doi:10.11887/j.issn.1001-2486.25120033　　http://journal.nudt.edu.cn

# 低空无人飞行器绝对视觉定位技术综述

叶熠彬[1,2],陈　硕[1,2],滕锡超[1,2],李　璋[1,2*],杨鸿睿[1,2],宋潇铠[1,2],于起峰[1,2]
(1. 国防科技大学 空天科学学院,湖南 长沙　410073;2. 图像测量与视觉导航湖南省重点实验室,湖南 长沙　410073)

**摘　要**:针对低空无人飞行器在全球卫星导航系统(global navigation satellite system,GNSS)拒止环境下自主定位的迫切需求,系统综述了基于"检索—匹配—位姿解算"框架的飞行器绝对视觉定位技术。讨论了低空观测带来的成像差异、场景尺度变化和地物遮挡等问题,阐明了该分层定位框架在解决大范围、长航时定位问题上的技术优势。在此基础上,分别从跨视角图像检索、像素级特征匹配及飞行器位姿解算三个核心模块,系统梳理了从传统手工特征到深度学习范式的技术发展趋势与研究现状。结合机载边缘计算平台的部署需求,分析了现有技术局限并展望了未来研究方向。本综述可为低空飞行器绝对视觉定位的技术研究与工程应用提供参考。

**关键词**:低空无人飞行器;视觉定位;跨视角图像检索;跨视角图像匹配;位姿解算;GNSS 拒止环境
**中图分类号**:V19;TP751　　**文献标志码**:A　　**文章编号**:1001-2486(2026)02-029-19

# Survey on absolute visual localization techniques for low-altitude unmanned aerial vehicles

*YE Yibin*[1,2], *CHEN Shuo*[1,2], *TENG Xichao*[1,2], *LI Zhang*[1,2*], *YANG Hongrui*[1,2], *SONG Xiaokai*[1,2], *YU Qifeng*[1,2]
(1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. Hunan Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China)

**Abstract**: To address the critical need for autonomous navigation of low-altitude UAVs (unmanned aerial vehicles) in GNSS (global navigation satellite system)-denied environments, a comprehensive survey was presented on absolute visual localization techniques based on a "retrieval-matching-pose estimation" framework. Key challenges inherent to low-altitude UAV observations—including significant imaging disparities, scale variations, and object occlusions—were analyzed, thereby elucidating the advantages of this hierarchical framework for large-scale, long-endurance localization tasks. Subsequently, the technological evolution and state-of-the-art advancements across three core components (cross-view image retrieval, pixel-level feature matching, and UAV pose estimation) were systematically reviewed, tracing the progression from traditional handcrafted features to deep learning paradigms. Finally, considering the deployment requirements of onboard edge computing platforms, the limitations of existing technologies were discussed, and promising future research directions were outlined. This survey is intended to serve as a valuable reference for both research and practical applications in absolute visual localization for low-altitude UAVs.

**Keywords**: low-altitude UAV; visual localization; cross-view image retrieval; cross-view image matching; pose estimation; GNSS-denied environments

无人飞行器因其成本低廉、机动灵活、操作便捷等突出优点,已广泛应用于物流配送、灾害救援、航拍摄影、军事侦察等诸多领域[1],在这些复杂多样的任务场景中,高精度的自主定位是确保无人飞行器安全、高效完成飞行作业的前提[2]。目前,主流的无人飞行器自主定位技术主要依赖于全球卫星导航系统(global navigation satellite system,GNSS)与惯性导航系统(inertial navigation system,INS)的组合。在开阔、无干扰环境中,GNSS 能够提供米级的定位精度[3]。然而,GNSS 信号易受地形干扰,导致无人飞行器在城市、山地等区域低空飞行时的定位精度下降严重。此外,在军事对抗等特定场景下,GNSS 信号还易受到干扰和欺骗,致使定位精度急剧下降[4]。INS 不

**收稿日期**:2025-12-15
**基金项目**:国家自然科学基金资助项目(12472189)
**第一作者**:叶熠彬(2001—),男,福建宁德人,博士研究生,E-mail:yeyibin18@nudt.edu.cn
***通信作者**:李璋(1985—),男,湖南岳阳人,副研究员,博士,硕士生导师,E-mail:zhangli_nudt@163.com
**引用格式**:叶熠彬,陈硕,滕锡超,等. 低空无人飞行器绝对视觉定位技术综述[J]. 国防科技大学学报,2026,48(2):29-47.
**Citation**: YE Y B, CHEN S, TENG X C, et al. Survey on absolute visual localization techniques for low-altitude unmanned aerial vehicles[J]. Journal of National University of Defense Technology, 2026, 48(2): 29-47.

依赖外部信号,但高精度惯导设备(如光纤或激光惯导)成本高昂,难以在消费级无人机上普及;而低成本的微机电系统(micro-electro-mechanical system,MEMS)惯导的位姿解算存在显著的时间累积误差[5],在 GNSS 拒止条件下无法满足飞行器长航时、高精度定位需求。

为了应对上述挑战,绝对视觉定位作为一种重要的自主定位方法,近年来受到学术界与工业界的广泛关注[4-9],该技术主要通过机载实时对地成像(实时图)与预先地理编码的卫星或航空影像(基准图)进行匹配,建立实时图与基准图之间的几何映射关系,最终解算出无人飞行器的绝对位置。与惯性导航和卫星导航相比,绝对视觉定位技术具有无漂移以及抗电磁干扰等优势,目前在军用和民用两个领域均得到应用验证。例如,美军"战斧"巡航导弹在末制导中采用景象匹配技术来修正惯性导航的累积误差、提升打击精度;美军"黑寡妇"无人机部署 Palantir 公司开发的视觉导航样机,完成了 GNSS 拒止下的自主导航试验验证[7];美国"毅力号"火星车在进入火星大气层的着陆阶段,利用视觉导航系统实时识别预定着陆区的地形以实现精准软着陆[8]。此外,商业航天公司 Maxar 开展了 P3DR(Precision 3D Registration)视觉辅助导航项目,实现了航空遥感数据的精确地理配准;国内 DJI 公司开发了 Guidance 视觉传感导航系统,可用于室内及低空场景自主避障和小范围导航[9]。

尽管无人飞行器绝对视觉定位研究已有数十年历史,但其应用主要集中于高空(飞行高度通常大于 500 m)、正下视成像条件[10-11](如图 1 中插图①所示)。在此条件下,由于飞行高度远大于地面起伏程度且观测视角接近垂直,地表场景可被简化为二维平面,实时图与基准图间的几何关系主要体现为缩放、旋转和平移,能够用相似变换或单应性变换模型进行描述[12]。在高空下视成像条件下,基于图像匹配的飞行器定位算法[4,6]已能实现较高的定位精度与较强的鲁棒性。

目前,无人飞行器应用场景正向低空、精细化方向发展,小型商业级无人机的飞行高度通常被限制在 500 m 以下,在执行目标侦察、基础设施巡检等任务时,常需进行低空倾斜观测以获取侧视信息(如图 1 中插图②所示)。在此条件下,地表场景的三维立体效应与透视形变显著,平面假设不再成立,实时图与基准图之间存在巨大视角差异。此外,低空成像还具有视场小、尺度变化快等特点,进一步增加了视觉定位的复杂度,原有针对高空下视成像的视觉定位方法已难以满足低空倾斜成像条件下的高精度视觉定位需求。近年来已有不少学者针对低空观测条件下的无人飞行器视觉定位问题提出了系列方法[12-15]和数据集[16-17],但现有技术的定位精度、鲁棒性和计算效率仍难以满足无人飞行器长航时自主飞行需求。此外,目前已有若干关于无人飞行器视觉定位的综述文献(见表 1),但它们尚未全面覆盖低空视觉定位的关键技术模块。文献[18]概述了基于异源图像匹配和惯性导航融合的飞行器定位框架所涉及的关键技术;文献[4]系统分析了 2020 年前的绝对视觉定位技术,但未讨论基于深度学习的视觉定位方法;文献[3]总结了无人飞行器跨视角地理定位任务的数据集与评估指标,文献[19]从神经网络架构的视角进行分类综述,这两篇综述均未对低空倾斜观测条件下的挑战进行深入探讨;文献[20]只讨论了视觉定位中景象匹配面临的挑战;文献[21]与文献[22]将视觉定位流程拆分为检索与匹配两个模块进行论述,但未能系统阐述"检索—匹配—位姿解算"这一完整定位框架,以及该定位框架相较于端到端定位等其他框架的优势。针对低空观测条件下的显著视角差异与地物立体效应,相比于存在误差累积


图 1 不同观测条件下的绝对视觉定位对比

Fig. 1 Comparasion of absolute visual localization under different observation conditions

表 1　现有无人飞行器视觉定位综述对比

Tab. 1　Review of existing visual localization methods for unmanned aerial vehicles

| 综述 | 年份 | 主要关注点 | 主要贡献 |
|---|---|---|---|
| 文献[18] | 2020 | 异源图像匹配辅助惯性导航的飞行器定位 | 从相机 – 惯导标定技术、异源图像匹配、姿态解算、数据融合和后端优化等方面阐述了异源图像匹配辅助惯性导航的飞行器定位关键技术 |
| 文献[4] | 2021 | 无人飞行器绝对视觉定位 | 综述了 2020 年之前的无人飞行器绝对视觉定位技术,涵盖模板匹配、特征匹配、深度学习和视觉里程计等方面 |
| 文献[3] | 2024 | 基于图像检索的无人飞行器跨视角地理定位 | 综述了针对飞行器跨视角地理定位任务设计的数据集、指标及方法 |
| 文献[19] | 2024 | 基于深度学习的无人飞行器视觉定位 | 综述了基于不同神经网络架构的无人飞行器视觉定位方法,总结了深度学习方法面临的挑战和未来研究方向 |
| 文献[20] | 2025 | 基于景象匹配(图像检索)的无人飞行器地理定位 | 综述了基于模板匹配、特征匹配以及场景语义学习的景象匹配视觉定位方法及评价指标,总结了基于视觉的组合定位算法,探讨了当前景象匹配方法面临的挑战并提出了多点解决思路 |
| 文献[21] | 2025 | 基于深度学习的无人飞行器绝对视觉定位 | 综述了基于深度学习的无人飞行器绝对定位方法和数据集,总结了图像级检索和像素级匹配方法的优缺点 |
| 文献[22] | 2025 | 基于图像检索和匹配的无人飞行器视觉定位以及基于避障和路径规划的视觉导航 | 将无人飞行器视觉定位拆分为检索和匹配模块,分别综述了领域内的主要方法、数据集、评价指标以及基准测试结果;综述了视觉导航中的避障和路径规划方法;总结了领域面临的主要难点 |
| 本文 | 2026 | 低空观测条件下基于“检索—匹配—位姿解算”框架的无人飞行器绝对视觉定位 | 系统分析了基于“检索—匹配—位姿解算”的视觉定位框架在低空倾斜观测条件下的必要性;以低空倾斜观测条件下视觉定位的难点为导向,综述了现有方法的思路、效果及缺陷,深入讨论现有测试基准存在的突出问题,为后续研究提供参考 |

问题的相对视觉定位框架[4]、泛化能力受限的端到端直接定位框架[2]、依赖特定场景训练的无地图定位框架[23],以及对视角和尺度变化敏感的传统模板匹配技术[18],“检索—匹配—位姿解算”框架通过“分层解耦、由粗到精”的策略,能够有效兼顾搜索效率、定位精度与场景泛化性,是低空长航时绝对定位的稳健路径。

与已有综述不同,本文聚焦于低空观测条件下的飞行器绝对视觉定位技术。需要说明的是,高空下视观测可视为低空观测的特殊形式,本文讨论的方法也同样适用。本文的主要贡献如下:

1)阐述了面向低空倾斜观测的飞行器绝对视觉定位框架。针对显著视角差异与地物立体效应带来的问题,详述了“检索—匹配—位姿解算”的由粗到精定位范式,论证了该框架在大范围初值获取与高精度位姿解算等方面的优势。

2)梳理了关键技术演进脉络与前沿研究现状。围绕图像级检索、像素级匹配及位姿解算三大模块,回顾了技术变革历程,重点分析了视觉基础模型、自监督学习等前沿技术在提升模型泛化能力与鲁棒性方面的最新进展。

3)剖析了工程化部署面临的挑战与未来突破方向。结合主流边缘计算平台的算力特性,揭示了现有算法在实时推理效率、系统鲁棒性以及硬件成本和生态依赖等方面存在的局限,提出了系列未来的重点研究方向。

## 1　基于“检索—匹配—位姿解算”的飞行器绝对视觉定位

图像级检索、像素级匹配以及基于匹配点对的位姿解算是实现飞行器在低空观测条件下高精度定位的三个关键技术模块(如图 2 所示),三者构成了本文讨论的低空飞行器绝对视觉定位框架,本章将围绕这一框架下的主要问题展开分析。

### 1.1　三个模块在视觉定位框架中的必要性

在卫星信号失效和惯导误差累积下,无人飞行器会偏离预定航线,绝对视觉定位技术通过建

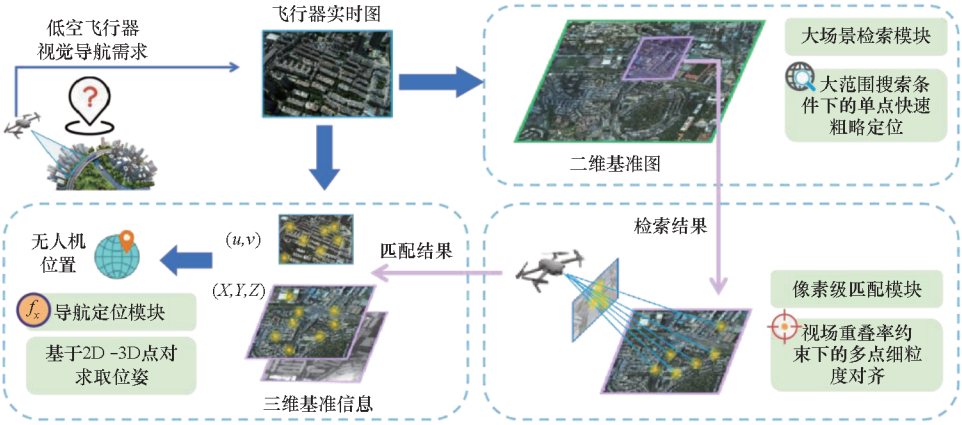


图 2　基于“检索—匹配—位姿解算”的无人飞行器绝对视觉定位框架

Fig. 2　UAV absolute visual localization framework based on " retrieval – matching – pose estimation"

立飞行器成像的实时图与带有地理编码的基准图之间的几何映射关系，解算飞行器的绝对位置。需要说明的是，当机载相机与飞行器机体的相对安装关系确定时，已知相机绝对位姿即可转换出机体绝对位姿，为论述方便，本文将相机位姿和机体位姿等同。

### 1.1.1　图像级检索模块

图像级检索可实现无人飞行器的快速初始视觉定位（见图 2），该技术一般基于图像全局特征的相似性度量[24]，其数学定义为：将基准图 $I_{\text{ref}}$ 划分为若干图像块，构成候选集 $G=\{I_k \mid k=1,\cdots,N\}$，通过特征提取函数 $f(\cdot)$ 将查询图像 $I_{\text{q}}$（即飞行器实时图 $I_{\text{cam}}$）与每个候选块映射为特征向量，并基于相似度函数 $s(\cdot,\cdot)$ 进行比对，寻找出数量为 $k$ 的最优候选子集 $C^*$。

$$C^*=\underset{C\subset G,\mid C\mid=k}{\operatorname{argmax}}\sum_{I_j\in C}s(f(I_{\text{q}}),f(I_j)) \tag{1}$$

图像级检索的搜索范围主要由飞行器当前位置的不确定度决定[25]。该不确定度通常取决于上一时刻主导航系统（如 GNSS/INS）的精度以及从该系统最后一次有效更新到视觉定位启动所经历的时间。位置不确定度越大，图像级检索所需的基准图地理范围也相应越大。

在低空观测条件下，机载相机视场范围（记为 $S_{\text{cam}}$）一般较小且随高度、角度调整而迅速变化，此时待搜索区域 $I_{\text{ref}}$ 的地理范围 $S_{\text{ref}}$ 可能远大于实时图 $I_{\text{cam}}$ 的视场范围 $S_{\text{cam}}$。假设飞行器高度为 $H$，相机俯仰角为 $\theta$，视场角为 $\varphi$，图像尺寸为 $w\times h$ 像素，且滚转角为零，则 $S_{\text{cam}}$ 可以依据式（2）进行粗略估计。以一典型场景为例：设 $H=100$ m，$\theta=30°$，$\varphi=70°$，$w\times h=4\,000\times3\,000$，$S_{\text{ref}}=1\,000\text{ m}\times1\,000\text{ m}$，此时 $S_{\text{ref}}$ 与 $S_{\text{cam}}$ 的面积比值高达 106。

$$S_{\text{cam}}=\left[\frac{H}{\sin(\theta)}\cdot\frac{\tan\left(\frac{\varphi}{2}\right)}{\sqrt{w^2+h^2}}\right]^2\cdot w\cdot h \tag{2}$$

当 $S_{\text{ref}}\gg S_{\text{cam}}$ 时，直接对 $I_{\text{cam}}$ 与 $I_{\text{ref}}$ 进行像素级匹配难以满足实时性要求[14,26]，且现有跨视角匹配算法无法实现鲁棒配准[27]。因此，为在大范围场景下快速定位 $I_{\text{cam}}$ 在 $I_{\text{ref}}$ 中的位置，一般会采用实时性能更好的图像检索技术。图像级检索将图像抽象为全局特征向量进行比对，避免了密集的像素级运算，因而在计算效率上具有显著优势[24,27]。此外，与侧重于局部细节纹理的像素级特征相比，图像级检索所依赖的全局特征能够捕捉场景的整体视觉结构和高级语义信息[28]，有助于在 $S_{\text{ref}}\gg S_{\text{cam}}$ 的情况下有效区分真值区域与冗余区域的干扰。

### 1.1.2　像素级匹配模块

图像级检索可实现飞行器的视觉快速定位，但其精度难以满足低空飞行器的高精度位姿解算需求[24]。此外，图像级检索仅能提供实时图与基准图的单点对应关系。在低空倾斜观测时，若无飞行高度及相机姿态先验，单点对应关系不足以解算飞行器六自由度位姿。因此，在图像级检索确定候选区域后，需进一步采用像素级匹配技术，实现实时图与基准图间的精细对齐，获取用于高精度位姿解算的同名点。

像素级匹配的数学定义为：给定实时图 $I_{\text{cam}}$ 和基准图 $I_{\text{ref}}$，寻找一个映射函数 $\mathcal{T}$，使得对于 $I_{\text{cam}}$ 中的像素点 $p$，能找到其在 $I_{\text{ref}}$ 中的同名点 $p'=\mathcal{T}(p)$，并满足特定的几何约束。在飞行器视觉导航这类实时性要求比较高的场景，一般会先采取“由粗到精”的定位策略。检索模块负责解决大范围搜索条件下的飞行器粗定位，定位精度一般在数十米量级；匹配模块则基于检索结果完成精

定位,满足米级的定位需求。

### 1.1.3　位姿解算模块

在像素级匹配建立了实时图和基准图的同名点对应关系后,即可采用透视 $n$ 点算法(perspective-$n$-point,PnP)求解飞行器在世界坐标系下的绝对位姿。PnP 问题的数学定义如下:给定一组三维世界坐标系中的点 $\{\boldsymbol{P}_i \in \mathbb{R}^3\}_{i=1}^n$ 及其在相机归一化成像平面对应的二维投影点 $\{\boldsymbol{p}_i \in \mathbb{R}^2\}_{i=1}^n$,目标是求解相机相对于世界坐标系的位姿,即旋转矩阵 $\boldsymbol{R} \in SO(3)$ 和平移向量 $\boldsymbol{t} \in \mathbb{R}^3$,使式(3)所示投影关系成立。

$$\boldsymbol{p}_i = \pi(\boldsymbol{R}\boldsymbol{P}_i + \boldsymbol{t}), i = 1, \cdots, n \tag{3}$$

其中,$\pi$ 表示相机的投影模型。对于针孔相机模型,该投影关系可具体表示为:

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \boldsymbol{K}[\boldsymbol{R} \quad \boldsymbol{t}] \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{4}$$

式中,$(u_i, v_i)$ 为像点 $\boldsymbol{p}_i$ 的像素坐标,$(X_i, Y_i, Z_i)$ 为世界点 $\boldsymbol{P}_i$ 的三维坐标,$\boldsymbol{K}$ 为相机内参矩阵,$\lambda_i$ 为深度尺度因子。基准图三维世界坐标的经纬坐标一般由基准图自带的地理编码信息提供,而高程信息一般可从数字表面模型(digital surface model,DSM)获取。

在高空正下视观测的视觉定位任务中,由于地面起伏远小于飞行高度,且图像主要捕获地物顶面信息,因此整个场景可被近似为同一高程的平面。在此假设下,即使不引入精确的高程数据,也可通过平面单应性变换等方式解算飞行器的二维位置[6](如图 1 所示)。然而,在低空观测条件下,实时图视场范围内的地形与地物高差显著,匹配点对可能分布于不同高度的表面(如建筑屋顶、立面与地面),此时"将地面场景近似为二维平面"的假设不再成立,高精度位姿解算需要引入 DSM 数据辅助。

## 1.2　与不同视觉定位框架的区别和联系

除本文所关注的基于"检索—匹配—位姿解算"的绝对视觉定位框架之外,相对视觉定位、端到端直接定位、无地图定位以及模板匹配—位置解算等框架也各有其特定的技术路线与适用领域(如图 3 所示)。

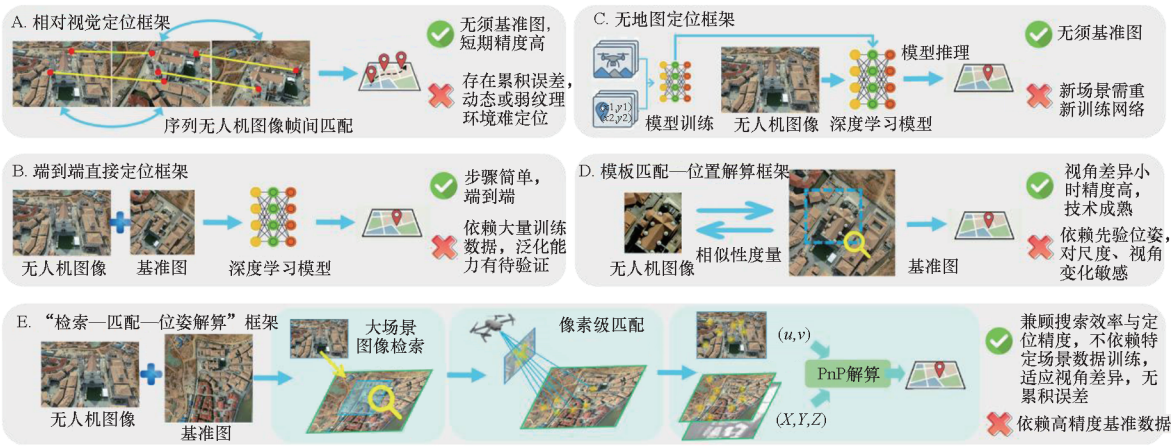


图 3　不同视觉定位框架对比

Fig. 3　Comparison of different visual localization frameworks

相对视觉定位框架通过对序列图像进行帧间匹配来解算图像间相对位姿,该方法无须基准图且短时间段内精度较高,常用于室内、洞穴等封闭或结构化场景的视觉里程计和同步定位与建图[29-30](simultaneous localization and mapping,SLAM)中,但其固有的累积误差使其难以胜任长航时、大范围的室外绝对定位任务,且在动态变化大或弱纹理环境下易失效。端到端直接定位框架[31-33]利用深度学习模型实现图像到位姿的直接映射,避免了复杂的匹配和位姿解算环节。然而,该类方法需要大量高质量、多样化的"图像-位姿"训练数据,模型的域外泛化能力面临挑战,其在无人机视觉定位任务上的应用暂处于前沿探索阶段。无地图定位框架[23,34]利用隐式神经表示替代显式地图存储,提升了存储效率。但其存在显著的场景依赖性,需针对每一个特定任务区域采集数据重新训练网络,这一特点使其适用于室内、园区等范围有限且结构相对稳定的区域而难以直接迁移至未知环境。模板匹配—位置解算框架[35]首先将无人机图像和基准图校正到相同尺度和视角,然后进行相似性度量。该方法在高空正下视条件下表现良好且技术成熟[20],但其无法克服低空观测下的巨大视角与尺度变化,且对先验位姿精度要求相比其他方法更高。

本文讨论的基于“检索—匹配—位姿解算”的框架[36]采用“由粗到精”的策略,在保证大范围搜索效率的同时兼顾定位精度,对视角与尺度变化具有良好的适应性。该框架具有多方面的比较优势:其绝对定位模式避免了累积误差;不依赖针对特定场景的训练数据,泛化能力强;对先验位姿的要求较低,能适应低空飞行中剧烈的视角与尺度变化。此外,从工程实践角度看,其模块化的设计便于独立优化与系统集成。这些综合优势为低空飞行器视觉定位提供了更为可行且稳健的技术路径。

## 2 跨视角图像检索技术

传统的图像粗定位方法主要包括基于模板匹配的景像匹配方法与基于局部特征聚合的检索方法。鉴于模板匹配本质上也依赖于图像全局特征的相似性度量,本文将其一并归入图像级检索范畴进行讨论。近年来,基于深度学习的无人机-卫星跨视角图像检索(又称跨视角地理定位)技术逐渐成为主流[37-38]。

### 2.1 基于人工设计特征的跨视角图像检索

#### 2.1.1 基于模板匹配的景象匹配方法

基于区域的匹配方法又被称为模板匹配,通过在卫星参考底图上滑动飞行器实时图像(模板),计算重叠区域的相似度以确定最佳匹配位置[39]。该类方法的研究核心在于设计鲁棒的相似性度量准则。常用的相似性度量有绝对误差和[40](sum of absolute differences,SAD)、误差平方和[41](sum of squared differences,SSD)以及归一化互相关[42](normalized cross-correlation,NCC)等。以上方法在同源图像匹配效果较好,但是难以适应飞行器实时图与基准图之间的模态差异,因此,一些研究设计了基于相位一致性[43-44]或图像局部自相似描述符[45-46]的模板匹配方法。上述方法常用于高空下视成像条件下的飞行器视觉定位任务,但是由于模板匹配方法需要依赖惯导提供的姿态和高度先验信息来将飞行器实时图校正到与基准图相似的尺度和角度,针对低空倾斜观测的场景,这类方法无法适应显著视角差异带来的透视畸变与复杂的几何形变,因而匹配性能急剧下降。

#### 2.1.2 基于局部特征聚合的检索方法

为了解决全局模板匹配对几何形变敏感的问题,研究者们提出了基于局部特征聚合的检索方法。该类方法提取图像的局部特征,并通过聚类或编码技术将其聚合为紧凑的全局特征描述符,进而通过特征向量的相似性度量实现图像检索。其中,视觉词袋(bag of words, BoW)模型[47]是最具代表性的方法之一。该方法通过聚类将局部特征量化为“视觉单词”,并将图像表示为视觉单词的直方图向量。BoW 利用了局部特征的旋转和尺度不变性,在一定程度上解决了视角变化带来的匹配难题,在宽基线检索匹配中得到了广泛应用[48-49],能够应用于无人飞行器低空观测条件下的图像检索。然而,该方法强制归类(硬量化)的过程会导致大量特征细节丢失。因此,Jégou 等[50]提出了局部聚合描述符向量(vector of locally aggregated descriptors,VLAD),其不再仅仅统计视觉单词的出现频率,而是累加了局部特征与聚类中心之间的残差向量,生成了一种既紧凑又保留了大量细节的图像表征。VLAD 能够更好地捕捉纹理细节,且计算效率优于传统的 BoW。

然而,当成像视角差异逐步加大时,底层特征的共视性急剧下降,导致检索性能显著衰退[51]。Lin 等[52]较早地尝试利用手工特征(如 HOG、颜色直方图等)结合典型相关分析来学习多视角图像间的特征转换关系。Castaldo 等[53]提出了一种基于语义布局的跨视角图像匹配方法,通过匹配图像中的几何结构布局来克服视角差异。然而,在面对大规模检索候选库时,该方法的 TOP1 召回率(Recall@1)不足 30%。已有研究表明基于手工特征或浅层统计模型的传统方法还无法适用低空视觉定位这类具有显著视角差异的应用场景。

### 2.2 基于深度学习的跨视角图像检索

2020 年,Zheng 等[16]发布首个包含无人机视角的跨视角地理定位数据集,正式提出了无人机跨视角定理定位任务。该工作基于 ResNet-50 构建了包含无人机分支的孪生网络基线算法,填补了低空飞行器视角数据缺失的空白,同时为深度学习在无人机跨视角地理定位中的应用奠定了基础。此后,学者针对这一任务展开了一系列研究,主要集中在鲁棒特征提取网络的设计以及网络模型训练策略的优化两个方面。

#### 2.2.1 特征提取网络的设计

在特征提取网络的设计中,特征图的重组与对齐是主要研究对象。为了让模型能够从不同视角的图像中提取出具备对应关系的局部特征,研究者们提出了多种特征重组策略。早期的 LPN 方法[54]设计了图像方环形切分策略,通过提取环状区域特征来适应图像间的旋转变化。然而,该

方法主要基于立体目标主体(例如建筑)在飞行器实时图中心的假设,其环状区域特征表述无法适应目标主体偏移实时图中心这一更普遍的应用场景。针对这一局限,SDPL[55] 设计了图像密集切分和偏移融合策略,提升了模型适应目标偏移中心的能力;MCFA[56] 通过多尺度级联模块动态调整中心区域,增加了对中心区域位置变化适应性的同时,提高了对尺度变化的适应性。

在目标位置偏移以外,特征重组也需要考虑图像切分导致的上下文割裂问题,研究者展开了对特征图柔性分割的探索。Ge 等[57] 引入了局部信息融合策略,旨在保留细粒度特征的同时强化全局语义关联。FSRA[58] 根据热力图对特征图进行均匀分割,实现了特征的语义级分割对齐。在此基础上,CAMP[59] 引入了位置感知分支,通过为特征图嵌入位置编码,进一步丰富了分割后特征的空间表达能力。

#### 2.2.2　训练策略的优化

在训练策略方面,研究者致力于设计更高效的损失函数与学习范式,以提升模型对特征的判别能力及其对真实复杂应用场景的适应能力。早期的工作多采用分类思想,将相同地理场景(如包含同一建筑物)视为一个类别进行训练,并在推理时使用分类层前的特征进行检索[16,54]。随着任务从离散场景的检索定位发展至连续场景,简单地将相似场景视为一类的思路无法有效区分连续空间下的近邻图像,这使得对比学习逐渐成为主流,经典的三元组损失和交叉熵损失被广泛应用于多任务驱动的训练中[58,60]。为了解决三元组损失挖掘难样本的能力有限的问题,Sample4Geo[61] 首次在该任务中使用结合对比的 InfoNCE 损失函数,并约束批次内相同场景的出现次数,显著提升了模型性能。对称 InfoNCE 损失成为多任务学习中对比学习方面的常用损失[59,62-63]。在此基础上,DAC[62] 引入跨批次场景一致性策略,进一步挖掘跨批次正样本信息,以实现特征空间中更紧凑的类内聚类。CAMP 通过增加同一批次相同成像平台样本间的对比,扩大了负样本范围。QDFL[64] 则通过多相似度损失与 JS 散度损失,从分布层面促使不同视角的特征分布趋于一致。

面对真实场景中标注数据匮乏及环境多变的挑战,训练策略逐渐向降低数据依赖的自监督方向演进。Li 等[65] 利用冻结的基础模型强大的泛化能力,设计了一种自监督适配框架。该方法通过期望最大化算法迭代估计跨视角匹配关系以优化适配器,并结合信息一致性模块防止特征在适配过程中退化,实现了从通用模型向跨视角任务的高效迁移。DMNL[66] 则提出了一种端到端的自监督学习框架,利用聚类算法生成伪标签,并设计了动态分层记忆模块与信息一致性演化机制,分别用于增强特征的判别能力与挖掘跨视角的潜在关联,在无标签条件下实现了特征的有效对齐。这标志着该领域正从依赖强监督的度量学习,向利用海量无标签数据进行自监督特征挖掘的方向转变。

相较于传统方法,基于深度学习的无人飞行器跨视角检索定位技术展现出了更强的视角差异适应能力,为后续像素级匹配提供了更精确的粗定位结果。

## 3　跨视角图像匹配技术

在经典的图像匹配论文中,跨视角图像匹配被归纳到广义的宽基线(wide baselines)图像匹配的概念范畴中进行讨论[67-68]。基线在立体视觉中表征两个摄像机光心之间的距离,一般而言,基线越宽,图像间的视角差异越大,场景重叠区域越小。飞行器视觉定位中的跨视角匹配在概念上与宽基线匹配相似,但其面临的问题更为复杂:实时图来自机载相机采集的倾斜视角图像,而基准图来源于卫星遥感或航空摄影测量生成的正射投影产品。成像模型的差异导致了几何特性的不同,实时图采用中心投影的小孔成像模型,遵循透视变换规律,存在近大远小现象;基准图基于平行投影模型,保持统一比例尺,不存在透视变形。同时,两类图像在空间分辨率上数量级差异进一步增大了跨视角图像间的尺度变化。上述差异使得飞行器视觉定位中的跨视角匹配相比传统宽基线立体匹配难度更大。

由于跨视角图像匹配技术不仅在飞行器视觉定位任务中发挥作用,还在三维重建、虚拟现实以及室内和街道场景的视觉定位任务中扮演重要角色[68-69],这些应用场景同样面临与低空倾斜观测类似的视角差异、立体效应、低重叠率等问题,因此本文在归纳跨视角图像匹配技术时不局限于飞行器视觉定位场景,而是同时关注通用的跨视角图像匹配技术。

### 3.1　基于人工设计特征的跨视角图像匹配

人工设计特征的跨视角图像匹配遵循特征检测—描述符构建—特征匹配的经典流程,首先利用手工设计的特征检测器在图像中定位关键点(或线段),然后为每个关键点(或线段)构建具有

判别力的描述符,最后基于描述符相似性建立待匹配图像间的对应关系。

尺度不变特征变换(scale-invariant feature transform, SIFT)算法[70]是跨视角匹配的经典工作,它通过高斯差分(difference of Gaussian, DoG)金字塔检测尺度不变的关键点,并生成基于梯度方向直方图的128维描述符,对旋转、尺度和亮度变化具有一定鲁棒性。加速稳健特征(speeded up robust features, SURF)算法[71]在SIFT基础上采用Hessian矩阵检测特征点,并使用积分图像和Haar小波响应生成描述符,在保持性能的同时显著提高了计算效率。定向快速特征点与旋转不变特征(oriented FAST and rotated BRIEF, ORB)算法[72]结合FAST角点检测与BRIEF二进制描述符,进一步优化了实时性能,成为移动端和嵌入式设备的常用方案[29]。上述方法可以适应一定程度的视角差异,但在应对由大视角变化引起的仿射或透视形变时存在明显局限性。

为应对跨视角图像间的复杂几何变换,研究者提出了多种改进策略。Yu等[51]提出的仿射SIFT(affine-SIFT, ASIFT)算法首先通过密集采样仿射参数(倾斜参数和经度角)对原始图像进行模拟,生成一系列覆盖仿射形变空间的模拟图像,然后使用SIFT算法对所有生成的模拟图像对进行特征提取与匹配,并通过筛选最优匹配对来获取可靠的匹配点。该策略通过在前端模拟仿射形变来有效缩小原始跨视角图像间的表观差异,进而提升大视角变化下的匹配鲁棒性,但密集的视角模拟与匹配也增加了计算成本。Wang等[13]在ASIFT的基础上设计了一种针对飞行器图像的跨视角匹配方法,首先利用ASIFT提取初始特征点,再通过自适应归一化互相关算法修正特征点位置并估计局部变换模型,最后结合加权最小二乘匹配优化配准结果,该方法虽可以提升配准精度但流程更为复杂。虽然ASIFT类方法可以应对跨视角观测带来的仿射变换,但是在相机焦距较短或场景深度变化显著的条件下(如低空飞行器实时图),仿射模型难以准确反映真实成像过程。因此Cai等[73]提出了透视SIFT(perspective-SIFT)方法,采用透视变换代替仿射变换来模拟相机视角变化,再使用SIFT进行匹配,该方法在高透视畸变情况下的鲁棒性更强。此外,为了提高此类基于几何变换模拟的匹配方法的计算效率,研究人员从不同方面进行了改进。例如,Yu等[74]提出了迭代SIFT(iterative-SIFT)方法,通过迭代估计图像间的相对视角和光照变换参数,逐步优化匹配结果,在保证精度的同时将计算复杂度控制在较低水平;Song等[12]则提出一种基于迭代仿真与单应性矩阵评估的优化方法,通过动态停止仿真过程(当匹配正确点数不再增加时终止)和几何约束验证单应性矩阵的合理性,在保持匹配精度的同时进一步提升了计算效率。

除了对特征提取与匹配算法本身的优化,还有一些研究充分利用成像平台的位姿先验信息,通过引入空间几何约束来提升匹配的稳健性。例如,Hu等[75]在标准特征匹配流程中引入了三种空间关系约束——循环角序约束、局部位置一致性约束和邻域保持约束来提升跨视角图像匹配鲁棒性;Roth等[76]则通过投影变换合成虚拟视图,结合本质矩阵估计对跨视角匹配结果进行几何验证;Zhang等[77]通过相机位姿计算极线约束来过滤误匹配点,提升了倾斜视角图像的匹配速度与正确率。此类方法虽然能够有效利用初始位姿信息,但其性能在很大程度上依赖于先验位姿的准确性。此外,图像间局部几何形变(如建筑物立面与屋顶的透视差异)往往难以通过单一的全局变换模型进行准确描述[78],进一步限制了此类方法在复杂场景下的适用性。

### 3.2 基于深度学习的跨视角图像匹配

近年来,基于深度学习的图像匹配方法取得了突破性进展,已成为解决跨视角图像匹配问题的主流技术路径。现有研究主要从匹配网络的设计、先验知识的融合与利用以及训练策略的优化等几个方面切入。

#### 3.2.1 匹配网络的设计

根据模型结构的差异,现有匹配网络可分为稀疏匹配网络、半稠密匹配网络和稠密匹配网络三大类,它们在匹配点密度、计算效率和适用场景上各有特点。

(1)稀疏匹配网络

稀疏匹配方法延续了传统的“检测—描述—匹配”流程,但采用神经网络进行特征提取与匹配。这类方法的核心特点是匹配点数量较少(通常数百到数千个)、计算效率高,适合实时性要求高的应用场景。SuperPoint[79]采用自监督方式联合训练关键点检测与描述符提取网络,通过单应性变换生成训练数据,在特征均匀性和重复性上显著优于手工特征。D2-Net[80]采用单一卷积神经网络同时完成特征检测和描述符构建,改变了传统先检测后描述的处理模式。R2D2[81]引入可靠性和重复性预测机制,进一步提升了匹配质量。

RDD[82]采用可变形 Transformer 的双分支架构,分别处理关键点检测和描述符提取,通过可变形注意力机制捕获几何不变性。在特征匹配阶段,图神经网络得到广泛应用。SuperGlue[83]利用注意力机制进行图匹配;LightGlue[84]通过自适应剪枝优化计算效率,显著提升了匹配速度。

(2)半稠密匹配网络

半稠密匹配方法采用“检测自由”的策略,直接在特征图上进行密集相关性计算,然后提取高置信度匹配。这类方法产生中等密度的匹配(通常数千个点),在保持较高效率的同时可改善弱纹理区域的匹配能力。LoFTR[85]摒弃显式检测器,直接在粗粒度特征图上使用 Transformer 进行全局感受野的特征聚合,通过粗到细架构提升定位精度。MatchFormer[86]通过交错自注意力和交叉注意力增强特征表示。ASpanFormer[87]引入自适应跨度机制动态调整注意力范围以处理尺度变化。RCM[88]引入了动态视角切换机制,自动选择尺度较大的图像作为稀疏分支进行处理,有效增加了可匹配点数量,同时采用多对一匹配策略解决尺度变化下的匹配冲突问题。

(3)稠密匹配网络

稠密匹配方法致力于建立图像间所有像素的完整对应关系,生成稠密流场。这类方法提供最全面的场景对应信息,但计算成本最高,适用于需要密集几何信息的应用场景。PDC-Net[89]提出一种概率性稠密匹配框架,通过约束混合模型建模预测分布,并结合扩散模型引导的跨图像交互提示模块,显著提升了对未见过域数据的泛化能力。DKM[90]提出基于高斯过程的核化回归匹配器,通过堆叠特征图和大核深度卷积进行形变细化。RoMa[91]在 DKM 的基础上使用视觉基础模型 DINO-v2 进行特征提取,提高了模型对特征的鲁棒表达能力,在极端视角变化下仍能保持稳定的匹配性能。

为量化分析不同匹配网络的性能,本文引用 Ye 等[27]在 AnyVisLoc 基准上的测试结果(见表 2)。该实验在 NVIDIA 3090 GPU 上统一进行,以定位精度($A@T$ m)和单帧匹配时间为评估指标。结果表明,三类网络在效率与精度间存在权衡:稀疏匹配网络(如 SuperPoint + LightGlue)速度最快(75 ms/帧)但匹配精度较低;半稠密匹配网络(如 LoFTR)的匹配精度有所提高但计算开销进一步增大;稠密匹配网络(如 RoMa)的匹配用时约是 SuperPoint + LightGlue 方法的 9 倍。工程应用中需要根据实时性要求和资源约束进行网络选型。

**表 2　不同匹配网络的精度和计算效率对比[27]**

Tab. 2　Comparison of accuracy and computational efficiency among different matching networks[27]

| 匹配网络 | 网络类型 | 匹配性能($A@5$ m)/% | 计算效率/(ms/帧) |
|---|---|---|---|
| SuperPoint + SuperGlue | 稀疏 | 52.1 | 92 |
| SuperPoint + LightGlue | 稀疏 | 55.8 | 75 |
| LoFTR | 半稠密 | 59.5 | 165 |
| DKM | 稠密 | 65.6 | 4 915 |
| RoMa | 稠密 | 70.1 | 659 |

### 3.2.2　先验知识的融合与利用

针对跨视角图像在几何与语义上存在的巨大差异,研究者的一个重要思路是将来自其他视觉任务的“先验知识”融合到匹配网络的设计与推理过程中,以引导网络克服视角、尺度与外观变化带来的困难。这主要包括两条技术路径。

(1)基于视觉基础模型的图像匹配网络

视觉基础模型(vision foundation models, VFM)因在海量数据上预训练而具备强大的泛化与语义理解能力,近年来被广泛用于增强跨视角匹配的鲁棒性。Cadar 等[92]提出了一种基于语义提示的匹配增强方法,该方法利用 DINO-v2 提取高层语义信息,并以此指导局部特征的匹配过程,通过语义一致性约束有效过滤错误匹配。OmniGlue[93]将 DINO-v2 直接集成到可学习的特征匹配网络中,使其泛化性强的特征能直接引导特征传播与聚合,从而在多个挑战性数据集上表现优异。SemaGlue[94]则进一步提出“感知式匹配”框架,将 SegNext 网络提取的细粒度语义信息融入匹配流程,通过动态语义聚合模块实现语义感知与几何匹配的底层统一。还有一些研究通过知识蒸馏技术对 VFM 进行轻量化,如 Yang 等[95]提出的 DistillMatch 框架,旨在将视觉基础模型的强大语义理解能力蒸馏至一个更轻量的专用匹配网络中,在保留泛化能力的同时提升计算效率。

(2)基于语义-深度先验的图像匹配网络

除了视觉基础模型,研究者也利用语义分割、单目深度估计等任务的模型输出作为先验,来克服跨视角匹配中的几何与外观差异。例如:Zhang 等[96]提出由面到点(region-to-point)的匹配框架,首先使用图像分割模型进行语义分割,然后进行

区域级匹配,最后进行像素级配准,利用高级语义信息为特征点匹配提供先验,有效缓解低纹理和重复纹理场景的匹配困难问题;Toft 等[97]通过单目深度估计将图像划分为多个平面,并分别进行基于单应性变换的匹配;Karpur 等[69]把归一化目标三维坐标作为三维信息融入特征以提升跨视角匹配网络的效果;Wang 等[98]将基于深度信息的曲率特征融入特征描述子;Huang 等[99]利用深度数据将图像分为不同小块并分别计算单应性变换矩阵以应对立体目标不满足平面单应性假设的问题;LiftFeat[100]利用单目深度估计网络获取图像法线,然后将法线方向特征显式融合到局部特征描述符中,提升了模型在弱纹理区域和光照极端变化区域的匹配效果。

3.2.3 训练策略的优化

在网络训练方面,研究者探索了多种训练策略。SuperPoint 采用大量模拟数据进行预训练后使用真实数据微调,使网络获得优异的泛化性能。许多研究在训练过程中通过对单帧图像施加已知几何变换来进行数据增强[79,84,101],此类方法可以生成平面场景的多视角仿真图像进而扩增数据,但无法模拟立体目标在不同视角下的真实成像差异。Shen 等[102]提出了基于视频帧传递的匹配真值构建方法,通过相邻帧间几何变换传递建立长时程匹配,实验表明模型匹配性能随着训练视频时长的增长而有效提升。该方法虽能扩展数据规模,但其输入限于同一视频序列,在飞行器视觉定位等跨模态图像场景中适用性受限。针对跨模态匹配问题,有研究采用风格迁移技术构建模态各异但几何对齐的图像对[103-104],然而,此类方法同样难以准确模拟真实空间视角变化带来的复杂几何形变。除数据增强策略外,无监督与弱监督学习也是重要方向。例如,DISK[105]采用了基于强化学习的训练范式,通过策略梯度优化关键点检测和描述,并利用从多视角图像中重建的深度信息作为奖励信号,引导网络学习更具辨别力的特征,而无须严格的像素级匹配真值。RIPE[106]针对特征点提取网络设计了一种基于强化学习的弱监督训练框架,仅需图像对是否描述同一场景的二元标签即可训练关键点检测器,大幅扩展了可用训练数据范围。

相较于人工设计的匹配方法,基于深度学习的跨视角图像匹配网络对显著视角差异具有更强鲁棒性,可为后续位姿解算提供更准确的匹配点对。未来研究可着力于引入视觉基础模型的强大语义先验与自监督学习范式,以提升模型的泛化能力;同时,需在保证匹配精度的前提下,设计更为轻量化的网络架构,以适应机载平台的算力约束。

## 4 飞行器位姿解算方法

在利用跨视角图像匹配模型获取飞行器实时图与基准图的匹配点对关系后,需要利用 PnP 方法求解飞行器位姿。在低空倾斜观测条件下,跨视角匹配算法产生的匹配点对往往存在空间分布不均匀、误匹配率较高、匹配点数量波动大等问题。这种匹配点的不确定性给后续的位姿解算带来困难。

### 4.1 经典的 PnP 方法

经典的 PnP 方法主要包括基于最小点集的 P3P 方法[107]、线性求解方法如直接线性变换(direct linear transformation, DLT),以及各种非线性优化方法。P3P 作为最小子集求解方案,通过 3 个 2D-3D 点对应关系构建三角形几何约束,利用余弦定理建立方程组,最多可产生 4 个可能解,需通过额外点进行解歧义,其计算效率高但噪声敏感度较大。EPnP 方法[108]通过引入 4 个虚拟控制点来表示所有 3D 点,将问题转化为线性求解,具有 $O(n)$ 的时间复杂度,适用于大规模点集。RPnP[109]采用旋转轴和角度分离估计的策略,在保持线性复杂度的同时提高了准确性。DLS 方法[110]将 PnP 问题形式化为多项式系统,通过矩阵特征值分解直接求解,避免了迭代优化可能陷入局部最优的问题。OPnP 方法[111]采用二次约束二次规划形式,通过半定规划松弛技术逼近全局最优解。MLPnP 方法[112]进一步考虑了观测数据的不确定性,通过最大似然估计提高在噪声环境下的估计精度。以上方法在精度、效率和稳定性方面各有侧重,为视觉定位提供了多样化的技术路线选择。

### 4.2 适用于飞行器定位的 PnP 方法

在飞行器绝对视觉定位任务中,解算稳定性受匹配点空间分布的影响显著。需要说明的是,匹配点若集中于近似平面(如高空正下视场景)可能导致 PnP 位姿解算的退化或不稳定[108],但在低空观测条件下,匹配点对通常分布于具有显著高差的建筑屋顶及地面等多维表面上,平面假设不再成立。这种丰富的三维空间分布特性虽然增加了同名点对齐的难度,但其提供的空间几何约束有助于克服纯平面假设下的解算歧义性,为高精度位姿解算提供了鲁棒性基础。基于此背

景,一些工作采用 DLT 和最小二乘迭代优化的策略来求解飞行器位置[14-15,113]。为了解决初始位置不准确及空间后方交会迭代收敛困难问题,Ye 等[15]利用了相机光心、飞行器实时图上二维点和空间三维点之间的角度相似性约束来建立三角锥几何模型,进而通过逐次修正位移差值来提升定位精度。针对许多 PnP 方法将观测噪声建模为各向同性高斯分布而实际观测数据中存在各向异性噪声的问题,Zhan 等[114]提出了广义最大似然 PnP (generalized maximum likelihood PnP, GMLPnP)算法。该算法通过迭代广义最小二乘过程,同时解算位姿和观测噪声的协方差矩阵参数,有效处理了实际数据中不同方向噪声强度不一致的问题,并在无人机视觉定位任务中验证了其对强噪声观测的鲁棒性。为了避免误匹配点对影响位姿解算精度,许多研究采用了基于 RANSAC 的 PnP 方法来获取满足一致几何约束的 2D-3D 点对[27,113,115]。

此外,一些研究利用姿态先验信息来缩小 PnP 问题的解空间。Wu 等[115]在 PnP RANSAC 过程中引入重力方向约束,计算机载相机姿态的重力方向 $\zeta_s^g$ 与假设位姿的重力方向 $\zeta_{hyp}^g$ 之间的偏差 $d_e = \arccos(\zeta_s^g \cdot \zeta_{hyp}^g)$,当该偏差小于预设阈值时提前终止 RANSAC 迭代,由此提升了位姿解算的稳定性。Li[116]利用惯导提供的俯仰角和横滚角信息,将 PnP 问题的自由度从 6 个降至 4 个,仅需 2 对匹配点即可求解偏航角和位置。这种基于惯导先验的姿态辅助 PnP 算法显著提升了在图像匹配存在较大误差时的定位鲁棒性。为了使 PnP 算法适应计算资源受限的机载嵌入式硬件平台,Jubran 等[117]提出了 Newton-PnP 算法。该算法将 PnP 问题形式化为半定规划问题,并通过求解其变量更少的对偶问题将待求变量从 241 个减少至 70 个。该算法还结合障碍函数法与牛顿法,在保证理论最优性的同时实现了对数级别的时间复杂度,最终使得算法能够在一款超小型机载计算机上实时运行。

## 5　低空视觉定位数据集

高质量的数据集是飞行器视觉定位算法发展与评估的基础,其数据样本的多样性、真值标注的精度以及任务的针对性直接影响算法的性能[102,106]。

### 5.1　跨视角图像检索数据集

表 3 总结了当前开源的无人飞行器跨视角图像检索数据集,继 University-1652 数据集发布后,本领域的公开数据集越来越丰富。这些数据集在数据来源(实景/仿真)、视角类型(正下视/多视角)、场景多样性及图像元数据丰富度等方面呈现不同的特点,反映了该领域从单一场景向复杂现实环境演进的趋势[11,16-17,118-122]。

早期的研究受限于数据获取成本,多采用仿真或有限的实景数据。University-1652[16]通过谷歌地球仿真生成了全球 72 所大学的建筑物场景,但成像时的天气、光照等环境条件较为单一。为了弥补环境多样性方面的不足,Multi-weather University[118]在此基础上引入了多种天气条件的仿真,GTA-UAV[121]则利用游戏引擎生成了包含丰富纹理和光照变化的仿真数据。然而,仅依靠仿真难以完全覆盖真实世界的复杂变化。随着无人机技术的普及,越来越多的实景数据集被提出。SUES-200[17]采集了不同高度(150 ~ 300 m)的无人机图像,重点关注飞行高度变化对检索的影响,但其场景主要局限于城郊,且视角变化相对有限。DenseUAV[119]涵盖了80 ~ 100 m 间不同高度的实拍数据且采样频率更高,UAV-VisLoc[120]包含城市、农村等 11 个不同场景,但这两个数据集仍以下视成像为主,无法满足低空倾斜观测的定位需求。针对低空倾斜观测这一更通用的场景,部分数据集开始关注多视角与复杂场景的结合,比如:UAVD4L[115]与 ComplexUAV[63]提供了多视角的实景无人机图像。最新的 AnyVisLoc[27]数据集则在多视角、多场景、多时相以及飞行高度覆盖范围(30 ~ 300 m)上进行了全面拓展,并提供了精确的无人机位置信息与 DSM。该数据集不仅支持粗粒度的图像检索,也为后续的精细化位姿解算提供了可能,是未来研究的重要基准。

### 5.2　跨视角图像匹配数据集

表 4 归纳了主流的跨视角图像匹配数据集[27,67,82,123-128],现有数据集在数据多样性和真值有效性方面做出诸多努力但仍存在一些问题。WxBS 数据集[67]通过人工标注保证真值精度,但是由于人工标注的高昂成本,该数据集仅有 37 对图像,难以支持大规模训练。为降低人工标注成本,MegaDepth[123]与 Air-to-Ground[82]等数据集利用网络开源图像和三维重建技术(如 COLMAP)生成数据。这种方法扩大了数据规模但其真值精度受三维重建质量限制。即便是 COLMAP 这类

**表 3 无人飞行器跨视角图像检索数据集对比**

Tab. 3 Comparison of cross-view image retrieval datasets for unmanned aerial vehicles

| 数据集名称 | 年份 | 无人机数据源 | 无人机类型 | 无人机位置信息 | 二维基准图 | DSM | 飞行高度/m | 视角 | 场景 | 时相 |
|---|---|---|---|---|---|---|---|---|---|---|
| University-1652[16] | 2020 | 仿真 | | × | 卫星 | × | 121.5 ~ 256 | 多视角 | 建筑 | — |
| VPAIR[11] | 2022 | 实景 | 1 | √ | 卫星 | √ | 300 ~ 400 | 正下视 | 多种 | — |
| SUES-200[17] | 2023 | 实景 | — | × | 卫星 | × | 150/200/250/300 | 多视角 | 城郊 | — |
| UAVD4L[115] | 2024 | 实景 | 1 | √ | 航空 | √ | 50 ~ 151 | 多视角 | 城郊 | — |
| Multi-weather University[118] | 2024 | 仿真 | | × | 卫星 | × | 121.5 ~ 256 | 多视角 | 建筑 | 多种 |
| DenseUAV[119] | 2024 | 实景 | 1 | √ | 卫星 | × | 80/90/100 | 正下视 | 城郊 | 多种 |
| UAV-VisLoc[120] | 2024 | 实景 | — | √ | 卫星 | × | 400 ~ 2 000 | 正下视 | 多种 | 多种 |
| GTA-UAV[121] | 2025 | 仿真 | | √ | 卫星 | × | 80 ~ 650 | 近正下视 | 多种 | — |
| ComplexUAV[63] | 2025 | 实景 | — | √ | 卫星 | × | 300 ~ 800 | 多视角 | 多种 | 多种 |
| Urban-500[122] | 2025 | 实景 | — | × | 卫星 | × | — | 多视角 | 城郊 | 多种 |
| CVGL-RGBT[122] | 2025 | 实景 | 1 | × | 卫星 | × | — | 多视角 | 城郊 | 多种 |
| AnyVisLoc[27] | 2025 | 实景 | 7 | √ | 航空 & 卫星 | √ | 30 ~ 300 | 多视角 | 多种 | 多种 |

**表 4 跨视角图像匹配数据集对比**

Tab. 4 Comparison of cross-view image matching datasets

| 数据集名称 | 数据集描述 | 真值来源 | 用途 | 应用场景 |
|---|---|---|---|---|
| MegaDepth[123] | 大型多视图立体视觉数据集,包含 196 个不同场景,图像来源于互联网,具有丰富的视角和外观变化 | SfM 三维重建(COLMAP) | 训练/测试 | 其他场景 |
| HPatches[124] | 用于评估局部特征描述符和匹配算法的基准数据集,包含视角和光照变化下的图像序列 | 单应性变换提供像素级对应关系 | 训练/测试 | 其他场景 |
| KITTI[125] | 自动驾驶场景数据集,包含在车辆行驶过程中采集的街景图像序列 | 高精度 GPS/IMU 系统提供相机位姿 | 测试 | 其他场景 |
| ScanNet[126] | 大规模的室内场景数据集,包含大量通过深度传感器采集的 RGB-D 视频序列 | SfM 三维重建 | 训练/测试 | 其他场景 |
| WxBS[67] | 宽基线图像匹配数据集,包含不同模态(如可见光、红外、素描等)、不同视角下的图像对 | 人工配准 | 测试 | 其他场景 |
| GL3D[127] | 图像数据主要由无人机从多尺度、多视角采集,具有广泛的几何重叠,覆盖城市、乡村等多种场景 | SfM 三维重建 | 训练/测试 | 无人机场景 |
| Air-to-Ground[82] | 图像来源于互联网照片,包含无人机视角与地面视角的匹配图像对,用于研究跨视角地理定位 | SfM 三维重建(COLMAP) | 训练/测试 | 无人机场景 |
| UAVScenes[128] | 提供多场景无人机图像以及对应的位姿真值和语义标注,还提供场景三维模型(可用于匹配定位) | SfM 三维重建(大疆智图) | 测试 | 无人机场景 |
| AnyVisLoc[27] | 包含中国 25 个区域的低空、多视角无人机图像和对应的航空/卫星基准图,提供无人机位姿真值 | SfM 三维重建 | 测试 | 无人机场景 |

先进软件,在处理存在显著模态和视角差异的图像对也容易引入误差。虽然一些研究会额外设计策略来滤除置信度较低的匹配点对(比如 RoMa 方法[91]在训练过程中引入深度连续性约束来滤除 MegaDepth 数据中不可靠的匹配点对),但这类后处理策略无法从根本上消除真值噪声。此外,一些无人机飞行模拟器(如 AirSim[129])以及三维渲染器(比如 Unreal Engine 和 Google Earth)也常被用来生成仿真图像,这些数据已在图像检索[120]以及视觉语言导航[130]等领域中被采用,但主流的跨视角图像匹配网络很少采用此类数据训练[82,90-91,101],可能原因在于其难以模拟真实场景中的复杂光照与瞬时变化,从而限制了模型的泛化性能[131]。

除数据规模以及真值精度的问题外,现有数据集还无法覆盖飞行器视觉定位的需求。现有大部分图像匹配数据集均以室内或街道场景为主[123-126],虽然部分数据集融入了空中视角数据[82,127-128],但其图像对多为无人机图像与街景,而非无人机图像与基准图。此外,一些多模态遥感数据集[132-133]也可用于匹配模型训练,但其图像除模态差异外缺乏显著视角变化。目前,包含低空倾斜实时图与正射基准图匹配真值的数据集仍较为缺乏,AnyVisLoc 数据集[27]虽尝试填补这一空白,但尚未完全开源。

## 6　现有飞行器视觉定位方法对比

为了分析现有绝对视觉定位方法的能力和局限,对近年来在边缘计算平台上部署的基于深度学习的飞行器视觉定位方法进行了调研和总结(见表 5)。从定位能力来看,大多数成功部署于边缘计算平台且精度达到米级的方法[36,38],均采用了“检索—匹配—位姿解算”的框架。Sui 等[134]的方案基于实时图初始帧位置已知的假设,因此未采用检索进行粗定位,这种策略虽然节省了计算资源,但依赖于准确的初始位姿先验。He 等[5]的方案未进行像素级匹配和 PnP 解算,直接在图像检索的基础上通过 Deep-LK(Lucas-Kanade)模块估计飞行器实时图与基准图间的单应性矩阵,进而推算飞行器的二维位置。然而,此类方法只适用于相机正下视成像且地面地形起伏不大的情况,在低空倾斜观测条件下通用性不足。从计算效率看,受限于边缘计算平台的算力与存储资源,现有方法的处理速度普遍较慢,单帧绝对视觉定位的最快处理速度通常在 0.3 s 以上。以 He 等[5]方法为例,尽管对基准图特征进行了离线提取与存储以降低计算负荷,但在 NVIDIA Jetson AGX Xavier 单帧处理时间仍需 0.396 s。目前主流的光学相机成像速率通常快于 0.02 s/帧,远高于后端视觉导航算法的处理能力,成像平台的实

**表 5　不同飞行器视觉定位方法在边缘计算平台上的推理性能**

Tab. 5　Inference performance of different UAV visual localization algorithms on edge computing platforms

| 方法 | 技术路径 | 边缘计算平台 | 平台最高算力 | 飞行高度/m | 机载相机分辨率 | 相机视场角/(°) | 其他传感器 | 定位场景尺度 | 定位精度/m | 单帧处理时间/s |
|---|---|---|---|---|---|---|---|---|---|---|
| Chen 等[36] | 检索—匹配—绝对位姿解算 | NVIDIA Jetson AGX Xavier | 32 TOPS | 70 | 640×480 | 84 | 无 | 400 m×400 m | 2.47 | 0.89 |
| LSVL[25] | 检索+视觉惯性里程计(VIO) | NVIDIA Jetson Nano | 0.47 TFLOPS | 50 | 未知 | 未知 | 惯导 | 1.68 km×3.82 km | 12.6~18.7 | 3.15 |
| Sui 等[134] | 匹配—绝对位姿解算 | NVIDIA Jetson Xavier NX | 21 TOPS | 300 | 1 920×1 080 | 84 | 无 | 6 km×6 km | 2.24 | 0.65~1.46 |
| FoundLoc[135] | 大范围检索+视觉惯性里程计(VIO) | NVIDIA Jetson Xavier NX | 21 TOPS | 50 | 2 464×2 056 | 100 | 惯导 | 142 000 $m^2$ | 19.38 | 0.517+0.125 |
| He 等[5] | 检索—绝对位姿解算+相对位姿解算 | NVIDIA Jetson AGX Xavier | 32 TOPS | 200 | 2 048×1 536 | 未知 | 无 | 2.23 km×1.29 km | 17.3 | 0.396 |
| GeoRVLF[38] | 大范围检索—匹配—绝对位姿解算<br>小范围匹配—绝对位姿解算+视觉里程计(VO) | NVIDIA Jetson Orin NX | 100 TOPS | 150/250 | 1 080 | 82.9 | 无 | 2.9 km×1.53 km | 2.73~4.58<br>5.68~7.15 | 36.41<br>0.747 |

时感知能力与绝对视觉定位算法计算效率之间存在脱节。在飞行速度较慢且对定位精度要求不高的应用场景中,现有算法尚可满足基本需求。然而,对于高速飞行器,例如第一人称视角(first person view, FPV)无人机(速度>50 m/s[136]),当前算法的处理速度仍无法满足实时定位的要求。从各模块时间占用来看,检索通常是耗时最多的环节[5,38,135],限制了整个系统的运行频率。

在硬件成本及部署方面,主流绝对视觉定位算法多基于NVIDIA Jetson系列平台实现。Jetson Nano的硬件成本为99~149美元,其FP16算力为0.47 TFLOPS,难以在保障实时性的前提下运行复杂的检索与匹配模型。而算力更为充沛的Jetson AGX Xavier(INT8算力为32 TOPS)与Orin系列(如AGX Orin 64 GB,INT8算力为275 TOPS)的开发套件成本通常在1 000~2 000美元,超过了多数小型FPV的硬件成本。此外,边缘计算单元的功耗(如AGX Orin功耗范围为15~60 W)与质量(核心模块及散热结构通常超过100 g)会占用小型FPV有限的载荷空间并缩短续航时间,给飞行器的总体结构设计与能源管理带来严峻挑战。

在系统鲁棒性方面,表5中方法多选择宽视场角(>83°)相机进行成像,可能原因在于宽视场角能够覆盖更大范围的场景,获取更丰富的环境上下文信息,进而提升无人机实时图与基准图之间检索匹配的鲁棒性。此外,表5中的方法在无人机-基准图检索与匹配流程之外均引入了额外的保障策略。例如,LSVL[25]与Sui等[134]所提方法利用成像姿态信息对飞行器实时图进行几何校正,以降低视角偏差并提升匹配准确率,但该方法依赖高精度的姿态先验。许多方法在绝对视觉定位的基础上集成了视觉里程计或视觉惯性里程计[5,25,38,135],利用帧间跟踪与惯导参数增加系统鲁棒性。此类方法需进行多源信息融合,但在飞行器机动转弯或俯冲等运动剧烈的时刻,非线性运动引起的图像模糊或剧烈视角变化可能导致跟踪漂移或失败。此外,Chen等[36]在检索模块后利用PnP算法RANSAC的内点数量对检索得到的备选区域进行二次筛选,缓解了地物重复性纹理导致的检索误匹配问题。总体而言,这些工程部署方法通过增加几何或物理约束提升了系统的环境适应能力,但这些策略也增加了算法的复杂度或对辅助传感器的依赖,需根据实际应用需求进行权衡。

尽管目前的方法在计算实时性、硬件部署成本及鲁棒性上还存在不足,但"检索—匹配—位姿解算"框架依然是解决低空观测问题的有效途径。因此,未来研究应致力于该框架下特征提取与检索匹配模型的轻量化设计及推理加速,在保持其应对复杂观测条件鲁棒性的同时大幅降低计算延迟,从而在资源受限的机载平台上真正实现满足高速飞行需求的高精度实时定位。

## 7 总结与展望

本文综述了基于"检索—匹配—位姿解算"的低空飞行器视觉定位技术。现有技术虽已取得显著进展,但仍面临多方面挑战。在图像检索环节,现有模型训练数据多基于仿真或特定场景,对真实世界数据的跨域泛化能力不足;在图像匹配环节,缺乏高质量的"无人机-卫星/航空影像"匹配真值数据,制约了算法性能评估,且现有匹配网络仍难以兼顾效率与鲁棒性能;在位姿解算环节,依赖高精度的三维基准数据(如正射影像和DSM),但开源数据往往分辨率不足或存在几何畸变,限制了技术广泛应用。从工程部署角度看,现有方案多依赖NVIDIA平台,鲜有研究在国产边缘计算设备上开展算法验证,制约了技术链的自主可控发展。

此外,本文的论述重点聚焦于低空倾斜观测引发的视角差异挑战,但在实际复杂任务场景中,视觉定位系统往往需在全天候环境下运行,这不可避免地涉及可见光图像与红外、合成孔径雷达(synthetic aperture radar,SAR)或激光雷达等异源数据的跨模态匹配。如何在同时存在"极端视角差异"与"传感器模态差异"的复杂场景下实现飞行器高精度绝对定位仍是当前视觉导航领域尚未解决的核心难题。

针对上述局限,未来的研究可从几个维度展开突破:在跨视角检索方面,研究基于自监督学习与生成式数据增强的训练方法[103,106],提升跨域泛化能力;在跨视角匹配方面,构建大规模、高精度的低空视觉定位基准数据集,并利用知识蒸馏等技术将大模型能力迁移至轻量化网络[137-138],平衡效率与鲁棒性。同时应针对全天时全天候任务需求,重点攻克跨视角多模态图像一致性特征深度表征与跨域关联问题[104,122],提升复杂环境下的检索匹配鲁棒性。在位姿解算方面,设计对低质量基准数据具有强鲁棒性的优化方法,并深入研究视觉与INS、高度计等多源信息紧耦合的组合导航算法[25,135],提升系统可靠性。此外,还应积极开展算法在国产化平台上的适配与加速研

究,推动软硬件协同发展。随着以上技术的进步,基于“检索—匹配—位姿解算”的视觉定位技术将为低空飞行器实现全天候、全地域自主导航提供有力支撑。

## 参考文献(References)


[1] XIANG T Z, XIA G S, ZHANG L P. Mini-unmanned aerial vehicle-based remote sensing: techniques, applications, and prospects [J]. IEEE Geoscience and Remote Sensing Magazine, 2019, 7(3): 29 – 63.

[2] ZHU J L, YAN S, WANG L, et al. Loc-Loc: aerial visual localization using Lod 3D map with neural wireframe alignment[EB/OL]. (2024 – 10 – 16) [2026 – 01 – 17]. https://arxiv.org/abs/2410.12269.

[3] AVOLA D, CINQUE L, EMAM E, et al. UAV geo-localization for navigation: a survey [J]. IEEE Access, 2024, 12: 125332 – 125357.

[4] COUTURIER A, AKHLOUFI M A. A review on absolute visual localization for UAV [J]. Robotics and Autonomous Systems, 2021, 135: 103666.

[5] HE M F, LIU J C, GU P F, et al. Leveraging map retrieval and alignment for robust UAV visual geo-localization [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 2523113.

[6] CHANG Y X, CHENG Y Q, MANZOOR U, et al. A review of UAV autonomous navigation in GPS-denied environments[J]. Robotics and Autonomous Systems, 2023, 170: 104533.

[7] JAMES S. Black Widow drone completes successful GPS-denied navigation test with Palantir software[EB/OL]. (2025 – 10 – 31) [2026 – 01 – 17]. https://www.defenseadvancement.com/news/black-widow-drone-completes-successful-gps-denied-navigation-test-with-palantir-software/.

[8] JOHNSON A E, CHENG Y, TRAWNY N, et al. Implementation of a map relative localization system for planetary landing [J]. Journal of Guidance Control and Dynamics, 2023, 46(4): 618 – 637.

[9] 大疆. 视觉传感导航系统[EB/OL]. [2026 – 01 – 17]. https://www.dji.com/cn/guidance.
DJI. Visual sensor navigation system [EB/OL]. [2026 – 01 – 17]. https://www.dji.com/cn/guidance. (in Chinese)

[10] MUGHAL M H, KHOKHAR M J, SHAHZAD M. Assisting UAV localization via deep contextual image matching [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 2445 – 2457.

[11] SCHLEISS M, ROUATBI F, CREMERS D. VPAIR-aerial visual place recognition and localization in large-scale outdoor environments[EB/OL]. (2022 – 05 – 23) [2026 – 01 – 17]. https://arxiv.org/abs/2205.11567.

[12] SONG W H, JUNG H G, GWAK I Y, et al. Oblique aerial image matching based on iterative simulation and homography evaluation[J]. Pattern Recognition, 2019, 87: 317 – 331.

[13] WANG C Y, CHEN J B, CHEN J S, et al. Unmanned aerial vehicle oblique image registration using an ASIFT-based matching method[J]. Journal of Applied Remote Sensing, 2018, 12(2): 025002.

[14] CHEN Y, JIANG J. An oblique-robust absolute visual localization method for GPS-denied UAV with satellite imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5601713.

[15] YE Q, LUO J Q, LIN Y. A coarse-to-fine visual geo-localization method for GNSS-denied UAV with oblique-view imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 212: 306 – 322.

[16] ZHENG Z D, WEI Y C, YANG Y. University-1652: a multi-view multi-source benchmark for drone-based geo-localization[C]//Proceedings of the 28th ACM International Conference on Multimedia, 2020: 1395 – 1403.

[17] ZHU R Z, YIN L, YANG M Z, et al. SUES-200: a multi-height multi-scene cross-view image benchmark across drone and satellite[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(9): 4825 – 4839.

[18] 罗世彬, 刘海桥, 胡茂青, 等. 无人飞行器异源图像匹配辅助惯性导航定位技术综述[J]. 国防科技大学学报, 2020, 42(6): 1 – 10.
LUO S B, LIU H Q, HU M Q, et al. Review of multi-modal image matching assisted inertial navigation positioning technology for unmanned aerial vehicle [J]. Journal of National University of Defense Technology, 2020, 42(6): 1 – 10. (in Chinese)

[19] AL-JARRAH O Y, SHATNAWI A S, SHURMAN M M, et al. Exploring deep learning-based visual localization techniques for UAVs in GPS-denied environments[J]. IEEE Access, 2024, 12: 113049 – 113071.

[20] 袁媛, 孙柏, 刘赶超. 景象匹配无人机视觉定位[J]. 自动化学报, 2025, 51(2): 287 – 311.
YUAN Y, SUN B, LIU G C. Drone-based scene matching visual geo-localization [J]. Acta Automatica Sinica, 2025, 51(2): 287 – 311. (in Chinese)

[21] 苗宗成, 周润玺, 柳杰, 等. 无人机绝对视觉定位的研究进展[J]. 液晶与显示, 2025, 40(6): 942 – 955.
MIAO Z C, ZHOU R X, LIU J, et al. Research progress on absolute visual localization of unmanned aerial vehicles[J]. Chinese Journal of Liquid Crystals and Displays, 2025, 40(6): 942 – 955. (in Chinese)

[22] 谷美颖, 李航, 张家伟, 等. 基于视觉的无人机定位与导航方法研究综述[J]. 电子学报, 2025, 53(3): 651 – 685.
GU M Y, LI H, ZHANG J W, et al. A review of vision-based UAV localization and navigation methods [J]. Acta Electronica Sinica, 2025, 53(3): 651 – 685. (in Chinese)

[23] CHEN S, CAVALLARI T, PRISACARIU V A, et al. Map-relative pose regression for visual re-localization [C]// Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 20665 – 20674.

[24] LI H L, WANG J N, WEI Z W, et al. Jointly optimized global-local visual localization of UAVs[EB/OL]. (2023 – 10 – 12) [2026 – 01 – 17]. https://arxiv.org/abs/2310.08082.

[25] KINNARI J, RENZULLI R, VERDOJA F, et al. LSVL: large-scale season-invariant visual localization for UAVs[J]. Robotics and Autonomous Systems, 2023, 168: 104497.

[26] ZHOU Q, TANG H C, ZHANG Z X, et al. A hierarchical absolute visual localization system for low-altitude drones in GNSS-denied environments [J]. Remote Sensing, 2025, 17(20): 3470.

[27] YE Y B, TENG X C, CHEN S, et al. Exploring the best way for UAV visual localization under low-altitude multi-view

observation condition: a benchmark[EB/OL]. (2025-03-12)[2026-01-17]. https://arxiv.org/abs/2503.10692.

[28] TANG B, LU R T, YANG X G, et al. R2PLoc: a region-to-point UAV visual geo-localization framework leveraging hierarchical semantic representation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5643818.

[29] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.

[30] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[31] SARLIN P E, UNAGAR A, LARSSON M, et al. Back to the feature: learning robust camera localization from pixels to pose[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 3246-3256.

[32] SARLIN P E, DETONE D, YANG T Y, et al. OrienterNet: visual localization in 2D public maps with neural matching[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 21632-21642.

[33] KHURSHID M, SHAHZAD M, KHATTAK H A, et al. Vision-based 3-D localization of UAV using deep image matching[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 12020-12030.

[34] LI W, LIU C, YU S S, et al. LightLoc: learning outdoor LiDAR localization at light speed[C]//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025: 6680-6689.

[35] VAN DALEN G J, MAGREE D P, JOHNSON E N. Absolute localization using image alignment and particle filtering[C]//Proceedings of AIAA Guidance, Navigation, and Control Conference, 2016: 0647.

[36] CHEN S X, WU X Y, MUELLER M W, et al. Real-time geo-localization using satellite imagery and topography for unmanned aerial vehicles[C]//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021: 2275-2281.

[37] GAN W J, ZHOU Y, HU X F, et al. Learning robust feature representation for cross-view image geo-localization[J]. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 6004405.

[38] ZHANG Z Y, CHU J K, SONG T, et al. GeoRVLF: a robust drone-satellite visual geo-localization framework for small unmanned aerial vehicle platforms[J]. IEEE Robotics and Automation Letters, 2025, 10(7): 7380-7387.

[39] 叶熠彬, 滕锡超, 于起峰, 等. 基于MatchNet和多点匹配约束的可见光-SAR图像匹配[J]. 航空学报, 2024, 45(10): 329162.
YE Y B, TENG X C, YU Q F, et al. Optical-SAR image matching based on MatchNet and multi-point matching constraint[J]. Acta Aeronauticaet Astronautica Sinica, 2024, 45(10): 329162. (in Chinese)

[40] ATALLAH M J. Faster image template matching in the sum of the absolute value of differences measure[J]. IEEE Transactions on Image Processing, 2001, 10(4): 659-663.

[41] HISHAM M B, YAAKOB S N, RAOF R A A, et al. Template matching using sum of squared difference and normalized cross correlation[C]//Proceedings of 2015 IEEE Student Conference on Research and Development (SCOReD), 2015: 100-104.

[42] SARVAIYA J N, PATNAIK S, BOMBAYWALA S. Image registration by template matching using normalized cross-correlation[C]//Proceedings of 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009: 819-822.

[43] YE Y X, SHAN J, BRUZZONE L, et al. Robust registration of multimodal remote sensing images based on structural similarity[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2941-2958.

[44] YE Y B, WANG Q W, ZHAO H, et al. Fast and robust optical-to-SAR remote sensing image registration using region-aware phase descriptor[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5208512.

[45] HEINRICH M P, JENKINSON M, BHUSHAN M, et al. MIND: modality independent neighbourhood descriptor for multi-modal deformable registration[J]. Medical Image Analysis, 2012, 16(7): 1423-1435.

[46] TENG X C, LIU X C, LI Z, et al. OMIRD: orientated modality independent region descriptor for optical-to-SAR image matching[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 4003405.

[47] SIVIC J, ZISSERMAN A. Video Google: a text retrieval approach to object matching in videos[C]//Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003: 1470-1477.

[48] MATAS J, CHUM O, URBAN M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and Vision Computing, 2004, 22(10): 761-767.

[49] PHILBIN J, CHUM O, ISARD M, et al. Object retrieval with large vocabularies and fast spatial matching[C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[50] JÉGOU H, DOUZE M, SCHMID C, et al. Aggregating local descriptors into a compact image representation[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 3304-3311.

[51] YU G S, MOREL J M. ASIFT: an algorithm for fully affine invariant comparison[J]. Image Processing on Line, 2011, 1: 11-38.

[52] LIN T Y, BELONGIE S, HAYS J. Cross-view image geolocalization[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013: 891-898.

[53] CASTALDO F, ZAMIR A, ANGST R, et al. Semantic cross-view matching[C]//Proceedings of 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 2015: 1044-1052.

[54] WANG T Y, ZHENG Z D, YAN C G, et al. Each part matters: local patterns facilitate cross-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(2): 867-879.

[55] CHEN Q, WANG T Y, YANG Z H, et al. SDPL: shifting-dense partition learning for UAV-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11810-11824.

[56] HOU K J, TONG Q, YAN N, et al. MCFA: multi-scale cascade and feature adaptive alignment network for cross-view geo-localization[J]. Sensors, 2025, 25(14): 4519.
[57] GE F W, ZHANG Y Z, LIU Y X, et al. Multibranch joint representation learning based on information fusion strategy for cross-view geo-localization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5909516.
[58] DAI M, HU J H, ZHUANG J D, et al. A transformer-based feature segmentation and region alignment method for UAV-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4376 - 4389.
[59] WU Q, WAN Y, ZHENG Z, et al. CAMP: a cross-view geo-localization method using contrastive attributes mining and position-aware partitioning [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1 - 14.
[60] SHEN T R, WEI Y M, KANG L, et al. MCCG: a convnext-based multiple-classifier method for cross-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(3): 1456 - 1468.
[61] DEUSER F, HABEL K, OSWALD N. Sample4Geo: hard negative sampling for cross-view geo-localisation[EB/OL]. (2023 - 08 - 29) [2026 - 01 - 17]. https://arxiv.org/abs/2303.11851.
[62] XIA P W, WAN Y, ZHENG Z, et al. Enhancing cross-view geo-localization with domain alignment and scene consistency[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(12): 13271 - 13281.
[63] CHEN J L, WEN G J, JIAN H J, et al. A visual localization benchmark for UAVs in complex multi-terrain environments[J/OL]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025: 1 - 15 (2025 - 01 - 07) [2026 - 01 - 17]. https://ieeexplore.ieee.org/document/10829853.
[64] HU S Y, SHI Z L, JIN T, et al. Query-driven feature learning for cross-view geo-localization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1 - 15.
[65] LI H Y, XU C, YANG W, et al. Learning cross-view visual geo-localization without ground truth[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5632017.
[66] CHEN Z W, YANG Z X, RONG H J, et al. Without paired labeled data: end-to-end self-supervised learning for drone-view geo-localization[EB/OL]. (2025 - 11 - 26) [2026 - 01 - 17]. https://arxiv.org/abs/2502.11381.
[67] MISHKIN D, MATAS J, PERDOCH M, et al. WxBS: wide baseline stereo generalizations[EB/OL]. (2015 - 05 - 12) [2026 - 01 - 17]. https://arxiv.org/abs/1504.06603.
[68] JIN Y H, MISHKIN D, MISHCHUK A, et al. Image matching across wide baselines: from paper to practice[J]. International Journal of Computer Vision, 2021, 129(2): 517 - 547.
[69] KARPUR A, PERROTTA G, MARTIN-BRUALLA R, et al. LFM-3D: learnable feature matching across wide baselines using 3D signals [C]//Proceedings of 2024 International Conference on 3D Vision (3DV), 2024: 11 - 20.
[70] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 - 110.
[71] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speeded up robust features [C]//Proceedings of Computer Vision-ECCV 2006, 2006: 404 - 417.
[72] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of 2011 International Conference on Computer Vision, 2011: 2564 - 2571.
[73] CAI G R, JODOIN P M, LI S Z, et al. Perspective-SIFT: an efficient tool for low-altitude remote sensing image registration[J]. Signal Processing, 2013, 93(11): 3088 - 3110.
[74] YU Y N, HUANG K Q, CHEN W, et al. A novel algorithm for view and illumination invariant image matching[J]. IEEE Transactions on Image Processing, 2012, 21(1): 229 - 240.
[75] HU H, ZHU Q, DU Z Q, et al. Reliable spatial relationship constrained feature point matching of oblique aerial images[J]. Photogrammetric Engineering & Remote Sensing, 2015, 81(1): 49 - 58.
[76] ROTH L, KUHN A, MAYER H. Wide-baseline image matching with projective view synthesis and calibrated geometric verification[J]. PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 2017, 85(2): 85 - 95.
[77] ZHANG Q Y, ZHENG S Y, ZHANG C, et al. Efficient large-scale oblique image matching based on cascade hashing and match data scheduling[J]. Pattern Recognition, 2023, 138: 109442.
[78] CHEN M, ZHU Q, YAN S, et al. LGS: local geometrical structure-based interest point matching for wide-baseline imagery in urban areas [J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019, IV - 2/W5: 13 - 20.
[79] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: self-supervised interest point detection and description [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018: 337 - 33712.
[80] DUSMANU M, ROCCO I, PAJDLA T, et al. D2-Net: a trainable CNN for joint description and detection of local features[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 8084 - 8093.
[81] REVAUD J, WEINZAEPFEL P, DE SOUZA C, et al. R2D2: reliable and repeatable detector and descriptor[EB/OL]. (2019 - 06 - 17) [2026 - 01 - 17]. https://arxiv.org/abs/1906.06195.
[82] CHEN G L, FU T W, CHEN H W, et al. RDD: robust feature detector and descriptor using deformable transformer[C]//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025: 6394 - 6403.
[83] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: learning feature matching with graph neural networks[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 4937 - 4946.
[84] LINDENBERGER P, SARLIN P E, POLLEFEYS M. LightGlue: local feature matching at light speed [C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 17581 - 17592.
[85] SUN J M, SHEN Z H, WANG Y, et al. LoFTR: detector-free local feature matching with transformers [C]//

Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 8918 – 8927.

[86] WANG Q, ZHANG J M, YANG K L, et al. MatchFormer: interleaving attention in transformers for feature matching[C]//Proceedings of Computer Vision – ACCV 2022, 2023: 256 – 273.

[87] CHEN H K, LUO Z X, ZHOU L, et al. ASpanFormer: detector-free image matching with adaptive span transformer[C]// Proceedings of Computer Vision – ECCV 2022, 2022: 20 – 36.

[88] LU X Y, DU S L. Raising the ceiling: conflict-free local feature matching with dynamic view switching [C]// Proceedings of Computer Vision – ECCV 2025, 2024: 256 – 273.

[89] TRUONG P, DANELLJAN M, TIMOFTE R, et al. PDC-Net+: enhanced probabilistic dense correspondence network [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10247 – 10266.

[90] EDSTEDT J, ATHANASIADIS I, WADENBÄCK M, et al. DKM: dense kernelized feature matching for geometry estimation[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 17765 – 17775.

[91] EDSTEDT J, SUN Q Y, BÖKMAN G, et al. RoMa: robust dense feature matching[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 19790 – 19800.

[92] CADAR F, POTJE G, MARTINS R, et al. Leveraging semantic cues from foundation vision models for enhanced local feature correspondence[C]//Proceedings of Computer Vision – ACCV 2024, 2024: 54 – 70.

[93] JIANG H W, KARPUR A, CAO B Y, et al. OmniGlue: generalizable feature matching with foundation model guidance[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 19865 – 19875.

[94] ZHANG S H, ZHU Z J, LI Z Z, et al. Matching while perceiving: enhance image feature matching with applicable semantic amalgamation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39 (10): 10094 – 10102.

[95] YANG M, FAN F, LI Z Z, et al. DistillMatch: leveraging knowledge distillation from vision foundation model for multimodal image matching[EB/OL]. (2025 – 09 – 19) [2026 – 01 – 17]. https://arxiv.org/abs/2509.16017.

[96] ZHANG Y S, ZHAO X. MESA: matching everything by segmenting anything[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 20217 – 20226.

[97] TOFT C, TURMUKHAMBETOV D, SATTLER T, et al. Single-image depth prediction makes feature matching easier[C]//Proceedings of Computer Vision – ECCV 2020, 2020: 473 – 492.

[98] WANG S Z, KANNALA J, POLLEFEYS M, et al. Guiding local feature matching with surface curvature [C]// Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 17935 – 17945.

[99] HUANG C W, PAN X, CHENG J C, et al. Deep image registration with depth-aware homography estimation [J]. IEEE Signal Processing Letters, 2023, 30: 6 – 10.

[100] LIU Y P, LAI W P, ZHAO Z, et al. LiftFeat: 3D geometry-aware local feature matching [C]//Proceedings of 2025 IEEE International Conference on Robotics and Automation (ICRA), 2025: 11714 – 11720.

[101] POTJE G, CADAR F, ARAUJO A, et al. XFeat: accelerated features for lightweight image matching [C]// Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 2682 – 2691.

[102] SHEN X L, CAI Z P, YIN W, et al. GIM: learning generalizable image matcher from internet videos[EB/OL]. (2024 – 02 – 16) [2026 – 01 – 17]. https://arxiv.org/abs/2402.11095.

[103] REN J W, JIANG X Y, LI Z Z, et al. MINIMA: modality invariant image matching[C]//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025: 23059 – 23068.

[104] HE X Y, YU H, PENG S D, et al. MatchAnything: universal cross-modality image matching with large-scale pre-training[EB/OL]. (2025 – 01 – 13) [2026 – 01 – 17]. https://arxiv.org/abs/2501.07556.

[105] TYSZKIEWICZ M J, FUA P, TRULLS E. DISK: learning local features with policy gradient[EB/OL]. (2020 – 10 – 27) [2026 – 01 – 17]. https://arxiv.org/abs/2006.13566.

[106] KÜNZEL J, HILSMANN A, EISERT P. RIPE: reinforcement learning on unlabeled image pairs for robust keypoint extraction[EB/OL]. (2020 – 07 – 14) [2026 – 01 – 17]. https://arxiv.org/abs/2507.04839.

[107] GAO X S, HOU X R, TANG J L, et al. Complete solution classification for the perspective-three-point problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 930 – 943.

[108] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: an accurate $O(n)$ solution to the PnP problem [J]. International Journal of Computer Vision, 2009, 81(2): 155 – 166.

[109] LI S Q, XU C, XIE M. A robust $O(n)$ solution to the perspective-n-point problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1444 – 1450.

[110] HESCH J A, ROUMELIOTIS S I. A direct least-squares (DLS) method for PnP [C]//Proceedings of 2011 International Conference on Computer Vision, 2011: 383 – 390.

[111] SCHWEIGHOFER G, PINZ A. Globally optimal $O(n)$ solution to the PnP problem for general camera models[C]// Proceedings of the British Machine Vision Conference 2008, 2008: 1 – 10.

[112] URBAN S, LEITLOFF J, HINZ S. MLPnP: a real-time maximum likelihood solution to the perspective-n-point problem[EB/OL]. (2016 – 07 – 27) [2026 – 01 – 17]. https://arxiv.org/abs/1607.08112.

[113] ZHAO C H, WU D W, HE J, et al. A visual positioning method of UAV in a large-scale outdoor environment[J]. Sensors, 2023, 23(15): 6941.

[114] ZHAN T, XU C F, ZHANG C, et al. Generalized maximum likelihood estimation for perspective-n-point problem [J]. IEEE Robotics and Automation Letters, 2025, 10(2):

1752 – 1759.

[115] WU R W, CHENG X Y, ZHU J L, et al. UAVD4L: a large-scale dataset for UAV 6-DoF localization [C]// Proceedings of 2024 International Conference on 3D Vision (3DV), 2024: 1574 – 1583.

[116] LI Y F. IMU-aided geographic pose estimation method for UAVs using satellite imageries matching[J]. IEEE Robotics and Automation Letters, 2025, 10(3): 2902 – 2909.

[117] JUBRAN I, FARES F, ALFASSI Y, et al. Newton-PnP: real-time visual navigation for autonomous toy-drones[C]// Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022: 13363 – 13370.

[118] WANG T Y, ZHENG Z D, SUN Y Q, et al. Multiple-environment self-adaptive network for aerial-view geo-localization[J]. Pattern Recognition, 2024, 152: 110363.

[119] DAI M, ZHENG E H, FENG Z H, et al. Vision-based UAV self-positioning in low-altitude urban environments[J]. IEEE Transactions on Image Processing, 2024, 33: 493 – 508.

[120] XU W J, YAO Y X, CAO J Q, et al. UAV-VisLoc: a large-scale dataset for UAV visual localization[EB/OL]. (2024 – 05 – 20) [2026 – 01 – 17]. https://arxiv.org/abs/2405.11936.

[121] JI Y X, HE B Y, TAN Z Y, et al. Game4Loc: a UAV geo-localization benchmark from game data[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(4): 3913 – 3921.

[122] ZHOU X, YANG X R, ZHANG Y C. CDM-Net: a framework for cross-view geo-localization with multimodal data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 4706416.

[123] LI Z Q, SNAVELY N. MegaDepth: learning single-view depth prediction from Internet photos[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 2041 – 2050.

[124] BALNTAS V, LENC K, VEDALDI A, et al. HPatches: a benchmark and evaluation of handcrafted and learned local descriptors[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 3852 – 3861.

[125] GEIGER A. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012: 3354 – 3361.

[126] DAI A, CHANG A X, SAVVA M, et al. ScanNet: richly-annotated 3D reconstructions of indoor scenes [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2432 – 2443.

[127] SHEN T W, LUO Z X, ZHOU L, et al. Matchable image retrieval by learning from surface reconstruction [C]// Proceedings of Computer Vision – ACCV 2018, 2019: 415 – 431.

[128] WANG S J, LI S Q, ZHANG Y W, et al. UAVScenes: a multi-modal dataset for UAVs[EB/OL]. (2025 – 07 – 30) [2026 – 01 – 17]. https://arxiv.org/abs/2507.22412.

[129] SHAH S, DEY D, LOVETT C, et al. AirSim: high-fidelity visual and physical simulation for autonomous vehicles[C]// Proceedings of Field and Service Robotics, 2017: 621 – 635.

[130] LIU S B, ZHANG H S, QI Y K, et al. AerialVLN: vision-and-language navigation for UAVs[C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 15338 – 15348.

[131] VUONG K, GHOSH A, RAMANAN D, et al. AerialMegaDepth: learning aerial-ground reconstruction and view synthesis [C]//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025: 21674 – 21684.

[132] LI J Y, HU Q W, ZHANG Y J. Multimodal image matching: a scale-invariant algorithm and an open dataset[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 204: 77 – 88.

[133] LIU Y X, LIU Y, YAN S, et al. A multi-view thermal-visible image dataset for cross-spectral matching[J]. Remote Sensing, 2023, 15(1): 174.

[134] SUI H G, LI J J, LEI J F, et al. A fast and robust heterologous image matching method for visual geo-localization of low-altitude UAVs [J]. Remote Sensing, 2022, 14(22): 5879.

[135] HE Y, CISNEROS I, KEETHA N, et al. FoundLoc: vision-based onboard aerial localization in the wild [EB/OL]. (2023 – 10 – 25) [2026 – 01 – 17]. https://arxiv.org/abs/2310.16299.

[136] EFAZ E T, MOWLEE M M, JABIN J, et al. Modeling of a high-speed and cost-effective FPV quadcopter for surveillance [C]//Proceedings of 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020: 1 – 6.

[137] QUAN D, WANG Z, LV C H, et al. LM-Net: a lightweight matching network for remote sensing image matching and registration [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5229313.

[138] EDSTEDT J, BÖKMAN G, WADENBÄCK M, et al. DaD: distilled reinforcement learning for diverse keypoint detection[EB/OL]. (2025 – 03 – 11) [2026 – 01 – 17]. https://arxiv.org/abs/2503.07347.