



国防科技大学学报

Journal of National University of Defense Technology

ISSN 1001-2486, CN 43-1067/T

《国防科技大学学报》网络首发论文

题目：成像先验驱动超分的倾斜视角目标检测
作者：王堃，刘敏豪，李璋，王梓，滕锡超，裴云祥，于起峰
收稿日期：2025-06-06
网络首发日期：2025-07-29
引用格式：王堃，刘敏豪，李璋，王梓，滕锡超，裴云祥，于起峰. 成像先验驱动超分的倾斜视角目标检测[J/OL]. 国防科技大学学报.
<https://link.cnki.net/urlid/43.1067.t.20250728.1555.004>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

doi: 10.11887/j.issn.1001-2486.25060007

成像先验驱动超分的倾斜视角目标检测

王 堃^{1,2}, 刘敏豪³, 李 璋^{1,2*}, 王 梓^{1,2}, 滕锡超^{1,2}, 裴云祥¹, 于起峰^{1,2}

(1. 国防科技大学 空天科学学院, 湖南 长沙 410073; 2. 国防科技大学 图像测量与视觉导航湖南省重点实验室, 湖南 长沙 410073; 3. 湖南先进技术研究院, 湖南 长沙 410073)

摘 要: 目标检测是遥感图像智能解译的关键技术, 现有方法主要针对卫星下视成像设计, 且对成像先验信息利用不足, 难以有效应对无人机倾斜成像导致的目标尺度显著变化问题。针对倾斜成像的高精度目标检测需求, 提出成像先验驱动超分 (imaging prior-driven super-resolution, IPSR) 方法。该方法利用相机焦距、成像俯仰角和飞行高度等多参数构建成像距离估计模型, 用于估计图像任意位置的成像距离; 基于此距离信息实现了全图自适应超分, 有效减小倾斜成像视角下的目标尺度差异。在多视角成像数据集 VSAI 和卫星图像数据集 DOTA-v1.5 上的实验表明: IPSR 可结合任意主流目标检测模型, 并显著提升目标检测模型在倾斜成像条件下的检测精度 (VSAI 上提升 6-7 mAP, DOTA-v1.5 上提升 3-4 mAP), 同时展现了该方法在多场景下的通用性。

关键词: 旋转目标检测; 超分; 成像先验; 遥感; 倾斜成像视角

中图分类号: TP751 文献标志码: A

Imaging Prior-Driven Super-Resolution for Object Detection under Oblique Views

Wang Kun^{1,2}, Liu Minhao³, Li Zhang^{1,2*}, Wang Zi^{1,2}, Teng Xichao^{1,2}, Pei Yunxiang¹, Yu Qifeng^{1,2}

(1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China; 2. Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, National University of Defense Technology, Changsha 410073, China; 3. Hunan Institute of Advanced Technology, Changsha 410073, China)

Abstract: Object detection is a key technology for intelligent interpretation of remote sensing (RS) images. Existing object detection methods were primarily designed for nadir-imaging satellite RS images. Additionally, their insufficient utilization of imaging prior information makes them struggle to effectively address the significant object scale variations caused by UAV oblique imaging. To address the demand for high-accuracy object detection under oblique imaging perspectives, an imaging prior-driven super-resolution (IPSR) approach was proposed. An imaging distance model was established based on multiple parameters including camera focal length, flight altitude, and pitch angle, enabling the estimation of imaging distances at arbitrary positions in the image. Then, full-image adaptive SR reconstruction was performed based on distance estimation, effectively mitigating intra-image scale disparities under oblique imaging perspectives. Experiments on the UAV multi-view imaging dataset VSAI and satellite imagery dataset DOTA-v1.5 demonstrate that IPSR can be integrated with any mainstream object detection model. It significantly enhances the detection accuracy of these models under oblique imaging conditions (achieving gains of 6-7 mAP on VSAI and 3-4 mAP on DOTA-v1.5). Besides, it exhibits the versatility of the proposed method across multiple scenarios.

Keywords: Oriented object detection, Super-resolution, Imaging prior, Remote Sensing, Oblique Views

* 收稿日期: 2025-06-06

基金项目: 国家自然科学基金资助项目 (12472189), 湖南省科技创新计划资助项目 (2022RC1196)

第一作者: 王 堃 (1996—), 男, 湖南湘潭人, 博士研究生, E-mail: wangkun21@nudt.edu.cn

* 通信作者: 李 璋 (1985—), 男, 湖南长沙人, 副研究员, 博士, 硕士生导师, E-mail: lizhang08@nudt.edu.cn

无人机近年来已成为航空领域的研究热点^[1,2],凭借高效、灵活和低成本的数据采集分析能力,已广泛应用于精准农业^[3]、灾害救援^[4]、情报侦察^[5]及目标搜索打击^[6,7]等领域。目标检测是上述领域的支撑技术,旨在识别无人机图像中的特定目标并回归其精确位置,其性能直接影响无人机的实用效果^[8-10]。

随着深度学习的出现,无人机目标检测近年来受到广泛关注^[11-13]。然而,当前主流的无人机目标检测方法主要源自卫星遥感目标检测技术体系,其设计初衷主要针对下视成像图像^[14-17],在无人机倾斜视角成像下的目标检测性能受限^[18]。不同于下视成像条件,无人机倾斜成像中图像的空间分辨率会随成像视角和高度等因素影响,导致单张图像中同类目标会存在剧烈的尺度差异。此外,成像远端的小目标检测也是难点问题,其特征信息匮乏,导致检测精度进一步降低^[19,20]。

针对目标检测中的尺度变化问题,现有研究主要分为两类方法,即网络级方法与数据级方法。网络级方法聚焦于设计多尺度网络结构以实现多尺度特征表示,例如多分支结构^[21-22]、特征金字塔网络^[23-29]。多分支结构通常将不同尺度的图像输入不同分支,或在各分支使用具有不同感受野的卷积核(例如不同扩张率的空洞卷积),其网络结构一般较为复杂。此外,在训练过程中,多分支结构通常采用静态配置策略,无法自适应调整以实现更好的优化。相比之下,特征金字塔更加简洁高效,已广泛应用于主流目标检测模型^[14-17]。

数据级方法则侧重于通过数据增强策略优化网络性能,例如多尺度训练和超分方法。多尺度训练^[30]通过将随机缩放输入图像至不同大小以实现数据增强;超分方法^[31-33]则旨在从低分辨率图像重建高分辨率图像,以丰富包括目标区域在内的全图结构信息。由于数据级方法与网络级方法解耦,因此可与网络级方法兼容互补。得益于此,先进的目标检测模型普遍结合数据级方法,并在多种场景下展现出优异性能^[14-17]。

然而,上述方法均未考虑无人机成像自有的先验信息(如相机焦距、飞行高度和成像俯仰角)。以数据级方法为例,多尺度训练的缩放范围通常有限(如 $[0.5, 1.5]$)^[34],且需要依赖专家经验针对特定任务进行设计,对小目标检测性能的提升效果有限,且显著增加了计算负担。超分方法对全图进行统一倍率的超分,不仅超分处理会带来高昂的计算复杂度与内存开销,而且重建后的高分辨率图像会进一步加剧检测模型的运算负载与

显存需求。此外,上述方法没有完全解决倾斜视角下的目标大尺度变化的问题,例如,多尺度训练随机缩放图像有可能进一步加剧尺度差异——小目标在随机下采样过程中存在信息丢失风险、大目标经上采样后可能超出有效感知范围,而超分方法的统一尺度放大机制本质上属于各向同性缩放范式,未能解决目标间的相对尺度差异问题。因此,亟需设计一种高效的自适应数据增强策略,以缩小目标间的尺度差异。

鉴于此,本文提出了一种新的数据级方法——成像先验驱动超分(imaging prior-driven super-resolution, IPSR)。该方法基于估计的成像距离实现全图自适应超分。首先,遵循分块裁剪策略^[18,35]将无人机大图像分割成一组固定大小的图像块以降低显存需求;其次,构建基于相机焦距、平台飞行高度、成像俯仰角等多参数的成像距离估计模型,以实现无人机图像任意坐标的成像距离估计;再次,基于成像距离引入超分模型对图像块进行自适应超分,从而减小目标间的尺度差异并恢复目标的结构信息;最后,将重建后的超分图像输入检测网络进行训练和测试。

1 基于 IPSR 的目标检测方法

如图 1 所示,本节详细阐述所提的 IPSR 方法,包括整体网络框架、基于成像先验的距离估计模型、基于成像距离的自适应超分三个部分。

1.1 整体网络框架

针对无人机图像目标检测任务,定义其训练样本为 $\{I_n, S_n, Y_n\}_{n=1}^N$,其中 $I_n \in \mathbb{R}^{H \times W \times 3}$ 、 S_n 、 Y_n 分别表示无人机原图、成像先验信息和目标真值, N 为训练样本总数。由于 GPU 显存限制,常规深度学习框架难以将无人机大图像直接作为输入进行端到端训练。现有方法多采用分块裁剪策略^[18,35],即将无人机大图像分割成一组固定大小的小图像块 $\{I_{n,m}, Y_{n,m}\}_{m=1}^{M_n}$ 作为模型输入。其中 $I_{n,m}$ 表示原图 I_n 的第 m 个图像块,其中心点在 I_n 中的坐标定义为 $c_{n,m}$; $Y_{n,m}$ 表示 $I_{n,m}$ 中包含的目标真值; M_n 表示 I_n 分割后的图像块总数。

因成像角度和高度的差异,同一无人机图像中各图像块的分辨率通常存在差异,导致目标尺度变化显著。如图 1 所示,远端红色框图像块的目标面积可能低于 10 像素,而近端绿色框图像块的目标面积甚至可达 10000 像素。这种显著的目标尺度差异会严重制约检测模型的精度。IPSR 正是针对该问题设计的自适应数据增强策略。

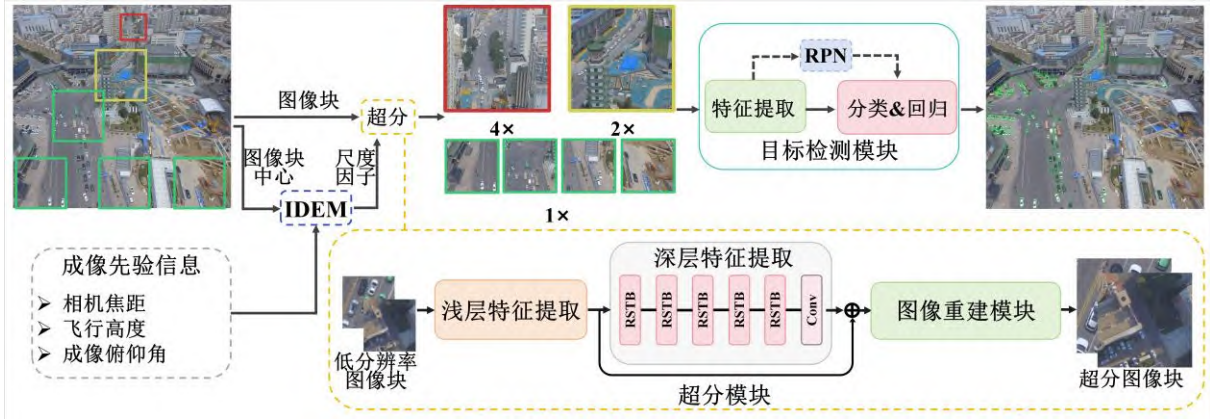


图 1 IPSR 整体结构图

Fig.1 Overall structure of IPSR

IPSR 可基于成像先验信息对图像块进行自适应超分处理, 从而有效缩小目标的尺度差异。作为一种数据增强方法, IPSR 可适用于任意的目标检测模型, 其主要包含两个模块:

1) **成像距离估计模型 (imaging distance estimate model, IDEM)**。基于图像块 $I_{n,m}$ 的中心坐标 $c_{n,m}$ 和成像先验信息 S_n , 通过 IDEM 可计算出各图像块的成像距离 (具体步骤见 1.2 节):

$$d_{n,m} = IDEM(c_{n,m}, S_n) \quad (1)$$

基于成像距离 $d_{n,m}$ 将图像块 $\{I_{n,m}\}_{m=1}^{M_n}$ 划分为近、中和远三组, 分别对应图 1 中的绿色、黄色和红色图像块区域, 从而可以自适应地根据图像块 $I_{n,m}$ 的成像距离设置超分尺度因子 $s_{n,m}$ 。

2) **自适应超分重建模块**。根据尺度因子 $s_{n,m}$ 对图像块 $I_{n,m}$ 执行超分重建 (具体步骤见 1.3 节):

$$\hat{I}_{n,m} = SR(I_{n,m}, s_{n,m}) \quad (2)$$

$$\hat{Y}_{n,m} = T(Y_{n,m}, s_{n,m}) \quad (3)$$

式中, $SR(\bullet, \bullet)$ 表示超分模型, $T(\bullet, \bullet)$ 表示对目标真值进行对应尺度缩放的函数。如图 1 所示, 分别对中距离和远距离图像块执行 2 倍和 4 倍超分, 再将这些超分图像块样本 $\{\{\hat{I}_{n,m}, \hat{Y}_{n,m}\}_{m=1}^{M_n}\}_{n=1}^N$ 输入到目标检测模型中进行训练或推理。

1.2 基于成像先验的距离估计模型

目前主流的无人机成像距离估计技术主要包括激光雷达测距^[36]、深度估计^[37], 但它们普遍存在成本高、计算资源消耗大等问题。激光雷达虽能提供较精确的距离信息, 但其成本随测量距离呈指数级增长, 难以普及。深度估计虽然能从单

幅图像中估计深度, 但其输出的是相对深度而非绝对距离, 需要通过传感器融合或特定标定来恢复尺度信息。此外, 深度估计依赖深度学习模型, 会带来大量的计算资源消耗。因此, 本文提出一种成像距离估计模型 IDEM, 无需复杂的计算过程, 即可快速估计出任意坐标的成像距离。

IDEM 利用无人机以及搭载传感器的元数据获取相机焦距 f 、成像时刻飞行高度 H 及俯仰角 θ 等成像先验信息, 随后基于光心-像点-物点共线的几何关系, 实现从二维图像坐标到三维空间距离的映射计算, 最后估算出无人机图像中任意坐标 (x, y) 的成像距离 d 。需要特别指出的是, 鉴于实验数据集 VSAI^[18] 中偏航角与翻滚角对成像几何的影响可忽略不计, 本研究在建模过程中将这两个姿态角参数默认置零。此外, 鉴于单目视觉在恢复三维信息时存在病态特性, 本研究对成像进行了合理简化: 忽略实际地表微地形起伏, 将成像区域简化为理想平面模型。基于上述成像先验与场景假设条件, 提出的 IDEM 方法如下:

如图 2 所示, 首先定义世界坐标系 $O-XYZ$, 其原点 O 设定于相机光心处, 且其 YOZ 平面与像平面保持垂直。基于相机光心 O 、像点 $p(x, y)$ 与对应三维地面点 $P(X, Y, H)$ 之间的共线关系, 可推导成像距离 $d = \overline{OP}$ 的计算公式:

$$|\overline{OP}| = \frac{H}{z} |\overline{Op}| \quad (4)$$

式中, $|\bullet|$ 表示向量模长, z 表示向量 $|\overline{Op}|$ 在世界坐标系 Z 轴方向的分量。 $|\overline{Op}|$ 可通过下式计算:

$$\overline{Op} = \overline{Oc} + \overline{cp} = \overline{Oc} + \overline{cq} + \overline{qp} \quad (5)$$

式中, c 表示图像中心坐标, $\overline{cq}$ 和 $\overline{qp}$ 分别表示 $p(x, y)$ 在图像坐标系 y 轴和 x 轴方向上的投影。

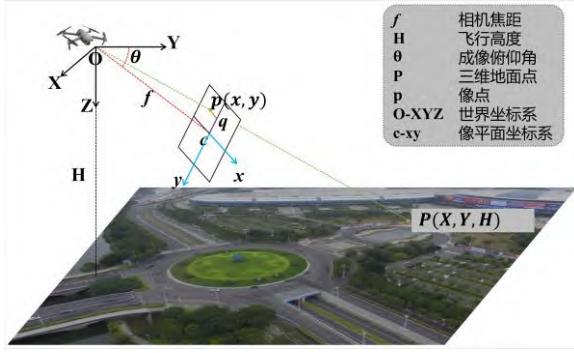


图2 成像距离估计模型示意图

Fig.2 Imaging distance estimation model

向量 $\overrightarrow{Oc}$ 、 $\overrightarrow{cq}$ 及 $\overrightarrow{qp}$ 可分别表示为:

$$\begin{cases} \overrightarrow{Oc} = [0, f \cos \theta, f \sin \theta] \\ \overrightarrow{cq} = [0, -y \sin \theta, y \cos \theta] \\ \overrightarrow{qp} = [x, 0, 0] \end{cases} \quad (6)$$

联立式(4)~(6), 即可建立任意像点坐标 $p(x, y)$ 对应的地面点成像距离 $|\overrightarrow{OP}|$:

$$|\overrightarrow{OP}| = \frac{H}{f \sin \theta + y \cos \theta} \sqrt{f^2 + x^2 + y^2} \quad (7)$$

相比于激光雷达测距依赖额外的高成本设备, 本文提出的 IEDM 只需利用成像先验信息即可估计图像中任意位置的成像距离。相比于深度估计方法需处理整张高分辨率无人机航拍图像的密集计算模式, IDEM 仅需对分割后的图像块进行中心点点距离估算, 显著提升了运算效率。

1.3 基于成像距离的自适应超分

倾斜成像条件下, 远端目标的尺度更小且缺乏清晰的视觉结构特征, 显著增加了目标检测难度。因此, 本文设计了一种基于成像距离的自适应超分策略, 通过动态调整超分重建的尺度因子, 实现对不同距离成像的针对性优化:

$$s = \begin{cases} 1, & \text{if } d < d_1 \\ 2, & \text{else if } d_1 \leq d < d_2 \\ 4, & \text{otherwise} \end{cases} \quad (8)$$

式中, s 和 d 分别表示超分尺度因子和成像距离; d_1 和 d_2 为预设的距离阈值。算法 1 给出了自适应超分的具体步骤: 基于 IDEM 估计的图像块成像距离将图像块划分为近、中和远三组, 对中距离和远距离的图像块分别实施 2 倍和 4 倍超分, 同时保持近距离组图像块的分辨率。该策略可以有效增大远端小目标的尺度, 恢复其结构信息, 从

而减小目标间的尺度差异, 提升检测精度。

IPSR 与具体超分算法解耦, 可适配任意超分模型, 本文选择了当前最具影响力的超分方法 SwinIR^[38] 作为本文实验的超分方法。SwinIR 在多个基准数据集上均展现出最优的性能。如图 3 所示, SwinIR 主要包含三个核心模块:

1) 浅层特征提取模块: 使用卷积层提取浅层特征, 并通过长连接直接传递至重建模块。

2) 深层特征提取模块: 主要由多个 residual swin transformer module (RSTB) 构成。每个 RSTB 利用多个 Swin transformer layer (STL) 实现局部自注意力和跨窗口特征交互, 末端包含一个用于特征增强的卷积层, 并通过残差连接实现特征聚合。

3) 重建模块: 融合浅层特征的低频信息和深层特征的高频信息, 最终生成高保真重建图像。

算法 1 基于成像距离的自适应超分

Alg.1 Adaptive SR based on imaging distance

输入: 图像块样本 $\{\{I_{n,m}, Y_{n,m}, C_{n,m} = (x_{n,m}, y_{n,m})\}_{m=1}^{M_n}\}_{n=1}^N$, 距离阈值 d_1 、 d_2 , 成像先验信息 $\{S_n = \{f_n, H_n, \theta_n\}\}_{n=1}^N$

输出: 超分图像块样本 $\{\{\hat{I}_{n,m}, \hat{Y}_{n,m}\}_{m=1}^{M_n}\}_{n=1}^N$

for $n = 1$ to N **do**

for $m = 1$ to M_n **do**

根据公式(7)计算 $I_{n,m}$ 对应的成像距离 $d_{n,m}$

基于 $d_{n,m}$, 根据公式(8)确定 $I_{n,m}$ 的超分尺度 $s_{n,m}$

根据公式(2)和(3)得到超分图像块样本 $\{\hat{I}_{n,m}, \hat{Y}_{n,m}\}$

end for

end for

return $\{\{\hat{I}_{n,m}, \hat{Y}_{n,m}\}_{m=1}^{M_n}\}_{n=1}^N$

2 实验与分析

2.1 数据集

本节将所提出 IPSR 与其他数据级方法在目标检测任务上进行对比, 并在多个主流基准检测模型上开展实验, 全面评估所提 IPSR 的有效性。实验评估采用目标检测领域最常用的量化指标: 平均精度均值 (mean average precision, mAP)。

2.1.1 无人机多视角成像目标检测数据集

当前主流的无人机目标检测数据集^[2,39-42]均未提供所需的完整成像先验。据调研, VSAI^[18]是当前唯一包含 IPSR 所需成像先验信息的无人机目标检测数据集。本文方法所需的成像先验信息均可通过 VSAI 提供的 EXIF 元数据获取。

VSAI 数据集通过倾斜摄影技术构建而成, 专注于车辆检测任务, 包含大车 (large vehicle,

LV) 和小车 (small vehicle, SV) 两类目标。

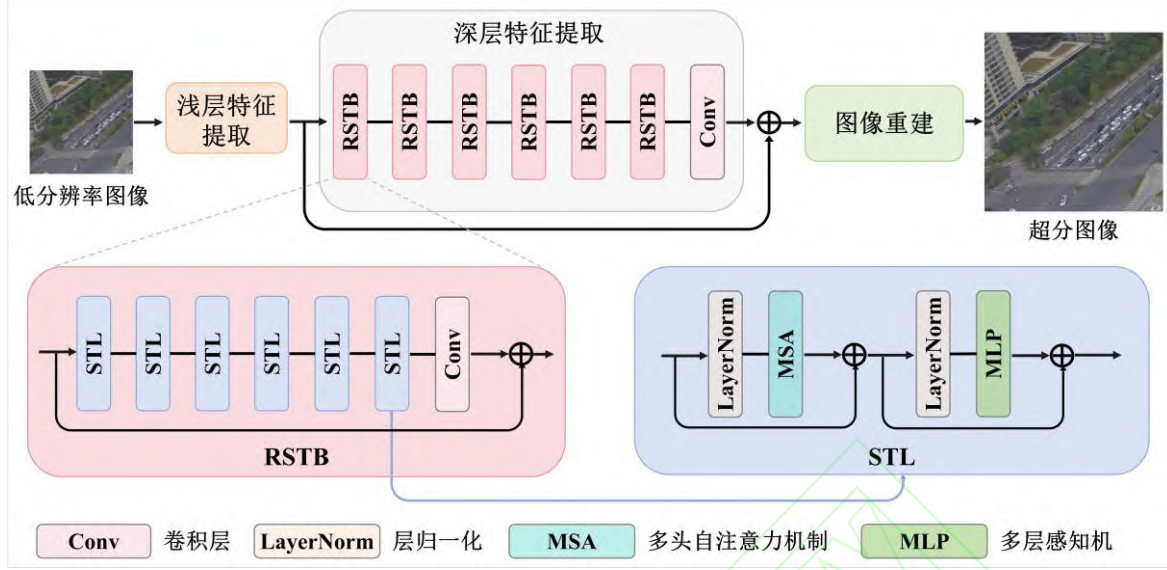


图3 SwinIR 网络结构图^[38]

Fig.3 Network structure of SwinIR^[38]

表 1 VSAI 数据集的相机参数表

Tab.1 Camera parameters of VSAI dataset

无人机型号	相机型号	图像分辨率	像素尺寸/ μm	焦距/mm
MAVIC PRO 2	L1D-20c	5472×3648	2.4	10.3
Mavic Air	FC2103	4056×3040	1.55	4.5
Phantom 4	FC330	4000×3000	1.55	3.6
Phantom 3 Pro	FC300X	4000×3000	1.55	3.6
Mavic Air 2	FC3170	4000×3000	0.8	4.5
Phantom 4 RTK	FC6310R	5472×3648	2.4	8.8

由于 GPU 显存限制以及防止目标被截断, 原始图像以 600 像素步长切割为 800×800 像素的图像块。除目标标注外, VSAI 还提供了完整的相机参数 (如表 1 所示) 与每张图像的成像俯仰角和飞行高度等成像先验信息。基于成像先验信息, 本文通过成像距离估计模型计算所有包含目标的图像块中心的成像距离并按排序, 随后将排序后三分之一和三分之二分位值作为距离阈值 (即式 (8) 中的 d_1 和 d_2)。最后依据式 (8) 为不同距离区间的图像块自适应设置对应的超分尺度因子。

2.1.2 卫星多分辨率成像目标检测数据集

为进一步验证本文方法的通用性, 本文将研究扩展至卫星图像目标检测领域。现有最常用的卫星图像目标检测数据集 DOTA-v1.5^[35] 提供了地面采样距离 (ground sample distance, GSD), 该信息可作为超分重建的尺度先验。

DOTA-v1.5 数据集包含 2,806 张卫星图像、403318 个目标, 涵盖 16 个类别。所有图像以 824 像素步长切割为 1024×1024 的图像块用于训练和测试。如图 4 所示, 目标尺度 (定义为目标区域面积的平方根) 与 GSD 呈现显著相关性: 当 GSD 增大时, 目标尺度分布范围收窄; 当 GSD 减小时, 分布范围显著扩展。特别地, 当 $\text{GSD} > 0.35$ 时出现大量极小目标, 而 $\text{GSD} < 0.35$ 时则包含更多超大目标。因此, 本文选择 0.35 作为 GSD 阈值, 对 $\text{GSD} > 0.35$ 的图像实施 2 倍超分重建。

2.2 实验设置

本文所有实验均基于以下硬件与软件配置开展: Intel i9-10920X 处理器、NVIDIA RTX 4090 显卡、Ubuntu 18.04 操作系统。

超分重建: 采用 SwinIR 模型^[38] 进行训练, 优化器选用 Adam 算法, 初始学习率、批处理大小、总迭代次数分别设为 5×10^{-5} 、8、500000。RSTB 的数量及其内部层数均设为 6。本文从 VSAI 数据集中精选出 470 张含清晰车辆纹理的图像块作为训练集。为适配后续处理, 将原始图像进一步裁剪为 1880 张 400×400 像素的图像块。参照 BSRGAN^[43] 的图像降质处理, 对每个图像块生成对应的低分辨率 (low resolution, LR) 图像, 构建 LR-SR 图像对数据集用于模型训练。

旋转目标检测: 基于 MMRotate 框架^[34], 本文在主流旋转目标检测框架上开展实验。训练设

置和检测模型均遵循 MMRotate 的默认配置。

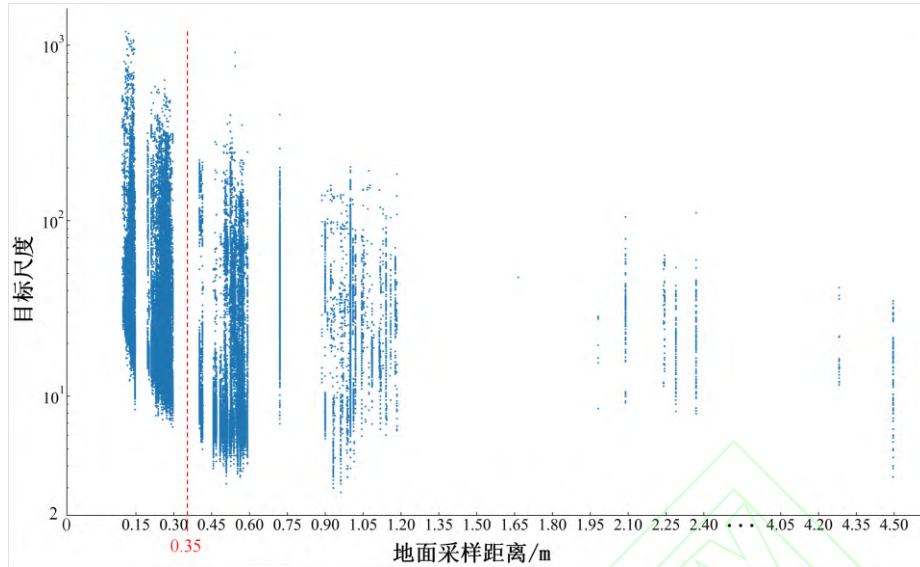
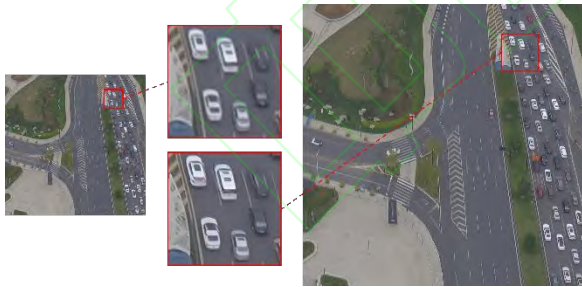


图4 DOTA-v1.5 数据集的目标尺寸与地面采样距离散点图

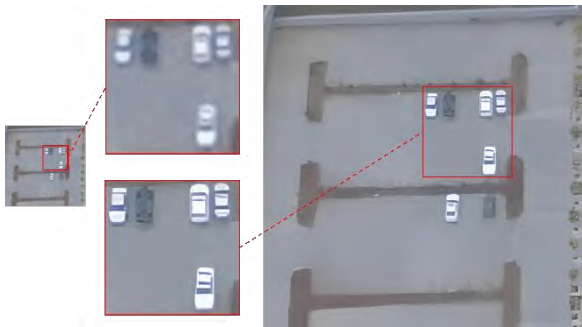
Fig.4 Scatter plot of object scale and ground sampling distance in DOTA-v1.5 dataset

2.3 超分实验可视化结果

图 5(a)和图 5(b)分别展示了 SwinIR 在 2 倍和 4 倍超分重建的可视化效果。左侧为原始图像，右侧为 SwinIR 输出的超分图像。通过视觉对比可以观察到，超分图像中车辆、道路标线及停车场区域等关键目标的细节清晰度较原始图像显著提升，其纹理结构与边缘轮廓均得到明显增强。对比分析图 6 可见，相比双线性插值，SwinIR 在保持图像结构完整性的同时，有效恢复了高频纹理细节并抑制了模糊现象。



(a) 2×



(b) 4×

图5 SwinIR 的可视化结果

Fig.5 Visualization results of SwinIR



(a) Bilinear

(b) SwinIR

图6 Bilinear 和 SwinIR 结果对比。

Fig.6 Comparison of results between Bilinear and SwinIR

2.4 旋转目标检测结果与分析

2.4.1 VSAI 数据集检测结果对比

为验证 IPSR 的有效性，将其与多种典型数据级方法开展对比实验。参与对比的方法简述如下：1) Baseline：表示标准单尺度训练测试的传统方法；2) Bilinear 和 SR：分别表示对所有图像块进行 2 倍尺度因子的双线性插值和超分重建方法；3) IPBilinear 和 IPSR：分别表示基于成像先验驱动的双线性插值和超分重建

方法。实验在 R3Det^[14]、S2Anet^[15]、Oriented RCNN^[16] 和 RoI Transformer^[17] 及 RoI Transformer+（将主干网络替换为 Swin transformer^[44]）等主流目标检测模型上进行。

实验结果如表 2 所示（表中各指标最优值已加粗标识），基于 IPSR 的目标检测模型均取得最高 mAP 值。其中，R3Det 结合 IPSR 后精度提升最显著（较 Baseline 提升 7.4 mAP）。在多数检测模型上，超分方法（SR 和 IPSR）的精度均优于双线性插值（Bilinear 和 IPBilinear），这主要由于超分重建在扩大目标尺寸的同时，能更有效地恢复小目标的纹理信息，有助于检测模型提取更具判别性的特征。

另一方面，基于成像先验驱动的方法（IPSR 和 IPBilinear）的检测精度优于无成像先验驱动的方法（SR 和 Bilinear），以 Oriented RCNN 模型为例，IPSR 较 SR 精度提升达 4.1 mAP。图 7 展示了 VSAI 数据集中原始图像、执行 IPSR 后和执行 SR 后的目标尺度分布对比的箱线图，其中目标尺度定义为目标区域面积的平方根。由于 SR/Bilinear 方法对所有图像进行固定倍数的超分或缩放，因此目标尺度差异保持不变。相比之下，IPSR 和 IPBilinear 则自适应的对不同成像距离的目标进行超分，因此其处理后的目标尺度分布更为紧凑，具体表现为：1）极值比显著降低：目标尺度的最大值与最小值之比显著降低（SV: 13.85 vs. 8.36; LV: 17.98 vs. 8.98）。2）分布集中度提升：箱体长度（即四分位距范围）更短，第三四分位数与第一四分位数的比值明显减小（SV: 2.03 vs. 1.58; LV: 2.52 vs. 1.73）。这表明 IPSR/IPBilinear 有效减小了同类目标间的尺度差异。

为进一步验证 IPSR 对目标定位精度的影响，本文采用更严格的 mAP75 评价指标进行结果评估。如表 2 右侧结果所示，IPSR 在各类别 AP75 及 mAP75 指标上均优于其他数据级方法。

2.4.2 DOTA-v1.5 数据集检测结果对比

DOTA-v1.5 数据集上的实验结果如表 3 所示，其中“-”表示基于 IPSR 的目标检测方法与 Baseline 方法在各类别 AP 指标的差值。由实验结果可知：1）IPSR 使得主流检测模型均实现显著的精度提升，增幅约 3-4 mAP。2）IPSR 对小目标（如小车 SV、大车 LV、储罐 ST）的提升尤为显著。例如，Oriented RCNN 和 RoI

Transformer 在 SV 类别上的提升分别达到 14.54 AP 和 14.50 AP。图 8 通过箱线图对比了 DOTA-v1.5 数据集中原始图像与经 IPSR 执行后各类目标的尺度分布，IPSR 处理后各类目标的尺度差异显著减小，同时小尺度目标的数量也明显减少，因此可以有效提升目标检测的精度。以 SV 类别为例，IPSR 处理后的目标尺度最小值由 2.83 提升至 6.32，最大值与最小值之比由 10.76 降低至 3.90。3）基于 IPSR 的模型在个别类别出现了检测性能的轻微下降，尤其在集装箱 CC 类别表现普遍较差，这主要归因于该类别样本量极少（仅 142 个），导致检测模型学习到的特征区分度不足。

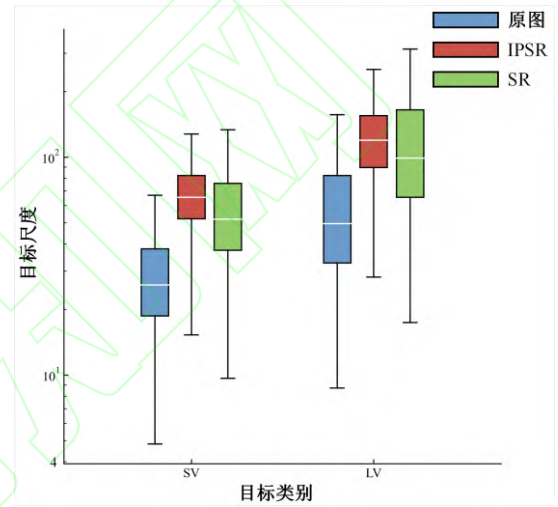


图 7 VSAI 数据集中目标尺度的分布对比

Fig.7 Comparison of object scale distributions across categories in the VSAI dataset

2.4.3 检测模型推理速度对比分析

本文采用平均每秒帧数（frames per second, FPS）作为模型推理的实时性评价指标。以 Oriented RCNN 为基准，Baseline、Bilinear、SR、IPBilinear、IPSR 的 FPS 分别为：43.0、19.6、19.8、24.9、23.3。相较于 Baseline，IPSR 因处理更高分辨率图像块导致 FPS 略有下降，但相较于无成像先验驱动的方法（SR 和 Bilinear），IPSR 在缩短运行时间的同时取得了更优的检测效果，主要因为无成像先验驱动的方法（SR 和 Bilinear）对所有图像块进行统一上采样，相比 IPSR 会产生更多高分辨率的图像，导致运算复杂度提高。虽然 IPSR 的 FPS 略低于 IPBilinear，但在检测精度与推理速度之间实现了更好的平衡。

表 2 IPSR 和其他数据级方法在多个主流检测模型上的比较

Tab.2 Comparison of IPSR and other data-level methods on multiple mainstream detectors

检测模型		评估指标(AP50)					评估指标(AP75)				
		Baseline	Bilinear	SR	IPBilinear	IPSR	Baseline	Bilinear	SR	IPBilinear	IPSR
R3Det	SV	78.4	84.5	85.4	84.8	86.0	48.1	58.7	58.8	58.5	59.3
	LV	51.5	58.0	58.2	57.4	58.9	30.4	37.9	38.3	37.5	41.3
	mAP	65.0	71.2	71.8	71.1	72.4	39.3	48.3	48.5	48.0	50.3
S2ANet	SV	79.3	84.9	85.2	85.2	85.8	48.4	56.7	56.9	57.2	58.5
	LV	52.1	56.1	57.3	57.2	58.7	25.9	28.6	29.9	29.4	33.1
	mAP	65.7	70.5	71.2	71.2	72.3	37.1	42.7	43.4	43.3	45.8
Oriented RCNN	SV	79.7	80.4	80.4	80.4	85.9	58.9	61.1	66.3	61.1	67.1
	LV	53.5	57.5	58.0	56.9	60.7	36.7	38.7	38.9	40.4	44.0
	mAP	66.6	69.0	69.2	68.7	73.3	47.8	49.9	52.6	50.8	55.6
RoI Transformer	SV	79.8	80.4	80.5	80.6	80.3	58.8	60.4	60.9	60.9	66.2
	LV	55.1	55.9	56.8	57.9	60.8	37.6	35.4	40.4	43.0	44.1
	mAP	67.5	68.1	68.7	69.2	70.6	48.2	47.9	50.6	52.0	55.1
RoI Transformer+	SV	80.2	87.3	86.4	87.1	87.0	59.9	67.7	66.1	67.8	67.9
	LV	60.5	63.0	60.6	63.0	64.1	40.9	46.5	44.0	47.5	48.2
	mAP	70.4	75.2	73.5	75.1	75.6	50.4	57.1	55.0	57.6	58.0

表 3 IPSR 和 Baseline 在 DOTA v1.5 数据集上的检测结果比较

Tab.3 Comparison of IPSR and baseline methods on multiple mainstream detectors in DOTA v1.5

Detectors		PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	CC	mAP
R3Det	Baseline	77.94	76.99	44.07	67.62	53.01	69.70	79.73	90.82	72.22	70.57	45.61	71.19	58.70	63.53	35.88	0.00	61.10
	IPSR	80.69	79.03	47.34	70.84	61.59	70.87	80.20	90.82	77.94	77.62	49.60	73.17	60.17	63.08	53.89	0.00	64.80
	-	+2.75	+2.04	+3.27	+3.22	+8.58	+1.17	+0.46	+0.00	+5.73	+7.05	+3.99	+1.98	+1.47	-0.45	+18.01	+0.00	+3.70
S2ANet	Baseline	77.23	77.56	52.49	70.59	57.64	74.66	84.25	90.88	76.95	75.14	55.38	72.08	66.02	63.46	45.20	0.01	64.97
	IPSR	80.65	77.54	53.20	73.40	65.10	78.15	88.20	90.83	80.36	83.45	59.24	72.19	66.52	66.07	62.99	0.00	68.62
	-	+3.42	-0.02	+0.71	+2.81	+7.46	+3.49	+3.95	-0.05	+3.41	+8.31	+3.86	+0.11	+0.50	+2.61	+17.79	-0.01	+3.65
Oriented RCNN	Baseline	78.91	78.45	53.82	71.99	52.48	76.28	87.42	90.77	77.86	68.26	59.89	72.13	67.58	66.13	56.25	4.49	66.42
	IPSR	80.66	83.37	54.62	70.22	67.02	81.17	88.41	90.74	81.59	77.09	58.33	73.70	67.57	63.95	68.31	6.13	69.56
	-	+1.75	+4.92	+0.80	-1.77	+14.54	+4.89	+0.99	-0.03	+3.73	+8.83	-1.56	+1.57	-0.01	-2.18	+12.06	+1.64	+3.14
RoI Transformer	Baseline	78.85	82.79	55.28	70.04	52.69	76.55	81.02	90.87	78.42	69.30	62.77	73.84	74.54	66.43	55.43	10.72	67.47
	IPSR	80.51	82.30	55.71	74.34	67.19	80.63	88.81	90.84	82.68	82.80	63.82	75.66	74.24	71.67	64.42	1.57	71.07
	-	+1.66	-0.49	+0.43	+4.30	+14.50	+4.08	+7.78	-0.03	+4.26	+13.50	+1.05	+1.82	-0.30	+5.24	+8.99	-9.15	+3.60

2.4.4 不同尺度下的检测结果比较

为验证基于 IPSR 目标检测模型对不同尺度目标的有效性,本文以 RoI Transformer 为基准模型,针对 VSAI 数据集,根据目标尺度范围比较目标检测的 AP 和平均召回率 (average recall, AR)。实验将目标尺度范围划分为三组:小尺度(0-16×16 像素)、常规尺度(16×16-128×128 像素)和大尺度(128×128 像素以上)。

目标尺度划分的依据如下:1)目标检测模型锚框设置的最小尺度为 32×32 像素,导致 0-16×16 像素范围内目标难以分配合适锚框作为正样本;2)为避免因大尺度目标样本数量不足影响评估结果,设置 128×128 像素作为大尺度阈值,确保足够数量的大尺度目标参与评估。

如表 4 所示,基于 IPSR 的 RoI Transformer 在所有尺度范围内均取得最优 AP 性能,且 AR

指标亦达到最优或次优水平（仅比最优结果低 0.2）。从表中可以看出，在小尺度范围内，基

表 4 IPSR 与其他数据级方法对不同尺度目标的检测精度比较

Tab. 4 Comparison of IPSR and other data-level methods on object detection with different scales

目标尺度		真值 数量	AP					AR				
			Baseline	Bilinear	SR	IPBilinear	IPSR	Baseline	Bilinear	SR	IPBilinear	IPSR
Small	SV	16552	56.8	69.2	71.0	70.0	71.4	67.8	80.5	82.4	83.7	83.5
	LV	137	4.4	3.6	11.0	14.4	16.5	18.3	21.9	23.4	29.2	34.3
	mAP/mAR	16689	30.6	36.4	41.0	42.4	44.0	43.0	51.2	52.7	56.5	58.9
Normal	SV	48364	88.4	89.0	89.1	88.6	89.1	90.8	91.7	91.6	91.8	91.8
	LV	1727	56.7	60.9	61.9	61.1	62.7	69.4	70.3	72.5	72.7	73.3
	mAP/mAR	50091	72.6	75.0	75.5	74.9	75.9	80.1	81.0	82.1	82.2	82.6
Large	SV	1252	90.8	90.4	90.8	90.8	90.8	92.7	90.4	91.7	95.7	95.5
	LV	167	64.1	48.6	57.5	72.8	75.9	79.0	61.1	71.9	80.2	84.4
	mAP/mAR	1419	77.5	69.5	74.2	82.8	83.3	85.8	75.8	81.8	88.0	90.0

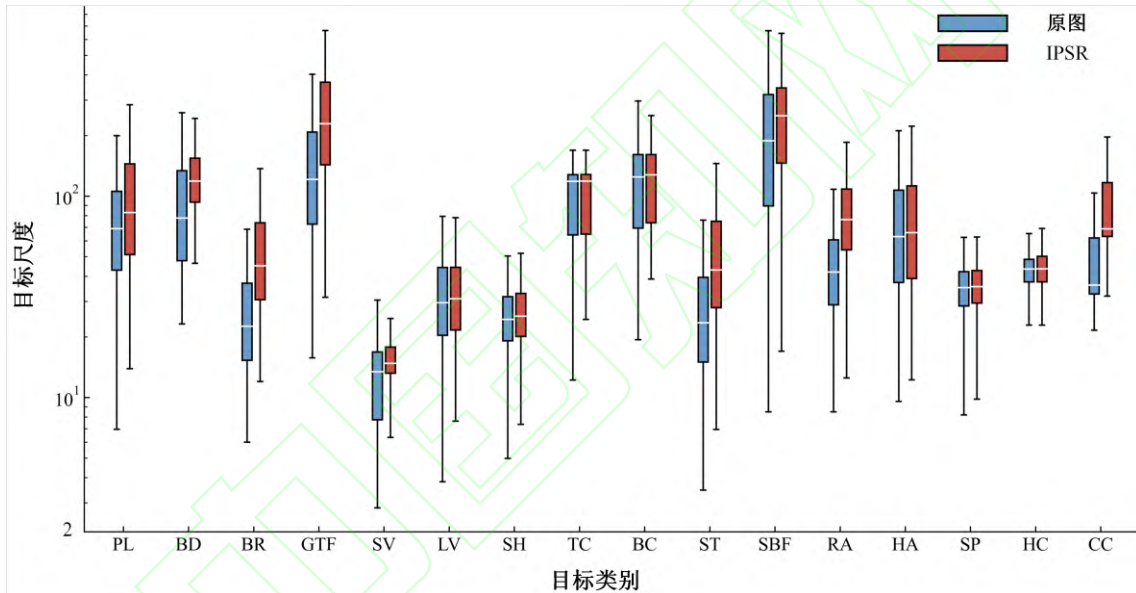


图 8 DOTA-v1.5 数据集各类别目标尺度分布对比（IPSR 实施前后）

Fig.8 Comparison of scale distribution of multi categories in DOTA-v1.5 dataset (before and after IPSR implementation)

于超分重建的方法(SR 和 IPSR)相比 Baseline, SV 提升约 13 AP 和 14 AR, 主要因为超分模块提升了小目标尺度, 有助于检测模型为其分配更合适的锚框, 并使特征提取的感受野更适配。在大尺度范围内, 无成像先验驱动的方法 (Bilinear 和 SR)的 LV 检测精度低于 Baseline, 这主要是由于检测模型的有效感受野小于某些大目标的尺度^[45], 而基于成像先验驱动的方法 (IPBilinear 和 IPSR) 通过自动选择目标适配尺度, 其性能显著优于 Baseline 和无成像先验驱动的方法 (Bilinear 和 SR)。在常规尺度范围内, 各方法性能接近, 因目标尺度均处于合理

匹配区间。

基于超分重建的方法 (SR 和 IPSR) 在 AP 和 AR 指标上均优于双线性插值方法 (Bilinear 和 IPBilinear), 这得益于超分技术可增强目标外观细节, 使检测模型提取更具判别性的特征。此外, LV 的检测性能显著低于 SV, 主要由于 VSAI 数据中 LV 样本数量远少于 SV, 少量误检或漏检对目标检测精度会产生显著影响; 此外, 小尺度 LV 目标外观特征匮乏, 导致预测结果准确性不足。

2.5 目标检测结果可视化

如图 9 所示, 本文通过可视化分析展示了部分远距离场景的检测效果。相较于 **Baseline**, **IPSR** 减少了小尺度目标的漏检数量。此外, 该方法还成功检测到众多模糊且特征不明显的 **SV** 目标, 验证了其在弱小目标检测中的有效性。尽管 **Bilinear**、**SR**、**IPBilinear** 能缓解漏检问题, 但其漏检率仍高于 **IPSR**。

为进一步解释本文方法的有效性, 采用 **ODAM**^[46]生成可解释性热力图。该热力图通过计算预测目标相对于特征图的梯度获得, 可反映不同像素对模型决策的影响程度。图 10 对比展示了 **Baseline** 与基于 **IPSR** 的 **RoI Transformer** 检测结果及对应热力图。相较于 **IPSR**, **Baseline** 模型存在大量误检和漏检问题, 且其热力图在背景区域呈现异常响应 (如蓝色框标示的误检 **LV**)。针对同一目标, 两种方法的关注区域也存在显著差异, **Baseline** 模型的热力图区分度较低, 导致目标定位出现偏差 (如第 1 行左侧和第 3 行右侧箭头所示的大型车辆); 基于 **IPSR** 的方法, 其热力图聚焦于目标核心区域, 实现了精准定位和可靠分类。

3 结论

本文提出一种基于成像先验驱动的超分重建方法 (**IPSR**), 用于解决无人机倾斜视角

下的目标尺度变化问题。该方法通过成像先验信息估算各图像块的成像距离, 从而自适应确定超分重建的尺度因子, 最后将重建后的高分辨率图像输入到目标检测网络。在无人机多视角成像目标检测数据集 **VSAI** 和卫星多分辨率成像目标检测数据集 **DOTA-v1.5** 上的对比实验表明, 基于 **IPSR** 的目标检测方法相较于基线方法表现出显著性能优势, 充分验证了成像先验在提升目标检测精度方面的有效性。

局限性与未来研究方向。 尽管 **IPSR** 展现出优越检测性能, 但其超分处理和输入图像放大过程不可避免地增加了计算开销。此外, 超分网络与目标检测网络采用分立训练策略, 未联合优化, 限制了检测性能。针对上述问题, 未来研究可从以下三个方向展开: 1) 探索端到端的联合框架, 将超分网络与目标检测网络整合为统一结构, 实现协同训练; 2) 将超分网络设计为辅助分支, 用于学习低分辨率图像的高分辨率特征表示, 使其可作为即插即用模块嵌入现有检测框架; 3) 拓展成像先验驱动方法的应用边界, 探索其在目标跟踪、语义分割等其他计算机视觉任务的泛化能力。特别地, 针对上述任务, 可构建包含成像参数标注的数据集, 为成像先验的深度利用提供数据支撑

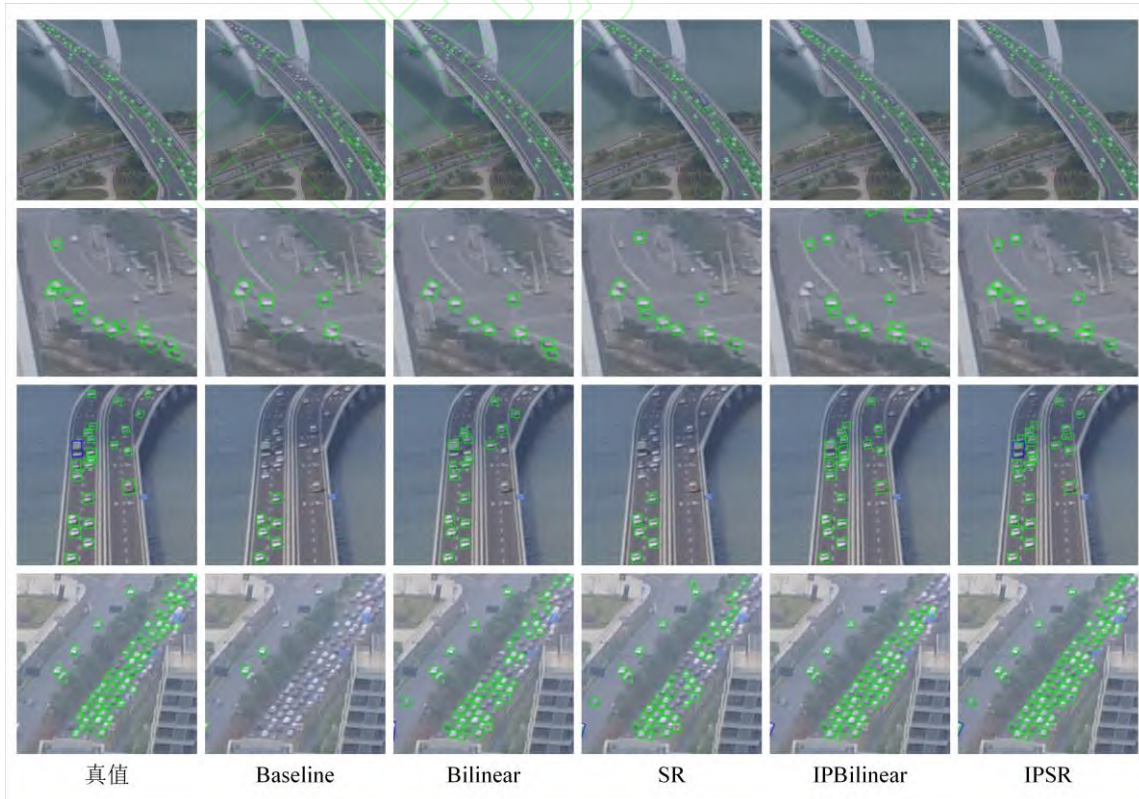


图 9 不同数据级方法在远端场景的检测结果可视化对比图

Fig.9 Visual comparison of detection results of different data level methods in distant scenes

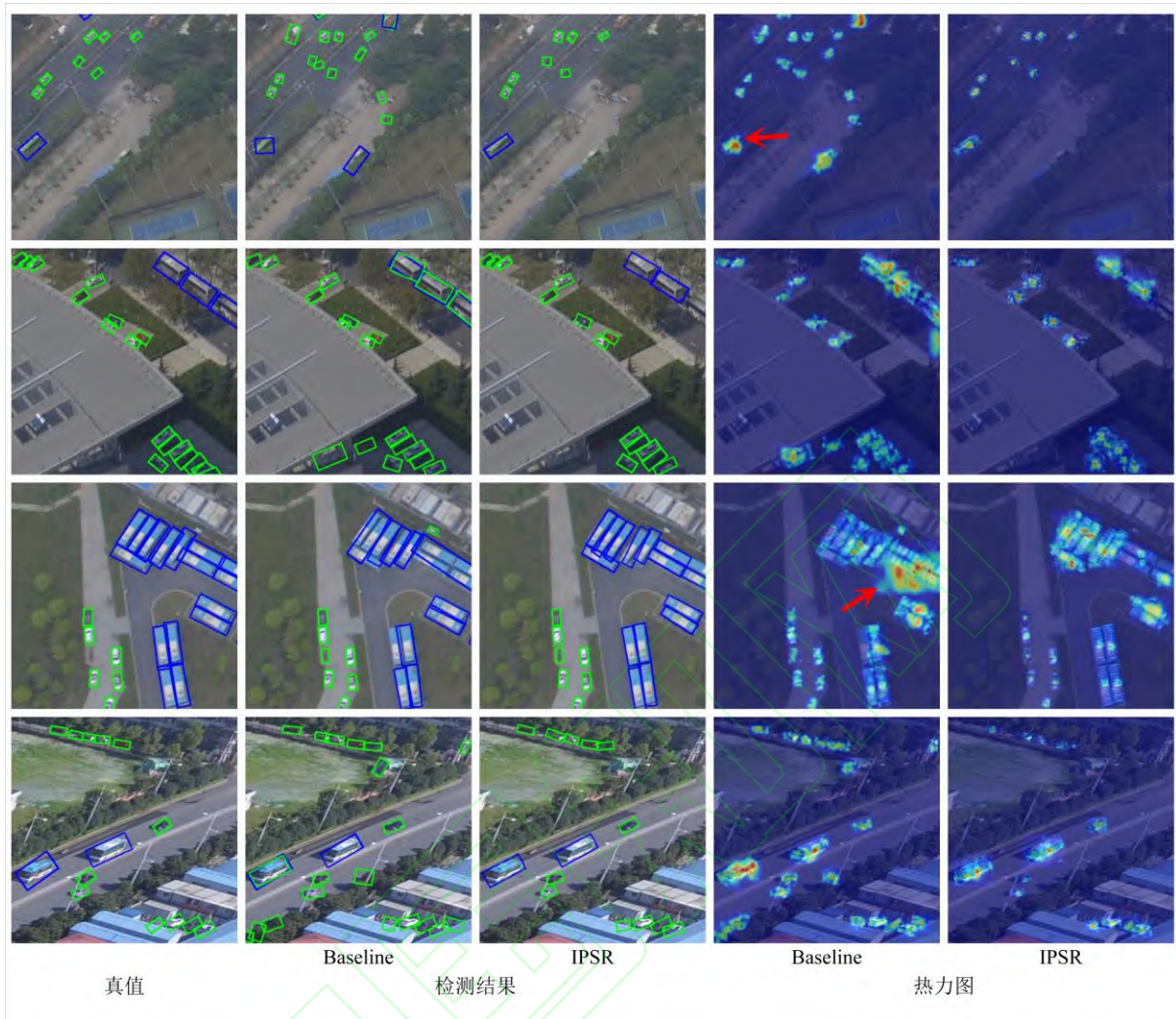


图 10 Baseline 和 IPSR 的检测结果和热力图的可视化比较

Fig.10. Visual comparison of the detection results and heatmaps between the Baseline and IPSR Vehicular Technology, 2024, 73(2): 1765-1776.

参考文献

- [1] WU X, LI W, HONG D F, et al. Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(1): 91-124.
- [2] ZHU P F, WEN L Y, DU D W, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7380-7399.
- [3] OSCO L P, DOS SANTOS DE ARRUDA M, GONÇALVES D N, et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 174: 1-17.
- [4] XU L J, YANG Q H, QIN M, et al. Collaborative human recognition with lightweight models in drone-based search and rescue operations[J]. Ieee Transactions on
- [5] LIU W, ZHANG T, HUANG S J, et al. A hybrid optimization framework for UAV reconnaissance mission planning[J]. Computers & Industrial Engineering, 2022, 173: 108653.
- [6] MA B D, LIU Z B, JIANG F H, et al. Reinforcement learning based UAV formation control in GPS-denied environment[J]. Chinese Journal of Aeronautics, 2023, 36(11): 281-296.
- [7] ZHOU W H, LI J, LIU Z H, et al. Improving multi-target cooperative tracking guidance for UAV swarms using multi-agent reinforcement learning[J]. Chinese Journal of Aeronautics, 2022, 35(7): 100-112.
- [8] LIU L, OUYANG W, WANG X G, et al. Deep learning for generic object detection: a survey [J]. International Journal of Computer Vision, 2020, 128(2): 261 - 318.
- [9] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [10] 蓝鑫, 吴淞, 伏博毅, 等. 深度学习的遥感图像旋转

- 目标检测综述[J]. 计算机科学与探索, 2024, 18(4): 861-877.
- LAN X, WU S, FU B Y, et al. Survey on deep learning in oriented object detection in remote sensing images[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(4): 861-877. (in Chinese)
- [11] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [14] YANG X, YAN J C, FENG Z M, et al. R3Det: refined single-stage detector with feature refinement for rotating object [C]// Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence. California: AAAI Press, 2021: 3163 - 3171.
- [15] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.
- [16] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection [C]// Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 3500 - 3509.
- [17] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images [C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 2844 - 2853.
- [18] WANG J H, TENG X C, LI Z, et al. VSAI: a multi-view dataset for vehicle detection in complex scenarios using aerial images[J]. Drones, 2022, 6(7): 161.
- [19] HAN W, CHEN J, WANG L Z, et al. Methods for small, weak object detection in optical high-resolution remote sensing images: a survey of advances and challenges [J]. IEEE Geoscience and Remote Sensing Magazine, 2021, 9(4): 8 - 34.
- [20] CHENG G, YUAN X, YAO X W, et al. Towards large-scale small object detection: survey and benchmarks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.
- [21] LI Y H, CHEN Y T, WANG N Y, et al. Scale-Aware Trident Networks for Object Detection [C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 6053 - 6062.
- [22] WANG K, WANG Z, LI Z, et al. Multi-scale cross distillation for object detection in aerial images [C]// Computer Vision-ECCV 2024. Cham: Springer, 2024: 452-471.
- [23] 吴洛冰, 谷玉海, 吴文昊, 等. 基于多尺度特征提取的遥感旋转目标检测[J]. 激光与光电子学进展, 2023, 60(12): 1228010.
- WU L B, GU Y H, WU W H, et al. Remote sensing rotating object detection based on multi-scale feature extraction[J]. Laser & Optoelectronics Progress, 2023, 60(12): 1228010. (in Chinese)
- [24] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936 - 944.
- [25] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759 - 8768.
- [26] 任洋, 陈绪君, 王磊. 基于跨空间多尺度的弱监督有向目标检测算法研究[J]. 激光杂志, 2024, 45(7): 63-70.
- REN Y, CHEN X J, WANG L. Research on weakly supervised directed target detection algorithm based on cross-space multi-scale[J]. Laser Journal, 2024, 45(7): 63-70. (in Chinese)
- [27] 李云红, 魏小双, 苏雪平, 等. 多尺度特征增强与交互融合的遥感小目标检测[J]. 西北大学学报(自然科学版), 2025, 55(2): 277-285.
- LI Y H, WEI X S, SU X P, et al. Remote sensing small target detection based on feature multi-scale enhancement and interaction fusion[J]. Journal of Northwest University (Natural Science Edition), 2025, 55(2): 277-285. (in Chinese)
- [28] 刘敏豪, 王堃, 金睿蛟, 等. 基于改进 RoI Transformer 的遥感图像多尺度旋转目标检测[J]. 应用光学, 2023, 44(5): 1010-1021.
- LIU M H, WANG K, JIN R J, et al. Multi-scale oriented object detection based on improved RoI Transformer in remote sensing images[J]. Journal of Applied Optics, 2023, 44(5): 1010-1021. (in Chinese)
- [29] 金睿蛟, 王堃, 刘敏豪, 等. 基于 DETR 和改进去噪训练的光学遥感图像多尺度旋转目标检测(特邀)[J]. 激光与光电子学进展, 2024, 61(2): 0211023.
- JIN R J, WANG K, LIU M H, et al. DETR with improved DeNoising training for multi-scale oriented object detection in optical remote sensing images(invited)[J]. Laser & Optoelectronics Progress,

- 2024, 61(2): 0211023. (in Chinese)
- [30] SINGH B, DAVIS L S. An Analysis of Scale Invariance in Object Detection – SNIP [C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3578 - 3587.
- [31] BAI Y C, ZHANG Y Q, DING M L, et al. SOD-MTGAN: small object detection via multi-task generative adversarial network[C]//Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 210-226.
- [32] NOH J, BAE W, LEE W, et al. Better to follow, follow to be better: towards precise supervision of feature super-resolution for small object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9724-9733.
- [33] ZHANG J Q, LEI J, XIE W Y, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J]. Ieee Transactions on Geoscience and Remote Sensing, .
- [34] ZHOU Y, YANG X, ZHANG G F, et al. MMRotate: A Rotated Object Detection Benchmark using PyTorch [EB/OL]// (2022-07-19)[2025-05-21]. <http://arxiv.org/pdf/2204.13317>.
- [35] XIA G-S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.
- [36] 邢星宇, 黄安, 姜为, 等. 降雨条件下车载激光雷达感知局限性[J]. 同济大学学报(自然科学版), 2023, 51(5): 785-793.
XING X Y, HUANG A, JIANG W, et al. Perception limitation of LiDAR sensor in autonomous vehicles under rainfall condition[J]. Journal of Tongji University (Natural Science), 2023, 51(5): 785-793. (in Chinese)
- [37] DONG Q L, ZHOU Z M, QIU X L, et al. A survey on self-supervised monocular depth estimation based on deep neural networks[J/OL]. IEEE Transactions on Neural Networks and Learning Systems[2025-05-21]. <https://ieeexplore.ieee.org/document/10947242>.
- [38] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE Press, 2021: 1833-1844.
- [39] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: a small target detection benchmark[J]. Journal of Visual Communication and Image Representation, 2016, 34: 187-203.
- [40] DU D, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: Object detection and tracking [C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2018: 375 - 391.
- [41] SUN Y M, CAO B, ZHU P F, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. Ieee Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6700-6713.
- [42] FENG C F, CHEN Z Y, LI X, et al. Hazydet: open-source benchmark for drone-view object detection with depth-cues in hazy scenes [EB/OL]. (2025-05-26)[2025-05-28]. <http://arxiv.org/abs/2409.19833>.
- [43] ZHANG K, LIANG J Y, VAN GOOL L, et al. Designing a practical degradation model for deep blind image super-resolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 4771-4780.
- [44] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9992-10002.
- [45] LUO W J, LI Y J, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: Curran Associates, Inc., 2016: 4905 - 4913.
- [46] ZHAO C Y, HSIAO J H, CHAN A B. Gradient-based instance-specific visual explanations for object specification and object discrimination[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(9): 5967-5985.