Vision-based Geo-Localization of Future Mars Rotorcraft in Challenging Illumination Conditions

Dario Pisanti[1], Robert Hewitt[1], Roland Brockers[1], Georgios Georgakis[1]
[1]Jet Propulsion Lab, California Institute of Technology

***Abstract*—Planetary exploration using aerial assets has the potential for unprecedented scientific discoveries on Mars. While NASA's Mars helicopter Ingenuity proved flight in Martian atmosphere is possible, future Mars rotocrafts will require advanced navigation capabilities for long-range flights. One such critical capability is Map-based Localization (MbL) which registers an onboard image to a reference map during flight in order to mitigate cumulative drift from visual odometry. However, significant illumination differences between rotocraft observations and a reference map prove challenging for traditional MbL systems, restricting the operational window of the vehicle. In this work, we investigate a new MbL system and propose Geo-LoFTR, a geometry-aided deep learning model for image registration that is more robust under large illumination differences than prior models. The system is supported by a custom simulation framework that uses real orbital maps to produce large amounts of realistic images of the Martian terrain. Comprehensive evaluations show that our proposed system outperforms prior MbL efforts in terms of localization accuracy under significant lighting and scale variations. Furthermore, we demonstrate the validity of our approach across a simulated Martian day.**

I. Introduction

The demonstration flights of NASA's Mars Helicopter, *Ingenuity*, have marked a groundbreaking milestone in Mars exploration and scientific discovery [13]. The next generation of Mars rotorcraft will require even more sophisticated autonomous navigation capabilities to access diverse terrains under challenging environmental conditions and to support long-range, fully autonomous flights, with the ultimate goal of enabling high-priority investigations in Martian astrobiology, climate, and geology [2].

One of the next evolutionary steps in advancing aerial mobility on the red planet is represented by the Mars Science Helicopter (MSH), a concept of a hexacopter with a payload capacity of up to 5 kg and lateral traverse capability over 10 km [33]. To achieve precision during such long-range traverses, it is crucial to minimize drift in position estimates generated by the on-board Visual Inertial Odometry (VIO) in a global navigation satellite system (GNSS)-denied environment like Mars. During flights demos on Mars, Ingenuity's VIO algorithm produced a position error drift of 2-6% over a flight envelope that includes flight with up to 625 meters of total distance traversed, at a maximum forward velocity of 5 m/s and a maximum altitude of 10 m. For MSH's long-range traverses within ∼10 km and altitudes up to 100 meters, the drift is expected to be considerably higher and onboard global localization needs to be performed online.

Global localization can be accomplished by registering images captured by the rotorcraft's navigation camera onto orbital maps that are pre-registered to a global reference frame followed by the derivation of the rotocraft's position and orientation relative to the orbital map. This inherently drift-free geo-localization technique is referred to as Map-based Localization (MbL). At the heart of every MbL system lies an image registration method that identifies distinctive features or landmarks in the onboard image and the map in order to enable localization. The image registration can be very challenging in this domain due to the large differences in lighting between the onboard image and the map. Another factor that may challenge MbL performance is that of scale difference, particularly relevant for the MSH which needs to operate in a wide range of altitudes up to 100 m. Furthermore, the localization accuracy can also be impacted by variation in terrain morphology and textures. Complex, high-relief terrain may strengthen visual disparity between the onboard and map images at different times of day, due to changes in shadow casting. Conversely, textureless terrain can hinder the identification of distinctive features necessary for matching the map and image.

In spite of the recent progress of deep learning in visual tasks, space applications still rely mostly on template-matching techniques or hand-crafted features to solve the image registration problem for two main reasons: 1) Low computational requirements, and 2) They solve a relatively narrow problem with strong assumptions being made regarding the viewpoint, scale, and lighting conditions. Examples include the Lander Vision System (LVS) [17] developed for the Mars2020 mission, the recently introduced rover global localization system [26], and initial studies on MbL for MSH [4]. However, such strong assumptions limit the operational capability of a future Mars rotocraft. For example, the lack of robustness to different illumination conditions would restrict the flights to the time of day when the orbital map was originally collected, thus posing stringent constraints to the operational mission envelope. While recent deep learning methods [32, 12] have demonstrated robustness to illumination and scale variations on in-the-wild datasets such as MegaDepth [19], the main bottleneck is the lack of large-scale datasets that would allow finetuning these methods in planetary domains.

In this paper, we explore a new MbL system that makes no assumptions regarding the illumination conditions or scale variation for vision-based geo-localization on Mars for a future Mars rotocraft. In particular, we incorporate a transformer-based method [32] for image registration into the MbL pipeline, and further enhance this method to use geome-

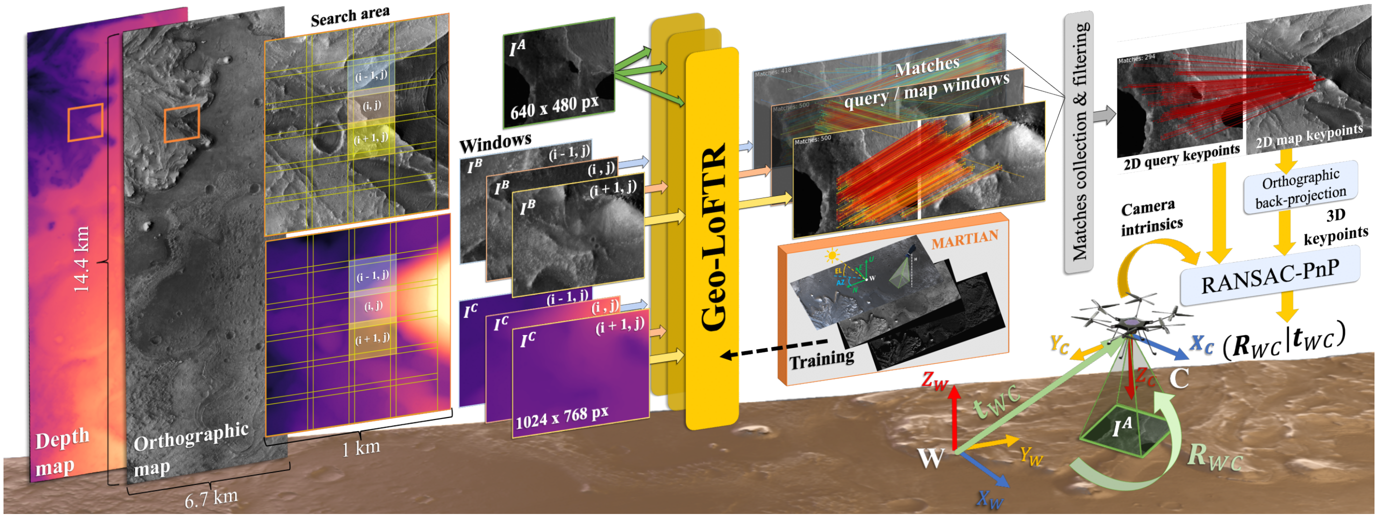

Fig. 1. Given an ortho-projected map of the terrain and a simulated onboard image we aim to estimate the pose of a rotocraft operating on Mars. Assuming a noisy pose prior, a search area is selected that is further divided into smaller regions and passed sequentially to our geometrically-enhanced Geo-LoFTR observation-to-map matcher. Geo-LoFTR is trained from data generated by our simulation framework MARTIAN. Finally, the matches are then filtered and passed to RANSAC-PnP to estimate the pose.

try in order to increase robustness to lighting variations. Furthermore, we introduce the Mars Aerial Rendering Tool for Imaging and Navigation (MARTIAN) which we use to generate a large-scale Mars dataset from orbital maps and train the image registration method. In summary, our contributions towards a robust MbL pipeline include:

- A new image matching method, Geo-LoFTR, that uses geometric context from digital terrain models and can improve localization accuracy compared to prior methods @1m up to 31.8% in challenging illumination conditions under low sun elevation angles.
- A custom simulation pipeline to generate maps and aerial observations of realistic Martian landscapes derived from HiRISE [24] data under a wide variety of illumination conditions.
- A comprehensive evaluation of our MbL approach validating its robustness under challenging environmental conditions which clearly demonstrates the advantage of incorporating geometric context.

II. Related work

Space-based applications of MbL traditionally rely on template-matching techniques, where an ortho-rectified onboard image slides over a reference map in order to estimate pixel-wise similarity using distance metrics such as Normalized Cross-Correlation [28], Phase-Correlation [34], and Mutual Information [1]. The Mars2020 Lander Vision System (LVS) integrated a coarse-to-fine template-matching approach in their terrain relative navigation pipeline to perform precise absolute localization on a Context Camera (CTX) map (6 m / pixel) during the mission's Entry Descent and Landing (EDL) phase [17]. The Censible framework proposed in [26] successfully performed global localization of the Perseverance rover by registering an ortho-mosaic of panoramic stereo images collected onboard to a HiRISE map (0.25m / pixel) using a modified census transform [38]. Even though template-matching approaches were successful in the aforementioned cases, they were applied under minimal lighting variations, and they are generally not robust to viewpoint and in-plane rotation variations without a correction step. Recently, the Lighting Invariant Matching Algorithm (LIMA) [29] proposed a new correlation-based method that is robust to challenging illuminations, but it assumes the existence of at least two pre-registered images with at least some lighting diversity.

Beyond template-matching, hand-crafted features such as SIFT [21], ORB [30], SURF [3], and SOSE [5] have been investigated and pitted against deep learning approaches for MbL and related problems. For example, the work of [4] demonstrated SuperPoint [9] to outperform SIFT under very low sun elevation angles in terms of localization accuracy, while AstroVision [10] showed ASLFeat [23] to be more robust than hand-crafted features for feature tracking under large shadows for the task of small body navigation. Similar to our work, JointLoc [22] proposes a vision-based system for UAV localization on Mars using SuperPoint [9] and LightGlue [20] for local feature matching. However, JointLoc does not investigate challenging illumination and scale variations during localization and uses mostly synthetic images of Mars with a small sample of real images.

Many deep learning methods have been introduced for the fundamental task of image matching that have shown robustness to real-world changes in scale, illumination, and viewpoint. LoFTR [32] uses Transformers [35] in a detector-free manner, RoMa [12] leverages features from the vision foundation model of DINOv2 [27], and DKM [11] estimates a dense warp to provide a match for every pixel. Other works such as GAM [37] and GoMatch [40] sought to introduce geometric information during the matching process for the task

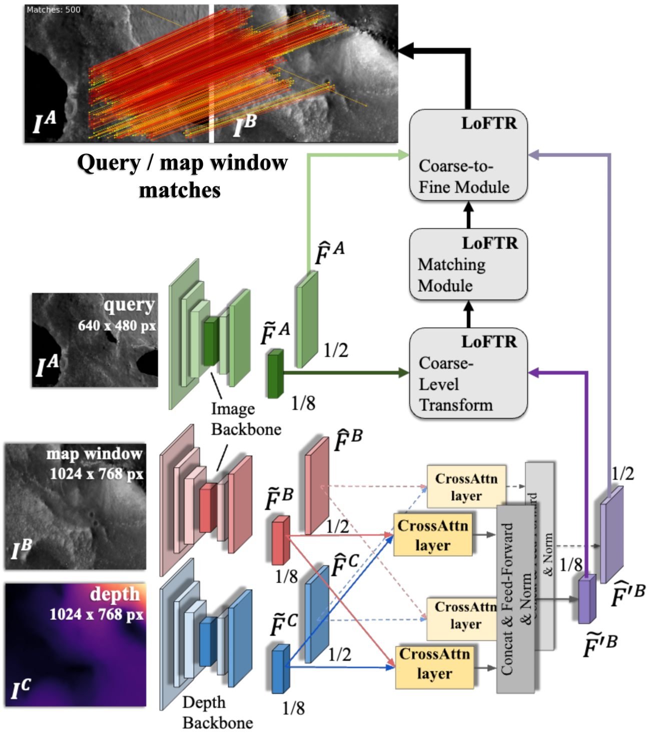

Fig. 2. Architecture of Geo-LoFTR that uses as inputs the query image I^A and a map crop image I^B along with its corresponding depth I^C. Geo-LoFTR learns to merge visual information from I^B with geometric from I^C using parallel $CrossAttn$ operations. The produced features then follow the coarse-to-fine approach proposed in LoFTR [32]. Our experimental evaluation shows that Geo-LoFTR is more robust than the original LoFTR under challenging illumination conditions.

of visual localization. Inspired by these recent developments, we aim to adapt the state-of-the-art method of LoFTR [32] for MbL on Mars and extend its architecture to leverage geometric context.

Finally, we acknowledge the extended body of work on vision-based global localization for UAVs. A recent survey that focuses on deep learning methods for UAV localization can be found here [7, 36]. A large part of this literature is devoted to Earth-based applications where typically a drone image is registered to geo-referenced satellite imagery. Given the abundance of data in this domain, progress was driven by the adaption of deep learning in existing pipelines that showed increased robustness to challenging conditions. For example, [14] incorporates SuperGlue [31] for solving this task for long distance flights in-the-wild, while [8, 16] rely on deep features for image retrieval for low altitude flights in urban environments. Multimodal inputs were also explored in this domain, with [41] including language descriptions of drone and satellite images to facilitate cross-view matching.

III. Methodology

We propose a new MbL method that is robust to challenging illumination conditions on Mars. Our approach focuses on improving the registration capabilities of MbL, and addresses the lack of training data in this domain. More specifically, we introduce Geo-LoFTR, an image matching model that learns to merge geometric and image features (Sec. III-A), based on the state-of-the-art method of LoFTR [32]. In addition, we present MARTIAN our simulation tool that uses real orbital maps (Sec. III-B), and our strategy for generating a training set for Geo-LoFTR (Sec. III-C). Finally, we discuss the MbL pipeline (Sec. III-D). An overview of our method along with a summary of the pipeline is illustrated in Figure 1.

A. *Geometry-aided Local Feature Matching*

Preliminaries: LoFTR. We first briefly introduce the image matching approach we use as basis for our geometry-aided observation to map registration. LoFTR follows a detector-free approach to produce semi-dense matches between two images I^A and I^B. Its strength lies on using transformers to process features from a CNN backbone at two scales $\tilde{F}^A \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times d}$ and $\hat{F}^A \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times d}$, where H, W are height and width of I^A and d is the feature dimension. This allows the already semantically rich pixel-wise representations to capture global image context. The corresponding features are also extracted from I^B. The method then follows a coarse-to-fine approach, where initial dense matching is performed at the coarse level, followed by refinement of the matches around a small window using the higher resolution feature maps.

We find LoFTR to be a suitable method for adopting into our MbL pipeline for two reasons: 1) The usage of transformers allows for some robustness to images with large shadows due to their ability to draw information from other parts of the image, and 2) it incorporates linear attention layers [18] that renders the method more computationally efficient than other approaches (e.g., RoMa [12]). For more details please see the original paper [32].

Geo-LoFTR. We aim to incorporate geometric context during feature matching between onboard observations and the reference map in order to increase robustness to challenging lighting scenarios where visual cues alone may lead to degeneracy. In the context of map-based localization on Mars, we take advantage of the 3D information from a digital elevation model and enrich the learned representation of the ortho-projected map.

To accomplish this, we extend the original LoFTR architecture to take as input a depth image I^C of a map crop along with the corresponding gray-scale image I^B of the crop from the ortho-rectified map and the gray-scale onboard image I^A. Each input depth map crop is normalized using the highest depth value in the crop, to avoid overfitting on absolute depth values of local areas, but rather associate relative geometry with visual information. In order to extract the corresponding coarse $\tilde{F}^C$ and fine $\hat{F}^C$ features for I^C we use the same ResNet-18 [15] backbone from the original LoFTR architecture.

Our objective is to learn how to merge the features $\tilde{F}^B$, $\hat{F}^B$ with the corresponding $\tilde{F}^C$, $\hat{F}^C$ to produce coarse- and fine-level feature maps that incorporate both visual and geometric information of the local map area we wish to localize the vehicle. To do so, we use cross-attention layers in both directions in the following manner:

$$\tilde{F}'^B = G\left(CrossAttn(\tilde{F}^B, \tilde{F}^C) \oplus CrossAttn(\tilde{F}^C, \tilde{F}^B)\right)$$

where the first argument for each *CrossAttn* layer is used as the query, $\oplus$ concatenates the outputs along the feature dimension, and G is a small feedforward network comprised of two linear layers followed by LayerNorm. The resulting $\tilde{F}'^B \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times d}$ is the merged coarse level feature map. The corresponding process is repeated for the fine level maps to produce $\hat{F}'^B \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times d}$. The rest of the pipeline follows the original LoFTR, with the merged feature representations being used in the coarse and fine transformer modules instead of the features extracted only from I^B. Figure 2 shows the architectural details of Geo-LoFTR.

B. MARTIAN: Mars Aerial Rendering Tool for Imaging and Navigation

Unlike Earth-based applications that have the privilege of abundant data [42, 39], relevant, annotated, and large-scale datasets are not readily available in the Martian domain. Data released from the Mars2020 mission[1] offer, among other, observations from the Lander vehicle during EDL, and from the navigation camera of Ingenuity. However, the data do not come with accurate pose annotations, and they are far from complete to perform a comprehensive study on the robustness of an MbL pipeline on scale and illumination variations.

Instead, we take advantage of real map products created from the Mars Reconnaissance Orbiter (MRO) High-Resolution Imaging Science Experiment (HiRISE) [24] to create a large-scale dataset suitable for training Geo-LoFTR and evaluating our MbL pipeline. We developed a python-based framework in the open-source 3D computer graphics software Blender to import HiRISE Digital Terrain Models (DTMs) and textured ortho-projected images in order to generate maps and aerial observations of a Martian site under various lighting conditions and at different altitudes.

HiRISE data. With its resolution capability at nadir of 25 centimeters per pixel from 300 km altitude, HiRISE has been serving as an indispensable orbital asset for identifying and selecting safe landing sites for robotics exploration missions. In this work, we utilize a 1 m / post DTM and a 0.25 m / pixel ortho-image with equirectangular projection generated from stereo pairs imaging of the Jezero Crater landing site for the Mars2020 Mission, over an area of 6.737 km by 14.403 km.

Terrain and texture modeling. The Jezero HiRISE DTM was imported in Blender 4.0 using a modified version of the original HiRISE import plug-in [6]. This add-on can load the terrain data at a desired resolution within a range (0., 100] % of the full model resolution and generate a mesh. By leveraging terrain metadata, including geographic information and resolution, a UV map is created for ortho-images to be draped over the companion DTM and used as a high-fidelity terrain texture. The MARTIAN framework initially imports the Jezero DTM with 10% of its original resolution, allowing for efficient terrain setup. Then, a material is added to the mesh, with its surface shading model being controlled via a Principled Bidirectional Scattering Distribution Function (BSDF). The 0.25 m / pixel Jezero ortho-image is then loaded in the shading editor in to serve as the base texture for the terrain surface. The ortho-image coordinates data are retrieved by a Texture Coordinate Node to ensure that the texture is properly mapped onto the 3D mesh. Finally, the mesh is reloaded with its full resolution of 1 m / post.

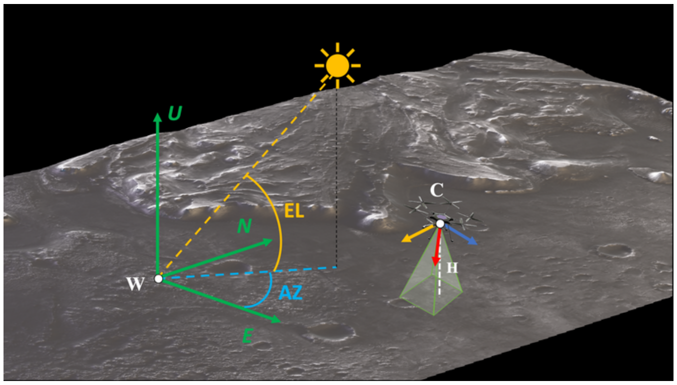

Fig. 3. View of the Jezero Crater's DTM in MARTIAN.

Scene setup and camera modeling. MARTIAN allows for setting multiple scenes for perspective and orthographic imaging with user-defined camera intrinsics and extrinsics (see Figure 3). Given the *world* reference frame W defined as a East-North-Up coordinate system with origin on the terrain map center, the camera can be located in W by providing the xy-coordinates and the altitude (in meters) with respect to the terrain at those coordinates. To position the camera object above the terrain mesh at the desired altitude, MARTIAN adopts a ray-tracing approach by using a Bounding Volume Hierarchy tree, a data structure used by Blender to efficiently organize geometric objects in 3D space. The camera frame C is centered at the camera's optical center, with its X-axis pointing to the right along the image width, the Z-axis pointing towards the terrain, and the Y-axis completing the orthogonal set. The attitude of the the C frame with respect to the world frame W can be specified as input, and the final pose $(\mathbf{R}_{WC}|\mathbf{t}_{WC})$ is saved, where $\mathbf{t}_{WC}$ is the location of the camera in world coordinated and $\mathbf{R}_{WC}$ is the rotation matrix that aligns W to C.

Lighting. MARTIAN provides the capability to render scenes in Blender with various illumination conditions by tuning Sun light source parameters such as the irradiance, the apparent disk diameter and orientation in the map frame.

[1] https://mars.nasa.gov/mars2020/multimedia/raw-images/

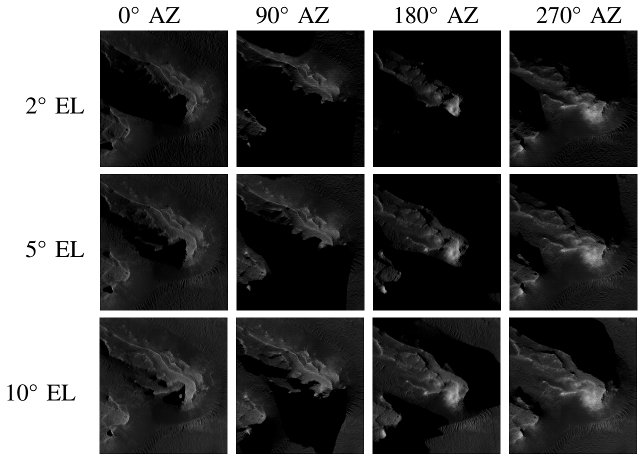

Fig. 4. Gray scale images of a map tile from the Jezero crater site, rendered with different combinations of sun AZ and EL. Generated with MARTIAN.

Parameters	**Maps**	**Observations**
N.o. gray images	17	4500
N.o. depth images	1	4500
Image size (H × W)	26949 × 57613	480 × 640
Pixel resolution	0.25 m / px	[0.25, 0.78] m / px
Projection type	Orthographic	Perspective
Orthographic scale	6737 m	/
Focal length	/	32 mm
Sensor width	/	80 mm
Location in W	(0,0) m	uniform random distribution
Orientation in W	nadir-pointing	nadir-pointing
Altitude	4000 m	[64, 200] m
Sun AZ	[0, 360]° with 45° steps	180°
Sun EL	30°, 60°, 90°	40°

TABLE I
PARAMETERS FOR THE MAPS AND OBSERVATIONS IMAGE DATASET SOURCED FOR THE GENERATION OF THE PAIRS FORMED BY QUERY OBSERVATIONS AND MAP WINDOWS, USED FOR LOFTR AND GEO-LOFTR TRAINING.

Higher irradiance values cast brighter illumination and shadows in the scene, while the angular diameter of the Sun disk as seen by the scene controls the softness versus harshness of the shadows. The Sun orientation simulates the scene time of the day and it is specified by the user though Sun elevation (EL) and azimuth angles (AZ) as shown in Figure 3. The lighting computations are performed using the Blender Cycles engine. This is a physically-based rendering engine that uses a ray tracing algorithm to accurately simulate light behavior. In this work, we set Sun irradiance to the maximum value of 590 W/m^2 and the angular diameter to 0.35°, coherently with actual estimations for Mars. We compared empirically the visual appearance of the real ortho-projected map to the MARTIAN rendered one, and tuned the BSDF and camera exposure such that they match visually. Figure 4 illustrates the effect of different combinations of sun azimuth and elevation angles on the visual appearance of an orthographic map tile rendered in MARTIAN.

C. Generating a training set with MARTIAN

We generated a training image dataset comprising 17 orthographic gray-scale maps rendered from combinations of Sun azimuth $(0° - 360°, 45° steps)$ and elevation $(30°, 60°, 90°)$, one corresponding orthographic depth map, and 4500 nadir-pointing aerial observations at fixed Sun angles AZ=180° and EL=40° along with their corresponding depth images. The query observations were randomly sampled from the HiRISE Jezero Crater's DTM with uniform distribution in the (x,y) coordinates in the world frame and within altitude range $[64, 200]$ m. The query camera intrinsics are given by a pinhole camera model characterized by a sensor width of 80 mm, a focal length of 32 mm, 0 lens shifts along the image width and height axes, and a unitary pixel aspect ratio. Camera extrinsics, along with altitude data, were stored for each observation and used as ground truth. Further details are reported in Table I.

Training examples for Geo-LoFTR are created by forming triplets (I_A, I_B, I_C) out of query observations and map windows crops (gray-scale and depth images) for multiple combinations of Sun azimuth and elevation angles' offsets between queries and the source maps. For a given combination of query and map lighting, each query observation is paired with map windows with at least 25% area overlap on the terrain. Therefore, the set of triplets is formed by different combination of image I_A with the mop window tuple (I_B, I_C). The map window sizes are carefully chosen to introduce an appropriate level of scale variance within the altitude range of the observations, ensuring a balance between model generalization and training efficiency. Geo-LoFTR has been fine-tuned from the original LoFTR pre-trained model on a total of 150K generated triplets. An independent validation set of 3400 triplets has been used to regularly assess the model performance during training and prevent over-fitting.

D. Map-based Localization Pipeline

The goal of map-based localization is to retrieve the onboard (query) camera pose ($\mathbf{R}_{WC_{\text{query}}}$, $|\mathbf{t}_{WC_{\text{query}}}$) in the world frame W, where $\mathbf{R}_{WC_{\text{query}}}$ is the rotation matrix that aligns W with the camera frame C_{query}, and $\mathbf{t}_{WC_{\text{query}}}$ is the camera location in W. This is preceded by the registration of the query observation captured by a camera with known intrinsic parameters over a geo-referenced ortho-projected map.

We assume that a hypothetical future Mars rotocraft is going to be equipped with onboard Visual Odometry (VIO) such that it would provide a noisy pose prior to the MbL pipeline. This allows to narrow down the registration of the query image to a smaller search area of the reference map based on the uncertainty of the VIO pose. In our work we assume a large predetermined search area of 1 km^2 centered at the query observation for two reasons. First, it allows us to simulate a conservative scenario with a high-uncertainty pose prior. Second, it prevents our evaluation strategy from being overly influenced by variations in the poses themselves, thus ensuring a more consistent assessment of our pipeline's performance across experiments. We note that the overall size of the map is 6.737 km by 14.403 km with a resolution of 0.25 m / pixel.

For each query, the map search area is further divided into multiple overlapping windows, each sized at 1024×768 pixels and with an overlap of 10%. The query image I^A is paired with each map window crop I^B and the corresponding depth image crop I^C. The formed triplet (I^A, I^B, I^C) is processed by Geo-LoFTR, which outputs matched keypoints on both the map windows and the query observations with their confidence scores. The top 500 matches are retained for each map window, with further filtering applied across the entire search area to include only matches meeting a 95% confidence threshold. Given a simulated orthographic map camera with pose $(\mathbf{R}_{WC_{\text{map}}}, |\mathbf{t}_{WC_{\text{map}}})$ in the world frame, each matched point of pixel coordinates $(u_{\text{map}}, v_{\text{map}})$ in the map image plane is back-projected to its 3D location in W using an inverse orthographic projection:

$$\begin{bmatrix} X^W \\ Y^W \\ Z^W \end{bmatrix} = p_{\text{map}} \mathbf{R}_{WC_{\text{map}}} \begin{bmatrix} 1 & 0 & -c_{x,\text{map}} \\ 0 & 1 & -c_{y,\text{map}} \\ 0 & 0 & Z/p_{\text{map}} \end{bmatrix} \begin{bmatrix} u_{\text{map}} \\ v_{\text{map}} \\ 1 \end{bmatrix} + \mathbf{t}_{WC_{\text{map}}}$$

where Z is the depth of the map keypoint, $(c_{x,\text{map}}, c_{y,\text{map}})$ is the optical center and p_{map} is the pixel resolution. The 2D matched points on the query images, their 3D correspondences, and the query camera intrinsics matrix $\mathbf{K}$ are then fed to a RANSAC-PnP algorithm to solve the perspective projection problem and retrieve the estimated pose $(\widetilde{\mathbf{R}}_{WC_{\text{query}}}, |\widetilde{\mathbf{t}}_{WC_{\text{query}}})$.

IV. Map-based Localization Evaluation

We evaluated the performance of our MbL pipeline on multiple image datasets generated in MARTIAN from HiRISE maps and DTMs of the Jezero Crater. To ensure an unbiased assessment, none of the test data overlaps with the data used for training. We conducted several experiments to assess the robustness of these methods under changes in lighting (Sec. IV-A) and scale (Sec. IV-B). Figure 5 shows areas on orthographic maps sampled from the test sets and rendered with two different illumination conditions, accompanied by three example observations at different altitudes. To further stress our MbL pipeline with challenging lighting in a real-case scenario, we tested it over a simulated Martian day on the Jezero Crater site, with aerial observations generated in MARTIAN at multiple simulated times of day from sunrise to sunset (Sec. IV-D). Furthermore, an evaluation of our method's performance under varying terrain morphologies is presented in the supplementary material.

We compared our results to the original LoFTR model fine-tuned on our training dataset (*Fine-tuned LoFTR*), and to the model pre-trained on the MegaDepth dataset (*Pre-trained LoFTR*). As for comparison with state-of-the-art feature matching in planetary aerial mobility, we also tested SIFT that proved to be one of the most accurate handcrafted methods for absolute localization over simulated Mars terrain [4]. In each experiment, we use the percentage of queries with localization error $\|\mathbf{t}_{WC_{\text{query}}} - \widetilde{\mathbf{t}}_{WC_{\text{query}}}\|$ below 1m (@1m) as our evaluation metric. Also, we plot the Cumulative Distribution Function (CDF) of the localization accuracy up to 10m.

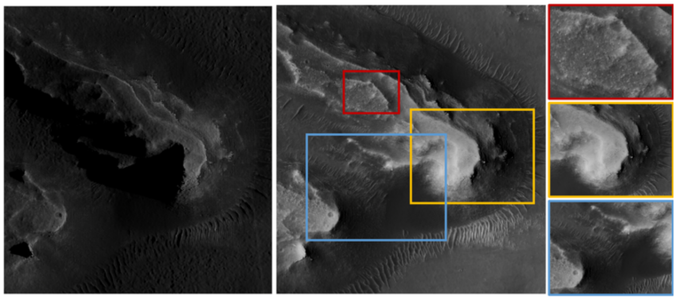

Fig. 5. Tiles from orthographic maps at sun (AZ=0°, EL=5°) (*left*) (AZ=180°, EL=40°) (*center*) with three sampled query observations (*right*).

A. Robustness to Changing Solar Angles

Robustness to challenging illumination variance is assessed through registering query observations onto orthographic maps rendered with varying sun elevation and azimuth angles, the effects of which are evaluated separately.

The dataset for the experiment addressing the robustness to sun elevation changes comprises orthographic maps rendered at EL=$\{2, 5, 10, 40, 60, 90\}$ and AZ=180, along with 500 nadir-pointing query observations at fixed sun angles (AZ=180, EL=40). The queries were taken in the 64-200 m altitude range which encompasses the nominal operation of the MSH. The minimum altitude of 64 m is set by the best resolution achievable in MARTIAN, which coincides with the 0.25 cm / pixel resolution of the HiRISE ortho-image used as texture. It is worth to note that elevations below 30° were not encountered during the model training. During the Martian day when our reference HiRISE map was collected, the sun elevation varied in the 7.9-82.6° range, from 6:00 to 17:00 Local Mean Solar Time (LMST), with sunrise and sunset occurring at 05:11 LMST and 17:32 LMST, respectively. Therefore, elevations below 10° occupy a very brief portion of sun's local trajectory, representing exceptional cases in the Martian surface operations. Nevertheless, we include these cases in our evaluation to assess the models' ability to generalize and perform effectively under challenging lighting conditions.

To evaluate sun azimuth effect we generated 500 nadir-pointing queries with sun AZ=0° and EL=10° to be registered onto maps with varying azimuth angles in the 0-360° and same elevation as the queries.

Robustness to sun elevation. Figure 6 shows the CDFs of the test observations' localization error onto maps at four different sun elevations and fixed 0 azimuth offset. Geo-LoFTR outperforms all the other methods in localization accuracy across the entire range of sun elevation offsets. In the case of zero sun angles' offsets, Geo-LoFTR is 87% @1m accurate, showing an improvement of +38% over the fine-tuned model, and +63% over SIFT. Below 10° EL of the map, the performance of all the methods is significantly impacted, with Geo-LoFTR being 17% @1m accurate at the very challenging case EL=2°, where the pre-trained model and SIFT completely fail.

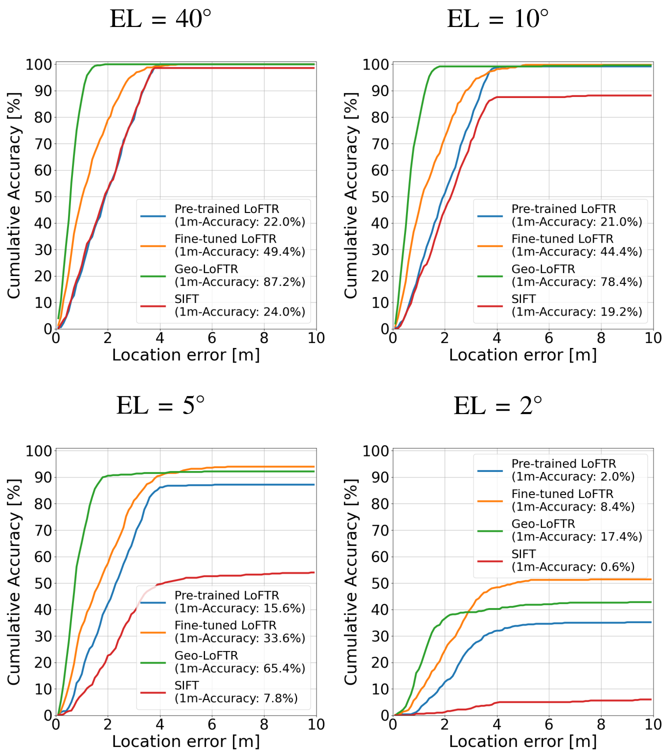

Fig. 6. Cumulative distributions of the localization error of simulated Mars observations at sun AZ=180° and EL=40°, registered onto maps at four different elevation angles and 0° azimuth offset.

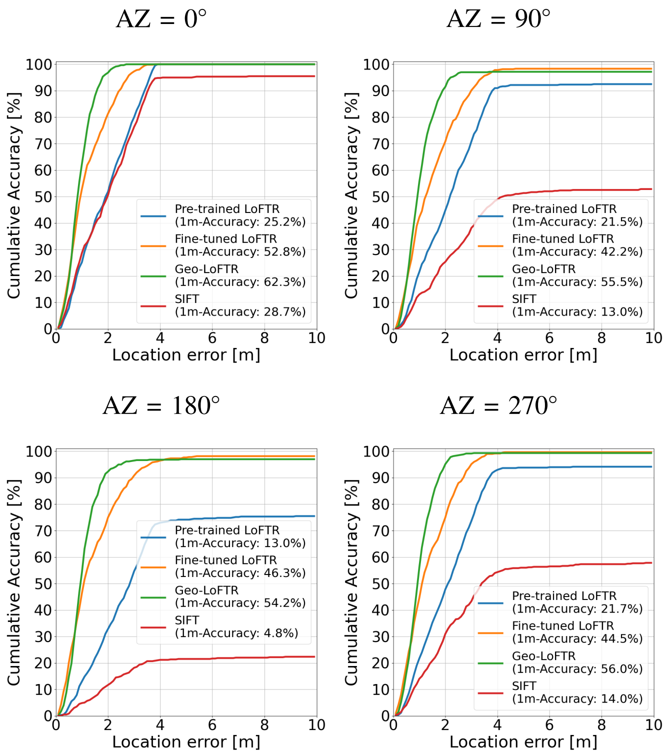

Fig. 7. Cumulative distributions of the localization error of simulated Mars observations at sun AZ=0° and EL=10°, registered onto maps at four different azimuth angles and 0° elevation offset.

Robustness to sun azimuth. The sun azimuth effect on MbL performance is presented through the cumulative distributions (Figure 7) of the localization error for the test observations registered onto maps at four different azimuth angles. Also in this experiment, Geo-LoFTR proved to be the most accurate model with a @1m accuracy being bound to the 54-63 % range in the entire map sun azimuth range, despite the relatively low elevation of 10°. The number and quality of the SIFT matched keypoints between query and map (Figure 8) decreases much faster than the LoFTR-based models as we depart from the zero azimuth offset case, with failure already at 90° offset.

B. Robustness to Scale Variation

We split the test observations from Sec. IV-A in three different altitude sub-ranges, and registered them onto maps with zero sun angle offsets to assess the pipeline's performance under scale changes. Figure 9 shows the CDF of the localization errors of observations taken within 64m-112m, 112m-155m, 155m-200m registered on a map with constant (AZ=180°, EL=40°) sun angles. With only a -7% @1m accuracy drop across the entire altitude range, Geo-LoFTR proved to be more robust than the fine-tuned model (-33% @1m). A similar degree of scale invariance is shown for the pre-trained model and SIFT, although being much less accurate.

C. Robustness to Combined Illumination and Scale Changes

Leveraging the test data in Sec IV-A, we performed a quantitative evaluation of the scale variation effects in conjunction with sun angle offsets. Figure 10, shows the @1m accuracy as a function of map sun EL and AZ for observations taken within three different altitude ranges. Although localization accuracy declines sharply at relatively low sun elevation angles, Geo-LoFTR maintains consistent localization performance across altitude variations within the 10–90° EL range. In contrast, the fine-tuned model demonstrates poor robustness in the same range. A similar trend is observed for azimuth variations, where localization accuracy remains stable with changing azimuth but decreases with altitude.

D. Localization Over a Simulated Martian Day

The MbL performance is investigated for observations taken at different LMSTs during a simulated Martian day on the Jezero Crater HiRISE map at coordinates (77.44°E, 18.43°N). We used the Mars24 [25] software developed by NASA Goddard Institute for Space Studies to compute the sun's local trajectory for the selected site on a given date. The chosen date, 2031-05-10, ensures that the sun zenith is at a relatively high elevation angle of 86.7°, allowing a broad range of elevation angles to be observed throughout the day (Figure 11).

We generated nadir-pointing observations in MARTIAN at multiple times of the day from 5:30 to 17:00 LMST, with a total of 3000 queries collected across the 64-200 m altitude range (Figure 12). We also rendered an orthographic map at 15:00 LMST, (AZ=175.1°, 39.9° EL), serving as a HiRISE-like reference. Figure 13 shows the @1m accuracy as

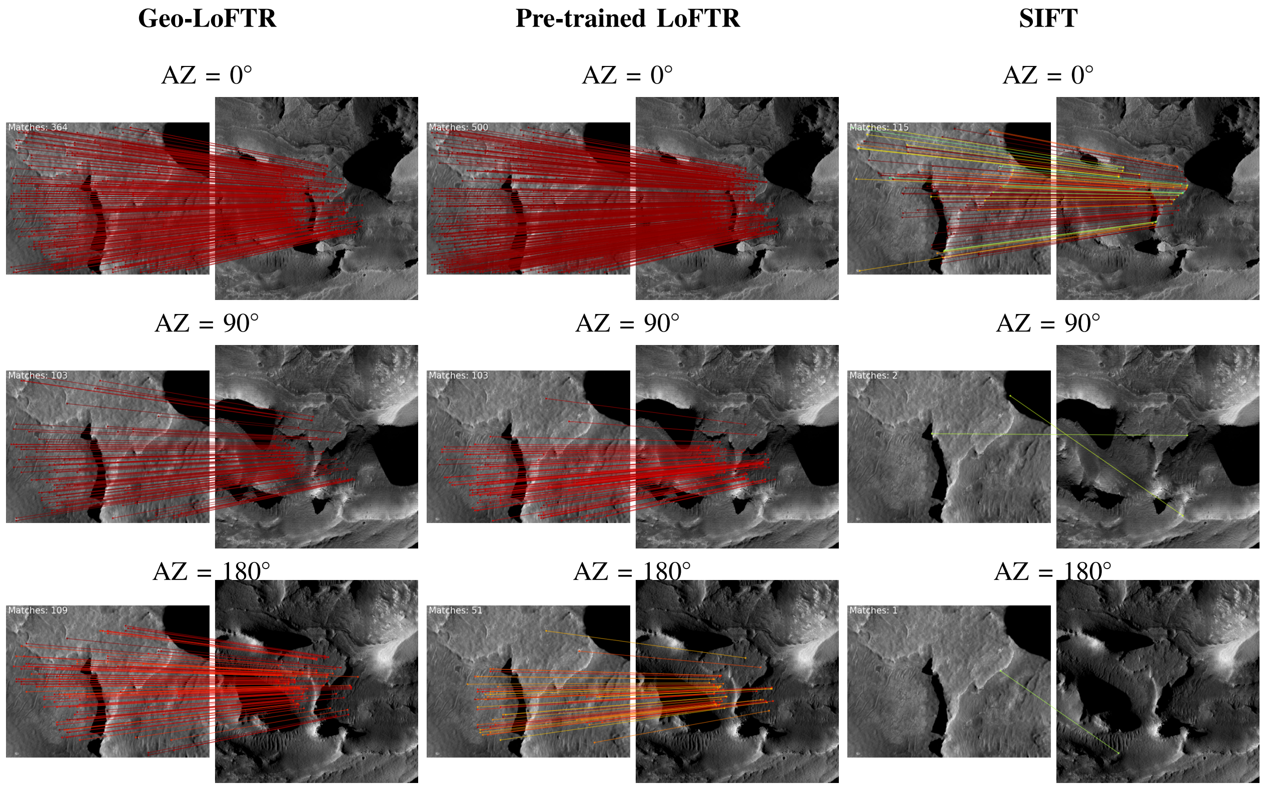

Fig. 8. Geo-LoFTR, Pre-trained LoFTR and SIFT matched keypoints displayed for a sample query image (*left side of each panel*) with (0° AZ, 10° EL) sun angles and a map search area image (*right side of each panel*) under three different sun elevations and 0° azimuth offset. Match lines are color-coded by confidence score, with redder indicating higher confidence. Despite still providing a localization solution in the 0-180° AZ range, the pre-trained LoFTR matches exhibit lower confidence with azimuth changes than Geo-LoFTR, resulting in a coarser localization.

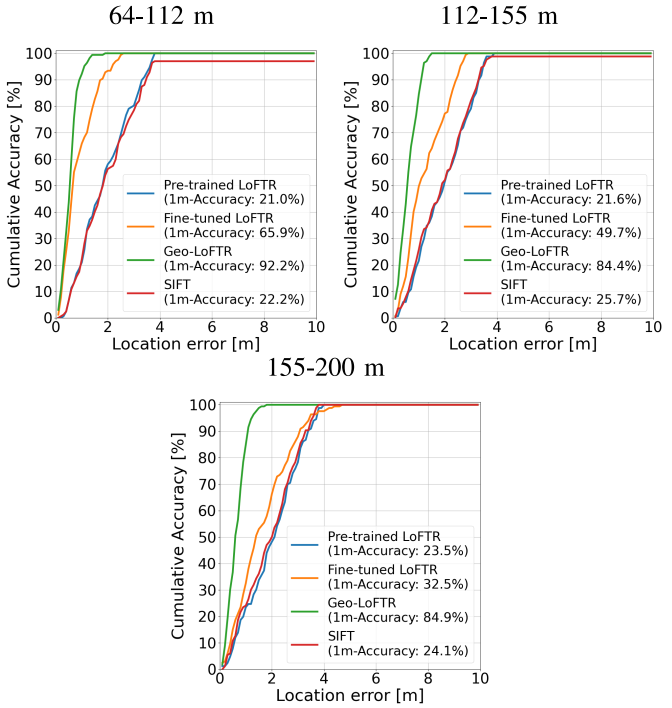

Fig. 9. Cumulative distributions of the localization error of simulated Mars observations at sun AZ=0° and EL=10°, registered onto maps with the same illumination condition for three different altitude ranges.

function of LMTs within three different altitude sub-ranges. Geo-LoFTR outperformed the other methods for most of the Martian day, except at 5:00 LMST, where the fine-tuned model shows better accuracy. However, the fine-tuned LoFTR experienced significant performance degradation with altitude, in contrast with the other methods that exhibited a certain grade of scale invariance also in this experiment.

E. Discussion

Geo-LoFTR demonstrated superior localization accuracy compared to other methods across a broad range of illumination conditions, indicating that incorporating depth information can mitigate degeneracies inherent to purely visual data. Robustness to sun elevation is maintained within a wide range of angles, except in extremely challenging cases (e.g., EL = 2°), where poor lighting and extensive shadow coverage might saturate the constraining power of the geometric information, leading to a rapid decline in localization accuracy. More stable is the behavior for changes in azimuth. Geo-LoFTR also showed greater robustness to changing observation altitude than the fine-tuned model across multiple experiments, suggesting that providing a geometric context contributes to scale invariance. A possible explanation is that depth data constrains matches by providing consistent pixel-to-pixel depth relationships across altitudes, reflecting terrain elevation alone. This added layer of geometric consistency likely enhances

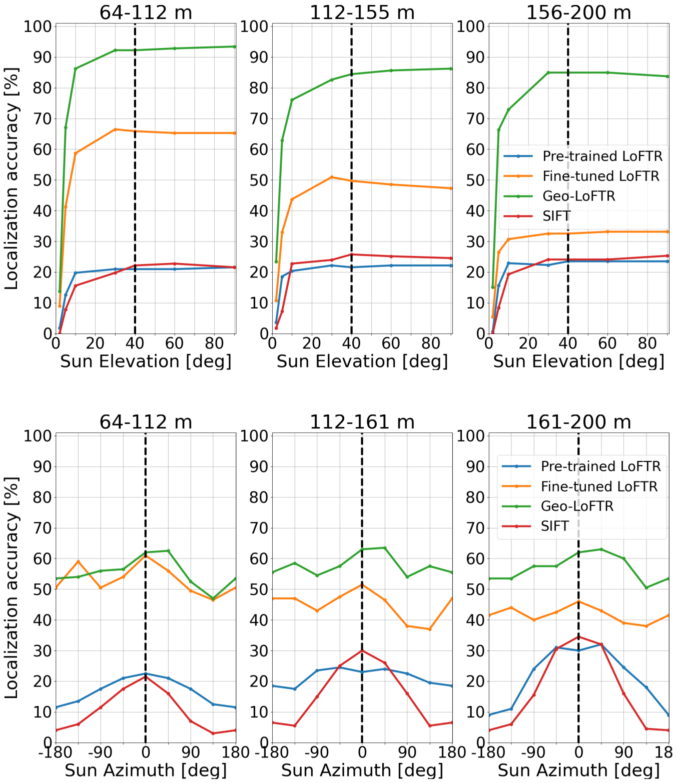

Fig. 10. Localization accuracy at 1m precision as a function of map sun elevation (*top*) and azimuth (*bottom*) for test observations across three altitude ranges. Sun azimuth angles are in the $[-180°, 180°]$ range. Map sun angles matching the observations are marked with a thick black vertical line.

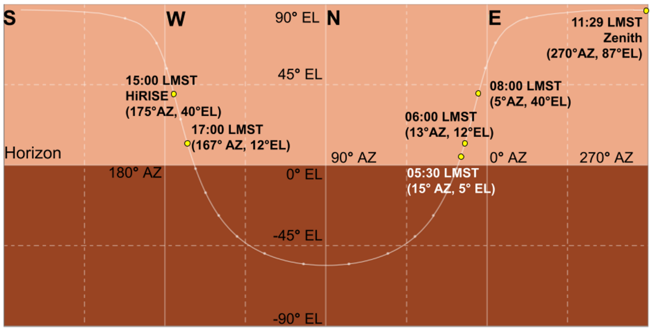

Fig. 11. Sun trajectory on a local panorama from 77.44°E longitude and 18.43°N latitude on Mars, on 2031-05-10, with positions shown at four Local Mean Solar Times (LMSTs). Adapted from Mars24 [25].

Geo-LoFTR's ability to learn accurate matches by reducing ambiguity from appearance-based features alone.

V. Limitations

Since our main motivation was to investigate robustness to illumination and scale variations, all observations in our datasets are nadir-pointing (i.e., we do not add any variation in the camera's viewpoint). While LoFTR [32] has shown sufficient invariance to viewpoint changes in in-the-wild datasets [19], it is still not clear whether this would transfer

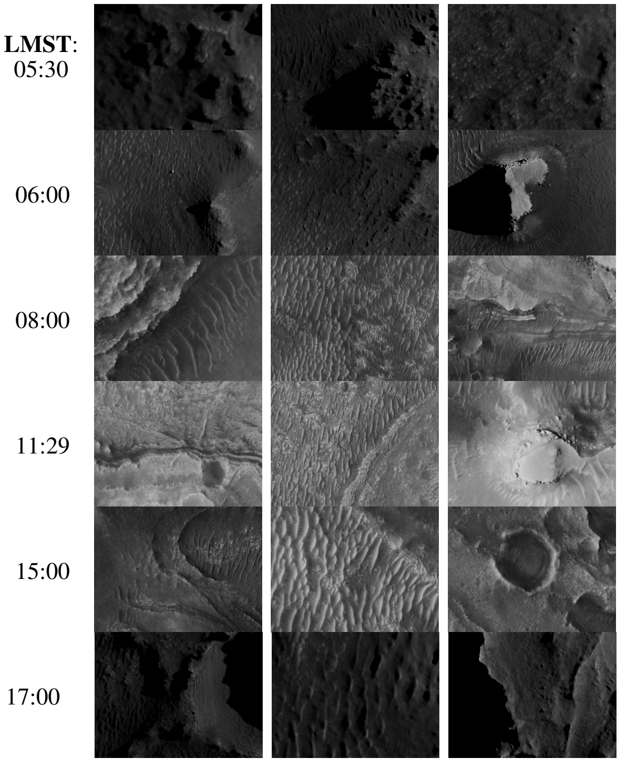

Fig. 12. Sample nadir-pointing observations rendered at different Local Mean Solar Times (LMSTs) taken on Mars on 2031-05-10. The reference HiRISE map is taken at 15:00 LMST. The sun Zenith is at 11:29 LMST.

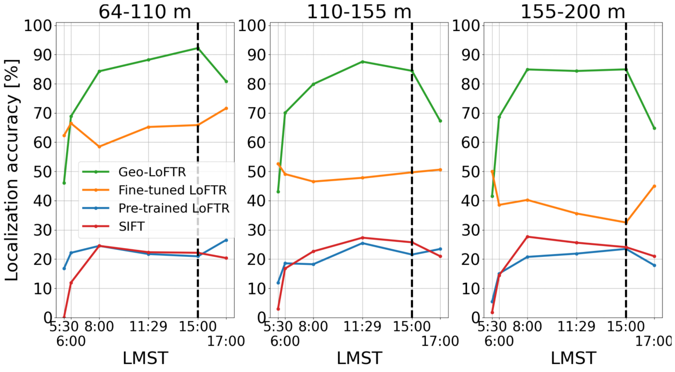

Fig. 13. Localization accuracy (@1m) as a function of Local Mean Solar Time (LMST) of simulated test observations from the Jezero Crater on 2031-05-10 across the 64-200 m altitude range. The reference HiRISE-like map is taken at 15:00 LMST (*dashed black line*). The sun Zenith is at 11:29 LMST.

in the Martian domain. Furthermore, we do not use any CTX orbital map products (6m/pixel) to generate our data and focus only on HiRISE (0.25m/pixel) maps which are of better quality and higher resolution. There is interest by the Mars exploration community to utilize CTX maps due to their almost 99% coverage of the planet. A separate investigation is warranted to determine whether current image matching methods can handle this large resolution difference. Finally, we did not focus on optimizing our pipeline in terms of computational efficiency as this was out-of-the-scope of this work.

VI. Conclusion

In this paper, we presented a new map-based localization pipeline that uses Geo-LoFTR, a geometry-aided feature matching model to register onboard images to reference maps, and MARTIAN, a custom simulation framework that uses real Digital Terrain Models from Mars to generate large-scale datasets. Our method has outperformed the baselines in terms of localization accuracy by a large margin, demonstrating that enhancing the feature matching with geometric context results in increased robustness to challenging environmental conditions. This robustness was witnessed across the board in terms of varying sun elevation and azimuth angles, along with altitude variation.

Acknowledgments

©2025 All rights reserved. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

References

[1] Adnan Ansar and Larry Matthies. Multi-modal image registration for localization in titan's atmosphere. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3349–3354. IEEE, 2009.

[2] Jonathan Bapst, T. J. Parker, J. Balaram, T. Tzanetos, L. H. Matthies, C. D. Edwards, A. Freeman, S. Withrow-Maser, W. Johnson, E. Amador-French, J. L. Bishop, I. J. Daubar, C. M. Dundas, A. A. Fraeman, C. W. Hamilton, C. Hardgrove, B. Horgan, C. W. Leung, Y. Lin, A. Mittelholz, and B. P. Weiss. Mars Science Helicopter: Compelling Science Enabled by an Aerial Platform. *Bulletin of the AAS*, 53(4), mar 18 2021. https://baas.aas.org/pub/2021n4i361.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-33833-8.

[4] Roland Brockers, Pedro Proença, Jeff Delaune, Jessica Todd, Larry Matthies, Theodore Tzanetos, and J. Bob Balaram. On-board absolute localization based on orbital imagery for a future mars science helicopter. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–11, 2022. doi: 10.1109/AERO53065.2022.9843673.

[5] Yang Cheng and Adnan Ansar. Simultaneous orientation and scale estimator. 2024.

[6] PhaseIV contributors. Blender hirise dtm importer, 2024. URL https://github.com/phaseIV/Blender-Hirise-DTM-Importer. Version 1.0.

[7] Andy Couturier and Moulay A Akhloufi. A review on deep learning for uav absolute visual localization. *Drones*, 8(11):622, 2024.

[8] Ming Dai, Enhui Zheng, Zhenhua Feng, Lei Qi, Jiedong Zhuang, and Wankou Yang. Vision-based uav self-positioning in low-altitude urban environments. *IEEE Transactions on Image Processing*, 2023.

[9] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018.

[10] Travis Driver, Katherine A Skinner, Mehregan Dor, and Panagiotis Tsiotras. Astrovision: Towards autonomous feature detection and description for missions to small bodies using deep learning. *Acta Astronautica*, 210:393–410, 2023.

[11] Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, and Michael Felsberg. Dkm: Dense kernelized feature matching for geometry estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17765–17775, 2023.

[12] Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. Roma: Robust dense feature matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19790–19800, 2024.

[13] Håvard Fjær Grip, Dylan Conway, Johnny Lam, Nathan Williams, Matthew P. Golombek, Roland Brockers, Michael Mischna, and Martin R. Cacan. Flying a helicopter on mars: How ingenuity's flights were planned, executed, and analyzed. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–17, 2022. doi: 10.1109/AERO53065.2022.9843813.

[14] Marius-Mihail Gurgu, Jorge Peña Queralta, and Tomi Westerlund. Vision-based gnss-free localization for uavs in the wild. In *2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR)*, pages 7–12. IEEE, 2022.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Mengfan He, Jiacheng Liu, Pengfei Gu, and Ziyang Meng. Leveraging map retrieval and alignment for robust uav visual geo-localization. *IEEE Transactions on Instrumentation and Measurement*, 2024.

[17] Andrew E. Johnson, Yang Cheng, Nikolas Trawny, James F. Montgomery, Steven Schroeder, Johnny Chang, Daniel Clouse, Seth Aaron, and Swati Mohan. Implementation of a map relative localization system for planetary landing. *Journal of Guidance, Control, and Dynamics*, 46(4):618–637, 2023. doi: 10.2514/1.G006780. URL https://doi.org/10.2514/1.G006780.

[18] Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pages 5156–5165. PMLR, 2020.

[19] Zhengqi Li and Noah Snavely. Megadepth: Learning single-view depth prediction from internet photos. In

Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2041–2050, 2018.

[20] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17627–17638, 2023.

[21] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. URL https://api.semanticscholar.org/CorpusID:174065.

[22] Xubo Luo, Xue Wan, Yixing Gao, Yaolin Tian, Wei Zhang, and Leizheng Shu. Jointloc: A real-time visual localization framework for planetary uavs based on joint relative and absolute pose estimation. *arXiv preprint arXiv:2405.07429*, 2024.

[23] Zixin Luo, Lei Zhou, Xuyang Bai, Hongkai Chen, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang, and Long Quan. Aslfeat: Learning local features of accurate shape and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6589–6598, 2020.

[24] Alfred S. McEwen, Eric M. Eliason, James W. Bergstrom, Nathan T. Bridges, Candice J. Hansen, W. Alan Delamere, John A. Grant, Virginia C. Gulick, Kenneth E. Herkenhoff, Laszlo Keszthelyi, Randolph L. Kirk, Michael T. Mellon, Steven W. Squyres, Nicolas Thomas, and Catherine M. Weitz. Mars reconnaissance orbiter's high resolution imaging science experiment (hirise). *Journal of Geophysical Research: Planets*, 112 (E5), 2007. doi: https://doi.org/10.1029/2005JE002605. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JE002605.

[25] NASA Goddard Institute for Space Studies. Mars24: Sunclock - a mars solar time and solar longitude calculator. URL https://www.giss.nasa.gov/tools/mars24/. Version 8.3.1, released on 2023-05-18. Accessed: 2024-10-28.

[26] Jeremy Nash, Quintin Dwight, Lucas Saldyt, Haoda Wang, Steven Myint, Adnan Ansar, and Vandi Verma. Censible: A robust and practical global localization framework for planetary surface missions. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8642–8648, 2024. doi: 10.1109/ICRA57147.2024.10611697.

[27] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[28] Tu-Hoa Pham, William Seto, Shreyansh Daftry, Barry Ridge, Johanna Hansen, Tristan Thrush, Mark Van der Merwe, Gerard Maggiolino, Alexander Brinkman, John Mayo, et al. Rover relocalization for mars sample return by virtual template synthesis and matching. *IEEE Robotics and Automation Letters*, 6(2):4009–4016, 2021.

[29] Noah Rothenberger, Georgios Georgakis, Yang Cheng, and Adnan Ansar. Illumination invariant image matching for lunar trn. In *AIAA SCITECH 2025 Forum*, page 2073, 2025.

[30] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International Conference on Computer Vision*, pages 2564–2571, 2011. doi: 10.1109/ICCV.2011.6126544.

[31] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.

[32] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8922–8931, 2021.

[33] Theodore Tzanetos, Jonathan Bapst, Gerik Kubiak, Luis Phillipe Tosi, Sam Sirlin, Roland Brockers, Jeff Delaune, Håvard Fjær Grip, Larry Matthies, J. Balaram, Shannah Withrow-Maser, Wayne Johnson, Larry Young, and Benjamin Pipenberg. Future of mars rotorcraft - mars science helicopter. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–16, 2022. doi: 10.1109/AERO53065.2022.9843501.

[34] Xue Wan, Jianguo Liu, Hongshi Yan, and Gareth L.K. Morgan. Illumination-invariant image matching for autonomous uav localisation based on optical sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119:198–213, 2016. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2016.05.016. URL https://www.sciencedirect.com/science/article/pii/S0924271616301113.

[35] A Waswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A Gomez, L Kaiser, and I Polosukhin. Attention is all you need. In *NIPS*, 2017.

[36] Yingxiao Xu, Long Pan, Chun Du, Jun Li, Ning Jing, and Jiangjiang Wu. Vision-based uavs aerial image localization: A survey. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 2018. URL https://api.semanticscholar.org/CorpusID:53428751.

[37] Hailin Yu, Youji Feng, Weicai Ye, Mingxuan Jiang, Hujun Bao, and Guofeng Zhang. Improving feature-based visual localization by geometry-aided matching. *arXiv preprint arXiv:2211.08712*, 2022.

[38] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6 1994 Proceedings, Volume II 3*, pages 151–158. Springer, 1994.

[39] Zhedong Zheng, Yunchao Wei, and Yi Yang. University-

1652: A multi-view multi-source benchmark for drone-based geo-localization. In *Proceedings of the 28th ACM international conference on Multimedia*, pages 1395–1403, 2020.

[40] Qunjie Zhou, Sérgio Agostinho, Aljoša Ošep, and Laura Leal-Taixé. Is geometry enough for matching in visual localization? In *European Conference on Computer Vision*, pages 407–425. Springer, 2022.

[41] Runzhe Zhu, Mingze Yang, Ling Yin, Fei Wu, and Yuncheng Yang. Uav's status is worth considering: A fusion representations matching method for geo-localization. *Sensors*, 23(2):720, 2023.

[42] Runzhe Zhu, Ling Yin, Mingze Yang, Fei Wu, Yuncheng Yang, and Wenbo Hu. Sues-200: A multi-height multi-scene cross-view image benchmark across drone and satellite. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(9):4825–4839, 2023.