

分类号：TP391

密 级：公开

学校代码：10363

学 号：2220320131



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 基于深度学习的遥感图像常见地物目标
检测与识别

论 文 作 者	<u>王喜辰</u>
校 内 导 师	<u>诸志龙</u>
校 外 导 师	<u>郝旭耀</u>
专业学位类别	<u>电子信息</u>
研究方向（领域）	<u>图像处理</u>

论文提交日期：2025年5月16日

分类号：TP391

单位代码：100363

密类级：公开

学位代号：2220320131

题 目 基于深度学习的遥感图像常见地物目标检测与识别

题 目 Recognition of Common Ground Objects in Remote Sensing Images Based on Deep Learning

论文作者：_____王喜辰_____

校内导师：_____诸志龙_____

校外导师：_____郝旭耀_____

专 业：_____电子信息_____

研究方向：_____图像处理_____

论文答辩日期：_____2025 年 5 月 12 日_____

安徽工程大学学位论文原创性声明

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

学位论文作者签名：王喜辰

日期：2025 年 5 月 17 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：王喜辰

指导教师签名：诸志书

日期：2025 年 5 月 17 日

日期：2025 年 5 月 17 日

基于深度学习的遥感图像常见地物目标检测与识别

摘 要

随着科学技术的不断发展,遥感图像目标检测逐渐成为计算机视觉中重要的研究方向,其在国防、灾害监测、生态保护等领域具有重要的应用价值。本文基于深度学习的方法,研究了两种改进的遥感图像常见地物目标检测算法,主要研究工作如下:

(1)在遥感图像目标检测中,数据质量对模型性能有重要影响。本文针对遥感影像中数据格式不匹配、目标数量不足以及分布不均等问题,开展了包括数据格式标准化转换、裁剪、平移等在内的数据预处理与增强工作,为后续改进算法的训练和评估提供了坚实基础。

(2)针对遥感目标检测任务中由于误检、漏检等现象导致的检测精度低的问题,设计了一种改进的 Faster RCNN 的遥感图像目标检测算法。首先,将原始的 VGG16 主干网络替换成 ResNet50,并在其中引入 CBAM 注意力机制。其次,在原始网络中引入特征金字塔网络,提升对不同尺度目标特征的提取能力。最后,在 NWPU VHR-10 数据集和 MAR20 数据集中进行实验验证,其结果如下:本文模型的 mAP@0.5,在 NWPU VHR-10 数据集上,达到了 86.66%,与基准模型相比提升了 10.69%,在 MAR20 数据集中,达到了 97.93%,提升了 3%。实验结果表明,ResNet50 网络与 CBAM 模块结合可以提升有效特征的关注度,从而有效缓解了遥感图像目标检测任务中误检现

象。此外，特征金字塔网络的加入提高了对不同尺度的遥感目标的检测能力，减少了漏检现象。

(3) 虽然改进的 Faster RCNN 算法在检测精度上得到了提升，但其仍是基于双阶段目标检测框架，计算复杂度高，资源消耗大，难以满足实际应用的需求。为此，本文还选取了更为轻量和高效率的 YOLOv5s 算法，并结合遥感图像特点对其进行优化，减少计算资源消耗的同时，减少误检和漏检的现象，进一步提升检测精度。具体改进包括：首先，在主干网络中引入 CBAM 模块以及 Swin-Transformer 模块。其次，在颈部网络中采用加权双向特征金字塔结构。最后，将损失函数替换为表现更加优秀的 α -IoU。在计算量以及参数量方面，改进后 YOLOv5s 算法相对于改进的 Faster RCNN 算法的计算量以及参数量减少了 127.32GFLOPS 和 22.4M。在检测精度方面，本文模型的 $mAP@0.5$ ，在 NWPU VHR-10 数据集上，达到了 87.09%，与基准模型相比提升了 3.7%，在 MAR20 数据集中，达到了 98.01%，提升了 0.9%。实验结果表明，改进的 YOLOv5s 算法有效降低了计算复杂度以及参数量。同时，在网络性能方面，CBAM 和 Swin-Transformer 模块增强了特征关注和语义理解，这减少了检测任务中的误检现象；加权双向金字塔结构提升了复杂背景下的检测能力，降低漏检现象；而 α -IoU 函数优化了框回归精度，全面提升了目标定位准确性。

关键字：遥感图像，地物目标，深度学习，Faster RCNN，YOLOv5s

DETECTION AND RECOGNITION OF COMMON GROUND OBJECTS IN REMOTE SENSING IMAGES BASED ON DEEP LEARNING

ABSTRACT

With the continuous development of science and technology, remote sensing image object detection has become an important research direction in computer vision, with significant applications in national defense, disaster monitoring, and ecological protection. This study explores two improved remote sensing image object detection algorithms based on deep learning. The main contributions are as follows:

(1) In remote sensing image object detection, data quality has a significant impact on model performance. This study addresses issues such as insufficient target quantity, uneven distribution, and inconsistent data formats in remote sensing imagery by conducting a series of data preprocessing and augmentation procedures, including standardized format conversion, image cropping, and translation. These efforts lay a solid foundation for the subsequent training and evaluation of the improved detection algorithm.

(2) To address low detection accuracy due to false positives and false

negatives, an improved Faster RCNN algorithm is proposed. First, the original VGG16 backbone network was replaced with ResNet50, and the CBAM attention mechanism was introduced. Second, a Feature Pyramid Network (FPN) was integrated into the original network to enhance the ability to extract features of objects at different scales. Finally, experiments were conducted on the NWPU VHR-10 and MAR20 datasets, with the following results: The $mAP@0.5$ of the proposed model reached 86.66% on the NWPU VHR-10 dataset, an improvement of 10.69% compared to the baseline model, and 97.93% on the MAR20 dataset, with a 3% increase. The experimental results indicate that combining the ResNet50 network with the CBAM module can enhance the focus on effective features, thereby effectively reducing false positives in remote sensing image object detection tasks. Additionally, the introduction of the FPN improved the detection capability for objects of different scales, reducing false negatives.

(3) Although the improved Faster RCNN algorithm achieved higher detection accuracy, it is still based on a two-stage object detection framework, which has high computational complexity and large resource consumption, making it difficult to meet the demands of practical applications. Therefore, the lightweight and efficient YOLOv5s algorithm was selected and optimized for remote sensing image characteristics to reduce computational resource consumption while further reducing false

positives and false negatives and improving detection accuracy. The specific improvements include: First, the CBAM module and Swin-Transformer module were introduced into the backbone network. Second, a weighted bidirectional feature pyramid structure was adopted in the neck network. Finally, the loss function was replaced with the more effective alpha-IoU. In terms of computational and parameter complexity, the improved YOLOv5s algorithm reduced the computational load and parameter count by 127.32GFLOPS and 22.4M, respectively, compared to the improved Faster RCNN algorithm. In terms of detection accuracy, the mAP@0.5 of the proposed model reached 87.09% on the NWPU VHR-10 dataset, a 3.7% improvement over the baseline model, and 98.01% on the MAR20 dataset, with a 0.9% increase. The experimental results show that the improved YOLOv5s algorithm is efficient in computational complexity and resource consumption. The CBAM and Swin-Transformer modules enhance feature attention and semantic understanding, reducing false positives. The weighted bidirectional pyramid structure improves detection in complex backgrounds, lowering false negatives. The alpha-IoU optimizes bounding box regression accuracy, enhancing target localization and overall performance.

KEY WORDS: Remote Sensing Image, Ground Object, Deep Learning, Faster RCNN, YOLOv5s

目 录

摘 要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 背景与意义	1
1.2 国内外研究现状	2
1.2.1 深度学习研究现状	2
1.2.2 目标检测研究现状	3
1.2.3 遥感图像目标检测研究现状	5
1.3 论文研究内容	7
1.4 论文的结构安排	8
第 2 章 基于深度学习的目标检测相关理论	10
2.1 卷积神经网络基础	10
2.1.1 卷积层	10
2.1.2 池化层	10
2.1.3 激活函数	11
2.1.4 全连接层	13
2.2 基于卷积神经网络的目标检测算法	14
2.2.1 单阶段的目标检测算法	14
2.2.2 双阶段的目标检测算法	18
2.3 本章小结	21
第 3 章 遥感影像数据分析与预处理	22
3.1 遥感影像数据分析	22
3.2 遥感影像数据预处理	25
3.3 目标检测常用的评价指标	26
3.4 本章小结	27
第 4 章 基于改进的 Faster RCNN 的遥感图像常见地表目标检测算法	28
4.1 引言	28
4.2 改进的 Faster RCNN 模型	28
4.2.1 残差网络模块	28

4.2.2 注意力机制的加入	31
4.2.3 特征金字塔网络模块	33
4.2.4 改进后的网络结构	35
4.3 实验结果与分析	35
4.3.1 实验平台	35
4.3.2 实验设计	35
4.3.3 消融实验结果与分析	36
4.3.4 对比实验结果与分析	41
4.4 本章小结	43
第 5 章 基于改进的 YOLOv5s 的遥感图像常见地表目标检测算法	44
5.1 引言	44
5.2 改进的 YOLOv5s 模型	45
5.2.1 主干网络的改进	45
5.2.2 BiFPN 网络的引入	47
5.2.3 损失函数的优化	47
5.2.4 改进后的网络结构	48
5.3 实验结果与分析	49
5.3.1 实验平台	49
5.3.2 实验设计	49
5.3.3 消融实验结果与分析	50
5.3.4 对比实验结果与分析	54
5.4 本章小结	57
第 6 章 总结与展望	58
6.1 总结	58
6.2 展望	59
参考文献	61
攻读学位期间的学术成果	68
致 谢	69

第 1 章 绪论

1.1 背景与意义

随着遥感技术的快速发展^[1]，遥感影像在国土资源管理、城市规划^[2]、环境保护^[3]、军事监控等多个领域得到了广泛的应用。遥感图像凭借着其覆盖范围广、时效性强以及多维度信息丰富等特点，已成为获取地表信息的重要手段。然而，对遥感图像目标的检测一直面临诸多复杂的挑战。一方面，遥感图像中目标种类繁多，包括建筑物、车辆、船舶等，这些目标在形状、大小、纹理及光谱特征上差异明显；另一方面，由于遥感图像的拍摄视角多样（如卫星、无人机等）、分辨率差异，以及光照变化、背景复杂等因素的干扰，使得遥感目标检测的精度和鲁棒性难以保障。

遥感图像目标检测在国防、灾害监测、生态保护等领域具有重要的应用价值。通过提升检测精度，可以更准确地获取地表信息，为科学决策提供有力支撑。随着应用需求的不断提升，遥感目标检测展现了巨大的发展潜力，对该领域的研究已经不仅仅局限于提高精度，还需要充分考虑检测方法的适用性以及鲁棒性。

传统的目标检测方法依赖于人工设计特征^[4]，如纹理、边缘、颜色等，但面对复杂的遥感场景时，其特征提取能力有限，难以实现对目标的准确定位和分类。随着深度学习技术^[5]的飞速发展，基于深度学习的目标检测方法展现出了显著优势。其强大的特征学习能力，不仅能够处理复杂背景，显著提升检测的鲁棒性与适应性，克服传统算法的局限，还能自动提取高维特征，为遥感图像的目标检测带来了新的突破。

近年来，遥感图像目标检测领域在深度学习领域的发展上取得了显著进展。两阶段检测方法以 Faster RCNN^[6]为代表，通过候选区域生成与分类回归分开进行，有效提升了检测精度。后续如 HOFA-Net^[7]、HODet^[8]等进一步优化了训练策略，显著改善了检测性能。另一方面，一阶段检测器以 YOLO 系列为代表，强调检测速度与端到端训练优势。REI-YOLOv8^[9]、YOLOv9^[10]通过结构改进与损失函数优化，进一步提升了小目标检测精度与模型泛化能力。这些研究不仅推动了

检测器在自然图像领域的应用，也为遥感图像中多尺度、小目标复杂场景下的检测任务提供了有力支持。

因此，对基于深度学习的遥感图像目标检测的研究，不仅具有理论研究意义，还能在实际应用中解决现实问题，助力遥感技术的发展与普及，推动国民经济和社会建设的高质量发展。

1.2 国内外研究现状

1.2.1 深度学习研究现状

深度学习是一种基于人工神经网络的机器学习方法^[11]，随着计算能力、数据获取能力以及算法创新等的共同推动下，深度学习取得长足的发展，其在计算机视觉^[12]、自然语言处理^[13]、自动驾驶领域^[14]等多个不同的领域都有着出色的表现。深度学习技术的发展逐渐成为了推动人工智能发展的核心驱动力。

深度学习的核心部分便是人工神经网络，早期研究者们通过对大脑中神经网络工作原理进行分析，并通过数学建模，使得计算机能够如同人脑一样可以对数据进行处理。感知机模型是最早的神经网络模型之一^[15]，该模型的提出主要是解决线性分类问题。在神经网络中单个感知机又被称作神经元，多个神经元之间进行链接并经过非线性单元后形成神经网络就可以对非线性问题。而多个神经元进行堆叠优化形成的深度神经网络就可以处理更加复杂的任务。

深度学习与传统的机器学习相比有着显著的特点，深度学习实现了端到端的映射，避免了传统深度学习复杂的过程，并且通过非线性变换在处理复杂任务时有着更加优秀的表现。目前常见的深度学习的模型主要包括卷积神经网络^[16]、循环神经网络^[17]以及生成对抗网络^[18]等，其中卷积神经网络主要通过局部感知以及权值共享机制对数据特征进行提取，其主要由卷积层、激活函数、池化层、全连接层以及损失函数等组成，在图像分类、目标检测和图像分割等方面的表现十分优异。循环神经网络则是通过在网络结构中引入循环连接，主要由输入层、隐藏层以及输出层组成，其核心是当前输出不仅与当前输入有关，还与前一时间状态相关。循环神经网络由于可以捕捉数据中的时间顺序以及上下文关系，使得其在自然语言处理、语音处理以及时间序列分析等问题中表现良好。而生成对抗网

络的核心则是通过博弈论中的对抗机制训练生成器和判别器,通过对抗交替优化生成器与判别器,使生成效果逐渐接近真实数据。该网络主要用于生成高分辨率图像,以及数据增强等。

目前深度学习作为一种强大的数据处理技术,已经对多个领域产生不同程度的促进作用。然而,其依旧存在对数据依赖性高、计算资源消耗巨大和较差的鲁棒性等问题。未来,对深度学习的研究依旧是一个热门的课题,将更加模型的高效性、适用性和可靠性。

1.2.2 目标检测研究现状

目标检测作为计算机视觉领域的重要任务之一,近年来伴随着科学技术的发展其也在向着更快速、更准确、更轻量化的方向发展。目标检测任务的内容通常分为目标的检测任务以及识别任务,对于检测任务需要完成对物体大小、位置的确定,而识别任务则是要确定物体种类,对目标进行分类。对于目标检测而言,无论是被检测目标自身外观、尺寸,还是所处环境的不同等因素,都会对检测的准确度产生影响。

目标检测的发展主要经过了两个时期,分别为传统目标检测时期^[19]以及基于深度学习的目标检测时期。传统目标检测算法主要依赖于手工特征提取以及传统机器学习算法,其主要流程包括特征提取、候选框生成以及分类检测等,其中特征提取主要是通过手工设计的特征提取图像中的关键信息,如 SIFT 检测器^[20]以及 HOG 检测器^[21]等。候选框生成的目的是确定图像中可能包含目标的区域,该步骤是通常是通过窗口滑动以及区域建议的方法实现,而分类检测主要是实现目标识别与定位,通过训练好的分类器对候选区域进行判断,得到初步检测结果,之后利用非极大值抑制等操作去除冗余检测框,最后输出最终结果。然而,随着计算能力的发展,传统目标检测算法由于其计算复杂度高和鲁棒性差等因素,逐渐被深度学习算法所取代。

目前,基于深度学习的目标检测算法主要被分为单阶段目标检测算法和双阶段目标检测算法两种^[22]。其中单阶段检测算法主要是通过卷积神经网络直接对目标特征进行提取,预测目标类别与位置,这是一种端到端的操作,省略了复杂的生成候选区域的步骤,这种算法的优点在于检测速度快,但检测精度相对较低。

而两阶段目标检测算法首先由区域建议网络生成一系列候选区域,其次是对候选框进行分类与边界框回归,通过这两步操作使得该算法具有较高的检测精度以及可靠性,但由于其存在候选区域生成这一步骤,使得算法的检测速度较低。

2012 年 AlexNet 分类网络^[23]的提出,使得研究者们开始探索将深度学习领域应用在目标检测任务等更加复杂的场景。2014 年,由 Girshick 等人提出的区域卷积神经网络 RCNN^[24],正式将深度学习应用在目标检测领域,通过将卷积神经网络、线性回归器以及支持向量机^[25]等算法的结合,实现了图像中目标自动定位与分类。2015 年, Girshick 等人通过改进 RCNN 算法,提出了 Fast RCNN 算法^[26],利用神经网络替代支持向量机来进行分类预测,提升检测的精度与速度。同年,在 Fast RCNN 的启发下, Ren 等人又提出 Faster RCNN 算法,其核心是区域建议网络的提出,该网络可以与检测网络共享卷积特征,进一步提升了网络的检测精度与速度。在 2016 年, liu 等人提出了 SSD 检测算法^[27],该算法通过利用不同特征层来检测不同尺寸的目标,提高了模型的泛化能力的同时,结构简单,检测速度高。同年, Redmon 等人提出 YOLO 算法^[28],这是一种端到端的算法,将目标检测任务转化为一个回归问题,通过单个神经网络直接在图像上进行推理,实现实时检测目标,该种算法快速、高效等优点,使得该算法在目标检测任务中被广泛应用。随着深度学习的快速发展,目标检测领域也在飞速发展。

Zhang Y 等^[29]人通过更改特征提取网络的同时,使用 K-means 聚类算法对边界框进行聚类,根据聚类结果改进锚定,既可以提高目标识别的准确性,又可以准确定位目标的位置。Zou X 等^[30]人提出另一种改进的 Faster RCNN 目标检测方法,采用 ResNet50 网络作为特征提取网络,并重新设计损失函数与极大值抑制函数,这不仅缩短了检测时间,还提升了检测精度。马宇等^[31]人提出一种改进的 Faster RCNN 目标检测方法,通过将 RestNet-50 作为主体网络并在 K-means 算法的基础上对聚类点进行均值优化,虽然相对于原始算法的性能提升较为明显,但从整体看,精度仍然较低。Tan L 等人^[32]提出 SP-SSD 算法,该算法将原始的 SSD 算法进行改进,采用深度可分离卷积进行下采样,同时引入 ECM 模块用于特征提取,并引入残差连接以及加入 SPA 模块,改进后的网络精度得到提升,参数量减小,检测速度也得到了提升。孙仁科等^[33]人提出基于一种改进的 SSD 目标

检测算法,将原始主干网络替换为轻量化网络 **PeeNet**^[34],并加入特征融合模块,减少网络的复杂程度,同时提升了网络对弱小目标的检测能力。**Karthikeyan M** 等^[35]人改进了原始的 **YOLOv3**^[36]网络,通过在检测层中加入空间金字塔和 **swish** 激活函数,提升了网络对不同类型目标的识别能力。**Tang S** 等^[37]人根据 **YOLOv5** 网络提出了一种改进方法,添加了一个特定于小目标的检测头,并在骨干网络与颈部网络之间加入一个反演模块增加特征图的通道信息,同时加入注意力机制模块,提升对关键特征的关注度。**王林枫**等^[38]人提出一种关于 **YOLOv4-tiny**^[39]的改进算法,该算法修改原始卷积模块并在网络中加入 **PANity** 模块,同时利用所提出的 **CSPConcat** 模块对网络进行优化,降低了检测的漏检率和误检率。**Gao W** 等^[40]人提出了 **YOLO V5-EFFICIENT** 算法,该算法在 **YOLOv5** 的基础上利用改进的 **K-means** 算法生成大小更加合适的初始锚点,并加入注意力机制模块使特征提取层更多的关注感兴趣区域,改进损失函数并改进原始网络的 **SPPF** 模块,对网络进行优化并确保检测的实时性。

如今,科技的发展正推动着深度学习飞速发展,研究者们对算法的改进从未停止,这使得目标检测算法不仅在检测精度与实时性上有所提升,同时在应对更多复杂场景中都能够展现出很好的检测效果。

1.2.3 遥感图像目标检测研究现状

遥感影像数据集相对于普通数据集而言,有着不同的特点。这些数据集通常是通过卫星或无人机在高空中拍摄得到的,拍摄高度从几百米到几千米不等,这导致图像中目标的大小差异显著,给目标检测带来了额外的难度。此外,由于遥感图像多采用俯视角度拍摄^[41],这常常使得那些在普通数据集中表现出色的算法在遥感影像数据集上的效果不尽如人意。在一张遥感图像中,我们可能会发现几个甚至几十个目标,这些目标的大小通常各不相同,这对检测算法来说无疑是一项挑战。在实际检测环境中,遥感图像中的目标往往处于一个非常复杂的背景之中,不仅可能面临目标被遮挡的问题,还因为图像视野广阔而包含了多种多样的背景元素,这些都可能对目标检测造成干扰。

在早期,对遥感图像目标检测任务主要是通过传统算法进行的,由于计算资源有限,只能通过手工设计特征实现对特定目标进行检测,但这种方法的不仅资

源消耗大、效率低，还由于遥感图像自身的特点导致实际的检测效果极差。随着科学技术的发展，深度学习在目标检测领域的飞速发展，为遥感图像目标检测带来了新的思路。为此，研究者在基于深度学习的遥感图像目标检测领域做出大量改进，为该领域的发展做出了巨大贡献。

Shi L 等^[42]人提出的一种改进的 Faster-RCNN 检测算法，将 EfficientNetb0 与 FPN 结合，使得原始网络不仅减少了模型大小，还提升了网络对遥感图像目标检测的性能。另外，凌晨等^[43]人优化 mask-RCNN^[44]算法，通过调节对目标边缘的学习以及参数的调整，从而获得更加丰富的船舶边缘信息，但该算法的适用性以及误检率仍然相较于其他算法来说并不理想。Xu S 等人^[45]提出改进的 YOLOv7 算法^[46]，在原始网络中加入小目标检测头，增加网络对目标的识别能力，然后再骨干网络融入 Transformers 模块^[47]，增强特征提取能力，之后加入空间通道注意力机制^[48]模块，以及更改极大值抑制函数为 DIoU-NMS^[49]，提升网络对小目标的检测能力。黄瑶^[50]提出基于改进 YOLOv5 算法的遥感目标检测方法，将原始网络与 MobileNetV2^[51]网络进行结合，并使用多路径融合网络，这样做不仅减少了原始网络的复杂程度，而且增强了网络多尺度特征融合能力，但其对于部分目标定位的精准度依旧有待提升。张亮^[52]则提出另一种改进 YOLOv5 的算法，对所使用的数据集进行 Copy-paste 数据增强，同时将原网络中的 C3 模块删除，使用 C3Ghost 模块进行代替。该方法在提升准确度以及减少计算量等方面存在一定优势，但该方法在适用性等方面也存在一定局限。Xu X 等^[53]人提出 ESOD-YOLO 算法，该算法针对航拍小目标进行检测，在原始算法中集成了可变形和全局-局部特征引导（DGLFG）、简单跨尺度和跨层特征融合（SCCFF）和基于部分卷积的逆残差共享参数（ISPP）三个模块，提升检测精度的同时减少了参数量。张新君等^[54]人也提出了一种改进 YOLOv5 算法的方法，其在原始算法的基础上加入不同的注意力机制模块，同时简化了骨干网络，减少网络的训练时间，提升了对小目标检测的精度，但其在漏检和误检方面依旧有优化的空间。为缓解在遥感图像中小目标检测任务带来的问题，Gao T 等^[55]人提出 MSNet 算法，设计了 P²CEM 将逐点卷积与部分卷积进行结合减少参数量，其次，集成上下文建模和残差模块，增强小目标的特征，最后改进的上采样运算符和完全连接的 FPN 减

少网络参数。陈欣等^[56]人提出采用多尺度特征融合 SSD 的遥感图像小目标检测的方法,在 SSD 算法的基础上,设计了特征融合机制并加入通道注意力机制^[57],使得其对小目标检测的精度有所提升,但同时检测的速度也有所下降。王玺坤等^[58]人提出基于 YOLOv3-tiny 的算法改进策略,在原始算法上增加了特征映射模块,并引用了残差网络,使得对遥感图像的精度提升明显,但对于模糊目标的检测精度及准确率较低。李志昂等^[59]人针对遥感图像中小目标船舶提出了一种改进的 YOLOv5s 算法,该算法引入了适应空间特征融合模块以及 BOTNet 注意力机制,同时修改原始损失函数为 EIoU,最后加入 Slim-neck 结构对网络整体进行轻量化,这样做不仅提升了目标检测的精度,还减少了原始模型的参数量以及计算量,同时权重大小也得到了改善。

1.3 论文研究内容

本文以遥感图像常见地表目标为检测对象,针对遥感图像中目标的特点,利用深度学习算法实现目标检测任务,同时从多个方面提升目标检测的准确度。本文的主要工作如下:

(1) 对遥感图像目标检测数据集进行分析与预处理,为解决由于数据集样本导致使用算法的检测效果差的问题,同时提高算法的泛化性与鲁棒性,本文使用数据增强的方法来扩充数据。通过设置多组实验来验证改进算法的性能。

(2) 为提高遥感图像目标的检测精度,本文对原始 Faster RCNN 算法进行改进。首先将原始网络使用的 VGG16 网络^[60]替换为 ResNet50 网络^[61],这样做的目的是增强了深层特征的提取能力从而提高检测精度。其次,对 ResNet50 网络进行改进,在网络中加入注意力机制模块,提升算法对遥感图像目标特征的关注度。最后,针对遥感图像目标尺度不一导致原始算法性能下降的问题,本文在网络的主干部分引入特征金字塔网络^[62],提升网络多尺度融合能力。

(3) 为满足实际检测的需求,减少计算参数以及复杂程度,同时进一步减少误检和漏检现象。本文还研究了另一种基于改进 YOLOv5 的遥感图像目标检测算法。首先,将主干网络中的 C3 模块进行改进,加入与改进 Faster RCNN 算法相同的注意力机制模块,这样做不仅解决了原始算法对遥感图像目标关注度不

足的问题，还检验了所采用注意力模块的高效性以及适用性。其次，针对遥感图像像素低以及目标密集的问题，本文在算法中加入 Swin-Transformer 模块^[63]，提升算法局部与全局特征提取能力。之后，本文借助加权双向特征金字塔^[64]的思想改进网络结构，融合不同特征层，提升网络的多尺度特征融合能力。最后，为提升目标框的准确度以及加快收敛速度，本算法更改 Loss 函数为 alpha-IoU^[65]，提升算法的整体性能，使得改进后的算法精度进一步得到提升。

1.4 论文的结构安排

本文的组织架构为：

第一章为绪论部分，该部分首先阐述了遥感图像目标检测的背景与意义，其次，对该领域相关的国内外研究现状做了分析与研究，最后，对本文的研究内容进行了简要阐述以及对本文的结构安排做出了具体规划。

第二章为基于深度学习的目标检测相关理论，该部分主要介绍了构成神经网络基本模块的相关概念，包括卷积层、池化层、激活函数等。之后对单阶段目标检测以及双阶段目标检测算法中几种代表算法做了具体介绍，分析了其结构，为后续改进打下坚实的理论基础。

第三章为数据分析与处理，该部分主要介绍了本文所使用的西北工业大学遥感影像数据集^[66]以及 MAR20 飞机遥感影像数据集^[67]，并在训练之前对数据集进行预处理，增加数据的丰富性，提高算法的鲁棒性，之后介绍了目标检测常用的检测指标用于评估改进模型的性能。

第四章为基于改进 Faster RCNN 的目标检测算法部分，该部分对现有的 Faster RCNN 算法进行改进，首先，将 VGG16 网络替换为 ResNet50 网络，减少计算量，增加检测速度与精度，并加入空间通道注意力机制模块，提高网络对目标特征的关注度。其次，引入特征金字塔网络，增加网络多尺度融合能力。最后在两种数据集上对结果进行验证。

第五章为基于改进 YOLOv5 的目标检测算法部分，该部分提供了另一种目标检测方法，在 YOLOv5 原始算法的基础上，对主干网络进行改进，引入 Swin-Transformer 模块，增加全局与局部特征提取能力。其次在加权双向特征金

字塔网络思想的启发下,修改颈部网络结构,提升网络多尺度融合的能力。之后为提升网络对目标特征的关注度,同样加入空间通道注意力机制,并通过更改 Loss 函数,来提升网络的整体性能。最后,依旧在两个数据集中对结果进行验证。

第六章为总结与展望,该部分总结了本文的全部工作内容,并对算法存在的问题进行分析,提出下一步改进的方案。

第2章 基于深度学习的目标检测相关理论

2.1 卷积神经网络基础

2.1.1 卷积层

卷积层是深度学习中卷积神经网络的核心组成部分之一，它主要用于处理具有网格结构的数据，例如图片等。卷积层主要由多个卷积核组成，而卷积核通常是一个小型矩阵，该矩阵在输入图片上滑动以检测特定的特征。对于卷积层来说不同的卷积核，其取得的特征也不相同，例如输出特征图的深度主要由卷积核的数量所决定，而卷积核的大小则会影响感受野的大小。同样的，对于滑动的步长则决定了卷积核在输入数据上滑动的间隔。卷积层具体的运算步骤如图2-1所示。

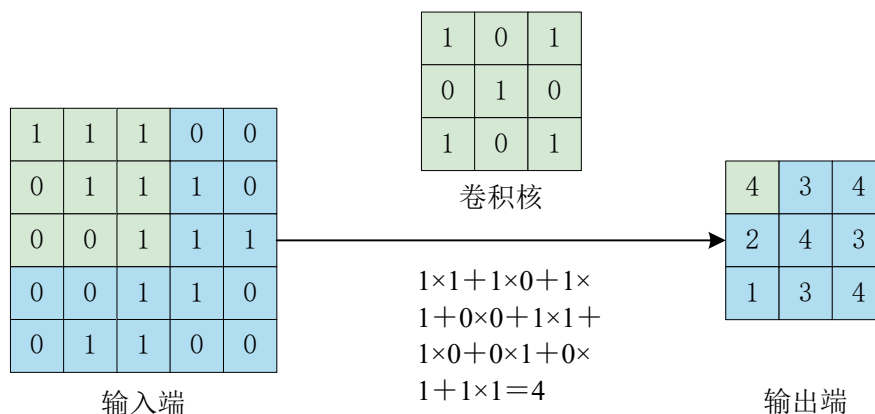


图 2-1 卷积运算示意图

以一个常见的 3×3 卷积核为例，通过在输入图片中以步长为 1 进行滑动，其具体的卷积运算过程就是用卷积核对图像进行覆盖，将覆盖的区域中的元素与对应位置卷积核的元素进行相乘，之后对所有相乘后的结果进行求和，最后得到输出特征图对应位置的特征值。

2.1.2 池化层

池化层，也称下采样层，在卷积神经网络中非常常见，该层主要的作用是降低特征图的尺寸，减少后续层的参数量以及计算量，同时对特征提取以及模型的平移不变性都有重要的作用。对于常用的池化操作主要有两种，一种为平均池化，

该池化操作是计算给定感受野内所有特征值的平均值，并将该值作为输出，该方法可以捕捉特征图局部区域的整体信息。另一种为最大池化，该池化操作是在给定局部感受野中取最大特征值作为最后输出，该方法可以捕捉局部区域中最明显的特征信息。

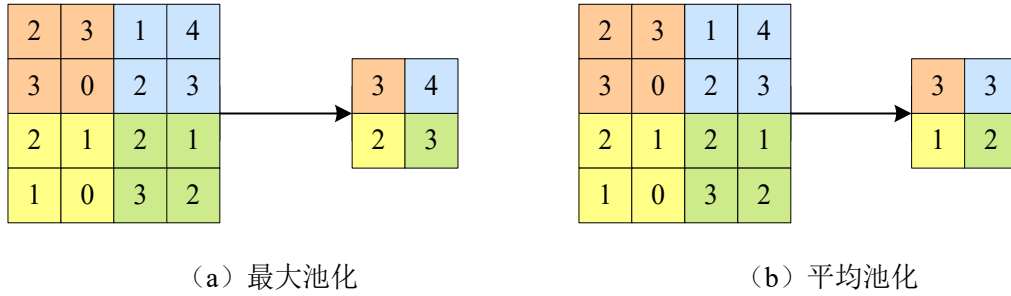


图 2-2 最大池化与平均池化示意图

图 2-2 为两种池化操作的示意图，通过图中信息可以发现，池化操作主要与窗口的大小有关，即给定感受野的大小，同时池化窗口在特征图上滑动间隔大小也会对池化后的特征图产生影响。

2.1.3 激活函数

激活函数是神经网络中不可或缺的组成部分，它的作用是在网络中加入非线性因素，使得神经网络可以获得学习和模拟复杂函数映射的能力。在未加入激活函数前，神经网络的输入与输出都是线性的，这极大抑制了网络的表达能力。在神经网络中，激活函数通常位于神经元的输出部分，具体来说，就是在神经网络的加权输入和偏置项之后，输出结果传递到下一层之前的位置。常用的激活函数有多个，例如 Sigmoid 函数、ReLU 函数、tanh 函数等，每个函数在不同的网络结构和任务中表现可能会有所不同，因此在实践中可能需要尝试不同的激活函数来找到最佳选择。以下是对部分激活函数及其特点的介绍。

(1) Sigmoid 函数

Sigmoid 函数是在神经网络中最先使用的激活函数，其函数表达式如下式(2-1)所示：

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

其图像如下图 2-3 所示：

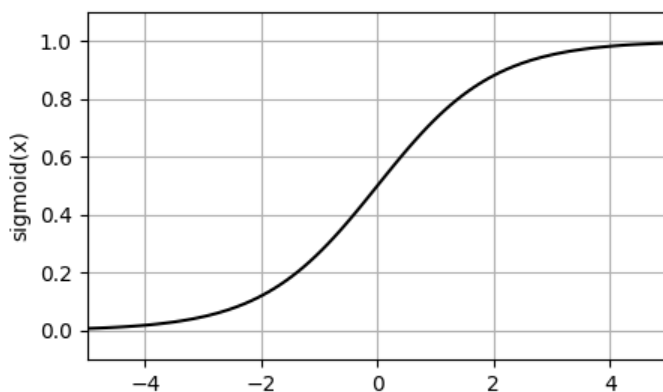


图 2-3 Sigmoid 函数示意图

由其公式和图像可以看出，Sigmoid 函数整体呈现单调递增的状态，并且当输入接近负无穷和正无穷时输出分别接近 0 与 1，这意味着可以被解释为概率值，这使得其在二分类问题中作为输出层的激活函数将非常合适。同时，其导数是平滑的，导致梯度不容易发生突变，提高了网络的稳定性。然而，可以发现当输入很小或者很大时，其梯度几乎为零，这可能会导致梯度消失的情况，使得训练过程中权重更新非常缓慢。由于其需要指数运算，会计算资源消耗较大。

(2) Tanh 函数

Tanh 函数是一种双曲正切函数，其函数表达式如下式(2-2)所示：

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

其图像如下图 2-4 所示：

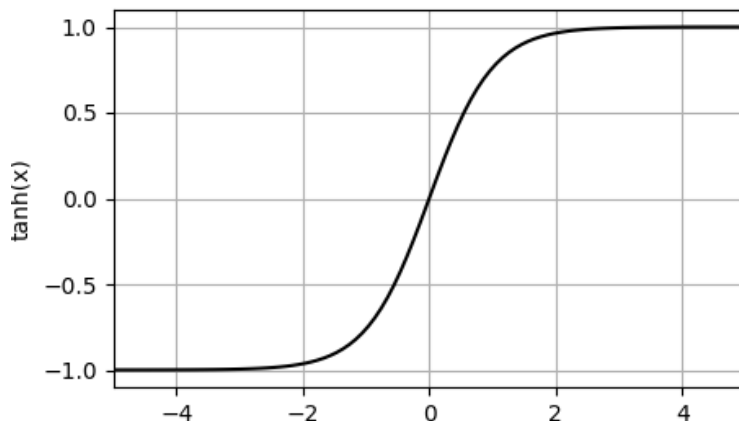


图 2-4 Tanh 函数图像示意图

由其公式和图像可以看出，Tanh 函数与 Sigmoid 函数一样呈单调递增的状态，其输出值在-1 到 1 之间，这意味着其输出为零中心化，这有助于处理数据与网络训练，同时由于其能够表达更广泛的值的范围，使得其在某些任务中有着更好的表达能力。但其依旧没有解决梯度消失以及计算复杂等问题。

（3）ReLU 函数

ReLU 函数是 Krizhevsky、Hinton 等人在 2012 年时提出，其函数表达式如下式(2-3)所示

$$f(x) = \max(0, x) \quad (2-3)$$

其图像如下图 2-5 所示：

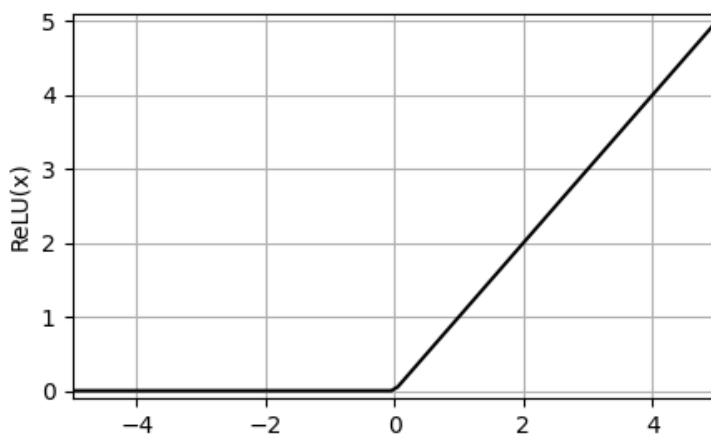


图 2-5 ReLU 函数示意图

由其公式和图像可以看出，当输入为负值时，其输出为零，当其输入大于 0 时，其导数出现恒定值 1。这样的特点导致该激活函数的计算非常简单，使用也更加高效，同时，当输入为正时，恒定的梯度可以很好的解决梯度爆炸的问题。但由于其输入为负值时，输出均为 0，激活函数的导数为 0，这导致神经元的参数不能进行有效更新，同时其部分神经元可能会被永久抑制，不能充分学习特征。

2.1.4 全连接层

全连接层是神经网络的一种基本结构，其位置通常位于网络的后端。该层的主要功能是将之前层提取的特征进行整合，并用于最终的分类或回归任务中。图 2-6 为全连接层的结构示意图。

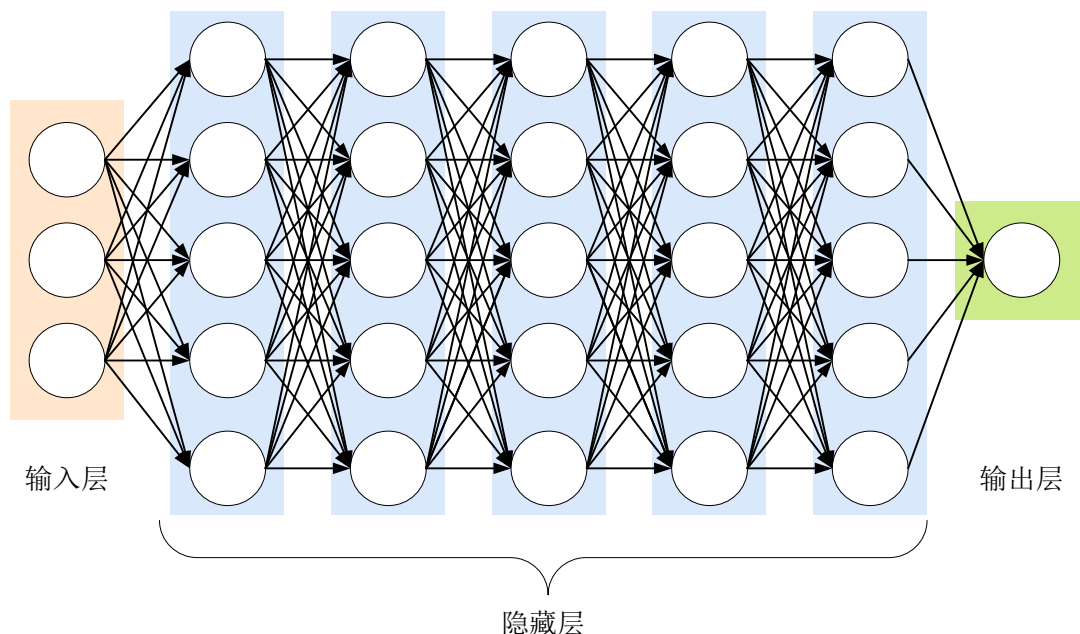


图 2-6 全连接层结构示意图

全连接层大致可以分为三个部分，首先是输入层，该层主要接受之前层提取的特征，其中每一个节点表示一个特征。其次是隐藏层，该层为神经网络中非常关键的部分，该层位于输入层与输出层之间，负责处理和计算信息。它通过全连接的方式捕捉复杂的数据特征，并通过权重和偏置最小化损失函数，同时每个隐藏层神经元应用非线性激活函数来增强网络处理能力。最后为输出层。该层主要是对神经网络预测结果进行输出。由于全连接层的结构特点使得其具有非常强的特征表达能力和较高的通用性，但也存在计算量大、训练成本过高等问题。

2.2 基于卷积神经网络的目标检测算法

2.2.1 单阶段的目标检测算法

(1) SSD

SSD 算法是一种高效的单阶段目标检测算法，其核心思想是通过保留不同尺度的特征图来对不同大小目标进行检测，同时通过调整先验框的大小更好的在不同尺度以及位置上预测目标框。图 2-7 为 SSD 算法的网络结构示意图。

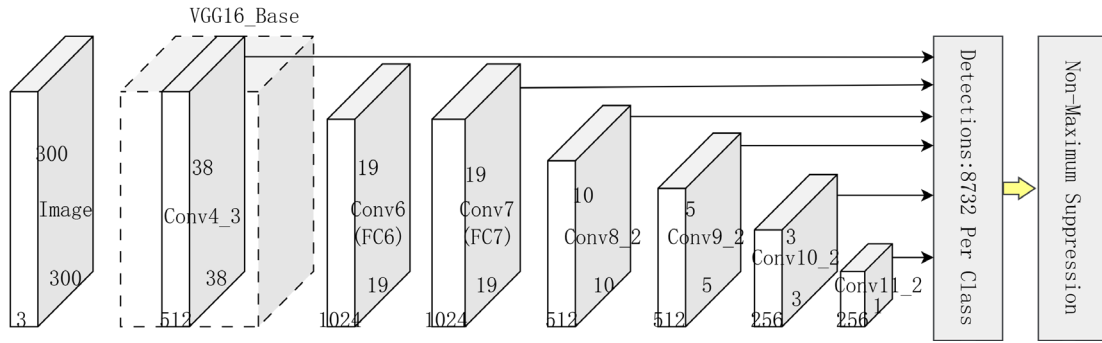


图 2-7 SSD 算法网络结构示意图

SSD 算法以 VGG16 模型作为其特征提取的主干网络，并在该模型上进行改进。首先，算法保留了原始 VGG16 模型中 Conv1 到 Conv4 的所有层，以及 Conv5 的第一个和第二个卷积层，这些层主要是为了高质量特征提取。其次，移除 VGG16 模型中 FC8 全连接层，将 FC6 与 FC7 两个全连接层被转换为卷积层，这样做保持了完整的空间信息。同时对网络中第五个最大池化层 Pool5 进行修改，将卷积核大小从 2×2 更改为了 3×3 ，步长也从 2 改为了 1，这样做是为了保留特征图分辨率。之后，在原先的 VGG16 的基础上增加了多个卷积层，生成多尺度特征图，用于检测不同大小的目标。最后，采用借助了锚点机制以及非极大值抑制函数，对目标进行类别与边界框的预测。该算法不仅提升了网络对不同尺度目标的检测精度，同时在检测速度上也有较大优势。

(2) YOLO

YOLO 算法作为单阶段目标检测算法的代表算法之一，以 YOLO 系列的开山之作 YOLOv1 为例，其核心思想是将图片划分成固定网格，每个网格负责预测目标的位置与类别，同时 YOLO 还将目标检测任务当做单一的回归问题，网络输出既包含了分类信息，也包含了边界框位置信息，实现了端到端检测。图 2-8 为 YOLOv1 的网络结构示意图。

图 2-8 YOLOv1 网络结构示意图

(3) YOLOv5

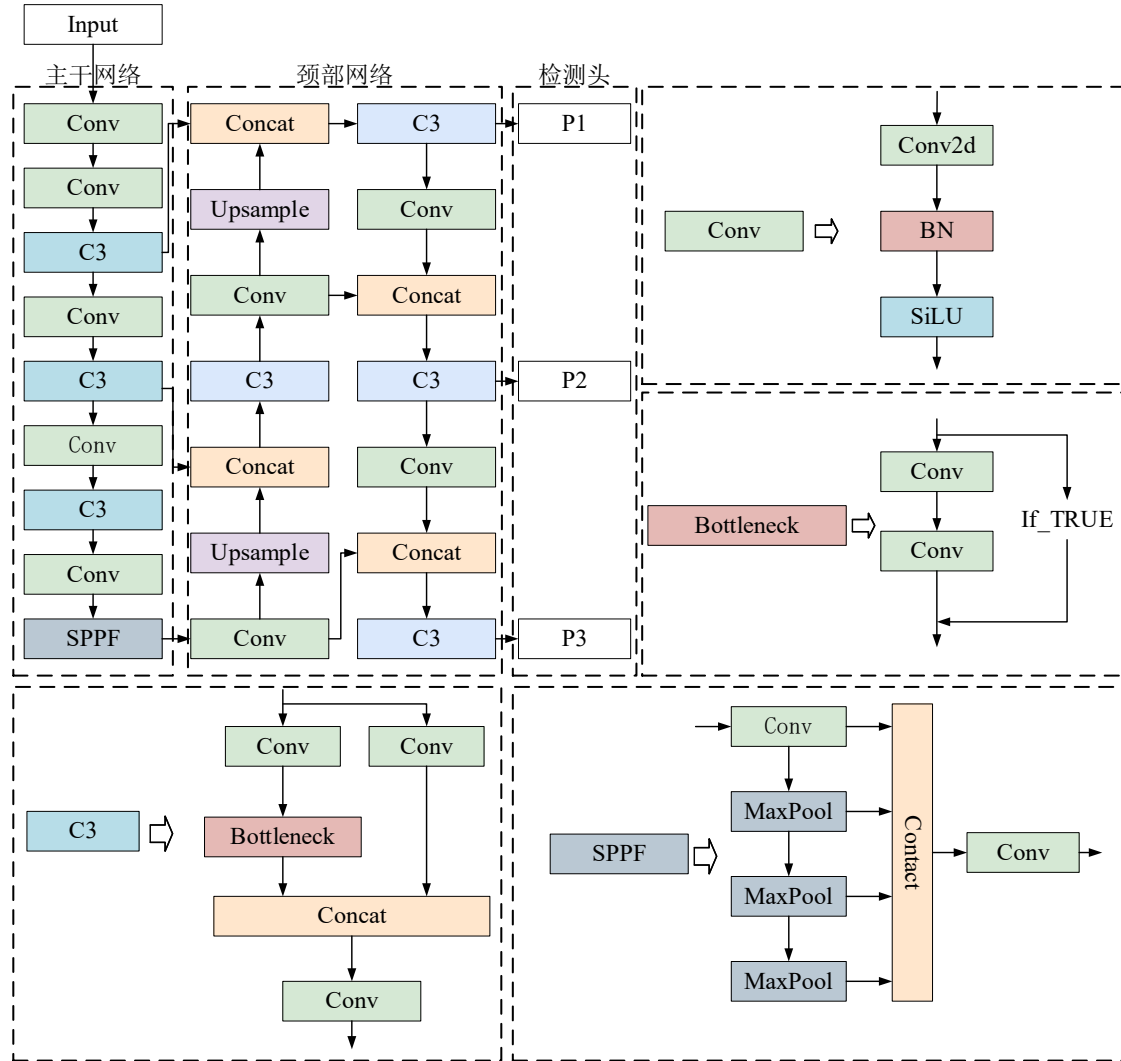


图 2-9 YOLOv5 网络结构示意图

YOLOv5s 的输入部分主要是数据增强、自适应描框计算以及自适应图片缩放。其中数据增强的方法有很多种，比如，MixUp 数据增强、Cutout 数据增强、CutMix 数据增强以及 Mosaic 数据增强，本文为丰富数据集和减少 GPU 用量选用 Mosaic 数据增强。自适应描框计算采用 `kmean_anchors` 函数计算最适合目标的描框，而自适应图片缩放根据不同尺寸的目标对图片进行缩放来适应后续的操作。

YOLOv5s 的主干网络部分主要是对目标的特征进行提取，其主要是以 Conv、C3 和空间金字塔池化（Spatial Pyramid Pooling Fast, SPPF）等基本模块组成，Conv 模块包含了卷积层、激活函数以及增加了 BN 层，C3 模块提高网络提取特征的能力，进一步提升网络的稳定性及检测准确率。而 SPPF 模块则为了融合多

尺度特征，进而丰富语义特征。

YOLOv5s 的颈部网络部分主要是采用了“特征金字塔网络（Feature Pyramid Network, FPN）+路径聚合网络（Path Aggregation Network, PANet）^[68]”结构，其中 FPN 采用自底而上和自顶而下的结构，进行多尺度特征融合。而 PANet 在 FPN 的基础上又添加了一个自底向上的结构以及自适应特征池化层和全连接融合，这样使得网络更加聚焦于目标物体，同时减少了冗余信息的干扰，使得检测更加精准。

YOLOv5s 的 Head 部分主要用来获得网络的输出，检测物体的位置与类别，做出预测。输出部分分为三层，分别为浅、中、深，对应检测小物体、较小物体与大物体。而输出存在的损失函数与非极大值抑制函数则是对网络的优化，进一步使检测的精度有所提升。

2.2.2 双阶段的目标检测算法

(1) RCNN

RCNN 算法是目标检测领域的一个重要的里程碑，它首次将深度学习与传统的目标检测算法相结合，该算法的核心思想是利用候选区域生成与卷积神经网络提取的特征相结合，实现较高精度的目标检测任务。图 2-10 为 RCNN 算法的网络流程图。

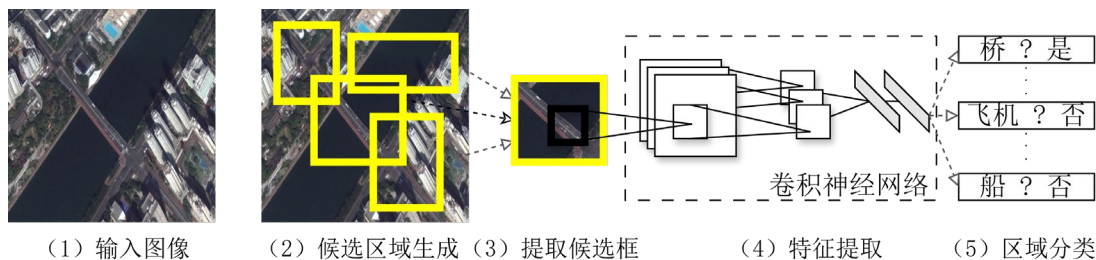


图 2-10 RCNN 算法流程示意图

RCNN 算法的具体工作流程可以分为四个部分，首先为候选区域提取，该部分主要是利用 Selective Search 算法在输入图像中生成一系列候选区域。其次为区域特征提取，该部分将每个候选区域调整为固定大小之后，将这些候选区域输入到预训练的卷积神经网络中进行前向传播，提取目标特征。第三部分为分类预测，该部分将提取好的特征向量送入 SVM 分类器中，对候选区域进行判断，区分背景与目标类别。最后一部分为边界框回归部分，该部分主要利用线性回归器

对候选区域的边界框进行优化与调整。该算法与传统的目标检测算法相比，具有准确度高、适用性强等特点，但由于其网络结构特点，又导致其检测速度慢、资源消耗大。

(2) Fast RCNN

Fast RCNN 算法是根据 RCNN 算法为原始模型进行改进，该算法的核心思想是利用卷积神经网络与候选区域生成结合，经过感兴趣区域池化层（ROI Pooling）对特征图进行固定，最后进行分类与回归预测。图 2-11 为 Fast RCNN 的网络结构流程示意图。

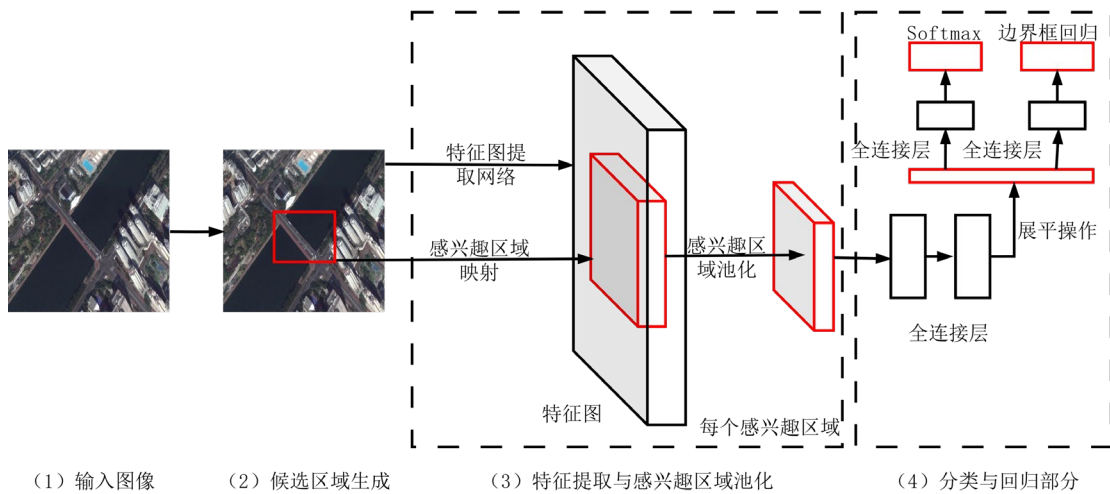


图 2-11 Fast RCNN 网络结构流程示意图

Fast RCNN 算法的具体工作流程与 RCNN 一样分为四个部分，首先是候选区域提取，依旧利用 Selective Search 算法对图片进行候选区域生成。其次为特征提取，与 RCNN 算法不同，Faster RCNN 算法将一整张图片直接输入到卷积神经网络中，生成一整张特征图。第三部分为 ROI Pooling 部分，该部分选择部分候选区域，将其映射到共享特征图中，再通过 ROI Pooling 操作将特征图划分成固定大小。最后一部分为特征分类与回归，通过将池化后的特征图输入到全连接层，分别对候选区域进行分类以及边界框的回归。该算法与原始 RCNN 相比，由于只对特征图进行一次特征提取处理，极大提升了网络的检测速度，同时对不同大小的候选区域特征进行固定大小的操作，解决了全连接层对输入尺度的要求。但该算法依旧依赖于 Selective Search 算法，导致候选框生成的速度较慢，使得其在性能上，特别是检测速度方面，依旧有很大的提升空间。

(3) Faster RCNN

Faster RCNN 是双阶段目标检测算法的经典算法之一，该算法主要是在 Fast RCNN 算法中进行改进，利用 RPN 网络取代 Selective Search 算法，这样可以有效的解决 Selective Search 算法选取候选框时存在的速度慢的问题。图 2-12 为 Faster RCNN 的结构示意图。

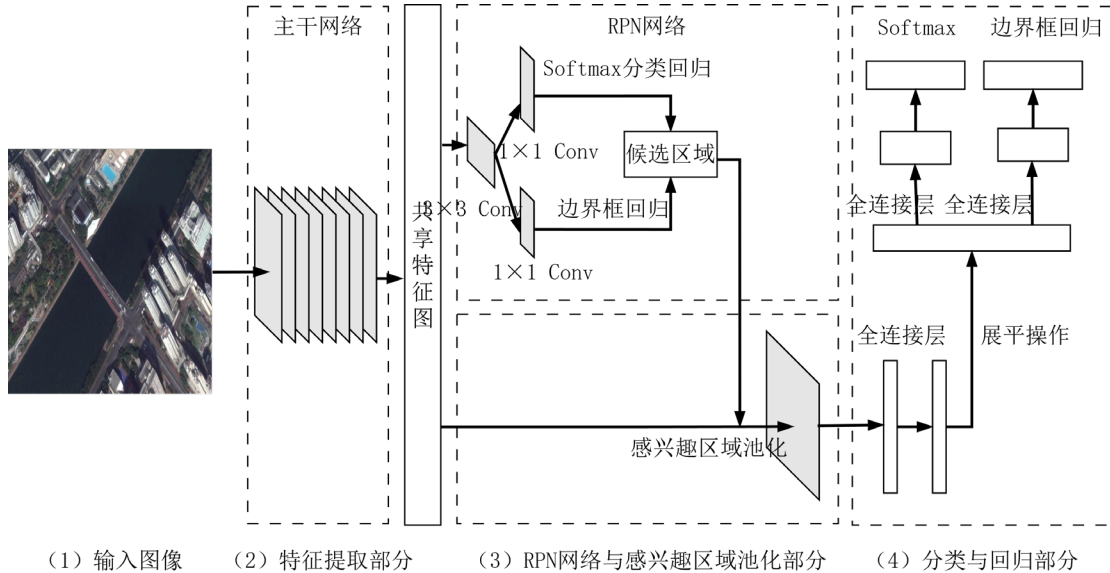


图 2-12 Faster RCNN 网络结构流程示意图

其主要流程可以分为“Fast RCNN”与“RPN”两个部分，首先是将输入图片送入特征提取网络中进行特征提取，将提取后的特征图输入到 RPN 网络中生成一系列可能包含目标的候选区域，之后通过 ROIpolling 层将候选区域映射到特征图上，并进行统一尺寸的调整，最后将固定大小后的特征图输入到全连接层中，并对其分别进行目标分类以及边界框回归，最终输出预测目标的类别及位置信息。

RPN 层为候选区域网络层，该层是原始 Faster RCNN 算法改进的核心，其主要工作流程为，首先将通过卷积神经网络得到的共享特征图作为输入，之后使用 3×3 的卷积核在特征图上进行滑动窗口操作，并通过计算得到窗口中心点对应原图中的中心点，从而按不同比例生成多个锚框，在分别对每个锚框进行前景与背景的二分类以及对锚框的坐标进行回归调整，最后通过非极大值抑制函数等处理筛选出最优的候选区域。该方法相较于 Selective Search 算法，计算效率有明显提升，同时生成的候选区域要更加贴合目标分布。

2.3 本章小结

本章节深入探讨了深度学习在目标检测领域的相关理论。首先介绍了卷积神经网络的基础知识，包括卷积层和池化层等核心概念，并剖析了它们的特性。继而，本章详细分析了当前流行的几种目标检测算法，涵盖了它们的架构、工作机制以及各自的优势和局限。这些理论探讨为后续章节提出的遥感图像目标检测算法奠定了坚实的理论根基。

第3章 遥感影像数据分析与预处理

3.1 遥感影像数据分析

数据集在深度学习目标检测任务中起着至关重要的作用，他将直接影响到模型训练的效果、性能以及最终应用的效果。对于普通数据集而言，图片的获取方式往往更加简单，同时目标的特征比较明显，这对数据集的采集与目标的标注工作都非常有利。而遥感图像数据的获取方式往往来自于无人机或者卫星等，这对于遥感图像的采集工作造成很大的困难，同时由于遥感图像自身的特点使得对目标的标注工作也比较困难。目前，研究者们为了能够满足对遥感图像目标检测任务的研究，通过现有资源以及技术制作并公开了一些遥感影像数据集。例如，NWPU VHR-10 数据集以及 MAR20 遥感军用飞机数据集等，本文也将采用这两个数据集对模型进行训练与评估。

NWPU VHR-10 数据集是在 2014 年由西北工业大学发布的，该数据集总共有 800 张图片，分别由 600 张为正样本图片以及 200 张为负样本图片组成，其中正样本图片中包含至少一个目标，而负样本中则完全由背景组成，不包含目标，整个数据集共有 10 类常见的地表目标，具体包括飞机、船舶、网球场、大桥等，这有利于检验模型在应对不同尺度大小的目标时的表现，图 3-1 所示为数据集中部分数据示意图。

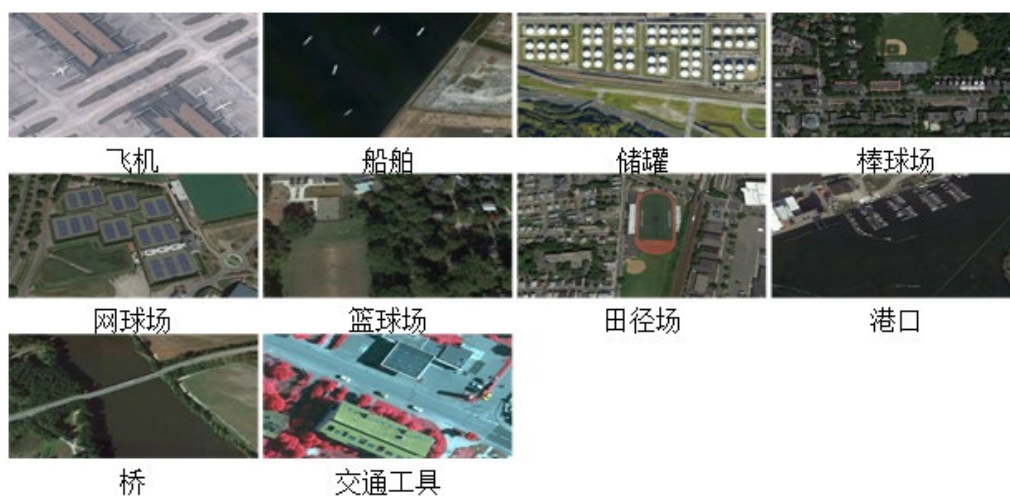


图 3-1 NWPU VHR-10 数据集部分数据示意图

在该数据集中共包含 3775 个目标，该数据集存在背景复杂、目标尺度不一以及目标各自的特征有明显区别等特点，这能够更加准确的检验模型的各项性能。对于数据集中每种目标的数量以及在整个数据集中所占比例如表 3-1 所示。本文在实验中将该数据集随机 7:1:2 的比例划分为训练集、验证集以及测试集。

表 3-1 NWPU VHR-10 数据集目标的数量以及所占比例表

目标类别	目标数量	比例 (%)
飞机	757	20.05
船舶	302	8.00
储罐	655	17.35
棒球场	390	10.33
网球场	524	13.88
篮球场	159	4.21
田径场	163	4.32
港口	224	5.93
桥	124	3.28
交通工具	477	12.64

需要说明的是，在本文设计实验中，“常见地物目标”主要指代 NWPU VHR-10 数据集中所包含的常见地表物体，如飞机、船舶、桥梁等类别。为了进一步验证所提出方法在遥感图像小目标检测任务中的性能，本文还引入了以军用飞机为代表的 MAR20 数据集。

MAR20 遥感军用飞机数据集为 2023 年由禹文奇等人向外公开发布，该数据集共有 3842 张图片，包含包括 A1 到 A20 在内的 20 类常见型号的军用飞机，图 3-2 为该数据集中部分数据示意图。

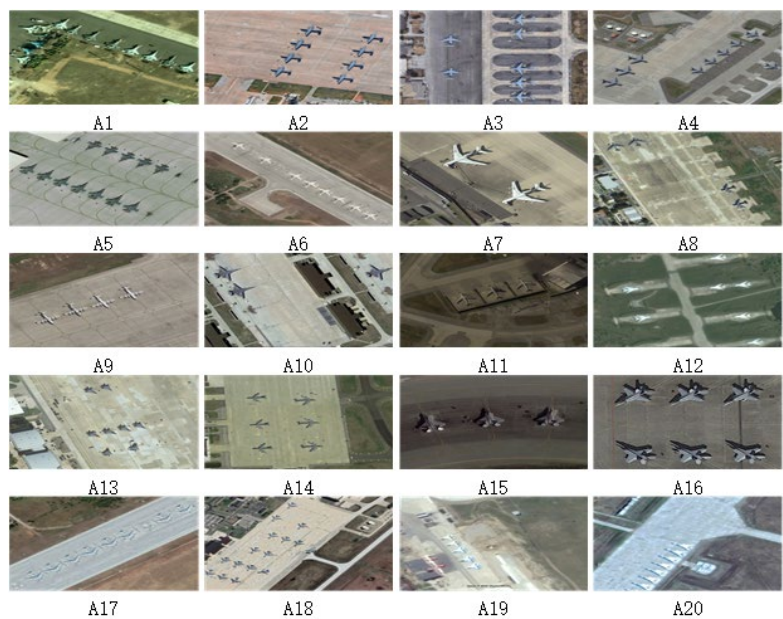


图 3-2 MAR20 数据集部分数据示意图

该数据集共有 22341 个目标，该数据集的特点有目标尺度均较小且目标的类型特征相似，同时由于目标所处环境使得目标与背景的相似度很高，这有利于检验模型对特征的提取能力以及对小目标的检测能力。该数据集每种目标的数目及在每种数目在数据集中所占比例均在表 3-2 中展示。本文同样采用 7:1:2 的比例对该数据集进行随机划分。

表 3-2 MAR20 数据集目标的数量以及所占比例表

目标类别	目标数量	比例 (%)	目标类别	目标数量	比例 (%)
A1	1646	7.41	A11	507	2.28
A2	1729	7.78	A12	702	3.16
A3	1176	5.29	A13	1625	7.32
A4	642	2.89	A14	1778	8.00
A5	1262	5.68	A15	618	2.78
A6	441	1.99	A16	2632	11.85
A7	580	2.61	A17	1397	6.29
A8	944	4.25	A18	308	1.39
A9	1086	4.89	A19	1236	5.56
A10	924	4.16	A20	981	4.42

3.2 遥感影像数据预处理

在目标检测任务中，一般使用 labeling 工具对目标数据进行标注工作，其工作界面如图 3-3 所示。本文所获取的原始数据集的数据格式均为 Pascal VOC 格式，该格式将被检测目标的信息均保存在 XMI 文件中，该文件中包含了目标种类、标签、图片大小以及边界框等信息，这种格式一般适用于 Faster RCNN 等算法，而 YOLO 算法却并不适用。YOLO 算法对于数据集格式有着其自身的要求，其所检测的目标一般保存在 txt 文件中，该文件只有标签信息以及边界框坐标参数。所以对于本文所使用 YOLOv5 算法，需要将数据集格式更改为 YOLO 格式，图 3-4 为 Pascal VOC 格式以及 YOLO 格式的示例图。

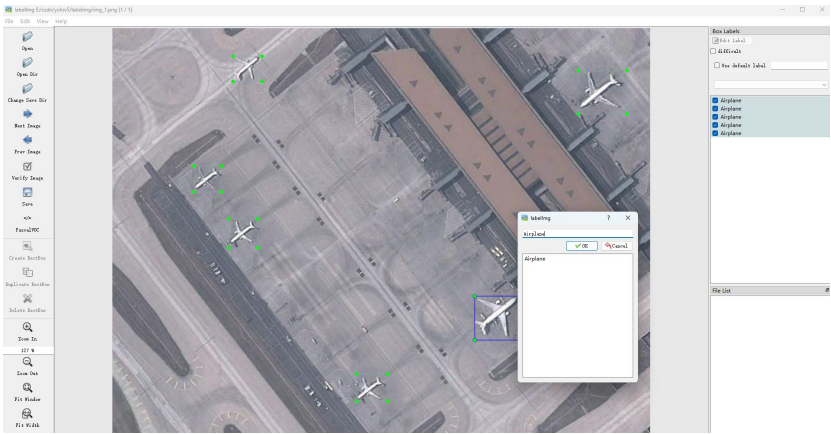


图 3-3 labeling 标注界面

```
<annotation>
  <folder>VOC2007</folder>
  <filename>000001.jpg</filename>
  <size>
    <width>958</width>
    <height>808</height>
    <depth>3</depth>
  </size>
  <object>
    <name>airplane</name>
    <pose>unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>563</xmin>
      <ymin>478</ymin>
      <xmax>630</xmax>
      <ymax>573</ymax>
    </bndbox>
  </object>
</annotation>
```

```
0 0.621607 0.649133 0.069937 0.117574|
```

(a) VOC数据格式

(b) YOLO数据格式

图 3-4 数据集格式

除对数据集格式的转换外，本文还对两种数据集进行数据增强操作。数据增强是在深度学习中常见的数据预处理方式之一，在目标检测任务中，想要训练出较好的神经网络模型离不开大量的数据支撑，然而，实际中获取足够数量且标注精准的数据往往成本较高，难以满足模型训练需求，故需要数据增强来对现有数据进行扩充。经过数据增强操作后，原有数据集不仅得到了扩充，还会增加样本的多样性，提高模型的性能。

本文所使用的两个数据集只图片数量与种类并不丰富，这可能会导致在训练过程中出现过拟合的现象，同时数据集中数据数量分布并不均匀，这会导致对部分目标的检测效果差。为解决这些问题，本文在数据预处理过程中对所使用的数据集进行数据增强操作，具体操作包括裁剪、平移、缩放、翻转以及添加椒盐噪声等方式。图 3-5 为具体操作的示意图。

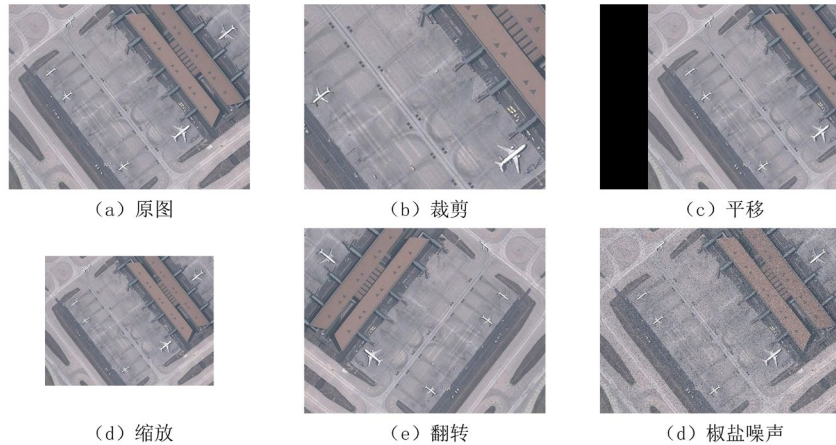


图 3-5 数据增强处理后对比图

3.3 目标检测常用的评价指标

在目标检测任务中，评价模型的指标通常有精确率（Precision, P）、召回率（Recall, R）、平均精度均值（mean Average Precision, mAP）以及 F1_score 等，其中精确率指的是在预测正确的正样本占有所有正样本的比例，召回率指的是预测正确的正样本占实际的正样本的比例，其计算表达式如式(3-1)与式(3-2)所示。

$$Precision = \frac{TP}{TP + FP} \quad (3-1)$$

$$Recall = \frac{TP}{TP + FN} \quad (3-2)$$

其中 TP 为在预测正确的正样本数量，FP 为预测错误的正样本数量，FN 为预测错误的负样本数量。

mAP 则更加全面的反映了模型的性能，其计算表达式如式(3-3)所示，mAP 包括 mAP@0.5 与 mAP@.5:.95，其中 mAP@0.5 分别表示交并比(intersection over union, IoU)的阈值设置为 0.5 时的平均精度，而 mAP@.5:.95 表示的是交并比阈值在 0.5 到 0.95 之间（步长为 0.05）十个不同阈值的平均 mAP。

$$mAP = \frac{1}{N} \sum_{i=0}^N AP_i \quad (3-3)$$

对于 F1_score 指标为衡量准确率与召回率的综合指标，该指标可以更好的反映模型在不均衡数据集集中的表现，其计算表达式如式(3-4)所示。

$$F1_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3-4)$$

此外，为全面评估模型的复杂度与资源消耗，本文还引入了浮点计算次数（FLOPS, Floating Point Operations Per Second）和参数量（Params, Parameters）两个指标。FLOPS 是指模型在一次前向推理过程中所需进行的浮点运算次数。FLOPS 越低，通常意味着模型计算开销越小，推理速度越快，在资源受限场景下具有重要意义。Params 是模型中所有可训练参数的总数量。参数量直接影响模型的存储需求和训练成本，参数量越小，模型越轻量，部署在边缘设备或移动终端上越具有优势。

3.4 本章小结

本章首先对遥感图像数据进行分析，着重介绍了本文所使用的两个数据集，并对数据集的特点进行了分析，为了适应不同算法的需求，本文对数据格式进行了转换，并通过几种数据增强操作扩展数据集。最后，本文还介绍了目标检测中的常用评价指标，如精确率和召回率等，帮助全面评估模型的性能。

第 4 章 基于改进的 Faster RCNN 的遥感图像常见地表目标检测算法

4.1 引言

对于遥感图像来说,目标检测一直是一个非常重要的研究方向。随着科学技术的发展,对遥感图像目标检测任务的标准也越来越高,传统得目标检测算法已经不能够满足要求,而随着深度学习的兴起为遥感图像目标检测打开了新的大门。目前基于深度学习的方法主要分为单阶段目标检测算法与双阶段目标检测,其中双阶段目标检测的检测精度相对较高,但由于其将检测算法分为特征提取以及候选区域生成两部分,导致其网络的复杂程度较高以及检测速度较慢,这使得其在遥感图像目标检测任务中存在了一定的局限性。

以本章所使用的 Faster RCNN 算法为代表,其特征提取部分使用的网络为 VGG16 网络,该网络仅仅通过不断对输入图像进行卷积与池化操作,这样的得到的最终特征图虽然能够提取到有用的图像特征,但其对多尺度特征的捕捉能力较弱,使得原始网络在处理不同尺度目标时效果并不理想。由于 VGG16 网络各个层之间采用的是堆叠方式,其随着网络的深度的提升,原始网络可能会出现梯度爆炸或者梯度消失的现象。同时原始的 VGG16 网络的参数较多,这可能会导致原始网络出现过拟合的现象,因此网络在训练过程中对数据量有一定要求。为此,本章针对这些问题,同时结合遥感图像目标自身的特点,设计改进的 Faster RCNN 算法,对原始算法进行针对性的改进,并通过设置多组实验来对改进的算法进行全面评估。

4.2 改进的 Faster RCNN 模型

4.2.1 残差网络模块

在 Faster RCNN 原始算法中,主干网络在整个网络承担着特征提取的作用,该部分的好坏将直接影响整个目标检测算法的性能。原始主干网络使用的是

VGG16 网络，该网络的优点是具有较高的准确率，但其缺点却相对明显，就是网络复杂、消耗资源大且训练与检测速度非常慢，对此，本文引入残差网络对其进行改进，残差网络是 2015 年由何恺明等人提出，其主要思想是引入了残差模块，该残差模块通过一系列卷积等操作以及一条捷径组成，其结构如图 4-1 所示：

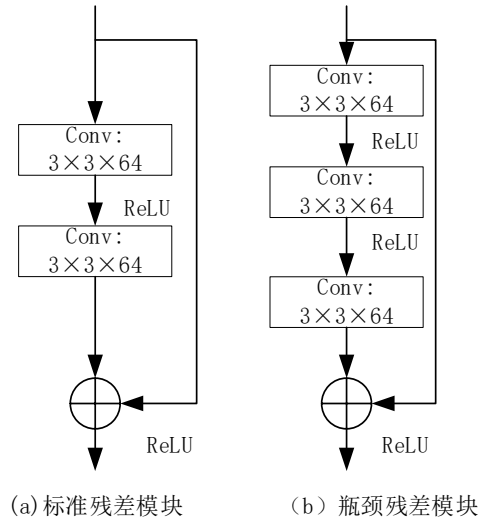


图 4-1 两种残差模块结构示意图

残差模块主要分为两种，一种为标准的残差模块，其将输入依次经过卷积层、激活函数以及相同的卷积层后，与原始输入进行相加，其中卷积层中卷积核大小为 3×3 ，步长为 1，该种残差模块主要用在一些低层网络中。另一种为瓶颈残差模块，该模块将输入依次经过 1×1 的卷积层、激活函数、 3×3 的卷积层、激活函数以及 1×1 的卷积层，之后与输入进行相加，在该结构中卷积层的步长均为 1，这种残差结构主要是为了减少计算量而被运用在具有较深层次的网络中。

残差网络有很多网络变体，例如 ResNet18、ResNet34、ResNet50、ResNet101 等网络，其区别主要是残差块的数量与网络数量的不同。以 ResNet50 为例，主要包含初始层、残差模块组与结尾层，其中初始层包含一个步长为 2 的 7×7 的卷积层和一个步长为 2 的 3×3 的最大池化层，残差模块组中有包含四个模块组，每个模块组中分别包含 3、4、6 和 3 个残差模块，而结尾层主要是由一个平均池化层与一个全连接层组成。在残差模组中，为了匹配每个模组所输入的特征图所需的大小以及深度，需要对输入进行修改。通过在捷径分支中加入一个步长为 2、卷积核大小为 1×1 的卷积层，其路径用线表示，以及修改另一条分支中卷积核

步长以及深度来实现，该残差模块在网络中被叫做 **Bottleneck1**，而该模组中其余的残差模块称为 **Bottleneck2**，具体的实现方式如图 4-2 所示。

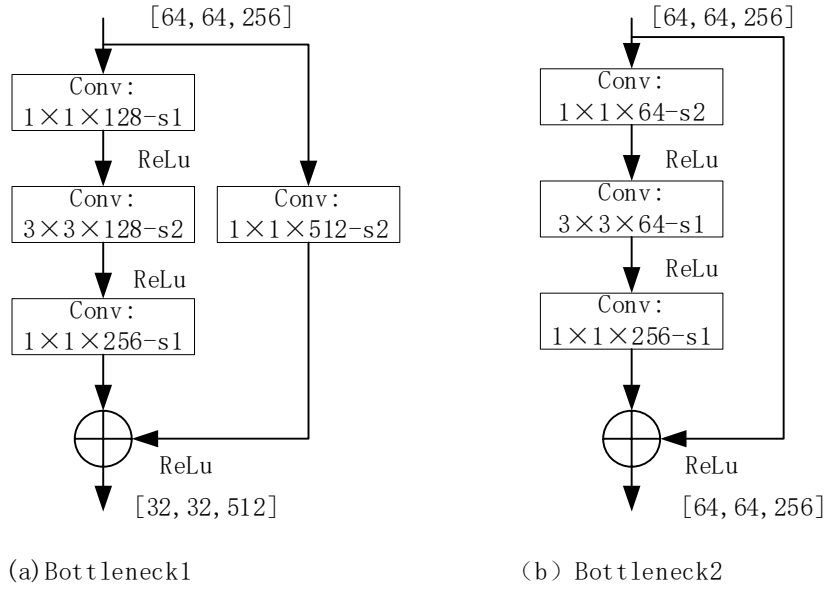


图 4-2 Bottleneck1 与 Bottleneck2 的网络结构示意图

在 ResNet50 的网络结构中，为了匹配每个残差模组的最佳大小与深度，需要对每个残差模块卷积层的卷积核大小、深度以及步长做不同的调整。具体调整在图 4-3 中进行了展示，其中 Stage1 表示网络中第一个残差模组。

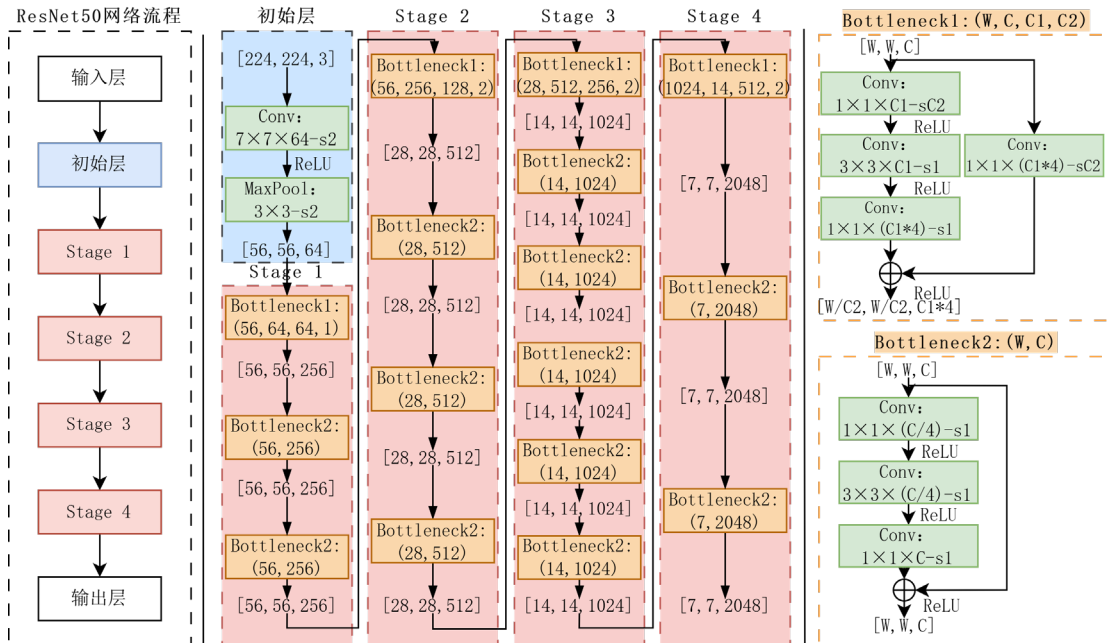


图 4-3 ResNet50 网络结构图

本文选择用 ResNet50 网络替换原始 Faster RCNN 算法的主干网络，相对于

原始 VGG16 网络其有着更加深层次的网络结构以及更好的泛化能力，并且在计算效率方面也有着不小优势，拥有较低的参数量和计算量。同时在训练的稳定性以及收敛速度方面的表现都比原始的 VGG16 网络要更加优秀。其与 ResNet34 不同点在于其拥有更加深的网络结构，这使得网络能更好的提取遥感目标更加深层的特征。而 ResNet101 网络，其从理论上可以学习到更多的特征，但其网络的深度也带来了训练与推理的一些挑战，同时 ResNet50 拥有更加少数参数量，在遥感图像目标检测任务中往往有着更加优秀的表现。

4.2.2 注意力机制的加入

注意力机制是一种深度学习技术^[69]，用于让模型关注输入数据的关键部分，从而更加高效的处理信息。注意力机制的核心思想是根据输入的上下文，动态的为不同部分分配权重，让模型更加关注于当前任务相关部分，而弱化不重要的部分。常见的注意力机制有 SE^[70]、CA、ECA^[71]以及 CBAM 等。

遥感图像因其背景复杂，使得原始算法对目标特征的关注度较低，因此此次改进算法引入更加有效的空间通道注意力机制，该注意力机制其结构主要由空间注意力模块与通道注意模块组成。其中空间注意力机制将输入的特征图进行全局池化（最大池化与平均池化），将结果相加后经过一个 7×7 的卷积层（激活函数为 Sigmoid 函数），得到权重矩阵，并将该矩阵与原特征图相乘，形成最终的新特征图。通道注意力则是将经过全局池化后的结果输入到共享多层感知器（Multi-Layer Perceptron ,MLP）中，得到通道权重，从而得到新的特征图。其结构分别如图 4-4 和图 4-5 所示。

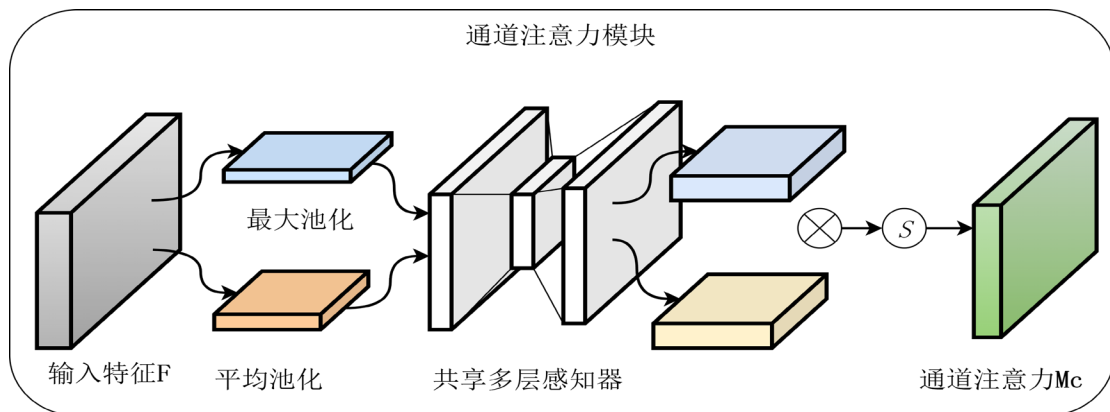


图 4-4 空间注意力机制模块示意图

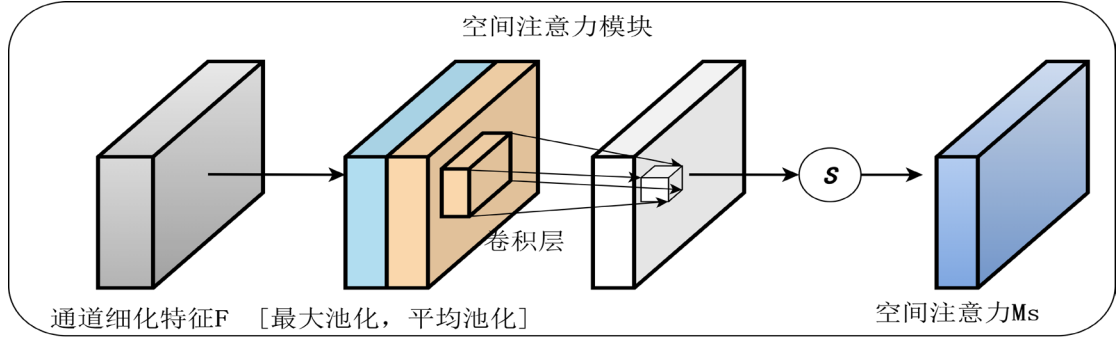


图 4-5 通道注意力机制模块示意图

空间注意力机制公式如下式(4-1):

$$\begin{aligned} M_s(F) &= \delta(f^{(7*7)}([AugPool(F), MaxPool(F)])) \\ &= \delta(f^{(7*7)}([F_{aug}^C; F_{max}^C])) \end{aligned} \quad (4-1)$$

通道注意力机制公式如下式(4-2):

$$\begin{aligned} M_c(F) &= \delta(MLP(AugPool(F)) + MLP(MaxPool(F))) \\ &= \delta(W_1(W_0(F_{aug}^C)) + W_1(W_0(F_{max}^C))) \end{aligned} \quad (4-2)$$

通道注意力主要关注有意义的特征，而空间注意力主要关注那些有意义特征的位置，CBAM 模块将两种模块串联起来，提升网络对全局与局部信息的把握，其结构如图 4-6 所示。

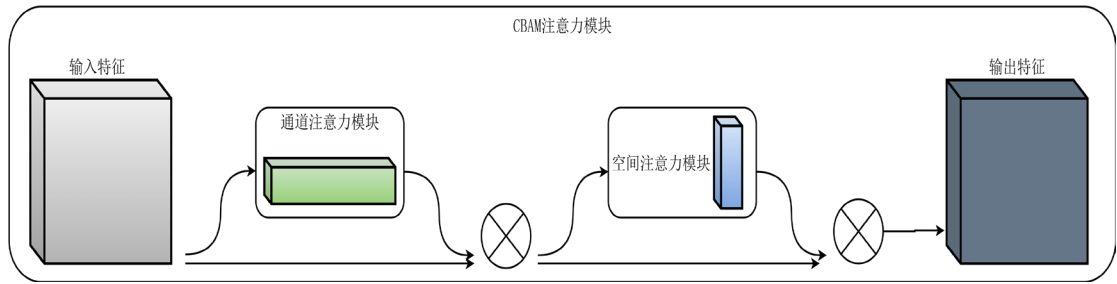


图 4-6 CBAM 注意力机制示意图

图 4-7 所示为融合 CBAM 注意力机制后的特征提取网络结构示意图，本文在 ResNet50 网络中每个残差模组后加入注意力机制，加入后的网络相较于原始网络，在性能上，在处理遥感图像任务时，CBAM 注意力机制能够自动调节关注区域，提高目标识别能力。在计算效率上，虽然 CBAM 模块增加了额外的计算，但其计算开销相对较小，同时在精度等方面带来的收益较大。在适用性方面，修改后的网络能够处理更加复杂的场景，同时由于 CBAM 本身具有即插即用的特

性使得其在其他网络中也拥有不错的效果。

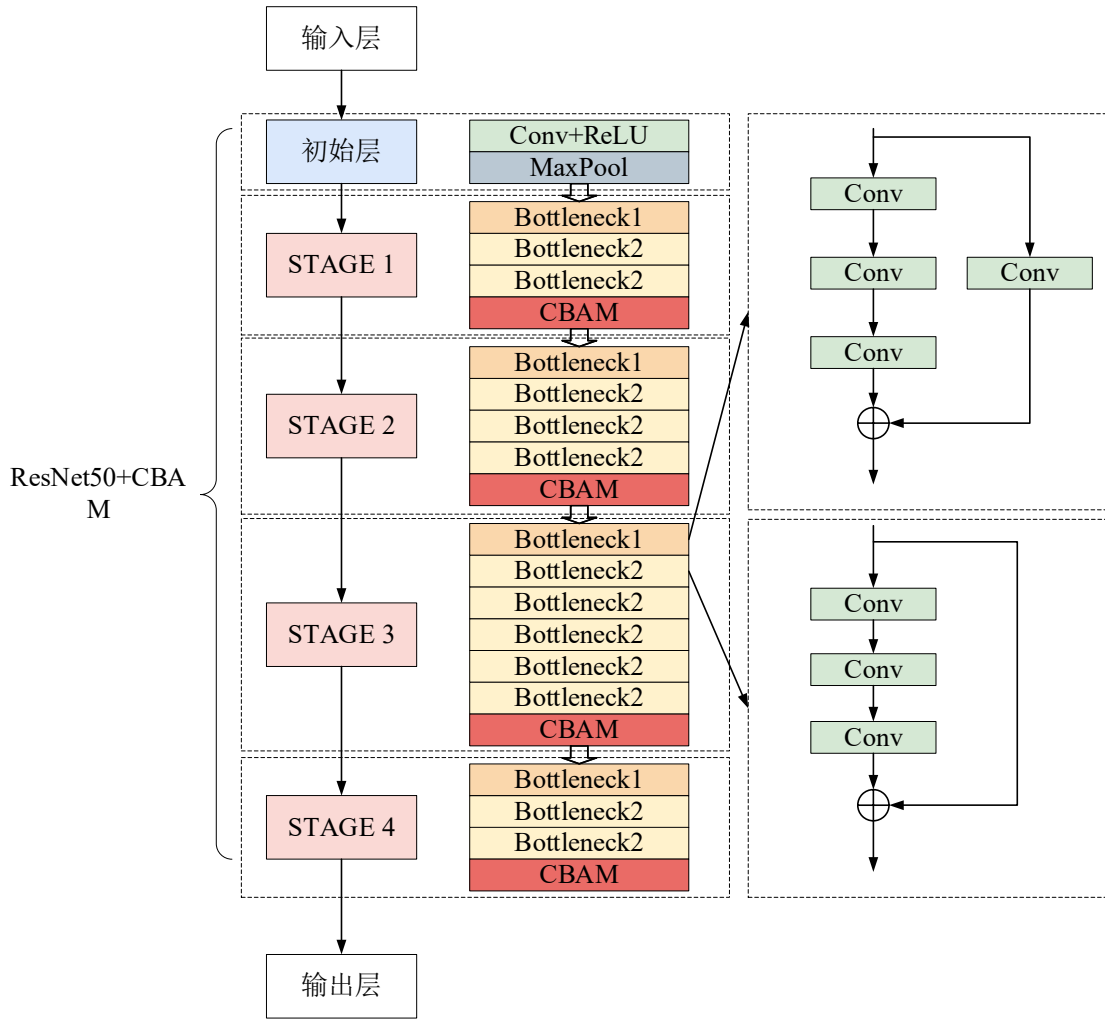


图 4-7 修改后的 ResNet50 结构示意图

4.2.3 特征金字塔网络模块

通过前两节残差网络以及注意力机制模块的引入，一定程度上缓解了原始算法存在的误检现象，但是对于遥感图像目标而言，由于其目标的尺度大小差异大且背景比一般图片复杂等原因，使得原始网络检测精度依然有待提升，为解决这一问题，本文通过在之前内容的基础上引入特征金字塔网络（FPN，Feature Pyramid Networks）。FPN 由 Tsung-Yi Lin 等人在 2017 年提出，其主要思想是利用“自顶向下”的结构，对不同特征层进行多尺度融合。其结构主要由自底向上模块、自顶向下模块以及横向连接组成，其中自底向上模块主要是网络进行的前向传播，主要由不同分辨率的特征图组成。自顶向下模块主要是进行上采样，上采样的进行使得不同层的特征图具有相同的大小，便于下一步的横向连接。横向连

接主要是进行不同尺度特征图之间进行特征融合，其主要操作是将上采样的结果与自底向上的不同层特征图之间进行特征融合。

本文将利用 ResNet50 网络与 FPN 相结合，将残差网络的每个残差模块与 FPN 进行结合，经过一个 1×1 的卷积，将高层特征降采样到统一的通道数，同时为消除上采样带来的混叠效应，会经过一个 3×3 的卷积对融合结果进行卷积，这样进行不同尺度特征进行特征融合使得改进后的算法与原始 Faster RCNN 相比，不仅减少了资源消耗和检测时间，而且提升了对于遥感图像目标检测的精度。图 4-8 为 ResNet50 与 FPN 相结合后的网络结构图。

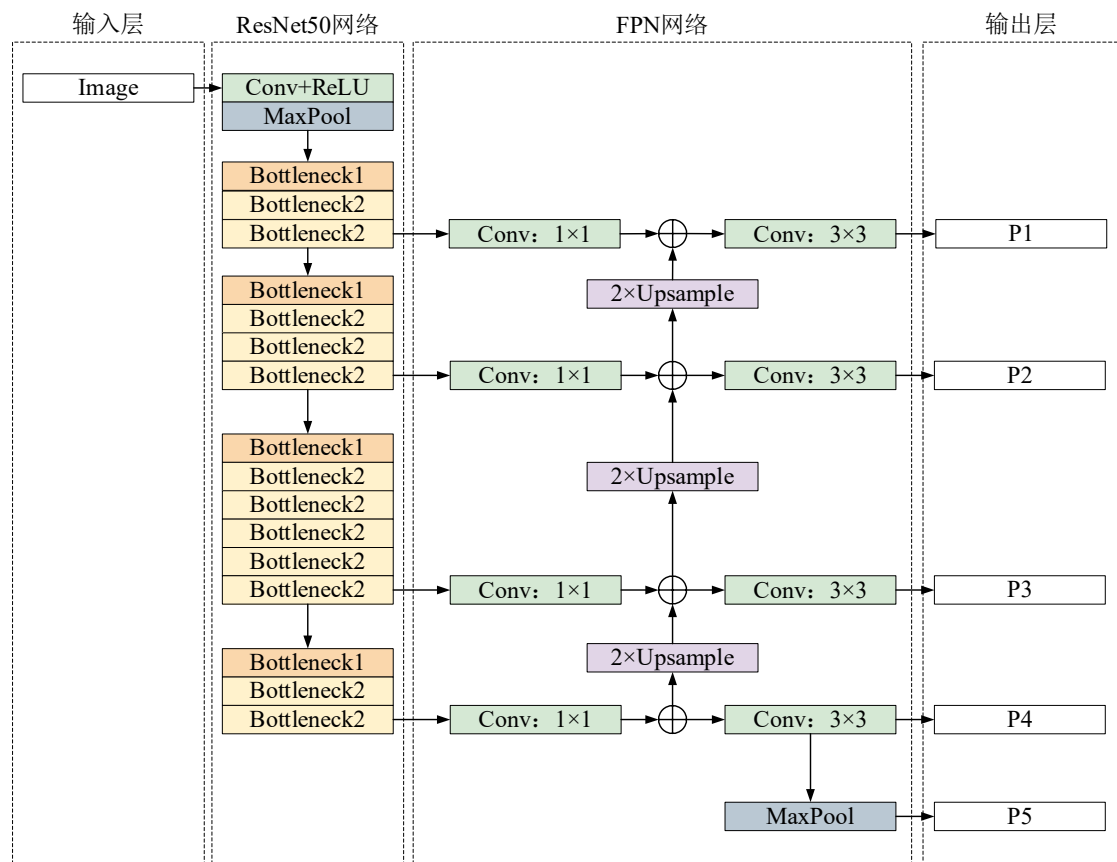


图 4-8 ResNet50 与 FPN 结合后结构示意图

结合 FPN 后的网络，与原始网络相比，在性能上，增加了多尺度特征表达能力，显著提升了网络对遥感图像中不同尺度目标的检测能力。在计算效率上，由于通过 1×1 的卷积降低了通道数，使得后续的 ROI Pooling 层以及 RPN 网络的参数量得到大幅减少。同时，对遥感图像而言，FPN 的加入提高了网络的泛化能力，增强了目标和背景的区别能力，使网络在不同场景中拥有更好的检测效果。

4.2.4 改进后的网络结构

针对 Faster RCNN 算法在遥感图像任务中的出现的各种不足,同时满足遥感图像目标检测精度等指标的要求,本文在上述分析的基础上对原始 Faster RCNN 算法进行改进,改进后算法的整体架构如图 4-9 所示。

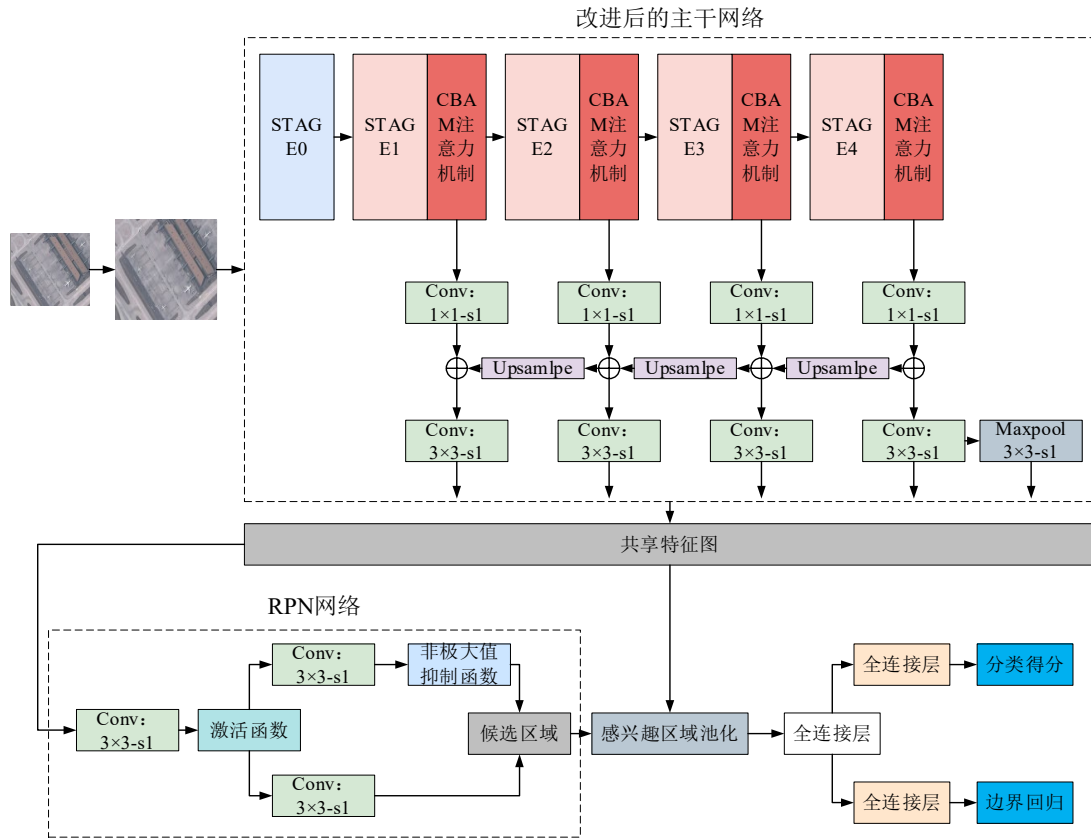


图 4-9 改进后 Faster RCNN 算法整体架构示意图

4.3 实验结果与分析

4.3.1 实验平台

本实验采用的实验平台为 64 位 windows 11 系统, CPU 为 13th Gen Intel(R)Core(TM) i5-13400F, GPU 为 NVIDIA GeForce 3060ti (8G)。系统内存容量为 16G。此次研究使用软件平台为 PyCharm Community Edition 2023.2, 深度学习框架使用 pytorch1.10.2, 编译语言采用 python3.6, 同时使用 Cuda11.8 用于 GPU 加速。

4.3.2 实验设计

本次全部实验采用 NWPU VHR-10 数据集以及 MAR20 遥感军用飞机数据集，对每个数据设计消融实验与对比实验来对改进算法进行验证。同时，为了实验的严谨性，本章对算法中的超参数进行统一设置，具体设置由表 4-1 所示。

表 4-1 Faster RCNN 模型超参数设置一览表

超参数名称	参数设置
batch size	2
workers	1
初始学习率	0.01
epoch	150
优化器	SGD

其中，由于网络对资源消耗较大，故本章设置 batch size 大小为 2，workers 设置为 1。同时由于本算法采用在 COCO 数据集集中的预训练权重进行辅助训练，设置初始学习率为 0.01，训练轮数 epoch 设置为 150，并使用 SGD 优化器对网络性能进行优化。

4.3.3 消融实验结果与分析

(1) 参数量以及计算量实验结果与分析

本文为验证改进算法的有效性，将每项改进方法逐一加入到原始 Faster RCNN 算法中，如表 4-2 表明了 NWPU VHR-10 数据集上计算复杂度以及参数量的情况展示，其中，如果采用哪种改进方法就在下方使用“√”作为标记，未采用的改进方法则标记为“-”，例如实验 1，表示在原始网络中仅将原始主干网络使用 VGG16 模型替换成 ResNet50 模型的方法。

表 4-2 参数量以及计算量结果表

实验	VGG16	ResNet50	FPN	CBAM	FLOPS/G	Params/M
1	√	-	-	-	215.59	43.92
2	-	√	-	-	95.42	70.55
3	-	√	√	-	80.57	40.74
4		√		√	75.12	70.59
5	-	√	√	√	134.53	41.40

如表 4-2 所示，由于原始 VGG16 网络对资源消耗较大，且设备资源有限，需要对其进行修改，只采用特征网络中第 42 层作为特征图输出，使得原始网络的参数量 Params 仅达到了 43.92M，即使如此，其计算量依旧达到了 215.59GFLOPS。随着将 VGG16 网络替换为 ResNet50 网络，虽然由于网络深度的提升，导致参数量加大，但计算量却大幅减少。之后随着 FPN 以及 CBAM 注意力机制的加入，使得最终算法的计算量以及参数量分别达到 134.53GFLOPS 和 41.40M，较原始算法分别减少了 81.06GFLOPS 以及 2.52M，这充分说明改进算法更加高效，且消耗的资源少。

(2) NWPU VHR-10 数据集消融实验与分析

表 4-3 NWPU VHR-10 数据集消融实验结果表

实验	VGG16	ResNet50	FPN	CBAM	mAP@0.5/%	mAP@.5:.95/%
1	√	-	-	-	75.97	40.23
2	-	√	-	-	76.88	41.25
3	-	√	√	-	80.58	43.57
4		√		√	83.48	44.14
5	-	√	√	√	86.66	50.02

由表 4-3 中数据可以得知，本章算法在 NWPU VHR-10 数据集的 mAP@0.5 可以达到 86.66%，相较于原始算法的 75.97%提升了 10.69%，而其 mAP@.5:.95 则为 50.02%，相较于原始算法的 40.23%提升了足足有 9.79%。其中，将原始算法主干网络替换为 ResNet50 后，mAP@0.5 与 mAP@.5:.95 分别提升了 0.91% 与 1.02%，这说明 ResNet50 网络相较于 VGG16 网络在处理遥感图像时具有一定优势，随着 FPN 网络的结构加入，其检测精度有着明显提升，其 mAP@0.5 与 mAP@.5:.95 达到了 80.58%与 43.57%，相较于原始模型提升了 3.70%与 2.32%，这说明在应对尺度变化较大的目标时，FPN 结构对算法性能有较大的提升。而随着 CBAM 注意力机制的加入，其检测精度有了进一步的提升。通过上述一系列的消融实验说明本章的改进思路对原始算法在遥感图像目标检测任务中有着较好的表现。

除了上述实验外，本文为进一步验证算法的有效性，还列出了改进前后数据

集中每类目标的检测精度以及其对比结果，表 4-3 详细的记录了算法对每一种类型的飞机检测精度。

表 4-4 每类目标检测结果对比表

种类	mAP@0.5/%		
	改前	改后	变化对比
飞机	95.6	97.7	+2.1 ↑
船舶	89.6	91.2	+1.6 ↑
储罐	50.4	70.0	+19.6 ↑
棒球场	97.9	69.7	-28.2 ↓
网球场	70.4	86.1	+15.7 ↑
篮球场	72.1	89.8	+17.7 ↑
田径场	98.8	99.9	+1.1 ↑
港口	75.1	83.0	+7.9 ↑
桥	47.1	71.0	+23.9 ↑
交通工具	62.5	81.2	+18.7 ↑

由表 4-4 数据可以看改进后算法对田径场的检测时 mAP@0.5 最高,为 99.9%,较原始算法提升了 1.1%。在整个数据集中提升最多的种类为桥,该种类从 47.1%提升到了 71.0%,提升了有 23.9%。同时,改进后的算法在检测棒球场时效果有所下降,这可能是因为棒球场与背景纹理相似程度较高,改进后的算法对目标特征的关注度较高,过度关注了目标周边的相似区域,造成误检现象,导致精度下降,但对于整体而言,改进后的算法对数据集中绝大部分目标的检测精度都有一定程度的提升。

图 4-10 为算法改进前后的训练 Loss 曲线的对比图,从图中可以看出,原始算法在开始训练时的 Loss 值较大,这说明原始算法一开始并不能很好的对数据集中目标的特征进行提取,随着训练轮数的增加, Loss 值很快趋于稳定,这说明改进后的算法训练结果还是较为理想的。同时对于最终趋于稳定的 Loss 值改进后的算法比原始算法要低,这说明改进后的算法更加适用于遥感图像目标检测任务。

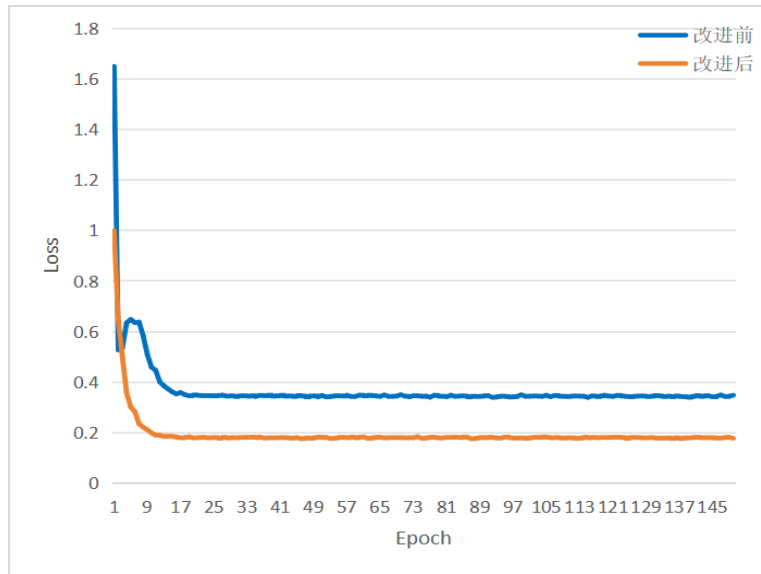


图 4-10 NWPU VHR-10 数据集 loss 曲线对比图

(3) MAR20 飞机遥感数据集消融实验与分析

为进一步检验改进后算法的效果以及适用性，本文除了使用 NWPU VHR-10 数据集，还增加了 MAR20 飞机遥感影像数据集，该数据集中存在 20 种常见飞机类型，对该数据同样进行消融实验，其实验结果如表 4-5 所示

表 4-5 MAR20 数据集消融实验结果表

实验	VGG16	ResNet50	FPN	CBAM	mAP@0.5/%	mAP@.5:.95/%
1	√	-	-	-	94.93	70.38
2	-	√	-	-	95.58	70.70
3	-	√	√	-	96.13	71.05
4	-	√	-	√	96.63	72.82
5	-	√	√	√	97.93	75.14

由表 4-5 中数据可以得知，本文对于遥感飞机数据集的检测效果较好，原始算法的 mAP@0.5 与 mAP@.5:.95 已经达到了 94.93%与 70.38%，将主干网络更换成 ResNet50 后分别达到了 95.58%与 70.70%。随着 FPN 以及 CBAM 注意力机制的加入，最终精度分别达到了 97.93%以及 75.14%，其中由于该数据集中目标尺度变化并不明显，FPN 结构加入后对于精度提升较小，但对于 CBAM 注意力机制的加入，网络在检测精度方面有较大的提升，mAP@0.5 与 mAP@.5:.95 精度提升了 1.3%与 4.09%，这充分说明了改进后的网络在应对更加复杂的场景时

也能有较好的表现。

表 4-6 MAR20 数据集中每类目标检测结果对比表

种类	mAP@0.5/%			种类	mAP@0.5/%		
	改前	改后	对比		改前	改后	对比
A1	96.1	97.8	+1.7 ↑	A11	93.6	98.5	+4.9 ↑
A2	97.9	99.9	+2.0 ↑	A12	94.2	96.8	+2.6 ↑
A3	97.9	99.0	+1.1 ↑	A13	95.0	99.5	+4.5 ↑
A4	88.0	93.3	+5.3 ↑	A14	97.9	98.6	+0.7 ↑
A5	91.3	97.0	+5.7 ↑	A15	86.4	96.5	+10.1 ↑
A6	98.7	99.8	+1.1 ↑	A16	97.6	98.7	+1.1 ↑
A7	98.4	98.7	+0.3 ↑	A17	99.2	99.9	+0.7 ↑
A8	99.3	99.0	-0.3 ↓	A18	83.6	94.6	+11.0 ↑
A9	98.7	99.0	+0.3 ↑	A19	95.3	93.8	-1.5 ↓
A10	97.9	99.8	+1.9 ↑	A20	93.0	98.5	+5.5 ↑

由表 4-6 数据可以看改进后算法在 A2 目标的检测上其 mAP@0.5 最高，为 99.9%，较原始算法提升了 2%。在整个数据集中提升最多的种类为 A18，该种类从 83.6%提升到了 94.6%，提升了有 11.0%。同时，除 A8 与 A19 外，其他种类的目标检测精度均有不同程度的提升。

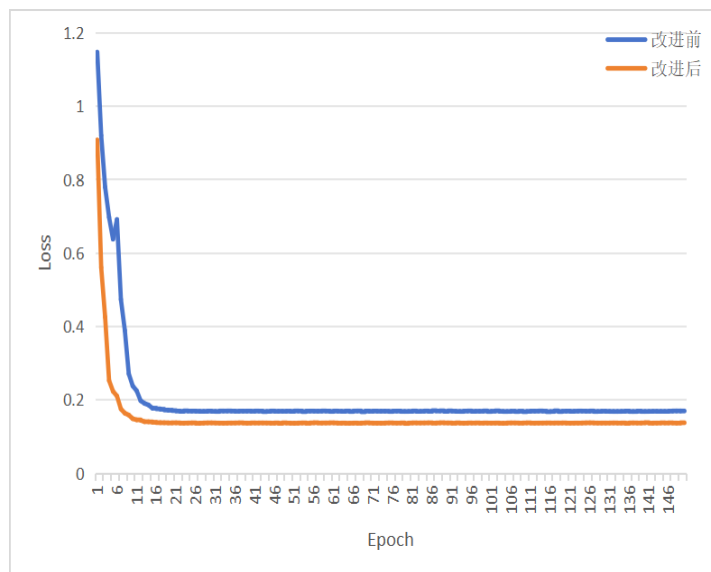


图 4-11 MAR20 数据集 loss 曲线对比图

图 4-11 为算法改进前后在 MAR20 数据集中的训练 Loss 曲线的对比图，该图中曲线的变化情况与在 NWPU VHR-10 数据集中的变化曲线基本一致，这不仅说明了改进后的算法有着良好适用性，更是充分的说明了算法在处理遥感图像目标检测任务中有着更好的表现。

4.3.4 对比实验结果与分析

(1) NWPU VHR-10 数据集对比实验与分析

本文除了通过算法的消融实验来验证算法的有效性，还通过一系列对比实验来进一步对算法的有效性进行验证。本次实验除了选择用 SSD、原始 Faster RCNN、YOLOv3、YOLOv5n 以及原始 YOLOv5s 等经典算法来对比本章算法。

表 4-7 NWPU VHR-10 数据集对比实验结果表

实验	Precision/%	Recall/%	F1_score/%	mAP@0.5/%
SSD	91.88	57.08	70.42	77.30
Faster RCNN（原始）	79.52	75.43	77.44	75.97
YOLOv3	84.76	70.32	76.87	84.58
YOLOv5n	93.00	78.50	85.14	85.40
YOLOv5s	86.50	80.70	83.50	84.10
Faster RCNN（改进）	86.73	83.65	85.16	86.66

表 4-7 是在 NWPU VHR-10 数据集上进行对比实验的结果，从表中可以看出本文算法的 mAP@0.5/% 比 SSD 的 mAP@0.5/% 高了 9.36%，召回率比 SSD 算法高了 26.57%，F1_score 高了 14.74%。除此之外，改进算法的 mAP@.5/% 相较于 YOLOv3、YOLOv5n 以及原始 YOLOv5s 分别高 2.08%、1.26% 与 2.56%。而除 Precision 外，Recall 与 F1_score 均比其他四种算法高。这充分说明了本文算法与其他常见算法相比具有一定的优势。

图 4-12 为在西北工业大学遥感影像数据集上的部分实验结果，可以看出面对不同的目标，原始算法存在漏检误检的现象，而本文算法则可以很好的缓解这些现象，使得检测更加精准。

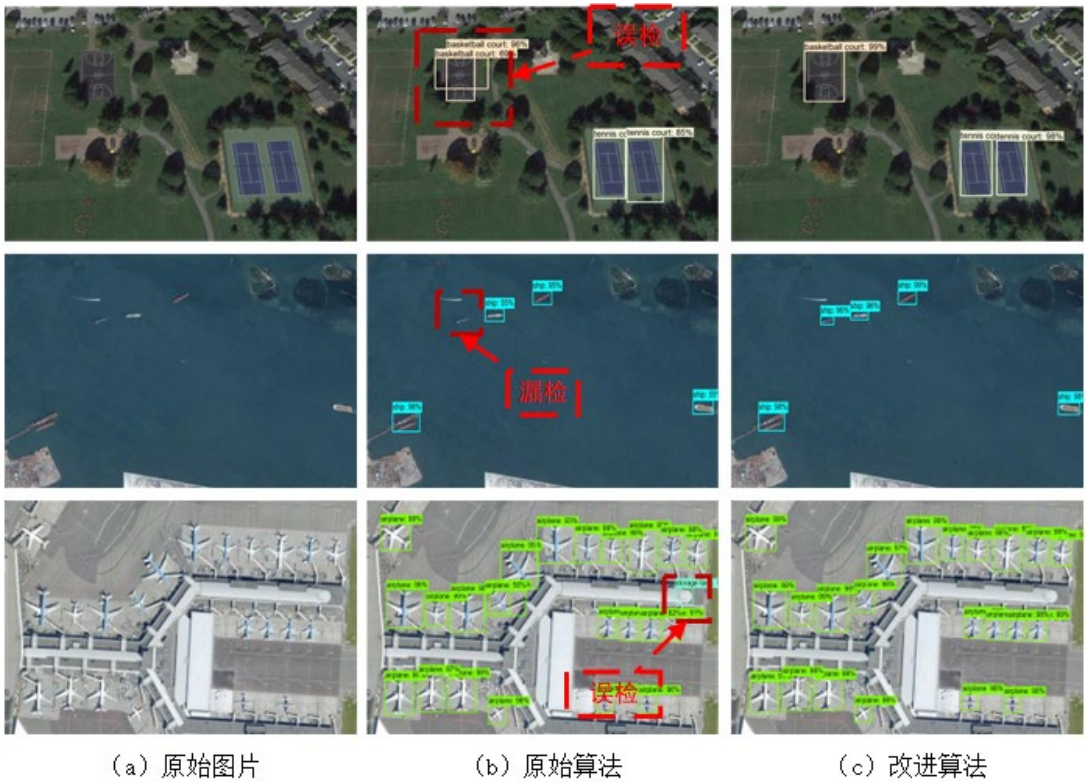


图 4-12 NWPU VHR-10 数据集对比实验结果图

(2) MAR20 数据集对比实验与分析

表 4-8 是在 MAR20 数据集上进行对比实验的结果，同样的由表可以看出，虽然改进算法在 MAR20 数据集上的准确率与 SSD 算法相比些许不足，但相对于其他算法而言有一定优势，并且本章改进算法的召回率、F1 及 mAP@0.5 等指标均优于其他算法。

表 4-8 MAR20 数据集对比实验结果表

实验	Precision/%	Recall/%	F1_score/%	mAP@0.5/%
SSD	97.95	87.62	89.59	94.46
Faster RCNN（原始）	95.56	91.63	93.55	94.93
YOLOv3	77.85	45.30	51.30	63.34
YOLOv5n	95.40	92.40	93.88	96.60
YOLOv5s	96.70	94.60	95.19	97.20
Faster RCNN（改进）	97.10	95.42	96.25	97.93

图 4-13 为在 MAR20 飞机遥感影像数据集上的部分实验结果，可以看出面对不同的目标，原始算法存在漏检误检的现象，而本文算法则可以很好的缓解这些现象，使得检测更加精准。

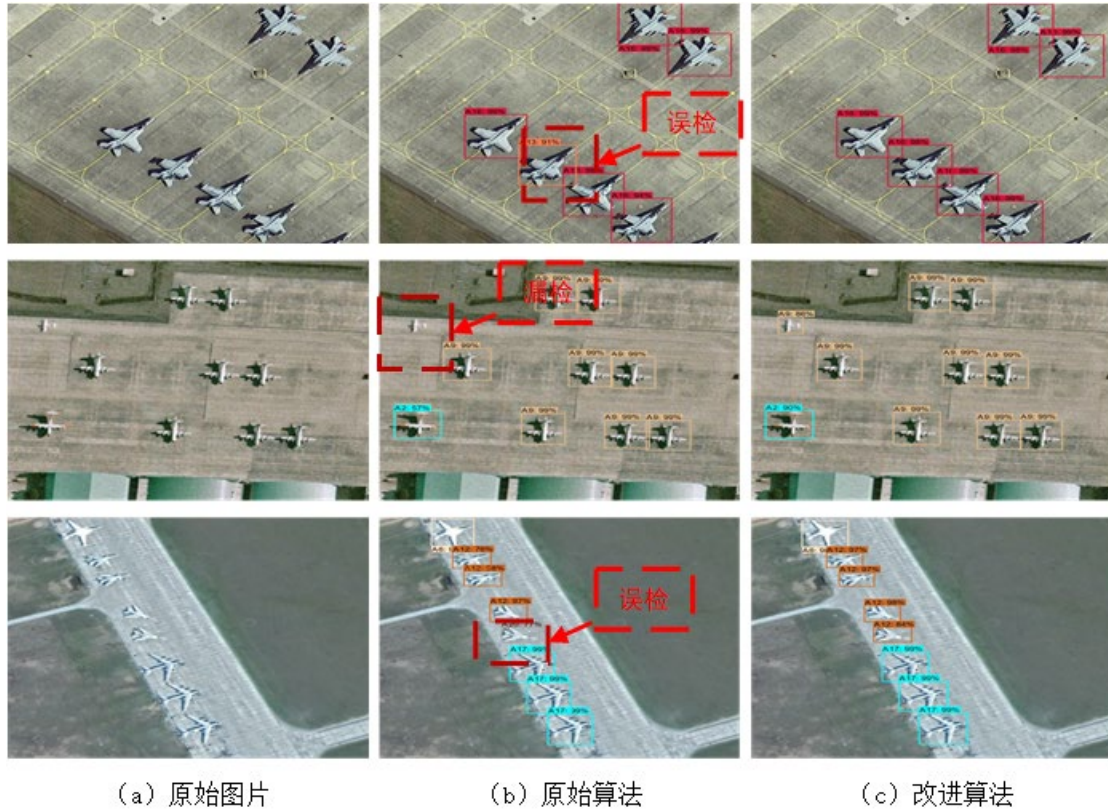


图 4-13 MAR20 数据集对比实验结果图

4.4 本章小结

本章主要针对 Faster RCNN 在遥感图像目标检测任务中存在的一些缺点，对 Faster RCNN 算法进行改进。首先，将原始主干网络替换为 ResNet50 网络，全面提升网络的特征提取能力。其次，为提升网络对遥感图像中目标特征的关注度，本章对 ResNet50 网络进行改进，在每层残差网络模组后加入 CBAM 注意力机制，这使得网络更加关注每层特征图中的特征，提升了网络在复杂背景下对目标的识别能力。最后，为提升网络对不同尺度目标的特征提取能力，本章在网络中引入 FPN 结构，这使得网络在检测不同大小目标时有着良好的表现，全面提高了模型对遥感图像目标的检测精度。除了理论性分析外，本章还通过在不同遥感影像数据集上进行多组实验来对改进算法进行验证，实验结果表明，改进算法在 NWPU VHR-10 数据集以及 MAR20 数据集中的表现良好，每一部分的改进相较于原始算法其检测精度都有不同程度的提升，同时与其他主流算法相比也有一定优势，并且改进算法在改善漏检与误检等方面都有很好的表现。

第 5 章 基于改进的 YOLOv5s 的遥感图像常见地表目标检测算法

5.1 引言

通过上一章的改进,基本使原始的 Faster RCNN 算法在遥感图像目标检测任务中有了不错的表现,改进后的算法虽然在检测精度上有所提升,但其使用的依旧是双阶段目标检测算法的框架,计算复杂,资源消耗大,这难以满足实际的使用需求。为解决这些问题,并进一步提升检测精度,本章设计了一种基于 YOLOv5s 的遥感图像目标检测的改进算法。YOLO 系列算法作为单阶段目标检测的代表算法,其网络的核心思想是将目标检测问题转化为回归问题,这类检测算法无需像双阶段目标检测算法那样进行候选框生成的操作,而是通过单一神经网络来同时对目标的类别、边界框和置信度进行预测,实现了端到端检测目标。正是因为如此,这类算法具有较高的检测速度以及较低的计算资源消耗,被广泛用于实时目标检测任务。然而该算法在处理遥感图像时依旧面临不小的挑战。

以本章所使用 YOLOv5s 算法为例,其在处理普通数据集时拥有较高的检测性能,但对于遥感图像依旧存在不足。YOLOv5s 的主干网络部分使用的依旧是传统的卷积神经网络架构,卷积核的感受野相对较小,这导致网络在提取特征时,难以捕捉到较小物体的细节。同时原始主干网络主要依赖于卷积层提取局部特征,这使得网络对全局特征的建模比较薄弱,导致模型有时难以区分背景与目标,致使模型整体性能下降。YOLOv5 算法的颈部网络主要是以“FPN+PANet”的形式组成,这种结构虽然可以融合多层特征图,但因信息传递路径单一,限制了特征融合效果,尤其在处理遥感图像中的多尺度目标时表现较差。最后,原始网络使用的损失函数是 CIoU,该损失函数虽然在大部分情况下可以满足对目标的定位要求,但对于遥感图像而言,目标往往较小,这会导致对目标边界框定位并不准确。本章针对这些问题,结合遥感图像的特点,设计了改进的 YOLOv5s 算法并通过在不同遥感影像数据集中设置多组实验验证其性能。

5.2 改进的 YOLOv5s 模型

5.2.1 主干网络的改进

在利用深度学习处理遥感图像目标检测任务时，算法的特征提取部分对于整个网络来说是非常重要的。而 YOLOv5s 算法的主干网络依旧采用传统的卷积神经网络来进行特征提取，这很大程度的限制了网络的性能，故本节针对遥感图像来对 YOLOv5s 网络的主干部分进行改进。

(1) C3 模块的修改

C3 模块是 YOLOv5s 算法主干网络的重要组成部分，其目的主要是为了特征提取以及多层次信息融合，主要由卷积层和 Bottleneck 模块组成，为防止出现梯度爆炸或者梯度消失，在 C3 模块内部往往采用残差连接。但由于遥感图像中包含大量背景信息和冗余特征，图像在经过 C3 模块时生成的特征图中包含很多其他信息，这使得网络在训练过程中难以关注到关键信息，进而对检测精度产生影响。为解决这一问题，同时验证上一章 CBAM 注意力模块的有效性，本节选择对 C3 模块进行改进，在 C3 模块中加入 CBAM 注意力机制模块，改进后的 C3 模块如图 5-1 所示。

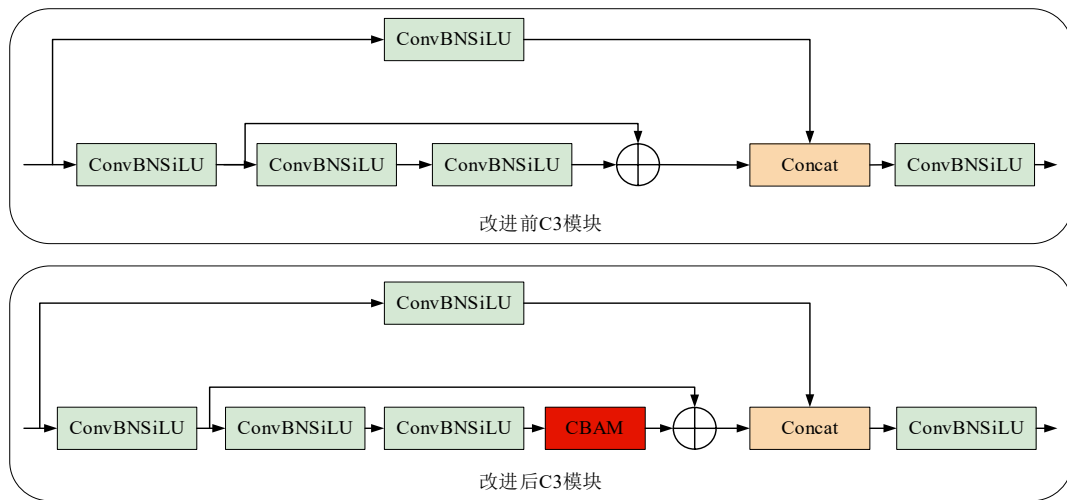


图 5-1 改进前后的 C3 模块对比

本文将 CBAM 注意力机制加入到 backbone 的 C3 模块中组成 C3_CBAM 模块，并将主干网络中所有的 C3 模块全部替换为 C3_CBAM 模块，改进后的模块与原始 C3 模块相比，在性能方面，由于增加了 CBAM 模块的缘故，提升了 C3

模块的特征表达能力。在计算效率方面，虽然加入 CBAM 模块会增加一些计算成本，但就提升的检测精度而言基本可以弥补这一点。在通用性方面，进一步验证了 CBAM 注意力机制在遥感图像目标检测任务中的有效性。

（2）Swin-Transformer 模块的加入

在 YOLOv5s 的主干网络中卷积层通常只关注局部特征，缺少跨区域的信息融合，对于遥感图像中尺度较大的目标或者结构复杂的目标，YOLOv5s 算法可能无法充分利用全局信息做出准确的预测。为解决这一问题，在主干网络的 C3_CBAM 模块与 SPPF 模块之间引入 Swin-Transformer 模块，图 5-2 为该模块的网络架构示意图。将经过 C3_CBAM 模块后的特征图输入到该模块中，经过归一化网络(Layer Normalization, LN)^[72]、多层感知机(Multilayer Perceptron, MLP)、基于窗口的自注意力机制(Window-base MSA, W-MSA)以及基于移动窗口的自注意力机制(Shifted Window-base MSA, SW-MSA)等模块进行输出，其中 W-MSA 将特征图分成不同大小窗口，处理不同窗口内像素的相关性，专注于局部信息。SW-MSA 通过循环位移以及掩码操作，恢复全局信息。

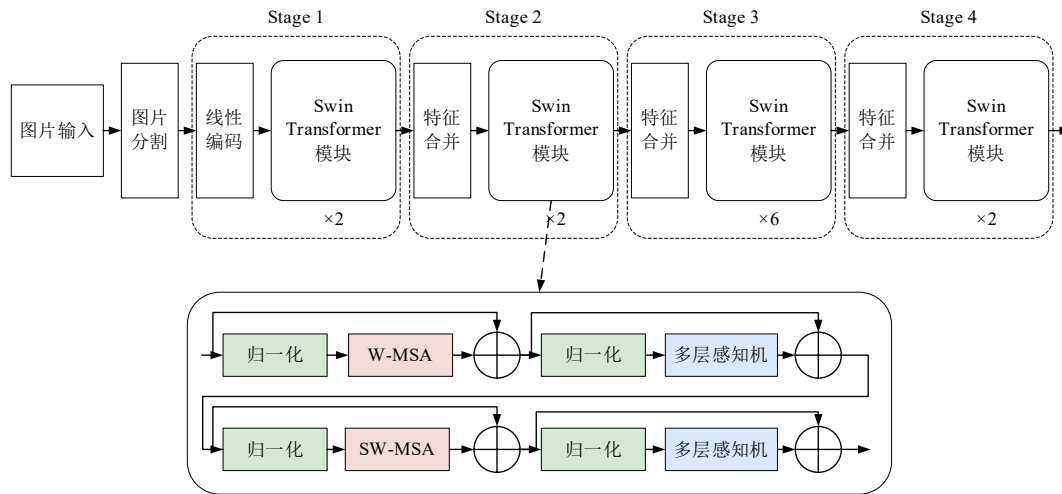


图 5-2 Swin-Transformer 模型架构

改进后的主干网络相较于原始网络优点在于：由于 Swin-Transformer 模块的加入，使得在其性能方面能够更加有效的捕获图片的语义信息以及空间关系，提升网络的感受野。在通用性方面，改进后的网络更加适合对遥感图像目标进行检测，且还具备较强的可迁移性。在计算效率方面，由于其将特征图进行窗口化，使得其计算复杂程度进一步降低，计算效率提高。

5.2.2 BiFPN 网络的引入

在 YOLOv5s 算法中颈部网络位于主干网络与输出层之间，其作用是对主干网络中提取的特征进行进一步融合，从而提高检测性能。但由于其单向信息流动的特点，限制了网络的性能，故本章借助 BiFPN 的思想对算法的颈部网络进行改进。BiFPN 是一种用于多尺度融合的网络结构，其结构主要通过设计自顶向下、自下向上的路径来进行双向融合，并通过删去部分对检测效果影响较小的连接来对网络进行简化网络，同时在相同尺度上对不同节点之间进行连接，该方式使得网络可以进行更高层次的特征融合。在原始 BiFPN 网络中，作者还通过对每个输入进行不同程度的加权，提网络的整体检测性能。本章利用该结构的特点，将原始颈部网络的第 13、第 17 和第 32 个模块替换为 BiFPN 网络中专有的 BiFPN_add2 模块，将第 20 个模块替换为 BiFPN_add3 模块，其中 BiFPN_add2 模块与 BiFPN_add3 模块的输入分别有两个输入与 3 个输入。同时结合上一节的改进，将第 6 层的 C3_CBAM 模块与第 14 层的 C3 模块、第 19 层的卷积层连接，进行特征融合。图 5-3 为 FPN、PANet 以及 BiFPN 的网络结构示意图

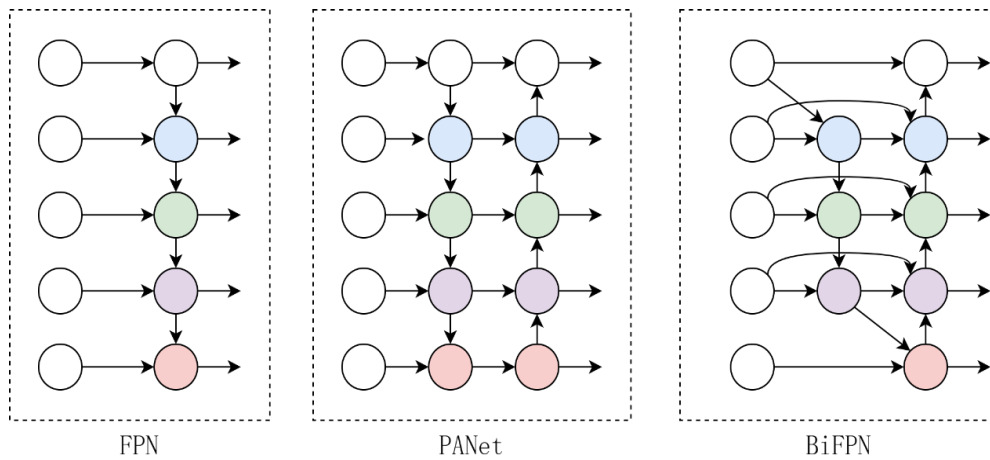


图 5-3 FPN、PANet 以及 BiFPN 的结构对比示意图

通过本章改进，与原始结构相比，改进算法在性能上，通过对不同层之间的交互，使得网络能够更精确的捕捉目标在不同尺度下的特征表现。在效率上，对不同层次的特征进行充分利用，提高网络在多尺度融合的效率。在通用性上，通过该方式使得网络在处理不同特征有着良好的表现，使得网络在处理复杂背景时更加有效。

5.2.3 损失函数的优化

在目标检测任务中，尤其在遥感图像目标检测中，确定目标位置以及尺度是非常重要的。然而，对于遥感图像而言，复杂的背景以及图片分辨率等问题使得对目标尺度以及定位要更加困难。在深度学习中，常见的损失函数有 EIoU^[73]、CIoU、DIoU、GIoU^[74]等，其中原始网络采用的是 CIoU 损失函数，该损失函数是 DIoU 函数的基础上进行改进，由于 DIoU 并未考虑到框的横纵比对 Bounding box 的回归的影响，CIoU 根据 DIoU 进行框的宽高比参数的引入，很好解决了 DIoU 的不足，其公式如式(5-1)所示：

$$L_{CIoU} = 1 - IoU + \frac{|\rho^2(B, B^{gt})|}{c^2} + \beta v \quad (5-1)$$

其中参数 B 与 Bgt 分别表示的是预测框与真实框，c 表示预测框与真实框最小的外接矩形的对角线距离，β 为权重函数，v 用来对长宽比相似性进行度量，其公式分别如式(5-2)和式(5-3)所示：

$$\beta = \frac{v}{(1 - IoU) + v} \quad (5-2)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^t} - \arctan \frac{\omega}{h} \right)^2 \quad (5-3)$$

但 CIoU 损失函数本身也存在缺点，即其虽然长宽比相同，但其宽高比不同，惩罚项仍为零，这样依旧会阻碍模型的优化。为解决该问题，选用的 alpha-IoU 损失函数通过引入一个参数 α 来根据实际的网络进行合理的调节，这个参数可以对网络进行优化，加快梯度的收敛速度，有效提升网络的性能。其公式如下式(5-4)所示：

$$\begin{aligned} L_{CIoU} &= 1 - IoU + \frac{|\rho^2(B, B^{gt})|}{c^2} + \beta v \\ \Rightarrow L_{(\alpha-CIoU)} &= 1 - IoU^\alpha + \frac{\rho^{2\alpha}(B, B^{gt})}{c^{2\alpha}} + (\beta v)^\alpha \end{aligned} \quad (5-4)$$

5.2.4 改进后的网络结构

根据上述的改进，从而得出最终本章基于 YOLOv5s 算法的改进模型，其结构如图 5-4 所示，之后将通过西北工业大学遥感影像数据集与 MAR20 遥感飞机影像数据集进一步验证该算法对于遥感图像目标检测任务的可行性。

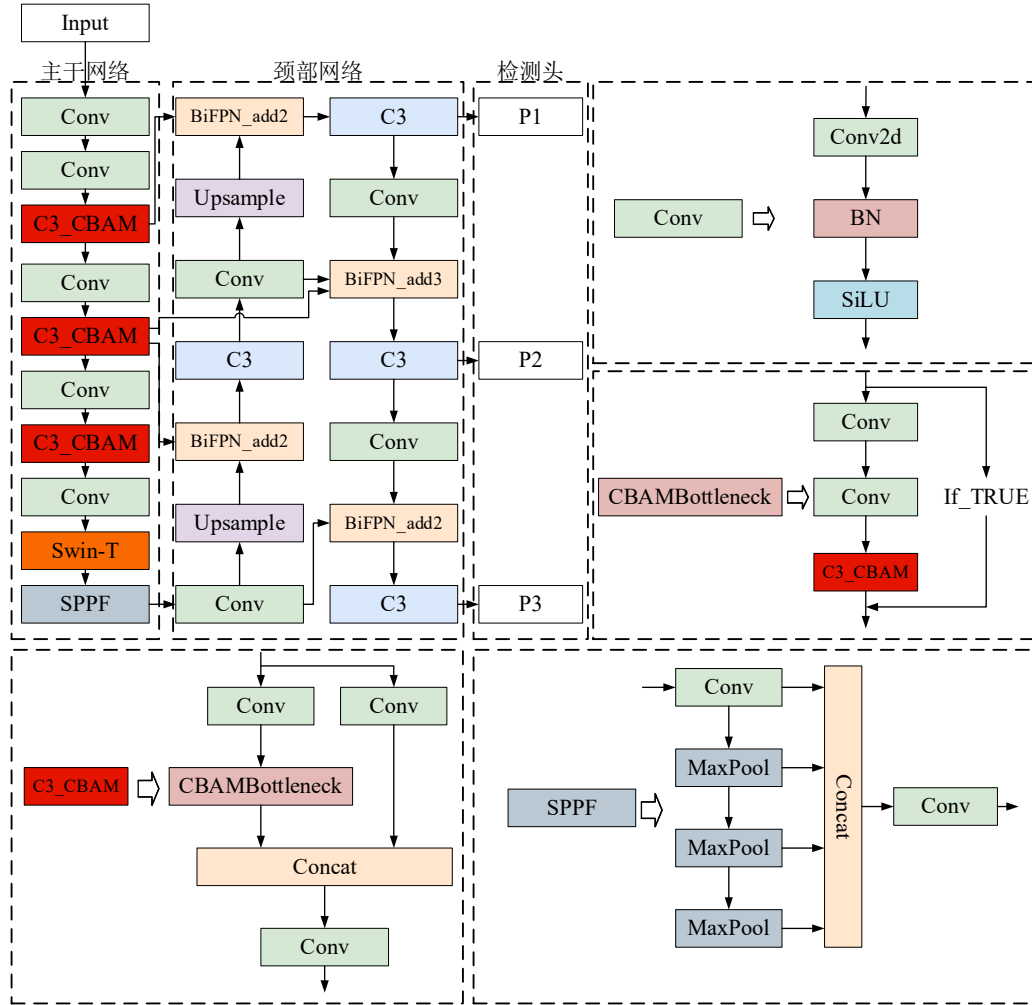


图 5-4 改进后 YOLOv5s 结构示意图

5.3 实验结果与分析

5.3.1 实验平台

对于本章采用的实验环境与实验设备，与 4.3.1 章节的实验平台相同。

5.3.2 实验设计

本次实验依旧采用 NWPU VHR-10 数据集以及 MAR20 遥感军用飞机数据集对网络进行验证，对于消融实验与对比实验的设置与 4.3.2 节相同。由于 YOLOv5s 网络的特点，在参数设置上与 4.3.2 节略有不同。具体设置如表 5-1 所

示。

表 5-1 YOLOv5s 模型超参数设置一览表

超参数名称	参数设置
batch size	16
workers	8
初始学习率	0.01
epoch	150
优化器	SGD

本章由于 YOLOv5s 对资源消耗较少，为减少训练所需的时间，提高整体实验的效率，故选择提升网络的 batch size 以及 workers，分别设置为 16 与 8，该设置对最终实验结果影响较小。为更好的对比实验结果，本章所有实验均未使用预训练权重进行辅助训练。

5.3.3 消融实验结果与分析

(1) 参数量以及计算量实验结果与分析

表 5-2 参数量以及计算量结果表

实验	模型	FLOPS/G	Params/M
1	Faster RCNN（原始）	215.59	43.92
2	Faster RCNN（改进）	134.53	41.40
3	YOLOv5s（原始）	7.04	16.00
4	YOLOv5s（改进）	7.21	19.00

由表 5-2 可以看出，本章改进后的算法，相对于本文第四章改进后的 Faster RCNN 算法在计算量以及参数量方面都有了非常大的改善。虽然相对于原始 YOLOv5s 算法，本章改进算法的计算量以及参数量有所增加，但其对遥感图像目标检测能力却得到了不错的提升。这表明本章的改进算法对资源的需求较少，同时，较少的计算量也会节省的训练以及推理的时间。

(2) NWPU VHR-10 数据集消融实验与分析

本文为验证改进算法的有效性，其在 NWPU VHR-10 数据集上的具体检测结果如表 5-3 所示。

表 5-3 NWPU VHR-10 数据集消融实验结果表

实验	CBAM	Swin-T	BiFPN	alpha-IoU	mAP@0.5/%	mAP@.5:.95/%
1	-	-	-	-	84.10	46.50
2	√	-	-	-	85.10	46.50
3	-	√	-	-	86.40	50.70
4	-	-	√	-	85.80	51.50
5	-	-	-	√	86.00	51.20
6	√	√	-	-	86.30	51.10
7	√	√	√	-	87.00	51.10
8	√	√	√	√	87.80	51.70

由表 5-3 中数据可以得知, 本文算法在 NWPU VHR-10 数据集的 mAP@0.5 可以达到 87.80%, 相较于原始算法的 84.10%提升了 3.7 个百分点, 而其 mAP@.5:.95 则为 51.70%, 相较于原始算法的 46.50%提升了足足有 5.2 百分点, 这说明本文算法更加满足日常遥感图像目标检测精度要求。而表 5-4 则详细的记录了本文算法对每一种类型的飞机检测精度。

表 5-4 每类目标检测结果对比表

目标种类	mAP@0.5/%		
	原始算法	本章算法	对比结果
飞机	99.5	99.5	-
船舶	81.7	82.9	+1.2 ↑
储罐	96.2	98.3	+2.1 ↑
棒球场	98.2	99.1	+0.9 ↑
网球场	68.2	70.6	+2.4 ↑
篮球场	67.8	73.0	+5.2 ↑
田径场	98.3	97.9	-0.4 ↓
港口	79.9	90.3	+10.4 ↑
桥	79.6	93.4	+13.8 ↑
交通工具	71.4	73.1	+1.7 ↑

由表 5-4 数据可以看本文算法在飞机目标的检测上其 $mAP@0.5$ 最高，为 99.5%，其中 $mAP@0.5$ 提升最多的为桥梁，从 79.6%提升到了 93.4%，提升了有 13.8 个百分点，其他类型目标绝大部分均有不同程度的提升，有效的提升了各个目标检测的精度。

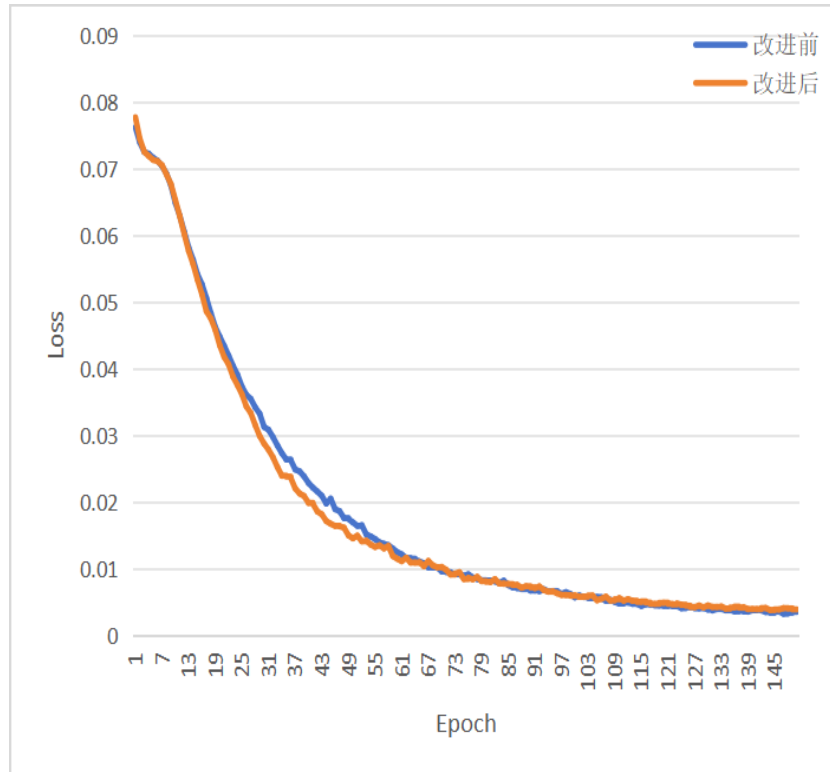


图 5-5 NWPU VHR-10 数据集 loss 曲线对比图

图 5-5 为 YOLOv5s 算法改进前后的训练 Loss 曲线的对比图，图中可以看到无论是原始算法还是改进算法，在训练时的 Loss 值都较低，并且在曲线收敛后 Loss 值相差不大，均趋近于一个非常低的数值，这说明虽然本文算法经过改进增加了计算量以及参数量，但在训练过程中依旧有着不错的表现。同时，在图中可以看到约在 120 轮后，改进后曲线下降缓慢，甚至略高于改进前的算法，这可能由于随着训练轮数增加，改进后的模型逐渐开始关注更复杂的特征分布，优化难度增大，使得 Loss 下降趋势减缓，甚至在 120 轮后略高于改进前，但这却提升了整体模型的泛化能力，使得模型对目标检测的准确度得到提升。

(3) MAR20 飞机遥感数据集消融实验与分析

为进一步检验改进后算法的效果以及适用性，同样在 MAR20 数据集进行消融实验，其结果如表 5-5 所示。

表 5-5 MAR20 数据集消融实验结果表

实验	CBAM	Swin-T	BiFPN	alpha-IoU	mAP@0.5/%	mAP@.5:.95/%
1	-	-	-	-	97.20	76.90
2	√	-	-	-	97.60	77.20
3	-	√	-	-	97.80	77.20
4	-	-	√	-	97.40	76.90
5	-	-	-	√	97.30	77.30
6	√	√	-	-	97.90	77.10
7	√	√	√	-	98.00	77.00
8	√	√	√	√	98.10	79.10

由表 5-5 中数据可以得知，本文算法在 MAR20 数据集的 mAP@0.5 可以达到 98.10%，相较于原始算法的 97.20%提升了 0.9 个百分点，而其 mAP@.5:.95 则为 79.10%，相较于原始算法的 76.90%提升了足足有 2.2 百分点，这说明本文算法更加满足日常遥感图像目标检测精度要求。而表 5-6 则详细的记录了本文算法对每一种类型的飞机检测精度。

表 5-6 MAR20 数据集中每类目标检测结果对比表

种类	mAP@0.5/%			种类	mAP@0.5/%		
	改前	改后	对比		改前	改后	对比
A1	94.8	97.7	+2.9 ↑	A11	95.9	96.8	+0.9 ↑
A2	99.4	99.5	+0.1 ↑	A12	99.5	99.5	-
A3	99.0	98.9	-0.1 ↓	A13	94.9	96.2	+1.3 ↑
A4	98.9	98.0	-0.9 ↓	A14	98.5	98.4	-0.1 ↓
A5	93.9	97.1	+3.2 ↑	A15	83.6	91.8	+8.2 ↑
A6	99.4	99.4	-	A16	99.2	99.2	-
A7	98.7	99.3	+0.6 ↑	A17	99.5	99.5	-
A8	99.5	99.5	-	A18	96.4	96.9	+0.5 ↑
A9	99.4	99.5	+0.1 ↑	A19	95.0	96.0	+1.0 ↑
A10	99.3	99.5	+0.2 ↑	A20	99.3	99.2	-0.1 ↓

从表 5-6 中可以看出,改进方法在大多数类别上均取得了精度提升。其中,最为显著的是 A16 类,由 83.6%提升至 91.8%,提升幅度高达 8.2%,说明改进方法对原本检测性能较弱的类别具有明显的优化效果。部分类别在改进后略有下降,例如 A3 和 A4 等,这可能是由于该类别样本数量不足、特征分布较为集中或受噪声影响导致。但总体而言,下降幅度较小,对整体性能影响不大。

综上所述,本文的改进策略在 MAR20 数据集上展现出良好的检测效果,该结果充分说明所提出方法不仅提升了模型的整体检测精度,同时增强了对复杂目标和多样化场景的适应能力,具有较强的推广与应用价值。

图 5-6 为算法改进前后在 MAR20 数据集中的训练 Loss 曲线的对比图,可以从图中看出,本章改进算法在训练过程中 Loss 函数也处于缓慢下降状态,最终基本保持不变,同时,在训练过程中 Loss 的值要低于原始算法,这更加说明,在复杂场景下,本章改进算法面对遥感图像目标检测任务时有着更好的表现。

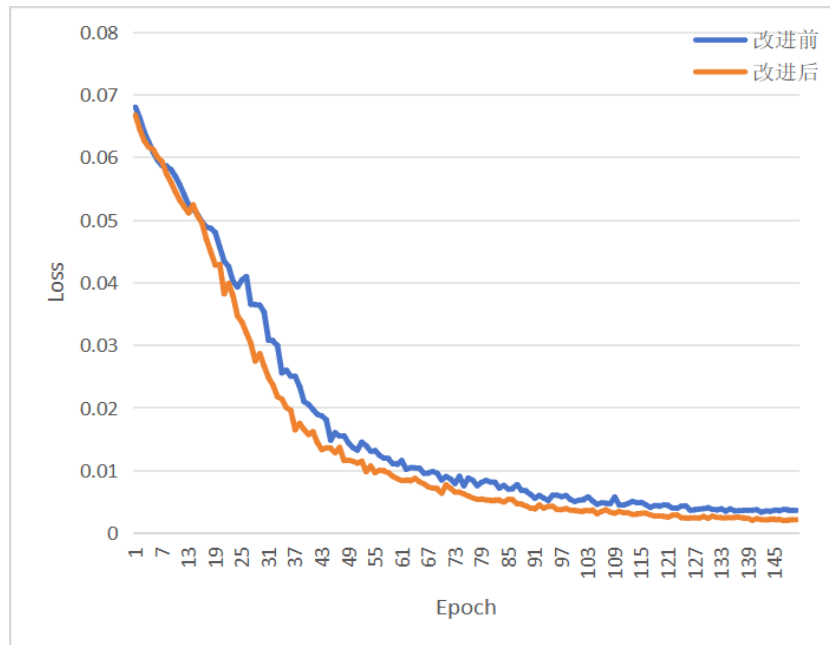


图 5-6 MAR20 数据集 loss 曲线对比图

5.3.4 对比实验结果与分析

(1) NWPU VHR-10 数据集对比实验与分析

本章除了通过算法的消融实验来验证算法的有效性,依旧还通过一系列对比实验来进一步对算法的有效性进行验证。本次实验在 4.3.4 节对比实验的基础上增加本章实验的改进算法来进行对比。

表 5-7 NWPU VHR-10 数据集对比实验结果表

实验	Precision/%	Recall/%	F1_score/%	mAP@0.5/%
SSD	91.88	57.08	70.42	77.30
Faster RCNN (原始)	79.52	75.43	77.44	75.97
YOLOv3	84.76	70.32	76.87	84.58
YOLOv5n	93.00	78.50	85.14	85.40
YOLOv5s (原始)	86.50	80.70	83.50	84.10
Faster RCNN (改进)	85.73	83.65	85.16	86.66
YOLOv5s (改进)	91.00	83.50	85.47	87.09

表 5-7 是在 NWPU VHR-10 数据集上进行对比实验的结果,从表中可以看出,与第四章改进的 Faster RCNN 相比,本章改进算法在准确率、召回率、F1_score 以及 mAP@0.5 指标上均有一定提升,同时,除准确率比 SSD 算法低外,与其他经典算法相比都有一定的优势。这说明了本文算法在处理遥感图像目标检测任务是有着更加优秀的表现。

图 5-7 为 NWPU VHR-10 飞机遥感影像数据集上的部分实验结果,可以看出面对不同的目标,原始算法存在漏检误检的现象,而本文算法则可以很好的缓解这些现象,使得检测更加精准。

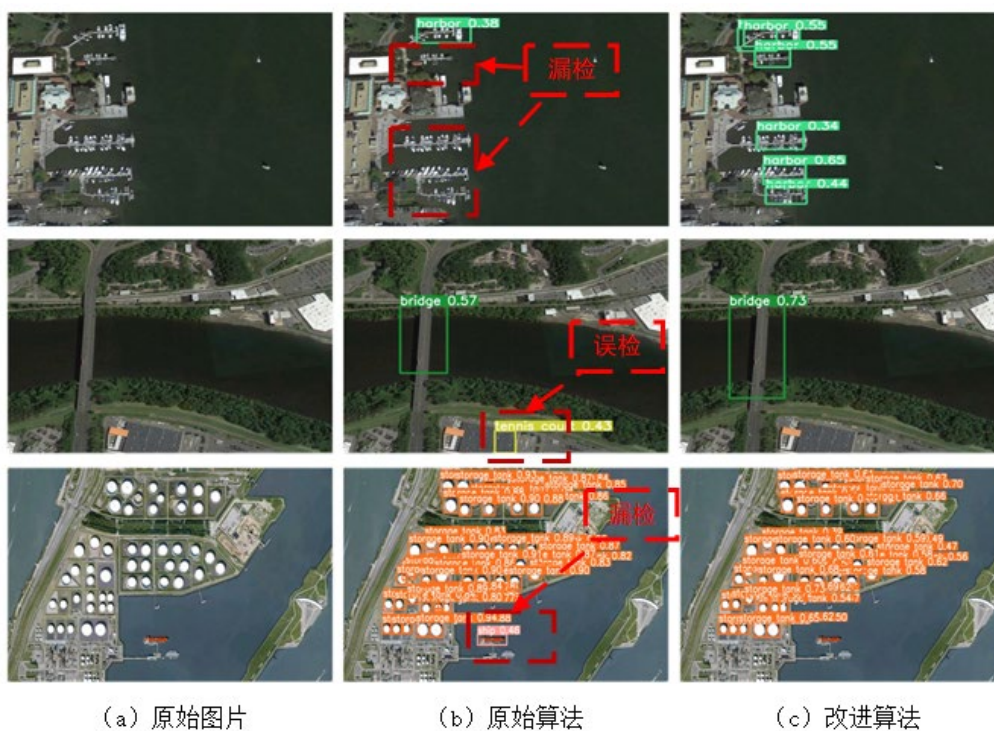


图 5-7 NWPU VHR-10 数据集对比实验结果图

(2) MAR20 数据集对比实验与分析

表 5-8 MAR20 数据集对比实验结果表

实验	Precision/%	Recall/%	F1_score/%	mAP@.5/%
SSD	97.95	87.62	89.59	94.46
Faster RCNN (原始)	95.56	91.63	93.55	94.93
YOLOv3	77.85	45.30	51.30	63.34
YOLOv5n	95.40	92.40	93.88	96.60
YOLOv5s (原始)	96.70	94.60	95.19	97.20
Faster RCNN (改进)	97.10	95.42	96.25	97.93
YOLOv5s (改进)	98.00	95.50	96.73	98.10

表 5-8 是在 MAR20 数据集上进行对比实验的结果，同样的由表可以看出，改进算法在 MAR20 数据集上的准确率、召回率、F1 及 mAP@0.5 等指标上都有不同程度的提升，特别地，mAP@0.5/%与其他算法相比有较明显的提升。

下图为在 MAR20 飞机遥感影像数据集上的部分实验结果，可以看出面对不同的目标，原始算法存在漏检误检现象，而本文算法则可以很好的缓解这些现象，使得检测更加精准。

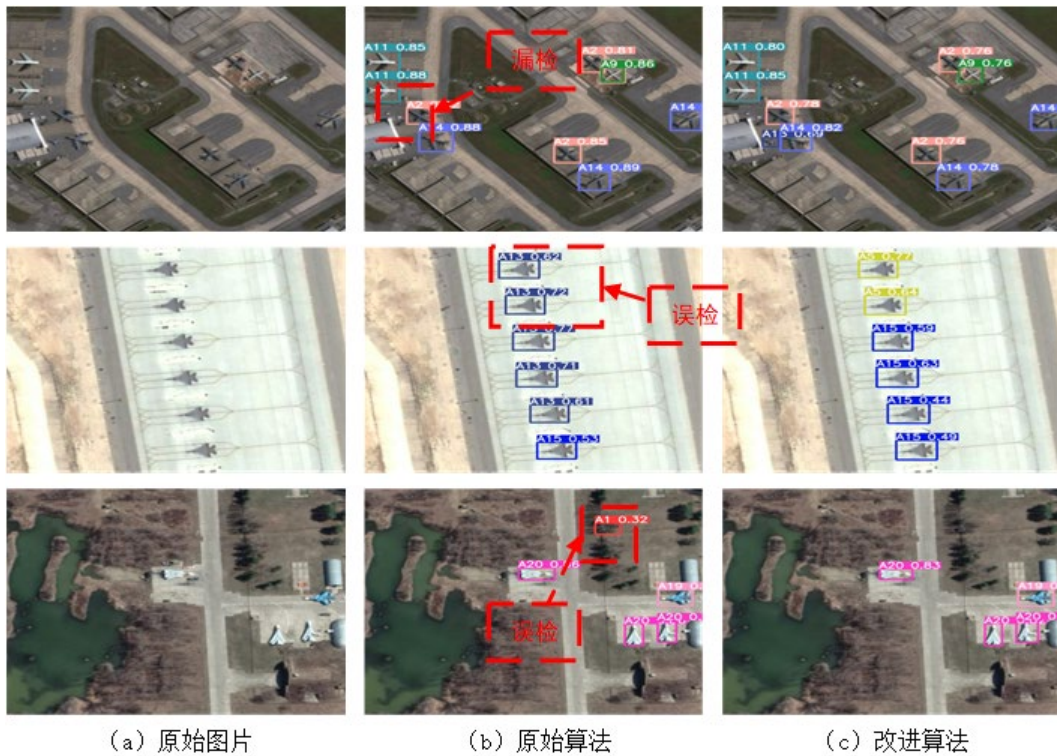


图 5-8 MAR20 数据集对比实验结果图

通过两种数据集的消融实验与对比实验所显示的结果可以得出，本文基于 YOLOV5s 的改进算法相对于原始算法性能有较明显提升，并且相较于其他改进算法也存在一定的优势，对遥感图像领域目标检测具有一定的现实意义。

5.4 本章小结

本章考虑到实际应用情况，在减少资源消耗以及计算复杂程度的同时，进一步提升深度学习算法在处理遥感图像目标检测任务中的性能，对原始 YOLOv5s 算法进行改进。首先，为全面提升网络的特征提取能力，本章选择对 YOLOv5s 算法的主干网络进行改进，在 C3 模块中加入 CBAM 注意力机制提升对目标特征的关注度，同时在主干网络中加入 Swin-Transformer 模块加强全局与局部特征的联系。其次，通过借助 BiFPN 思想改变算法颈部网络的结构来对网络进行优化，促进更高层次的特征融合能力，提升网络对遥感图像复杂场景中目标的检测能力。之后，为更加精准定位遥感图像中的目标，本章通过优化损失函数，提高网络的检测性能。最后，在两个不同的遥感影像数据集中设置多组实验进行验证，实验结果表明，本章算法对资源消耗较少，计算复杂程度也较低，同时，每部分改进在提升网络性能上都有一定效果，与其他算法相比也具备一定优势，充分验证了本章算法的有效性。

第 6 章 总结与展望

6.1 总结

遥感图像目标检测任务作为遥感图像研究的重要领域之一，在军事、农业以及城市规划等领域中均发挥着重要作用。但遥感图像却有着其自身的特点，如数据获取困难、背景复杂以及目标尺度大小不一等特点，使得对于该方面的研究颇具挑战。因此，本文选择对遥感图像目标检测领域进行研究，以遥感图像中常见的地表目标作为研究对象，利用公开数据集 NWPU VHR-10 数据集以及 MAR20 遥感军用飞机数据集对深度学习中常见的两种双阶段与单阶段目标检测算法进行分析与优化。最后设置多组实验对算法进行验证，证明本文改进算法的有效性。本文所做的研究工作总结如下：

（1）本文对遥感图像目标检测领域的国内外研究现状进行具体分析，并对深度学习算法基本理论部分进行了阐述，介绍了几种较为经典的双阶段与单阶段目标检测算法，分析了它们的优缺点。本文还对遥感影像数据集进行了介绍，分析了其特点，介绍了本文所使用的两种数据集，并对其进行预处理。采用数据增强的方式对数据集进行扩充，增加样本的多样性，提高模型鲁棒性。同时，对遥感图像目标检测中常见的评价指标进行了介绍。

（2）针对遥感图像的特点，结合双阶段目标检测算法缺点，对 Faster RCNN 算法进行改进。针对原始 Faster RCNN 算法主干网络对遥感图像目标特征提取能力不足的问题，本文将原始网络的 VGG16 网络替换为特征提取能力更强、效率更高的 ResNet50 网络。为提升网络对遥感图像目标的关注度，增强对小尺度目标的检测性能，本文对 ResNet50 网络结构进行修改，在原始结构每个残差模组后加入 CBAM 注意力机制，不仅提升了网络的性能，更带来了精度上的收益。为处理遥感图像中目标尺度变化大导致的检测精度低的情况，在原始网络中加入特征金字塔网络，充分利用每层特征图，提升网络的多尺度表达能力，进一步提升网络在遥感图像目标检测任务中的性能。最后实验验证，在消融实验中，改进算法在 NWPU VHR-10 数据集中 $mAP@0.5$ 提升了 10.69%，在 MAR20 数据集中

mAP@0.5 提升了 3.00%，同时，相对与其他算法也有一定优势，充分说明了改进算法的有效性。

(2) 针对深度学习算法在完成遥感图像任务时出现的占用资源大、计算复杂等问题，同时进一步提升目标检测的精度，本文还选用 YOLOv5s 算法，并对针对遥感图像特点对其进行改进。对算法的主干部分进行修改，通过加入 CBAM 注意力机制以及 Swin-Transformer 模块提升网络对感兴趣区域的关注以及增强局部与全局特征关联。另外，对算法的颈部网络也进行修改，通过增加不同层之间连接，对特征图进行充分利用，提升算法对遥感图像多尺度目标的检测能力。在最后，为提升 YOLOv5s 算法对遥感图像目标的定位能力，本文对算法的损失函数进行优化，将原始算法替换为更加先进 α -IoU，是网络可以更加准确对目标的位置与大小进行判断。最后对改进算法进行实验验证，改进算法相较于改进的 Faster RCNN 算法其计算复杂程度降低了 127.32GFLOPS，参数量减少了 22.4M。同时，在消融实验与对比实验中，改进算法在 NWPU VHR-10 数据集中 mAP@0.5 提升了 3.7%，在 MAR20 数据集中 mAP@0.5 提升了 0.9%，与本文改进的 Faster RCNN 算法相比，也具有一定优势，这充分说明了本文算法的有效性。

6.2 展望

本文主要针对遥感图像中背景复杂、目标特征不明显以及尺度变化较大等特点造成的检测精度低的问题，对深度学习中目标检测算法进行改进，虽然在检测效果上相对于原始算法以及其他经典算法上有所提升，但由于设备与时间等原因，本文的研究依旧有很大的优化空间，在未来的研究工作中可以从以下几方面考虑：

(1) 优化与扩充数据集。对于遥感图像而言，在进行检测时，目标的方向往往也会成为检测的重点，然而现在的遥感影像数据大多只包含传统的矩形框信息，并不包含对目标方向的描述。为了满足越来越复杂的遥感图像目标检测的要求，在采集和标注数据集时，增加目标的方向信息是必不可少。同时，遥感影像的采集与标注工作往往十分困难，需要大量的人力以及物力，故需要对现有的数据进行扩充，除了使用本文采用的数据增强的方式。还可使用生成对抗网络，来

生成大量的高质量的遥感图像样本，丰富数据集，增加算法的鲁棒性。

（2）网络的轻量化。在本文第四章时，虽然通过改进提升了网络对遥感图像的检测性能，但不可避免的产生了额外的计算开销，为此，下一步可以通过对算法进行轻量化，采用更加轻便、有效的主干网络，例如 MobileNetV3 等网络。除此之外，还可以对网络进行知识蒸馏，减少资源消耗，提升网络模型的推理速度。

（3）部署与实际应用。在本文的第五章，针对遥感图像特点对 YOLOv5s 算法进行优化后，算法在资源占用、推理速度与检测精度之间达到了较好的平衡。下一步可以对算法进行实际的部署，将优化后的算法移植到无人机等设备上，提升遥感影像目标检测智能化水平，为相关实际应用提供更加强大的技术支持。

参考文献

- [1] 张尧,张艳,王涛,等.大场景SAR影像舰船目标检测的轻量化研究[J].地球信息科学学报,2025,27(01):256-270.
- [2] WELLMANN T, LAUSCH A, ANDERSSON E, et al. Remote sensing in urban planning: Contributions towards ecologically sound policies?[J]. Landscape and urban planning, 2020, 204: 103921.
- [3] ASADZADEH S, DE OLIVEIRA W J, DE SOUZA FILHO C R. UAV-based remote sensing for the petroleum industry and environmental monitoring: State-of-the-art and perspectives[J]. Journal of Petroleum Science and Engineering, 2022, 208: 109633.
- [4] 管含宇,凌云,汪舒磊.单阶段安全帽检测深度学习算法综述[J].计算机工程与应用,2024,60(16):61-75.
- [5] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553):436-444.
- [6] 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, 2016, 39(6): 1137-1149.
- [7] XU Y P, WU X, WANG L, et al. HOFA-Net: A High-Order Feature Association Network for Dense Object Detection in Remote Sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 17: 1513-1522.
- [8] XU L, YU J, MING D. HODet: A New Detector for Arbitrary-Oriented Rectangular Object in Optical Remote Sensing Imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 2918-2926.
- [9] 陈清江,李璐.基于改进YOLOv8 的遥感小目标检测算法[J/OL].激光与光电子学进展,2025: 1-18

- [10] WANG C Y, YEH I H, MARK LIAO H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2024: 1-21.
- [11] 潘宏刚,李员禄,郭宝仁,等.基于卷积神经网络的转子系统故障诊断与研究[J].汽轮机技术,2024,66(02):145-148.
- [12] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: A brief review[J]. Computational intelligence and neuroscience, 2018, 2018(1): 7068349.
- [13] CHOWDHARY K R, CHOWDHARY K R. Natural language processing[J]. Fundamentals of artificial intelligence, 2020: 603-649.
- [14] ATAKISHIYEV S, SALAMEH M, YAO H, et al. Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions[J]. IEEE Access, 2024,12:1603-1620.
- [15] 徐向阳.基于有限元模拟的机器学习触觉材质识别算法研究[D].哈尔滨:哈尔滨工业大学,2022.
- [16] LI Z, LIU F, YANG W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. IEEE transactions on neural networks and learning systems, 2021, 33(12): 6999-7019.
- [17] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: An overview[J]. IEEE signal processing magazine, 2018, 35(1): 53-65.
- [18] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE transactions on neural networks, 1994, 5(2): 157-166.
- [19] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 32(9): 1627-1645.
- [20] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.

- [21] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.
- [22] 张蓬鹤,杨艺宁,王璧成,等.基于D-DADA数据增广与DBE-YOLO网络的电表异常检测方法[J/OL].计算机工程 2025,1-13.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [24] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015,38(1): 142-158.
- [25] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and their applications, 1998, 13(4): 18-28.
- [26] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Santiago: IEEE ,2015: 1440-1448.
- [27] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands: Springer International Publishing, 2016: 21-37.
- [28] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 779-788.
- [29] ZHANG Y, SONG C, ZHANG D. Deep learning-based object detection improvement for tomato disease[J]. IEEE access, 2020, 8:56607-56614.
- [30] ZOU X, WU C, LIU H, et al. An accurate object detection of wood defects using an improved Faster R-CNN model[J]. Wood Material Science & Engineering, 2024: 1-7.
- [31] 马宇,单玉刚,袁杰.基于改进Faster R-CNN的遥感目标检测算法[J].现代电子技术,2022,45(03):58-63.

-
- [32] TAN L, WU H, XU Z, et al. Multi-object garbage image detection algorithm based on SP-SSD[J]. Expert Systems with Applications, 2025, 263: 125773.
 - [33] 孙仁科, 营鹏, 李仲年, 等. 基于轻量化SSD的弱小目标检测[J]. 计算机仿真, 2024, 41(10): 355-361.
 - [34] WANG R J, LI X, LING C X. Pelee: A real-time object detection system on mobile devices[J]. Advances in neural information processing systems, 2018, 31: 1967-1967.
 - [35] KARTHIKEYAN M, SUBASHINI T S, SRINIVASAN R, et al. YOLOAPPLE: Augment Yolov3 deep learning algorithm for apple fruit quality detection[J]. Signal, Image and Video Processing, 2024, 18(1): 119-128.
 - [36] FARHADI A, REDMON J. Yolov3: An incremental improvement[C]//Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
 - [37] TANG S, ZHANG S, FANG Y. HIC-YOLOv5: Improved YOLOv5 for small object detection[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024: 6614-6619.
 - [38] 王林枫, 左云波, 徐小力, 等. 改进YOLOv4-tiny网络的日用商品目标检测算法[J]. 计算机应用与软件, 2024, 41(11): 319-326+365.
 - [39] PINCA R C B. Development of Real-Time Detection of Philippine Traffic Signs Using YOLOv4-Tiny[C]//2024 IEEE 6th Symposium on Computers & Informatics (ISCI). IEEE, 2024: 218-223.
 - [40] GAO W, ZONG C, WANG M, et al. Intelligent identification of rice leaf disease based on YOLO V5-EFFICIENT[J]. Crop Protection, 2024, 183: 106758.
 - [41] 柯静. 基于无锚框的遥感图像目标检测研究[D]. 西安: 西安电子科技大学, 2023.
 - [42] SHI F, LIU Y, WANG H. Target Detection in Remote Sensing Images Based on Multi-scale Fusion Faster RCNN[C]//2023 35th Chinese Control and Decision Conference (CCDC). Yi Chang :IEEE, 2023: 4043-4046.

-
- [43] 凌晨,张鑫彤,马雷.基于Mask R-CNN算法的遥感图像处理技术及其应用[J].计算机科学,2020,47(10):151-160.
 - [44] K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
 - [45] XU S, CHEN Z, ZHANG H, et al. Improved remote sensing image target detection based on YOLOv7[J]. Optoelectronics Letters, 2024, 20(4): 234-242.
 - [46] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.
 - [47] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 16519-16529.
 - [48] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
 - [49] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.
 - [50] 黄瑶.结合MobileNetV2 与YOLOv5s的轻量级遥感目标检测方法研究[J].经纬天地,2024,(02):5-8.
 - [51] 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.
 - [52] 张亮.基于遥感图像和改进YOLOv5s的舰船识别技术[J].舰船电子对抗,2024,47(02):67-70.
 - [53] XU X, LI Q, PAN J, et al. ESOD-YOLO: an enhanced efficient small object detection framework for aerial images[J]. Computing, 2025, 107(2): 1-19.

- [54] 张新君,赵春霖.改进的YOLOv5s遥感影像机场场面飞机小目标识别[J].电光与控制,2024,31(07):104-111.
- [55] GAO T, XIA S, LIU M, et al. Msnet: Multi-scale network for object detection in remote sensing images[J]. Pattern Recognition, 2025, 158: 110983.
- [56] 陈欣,万敏杰,马超等.采用多尺度特征融合SSD的遥感图像小目标检测[J].光学精密工程,2021,29(11):2672-2682.
- [57] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- [58] 王玺坤,姜宏旭,林珂玉.基于改进型YOLO算法的遥感图像舰船检测[J].北京航空航天大学学报,2020,46(06):1184-1191.
- [59] 李志昂,肖小玲,周绍发.基于改进的YOLOv5s小目标船舶遥感图像检测[J].科学技术与工程,2025,25(02):657-666.
- [60] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [61] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [62] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [63] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [64] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

- [65] HE J, ERFANI S, MA X, et al. α -IoU: A family of power intersection over union losses for bounding box regression[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 20230-20242.
- [66] SU H, WEI S, LIU S, et al. HQ-ISNet: High-quality instance segmentation for remote sensing imagery[J]. *Remote Sensing*, 2020, 12(6): 989.
- [67] 禹文奇,程堃,王美君,等. MAR20: 遥感图像军用飞机目标识别数据集[J]. *遥感学报*, 2023, 27(12): 2688-2696.
- [68] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. The city of Saline Lake: IEEE, 2018: 8759-8768.
- [69] 陈艳. 面向混合时空网络的交通流预测方法[D]. 贵州民族大学, 2024.
- [70] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 7132-7141.
- [71] 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*. 2020: 11534-11542.
- [72] LEI BA J, KIROUS J R, HINTON G E. Layer normalization[J]. *ArXiv e-prints*, 2016: arXiv: 1607.06450.
- [73] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.
- [74] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seoul: IEEE/CVF 2019: 658-666.

攻读学位期间的学术成果

- [1] 王喜辰, 诸志龙, 赵发. 基于YOLOv5 的遥感图像常见地物目标检测与识别[J]. 安徽工程大学学报, 2025, 40 (5) . (已录用)

致 谢

时光匆匆，三年的研究生生涯即将结束，在此期间，有过迷茫与焦虑，也有欢笑与收获，在本论文完成的过程中，我得到过许多人的帮助，在此，我向所有关心和帮助过我的人表示衷心的感谢。

首先，我要感谢我的导师诸志龙老师，在论文的选题、研究思路和写作过程中，导师给予了我悉心的指导与宝贵的建议，使我能够顺利的完成毕业论文的撰写。导师深厚的学术造诣以及严谨的治学态度让我受益匪浅。

其次，感谢课题组的老师和同学们，他们在研究过程中给了我很多的指导与帮助，尤其是胡程、李志晋和余同阳三位学长，在实验设置、数据分析以及论文修改等方面给予我很多有益的建议与帮助。同时，我要感谢安徽工程大学提供的学习平台和资源，使我能够顺利的展开研究。此外，还要感谢所有帮助我的朋友和家人，他们在我的求学之路上在精神与物质上都给予了充分的鼓励与帮助，使我能够克服困难，坚持到底。

最后，我要感谢所有参考文献的作者，他们的研究成果为本文提供了重要的理论支持。

本文论文虽然已经完成，但由于本人水平有限，文中难免出现疏漏与不足，恳请各位专家批评指正。