

分类号: TP391

学校代码: 10363

密 级: 公开

学 号: 2220320144



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于无人机航拍图像的
海上目标检测与跟踪研究

论 文 作 者 冯浩宇

校 内 导 师 葛愿 教授

校 外 导 师 叶刚 高级工程师

专业学位类别 电子信息

研究方向(领域) 控制工程

论文提交日期: 2025 年 5 月 19 日

分类号：TP391

单位代码：10363

密 级：公开

学 号：2220320144

题 目： 基于无人机航拍图像的
海上目标检测与跟踪研究

题 目： RESEARCH ON MARITIME TARGET DETECTION
AND TRACKING BASED ON UAV AERIAL IMAGES

学 生 姓 名： 冯浩宇

校 内 导 师： 葛愿 教授

校 外 导 师： 叶刚 高级工程师

专 业 学 位 类 别： 电子信息

研 究 方 向（领 域）： 控制工程

论文答辩日期： 2025 年 5 月 11 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：冯浩宇

日期：2025 年 5 月 19 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：冯浩宇

日期：2025 年 5 月 19 日

指导教师签名：[Signature]

日期：2025 年 5 月 19 日

基于无人机航拍图像的海上目标检测与跟踪研究

摘 要

随着各国对海洋资源的深度开发以及航运活动的日益频繁，海上交通运输安全事故风险随之攀升。无人机作为一种速度快且视野范围广的救援设备，能够通过感知设备快速定位到目标位置，从而提升应急救援的效率和成功率。然而，由于海洋环境的复杂性以及无人机拍摄图像中目标像素比低，直接影响了目标检测与跟踪的性能。为此，本文以提升海上目标检测精度与跟踪稳定性为目标，基于 YOLOv7 与 YOLOv7-tiny 算法展开研究，设计了面向不同天气条件下的海上目标检测优化方法，并基于 DeepSORT 构建多目标跟踪框架。主要研究内容如下：

（1）针对晴天强光照及目标遮挡导致检测精度下降的问题，本文设计一种基于 YOLOv7 的晴天目标检测优化方法。首先，引入坐标卷积模块并替换掉部分标准卷积，有助于网络更好地理解图像中目标的相对位置关系，从而更精确地定位目标。其次，将 CBAM 注意力机制融入到特征增强提取网络中，不仅有助于提取目标特征，还能帮助网络关注全局以及局部信息，有效解决海洋背景干扰下小目标检测困难的问题。同时，采用 NWD 损失函数优化 CIOW 损失函数，以应对目标遮挡的问题。最后，在自制的数据集上对改进后的模型进行了消融实验与对比实验，验证了改进后的算法的有效性。实验结果表

明，改进后的 YOLOv7 模型显著提升了检测精度。

(2) 针对雾天图像退化导致检测精度和召回率偏低的问题，本文设计基于 YOLOv7-tiny 的雾天目标检测优化方法。首先，使用 AODNet 去雾网络进行图像预处理，有效改善图像质量。其次，对检测网络进行改进。通过引入 SimAM 注意力机制和自适应特征融合检测头，增强网络对目标特征的提取与融合能力。同时，采用前文提出的优化后的损失函数提升遮挡目标的检测效果。最后，在基于浓雾场景去雾数据集上进行消融和对比实验，验证了本文所提算法的有效性。改进后的检测算法解决了 YOLOv7-tiny 算法在大雾天气下检测精度低和漏检率高的问题，提升了自身的检测性能，更适用于复杂场景下的目标检测任务。

(3) 针对海上目标跟踪过程中因复杂海况、遮挡等因素致使目标无法被持续跟踪的问题，本文将前一章节改进后的 YOLOv7-tiny 检测网络作为检测器，基于 DeepSORT 算法设计一种融合自适应噪声抑制卡尔曼滤波与 DIOU 匹配机制的海上多目标跟踪算法。通过动态调整噪声协方差矩阵提升跟踪稳定性，结合 DIOU 匹配策略减少目标丢失次数，使跟踪器能够更好地适应目标的动态变化。实验结果表明，改进后的跟踪算法性能得到明显提升。最后，将本文提出的算法与 PyQt5 相结合，开发出带有人机交互界面的海上目标检测与跟踪系统。经系统测试，该系统能够准确地检测和跟踪多个目标。

关键词：无人机；海上应急救援；目标检测；多目标跟踪；人机交互

RESEARCH ON MARITIME TARGET DETECTION AND TRACKING BASED ON UAV AERIAL IMAGES

ABSTRACT

With the increasing exploitation of marine resources and the growing frequency of shipping activities worldwide, the risks of maritime transportation safety accidents have risen significantly. As rapid-response platforms with wide-area surveillance capabilities, unmanned aerial vehicles (UAVs) leverage advanced sensing technologies to enable swift target localization, thereby enhancing the efficiency and success rates of maritime emergency rescue operations. However, the intricate nature of marine environments and the low pixel-to-target ratios in UAV-captured imagery substantially degrade the performance of target detection and tracking systems. Therefore, this paper aims to enhance the accuracy of maritime target detection and the stability of tracking. Based on the YOLOv7 and the YOLOv7-tiny algorithms, this study proposes optimization methods for maritime target detection under different weather conditions and constructs a multi-target tracking framework based on DeepSORT. The main research contents are as follows:

(1) To address the issue of decreased detection accuracy caused by strong sunlight and target occlusion on clear days, this paper designs a clear-weather target detection optimization method based on YOLOv7. First, the Coordinate Convolution (CoordConv) module is introduced in

parallel with the partial replacement of standard convolutional layers, jointly enhancing the network's spatial awareness of inter-target relationships in images and consequently improving localization precision. Second, the Convolutional Block Attention Module (CBAM) is integrated into the feature enhancement network to optimize target feature extraction. This integration enables the network to simultaneously capture global contextual patterns and local discriminative features, thereby effectively addressing small target detection challenges in marine environments with complex background interference. Meanwhile, the Normalized Wasserstein Distance (NWD) metric is introduced to optimize the CIOU loss function, addressing the issue of target occlusion. Finally, ablation and comparative experiments are conducted on a self-constructed dataset to validate the effectiveness of the improved algorithm. Experimental results demonstrate that the modified YOLOv7 model significantly enhances detection accuracy.

(2) To address the issue of low detection accuracy and recall rate caused by foggy image degradation, this paper designs an optimized foggy object detection method based on YOLOv7-tiny. First, the AODNet dehazing network is employed for image preprocessing to effectively enhance image quality. Second, improvements are made to the detection network by introducing the SimAM attention mechanism and an adaptive feature fusion detection head, which enhances the network's capability for target feature extraction and fusion. Additionally, the optimized loss function proposed above is applied to improve the detection performance for occluded targets. Finally, ablation and comparative experiments are conducted on a dataset composed of images dehazed from heavy fog, validating the effectiveness of the proposed algorithm. The improved detection algorithm addresses the low detection accuracy and high missed detection rate observed in the YOLOv7-tiny algorithm under heavy fog

conditions. It significantly enhances detection performance while demonstrating robust applicability to object detection tasks in complex scenarios.

(3) To resolve tracking discontinuities in maritime scenarios under complex conditions (e.g., occlusions, environmental interference), this paper integrates an improved YOLOv7-tiny detector with DeepSORT, proposing a multi-object tracker that combines adaptive noise-suppressed Kalman filtering and a Distance-IOU (DIOU) matching mechanism. By dynamically adjusting the noise covariance matrix to stabilize tracking and employing DIOU-based matching to minimize target loss, the tracker effectively adapts to dynamic target behaviors. Finally, the proposed algorithm is integrated with PyQt5 to develop a maritime target detection and tracking system featuring a human-computer interface. Controlled system testing validates its reliable multi-target detection and tracking performance.

Haoyu Feng(Electronic Information)

Supervised by Prof. Yuan Ge

Keywords: unmanned aerial vehicles; maritime emergency rescue; target detection; multi-target tracking; human-computer interface

目录

摘 要	I
ABSTRACT.....	III
第 1 章 绪论	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 海上应急救援研究现状.....	2
1.2.2 海上应急救援目标检测研究现状.....	4
1.2.3 海上应急救援目标跟踪研究现状.....	8
1.3 论文主要研究内容及章节安排.....	9
1.3.1 主要研究内容.....	9
1.3.2 具体章节安排.....	10
第 2 章 海上目标检测与跟踪理论	12
2.1 目标检测理论.....	12
2.1.1 卷积神经网络相关知识.....	12
2.1.2 经典卷积神经网络.....	20
2.1.3 基于深度学习的两阶段目标检测算法.....	23
2.1.4 基于深度学习的一阶段目标检测算法.....	24
2.2 目标跟踪理论.....	26
2.2.1 卡尔曼滤波.....	27
2.2.2 匈牙利匹配.....	27
2.3 本章小结.....	28
第 3 章 基于 YOLOv7 的晴天海上目标检测算法	29
3.1 YOLOv7 检测模型.....	29
3.2 YOLOv7 检测模型的改进.....	30
3.2.1 引入坐标卷积模块.....	31
3.2.2 加入 CBAM 注意力机制.....	32
3.2.3 优化损失函数.....	33
3.3 晴天海上目标检测数据集制作.....	37
3.3.1 图像采集.....	37
3.3.2 图像标注.....	37
3.4 实验结果与分析.....	38
3.4.1 实验环境及配置.....	38
3.4.2 目标检测评价指标.....	38
3.4.3 改进前后检测模型对比实验分析.....	39
3.4.4 消融实验分析.....	40
3.4.5 不同检测模型对比实验分析.....	40
3.5 本章小结.....	42
第 4 章 基于 YOLOv7-tiny 的雾天海上目标检测算法	43
4.1 YOLOv7-tiny 检测模型	43
4.2 基于去雾算法的图像预处理.....	44

4.2.1 单尺度 Retinex 去雾算法	44
4.2.2 暗通道先验去雾算法	46
4.2.3 AODNet 去雾算法	47
4.3 YOLOv7-tiny 检测模型的改进	50
4.3.1 引入 SimAM 注意力机制	51
4.3.2 加入 ASFF 检测头	52
4.4 雾天海上图像数据集制作	53
4.4.1 海上图像采集	53
4.4.2 海上图像标注	53
4.4.3 图像加雾算法	54
4.5 实验结果与分析	56
4.5.1 实验环境及配置	56
4.5.2 目标检测评价指标	56
4.5.3 去雾前后检测对比实验	57
4.5.4 改进前后检测模型对比实验分析	58
4.5.5 消融实验分析	59
4.5.6 不同检测模型对比实验分析	60
4.6 本章小结	62
第 5 章 基于检测器的海上多目标跟踪算法	63
5.1 DeepSORT 目标跟踪算法	63
5.2 DeepSORT 目标跟踪算法的优化	64
5.2.1 自适应噪声抑制卡尔曼滤波器	64
5.2.2 DIIOU 匹配机制	66
5.3 基于检测器的改进跟踪模型结构	66
5.4 实验结果与分析	67
5.4.1 实验环境及配置	67
5.4.2 跟踪数据集制作	67
5.4.3 多目标跟踪评价指标	68
5.4.4 实验结果分析	68
5.5 海上目标检测与跟踪系统	70
5.5.1 系统设计	70
5.5.2 系统软硬件平台	71
5.5.3 系统界面	71
5.5.4 系统测试	72
5.6 本章小结	73
第 6 章 总结和展望	74
6.1 总结	74
6.2 展望	75
参考文献	76
攻读硕士学位期间主要研究成果	84
致谢	85

第 1 章 绪论

1.1 研究背景与意义

海洋经济作为全球经济增长的动力，备受世界各国的关注。随着世界各国持续加大对海洋资源的深度开发与海上贸易的投入，海上航运变得愈发繁忙，海上安全风险^[1]随之上升，导致海上搜救任务日益繁重。根据欧洲海事安全局统计，在 2011 年至 2018 年期间共发生了 23073 起海上事故，其中有 7694 人受伤^[2]。据国家交通运输部去年统计，截至 2024 年 11 月中国海上搜救中心共组织、协调搜救行动约为 1270 次，其中约 6630 人获救，1200 艘船只获救。从这些数据可以看出海上事故不仅严重影响了全球海洋经济的共同发展，而且给各国造成了重大的财产损失，同时给人员带来了安全威胁。因此，如何高效确定事故船只以及落水者的位置信息是海上救援任务的关键一环，从而为救援人员抢救生命赢得时间。

传统海上搜救主要依靠救援人员搭乘巡逻船和直升机人工搜寻遇险目标，成功定位后再开展救援行动。这种方式不仅耗费过多的成本和人力^[3]，而且搜寻能力有限。近年来随着视觉技术的进步，无人机凭借着高机动性和监控范围广等优势，逐渐成为新型的海上救援方式，世界各国都在积极研发相关无人机设备。2015 年英国 Marlek Aviation 公司研发的无人机海上搜救系统能够侦测到肉眼看不到的落水人员以及小艇^[4]。法国 Helper 公司的救援无人机^[5]截至 2016 年 9 月在南部地区比斯卡罗海滩累计执行了 50 次的海上救助任务^[6]。我国自主研发的“Z-3”型和“WD-50”型无人机也成功应用于水上交通监管和水上应急救援^[7]。无人机机动性高、适应性强、视野广阔^[8]，通过集成环境感知^[9]、目标识别^[10]、定位和跟踪^[11]等技术，实现了对遇险目标的快速发现与持续跟踪，可替代人工执行救援任务，成为目前海上应急救援的首要选择。

无人机执行海上搜索救援任务需要精准定位遇险目标^[12]。这一过程借助视觉设备采集图像，依靠检测与跟踪算法识别和追踪目标，同时结合无人机 GPS 信息以及自身与目标的角度关系确定其精确坐标^[13]，这对系统要求较高。检测与跟踪的效果受到多种因素的影响。无人机的飞行高度和拍摄角度的差异^[14]会使航拍图像中目标物体所占的像素占比不同，给检测与跟踪任务增加了难度。复

杂多变的海洋环境也会阻碍无人机完成检测与跟踪任务。多变的天气条件、海面的动态变化以及光线变化等干扰，易造成无人机误检目标，不利于准确检测与跟踪遇险目标，极大程度上增加了无人机搜救任务的难度。从救援目标的特征看，航拍图像中的海上救援目标多为落水者、遇险船只等，在宽阔海面上特征细节少。它们的颜色、形状等特征受海洋背景干扰严重，使得无人机难以准确检测和跟踪到目标。

在海上应急救援^[15]的研究背景下，本文针对无人机执行海上搜救任务时目标检测与跟踪面临的难题，开展基于无人机航拍图像的相关算法研究。

1.2 国内外研究现状

随着海上运输以及海洋资源开发等各类活动日益频繁，海上突发事故风险呈指数级上升。在海上运输安全体系中，海上搜救行动是实施救援的核心环节。2018年国际海事组织针对《国际海上生命安全公约》做出了重大的法律修订，着重强调了海上搜救行动的重要意义。传统搜救设备如水面船、水下航行器及空中平台，因响应速度慢、能源消耗大且成本高昂，难以满足现代搜救需求。相比之下，以无人机、无人船、无人水下航行器及异构合作智能体集群为代表的无人系统具有高机动性、快速部署和协同合作的优点，在搜救工作里扮演着越来越重要的角色。尤其是无人机已成为现代海上搜救的核心装备。

当前国内外很多学者以提升无人机搜救效率为研究方向，通过优化搜索路径规划、目标检测、目标跟踪等算法并集成在无人机身上，从而有效提升了海上搜救的搜救效率和成功率。

1.2.1 海上应急救援研究现状

当下，海上搜救正从有人化朝着无人化、多样化发展。现有的海上搜救系统主要分为有人搜救系统(搜救船、搜救直升机)和无人搜救系统(水下无人航行器-UUV、无人机-UAV、水面无人船-USV、水下无人航行器-无人机、水下无人航行器-无人机-水面无人船)。有人搜救系统依托专业海上救援船和直升机展开作业，存在着搜救效率低、成本高昂的问题。而无人搜救系统借助先进的路径规划、自动控制、信息传输等技术，大幅度提升了搜救速度。搜寻效率与无人系统所具有搜索半径能力的大小密不可分，而搜索半径主要由相机性能、空中系统的飞行高度、水下系统下潜深度等因素决定。从表 1.1 的信息可知，无人机凭借高空飞行

优势,搜索半径大于水面无人船,而水面无人船的搜索范围大于水下无人航行器。将不同类型的无人系统协同作业,能够扩大搜索半径、增强救援载荷能力,显著提升海面及水下的搜索效率。这种发展趋势凸显了传统救援模式在应对复杂多变的海况时的局限性,使得人们对于无人机等无人装备参与海上救援行动寄予更多的期许,它们为现代海上应急救援提供新的解决方案。

表 1.1 部分海上无人搜索系统信息

系统类型	系统名称	工作范围	续航时间	载荷能力	搜索半径
UAV	SK-62	25km	0.7h@10m/s	6kg	150m(海面)
USV	SE40	65km	6h@3m/s	25kg	120m(海面)
UUV	A9-M	54km	10h@1.5m/s	2.3kg	80m(水下)
UAV-USV	INAC4	90km	5h@5m/s	26kg	250m(海面)
UAV-USV-UUV	I4@Drones	324km	15h@6m/s	65kg	250m(海面) 80m(水下)

当前应急救援体系正朝着智能化、快速化的方向加速迈进,许多研究人员针对传统搜救模式存在效率低的问题进行了技术创新。

在多智能体协同合作方面, Yang T 等人^[16]提出了基于强化学习的无人机-无人船异构集群搜救系统,通过采用网格建模方法与马尔可夫决策算法提升了区域搜索效率,同时解决了通信吞吐量受限的问题。Jin Y 等人^[17]利用地理信息系统对海上救援任务进行建模,然后使用 K-means 算法精准获取救援点的位置,最后将得到的数据输入到优化算法模块,从而得到最优的任务分配方案。Ferrari J F 和 Chen M^[18]将一支临时舰队在公海执行搜索任务的调度问题视为动态资源分配问题,通过改进二进制整数规划模型的目标函数,实现了搜救任务中的响应速度、覆盖范围、能耗效率的多目标优化。美国海岸警卫队^[19]开发了无人机与直升机协同搜救系统,通过视觉跟踪算法实现了两机的轨迹同步,同时结合神经网络技术构建了导航系统。Ebrahimi E 等人^[20]提出了一种分布式无人机集群搜救路径规划方案,通过集群中个体之间的合作关系完成搜寻与定位任务,使得被救人员的获救成功率得到了明显提升。在路径规划方面, Ai B 等人^[21]考虑到完全覆盖海上区域和躲避障碍物的情况下,规划了一条耗时最短的搜索路径。Cho S W 等人^[22]提出了多无人机协同覆盖路径规划的两阶段算法,第一阶段对海上搜救区域进行基于网格化的区域最小化分解,第二阶段使用随机搜索启发式算法推导出用时最少的覆盖搜索路径。Guo Y 等人^[23]针对无人机能耗大的问题,提出了最高能源利用率的无人机海上搜救路径规划算法。首先基于改进 A*-DWA 算法实现动

态避障与低能耗转弯控制, 然后结合 RCPP 算法进行区域覆盖路径规划, 最后通过优化策略使得能源利用率有所提高。在定位与通信方面, 朱武斌^[24]开发的基于射影交比不变性原理的无人机定位将定位误差控制在 0.2% 以内。该系统在海上船舶搜救的视觉范围内具有显著优势, 能够有效提升船舶搜救的效率, 为海上应急救援工作提供了可靠的技术支持。韦定江等人^[25]通过无人机与搭载 5G 基站的救援船建立 5G 网络通信, 提高无人机在远海区域的搜救范围。

随着海上无人机相关技术在多领域不断取得新进展, 它已从辅助工具转变为海上应急救援体系的核心力量, 推动着救援模式向无人化发展。高精度视觉传感器突破了传统海上搜救的感知限制, 路径规划与集群算法可以实现救援无人机的最优调度。这些技术已从研究成果转化为实践成果, 在全世界得以体现。2018 年 1 月 18 日英国的《每日邮报》报道了世界上第一次无人机自主完成海上救援任务的事件^[26], 成功救助了澳大利亚 2 名被卷入大海的孩子, 该事件标志着海上救援进入了无人化时代。2019 年 3 月 3 日美国的《纽约先驱报》新闻报道了澳大利亚的史密斯海滩已部署了 25 架无人机对海滩人员安全进行监控^[27]。我国在救援无人机应用上展现出自身的优势。在 2020 年 5 月我国山东省枣庄市使用几十架 Phantom4 大疆无人机协同合作成功完成了对演练海域的区域搜索和遇险人员的识别等任务, 备受国内外无人机爱好者和新闻媒体的关注。2022 年广东省应急救援部门使用大疆无人机和云平台共同作业, 极大程度提升了防汛救灾工作效率。

在未来, 通过不断优化无人机的路径规划、任务分配、目标检测与跟踪等算法, 建立起更智能化的应急救援决策机制, 能够让无人机在救援领域起到不可估量的作用, 更好地造福于人类。

1.2.2 海上应急救援目标检测研究现状

船只倾覆与搁浅等事件是较为常见的海上事故类型, 如图 1.1(b)所示。面对这种紧急事件需要进行快速高效的人道主义救援, 依靠大量的人力在茫茫大海中搜寻目标, 会浪费宝贵的时间。因此, 部署具有海上目标检测功能的无人机, 对于挽救更多生命意义重大, 如图 1.1(a)所示。然而, 在无人机执行海上搜救任务时, 存在着诸多加大搜索难度的因素, 如图 1.1(d)和图 1.1(e)所示。它们会产生视觉干扰, 使得识别待救助人员的难度加大, 进而出现搜救效率低、难以准确识

别出目标、难以检测出目标等问题。面对这些挑战，不少研究人员研发出搭载热成像、多光谱和高分辨率可见光摄像头且具有图像检测功能的无人机，如图 1.1(c) 所示。

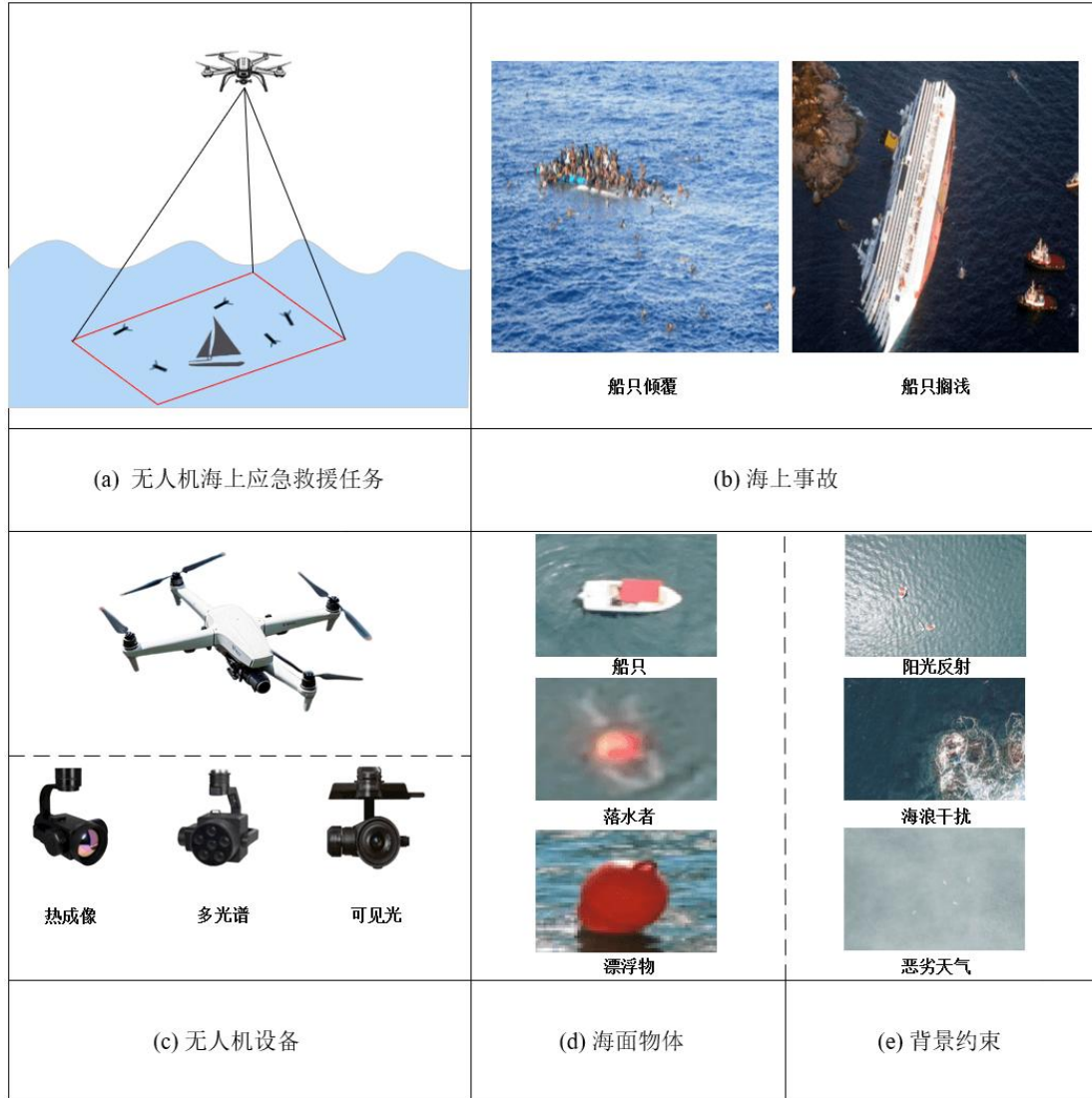


图 1.1 无人机海上救援目标检测

在海上应急救援目标检测领域，按照时间先后顺序，检测方法大概可分为两类：传统目标检测方法和基于卷积神经网络的目标检测方法。早期的海上定位目标检测方法主要依赖于传统的计算机视觉(Computer Vision, CV)算法。这一阶段多数人采用如图 1.2 的流程：先使用图像特征提取器对图像进行处理，然后利用分类器对不同目标进行分类。这种方法不仅增加了计算量，而且由于依赖人工提取目标特征，致使检测性能欠佳。Leira F S 等人^[28]率先使用这种方法，开发出一种融合传统 CV 技术的无人机目标识别系统。他们使用高斯滤波器降低热成像图

片中的噪声，使用普雷维特算子进行边缘检测。此外，他们应用一种阈值化方法去除海浪引起的边缘干扰。最后，使用连通分量算法来提取目标轮廓，从而计算出边界框。Hoai D K 和 Van Phuong N^[29]针对异常检测算法在各种海上搜救场景中的实用性进行了研究。他们发现，用多个颜色空间模拟不同场景时，需要采用不同方法，这表明传统 CV 算法缺乏泛化能力。对于海上搜救场景，Reed I S 和 Yu X^[30]提出的 RX 检测器能够较好地适应这一特定背景，且取得了最佳效果。Zhao B 等人^[31]的最新研究中利用了海洋场景的独特特征，运用聚类方法处理目标大小问题，提出将锚框作为待候选目标，为基于传统 CV 算法提供了一种启发式探索思路。该方法所得结果与近期基于神经网络的方法相比，具有一定的竞争力。



图 1.2 传统 CV 检测流程

随着卷积神经网络的热潮席卷整个目标检测领域，众多研究者将海上目标检测方法由传统方法转变为深度学习方法。基于卷积神经网络的目标检测器包括一阶段检测器和两阶段检测器，如图 1.3 所示。一阶段检测器由基于 CNN 或者 Transformer 的主干网络和头部网络共同构成，如图 1.3(a)所示。当图像输入进来后，该检测器可直接对边界框和目标类别进行预测。当前比较主流的一阶段算法有 RetinaNet^[32]，You-Only-Look-Once(YOLO)^[33]以及 Single-Shot Detector (SSD)^[34]。另一种是两阶段检测器，如图 1.3(b)所示。它与一阶段检测器的不同点在于在头部网络增加了两个额外操作，即区域提议以及分类与定位。这种两阶段模型的典型代表有 Fast R-CNN^[35]，Faster R-CNN(FR-CNN)^[36]以及 Cascade R-CNN(CR-CNN)^[37]。为了解决海上目标难以检测和精确定位的问题，不少研究人员最为关注的是 YOLO 系列模型，并采用多种策略对 YOLO 系列算法进行改进。Lygouras E 等人^[38]作为该领域的开拓者，利用无人机拍摄的 9000 张游泳者图像，训练了一个低成本的 YOLOv3 模型，用来满足无人机海上救援巡检的需求。但是，存在着检测精度较低的缺点。Sharafaldeen J 等人^[39]利用可见光相机和红外传感器，使用 YOLOv4 系列模型验证了在海洋环境中能够准确地检测出人员和船只，这表明该模型能应用于救援无人机上面。Goncalves L 和 Damas B^[40]

以及 Rizk M 等人^[41-42]同样使用 YOLOv4 模型，在海上救援场景下对人员检测取得良好的效果。Bai J 等人^[43]提出了轻量型的 YOLOv5 改进模型，针对海洋救援场景中小目标物体检测困难的问题，添加了包含浅层特征提取层、特征融合层以及一个额外预测头的全新模块。这些改进不仅提升了小目标的检测性能，而且提高了检测效率。

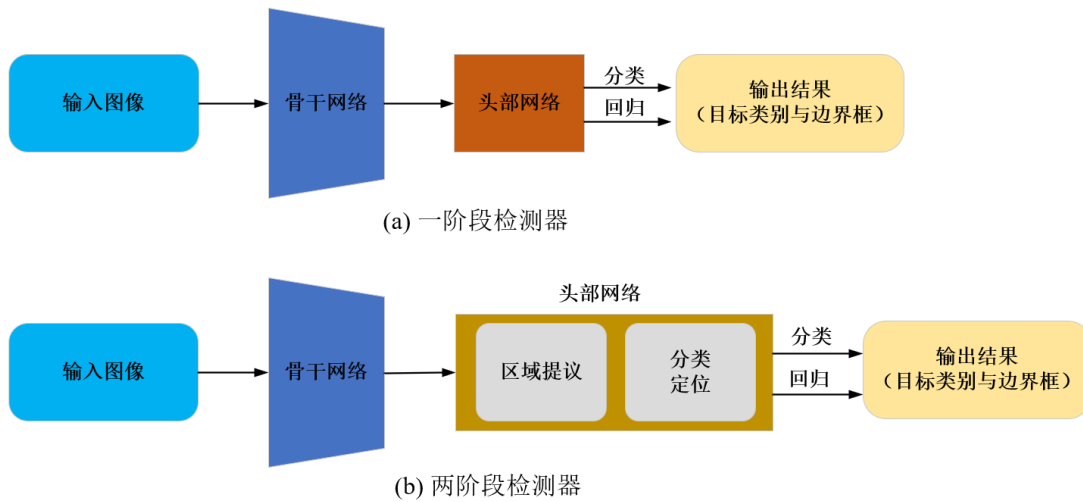


图 1.3 基于卷积神经网络的目标检测器

目前，视觉注意力机制与基于 Transformer 的方法在目标检测领域备受关注，它们在多个检测任务中体现出显著的价值。例如，在海事计算机视觉研讨会^[44]上举办的相关比赛中，部分研究者应用这种方法。Yang Z 等人^[45]提出了 YOLO-BEV 检测算法，在 Liu S 等人^[46]的研究成果的基础上不仅融入了双向特征融合模块，还专门设计了一个用于预测极小目标的头部网络。并提出了 C2fSESA 模块，用来强化特征提取。相较于原始的 YOLOv5，检测性能有了明显提升。Wang X 等人^[47]同样以 YOLOv5 为基础模型，融入了卷积块注意力模块，能够在高分辨率图像中确定注意力区域，使模型在复杂海上场景能更好地聚焦关键区域。Swin-Transformer 也是一种被广泛应用的方法。Liu Z 等人^[48]提出 YOLOv5s-SwinDS 模型，在主干网络中使用 Swin-Transformer 模块替代原有 YOLOv5s 的部分模块，此外，引入了多级特征融合特性网络，同时在特征输出端采用可变形卷积网络^[49]替代传统卷积，增强了对不规则形状物体的识别能力。Wang X 等人提出的 Sea-YOLOv5s 模型则将 Swin-T 架构引入到小目标检测头，增强了网络的上下文信息恢复能力，使得改进后 YOLOv5 模型更适合检测海上小目标。Zhao H 等人^[50]提出了 YOLOv7-sea 算法，通过采用了新的预测头和一

种无参数注意力模块，更好地检测出小目标。Zhu Q 等人^[51]提出了 YOLOv7-CSAW 的检测方法，先利用 K-means++ 算法优化先验锚框尺寸，再采用 C2f 模块构建轻量级模型，并使用自适应特征融合网络增强小目标的语义特征。此外，他们还采用 WIOU 损失函数替代 CIOU 损失函数，有效应对较大的定位误差和漏检问题，减少负样本误判的情况。Gao S 等人^[52]设计了一种基于多尺度注意力的检测模型。该模型主要由多尺度注意力增强模块和多尺度特征融合模块两个模块组成，在两者的共同作用下能够使模型聚焦于不同尺度物体的相关特征。实验表明，这种方法能有效提升在复杂的海洋环境中对小目标检测的准确性。Zhang Y 等人^[53]提出了 ABT-YOLOv7 算法。他们通过在网络中融入 BiFormer 注意力模块和加入渐进式特征金字塔网络，有效增强了网络对小尺度目标的检测能力。尽管上述方法在海上目标检测中取得了一定进展，但在像雾天这样的复杂场景下，现有方法的检测准确率和鲁棒性仍存在显著不足。因此，本文后续章节先采用 AODNet 算法对输入图像进行去雾预处理，以增强目标特征的可见性，进而提升检测精度。

1.2.3 海上应急救援目标跟踪研究现状

在海上搜救领域的另一个重要研究方向是目标跟踪技术，其核心任务是对目标进行持续监测与实时追踪。海洋环境具有动态特性，海况可能瞬息万变，洋流、海浪以及其他环境因素会对目标的精确定位产生较大的干扰。借助有效的跟踪系统，一旦发现目标后便能持续监控，可有效降低目标丢失风险。要想实现对特定目标的跟踪，首先需要处理视频序列数据，通过从连续视频流中提取关键帧作为输入进行分析，接着利用检测器逐帧提取目标位置信息，再由跟踪器对各帧之间的对应关系进行匹配，为每个目标分配唯一 ID 并预测其随时间变化的运动轨迹，图 1.4 详细展示了该流程的具体步骤。目前针对海上目标跟踪所提出的技术方法，主要分为传统跟踪方法和基于深度学习的跟踪方法这两类。

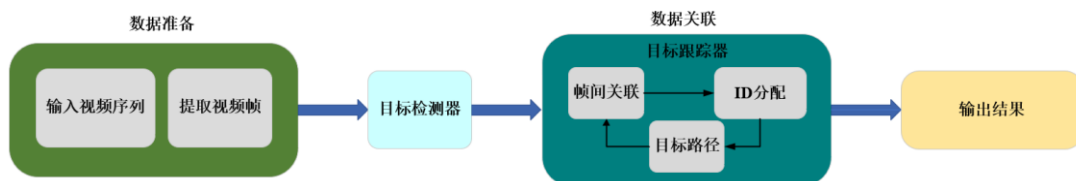


图 1.4 目标跟踪总体流程

传统跟踪模型主要基于数学统计理论与传统计算机视觉技术，通过分析视频

序列中目标特征在时间维度上的关联性实现目标跟踪任务。在早期研究中, Westall P 等人^[54]系统地对四种不同色彩空间、四种点目标检测技术以及两种时序跟踪方法进行了评估。该团队在后续的研究^[55]中引入数据融合算法来优化检测流程。与之形成鲜明对比的是, Kaiyu X 与 Chaojian S^[56]根据红外视频帧特性与海洋环境特征,将帧差法与目标区域扩展的方法相结合,实现了对海面目标的快速且精准的跟踪。

近年来,深度学习技术迅猛发展,为复杂的海洋环境下的目标跟踪研究带来了新的突破。Kiefer B 和 Zell A^[57]提出了一种实时端到端的未来帧预测编码器,该编码器能高效识别视频流中的兴趣区域,显著提升了处理速度。并经过实验验证,即便在资源受限的设备上也具有良好的应用潜力。Kiefer B 等人^[58]构建了时间记忆图,在不显著增加计算开销的情况下融入元数据。他们将此方法与 YOLOv7-Tiny 相结合用于视频目标检测,并进一步整合 DeepSORT 算法和重识别(ReID)模块,成功实现对视频序列中人物的跟踪。Yang C Y 等人^[59]提出了元数据引导多目标跟踪(MG-MOT)的算法,该算法融合无人机提供的 GPS 位置、飞行高度、相机角度等元数据,将短期跟踪数据整合为连贯的长期轨迹。同时使用 YOLOv8 的不同的骨干网络作为检测器,与其他方法相比,性能上有了显著提高。深度学习技术在传统跟踪方法的基础上,推动着目标跟踪技术不断向前发展。

1.3 论文主要研究内容及章节安排

1.3.1 主要研究内容

本文以无人机海上应急救援为研究背景,鉴于特殊的海洋环境可能会对目标检测与跟踪效果产生影响,进而导致救援效率降低。因此,本文针对目标检测精度欠佳、召回率低以及目标跟踪效果不好的情况,以 YOLOv7 系列算法和 DeepSORT 算法为基础,对基于无人机航拍图像的海上目标检测与跟踪进行深入研究,主要研究内容如下:

(1) 基于 YOLOv7 的晴天海上目标检测算法

针对海面存在强光照致使多目标检测精度欠佳的问题,本文设计一种基于 YOLOv7 的晴天海上目标检测优化方法。通过在网络中引入坐标卷积模块并融入 CBAM 注意力机制,有效增强了原网络的空间感知能力与特征提取能力。使用

NWD 损失函数优化原先的 CIOU 损失函数，有效提升了多个目标相互遮挡情形下的检测效果。最后，在自制的晴天数据集上开展了消融实验和对比实验，综合验证了改进后的算法具备高精度检测性能。

（2）基于 YOLOv7-tiny 的雾天海上目标检测算法

针对雾天图像画质下降导致多目标检测精度不高的问题，本文设计一种基于 YOLOv7-tiny 的雾天海上目标检测优化方法。首先，选用 AODNet 去雾网络作为图像预处理模块。在通过加雾算法生成的自制的浓雾数据集上进行了对比实验，实验结果表明这种去雾预处理方法能够提高目标检测精度。在检测网络方面，通过引入 SimAM 注意力机制，同时引入自适应特征融合 ASFF 检测头替换掉原检测头，有效增强了网络对目标特征的提取与融合能力。使用 NWD 损失函数优化 CIOU 损失函数，提升了存在相互遮挡目标时的检测效果。最后，在基于浓雾去雾后的数据集上与多个目标检测模型进行了对比实验，验证了改进后的算法具有较高的检测精度，同时兼顾良好的检测速度。

（3）基于检测器的海上多目标跟踪算法

考虑到改进后的 YOLOv7-tiny 算法兼具相对较高的精度和相对较快的速度，本文将改进后的 YOLOv7-tiny 检测网络作为海上多目标跟踪的检测器。通过使用自适应噪声抑制卡尔曼滤波器，有效处理不同尺度的噪声，使改进后的跟踪器能够更稳定地适应目标的动态变化。此外，采用 DIOU 匹配机制替换掉原有的 IOU 匹配机制，减少了目标丢失的次数。随后，与原跟踪算法进行对比实验，实验结果表明改进后的跟踪算法有效提升了跟踪性能。最后，将改进后的算法与 PyQt5 集成在一起，构建了带有人机交互界面的海上目标检测与跟踪系统。

1.3.2 具体章节安排

本文主要围绕基于无人机航拍图像的海上目标检测与跟踪进行研究，共分为六个章节，具体章节安排如下：

第 1 章：绪论。首先介绍本课题的研究背景与意义，提出海上目标检测与跟踪算法对无人机完成海上救援任务的重要性；然后分别从海上搜救、海上目标检测以及海上目标跟踪对国内外研究现状进行了详细的介绍；最后对本文的研究内容和章节安排进行了说明。

第 2 章：海上目标检测与跟踪理论。首先介绍基于深度学习的目标检测理论

知识，主要包含卷积神经网络、两阶段目标检测和一阶段目标检测相关理论。然后介绍了目标跟踪相关理论，包括卡尔曼滤波原理与匈牙利匹配方法。

第 3 章：基于 YOLOv7 的晴天海上目标检测算法。本章首先对 YOLOv7 算法进行了具体介绍，然后针对基准模型在晴天环境下海上目标检测表现欠佳的问题，设计了一种高精度的海上目标检测算法。该算法通过在 YOLOv7 网络中引入新的卷积 CoordConv 模块、融入 CBAM 注意力机制以及优化原始损失函数，使改进后的模型具有更好的检测性能。通过在自制的晴天海上目标检测数据集上对改进后的模型进行消融实验和对比实验，验证了所提算法的有效性。

第 4 章：基于 YOLOv7-tiny 的雾天海上目标检测算法。本章首先对 YOLOv7-tiny 算法进行了具体分析，针对雾天图像画质下降致使基准模型检测精度不高的问题，设计了一种基于 YOLOv7-tiny 的雾天海上目标检测优化方法。然后，本文将 AODNet 作为图像预处理模块，通过在检测网络中引入 SimAM 注意力机制、使用 ASFF 检测头以及优化原网络的 CIOU 损失函数，提升模型的检测性能。最后，在基于浓雾去雾后的数据集上，通过消融实验和对比实验加以验证。

第 5 章：基于检测器的海上多目标跟踪算法。首先介绍 DeepSORT 跟踪算法；其次对 DeepSORT 进行优化，改进的地方包括引入自适应噪声抑制卡尔曼滤波和使用 DIOU 替换原先 IOU 匹配机制；然后，将前一章改进的检测网络与优化后的 DeepSORT 相结合作为改进后的跟踪算法，并在自制的跟踪数据集上通过对比实验进行验证；最后，将改进后的算法与 PyQt5 相结合，开发出具备人机交互界面的海上目标检测与跟踪系统。

第 6 章：总结与展望。总结本文的研究内容，指出当前存在的不足，并明确未来的研究方向。

第 2 章 海上目标检测与跟踪理论

近年来，基于深度学习的目标检测与跟踪算法在海洋监测、舰船识别、海上搜救等领域展现出优秀的性能。将无人机和这类算法结合起来，能够显著提升搜索效率，为救援人员开展海上救援任务赢得宝贵时间。因此，深入理解深度学习的核心理论与方法，对于实现精准的海上目标检测与跟踪十分重要，是无人机高效完成海上救援任务的基本前提。本章首先介绍卷积神经网络相关知识，其中包括以 LeNet 网络、AlexNet 网络和 ResNet 网络为代表的经典卷积神经网络。接着介绍以 R-CNN、Fast R-CNN、Faster R-CNN 为代表的两阶段目标检测方法，以及以 SSD 和 YOLO 为代表的一阶段目标检测方法。最后，介绍目标跟踪理论的相关知识与方法，包括卡尔曼滤波方法和匈牙利匹配方法。

2.1 目标检测理论

2.1.1 卷积神经网络相关知识

在实际应用场景里，从工业生产中的产品缺陷检测到交通领域的自动驾驶场景识别，卷积神经网络展现出了强大的适应性与高效性。

卷积神经网络(Convolutional Neural Networks, CNNs)是深度学习中极为重要的计算模型，其设计灵感源于人类的视觉系统从复杂场景中提取出有用信息的过程。在人类的视觉系统中，人眼获取的视觉图像从视网膜开始，经过多个神经元处理，最终在大脑皮层完成认知判断。最初的设计者将这一感知过程构建成一个包含输入层、卷积层、池化层、激活函数、全连接层以及输出层的神经网络，其框架结构如图 2.1 所示。首先，将摄像头获取的图像输入到卷积神经网络的输入层，接着依次经过多个卷积层的卷积操作和多个池化层的池化操作，实现有效的高层次特征提取。提取出来的特征信息经全连接层进行相关的分类和预测，最后由输出层输出检测结果。

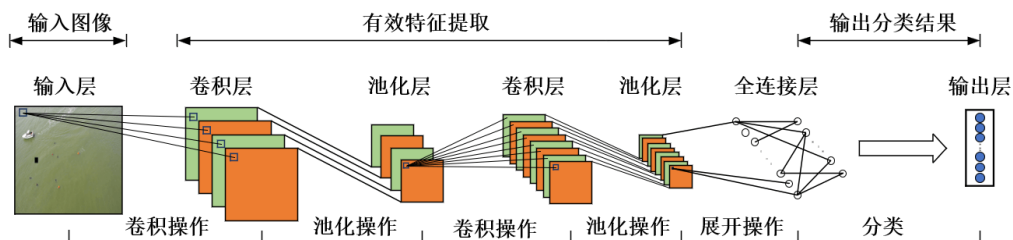


图 2.1 卷积神经网络框架结构图

(1) 输入层

输入层作为卷积神经网络的第一层，接收外部输入的图像数据，将这些数据以数值形式传递给后续的网络层。输入层所接受常见的数据形式为二维矩阵或三维张量。对于一张高度为 H ，宽度为 W 的三通道 RGB 彩色图像而言，输入层需要将图像转换为 $H \times W \times 3$ 的三维张量。以一张高度为 H ，宽度为 W 的灰色图像为例，输入层则要将图像转换为 $H \times W$ 的二维矩阵。在输入层对原始图像进行标准化、归一化等预处理操作，能够有效改善网络的训练效果。为了增强检测模型的泛化能力，还会在输入层对原始图像进行水平翻转、平移变换、裁剪等数据增强操作。

(2) 卷积层

卷积层作为卷积神经网络的核心组成部分。卷积运算是卷积层获得特征信息的关键操作，通过在图像上滑动卷积核来生成特征图，能有效捕捉图像中不同的特征信息。在卷积层对图像进行特征提取时，输入图像经过卷积运算后，通过激活函数进行非线性变换，得到输出特征图。输入图像经卷积运算后输出特征信息的表达式为：

$$x_j^i = f(\sum_{a \in M_j} x_a^{i-1} \times k_{aj}^i + b_j^i) \quad (2-1)$$

其中， i 表示卷积层的当前层数， M_j 表示输出特征图的特定集合， x_j^i 表示第 i 层的第 j 个输出特征图， k_{aj}^i 表示第 i 层的卷积核，用于将第 $i-1$ 层的第 a 个特征图连接到第 j 个输出特征图。 b_j^i 表示第 i 层第 j 个特征图的偏置项， f 表示引入的非线性激活函数。

在卷积层中存在着大量神经元。若卷积层数量越多，网络的特征提取能力越强，因此其能够获取到的图像特征信息越丰富。每个卷积层包含五个基本参数，分别为卷积核大小、边界填充、步长、输入通道数和输出通道数，共同决定着卷积层的特征提取能力。其中，卷积核大小、边界填充和步长三者决定了卷积层输出特征图的尺寸。不同大小的卷积核捕捉到不同类型的图像特征，较小的卷积核侧重于提取局部细节特征，较大的卷积核则能够捕捉更广泛的特征信息。常见的卷积核有 3×3 和 5×5 。边界填充用于解决特征图与卷积核尺寸不匹配的问题，避免图像边缘像素丢失。通过合理填充的操作，能确保卷积操作在图像边缘区域能正常进行，这样可以防止因边缘信息丢失而影响特征提取的完整性。按照填充

层数和目的,可分为有效填充、相同填充、全填充和任意填充四类。步长指上一次卷积到下一次卷积所滑动的距离,决定了卷积输出矩阵的大小。步长增加,特征图分辨率会降低,参数量相对减少。输入通道数取决于输入图像的通道维度。例如,输入图像是 RGB 类型,通道数为 3。输入的是灰色图像,则通道数为 1。输出通道数等同于卷积核的数量,决定了卷积层输出的特征图数量,这些特征图作为后续网络层的输入。

图 2.2 是无填充情况下步长为 1 且卷积核为 2×2 的卷积运算过程,卷积窗口从输入张量的左上角区域 1 开始,滑动到区域 4,每次滑动一个元素。输入图像的大小为 3×3 ,卷积核大小为 2×2 ,输出特征图大小为 2×2 。

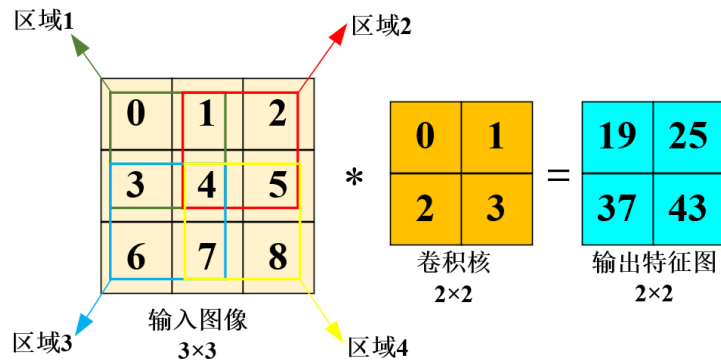


图 2.2 步长为 1 的卷积运算

图 2.3 是填充数为 1 的情况下步长为 2 且卷积核为 3×3 的卷积运算过程,卷积窗口从输入张量的左上角区域 1 开始,滑动到区域 4,每次滑动一个元素。输入图像的大小为 3×3 ,填充完后输入特征图变为 5×5 ,卷积核大小为 3×3 ,输出特征图大小为 $(2,2)$ 。

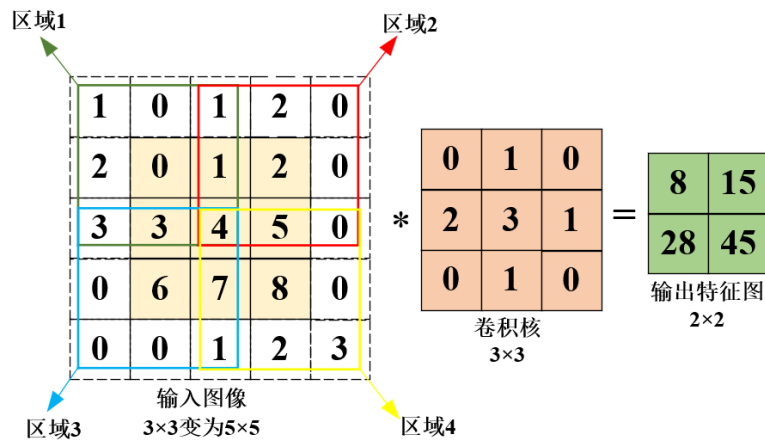


图 2.3 带有填充且步长为 2 的卷积运算

结合两幅图可以得出输出特征图大小与输入特征图大小、卷积核大小、步长

和填充数都有关系。假如输入特征图大小为 (H, W) ，卷积核大小为 (F_H, F_W) ，填充数量为 P ，步长为 S ，则输出特征图大小为 (O_H, O_W) 的公式可以表示为：

$$O_H = \frac{H + 2P - F_H}{S} + 1 \quad (2-2)$$

$$O_W = \frac{W + 2P - F_W}{S} + 1 \quad (2-3)$$

(3) 池化层

池化层(Pooling Layer)是卷积神经网络的重要组成部分，其作用是对卷积层输出的特征图进行下采样操作，因此也被称为降采样层。借助池化层，能够有效降低特征维度，进而减少后续网络层的计算量与参数数量。同时，这有助于缓解过拟合现象，增强模型的泛化能力。依据具体的池化策略，池化层可分为多种类型，其中较为常见的是最大池化和平均池化这两种操作。

最大池化通过在每个池化窗口内选取最大值作为输出结果，来有效保留目标最具辨识度的显著特征。对应的数学表达式表示为：

$$Y_{i,j} = \max_{(p,q) \in R_{i,j}} X_{p,q} \quad (2-4)$$

其中， $R_{i,j}$ 表示输入特征图上的局部感受野区域， $Y_{i,j}$ 表示输出特征图 Y 上坐标为 (i,j) 处的值， $Y_{i,j}$ 是输入特征图 X 中对应池化窗口 $R_{i,j}$ 内所有元素的最大值。

图 2.4 展示了最大池化的过程，输入特征图大小为 5×5 ，通过 3×3 的池化窗口以步长为 1 进行滑动扫描，在每个感受野区域内选取最大值作为输出特征值，最后得到大小为 2×2 的输出特征图。

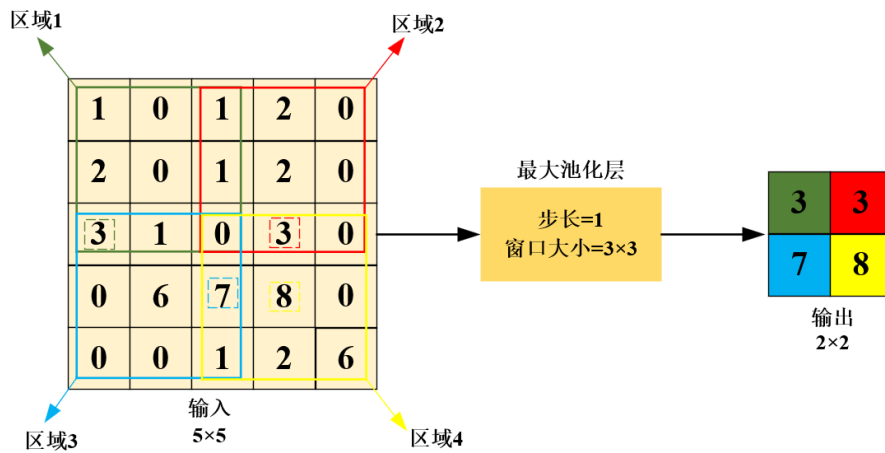


图 2.4 最大池化过程

平均池化的输出结果是通过计算池化窗口内所有值的平均值来获得。与最大

池化相比，它保留目标特征的全局分布特性。平均池化的数学表达式表示为：

$$Z_{i,j} = \frac{1}{m \times n} \sum_{(p,q) \in R_{i,j}} X_{p,q} \quad (2-5)$$

其中， $R_{i,j}$ 表示输入特征图上的局部感受野区域， $Z_{i,j}$ 表示输出特征图 Z 上坐标为 (i,j) 处的值， $Z_{i,j}$ 是输入特征图 X 中对应池化窗口 $R_{i,j}$ 内所有元素求和后的平均值。

图 2.5 展示了平均池化的过程，输入特征图大小为 5×5 ，利用 3×3 的池化窗口以步长为 1 进行滑动扫描，在每个感受野区域内取平均值作为输出特征值，最后得到大小为 2×2 的输出特征图。

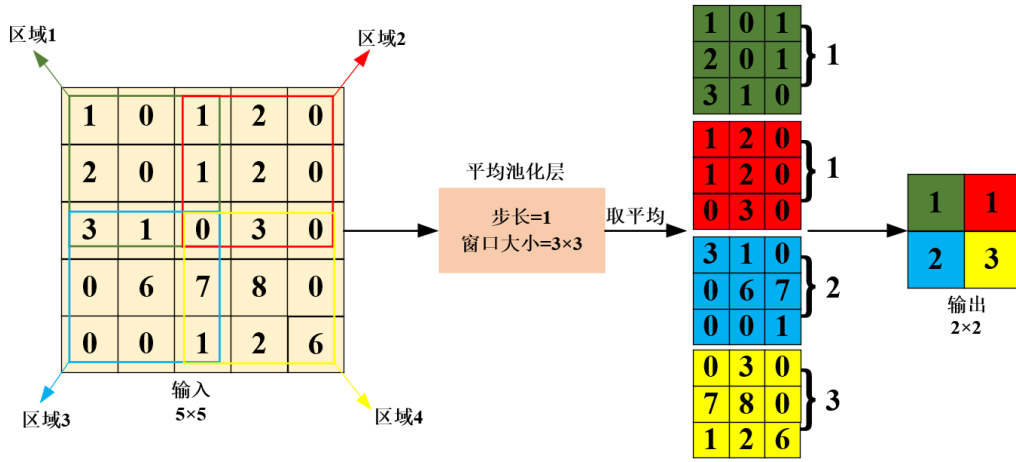


图 2.5 平均池化过程

(4) 全连接层

全连接层(Fully Connected Layers, FC)是卷积神经网络的重要组成部分，位于网络结构的后几层。在目标分类任务中，全连接层被用作输出层，发挥着关键作用。全连接层的核心作用在于其神经元与前一层的所有神经元均相互连接。这种结构使前一层的每个输入都能作用于全连接层的所有输出，实现对前面层所提取特征的深度融合。在计算过程中将输入向量 e 与权重矩阵 w 进行线性变换，通过添加偏置项 c ，来增强模型的表达能力。最后，经 ReLU、Sigmoid 或 Tanh 等类型激活函数 σ 进行非线性变换，使模型具备学习复杂非线性关系的能力，进而提升在各类任务中的性能。全连接层的数学表达式为：

$$y_j = \sigma\left(\sum_{i=1}^n w_{ji}e_i + c_j\right) \quad (2-6)$$

其中, y_j 表示当前层第 j 个神经元的输出值, w_{ji} 表示连接第 i 个输入节点与第 j 个输出节点的权重, c_j 表示第 j 个输出节点的偏置项。图 2.6 展现了全连接层的前向传播过程, 蓝色方块表示输入层, 输入值为 e , 隐藏层由多个黄色圆圈表示为神经元, 神经元之间的连接的权重用 w 表示。绿色圆圈表示输出层, 输出值为 y 。且前一层的每个神经元都与下一层的每个神经元相连。

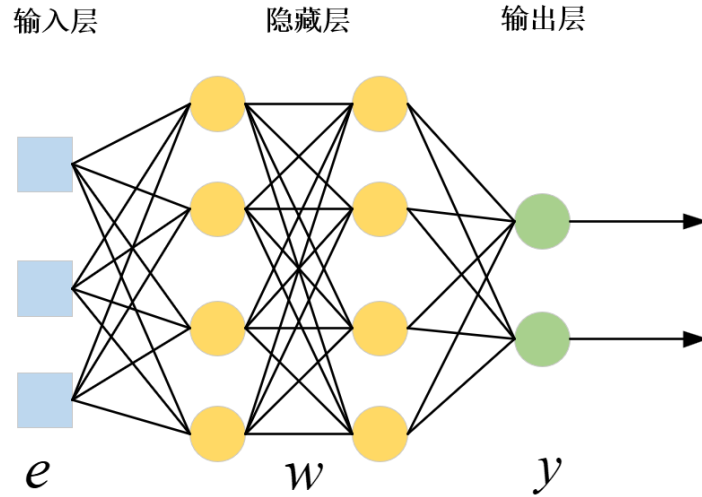


图 2.6 全连接层的前向传播过程

(5) 激活函数

在神经网络中, 神经元之间的运算本质上是线性变换。然而, 在实际应用场景里大量数据特征极为复杂, 呈现出高度的非线性特征。以图像数据为例, 其不仅包含丰富的像素信息, 还存在复杂的纹理、形状以及颜色等特征。这些因素会使原本基于线性回归的网络结构受到一定的局限, 无法有效地提取图像中的复杂特征。为了突破这种局限性, 在卷积层和全连接层之间引入具有非线性特性的激活函数是十分必要的, 通过引入激活函数, 网络能够学习和捕捉到更为复杂的特征信息。在深度学习领域, 常见的激活函数有 Tanh 函数、ReLU 函数、Leaky ReLU 函数和 Sigmoid 函数, 这几种激活函数各有特点。

① Tanh 激活函数

Tanh 激活函数常应用于神经网络的隐藏层中, 能够有效激活神经元, 使其输出限制在 $(-1,1)$ 范围内。与 Sigmoid 函数相比, Tanh 激活函数的输出均值更接近 0, 收敛速度相对更快。然而, Tanh 函数容易出现梯度消失的现象。Tanh 函数公式如下:

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-7)$$

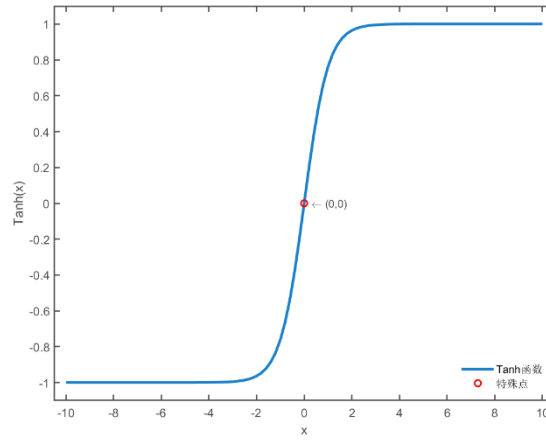


图 2.7 Tanh 函数

由图 2.7 可知，Tanh 激活函数的输出以特殊点(0,0)为中心，相比于一些激活函数，解决输出非零中心问题，使得在训练神经网络时的数据分布更有利于模型收敛。

②ReLU 激活函数

ReLU 激活函数在深度学习领域应用极为广泛。相较于 Sigmoid 和 Tanh 函数，它呈现出线性、非饱和的特性，这使得模型能够快速收敛。ReLU 函数公式如下：

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2-8)$$

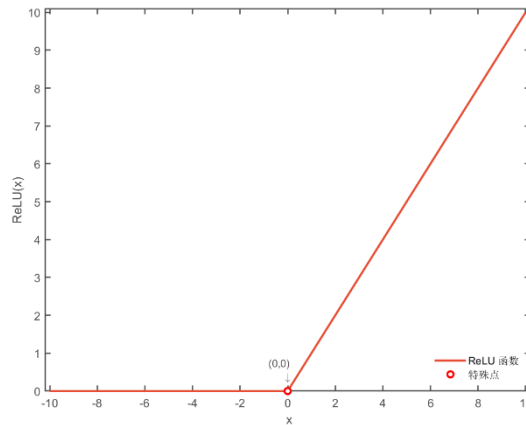


图 2.8 ReLU 函数

由图 2.8 可以看出，ReLU 函数能够在 $x \geq 0$ 时保持梯度不衰减，从而有效缓解了梯度消失的问题。当 $x < 0$ 时，ReLU 函数使一部分神经元的输出为 0，为神经网络提供了稀疏表达能力，避免在训练过程中出现过拟合现象。因此，这种激

活函数被广泛应用在目标检测任务中。

③Leaky ReLU 激活函数

Leaky ReLU 激活函数是在 ReLU 激活函数的基础上改进而来，其目的是解决由于神经元的输入长时间处于负值状态，导致 ReLU 函数无法处理“神经元死亡”的问题。对于 Leaky ReLU 函数，即便输入为负值，神经元依然能有一定的输出，从而避免了“神经元死亡”的情况，这使得网络在训练时更加稳定。Leaky ReLU 函数公式如下：

$$Leaky\ ReLU(x) = \max(\alpha x, x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases} \quad (2-9)$$

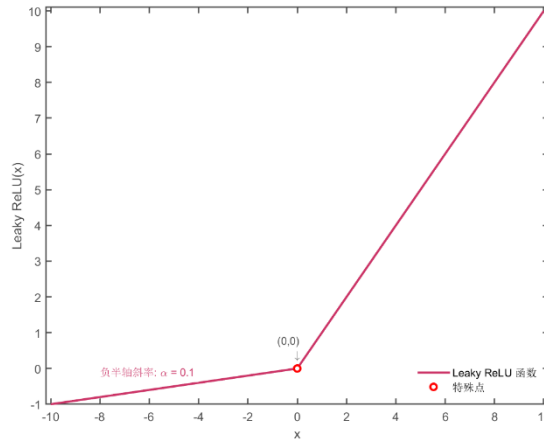


图 2.9 Leaky ReLU 函数

从图 2.9 的曲线变化情况可以看出，当输入值小于 0 时，Leaky ReLU 函数输出一个很小的负数。这样有效地避免“神经元死亡”的情况，帮助神经网络更好地学习数据中各种特征。当输入值大于等于 0 时，Leaky ReLU 函数的输出值与 ReLU 函数一致。这种方式保留了 ReLU 函数的特性，有效避免了出现梯度消失的现象。Leaky ReLU 函数能够适用于复杂的目标检测任务场景。

④Sigmoid 激活函数

在早期神经网络研究中广泛使用 Sigmoid 函数。作为一种非线性函数，对应的输出值范围在(0,1)内，所以它被适用于二分类任务。Sigmoid 函数公式如下：

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (2-10)$$

由图 2.10 可知，Sigmoid 函数的输出值范围在(0,1)内，因而用在二分类任务中用来表示目标的概率。但是，它存在一定的缺陷。由于导数范围在(0,0.25)以内，在神经网络的反向传播过程中经过多层传递后梯度值会不断变小，导致权重

更新变得越来越慢，容易出现梯度消失问题，这样使得模型不能收敛。

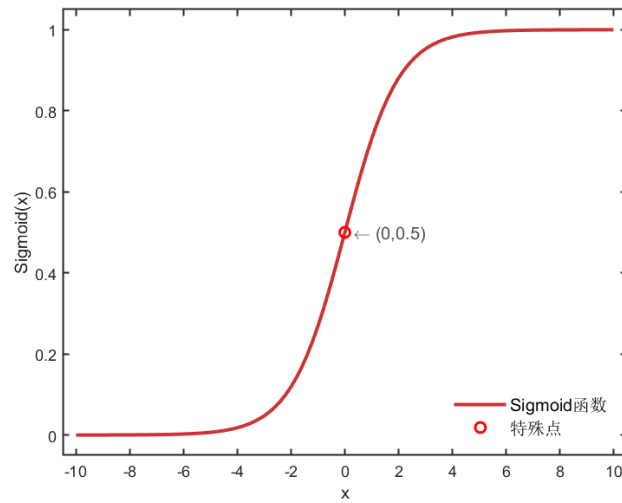


图 2.10 Sigmoid 函数

(6) 输出层

输出层是神经网络的最后一层，在整个模型中起着关键作用。它将前面卷积层、池化层和全连接层所提取的特征转化为最终的输出结果，以适应不同的任务需求。输出层与前面各层相互影响。一方面，输出层接收前面各层提取的特征信息，其性能取决于卷积层、池化层和全连接层对图像特征的有效提取。另一方面，在反向传播过程中，输出层计算得到的梯度会反向传播到前面各层，用于权重更新，进而优化网络参数。由此可见，输出层的性能对整个网络的训练和优化效果极为重要。

2.1.2 经典卷积神经网络

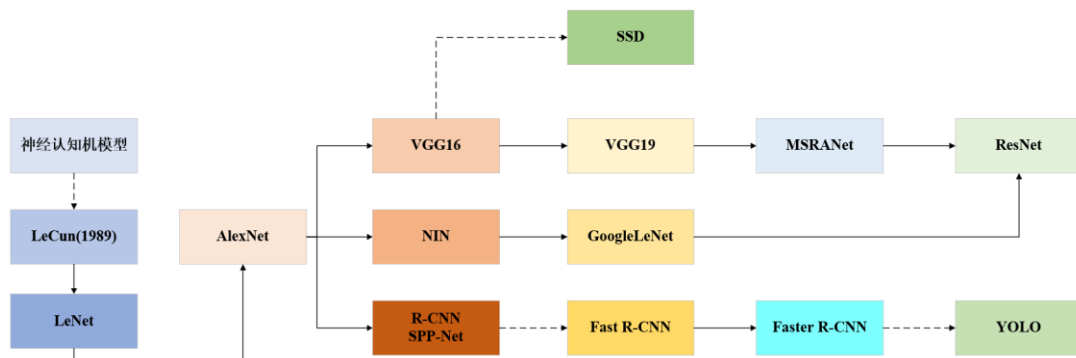


图 2.11 卷积神经网络的发展历程

从最早的神经认知机模型到现阶段比较主流的一阶段和两阶段目标检测方法，这些技术上的进步都离不开卷积神经网络的持续演变与发展。图 2.11 展示了卷积神经网络的发展历程，图中实线代表改进后的模型，虚线表示具有突破性

的模型。

卷积神经网络的出现,为人们在日常生活中各个行业的应用提供了无限可能。其中,以 ResNet、AlexNet 和 VGGNet 为典型代表的网络为后面的一阶段和两阶段目标检测技术奠定了坚实的基础,以下是这些经典网络的特点。

(1) ResNet

残差网络(ResNet)由 He K 等人^[60]提出,在当年的目标检测比赛中取得第一名的好成绩。ResNet 作为经典卷积神经网络之一,其核心在于残差模块的设计。它有效缓解了神经网络中因梯度消失、梯度爆炸以及网络退化所导致的模型性能下降问题。图 2.12(a)为残差块结构图,相较于传统网络结构,它添加了跳跃连接并连接到第二层激活函数之前。这样的设计使得网络在不理想的情况下也能获得与输入相同的输出,不易出现网络退化的问题。图 2.12(b)为 ResNet 网络结构图,它由五个含有不同数量卷积核的卷积层、一个最大池化层、一个全局平均池化层、一个全连接层组成,其结构中包含残差模块。ResNet 被广泛应用于计算机视觉领域,涵盖图像分类、目标检测、人脸识别等应用场景,而且它在自然语言处理任务中也有着出色的表现,例如文本分类和机器翻译任务。

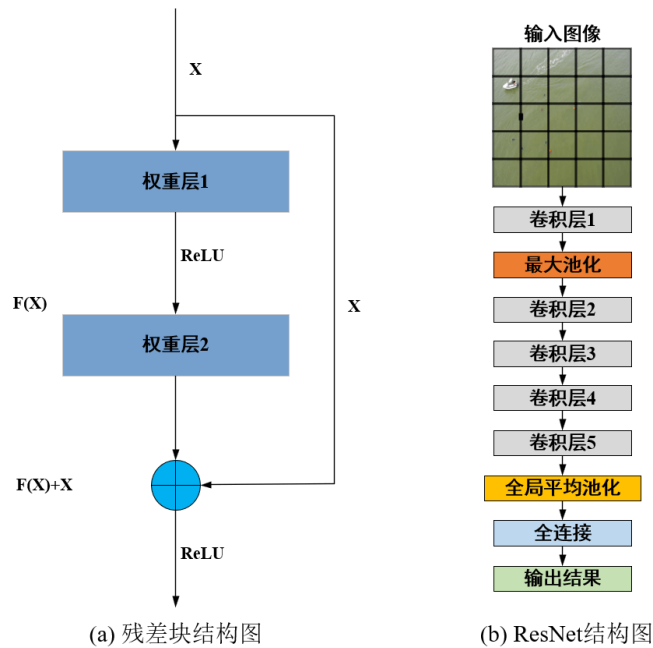


图 2.12 残差网络示意图

(2) AlexNet

AlexNet 由 Krizhevsky A 等人^[61]提出,在当年的 ImageNet 大规模视觉识别挑战赛中表现优异,将卷积神经网络带入了一个新的阶段。与基准模型 LeNet

相比，AlexNet 在其基础上增加了 3 层权重层，总共有 8 层权重层，其中包含 5 层卷积层和 3 层最大池化层。它还有不少创新性应用，例如将 ReLU 激活函数、Dropout 正则化和重叠池化应用于网络中，这不仅成功降低了过拟合的概率，还加快了训练进程。图 2.13 为 AlexNet 网络结构图，从图中可以看出网络参数量较大，因此需要在多个 GPU 上训练模型，以加快计算速度。

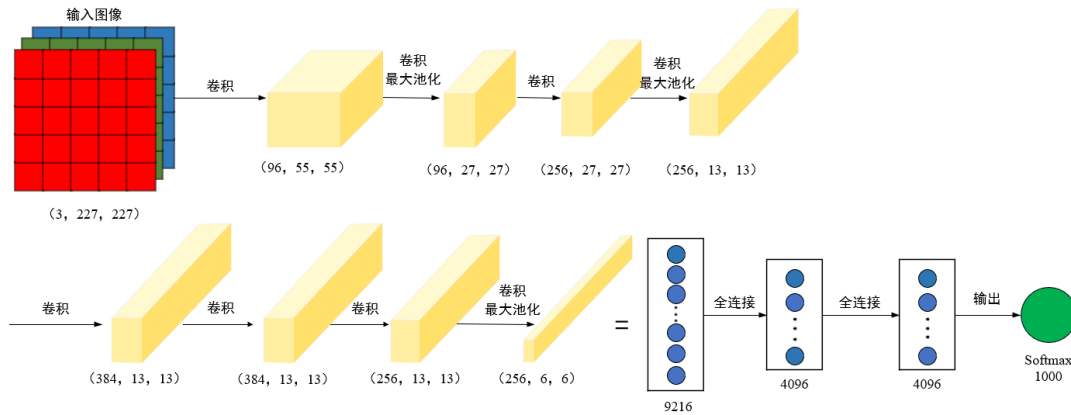


图 2.13 AlexNet 网络结构图

(3) VGGNet

VGGNet 由牛津大学视觉几何组的 Simonyan K 等人^[62]提出，在 ILSVRC2014 比赛中取得不错的成绩。它是在 AlexNet 的基础上进行了改进和扩展，网络结构被分成 5 段，每一段都由多个卷积核串联而成，每段卷积后都接一个最大池化层，在网络的末端串接着三个全连接层和一个 softmax 层。VGGNet 有两种结构，分别是 VGGNet-16 和 VGGNet-19，二者的区别在于 VGGNet-19 的层数比 VGGNet-16 更多，深度更深。

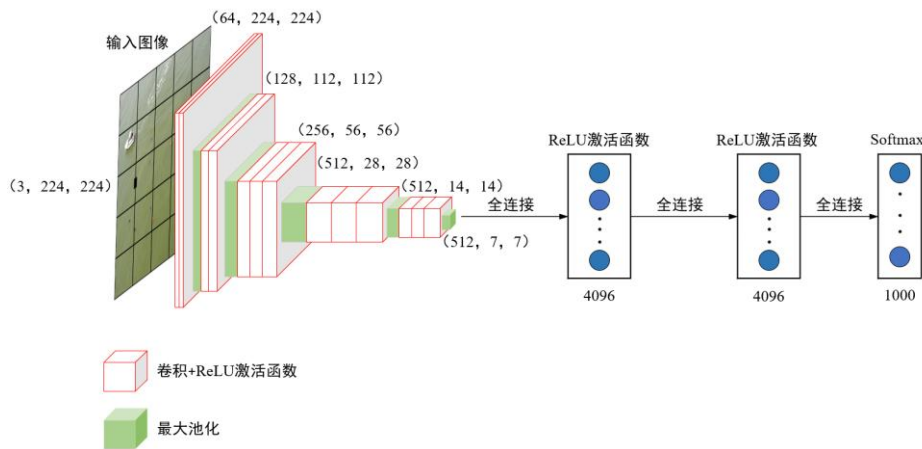


图 2.14 VGGNet 网络结构图

图 2.14 为 VGGNet-16 网络结构图，一方面它有效减少了网络的参数量，另

一方面它有更多的非线性变换，增强了自身的特征学习能力，VGGNet 为后续的一阶段目标检测 SSD 方法提供了理论基础。

2.1.3 基于深度学习的两阶段目标检测算法

两阶段目标检测模型的典型代表有 R-CNN、Fast R-CNN、Faster R-CNN 等。基于深度学习的两阶段目标检测算法包含两个主要步骤，其具体流程如下：第一步生成目标候选区域；第二步将生成的目标候选区域映射到特征图上，根据这些特征对目标进行精确分类，同时对边界框的位置和大小进行微调，进而得到更精准的检测结果。

(1) R-CNN

R-CNN 首次目标检测领域引入深度学习，并运用卷积神经网络。其实现过程分为三个步骤：第一步，利用 Selective Search 算法在每个图像上生成多个候选区域；第二步，将生成的多个候选区域缩放为固定尺寸的向量，以便进行特征提取；第三步，使用 SVM 分类器进行分类，同时使用回归器对边界框进行微调。R-CNN 网络结构图如图 2.15 所示。

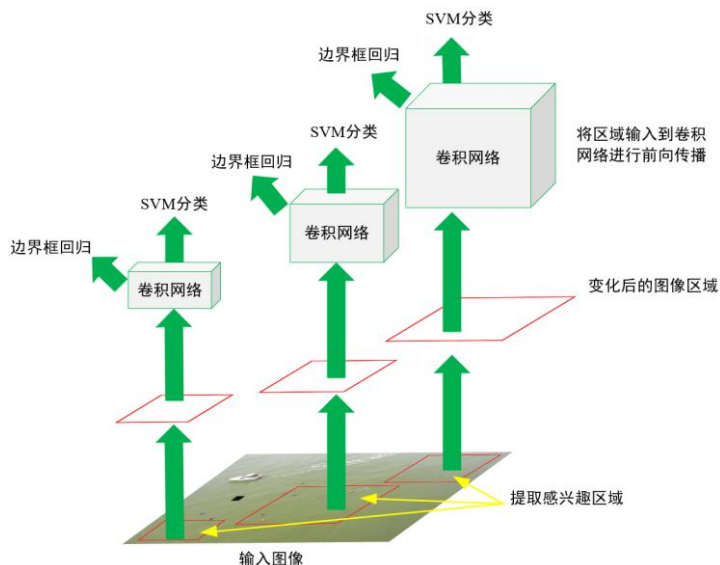


图 2.15 R-CNN 网络结构图

(2) Fast R-CNN

Ross Girshick 在 R-CNN 的基础上提出了 Fast R-CNN，很大程度上提高了目标检测的速度。Fast R-CNN 与 R-CNN 的区别在于，其网络中的卷积层不再针对图片中每个候选区域进行特征提取，而是对整个图像进行特征提取。除此之外，它还使用 Roi Pooling 进行特征的尺寸变换，并利用 softmax 替代了 R-CNN 中的

SVM 分类器。这些改进虽然减少了提取特征操作中的大量冗余，然而在选择目标区域时耗费大量时间。Fast R-CNN 网络结构图如图 2.16 所示。

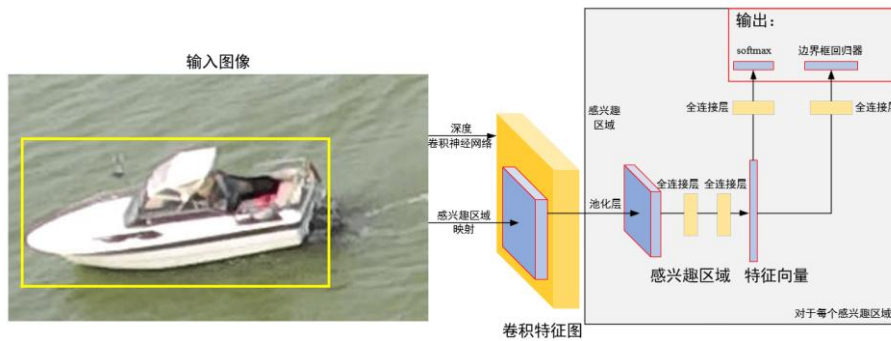


图 2.16 Fast R-CNN 网络结构图

（3）Faster R-CNN

Ross Girshick 时隔一年在 Fast R-CNN 的基础上提出了新的目标检测方法 Faster R-CNN，它不再使用传统的滚动窗口和选择性搜索方法，而是创造一种新的模块 RPN(Region Propose Network, RPN)对提取的卷积特征图进行处理，并生成检测框，这种方法极大程度地提升了目标检测的速度。Faster R-CNN 网络结构图如图 2.17 所示。

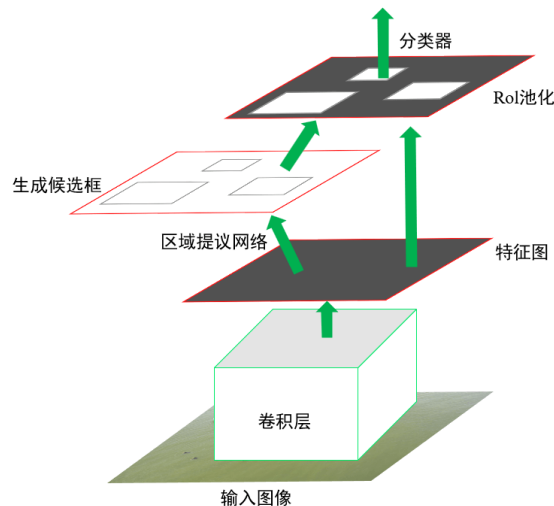


图 2.17 Faster R-CNN 网络结构图

2.1.4 基于深度学习的一阶段目标检测算法

由于两阶段目标检测方法步骤过于繁琐，研究者后期提出了一阶段目标检测方法。它通过简化流程，直接生成物体识别概率和位置坐标信息，只需一次操作得出检测结果。其中，典型的模型有 YOLO 和 SSD 两类，它们在检测速度、精度等方面各有特点。

YOLOv5 中的 C3 模块替换为 C2f 模块，检测头部分从 Anchor-Based 转变为 Anchor-Free。研究者在 Focal Loss 的基础上提出了 Varifocal Loss，这种方法有效保留了正样本的有用信息。

随着 YOLO 算法持续迭代发展，YOLO 系列目标检测算法如今广泛应用于道路监控、工业检测、自动驾驶等各个场景，为人们的生活带来了极大便利。

(2) SSD

SSD 检测算法是另一种一阶段目标检测方法，它是在 VGGNet-16 基础上增加了卷积层的数量来获取更多的特征图，以便更好地检测目标。首先，将原网络中第六和第七的全连接层分别替换成 3×3 的卷积层和 1×1 的卷积层，再将 VGGNet-16 的最后一个全连接层去除，并把最末端步长为 2， 2×2 的池化层换成步长为 1， 3×3 的池化层。

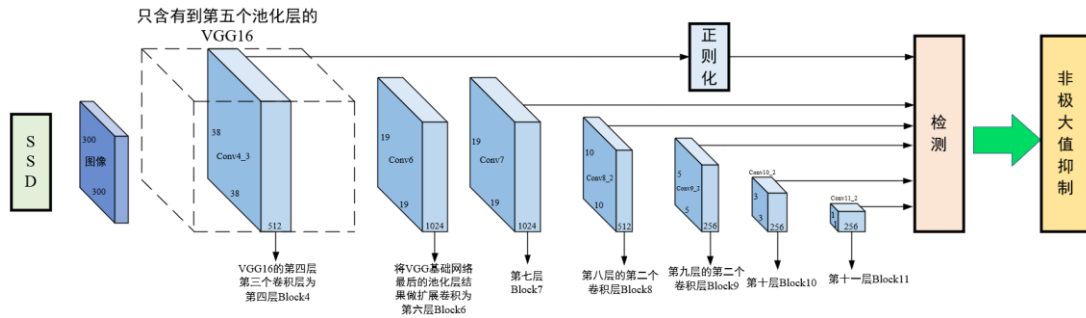


图 2.19 SSD 网络结构图

图 2.19 为 SSD 网络结构图。从图中可以看出，检测过程由三个步骤来完成：第一步将原始图像输入到改进后的 VGGNet-16 网络中，提取不同层级的特征图，其中深层特征图用于检测大目标，浅层特征图用于检测小目标；第二步在每个特征图的每个位置上，预先生成多个不同比例和大小的先验框，并通过卷积层为每个框预测类别概率和调整参数；最后一步对预测结果进行置信度阈值过滤，再通过非极大值抑制去除重复检测框，输出最终的检测边界框及类别。

2.2 目标跟踪理论

目标跟踪作为计算机视觉领域的重要研究方向，被广泛应用于智能交通、安防监控、军事行动等场景。通过在连续的视频帧序列中进行定位和状态估计，实现对特定目标的跟踪。其难点是如何在存在光照变化、目标遮挡、目标外观变化等因素的复杂场景下，准确且稳定地跟踪目标。考虑到上述因素会干扰目标跟踪的过程，研究者常用滤波方法和优化目标关联的方法来解决此类问题。因此，本

节主要介绍经典滤波方法-卡尔曼滤波和经典优化方法-匈牙利匹配。

2.2.1 卡尔曼滤波

卡尔曼滤波(Kalman Filter)^[63]是一种基于线性动态系统和高斯噪声假设的最优状态估计算法。它具有高效性和稳定性，所以广泛应用于目标跟踪领域。其核心步骤是通过预测和更新这两个阶段，来融合目标的历史运动状态和当前观测数据，实现对目标位置的连续估计。

在预测阶段，输入数据为检测器输出的结果，跟踪系统模型依据公式(2-11)和公式(2-12)构建：

$$\dot{x}_k = A_k \dot{x}_{k-1} + B_k u_k \quad (2-11)$$

$$P_k^* = A_k P_{k-1}^* A_k^T + Q_k^* \quad (2-12)$$

其中， $\dot{x}_k$ 表示第 k 帧的状态估计值， A_k 表示状态转移矩阵， B_k 表示控制输入矩阵， u_k 表示当前控制输入量， P_k^* 表示第 k 帧的状态估计误差协方差矩阵， Q_k^* 表示服从正态分布的过程噪声的协方差矩阵，代表外部不确定的动态因素。通过建立状态估计方程和过程噪声方程，实现对目标状态的动态预测。

在状态更新阶段先求解卡尔曼增益参数，用来平衡模型预测结果与实际测量数据之间的差异。利用卡尔曼增益对先前的预测状态进行调整，使其更接近当前观测值。通过含有预测误差和过程噪声的协方差矩阵来提升状态估计的精度。通过融合系统动力学模型与观测信息，最终输出最优的估计结果。数学表达式如公式(2-13)所示：

$$\begin{cases} K_k = P_k' H_k^T (H_k P_k' H_k^T + R)^{-1} \\ \dot{x}_k = \dot{x}_k' + K_k (z_k - H_k \dot{x}_k') \\ P_k^* = (I - K_k H_k) P_k' \end{cases} \quad (2-13)$$

其中， K_k 表示卡尔曼增益，用来决定预测与测量的权重分配。 P_k' 表示第 K 时刻的先验估计协方差， H_k 表示观测矩阵， R 表示测量噪声的协方差矩阵。 $\dot{x}_k'$ 表示第 K 时刻的先验状态估计量， z_k 表示第 K 时刻的观测值， I 表示单位矩阵。

2.2.2 匈牙利匹配

匈牙利匹配^[64]是一种基于图论的优化算法，主要用来解决目标 ID 分配问题。

在多目标跟踪场景下，通过对预测目标与检测目标进行关联，来保证目标 ID 的连续性。多目标跟踪的主要步骤如下：

（1）创建代价矩阵

将预测目标与检测目标之间的关联度量作为矩阵元素，构建成 $n \times m$ 的代价矩阵 C 。其中， c_{ij} 代表第 i 个预测目标与第 j 个检测目标之间的匹配成本。

（2）求解最优匹配

通过匈牙利算法对代价矩阵进行最优化求解。通过增广路径搜索策略和动态调整标记值，逐步消除行列冲突，找到成本最低的最优方案。最终，输出一个 $n \times m$ 的二元矩阵 M 。若 $M_{ij}=1$ ，则表明第 i 个预测目标与第 j 个检测目标成功关联。

（3）更新目标状态

在目标成功匹配后，需要利用卡尔曼滤波对目标状态进行动态更新。假设在当前跟踪过程中， $\hat{x}_i$ 为第 i 个目标的预测状态向量， z_j 为第 j 个检测目标的观测向量， H 为观测模型矩阵，则更新后的目标状态可以表示为：

$$\hat{x}_i = \hat{x}_i' + K(z_j - H\hat{x}_i') \quad (2-14)$$

其中，卡尔曼增益 K 是通过对方差矩阵进行动态调整后计算得出的。

（4）产生新的跟踪目标

对于未匹配成功的检测目标，它会将其被认定为视频中新出现的跟踪目标，并将其初始化后纳入跟踪序列，以便后续进行持续观测。

2.3 本章小结

本章首先介绍卷积神经网络，包含 LeNet、AlexNet、ResNet 等经典网络理论。接着阐述目标检测方法，包括以 R-CNN、Fast R-CNN 和 Faster R-CNN 为代表的两阶段算法，以及以 SSD、YOLO 为代表的一阶段算法。最后，介绍了目标跟踪理论的相关知识，包括卡尔曼滤波方法与匈牙利匹配方法。这些内容为本文提出的无人机海上目标检测与跟踪算法提供了必要的理论和技术支撑。

第3章 基于 YOLOv7 的晴天海上目标检测算法

无人机海上目标检测在船舶监控、海上搜救、海上巡逻等实际场景中具有重要的应用价值。然而，在无人机航拍视角下开展救援目标检测时，海上环境的特殊性带来了光照剧烈变化、海浪杂波干扰、目标尺度多样等不确定性因素，这使得海上目标检测极具挑战性。晴天时光照充足，受强反射光与海浪杂波的影响，海上目标的特征信息难以被检测模型捕捉。目前通用的目标检测算法在陆地场景中表现不错，但直接应用于海洋环境时会出现特征提取不充分、精度欠佳、漏检率高等问题。针对晴天时的目标检测场景适应性不足的问题，本章设计了一种基于 YOLOv7 的适用于晴天海上环境的高精度目标检测优化方法。

3.1 YOLOv7 检测模型

YOLOv7^[65]算法的网络结构由输入端(Input)、主干网络(Backbone)、头部网络(Head)三个部分构成。

在输入端处理阶段，对输入图像进行 Mosaic 数据增强操作，能够提升模型的泛化能力。同时，将输入图像缩放至统一尺寸，能确保批量处理时张量形状一致，便于 GPU 进行快速计算。

主干网络主要用来提取经输入端处理后的图像特征，其核心组成模块包括 ELAN-A、CBS 和 MPConv 模块。ELAN-A 由多个 CBS 模块组成，能有效增强网络学习特征信息的能力。MPConv 卷积层采用上下分支结构，增加用于下采样操作的 Maxpool 层，并通过拼接(Concat)操作融合上下分支所提取的特征，显著提升了网络的特征提取能力。YOLOv7 的头部网络由 SPPCSPC 和 ELAN-B、CBS 和 MPC 等模块组成，采用自上而下的特征金字塔结构(FPN)和自下而上的特征金字塔结构(PAN)。这些设计在一定程度上提升网络对不同尺度目标的特征提取能力。ELAN-B 通过增加拼接操作来扩大感受野，SPPCSPC 进一步扩展网络的感受野。FPN 和 PAN 分别通过上采样和定位信息传递来实现多尺度特征的融合，增强了网络在语义表达和目标定位方面的性能。通过在头部的三个检测层输出大、中、小三种不同尺度的特征图，在最后的输出特征图上进行结果预测，最终输出相应的目标框与类别信息。YOLOv7 网络结构图如图 3.1 所示。

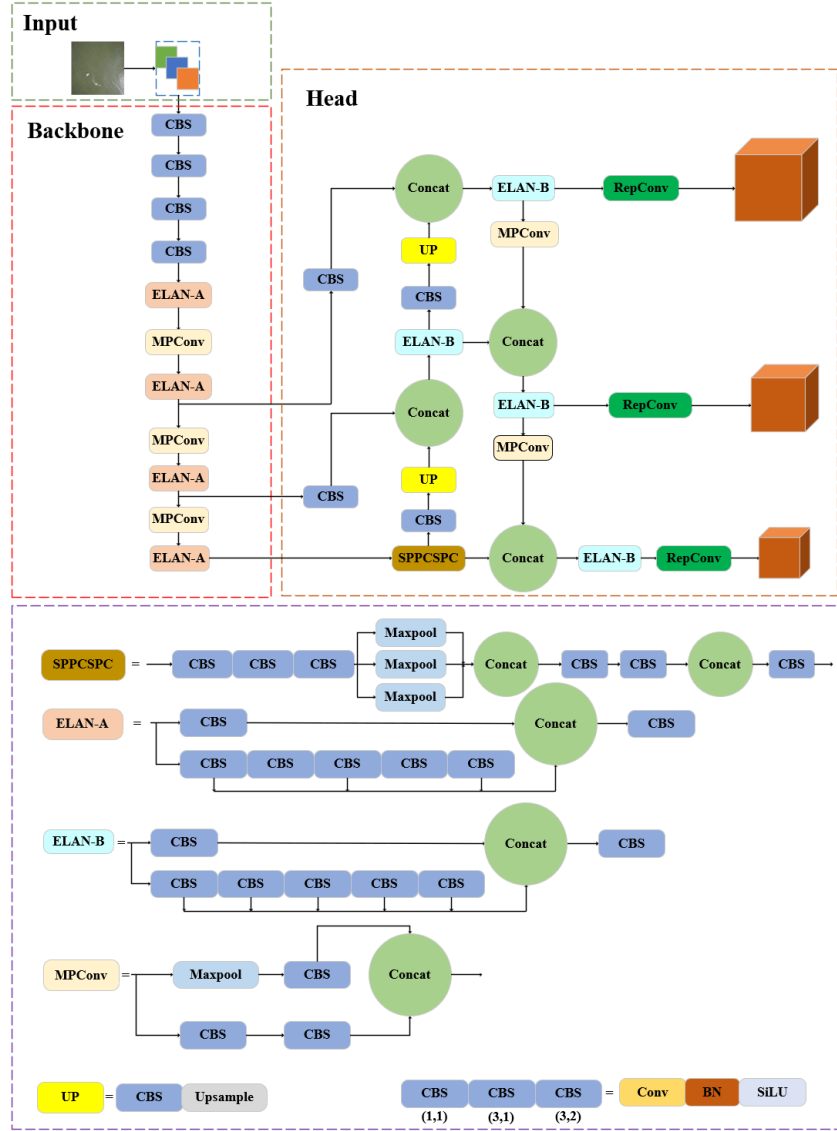


图 3.1 YOLOv7 网络结构图

3.2 YOLOv7 检测模型的改进

在光线充足时，海面上会投射太阳耀斑，加上海浪杂波的存在，可能导致图像中目标区域的特征被掩盖。此外，无人机在航拍过程中由于视角变换会造成多个目标相互遮挡。这些干扰因素会严重降低 YOLOv7 模型的检测精度。针对上述问题，在 YOLOv7 模型的基础上从特征提取优化、特征增强和损失函数优化这三个方面进行改进。首先，在网络中引入坐标卷积模块并替换掉部分标准卷积，这样能够弥补标准卷积无法确定目标坐标信息的缺陷，使模型能够更好地理解图像中的相对位置关系和空间分布，有助于更精确地定位海上目标。其次，引入注意力机制 CBAM，让模型更加聚焦于图像中的关键区域，进而抑制海浪杂波的

干扰。最后，运用 NWD 损失函数对 CIOU 损失函数进行优化，以更好地度量重叠且相似的小目标，提高模型对小目标的检测精度。改进后的 YOLOv7 模型结构如图 3.2 所示。

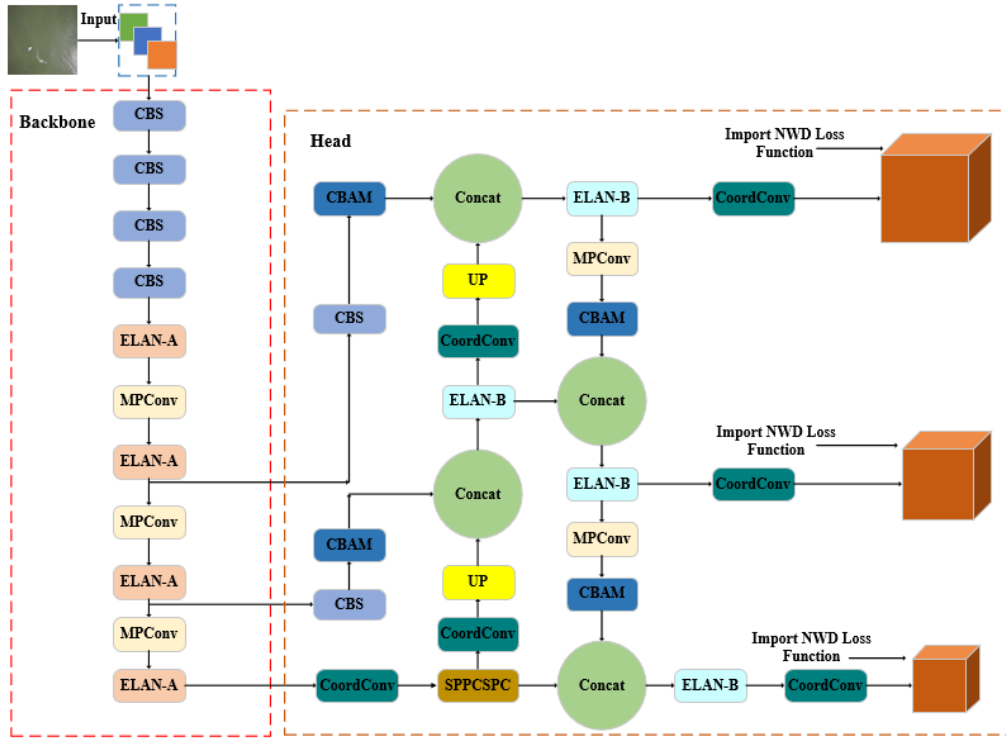


图 3.2 改进后的 YOLOv7 网络结构图

3.2.1 引入坐标卷积模块

晴天时，在强烈阳光照射下海面易产生反射光斑，致使图像中的目标特征变得模糊，极大程度上影响了检测效果。例如，海面上漂浮的小目标，如人、浮标等，在反射光的影响下难以被定位。此外，波浪纹理易被误检成诸如人这样的小目标。通过实验发现，YOLOv7 的标准卷积(Conv)只能关注局部信息，无法在特征图中感知确切的位置信息。例如，当小目标与反射光斑或波浪纹理相互交织时，标准卷积无法准确区分目标与干扰因素，进而导致检测精度下降。因此，引入坐标卷积(CoordConv)来帮助网络更好地定位小目标位置，进而优化特征提取。在 Head 网络中，使用坐标卷积替代 FPN 中的 1×1 Conv 以及检测头前的 RepConv。YOLOv7 的 Backbone 虽适合捕捉全局特征信息，但容易丢失细微的位置信息。所以在 Backbone 末尾添加了坐标卷积，为后续的 FPN 特征融合模块提供更丰富的空间信息，从而改善多尺度目标的检测效果。

坐标卷积是一种通过在标准卷积中添加坐标信息的卷积。其核心是在输入特

征图中增加了两个坐标通道，分别记录每个像素的水平坐标 i 和垂直坐标 j 。经过这样的预处理后，标准卷积能够处理原始数据和位置信息。对扩展后的特征图进行卷积操作，使得网络具备空间感知能力。如果在坐标通道中学习到特定信息时，坐标卷积具有平移依赖性。若在坐标通道中未学习到任何特定信息时，坐标卷积的特性与传统卷积相似，保持平移不变性。标准卷积如图 3.3(a)所示，坐标卷积如图 3.3(b)所示。从图中可以看出，输入特征图的尺寸为 $H \times W \times C$ ，其中 H 表示高度， W 表示宽度， C 表示通道数。添加坐标通道后，输入特征图的尺寸变为 $H \times W \times (C+2)$ 。

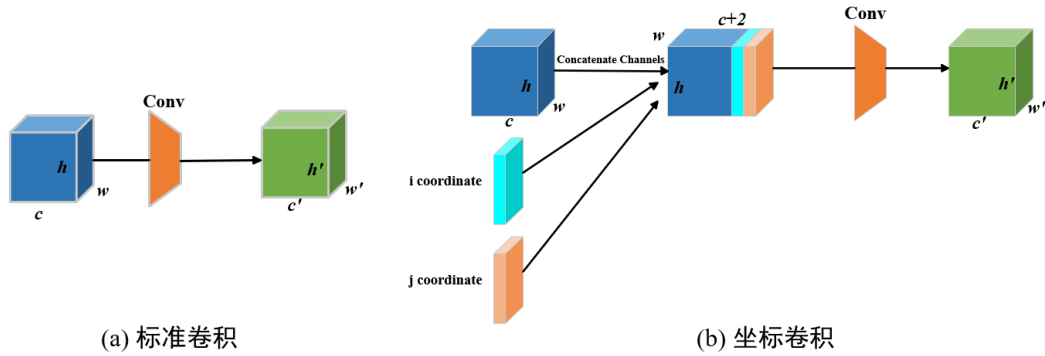


图 3.3 两种卷积结构图

3.2.2 加入 CBAM 注意力机制

注意力机制的设计灵感源自人类视觉系统的选择性注意特性。人眼在采集图像信息时，会主动将注意力聚焦于图片中的关键区域，忽略其他区域的次要信息，从而高效且精准地获取图像中的信息。这种生物视觉机制为计算机视觉中的特征优化提供了重要启示。

在无人机视角下的海上目标检测任务中，存在海面上的浮标、海洋中的人等小目标。它们只占较小的像素面积，容易出现错检，漏检和误检等情况。而且海上目标往往受到尺度变化、海浪等因素的影响，导致目标特征难以提取。本章将 CBAM 注意力机制模块融入到特征增强提取网络中，有助于提取更具区分度的小目标特征，从而针对性解决海洋背景干扰下难以提取目标特征的问题。

CBAM 注意力机制由通道注意力和空间注意力两个子模块组成。它通过在卷积操作中调整注意力权重，让网络能够学习图像中的特征信息，进而提高模型对目标特征的关注程度。

在通道注意力模块中，首先对输入的特征映射进行两种不同的池化操作，分

别是最大池化和全局平均池化。假设输入特征图的尺寸为 $H \times W \times C$ ，经过这两种池化操作后，会得到两个维度均为 $1 \times 1 \times C$ 的特征映射。然后，将这两个特征映射输入到一个两层的感知器中。该感知器(MLP)的第一层有 C/r 个神经元，并采用 ReLU 作为激活函数；第二层有 C 个神经元，且两层神经网络共享相同的权值参数。接着，对这两层神经网络的输出特征进行逐元素相加操作，再通过 sigmoid 激活函数进行处理，最终生成通道注意力特征。最后，将原始的输入特征映射与生成的通道注意力特征按元素相乘，所得结果将作为后续空间注意力模块的输入特征。图 3.4 为通道注意力模块结构图，图中 M_c 表示通道注意力向量。

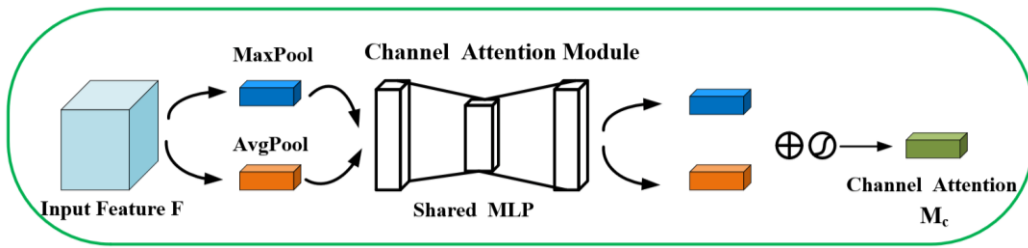


图 3.4 通道注意力结构图

空间注意力模块以前一步得到的特征映射作为输入。对该输入特征映射分别进行全局最大池化(Global MaxPool)和全局平均池化(Global AvgPool)操作，得到两个维度均为 $H \times W \times 1$ 的特征图。随后，将这两个特征图在通道维度上进行拼接。拼接后的特征图会经过卷积操作进行降维处理，再通过激活函数处理，生成空间注意力特征。最后，将输入的特征图与生成的空间注意力特征按元素相乘，得到的结果即为经过 CBAM 注意力机制处理后的最终特征图。图 3.5 为空间注意力模块结构图，图中 M_s 表示空间注意力向量。

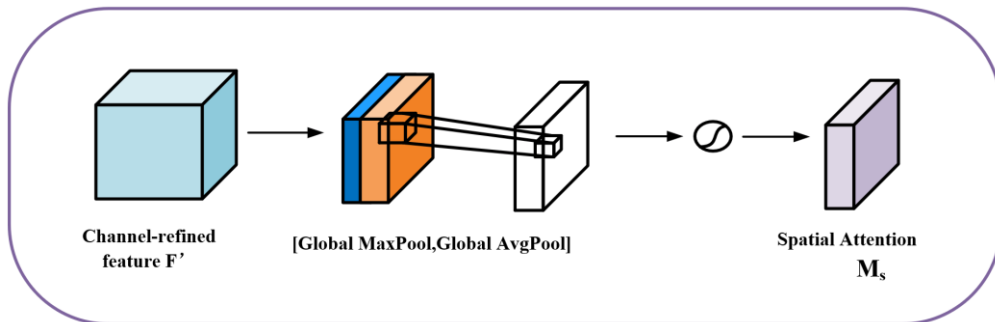


图 3.5 空间注意力结构图

3.2.3 优化损失函数

损失函数的核心作用是衡量模型预测结果与真实标签之间的差异程度。在目

标检测任务中，选择恰当的损失函数尤为重要，它能够显著加快神经网络的收敛进程，同时提升目标检测的精准度。在 YOLOv7 模型的训练过程中，需要借助三个损失函数来全面评估模型的性能表现，分别是边界损失、置信度损失以及分类损失。在计算边界损失和置信度损失时，采用的是交叉熵损失函数，能够有效地衡量预测值和真实值之间的概率分布差异。而在计算边界框损失时，则运用了 CIOU 损失函数，这种方法充分考虑了边界框的重叠面积、中心点距离以及长宽比等因素，能更精准地评估边界框的预测效果。CIOU 的公式如下：

$$L_{CIOU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \left(\frac{v}{(1 - IoU) + v} \right) v \quad (3-1)$$

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

$$\frac{\partial v}{\partial w} = \frac{8}{\pi^2} \left(\tan^{-1} \frac{w}{h} - \tan^{-1} \frac{w_{gt}}{h_{gt}} \right)^2 \frac{h}{h^2 + w^2} \quad (3-3)$$

$$\frac{\partial v}{\partial h} = -\frac{8}{\pi^2} \left(\tan^{-1} \frac{w}{h} - \tan^{-1} \frac{w_{gt}}{h_{gt}} \right)^2 \frac{w}{h^2 + w^2} \quad (3-4)$$

在公式(3-1)中 b 表示预测框的中心点， b_{gt} 表示真实框的中心点。 ρ 表示这两个中心点之间的欧氏距离，即在空间中这两个点间的直线距离。 c 表示真实框和预测框的最小外接矩形的对角线长度， v 用于衡量预测框和真实框宽高比的一致性。预测框的宽度 w 和高度 h 相对于 v 的梯度分别如公式(3-3)和公式(3-4)所示。从这两个公式可以看出，它们成反比关系， w 和 h 不能同时朝着增大或者减小的方向变化。当预测框的宽高比满足特定条件时，CIOU 的惩罚函数会失效，该损失函数可能无法有效地促使网络收敛。

在无人机所采集的图像中有些小目标仅占少量像素，导致检测模型难以获取目标特征信息。CIOU 在处理这类小目标时，对位置偏差极为敏感，这使得模型在小目标检测方面的性能大幅下降。为了有效解决小目标自身对 IOU 的高敏感性问题，本章采用了一种优化策略。具体来说，将归一化 Wasserstein 距离 (Normalized Wasserstein Distance, NWD) 与 CIOU 相结合，对边界框的回归损失函数进行优化。

NWD 是一种基于归一化高斯 Wasserstein 距离的目标框相似性度量方法。该

方法主要包含两个核心步骤：将预测框和真实框分别构建为二维高斯分布模型；利用 NWD 计算两者之间的相似度。与传统度量方式相比，NWD 具有两个显著优势：不受小目标之间是否重叠的限制，均能对目标相似性进行有效度量；其对目标的尺度变化敏感度较低，能够更全面地捕捉小目标的特征信息。这些特性使得 NWD 能够提供更为精确的度量结果。因此，在衡量小目标的真实框和预测框之间的相似性时，该方法是一个较为合适的选择。

在进行海洋场景中的小型目标检测任务时，传统矩形标注框在定位精度方面存在缺陷。由于物体尺寸较小，检测器生成的预测框难以精准且完全地贴合目标真实轮廓，不可避免地引入部分无关背景元素的干扰。考虑到海上救援场景中人、船只物体形状丰富多样，使用矩形边界框完整地框住所有物体并不现实。这种局限性是客观存在的，目标的有效像素大多相对集中地分布于检测框的几何中心区域，而背景像素更倾向于散布在标注框的边缘区域。为了充分地挖掘边界框内每个像素的价值，可为中心区域的像素赋予较高的权重，并且会随着像素点与中心点欧式距离的增加而逐渐降低，直至边界框的边界位置。将边界框 $B=(c_x, c_y, w, h)$ 建模为一个二维高斯分布，其二维高斯分布 $N(\mu, \Sigma)$ 的公式为：

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

其中 (c_x, c_y) 表示边界框的中心坐标， w 和 h 表示分别对应着边界框的宽度与高度。

在计算边界框之间的分布距离时，选用 Wasserstein 距离作为度量方式。对于基于两个不同边界框构建的二维高斯分布 N_1 和 N_2 ，它们之间的 Wasserstein 距离如公式 (3-6) 所示。

$$W_2^2 = \left\| \left([cx_1, cy_1, \frac{w_1}{2}, \frac{h_1}{2}]^T, [cx_2, cy_2, \frac{w_2}{2}, \frac{h_2}{2}]^T \right) \right\|_2^2 \quad (3-6)$$

归一化高斯 Wasserstein 距离可表示为：

$$NWD(N_1, N_2) = \exp\left(-\frac{\sqrt{W_2^2(N_1, N_2)}}{C}\right) \quad (3-7)$$

公式 (3-7) 表示 NWD 度量方法，其中 C 表示与数据集相关的常量。基于此，基于

NWD 的损失函数公式可以表示为:

$$Loss_{NWD} = 1 - NWD(N_1, N_2) \quad (3-8)$$

考虑到在本章所使用的数据集中,目标并非均为小目标,还包含其他尺寸类型的目标。因此,将 $Loss_{NWD}$ 与 $Loss_{CIOU}$ 结合起来构建新的边界框损失函数 $Loss_{NWD-CIOU}$ 。同时,为 $Loss_{NWD}$ 和 $Loss_{CIOU}$ 分别赋予不同的权重。 $Loss_{NWD-CIOU}$ 的公式为:

$$Loss_{NWD-CIOU} = \lambda Loss_{CIOU} + (1 - \lambda) Loss_{NWD} \quad (3-9)$$

其中 λ 表示 $CIOU$ 所占的比例值。

然而,对于这两种度量方式所设定的不同比例系数,会对检测结果产生不同的影响。为了验证本章提出的 $Loss_{NWD-CIOU}$ 损失函数的有效性,在添加了 CoordConv 和 CBAM 两个模块的 YOLOv7 模型上通过分配不同的比例值 λ 进行对比实验。实验结果如表 3.1 所示。

表 3.1 不同分配比例下的实验结果

比例值 λ	Precision	Recall	mAP@0.5
0.00	95.2%	88.3%	92.7%
0.15	95.5%	88.7%	93.1%
0.25	95.8%	88.9%	93.4%
0.35	95.4%	88.5%	92.9%
0.45	95.1%	88.3%	92.6%
0.55	94.5%	87.5%	92.2%
0.65	94.2%	86.8%	91.7%
0.75	93.9%	86.2%	91.3%
0.85	93.4%	85.5%	90.7%
0.95	93.1%	84.7%	90.0%
1.00	92.7%	83.9%	89.4%

从表 3.1 可以看出, $CIOU$ 损失在损失函数中所占比例对检测结果影响极为明显。当将 $CIOU$ 损失的比例值设定为 0.25 时, NWD 损失的比例值为 0.75 时,模型的检测效果较为不错,检测精度达到 95.8%,召回率为 88.9%, $mAP@0.5$ 为 93.4%。综合上述实验结果能够得出, NWD 能够有效地平衡样本之间的差异。将它与 $CIOU$ 结合使用,能够更为精准地度量真实边界框和预测边界框之间的相似性,进而显著提升模型在小目标检测方面的性能,使模型在识别小目标时更加准确。

3.3 晴天海上目标检测数据集制作

3.3.1 图像采集

本章实验的数据集来源于 SeadronesSecV2 数据集，该数据集是一个基于无人机航拍视角下的海上救援大规模目标检测数据集。它为 WACV'2023 官方指定的海事挑战赛数据集，开放在 Kaggle 网站上。其中包含了数万张图像，有 3736 张图像拍摄于强光照条件下。因此，从整个数据集中挑选出这 3736 张图像，制作成了晴天海上目标检测数据集。此数据集中待检测目标数量达到一万以上，共包括四个类别，分别为船只、水上摩托艇、人和漂浮物。在自制的数据集中，小目标中超过一半的都是人。最后，使用 Labelimg 标注软件对图像中的目标进行标注，生成 YOLO 格式的标签文件。本章实验数据集按照 8:1:1 的比例划分训练集、验证集和测试集。

3.3.2 图像标注

图像标注主要包括人工标注、半自动标注和自动标注这几种方式，其中半自动标注和自动标注容易产生标注误差。因此，本章选用 Labelimg 标注工具对海上目标数据集进行人工标注。Labelimg 是一款专为图像标注而开发的开源工具，在图像标注领域应用广泛。它具有强大的功能，可生成三种不同格式的标签。第一种为 Pascal VOC 格式，文件以 XML 形式存储，其中含目标类别和边界框坐标信息。第二种是 YOLO 格式，采用 TXT 文本文件存储数据，每行有目标类别索引和归一化的边界框坐标，常用于 YOLO 检测算法。第三种以 JSON 文件保存，含类别名称和边界框坐标等信息。Labelimg 标注数据的过程如图 3.6 所示。

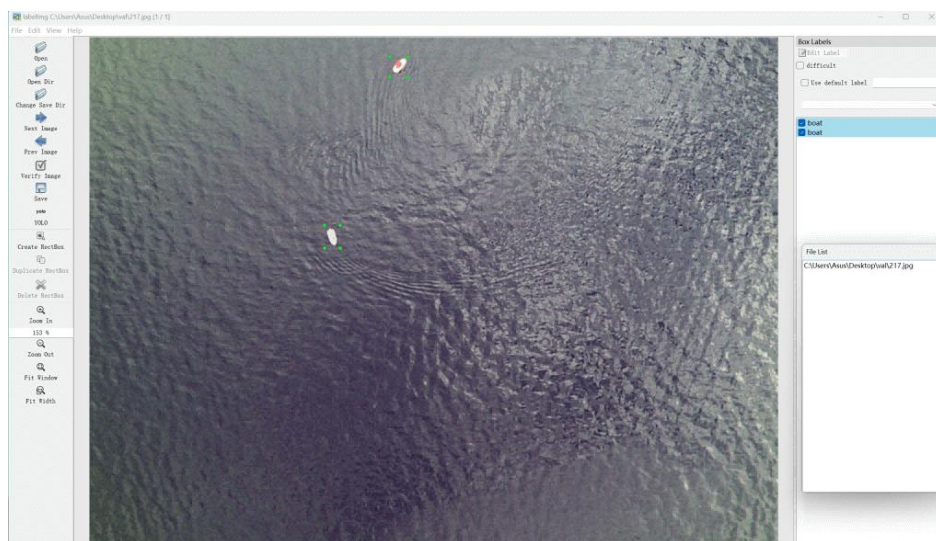


图 3-6 Labelimg 数据标注

3.4 实验结果与分析

3.4.1 实验环境及配置

本章以 Pycharm 作为实验平台，运用 Python 编程语言，并借助深度学习开发框架 Pytorch，对改进前后的 YOLOv7 模型展开训练。整个实验所需的环境配置如表 3.2 所示。

表 3.2 实验环境配置

实验环境	配置
操作系统	Ubuntu20.04
开发框架	PyTorch1.10.0
编程语言	Python3.8
GPU	RTX3090(24GB)
CUDA	11.3

实验前需要进行超参数的设定。其中，训练总轮数设为 300，批大小为 16，优化器选用 SGD，初始学习率为 0.01，权重衰减系数设置为 0.0005。

3.4.2 目标检测评价指标

为了准确评估模型在晴天海上目标检测图像任务中的检测性能，本文采用了认可度极高的 mAP(map@0.5)、精确率 P(Precision)、召回率 R(Recall)、F1 分数 (F1-Score)作为评价指标。其中，mAP@0.5 表示当 IOU 阈值达到 0.5 时所有类别的平均精度；P 指在预测为正样本中，真实为正的样本所占的比例；R 指在实际正样本中，预测为正样本的比例；F1 分数是一个综合衡量 P 和 R 的指标，常用于目标分类任务，F1 越高，说明分类效果越好。P、R、F1、mAP 的计算公式如下：

$$P = \frac{TP}{TP + FP} \quad (3-10)$$

$$R = \frac{TP}{TP + FN} \quad (3-11)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (3-12)$$

$$AP = \int_0^1 P(R) dR \quad (3-13)$$

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \quad (3-14)$$

在公式(3-10)到公式(3-14)中，其中 TP 为正确检测框的数目，FP 为误检框的数目，

FN 为漏检框的数目。 AP 为单个类别的检测精度， k 是检测目标的类别数量。

3.4.3 改进前后检测模型对比实验分析

在本章的研究中，采用 YOLOv7 模型对晴天海上的目标进行检测。然而，由于受到光照强度、拍摄角度以及距离远近等条件的影响，导致该模型的检测效果不佳。为解决上述问题，本章从特征提取优化、特征增强、优化损失函数这三个方面对 YOLOv7 模型进行改进，以提升检测效果。为了验证改进后模型的检测效果，分别对改进前后的 YOLOv7 模型进行训练。训练结束后，选取最优权重在测试集上进行检测实验，并将模型改进前和改进后的检测结果进行详细对比分析。为了确保实验结果的准确性与可靠性，在整个实验过程中严格控制变量，实验环境、数据集以及模型超参数设置均保持一致。改进前后的实验结果如下表 3.3 所示。

表 3.3 改进前后的实验结果

Model	Precision	Recall	F1	mAP@0.5
YOLOv7	89.6%	79.5%	0.842	86.7%
YOLOv7-Improved	95.8%	88.9%	0.922	93.4%

从表 3.3 可以看出，改进后的精确率提升了 6.2%，召回率提升了 9.4%，F1 分数提升了 0.08，mAP@0.5 提升了 6.7%。由此可见，改进后的模型在检测精度方面有显著提升。

改进前后的 YOLOv7 模型的 Precision-Recall 曲线图分别如图 3.7(a)和图 3.7(b)所示。

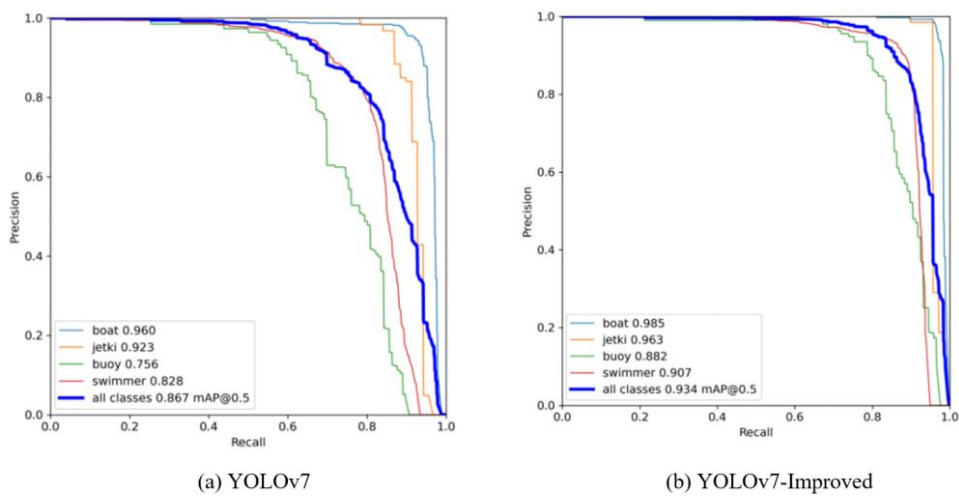


图 3.7 Precision-Recall 曲线对比图

通过观察图中曲线所围成的面积大小可知，图 3.7(b)中曲线包围的面积更大，这表明所有类别的 mAP 值更高。当 IOU 阈值为 0.5 时，相较于改进前，mAP 值

增长了 6.7%，这表明了改进后的模型检测性能优于原始模型。

3.4.4 消融实验分析

为了验证本章所提出的改进算法在海上目标检测中的有效性，在保持实验环境、数据集以及模型超参数设置一致的前提下，将坐标卷积、CBAM 注意力机制和 NWD 损失函数融入到初始 YOLOv7 网络中，并设计了四组消融实验，用于评估改进算法在海面目标检测中的性能表现。消融实验结果如表 3.4 所示。

表 3.4 消融实验结果

Group	CoordConv	CBAM	NWD	Precision	Recall	F1	mAP@0.5
1				89.6%	79.5%	0.842	86.7%
2	√			89.6%	82.3%	0.859	87.1%
3	√	√		92.7%	83.9%	0.881	89.4%
4	√	√	√	95.8%	88.9%	0.922	93.4%

从表 3.4 中可以看出，第 1 组是原始 YOLOv7 模型的实验结果。第 2 组实验将坐标卷积添加到 YOLOv7 网络中，赋予网络空间位置信息感知能力，使其能够更有效地定位图像中的目标。相比于基准实验(第 1 组)，各项指标均有一定程度的提高。第 3 组实验是在第 2 组实验的基础上引入 CBAM 注意力机制，mAP@0.5 值提高了 2.3%，准确率显著提高，召回率提高了 1.6%。实验结果表明，CBAM 模块有助于在检测过程中减少背景信息的干扰，使模型能够更加聚焦于目标自身，进而提高目标检测的准确性，减少目标漏检数量。第 4 组实验是在第 3 组实验的基础上，使用了 NWD 损失函数对原始损失函数进行优化。优化后，mAP@0.5 值增加了 4%，准确率提高了 3.1%，召回率提高了 5%，F1-Score 也显著提升。其中，召回率的提升表明优化后的损失函数能够使预测框更接近于真实框，这对于解决相互遮挡场景下的目标检测问题尤为适用。通过对消融实验结果中的各项评价指标进行综合分析，充分验证了本章改进算法的有效性。

3.4.5 不同检测模型对比实验分析

为了进一步验证本文改进算法在海上目标检测任务中的性能表现，将其与目前主流的目标检测算法(SSD、YOLOv3-tiny、YOLOv5-n、YOLOv7-x、YOLOv8-n、YOLOX-s)在同一数据集上进行了对比实验，并以 mAP@0.5、Precision、Recall、F1 作为对比实验的评价指标。

实验结果如表 3.5 所示。由表 3.5 可知，相较于其他算法，本文算法在 Precision、Recall、F1-Score、mAP@0.5 指标上都有了明显的提升。这充分表明了本文算法

在晴天场景下对海上目标具有更好的识别能力，展现出更优异的检测效果。

表 3.5 对比实验结果

Model	Precision	Recall	F1	mAP@0.5
YOLOv3-tiny	0.877	0.746	0.806	0.801
YOLOv5-n	0.889	0.793	0.838	0.833
SSD	0.957	0.455	0.617	0.856
YOLOv7	0.896	0.795	0.842	0.867
YOLOv8-n	0.899	0.828	0.862	0.874
YOLOX-s	0.916	0.853	0.883	0.882
YOLOv7-x	0.914	0.814	0.861	0.884
YOLOv7-Improved	0.958	0.889	0.922	0.934

为了更直观地展现改进方法的检测效果，本文对改进后的 YOLOv7 与表 3.5 中的主流算法进行了可视化对比实验。实验选取了包含相互遮挡的物体以及投射在海面上的眩光的图像进行检测，将置信度阈值设为 0.5，可视化对比结果如图 3.8 所示。

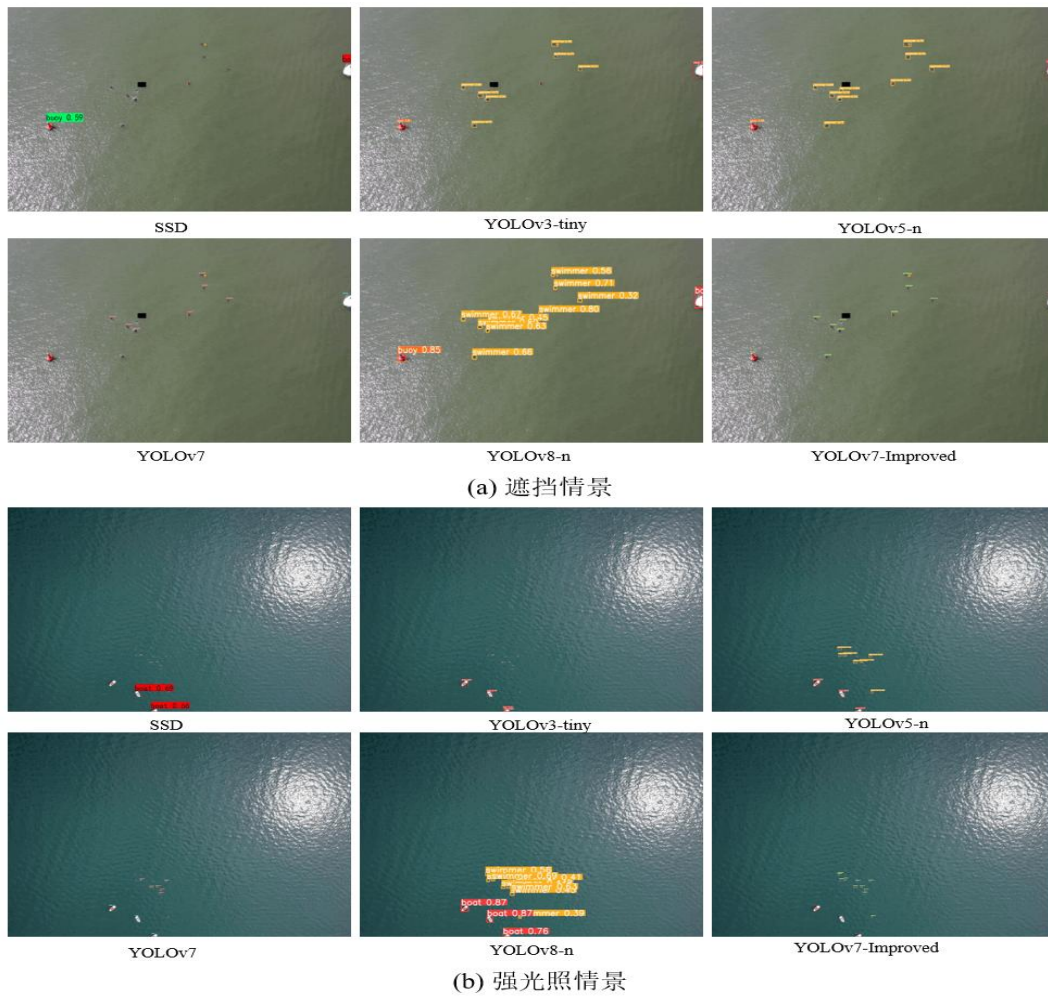


图 3.8 可视化结果对比图

从图 3.8(a)可以看出，与其他算法相比，改进后的方法能够检测出相互遮挡

的物体，例如游泳者周围的静态漂浮物。图 3.8(b)呈现的是一束强光照射到大海场景下的检测结果。与其他算法相比，改进后的方法可以检测到更多的目标。可视化结果表明，本文提出的检测方法能够有效解决本章提及的检测问题。

3.5 本章小结

本章首先详细介绍了 YOLOv7 的网络模型结构。然后，针对 YOLOv7 模型在晴天环境下海上目标检测效果欠佳的问题，设计了一种更适用于晴天环境下的高精度海上目标检测算法。该算法通过在 YOLOv7 网络中引入新的卷积 CoordConv、引入 CBAM 注意力机制，以及采用 NWD 损失函数优化原始损失函数，显著提升了检测性能。为了综合验证改进算法的有效性，在自制的晴天海上目标检测数据集上对改进后的模型进行消融实验，并与主流模型进行了对比实验。实验结果证明，改进后的模型能够实现对晴天环境下海上目标的高精度检测。

第 4 章 基于 YOLOv7-tiny 的雾天海上目标检测算法

本文第三章所设计的改进方法实现了晴天条件下海上目标的高精度检测，充分体现了模型在图像清晰的条件下具有较好的检测能力。然而，在实际检测任务中图像不局限于晴天，雾天等恶劣天气对图像质量的影响更为显著。特别是海雾天气，这种环境会使无人机采集的图像画质下降，会影响目标检测效果。如何在雾天这种复杂场景下仍能保持较高精度，又能兼顾一定的检测速度，成为一个棘手的问题。因此，本章设计了一种基于 YOLOv7-tiny 的雾天目标检测优化方法。

4.1 YOLOv7-tiny 检测模型

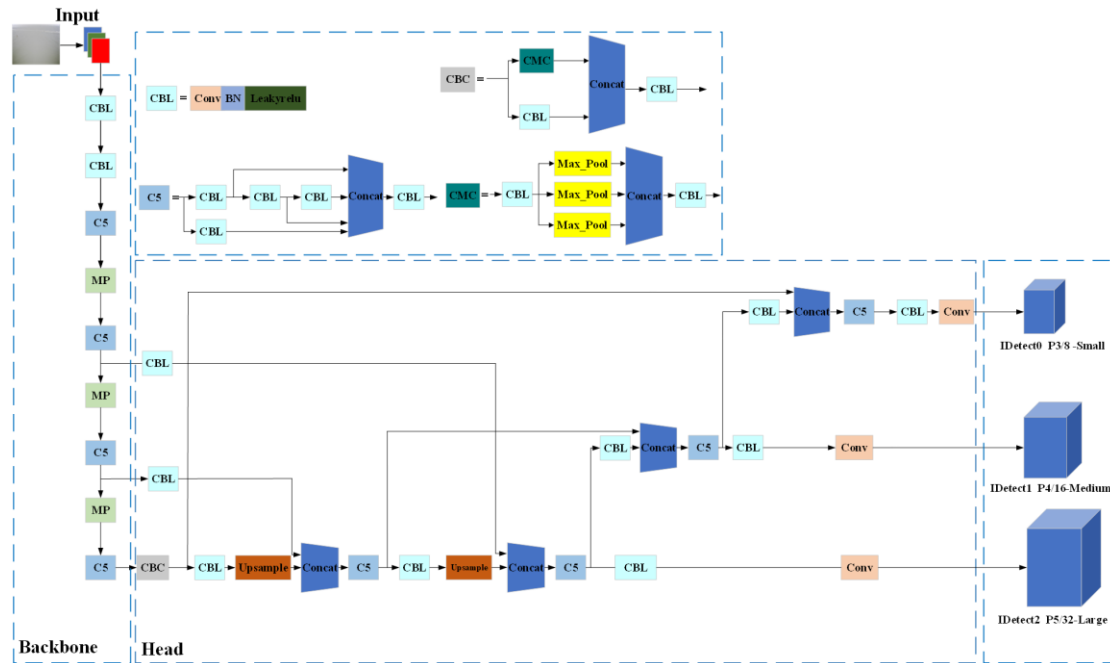


图 4.1 YOLOv7-tiny 网络结构图

YOLOv7-tiny 与 YOLOv7 的整体结构相似，均由输入端(Input)、骨干网络(Backbone)和头部网络(Head)三部分构成。输入图像经输入端进入骨干网络，再通过头部网络输出三个不同大小的特征图，最后调整通道数输出预测结果。YOLOv7-tiny 算法采用了多尺度预测策略，在不同的特征图上进行目标检测，进而有效地检测不同尺寸和比例的目标。YOLOv7-tiny 模型采用 CBL 卷积块，将 YOLOv7 模型 CBS 模块中的 SiLu 激活函数替换为 Leaky ReLU。在主干网络中去掉了两个卷积块，还用传统的最大池化 MaxPool 替代原网络的 MPconv 进行下采样操作。YOLOv7-tiny 的模型结构图如上图 4.1 所示。

与晴天检测中采用的 YOLOv7 模型不同的是, YOLOv7-tiny 是 YOLOv7 系列中的轻量级模型, 在计算量上进行了优化。雾天图像需去雾预处理以改善画质, 该环节会增加所用时间, 需要在检测模型上面使用轻量化模型, 可以节约时间, 提高效率。YOLOv7-tiny 推理速度更快、计算量更低, 能显著节约检测时间, 与预处理环节形成效率协同, 在保证缩短检测时间的同时又兼顾一定精度。因此, 不再使用 YOLOv7 而采用 YOLOv7-tiny, 并对其改进以应对雾天检测挑战。

4.2 基于去雾算法的图像预处理

海洋环境复杂, 海面水汽充足, 阳光照射后水汽发生折射、散射的现象, 光线与目标反射光交织, 极易形成雾气。当无人机执行海上搜救任务时, 雾气干扰成像过程, 导致目标模糊、图像失真, 这严重影响了后续的目标检测工作, 降低了救援行动的准确性与效率。因此, 对采集到的图像进行去雾处理是十分必要的。

本节主要介绍单尺度 Retinex 去雾算法^[66]、暗通道先验去雾算法^[67]和 AODNet 去雾算法^[68]。同时, 我们对这三种算法的去雾效果以及去雾速度进行对比。通过对比实验, 找出一种最优去雾算法, 将其作为图像预处理方法。

4.2.1 单尺度 Retinex 去雾算法

Retinex 理论认为, 人眼对物体颜色和亮度的感知取决于物体对不同波长光线的反射能力, 而不是由入射光的强度决定。基于此理论, 单尺度 Retinex 去雾算法通过对图像的光照分量 $L(x, y)$ 和反射分量 $R(x, y)$ 进行分离, 增加反射分量, 抑制光照分量中的雾气成分, 进而达到去雾的目的。单尺度 Retinex 去雾算法具体步骤如下:

(1) 首先, 对有雾图像 $S(x, y)$ 进行对数操作, 其表达式为:

$$S(x, y) = R(x, y) \cdot L(x, y) \quad (4-1)$$

$$\log(S(x, y)) = \log(R(x, y)) + \log(L(x, y)) \quad (4-2)$$

(2) 其次, 提取出反射分量, 其表达式为:

$$\log(R(x, y)) = \log(S(x, y)) - \log(L(x, y)) \quad (4-3)$$

(3) 然后, 分离光照分量和反射分量以实现去雾。其表达式为:

$$R(x, y) = \exp[\log(S(x, y)) - \log(L(x, y))] \quad (4-4)$$

(4) 运用高斯函数，对有雾图像中的光照分量进行预估，其表达式为：

$$L(x, y) = S(x, y) * G(x, y) \quad (4-5)$$

其中， $*$ 表示卷积运算， $G(x, y)$ 为高斯函数，其表达式如公式(4-6)所示。

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x^2 + y^2)}{2\sigma^2}\right] \quad (4-6)$$

其中， σ 为滤波尺度。

(5) 结合公式(4-4)、公式(4-5)和公式(4-6)，可得单尺度 Retinex 去雾算法公式：

$$R(x, y) = \exp[\log(S(x, y)) - \log(S(x, y) * G(x, y))] \quad (4-7)$$

本小节采用单尺度 Retinex 去雾算法，分别对薄雾和浓雾条件下的图像分别进行去雾处理。去雾效果如图 4.2 所示，其中图 4.2(a)展示的是薄雾图像去雾后的效果，图 4.2(b)呈现的是浓雾图像去雾后的效果。



(a) 薄雾去雾效果图



(b) 浓雾去雾效果图

图 4.2 单尺度 Retinex 去雾效果图

4.2.2 暗通道先验去雾算法

He 等人提出了暗通道先验去雾理论，该理论指出图像中部分像素至少在一个颜色通道上具有较低的灰度值。此外，对于无雾图像的非天空区域，其暗通道值趋近于 0。

暗通道模型的表达式为：

$$J^{dark}(x, y) = \min_{C \in \{R, G, B\}} (\min_{x, y \in \Omega(x, y)} (J^C(x, y))) \rightarrow 0 \quad (4-8)$$

其中， $J^{dark}(x, y)$ 为暗通道图像， $C \in \{R, G, B\}$ 为彩色图像三通道中的一个通道， $\Omega(x, y)$ 表示以像素点 (x, y) 为中心的局部范围。 $J^C(x, y)$ 为像素点 (x, y) 处的亮度值。

暗通道先验去雾方法以大气散射模型^[69]为基础，通过与暗通道模型相结合来实现去雾的目的。其中，大气散射模型的表达式为：

$$I(x, y) = J(x, y)t(x, y) + A(1 - t(x, y)) \quad (4-9)$$

其中， $I(x, y)$ 为有雾图像， $t(x, y)$ 为透射率， A 为大气光照值。

暗通道先验去雾推导公式的过程如下：

(1) 将公式(4-9)代入到公式(4-8)并取最小值，可得：

$$\min_{\{x, y\} \in I} (\min_{C \in \{R, G, B\}} \frac{I^C(x, y)}{A^C}) = t(x, y) \min_{\{x, y\} \in I} (\min_{C \in \{R, G, B\}} \frac{I^C(x, y)}{A^C}) + (1 - t(x, y)) \quad (4-10)$$

(2) 由公式(4-8)可知无雾图像的暗通道值为 0，将其代入公式(4-10)可得：

$$t(x, y) = 1 - \min_{\{x, y\} \in I} (\min_{\{x, y\} \in I} (\min_{C \in \{R, G, B\}} \frac{I^C(x, y)}{A^C})) \quad (4-11)$$

(3) 最后，结合公式(4-9)和公式(4-11)，可得无雾图像 $J(x, y)$ 的表达式：

$$J(x, y) = \frac{I(x, y) - A}{\max[t(x, y), t_0]} + A \quad (4-12)$$

其中， t_0 是阈值下限，设置该下限是为了防止透射率为零。

本小节采用暗通道先验去雾算法，分别对薄雾和浓雾条件下的图像进行去雾处理。去雾效果呈现于图 4.3，其中图 4.3(a)展示的是薄雾图像经去雾处理后的效果，图 4.3(b)呈现的是浓雾图像去雾后的效果。



(a) 薄雾去雾效果图



(b) 浓雾去雾效果图

图 4.3 暗通道先验去雾效果图

4.2.3 AODNet 去雾算法

AODNet 去雾算法的网络模型由 5 个卷积模块和 3 个池化层组成，其模型结构如图 4.4 所示。它不需要分别估计传输矩阵和大气光值，而是借助卷积网络生成去雾图像。该网络主要由透射率值 k 估计模块与图像生成模块构成。其中， k 估计模块会先对透射率图进行估计，进而算出每个像素所包含的雾气深度以及相对雾度水平，最后凭借得到的透射率图精确地去除图像中的雾霾，而图像生成模块则负责生成最终的去雾图。

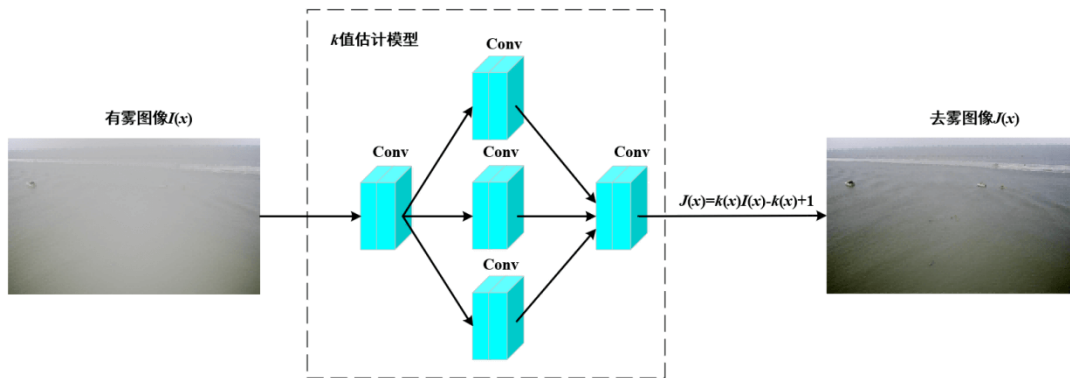


图 4.4 AODNet 去雾模型示意图

AODNet 去雾算法与传统去雾方法有所不同，传统去雾方法需要分别估计大气光值 A 与透射率 $t(x)$ ，而 AODNet 去雾算法则引入新的参数值 k ，并使用 k 作替换掉 A 与 $t(x)$ ，以此实现对大气光值与透射率的一次性估计。其中，参数值 k

的计算公式如下：

$$k(x) = \frac{\frac{I(x)-A}{t(x)} + (A-1)}{I(x)-1} \quad (4-13)$$

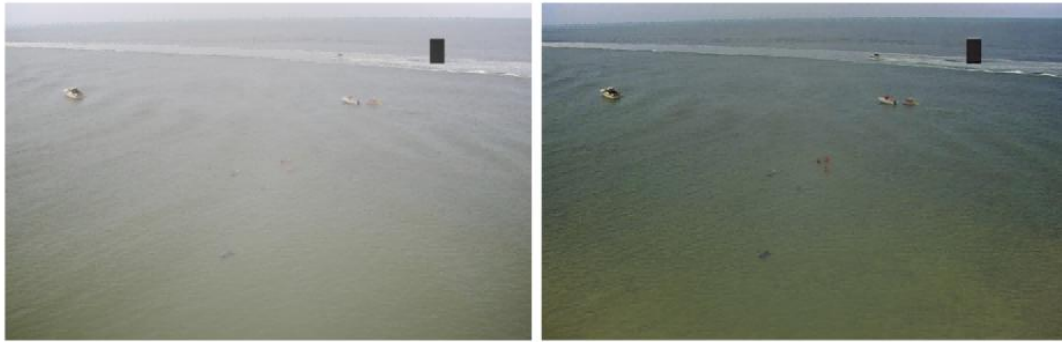
已知大气散射模型的表达式为：

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (4-14)$$

将公式(4-13)代入公式(4-14)，经过推导可得生成的无雾图像 $J(x)$ 为：

$$J(x) = k(x)I(x) - k(x) + 1 \quad (4-15)$$

基于上述理论基础，本小节使用 AODNet 去雾算法分别对薄雾和浓雾条件下的图像进行去雾处理。去雾效果呈现于图 4.5，其中图 4.5(a)展示的是薄雾图像经去雾处理后的效果，图 4.5(b)呈现的是浓雾图像去雾后的效果。



(a) 薄雾去雾效果图



(b) 浓雾去雾效果图

图 4.5 AODNet 去雾效果图

在海上无人机进行图像采集时，海雾等复杂环境因素严重影响成像质量。为了解决这一难题，本节在自制的海雾模拟数据集上开展去雾实验，并对比分析了这三种去雾算法的去雾性能，分别从图片平均耗时、去雾后图像的还原度以及失真度等方面评判算法性能的好坏。

为了评估去雾算法的效率，进行了三组实验。实验中使用上述三种算法，对原始尺寸均为 1920×1080 的数据集进行去雾处理。在处理过程中，记录处理整个数据集所耗费的总时长，再通过将总时长除以数据集中图像的数量，从而计算出每种算法的平均处理时间。

表 4.1 三种去雾算法平均耗时

去雾算法	图片数量	图片尺寸	平均耗时
单尺度 Retinex 去雾算法	2804	1920×1080	0.70s
暗通道去雾算法	2804	1920×1080	1.51s
AODNet 去雾算法	2804	1920×1080	0.23s

从表 4.1 看出，AODNet 去雾算法去雾速度最快，暗通道去雾算法去雾速度最慢，单尺度 Retinex 去雾算法的处理速度处于中间水平。

在浓雾条件下，三种算法的去雾效果如图 4.6 所示。通过去雾效果图可以看出，单尺度 Retinex 去雾算法去雾后雾气有所减弱，但是图像出现了光晕。暗通道去雾算法处理后，雾气去除得相对更明显，然而画面整体亮度偏暗，而且细节保存不太完整。AODNet 去雾算法相较于前两种去雾算法，雾气被有效清除，画面细节呈现得更丰富，例如海面上一些细微的物体和纹理都更清晰可见。同时，该算法的色彩还原度也较高，整体视觉效果更加自然，去雾效果表现较为出色。此外，该算法去雾速度较快，适合用于无人机雾天目标检测场景。

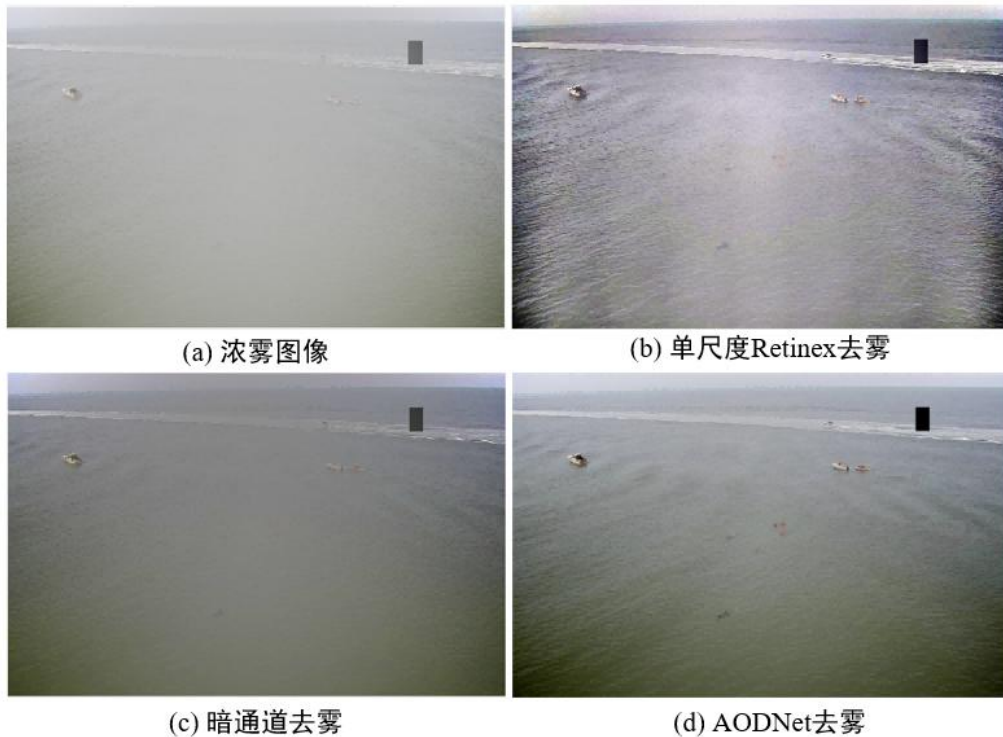


图 4.6 去雾算法效果对比图

因此,综合前文对不同去雾算法在速度、去雾效果等方面的分析,本章选用 AODNet 去雾算法作为 YOLOv7-tiny 目标检测的图像预处理算法。

4.3 YOLOv7-tiny 检测模型的改进

在雾天环境中,图像的能见度会显著下降,雾气会导致图像中目标区域的特征变得模糊不清。此外,在无人机在雾天航拍时,因视角的动态变化,极有可能出现多个目标相互遮挡的情形。这些干扰因素会对 YOLOv7-tiny 模型的检测精度造成严重影响。针对上述问题,本节基于 YOLOv7-tiny 模型从特征增强、检测头优化和损失函数优化这三个关键方面进行改进,以提升模型在雾天复杂环境下的检测性能。

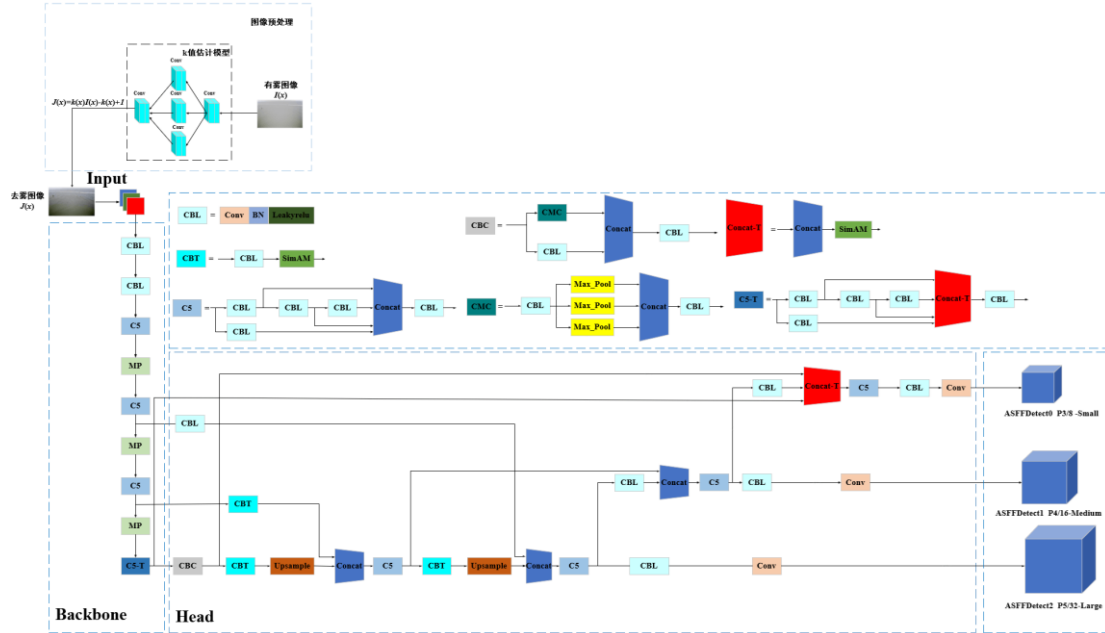


图 4.7 改进的 YOLOv7-tiny 网络结构图

首先,在模型前端将 AODNet 去雾算法应用于图像预处理阶段。AODNet 能够有效去除图像中的雾气,增强图像的对比度和清晰度,进而为后续的特征提取提供更优质的图像数据,保障特征提取的准确性与完整性。在特征增强方面,通过融入 SimAM 注意力机制,能够让模型更加聚焦于图像中的关键区域,抑制海洋背景带来的干扰,凸显目标的关键特征,从而有助于模型更准确地识别和定位目标。为了提升网络的多尺度特征融合能力,引入自适应特征融合 ASFF 模块的检测头替换掉原先模型的检测头,从而有效融合不同尺度目标的特征,让模型对不同大小的目标均能实现较好的检测效果。在损失函数优化方面,继续使用

NWD 损失函数对原始模型的损失函数进行优化，相关内容在本章后续小节不再介绍。改进后的损失函数能够更好地度量目标间的关系，提升模型对小目标以及边界模糊目标的检测精度。经过上述改进，并结合图像预处理阶段，改进后的 YOLOv7-tiny 模型能够满足雾天环境下海上目标的检测需求。改进后的网络结构如上图 4.7 所示。

4.3.1 引入 SimAM 注意力机制

SimAM 注意力机制与传统注意力机制不同的是，它是一种无参数注意力机制。在计算过程中无需额外参数，不会增加模型的参数量。SimAM 根据图像自身特征自适应地生成 3D 注意力权重，在面对不同类型的图像数据以及各类任务时，它能表现出较好的适应性，更灵活地聚焦于目标关键区域，有效抑制背景干扰。SimAM 的结构如图 4.8 所示。

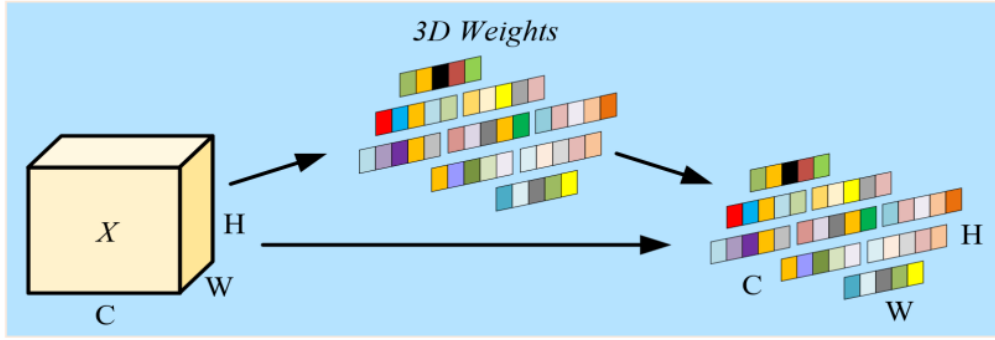


图 4.8 SimAM 注意力机制结构图

SimAM 的设计灵感源自神经科学理论，它通过为每个神经元定义一个能量函数，实现对神经元的优化。最小能量函数的公式为：

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (4-16)$$

其中， e_t^* 表示目标神经元的能量大小， $\hat{\mu}$ 表示所有神经元的平均值， λ 表示正则化系数， t 表示目标神经元， $\hat{\sigma}^2$ 表示所有神经元的方差。根据公式(4-16)能够计算出神经元的能量值，再调整特征图，最终找出更为关键的特征信息。通过 Sigmoid 激活函数对能量矩阵 E 进行抑制，得到注意力权重，从而提升模型的注意力机制的效果。经过注意力加权后的特征图 $\tilde{X}$ ，其表达式如公式(4-17)所示。

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X \quad (4-17)$$

其中， X 表示输入特征图。

为了提升 YOLOv7-tiny 的特征提取能力，本章对 YOLOv7-tiny 网络结构进行改进。首先，在 Backbone 尾端位置 C5 模块的拼接层之后添加 SimAM；其次，在 Head 的多个卷积层之后融入 SimAM；最后，将第 28 层输出、第 37 层以及第 66 层的输出，经过一个卷积层处理后再融入 SimAM，以此让特征提取网络更好地提取目标特征。改进的位置如上图 4.7 中的 C5-T、CBT 以及 Concat_T 可见。

4.3.2 加入 ASFF 检测头

检测头作为目标检测模型的关键组成部分，直接影响目标的定位和分类精度。YOLOv7-tiny 自带的检测头 IDetect 对不同尺度目标特征的融合方式较为简单，难以充分发挥不同尺度特征的优势。在海上目标检测场景中，不同大小的目标在特征表达上差异显著，使原检测头无法自适应地对这些不同尺度的特征进行有效融合，进而导致对小目标和大目标的检测性能不佳。针对原模型无法充分进行特征融合的问题，本节将自适应特征融合(ASFF)模块引入到原模型的检测头网络中，以提升对不同尺度目标的检测性能。改进的位置如上图 4.7 的 Head 尾端所示。

ASFF 的结构如图 4.9 所示。在图中 Level1、Level2 和 Level3 分别代表不同层级的融合策略。Level1 通过高分辨率特征图，精准捕捉小目标的特征细节；Level2 使用中分辨率特征图，实现对中等尺度目标特征细节的有效捕捉；Level3 通过低分辨率特征图来关注大目标的整体特征。ASFF 还能自适应地调整每个特征层的权重，有效优化特征融合效果。

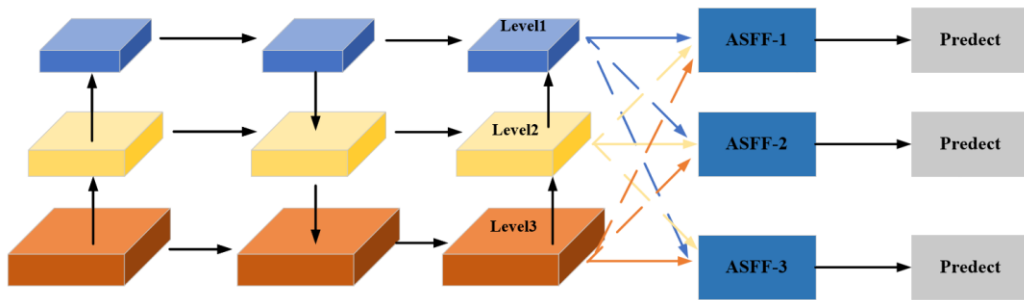


图 4.9 ASFF 结构图

本节将 ASFF 模块引入到去雾后的海上目标检测网络中，使其能够在不同层级对去雾后图像的特征进行空间滤波操作。在海上多雾环境中，雾气会严重干扰图像信息。即便经过去雾处理，图像中依然会存在因去雾算法所产生的噪声、残

留的模糊痕迹以及受雾气影响原本就复杂的背景纹理，致使特征信息依旧繁杂。而 ASFF 模块能够精准筛选这些特征，仅保留与船舶、浮标等海上目标相关的有用信息进行融合，有效甄别出对海上目标检测具有价值的特征，实现去雾后特征的高效整合。

4.4 雾天海上图像数据集制作

4.4.1 海上图像采集

本章实验的数据集选自 SeadronesSeeV2 数据集，从中选取了 2804 张图片。数据集中的检测类别包含多个种类，分别为船只、半身露出海面的人员、漂浮物、摩托艇、救生设备。在筛选图片时，重点关注拍摄高度、拍摄角度以及天气具有海洋背景。图 4.10 展示了该数据集的部分图片。



图 4.10 部分海上图片

4.4.2 海上图像标注

本章和上一章一样均使用 Labelimg 软件对图片中的目标进行标注，从而生成 YOLO 格式的标签文件。Labelimg 标注数据的过程如图 4.11 所示。



图 4.11 Labelimg 数据标注

4.4.3 图像加雾算法

海面水汽蒸发剧烈，当水汽压达到饱和状态时，水汽便会凝结成雾。海风冲击海水形成的微小海雾颗粒聚集也会有雾，这使得海洋容易出现大雾天气。然而，在雾天环境下操纵无人机收集海上图像极为困难，更新数据标签的成本高。考虑到生成对抗网络(GAN)需要依赖大量的海洋场景下真实雾天-无雾数据才能生成对应的雾天图像，且训练成本高昂。相比于 GAN 对雾天图像的依赖，采用人工加雾算法根据原始海洋图像的深度信息来生成海雾图像，为去雾网络的训练提供样本。本章采用基于大气散射模型和暗通道先验的海洋场景雾天图像模拟方法来生成雾景图。该方法图 4.12 为加雾原理模型图。

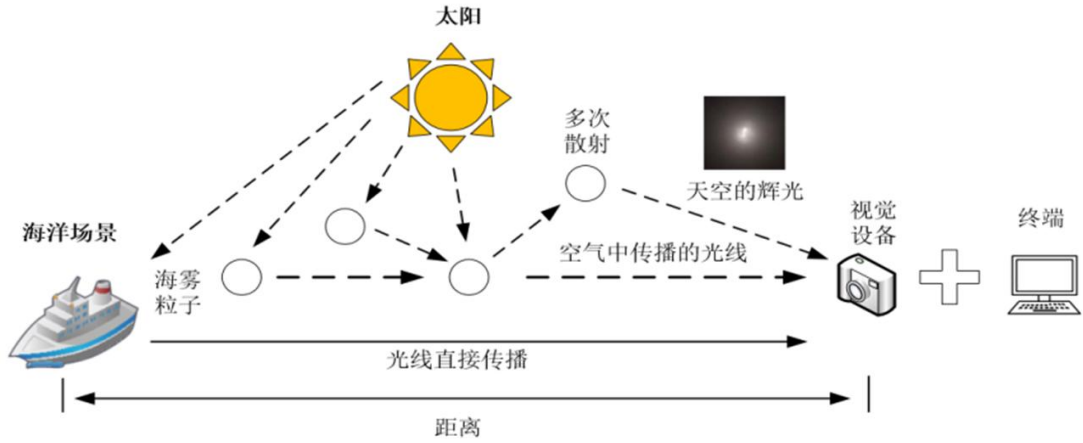


图 4.12 海洋场景加雾原理图

大气散射模型模拟了光线通过大气时的散射过程，其公式如下：

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (4-18)$$

其中， $I(x)$ 表示有雾图像， $J(x)$ 表示无雾图像， $t(x)$ 表示大气透射率， A 表示大气光值。因此，当 $t(x)$ 和 A 已知时，图像可从无雾空间映射到有雾空间。当大气均匀时，大气透射率 $t(x)$ 表示为：

$$t(x) = e^{-\beta d(x)} \quad (4-19)$$

$$d(x) = s - \frac{\rho(L, C)}{20} \quad (4-20)$$

在公式(4-19)中， x 为雾化中心， β 为大气散射系数， $d(x)$ 由公式(4-20)计算出的散射场景深度。在公式(4-20)中， ρ 是两个变量之间的欧氏距离， s 是雾化空间

的维数， L 是原始图像的高度和宽度信息， C 是雾化空间的中心点坐标。

在大多数无雾图像的非天空区域中，像素在三个 RGB 颜色通道中的至少一个具有暗像素。由这些暗像素组成的通道被称为暗通道，可定义为：

$$J^{dark}(x) = \min_{C \in \{r, g, b\}} (\min_{y \in \Omega(x)} (J^C(y))) \quad (4-21)$$

其中， J^C 是 J 的颜色通道， $\Omega(x)$ 是一个以 x 为中心的局部区域。假设 $J(x)$ 是一个无雾的户外图像，基于暗通道优先理论，除天空区域外， J^{dark} 的强度极低，趋于零。

因此，本小节利用暗通道先验理论对雾天图像的大气光值 A 进行估计。计算大气光值主要分为三个步骤：首先，计算暗通道；然后按照与图像像素数成一定比例，对最亮的暗像素进行采样；最后，将采样得到的暗像素对应到原始图像，该像素三个 RGB 通道的平均值为大气光值。

在计算暗通道时，首先将整个图像分成多个 7×7 的区域。通过将图像中的所有通道与所有区域一起扁平化处理，得到 $P \in R^{147}$ 的 N 个向量。由于采用类似于卷积操作的滑动窗口将图像分割成子区域， N 等于图像的长度乘以它的宽度。然后，在 $\Phi \in R^{147 \times N}$ 上进行矩阵最小化操作，如公式(4-22)所示：

$$J^{dark}(x) = \min_{y \in \Phi(x), d=0} (y) \quad (4-22)$$

在公式(4-22)中， $\Phi(x)$ 表示一个矩阵 $\Phi \in R^{147 \times N}$ ，它包含三个通道的所有区域像素， d 是进行矩阵最小化操作的维数。

当采样暗像素时，选取暗通道中最亮的暗像素。由于采集到的图像分辨率为 1920×1080 ，因此对暗像素采样数进行了调整。选择原始图像中像素数的 0.9% 作为采样数，即采样所有暗通道中最亮暗像素的前 0.9%，以此完成暗像素的采样操作。最后，选取与原始图像相同位置对应的采样暗像素，并将采样像素的最大值限制为 220(0.89)。大气光值 A 通过对该像素位置的 RGB 三个通道取平均值来确定的。

根据最终成像效果，选取大气散射系数 β 分别为 0.02 和 0.06 来模拟薄雾和浓雾天气。图 4.13 展示了加雾效果图。其中左侧为原始图像，中间为薄雾图像，右侧为浓雾图像。左侧原始图像清晰，细节完整。中间薄雾图像中部分细节虚化，

色调柔和,对比度稍降。右侧浓雾图像里,海洋场景被浓雾遮蔽,细节大量丢失,物体轮廓模糊,能见度低,色彩不饱和。对比这三张图,能直观看到不同大气散射系数对雾天模拟效果的显著差别。



图 4.13 加雾效果图

本节通过加雾算法制作了本章的雾天数据集,进而为后续在雾天环境下海上目标检测算法提供了样本支撑。

4.5 实验结果与分析

4.5.1 实验环境及配置

本章以 Pycharm 作为实验平台,采用 Python 编程语言,并借助深度学习开发框架 Pytorch,对改进前后的 YOLOv7-tiny 模型进行训练。整个实验所需的环境配置如表 4.2 所示。

表 4.2 实验环境配置

实验环境	配置
操作系统	Ubuntu20.04
开发框架	PyTorch1.10.0
编程语言	Python3.8
GPU	RTX3090(24GB)
CUDA	11.3

实验前需进行超参数的设定。其中,训练总轮数为 300,批大小为 16,优化器选用 SGD,初始学习率为 0.01,权重衰减系数设置为 0.0005。

4.5.2 目标检测评价指标

本节实验采用平均精度均值 $mAP(mAP@0.5)$ 、精确率 $P(Precision)$ 、召回率

R(Recall)以及 F1 分数(F1-Score)作为评价指标,以此来衡量雾天海上目标检测性能的优劣。

4.5.3 去雾前后检测对比实验

本章为验证 AODNet 去雾算法能否提升 YOLOv7-tiny 在雾天的目标检测效果,首先对浓雾数据集进行了去雾处理,从而得到清晰图像。然后使用 YOLOv7-tiny 分别对雾天数据集和清晰数据集进行检测实验,并使用上述评价指标评估检测效果。在实验过程中,确保数据集的划分方式保持一致,以此保证实验的严谨性。实验结果如表 4.3 所示。

表 4.3 去雾前后的实验结果

Type	Precision	Recall	F1	mAP@0.5
Fog	72.8%	68.7%	0.707	70.2%
Defog	78.1%	75.4%	0.767	74.6%

由表 4.3 的实验结果可知,经过 AODNet 图像预处理后,检测指标均有提升。这表明 AODNet 去雾算法有助于提升 YOLOv7-tiny 在雾天的目标检测效果。去雾前后的检测模型的 Precision-Recall 曲线图分别如图 4.14(a)和图 4.14(b)所示。通过对比两幅图中不同颜色曲线与横轴围成的面积,可以看出右图所围成的面积更大,说明经去雾预处理后检测效果有所改善。

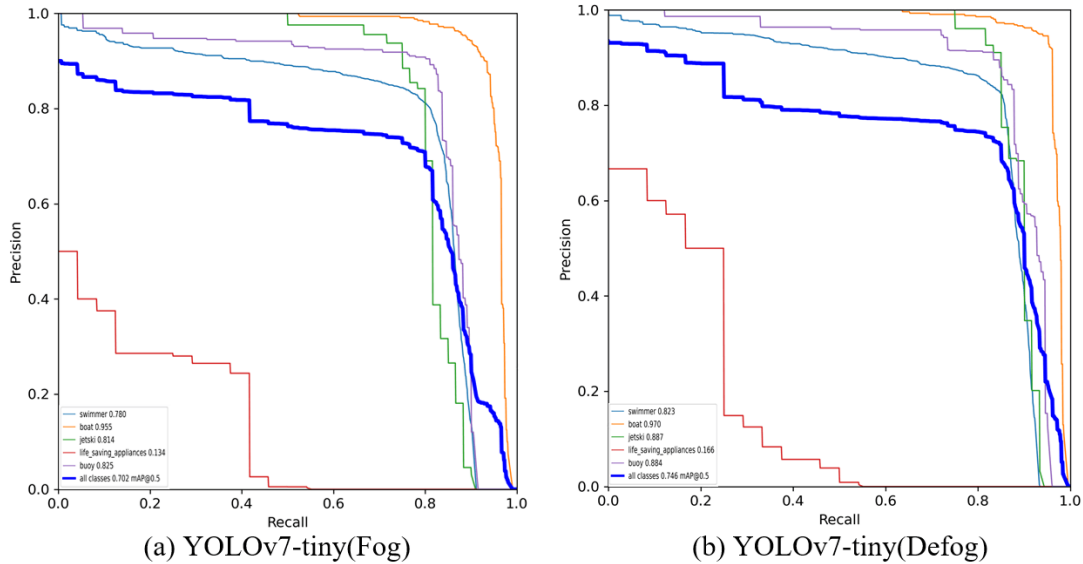


图 4.14 去雾前后 Precision-Recall 曲线对比图

为了更直观地展现去雾后检测效果的提升情况,本节开展了去雾前后可视化实验结果的对比分析。可视化结果如下图 4.15 所示。

图 4.15(a)展示的是 YOLOv7-tiny 在大雾天气下对海上目标的检测情况。原

本图像中有两个目标，然而该模型仅检测出船只，未能检测出漂浮物。图 4.15(b) 呈现的是经 AODNet 去雾处理后 YOLOv7-tiny 准确检测出两个目标的结果。对比这两张图可以发现，AODNet 去雾算法有助于提升 YOLOv7-tiny 对雾天环境下的目标检测效果。这种提升不仅体现在检测目标数量的增加上，从检测精度方面来看，去雾后的图像使模型能够更准确地识别目标边界，有效减少了因雾气干扰而导致的漏检情况，为无人机海上目标检测任务提供了保障。

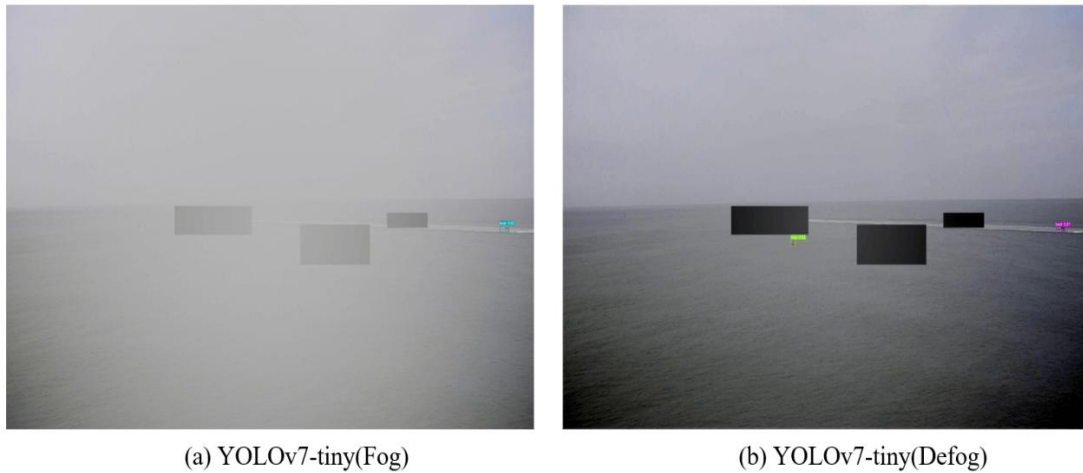


图 4.15 去雾前后可视化结果对比图

4.5.4 改进前后检测模型对比实验分析

虽然经 AODNet 去雾处理后，YOLOv7-tiny 对雾天环境下的检测效果有所提升，然而提升幅度并不显著。这是由 YOLOv7-tiny 网络结构自身的局限性所决定的。鉴于此，为了进一步增强检测模型的性能，本章从特征增强、检测头优化和损失函数优化这三个方面，对 YOLOv7-tiny 检测网络进行了改进。为了验证改进后模型的检测效果，分别对改进前后的 YOLOv7-tiny 模型进行训练，训练结束后选取最好的权重在测试集上展开检测实验，并将改进前与改进后的模型检测结果进行详细的对比分析。整个实验过程严格控制变量，实验环境、数据集以及模型超参数设置均保持一致。(实验结果均基于同一个浓雾数据集进行去雾图像预处理)改进前后的实验结果如下表 4.4 所示。

表 4.4 改进前后的实验结果

Model	Precision	Recall	F1	mAP@0.5	Weights	FPS
YOLOv7-tiny	78.1%	75.4%	0.767	74.6%	12.0MB	46.1
YOLOv7-tiny-Improved	83.3%	80.9%	0.821	81.5%	12.4MB	52.3

从表 4.4 可知，改进后模型的精确率提升了 5.2%，召回率提升了 5.5%，F1 分数提升了 0.054，mAP@0.5 提升了 6.9%，模型权重大小略微增加。FPS 提升

了 6.2, 说明检测速度有所提升。改进前后的 YOLOv7-tiny 模型的 Precision-Recall 曲线图分别如图 4.16(a)和图 4.16(b)所示。通过对比这两张图可以看出, 改进后的模型在 Precision-Recall 曲线所反映的性能指标上均优于原始模型, 进一步表明改进后的检测性能得到了显著优化, 能够更精确地检测目标。

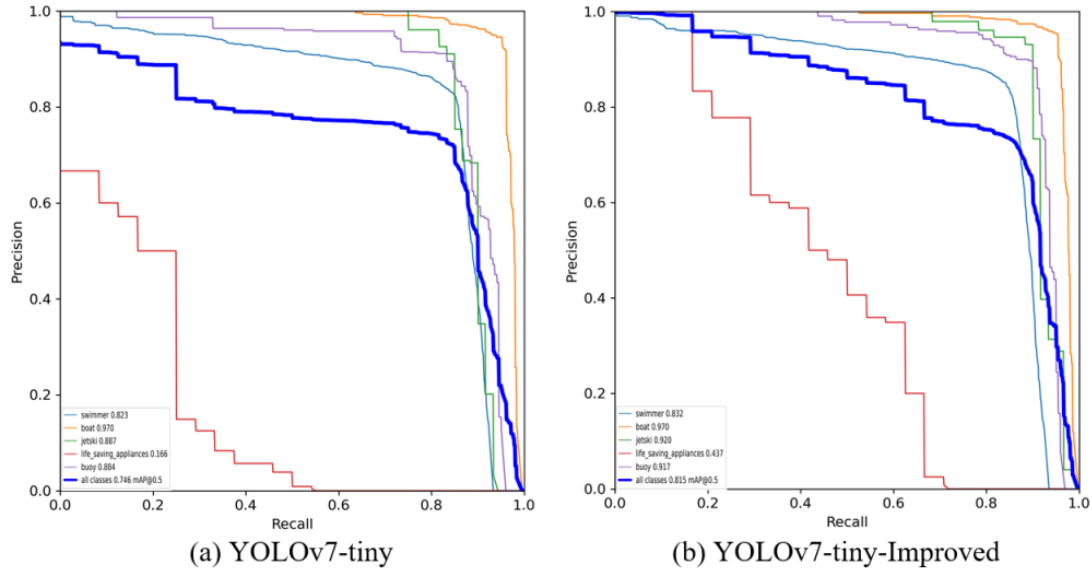


图 4.16 Precision-Recall 曲线对比图

4.5.5 消融实验分析

为了验证本章所改进的检测网络模型在海上目标检测中的有效性, 在确保实验环境、数据集以及模型超参数设置均保持一致的前提下, 将 SimAM 注意力机制、ASFF 检测头和 NWD 损失函数融入初始 YOLOv7-tiny 网络中, 并设计了四组消融实验, 以此评估所提出的改进算法在海上目标检测方面的性能。消融实验结果如表 4.5 所示。

表 4.5 消融实验结果

Group	SimAM	ASFF	NWD	Precision	Recall	F1	mAP@0.5
1				78.1%	75.4%	0.767	74.6%
2	√			79.7%	78.6%	0.791	78.7%
3	√	√		80.5%	80.8%	0.806	80.2%
4	√	√	√	83.3%	80.9%	0.821	81.5%

从表 4.5 中可以看出, 第 1 组为 YOLOv7-tiny 模型的实验结果, 作为基准实验。第 2 组实验在 YOLOv7-tiny 网络中添加了 SimAM 模块, 该模块显著提升了模型对目标特征的捕捉和识别能力。相较于基准实验, 各项指标均有很大的提升。第 3 组实验是在第 2 组实验的基础上在网络中引入 ASFF 检测头, mAP@0.5 值

提高了 1.5%，准确率略有提升，召回率提高了 2.2%。实验结果表明，ASFF 检测头能够更有效地整合多尺度特征信息，从而有效提升对不同目标的召回率。第 4 组实验是在第 3 组实验的基础上，使用 NWD 损失函数对原始 CIOU 损失函数进行优化。其中，将 CIOU 损失的比例值设定为 0.3 时，NWD 损失的比例值为 0.7 时，模型的检测效果最佳。mAP@0.5 值增加了 1.3%，准确率提高了 2.8%。准确率的提升体现了 NWD 损失函数针对复杂场景下小目标检测优化的有效性，使得模型能够更精准地识别出目标。而召回率几乎不变，这说明第 3 组实验中的模型已经较为充分地挖掘出图像中的不同尺度目标的特征信息了，最大程度地减少了漏检情况。改进后的 F1-Score 相较于基准有了很大的提升。通过对消融实验结果中的每个评价指标进行综合分析，验证了本章所提出的改进算法的有效性。

4.5.6 不同检测模型对比实验分析

为了进一步验证本文所改进的检测算法在海上目标检测方面的性能表现，将其与表 4.6 中目前主流的目标检测算法在同一个去雾后的数据集上(基于浓雾数据集进行去雾处理)进行了对比实验，评价指标包括 Precision、Recall、F1、mAP@0.5、模型权重(Weights)、FPS，实验结果如表 4.6 所示。

表 4.6 对比实验结果

Model	Precision	Recall	F1	mAP@0.5	Weights	FPS
YOLOv3-tiny	73.8%	71.2%	0.725	71.1%	33.9MB	32.5
YOLOv5-n	82.3%	80.8%	0.815	81.4%	3.8MB	29.2
SSD	75.5%	32.3%	0.452	74.7%	92.1MB	42.8
YOLOv7-tiny	78.1%	75.4%	0.767	74.6%	12.0MB	46.1
YOLOX-s	80.7%	80.1%	0.804	79.8%	34.7MB	44.3
YOLOOW	76.2%	73.8%	0.749	76.3%	71.1MB	58.7
YOLOv9-s	83.5%	81.2%	0.823	81.9%	14.5MB	60.1
YOLOv10-s	85.2%	82.7%	0.839	83.3%	15.7MB	62.5
YOLOv7-tiny-Improved	83.3%	80.9%	0.821	81.5%	12.4MB	52.3

从表 4.6 可以看出，本文算法在 Precision、Recall、F1、mAP@0.5、FPS 这几个指标上，相较于表 4.6 中前五种算法均有了一定的提升。基于这些指标的提升情况可以表明，本文算法在海上目标检测方面具有更好的检测性能，其优势显著。与大部分算法相比，改进后的算法模型体积更小，检测速度较快，更适合部署到资源有限的设备终端。这表明本文算法在提升检测性能的同时，并未过度牺牲检测速度，仍然能满足实时检测的需求，达到了检测性能与速度上的平衡，从而为下一章的基于检测器的海上目标跟踪研究奠定了良好基础。

为了更直观地展现改进后方法的检测效果，本文对改进后的 YOLOv7-tiny 与表 4.6 中的主流算法进行了可视化对比实验。从测试集中选取若干张具有代表性的图像进行检测，将置信度阈值设定为 0.5，可视化对比结果如图 4.17 所示。由可视化结果可知，相较于其他算法，改进后的检测方法能够更为精确、有效地检测出图像中的所有目标，无论是船只、被遮挡的目标，还是浮在海面上的目标，都能被检测出来。

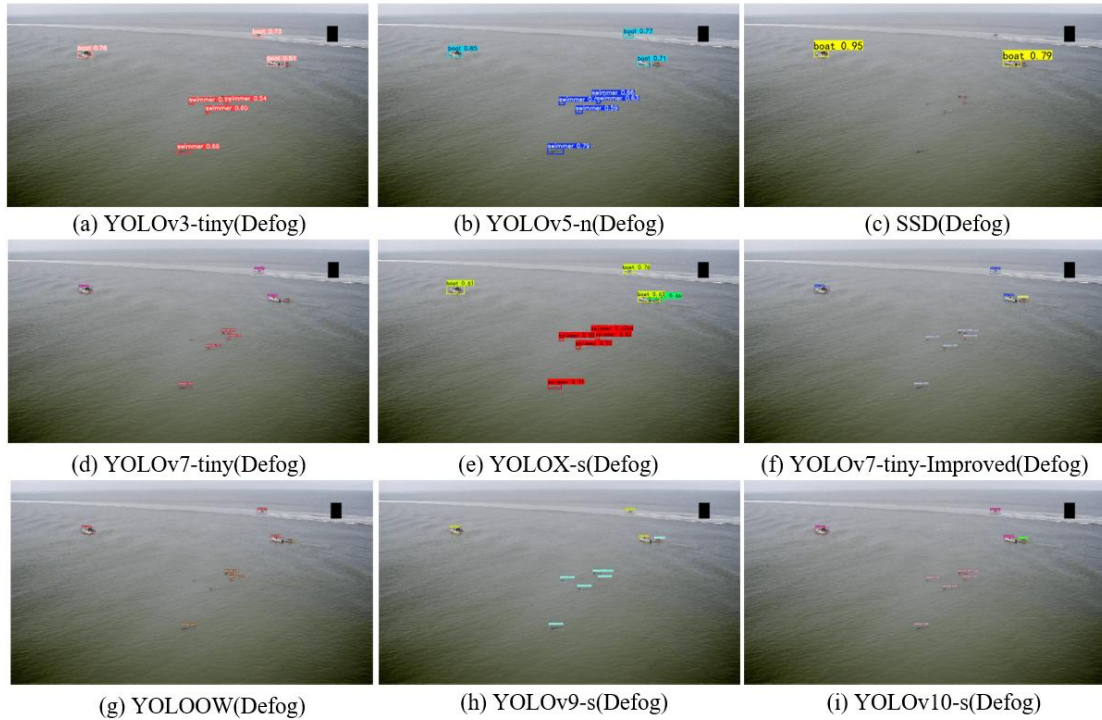


图 4.17 可视化结果对比图

为了更充分地展现基于 YOLOv7-tiny 的雾天目标检测优化方法的优越性，本节做了可视化检测结果对比分析，可视化对比结果如图 4.18 所示。对比图 4.18(a)，图 4.18(b)以及图 4.18(c)这三张图可以清晰地发现，改进后的检测算法有效解决了 YOLOv7-tiny 算法在大雾天气下出现的漏检问题，显著提升了自身的检测性能，能够更好地满足复杂海洋场景中海上目标检测的需求。



图 4.18 可视化结果对比图

4.6 本章小结

本章首先详细介绍了 YOLOv7-tiny 的网络模型结构。然后, 针对该模型在雾天环境下海上目标检测精度欠佳的问题, 设计了一种更适用于雾天环境的海上目标检测算法。在图像预处理阶段, 采用 AODNet 去雾算法对雾天图像进行去雾处理, 随后将去雾后的图片输入到改进后的检测网络。为了让检测网络能更聚焦于图像中的目标, 在原有网络中嵌入 SimAM 模块。通过使用 ASFF 检测头替代原有网络的检测头, 增强网络的多尺度特征融合能力, 进而提升对不同大小目标的检测效果。考虑到本章自建雾天数据集和前一章的数据集都是源于 SeaDronesSeeV2, 其中存在着形状相似的小目标, 因此沿用 NWD 损失函数对 CIOU 损失函数进行优化, 以便更精准地度量小目标物体。通过在自制的雾天海上目标检测数据集上, 对改进后的模型进行消融实验以及与主流模型进行对比实验, 综合验证改进算法的有效性。实验结果表明, 改进后的检测模型能够显著提升雾天环境下的海上目标检测精度。

第 5 章 基于检测器的海上多目标跟踪算法

前两章针对不同天气状况，分别设计了两种海上目标检测算法。其中，第三章的算法侧重于高精度检测，而第四章的算法能够兼顾精度和速度，为海上多目标跟踪奠定了坚实基础。相较于 YOLOv7，YOLOv7-tiny 检测速度更快，因此本章选用第四章改进的检测网络作为检测器，并以 DeepSORT 目标跟踪算法为基础进行研究。在海上多目标跟踪过程中，常面临目标消失、新目标出现以及遮挡等问题。考虑到 DeepSORT 目标跟踪算法中的卡尔曼滤波器在预测目标运动状态时容易受到低置信度检测框的干扰，且在处理不同尺度噪声方面存在不足。因此，本章采用一种自适应抑制噪声的卡尔曼滤波器，使其更稳定地适应目标的动态变化。同时，通过使用 DIOU 替代原先的 IOU 来优化匹配机制，有效解决因目标尺度变化导致跟踪效果不佳的问题。通过与原跟踪算法进行对比实验，结果表明改进后的 Deepsort 算法具有更稳定的目标跟踪能力。

5.1 DeepSORT 目标跟踪算法

DeepSORT^[70]算法对 SORT 算法进行了改进。通过引入图像分类网络来提取目标的外观信息，再将这些信息应用于代价矩阵的计算过程。在遮挡场景下，当目标再次出现时，DeepSORT 算法借助外观特征匹配方法，实现对目标的持续跟踪，一定程度解决了因遮挡导致的 ID 切换问题。DeepSORT 算法还提出了级联匹配机制，优化了不确定轨迹与目标的匹配顺序，让不确定轨迹能更顺利地为目标重新匹配，快速恢复跟踪状态。DeepSORT 算法的具体步骤如下：

（1）开始时，依据第一帧检测到的图像构建对应的轨迹，然后运用卡尔曼滤波对所构建的轨迹进行初始化。初始化完成后，借助卡尔曼滤波预测下一帧图像中的检测框。此时，所有已构建的轨迹均处于不确定状态。

（2）使用构建好的轨迹对当前帧和前一帧的检测框进行匹配操作，然后计算代价矩阵。

（3）通过匈牙利算法对目标进行线性匹配，会得到三种不同的匹配结果：一是部分轨迹未能与检测框成功匹配；二是部分检测框未能与轨迹完成匹配；三是匹配成功，即轨迹和检测框成功对应。

(4) 对于匹配成功的轨迹，采用卡尔曼滤波进行更新，以获取其运动相关的变量，同时将该轨迹的状态转变为确认态。而对于失配的轨迹，则需根据自身状态进行不同处理。若轨迹处于确认态，只有当连续失配达到 30 次以上时，才将其删除；若轨迹原本处于不确定态，则直接将其删除。

(5) 循环执行第二步至第四步的过程，直到满足以下两个条件之一：一是视频的所有帧都已处理完毕；二是出现了处于确认态的轨迹。

(6) 运用卡尔曼滤波方法，分别针对与确定态以及不确定态轨迹相对应的检测框进行预测。其中，确定态轨迹会与检测框进行匹配操作。级联匹配通过将此前保存的外观特征以及动作信息与检测框进行匹配，以此提升整体的匹配成功率。

(7) 根据级联匹配得出的结果，利用卡尔曼滤波更新匹配成功的轨迹状态。对于在级联匹配中失配的轨迹，使用 IOU 和匈牙利算法进行匹配和计算，得到新的匹配结果。

(8) 重复第六步和第七步操作，直到视频结束。

DeepSORT 算法的流程图如图 5.1 所示。

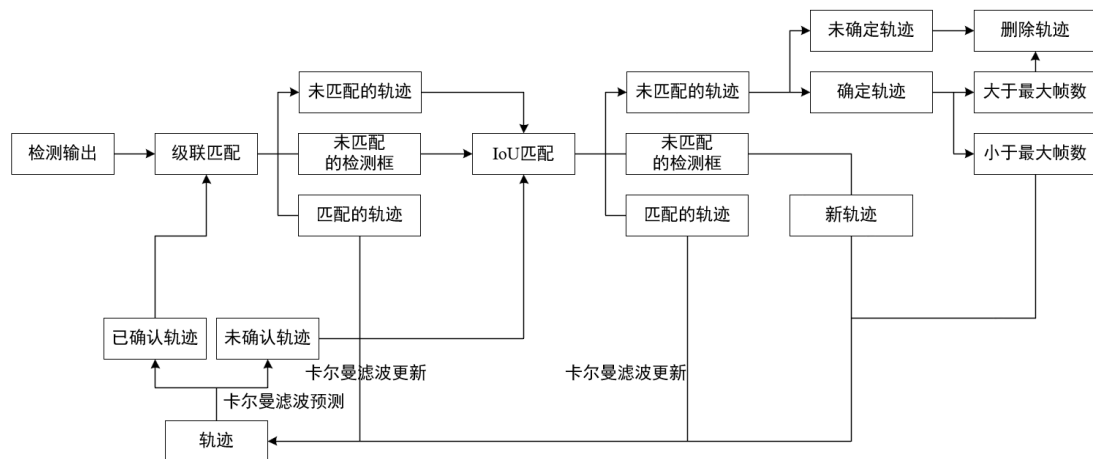


图 5.1 DeepSORT 算法流程图

5.2 DeepSORT 目标跟踪算法的优化

5.2.1 自适应噪声抑制卡尔曼滤波器

在 DeepSORT 算法中使用的是具有线性关系的卡尔曼滤波器，它将一个恒定不变的常数矩阵 R 作为测量噪声协方差矩阵，度量视频当前检测帧中的噪声尺度。通过采用这种度量方式，会使所有待检测目标受到的测量噪声影响程度相同。然

而，在海上检测与跟踪任务中，由于目标特征存在差异以及环境因素的干扰，不同大小的目标检测框呈现出不同的噪声尺度。

本节为了消除多尺度噪声对目标跟踪效果的干扰，采用自适应噪声抑制卡尔曼滤波器(Adaptive Noise Suppressing Kalman Filter, ANSKF)。通过在原先的测量噪声协方差矩阵基础上加入了自适应计算噪声协方差的公式，能够根据目标检测框的质量差异，动态调整噪声尺度。通过这种方法，它在复杂多变的实际场景中能够更精确地估计当前目标的状态变量，进而提升目标跟踪的性能。

自适应噪声抑制卡尔曼滤波器的具体计算过程如下：

(1) 计算预拟合残差值，如公式(5-1)所示：

$$\tilde{y}_h = z_h - H_h \hat{x}_{h|h-1} \quad (5-1)$$

其中， z_h 为当前测量值， H_h 为观测模型矩阵， $\hat{x}_{h|h-1}$ 为上一时刻的状态预测值。

(2) 动态调整测量噪声协方差，然后通过公式(5-2)对测量噪声协方差进行调节：

$$\tilde{R}_h = (1 - c_h) R_h \quad (5-2)$$

其中， R_h 表示预设好的常数测量噪声协方差， c_h 表示 h 时刻的检测置信度，即动态调整系数。

(3) 结合预测协方差与调整后的测量噪声协方差来计算残差协方差矩阵，如公式(5-3)所示：

$$S_h = H_h P_{h|h-1} H_h^T + \tilde{R}_h \quad (5-3)$$

其中， $P_{h|h-1}$ 为预测状态协方差，反映状态估计的不确定性。

(4) 利用残差协方差矩阵的逆矩阵，计算更新后的最优卡尔曼增益，如公式(5-4)所示：

$$K_h = P_{h|h-1} H_h^T S_h^{-1} \quad (5-4)$$

其中，增益矩阵 K_h 决定测量残差对状态更新的权重。

(5) 通过公式(5-5)更新状态估计值，得到更精确的当前状态估计：

$$\hat{x}_{h|h} = \hat{x}_{h|h-1} + K_h \tilde{y}_h \quad (5-5)$$

(6) 通过公式(5-6)更新估计协方差矩阵:

$$P_{h|h} = (I - K_h H_h) P_{h|h-1} \quad (5-6)$$

其中, I 表示单位矩阵。

5.2.2 DIOU 匹配机制

在 DeepSORT 算法的匹配过程中, 依据检测框与跟踪轨迹之间的 IOU 值构成本矩阵, 再使用匈牙利算法完成最优匹配, 按照 IOU 阈值筛选出有效的关联方案, 实现对目标的持续跟踪。然而, 在实际场景中 IOU 存在一定的局限性。例如, 当目标在镜头中的尺寸发生较大变化时, IOU 值可能由于框体比例失衡而上下波动, 导致跟踪效果变差。当检测框与预测框之间无任何重叠区域时, IOU 值会变为零, 这可能造成目标丢失。针对上述问题, 本节采用 DIOU 替代原有的 IOU, 优化检测框与预测框之间的匹配程度。DIOU 在保留 IOU 对重叠区域敏感性的基础上引入中心点距离惩罚项, 有效避免因框体无重叠而导致匹配完全失效的问题。此外, 使用 DIOU 能降低因目标尺度变化而导致错误匹配的概率。DIOU 损失函数表示为:

$$DIOU = IOU - \frac{\rho^2(b, b^g)}{c^2} \quad (5-7)$$

$$L_{DIOU} = 1 - IOU + \frac{\rho^2(b, b^g)}{c^2} \quad (5-8)$$

其中, 公式(5-7)中 $\rho^2(b, b^g)$ 表示目标框与预测框的中心点之间的欧氏距离。公式(5-8)中 L_{DIOU} 表示 DIOU 损失函数, c 表示覆盖预测框和目标框的最小外接矩形的对角线长度。

5.3 基于检测器的改进跟踪模型结构

本章对 DeepSORT 跟踪算法的改进点进行整合, 把第四章中改进后的检测网络用作海上目标位置信息提取的检测器。图 5.2 展示的是基于该检测器的改进 DeepSORT 算法流程图。在跟踪模型的运行过程中, 视频序列作为导入后的输入数据。利用目标检测算法处理视频中的目标, 获取对应的边界框与尺寸信息。同时, 运用 ResNet 网络提取目标的外观特征, 并存储这些特征信息。将上述信息与卡尔曼滤波预测得到的预测框一起进行级联匹配。完成匹配后, 对匹配结果进

行判定。若匹配成功，便采用自适应噪声抑制卡尔曼滤波更新轨迹信息；若匹配失败，则继续执行 DIOU 匹配操作。DIOU 匹配结束后会产生三种判定结果，即保存轨迹、更新轨迹和删除轨迹。整个流程会持续循环，直至出现轨迹匹配的情况或者视频序列全部被遍历，最终输出跟踪轨迹的结果。

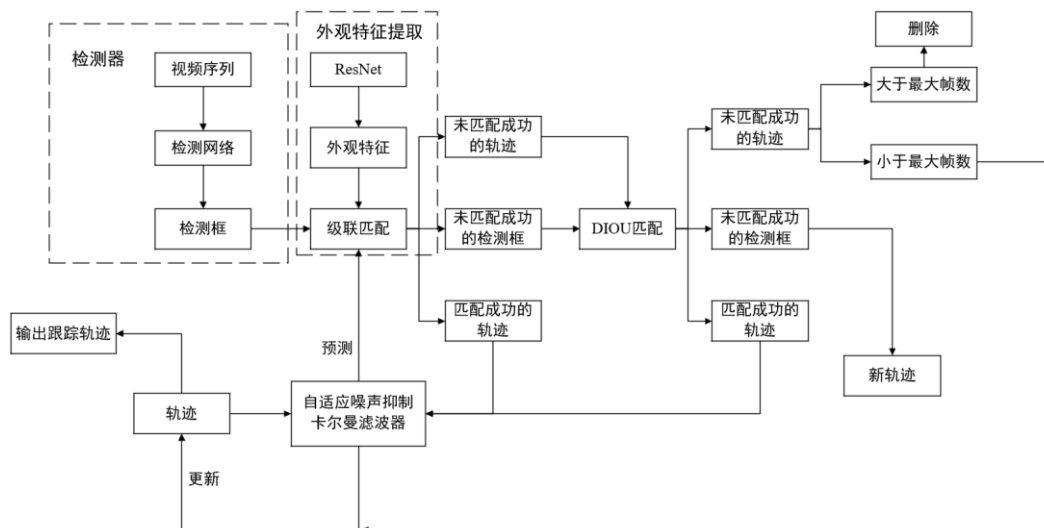


图 5.2 基于检测器的改进 DeepSORT 算法流程图

5.4 实验结果与分析

5.4.1 实验环境及配置

本节实验采用了与前两章海上目标检测实验相同的硬件配置与软件环境。整个实验所需的环境配置如表 5.1 所示。

表 5.1 实验环境配置

实验环境	配置
操作系统	Ubuntu20.04
开发框架	PyTorch1.10.0
编程语言	Python3.8
GPU	RTX3090(24GB)
CUDA	11.3

5.4.2 跟踪数据集制作

本节的数据集来源于 SeaDronesSee MOT 挑战赛，从中挑选了 6 个海上目标运动视频用来制作测试视频数据集。使用 Darklabel 标注软件对视频的每一帧进行标注，Darklabel 是一款较为好用的视频标注软件，支持目标检测和目标跟踪中各种标注文件格式。标注过程如图 5.3 所示，不同颜色的标注框代表着不同的目标 ID 编号。



图 5.3 Darklabel 标注过程图

5.4.3 多目标跟踪评价指标

多目标跟踪准确度(MOTA)、多目标跟踪精确度(MOTP)以及 ID 切换总数(IDs)是衡量多目标跟踪模型性能的常用指标。因此,本节将运用这些指标对所提出的改进跟踪算法的性能进行评估。

MOTA 表示多目标跟踪准确率,数值越大,说明其跟踪效果越好。MOTA 可以表示为:

$$MOTA = 1 - \frac{\sum(FN + FP + IDs)}{\sum GT} \quad (5-9)$$

其中, FN 表示漏检次数, FP 表示误检次数, GT 表示实际目标的数量。

MOTP 表示多目标跟踪精度,其数值越大,说明其跟踪器的定位效果越好。MOTP 可以表示为:

$$MOTP = \frac{\sum_{t,j} d_{t,j}}{\sum_t c_t} \quad (5-10)$$

其中, $d_{t,j}$ 表示第 t 帧中的第 j 个目标的预测框与真实边界框之间的距离, c_t 表示第 t 帧中成功匹配的目标个数。

IDs 表示该目标 ID 变化的总次数,数值越小,说明跟踪越稳定。

5.4.4 实验结果分析

本节实验采用多目标跟踪准确率(MOTA)、多目标跟踪精度(MOTP)和 ID 切换次数(IDs)作为评价指标,对改进后的跟踪模型进行评估。改进前后的对比实

验结果如表 5.2 所示。

表 5.2 对比实验结果

跟踪模型	MOTA	MOTP	IDs
DeepSORT	68.2%	78.3%	753
SORT	62.7%	75.1%	879
ByteTrack	70.3%	79.6%	685
DeepSORT-Improved	73.5%	80.1%	597

从表 5.2 可知,相较于原先的 DeepSORT 跟踪算法,改进后的算法在准确率上提升了 5.3%,多目标跟踪精度提升了 1.8%,ID 切换次数下降了 156 次。相较于 SORT 跟踪算法,在准确率上提升了 10.8%,多目标跟踪精度提升了 5%,ID 切换次数下降了 282 次。相较于 ByteTrack 跟踪算法,准确率提升了 3.2%,多目标跟踪精度提升了 0.5%,ID 切换次数下降了 88 次。实验结果表明,改进后的跟踪算法有效地提升了多目标跟踪的性能。

为验证所加入的自适应噪声抑制卡尔曼滤波器和改进 IOU 匹配的有效性,进行了消融实验,评估不同模块在实验条件下对跟踪性能的影响。消融实验结果如表 5.3 所示。

表 5.3 消融实验结果

跟踪模型	MOTA	MOTP	IDs
DeepSORT	68.2%	78.3%	753
DeepSORT+ANSKF	70.8%	79.4%	621
DeepSORT+DIOU	71.5%	78.9%	658
DeepSORT-Improved	73.5%	80.1%	597

据表 5.3 可知,相较于原始的 DeepSORT 跟踪算法,采用自适应噪声抑制卡尔曼滤波器后,跟踪准确率上提升了 2.6%,达到了 70.8%,跟踪精度提升了 1.1%,达到了 79.4%,同时 ID 切换次数减少了 132 次。采用改进后的 DIOU 后,跟踪准确率上提升了 3.3%,达到了 71.5%,跟踪精度提升了 0.6%,达到了 78.9%,同时 ID 切换次数减少了 95 次。实验结果表明,经过改进的跟踪算法能够有效地优化了多目标跟踪效果。

为了更直观地展现改进后跟踪算法的有效性,对改进前后的算法在一段测试视频中进行了可视化对比,结果如图 5.4 所示。从图中可以看出,在改进前的第 253 帧中最右边的船只在运动过程中目标 ID 出现了丢失情况;在改进前的第 385 帧中右上角的船只在运动过程中也出现了目标 ID 丢失的情况。而改进后的算法能够对视频中的所有目标进行持续跟踪,且全程维持所有目标的 ID 编号,没有

发生 ID 切换情况。对比改进前后的可视化结果，可以充分表明本章所提出的改进后的跟踪算法在跟踪性能的稳定性方面表现更为出色。

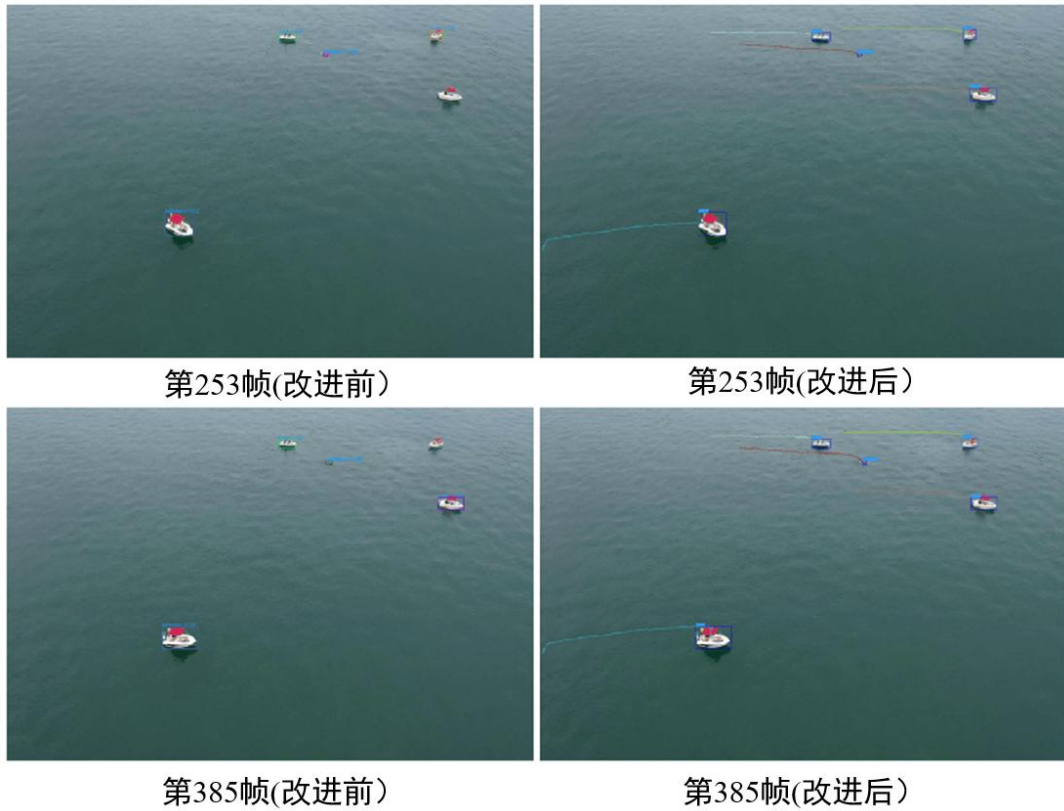


图 5.4 改进前后的跟踪效果图

5.5 海上目标检测与跟踪系统

5.5.1 系统设计

本章主要通过 PyQt5 完成海上目标检测与跟踪系统的界面设计，并使用前面章节提出的改进的 YOLOv7-tiny 检测网络进行识别，采用改进后的 DeepSORT 算法进行多目标跟踪。最后，将检测结果和跟踪过程中的预测轨迹关联起来，把跟踪视频结果在界面上显示出来。系统的整体设计框架图如图 5.5 所示。



图 5.5 系统整体设计框架图

从图 5.5 可以清楚地看出，海上目标检测与跟踪系统由视频处理、检测与跟踪以及界面显示三个大模块组成。系统首先通过无人机摄像头或本地终端获取原始数据，经视频编码模块解码后送入预处理模块进行分辨率统一，以优化输入质量。然后采用改进的 YOLOv7-tiny 模型进行目标检测，检测结果随后通过运用 ResNet 网络提取目标的外观特征，为跟踪提供外观信息。再采用改进后的 DeepSORT 算法完成跟踪任务。最终，视频中的检测与跟踪结果显示在 PyQt5 构建的可视化界面上。支持用户通过“开始/暂停/结束”等控件交互操作，同时提供模型类型切换及检测记录保存功能。

5.5.2 系统软硬件平台

系统的硬件配置为：处理器采用 Intel Core i7-13650HX CPU，显卡为 NVIDIA GeForce RTX4060 显卡，内存大小为 16GB，操作系统为 Windows11。软件开发方面，使用 Python 语言，并借助 opencv4.6 和 PyQt5 等库来实现相关功能。

5.5.3 系统界面

海上目标检测与跟踪系统界面如图 5.6 所示。

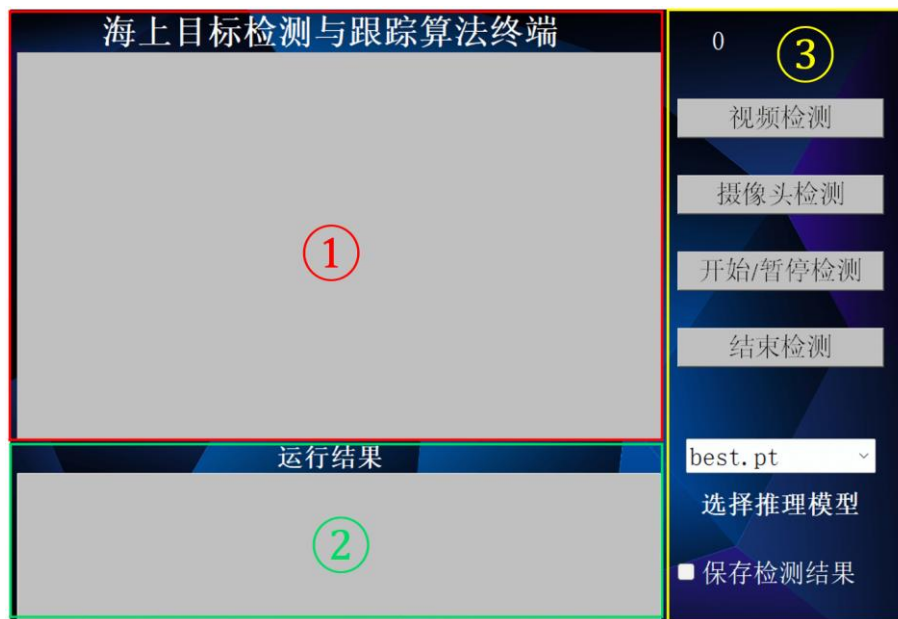


图 5.6 系统界面图

从图 5.6 中可以看出系统界面由红色、绿色和黄色三个区域构成。红色区域用于显示检测与跟踪可视化结果，包括目标检测框、运动轨迹及视频画面。绿色区域用于展示检测结果的具体信息，包含目标类型和数量。黄色区域用于用户交互控制，集成了“开始/暂停检测结束”“结束检测”等操作控件，支持切换检

测推理模型，并提供检测与跟踪记录的保存功能。

5.5.4 系统测试

通过将改进后的 YOLOv7-tiny 目标检测算法与改进后的 DeepSORT 算法集成在一起，并在基于 PyQt5 设计的可视化界面中对离线视频中的目标进行检测与跟踪，来测试系统的整体性能。

首先，点击“视频检测”控件选取特定的离线视频，如图 5.7 所示。



图 5.7 离线视频选择界面

然后跳出离线视频选取的界面，选择需要的离线视频，点击“打开”控件，如图 5.8 所示。

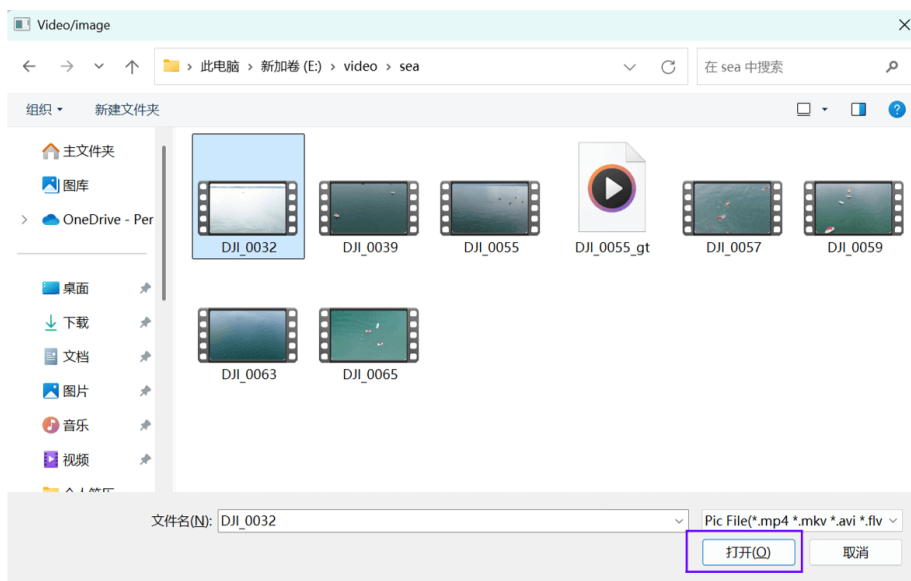


图 5.8 离线视频输入界面

最后，选择推理模型并点击“开始检测”控件。此时，系统会对该视频中的目标进行检测与跟踪，如图 5.9 所示。

图 5.9 展示的是测试后的可视化界面，该系统界面不仅能够实时显示检测结果，还能呈现画面中目标的检测跟踪效果。其中，FPS 代表当前的检测速度。本章所采用的算法的 FPS 在本地电脑上运行始终保持在 30 以上，表明系统能够满足实时检测的需求。经过对系统进行完整测试，系统的整体性能表现优异。



图 5.9 系统测试效果图

5.6 本章小结

本章将改进后的 YOLOv7-tiny 算法作为 DeepSORT 跟踪算法的检测器，并在 DeepSORT 的基础上进行优化，具体改进的地方包括引入自适应噪声抑制卡尔曼滤波以及采用 DIOU 匹配机制。随后，在自制的多目标跟踪数据集上进行对比实验，实验结果表明，本章改进的跟踪算法的跟踪效果优于改进前的算法。最后，将改进后的算法与人机交互界面相结合，构建了海上目标检测与跟踪系统，并选取离线视频对该系统的整体性能进行测试，结果显示其性能表现良好。

第 6 章 总结和展望

6.1 总结

在海上救援场景中，将智能无人机与目标检测跟踪技术相结合能够显著提升救援速度和救援成功率。然而，特殊的海洋环境对目标检测跟踪造成了限制，使得目标检测跟踪效果不佳。为了进一步提升目标检测与跟踪的效果，本文对 YOLOv7 系列中两种目标检测算法以及 DeepSORT 跟踪算法进行了改进，有效提升了海洋搜救场景下的目标检测跟踪效果。本文的主要研究内容如下：

(1) 针对晴天海面因光线强导致多目标检测精度欠佳的问题，本文设计了一种基于 YOLOv7 的晴天海上目标检测优化方法。通过在网络中引入坐标卷积模块并融入 CBAM 注意力机制，有效增强了原网络的空间感知能力与特征提取能力。采用 NWD 损失函数对原有的 CIOU 损失函数进行优化，有效提升了多个目标相互遮挡场景下的检测效果。最后，通过在自制的晴天数据集上与多个主流模型进行了对比实验，验证了改进后的算法具备高精度检测的效果。

(2) 针对雾天图像画质下降致使多目标检测精度欠佳的问题，本文设计了一种基于 YOLOv7-tiny 的雾天海上目标检测优化方法。首先，为选出一种最优的去雾算法作为图像预处理方法，本文将目前常用的三种去雾算法在去雾效果和效率上进行了对比，最终选用 AODNet 进行去雾处理。为验证去雾后的检测效果，在自制的浓雾数据集上进行了对比实验，实验结果表明，这种去雾预处理方法能提高目标检测精度。在检测网络方面，通过引入 SimAM 注意力机制，进一步突出目标的关键特征，有助于提升网络的特征提取能力。在原网络中引入具备自适应特征融合功能的 ASFF 模块的检测头，并替换掉原模型的检测头，有效融合不同尺度目标的特征。在损失函数优化上，继续使用 NWD 损失函数优化 CIOU 损失函数，以改善存在相互遮挡目标时的检测效果。最后，通过在基于浓雾去雾后的数据集上与多个目标检测模型进行了对比实验，验证了改进后的算法具有较高的检测精度。

(3) 考虑到改进后的 YOLOv7-tiny 算法兼具相对较高的精度和相对较快的速度，本文将改进后的 YOLOv7-tiny 检测网络用作海上多目标跟踪的检测器。

通过运用自适应噪声抑制卡尔曼滤波器，能够有效处理不同尺度的噪声，使改进后的跟踪器更稳定地适应目标的动态变化。此外，使用 DIOU 匹配机制替换掉原有的 IOU 匹配机制，有效优化了检测框与预测框之间的匹配度。随后，与原跟踪算法做了对比实验，实验结果表明改进后的跟踪算法有效提升了跟踪性能。最后，将改进后的算法与 PyQt5 相结合，开发出了带有人机交互界面的海上目标检测与跟踪系统。

6.2 展望

本文虽在无人机海上目标检测与跟踪方面取得一定进展，但仍存在不足之处，有待进一步优化：

（1）本文中的基于 YOLOv7 的晴天海上目标检测优化方法虽然检测精度较高，但是增加了网络的参数量，导致计算量有所上升。因此，在未来的工作中需考虑在不降低检测精度的前提下，减少网络模型参数量，提升检测速度，以实现轻量化的目标检测。

（2）本文在进行雾天海上目标检测时，自制的数据集是通过加雾算法来模拟雾天海上场景，但是与实际海雾天气场景存在一定差异。因此，在未来的研究中需利用无人机采集真实海雾图像，进一步验证该算法在恶劣天气海上目标检测的有效性。

（3）本文在开展海上多目标跟踪研究时，未对 DeepSORT 算法中特征提取重复的缺陷进行优化。所以，在未来的研究中需采用更优更快的重识别网络来解决此问题。此外，本文仅将人机交互界面部署到本地设备上，后续考虑将人机交互界面部署至无人机上，以完成实时检测与跟踪任务。

参考文献

- [1] 葛兴旺. 关于加强我国交通运输业航空应急救援能力建设的思考[J]. 城市与减灾, 2025, (01): 1-4.
- [2] Issa Zadeh S B. Analysis of European Union maritime casualties crisis management prospective[J]. Available at SSRN 4671217, 2023.
- [3] 赵德灿. 无人机在海上搜救中的应用[J]. 中国水运, 2018(11): 29-30.
- [4] Stewart D. Vidar optical radar provides new maritime search capability[EB/OL]. (2017-03-01) [2025-03-01]. <https://www.techbriefs.com/component/content/article/26476>.
- [5] Cassoulat A. Cet été, le drone helper surveillera trois plages en aquitaine [EB/OL]. (2017-06-17) [2025-03-01]. <https://www.latribune.fr/technos-medias/cet-ete-le-drone-helper-surveillera-trois-plages-en-aquitaine-741643.html>.
- [6] Seguin C, Blaqui re G, Loundou A, et al. Unmanned aerial vehicles (drones) to prevent drowning[J]. Resuscitation, 2018, 127: 63-67.
- [7] 周伟, 宋军, 成旭. 无人机在海事监管救助中的应用与选型[J]. 水运管理, 2012, 34(04): 11-13.
- [8] 祁圣君, 井立, 王亚龙. 无人机系统及发展趋势综述[J]. 飞航导弹, 2018, (04): 17-21.
- [9] 栗智. 无人机在海上救援行动中的应用现状及发展探究[J]. 中国设备工程, 2021, (11): 242-243.
- [10] 闫若璇, 廖子薇, 陆叶. 基于图像处理的无人机海上巡检救援研究[J]. 科技创新导报, 2020, 17(06): 119-120.
- [11] 安敬蕊. 海上搜寻无人机移动目标识别与跟踪[D]. 南京: 南京航空航天大学, 2014.
- [12] 费凡. 基于视觉感知的无人机海上目标自主跟踪研究[D]. 大连: 大连海事大学, 2023.
- [13] 程翔, 杜军. 基于红外感应技术的海上即时搜救无人机设计[J]. 设计, 2024, 37(12): 112-115.

- [14] 钟帅, 王丽萍. 无人机航拍图像目标检测技术研究综述[J/OL]. 激光与光电子学进展, <http://kns.cnki.net/kcms/detail/31.1690.TN.20241209.0956.042.html>, 2025-03-13.
- [15] 窦郭庆. 复杂环境下基于无人机视频分析的海上小目标检测研究[D]. 上海: 东华大学, 2022.
- [16] Yang T, Jiang Z, Sun R, et al. Maritime search and rescue based on group mobile computing for unmanned aerial vehicles and unmanned surface vehicles[J]. IEEE Transactions on Industrial Informatics, 2020, 16(12): 7700-7708.
- [17] Jin Y, Wang N, Song Y, et al. Optimization model and algorithm to locate rescue bases and allocate rescue vessels in remote oceans[J]. Soft Computing, 2021, 25: 3317-3334.
- [18] Ferrari J F, Chen M. A mathematical model for tactical aerial search and rescue fleet and operation planning[J]. International Journal of Disaster Risk Reduction, 2020, 50: 101680.
- [19] Wang S, Han Y, Chen J, et al. A deep-learning-based sea search and rescue algorithm by UAV remote sensing[C]//Proceedings of the 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC), Xiamen, China, August. IEEE, 2018: 1-5.
- [20] Ebrahimi E, Page J. UAV swarm search strategy applied to chaotic ship wakes[C]//Proceedings of the 2013 15th Australian International Aerospace Congress (AIAC), Melbourne, Australia, February. Australian International Aerospace Congress, 2013: 565-573.
- [21] Ai B, Jia M, Xu H, et al. Coverage path planning for maritime search and rescue using reinforcement learning[J]. Ocean Engineering, 2021, 241: 110098.
- [22] Cho S W, Park H J, Lee H, et al. Coverage path planning for multiple unmanned aerial vehicles in maritime search and rescue operations[J]. Computers & Industrial Engineering, 2021, 161: 107612.
- [23] Guo Y, Yi J, Hao J. UAV maritime search and rescue based on improved A* and RCPP algorithms for path planning[C]//Proceedings of the 2023 IEEE 5th

- International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, July. IEEE, 2023: 873-879.
- [24] 朱武斌. 海上无人机搜救技术研究[J]. 价值工程, 2019, 38(31): 161-163.
- [25] 韦定江, 魏霄冉, 蒋晓刚, 等. 基于 5G 通信网络条件下的无人机海上搜救方案[J]. 软件, 2021, 42(06): 145-150.
- [26] 付岩果. 海上搜救用无人机姿态控制研究[J]. 机械工程师, 2019, (05): 8-9+12.
- [27] 王四洋. 无人机集群在海上救援中的应用[J]. 航海技术, 2023, (06): 73-75.
- [28] Leira F S, Johansen T A, Fossen T I. Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera[C]//Proceedings of the IEEE International Conference on Aerospace (ICCA), Montana, USA, March. IEEE, 2015: 1-10.
- [29] Hoai D K, Van Phuong N. Anomaly color detection on UAV images for search and rescue works[C]//Proceedings of the 2017 9th International Conference on Knowledge and Systems Engineering (ICKSE), Hue, Vietnam, October. IEEE, 2017: 287-291.
- [30] Reed I S, Yu X. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, 38(10): 1760-1770.
- [31] Zhao B, Song R, Zhou Y, et al. Heuristic data-driven anchor generation for UAV-based maritime rescue image object detection[J]. Heliyon, 2024, 10(10).
- [32] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October. IEEE, 2017: 2980-2988.
- [33] 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 (CVPR), Las Vegas, NV, USA, June. IEEE, 2016: 779-788.
- [34] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Proceedings of the 2016 14th European Conference on Computer

- Vision (ECCV), Amsterdam, the Netherlands, October. Springer International Publishing, 2016: 21-37.
- [35] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December. IEEE, 2015: 1440-1448.
- [36] 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.
- [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 (CVPR), Salt Lake City, UT, USA, June. IEEE, 2018: 6154-6162.
- [38] Lygouras E, Santavas N, Taitzoglou A, et al. Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations[J]. Sensors, 2019, 19(16): 3542.
- [39] Sharafaldeen J, Rizk M, Heller D, et al. Marine object detection based on top-view scenes using deep learning on edge devices[C]//Proceedings of the International Conference on Smart Systems and Power Management (IC2SPM), Beirut, Lebanon, November. IEEE, 2022: 35-40.
- [40] Gonçalves L, Damas B. Automatic detection of rescue targets in maritime search and rescue missions using UAVs[C]//Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), Dubrovnik, Croatia, June. IEEE, 2022: 1638-1643.
- [41] Rizk M, Heller D, Douguet R, et al. Optimization of deep-learning detection of humans in marine environment on edge devices[C]//Proceedings of the 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, UK, October. IEEE, 2022: 1-4.
- [42] Rizk M, Slim F, Baghdadi A, et al. Towards real-time human detection in maritime environment using embedded deep learning[C]//Proceedings of the International Conference on System-Integrated Intelligence (ICSII), Genova, Italy, September. Cham: Springer International Publishing, 2022: 583-593.

- [43] Bai J, Dai J, Wang Z, et al. A detection method of the rescue targets in the marine casualty based on improved YOLOv5s[J]. *Frontiers in Neurorobotics*, 2022, 16: 1053124.
- [44] Kiefer B, Kristan M, Perš J, et al. 1st workshop on maritime computer vision (MACV) 2023: Challenge results[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Palm Springs, CA, USA, January. IEEE, 2023: 265-302.
- [45] Yang Z, Yin Y, Jing Q, et al. A high-precision detection model of small objects in maritime UAV perspective based on improved YOLOv5[J]. *Journal of Marine Science and Engineering*, 2023, 11(9): 1680.
- [46] 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 (CVPR)*, Salt Lake City, UT, USA, June. IEEE, 2018: 8759-8768.
- [47] Wang X, Pan Z, He N, et al. Sea-yolov5s: A UAV image-based model for detecting objects in seadronessee dataset[J]. *Journal of Intelligent & Fuzzy Systems*, 2023, 45(3): 3575-3586.
- [48] 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 (ICCV)*, Montreal, Canada, October. IEEE, 2021: 10012-10022.
- [49] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: More deformable, better results[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June. IEEE, 2019: 9308-9316.
- [50] Zhao H, Zhang H, Zhao Y. Yolov7-sea: Object detection of maritime UAV images based on improved yolov7[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Palm Springs, CA, USA, January. IEEE, 2023: 233-238.

- [51] Zhu Q, Ma K, Wang Z, et al. YOLOv7-CSAW for maritime target detection[J]. *Frontiers in Neurorobotics*, 2023, 17: 1210470.
- [52] Gao S, Liu C, Zhang H, et al. Multiscale attention-based detection of tiny targets in aerial beach images[J]. *Frontiers in Marine Science*, 2022, 9: 1073615.
- [53] Zhang Y, Yin Y, Shao Z. An enhanced target detection algorithm for maritime search and rescue based on aerial images[J]. *Remote Sensing*, 2023, 15(19): 4818.
- [54] Westall P, Ford J J, O'Shea P, et al. Evaluation of maritime vision techniques for aerial search of humans in maritime environments[C]//*Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Perth, Australia, December. IEEE, 2008: 176-183.
- [55] Westall P, O'Shea P, Ford J J, et al. Improved maritime target tracker using colour fusion[C]//*Proceedings of the International Conference on High Performance Computing & Simulation (ICHPCS)*, Leipzig, Germany, June. IEEE, 2009: 230-236.
- [56] Kaiyu X, Chaojian S. Vision enhancement system for SAR based on infrared video[C]//*Proceedings of the International Conference on Image Analysis and Signal Processing (ICIASP)*, Linhai, China, April. IEEE, 2009: 196-199.
- [57] Kiefer B, Zell A. Fast region of interest proposals on maritime UAVs[C]//*Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, London, UK, May. IEEE, 2023: 3317-3324.
- [58] Kiefer B, Quan Y, Zell A. Memory maps for video object detection and tracking on UAVs[C]//*Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Huntington Place, Detroit, USA, October. IEEE, 2023: 3040-3047.
- [59] Yang C Y, Huang H W, Jiang Z, et al. Sea you later: Metadata-guided long-term re-identification for UAV-based multi-object tracking[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, Hawaii, USA, January. IEEE, 2024: 805-812.

- [60] 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 (CVPR), Las Vegas, Nevada, USA, June. IEEE, 2016: 770-778.
- [61] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25.
- [62] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May. 2015, <https://arxiv.org/abs/1409.1556>.
- [63] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960: 35-45.
- [64] Kuhn H W. The Hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955, 2(1-2): 83-97.
- [65] 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 (CVPR), Vancouver, Canada, June. IEEE, 2023: 7464-7475.
- [66] Jobson D J, Rahman Z, Woodell G A. Properties and performance of a center/surround retinex[J]. IEEE Transactions on Image Processing, 1997, 6(3): 451-462.
- [67] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(12): 2341-2353.
- [68] Li B, Peng X, Wang Z, et al. Aod-net: All-in-one dehazing network[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October. IEEE, 2017: 4770-4778.
- [69] McCartney E J. Optics of the atmosphere: scattering by molecules and particles[J]. Optica Acta: International Journal of Optics, 1977, 24(7): 779.

- [70] Veeramani B, Raymond J W, Chanda P. Deepsort: Deep convolutional networks for sorting haploid maize seeds[J]. BMC Bioinformatics, 2018, 19: 1-9.

攻读硕士学位期间主要研究成果

- [1] Feng H Y, Ge Y, Ye G, et al. UAV maritime target detection algorithm based on improved YOLOv7[C]//Proceedings of the 36th Chinese Control and Decision Conference (CCDC), Xi'an, China, May. IEEE, 2024: 5488-5493.

致谢

光阴似箭，三年的硕士求学之旅即将结束。回首这段充满挑战与成长的时光，心中不禁感慨万千。在此，我要向所有在求知路上为我遮风挡雨、助力前行的老师、同学和家人们致以最深切的感激。

首先，要衷心感谢我的导师葛愿教授。葛老师治学严谨、工作尽责，成为我求学道路上的引路人。每当我在学习或生活中陷入困境，他总是耐心答疑解惑，为我拨开迷雾，指引前行方向。在学术探索的过程中，葛老师以其渊博知识为我照亮前路，引导我逐步成长为能够独立钻研的硕士研究生。从论文选题、理论钻研、到实验验证再到毕业论文的精心打磨，每个环节都离不开他的悉心指导与严格把关。值此之际，千言万语汇聚成一句诚挚的感谢，感恩葛老师一路以来的关怀与教导，这段师生情谊学生将铭记于心。

其次，要诚挚感谢实验室的伙伴们以及读研期间朝夕相伴的室友。在科研和生活上遇到困难时，是你们给予我无尽的鼓励与温暖的帮助。我们共同成长的这段时光，是我读研过程中最温暖的回忆。

最后，要特别感谢含辛茹苦养育我的父母。是你们用爱与呵护为我筑起坚实的后盾，让我能够心无旁骛地求学。未来的日子，我定不负家人们的期许，更加努力地美好幸福的生活奋斗！

尚模数学 唯实惟新

