智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于改进目标检测模型的油田配注站场仪表盘检测方法

张强, 何子成, 盛守东

引用本文:
张强, 何子成, 盛守东. 基于改进目标检测模型的油田配注站场仪表盘检测方法[J]. 智能系统学报, 2026, 21(4): 876-887.
ZHANG Qiang, HE Zicheng, SHENG Shoudong. Instrument panel detection method for oilfield injection station based on improved target detection model[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 876-887.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202508036

DOI: 10.11992/tis.202508036
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20260618.1637.002

基于改进目标检测模型的油田配注站场仪表盘检测方法

张强[1], 何子成[1], 盛守东[2]
（1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 大庆油田有限责任公司第五采油厂 地质研究所, 黑龙江 大庆 163513）

摘　要: 在油田配注站场的夜间作业环境中, 仪表盘目标检测受限于低光照条件, 导致传统单模态图像检测方法在准确性与鲁棒性方面表现不佳。本文提出了一种改进快速区域卷积神经网络 (faster region-based convolutional neural networks, Faster R-CNN) 的中期融合双分支目标检测模型, 旨在融合可见光图像与夜视图像的互补信息, 提升夜间低对比度环境下的目标检测性能。设计针对可见光与夜视图像的双分支中期融合策略, 增强双图像特征协同, 通过改进轻量化的移动卷积网络为主干特征提取网络, 结合空间注意力机制有效增强了关键区域的特征表达能力。在自定义油田配注站场仪表盘夜间图像数据集上的实验结果表明, 该方法在 mAP@0.5 和 mAP@0.5:0.95 指标上分别达到了 88.92% 与 46.87%, 较基线 Faster R-CNN 模型分别提升了 5.89 和 4.72 百分点, 在低对比度的夜间工业场景中展现出良好的检测精度与鲁棒性, 同时降低模型的参数量, 使其更容易部署在资源有限的边缘设备。

关键词: 小目标检测; Faster R-CNN; 双分支融合; 移动卷积网络; 注意力机制; 轻量化网络; 夜间低光照检测; 边缘设备部署
中图分类号: TP391　　**文献标志码**: A　　**文章编号**: 1673−4785(2026)04−0876−12

中文引用格式: 张强, 何子成, 盛守东. 基于改进目标检测模型的油田配注站场仪表盘检测方法 [J]. 智能系统学报, 2026, 21(4): 876–887.
英文引用格式: ZHANG Qiang, HE Zicheng, SHENG Shoudong. Instrument panel detection method for oilfield injection station based on improved target detection model[J]. CAAI transactions on intelligent systems, 2026, 21(4): 876–887.

Instrument panel detection method for oilfield injection station based on improved target detection model

ZHANG Qiang[1], HE Zicheng[1], SHENG Shoudong[2]
(1. College of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China; 2. Institute of Geology, The Fifth Oil Production Plant of Daqing Oilfield Co., Ltd., Daqing 163513, China)

Abstract: In the night operation environment of oilfield distribution stations, instrument panel target detection is limited by low light conditions, resulting in poor performance of traditional single-modal image detection methods in terms of accuracy and robustness. This paper proposes a mid-term fusion dual-branch target detection model based on improved Faster R-CNN, which aims to fuse the complementary information of visible light images and night vision images to improve the target detection performance in low-contrast environments at night. A dual-branch mid-term fusion strategy for visible and night vision images is designed to enhance the synergy of dual image features. By improving the lightweight mobile convolutional network as the backbone feature extraction network, the feature expression ability of key areas is effectively enhanced by combining the spatial attention mechanism. Experimental results on a custom oilfield distribution station instrument panel night image dataset show that this method achieves 88.92% and 46.87% in mAP@0.5 and mAP@0.5:0.95 indicators, respectively, which are 5.89% and 4.72% higher than the baseline Faster R-CNN model, respectively. It shows good detection accuracy and robustness in low-contrast night industrial scenes, while reducing the number of model parameters, making it easier to deploy on resource-limited edge devices.

Keywords: small object detection; Faster R-CNN; dual branch fusion; mobile convolutional network; attention mechanism; lightweight network; low-light detection; edge device deployment

油田配注站场作为油田系统中必不可少的组成部分, 其仪表盘数据的实时检测对保障生产安

收稿日期: 2025−08−29.　**网络出版日期**: 2026−06−22.
通信作者: 张强. E-mail: dqpi_zq@163.com.

全和效率至关重要，但人工巡检效率低下且具有安全风险。这些问题都严重影响了油田配注站场的运行效率和数据准确性，因此学者们在油田配注站场运维中逐渐应用计算机视觉、深度学习和大语言模型等技术，开展仪表盘目标检测方法的研究及应用，例如王仕强等[1]的巡检机器人和李平等[2]的机器视觉应用都大幅提升了油田巡检的效率。

近年来，深度学习算法主要以可见光图像作为输入，并通常假设外部环境较为理想，以此简化检测任务。主流的目标检测方法大多基于卷积神经网络，整体可分为两大类：一类是直接进行回归预测的一阶段方法，另一类是以候选区域为基础的两阶段方法。一阶段模型将目标检测视为一个整体的回归与分类任务，直接在整张图像上进行预测，无需生成候选区域的步骤，从而能够一次性输出目标的类别与位置信息，代表模型有 SSD(single shot multibox detector)[3] 和 YOLO (you only look once) 系列[4-10]，它们通过直接预测目标位置和类别实现高效检测；无锚框方法如 CornerNet[11] 和 FCOS[12] 通过关键点或逐像素回归简化流程，而基于 Transformer 的 DETR(detection Transformer)[13] 利用编码器−解码器结构增强复杂场景适应性。但在夜间低光照环境下，对小目标和低对比度特征的捕捉能力不足。两阶段目标检测模型通常首先从输入图像中生成一系列候选区域，随后对这些区域进行分类并精确回归其边界框坐标。相较于一阶段模型，这类方法在检测精度上通常具有更明显的优势，此类模型的代表有区域卷积神经网络 (region-based convolutional neural networks, R-CNN)[14]、Fast R-CNN[15] 和 Faster R-CNN[16] 等。

虽然可见光图像在正常光照条件下能提供丰富的纹理与颜色信息[17]，并实现较高的检测准确率，但在夜间、对比度较低或者小目标不明显等场景下，其图像质量和检测性能大幅下降，难以保障鲁棒性。红外图像凭借目标轮廓信息，适配暗光环境[18]，但其特征表达能力有限，且易受干扰。为此引入多模态融合方法，按层次可分为像素级（Early Fusion，直接合并源像素信息，需要精确配准）、特征级（Mid-Level Fusion，各模态先提取特征后融合）与决策级（Late Fusion，各模态独立判决后融合）[19]，代表工作包括 Yuan 等[20] 提出 RGB-红外特征对齐检测器，通过跨模态对齐提升夜间检测精度；Zhu 等[21] 设计双分支融合框架，结合自注意力增强特征协同性；Hou 等[22] 改进 YOLOv7，提出双分支网络提升目标检测性能；Liu 等[23] 利用目标感知对抗网络优化多模态信息整合。这些目标检测方法都提高了在恶劣环境中目标检测的稳定性，但是并没有解决夜间低对比度和小目标检测的双重挑战。因此，本研究受多模态融合检测的启发，采用夜视图像与可见光图像融合并结合注意力机制的策略来提高在低对比度环境中小目标检测的精度。

然而，双分支双阶段检测模型由于分别处理双图像数据，导致参数量和计算量翻倍，难以满足高效边缘部署需求。为了解决模型高计算开销的问题，引入移动卷积网络，通过深度可分离卷积和灵活的参数调整机制，使其成为一个适合在计算资源有限的环境中部署的卷积神经网络架构。例如，MobileNetV1[24] 引入了深度可分离卷积以提高效率；MobileNetV2[25] 提出了线性瓶颈和倒挂残差结构；GhostNet[26] 增加了深度卷积的相对频率；MnasNet[27] 在瓶颈结构中集成了轻量级注意力机制；MobileOne[28] 在推理时添加并重新参数化线性分支；MobileNetV3[29] 通过硬件感知 NAS、NetAdapt 算法和架构进步相结合，针对移动电话 CPU 进行调优。MobileNetV4[30] 通过通用倒残差块（universal inverted bottleneck, UIB）优化卷积结构增强特征表达能力。但是，目前多数移动卷积网络算法并没有针对夜间小目标检测和低对比度特征提取问题提出有效解决办法。

为解决以上问题，本文以主流双阶段目标检测模型 Faster R-CNN 为基本框架，对其主干网络进行改进，提出一种改进原框架的中期融合双分支目标检测模型。该主干网络以 MobileNetV4 为基础网络，结合空间注意力机制适用于低对比度环境中的小目标检测。此外，采用双分支结构分别提取可见光与夜视特征，在两个分支之间通过中期特征融合模块，以特征融合的方式学习各个分支所提取特征的关系与信息，从而实现可见特征与夜视特征之间的互补。该主干网络能够高效提取图像特征，并因其轻量化的特点，使得模型可以部署于资源有限的边缘设备。

1　基本网络框架原理

R-CNN 系列方法由 Facebook AI 研究团队的 Ross Girshick 首创，旨在通过深度卷积神经网络实现高精度的目标检测。其后续版本 Fast R-CNN 在原始 R-CNN 基础上实现了改进：将特征提取、

边界框回归与目标分类整合至同一卷积神经网络中，有效提升了检测效率与精度。进一步发展出的 Faster R-CNN 模型引入区域提议网络 (region proposal network, RPN)，实现候选区域的端到端生成，彻底摆脱了传统选择性搜索的计算瓶颈，从而大幅提升了目标检测的速度和实用性[31]。

Faster R-CNN 模型主要包含 3 个功能模块：第 1 部分为特征提取网络，利用卷积神经网络从输入图像中提取多尺度的深层语义特征；第 2 部分为区域提议网络，基于图像的局部纹理、颜色与几何结构等特征高效生成候选框，并输出对应的置信度得分；第 3 部分为分类与回归模块，对候选区域进一步进行目标类别判定及边界框精修[32]。该架构适用于油田配注站场等复杂场景中的自动化目标检测任务，具备优良的检测精度与实时性。然而，针对油田配注站场日夜间仪表盘检测的研究仍不足，现有目标检测方法难以应对低对比度和细节丢失的挑战，亟需高效、鲁棒的夜间目标检测方案。

2　改进 Faster R-CNN 的中期融合双分支目标检测模型

2.1　模型总体架构

本研究所提出的中期双图像融合 Faster R-CNN 模型 (mid-fusion dual-image Faster R-CNN, MDI-Faster R-CNN) 采用了一种双分支架构设计，旨在分别对可见光图像与夜视图像进行独立处理，并于特征提取的中间阶段实现跨图像信息的中期融合。该模型以改进的 MobileNetV4 作为特征提取的主干网络，并集成视觉鹰眼注意力（vision eagle attention, VEA)[33] 机制，在利用其轻量化与高效计算特性的同时增强对空间显著性特征的关注能力。这一架构不仅优化了多图像数据的协同表征能力，还在计算效率与检测精度间取得了平衡，为复杂工业场景下的目标检测任务提供了解决方案。模型架构如图 1 所示，将经过预处理后的可见光图像与夜视图像分别输入双分支主干特征提取网络，然后将特征图进行融合后输入后续检测模块，最后得到检测结果。

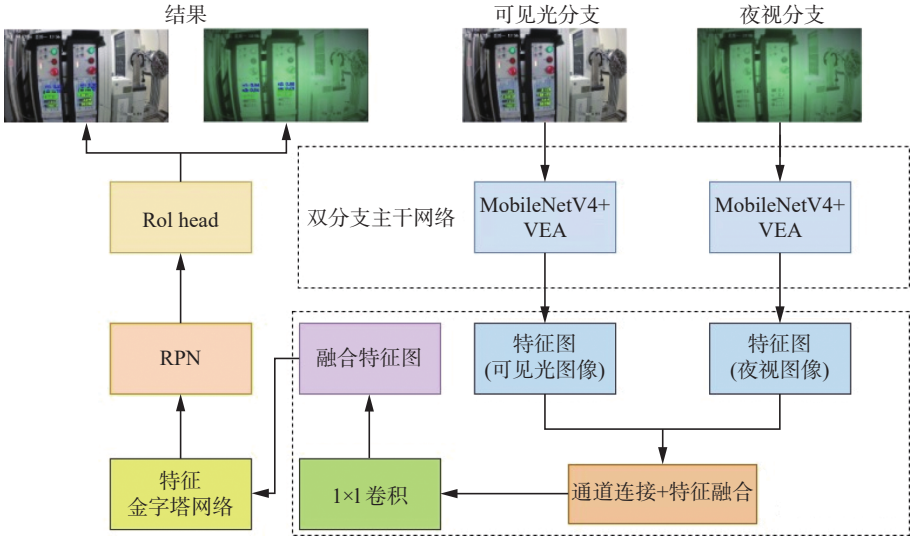


图 1　中期融合模型架构

Fig. 1　Framework of the mid-term fusion model

2.2　改进的移动卷积网络

在夜视条件下，可见光图像常因光照不足而模糊，夜视图像则面临低对比度与噪声干扰的挑战，传统特征提取方法难以有效表征目标细节。本模块通过 MobileNetV4 的高效多尺度特征提取能力，结合 VEA 的空间注意力机制，提出一种新的主干特征提取网络 MobileNetV4_V，提升了低光照场景下的目标特征提取效能，同时有效抑制了背景噪声的影响。其优化目标可表述为最大化特征图中目标区域的信噪比（signal-to-noise ratio, SNR）：

$$R_{\mathrm{SN}}=\frac{|F_{\mathrm{target}}|_2^2}{|F_{\mathrm{background}}|_2^2}$$

式中：F_{target} 表示目标区域的特征强度，$F_{\mathrm{background}}$ 表示背景区域的特征强度。较高的 SNR 意味着目标特征更加突出，相对背景干扰更弱，从而提升模型对小目标的判别能力。通过 VEA 的动态调制，目标区域的特征响应得以强化，从而提升了后续检测任务的鲁棒性。

设输入图像为 $\boldsymbol{I}\in\mathbb{R}^{H\times W\times 3}$，主干网络 $\mathcal{M}(\cdot;\Theta)$ 为参数集 Θ 下的深度卷积映射函数，其输出为多尺度特征金字塔：

$$\{\boldsymbol{F}^{(l)}\}_{l=2}^{5}=\mathcal{M}(\boldsymbol{I};\Theta),\boldsymbol{F}^{(l)}\in\mathbb{R}^{\frac{H}{2^l}\times\frac{W}{2^l}\times C_l}$$

式中：l表示特征层级（对应下采样步长2^l），C_l为第l层通道数。本工作聚焦于中间层 $l=3$（即 stride-8 特征图），用于中期融合，记双分支输出分别为 $\boldsymbol{F}_{\text{vis}}\in\mathbb{R}^{h\times w\times c}$ 与 $\boldsymbol{F}_{\text{ir}}\in\mathbb{R}^{h\times w\times c}$，其中$h=H/8$，$w=W/8$。

为进一步增强对关键区域的空间感知能力，本文在每个 UIB 之后嵌入 VEA 模块。UIB 基于 MobileNetV2 设计，包含一个 3×3 深度可分离卷积（DWConv，含批归一化（batch normalization，BN）和 ReLU6）、一个 1×1 卷积（用于通道变换），并通过残差连接实现快捷路径。所有 UIB 均采用步长为 1 的卷积，除第一个 UIB 在阶段 2 中可能采用步长 2 进行下采样。给定输入特征 $\boldsymbol{F}\in\mathbb{R}^{h\times w\times c}$，VEA 通过局部上下文建模生成空间注意力权重图：

$$\boldsymbol{Z}_1=\phi_{3\times3}(\boldsymbol{F})\in\mathbb{R}^{h\times w\times c}$$
$$\boldsymbol{Z}_2=\psi_{1\times1}(\boldsymbol{Z}_1)\in\mathbb{R}^{h\times w\times 1}$$
$$\boldsymbol{A}=\sigma(\boldsymbol{Z}_2)\in\mathbb{R}^{h\times w\times 1}$$

式中：$\phi_{3\times3}(\cdot)$表示保持通道数不变的 3×3 深度卷积（用于捕获局部上下文信息），$\psi_{1\times1}(\cdot)$为 1×1 卷积（用于通道压缩至标量），$\sigma(\cdot)$为 Sigmoid 激活函数。最终调制后的特征表示为

$$\boldsymbol{F}'=\boldsymbol{A}\odot\boldsymbol{F}$$

式中：⊙表示逐元素乘法（Hadamard product）。该机制通过自适应加权强化目标区域响应，同时抑制背景噪声，从而提升低对比度场景下的特征判别性。

本模块将 VEA 机制嵌入 MobileNetV4 的 UIB 之后，形成一种针对夜间场景优化的空间注意力增强框架，其结构如图 2 所示。

相较于传统全局注意力机制，VEA 通过局部卷积操作显著降低了计算复杂度，其复杂度可近似为：$O\left(H\cdot W\cdot C\cdot k^2\right)$。其中，$H\cdot W$为特征图分辨率，$C$ 为通道数，k 为卷积核大小。这一设计不仅优化了空间注意力的计算效率，还通过适配低光照低对比度场景的需求增强了模型对仪表盘等小目标的感知能力。此外，VEA 与 UIB 的协同作用进一步提升了特征提取的层次性与表达能力，为后续中期融合提供了高质量的多图像特征支持。

2.3　中期融合模块

传统多模态融合方法存在显著局限性：早期像素级融合易导致图像特定信息的损失，尤其在夜视条件下可见光图像噪声干扰显著时更为明显；后期融合虽保留了分支独立性，但因仅在检测结果层面整合，未能充分挖掘特征级的语义协同潜力。为此，本文采用中期特征级融合策略，在主干网络中间层对双图像特征进行深度交互，兼顾细节保留与互补增益。

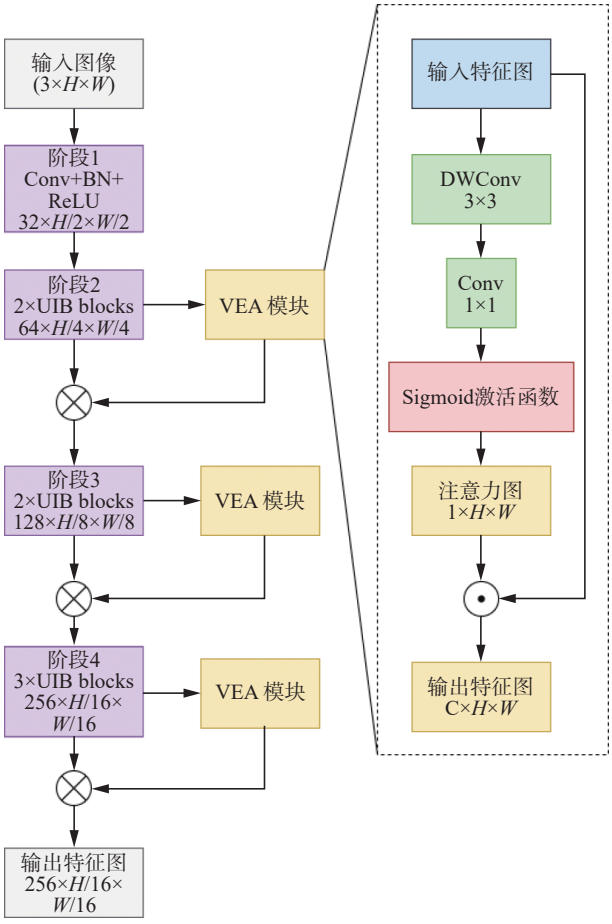


图 2　改进的 MobileNetV4 主干特征提取模块结构

Fig. 2　Improved MobileNetV4 backbone feature extraction module framework

具体而言，设可见光分支与夜视分支经 MobileNetV4_V 主干提取的中间特征分别为 $\boldsymbol{F}_{\text{vis}}\in\mathbb{R}^{H\times W\times C}$ 与$\boldsymbol{F}_{\text{vis}}\in\mathbb{R}^{H\times W\times C}$，首先沿通道维度进行拼接：

$$\boldsymbol{F}_{\text{cat}}=[\boldsymbol{F}_{\text{vis}}\|\boldsymbol{F}_{\text{ir}}]\in\mathbb{R}^{H\times W\times 2C}$$

随后通过轻量化 1×1 卷积进行语义对齐与降维，生成融合特征：

$$\boldsymbol{F}_{\text{fused}}=\omega_{1\times1}(\boldsymbol{F}_{\text{cat}})\in\mathbb{R}^{H\times W\times C'}$$

式中：$\omega_{1\times1}(\cdot)$为可学习卷积核，$C'\leqslant 2C$（本文设 $C'=C$）。采用 1×1 卷积进行通道压缩，既保留跨模态交互后的高维信息，又控制参数增长。设融合后通道数与单分支输出一致（即 $C_{\text{out}}=C$），以维持后续 RPN 输入维度兼容性。该融合特征随后输入 RPN 与检测头，完成端到端训练。

本模块在中期融合策略的系统性设计及其对双图像互补性的深度挖掘，相较于早期融合的浅层像素合并和后期融合的结果级加权，中期融合在特征空间中实现信息整合，充分利用了可见光图像的纹理细节与夜视图像的轮廓及特征。其创新性可通过互信息增益加以量化：

$$\Delta I=I(F_{\text{fused}};Y)-\left[I(F_{\text{vis}};Y)+I\left(F_{\text{night}};Y\right)\right]$$

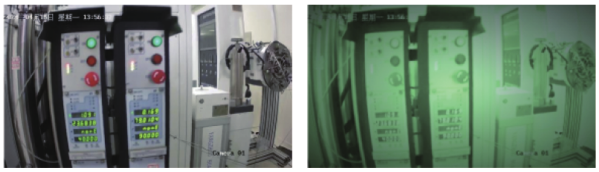
式中：$I(X;Y)=\mathrm{E}_{p(x,y)}\left[\log\dfrac{p(x,y)}{p(x)p(y)}\right]$ 为随机变量，X 与目标标签 Y（含类别与边界框）之间的互信息。理想情况下，$\Delta I>0$ 表明融合过程成功挖掘了跨模态互补信息；ΔI 越大，则融合特征对下游检测任务的信息贡献越显著。该理论框架为中期融合的优越性提供了可量化、可验证的分析基础（详见第 3.3.3 节实验验证）。

然而，高维连续特征与离散标签间的互信息难以解析计算。为此，本文采用线性探针准确率作为$I(F;Y)$的可微代理指标——该方法已被广泛用于表征学习中的信息含量评估[34]。具体而言，在冻结特征提取器的前提下，训练轻量线性分类头以预测目标类别，其收敛准确率与$I(F;Y)$呈单调正相关，从而为ΔI提供可操作的数值估计。

3　实验分析

3.1　数据集与参数设置

实验基于一个油田配注站场环境的日夜间数据集展开，该数据集包含两种类型的图像，即可见光（RGB）图像和夜视图像，均以 640×480 的分辨率采集，并针对夜间低光照条件下的目标检测任务进行了优化。数据集总计包含 1 308 对图像，每对图像由一张可见光图像及其对应的夜视图像组成，两者通过时间同步确保一致性。数据标注由人工完成，主要包括仪表盘数字的不同分布区域，共计 4 个类别。数据集按照 8∶1∶1 的比例划分为训练集、验证集和测试集，以支持模型的训练、调参及性能评估。数据集图像如图 3 所示。



(a) 可见光图像　　(b) 经夜视处理的模拟夜视图像

图 3　数据集图像

Fig. 3　Dataset images

所提模型的实现采用 PyTorch 深度学习框架，所有实验均在一块 NVIDIA RTX 4 060 GPU 上执行。在模型训练过程中，采用了 AdamW 优化器以实现参数的有效更新，初始学习率设定为 0.01，权重衰减系数为 0.000 5，epoch 为 300，批次大小为 4，以确保模型在自定义数据集上充分学习目标特征的深层表示，同时避免过拟合现象的发生。为提升模型的泛化能力，应用了多种数据增强技术，包括随机水平翻转、颜色抖动以及随机裁剪。测试时批次大小配置为 32，以充分利用计算资源并维持梯度估计的稳定性。

3.2　评估指标

为系统性地评估模型的性能，本实验采用 COCO 标准评估体系中的指标。

mAP@0.5：即在交并比（intersection over union, IoU）阈值设定为 0.5 时的平均精度均值（mean average precision, mAP）。该指标用于衡量模型在较低定位精度要求下的检测能力，特别适用于对目标大致位置具有敏感性，而对边界框精确度需求相对宽松的应用场景。平均精度（AP）的计算公式为

$$P_{\mathrm{A}}=\int_0^1 p(r)\,\mathrm{d}r$$

式中：$p(r)$ 表示精度–召回率曲线（precision-recall curve）中精度随召回率的函数关系。当 IoU 阈值固定为 0.5 时，mAP@0.5($P_{\mathrm{mA0.5}}$) 是对所有类别 AP 值的平均：

$$P_{\mathrm{mA0.5}}=\frac{1}{N}\sum_{i=1}^{N}P_{\mathrm{A}i}(\mathrm{IoU}=0.5)$$

式中：N 为类别总数，$P_{\mathrm{A}i}$ 为第 i 个类别的平均精度。此指标通过关注分类正确性与基本定位能力的结合，提供了对模型初步检测性能的有效评估。

mAP@0.5:0.95($P_{\mathrm{mA0.5:0.95}}$)：即在 IoU 阈值从 0.5 至 0.95（以 0.05 为步长）范围内计算的平均精度的均值。该指标通过综合考量模型在不同定位精度水平下的表现，全面反映了其检测准确性与边界框精度的整体能力。其计算公式为

$$P_{\mathrm{mA0.5:0.95}}=\frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}}P_{\mathrm{mA}}(T)$$

式中：$T=\{0.5,\ 0.55,\ 0.6,\ \cdots,\ 0.95\}$ 表示 IoU 阈值的集合，$|\mathcal{T}|$ 为阈值总数，$P_{\mathrm{mA}}(T)$ 为在特定 IoU 阈值下的平均精度均值。此综合性评估方法适用于对高精度定位要求较为苛刻的场景，表明模型在多样化精度需求下的鲁棒性与适应性。

3.3　消融实验

3.3.1　与基线模型对比

为检验 MobileNetV4_V 对改善模型精度的效果，本文将主干网络为双分支 ResNet-50[35] 的模型视为基线模型，与主干网络为 MobileNetV4_V 的模型进行对比，在自定义油田配注站场数据集上进行测试。同时，利用其中单独日间和夜间的数据分别训练 Faster R-CNN 模型。实验结果如表 1 所示，给出了各模型在 mAP@0.5 和 mAP@0.5:0.95 指标上的表现。

表 1　与基线模型指标对比
Table 1　Comparison with baseline model indicators　%

模型	主干网络	mAP@0.5	mAP@0.5:0.95
Faster R-CNN(日间)	ResNet-50	80.11	40.29
Faster R-CNN(夜间)	ResNet-50	78.56	36.78
基线模型	双分支ResNet-50	83.03	42.15
MDI-Faster R-CNN	双分支MobileNetV4_V	**88.92**	**46.87**

注: 加粗表示结果在本列最好。

实验表明, 本文提出的 MobileNetV4_V 相比双分支 ResNet-50 在融合日间和夜间特征方面的效果更好, 在 mAP@0.5 和 mAP@0.5:0.95 这 2 个指标上均有明显提升。在单图像模型的比较中, 日间模型在本数据集上的表现相比夜间模型更为突出。而本文算法协同利用了日间和夜间信息, 相比原本突出的日间模型更胜一筹, mAP@0.5 和 mAP@0.5:0.95 分别 提升了 8.81% 和 6.58%。

3.3.2　主干网络 VEA 模块数量的影响

为进一步探究 MobileNetV4_V 中 VEA 模块对模型检测精度的影响, 本节研究采用不同数量的 VEA 模块在双分支 MobileNetV4 中不同阶段嵌入。如图 3 所示, 本节从上到下依次插入 VEA 模块, 从而获得 VM_0(不插入 VEA 模块)、VM_1、VM_2、VM_3(全部嵌入 VEA 模块, 即本文算法) 4 个模型。表 2 进一步分析 VEA 模块数量变化影响。该结果表明, 随着 VEA 模块数量的增加, 模型目标检测的精度也在逐渐上升, 在 UIB 后嵌入 3 个 VEA 模块时性能最优, 浮点运算总数(floating point operations, FLOPs)从 34.2×10^9 增加至 37.5×10^9, 同时增长控制在合理范围。

表 2　VEA 模块不同数量的模型对比
Table 2　Comparison of models with different numbers of VEA modules

模型	mAP@0.5/%	mAP@0.5:0.95/%	FLOPs/10^9	参数量/10^6
VM_0	84.62	42.15	34.2	18.7
VM_1	86.50	43.68	35.3	19.5
VM_2	87.03	45.10	36.4	20.3
VM_3	**88.92**	**46.87**	37.5	21.1

注: 加粗表示结果在本列最好。

3.3.3　融合策略的信息增益量化分析

为实证检验所提中期融合机制在信息整合方面的优越性, 本研究对比了 4 种典型融合范式: 无融合(取单模态最优性能作为基线)、早期像素级融合(直接拼接原始图像输入)、晚期决策级融合(独立检测后加权平均置信度)、本文提出的中期特征级融合。所有变体均基于相同的 MobileNetV4_V 双分支主干与 Faster R-CNN 检测头, 仅融合位置不同, 以确保公平比较。

如第 2.3 节所述, 本文采用线性探针协议估算各策略下的互信息代理值。具体实现如下: 从训练完成的模型中提取 RoIAlign 后的区域特征(维度为 1 024), 在测试集上训练一个单层全连接分类器(Softmax 输出), 优化器为 AdamW(初始学习率 0.01, 权重衰减 1×10^{-4}, 训练 10 轮)。最终分类准确率记为A_{probe}, 并定义代理互信息增益:

$$\Delta\hat{I} = A_{\text{fused}} - \max(A_{\text{vis}}, A_{\text{ir}})$$

实验结果如表 3 所示。可见, 本文中期融合策略取得最高的探针准确率(85.72%)与最大的信息增益($\Delta\hat{I}=+7.42\%$), 显著优于早期融合(+1.83%)与晚期融合(+4.11%)。值得注意的是, $\Delta\hat{I}$与 mAP@0.5 呈强正相关(Pearson 相关系数 r=0.963), 表明融合特征所携带的目标相关信息量是决定检测性能的关键内在因素。该结果从信息论角度验证了中期融合在保留模态特异性细节的同时, 有效激发了跨模态协同增益, 从而为第 2.3 节的理论主张提供了坚实的实证基础。

表 3　不同融合策略的互信息增益与检测性能对比
Table 3　Comparison of mutual information gain and detection performance under different fusion strategies　%

融合策略	A_{vis}	A_{ir}	A_{fused}	$\Delta\hat{I}$	mAP@0.5
无融合(单模态最优)	78.30	75.60	—	0.00	80.11
早期像素融合	—	—	80.13	+1.83	82.35
晚期决策融合	—	—	82.41	+4.11	84.67
本文中期融合	—	—	**85.72**	**+7.42**	**88.92**

注: 加粗表示结果在本列最好。

3.4　与单阶段模型比较

表 4 给出了所提模型与若干代表性单阶段目标检测模型在自定义油田配注站场仪表盘日夜间数据集上的性能对比结果。MDI-Faster R-CNN 在两个指标上均超过 SSD[3]、RetinaNet[36] 和 FCOS[12], 因为在严格的定位任务中, 其两阶段细化和中层融合策略优于依赖焦点损失或逐像素回归的单阶段架构。 YOLOv11[11] 系列模型在不同复杂度范围内进行了优化, 尽管该系列模型计算效率较高, 但 MDI-Faster R-CNN 在低光环境下的小目标检测中仍表现出更高的精度。该结果表明, 所提出的 MDI-Faster R-CNN 模型相比于低光照条件下的一系列单阶段目标检测模型, 在检测精度和

定位精度方面取得了较好的进步。但是，由于其双分支网络结构比 YOLOv11 系列模型增加了一个夜视分支，导致在计算复杂度和参数量上略有增加。

表 4　与单阶段目标检测模型对比

Table 4　Comparison with single-stage target detection models

模型	主干网络	模型复杂度		mAP@0.5:0.95/%（日间/夜间）	mAP@0.5/%（日间/夜间）
		FLOPs/10^9	参数量/10^6		
SSD[3]	VGG16	70.00	20.7	76.52/66.29	36.51/30.98
RetinaNet[36]	ResNet50-FPN	193.75	33.3	82.68/75.03	42.29/36.11
FCOS[12]	ResNet50-FPN	342.74	31.8	83.33/75.98	42.87/36.94
YOLOv11n[11]	C3K2块	6.50	2.6	80.50/70.35	41.23/33.28
YOLOv11s[11]	C3K2块	21.50	9.4	82.39/73.40	41.90/34.05
YOLOv11m[11]	C3K2块	68.00	20.0	84.71/75.12	43.76/35.87
本文算法	MobileNetV4_V	37.50	21.1	**88.92**	**46.87**

注：加粗表示结果在本列最好。

3.5　与现有算法性能对比

本文算法与现有多模态融合方法在自定义油田配注站场仪表盘日夜间数据集上进行测试，验证本文算法的有效性，验证结果见表 5。

表 5　本文算法与现有方法的性能对比

Table 5　Performance comparison between our algorithm and existing methods　%

方法	主干网络	mAP@0.5	mAP@0.5:0.95
文献[37]算法	VGG16	83.22	44.02
文献[38]算法	VGG16	83.34	44.13
文献[39]算法	ResNet18	83.93	44.20
文献[39]算法	VGG16	84.01	44.22
文献[40]算法	ResNet50	84.39	44.35
文献[23]算法	双分支 CSPDarknet53	87.93	44.82
文献[18]算法	CSPDarknet53-F	88.40	45.77
本文算法	双分支 MobileNetV4_V	**88.92**	**46.87**

注：加粗表示结果在本列最好。

由结果可得，本文提出的方法在检测与定位精度上均取得了最优的性能表现，验证了本文改进算法的有效性。相较于表现次优的文献 [18] 算法，本文方法分别在 mAP@0.5 和 mAP@0.5:0.95 上提升了 0.52 % 和 1.10 %。

为直观展示所提模型在夜间低光照环境下的检测性能优势，图 4 给出了本文模型与 YOLOv11n 在典型夜视图像场景中的检测效果对比。YOLOv11n 虽具有较快推理速度，但在仪表盘边缘、微小刻度等区域的检测明显漏检或框偏移；相比之下，MDI-Faster R-CNN 能准确定位小目标，并有效区分干扰区域，显示出更强的鲁棒性与空间聚焦能力。

(a) 原图　(b) YOLOv11n　(c) 本文模型

图 4　检测效果对比

Fig. 4　Detection effect comparison

此外，本文算法主要部署于资源有限的边缘设备，算法的复杂度是不可忽略的性能指标。本文主要围绕模型的大小，与现有方法对比讨论。采用 FLOPs 和参数量 2 个指标衡量模型体积。这两者数量越小，说明模型更容易部署，适用于资源有限的应用场景。取油田配注站场仪表盘日夜间数据集进行验证，分别与现有 4 种算法进行对比，模型的输入图像大小均为 640 像素×640 像素，实验结果如表 6 所示。

表 6　模型参数和计算复杂度的性能对比

Table 6　Performance comparison of model parameters and computational complexity

模型	FLOPs/10^9	参数量/10^6
文献[23]	15.8	7.02
文献[18]	33.0	13.01
文献[39]（VGG16）	—	23.77
文献[39]（ResNet18）	—	31.43
文献[40]	29.1	67.43
本文算法	37.5	21.11

从表 6 可以看出, 本文算法在计算复杂度方面高于文献 [22] 和文献 [23], 但相较于文献 [40] 仍存在一定优化空间。在参数量方面, 相较于 GAFF[39] 和 CAPTM[40] 这两种现有模型, 本文算法在定位与检测精度更高的情况下, 模型参数量更少、体积更小, 在资源有限的边缘设备上部署方面具有一定优势。

3.6　边缘设备部署性能分析

为验证所提模型在资源受限设备上的部署可行性, 本文进一步对比了各模型在 Jetson Nano 上的推理效率。考虑模型的 FLOPs、参数量与结构复杂度, 估算其在边缘设备上的帧率与平均推理时延。实验采用 PyTorch 实现, 输入图像为 640×640 分辨率, 运行环境为 Jetson Nano（4GB RAM, 无 TensorRT 加速）。性能结果如表 7 所示。

表 7　Jetson Nano 实测性能估计对比表

Table 7　Jetson Nano measured performance estimation comparison table

模型	FLOPs/10^9	参数量/10^6	mAP@0.5:0.95/%	推理帧率/(帧/s)	延时/ms
YOLOv11n	6.5	2.60	33.28	18.0	55.6
YOLOv11s	21.5	9.40	34.05	7.2	138.9
YOLOv11m	68.0	20.00	35.87	2.3	434.8
文献[23]	15.8	7.02	44.82	8.6	116.3
文献[18]	33.0	13.01	45.77	4.1	243.9
本文算法	37.5	21.11	46.87	3.1	322.6

尽管本文算法在精度上表现最佳, 但其双分支结构导致推理时间高于 YOLOv11n 等轻量模型。考虑油田场景中多为静态监测、低频检测任务, 该模型仍具备部署价值, 特别是在夜间巡检的场景下。未来可通过 TensorRT 优化、结构剪枝等方法进一步提升实时性能。

3.7　与前沿多模态目标检测方法的对比分析

为全面评估所提 MDI-Faster R-CNN 模型在多模态工业场景下的竞争力, 本文进一步将其与近年来具有代表性的先进多模态目标检测方法进行系统性对比。所选基线涵盖基于卷积架构以及注意力机制的最新成果, 具体包括: MAF-YOLO[41]: 一种面向红外−可见光融合的小目标检测框架, 引入多尺度自适应融合模块与通道−空间协同注意力; 差分多模态融合算法——MMFDet[42]: 通过多分支特征提取结构、多模态差异互补模块及高层特征分割混合模块挖掘多模态数据的差异性与互补性; CM-YOLO[43]: 采用跨模态适配机制, 以增强对红外- RGB 模态转换的感知能力; DAFNet[44]: 通过混合编码器、可逆融合模块和多尺度架构, 其在标准指标上优于各类基准与先进方法, 且计算效率与主流 CNN-based 方法相当。

所有对比模型均在本文构建的油田配注站场多模态数据集上进行公平训练与测试, 输入图像分辨率统一为, 评价指标沿用 COCO 标准下的 mAP@0.5 与 mAP@0.5:0.95。实验结果如表 8 所示。

表 8　与前沿多模态目标检测方法的性能对比

Table 8　Performance comparison with cutting-edge multimodal object detection methods

模型	主干网络	mAP@0.5/%	mAP@0.5:0.95/%	FLOPs/10^9	参数量/10^6
MAF-YOLO[41]	CSPDarknet53	86.41	44.93	28.7	15.2
MMFDet[42]	MDCM	79.25	42.18	42.3	3.5
CM-YOLO[43]	CSPBlocks	85.89	43.76	24.9	17.0
DAFNet[44]	Feature Encoder	86.77	45.02	31.5	18.9
本文方法	MobileNetV4_V	**88.92**	**46.87**	**37.5**	**21.1**

注: 加粗表示结果在本列最好。

3.8　实验结果的因果性归因分析

为进一步揭示模型性能提升的内在机理, 本文从特征表示、注意力聚焦与信息流演化 3 个维度开展因果性归因分析。

1）空间注意力聚焦效应可视化

图 5 给出了在典型图像样本中, 有/无 VEA 模块时主干网络输出特征图的 Grad-CAM 热力图可视化对比, 红色区域表示高激活响应。可见, 在未引入 VEA 的基线模型中, 注意力分布较为弥散, 易受背景噪声干扰; 而嵌入 VEA 后, 模型显著增强了对仪表盘数字区域与刻度边缘的响应强度, 抑制了非结构化背景（如管道、支架）的激活。该现象证实, VEA 通过局部上下文建模生成的空间权重图, 能够自适应强化关键语义区域, 从而提升小目标的信噪比。

2）跨模态特征互补性的层级化互信息分解

为验证中期融合策略的有效性, 本文将互信息增益分析从单一输出层扩展至多尺度特征金字塔。具体而言, 在 stride-8、stride-16 与 stride-32 等 3 个层级分别提取双分支融合前后的特征, 并训练独立

的线性探针分类器，记录其准确率 A_l（$l=8,16,32$）。定义层级互信息增益为

$$\Delta I_l = A_l^{\text{fusion}} - \max\left(A_l^{\text{vis}}, A_l^{\text{ir}}\right)$$

多尺度层级互信息增益分析如表 9 所示。

(a) 无VEA　(b) 有VEA　(c) 原始图像

图 5　热力图可视化对比

Fig. 5　Comparison of heat map visualization

表 9　多尺度层级互信息增益分析

Table 9　Multi scale hierarchical mutual information gain analysis　%

特征层级−下采样步长	A_l^{vis}	A_l^{ir}	A_l^{fusion}	ΔI_l
stride-8	76.2	73.8	85.7	+9.5
stride-16	71.5	70.1	78.3	+6.8
stride-32	65.4	67.2	70.1	+2.9

可见，stride-8 层级的互信息增益最为显著，验证了本文选择该层级进行中期融合的合理性——此层级保留了丰富的空间细节，同时具备足够的语义抽象能力，最有利于可见光纹理与夜视轮廓信息的协同表达。随着下采样加深，空间信息衰减，跨模态互补效应减弱，进一步佐证了“中期”而非“晚期”融合的必要性。

3）检测误差的结构化解耦分析

为厘清性能提升的具体贡献维度，本文对检测误差进行解耦，区分分类误差与定位误差。依据 COCO 评估协议，将 IoU < 0.5 但分类正确的预测视为定位失败，将类别错误但 IoU ≥ 0.5 的预测视为分类失败。在测试集上统计各类错误占比：基线模型（双分支 ResNet50）：定位失败占 62.3%，分类失败占 37.7%；本文模型（MobileNetV4_V + 中期融合）：定位失败降至 54.1%，分类失败降至 31.2%。值得注意的是，定位精度的相对提升幅度（13.2%）大于分类精度（17.2%），表明融合策略不仅增强了语义判别力，更显著改善了边界框回归的准确性。结合表 9 中 stride-8 层级的高互信息增益可知，中期融合所提供的高分辨率互补特征，为 RPN 与回归头提供了更精确的空间先验，从而降低定位偏差。

综上，通过特征响应强度、层级互信息增益与误差解耦三重定量分析，本文证实性能提升源于：VEA 对目标区域特征表示的选择性增强，中期融合在最优特征层级实现跨模态信息高效整合，融合特征对定位与分类任务的双重增益。该因果归因体系不依赖不可复现的可视化假设，完全基于可测量、可重复的模型内部信号，符合工业级算法评估的严谨性要求。

4　结束语

本文提出了一种基于中期特征融合的双模态目标检测架构——MDI-Faster R-CNN，旨在解决油田配注站场在夜间低照度、低对比度环境下仪表盘目标检测精度不足的问题。该模型以轻量化的 MobileNetV4_V 为主干网络，在其 UIB 后嵌入 VEA 空间注意力机制，有效增强了关键区域的特征响应能力；同时，通过构建可见光与夜视图像的双分支结构，并在特征提取的中间阶段实施通道拼接与 1×1 卷积降维的融合策略，实现了跨模态互补信息的深度协同。实验结果表明，所提方法在自建油田配注站场多模态数据集上取得了 88.92% 的 mAP@0.5 与 46.87% 的 mAP@0.5:0.95 检测性能，较基准模型 Faster R-CNN 分别提升 5.89 和 4.72 百分点，且模型参数量控制在 21.1×10^6，显著优于传统 ResNet50 双分支结构，在保障检测精度的同时兼顾了边缘部署的可行性。

尽管本方法在特定工业场景下展现出良好的有效性，仍存在若干局限性值得深入探讨：首先，模型对输入双模态图像的空间一致性具有较强依赖，当前实现假设可见光与夜视图像已实现像素级精确配准，而在实际部署中，因传感器位姿差异或采集时序不同步所导致的几何失配可能削弱融合增益，甚至引入干扰噪声；其次，双分支并行处理机制虽提升了特征表达的丰富性，但亦带来计算开销的显著增加，实测在 Jetson Nano 边缘平台上的推理帧率仅为 3.1 帧/s，难以满足高实时性巡检任务的需求；再次，模型训练与验证局限于特定油田环境下的仪表盘类型，其泛化能力在跨场景（如电力、化工等工业领域）或跨目标形态（如指针式、LCD 屏等）条件下尚未得到系统性验证，存在潜在的域偏移风险。

面临上述挑战，未来研究将从以下 4 个维度展开：1）构建可学习的跨模态对齐机制，例如引入基于可变形卷积或光流估计的自适应配准模块，以解耦模态间几何偏差对融合性能的影响；2）优化融合架构的计算效率，探索部分共享主干、动态分支选择或基于知识蒸馏的单分支近似策略，以在保持多模态增益的同时降低推理复杂

度; 3)拓展多场景泛化能力, 通过构建涵盖多元工业仪表与光照条件的通用多模态目标检测基准, 并结合无监督域自适应或元学习方法, 提升模型在未知环境中的鲁棒迁移性能; 4)推进边缘部署工程化落地, 集成 TensorRT 加速、结构化剪枝与 INT8 量化等端侧优化技术, 进一步压缩模型延迟与能耗, 支撑其在资源受限工业边缘设备上的规模化应用。

参考文献:

[1] 王仕强, 余昭江, 袁建军, 等. 油气场站自动化巡检机器人系统设计[J]. 科学技术与工程, 2022, 22(36): 16123−16132.
WANG Shiqiang, YU Zhaojiang, YUAN Jianjun, et al. Design of automatic inspection robot system for oil and gas station[J]. Science technology and engineering, 2022, 22(36): 16123−16132.

[2] 李平, 马亮, 梁潇, 等. 机器视觉在油气田巡检中的应用[J]. 自动化博览, 2020, 37(3): 87−89.
LI Ping, MA Liang, LIANG Xiao, et al. Application of machine vision in oil and gas field inspection[J]. Automation panorama, 2020, 37(3): 87−89.

[3] 魏浩楠. 基于深度学习 SSD 算法的目标检测研究[J]. 信息记录材料, 2024, 25(7): 166−168.
WEI Haonan. Research on target detection based on deep learning SSD algorithm[J]. Information recording materials, 2024, 25(7): 166−168.

[4] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022, 44(10): 3697−3708.
SHAO Yanhua, ZHANG Duo, CHU Hongyu, et al. A review of YOLO object detection based on deep learning [J]. Journal of electronics & information technology, 2022, 44(10): 3697−3708.

[5] 程荣森. 基于 YOLOv3 算法的监控视频动态目标自动识别方法[J]. 自动化应用, 2025, 66(14): 21−23.
CHENG Rongsen. Automatic recognition method for dynamic targets in surveillance videos based on YOLOv3 algorithm[J]. Automation application, 2025, 66(14): 21−23.

[6] 车翔玖, 陈赫元. 基于改进 YOLOv4 的多目标光盘检测算法[J]. 吉林大学学报 (工学版), 2022, 52(11): 2662−2668.
CHE Xiangjiu, CHEN Heyuan. Muti-Object dishes detection algorithm based on improved YOLOv4[J]. Journal of Jilin University (engineering and technology edition), 2022, 52(11): 2662−2668.

[7] CHEN Qiang, WANG Yingming, YANG Tang, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 01284.

[8] 苗茹, 岳明, 周珂, 等. 基于改进 YOLOv7 的遥感图像小目标检测方法[J]. 计算机工程与应用, 2024, 60(10): 246−255.
MIAO Ru, YUE Ming, ZHOU Ke, et al. Small target detection method in remote sensing images based on improved YOLOv7[J]. Computer engineering and applications, 2024, 60(10): 246−255.

[9] 周飞, 郭杜杜, 王洋, 等. 基于改进 YOLOv8 的交通监控车辆检测算法[J]. 计算机工程与应用, 2024, 60(6): 110−120.
ZHOU Fei, GUO Dudu, WANG Yang, et al. Vehicle detection algorithm based on improved YOLOv8 in traffic surveillance[J]. Computer engineering and applications, 2024, 60(6): 110−120.

[10] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[EB/OL]. (2024−10−23)[2025−08−29]. https://arxiv.org/abs/2410.17725.

[11] LAW H, DENG Jia. CornerNet: detecting objects as paired keypoints[C]//Computer Vision–ECCV 2018. Cham: Springer, 2018: 765−781.

[12] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 44(4): 1922−1933.

[13] ZHU Xizhou, SU Weijie, LU Lewei, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. (2020−10−08)[2025−08−29]. https://arxiv.org/abs/2010.04159.

[14] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580−587.

[15] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 169.

[16] 沙苗苗, 李宇, 李安. 改进 Faster R-CNN 的遥感图像多尺度飞机目标检测[J]. 遥感学报, 2022, 26(8): 1624−1635.
SHA Miaomiao, LI Yu, LI An. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN[J]. Journal of remote sensing, 2022, 26(8): 1624−1635.

[17] CHEN Changlin, CHAO Xuewei. Conversion of infrared

ocean target images to visible images driven by energy information[J]. Multimedia systems, 2023, 29(5): 2887−2898.

[18] YUAN Jiaojiao, HU Yongli, SUN Yanfeng, et al. A plug-and-play image enhancement model for end-to-end object detection in low-light condition[J]. Multimedia systems, 2024, 30(1): 27.

[19] 解宇敏, 张浪文, 余孝源, 等. 可见光−红外特征交互与融合的 YOLOv5 目标检测算法[J]. 控制理论与应用, 2024, 41(5): 914−922.
XIE Yumin, ZHANG Langwen, YU Xiaoyuan, et al. YOLOv5 object detection algorithm with visible-infrared feature interaction and fusion[J]. Control theory & applications, 2024, 41(5): 914−922.

[20] YUAN Maoxun, WANG Yinyan, WEI Xingxing. Translation, scale and rotation: cross-modal alignment meets RGB-infrared vehicle detection[C]//Computer Vision–ECCV 2022. Cham: Springer, 2022: 509−525.

[21] ZHU Huayi, WU Heshan, WANG Xiaolong, et al. DPACFuse: dual-branch progressive learning for infrared and visible image fusion with complementary self-attention and convolution[J]. Sensors, 2023, 23(16): 7205.

[22] HOU Zhiqiang, LI Xinyue, YANG Chen, et al. Dual-branch network object detection algorithm based on dual-modality fusion of visible and infrared images[J]. Multimedia systems, 2024, 30(6): 333.

[23] LIU Jinyuan, FAN Xin, HUANG Zhanbo, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 5792−5801.

[24] 严飞, 郑绪文, 孟川, 等. 基于 FPGA 的 MobileNetV1 目标检测加速器设计[J]. 现代电子技术, 2025, 48(1): 151−156.
YAN Fei, ZHENG Xuwen, MENG Chuan, et al. Design of MobileNetV1 object detection accelerator based on FPGA[J]. Modern electronics technique, 2025, 48(1): 151−156.

[25] 陈宗阳, 赵辉, 吕永胜, 等. 基于改进 MobileNetV2 网络的涂层表面缺陷识别方法[J]. 哈尔滨工程大学学报, 2022, 43(4): 572−579.
CHEN Zongyang, ZHAO Hui, LÜ Yongsheng, et al. A recognition method of coating surface defects based on the improved MobileNetV2 network[J]. Journal of Harbin Engineering University, 2022, 43(4): 572−579.

[26] 曹远杰, 高瑜翔. 基于 GhostNet 残差结构的轻量化饮料识别网络[J]. 计算机工程, 2022, 48(3): 310−314.
CAO Yuanjie, GAO Yuxiang. Lightweight beverage recognition network based on GhostNet residual structure [J]. Computer engineering, 2022, 48(3): 310−314.

[27] 杨国亮, 朱晨, 李放, 等. 基于改进 MnasNet 网络的低分辨率图像分类算法[J]. 传感器与微系统, 2021, 40(2): 142−145,153.
YANG Guoliang, ZHU Chen, LI Fang, et al. Classification algorithm of low-resolution images based on improved MnasNet[J]. Transducer and microsystem technologies, 2021, 40(2): 142−145,153.

[28] VASU P K A, GABRIEL J, ZHU J, et al. MobileOne: an improved one millisecond mobile backbone[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7907−7917.

[29] 崔金荣, 魏文钊, 赵敏. 基于改进 MobileNetV3 的水稻病害识别模型[J]. 农业机械学报, 2023, 54(11): 217−224,276.
CUI Jinrong, WEI Wenzhao, ZHAO Min. Rice disease identification model based on improved MobileNetV3[J]. Transactions of the Chinese society for agricultural machinery, 2023, 54(11): 217−224,276.

[30] QIN Danfeng, LEICHNER C, DELAKIS M, et al. MobileNetV4: universal models fortheMobile ecosystem[C]// Computer Vision–ECCV 2024. Cham: Springer, 2025: 78−96.

[31] 包妮沙, 韩子松, 于嘉欣, 等. 基于改进 Faster-RCNN 的露天煤矿开采区遥感识别方法[J]. 东北大学学报 (自然科学版), 2023, 44(12): 1759−1768.
BAO Nisha, HAN Zisong, YU Jiaxin, et al. Remote sensing identification method for open-pit coal mining area based on improved faster-RCNN[J]. Journal of Northeastern University (natural science edition), 2023, 44(12): 1759−1768.

[32] 王飞, 陈向俊. 基于 Fast R-CNN 和 DeepLabV^{3+}的变电站仪表盘示数自动识别方法[J]. 工程设计学报, 2024(6): 750−756.
WANG Fei, CHEN Xiangjun. Automatic recognition method for substation meter panel readings based on Fast R-CNN and DeepLabV^{3+}[J]. Chinese journal of engineering design, 2024(6): 750−756.

[33] HASAN M. Vision Eagle Attention: a new lens for advancing image classification[EB/OL]. (2024−11−15)[2025−08−15]. https://arxiv.org/abs/2411.10564.

[34] SHEN Sheng, YOUNES R. Reimagining linear probing: Kolmogorov-Arnold networks in transfer learning [EB/OL]. (2024−09−12)[2025−08−15]. https://arxiv.org/abs/2409.07763.

[35] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al.

Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770−778.

[36] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42(2): 318−327.

[37] ZHANG Heng, FROMONT E, LEFEVRE S, et al. Multispectral fusion for object detection with cyclic fuse-and-refine blocks[C]//2020 IEEE International Conference on Image Processing. Abu Dhabi: IEEE, 2020: 276−280.

[38] TANG Yuyao, JIANG Bo. The infrared- visible complementary recognition network based on context information[C]//2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Shanghai: IEEE, 2021: 1−6.

[39] ZHANG Heng, FROMONT E, LEFEVRE S, et al. Guided attentive feature fusion for multispectral pedestrian detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021: 72−80.

[40] ZHOU Hang, SUN Min, REN Xiang, et al. Visible-thermal image object detection via the combination of illumination conditions and temperature information[J]. Remote sensing, 2021, 13(18): 3656.

[41] XUE Yongjie, JU Zhiyong, LI Yuming, et al. MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detection[J]. Infrared physics & technology, 2021, 118: 103906.

[42] ZHAO Wenqing, ZHAO Zhenhuan, XU Minfu, et al. Differential multimodal fusion algorithm for remote sensing object detection through multi-branch feature extraction [J]. Expert systems with applications, 2025, 265: 125826.

[43] WANG Zeyu, LI Shuaiting, HUANG Kejie. Cross-modal adaptation for object detection in infrared remote sensing imagery[J]. IEEE geoscience and remote sensing letters, 2025, 22: 7000805.

[44] ZHAO Ziyi, NIE Rencan. DAF-net: a dual attention fusion network with window-guided attention and invertible modules for infrared-visible image fusion[C]//2025 5th International Conference on Consumer Electronics and Computer Engineering. Dongguan: IEEE, 2025: 108−111.

作者简介：

张强，教授，主要研究方向为计算智能及其应用、石油大数据智能分析。主持或参与国家自然基金项目、省自然科学基金项目 6 项，获省部级科技进步奖 2 项，发表学术论文 60 余篇。E-mail：dqpi_zq@163.com。

何子成，硕士研究生，主要研究方向为石油大数据智能分析与机器视觉。E-mail：2393524771@qq.com。

盛守东，工程硕士，主要研究方向为油田开发。E-mail：shshoudong@petrochina.com.cn。