

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2220320149



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目

基于深度学习的无人机航拍行人检测

算法研究

论 文 作 者

居鸿源

校 内 导 师

江明（教授）

校 外 导 师

王鸿浩（高级工程师）

专业学位类别

电子信息

研究方向（领域）

控制工程

论文提交日期: 2025 年 5 月 15 日

分类号：TP391

单位代码：10363

密类级：公开

学 号：2220320149

题 目 基于深度学习的无人机航拍行人检测算法研究

题 目 RESEARCH ON PEDESTRIAN DETECTION
ALGORITHMS IN UAV AERIAL IMAGERY
BASED ON DEEP LEARNING

学 生 姓 名：居鸿源

校 内 导 师：江明（教授）

校 外 导 师：王鸿浩（高级工程师）

专业学位类别：电子信息

研究方向(领域)：控制工程

论文答辩日期：2025 年 4 月 29 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： 屠明强

日期： 2025 年 5 月 15 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：周鸿源

指导教师签名：江斌

日期： 2025 年 5 月 15 日

日期： 2025 年 5 月 15 日

基于深度学习的无人机航拍行人检测算法研究

摘 要

近年来, 基于深度学习的目标检测算法在无人机航拍行人检测领域展现出重要的应用价值, 尤其在公共安全、人流管理及安防监控等方面得到广泛应用。然而, 无人机航拍过程中, 飞行高度的变化导致行人目标尺度差异显著。在密集场景下, 行人之间易产生遮挡, 且小尺度行人目标易与背景混淆, 增加了特征提取难度, 进而影响检测精度。此外, 受限于无人机的载重能力, 其搭载的嵌入式设备通常性能有限, 而高精度算法往往伴随较大的模型权重和计算量, 导致资源利用率较高, 增加了在嵌入式设备上的部署难度。

本文针对上述问题, 分别提出了两种方法。首先, 针对航拍图像中行人多尺度变化、遮挡及背景复杂等情况导致目标检测算法精度低和漏检的问题, 提出了一种基于特征增强的航拍行人检测方法。该方法以通用目标检测算法中检测精度和速度都较为优异的YOLOv7 算法为基础。增加基于浅层特征图的小目标检测层, 增强了对小目标细节信息的捕捉能力。同时, 对 SPPCSPC(Spatial Pyramid Pooling Cross Stage Partial Channel)模块进行了重构, 提出了一种基于无参注意力机制 SimAM(Simple Attention Module)的特征增强模块 SPPFCSPCS7, 聚焦航拍图像中行人信息, 进一步提升算法的特征表达能力。进一步, 利用 NWD(Normalized Wasserstein Distance)损失函数作为边界框相似度的度量指标, 改善算法对小目标定位偏差的敏感性, 增强对行人小目标检测能力。

其次, 算法的检测精度通常与模型权重大小和计算量成正比相关, 针对高精度算法在嵌入式设备上的资源利用率高, 导致难以部

署到资源有限的嵌入式设备的问题，提出了一种航拍行人轻量化检测方法。具体而言：将深度可分离卷积与注意力机制相结合，设计了一种基于注意力机制的轻量化模块，旨在降低模型计算复杂度并保持特征提取能力；进一步，通过对算法中的冗余部分进行剪枝操作，在保证检测精度的同时减少了参数量和计算量。基于上述轻量化模块和剪枝技术，对基于特征增强的航拍行人检测方法进行了轻量化处理，使其更适合部署在资源受限的嵌入式设备上。

最后，本文设计并实现了基于 PyQt5 的航拍行人检测系统，对特征增强方法和轻量化方法进行了可视化分析。并且，根据无人机载重的能力，选用了英伟达系列的 Jetson Xavier NX 嵌入式设备搭载无人机进行模型部署实验。

关键词：深度学习，无人机，航拍行人，特征增强，轻量化

RESEARCH ON PEDESTRIAN DETECTION ALGORITHMS IN UAV AERIAL IMAGERY BASED ON DEEP LEARNING

ABSTRACT

In recent years, deep learning-based object detection algorithms have demonstrated significant application value in the field of UAV aerial pedestrian detection, particularly in public safety, crowd management, and security surveillance. However, during UAV aerial imaging, variations in flight altitude lead to substantial scale differences in pedestrian targets. In crowded scenes, pedestrians are prone to occlusion, and small-scale pedestrian targets are easily confused with the background, increasing the difficulty of feature extraction and consequently affecting detection accuracy. Additionally, due to the limited payload capacity of UAVs, the embedded devices they carry typically have constrained computational capabilities. High-precision algorithms often come with large model weights and high computational demands, resulting in increased resource utilization and greater deployment challenges on embedded devices.

This paper addresses the aforementioned issues by proposing two distinct methods. Firstly, to tackle the problems of low accuracy and missed detections in general object detection algorithms caused by multi-scale variations, occlusions, and complex backgrounds in aerial images, a feature-enhanced aerial pedestrian detection method is introduced. This method is based on the YOLOv7 algorithm, which is known for its superior detection accuracy and speed in general object detection tasks. An additional detection layer for small objects, based on shallow feature maps, is incorporated to enhance the capability of capturing detailed information of small targets.

Moreover, the SPPCSPC (Spatial Pyramid Pooling Cross Stage Partial Channel) module is restructured, and a feature enhancement module named SPPFCSPCS7, based on the parameter-free attention mechanism SimAM (Simple Attention Module), is proposed. This module focuses on pedestrian information in aerial images, further improving the feature representation capability of the algorithm. Additionally, the NWD (Normalized Wasserstein Distance) loss function is utilized as a metric for bounding box similarity, reducing the algorithm's sensitivity to localization deviations of small targets and enhancing its detection capability for small pedestrian objects.

Secondly, the detection accuracy of algorithms is often positively correlated with the size of model weights and computational complexity. To address the issue of high resource utilization of high-precision algorithms on embedded devices, which makes deployment on resource-constrained embedded devices challenging, a lightweight aerial pedestrian detection method is proposed. Specifically, depthwise separable convolution is combined with an attention mechanism to design a lightweight module based on the attention mechanism, aiming to reduce the computational complexity of the model while maintaining feature extraction capabilities. Furthermore, by pruning redundant parts of the algorithm, the number of parameters and computational load are reduced without compromising detection accuracy. Based on the aforementioned lightweight module and pruning techniques, the feature-enhanced aerial pedestrian detection method is further optimized for lightweight deployment, making it more suitable for resource-constrained embedded devices.

Finally, this paper designs and implements an aerial pedestrian detection system based on PyQt5, providing a visual analysis of the feature enhancement and lightweight methods. Furthermore, considering the payload constraints of the drone, an NVIDIA Jetson Xavier NX embedded device

was selected for onboard model deployment experiments.

Ju Hongyuan (Electronic Information)

Supervised by Professor Jiang Ming

KEY WORDS: deep learning, unmanned aerial vehicle, aerial pedestrian, feature enhancement, lightweight

目录

摘要.....	I
ABSTRACT.....	III
第 1 章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.2.1 通用目标检测算法.....	3
1.2.2 小目标检测算法.....	6
1.2.3 轻量化检测算法.....	7
1.3 论文研究内容及架构安排.....	8
1.3.1 主要研究内容.....	8
1.3.2 本文组织结构.....	10
第 2 章 无人机航拍行人检测相关理论基础.....	11
2.1 航拍行人检测的问题.....	11
2.1.1 航拍图像中行人特点.....	11
2.1.2 航拍行人检测任务难点.....	12
2.2 目标检测算法研究.....	13
2.2.1 双阶段目标检测算法.....	13
2.2.2 单阶段目标检测算法.....	14
2.3 特征增强算法研究.....	17
2.3.1 多尺度特征融合.....	17
2.3.2 注意力机制.....	19
2.4 轻量化算法研究.....	20
2.5 数据集与评价指标.....	22
2.5.1 数据集.....	22
2.5.2 评价指标.....	23
2.6 本章小结.....	24
第 3 章 基于特征增强的航拍行人检测方法.....	25

3.1 改进 YOLOv7 模型结构.....	25
3.2 多尺度检测策略.....	27
3.2.1 小目标检测层.....	27
3.2.2 特征增强模块.....	28
3.3 NWD 损失函数	31
3.4 实验结果与分析.....	32
3.4.1 实验环境及参数配置.....	32
3.4.2 消融实验.....	33
3.4.3 YOLO 系列算法对比实验.....	33
3.4.4 双阶段算法对比实验.....	35
3.4.5 VisDrone2019 数据集通用性实验	36
3.5 航拍行人检测系统的可视化分析.....	37
3.5.1 PyQt5 简介.....	37
3.5.2 系统功能与检测流程.....	38
3.5.3 可视化分析.....	38
3.6 本章小结.....	40
第 4 章 航拍行人轻量化检测方法.....	41
4.1 轻量化模块.....	41
4.1.1 FCBAM 模块.....	41
4.1.2 轻量化卷积模块.....	43
4.2 冗余网络结构剪枝.....	44
4.3 航拍行人轻量化检测模型.....	46
4.4 实验与分析.....	47
4.4.1 消融实验.....	47
4.4.2 不同检测算法对比实验.....	47
4.4.3 改进方法通用性实验.....	49
4.4.4 检测系统可视化分析.....	50
4.5 本章小结.....	50
第 5 章 航拍行人检测算法部署实验.....	51

5.1 嵌入式设备选择.....	51
5.2 Jetson Xavier NX 嵌入式硬件介绍	52
5.3 嵌入式系统环境搭建.....	52
5.4 模型部署.....	53
5.4.1 TensorRT 加速推理实现.....	53
5.4.2 模型大小分析.....	54
5.4.3 模型检测性能分析.....	55
5.4.4 无人机航拍行人实时检测.....	56
5.5 本章小结.....	57
第 6 章 总结与展望.....	58
6.1 工作总结.....	58
6.2 未来展望.....	59
参考文献.....	60
攻读硕士学位期间所发表的论文及其它研究成果.....	66
致谢.....	67

第 1 章 绪论

1.1 研究背景与意义

近年来，随着计算机视觉在人工智能领域的快速发展，目标检测算法取得了巨大的发展。它在自动驾驶^[1]、图像分类^[2]、目标跟踪与检测^[3]等领域得到了广泛的应用。对于无人机航拍图像的目标检测算法也广泛应用于测绘航测^[4]、城市规划^[5]、森林防护^[6]等领域。

其中，无人机针对行人的目标检测任务的应用场景如图 1-1 所示，在交通管理领域，利用无人机航拍检测行人的流量和走向，结合车辆流量信息，交通管理部门可以更全面地了解交通状况，优化信号灯配时，提高道路的通行效率，缓解交通拥堵。在应急救援领域，无人机航拍行人目标检测可以迅速对受灾区域进行空中侦察，快速定位被困行人的位置和数量，为救援人员制定救援方案提供重要信息，提高救援效率，争取在最短时间内营救更多生命。在安防监控领域，无人机航拍可以从空中俯瞰，快速检测行人数量、分布及行为，提前发现潜在的安全隐患，如人群聚集引发的拥挤踩踏风险，相关部门可及时采取措施进行疏导。在商业服务领域，通过无人机检测行人分布情况，景区管理者可以实时了解各景点的游客流量，合理引导游客游览，避免局部区域游客过度集中，提升游客的游览体验。



图 1-1 航拍行人检测的应用场景

在无人机航拍图像中，由于飞行高度的限制，行人目标通常呈现为小目标。小目标是指目标在图像中的成像尺寸较小。目前，学术界对小目标的定义主要有两种广泛认可的标准：绝对尺度和相对尺度。绝对尺度定义基于 COCO 数据集^[7]将像素面积小于 32×32 的目标归类为小目标，具体划分标准如表 1-1 所示。相对尺度定义则由国际光学工程学会提出，在 256×256 像素的图像中，若目标的像素面积小于 80 像素，或目标面积占图像总面积的比例低于 0.12%，则该目标被视为小目标。

表 1-1 绝对尺寸大小的划分

目标大小	尺寸大小
小目标	小于 32×32
中目标	大于 32×32 且小于 96×96
大目标	大于 96×96

无人机航拍图像中的行人通常具有以下特点：

(1) 目标尺寸变化大：无人机拍摄高度不同，导致行人在画面中的尺寸多样。

(2) 分布密集、易受遮挡^[8]：在城市街道、广场等公共场所进行航拍时，行人往往较为集中，分布密集，且行人之间、行人与建筑物、树木等物体之间容易发生遮挡。

(3) 背景复杂、光照变化大：航拍场景中背景复杂多样，包含建筑物、街道、车辆、植被等多种物体，背景信息可能与行人特征相似，且不同时间、不同天气条件下，光照强度和角度变化明显，会导致行人的颜色、亮度等特征发生变化。

根据以上可知，航拍图像中的行人具有尺度变化大、目标遮挡严重、背景复杂等特点，这对目标检测算法提出了更高的要求。目前，在基于深度学习的目标检测领域，大多数目标检测模型主要是针对自然图像数据集开发的，这些模型在无人机航拍行人检测任务中表现不佳。因此，许多学者通过改进传统基于深度学习的通用目标检测算法来提高航拍目标的检测精度。然而，精度的提升通常伴随着模型权重大小和计算量的增加，导致模型难以部署在资源受限的嵌入式设备上。

因此，针对多数通用目标检测算法存在的检测精度低和模型部署难度大等问题，本文分别提出了两种改进方法。具体而言，针对无人机航拍图像中行人

目标多尺度、目标遮挡以及背景复杂等导致的检测精度低的问题，本文主要通过多尺度特征融合和注意力机制等技术，提升了通用目标检测算法对于航拍图像中行人的检测精度。其次，针对模型权重大小和计算复杂度增加导致的难以部署到资源有限的嵌入式设备的问题，本文利用轻量化模块和模型剪枝等方法，对改进后的算法进行优化，有效降低了模型的参数量和计算量，从而减少了模型对硬件资源的需求，提升了其在嵌入式设备上的部署可行性。

1.2 国内外研究现状

图 1-2 展示了目标检测算法的分类情况，主要划分为传统方法和基于深度学习的方法两大类。传统目标检测算法通常遵循三个关键步骤：首先生成候选区域，接着进行人工特征提取，最后选择合适的分类器。这种方法高度依赖人工设计的特征，导致检测精度和速度相对较低，同时模型的泛化能力和鲁棒性也较为有限。相比之下，基于深度学习的目标检测算法利用神经网络自动学习特征，极大地提升了检测性能，使其成为当前研究的主流方向。本文主要关注基于深度学习的目标检测算法，重点从通用目标检测、小目标检测以及轻量化检测三个方面展开深入研究。

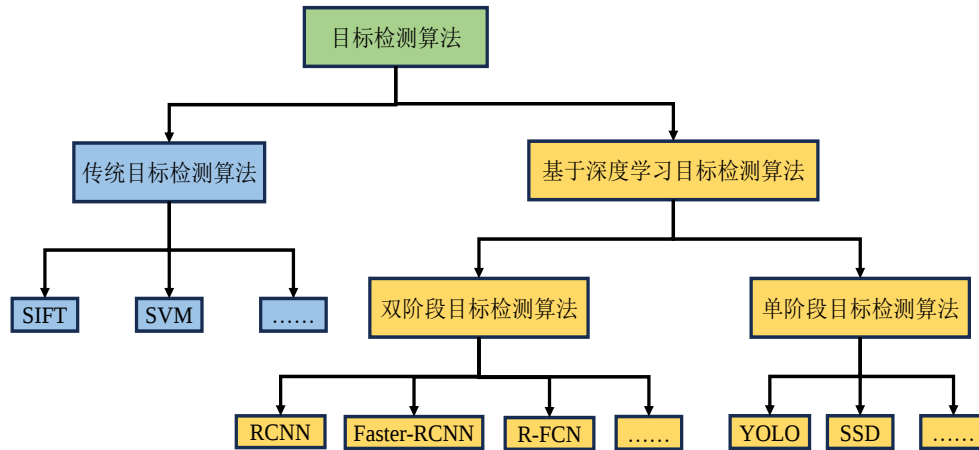


图 1-2 目标检测算法分类

1.2.1 通用目标检测算法

目前，基于深度学习的目标检测算法主要分为单阶段目标检测算法和双阶段目标检测算法。二者的核心区别在于是否生成候选区域^[9](Region Proposals)。

双阶段算法首先生成候选框，再对其进行分类和回归，因此检测精度较高，但检测速度相对较慢；而单阶段算法则直接在特征图上进行回归预测，无需生成候选区域，因而具有更低的计算复杂度和更快的检测速度。针对这两类算法，研究者们已开展了大量深入的研究，并取得了丰硕的成果。

在单阶段目标检测算法领域。Redmon 等人^[10]首次提出了 YOLO 算法，其网络结构主要由卷积层和全连接层组成。该算法的创新之处在于摒弃了传统方法中生成候选框的步骤，从而在检测速度上实现了明显提升，同时仍能保持较高的检测精度。然而，YOLOv1 算法在目标定位和检测精度方面仍存在不足。为了解决这些问题，Redmon 等人^[11]进一步提出了 YOLOv2 算法，该算法引入了多尺度训练策略，不仅提升了检测精度和速度，还保留了 YOLO 算法高效简洁的优势。随后，Redmon 等人^[12]又提出 YOLOv3 算法，该版本通过增加边界框数量并采用三个不同尺度的特征图进行预测，进一步优化了检测性能。这一改进使得 YOLOv3 在多尺度目标检测中表现尤为突出，尤其是在小目标检测任务中展现了更强的适应性和鲁棒性。BOCHKOVSKIY 等人^[13]提出 YOLOv4 算法，该算法主要改进包括：引入了 mosaic 数据增强策略；采用 SPP(Spatial Pyramid Pooling)模块有效扩大模型感受野；采用全新的骨干网络结构以及激活函数；采用全新的特征金字塔与路径聚合网络结构；采用 DropBlock 正则化。相较于 YOLOv3，YOLOv4 在维持检测速率的基础上提升了检测精度。此后，YOLO 系列算法不断被推出，其检测性能在通用数据集上不断得到改进。单阶段目标检测算法的另一经典之作为 Liu 等人^[14]提出的 SSD(Single Shot MultiBox Detector)算法，该算法采用多尺度特征提取策略，通过构建分层检测网络，在不同层级同时执行特征提取、类别预测和边界框回归任务。此后，DSSD(Deconvolutional single shot detector)算法^[15]被提出，通过将残差网络与 SSD 算法进行结合，获得额外的上下文信息，增强 SSD 低层特征图的表达能力，从而得到更加精准的检测结果。Lin 等人^[16]提出 RetinaNet 算法，该算法通过引入 Focal Loss 并结合 FPN(Feature Pyramid Network)，成功解决了目标检测中的类别不平衡问题，提高了在目标密集的场景下检测的效果。

在双阶段目标检测算法领域。Girshick 等人^[17]提出了 R-CNN(Region-based Convolutional Neural Networks)算法，该算法在卷积神经网络的基础上，使用选

择性搜索等方法从输入图像中提取可能包含目标的候选区域，然后对每个候选区域分别进行特征提取和分类，显著提升了检测精度。其训练过程需要分阶段进行，包括候选区域生成、特征提取、分类和边界框回归，导致训练速度缓慢。Girshick 等人^[18]提出了 Fast R-CNN(Fast Region-based Convolutional Neural Network)算法，该算法通过引入 Softmax 分类器取代传统的 SVM(Support Vector Machine)方法，与 R-CNN 相比，Fast R-CNN 不仅简化了训练流程，还实现了端到端的训练和测试，避免了 R-CNN 中重复的特征提取过程，提升了目标分类的效率。Ren 等人^[19]提出了 Faster R-CNN(Faster Region-based Convolutional Neural Network)算法，该算法通过引入区域候选网络，能够直接在特征图上生成高质量的候选区域，取代了传统方法中耗时的候选框生成过程，提升了检测的速度和精度。Dai 等人^[20]提出 R-FCN(Region-based Fully Convolutional Networks)算法，R-FCN 的核心思想是将卷积特征的计算共享到整个图像上，而不是对每个候选区域单独计算特征，从而大幅提升了计算效率。实验结果表明，R-FCN 在检测精度和速度方面均优于 Faster R-CNN，尤其是在处理大规模数据集时表现尤为突出。He 等人^[21]提出 Mask R-CNN(Mask Region-based Convolutional Neural Network)算法，对于待检测目标，为其生成分割掩码，增加一个网络分支用于 ROI 预测分割，并以 ROI Align 层代替 ROI 层，提高分支的检测精度，Mask R-CNN 在小目标和复杂场景中取得了不错的效果。

除此之外，还有其它优化算法同样提高了目标检测的效果。He 等人^[22]提出了空间金字塔池化(Spatial Pyramid Pooling, SPP)结构，SPP 允许网络接受任意尺寸的输入图像，并生成固定长度的特征向量。这一改进使得卷积神经网络的应用更加灵活，能够适应不同分辨率的输入数据，从而提升网络的多尺度检测能力。Zhang 等人^[23]提出了 RefineDet 算法，RefineDet 先利用单阶段算法生成高质量的候选框，再通过双阶段算法对候选框进行精细调整，从而实现精度与速度的平衡。该算法结合了双阶段目标检测算法和单阶段目标检测算法的优点，在保持较高检测速度的同时提升了检测精度。Li 等人^[24]提出 TridentNet 算法，设计了具有不相同感受野的多个分支，通过多分支结构和权重共享，提升了目标检测的尺度适应能力。

1.2.2 小目标检测算法

上述通用目标检测算法大多针对常规数据集提出，对航拍小目标检测效果并不理想。小目标检测是目标检测领域中的一个重要研究方向，由于小目标在图像中占据较少的像素、特征信息不足，导致其检测难度较大。近年来，随着深度学习技术的发展，研究者在通用目标检测算法的基础上，提出了许多针对小目标检测的改进算法。

大多数单阶段检测算法在提取图像特征时未考虑到小目标信息易丢失的问题，因此对小目标的检测效果不理想。为了提升对小目标的检测效果，Yin 等人^[25]提出了 FD-SSD(Feature Fusion and Dilated Convolution Single Shot Multibox Detector)算法。该算法通过以下三方面改进提升检测性能：利用多层特征融合模块，增强浅层特征图的语义信息；采用多分支残差空洞卷积模块，提升上下文信息的捕捉能力；引入可变形卷积，更好地适应小目标的形状变化。Liang 等人^[26]提出了一种基于特征融合和缩放的单次检测器，通过融合不同层级的特征图，增强模型对多尺度目标的检测能力。实验结果表明，FS-SSD 在保持较高检测速度的同时，明显提升了检测精度，尤其是在小目标检测和复杂场景下表现优异。赵^[27]在 YOLOv5 的 Neck 部分引入了 Swin Transformer 模块，使网络模型能够将得到的全局信息与卷积神经网络提取的特征信息高效融合，解决了航拍图像中目标尺寸小、差异大的问题。Zhu 等人^[28]提出了 TPH-YOLOv5，设置四个检测端来应对不同尺度的物体，集成注意力模型，以在具有密集对象的场景中找到重点区域。余等人^[29]在 YOLOv7 中引入了集中特征金字塔模块，并添加了 ACmix(Attention Convolution Mixer)注意力机制，解决遥感图像因目标分布密集，背景复杂所导致的漏检误检问题。Zhao 等人^[30]采用多检测头和 CBAM(Convolutional Block Attention Module)注意力机制相结合用于提取不同尺度的特征，在 YOLOv7 网络中添加了新的金字塔池化模块，并结合了 Soft NMS 和 Mish 激活函数，提高网络识别重叠和遮挡目标的能力。双阶段算法在检测小目标方面也存在特征信息易丢失的问题。为解决这一问题，Zhang 等^[31]分析了 Faster R-CNN 算法，并发现了检测小目标结果不理想的原因：神经网络的特征图分辨率不足，导致网络在学习过程中容易忽略小目标位置信息。为解决这个问题，Liu 等人^[32]提出了一种小目标检测算法，从不同网络层中提取多分辨率特

征图，并将这些特征与卷积特征集成在一起，从而增强网络对小目标的特征提取能力。

1.2.3 轻量化检测算法

上述算法在目标检测领域取得了显著进展，有效提升了检测精度和性能。然而，这些方法通常以增加模型计算复杂度为代价，导致计算资源和存储需求上升，难以部署在资源受限的设备上。为此，研究者提出了轻量化神经网络的设计思路，通过优化网络结构和压缩模型，在保持较高性能的同时降低计算和存储开销，从而满足实际应用需求。

在优化网络结构方面，谷歌团队提出的 MobileNet 系列网络成为了研究者设计轻量化网络的首选。MobileNetV1 网络将传统卷积替换为深度可分离卷积，使得卷积操作的计算量降低到原来的 1/8 左右。MobileNetV2^[33]在 MobileNetV1 的基础上，采用倒残差结构作为网络的主要模块，并增加了捷径分支，提高了检测性能。MobileNetV3^[34]在 MobileNetV2 网络的基础上，引入通道注意力机制，让网络权重参数集中于更重要的通道，进一步减少网络的参数量。此外，ShuffleNet 系列^[35-36]利用分组卷积来降低参数量，并采用通道混洗(Channel Shuffle)操作重新排列通道信息，促进不同组之间的信息交互，从而提升模型的表现能力。这些轻量化网络的提出，为更多更优秀的网络模型提供了思路。

在模型压缩方面，可以分为模型剪枝^[37]、模型量化、知识蒸馏等几个方向。模型剪枝是通过移除神经网络中的冗余参数或结构，减少模型的计算复杂度和存储需求，同时尽量保持模型的性能。Liu 等^[38]提出了 Network Slimming 方法，通过学习通道缩放因子来自动识别并剪枝不重要的通道，从而实现模型的轻量化，并且在训练过程中引入 L1 正则化(Lasso Regularization)，使得通道缩放因子稀疏化，进而移除缩放因子接近于零的通道。该方法在保持较高模型性能的同时减少计算量和参数量。模型量化是通过降低模型参数的数值精度，从而减少模型的计算复杂度和存储需求，同时尽量保持模型的性能。石等人^[39]针对无人机航拍检测提出了一种改进后的 YOLOv4 模型。他们成功运用感知量化技术，对权重实施了 INT8 量化处理，在保持精度的前提下缩减了网络模型的规模和参数量。知识蒸馏^[40]是一种模型轻量化技术，通过将高性能复杂教师网络的

知识迁移至轻量级学生网络，实现模型压缩与性能优化的双重目标。其核心在于引导学生网络模仿教师网络的输出分布或中间特征，从而在降低参数量的同时维持较高的检测精度。该技术在目标检测领域成效显著，尤其适用于移动端及嵌入式设备等资源受限场景。

综上所述，大多数通用的目标检测算法在无人机航拍场景中面临两大主要挑战：其一，由于航拍图像中目标尺寸普遍较小且背景复杂，通用算法在特征提取过程中容易丢失小目标的细节信息，导致检测精度不足，难以满足无人机航拍领域对高精度检测的需求；其二，无人机平台的计算资源、存储空间和功耗均有限，而复杂的航拍目标检测模型通常具有较高的计算复杂度和参数量，难以在资源受限的嵌入式设备上高效部署。因此，如何在提高小目标检测精度的同时，保持网络模型的轻量化特性，成为当前无人机航拍目标检测领域的重要研究问题。

1.3 论文研究内容及架构安排

1.3.1 主要研究内容

本文主要研究内容如图 1-3 所示。针对无人机航拍行人检测任务，主要面临两个问题：多数通用目标检测算法对航拍图像中行人的特征提取能力较差带来的检测精度低的问题；模型权重大小和计算复杂度较高，难以部署到资源有限嵌入式设备上的问题。本文以通用目标检测算法中的 YOLOv7 算法^[41]为基准，提出两种方法来解决上述两种问题，

针对多数通用目标检测算法对航拍图像中行人特征表达能力不足的问题，本文提出了一种基于特征增强的航拍行人检测方法。该方法增加了基于浅层特征图的小目标检测层，可获得航拍图像中更多行人目标的细节信息，应对多尺度检测任务。其次，对 SPPCSPC^[42]模块进行了重构，提出了一种基于注意力的特征增强模块 SPPFCSPC7，该模块利用注意力获取图像中感兴趣区域并加强特征信息融合，提升了对密集目标的检测能力，减少了背景信息的干扰。对于网络的后端处理，采用 NWD^[43]替换 IOU 度量边界框之间的相似度，有效改善了 IOU 对小目标位置偏移敏感的问题。

针对模型权重大小和计算复杂度较大，导致难以部署到资源有限的嵌入式设备的问题，本文提出了一种轻量化航拍行人检测方法。首先，该方法引入了一种基于注意力机制的轻量化模块，通过结合深度可分离卷积和空间、通道注意力机制，在减少模型计算量和参数量的同时，保留了对航拍图像中行人目标的特征提取能力。此外，针对原始网络在航拍图像任务中存在的冗余问题，本文对大目标的特征提取与特征融合部分以及预测分支进行了剪枝，移除了不必要参数。最后，对基于特征增强改进的 YOLOv7 算法进行轻量化处理，构建轻量级特征融合网络，降低了模型权重大小和计算复杂度。

最后，本文设计并实现了基于 PyQt5 的航拍行人检测系统，对特征增强方法和轻量化方法进行了可视化分析。通过该系统，能够直观地展示两种方法在航拍图像中的检测效果。并且，本文还将改进后的算法部署到 Jetson Xavier NX 嵌入式平台上，并搭载无人机进行航拍行人实时检测。

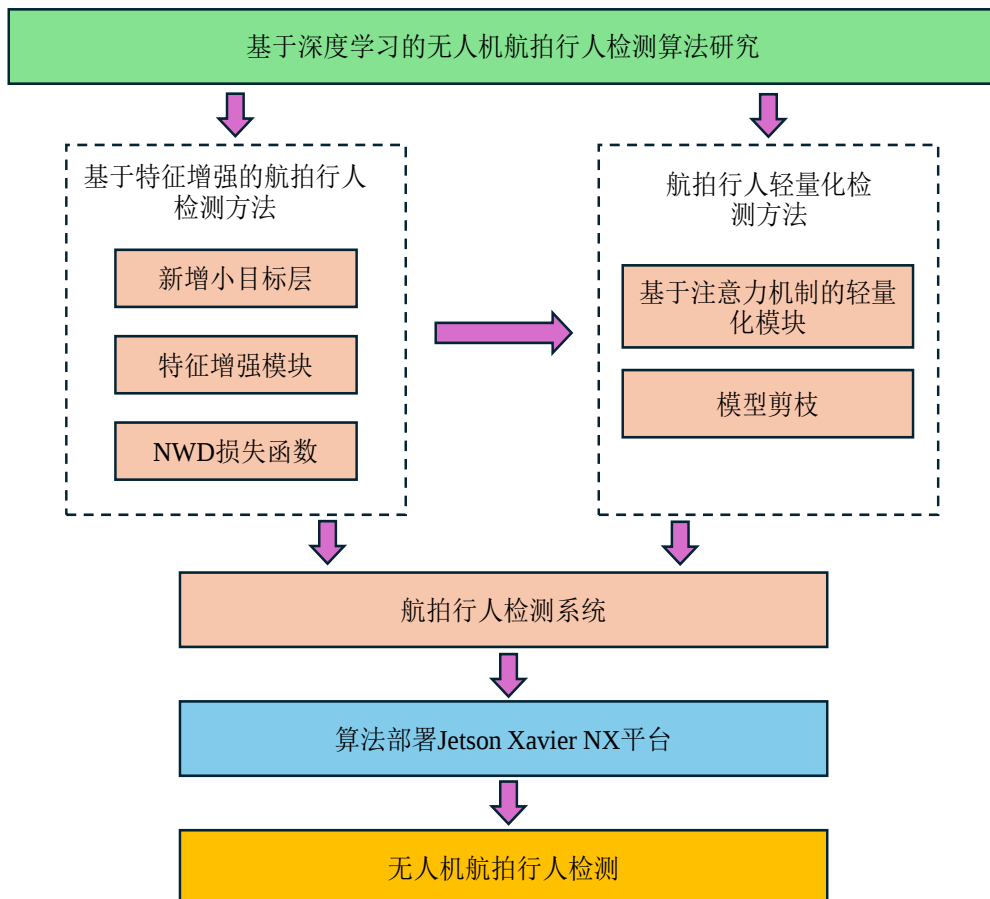


图 1-3 研究内容框图

1.3.2 本文组织结构

本文主要分为六个章节，具体内容如下：

第 1 章概述了航拍图像行人检测的研究现状及意义，系统梳理了通用目标检测、小目标检测及轻量化检测算法的研究进展，并阐明了本文的研究内容与结构安排。

第 2 章主要通过对无人机航拍行人目标检测的技术难点进行了分析，引出了单阶段和双阶段目标检测算法以及模型相关的特征增强和轻量化方法，最后介绍了本文采用的数据集和评价指标。

第 3 章提出了一种基于特征增强的航拍行人检测方法，旨在改善通用目标检测算法 YOLOv7 在航拍行人检测任务中精度低、漏检率高等问题。在航拍行人数据集上进行了多组实验，实验结果表明所提方法明显提升了模型的检测精度。此外，设计了无人机航拍行人检测系统，通过该系统可视化分析模型的实际检测效果，进一步验证了算法的实用性。

第 4 章提出了一种航拍行人轻量化检测方法，旨在改善目标检测算法因模型权重较大、计算复杂度较高而难以在资源受限的嵌入式设备上高效部署的问题。最后，在航拍行人数据集上进行了实验验证，并在航拍行人检测系统中进行了可视化分析。实验结果表明，改善后算法有效的降低了模型权重大小和计算复杂度，使其更适合部署在资源受限的嵌入式设备上。

第 5 章实现了改进算法在嵌入式平台的部署。基于无人机载重能力分析，选用 Jetson Xavier NX 模块作为硬件平台，开展航拍行人检测算法的部署实验，并搭载无人机进行了实际测试。

第 6 章主要对所做工作进行回顾与总结，并对未来的工作方向提出相应的规划设想。

第 2 章 无人机航拍行人检测相关理论基础

在第 1 章中，本文探讨了航拍小目标检测的研究背景、研究意义以及当前的研究现状，为后续提出的方法奠定了坚实的理论基础和实验依据。本章将继续作为后续章节的理论基石，从无人机航拍行人检测的问题出发，探讨深度学习的相关基础理论。

2.1 航拍行人检测的问题

2.1.1 航拍图像中行人特点

无人机航拍图像幅面较大，行人在图像中所占像素比例较小，导致多数通用目标检测算法在特征提取过程中，容易导致行人信息的丢失。此外，如图 2-1 所示，无人机航拍图像中行人目标还面临多尺度变化、遮挡以及背景复杂等特点，这些因素进一步增加了行人目标的特征提取难度。

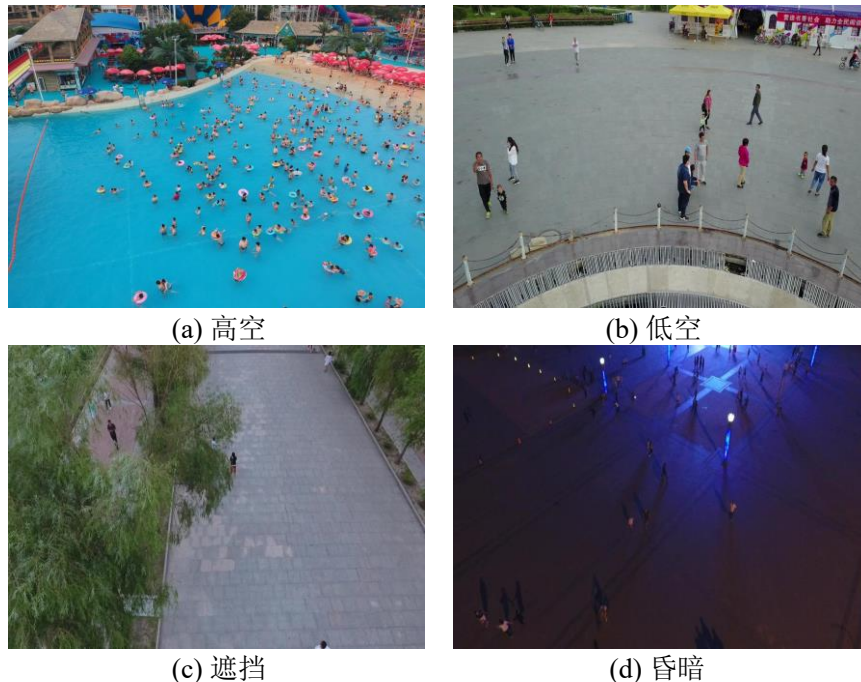


图 2-1 无人机航拍图像

图中 a、b、c、d 所示，无人机航拍图像中行人呈现为小目标，由于目标尺寸较小，其特征信息相对有限，增加了通用目标检测算法的漏检概率。如图 2-1 中 a、b 所示，当无人机飞行高度发生改变时，航拍图像中行人目标所占像素也

在变化，这种尺度不一的现象导致算法难以对不同尺度的行人目标进行精确识别和定位。如图 2-1 中 b、c 所示，在无人机航拍视角下，行人往往以聚集形式出现，容易发生遮挡现象。遮挡主要分为两类：一是行人之间的相互遮挡，二是背景信息对行人的遮挡。这些遮挡情况会明显增加算法漏检的概率。如图 2-1 中 d 所示，无人机航拍场景复杂多样，背景信息容易与行人目标混淆。例如，在光线昏暗的场景中，行人与背景的对比度较低，难以区分。这种复杂性使得通用目标检测算法在无人机航拍图像中也容易出现误检和漏检问题。

2.1.2 航拍行人检测任务难点

从上述航拍图像中行人特点的分析可知，对于通用目标检测算法，在航拍场景下的行人检测面临诸多挑战，主要体现在多尺度变化、遮挡问题以及背景复杂性等方面。以下从这三个特点出发，详细分析航拍行人检测的难点：

（1）多尺度变化：要求检测算法具备较强的多尺度检测能力，能够同时处理不同尺度目标，尤其需要改善对小目标特征提取不足的问题。

（2）遮挡：要求检测算法具备较强的局部特征提取能力和上下文推理能力，能够通过可见部分推断被遮挡目标的存在和位置。

（3）背景复杂：要求检测算法具备强大的特征区分能力，能够有效分离行人目标与背景干扰，同时应对光照变化和动态背景的挑战。

此外，对于无人机航拍行人检测任务，算法的实用性还受到部署条件的限制。为了提升检测性能，通常需要增强通用目标检测算法的特征表达能力，但这往往会增加算法的权重大小和计算复杂度。在资源受限的无人机设备上，计算能力、存储空间均受到严格限制，这使得在保证检测精度的同时实现高效部署变得尤为困难。

综上，本文主要结合特征增强、轻量化等策略提出了两种航拍行人检测方法：结合多尺度特征融合、注意力机制等技术，设计了一个基于特征增强的航拍行人检测方法，提高模型对航拍行人的检测精度；结合轻量化网络设计、模型剪枝等技术，设计了一个轻量化的航拍行人检测方法，降低无人机设备的资源利用率。

2.2 目标检测算法研究

目标检测作为计算机视觉的核心任务之一，旨在从图像或视频中定位并识别出感兴趣的目标。近年来，随着深度学习的快速发展，目标检测算法取得了显著进步，涌现出众多性能优异的通用检测算法。然而，面对种类繁多的算法，如何根据具体应用场景和需求选择合适的算法，成为研究者面临的关键问题。本文将重点分析几种具有代表性的通用检测算法，从而选取合适的目标检测算法用于无人机航拍行人检测研究。

2.2.1 双阶段目标检测算法

双阶段目标检测算法将目标检测任务分为两个步骤：首先生成候选区域，随后执行类别判定与边界框回归。这种两阶段的策略提升了目标检测的准确性和稳定性，使模型能够更精准地捕捉目标信息。典型的双阶段目标检测算法包括 Faster R-CNN 和 Mask R-CNN。下文将通过对这两种算法的详细介绍，阐述双阶段目标检测算法的具体流程。

（1）Faster R-CNN 算法

Faster R-CNN 算法的提出旨在提升 Fast R-CNN 模型的目标区域筛选效率。该算法采用 RPN 替代传统的选择性搜索方法，用于生成候选区域。RPN 能够灵活调整候选框的数量，并根据候选框的得分和位置回归参数筛选出更高质量的候选目标区域，从而为后续的目标分类和位置精确定位提供支持。Faster R-CNN 实现了端到端的训练，将区域选择、特征分类和边界回归等环节统一到一个深度学习框架中进行联合优化，提升了目标检测系统的速度和精度。算法流程如图 2-2 所示。首先，将待检测图像输入 Faster R-CNN 模型，通过共享卷积神经网络(CNN)提取特征图，并利用区域候选网络从特征图中生成候选区域。接着，ROI 池化层对特征图和 RPN 生成的候选区域进行池化处理，输出固定大小的特征图。最后，通过全连接层完成目标分类和边界框回归。尽管 Faster R-CNN 明显提升了检测速度，但仍存在一些局限性：由于 Faster R-CNN 的锚框尺寸固定，在处理多尺度目标时灵活性不足，尤其是对小目标的检测效果较差；ROI 在获取目标边界框时的性能不足，影响了网络的定位精度。

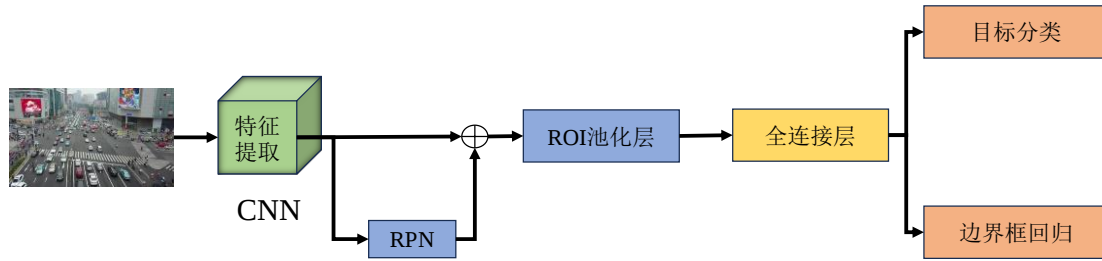


图 2-2 Faster R-CNN 目标检测算法流程

(2) Mask R-CNN 算法

Mask R-CNN 模型是对 Faster R-CNN 算法的进一步优化。该模型通过引入 mask 掩码分支，实现了更精细的实例分割和目标检测。同时，Mask R-CNN 采用 ROI Align 替代 ROI Pooling，以减少特征信息的丢失，从而获取更精确的特征表示。此外，模型使用二分类损失函数取代多任务损失函数，有效消除了不同类别掩码之间的竞争关系。算法流程如图 2-3 所示，Mask R-CNN 算法通过引入额外的 mask 掩码分支网络，提升了对目标的定位精度和识别性能，从而进一步提高了检测的精确度。

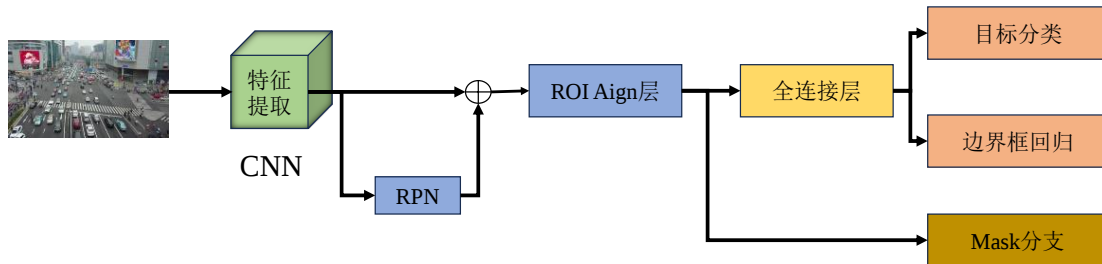


图 2-3 Mask R-CNN 目标检测算法流程

2.2.2 单阶段目标检测算法

对于航拍行人检测任务，涉及移动的目标检测和需要实时反馈的情况较多，对算法速度方面有所要求，单阶段算法更符合本文的要求，在实际应用中，能够更好地满足这种实时性需求。在单阶段目标检测算法中，YOLO 系列算法在检测速度和精度之间有着很好的平衡，并被广泛应用于航拍图像目标检测中。YOLO 模型具有灵活的参数配置，对于航拍图像中行人的检测，其误检与漏检率明显降低。因此，本文选择 YOLO 系列算法作为课题研究的模型基础，针对航拍图像特点进行模型改进。

目前，YOLO 系列算法已拥有 YOLOv5、YOLOv7、YOLOv8 等多个版本。

本文在开展无人机航拍行人检测研究时，相较于其它 YOLO 算法，YOLOv7 对于小目标检测具有更好的检测性能。因此，本文以 YOLOv7 算法为基础，重点研究了以下两种方法：一是基于特征增强的航拍行人检测方法，旨在提升算法对行人的检测能力；二是航拍行人检测算法的轻量化方法，以解决模型在无人机设备上部署时的资源限制问题。此外，为了验证所提出方法的通用性，本文在后续新的 YOLO 系列算法发布时，也进行了相关实验，进一步验证了所提方法在不同模型上的适用性。下文将通过对 YOLOv5 和 YOLOv7 两种模型结构的对比，阐述 YOLOv7 在航拍行人检测任务中的适用性。

(1) YOLOv5

YOLOv5 提供了四种不同规模的网络模型，分别是 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x。其中 YOLOv5s 的模型参数量最小，其余三种模型均在其基础上通过增加网络宽度和深度进行扩展，以满足不同场景下的性能需求。YOLOv5s 网络结构如图 2-4 所示，分为输入端、特征提取网络、特征融合网络和检测端四个部分。

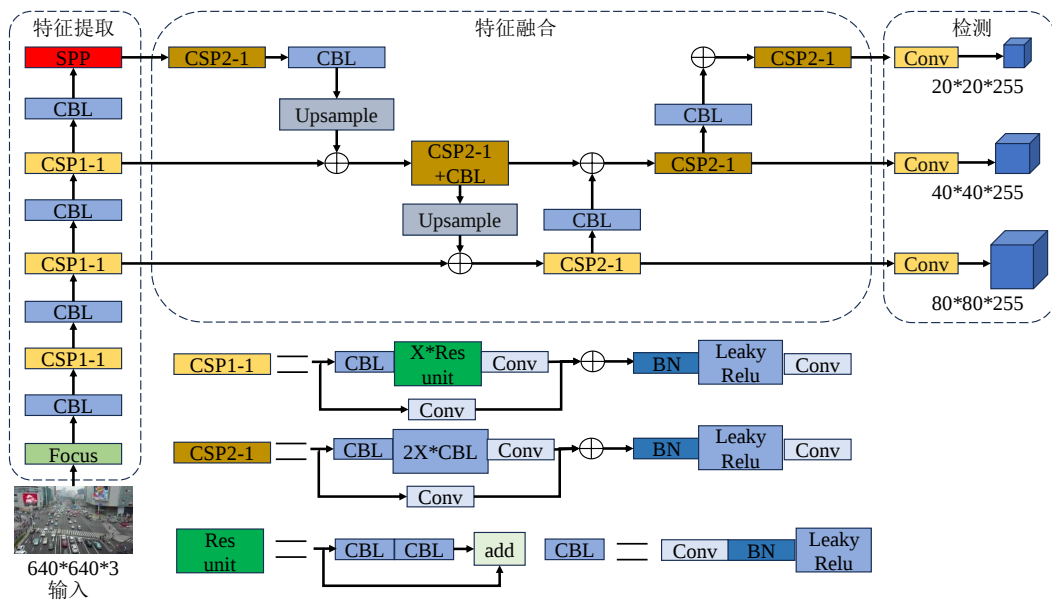


图 2-4 YOLOv5s 网络结构示意图

(a) 输入端：YOLOv5 采用 Mosaic 数据增强技术，丰富了数据集的多样性，增强了模型的鲁棒性。

(b) 特征提取网络：在主干网络中，YOLOv5 使用 Focus 模块进行分片处理，降低了计算复杂度并提升了特征表达能力。同时，引入 CSP(Cross Stage

Partial)和 SPP 模块，通过深度卷积操作捕获多尺度特征信息，增强了模型的感知能力，减少了信息丢失，并提高了对不同尺度目标的检测效果。

(c) 特征融合网络：在特征融合网络中，YOLOv5 引入了 PAN(Path Aggregation Network)结构，融合不同层级的特征图，使网络能够更好地理解目标的上下文信息和语义内容。此外，结合特征金字塔网络，在不同网络层次之间建立连接，实现了跨层级的特征融合。

(d) 检测端：YOLOv5 采用 GIoU^[44]作为损失函数，通过计算预测边界框与真实边界框之间的广义交并比，优化目标框的匹配精度，从而提高了模型的鲁棒性。

(2) YOLOv7

YOLOv7 模型型号主要包括 YOLOv7-Tiny、YOLOv7 基础版本、YOLOv7-W6、YOLOv7-X、YOLOv7-E6、YOLOv7-D6。其中，YOLOv7 基础版本，模型结构适中，在速度和精度之间取得了较好的平衡，更适用于航拍行人检测任务。YOLOv7 基础版本的模型结构如图 2-5 所示。

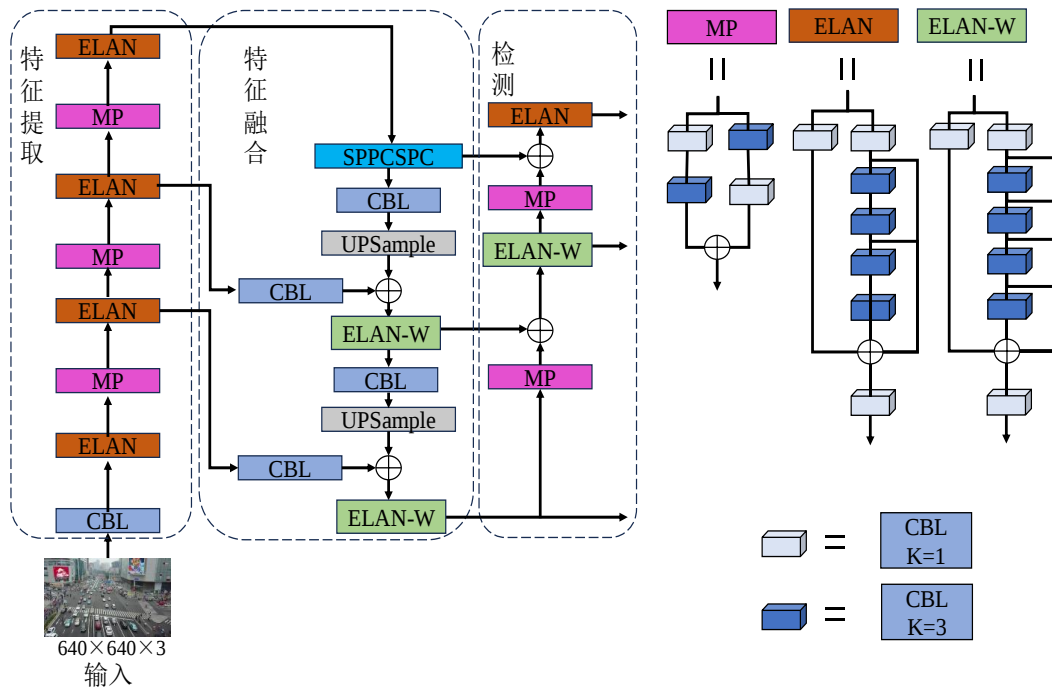


图 2-5 YOLOv7 网络模型结构图

YOLOv7 是基于 YOLOv5 的进一步优化版本。YOLOv7 在模型输入端、特征提取网络、特征融合网络、检测端做出了以下改进：

(a) 输入端: YOLOv7 在 YOLOv5 的基础上进一步优化了数据增强策略, 引入了更强的正则化方法, 提升了模型对小目标和复杂场景的适应能力。

(b) 特征提取网络: YOLOv7 引入了 ELAN 网络(Efficient Layer Aggregation Network), 通过更高效的特征聚合机制提升了特征提取能力。此外, YOLOv7 移除了 YOLOv5 中的 Focus 模块, 直接使用卷积层进行下采样, 减少了计算复杂度并提升了训练稳定性。YOLOv7 的主干网络更深, 能够提取更丰富的语义信息, 尤其是在处理复杂场景时表现更好。

(c) 特征融合网络: YOLOv7 在 YOLOv5 的基础上进一步优化了特征融合路径, 引入了更高效的特征聚合机制, 提升了多尺度特征的融合效果。

(d) 检测端: YOLOv7 引入了动态损失函数, 根据目标的尺度和难度动态调整分类和回归任务的权重, 进一步优化了模型的训练过程。

2.3 特征增强算法研究

2.3.1 多尺度特征融合

在目标检测领域, 多尺度检测始终是一个重要的研究方向, 尤其是在小目标检测任务中, 由于小目标在图像中的像素占比较低, 其在特征提取过程中容易丢失关键的细节信息, 从而导致检测性能下降。早期的目标检测方法主要可以分为以下三类:

(a) 单特征层。单特征层目标检测方法直接使用特征提取网络最后一层输出的特征图进行检测或识别。然而, 单特征层方法存在明显的局限性: 多次下采样操作导致最后一层特征图的空间分辨率大幅降低, 从而丢失目标的空间信息, 尤其对小目标或密集目标的定位精度较差。

(b) 多尺度输入。多尺度输入通过将图像调整为不同尺度并分别提取特征, 能够有效应对目标尺度变化问题, 尤其适用于小目标和大目标检测。然而, 该方法需要对每个尺度单独运行特征提取网络, 导致计算量和内存占用增加, 特别是在处理高分辨率图像或大规模数据集时, 计算成本急剧上升。

(c) 多尺度特征。多尺度特征方法通过保留并利用中间层的多尺度特征图, 能够有效应对目标尺度变化问题, 尤其适用于小目标和大目标检测。然

而，传统方法存在局限性：高层特征图语义丰富但分辨率低，定位精度不足；低层特征图分辨率高但语义信息弱，分类能力较差。

（1）FPN 网络

为了解决上述三类问题，LIN^[45]等人提出了一种用于多尺度特征融合的经典网络结构 FPN，旨在提升目标检测算法对多尺度目标的识别能力。如图 2-6 所示，FPN 的核心思想是通过构建特征金字塔结构，将浅层特征图的空间信息与深层特征图的语义信息进行有效融合，从而增强模型对不同尺度目标的检测性能。然而，FPN 仅通过单一路径将深层语义信息传递至浅层，可能导致浅层特征的空间信息未能充分反馈到深层，从而限制了特征融合的充分性。

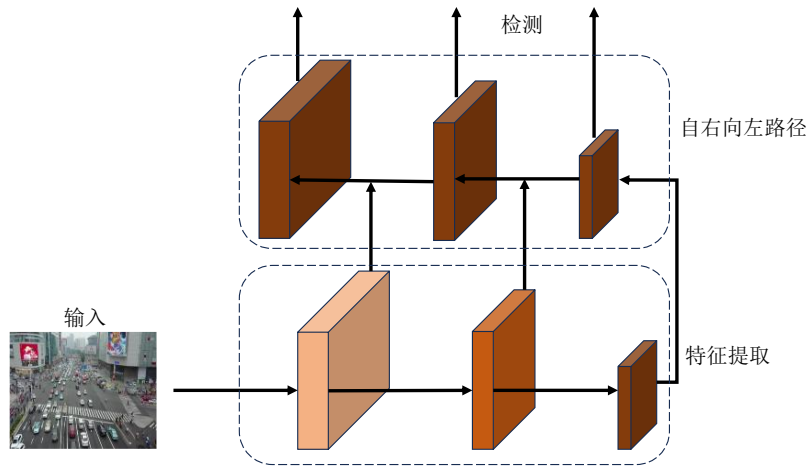


图 2-6 FPN 网络结构图

（2）PAN 网络

为了弥补 FPN 的不足，Liu 等人^[46]提出了一种新的特征融合网络 PAN，该网络结构如图 2-7 所示。

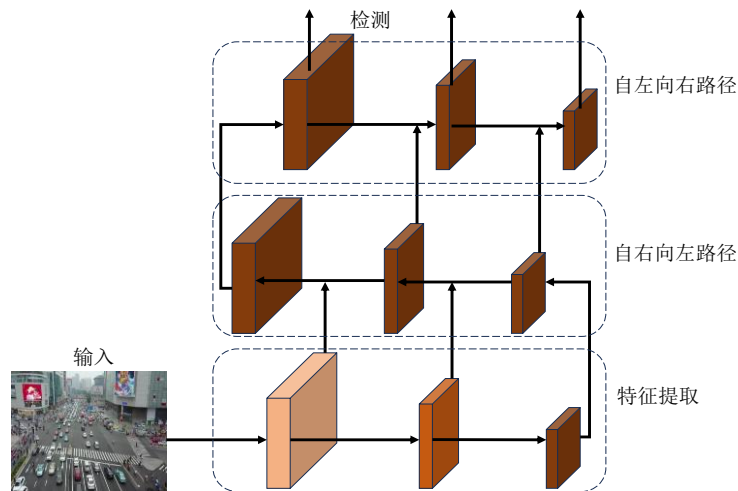


图 2-7 PAN 网络结构图

与 FPN 类似, PAN 同样采用金字塔式的特征提取结构。与 FPN 不同的是,在 FPN 的基础上增加了一条新的路径,它不仅将浅层特征向左传递,还将深层特征向右传递,使得特征在不同层次之间能够进行更充分的交互和融合,进一步增强了模型对不同尺度目标的适应性。这样可以减少在特征融合过程中信息的损失,同时还可以保留更多的细节信息,从而提高检测精度。

2.3.2 注意力机制

注意力机制最初在自然语言处理领域提出,其核心在于引导模型聚焦于关键特征。随着该机制向计算机视觉领域的迁移,其在目标检测、语义分割等视觉任务中展现出广泛的应用价值。例如, Vaswani 等人^[47]提出了 Transformer 模型,该模型通过自注意力机制实现了高效的序列建模,极大地推动了注意力机制的发展,并使其在各类任务中的应用更加便捷。随着研究的深入,许多相关的注意力机制的被提出,并成功应用于计算机视觉、自然语言处理等多个领域。本文将主要介绍以下三种注意力机制:

(1) 通道注意力机制。该机制基于特征图的通道维度特性,针对不同通道所承载的差异化语义信息,通过自适应权重分配来增强特征表达能力。这种通道维度的特征选择策略能够有效提升模型对多通道特征的利用效率,从而优化网络性能。该机制通过学习每个通道在特定任务中的重要性权重,使网络能够自适应地增强关键通道的特征表达,同时抑制不重要的通道,从而更好地捕捉数据中的区分性特征。典型的通道注意力有 CA 注意力(Coordinate Attention)^[48],其主要是在网络中插入两个并行的分支,一个分支执行全局池化,另一个分支通过卷积操作学习到的通道相关性对不同通道的特征图加权,最后将两者相乘得到最终的特征图。这种方法可以引入通道感知能力,使网络能够自适应地关注不同通道的重要性,从而具有更好的特征表达能力。

(2) 空间注意力机制。其是一种基于位置信息的特征优化方法。通过将注意力集中在感兴趣的区域上,空间注意力机制能够有效增强模型对目标特征的提取能力,尤其是在复杂场景和小目标检测任务中表现突出。例如 RetinaNet^[49]利用特征金字塔和聚类损失结合,提高小目标的注意力权重,从而提高检测性能。

(3) 通道空间注意力机制。通过融合通道维度和空间位置的双重信息，实现了对关键特征的自适应选择与增强，从而提高整个模型的效果和准确度。由残差注意力(Residual Attention)^[50]网络首先融合通道和空间注意力内容，在残差注意力之后，为了有效利用全局信息，后来的工作又相继引入全局平均池化(Global Average Pooling)等内容，充分利用全局上下文信息，增强模型对全局特征的感知能力。典型的通道空间注意力网络有 BAM^[51]，CBAM^[52]，DA-Net^[53]等。

2.4 轻量化算法研究

早期深度学习目标检测模型因结构复杂、参数量庞大及计算成本高，难以在资源受限的设备上高效运行。为此，模型轻量化成为重要研究方向。研究者通过优化网络架构、减少冗余参数及引入高效计算模块等方法，提升模型效率，使其更适用于嵌入式或移动终端设备的部署。例如，通过减少卷积层数量或缩小卷积核尺寸等方法降低模型计算复杂度^[54]，但此类简单的轻量化方法常导致模型精度下降。模型轻量化的核心目标是在保持性能稳定的前提下，通过降低计算复杂度与参数量，提升模型的运行效率与部署灵活性。当前，模型轻量化方法主要分为以下三类：

(1) 轻量化网络设计

轻量化网络设计在算法移植到设备上的过程中展现了非常重要的作用，例如 Chang 等人^[55]采用深度可分离卷积构建了多尺度特征轻量级网络结构，将网络结构中的普通卷积替换为深度可分离卷积显著降低参数量和计算量。普通卷积的过程如图 2-8 所示。深度可分离卷积的过程如图 2-9 所示。与普通卷积不同的是深度可分离卷积可分为逐通道卷积 (Depthwise Convolution, DWC) 和逐点卷积 (Pointwise Convolution, PWC) 两个过程。

普通卷积和深度可分离卷积的参数量分别如式(2-1)，(2-2)所示，式(2-3)表示的是它们的参数量比：

$$P_1 = H \times W \times K \times K \times C1 \times C2 \quad (2-1)$$

$$P_2 = (H \times W \times K \times K \times C1) + (H \times W \times 1 \times 1 \times C1 \times C2) \quad (2-2)$$

$$\frac{P_2}{P_1} = \frac{1}{C_2} + \frac{1}{K^2} \quad (2-3)$$

其中 H 、 W 、 $C1$ 为输入特征图的高、宽、通道数。 K 为卷积核的尺寸大小。 $C2$ 为输出特征图的通道数。通常情况下， $C2$ 、 K^2 都远大于 1，因此，深度可分离卷积的参数量远小于普通卷积。

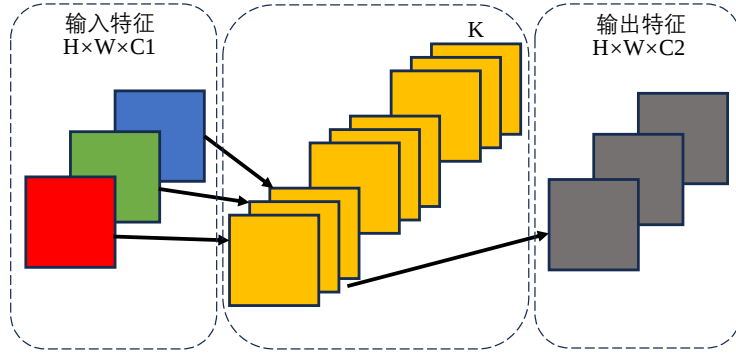


图 2-8 普通卷积

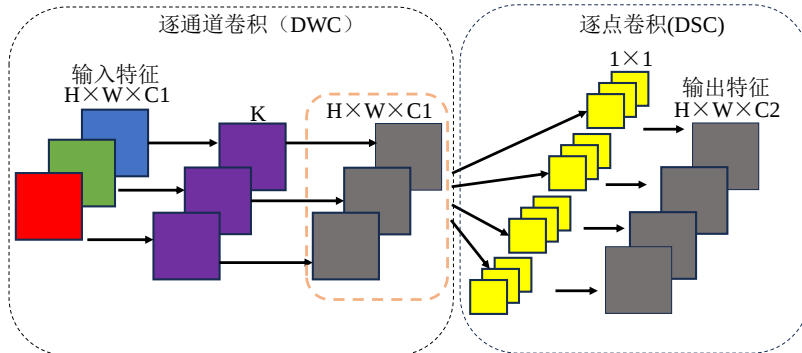


图 2-9 深度可分离卷积

(2) 模型剪枝

模型剪枝的理论基础在于神经网络训练过程中存在参数冗余现象，即部分参数对模型性能贡献较小，表现为权重值低或对输出影响微弱，甚至可忽略。通过识别并移除这些冗余参数，模型剪枝能够在保持性能基本稳定的前提下，简化模型结构，从而使其更加紧凑和高效。例如 Deng 等人^[56]通过对网络中不重要通道进行剪枝，极大地减少了模型的参数量。此外，Cui 等人^[57]提出了一种联合模型压缩方法，融合结构化剪枝与密集知识蒸馏技术，在保持模型精度的同时实现了明显的压缩效果。

(3) 知识蒸馏

知识蒸馏本质上是一种迁移学习，将复杂的高性能模型称为教师神经网络，将待压缩的简单模型称为学生神经网络。“知识”是从教师神经网络中抽取出的有用的目标特性信息，而“蒸馏”就是从教师神经网络中把有用的特性信息导入到学生神经网络。例如，Zhou 等人^[58]出了一种基于知识蒸馏的轻量化小目标检测算法，基于师生知识蒸馏的思想对 YOLOv5 算法进行知识蒸馏，该方法在提升小目标检测精度的同时，保持了模型的轻量化特性。

通过合理利用上述三种不同的模型轻量化方法，可以很好的平衡模型计算复杂度和检测精度之间的关系。表 2-1 对比了它们之间的优缺点。

表 2-1 VisDrone2019 数据集各类标签数量

方法	优点	缺点
模型优化	提高网络效率，减少计算量和网络参数	检测精度降低
模型剪枝	删除冗余参数，降低网络的计算量	剪枝策略的通用性差
知识蒸馏	模型尺寸不变，检测性能接近教师网络	轻量化结果依靠教师网络的性能

2.5 数据集与评价指标

2.5.1 数据集

本文所有实验采用的数据集均来源于 VisDrone2019^[59]。该数据集的制作是由 AISKYEYE 团队负责，基准数据集包含 288 个视频片段、261908 帧和 10209 幅静态图像。数据集中所有图片均在不同环境背景下采用无人机拍摄，共包括十种不同的标签类别，可以支持无人机航拍目标检测的研究，数据集的各类标签与数量如表 2-2 所示。

表 2-2 VisDrone2019 数据集各类标签数量

种类	数量
行人(pedestrian)	79055
人(People)	26962
自行车(bicycle)	10389
汽车(car)	144620
货车(van)	24899
卡车(truck)	12875
三轮车(tricycle)	4812
遮阳三轮车(awning-tricycle)	3245
公交车(bus)	5917
摩托车(motor)	29618

从表 2-2 中可以看出，该数据集包含 pedestrian、people 两个类别，且标签数可以满足无人机航拍行人目标检测所需，故在 VisDrone2019 数据集的基础上构建了无人机航拍行人检测数据集。构建流程如图 2-10 所示，总共分为两个步骤：首先，提取了 VisDrone2019 数据集中包含行人、人的图片与标签文件。随后，删除标签文件中另外八个类别，并将行人和人两类标签并为同一类别。

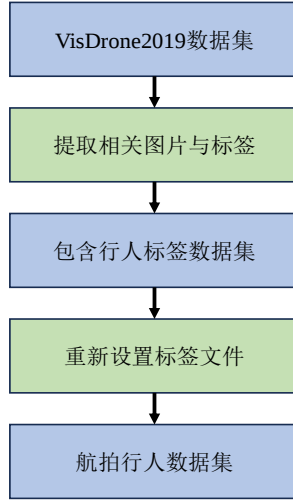


图 2-10 无人机航拍行人检测数据集构建流程

2.5.2 评价指标

为验证算法对目标检测任务的精度大小，采用的评价指标有 P （准确率）、 R （召回率）、 mAP （平均精度均值）、 FPS （每秒推理图片数）。其 P 、 R 、 mAP 计算公式分别如式(2-4)、(2-5)、(2-6)所示：

$$P = \frac{TP}{TP + FP} \quad (2-4)$$

$$R = \frac{TP}{TP + FN} \quad (2-5)$$

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i \quad (2-6)$$

其中 TP 为对正样本的正确预测， FP 为对负样本的错误预测， FN 为对正样本的错误预测， c 为检测的类别数， AP 为 P-R 曲线与坐标轴所夹的面积大小。 AP 的计算公式如式(2-7)所示：

$$AP = \int_0^1 P(R) dR \quad (2-7)$$

为验证本文算法对无人机航拍行人检测的可行性，除 P 、 R 、 mAP 、 FPS 外，还采用计算量（GFLOPs）、算法权重大小作为评价指标。通过对比实验、

消融实验等方法验证本文提出的方法在无人机航拍行人检测任务的优越性。

2.6 本章小结

本章首先详细阐述了无人机航拍行人检测任务所面临的问题。随后，通过对单阶段和双阶段目标检测算法的分析，选择 YOLOv7 作为本文基础模型进行研究，并分析了目前主要的模型特征增强、轻量化方法。最后，介绍了本文采用的数据集和评价指标。

第 3 章 基于特征增强的航拍行人检测方法

从第 2 章的分析可知，无人机航拍图像中的行人目标通常表现出多尺度变化、背景复杂、密集遮挡等特点，这些因素增加了多数通用目标检测算法对航拍图像中特征信息的提取难度。因此，本章针对航拍图像中行人的特点，重点研究改善航拍行人检测能力的方法，以提升当前通用算法在航拍场景中的检测能力。

3.1 改进 YOLOv7 模型结构

本文选取 YOLOv7 基础版本进行改进，尽管 YOLOv7 算法在目标检测任务中表现出色，但在无人机航拍行人检测场景中仍存在一些不足，主要体现在以下几个方面：

（1）小目标检测能力有限：无人机航拍图像中的行人目标通常占据较少的像素，属于小目标范畴。YOLOv7 虽然在多尺度检测上有所改进，但对于极端小目标的特征提取能力仍然不足，容易导致漏检或误检。

（2）多尺度适应性不足：无人机航拍图像中的行人目标尺度变化较大，尤其是远距离拍摄时目标尺寸较小，而近距离拍摄时目标尺寸较大。YOLOv7 虽然通过多尺度特征融合提升了检测能力，但在极端尺度变化下的表现仍有改进空间。

（3）密集遮挡问题：在无人机航拍图像中，行人目标之间或与周围环境之间可能发生密集遮挡，导致目标特征信息不完整。YOLOv7 在处理遮挡情况时，检测精度可能会明显下降。

（4）复杂背景干扰：无人机航拍图像背景复杂，可能存在与行人目标相似的纹理或颜色的背景信息，这会对检测算法造成干扰，增加漏检率。

针对 YOLOv7 算法对航拍行人检测任务的不足，本章在 YOLOv7 的基础上，主要通过特征增强的方法对算法进行改进。改进的 YOLOv7 算法网络结构如图 3-1 所示。

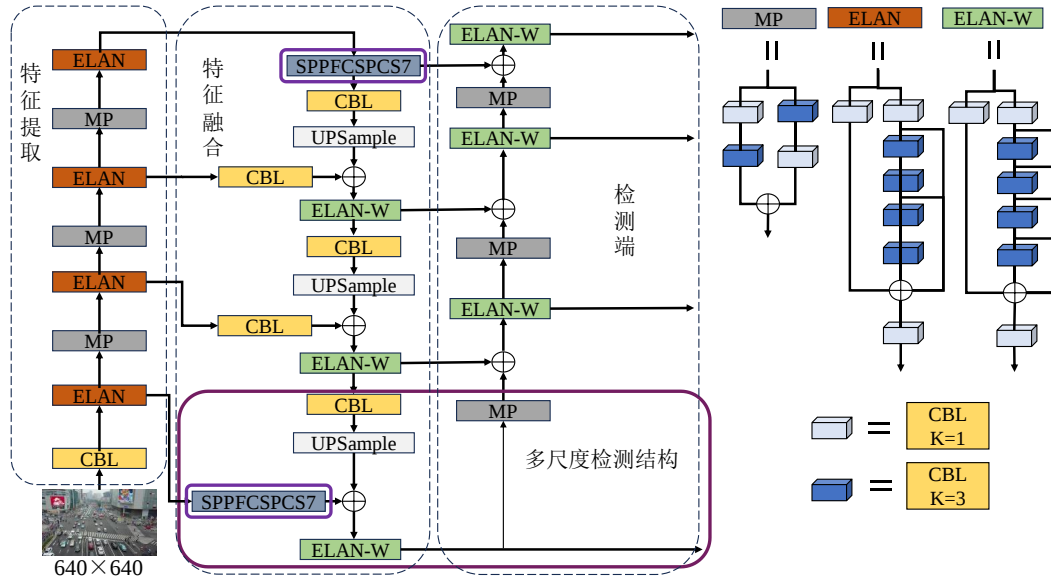


图 3-1 改进的 YOLOv7 网络模型结构图

本章主要从多尺度特征融合、注意力机制、后端处理的方向出发，增强 YOLOv7 算法对航拍图像中行人信息的特征表达能力，提高对航拍图像中行人的检测精度，主要做了以下改进：

(1) 从多尺度特征融合角度。本文对目标检测层进行了重新设计。具体而言，将原来的三个目标检测层扩展为四个，通过引入四倍下采样的特征图，并将其与原有的三个尺度特征图进行深度融合，从而新增了一个专门用于小目标检测的检测层。该小目标检测层对应的特征图分辨率更高，能够保留更多的细节信息，提升了算法对多尺度行人的检测能力。

(2) 从注意力机制角度。本文对 SPPCSPC 模块进行了重构，设计了一种基于注意力机制的特征增强模块 SPPFCSPCS7。该模块融合了无参注意力机制 SimAM，通过细化池化核的设计，进一步增强了特征提取能力。此外，基于小目标检测层和 SPPFCSPCS7 模块，提出了一种多尺度检测策略。这一设计不仅能够有效获取多尺度特征信息，还能通过 SimAM 注意力机制聚焦于小目标的关键信息，同时抑制无用背景信息的干扰。提升了算法在复杂背景、遮挡等场景下的检测精度，减少了漏检现象。

(3) 从后端处理角度。提出用 NWD 代替 IoU 作为度量标准，将相似度的计算换算成了距离的计算，NWD 能够有效减少由于边界框之间微小位置偏移，导致误差较大的问题。

3.2 多尺度检测策略

3.2.1 小目标检测层

在特征提取过程中，随着网络的不断加深，特征图尺度不断减小，会丢失小目标的位置信息。YOLOv7 在特征提取过程中总共进行了五次下采样，每次下采样尺寸都变为原来的 $1/2$ ，获得了五种不同尺寸的特征图(p1,p2,p3,p4,p5)。YOLOv7 中用于检测的特征图为 p3、p4、p5，其中深层特征图 p5 有着丰富的语义信息，但由于特征图尺寸不断减小，失去了很多小目标的位置信息，而浅层特征图 p2 正好有着丰富的小目标位置信息。为了能够更好的提取出小目标的位置信息，采用了 FPN+PAN 的结构，基于浅层特征图增加 p2 小目标检测层。如图 3-2 所示，将原来的三个目标检测层增加到四个，p2 检测层可以与更浅层的细节信息相结合，获取更多的关于小目标的位置信息，从而改善检测算法对航拍行人不同尺度的检测能力，尤其对于小目标。

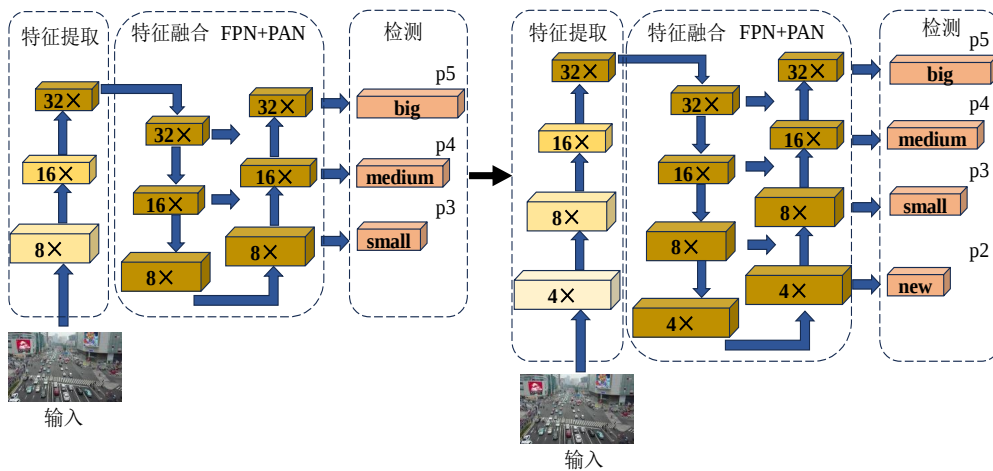


图 3-2 新增小目标检测层

为验证该方法的有效性，对比 YOLOv7 模型新增 p2 检测层前后的检测效果图。如图 3-3 所示，采用高空、低空两种不同情况下的图片进行实验。实验结果表明在这两种情况下，改进算法增强了多尺度检测能力，漏检率得到了明显改善。尤其在高空情况下，图中行人目标变得非常小，YOLOv7 原模型几乎无法检测到任何目标，而新增 p2 检测层后，几乎可以识别图中所有的目标，检测效果得到了明显的提升。

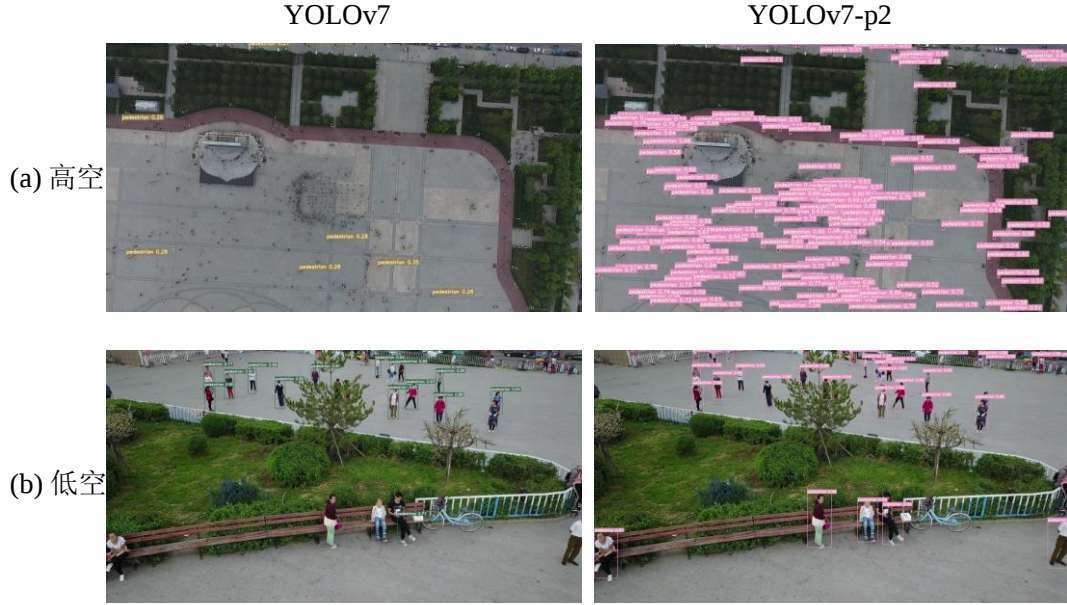


图 3-3 多尺度检测结果

3.2.2 特征增强模块

目标检测算法通过合理利用注意力机制，为感兴趣区域分配更多资源，并扩大感受野以增强多尺度特征信息，能够有效降低复杂背景和遮挡等因素的干扰，从而提升算法的检测性能。

SPP 模块的提出实现了特征图局部信息和全局信息的融合，丰富特征图的表达能力。之后又提出了 SPPF 结构，该结构采用串联池化层的方式加快了结构的推理速度。再到 YOLOv7 中引入了 SPPCSPC 结构，如图 3-4 所示，都达到了提高多尺度目标检测能力的目的。为进一步扩大模型的感受野范围，忽略背景信息的干扰，本文对 YOLOv7 模型中的 SPPCSPC 模块进行重构，提出了一种基于注意力机制的特征增强模块 SPPFCSPCS7，该模块进一步增强了航拍图像中目标特征信息的表达。该模块结构如图 3-5 所示。

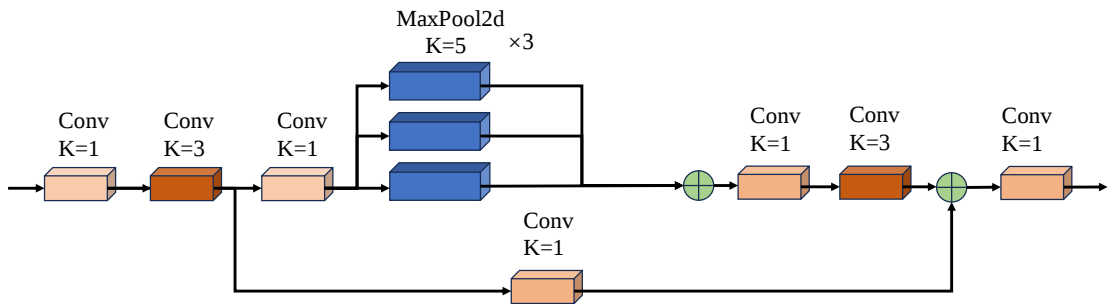


图 3-4 SPPCSPC 模块结构

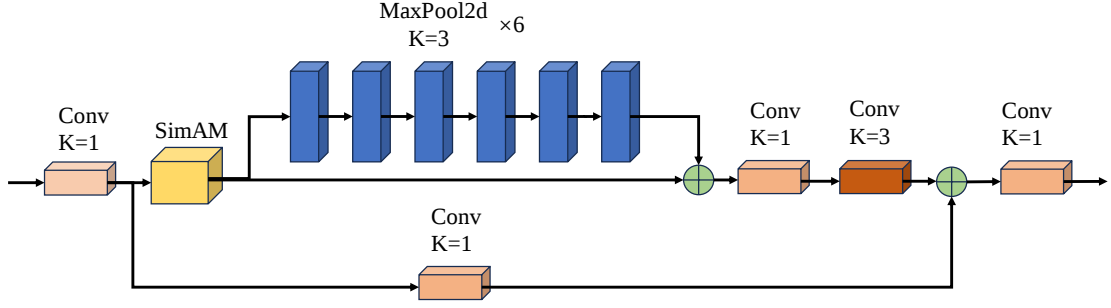


图 3-5 SPPFCSPCS7 模块结构

该结构主要改进如下：

(1) 首先，该模块引用了六个 3×3 的最大池化层，从而扩大对航拍行人目标的感受野范围。该方法可以更有效地捕捉目标的上下文信息，获取不同尺度的特征信息，从而提升网络的多尺度特征融合能力。

(2) 为减少前者带来的计算成本，将池化层之前两个卷积模块替换为无参数注意力机制 SimAM，在忽略无用背景信息的同时，增加对航拍图像行人特征信息表达，并且降低了该模块因多池化层而带来的计算成本。

SimAM 为无参数注意力机制，与现有的通道和空间注意力模块不同之处在于该模块无需向原始网络添加参数，而是在一层中推断特征图的三维关注权重。它通过能量函数衡量目标像素点 (t) 和其它像素点 (x_i) 的线性可分性，当输入特征图为 $X \in R^{C \times H \times W}$ 时，像素点线性变换如式(3-1)、(3-2)所示：

$$\hat{t} = w_t t + b \quad (3-1)$$

$$\hat{x}_i = w_t x_i + b_t \quad (3-2)$$

其中 w_t 为权重， b_t 为偏置值，带入能量函数如式(3-3)所示：

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (3-3)$$

其中 e_t 为能量函数值 y 为某一像素点的标签值， y_t 和 y_0 为目标像素点 (t) 和其它像素点 (x_i) 的标值， i 为空间维度上的索引， M 为某一通道上像素点个数，其计算公式如式(3-4)所示：

$$M = H \times W \quad (3-4)$$

为充分利用特征金字塔结构，在 p2 检测层处增加第二个 SPPFCSPCS7 特征

增强模块，提出了一种多尺度检测策略(YOLOv7-SPPFCSPCS7-p2)，进一步加强了网络对航拍图像的特征表达能力

YOLOv7 采用多尺度检测策略后，对于航拍图像中像素占比小的行人具有更好的关注度。为验证该方法对于航拍行人检测的优越性，采用可视化热力图的方法观察 YOLOv7 应用 p2 检测层、SPPFCSPCS7 模块前后对航拍图像中行人的敏感度情况。如图 3-6 所示，采用了明亮和昏暗两种情况下的图片进行实验，结果表明采用多尺度检测策略后，模型明显对于行人的敏感度更强，关注度更高。验证了本章所提方法的可靠性。

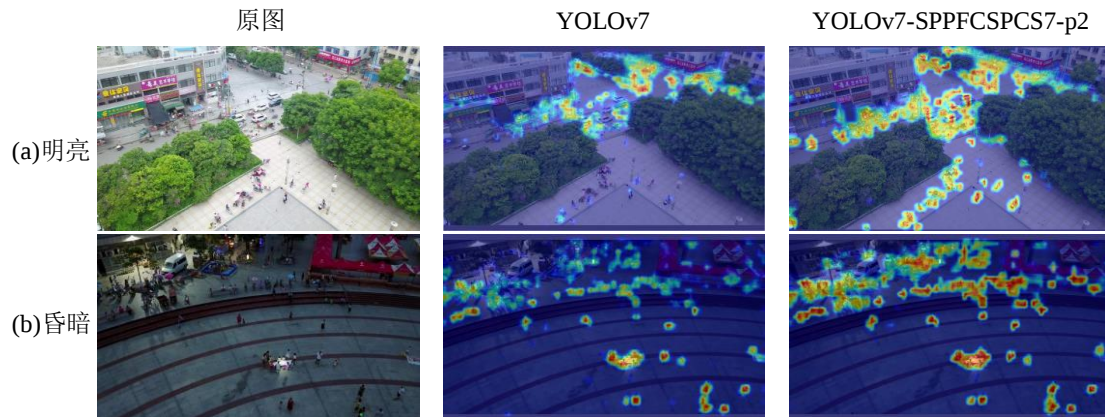


图 3-6 热力图可视化

为直观验证该方法对复杂背景场景、遮挡问题的改善效果，将原算法和本文算法进行检测对比。结果如图 3-7 所示。



图 3-7 检测效果对比

a、b 分别为复杂背景场景图和遮挡图的对比结果。其中 a 受光照的影响，图中间目标所处场景光线较亮，原算法和改进算法都可以准确识别目标，然后，对于图右上角所处黑暗场景，改进后算法具有更强的特征表达能力，明显改善了原算法的漏检问题。b 中目标人群受路边树木的影响，造成了严重遮挡，从图中可看出改进后算法对于遮挡目标的检测效果也明显优于原算法。由分析可知，本章提出的多尺度检测策略可以帮助 YOLOv7 模型更好的应对复杂背景、遮挡等情况的航拍行人检测任务。

3.3 NWD 损失函数

航拍图像中行人目标多数为小目标，然而对于小目标检测，除了加强网络的特征表达能力外，网络模型的后端处理也极为重要。如图 3-8 所示，A 为行人的真实边界框，B、C 分别为偏移 1 个像素和 4 个像素的预测框，仅 3 个像素的位置偏移导致 IoU 下降 0.45。这表明，对于小目标而言，较小的位置偏移会显著影响 IoU 值。为改善这一问题，本文采用 NWD 替代 IoU 作为度量标准。

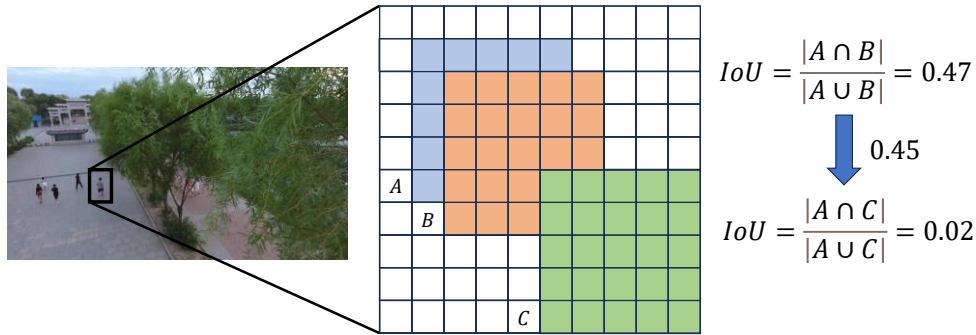


图 3-8 像素偏移导致 IoU 变化示意图

在真实场景中，物体通常并非严格的矩形形状，边界框中往往带有背景像素，目标物体像素和背景像素分别集中在边界框的中心和边界上，为了更准确地描述不同像素点的重要性，对边界框 $R=(cx,cy,w,h)$ 拟合成二维高斯分布 $N(\mu,\Sigma)$ ，边界框的中心具有最高权重，重要性从中心到边界逐步减小，用这种方式将边界框的相似性转换为两个高斯分布之间的分布距离。其中， (cx,cy) 为边界框的中心坐标， w 、 h 为边界框的宽和高， μ 、 Σ 的表达式如式(3-5)所

示:

$$\boldsymbol{\mu} = \begin{bmatrix} cx \\ cy \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (3-5)$$

用 Wasserstein 距离来计算分布距离，两个不同的二维高斯分布 $\mu_1 \sim N(\mathbf{m}_1, \boldsymbol{\Sigma}_1)$ 、 $\mu_2 \sim N(\mathbf{m}_2, \boldsymbol{\Sigma}_2)$ 之间的距离如式(3-6)为:

$$W_2^2(\mu_1, \mu_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_2^2 + \text{Tr}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{1/2})^{1/2}\right) \quad (3-6)$$

对于从边界框 $A = (cx_a, cy_a, w_a, h_a)$ 和 $B = (cx_b, cy_b, w_b, h_b)$ 建模的二维高斯分布 N_a 和 N_b ，等式可进一步化简，如式(3-7)所示:

$$W_2^2(\mu_1, \mu_2) = \left\| \begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right\|_2^2 \quad (3-7)$$

因 $W_2^2(N_a, N_b)$ 是距离度量，不能直接用作相似性度量，因此对其进行归一化，其中 C 为常数。最终得到 NWD，如式(3-8)所示:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (3-8)$$

NWD 在检测小目标物体时，具有对位置偏差的平滑性，可以有效改善网络对小目标位置偏移敏感而造成检测精度较低的问题。

3.4 实验结果与分析

3.4.1 实验环境及参数配置

本文实验使用 Python 语言编写的基于 Pytorch 深度学习框架进行模型搭建与训练，硬件环境的 GPU 型号为 NVIDIA GeForce RTX3060，CPU 为 Intel i7-10700K。软件环境操作系统 Windows10，Cuda11.2，深度学习框架采用 Pytorch1.10.2。其次，由于目前面向航拍行人检测任务的公开数据集较少，本文基于 VisDrone2019 数据集，通过数据筛选构建了航拍行人检测数据集。将数

数据集按 6:2:2 的比例划分为训练集、验证集和测试集，模型训练的 batch-size 设置为 4，训练持续至 mAP 值基本稳定为止，确保模型达到最优性能。

3.4.2 消融实验

利用消融实验来验证不同改进方法对模型性能的影响。实验结果如表 3-1 所示，各个模块对 YOLOv7 模型的检测精度、召回率都起促进作用。其中增加 p2 小目标检测层的提升最大，其 mAP0.5 增加了 4.1%。分析问题可知，航拍图像中行人目标都很小，然而，由于卷积模块的局限性，随着网络的加深，特征图的尺寸不断减小，容易丢失小目标的信息，而增加 p2 检测层可以获得更多的关于小目标的细节信息。其次在增加 p2 层的基础上引入 SPPFCSPCS7 特征增强模块，构建了多尺度网络结构(YOLOv7-p2-SPPFCSPCS7)，其 mAP0.5 较原网络提升了 7.2%，FPS 为 42，说明该方法提高了网络的检测精度并保证了检测速度。最后为改善传统的 IoU 度量对小目标位置偏移敏感的问题，采用 NWD 度量边界框损失，mAP0.5 提高了 1.4%。其最终网络模型的 mAP0.5 达到了 56.4%，较原网络提高了 9.1%。然而，模型的权重大小和计算量得到一定的上升，增加了部署嵌入式设备时的资源利用率。

表 3-1 消融实验

model	GFLOPs	模型权重	mAP0.5/%	R/%	FPS
YOLOv7	105.4	72.7	47.3	47.5	40
YOLOv7-p2	120.8	77.6	51.4	49.6	40
YOLOv7-SPPFCSPCS7	109.3	74.3	50.1	48.9	44
YOLOv7-p2-SPPFCSPCS7	127.2	80.5	54.5	53.2	42
YOLOv7-NWD	106.9	73.6	48.7	48.2	44
Ours	129.2	82.1	56.4	54.9	41

3.4.3 YOLO 系列算法对比实验

(1) 检测精度对比

为验证本章所提算法相较于其它算法的优越性，采用目前具有代表性的 YOLO 系列算法在相同的实验环境下进行检测对比。实验结果如表 3-2 所示，改进算法和基准模型 YOLOv7 的检测精度比其它 YOLO 系列算法都要高。说明 YOLOv7 在航拍行人检测中较其它对比算法有着更好的检测性能，且基于特征

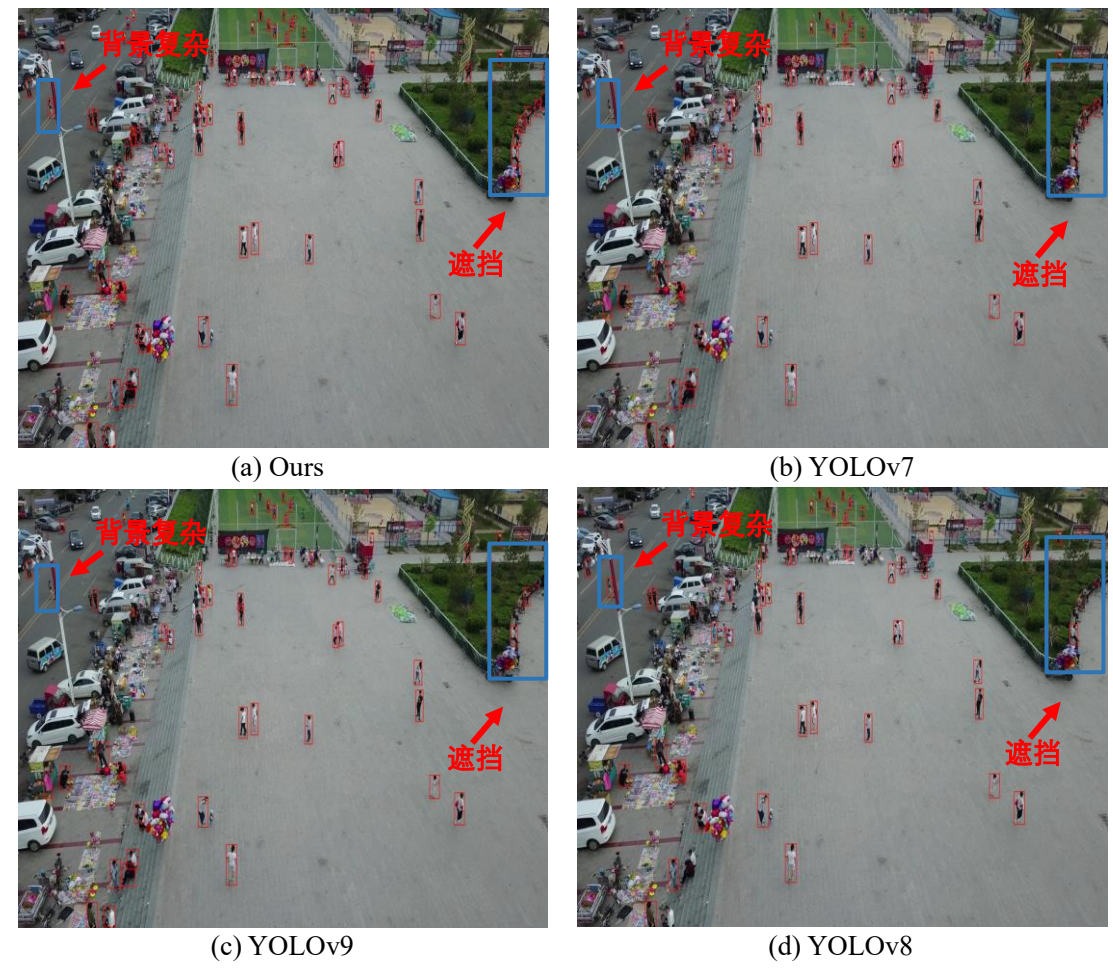
增强改进算法相较于对比实验中其它算法，在检测精度方面具有明显的优势。

表 3-2 YOLO 系列目标检测算法的对比试验

model	mAP0.5/%
YOLOv5	36.2
YOLOv6	33.5
YOLOv8	39.7
YOLOv9	45.5
YOLOv7	47.3
Ours	56.4

(2) 检测效果图对比

为直观的验证本章所提算法的检测能力，对不同算法进行了检测效果图对比。如图 3-9 所示，对于尺度较小且被严重遮挡的行人目标，本章算法相较于其它 YOLO 系列算法，可以有效地捕捉被遮挡目标的局部特征信息，准确检测出被遮挡目标。针对尺寸小且与背景颜色相近的行人，本章算法在复杂背景干扰下仍可以检测出目标。实验结果表明，本章算法在应对目标遮挡和复杂背景等挑战性场景时具有明显优势。



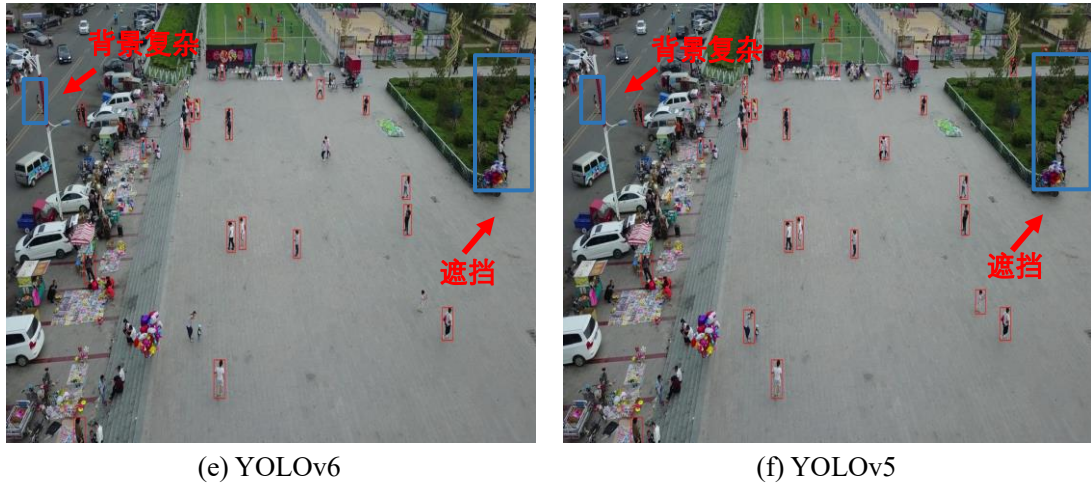


图 3-9 YOLO 系列算法检测效果对比图

3.4.4 双阶段算法对比实验

(1) 检测精度对比

为充分验证本章所提算法相较于其它算法的优越性，利用航拍行人检测数据集，将所提算法与多种经典双阶段目标检测算法进行了对比实验。实验结果如表 3-3 所示，实验结果表明，在检测精度方面，本章所提算法的 mAP0.5 分别比 Grid R-CNN^[60]、Mask R-CNN、DH R-CNN^[61]高 8.9%、12.2%、9.5%。在检测速率方面，本章所提算法同样表现出明显优势。

表 3-3 其它类型目标检测算法的对比试验

model	mAP0.5/%	FPS
Grid R-CNN	47.5	10
Mask R-CNN	44.2	12
DH R-CNN	46.9	20
Ours	56.4	41

(2) 检测效果图对比

如图 3-10 所示，为进一步验证本章算法在复杂场景下的优越性，采用与 YOLO 系列对比实验相同的航拍行人检测数据集图片，与 Grid R-CNN、Mask R-CNN、DH R-CNN 等经典双阶段目标检测算法进行了可视化对比分析。实验结果表明，在目标遮挡场景中，本章算法通过引入基于注意力机制的特征增强模块，能够有效区分被遮挡目标的部分可见区域，从而降低漏检率，而本文对比的双阶段算法在目标重叠严重时检测精度明显下降。此外，在复杂背景场景

中，本章算法增强了对关键特征的聚焦能力，能够有效抑制背景干扰，而双阶段算法受复杂背景影响，导致漏检率升高。

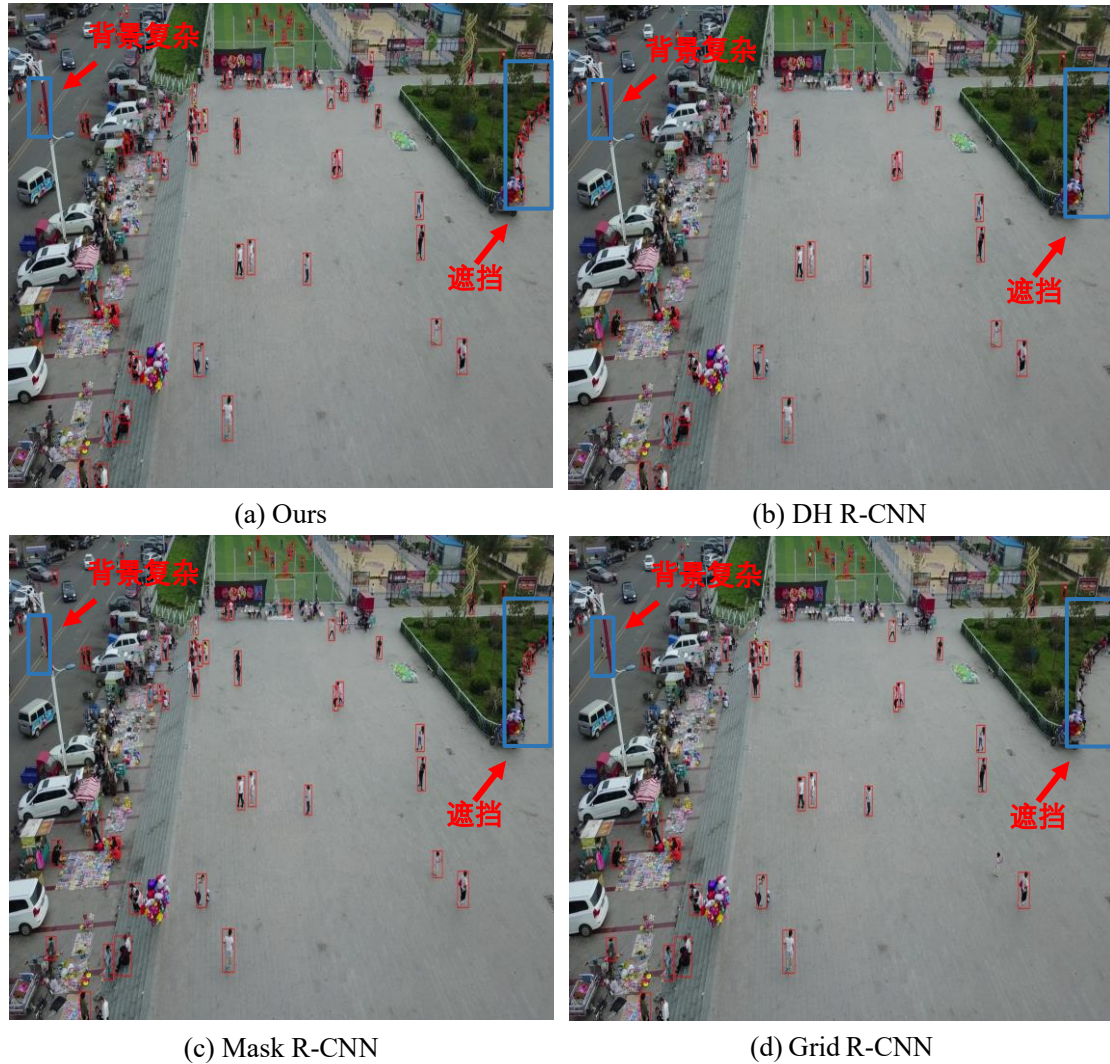


图 3-10 双阶段算法检测效果对比图

3.4.5 VisDrone2019 数据集通用性实验

为了验证本章所提方法在其它航拍数据集上的通用性，利用改进算法与 YOLOv7 基准算法在 VisDrone2019 数据集中进行对比实验。该数据集中共标注了十种不同类别的目标，可以满足各种无人机航拍检测任务。从中选取了 7019 张航拍图像，同样将数据集按照 6:2:2 的比例划分为训练集、验证集和测试集，对模型进行训练。实验结果如表 3-4 所示，从中可以看出数据集中各种检测目标在本文模型上都得到了明显的提升，其中 pedestrian 的提升最大，其 mAP0.5 较 YOLOv7 提升了 7.8%。这说明本章所提方法可有效增强对航拍图像中行人目标

的特征表达，且对其它不同类型目标的信息提取能力也起促进作用，验证了本章所提方法在无人机航拍检测任务中的广泛适用性。

表 3-4 VisDrone2019 数据集上的通用性效果

种类	YOLOv7(mAP0.5/%)	Ours(mAP0.5/%)
all	47.1	51.1
人(People)	48.3	56.1
自行车(bicycle)	33.1	38.6
汽车(car)	19.6	23.5
货车(van)	76.9	82.3
卡车(truck)	54.8	58
三轮车(tricycle)	59.2	59.5
遮阳三轮车(awning-tricycle)	34.8	37
公交车(bus)	25.2	27
摩托车(motor)	69.7	73
行人(pedestrian)	48.2	54.7

3.5 航拍行人检测系统的可视化分析

设计并实现了一套基于 PyQt5 框架的航拍行人检测可视化系统。该系统在硬件配置为 i5-8300H CPU、NVIDIA GeForce GTX 1060 的笔记本电脑平台上开发完成,能够有效支持用户进行直观的航拍行人检测分析。在系统实现过程中,通过 PyQt5 的图形界面组件实现了友好的用户交互界面,用户可通过简单的操作完成模型加载、检测执行、数据保存等任务。系统还集成了检测结果的可视化分析功能,可直接计算出模型检测到的目标个数以及推理速度。

3.5.1 PyQt5 简介

PyQt5 是 Qt5 框架与 Python 的集成,为开发者提供了丰富的图形用户界面(Graphical User Interface, GUI)应用程序框架,可以在 Windows 和 Linux 等系统上进行开发使用。PyQt5 通过将 Qt 的 C++类转换为 Python 类,并充分利用 Python 的语法特性,简化了 GUI 应用程序的开发流程。其与 Qt Designer 等图形用户界面设计工具的兼容性,使开发者能够通过直观的可视化界面高效设计应用程序界面。此外,PyQt5 提供了丰富的功能模块如 QtXml、QtCore 等,为构建高性能、高质量的 GUI 应用提供了坚实支持。

综上所述，PyQt5 是一个功能强大且灵活的 GUI 开发工具，适用于多种应用场景。其跨平台特性和丰富的组件库使得开发者能够高效地创建高质量的桌面应用程序。

3.5.2 系统功能与检测流程

本文设计的系统主要包含三个功能，对无人机航拍本地图片、视频进行检测，以及调用本地摄像头模拟无人机视角进行实时检测。检测流程如图 3-11 所示，进入检测系统用户需要选择已经训练好的模型权重，权重加载完成后，系统会调用网络参数对后续输入的内容进行检测，此时可以根据自己的需求选择合适的功能完成检测。

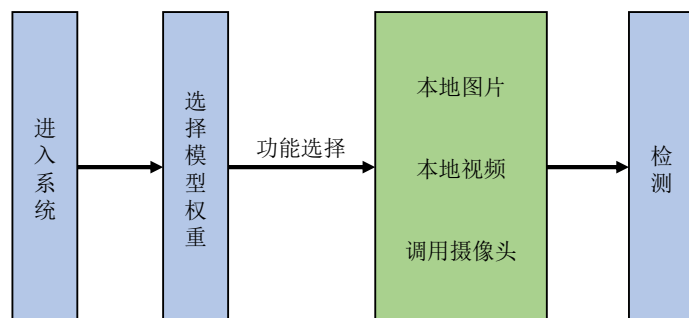


图 3-11 航拍行人检测系统检测流程

3.5.3 可视化分析

(1) 检测图片

利用该系统加载训练好的算法权重文件对本地图片进行检测。选择的本地图片具有以下特点：行人尺寸较小，在行人密集区域存在严重的遮挡，边缘区域行人与背景信息混淆。载入 YOLOv7 算法权重文件的检测结果如图 3-12 所示，从右侧的统计结果可以看出，YOLOv7 算法仅检测到 48 个行人目标，对比原图可以发现存在明显的漏检现象。



图 3-12 YOLOv7 算法检测结果

为了验证本章所提算法的优越性，重新载入训练好的改进 YOLOv7 算法的权重文件，随后对相同的航拍图片进行检测，检测结果如图 3-13 所示，根据右侧统计结果可知，算法检测出行人 87 个。相较于 YOLOv7，因行人目标小、遮挡、背景复杂带来的漏检问题得到了明显改善。实验结果表明，本章提出的基于特征增强改进的算法增强对复杂背景中密集遮挡小目标的特征表达能力。

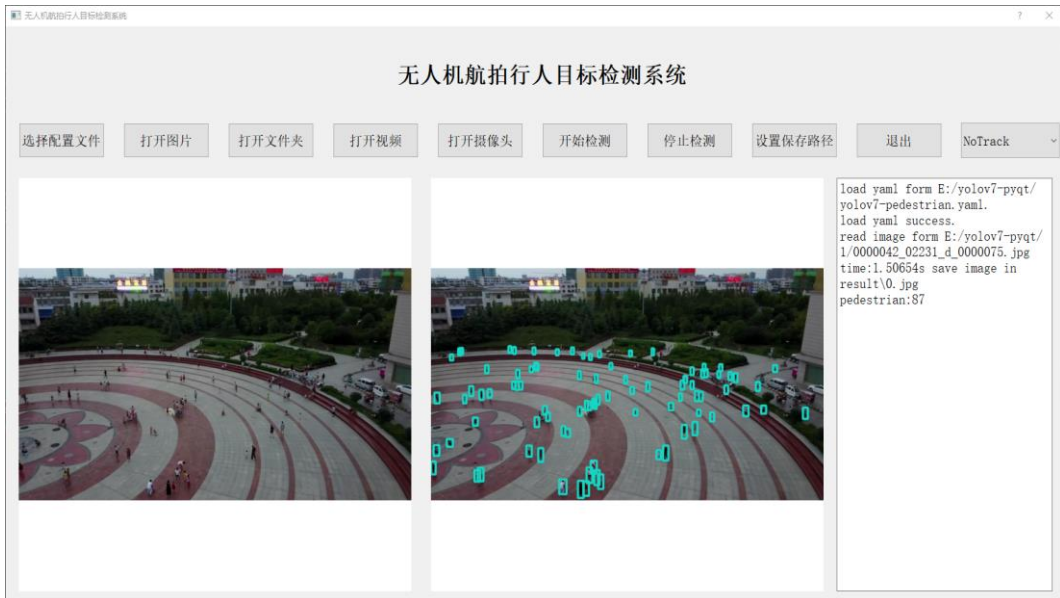


图 3-13 基于特征增强改进的算法检测结果

(2) 检测视频

为验证本章所提算法对处于移动状态下行人目标的检测能力，利用航拍行

人检测系统对本地视频进行检测。载入本章改进算法的权重文件后的检测结果如图 3-14 所示。



图 3-14 视频检测效果

本地视频中足球场上的行人由于处于移动状态，很容易出现遮挡的情况，且远处的行人容易与背景信息混淆。改进后算法几乎可以完整的检测出视频中所有行人目标，该结果表明本章所提方法可以有效聚焦移动状态下行人位置信息，在复杂场景下具有不错的检测性能。然而，由于无人机航拍检测系统所使用的笔记本硬件原因，每秒只能推理 13 张图片(FPS=13)，检测的实时性较差。不过，该系统已经可以基本满足航拍行人目标检测的要求，后续为了提高检测速度，可以考虑更换高性能的硬件设备。

3.6 本章小结

本章针对通用目标检测算法在行人尺度变化大、遮挡及复杂背景等场景下的不足，在 YOLOv7 的基础上，提出了一种基于特征增强的航拍行人改进算法。通过在无人机航拍数据集上进行多组实验验证，并结合设计的航拍行人检测系统进行可视化分析。实验结果表明，改进后的 YOLOv7 在航拍行人检测任务中的精度达到 56.4%，优于多数经典的单阶段和双阶段算法。此外，本章还对比了该算法与其他算法在极端环境下的检测效果，进一步验证了本文所提方法的有效性。

第 4 章 航拍行人轻量化检测方法

在第 3 章中以 YOLOv7 为基准模型，利用多尺度检测策略等特征增强方法，有效提高了对航拍图像行人信息的特征表达能力，但也增加了模型复杂度。为了使检测模型更易进行嵌入式设备的部署，缓解模型精度与权重大小、计算量之间的矛盾，本章提出一种网络模型轻量化策略：设计了基于注意力机制的轻量化模块，并结合网络结构剪枝方法，对基于特征增强改进的 YOLOv7 目标检测算法进行轻量化改造。

4.1 轻量化模块

本节提出了一种高效的轻量化卷积模块，该模块通过将深度可分离卷积与注意力机制相结合，实现了模型性能与计算效率的平衡。具体而言，首先设计了 FCBAM 模块，该模块通过融合 CBAM 与 FReLU^[62]激活函数，进一步增强了特征提取能力。其中，CBAM 模块通过通道注意力和空间注意力的双重机制，使网络能够自适应地聚焦于关键特征区域。而 FReLU 则通过引入空间上下文信息，增强了模型对复杂特征的表达能力。在此基础上，将 FCBAM 与深度可分离卷积有机结合，构建了轻量化卷积模块。深度可分离卷积通过将普通卷积分解为深度卷积和逐点卷积两个步骤，有效减少了参数量和计算量。后续实验结果表明，所提出的轻量化卷积模块在保持检测精度的同时，降低了模型的权重大小和计算量。

4.1.1 FCBAM 模块

注意力机制的引入为增强网络表征能力提供了新的思路。使网络能够自主聚焦于输入数据中的关键特征区域，同时抑制冗余或无关信息，从而提升模型的特征学习能力。在众多注意力机制中，卷积注意力模块 CBAM 是一种高效且轻量级的模块，满足本文的研究要求。其结构如图 4-1 所示，该模块通过集成两个并行的注意力子模块：通道注意力模块 CAM（Channel Attention Module）和空间注意力模块 SAM（Spatial Attention Module），分别从通道维度和空间维度

对特征进行自适应加权，从而实现了更精确的特征选择与增强。这种双路注意力机制能够有效提升模型对重要特征的捕获能力，同时抑制无关特征的干扰。

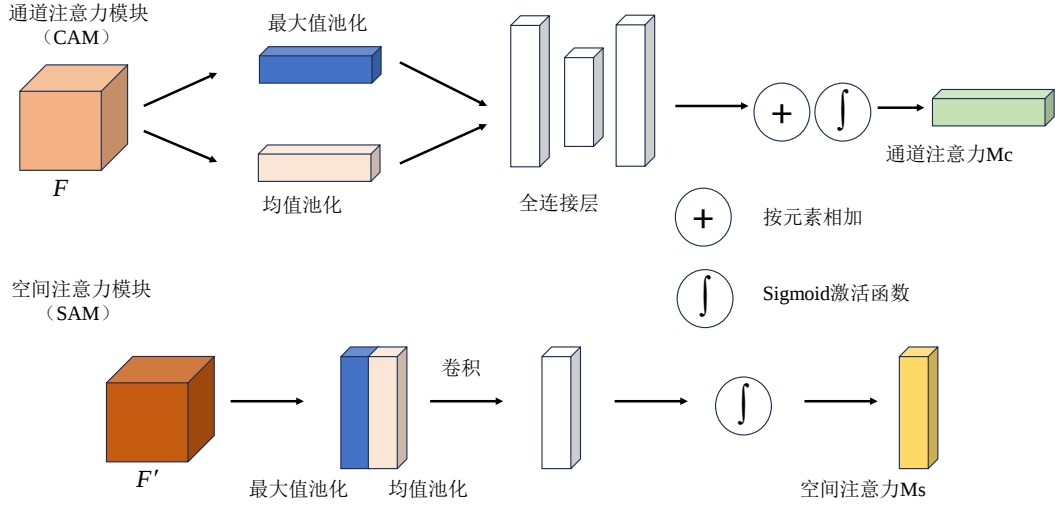


图 4-1 通道和空间的注意力机制模块

给定输入特征 $F \in R^{C \times R \times W}$ ，其通道注意力模块的操作为：对输入特征按照通道进行最大值池化和均值池化，然后池化后的两个一维向量通过全连接层生成通道注意力 $M_c \in R^{C \times 1 \times 1}$ ，最后将通道注意力与输入元素相乘得到通道特征图 F' 。空间注意力模块的操作为：将 CAM 输出结果作为 SAM 的输入，对输入特征图 F' 根据空间进行最大值池化和均值池化，对生成了两个二维向量进行拼接和卷积运算，最后将空间注意力与 F' 相乘。过程公式如式(4-1)、(4-2)所示：

$$F' = M_c(F) \otimes F \quad (4-1)$$

$$F'' = M_s(F') \otimes F' \quad (4-2)$$

由上述可知，CBAM 是一种混合注意力机制模块，通过 CAM、SAM 两个模块，能够根据输入特征自适应地学习通道和空间的重要性，增强模型的特征表达能力。为进一步增强 CBAM 对网络的空间敏感度，提高网络鲁棒性，提出了一个新的模块 FCBAM。如图 4-2 所示，在 CBAM 模块中引入卷积层，并采用了 FReLU 激活函数。FReLU 是对 ReLU 和 PReLU^[63]的扩展，相较于 Leaky ReLU，FReLU 可以将像素化的建模能力进行参数化，在卷积的帮助下捕获更为复杂的视觉布局，以此在复杂背景条件下获得更好的性能表现，增强网络鲁棒性。

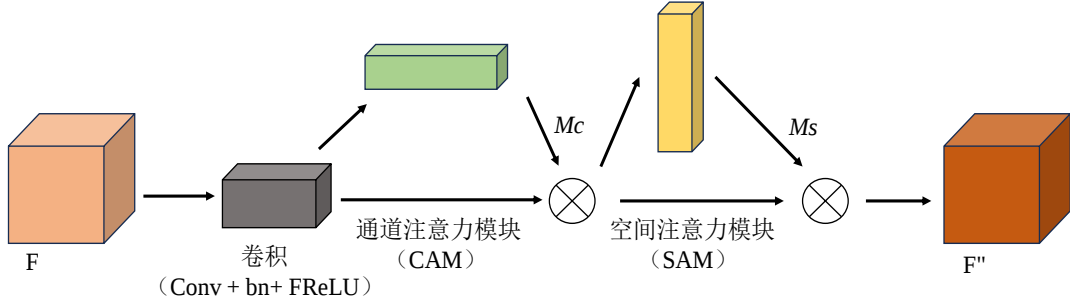


图 4-2 FCBAM 模块

与 ReLU 相同, FReLU 也使用 $\max(x, 0)$ 来获得 x 和条件之间的最大值, 为了实现空间条件, 采用一个参数池窗口来捕捉不同像素之间的空间关系, 根据不同位置的像素进行适应性调整。ReLU 和 PReLU 的形式分别是 $f(x) = \max(x, 0)$ 和 $f(x) = \max(x, P_x)$, FReLU 可以用式(4-3)、(4-4)所示:

$$f(x_{c,i,j}) = \max(x_{c,i,j}, T(x_{c,i,j})) \quad (4-3)$$

$$T(x_{c,i,j}) = x_{c,i,j}^\omega \cdot P_c^\omega \quad (4-4)$$

其中, $x_{c,i,j}$ 表示通道 c 上非线性激活函数 $f(\cdot)$ 在二维空间位置 (i, j) 的输入像素; 漏斗条件用函数 $T(\cdot)$ 来表示; $x_{c,i,j}^\omega$ 表示以 $x_{c,i,j}$ 为中心的参数池化窗口; 此窗口在同一通道中共享的系数用 P_c^ω 来表示。

4.1.2 轻量化卷积模块

为了在降低模型复杂度的同时确保模型的特征表达能力不受影响, 设计了一种轻量化卷积模块, 该模块的设计充分融合了注意力机制和深度可分离卷积的核心思想, 实现了计算效率与特征提取能力的平衡。在减少计算成本的同时, 避免了精度的显著下降。如图4-3所示, 该模块的设计流程可分为以下四个步骤:

首先利用 FCBAM 模块将输入特征通道数($c1$)压缩为最终输出特征通道数($c2$)的一半, 这一操作不仅降低了后续计算的开销, 还通过注意力机制增强了关键特征的提取能力, 抑制了冗余信息。接着, 对压缩后的特征图应用逐通道卷积, 进一步减少参数计算量。逐通道卷积通过对每个输入通道独立进行卷积操作, 降低了计算复杂度, 同时保留了空间特征信息。然后, 将得到的两个通

道数为 $c2/2$ 的特征图进行拼接，恢复为 $c2$ 通道数的特征图。最后，为了促进不同通道之间的信息交互，避免特征图在通道维度上的信息孤立，本文在模块的最后引入了通道混洗操作（Channel Shuffle）。该操作通过打乱通道顺序，增强了不同通道间的特征融合能力，从而提升了特征的多样性和表达能力。最终，输出通道数为 $c2$ 的特征图。

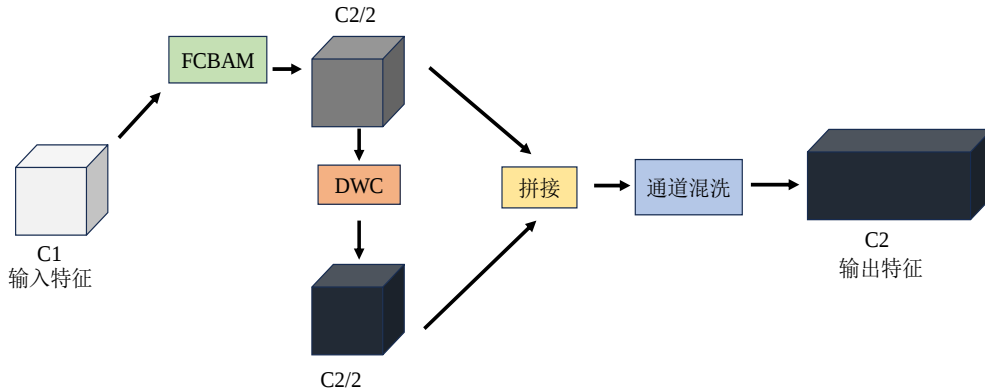


图 4-3 轻量化卷积模块

为了验证 CBAM、FCBAM、轻量化卷积模块等方法对检测算法的性能影响。如表 4-1 所示，基于航拍行人数据集，将多种方法替换 YOLOv7 特征融合网络中普通卷积模块，进行对比试验。

表 4-1 实验对比

	CBAM	FCBAM	轻量化卷积模块	GFLOPs/G	Weight/M	mAP0.5/%
YOLOv7				105.4	72.7	47.3
a	√			101.5	70.3	47.5
b		√		102.3	71.5	47.9
c			√	94.2	65.5	47.6

从表 4-1 中可以看出，CBAM、FCBAM、轻量化卷积模块，都提升了 YOLOv7 算法的检测精度，然而，CBAM 和 FCBAM 在模型轻量化方面的效果并不显著，相比之下，本文提出的轻量化卷积模块在提升检测精度的同时，优化了模型的计算复杂度。

4.2 冗余网络结构剪枝

第 3 章中，为了 YOLOv7 检测算法能够更好的提取出小目标的位置信息，增强多尺度检测能力，重新设置了检测层。如图 4-4 所示，基于浅层特征图增加了 p2 检测层，将原来的三个目标检测层增加到四个 p2 检测层可以与更浅层的

细节信息相结合，获取更多的关于小目标的位置信息，从而改善检测过程中的漏检问题，提升对航拍小目标的检测能力。

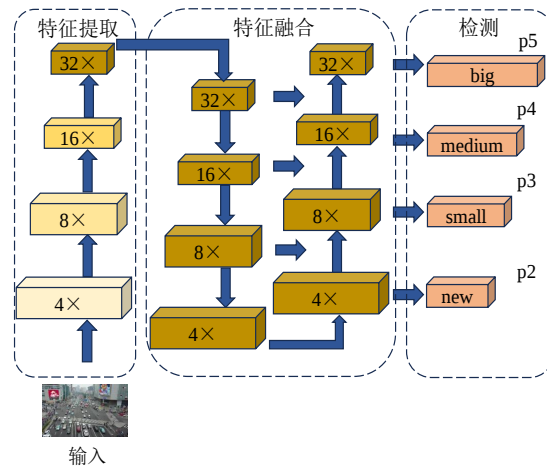


图 4-4 新增检测层网络模型示意图

然而，对于无人机航拍图像而言，其数据集中目标主要以小尺寸行人为主，几乎不存在大尺寸目标。因此，模型中针对大尺寸目标设计的 32 倍下采样相关模块在实际应用中失去了其传统意义。这些模块原本旨在捕捉大尺寸目标的全局特征信息，通过扩大感受野来提取目标的上下文语义信息。但在航拍行人检测任务中，由于目标尺寸普遍较小，高层特征的语义信息对小目标检测的贡献有限，反而增加了不必要的计算开销。

如图 4-5 所示，本文针对上述问题对第 3 章所提算法进行了优化，剔除了 32 倍下采样的特征提取模块、对应的特征融合模块以及检测端。这一改进改善了原始网络结构存在的冗余问题，减少了模型的计算量和模型权重大小，为后续嵌入式设备的部署提供了便利。

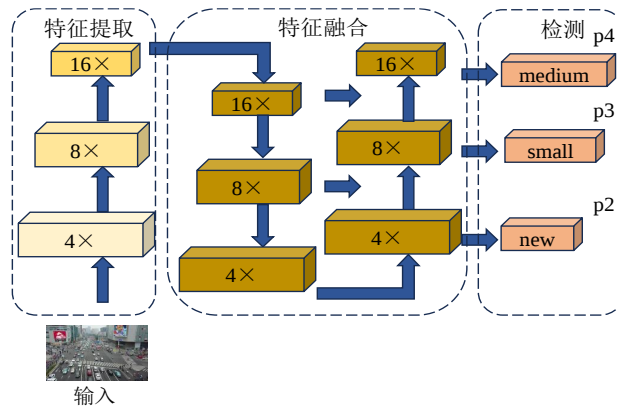


图 4-5 剪枝网络模型示意图

4.3 航拍行人轻量化检测模型

本章针对无人机航拍行人检测任务，提出了两种模型轻量化优化方法。并将轻量化方法作用于第 3 章提出的改进算法中，提出了一种航拍行人轻量化检测算法，航拍行人轻量化检测算法的网络结构，如图 4-6 所示。

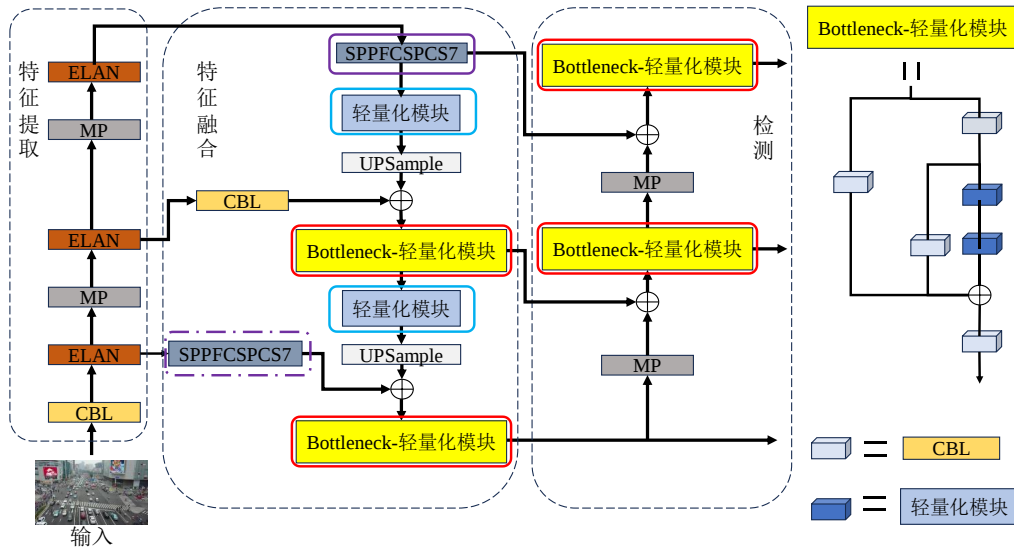


图 4-6 航拍行人轻量化检测模型网络结构图

该算法主要以第 3 章提出的改进 YOLOv7 目标检测算法为基础，进行了轻量化处理。虽然基于特征增强改进的 YOLOv7 算法提高了对航拍图像中行人目标的检测精度，但在模型部署方面，仍然存在以下不足：

- (1) 普通卷积。在特征融合网络中，仍然利用普通卷积进行多尺度融合。普通卷积需要对输入特征图的每个位置进行密集的乘加运算，计算复杂度较高。特别是在输入特征图尺寸较大或通道数较多时，计算量会显著增加。
- (2) 增加检测层。虽然可以提升模型对多尺度目标的检测能力，但这样增加了特征融合网络和检测端的计算复杂度。

上述不足会增加嵌入式设备的资源利用率，包括计算资源、存储空间和功耗等方面，从而限制模型在无人机等资源受限平台上的实际部署。为改善这一问题，本章从模型结构优化和剪枝两个方面提出了针对性的解决方案。首先，采用本章提出的轻量化卷积模块替换了特征融合网络中的普通卷积模块，搭建了轻量级的特征融合网络。该轻量化卷积模块通过结合深度可分离卷积和注意力机制，在减少参数量和计算复杂度的同时，保留了模型对关键特征的提取能

力。其次，针对航拍图像中目标尺寸普遍较小的特点，本文对大目标的特征提取与融合部分以及检测端进行了剪枝操作，移除了对检测性能影响较小的冗余模块，并优化了特征融合路径。

4.4 实验与分析

4.4.1 消融实验

为验证本文所提方法对模型性能影响，基于无人机航拍行人检测数据集，以 YOLOv7 算法为基准，将各种方法依次加入，利用消融实验评估本章提出的轻量化方法对检测算法最终性能的影响。从表 4-2 可以看出，相较于 YOLOv7 基准算法，方法 a 通过引入轻量化模块，在检测精度 mAP0.5 上提升了 0.3%，同时减少了模型权重大小和计算量。方法 b 通过增加 p2 小目标检测层，使 mAP0.5 提升了 4.1%，但增加了权重大小和计算量。方法 c 在方法 b 的基础上，采用模型剪枝技术剔除了大目标检测层，虽然 mAP0.5 下降了 0.5%，但权重大小降低至 36.3M。方法 e 结合轻量化模块和模型剪枝技术，应用于第 3 章提出的基于特征增强改进的 YOLOv7 检测算法，其 mAP0.5 较 YOLOv7 提升了 8.8%，达到 56.1%，同时降低了模型权重大小和计算量。

表 4-2 消融实验

方法	轻量化模块	p2	剪枝	SPPFCSPCS7	NWD	GFLOPs /G	Weight /M	mAP0.5 /%
YOLOv7						105.4	72.7	47.3
a	√					94.2	65.5	47.6
b		√				120.8	77.6	51.4
c		√	√			121.7	36.3	50.9
e	√	√	√	√	√	98.3	29.8	56.1

4.4.2 不同检测算法对比实验

(1) 检测精度对比

为验证本章检测算法在无人机航拍行人检测任务中的有效性，本小节与多种检测算法进行了对比实验。从表 4-3 中可以看出，本章提出的航拍行人轻量化检测算法（YOLOv7-特征增强+轻量化）在检测精度上表现优异，mAP0.5 达到 56.1%，优于通用目标检测算法 Faster-RCNN、Deformable DETR^[64]，以及针对

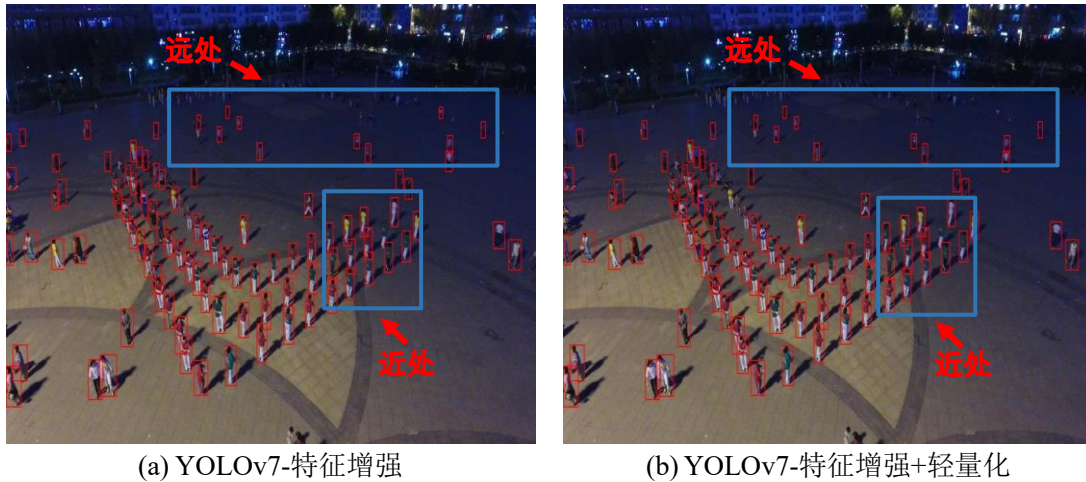
小目标的检测算法 QueryDet^[65]和 TPH-YOLOv5。尽管相较于第 3 章提出的改进算法（YOLOv7-特征增强），检测精度 mAP0.5 略低 0.3%，但其计算复杂度明显降低，模型权重大小仅为第 3 章算法的 36.3%。说明了本章所提两种轻量化方法的有效性，可以在保证算法对航拍图像中行人特征表达能力的同时，降低算法的复杂度。此外，该算法的检测速度为每秒 42 张图片，能够有效满足无人机航拍场景下行人检测任务对实时性的需求。

表 4-3 不同检测模型对比实验

model	GFLOPs/G	Weight/M	mAP0.5/%	FPS
Faster-RCNN	184.5	204.7	33.2	26
Deformable DETR	203.2	187.9	40.1	31
QueryDet	18.3	27.2	30.7	36
TPH-YOLOv5	113.1	171.4	45.5	28
YOLOv7-特征增强（第 3 章）	129.2	82.1	56.4	41
YOLOv7-特征增强+轻量化	98.3	29.8	56.1	42

（2）检测效果图对比

为验证本文算法在极端环境下的检测效果，对不同目标检测算法进行了可视化对比分析。如图 4-7 所示，展示了在行人目标尺寸小、光线昏暗、背景复杂等挑战性场景下不同算法的检测结果。实验结果表明，本文提出的两种改进算法在检测性能上均优于对比算法，改善了极端环境下的漏检问题。针对小目标场景，本文改进算法通过引入特征增强模块和多尺度特征融合策略，提升了小目标的定位与识别精度，避免了因特征提取不足导致的漏检。在低光照条件下，改进算法通过增强对关键特征的提取能力，克服了光照不足的影响，而对比算法性能明显下降；此外，针对复杂背景干扰，本文算法采用注意力机制有效抑制了背景噪声，进一步优化了模型的检测性能。



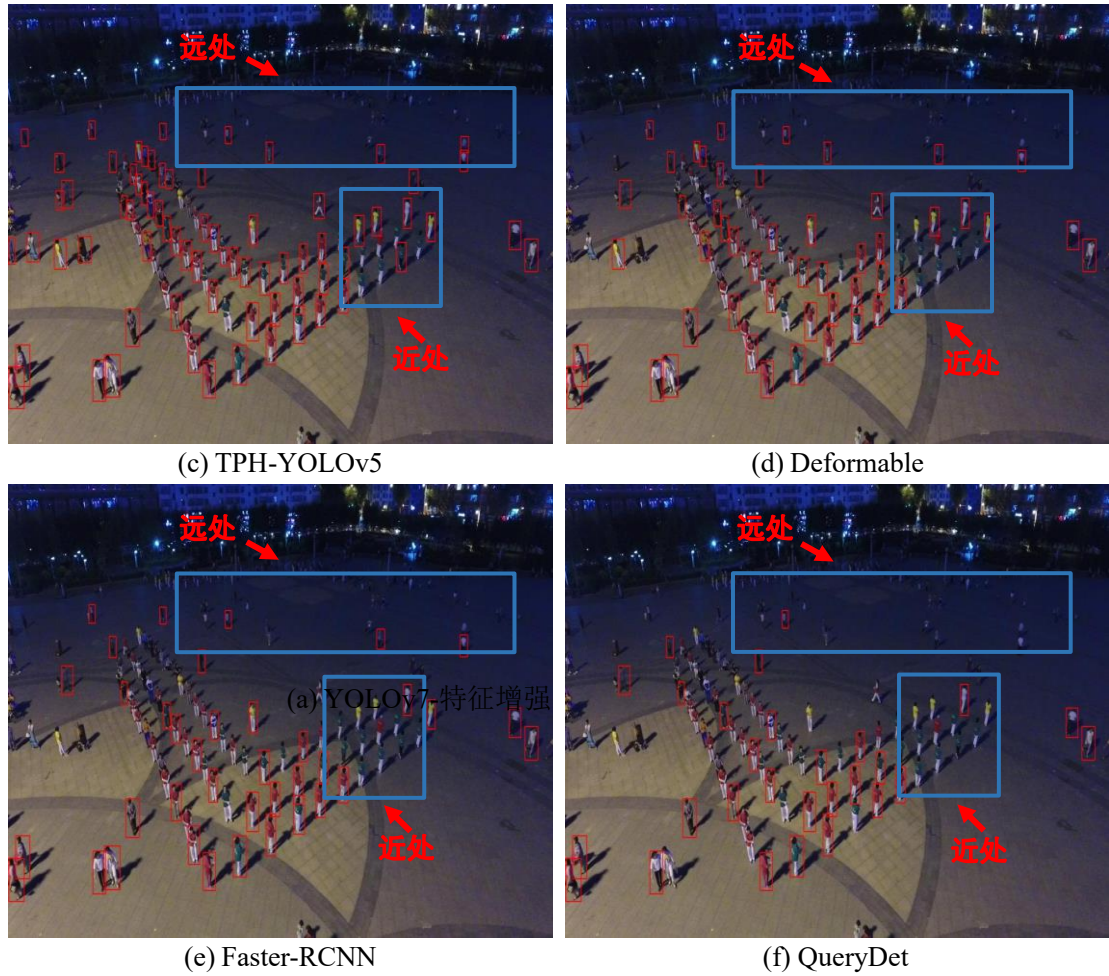


图 4-7 不同检测算法对比结果

4.4.3 改进方法通用性实验

为验证本文所提出的特征增强和轻量化方法对其它目标检测算法的通用性和有效性，选取了 YOLOv5 和 YOLOv8 两种经典算法作为基准模型，并利用航拍行人检测数据集进行了对比实验。将本文提出的优化策略分别集成到 YOLOv5 和 YOLOv8 的原始结构中，并对改进前后的模型性能进行了评估。实验结果如表 4-4 所示，经过本文方法改进后，YOLOv5 和 YOLOv8 算法的检测性能均明显提升，其检测精度 $mAP_{0.5}$ 分别提升了 6.6%、8.6%，同时模型权重大小和计算复杂度也得到降低。实验结果表明，本文提出的特征增强和轻量化方法对模型的检测精度和复杂度都有着优化作用，在增强对航拍图像中行人特征表达能力的同时，减少模型对计算机设备的资源消耗。

表 4-4 检测方法通用性实验

model	GFLOPs/G	Weight/M	mAP0.5/%	FPS
YOLOv5	15.8	9.3	35.2	97
改进 YOLOv5	12.3	4.5	41.8	100
YOLOv8	42.5	27.8	38.2	106
改进 YOLOv8	38.1	13.4	46.8	108

4.4.4 检测系统可视化分析

为验证本章所提算法在航拍视角下的实际检测性能，基于第 3 章设计的航拍行人检测系统，载入本章改进算法的模型权重，通过调用其摄像头功能模拟无人机高空视角进行实验。实验结果如图 4-8 所示，共检测到 11 个目标，未出现明显的漏检现象。此外，系统实时检测速度达到 14FPS，满足无人机航拍行人检测任务的实时性要求。



图 4-8 调用摄像头实时检测效果

4.5 本章小结

本章提出了一种航拍行人轻量化检测方法，旨在降低模型复杂度，便于后续部署。首先，设计了轻量化模块，结合注意力机制、深度可分离卷积和 FReLU 激活函数的优势，在保持特征表达能力的同时降低了模型复杂度。为进一步减少冗余信息干扰，提出了针对航拍数据集的模型剪枝技术。实验部分，通过对比试验和消融实验，验证了本文所提改进算法对航拍行人检测的有效性。随后验证本文方法的通用性，将特征增强和轻量化方法移植到 YOLOv5 和 YOLOv8，改进方法取得了很好的检测效果。

第 5 章 航拍行人检测算法部署实验

在进行无人机航拍检测任务时，无人机通常需要搭载嵌入式设备完成模型部署，本章选用具有“小尺寸、高性能”特点的英伟达 Jetson Xavier NX 小型嵌入式设备，进行算法部署实验。前面两章在 YOLOv7 的基础上，针对无人机航拍行人检测提出了特征增强和轻量化方法，在不增加模型复杂度的前提下提高了检测精度。本章的主要工作是：将改进算法部署到嵌入式平台，对模型推理速度、权重大小、检测精度进行实验论证，并利用 Jetson Xavier NX 模块搭载无人机进行航拍实时检测。

5.1 嵌入式设备选择

无人机按应用领域可分为军用无人机和民用无人机，其中民用无人机进一步分为消费级和工业级两类。军用无人机最早应用于军事领域，对性能要求较高，主要包括战斗机和侦察机等。近年来，无人机技术在民用领域快速发展，广泛应用于农业、环境保护、安全与救援等领域。工业级无人机具有强劲的动力和载重能力，而消费级无人机则因价格低廉、操作简便而广受欢迎，但其设计小巧轻便的特点限制了载重能力。

在航拍检测任务中，嵌入式设备的选取需考虑无人机的载重能力。如表 5-1 所示，通过对当前市场最新消费级无人机的载重能力分析，发现其载重能力普遍有限。因此，需选择重量适宜的嵌入式设备以适配无人机，确保航拍检测任务的顺利执行。

表 5-1 消费级无人机载重能力分析

型号	最大载重
大疆 Mavic3	1200g 左右
大疆 Mavic2Pro	1100g 左右
大疆 Air3S	850g 左右
大疆 Air2S	850g 左右

本文选用 Jetson Xavier NX 小型嵌入式设备进行模型部署实验，其重量约为 50g。该设备能够适配多数消费级无人机，几乎满足市面上所有无人机的航拍检测任务需求。

5.2 Jetson Xavier NX 嵌入式硬件介绍

Jetson Xavier NX 是由 NVIDIA 公司推出的一款高性能人工智能计算平台，专为嵌入式和边缘计算系统设计。该设备在性能、功耗和尺寸之间实现了良好的平衡，其紧凑的尺寸仅为 70×45mm，使其能够广泛应用于各类人工智能场景。在功耗方面，Jetson Xavier NX 在 10W 的功率下可提供高达 14TOPS 的计算性能，而在 15W 的功率下，其计算性能进一步提升至 21TOPS。该处理器集成了 64 位 NVIDIA CPU、384 个 CUDA 核心的 GPU 以及 48 个 Tensor Core，并配备了两个 NVIDIA 深度学习加速器，能够高效支持深度学习推理任务。Jetson Xavier NX 的详细参数如表 5-2 所示。

表 5-2 Jetson Xavier NX 具体参数

名称	型号和参数
尺寸	70mm×45mm
CPU	6-core NVIDIA Carmel ARM®v8.2 64-bit CPU 6MB L2 + 4MB L3
GPU	384core NVIDIA Volta TM GPU with 48 Tensor Cores
显存	8GB/16GB
存储	16 GB eMMC 5.1
功耗	10W/15W/20W
算力	21 TOP

5.3 嵌入式系统环境搭建

Jetson Xavier NX 嵌入式平台具有相对完善的软硬件开发环境，本次实验所用到的硬件设备 Jetson Xavier NX 如图 5-1 所示。

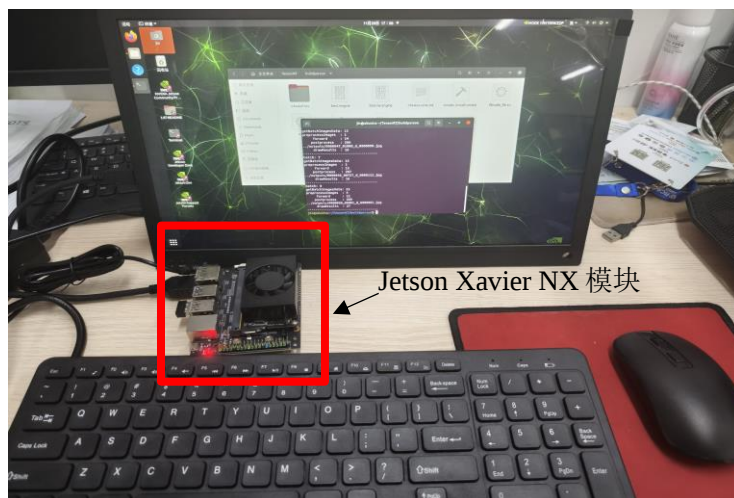


图 5-1 Jetson Xavier NX 硬件配置

模型部署到 Jetson Xavier NX 嵌入式平台，需要配置相关的环境。首先需要安装 JetPack SDK。JetPack SDK 是 NVIDIA 提供的开发套件，为模块提供了一体化的软件包，包含了英伟达 Jetson 系列平台所需的软件工具和驱动程序的套件，主要用于烧录系统镜像、开发和部署应用程序。JetPack SDK 可以根据不同的应用场景选择不同的软件工具，使硬件与软件充分发挥作用，实现各种功能，同时 JetPack 安装了 CUDA、cuDNN、OpenCV、Python 等软件环境。Jetson Xavier NX 模块内部采用的是 Linux 操作系统，并且支持多种深度学习框架和工具，如 TensorFlow、PyTorch、Caffe、MXNet、OpenCV 等。同时，还集成了 NVIDIA TensorRT 深度学习推理引擎，可提供高效的深度学习推理加速。目前常用的卷积神经网络框架有 TensorFlow、Pytorch 等。本章所搭建的软件环境如表 5-3 所示。

表 5-3 软件环境

环境	版本号
Linux 系统	Ubuntu20.04
Python	3.8.10
Pytorch	2.0.0
OpenCV	4.5.4
CUDA	11.4
cuDNN	8.6

5.4 模型部署

5.4.1 TensorRT 加速推理实现

TensorRT 的优化是通过以下几个步骤实现的：

（1）模型解析和优化：将训练好的深度学习模型导入 TensorRT 中，并对模型进行解析和优化，包括层级融合、卷积核优化等。

（2）引擎构建：将优化后的模型转换为 TensorRT 引擎，包括将模型参数、权重和结构保存到硬盘，以便在之后的推理中使用。

（3）数据预处理：将输入数据预处理成 TensorRT 支持的格式，包括将数据转换为 GPU 内存格式、调整数据大小等。

（4）推理运行：预处理后的数据被送入 TensorRT 引擎中，进行推理计算，最终输出模型的检测结果。

在深度学习模型的部署过程中，模型的解析与优化、引擎构建及数据预处理等步骤可在离线环境中预先完成，从而在实际推理时快速加载已处理好的模型和数据。推理过程可根据应用场景需求选择在线或离线执行。TensorRT 提供了一套完整的 API 和命令行工具，支持用户灵活配置和调整优化策略，以满足不同模型和应用需求。此外，TensorRT 兼容多种架构和数据格式，如 TensorFlow、PyTorch 和 ONNX 等，便于用户将深度学习模型整合进 TensorRT 进行高效优化和加速。

将航拍行人轻量化检测模型部署到 Jetson Xavier NX 平台上，采用 200 张测试数据图像测试模型的推理速度。模型在嵌入式设备上平均推理时间如表 5-4 所示，从表中可以看出采用 TensorRT 加速可显著提升模型在 Jetson Xavier NX 的推理速度，且与 PC 的检测速度不相上下。

表 5-4 TensorRT 加速推理实验对比

处理方式	输入尺寸	Infer(ms)
无 TensorRT 加速	640×640	352
有 TensorRT 加速	640×640	23.4
NVIDIA GeForce GTX 3060	640×640	23.8

5.4.2 模型大小分析

模型从 PC 端部署到 Jetson Xavier NX 平台上，因为操作系统不同，模型权重文件格式需要进行转换。首先，本文模型通过在 PC 端的训练得到 pt 格式的权重文件，然后将其转换为 onnx 格式。随后，本文采用 TensorRT 加速部署模型，将部署到 Jetson Xavier NX 端的权重文件 onnx 格式转换为 enging 格式。此操作往往使得模型权重文件增大，因此对模型进行轻量化处理可以有效节约 Jetson Xavier NX 存储空间，转换后模型大小如表 5-5 所示。

表 5-5 不同检测端模型权重大小对比

序号	模型	PC (M)	嵌入式 (M)
a	YOLOv7	72.7	141
b	YOLOv7-特征增强 (第 3 章)	82.1	163.66
c	YOLOv7-特征增强+轻量化 (第 4 章)	29.8	56.6

其中 b、c 分别表示第 3 章和第 4 章所提出的改进算法。模型 a、b、c 从 PC 端部署到 Jetson Xavier NX 平台上，其模型的权重大小都增加了近一倍。模型 c 在 PC 端针对模型 b 进行了轻量化处理，其模型权重大小仅为模型 a、b 的 41%、

36.3%，且部署到 Jetson Xavier NX 平台后，其模型权重大小也仅有 56.6M，大大节省了 Jetson Xavier NX 平台的内存空间。

5.4.3 模型检测性能分析

针对无人机航拍行人检测任务，第 3 章提出了一种基于特征增强改进的 YOLOv7 算法，并在 PC 端验证了其改进效果。第 4 章进一步利用轻量化方法对该算法进行优化，提出了一种航拍行人轻量化检测算法，在不损失精度的前提下，降低了模型大小和计算复杂度。前两章的实验均在 PC 端完成。本小节将 YOLOv7、基于特征增强的改进算法以及航拍行人轻量化检测算法的训练权重文件移植至 Jetson Xavier NX 平台，并在该设备上对模型检测精度进行测试。所有算法的权重都是在 PC 端进行训练得到，且 Jetson Xavier NX 平台与 PC 端硬件不一样。故理论上来说，模型在小型嵌入式设备上的检测效果略逊于 PC 端。表 5-6 对三种不同模型在 Jetson Xavier NX 和 PC 端进行了精度对比。从表中可以看出，模型部署后，其检测精度都略微下降，但其对实际应用效果的影响可忽略不计。

表 5-6 不同检测端模型精度对比

模型	PC (mAP0.5)	嵌入式 (mAP0.5)
YOLOv7	47.3	47.1
YOLOv7-特征增强 (第 3 章)	56.4	56.1
YOLOv7-特征增强+轻量化 (第 4 章)	56.1	55.9

为了直观验证模型在 Jetson Xavier NX 端的检测性能，如图 5-2 所示，采用了昏暗、密集场景的航拍图像进行对比实验。



图 5-2 Jetson Xavier NX 端检测效果对比

图 5-2 中展示了三种不同算法在 Jetson Xavier NX 小型嵌入式设备上的检测效果。如图 5-2 中 a、b 所示，在行人目标尺寸较小，且存在密集遮挡、背景复杂等情况下，本文提出的算法表现均优于原模型，对行人检测的检测效果得到了明显改善。

5.4.4 无人机航拍行人实时检测

为进一步验证本文提出的航拍行人检测算法在实际场景中的检测效果，利用无人机搭载 Jetson Xavier NX 嵌入式设备，在大学操场、篮球场等多种场景下进行实时检测。

采用了行人多尺度、遮挡和背景复杂度情况下的检测效果图进行分析。如图 5-3 中 a、b、c、d 所示，无人机在低空和高空时都取得了很好的检测效果，特别在高空状态下目标尺度较小的情况，本文算法几乎可以检测出图中的所有行人。如图 5-3 中 c、d 所示，在行人之间、行人与树木之间存在遮挡，且由于行人目标尺寸较小、与篮球场地的背景颜色和纹理容易混淆的复杂场景下，本文提出的算法仍表现出较好的检测效果。由实验结果可知，本文提出的无人机航拍行人检测算法可以很好的完成无人机航拍行人检测任务。

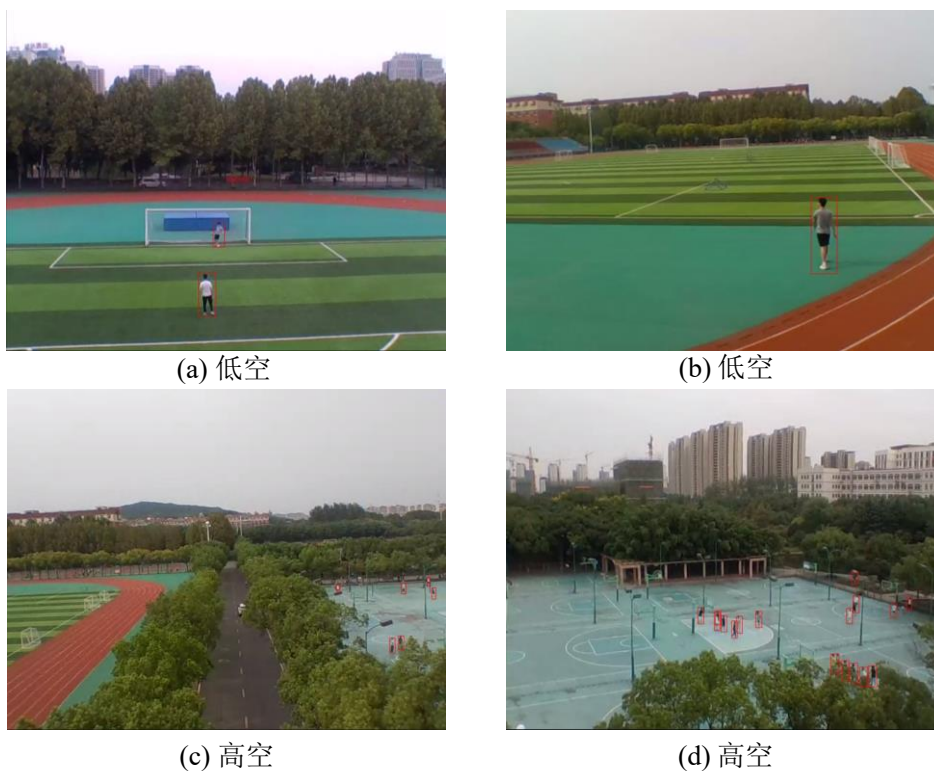


图 5-3 无人机航拍行人检测结果

5.5 本章小结

本章主要验证了本文算法在小型嵌入式设备上的检测性能。首先，通过对无人机载重能力的分析，选择了英伟达 Jetson Xavier NX 模块作为硬件平台。随后，基于 Jetson Xavier NX 平台对本文提出的方法进行了多项实验验证，包括模型检测速率、权重大小以及检测精度的测试。最后，利用大疆无人机搭载 Jetson Xavier NX 嵌入式设备，在真实场景中进行了航拍行人实时检测。实验结果表明，算法能够很好地完成航拍行人检测任务，验证了其在实际应用中的可行性。

第 6 章 总结与展望

6.1 工作总结

本文致力于无人机航拍行人检测任务。特别关注了航拍图像中行人多尺度、遮挡、复杂背景等因素的影响，此外，模型权重和计算量较大对部署到资源有限嵌入式平台也带来了挑战。本文引入深度学习的方法，提高了行人检测的精度，并对模型进行了轻量化处理，方便模型部署到嵌入式设备上，完成无人机航拍行人检测任务。全文内容可总结为以下三个部分：

(1) 针对航拍图像中行人多尺度变化、遮挡及背景复杂等情况导致目标检测算法精度低和漏检的问题，本文提出了一种基于特征增强的航拍行人检测方法。该方法以精度和检测速度都较为优异的 YOLOv7 检测算法为基准。增加了针对小目标的 p2 检测层，并提出了一种基于注意力机制的特征增强模块，明显改善了对航拍行人的特征表达能力。其次，损失函数方面采用 NWD 替换 IOU 度量边界框之间的相似度，改善了对小目标定位偏差敏感的问题。在航拍行人检测数据集上，改进模型 Map0.5 达到了 56.4%，较 YOLOv7 原模型提升了 9.1%。

(2) 为平衡算法的检测精度与计算复杂度之间的关系，方便算法部署到资源有限的嵌入式平台，本文提出了一种航拍行人轻量化检测方法。首先，设计了一种基于注意力机制的轻量化模块。其次，针对大目标的特征提取与融合部分以及预测分支进行了裁剪。最后，将本文提出的轻量化方法应用于基于特征增强改进的 YOLOv7 算法上，构建了轻量级的特征融合网络。在航拍行人检测数据集上，轻量化后的算法 Map0.5 达到了 56.1%，较 YOLOv7 算法提升了 8.8%，模型权重大小降为原来的 41%，计算量也得到一定的下降，更适用于资源有限的嵌入式设备。

(3) 为验证本文提出的两种方法实际应用效果。首先，本文设计了基于 PyQt5 的航拍行人检测系统，对特征增强方法和轻量化方法进行了可视化分析。其次，选用英伟达 Jetson Xavier NX 小型嵌入式设备对本文的算法进行部署，在该设备上验证了本文算法对于航拍行人检测任务的有效性，并利用嵌入

式设备搭载无人机完成了航拍实时检测。

综合而言，本文引入了深度学习的方法，分别从模型特征增强、模型轻量化的角度出发对通用目标检测算法进行改进。随后，将改进后算法部署到嵌入式设备并搭载无人机完成航拍行人检测任务。为无人机航拍行人检测任务提供了有效解决方案。

6.2 未来展望

对于无人机航拍行人检测任务，本文提出的算法在检测检测、模型权重大小和计算复杂度方面都取得了不错的效果，并且完成了在小型嵌入式平台的部署。但是，仍然存在以下几点问题：

（1）本文在研究中并未考虑到行人姿态的问题，行人的各种姿态可能会导致不同程度的误检和漏检情况，今后需要对姿态进行分析，进一步提高识别行人的准确率。

（2）随着无人机技术不断发展，从不同的角度航拍的数据集相差甚大。训练后的模型在别的航拍数据集上仍有较好的检测效果，将是以后在不同的数据集上进行迁移学习的核心内容。

参考文献

- [1] Grigorescu S, Trasnea B, Cocias T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.
- [2] 陈宣池.基于自监督学习的小样本图像分类方法研究[D]. 山东: 山东师范大学,2024.
- [3] 冯彦卿. 基于深度学习的视觉目标检测与跟踪技术研究[D]. 湖南: 国防科技大学, 2020.
- [4] 郑凯枫. 无人机航测在大比例尺地形图测绘中的应用探究[J]. 智能城市, 2020, 6(08): 71-72.
- [5] 杜坤. 基于街景影像的人本尺度街道空间规划理论与方法[D]. 湖北: 武汉大学, 2020.
- [6] 吴福明. 基于无人机视频图像的崇礼地区林火检测研究[D]. 江苏: 南京林业大学, 2022.
- [7] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO:Common objects in context[C]//Proceedings of the European Conference on Computer Vision, 2014: 740-755.
- [8] Cui Z, Wang X, Liu N, et al. Ship Detection in Large-Scale SAR Images Via Spatial Shuffle-Group Enhance Attention[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 379-391.
- [9] Amin S M . 基于深度学习网络的时尚物品分类和检索方法研究[D]. 上海: 华东师范大学, 2024.
- [10]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.
- [11]Redmon J, Farhadi A. YOLOv2:better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [12]Redmon J, Farhadi A. YOLOv3:an incremental improvement[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95.

- [13]Bochkovski A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arxiv: 2004.10934, 2020.
- [14]Liu W, Anguelov D, Erhan D, et al. SSD:Single shot multibox detector[C]// Proceedings of the European Conference on Computer Vision, 2016: 21-37.
- [15]周文军.基于 DSSD 的静态手势识别方法研究[D]. 安徽: 合肥工业大学, 2019.
- [16]Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision, 2017: 2980-2988.
- [17]Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [18]Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [19]Ren S, He K, Girshick R, et al. Faster R-CNN:Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [20]Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[J]. Advances in neural information processing systems, 2016, 29.
- [21]He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 386-397.
- [22]He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [23]Zhang S, Wen L, Bian X, et al. Single-shot refinement neural network for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4203-4212.
- [24]Li Y, Chen Y, Wang N, et al. Scale-aware trident networks for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6054-6063.
- [25]Yin Q, Yang W, Ran M, et al. FD-SSD: An improved SSD object detection

- algorithm based on feature fusion and dilated convolution[J]. Signal Processing: Image Communication, 2021, 98: 116402.
- [26]Liang X, Zhang J, Zhuo L, et al. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(6): 1758-1770.
- [27]赵立敏. 基于深度学习的航拍图像目标检测算法研究[D]. 陕西: 延安大学, 2023.
- [28]Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 2778-2788.
- [29]余俊宇, 刘孙俊, 许桃. 融合注意力机制的 YOLOv7 遥感小目标检测算法研究[J]. 计算机工程与应用, 2023, 59(20): 167-175.
- [30]Zhao L L, ZHU M L. MS-YOLOv7:YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography[J]. Drones, 2023, 7(3): 188-188.
- [31]Zhang L, Lin L, Liang X, et al. Is Faster R-CNN doing well for Pedestrian Detection[C]//Proceedings of the European Conference on Computer Vision, 2016: 443-457.
- [32]Liu T, Sathaki T. Faster R-CNN for robust pedestrian detection using semantic segmentation network[J]. Frontiers in neurorobotics, 2018, 12: 64.
- [33]Sandler M, Howard A, Zhu M, et al. Mobilenetv2:Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [34]Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [35]Zhang X, Zhou X, Lin M, et al. ShuffleNet:An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [36]Ma N, Zhang X, Zheng H T, et al. ShuffleNetV2:Practical guidelines for efficient

- CNN architecture design[C]//Proceedings of the 15th European Conference on Computer Vision, 2018: 122-138.
- [37]Zang L, Chen G, Shi Y, et al. Anonymous model pruning for compressing deep neural networks[C]//Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval, 2020: 157-160.
- [38]Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the IEEE international conference on computer vision, 2017: 2736-2744.
- [39]石博雅, 董学峰. 融合 GhostNet 的 YOLOv4 轻量化网络设计与实现[J]. 小型微型计算机系统, 2024, 45(03): 651-656.
- [40]李秀芳. 深度知识蒸馏网络感知与学习分类[D]. 陕西: 西安电子科技大学, 2024.
- [41]Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [42]张润梅, 肖钰霏, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法[J]. 光电工程, 2024, 51(05): 92-103.
- [43]梁海翔, 唐艳慧, 王宇庆, 等. 高分辨率遥感图像的目标检测[J]. 液晶与显示, 2024, 39(10): 1350-1360.
- [44]侯志强, 刘晓义, 余旺盛, 等. 使 GIoU 改进非极大值抑制的目标检测算法[J]. 电子学报, 2021, 49(04): 696-705.
- [45]Lin T Y, DOLLAR 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.
- [46]Liu S, Qi L, Qin H F, et al. Path Aggregation Network for Instance Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [47]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31th Annual Conference on Neural Information Processing Systems, 2017: 5998-6008.

- [48]Hou Q, Zhou D, Feng, J. Coordinate attention for efficient mobile network design[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition, 2021: 13708-13717.
- [49]Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 42: 318-327.
- [50]Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6450-6458.
- [51]唐东林, 吴续龙, 周立, 等. 结合深度学习与注意力机制的墙体安全检测模型[J]. 安全与环境学报, 2022, 22(01): 8-15.
- [52]Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [53]Li Z Z, Qu N, Li X X, et al. Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms[J]. Journal of Power Electronics, 2021, 21: 1030-1040.
- [54]Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
- [55]Chang L, Zhang S, Du H, et al. Position-aware lightweight object detectors with depthwise separable convolutions[J]. Journal of Real-Time Image Processing, 2020, 18(3): 1-15.
- [56]Deng C W, Jing D L, Ding Z H, et al. Sparse Channel Pruning and Assistant Distillation for Faster Aerial Object Detection[J]. Remote Sensing, 2022, 14(21): 5347-5347.
- [57]Cui B Y, Li Y M, Zhang Z F. Joint structured pruning and dense knowledge distillation for efficient transformer model compression[J]. Neurocomputing, 2021, 458: 56-69.
- [58]Zhou W, Wang X, Fan Y, et al. KD SMALL: A lightweight small object detection algorithm based on knowledge distillation [J]. Computer Communications, 2024, 219: 271-281.

- [59]Du D, Zhu P, Wen L, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 213-226.
- [60]Lu X, Li B, Yue Y, et al. Grid R-CNN[C]. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7363-7372.
- [61]Wu Y, Chen Y, Yuan L, et al. Rethinking classification and localization for object detection[C]. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10186-10195.
- [62]Qiu S, Xu X, Cai B. FReLU: flexible rectified linear units for improving convolutional neural networks[C]//Proceedings of the 24th International Conference on Pattern Recognition, 2018: 1223-1228.
- [63]He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026-1034.
- [64]江志鹏, 王自全, 张永生等. 基于改进Deformable DETR的无人机视频流车辆目标检测算法[J]. 计算机工程与科学, 2024, 46(01): 91-101.
- [65]Yang C, Huang Z, Wang N. Querydet:Cascaded sparse query for accelerating high-resolution small object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 13668-13677.

攻读硕士学位期间所发表的论文及其它研究成果

学术论文：

居鸿源,江明,李云飞.重构SPPFCSPC多尺度航拍小目标检测算法[J].安徽工程大学学报（已录用）

致谢

行文至此，落笔为终。当毕业论文的最后一个句点画下，我的三年硕士求学生涯也即将落幕。此刻回首，一路走来虽有坎坷，但更多的是成长与收获。在此，我谨以最诚挚的谢意，向所有给予我支持、帮助和陪伴的师长、亲友和同窗致以最深的感激。

首先，我要衷心感谢我的导师江明教授。感谢您以渊博的学识和严谨的治学态度为我指明研究方向。从论文选题到最终成稿，您始终给予我耐心的指导和宝贵的建议。您不仅教会我如何做研究，更让我明白为人处世的道理。

感谢课题组的全体老师与同门。你们让我明白，科研从来不是孤独的跋涉，而是团队协作的征程。那些并肩作战的时光，将成为我人生中最温暖的记忆。

感谢我的父母与女朋友小徐同志。是你们用无条件的爱与支持为我筑起最坚实的后盾。你们教会我善良与坚韧，让我无论顺逆都能保持前行的勇气。未来的路或许仍有荆棘，但这段淬炼成长的岁月，已让我有底气面对任何挑战。

山水一程，三生有幸。谨以此文献给所有照亮我求学之路的星辰，愿我们都能在各自的人生中奔赴山海，静待春华。

尚德敏学 唯实惟新

