ETH 数据集包含两个具有空间异构特性的观测场景：其一是建筑入口监控视角，通过固定俯仰

角摄像头记录酒店主入口行人流动；其二是交通枢纽周边区域，采用相同采集方案获取火车站周边行人运动模式。该数据集提供多维运动学参数，包括时序帧索引、行人唯一标识符及地平面坐标系下的(x,y)坐标对。UCY 数据集则构建了空间异质性的环境对比框架，涵盖学术机构开放空间与商业街区密集人流区域两类典型城市空间场景下的行人轨迹数据，所有观测数据均包含高精度时空标注。

5.3.4 ETH 和 UCY 数据集实验结果

在运动预测模型的性能评估框架中，主流研究通常采用基于 8 秒历史观测序列预测 12 秒未来轨迹的基准测试，并借助 ADE（平均位移误差）和 FDE（最终位移误差）两项指标进行量化评估。但实证研究表明，工业动态场景下的轨迹预测需求具有显著的时间敏感性特征：以动态障碍物规避系统为例，其实时决策引擎仅需 3-5 秒轨迹数据即可完成运动意图推断，无需进行长时域预测。针对此矛盾，本研究构建双阶段验证体系：首先遵循领域规范实施传统长时域评估流程，同时创新性地设计短时域，基于前 6 秒轨迹数据预测后 3 秒轨迹的方案。该模块整合通用评估指标与场景特异性评价参数，重点验证算法在工业实时决策场景下的时空特征提取效率与运动模式泛化能力。

为了进一步验证本文算法在性能和精度方面的优越性，此处以 6 秒时间序列坐标数据为基础，预测后续 3 秒轨迹，并针对 ADE 和 FDE 指标，与现有的多种轨迹预测模型开展了对比实验。这些模型包括长短期记忆网络（LSTM）、社会长短期记忆网络（S-LSTM）、社会对抗网络（S-GAN）以及时空图注意力网络（STGAT）。

实验结果在表 5-4 和表 5-5 中展示，涵盖了五个数据集的表现，数据单位为米。数值越小，表明预测轨迹与真实轨迹的吻合度越高，模型的预测精度也越优异。表中以加粗形式标示的数据表示在各场景下性能最佳的模型。

表 5-4 各模型 ADE 预测结果

		UNIV	HOTEL	ETH	ZARA1	ZARA2	AVG
ADE	S-GAN	0.76	0.67	0.87	0.35	0.42	0.61
	Social-LSTM	0.67	0.79	1.09	0.47	0.56	0.72
	Vanilla-LSTM	0.61	0.86	1.09	0.41	0.52	0.70

		UNIV	HOTEL	ETH	ZARA1	ZARA2	AVG
	Liner	0.82	0.39	1.33	0.62	0.77	0.79
	STGAT	0.52	0.56	0.88	0.41	0.31	0.54
	本文方法	0.24	0.10	0.28	0.17	0.13	0.18

根据表 5-4 的数据分析可知，在以 ADE 作为算法性能的评估标准时，本章提出的算法在除 ETH 场景之外的所有数据集中均表现出更高的精度，优于其他对比模型。表 5-5 的结果显示，当采用 FDE 作为评估指标时，本章算法在全部测试数据集上均展现出显著的优越性。综合实验结果表明，基于图卷积神经网络的方法在预测精度上明显超越了传统的时序网络模型。此外，S-LSTM 的测试表现揭示，当算法对行人轨迹数据的挖掘不足时，其预测效果会相对较弱。而继 S-LSTM 之后提出的多种方法，则从不同角度深入分析了轨迹数据，相较于 S-LSTM 实现了显著的精度提升。本研究提出的方法通过时空图的视角，进一步强化了对轨迹数据的挖掘与处理，经过重构后取得了实验中的最佳效果。

基于表 5-4 的量化分析，本研究构建的时空图卷积架构在 ADE 指标维度呈现出显著的环境适应性差异：除具有显著空间异质性的 ETH 场景外，该模型在其余四个基准数据集上的预测误差均值较对比模型降低。表 5-5 的 FDE 评估结果进一步验证模型的终态预测优势，其在所有测试场景中均保持统计显著性优势。横向对比实验表明，基于图神经网络（GNN）的方法在空间关系建模能力上较传统时序模型（LSTM）具有本质性提升。本研究所提出的时空图嵌入机制，通过动态邻域关系建模与多尺度运动模式解耦的协同优化，验证了时空图表示在复杂运动预测任务中的理论优越性。

表 5-5 各模型 FDE 预测结果

		UNIV	HOTEL	ETH	ZARA1	ZARA2	AVG
FDE	S-GAN	1.52	1.37	1.62	0.68	0.84	1.21
	Social-LSTM	1.40	1.76	2.35	1.00	1.17	1.54
	Vanilla-LSTM	1.31	1.91	2.41	0.88	1.11	1.52
	Liner	1.59	0.72	2.94	1.21	1.48	1.59
	STGAT	1.13	1.15	1.65	0.91	0.68	1.11
	本文方法	0.41	0.14	0.33	0.27	0.22	0.27

在算法性能的多维度验证框架中，本研究突破传统精度评价的单维局限，构建包含计算效率与部署可行性的综合评价体系。如表 5-6 所示，通过跨场景基准测试对比分析表明：本模型在架构设计层面采用拓扑精简策略，其参数空间复杂度较主流方法实现数量级压缩，显著降低边缘计算设备的存储压力。特别地，通过解耦传统时序递归结构并引入并行化特征提取机制，系统推理延迟较基于 LSTM 的基线模型优化两个数量级，成功突破工业实时系统的时间窗口约束阈值。该设计方法验证了轻量化图神经网络架构在动态场景预测任务中的工程落地优势。

表 5-6 模型参数对照表

模型	模型大小（M）	运行速度（ms）
Social-LSTM	110.6	393
S-GAN	109.5	383
GCN	80.5	26.7
本文算法	22.8	12.6

5.3.5 车间内行人轨迹预测结果

在实时碰撞预警方面，本文截取一段厂区中监控视频，视频帧率为 30FPS，视频时长为 15 秒。根据本文 3.4 节所提出的人机防碰撞分级预警，当画面中行人出现在距离无人叉车 8 米范围内时，对其未来 3 秒运动轨迹进行预测。本文以进入预测范围的上一帧为起始帧；以到达预测终点为第 0 帧；以预警后行人进入危险状态为最后帧，如图 5-4 所示。

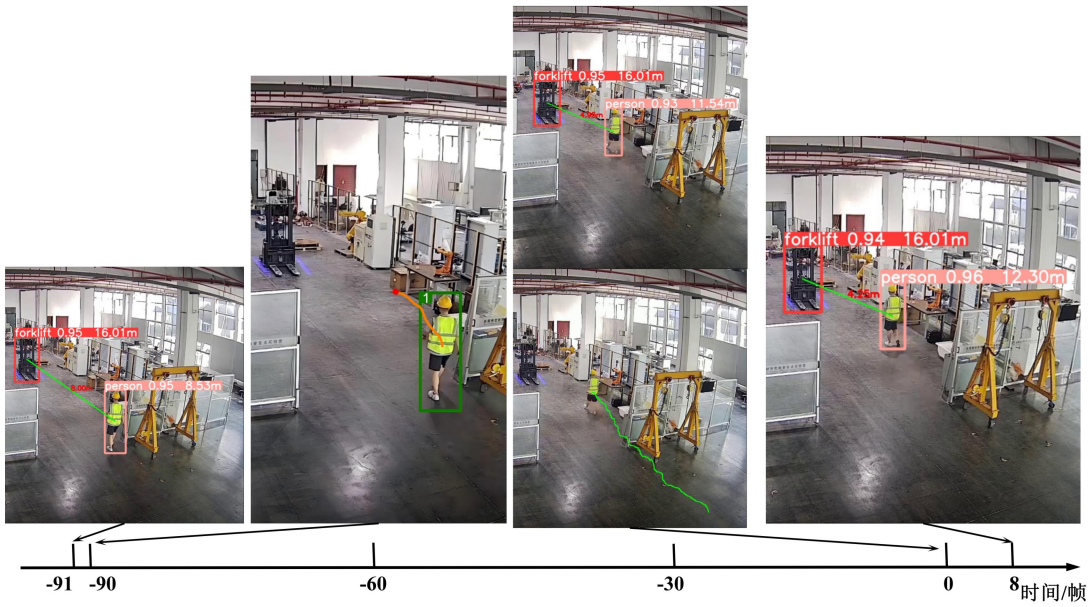


图 5-4 预警时间序列分析图

由图 5-4 可以看出在第-90 帧时，本文所提轨迹预测算法预测出三秒后行人轨迹端点（图中红点所示），第-90 帧到第 0 帧行人轨迹如图中绿线所示，可以看出本文算法在行人进入预警范围时，准确预测出行人轨迹，可以在人机碰撞前发进行预警。可见，在行人进入预警范围时，利用本文所提出的实时轨迹预测模型和分级预警机制，可准确在人机碰撞前发出预警，确保车间生产安全。

5.4 本章小结

针对工业车辆智能化场景下的轨迹预测需求，本研究创新性地构建了基于 Y-Net 架构的动态轨迹预测模型，并建立了面向工业场景的轨迹预测评价体系。区别于传统研究方法，本工作首次设计了基于历史轨迹时序特征的预测精度评估机制，通过构建特定时间窗口的轨迹映射关系实现预测模型的性能验证。为增强研究结论的普适性，实验环节同时采用公开基准数据集与自主采集的工业场景轨迹数据进行对比验证，其中自主数据源自真实仓储环境下的行人运动观测。研究结果表明，相较于现有主流算法，本模型在轨迹误差等核心评价维度上展现出明显优势，其网络结构设计有效平衡了计算效率与模型复杂度。在工业场景的实测验证中，模型输出的预测轨迹呈现出与实际运动模式高度一致的空间分布特性。该研究成果在保持实时响应能力的同时显著提升了预测准确性，为无人叉车系统的安全决策提供了可靠的技术支撑。

第 6 章　总结与展望

6.1 总结

随着智能制造的快速发展，车间场景下人机协同作业的需求日益增加，其高动态性与复杂性对安全保障机制提出了更高要求。本研究聚焦工业场景中智能搬运设备与作业人员共存的安全挑战，系统性地构建了面向智能制造的车间人机协同安全保障技术体系，为提升作业安全性和物流效率提供了创新性解决方案。本研究的主要技术突破包括以下几个方面：

（1）目标检测增强技术：通过架构优化型 YOLOv5 网络设计，创新性地引入多尺度特征融合机制与改进损失函数，有效克服传统检测方法在复杂工业场景中的目标遮挡与密集分布问题，实现复杂人机共存环境下的鲁棒检测性能优化。

（2）轻量化协同检测框架：融合轻量化网络架构与特征增强策略，构建安全装备状态识别系统。该框架通过特征提取模块的跨层连接优化，在保持实时处理能力的同时提升细粒度特征的表达能力，为作业人员安全装备佩戴检测提供可靠技术方案。

（3）人机交互预警系统：研发多级人机交互预警体系，集成运动轨迹预测模型与空间关系解析方法。通过建立动态风险场模型实现人机运动态势的实时推演，基于时空约束条件构建分级预警决策机制，显著提升工业现场的安全防护水平。

经实验验证，本文所研究的基于深度学习的车间人员安全监测与预警，在目标检测鲁棒性、识别准确率及人机冲突预警时效性等关键指标上均具有良好效果。各功能模块的协同作用有效提升了复杂工业场景下的整体安全性，为智能制造系统的安全运行提供了新思路。

6.2 不足之处与展望

本研究仍存在以下亟待突破的学术问题和技术难点，需在后续工作中重点攻关：

首先，在环境感知维度，现有方法主要依赖单一视觉模态的影像分析，存在复杂工况下的识别局限性。建议构建多源异构数据融合框架，整合激光点云、热成像图谱等多维度传感信息，通过跨模态特征对齐与知识蒸馏技术，建立鲁棒性更强的环境感知模型。这将有效解决光照变化、遮挡干扰等工业现场常见问题，提升目标检测的召回率与定位精度。

其次，在风险干预机制方面，需突破现有单向预警模式的技术瓶颈。可基于数字孪生技术构建人机环协同决策平台，实现安全系统与 AGV 调度、物流管控等工业物联网子系统的双向数据交互。通过开发具有边缘计算能力的自适应调控算法，构建"感知-评估-响应"的闭环控制体系，实现从风险预警到动态避障的智能决策闭环，例如通过轨迹重规划算法实时调整无人叉车的运动参数。

最后，在应用推广层面，应建立面向不同制造业场景的柔性化技术适配方案。针对离散制造与流程工业的差异性需求，重点攻克模块化架构设计、领域自适应迁移学习等关键技术，通过与典型企业共建工业验证平台，加速技术成果在汽车制造、化工生产等典型场景的工程化落地，推动智能安全系统从实验验证向产业化应用转化。

参考文献

[1] 唐小林,甘露,李国法,等.面向自动驾驶的大模型对齐技术：综述[J].汽车工程,2024,46(11):1937-1951.

[2] Katrakazas C, Quddus M, Chen W H, et al. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions[J]. Transportation Research Part C: Emerging Technologies, 2015, 60: 416-442.

[3] Alessandrini A, Cattivera A, Holguin C, et al. CityMobil2: challenges and opportunities of fully automated mobility[J]. Road vehicle automation, 2014: 169-184.

[4] Schwarting W, Alonso-Mora J, Rus D. Planning and decision-making for autonomous vehicles[J]. Annual Review of Control, Robotics, and Autonomous Systems, 2018, 1(1): 187-210.

[5] 魏凌涛,王翔宇,邱彬,等.基于自适应预瞄路径的自动驾驶车辆寻迹和避障控制[J].机械工程学报,2022,58(06):184-193..

[6] 于赫年,白桦,李超.仓储式多 AGV 系统的路径规划研究及仿真[J].计算机工程与应用,2020,56(02):233-241.

[7] 王雷雄,王波,马富齐,等.基于双目立体匹配和场景元素识别的变电人员近电安全距离检测方法研究[J].电网技术,2023,47(03):1010-1021.

[8] 周玉,赵小锋,汪一,等.关键细粒度信息指导的多尺度遮挡行人重识别[J].电子与信息学报,2024,46(06):2578-2586.

[9] 何智敏,钱江波,严迪群,等.基于长短期时间关系网络的视频行人重识别[J].电子学报,2024,52(08):2746-2757.

[10] 杨静,张灿龙,李志欣,等.集成空间注意力和姿态估计的遮挡行人再辨识[J].计算机研究与发展,2022,59(07):1522-1532.

[11] 廉继红,薛维哥,王延年,等.基于卷积注意力模块的人体姿态估计研究[J/OL].西安工程大学学报,1-9[2025-03-23].

[12] 张海燕,张富凯,袁冠,等.多姿态图像生成的行人重识别算法研究[J].计算机工程与应用,2023,59(02):143-152.

[13] 李佳宁,王东凯,张史梁.基于深度学习的二维人体姿态估计:现状及展望[J].计算机学报,2024,47(01):231-250.

[14] 谢明鸿,康斌,李华锋,等.Anchor free 与 Anchor base 算法结合的拥挤行人检测方法[J].电子与信息学报,2023,45(05):1833-1841.

[15] ChenZ ,LinY ,XuJ , et al.A fused score computation approach to reflect the over lap between the predicted box and the ground truth in pedestrian detection[J].IE T Image Processing,2024,18(13):4287-4296.

[16] 汤书苑,周一青,李锦涛,等.基于特征校准的双注意力遮挡行人检测器[J].西安电子科技大学学报,2024,51(06):25-39.

[17] Feng Y, Wu R, Li P, et al. Multi-view high-dynamic-range 3D reconstruction an d point cloud quality evaluation based on dual-frame difference images[J]. Appl ied Optics, 2024, 63(30): 7865-7874.

[18] Savitsky V, Schmies L, Gumenyuk A, et al. Comparative performance of DIC a nd optical flow algorithms for displacement and strain analysis in laser beam we lding[J]. Optics and Lasers in Engineering, 2025, 187: 108870.

[19] Alawode B, Javed S. Learning spatial–temporal regularized tensor sparse RPCA for background subtraction[J]. IEEE Transactions on Neural Networks and Lea rning Systems, 2025.

[20] Kamel M M, Hussein S I, Salama G I, et al. Efficient Target Detection Techniq ue Using Image Matching Via Hybrid Feature Descriptors[C]//2020 12th Intern ational Conference on Electrical Engineering (ICEENG). IEEE, 2020: 102-107.

[21] Lipton A J, Fujiyoshi H, Patil R S. Moving target classification and tracking fro m real-time video[C]//Proceedings fourth IEEE workshop on applications of co mputer vision. IEEE, 1998: 8-14.

[22] 盖绍彦,黄妍妍,达飞鹏.基于通道注意力和特征切片的图像快速匹配算法[J].光学学报,2023,43(22):158-166.

[23] Wang F, Yu S, Yang J. Robust and efficient fragments-based tracking using me an shift[J]. AEU international Journal of Electronics and Communications, 2010, 64(7): 614-623.

[24] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part based models[J]. IEEE Transactions on Software Engineering, 2010, 32(9): 1627-1645.

[25] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. 2015: 91-99.

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

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

[28] 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.

[29] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[30] REDMON J, FARHADI A. Yolo v3: An incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.

[31] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: Optimal speed and accuracy of object detection [J]. arXivPreprint arXiv:2004.10934, 2020.

[32] Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic[J]. Nature, 2000, 407(6803): 487-490.

[33] Johansson A, Helbing D, Shukla P K. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data[J]. Advances in complex systems, 2007, 10(supp02): 271-288.

[34] Koehler S, Goldhammer M, Bauer S, et al. Stationary detection of the pedestrian's intention at intersections[J]. IEEE Intelligent Transportation Systems Magazine, 2013, 5(4): 87-99.

[35] Kooij J F P, Schneider N, Flohr F, et al. Context-based pedestrian path predictio n[C] //European Conference on Computer Vision. Springer, Cham, 2014: 618-6 33.

[36] Bera A, Kim S, Randhavane T, et al. GLMP-realtime pedestrian path prediction using global and local movement patterns[C]//2016 IEEE International Confere nce on Robotics and Automation （ICRA）. IEEE, 2016: 5528-5535.

[37] Hochreiter S, Schmidhuber J. Long Short-term Memory[J]. Neural Computation, 1997, 9（8）: 1735-1780.

[38] Liu J, Shahroudy A, Xu D, et al. Spatio-temporal lstm with trust gates for 3d hu man action recognition[C]//European Conference on Computer Vision. Springer, 2016: 816-833.

[39] Alahi A, Goel K, Ramanathan V, et al. Social lstm: Human trajectory prediction in crowded spaces[C]//Proceedings of the IEEE Conference on Computer Visio n and Pattern Recognition. 2016: 961-971.

[40] Gupta A, Johnson J, Fei-Fei L, et al. Social gan: Socially acceptable trajectories with generative adversarial networks[C]//Proceedings of the IEEE conference o n computer vision and pattern recognition. 2018: 2255-2264.

[41] Hecht-Nielsen R. Theory of the backpropagation neural network[M]//Neural net works for perception. Academic Press, 1992: 65-93.

[42] Cun Y L , Boser B , Denker J S , et al. Handwritten digit recognition with a bac k-propagation network[J]. Advances in neural information processing systems, 1990, 2(2):396--404.

[43] Lecun Y , Bottou L . Gradient-based learning applied to document recognition[J] . Proceedings of the IEEE, 1998, 86(11):2278-2324.

[44] Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belie f Nets[J]. Neural Computation, 2014, 18(7):1527-1554.

[45] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep conv olutional neural networks[J]. Advances in neural information processing system s, 2012, 25.

[46] Jiang N, Xu J, Goto S. Pedestrian detection using gradient local binary patterns[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2012, 95(8): 1280-1287.

[47] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.

[48] Mohamed A, Qian K, Elhoseiny M, et al. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14424-14432.

[49] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, realtime object detectiono[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 779-788.

[50] 彭浩. 基于 YOLOv5 的无人机巡检图像绝缘子检测技术的研究[D].中国矿业大学,2021.

[51] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE Computer Society, 2017: 6517-6525.

[52] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[EB/OL].https: //arxiv.org/abs/1804. 02767, 2018-04-08.

[53] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[54] Rezatofighi H , Tsoi N , Gwak J Y ,et al.Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2019.

[55] 杜鹏,宋永红,张鑫瑶. 基于自注意力模态融合网络的跨模态行人再识别方法研究[J]. 自动化学报,2022,48(6):1457-1468.

[56] Zheng Z , Wang P , Liu W ,et al.Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J].arXiv, 2019.

[57] 李润东,曲英伟,殷丽凤,等.YOLO-sea：改进 YOLOv7-tiny 的复杂海底目标检测算法研究[J].计算机工程与应用,2025,61(02):247-258.

[58] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 3-1

[59] 董红召,林少轩,佘翊妮.交通目标 YOLO 检测技术的研究进展[J].浙江大学学报(工学版),2025,59(02):249-260.

[60] 国家体育总局体育科学研究所.国家国民体质监测中心发布《第五次国民体质监测公报》[EB/OL].(2022-06-07)[2023-10-30].https: // www.sport.gov.cn/n315/n329 /c24335066/content.html.

[61] Maddern W, Pascoe G, Linegar C, et al. 1 year, 1000 km: The Oxford RobotCar dataset[J]. The International Journal of Robotics Research, 2017, 36(1): 3-15.

[62] Pellegrini S, Ess A, Schindler K, et al. You'll never walk alone: Modeling social behavior for multi target tracking[C]//2009 IEEE 12th international conference on computer vision. IEEE, 2009: 261-268.

[63] Lerner A, Chrysanthou Y, Lischinski D. Crowds by example[C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2007, 26(3): 655-664.

[64] Robicquet A, Sadeghian A, Alahi A, et al. Learning social etiquette: Human trajectory understanding in crowded scenes[C]//European conference on computer vision. Springer, Cham, 2016: 549-565.

攻读硕士学位期间取得学术成果

一、发表学术论文

[1] Liu Yiping, Jiang Benchi, He Huan, Chen Zhijun, Xu Zhenfa. Helmet wearing detection algorithm based on improved YOLOv5[J]. Scientific reports, 2024, 14(1): 8768.（SCI 收录）

[2] Jiang Benchi, Liu Yiping, Xu Zhenfa, Chen Zhijun. Enhancing AGV path planning: An improved ant colony algorithm with nonuniform pheromone distribution and adaptive pheromone evaporation[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2024: 09544070251327268.（SCI 收录）

[3] 刘懿平, 何欢, 江本赤. 基于改进 YOLOv5 算法的密集动态目标检测方法[J]. 安徽科技学院学报,2024,38(02):79-86.

二、申请知识产权

[1] 第一发明人.一种用于 3D 打印的成型平台[P]. 中国专利：CN202221753070.0, 2022.

[2] 第二发明人（导师第一发明人）.一种基于视觉焊缝跟踪系统的焊缝纠偏装置[P]. 中国专利：CN202421299231.2, 2025.

致谢

时光荏苒，硕士阶段的学习和研究即将画上圆满的句号。在完成这篇论文的过程中，我得到了许多人的支持与帮助，在此向他们致以最诚挚的谢意。

首先，我要衷心感谢我的导师江本赤教授。您的悉心指导、严谨的治学态度和渊博的学识让我受益匪浅。您不仅在学术上为我指点迷津，还在生活中给予我关怀与鼓励，使我能够在科研的道路上不断前行。

感谢我的父母和家人，感谢你们无条件的支持与理解。你们是我人生道路上最坚实的后盾，无论我身处何地，你们的爱与关怀始终伴随着我，激励我克服困难、追求卓越。

感谢我的同门孙笑笑、何欢，与你们一起探讨问题、分享经验的时光让我受益良多。你们的陪伴让我的研究生生活充满了温暖与欢乐。

感谢课题组的全体成员，感谢你们在实验设计、数据分析等方面提供的帮助与建议，你们的专业精神和团队合作让我深受启发。

最后，感谢所有在我研究生阶段给予我帮助和支持的人。正是因为有了你们，这段旅程才变得充实而有意义。这篇论文不仅是我个人努力的结晶，更是所有人共同支持的成果。

在此，我亦怀一颗赤子之心，愿伟大祖国繁荣昌盛！

2025 年 3 月

尚德教学　唯实惟新