Enhancing Visual Odometry Estimation Performance Using Image Enhancement Models

Hajira Saleem[1,2][a], Reza Malekian[1,2][b] and Hussan Munir[1,2][c]
[1]Department of Computer Science and Media Technology, Malmö University, Malmö, 20506, Sweden
[2]Internet of Things and People Research Centre, Malmö University, Malmö, 20506, Sweden
{hajira.saleem, reza.malekian, hussan.munir}@mau.se

Keywords: Visual Odometry, Image Enhancement, Low-Light Images, Localization, Pose Estimation.

Abstract: Visual odometry is a key component of autonomous vehicle navigation due to its cost-effectiveness and efficiency. However, it faces challenges in low-light conditions because it relies solely on visual features. To mitigate this issue, various methods have been proposed, including sensor fusion with LiDAR, multi-camera systems, and deep learning models based on optical flow and geometric bundle adjustment. While these approaches show potential, they are often computationally intensive, perform inconsistently under different lighting conditions, and require extensive parameter tuning. This paper evaluates the impact of image enhancement models on visual odometry estimation in low-light scenarios. We assess odometry performance on images processed with gamma transformation and four deep learning models: RetinexFormer, MAXIM, MIRNet, and KinD++. These enhanced images were tested using two odometry estimation techniques: TartanVO and Selective VIO. Our findings highlight the importance of models that enhance odometry-specific features rather than merely increasing image brightness. Additionally, the results suggest that improving odometry accuracy requires image-processing models tailored to the specific needs of odometry estimation. Furthermore, since different odometry models operate on distinct principles, the same image-processing technique may yield varying results across different models.

1 INTRODUCTION

Odometry estimation is an important process for the navigation of autonomous robots, particularly in environments that lack pre-existing maps. Odometry involves estimating the self-motion of an autonomous vehicle based on sensor measurements, predicting its pose over time. Pose estimation aims to determine the robot's position and orientation relative to a reference frame. Visual odometry offers several advantages, including lower computational complexity compared to other odometry methods. However, it tends to perform sub-optimally in low-light or dark conditions (Zhao et al., 2021; Wisth et al., 2021; Lee et al., 2024). Various enhancement techniques have been investigated to address this limitation, each presenting unique advantages and trade-offs.

In this study, we evaluate the impact of gamma transformation and four deep learning-based image enhancement models—RetinexFormer (Cai et al., 2023), MAXIM (Tu et al., 2022), MIRNet (Zamir et al., 2022), and KinD++ (Zhang et al., 2021)—on visual odometry estimation under low-light conditions. For our experiments, we used four sequences (01, 06, 07, and 10) from the KITTI dataset (Geiger et al., 2012), which consists of 11 sequences of images with ground truth poses. We randomly chose these sequences without bias towards any particular result. Although the KITTI sequences include images with mixed lighting conditions, they do not represent extremely dark conditions, such as those in the evening. Therefore, we artificially darkened the KITTI images to simulate low-light conditions for our testing.

We chose TartanVO (Wang et al., 2021) and Selective VIO (Yang et al., 2022) for the odometry estimation evaluation. We selected TartanVO based on the 'its designers' claim that it can generalize to various environmental conditions. We chose Selective VIO for its ability to achieve near ground truth odometry while being resource-efficient due to its lower computational cost.

[a] https://orcid.org/0000-0002-9596-2688
[b] https://orcid.org/0000-0002-2763-8085
[c] https://orcid.org/0000-0001-9376-9844

Saleem, H., Malekian, R. and Munir, H.
Enhancing Visual Odometry Estimation Performance Using Image Enhancement Models.
DOI: 10.5220/0012932600003822

In *Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2024) - Volume 1*, pages 293-300
ISBN: 978-989-758-717-7; ISSN: 2184-2809

Low-light conditions exacerbate image degradation issues such as noise and color distortion, common in settings with limited camera quality. Simply increasing brightness can worsen these issues by amplifying image artifacts. Therefore, effective low-light enhancement requires not only brightening shadows but also reducing noise and preserving trackable features for accurate pose estimation.

RetinexFormer (Cai et al., 2023) enhances images by decomposing them into illumination and reflectance components, adjusting light and removing degradation separately. It uses an Illumination-Guided Transformer to manage long-range dependencies, outperforming 17 other methods on 13 low-light benchmarks. Similarly, KinD++ (Zhang et al., 2021) uses a retinex-based approach outperforming 12 other models on seven datasets, though DUPE (Wang et al., 2019) showed comparable results in some cases.

MAXIM (Tu et al., 2022) enhances dark regions using a UNet-shaped framework with spatially-gated MLPs, combining local and global visual cues. It performed well on low-light enhancement tasks, though MIRNet (Zamir et al., 2022) had a higher Peak Signal-to-Noise Ratio but comparable Structural Similarity Index. MIRNet, with its multi-scale information retention and attention mechanisms, preserves spatial details while enriching features across scales, making it highly effective on low-light benchmarks.

Instead of aiming to perfectly restore original image quality, our study focuses on improving odometry and pose estimation under low-light conditions. Therefore, we used odometry metrics like absolute trajectory error, relative translational error, and relative rotational error for evaluating and comparing the results in this study.

This paper makes the following contributions:

- It presents a comparative analysis of odometry performance acheived by gamma transformation and four state-of-the-art image enhancement models.
- It demonstrates that none of the models performed optimally in all scenarios, highlighting the need for a model that enhances features useful for odometry while removing artifacts that decrease performance.
- It identifies the strengths and weaknesses of each enhancement model in the context of odometry, offering practical recommendations for their use in specific scenarios.

The rest of the paper is organized as follows: Section II presents the related works. Section III introduces the methodology. Section IV presents the results and discussion. Finally, Section V concludes the paper, summarizing the key findings and implications of the research.

2 RELATED WORK

Visual odometry faces challenges in low-light and blurry conditions due to haze, motion blur, and similar factors. Various approaches, including sensor-fusion methods with LiDAR, have been explored to enhance performance in such conditions (Zhao et al., 2021; Wisth et al., 2021; Lee et al., 2024). Multi-camera systems outperform monocular setups in dark environments Liu et al.'s (Liu et al., 2018) , but multi-sensor odometry increases computational costs. To mitigate these issues, recent research focuses on enhancing monocular odometry through position-aware optical flow and geometric bundle adjustment (Cao et al., 2023). Despite achieving superior results in low-light settings, these methods often struggle with depth estimation and object proximity issues in high-luminance and dynamic environments.

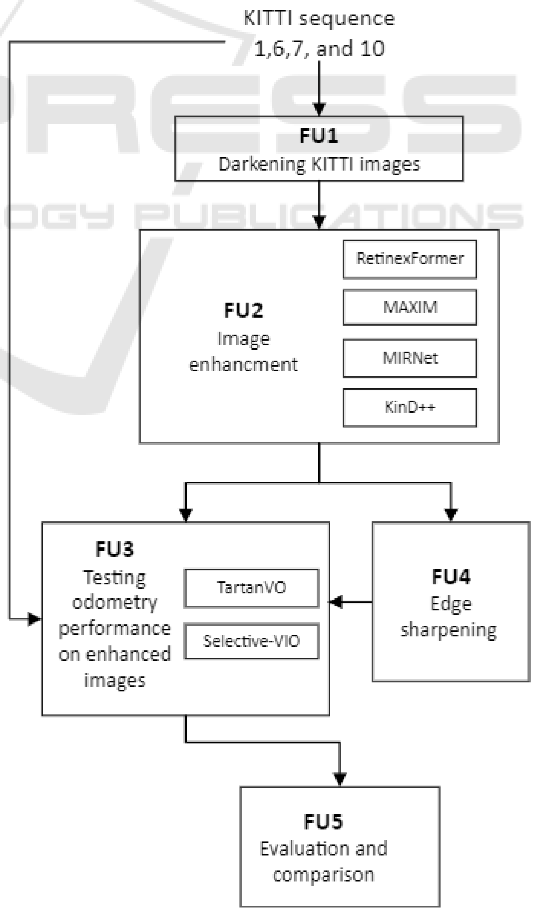


Figure 1: Methodology diagram.

Traditional algorithms like adaptive histogram equalization have been employed to enhance low-light images for better odometry performance (Hao et al., 2019; Zhang et al., 2022; Gao et al., 2022). However, these methods often fail to account for varying illumination within the same image, leading to suboptimal results. Moreover, contrary to learning-based approaches, their parameters must be tuned for specific lighting conditions in most cases. Conversely, deep learning approaches like CycleGAN and generative adversarial networks have been proposed to enhance low-light images while maintaining structural consistency between frames (You et al., 2023). Efforts have also been made to integrate low-light capabilities directly into odometry estimation neural networks using binary and deep descriptors (Alismail et al., 2016), though these studies often lack generalizability.

3 METHODOLOGY

The methodology employed in this study consists of several key steps, organized into five functional units (FUs), as illustrated in Figure 1.

3.1 FU1: Image Darkening

The original KITTI dataset sequences (01, 06, 07, and 10) were predominantly recorded during daylight, which does not represent the low-light conditions this study aims to investigate. To address this, we applied a gamma transformation to darken the images to simulate night-time conditions. This method effectively adjusts the luminance through a non-linear mapping of pixel intensities, allowing us to retain fine details while creating the desired low-light effect. The general form of gamma transformation is expressed as: $I_{\text{out}} = c \cdot I_{\text{in}}^{\gamma}$, where I_{out} denotes the output pixel intensity, c is a scaling constant, typically set to 1 for simplicity, γ is the gamma correction parameter (we used gamma value 0.3), and I_{in} represents the input pixel intensity, normalized to the range [0, 1].

3.2 FU2: Image Enhancement

The darkened images were then processed using four state-of-the-art image enhancement methods: RetinexFormer, MAXIM, KinD++, and MIRNet. The purpose of this step was to assess the effect of these enhancement models on the odometry estimation performance. In Figure 2, one image from KITTI dataset and its enhanced versions using image processing methods is shown.

3.3 FU3: Odometry Estimation

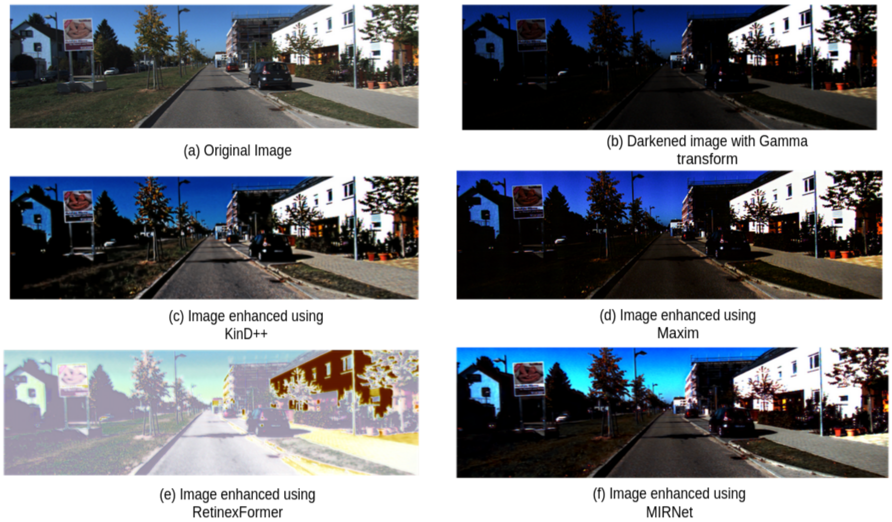
We tested the enhanced images using two odometry estimation models, TartanVO and Selective VIO, to evaluate their performance in tracking and estimating pose. These models were chosen for their robustness in varying environmental conditions and their ability to handle different image qualities.

3.4 FU4: Edge Enhancement

We observed that in RetinexFormer-enhanced images, the color channels and image features seemed to deteriorate, yet they consistently demonstrated strong performance in many cases. We attributed this to the enhancement of edges. To investigate this further, we explored the impact of edge enhancement on odometry estimation. As the next step in our research, we applied edge enhancement techniques to the images. Examples of images enhanced by RetinexFormer are shown in Figure 3. To enhance edges, we first applied a Gaussian blur with a sigma value of 2 to reduce noise and smooth the image. Then, we used the Canny edge detector with threshold values of 100 and 200 to identify edges. To make the detected edges more pronounced, we dilated them using a 3x3 kernel. The edge map, initially in grayscale, was converted to a three-channel image to match the original image. Finally, we combined the original image with the edge map by blending them with weights of 1.5 for the original image and -0.5 for the edges, resulting in a sharpened image with enhanced edges.

3.5 FU5: Performance Comparison

Finally, the odometry estimation performance across different image versions was compared the using metrics: absolute trajectory error (ATE), relative translational error (t_{rel}), and relative rotational error (r_{rel}). ATE assesses the global accuracy of the estimated trajectory by comparing it to the ground truth, providing a single error value that summarizes the deviation. t_{rel} and r_{rel} measure the translation and rotation errors over specific distances or time intervals, respectively, representing the local accuracy over short trajectory segments. Lower values of ATE, t_{rel}, and r_{rel} indicate better performance. The analysis in this paper involved a thorough examination of error curves, evaluation metrics, and speed maps.



(a) Original Image

(b) Darkened image with Gamma transform

(c) Image enhanced using KinD++

(d) Image enhanced using Maxim

(e) Image enhanced using RetinexFormer

(f) Image enhanced using MIRNet

Figure 2: Sample images enhanced from KITTI dataset using four enhancement models.

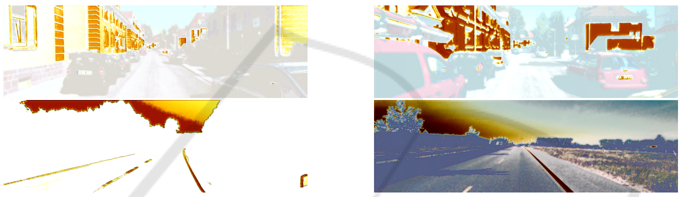


Figure 3: Images enhanced using RetinexFormer model.

4 RESULTS AND DISCUSSION

This section evaluates the impact of various image enhancement techniques on visual odometry performance using the TartanVO and Selective VIO techniques. The discussion is organized based on the two odometry models.

4.1 Odometry Performance with TartanVO

Figure 4 shows trajectories produced using TartanVO using images processed using various image processing methods. Figure 5 and 7 show the results of odometry estimation using TartanVO on KITTI sequences 01, 06, 07, and 10. These figures shows that MAXIM-enhanced images generally provided the best overall odometry performance when tested with TartanVO, especially noticeable improvements in ATE for sequences 01 and 10. However, a significant t_{rel} value for MAXIM-enhanced Seq-01 suggests it may struggle with accurate translations over smaller segments. This indicates that while MAXIM enhances global trajectory consistency, it may not reliably estimate smaller segment translations. Additionally, edge-enhanced images with MAXIM show consistent performance but highlight issues with Seq-01. Rotation estimates exhibited minimal variation across images processed with all four different methods as can be seen in Figure 7.

The ATE results for Seq-06 and Seq-10 demonstrate that odometry performance is negatively affected in dark conditions. Conversely, improved performance using dark images in Seq-01 and Seq-07, compared to original KITTI images is due to enhanced image contrast. Despite the general improvement with image enhancement models, maintaining brightness consistency across the sequence is crucial for better tracking and odometry estimation.

4.2 Odometry Performance with Selective VIO

Figure 6 and 8 shows the results of odometry estimation using Selective-VIO on KITTI sequences 01, 06, 07, and 10. MIRNet edge-enhanced images offered the best performance for odometry estimation with Selective VIO, with significantly lower ATE values across most sequences, making it highly effective for this model. Conversely, RetinexFormer showed

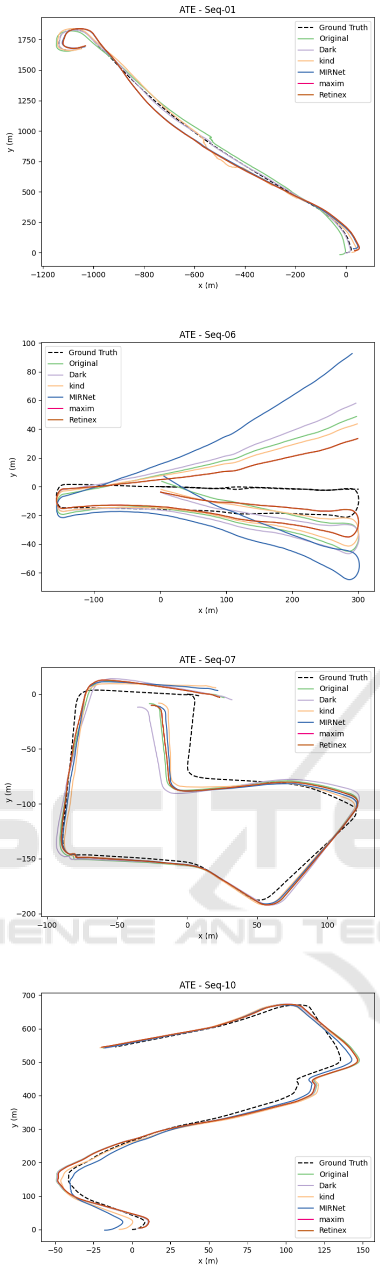


Figure 4: Odometry estimation trajectories using TartanVO.

the worst performance, especially when using edge-enhanced images, indicating its unsuitability under dark conditions. MAXIM-enhanced images provided moderate improvements but lacked consistency, particularly in edge-enhanced scenarios. Overall, dark images enhanced using MIRNet and KinD++, and edge-enhanced versions of images enhanced using MIRNet appeared promising for Selective VIO.

4.3 Impact of Edge Enhancement

The analysis showed that not all methods benefited from edge enhancement. While RetinexFormer deteriorated the visual appearance of images, it still performed better in some cases than other models even though the images processed by it appeared to keep edges of objects and rest of the image content vanished. This prompted an investigation into edge enhancement's effect on odometry estimation. Edge-enhanced original images did not significantly improve accuracy with TartanVO, while they did improve performance with Selective VIO, except when the images were darkened. In sequences where the original images were already well-lit or had a lower contrast, edge enhancement sometimes degraded performance. For example, in well-lit sequences, the additional emphasis on edges introduced by enhancement techniques like RetinexFormer occasionally led to over-sharpening, which in turn reduced the overall quality of feature matching and tracking. Morever, edge enhancement method we used did not ensure that edge sharpening is consistently done across the image sequence. This result highlights the importance of carefully selecting when and how to apply edge enhancement, depending on the specific characteristics of the image and the odometry model being used. Compared to TartanVO, Selective VIO showed more consistent benefits from edge enhancement, especially when combined with MIRNet.

4.4 Speed and Lighting Variation

Rotation estimates exhibited minimal variation across images processed with all models. However, error curves indicated a consistent trend where sequences with abrupt lighting changes and higher speeds showed higher errors. This suggests that image enhancement methods should focus on improving reliable feature extraction and maintaining consistent lighting conditions rather than indiscriminately brightening images.

5 CONCLUSION AND FUTURE WORK

This study addressed the decline in odometry performance under dark conditions by evaluating four deep learning-based image enhancement techniques—MAXIM, MIRNet, RetinexFormer, and KinD++—on darkened images from the KITTI dataset. Our findings indicate that while MAXIM-enhanced KITTI sequences generally performed well

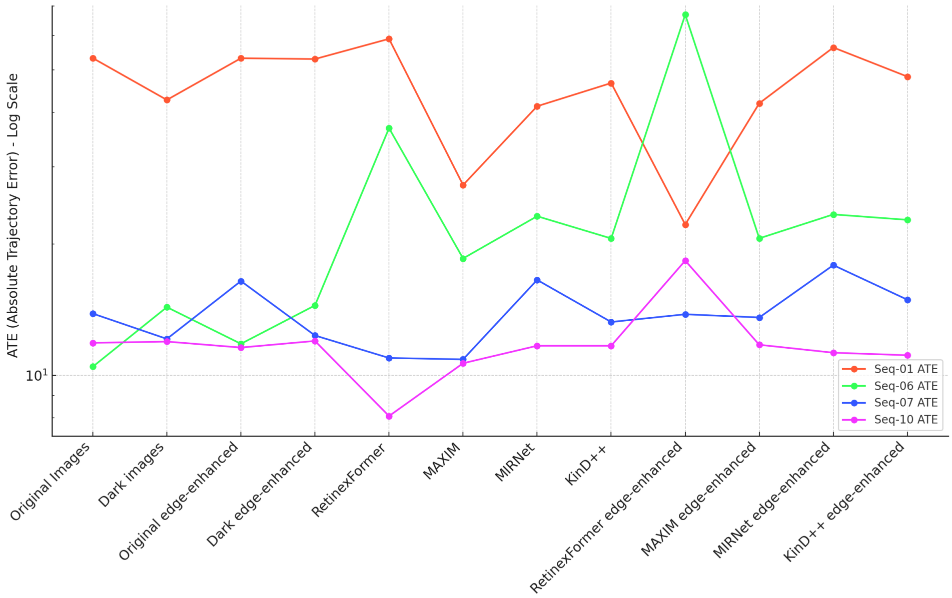


Figure 5: ATE comparison for TartanVO on different KITTI sequences.

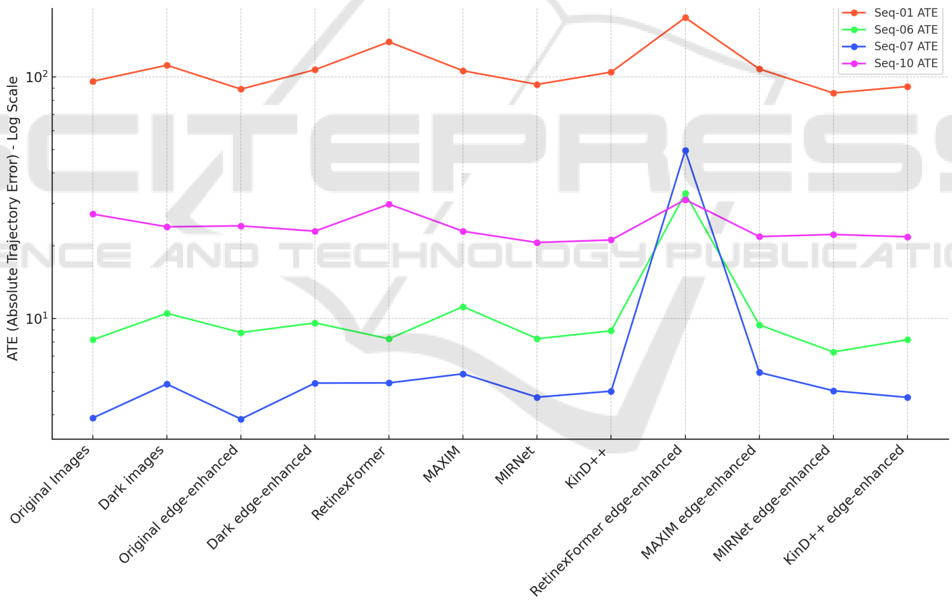
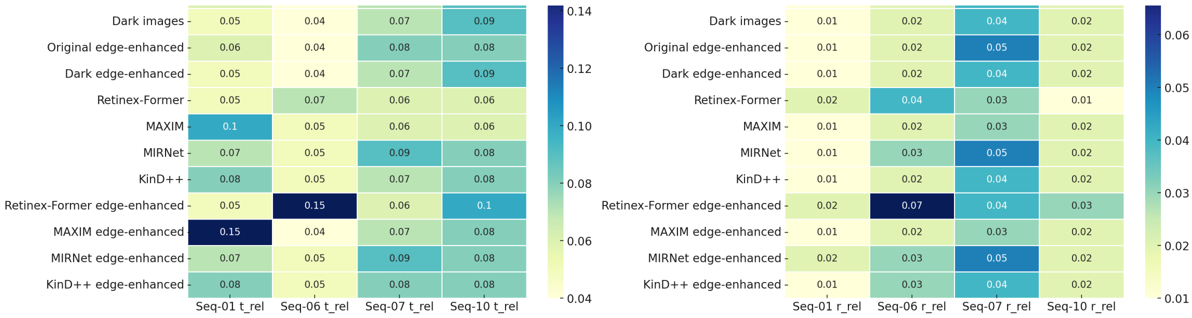


Figure 6: ATE comparison for Selective-VIO on different KITTI sequences.

with TartanVO, not all sequences yielded optimal results. For Selective VIO, MIRNet, its edge-enhanced versions, and KinD++ showed promise in improving odometry performance. However, maintaining brightness consistency across image sequences remains crucial for reliable tracking and odometry estimation.

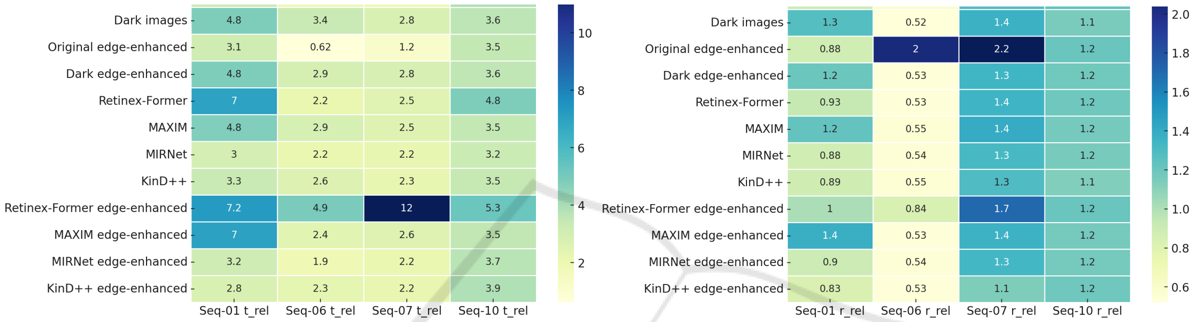
Most existing image enhancement methods are general-purpose models that do not account for their impact on odometry performance. As such, our future work will focus on integrating odometry-aware loss functions into the training of image enhancement models. We also aim to validate these techniques in real-time scenarios across diverse low-light datasets.

This study underscores the need for adaptive enhancement strategies tailored to the specific requirements of different odometry algorithms, particularly in challenging lighting conditions. Given that Selective VIO and TartanVO models respond differently to image enhancements, it is essential to test multiple methods to identify the best combination for



(a) TartanVO t_rel (b) TartanVO r_rel

Figure 7: Heatmaps of t_rel and r_rel of TartanVO.



(a) Selective-VIO t_rel (b) Selective-VIO r_rel

Figure 8: Heatmaps of t_rel and r_rel Selective VIO.

real-world applications. Future research should prioritize developing robust, reliable navigation systems for autonomous vehicles operating in low-light environments by incorporating odometry-aware training approaches. Additionally, future work should also emphasize the importance of maintaining brightness consistency across image sequences to improve odometry estimation.

ACKNOWLEDGEMENTS

This work was supported by the Department of Computer Science and Media Technology and Internet of Things and People Research Centre at Malmö University.

AUTHORSHIP CONTRIBUTION STATEMENT

Hajira Saleem: Conceptualization, investigation, data curation, Formal analysis, writing—original draft preparation, Reza Malekian: Conceptualization, review and editing, methodology, supervision, funding acquisition, Hussan Munir: Conceptualization, review and editing, methodology, supervision.

REFERENCES

Alismail, H., Kaess, M., Browning, B., and Lucey, S. (2016). Direct visual odometry in low light using binary descriptors. *IEEE Robotics and Automation Letters*, 2(2):444–451.

Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., and Zhang, Y. (2023). Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513.

Cao, Y.-J., Zhang, X.-S., Luo, F.-Y., Peng, P., Lin, C., Yang, K.-F., and Li, Y.-J. (2023). Learning generalized visual odometry using position-aware optical flow and geometric bundle adjustment. *Pattern Recognition*, 136:109262.

Gao, W., Yang, G., Wang, Y., Ke, J., Zhong, X., and Chen, L. (2022). Robust visual odometry based on image enhancement. In *Journal of Physics: Conference Series*, volume 2402, page 012010. IOP Publishing.

Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE.

Hao, L., Li, H., Zhang, Q., Hu, X., and Cheng, J. (2019). Lmvi-slam: Robust low-light monocular visual-inertial simultaneous localization and mapping. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 272–277. IEEE.

Lee, D., Jung, M., Yang, W., and Kim, A. (2024). Lidar odometry survey: recent advancements and remaining challenges. *Intelligent Service Robotics*, pages 1–24.

Liu, P., Geppert, M., Heng, L., Sattler, T., Geiger, A., and Pollefeys, M. (2018). Towards robust visual odometry with a multi-camera system. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1154–1161. IEEE.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., and Li, Y. (2022). Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780.

Wang, R., Zhang, Q., Fu, C.-W., Shen, X., Zheng, W.-S., and Jia, J. (2019). Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857.

Wang, W., Hu, Y., and Scherer, S. (2021). Tartanvo: A generalizable learning-based vo. In *Conference on Robot Learning*, pages 1761–1772. PMLR.

Wisth, D., Camurri, M., Das, S., and Fallon, M. (2021). Unified multi-modal landmark tracking for tightly coupled lidar-visual-inertial odometry. *IEEE Robotics and Automation Letters*, 6(2):1004–1011.

Yang, M., Chen, Y., and Kim, H.-S. (2022). Efficient deep visual and inertial odometry with adaptive visual modality selection. In *European Conference on Computer Vision*, pages 233–250. Springer.

You, D., Jung, J., and Oh, J. (2023). Enhancing low-light images for monocular visual odometry in challenging lighting conditions. *International Journal of Control, Automation and Systems*, 21(11):3528–3539.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., and Shao, L. (2022). Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):1934–1948.

Zhang, S., Zhi, Y., Lu, S., Lin, Z., and He, R. (2022). Monocular vision slam research for parking environment with low light. *International Journal of Automotive Technology*, 23(3):693–703.

Zhang, Y., Guo, X., Ma, J., Liu, W., and Zhang, J. (2021). Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037.

Zhao, S., Zhang, H., Wang, P., Nogueira, L., and Scherer, S. (2021). Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8729–8736. IEEE.