174B Senior Design Final Report:

## Low-Light Vision Enhancement and Restoration

### Yui Ishihara | [yuiishi@ucdavis.edu](mailto:yuiishi@ucdavis.edu)

### Huy Nguyen | [hkhnguyen@ucdavis.edu](mailto:hkhnguyen@ucdavis.edu)

### Tam Nguyen | [tamnguyen@ucdavis.edu](mailto:tamnguyen@ucdavis.edu)

### Adrian Rivera | [amrivera@ucdavis.edu](mailto:amrivera@ucdavis.edu)

**I. Abstract**

This project focuses on enhancing image and video quality in low-light environments, particularly targeting the improvement of object detection in security camera footage. By addressing the challenges of lack of color and increased noise in low-light imagery, our solution aims to significantly enhance object recognition in dimly lit settings. Through the integration of noise reduction and color enhancement processes, coupled with object detection algorithms like You-Only-Look-Once (YOLO) object detection, we aim to generate color-enhanced images with accurately detected object classes, thereby facilitating improved surveillance capabilities for security applications.

To assess the efficacy of our methodology, we plan to employ metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Average Precision (mAP) to evaluate image quality improvements quantitatively. Additionally, we consider the compatibility of different image enhancement models with the YOLO object detection framework, ensuring optimal performance and accuracy in detecting objects within the reconstructed images. Our comprehensive approach aims to not only enhance image quality but also to optimize object detection capabilities, contributing to advancements in security surveillance effectiveness, and comprehension in low-light conditions.

**II. Introduction / Motivation**

Our project addresses the issue of enhancing image and video in low-light environments. Specifically, we aim to strengthen pre-existing low-light image-enhancing models for object detection in security camera footage. Utilizing deep learning methods, we aim to enhance the clarity of night vision images, typically lacking sufficient light and color data, by generating authentic and lifelike color-enhanced outputs.

The significance of solving this problem becomes apparent when considering that conventional low-light images appear dark and hazy to the human eye, rendering objects indistinguishable. By reducing noise and enhancing color in low-light images, we hope to significantly enhance object recognition in dimly lit environments. An image enhancement model can help in pinpointing suspicious activities and facilitating forensic analysis of surveillance footage. Our solution aims to foster advancements in industry-wide safety, efficiency, and comprehension by addressing these challenges.

Low-light image enhancement techniques have evolved significantly, transitioning from traditional methods like histogram equalization and Retinex theory to more sophisticated deep learning. While traditional methods are often quick and simple, they can struggle with over-enhancement and noise introduction, making them less effective for complex images. Deep learning models, although more powerful in capturing details and reducing noise, require large datasets for training and substantial computational resources, limiting their accessibility. Furthermore, the continuous development of hybrid models aims to leverage the strengths of these approaches, seeking a balance between enhancement quality, computational efficiency, and adaptability to various lighting conditions. However, challenges in achieving consistent performance across diverse scenarios and minimizing computational demands for practical applications remain significant hurdles. Future advancements may focus on optimizing these models for real-time processing and improving their generalizability to unseen lighting conditions.

Initially, we planned to apply an enhancement model to mitigate the dark and grainy quality of the low-light data, transitioning from still images to video footage. This process involves employing noise reduction and color enhancement techniques to enhance image clarity and color. Then, we plan to utilize object detection algorithms such as YOLO to identify and classify objects within the enhanced images For our project, we employed YOLOv5. The output of this pipeline comprises color-enhanced images with accurately detected object classes, along with a comprehensive data dictionary detailing the detected classes and the corresponding number of entities, providing valuable information for security surveillance purposes.



Fig. 1 Proposed Pipeline

**II. Previous Work**

1. [Retinexformer](https://github.com/caiyuanhao1998/Retinexformer)

Huy Nguyen

Retinexformer [1] is a one-stage Retinex-based model to enhance low-light images. One component of the Retinex Former model is the Retinex-based One-stage Transformer (ORF), which estimates illumination information to brighten low-light images and restore corruptions to enhance the overall image quality. They introduce an Illumination-Guided Transformer (IGT) and the Illumination-Guided Multi-head self-attention to model non-local interactions between regions with different lighting conditions, improving the algorithm's ability to capture long-range dependencies.

The model was evaluated on the LOL(v1 and v2), SID, SMID, SDSD, and FiveK datasets. The model’s performance was measured using both PSNR and SSIM for quantitative results, and the model showed performance increases when compared to other various recent models while also having fewer parameters. Qualitative results were achieved using an image detection model. For the low-light images, the detector missed several objects or misidentified objects. However, in the enhanced image, all the objects were correctly identified.

1. [Global Structure-Aware Diffusion Process for Low-Light Image Enhancement](https://github.com/jinnh/GSAD/tree/main?tab=readme-ov-file)

Adrian Rivera

Global Structure-Aware Diffusion Process for Low-Light Image Enhancement (GASD) [2] is a novel diffusion-based method to boost the performance of low-light enhancement from the perspective of regularizing the ODE-trajectory. Diffusion-based generative models have delivered outstanding results with the advancements in denoising diffusion probabilistic models (DDPM) making them increasingly influential in low-level vision tasks. Most of the existing works tend to adopt pixel-wise objective functions to optimize a deterministic relationship, but such regularization frequently produces suboptimal reconstructions resulting in visibly lower reconstruction quality. The DDPM introduces a global structure-aware regularization scheme into the diffusing-based framework by gradually exploiting the intrinsic structure of image data. Introducing this structure-aware regularization scheme, promotes structural consistency and content coherence across similar regions, contributing to a more natural and aesthetically pleasing image enhancement while preserving the image’s fine detail and textures. The authors tested their model with the LOLv1 and LOLv2 datasets. In their paper, the authors have stated that their PSNR is 27.839, and their SSIM is 0.877, respectively.

1. [Implicit Neural Representation for Cooperative Low-light Image Enhancement](https://openaccess.thecvf.com/content/ICCV2023/html/Yang_Implicit_Neural_Representation_for_Cooperative_Low-light_Image_Enhancement_ICCV_2023_paper.html)

Yui Ishihara

Implicit Neural Representation for Cooperative Low-light Image Enhancement (NeRCo) [7] is an innovative method for enhancing low-light images, aiming to overcome several existing limitations. NeRCo employs an implicit Neural Representation approach to enhance images in an unsupervised manner, effectively addressing real-world degradation factors. By incorporating semantic-oriented supervision and a dual-closed-loop constrained enhancement module, NeRCo not only improves robustness but also reduces reliance on paired data, leading to more visually pleasing results. The reported SSIM of this paper is around 0.5360 which shows decent results but with some minor distortion.

1. [Low-Light Image Enhancement with Multi-stage Residue Quantization and Brightness-aware Attention](https://openaccess.thecvf.com/content/ICCV2023/papers/Liu_Low-Light_Image_Enhancement_with_Multi-Stage_Residue_Quantization_and_Brightness-Aware_Attention_ICCV_2023_paper.pdf)

Yui Ishihara

Low-Light Image Enhancement with Multi-stage Residue Quantization (RQ-LLIE) [3] presents a novel approach for enhancing low-light images by using normal-light image priors through a brightness-aware network. By incorporating a query module to extract reliable normal-light features and a brightness-aware attention module, the proposed method achieves more natural and realistic enhancements. Its utilization of prior data from normal-light images improves the robustness of the network to brightness variations. Experimental results demonstrate superior performance over existing state-of-the-art methods on both real-captured and synthetic data, suggesting its potential for enhancing visibility and quality in low-light image scenarios.

1. [Low-Light Image Enhancement with Normalizing Flow](https://github.com/wyf0912/LLFlow?tab=readme-ov-file)

Tam Nguyen

LLFlow [6] introduces an interesting approach for low-light image enhancement using a normalizing flow model, which addresses the challenge of mapping low-light images to their normally exposed counterparts. This model captures the complex conditional distribution of normally exposed images through an invertible network, facilitating the enhancement process. It significantly outperforms traditional methods on benchmark datasets such as LOL and VE-LOL, using metrics like PSNR, SSIM, and LPIPS, indicating superior restoration of color, detail, and reduction of noise and artifacts. On the LOL dataset, LLFlow achieved a PSNR of 25.19 and an SSIM of 0.93. In a cross-dataset evaluation on the VE-LOL dataset, it recorded a PSNR of 23.85 and an SSIM of 0.8986, when trained on LOL and tested on VE-LOL. With intra-dataset evaluation on VE-LOL, after retraining, it showed a PSNR of 26.02 and an SSIM of 0.9266. Additionally, LLFlow incorporates an illumination-invariant color map, enhancing image saturation and reducing color distortion. Extensive experiments validate the effectiveness of each component of LLFlow, showcasing its ability to produce high-quality, naturally enhanced images.

**IV. Evaluation Criteria:**

In evaluating our proposed solution for enhancing image quality in low-light environments, we hope to employ a comprehensive methodology focused on two key metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). We will also employ Mean Average Precision (mAP) to quantitatively assess the effectiveness of

1. **Peak-to-Signal Noise Ratio (PSNR):**

To calculate PSNR, we can begin by determining the dimensions of both the original and enhanced images. We then compute the pixel-wise squared difference between corresponding pixels in the original and enhanced images, subsequently calculating the Mean Squared Error (MSE). From this, we derived the PSNR values, considering both greyscale intensity and RGB values independently for each color channel. PSNR represents a ratio of quality and color intensity between two images

Table I. PSNR Scaling

| < 20 dB | Low quality, significant distortion |
| --- | --- |
| 20 - 30 dB | Acceptable quality, some distortion |
| 30 - 40 dB | Good quality, minor distortion |
| > 40 dB | Excellent quality, no perceptible distortion |

1. **Structural Similarity (SSIM):**

Additionally, we will employ SSIM, which compares luminance, contrast, and structure between the original and reproduced images. Similarly to PSNR, SSIM compares pixels of the initial image and the enhanced image. However, this metric is potentially more accurate in assessing model outputs than PSNR due to its broader focus and more inputs, which will help us evaluate the success of our models.

Table I. PSNR Scaling

| < 0.2 | Poor quality, significant distortion |
| --- | --- |
| 0.2 - 0.4 | Fair quality, noticeable distortion |
| 0.4 - 0.6: | Good quality, minor distortion |
| 0.6 - 0.8 | Very good quality, slight distortion |
| ≥ 0.8 | Excellent quality, no perceptible distortion |

**Mean Average Precision (mAP):**

The mean average precision of a model is the average of its precision results across all detected classes. The precision of a model for a class for a singular image is the number of correct detections over the number of total ground truth objects in the image. A correct detection is defined as a detection bounding box that reaches a specified Intersection over Union (IoU) threshold. The average precision of the class for a model is the average of these precision values of a whole dataset.

**V. Implementation and Results:**

1. **PSNR and SSIM Data**

Initially, we replicated the experiments from the models we researched. Each model was tested on the LOw-Light Dataset (LOLv1) dataset as they were initially trained on LOLv1. Based on our previously researched PSNR and SSIM metric scale, the majority of our models performed with fair quality and some distortion.

Table 1: PSNR and SSIM Comparisons

| **Model** | **PSNR(dB)** | **SSIM** |
| --- | --- | --- |
| **RQ-LLIE** | 28.27 | 0.11 |
| **GSAD** | 27.51 | 0.87 |
| **RetinexFormer** | 25.15 | 0.85 |
| **LLFlow** | 24.06 | 0.91 |

1. YOLO Benchmarking Data

For benchmarking YOLOv5 results, we employed the Exclusively-Dark-Image Dataset (ExDark), as the dataset contains object and bounding box ground-truth data necessary for mAP calculations. The ExDark dataset contains 7,363 images of twelve classes including bus, cat, people, etc. of which 414 were used in a simplified dataset for our project. Our mAP metric was implemented with an IoU threshold of 0.5 or higher. We obtained mAP for each model executed on the ExDark Dataset.



Fig 2: mAP Values Retinexformer, RQ-LLIE, LL-Flow, GSAD (from top left to bottom right)

The total number of classes detected by YOLOv5 was also measured. All models suffered from overestimation on some classes. The total number of objects present in the 441 images we tested was 1,528 objects, and over-estimated classes recorded an average of 42.05 objects overestimated. For classes that were not over-estimated, LL-Flow reached the highest number of classes detected, showing the strongest YOLOv5 results post-enhancement.

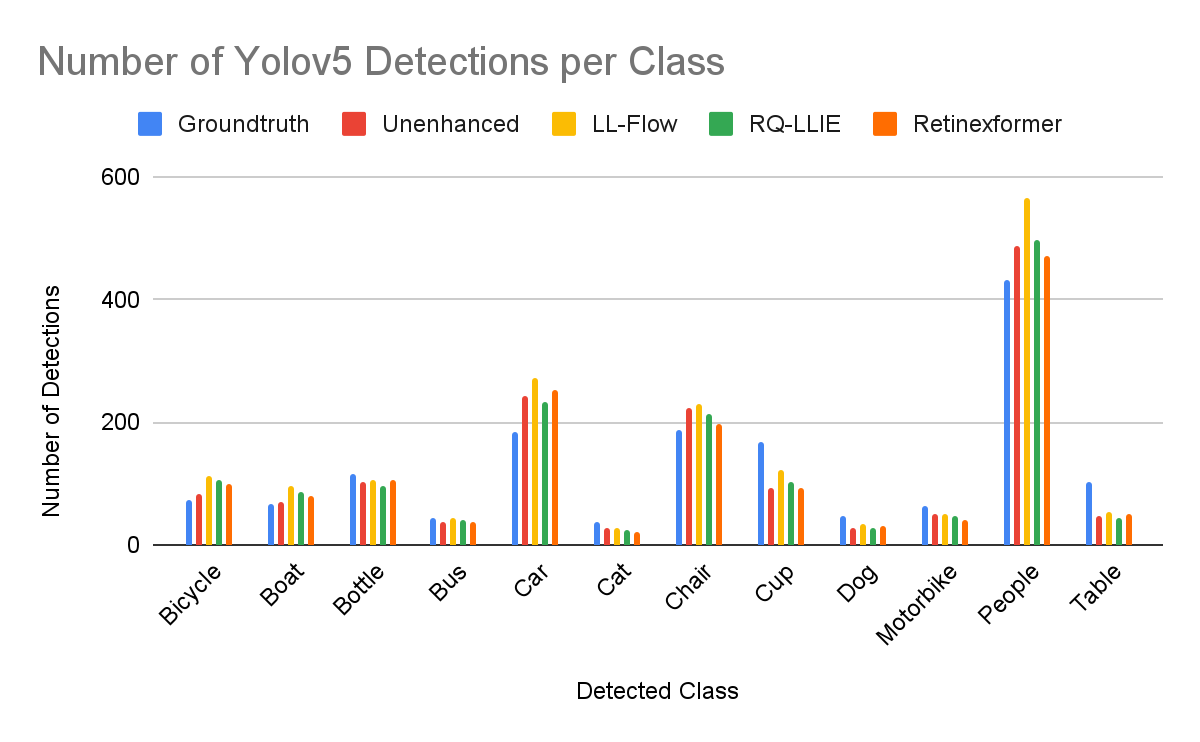


Fig. 3 YOLOv5 Detection Data

**RQ-LLIE Implementation**

*Yui Ishihara*

The RQ-LLIE model exhibits a commendable PSNR of 28.27 dB, reflecting its proficiency in image reconstruction tasks. Despite a moderate SSIM of 0.11, indicating room for improvement in capturing finer details and preserving image structure, the model excels in object detection. Leveraging the YOLO framework on the ExDark dataset, the RQ-LLIE model detected 1,523 objects with an mAP score of 39.17%. The mAP results do not reflect high detection accuracy, but with some optimization of the model, I believe this model can produce more robust results.

**LL-Flow Implementation**

*Tam Nguyen*

LLFlow delivering the highest SSIM score of 0.91, indicates a superior structural similarity to the target images compared to the other models, despite having a lower PSNR value of 24.06 dB. When further evaluating the performance using YOLO object detection on the ExDark dataset, LLFlow's enhanced images achieved the most accurate results, leading to the highest mean Average Precision (mAP) score of 61.86% among the four models. This suggests that LLFlow, while not the top performer in terms of PSNR, excels in preserving image structure and enhancing object detection capabilities in low-light conditions.

**GSAD Implementation**

*Adrian Rivera*

The GSAD model demonstrates its excellence with a commendable PSNR of 27.51 dB, indicative of its proficiency in reconstructing images with high fidelity. Despite facing competition from LLFlow, which boasts an SSIM of 0.91, GSAD still maintains an impressive SSIM score of 0.87, affirming its ability to preserve image structure and details effectively. Utilizing the YOLO object detection framework on the dataset, GSAD's enhanced images exhibit remarkable accuracy, achieving a mAP score of 61.58%.

**RetinexFormer Implementation**

*Huy Nguyen*

The Retinex former model showcases its capabilities with a PSNR of 25.15 dB, indicating its proficiency in image reconstruction tasks. Despite facing stiff competition from models like LLFlow and GSAD, which boast higher SSIM scores, Retinex former maintains a respectable SSIM of 0.85, reflecting its ability to preserve image structure and details effectively. Moreover, when evaluated using YOLO object detection on the dataset, Retinex former's enhanced images demonstrate promising accuracy, achieving a mAP score of 39.86%.

**Video Footage**

From here, we proceeded to incorporate a demonstration of how our proposed later would integrate into a video footage input. The Retinex Former model was selected for the video demonstration based on ease of implementation and clean-looking results. The highest FPS we achieved was 7fps. With optimization of the model, we hope to achieve frames fast enough for a video feed.

**Future Work**

For the future development of our model, several directions are considered to broaden its utility and efficiency. First, testing and adapting the model for challenging conditions such as underwater or foggy environments will significantly expand its applicability, catering to specialized photography and critical applications like surveillance or underwater research. Additionally, a focus on developing a portable version of the model with reduced parameters or layers could optimize its deployment on mobile and embedded devices, making it accessible for real-time applications on smartphones, drones, and compact cameras. Lastly, a more comprehensive benchmarking approach, beyond PSNR and SSIM, could include evaluating the model on metrics such as LPIPS for perceptual quality, color fidelity measures, and computational efficiency to ensure balanced performance across diverse scenarios. This holistic development strategy aims to make the model a versatile tool for various imaging conditions while ensuring it remains practical for widespread adoption. One notable limitation encountered during our work was budgetary and computational constraints, which restricted the scope of experimentation and the complexity of models we could explore. Future efforts will aim to address these limitations by seeking efficient model architectures and leveraging advancements in computational resources.

**Acknowledgements**

We’d love to express our gratitude to Kartik and Professor Chen-Nee Chuah. Thank you for your help and support.

**References**

1. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., & 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 (ICCV) (pp. 12504-12513).
2. Jinhui Hou, & Hui Yuan (2023). Global Structure-Aware Diffusion Process for Low-Light Image Enhancement*. Advances in Neural Information Processing Systems*.
3. Liu, Y., Huang, T., Dong, W., Wu, F., Li, X., & Shi, G. (2023). Low-Light Image Enhancement with Multi-Stage Residue Quantization and Brightness-Aware Attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 12140-12149).
4. Loh, Y., & Chan, C. (2019). Getting to Know Low-light Images with The Exclusively Dark Dataset*. Computer Vision and Image Understanding, 178, 30-42.*
5. Wang J, Yang P, Liu Y, Shang D, Hui X, Song J, Chen X. Research on Improved YOLOv5 for Low-Light Environment Object Detection. Electronics. 2023; 12(14):3089. <https://doi.org/10.3390/electronics12143089>
6. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.-P., & Kot, A. C. (2021). Low-Light Image Enhancement with Normalizing Flow. \*arXiv:2109.05923v1 [eess.IV]\*.
7. Yang, S., Ding, M., Wu, Y., Li, Z., & Zhang, J. (2023). Implicit Neural Representation for Cooperative Low-light Image Enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 12918-12927).

**Appendix**

**Tam:** My role involved running LLFlow and applying YOLOv5 on the enhanced results. I implemented PSNR and SSIM calculations on LLFlow’s enhancement results to benchmark the output images. I also contributed in writing a detection summary script for YOLO outputs.

**Adrian:** My role involved running Global Structure-Aware diffusion model. I helped in benchmarking PSNR and SSIM for my model.

**Yui:** My role involved running NeRCo and RQ-LLIE models. I implemented PSNR and SSIM alongside the other models and structured the repository to contain all models and benchmarking tools. I tested YOLOv5 on LOLv1 dataset, self-made ExDark dataset, and on video enhancement results. I also worked closely with Huy in implementing Retinexformer image enhancement on test videos from the Seeing Dynamic Scene in the Dark Video dataset.

**Huy:** My role involved running the Retinexformer model. I successfully ran a video feed through my model using OBS with a maximum FPS of 7. I also helped by benchmarking PSNR and SSIM results on my LOLv2 results. I then implemented a video input script for Retinexformer to run video instead of image inputs. Using the YOLOv5 results, I ran and tested mAP values, successfully plotting all mAP results for the 4 models.