

# 基于语义地图的灾探无人机重定位方法\*

黎容熙<sup>1</sup>, 唐家成<sup>1</sup>, 胡天江<sup>1,2</sup>

1. 中山大学航空航天学院, 广东 深圳 518107
2. 中山大学人工智能学院, 广东 珠海 519080

**摘要:** 针对灾难环境可能造成全局定位系统的失效及可见光图像的退化, 以及传统的基于计算机视觉的重定位算法因图像特征点不足而成功率较低的问题, 提出了一种基于语义地图的无人机重定位方法。该方法依赖 RGB-D 图像来识别并构建受灾环境中的路标点, 通过与先验地图间的路标点匹配, 进而优化求解得到无人机的相对位姿。通过减少对目标识别网络潜在的广义物体识别能力的抑制, 得到图像中的高级别特征点, 有效解决了因特征点不足而难以重定位的问题。基于广义物体重建的路标点, 提出了一种快速的路标点检索与匹配方法。实验结果表明, 与基于目标识别网络的路标点构建方法相比, 本方法能在未知环境中重建更丰富的路标点, 并能有效地基于这些路标点进行重定位。在图像退化的灾难场景中, 本方法展现出比目前被广泛使用的图像检索方法更高的召回率和鲁棒性。

**关键词:** 城市火灾; 无人机; 重定位; 语义地图; 回环检测

**中图分类号:** TP242 **文献标志码:** A **文章编号:** 2097-0137(2025)03-0109-10

## A semantic map-based drones relocalization method for UAV in disaster exploration

LI Rongxi<sup>1</sup>, TANG Jiacheng<sup>1</sup>, HU Tianjiang<sup>1,2</sup>

1. School of Aeronautics and Astronautics, Sun Yat-sen University, Shenzhen 518107, China
2. School of Artificial Intelligence, Sun Yat-sen University, Zhuhai 519080, China

**Abstract:** In response to the potential failure of global positioning systems in disaster environments and the degradation of visible light images, as well as the low success rate of traditional computer vision-based relocalization algorithms due to insufficient image feature points, a semantic map-based drone relocalization method for unmanned aerial vehicle (UAV) is proposed. This method relies on RGB-D images to identify and construct landmark points in the disaster-affected environment. These landmark points are then matched with prior maps to optimize and estimate the relative pose of the drone. By reducing the suppression of the potential general object recognition capability within object recognition networks, high-level feature points in the image are obtained, effectively addressing the problem of difficult relocalization due to insufficient feature points. Building on the generalized object-based reconstruction of landmark points, an efficient method for retrieving and matching these points is proposed. Experimental results demonstrate that the approach can reconstruct a richer set of landmarks in unknown environments and effectively utilize them for localization, compared to other object recognition-based landmark point construction methods. In disaster scenarios with image degradation,

\* 收稿日期: 2024-12-11

录用日期: 2024-12-23

网络首发日期: 2025-01-24

基金项目: 广东省重点领域研发计划(2024B1111060004)

作者简介: 黎容熙(2000年生), 男; 研究方向: 视觉感知; E-mail: lirx67@mail2.sysu.edu.cn

通信作者: 胡天江(1979年生), 男; 研究方向: 群体智能, 集群系统; E-mail: hutj3@mail.sysu.edu.cn

全文阅读



ZR20240348

this method exhibits higher recall rates and robustness than widely used image retrieval methods.

**Key words:** urban fire; UAV; relocalization; semantic map; loop closure detection

为了保证救援工作中的准确性与实时性,利用无人机协同侦查(刘宇轩等,2020;潘礼规等,2023)成为提高受灾区域勘察效率的重要途径。然而,在灾难环境中全局定位方法,如GPS,通常受限严重。因无人机感知信息基于各自独立的局部坐标系,当两架无人机协作探索一个场景或单架多次探索同一场景时,仅共享基于局部坐标系的感兴趣目标信息是难以实现信息融合的。如何对齐不同架次无人机间的局部坐标系,成为亟待解决的问题。

可见光图像受环境的限制较小,能够提供丰富的灾区信息,是无人机感知信息的重要组成部分。因此,利用当前的视觉感知信息,与历史或其他无人机构建的先验场景模型进行匹配得到路标点对,重构出当前无人机相对于先验场景模型的相对位置和姿态。然后,针对上述问题,通过在共同环境中对无人机进行重定位,进而可以确立不同局部坐标系间的相对变换关系。

现有工作根据场景模型表达方式可分为基于图像特征数据库、隐式模型和三维模型三类。基于图像特征数据库的工作(Galvez-Lopez et al., 2011; 吴俊君等, 2013)是当前主流的重定位方法。以广泛用于SLAM算法中的回环检测方法为例,其聚合图像内的特征描述子作为全局描述以构建词袋模型。在重定位时,该算法基于词袋模型查询相似的历史图像,通过RANSAC算法匹配关键点,并使用PnP算法估计相机的位姿。近年,基于神经网络的图像匹配方法(Detone et al., 2018; Sarlin et al., 2020)极大地提高了图像匹配的性能。此类方法需要不断存储新增的图像或关键点及其描述子,增加了内存的负担,同时随着数据规模上升也导致了查询效率的降低。另外,此类算法描述子关注图像中的角点、边缘等低级别特征,缺乏独特性且易受干扰。在火灾等灾难场景中,由于火焰和烟雾造成画面纹理弱、光照变化大等情况,削弱了低级别图像特征的稳定性,因此该方法在实际灾难场景应用中成功率较低。尽管谢红玉(2024)和丁静(2024)针对烟雾场景中图像的退化引入了图像增强技术,其仍未改变低级别特征在该场景中的局限性。

基于隐式模型的方法通过场景数据训练一个

神经网络,使其能端到端的输出相机位姿(Kendall et al., 2015)或场景特征点(Do et al., 2022, 2024)。这类方法的泛化性较差,通常需要先获取工作场景的环境图像数据,并进行针对性的训练,难以直接投入消防救援等各种未知环境中完成重定位任务。

基于三维模型的方法首先建立环境模型,并将所查询图像的2D或3D特征点与模型的3D特征点进行匹配,根据特征点之间的匹配关系即可进行位姿估计(Sattler et al., 2011)。最近, Liu et al. (2019, 2023)通过语义分割网络将环境语义嵌入三维模型当中,提升了算法在复杂场景中的鲁棒性。一些研究(Li et al., 2020; Zins et al., 2022; Mahattansin et al., 2022; Qian et al., 2022)构建了物体级别的地图以估算相机位姿,但这些方法多面向AR场景,关注小范围特定物体的持续跟踪。灾难救援工作更侧重于不确定环境中,对未预定义物体进行有限次数观测,并构建物体语义地图以供无人机重定位。另外,受限于前端物体识别网络有限的物体类别识别能力,在灾难场景中可能无法提供准确且足够的物体类别信息,导致缺乏有效的物体语义地图用于无人机重定位。

因此,针对灾难场景中由于图像退化导致的低层级特征的局限性问题(李海顺等,2024),本文提出了一种基于高级别特征点的无人机重定位方法,使用神经网络提取图像中的高级别特征点作为路标点。同时,因火灾、地震等未知灾难场景的复杂性和不可预测性,以及预先采集场景中物体进行训练存在困难且场景中普遍存在的烧毁、破碎物体难以识别等,本文在训练与推理阶段分别采用了不同的网络头部层,以实现灾难场景中广义物体的识别。最后,基于此设计了路标点检索与匹配方法,以及无人机重定位方法。

## 1 灾难环境的重定位原理

### 1.1 基于三维模型的重定位原理

本文面向灾难环境的重定位方法是基于三维模型的重定位方法,工作原理如图1所示。其输入包含两个部分:先验三维模型 $\{P\}$ 是历史构建的模型或其他无人机实时构建的模型;环境感知信息则涉及可见光图像 $I_{rgb}$ 、深度图像 $I_d$ 以及惯性传感

器信息  $I_{\text{imm}}$ 。基于环境感知信息构造无人机的视觉惯性里程计, 其记录了该无人机在局部坐标系下的位置和姿态变化。

对于一帧需要被重定位的图像  $I$ ,  $I$  首先被输入至广义物体点识别网络, 该网络能够提取  $I$  中广义物体点  $K$  的像素坐标、像素大小及其特征描述。随后, 通过对连续多帧的物体点  $K$  进行跟踪, 筛选出稳定的物体点。结合深度信息以及无人机的里程计信息, 重建  $K$  的三维位置和尺度, 从而构建出一系列的路标点  $\{\hat{P}\}$ 。在无人机的重定位阶段, 通过在先验三维模型  $\{P\}$  中检索当前重建的路标点  $\{\hat{P}\}$ , 确定对应的匹配区域。当确认了位置初值后, 对当前  $\{\hat{P}\}$  与  $\{P\}$  匹配区域中的路标点进行重新匹配, 形成路标点  $\{\hat{P}_i, P_i\}$ 。最终, 通过优化路标点之间的位置误差, 确定相机相对于先验三维模型的相对位置和姿态。

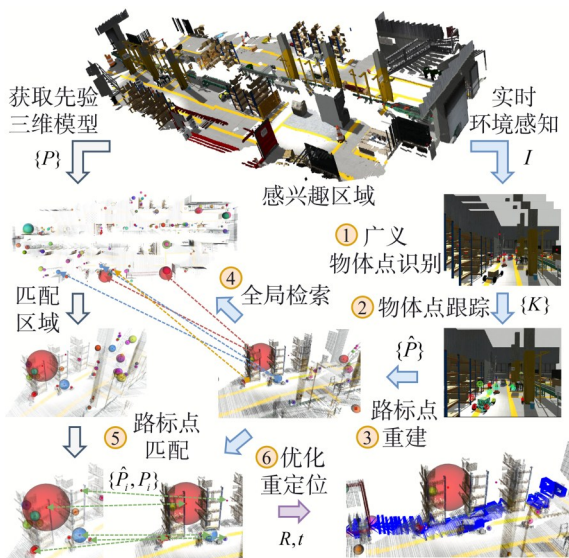


图1 灾难环境的重定位方法

Fig. 1 Relocalization method for disaster environments

在上述过程中, 路标点重建的特征质量与数量是重定位算法性能的关键因素。本文引入了一个神经网络来提取图像中的高级别特征点, 这些特征点更加鲁棒, 有助于在复杂环境中进行准确的检索和匹配, 解决灾难场景中烟雾、火光等因素导致的低级别特征不稳定问题。物体作为灾难场景中的关键组成部分, 其在环境中的独特性使其能够稳定地被重新识别, 并有效地表征当前环境的地理位置, 因此本文采用物体特征作为高级别特征。

## 1.2 广义物体定义

由于传统目标识别网络在物体类别识别能力上的限制, 本文不直接使用网络对特定物体的识别结果, 而是通过调整网络的输出, 减少对非训练数据集包含的标签类别的抑制, 从而允许网络在推理阶段识别更广泛的物体。这些被识别的物体点可能不属于训练时使用的特定类别, 本文将这种不局限于特定类别的物体点称为广义物体。目标识别网络训练过程中的数据集信息记为

$$\Omega = \{B_i, C_i\}_{i=1}^N,$$

$\Omega$  为整体数据集,  $\{\hat{B}_i\}$  与  $\{\hat{C}_i\}$  分别代表所识别出的目标物体的外接矩形和其所对应的标注;  $N$  表示训练数据集大小。定义网络输出的广义物体

$$\hat{Y} = \{\hat{B}_i, \hat{C}_i\}_{i=1}^k,$$

其中  $\{\hat{C}_i\} \subseteq \{C_i\}$ , 即神经网络对物体类别识别能力大于等于传统的目标识别输入。由于  $\{C_i\}$  无法被有监督训练, 故广义物体通常使用其长度为  $N$  的特征  $D_i$  来表述, 即  $\hat{Y} = \{\hat{B}_i, D_i\}_{i=1}^k$ 。

## 2 广义物体点识别方法

目前被广泛应用于物体的重定位工作中的基于先验锚框的方法, 如 YoloV3 (Redmon et al., 2018) 引入了物体大小的先验知识, 同时也为广义物体识别带来了额外的约束。本文利用上述对广义物体的定义, 受无先验锚框目标识别网络 CenterNet (Duan et al., 2019) 的启发, 提出了一种广义物体点识别网络, 如图 2 所示。CenterNet 目标识别网络通过骨干网络、编码器和解码器这三个模块, 提取图像的普适性特征, 并在头部层对目标的位置、尺寸及类别进行预测。在常规的目标识别任务中, 往往侧重于物体置信度的绝对值, 通过比较各分类置信度是否超越既定阈值来筛选目标。由于有监督训练过程中会降低热图中非目标分类的物体概率, 故而阈值策略能有效排除数据集中未出现的物体。这意味着经过有限类别数据集的有监督训练, 网络已具备识别一定通用物体的能力。本文旨在实现的广义物体识别, 目的是防止在推理阶段对数据集中未包含物体的置信度进行抑制。

下面, 详述该网络的总体架构、各模块深度神经网络的设计以及推理头部层实现广义物体识别的具体策略。

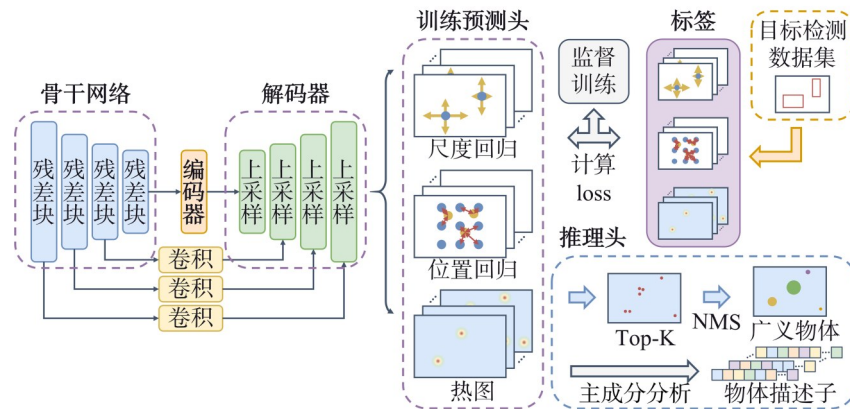


图2 广义物体识别的网络架构

Fig. 2 Generalized object recognition network framework

## 2.1 网络架构与训练

在训练阶段, 广义物体点识别网络架构包括4个主要模块: 骨干网络、编码器、解码器和训练预测头。首先, 该网络以RGB图像为输入, 经过前3个模块提取物体的通用深度特征, 最后通过训练预测头得到有限类别的目标识别推理结果。

物体的通用深度特征的提取首先输入骨干网络中, 由于本文更侧重于提取场景中物体的共性特征, 因此采用了较浅的ResNet-18作为图像特征提取的骨干网络, 避免增加额外的参数量和训练集标签的过拟合问题。在编码器层, 为了识别图像中不同尺度的物体, 特征图金字塔网络(FPN, feature pyramid networks)常被应用于目标识别网络中。然而, FPN引入了多个不同尺度的特征图, 带来了目标分配问题, 不利于端到端地学习物体的共性特征与广义物体的输出。受YOLOF(Chen et al., 2021)的启发, 采用了基于空洞卷积的编码器。通过堆叠空洞率依次递增的空洞卷积(dilation convolution), 获得不同大小感受野的深度特征编码。随后, 解码器通过最近邻插值对深度特征进行上采样。在编码器和解码器对应层中加入了长跳跃连接, 融合不同尺度的上下文信息, 并提供了直接的梯度路径, 缓解在训练深层网络时可能出现的梯度消失问题。

训练预测头根据上文中计算的通用物体深度特征预测各类物体在图像上的热力图、目标的偏移量和目标的尺度, 表示为 $\hat{Y} = \{\hat{M}_{i=1}^N, \hat{O}, \hat{S}\}$ 。训练集 $\Omega$ 也将被构造为 $Y = \{\hat{M}_{i=1}^N, O, S\}$ , 在每一个数据集中物体的外接矩形中心位置创建一个正样本点, 并在其他位置创建负样本点。生成正样本点时使用高斯核函数, 使得该像素为该类别物体在中

心点处达到峰值, 随着距离的增加而逐渐减小。

## 2.2 广义物体推理头

在推理阶段, 推理头综合训练预测头中预测的所有热图, 以获得图像中每个像素点为物体的概率, 并输出仅考虑置信度的相对值。从热图中选择置信度最高的 $k$ 个点作为广义物点, 具体过程如下:

首先, 对预测的热图 $H_{whc}$ (下标 $w, h, c$ 分别表示宽、高和通道数)沿通道方向进行聚合, 以得到广义物体置信度热图

$$M_{wh} = \text{sigmoid} \left( \sum_{c=1}^N H_{whc} \right);$$

其次, 对这个热图进行最大池化处理, 即在每个像素点的邻域中选取最大值作为该点的热值, 以抑制热图中的非极大值

$$M_{wh} = \max_{\substack{i \in [w-k, w+k], \\ j \in [h-k, h+k]}} (M_{ij}).$$

从热图 $M$ 中选择置信度最高的 $k$ 个点, 得到 $q_i^k = \{u_i, v_i, p_i\}$ , 其中 $u, v, p$ 分别代表该点在图像中的横坐标、纵坐标和概率。尺度回归为根据训练预测头中的亚像素偏移量和尺度预测构建广义物体外接矩形

$$B_i = \left( u_i + \hat{O}_u - \frac{\hat{S}_u}{2}, v_i + \hat{O}_v + \frac{\hat{S}_v}{2} \right).$$

$B_i$ 表示为该外接矩形的左上角和右下角。

最后采用极大值抑制技术进一步去除重复检测的广义特征点, 并将最终剩下的点作为广义物体路标点。这种方法有效避免了非极大值抑制(NMS)和标签分配对训练的影响, 以及对不在训练数据集中的物体( $c \notin C$ )的抑制。

### 3 基于广义路标点的重定位算法

利用上述方法确定广义物体点之后, 本研究将其转换为三维路标点并应用于无人机重定位任务中。基于广义物体路标点的重定位技术核心在于快速实现路标点的检索与匹配, 构建局部路标点与三维模型路标点之间的映射关系, 从而推算出相机的相对位姿。如图3所示, 本文提出了一种基于KD树和路标点哈希表的特征点索引与匹配方法, 该方法综合考虑了物体点及其邻居的三维坐标位置以及深度特征的相似性, 以提升索引与匹配的效率 and 准确性。

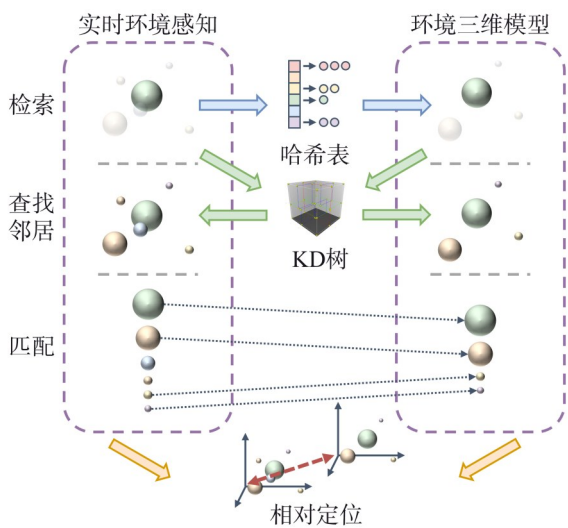


图3 基于广义物体的重定位方法

Fig. 3 Relocalization method based on generalized objects

#### 3.1 广义路标点检索与匹配

在灾难环境搜索任务中, 通常不只关注某个特定位置, 而是涉及对整个区域的迅速覆盖。然而, 随着搜索范围的扩大, 对广义物体特征描述子进行逐一比对的开销亦随之增加。为了在灾难场景的三维模型中实现广义路标点的高效索引, 针对基于深度特征索引和基于位置索引两种情况, 分别构建了一套基于哈希表的索引系统, 用以检索三维模型中与广义路标点可能匹配的所有点; 同时构建了一个KD树结构, 旨在快速查找广义路标点的邻近路标点。其中, 在构建索引表过程中, 针对三维模型中每一个广义物体路标点, 将选取其深度特征描述子中相似度最大标签作为其在全局路标点索引中的键值。

在上述索引机制的基础上, 本文通过目标广义物体路标点的最大相似标签, 从三维模型中提取了可能匹配的点的集合

$$H(L(\hat{P})) = \{P_0, P_1, \dots, P_n\},$$

$H(\cdot)$ 代表特征索引哈希表,  $L(\hat{P})$ 表示点 $\hat{P}$ 的最大相似标签,  $P_i$ 是候选匹配点。

随后, 查找目标广义路标点的邻域点集 $N(\hat{P})$ 以及候选点邻域集 $\{N(P_0), N(P_1), \dots, N(P_n)\}$ , 此时匹配目标邻域点集 $N(\hat{P})$ 与候选邻域点集 $N(P_i)$ 得到相似度分数, 从而选择最优的候选点作为正确匹配。该匹配问题实际上可以视为一个最优分配问题, 给定一个相似度矩阵, 最大化其匹配物点对的相似度分数之和, 本文采用匈牙利算法。物体相似度将从位置、尺度和特征描述子综合考虑。另外, 考虑到匈牙利算法尽可能多匹配的特点会造成物点的错误匹配, 当位置、尺度、特征三方面任意一个指标小于其对应阈值 $T$ , 该点将不被匹配。相似度

$$s = \begin{cases} \alpha s_{pos} + \beta s_{scale} + \gamma s_{desc}, & s_{pos} > T_{pos}, s_{scale} > T_{scale}, s_{desc} > T_{desc} \\ 0, & \text{其他} \end{cases}$$

其中 $s_{pos}$ 为位置误差,  $s_{scale}$ 为尺度误差,  $s_{desc}$ 为特征误差。

为进一步加速重定位过程, 本文根据目标点与候选点的特征相似度对候选点邻域 $N(P_i)$ 进行排序, 并在成功匹配到 $m$ 对广义物体路标点匹配 $\{\hat{P}_i, P_i\}$ 后停止匹配过程, 此时获取了两个局部坐标系的相对位姿初值。利用这一相对位姿初值, 本文将再次进行广义物体路标点的匹配, 以加入由于提前终止匹配而未匹配的潜在点对, 从而平衡路标点匹配个数与匹配时间。

#### 3.2 基于路标点重的重定位

无人机在计算相对定位过程中, 首先是对异常匹配的点对进行检测与剔除。在这一阶段, 假设两个坐标系间点对的位移应表现出较高的一致性。本文采用迭代方法, 对所有点对内部距离的中值进行分析, 从而逐个识别并移除那些距离超出预定阈值的点对。在此基础上, 通过计算所有剩余点对的位移均值, 得到了当前无人机在三维模型中位置的初值。

随后, 为了进一步优化无人机的位移和旋转参数。本文通过最小化当前观测路标点集 $p_i$ 和三维模型中对应路标点集 $q_i$ 之间的观测误差, 采用Levenberg-Marquardt算法迭代求解最优的无人机的位姿。最小化路标点间位置误差为

$$E(R, t) = \sum_{i=1}^n \|q_i - (Rp_i + t)\|^2,$$

其中  $R, t$  为待优化的无人机相对于三维模型的姿态和位置。在完成优化后, 若定位结果与初步估计值的误差小于预设的阈值, 则认为此次重定位结果是可靠的。

#### 4 面向救援场景的算法验证

本节将从路标点的重建以及基于路标点的重定位两个部分对算法进行验证。在路标点重建部分, 算法对各类物体的重建能力及其所重建路标点的数量, 将直接影响后续重定位步骤的成功率。因此, 重建路标点的数量是评价该算法性能的关键指标。在重定位部分, 召回率作为衡量重定位成功率的重要指标, 直接反映了算法在灾难环境中的性能表现。此外, 重定位误差也是评估算法有效性的重要依据。定义各轴向定位误差分别为  $e_x, e_y, e_z$ , 整体位置误差为  $e_p$ , 角度误差为  $e_r$ 。由于

本文每构建一个关键帧将会尝试进行一次重定位, 召回率被定义为成功重定位的关键帧数量与全部关键帧数量之比。位置误差与旋转误差将比较重定位结果与真值的误差。

通过 Gazebo 平台搭建受灾工厂环境, 包含了消防用具、工厂流水线、办公区域、杂物区、货架等部分, 其中消防用具不在目标识别数据集中, 且货架中摆放的箱子相似性较高, 此外物体存在破碎的情况, 这都为重定位工作带来了挑战。无人机的飞行控制单元采用 PX4, 并配备了 Realsense D435i 深度相机, 能够输出可见光及深度图像, 该组合能够在仿真与实物环境中提供统一的对外接口。此外, 利用基于 RGB-D 的视觉惯性里程计 (Liu et al., 2022) 获取无人机的轨迹。为了进一步验证该算法在火灾场景中的有效性, 本文基于 Sakaridis et al. (2018) 的方法模拟了火灾的烟雾, 并通过随机改变图像亮度值与色相模拟火焰对图像的影响, 如图 4 所示。

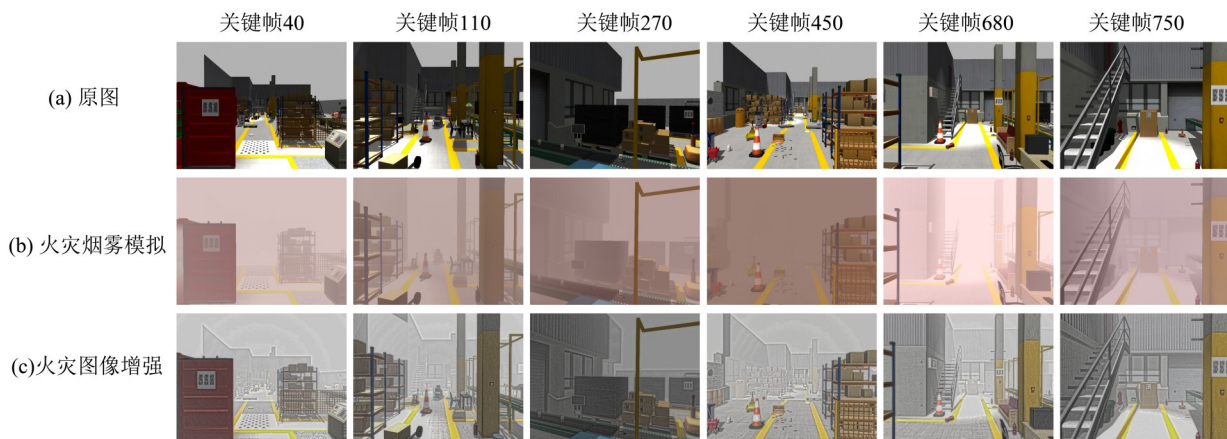


图 4 仿真场景及烟雾模拟效果

Fig. 4 Simulation scenario and smoke simulation effects

##### 4.1 广义物体识别与重建结果

本文在仿真与真实场景中对比了基于传统目标识别重建方法和本文的广义路标点重建方法。实验中, 传统的目标识别网络与本文的广义物体识别网络均采用了 COCO 数据集 (Lin et al., 2015) 进行训练。

在仿真场景的测试中, 如图 5 所示, 对基于目标识别的方法与本文提出的广义物体识别方法在路标点重建方面进行了对比。结果显示, 受限于训练集类别, 传统目标识别方法在未知灾难环境中难以检测到足够的路标点, 而本文的广义物体识别网络不受此限制, 重建的路标点数量显著多

于传统方法。

图 6 展示了灾难场景中不同物体的重建效果, 其中带有星号 (\*) 标记的物体不包含在训练用的 COCO 数据集中。本文提出的广义物体识别网络展现了对于数据集外物体的识别能力, 例如灭火器、安全出口指示灯、路障等。同时能对这些物体进行描述, 这在图中路标点的颜色编码上得到了体现。尽管数据集中包含了完整的花瓶类别, 但传统目标识别方法未能识别并重建破碎的花瓶, 而本文所提出的方法则能够有效地将其主要碎片重建为路标点。

在真实场景的实验中, 通过在可见光图像中

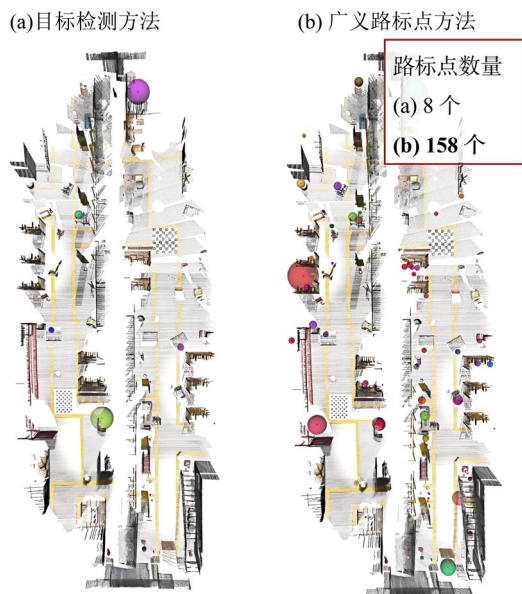


图5 两种重建方法下的三维地图以及路标点数量

Fig. 5 The number of reconstructed landmark and 3D map under the two reconstruction methods

加入不同浓度的烟雾来模拟火灾环境。如图7所示, 本文方法对烟雾引起的图像退化、光照变化表现出较强的鲁棒性。

实验结果证实, 所提出的广义物体识别网络能够从有限类别数据集中学习到泛化能力较强的物体特征, 对环境变化具有较强的抗干扰性, 对物体类别与形态展现出良好的鲁棒性。与传统目标识别网络相比, 本文方法在重建路标点数量上更具优势, 更适应灾难环境下的路标点识别与重建, 为后续重定位任务提供了有效的路标点支撑。

#### 4.2 单机多架次重定位结果

单机多架次实验旨在检验在路标点重建任务完成后, 无人机在面对离线构建的先验三维模型时的重定位性能表现。首先, 无人机进入未知灾难区域构建三维模型。在后续作业中, 该无人机将再次进入同一环境, 执行重定位任务以验证模型的准确性。如图8所示, 场景中分布的不同尺寸的球体代表了重建出的广义物体, 无人机重定位的位置和姿态以蓝色标记显示。

实验对比了本文提出的重定位方法与基于词袋模型的图像检索方法在重定位召回率方面的表现, 其中图像检索方法采用了ORB描述子(Mur-Artal et al., 2017)。针对火灾、极端天气等退化场景, 基于传统的图像检索方法通常使用图像增强方法减少图像退化对召回率的影响, 图4(a)展示了

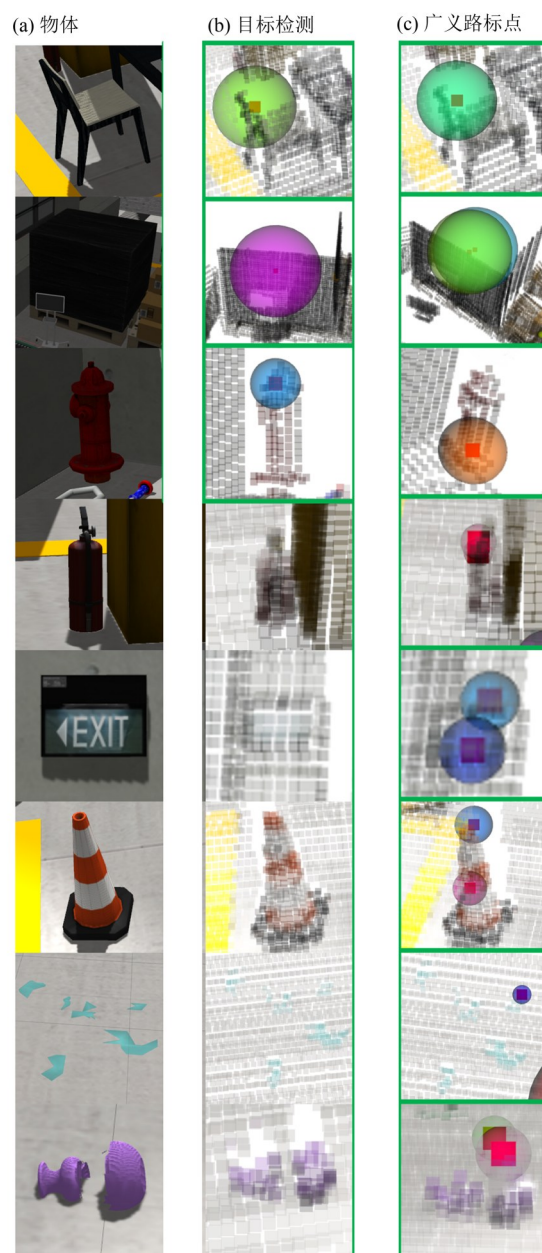


图6 灾难场景中物体的重建结果

Fig. 6 Reconstruction results for objects in a disaster scene

增强后的图像。

实验结果如表1所示, 其中标注“\*”的数据代表火灾模拟环境下的实验结果。相较于传统的基于词袋模型的图像检索方法, 本文所提方法在召回率上有显著优势, 并在火灾模拟环境中更为鲁棒。火灾场景下, 传统方法的ORB特征点的FAST角点在可见光图像中数量锐减, 同时BRIEF描述符易受环境亮度波动的影响, 使得即便图像经过增强处理召回率依旧偏低。本文方法利用环境中更为稳定的高级特征点, 即使在火灾等退化环境下也具有较高的鲁棒性。

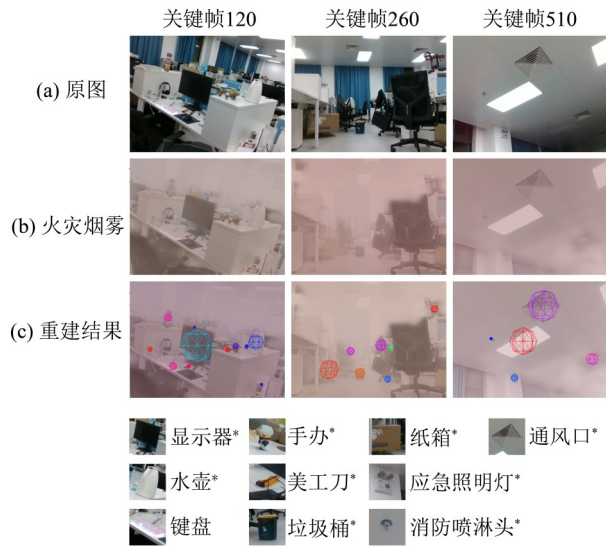


图7 实验场景及路标点的重建结果

Fig. 7 Experimental scenario and landmark point reconstruction results

在路标点较为丰富的情况下，本文提出的重定位方法能够达到较高的精度，如表2所示。然而，在区间II和IV中，成功识别并重建的路标点数量有限，导致重定位失败。因此，增加路标点的数量对于提高本文重定位方法的召回率至关重要。

表1 单机多架次实验的重定位召回率

Table 1 Recall rate of relocalization in single-drone multi-flight experiments

项目	全局	区间I	区间III	区间V
ORB	0.288	0.236	0.284	0.382
本文方法	0.566	0.920	0.682	0.829
增强ORB*	0.061	0.023	0.119	0.074
本文方法*	0.164	0.321	0.240	0.113

\*标记代表火灾模拟环境下的实验结果。

量有限，导致重定位失败。因此，增加路标点的数量对于提高本文重定位方法的召回率至关重要。

### 4.3 双机重定位结果

双机重定位实验旨在验证两架无人机中在线构建先验模型与进行环境感知的重定位可行性。两架无人机协同进入未知灾难环境，无人机1与无人机2能够实时构建各自的广义路标点三维模型，并实时交换这些模型信息。在探索过程中，每架无人机都会尝试将自身的观测结果在对方的三维

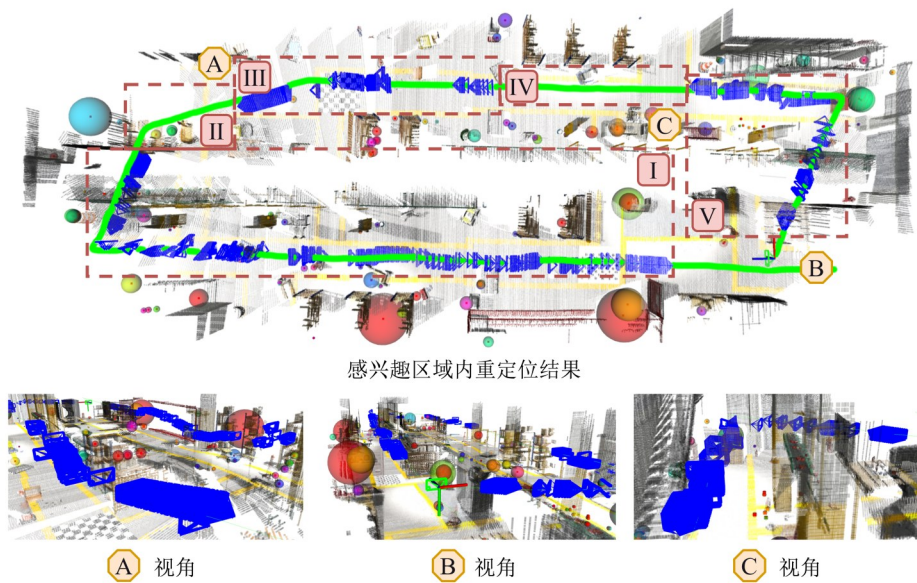


图8 单机多架次实验的重定位结果

Fig. 8 Single-drone multi-flight relocalization results

表2 单机多架次实验重定位精度

Table 2 Relocalization accuracy in single-drone multi-flight experiments

项目	x方向误差/m	y方向误差/m	z方向误差/m	定位误差/m	方向误差/(°)
区间I	0.150±0.331	0.042±0.099	0.205±0.340	0.279±0.436	0.707±0.945
区间III	0.293±0.396	0.383±0.599	0.072±0.116	0.520±0.633	1.587±1.725
区间V	0.036±0.087	0.175±0.290	0.097±0.146	0.210±0.320	0.526±0.823
全局	0.149±0.348	0.149±0.412	0.148±0.288	0.310±0.507	0.840±1.360

模型中进行重定位。

如图9所示, 无人机2在标记为绿色的区域成功地无人机1的三维模型中实现了重定位。在实验过程中, 无人机1的起始点靠近绿色框区域, 优先完成了蓝色框内广义物体的重建, 并将其作为路标点存储于三维模型中, 随后传输至无人机2; 无人机2抵达绿色框位置后同样完成对蓝色框广义物体的重建, 并通过检索确认了无人机1的三维模型中存在相应的路标点, 利用路标点完成了重定位。

重定位的误差如表3所示, 其中 $x$ 轴与 $y$ 轴误差处于正常范围,  $z$ 轴即高度方向误差较大, 这是由于路标点 $z$ 轴差异较小导致的 $z$ 轴上定位退化。实验结果表明, 本文提出的无人机重定位算法能够在实时更新的先验三维模型中完成重定位。

## 5 结语

本文提出了一种基于广义路标点的无人机重定位技术。针对灾难场景下特征缺失的问题, 引入深度神经网络以识别图像中的高级别特征, 提出了一种广义物体点识别与重建方法。该方法侧重于图像中物体的相对置信度, 自适应地选取受灾环境中具有高辨识度的广义物体点, 并将其重建为路标点, 从而无需针对特定环境重新训练识

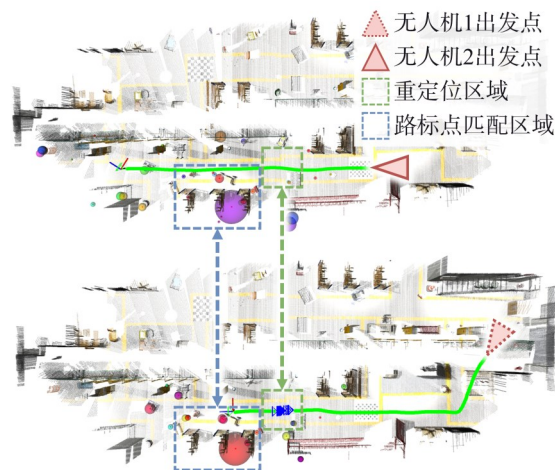


图9 双机重定位结果

Fig. 9 Dual-drone relocalization results

别网络。在此基础上, 本研究基于广义物体语义地图, 设计了一套广义物体检索与匹配方法以及无人机重定位策略。

仿真与实验结果证实, 本方法在灾难环境中具备更强的路标点重建能力, 且重定位算法在离线与在线场景中具备更强的有效性。相较于传统的基于词袋模型的图像检索方法, 本文在灾难场景中重定位的成功率展现出明显优势, 能够有效解决灾难环境中无人机的重定位问题。

表3 双机实验重定位精度

Table 3 Relocalization accuracy in dual-drone experiments

$x$ 方向误差/m	$y$ 方向误差/m	$z$ 方向误差/m	定位误差/m	方向误差/(°)
0.128	0.035	0.519	0.536	1.056

## 参考文献:

- 丁静, 2024. 烟雾环境下运动目标双目视觉定位技术研究[J]. 测绘学报, 53(3):582.
- 李海顺, 李兴东, 2024. 基于改进 CycleGAN 的林火图像烟雾滤除算法研究[J]. 消防科学与技术, 43(11):1596-1602.
- 刘宇轩, 刘虎, 田永亮, 等, 2020. 面向林火持续侦察的多无人机分布式控制方法[J]. 航空学报, 41(2): 272-287.
- 潘礼规, 尹佳琪, 徐春光, 2023. 基于载波相位观测的无人机集群相对定位方法[J]. 中山大学学报(自然科学版中英文), 62(3): 125-136.
- 吴俊君, 胡国生, 2013. 室内环境仿人机器人快速视觉定位算法[J]. 中山大学学报(自然科学版), 52(4):7-13.
- 谢红玉. 2024. 多源信息融合的移动机器人SLAM方法研究[D]. 北京:北京化工大学.
- CHEN Q, WANG Y, YANG T, et al, 2021. YOLOLF: You only look one-level feature[J/OL]. <https://arxiv.org/pdf/2103.09460>.
- DETONE D, MALISIEWICZ T, RABINOVICH A, 2018. SuperPoint: Self-supervised interest point detection and description [C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, UT, USA.
- DO T, MIKSIK O, DEGOL J, et al, 2022. Learning to detect scene landmarks for camera localization [C]//IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA.
- DO T, SINHA S N, 2024. Improved scene landmark detection for camera localization [C]//2024 International Conference on 3D Vision. Davos, Switzerland.
- DUAN K, BAI S, XIE L, et al, 2019. CenterNet: Keypoint triplets for object detection [C]//IEEE/CVF International Conference on Computer Vision. Seoul, Korea.
- GALVEZ-LOPEZ D, TARDOS J D, 2011. Real-time loop detection with bags of binary words [C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, California, USA.
- KENDALL A, GRIMES M, CIPOLLA R, 2015. PoseNet: A convolutional network for real-time 6-DOF camera relocalization [C]//IEEE International Conference on Computer Vision. Santiago, Chile.
- LI J, KOREITEM K, MEGER D, et al, 2020. View-invariant loop closure with oriented semantic landmarks [C]//IEEE International Conference on Robotics and Automation. Paris, France.
- LIN T Y, MAIRE M, BELONGIE S, et al, 2015. Microsoft COCO: Common objects in context [EB/OL]. arXiv.1405.0312.
- LIU C, SHEN S, 2023. Towards view-invariant and accurate loop detection based on scene graph [C]//IEEE International Conference on Robotics and Automation. London, UK.
- LIU J, LI X, LIU Y, et al, 2022. RGB-D inertial odometry for a resource-restricted robot in dynamic environments [J]. IEEE Robotics and Automation Letters, 7(4): 9573-9580.
- LIU Y, PETILLOT Y, LANE D, et al, 2019. Global localization with object-level semantics and topology [C]//International Conference on Robotics and Automation. Montreal, Canada.
- MAHATTANSIN N, SUKVICHAI K, BUNNUN P, et al, 2022. Improving relocalization in visual SLAM by using object detection [C]//19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Prachuap Khiri Khan, Thailand.
- MUR-ARTAL R, TARDOS J D, 2017. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 33(5): 1255-1262.
- QIAN Z, FU J, XIAO J, 2022. Towards accurate loop closure detection in semantic SLAM with 3D semantic covisibility graphs [J]. IEEE Rob Autom, 7(2): 2455-2462.
- REDMON J, FARHADI A, 2018. YOLOv3: An incremental improvement [J/OL]. arXiv.1804.2767.
- SAKARIDIS C, DAI D, van GOOL L, 2018. Semantic foggy scene understanding with synthetic data [J]. J Computer Vision, 126(9): 973-992.
- SARLIN P E, DETONE D, MALISIEWICZ T, et al, 2020. SuperGlue: Learning feature matching with graph neural networks [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA.
- SATTLER T, LEIBE B, KOBELT L, 2011. Fast image-based localization using direct 2D-to-3D matching [C]//International Conference on Computer Vision. Colorado Springs, CO, USA.
- ZINS M, SIMON G, BERGER M O, 2022. OA-SLAM: Leveraging objects for camera relocalization in visual SLAM [C]//International Symposium on Mixed and Augmented Reality. Daejeon, South Korea.

(责任编辑 王海蓉)