

AI-Driven Dim-Light Adaptive Camera (DimCam) for Lunar Robots

Ran Duan, Bo Wu*, Long Chen, Hao Zhou, Qichen Fan

Department of Land Surveying and Geo-informatics | Research Centre for Deep Space Explorations, The Hong Kong Polytechnic University, Hung Hom, Hong Kong – bo.wu@polyu.edu.hk

Keywords: DimCam, Deep Learning, Dim Light, Lunar Robots, Lunar South Pole

Abstract:

The past decade has been a boom in lunar exploration. China, India, Japan and other countries have successfully landed landers or rovers on the lunar surface (Wu et al., 2014, 2018, 2020; Prasad et al., 2023). Future missions to explore the Moon are focusing on the lunar south pole (Peña-Asensio et al., 2024). The solar altitude angle at the lunar south pole is extremely low, resulting in low solar irradiance and large areas often in dim light or shadows. The permanently shadowed regions (PSRs) at the lunar south pole are also likely containing substantial amounts of water ice (Li et al., 2018). Future lunar robots exploring the lunar south pole will need to operate in low light or shadowed regions, making camera sensors sensitive to the dim-light environments necessary for these robots. Common night vision sensors usually use near-infrared cameras. However, sensors based on passive infrared technology have image resolution limited by several factors, including the intensity of infrared radiation emitted by the object, the sensitivity of the camera, and the performance of the optical system. For instance, thermal imagers typically have a resolution of 388*284 pixels only.

We have developed a dim-light adaptive camera (DimCam) that is ultra-sensitive to the varying illumination conditions driven by AI to achieve high-definition imaging of 1080P or above, for future lunar robots operating in shadows or dim-light regions. The DimCam integrates two starlight-level ultra-sensitive imaging sensors connected by a rigid base to provide stereo vision in low illumination environment. An AI edge computing unit is embedded inside the DimCam to adaptively denoise and enhance image quality. The AI module uses an end-to-end image denoising network to identify and remove noises in the images more accurately by utilizing depth information from the stereo sensors. Compared with traditional monocular denoising algorithms, the denoising network based on stereo vision can significantly improve denoising effects and efficiency by enhancing the signal-to-noise ratio of the data input in the front end. The superposition of overlapping scenes can be regarded as a delayed exposure. Concurrently, the residual analysis of the aligned images aids in noise identification. In addition, for pixels obscured by noise, more accurate pixel values can be restored through interpolation or replacement using depth information obtained from the stereo sensors. Subsequently, a pre-trained lightweight deep network modified from Zero-DCE (Guo et al., 2020) is used for image quality enhancement in terms of brightness and contrast, providing high-quality images even in low-light environments for subsequent applications, such as positioning and navigation of robots, 3D mapping of the surrounding environment, and autonomous driven. We have tested the DimCam in a simulated environment in the laboratory, and the results show that the DimCam has promising performances and great potential for various applications.

* Corresponding author. Email: bo.wu@polyu.edu.hk

1. Introduction

The focus of future lunar missions is gravitating toward the lunar south pole, an area characterized by an extremely low solar altitude angle and predominantly dim-light conditions, factors that have contributed to its largely unexplored status (Peña-Asensio et al., 2024). The lunar south pole is believed to contain substantial quantities of water ice, especially in its perpetually shadowed regions (PSRs). This water ice could serve as a vital life support resource. It could also be used to produce hydrogen and oxygen for rocket fuel. These distinctive environmental and geological attributes position the lunar south pole as an ideal candidate for the establishment of a future lunar base. However, to actualize this, the development of novel vision sensors that can operate in shadowy or dimly lit regions stands out as one of the demanded technologies.

Improving the quality of images taken under the dimly lit conditions at the lunar south pole, while simultaneously preserving their natural colors, presents a significant challenge. This complexity stems from the need to recover information that was either lost or never captured in the first place, a characteristic of typical inverse problems. Traditional image enhancement methods, such as histogram equalization, gamma correction, and Retinex-based techniques, have been extensively utilized (Chang et al., 2018). However, these methods often fall short of delivering satisfactory results under challenging lighting conditions, such as those encountered at the lunar south pole. They tend to either over-enhance or under-enhance the images, leading to problems like color distortion or loss of detail. Machine learning-based methods have also been investigated, with models being trained on specific datasets (Li et al., 2021). However, these models frequently struggle to perform optimally on images captured by different cameras due to variations in individual camera processing techniques. The task of compiling a comprehensive set of paired images for a multitude of devices is not only formidable but also expensive. Moreover, the dependence on a single accurate lighting condition for each pair limits the model's ability to accurately learn how to adjust the brightening factor. This underscores the need for more robust and adaptable image enhancement techniques for images captured under the dimly lit conditions of the lunar south pole. A review of existing literature reveals that numerous studies have reported promising results using paired datasets comprising both well-lit and poorly-lit images (Koohestani et al., 2023). Deep learning has gained significant popularity in recent years, particularly for its robust feature representation and non-linear mapping capabilities. These attributes have found numerous applications in the field of image processing, leading to substantial advancements. The research by (Lore et al., 2017) was among the pioneers of leveraging deep learning for enhancing low-light images, resulting in the development of LLNet. They introduced a method based on deep autoencoders, which identifies signal features from low-light images and adaptively brightens them in high dynamic range scenarios without excessively amplifying or saturating the lighter sections. They demonstrated that a variant of the stacked sparse denoising autoencoder could learn from synthesized dark and noise-added training examples. This learning process enabled the adaptive enhancement of images from natural low-light environments and those degraded by hardware limitations. Their experiments on real low-light images validated the effectiveness of models trained with synthetic data. Shen et al. (2017) proposed a low-light image enhancement algorithm that combines convolutional neural networks (CNN) and Retinex theory. They demonstrated that the Multi-Scale Retinex (MSR) is

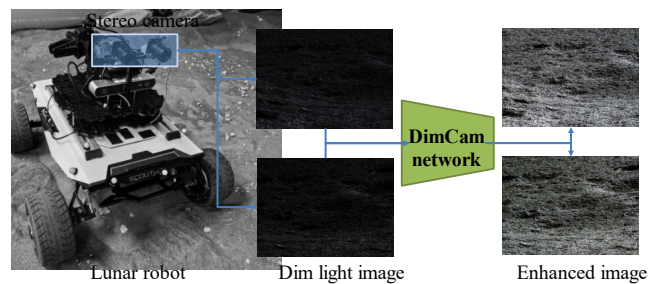


Figure 1. Dim-light adaptive camera (DimCam) that uses stereo vision and image enhancement network for lunar robots.

equivalent to a feed-forward convolutional neural network with varying Gaussian convolution kernels. Inspired by MSR, they introduced the MSR-net to learn the end-to-end mapping between dark and bright images directly. Unlike traditional methods, the majority of parameters in this model are optimized through backpropagation, which is a significant advantage. Cai et al. (2018) introduced a method for training single-image contrast enhancement using CNNs. Their approach outperforms existing methods by revealing more image details. A notable work RetinexNet is one of the most successful models. Its fundamental principle involves using convolutional layers, non-linear mapping, and adaptive filters to decompose the original low-light image at multiple scales and enhance the image at the feature level. But it also requires preprocessing the image using prior knowledge to eliminate interference factors such as uneven illumination and noise. Currently, most of these algorithms are designed for everyday scenarios, with abundant data available for network training. However, data for planetary robots exploring dim light environments and algorithms specifically designed for such purposes are still few and far between.

Dim-light environment imaging has always been a popular research topic due to its huge practical application demands, and there are many mature industrial implementation solutions. However, most of the current solutions are based on infrared imaging or thermal imaging. Infrared imaging often uses additional LEDs to enhance the illumination of the environment in the infrared band. On the other hand, thermal imaging does not require fill light, but its imaging resolution is very limited. With the rapid advancement of AI technology in recent years, using deep neural networks to directly restore color images from the visible light band from low-light to a comfortable illumination range for human eyes has become an important research direction to solve the problem of enhancing perception in low-light environments. However, restoring high-definition images directly from low-light images captured by ordinary visible-light cameras faces physical limitations. This is because most camera sensors have a low sensitivity to dim light, and the noise inherent in the camera system itself results in a very low signal-to-noise ratio (SNR) of image information under low-light conditions. Directly enhancing the original image will also amplify these noises, leading to a loss of image details and degrade the performance of subsequent image processing algorithms, such as scene segmentation and object detection.

This paper presents an AI-driven dim-light adaptive camera (DimCam) based on a stereo vision system. The DimCam uses the Sony IMX291 CMOS as the imaging chip to address the issues of

camera sensitivity and system noise. The IMX291 chip supports the capture of approximately 2 million pixels, with a clarity up to 1080P, and it also supports WDR (Wide Dynamic Range) for enhanced image capture in varying lighting conditions, which is particularly important for applications especially in scenarios requiring strong night vision. The configuration of stereo vision provides data source-based guarantees for enhancing the signal-to-noise ratio of images and repairing images degraded by noise. We use stereo vision to emulate the effect of a photo taken with a longer exposure. This is achieved by eliminating noise, boosting brightness, and maintaining the integrity of natural colors. For the algorithm part, we propose a DimCam deep network, which is an improved version of Zero-DCE (Guo et al., 2020) model and capable of process stereo vision. The improvements include the ability to understand depth information of the scene and the ability to eliminate the impact of noise through images from both the left and right views, thereby achieving better enhancement results.

2. AI-Driven Dim-Light Adaptive Camera (DimCam)

We show the overview of the proposed system in Fig. 2. We use two starlight-level low-light camera modules as a stereo vision system. The images captured by them are first synchronized through software. Then, we use the camera parameters to perform rectification and homography transformation on the stereo views to extract the overlapping areas. Based on the overlapped region, we extract noise masks from the stereo image pairs. The pre-processed data are subsequently fed into the proposed deep neural network DimCam, which incorporates two parallel Zero-DCE models to accommodate stereo vision input, and then feed the output tensors to a transformer block with cross attention so that the network can learn the correlation of stereo vision images. We add a pre-trained DepthNet (Kumar et al., 2018) to the output end of our network, which can generate depth maps from enhanced stereo vision images. This adaptation enables the network to use depth information from stereo vision or laser measurements (Wu et al., 2015) for better denoising images.

2.1 Camera System

Our camera system adopts two SONY 2MP IMX291 color COMS sensors, achieves 60FPS in 1080P resolution, making it easy to synchronize in the software level. We use two CSI interfaces of the NVIDIA Jetson Orin NX edge AI computing unit as the receiver for the stereo camera system. When the program receives image data from the two cameras with a time difference smaller than the

pre-defined threshold, it is considered as synchronized stereo image pairs.

Light condition	Lux
Sunlight	107,639
Twilight	10.76
Full Moon	0.108
Starlight	0.0011
Overcast Night	0.0001

Table 1. The illuminance levels (in Lux) under different outdoor light conditions

The IMX291 COMS sensor has an 80dB wide dynamic range, can capture the minimum illumination as low as 0.001Lux. The illuminance levels (in Lux) under different outdoor light conditions are given in Table 1. For the camera in a normal working illuminance range, the minimum illumination value is between 0.1-1Lux. Some night vision cameras can reach the minimum illumination value of about 0.01Lux. In order to adapt the dim light conditions, the camera should have a very low minimum illumination value.

2.2 Noise Mask Extraction

Since the input images in our scenario are often extremely low-light images, we first suppress the areas with high image noise through a noise mask before enhancing the brightness, in order to improve the signal-to-noise ratio. Denote the rectified stereo image pairs I_l and I_r , where we only consider the overlap region. Assume that the noise N is directly superimposed on the original image:

$$I_l = \hat{I}_l + N_l,$$

where \hat{I}_l is the original scene of left view I_l and $\hat{I}_l = H(\hat{I}_r)$ or $\hat{I}_r = H^{-1}(\hat{I}_l)$, $H(\cdot)$ is the homography transformation. Then, we can obtain the noise masks for left and right views respectively by calculating the residuals of the aligned images:

$$M_l = \text{Bin}(I_l - H(I_r)) = \text{Bin}(N_l - H(N_r)),$$

$$M_r = \text{Bin}(I_r - H^{-1}(I_l)) = \text{Bin}(N_r - H^{-1}(N_l)),$$

where $\text{Bin}(\cdot)$ is the image binarization. Subsequently, we enhance the masked images through DimCam network.

2.3 DimCam Network

The Zero-DCE (Zero-Reference Deep Curve Estimation) algorithm is an efficient low-light enhancement method that utilizes a

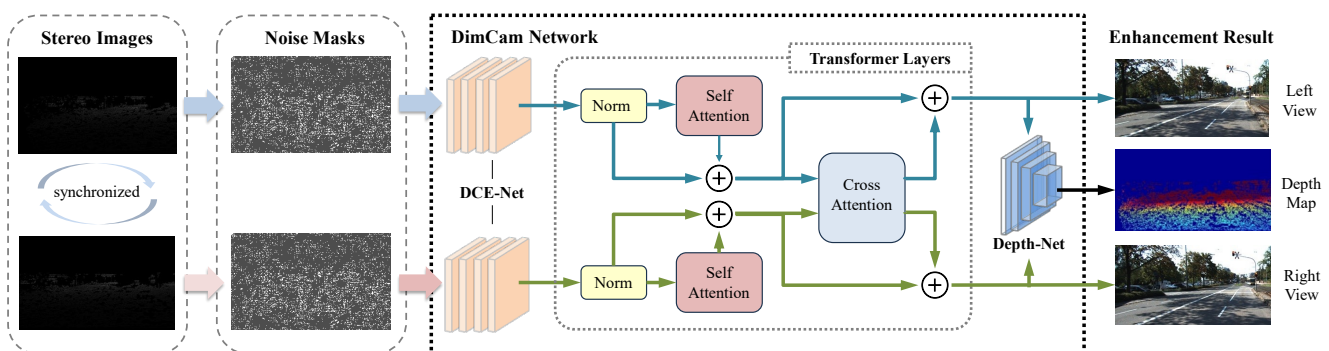


Figure 2. Workflow of the DimCam system.

lightweight deep convolutional neural network, DCE-Net, to predict a pixel-wise high-order curve for dynamic range adjustment of input images. It requires no paired or unpaired data for training, preventing overfitting, and can handle various illumination conditions, making it widely applicable in real-world scenarios. To modify the Zero-DCE network for stereo vision, we design two parallel input branches for images from each camera perspective. During the enhancement process, it is crucial to ensure that the stereo consistency between the two images is not disrupted. To achieve this, we integrate Vision Transformer (ViT) layers in series within each branch and introduce a cross-attention mechanism. Additionally, we incorporate a stereo consistency loss term into the loss function to encourage the network to maintain correspondence between the left and right images during enhancement. The stereo consistency loss is given by:

$$L_{stereo} = \frac{1}{K} \sum_{i=1}^K \sum_{j \in \Omega(i)} (|I_l(i) - I_l(j)| - |H(I_r)(i) - H(I_r)(j)|),$$

where K is the number of pixels, i is the index of pixels. $\Omega(i)$ is the 4-neighborhood of the i -th pixel. I_l and I_r are the input stereo images, $H(\cdot)$ is the homography transformation. At this point, the two images have been enhanced while ensuring illuminance consistency in stereo vision. To further improve image quality, we input the enhanced stereo images into DepthNet for depth map estimation. We utilize the depth map by adding terms in

the loss function to enhance the perception of the continuity of the scene in three-dimensional space. The fundamental principle behind the design of these loss terms is to make consecutive pixels as smooth as possible in three-dimensional space. The depth consistency loss is given as follows:

$$L_{depth} = \frac{1}{K} \sum_{i=1}^K \sum_{j \in \Omega(i)} (|Y(i) - Y(j)| - |D(i) - D(j)|),$$

where Y and D are the enhanced image and depth map, respectively. We choose Y from one of the enhanced stereo images as a fixed reference frame which the depth map is registered.

2.4 Network Training

We utilize the proposed low-light stereo camera system to capture images of various scenes under different lighting conditions for training and testing purposes. Both Zero-DCE and DepthNet employed pre-trained models, while the network that be trained and tuned is the transformer block with cross-attention. The main objective of the training is to improve the Zero DCE model, which can only enhance single images, with additional stereo and depth perception and correlation capabilities.

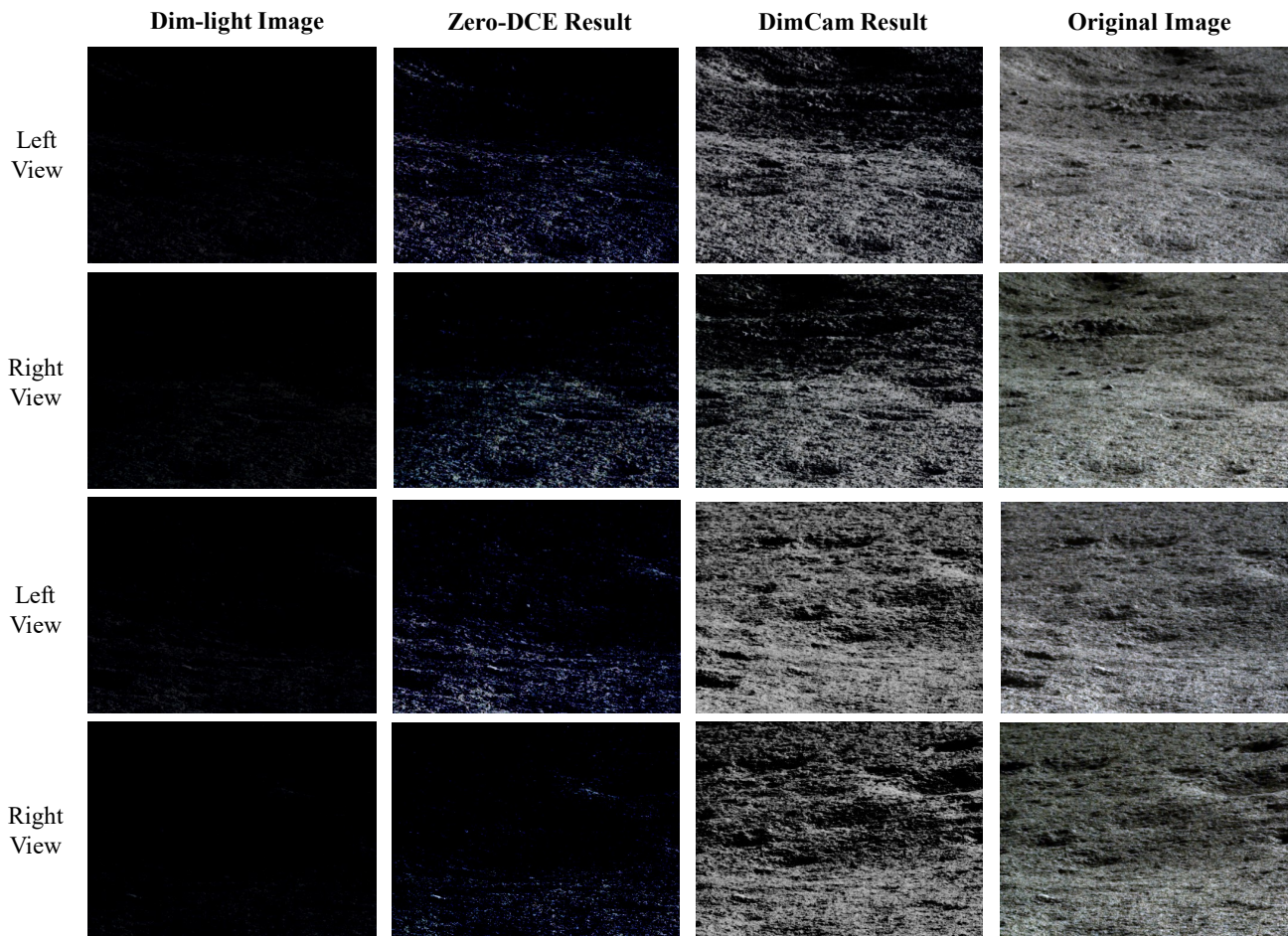


Figure 3. Test on simulated dim-light images from real images of the lunar surface.

3. Experimental Evaluation

Due to the lack of low-light scene images collected by lunar rovers, we conducted tests on simulated dim light images of real moon surface images. These test images were generated using a degradation model. This degradation model includes a sharp decrease in simulated brightness and contrast, as well as the addition of random noise with similar intensity to the degraded light amplitude. The original images were captured by the Chang'E-4 rover. We extracted overlapping high-definition lunar surface images taken from different angles to simulate stereo vision. We used these simulated degraded images as the input to test the performance of the proposed DimCam network and used the state-of-the-art model Zero-DCE as a control test. The sampled experimental results are shown in Fig. 3. The results indicate that our proposed model significantly improves the quality of image enhancement by integrating stereoscopic vision information. Particularly in extremely low-light conditions, our model has shown significant improvement compared to the state-of-the-art single-image-based model Zero-DCE.

To further validate the proposed model, we deployed the entire system on a rover platform and conducted tests in a simulated environment in the laboratory as Fig. 5 shows. Although the simulated scenario we use for testing the algorithm is the surface of Mars, it does not significantly impact the evaluation of the system's adaptive performance in dim light. Moreover, it provides a simpler method to assess the algorithm's ability to restore the colors of real objects. In addition to the proposed system, we also installed an industrial color night vision camera at a fixed position to monitor the entire test site. We simulated a dim light environment by turning off all light sources in the laboratory environment, leaving only some tiny LED lights from equipment and weak EXIT sign lights. Under extremely low-light conditions, images captured by regular cameras barely display any information recognizable to the human eye. However, our DimCam system can effectively output images with clear scene information discernible to the human eye.

Furthermore, we conducted an image segmentation test on the enhanced images, and the results demonstrate that the enhanced images can be effectively recognized by other image processing models. This illustrates its great potential in improving the intelligent scene perception capability of autonomous robots for planetary exploration missions under extremely low-light conditions.

It's worth mentioning that even a sensitive CMOS like IMX291 does not fundamentally solve the dim light problem in the real application of lunar exploration. The lux level in PSRs at the lunar south pole would essentially be zero, as there is no direct sunlight and very little reflected light. In the actual exploration process, active light sources used for illumination are indispensable. These active light sources can help visualize scenes at close range. However, for distant sensing and visualization, as well as considering the energy-saving of the light source, the proposed work could be a necessary tool.

4. Conclusions and Discussion

The low light conditions in the lunar south pole regions pose a challenge for lunar robot automation. To address this, a dim light adaptive camera (DimCam) system has been proposed. The DimCam uses two low light sensitive imaging sensors (SONY IMX291 CMOS) for stereo perception and a DimCam deep network for stereo image denoising and quality enhancement. The images captured are synchronized and processed to extract overlapping areas and depth information. This data is then fed into an improved version of the Zero-DCE neural network, which has been adapted to better correlate depth information and noise masks with the denoised images via a Transformer block and a DepthNet. This innovative approach allows for high-quality imaging in low-light conditions, such as those found at the lunar south pole, enhancing the capabilities of lunar robots in these challenging environments. The experimental results have shown promising

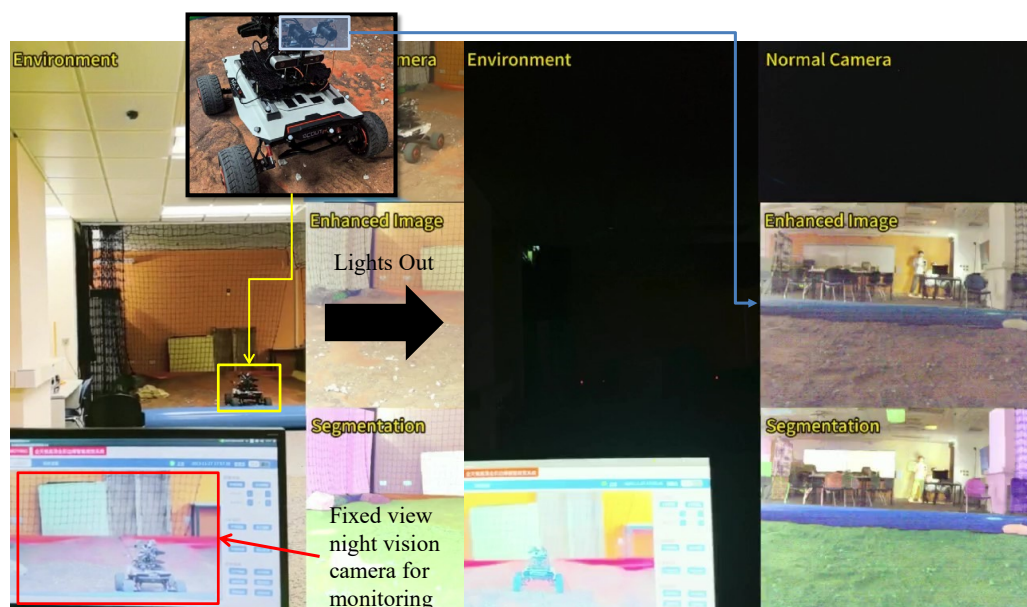


Figure 4. Real-time on-board testing of DimCam in a laboratory simulation environment.

outcomes, indicating the DimCam's potential for various applications in lunar exploration.

Acknowledgements

This work was supported by grants from the Research Grants Council of Hong Kong (RIF Project No: R5043-19, Project No: PolyU 15215822, Project No: 15219821) and grants from the Hong Kong Polytechnic University (Project No. P0046112, Project No. P0044685).

References

Cai, J., Gu, S., and Zhang, L., 2018. Learning a deep single image contrast enhancer from multi-exposure images. In *IEEE Transactions on Image Processing*, 27(4), (pp. 2049-2062).

Chang, Y., Jung, C., Ke, P., Song, H., and Hwang, J., 2018. Automatic contrast-limited adaptive histogram equalization with dual gamma correction. *Ieee Access*, 6, (pp. 11782-11792).

CS Kumar, A., Bhandarkar, S. M., and Prasad, M., 2018. Depthnet: A recurrent neural network architecture for monocular depth prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 283-291).

Gläser, P., Oberst, J., Neumann, G.A., Mazarico, E., Speyerer, E.J. and Robinson, M.S., 2018. Illumination conditions at the lunar poles: Implications for future exploration. In *Planetary and Space Science*, 162, (pp. 170-178).

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S. and Cong, R., 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1780-1789).

Hernandez-Juarez, D., Chacón, A., Espinosa, A., Vázquez, D., Moure, J.C., López, A.M., 2016. Embedded real-time stereo estimation via semi-global matching on the GPU. *Procedia Computer Science* 80, 143-153.

Koohestani, F., Karimi, N., & Samavi, S., 2023. Revealing Shadows: Low-Light Image Enhancement Using Self-Calibrated Illumination. *arXiv preprint arXiv:2312.15199*.

Li, C., Guo, C., Han, L., Jiang, J., Cheng, M. M., Gu, J., & Loy, C. C., 2021. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(12), (pp. 9396-9416).

Li, S., P. G. Lucey, R. E. Milliken, P. O. Hayne, E. Fisher, J. Williams, D. M. Hurley, R. C. Elphic, 2018, Direct evidence of surface exposed water ice in the lunar polar regions, *PNAS*, (pp. 115).

Lore, K. G., Akintayo, A., and Sarkar, S., 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61, (pp. 650-662).

Prasad, D., M. Bhatt, Amitabh, G Ambily, S. Sathyan, D. Misra, N. Srivastava, A. Bhardwaj, 2023. Contextual characterization study of Chandrayaan-3 primary landing site, *Monthly Notices of the Royal Astronomical Society: Letters*, 526, (pp. L116–L123).

Peña-Asensio, E., Sutherland, J., Tripathi, P., Mason, K., Goodwin, A., Bickel, V. T., Kring, D. A., 2024. Automated astronaut traverses with minimum metabolic workload: Accessing permanently shadowed regions near the lunar south pole. *Acta Astronautica*, 214, (pp. 324-342).

Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., and Ma, J., 2017. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*.

Wu, B., Fei Li, Lei Ye, Si Qiao, Jun Huang, Xueying Wu, He Zhang, 2014. Topographic Modeling and Analysis of the Landing Site of Chang'E-3 on the Moon, *Earth and Planetary Science Letters*, 405, (pp. 257-273).

Wu, B., Tang, S., Zhu, Q., Tong, K., Hu, H., Li, G., 2015. Geometric integration of high-resolution satellite imagery and airborne LiDAR data for improved geopositioning accuracy in metropolitan areas, *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 139-151.

Wu, B., J. Huang, Y. Li, Y. Wang, J. Peng, 2018. Rock Abundance and Crater Density in the Candidate Chang'E-5 Landing Region on the Moon, *Journal of Geophysical Research - Planets*, 123(12), (pp. 3256-3272),

Wu, B., F. Li, H. Hu, Y. Zhao, Y. Wang, P. Xiao, Y. Li, W. C. Liu, L. Chen, X. Ge, M. Yang, Y. Xu, Q. Ye, X. Wu, H. Zhang, 2020. Topographic and Geomorphological Mapping and Analysis of the Chang'E-4 Landing Site on the Far Side of the Moon, *Photogrammetric Engineering & Remote Sensing*, 86(4), (pp. 247-258).