

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2210330127



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 面向航拍图像的
小目标检测算法研究

论 文 作 者 张健

校 内 导 师 江明（教授）

校 外 导 师 陈家龙

专业学位类别 电子信息

研究方向（领域） 人工智能与系统集成

论文提交日期: 2024 年 5 月 31 日

分类号：TP391

单位代码：10363

密类级：公开

学 号：2210330127

题 目 面向航拍图像的小目标检测算法研究

题 目 RESEARCH ON SMALL TARGET
DETECTION ALGORITHM FOR
AERIAL IMAGES

学 生 姓 名 ： 张健

校 内 导 师 ： 江明（教授）

校 外 导 师 ： 陈家龙

专业学位类别 ： 电子信息

研究方向（领域）： 人工智能与系统集成

论文答辩日期 ： 2024 年 5 月 19 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：张健

日期：2024 年 5 月 25 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：张健

指导教师签名：



日期：2024 年 5 月 25 日

日期：2024 年 5 月 25 日

面向航拍图像的小目标检测算法研究

摘 要

随着人工智能技术的飞速发展,深度学习在计算机视觉领域取得显著进步,航拍图像小目标检测作为其重要分支,广泛应用于军事和民用领域。但航拍图像小目标检测面临着一些难点。

(1) 多尺度目标及目标密集分布。航拍图像因远距离拍摄导致目标像素少,且不同高度航拍造成目标尺度多变,形成多尺度目标共存。同时,目标数量多且紧密排列,影响特征信息提取。

(2) 目标方向任意及复杂背景。航拍图像中目标角度多样,水平框检测易受倾斜目标背景的干扰,并可能包含过多临近目标信息。此外,航拍场景与背景易混淆,增加检测难度。

本文针对上述问题,分别提出了两种方法。首先,针对航拍图像中存在被测对象在整幅图像中占比较小、多尺度目标共存、密集分布等情况,导致目标检测网络获取到有效信息较少的问题,本文提出一种基于特征增强的航拍图像目标检测方法。该方法在主干特征提取网络中引入 SPD-Conv 模块,解决下采样时图像信息大量丢失的问题。在此基础上加入注意力机制增强对密集目标感兴趣区域的获取,提高主干特征提取能力。改进多尺度目标检测模块,保持模型对大目标检测能力的同时,提升网络对航拍图像中小目标的检测性能,进一步强化特征表达能力。

其次,针对航拍图像中目标在俯视下呈任意方向排列,水平目标

框里存在大量背景信息干扰的问题,本文提出一种融合注意力的航拍图像旋转目标检测方法。首先设计了一种增强注意力机制,可以对主干特征结构的特征信息进行再处理,增强在复杂背景下对航拍图像小目标的特征提取能力。然后提出了双线性注意力平行融合机制,集成不同层次的特征图,结合多尺度的上下文信息,积累多个特征,加强特征信息的融合。最后,为了解决物体任意角度检测的问题,使用 KLD 改进分类的损失函数,提高任意角度物体的检测精度。

最后,为评价本文所提出方法的优势,本文进行了多组实验,验证网络模型的检测效果。同时通过 PyQt5 设计航拍图像小目标检测客户端软件,在满足软件需求的前提下,设计了功能模块和底层代码。并将其与改进的方法相融合,以验证在实际应用中的效果。

关键词: 航拍图像, 小目标, 特征增强, 旋转目标检测, 注意力机制

RESEARCH ON SMALL TARGET DETECTION ALGORITHM FOR AERIAL IMAGES

ABSTRACT

With the rapid development of artificial intelligence technology, deep learning has made significant progress in the field of computer vision, and small target detection in aerial images, as its important branch, is widely used in military and civilian fields. However, aerial image small target detection faces some difficulties.

(1) Multi-scale targets and dense target distribution. Aerial images have fewer target pixels due to long-distance shooting, and aerial photography at different heights results in variable target scales, resulting in the coexistence of multi-scale targets. At the same time, the number of targets is large and closely arranged, which affects the extraction of feature information.

(2) Arbitrary target direction and complex background. Targets in aerial images have various angles, and horizontal frame detection is easily interfered with by the tilted target background and may contain too much information about neighboring targets. In addition, the aerial scene and the background are easily confused, which increases the detection difficulty.

In this paper, two methods are proposed respectively for the above problems. Firstly, for the aerial image, there are situations such as the object under test occupies a relatively small proportion in the whole image, the coexistence of multi-scale targets, and the dense distribution. Resulting in the problem that the target detection network acquires less effective information, this paper proposes a feature enhancement-based target detection method for aerial images. The method introduces the SPD-Conv module in the backbone feature extraction network to solve the problem of large loss of image information during downsampling. Based on this, an attention mechanism is added to enhance the acquisition of dense target regions of interest and improve the backbone feature extraction capability. Improve the multi-scale target detection module. While maintaining the model's ability to detect large targets, improve the network's performance in detecting small targets in aerial images. Further, strengthens the feature expression ability.

Secondly, the problem is that targets in aerial images are arranged in arbitrary directions in the top view and there is a lot of background information interference in the horizontal target frame. In this paper, we propose a fusion attention method for rotating target detection in aerial images. Firstly, an enhanced attention mechanism is designed, which can reprocess the feature information of the main feature structure and enhance the feature extraction ability of small targets in aerial images

under complex backgrounds. Then a bilinear attention parallel fusion mechanism is proposed, which integrates different levels of feature maps, combines multi-scale contextual information, accumulates multiple features, and enhances the fusion of feature information. Finally, to solve the problem of object detection at arbitrary angles, the loss function of classification is improved using KLD to enhance the detection accuracy of objects at arbitrary angles.

Finally, to evaluate the advantages of the proposed method in this paper. In this paper, several sets of experiments are conducted to verify the detection effect of the network model. At the same time, the client software for small target detection of aerial images is designed through PyQt5, and the functional modules and underlying code are designed under the premise of meeting the software requirements. And it is integrated with the improved method to verify the effect in practical applications.

Zhang Jian (Electronic Information)

Supervised by Professor Jiang Ming

KEY WORDS: aerial image, small target, feature enhancement, rotating target detection, attention mechanism

目录

摘 要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	2
1.2.1 传统航拍图像小目标检测算法研究现状	3
1.2.2 基于深度学习的航拍图像小目标检测算法研究现状	3
1.3 论文研究内容及架构安排	5
1.3.1 主要研究内容	5
1.3.2 论文组织与架构	6
第 2 章 航拍图像小目标检测相关理论	9
2.1 卷积神经网络理论基础	9
2.2 基于深度学习的目标检测方法	11
2.2.1 二阶段目标检测算法	11
2.2.2 一阶段目标检测算法	12
2.2.3 模型选择	15
2.3 神经网络中的注意力机制	16
2.4 航拍图像小目标定义及技术难点	18
2.4.1 小目标的定义	18
2.4.2 多尺度目标共存及目标密集分布的问题	19
2.4.3 目标方向任意及复杂背景的问题	20
2.5 本章小结	20
第 3 章 基于特征增强的航拍图像目标检测方法	21
3.1 模型整体结构	21
3.2 主干网络特征增强方法	23
3.2.1 Coordinate Attention 增强特征信息获取	23
3.2.2 Transformer 强化全局信息提取	24

3.2.3 SPD-Conv 优化特征提取方法	26
3.3 多尺度检测策略	27
3.4 检测图像对比实验	28
3.5 本章小结	28
第 4 章 融入注意力的航拍图像旋转目标检测方法	29
4.1 网络整体结构	29
4.2 融入注意力机制的目标检测方法	31
4.2.1 增强注意力机制强化特征信息提取	31
4.2.2 双线性注意力平行融合机制促进信息融合	32
4.3 旋转目标检测策略	35
4.4 检测可视化实验	36
4.5 本章小结	37
第 5 章 实验分析及客户端软件设计	38
5.1 基于特征增强的航拍图像目标检测方法的实验分析	38
5.1.1 特征增强目标检测方法的实验设置	38
5.1.2 特征增强目标检测方法的消融实验分析	39
5.1.3 特征增强目标检测方法的对比实验分析	41
5.2 融入注意力的航拍图像旋转目标检测方法的实验分析	43
5.2.1 注意力旋转目标检测方法的实验设置	43
5.2.2 注意力旋转目标检测方法的消融实验分析	44
5.2.3 注意力旋转目标检测方法的对比实验分析	45
5.3 客户端软件设计	48
5.3.1 PyQt5 简介	48
5.3.2 客户端软件架构	49
5.3.3 客户端软件展示	50
5.4 本章小结	51
第 6 章 总结与展望	53
6.1 总结	53
6.2 展望	54

参考文献.....	55
攻读学位期间发表的学术论文与研究成果.....	61

第 1 章 绪论

1.1 研究背景和意义

随着科技发展,人工智能时代已经到来。计算机视觉作为代表之一,能够高效提取并分析视频或图像中的关键信息,从而在生活中多个领域以及工作场景中有效替代人眼。目标检测是计算机视觉领域的重要分支,近年来得到了快速发展。其主要任务是获取图像中目标的位置以及类别^[1],在自动驾驶、人脸检测、医疗诊断等领域得到广泛地应用^[2]。

航拍图像小目标检测是指利用航拍设备在高空中对目标进行拍摄,通过网络模型对获取到航拍图像中的目标实施准确识别和定位。航拍图像小目标检测的应用场景如图 1-1 所示。在救灾方面^[3-4],通过对航拍救援现场的图像进行分析,确定受困者的位置信息,进而及时安排救援。在交通管理方面^[5],可以通过航拍实时监控当前道路上车辆通行情况,在发生交通事故、车辆拥堵等情况时,及时提醒相关人员。在农业方面^[6-7],利用航拍图像检测实现高精度农药喷洒作业,有助于提高工作效率。除此之外,也可以进行农作物监控,检测航拍图像中的农作物是否存在病虫害情况。在高空检查方面^[8-9],使用航拍图像对高压电线等高空目标进行检测,帮助工作人员快速发现异常情况并及时维修。此外,航拍图像目标检测在环境治理^[10]、城市管理^[11]以及行人监控^[12]等方面也有应用价值。



图 1-1 航拍图像小目标检测的应用场景

由于航拍角度的原因，航拍图像目标通常具有以下特点：

(1) 背景复杂多变。航拍场景多样，环境复杂，导致航拍图像中的小目标易与背景混淆，且受到遮挡、物体群聚、光照和阴影等多种因素干扰，从而增加了检测难度。

(2) 目标尺寸较小，特征模糊。由于需要远距离探测，航拍的目标在图像中尺寸微小，像素有限，导致特征模糊，难以准确检测和区分，易出现漏检或误检情况。

(3) 目标分布不规律。航拍轨迹的灵活多变，使得目标在图像中的分布和聚集方式杂乱无章，增加了检测任务的搜索难度和成本。

深度学习技术的出现，使得目标检测得到了巨大的进步，其强大的特征提取能力，解决了传统目标检测算法特征信息提取不足的问题。同时算力设备的进步，也推动了深度学习技术的发展。但现有的深度学习目标检测模型大多是针对自然图像数据集开发的，面对航拍图像小目标的特殊性，现有检测方法对航拍图像中小目标的特征处理还不够完善，不能完全应对实际复杂场景。

本文主要是研究面向航拍图像小目标的检测方法。针对航拍图像中多尺度目标共存并且目标密集等问题，结合增强特征表达的方法来优化目标检测网络结构。由于航拍图像中的部分目标是以任意的角度分布在图像各处，在检测时更易受到背景的干扰。利用融入注意力机制的旋转目标检测方法，减少背景的干扰，使物体在倾斜情况下也能准确识别出来。解决上述问题，提高航拍图像小目标的检测精度，具有显著的研究意义。其可以推动目标检测领域的发展，提高我国的科技创新水平。

1.2 国内外研究现状

目标检测算法的发展历程如图 1-2 所示，一般可分为传统目标检测算法和深度学习目标检测算法。传统目标检测算法在特征提取上依赖手动设计，对丰富的特征处理没有较好的鲁棒性。深度学习目标检测算法通过神经网络自动学习并提取图像中的特征，能够学习到更高级、更复杂的特征表示。人工智能技术的崛起，使得基于深度学习的目标检测方法开始成为主流。

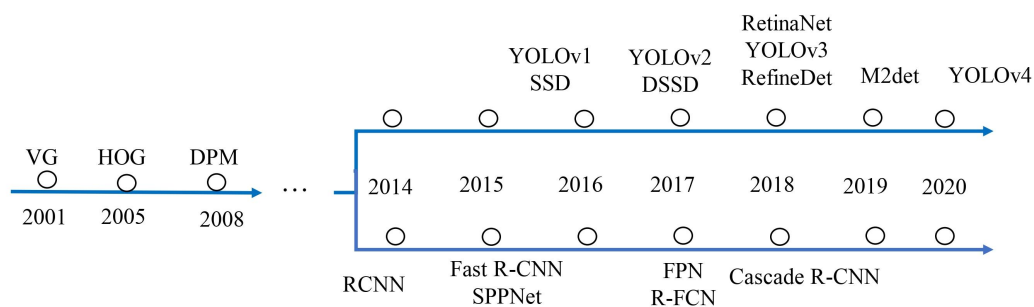


图 1-2 目标检测算法发展历程

1.2.1 传统航拍图像小目标检测算法研究现状

传统目标检测方法有区域选择、特征提取和分类器三个环节。首先通过滑动窗口对输入图像进行遍历，筛选出候选目标区域。然后在候选区域中提取关键特征信息，分类器接受关键信息进行训练。最后输出结果进行目标的分类。在不同图像中目标的位置和大小是不确定的，因此对图像进行遍历时，就需要设计不同尺度的滑动窗口。然而，这种穷举的策略存在冗余和过程复杂的问题，影响提取过程。在提取时，一般使用局部二值模式^[13]、尺度不变特征变换^[14]和方向梯度直方图^[15]等方法。目标形态的各异、环境的多变，这使得设计具有可靠稳定的特征极具挑战性。特征的质量直接决定了后续分类的精准度，传统目标检测方法实际使用情况较差，因此逐渐被深度学习目标检测算法取代。

1.2.2 基于深度学习的航拍图像小目标检测算法研究现状

针对传统目标检测方法效果不佳的情况，有学者开始利用深度学习方法进行研究。近年来，对于航拍图像小目标存在的检测难题，学者们相继提出优化算法。由于像素的限制，大步幅的下采样因子往往会使小物体的特征信息在低分辨率特征图中成为小点甚至消失。为此，Liu 等^[16]提出一种多分支并行特征金字塔网络结构，旨在提取小尺寸物体更丰富的特征信息。利用并行分支恢复深层中遗漏的特征，应用监督空间注意力模块来削弱背景噪声对目标信息的影响。因航拍图像中目标尺度差异大，Lin 等^[17]提出 Trident-FPN 骨干结构。设计了一种新的注意力机制，解决航拍图像中多尺度目标预测的问题。对于航拍图像中存在目标尺寸小的问题，Li 等^[18]提出一种密集注意力残差网络。该网络采用的新型密集瀑布残差块可有效保存空间信息，提取高层语义信息。设计多尺度感受场注意力模块，

从特征图中选择信息特征，增强多尺度感知能力。针对在不同的高度航拍时，目标尺度变化大的问题，Zhu 等^[19]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。由于拍摄距离偏大，导致航拍图像中的物体通常很小，Liang 等^[20]提出了一种基于特征融合和缩放的单次检测器(Feature Fusion and Scaling-based Single Shot Detector, FS-SSD)。通过设置反卷积模块，调整了原有的特征融合分支，以更好地适应航拍图像目标检测任务。针对航拍图像中目标大尺度变化和密集的目标分布特性给目标检测器带来的挑战，Wang 等^[21]提出一种名为 SPB-YOLO 的高效端到端探测器。设计了一个条带瓶颈(Strip Bottleneck, SPB)模块，通过注意力机制提高航拍图像中不同尺度物体的检测灵敏度，以更好地应对宽高依赖关系。并提出一种基于路径聚合网络的上采样策略，增加一个检测头，专门处理密集目标分布的检测任务。

航拍图像大都是俯视成像，导致图像中目标排列方向具有多样性，水平框内倾斜的目标会存在大量的背景信息，以及临近目标信息的干扰，影响检测效果。针对这一特殊性，研究者们又开始探索采用旋转框对目标进行检测，因此基于旋转框(Oriented Bounding Boxes, OBB)的方法被广泛应用于航拍图像的目标检测。

大多数现有的基于 OBB 检测方法都存在边界不连续性问题，无法准确预测边界框角度，Gao 等^[22]提出一种基于 OBB 的新型检测算法。该算法的椭圆编码有效地利用了船舶目标的几何和散射特性，通过将弓角区间与译码过程相结合，避免边界不连续问题的负面影响。并提出椭圆高斯分布热力图和双峰最大池化的策略，以应对在近海场景中分离密集分布船舶的挑战。针对航拍图像中目标方向任意、背景复杂等问题，Zhang 等^[23]提出一种更有效的旋转框目标检测算法。该算法引入了多分支卷积注意力，通过在传统的卷积层和激活函数中加入恒等式、残差分支和注意力机制来构建模块，以替换网络中的简单卷积模块。在骨干网络末端引入 Transformer 模块，增强全局感知能力。通过双向特征金字塔网络来改进颈层结构以加快网络运行速度，并在 CspLayer 之后添加注意力机制以突出显示重要的特征信息。小至 2×2 像素的微小对象出现在航拍图像中，对目标检测算法提出了挑战。更重要的是，当前的目标检测方法最初设计用于适应轴对齐的边界框检测，因此无法准确定位最能描述自由旋转物体的定向框。Zand 等^[24]提出

了基于卷积神经网络的方法，在多个尺度级别使用潜在的像素信息，而无需任何外部资源。该方法对目标特征的精确位置和方向进行编码，与现有回归边界框的位置和维度的方法不同，所提出的方法通过分类学习所有必需的信息，无需任何额外计算即可实现定向边界框检测。现有的高性能无锚旋转目标检测方法相对复杂，推理速度也较慢。Liu 等^[25]提出了一种基于 YOLOX 方法的无锚旋转探测器，用于航拍图像中的物体检测。该方法包括两个改进：一个细化旋转模块(Refined Rotated Module, RRM)和一种高斯分布抽样最优传输分配(Gaussian Distribution Sampling Optimal Transport Assignment, GSOTA)方法。RRM 可以对齐特征，并为探测器获得更多有用的先验信息，GSOTA 利用高斯分布对定向边界框进行建模。

为减少背景的影响，一些研究者们将注意力机制整合到网络模型中。Chen 等^[26]提出一种多尺度空间和通道注意力机制(Multi-scale Spatial and Channel-wise Attention, MSCA)。MSCA 通过集成多尺度信息的注意力分布图，将其应用于深度网络的特征图，在 MSCA 的作用下，网络模型可以有效地聚焦航拍图像中不同背景和大小物体。Liu 等^[27]提出全局上下文平行注意力机制，由一个并行的全局上下文通道自注意力块和一个空间自注意力块组成。能够同时捕获全局空间和信道信息，显著提高分割精度。Xu 等^[28]提出了一种混合注意力机制来分类和识别不同类型的视网膜病变图像，该模块包括平行空间注意力机制和通道注意力机制，提取视网膜病变图像通道和空间维度的关键特征，减少背景信息的影响。

上述的改进方法取得了一些效果，但是航拍图像中目标像素偏小、多尺度目标共存、背景复杂等问题影响依旧。因此，提高航拍图像小目标检测精度还是一个需要不断努力的方向。

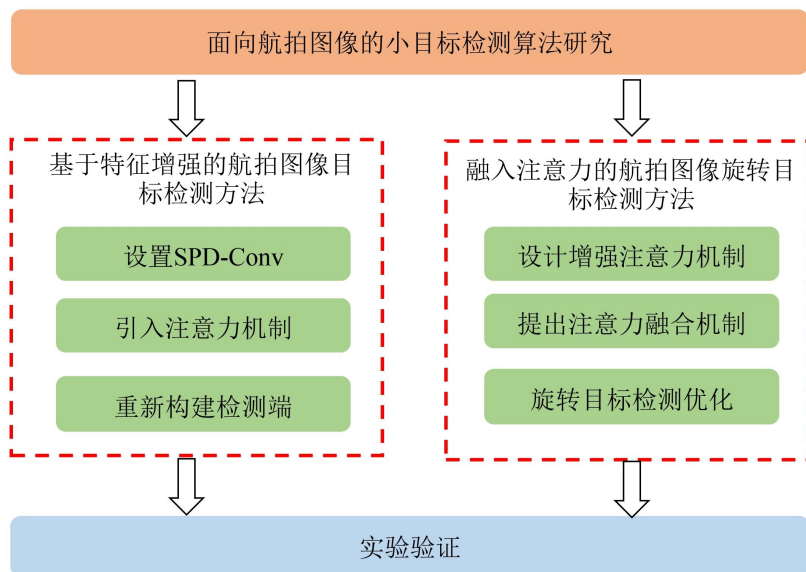
1.3 论文研究内容及架构安排

1.3.1 主要研究内容

本文以航拍图像小目标检测方法为研究对象，针对航拍图像中通用模型对小目标特征处理较差和目标方向任意，导致检测效果差的问题。通过提出两种方法来解决上述问题，本文研究内容如图 1-3 所示，重点研究内容如下：

(1) 针对航拍图像中存在多尺度目标共存以及目标密集分布等问题, 本文提出一种基于特征增强的航拍图像目标检测方法。该方法在 YOLOv5(You Only Look Once version 5)的主干网络设置对称正定卷积(Symmetric Positive Definite Convolution, SPD-Conv)模块, 减少下采样时对小目标特征信息的丢失。同时引入注意力机制, 利用注意力获取图像感兴趣区域, 提高主干网络对特征信息的提取能力。重新设置模型的检测端, 以应对多尺度目标共存的检测任务。进而增强航拍图像中目标的特征表达, 提高密集目标检测能力。

(2) 针对航拍图像中目标呈任意方向排列, 并且存在大量背景和临近目标信息等干扰的问题, 本文提出了一种基于注意力的航拍图像旋转目标检测方法。该方法对 YOLOv7 网络进行旋转目标检测优化, 采用相对熵(Kullback-Leibler Divergence, KLD)改进模型的损失函数, 提高对旋转目标检测精度。提出了增强注意力机制, 加强对航拍图像小目标的特征提取能力。设计了双线性注意力机制, 并构建平行融合机制, 集成不同层次的特征图, 加强特征信息的融合, 进而减少背景信息的干扰。



1.3.2 论文组织与架构

本文共分为 6 章, 具体结构如下:

第 1 章, 绪论。对本文的研究背景与意义进行分析, 概述航拍图像小目标检测的发展, 分析在航拍图像中小目标检测领域的国内外发展现状, 最后提出本文

的重点研究内容与架构安排。

第 2 章，航拍图像小目标检测相关理论。先论述卷积神经网络的相关概念，然后分析两阶段和一阶段目标检测方法的框架，为后续研究提供理论支撑。最后概述注意力机制的发展历程以及阐述航拍图像小目标检测存在的问题，以此引出本文的后续内容。

第 3 章，基于特征增强的航拍图像目标检测方法。针对航拍图像中存在多尺度目标共存和目标密集分布等问题，本文提出改进的 YOLOv5 目标检测方法。在主干网络中引入坐标注意力(Coordinate Attention, CA)模块、Transformer 模块以及 SPD-Conv 模块来增强特征信息提取能力。改进网络检测端，通过输出新的三个检测端对目标进行检测，并在中目标检测端添加 Transformer 模块，克服网络末端图像分辨率较低的问题。在 3.1 节对提出的模型结构进行整体概述。在 3.2 节介绍增强特征提取网络的方法，以解决小目标信息少，提取困难等问题。在 3.3 节提出多尺度检测模块，以实现多尺度目标的检测。在 3.4 节展示检测图像对比实验。最后在 3.5 节总结内容。

第 4 章，融入注意力的航拍图像旋转目标检测方法。因为航拍图像中目标方向任意，水平框会包括背景和临近目标信息，影响检测效果，本文提出一种改进的 YOLOv7 目标检测方法。设计增强的注意力模块用于主干特征提取网络，积累多层特征，加强特征信息提取。利用双线性注意力平行融合机制增强航拍图像中小目标信息的特征融合。利用 KLD 方法提高任意角度目标的检测精度，重新定义分类损失函数，并在损失计算中加入角度问题。在 4.1 节对所提出的模型结构进行概述。在 4.2 节提出两种注意力机制，应用注意力机制构成一种特征融合机制，增强网络对背景信息的过滤能力。在 4.3 节中提出旋转目标检测方法，以实现旋转目标的检测。在 4.4 节展示可视化检测实验。最后在 4.5 节总结内容。

第 5 章，实验分析及客户端软件设计。本章对研究所提的方法在 VisDrone-DET2019 数据集、NWPUVHR-10 数据集以及自建数据集上进行了实验，验证了所提方法的有效性。利用消融实验证明了本研究所提出的各个模块对网络的提升效果，通过对比实验证实了本研究提出方法相较于其他目标检测网络的优越性。同时利用 PyQt5 设计了航拍图像小目标检测客户端软件，实现对航拍图像小目标的识别。

第 6 章，总结与展望。本章对全文工作进行简要的概括，并对存在的不足进行剖析。对航拍图像中小目标检测领域的未来发展趋势进行展望，以期为后续研究提供有价值的参考。

第 2 章 航拍图像小目标检测相关理论

在第 1 章中, 本文对航拍图像小目标检测的研究背景、意义以及研究现状进行了深入分析。为给本文所提方法提供有力的理论支撑和夯实的实验基础, 本章将作为后续章节的基石, 深入探讨深度学习的相关基础理论, 以及对航拍图像小目标定义和技术难点进行阐述。

2.1 卷积神经网络理论基础

作为前馈神经网络^[29]的一种, 卷积神经网络是由多层感知机^[30]演化而来。具有局部连接, 权值共享的特点, 具体组成如图 2-1 所示。

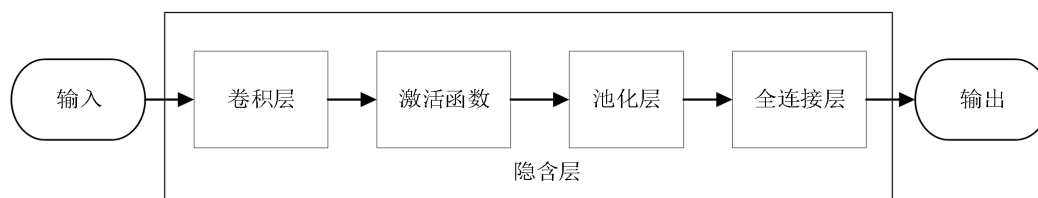


图 2-1 卷积神经网络框架图

(1) 卷积层

卷积层从输入图像中提取特征信息, 揭示图像中的空间关系。在执行卷积运算时, 卷积核(一个预先定义的矩阵)会与图像上的点进行运算, 卷积核的尺寸、数量和填充大小等参数需预先设定。以二维图像为例, 若输入图像为 5×5 大小, 卷积核为 3×3 大小。则从图像的左上角开始, 卷积核会按照一定顺序在图像上进行卷积运算, 卷积操作如图 2-2 所示。

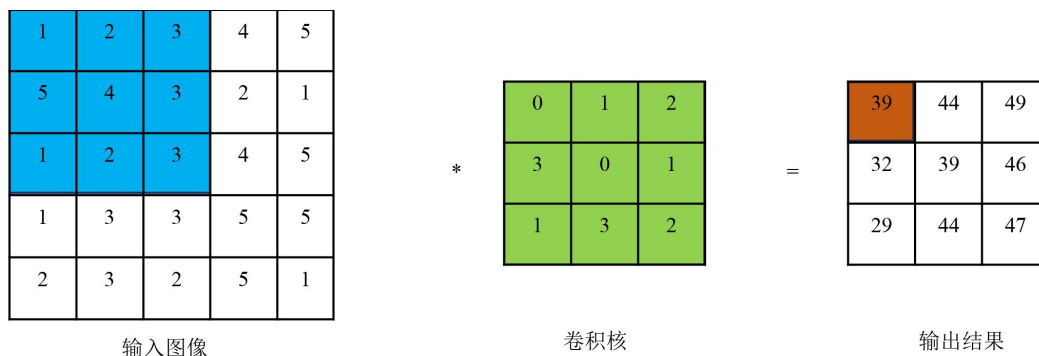


图 2-2 卷积操作示意图

输出结果尺寸大小的计算如式(2-1):

$$Output = \frac{(Input - Kernel + 2 \times Padding)}{Stride} + 1 \quad (2-1)$$

其中 $Input$ 表示输入图像大小， $Kernel$ 表示卷积核大小， $Padding$ 表示原矩阵进行边界填充，也就是在矩阵的边界上填充一些值，以增加矩阵的大小，通常都用“0”来进行填充的， $Stride$ 表示为步长，滤波器在输入数据上每次滑动的距离。

(2) 激活函数

在卷积神经网络中，卷积和池化是进行线性变换的图像操作，其输出与输入保持线性关系。为了增强网络处理非线性特征的能力，激活函数被引入。提升非线性特征的表达能力，有助于学习图像等数据的复杂特征，常用的激活函数包括 Sigmoid、ReLU、Leaky-ReLU 等。

(3) 池化层

池化层又称下采样，其作用到特征图上，可降低特征图的维度，减少参数量。也能改善过拟合的现象^[31]，是卷积神经网络中的重要组成。如图 2-3 所示为最大池化示意图，选取区域内的最大像素值作为输出，而平均池化则计算所选区域内像素的平均值。

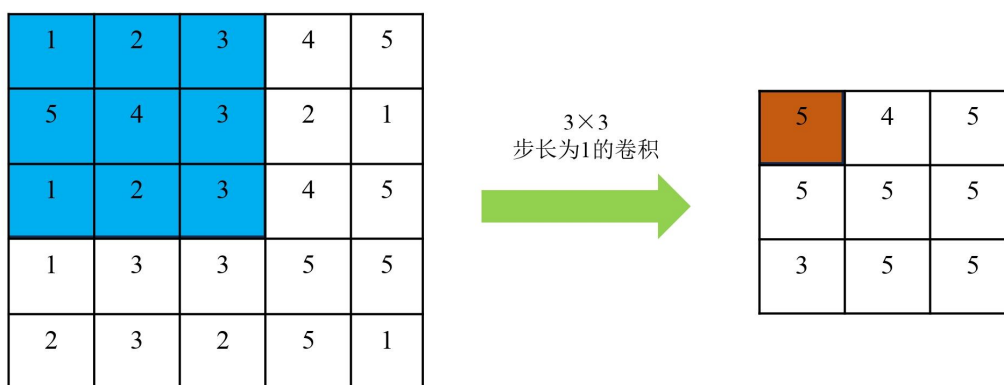


图 2-3 最大池化示意图

(4) 全连接层

全连接层是卷积神经网络的关键部分，通常在网络末端，通过处理局部特征图，从而具有获取全局特征信息的作用。在这一过程中，特征图被展平为一维向量，并传递给隐含层，经过隐含层处理后最终完成预测任务。全连接层的加入，会增加模型的参数量，训练难度也会提升，针对这个问题会用卷积层来代替全连接层。

2.2 基于深度学习的目标检测方法

在目标检测中，基于深度学习的目标检测方法分为一阶段和二阶段两种。在二阶段检测中，网络首先处理候选框，然后进行分类和回归。相比之下，一阶段具有较快的训练，由于其简化流程，可以直接提取特征^[32]。

2.2.1 二阶段目标检测算法

二阶段方法会先获取候选区域，再对得到的区域进行特征提取，最后实现目标检测框的回归。二阶段目标检测方法是分阶段完成的，提高识别的质量，两阶段网络结构如图 2-4 所示。

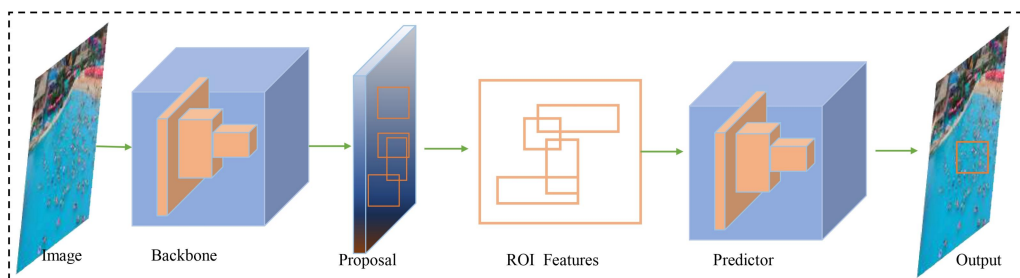


图 2-4 二阶段网络结构

Faster R-CNN^[33]属于二阶段中经典的方法，其提出区域建议网络(Region Proposal Network, RPN)及感兴趣的区域(Region Of Interest, ROI)Polling 层，在当时存在的方法中获得领先的检测效果。Faster R-CNN 会先通过 RPN 对目标进行粗分类，再利用 ROI Polling 实现精细识别，其结构如图 2-5 所示。之后许多二阶段目标检测算法都以 Faster R-CNN 为基础，从图像中目标的特点出发，进行改善。

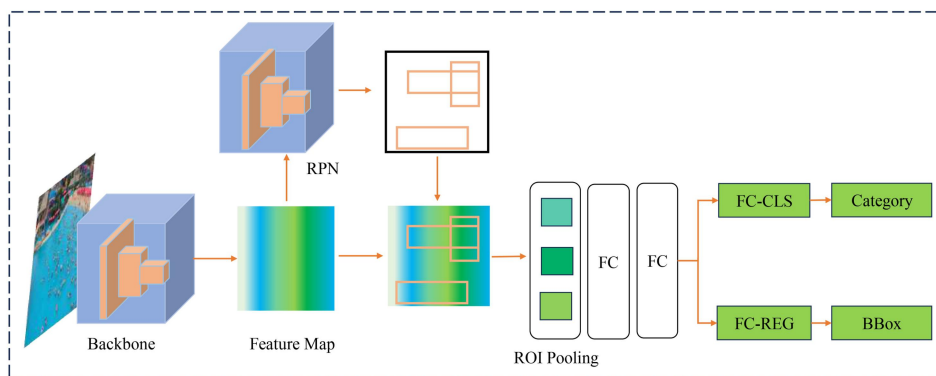


图 2-5 Faster R-CNN 网络结构^[33]

Mask R-CNN^[34]在 Faster R-CNN 基础上，用 ResNet^[35]替换了视觉几何组

(Visual Geometry Group, VGG), 并提出了 ROI Align, 提升了分割的准确率。Mask R-CNN 采用了特征金字塔网络^[36](Feature Pyramid Networks, FPN)的结构来进行特征的融合。在增加较少计算量的背景下, 融合了低层特征图的空间信息和高层特征图的语义信息, 增强特征的表达能力。并且针对 Faster R-CNN 在多尺度目标共存情况下, 导致检测效果偏差的问题, 利用输出各个层次的特征图用来检测多尺度的目标, FPN 结构如图 2-6 所示。

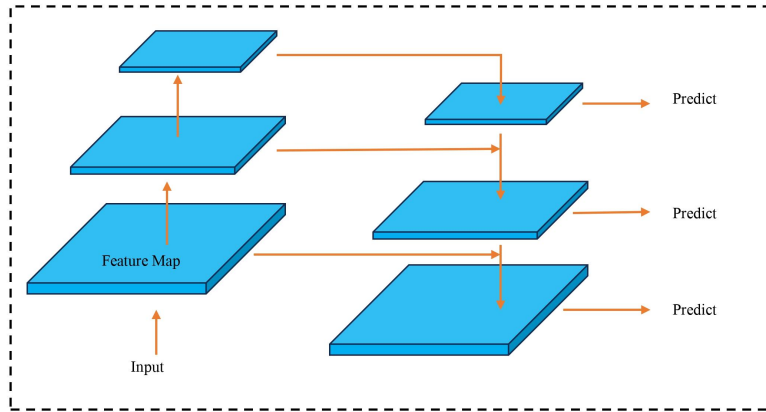


图 2-6 FPN 结构^[36]

Cascade R-CNN^[37]结构如图 2-7 所示, 在第二阶段中, 网络检测器的输入是上一个检测器的输出。相比于 Faster RCNN, Cascade RCNN 达到了更好的检测结果。

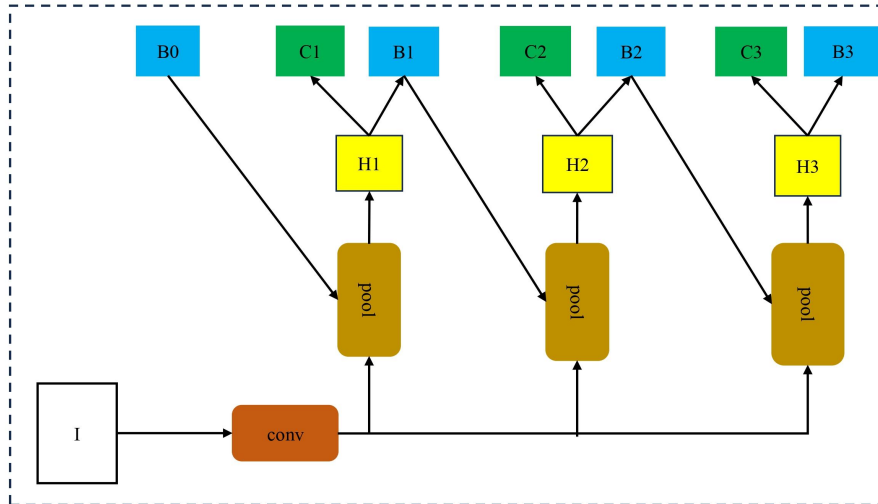


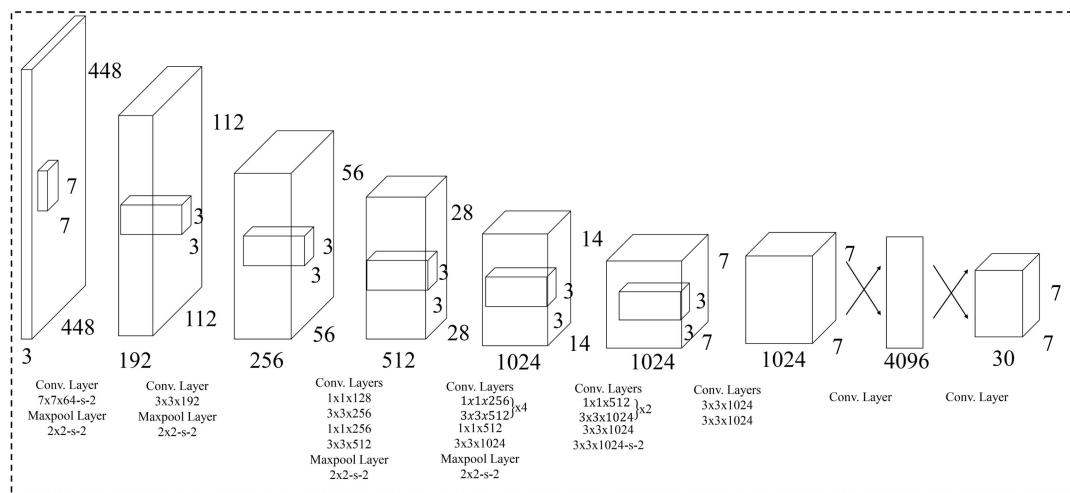
图 2-7 Cascade R-CNN 级联结构^[37]

2.2.2 一阶段目标检测算法

二阶段算法虽检测精度高, 但计算量大, 难以满足时效性要求。为提高检测

YOLO 作为一阶段目标检测算法，结构如图 2-8 所示，使用卷积层来提取图像特征，利用全连接层来对目标进行识别定位。YOLO 算法会把输入图像分成 $S*S$ 的网格，然后利用每个网格去负责预测物体，每个区域有 B 个预测框 (Bounding box) 及各自的置信度。在各个预测框里，要预测 C 个类别概率 $Pr(Class_i|Object)$ 。Bounding box 包括 x ， y ， w ， h 和 confidence 数据， (x,y) 指中心坐标， (w,h) 为宽和高，confidence 为目标识别的准确度，confidence 计算如式(2-2)所示：

其中, 若 bounding box 包含物体, 则 $Pr(Object)=1$, 否则 $Pr(Object)=0$ 。



Liu 等^[38]提出了 SSD, 其与 YOLO 相比, 主要有几个方面改进: 利用多尺度的特征图进行预测和预设先验框, SSD 基本框架如图 2-9 所示。

其中首个得分就是针对背景的。第二部分为检测框的中心坐标以及宽高 (cx, cy, w, h) 。 $d = (d^{cx}, d^{cy}, d^w, d^h)$ 代表预设锚定框， $b = (b^{cx}, b^{cy}, b^w, b^h)$ 代表回归锚定框，得到的 l 为 b 和 d 的偏差分别如式(2-3)，(2-4)所示：

$$l^{cx} = \frac{b^{cx} - d^{cx}}{d^w}, l^{cy} = \frac{b^{cy} - d^{cy}}{d^h} \quad (2-3)$$

$$l^w = \log\left(\frac{b^w}{d^w}\right), l^h = \log\left(\frac{b^h}{d^h}\right) \quad (2-4)$$

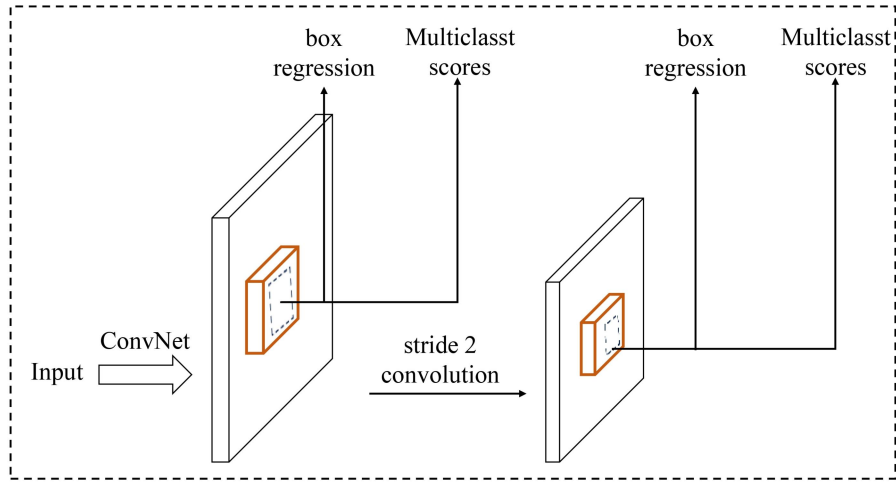


图 2-9 SSD 基本框架^[38]

SSD 和 YOLO 相较于二阶段方法在检测效果上仍有改进的空间，缺少从候选区域提取特征的步骤。Lin 等^[39]提出了 RetinaNet 目标检测模型，解决训练时负样本量远超正样本量，使得负样本控制训练方向的问题，该模型引入 Focal loss 的损失函数。对标准的交叉熵损失改进，以此来得到对正负样本以及难易样本的平衡。这在原始的交叉熵损失中是不容易实现的，二分类如式(2-5)所示：

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise} \end{cases} \quad (2-5)$$

$y \in \{\pm 1\}$ 表示类别归属， $y = 1$ 的得分由 $p \in [0, 1]$ 表示。 p_i 如式(2-6)所示：

$$p_i = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise} \end{cases} \quad (2-6)$$

交叉熵损失又能被表示如式(2-7)所示：

$$CE(p, y) = CE(p_i) = -\log(p_i) \quad (2-7)$$

这种损失的特点在于，即便是容易分类的目标，也会产生较大的损失。当损

失累积后，其总量可能超过难分类样本的损失。因此，模型训练时更倾向于关注容易分类的样本，而忽视难分类的样本。

对于产生的这个问题，通过设置权重系数 α ，具体如式(2-8)所示。使用这种方法可以解决正负样本不平衡的问题，但在难分类的样本还存在不足。

$$CE(p_i) = -\alpha_i \log(p_i) \quad (2-8)$$

因此引入 $(1-p_i)^\gamma$ ，其中 γ 为 Focusing Parameter，且 $\gamma \geq 0$ ，其式如式(2-9)所示：

$$FL(p_i) = -\alpha_i (1-p_i)^\gamma \log(p_i) \quad (2-9)$$

当 γ 值逐渐增大时，易分类样本的 loss 通常接近零，而难分类样本的 loss 值则相对较高。这种差异确保在类别不平衡的情况下，难例样本也能对总 loss 产生较大影响，使模型在训练时更加关注对难例样本的优化，RetinaNet 的结构如图 2-10 所示。

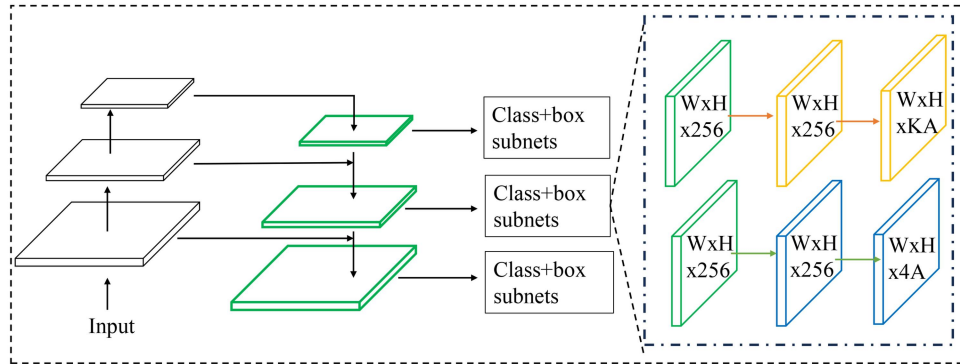


图 2-10 RetinaNet 网络结构^[39]

2.2.3 模型选择

经上面概述可知，基于深度学习的目标检测方法主要分为一阶段、二阶段两类，其各有优缺点。在航拍场景中，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，一阶段算法能够更好地满足这种实时性需求，在实际应用中更符合本文的要求。在一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小目标检测中。YOLO 模型具有灵活的参数配置，对于航拍图像中小目标的检测，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。

2.3 神经网络中的注意力机制

在自然语言处理任务中注意力机制首先被提出^[40],该机制使得网络更加关注于重要的特征信息。被引入到计算机视觉后,在目标检测、语义分割等领域都有所应用。注意力机制发展历程如图 2-11 所示,经历了空间注意力,通道注意力,混合注意力以及自注意力。



图 2-11 注意力机制发展历程

空间注意力可以筛选特征图中重要的空间区域,是一个自适应选取的过程。通过加强网络对重要空间特征的捕获,实现对不同空间区域的加权操作,使网络更专注于有效的特征信息。

2014 年循环注意模型(Recurrent Attention Model, RAM)^[41]模型被提出,这是首个应用空间注意力的模型,该网络采用循环神经网络结构。其关注图像的不同空间区域,整合区域信息,给网络赋予不同区域相应的权重。通过这种方式,网络能够多次提取局部特征,获取多样化的空间特征信息,其整体结构如图 2-12 所示。RAM 通过端到端的训练方式,更加专注于图像中不同位置的信息。

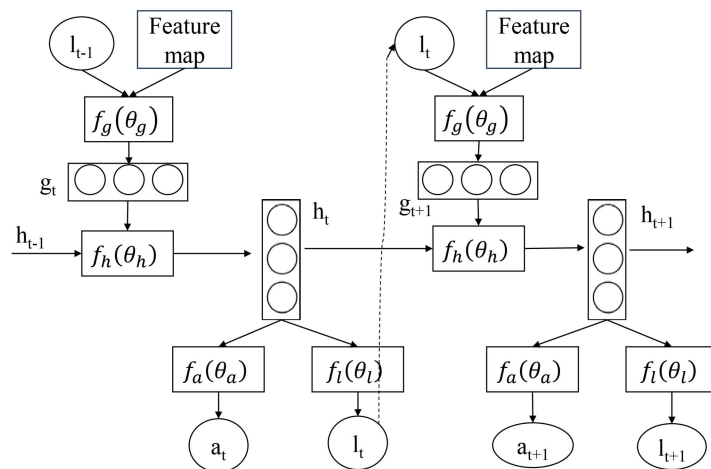


图 2-12 RAM 结构^[41]

通道注意力机制通过对局部和全局信息的获取，有助于提取特定类别的特征。通过映射特征中不同通道的权重进行学习，衡量特征的重要性，其中权重大小反映了特征相关性的高低。这种机制能够优化通道特征的关联性，通过权重信息对不同通道特征的处理，实现重要特征的强化，帮助网络更好地捕捉不同通道的特征，提升模型在处理复杂数据时的表达能力。

挤压和激励(Squeeze-and-Excitation, SE)^[42]是典型的通道注意力机制代表，其可以突出特征图中重要通道特征的权重，建立相互间的依赖关系。通过学习通道间特征的权重，增强全局信息中有效特征。并利用其获取的重要权重信息对原始特征信息进行动态调整，从而提高整个网络的准确性，SE 结构如图 2-13 所示。

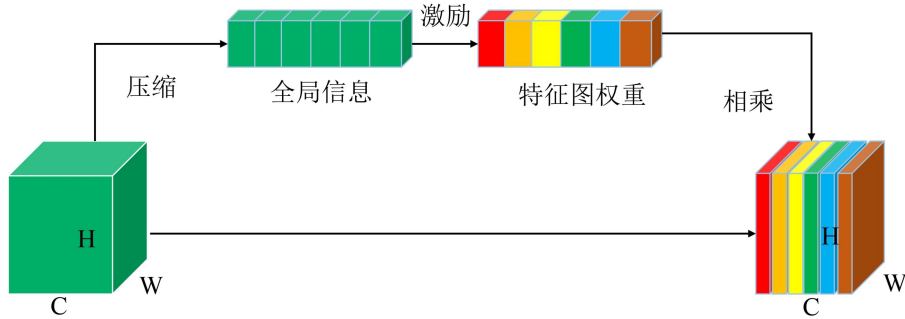


图 2-13 SE 结构^[42]

通道与空间混合注意力机制结合了通道与空间维度的信息优势，自适应地提取重要权重，并加权求和生成混合特征。该机制利用特征在通道与空间维度上的交错关系，获得更综合的权重信息。

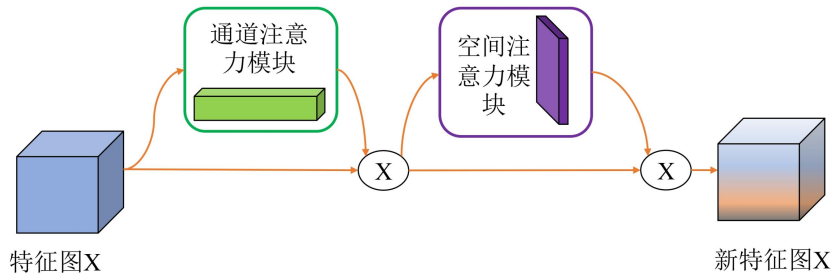


图 2-14 CBAM 结构^[43]

卷积块注意力模块(Convolutional Block Attention Module, CBAM)^[43]是一种典型的通道与空间混合注意力机制，首先经过通道注意力机制处理通道信息的权重，突出图像中重要通道特征。然后再利用空间注意力机制对空间信息进行筛选，强化图像中有效的空间区域信息。自适应地处理特征的细化，通过通道和空间特征信息的映射以完善特征图，CBAM 整体结构如图 2-14 所示。

自注意力机制是注意力机制的一种特例，灵感来自生物视觉系统的特性。利

用对输入信息进行特殊的计算，从而实现对不同元素分配不同权重的能力，便于捕捉序列内部的联系。自注意力机制通过学习查询向量 Q 、键向量 K 、值向量 V 之间的权重映射，获取全局特征关联性，实现全面关注。每个输入单元都与所有单元进行计算，自适应学习关键信息，自动捕获不同单元间的关联性，具有强大的表达能力和自适应性。

视觉变形器(Vision Transformer, ViT)^[44]是一种多头自注意力机制，于 2020 年被提出，在计算机视觉任务中取得很好的优势，多头自注意力机制整体结构如图 2-15 所示。ViT 中的多头自注意力机制作用于图像，提高了对像素的关注，在 ImageNet 数据集上取得优异成绩，打破了 CNN 在视觉领域的领先地位。通过独立学习不同的特征，以达到相互补充，提升模型表达的作用。关注不同位置和通道的信息，减少信息丢失。

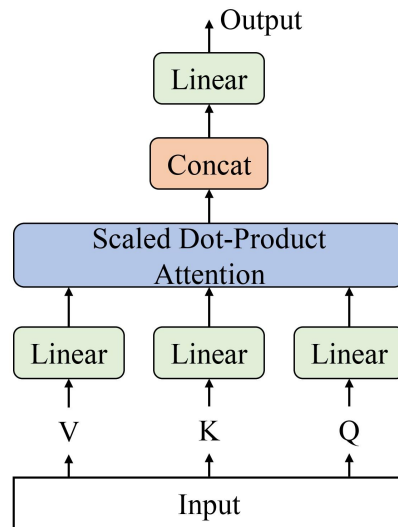


图 2-15 多头注意力机制结构^[44]

注意力机制对输入信息进行加权处理，使模型聚焦于图像中的关键区域，提高视觉任务精度。因此，本文引入并设计多种注意力机制，以提升网络检测精度。

2.4 航拍图像小目标定义及技术难点

2.4.1 小目标的定义

在目标检测领域，小目标的定义尚无定论，主要有相对尺度和绝对尺度两种观点。

相对尺度小目标以图像为参照，通过目标边框大小与图片面积之比或目标边

框与图像宽高比例来界定。然而，这种定义易受数据集影响，模型评估结果可能因数据集图像大小不同而有所偏差。

绝对尺度小目标通常依据物体像素大小来界定，例如在 City-Persons 数据集中，小目标的定义为高度小于 75 像素的目标，不同数据集如 DOTA、WIDERFACE 和 Tiny-Person 也各自设定了不同的像素大小作为小目标的标准。

总结下来这些小目标通常像素较少，信息特征有限，导致检测进展相对缓慢。在航拍图像小目标检测中，还存在以下两个技术难点需要克服。

2.4.2 多尺度目标共存及目标密集分布的问题

由于航拍图像拍摄距离较远，图像幅面较大，单个目标在图像中所占像素较少。从图 2-16 可以看出，航拍图像中存在许多小目标，这些目标占据像素较少。在现有大部分目标检测网络中，其网络层数都比较多，随着层数的不断加深，网络在提取特征信息时，特征图的分辨率也会逐渐减小。这会导致小目标本来特征信息就偏少的情况更加突出，获取到的特征信息很少甚至消失。并且航拍图中夹杂着一些大目标，最终使图像中出现大小不一的多尺度目标共存现象，进而影响网络对不同尺度目标的精准识别定位。同时在航拍图像中，存在的目标数量偏多，会出现目标紧密排列在一起的情景，它们的特征信息会相互干扰和重叠，使得模型难以准确区分和提取每个目标的特征。这也进一步影响通用模型对小目标的特征信息的提取。

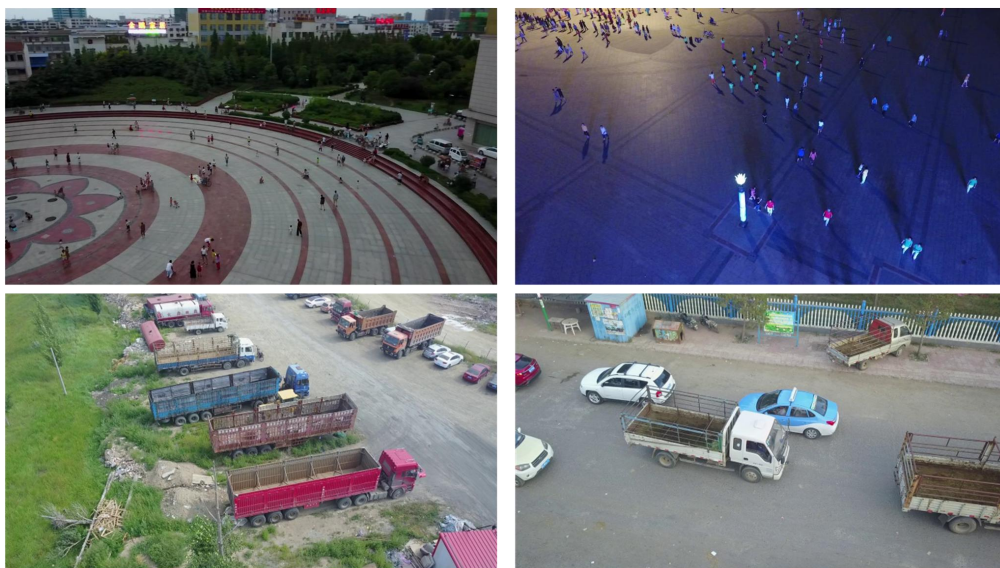


图 2-16 航拍图像展示图 1

2.4.3 目标方向任意及复杂背景的问题

由于航拍场景的复杂多样，周围的背景容易与目标造成混淆。此外，在现实生活中还存在目标遮挡、目标的不同方向排布等情况。如图 2-17 所示，航拍图像中的物体往往随机分布，倾斜的物体更易受背景干扰。水平预测框作用于倾斜的物体可能包含过多的背景信息以及临近目标的信息，干扰模型训练，导致误检和漏检，这都增加了网络对目标特征提取的难度。



图 2-17 航拍图像展示图 2

针对以上两个难点，本文分别在第 3 章和第 4 章提出改进方法对其进行优化。提高航拍图像小目标检测效果，为后续的工作提供参考价值。

2.5 本章小结

本章对卷积神经网络的概念和一些理论基础做出论述，并引出二阶段检测和一阶段检测算法，以及各阶段检测算法对应的方法做较为详细的分析，并对注意力机制的发展做出概述，最后对航拍图像小目标的定义及技术难点进行了分析。

第3章 基于特征增强的航拍图像目标检测方法

目前已有的许多通用算法在通用数据集基准测试上表现出色。然而，在处理特殊场景，如航拍图像时，现有大部分算法在特征提取阶段未能充分考虑到一些独特挑战：如目标偏小，多尺度目标共存，密集分布等，而此类难点在航拍图像中尤其突出。因此，针对上述航拍图像中目标的特点，研究有效改善航拍目标特征提取的方法，提高当前通用算法在航拍图像上目标检测精度。

本章在 YOLOv5 算法的基础上，基于上述问题，提出一种基于特征增强的航拍图像目标检测方法。改进主干特征提取网络，以增强在复杂背景下的目标特征表达，尤其是对小目标的特征进行强化。结合多尺度检测的改进方法，实现航拍图像多尺度目标的高精度检测。

3.1 模型整体结构

YOLOv5 的结构如图 3-1 所示。首先，提取输入图像的特征信息，随后融合提取的特征信息，最终通过输出三个检测端进行目标预测。YOLOv5 有五个模式，有 x, s, n, m, l 五种，它们的不同表现在网络模型的深度和宽度上。

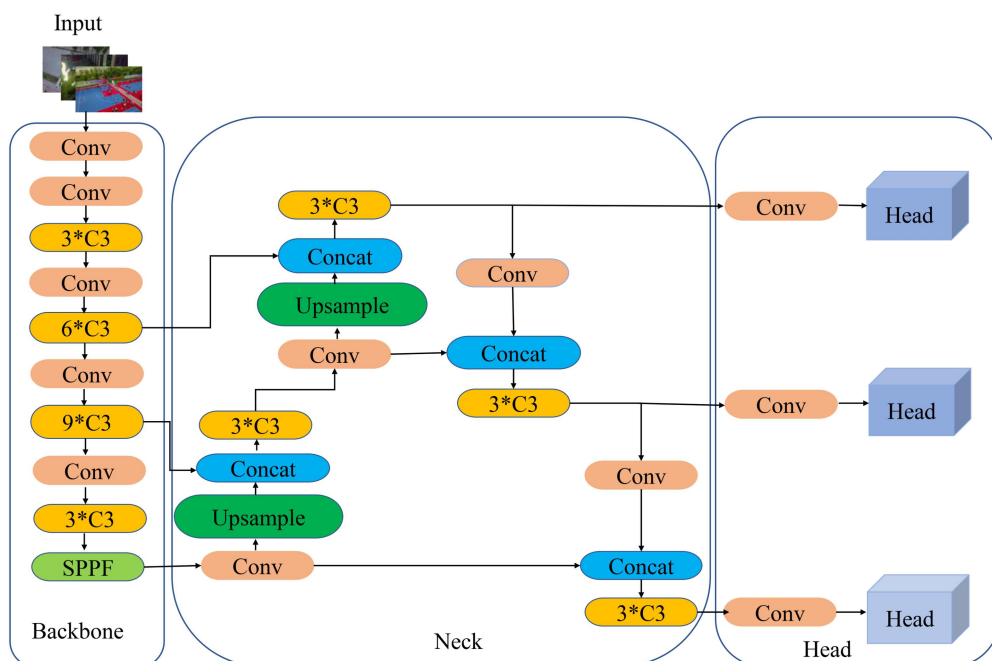


图 3-1 YOLOv5 网络结构

本章所提出的模型以 YOLOv5s 为基础框架，着眼于在航拍图像目标检测中

平衡检测速度与精度的需求。YOLOv5 的网络结构包括 Input、Backbone、Neck 和 Head 四个部分，在 Input 部分，利用 Mosaic 方法对输入图像进行数据增强，Backbone 层提取图像的特征信息，Neck 层对提取的特征进行融合，Head 层从 Neck 层提取的 80×80 、 40×40 和 20×20 三个尺度进行预测(输入图像尺寸 640×640)，最终输出检测结果。

本章提出的一种基于特征增强的航拍图像目标检测方法的结构图如图 3-2 所示。

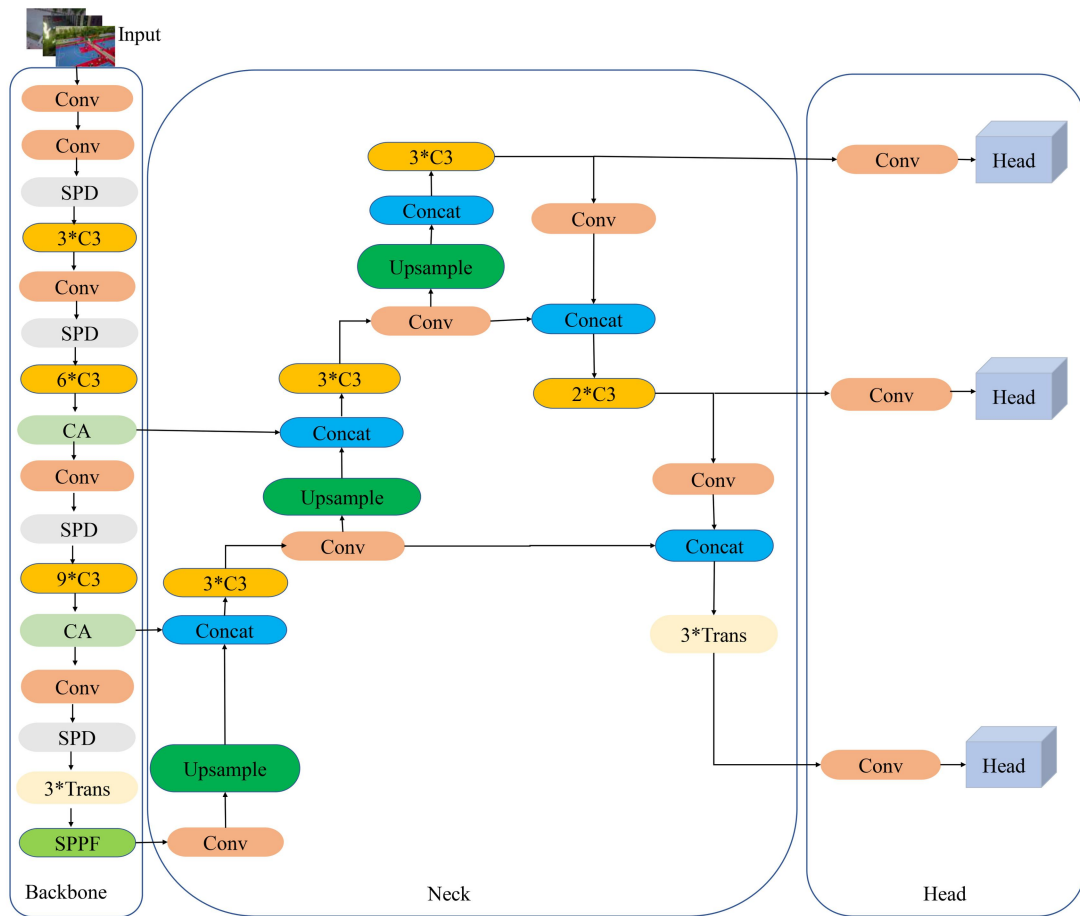


图 3-2 基于特征增强的航拍图像目标检测模型结构

在 YOLOv5 的基础上，针对航拍图像特点，本章做如下改进：

(1) 图像信息获取角度。采用多种注意力模型，引入 Transformer 模块和 CA 模块，利用注意力机制来学习该区域的特征信息，从而获取图像中感兴趣区域，使网络能更好地获取特征信息。

(2) 特征信息提取角度。在主干特征网络引入 SPD-Conv 模块，克服卷积网络中使用跨步卷积和池化层，导致图像细节信息丢失以及特征学习较低效率表示

的问题，提高对航拍图像小目标细节信息的提取。

(3) 多尺度检测角度。在网络 Head 端添加小目标检测端，移除大目标检测端，提升模型对航拍图像多尺度目标的检测准确率。在检测端中引入 Transformer 模块，利用自注意力机制，有效获得全局信息。多头注意力具有将其映射多个空间的特性，可以解决网络末端图像分辨率较低的问题。

3.2 主干网络特征增强方法

3.2.1 Coordinate Attention 增强特征信息获取

注意力机制被广泛用于各种检测任务，如 SE、CBAM 等。CA^[45]将位置信息加入模块，利用简洁高效的方法，使网络模型能够获取更广泛区域的特征信息，同时也不增加额外的计算负担。

网络模型在获取航拍图像特征信息过程中，随着网络的加深会造成信息丢失，从而忽略小目标、密集目标的信息。因此本章引入 CA 模块，减少图像信息的丢失。位置和通道信息的同时作用，可以实现显著信息的获取，帮助网络更好地识别目标。

CA 注意力机制主要分坐标信息嵌入与生成，信息嵌入相关公式如式(3-1)，(3-2)所示：

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

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

其中， $Z_c(h)$ 表示为第 c 个通道的输出，而 h 为高度。 H ， W 分别为对应输入特征图的高度和宽度。 $x_c(h, i)$ 为第 c 通道对应位置特征值。通过上述两种方式嵌入，聚合两个空间方向特征，形成一对方向感知的特征图。网络获取一个空间方向关系时，也可以保存另一个空间方向的位置信息。生成的相关公式如式(3-3)所示：

$$f = \delta(F_l([Z^h, Z^w])) \quad (3-3)$$

$F1$ 为卷积变换函数, $[Zh, Zw]$ 为特征层沿空间维度的堆叠操作, δ 为 sigmoid 函数, f 为水平或垂直方向对空间信息进行编码的中间特征图。利用两个 1×1 卷积变换函数将 f^h 和 f^w 变换为具有相同通道数的张量, 分别得到了式(3-4), (3-5):

$$g^h = \sigma(F_h(f^h)) \quad (3-4)$$

$$g^w = \sigma(F_w(f^w)) \quad (3-5)$$

g 为不同空间维度拆分特征图的注意力输出结果, F 为卷积计算函数。最终输出如式(3-6)所示, CA 模块结构如图 3-3 所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (3-6)$$

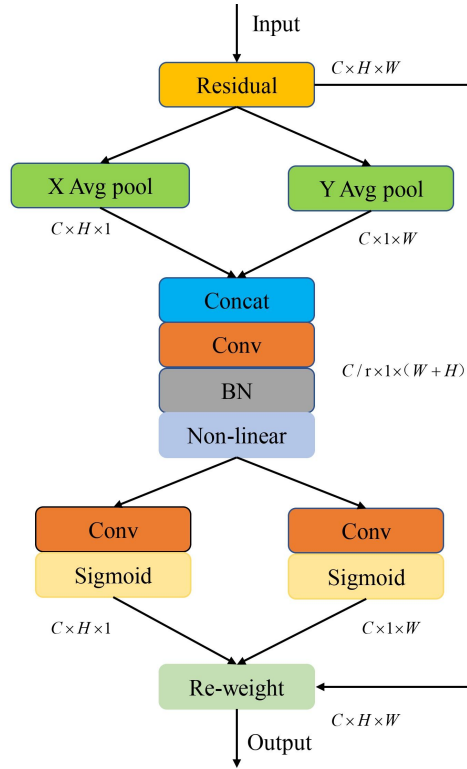


图 3-3 Coordinate Attention 模块结构^[45]

3.2.2 Transformer 强化全局信息提取

Transformer^[46]模块目前在视觉方面变得热门, ViT 是第一个在视觉识别任务中引入 Transformer 的模块, 并取得很好的检测效果, 本文将主干提取网络中最后一个 C3 模块替换为 Transformer 与 C3 结合的模块。改进模块与原始网络的

C3 模块相比，提高了全局信息获取能力，丰富了上下文特征信息，在航拍高密度分布的目标检测任务上具有更好的效果。Transformer 模块包含两个部分，第一部分为多头注意力模块，第二部分为全连接层。计算如式(3-7)所示：

$$Attention(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d_K}} \right) V \quad (3-7)$$

在上述公式中， Q 、 K 、 V 分别为查询向量、键向量、值向量， d_k 为输入特征 K 向量的维度。这样使输入的图像的每一个像素可以与其他像素产生关联性，融合了其他像素的特征。最后，把得到的不同部分结果进行融合，即可得到多头注意力的输出，如式(3-8)所示：

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h) W^O \quad (3-8)$$

其中 $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$ ， $W_i^Q \in \mathbb{R}^{d_{model} \times d_q}$ ， $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$ ， $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ ， $W^O \in \mathbb{R}^{d_{model} \times d_{hd_v}}$ 为参数矩阵， d_{model} 为向量维度。

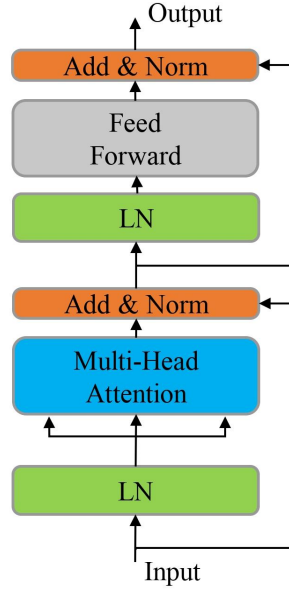
Transformer 模块中的多层感知器，主要是对输入信息进行映射变化，进而实现对输入信息的特征提取及筛选，输出表示如式(3-9)所示。

$$f_{FFN} = MLP(LN(MultiHead(Q, K, V) + \bar{z})) \quad (3-9)$$

其中， $\bar{z}$ 作为输入， $LN(MultiHead(Q, K, V) + \bar{z})$ 表示包含了两层全连接层的多层感知器，Transformer 编码器结构如图 3-4 所示，最终 Transformer 模块的输出表示如式(3-10)所示：

$$Transformer(x_{in}) = f_{FFN} + MultiHead(Q, K, V) \quad (3-10)$$

多头自注意力机制增加了特征子空间的多样性，提升不同位置捕获信息的能力。通过同时处理多头注意力，模型可以更全面地捕捉输入序列中的关联性和上下文信息，从而使得编码器更为灵活和有效地学习序列中的特征，这种并行的多头机制有助于提高模型的性能和表达能力。Self-Attention 机制作为 Transformer 中的重点模块，其能够使图像中的每个像素可以关联到其他像素的特征。利用获取到的各个像素之间的关联性权重，可以建立不同目标特征的联系，从全局中获取更多有效特征。引入 Transformer 模块可以增强主干网络对航拍小目标的特征提取能力。

图 3-4 Transformer 编码器结构^[46]

3.2.3 SPD-Conv 优化特征提取方法

卷积网络用于常见场景目标检测时，因为一般图像的分辨率良好，图像中目标大小适中，所以跨步卷积和池化跳过冗余信息，模型仍然可以提取足够的特征信息。然而，面对航拍图像小目标检测的任务，冗余信息变得相对较少，跨步卷积和池化层的作用，就会导致小目标的信息丢失以及特征表示的学习效率低。为此，本文引入 SPD-Conv 模块^[47]，以代替跨步卷积层和池化层。

对于给定任何特征图 X ，每个子图可以按比例因子对 X 进行下采样。图 3-5 给出了 $\text{scale}=2$ 时的示例图，SPD-Conv 模块将特征图转换为中间特征图

$X' \left(\frac{S}{\text{scale}}, \frac{S}{\text{scale}}, \text{scale}^2 C_1 \right)$ 。在 SPD-Conv 特征变换层之后，再添加了一个非跨步

卷积层 (即 $\text{stride}=1$)，其中 C2 模型的过滤器，通过这样进一步变换，

$X' \left(\frac{S}{\text{scale}}, \frac{S}{\text{scale}}, \text{scale}^2 C_1 \right) \rightarrow X'' \left(\frac{S}{\text{scale}}, \frac{S}{\text{scale}}, C_2 \right)$ ，尽可能地保留图像所有的特

征信息。

使用卷积步长和池化层，这将导致细粒度特征信息的丢失，特别是在低分辨率图像和小目标上。使用 SPD-Conv 消除了步长和池化操作，取而代之的是一个空间到深度卷积和一个无步长卷积。这种新设计在保留判别特征信息的同时，具

有向下采样特征图的优点。

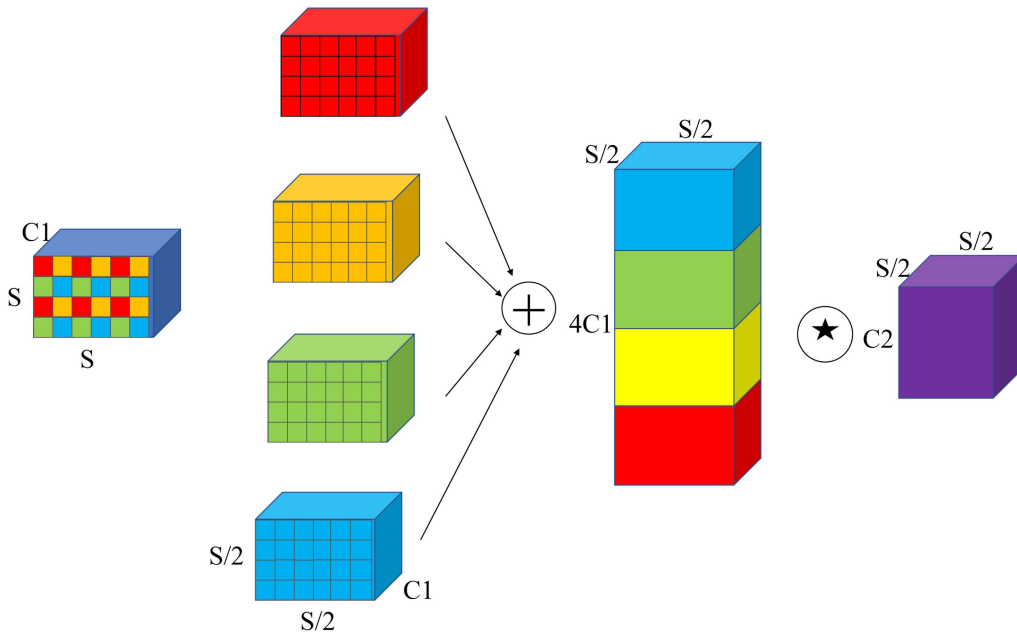


图 3-5 SPD-Conv 结构示例图(scale=2)^[47]

3.3 多尺度检测策略

YOLOv5 算法有 3 个不同尺度的检测层。输入图像经过 Backbone 层的处理后, Neck 层会先进行上采样,然后再分别与 Backbone 层中的第 4 层和第 6 层融合,之后再进行下采样。Head 层会提取 Neck 层三个尺寸特征图进行输出,用以检测图像中的小、中、大目标。但是航拍图像中小目标居多,这导致 YOLOv5 模型对航拍图像小目标的检测能力结果偏差。

本文通过改进 Head 结构,增加一个检测头,输出大小为 160×160 的检测特征图(输入图像大小为 640×640),提高对小目标的敏感度,提升对航拍图像中小目标的检测能力。由于增加检测头会提高计算成本,并考虑到航拍图像中也会存在其他非小目标物体,因此本文在输出四个检测头的基础上移除大目标检测头,保留检测中,小,微小的检测头,减少不必要的计算量,保留网络具有检测速度优势,应对航拍图像中的多尺度目标检测任务的挑战。在处理航拍图像这类大规模、高复杂度的数据时,网络末端的特征图分辨率往往较低。因而本文只在低分辨率特征图上引用 Transformer 模块,进一步挖掘语义信息之间的关联,从而提高小目标的特征表达能力。

3.4 检测图像对比实验

将本章网络的检测结果图与基准检测方法得到的检测结果图进行比较,如图 3-6 所示。其中图 3-6 中(a)、(c)图片为 YOLOv5 算法的可视化检测结果, (b)、(d) 图片为本文提出方法的结果图。观察图 3-6 的(a)和(b)两个图,可以发现 YOLOv5 出现了误检,而本章方法没有。图 3-6 的(c)和(d)图是密集检测环境,从中可以看出本章算法可以检测出 YOLOv5 漏检的 people。上述分析说明改进方法在多尺度目标场景下鲁棒性较强,对于目标密集分布下也能有较好的检测精度。

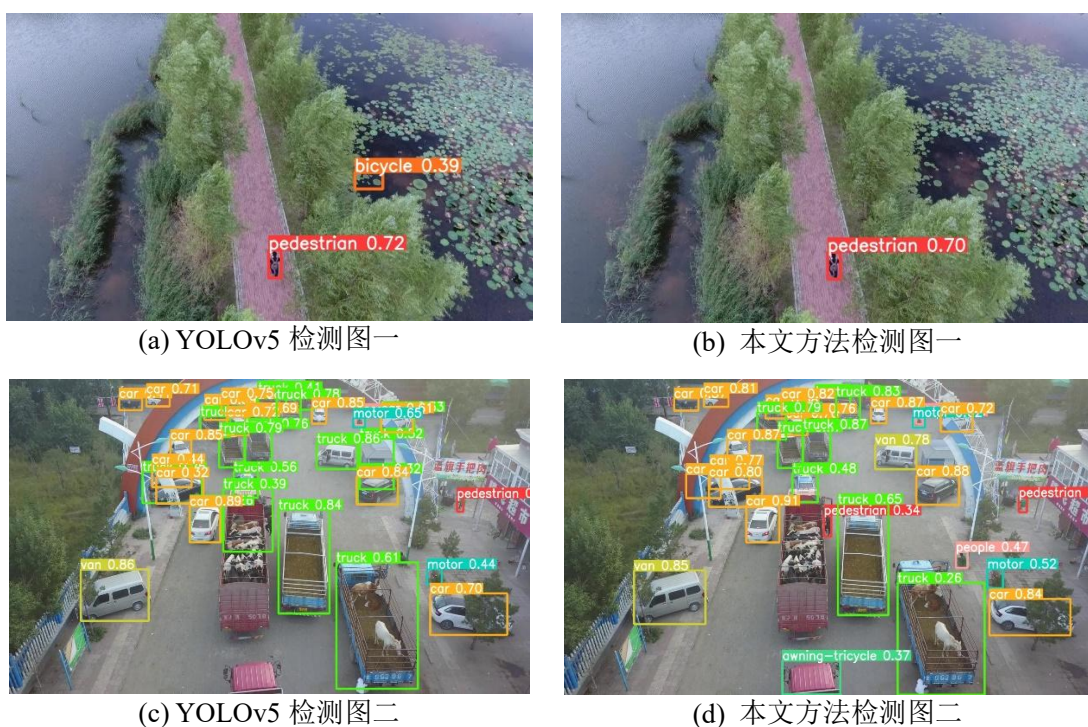


图 3-6 本文方法与 YOLOv5 的可视化结果对比

3.5 本章小结

本章针对现有的通用算法并没有过多关注航拍图像目标检测中的目标尺度差别大以及存在大量小目标密集排列等情况,在 YOLOv5 的基础上,提出了一种基于特征增强的航拍图像目标检测方法。在主干特征提取网络引入 CA 模块、Transformer 模块以及 SPD-Conv 模块来增强航拍图像特征信息提取的能力,增强复杂背景中密集小目标的特征。改进网络检测端,通过输出新的三个检测端为多尺度航拍图像中目标的高精度检测提供强大的特征表达。

第 4 章 融入注意力的航拍图像旋转目标检测方法

第 3 章所提方法主要针对多尺度分布且密集的目标，但在实验过程中发现，一些方向任意分布的目标，常用的基于水平框(Horizontal Bounding Box, HBB)方法会提取到过多的背景噪声和相近目标的特征信息，干扰模型的训练，导致检测效果差。同时基于水平检测框的方法无法准确估计目标的方向，而在航拍图像目标检测中，准确的方向信息是至关重要的，旋转目标检测方式具有更准确的目标方向估计的优点。因此，引入航拍图像旋转目标检测方法，结合注意力机制以减少背景信息的干扰，以此来准确地检测目标。

本章在 YOLOv7 算法的基础上，考虑到上述情况，提出了一种融入注意力的航拍图像旋转目标检测方法。通过引入注意力机制和旋转目标检测方法，减少背景信息的干扰，提高对方向任意目标的检测精度。

4.1 网络整体结构

YOLOv7 的整体结构如图 4-1 所示。

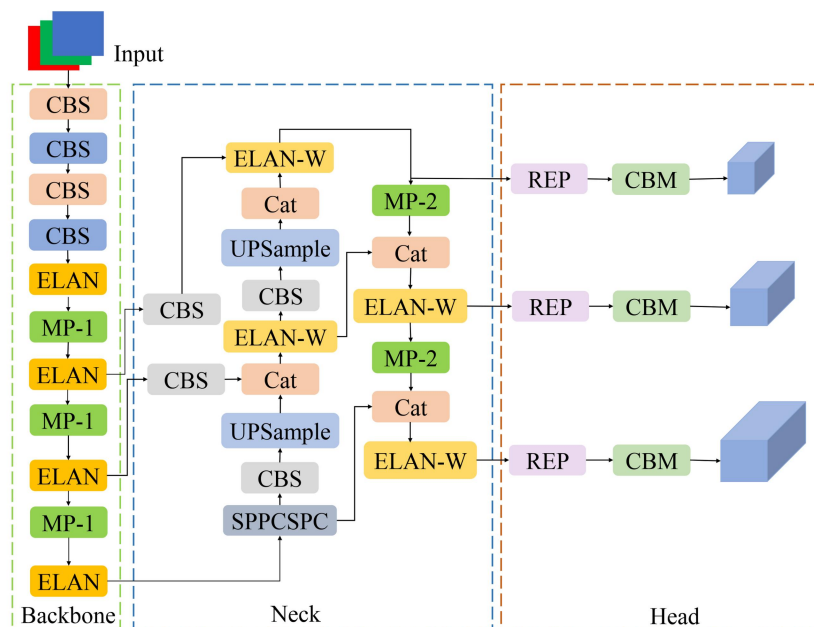


图 4-1 YOLOv7 网络结构

其结构与 YOLOv5 相同，包括 Input、Backbone、Neck 和 Head 四个部分。整体而言，YOLOv7 通过这些组成部分的协同作用，实现了对输入图像中目标的

准确检测。

针对航拍图像小目标检测存在的难点，本章提出一种改进的 YOLOv7 目标检测方法，整体结构如图 4-2 所示。设计增强注意力机制，将其引入 Backbone 中，增强网络的特征提取能力。提出一种双线性注意力机制，同时利用其在 Neck 构建平行融合机制，帮助网络更好融合航拍图像目标的特征信息。引入一种损失函数，使旋转框可以更好地匹配目标的方向。

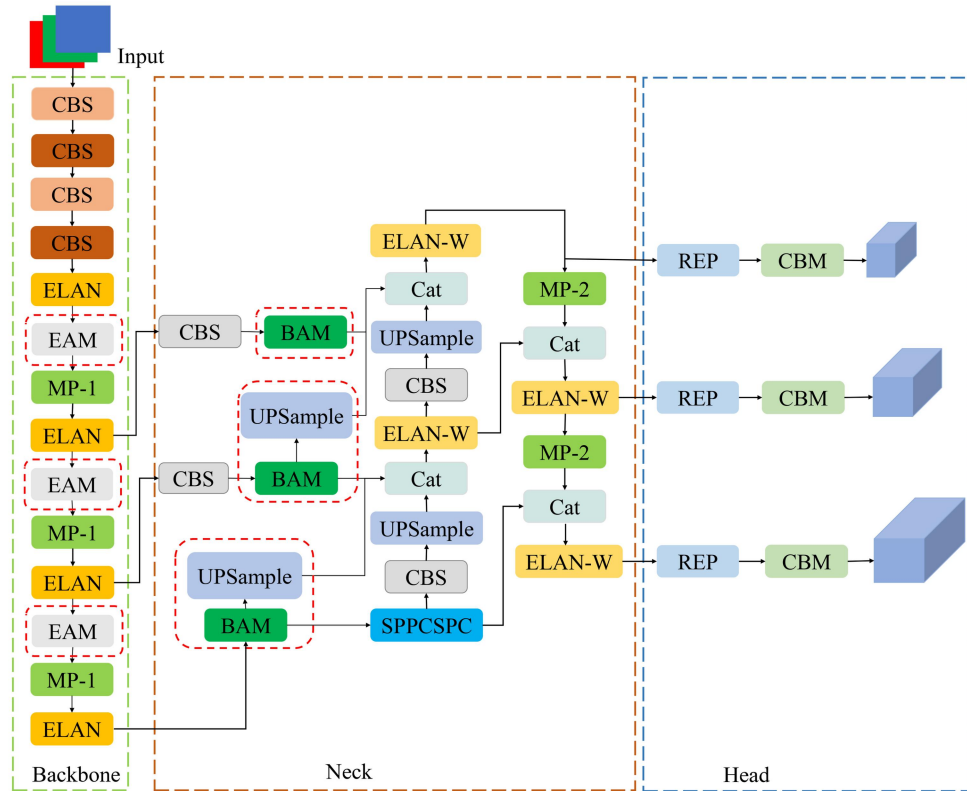


图 4-2 融入注意力的航拍图像旋转目标检测模型结构

在 YOLOv7 的基础上，针对航拍图像特点，本章做如下改进：

(1) 背景信息干扰方面。设计增强注意力机制，利用注意力机制来学习该区域的特征信息，从而获取图像中感兴趣区域。使网络在有限的计算能力下更好地获取特征信息，减少图像信息的丢失以及背景信息的干扰。

(2) 特征信息融合方面。设计双线性注意力平行融合机制，利用双线性注意力获取更多主干的特征信息，并通过上采样方式进行融合，提高特征表达能力。

(3) 旋转角度检测方面。采用旋转目标框进行目标检测，增加检测目标的角度信息。通过 KLD^[48]作为损失值来训练网络，优化旋转检测效果，进而提升网络在航拍图像旋转目标的检测精度。

4.2 融入注意力机制的目标检测方法

4.2.1 增强注意力机制强化特征信息提取

网络通过注意力机制获取到的可学习权重，使模型有效聚焦重要区域，降低噪声对航拍图像小目标特征提取的影响。针对航拍图像小目标存在的问题，本章设计一个增强注意力机制，利用特征之间的空间关系生成空间注意力图，以获得丰富的上下文特征。接下来，本章将分析聚合上下文特征的过程，如式(4-1)所示。

$$F = F_A^{Avg} + F_A^{Max} \quad (4-1)$$

其中 $F_A^{Avg} = AvgPool(F_{local})$ ， $F_A^{Max} = MaxPool(F_{local})$ ， F_{local} 为给定特征，本章使用 Maxpool 和 Avgpool 来压缩输入特征图的空间维度信息并集中处理。Maxpool 将全通道中最大值的空间像素点提取到输出空间维度中，突出检测图中一些中小目标的存在，Avgpool 计算全通道的平均空间像素输出到空间维度，突出显示所有空间维度中平均像素较高的点，过滤背景噪点。更值得一提的是，Maxpool 模块和 Avgpool 模块没有引入其他参数，因此不会给网络训练带来负担。使用平均池化和最大池化的目的是为在平均和最大池化之间找到最佳池化，收集更全面的空间特征信息。

从复杂的特征图中人为地提取通道注意力较为困难，同时提取操作也受网络复杂性的影响。为了解决这两个问题，本文引入深度可分离卷积(Depthwise Convolution, DWConv)，使用 DWConv 有 2 个优点：

(1) 可以通过深度可分离卷积直接从特征图中的所有空间信息中提取通道注意力。

(2) 深度卷积显著降低了计算成本，因此本文使用 DWConv，让网络可以降低输入特征图的维度，保持特征信息的不变性，以响应特征图中的所有空间信息，同时尽可能降低计算成本。

本文进行三次深度可分离卷积，使模型看到的感受野变大，提取的特征更丰富。将三次深度可分离卷积的结果进行融合，生成不同的特征，具体如式(4-2)，(4-3)所示。

$$M_{conv} = \sigma(DWConv(F) + DWConv^2(F) + DWConv^3(F)) \quad (4-2)$$

$$M_s = M_{conv} \odot F_{local} \quad (4-3)$$

其中, $DWConv(F)$ 表示对特征图进行深度可分离卷积操作, cat 表示三个矩阵相加, $\odot$ 表示哈达玛德乘积, M_s 表示最终的输出。注意力权值与特征图做张量积, 得到加权特征图。在复杂场景中, 该模块更侧重分配重要特征的通道权重, 使神经网络资源更多关注前景, 提升目标检测性能, 增强注意力机制如图 4-3 所示。

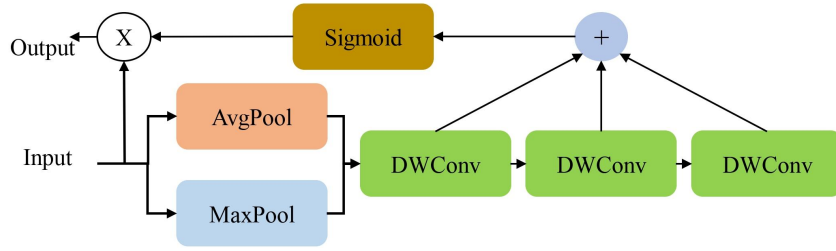


图 4-3 增强注意力结构

4.2.2 双线性注意力平行融合机制促进信息融合

文献[49]介绍一种有效的通道注意力(Efficient Channel Attention, ECA)模块, 其主要是对 SENet 模块改进。因该模块只用到了通道注意力, 对空间域信息特征的利用仍不够全面。因此, 本节以 ECA 模块为基础, 提出一种双线性注意力机制。学习目标空间结构信息与通道信息, 引导神经网络更有效地提取目标特征。提高网络对航拍图像目标的特征表达能力, 实现目标信息的有效提取, 双线性注意力机制结构如图 4-4 所示。

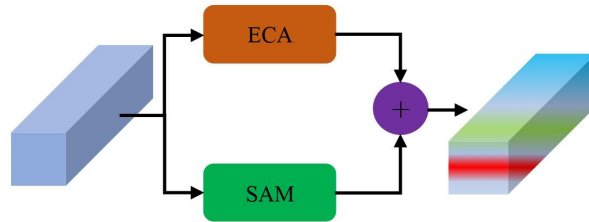


图 4-4 双线性注意力机制结构

首先, 对输入图像进行全局池化处理, 然后通过自适应卷积核来分配通道比重, 再利用 sigmoid 函数获取权重, 最后将输入特征与权重相乘, 加强重要区域的关注, 过程如式(4-4)所示:

$$ECA(x) = \sigma(\text{Conk}(\text{Gap}(F(x)))) \times F(x) \quad (4-4)$$

其中, $Conk()$ 表示大小为 k 的卷积函数, $Gap()$ 表示全局池化, $F(x)$ 表示当前时刻的输入。采用大小为 k 的一维卷积替换全连接层, 降低了全连接层复杂度对模型的影响。特征图通道数 C 如式(4-5)所示, 局部交互 k 值的计算如式(4-6)所示。

$$C = 2^{(r \times k - b)} \quad (4-5)$$

$$k = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (4-6)$$

其中 $|x|_{odd}$ 表示 x 最接近的奇数。在本节实验中分别将 γ 和 b 设为 2 和 1, ECA 模块的结构图如图 4-5 所示。通过池化操作对通道进行压缩, 利用 Maxpool 提取通道的最大值, 利用 Avgpool 提取通道的平均值。将提取到的通道特征图合并, 如式(4-7)所示, 图 4-6 展示了空间注意力模块的结构图。

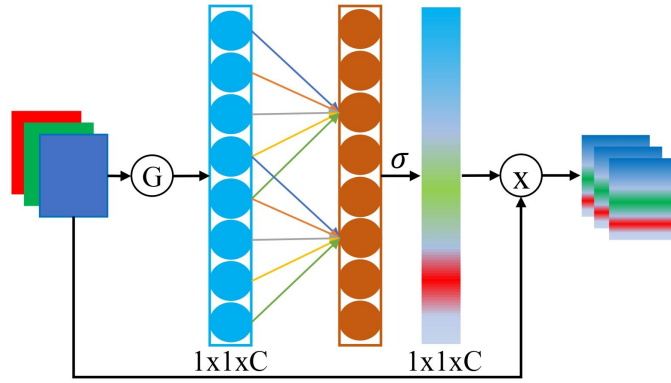


图 4-5 ECA 模块结构

$$SAM(F) = \sigma(f^{7 \times 7}([AvgPool(F(x)); MaxPool(F(x))])) \times F(x) \quad (4-7)$$

其中, $f^{7 \times 7}$ 表示卷积核的大小, 7×7 的卷积核比 3×3 的卷积核的效果更好。将 ECA 注意力和空间注意力的结果进行融合, 如式(4-8)所示。

$$A(x) = ECA(x) + SAM(x) \quad (4-8)$$

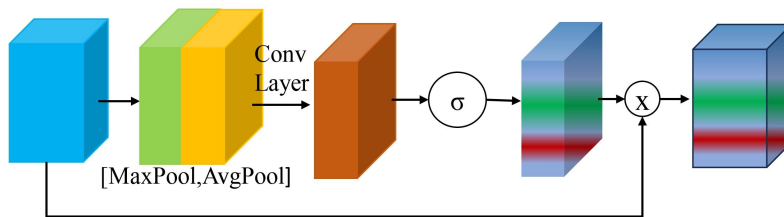


图 4-6 空间注意力模块结构

FPN 通过自顶向下的途径融合高层和低层语义特征,提供更准确的目标识别能力。路径聚合网络(Path Aggregation Network, PANet)增加了一条自底向上的路径,加强底层和顶层之间特征信息融合的过程,有助于更好地整合不同层次的特征,提高模型对目标的感知和识别能力。高效的特征融合模块能够实现高低层信息的充分融合,提高特征表达。YOLOv7 所用的网络结构并未对 Backbone 输出给 FPN 的特征进行细致的处理,其特征表达明显不足。为增强特征表达,本节设计基于注意力的特征融合模块对 FPN 进行改进。双线性注意力平行特征融合机制结构如图 4-7 所示,增加了特征尺度的跳跃连接,并引入本文提出的双线性注意力机制。提高对贡献度较高节点的关注,并将不同感受野大小的特征图进行耦合,提升特征融合的效率。

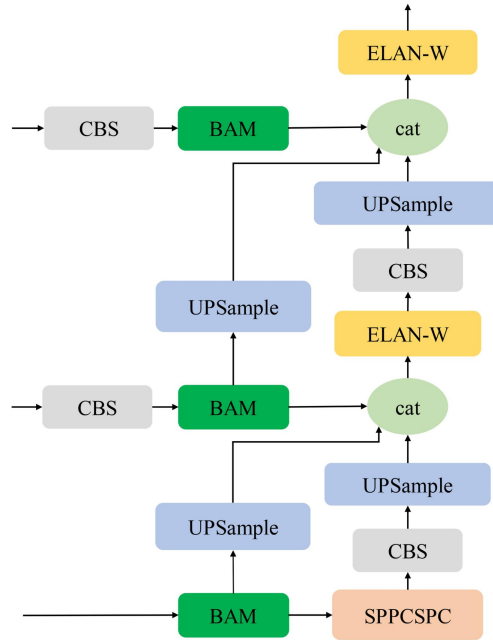


图 4-7 双线性注意力平行特征融合机制结构

双线性注意力机制加强了 Backbone 传给 FPN 顶层特征图的语义和定位特征,尤其语义特征。在此机制中,双线性注意力模块中的平均池化操作在顶层的小特征图上对定位信息的聚合能力强于对语义信息的聚合能力,增强了小特征图上的定位特征,有利于检测复杂场景下的大目标。最大池化操作在底层大特征图上对语义信息的聚合能力强于对定位信息的聚合能力,增强了大特征图上的语义特征,有利于检测复杂场景下的小目标。双线性注意力模块通过在不同层级的特征图上应用平均池化和最大池化操作,巧妙地平衡了定位信息和语义信息的聚合。这种机制使得模型在检测不同尺度和复杂度的目标时更加灵活和高效。

4.3 旋转目标检测策略

在一般场景下都使用水平框进行目标检测，用 (x, y, w, h) 参数来进行表示，分别代表框的中心点坐标、宽度和高度。但在航拍图像中任意方向分布的目标，水平框对其的应用效果偏差，存在背景信息和临近目标等干扰。基于此，旋转目标检测被提出，其增加旋转角度的信息，感知到目标更完整、更全面的特征信息。如图 4-8 所示，本文通过 (x, y, w, h, θ) 五个参数来确定旋转目标的边界框， θ 由 x 轴正方向和 w 的夹角来决定的，其范围在 $[-90^\circ, 90^\circ]$ ，利用 θ 把角度信息引入 YOLOv7 模型中。

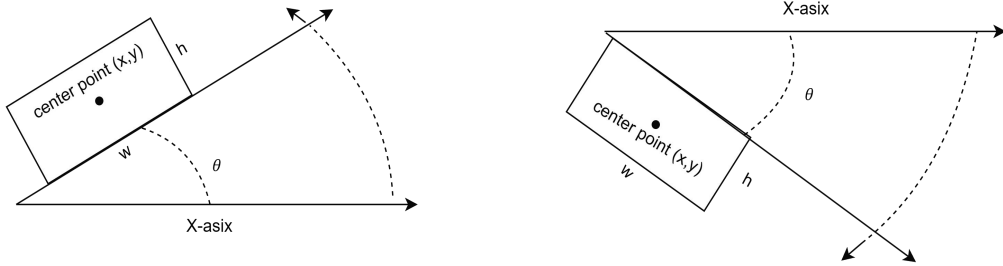


图 4-8 旋转方法表示

YOLOv7 模型的损失函数包括类别损失($Loss_{cls}$)、边框回归损失($Loss_{reg}$)和置信度损失($Loss_{obj}$)，具体如式(4-9)所示：

$$Loss = Loss_{reg} + Loss_{cls} + Loss_{obj} \quad (4-9)$$

$Loss_{cls}$ 和 $Loss_{obj}$ 采用的是二元交叉熵损失函数，如式(4-10)所示：

$$BCELoss = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(\sigma(x_i)) + (1 - y_i) \cdot \log(1 - \sigma(x_i)) \quad (4-10)$$

其中 $\sigma(x_i)$ 为 Sigmoid 函数，如式(4-11)所示：

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (4-11)$$

在 YOLOv7 中使用角度参数检测时，旋转边框回归可能引发旋转敏感度误差，降低了检测精度。为解决此问题，本章引入 KLD 损失函数计算边框回归损失。其原理是将旋转边界框转换为二维高斯分布，通过计算高斯分布间的相对熵

来实现，具体如式(4-12)所示：

$$D_{KL}(N_p \parallel N_t) = \frac{1}{2}(u_p - u_t)^T \sum_t^{-1} (u_p - u_t) + \frac{1}{2} \text{Tr}(\sum_t^{-1} \sum_p) + \frac{1}{2} \ln \frac{|\sum_t|}{|\sum_p|} \quad (4-12)$$

式中的每一项分别如式(4-13)，(4-14)，(4-15)所示：

$$(u_p - u_t)^T \sum_t^{-1} (u_p - u_t) = \frac{4(\Delta x \cos \theta_t + \Delta y \sin \theta_t)^2}{w_t^2} + \frac{4(\Delta y \cos \theta_t - \Delta x \sin \theta_t)^2}{h_t^2} \quad (4-13)$$

$$\text{Tr}(\sum_t^{-1} \sum_p) = \frac{h_p^2}{w_t^2} \sin^2 \Delta \theta + \frac{w_p^2}{h_t^2} \sin^2 \Delta \theta + \frac{h_p^2}{h_t^2} \cos^2 \Delta \theta + \frac{w_p^2}{w_t^2} \cos^2 \Delta \theta \quad (4-14)$$

$$\ln \frac{|\sum_t|}{|\sum_p|} = \ln \frac{h_t^2}{h_p^2} + \ln \frac{w_t^2}{w_p^2} \quad (4-15)$$

其中， $\Delta x = x_p - x_t$ ， $\Delta y = y_p - y_t$ ， $\Delta \theta = \theta_p - \theta_t$ ，分别是预测框和真实框之间的中心点坐标和角度差值。KLD 的计算如式(4-16)所示。

$$L_{KLD} = (N_p \parallel N_t) = 1 - \frac{1}{1 + \ln(1 + D_{KL}(N_p \parallel N_t))} \quad (4-16)$$

KLD 损失函数具备尺度自适应的权重修正功能，能有效降低微小角度误差对精度的影响，实现高精度的旋转检测。

4.4 检测可视化实验

将本章方法分别在两个数据集进行实验，进行可视化结果展示，结果如图 4-9，4-10 所示。



图 4-9 自建数据集上的可视化检测结果



图 4-10 NWPUVHR-10 数据集上的可视化检测结果

图 4-9 显示了该方法在自建数据集上的部分可视化结果，图 4-10 显示了该方法在 NWPUVHR-10 数据集上的部分可视化结果。从图 4-9 和 4-10 可以看出，本文提出的模型在处理复杂环境下的噪声问题方面展现出了较好的性能。在复杂场景中，该模型能够有效地滤除噪声，凸显出有用的信息，使得目标检测更加准确可靠。

4.5 本章小结

本章针对航拍图像存在视角特殊性，导致大量背景信息以及任意方向分布影响检测效果的问题，构建了融入注意力的航拍图像旋转目标检测方法。设计增强注意力提高主干网络的提取能力，使用双线性注意力构成平行融合机制，提高特征信息的融合，帮助网络有效聚集特征图上的语义特征。这两种注意力有利于检测复杂背景场景下的目标。增加检测目标的角度信息，并通过 KLD 作为损失值训练网络，优化旋转检测效果，提升网络在航拍图像中旋转目标的检测精度。

第 5 章 实验分析及客户端软件设计

前两章的内容针对航拍图像小目标检测中存在多尺度目标共存、复杂背景、目标密集以及方向任意分布等问题展开讨论研究,提出两种改进检测方法应对不同场景下的航拍图像小目标检测。本章将对提出的两种方法展开进行详细的实验,以验证方法的有效性。并设计客户端软件,给用户提供直观简洁的可视化操作界面。

5.1 基于特征增强的航拍图像目标检测方法的实验分析

5.1.1 特征增强目标检测方法的实验设置

本部分方法的实验环境是在 Windows 10 系统下进行的,使用了 PyTorch 1.10.2 深度学习框架, CUDA 版本为 11.3。所用硬件配置包括 NVIDIA GeForce RTX 3060 GPU(12GB 显存)、Intel(R) Core(TM) i7-10700 CPU(2.90GHz)、16GB RAM。

训练过程中的参数设置如下:输入图像大小为 640×640 ,采用 SGD 优化器,初始学习率为 0.01,学习率采用余弦退火衰减。

采用 VisDrone-DET2019^[50]数据集(以下简称 VisDrone2019)进行实验,该数据集包含 10 个类别,10209 张图像,其中 6471 张用于训练,548 张用于验证,3190 张用于测试。

性能评价指标采用每秒传输帧数(Frames Per Second, FPS),网络模型每秒可以检测多少张图片,以及平均精度均值(mean average precision, mAP),其计算方式如式(5-1)、(5-2)、(5-3)、(5-4)所示。

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detctions}} \quad (5-1)$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}} \quad (5-2)$$

$$AP = \int_0^1 P dR \quad (5-3)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (5-4)$$

在上述公式中，各项的含义如下：TP 为正确检测数量，FP 为误检数量，FN 为漏检数量，P 表示正确检测数量在全部检测结果中的比率，R 表示正确检测的数量占有所有目标数量的比率。

5.1.2 特征增强目标检测方法的消融实验分析

为验证各项改进模块的有效性，以 YOLOv5 为基准进行消融实验，实验结果如表 5-1 所示。

改进主干特征提取模块的有效性：为了证明改进主干特征提取模块的有效性，本节在 YOLOv5 的基础上加入该改进模块进行实验。如表 5-1 所示，与 YOLOv5 相比，在 VisDrone2019 数据集上，mAP 提高了 2.5 个百分点，FPS 则降低了 15。这说明改进的主干特征提取模块可以有效地提升多尺度目标的信息提取能力，证明该模块的有效性。

改进检测头模块的有效性：为了证明改进检测头模块的有效性，本节在 YOLOv5 的基础上改进该模块，并进行实验。如表 5-1 所示，与 YOLOv5 模型和改进主干特征提取模块相比，在 VisDrone2019 数据集上，mAP 分别提高了 6.7 和 4.2 个百分点，FPS 分别降低 17 和 2。这说明改进的检测头模块，通过增设小目标检测端可以有效地提升对航拍图像小目标的检测性能，加入 Transformer 模块在一定程度上提升了中尺度目标的检测性能，进而提升整体检测的性能。证明改进多尺度检测模块的不可缺失性。

整体结构的有效性：为证明整体结构的有效性，本节将整体方案加入 YOLOv5 中进行实验。如表 5-1 所示，在 VisDrone2019 数据集上与 YOLOv5 比，mAP 提高了 7.8 个百分点，FPS 降低了 28。较改进的两个模块 mAP 分别提高了 5.3 和 1.1 个百分点。这说明，本文提出的算法在抑制背景信息影响、获取小目标特征信息以及检测多尺度目标物体等方面取得较好的成果，提升航拍图像小目标检测性能。证明了本文所提算法的有效性，也说明改进的两个模型具有互补性与不可分割性。

可以从表 5-1 中看出，网络加入各个模块后 FPS 较 YOLOv5 都有所下降，本文提出算法的 FPS 为 54，但可以满足航拍目标检测要求。

表 5-1 消融实验对比

Methods	mAP/%	Frame Rate /FPS
YOLOv5	34	82
YOLOv5+CA	34.3	69
YOLOv5+CA+SPD	36.4	59
YOLOv5+CA+SPD+Transformer	36.5	67
YOLOv5+Transformer+Head	40.7	65
Our	41.8	54

本节详细列出各个改进模型在 VisDrone2019 数据集上对 10 个目标类别的检测精度，结果如表 5-2 所示。其中 YOLOv5+CA+SPD+Transformer 在表中简写为 YOLOv5+CST，YOLOv5+Transformer+Head 在表中简写为 YOLOv5+TH。从表 5-2 中可以看出，本文提出的各个改进模块较 YOLOv5 在各个类别 mAP 都有所上升。在 YOLOv5 网络检测结果中，数据集的 bicycle 和 awning 这两类的 mAP 仅达到 13.5%和 10.8%，检测的精度很差，这是因为数据集中这两类在图像中所占尺寸过小，且存在背景、视野等因素影响。本文方法将 bicycle 和 awning 的 mAP 提高了 3.7%和 4.9%，表明本文提出的方法具有较强的小目标检测能力，对于多尺度目标的检测效果也表现不错，适用于航拍图像小目标的检测方面的应用。

表 5-2 VisDrone2019 数据集各个类别 mAP 对比

	YOLOv5	YOLOv5+CST	YOLOv5+TH	our
pedestrian	41.4%	44%	49.5%	51.4%
people	32.9%	34.7%	37.4%	38.9%
bicycle	13.5%	14.1%	16.3%	17.2%
car	74.5%	75.7%	81.3%	82.5%
van	37.5%	39.5%	43.5%	45.1%
truck	29.2%	32.4%	33.4%	35.8%
tricycle	19%	21.7%	28%	27.6%
awning	10.8%	12.1%	14.8%	15.7%
bus	45.7%	48.5%	55.9%	56.1%
motor	39.4%	42%	46.4%	48.7%

5.1.3 特征增强目标检测方法的对比实验分析

为验证本部分所提方法的优越性,选取了具有代表性的深度学习目标检测模型作对比实验。分别选取 SSD512、Faster-RCNN、RetinaNet、YOLOv4 网络和本文算法进行测试,实验结果如表 5-3 所示。从表 5-3 可知,SSD512 和 RetinaNet 在航拍图像数据集上的检测效果欠佳。虽然 Faster-RCNN 效果稍好,但因其复杂的网络结构,导致帧率偏低。相较之下,YOLOv4 的检测效果有所提升,且比 YOLOv5s 高,因为本文选取的 YOLOv5s 兼顾了速度,所以降低了检测效果。但 YOLOv4 检测效果还是稍差于本文提出的模型,并且帧率也相差不大。由此可看出本文所提出方法在航拍图像目标检测能力较上述对比方法的优越性。

表 5-3 不同算法检测性能对比

Model	mAP/%	Frame Rate /FPS
YOLOv4	38.6	57.6
YOLOv5s	34	82
SSD512(VGG-16)	26.8	37
RetinaNet(ResNet-50)	22.7	32
Faster-RCNN(ResNet-50)	29.6	36
Ours	41.8	54

为验证本节算法对航拍图像目标检测的性能,选取针对航拍图像目标检测提出的深度学习模型进行对比实验,分别选取了 YOLOv3_ReSAM^[51], mSODANet^[52], TPH-YOLOv5, UCGNet^[53], 与 YOLOv8 进行测试,实验结果如表 5-4 所示。从表 5-4 的结果可以得出,本节提出的改进算法较上述网络模型的 mAP 都有提升。YOLOv3_ReSAM 模型利用残差结构的 SAM 注意力机制提升主干信息获取能力,帮助特征信息更好地融合,在多个类别目标检测精度优于本节算法。但在微小目标检测方面效果较差,具体表现为对行人和人两类目标的检测。主要是因为本节方法输出针对微小目标的预测头,使网络更有利于检测微小目标。mSODANet 模型利用 Hierarchical dilated network 充分学习上下文信息,能够在检测过程中获取更难检测目标的特征,在本文算法检测效果最低的遮阳棚中取得 24.67%的 mAP,提高了 8.97%的 mAP。但该网络只输出两个预测头,与本节改进的检测端相比,无法对微小目标实现精准检测。TPH-YOLOv5 模型在检测端也输出微小目标预测头,但对微小目标的检测精度还是偏差,这由于本文对主干提取网络进行了改进,使网络可以提取更多的目标特征信息,进而利用更多

特征信息进行预测。UCGNet 提出一个 LLM 来预测对象的分布图和一个无监督的聚类模块来生成密集的子区域，但骨干网络由三个 BottleNeckCSP 和一个 SPP 模块组成，对于小目标的细节信息提取不够，从而导致检测低于本节提出的算法。选取 YOLOv8 进行对比，可以看出 YOLOv8 的检测精度还是稍低于本文提出的算法，主要是因为本文算法设置了更利于检测小目标的检测端。

表 5-4 航拍图像检测算法 mAP 对比

	YOLOv3_ReSAM	mSODANet	TPH-YOLOv5	UCGNet	YOLOv8	our
mAP	33.15%	36.89%	39.18%	32.8%	40.3%	41.8%
pedestrian	17.28%	38.72%	39.2%	34%	43.8%	51.4%
people	15.78%	21.43%	27.1%	22.8%	33.3%	38.9%
bicycle	28.70%	20.28%	23.3%	19.4%	13.3%	17.2%
car	84.31%	59.61%	68.1%	58.6%	79.9%	82.5%
van	53.77%	44.32%	45.4%	39.6%	45.5%	45.1%
truck	42.96%	41.73%	39.6%	33%	36.4%	35.8%
tricycle	30.48%	35.75%	29.5%	23.1%	29.7%	27.6%
awning	24.38%	24.67%	18.4%	14.5%	18%	15.7%
bus	62.35%	49.93%	56.3%	52.4%	56.3%	56.1%
motor	61.57%	32.52%	36.1%	30.9%	46.7%	48.7%

为验证本文方法的泛化性，将改进方法放入 YOLOv3 进行实验，实验结果如表 5-5 所示。从表 5-5 可以看出在 YOLOv3 中加入本节提出的改进方法后，YOLOv3 对于 Visdrone2019 数据集检测的 mAP 有很好地提升，同时也可以看出数据集中各个类的 mAP 都有所提高，由此可以证明本文提出方法的普适性，本文改进的方法可以帮助网络模型更好地提取小目标的细节信息，进而提高对小目标的检测精度。

表 5-5 改进 YOLOv3 的检测精度对比

	YOLOv3	Improve YOLOv3
mAP	28.1%	35.1%
pedestrian	38%	46.2%
people	31.2%	37.1%
bicycle	7.15%	10.1%
car	71.1%	79%
van	26.6%	35.3%
truck	21.3%	26.8%
tricycle	15%	21%

续表 5-5 改进 YOLOv3 的检测精度对比

	YOLOv3	Improve YOLOv3
awning	7.12%	11.8%
bus	28.5%	41%
motor	35.5%	42.8%

通过以上对比实验，证实本文提出算法的优势。本文改进方法有效地提高网络的提取能力，模型在面对航拍图像目标检测任务时具有更大的优势。

5.2 融入注意力的航拍图像旋转目标检测方法的实验分析

5.2.1 注意力旋转目标检测方法的实验设置

本部分方法的实验环境与 5.1.1 节实验环境相同。使用的自建数据集，共有船舶和飞机两个类。利用 Google Earth 进行截图，并使用 roLabelImg 对获取的数据集进行旋转目标标记，最后利用分割方法，获得 618 张 1024*1024 大小的图像数据集，其中 498 张图片作为训练集，60 张图片作为验证集，60 张图片作为测试集。部分数据如图 5-1 所示。



图 5-1 自建数据集示例图

本文使用的 NWPUVHR-10 数据集^[54-55]包含 800 张高分辨率的卫星图像，数据集共分为 10 类。图像共计 650 张。所有图像的标签由手动进行标注。由于采用监督学习，因此只选用正样本数据集进行实验。按照 8 : 1 : 1 的比例划分为训练集、验证集和测试集。最终得到 520 张训练图像，65 张验证图像，65 张测试图像。部分数据集的示例如图 5-2 所示。



图 5-2 NWPUVHR-10 数据集示例图

5.2.2 注意力旋转目标检测方法的消融实验分析

本节以 R_YOLOv7 代表在 $YOLOv7$ 引入 KLD 损失函数旋转目标检测方法，作为基准模型，在自建数据集进行消融实验。消融实验结果如表 5-6 所示，其中 GFLOPs 为每秒 10 亿次的浮点运算，用于表示模型的复杂度。EAM 表示本文提出的增强注意力模块，BABFM 表示本文提出的双线性注意力平行融合机制。

增强注意力模块的有效性：为了证明增强注意力模块的有效性，本节在 R_YOLOv7 基础上加入该改进模块进行实验。结果如表 5-6 所示，与基准网络相比，在自建数据集上，mAP 提高了 2.8，提升了 3.1%，GFLOPs 上升 2.4，增加了 2.3%。这说明提出的注意力模块在主干特征提取网络中可以有效地提升航拍图像小目标的信息提取能力，证明了该模块的有效性。

双线性注意力平行融合模块的有效性：为了证明双线性注意力平行融合模块的有效性，本节在 R_YOLOv7 基准网络的基础上改进该模块，并进行实验。如表 5-6 所示，在自建数据集上，双线性注意力平行融合模块和基准网络相比，mAP 提高了 2.6，提升了 2.9%，GFLOPs 上升了 1，增加了 1%。这说明该模块的检测效果较好。这种融合方式充分发挥了注意力机制强化特定特征的能力，使得模型更加注重对关键信息的捕捉和利用，从而提升目标检测的性能，证明了双线性注意力平行融合模块的不可缺失性。

整体结构的有效性：为了证明整体结构的有效性，本节将整体方案进行实验。如表 5-6 所示，在自建数据集上与基准网络相比，mAP 提高了 3.2，提升了 3.6%，GFLOPs 上升了 3.3，增加了 3.1%。较改进的两个模块 mAP 分别提高了 0.4 和 0.6。这说明本文提出的算法在抑制背景信息影响和检测航拍图像中目标物体等方面取得较好的成果，从而提升航拍图像目标检测性能。通过消融实验证明本文

所提算法的有效性，也说明了两个模型的互补性与不可分割性。

表 5-6 消融实验对比

Model	mAP/%	GFLOPs
R_YOLOv7	89.9%	105.1
R_YOLOv7+EAM	92.7%	107.5
R_YOLOv7+BABFM	92.5%	106.1
Our	93.1%	108.4

5.2.3 注意力旋转目标检测方法的对比实验分析

为验证所设计增强注意力模块的有效性，本节以 R_YOLOv7 为基准模型，在自建数据集上进行实验，并以 CA、SE、CBAM 模块替换 EAM 模块进行对比实验。实验结果如表 5-7 所示。EAM 模块引入的附加参数和计算成本适中，与之前提出的注意力模块相比，EAM 模块在准确性和计算成本之间提供了有竞争力的权衡。

SENet 只关注通道区域的权重分布，而忽略了空间区域的权重分布。本文提出的 EAM 生成了一个三维注意力矩阵，不仅对通道注意力表示进行建模，而且突出空间注意力的多样性分布。CBAM 依次通过通道注意力模块和空间注意力的模块，分别生成二维响应，但值得注意的是，其不是分别产生通道和空间注意力的最佳解决方案，因为忽略了通道和空间分布之间的协同作用。本节提出的 EAM 将通道注意力纳入空间注意力，这更有利于嵌入特征图的相关性。CA 提升注意力模块的运行效率，但该模块对空间域信息特征的利用仍不全面，无法捕捉长距离的依赖关系。本文提出的 EAM 不仅专注于信道注意力建模，而且使网络获取到特征图中的哪些信息在空间层面具有更高的响应。

表 5-7 不同注意力机制与 EAM 对比

Model	mAP/%	GFLOPs
R_YOLOv7+EAM	92.5	106.1
R_YOLOv7+SE	90.2	105.3
R_YOLOv7+CBAM	91	105.5
R_YOLOv7+CA	89.1	105.5

为验证双线性注意力模块(BA)的有效性，本节以 R_YOLOv7+注意力平行融合机制为基准模型，在自建数据集上进行实验，并以 ECA、SA、CBAM 模块替

换双线性注意力模块进行对比实验，实验结果如表 5-8 所示。

ECA 模块可以避免降维，但它仅是对 SE 的改进，只关注通道注意力，而忽略空间注意力。SA 为空间注意力，主要是突出哪里的特征重要，但其缺少通道注意力。双线性注意力机制在 ECA 和 SA 的基础上，对它们进行并联融合，提升对通道和空间注意力的利用。CBAM 注意力顺序组合通道注意力和空间注意力机制，有效地捕获更多的空间环境，有助于集中空间位置信息。但检测效果还是低于本文方法，进一步证明双线性注意力可以更好地利用通道与空间注意力机制，充分对特征信息的利用，CBAM 对特征的空间信息的利用率不高，缺乏建立远程通道依赖关系的能力。

表 5-8 不同注意力与双线性注意力对比

Model	mAP/%	GFLOPs
baseline+BA	92.7	107.5
baseline +ECA	90.5	107.5
baseline +SA	91	107.5
baseline +CBAM	91.6	107.7

为验证本文旋转检测方法的性能，在自建数据集上完成本文方法与 rotated reppoints^[56]、sasm reppoints^[57]、rotated faster rcnn 和 rotated atss^[58]等旋转目标检测方法的对比实验，结果如表 5-9 所示。由表 5-9 可知，本文提出的方法取得了最好的结果，在所有类别中都取得了相对稳定的检测结果，其 mAP 结果为 93.1%。rotated atss 和 rotated reppoints 网络检测效果较差，对密集旋转目标的检测效果相对较低，其较差的特征提取网络性能有待改进，sasm reppoints 和 rotated faster rcnn 网络模型均取得了较好的检测效果。与其他方法相比，本文方法对角度损失函数和利用多种注意力对特征信息的处理，在背景复杂、分布密集目标的图像检测中，该模型取得较好的分类效果。与次优方法相比，该方法的 mAP 值提高了 7.6%。

表 5-9 旋转算法 mAP 对比

Module	Plane/%	Ship/%	mAP/%
rotated reppoints	98	70.1	65.8
sasm reppoints	90.3	73.8	82.1
rotated faster rcnn	99.6	71.4	85.5
rotated atss	98.9	36.7	67.8
our	94.6	91.6	93.1

证明所提出的注意力方法的性能，在 NWPUVHR-10 数据集上将所提出的方

法加入 YOLOv7 与 YOLO 系列的方法进行比较, 表 5-10 展示了比较结果。由表 5-10 可知, 本文提出的模型取得了最好的结果, mAP 为 91.6%, 并且本文提出方法的 GFLOPs 值相比于其他四种算法也偏小。YOLOv6 和 YOLOv7 网络检测效果较差, 对航拍图像小目标的检测效率相对较低。

表 5-10 YOLO 系列算法对比

module	mAP/%	GFLOPs
YOLOv5	89.7	108.4
YOLOv6	74.4	144.4
YOLOv7	88.3	105.3
YOLOv8	88.4	192.5
our	91.6	108.5

为验证 EAM 模块的有效性, 本节以 YOLOv7 为基准模型, 在 NWPUVHR-10 数据集上进行实验, 并以 CA、SE、CBAM 模块替换 EAM 模块进行对比实验, 实验结果如表 5-11 所示, 列出在 NWPUVHR-10 数据集上 10 个类别的检测精度。从表 5-11 中可以看出, 本文提出的增强注意力模块较其他注意力 mAP 都有所上升。baseball diamond 取得了 96.7% 的 mAP, 高于其他注意力模块。

表 5-11 不同注意力与 EAM 对比

	YOLOv7	YOLOv7+EAM	YOLOv7+SE	YOLOv7+CBAM	YOLOv7+CA
airplane	99.5%	99.5%	99.5%	99.5%	99.5%
ship	94.4%	93%	96.5%	95.4%	99.1%
storage tank	96.8%	93.4%	97.5%	97%	97.15
baseball diamond	91.9%	96.7%	94.9%	95.5%	94.9%
tennis court	89.3%	89.4%	89.5%	89.4%	87.3%
basketball court	41.3%	77.8%	51%	59.4%	61.3%
ground track field	99.5%	99.3%	99.1%	99.5%	99.5%
harbor	99.5%	98%	99.5%	99.5%	97.6%
bridge	75.9%	61.8%	66.5%	73.8%	70.2%
vehicle	94.8%	95.8%	94.9%	94.3%	98%
all	88.3%	90.5%	88.9%	90.3%	90.4%

为验证双线性注意力模块的有效性, 本文以 YOLOv7+注意力平行融合机制为基准模型, 在 NWPUVHR-10 数据集上进行实验, 并以 ECA、SPA、CBAM 模

块替换双线性注意力模块进行对比实验，实验结果如表 5-12 所示，列出在 NWPUVHR-10 数据集上 10 个类别的检测精度。从表 5-12 中可以看出，本文提出的双线性注意力在 baseball diamond, basketball court 取得很好的检测精度，同时与其他注意力对比，该模块对各个类的检测精度比较平均，没有出现个别检测精度很低的现象。

表 5-12 不同注意力与 BA 对比

	YOLOv7	baseline+BA	baseline+ECA	baseline+SPA	baseline+CBAM
airplane	99.5%	99.5%	99.5%	99.5%	99.5%
ship	94.4%	89.6%	93.8%	95.6%	97.6%
storage tank	96.8%	96.5%	98%	94.4%	98.4%
baseball diamond	91.9%	97.7%	95.7%	94.5%	93.8%
tennis court	89.3%	89.5%	91.4%	87.2%	88.7%
basketball court	41.3%	74.1%	45.6%	68.9%	43%
ground track field	99.5%	99.1%	99.3%	94.9%	98.6%
harbor	99.5%	99.5%	99.5%	98%	99.5%
bridge	75.9%	70.2%	76.4%	77.4%	74.9%
vehicle	94.8%	89.3%	95.5%	90.1%	94.4%
all	88.3%	90.5%	89.5%	90.1%	88.8%

通过以上对比实验，证实本文提出注意力方法的优势。对比其他旋转目标检测算法，本文方法也取得了不错的成绩，证实本文提出的基于注意力的航拍图像旋转目标检测方法的优越性。

5.3 客户端软件设计

5.3.1 PyQt5 简介

PyQt5 是 Qt5 框架与 Python 的集成，Qt 作为图形用户界面(Graphical User Interface, GUI)应用程序框架，提供多样化的用户界面控件和功能，助力创建风格各异的图形用户界面应用。PyQt5 使 Python 开发者能够便捷地利用 Qt5 的 GUI 特性来打造应用。它全面提供 Python 接口，将 Qt 的 C++类转化为 Python 类，

同时融合 Python 的语法特性，优化开发流程。PyQt5 不仅承袭 Qt 的跨平台特性，还可在 Windows 和 Linux 等系统上进行开发使用。还兼容多种图形用户界面设计工具，如 Qt Designer，从而便于开发者通过直观易用的界面，轻松地设计应用程序。在 PyQt5 中，常用的模块如 QtXml、QtCore 等，都为构建高质量的 GUI 应用提供了坚实支持。

综上所述，PyQt5 是一款功能丰富、易于上手且强大的库，适合开发各类 GUI 应用程序。

5.3.2 客户端软件架构

本节通过 PyQt5 设计面向航拍图像小目标检测的客户端软件^[59]，应用本文提出的基于特征增强的航拍图像目标检测方法对航拍小目标进行检测并及时将检测结果进行呈现^[60]。客户端软件组成图如图 5-3 所示。

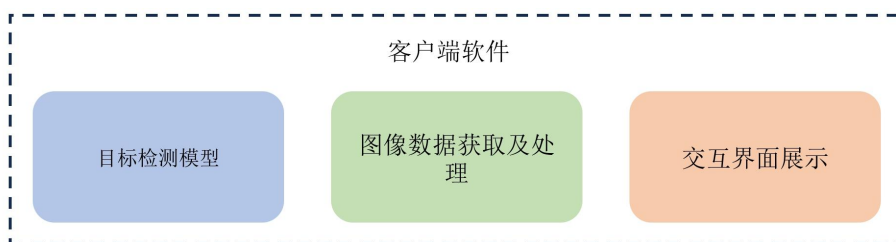


图 5-3 客户端软件组成图

客户端软件主要由图像数据获取及处理、目标检测模型和交互界面展示^[61]构成。图像数据获取及处理模块负责捕获和处理图像数据，为目标检测模型提供输入。目标检测算法模块专门用于识别和分类航拍图像中的目标，并输出检测结果。而交互界面则为用户提供了一个直观、简洁的可视化平台，方便用户进行交互和控制。

用户与交互界面互动时，界面发出信号，目标检测模型随即捕获并响应。模型根据信号调用相应的业务逻辑，准备所需环境和资源。随后，模型向图像数据获取模块发出请求，获取测试数据集。模块获取数据后返回给目标检测模型，目标检测模型利用已训练模型进行验证，并将结果传回交互界面。界面即时显示结果，确保用户轻松触发检测操作并获取即时反馈。这一流程简化了用户操作，提高了检测效率。

本文的客户端软件设计流程图详细展示了整个开发过程，如图 5-4 所示，主

要分为两大部分：航拍图像小目标检测模型设计和可视化软件界面设计。在模型设计部分，首先分析了航拍图像小目标检测的现状和挑战，选择了合适的网络模型。随后，利用航拍小目标数据集进行模型训练，得到了基准模型。在此基础上，对基准模型进行了优化，得到了性能更优的模型。在界面设计部分，根据需求使用 PyQt5 设计了简洁明了的软件界面，并划分各个功能模块。最后，将最优模型嵌入到可视化界面的功能区，实现用户通过界面调用模型进行目标检测的功能。

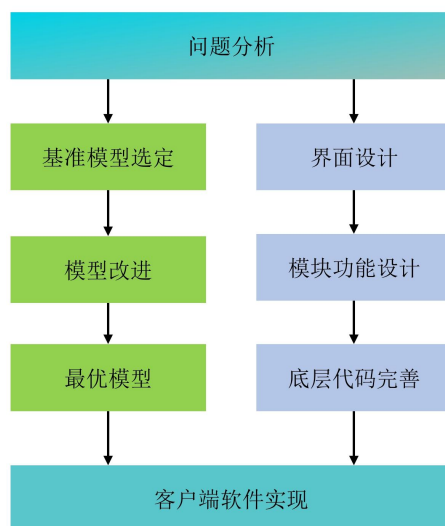


图 5-4 软件设计框图

5.3.3 客户端软件展示

图 5-5 所示为航拍图像小目标检测客户端的首页页面。



图 5-5 客户端检测首页

首先从界面中选取文件夹中航拍图像,也可以选取航拍视频以及打开电脑的摄像头,上传好选定的航拍文件后,然后选取本文提出方法训练好的网络模型,加载到客户端的底层,也可以进行部分参数的设置,最后进行航拍图像小目标检测。

图 5-6 为图像检测结果页面展示,本文选取了 VisDrone-DET2019 数据集中图像进行检测,利用本文提出基于特征增强的航拍图像目标检测方法进行实验。

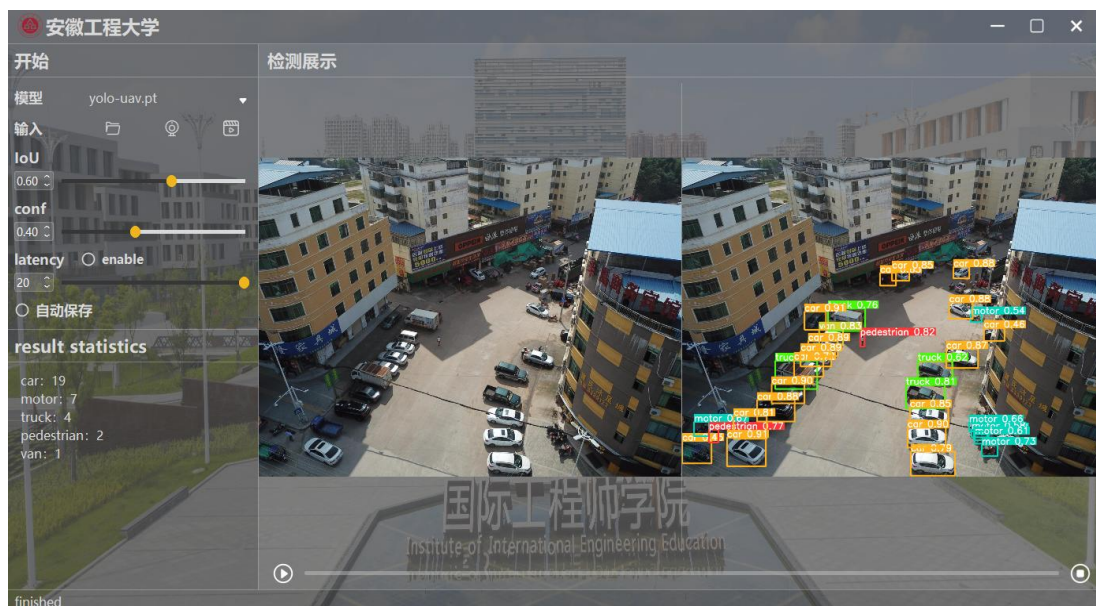


图 5-6 客户端检测结果

5.4 本章小结

本章通过实验论证了本文所提方法的有效性和优越性并利用 PyQt5 设计客户端软件。

基于特征增强的航拍图像目标检测方法。首先对实验环境和 VisDrone-DET2019 数据集进行概述,然后通过消融实验和对比实验,验证了本文所提算法对航拍图像小目标检测的有效性。最后验证本文方法的通用性,将改进方法移植到 YOLOv3,实验结果表明,改进的方法取得了很好的检测效果。

基于注意力的航拍图像旋转目标检测方法。首先对实验环境、自建数据集和 NWPUVHR-10 数据集进行简要的介绍,然后分别利用两组数据进行对比实验,验证了本文提出注意力的高效性。最后通过对比实验,证明本文方法在旋转目标检测方面的有效性。

构建客户端软件。首先论述了 PyQt5 框架的知识，详细探讨了如何借助该框架构建航拍图像小目标检测客户端软件。随后，对软件的模块功能和整体架构进行讨论，并最终展示软件的操作界面。

第 6 章 总结与展望

6.1 总结

航拍图像小目标检测是一个重要且广泛应用的方向，涉及军事和民用领域，具有巨大价值。然而，其还面临诸多挑战，包括背景复杂、目标尺度差异大、目标密集排列以及目标方向多样等问题。传统方法在处理航拍图像时效率低且鲁棒性差，深度学习的兴起，利用深度学习进行目标检测取得很好效果。但直接应用这些方法对于航拍图像通常难以取得良好效果，因此，本文在深度学习的基础上，分析航拍图像特点，有针对性地提出了适用于该场景的目标检测方法。本文所做工作如下：

(1) 针对航拍图像中存在小目标居多，多尺度目标共存以及目标密集等问题，本文提出一种基于特征增强的航拍图像目标检测方法。在主干特征提取网络引入 CA 模块、Transformer 模块以及 SPD-Conv 模块来减少模型在下采样时对特征信息的丢失，提高特征信息的获取。同时改进网络检测端，通过新的网络检测端来实现对多尺度目标的检测。为验证本文提出方法的性能，分别进行了消融实验和对比实验，通过实验得出，本文提出的方法都取得了较好的性能。

(2) 针对航拍图像中目标存在方向任意，复杂背景，进而影响检测效果的问题，本文提出了融入注意力的航拍图像旋转目标检测方法。该方法设计了旋转目标检测方法，并利用 KLD 方法提高了任意角度目标的检测精度，重新定义分类损失函数，在损失计算中加入角度问题。同时该方法还设计了两种注意力机制用于网络，减少背景对特征提取以及融合的影响，进而提高目标的检测效果，使该模型具有较佳的检测性能。

(3) 设计了一款基于 PyQt5 的目标检测客户端软件，该软件结合本文提出的方法，通过用户界面实现了对场景中目标的有效检测。整个系统由三大核心模块组成：图像数据获取与处理模块、目标检测算法模块和用户交互界面。图像数据获取与处理模块的主要职责是从文件中捕获图像信息，并将其转换为数字化数据流，为后续的目标检测提供必要的技术支持。目标检测算法模块则专注于对图像中的物体进行精确检测，该模块利用先进的算法对图像进行深度分析，识别出目

标物体，并输出相应的检测结果。用户交互界面采用了 PyQt5 技术，它具备跨平台的特性，且拥有丰富的图形界面元素，使得用户能够更直观地操作和控制整个系统。界面设计简洁明了，提供了便捷的可视化操作，使用户能够轻松地进行目标检测操作。在测试阶段，本文使用真实的图像数据对系统进行了验证。结果表明，该系统能够有效地检测出图像中的目标物体，并具备较高的实用性和准确性。通过这款基于 PyQt5 的目标检测客户端软件，用户能够轻松实现对场景中目标的检测，为实际应用提供了强有力的支持。

6.2 展望

本文在计算机视觉领域的重要研究方向——目标检测技术上进行深入的研究，尤其在航拍图像目标检测领域的应用具有巨大的价值，无论是在军事还是民用方面。尽管本文做了大量的研究工作，但由于受到时间限制、实验条件制约以及个人能力有限等因素影响，所提出的方法仍存在一些不足，需要进一步探索和改进。

首先，本文研究的模型采用的是有监督学习，对于供网络训练所需的数据集，要求通过人工标记的方式对图像中的目标进行标注。然而，在计算机视觉研究中，每个人的标注习惯可能存在不同，进而标注后的结果也会存在显著的标准偏差，这对模型的训练和学习造成了一定影响。为解决这一问题，后续研究可考虑采用半监督或弱监督学习的方式处理训练数据，以减少人为因素导致的训练误差，提高模型的准确性和稳定性。

其次是数据集的不足。现有的公开航拍图像数据集偏少，并且这些数据集在类别细粒度和标注精度还存在不足。如类别缺失，对于车辆、船只等目标没有进一步细分。还有存在模糊目标未被标注，这都一定程度限制航拍图像小目标检测发展。

最后是算法的实时性。虽然本文的方法在精准率上表现较好，但模型运行速度不高。未来研究可以聚焦于模型速度的优化，利用轻量化的方法改善模型，使其可以在无人机设备部署使用。

参考文献

- [1] 徐印赞, 江明, 李云飞, 等. 基于改进 YOLO 及 NMS 的水果目标检测[J]. 电子测量与仪器学报, 2022, 36(04): 114-123.
- [2] 徐印赞. 基于深度学习的移动机器人目标物体识别与定位技术研究[D]. 安徽: 安徽工程大学, 2022.
- [3] Sambolek S, Marina I. Automatic person detection in search and rescue operations using deep CNN detectors[J]. IEEE Access, 2021, 9: 37905-37922.
- [4] Božić Š, Dunja, Zeljko M, et al. Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions[J]. International Journal of Computer Vision, 2019, 127: 1256-1278.
- [5] 朱子健, 刘琪, 陈红芬, 等. 基于并行融合网络的航拍红外车辆小目标检测方法[J]. 光子学报, 2022, 51(2): 190-202.
- [6] Osco L P, De A M S, Goncalves D N, et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 174: 1-17.
- [7] Eeranampalayam S A N, Li J, Scott S, et al. Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery[J]. Remote Sensing, 2020, 12(13): 2136-2158.
- [8] Guo J, Xie J, Yuan J, et al. Fault identification of transmission line shockproof hammer based on improved YOLOv4[C]//2021 International Conference on Intelligent Computing, Automation and Applications (ICAA), Nanjing, China, 2021: 826-833.
- [9] Liu C, Wu Y, Liu J, et al. MTI-YOLO: A light-weight and real-time deep neural network for insulator detection in complex aerial images[J]. Energies, 2021, 14(5): 1426-1445.
- [10] 李澜. 基于无人机高光谱遥感的城市中小河流水质等级判别研究[D]. 上海: 中国科学院大学(中国科学院上海技术物理研究所), 2021.
- [11] Liu Y C, Shi G, Li Y X, et al. M-YOLO based detection and recognition of

- highway surface oil filling with unmanned aerial vehicle[C]//2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022: 1884-1887.
- [12] 孟文文. 面向铁路视频监控的小尺度行人检测算法研究与 ARM 平台部署[D]. 北京: 北京交通大学, 2022.
- [13] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition, 1996, 29(1): 51-59.
- [14] Lowe, David G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60: 91-110.
- [15] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, 2005: 886-893.
- [16] Liu Y J, Yang F B, Hu P. Small-Object Detection in UAV-Captured Images via Multi-Branch Parallel Feature Pyramid Networks[J]. IEEE Access, 2020, 8: 145710-145750.
- [17] Lin Q Z, Ding Y, Xu H, et al. ECascade-RCNN: Enhanced Cascade RCNN for Multi-scale Object Detection in UAV Images[C]//Proceedings of the 2021 7th International Conference on Automation, Robotics and Applications (ICARA), Prague, Czech Republic, 2021: 268-272.
- [18] Li K, Wang B. DAR-Net: Dense Attentional Residual Network for Vehicle Detection in Aerial Images[J]. Computational Intelligence and Neuroscience, 2021, 2021: 6340823-6340842.
- [19] 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.
- [20] 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.
- [21] Wang X R, Li W H, Gou W, et al. SPB-YOLO: An Efficient Real-Time Detector For Unmanned Aerial Vehicle Images[C]//Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Jeju, Island, 2021: 99-104.
- [22] Gao F, Huo Y, Sun J, et al. Ellipse Encoding for Arbitrary-Oriented SAR Ship Detection Based on Dynamic Key Points[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-28.
- [23] Zhang Q, Tang J, Zheng H, et al. Efficient object detection method based on aerial optical sensors for remote sensing[J]. Displays, 2022, 75: 102328-102341.
- [24] Zand M, Etemad A, Greenspan M. Oriented Bounding Boxes for Small and Freely Rotated Objects[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.
- [25] Liu F, Chen R, Zhang J, et al. R2YOLOX: A Lightweight Refined Anchor-Free Rotated Detector for Object Detection in Aerial Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.
- [26] Chen J, Wan L, Zhu J, et al. Multi-Scale Spatial and Channel-wise Attention for Improving Object Detection in Remote Sensing Imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(4): 681-685.
- [27] Liu X, Di X. Global Context Parallel Attention for Anchor-Free Instance Segmentation in Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [28] Xu L, Wang L, Cheng S, et al. MHANet: A hybrid attention mechanism for retinal diseases classification[J]. Plos one, 2021, 16(12): 1-20.
- [29] 江学军, 唐焕文. 前馈神经网络泛化性能的系统分析[J]. 系统工程理论与实践, 2000, 20(8): 36-40.
- [30] Tang J, Deng C, Huang G B. Extreme learning machine for multilayer perceptron[J]. IEEE transactions on neural networks and learning systems, 2015, 27(4): 809-821.

- [31]李俭川, 秦国军, 温熙森, 等. 神经网络学习算法的过拟合问题及解决方法[J]. 振动测试与诊断, 2002, 22(4): 260-264.
- [32]Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 779-788.
- [33]Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [34]He K M, Gkioxari G, Dollar P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2980-2988.
- [35]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA, 2016: 770-778.
- [36]Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 2117-2125.
- [37]Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UAS, 2018: 6154-6162.
- [38]Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]//Proceeding of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 21-37.
- [39]Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999-3007.
- [40]Cho K, Van M B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Toronto, Canada, 2014: 1724-1734.

- [41] Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems, California, USA, 2014: 2204-2212.
- [42] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, USA, 2018: 7132-7141.
- [43] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 2018: 3-19.
- [44] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]// International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020:134-142.
- [45] 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, USA, 2021: 13713-13722.
- [46] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// 31th Annual Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 5998-6008.
- [47] 邱晓凤. 基于 YOLOv5 航拍视角下的小目标检测研究[D]. 西藏: 西藏大学, 2023.
- [48] Yang X, Yang X, Yang J, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence[J]. Advances in Neural Information Processing Systems, 2021, 34: 18381-18394.
- [49] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, UAS, 2020: 11534-11542.
- [50] Du D, Zhu P, Wen L, et al. VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results, 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, South Korea, 2019: 213-226.
- [51] Liu B, Luo H, Wang H, et al. YOLOv3_ReSAM: A Small-Target Detection

- Method[J]. Electronics, 2022, 11(10): 1635-1650.
- [52]Chalavadi V, Jeripothula P, Datla R, et al. mSODANet: A network for multi-scale object detection in aerial images using hierarchical dilated convolutions[J]. Pattern Recognition, 2022, 126: 108548-108558.
- [53]Barnwal R P, Bharti S, Misra S, et al. UCGNet: wireless sensor network based active aquifer contamination monitoring and control system for underground coal gasification[J]. International Journal of Communication Systems, 2017, 30(1): 2852-2870.
- [54]Cheng G, Han J, Zhou P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.
- [55]Cheng G, Han J. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.
- [56]Yang Z, Liu S, Hu H, et al. Reppoints: Point set representation for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision, Long Beach, USA, 2019: 9657-9666.
- [57]Hou L, Lu K, Xue J, et al. Shape-adaptive selection and measurement for oriented object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2022: 923-932.
- [58]Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, USA, 2020: 9759-9768.
- [59]徐笑锋, 肖英杰, 章学来. 基于避碰规则自动驾驶船舶路径优化的研究进展[J]. 交通信息与安全, 2021, 39(02): 1-8.
- [60]许志广, 刘建卿. 基于大数据分析的船舶航行安全自动评估系统设计[J]. 舰船科学技术, 2018, 40(06): 31-33.
- [61]郭栋梁. 基于红外图像的内河船舶目标检测技术研究[D]. 重庆: 重庆大学, 2009.

攻读学位期间发表的学术论文与研究成果

一、发表学术论文

- [1] Zhang J, Wan G, Jiang M, et al. Small object detection in UAV image based on improved YOLOv5[J]. Systems Science & Control Engineering, 2023, 11(1): 2247082. (EI 收录)
- [2] Zhang J, Li Y, Wan G, et al. Small Target Detection Algorithm for UAV Based on Improved YOLOv5[C]//2023 8th International Conference on Signal and Image Processing (ICSIP). IEEE, 2023: 78-82. (EI 会议)
- [3] Zhang J, Jiang M, Bai S, et al. Remote Sensing Image Rotation Target Detection Algorithm Based on Attention Mechanism[C]//2023 China Automation Congress (CAC). IEEE, 2023: 805-810. (EI 会议)。

二、实审的发明专利

- [1] 江明,张健,柏受军,等. 基于无人机航拍的电动车骑手头盔佩戴监测系统.CN202310709440.3.
- [2] 江明,张健,万国扬,等. 基于无人机航拍的高速道路交通监测系统及方法.CN202211467442.8.

三、参与的科研项目

- [1] 电气传动与控制安徽省重点实验室开放基金资助项目《面向金属铸件的工业机器人单目视觉定位策略研究》(DQKJ202206)。
- [2] 安徽省教育厅重点项目《基于深度学习的汽车流水灯缺陷检测策略研究》(2023AH050924)。

致谢

闲云潭影日悠悠，物换星移几度秋。三年前我走进 A 座的景象历历在目，但是美好的时光总是匆匆，转眼间，在安徽工程大学电气工程学院的求学时光便要结束了。如今在论文即将完成之际，我谨向所有帮助过我的老师、同学们致以真诚的感谢和祝福，也感谢母校对我的培养，祝福母校越来越好！

令公桃李满天下，何用堂前更种花。衷心感谢我的导师江明教授，在研究生学习期间，您以丰富的人生阅历和严谨的学术态度，深刻地影响并塑造了我的思维方式。您的严谨、深入、直面、比较的科研思维，如同明灯一般照亮我前行的道路，使我在学术探索中受益匪浅。即便在繁忙的工作间隙，您依然严格督促我，耐心指导我，为我的学习与论文倾注了无尽的关心与付出。您每一次细致的审阅、每一个中肯的建议，都极大地提升了我的学术水平和研究能力。您的付出与关怀，让我感受到了学术的严谨与温暖，也坚定了我追求卓越的信念。此外，我还要特别感谢您对我工作的推荐与帮助。您对我的鼓励，让我更加坚定地走在自己的职业道路上，追求更高的成就。在此，真挚地像向您表达真挚的感谢和祝福。

忽遇文殊开慧眼，他年应记老师心。感谢柏受军副教授对我的悉心指导，三年来，不仅在学术上给予我严谨的指导和督促，而且在日常生活中时刻关注我的成长和进步。在课题研究中，您始终关注着我的进展，耐心解答我的疑惑。在我的论文撰写过程中，您认真审阅了我的论文，还提出了许多宝贵的修改意见，使我的论文质量得到了提升。感谢万国扬老师在我研究课题专业内容方面所给予的耐心指导，在您的悉心引领下，让我对课题的研究内容有了更深入的理解。在课题探索的道路上，是您引导我入门，为我指明了前行的方向。在论文撰写过程中，您不仅对我的小论文提出了宝贵的修改意见，还对大论文的完善给予了细致入微的指导。感谢李云飞老师在学术和生活上的帮助，让我在人生道路上不断成长。在组会中，您都会针对我的研究内容提出宝贵的指导意见，在撰写论文中，您对我的论文提出了许多具有针对性的建议，这些让我受益匪浅。感谢从小到大所有教导过我的老师，正是老师们的悉心指导和无私付出，让我在知识的海洋中得以茁壮成长。

希君生羽翼，一化北冥鱼。感谢已经毕业的师兄们，徐印赞、卢志远、吴云

飞、刘富春、高宝林、葛凌枫、赵朝朝、贾凯、殷宗琨师兄，你们无论是在科研还是在生活上都很照顾我，给予了我莫大的帮助与支持。感谢课题组的小伙伴，陶秀文、黄志远、刘鹏、周阳、陈金城，很开心有缘与你们一起在这度过研究生生涯，一起成长与进步。感谢师弟师妹们的关心和帮助，才使得我的研究生日常不那么枯燥。初见乍惊欢，久处亦怦然。感谢女朋友小褚同学的陪伴与支持，她的理解与鼓励让我勇往直前。她的关心与照顾，让我的生活充满温暖，她的存在让我的研究生生涯更加美好。感谢一路走来曾给予我温暖的人们和一直支持鼓励我的朋友们，谢谢！

谁言寸草心，报得三春晖。衷心感谢我的家人，他们一直是我人生旅途中最坚实的后盾。尽管父母不能时常陪伴在我身边，但离家求学的日子里，我始终铭记着他们的养育之恩。感谢父母一直以来对我的关心和期望，正是他们的支持和鼓励，让我坚定了继续深造、考取研究生的决心。他们的爱是我前行的动力，也是我不断追求卓越的源泉。

岁月匆匆留旧痕，前程似锦待新篇。始终相信努力与真诚，永远不缺再出发的勇气与信心，奔赴下一个山海。再见了，安徽工程大学！再见！

尚德敏学 唯实惟新

