LapLoss: Laplacian Pyramid-based Multi-scale loss for Image Translation

Krish Didwania, Prakhar Arya[†]**, Sanskriti Labroo**[†]
Department of Computer Science and Engineering
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, Karnataka, India
{krishdidwania0674,prakhararya13,sanskritilabroo}@gmail.com

Ishaan Gakhar[†]
Department of Information and Communication Technology
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, Karnataka, India
ishaangakhar04@gmail.com

Abstract

Contrast enhancement, a key aspect of image-to-image translation (I2IT), improves visual quality by adjusting intensity differences between pixels. However, many existing methods struggle to preserve fine-grained details, often leading to the loss of low-level features. This paper introduces LapLoss, a novel approach designed for I2IT contrast enhancement, based on the Laplacian pyramid-centric networks, forming the core of our proposed methodology. The proposed approach employs a multiple discriminator architecture, each operating at a different resolution to capture high-level features, in addition to maintaining low-level details and textures under mixed lighting conditions. The proposed methodology computes the loss at multiple scales, balancing reconstruction accuracy and perceptual quality to enhance overall image generation. The distinct blend of the loss calculation at each level of the pyramid, combined with the architecture of the Laplacian pyramid enables LapLoss to exceed contemporary contrast enhancement techniques. This framework achieves state-of-the-art results, consistently performing well across different lighting conditions in the SICE dataset.

1 Introduction

Image-to-image translation (I2IT) (Isola et al., 2017) is a popular task in the field of Computer Vision, which targets the transfer of images mapped from an input domain to an output domain. It is a critical challenge in modern industries, where the demand for complex and precise image transformations is rapidly growing requiring diverse image transformations, from minor augmentations to major format alterations. As the field advances, I2IT has proven highly effective in tasks such as colourizing grayscale images, image illumination, style transfer, and contrast enhancement, with the latter being a key challenge for improving visual clarity in underexposed or overexposed images, particularly in critical applications like autonomous driving (Xia et al., 2022), where accurate and realistic visual representations are essential.

Many traditional methods (Liu et al., 2017) (Wang et al., 2018) are available for performing translation, but they are computationally heavy, often requiring large inference times due to the complexity of the algorithms involved. This work focuses on tackling contrast enhancement and combatting the existing limitations with a novel approach. To this end, the Laplacian Pyramid (Liang et al., 2021), a powerful multi-scale image processing approach that decomposes any image into a series of levels

[†] Equal Contribution.

that represent distinct low and high-level features and from which the original image can be reconstructed is employed. Initially, it involves the creation of a Gaussian pyramid from the original image through downsampling and at each level, the difference between the original and the smoothened version is calculated to capture finer structural intricacies. This method preserves important features across different scales, enabling models to better handle the structure of an images, which ultimately enhances the quality of image translation. It addresses the previous challenges through its lightweight architecture (Huang et al., 2022), which supports faster inference times while delivering results comparable to other state-of-the-art models (SOTA).

In this work, Generative Adversarial Networks employed (GANs) to enhance the translational networks for I2IT (Denton et al., 2015). This competitive process between the generator and discriminators encourages incremental improvements, refining the generator's outputs to address subtle changes like variations in saturation and other visual attributes, successfully fooling the discriminator over time. Due to this competitive framework, GANs prove to be highly versatile with their use cases (Islam et al., 2024); (Sauer et al., 2023), and produce highly realistic outputs. However, GANs are susceptible to unstable training and mode collapse (Durall et al., 2020), where the generator fails to capture the full data distribution and repeatedly produces similar outputs. Additionally, GANs may suffer from vanishing gradients, making optimization difficult. Generative tasks in high-dimensional spaces demand more robust solutions for ensuring efficiency and latent space leads to further instability of training.

Major contributions in this work are summarized as:

- Proposed a novel approach that integrates pixel-wise loss with adversarial losses across multiple scales of a Laplacian pyramid, achieving state-of-the-art results.
- Performed extensive experiments to analyze the impact of the translational network at each pyramid level, demonstrating its adaptability and generalizability across nine different contrast levels, effectively handling underexposed, overexposed, and mixed-exposure conditions.
- Showcased the robustness of the proposed loss function through cross-validation, achieving competitive performance to other frameworks for contrast enhancement.

2 Related Works

2.1 Traditional Low-Light Enhancement Methods and Contrast Enhancement

The low-light image enhancement (LLIE) discipline has transitioned significantly over the years. Elementary approaches relied upon histogram equalization (Woods & Gonzalez, 2018), which redistributes pixel intensities across the entire range. Further contrast enhancement in each tile and modifying the histogram equalization process (Reza, 2004; Tian & Cohen, 2017). Despite their utility, the persistent shortcomings of histogram-based methods in achieving accurate colour restoration—particularly under non-linear illumination distortions—] motivated the adoption of gamma correction (Huang et al., 2012; 2013) by applying power-law transformation compensated for the non-linear response of display devices and for enhancing details in dark or bright regions of an image. However, it does not address issues like noise or colour distortion, which are common in low-light conditions.

Deep learning has improved LLIE by addressing colour shifts, noise, and uneven illumination. RetinexNet (Chen et al., 2018; Wang et al., 2019) and Bread (Guo & Hu, 2023) tackle noise but introduce colour distortion. SNR-Aware (Xu et al., 2022) tries to tackle noise using transformers but struggles with extreme lighting. Recent approaches, including DRBN (Yang et al., 2020) and self-supervised methods (Li et al., 2021), have improved noise suppression and adaptive illumination adjustment. Another direction of using Diffusion models led to innovations including DDPMs (Ho et al., 2020), PyDiff (Zhou et al., 2023), and Diff-Retinex (Yi et al., 2023) which improved noise handling but suffered from inefficiency and over-exposure. KinD (Zhang et al., 2019) introduces noise-aware priors, yet the high costs and distortions remain.

Generative models, especially GANs and Cycle GANs, have shown incredible prowess in contrast enhancement across domains. In the medical domain, GANs have been applied to generate contrast-enhanced MRI and CT images (Cheng et al., 2024). For example, CGAN-based models yield notable

SSIM values for the synthetic T1-weighted brain MRI images (Solak et al., 2024); the deep learning frameworks for sCECT enhance mediastinal lymph nodes lesion conspicuity and contrast-to-noise ratios (Choi et al., 2021). Nonetheless, there are significant limitations for generative models (Skandarani et al., 2023), which limit their wider applicability. For instance, in medical imaging, their performance mostly relies on homogeneous data sets and often shows less generalization (Choi et al., 2021). Moreover, their heavy computational nature postpones real-time deployment (Solak et al., 2024) (Wang et al., 2023) hindering practical capabilities.

2.2 Laplacian Pyramids in I2IT

Pyramid decomposition, originally used for multi-resolution analysis like DWT (Burt & Adelson, 1983), has been widely adopted in machine learning. Techniques like using convolutional networks in a Laplacian pyramid for image generation (Denton et al., 2015) and SPD for textures (Thakur & Chubach, 2015) have advanced this approach. Our work builds on this by employing a Laplacian pyramid to decompose images, enabling better high-resolution detail restoration. Parallely, The fastFF2FFPE method (Fan et al., 2022), which uses a Laplacian Pyramid to decompose FF histopathological images into low- and high-frequency components for efficient FFPE-style translation, offers notable computational advantages, including faster inference and lower memory usage compared to methods like vFFPE and AI-FFPE. However, it is quantitative results and perceptual quality remain comparable without surpassing existing approaches, limiting its appeal for high-fidelity applications and high-frequency details.

A recent development of the Laplacian Pyramid Translation Network (LPTN) (Liang et al., 2021) is a scalable deep learning framework for image enhancement tasks such as low-light enhancement. LPTN leverages a Laplacian Pyramid to decompose images into low-frequency global features and high-frequency components independently processed by lightweight neural networks, enhancing global and local features while minimizing computational costs. The Laplacian Pyramid Super-Resolution Network (LapSRN) (Lai et al., 2017) progressively reconstructs high-resolution images by predicting residuals in a coarse-to-fine manner using a Laplacian pyramid framework. Despite the improvements made, LPTN suffers from limitations in addressing mixed exposures, where it fails to focus on localized behaviour and does not adapt to changes in dynamic lighting (Zhou et al., 2019), curtailing its ability to perform at an optimal level on challenging datasets. The research by (Rathore et al., 2025) highlights the effectiveness of the LPTN architecture when adapted to process both underexposed and overexposed images, albeit with a significant computational cost.

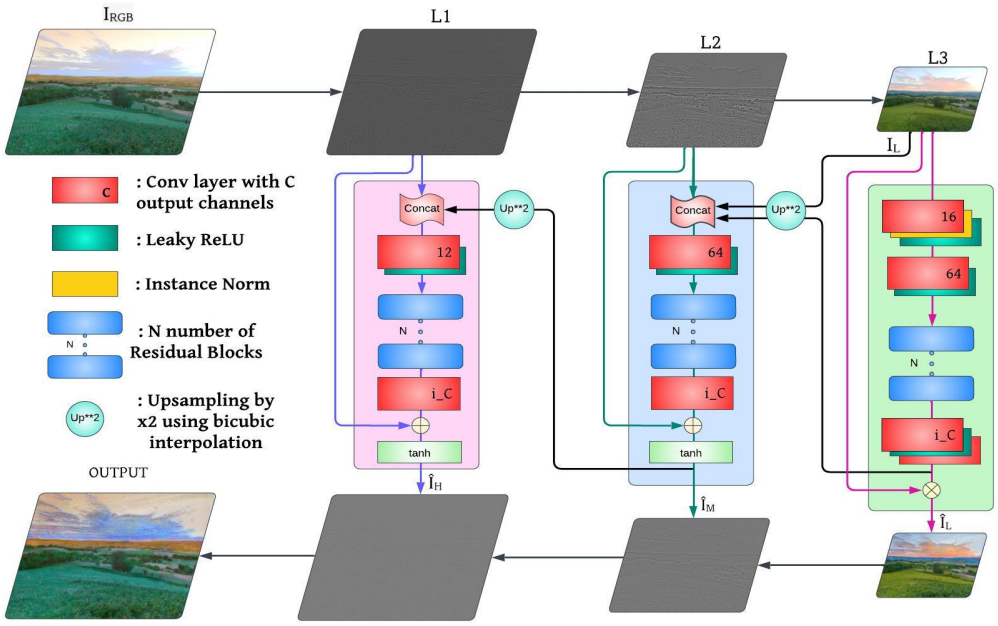
3 Proposed Methodology

3.1 Architecture

This section describes the architecture employed to generate the Laplacian pyramids of the contrast-enhanced image. LapGSR (Kasliwal et al., 2024), a lightweight model designed for Guided Super Resolution is employed for this task. With some minor changes to the model, it is configured for single-image processing and reconstruction of Laplacian pyramids. As evident in Fig. 1, LapGSR merges features from various levels using residual blocks in 3 different branches to reconstruct a Laplacian pyramid. Each branch of the model extracts features from the Laplacian pyramid and reconstructs the corresponding pyramid layer. The number of residual blocks in each layer is represented as N_{Top}, N_{Middle} and N_{Low}. The ideal configuration of these residual blocks is decided by exhaustive experimentation. This pyramid is responsible for the final non-parametric reconstruction of the output image by an inverse Laplacian operation. Here, we modify the pipeline to preserve each layer of the pyramid and use it to compute loss against the Laplacian pyramid of the Ground Truth, as represented in Fig. 2. A more detailed explanation of the architecture is included in the appendix A.

Each discriminator, corresponding to a level of the translational network's output, contains 4 residual blocks for level 0, 4 residual blocks for level 1, and 3 residual blocks for level 2. Its architecture also includes instance normalization Ulyanov et al. (2016) and Leaky ReLU Maas et al. (2013).

Figure 1: Schematic overview of the LapGSR model (Kasliwal et al., 2024) employed for multi-level adversarial processing (Fig. 2). Each branch helps to extract features to finally non-parametrically reconstruct the output image. The instance in the figure is taken from the test set of SICE V1 dataset (Zheng et al., 2024).



3.2 Loss Function

In the proposed method, a distinct discriminator is introduced at each level of the Laplacian pyramid, where it is trained to identify whether the images at that level of the reconstructed output from the generator are real or fake. Each level contributes to the final loss, as a result of a weighted average of the losses from different levels. Adversarial loss was combined with a pixel-wise $\mathcal{L}_{\text{MSE}}$ Loss at various scales to demonstrate results and robustness across datasets.

Mean Squared Error Loss (MSELoss) (Goodfellow et al., 2016) is a fundamental loss function commonly used in image-based tasks. It calculates the pixel-wise difference between the squares of the ground truth and the predicted image, averaged across all pixels. This loss ensures that pixel-wise accuracy between the output and the target images is preserved by penalizing larger deviations more severely to maintain structure.

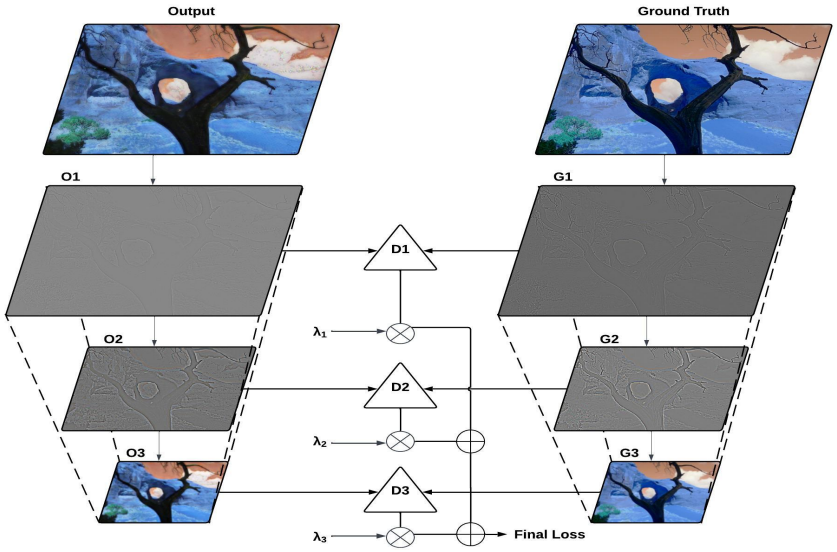
Adversarial Loss is critical to generating realistic images through a Generative Adversarial Network (GAN) framework, it helps both the generator and the discriminator learn from each other. We employ the Least-Square Generative Adversarial Network (LSGAN) loss (Mao et al., 2017) along with pixel-wise MSE to enhance image fidelity by preserving critical spatial attributes. The LSGAN loss for the discriminator and generator is expressed as:

$$\mathcal{L}_{\text{D}} = \frac{1}{2}\mathbb{E}_{x_{\text{real}}}[(D(x_{\text{real}}) - 1)^2] + \frac{1}{2}\mathbb{E}_{x_{\text{fake}}}[D(x_{\text{fake}})^2] \tag{1}$$

$$\mathcal{L}_{\text{GAN}} = \frac{1}{2}\mathbb{E}_{x_{\text{fake}}}[(D(x_{\text{fake}}) - 1)^2] \tag{2}$$

As illustrated in Fig. 2, both the output and ground truth images are decomposed into hierarchical Laplacian pyramid representations. The loss at each level is computed between corresponding pyramid layers, with D_1, D_2, D_3 discriminators. These losses are weighted using $\lambda_1, \lambda_2, \lambda_3$, allowing finer control over how much each scale contributes to the final optimization. This structure aligns well with the LapGSR method, which employs multi-scale, lightweight translational networks

Figure 2: Schematic overview of the multi-scale GAN paradigm. The affine effect is only applied for visual purposes and is not a preprocessing step. It shows the decomposed pyramid of the predicted image and the ground truth across 3 levels. In the final loss, λ_1, λ_2, and λ_3 are the hyperparameters to weight the level-wise loss explained in Section 3.2.



where higher-level branches rely on outputs from lower-level branches. Specifically, these values are weighted and summed to enable the generator network to optimize both the precision of the generated images (measured by MSELoss) and its ability to deceive the discriminator (indicated by the adversarial loss). This balances the loss at each level, which is later scaled using a weighting parameter w. Therefore, the final loss for the generator is defined as:

$$\mathcal{L}_{\text{total}} = \sum_{i=0}^{N} \lambda_i (\mathcal{L}^i_{\text{GAN}} + w\mathcal{L}^i_{\text{MSE}}) \tag{3}$$

where N represents the number of pyramid levels, λ is the weight assigned to each level, $\mathcal{L}^i_{\text{MSE}}$ denotes the pixel-wise MSELoss at level i, and $\mathcal{L}^i_{\text{GAN}}$ represents the adversarial loss for that level. Extensive experimentation was conducted with hyperparameters as discussed in Section 4.

4 Experiments

4.1 Datasets

The dataset used in this work is the publicly available **SICE dataset** (Cai et al., 2018), which comprises 589 high-resolution multi-exposure sequences with a total of 4,413 images. Each sample in the dataset contains either 7 or 9 different contrast levels of the same scene. For this study, we utilized the **SICE V2** dataset, which includes 1,458 images derived from the 229 unique (out of the 589) samples for training.

For testing, images from the **SICE V1** dataset were used, selecting the -1EV (Exposure value) image as the low-light input for underexposure and the +1EV image for overexposure, creating two separate test sets, as per the testing indices provided with the dataset. Additionally, we utilized the **SICEGrad** and **SICEMix** (Zheng et al., 2024) datasets, each containing 529 unique images. These datasets were employed for testing, as they include all possible contrast variations within a single image, replicating the training set. The SICEGrad dataset arranges contrast in increasing or decreasing strips, while the Mix dataset has unordered contrast variations.

4.2 HYPERPARAMETER TUNING

The images used for training were resized to a shape of 608×896, and Vertical Flip, Horizontal Flip, and Shift Scale Rotate augmentations were employed to avoid overfitting of the generator.

Through extensive experimentation, we determined that the optimal ratio of adversarial to reconstruction weighting to be w=4.5×10^3 which effectively stabilized the translation network. Additionally, we found that a learning rate of 10^{-3} was optimal for achieving stable convergence, ensuring robust results.

Table 1: Impact of Level-wise weights on performance across all subsets and datasets. The values show the levels included for loss calculation and their respective weights. The metrics in this table are given as PSNR/SSIM, with higher values indicating better performance.

Levels	Weights	Overexposure	Underexposure	SICEGrad	SICEMix
[0]	[1]	16.54/0.767	16.97/0.726	16.14/0.691	16.09/0.680
[1]	[1]	18.68/**0.774**	17.74/**0.748**	16.49/**0.699**	16.31/**0.726**
[2]	[1]	**18.79**/0.530	**18.94**/0.625	**17.00**/0.648	**16.79**/0.635
[0,1,2]	[4/7, 2/7, 1/7]	19.91/0.714	18.79/0.751	16.74/0.678	16.63/0.668
[0,1,2]	[1/7, 2/7, 4/7]	19.71/0.691	18.87/0.746	16.72/**0.683**	16.57/0.671
[0,1,2]	[1/3, 1/3, 1/3]	**20.33/0.745**	**18.96/0.754**	**16.76**/0.671	**16.64/0.681**

The interplay between pyramid-level weighting and performance is shown in Table 1. Training with only a single-level loss resulted in higher pyramid levels performing better on PSNR, as these focus on fine-grained details important for pixel-wise accuracy. However, SSIM peaked at intermediate levels, since mid-frequency features control structural coherence. In this setup, training with a single-level loss means that only one discriminator was active at a specific pyramid level during GAN training, with both reconstruction and adversarial losses computed exclusively at that level, resulting in N=1 for Equation 3. Additionally, i represents the corresponding level, as shown in the first three rows of 1.

Weighting schemes like $\frac{1}{7}, \frac{2}{7}, \frac{4}{7}$, which emphasize finer levels, degraded performance because they disproportionately prioritize high-frequency details at the cost of mid-frequency textures and global illumination corrections.This shows that multi-level integration is indeed necessary and the equal weights of $\frac{1}{3}$ per level optimally allow balanced contributions from all levels, achieving state-of-the-art metrics.

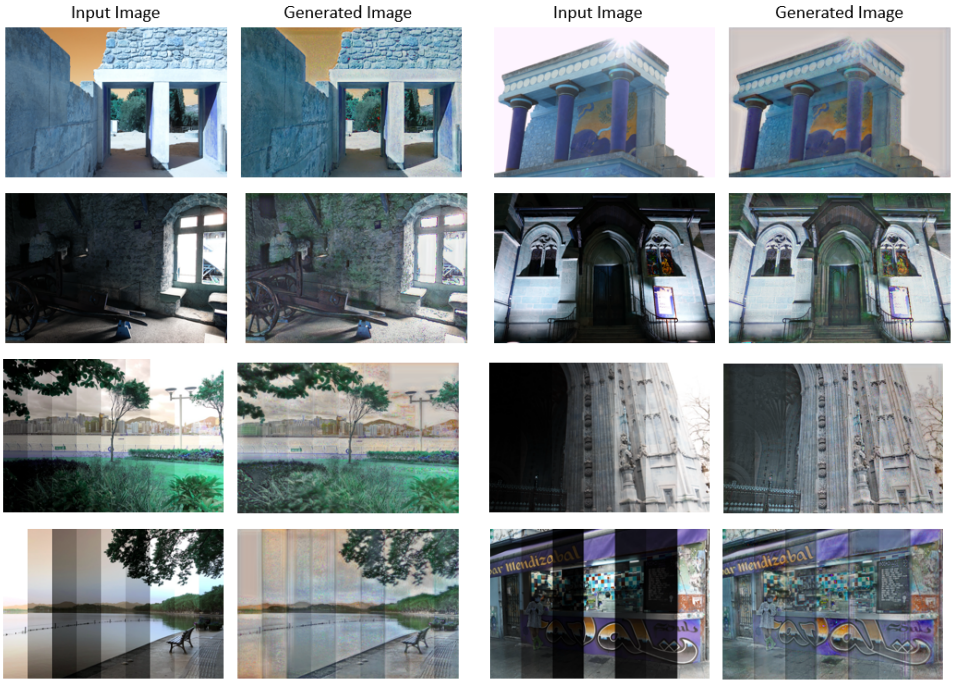
Thorough experimentation was conducted to evaluate the various GAN variants, revealing that LSGAN outperformed WGAN (Wasserstein GAN)Arjovsky et al. (2017), WGAN-Softplus Ding et al. (2020), and HingeGANLim & Ye (2017). Its stable training dynamics and features, such as preventing gradient vanishing, made LSGAN the optimal choice.We employed SOAP (Shampoo with Adam in the Preconditioner's eigenbasis) (Vyas et al., 2025) for optimizing the generator and AdamW Kingma & Ba (2014) for each discriminator in the GAN training process.

Furthermore, the number of residual blocks in LapGSR was systematically adjusted after testing configurations ranging from 3 to 5 blocks for each translational network, with trainable parameters increasing from 694K to 1.13M, as detailed in the apendix A. The explanation and metrics related to these configurations are provided in the ablation section. Ultimately, we proceeded with LSGAN and the 3, 3, 3 lightweight framework for further experiments, comparing our results with those of other approaches.

Structural Similarity Index Measure (SSIM) (Wang et al., 2004) and Peak Signal-to-Noise Ratio (PSNR) (Brooks et al., 2008) were used as the metrics of evaluation due to their complementary strengths in measuring the quality of the image. PSNR is pixel-level reconstruction accuracy, quantifying noise and distortion. SSIM evaluates perceptual quality based on luminance, contrast, and structural similarity, making it closer to human visual perception.

Through extensive hyperparameter tuning, our architecture achieved state-of-the-art results on the SICE dataset. This systematic optimization highlights the importance of customized hyperparameter strategies in improving low-light and high-light image enhancement, showcasing the effectiveness of our Laplacian pyramid-based approach.

Figure 3: Input and output taken for various samples across all datasets. The images in the 1st row are taken from the Overexposure set, 2nd row is taken from Underexposure, 3rd are taken from SICEMix and 4th are taken from SICEGrad. All images are from the test sets.



5 RESULTS

The proposed method, LapLoss, effectively mitigates illumination inconsistencies such as non-uniform noise and abrupt luminance transitions. Exhaustive experiments conducted on the SICE mixed-illumination testing sets (SICEGrad and SICEMix) demonstrate our method's ability to balance striated darkness—alternating bands of underexposed and well-lit regions found in cloud-shadowed landscapes or unevenly lit interiors.

Trained on images with nine contrast levels, ranging from extreme underexposure to overexposure, it effectively corrects contrast across diverse lighting conditions. As demonstrated in Fig. 3, the framework enhances both brightly lit and dimly lit conditions, producing outputs closely aligned with ground truth images. Integrating Laploss with LapGSR ensures the preservation of textures, colours, and details across varying illumination, thereby demonstrating excellent generalizability.

As shown in Table 2, our proposed model achieves the highest SSIM across all lightweight models in recent years. While it does not produce state-of-the-art PSNR on the overexposure and underexposure test sets, it outperforms other methods in SSIM, demonstrating superior image structure preservation. This improvement over previous models highlights its ability to maintain texture consistency and natural luminance gradients, which are crucial for human perception. Results of LapGSR outdo LPTN with Laploss, which are further outperformed by LapGSR and Laploss. The significant increase of **10%** in a few metrics between LapGSR and LapGSR with LapLoss thereby validates our methodology.

In the Table 3, we aim to establish the robustness of our method and achieve notable enhancements in both PSNR and SSIM across the SICEGrad and SICEMix at only **694K** trainable parameters. An increase of **30%** compared to the previous SOTA is observed in the PSNR on the SICEGrad dataset. Similarly, a **20%** improvement is noted in PSNR on the SICEMix dataset. Along with this, a steady increase in SSIM is observed across all datasets. Thus, the proposed loss propels us to excel at scenarios where abrupt lighting transitions or spatially varying exposures challenge conventional methods. A notable difference in performance is noted between LapGSR and LapGSR with LapLoss, validating our proposed methodology.

Table 2: Comparing our results on SICE test sets against other models. The best, second-best, and third-best metrics per column are highlighted in blue, orange, and red, respectively. †Experiments conducted by authors.

Method	Underexposure		Overexposure		Average	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
LCDPNet (Wang et al., 2022)	17.45	0.562	17.04	0.646	17.25	0.604
DRBN (Yang et al., 2020)	17.96	0.677	17.33	0.683	17.65	0.680
DRBN+ERL (Huang et al., 2023)	18.09	0.674	17.93	0.686	18.01	0.680
DRBN-ERL+ENC (Huang et al., 2023)	22.06	0.705	19.50	0.721	20.78	0.713
ELCNet (Huang & Belongie, 2017)	22.05	0.689	19.25	0.687	20.65	0.686
IAT (Cui et al., 2022)	21.41	0.660	22.29	0.681	21.85	0.671
ELCNet+ERL (Huang et al., 2023)	22.14	0.691	19.47	0.692	20.81	0.695
FECNet (Huang et al., 2019)	22.01	0.674	19.91	0.696	20.96	0.685
FECNet+ERL (Huang et al., 2023)	22.35	0.667	20.10	0.689	21.22	0.678
U-EGformer (Adhikarla et al., 2024)	21.63	0.711	19.74	0.705	20.69	0.707
U-EGformereaf (Adhikarla et al., 2024)	22.98	0.719	21.84	0.710	22.41	0.717
LPTN+LapLoss†	18.94	0.653	20.26	0.698	19.60	0.676
LapGSR†	19.33	0.662	20.62	0.708	19.97	0.685
LapGSR+LapLoss†	19.42	0.732	21.32	0.766	20.37	0.749

Table 3: Comparison of results of the proposed method on SICEGrad and Mix sets against other models. The best, second-best, and third-best metrics per column are highlighted in blue, orange, and red, respectively. †Experiments conducted by authors.

Method	SICEGrad		SICEMix	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
RetinexNet (Wei et al., 2018)	12.40	0.606	12.45	0.619
ZeroDCE (Guo et al., 2020)	12.43	0.633	12.48	0.644
RAUS (Zhang et al., 2021)	0.86	0.493	0.86	0.494
SGZ (Zheng & Gupta, 2021)	10.86	0.607	10.99	0.621
LLFlow (Wang et al., 2021)	12.74	0.617	12.74	0.617
URetinexNet (Wu et al., 2022)	10.90	0.600	10.89	0.610
SCI (Ma et al., 2022)	8.64	0.529	8.56	0.532
KinD (Zhang et al., 2021)	12.99	0.656	13.14	0.668
KinD++ (Zhang et al., 2021)	13.20	0.657	13.24	0.666
U-EGformer (Adhikarla et al., 2024)	13.27	0.643	14.24	0.652
U-EGformer† (Adhikarla et al., 2024)	14.72	0.665	15.10	0.670
LPTN+LapLoss†	17.32	0.657	16.67	0.624
LapGSR†	17.27	0.629	16.71	0.623
LapGSR+LapLoss†	17.33	0.657	17.13	0.648

6 Conclusion and Future Works

In this work, we introduced **LapLoss**, a novel approach leveraging Laplacian Pyramid Translational Networks (LPTNs) for various I2IT tasks. Our findings highlight its adaptability in restoring images across diverse contrast levels, emphasizing the significance of multi-level processing. We also analyze each level's contribution to performance.

For future work, we aim to explore broader generator and pixel-wise loss functions to enhance robustness and accuracy. The insights from this study can extend to super-resolution, debanding, and denoising, promoting a unified approach to image restoration using LPTNs and advanced loss functions.

ACKNOWLEDGMENTS

The authors would like to thank the Research Society Manipal, a student-run organization at Manipal Institute of Technology, Manipal, for providing the necessary resources for this work. We also express our gratitude to Aditya Kasliwal and Pratinav Seth for their valuable contributions in refining the quality of the final manuscript.

REFERENCES

Suraj Adhikarla et al. Unified-egformer for mixed-exposure image enhancement. *Journal of Image Processing*, 2024.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Alan C. Brooks, Xiaonan Zhao, and Thrasyvoulos N. Pappas. Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. *IEEE Transactions on Image Processing*, 17(8):1261–1273, 2008. doi: 10.1109/TIP.2008.926161.

P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. doi: 10.1109/TCOM.1983.1095851.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.

Wei Chen, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018.

Ka-Hei Cheng, Wen Li, Francis Kar-Ho Lee, Tian Li, and Jing Cai. Pixelwise gradient model with gan for virtual contrast enhancement in mri imaging. *Cancers*, 16(5):999, 2024. doi: 10.3390/cancers16050999.

Jae Won Choi, Yeon Jin Cho, Ji Young Ha, Seul Bi Lee, Seunghyun Lee, Young Hun Choi, Jung-Eun Cheon, and Woo Sun Kim. Generating synthetic contrast enhancement from non-contrast chest computed tomography using a generative adversarial network. *Scientific Reports*, 11(1): 20403, 10 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-00058-3. URL https://doi.org/10.1038/s41598-021-00058-3.

Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada. You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press, 2022.

E. L. Denton, S. Chintala, and R. Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, pp. 1486–1494, 2015.

Xin Ding, Z Jane Wang, and William J Welch. Subsampling generative adversarial networks: Density ratio estimation in feature space with softplus loss. *IEEE Transactions on Signal Processing*, 68:1910–1922, 2020.

Ricard Durall, Avraam Chatzimichailidis, Peter Labus, and Janis Keuper. Combating mode collapse in gan training: An empirical analysis using hessian eigenvalues, 2020. URL https://arxiv.org/abs/2012.09673.

Lei Fan, Arcot Sowmya, Erik Meijering, and Yang Song. Fast ff-to-ffpe whole slide image translation via laplacian pyramid and contrastive learning. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pp. 409–419, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16434-7.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023. ISSN 1573-1405. doi: 10.1007/s11263-022-01667-9. URL https://doi.org/10.1007/s11263-022-01667-9.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Jie Huang, Zhiwei Xiong, Xueyang Fu, Dong Liu, and Zheng-Jun Zha. Hybrid image enhancement with progressive laplacian enhancing unit. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, pp. 1614–1622, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368896. doi: 10.1145/3343031.3350855. URL https://doi.org/10.1145/3343031.3350855.

Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6043–6052, 2022.

Jie Huang, Feng Zhao, Man Zhou, Jie Xiao, Naishan Zheng, Kaiwen Zheng, and Zhiwei Xiong. Learning sample relationship for exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9904–9913, June 2023.

S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 22(3):1032–1041, 2013.

Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 22(3):1032–1041, 2012.

X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

Showrov Islam, Md. Tarek Aziz, Hadiur Rahman Nabil, Jamin Rahman Jim, M. F. Mridha, Md. Mohsin Kabir, Nobuyoshi Asai, and Jungpil Shin. Generative adversarial networks (gans) in medical imaging: Advancements, applications, and challenges. *IEEE Access*, 12:35728–35753, 2024. doi: 10.1109/ACCESS.2024.3370848.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, 2017. doi: 10.1109/CVPR.2017.632.

Aditya Kasliwal, Ishaan Gakhar, Aryan Kamani, Pratinav Seth, and Ujjwal Verma. Lapgsr: Laplacian reconstructive network for guided thermal super-resolution. *arXiv preprint arXiv:2411.07750*, 2024.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conferene on Computer Vision and Pattern Recognition*, 2017.

Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2021.

Jie Liang, Hui Zeng, and Lei Zhang. High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9387–9395, June 2021. doi: 10.1109/CVPR46437.2021.00927.

Jae Hyun Lim and J. C. Ye. Geometric gan. *ArXiv*, abs/1705.02894, 2017. URL https://api.semanticscholar.org/CorpusID:9010805.

Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 700–708, 2017.

Yuanming Ma et al. Toward fast, flexible, and robust low-light image enhancement. *IEEE Transactions on Image Processing*, 2022.

Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, pp. 3, 2013.

Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, 2017. doi: 10.1109/ICCV.2017.304.

Shaurya Singh Rathore, Aravind Shenoy, Krish Didwania, Aditya Kasliwal, and Ujjwal Verma. Hipyrnet: Hypernet-guided feature pyramid network for mixed-exposure correction, 2025. URL https://arxiv.org/abs/2501.05195.

Ali M. Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *IEEE Transactions on Multimedia*, pp. 35–44, 2004.

Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. Stylegan-t: Unlocking the power of gans for fast large-scale text-to-image synthesis, 2023. URL https://arxiv.org/abs/2301.09515.

Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. Gans for medical image synthesis: An empirical study. *Journal of Imaging*, 9(3):69, 2023. doi: 10.3390/jimaging9030069.

Merve Solak, Murat Tören, Berkutay Asan, Esat Kaba, Mehmet Beyazal, and Fatma Beyazal Çeliker. Generative adversarial network based contrast enhancement: Synthetic contrast brain magnetic resonance imaging. *Academic Radiology*, 2024. ISSN 1076-6332. doi: https://doi.org/10.1016/j.acra.2024.11.021. URL https://www.sciencedirect.com/science/article/pii/S1076633224008651.

U. S. Thakur and O. Chubach. Texture analysis and synthesis using steerable pyramid decomposition for video coding. In *Proceedings of the International Conference on Systems, Signals, and Image Processing (IWSSIP)*, pp. 204–207, London, U.K., Sep 2015.

Qi-Chong Tian and Laurent D. Cohen. Global and local contrast adaptive enhancement for non-uniform illumination color images. In *ICCV Workshops*, pp. 3023–3030, Oct 2017.

Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *ArXiv*, abs/1607.08022, 2016. URL https://api.semanticscholar.org/CorpusID:16516553.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Mujin Kwun, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. Soap: Improving and stabilizing shampoo using adam, 2025. URL https://arxiv.org/abs/2409.11321.

H. Wang, K. Xu, and R. W. Lau. Local color distributions prior for image enhancement. In *European Conference on Computer Vision*, pp. 343–359. Springer, 2022.

Tenghui Wang, Lili Wang, En Zhang, Yan Ma, Yapeng Wang, Haijun Xie, and Mingchao Zhu. Underwater image enhancement based on optimal contrast and attenuation difference. *IEEE Access*, 11:68538–68549, 2023. doi: 10.1109/ACCESS.2023.3292275.

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 8798–8807, 2018.

Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, Zhiwei Xiong, Wei Zhang, and Feng Wu. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In *MM '19: Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2015–2023. ACM, 2019.

Zheng Wang et al. Low-light image enhancement with normalizing flow. *IEEE CVPR*, 2021.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Chen Wei et al. Deep retinex decomposition for low-light enhancement. *British Machine Vision Conference*, 2018.

Richard E. Woods and Rafael C. Gonzalez. *Digital Image Processing*. Pearson, 2018.

Yifan Wu et al. Uretinex-net: Retinex-based deep unfolding network for low-light enhancement. *IEEE Transactions on Image Processing*, 2022.

Youya Xia, Josephine Monica, Wei-Lun Chao, Bharath Hariharan, Kilian Q Weinberger, and Mark Campbell. Image-to-image translation for autonomous driving from coarsely-aligned image pairs, 2022. URL `https://arxiv.org/abs/2209.11673`.

X. Xu, R. Wang, C.-W. Fu, and J. Jia. Snr-aware low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17693–17703, 2022.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semisupervised approach for low-light image enhancement. In *CVPR*, pp. 3063–3072, 2020.

X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12302–12311, 2023.

Ying Zhang et al. Beyond brightening low-light images. *ACM Transactions on Graphics*, 2021.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pp. 1632–1640. ACM, 2019.

Shen Zheng, Yiling Ma, Jinqian Pan, Changjie Lu, and Gaurav Gupta. Low-light image and video enhancement: A comprehensive survey and beyond, 2024. URL `https://arxiv.org/abs/2212.10772`.

Yu Zheng and Anurag Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. *IEEE Transactions on Neural Networks*, 2021.

D. Zhou, Z. Yang, and Y. Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint*, 2023.

Jingchun Zhou, Dehuan Zhang, Peiyu Zou, Weidong Zhang, and Weishi Zhang. Retinex-based laplacian pyramid method for image defogging. *IEEE Access*, 7:122459–122472, 2019. doi: 10.1109/ACCESS.2019.2934981.

A APPENDIX

In this study, we detail the exact architecture, effect and configuration of the residual blocks. Additionally, we display results with various Adversarial Losses, as well as their effects across datasets.

A.1 RESIDUAL BLOCKS

A detailed explanation of the architecture used for experiments as shown in Fig. 1 is given below:
Lower Transformation Branch (LTB):
The LTB extracts fundamental features such as luminance, texture, and illumination from lower most level of the pyramid. It starts with a convolutional layer, followed by instance normalization and a leaky ReLU activation to prevent vanishing gradients. After this, another convolutional layer with leaky ReLU is applied, followed by several residual blocks consisting of convolutional layers with skip connections. The final feature map, denoted as $\hat{I}_L$, is the output of the LTB. This feature map is upsampled and combined with the second-to-last layer (L_2) of the branch before being passed to the Middle Transformation Branch (MTB). The output of the residual blocks is multiplied with L_1, yielding the final product of the LTB.

Middle Transformation Branch (MTB):
The MTB serves to bridge the low-level features from the LTB and the high-level abstractions for subsequent tasks. It begins with a convolutional layer followed by leaky ReLU activation to extract features while avoiding vanishing gradients. The MTB contains several residual blocks, followed by a final convolutional layer. The feature maps from the LTB and Laplacian pyramid (L_2) are fused and passed through a tanh activation, yielding the intermediate representation, $\hat{I}_M$. This output is upsampled by 2x and concatenated with L_3, before being passed into the High Transformation Branch (HTB).

High Transformation Branch (HTB):
The HTB is the final stage of the transformation pipeline, specializing in synthesizing the contrast enhanced output image. It receives the 2x upsampled output from the MTB, $\hat{I}_M$, and processes it through a convolutional layer, followed by leaky ReLU activation. The residual blocks refine the upsampled features, and a final convolutional layer consolidates them into a corrected contrast feature map. This output is then added to L_1 and passed through a tanh activation to generate the top layer, $\hat{I}_H$, of the translated pyramid. This layer combines detailed texture and abstract features for the final visual output.

Through systematic evaluation as shown in Table 4, employing 5 residual for the lowest level branch ($NLow$), 5 for the intermediate level (NMid), and 5 for the top level (NTop) in the network for balancing the feature depth and computational efficiency. Although the configuration of 5,5,5 in low, middle and top branches achieved the best performance in terms of metrics, we present results using the 3,3,3 configuration while comparing against other models, as the difference in metrics was not substantial despite the increased framework complexity. We also observed that the number of residual blocks in the low-frequency component played a crucial role in the overall quality of the generated images, as it serves as the foundation for translation in our interconnected network.

Params	**Low**	**High**	**Top**	**Over**		**Under**	
				PSNR	**SSIM**	**PSNR**	**SSIM**
768k	4	3	3	21.08	0.763	19.71	0.741
842k	5	3	3	21.21	0.742	19.31	0.7206
916k	5	4	3	21.12	0.773	19.46	0.7239
990k	5	5	3	21.06	0.771	18.81	0.7199
1.06M	5	5	4	21.09	0.766	18.91	0.7049
1.14M	5	5	5	21.32	0.765	19.42	0.7324

Table 4: Test set results for underexposure and overexposure sets.

Another observation we made is that increasing the number of residual blocks in the intermediate layers can lead to a decrease in metrics. This could be due to the fact that, in this particular case of I2IT, it is crucial to preserve the frequency components. In contrast, enhancement techniques

Params	Low	High	Top	Grad		Mix	
				PSNR	SSIM	PSNR	SSIM
768k	4	3	3	17.25	0.658	16.94	0.644
842k	5	3	3	17.33	0.657	17.12	0.647
916k	5	4	3	17.28	0.654	17.03	0.642
990k	5	5	3	17.26	0.656	17.03	0.647
1.06M	5	5	4	17.05	0.642	17.33	0.658
1.13M	5	5	5	17.28	0.659	17.05	0.648

Table 5: Test set results with two metrics for Grad and Mix datasets.

like the intermediate level of a Laplacian pyramid effectively manage this by focusing on the finer details of the frequency spectrum. The intermediate level of a pyramid captures crucial frequency components that help maintain the balance between high-level features and low-level details, which can lead to improved performance in tasks that require fine-grained information preservation.

A.2 Adversarial Losses

Table 6: Performance metrics (PSNR/SSIM) for different GAN types across four testing sets. Best metrics are highlighted.

Loss Type	Overexposure	Underexposure	SICEMix	SICEGrad
LSGAN	**20.54/0.75**	**18.88/0.75**	16.79/0.67	16.66/0.66
WGAN	20.00/0.73	18.78/0.73	16.75/**0.68**	16.62/**0.67**
WGAN_SOFT+	20.13/0.75	18.72/0.72	**16.80**/0.67	16.61/0.66
HINGE	19.02/0.68	19.65/0.72	16.80/0.65	**16.64**/0.64

Our proposed loss function for Laplacian pyramid-centric generators primarily focuses on adversarial losses at each level. To evaluate its effectiveness, we experimented with various GAN loss functions, as summarized in Table 6. The results indicate that most loss functions performed similarly, with minor variations in performance metrics.

Based on our experiments, LSGAN consistently produced the highest-quality images, achieving the best average metrics across all four testing sets. As shown in Table 6, LSGAN outperformed other GAN loss functions, particularly excelling in the Overexposure test set. This highlights its robustness in handling bright regions and its overall effectiveness in image enhancement tasks.