HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Artificial-to-Real Domain Adaptation for Lunar Landscape Semantic Segmentation

Jeonghyeon Seo[1] and Donghyuk Yang[#]

[1]Korean Mink Leadership Academy, Republic of Korea
[#]Advisor

ABSTRACT

From Apollo 11 to India's Chandrayaan-3, humans have put a vast amount of effort into exploring the moon. While analysis on the lunar surface provides valuable information for humans, costly and risk-taking problems of lunar exploration resulted in such scant amount of data accessible to the lunar landscape to this day. Nonetheless, study on lunar landscape remains pivotal for probing resources, detecting hazards, and studies on moon evolution. In addition, analysis on the lunar surface is the ground for future development on the moon as a potential site of resources, identifying safe landing sites, and further civilization. In this study, I propose a machine learning-based lunar landscape image semantic segmentation system. Given lunar landscape images, the proposed method outputs semantic segmentation maps that separate different types of objects such as the ground, sky, and rocks. These segmented objects potentially provide valuable insights for guiding autonomous investigation rovers. The proposed method is trained on a synthetic lunar landscape dataset and evaluated on both synthetic and real lunar landscape samples. Through comprehensive experiments, it is demonstrated that the proposed method exhibits domain adaptation capabilities, achieving state-of-the-art performance on real lunar landscape images.

Introduction

From Apollo 11(Gisler and Sornette 2009) to India's Chandrayaan-3 (Sinha et al. 2023), humans have put a vast amount of effort into exploring the moon. While analysis on the lunar surface provides valuable information for humans, costly and risk-taking problems of lunar exploration resulted in such scant amount of data accessible to the lunar landscape to this day. Nonetheless, study on lunar landscape remains pivotal for probing resources, detecting hazards, and studies on moon evolution. In addition, analysis on the lunar surface is the ground for future development on the moon as a potential site of resources, identifying safe landing sites, and further civilization.

In this research paper, I present a novel domain adaptation technique aimed at reducing the distribution gap between synthetic and real datasets. I introduce a gradient reversal layer to reverse the gradient during backpropagation. This gradient reversal layer guarantees that the trained network focuses on generating consistent activation maps. The structure of this research paper is as follows: Section 2 provides background knowledge to facilitate an understanding of the research. Section 3 explains every detail of the proposed approach, including implementation specifics. Section 4 offers comprehensive experimental results and insights into the proposed approach, while Section 5 provides a summary of the paper.

Related Work

Lunar Landscape Dataset

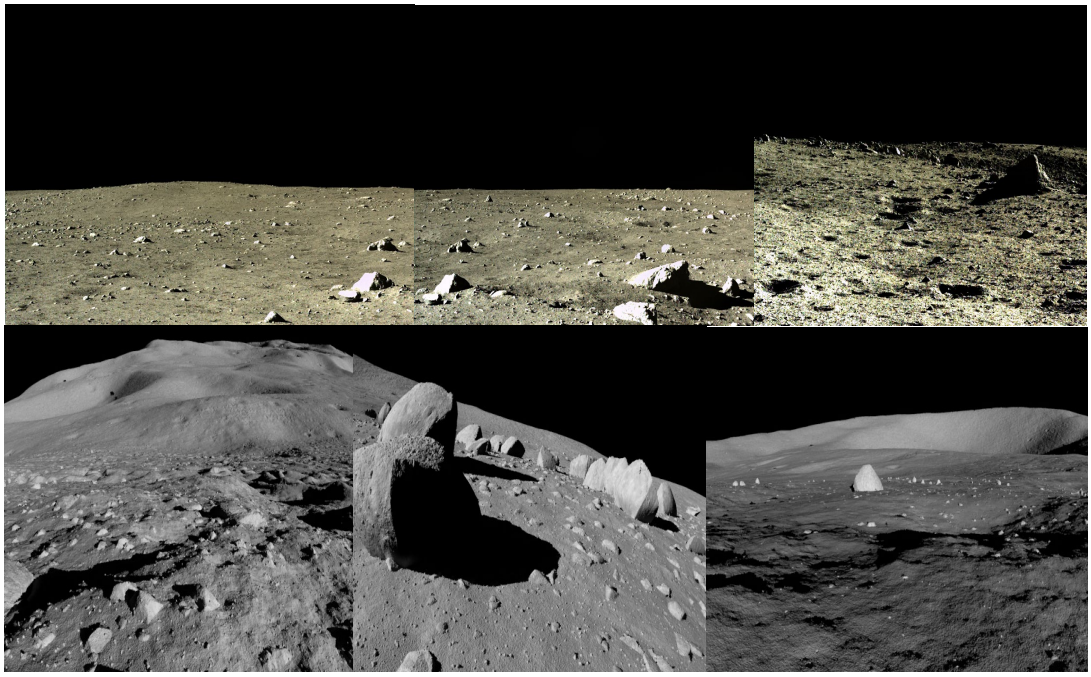


Figure 1. Samples of lunar landscape dataset (real images above and synthetic images below)

Figure 1 displays samples from the lunar landscape dataset. The focus of this research lies in the domain of image semantic segmentation specifically tailored for lunar landscapes, a critical aspect for advancing the capabilities of autonomous rovers and enhancing resource investigation on celestial bodies. Creating an accurate machine learning model demands a substantial dataset which allows the algorithm to learn diverse patterns in lunar landscape imagery. However, the practicality of acquiring lunar landscape image samples is severely constrained by the current limitations in space exploration technology (very costly). Given the current inability to physically visit the Moon at will, constructing a dataset solely from real lunar images becomes an unattainable task.

To address this problem, a practical approach involves the careful synthesis of artificial lunar landscape images. This synthetic dataset becomes a feasible alternative which enables the training of machine learning models in a controlled environment. Nonetheless, a notable hurdle arises from the inherent differences in distribution between real and synthetic datasets. The dissimilarities in visual characteristics, lighting conditions, and surface textures impact the model's ability to generalize effectively, potentially leading to suboptimal performance in real-world lunar scenarios.

The objective of this research is to minimize the distribution gap between the synthetic and real datasets. I propose a novel training approach designed to enforce the trained model to extract consistent image features. A thorough explanation of the proposed approach will be presented in Section 3.

Image Semantic Segmentation

Figure 2. Samples of image semantic segmentation (Lin et al. 2017)

Figure 2 illustrates the output of the image semantic segmentation system. Image semantic segmentation is a computer vision task that divides an inputted image into semantically meaningful regions. In simpler terms, the goal is to understand and categorize the content of an image at the pixel level. Unlike object detection, which provides bounding boxes around objects, semantic segmentation aims to precisely classify each pixel in the image.

Image semantic segmentation finds applications across various domains due to its ability to provide detailed and precise understanding of image content. In medical imaging, semantic segmentation aids in the identification and delineation of anatomical structures and abnormalities (Taghanaki et al. 2021). It is used in tasks such as tumor detection, organ segmentation, and disease diagnosis. Analyzing satellite images for land cover classification, urban planning, and environmental monitoring benefits from semantic segmentation (Neupane et al. 2021). It allows the identification of different types of terrain, vegetation, and land use.

Proposed Method

In this section, I provide a comprehensive overview of the proposed lunar landscape semantic segmentation method. The subsequent chapters are organized into three parts: baseline image semantic segmentation, the proposed domain adaptation method, and implementation.

Baseline Image Semantic Segmentation

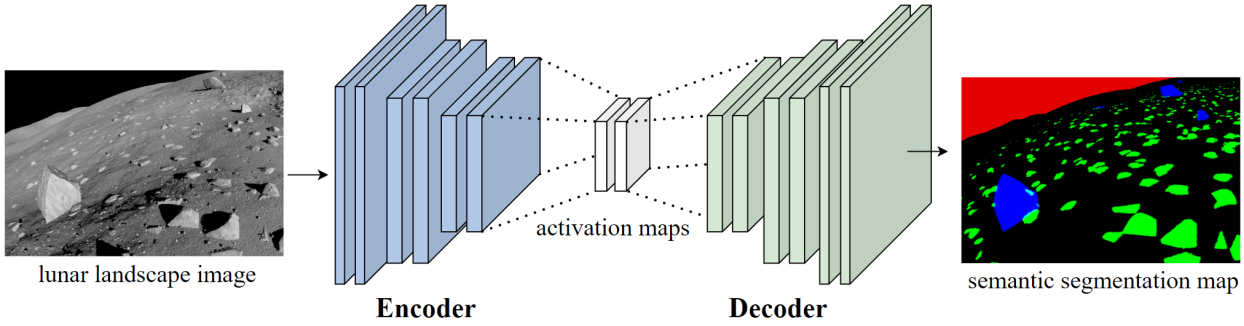


Figure 3. Architecture of the naïve approach for lunar landscape segmentation.

Figure 3 illustrates the naive approach to image semantic segmentation. Image semantic segmentation networks often adopt the UNet (Ronneberger et al. 2015) architecture due to its demonstrated promise and comparable performance in tasks related to image semantic segmentation. The UNet architecture consists of an encoder and a decoder. The encoder receives the input image and generates activation maps containing visual characteristics of the input image, which, in this research, represents lunar landscape images. The decoder utilizes these activation maps as input to generate the semantic segmentation map. Both the encoder and decoder are implemented using convolutional neural networks. This well-established semantic segmentation method typically ensures comparable results when the distribution of training and testing data is similar, a common scenario in general machine learning problems.

However, as mentioned earlier, due to the unique characteristic of the lunar landscape dataset (unavailability of direct visits to the moon for data collection), synthetic datasets are employed. Despite the careful generation of these datasets, a distribution gap remains between real and synthetic data, leading to performance degradation when evaluating the trained model with the synthetic dataset. For this reason, it is necessary to develop domain adaptation techniques to narrow the distribution gap during the training process, aiming for a more accurate model.

Proposed Domain Adaptation Method

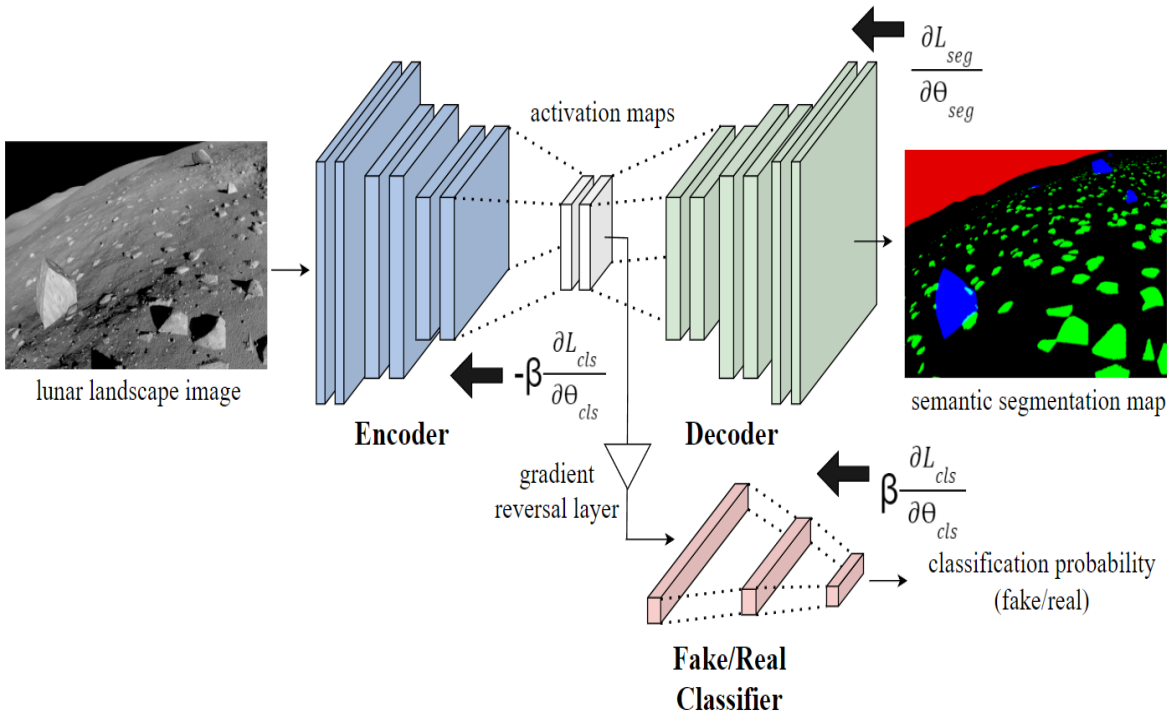


Figure 4. Architecture of the proposed domain adaptation method

Figure 4 illustrates the proposed domain adaptation method for lunar landscape semantic segmentation. This method extends the aforementioned UNet-like architecture. The proposed method is composed of an encoder, a decoder and a fake/real classifier with a gradient reversal layer. The encoder processes input images to produce activation maps, and the decoder generates semantic segmentation maps.

Throughout this process, I introduced an additional fake/real classifier that takes the activation maps as input. This module is responsible for determining whether the activation maps are generated from real or synthetic images. As the model undergoes training, it is expected that the fake/real classifier becomes more accurate. Here, the objective is to ensure the consistent generation of activation maps from both synthetic and real images, making it challenging for the fake/real classifier to distinguish between the two sources. More specifically, if the encoder successfully generates consistent activation maps regardless of the source of input images, the real/fake classifier cannot accurately determine the classification.

To accomplish this, I propose a gradient reversal layer between the encoder and the fake/real classifier, reversing the gradient during backpropagation (gradient descent). This gradient reversal layer guarantees that the encoder focuses on enhancing its proficiency in generating consistent activation maps (domain adaptation ability), while concurrently aiding the fake/real classifier in improving its capacity to distinguish between real and generated data. The effectiveness of this approach is further elaborated in Section 4 through comprehensive experimental results.

Implementation

The proposed domain adaptation approach is versatile, allowing its application to various network architectures. I have applied this ad-hoc approach to train several well-established image semantic segmentation architectures, including PSPNet (Zhao et al. 2017), Multipath-RetineNet (Lin et al. 2017), IDW-CNN (Wang et al. 2017), CASIA_IVA_SDN (Fu et al. 2019), and DeepLabV3 (Chen et al. 2017). The training process involves employing the dice coefficient loss function which is commonly used for training image semantic segmentation networks. Equation 1 provides an explanation of the dice coefficient loss function.

Equation 1: Dice coefficient loss function

$$L_{dice} = 1 - score_{dice}$$

In equation 1, $score_{dice}$ denotes the dice coefficient score calculated by equation 2 below.

Equation 2: Dice score

$$score_{dice} = \frac{2 * |Pred \cap Gt|}{|Pred| + |Gt|}$$

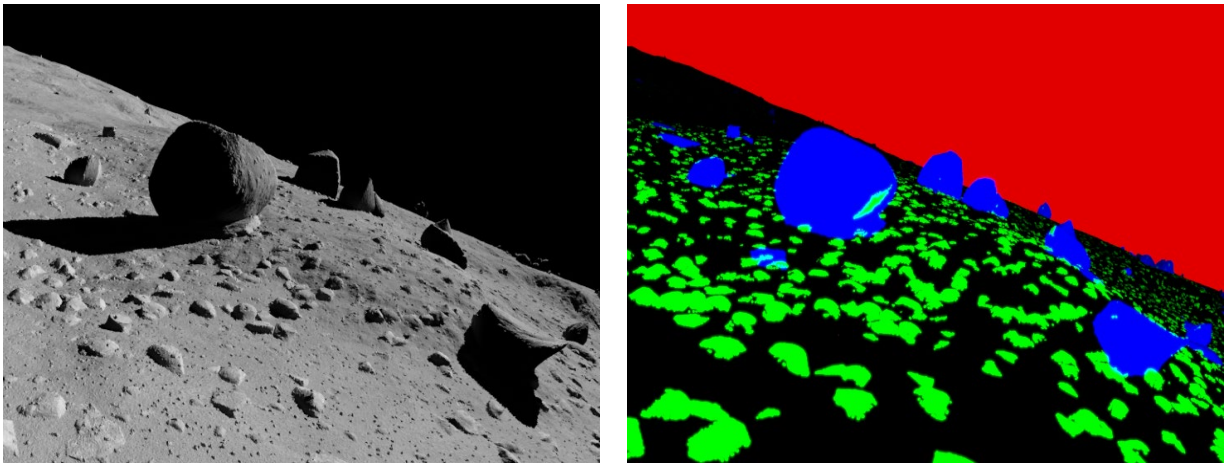
Where, *Pred* and *Gt* represent the prediction of the trained model and its ground truth. The dice coefficient loss function quantifies the similarity between prediction and its ground truth pixel sets. It measures the agreement between the predicted segmentation and the ground truth. During the training process, the model learns to maximize the overlap between the predicted and true segmentation masks. Further investigation into the detailed performance comparison and effectiveness of the proposed gradient reversal layer is presented in Section 4.

Experimental Results

Dataset

The synthetic lunar landscape dataset comprises a total of 9,766 realistic images generated through a simulation process. These synthetic images aim to emulate the visual characteristics and diversity of lunar terrains. In addition to the synthetic dataset, 36 real moon images are incorporated into the study. To train and evaluate the

proposed semantic segmentation model, 80% of the synthetic dataset, amounting to 7,812 images, is utilized for training, while the remaining 20%, consisting of 1,954 images, is reserved for model evaluation. The synthetic dataset is annotated with three distinct labels such as sky, small and large obstacles. Some visual examples of the dataset are illustrated in Figure 5.



(a) (b)

Figure 5. Visual example of dataset, (red: sky, blue: big obstacles, and green: small obstacles)

Intersection over Union

The performance of the proposed semantic segmentation model is quantitatively assessed using the Intersection over Union (IoU). IoU measures the overlap between predicted and ground truth segmentation masks and is defined as the size of the intersection divided by the size of the union of the predicted and ground truth regions as shown in Figure 6. The IoU is computed separately for each class, allowing for a detailed understanding of how well the model performs in segmenting specific lunar landscape features, such as the sky, small obstacles, and large obstacles.

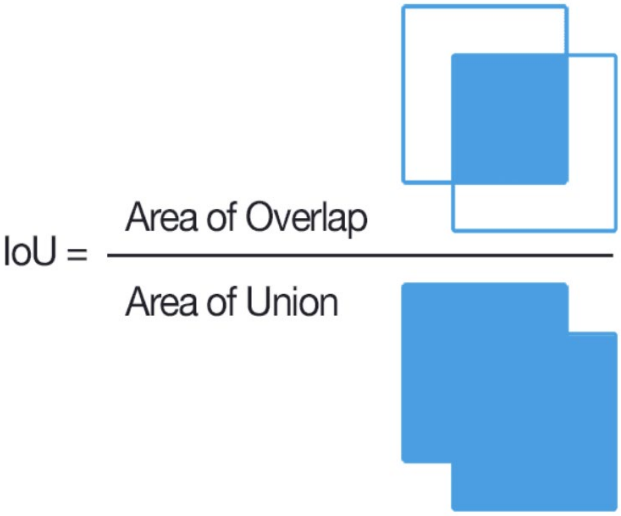


Figure 6. Illustration of intersection over union

High IoU values indicate a strong alignment between the predicted and ground truth segmentation masks, signifying accurate and reliable semantic segmentation.

Performance Comparison

Table 1. IoU comparison with the state-of-the-art semantic segmentation methods

Method	Sky	Small Obstacle	Large Obstacle	Total
PSPNet (Zhao et al. 2017)	84.4	80.4	82.8	82.7
Multipath-RetineNet (Lin et al. 2017)	84.5	81.2	82.8	82.8
IDW-CNN (Wang et al. 2017)	84.7	81.1	83.9	83.8
CASIA_IVA_SDN (Fu et al. 2019)	86.1	82.8	84.2	84.6
DeepLabV3 (Chen et al. 2017)	**86.8**	**83.3**	**85.2**	**85.6**

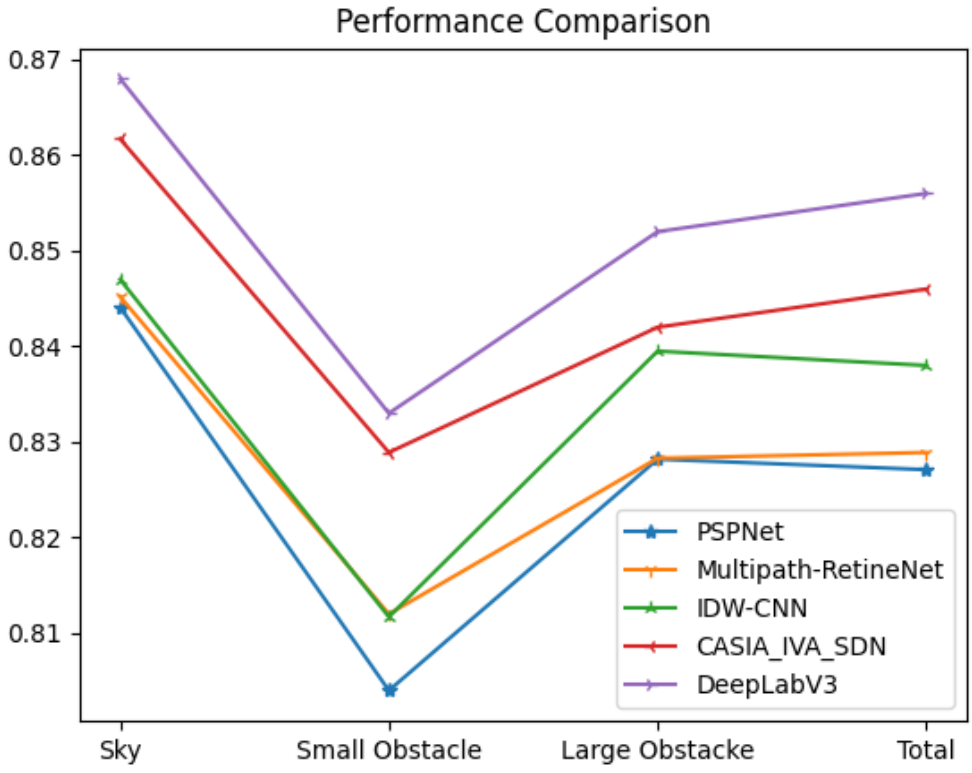


Figure 7. IoU comparison with the state-of-the-art semantic segmentation methods (graph)

I initially conducted a performance comparison experiment involving five well-established image semantic segmentation methods: PSPNet (Zhao et al. 2017), Multipath-RetinaNet (Lin et al. 2017), IDW-CNN (Wang et al. 2017), CASIA_IVA_SDN (Fu et al. 2019), and DeepLabV3 (Chen et al. 2017). As shown in Table 1 and Figure 7, DeepLabV3 demonstrated the highest accuracy, outperforming the other methods, which yielded comparatively inaccurate results.

All these experimental results were derived from a trained model that employed the proposed training strategy. To assess the effectiveness of the proposed training approach (particularly the use of gradient reversal), I further conducted an ablation study.

Ablation Study

Table 2. Ablation study result (proposed gradient reversal)

	Ablation Model		**Proposed Model**	
Method	**Synthetic**	**Real**	**Synthetic**	**Real**
PSPNet (Zhao et al. 2017)	83.0	69.5	82.7	79.5
Multipath-RetineNet (Lin et al. 2017)	82.9	71.4	82.8	80.2
IDW-CNN (Wang et al. 2017)	83.8	71.0	83.8	81.4

CASIA_IVA_SDN (Fu et al. 2019)	84.9	76.7	84.6	82.5
DeepLabV3 (Chen et al. 2017)	85.8	83.8	85.6	83.8

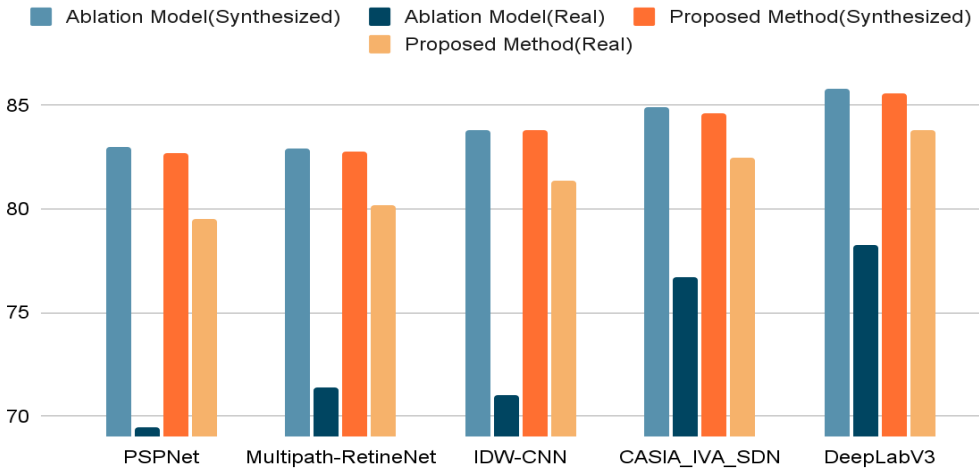


Figure 8. Ablation study result (proposed gradient reversal)

To examine the impact of the proposed method, I initially trained each semantic segmentation method without employing the proposed approach, which I refer to as ablation models in Table 2. Subsequently, I assessed the trained method on both synthetic and real datasets, as depicted in Table 2. I then quantified the performance gap for each ablation model. I trained the identical semantic segmentation model using the proposed training approach and measured the performance gap exactly as before. Finally, I measured how this performance gap differs between models trained without and with the proposed training approach.

As illustrated in Figure 8, the performance of each model trained with the proposed approach shows a slight decrease during real data evaluation. However, the reduction is relatively small compared to models trained without the proposed approach. These experimental results clearly demonstrate that the proposed gradient reversal approach functions as a domain adaptation technique, mitigating the performance degradation when applied to real-world data samples.

Conclusion

In this research, I proposed a novel domain adaptation method for lunar landscape semantic segmentation. The proposed machine learning-based approach, leveraging a synthetic lunar landscape dataset, showcased promising results. The comprehensive evaluation, encompassing both synthetic and real lunar landscape samples, underlined the model's adaptability and domain generalization capabilities. Notably, the incorporation of the proposed gradient reversal training approach demonstrated a crucial role in preventing performance collapse when transitioning to real-world data samples. This research makes a valuable contribution to the domain of lunar landscape semantic segmentation. In future studies, I plan to expand the scope by introducing additional segmentation categories, specifically focusing on applications related to resource search and the autonomous navigation of rovers.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Asgari Taghanaki, S., Abhishek, K., Cohen, J. P., Cohen-Adad, J., & Hamarneh, G. (2021). Deep semantic segmentation of natural and medical images: a review. Artificial Intelligence Review, 54, 137-178.

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

Fu, J., Liu, J., Wang, Y., Zhou, J., Wang, C., & Lu, H. (2019). Stacked deconvolutional network for semantic segmentation. IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2019.2895460

Gisler, M., & Sornette, D. (2009). Exuberant innovations: the apollo program.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1925-1934). https://doi.org/10.48550/arXiv.1611.06612

Lin, G., Shen, C., Van Den Hengel, A., & Reid, I. (2017). Exploring context with deep structured models for semantic segmentation. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1352-1366.

Neupane, B., Horanont, T., & Aryal, J. (2021). Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. Remote Sensing, 13(4), 808.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing. https://doi.org/10.48550/arXiv.1505.04597

Sinha, R. K., Rani, A., Ruj, T., & Bhardwaj, A. (2023). Geologic investigation of lobate scarps in the vicinity of Chandrayaan-3 landing site in the southern high latitudes of the moon. Icarus, 402, 115636.

Wang, G., Luo, P., Lin, L., & Wang, X. (2017). Learning object interactions and descriptions for semantic image segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5859-5867). https://doi.org/10.1109/CVPR.2017.556

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890). https://doi.org/10.48550/arXiv.1612.01105