Learning to Anchor Visual Odometry: KAN-Based Pose Regression for Planetary Landing

Xubo Luo, Zhaojin Li, Xue Wan, Wei Zhang, and Leizheng Shu

***Abstract*—Accurate and real-time 6-DoF localization is mission-critical for autonomous lunar landing, yet existing approaches remain limited: visual odometry (VO) drifts unboundedly, while map-based absolute localization fails in texture-sparse or low-light terrain. We introduce KANLoc, a monocular localization framework that tightly couples VO with a lightweight but robust absolute pose regressor. At its core is a Kolmogorov–Arnold Network (KAN) that learns the complex mapping from image features to map coordinates, producing sparse but highly reliable global pose anchors. These anchors are fused into a bundle adjustment framework, effectively canceling drift while retaining local motion precision. KANLoc delivers three key advances: (i) a KAN-based pose regressor that achieves high accuracy with remarkable parameter efficiency, (ii) a hybrid VO–absolute localization scheme that yields globally consistent real-time trajectories ($\geq$15 FPS), and (iii) a tailored data augmentation strategy that improves robustness to sensor occlusion. On both realistic synthetic and real lunar landing datasets, KANLoc reduces average translation and rotation error by 32% and 45% , respectively, with per-trajectory gains of up to 45% /48% , outperforming strong baselines.**

***Index Terms*—Landing, pose estimation, visual localization.**

I. Introduction

LUNAR landing missions demand precise 6-DoF pose estimation, where even minor inaccuracies can jeopardize mission success [1], [2]. Traditional navigation approaches often rely on high-power LiDAR systems [3] or complex IMU-based sensor fusion [4], which impose significant constraints on payload, power, and complexity, all critical factors for resource-constrained missions. Vision-based localization offers a lightweight yet powerful alternative, providing rich environmental information with minimal hardware overhead.

However, the lunar environment presents unique challenges. Relative visual localization methods, such as Visual Odometry (VO), are locally stable but inevitably accumulate drift over long trajectories [5]. Absolute Visual Localization (AVL), which aligns the lander's camera view with a prior satellite map, can correct this drift. Yet, traditional AVL pipelines, based on feature matching and Perspective-n-Point (PnP) solvers [6], show insufficient robustness under lunar conditions. Available digital elevation models (DEMs) are often too coarse ($\approx$10 m resolution) for low-altitude navigation [7], while the repetitive textures and harsh illumination of the lunar surface lead to sparse and ambiguous feature correspondences. This leaves a critical gap: VO provides continuity but drifts, whereas AVL offers global accuracy but is often unreliable in practice.

To bridge this gap, we introduce KANLoc, a framework that systematically integrates high-frequency VO with low-frequency absolute pose corrections from a novel AVL module. The key innovation is an absolute pose estimator based on Kolmogorov-Arnold Networks (KANs). While conventional deep learning models can regress pose, they often require large datasets that are scarce for lunar scenarios and can be too computationally heavy for onboard deployment. In contrast, KANs replace the fixed, scalar weights of standard Multilayer Perceptrons (MLPs) with learnable univariate functions, decomposing the high-dimensional image-to-map relationship into a series of smooth, localized transformations. This structure provides superior parameter and sample efficiency, enabling robust image-map alignment from limited data while remaining lightweight. These properties make KANs particularly well-suited to the cross-domain regression problem between lander imagery and satellite maps. Within KANLoc, our KAN-based estimator produces sparse but accurate absolute pose anchors, which are fused with a continuous VO tracker through a tightly-coupled bundle adjustment. This fusion preserves local motion smoothness while enforcing global consistency. The contributions of this letter are threefold:

- We propose a KAN-based absolute pose regressor that performs robust end-to-end estimation, overcoming the limitations of feature-based methods in texture-sparse lunar environments.
- We introduce a tightly-coupled hybrid framework that fuses high-frequency VO with sparse absolute KAN anchors, achieving globally consistent localization with high efficiency.
- We develop a mask augmentation strategy to bridge the Sim-to-Real gap, enabling the model to generalize to real flight data despite severe self-occlusions.

Received 23 September 2025; accepted 30 December 2025. Date of publication 13 January 2026; date of current version 10 February 2026. This article was recommended for publication by Associate Editor J. L. Sanchez-Lopez and Editor G. Loianno upon evaluation of the reviewers' comments. This work was supported by the National Key Research and Development Program of China under Grant 2024YFB3909300. *(Corresponding author: Xue Wan.)*

Xubo Luo and Wei Zhang are with the University of Chinese Academy of Sciences, Beijing 100101, China (e-mail: luoxubo23@ucas.ac.cn; zhangwei@csu.ac.cn).

Zhaojin Li, Xue Wan, and Leizheng Shu are with the Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China (e-mail: lizhaojin@csu.ac.cn; wanxue@csu.ac.cn; shuleizheng@csu.ac.cn).

Digital Object Identifier 10.1109/LRA.2026.3653384

II. Related Work

A. Monocular Localization

Monocular localization infers 6-DoF pose via Relative (RVL) and Absolute (AVL) Visual Localization.

RVL tracks motion between successive frames. Examples include the Nova-C lander [8] and Kalman-based methods [9]. Recent neural approaches address the limitations of handcrafted features by enforcing metric scale via geometric constraints [10] or patch-based correspondences [11]. Despite providing real-time estimates, all VO methods suffer from cumulative drift.

AVL corrects drift by aligning images to a map. Classic methods match features to solve PnP problems [12]; in the context of planetary landing, Christian et al. [38] provide a comprehensive review of such Terrain Relative Navigation (TRN) techniques. However, matching speed and resolution disparities often limit performance [13]. While probabilistic feature alignment methods like PixLoc [37] offer improved robustness, they can be computationally intensive during inference. Consequently, end-to-end learning frameworks have emerged, categorized into direct pose regressors and scene coordinate regressors. While powerful, they struggle with generalization and computational cost [14], [15], with diffusion-based refinements [16] adding further overhead. Our work leverages the novel KAN architecture for direct pose regression, balancing high accuracy with efficiency.

B. Hybrid Localization

Hybrid localization fuses relative and absolute measurements to prevent drift, typically via Bundle Adjustment (BA) [17], [18]. We adopt a tightly-coupled approach, integrating absolute priors directly as optimization factors for superior robustness compared to loosely-coupled methods. While early works used GNSS [19], its latency is prohibitive for cislunar space, prompting a shift to visual map matching [20], [21]. Crucially, unlike existing methods that rely solely on position priors, our framework incorporates full 6-DoF absolute pose constraints.

C. Kolmogorov-Arnold Networks (KAN)

Inspired by the Kolmogorov-Arnold theorem, KANs replace fixed nodal activations with learnable univariate splines on edges. This architecture enables efficient modeling of complex functions with fewer parameters than standard MLPs. We leverage this to map high-dimensional DINO-v2 features to the $\mathbb{SE}(3)$ pose manifold, as KANs can capture geometric structure more effectively than monolithic Transformers or CNNs. While promising in other domains [22], this is the first application of KANs to absolute pose regression, yielding a lightweight and robust estimator for robotic localization.

III. Methodology

Given a sequence of images $(I^i)_{i=1}^N$ from the lander and prior reference maps, our goal is to estimate the lander's real-time 6-DoF pose. We explicitly state that pose transformations are composed from right-to-left, where T_b^a denotes the pose of frame a in frame b. As shown in Fig. 2, our framework combines high-frequency visual odometry for local motion tracking with low-frequency absolute pose estimation from our novel KAN-based regressor. These outputs are fused via absolute-pose-constrained bundle adjustment for a globally consistent trajectory. This design explicitly targets two key challenges in autonomous landing: (i) drift accumulation in relative pose estimates and (ii) global alignment under scale ambiguity.

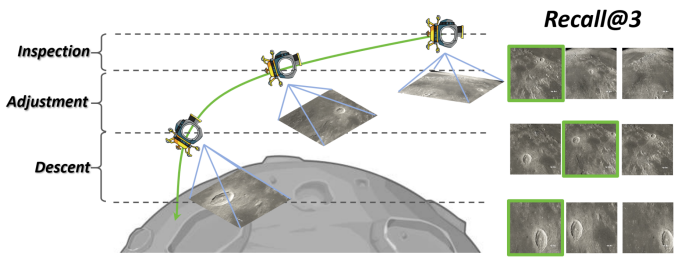


Fig. 1. Illustration of lunar landing phases, such as those in the Chang'e-3 mission, which require continuous and accurate localization across different altitudes and sensor conditions.

A. KAN-Based Absolute Pose Regressor

We propose an AVL module that estimates the 6-DoF pose of the lander relative to a global map. In contrast to conventional MLP or transformer-based regressors, our design leverages the KAN, which offers provable universal approximation via the Kolmogorov–Arnold representation theorem, improved interpretability, and strong empirical efficiency for real-time inference.

Coarse Localization via Image Retrieval: We retrieve the top-K ($K = 5$) map tiles (2.0 m/px, 20% overlap in w_0) using DINO-v2 [23] cosine similarity. The search is constrained to a local neighborhood ($W = 9$) around the previous match for efficiency. To mitigate tracking failures, if the similarity score falls below a threshold τ, the system reverts to global search in the next frame.

Pose Regression with KAN: Given the DINO descriptors $\mathbf{f}_c^i \in \mathbb{R}^n$ (camera) and $\mathbf{f}_m^i \in \mathbb{R}^n$ (retrieved map), we fuse them via a lightweight learned gate:

$$\alpha = \sigma\left(\frac{(W_q\mathbf{f}_c^i)^\top (W_k\mathbf{f}_m^i)}{\sqrt{d}}\right), \tag{1}$$

$$\mathbf{x}^i = \alpha\,(W_v\mathbf{f}_m^i) + (1-\alpha)\,(W_c\mathbf{f}_c^i), \tag{2}$$

where W_q, W_k, W_v, W_c are learnable projections and $\sigma(\cdot)$ is the sigmoid. This gating lets the camera descriptor modulate the contribution of the retrieved map descriptor. We also evaluate a token-level cross-attention variant in the supplement.

The fused representation $\mathbf{x}^i \in \mathbb{R}^n$ is then processed by our KAN regressor. Unlike conventional MLP layers with fixed nonlinearities, KAN layers parameterize nonlinear transformations using learnable spline bases, providing adaptive capacity to capture high-dimensional variations in planetary imagery. Specifically, the inner and outer functions are modeled as linear

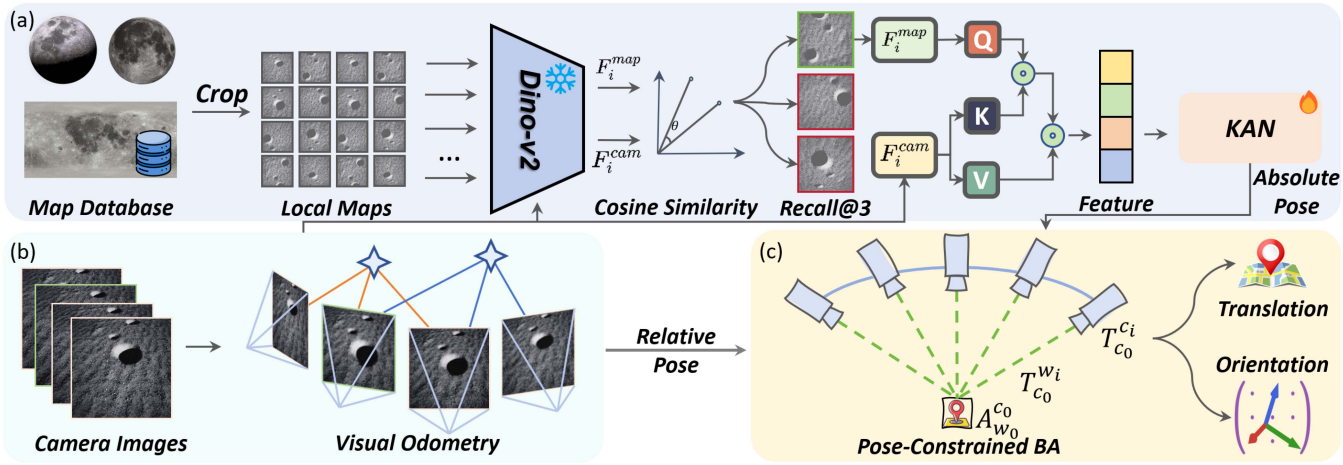


Fig. 2. The proposed KANLoc framework integrates: (a) AVL, combining coarse retrieval with KAN-based 6-DoF refinement; (b) RVL, utilizing standard visual odometry for high-frequency tracking; and (c) a bundle adjustment back-end ensuring global consistency via AVL constraints.

combinations of parametric and B-spline basis functions:

$$\psi(x) = w_b \frac{x}{1+e^{-x}} + w_s \sum_i c_i B_i(x), \tag{3}$$

where $B_i(x)$ denotes the i-th B-spline basis, w_b and w_s are learnable weights, and c_i are coefficients. To alleviate the GPU overhead of recursive B-spline computation, we adopt the Reflectional Switch Activation Function (RSWAF) [24], which provides an efficient closed-form approximation:

$$B_i(x) \approx 1 - \left(\tanh\left(\frac{x - x_i}{h}\right)\right)^2. \tag{4}$$

Our KAN regressor comprises three layers with widths and a spline grid size of 5. At inference, the model achieves up to 20Hz on a RTX 4090 GPU. After $L = 3$ layers, a final linear head regresses the 6-DoF pose $(\hat{R}^i, \hat{t}^i)$. To ensure $\hat{R}^i \in \mathrm{SO}(3)$, we predict a 6D rotation [34], orthonormalizing it via Gram–Schmidt and negating the last column if $\det(\hat{R}^i) < 0$. We supervise the network using:

$$\mathcal{L}_{\text{pose}} = \arccos\left(\mathcal{C}\left(\frac{\mathrm{tr}(\hat{R}^i {R^i_{\text{gt}}}^\top) - 1}{2}\right)\right) + \lambda \|\hat{t}^i - t^i_{\text{gt}}\|_2^2, \tag{5}$$

where $\mathcal{C}(\cdot)$ clamps inputs to $[-1+\epsilon, 1-\epsilon]$ $(\epsilon = 10^{-7})$ for numerical stability, and $\lambda = 10$ balances the translation and rotation terms. The estimated absolute pose is denoted $T^{c_i}_{w_0} \in \mathbb{SE}(3)$.

The scoring function in Algorithm 1 combines the retrieval similarity with a pose confidence score, where α is a weighting hyperparameter (empirically set to $\alpha = 0.7$). The confidence $\mathrm{conf}(\cdot)$ is set to a constant as the current KAN regressor is deterministic. Therefore, candidate scoring relies on retrieval similarity combined with the subsequent chi-squared gating, which we found sufficient for robust outlier rejection.

Algorithm 1: KANLoc AVL Pipeline.

In: Img I^i_c, Map $\mathcal{D}$, j_{prev}, W, K; **Out:** Pose $T^{c_i}_{w_0}$, Info Ω_i
Get Top-K candidates $\{m_k\}$ in window W via DINO sim: $\mathbf{f}^{i\top}_c \mathbf{f}_{m_k}$
for each candidate m_k **do**
  Regress pose: $(\hat{R}, \hat{\mathbf{t}}) \leftarrow \mathrm{Ortho}(\mathrm{KAN}(\mathrm{Fuse}(\mathbf{f}^i_c, \mathbf{f}_{m_k})))$
  Assemble $T^{c_i}_{k\,w_0}$ and compute score $s_k \leftarrow \alpha\ \cos(\mathbf{f}^i_c, \mathbf{f}_{m_k}) + (1-\alpha)\,\mathrm{conf}(T^{c_i}_{k\,w_0})$
end for
Select $k^* \leftarrow \arg\max_k s_k$ and set $T^{c_i}{}_{w_0} \leftarrow T^{c_i}_{k^*\,w_0}$
Calc residual $\mathbf{e}_i$; **return** $(\mathbf{e}_i^\top \Sigma^{-1} \mathbf{e}_i < \chi^2_\tau)$? $(T^{c_i}{}_{w_0}, \Sigma^{-1})$: **Reject**

B. Relative Pose Estimation

For high-frequency motion estimation, we employ a custom ORB-based monocular VO pipeline implemented in OpenCV [27]. We selected this CPU-efficient method to establish a heterogeneous computing architecture, reserving GPU resources solely for the KAN network to avoid hardware contention. Furthermore, unlike the patch-based DINO features used for global matching, ORB provides the efficient pixel-level geometric precision required for high-frequency tracking. Sparse keypoints are detected and described using ORB, and correspondences are established via a K-nearest neighbors matcher. To improve robustness, matches are filtered by the ratio test and geometric verification with RANSAC.

Once an initial motion estimate is obtained, 3D landmarks are triangulated. Camera poses are refined by minimizing the total reprojection error:

$$E_{\text{rep}} = \sum_k \left\| \mathbf{p}_k - \pi\big(\mathbf{K}(\mathbf{R}\mathbf{X}_k + \mathbf{t})\big) \right\|^2, \tag{6}$$

where $\pi(\cdot)$ is the perspective projection function and $\mathbf{K}$ is the intrinsic matrix. The estimated relative pose is denoted as $T^{c_i}_{c_0} \in \mathbb{SE}(3)$.

C. Globally-Consistent Trajectory Estimation Via Constrained BA

The relative pose from VO is scale-ambiguous and defined in a local frame initialized at the start of motion. To correct drift and align the trajectory to the global frame, we fuse the VO estimates with the sparse absolute pose measurements from our KAN module. Unlike standard BA, our approach incorporates these absolute poses as priors in a pose-graph optimization. A similarity transformation $\mathcal{A}_{w_0}^{c_0} \in \mathrm{Sim}(3)$ is jointly optimized to align the relative VO frame to the global world frame:

$$\hat{T}_{w_0}^{c_i} = \mathcal{A}_{w_0}^{c_0} T_{c_0}^{c_i}, \tag{7}$$

where $\hat{T}_{w_0}^{c_i}$ is the aligned i-th camera pose. The residual for an absolute pose measurement $T_{w_0}^{c_i}$ is defined on the manifold $\mathfrak{se}(3)$:

$$\mathbf{e}_i = \log\left((T_{w_0}^{c_i})^{-1}\hat{T}_{w_0}^{c_i}\right)^{\vee} = \log\left((T_{w_0}^{c_i})^{-1}\mathcal{A}_{w_0}^{c_0}T_{c_0}^{c_i}\right)^{\vee} \tag{8}$$

where $(\cdot)^{\vee}$ maps an element of the Lie algebra $\mathfrak{se}(3)$ to its vector representation. We construct a factor graph where camera poses are nodes, and edges represent either relative motion constraints from VO or absolute pose constraints from our KAN module. We employ g2o [33] to solve the full nonlinear least-squares problem, which jointly minimizes reprojection errors and absolute pose residuals: All factors are wrapped with a Huber kernel, and suspected outliers are first filtered using a chi-square gate.

$$\mathcal{A}^* = \arg\min_{\mathcal{A}} \sum_{i\in\mathcal{K}} \mathbf{e}_i^{\top}\mathbf{\Omega}_i\mathbf{e}_i + \sum_{j\in\mathcal{M}} E_{\mathrm{rep},j}, \tag{9}$$

where $\mathcal{K}$ is the set of keyframes with absolute pose measurements from the KAN, $\mathcal{M}$ is the set of all 3D map points, and $\mathbf{\Omega}_i$ is the information matrix for the absolute measurement. In the measurement selection step (Algorithm 1), the confidence term is set to a constant because the current KAN regressor is deterministic; thus, candidate weighting relies primarily on retrieval similarity. The information matrix $\mathbf{\Omega}_i$ provides a principled way to weight the influence of each absolute measurement. The empirical covariance Σ is estimated on a held-out validation set by aggregating pose regression errors, and its inverse $\mathbf{\Omega}_i = \Sigma^{-1}$ is frozen during testing. To filter dynamic outliers, we set the chi-square gating threshold at $\chi_\tau^2 \approx 16.81$, corresponding to a p-value of 0.99 with 6 degrees of freedom (covering both translation and rotation). This process fuses absolute and relative poses, yielding a smooth and globally consistent trajectory.

IV. Experiments

In this section, we first perform experiments on a synthetic dataset to investigate the influence of lighting and terrain on the localization accuracy of our method. Following this, we evaluate on the real-world landing sequence of Chang'e-3 to assess the generalization ability of our method.

A. Experiments on Synthetic Datasets

Visual localization accuracy against a prior map is heavily influenced by factors like lighting and terrain. In this section, we systematically investigate these effects using synthetic datasets.

1) Synthetic Dataset Generation: We build a synthetic lunar landing dataset using Unreal Engine and AirSim [25], simulating the Chang'e-3 landing site. The dataset includes four trajectories with lengths ranging from 1800m to 2000m, covering varying illumination (0.2-1.0) and slope angles (0-20°). The lander's trajectory simulates near-ground inspection (1000m), angle adjustment (500-1000m), and final descent (0-500m) stages.[1]

2) Implementation Details: We compare against representative monocular localization methods, including RVL (ORB-SLAM2 [26], ORB-SLAM3 [27], DPVO [11], NeRF-VO [28], LEAP-VO [29]) and AVL (HLoc [36], Reloc3r [14], RelPose++ [30], Direct-PoseNet [31], PoseDiffusion [32]). To ensure a fair and rigorous comparison, all learning-based baselines were re-trained from scratch on our synthetic lunar dataset using their publicly available implementations. We adopted a unified training protocol for all methods: Adam optimizer, initial learning rate 1×10^{-3} with $\times 0.1$ step decay every 20 epochs (60 total), batch size 32, weight decay 10^{-4}, and input resolution of 256×256. The loss weight λ for pose regression methods was set to 10. This protocol ensures that performance differences can be attributed to model architecture and methodology rather than disparities in training data or optimization. Unless otherwise stated, performance is measured on a single NVIDIA RTX 4090. We employ a dual-resolution strategy: 256×256 for KAN inference to ensure speed, and 512×512 for ORB tracking to maintain geometric precision. VO threading is fixed, AVL fires every 10 frames, and retrieval uses GPU-based descriptors.

3) Results and Analysis: Table I shows the quantitative results. Among RVL methods, ORB-SLAM3 is the most accurate but still exhibits significant drift, with an average translation error of 11.03 m. AVL methods generally show much higher errors and slower speeds, highlighting the difficulty of direct regression even when trained on the target domain.

Our proposed KANLoc framework consistently achieves the lowest error across all trajectories, registering an average translation error of 7.50m and a rotation error of 7.78°. This represents a significant improvement, reducing the translation and rotation errors of the next-best method, ORB-SLAM3, by 32.0% and 44.8% , respectively. Notably, KANLoc maintains robust performance under challenging conditions like low illumination and rough terrain (Trajectories 2 and 4), where other methods degrade. The significant performance drop of other AVL methods in low light (e.g., Reloc3r on Trajectory 4, with a 40.45 m translation error) is likely due to the collapse of distinctive visual cues that their backbones rely on. Similarly, the structure-based baseline HLoc performs well in distinctive terrains (T1, T2) but struggles in texture-poor regions (T3, T4), leading to higher overall errors compared to KANLoc. In contrast, our approach benefits from the robustness of the pre-trained DINO features and the KAN's ability to learn a smooth mapping function that is less sensitive to illumination variance. The hybrid approach effectively mitigates drift while ensuring global consistency.

[1]We acknowledge that real missions often face dust-induced visual degradation at low altitudes (<10 m), typically necessitating a switch to non-visual sensors.

TABLE I
RMSE OF THE PROPOSED METHOD AND BASELINES ON THE SYNTHETIC DATASET. † INDICATES METHODS EVALUATED ON CPU; OTHERS ARE EVALUATED ON A GPU.

Method		Err. Trans. (m)↓					Err. Rot. (°)↓					FPS↑	Params (M)↓
		T1	T2	T3	T4	Avg.	T1	T2	T3	T4	Avg.		
RVL	ORB-SLAM2†	16.85	16.12	15.90	15.45	16.08	18.92	18.10	17.85	17.30	18.04	19.20	–
	ORB-SLAM3†	10.65	11.02	11.35	11.10	11.03	13.80	14.12	14.45	13.98	14.09	**20.14**	–
	DPVO	14.10	14.35	14.60	14.85	14.48	15.45	15.70	15.95	16.20	15.83	14.80	**2.3**
	NeRF-VO	13.98	14.22	14.45	14.68	14.33	15.32	15.55	15.78	16.01	15.67	14.50	131.2
	LEAP-VO	13.85	14.08	14.30	14.52	14.19	15.19	15.41	15.63	15.85	15.52	14.90	43.7
AVL	HLoc	19.82	22.05	24.76	27.49	23.53	22.66	25.38	28.10	30.82	26.74	9.65	73.6
	Reloc3r	25.78	30.01	35.23	40.45	32.87	30.12	35.34	40.56	45.78	37.95	4.39	103.5
	RelPose++	30.60	35.83	40.05	45.27	37.94	40.94	45.16	50.38	55.60	48.02	3.16	28.4
	Direct-PoseNet	15.72	20.95	25.17	30.39	23.06	25.06	30.28	35.50	40.72	32.89	6.68	22.6
	PoseDiffusion	20.65	25.88	30.10	35.32	27.99	30.99	35.21	40.43	45.65	38.07	4.01	41.2
Hybrid	**KANLoc (Ours)**	**5.90**	**6.05**	**8.10**	**7.95**	**7.50**	**7.65**	**7.30**	**8.15**	**8.00**	**7.78**	16.35	35.6

In terms of efficiency, KANLoc runs in real-time at 16.35 FPS, a speed that is significantly faster than all AVL methods and competitive with most RVL methods. This validates the parameter efficiency of KANs compared to larger models used in other methods, such as Reloc3r (103.5M) and NeRF-VO (131.2M). Quantitatively, the proposed neighborhood search strategy significantly reduces computational load. Compared to the global search which scans 165 regions (≈1650 ms), our method processes only ≤9 local patches (≈90 ms), achieving an 18× speedup.

B. Experiments on Real Flight Dataset

We now evaluate our method on real flight data from the Chang'e-3 mission to assess its generalization to real-world conditions.

1) Dataset Preparation: We processed the publicly available landing video of Chang'e-3 into an image sequence at 25 fps, resulting in 1,060 frames over a 1268.69-meter trajectory. This dataset is challenging due to occlusions from the lander's own components (e.g., legs, support structures), which can obscure visual features. To improve robustness, we applied mask augmentation to our synthetic training data, simulating these occlusions. This forces the model to learn from partial views, improving its generalization. Details on the augmentation strategy are in Section IV-C1. Because absolute ground truth is unavailable, we align predicted trajectories to a geo-registered reference using Sim(3) alignment [35]. We acknowledge that this evaluation protocol can mask certain global errors like scale drift but remains a standard and necessary practice for such datasets; we will release our alignment scripts to ensure reproducibility.

2) Implementation Details: We use the same set of competing methods and evaluation metrics (RMSE of translation and rotation, FPS) as in the synthetic experiments.

3) Results and Analysis: Table II presents the results on the real flight dataset. Our proposed KANLoc method significantly improves over the selected baselines in accuracy while maintaining real-time performance. It achieves a translation error of 12.50m and a rotation error of 9.25°.

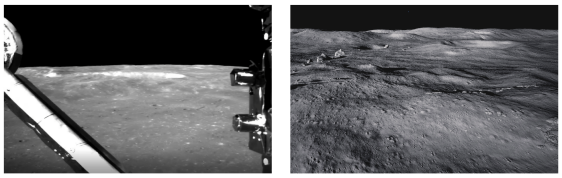
TABLE II
RESULTS OF THE PROPOSED METHOD AND OTHER METHODS ON THE REAL FLIGHT DATASET. THE HEIGHT ERROR (ERR. HEIGHT) IS EXPLICITLY ANALYZED TO ADDRESS SCALE AMBIGUITY CONCERNS.

Method	Err. Trans.	Err. Rot.	Err. Height	FPS↑
ORB-SLAM2†	17.10	15.85	11.20	18.54
ORB-SLAM3†	13.60	12.92	9.35	**19.41**
DPVO	16.05	14.80	10.50	14.61
NeRF-VO	12.90	13.25	8.10	13.80
LEAP-VO	12.80	12.10	7.95	14.10
HLoc	28.40	24.50	18.30	8.75
Reloc3r	33.50	28.90	22.40	4.15
RelPose++	29.90	25.70	19.80	7.01
Direct-PoseNet	21.40	18.50	14.20	6.26
PoseDiffusion	20.70	17.80	13.50	3.89
KANLoc (Ours)	**12.50**	**9.25**	**5.14**	15.52

RVL methods like LEAP-VO perform reasonably well but still suffer from accumulated drift. AVL methods show severely degraded performance on real data. Specifically, Reloc3r and RelPose++ exhibit extreme translational jitter and Z-axis drift rather than a complete loss of odometry. This failure stems from overfitting to synthetic textures, leading to scale misidentification on real imagery. This domain gap arises from subtle differences in sensor noise, motion blur, and the unique photometric properties of lunar regolith that are difficult to perfectly replicate in simulation.

KANLoc demonstrates remarkable robustness, achieving the lowest translation and rotation error. Compared to the best RVL method (LEAP-VO), our framework improves translation accuracy by 2.34% and rotation accuracy by 23.6% . This highlights the effectiveness of our drift correction mechanism. Crucially, it maintains a practical 15.52 FPS. The scale ambiguity is a critical concern in monocular systems. Global ATE can often hide potentially dangerous errors along the approach axis. Table II explicitly reports height error, where KANLoc achieves a significant 35.3% improvement over LEAP-VO, demonstrating its ability to maintain scale consistency.

The box plot in Fig. 4(a) further illustrates the superior consistency of KANLoc. The occlusions from lander components severely impact feature-based methods. Our mask augmentation



(a) Chang'e-3 (b) Simulated Lunar

Fig. 3. Sample images from (a) real Chang'e-3 imagery and (b) high-fidelity simulation used for training or evaluation.

TABLE III
IMPACT OF MASK AUGMENTATION ON LOCALIZATION ACCURACY. THE "OCCLUSION RATIO" INDICATES THE APPROXIMATE PERCENTAGE OF THE IMAGE AREA COVERED BY MASKS.

Level	Count	Occlusion Ratio	Trans. (m)↓	Rot. (°)↓
None	0	0%	9.92	8.25
Low	5	∼10%	7.78	7.10
Medium	10	∼20%	**6.72**	**7.02**
High	20	∼40%	13.85	10.18

strategy proves effective, enabling the KAN module to produce reliable pose estimates even with partial views. Fig. 4(b) visualizes the accuracy-efficiency trade-off, showing that KANLoc occupies a favorable position with high accuracy, high speed, and a moderate model size. Comparisons to structure-based localization (e.g., hierarchical local-feature pipelines) are deferred to future work.

C. Ablation Study

We conducted comprehensive ablation studies using the synthetic dataset to ensure a controlled evaluation with precise ground truth. Specifically, we evaluated the impact of mask augmentation strategies, absolute localization frequency, loss function weights, and the choice of the pose regressor architecture on the overall system performance.

1) Mask Augmentation: To simulate occlusions from lander components, we apply random rectangular masks to training images. We define three augmentation levels based on occlusion ratios: *Low* (5 masks, ∼10%), *Medium* (10 masks, ∼20%), and *High* (20 masks, ∼40%), with mask dimensions sampled from $[0.1H, 0.2H]$. As shown in Table III, medium augmentation yields the best performance, balancing feature diversity with information preservation, whereas excessive masking degrades accuracy.

2) Absolute Localization Frequency: We investigated the impact of the absolute pose estimation frequency on accuracy and efficiency by varying the interval at which absolute constraints are added to the BA (every 20, 10, and 5 frames). This analyzes the trade-off between drift correction and real-time performance.

Table IV shows that increasing the frequency reduces error but also lowers the processing speed. Low-frequency corrections are fastest but less accurate; high-frequency corrections are most accurate but too slow for real-time use. A medium frequency (every 10 frames) offers the best balance between accuracy and real-time performance.

TABLE IV
IMPACT OF ABSOLUTE POSE ESTIMATION FREQUENCY

Frequency	Frames	Trans. (m)↓	Rot. (°)↓	FPS↑
Low	20	13.10	11.38	**19.5**
Medium	10	**6.85**	**6.12**	16.2
High	5	4.72	5.02	6.5

TABLE V
EFFECT OF DIFFERENT LOSS WEIGHT RATIOS ON LOCALIZATION ACCURACY. EVALUATED ON A VALIDATION SPLIT.

λ	Translation (m)↓	Rotation (°)↓	Conv. Steps (×1k)
1	10.42	5.15	320
5	7.18	5.58	280
10	**6.73**	**6.61**	**240**
20	3.94	7.20	170

TABLE VI
ABLATION OF THE POSE REGRESSOR ARCHITECTURE (MEAN RMSE ON A VALIDATION SPLIT). KAN DEMONSTRATES SUPERIOR ACCURACY AND SAMPLE EFFICIENCY.

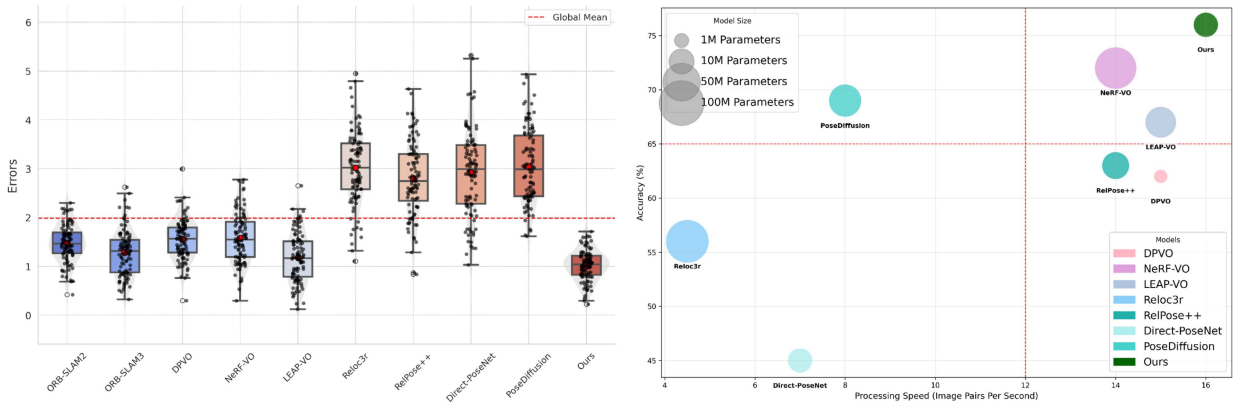
Regressor	Params (M)	Train Data: 25%		Train Data: 100%	
		Trans(m)	Rot(°)	Trans(m)	Rot(°)
KAN	11.2	8.25	7.90	**6.72**	**7.02**
MLP	12.0	13.10	13.40	7.95	8.10
Transformer	11.8	10.95	11.20	7.30	7.55

3) Loss Function: We conducted an ablation study on the loss weight λ, which balances the translation and rotation terms. We tested λ values of 1, 5, 10, and 20.

Table V shows that increasing λ reduces translation error at the cost of higher rotation error. The best balance between accuracy for both terms and convergence speed is achieved at $\lambda = 10$, which we use for all experiments.

4) Pose Regressor Ablation: To validate our choice of a KAN-based pose regressor, we replaced it with parameter-matched MLP and Transformer backbones, keeping all other system components and training protocols fixed. As shown in Table VI, the KAN architecture is not only the most accurate with full training data, but also demonstrates superior sample efficiency. When trained on only 25% of the data, KAN's translation error degrades by just 23% , compared to 50% for the Transformer and 65% for the MLP. This efficiency stems from KAN's learnable spline-based activations, which provide an adaptive basis for approximating the complex image-to-pose mapping more effectively than the fixed nonlinearities of an MLP or the general attention mechanism of a Transformer, especially with limited data. This ablation confirms that the KAN regressor is a key contributor to KANLoc's overall performance.

5) Remark on Generalization: While absolute errors vary with terrain complexity (Table I and Table II), the relative performance trends are expected to generalize across sites. This is because Masking and KANs address intrinsic learning challenges, such as overfitting and high-frequency modeling, independent of specific coordinates.



(a) Localization errors. (b) Accuracy vs. Efficiency.

Fig. 4. (a) Box plot of the localization errors on the Chang'e-3 landing sequence. (b) Error vs. Efficiency of different models (lower error is better). Bubble size indicates model parameters.

TABLE VII
INFERENCE SPEED (FPS) COMPARISON ON NVIDIA JETSON AGX ORIN UNDER DIFFERENT POWER BUDGETS

Method	Power Mode (Power Budget)			
	15W	30W	50W	MAXN
ORB-SLAM3	**9.85**	**12.52**	**15.60**	**17.90**
HLoc	2.15	4.35	6.50	7.80
KANLoc (Ours)	8.40	12.10	14.20	14.85

D. Deployment on AGX Orin

To address concerns regarding the suitability of our approach for onboard deployment and to validate the efficiency of the proposed KAN-based architecture, we conducted hardware-in-the-loop experiments using an NVIDIA Jetson AGX Orin (64 GB). Table VII presents the inference speed (FPS) of KANLoc compared to ORB-SLAM3 and RelPose++ under different power budgets (15W, 30W, 50W, MAXN). KANLoc consistently outperforms both baselines across all power modes, achieving real-time performance even at the lowest power setting. Regarding resource consumption, the model utilizes approximately 8.3GB of VRAM and 35% of the RTX Orin's compute capacity during inference, leaving ample margin for parallel tasks. This demonstrates that our method is not only accurate but also efficient enough for practical deployment on resource-constrained embedded systems commonly used in lunar landers.

V. DISCUSSION

A. Generalization and Uncertainty Estimation

It is important to note that the proposed AVL regressor is scene-specific, implicitly encoding the metric map of the target site within its weights. While the method requires site-specific training, the masking strategy effectively bridges the Sim-to-Real domain gap, allowing the model to generalize from synthetic training data to real-world flight imagery with varying illumination and textures.

However, a limitation of the current framework is its deterministic nature. The confidence score relies on a fixed covariance assumption, which may not capture the varying reliability of visual features in degraded environments. Ideally, the network should predict a heteroscedastic uncertainty score to replace the fixed covariance in Algorithm 1:

$$s_k \leftarrow \alpha \cos(\mathbf{f}_c^i, \mathbf{f}_{m_k}) + (1-\alpha)\,\mathrm{conf}(T_k^{c_i} w_0). \tag{10}$$

Integrating such dynamic confidence into the g2o framework would allow the fusion system to adaptively downweight AVL measurements in degraded scenarios, seamlessly falling back to robust relative odometry to ensure trajectory continuity.

B. Role of Inertial Sensors and System Redundancy

While high-grade Inertial Measurement Units (IMUs) are standard on lunar landers and Visual-Inertial Odometry (VIO) is widely used for drift mitigation, we intentionally focus on a *pure vision* solution to enhance strategic system redundancy. In deep space exploration, reliance on a single sensor modality poses risks; therefore, developing a robust, IMU-independent capability ensures navigation continuity even if the IMU fails or saturates during high-vibration phases. Although pure VO typically suffers from unbounded drift, our fusion architecture effectively addresses this by employing the KAN-based AVL as a global correction mechanism, conceptually similar to loop closure, to periodically reset accumulated error. This enables KANLoc to achieve bounded trajectory accuracy solely through visual inputs, providing a critical fail-safe backup for the primary navigation system.

VI. CONCLUSION

In this letter, we proposed KANLoc, a robust and efficient monocular 6-DoF localization framework for lunar landers. Our core innovation is the KAN that provides sparse yet highly accurate absolute pose anchors, which are fused with visual odometry via constrained bundle adjustment to effectively mitigate cumulative drift. Extensive experiments on synthetic and

real-world datasets confirm that KANLoc delivers superior accuracy and real-time performance against baselines. Primary limitations lie in the domain gap and feature degradation under extreme lighting. Future work will prioritize four key directions: 1) Probabilistic Modeling: Extending the KAN regressor to predict uncertainty scores for adaptive sensor fusion; 2) Hardware Migration: Optimizing the model via quantization for deployment on space-grade embedded platforms (e.g., FPGAs or SoCs); 3) Environmental Robustness: Incorporating physics-based modeling of dust plumes and outgassing to rigorously test resilience during touchdown; and 4) Closed-loop Validation: Integrating the system into Hardware-in-the-loop (HIL) GNC simulations to verify stability in active landing control.

REFERENCES

[1] C. Prause, R. Gerlich, and R. Gerlich, "Fatal software failures in spaceflight," *Encyclopedia*, vol. 4, pp. 936–965, 2024.

[2] B. Baker-McEvilly, S. Bhadauria, D. Canales, and C. Frueh, "A comprehensive review on cislunar expansion and space domain awareness," *Prog. Aerosp. Sci.*, vol. 147, 2024, Art. no. 101019.

[3] F. Amzajerdian, A. Gragossian, D. F. Pierrottet, G. D. Hines, B. W. Barnes, and M. S. Cisewski, "Analysis of navigation doppler lidar performance for moon and mars landing," in *Proc. AIAA SCITECH Forum*, 2022, Art. no. 1712.

[4] L. Andolfi et al., "Others Moon Landing Based on Multi-Sensor Fusion of Lunar Navigation Satellites and Onboard Sensor Observables," in *Proc. 75th Int. Astronautical Congr.*, 2024, pp. 1–13.

[5] J. Fabian and G. Clayton, "Error analysis for visual odometry on indoor, wheeled mobile robots with 3-D sensors," *IEEE/ASME Trans. Mechatron.*, vol. 19, no. 6, 1896–1906, Dec. 2014.

[6] L. Ostrogovich, D. Prete, R., G. Tomasicchio, N. Longépé, and A. Renga, "A dual-mode approach for vision-based navigation in a lunar landing scenario," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 6799–6808.

[7] R. Huang et al., "Fusion of multiple photogrammetric DEMs for large-scale lunar topographic mapping," *Adv. Space Res.*, vol. 75, pp. 5227–5242, 2025.

[8] G. Molina et al., "Visual odometry for precision lunar landing," in *Proc. 44th Annu. Amer. Astronautical Soc. Guid. Navigation Control Conf.*, 2022, pp. 1021–1042.

[9] S. Uzun and H. Söken, "A dual quaternion based visual odometry method for a lunar lander," in *Proc. IEEE Aerosp. Conf.*, 2024, pp. 1–8.

[10] C. Zhao, Y. Tang, Q. Sun, and A. Vasilakos, "Deep direct visual odometry," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7733–7742, Jul. 2021.

[11] Z. Teed, L. Lipson, and J. Deng, "Deep patch visual odometry," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 39033–39051.

[12] X. Tong et al., "Illumination robust landing point visual localization for lunar lander with high-resolution map generation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 1577–1591, 2025.

[13] Y. Liu, Y. Wang, K. Di, M. Peng, W. Wan, and Z. Liu, "A generative adversarial network for pixel-scale lunar DEM generation from high-resolution monocular imagery and low-resolution DEM," *Remote Sens.*, vol. 14, 2022, Art. no. 5420.

[14] S. Dong et al., "Reloc3r: Large-scale training of relative camera pose regression for generalizable, fast, and accurate visual localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2025, pp. 16739–16752.

[15] J. Zheng et al., "Scene-agnostic pose regression for visual localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2025, pp. 27092–27102.

[16] S. Yang and L. Ma, "Diffusion model-based pedestrian de-occlusion method using a monocular camera sensor for indoor visual localization," *IEEE Sensors J.*, vol. 23, no. 22, pp. 28421–28429, Nov. 2023.

[17] L. Chang, X. Niu, T. Liu, J. Tang, and C. Qian, "GNSS/INS/LiDAR-SLAM integrated navigation system based on graph optimization," *Remote Sens.*, vol. 11, 2019, Art. no. 1009.

[18] C. Holmes and T. Barfoot, "An efficient global optimality certificate for landmark-based SLAM," *IEEE Robot. Automat. Lett.*. vol. 8, no. 3, 1539–1546, Mar. 2023.

[19] W. Huang, S. Jiang, and W. Jiang, "Camera self-calibration with GNSS constrained bundle adjustment for weakly structured long corridor UAV images," *Remote Sens.*, vol. 13, 2021, Art. no. 4222.

[20] X. Wan, Y. Shao, S. Zhang, and S. Li, "Terrain aided planetary UAV localization based on geo-referencing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4602018.

[21] X. Luo, X. Wan, Y. Gao, Y. Tian, W. Zhang, and L. Shu, "JointLoc: A real-time visual localization framework for planetary UAVs based on joint relative and absolute pose estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2024, pp. 3348–3355.

[22] S. Somvanshi, S. A. Javed, M. M. Islam, D. Pandit, and S. Das, "A survey on kolmogorov-arnold network," *ACM Comput. Surv.*, vol. 58, no. 2, pp. 1–35, 2025.

[23] M. Oquab et al., "DINOV2: Learning robust visual features without supervision," 2023, *arXiv:2304.07193*.

[24] H. Ta, "BSRBF-KAN: A combination of B-splines and radial basis functions in kolmogorov-arnold networks," in *Proc. Int. Symp. Inf. Commun. Technol.*, 2024, pp. 3–15.

[25] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," in *Proc. Field Serv. Robot. Results 11th Int. Conf.*, 2018, pp. 621–635.

[26] R. Mur-Artal and J. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*. vol. 33, no. 5, 1255–1262, Oct. 2017.

[27] C. Campos, R. Elvira, J. Rodriguez, J. Montiel, and J. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, 1874–1890, Dec. 2021.

[28] J. Naumann, B. Xu, S. Leutenegger, and X. Zuo, "NeRF-VO: Real-time sparse visual odometry with neural radiance fields," *IEEE Robot. Automat. Lett.*, vol. 9, no. 8, pp. 7278–7285, Aug. 2024.

[29] W. Chen, L. Chen, R. Wang, and M. Pollefeys, "LEAP-VO: Long-term effective any point tracking for visual odometry," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 19844–19853.

[30] A. Lin, J. Zhang, D. Ramanan, and S. Tulsiani, "Relpose: Recovering 6D poses from sparse-view observations," in *Proc. Int. Conf. 3D Vis.*, 2024, pp. 106–115.

[31] S. Chen, Z. Wang, and V. Prisacariu, "Direct-posenet: Absolute pose regression with photometric consistency," in *Proc. Int. Conf. 3D Vis.*, 2021, pp. 1175–1185.

[32] J. Wang, C. Rupprecht, and D. Novotny, "Posediffusion: Solving pose estimation via diffusion-aided bundle adjustment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 9773–9783.

[33] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 34–40.

[34] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5744–5753.

[35] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. vol. 13 no. 4, pp. 376–380, Apr. 1991.

[36] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 12716–12725.

[37] P.-E. Sarlin et al., "Back to the feature: Learning robust camera localization from pixels to pose," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 3247–3257.

[38] J. Christian, L. Hong, P. McKee, R. Christensen, and T. Crain, "Image-based lunar terrain relative navigation without a map: Measurements," *J. Spacecraft Rockets*, vol. 58, pp. 164–181, 2021.