LEARNING ILLUMINATION INVARIANT FEATURES FOR LUNAR SOUTH POLE WITH DEEP LEARNING

Georgios Georgakis, Adnan Ansar
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109

Abstract. *A major challenge for vision-based applications for the Moon is large illumination variance that is caused by a combination of the sun position, topography, and lack of sky illumination due to the absence of an atmosphere. This diminishes the capability to recognize terrain features or landmarks that are critical for autonomous robotic operations, including spacecraft pinpoint landing and navigation. In this paper we explore deep learning for learning illumination invariant features in the challenging Lunar domain.*

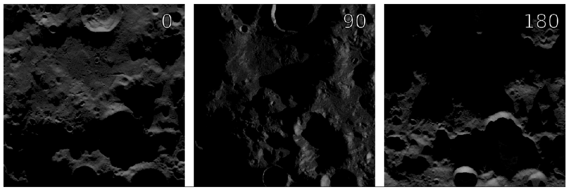

Figure 1. Renderings of the Lunar south pole with sun elevation angle of 2^o ***and various sun azimuth angles (top right of each map) that create extreme illumination discontinuities and large shadows.***

Introduction. The success of Terrain Relative Navigation (TRN) systems often hinges on consistent lighting conditions between the stored orbital imagery and the real-time conditions during landing. The Mars2020 mission benefited from landing during the late afternoon, closely matching the lighting conditions of its reference imagery. This alignment is crucial as variations in lighting can significantly alter terrain appearance, potentially compromising the TRN system's performance. Consequently, the operational window of the spacecraft is severely constrained to a brief amount of time where this alignment is present. This is particularly problematic on the Lunar south pole where the very low sun elevation angles create extreme illumination discontinuities and large shadows that are very difficult to match at different times of day (see Figure 1).

Such conditions can make methods like normalized cross-correlation (NCC) and descriptor-based feature matching less effective. While these classical methods (such as SIFT[1]) address scale and rotation variations between images with some level of robustness, they tend to struggle in low-texture situations and with visual appearance gaps that are usually the product of large illumination variances. In response to the shortcomings of these methods, deep learning approaches[2,3] produce more robust features via the exchange of spatial and visual information between keypoints, either through convolutions or Graph Neural Networks. More recently,[4,5] keypoint detection has been replaced with dense matching that allows the combination of information in a global context.

This paper explores a deep learning method for image matching, and introduces a training strategy to learn robust illumination invariant matching under very challenging lighting conditions. Out of the large number of methods in the literature we choose to adopt Local Feature Matching with Transformers (LoFTR)[4] for three reasons: 1) It is a modular and relatively simple architecture that can be easily built upon, 2) it uses Transformers[6] which are currently the main building block of many state-of-the-art models in computer vision, and 3) it has a good computational complexity vs performance trade-off as it is a rather modestly sized model (11.5M parameters). We view LoFTR as a proof of concept model for addressing extreme illumination variation in planetary applications with the opportunity for further development.

The LoFTR off-the-shelf model is pre-trained on the large in-the-wild Megadepth[7] dataset that offers challenging viewpoint, scale, and illumination variations. While the model trained on Megadepth shows some level of generalization to previously unobserved data, it is typically a challenge for data-driven methods to perform well outside the distribution of the data that they were trained on. This issue is even more pronounced in our settings as the extreme illumination changes present in the Lunar south pole imagery appear drastically different from typical outdoor images. Therefore, it is necessary to fine-tune the model to data from the Lunar domain. However, the main bottleneck is the absence of relevant, large-scale data that would allow the fine-tuning of deep learning models in this domain.

Synthetic Datasets with Lunar Imagery. There are several challenges in using real data from the Lunar domain for training. First, the orbit determination solution for the Lunar Reconnaissance Orbiter (LRO) lacks the accuracy with respect to ephemeris and attitude that would produce pixel-level correspondences necessary for supervision. Second, traditional Structure-from-Motion (SfM) methods, that are typically used to produce ground-truth, fail at the south pole due to the challenging lighting conditions. Finally, coverage over locations and illumination conditions is not complete. We aim to circumvent these issues by using Digital Elevation Maps (DEMs) of the Lunar south pole in simulation software that provide control over illumination conditions and camera poses. In this work, we present two simulated datasets and demonstrate that they are suitable for train-

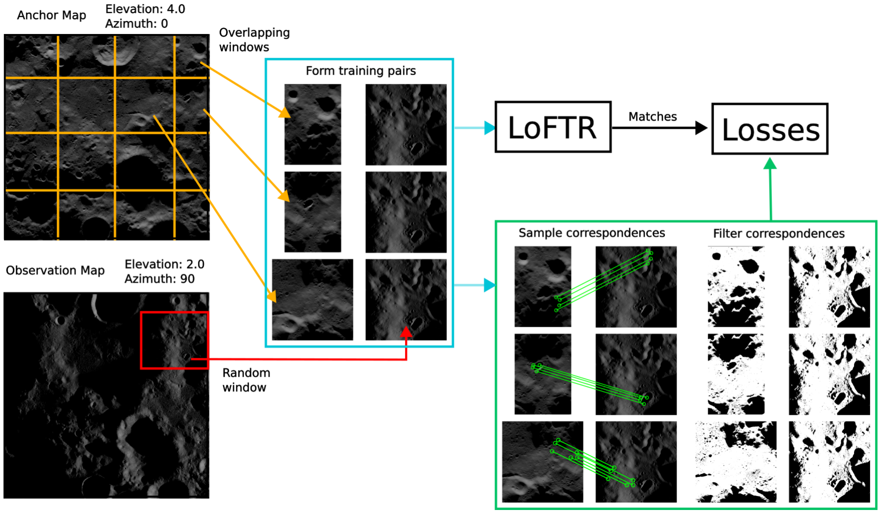

Figure 2. Procedure for creating Ortho-to-Ortho training pairs. We first divide every anchor map into a predetermined set of windows that we keep constant throughout this process. Then we randomly sample a window from an observation map and form training pairs with the overlapping windows in the anchor. This allows us to produce pairs with varying translation differences. Finally we sample ground-truth correspondences for each pair, which are filtered out from large shadow areas.

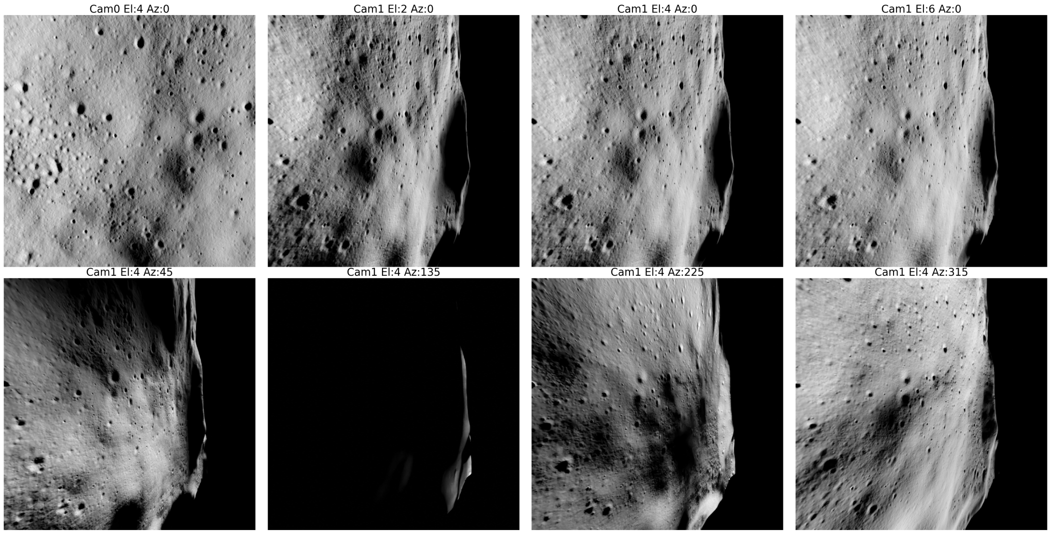

Figure 3. Examples of Perspective-to-Perspective pairs with sun angles shown on top of each image. The top left image (cam0) is paired with all other frames (cam1) that exhibit variations in illumination conditions. Note that frames with increased amount of shadows are discarded (second in bottom row).

ing illumination invariant image matching models.

Ortho-to-Ortho Pairs. First, we are interested in a quantitative evaluation of the model for its robustness to extreme illumination variations without the presence of other challenging factors (e.g., scale and viewpoint variations). To do so we use JPL-developed simulation software that renders Lunar imagery by hillshading DEMs from the Lunar Orbiter Laser Altimeter (LOLA[8]). We generate maps of size $150km \times 150km$ with $10m$ per pixel resolution from the Lunar south pole using orthographic projection. In order to simulate the illumination conditions in the Lunar domain, the maps are rendered with sun elevation angles between 0^o to 4^o with step of 1^o, and with sun azimuth angles between 0^o to 330^o with step of 30^o.

Our objective is to train the model to produce robust matches regardless of the sun angle differences between two images. A naive training approach would be to create training pairs with all illumination combinations and supervise the model to produce accurate matches. However, such an approach has two main problems: 1) it forces the model to fit a solution to any illumination variation, which is very challenging due to the very large number of combinations, and 2) it is impractical to generate sufficient

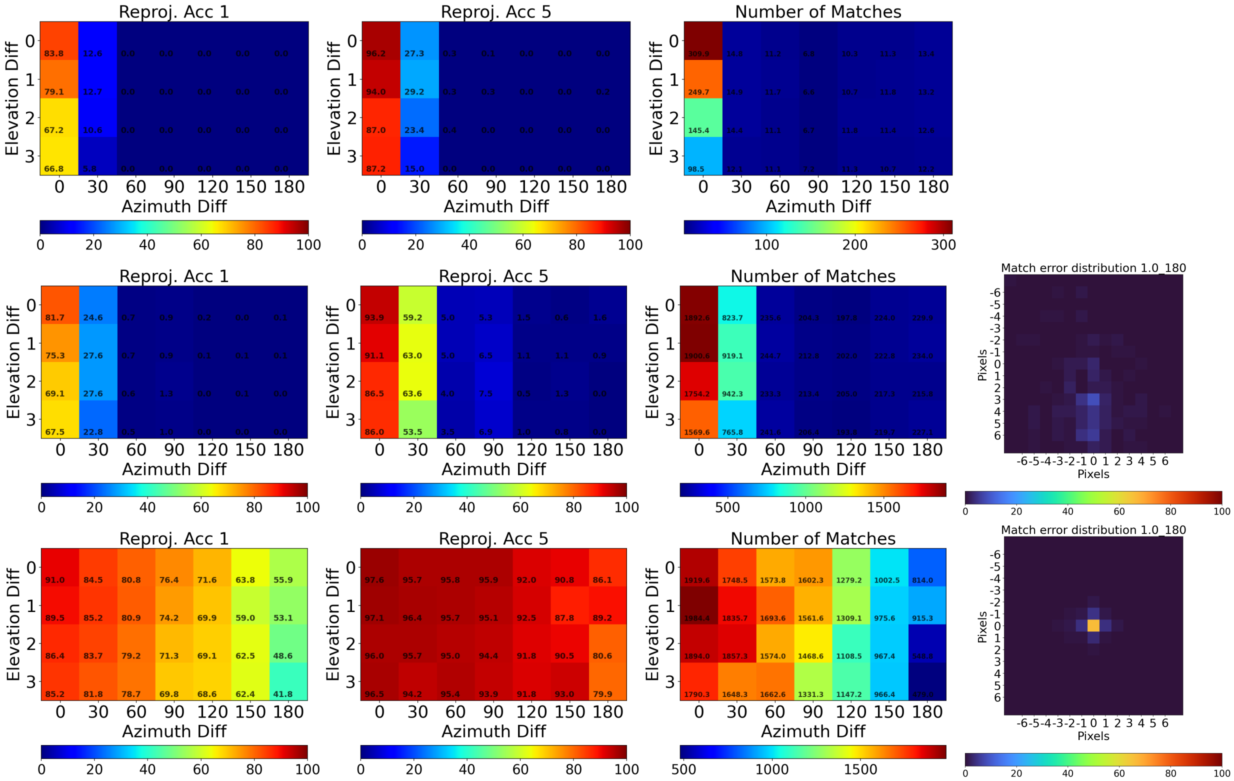

Figure 4. Quantitative comparison between SIFT (top row), off-the-shelf LoFTR model (middle), and trained LoFTR-fine (bottom) on the seen "180E" map of the Ortho-to-Ortho dataset. The last column shows the error distribution of the matches (in pixels) when azimuth and elevation difference is 180^o ***and*** 1^o ***respectively. The off-the-shelf model is showing a vertical bias which is fixed when the model is trained on Lunar imagery. SIFT had no matches with reprojection error less than 5 for this illumination configuration.***

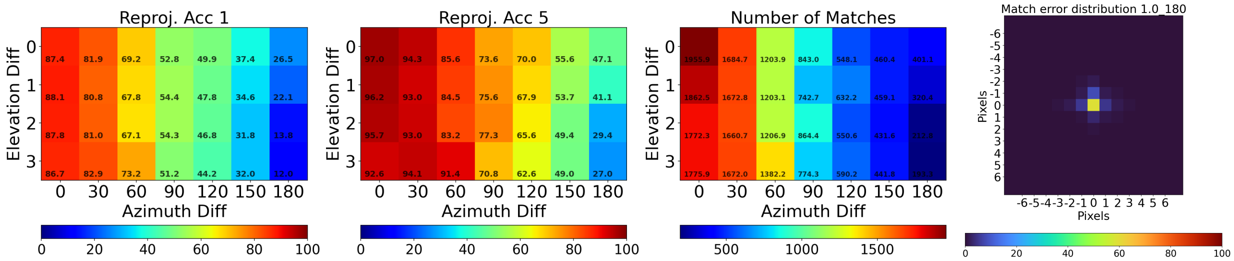

Figure 5. Performance of the trained LoFTR-fine on the unseen "0E" map of the Ortho-to-Ortho dataset. Even though there is a drop in matching accuracy compared to the evaluation on the seen "180E" map (see Figure 4), the model still produces a very large number of accurate matches, suitable for downstream tasks.

training pairs using all possible combinations. Instead, we create a training dataset over a subset of illumination conditions and show that the model is able to generalize to arbitrary illumination conditions.

Specifically, we define two sets of maps, the "anchors" and the "observations". Every training pair is created by sampling one map from each set. The "anchors" are rendered with a constrained set of illumination conditions (sun elevation angles of 2^o and 4^o and sun azimuth of 0^o, 90^o, 180^o, and 270^o for a total of 8 maps), while "observations" are unconstrained. Using the "anchor" map set significantly reduces the amount of training pairs required to learn illumination invariant features. Three of the anchor maps are shown in Figure 1. In practice, the training pairs are created by sampling windows over the maps in order to vary the overlap between the images. We generated approximately 25K training pairs. This procedure is illustrated in more detail in Figure 2.

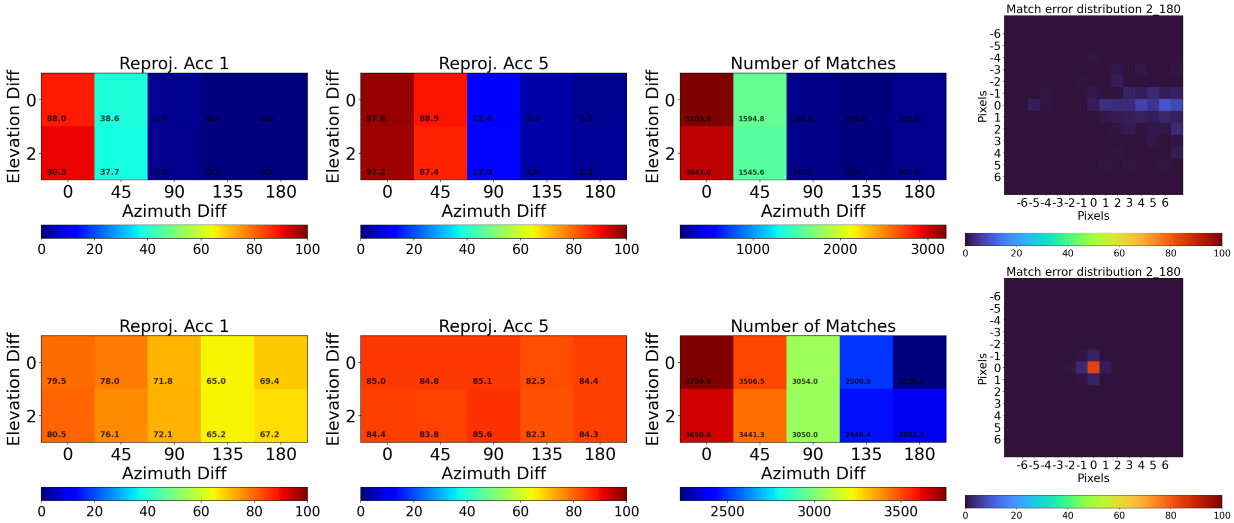

Figure 6. Results of the off-the-shelf LoFTR model (top), and trained LoFTR-fine (bottom) on the PerspA test set. Following a similar trend to the Ortho-to-Ortho experiments, the trained LoFTR is able to produce a large number of accurate matches even at the most difficult case of azimuth difference of 180^o***.***

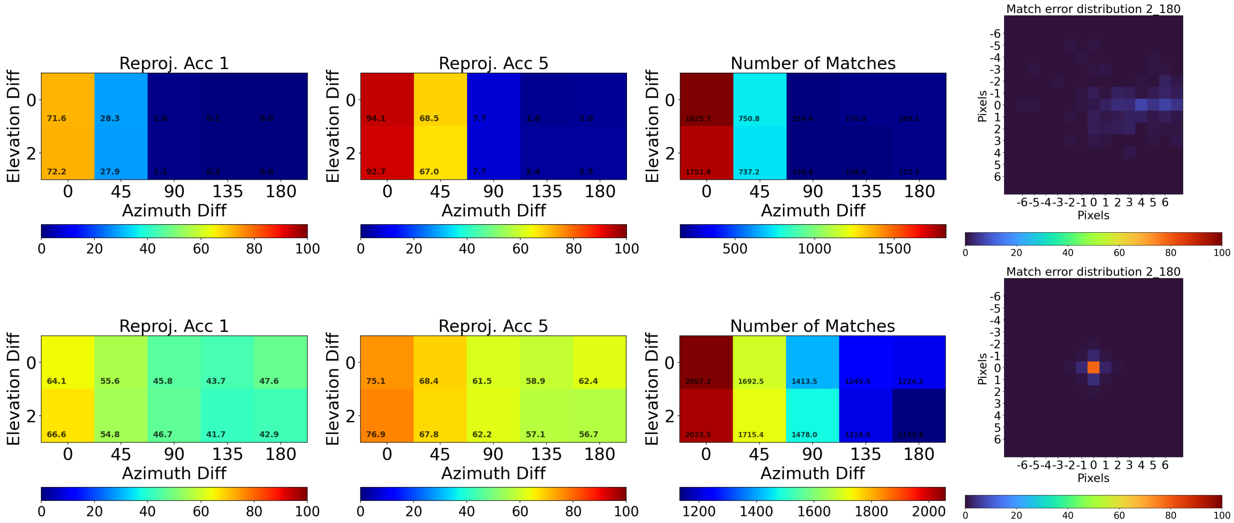

Figure 7. Results of the off-the-shelf LoFTR model (top), and trained LoFTR-fine (bottom) on the PerspB test set. PerspB contains larger attitude changes than the training set LoFTR-fine was trained on. Despite the drop in performance, the model is able to generalize on this test set and produce a large number of accurate matches.

Perspective-to-Perspective Pairs. While the Ortho-to-Ortho dataset is suitable for evaluating robustness to various illumination conditions, it is not representative of the conditions during entry, descent, and landing, where the spacecraft is using a downward-looking perspective camera without perfect pointing knowledge.

In order to generate a representative dataset, we employ the open-source simulation software Blender[9] that offers the physically-based rendering engine Cycles. Blender can produce photorealistic renders by simulating global illumination effects, by incorporating realistic shader functions and by modeling the camera's intrinsic parameters. In addition, it offers a Python interface, that allowed us to implement a Python wrapper to provide inputs to the simulation software.

In particular, we imported a $70km \times 70km$ with $10m$ per pixel resolution DEM from the Lunar south pole and defined two perspective cameras (*cam0*, *cam1*) with horizontal field-of-view of 100^o. During generation of the pairs, we kept *cam0* pointing downwards with no pose variation (i.e., $yaw = 0^o$, $pitch = 0^o$, $roll = 0^o$) and with constant sun angles ($elevation = 4$, $azimuth = 0$). Then

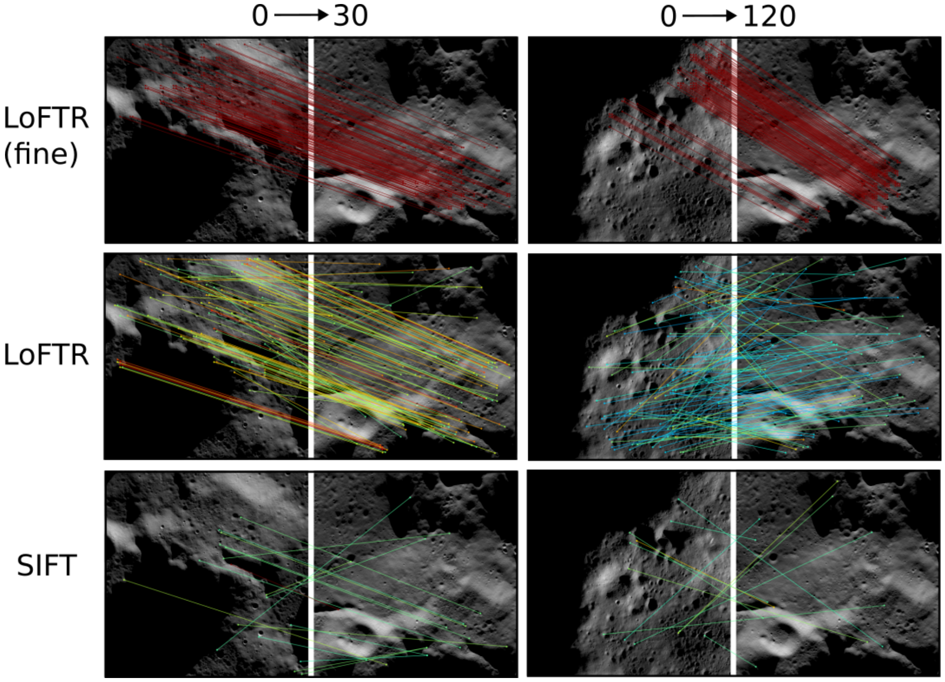

Figure 8. Matching examples on the Ortho-to-Ortho test set using trained LoFTR-fine, the off-the-shelf LoFTR model, and SIFT, a classical method. The images were rendered with sun elevation angle of 2^o and the sun azimuth angle difference is shown at the top for each pair. LoFTR-fine produces much more consistent and accurate matches between the images in spite of the large illumination difference.

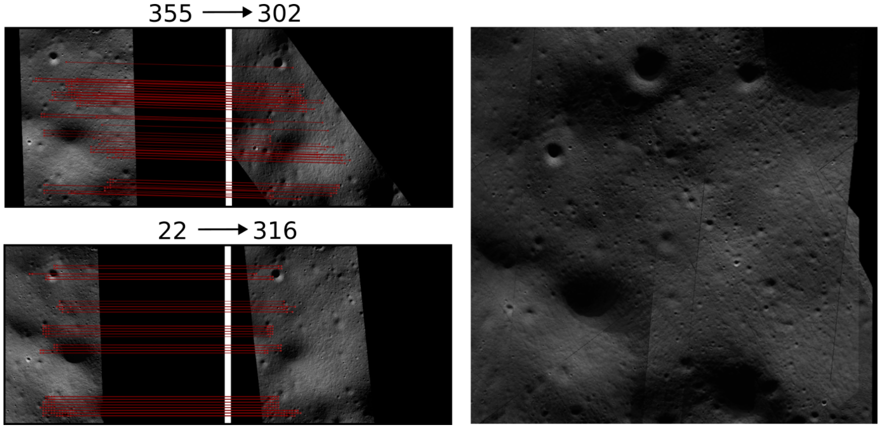

Figure 9. LoFTR is able to generalize to LROC-NAC images after it was trained on the synthetic Ortho-to-Ortho Lunar images. For the matching examples (left) we show the sun azimuth angle difference at the top of each pair. On the right, we demonstrate a mosaic created from NAC images captured at different illumination conditions.

we sample *cam1* poses in terms of orientation ($yaw = 0^o$, $15^o > pitch > 0^o$, $15^o > roll > 0^o$) and translation with respect to *cam0* of maximum of 500m. Altitude for both cameras is set to 3km. For every sampled $cam1$ we check the co-visibility with $cam0$ in order to ensure enough overlap between the paired renderings. Finally, every $cam1$ pose is rendered with different combinations of sun angles with elevation between 2^o to 6^o with step of 2^o, and with azimuth between 0^o to 315^o with step 45^o. Thus, for every $cam0$ frame, we pair it with 24 $cam1$ frames. In total we generate 8920 training pairs. A few examples of these pairs can be seen in Figure 3.

Experimental Evaluation. We present experiments on simulated and real data. In particular, we compare an off-the-shelf deep learning-based model (LoFTR[4]) to a traditional method (SIFT[1]) and to a fine-tuned model that we refer to as LoFTR-fine. The datasets are challenging and contain large illumination variations. The reported results show that traditional methods typically fail under these conditions, and that while the off-the-

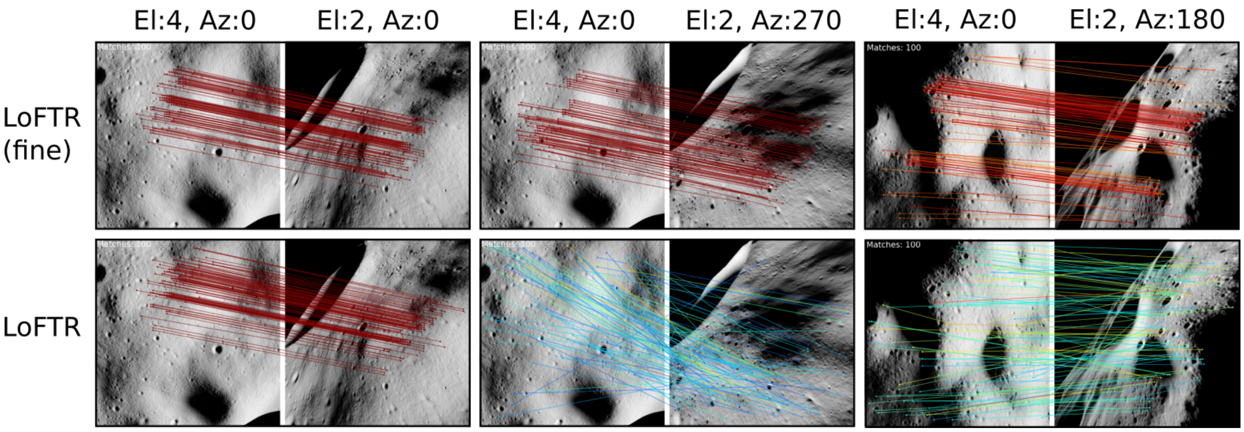

Figure 10. Matching examples on the PerspB test set that contains large attitude variations. Elevation and azimuth sun angles are shown above each image. LoFTR-fine is able to robustly find matches regardless of attitude and illumination change, while the off-the-shelf model starts failing in the presence of large illumination difference.

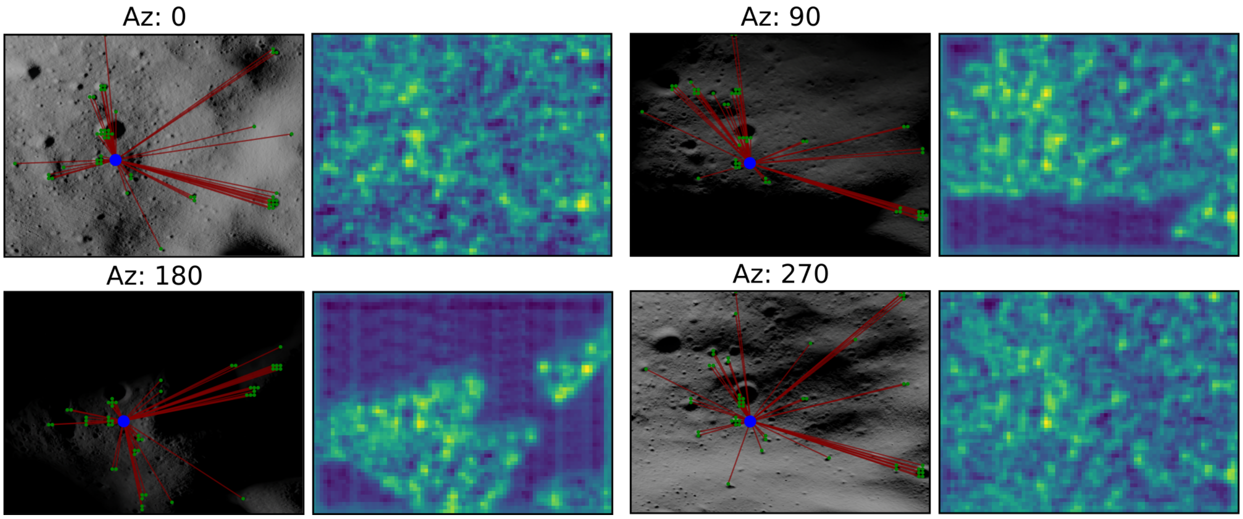

Figure 11. Visualization of how transformer blocks in LoFTR-fine incorporate context from the entire image when estimating the feature representation for a particular point. Here we show these for corresponding points (shown in blue) on images from the PerspA dataset with sun azimuth angles shown at the top of each image. The top 50 locations with the highest attention weights (green points) are shown to the left of each pair of images, while the entire heatmap is shown to the right. There are two important observations: 1) The model focuses on salient locations such as craters, and 2) These locations are consistent between varying lighting conditions.

shelf model shows some level of generalization, it does not consistently produce accurate results and is convincingly outperformed by its fine-tuned counterpart. For quantitative evaluation, we use Reprojection Accuracy (RA) that represents the percentage of matches under reprojection error of 1 and 5 pixels (RA-1, RA-5) along with the total number of matches produced. In the follow-up figures that show image matching examples, the color of the matches signifies the confidence of the model (red is more confident), which typically correlates with the accuracy of each match.

Synthetic Ortho-to-Ortho Dataset. We perform an analysis of the robustness of the matching on a wide range of illumination conditions on the Ortho-to-Ortho dataset. To do so, we generated a test set by randomly sampling locations from the same map used to create the training pairs, and from a previously unseen map. Testing on the unseen map is important to demonstrate that the model did not overfit on a specific region and that it can generalize to unseen regions of the Lunar south pole. We refer to the seen map as “180E” and the unseen as “0E”. The test sets were generated following the procedure in Figure 2 resulting in approximately 2700 pairs for each map.

Results of matching accuracies over 28 combinations of sun elevation and azimuth angle differences are shown in Figure 4 (seen “180E”) and Figure 5 (unseen “0E”).

LoFTR-fine outperforms the other baselines by a large margin and is able to produce a large number of matches with RA-1 (41.8% of 479 matches) even on the most difficult illumination condition of 180^o azimuth and 3^o elevation difference. On the other hand, the off-the-shelf model starts struggling when azimuth difference is beyond 90^o with RA-5 falling at or below 1.5%, and SIFT starts failing at 60^o of azimuth difference. Finally, Figure 8 illustrates matching examples between frames of large sun azimuth difference.

LRO-NAC Images. In this experiment we apply the LoFTR-fine model, that was trained on the synthetic Ortho-to-Ortho Lunar imagery, on pairs of real Lunar Reconnaissance Orbiter (LRO) Narrow Angle Camera (NAC) images captured at different sun azimuth angles. Figure 9 shows matching examples and a mosaic created from NAC images, demonstrating the ability of the fine-tuned model to generalize from synthetic to real imagery of the same domain. The images in the mosaic were registered with a simple blending function along the seams, with the illumination differences between the registered images easily noticeable.

Synthetic Perspective-to-Perspective Dataset. We test the robustness of the trained LoFTR model in the presence of pose differences between perspective images along with different illumination conditions. Two test sets were generated, *PerspA* contains 460 pairs and follows the generation parameters of the training set, while *PerspB* contains 507 pairs that are rendered with larger pose variation ($yaw = 0^o$, $45^o > pitch > 0^o$, $45^o > roll > 0^o$). Note that the model evaluated on *PerspB* is trained with the smaller pose variations of up to 15^o. We present quantitative results over 10 combinations of sun angle differences in Figure 6 and Figure 7 for the *PerspA* and *PerspB* test sets respectively. Similarly to the experiment on the Ortho-to-Ortho dataset, the off-the-shelf model starts failing when the azimuth difference is 90^o, while LoFTR-fine is able to learn an illumination invariant distribution and have a much smoother degradation of performance when the elevation and azimuth angle difference increases. Examples of matches in the presence of both lighting and attitude variations are shown in Figure 10.

Deep Feature Visualization. The rapid increase in complexity of deep learning-based models, has made these methods less interpretable[10] as often it is difficult for a user to explain why a model predicted a certain output. A popular method for providing some insight into the representation that a model has learned is to visualize features from intermediate layers of the model.[11] Since LoFTR uses transformers to incorporate information from the entire image and learn the features of individual pixels, it is meaningful to visualize the attention weights of these blocks. We provide a few examples in Figure 11.

Conclusion. We presented a study on learning illumination invariant features for the challenging domain of the Lunar south pole. Our work involved generating synthetic datasets suitable for training deep learning methods for image matching, and demonstrated the efficacy of both the datasets and the chosen image matching method on robust registration. Our experimental evaluation showed that deep learning models such as LoFTR are a viable solution for vision-based applications in the planetary exploration domain including those under extreme illumination conditions such as the Moon.

Acknowledgment. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). The primary funding source for this work was the LuNaMaps project under the NASA Space Technology Mission Directorate's Game Changing Development Program.

References.

[1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, Ieee, 1999.

[2] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 224–236, 2018.

[3] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4938–4947, 2020.

[4] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "Loftr: Detector-free local feature matching with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8922–8931, 2021.

[5] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, and M. Felsberg, "Roma: Revisiting robust losses for dense feature matching," *arXiv preprint arXiv:2305.15404*, 2023.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[7] Z. Li and N. Snavely, "Megadepth: Learning single-view depth prediction from internet photos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2041–2050, 2018.

[8] M. K. Barker, E. Mazarico, G. A. Neumann, D. E. Smith, M. T. Zuber, and J. W. Head, "Improved lola elevation maps for south pole landing sites: Error estimates and their impact on illumination conditions," *Planetary and Space Science*, vol. 203, p. 105119, 2021.

[9] B. O. Community, *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[10] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.