Night Vision Footage Enhancement and Colorization using Hierarchical Transformer Architecture

Submitted to

Dr. Md. Ashraful Alam

Associate professor, CSE

01. Introduction

- **Importance of Night Vision Technology:**
  - Traditional image colorization relies on reference images or user guides, often leading to unsatisfactory results due to the ill-posed and multimodal nature of the task.
- **Research Gaps:**
  - Existing solutions face challenges in achieving semantic consistency and color richness while balancing computational efficiency and temporal consistency for videos.
- **Goal:**
  - Restore realistic color and detail in low-light conditions.
  - Bridge the gap between enhanced imagery and practical usability.
  - Generate vibrant, semantically accurate, and high-quality visual outputs.

02. Literature Review

- **Manual Methods:**
  - CNN-based approaches outperform traditional methods in realism but require more training data.
  - Poor results on LWIR-only images.
- **Dual Decoder Framework (DDColor):**
  - Separates semantic and texture processing using a guiding fusion module, so that enhanced visual appeal and color consistency.
  - Have high computational cost and Dataset diversity and user evaluation issues.
- **Color-UNet++:**
  - Modified UNet++ with YUV color space to reduce artifacts and improve gradients and Validated on LSUN and LFW datasets.
  - Shallow dataset reduces generalizability.

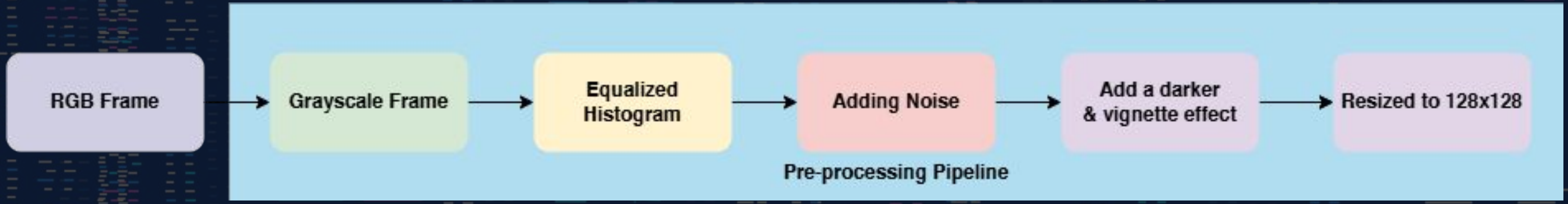
02. Literature Review

- **ColorFormer:**
  - Hybrid-Attention Transformer with a Color Memory Module.
  - Balances local-global dependencies and vivid, semantically rich outputs at real-time speeds (40 FPS).
  - Lowest FID scores and superior results across diverse datasets (ImageNet, COCO-Stuff).
- **Low-Light Image Enhancement (LLIE):**
  - Retinex-based models and GANs but struggles with scene diversity and computational efficiency.
  - Improve significant low-light performance.
  - NTIRE 2024 Challenge highlights the role of hybrid models and diverse datasets.

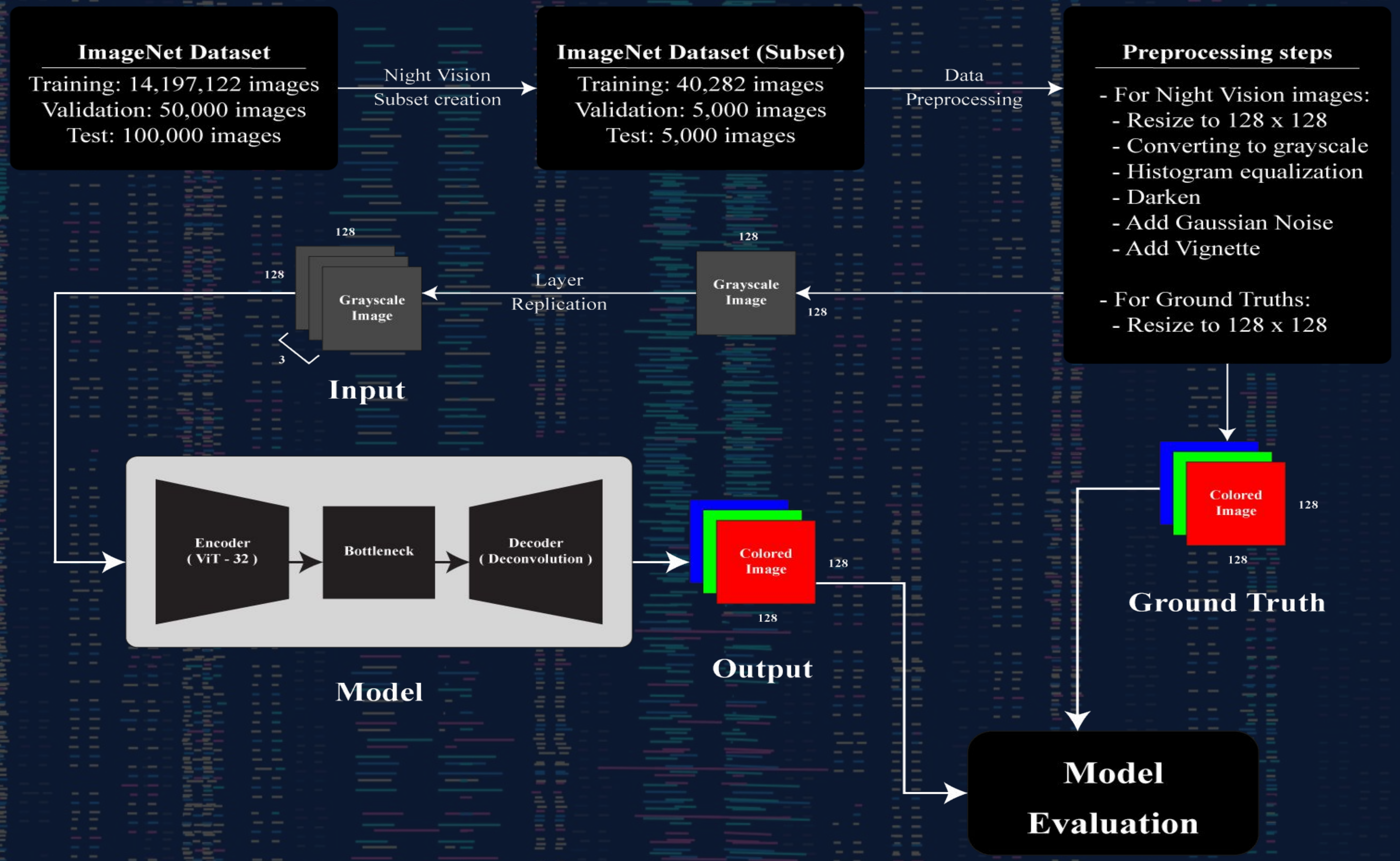
03. Data Pre-processing

- **Dataset Curation:**
  - ImageNet subset reduced to:
    - 40,282 training images.
    - 5,000 testing images.
    - 5,000 brightness-based validation images.

- **Preprocessing Steps:**



04. Architecture Overview

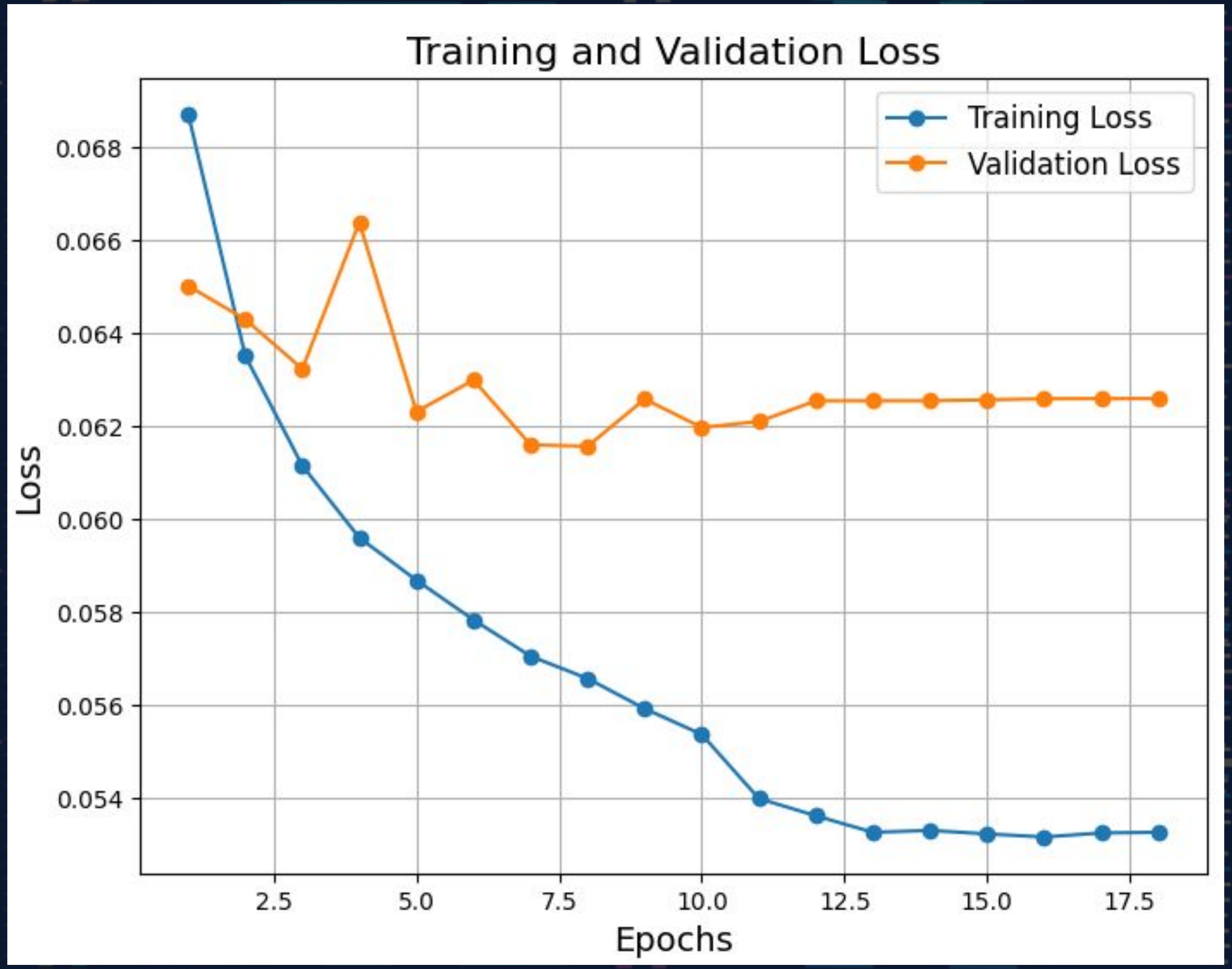


04. Architecture Overview

- **Encoder**:
  - Vision Transformer (ViT-32) structure.
  - Converts hierarchical outputs into single-dimensional vectors.
- **Bottleneck**:
  - Two layers:
    - 1024-unit layer → 8192-unit layer.
    - ELU activation & dropout for learning and overfitting mitigation.
  - Reshaped to 128 channels with an 8x8 shape
- **Decoder**:
  - Four deconvolution stages: 8x8→ 128x128 resolution.
  - ELU activation with final sigmoid layer for output normalization.

05. Model Evaluation

- The Training vs Validation loss curve shows the history of the training epochs. It helps to understand overfitting and underfitting.



05. Model Evaluation

- The model is being evaluated on MSELoss (Mean Squared Error Loss). The loss should be as low as possible.
- Comparing to ResNet and Xception bases, the ViT base is working better in this use case.

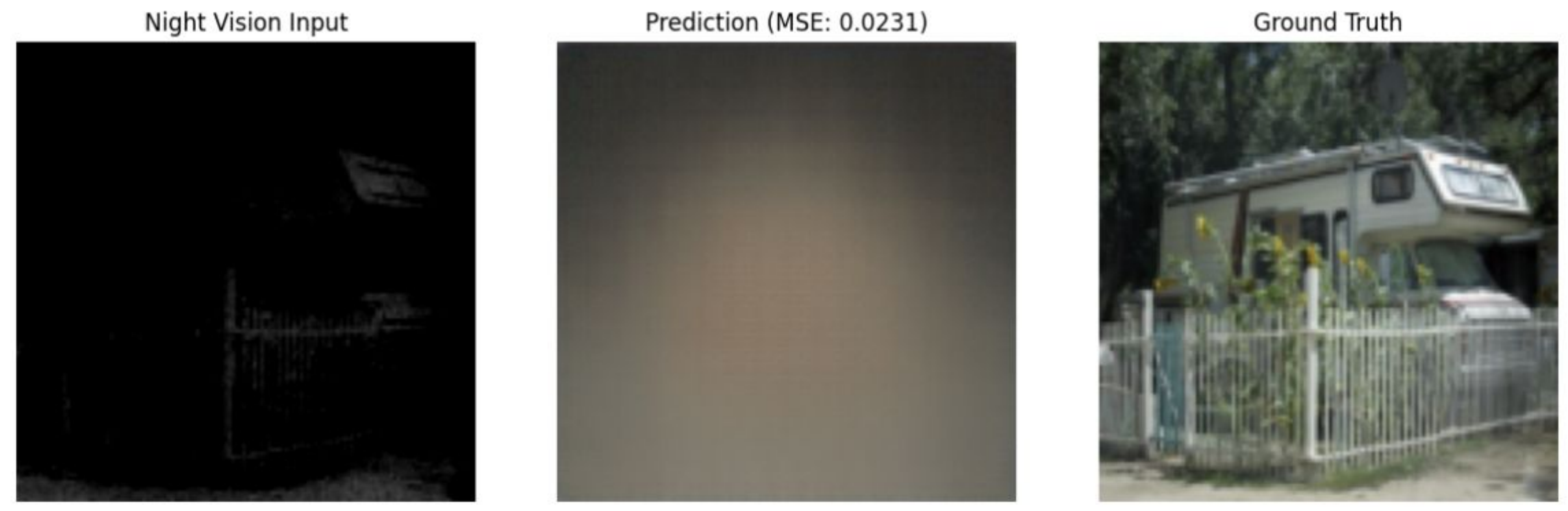
05. Model Evaluation

- We additionally use PSNR (Peak Signal-to-Noise Ratio) to compare our model with other existing models.

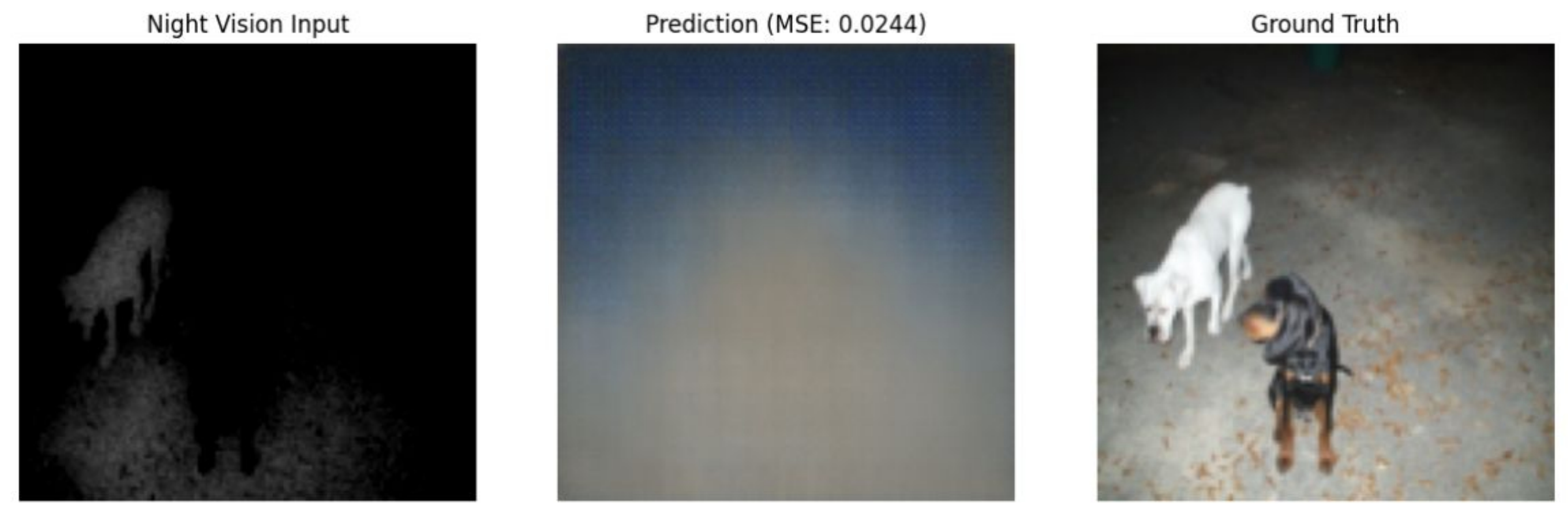
Method	ImageNet			
	FID ↓	CF ↑	Δ CF ↓	PSNR ↑
CIC [2]	19.17	**43.92**	4.83	20.86
Zhang et al. [14]	7.30	27.23	11.86	**24.13**
Instcolor [15]	7.36	27.05	12.04	22.91
ChromaGAN [16]	5.16	27.49	11.60	23.12
DeOldify [17]	3.87	22.83	16.26	22.97
ColTran [18]	6.14	35.50	3.59	22.30
GCP [3]	3.62	35.13	3.96	21.81
BigColor [19]	1.24	40.01	0.92	21.24
Colorformer [4]	1.71	39.76	0.67	23.00
DDColor [5]	1.23	37.72	1.37	23.63
Ours	**1.21**	39.33	**0.24**	23.37

05. Model Evaluation

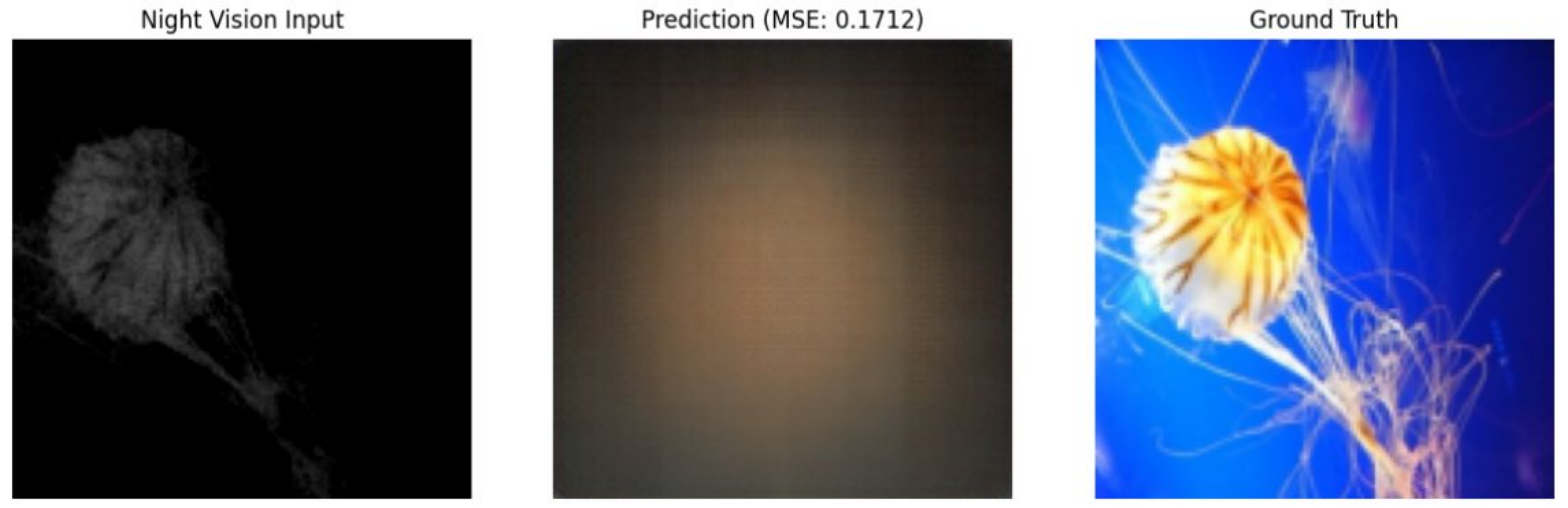
05. Model Evaluation



05. Model Evaluation



05. Model Evaluation



06. Future applications

- **Surveillance and Security**
  - Enhanced monitoring in low-light conditions for critical infrastructure, military bases, and urban security systems.
  - Improved identification and tracking of objects and individuals during nighttime operations
- **Autonomous Vehicles**
  - Safer navigation in low-visibility scenarios, such as nighttime driving or foggy environments.
  - Improved object detection and scene understanding for self-driving cars.
- **Aerospace and Defense**
  - Real-time night vision enhancement for drones, aircraft, and reconnaissance missions.
- **Entertainment and Media**
  - Post-production enhancement of low-light scenes in movies, gaming, and VR applications.

06. Future Work

- **Dataset Expansion**
  - Increase the diversity and size of the training dataset to encompass more complex and varied low-light scenarios, improving model generalization across real-world conditions.
- **Advanced regularization Techniques**
  - Implement advanced regularization strategies such as DropConnect or stochastic depth to mitigate overfitting and enhance model robustness.
- **Edge evaluation with SSIM:**
  - Incorporate Structural Similarity Index (SSIM) as a loss function for structure detection to preserve fine-grained details and improve image clarity.
- **Generator and discriminator for better accuracy**
  - Integrate a generator-discriminator layers to enhance image quality further, ensuring higher fidelity and realism through adversarial learning.

07. Conclusion

Night vision image colorization, leveraging advanced deep learning models like Hierarchical Transformers, has emerged as a transformative tool in enhancing low-light imagery. By improving visibility, semantic accuracy, and color richness, it holds immense potential across diverse fields, from **surveillance** to **autonomous systems**, ensuring robust performance and real-world adaptability for future applications.

07. References

[1] Xiaoyang Kang et al., DDColor: Towards Photo-Realistic Image Coloriza-tion via Dual Decoders, Computer Vision and Pattern Recognition , v(5), 1–10, 2023.

[2] Yide Di et al., Color-UNet++: A resolution for colorization of grayscale images using improved UNet++, Advances of machine learning in data an-alytics and visual information processing , Volume 80, pages 35629–35648, 2021.

[3] Zheng et al., Night Vision Colorization from Color Mapping to Color Trans-ferring, International Conference on Information Fusion (FUSION), Ottawa, ON, Canada 2019, pp. 1-7, doi: 10.23919/FUSION43075.2019.9011375.

[4] Cao et al., Image segmentation for night-vision surveillance camera based on deep learning,13th International Conference on Information Optics and Photonics (CIOP 2022);1247836 (2022) https://doi.org/10.1117/12.2654811

[5] Ji et al., ColorFormer: Image Colorization via Color Memory Assisted Hybrid-Attention Transformer,Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13676. Springer, Cham. https://doi.org/10.1007/978-3-031-19787

[6] Zhang et al., Real-time user-guided image colorization with learned deep priors,ACM Transactions on Graphics;Vol. 36, No. 4, https://doi.org/10.1145/3072959.3073703

[7] Zheng et al., A local-coloring method for night-vision colorization utiliz-ing image analysis and fusion,Information Fusion, Volume 9; https://doi 10.1016/j.inffus.2007.02.002

[8] Toet et al., Colorizing single band intensified nightvision images,Displays 26(1):15-21 https://DOI:10.1016/j.displa.2004.09.007

[9] Tran et al., Low-Light Image Enhancement Framework for Improved Object Detection in Fisheye Lens Datasets,Computer Vision and Pattern Recognition (CVPR); Workshops, 2024, pp. 7056-7065

THANK YOU