Twilight SLAM: A Comparative Study of Low-Light Visual SLAM Pipelines

Surya Pratap Singh, Billy Mazotti, Sarvesh Mayilvahanan, Guoyuan Li, Dhyey Manish Rajani, Maani Ghaffari
{suryasin, bmazotti, smayil, lguoyuan, drajani, maanigj}@umich.edu

***Abstract*—This paper presents a comparative study of low-light visual SLAM pipelines, specifically focusing on determining an efficient combination of the state-of-the-art low-light image enhancement algorithms with standard and contemporary Simultaneous Localization and Mapping (SLAM) frameworks by evaluating their performance in challenging low-light conditions. In this study, we investigate the performance of several different low-light SLAM pipelines for dark and/or poorly-lit datasets as opposed to just partially dim-lit datasets like other works in the literature. Our study takes an experimental approach to qualitatively and quantitatively compare the chosen combinations of modules to enhance the feature-based visual SLAM.**

Our code pertaining to this study is publicly available at https://github.com/TwilightSLAM

I. Introduction

In recent years, Visual SLAM (VSLAM) has received extensive attraction from researchers working in SLAM. This is primarily due to the availability of low-cost sensors, the ability to capture abundant environmental information, and easy fusion with other sensors [1].

VSLAM has proven to be at least as effective as its counterparts such as Lidar SLAM, but certain challenges persist with existing VSLAM methods [2]. One such challenge is to make these algorithms robust to illumination since the feature extractor's performance falls sharply under dark conditions [3]. The current approaches to deal with this challenge do improve the feature extraction performance but fail under extremely dark environments.

II. Related Works

The VSLAM formulations are primarily bifurcated into two groups: (i) Feature-driven (or direct) [4] and (ii) Image dependent (or indirect) approaches [5]. The former leverages descriptors that identify an image as a collection of features, whereas the latter processes overall images (for instance using optical flow-based techniques) rather than extracting characteristics or key features.

We observe that every group, and any sub-categorical formulation within the corresponding group, has its own limitations. For instance, the image-dependent approaches fail when there is an unprecedented or sudden change of the viewing perspective (mostly sharp turns) or angle, and feature-dependent frameworks are highly susceptible to the lighting and/or contrast variation level considered by the descriptor. These method-specific drawbacks lead to lower mapping and localization accuracy. The majority of feature-based SLAM techniques, such as those described in the papers [4], [6] and [7], include SLAM front-end formulations completely based on feature extractors which are mostly key-points. Although these techniques perform extremely well in well-lit settings, conventional descriptors perform poorly in low-light environments, consequently causing the accuracy of localization to suffer.

Hence, the feature-based methods are constrained by the descriptor attributes that they leverage. Unsymmetrical SIFT extractor demonstrates the best illumination invariance amongst standard feature extractors, like SIFT, SURF, and FAST, from the investigations done by [8] and [9]. However, these studies miss out on the classical ORB feature extractor [10], omnipresent in modern SLAM frameworks. There are various learning-based feature attribute description approaches like SuperGlue [11] and SuperPoint [12], but they haven't been incorporated into currently widely-used algorithms used in SLAM due to their computational inefficiency and a lack of direct plugins into standard algorithms.

Numerous research papers deal with the problem of appearance transformation, such as the study done in [13], which illustrates how to determine the adaptability and invariance of a feature. For instance, the system is able to identify compromising attributes of the feature which, for SLAM specifically, form the locations of any transient or dynamic objects in the environment. Therefore, the suggested technique is somewhat resilient to seasonal fluctuations, despite the fact that it cannot be applied directly to scenarios with low-light or normal-light contrast disturbances. Another study [14] uses basic image processing and warping with histogram equalization to deal with poorly-lit image/video sequences. However, this study is not very holistic as the proposed algorithm was tested primarily in partially dim-light conditions (which constitutes of more than fifty percent well-lit environment) and lacks robust testing under low-light (and dark) environments.

Research done in [15] depicts the use of GANs in the image pre-processing pipeline for modifying the localization results positively. This research generates images by using neural style transfer on input image/video sequence to make season changes as a part of the corresponding input images, which can be further used for learning the mapping in different environments. The study also proved that by using GAN-assisted unpaired image-to-image translation on input images led to more accurate localization results on different unseen input image datasets. The drawback here is this particular study focuses on the changing seasons, and on a thorough review, we found that the ratio of low-light to well-lit images

in this dataset is pretty low than what is expected out of low-light SLAM system datasets.

On further review and analysis, we found that Savinykh et. al [3] published DarkSLAM a novel GAN-based SLAM method for low-light conditions, which is a difficult task for traditional VSLAM methods. The authors of DarkSLAM found EnlightenGAN [16] enhances images better than deterministic image enhancing algorithms with respect to ORB-SLAM2 [10] performance, however, they did not further investigate alternative SOTA image enhancement methods like Zero-DCE [17] and Bread [18], to use in SLAM pipeline and feature matching algorithms.

In order to fully compare different low-light image enhancement models and SLAM frameworks, we present a comparative study with two main outcomes:

- Analyse various SOTA image enhancement modules (like EnlightenGAN, Bread, Zero-DCE & Dual) and their image enhancement capabilities for VSLAM.
- Evaluate a standard and a SOTA SLAM framework on relevant low-light datasets, along with various image enhancement modules to draw conclusions about the best configurations to use.

The development of the above motivation and outcomes is described at length in the following sections.

III. Datasets

A. KITTI

KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) is one of the most widely used datasets for mobile robots and autonomous driving. It consists of several hours' worth of traffic scenarios that were recorded using a variety of sensor modalities, including high-resolution RGB, grayscale stereo, and 3D laser scanner cameras. The dataset contains sequences of images with a resolution of 1,242×375. These images were captured at 10 FPS (Frames per second) from an automobile traveling in urban and highway environments. In our project, we have used sequences 04, 06, and 07. Sequence 04 doesn't contain a closed loop whereas sequences 06 and 07 have closed loops. All these sequences have daylight images, which will be helpful in evaluating how the image enhancement modules work under non-low-light environments.

B. ETH3D

ETH3D is a research group at ETH Zurich that focuses on 3D computer vision and robotics. They have released various datasets for SLAM research, which are available for download on their website. Here we use two image sequences: *sfm_house_loop* and *sfm_lab_room2* from a group of SLAM datasets collected specifically from a handheld stereo camera in both indoor and outdoor environments. Within this group of datasets, each dataset has a raw image sequence (with and/or without loop closure) and segregated data for the corresponding sequence, in monocular, stereo, RGBD, and IMU formats. Now each format consists of calibration, ground truth data, and the RGB images (which is the image sequence).

The *sfm_house_loop* dataset consists of a camera moving in a large loop circumscribing a house, whereas the *sfm_lab_room2* dataset consists of a camera moving through a cluttered lab room. Both of the datasets were filmed outside the availability of a Vicon mocap system, hence the ground truth is determined with Structure-from-Motion. We specifically choose the *sfm_house_loop* and *sfm_lab_room2* datasets because the major percentage of the image sequence consisted of traversing in dark and/or poorly-lit environment, which also helps us decide upon a robust configuration during framework design and selection.

IV. Image Enhancement Modules

A critical part of the model pipeline is the image enhancement module, which enlightens the low-light image for use in the SLAM methods described above. For this study, four different state-of-the-art enlightening models were rigorously validated before being inserted into the Twilight SLAM pipeline for comprehensive testing. These four models were chosen in particular due to their performance on traditional low-light image enhancement datasets.

A. EnlightenGAN

EnlightenGAN [16] is an unsupervised generative adversarial network trained without the need for a low/normal-light image pair dataset. Many of the state-of-the-art deep learning models rely on synthetic or highly cleaned datasets with ground-truth enlightened images. However, this data's availability and quality are insufficient for real-world applications.

The training process circumvents the need for ground truth enlightened images by building an unpaired mapping between low/normal-light images that don't rely on perfectly paired images. This is accomplished through the use of a global and local discriminator. The global discriminator compares the model's outputted enlightened image to another relatively similar normal light image. The local discriminator samples random cropped sections of the enlightened and normal light images and compares the sections. If the model is unable to discriminate between the two sections, the loss is low, and vice versa. The use of global and local discriminators ensures that EnlightenGAN has good performance in smaller regions as well as the image as a whole.

Due to the robustness of this unpaired training, EnlightenGAN performs very well in enhancing real-world images from a wide range of domains, which makes it very applicable for a general-purpose low-light SLAM.

B. Bread

Bread [18] is a novel low-light enhancement model that performs enhancement in both color and noise separately. A key advancement illustrated in this paper is the decoupling of both noise and color, which is achieved by converting the image from the RGB space to a luminance (brightness) and chrominance (color) space.

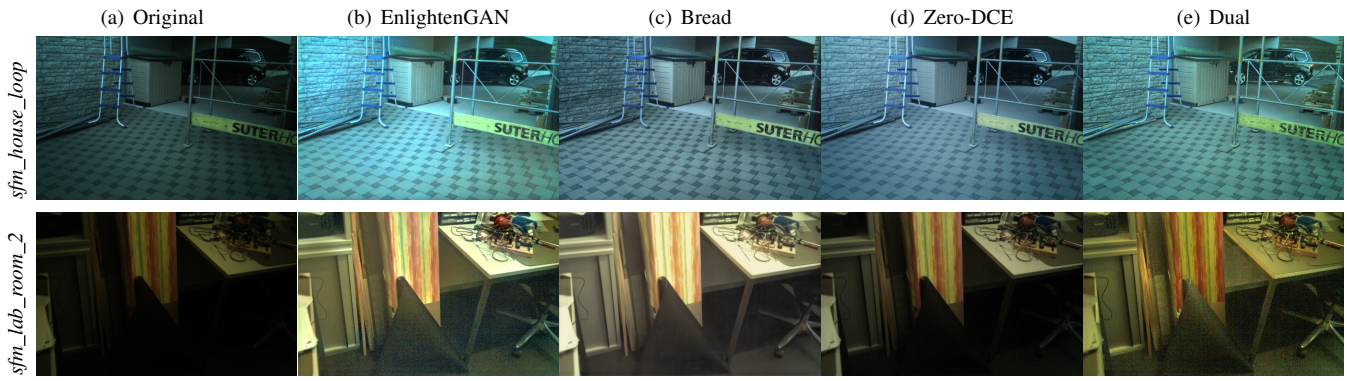

Fig. 1. Comparing image enhancement modules for two different ETH3D low-light SLAM datasets.

First, an Illumination Adjustment Network (IAN) is used to brighten the image, which is then passed to the noise suppression module to eliminate noise in the brightened luminance. This enhancement luminance is then passed to the Color Adaptation Network (CAN), which uses the illumination map to generate realistic colors globally and locally. When compared to other models on the LOw-Light dataset (LOL), Bread outperforms all models on all metrics, including peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

C. Zero-DCE

Zero-Reference Deep Curve Estimation (Zero-DCE) [17] is a novel deep-learning-based low-light image enhancement method that trains a lightweight deep network to estimate pixel-wise and high-order curves for dynamic range adjustment of the given images. Specifically, unlike the traditional image-to-image mapping enhancement methods, Zero-DCE reformulates the enhancement task as an image-specific curve estimation problem that takes low-light images as inputs into a Deep Curve Estimation Network (DCE-Net) and estimates a set of high-order Light-Enhancement curves (L-E curves) as its output. Then, the non-reference loss function performs a pixel-wise adjustment on the dynamic range of the input images' RGB channels by applying the high-order curves to eventually obtain an enhanced image.

This method has several unique advantages, making it a robust and desirable image enhancement technique for low-light visual SLAM projects. First, the training process is more convenient than CNN and GAN-based methods as it forgoes the requirement of paired and unpaired reference data. Second, it applies well to various lighting settings, including nonuniform and poor light cases, and brightens low-light images without losing their inherent features or quality. Finally, it is computationally effective and efficient.

D. Dual

Dual, or Dual Illumination Estimation for Robust Exposure Correction [19], is a novel automatic exposure correction method, and here we used it as an image-enhancement technique. In this framework, the dual illumination estimation is first performed on the original and inverted input low-light images to obtain the forward and reverse illuminations, from which it then recovers the intermediate under-exposure and over-exposure corrected images of the input by solving related objective functions for refined illumination map with well-tuned correction parameters. Next, an effective multi-exposure image fusion is applied to effectively blend visually best-exposed parts in the two intermediate exposure corrected images and the input image using well-tuned influence parameters into the final globally well-exposed image.

This method has a simple framework and can generate high-quality exposure-corrected results for images of various exposure conditions, such as underexposed, overexposed, partially under-exposed, and partially over-exposed cases. Therefore, it is considered an ideal and reliable low-light image enhancement technique for our low-light visual SLAM pipelines.

E. Comparison

Figure 1 compares the outputs of these image enhancement modules for select frames from three different low-light datasets. Qualitatively comparing their performance, EnlightenGAN, Bread, and Dual appear to perform well in enlightening the images. However, Zero-DCE appears to struggle in particular with the dark indoor images of *sfm_lab_room2*.

V. SLAM Methods

We evaluate the impacts of the image enhancement modules discussed in section IV on two different SLAM frameworks to expand the scope of this study. Each algorithm, ORB-SLAM3, and SuperPoint-SLAM, build off the ORB-SLAM architecture yet use deterministic and learning-based features respectively. The differences in the nature of the feature detectors and descriptors between the two SLAM frameworks allow us to observe and draw empirical-based conclusions about the effects of various image enhancement modules on state-of-the-art feature detectors and descriptors for SLAM applications.

A. ORB-SLAM3

ORB-SLAM3 is a visual sensor-based SLAM solution designed for Monocular, Binocular, Stereo, and RGB-D cameras [20]. In addition to its adaptability and high precision, it can

TABLE I
NCPP, RMSE, AND MAX ERROR FOR ALL SLAM PIPELINES
(RED = BEST PERFORMANCE FOR SLAM FRAMEWORK, BLUE = BEST PERFORMANCE OVERALL)

		Superpoint-SLAM					ORB-SLAM3				
Sequence	Metric	Original	Bread	Dual	EnGAN	Zero-DCE	Original	Bread	Dual	EnGAN	Zero-DCE
KITTI 04	NCPP (pairs) ↑	243	223	236	**244**	241	159	188	**205**	192	167
	RMSE (m) ↓	**0.26**	0.40	0.47	0.37	0.52	1.72	1.30	1.52	**1.19**	1.50
	Max AE (m) ↓	**0.50**	0.80	1.08	0.77	1.13	3.59	3.01	3.27	**2.71**	3.85
KITTI 06	NCPP (pairs) ↑	603	412	589	586	**604**	575	618	**717**	625	607
	RMSE (m) ↓	14.73	**11.00**	13.03	14.50	14.12	17.65	14.56	13.91	15.22	**11.98**
	Max AE (m) ↓	23.68	22.13	**21.70**	22.90	27.22	35.43	28.49	24.24	22.79	**18.39**
KITTI 07	NCPP (pairs) ↑	555	564	568	552	**598**	589	613	**721**	615	603
	RMSE (m) ↓	2.40	2.57	8.07	3.22	**2.08**	2.49	2.85	4.85	**1.81**	2.14
	Max AE (m) ↓	5.53	6.90	21.48	6.61	**4.51**	4.79	4.97	8.01	**3.36**	3.65
sfm_house	NCPP (pairs) ↑	27	22	24	29	**34**	**313**	289	290	281	285
	RMSE (m) ↓	**0.23**	0.25	0.45	0.57	0.67	0.43	0.29	0.54	**0.15**	0.36
	Max AE (m) ↓	**0.44**	0.47	0.49	0.88	1.01	0.96	0.92	1.22	**0.46**	0.95
sfm_lab	NCPP (pairs) ↑	11	**13**	**13**	12	10	63	88	**96**	90	90
	RMSE (m) ↓	7.28	7.27	7.22	**7.21**	7.26	1.13	0.53	0.53	**0.52**	0.56
	Max AE (m) ↓	7.29	7.29	7.26	**7.23**	7.28	1.51	0.87	1.03	**0.78**	0.96

compute the camera trajectory and a sparse 3D reconstruction of the world in real-time for a variety of situations, from brief desk-related hand-held sequences to car driving through many city blocks [6].

The ORB-SLAM3 architecture's foundation is built using ORB-SLAM2 and ORBSLAM-VI [21]. It is a complete multi-map and multi-session system that may operate in either visual-inertial or pure visual modes, that supports both pinhole and fisheye lens models. The ORB-SLAM3 pipeline, with its primary system components, is depicted in Figure 2.

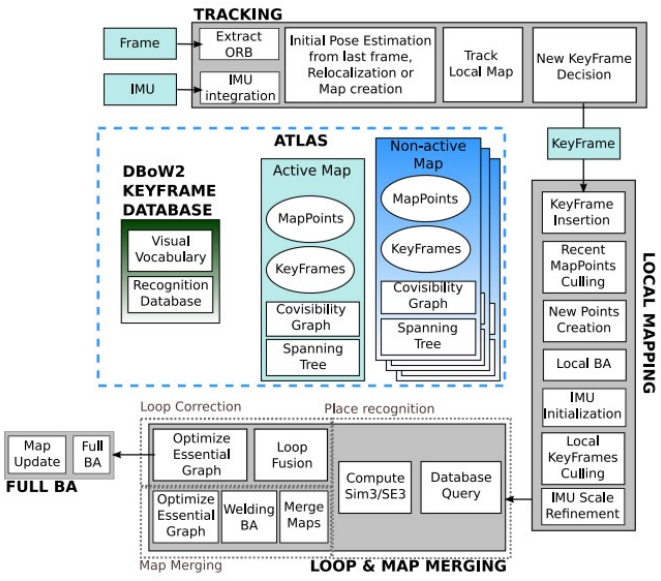

Fig. 2. Main components of ORB-SLAM3 [20]

Over the years, ORB-SLAM3 has proven to be more useful in low-light environments compared to ORB-SLAM2 because it incorporates improvements in several aspects that enhance its ability to operate in challenging lighting conditions. Firstly, ORB-SLAM3 includes a new feature descriptor called COdes of Low Light Features (COLLF), which is designed specifically to capture and match features in low-light environments. This feature descriptor is able to extract more distinctive features in low-light conditions than the feature descriptor used in ORB-SLAM2. Secondly, ORB-SLAM3 uses a new strategy for keyframe selection that is less sensitive to lighting changes. This strategy, called the Asymmetric Keyframe Selection (AKS), takes into account the visual information in the current frame and the surrounding keyframes to select the most suitable keyframe for the current frame. This approach is more robust to changes in lighting conditions, allowing ORB-SLAM3 to operate effectively in low-light environments. Lastly, ORB-SLAM3 incorporates a new loop closure detection method that is more robust to changes in illumination. The new method, called Invariant Covariant Feature Sets (ICFS), is able to detect loop closures even when the lighting conditions have changed significantly between the two views.

In short, ORB-SLAM3 is more useful in low-light environments compared to ORB-SLAM2 due to its improved feature descriptor, keyframe selection strategy, and loop closure detection method, all of which are designed to be more robust to changes in lighting conditions.

B. SuperPoint-SLAM

SuperPoint-SLAM builds off the ORB-SLAM2 architecture and replaces ORB-SLAM2's deterministic feature detector and descriptor, ORB, with SuperPoint, a learning-based feature detector and descriptor. A pre-trained interest point detector and Homographic Adaptation are used to generate pseudo-ground truth interest points in real images to train SuperPoint in a a self-supervised manner, as seen in Figure 3.

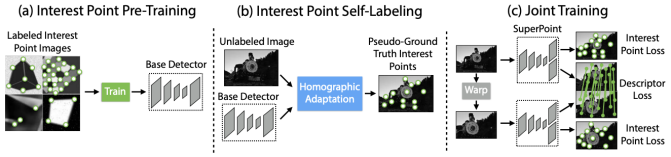

Fig. 3. SuperPoint-SLAM Self-Supervised Training Pipeline [12].

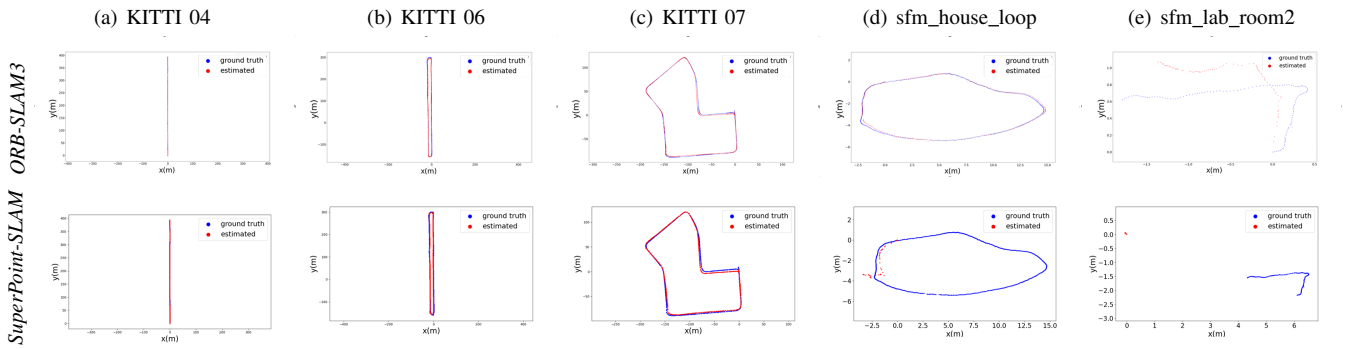

Fig. 4. Comparing the best combination of SLAM architectures and image enhancement modules, over the KITTI and ETH3D low-light SLAM datasets.

As a result, SuperPoint is used to achieve competitive performance in the Berlin Kudamm and MC PhotoTourism benchmarks for Visual Place Recognition and Image Matching tasks respectively when paired with the SuperGlue matching algorithm [11]. Furthermore, SuperPoint computationally efficient architecture runs real-time at 70 FPS with a Titan X GPU [12].

As a result of SuperPoint's appeal along with popularity in the ORB and GCNv2 feature extractors, the authors of SuperPoint-SLAM studied the differences in performances of the ORB-SLAM architecture by switching out each feature detector and their associated binary bag of words originally proposed by [22]. As part of the algorithm's introduction, the authors of SuperPoint-SLAM reported that SuperPoint-SLAM achieved competitive performances with ORB-SLAM on the KITTI SLAM benchmark dataset and failed in dataset sequences with more wild vegetation in the scene. In addition to SuperPoint's invariance to lighting with respect to features and as a result wide use in visual optometry tasks, SuperPoint-SLAM provides a SLAM foundation to provide effective SLAM performance in dynamic lighting environments.

VI. Results

In Table I, we have provided a comparison of the effect of different image enhancement modules on KITTI sequences, and ETH3D datasets, with the Superpoint-SLAM and ORB-SLAM3 architectures. The very first metric we use to compare the image enhancement modules is the number of compared pose pairs (NCPP), which signifies the number of features extracted throughout the sequence. Typically, low-light images will have fewer features extracted, and with image enhancement modules we expect the number of features extracted to increase. Two more evaluation metrics are used, which are, root mean square error (RMSE) and maximum absolute error (Max AE) between the SLAM predicted position and the ground truth. Due to the limited computing power of our system, SuperPoint-SLAM took significantly larger time than ORB-SLAM3 to run the same dataset sequences. Moreover, both ORB-SLAM3 and SuperPoint-SLAM did fail to detect features, with all image enhancement modules, in some of the runs over the ETH3D datasets. So Table I presents the best results among five runs for each of the sequences. After determining which image enhancement module condition had the lowest average RMSE for a dataset, the run with the highest NCPP for the corresponding module condition was chosen as the best run for the entire dataset. We found prioritizing the lowest average RMSE before the highest NCPP found runs with overall better results with respect to NCPP, RMSE, and MAE, as well as qualitatively produced more accurate and full trajectory maps.

Based on these metrics, from the point of features extracted (using NCPP metric), all the enhancement modules resulted in more number of features extracted, when compared with the original dataset, except for the Semi-dark sequence (*sfm_house*) in ORB-SLAM3. On the other hand, it can also be observed that the image enhancement modules led ORB-SLAM3 to achieve the lowest RMSE and Max AE, for all the dataset sequences. However, for SuperPoint-SLAM, the image enhancement modules improved RMSE and Max AE, only for KITTI 06, *sfm_house*, and *sfm_lab* sequences. Comparing the overall performance of the image enhancement modules, EnlightenGAN performed the best, and the same is summarized in Table II, which shows the best combinations of SLAM architectures and image enhancement modules, under different lighting environments. Just to note, the daylight dataset consists of KITTI 04, 06, and 07 sequences, whereas the semi-dark dataset is the *sfm_house* sequence, and the dark dataset is *sfm_lab* sequence.

TABLE II
Best Overall Combinations

Dataset Type	ORB-SLAM3	Superpoint-SLAM
Daylight	EnlightenGAN	Original
Semi-Dark	EnlightenGAN	Original
Dark	EnlightenGAN	EnlightenGAN

Figure 4 shows the best run trajectory maps for ORB-SLAM3 and SuperPoint-SLAM. Among these plots, if we first consider the sequences having loop-closure, which are KITTI sequences 06 and 07, and the *sfm_house* loop, we can notice the SLAM-generated poses are pretty close to the ground truth. However, if we look at KITTI sequence 04 and *sfm_lab* sequence, which do not have loop-closures, we notice the SLAM-generated poses closely match the ground truth only under daylight environments.

Our video is available on YouTube at https://www.youtube.com/watch?v=qe87hcqmZm0

VII. Conclusion

In this project, we did a thorough comparative study on low-light visual SLAM pipelines by evaluating the performance of different combinations of low-light image-enhancement modules and SLAM architectures on various SLAM datasets. By experimenting with different evaluation metrics, we explored the strengths and weaknesses of different combinations and eventually obtained the overall optimal combination for each dataset.

Several limitations are worth noticing. First, the sizes of selected low-light datasets are very small compared to the daylight dataset, making our findings may only apply to limited-scale problems. Second, the low-light image enhancement modules can rely heavily on correction parameters, and they can underestimate or overestimate under extreme lighting conditions and are subject to noise amplification. Finally, the SLAM algorithms may fail to detect features, break, and never recover, under dark environments.

Therefore, in future work, we can test larger or outdoor low-light datasets and experiment with more parameter-tuning in low-light enhancement or consider alternative enhancement models, such as Low-Light Image Enhancement via Illumination Map Estimation (LIME) [23], and other SLAM architectures.

References

[1] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich, "The vslam algorithm for robust localization and mapping," in *Proceedings of the 2005 IEEE international conference on robotics and automation*. IEEE, 2005, pp. 24–29.

[2] A. Beghdadi and M. Mallem, "A comprehensive overview of dynamic visual slam and deep learning: concepts, methods and challenges," *Machine Vision and Applications*, vol. 33, no. 4, p. 54, 2022.

[3] A. Savinykh, M. Kurenkov, E. Kruzhkov, E. Yudin, A. Potapov, P. Karpyshev, and D. Tsetserukou, "Darkslam: Gan-assisted visual slam for reliable operation in low-light conditions," in *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*. IEEE, 2022, pp. 1–6.

[4] M. Ferrera, A. Eudes, J. Moras, M. Sanfourche, and G. Le Besnerais, "Ov2slam: A fully online and versatile visual slam for real-time applications," *IEEE robotics and automation letters*, vol. 6, no. 2, pp. 1399–1406, 2021.

[5] Z. Min and E. Dunn, "Voldor+ slam: For the times when feature-based or direct methods are not good enough," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 813–13 819.

[6] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[7] Y. Zhao and P. A. Vela, "Good feature matching: Toward accurate, robust vo/vslam with low latency," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 657–675, 2020.

[8] P. Ross, A. English, D. Ball, B. Upcroft, G. Wyeth, and P. Corke, "A novel method for analysing lighting variance," in *Australian Conference on Robotics and Automation*, 2013.

[9] P. Ross, A. English, D. Ball, and P. Corke, "A method to quantify a descriptor's illumination variance," in *Proceedings of the 16th Australasian Conference on Robotics and Automation 2014*. Australian Robotics and Automation Association (ARAA), 2014, pp. 1–8.

[10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*. Ieee, 2011, pp. 2564–2571.

[11] 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*, 2020, pp. 4938–4947.

[12] 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*, 2018, pp. 224–236.

[13] P.-E. Sarlin, A. Unagar, M. Larsson, H. Germain, C. Toft, V. Larsson, M. Pollefeys, V. Lepetit, L. Hammarstrand, F. Kahl *et al.*, "Back to the feature: Learning robust camera localization from pixels to pose," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3247–3257.

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

[15] H. Porav, W. Maddern, and P. Newman, "Adversarial training for adverse conditions: Robust metric localisation using appearance transfer," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 1011–1018.

[16] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

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

[18] X. Guo and Q. Hu, "Low-light image enhancement via breaking down the darkness," *International Journal of Computer Vision*, vol. 131, no. 1, pp. 48–66, 2023.

[19] Q. Zhang, Y. Nie, and W.-S. Zheng, "Dual illumination estimation for robust exposure correction," in *Computer Graphics Forum*, vol. 38, no. 7. Wiley Online Library, 2019, pp. 243–252.

[20] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[21] C. Theodorou, V. Velisavljevic, and V. Dyo, "Visual slam for dynamic environments based on object detection and optical flow for dynamic object removal," *Sensors*, vol. 22, no. 19, p. 7553, 2022.

[22] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[23] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.