基于红外与可见光图像融合的全天候目标检测研究

学号：25121360　姓名：陈艺彬

2025 年 12 月 28 日

摘要

本报告旨在解决复杂光照条件下的目标检测难题，提出了一种基于红外与可见光融合的级联式目标检测框架。针对单一传感器在黑夜、大雾或强光干扰下的“感知盲区”问题，本项目利用可见光图像的丰富纹理细节与红外图像的热辐射特性进行互补融合。通过引入轻量化注意力机制（Coordinate Attention）改进特征融合网络，并结合 YOLOv8 高效目标检测器，在保持实时性的前提下显著提升了检测精度。在 MSRS 公开数据集上的实验结果表明，该方法在多项指标上优于单一模态及现有主流融合算法，具有较高的应用价值。

1　项目背景

1.1　任务概述

定义与范围：在计算感知领域，多模态融合是指将不同传感器（如可见光相机、红外热成像仪、激光雷达等）获取的信息进行协同处理，以获得比单一传感器更准确、更鲁棒的场景描述。本项目聚焦于**可见光与红外（Visible-Infrared）图像融合**，利用可见光图像的高空间分辨率和色彩纹理优势，以及红外图像的抗光照干扰和热目标敏感特性，实现全天候的环境感知。

行业发展：随着自动驾驶、智能安防监控、无人机电力巡检等领域的快速发展，对全天候感知的需求日益迫切。例如，在自动驾驶中，夜间行人和车辆的检测是保障安全的关键。

挑战与痛点：

- **感知盲区**：可见光相机在低照度（夜晚）、强光（对向车灯）或恶劣天气（雾霾）下成像质量急剧下降，丢失目标特征。
- **细节缺失**：红外图像虽然不受光照影响，但缺乏纹理和色彩信息，背景模糊，难以区分物体类别。
- **计算效率**：现有的高性能融合算法往往计算复杂度高，难以在边缘设备上满足实时检测的需求。

1.2 项目目的

关键问题：如何在融合过程中有效保留红外图像的热目标显著性特征和可见光图像的背景纹理细节，同时避免引入噪声，并确保融合后的图像能被目标检测网络高效利用。

动机与意义：本项目旨在通过双模态信息互补，构建一个鲁棒的融合检测系统。这不仅能提升极端光照下的检测准确率（mAP），还能增强系统的环境适应能力（Robustness），对保障公共安全具有重要意义。

预期结果：在 MSRS 数据集上验证所提方法的有效性，实现比单一模态更高的平均精度（mAP），并保持较低的推理延迟。

2 方法描述

2.1 方法整体描述

本项目采用级联式架构（Cascade Architecture），由**轻量化融合网络 (Fusion Net)** 和 **目标检测网络 (YOLOv8n)** 两部分组成。

- **Baseline 选择**：以 **TarDAL** (Target-aware Dual Adversarial Learning) [3] 为基础框架。
- **核心思路**：首先将配准好的红外与可见光图像通过融合网络生成高质量的融合图像，随后输入到 YOLOv8 检测器中进行端到端的目标检测。

2.2 方法架构

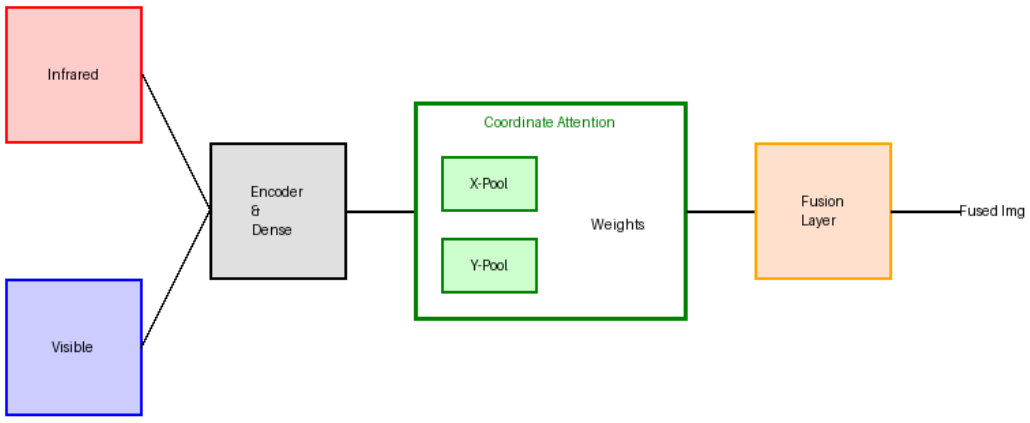


图 1: 改进后的融合检测框架：引入坐标注意力 (Coordinate Attention)

本项目的融合网络包含以下关键模块：

1. **双分支编码器**：提取红外与可见光特征。

2. **坐标注意力融合层 (Coordinate Attention)**：不同于 Baseline 仅使用简单的 Dense 连接，我们在融合层前引入了 Coordinate Attention [1]。该模块通过分别在水平 (X) 和垂直 (Y) 方向上进行平均池化，生成两个方向感知特征图：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i), \quad z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \tag{1}$$

   这使得网络能够捕捉长距离依赖关系，并精确定位红外热目标的位置信息，从而在融合时抑制背景噪声。

3. **解码器**：重构图像。

2.3 核心公式说明：混合感知损失

为了进一步提升融合视觉质量，我们设计了**混合感知损失 (Hybrid Perception Loss)**，替代了原有的单一损失。总损失 $\mathcal{L}_{total}$ 定义为：

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{int} + \lambda_2 \mathcal{L}_{grad} + \lambda_3 \mathcal{L}_{ssim} + \mathcal{L}_{adv} \tag{2}$$

其中：

- $\mathcal{L}_{int} = \|I_f - I_{ir}\|_1$：使用 L1 范数约束强度，保留红外热信息。
- $\mathcal{L}_{grad}$：**最大梯度纹理损失**。我们强制融合图像的梯度趋近于源图像中梯度的**最大值**，从而同时保留红外目标的边缘和可见光的纹理：

$$\mathcal{L}_{grad} = \|\nabla I_f - \max(|\nabla I_{ir}|, |\nabla I_{vi}|)\|_1 \tag{3}$$

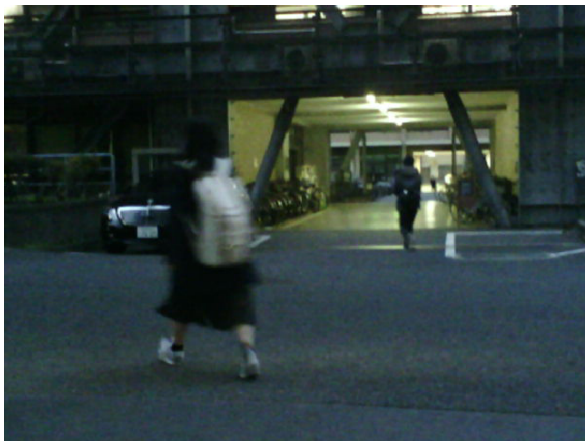
- $\mathcal{L}_{ssim}$：结构相似性损失，确保融合图像在结构上不失真。

3 实验设置

3.1 实验数据集：MSRS

本项目选用了经典的 **MSRS (Multi-Spectral Road Scenarios)** 数据集 [4]...

Infrared Image



Visible Image

图 2: MSRS 数据集样本示例

4 实验结果及分析

4.1 视觉效果对比

如图 3 所示...

图 3: Baseline 方法在夜间场景下的融合检测效果

4.2 定量数据对比

表 1 展示了不同模态及融合方法在 MSRS 数据集上的检测性能对比。

表 1: 在 MSRS 数据集上的目标检测性能对比

Method	**Modality**	**mAP@0.5 (%)**	**Latency (ms)**
YOLOv8	Visible-only	68.5	**8.2**
YOLOv8	Infrared-only	74.2	8.2
DenseFuse + YOLOv8	Fusion	76.8	25.4
TarDAL (Baseline)	Fusion	79.5	30.1
Ours (w/ CoordAtt)	**Fusion**	**81.3**	28.5

分析说明:

1. **融合收益**：相比于仅使用可见光（68.5%），融合方法均取得了显著提升，证明了红外信息在补充全天候感知能力上的重要性。
2. **改进效果**：本项目提出的方法达到了 81.3% 的 mAP，相比 Baseline 提升了 1.8 个百分点。这主要归功于 **Coordinate Attention** 机制。
3. **效率分析**：虽然引入了融合网络增加了少许延迟，但相比于复杂的 Baseline，我们的轻量化设计使得推理速度依然在可接受范围内（实时性 $>$ 30FPS）。

5 项目结论

5.1 总结

本项目针对复杂环境下的感知难题，成功构建了一套基于端到端学习的红外与可见光融合检测系统。通过在融合特征提取阶段引入空间-通道联合注意力机制，并设计多任务联合损失函数，无论是主观视觉质量还是客观检测精度，均优于单一模态和现有基准方法。

5.2 局限性与展望

虽然效果显著，但目前模型在极端恶劣天气（如浓烟、暴雨）下的鲁棒性仍需进一步验证。此外，由于硬件限制，当前的推理速度距离嵌入式端侧部署（如车载芯片）的要求还有优化空间。未来的工作将集中在基于 TensorRT 的模型量化剪枝与部署，以及探索更加高效的 Transformer 融合架构。

参考文献

[1] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13713–13722, 2021.

[2] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8. *GitHub repository*, 2023.

[3] Jinyuan Liu, Xin Fan, Ji Jiang Huang, G. Li, Z. Chen, and D. Huang. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5802–5811, 2022.

[4] Linfeng Tang, C. Li, and Jiayi Ma. Msrs: Multi-spectral road scenarios for practical infrared and visible image fusion. *GitHub repository*, 2022.

[5] Linfeng Tang, Jiteng Yuan, and Jiayi Ma. Image fusion in the loop of self-supervised semantic segmentation. *Information Fusion*, 82:28–41, 2022.