

DOI: 10.20189/j.cnki.CN/61-1527/E.202601007

多模态巡飞器时敏目标检测数据集构建

申 通¹, 王思宇², 杨小冈¹, 卢瑞涛¹

(1. 火箭军工程大学, 西安 710025;

2. 徐州工程学院 电气与控制工程学院, 徐州 221018)

摘 要: 针对现有多模态目标检测数据集目标种类单一、场景覆盖有限、图像视角与巡飞器不适配等问题, 构建了一套面向巡飞器应用的多模态时敏目标检测数据集, 旨在为巡飞器目标检测算法的训练评估提供数据支撑。首先, 综合利用实景拍摄和遥感影像平台生成 2 种方式, 采集种类丰富、场景多样的时敏目标图像; 其次, 基于 SuperPoint 和 SuperGlue 算法实现多模态图像的空间对齐, 确保目标在跨模态数据中的位置一致性; 进而采用图像合成和风格迁移方法进行数据增广, 以平衡模态分布并扩充数据规模; 最后, 通过语义分割算法实现时敏目标的自动标注。实验结果表明: 所构建的数据集相较于其他多模态目标数据监测数据集, 数据量更加充足, 目标种类更加多样, 场景覆盖更加广泛。

关键词: 多模态目标检测; 数据集构建; 图像配准; 数据增广; 图像标注

中图分类号: TP391

文献标志码: A

开放科学标识码(OSID):

文章编号: 2097-2741(2026)01-0068-11



听语音
聊科研
与作者互动

在现代军事战争中, 巡飞器对时敏目标的有效探测依赖于先进的传感设备和精确的目标检测算法^[1]。随着传感技术的不断发展, 具备光学、红外、雷达等成像技术的多模态传感器正逐渐应用于巡飞器平台。联合不同物理传感器的信号, 可以解决传统单一模态传感器在实际应用中存在的场景适应性差、探测能力不足等问题, 为巡飞器赋予了更为全面、准确的目标探测感知能力^[2]。因此, 与多模态传感器相适应的多模态目标检测算法成为当前目标检测领域的研究热点。相较于单模态检测算法, 多模态检测可以充分利用不同模态数据的互补特性, 进一步提升目标检测的鲁棒性与准确性, 从而实现对时敏目标的快速锁定与精准识别, 极大地提升巡飞器在现代战争中的作战效能^[3]。

在目标检测领域, 大量精准、有效的图像数据对于推动检测算法性能迭代突破极为重要。目前, 基于视觉的多模态时敏目标检测仍处于起步

阶段, 尚未形成一套被广泛认可的基准数据集^[4]。常用的公开数据集包括 VEDAI^[5]、DroneVehicle^[6]、LLVIP^[7]、KAIST^[8]等, 主要包含光学和红外 2 种模态, 各数据集示例如图 1 所示。VEDAI 数据集主要包含船、车、飞机等 9 个目标类别, 共 1 246 对光学-红外图像, 每张图像的分辨率为 1 024×1 024 像素, 且均在白天时段拍摄, 由于该数据集中物体在图像中的像素占比很小, 所以特别适用于弱小目标检测算法的训练与验证。DroneVehicle 数据集主要用于空地图像车辆目标检测, 包含多种拍摄条件和场景, 涵盖不同的视角、高度和光照条件, 共 28 439 对图像, 其目标尺度变化较大, 目标间重叠遮挡且存在大量低光照场景, 对于模态失衡条件下多模态目标检测问题具有重要价值。LLVIP 数据集主要用于低光照条件下多模态行人检测, 共包含超过 15 488 对图像, 大多使用道路监控在夜间、光线昏暗的环境和有雾的条件下拍摄。不同于 LLVIP 的监控视

收稿日期: 2025-08-29

修回日期: 2025-09-26

录用日期: 2025-10-09

基金项目: 国家自然科学基金项目 (62276274)

第一作者: 申通 (1995—), 男, 讲师, 博士, 主要从事图像识别、导航制导控制研究。E-mail: shentong521@live.com

引用格式: 申通, 王思宇, 杨小冈, 等. 多模态巡飞器时敏目标检测数据集构建[J]. 火箭军工程大学学报, 2026, 40 (1): 68-78.
SHEN Tong, WANG Siyu, YANG Xiaogang, et al. Construction of an aircraft multimodal time-sensitive object-detection dataset[J].
Journal of Rocket Force University of Engineering, 2026, 40 (1): 68-78.

角, KAIST 数据集为车辆前视视角下的多模态行人检测数据集, 在全时段采集, 主要包括校园、街道以及乡村的各种常规交通场景, 共 95 328 对图像。

上述多模态目标检测数据集虽然能够提供数量充足的图像, 但大部分是在非巡飞器视角下拍摄, 且覆盖的目标种类及场景有限, 无法有效支撑巡飞器多模态目标检测算法的训练和验证。例如, KAIST 数据集中均为驾驶场景下的行人或车辆目标, 目标尺度和视角变化较小, 外观稳定且无明显畸变, 而巡飞器视角下的多模态检测数据通常从高空或中低空的动态视角采集, 具有俯视特性。VEDAI 数据集为卫星航拍图像影像, 虽然场景较为丰富, 但均在光照充足条件下拍摄, 缺少雾天、夜晚等复杂环境下的图像数据, 且目标尺度变化较小。DroneVehicle 数据集虽为空地拍摄数据, 且具有一定的尺度变换, 但其目标种类单一, 仅包含车辆目标数据, 且图像缺少电子噪声、烟雾等干扰场景。针对上述数据集存在的局限性, 本文将探索面向巡飞器应用的多模态图像数据集构建方法, 并制备一套光学和红外融合的多模态巡飞器时敏目标检测数据集, 从而为巡飞

器搭载目标检测算法的训练与评估提供数据支撑。

1 多模态巡飞器时敏目标检测数据集构建流程

多模态巡飞器时敏目标检测数据集构建流程如图 2 所示, 包括多模态时敏目标图像采集、多模态图像配准、多模态数据增广以及数据标注等主要环节。首先, 多模态时敏目标图像采集主要通过搭载多模态传感器的巡飞器实景拍摄进行。由于军用高价值目标数量稀缺, 实景拍摄难度大, 因此, 此类目标的光学图像主要借助遥感影像平台进行采集。其次, 由于不同模态的传感器通常安装在设备的不同位置, 其视角和成像范围不完全一致, 因此, 对于实景拍摄的数据需要通过多模态图像配准使其在空间上对齐, 从而确保目标在不同模态数据中的位置一致。对于遥感影像平台采集到的光学图像, 可使用图像合成方法进一步扩充数据量, 再通过风格迁移方法生成与之空间对齐的红外图像。最后, 对已配对的光学-红外图像进行数据标注, 得到多模态巡飞器时敏目标

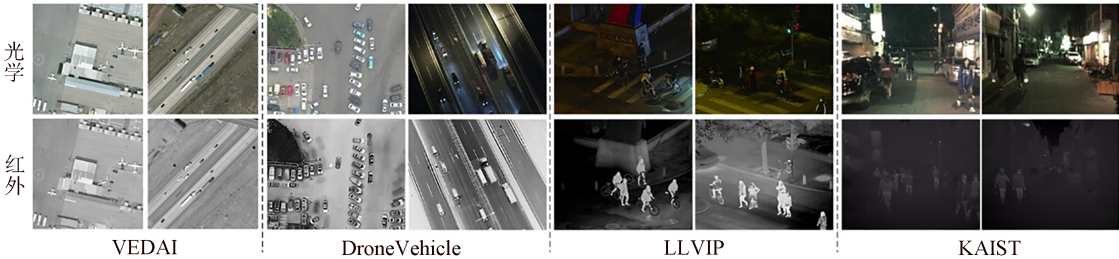


图 1 多模态目标检测常用数据集

Fig. 1 Commonly used datasets for multimodal object detection

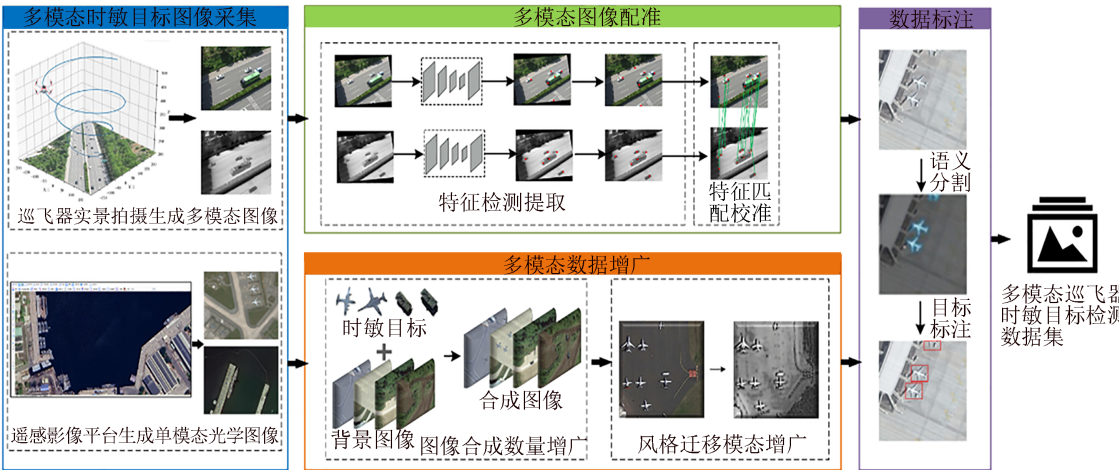


图 2 多模态巡飞器时敏目标检测数据集构建流程

Fig. 2 Construction process of an aircraft multimodal time-sensitive object-detection dataset

检测数据集。

2 多模态巡飞器时敏目标检测数据集构建方法

2.1 多模态时敏目标图像采集

时敏目标即时间敏感目标，指必须在有限的“攻击窗口”或“交战机会”内发现、定位、识别、瞄准和攻击的目标^[9]。这类目标通常具备高度的机动性和动态性，其可观测性和可操作性在时间维度上具有显著的敏感性。时敏目标广泛存在于军事和战术场景中，比如待机飞行器、近岸驻舰船、高速移动特种车辆等。考虑目标的典型性和实用性，选取战斗机、军舰、坦克、装甲车、军用卡车、轿车、货车、飞机和船等9类时敏目标作为采集对象，覆盖军用及民用领域。

对时敏目标图像的采集主要通过巡飞器实景拍摄和遥感影像平台生成2种方式进行。实景拍摄通常使用搭载集成光学和红外传感器的巡飞器进行。拍摄过程中，为了获得时敏目标多尺度、多视角图像，在确定某一采集区域后，设计巡飞器飞行路线为螺旋上升曲线，飞行高度从200 m升

高至500 m，围绕该区域飞行3周，并以每秒30帧的速度记录飞行视频，从而获得大量的目标及其周边环境的光学和红外图像，实景拍摄示意如图3所示。对于军用高价值稀缺目标，主要通过遥感影像平台生成，借助Google Earth^[10]等平台，在选定目标类型和区域后，通过调整视角高度及可视范围，模拟巡飞器视角生成大量光学图像，后续再通过数据增广环节中风格迁移产生对应的红外图像。此外，少部分稀缺目标通过自行搭建的实景仿真平台采集，该平台构建了水域、场坪、草地等多种真实场景，包含军机、军舰、坦克等仿真模型，如图4所示。

2.2 基于自监督学习的图像配准方法

现有的图像采集设备在拍摄时，由于不同模态的传感器如光学摄像头、红外摄像头、激光雷达等通常安装在设备的不同位置，因此，采集得到的光学-红外图像视角和成像范围不完全一致。然而，多模态目标检测算法依赖于不同模态的特征融合，未对齐的图像会导致特征在空间上不匹配，降低融合效果^[11]。因此，图像配准是多模态目标检测数据集制备的重要环节，其主要作用是把这些多模态图像在空间上对齐，从而确保目标

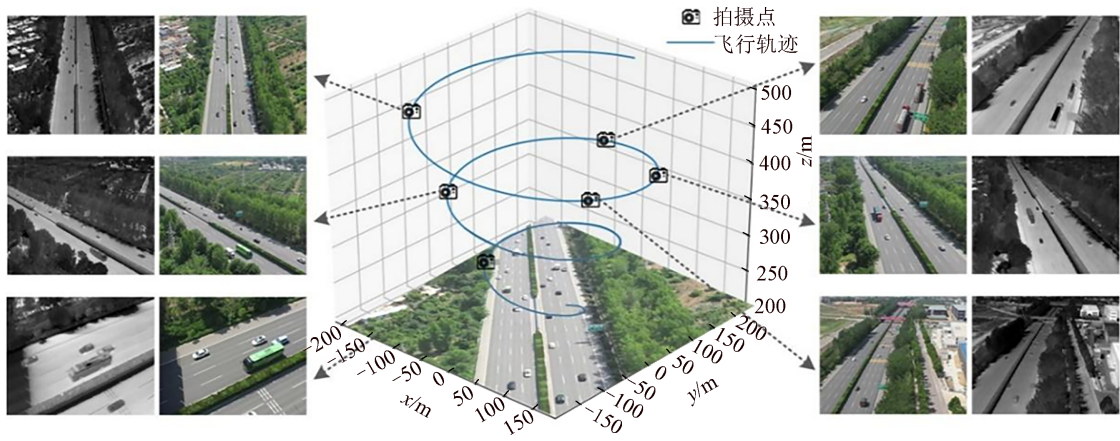


图 3 巡飞器实景拍摄示意
Fig. 3 Schematic of real-scene shot by aircrafts

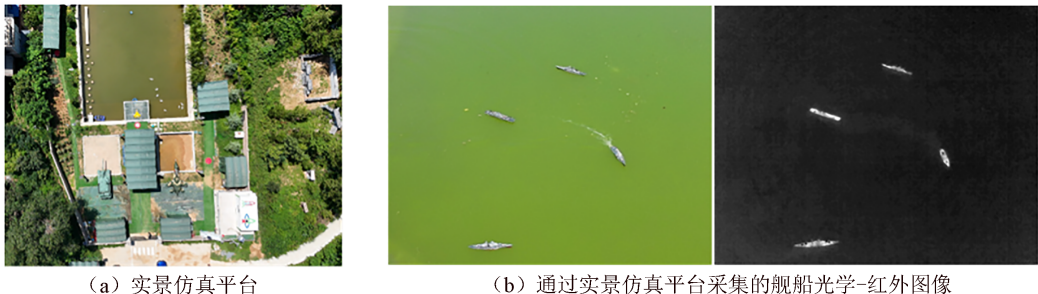


图 4 实景仿真平台及其采集的光学-红外图像

Fig. 4 Schematic of a real-scene simulation platform and optical-infrared images obtained by it

在不同模态中的位置一致。

针对多模态图像在特征分布、对比度和纹理上的差异, 采用一种基于自监督学习的图像配准方法, 实现光学和红外图像的精确对齐, 具体原理如图5所示。模型主要由特征检测提取模块、特征匹配校准模块组成。特征检测提取模块使用SuperPoint^[12]特征提取器, 其核心优势在于能够同时提取图像中的关键点及其描述子, 尤其适用于多模态场景下的跨域特征表达。特征匹配校准模块采用SuperGlue^[13]匹配算法, 对输入图片的多个描述子进行匹配。经过SuperPoint网络进行特征点检测和描述子提取后可能会产生错误的检测和提取, 而SuperGlue算法可以过滤掉错误点, 实现图像间的精确对齐。

在图像配准前, 首先采用包含丰富角点的人工合成数据集对SuperPoint模型进行预训练, 对待配准的图像生成伪真值的特征点标签。配准过程中, 将图像及其伪真值特征点标签进行随机单应性变换, 并输入至预训练的SuperPoint网络。网络分别输出提取到的特征点, 通过计算与伪标签的损失值, 监督模型参数进行更新和调整。然后, 对所有变形图像的特征点进行融合, 得到待配准图像的特征点。最后, 通过SuperGlue算法得到待配准图像对之间特征点的匹配关系, 同时剔除非共有的特征点, 完成特征点对的自动配准。

图6展示了部分不同光照条件、不同环境特征的图像配准结果, 包括源图像(光学)、目标图像

(红外)、配准结果图、棋盘格图以及重叠图。3组数据分别展示了战斗机跑道场景、坦克草地场景、车辆高速公路场景的配准结果。可以看出, 不同环境特征的复杂图像对都实现了较为精确的空间对齐。在草地、植物等边缘不明显的区域, 也达到了很好的对齐效果, 进一步验证了该方法在多模态配准中的有效性。

2.3 多模态数据增广

2.3.1 基于泊松梯度一致性合成的数量增广算法

在目标检测领域中, 数据集往往存在规模较小、分布不均匀等问题^[14]。特别是在具有军事背景的高价值时敏目标检测任务中, 由于数据采集成本高昂、红外模态数据获取难度大, 导致样本数量不足、模态分布不均衡等问题尤为突出。数据增广作为一种有效扩充训练数据的手段, 能够缓解模型训练中的过拟合问题, 并在一定程度上提升模型的检测精度与泛化能力^[15]。传统的增广方法如旋转、缩放、裁剪等难以生成足够多样化的样本, 限制了模型性能的进一步提升^[16]。

对此, 本文采用一种基于泊松梯度一致性合成^[17]的算法, 通过泊松梯度求解, 使合成区域的梯度不仅与源图像内部结构一致, 还与目标图像的背景纹理和结构相融合。由此, 可以在保持图像细节和自然性的同时, 生成高质量且多样化的合成样本, 实现图像数量的大幅扩充, 有效解决样本数量不足的问题。数量增广算法原理如图7所示, 主要包括提取目标和合成图像2个阶段。

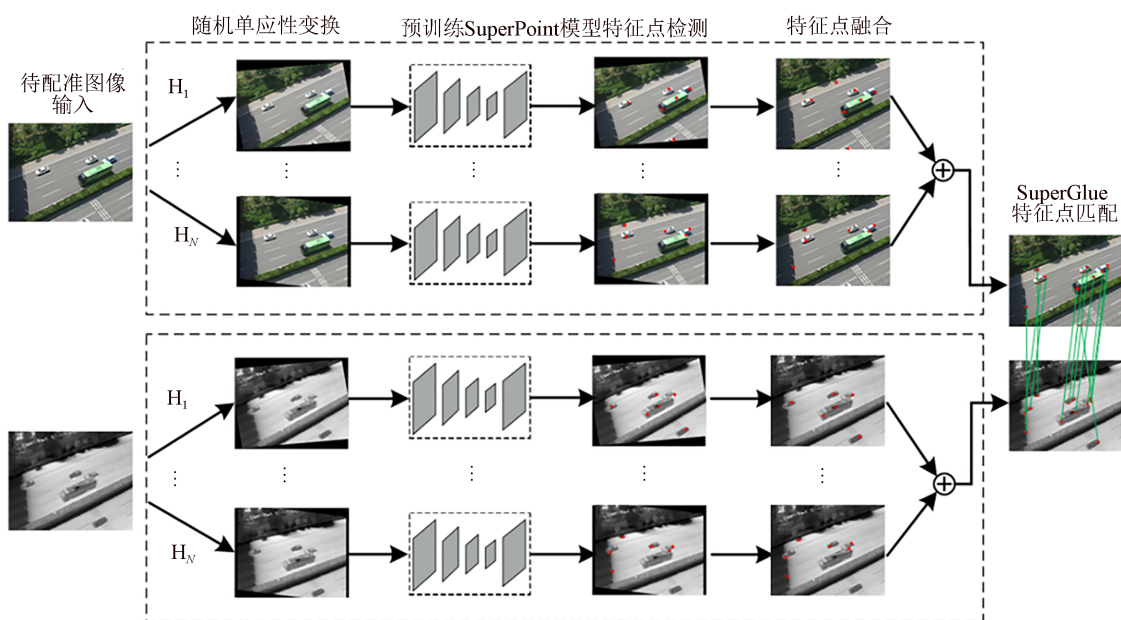


图5 基于自监督学习的图像配准示意

Fig. 5 Schematic of image registration based on self-supervised learning

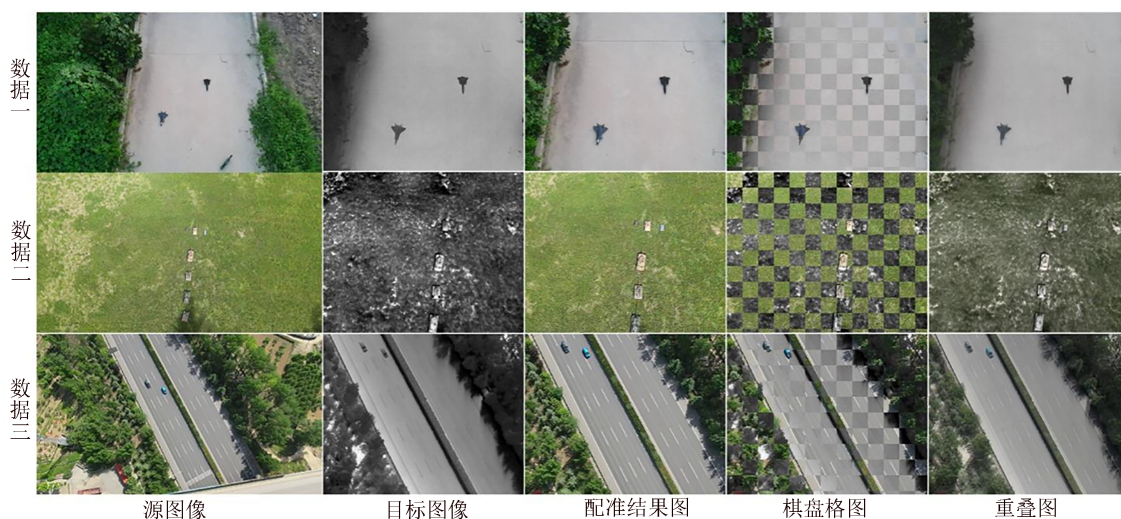


图 6 多模态图像配准可视化结果

Fig. 6 Visualization results of multimodal image registration

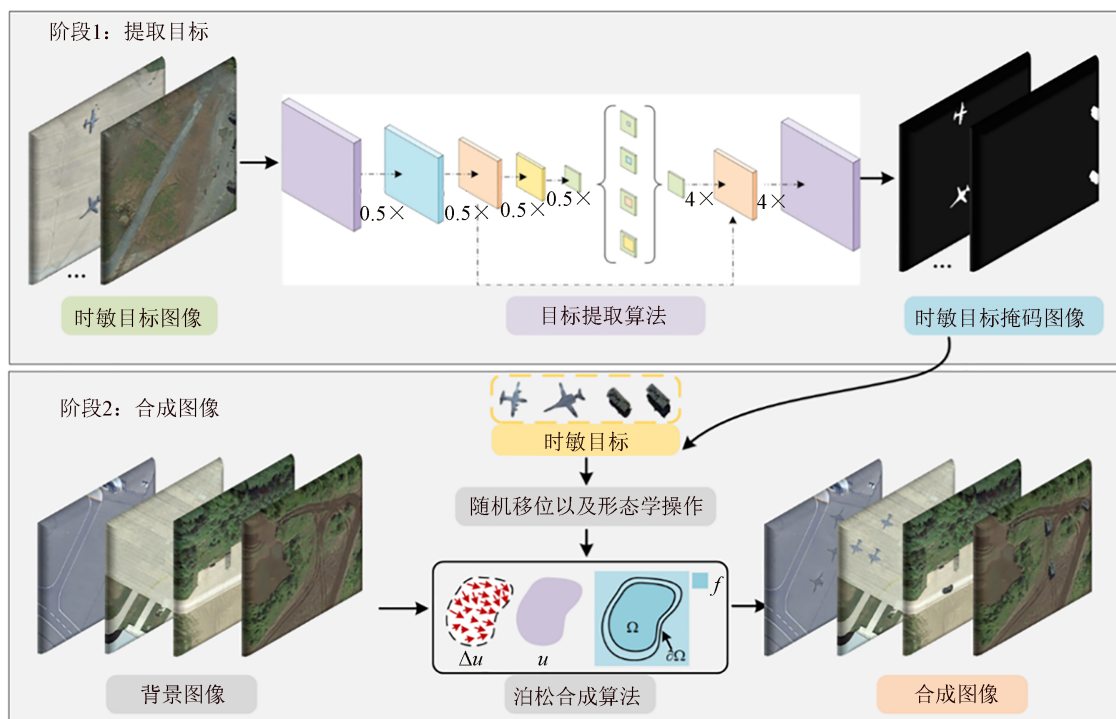


图 7 基于泊松梯度一致性合成的数量增广算法原理

Fig. 7 Schematic of principle of quantity augmentation algorithm based on Poisson gradient consistency synthesis

首先, 利用基于空洞空间金字塔池化的时敏目标提取模块^[18], 在原始图像中提取目标区域及其对应的分割掩码作为前景输入。然后, 在数据集中选取多样化的背景图像, 将不同规模的时敏目标嵌入其中, 嵌入过程中时敏目标会随机进行移位和形态学操作。最后, 利用泊松方程的图像融合技术, 在保持源图像的梯度场并匹配目标图像边界的条件下, 对目标的光照、颜色以及边缘进行自适应调整, 使其与背景的光学特性一致, 消除拼接痕迹, 实现图像间的无缝融合, 提升数

据的真实性与一致性。图8展示了使用所提方法合成的军事高价值时敏目标光学图像。可以看出, 合成目标在纹理、光照及几何特征上与真实目标高度一致, 目标边界通过自然过渡实现平滑融合。

2.3.2 基于风格迁移的模态增广方法

通过遥感影像平台生成和基于泊松梯度一致性合成所产生的图像均为光学图像, 对应的红外图像获取难度大, 因此, 数据集模态呈现严重的非均衡性^[19]。针对此问题, 采用一种基于Cycle-GAN-Turbo的光学-红外风格迁移方法, 利用实

景拍摄的光学-红外配对数据训练网络, 实现从光学图像到红外图像的跨模态映射。CycleGAN-Turbo 模态增广模型网络结构如图 9 所示, 该模型在原始扩散模型的基础上进行了深度优化, 通过整合将编码器、U-Net 以及解码器 3 个独立训练的子模块统一至参数规模优化网络中, 有效减少了冗余计算, 并显著提升了整体推理效率^[20]。

训练过程中, 以实拍光学图像 X 作为输入, 以真实的红外图像 Y 作为标签, 网络可在无配对

监督条件下通过对目标外观特征、纹理分布及光照特性的学习进行跨模态重构, 学习 2 个不同方向的映射函数 $G(x): X \rightarrow Y$ 和 $G(y): Y \rightarrow X$, 从而实现源域和目标域之间的相互转换, 2 种转换均采用相同的生成网络 G 。迁移过程中, 模型在保持输入光学场景结构的前提下, 按照训练的映射函数生成红外模态并进行细节强化, 最终生成样本与光学图像具有空间一致性。模态增广的可视化结果如图 10 所示。

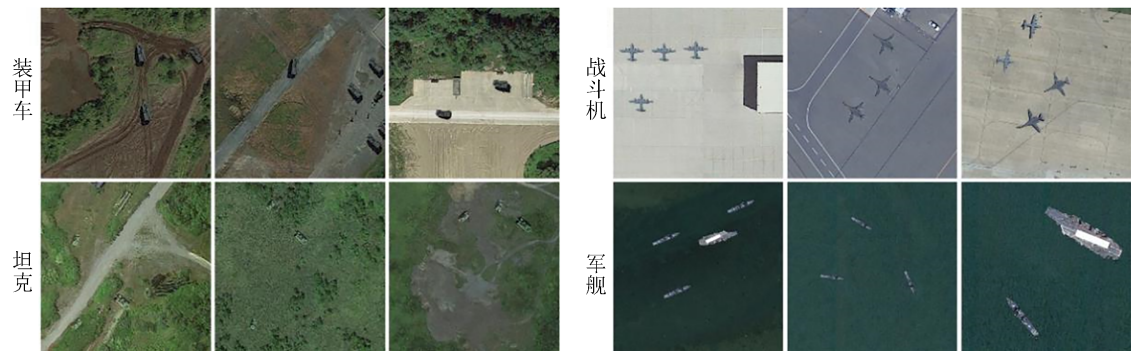


图 8 高价值时敏目标图像增广结果

Fig. 8 Augmentation results of high-value time-sensitive object images

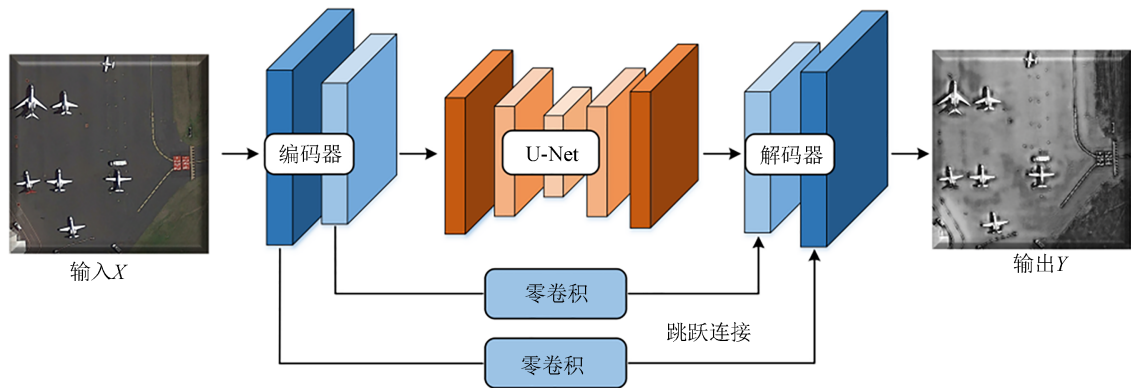


图 9 CycleGAN-Turbo 模态增广模型网络结构

Fig. 9 Structural diagram of modal augmentation model based on CycleGAN-Turbo

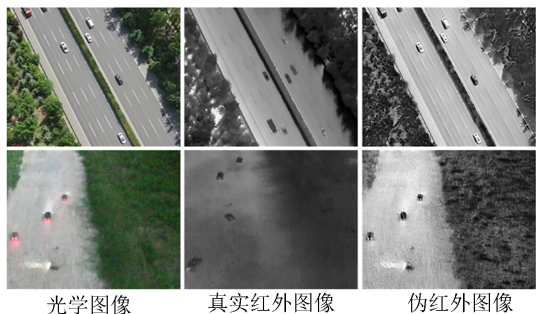


图 10 模态增广结果

Fig. 10 Modal augmentation results

由图 10 增广结果表现特性可以看出, 光学图像通常具有较高的空间分辨率, 能够精细捕捉物

体的细节, 如目标的形状、颜色和边缘纹理信息等, 从而提供丰富的视觉特征。真实的红外成像不受光照条件限制, 可在夜间、烟雾、雾霾等恶劣环境下实现全天候探测, 但其分辨率通常低于光学图像, 细节表现能力较弱, 难以精确识别目标的精细形状和纹理。由于伪红外图像是在光学图像的基础上通过风格迁移反演生成, 因此其边缘纹理会继承光学图像的密度优势, 在边缘细节及纹理信息方面与真实红外图像有所差异。为了验证伪红外图像的有效性, 本文采用计算训练集中真实红外图像与伪红外图像的均值、标准差、方差、信息熵、对比度和平均梯度并进行对比。

计算方法分别为

$$\mu = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H I(i, j) \quad (1)$$

$$\sigma = \sqrt{\frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H (I(i, j) - \mu)^2} \quad (2)$$

$$\sigma^2 = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H (I(i, j) - \mu)^2 \quad (3)$$

$$E = - \sum_{k=0}^{L-1} p(k) \log_2(p(k)) \quad (4)$$

$$C = \max(I) - \min(I) \quad (5)$$

$$\bar{G} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H \sqrt{G_x^2(i, j) + G_y^2(i, j)} \quad (6)$$

式中： μ 、 σ 、 σ^2 、 E 、 C 、 $\bar{G}$ 分别为图像的均值、标准差、方差、信息熵、对比度和平均梯度； W 、 H 分别为图像的宽度和高度； $I(i, j)$ 为位置 (i, j) 处的像素值； $p(k)$ 为灰度值 k 在图像中出现的频率； L 为图像灰度级总数； $G_x(i, j)$ 、 $G_y(i, j)$ 分别为位置 (i, j) 处的水平方向梯度和垂直方向梯度，通常使用Sobel算子卷积^[21]计算。

2种图像的模态增广效果对比如表1所示。可以看出，伪红外图像的均值和标准差均略高于真实红外图像，表明其在亮度、边缘清晰度等方面高于真实图像；信息熵和平均梯度反映了图像中的微小细节反差和纹理变化特征，二者较为接近。尽管伪红外图像在对比度方面相较于真实红外图像还有待提升，但整体误差水平处于可控范围内。

2.4 基于语义分割的数据集自动标注方法

图像标注作为数据集制备的最终环节，其主

要目的是对图像进行详细地描述与解释，并生成适用于目标检测算法训练的标签文件，该文件中通常要求给出图像中所包含的目标数量、大小、种类、位置等信息^[22]。传统的人工标注方法，由于任务量巨大，通常需要耗费大量的人力与时间^[23]。为了在保证数据标注准确的同时，减少人力时间成本，采用一种基于SAM的图像语义分割算法实现时敏目标精细化提取和自动标注^[24]。该模型可以在给定任何分割提示的情况下返回有效分割掩码，提示符可以包括识别对象的空间或文本信息，无需针对特定任务进行模型微调。基于SAM的语义分割模型结构如图11所示。可以看出，模型主要包括图像编码器、提示编码器、掩码解码器3个部分。考虑可扩展性和巨大的预训练需求，SAM使用预训练视觉变换器（vision transformer, ViT）作为图像编码器。输入图像经过图像编码器后获得图像的编码信息，生成图像的特征嵌入表示。提示编码器为双通道体系，包括稀疏的点、框、文本类型和密集的掩码类型^[25]，稀疏提示通道将点、框等几何要素与文本语义编码作为特征向量，而密集提示通道通过卷积层对输入掩码进行空间压缩，实现多模态提示信息与图像特征的深度融合。最后，通过轻量化Transformer掩码解码器，利用提示自注意力、交叉注意力迭代优化图像和提示的嵌入表示，输出像素级精度的语义分割结果。在分割结果基础上，筛选出符合要求的时敏目标区域，实现多模态目标检测数据的自动标注。

表 1 模态增广效果对比

Table 1 Comparison of modal augmentation effects

图像类型	均值	标准差	信息熵	对比度	平均梯度
真实红外图像	126.049 31	43.281 21	7.231 47	147.147 56	5.135 38
伪红外图像	137.884 15	45.599 76	7.188 54	43.930 82	3.158 52

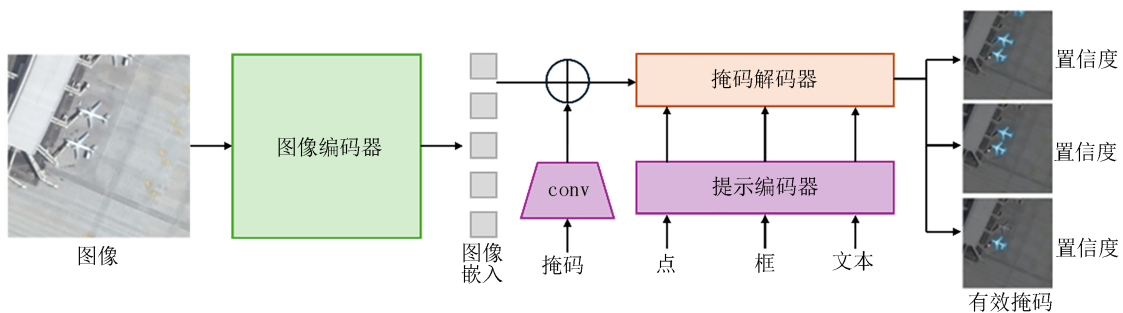


图 11 基于SAM的时敏目标语义分割模型结构

Fig. 11 Structural diagram of the SAM-based time-sensitive object semantic segmentation model

3 数据集评估分析

所构建的多模态巡飞器时敏目标检测数据集 (aircraft multimodal time-sensitive object-detection dataset, AMTOD) 包含战斗机、军舰、坦克、装甲车、军用卡车、轿车、卡车、飞机和船 9 类时敏目标。为了强化图像数据的结构化表征，并根据不同应用场景的任务需求针对性地检索数据，构建了四级递进式的目标检测数据架构^[26]，如图 12 所示。

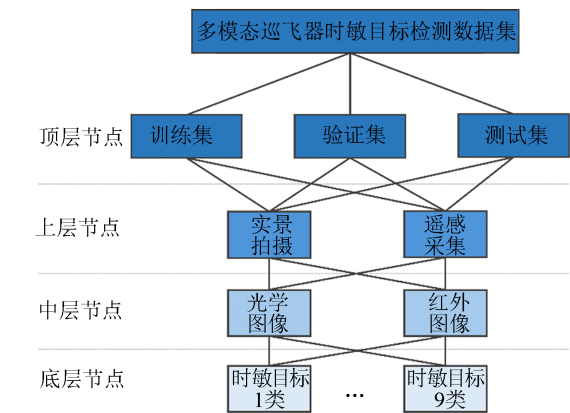


图 12 多模态巡飞器时敏目标检测数据集结构
Fig. 12 Structure of the multimodal time-sensitive object-detection dataset for aircrafts

其中，顶层节点根据数据应用目的划分为训练集、验证集和测试集，上层节点根据图像采集方式划分为实景拍摄和遥感采集，中层节点根据图像模态进行分类，底层节点按照时敏目标类型划分为 9 类，形成涵盖多模态、多视角、多类型的多维数据体系。每类数据都包含图像和标注 2 个类型文件，图像为包含时敏目标的 JPEG 格式文件，标

注则为 TXT 文本文件，清晰、准确地描述了对应图像所包含的目标信息，包括图像名称、图像尺寸、目标实例属性、目标框边界坐标等。

使用 COCO^[27] 数据集格式根据像素面积对总体数据集目标尺寸进行分类，面积小于 32×32 像素的目标被定义为小目标，大于 32×32 像素但小于 96×96 像素的目标为中目标，大于 96×96 像素的目标为大目标。其中，小目标占比 48.9%，中目标占比 36.8%，大目标占比 14.3%。分布特征表明，小目标在自建数据集中具有显著的数量优势，会对检测算法中小目标检测性能上的优化需求设计提出挑战。其次，目标类别数量分布方面，数据集呈现出类别不平衡特征。例如，标注框数量排名前三的目标类别分别为船、战斗机、轿车，数量分别为 58 236、23 118、20 909，明显高于其他类别。这种不平衡性为测试模型的鲁棒性提供了条件，能够有效评估模型在不同类别上的泛化性能。所构建的数据集与其他多模态目标检测数据集对比如表 2 所示。通过表 2 可以看出，所构建的数据集具有以下优势。①图像及目标数据量充足，包含 18 294 对光学-红外配对图像及 148 514 个目标标注框，可有效支撑巡飞器时敏目标检测算法的训练和验证；②目标种类丰富，包括战斗机、军舰、轿车等 9 大类时敏目标，除常见的民用时敏目标外，还包含大量的军事高价值时敏目标，拓宽了数据集的适用范围；③目标尺度、视角、场景分布多样，包含不同目标尺度、不同视角目标形态以及城市、沙漠、海洋等多种场景，数据采集覆盖全时段，可进一步增强目标检测算法在未知环境中的泛化能力。然而，与公开数据集相比，所构建的数据集还存在一些局限性。例如，

表 2 本文数据集与其他多模态时敏目标检测数据集对比						
Table 2 Comparison between the proposed dataset and other multimodal time-sensitive object-detection datasets						
数据集名称	主要用途	图像目标数量	图像分辨率	目标类别	模态	覆盖场景/时段
AMTOD	巡飞器时敏目标检测	18 294 对图像 148 514 标注框	640×512	9 大类	光学 红外	城市、乡村、港口、沙漠、海洋、田野、机场等/全时段
VEDAI	卫星航拍视角小目标车辆检测	1 246 对图像 3 640 标注框	1 024×1 024	1 大类	光学 红外	城市、港口、田野等/白天
DroneVehicle	无人机视角多尺度车辆检测	28 439 对图像 428 086 标注框	640×512	1 大类	光学 红外	城市道路/全时段
LLVIP	弱光环境安防监控视角行人检测	15 488 对图像 34 407 标注框	640×480	1 大类	光学 红外	城市人行道/夜晚
KAIST	车辆前视视角行人检测	95 328 对图像 103 128 标注框	640×480	1 大类	光学 红外	校园、街道和乡下等/全时段

部分红外图像基于风格迁移生成,与真实的红外传感器拍摄图像相比还有一定差距;标注环节还需要结合人工选择和修正,尚未实现全自动标注,后续仍需进一步完善。

4 结论

为满足巡飞器目标探测和态势感知能力的发展需求,支撑目标检测算法的迭代突破,设计了一种面向巡飞器的多模态时敏目标检测数据集构建方法,并通过多模态图像数据采集、图像配准、数据增广、目标标注等主要环节,制备了一个数量充足、格式规范、场景丰富的多模态时敏目标检测数据集。该数据集包含 9 类时敏目标,覆盖民用和军用领域,光学-红外配对图像达 18 294 对,涉及城市、沙漠、海洋等多种场景。与现有公开数据集相比,在数据量、目标类型、场景丰富度等方面更具优势,可为巡飞器目标探测、认知导航等技术发展提供数据支撑。后续,可考虑在已有数据基础上增加烟雾、微光和电子噪声等多种干扰因素,模拟更多实际应用中可能遇到的复杂情况,为多模态时敏目标检测算法研究提供更贴近真实场景的测试基准。

参考文献

- [1] 贾永楠,田似营,李擎. 无人机集群研究进展综述[J]. 航空学报, 2020, 41(S1): 4-14.
JIA Y N, TIAN S Y, LI Q. Development of unmanned aerial vehicle swarms[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(S1): 4-14.
- [2] 孙晓永,孙备,郭润泽,等. 面向多无人机协同的多模态目标检测方法[J]. 仪器仪表学报, 2025, 46(2): 209-220.
SUN X Y, SUN B, GUO R Z, et al. Multi-modal target detection method for multi-UAV coordination[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 209-220.
- [3] 郭润泽,孙备,孙晓永,等. 无人机弱光条件下多模态融合目标检测方法[J]. 仪器仪表学报, 2025, 46(1): 338-350.
GUO R Z, SUN B, SUN X Y, et al. Multi-modal fusion object detection method for UAVs under low light conditions[J]. Chinese Journal of Scientific Instrument, 2025, 46(1): 338-350.
- [4] LI J J, JI W, BI Q, et al. Joint semantic mining for weakly supervised RGB-D salient object detection[C]//Neural Information Processing Systems. La Jolla, CA, USA: NIPS, 2021, 11945-11959.
- [5] ZOU Z X, SHI Z W. Random access memories: a new paradigm for target detection in high resolution aerial remote sensing images[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1100-1111.
- [6] 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.
- [7] JIA X Y, ZHU C, LI M Z, et al. LLVIP: a visible-infrared paired dataset for low-light vision[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ, USA: IEEE, 2021: 3489-3497.
- [8] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: benchmark dataset and baseline[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 1037-1045.
- [9] 张开德,赵书斌. 时敏打击指挥控制技术初探[J]. 指挥控制与仿真, 2006, 28(2): 1-5.
ZHANG K D, ZHAO S B. The command and control technology of time critical strike[J]. Command Control & Simulation, 2006, 28(2): 1-5.
- [10] 黄仲良,何敬,刘刚,等. 面向GEE平台的遥感影像分析与应用研究进展[J]. 遥感技术与应用, 2023, 38(3): 527-534.
HUANG Z L, HE J, LIU G, et al. Research progress of remote sensing image analysis and application oriented to GEE platform[J]. Remote Sensing Technology and Application, 2023, 38(3): 527-534.
- [11] 占祥慧. 复杂场景下红外与可见光图像配准算法研究[D]. 北京: 中国科学院大学, 2022: 13-20.
ZHAN X H. Research on registration algorithm of infrared and visible images in complex scenes[D]. Beijing: University of Chinese Academy of Sciences, 2022: 13-20.
- [12] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2018: 224-236.
- [13] DENG Y X, MA J Y. Resmatch: residual attention learning for local feature matching[EB/OL]. (2023-07-11)[2025-05-26]. <https://arxiv.org/abs/2307.05180>.

- [14] MADHAN E S, KANNAN K S, RANI P S. A distributed submerged object detection and classification enhancement with deep learning[J]. Distributed and Parallel Databases, 2021, 23(3): 1-17.
- [15] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6: 60.
- [16] CUBUK E D, ZOPH B, MANÉ D, et al. Auto-Augment: learning augmentation strategies from data[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 113-123.
- [17] SUN J, ZHU H Y, XU Z B, et al. Poisson image fusion based on Markov random field fusion model[J]. Information Fusion, 2013, 14(3): 241-254.
- [18] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 390-391.
- [19] 张雨森, 李敏, 苟瑶, 等. 基于自注意相对鉴别对比学习的光学到SAR图像转换方法[J]. 火箭军工程大学学报, 2023, 37(1): 57-64.
- ZHANG Y S, LI M, GOU Y, et al. Contrastive learning for optical-SAR image translation with self-attention and relativistic discrimination[J]. Journal of Rocket Force University of Engineering, 2023, 37(1): 57-64.
- [20] PARMAR G, PARK T, NARASIMHAN S, et al. One-step image translation with text-to-image models[EB/OL]. (2024-03-18) [2025-05-26]. <https://arxiv.org/abs/2403.12036>.
- [21] 潘宏亮, 孙金霞, 韩希珍. 图像清晰度评价与变步长融合调焦方法[J]. 红外与激光工程, 2023, 52(1): 248-253.
- PAN H L, SUN J X, HAN X Z. Image sharpness evaluation and variable-step fusion focusing method[J]. Infrared and Laser Engineering, 2023, 52(1): 248-253.
- [22] ENSER P G B, SANDOM C J, LEWIS P H. Automatic annotation of images from the practitioner perspective[M]//Image and Video Retrieval. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 497-506.
- [23] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2009: 248-255.
- [24] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 3992-4003.
- [25] 蔡伟, 高蔚洁, 蒋昕昊, 等. 可见光条件下伪装目标智能化分割算法综述[J]. 火箭军工程大学学报, 2024, 38(2): 72-87.
- CAI W, GAO W J, JIANG X H, et al. Review of intelligent camouflaged object segmentation algorithms under visible light[J]. Journal of Rocket Force University of Engineering, 2024, 38(2): 72-87.
- [26] 陈璐, 李清格, 杨小冈, 等. 多源多视角飞行器认知导航数据集制备[J]. 火箭军工程大学学报, 2025, 39(3): 1-9.
- CHEN L, LI Q G, YANG X G, et al. Multi-source and multi-view aircraft cognitive navigation dataset construction[J]. Journal of Rocket Force University of Engineering, 2025, 39(3): 1-9.
- [27] LIN T Y, MAIRE M, BELONGIE S J, et al. Microsoft COCO: common objects in context[EB/OL]. (2014-5-01) [2025-05-26]. <https://arxiv.org/pdf/1405.0312>.

Construction of an Aircraft Multimodal Time-Sensitive Object-Detection Dataset

SHEN Tong¹, WANG Siyu², YANG Xiaogang¹, LU Ruitao¹

(1. Rocket Force University of Engineering, Xi'an 710025, China;

2. School of Electrical and Control Engineering, Xuzhou University of Technology, Xuzhou 221018, China)

ABSTRACT: To address the problems of existing multimodal object detection datasets, such as single target category, insufficient scene coverage, and image perspective mismatch with aircrafts, a multimodal time-sensitive object-detection dataset specifically designed for aircraft applications was constructed. This dataset aimed to provide data support for the training and evaluation of target detection algorithms. First, diverse time-sensitive target images were acquired by integrating real-scene shooting and remote sensing image platform generation. Second, spatial alignment among multimodal images was achieved using the SuperPoint and SuperGlue algorithms to ensure the positional consistency of targets across modalities. Furthermore, image synthesis and style transfer techniques were employed for data augmentation to balance the inter-modal distribution and expand the dataset scale. Finally, automatic annotation of time-sensitive targets was accomplished through a semantic segmentation algorithm. The experimental results demonstrate that compared with other multimodal target monitoring datasets, the constructed dataset is substantial in volume, diverse in target categories, and extensive in scene coverage.

KEYWORDS: multimodal object detection; dataset construction; image registration; data augmentation; image annotation

(编辑: 于福兴)