分类号:TP391　　学校代码:10363
密　级:公开　　学　　号:2220320166

安徽工程大学

AnhuiPolytechnicUniversity

硕士学位论文

题目　基于深度学习的红外图像行人检测算法研究

论 文 作 者　温福新
校 内 导 师　许钢 教授
校 外 导 师　朱标 高级工程师

专业学位类别　电子信息
研究方向（领域）　智能信息处理及应用

论文提交日期：2025 年 5 月 19 日

分类号：TP391　　单位代码：10363

密　级：公开　　学　　号：2220320166

题　　目 基于深度学习的红外图像行人检测算法研究

题　　目 Research on Deep Learning based Pedestrian Detection Algorithms with Infrared Images

论文作者：温福新

校内导师：许钢 教授

校外导师：朱标 高级工程师

专　　业：电子信息

研究方向：智能信息处理及应用

论文答辩日期：2025 年 5 月 12 日

安徽工程大学学位论文原创性声明

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

学位论文作者签名：温福新

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：温福新　　指导教师签名： 

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

基于深度学习的红外图像行人检测算法研究

摘 要

近年来，基于深度学习的行人检测算法研究成为计算机视觉领域的热点问题之一。但当前的行人检测算法主要集中在可见光图像上，对于红外图像行人检测问题仍然存在挑战。红外图像由于其在低光照、复杂环境下的优势，已广泛应用于安全监控、自动驾驶等领域。然而，红外图像行人检测仍存在一些问题。其一，目前开源的红外图像行人数据集匮乏，无法充分满足本文算法训练的需求；其二，现有的红外图像行人检测算法往往面临着模型体积大、检测精度低以及目标丢失等问题；其三，在红外图像中存在行人遮挡问题，导致行人的特征信息不完整，增加了行人检测的难度。针对以上问题，本文的主要工作内容和创新点如下：

(1) 从红外图像数据集 KAIST 中分别精心筛选出包含行人与包含行人遮挡情况的红外图像，分别建立本文使用的红外图像行人数据集与红外图像行人遮挡数据集。同时，为了进一步提升图像质量，提高数据集的多样性，对建立的数据集使用 CLAHE 算法对图像进行增强处理。最后，对预处理后的数据集使用 LabelImg 标注软件进行人工标注，确保数据集的质量和可用性。

(2) 针对 YOLOv5s 目标检测算法在红外图像行人检测中模型体积大、检测精度低等问题，提出一种轻量化的红外图像行人检测算法 YOLOv5s-CSBS。该算法通过构建 C3CA 模块、添加 SimAM 注意力

机制模块两种方法，对其主干网络进行改进，从而减少对行人错检、漏检的情况；其次，引入加权双向特征金字塔网络到颈部网络中，增强网络特征表达能力；最后，为了降低模型的参数量与复杂性，引入包含 GSConv 的 Slim-Neck 设计范式，更好地平衡模型准确性和速度，实现轻量化设计。实验结果表明：YOLOv5s-CSBS 在红外图像行人数据集上，mAP@0.5 值比 YOLOv5s 提高 2%，同时参数量与计算量分别降低了 10.7%和 18.4%。

(3) 针对红外图像行人遮挡问题，提出了一种基于 YOLOv10n 改进的检测网络模型 YOLOv10n-SCM。首先，引入 SCConv 卷积模块与主干网络中的 C2f 模块相结合，增强模型对细节特征的捕捉能力，提高整体的检测效率和性能；其次，在颈部网络引入 CCFM 模块，更好地处理尺度变化带来的挑战，从而提升模型在复杂场景下的检测性能；最后，构建 C2f-EMA 模块，C2f-EMA 模块通过更好地获取上下文信息，进一步提升了模型对红外图像中行人特征信息的提取能力。实验结果表明：在红外图像行人遮挡数据集上，改进后的算法模型 mAP@0.5 值比 YOLOv10n 提高 1.7%，同时参数量与计算量分别降低了 38.3%和 26.8%。

关键词：深度学习，红外图像行人检测，遮挡行人检测，轻量级网络，注意力机制

RESEARCH ON DEEP LEARNING BASED PEDESTRIAN DETECTION ALGORITHMS WITH INFRARED IMAGES

ABSTRACT

In recent years, research on pedestrian detection algorithms based on deep learning has become one of the hot issues in the field of computer vision. However, current pedestrian detection algorithms mainly focus on visible light images, and there are still challenges in pedestrian detection in infrared images. Due to their advantages in low-light and complex environments, infrared images have been widely used in security monitoring, autonomous driving, and other fields. However, there are still some problems in infrared image pedestrian detection. First, the current open-source infrared image pedestrian datasets are scarce and cannot fully meet the training needs of the algorithm in this paper. Second, existing infrared image pedestrian detection algorithms often face problems such as large model size, low detection accuracy, and target loss. Third, there is a problem of pedestrian occlusion in infrared images, which leads to incomplete feature information of pedestrians and increases the difficulty of pedestrian detection. To address these issues, the main work and innovations of this paper are as follows:

(1) Carefully select infrared images containing pedestrians and those with pedestrian occlusion from the KAIST infrared image dataset to establish the infrared image pedestrian dataset and the infrared image pedestrian occlusion dataset used in this paper. At the same time, to further improve image quality and increase the diversity of the dataset, the CLAHE algorithm is used to enhance the images in the established datasets. Finally, the preprocessed datasets are manually labeled using the LabelImg annotation software to ensure the quality and usability of the datasets.

(2) To address the problems of large model size and low detection accuracy of the YOLOv5s object detection algorithm in infrared image pedestrian detection, a lightweight infrared image pedestrian detection algorithm, YOLOv5s-CSBS, is proposed. This algorithm improves the backbone network by constructing the C3CA module and adding the SimAM attention mechanism module, thereby reducing the false detection and missed detection of pedestrians. Secondly, the weighted bidirectional feature pyramid network is introduced into the neck network to enhance the network's feature expression ability. Finally, to reduce the model's parameter quantity and complexity, the Slim-Neck design pattern containing GSConv is introduced to better balance model accuracy and speed, achieving lightweight design. Experimental results show that YOLOv5s-CSBS achieves a 2% increase in mAP@0.5 on the infrared image pedestrian dataset compared to YOLOv5s, while reducing the

parameter quantity and computational cost by 10.7% and 18.4%, respectively.

(3) To address the problem of pedestrian occlusion in infrared images, a detection network model based on the improved YOLOv10n, YOLOv10n-SCM, is proposed. Firstly, the SCConv convolution module is combined with the C2f module in the backbone network to enhance the model's ability to capture detailed features and improve overall detection efficiency and performance. Secondly, the CCFM module is introduced into the neck network to better handle the challenges brought by scale changes, thereby enhancing the model's detection performance in complex scenes. Finally, the C2f-EMA module is constructed. The C2f-EMA module further improves the model's ability to extract pedestrian feature information in infrared images by better obtaining context information. Experimental results show that the improved algorithm model achieves a 1.7% increase in mAP@0.5 on the infrared image pedestrian occlusion dataset compared to YOLOv10n, while reducing the parameter quantity and computational cost by 38.3% and 26.8%, respectively.

Wen Fuxin(Electronic Information)

Supervised by Professor Xu Gang

KEY WORDS: deep learning, infrared image pedestrian detection, occlusion pedestrian detection, lightweight network, attention mechanism

目录

第 1 章 绪论

1.1 研究背景及意义

随着科技的不断发展，图像处理与计算机视觉在现代社会中发挥着越来越重要的作用。其中，行人检测[1]作为计算机视觉中的一个重要任务，已经引起了国内外学者广泛的研究兴趣。行人检测技术能够自动识别视频或图像中出现的人体目标，具有重要的现实意义，例如智能监控[2-3]、自动驾驶[4]与智能交通管理[5]等领域。这些领域对于公共安全、交通效率以及个人安全防护等方面具有至关重要的作用。在智能监控系统中，行人检测技术能够实现对视频中行人目标的准确识别与跟踪，为安全监控提供重要信息，有助于预防犯罪和维护社会秩序。自动驾驶技术的发展同样离不开行人检测，它能够帮助车辆实时感知周围环境中的行人，从而做出安全的驾驶决策，减少交通事故的发生。在智能交通管理领域，行人检测技术可以应用于交通流量分析、行人行为预测等方面，为城市交通规划和管理提供科学依据。此外，安全监控系统中，行人检测技术的运用也极大地提升了安全防护的效率和准确性。图 1-1 为行人检测示意图。

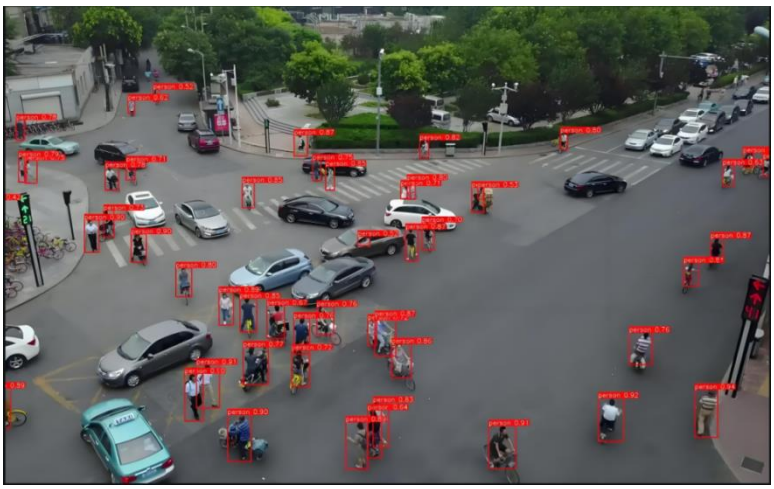


图 1-1 行人检测示意图

传统的行人检测技术主要依赖于可见光图像，虽然这些方法已经取得了一些成果，但在复杂环境下仍面临诸多挑战。例如，在低光照、雨雪、雾霾等恶劣天气条件下，基于可见光的行人检测精度降低。为了突破这些限制，红外图像行人检测[6-7]作为一种新兴的研究方向，正逐渐受到重视。红外图像能够在低光照和

恶劣环境中保持较好的可见性，特别是在夜间或遮蔽物较多的环境中，红外图像展现出较高的识别能力。红外图像行人检测研究不仅能弥补传统可见光检测技术的不足，还能为智能交通、安全监控等领域的应用提供技术支撑。具体来说，本研究的意义在于：(1) 提高低光照环境下行人检测的准确性：传统检测方法在光照不佳的情况下检测性能大幅下降，而红外成像系统通过捕捉热辐射图像，在低光或无光条件下依然能进行检测。通过深度学习模型对红外图像的特征学习，能够提高在恶劣光照条件下的行人检测精度。(2) 提升复杂环境中的检测效果：红外图像在雨雾天气、强光逆光等复杂环境中具有较强的穿透能力，使得行人检测不易受到天气和光照变化的干扰。研究红外图像行人检测有助于在复杂场景下提升检测效果，从而提高整体系统的适用性和鲁棒性。(3) 推动自动驾驶和智能监控系统的应用发展：行人检测是自动驾驶和智能监控系统中的关键环节，可靠的行人检测结果对避免事故和提升行人安全至关重要。红外图像结合深度学习方法，可改善检测系统在特殊环境下的检测能力，为自动驾驶和监控系统的安全性提供更高保障。

尽管红外图像在特定环境下展现出优势，但其行人检测技术仍面临众多挑战。例如，红外图像的分辨率通常较低，且图像质量在很大程度上受到传感器性能的影响，这可能会限制检测的精度。此外，红外图像中背景噪声较多，且背景与人体的热量差异并不总是十分明显，这导致行人检测算法在红外图像中的表现往往不及在可见光图像中的效果。因此，提升红外图像行人检测的准确性和鲁棒性，已成为研究领域的焦点。

综上所述，红外图像行人检测研究不仅具有实际的应用价值，而且对目标检测领域的发展具有深远的理论意义。通过这项研究，期望能够在复杂环境下构建更为高效、稳定的行人检测系统，为相关领域的技术应用提供有力支持。总的来说，红外图像行人检测技术在多个领域中展现了广阔的应用前景。随着红外传感技术、计算机视觉和深度学习算法的持续进步，未来红外图像行人检测有望在智能交通、安全监控、无人驾驶等领域扮演更加关键的角色，成为提升公共安全和自动化水平的核心技术之一。图 1-2 为可见光图像与红外图像。

(a) 可见光图像　　(b) 红外图像

图 1-2 可见光图像与红外图像

1.2 国内外研究现状与难点

1.2.1 国内外研究现状

随着深度学习和计算机视觉技术的不断进步，特别是在智能交通和公共安全等对高效行人检测技术需求日益增长的背景下，红外图像行人检测技术受到了广泛关注并得到了深入研究与广泛应用。红外图像技术在夜间或在复杂气候条件下，能够精准捕捉人体的热辐射特征，因此在行人检测领域展现出了其独特的优势。无论是国内还是国际上，红外图像行人检测的研究都取得了一系列研究成果。

行人检测方法主要有基于特征提取的传统模型和深度学习模型。传统的行人检测方法主要依赖于手工设计的特征和分类器，例如 Haar[8]特征分类器、HOG[9]特征分类器和 Adaboost[10]分类器等。

Haar 特征是基于计算图像区域内不同亮度变化的特征，最早由 Viola 和 Jones 在其提出的快速人脸检测方法中使用。该方法通过使用矩形区域的亮度对比来提取图像的特征。Haar 特征分类器广泛应用于行人检测，特别是在一些实时应用场景中，因为它计算效率较高。尽管 Haar 特征分类器计算速度较快，但其精度相对较低，且对噪声、光照变化等较为敏感，在复杂背景下的表现并不理想。

HOG 特征是由 Dalal 和 Triggs 于 2005 年提出的，是行人检测中应用广泛的传统特征之一。HOG 特征通过计算图像区域的梯度方向和幅值，并将其转化为直方图，从而获得该区域的描述信息。HOG 特征能有效捕捉到图像中的局部形状信息，尤其对于行人的姿态和轮廓具有较好的敏感性。HOG 特征通常与支持向量机(Supported Vector Machine, SVM)[11]分类器结合使用。SVM 是一个强大的

二分类器，可以通过构建最优超平面来区分目标和背景。然而，HOG 特征方法对视角变化、遮挡等情况仍较为敏感，且在复杂场景中，背景干扰可能影响检测精度。

Adaboost 是一种经典的集成学习方法，主要用于提高弱分类器的性能。在行人检测中，Adaboost 通常与简单的特征一起使用。Adaboost 通过加权的方式不断调整分类器的权重，逐步增强检测效果。其优势在于能够将多个弱分类器组合成一个强分类器，从而提升检测精度。Adaboost 在早期的行人检测算法中得到了广泛应用。通过不断迭代学习，Adaboost 能够选择最优的特征组合，从而提高分类准确性。尽管该方法能够在较低复杂度下进行行人检测，但对于复杂环境中的遮挡、光照变化等问题，Adaboost 仍然缺乏较强的鲁棒性。

David G Lowe[12-13]于 2004 年提出的尺度不变特征变换算法，具有旋转不变性、尺度不变性及对亮度变化的鲁棒性，是一种非常可靠的局部特征描述方法。

Olmeda[14]等在 2010 年提出了一种新的红外行人检测特征，即利用相位一致的局部能量信息直方图(Histograms of Oriented Phase Energy, HOPE)特性描述红外行人轮廓。

2016 年，胡庆新[15]等人采用 HOG 特征和 ISS 特征联合训练 SVM 分类器，并通过滑动窗口法进行行人分类检测。

胡燕伟[16]等人在 2019 年提出了一种基于 BING-casDPM 的快速行人检测方法。该方法通过训练 SVM 分类器，并利用 casDPM 模型来进行行人检测。

Xue[17-18]针对夜间光线不足导致传统可见光算法在行人和车辆检测中的效果下降，提出了一种多特征融合算法，利用红外图像中行人和车辆目标的特征进行优化。

基于手工特征的行人检测算法通常适用于行人特征明显的简单场景。然而，在实际环境中，行人的特征容易受到周围环境的干扰，难以通过手工设计的特征进行有效提取。相比之下，使用深度神经网络模型通过对行人数据集的持续训练，可以获得更为精准的行人特征，因此更适合应对复杂多变的环境中的行人检测任务。

随着深度学习的兴起，基于深度学习的行人检测方法逐渐成为主流，但传统模型仍具有一定的研究价值和应用前景，尤其在对计算资源有限制的场景中。深度学习技术特别是卷积神经网络(Convolutional Neural Network, CNN)[19-21]在计算机视觉领域取得了进展，行人检测作为深度学习方法的一个重要应用领域，受到了广泛关注。与传统方法相比，深度学习方法通过自动学习图像中的特征，能够有效应对复杂环境、遮挡、尺度变化等挑战，提升了检测精度。卷积神经网络是深度学习中最基本且最常用的神经网络结构。它通过一系列卷积层、池化层和全连接层，能够自动提取输入图像中的特征。CNN 特别适合图像数据的处理，因为它能够通过局部感知区域和共享权重的设计，减少参数数量，并自动从数据中学习到多层次的特征表达。在行人检测中，CNN 通常用于图像分类和特征提取。早期的行人检测方法采用浅层的 CNN 结构，在简单场景中能够取得不错的效果，但面对复杂的背景、尺度变化、姿态变化等问题时，CNN 的性能常常不尽如人意。因此，随着深度网络的进一步发展，许多更复杂的模型应运而生。区域卷积网络(Region-based CNN, R-CNN)是一种用于目标检测的深度学习算法，由 Ross Girshick[22]等人在 2014 年提出。该算法通过结合 CNN 和区域提议的思想，提高了目标检测的精度，但该算法的检测速度较慢。2015 年，Girshick[23]提出 Fast R-CNN，它是对 R-CNN 算法的优化，解决了 R-CNN 在训练和推理时的低效问题，但 Fast R-CNN 网络计算能力较差。为了解决这个问题，Shaoqing Ren[24]等人在 2015 年提出 Faster R-CNN，该算法基于 R-CNN 方法，结合了区域提议网络和卷积神经网络，极大地提高了检测速度和精度。Faster R-CNN 的工作流程包括两个主要阶段：第一个阶段是通过 RPN 生成候选区域，第二个阶段则使用 CNN 对这些候选区域进行分类和回归，从而完成目标的定位和分类。Faster R-CNN 通过共享卷积特征，使得区域提议和目标检测过程可以联合训练，从而提高了整体效率。对于行人检测，Faster R-CNN 通过生成候选框并对其进行精确分类，能够在多种尺度和背景下有效地检测出行人。随后，研究者们进行了进一步的探索，Joseph Redmon[25]等人于 2016 年提出 YOLO(You Only Look Once, YOLO)实时目标检测方法，YOLO 采用单一的神经网络进行端到端的回归问题求解。YOLO 将整个图像作为输入，并在网络的输出中直接预测目标的类别和位置。它通过将图像划分

为网格，并在每个网格内预测边界框和类别概率，能够同时完成目标的定位和分类。

YOLO系列模型具有较高的检测速度，非常适合实时检测应用，特别是在资源有限的情况下。自从其首次提出以来，YOLO经历了从YOLOv1到YOLOv11的不断迭代，每个版本都在性能、准确性和效率方面进行了改进。其中，YOLOv3[26]和YOLOv4[27]等版本不断改进，增加了对小目标和不同尺度目标的检测能力，因此在行人检测中，YOLO也表现出了良好的性能。尽管YOLO的精度相比Faster R-CNN稍低，但其优越的速度和实时性使得它在许多实际应用中成为首选。SSD(Single Shot Multibox Detector, SSD)是另一种流行的目标检测框架，由Wei Liu[28]等人于2016年提出。它采用单一的卷积神经网络来同时进行目标的位置回归和类别分类。SSD通过多个尺度的特征图来预测边界框和类别，从而使其能够检测不同尺寸的目标。SSD相比于YOLO具有更好的准确性，尤其是在小目标检测上表现突出。SSD的优点是检测速度快，且可以在多个尺度上进行检测，适应不同大小的目标。在行人检测中，SSD通过高效的卷积网络和多尺度特征预测，能够准确识别不同尺寸的行人。SSD和YOLO在实时检测任务中都表现得非常优秀，但SSD在精度上略有优势，尤其是在复杂场景下，对行人的检测能力较强。吕传龙[29]等人针对遮挡行人问题对Faster R-NN模型进行相应的改进，通过将ResNet50替换为HRNet与引入NMS-Loss等方法，解决遮挡行人目标在特征提取过程中信息容易丢失的问题。李彤晖[30]等研究人员针对有遮挡的行人检测问题，通过对初始中心点的选取进行优化，并引入注意力机制模块对YOLOv5m进行改进，从而提高了检测精度。然而，他们在改进过程中并未充分考虑到模型轻量化的需求，这在实际应用中可能会带来一定的局限性。

梁秀满[31]等人通过采用轻量化的 MESNet 网络替代 CSPDarkNet53，嵌入注意力机制 SA 模块，并利用轻量化的 DSASFF 结构优化特征融合网络等方法对 YOLOv5s 进行了改进。虽然这种方法在一定程度上满足了模型轻量化的要求，但检测精度有所下降，这在实际应用中可能会对检测效果产生负面影响。

综上所述，基于深度学习的行人检测方法已经取得了一些进展。经典的卷积神经网络和区域提议网络方法为深度学习行人检测奠定了基础，YOLO、SSD、

Faster R-CNN 等方法则进一步提高了检测精度与速度。同时，基于 Transformer 的检测方法也展示了新的发展方向。随着深度学习技术的不断进步，行人检测的性能有望不断提升，未来的研究将侧重于提高检测精度、实时性以及在复杂环境中的鲁棒性。

1.2.2 红外图像行人检测的技术难点

红外图像行人检测面临许多技术难点，主要源于红外图像的特点、环境的复杂性以及目标检测任务的挑战。以下是一些常见的技术难点：

(1) 缺乏大规模标注数据集。与可见光图像相比，红外图像的数据集较为匮乏，尤其是在大规模标注数据集方面。标注红外图像中的行人需要人工介入，这不仅费时费力，而且在不同的环境条件下，红外图像的特征差异较大，进一步增加了标注的复杂性。

(2) 低对比度和低信噪比。红外图像通常具有较低的对比度和较高的噪声，特别是在低温或复杂环境条件下。这使得在红外图像中检测行人比在可见光图像中更加困难。低对比度使得背景和行人的区分变得模糊，而高信噪比会影响特征的提取和目标的定位。

(3) 背景干扰。在夜间或低光环境中，红外图像的背景往往会包含各种复杂的物体（如建筑物、树木或其他动态物体），这些背景干扰会与行人目标相似，导致检测算法难以准确区分目标与背景。

(4) 遮挡问题。在实际场景中，行人经常会部分遮挡，如背对摄像头、与其他物体或行人发生遮挡等。这种情况下，传统的目标检测算法往往难以准确检测到被遮挡的行人。

(5) 行人特征不明显。相比于在可见光图像中，红外图像中行人的特征可能不如在人类眼睛中看到的那么明显。由于红外图像是基于热辐射获取的，行人的形状、细节和纹理可能不如可见光图像中的那样突出，这给检测算法的特征提取和分类任务增加了难度。

1.3 本文主要研究内容和结构安排

综上所述，本文针对红外图像行人检测技术展开研究，通过优化算法网络模

型来提升检测精度和模型轻量化设计。

本文的具体研究内容如下：

第 1 章绪论。本章首先阐述了课题的研究背景及意义，然后概述了行人检测技术的国内外研究现状，最后，介绍了本文的主要研究内容及章节结构安排。

第 2 章相关理论基础介绍。本章首先详细介绍了红外成像的基本原理以及可见光图像与红外图像的特性分析；其次，对深度学习与卷积神经网络进行了阐述；最后，介绍了行人检测算法的评价指标。

第 3 章红外图像行人检测算法的研究。本章针对 YOLOv5s 目标检测算法在红外图像行人检测中模型体积大、检测精度低等问题，提出了一种基于改进 YOLOv5s 算法的红外图像行人检测算法 YOLOv5s-CSBS。首先，将 Backbone 网络中的 C3 模块全部替换成 C3CA 模块，以增强对行人的长距离检测能力；其次，在 Backbone 网络中的 SPPF 下一层引入了 SimAM 注意力机制模块，以提升模型对行人特征的关注度，并增强对不同深度特征信息的感知能力；再次，在 Neck 网络中引入 BiFPN，不仅加强了低层与高层特征间的信息交流，还促进了不同尺度特征的有效整合；最后，引入包含 GSConv 的 Slim-Neck 设计范式降低模型复杂性，更好地平衡模型准确性和速度。

第 4 章红外图像行人遮挡检测算法的研究。本章针对红外图像行人遮挡问题，提出一种基于改进 YOLOv10n 算法的红外图像行人遮挡检测算法 YOLOv10n-SCM。首先，在网络中引入了 CCFM 模块，提高模型对尺度变化的适应性和对小尺度对象的检测能力；其次，引入 SCConv 卷积模块，并将其融合到 Backbone 部分的 C2f 模块中。SCConv 模块通过捕捉不同尺度和通道的特征信息，进一步提升了模型的特征提取能力；最后，构建 C2f-EMA 模块，并用它来替换原有 Neck 端中的 C2f 模块。C2f-EMA 模块通过更好地获取上下文信息，进一步提升了模型对红外图像中行人特征信息的提取能力。

第 5 章为总结与展望。对全文的工作内容和研究成果进行总结，分析目前存在的不足之处，对未来红外图像行人检测做了进一步的研究和探讨。

第 2 章 相关理论基础

2.1 引言

行人检测一直是智慧交通、自动驾驶等领域研究的热点。它不仅关乎公共安全，还是智能交通系统实现车辆自动导航、行人行为分析、交通流量控制等功能的关键技术之一。随着城市化进程的加速和智能交通系统的发展，行人检测的重要性日益凸显。它要求算法能够准确、快速地识别出视频或图像中的行人，并给出其位置信息，为后续的行人跟踪、行为识别等任务提供基础。因此，行人检测技术的持续研究和改进对于推动智慧交通、自动驾驶等领域的发展具有重要意义。

目前对于行人检测的研究主要建立在可见光图像上，对于红外图像的研究很少。然而，在实际应用中，特别是在夜间或光线条件较差的环境下，可见光图像的行人检测效果会受到影响。相比之下，红外图像不受光线条件的限制，能够在这些环境下提供更为清晰的行人信息。因此，本文研究基于红外图像的行人检测技术，对于提升全天候行人检测能力具有至关重要的意义。展望未来，随着技术的不断进步和应用需求的不断增长，红外图像行人检测有望成为智慧交通、自动驾驶等领域的关键研究方向之一。本章节首先对算法涉及的相关理论进行介绍。

2.2 红外成像基本原理

热红外成像[32]的基本原理是基于物体所发出的红外辐射来检测温度分布，并将这些信息转化为可视化图像。每个物体在其温度高于绝对零度时，都会发出一定波长范围的红外辐射。红外成像技术通过捕捉这种辐射，将其转化为图像，帮助查看物体的温度状态和热分布。红外成像系统结构如图 2-1 所示。

(1) 红外辐射与物体温度的关系：根据斯特藩-玻尔兹曼定律，所有温度高于绝对零度的物体都会发出红外辐射。物体的辐射强度与其表面温度直接相关，具体来说，辐射强度和物体的绝对温度的四次方成正比，如式(2-1)所示：

$$E = \sigma T^4 \tag{2-1}$$

式(2-1)中，E 是辐射的能量，σ 是斯特藩-玻尔兹曼常数，T 是物体的绝对温度（单位：K）。

(2) 红外探测器：热红外成像设备通常配备红外探测器（如焦平面阵列），这些探测器能够捕捉不同波长的红外辐射。探测器将捕捉到的红外辐射转化为电信号。

(3) 信号处理：探测器将红外辐射转化为电信号后，通过图像处理系统进行处理，最终生成热图像。图像中的每个像素表示一个位置的温度，显示为不同的颜色或灰度。

(4) 热图像的生成：热红外图像是根据热辐射强度生成的，其中每个像素的颜色或亮度值表示该区域的温度。通常，热成像仪会用不同的颜色（如红色、蓝色、绿色等）来表示不同的温度值，这样可以清晰地展示物体的热分布。常见的热图像表示为热图，其中每个像素的值直接对应一个温度，通过颜色标尺显示温度的不同范围。

(5) 显示与分析：最终的热成像图像或视频会被显示在监视器上供人们观察。通过分析，可以得知物体表面的温度分布情况，从而用于识别问题区域或进行热量分析。

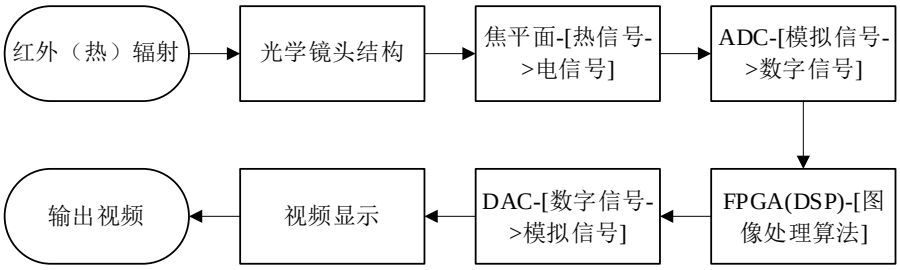


图 2-1 红外成像系统结构

2.3 可见光图像与红外图像特性分析

红外图像与可见光图像在图像获取原理、图像特征以及应用场景方面有明显的差异。了解这两种图像的特性，对于选择合适的图像处理方法和算法非常重要。下面将从多个方面对红外与可见光图像的特性进行对比分析。

(1) 成像原理

可见光图像是由人眼可见波长范围内的光线（约 400-700nm）形成的图像。光线通过场景中的物体后，反射或透过物体并进入相机的镜头，形成图像。在可见光成像中，物体的颜色、纹理、形状等视觉信息是图像中的主要特征。红外图

像则是通过捕捉物体发出的红外辐射（波长范围通常在 700nm 至 1mm 之间）来形成的图像。物体根据其温度发射不同强度的红外辐射，温度较高的物体会发出更多的红外辐射。因此，红外图像更多地呈现出温度分布而非物体的表面特征。图 2-2 为电磁辐射波谱。

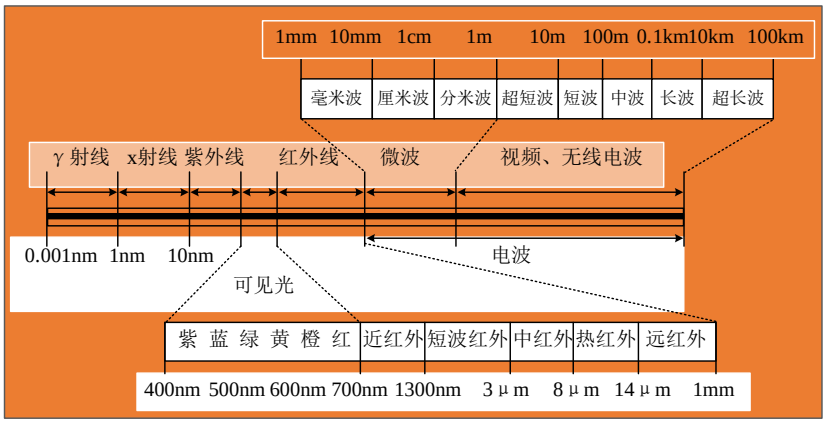


图 2-2 电磁辐射波谱

(2) 光照和条件影响

可见光图像对光照条件非常敏感。在白天和夜间，光照强度差异巨大，白天的图像比夜间的图像更加清晰明亮。在夜间，缺乏足够的光源时，可见光图像可能无法有效捕捉场景信息。此外，阴天、雾霾、雨雪等天气条件也会影响可见光图像的质量，导致成像效果下降。与可见光图像不同，红外图像对光照条件不敏感。即使在完全黑暗的环境中，红外成像设备依然能够通过感知物体发出的热辐射形成图像。因此，红外图像在夜间和低光照环境下表现优异。红外图像还能够穿透雾霾、烟雾等气象干扰，在复杂天气条件下保持较好的成像效果。

(3) 目标检测和识别

可见光图像在日常场景中表现良好，特别适合于物体的外观识别、颜色分类等任务。许多传统的目标检测方法在可见光图像中能够获得较好的性能。然而，在低光照、夜间或者天气恶劣的环境中，可见光图像的检测能力受到严重限制。红外图像通常用于检测物体的热量分布，因此，它特别适合于夜间、低光照或者视距受限的环境。红外图像能够在完全黑暗的环境中有效地检测到人体等热源，即使在遮挡、模糊或背景复杂的情况下也能维持较高的检测准确性。红外图像在行人检测、安防监控、无人驾驶等领域中具有重要应用，尤其是在夜间或恶劣天气条件下。

(4) 图像质量和分辨率

由于可见光图像采用的是普通的照相机成像技术，分辨率通常较高，能够展现清晰的细节和丰富的颜色信息。高分辨率的可见光图像能够清楚地展示场景中的物体和纹理细节，因此，在物体识别和分类中具有较好的性能。红外图像的分辨率通常较低，且受到红外传感器性能的限制，图像的清晰度和细节较差。红外传感器的性能直接影响成像质量，低端传感器可能会产生较大的噪声，影响图像质量。因此，红外图像通常比可见光图像在细节表现上较为模糊，不易分辨细小的物体。

总之，红外图像与可见光图像各自有其独特的优势和适用领域。可见光图像在日常场景和物体识别中表现优异，但在低光和复杂环境中受限；而红外图像在低光、夜间以及恶劣环境下的表现更为出色，适用于热源检测和夜间监控等应用。

2.4 深度学习与卷积神经网络概述

随着数据量的不断增长和计算能力的提升，深度学习在人工智能领域实现了突破，特别是在图像处理、语音识别以及自然语言处理等关键领域。深度学习，作为一种以多层神经网络为基础的机器学习技术，能够自动地通过多层非线性变换提取数据的深层次特征。作为深度学习关键技术之一的卷积神经网络[33]，因其在图像处理方面的出色性能而备受瞩目。CNN 利用卷积操作在图像的局部区域提取特征，极大地提高了图像分类、检测和分割等任务的准确性和效率。

2.4.1 卷积神经网络的基本结构

卷积神经网络是一种特别适用于图像数据的深度学习模型，最早由 Yann LeCun 等人提出，并在手写数字识别中取得成功。CNN 的设计理念来源于生物视觉系统，主要依靠卷积操作和池化操作实现特征提取。CNN 的典型结构包括以下几层：

输入层：输入层是卷积神经网络的起点，负责接收原始数据。其主要功能是将原始数据转换为神经网络可以处理的格式，为后续的卷积、激活和池化操作做准备。

卷积层：卷积层[34-35]是 CNN 的核心层，通过卷积核（即滤波器）在图像的

局部区域内进行滑动计算，提取局部特征。卷积层的作用是学习图像中的边缘、纹理等低层次特征，随着网络的加深，可以提取更复杂的模式。卷积层的参数通过反向传播算法在训练中自动学习。常用的卷积操作有一维卷积和二维卷积。

一维卷积通常用于处理一维数据，如时间序列、音频信号、文本数据等。与二维卷积（用于处理图像）不同，一维卷积只在一个维度上进行卷积操作。假设输入信号为 X，卷积核为 W，输出为 Y，一维卷积公式如式(2-2)所示：

$$Y(t)=\sum_{i=1}^{K}W(i)\bullet X(t+i-1) \tag{2-2}$$

式(2-2)中，$Y(t)$ 表示输出序列 Y 在位置 t 的值；$X(t)$ 表示输入序列 X 在位置 t 的值；$W(i)$ 表示卷积核 W 在位置 i 的值，W 的长度为 K；K 表示卷积核 W 的长度；t 表示卷积操作的当前时间步（或位置）。

卷积核 W 在输入序列 X 上滑动，沿着输入数据进行计算；对于每一个输出 $Y(t)$，卷积核与输入数据序列的当前窗口（大小为 K）进行逐元素乘积，并求和，得到输出值；这个过程会为输入序列的每个位置生成一个新的输出值，从而构成整个输出序列 Y。

二维卷积通常用于图像等二维数据的处理。假设输入图像为 I，卷积核为 K，输出为 O，二维卷积公式如式(2-3)所示：

$$O(i,j)=\sum_{m=1}^{M}\sum_{n=1}^{N}K(m,n)\bullet I(i+m-1,j+n-1) \tag{2-3}$$

式(2-3)中，$I(i,j)$ 表示输入图像 I 在位置 (i,j) 的像素值；$K(m,n)$ 表示卷积核 K 在位置 (m,n) 的值；$O(i,j)$ 表示输出图像 O 在位置 (i,j) 的值；M 和 N 分别是卷积核的高度和宽度；i 和 j 表示输出图像 O 中的位置，通常表示该位置的行和列；m 和 n 表示卷积核 K 中的元素位置；$I(i+m-1,j+n-1)$ 是输入图像的局部区域（由卷积核大小决定）的像素值。

卷积核 K 在输入图像 I 上滑动，每次在当前位置 (i,j) 与图像的一块局部区域进行卷积运算；卷积核与图像的局部区域的每个元素逐一相乘，并将乘积结果求和。对于输入图像的每一个位置 (i,j)，计算并生成对应的输出值 $O(i,j)$，最终得到卷积后的输出图像。二维卷积过程如图 2-3 所示。

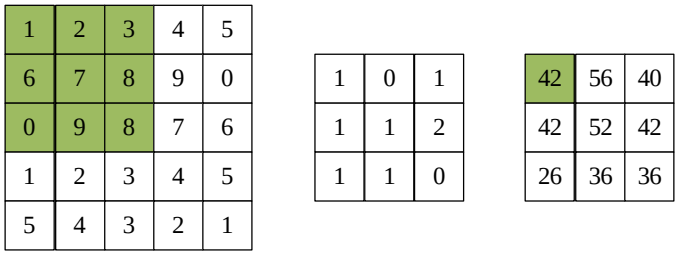


图 2-3 二维卷积过程

池化层：池化层[36]通常用于下采样，减少数据的维度，降低计算量，同时提高特征的平移不变性。常用的池化操作有最大池化和平均池化，如图 2-4 所示。最大池化会选择局部区域内的最大值，而平均池化会计算局部区域的平均值。池化层帮助模型在保留主要特征的同时，减少特征图的大小。

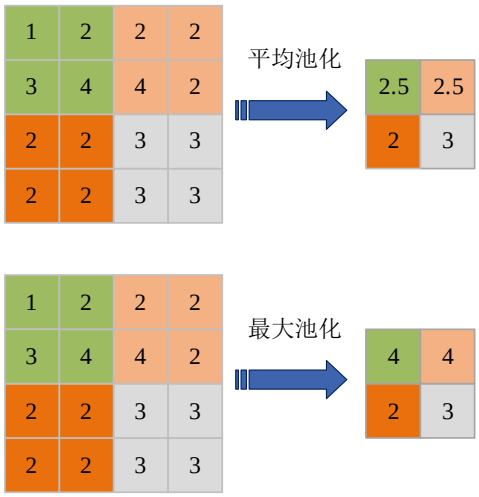


图 2-4 平均池化与最大池化

激活函数层：激活函数层通常位于卷积层之后，是用来引入非线性的关键组件。它的主要作用是对输入进行非线性变换，使得神经网络能够学习和拟合更复杂的数据模式，进而提升其表达能力。常用的激活函数包括 ReLU[37]、Sigmoid[38]和 Tanh[39]等。接下来将对这些函数做详细介绍。

(1) ReLU 函数：ReLU 是最常见的激活函数之一，公式如式(2-4)所示：

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

即输入大于零时输出原值，小于零时输出零。ReLU 非常简单，但效果非常好，因为它在正区间内不做任何变换，并且可以有效避免梯度消失问题，使得网络能够更深。函数图像如图 2-5 所示。

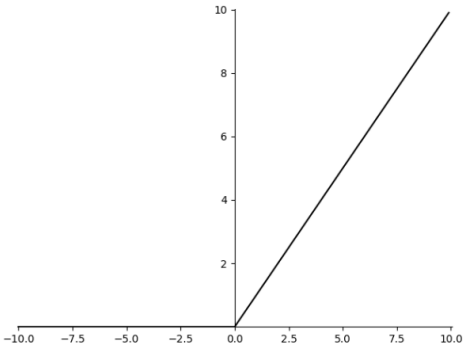


图 2-5 ReLU 函数

(2) Sigmoid 函数：Sigmoid 函数将输入映射到 0 和 1 之间，公式如式(2-5)所示：

$$f(x)=\frac{1}{1+e^{-x}} \tag{2-5}$$

它的优点是输出值在(0,1)区间内，适用于二分类问题。函数图像如图 2-6 所示。

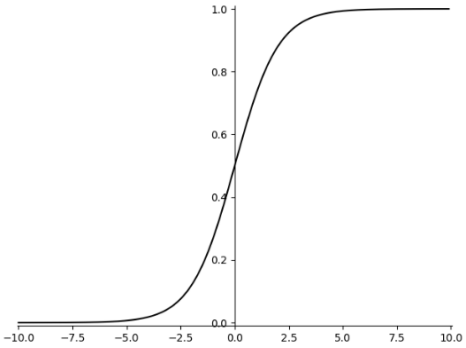


图 2-6 Sigmoid 函数

(3) Tanh 函数：Tanh 是另一个常见的激活函数，它将输入值映射到-1 到 1 之间，公式如式(2-6)所示：

$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2-6}$$

Tanh 比 Sigmoid 具有更好的性质，因为它的输出范围是-1 到 1，因此均值为零，避免了 Sigmoid 函数中的偏移问题。函数图像如图 2-7 所示。

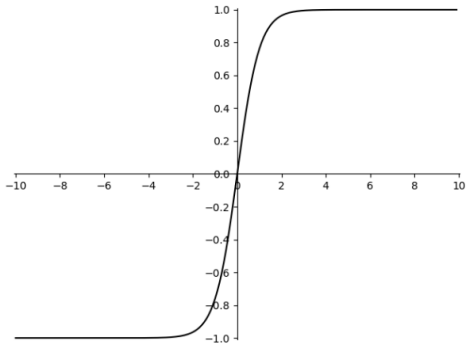


图 2-7 Tanh 函数

全连接层：全连接层[40]通常用于分类任务的输出阶段，负责将卷积层和池化层提取的特征进行展平处理，并通过一个或多个全连接层实现目标类别的预测。全连接层的节点与前一层的所有节点相连，学习数据的全局特征，如图 2-8 所示。

在全连接层中，每个神经元都与前一层的所有神经元相连接，这种密集的连接方式使得全连接层能够捕捉输入数据的全局特征。通过全连接层，模型可以将卷积层和池化层提取到的局部特征整合起来，形成对输入数据的全局理解。在分类任务中，全连接层的输出通常被用作分类器的输入，通过 softmax 等函数将输出转换为概率分布，从而实现对目标类别的预测。

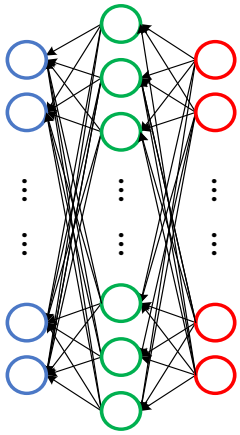


图 2-8 全连接层

2.4.2 卷积神经网络的特点

(1) 参数共享：卷积层的卷积核在整个图像上进行滑动，采用共享参数的方式，避免了为每个像素单独设置权重的需要。这种参数共享机制减少了 CNN 的参数数量，从而提高了模型的计算效率。同时，参数共享也使得 CNN 具有更强的泛化能力，因为它促使网络学习到图像中局部特征在不同位置的共享表示。这

一特性使得 CNN 对于图像的平移、旋转等变换表现出一定的鲁棒性。此外，卷积层通过卷积核的滑动操作，能够自动地提取图像中的层次化特征，从基础的边缘、纹理特征到高级的语义特征，为后续的分类或回归任务提供了丰富的特征表示。

(2) 局部连接：每个卷积核仅与输入图像的局部区域像素相连结，这使得网络能够专注于捕捉图像中的局部特征。局部连接的设计能够有效提取空间结构信息，同时减少模型参数的数量。与全连接网络相比，局部连接降低了模型的复杂性，提高了卷积神经网络处理高维数据的效率。此外，局部连接还使得卷积神经网络能够关注图像中的关键局部信息，这对于图像识别、目标检测等任务至关重要。通过组合多个卷积层，CNN 能够逐步构建出对图像整体结构的理解，从而实现对复杂图像任务的精确处理。

(3) 平移不变性：通过池化层和卷积核的滑动操作，CNN 能够掌握一定程度的平移不变性。即便目标在图像中发生轻微位移，模型依然能够进行有效识别。池化层通过对局部特征进行下采样，降低了数据的空间维度，同时保留了关键特征，提升了模型对输入图像中目标位置变化的适应能力。这种特性在处理图像时尤为重要，因为在实际应用中，目标物体可能出现在图像的任何位置。通过结合卷积层的特征提取能力和池化层的平移不变性，CNN 能够更有效地应对图像中的目标位置变化，提高了图像识别和目标检测的准确性。

2.5 评价指标

本文采用平均准确率 *AP*（Average Precision）、平均精度 *mAP*、模型参数量（Parameters）、浮点型计算量（GFLOPs）与每秒处理的帧数（FPS）作为评价指标。

平均准确率 *AP* 和平均精度 *mAP* 是衡量模型在分类任务中表现的重要指标。*AP* 反映了模型在不同阈值下的平均精度表现，其表达式如式(2-7)所示。而 *mAP* 则是对多个类别进行平均后的 *AP* 值，能够更全面地评估模型在多类别任务中的整体性能。*mAP* 数值越高，代表模型的性能越好，本文使用 IoU 阈值为 0.5 时 *mAP* 的值，其表达式如式(2-8)所示。由于本文只针对行人一个类别进行研究，因此 *AP* 值与 *mAP* 值相同。

$$AP = \int_0^1 PdR \tag{2-7}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{2-8}$$

式中：P 代表准确率；R 代表召回率；N 代表类别个数。

模型参数量（Parameters）是指模型中所有可训练参数的数量。参数量的多少直接影响模型的复杂度和计算资源的需求。较小的参数量有助于减少模型的存储空间和提高计算效率。

浮点型计算量（GFLOPs）是衡量模型计算复杂度的一个重要指标。它表示模型在进行一次前向传播时所需的浮点运算次数。GFLOPs 越高，模型的计算负担越重，对硬件的要求也越高。

每秒处理的帧数（FPS）是也是衡量模型实时性能的一个关键指标。FPS 越高，表示模型在单位时间内能够处理更多的帧数，从而具有更高的实时性。这对于需要快速响应的应用场景尤为重要。

通过综合这些评价指标，能够全面地评估改进后模型在准确率、复杂度、计算资源需求以及实时性能方面的表现，为进一步的优化和改进提供依据。

2.6 本章小结

本章首先对红外成像基本原理以及可见光图像与红外图像特性分析进行了介绍；其次，阐述了深度学习与卷积神经网络的相关内容；最后，对行人检测的评估指标进行简要说明。

第3章 基于改进 YOLOv5s 算法的红外图像行人检测

3.1 引言

行人检测要解决的问题是：识别图像或视频中的行人并准确地将其标注出来。例如，交通监控系统可以利用行人检测来监测交通流量和行人行为，以提供更安全和高效的交通环境。红外图像在夜间和低光条件下具有独特的优势，因此在安防监控、自动驾驶等领域具有广泛的应用前景。然而，现有的红外图像行人检测算法往往面临着检测困难、检测精度低以及目标丢失等问题，这主要是由于红外图像的特殊性质以及环境条件的限制所导致的。因此，对于红外图像行人检测技术仍然具有重要的研究意义。

3.2 YOLOv5s 算法介绍

YOLOv5[41]是由 Ultralytics 开发的目标检测模型，它是 YOLO 系列的一个重要改进。YOLOv5s 是 YOLOv5 的最小版本，专门优化了模型的体积和推理速度，它以体积小巧、推理速度快而著称，非常适合用于红外图像的行人检测。在实际应用中，YOLOv5s 通过优化网络结构和减少参数量，提升了检测速度，同时维持了较高的准确率。YOLOv5s 的网络结构由输入端、主干网络、颈部网络和检测端组成。输入端主要负责处理输入图像并将其转化为模型可以处理的格式。具体来说，YOLOv5s 输入端的作用是对输入的图像进行预处理，并根据模型的需求进行必要的调整。其主干网络采用 CSPDarknet53 结构，这是一种轻量级网络结构，它结合了跨阶段部分连接和残差连接的优点，能够有效地提取图像特征，并将这些特征应用于目标检测任务。CSPDarknet53 主干网络的引入，不仅降低了计算量，还增强了特征提取能力，确保了模型即使在复杂背景下也能准确识别行人。

颈部网络在特征融合阶段采用了 FPN+PAN[42-43]结构，目的是整合来自不同层级特征图的信息，以提高模型对各种尺度目标的检测准确性，有效解决了多尺度目标检测中的信息丢失问题，提升了模型对不同大小目标的检测性能。FPN（特征金字塔网络）通过自底向上的路径提升了特征图的空间分辨率，而 PAN（路径

聚合网络）则通过自顶向下的路径进一步整合了不同层次的特征，增强了模型对目标位置信息的敏感度。这种结构的设计使得 YOLOv5s 能够更好地捕捉不同尺度的目标特征，尤其是在处理小目标和遮挡目标时表现出色。通过 FPN+PAN 结构的结合，YOLOv5s 实现了对多尺度目标的精确检测，进一步提升了模型在实际应用中的性能。YOLOv5s 的检测端包含三个不同尺度的输出层，每个输出层都针对不同的目标尺度进行优化，通过综合不同层级的特征信息，确保了在各种场景下对行人的全面覆盖，特别是在小目标检测方面表现出色。其网络结构如图 3-1 所示。

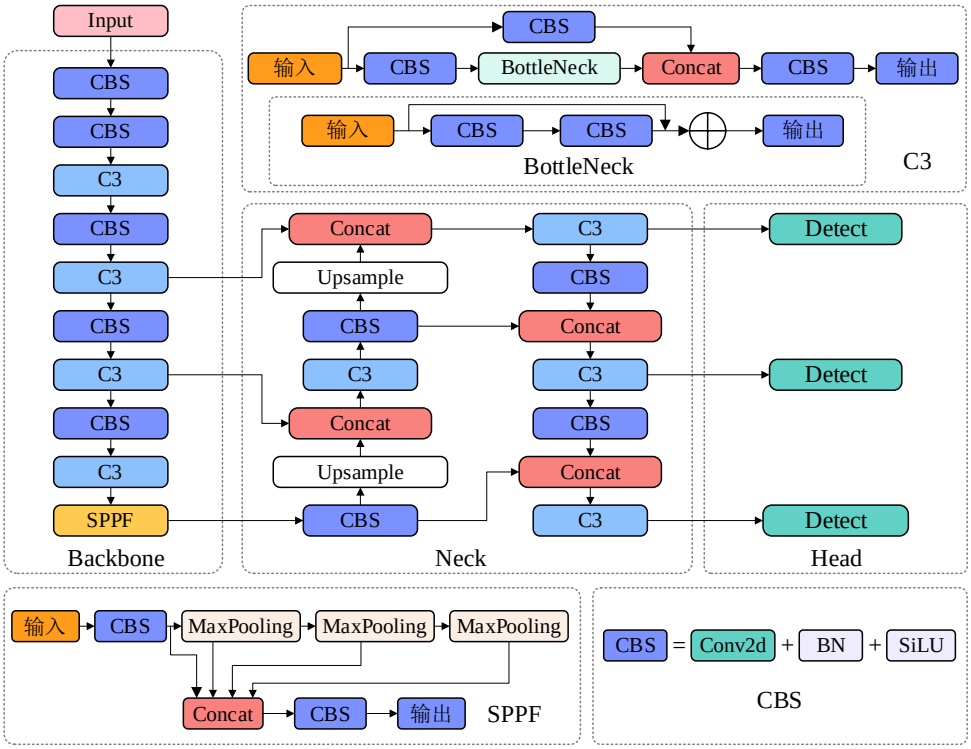


图 3-1 YOLOv5s 网络结构图

(1) Input 指的是输入端，它包括了固定的输入图像尺寸、数据预处理以及数据增强等关键步骤。这些设置对于模型而言至关重要，它们确保了模型能够高效地接收和处理输入数据，从而提高目标检测的准确性和效率。在 YOLOv5s 的输入阶段，采用固定尺寸的输入图像确保了模型能够以统一的标准处理数据，从而避免了因图像尺寸不一致而增加的处理复杂性和计算资源的额外开销。数据预处理则涵盖了图像的归一化、缩放等操作，这些步骤有助于加快模型的收敛速度并提高其稳定性。而数据增强技术，如随机裁剪、旋转、翻转等，能够增加训练数

据的多样性，有效防止模型过拟合，进一步提升模型的泛化能力。通过这些设计的输入端设置，YOLOv5s 得以在处理复杂多变的行人检测任务时展现出卓越的性能。

(2) Backbone 构成了网络结构的核心，主要负责提取输入图像的特征。它采用了 CSPDarknet53 作为其主干网络，通过一系列卷积层、池化层和残差块来深入挖掘图像中的特征信息。这些层不仅能够逐步减小特征图的尺寸，而且能够在不同层次提取出不同尺度的特征，这对于模型捕捉目标的各种尺度信息至关重要。

CBS 模块是特征提取的基本单元，它由二维卷积、批量归一化和激活函数构成。这些模块通过卷积操作提取图像特征、批量归一化加速训练过程、激活函数引入非线性，它们共同作用于提升模型的性能和准确性。C3 模块是一种特殊的卷积模块，它具备跨阶段部分连接和残差连接的特点，有助于提升特征提取的深度和网络的表达能力。C3 模块主要由 CBS 模块以及 Bottleneck 模块构成。C3 模块的设计增强了特征的表达能力，这对于目标检测任务至关重要，从而提升了模型的整体性能和效率。SPPF 模块是一种融合了空间金字塔池化和特征金字塔网络的结构，它能够提取不同尺度的特征并增强模型对不同大小目标的检测能力。这种设计有助于提升目标检测的精确度，使模型能够更好地适应各种尺度的目标。CSPDarknet53 的设计来源于 CSPNet，它通过引入跨阶段局部密集连接来增强特征学习能力，同时减少了计算量和内存占用。这种结构使得 Backbone 在处理大规模图像数据时更加高效，能够在保持高性能的同时降低硬件资源的需求。此外，CSPDarknet53 还融合了残差学习的思想，通过引入跳跃连接来缓解深层网络中的梯度消失问题，从而提升了网络的训练稳定性和收敛速度。这些特性共同构成了 YOLOv5s 强大特征提取能力的基础，为后续的目标检测任务提供了坚实保障。

(3) Neck 部分专注于特征融合，采用了 FPN+PAN 结构。通过路径聚合、特征融合和上下文信息提取等操作，Neck 部分帮助提升模型对目标的检测能力。Neck 部分的设计能够有效地整合主干网络提取的特征，并为检测头部提供更丰富信息和更高语义理解能力的特征表示，从而提高目标检测的准确性。FPN（Feature Pyramid Network）通过自顶向下的方式提升了高层特征的空间分辨率，而 PAN（Path Aggregation Network）则通过自底向上的路径进一步整合了不同尺

度的特征，实现了特征的双向流动。这种结构不仅增强了特征的丰富性和多样性，而且有助于模型更深入地理解目标的上下文信息，从而在复杂场景中实现精确的目标检测。Neck 部分的设计充分体现了 YOLOv5s 在网络结构上的创新和优化，为提升目标检测性能奠定了坚实基础。

(4) YOLOv5s 网络的 Head 阶段是其最终阶段，负责从主干网络和 Neck 部分提取的特征中生成目标检测的预测结果。该部分主要由三个不同尺度的输出层组成，检测不同尺寸的目标。每个输出层都针对特定尺寸范围的目标进行优化，通过应用适当的锚框和预测框，能够精确地定位和分类各种尺寸的目标。Head 部分还结合了分类和回归任务，同时输出目标的类别信息和边界框坐标，实现了端到端的目标检测。这种设计使得 YOLOv5s 在处理多尺度目标时具有出色的性能和鲁棒性，能够适应复杂多变的检测场景。

3.3 YOLOv5s 算法改进

为了解决当前红外图像行人检测算法在检测速度与精度之间难以取得平衡的问题，从而导致检测困难、精度低下、目标丢失等问题，提出一种改进 YOLOv5s 的轻量化红外图像行人检测算法 YOLOv5s-CSBS，通过改进 YOLOv5s 算法的 Backbone 网络和 Neck 网络，提升了红外图像中行人检测的精度。其网络结构图如图 3-2 所示。

(1) 建立本章使用的红外图像行人数据集。从红外图像数据集 KAIST 中筛选出 4000 张夜间红外图像作为本章使用的实验数据集。

(2) 改进 Backbone 网络。将 Backbone 网络中的 C3 模块全部替换成 C3CA 模块，以增强对行人的长距离检测能力。这一改进不仅提高了模型对远处行人的检测精度，还提升了其在复杂场景下的适应性。同时，在 Backbone 网络中的 SPPF 下一层引入了 SimAM 注意力机制模块，以提升模型对行人特征的关注度，并增强对不同深度特征信息的感知能力。SimAM 注意力机制能够自适应地学习并突出图像中的关键特征区域，同时抑制不相关的背景信息，这使得模型在复杂的红外图像环境中能更加专注于行人目标，提升了 YOLOv5s-CSBS 算法在行人检测任务中的准确性和鲁棒性。

(3) 改进 Neck 网络。在 Neck 网络中引入加权双向特征金字塔网络

(Bidirectional Feature Pyramid Network, BiFPN)，BiFPN 的引入，不仅加强了低层与高层特征间的信息交流，还促进了不同尺度特征的有效整合。这种设计使得模型能够更好地捕捉到行人的多尺度特征，提升检测效果。最后，引入包含 GSConv 的 Slim-Neck 设计范式降低模型复杂性，更好地平衡模型准确性和速度。

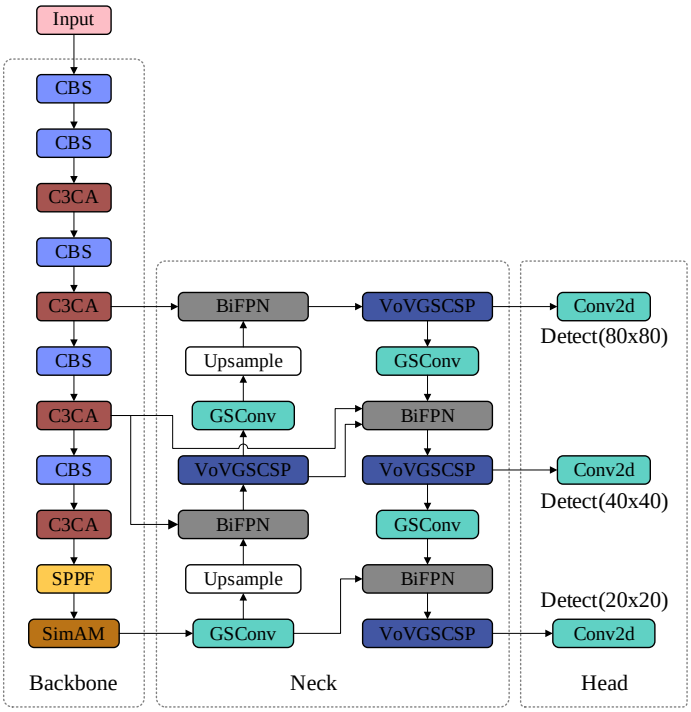


图 3-2 YOLOv5s-CSBS 网络结构图

3.3.1 CA 注意力机制模块

近年来，注意力机制已成为深度学习领域的研究焦点，它能够助力模型更精准地聚焦于关键特征，进而提高模型对不同尺度目标的敏感性和检测准确性。诸如 SE[44]、ECA[45]、CBAM[46]和 CA[47]等注意力机制，已成为当前广泛采纳的注意力机制。这些注意力机制通过各自独特的方式强化了模型对关键特征的识别。SE 是一种通道注意力机制，如图 3-3 所示，它通过学习通道间的依赖关系来强化重要特征，专注于通道维度的相互作用，但忽略了空间维度的特征。

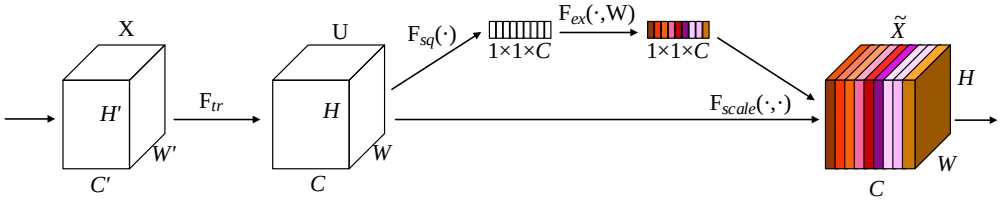


图 3-3 SE 通道注意力机制

ECA 在 SE 的基础上进行了优化，它利用局部跨通道的交互信息来增强特征表达，并移除了全连接层，通过在全局平均池化后的特征上应用 1×1 卷积，使其在性能提升的同时计算更高效。然而，ECA 主要侧重于通道信息，相对轻视了空间信息。

CBAM 则结合了空间和通道维度的注意力，引入了大尺度卷积核以捕捉空间特征，但其仅能捕捉局部相关性，忽略了远距离依赖问题，其原理图如 3-4 所示。

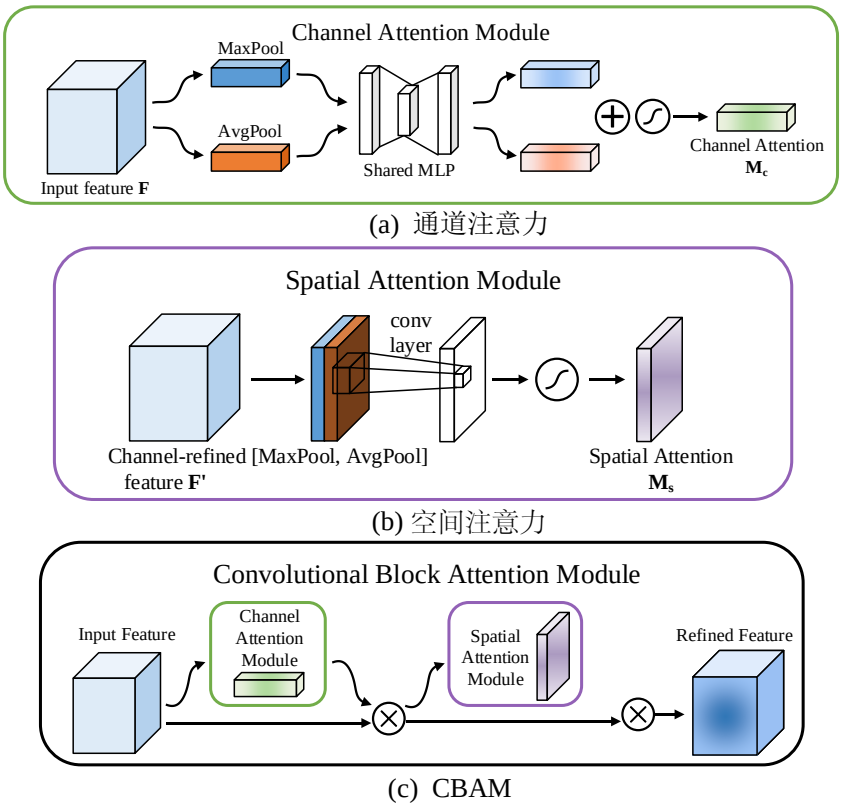


图 3-4 CBAM 注意力机制

CA 注意力机制则通过自适应地调整特征权重来优化特征表示，它不仅能够获取方向和位置的感知信息，还能捕获通道间的相互作用，有助于模型更准确地定位和识别目标。CA 注意力机制模块如图 3-5 所示。

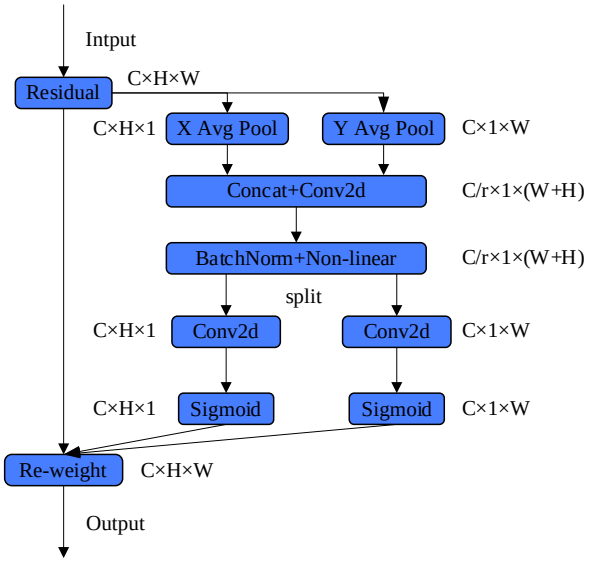


图 3-5 CA 注意力机制模块

图 3-5 中：C 为特征图通道数；H 代表输入特征图的高度；W 代表输入特征图的宽度；r 为通道压缩率。

图 3-5 中，首先，将输入尺寸为 C×H×W 的特征图在宽度和高度两个方向上进行分割。然后分别对每个通道执行沿宽度方向的全局平均池化与沿高度方向的全局平均池化，最后分别会得到一个尺寸大小为 C×H×1 的特征图与一个尺寸大小为 C×1×W 的特征图。公式分别如式(3-1)、(3-2)所示：

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i\le W}x_c(h,i) \tag{3-1}$$

$$Z_c^w(w)=\frac{1}{H}\sum_{0\le j\le H}x_c(j,w) \tag{3-2}$$

式(3-1)中：$Z_c^h(h)$ 代表第 c 通道在高度 h 处的输出；$x_c(h,i)$ 代表水平方向第 c 通道的输入。

式(3-2)中：$Z_c^w(w)$ 代表第 c 通道在宽度 w 处的输出；$x_c(j,w)$ 代表垂直方向第 c 通道的输入。

将得到的两个特征图进行拼接操作，得到尺寸大小为C/r×1×(W+H)的特征图，接着利用卷积、批量归一化与非线性激活函数获得特征图，公式如(3-3)所示：

$$f=\delta\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{3-3}$$

式(3-3)中：F_1 代表一个共享的 1×1 卷积转换函数。

沿着空间维度，再将特征图 f 进行 split 操作，获得两个尺寸大小分别为

C×H×1，C×1×W 的特征图，然后分别利用 1×1 卷积进行升维操作，获得了与原始特征图具有相同通道数的特征图 F_h 与 F_w，再利用激活函数分别计算得到高度方向和宽度方向上的注意力权重 g_h 与 g_w，公式分别如式(3-4)、(3-5)所示：

$$g_h = \sigma\left(F_h\left(f^h\right)\right) \tag{3-4}$$

$$g_w = \sigma\left(F_w\left(f^w\right)\right) \tag{3-5}$$

式(3-4)与式(3-5)中：σ 表示 Sigmoid 激活函数。

最后，将得到的注意力权重与原始特征图进行乘法加权计算，来增强特征图的表示能力，公式如(3-6)所示：

$$y_c(i,j) = x_c(i,j) \bullet g_c^h(i) \bullet g_c^w(j) \tag{3-6}$$

3.3.2 SimAM 注意力机制模块

为了更好地准确捕捉到红外图像中的行人信息，有效提升模型性能，本章引入SimAM[48]注意力机制，在YOLOv5s主干网络的SPPF下方添加SimAM注意力机制，以增强对多个深度特征的感知能力。SimAM注意力机制结合了通道注意力机制和空间注意力机制的优势，能够在不增加过多计算负担的前提下，提升模型对关键信息的捕捉能力。在YOLOv5s的主干网络中嵌入SimAM注意力机制后，模型在处理复杂场景下的红外图像时，能够展现出更强的特征提取和目标识别能力。

SimAM 是一种三维无参数注意力机制，该注意力机制的提出来源于神经科学空域抑制理论。在神经科学领域，具有空域抑制效应的神经元往往被认为更加重要，通过检测目标神经元与其他神经元之间的线性可分性可以来辨别这些神经元。因此，对每个神经元定义一个能量函数，通过使用二值标签并添加正则化的方式，最终简化得最小能量公式如式(3-7)所示：

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{\left(t - \hat{u}\right)^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{3-7}$$

式(3-7)中：t 为目标神经元；λ 为超参数；$\hat{u}$ 代表同一通道内其他神经元的平均值，公式如式(3-8)所示；$\hat{\sigma}^2$ 代表同一通道内其他神经元的方差，公式如式(3-9)所示。

$$\hat{u}=\frac{1}{M}\sum_{i=1}^{M}x_i \tag{3-8}$$

$$\hat{\sigma}^2=\frac{1}{M}\sum_{i=1}^{M}\left(x_i-\hat{u}\right)^2 \tag{3-9}$$

式中：M 为神经元个数；x_i 为同一通道内其他神经元个数，i 为空间维度指数。

分析式(3-7)可知，能量函数 e_t^* 的值越低，表示目标神经元与周围神经元的差异性越明显，其重要性也越高。

然后，对特征 X 进行增强处理，得到增强后的特征 $\tilde{X}$ ，公式如式(3-10)所示：

$$\tilde{X}=sigmoid\left(\frac{1}{E}\right)\odot X \tag{3-10}$$

式(3-10)中：E 为所有通道和空间维度中最小能量集合；X 为输入特征；$\odot$ 为点积运算；Sigmoid函数用于对特征进行增强处理。

SimAM注意力机制如图3-6所示。

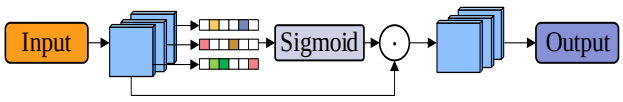
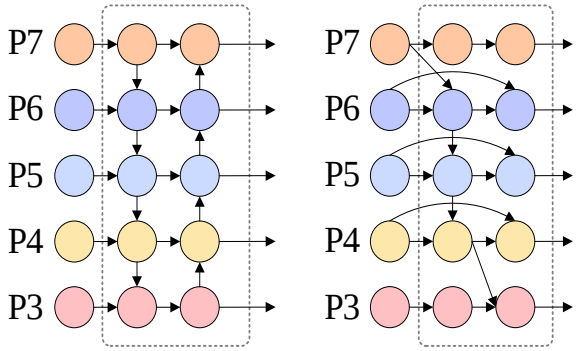


图 3-6 SimAM 注意力机制模块

3.3.3 加权双向特征金字塔网络

YOLOv5s模型在Neck部分采用PANet结构，其网络结构图如图3-7（a）所示，该结构通过金字塔形式连接不同尺度的特征图，实现了高级与低级特征信息的融合。但PANet结构仅包含一条自上而下和一条自下而上的路径，在特征提取过程中可能会导致部分原始信息的丢失，从而出现漏检与误检的情况，影响检测行人的准确性。为了解决这个问题，加权双向特征金字塔网络(Bidirectional Feature Pyramid Network, BiFPN)应运而生。BiFPN在PANet的基础上进行了创新性改进，不仅保留了原有的自上而下和自下而上的路径，还引入了加权机制，使得不同路径上的特征信息得以更高效地融合。这种加权机制可以根据特征的重要性动态调整权重，从而保留了更多的原始信息，减少了漏检和误检现象，提升了行人检测的准确性。BiFPN的网络结构图如图3-7（b）所示，通过对比可以看出，BiFPN在特征融合方面更加灵活和高效。

通过引入加权双向特征金字塔实现更高效的多尺度特征融合。首先，由于只有一条输入边的节点在特征融合的过程中贡献比较小，所以为了简化双向网络，同时降低计算开销，将此类节点去除。其次，当原始输入节点与输出节点处在相同的层级时，会在它们之间建立一条融合路径，从而融合更多特征。最后，通过构造双向通道，通过自顶向下和自底向上的结构在不同尺度之间建立连接，将每个双向路径视为网络中的一个特征层，并多次重复相同的层，以实现更高级别的特征整合，BiFPN通过上下两个方向的信息传递，加强了特征金字塔网络的特征提取能力，提高了对行人检测的准确率。



(a) PANet　　(b) BiFPN

图 3-7 PANet、BiFPN 网络结构图

3.3.4 Slim-Neck 设计范式

鉴于行人检测算法模型的复杂性，研究人员通常会采用深度可分离卷积层来构建轻量级模型，来减少参数数量和降低计算复杂度。然而，这种简单的替换往往会导致模型性能的下降。为了在模型轻量化和性能之间取得平衡，本章引入包含GSConv[49]的Slim-Neck设计范式。Slim-Neck设计范式采用了融合通道注意力机制和特征融合策略，以强化模型的特征表达能力。通过引入通道注意力机制，模型能够自适应地调整各个特征通道的重要性，从而增强对关键特征的敏感性。同时，特征融合策略使得模型能够更好地融合不同尺度和层次的特征，进一步提高检测的准确性和鲁棒性。

本章引入GSConv卷积来代替Neck部分中的CBS模块，以在降低模型复杂性的同时，也能够保证模型检测的精度。GSConv卷积结构如图3-8所示。

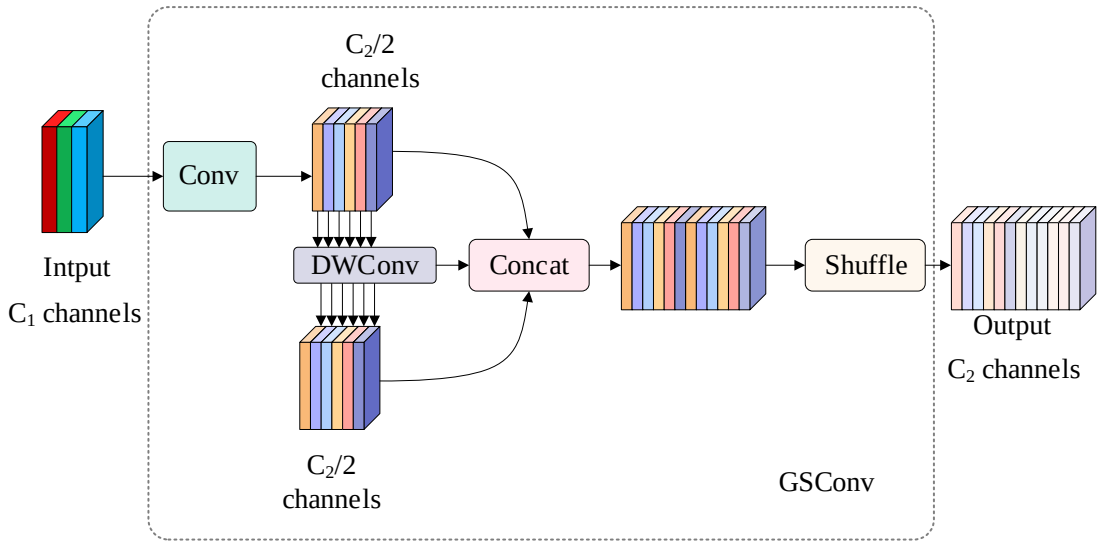


图3-8 GSConv卷积结构图

首先，对输入特征图进行标准卷积操作，然后将其与深度可分离卷积操作得到的特征图进行拼接和混合操作，最终得到输出的特征图。这种设计不仅能够充分利用标准卷积的特征提取能力，还能借助深度可分离卷积的低计算复杂度优势。通过拼接和混合两种卷积操作得到的特征图，GSConv卷积能够在不增加过多计算负担的情况下，融合多种特征信息，从而提升模型的检测精度。此外，GSConv卷积的引入还使得Slim-Neck设计范式在保持轻量级的同时，进一步增强了模型的泛化能力，使其在不同场景和光照条件下都能表现出良好的检测性能。

Slim-Neck设计方式范式是在GSConv的基础上通过引入GS bottleneck与VoVGSCSP而提出的一种轻量化结构。GS bottleneck模块由两个GSConv层组成，其结构如图3-9（a）所示。VoVGSCSP模块是在GS bottleneck基础上使用一次性聚合方法设计的跨阶段部分网络模块，在保持足够的准确性的同时，简化了模型的复杂性，其结构如图3-9（b）所示。

具体来说，GS bottleneck模块通过堆叠两个GSConv层，进一步增强了特征的提取和融合能力。第一个GSConv层主要负责初步的特征提取，而第二个GSConv层则对初步提取的特征进行进一步的细化和增强。这种堆叠设计不仅提高了特征的丰富性，还有助于模型更好地学习到不同层次的特征信息。

而VoVGSCSP模块则采用了跨阶段的部分网络设计，通过一次性聚合方法将不同阶段的特征进行有效地融合。这种设计不仅简化了模型的复杂性，还保持了足够的准确性。在VoVGSCSP模块中，不同阶段的特征被看作是不同的特征源，通过一次性聚合方法将这些特征源进行融合，可以得到更为全面和准确的特征表

示。这种特征表示有助于模型在不同场景和光照条件下都能表现出良好的检测性能。

总的来说，Slim-Neck设计范式通过引入GS bottleneck与VoVGSCSP模块，成功地实现了一种轻量化的结构设计。这种设计不仅降低了模型的复杂性，还保证了模型的检测精度和泛化能力。

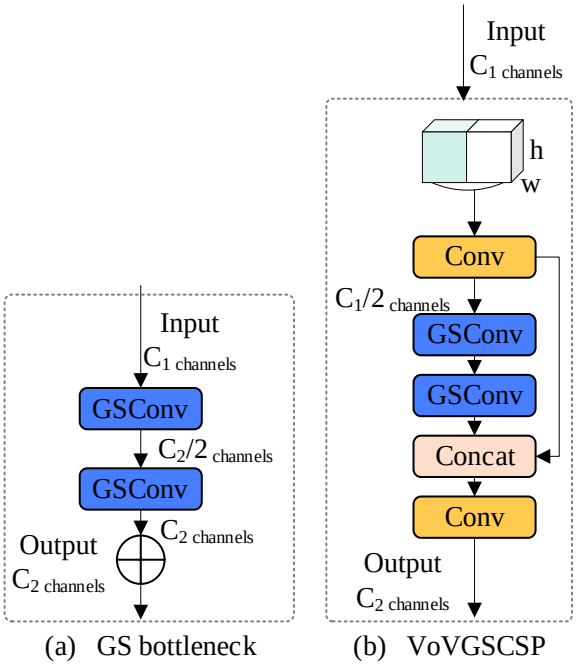


图3-9 GS bottleneck和VoVGSCSP网络结构

3.4 实验结果与分析

3.4.1 实验环境

实验环境采用Windows10操作系统，使用PyTorch1.9深度学习框架，利用GPU加速软件进行模型训练，实验配置环境如表3-1所示：

表 3-1 实验配置环境

项目	版本环境
操作系统	Windows10
GPU	NVIDIA TITAN X
PyTorch	1.9
Python	3.7
CUDA	11.1

实验超参数设置如下表3-2所列：

表 3-2 超参数设置

超参数	值
输入图像尺寸	640×640
batch-size	8
优化器	SGD
初始化学习率	0.01
动量	0.937
epoch	200

为了验证本章提出改进方法的有效性，分别对 YOLOv5s 主干网络和颈部网络进行改进，通过实验对比各种改进方法的效果。实验结果最优数据用加粗字体标注，次优数据用下划线标注。

3.4.2 数据集制作

本章使用 KAIST 红外图像数据集，该数据集涵盖了多种复杂场景和行人姿态，为算法的训练和测试提供了丰富的样本，数据集的标签中包含 person、people 和 cyclist 三个类别，如图 3-10 所示。

图 3-10 KAIST 数据集中部分图像

首先，从红外图像数据集 KAIST 中筛选出 4000 张包含行人的夜间红外图像作为本章使用的实验数据集。针对红外图像的特点，对数据集进行预处理，采用 CLAHE 算法对图像进行增强处理，以提高算法对红外图像的适应性和鲁棒性。CLAHE 算法的应用提升了图像的对比度，使行人的轮廓更加清晰，特征更加鲜明，为后续的行人检测提供了更为坚实的基础。图像增强前后的对比图如图 3-11 所示。

（a）原始图像

（b）CLAHE 图像

图 3-11 图像增强前后的对比图

然后，对预处理后的数据集使用标注软件 LabelImg 软件进行人工标注，如图 3-12 所示，标注标签为 person。标注过程确保了每张图像中的行人都能被准确识别和定位，为后续的算法训练和测试提供了精确的数据支持。

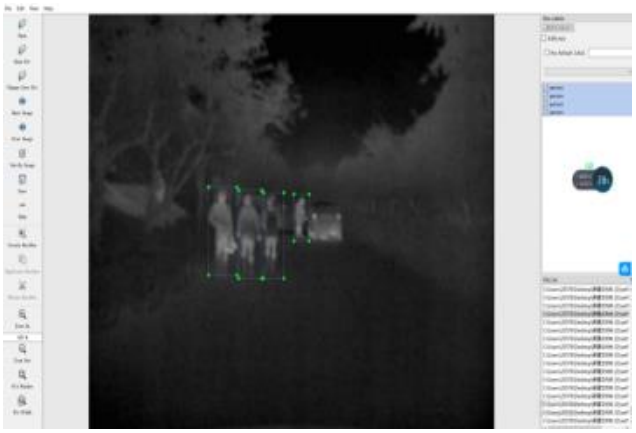


图 3-12 LabelImg 软件标注

最后，在完成标注后，对数据集按照 7:2:1 的比例将其分为训练集、验证集和测试集，以确保算法在训练过程中能够充分吸收数据特征，同时在验证和测试阶段能够准确评估其性能。这一系列的预处理和标注工作，为 YOLOv5s-CSBS 算法在红外图像行人检测领域的应用提供了重要的支持。

3.4.3 注意力机制实验

为了验证CA注意力机制在红外图像行人检测中的优势，在本课题建立的数据集上对比了SE、ECA、CBAM和CA四种注意力机制分别添加到主干网络的C3模块的实验结果，实验结果见表3-3。

表 3-3 注意力机制对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv5s	89.9	**7.01**	**15.8**
YOLOv5s+C3SE	90.4	7.32	17.0
YOLOv5s+C3ECA	91.1	7.31	17.0
YOLOv5s+C3CBAM	90.2	7.33	17.1
YOLOv5s+C3CA（Ours）	**91.6**	7.33	17.1

由表3-3可知，在基准模型YOLOv5s的基础上，分别将SE、ECA、CBAM和CA四种注意力机制融入C3模块，其参数量、计算量与mAP@0.5值均有不同程度的增加。其中，将CA注意力机制融入C3模块时，其mAP@0.5的值提升最大，提升1.7%，SE、ECA、CBAM三种注意力机制分别提升0.5%、1.2%、0.3%，验证了在红外图像行人检测中，将CA注意力机制融入C3模块的优势。虽然参数量与计算量有所增加，但考虑到检测精度的提升，这种增加是可以接受的。注意力机制对比图如图3-13所示。

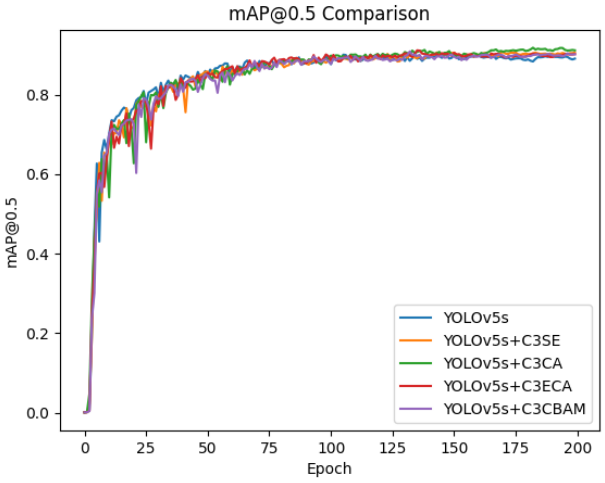


图3-13 注意力机制对比图

3.4.4 消融实验

为验证本章提出的四种改进方法的有效性，在本课题建立的数据集上进行了8 组不同的消融实验，以确认它们的影响。消融实验结果如表 3-4 所示。

表 3-4 消融实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs	FPS
YOLOv5s	89.9	7.01	15.8	85
YOLOv5s+C3CA	91.6	7.33	17.1	61
YOLOv5s+SimAM	90.1	7.01	15.8	83
YOLOv5s+BiFPN	91.0	7.16	16.4	81
YOLOv5s+Slim-Neck	90.3	6.47	13.1	**87**
YOLOv5s+C3CA+BiFPN	91.2	7.48	17.7	58
YOLOv5s+C3CA+BiFPN+Slim-Neck	90.5	**6.26**	**12.9**	65
YOLOv5s-CSBS(Ours)	**91.9**	**6.26**	**12.9**	70

由消融实验结果可以看出，以 YOLOv5s 为基准模型分别引入 C3CA、SimAM、BiFPN、Slim-Neck 四种改进方法，模型 mAP@0.5 值分别提升 1.7%、0.2%、1.1%、0.4%，验证了四种改进方法对行人检测的有效性。

当以 YOLOv5s 为基准模型同时引入 C3CA 和 BiFPN 时，mAP@0.5 值虽然提升 1.3%，但模型的参数量与计算量分别增加了 6.7%和 12.0%，没有实现模型的轻量化设计。

为了进一步实现模型轻量化，以 YOLOv5s 为基准模型同时引入 C3CA、BiFPN 和 Slim-Neck，模型的参数量与计算量分别降低了 10.7%和 18.4%，但 mAP@0.5 值只提升了 0.6%。

因此，本研究将四种改进方法融合，提出了一种轻量级红外图像行人检测算法模型 YOLOv5s-CSBS，该模型在提高检测精度的同时实现了轻量化设计，最终改进后的模型在 mAP@0.5 方面比 YOLOv5s 模型提高了 2%，同时参数量减少了 10.7%，计算量减少了 18.4%。尽管 FPS 略有下降，但仍然满足实时性要求。改进算法前后对比图如图 3-14 所示。

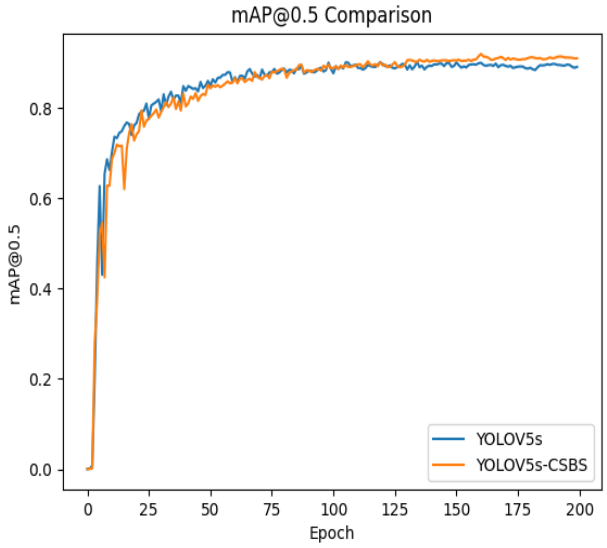


图 3-14 改进算法前后对比图

3.4.5 对比实验

(1) 本课题红外图像行人数据集对比实验

在实验环境及训练参数一致的情况下，在本课题建立的数据集上，对本章改进的算法和 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 等主流算法进行了训练和测试，以验证本章提出的模型的有效性，具体实验结果见表 3-5。

表 3-5 对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3	84.5	61.50	154.5
YOLOv5s	89.9	7.02	15.8
YOLOv8s	89.5	11.13	28.4
YOLOv10s	90.4	8.07	24.8
YOLOv11s	90.1	9.46	21.7
YOLOv5s-CSBS(Ours)	**91.9**	**6.26**	**12.9**

从表 3-5 可得，在数据集、实验环境及训练参数一致的情况下，本章改进的模型 YOLOv5s-CSBS 与 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 算法相比，mAP@0.5 的值分别高出 7.4%、2%、2.4%、1.5%和 1.8%，参数量与计算量都少于 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s，验证了本章提出模型的有效性。

(2) FLIR 公共红外图像行人数据集对比实验

为了进一步验证本章方法的优越性，在公共红外图像数据集 FLIR 上筛选出

包含行人的红外图像构成 FLIR 红外图像行人数据集进行对比实验。FLIR 公共红外图像数据集由 FLIR Systems 公司发布，为红外图像处理领域的研究和开发提供高质量的数据资源。该数据集包含大量红外图像，涵盖各种场景和目标，可用于目标检测、分类、跟踪和分割等任务，部分图像如图 3-15 所示。

图 3-15 FLIR 数据集中部分图像

FLIR 数据集对比实验结果见表 3-6。

表 3-6 FLIR 对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3	85.6	61.50	154.5
YOLOv5s	90.1	7.02	15.8
YOLOv8s	89.7	11.13	28.4
YOLOv10s	90.2	8.07	24.8
YOLOv11s	89.8	9.46	21.7
文献[50]	90.3	-	-
YOLOv5s-CSBS(Ours)	**91.7**	**6.26**	**12.9**

由表 3-6 实验结果可得，本章改进的模型 YOLOv5s-CSBS 在 FLIR 数据集上同样取得了不错的检测效果。从 mAP@0.5 上看，YOLOv5s-CSBS 的实验结果为 91.7%，比 YOLOv5s 高出 1.6%，比文献[50]高出 1.4%。综合各项评价指标可得，本章提出的算法模型 YOLOv5s-CSBS 在轻量化的同时提高了检测精度，满足检测红外图像行人的要求。

(3) 混合数据集对比实验

从 KAIST 数据集与 FLIR 数据集中分别挑选出 2000 张部分包含行人的红外图像组成混合数据集，来进一步验证本章所提算法的有效性。混合数据集中部分图像如图 3-16 所示。

(a) FLIR数据集图像　　(b) KAIST数据集图像

图 3-16 混合数据集中部分图像

混合数据集对比实验结果见表 3-7。

表 3-7 混合数据集对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3	86.7	61.50	154.5
YOLOv5s	90.6	7.02	15.8
YOLOv8s	90.2	11.13	28.4
YOLOv10s	90.5	8.07	24.8
YOLOv11s	90.3	9.46	21.7
YOLOv5s-CSBS(Ours)	**92.1**	**6.26**	**12.9**

由表 3-7 实验结果可得，本章改进的模型 YOLOv5s-CSBS 在混合数据集上也取得了不错的效果。同时可知，用混合数据集检测精度比单个数据集的检测精度要高。从 mAP@0.5 上看，YOLOv5s-CSBS 的实验结果为 92.1%，分别比 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 提高了 5.4%、1.5%、1.9%、1.6%、1.8%。

(4) 可见光数据集对比实验

上述对比实验均建立在夜间红外图像的基础上，实验结果也验证了本章提出的算法模型在训练红外数据集时表现出良好的性能。KAIST 红外图像数据集中每张图片都包含 RGB 彩色图像和红外图像两个版本，从其中挑选出 4000 张可见光图像构成可见光数据集来进一步验证本章算法的效果。可见光数据集中部分图像如图 3-17 所示。

图 3-17 可见光数据集中部分图像

可见光数据集对比实验结果见表 3-8。

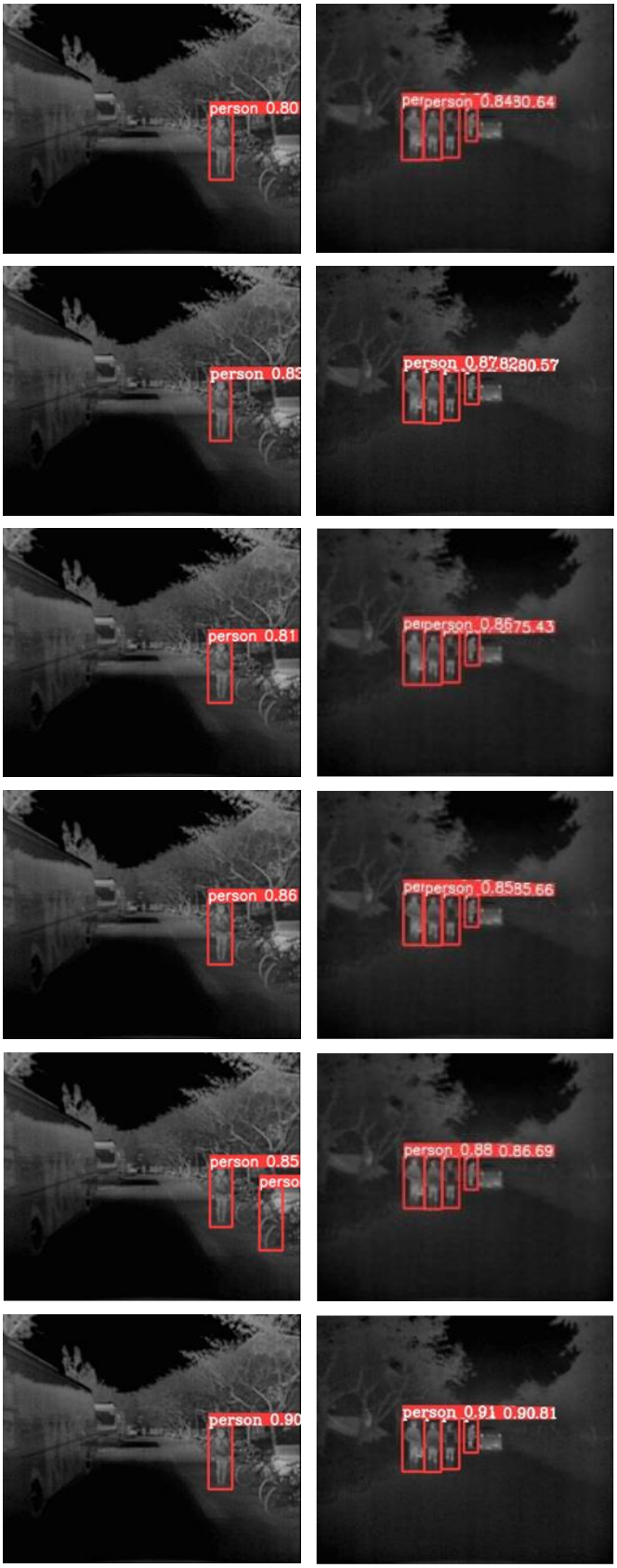
表 3-8 可见光数据集对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3	75.3	61.50	154.5
YOLOv5s	88.3	7.02	15.8
YOLOv8s	88.5	11.13	28.4
YOLOv10s	88.8	8.07	24.8
YOLOv11s	89.2	9.46	21.7
YOLOv5s-CSBS(Ours)	**90.1**	**6.26**	**12.9**

由表 3-8 实验结果可以得到，本章改进的模型 YOLOv5s-CSBS 在可见光数据集上也同样适用。从 mAP@0.5 上看，YOLOv5s-CSBS 的实验结果为 90.1%，分别比 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 提高了 14.8%、1.8%、1.6%、1.3%、0.9%。

3.4.6 检测结果可视化

为了更直观地比较本章提出的算法 YOLOv5s-CSBS 与 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 算法对红外图像行人的检测效果，从测试集中挑选出不同场景下的行人进行检测。图 3-18 从上到下分别展示了 YOLOv3、YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s、YOLOv5s-CSBS 对于同一场景的红外图像行人检测结果。



(a) 场景一　　(b) 场景二

图 3-18 不同算法检测结果

通过对比图 3-18 可知，本章所提出的算法模型 YOLOv5s-CSBS 相比 YOLOv5s 算法模型检测精度有所提升，验证了本章所提出的算法对红外图像行人检测的有效性。通过上述实验结果表明，本章所提出的算法模型 YOLOv5s-CSBS 相比 YOLOv5s 算法模型在检测精度上有所提高，同时解决了漏检的问题，证明了改进模型的优越性。

3.5 本章小结

本章提出一种基于 YOLOv5s 改进的轻量化红外图像行人检测算法 YOLOv5s-CSBS，解决了 YOLOv5s 在红外图像行人检测中存在的问题。首先，通过融合 CA 注意力机制，增加对行人的长距离检测能力。其次，通过引入 SimAM 注意力机制，进一步提高模型对行人检测精度。再次，基于加权双向特征金字塔网络重新构建了颈部网络，增强网络特征表达能力。最后，引入包含 GSConv 的 Slim-Neck 设计范式降低模型复杂性，进一步实现轻量化设计。实验结果表明：YOLOv5s-CSBS 在实验建立的红外图像行人数据集上，mAP@0.5 值比 YOLOv5s 提高 2%，同时参数量与计算量分别降低了 10.7%和 18.4%，验证了改进算法在增强红外图像行人检测精度的同时，实现模型的轻量化设计。

第 4 章 基于改进 YOLOv10n 算法的红外图像行人遮挡检测

4.1 引言

通过第 3 章提出的轻量化红外图像行人检测算法 YOLOv5s-CSBS，有效解决了 YOLOv5s 在红外图像行人检测中存在的误检、漏检等问题。然而，该算法只是对无遮挡的行人有较好的检测效果，在处理红外图像行人遮挡问题时仍存在一些不足之处。本章节针对红外图像中行人目标因部分或完全遮挡导致的检测性能下降问题，通过改进 YOLOv10n 网络架构，使其有效提升复杂遮挡环境下行人目标的检测精度与鲁棒性。

通常情况下，在人流量较大的公共场所，监控视频所记录的行人大部分都存在不同程度的遮挡问题，这导致了行人漏检的情况频繁发生，从而严重影响了行人检测的精度和准确性。随着城市人口的快速增长，夜晚的犯罪率也在不断上升。在这种情况下，通过在大型商场、游乐园、火车站等公共场所安装监控设备，可以实时监控人员的流动情况，从而有效预防和减少安全隐患问题的发生。因此，对于红外图像行人遮挡问题的研究具有极其重要的意义。

4.2 YOLOv10n 算法介绍

YOLOv10[51]算法是 YOLO 系列目标检测算法中的又一力作，共涵盖了 n、s、m、b、l、x 六大模型变体，这些变体各自针对不同的应用场景进行目标检测。鉴于红外图像行人遮挡问题对模型高效性与轻量化的需求，本章选取 YOLOv10n 作为基准模型，解决红外图像行人遮挡问题。该模型的网络架构如图 4-1 所示。图 1 中的网络架构图详细展示了 YOLOv10n 模型的各个层次和连接方式，从输入层到输出层，每一层都有其特定的功能和作用。YOLOv10n 算法由 Input，Backbone，Neck 和 Head 共 4 个部分组成。Input 为输入端，它主要负责处理输入图像的各种操作。这些操作包括但不限于调整输入图像的尺寸，以确保图像符合模型的输入要求。此外，Input 端还执行图像增强任务，通过一系列预处理步骤，如归一化、标准化等，来提高图像的质量和特征的可提取性。这些操作为后续的处理步骤奠定了坚实的基础。Backbone 部分在 YOLOv10n 中承担着特征提

取的核心任务。它通过一系列复杂的卷积层和池化层，从输入图像中提取丰富的特征信息，使得模型能够更好地捕捉图像中的关键信息。Neck 端的主要职责是进行特征融合，将不同层次的特征信息进行有效整合。在保留了 FPN 和 PANet 的经典结构的同时，Neck 端通过巧妙的设计，实现了浅层特征与深层特征的有效融合。这种融合不仅增强了特征的表达能力，还提高了模型对复杂场景的适应性。Head 端则采用了创新的解耦头结构，将回归分支和预测分支分离，使得模型在处理目标检测任务时更加高效和准确。通过这种解耦设计，回归分支专注于定位任务，预测分支则专注于分类任务，两者相互独立但又紧密协作，最终汇聚输出，生成精确的目标检测结果。这种结构的设计不仅提高了模型的性能，还为后续的研究和应用提供了更多的可能性。

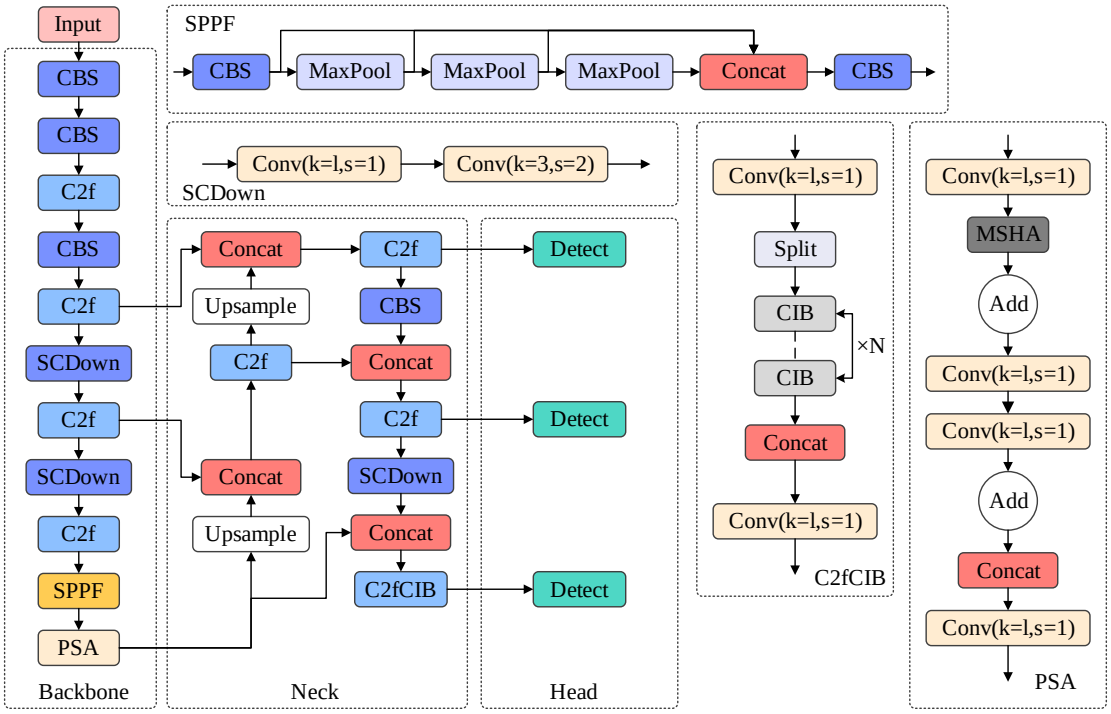


图 4-1 YOLOv10n 网络结构图

4.3 YOLOv10n 算法改进

针对 YOLOv10n 在检测红外图像行人遮挡时存在错检、漏检、模型参数量大等问题，受文献[52]启发，对 YOLOv10n 模型中 Backbone 端与 Neck 端进行改进，提高对红外图像行人遮挡检测精度，同时达到模型轻量化要求。改进后的模型为 YOLOv10n-SCM，如图 4-2 所示。

(1) 建立本课题使用的红外图像行人遮挡数据集。为了确保数据集的质量和代表性，从红外图像数据集 KAIST 中精心挑选了 2500 张含有行人遮挡的夜间红外图像作为本章使用的实验数据集。

(2) 为了提高模型对尺度变化的适应性和对小尺度对象的检测能力，在网络中引入了 CCFM 模块。该模块通过降低计算量，使得模型在保持高效率的同时，能够更好地处理尺度变化带来的挑战，从而提升了模型在复杂场景下的检测性能。

(3) 引入 SCConv 卷积模块，并将其融合到 Backbone 部分的 C2f 模块中。SCConv 模块通过捕捉不同尺度和通道的特征信息，进一步提升了模型的特征提取能力。这种融合不仅增强了模型对细节特征的捕捉能力，还提高了整体的检测效率和性能。

(4) 构建 C2f-EMA 模块，并用它来替换原有 Neck 端中的 C2f 模块。C2f-EMA 模块通过更好地获取上下文信息，进一步提升了模型对红外图像中行人特征信息的提取能力。这一改进有助于模型在处理遮挡和复杂背景时，能够更准确地定位和识别行人目标。

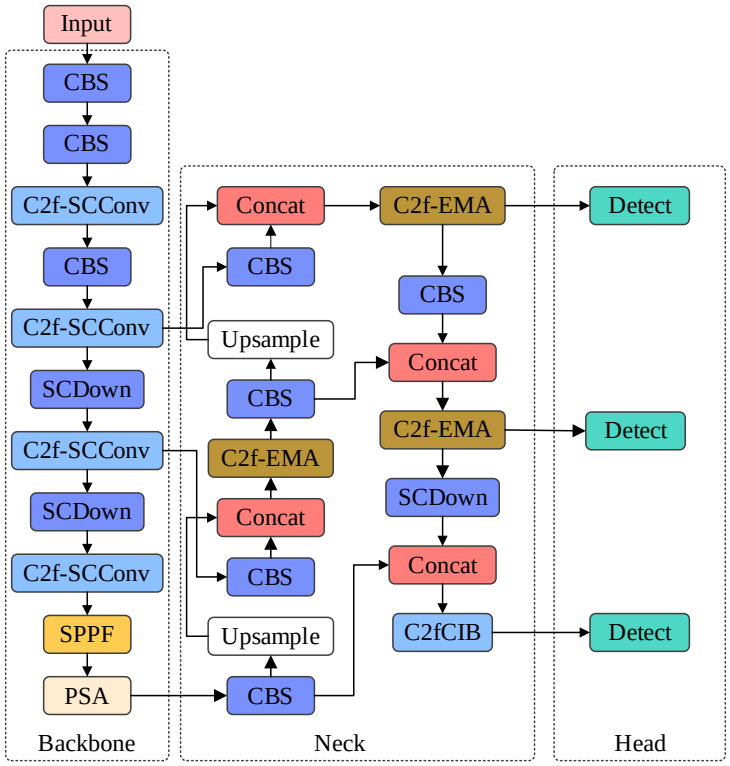


图4-2 YOLOv10n-SCM网络结构图

通过上述改进，YOLOv10n 检测网络模型在处理红外图像遮挡情况下的行人

检测问题上，展现出了更高的准确性和鲁棒性。

4.3.1 跨尺度特征融合模块 CCFM

CCFM[53]即跨尺度特征融合模块，是 RT-DETR 框架内的一项重要创新。该模块能够巧妙地融合跨尺度的特征信息，极大地增强了模型在目标特征提取方面的能力，使其对尺度变化具备更高的适应性，并提升了对小尺度目标的检测精度。此外，CCFM 通过有效整合细节特征与上下文信息，进一步提升了模型的整体性能表现。

在 YOLOv10nNeck 中引入轻量级的 CCFM 模块，增强了模型对大小不同目标的识别能力，同时拓宽了模型的感受野，深化了其语义理解能力。尤其是在目标检测的小目标检测精度和多尺度目标检测性能方面，提升模型的感受野和语义理解能力。CCFM 结构如图 4-3 所示。

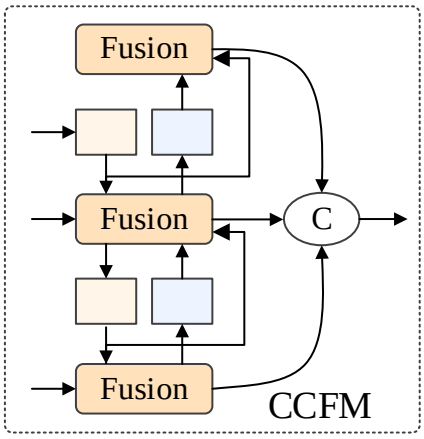


图 4-3 CCFM 网络结构图

4.3.2 空间和通道重构卷积 SCConv

为了进一步提升主干网络在特征提取方面的表现，并且减少模型中不必要的冗余信息，从而提高模型在目标检测任务中的准确率，本章提出将 SCConv[54-55]模块与主干网络中的 C2f 模块相结合，形成了一个名为 C2f-SCConv[56]的复合模块。通过这种方式，能够在保持模型精度的前提下，降低模型的计算量和参数量。C2f-SCConv 模块的引入是为了在特征提取过程中更有效地利用信息，同时避免不必要的计算负担。通过将 SCConv 模块与 C2f 模块相结合，能够更好地平衡模型的精度和效率。

SCConv 模块的核心在于其独特的结构设计，它主要由两个关键部分组成：空间重建单元（SRU）和通道重建单元（CRU）。这两个单元相互配合，协同工作，以实现对特征图的空间和通道维度的有效重建。空间重建单元专注于减少空间冗余，通过重新组织特征图的空间结构，使得模型能够更加关注于重要的特征区域，从而提高特征提取的效率。而通道重建单元则专注于减少通道冗余，通过优化特征通道的表示，进一步提升模型的性能。通过这种双重冗余减少机制，SCConv 模块能够提升模型的性能和速度，其网络结构如图 4-4 所示。在实际应用中，C2f-SCConv 模块的引入不仅能够提高模型的检测准确率，还能够在保持高精度的同时，减少模型的计算量和参数量，从而使得模型更加轻量级，更适合在计算资源受限的环境中部署。

如图 4-5 所示，C2f-SCConv 模块的结构设计充分考虑了特征提取的高效性和准确性，通过巧妙地结合 SCConv 模块与 C2f 模块，实现了在目标检测任务中的出色表现。

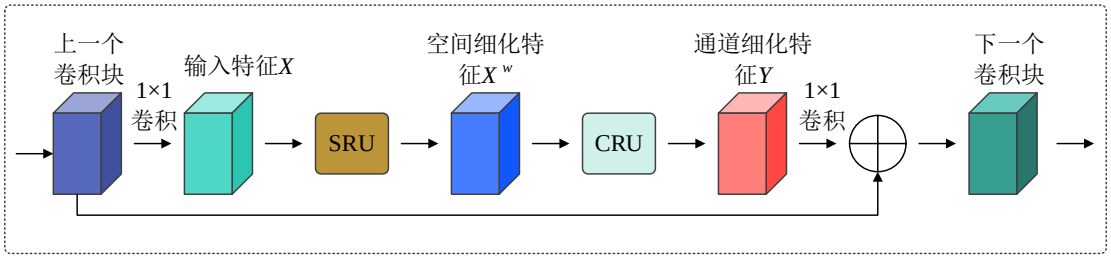


图 4-4 SCConv 网络结构图

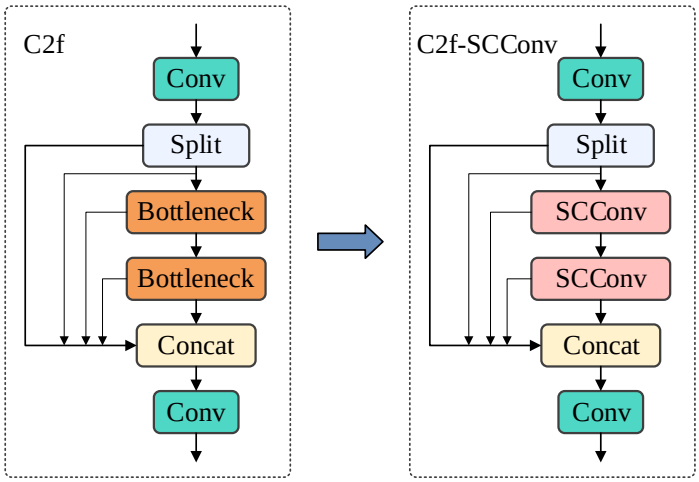


图 4-5 C2f-SCConv 网络结构图

(1) SRU 结构

SRU结构，即空间重建单元，主要通过分离和重构两个关键步骤来实现特征冗余的减少，进而降低模型的参数量和计算量，其结构如图4-6所示。首先，SRU结构接收输入特征 X ，然后通过归一化处理，通常使用组归一化(Group Normalization, GN)，得到输出特征 X_{out} ，公式如(4-1)所示。这一过程不仅简化了模型的复杂度，还提高了计算效率，使得模型在处理大规模数据时更加高效和轻量。通过这种方式，SRU结构能够在保持模型性能的同时，有效减少计算资源的消耗，从而在实际应用中具有更高的实用性和灵活性。

$$X_{out} = GN\left(X\right) = \gamma \frac{X-u}{\sqrt{\sigma^2+\varepsilon}} + \beta \tag{4-1}$$

式(4-1)中：u 为 X 平均值，σ 为 X 标准差。γ 和 β 是仿射变换的训练参数，ε 是为除法稳定性加入的一个较小的正常数。

在经过归一化处理之后，得到的结果会进一步通过一个名为 W_γ 的权重矩阵进行重新加权，具体如公式(4-2)所示：

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^{C}\gamma_j} \qquad i, j = 1, 2, ..., C \tag{4-2}$$

接下来将经过归一化处理后的数据与重新加权后的 W_γ 进行相乘操作。这一过程是为了将两个数据集进行融合，以便进一步处理。为了确保相乘的结果能够被有效地控制在一个合理的范围内，引入Sigmoid函数。Sigmoid函数的作用是将相乘的结果压缩并限制在(0,1)这个区间内，从而确保数据的稳定性和可操作性。在得到Sigmoid函数处理后的结果后，通过阈值控制进行门控操作。门控操作的目的是为了筛选出有用的信息，同时剔除冗余的数据。对于高于阈值的特征权重 W_1，认为它们是有用的特征，因此将它们的权重设为1，以保留这些重要的信息。相反，对于那些低于阈值的特征权重 W_2，认为它们是冗余的特征，因此将它们的权重设为0，以排除这些不重要的信息。整个运算过程可以用公式(4-3)来表示，其中Gate代表门控操作，用于进行阈值控制。通过这种方式，可以有效地筛选出有用的信息，同时剔除冗余的数据，从而提高数据处理的效率和准确性。

$$W = Gate\left\{Sigmoid\left(W_\gamma\left(GN\left(X\right)\right)\right)\right\} \tag{4-3}$$

最后将输入特征 X 与权重矩阵 W_1 、W_2 进行交叉运算，生成更加丰富和详细

的信息特征，这种交叉运算的过程有助于提取和融合更多的特征信息，从而使得最终获得的空间细化特征 X^w 更加精确和具有代表性，公式如(4-4)所示：

$$\begin{cases} X_1^w = W_1 \otimes X \\ X_2^w = W_2 \otimes X \\ X_{11}^w \oplus X_{22}^w = X^{w1} \\ X_{21}^w \oplus X_{12}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \tag{4-4}$$

式(4-4)中：$\otimes$ 表示元素相乘，$\oplus$ 表示元素相加，$\cup$ 代表求并集。

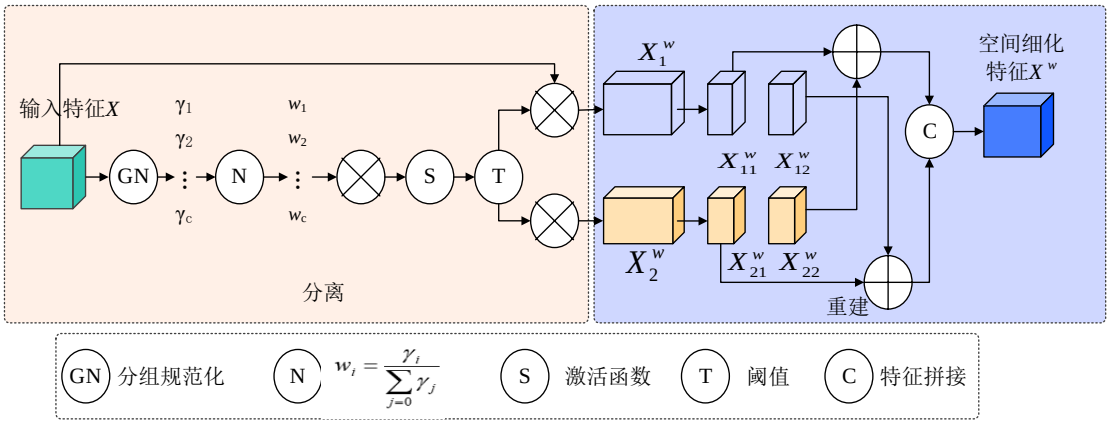


图 4-6 SRU 结构

(2) CRU 结构

CRU结构通过分割、转换和融合三个关键步骤来进一步消除 X^w 在通道维度上的特征冗余。如图4-7所示，这一过程可以有效地提升特征表达的准确性和鲁棒性。具体来说，在分割阶段，输入的特征图 X^w 被细分为两个部分：X_{up} 和 X_{low}。其中，X_{up} 部分被分配到上转换通道，通过使用高效的卷积操作来提取具有代表性的特征 Y_1。这些特征通常包含高层次的语义信息，有助于捕捉图像的主要内容。与此同时，X_{low} 部分则被分配到下转换通道，通过采用相对廉价的卷积操作来获取细节性特征 Y_2。这些特征通常包含丰富的局部信息，有助于捕捉图像的细微差别。通过这种分割策略，CRU结构能够有针对性地处理不同层次的特征，从而有效地减少通道间的冗余。在转换过程中，上转换通道和下转换通道分别提取出具有不同特性的特征，使得最终融合得到的特征更加丰富和全面。最后，在融合阶段，将提取出的代表性特征 Y_1 和细节性特征 Y_2 进行融合，生成通道细化特征

Y。这一融合过程不仅保留了高层次的语义信息，还增强了细节信息的表达能力，从而进一步提高了特征表达的准确性和鲁棒性。通过这种方式，CRU结构能够有效地优化特征表示，为后续的任务提供更为强大的特征支持。

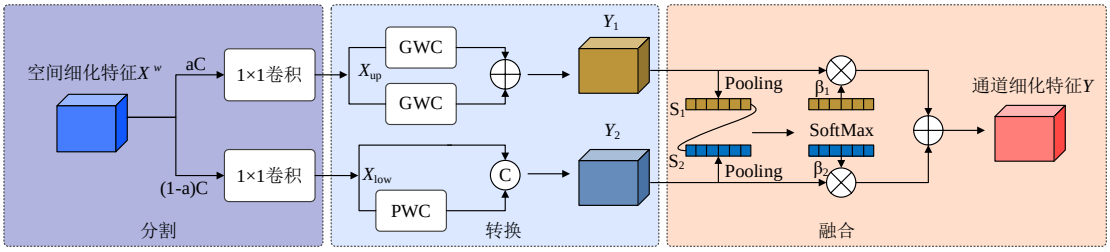


图 4-7 CRU 结构

4.3.3 高效多尺度注意力机制 EMA

在处理红外图像行人检测任务时，常常会面临行人之间相互遮挡的问题，这使得模型在定位行人时遇到困难。传统的 C2f 模块在处理这种遮挡问题时，往往难以准确地提取和筛选红外图像中行人的特征信息，从而导致漏检和错检的现象。为了解决这一问题，本章提出了一种改进的方法，即在 C2f 模块的基础上引入了高效多尺度注意力机制(Exponential Moving Average, EMA)，从而得到了 C2f-EMA[57]模块。这一改进增强了模型对目标的感知能力，提高红外图像行人检测的准确性。EMA[58]注意力机制是一种高效的多尺度注意力机制，它通过跨空间学习来实现对通道特征的学习，而无需进行降维操作。这种机制能够有效地增强模型对目标特征的感知能力。EMA 注意力机制模块网络结构如图 4-8 所示，H 和 W 分别为输入特征的空间维度，C 为通道数。其具体的工作流程如下：首先，对于任意输入的图像，EMA 注意力机制会按照跨通道维度方向将其切分为 G 个子特征。接着，通过两条 1×1 卷积分支和一条 3×3 卷积分支来获取不同的语义信息。在 1×1 分支中，采用两个一维全局平均池化操作对通道进行编码，以实现跨通道信息的交互。然后，使用二维全局平均池化操作分别将 1×1 分支和 3×3 分支的输出进行全局空间信息编码。将这两个编码特征连接起来，经过 1×1 卷积分解为两个向量，再利用非线性 Sigmoid 函数进行线性拟合。最后，将两条分支的注意力特征进行聚合，从而得到最终的注意力特征图。通过这种方式，C2f-EMA 模块能够更好地处理红外图像行人检测中的遮挡问题，提高检测的准确性和鲁棒性。

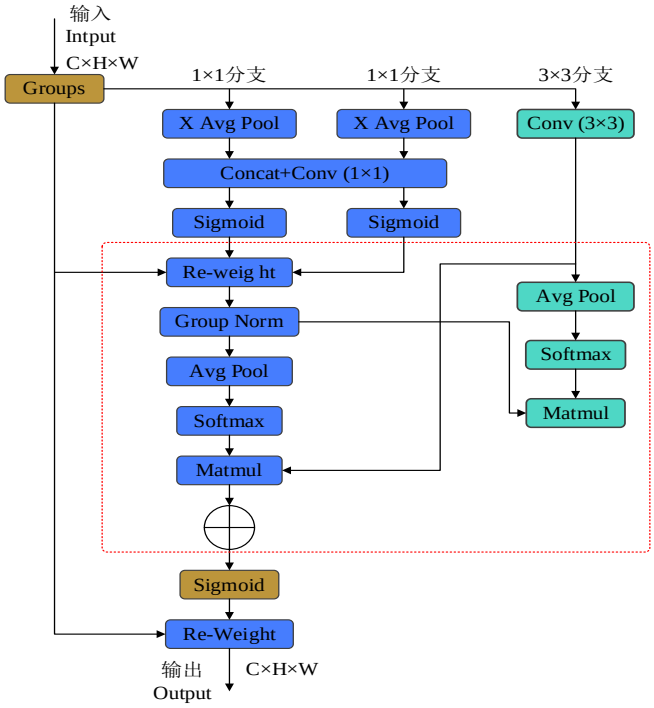


图 4-8 EMA 网络结构图

本章将 EMA 注意力机制融入了 C2f 模型的 Bottleneck 结构中，通过这种方式，构建了 C2f-EMA 模块。如图 4-9 所示，C2f-EMA 模块通过引入 EMA 注意力机制，增强了 Bottleneck 结构捕捉多尺度特征信息的能力。这种改进使得模型在处理红外图像时，能够更加有效地识别和定位行人目标，从而降低了行人漏检和误检的发生率。

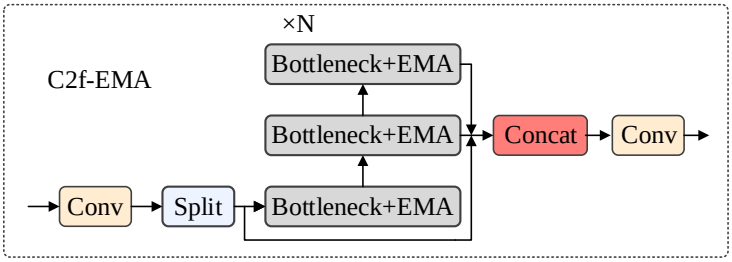


图 4-9 C2f-EMA 网络结构图

4.4 实验结果与分析

4.4.1 实验环境

实验环境采用 Windows10 操作系统，硬件设备 NVIDIA TITAN X 显卡，使用 pytorch1.9 框架，进行模型训练。设置动量为 0.937，batch-size 为 8，epoch 为 200，确保模型能够充分学习并达到较好的性能。

为了验证本章所提出的改进方法的有效性，分别对 YOLOv10n 的主干网络、颈部网络以及 C2f 模块进行了细致的改进。通过一系列的实验，对比了各种改进方法所带来的效果。在实验结果中，最优的数据用加粗字体进行标注，以突出其最佳性能；而次优的数据则用下划线进行标注，以便于区分和识别，从而更直观地评估本章提出方法的有效性。

4.4.2 数据集制作

数据集使用 KAIST 红外图像行人数据集。为了更好地研究行人遮挡问题，从 KAIST 庞大的数据集中精心筛选出了 2500 张包含行人遮挡情况红外图像，建立本课题使用的红外图像行人遮挡数据集，数据集中部分行人遮挡图像如图 4-10 所示，并按照 7:2:1 的比例划分训练集、验证集和测试集。与第 3 章数据集预处理方式相同，为了进一步提升图像质量，增强数据集的多样性，对建立的红外数据集采用 CLAHE 算法对图像进行增强处理。最后对预处理后的数据集使用标注软件 LabelImg 软件进行人工标注，标注标签为 person，确保数据集的质量和可用性。

(a) 轻度遮挡　　(b) 中度遮挡　　(c) 重度遮挡

图 4-10 KAIST 数据集中部分行人遮挡图像

4.4.3 C2f 模块改进实验

实验一：C2f 模块卷积改进实验

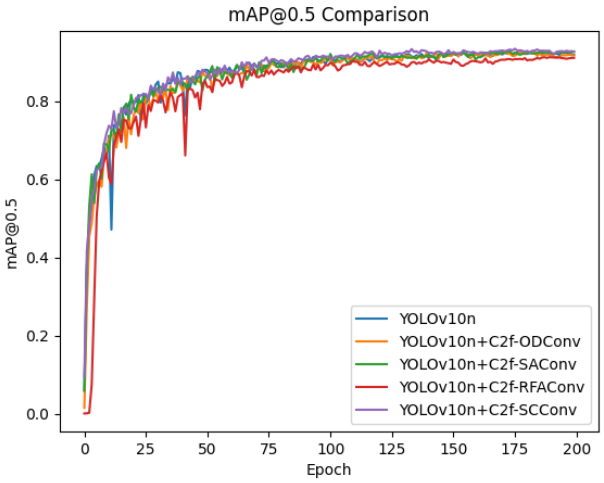
分别将全维度动态卷积[59]（ODConv）、可切换空洞卷积[60]（SAConv）、感受野注意力卷积[61]（RFAConv）、空间和通道重构卷积（SCConv）融入到主干网络中的 C2f 模块中，分别构建 C2f-ODConv、C2f-SAConv、C2f-RFAConv、C2f-SCConv 四种模块，C2f 模块改进后的模型性能对比结果如表 4-1 所示。

表 4-1 C2f 模块改进后的模型性能对比结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv10n	92.2	2.69	8.2
YOLOv10n+C2f-ODConv	92.3	3.70	**7.2**
YOLOv10n+C2f-SAConv	92.1	2.99	7.5
YOLOv10n+C2f-RFAConv	91.0	2.73	8.5
YOLOv10n+C2f-SCConv（Ours）	**92.6**	**2.50**	7.7

由表 4-1 实验结果可知，当以 YOLOv10n 作为基准模型，分别引入 C2f-SAConv、C2f-RFAConv 两种模块时，模型在 mAP@0.5 指标上有所下降，分别下降了 0.1%与 1.2%,而在参数量和计算量方面，C2f-RFAConv 相较于基准模型 YOLOv10n 略有增加，分别增加了 1.49%与 3.66%，但增加幅度不大。C2f-SAConv 相较于基准模型 YOLOv10n，其参数量增加了 11.15%，计算量降低了 8.54%。这表明，尽管这两种模块在提升模型精度方面的表现未达预期，但它们并未增加模型的计算负担。

当以 YOLOv10n 作为基准模型，分别引入 C2f-ODConv、C2f-SCConv 两种模块时，模型在 mAP@0.5 指标上有所提升，分别增加了 0.1%与 0.4%。尤其是引入 C2f-SCConv 模块时，模型的参数量和计算量也得到了进一步的减少，分别减少了 7.06%和 6.10%。这一结果表明，C2f-SCConv 模块在提升模型精度的同时，还能有效减轻模型的计算负担。图 4-11 为 C2f 模块卷积改进实验结果对比图。



4-11 C2f 模块卷积改进实验结果对比图

实验二：C2f 模块注意力机制改进实验

分别将自注意力机制[62]（ACmix）、混合局部通道注意力机制[63]（MLCA）、

三重注意力机制[64]（Triplet）、高效多尺度注意力机制（EMA）融入到颈部网络中的 C2f 模块中，分别构建 C2f-ACmix、C2f-MLCA、C2f-Triplet、C2f-EMA 四种模块，并进行了一系列对比实验。C2f 模块改进后的模型性能对比结果如表 4-2 所示。

表 4-2 C2f 模块改进后的模型性能对比结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv10n	92.2	**2.69**	**8.2**
YOLOv10n+C2f-ACmix	91.9	2.74	8.5
YOLOv10n+C2f-MLCA	92.1	2.70	8.3
YOLOv10n+C2f-Triplet	90.4	2.71	8.4
YOLOv10n+C2f-EMA（Ours）	**92.4**	2.70	8.3

由表 4-2 实验结果可知，当以 YOLOv10n 作为基准模型，分别引入 C2f-ACmix、C2f-MLCA、C2f-Triplet 三种模块时，模型在 mAP@0.5 指标上都有所下降，分别下降了 0.3%、0.1%与 1.8%；当以 YOLOv10n 作为基准模型，引入 C2f-EMA 模块时，模型在 mAP@0.5 指标上有所提升，增加了 0.2%。C2f-EMA 模块在提升模型检测精度方面的效果，进一步验证了其多尺度注意力机制的有效性。图 4-12 为 C2f 模块注意力机制改进实验结果对比图。

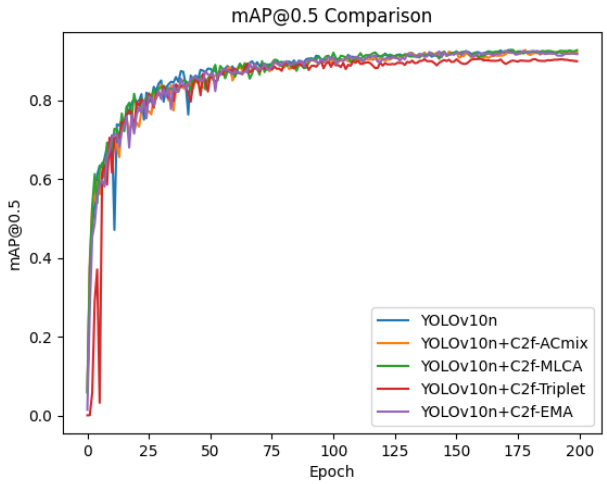


图 4-12 C2f 模块注意力机制改进实验结果对比图

4.4.4 消融实验

为验证本章提出的三种改进方法的有效性，在本课题建立的数据集上进行了不同的消融实验，确认这些改进方法对整体性能的影响。表 4-3 详细列出了不同改进方法组合下的实验结果，通过对比分析，可以清晰地看到每种改进方法对模

型性能的具体影响。消融实验结果如表 4-3 所示。

表 4-3 消融实验结果

YOLOv10n	CCFM	C2f-SCConv	C2f-EMA	mAP@0.5/%	Parameters/M	GFLOPs/G	FPS
√				92.2	2.69	8.2	**294**
√	√			92.9	1.85	6.5	227
√		√		92.6	2.50	7.7	232
√			√	92.4	2.70	8.3	263
√	√	√		93.3	**1.63**	**5.9**	222
√	√	√	√	**93.9**	1.66	6.0	208

根据消融实验的结果可知，当以 YOLOv10n 作为基准模型，分别引入 CCFM 模块与 C2f-SCConv 模块时，模型在 mAP@0.5 指标上实现了提升，分别增加了 0.7%与 0.4%。此外，这一改进还带来了额外的好处，即模型的参数量分别减少了 31.2%和 7.06%，计算量分别减少了 20.7%和 6.10%。这一结果表明，CCFM 模块与 C2f-SCConv 模块不仅提高了模型的精度，还有效地优化了模型的计算效率和参数规模。同理，当以 YOLOv10n 作为基准模型，引入 C2f-EMA 模块时，虽然模型的计算量与参数量都略有增加，但 mAP@0.5 值依然增加了 0.2%。

进一步地，当以 YOLOv10n 为基准模型，同时引入 CCFM 和 C2f-SCConv 这两种改进方法时，模型的性能得到了进一步的提升。具体来说，mAP@0.5 值达到了 1.1%的提升。与此同时，模型的计算量和参数量也得到了进一步的减少，这表明这两种改进方法在协同工作时能够相互补充，进一步优化模型的性能和效率。

最后，当以 YOLOv10n 为基准模型，同时引入 CCFM、C2f-SCConv 以及 C2f-EMA 这三种改进方法时，尽管模型的计算量和参数量相比仅引入 CCFM 和 C2f-SCConv 时有所增加，但模型的 mAP@0.5 值实现了更大的提升，达到了 1.7%。这一结果表明，尽管引入 C2f-EMA 会增加一定的计算和参数负担，但其性能提升了，从而在整体上提高了模型的综合表现。

综上所述，通过引入 CCFM、C2f-SCConv 以及 C2f-EMA 这三种改进方法，能够在保持模型计算效率和参数规模优化的同时，提升模型在 mAP@0.5 指标上的表现。这些改进方法的协同作用不仅提高了模型的精度，还优化了模型的整体性能，使其在实际应用中更具竞争力。图 4-13 为改进算法前后对比图。

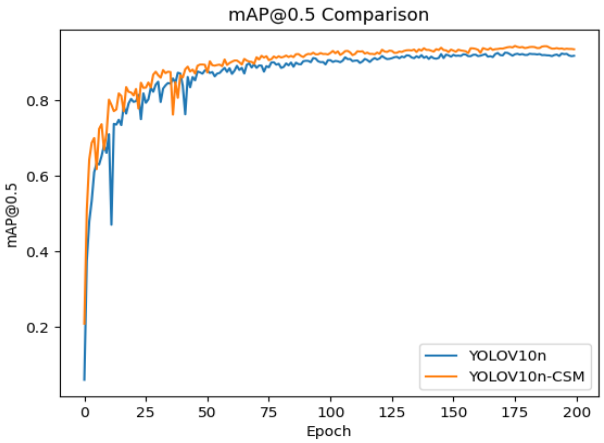


图 4-13 改进算法前后对比图

4.4.5 对比实验

(1) 本课题红外图像行人遮挡数据集对比实验

在实验环境及训练参数一致的情况下，在本课题建立的数据集上，对本章改进的算法和 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 等主流算法进行了训练和测试，以验证本章提出的模型的有效性，具体实验结果见表 4-4。

表 4-4 对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3-tiny	88.7	12.12	18.9
YOLOv7-tiny	89.5	6.01	13.1
YOLOv8n	92.0	3.01	8.1
YOLOv10n	92.2	2.69	8.2
YOLOv11n	89.9	2.58	6.3
YOLOv10n-SCM(Ours)	**93.9**	**1.66**	**6.0**

从表 4-4 可得，本章算法 YOLOv10n-SCM 与 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 算法相比，mAP@0.5 值分别高出 5.2%、4.4%、1.9%、1.7%、4.0%；参数量分别降低 86.3%、72.4%、44.9%、38.3%、35.7%；计算量分别降低 68.3%、54.2%、25.9%、26.8%、4.8%。验证了改进算法的有效性，提高算法检测精度的同时，实现模型的轻量化设计。

(2) 混合数据集对比实验

为了进一步验证本章方法的优越性，从公共红外图像数据集 KAIST 与 FLIR 上分别筛选出所有包含遮挡行人的红外图像构成混合数据集进行对比实验。混合

数据集中部分图像如图 4-14 所示。

(a) FLIR数据集图像　　(b) KAIST数据集图像

图 4-14 混合数据集中部分图像

具体实验结果见表 4-5。

表 4-5 对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3-tiny	89.1	12.12	18.9
YOLOv7-tiny	90.0	6.01	13.1
YOLOv8n	92.1	3.01	8.1
YOLOv10n	92.3	2.69	8.2
YOLOv11n	90.2	2.58	6.3
YOLOv10n-SCM(Ours)	**94.3**	**1.66**	**6.0**

由表 4-5 实验结果可知，用混合数据集检测精度比单个数据集的检测精度要高。从表 4-5 可得，本章改进的模型 YOLOv10n-SCM 与 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv11n 算法相比，mAP@0.5 值分别高出 5.2%、4.3%、2.2%、4.1%；与基线模型 YOLOv10n 相比，mAP@0.5 值提高 2.0%，同时参数量与计算量更低，具有高精度和低成本的优势。综合各项评价指标可得，本章提出的算法模型 YOLOv10n-SCM 在轻量化的同时也提高了检测精度，满足检测红外图像行人遮挡的要求。

(3) 可见光数据集对比实验

使用 KAIST 红外图像数据集中白天场景的包含行人遮挡的可见光数据集来进一步验证本章算法的效果。KAIST 红外图像数据集中每张图片都包含 RGB 彩色图像和红外图像两个版本，从其中挑选出 2500 张可见光图像构成可见光数据集来进一步验证本章算法的效果。可见光行人遮挡部分图像如图 4-15 所示。

图 4-15 可见光数据集中部分行人遮挡图像

可见光数据集对比实验结果见表 4-6。

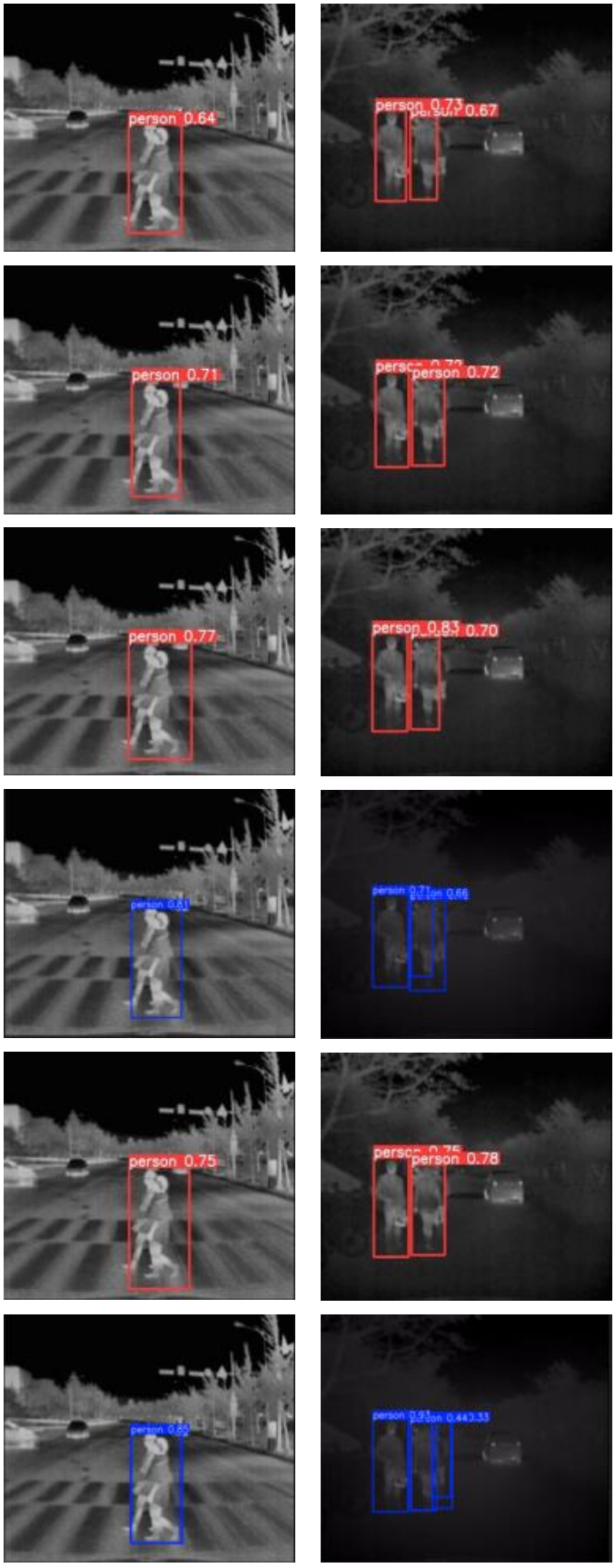
表 4-6 可见光数据集对比实验结果

算法	mAP@0.5/%	Parameters/M	GFLOPs
YOLOv3-tiny	81.5	12.12	18.9
YOLOv7-tiny	85.9	6.01	13.1
YOLOv8n	86.2	3.01	8.1
YOLOv10n	87.3	2.69	8.2
YOLOv11n	85.8	2.58	6.3
YOLOv10n-SCM(Ours)	**88.3**	**1.66**	**6.0**

从表 4-6 可得，本章改进的模型 YOLOv10n-SCM 在可见光数据集上也同样适用。从 mAP@0.5 上看，YOLOv10n-SCM 的实验结果为 88.3%，分别比 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 提高了 6.8%、1.7%、2.4%、2.1%、1.0%、2.5%。

4.4.6 检测结果可视化

为了比较本章提出的算法 YOLOv10n-SCM 与 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 等算法对红外图像行人遮挡的检测效果，从测试集中挑选出不同场景下的行人遮挡进行检测。图 4-16 从上到下分别展示了 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n、YOLOv10n-SCM 对于同一场景的红外图像行人遮挡检测结果。



(a) 场景一　　(b) 场景二

图 4-16 YOLOv10n-SCM 检测结果

通过审视图 4-16（a），可以观察到场景一中存在两位行人，行人之间存在中度遮挡问题。在图 4-16（a）中，YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 等算法只检测到了一位行人，未能准确识别出两位存在中度遮挡的行人，出现了漏检现象。相反，YOLOv10n-SCM 算法成功地检测到了重叠的两位行人，并且相较于 YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv10n、YOLOv11n 的检测结果，其检测精度也有了提升。

同理，观察图 4-16（b）可以发现场景二中存在四位行人，与图 4-16（a）相比，图 4-16（b）中的四位行人之间存在严重遮挡问题。在图 4-16（b）中，YOLOv3-tiny、YOLOv7-tiny、YOLOv8n、YOLOv11n 等算法仅识别出两位行人，YOLOv10n 仅识别出三位行人，同样存在漏检的情况。而在图 4-16（b）中，YOLOv10n-SCM 算法成功检测出所有行人，其检测精度相较于图 4-16（b）也有了提高。

综上所述，YOLOv10n 在检测被遮挡的行人时，存在漏检和检测精度不高的问题。这主要是由于 YOLOv10n 倾向于关注局部特征，而忽视了其他重要的细节特征。相比之下，YOLOv10n-SCM 通过集成 CCFM 模块、SCConv 卷积模块以及 EMA 注意力机制模块，提升了模型对细节特征的关注。这使得模型在定位行人特征信息方面表现更佳，加强了细节特征的提取和融合，从而提高了对被遮挡行人的检测精度。

4.5 本章小结

本章提出一种改进 YOLOv10n 红外图像行人遮挡检测算法。该算法有效解决了传统 YOLOv10n 在红外图像行人遮挡检测中遇到的问题。首先，通过引入 SCConv 卷积构建 C2f-SCConv 模块，不仅提升了模型对行人特征的整体提取能力，还实现了模型体积的缩减。其次，通过集成 CCFM 模块，进一步增强了模型在复杂场景下的检测性能。最后，在颈部网络中引入了 EMA 注意力机制，以更精确地定位和识别行人目标。实验结果表明，相较于现有模型，改进后的 YOLOv10n-SCM 算法在保持体积轻量化的同时，提升了检测精度，满足了实时性与准确性之间的平衡需求，展现出更高的实用价值。

第 5 章 总结与展望

5.1 全文总结

行人检测是一种计算机视觉技术，其从图像或视频中自动识别和定位行人，广泛应用在智能监控、自动驾驶、智能交通、人体行为分析等领域。本文在现有的行人检测模型基础上，对红外图像行人检测与行人遮挡问题进行了研究，通过改进 YOLO 检测算法模型，提高了对行人的检测精度。本文的主要工作内容如下：

(1) 通过阅读相关文献，对行人检测算法进行初步了解，并分析每种算法的优缺点，最终确定改进 YOLO 检测算法模型。针对目前开源的红外图像行人数据集匮乏，无法充分满足本文算法训练的需求，本文通过自制红外数据集，以供后续研究使用。

(2) 针对红外图像行人检测，解决 YOLOv5s 算法中模型体积大、检测精度低等问题，改进其网络结构。首先，将 CA 注意力机制模块添加到 Backbone 网络的 C3 模块中，以增强对行人的长距离检测能力；其次，在 Backbone 网络中的 SPPF 下一层引入了 SimAM 注意力机制模块，以提升模型对行人特征的关注度，并增强对不同深度特征信息的感知能力；再次，在 Neck 网络中引入 BiFPN，BiFPN 的引入，不仅加强了低层与高层特征间的信息交流，还促进了不同尺度特征的有效整合。最后，引入包含 GSConv 的 Slim-Neck 设计范式降低模型复杂性，更好地平衡模型准确性和速度。实验结果表明，改进后的模型不仅能够提高网络的检测精度，同时降低了模型的复杂度，实现了模型轻量化设计。

(3) 本文针对红外图像行人遮挡的问题，在 YOLOv10n 算法的基础上进行以下改进：首先，为了提高模型对尺度变化的适应性和对小尺度对象的检测能力，在网络中引入了 CCFM 模块；其次，引入 SCConv 卷积模块，并将其融合到 Backbone 部分的 C2f 模块中。SCConv 模块通过捕捉不同尺度和通道的特征信息，进一步提升了模型的特征提取能力；最后，构建 C2f-EMA 模块，并用它来替换原有 Neck 端中的 C2f 模块。C2f-EMA 模块通过更好地获取上下文信息，进一步提升了模型对红外图像中行人特征信息的提取能力。实验结果表明，改进后的算法模型在处理红外图像行人遮挡问题上，展现出了更高的准确性和鲁棒性。

综上所述：第 3 章所提出的基于改进 YOLOv5s 的轻量化红外图像行人检测算法 YOLOv5s-CSBS 适用于行人较少的场景，例如家庭或小型办公室等环境。在这些环境中，由于人员流动较少，遮挡情况相对不常见，YOLOv5s-CSBS 算法能够充分发挥其优势，快速且准确地检测行人，从而为安全监控和智能管理提供可靠支持。此外，该算法具备低计算复杂度的特点，使其在资源受限的环境中表现优异，实现了实时性与准确性的平衡。

第 4 章所提出的基于改进 YOLOv10n 的红外图像行人遮挡检测算法 YOLOv10n-SCM 适用于行人密集的场景，例如商场、大型公共场所等环境。在这些区域，由于人员流动频繁且遮挡现象较为普遍，YOLOv10n-SCM 算法凭借其卓越的检测能力和优化的处理策略，能够快速且准确地识别多个行人，即便在复杂环境下也能维持高检测精度。此外，该算法在处理大规模数据时展现出的高效性，使其在行人密集的环境中依然能够稳定运行，为公共安全管理与商业智能分析提供了可靠的技术支持。

5.2 工作展望

本文以红外图像行人检测为研究内容，分别对红外图像行人检测和红外图像遮行人遮挡检测提出了两种改进算法，虽然取得了一定的研究成果，但仍然有一些的不足之处，还需要去研究与探索。

(1) 数据集不足。现有红外行人检测数据集较少，标注成本高，难以支持深度学习模型训练。因此，建立一个大规模红外图像行人检测数据集是很有必要的，便于研究人员针对红外图像行人检测进行后续深入研究。

(2) 多模态融合。利用来自多个不同类型传感器的数据进行信息融合，以提升目标检测、识别、跟踪等任务的准确性和鲁棒性。在行人检测任务中，单一模态（如红外或可见光）可能会受环境因素影响，而多模态融合可以弥补不同传感器之间的不足，提高系统性能。通过融合可见光与红外、激光雷达、毫米波雷达等多种传感器信息，提高检测准确性。

(3) 强化学习与自适应检测。由于红外成像可以在低光、复杂天气环境下稳定工作，其在传统可见光行人检测难以应用的场景中具有独特优势。然而，红外行人检测仍面临目标对比度低、背景干扰大、行人检测场景复杂等挑战，未来可结合强化学习进行动态调整，提高检测稳定性。

参考文献

[1] 王清芳,胡传平,李静.面向交通场景的轻量级行人检测算法[J].郑州大学学报(理学版),2024,56(04): 48-55.

[2] 杨磊,王少云,刘力冉,等.一种智能视频监控系统中的行人检测方法[J].计算机与现代化,2019,(11): 69-74.

[3] 钟明.基于深度学习的智能视频监控系统设计[J].电子产品世界,2024,31(12): 38-42.

[4] Ye L, Duan T, Zhu J. Neural network-based semantic segmentation model for robot perception of driverless vision[J]. IET Cyber-Systems and Robotics, 2020, 2(4): 190-196.

[5] 武慧荣,张敬宜,郭春敏.基于深度学习的公交客流检测算法[J/OL]. 控制与决策,1-11[2025-03-22].

[6] Heo D, Lee E, Chul Ko B .Pedestrian Detection at Night Using Deep Neural Networks and Saliency Maps[J].Journal of Imaging Science and Technology, 2017, 61(6).

[7] 姜小强,陈骋,朱明亮.基于红外传感的视频监控行人检测方法[J].煤炭技术,2022,41(10): 223-225.

[8] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Ieee, 2001, 1: I-I.

[9] 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.

[10] Zhang C X, Zhang J S, Kim S W. PBoostGA: pseudo-boosting genetic algorithm for variable ranking and selection[J]. Computational Statistics, 2016, 31: 1237-1262.

[11] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20: 273-297.

[12] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J].

International Journal of Computer Vision, 2004, 60(2): 91-110.

[13]花成才.基于红外图像的行人检测与跟踪算法研究[D].天津职业技术师范大学,2024.

[14]Olmeda D, de la Escalera A, Armingol J M. Contrast invariant features for human detection in far infrared images[C]//2012 IEEE Intelligent Vehicles Symposium. IEEE, 2012: 117-122.

[15]胡庆新,吕鹏.基于多特征融合的红外图像行人检测[J].计算机应用,2016,36(S1): 157-160+195.

[16]胡燕伟,徐美华,郭爱英.基于 BING-casDPM 的快速行人检测算法[J].上海大学学报(自然科学版),2019,25(05): 712-721.

[17]Xue T, Zhang Z, Ma W, etc. Nighttime Pedestrian and Vehicle Detection Based on a Fast Saliency and Multifeature Fusion Algorithm for Infrared Images[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16741-16751.

[18]李锐.基于红外与可见光双模图像融合的车辆行人检测算法研究[D].重庆交通大学,2024.

[19]Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex[J]. The Journal of physiology, 1968, 195(1): 215-243.

[20]Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980, 36(4): 193-202.

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

[22]Girshick R, Donahue J, Darrell T, etc. 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.

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

[24]Ren S Q, He K M, Girshick R, etc. Faster R-CNN: Towards Real-time Object Detection With Region Proposal Networks[C]//Advances in Neural Information Processing Systems, December 7-12, 2015, Montréal, Quebec, Canada. New York:

Curran Associates, 2015: 91-99.

[25] Redmon J, Divvala S, Girshick R. B, etc. You Only Llook Once: Unified, Real-time Object Detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.

[26] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.

[27] Bochkovskiy A, Wang C Y, Liao H Y. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. (2020-06-07).

[28] Liu W, Anguelov D, Erhan D, etc. SSD: Single Shot Multibox Detector[C]// Lecture notes in computer science2016. Cham: Springer, 2016, 9905: 21-37.

[29] 吕传龙,许玉格.基于改进 Faster R-CNN 的遮挡行人检测[J].惠州学院学报,2024,44(03): 10-15.

[30] 李彤晖,宋晓茹.基于改进 YOLOv5 的有遮挡行人检测方法研究[J].现代计算机,2022,28(24): 47-51.

[31] 梁秀满,周佳润,杨若兰.LPD-YOLO:轻量级遮挡行人检测模型[J].计算机工程与科学,2023,45(12): 2197-2205.

[32] 肖学文,王亚淑,刘康林.基于红外热像技术的过程设备无损检测[J].化工机械,2020,47(06): 742-746.

[33] Ge Z, Liu S, Wang F, etc. Yolox: Exceeding yolo series in 2021[J]. ArXiv Preprint ArXiv: 2107, 0830, 2021.

[34] 李敬兆,何娜,张金伟,等.基于 VMD 和 CNN-BiLSTM 的矿井提升电动机故障诊断方法[J].工矿自动化,2023,49(07): 49-59.

[35] Chen J, Huang R, Zhao K, etc. Multiscale convolutional neural network with feature alignment for bearing fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-10.

[36] 王婷,王其兵,何志方,等.基于改进的一维级联神经网络的异常流量检测[J].计算机应用与软件,2023,40(09): 320-326.

[37] Han J, Moraga C. The Influence of the Sigmoid Function Parameters on the Speed

of Back Propagation Learning[C]//Proc of the International Workshop on Artificial Neural Networks, Springer Verlag, 1995: 195-201.

[38] Abhishek L, Aravindhan S, Kumar T R. Accident Detection and Number Plate Recognition using Image Processing and Machine Learning[J]. International Journal of Recent Technology and Engineering (IJRTE), 2020, 8(5): 5079-5083.

[39] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533-536.

[40] Anand V, Gupta S, Koundal D, etc. Deep learning based automated diagnosis of skin diseases using dermoscopy[J]. Computers, Materials Continua,2022, 71(2): 3145-3160.

[41] 王维锋,邵琳骞,黄建鑫,等.基于改进 YOLOv5 的双模态融合行人检测方法[J/OL].吉林大学学报(工学版),1-11[2025-03-13].

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

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

[44] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7132-7141.

[45] Wang Q, Wu B, Zhu P, etc. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

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

[47] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[48] Yang L, Zhang R Y, Li L, etc. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR, 2021: 11863-11874.

[49] Li H L, Li J, Wei H B, etc. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[J]. Computer Science, 2022: 1-17.

[50] 王晓红,陈哲奇.基于 YOLOv5 算法的红外图像行人检测研究[J].激光与红外,2023,53(01):57-63.

[51] Hussain M, Khanam R. In-depth review of yolov1 to yolov10 variants for enhanced photovoltaic defect detection[C]//Solar. MDPI, 2024, 4(3): 351-386.

[52] 金莉莉,魏利胜.基于改进 YOLOv10n 的太阳能电池板缺陷检测方法[J/OL].重庆工商大学学报(自然科学版),1-9[2024-11-27].

[53] 成顺,李建荣,王志乾,等.基于 YOLOv8 轻量化水下光学图像识别算法[J/OL].激光与光电子学进展,1-20[2024-11-02].

[54] LI J, WEN Y, HE L. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.

[55] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 734-750.

[56] 闫世洋,罗素云.基于 SC-YOLOv8 的交通标志检测算法研究[J/OL].电子测量技术,1-9[2024-11-02].

[57] 刘忠,卢安舸,崔浩,等.基于改进 YOLOv8 的轻量化荷叶病虫害检测模型[J].农业工程学报,2024,40(19):168-176.

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

[59] Xue P, Chen T, Zhao X, etc. A Study of Lightweight Classroom Abnormal Behavior Recognition by Incorporating ODConv[C]//2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC). IEEE, 2023: 491-500.

[60] Qiao S, Chen L C, Yuille A. Detectors: Detecting objects with recursive feature

pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10213-10224.

[61] Zhang X, Liu C, Yang D, etc. RFAConv: Innovating spatial attention and standard convolutional operation[J]. arXiv preprint arXiv:2304.03198, 2023.

[62] Pan X, Ge C, Lu R, etc. On the integration of self-attention and convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 815-825.

[63] Wan D, Lu R, Shen S, etc. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.

[64] Misra D, Nalamada T, Arasanipalai A U, etc. Rotate to attend: Convolutional triplet attention module[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021: 3139-3148.

攻读硕士期间取得的学术成果

[1] **温福新**,许钢,凌成,等.基于改进 YOLOv10n 的红外图像遮挡行人检测算法[J].重庆工商大学学报(自然科学版),1-11[2025-05-16].

[2] **温福新**,许钢,凌成,等.基于改进 YOLOv5s 的红外图像行人检测算法[J].贵州大学学报(自然科学版),2025-03.(录用)

[3] 凌成,许钢,**温福新**.红外图像行人检测实时性与准确性的平衡研究[J].红外技术(CSCD),2024-07-19.(录用)

致谢

时光飞逝，岁月如梭，三年的时间转瞬即逝。回首往事，点点滴滴涌上心头，非常有幸能够在 2022 年进入安徽工程大学求学，在这里，我不仅学到了专业的知识，还认识了许多志同道合的朋友。这三年的求学生涯将会成为我人生中弥足珍贵的一段时光。在这里，我想向一路上帮助过我的人致以诚挚的感谢！

首先，我要衷心感谢我的导师许钢教授，非常荣幸能够成为您的学生，从论文选题到论文完成这一过程中，每一步都离不开许老师的细心指导。感谢您在我的学术道路上给予的悉心指导和无私帮助。您不仅传授了我宝贵的专业知识，还在科研思维和方法上给予了我极大的启发。您的严谨学风和敬业精神将永远是我学习的榜样。

其次，我要感谢实验室的所有同学们，苏世林、王博、凌成，在整个研究过程中，非常感谢您们的支持与帮助，我才得以完成课题研究。

最后，我要感谢我的父母，无条件的支持我的一切，做我最坚强的后盾，正因为如此，我才能继续完成我的梦想！

尚德敏学 · 唯实惟新