

DOI:10.19344/j.cnki.issn1671-5276.2026.04.019

基于深度学习的无人机三维影像处理方法研究

陈伟江

(上海工艺美术职业学院, 上海 201808)

摘要:随着无人机遥感技术的快速发展,基于深度学习的三维影像处理方法已成为研究热点。针对传统三维重建算法在处理效率、特征提取与模型泛化能力方面的不足,探讨如何利用深度学习技术来提升无人机影像三维建模的精度与自动化水平。研究包括特征匹配、点云生成与表面重建等关键环节的深度学习实现方法,并通过实验验证所提方法在复杂场景下的有效性和鲁棒性。结果表明:深度融合学习模型能够显著改善三维模型的细节表现与整体结构一致性,为高精度三维重建提供可行的技术路径,在智慧城市、地形测绘及灾害评估等领域具有广泛的应用前景。

关键词:无人机;深度学习;人工智能;三维重构

中图分类号:TP183 **文献标志码:**A **文章编号:**1671-5276(2026)04-0098-05

Research on 3D Image Processing Method of Unmanned Aerial Vehicles Based on Deep Learning

Chen Weijiang

(Shanghai Art & Design Academy, Shanghai 201808, China)

Abstract: With the rapid development of unmanned aerial vehicle (UAV) remote sensing technology, 3D image processing methods based on deep learning have become a research hotspot. Addressing the shortcomings of traditional 3D reconstruction algorithms in terms of processing efficiency, feature extraction, and model generalization capability, this paper discusses how to utilize deep learning technology to improve the accuracy and automation level of 3D modeling from UAV imagery. The research covers deep neural network implementation methods for key stages such as feature matching, point cloud generation, and surface reconstruction, and Experiments validate the effectiveness and robustness of the proposed methods in complex scenarios. The results show that the deep fusion learning model can significantly improve the detail representation and overall structural consistency of 3D models, providing a feasible technical pathway for high-precision 3D reconstruction. It has broad application prospects in fields such as smart cities, topographic mapping, and disaster assessment.

Keywords: drone; deep learning; artificial intelligence; 3D reconstruction

0 引言

无人机遥感技术的蓬勃发展为三维地理空间信息的高效、低成本获取开辟了全新途径,它已成为智慧城市、数字孪生、实景三维中国建设及地质灾害应急响应等领域不可或缺的技术支撑^[1-2]。其核心优势在于能够快速获取高重叠度、超高分辨率的序列影像,为后续高精度三维建模奠定了坚实的数据基础。传统的三维重建技术体系,如运动恢复结构(structure from motion, SFM)和密集多视图立体(deuse multi-view stereo, DMVS),经过数十年发展已形成较为成熟的工艺流程^[3-6]。这些方法依赖于手工设计的局部特征描

述符(如 SIFT、ORB)和复杂的全局束调整(bundle adjustment)优化,在光照理想与纹理丰富的场景中能够取得令人满意的重建效果。然而,随着应用场景的不断拓展与深化,其固有的局限性也日益凸显。

尽管研究者对传统 SfM/MVS pipeline 进行了大量优化,但面对实际应用中常见的挑战,如大范围弱纹理区域(如墙面、水面)、重复性结构、剧烈光照变化以及复杂遮挡等情况时,基于手工特征的方法,其鲁棒性与泛化能力显著下降,极易导致匹配失败、点云稀疏甚至重建中断^[7]。近年来,深度学习技术的崛起为计算机视觉领域带来了革命性变化,也为解决上述技术瓶颈提供了新的思路。

基金项目:上海市高等教育学会 2024-2026 年度规划研究课题项目(1QZD2423)

作者简介:陈伟江(1975—),男,上海人,副教授,硕士,研究方向为数字媒体艺术、摄影摄像艺术、艺术设计、影视特效, chjj7768@126.com。

众多研究者尝试将卷积神经网络(CNN)、图神经网络(GNN)和Transformer等架构引入三维重建的各个环节,例如利用深度学习进行更具判别力的特征提取与匹配(如SuperPoint、LoFTR)、直接从多视图图像中回归深度图(如MVSNet系列方法)、以及对生成的点云进行智能优化与补全。这些学习驱动的方法在特定数据集上展现出了超越传统方法的精度和稳定性,标志着三维重建研究正逐步从“几何驱动”迈向“数据与几何联合驱动”的新范式^[8-13]。

然而,现有的基于深度学习的解决方案仍面临一系列挑战。首先,许多方法处于“孤立”状态,仅优化重建流程中的某一个环节(如只改进匹配或只优化深度估计),缺乏一个端到端、全局协同优化的统一框架^[14-17]。其次,网络模型对训练数据分布极为敏感,在无人机航拍数据(存在大幅面地形起伏、视角变化大等特点)上的泛化能力仍有待进一步验证。此外,如何有效地融合几何先验知识与深度学习模型的内在表征能力,以同时保障结果的精确度与可靠性,是一个关键难题。鉴于此,本文旨在系统性地研究一套深度融合学习技术的无人机影像三维处理全流程方法^[18]。本研究通过构建一个集成化的网络架构,将特征匹配、稠密深度估计及点云滤波优化等步骤有机统一,并引入注意力机制与多尺度特征融合策略,以期显著提升在复杂场景下的重建完整性、细节精度与自动化水平,最终推动无人机三维重建技术向更智能、鲁棒性更强的方向发展。

1 基于深度学习的无人机三维地面目标视觉导航研究

1.1 数据集的训练与验证

针对基于深度学习的无人机三维地面目标视觉导航研究,本文构建了一个融合真实采集与仿真生成的大规模专用数据集,通过搭载多传感器的无人机平台在多种典型环境与条件下采集目标图像及高精度位姿真值,并利用AirSim等仿真平台补充极端场景数据以增强其多样性;采用迁移学习策略,使用在Unity^[19-22]大型数据集上预训练的主干网络进行权重初始化,并在自建数据集上对导航模型进行精细调优。通过构建回归损失函数联合优化位置与姿态预测精度,最终利用严格划分的验证集对模型性能进行系统评估,以确保其泛化能力与可靠性。

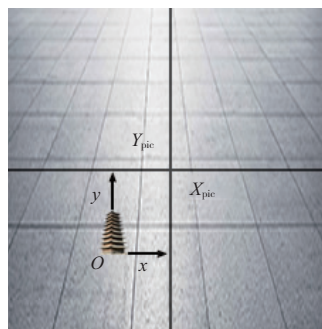
1.2 无人机相机位置和姿态的确定

为解决原始图像尺寸(2 000×2 000 像素)过大导致的卷积神经网络处理困难,本文采用分阶段处理策略^[23]。第1阶段使用第1个人工神经网络,对采样图像降至227×227 像素进行处理,以估计物体原点在图像中的投影坐标(x, y)。第2阶段基于第1阶段输出的坐标信息,对原始图像进行裁剪和重采样,生成以目标投影点为中心的227×227 像素修正图像,并输入到第2个人工神经网络中,用于估计无人机相机的空间位置与姿态参数,包括相机光学中心与物体原点间的距离 d 、表示相机朝向的四元数分量(W_q, X_q, Y_q, Z_q),以及投影点坐标的refined估计值(x', y')。通过这种两级串联的神经网络架构,在降低计算复杂度的同时,实现了对相机位姿的高精度估计。

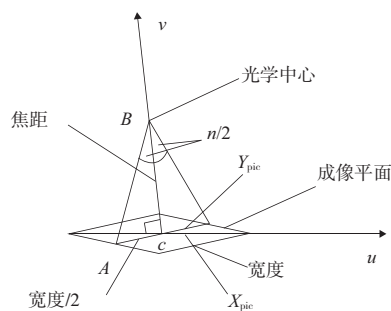
1) 无人机相机定位的第1阶段

a) 第1阶段的网络架构及其训练

本部分阐述第1阶段的任务目标,即在大尺度图像中定位物体原点投影的像素坐标,其像素坐标图详见图1所示。



(a) 像素坐标网格示意图



(b) 相机成像几何模型示意图

图1 大尺度图像中定位物体原点投影的像素坐标图

该投影点的水平与垂直坐标(u, v)需根据图像中心进行归一化处理,具体计算方法为将其像素坐标值除以相机的焦距 F ,最终表示为

$$u = X_{pic}/F \quad (1)$$

$$v = Y_{\text{pic}} / F \quad (2)$$

式中: $(X_{\text{pic}}, Y_{\text{pic}})$ 为第 1 阶段神经网络预测的、物体原点在二维图像平面上的投影坐标; F 为以像素为单位的相机焦距, 计算公式如下所示

$$F = \frac{W_{\text{idth}}}{2} \text{ctg}\left(\frac{\alpha}{2}\right) \quad (3)$$

式中: W_{idth} 为以像素为单位的图像宽度; α 为相机在 x 方向和 y 方向上的视野范围。

为了计算 u 和 v , 本文使用了人工神经网络, 该网络是在 Alex-Net 的基础上构建的, 在结构和训练方法上做了一些小的改动。

本文为适配相机坐标回归任务, 对预训练的 Alex-Net 网络结构进行了调整: 移除最后两个全连接层, 并替换为一个具有两个输出节点 (对应 x 坐标和 y 坐标) 的全连接层及回归层。所有输入图像均统一缩放至 227×227 像素分辨率, 以满足网络输入要求。

网络训练采用 Adam 优化器, 设置训练周期为 200 轮, 批量大小为 60, 初始学习率为 0.001, 并采用动态学习率衰减策略, 每训练 20 个周期将学习率减半。训练与验证数据来源于国际公开的 UCI 数据集, 其中训练集用于神经网络参数学习, 验证集用于模型性能评估。

为量化模型精度, 针对训练集图像, 计算神经网络预测的 x 坐标、 y 坐标与 Unity 渲染图像时使用的真实坐标之间的误差, 并汇总其均方根误差, 结果记录于表 1 首行。

具体误差详见表 1

表 1 物体原点投影位置坐标 u 和 v 的误差 (对应角度 ϕ_u 和 ϕ_v)

数据集	单位: ($^{\circ}$)	
	ϕ_u	ϕ_v
训练集	0.218	0.235
验证集	0.533	0.481

针对验证集图像, 同样通过计算神经网络预测的角度 (ϕ_u 和 ϕ_v) 与 Unity 渲染所用的真实角度之间的偏差, 得到所有误差值, 其均方根误差记录于表 1 第 2 行。至此, 第 1 阶段目标—在大尺度图像中精准定位物体原点的投影位置已成功实现。

b) 将物体原点投影平移至图像中心的单应性变换

在找到坐标 u 和坐标 v 之后, 可以进行单应性变换, 该变换对应于相机在水平方向上旋转角

度 ϕ_u 以及在垂直方向上旋转角度 ϕ_v 。经过这种变换后, 物体原点的投影会移动到图像的中心。其中相机的投影过程如图 2 所示。

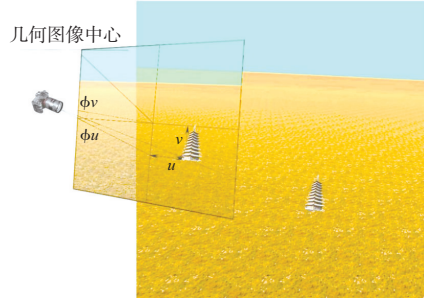


图 2 相机投影过程的坐标变换

假设焦距等于 1, 本文可以得到 ϕ_u 和 ϕ_v 的值

$$\cos \phi_v = \sqrt{1+u^2} / \sqrt{1+u^2+v^2} \quad (4)$$

$$\sin \phi_v = v / \sqrt{1+u^2+v^2} \quad (5)$$

最后, 得到了两个旋转矩阵 R_u 和 R_v , 它们描述了单应性变换 H (也就是将齐次坐标为 $(u_0, v_0, 1)$ 的点进行线性变换, 移动到点 $(u_1, v_1, 1)$), 使得物体原点投影点 $(u, v, 1)$ 移动到点 $(0, 0, 1)$:

$$(0, 0, 1) = (u, v, 1) \cdot H \quad (6)$$

$$H = R_v R_u \quad (7)$$

$$R_v = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi_v & -\sin \phi_v \\ 0 & \sin \phi_v & \cos \phi_v \end{bmatrix} \quad (8)$$

$$R_u = \begin{bmatrix} \cos \phi_u & 0 & \sin \phi_u \\ 0 & 1 & 0 \\ -\sin \phi_u & 0 & \cos \phi_u \end{bmatrix} \quad (9)$$

在本文定义的坐标系中, OX 轴与图像宽度方向一致, OY 轴与图像高度方向一致。旋转矩阵 H 由两次连续旋转变换构成: 首先绕 OX 轴旋转一定角度, 随后绕 OY 轴旋转另一角度。

为保持投影一致性, 选取点 $(0, 0, 1)$ 附近一个 227×227 像素的邻域作为处理区域, 该点对应于相机旋转后物体原点在图像中的投影位置。通过将该邻域中每一点的齐次坐标与矩阵 H 相乘, 可借助单应性变换回溯至原始图像中的对应位置, 并获取其颜色信息。最终, 通过这一变换过程生成一组新的图像序列, 其中每一幅图像的光轴均精确指向物体原点, 且该原点投影严格位于图像中心。

2) 无人机相机定位的第 2 阶段

a) 在新修正图像上确定物体原点投影的坐标
通过单应性变换生成修正图像后, 本文进一

步提取了物体原点投影在图像中的实际坐标位置 u 和坐标位置 v 。这些坐标数据将作为第 2 阶段训练过程中人工神经网络的监督信息,用于提升模型的定位精度。

b) 将相机坐标转换为四元数形式

为规避笛卡儿、柱坐标及球坐标系在描述角度时可能出现的“特殊点”问题(即同一相机位姿对应多种参数表示),本研究最终采用四元数进行角度参数化。在三维空间坐标系中,为精确描述相机相对于目标物体原点的空间位姿关系,可采用一套基于球坐标与自旋参数的表示体系。该体系将相机光学中心与物体原点之间的空间关系分解为以下四个基本参数:首先是二者之间的欧氏距离 r ,用以确定相机在空间中的径向位置;其次是由该连线向量所定义的两个角度参量——方位角 φ 用于表征该向量在水平面上的投影与参考方向之间的夹角,天顶角 θ 则描述该向量偏离垂直方向的程度;最后还需引入一个绕该连线轴旋转的相机自旋角 ω ,用于表达相机光学轴相对于该基准连线的内部旋转状态。上述参数共同完整定义了相机在球坐标系下的外部位姿,其对应关系与几何含义如图 3 所示。

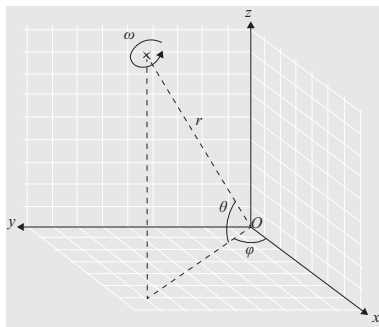


图 3 相机在球坐标系下的外部位姿关系图

c) 最终位置的确定

第 2 阶段构建了一个结构近似的改进型 Alex-Net 网络。具体网络改进包括:移除最后两个全连接层,代之以一个输出维度为 7 的全连接层及回归层。该网络的输出目标包括相机到物体原点的距离 d 、相机朝向四元数的四个分量 W_q , X_q , Y_q , Z_q , 以及物体原点在图像上的投影坐标 v_{gom} 和 u_{gom} 。鉴于距离 d 的数值范围(200~1 000)远大于其他输出(0~1),本文对 v_{gom} 和 u_{gom} 进行了尺度缩放以平衡量级。模型训练采用 Adam 优化器,迭代 200 轮,批大小为 60,初始学习率设为 0.001 并每 20 轮衰减 50%。

3) 训练结果

本研究采用训练集与验证集分别对预训练人工神经网络进行训练与性能评估。对于训练集,通过计算神经网络预测的位置及朝向参数与 Unity 环境生成图像时所用的真实位姿参数之间的偏差,得到各项误差指标,其均方根值如表 2 第 1 行所示。在验证集上,采用相同方法评估训练后模型的泛化能力,通过对比模型输出与真实位姿的差异,计算出相应的误差均方根值,结果记录于表 2 第 2 行,以客观反映模型在未知数据上的位姿估计精度。

表 2 无人机相机的坐标和旋转角度误差

数据集	σ_r/m	σ_w/m	σ_x/m	σ_y/m	σ_z/m	$\sigma_{u_{\text{gom}}}^\circ/^\circ$	$\sigma_{v_{\text{gom}}}^\circ/^\circ$
训练集	3.2	3.1	3.9	3.1	3.2	0.08	0.08
验证集	4.7	6.5	13.7	16.2	16.3	0.29	0.41

实验结果表明(如表 2 所示),在方法实施的第 2 阶段,系统达到了较高的位姿估计精度。其中,相机空间定位误差不超过 3 m,朝向估计误差在 10% 以内,体现出该方法在三维空间定位与姿态解算方面的有效性与实用性。

2 结语

本文提出了一种融合无人机倾斜摄影与深度学习的三维影像处理框架。该方法以无人机采集的高分辨率多视角影像为输入,借助深度学习技术实现影像的智能预处理、高精度配准与无缝拼接,并基于神经网络驱动的点云生成与优化算法,重建出细节丰富、结构完整的三维模型。实验表明,本方法不仅显著提高了三维模型的整体重建精度,尤其在地物边缘和复杂结构处方面表现出良好的完整性,且有效缓解了传统方法中常见的边缘模糊与结构缺失问题,该研究为三维场景中目标位姿估计与精细建模提供了新的技术路径,在提升模型可用性与自动化程度方面具有明显优势。

参考文献:

- [1] 滕威. GPS 物流车辆导航系统中地图匹配算法的实现[J]. 微型电脑应用, 2021, 37(9): 123-125.
- [2] 杨逸夫, 叶晓芬. 地面信标与惯性导航系统结合的无人机定位技术研究[J]. 数字通信世界, 2025(2): 21-23.
- [3] 陶磊, 郝晓丽, 王松涛, 等. 基于相位相关和 Harris 纹理分类的 SIFT 图像拼接算法[J]. 计算技术与自动

- 化,2025,44(1):12-18.
- [4] 崔建红,李银兵,赵海茹,等. 基于 Lucas-Kanade 算法的目标跟踪研究[J]. 电脑知识与技术,2023,19(31):25-27.
- [5] 吴岸平,鲁亚飞,郭正,等. 采用改进 Sage-Husa 自适应滤波的运动目标三维定位方法[J]. 国防科技大学学报,2023,45(2):146-154.
- [6] Kendall A, Cipolla R. Geometric loss functions for camera pose regression with deep learning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6555-6564.
- [7] Walch F, Hazirbas C, Leal-Taixé L, et al. Image-based localization using LSTMs for structured feature correlation[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 627-637.
- [8] Clark R, Wang S, Markham A, et al. VidLoc: A deep spatio-temporal model for 6-DoF video-clip relocalization[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2652-2660.
- [9] Melekhov I, Ylioinas J, Kannala J, et al. Image-based localization using hourglass networks[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Venice, Italy: IEEE, 2017: 870-877.
- [10] Brahmabhatt S, Gu J W, Kim K, et al. Geometry-aware learning of maps for camera localization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 2616-2625.
- [11] 史亚锋,韩可伟,高树论,等. 基于深度学习的轻量化无人机航拍小目标智能检测系统设计[J]. 计算机测量与控制,2025,33(9):100-108.
- [12] 宋留斌,郭道省. 多跳无人机自组网接入控制协议: 深度强化学习时隙分配方法[J]. 电子与信息学报,2025,47(5):1356-1367.
- [13] 易磊,黄哲玮,陆韬宇. 基于 SLAM 与深度学习融合的室内无人机系统[J]. 现代电子技术,2025,48(9):154-160.
- [14] 陈明强,张勇,刘俊杰,等. 基于深度学习特征匹配的无人机景象匹配导航[J]. 电光与控制,2025,32(5):60-66.
- [15] 李晓信,刘志宏,王冠政,等. 基于深度强化学习的四旋翼无人机双向推力控制[J]. 机器人,2025,47(3):305-314.
- [16] 刘伯凯,殷雪峰,孙传昱,等. 基于深度强化学习的无人机三维场景导航方法研究[J]. 图学学报,2025,46(5):1010-1017.
- [17] 魏潇龙,吴亚荣,姚登凯,等. 基于深度强化学习的无人机空战机动分层决策算法[J]. 系统工程与电子技术,2025,47(9):2993-3003.
- [18] 周翰祺,方东旭,张宁波,等. 基于深度学习的多无人机多目标跟踪[J]. 计算机工程,2025,51(4):57-65.
- [19] 郭天昊,尹瑞龙,李仙钟,等. 一种基于可移动天线和深度强化学习的无人机中继通信方案[J]. 移动通信,2025,49(4):96-103.
- [20] 方思雨,张振华,牛经龙,等. 多无人机空战机动决策中的深度强化学习:实践中的关键技术综述与未来展望[J]. 航空兵器,2025,32(2):19-33.
- [21] 顾同成,徐东伟,孙成巨. 无人驾驶深度强化学习决策模型性能评测方法综述[J]. 计算机工程与应用,2025,61(19):12-42.
- [22] 孔繁辉,姜斌. 基于深度强化学习的多无人机协同配送路径组合优化研究[J]. 科技管理研究,2025,45(7):194-206.
- [23] 赵军,何家政,孙冰寒,等. 基于深度强化学习的四旋翼无人机姿态控制[J]. 中国惯性技术学报,2025,33(3):284-292,300.

收稿日期:2025-05-14