学校代码 10702 密级 公开

中图分类号 TP391.4 学号 2206310838

西安工业大学
Xi'an Technological University

专业硕士学位论文

基于深度学习的红外微小目标检测方法研究

学位申请人：王雷

校内导师：喻钧 教授

企业导师：王振 高工

专业领域：计算机技术

学位类别：工 程

2025 年 5 月

西安工业大学
Xi'an Technological University

专业硕士学位论文

（学 位 研 究 生）

题目： 基于深度学习的红外微小目标检测方法研究

作者 王雷

校内导师 **喻钧** 专业技术职务 **教授**

企业导师 **王振** 专业技术职务 **高工**

专业领域 **计算机技术**

西安工业大学

2025 年 5 月 中国.西安

Research on Infrared Micro Target Detection Method Based on Deep Learning

by

Wang Lei

Thesis Submitted to the faculty of the Xi'an Technological University in Partial Fulfillment of the Requirements for the Degree of

MASTER

in

Software Engineering

Supervisor: Professor Yu Jun

Supervisor: Senior Engineer Wang Zhen

Xi'an Technological University

April 2025

Xi'an, Shaanxi, P.R.China

摘　要

在目标远距离和背景杂乱的红外成像场景中，由于噪声和背景纹理信息的干扰，导致图像中目标与背景对比度低、特征混淆，从而难以准确提取和检测目标。为应对这些难题，本文基于深度学习理论探索了一系列创新的红外微小目标检测方法。主要包括以下三个方面：

（1）为了提高红外图像的对比度和细节可见性，本文在图像预处理阶段引入了改进的双边滤波算法对原始图像进行分解，有效保留边缘细节的同时平滑噪声。并结合动态范围增强（DDE）与多尺度 Retinex（MSR）算法的优点对图像进行增强。实现图像动态范围的扩展和多尺度亮度信息的融合，从而强化了目标区域的特征表现。该改进方法显著提升了目标区域的可分辨性，有效抑制了背景干扰，为后续检测模块提供了更加清晰和高质量的输入数据。

（2）针对现有检测网络在红外微小目标检测任务中存在的特征提取能力不足、定位精度不高等问题，本文对 RT-DETR 网络进行了结构性优化，提出了三项关键改进措施：首先，在 Backbone 中引入 EMA 注意力机制，增强多尺度特征表示能力，改善微小目标在不同尺度下的感知效果；其次，在级联上下文融合模块（CCFM）中引入 CAMixing 卷积注意力机制，从空间和通道两个维度提升对细粒度目标的关注能力；最后，改进预测框筛选规则，引入 Shape-IoU 优化回归过程，采用 ATFL 提升网络分类能力。实验结果表明，改进后的算法相比原算法，平均精度均值（mAP）提升了 3.2%，并在不同复杂背景下展现出良好的鲁棒性和适应性。

（3）为了进一步提高模型的运行效率和适应性，本文进一步进行轻量化设计，提出了一种基于轻量化 RT-DETR 的目标检测模型。通过替换主干网络和特征融合模块，显著降低了模型的计算复杂度。实验结果显示，改进后的方法在红外图像目标检测任务中的最高推理速度达到了 157FPS，参数量减少了 55%，FPS 提升了 26%，在实现高效检测的同时保留了较高的检测准确率，具备良好的工程实用价值。

综上所述，本研究提出的方法在提升图像质量、优化网络结构和提高运行效率方面取得了显著成效，为红外弱小目标的准确检测提供了新的解决方案。

关键词：RT-DETR;EMA;CAMixing;Shape-IoU;轻量化

Abstract

In the infrared imaging scene where the target is at a long distance and the background is cluttered, due to the interference of noise and background texture information, the contrast between the target and the background in the image is low, and the feature is confused, so that it is difficult to accurately extract and detect the target. In order to solve these problems, this thesis explores a series of innovative infrared micro-target detection methods based on deep learning theory. It mainly includes the following three aspects:

(1) In order to improve the contrast and detail visibility of infrared images, an improved bilateral filtering algorithm is introduced in the image preprocessing stage to decompose the original image, which effectively retains the edge details and smooths the noise. Combined with the advantages of dynamic range enhancement (DDE) and multi-scale Retinex (MSR) algorithm, the image is enhanced. The expansion of the dynamic range of the image and the fusion of multi-scale luminance information are realized, so as to enhance the characteristic performance of the target area. The improved method significantly improves the resolution of the target region, effectively suppresses the background interference, and provides clearer and high-quality input data for the subsequent detection module.

(2) In order to solve the problems of insufficient feature extraction ability and low positioning accuracy in the infrared micro-target detection task of the existing detection network, this thesis optimizes the structure of the RT-DETR network and proposes three key improvement measures: firstly, the EMA attention mechanism is introduced into the backbone to enhance the multi-scale feature representation ability and improve the perception effect of micro-targets at different scales; The Cascaded Context Fusion (CCFM) module introduces the CAMixing convolutional attention mechanism to improve the ability to pay attention to fine-grained targets from the spatial and channel dimensions. Finally, the prediction box screening rules are improved, the Shape-IoU optimization regression process is introduced, and ATFL is used to improve the network classification ability. Experimental results show that the improved algorithm has an average precision mean (mAP) of 3.2% higher than the original algorithm, and shows good robustness and adaptability in different complex backgrounds.

(3) In order to further improve the operation efficiency and adaptability of the model, this thesis further carries out the lightweight design, and proposes an object detection model based on lightweight RT-DETR. By replacing the backbone network

and feature fusion modules, the computational complexity of the model is significantly reduced. The experimental results show that the maximum inference speed of the improved method in the infrared image target detection task reaches 157FPS, the number of parameters is reduced by 55%, and the FPS is increased by 26%, which retains the high detection accuracy while achieving efficient detection, and has good engineering practical value.

In summary, the method proposed in this thesis has achieved remarkable results in improving image quality, optimizing network structure and improving operational efficiency, and provides a new solution for the accurate detection of weak and small infrared targets.

Keywords: RT-DETR;EMA;CAMixing;Shape-IoU; Lightweight

目　　录

1 绪　论

1.1 研究背景

红外图像小目标检测技术是军事、航空航天和智能监控等领域的研究热点，其核心在于提升红外搜索与跟踪系统的性能，目前已广泛应用于红外预警和防御警戒任务[1]。相较于可见光成像，红外成像能够在低照度、恶劣天气和复杂环境下提供稳定的目标信息，使其在全天候目标探测与跟踪任务中具有独特优势。在军事领域，红外小目标检测技术可以探测入侵目标并实施拦截；在民用领域，该技术应用于无人机监管、大气分析、遥感探测、医疗成像及灾害救援等场景，对社会发展和公共安全同样具有重要意义[2]。

红外成像技术以热测量为基础，通过红外探测器接收目标表面的热辐射并转换为图像，具备无探测源、远距离感知、高隐蔽性和全天候适用的优点。然而，由于目标与背景间对比度低，且复杂场景下的信噪比较低，红外图像中的小目标通常表现为面积小、形状不完整、特征稀疏的目标点[3]，增加了检测的技术难度。

在早期研究中，传统算法主要通过人工设计特征进行单帧或多帧检测[4]。单帧检测算法通过增强目标与背景的对比度实现快速检测，尽管复杂度较低、实时性较强，但对复杂背景的适应性不足；多帧检测算法利用时域和空域信息预测目标轨迹，性能较好，但实时性较差，且在强杂波和弱目标条件下仍存在局限性。此外，各种基于滤波器的方法在检测概率和虚警概率上也存在瓶颈[5]。

随着神经的快速发展，研究者引入了基于深度神经网络的算法，这些方法通过自动学习特征信息展现出强大的特征提取和泛化能力。目前，深度学习目标检测算法主要分为两类：两阶段算法（如 R-CNN 系列）和单阶段算法（如 YOLO 系列）。前者通过候选区域生成与分类流程，获得高检测精度；后者通过直接定位与分类，具备更高的检测速度。然而，对于红外图像中的小目标，这些基于先验框的算法面临显著挑战。小目标通常具备分散性强、面积小、对比度低等特性，容易被复杂背景中的噪声干扰或遮挡。同时，深度学习模型在处理稀疏特征时可能丢失关键信息，导致检测性能下降。

为了缓解上述问题，研究者引入了注意力机制对网络进行改进，以增强模型对全局特征的捕获能力。注意力机制可以自适应调整模型对不同特征的关注程度，有效抑制背景干扰，提高了小目标检测的性能。然而，其特征表征能力在目标遮挡严重或分散时仍显不足。近年来，基于 Transformer 的目标检测框架逐渐受到关注。以 DETR（Detection Transformer）系列算法为代表的无先验框方法，通过

全局特征建模和自注意力机制显著提升了检测的鲁棒性和精度。与基于 CNN 的传统算法相比，DETR 系列算法能够直接建模目标间的全局关系，在处理分散小目标和遮挡问题时表现出良好的适应性[6]。

尽管基于 Transformer 的检测方法已取得显著进展，但在特征表征效率、计算复杂度和模型轻量化方面仍有改进空间。基于此，本文以 DETR 系列算法为基础，深入探索其在红外图像小目标检测中的应用与优化，旨在通过增强特征表征能力、提高鲁棒性并降低计算成本，为红外图像弱小目标检测问题提供高效解决方案。

1.2 国内外研究现状

根据提取方式的不同，目前的红外小目标检测算法主要可以分为两大类：依赖人工设计的传统单帧与多帧检测方法、自动从数据中学习特征表示的深度学习检测方法。

1.2.1 传统的红外微小目标检测方法

红外小目标检测的传统方法主要依赖图像的物理特性和手工设计特征。由于红外图像中目标常呈点状特征，且易受到背景噪声、遮挡及模糊的影响，如何从复杂背景中提取微弱目标信号始终是研究的难点。

单帧检测方法通常基于目标与背景特征的分离，旨在通过局部对比度的提升抑制背景噪声。例如，袁帅[7]等人基于双邻域差值放大检测思路，通过计算目标区域与内外双层邻域的差异，以增强局部对比度，提升亮、暗弱小目标的可检测性，并有效抑制复杂背景及噪声；吴文怡[8]等人利用 Contourlet 变换多方向、多尺度的将图像分解为不同方向上的子带，以捕捉图像中的边缘和纹理等细节信息，更为精准提取小目标特征；潘胜达[9]等人提出基于双层局部对比度机制（DLCM），该机制利用双层对角灰度差分析，结合小目标对比度先验信息，通过自适应阈值分割法提取真实目标，在提高目标对比度的同时有效抑制背景杂波和噪声；Bae[10]等人提出基于双边滤波器和时域交叉积结合算法，通过计算像素点在时域上的交叉积，分析时间轮廓特征，进而有效区分目标像素和背景像素；Li[11]等改进传统 DoG 方法，采用多尺度高斯滤波生成多个尺度的平滑图像，通过自适应调整局部对比度，增强弱小目标的局部对比特征。同时构建光谱尺度分析模型，利用自适应高通滤波方法，有效抑制背景干扰及噪声。这些单帧算法的改进具有计算相对简单、复杂度较低等优点，但在复杂多变的现实场景下，目标对比度特征不明显时，检测性能有限。

多帧检测方法利用时间序列信息，通过目标运动的连续性来增强检测性能，按照目标特性处理顺序的不同，该类算法可进一步分为跟踪前检测(DBT)算法与

检测前跟踪(TBD)算法[12]。

DBT 算法通常首先对每一帧图像进行目标检测，提取出图像中所有可能的目标区域，然后基于检测到的目标区域，通过目标跟踪算法在后续帧中跟踪目标的运动轨迹。如娄康[13]等人基于卡尔滤波方法，利用目标的运动轨迹和速度信息，结合背景建模和前景检测技术，增强目标与背景的区分度并预测下一帧的目标位置，提取目标轨迹；Wan[14]等人提出了一种基于全变分的帧间红外图像块分割模型，该模型通过将图像分解为稀疏目标矩阵和低秩背景矩阵，实现目标与背景的有效分离。通过施加帧间相似性约束，保证目标在时间序列中的连续性，同时引入全变分正则化项来缓解噪声引起的虚警。该类算法通过独立的目标检测步骤，并结合多种检测方法和跟踪策略，因此检测精度高、适应性强，但由于每一帧都需要重新进行目标检测，计算开销大、实时性差，且在目标和背景相似时，算法容易出现误检和漏检问题。

TBD 算法是首先进行目标跟踪的多帧处理方法，该类算法的核心思想为在多个帧中通过跟踪的方式预测目标的位置，然后在目标所在的区域进行局部检测，进一步确认目标。如刘德连[15]等人提出基于时间轮廓的小型运动目标检测方法，通过分析像素值的时间行为，利用时间轮廓的停滞点连接线（CLSP）作为基线，计算残差时间轮廓，建立目标和背景的时间轮廓模型，在复杂背景下能有效的提取运动目标；Pang[16] 提出 FDMDEA-STT 红外小目标检测模型。模型通过充分利用目标和背景的时空先验信息，构建时空张量，将目标检测问题转化为低秩稀疏张量优化问题，并采用交替方向乘子法求解，从而提高红外小目标的检测精度。此类方法通过对多个帧的目标跟踪进行优化，从而在多帧图像中提供更为准确的目标检测，但其对于目标跟踪的准确性要求较高，易受噪声和目标运动变化的影响。

因此，多帧方法在检测性能上优于单帧方法，尤其是在运动目标的轨迹推测方面表现突出。然而，多帧方法对计算资源需求较高，实时性不足，且依赖于单帧检测的初步性能，对低信噪比场景的适应性仍有局限。

1.2.2 基于深度学习的红外微小目标检测方法

近年来，深度学习技术在红外小目标检测中的应用取得了显著进展，主要分为 One-Stage 方法、Two-Stage 方法以及基于注意力机制和 Transformer 的框架。

One-Stage 方法具有较高的实时性，其通过回归直接实现目标分类与定位，不生成候选框而直接对物体进行分类和候选框预测，简化了网络结构，准确度虽较 Two-Stage 的目标检测框架低，但实时性较好。代表算法有 YOLO 系列、SSD 系列、Anchor-free 系列等。针对红外小目标检测，李慕锴[17]等人借鉴 SENet 中对特征进行权重重标定的思路，通过对特征图中的各个通道进行加权来增强小目

标的关注度，同时抑制背景干扰。其将 SEblock 模块引入 YOLOv7，实现对小目标检测的准确率大幅提升。徐延想[18]等人针对红外小目标检测中常见的特征不明显和提取困难的问题，基于 RefineDet 网络设计出 IoU 预测模块，并提出目标搬移算法。通过间接增加小目标的数量并对小目标进行位置移动，增加目标的多样性，促进了网络在训练过程中的小目标识别能力，提升检测效果。

Two-Stage 目标检测算法通常包括两个阶段，先通过选择性搜索生成一组可能包含目标的候选区域，再对这些区域进行分类和精确定位。典型的 Two-Stage 算法即 R-CNN 系列，该类算法具有较高的检测精度，但计算开销较大，实时性差。杨子轩[19]等人针对红外小目标纹理信息稀疏的问题，将通道及空间注意力机制引入 Casacd-CNN，同时，减少锚框尺寸以匹配小目标，优化小目标检测效果。，蒋志新[20]等人针对红外图像本身存在的低对比度、低信噪比问题，在图像预处理阶段结合直方图均衡化和 Retinex 算法进行图像增强，以提高图像质量，同时，改进 Faster-CNN 网络的损失函数，使得模型在小目标检测任务中的表现得到了显著提升。

而除了 One-Stage 的方法与基于 Two-Stage 的检测方法，近年来逐渐兴起的基于注意力机制的检测框架在对全局特征提取起到了更高的性能。其不仅可以将注意力机制作为辅助块加入网络模型，基于 transformer 的检测框架 DETR 对小目标检测方面起到了更好的效果。崔莹[21]等人使用空间注意力模块结合标准 Transformer 块设计了一个用于增加模型深度的 DFPN 块，将其嵌入到 Deformable DETR 模型中，提高模型了对深层纹理信息的提取能力。Wei[22]等人提出的 CG-Net，其利用 Transformer 框架改进小目标检测，自适应的为每个通道分配校准权重，根据输入图像的特征动态调整各个通道的重要性，从而在多个通道间实现加权聚合。加权后的特征通道被重新整合，形成更加精确的特征表示，进而更好地处理复杂背景下的小目标检测性能。Pang[23]等人提出了一种名为 R2-CNN 的高效算法，用于大规模遥感图像中的微小物体检测，R2-CNN 借鉴 transfomer 全局提取思路，通过引入一个区域候选生成模块来迅速确定可能包含小目标的区域，利用卷积神经网络（CNN）来对候选区域进行特征提取，并结合一种新的候选区域评估机制，以提高小目标的检测精度。LiuShilong[24]等人提出了一种基于动态锚框的改进算法 DAB-DETR，该模型在 DETR 目标检测方法的基础上，借鉴锚框预测思路，将解码器中的 Query 视为锚框的四维坐标，通过动态调整锚框位置，使模型更高效地学习目标的空间信息，从而提升检测性能。Wang[25]等人基于 ViT[26]提出了一种新的旋转可变窗口注意力机制，该机制通过对输入图像进行局部窗口划分，每个窗口可以根据其特定的旋转角度进行自适应调整，同时，不同窗口之间共享上下文信息，使得模型能够更好地捕捉到目标的局部特征。这种改

进在复杂背景下，尤其是在处理旋转、缩放等变换的小目标检测任务中检测效果显著。

尽管上述方法在红外小目标检测领域取得了一定成果，但仍存在以下挑战：实时性与精度的平衡、复杂背景下的鲁棒性不足以及特征提取难度较高。因此，探索更高效、更精确的小目标检测方法仍具有重要意义。

1.3 主要研究内容

本文聚焦于红外图像中的飞机微小目标检测问题，围绕低对比度、成像模糊以及复杂背景等挑战展开研究。针对这些问题，本文从算法优化、网络模型设计与轻量化改进三个方面进行深入探讨，以提升检测精度并满足工程应用需求。

主要研究有以下三点：

（1）图像增强与细节提升

红外图像中的微小目标通常表现为低对比度、边缘模糊和低信噪比，这使得目标在复杂背景中难以被有效识别。为此，本文通过滤波分解技术，结合 DDE 与 MSR 局部增强算法，通过多尺度处理和局部增强方法，突出微小目标的细节信息，增强目标与背景的对比度，提升后续检测和特征提取的效果。

（2）基于 RT-DETR 的小目标检测算法

针对现有深度学习方法在红外图像小目标检测中的精度不足问题，本文提出了一种基于改进 RT-DETR 算法的红外小目标检测方法。通过引入新的注意力机制，如多通道注意力模块，提升特征图提取和目标定位能力；优化传统的回归损失函数，抑制背景噪声的干扰，提升网络对小目标特征的捕捉能力。

（3）RT-DETR 轻量化网络模型

为满足实际工程中的红外图像目标检测，研究一种轻量化的 RT-DETR 目标检测网络模型，采用轻量化主干网络结构，通过减少网络参数量、简化计算操作来降低计算复杂度，同时保持较高的检测精度，能够在保证性能的同时，显著提升推理速度和减少内存占用。

1.4 论文结构安排

本文的结构安排如下：

第一章绪论主要介绍了研究的背景和意义，并系统总结了红外小目标检测领域的传统检测算法和基于深度学习算法的相关知识。接着，对国内外红外小目标检测领域的研究现状进行了总结与分析，明确了当前红外小目标检测的应用场景及面临的挑战。最后，提出了本文的研究主题，并对文章的整体结构进行了概述。

第二章主要概述了红外图像特性及相关检测模型。首先，探讨了红外图像的独特特性及其增强技术，其次，介绍了两种典型的基于 Transformer 的目标检测模型，包括首次使用 Transformer 架构进行特征提取和分类的 ViT 检测模型，以及将 Transformer 架构融入目标检测任务的端到端 DETR 模型。为后续研究提供了理论支持。

第三章为改进 RT-DETR 的红外小目标检测。针对红外图像中目标信号弱、对比度低导致检测精度较低的问题，本章提出了一种改进的 RT-DETR 目标检测算法。该算法首先将 EMA 注意力机制应用于 RT-DETR 的骨干网络，以增强其特征提取能力；其次，在 CCFM 进行特征融合时，加入 CAMixing 卷积注意力模块，以提高检测过程中的目标关注度；接着，采用了 ATFL 及 Shape-IOU 改进损失函数，以优化分类效果并更精准地衡量预测框与真实框之间的匹配程度。

第四章为针对 RT-DETR 的轻量化研究，本章针对 RT-DETR 在计算量较大的问题进行分析，提出了一种轻量化的优化方案。具体包括：将 MobileNetV4 作为骨干网络以减少计算开销，并利用轻量化注意力模块替换原特征融合模块，从而显著提高了模型在红外图像目标检测中的速度和准确度。最终，基于该优化方案，设计红外微小目标检测系统，进一步完善研究流程。

第五章主要围绕实验设计与结果分析展开。本章详细介绍了实验方案和具体实施流程。通过一系列消融实验验证了各优化策略对模型性能的具体贡献。同时，将改进后的 RT-DETR 算法与多种主流目标检测模型在相同数据集上进行对比分析，实验结果表明，改进后的 RT-DETR 算法在红外小目标检测任务中表现出显著的优势，具有较高的精度和鲁棒性。

第六章对本文的主要研究工作进行了总结。归纳了改进 RT-DETR 在红外小目标检测中的方法与成果。最后，结合当前研究中的局限性与挑战，提出了未来可能的研究方向和发展趋势，为后续研究提供了参考。

2 相关技术与理论基础

本章首先介绍了红外图像的成像原理及其特性，分析了红外图像在目标检测中的独特挑战。然后介绍了本文所用的红外图像增强方法，通过不同的技术手段提高图像的对比度与清晰度，以便更好地提取目标特征。接着介绍了以transformer为核心的目标检测框架，详细讲解了 Transformer 在特征提取、上下文建模等方面的优势，为本文后续提出的红外微小目标检测网络提供理论基础。

2.1 红外成像原理

根据物体的热辐射特性原理，在自然界中，任何物体的温度高于绝对零度（-273°C），都会散发红外辐射。物体的温度越高，辐射的辐射强度越大，这个过程遵循普朗克辐射定律，如式 2.1 所示，其描述了一个黑体在某一温度下单位面积内单位波长范围内辐射的强度。

$$I(\lambda,T)=\frac{2hc^2}{\lambda^5}\cdot\frac{1}{e^{\frac{hc}{\lambda kT}}-1} \tag{2.1}$$

其中，I(λ，T)是辐射波长为λ时，温度为 T 的物体辐射的辐射强度，h 是普朗克常数，c 是光速，k 是玻尔兹曼常数。该公式揭示了温度越高，辐射强度越大，且波长较短的辐射越强。

不同温度、表面特性和材质的物体辐射红外线的强度和波长有所不同。通过探测器检测目标与背景之间的红外辐射差异，可以获取不同的红外图像。热红外图像是将物体表面的温度分布转化为人眼能够看到的图像。通过这种方式，即使在夜间或可见光不足的环境中，也可以对目标进行远程热状态成像和温度测量，进而实现对目标的智能分析和判断。与热红外图像不同，灰度图像在红外成像系统中的成像过程并不直接显示温度信息，而是反映图像中不同区域的红外辐射强度。探测器接收到红外辐射信号并将其转换为电信号后，成像系统对这些电信号进行处理并映射成灰度图像。灰度图像的亮度等级通常反映了辐射强度的差异，高辐射强度的区域会显示为亮色，低辐射强度的区域会显示为暗色。因此，这类图像通常用于目标检测，物体的形状、轮廓以及与背景的对比度都能通过图像的灰度特性进行分析。

红外成像系统作为红外图像形成的基石，主要由红外探测器、红外光学系统、信号处理器和显示设备四部分组成[28]。其工作流程如图 2.1 所示。首先，红外探测器负责捕获物体发出的红外辐射信号，将其转化为电信号。红外光学系统则将这些辐射信号聚焦到探测器的感光元件上，确保信号的准确采集。接着，信号处

理器对原始电信号进行放大、去噪和图像增强，以提高图像质量。最后，处理后的信号通过显示设备转化为图像，通常为灰度图像或热图像，后者使用伪彩色映射将物体表面温度的空间分布显示出来。

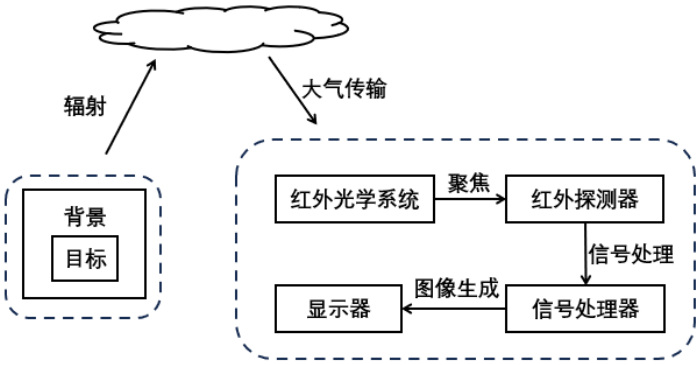


图 2.1 红外成像过程

2.2 红外图像特性分析

红外图像是表示物体红外光强度的一种图像形式。无论是灰度图像或热图像，一幅红外图像$I(m,n)$主要包含目标、背景、噪声三部分，如图 2.2 所示。

$$I(m,n) = T_{\text{target}}(m,n) + T_{\text{background}}(m,n) + N(m,n) \tag{2.2}$$

其中，$T_{\text{target}}(m,n)$表示目标的辐射信息，$T_{\text{background}}(m,n)$表示背景的辐射信息，$N(m,n)$表示噪声。

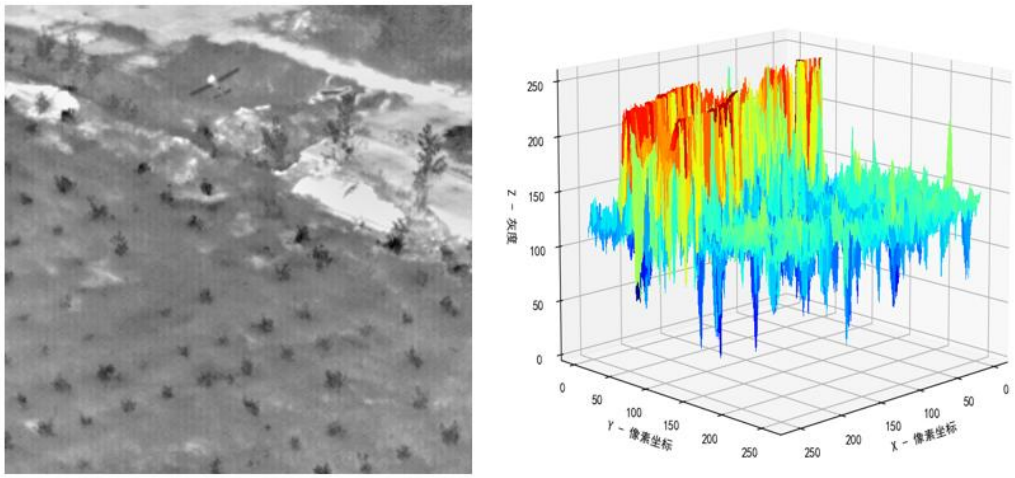


图 2.2 红外图像及其 3D 图像

通过红外灰度图像及其辐射强度 3D 图能看到：对于背景，红外辐射强度分布不均匀，但大部分区域（尤其是林地植被背景部分）辐射强度较低且分布均匀。对于目标，由于背景也存在大量高辐射强度区域，导致目标与背景对比

度较低，且特性不够显著，与背景的差异不明显。因此，增强目标特征的同时，抑制背景的影响是图像增强的主要方向。

2.2.1 红外图像点目标特性

在灰度图像中，红外点目标通常是通过红外辐射强度差异表现出来的。与热图像不同，灰度图像反映的是场景中各部分的辐射强度，通过图像的灰度值变化来表达。这类图像的目标区域显示为亮度较高的部分，而背景则通常为较低亮度的区域。通过分析图像点目标的特性，可以有效地为后续的图像增强和目标检测提供有价值的信息。

（1） 几何特性

在红外灰度图像中，点目标通常表现为一个小而明亮的区域，具有点状扩散的特征。其能量分布类似于二维高斯函数，如式（2.3）所示。这是由于目标的辐射强度较高，而与背景的差异较小，使得目标在图像中往往呈现为一个亮斑，并且亮度随着离目标中心的远离而衰减。对于远距离目标，辐射强度通常较低，经过成像系统处理后，目标在图像上显现为一个微小的亮点。

$$I(x,y) = I_0 \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2.3}$$

其中，I_0是目标中心的灰度值，σ是高斯分布的标准差，控制目标在图像中的扩散程度。目标的辐射强度在图像中逐渐衰减，导致它的形态表现为一个亮度较高的圆形区域，随着距离中心的增大，灰度值逐渐减小。

（2） 辐射强度特性

在灰度图像中，红外点目标的表现是通过目标的辐射强度来实现的。如图 2.3 所示。物体的温度较高时，其辐射的红外线强度较大，这在红外成像系统中转化为更高的灰度值。在远距离或者低温的情况下，目标的辐射强度较低，因而灰度值也较低。如图 2.3 所示。

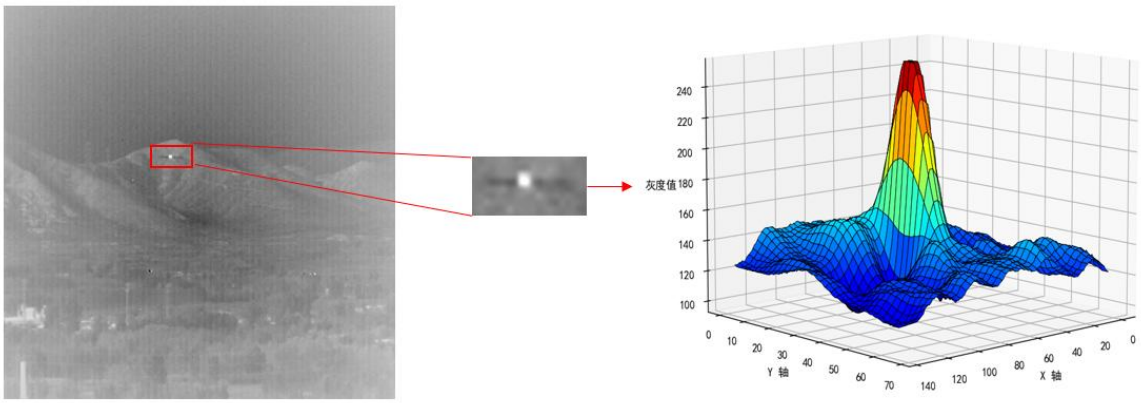
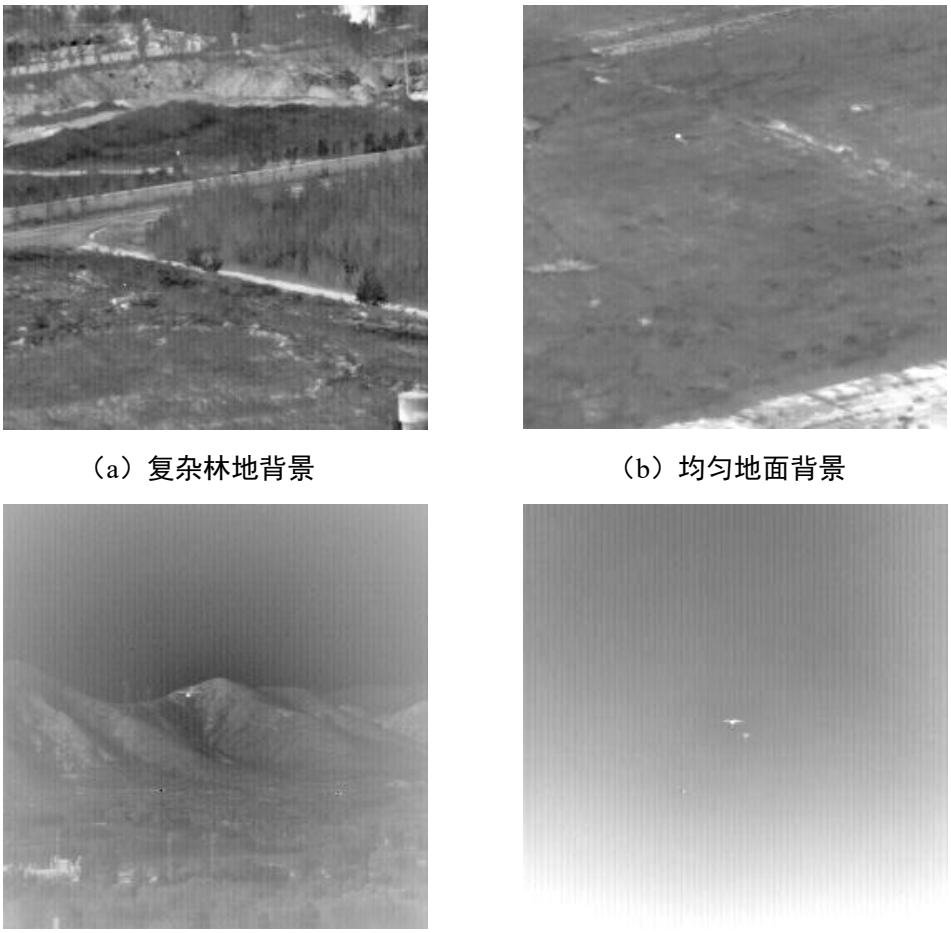


图 2.3 红外点目标灰度 3D 图

2.2.2 红外图像背景特性

红外成像技术依赖于物体表面辐射的差异来形成图像，这种成像方式使得红外图像的背景往往包含大量复杂的信息和噪声，给目标检测带来了极大的挑战。如图 2.4 所示红外背景图像。



（a）复杂林地背景　　（b）均匀地面背景

（c）复杂山区背景　　（d）均匀天空背景

图 2.4 红外背景图像

图 2.4(a)中，图像中主要是山地地形，轮廓清晰，高度起伏不平，部分山坡区域亮度较高，山体背景的灰度分布较为均匀，但由于植被覆盖，部分区域存在纹理细节，道路等线性结构较为明显，与周围的植被形成对比。而目标目标红外特征较弱，与背景对比度较低。因此，具有复杂地面背景的红外图像对检测算法的挑战较大，尤其是当目标与背景有较低对比度时，检测的准确性容易受到影响。

图 2.4(b)中，整体呈现为灰度较低的区域，具有均匀且连续的灰色分布。地面上分布着较浅的线性结构和一些零散的高亮点，地面纹理和高亮区域的形

态与点目标具有一定相似性，目标与背景之间的亮度差异不明显。地面整体灰度值变化缓慢，低对比度使目标在图像中较难被直接辨识。

图 2.4(c)中，山体的轮廓清晰且高度起伏不平，部分山坡存在高亮区域，但与目标对比度较高，较模糊的植被或地面细节，与山体背景叠加，使得整体图像具有一定的层次感，山体上方的天空区域较为均匀，灰度值较高，无明显的纹理或干扰物，整体检测难度较低。

图 2.4(d)中，远距离天空背景呈现均匀灰度分布，存在轻微的条纹噪声，但整体均匀性较好，目标较为突出，检测难度低。

因此，红外图像背景的复杂性对目标检测算法的性能具有重要影响。复杂地面背景和低对比度场景（如图 2.4a、b）容易导致虚警和漏检，而具有清晰轮廓和高对比度特征的山区背景（如图 2.4c）则有助于目标检测的实现。在均匀且无干扰物的天空背景（如图 2.4d）下，目标检测效果最佳。为提升红外目标检测的鲁棒性，针对不同背景特性，需设计具有自适应能力的检测算法，增强对复杂场景的抗干扰能力，并充分利用目标与背景的对比度和形态特征进行有效区分。

2.2.3 噪声特性

红外图像的噪声特性表现为多源性、随机性和非均匀性，且易受外界环境的影响。噪声特性主要来源于成像系统、传感器特性以及外界环境因素。

（1）成像系统噪声：红外成像系统在工作过程中会引入系统噪声，主要包括电子噪声和量化噪声。电子噪声是由于系统内部电子元件引起的随机信号波动，量化噪声是在模拟信号的数字化量化过程中产生的，这类噪声都与成像设备的性能密切相关，尤其在低信噪比情况下更为显著。

（2）传感器噪声：红外传感器在探测过程中会引入热噪声、暗电流噪声和读出噪声。其中，热噪声是由于传感器自身温度变化引起的随机噪声，而暗电流噪声与传感器内部结构及工作温度相关，通常在长时间曝光或低温条件下更为明显。读出噪声通常是由于传感器内部电子放大器的增益引起的误差。它会影响图像的质量，使得在低光照条件下的图像更加噪声化，特别是图像的高频部分。

（3）环境噪声：环境噪声主要来源于外部因素，包括气候、周围环境、地形以及其他干扰源等。大气湍流造成热辐射的不稳定和图像的模糊，地面上的反射、散射或光滑表面都会导致热辐射的反射现象，尤其是对于某些建筑表面或湿润的区域，反射的热辐射可能与目标的辐射特征相似，导致目标和背景难以分辨。

这些噪声不仅降低了图像的清晰度，还会对目标检测和识别算法的准确性造成影响。因此，针对红外图像噪声特性，通过结合去噪算法和图像增强技术，提高信噪比，增强目标的可检测性。

2.3 红外图像增强

红外图像通常存在对比度低、噪声高、细节不清晰等问题，影响目标的检测和识别效果。现有对图像预处理的算法大多在可见光领域成果显著。但是在对红外图像的处理中存在不足。例如，AHE 在局部区域内应用直方图均衡化来增强对比度，克服了传统直方图均衡化的部分不足，但在图像亮度变化剧烈的区域，容易产生伪影，影响图像的自然性。拉普拉斯算法通过二阶微分来突出图像的细节和边缘，但是对噪声敏感，容易导致图像过度锐化。

在对红外图像的预处理中，动态细节增强(DDE)和 Retinex 算法是两种常用的图像增强技术。DDE 通过滤波算法分离图像的基础层和细节层，并增强细节层，实现细节的显著增强。但是对基础层的亮度与对比度处理效果有限，同时在放大细节的同时也容易放大噪声。多尺度 Retinex 算法则通过对数变换和平滑处理，改善图像的亮度和全局对比度。但在细节层次的处理效果有限，在复杂场景下，局部对比度的提升效果不明显。因此，将 DDE 与 Retinex 算法结合使用，能根据图像内容的复杂度动态调整增强图像的细节信息与清晰度。

2.3.1 滤波算法改进

DDE 算法的核心在于，利用边缘保持滤波器将图像有效分解为低通分量（基础层）和高通分量（细节层），通过对高通分量的增强或抑制，实现图像的增强、去噪与细节提取。针对红外图像显著的空间结构特征和像素强度差异，本文采用双边滤波算法进行图像分解，并在此基础上对传统双边滤波方法进行改进，以提升其在红外图像处理中的分解性能与细节保留能力。

原始双边滤波算法如式 2.4 所示。

$$I_B = \frac{1}{w_q}\sum_{p\in S} G_S(x_i, x) * G_r(x_i, x) * I(x_i) \tag{2.4}$$

其中，$I(x_i)$表示原始红外图像，w_q表示滤波窗口内每个像素值的权重和，用于权重的归一化，G_S表示像素值权重，G_r表示空间距离权重，采用值域高斯权重函数计算，其公式计算如式 2.5 所示。

$$G_r(x_i, x) = \exp\left(-\frac{I(x_i) - I(x)^2}{2\sigma_r^2}\right) \tag{2.5}$$

但在红外图像中，尤其是微小目标检测中，值域高斯权重函数仅考虑像素间的灰度差，当目标灰度与背景接近，$|I(x_i) - I(x)|$值很小，导致滤波器难以有效区分边缘与非边缘，从而造成边缘模糊、目标弱化。因此，传统双边滤波在红外图像分解中对低对比度区域的平滑处理能力存在不足，易导致基础层提取不准确、细节层噪声残留，从而影响整体分解质量。为增强对图像结构信息的感知，本文

引入梯度感知机制，将像素灰度差与梯度差联合建模，构建梯度感知的值域权重函数。该方法通过结合像素灰度与梯度信息，实现对边缘与噪声的智能区分。

改进后的值域高斯权重改进滤波算法，公式如式 2.6 所示。

$$G_r^{new}(x,x_i)=\exp\left\{-\frac{\left(I(x_i)-I(x)\right)^2+\lambda\|\nabla I(x_i)-\nabla I(x)\|^2}{-2\sigma_r^2}\right. \tag{2.6}$$

其中$\nabla I(x)$表示图像在x点的梯度信息，由 Sobel 算子计算。$\|\nabla I(x_i)-\nabla I(x)\|^2$是梯度向量的欧氏距离。λ是权衡系数，用于平衡灰度差与梯度差的影响，当两个像素灰度相近但梯度差异较大时，则认为在边缘区域，降低权重，以减少边缘模糊。该方法在增强边缘保持能力的同时，有效抑制了高频噪声对基础层提取的干扰，从而提升红外图像分解的整体性能。最终分解效果对比如图 2.5 所示。

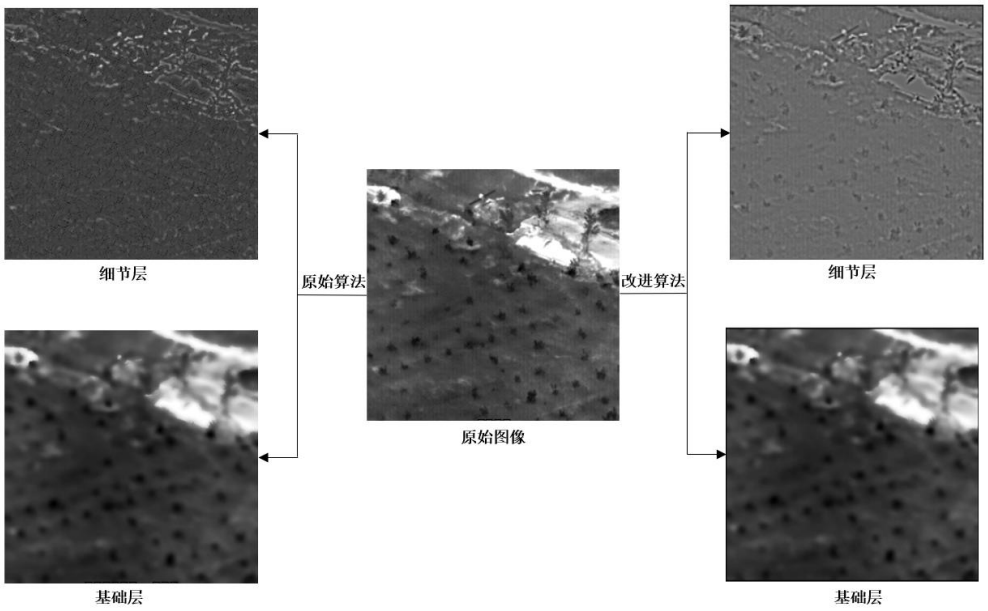


图 2.5 算法分解对比

从图中可以看出，改进后的算法在细节层中保留了更多微小结构信息，同时基础层更加平滑、噪声更少，整体分解效果更加清晰准确，验证了该方法在红外图像处理中的有效性。

2.3.2 图像增强

最终利用改进的双边滤波算法实现图像分解并进一步增强流程如图 2.6 所示。

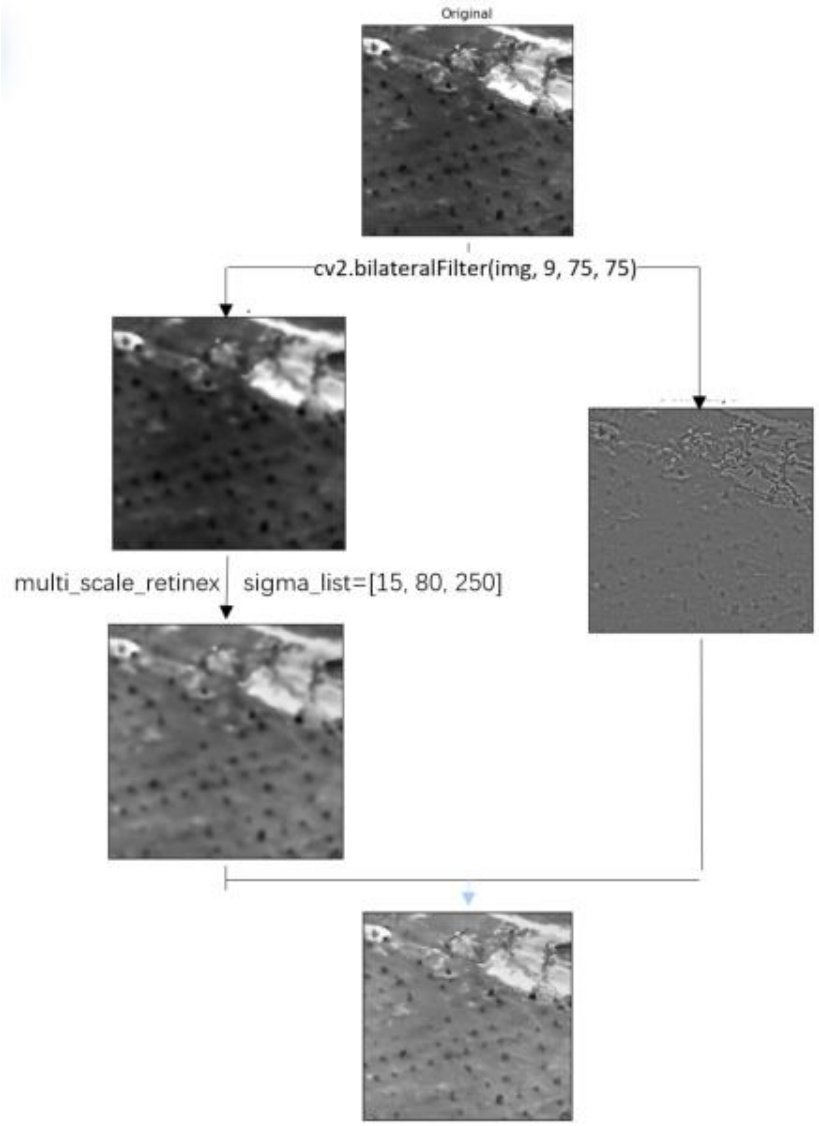


图 2.6 图像增强流程

首先，利用双边滤波将原始图像分解为基础分量I_B，而细节分量I_D为原始图像减去基础分量的结果，输出公式如式 2.7、2.8 所示：

$$I_B = \frac{1}{w_q}\sum_{p\in S} G_S(p) * G_r^{new}(p) * I_p \tag{2.7}$$

$$I_D = I_p - I_B \tag{2.8}$$

基础分量I_B包含图像的大尺度结构和光照信息。对基础分量采用 MSR 增强局部细节，MSR 公式如式 2.9 所示。

$$Log\big(R(x,y)\big) = Log\big(R(x,y)\big) + Weight(i) * \Big(Log\big(I(x,y)\big) - Log\big(Li(x,y)\big)\Big) \tag{2.9}$$

MSR 将原始图像分为三个维度分别进行三次不同 sigma 参数的 SSR 操作。对三次的 SSR 结果加权求平均，然后把三个维度合并，得到 R。其中，双边滤

波滤波器核大小为 9×9，空间标准差和色彩标准差采用均衡值 75。MSR 的三次 SSR 操作 sigma 因子分别为 15，80，250，$Weight(i)$表示每个尺度对应的权重，要求各尺度权重之和必须为 1，经典的取值为等权重 1/3。

经过 MSR 增强基础层之后得到R_B如式 2.9 所示。

$$R_B = MSR(I_B) \tag{2.9}$$

最终将细节分量重新与增强后的基础分量进行结合得到最终的红外图像I。

$$I = R_B + I_D \tag{2.10}$$

2.4 基于深度学习的目标检测算法

目标检测技术是计算机视觉中的核心任务之一，其核心在于从图像或视频中准快速识别并定位物体。传统方法依赖手工特征和机器学习，检测速度受限，且难以充分利用大规模数据。而深度学习的引入，特别是卷积神经网络（CNN），显著提升了检测速度与精度。近年来，随着自然语言处理（NLP）领域 Transformer 架构的兴起，人们开始探索将其应用于计算机视觉任务的可能性。

2.4.1 ViT 网络模型

Vision Transformer(ViT)是一种基于 Transformer 架构的视觉模型，由 Alexey Dosovitskiy 等人在 2020 年提出。ViT 首次用 Transformer 模型直接处理图像数据，突破了传统卷积神经网络（CNN）在计算机视觉任务中的主导地位。尽管 ViT 并非首个在视觉任务中应用 Transformer 的网络结构，但其卓越的效果和强大的泛化能力，使其成为 Transformer 在计算机视觉领域应用的里程碑。

ViT 的核心思想是将图像视为一个序列的补丁（patch）进行处理，与 Transformer 在自然语言处理（NLP）任务中将句子视为单词序列类似。模型架构如图 2.7 所示。对于目标检测任务，其工作流程主要通过将图像分割为 patch、转换为 token 序列、添加位置嵌入和类别 token、通过 Transformer Encoder 进行编码以及通过 MLP Head 进行预测。

ViT 模型主要由 Embedding 层、Transformer Encoder、MLP Head 三个模块构成。

Embedding 层主要完成对图像的分割、对数据进行线性嵌入并添加位置信息，构成 Transformer 的嵌入输入序列。对于标准的 Transformer 模块，要求输入的是 token（向量）序列，因此 Embedding 层的输入首先将图像进行分块，当输入图像的大小为 H×W×C，将其分割为固定大小的补丁（patch），每个补丁的大小为 P×P。这样，图像可以被分割成 N=$(H \times w)/P^2$个补丁，每个补丁展平成一维向量，维度为P^2×C。每个补丁通过线性变换映射到 D 维特征空间，得到大小为 N×D

的补丁嵌入。同时为补丁序列添加固定或可学习的位置编码（Position Embedding），保留图像的空间信息。

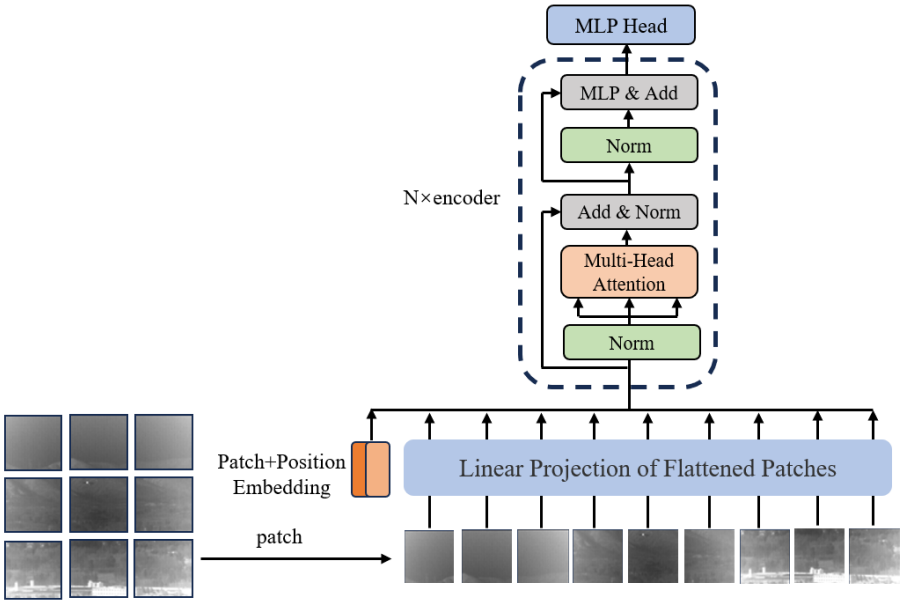


图 2.7 ViT 模型结构

Transformer Encoder 层由多个 Transformer 层堆叠组成，逐步提取全局特征。与标准 Transformer 层相同，每层主要由多头注意力机制捕获序列中任意两补丁之间的全局关系、前馈网络提升特征的表达能力、残差连接避免梯度消失问题，同时在每个 Transformer 层之间有层归一化（Layer Norm），以保持特征分布的一致性。此外引入 MLP Block 用于进一步增强特征的非线性表示能力。每个 MLP Block 包含两个全连接层，中间使用 GELU 激活函数和 dropout 机制来提高模型的表达能力并防止过拟合。Encoder 结构如图 2.8 所示。

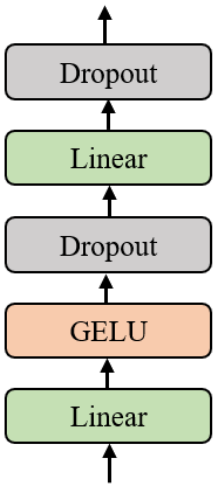


图 2.8 MLP Block 结构

最终，通过 Transformer Encoder 输出包含 N+1（其中第一个特征向量为[CLS]token 的嵌入，表示整幅图像的全局特征）个特征向量的序列，并将序列输入 MLP Head 层。MLP Head 通过一系列全连接层和激活函数对[CLS]token 进行处理，用于完成最终的任务（如分类或回归）。

2.4.2 DETR 模型

DETR（DEtection TRansformer）是首个基于 Transformer 的端到端目标检测框架，由 Facebook AI 提出。其创新点在于将目标检测问题转化为集合预测（Set Prediction）问题，并使用 Transformer 进行特征建模和预测，从而避免了传统目标检测方法中复杂的后处理（如非极大值抑制，NMS），极大的简化了目标检测流程。

DETR 使用标准的 Transformer 编码器-解码器架构，其网络结构主要包括四个模块：特征提取的 Backbone、特征建模的编码器（Encoder）、目标解码的解码器（Decoder）和预测层（FFN）。模型结构如图 2.9 所示。

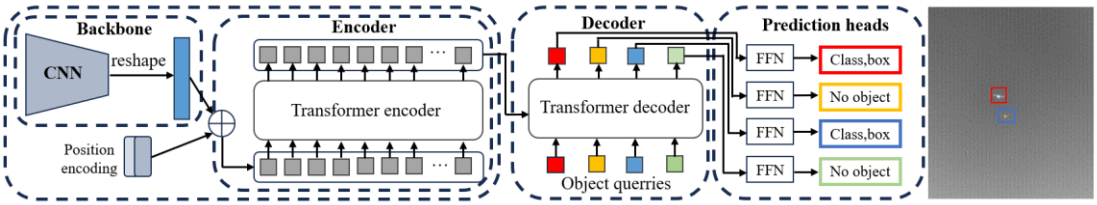


图 2.9 DETR 模型结构图

（1）Backbone

Backbone 用于从输入图像中提取特征图。作为 Transformer 模块的特征输入，DETR 常用的 Backbone 网络包括 ResNet 系列和 HGNet 系列等预训练卷积神经网络。这些网络通过多层卷积和池化操作，将输入图像下采样至 1/32 的尺度。尽管空间分辨率有所降低，但语义表达更加丰富。同时，为弥补 Transformer 对输入顺序不敏感的局限性，特征图中显式加入了位置编码，以嵌入空间位置信息。位置编码采用正弦和余弦函数生成，其公式如式 2.11、2.12 所示。

$$PE(pos, 2i) = sin\left(\frac{pos}{10000^{2i/D}}\right) \tag{2.11}$$

$$PE(pos, 2i+1) = cos\left(\frac{pos}{10000^{2i/D}}\right) \tag{2.12}$$

其中，pos 是特征图的位置索引，i 是维度索引，D 是特征索引。通过将位置编码添加到每个像素的特征通道中，模型能够显式地感知特征的空间位置信息，从而增强对空间关系的建模能力。

（2）Encoder

DETR 的 Encoder 基于标准 Transformer 编码器结构，旨在捕获输入特征图中任意像素之间的全局上下文关系。Backbone 提取的特征图被展平并通过 1×1 卷积降维，以降低计算复杂度并优化特征表达。降维后的特征输入到 Encoder 中，利用多头自注意力机制建模全局上下文信息。与标准 Encoder 不同，DETR 在每个多头自注意力模块之前都嵌入了位置编码，并针对特征图的 x 和 y 维度分别计算位置编码后进行拼接。Encoder 具体结构如图 2.10 所示。这种设计显著提升了模型对图像空间布局的感知能力。

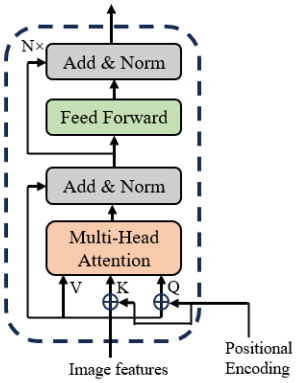


图 2.10 Encoder

（3）Decoder

DETR 的解码器同样基于标准 Transformer 解码器架构，且引入了一些关键设计来实现集合预测任务,具体结构如图 2.11 所示。

首先，为了解决同时检测多个目标的问题，Decoder 不再使用单一的 class token，而是采用了固定数量的查询 tokens(Object Queries)作为输入。其选择了固定的 N=100 个查询 tokens，这一数量通常大于图像中实际目标的数量，确保模型能够覆盖所有潜在目标。这些查询 tokens 是 Learnable Embeddings，通过并行处理后与图像特征一一对应。

Decoder 的输出是 N 个经过注意力机制和映射后的 tokens，这些 tokens 被传递给一个前馈网络（FFN），预测出对应的 N 个目标的类别分数（包括背景类别）和边界框坐标。与分类任务中的 class token 不同，DETR 的 Object Queries 在每层 Decoder 中都加上了 Positional Encodings，以增强每个查询 token 的空间表示能力。这种位置编码的设计使得每个+查询 token 能够关注特定的空间区域，预测该区域内是否存在目标，以及该目标的类别和位置。此外，为了进一步提升位置信息的建模能力，Decoder 在每层的 key 和查询 tokens 上都加上了 Positional Encodings。

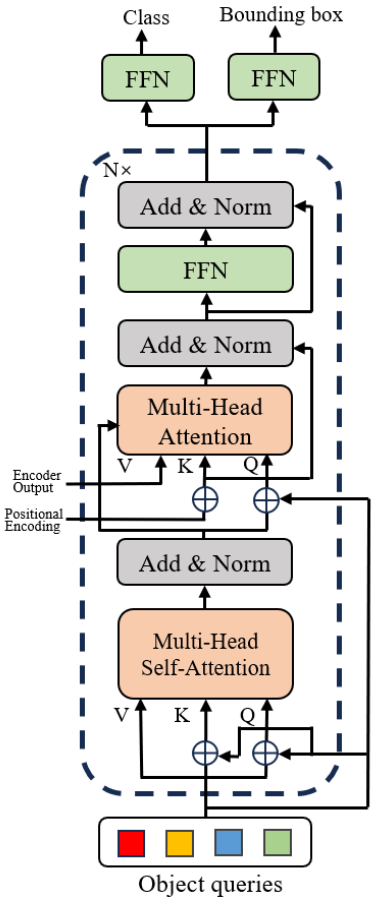


图 2.11 Transformer Decoder

这一设计的固定预测数量（N）简化了模型输出，方便显存对齐，但也带来了一些限制。当图像中的目标数量过多时（例如接近 N 个），模型可能会出现目标丢失的问题，因为网络的检测能力在 N/2 个目标左右就趋于饱和。在图像实例较少时，这种设计表现良好，但在实例密集的情况下，检测结果的上限可能受限于固定数量的查询 tokens。总体而言，这一设计通过简单直观的方式实现了并行解码和集合预测，是 DETR 实现端到端目标检测的一大核心。

（4）Prediction Heads

最后的预测头（Prediction Heads）由具有 ReLU 激活函数且具有隐藏层的 3 层线性层计算，对每个查询嵌入进行分类和回归，输出目标类别和边界框位置。其分类分支输出类别的 softmax 概率分布，回归分支则输出归一化边界框坐标。

2.5 本章小结

本章首先详细介绍了红外图像的成像原理和其辐射特性。然后介绍了红外图像的增强算法。接下来，详细介绍了 Transformer 架构的基本原理，特别是其在自然语言处理中的成功应用及其迁移到计算机视觉领域的可行性。结合目标检测

的任务需求，分析了基于 Transformer 的目标检测模型，包括首次使用 Transformer 架构进行图像特征提取和分类的 ViT 模型，以及将 Transformer 与目标检测端到端融合的 DETR 模型，重点讨论了它们的网络结构、工作原理和适用场景，为后续对 RT-DETR 的研究和改进提供了基础。

3 改进 RT-DETR 的红外微小目标检测

本章首先介绍了 RT-DETR 网络模型，然后介绍了改进的 RT-DETR 网络模型。改进思路主要基于两方面，一方面针对小目标检测进行网络增强：通过在 RT-DETR 的骨干网络中添加 EMA 注意力机制模块，采用 Shape IoU 代替 GIOU。另一方面针对红外目标检测进行改进：为 CCFM 引入 CAMixing 卷积-注意力模块，采用 ATFL 损失函数代替原本的 VFL 分类损失。最终形成本文所改进的红外微小目标检测方法。

3.1 RT-DETR 模型

RT-DETR (Real-Time Detection Transformer) 是百度飞桨（PaddlePaddle）团队提出的一种基于 Transformer 架构的目标检测模型，是 DETR 系列的第一款实时目标检测器。其包含四种网络模型，不同网络采用的 Backbone 不同，其中 rtdetr-l 主干网络采用 HGNetv2 作为主干网络，以相对较少的参数及计算量，在同等条件下，表现性能最佳。其网络模型如图 3.1 所示。

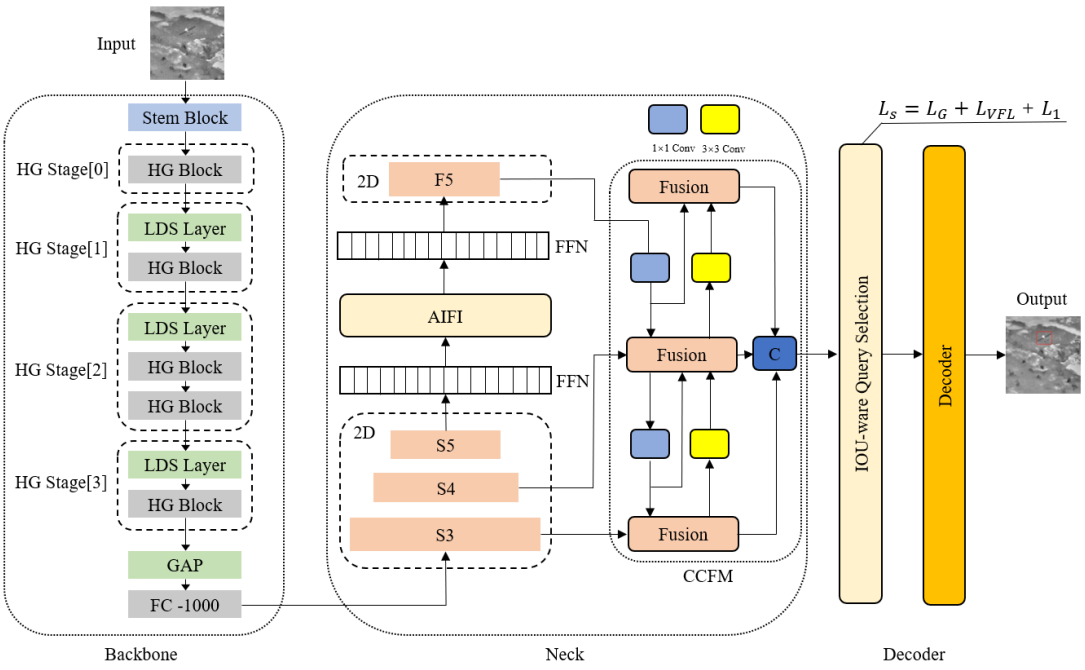


图 3.1 RT-DETR 网络结构图

rtdetr-l 网络结构由特征提取网络(Backbone)、混合编码器(Hybrid Encoder)、IOU-ware Query Selection 和解码器(Decoder)四个部分组成。

3.1.1 主干网络 HGNetv2

HGNetv2 是 HGNet 系列的改进版本，专为 GPU 设备优化设计。HGNet 系列通过尽可能多地采用计算密度最高的 3×3 标准卷积，从而构建一个高效的、适用于 GPU 推理的骨干网络。相较于前代版本，HGNetv2 的结构进一步优化，主要由一个 Stem Block 四个 HG Stage 模块组成，每个 HG Stage 通过大量的标准卷积操作实现多尺度特征的提取，同时避免了复杂的多分支设计。网络在每个阶段逐步增加特征通道数，并通过 LDS（可学习下采样层）对通道进行压缩，以显著降低计算复杂度。HGNetv2 的具体结构如图 3.2 所示。

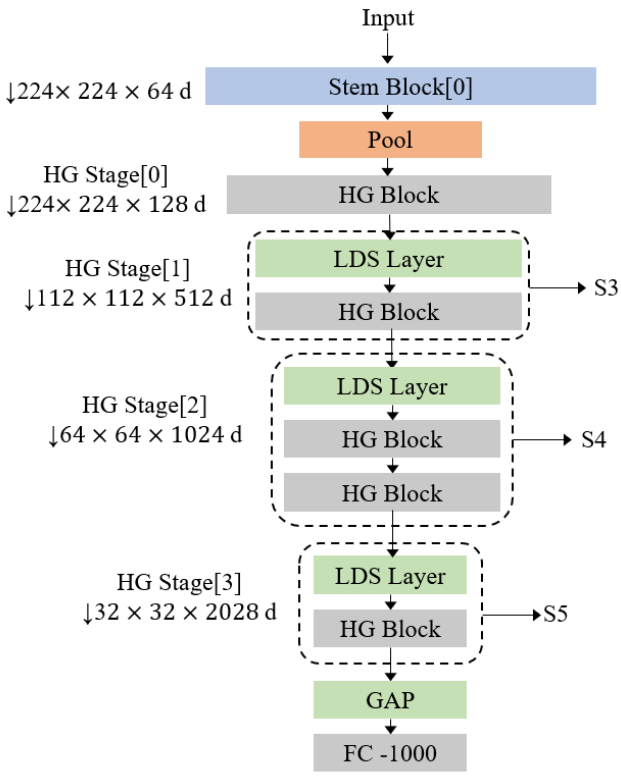


图 3.2 HGNetv2

HGNetv2 的核心结构为 HG Stage 模块，HG Stage 主要 HG Block 完成对特征的提取，其基本结构如图 3.3 所示。每个 HG Block 包括以下三个主要部分：

（1）输入处理模块（ConvBNAct）

ConvBNAct 模块是 HG Block 的基础单元，由两层卷积层（Conv）、批归一化层（BN）和激活函数（ReLU）组成，用于初步提取特征并确保网络训练的稳定性。同时引入 LearnableAffineBlock 模块，当 use_lab 为 True 时，对融合后的最终输出特征进行仿射变换，增强网络的表达能力。仿射变换公式如式 3.1 所示。

$$y = scale \cdot x + bias \tag{3.1}$$

其中，scale 是可学习参数，用于控制特征分流的大小；bias 是偏置参数，用于线性平移特征。

（2）多分支并行 CBA（Conv-BN-Activation）模块：

ConvBNAct 模块的输出被输入到多个并行分支中，分支的数量由网络的层次参数 Layer_num 决定。每个分支通过独立的卷积块提取特定特征。在完成并行 CBA 处理后，各分支的输出特征通过逐元素相加进行特征整合，形成统一的特征表示。这种设计有效提升了网络的特征表达能力，同时保持了结构的简单性。

（3）ESE（Efficient Squeeze-and-Excitation Module）模块

在特征融合后，ESE 模块对整合的特征进行通道注意力增强。ESE 模块是一种轻量化的通道注意力机制，旨在提升特征的表达能力和重要性。通过全局平均池化计算每个通道的全局特征，接着将其输入一个轻量化的线性变换（如单层全连接层），生成通道权重。随后，利用激活函数（ReLU）对权重进行归一化，并将权重重新作用于输入特征，从而调整各通道的特征强度，突出重要特征。ESE 模块以较低的计算成本显著增强了网络对特征的选择性和表达能力。

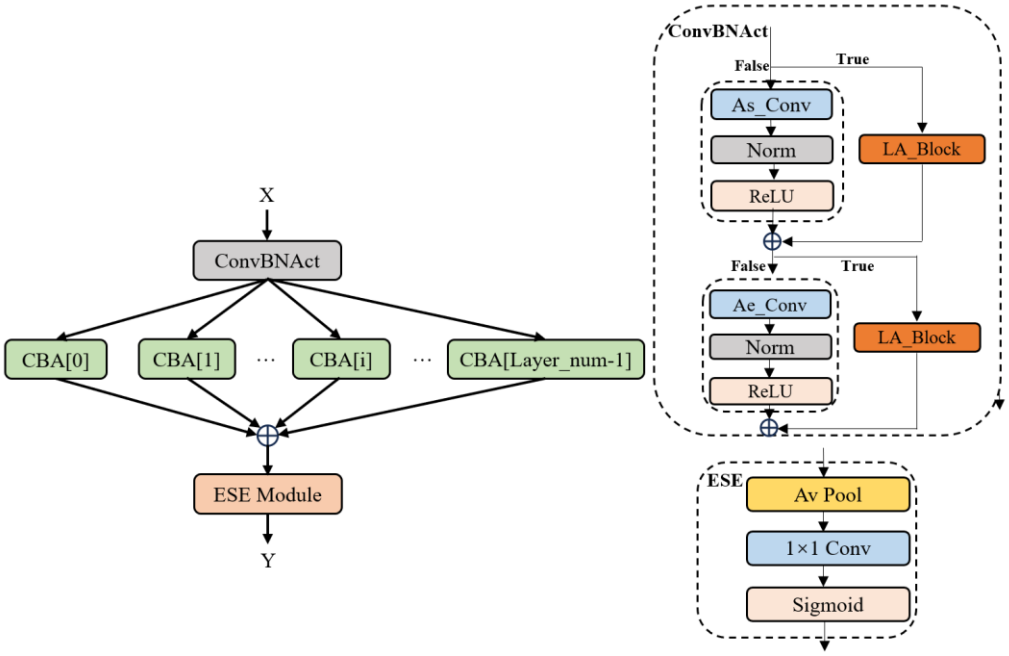


图 3.3 HG Block

HGNetv2 主干网络在不同的层次提取了 3 种尺度的特征图（S3、S4、S5），这些特征图被进一步用作混合编码器的输入，为后续检测任务提供丰富的特征信息。

3.1.2 混合编码器(Efficient Hybrid Encoder)

混合编码器由 AIFI 编码器与 CCFM 特征融合模块两大部分构成。它结合了多头自注意力机制与特征融合网络，旨在通过跨尺度特征融合提升模型的表现。

AIFI 模块结构与 DETR 中的 Encoder 模块相同。其核心思想是通过多头自注意力（multi-head self-attention）和前馈神经网络（FFN）来处理输入的特征，其数学过程如式 3.2、3.3、3.4 所示。

首先，进行输入特征的预处理，将二维输入特征（S_5特征图）拉平成向量，并设定：

$$\mathrm{Q} = \mathrm{K} = \mathrm{V} = Flatten(S_5) \tag{3.2}$$

其次，通过多头自注意力机制，计算出加权后的注意力输出，注并将意力机制的输出特征F_5重新（reshape）以适应后续处理：

$$F_5 = Reshape\big(Attn(\mathrm{Q},\mathrm{K},\mathrm{V})\big) \tag{3.3}$$

最后，输出通过 AIFI 模块的处理后，再与 S_3,S_4,F_5 特征共同输入到 CCFM 模块，进行特征融合。

$$Output = CCFM(S_3, S_4, F_5) \tag{3.4}$$

其中，S_3, S_4为不同尺度的特征，F_5是经过 AIFI 编码器处理后的特征。

CCFM 模块本质上一个 FPN/PAN 结构，用于将不同尺度的特征融合起来，其分别进行两次上特征融合（upsample）和两次下特征融合（downsample），详细结构如图 3.4 所示。

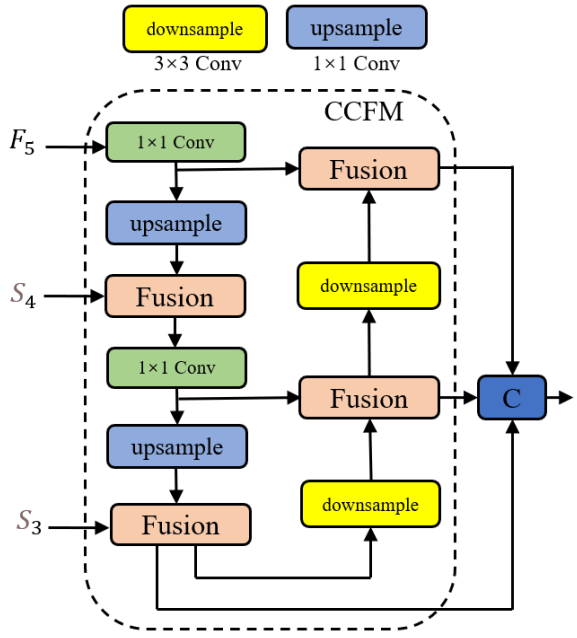


图 3.4 CCFM 模块

CCFM 融合模块的核心为 fusion 模块，其结构如图 3.5 所示，包含两个 1x1 卷积层和多个 RepBlock。每个输入的特征首先通过拼接操作进行合并，然后送入后续的分支处理。其中，RepBlock 是一种重参化卷积模块，它允许网络在训练和推理阶段使用不同的结构。在训练阶段，RepC3 可以表示为一个标准的卷积层，进行常规的卷积计算。但在推理阶段，它可以被重参化为一个更高效的结构，从而减少计算量和提高推理速度。通过调整 CCFM 中 RepBlock 的数量和 Encoder 的编码维度分别控制 Hybrid Encoder 的深度和宽度，同时对 backbone 进行相应的调整即可实现检测器的缩放。在原始网络代码中 RepBlock 数量 N=2。

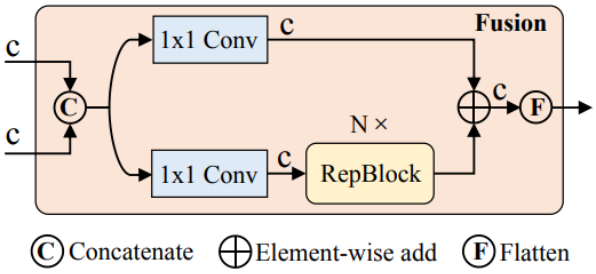


图 3.5 Fusion 模块

3.1.3 IOU-ware Query Selection

IoU-aware Query Selection 的目的是确保位置置信度高的特征能够预测出更准确的类别，从而提高模型的整体性能。该模块的主要作用是从 Encoder 输出的特征序列中选择固定数量的特征作为 object queries，这些查询经过 Decoder 处理后，通过预测头映射为置信度和边界框。模块具体结构如图 3.6 所示。

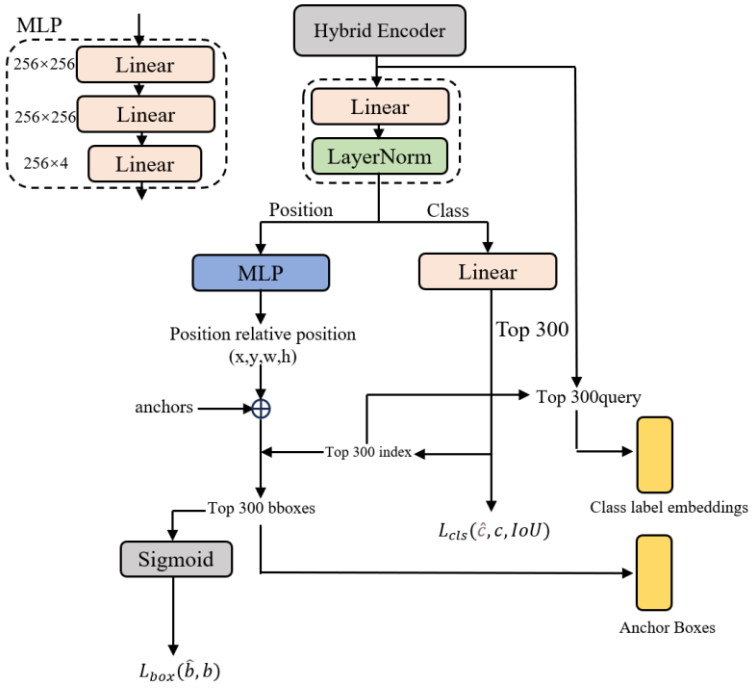


图 3.6 IOU-ware Query selection

在原始的 DETR 中，输入 Decoder 的查询是权重为零的标签查询（label queries），而位置部分的查询嵌入（query embedding）是随机生成的，用于初始化查询特征。然而，这两组信息缺乏有效的先验信息，因此模型收敛较慢。为了解决这个问题，RT-DETR 通过从 Hybrid Encoder 输出的特征中，挑选出表现较好的查询作为 Decoder 的输入，以提高模型的收敛速度。其具体处理流程如下：

首先，将 Hybrid Encoder 输出输入一个位置检测头（position detection head），该头使用 MLP 来预测每个 anchors（先验框）的相对位置。同时结合 anchors 的平铺位置信息，进一步预测出 anchors 的绝对位置。anchors 是由三个不同大小的 feature map 迭代生成的，这样可以覆盖不同大小的目标，确保模型能够识别不同尺度的物体。

另一方面，Hybrid Encoder 输出还会经过一个类别检测头（category detection head），该头是一个全连接层（Linear layer），用于预测每个查询特征的类别信息。每个查询会根据其对应的类别置信度进行排序，从中挑选出 Top-K（300）作为最终的查询目标。

在 Top-K 选择过程中，先对于每一个查询，选择置信度最高的类别作为该查询的预测类别。接着，在所有查询的类别预测中，选择置信度最高的 300 个查询作为最终的 Top-K300 查询。为了保证选择的查询在类别和位置上的一致性，IoU-aware Query Selection 引入了 IoU 作为额外的考量因素。IoU 分数较高的预测框将被赋予更高的类别置信度，而 IoU 分数低的预测框则会具有较低的类别置信度。这样一来，Top-K300 的查询将基于其类别和位置信息计算损失。类别损失不仅仅考虑类别预测的准确性，还会考虑 IoU 来进一步提升类别与位置预测的协调性。这样，IoU 作为一个关键的额外因素，保证了低 IoU 框具有低类别置信度，而高 IoU 框则具有高类别置信度，从而优化了类别和位置的预测一致性。

最终，IoU-aware Query Selection 通过结合类别置信度和位置置信度，利用 IoU 引导查询选择过程，选择出 300 个 query 作为 Decoder 的输入，这样既保证了类别和位置的一致性和准确性，还有效提高了 RT-DETR 模型的收敛速度和检测精度。

3.1.4 解码器(Decoder)

RT-DETR 相较于 DETR 在解码器（Decoder）部分做出了几处关键改进，首先，解码器的输入不再是随机生成的查询，而是由两部分模块的输出构成，一部分是 IoU-aware Query selection 输出的 Top-K300 的 query，其包含 bounding box 类别信息和坐标信息,将这两部分信息分别与 ground truth 加噪（denoising 模块）生成的 query 拼接到一起作为 Decoder 的输入。另一部分则是 Hybrid Encoder 的直接输出。其次，在 Decoder 结构中，RT-DETR 使用了 Multi-scale Deformable-

Attention 结构。其详细结构如图 3.7 所示

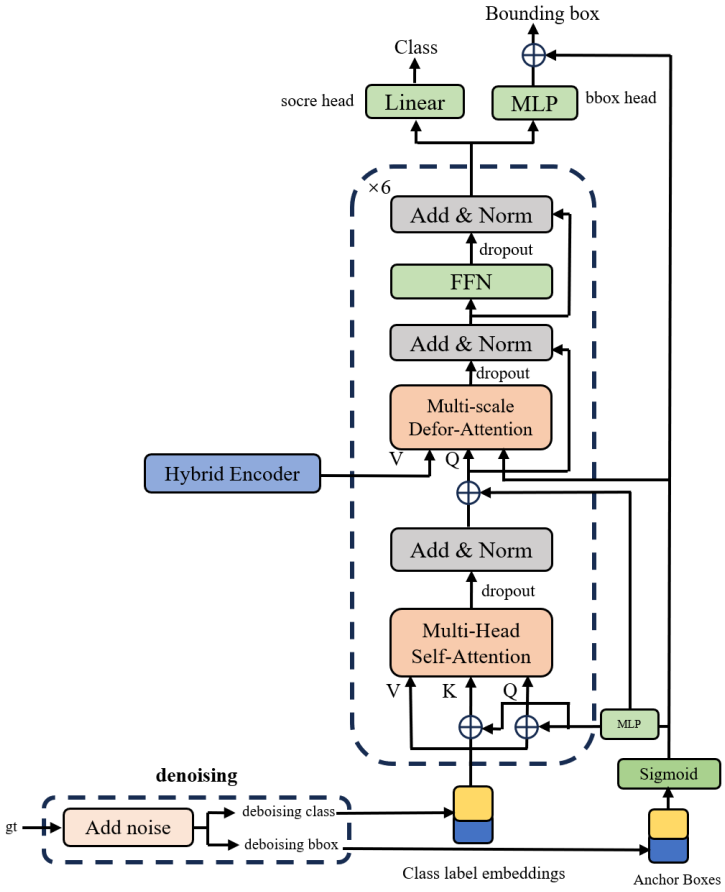


图 3.7 Decoder 结构图

在 RT-DETR 中，Denoising 模块取自 DN-DETR。其核心目标是通过将 Ground Truth(GT)数据加噪（类别噪声和坐标噪声），然后将加噪的数据作为 Decoder 输入，迫使 Decoder 学习去噪。类别噪声和坐标噪声会使得模型的输入在初始时带有一定的不确定性，这样 Decoder 需要学习如何从这些噪声中恢复出真实的目标信息，通过噪声干扰，模型在预测时明确知道自己在预测哪个目标，从而使得预测结果更接近 Ground Truth。

Multi-scale Deformable-Attention 是一种更加灵活的注意力机制，通过动态地选择输入特征中的一部分进行关注，从而减少了传统全局自注意力（global self-attention）的计算复杂度，Multi-scale 的设计则进一步增强了该模块对不同尺度物体 的检测能力。该机制利用多个尺度的特征图来进行多层次注意力计算，从而使得模型能够更加精确地关注到不同大小的目标。

3.1.5 损失函数

通过对 RT-DETR 结构的详细分析，发现 IoU-aware Query Selection 和 Decoder

都涉及位置坐标和类别的预测输出，并且各自都有相应的损失。因此，损失函数可以分为三个部分，第一部分针对 Decoder 最后一层输出的预测结果进行计算，其中包含了类别和位置的预测；第二部分针对 Decoder 中间五层的输出以及 IoU-aware Query Selection 模块输出的 Top-K 查询结果的损失；第三部分则是 Denoising 模块的损失值。通过独立计算每部分的损失，可以更精细地控制每一部分的优化。最终的总损失值由这三部分损失的加权求和得到，并通过反向传播优化网络参数。计算流程如图 3.8 所示。

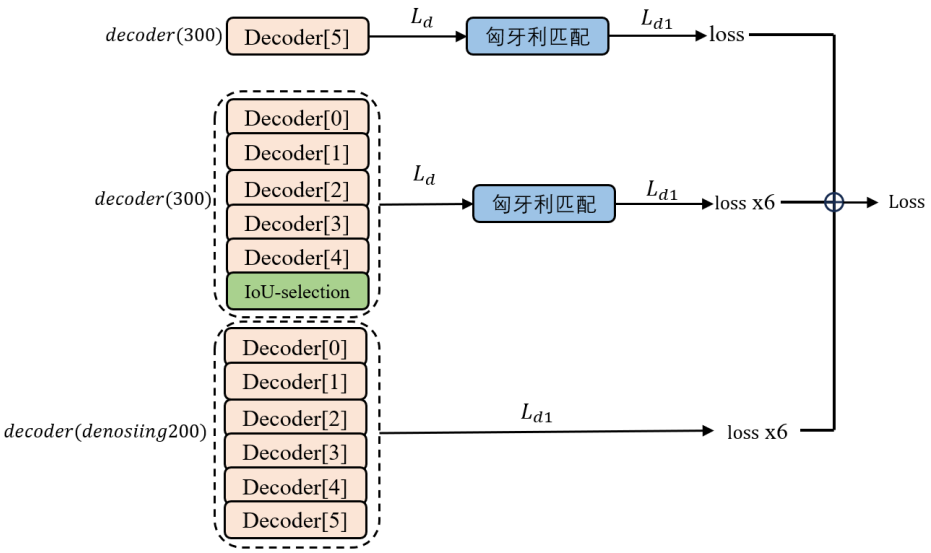


图 3.8 损失计算流程

第一部分和第二部分的损失计算方法相同。首先，对 Decoder 输出的 300 个边界框（bbox）的预测值和真实值之间的损失进行计算。计算公式如式 3.5 所示。

$$L_d = 2 \cdot focal\_Loss + 5 \cdot L1_loss + 2 \cdot GIoU\_Loss \tag{3.5}$$

其中，类别损失采用$focal_loss$损失函数，定位损失则由 GIoU 损失和L_1损失共同构成，对每种损失进行加权求和，得出损失值L_d。

再将得到的损失值L_d进行匈牙利匹配,得出最优解。将最优解作为正样本，其它值作为负样本再次进行加权求和，得到最终的损失值。计算公式如式 3.6 所示。

$$L_{d1} = 1 \cdot varfocal\_Loss + 5 \cdot L1_loss + 2 \cdot GIoU\_Loss \tag{3.6}$$

不同之处在于，varfocal loss 取代了普通的 focal loss，从而使得类别损失的计算更加灵活。

第三部分损失函数的计算与前两部分不同，不采用匈牙利匹配区分正负样本，而是采用 Denoising 模块加噪数据的正负样本直接参与运算，同样利用式 3.7 计算第三部分损失的损失值。最终，将三部分的损失值相加组成总损失值。最终的损失函数公式如式 3.7 所示。

$$L = L_{cls}(\hat{c}, c, GIoU) + L_{box}(\hat{b}, b) + L_1 = VFL + GIoU + L_1 \quad (3.7)$$

其中，分类损失为 VFL，定位损失为 GIoU 和L_1。这种多层次的损失设计不仅能够加速模型收敛，还能提升模型在不同目标尺度上的表现，从而更好地适应复杂的目标检测任务。

3.2 改进 RT-DETR 目标检测模型

针对红外图像目标检测中小目标难以检测的问题，本文改进思路主要基于两方面，一方面针对小目标检测进行网络增强，一方面针对红外目标检测进行改进。

针对小目标检测的改进，通过在 RT-DETR 的骨干网络中添加 EMA 注意力机制模块，使其能提取丰富全局上下文信息和差异化特征的多尺度特征信息，特别是小目标特征信息。同时改进损失函数，对于小目标检测，目标的轻微偏移可能导致其位置信息发生显著变化。GIoU 并没有直接考虑预测框与真实框之间的长宽比差异，无法充分捕捉到位置信息的细微变化。因此，针对微小目标采用 Shape-IoU 代替 GIOU，通过关注边界框本身的形状和尺度来计算损失，优化小目标检测精度。

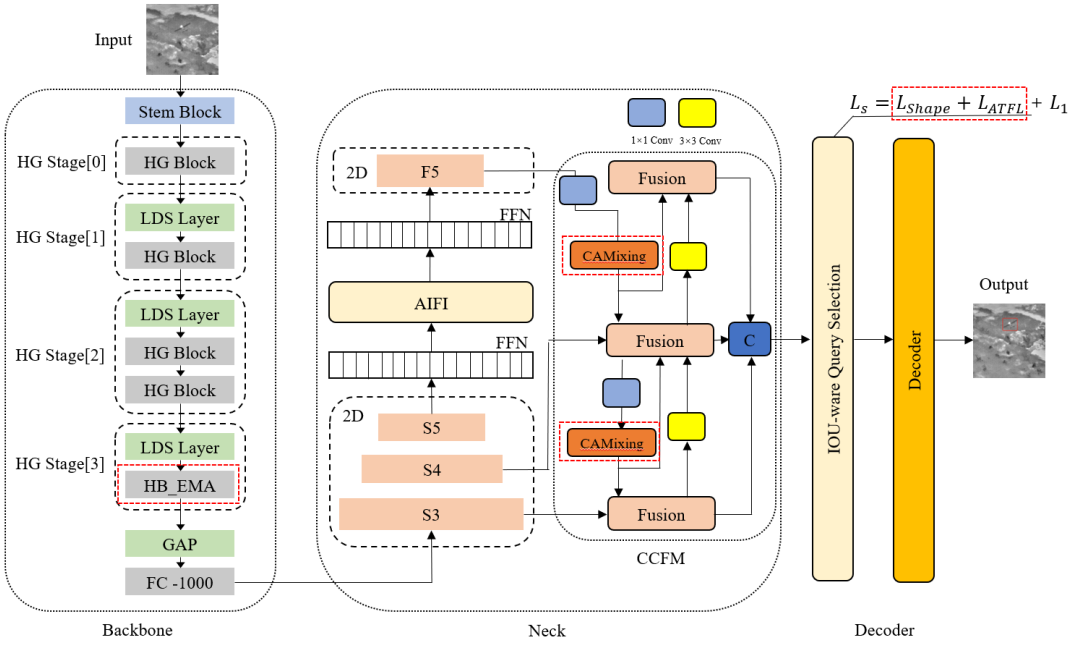


图 3.9 改进后的 RT-DETR 网络结构图

针对红外目标检测的改进，为了抑制红外图像中背景和噪声等信息的干扰，提高小目标的关注度。引入 CAMixing 卷积-注意力模块使得 CCFM 在多尺度特征融合时，能抑制无关信息的干扰，提升去噪性能，有利于增强全局和局部特征的建模，同时提高小目标的检测率。而在红外小目标图像中，由于图像主要由背

景组成，只有小部分被目标占据。在训练过程中学习背景的特征比学习目标的特征更容易。因此采用 ATFL 损失函数代替原本的 VFL 分类损失，该函数将目标和背景解耦，并利用自适应机制来调整损失权重，迫使模型将更多的注意力分配给小目标特征。改进后的 RT-DETR 网络结构如图 3.9 所示。在 rt-detr-l 的主干网络 HGNetv2 中添加 EMA 模块，在 CCFM 中添加 CAMixing 模块，并改进 IOU-ware Query Selection 的锚框筛选函数，最终形成本文所改进的红外微小目标检测方法。

3.2.1 EMA 模块

在进行红外小目标检测时，由于目标在图像中占的像素少，且目标信号弱，背景噪声强，基于 CNN 的骨干网络在特征提取时，在深层次采样过程中容易丢失小目标信息。因此，在骨干网络中加入注意力机制，在特征提取时保持分辨率或采用多尺度特征金字塔结构是目前常用的解决方法。

高效多尺度注意力（EMA）是一种专门针对小目标检测优化的注意力机制，其通过多尺度特征提取、跨空间学习和通道混合卷积相结合，有效增强目标区域特征表达，抑制背景干扰，同时提高计算效率，将其集成到 RT-DETR 骨干网络中，可以有效的提高红外小目标检测的精度。

EMA 的具体网络结构如图 3.10 所示，主要由三个主要部分组成。

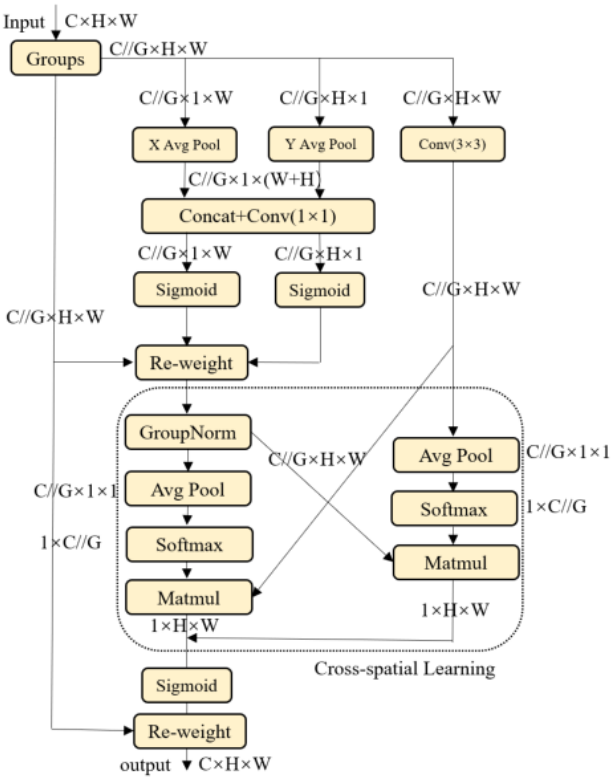


图 3.10 EMA 模块结构图

首先，EMA 采用多尺度卷积，通过不同大小的感受野同时提取目标的局部和全局特征。对于任意给定的输入特征图$X \in R^{C\times H\times W}$， EMA 通过三条平行路线来提取分组特征图的注意力权重描述符。其中，两条平行路径F_1，F_2在 1x1 分支上进行通道变换，其分别在水平、垂直两个空间方向对通道进行一维平均池化，以保持局部特征，增强通道信息。第三条路径F_3在 3x3 分支上，增加感受野，获取全局信息。最终，多尺度特征通过 1x1 分支通道维度拼接，并采用 sigmoid 函数激活，使其符合二维二项分布。多尺度卷积过程如式 3.8、3.9、3.10 所示。

$$F_1 = Sigmoid(Avg1D(Conv_{1\times1}(X_{1\times W}))) \tag{3.8}$$

$$F_2 = Sigmoid(Avg1D(Conv_{1\times1}(X_{H\times1}))) \tag{3.9}$$

$$F_3 = Conv_{3\times3}(X_{H\times W}) \tag{3.10}$$

其次， EMA 采用自适应注意力加权（Re-weight）机制，为不同尺度和空间的特征分配动态权重，使得网络在推理过程中可以自动调整对不同特征的关注度。采用 Softmax 归一化，计算每个分支的权重α_i，确保权重总和为 1，提高稳定性，避免数值溢出问题。其计算过程如式 3.11、3.12 所示。

$$F_{final} = \sum_{i=1}^{n} \alpha_i F_i \tag{3.11}$$

其中，α_i由 softmax 归一化计算：

$$\alpha_i = \frac{e^{w_i}}{\Sigma_{u=1}^{n} e^{w_j}} \tag{3.12}$$

其中，w_i，w_j为可学习参数，用于调整不同路径的权重。该机制确保网络在不同任务情况下能够自动调整对特征的关注度，使网络能根据不同输入情况，调整局部和全局特征的比重，从而提高检测效果。

最终，将三条路径提取的特征进行跨空间学习（Cross-Spatial Learning）。小目标检测的关键挑战之一是目标区域的特征容易被背景淹没，导致检测精度下降。因此，EMA 采用跨空间学习机制，使远离目标的区域能够提供更多语义信息。其利用非局部注意力或交叉通道混合卷积，计算特征图上不同位置之间的相关性并特征更新，通过跨空间特征融合，使得不同区域的信息能够交互，提高目标区域的分辨能力。计算过程如式 3.13、3.14 所示。

$$A(i,j) = Softmax(QK^T) \tag{3.13}$$

$$F_{CS} = AV \tag{3.14}$$

其中，$Q = W_Q F_{MS}$，$K = W_K F_{MS}$，$V = W_V F_{MS}$，$A(i,j)$表示第i个像素点与第j个像素点之间的相似度。

3.2.2 CAMixing 模块

CAMixing 模块源于 HSI 去噪的混合卷积和注意网络(HCANet)，是一个

transformer 与 CNN 的混合模型。CAMixing 主要由 CAFM、MSFN 两部分组成，为了增强全局和局部特征的建模，其设计了一个卷积和注意力融合模块(CAFM)，旨在捕获远程依赖关系和邻域光谱相关性。为了改善 FFN 的多尺度信息聚合，其设计了一个多尺度前馈网络(MSFN)，在 MSFN 中使用了三个不同步长的平行扩展卷积，通过提取不同尺度的特征来增强去噪性能。模块结构如图 3.11 所示。

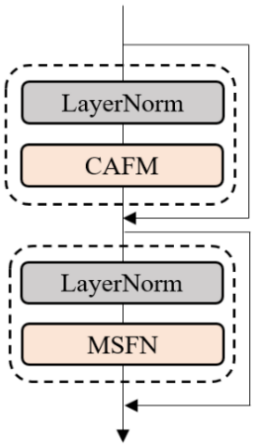


图 3.11 CAMixing 模块

（1）CAFM 模块

卷积运算受局部性质和感官场的限制，在建模全局特征方面存在不足。而由注意力机制支持的 Transformer 擅长提取全局特征和捕获远程依赖关系。因此卷积和注意力融合模块(CAFM)采用分支结构，一个分支中采用自关注机制来捕获捕获远程依赖关系，另一个分支利用 CNN 进行局部特征提取。CAFM 模块结构如图 3.12 所示。

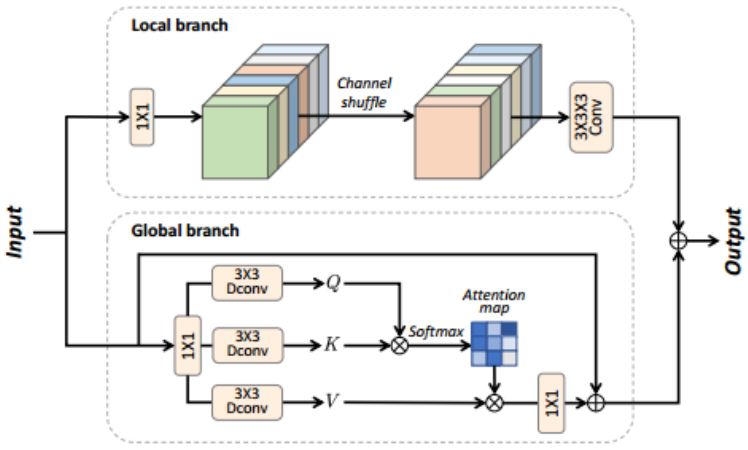


图 3.12 CAFM 模块

在局部分支中，为了加强多尺度融合，首先使用 1×1 卷积对输入特征 Y 进行通道调整。以适应后续的通道划分操作。对调整好的通道按通道维度划分为多个组，每组进行单独的处理，组内采用深度可分卷积来融合通道信息。经过通道融合后，各组的输出沿通道维度进行拼接，并通过 3×3×3 卷积提取局部特征。局部分支结构的表达式如式 3.15 所示。

$$\mathrm{F_{conv}} = W_{3\times3\times3}\left(\mathrm{CS}\left(W_{1\times1}(\mathrm{Y})\right)\right) \tag{3.15}$$

式中，$\mathrm{F_{conv}}$为局部支路输出，$W_{1\times1}$为 1×1 卷积，$W_{3\times3\times3}$为 3×3×3 卷积，CS 代表通道划分操作。

在全局分支中，首先通过 1×1 卷积和 3×3 深度卷积生成 Q、K、V 三个键值，得到形状为 H×W×C 的特征张量。接下来，将 Q 重构为: $\widehat{\mathbf{Q}} \in \mathrm{R}^{\hat{H}\hat{W}\times\hat{C}}$, K 重构为:$\widehat{K} \in \mathrm{R}^{\hat{C}\times\hat{H}\hat{W}}$。然后，我们通过$\widehat{\mathbf{Q}}$和$\widehat{K}$的相互作用计算注意力图 A∈$\mathrm{R}^{\hat{C}\times\hat{C}}$。减少了计算负担，而不是计算大小为$\mathrm{R}^{\hat{H}\hat{W}\times\hat{H}\hat{W}}$的巨大规则注意力图。全局分支的输出 Fatt 定义为式 3.16、3.17 所示。

$$\mathrm{F_{att}} = W_{1\times1}\mathrm{Attention}\left(\hat{\mathrm{Q}},\hat{\mathrm{K}},\hat{\mathrm{V}}\right) + Y \tag{3.16}$$

$$Attention\left(\hat{\mathrm{Q}},\hat{\mathrm{K}},\hat{\mathrm{V}}\right) = \hat{\mathrm{V}}Softmax\left(\hat{\mathrm{K}}\hat{\mathrm{Q}}/\alpha\right) \tag{3.17}$$

其中 α 是一个可学习的缩放参数，用于在控制矩阵乘法的尺度。最终，CAFM 模块计算输出计算为式 3.18 所示:

$$\mathrm{F_{out}} = \mathrm{F_{att}} + \mathrm{F_{conv}} \tag{3.18}$$

这样，局部分支提取局部细节信息，而全局分支关注远程依赖关系，两者互补，提高去噪效果。

（2）MSFN 模块

MSFN 主要用于增强非线性特征变换，并解决 FFN 单尺度特征聚合的局限性。该模块的特点是引入门控机制，并结合多尺度扩张卷积来增强特征提取能力。MSFN 结构如图 3.13 所示。

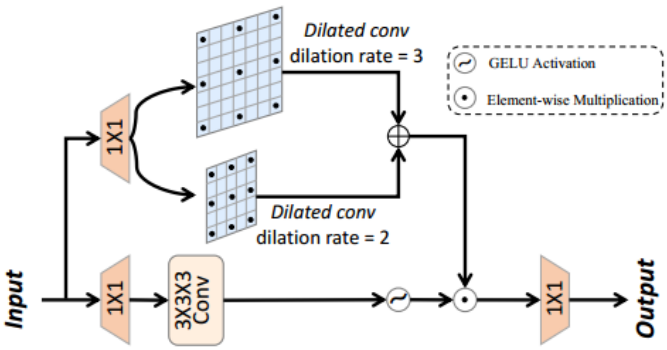


图 3.13 MSFN 结构图

在每个混合块之后，将 CAFM 的输出输入到 MSFN 中，以聚合多尺度特征，增强非线性信息转换。MSFN 使用两个 1×1 卷积扩展特征通道，扩展比为 $\gamma=2$。输入特征在两个并行路径上处理，引入门控机制，通过两个路径特征的元素积来增强非线性转换。在下主路中，使用深度卷积进行特征提取。

在上支路，采用多尺度扩展卷积进行多尺度特征提取。使用两个 3×3 的扩张卷积，扩张率分别为 2 和 3。给定一个输入张量 $X\in R^{\hat{H}\times\hat{W}\times\hat{C}}$，MSFN 表达式为 3.19、3.20 所示。

$$\begin{aligned}\mathrm{Gating}(X)=&\phi\left(W_{3\times3\times3}W_{1\times1}(X)\right)\\&\odot\left(W_{3\times3}^{2}W_{1\times1}(X)+W_{3\times3}^{3}W_{1\times1}(X)\right)\end{aligned}\tag{3.19}$$

$$X_{out}=W_{1\times1}\mathrm{Gating}(X)\tag{3.20}$$

其中⊙表示单元乘法，φ 表示 GELU 非线性，$W_{3\times3}^{2}$表示 3×3 膨胀卷积，膨胀率为 2, $W_{3\times3}^{3}$表示 3×3 膨胀卷积，膨胀率为 3。与 CAFM 相比，MSFN 专注于提取上下文信息。

RT-DETR 中将编码器仅用于 S5，采用对比实验验证了不仅有助于显著减少计算量和提高计算速度，同时不会对模型的性能造成明显的损害。因此将 CAMixing 添加进作于与 S5 的两次向下的特征融合，有助于提升检测精度的同时减少计算量。

3.2.3 Shape-IoU 损失函数

原本的损失函数采用 GIoU 作为 bbox 回归损失函数，其引入了外接矩形来考虑预测框和真实框之间的空间关系，旨在解决传统 IoU 在框之间没有交集的情况下无法有效衡量的缺陷。但是红外小目标对 IoU 度量高度敏感，由于目标的低信噪比、低对比度甚至目标边界模糊的影响，即便是很小的框位移或尺度变化，都可能导致 IoU 值的显著变化，从而影响目标检测精度。GIoU 在计算过程中没有明确考虑预测框和真实框之间的长宽比差异，因此在目标的形状（特别是长条形目标）差异较大的情况下，无法准确反映目标形状的变化。

现有的边框回归方法通常关注 GT 框与预测框之间的几何关系，如相对位置和相对形状，并据此计算损失。然而，这些方法往往忽略了边框自身的形状和尺度等固有属性对回归过程的影响。如图 3.14，展示了边框回归过程中 deviation 和 shape-deviation 对 IoU 变化的影响。

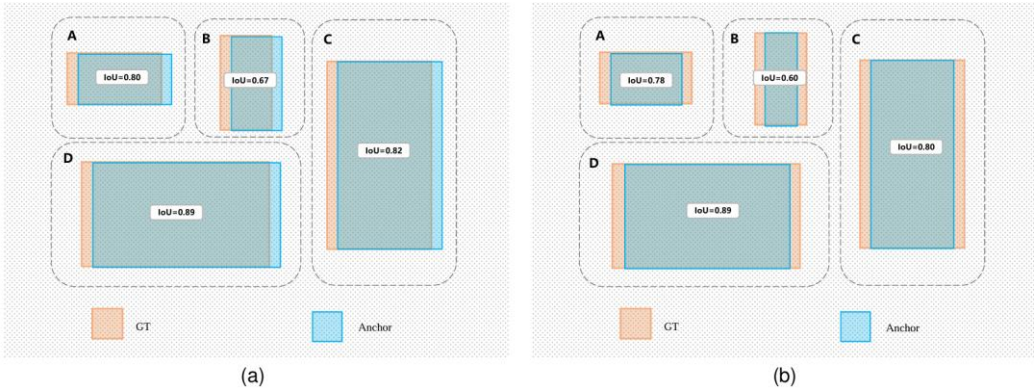


图 3.14 回归损失对比图

其中，图（a）中，A 与 B 的 GT 框尺度相同，但形状不同，deviation 方向分别沿长边和短边，由于短边方向的 deviation 对 IoU 影响更大，B 的 IoU 变化比 A 更显著，C 与 D 的 GT 框尺度较大，相较于 A 与 B，IoU 变化较小。图（b）中，A 与 B（C 与 D）shape-deviation 相同，但 IoU 仍存在差异。因此，GT 框的尺度和形状影响 IoU 变化，小尺度边框更敏感，形状影响更显著。为了提高小目标检测的精度，针对微小飞机目标采用 Shape-IoU 代替 GIOU，通过关注边界框本身的形状和尺度来计算损失，优化小目标检测精度。

Shape-IoU 所计算回归损失L_S如式 3.21 所示。

$$L_S = 1 - IOU + D^S + 0.5\Omega^S \tag{3.21}$$

其中，IOU 为交并比，S 为尺度因子，与数据集中目标的尺度有关，D^S为距离损失，其计算公式如式 3.22、3.23、3.24 所示。

$$ww = \frac{2 \times (w^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \tag{3.22}$$

$$hh = \frac{2 \times (h^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \tag{3.23}$$

$$D^S = hh \times \frac{\left(x_c - x_c^{gt}\right)^2}{c^2} + ww \times \frac{\left(y_c - y_c^{gt}\right)^2}{c^2} \tag{3.24}$$

其中，w^{gt}，h^{gt}是 GT 框的宽高，x_c，y_c，x_c^{gt}，y_c^{gt}作为先验框的与 GT 框的中心点坐标。通过坐标计算出ww与hh作为水平方向和垂直方向的权重系数，调整 s 的取值即可调整D^S的损失大小。

Ω^S为形状损失。其沿用了 SIOU 的计算公式 3.25。

$$\Omega^S = \sum_{t=w,h} (1 - \mathrm{e}^{-w_t})^{\theta} \tag{3.25}$$

其中，w,h分别为先验框的宽和高，θ控制对形状损失的关注程度，为了避免过于关注形状损失而降低对预测框的移动，其取值通过遗传算法确定为 4。

3.2.4 ATFL

ATFL 源于增强特征学习网络 EFLNet，对于红外小目标图像，其主要由背景组成，只有小部分被目标占据。在训练过程中学习背景的特征比学习目标的特征更容易。因此采用 ATFL 损失函数代替原本的 VFL 分类损失，该函数利用阈值p_t设置将易识别的背景与难识别的目标解耦;通过强化与目标相关的损失，减轻与背景相关的损失，迫使模型将更多的注意力分配到目标特征上，从而缓解目标与背景之间的不平衡。最后，将自适应设计应用于超参数，以减少调整超参数所带来的时间消耗。

BCE 损失函数用于衡量模型的预测概率与真实标签之间的差异，能很好地度量概率预测的准确性，并促使模型对正负样本做出区分。计算公式如式 3.26 所示。

$$\mathcal{L}_{\mathrm{BCE}} = -(y\log(p) + (1-y)\log(1-p)) \tag{3.26}$$

其中，p表示预测概率，y 表示真实标签,当真实标签为 1（正样本）时，BCE 公式简化为式 3.27 所示。

$$\mathcal{L}_{\mathrm{BCE}} = -\log(p_t) \tag{3.27}$$

其中，p_t表示当前平均预测概率值，取值如式 3.28 所示。

$$p_t = \begin{cases} p, y = 1 \\ 1-p, \text{others} \end{cases} \tag{3.28}$$

对于当真实标签为 1（正样本）时，BCE 损失函数中的-ylog(p)部分会起主要作用。为了最小化损失，模型需要最大化p（即预测为正样本的概率），因为当p趋近于 1 时，- log (p)趋近于 0。当真实标签为 0（负样本）时，BCE 损失函数中的-$(1-y)$log $(1-p)$部分会起主要作用。为了最小化损失，模型需要最大化 1-p（即预测为负样本的概率，或者等价的最小化p，因为当p趋近于 0 时，- log $(1-p)$趋近于 0。但是 BCE 无法解决正负样本之间的不平衡问题。

因此交叉熵损失函数（FL）引入了一个调制因子$(1-p_{\mathrm{t}})^{\gamma}$，通过调整聚焦参数 γ 来降低易于分类的样本的损失贡献。FL 公式如式（3.29）所示。

$$FL(p_{\mathrm{t}}) = (1-p_{\mathrm{t}})^{\gamma}\mathcal{L}_{\mathrm{BCE}} \tag{3.29}$$

焦损失函数可以调节 γ 的值，以降低易试样的损失权重。但是，调制因子在降低易样本损失的同时，也降低了难样本损失的值，不利于难样本的学习。

因此，ATFL 通过减少简单样本的损失权重，有效地减轻了简单样本的影响，同时增加了分配给困难样本的损失权重。其表达式如式 3.30 所示。

$$\begin{cases} -(\lambda - p_t)^{-\ln(p_t)}\log(p_t) & p_t \le 0.5 \\ -(1-p_t)^{-\ln(\hat{p}_c)}\log(p_t) & p_t > 0.5 \end{cases} \tag{3.30}$$

其中p_t表示当前平均预测概率值，$\hat{p}_c$表示下一个 epoch 的预测值，λ(>1)为超参数，在本次实验中λ =3.5。将 0.5 概率以上的样本认为是易识别样本，0.5 以下的样本值是困难样本，ATFL 通过改进 FL 中的自适应因子γ为$-\ln(p_t)$来平衡样本的权重，增加小目标的贡献，同时减少调整超参数所带来的时间消耗。

3.3 本章小结

本章首先详细介绍了 RT-DETR 网络模型，然后针对目前红外微小目标检测的难点设计一种基于改进 ER-DETR 的红外微小目标检测算法。

一方面，针对小目标检测的难点进行网络增强。在 RT-DETR 的骨干网络中引入 EMA 注意力机制模块，通过捕获更丰富的上下文信息来增强对微小目标的特征提取能力。将 Shape IoU 替代传统的 GIoU 作为回归指标，聚焦目标边界框的固有属性，从而提升对微小目标的检测精度。

另一方面，结合红外目标检测的特性，对网络结构进行优化。在 CCFM 模块中引入 CAMixing 卷积-注意力模块，利用多尺度卷积自注意力机制更好地捕获红外图像的局部信息和多尺度特征。此外，将 ATFL 损失函数替代原有的 VFL 分类损失，进一步提升分类性能和小目标检测的鲁棒性。

通过上述改进，最终形成了一种针对红外微小目标检测的高效方法，为复杂场景下的红外目标检测提供了技术支持。

4 轻量化红外微小目标检测方法

在红外微小目标检测领域，传统检测方法依赖于高分辨率成像和复杂的计算模型，但计算资源需求高，难以满足实时性要求。因此，如何在保证检测精度的同时减少计算量，实现轻量化优化，成为当前研究的关键问题。目前，轻量级深度学习模型通过减少参数量和计算复杂度，提高推理速度并降低存储需求，使其适用于嵌入式设备、移动端和边缘计算场景。此外，结合 CNN 与 Transformer 结构的轻量化设计，能兼顾全局信息建模与局部特征提取，广泛应用于目标检测、语音识别、图像分类等任务，成为实时智能应用的理想选择。

4.1 MobileNetv4 轻量化网络

MobileNet 是 Google 研究团队开发的一种轻量级深度卷积神经网络架构，专为移动设备和资源受限环境设计，具有模型体积小、计算效率高和可调整性强的特点。其核心创新在于引入深度可分离卷积，将标准卷积拆分为深度卷积和逐点卷积两个步骤。可显著减少参数量和计算复杂度，缩小模型体积，提高推理速度。

MobileNetv4 作为 MobileNet 系列最优的网络结构，相比于 MobileNetv2，是一种专为移动设备设计的高效神经网络架构，通过引入统一灵活的通用倒置瓶颈（UIB）搜索块和移动版多头注意力（Mobile MQA）模块，并改进神经架构搜索（NAS）配方，实现了在多种硬件平台上的优异性能。

（1）深度可分离卷积与轻量化设计

MobileNetv4 保持了 MobileNet 系列的核心设计思想—深度可分离卷积，结构如图 4.1 所示。在深度可分离卷积中，传统卷积操作被拆分为逐通道卷积和逐点卷积，在减少计算量的同时又保持了其特征提取能力。MobileNetv4 进一步优化了其结构，特别是在倒置瓶颈（IB）模块和注意力机制上的改进，使得网络在计算效率上得到了进一步提升。

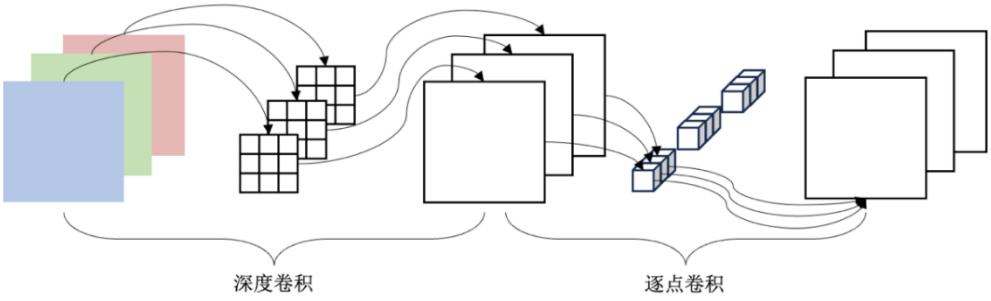


图 4.1 深度可分离卷积

（2）通用倒置瓶颈（UIB）模块

MobileNetv4 引入了通用倒置瓶颈（UIB）模块，这是基于 MobileNetv2 的倒置瓶颈（IB）模块的扩展。UIB 模块结合了 ConvNext、前馈网络（FFN）和额外深度卷积（ExtraDW）变体，为网络架构提供了更高的灵活性。与 MobileNetv3 的倒置瓶颈相比，UIB 模块能够在不显著增加计算开销的情况下，提供更大的模型容量和更强的特征表示能力。其具体结构如图 4.2 所示。

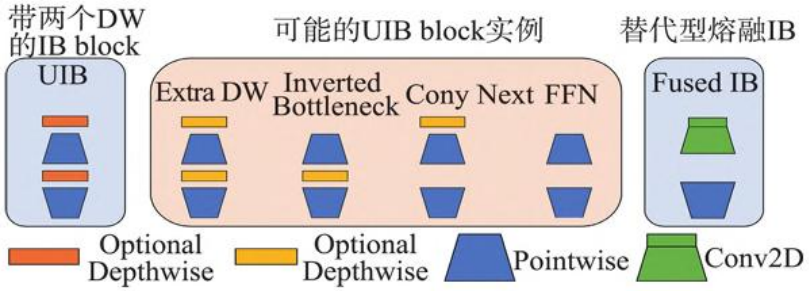


图 4.2 UIB 模块

UIB 块中的两个可选深度卷积有四种可能的实例化，导致不同的权衡。倒置瓶颈（IB）在扩展特征激活上执行空间混合，提供更大的模型容量，但成本增加。ConvNext 允许通过在扩展之前执行空间混合以较大的内核大小实现更便宜的空间混合。ExtraDW 是 MobileNetv4 引入的一种新变体，可实现网络深度和感受野的廉价增加，提供了 ConvNext 和 IB 的综合优势。FFN 是两个 1×逐点卷积（PW）的堆叠，中间带有激活和归一化层。PW 是最适合加速器的操作之一，但在与其他块一起使用时效果最佳。这些深度卷积层的引入与否由神经架构搜索（NAS）优化过程自动决定，进而生成了一系列新颖且高效的架构。

（3）移动版多头注意力（Mobile MQA）模块

传统的 MHSA 通过对查询（query）、键（key）和值（value）分别进行投影，以捕获不同信息维度的多头自注意力，但这导致了大量的内存访问开销。Mobile MQA 是一种专为移动加速器设计的新型注意力模块。其通过在混合视觉模型中采用多查询注意力（MQA），共享所有头部的键（keys）和值（values），从而减少了数据访问需求，同时保留多个查询头部以捕获信息的多样性。Mobile MQA 计算过程如式 4.1、4.2 所示。

$$MQA(X) = Concat(attention_1, \ldots, attention_n)W^O \tag{4.1}$$

$$attention_j = softmax\left(\frac{(xw^{Q_j})(SR(X)W^K)^T}{\sqrt{d_k}}\right)(SR(X)W^V) \tag{4.2}$$

其中 SR 表示空间缩减，即步幅为 2 的 DW，或者在不使用空间缩减的情况

下表示恒等函数。

（4）神经架构搜索（NAS）的改进

MobileNetv4 在神经架构搜索（NAS）方面进行了优化，进一步提高了架构的灵活性和自动化程度。与 MobileNetv3 中的传统 NAS 方法不同，MobileNetv4 的 NAS 过程更加注重架构的高效性和硬件适应性。通过自动化调整网络的复杂度，MobileNetv4 能够生成适合不同硬件平台和应用需求的最佳架构，从而实现跨平台的优异性能。

4.2 StarNet 网络结构

StarNet 是由 Microsoft 研究人员提出的轻量、高效的神经网络架构，其核心是使用星操作（star operation）机制，采用逐元素乘法作为主分支与跳跃连接之间的融合方式，替代传统的加性残差连接，在保持网络紧凑结构的前提下，实现更高的特征表示能力，且堆叠层数越多，提升效果越明显。

StarNet 的整体架构整体网络结构与传统 CNN 架构类似，由输入 Stem、多个阶段（Stage）以及最终的分类 Head 组成。其整体网络结构如图 4.3 所示。

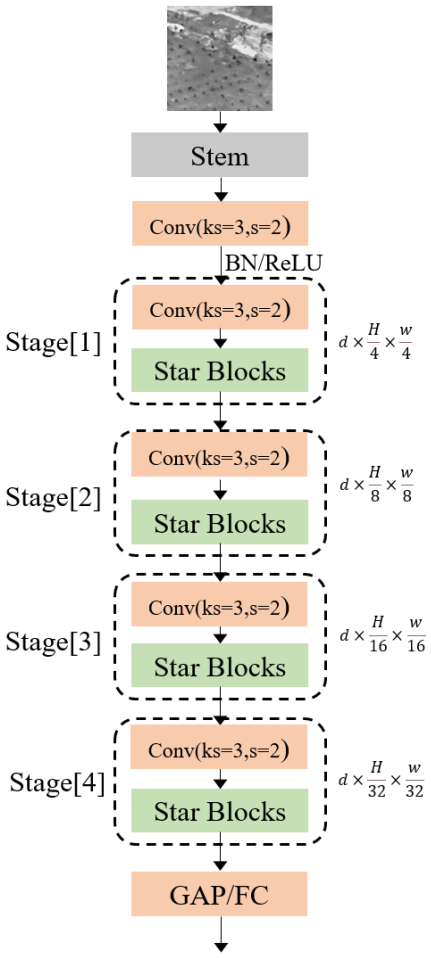


图 4.3 starnet 网络结构

输入图像首先经过一个带有步长为 2 的 3×3 卷积层进行初步特征提取与下采样，随后依次通过四个阶段，每个阶段内部由多个 Star Block 构成，并在阶段之间通过 stride=2 的卷积实现空间尺寸的减半。在最后阶段之后，采用全局平均池化（GAP）和全连接层进行分类。Star Block 网络结构如图 4.4 所示。

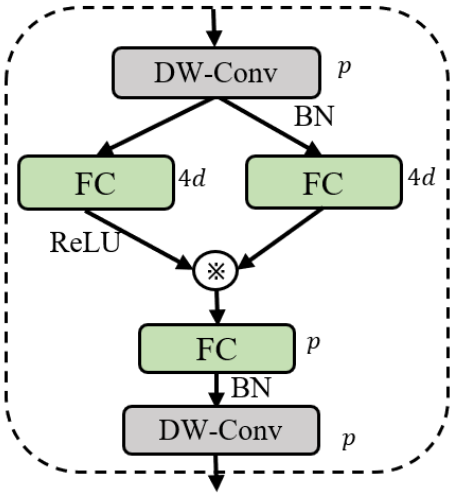


图 4.4 Star Block

其结构包含三个卷积层与一个乘法残差连接， 第一层 1×1 卷积用于通道压缩，降低计算成本，同时配合批归一化（BN）和 ReLU 激活函数提高非线性表达能力，第二层 3×3 卷积提取局部空间特征，第三层 1×1 卷积用于恢复通道维度，输出后再次接 BN，但不加激活，而残差连接部分直接将输入传递，与主分支输出进行逐元素乘法融合。最终输出如式 4.3 所示。

$$y = f_3\left(f_2\left(f_1(x)\right)\right) \odot x \tag{4.3}$$

其中，f_1，f_2，f_3分别表示三个卷积模块组成的函数，$\odot$表示逐元素乘法操作。与传统残差模块使用加法 y=f(x)+x 相比，乘法融合机制可以捕捉更复杂的特征组合关系，增强网络的表示能力，并在不引入额外参数的情况下实现非线性增强。与多数注意力机制（如 SE、CBAM）相比，乘法残差结构则无需显式引入注意力模块，即可隐式实现通过学习一个“重要性权重”对特征进行逐元素缩放，因而具备更高的结构简洁性与计算效率。

4.3 RT-DETR 网络模型的轻量化

在 RT-DETR 中，网络的计算量主要集中在特征提取阶段的 Backbone 和颈部的特征金字塔融合模块。特征提取的 Backbone 是网络计算量的主要来源，其复杂度取决于卷积操作的数量、特征图分辨率以及通道数，而特征金字塔融合模块则通过整合不同尺度的特征进一步增加了计算开销。

为了降低整体计算量，通过将 Backbone 替换为轻量化网络结构并在特征融合阶段中引入高效的注意力模块，以减少参数量和计算复杂度，能够在保证精度的同时，进一步降低计算资源的占用，从而提高在资源受限环境中的应用性能。

因此，本章将 MobileNetv4 模块融合到主干网络之中，并探究合适的注意力模块替换 fusion 模块中的 repc3 结构，以达到红外微小目标检测时所需的较高的精确度和实时检测的要求。

4.3.1 骨干网络替换

（1）MobileNetV4 骨干替换

原算法 rtdetr-l 采用 HGnetV2 作为特征提取主干网络，网络层次较深，计算量较大。采用 MobileNetV4 网络作为改进 RT-DETR 的特征提取主干网络，降低算法参数量，提升算法计算效率。在更换骨干网络的同时保留 RT-DETR 本身的 CCFM 的输入结构，新的网络结构如图 4.5 所示。

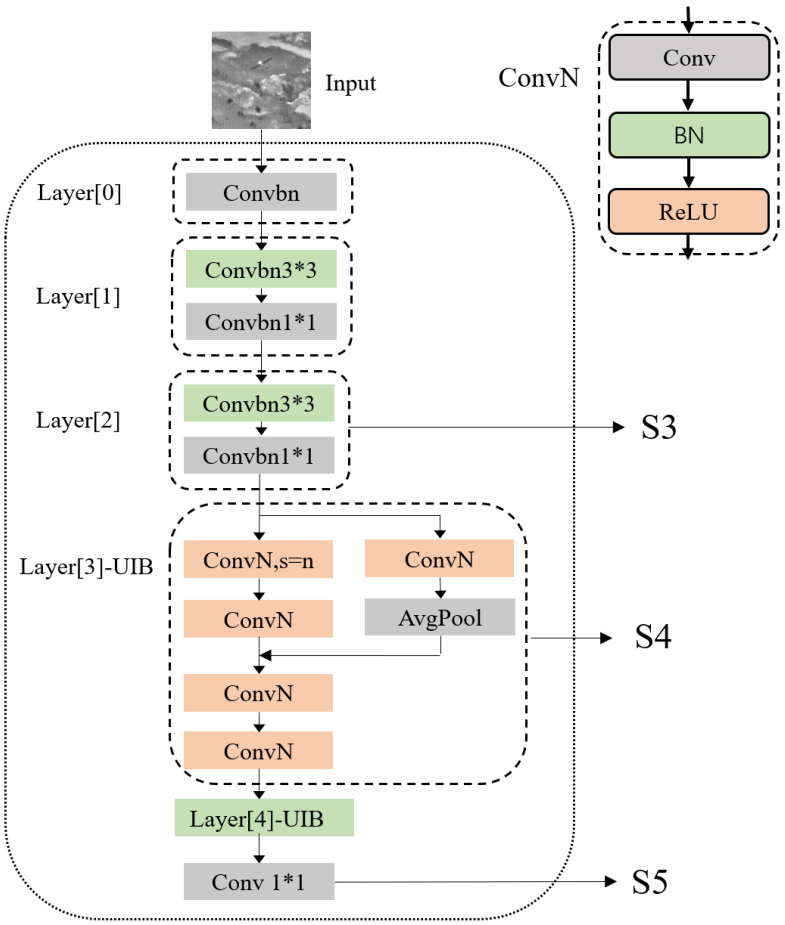


图 4.5 MobileNetv4_RT-DETR 网络结构图

MobileNetV4 核心是引入了通用反转瓶颈（UIB）搜索模块，由 UIB 模块统一反转瓶颈（IB）、ConvNext、视觉变换器（VIT）的前馈网络（FFN），并引入

了额外深 IB (ExtraDW)块，集成上述模块的优势，以较少的计算代价增加网络的深度和感受野并最大化计算利用率，使其对具有轻量化需求的识别任务更具备优势，同时保持原有的 CCFM 结构，从三个尺度分别提取图形特征，在定义输入图像大小为(640，640)情况下，本文所用的 MobileNetV4 特征提取主干的详细网络架构如表 4.1。使用 MobileNetV4 替换 HGNetv2 作为特征提取主干网络，可以使得算法在保持准确率的情况下提升实时响应能力、速度和精度要求。

通过将 MobileNetV4 网络架构引入 RT-DETR 的主干网络，可以有效地降低计算量和参数量，提升计算效率。同时，保持原有的多尺度特征提取能力和特征融合模块，在不显著影响精度的前提下提升算法的实时响应能力。这一改进为红外微小目标检测在资源受限环境下的应用提供了更好的性能表现，尤其在实时性要求较高的场景中。

（2）StarNet 骨干替换

StarNet 提供了多个规模的网络版本（S1-S4），其区别体现在每个 Stage 中所使用的 Block 数量及通道维度的不同，本文最终采用 StarNet-S4 网络结构，其参数量较高，但计算量及准确率也较高，网络各阶段的结构配置如表 4.1 所示。

表 4.1 StarNet-S4 网络结构配置

Stage	输出尺寸	Block 数量	通道数
Stem	112×112×32	1	32
S1	112×112×64	1	64
S2	56×56×128	2	128
S3	28×28×256	3	256
S4	14×14×512	3	512
Head	1×1×1000	-	-

相比于使用深度可分离卷积（DWConv）的轻量网络，StarNet 采用标准卷积进行特征提取，避免了通道信息分离导致的表达能力损失；而其乘法残差结构无需引入注意力机制即可增强特征选择性。此外，整个网络架构保持了良好的模块化与并行性，适用于在嵌入式设备或边缘端部署。

4.3.2 Fusion 模块改进

特征融合时，由于 CCFM 模块需要对跨尺度特征图分别进行两次上采样及两次下采样，且 Fusion 模块的核心之一是 N 个 RepBlock，RepBlock 是一种模块化的子结构，其结构由卷积操作、激活函数和残差连接组成，如图 4.6 所示。

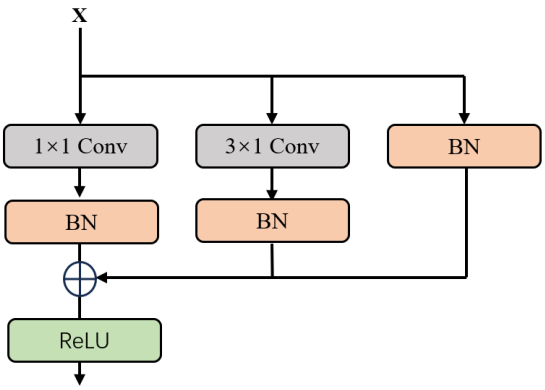


图 4.6 Repc3 结构

RepBlock 中的大量逐通道操作（卷积、激活函数）虽然计算量低于全卷积，但随着通道数的增加，其累积计算量也显著提升。且红外图像普遍为高分辨率特征图（尤其是用于小目标检测的特征层）会显著增加每次卷积、注意力计算和融合操作的计算量。将原 fusion 模块的 Repc3 分别用新的轻量注意力模块替换，通过对比选择最合适的轻量化结构。

（1）UIB 模块替换

将 Repc3 直接替换为 UIB 模块，构成了新的 UIB_C3 模块。改进后的 fusion 模块的结构如图 4.7 所示。

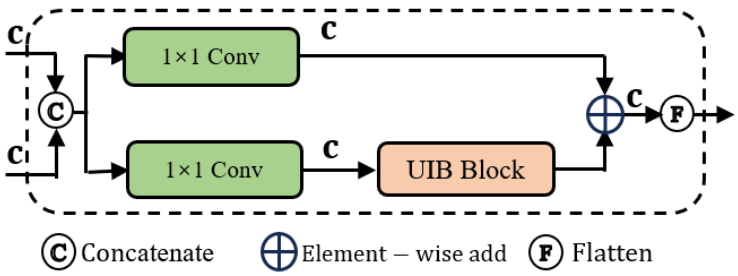


图 4.7 UIB_fusion 模块结构

UIB 继承了倒置瓶颈的思想，利用深度可分离卷积显著降低计算成本，且通过引入额外的 ExtraDW 和 FFN，使其在捕捉局部与全局特征方面表现更强，最终通过 NAS 自动搜索可以生成适配不同设备的最佳架构，在参数共享和表达能力之间取得平衡。

（2）EMA 模块替换

EMA 融合通道与空间注意力机制，不仅可以作为独立注意力机制与 HGBlock 高效融合，也可以将 Repc3 与 EMA 模块相结合，构成了新的 EMA_C3 模块。改进后的模块的结构如图 4.8 所示。

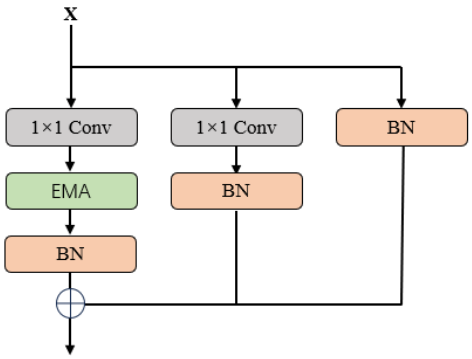


图 4.8 EMA_fusion 模块结构

该结构利用 EMA_attention 提升注意力建模能力，借助多路径结构和 nn.Identity()优化计算效率，没有使用重参数化，体现 RepBlock 在推理时的简化思想，两者的结合使得模块在训练时具有更强的表达能力，同时在推理时保持高效性。

4.4 本章小结

本章深入探讨了基于轻量化思想的红外微小目标检测优化方法。首先，介绍了两种典型的轻量化网络结构（MobileNetv4 和 StarNet），包括它们的结构特点与创新设计。随后，围绕 RT-DETR 模型，开展了基于轻量化骨干网络的检测算法研究，分别将原始骨干网络替换为 MobileNetv4 和 StarNet，以提升模型在红外微小目标检测中的计算效率和推理速度。同时，在 fusion 模块中引入 UIB 和 EMA 结构，替代原有的 RepC3 模块，并进一步结合改进的损失函数，以实现在参数量和检测速度基本不变的前提下提升检测精度。。

5 实验及结果分析

实际测试中，模型的性能决定了算法改进的有效性，评价体系也需要根据不同的场景和目标进行调整。因此，确保实验设计的客观性和选择恰当的评价指标对于准确评估算法表现至关重要。本章首先利用相关指标对红外图像增强算法进行分析。其次，针对改进的算法模型，基于红外图像数据集和评价指标进行评估。最后通过对比实验验证改进算法的优势。

5.1 实验准备

5.1.1 实验数据集及训练参数

本实验所使用红外飞机小目标数据集图像采用国科大红外目标开源数据集，内容主要以地面背景、天空背景、多架飞机、飞机远离、飞机靠近等情景为主。选取其中 11371 张红外图像包括 13944 个飞机实例，图像分辨率 256*256，通道数为 1，位深 24。该数据集被广泛应用于目标检测、目标追踪等任务。标签格式为 txt，在该数据集中，训练集占比 80%，测试集占比 20%。训练所用参数如表 5.1 所示。

表 5.1 训练参数

参数	参数
初始学习率	0.01
批处理大小	8
迭代次数	100
优化器	SGD

5.1.2 实验环境

实验在 Ubuntu 18.04 系统上进行，硬件配置为 RTX 2080Ti 显卡和 16GB 内存，采用 Python 语言与 PyTorch 深度学习框架构建模型，详细信息如表 5.2 所示。

表 5.2 实验环境

配置	参数
操作系统	Ubuntu18.04
内存	16G
GPU	RTX 2080 Ti

CUDA	CUDA 11.0
编程语言	Python 3.9
框架	Pytorch 1.8.0

5.1.3 网络评价指标

本实验使用准确率（P）、召回率（R）和平均精度（AP）三个评价指标来衡量模型性能，其数学表达式为式 5.1。

$$Precision = \frac{TP}{TP + FN} \tag{5.1}$$

其中 TP 表示真正例（True Positives），FP 表示假正例，Precision 衡量的是模型预测为正样本的实例中，真正为正样本的比例。用于评估预测为正样本的实例的准确性。

召回率又称查全率，是预测为正例的样本占预测样本的比例，其数学表达式为：

$$Recall = \frac{TP}{TP + TN} \tag{5.2}$$

平均精度（AP）的数学表达式为：

$$AP = \int P(R) d_R \tag{5.3}$$

用于评估模型在不同阈值下的 Precision-Recall 曲线，是对 Precision-Recall 曲线进行积分得到的平均值。它衡量了模型在不同阈值下预测结果的平均准确性。此外，为了评估算法的实时处理能力，引入 FPS（F/s）评价指标。计算每秒可以处理的图片数量。

5.2 图像处理实验结果及分析

5.2.1 图像增强算法评价指标

图像质量评价采用客观评价指标衡量图像的增强效果，其中包括：信息熵（Entropy）用于评价图像的平均信息量；平均梯度（Average Gradient）用于评价图像的清晰度；熵增强量（EME）用于评价图像对比度。

其中，MSE 为两幅图像的均方误差，Peak 表示图像像素强度的最大取值，由于图像中每个采样点用 8 位表示， Peak= 255。

（1）信息熵（Entropy）：图像的各灰度值是独立的，定义 P(i)是图像的灰度分布，P(i)表示灰度值为 i 的像素占比，L 为总灰度级数。则该图像信息熵定义如式 5.4 所示。

$$E = -\sum_{i=0}^{L-1} p(\mathrm{i}) log_2 \cdot p(\mathrm{i}) \tag{5.4}$$

图像信息熵表示图像中所包含的信息量的多少。信息熵值越大，表示图像信息量越丰富，增强效果越好。

（2）平均梯度（Average Gradient）：平均梯度的值可以反映融合图像微小细节的变化程度，对一幅大小为 M×N 的融合图像，ΔI_x和ΔI_y，分别是图像在行和列方向上的一阶差分，其平均梯度定义如式 5.5 所示。

$$\bar{g} = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(\Delta I_x^2 + \Delta I_x^2)/2} \tag{5.5}$$

平均梯度值越大，图像在某一方向的灰度级变化率就越大，层次会越多，图像越清晰，质量越好。

（3）熵增强量（EME）：

EME 指标将图像分为多个图像块，统计图像块内的像素值，根据像素块内最大最小值的关系评价对比度。计算公式如式 5.6 所示。

$$EME = \frac{1}{k1 * k2} \sum_{m=1}^{k1} \sum_{n=1}^{k2} 20 \log\left(I_{max,m,n}^{w} / I_{min,m,n}^{w}\right) \tag{5.6}$$

其中，$k1$、$k2$ 是图像横纵方向的分块数量，$I_{max,m,n}^{w}$表示块内最大值，$I_{min,m,n}^{w}$表示块内最小值。EME 数值越大，表示图像对比度越高。

5.2.2 增强结果及分析

为验证本文方法的有效性，选取 infrared_COCO 数据集作为实验数据源，同时选取 SSR、MSR、DDE、双边滤波等算法作为对照组进行比对分析。

从主观视觉效果来看，SSR 和 MSR 算法在一定程度上提升了图像的整体亮度与清晰度，但目标与背景之间的对比度提升有限，无法显著增强目标的辨识度。而本文提出的改进算法在图像增强方面表现优越，不仅提升了目标与背景的对比度，还显著保留了边缘信息，增强了图像细节表现力，具有更好的视觉效果。图 5.1 展示了几种图像增强算法的主观对比结果。

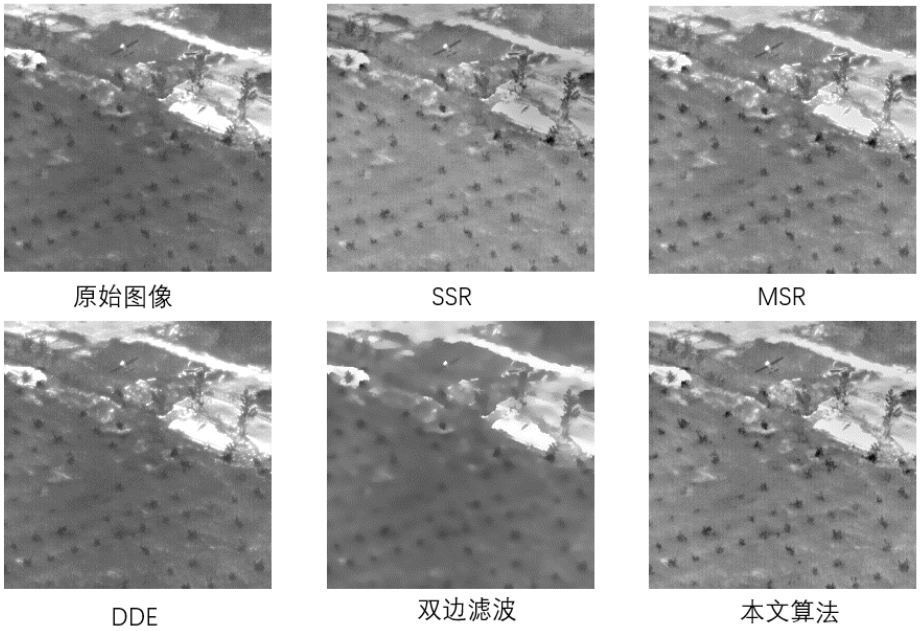


图 5.1 几种增强算法对比结果

客观指标方面，通过信息熵（Entropy）、边缘梯度（AG）、和增强对比度度量（EME）对算法进行定量评估，结果表明本文算法在同类方法中表现最佳。相比 DDE 和 MSR 算法，本文方法在整体图像清晰度和目标对比度上取得了显著提高。实验数据具体见表 5.3。

表 5.3 算法增强指标对比

指标 / 算法	Entropy	AG	EME
原始图像	6.2750	41.9135	2.6388
SSR	5.7150	44.2675	2.8967
MSR	6.2176	44.9135	2.8932
DDE	6.1594	44.2206	2.6857
双边滤波	6.3155	21.7407	1.4891
本文算法	6.2793	46.8969	2.8882

由表中数据可以看出，本文算法在 Entropy 和 AG 上均处于最高水平，同时熵信息量也显著优于其他对比算法，充分证明了本文方法在增强图像清晰度和对比度方面的优越性。

5.3 目标检测实验结果及分析

5.3.1 IoU 对比实验

本文在网络模型中引入了 Shape-IoU 损失函数，该函数包含一个与数据集中目标尺度相关的 scale 参数。为了确定最佳参数值，本文在不同的 scale 参数设置下进行了对比实验，相关实验结果如表 5.4 所示。

表 5.4 Shape-IoU 不同参数的对比实验

s	0.1	0.2	0.3	0.4	0.5	0.6	1.0
mAP(%)	83.5	83.4	84.9	**85.3**	84.8	83.5	83.4

检测结果变化散点图如图 5.2 所示。

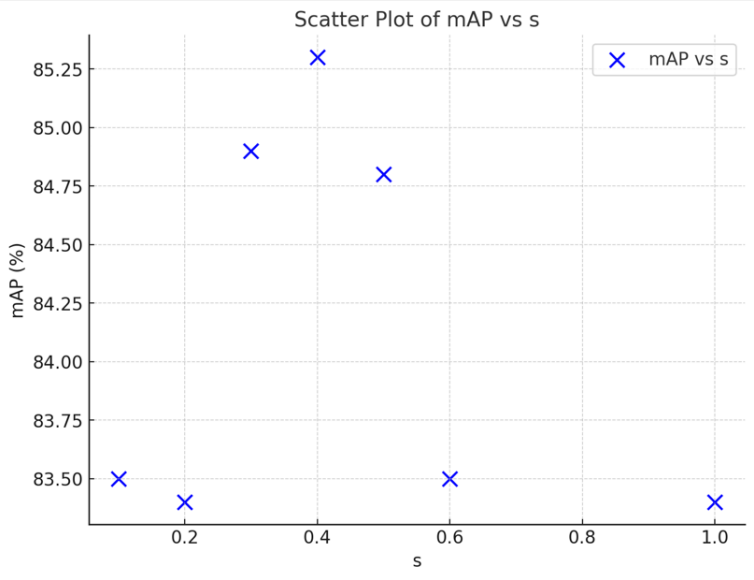


图 5.2 Shape-IoU 检测对比图

根据实验对比结果，随着尺度参数 s 的增大，模型的平均精度（mAP）呈现先上升后下降的趋势，这表明尺度参数对模型的检测能力具有非线性影响。当 s 值接近两端（0.1 和 1.0）时，mAP 均较低，分别为 83.5%和 83.4%，说明过小或过大的尺度参数均对模型性能不利，过小的参数导致模型对形状信息的权重不足，从而影响检测精度，过大的参数可能导致模型对形状信息过于敏感，反而削弱了整体性能。而在本实验数据集上，尺度参数 s=0.4 是最优选择，模型的综合性能达到最高。在后续优化和扩展实验中，可以进一步细化参数调节范围，验证其在不同数据集上的通用性。

同时为了验证 IoU 改进的有效性，采用不同的 IoU 损失函数进行对比，对比结果如表 5.5 所示。可视化结果如图 5.3 所示。

表 5.5 不同 IoU 损失函数的对比结果

Loss	P(%)	R(%)	AP(%)
CIoU	71.5	74.4	84.7
WIoU	73.7	75.8	83.9
EIoU	74.5	75.1	84.2
GIoU	73.2	75.2	84.6
Shape-IoU	73.5	75.6	85.3

从实验结果可以看出，当 scale 参数设置为 0.4 时，相比其他损失函数，Shape-IoU 损失函数的改进在 AP 上取得了显著提升，达到 85.3%，比原始网络采用的 GIoU 提升了 0.7%，并在精度和召回率上也表现得相对均衡（精度为 73.5%，召回率为 75.6%），这表明，Shape-IoU 在综合性能上具有明显的优势。

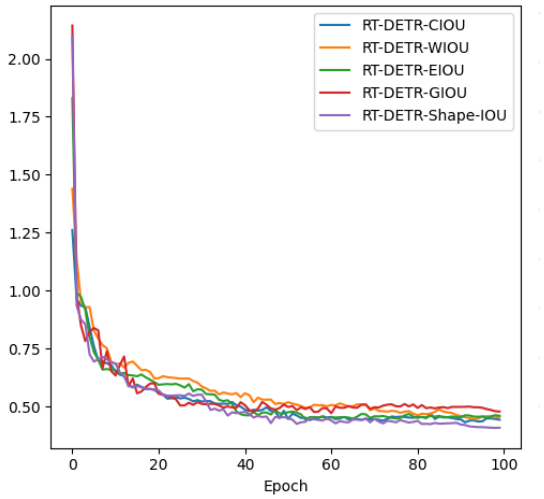


图 5.3 IOU 对比结果

5.3.2 消融实验

为了验证所提出改进方法的有效性及其效果，本文在原始 RT-DETR 模型的基础上，使用相同的数据集进行消融实验，并保持实验参数的一致性。表 5.5 展示了改进后的 RT-DETR 模型在消融实验中的结果。

表 5.6 改进点消融实验

EMA	CAMixing	Shape-IoU	ATFL	P(%)	R(%)	AP(%)	Param(10^6)
				73.2	75.2	84.6	32.81
√				72.2	76.3	85.5	33.40
	√			74.8	75.2	85.6	34.97
√	√			74.1	77.0	86.2	35.23
		√	√	75.5	76.2	85.9	32.81
√	√	√	√	75.4	77.1	87.8	35.23

根据实验结果可以得出：

基准模型性能在未应用改进方法的情况下，精度（P）为 73.2%，召回率（R）为 75.2%，平均精度（AP）为 84.6%，参数量为 32.81M。

引入 EMA 后，召回率提高了 1.1%，AP 提升了 0.9%，参数量从 32.81M 增加到 33.40M（+1.8%），EMA 的改进主要体现在召回率的提升上，说明其对特征提取的稳定性和一致性具有较大帮助。

引入 CAMixing 后，精度提升了 1.6%，AP 提升了 1.0%，参数量从 32.81M 增加到 34.97M（+6.6%），CAMixing 显著增强了精度和平均精度，说明该方法对背景和目标区域的特征对比度提升有较大帮助，但相对增加了较多的参数量。

当 EMA 和 CAMixing 联合使用时，精度提升了 0.9%，平均精度提升了 1.6%，参数量从 32.81M 增加到 35.23M（+7.4%），对比单独引入 EMA 与 CAMixing，两者的联合改善了精度和 AP，虽然召回率略有下降，但性能整体优于单独使用任一方法的效果。

当 Shape-IoU 和 ATFL 联合使用时，精度提升了 2.3%，召回率提升了 1.0%，AP 提升了 1.3%，由于是对损失函数的改进，参数量保持稳定。Shape-IoU 和 ATFL 的改进集中在性能提升与参数量稳定之间取得了平衡。说明两者的优化重点在于提升特征匹配和目标框的定位精度。

当四种改进方法（EMA +CAMixing +Shape-IoU+ATFL）联合使用时，AP 提升了 3.2%，精度提升了 2.2%，召回率提升了 1.9%，参数量从 32.81M 增加到 35.23M（+7.4%），所有方法的联合使用取得了最佳性能，验证了各方法之间的协同作用。参数量增加幅度适中，性能提升显著，表现出了改进方案在复杂度与性能之间的良好平衡。

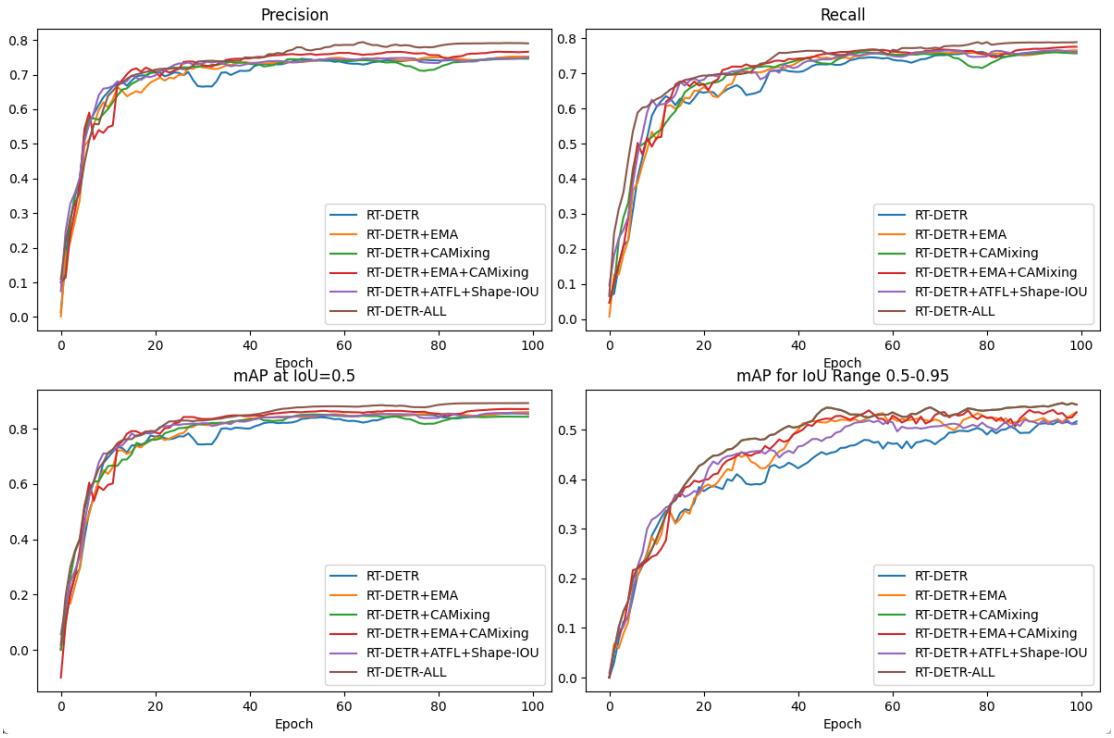


图 5.4 不同改进点的评价指标对比图

图 5.4 展示了各种不同模型在四个评价指标上的性能比较：Precision、Recall、mAP at IoU=0.5 和 mAP for IoU Range 0.5-0.95。由于数据集中复杂背景较多，正负样本差别较大，检测收到背景干扰，所有模型的精度变化曲线波动较为显著。

对于 Precision，不同方法的精度在前 20 个 epoch 快速提升，随后逐渐趋于平稳。RT-DETR-ALL 的精度整体上略高于其他方法，特别是在后期。RT-DETR+EMA 和 RT-DETR+ATFL+Shape-IOU 的精度提升较快，且在中期（20~40 个 epoch）处于领先地位。

对于 Recall，各方法的 Recall 在训练早期（前 10 个 epoch）提升迅速，之后逐步趋于饱和。RT-DETR+EMA+CAMixing 在 Recall 上前期表现最好，能更有效地捕捉到正样本，从而提升召回率，后趋于平稳，RT-DETR 基准网络的 Recall 最低，表明其漏报率较高。

IoU=0.5 表示检测框与真实框的交并比阈值较低，容易达成较高的 mAP。对于 mAP at IoU=0.5 的总体趋势，所有方法在早期（0~20 个 epoch）迅速提升，后期趋于平稳，RT-DETR-ALL 的 mAP 在大部分训练过程中略优于其他方法，且稳定性较强。

对于 mAP for IoU Range 0.5-0.95，RT-DETR 的性能最低，其他改进方法均有明显提升，RT-DETR+EMA 和 RT-DETR+ATFL+Shape-IOU 在中期有较高的 mAP，但后期略低于 RT-DETR-ALL，RT-DETR-ALL 在后期表现显著优于其他方法，表明其综合能力更强。

对于此次模型的对比，各自的 Loss 变化曲线如图 5.5 所示。

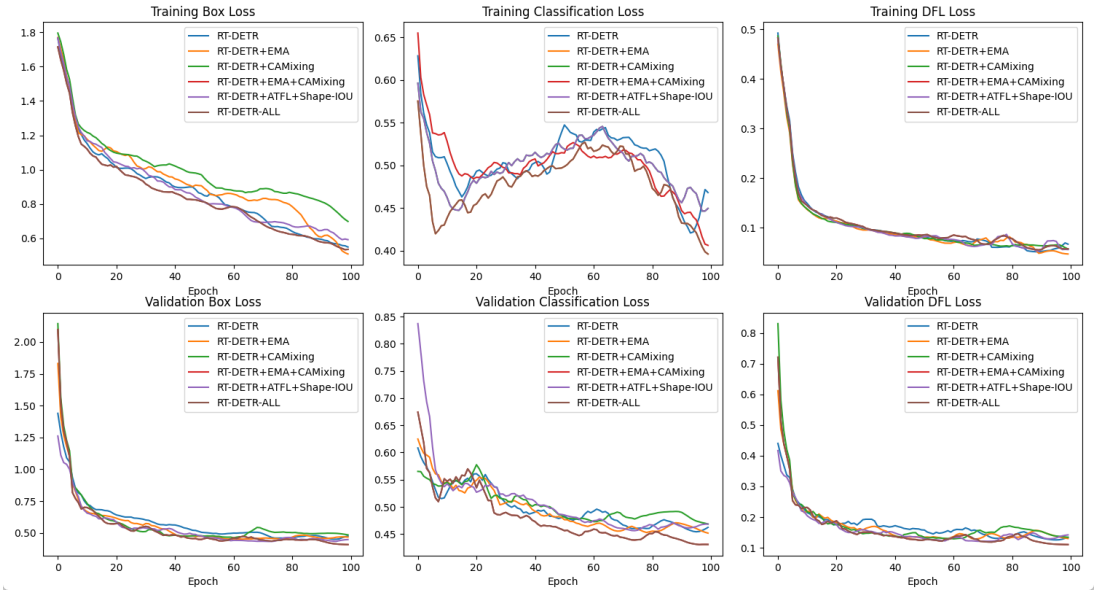


图 5.5 Loss 变化曲线

在训练阶段，Box Loss 采用 GIOU 衡量预测框与真实框的边界差异，损失越小表示模型定位越准确。Classification Loss 采用 VFL 区分正负样本，表示模型对类别分类的准确性，损失越小说明分类能力越强，由于红外数据集本身的差异性，各方法在训练集上的分类损失波动明显。RT-DETR+ATFL+Shape-IOU 与 RT-DETR-ALL 方法在整个过程中表现出更低的 Box Loss 和 Classification Loss，说明了损失函数改进的有效性。DFL Loss 旨在优化预测框的边界分布，使预测框更接近真实框，在整个训练过程中，不同方法之间的差异较小，各方法的 DFL Loss（分布式焦点损失）在训练初期快速下降，后期几乎趋于一致，说明该部分损失对模型性能的差异影响较小。

在验证阶段，loss 变化总体趋势与训练阶段类似，但波动较为平缓。RT-DETR+ATFL+Shape-IOU 和 RT-DETR-ALL 在定位损失和分类损失的变化中较优，RT-DETR+CAMixing 在部分阶段损失较高。

因此，RT-DETR-ALL 在多个指标上表现最佳，尤其是分类损失和验证阶段的 Box Loss，说明它在综合优化后具有较强的性能和泛化能力。RT-DETR+ATFL+Shape-IOU 在分类损失上表现优异，表明其对分类的针对性优化有效。RT-DETR+EMA+CAMixing 在训练和验证阶段均表现稳定，尤其在 BoxLoss 和 DFL Loss 方面表现出色，说明 EMA+CAMixin 策略对损失的平滑效果显著。

对于各个网络的预测结果如图 5.6 所示。

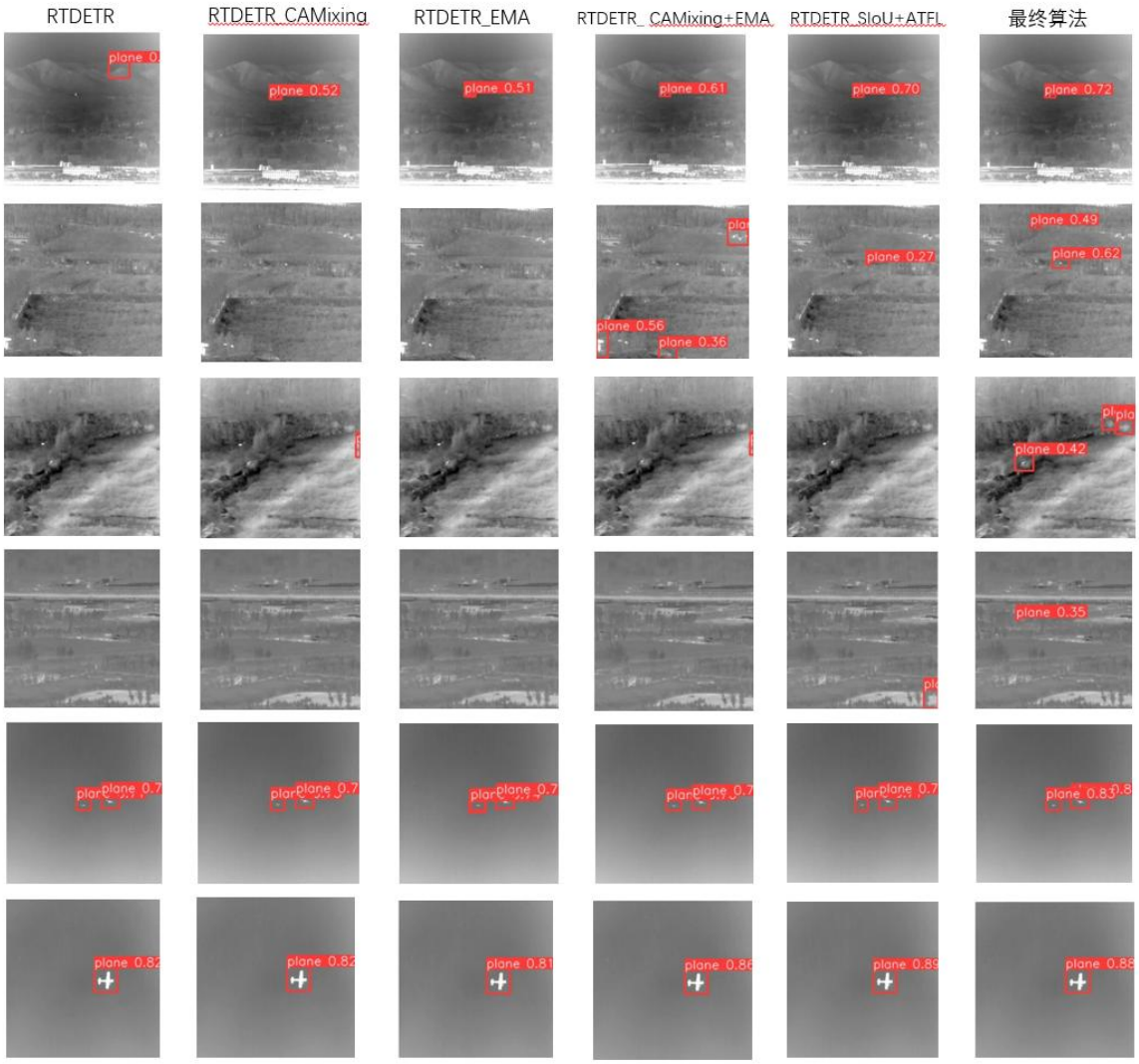


图 5.6 网络改进预测图

通过对预测结果的观察与分析，能看到网络的改进显著提升了检测精度，一定程度上解决了原始网络的漏检，误检等问题。

在简单背景下，各网络对微小目标的检测均表现优异，目标置信分数显著上升，说明背景复杂性是影响检测性能的主要因素。

对于复杂背景的图像，目标与背景对比度低，背景干扰性强，原始网络对复杂背景的干扰信息适应能力较弱，且对小目标特征学习不足，容易出现现漏检和误检现象。RT-DETR+EM/CAMixing 一定程度上缓解了误检问题，但由于对复杂背景特征的抑制能力有限，漏检现象依然存在。而 RT-DETR+ATFL+Shape-IOU 以及 RT-DETR+EM+CAMixing 通过两两结合，针对小目标特征进行有效的进一步学习，显著改善了漏检问题，尤其是在复杂背景下对小目标的检测率有所提升。但由于背景干扰信息中存在与目标类似的特征，误检现象较为明显。而 RT-DETR-ALL 通过结合所有改进，将所有策略结合，进一步提升了模型对复杂背景和小目

标的检测能力，在解决漏检问题的同时显著降低了误检率，特别是在复杂背景中展现了更好的鲁棒性。

5.3.3 与其他算法对比实验及分析

为了验证本文方法在红外图像目标检测中的有效性，将本文改进算法（RT-DETR-ECSA）与今年前沿的目标检测算法（YOLOv9n、YOLOv10n 等）进行了系统对比。对比指标包括平均精度（AP）、参数量（Param，单位：百万）、检测速度（FPS）以及浮点运算量（GFLOPs），在相同的训练条件和方法下完成实验。表 5.7 总结了各算法的性能对比结果。

表 5.7 不同算法的检测指标对比

Model	AP/%	Param/10^6	FPS	GFLOPs
YOLOv5n	79.7	4.51	65	5.8
YOLOv9n	81.2	1.27	67	1.2
YOLOv10n	84.8	3.85	75	8.6
RT-DETR	84.6	32.81	118	108
本文算法	87.8	35.23	106	110

实验结果表明，本文提出的改进算法在红外图像目标检测中表现最佳，其平均精度（AP）达到了 87.8%，较 YOLOv8n 提高 3.4%，较 RT-DETR 提高 3.2%，验证了其在复杂场景下的检测能力。在参数量方面，本文算法为 35.23M，略高于 RT-DETR 的 32.81M。检测速度上，DETR-ECSA 达到了 106FPS，能够满足实时性要求。尽管浮点运算量（GFLOPs）增加至 110，相比 RT-DETR 略有提升，但综合考虑检测精度与实时性能的提升，本文算法展现了更优的性能平衡，进一步证明了算法改进的有效性。

5.3.4 轻量级网络比较分析

在本文的研究中，为了深入探索不同神经网络架构对红外微小目标检测性能的影响，基于 RT-DETR 模型，采取对比实验，分别对其整体的骨干网络进行了替换，并针对 fusion 模块进行了优化改造。同时，在保持计算量和检测速率不降低的前提下进一步提升检测精度，将轻量化网络结构与改进的损失函数相结合。具体实验结果如表 5.8 所示。

表 5.8 轻量化实验结果

Model	P/%	R/%	mAP/%	Param/10^6	FPS
RT-DETR	73.2	75.2	84.6	32.81	118
StarNet	70.1	69.3	79.0	11.94	126
MobileNetv4	70.5	67.7	79.4	12.04	132
UIB_C3	71.4	69.4	78.2	16.73	157
EMA_C3	71.3	70.4	80.2	14.93	149
EMA_C3+Shape-IoU+ATFL	71.3	70.4	80.9	14.93	149

表 5.8 轻量化实验结果表明，针对骨干网络的替换，采用 StarNet 后，模型参数量显著降低至 11.94M，推理速度提升至 126 FPS，表现出良好的轻量化能力。将骨干网络替换为 MobileNetv4 后，推理速度进一步提升至 132FPS，模型参数量降至 12.04M，mAP 为 79.4%。

针对 fusion 模块的改进，采用 UIB_C3 网络后，参数量减少至 16.73M，推理速度提升至 157FPS，但 mAP 略降至 78.2%。使用 EMA_C3 结构后，参数量为 14.93M，推理速度为 149FPS，mAP 提升至 80.2%。在此基础上结合 Shape-IoU 与 ATFL 损失函数，最终模型的 mAP 达到 80.9%，在保持速度和参数基本不变的情况下进一步提升了检测精度。

综合来看，EMA_C3 融合了 EMA 模块与 RepBlock 分支的设计思路，表现出较为显著的轻量化效果。轻量化网络有效提升了检测速度和减少参数量，但在检测性能上有所牺牲，轻量化网络更适合资源受限的应用场景。

5.4 红外微小目标检测系统设计

为实际应用于红外检测系统并验证 RT-DETR 目标检测算法在精度和轻量化方面的有效性，本节详细介绍了集成 RT-DETR 的红外微小目标检测系统的设计与实施方案。该系统能够完成微小目标检测任务，并将检测结果同步显示在主界面上。每次检测的飞机坐标和数量会直观地显示在信息栏中，方便用户查看。系统操作简单、界面整洁、功能丰富，适合不同类型的用户使用。

5.4.1 开发环境

本文采用开发环境如下：

（1）软件开发环境：PyCharm2022、Windows11。PyCharm 是 JetBrains 开发的一款专业 Python 集成开发环境。Windows11 作为微软推出的最新操作系统，提供了更好的性能优化、窗口管理和兼容性。

（2）界面开发框架： PyQt。PyQt 框架是一个用于开发图形用户界面的

Python 库，其结合 Python 语言的简单高效及 Qt 框架的稳定性，并且可以在多种不同的操作系统上进行设计和使用。

（3）图像处理工具：OpenCV 技术。本文采用 Python 代码结合 OpenCV 库，实现红外图像增强技术。通过 OpenCV 提供的图像滤波、DDE、MSR、直方图均衡化等算法，有效提升红外图像的清晰度和目标可见性。

5.4.2 GUI 可视化界面搭建

本文设计的红外目标检测软件系统结构主要由模型选择模块、图像输入模块、检测模块和显示模块组成。软件系统的结构图如图 5.7 所示。

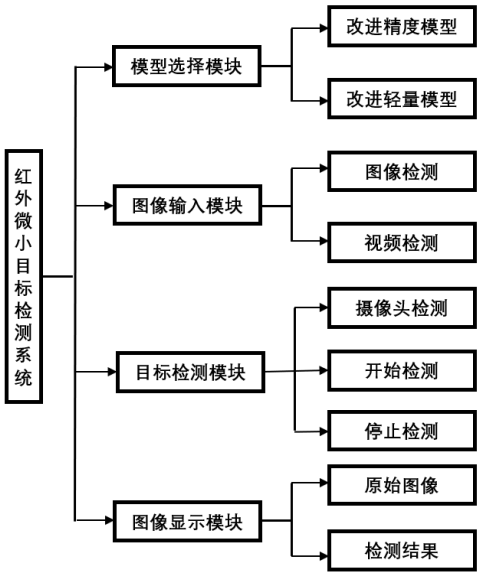


图 5.7 红外微小目标检测软件系统结构图

模型选取模块允许用户选择载入到检测系统的权重模型，包括改进后的高精度模型以及轻量化模型，以满足实际检测任务需求，提高检测系统的适用性；图像输入模块支持实时摄像头输入或本地视频、图像文件导入；控制模块负责操作开始检测和停止检测；显示模块包括输入输出显示区域和检测飞机的位置与数量结果统计部分。最终设计的软件窗口如图 5.8 所示。

图 5.8 系统功能界面

5.4.3 系统测试

系统基于改进后的网络模型，主要功能针对检测设计。模型选择训练完成后的 pt 文件，不同改进点训练出的 pt 文件不同，检测结果也不同。

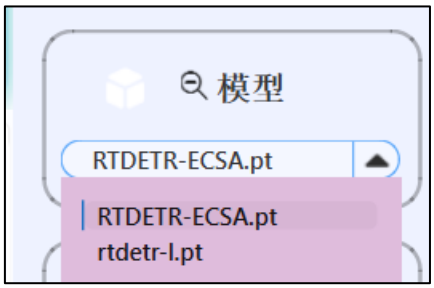


图 5.9 模块选择区域

图像/视频来源主要包括本地导入，用户可以通过点击按钮导入图像、视频并进行后续的目标检测。摄像功能如图 5.10 所示。

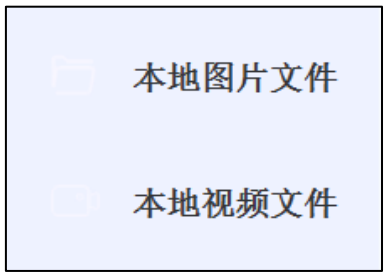


图 5.10 数据输入模块

操作过程区域记录了用户与系统的交互过程，检测结果区域展示检测结果的

详细信息，包括检测的类别数量、目标数量（plane number）以及置信分数（socre）。最终，将数据输入的图像或者视频进行检测，最终检测结果如图 5.11 所示。

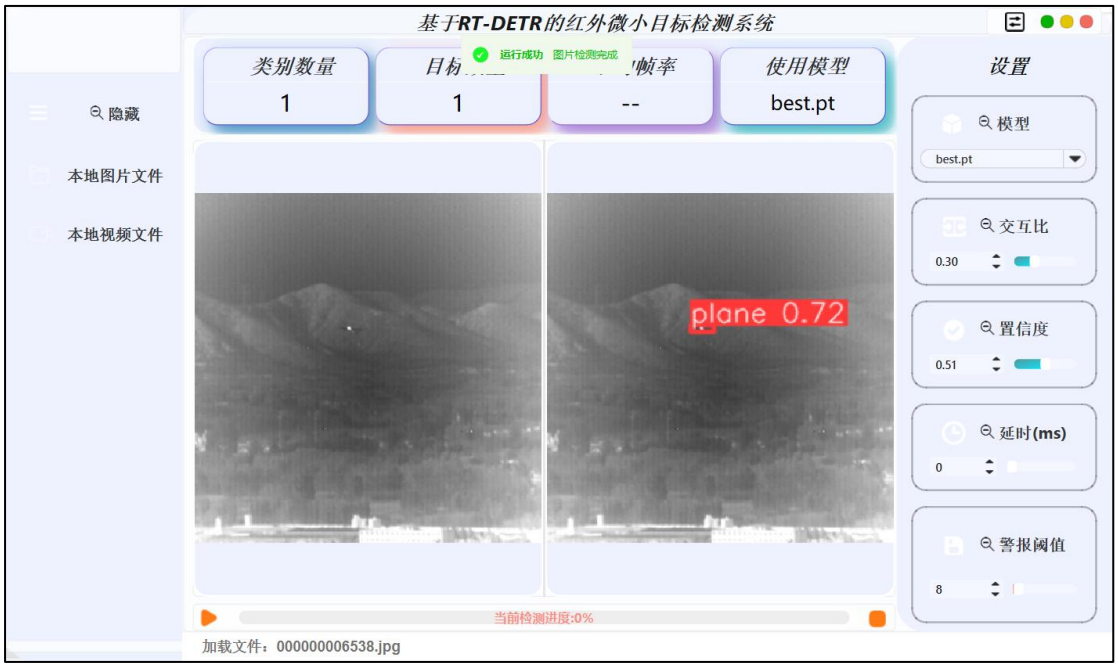


图 5.11 系统检测界面

系统从图像或视频输入开始，通过选择合适的模型进行处理，最终反馈检测结果，帮助用户评估模型的性能并做出决策。

5.5 本章小结

本章首先开展了红外图像增强实验，并通过对比试验验证了所采用增强算法在对比度与清晰度提升方面的有效性。随后，详细介绍了实验所使用的数据集、训练参数、实验环境及评价指标。为评估改进模块的有效性，本章分别进行精度消融实验与轻量化对比试验，结果表明不同改进模块在提升检测精度与模型轻量化方面具有显著优势，进一步证明了其必要性。最后，通过与同期经典 YOLO 系列算法的对比实验，从检测精度与速度等指标综合评估本文方法的性能。实验结果充分验证了所提出改进算法在红外图像处理与目标检测任务中的有效性与优越性。最后，设计并实现了具备图形用户界面的检测系统，构建了完整的红外微小目标检测平台。

6 总结与展望

6.1 研究工作总结

本文在深入红外图像特性的基础上，利用滤波算法及 DDE-MSR 算法进行红外图像增强，针对远景小目标及复杂林地背景导致的检测精度下降问题，提出了一种基于改进 RT-DETR 的目标检测算法，实现了对红外飞机微小目标快速精准的检测。并针对检测速度低、参数计算量大的问题，研究了 RT-DETR 网络的轻量化，本文具体工作总结如下：

1）深入研究了红外图像的辐射特性及现有红外图像增强技术，并通过滤波算法结合 DDE 与 MSR 算法。通过实验对比，本文使用的图像增强方法在保留边缘信息、增强目标与背景对比度方面表现优越，同时具备更佳的去噪效果和视觉质量，优于多种常见算法。

2）针对红外目标微弱、背景复杂影响检测精度的问题，提出了一种改进 RT-DETR 的检测算法。在 RT-DETR 骨干网络 HGNetV2 中引入 EMA 注意力模块，并在 Neck 结构中集成 CAMixing 注意力机制，以增强目标特征的表达能力。同时，对损失函数进行了优化，采用 Shape-IOU 替换 GIOU，VTL 分类损失调整为 ATFL，并重新定义锚框筛选规则。实验结果表明，该方法有效提升了检测性能。

3）构建了一种轻量化的 RT-DETR 目标检测模型。该模型使用 MobileNetv4 作为骨干提取网络，并用优化的注意力模块替换原始 repc3 结构。通过对比实验筛选最优轻量化策略，使模型在降低计算复杂度和参数量的同时，保持较高的检测精度。

为验证改进算法的有效性，本文与其他主流深度学习算法进行了对比分析。实验结果表明，本文方法在检测精度和计算效率上均表现出色。相较于现有算法，本方法在保障高精度的同时，显著提升了检测速度，满足工程应用中对红外微小目标快速检测的需求。

6.2 未来工作展望

深度学习技术的应用不仅优化了红外微小目标检测的流程，还显著提升了图像处理效率。本文针对检测过程中面临的关键挑战进行了深入分析，并提出了相应的解决方案。未来可针对于以下方向进行研究，以进一步提升检测精度和算法鲁棒性。

（1）提高红外图像目标检测算法的泛化能力

当前检测算法在特定场景和目标类型下表现较好，但在复杂环境或未见场景中，识别效果可能显著下降。因此，未来研究应注重开发具有更强泛化能力的算法。例如，通过构建更大规模的多样化红外图像数据集，以及采用跨场景迁移学习方法，使算法能够适应不同背景、不同目标尺度及不同气象条件下的目标识别任务。

（2）红外微小飞机目标的漏检率和虚警率问题

红外微小目标通常具有低信噪比、低灰度对比度和不完整的结构特征，使其易受漏检或虚警的影响。降低检测门限虽能提高目标检出率，但会导致大量虚警；相反，提高检测门限可减少虚警，但可能漏检微弱目标，形成检测性能的权衡挑战。为优化检测效果，未来研究可聚焦于目标特征提炼与增强，如利用自适应滤波、多尺度分析提升弱目标可见性，结合时空特征约束目标轨迹，或基于区域分布特征优化筛选策略，以提升目标检测的精准性与鲁棒性。

（3）应对实际应用中的复杂场景

红外图像通常受外界条件（如大气扰动、传感器噪声、极端天气等）的影响，识别性能可能显著下降。未来研究需要针对这些复杂场景提出有效的解决方案：比如引入对抗性学习和自适应增强技术，提高模型对环境干扰的鲁棒性。研究非均匀大气衰减和背景抑制的物理建模方法，以提升红外图像预处理的质量。

参考文献

[1] Qing-BO Ji, etal. A detection algorithm for the small moving target in infrared image sequences with the dynamic background[C]. Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition:2007.

[2] Gao Kanglin, etal. A Small Target Detection Algorithm Based on Immune Computation and Infrared Background Suppression[C]. Third International Conference on Natural Computation. 2007.

[3] 郭雨洁. 基于信息增强的微小目标检测与识别方法研究[D].广东技术师范大学,2023.

[4] 刘颖,孙海江,赵勇先.基于注意力机制的复杂背景下红外弱小目标检测方法研究[J].计算机软件与计算机应用,2023,38(11):1455-1467.

[5] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2022: 580-587.

[6] 周良钰,杨硕.基于改进的 DETR 的目标检测与测距实现[J].信息技术与信息化,2023(06):75-78.

[7] 袁帅,延翔,张昱赓,等.双邻域差值放大的高动态红外弱小目标检测方法[J].红外与激光工程,2022,51(4):20221071.

[8] 吴文怡. 红外图像序列中弱小目标检测与跟踪技术研究[D].南京:南京航空航天大学,2008.

[9] 潘胜达,张素,赵明,等.基于双层局部对比度的红外弱小目标检测方法[J].光子学报,2020,49(1):0110003.

[10] Bae T W. Small target detection using bilateral filter andtemporal cross product in infrared images [J]. InfraredPhysics & Technology ,2021 ,54 (5) :403 - 411.

[11] LI J,GU J N,HUANG Z D,et al. Application research of improved DETR algorithmin PCB electronic compo-nent detection[J].Applied Sciences,2019,9(18):3738-3750.

[12] 李文博,高尚.基于深度学习的红外小目标检测算法综述[J].激光与红外.2023,53(10)1476-1484.

[13] 娄康,朱志宇,葛慧林.基于目标运动特征的红外目标 检测与跟踪方法[J].南京理工大学学报,2019,43(4):455-461.

[14] Wan. Radar CFAR Thresholding in Clutter and Multiple object Situations[J]. IEEE Transactions on Aerospace and Electronic Systems, 1983, 19(4):608-621.

[15] Liu D,Zhang J,Dong W. Temporal profile based smallmoving target detection algorithm in infrared image sequences[J]. International Journal of Infrared and Milli-meter

Waves,2020,28(5):373-381.

[16] Pang X,etal. Infrared Image Semantic Segmentation Based on Improved DeepLab and Residual Network[C]. 10th International Conference on Modelling, Identification and Control (ICMIC). 2018.

[17] 李慕锴.基于深度学习的小尺度红外行人检测技术研究[D].上海:中国科学院大学,中国科学院上海技术物理研究所,2022.

[18] 徐延想.基于深度学习的红外小目标检测研究与实现[D].杭州：杭州电子科技大学，2021.

[19] 杨子轩,肖嵩,董文倩,等.一种引入注意力机制的红外目标检测方法[J].西安电子科技大学学报,2022,49(3):28-35.

[20] 蒋志新.基于深度学习的海上红外小目标检测方法 究[D].大连:大连海事大学,2019.

[21] 崔颖,韩佳成,高山,陈立伟.基于改进 Deformable-DETR 的水下图像目标检测方法[J/OL].应用科技.2023.

[22] WEI Y L,YOU X G,LI H. Multiscale patch-based con- trast measure for small infrared target detection［J］. Pattern Recognition，2016，58：216-226.

[23] PANG J,LI C, SHI J,et al.R2-CNN: fast tiny objectdetection in large-scale remote sensing images[J].IEEETransactions on Geoscience and RemoteSensing , 2021.57:5512-5524.

[24] Liu S,Li F,Zhang H,et al. Dab - detr: Dynamic anchor boxes are better queries for detr[J]. arXiv preprint arXiv:2201.12329,2022.

[25] WANG D,ZHANG Q,XU Y,et al.Advancing plain visiontransformer towards remote sensing foundation modelJ]arXiv:2208.03987,2022.

[26] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,ct al.An image is worth 16x16 Words: transformers for imagerecognition at scale[J].arXiv:2010.11929,2020.

[27] 孔轩. 基于局部结构张量分析的红外目标检测方法[D]. 电子科技大学, 2022.

[28] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. arXiv, 2017.

[29] 洪季芳. Transformer 研究现状综述[J]. 信息系统工程, 2022(2):4.

[30] He K, Gkioxari G, Doll´ar P, et al. Mask r-cnn. in: Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.

[31] Kaiwen Duan, Song Bai, Lingxi Xie, et al. Centernet: Keypoint triplets for object detection[J]. International Conference on Computer Vision, Seoul, 2019.

[32] Z Tian, C Shen, H Chen, FCOS: Fully convolutional one-stage object detection, International Conference on Computer Vision (ICCV), IEEE, 2020.

[33] Kumar D , Zhang X . Ship Detection Based on Faster R-CNN in SAR Imagery by Anchor Box Optimization[J]. International Conference on Control Automation and Information Sciences (ICCAIS), Chengdu, 2019: 1-6.

[34] 蔡娟.基于深度学习的红外目标检测方法[J].计算机测量与控制,2023,31(1):45-50

[35] 苏娟，杨龙，黄华. 用于红外图像小目标舰船检测的改进 SSD 算法[J].系统工程与电子技术, 2020, 42(05): 1026-1034.

[36] Hong Z, Yang T, Tong X, et al. Multi-Scale Ship Detection from SAR and Optical Imagery via A More Accurate YOLOv3[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021,1(04):122-127.

[37] 焦军峰,靳国旺,熊新,等.旋转矩形框与 CBAM 改进 RetinaNet 的红外图像近岸舰船检测[J].测绘科学技术学报,2020,37(06):603-609.

[38] 王英. 基于动态规划的弱目标检测前跟踪算法的研究[D].2017.

[39] 张梦,邓小颖,曹海涛,张剑云,贺翔,朱金荣.基于改进 GoDec 算法的红外与可见光图像融合[J].激光杂志,2022,43(08):135-140.

[40] JIN Fenglai, ZHONG Heping. An Improved Lee Filter Al-gorithm for Synthetic Aperture Sonar Image[J]. Ship Elec- tronics Engineering，2017,37(03):35-37.

[41] HUANG Haiyan, WANG Ying. A Improved LEE Filtering Algorithm for SAR Images[J]. Journal of Remote Sensing Information，2010(05):26-29.

[42] LANG Fengkai, YANG Jie,LI Deren.Adaptive Enhance- ment Lee Filtering Algorithm for Polarimetric SAR Images[J]. Journal of Mapping，2014,43(07):690-697.

[43] ZHANG Xiaoyan, TU Honghong·Abdukki Riki. Improve- ment of Threshold Image Denoising Algorithm Based on Wavelet Transform[J]. Journal of Computer Technology and Development,2017,27(03):81-84.

[44] YANG Xiufang, ZHANG Wei,YANG Yuxiang.Detection Technique of Radar Life Signal Based on Lifting Wavelet Transform[J]. Acta Optica Sinica,2014,34 (03): 300-305.

[45] LIU Bosen,ZHANG Ye. Experimental Modal Decomposi- tion and Sparse Representation for SAR Image Denoising[J]. Journal of Harbin Engineering University,2016,37(09):1297-1301.

[46] LI Hongjun. Research and application of image denoising based on multi-scale geometric analysis and partial differ- ential equation[D]. Nanjing:Nanjing University of Aero- nautics and Astronautics,2012.

[47] LIU Rong, LOU Xiaoguang. Improved Algorithm of Polar- ization Lee Filter Based on Edge Characteristics[J]. Sci- ence Technology and Engineering,2011,11(11): 2497-2501.

[48] Arsenault H H,April G. Properties of speckle integrated with a finite aperture and logarithmically trans-formed[J]. Journal of the Optical Society of America,1978,66:1160-1163.

[49] LI Ying, ZHENG Yongguo. An improved and improved algorithm of Lee filtering[J]. Journal of Computer Appli- cations and Software,2012,29(07):243-245.

[50] YANG Dahai, MA Debao. Optimization Algorithm of Po- larization Lee Filter Based on

Polarization Vector Similar- ity Coefficients[J]. Journal of Information Engineering University,2010,11(06):737-740.

[51] Paul Viola, Michael J. Jones. Robust Real-Time Face Detection. 2004, 57(2): 137-154.

[52] Felzenszwalb, Pedro, F, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9):1627-1645.

[53] OUYANG D L, HE S, ZHANG J, et al. Efficient multi-scale attention module with cross-spatial learning[J]. 2023 International Conference on Acoustics, Speech, and Signal Processing, 2023,26(05):127-133.

攻读硕士学位期间发表的论文及成果

发表论文：

[1] 《Infrared weak and small target detection algorithm based on deep learning》- International Journal of Advanced Network, Monitoring and Controls

取得成果：

[2] 参加“全 JHW 特征校核计算及 YS 方案评估研究”科研项目

[3] 参加“伪装效果图像分析”科研项目

致　　谢

本论文的完成离不开许多人的支持与帮助，在此谨向所有给予我指导、关心和支持的人表示最诚挚的感谢。

首先，我要衷心感谢我的导师喻钧老师。老师不仅在学术上给予我细致入微的指导，从论文选题、研究方法到最终撰写都倾注了大量心血，还在我遇到困难和疑惑时耐心引导，为我指明方向。老师严谨的治学态度、敏锐的学术洞察力和宽厚的为人风范让我受益匪浅。

同时，我要感谢计算机学院提供的良好学习和研究环境，以及各位老师在课程教学中给予的悉心讲解和指导。各位同学在讨论中提出的建议和观点，也为我的研究工作带来了诸多启发。特别感谢实验室的各位同学，大家在学习、生活中互帮互助，是我在研究生阶段宝贵的财富。

此外，我还要感谢我的家人。你们始终如一的支持和鼓励是我不断前行的动力源泉。每当我感到迷茫或疲惫时，你们总是给我温暖和力量，使我能重新振作，坚定前行。

最后，感谢在研究过程中曾经提供帮助的所有人。因为你们的帮助，我才能顺利完成本论文的写作。今后我将继续努力，不忘初心，砥砺前行，以更加饱满的热情投入到学习与工作中。

再次向所有给予我关心和帮助的人致以诚挚的谢意！

。

学位论文独创性与知识产权声明

秉承学校严谨的学风与优良的科学道德，本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，学位论文中不包含其他人已经发表或撰写过的成果，不包含本人已申请学位或他人已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了致谢。

学位论文与资料若有不实之处，本人承担一切相关责任。

本人完全了解西安工业大学有关保护知识产权的规定，即：研究生在校攻读学位期间学位论文工作的知识产权属于西安工业大学。本人保证毕业离校后，使用学位论文工作成果或用学位论文工作成果发表论文时署名单位仍然为西安工业大学。学校有权保留送（提）交的学位论文，并对学位论文进行二次文献加工供其他读者查阅和借阅；学校可以在网络上公布学位论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存学位论文。

（保密的学位论文在解密后应遵守此规定）

学位论文作者签名：

校内导师签名：　　日期：

企业导师签名：　　日期：