Illuminating Darkness: Learning to Enhance Low-light Images In-the-Wild

S. M. A. Sharif[1] Abdur Rehman[1] Zain Ul Abidin[2]
Fayaz Ali Dharejo[3] Radu Timofte[3*] Rizwan Ali Naqvi[2]
[1]Opt-AI Inc. [2]Sejong University [3]University of Würzburg
Code and dataset: github.com/sharif-apu/LSD-TFFormer

This supplementary document details the LSD preparation and insights of TFFormer, provides additional experimental results, and discusses the limitations of the proposed method. We organized the document as follows:

- Section 1 details LSD preparation
- Section 2 details the distribution of the LSD and illustrates the limitation of the existing dataset visually.
- Section 3 provides learning details and performance of the TFFormer in existing SLLIE.
- Section 4 provides additional experimental results

1. LSD Preparation

We designed our LSD app carefully to collect a large-scale SLLIE dataset leveraging multiple sensors for diverse real-world data collection. We also developed our scene classifier by utilizing the capability of SOTA image classification models.

1.1. Sensor Details

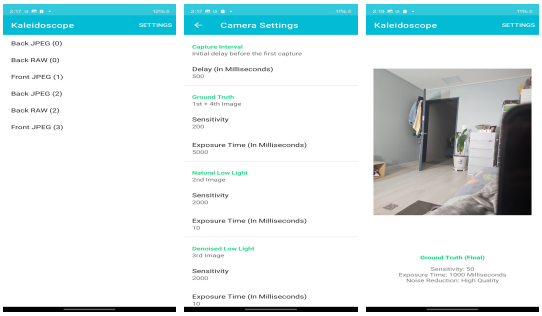
To diversify our proposed LSD dataset, we utilize 15 distinct image sensors from six different smartphones, encompassing a variety of imaging principles, resolutions, pixel sizes, and sensor sizes. These sensors respond uniquely to various lighting conditions, particularly noise, dynamic range, and color reproduction. Table 1 provides the specifications of these smartphones and their cameras, with the resolution reflecting the maximum native output of each sensor. However, under low-light conditions, these sensors typically generate 4K images by leveraging pixel-binning techniques to create larger effective pixels, enhancing brightness and reducing noise.

This study replicates real-world scenarios using the default pixel-binning configurations of modern smartphones. Only the rear cameras were employed, as they offer superior imaging performance compared to front-facing cameras. The diversity in sensor types, field of view, and imaging characteristics substantially enriches the LSD dataset, ensuring its robustness and relevance for low-light image enhancement research.

*Radu Timofte and Rizwan Ali Naqvi are the corresponding authors.

1.2. LSD App

Our LSD app was developed by leveraging android-api [7] to capture aligned low-light and corresponding reference images. Before fixing the final version, we experimented with the available APIs to find the best and easy-to-use interface for collecting low-light samples. Fig. 1 illustrates the sample screenshot of our LSD app. To capture a scene, we first selected a random camera sensor for the device; we calibrated the settings based on our capture calibration strategy. Before saving the scenes, we visually inspected the captured scenes to ensure their usability.



(a) Sesnor Selection (b) Capture Settings (c) Preview

Figure 1. User interface and detail of LSD APP. (a) Selecting Sensor. (b) Settings adjustment of low-light and reference images. (c) Preview of capture scenes

1.3. VGGScene Classifier

Refining the collected patches from our image pairs posed significant challenges, mainly due to unwanted imaging limitations such as blurs, over-exposed light sources, dark under-exposed regions, and defocus in reference images. Developing individual algorithms to address each issue proved inefficient and complex, while manually inspecting every image pair was impractical.

To overcome these hurdles, we developed a learning-based scene classifier to automate and accelerate the refinement process. This approach enabled us to identify and exclude imperfect images from training, ensuring higher dataset quality. Before beginning the LSD scene collection,

Device	Camera	Sensor	Resolution (Max.)	Sensor Size	Pixel Size	Aperture	Year
Samsung Galaxy S10	Main (Wide)	S5K2L4	12 MP (4032 x 3024)	1/2.55 in.	1.4 µm	f/1.5-2.4	2019
	Telephoto	S5K3M3	12 MP (4032 x 3024)	1/3.6 in.	1.0 µm	f/2.4	2019
	Ultra-Wide	S5K4H5	16 MP (4608 x 3456)	1/3.1 in.	1.0 µm	f/2.2	2019
Samsung Galaxy Z Flip3	Main (Wide)	IMX555	12 MP (4032 x 3024)	1/2.55 in.	1.4 µm	f/1.8	2021
	Ultra-Wide	IMX258	12 MP (4000 x 3000)	1/3.0 in.	1.12 µm	f/2.2	2021
Xiaomi Redmi 10C	Main (Wide)	ISOCELL JN1	50 MP (8160 x 6144)	1/2.76 in.	0.64 µm	f/1.8	2022
Samsung Galaxy S22 Ultra	Main (Wide)	HM3	108 MP (12032 x 9024)	1/1.33 in.	0.8 µm	f/1.8	2022
	Telephoto (3x)	IMX754	10 MP (3648 x 2736)	1/3.52 in.	1.12 µm	f/2.4	2022
	Telephoto (10x)	IMX754	10 MP (3648 x 2736)	1/3.52 in.	1.12 µm	f/4.9	2022
	Ultra-Wide	IMX563	12 MP (4000 x 3000)	1/2.55 in.	1.4 µm	f/2.2	2022
Samsung Galaxy Z Flip5	Main (Wide)	IMX563	12 MP (4032 x 3024)	1/1.76 in.	1.8 µm	f/1.8	2023
	Ultra-Wide	IMX258	12 MP (4000 x 3000)	1/2.55 in.	1.12 µm	f/2.2	2023
Samsung Galaxy Z Fold5	Main (Wide)	GN5	50 MP (8160 x 6120)	1/1.57 in.	1.0 µm	f/1.8	2023
	Telephoto	IMX754	10 MP (3648 x 2736)	1/3.94 in.	1.0 µm	f/2.4	2023
	Ultra-Wide	IMX563	12 MP (4000 x 3000)	1/3.0 in.	1.12 µm	f/2.2	2023

Table 1. Specifications of the camera sensors utilized to collect the proposed LSD. Utilizing such diverse sensors helps us generalize the SLLIE in the real world for practical use.

we extensively evaluated state-of-the-art (SOTA) classification methods and constructed a separate scene classification dataset to fine-tune our classifier. This allowed us to handle challenging scenarios and streamline the refinement process robustly.

1.3.1. Data Preparation.

Before start collecting the LSD dataset, we captured around 200 random scenes to find the limitations and probable obstacles of preprocessing our LSD. Also, we made a list of unwanted imaging consequences that may affect our learning strategy. Based on the initial study, we made a classification dataset and separated perfect and imperfect scenes into two categories. Tab. 2 illustrates the detail of our scene classification dataset. We collected 12,770 image patches to learn scene classifiers with SOTA classification methods. Fig. 2 demonstrates the sample images from our scene classification dataset.

Class	Train	Test
Perfect	5,580	755
Imperfect	5,456	979
Total	**11,036**	**1,734**

Table 2. Detail of LSD scene classification dataset.

1.3.2. Experiments and Results.

We studied the existing classification method to learn the perfect scene identification. We alter the final classification layer of SOTA models to fit our objective [15]. Additionally, we leverage Imagenet's [6] pre-trained weights of these methods from the torch-vision [4] library to accelerate the learning process and achieve faster convergence. We trained existing methods for 25 epochs with their suggested hyperparameters. Fig. 3 details the training loss and validation accuracy of the SOTA classification method on LSD scene classification. VGG13 [12] illustrates the maximum validation accuracy among the SOTA classification methods. Based on the experimental results, we leverage the best weight of VGG13 for making our LSD scene classifier. Tab. 3 compares numerous deep methods in LSD scene classification.

Model	Accuracy (%)
Widerresnet [20]	95.13
VGG16 [12]	97.02
VGG16_bn [12]	95.67
Densenet161 [9]	96.92
VGG13 [12]	**97.57**

Table 3. Comparison between SOTA deep image classification methods on LSD scene classification.

Fig. 4 illustrates the sample images eliminated by our patch filtering strategy.

2. LSD vs. Existing Datasets

2.1. Training Scenes

Capturing images in real-world, uncontrolled environments offers significant advantages, particularly in replicating scenarios where images are taken with handheld cameras. However, such conditions often introduce imaging limitations, such as blurs, over-exposed highlights, and under-exposed regions, especially in the reference images. These imperfections can mislead deep learning models and result in unusable outputs.

To address this, we refined our training dataset by filtering out images with these limitations. This was achieved by analyzing the global intensity of reference images and

Figure 2. Sample images from scene classification dataset. (a) Imperfect scenes. (b) Perfect scenes.

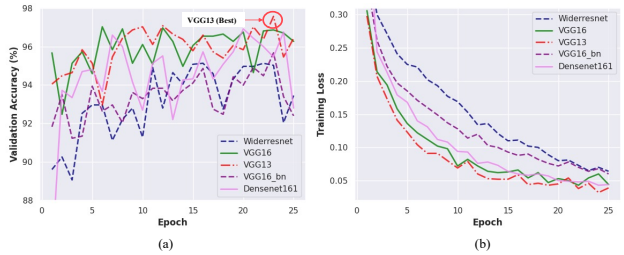


Figure 3. Learning scene classification with SOTA image classification methods. (a) Validation accuracy vs epoch. (b) Training loss vs epoch.

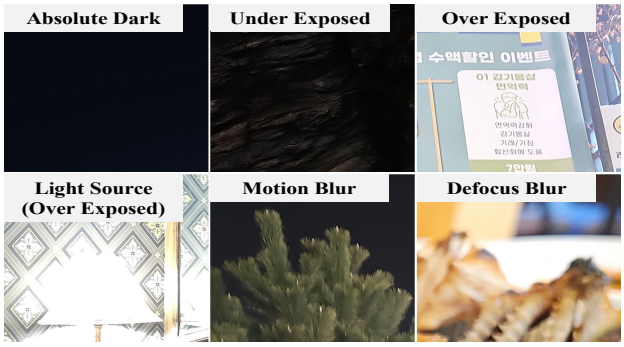


Figure 4. Sample images were eliminated through our patch filtering strategy.

leveraging our VGG-based scene classifier for robust filtering. Tab .4 provides a detailed breakdown of the device-wise distribution of the refined LSD training dataset, highlighting the diversity and quality ensured through this process. It is worth noting that some devices may use the same image sensor. However, these setups employ different focal lengths and image processing techniques, ensuring that the resulting samples remain diverse even when the same sensor is utilized.

2.2. Testing Scenes

One major limitation of the current SLLIE datasets is the absence of diverse testing scenes. To counter this, we curated the most extensive benchmarking test set, incorporating a variety of hardware, scenes, and sources. Tab. 5 details our LSD test set. For comprehensive evaluation, our test set includes paired scenes for quantitative and qualitative assessments and unpaired images for subjective evaluation.

2.2.1. Pair Images.

We collected 400 testing scenes for pair testing, including DLL and NLL pairs captured in different lighting conditions. Each of these categories also includes subsets for indoor and outdoor samples. For testing scenes, we selected the full scenes with minimal unexpected imaging instances.

2.2.2. Unpair Images.

Image enhancement invariably remains debatable, contingent upon personal preferences. Therefore, the proposed LSD offers 2,117 unique test cases for extensive subjective assessment. Our unpaired testing set encompasses data samples captured with DSLR cameras, smartphones, frames from 10 low-light videos, and social media images. We observed millions of images stored on social media, which can still benefit from SLLIE techniques, breathing new life into them.

2.3. Limitation of Existing SLLIE Datasets

Existing SLLIE datasets predominantly consist of scenes captured under well-lit conditions, with low-light inputs artificially generated using low ISO settings and short exposure times. As shown in Fig. 5, these samples are often taken in high-brightness outdoor environments, making the dataset semi-synthesized rather than representative of actual low-light scenarios. While useful for controlled experiments, this approach fails to authentically replicate the challenges faced in natural low-light environments, such as complex lighting distributions, varying noise levels, and diverse low-illumination conditions.

Moreover, these brighter scenes often contain naturally underexposed regions better suited for HDR mapping than developing and evaluating dedicated SLLIE methods. This mismatch can mislead model development, as the dataset does not fully capture the nuanced challenges of real low-light imaging. To advance SLLIE research, it is essential to focus on datasets that authentically represent diverse low-

Device	DLI	NLI	Combine	Raw Patches	Filtered Patches	Usable Patch (%)
Samsung Galaxy S10	84	67	151	5,185	3,847	74.19
Samsung Galaxy Z Fold 5	759	689	1448	56,577	50,416	89.11
Samsung Galaxy S 22 Ultra	1,270	1,420	2,690	85,471	75,514	88.35
Samsung Galaxy Z Flip 3	329	347	676	22,800	20,312	89.09
Xiaomi Redmi10C	166	161	327	9,163	8,439	92.10
Samsung Galaxy Z Flip5	367	366	733	29,657	21,167	71.37
Total	**2,975**	**3,050**	**6,025**	**208,853**	**179,695**	**86.04**

Table 4. Details of LSD training set (filtered). We filtered out around 15% of patch pairs to make our training set robust.

Type	Lighting Condition	Category	Scenes
	Low-light	DLI	100
Pair	Extreme	DLI	100
	Low-light	NLI	100
	Extreme	NLI	100
		Android	300
		iPhone	250
Unpair	Mix	DSLR	300
		Social Media	50
		Video Frames	1,217
	Total		**2,517**

Table 5. Detail of our LSD benchmarking set. We developed the largest and most diverse SLLIE benchmarking dataset throughout this study.

light conditions rather than relying on semi-synthetic approximations.

In contrast to the existing SLLIE datasets, we collected our datasets in actual low-light scenarios (0.1-200 lux). Fig. 6 illustrates the samples from the proposed LSD. We collected data over the years in different seasons, such as winter, autumn, and summer.

3. TFFormer

Due to page limits, our main article could not provide details on implementing TFFormer, LC loss, and complexity analysis. This section details the missing part to ensure the reproducibility of our proposed method.

3.1. Implementation Details

LC Extraction: Luminance-Chrominance (LC) extraction separates an image's intensity and color information. Luminance (L) is computed as L = 0.299R + 0.587G + 0.114B, representing brightness based on human visual perception. Where R, G, and B are color channels of an image. Chrominance (C) captures color details by subtracting luminance from the original Image: C = I - L. This decomposition is widely used in image processing for tasks like compression, enhancement, and object tracking, enabling better independent handling of brightness and color.

LC Encoding: We expanded the luminance-boosted image ($\mathbf{I_{BL}}$), luminance features ($\mathbf{F_L}$), chrominance-boosted image ($\mathbf{I_{BC}}$) and chrominance features ($\mathbf{F_C}$) into same feature dimension as $\mathbf{F_x} \in \mathbb{R}^{H\times W\times 40}$. Later, we feed our luminance and chrominance encoder with these expanded feature maps to obtain $\mathbf{I_{L_{enc}}}$ or $\mathbf{I_{C_{enc}}}$. Our luminance and chrominance encoders comprise the LCGAB, followed by two convolution operations with stride 2 to down-sample the LC attributes and boosted-image maps. We expanded the dimension of the feature channel by a factor of 2 in every recurring encoder block. Also, we reduced the spatial dimension by a factor of 2 while expanding the feature maps in both encoders to obtain $\mathbf{I_{BL_{down}}}$ and $\mathbf{I_{FL_{down}}}$.

$$\mathbf{I_{L_{enc}}} = LCGAB(\mathbf{I_{BL}}, \mathbf{F_L}) \tag{1}$$

$$\mathbf{I_{BL_{down}}} = Conv(\mathbf{I_{enc}}, s = 2) \tag{2}$$

$$\mathbf{I_{FL_{down}}} = Conv(\mathbf{F_L}, s = 2) \tag{3}$$

The Chrominance branch follows the same equation to obtain their respective images $\mathbf{I_{C_{enc}}}$, $\mathbf{I_{BC_{down}}}$ and $\mathbf{I_{FC_{down}}}$.

LC Decoding. In the decoder part, we combined the luminance and chrominance features from the encoder by adding them as a skip connection to guide the decoder's LCGAB. The LCGAB in the decoder is followed by a transpose convolution operation to decode the RGB images into their actual input dimension.

$$\mathbf{I_{L_{dec}}} = LCGAB(\mathbf{I_{BL_{down}}}, \mathbf{I_{FL_{down}}}) \tag{4}$$

$$\mathbf{I_{UP}} = ConvTransposed(\mathbf{I_{L_{dec}}}) \tag{5}$$

Chrominance features are also up-sampled similarly. We refined the final decoded features with our reconstructed luminance and chrominance attributes. Additionally, our TFFormer is designed as a fully convolutional network and can take images of any dimension as input. It can generate the

Figure 5. Sample images from existing SLLIE dataset. Many of the collected scenes of these datasets are captured in bright outdoor scenes and replicated in the low-light input by controlling camera parameters.

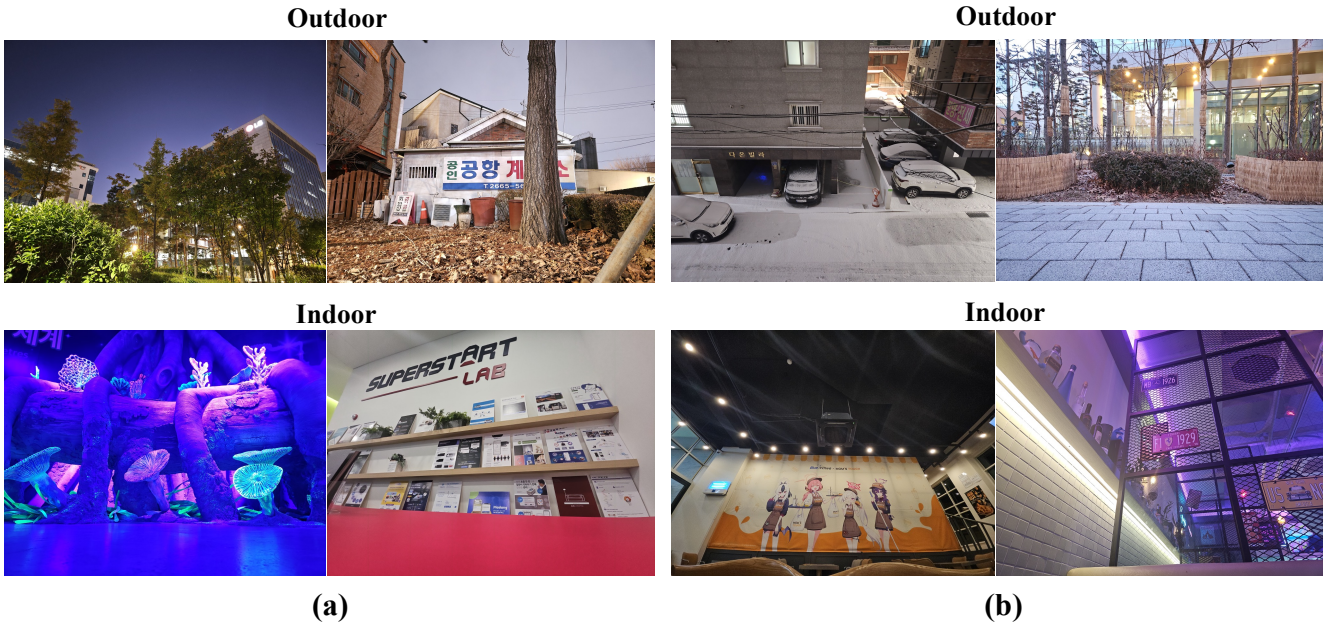


Figure 6. Sample images LSD. We collected all samples in 0.1-200 lux for over two years. (a) Samples from DLI scenes. (b) Samples from NLI scenes

output with the same dimension as the input without losing any spatial dimension.

3.2. Loss calculation

To perceive visually plausible images, we leverage an additional L1 loss along with the proposed LCCG. Thereby, we defined final LC-guided loss as:

$$\mathcal{L} = \lambda_R(\mathcal{L}_R + \mathcal{L}_{LC}) \tag{6}$$

where $\mathcal{L}_R$ represents the traditional L_1 loss, λ_R is a low-light regularization coefficient.

3.3. Training Details

We implemented our proposed TFFormer using PyTorch framework [13]. Our TFFormer optimized with an Adam

optimizer [10] with a hyperparameter of $\beta_1 = 0.9$, $\beta_2 = 0.99$. We set the initial learning rate = $1e-4$ and adjusted it after training 100,000 steps using `ReduceLROnPlateau` scheduler [3]. We trained our model for 250,000 steps, utilizing a batch size of 12 and image dimensions $128 \times 128 \times 3$. Additionally, we leverage a low-light regularization coefficient $\lambda_R = 0.2$ empharically to tune the proposed LC guidance. The training process spanned approximately 100 hours, executed on low-end hardware featuring an AMD Ryzen 3200G central processing unit (CPU) operating at 3.6 GHz, complemented by 32 GB of random-access memory, and an Nvidia GeForce GTX 3060 (12GB) graphical processing unit (GPU).

3.4. TFFormer on SLLIE Datasets

To assess the performance of the proposed LSD dataset against existing SLLIE datasets, we leverage our TFFormer model throughout this study. Prior to conducting cross-dataset comparisons, we validated the effectiveness of TFFormer on existing SLLIE datasets to ensure a fair and consistent evaluation. For this purpose, we compared TFFormer with state-of-the-art (SOTA) SLLIE methods that utilize Retinex theory or transformer-based architectures. Tab. 6 summarizes these comparisons, focusing on methods like Retinexformer, recognized as one of the best single-stage SLLIE models and the winner of the NCLLIE-CVPR24 challenge. Additionally, we included only models for which reported results were reproducible.

As shown in Tab. 6, TFFormer consistently outperformed all SOTA methods across LOL-V1 [19] and LOL-V2 (real) [24] datasets. Specifically, TFFormer achieved the highest PSNR (26.13 dB on LOL-V1 and 31.55 dB on LOL-V2), SSIM (0.8875 on LOL-V1 and 0.9147 on LOL-V2), and the lowest LPIPS (6.12 and 3.79, respectively). These results highlight the superior performance of TFFormer in enhancing real-world low-light images, surpassing strong baselines such as Retinexformer and MIRNet. This establishes TFFormer as a robust model for LSD-based evaluations and a strong contender for advancing SLLIE research on diverse datasets.

In addition to the quantitative evaluation, the qualitative results in Fig. 7 and Fig. 8 further demonstrate the practicability of TFFormer for generic SLLIE tasks. The proposed TFFormer consistently produces cleaner images with enhanced detail preservation and improved color accuracy, resembling the reference images. Additionally, the superior performance of TFFormer on LOL-V1 and LOL-V2 ensures its reliability and robustness for performing cross-dataset evaluations.

3.5. Inference Analysis

Tab. 7 demonstrates the inference speed and computational complexity of the proposed TFFormer on our hardware setup. Notably, the proposed TFFormer comprises only 5.87M trainable parameters—substantially fewer than those in well-known transformer models such as Uformer, MIRNet, and SNRNet. Moreover, as a single-stage network, TFFormer enables end-to-end optimization and efficient inference, enhancing performance while significantly reducing computational overhead. It takes just over 0.5 sec to enhance a large dimension low-light input on a mid-end GPU like GTX-3060. It is worth noting we evaluated our method with Float32 precision without performing any optimization. Therefore, the inference speed of TFFormer can be further improved by adopting model compression techniques such as quantization [23], pruning [11], etc., for future usage.

4. LSD-TFFormer In Real-world

This section illustrates more results of LSD-TFFormer on diverse scenarios.

4.1. TFFormer on LSD

Our TFFormer can perform evenly in numerous lighting conditions and scene types. We perform an extensive evaluation of TFFormer in eight subsets of LSD testing set. Fig. 9 and 10 illustrate the performance of our TFFormer in all subcategories of the proposed LSD pair benchmark dataset. Please note that the LSD benchmark dataset was collected using various camera sensors of different sizes. These sensors exhibit distinct responses to different lighting conditions. As a result, certain scenes within the LSD testing benchmark dataset may appear either darker or brighter than their actual lighting conditions.

4.2. TFFormer vs. Existing Method (More Results)

As we mentioned earlier, the challenges posed by over-enhancement, noise amplification, and color distortion in existing SLLIE methods. We present qualitative comparisons in Figures 11 and 12. These examples underscore TFFormer's ability to generalize to complex, real-world scenes across varying illumination levels and conditions.

In Figure 11, we visually compare TFFormer against three representative SLLIE architectures: Diff-Retinex [25], HVI [22], and RetinexFormer [2], on both standard low-light (DLI) and challenging noisy low-light (NLI) scenarios. Despite leveraging Retinex priors [2, 25] or biologically inspired feature fusion [22], these baselines frequently suffer from overexposure, structural artifacts, and color distortions—especially under NLI conditions. In contrast, TFFormer consistently delivers results with more natural tone rendering and finer structure preservation, validating the effectiveness of its luminance–chrominance (LC) decoupling and guided refinement mechanism.

Figure 12 extends the comparison to a broader set of 10 state-of-the-art methods, including classical

Method	LOL-V1			LOL-V2			Average		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Input	7.77	0.4186	37.08	9.72	0.4370	27.24	8.75	0.4278	32.16
Kind [27]	19.66	0.8519	10.42	18.06	0.8571	12.59	18.86	0.8545	11.51
MIRNet [26]	24.14	0.8675	7.45	28.10	0.9012	5.13	26.12	0.8844	6.29
SNRNet [21]	24.61	0.8660	6.84	21.48	0.8717	9.15	23.05	0.8689	7.99
Retinexformer [2]	25.15	0.8675	6.54	22.79	0.8637	8.20	23.97	0.8656	7.37
TFFormer	26.13	0.8875	6.12	31.55	0.9147	3.79	28.84	0.9011	4.95

Table 6. Quantitive comparison between existing SLLIE methods and TFFormer on LOL-V1 and LOL-V2. The proposed TFFormer outperforms the existing methods on existing benchmarking datasets as well. The best and the second-best results are highlighted in red and blue, respectively.

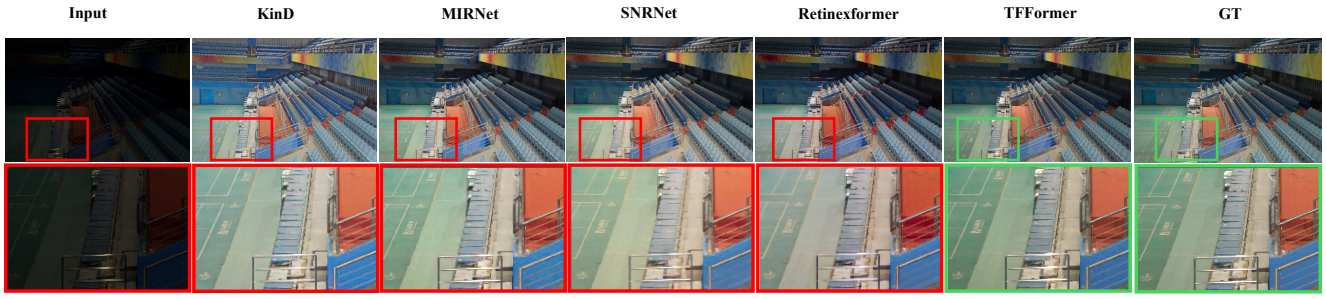


Figure 7. Qualatitive comparison between TFFormer and SOTA methods. The proposed method can produce cleaner and more color-accurate indoor scene images.

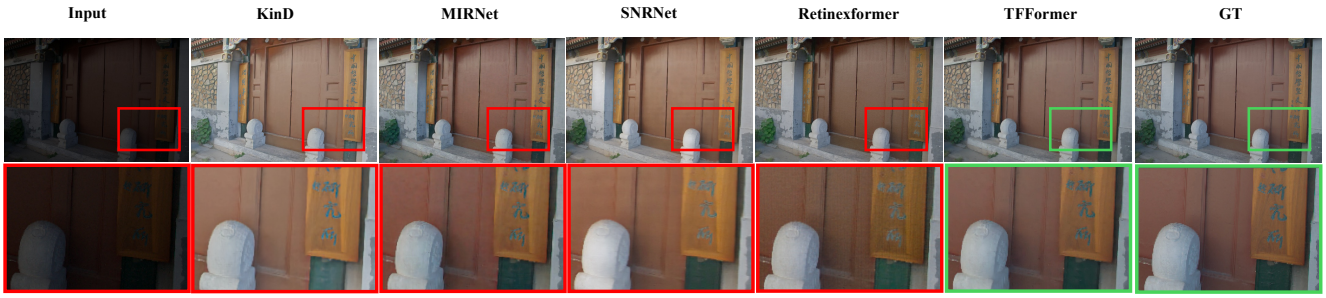


Figure 8. Qualatitive comparison between TFFormer and SOTA methods. The proposed method can produce cleaner and more color-accurate images on outdoor scenes.

Input Size	Param. (M)	Compl. (GMac)	Inf. Time (ms)
$256 \times 256 \times 3$		34.52	62.21
$512 \times 512 \times 3$	5.87	138.09	180.37
$1024 \times 1024 \times 3$		552.35	657.78

Table 7. Inference analysis of proposed TFFormer.

RetinexNet [19], Kind(+)[27, 28], DeepUPE[17], MIRNet [26], IAT [5], Uformer [18], SNRNet [21], GSAD [8], and LYT [1]. Across both DLI and NLI cases, TFFormer produces perceptually coherent, cleaner outputs with better structure and fewer color artifacts. Prior RGB-based transformer and CNN methods often suffer from over-smoothing (e.g., MIRNet, Uformer) or produce strong unnatural color casts (e.g., DeepUPE, SNRNet). Retinex-based designs, despite strong PSNR, tend to yield inconsistent tones under challenging lighting, likely due to excessive or misaligned enhancement. These observations, supported by both qualitative and quantitative evidence, underscore the robustness and generalization capacity of TFFormer in complex real-world scenes.

Together, these visual results reinforce the paper's central hypothesis: existing SLLIE models struggle to preserve structure and realism under challenging real-world scenarios, while TFFormer, trained on the proposed LSD dataset

Figure 9. TFFormer performance on DLI subsets, including indoor and outdoor scenes from low-light and extreme lighting conditions. In each pair, on the left is the input image, and on the right is the enhanced image by LSD-TFFormer.

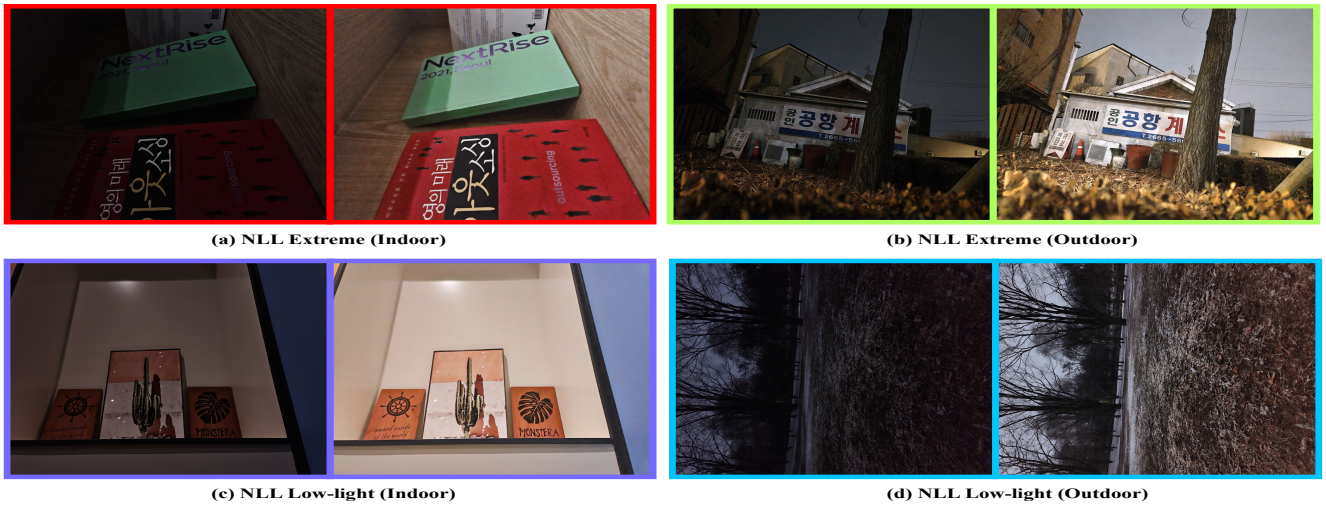


Figure 10. TFFormer performance on NLI subsets, including indoor and outdoor scenes from low-light and extreme lighting conditions. In each pair, on the left is the input image, and on the right is the enhanced image by LSD-TFFormer.

and guided by LC-aware modules, produces high-fidelity enhancements with superior generalization.

4.3. LSD on Deep SLLIE Methods

The primary motivation of this study is to advance SLLIE methods by providing a reliable dataset for training deep-learning models on real-world scenarios and benchmarking their performance across diverse conditions. To evaluate this, we trained the state-of-the-art Retinexformer [2] on existing LOL-V1, LOL-V2, and our proposed LSD dataset. Subsequently, we tested the model on complex scenes out of this dataset. Fig. 13 illustrates the performance of Retinexformer in enhancing these real-world complex scenes.

Notably, Retinexformer is one of the top-performing methods on LOL-V1 and LOL-V2. However, we observed that even its pretrained model (obtained directly from the official repository) often produces severe artifacts and over-exposed regions in challenging scenarios trained with existing SLLIE datasets. In contrast, training Retinexformer on the proposed LSD dataset enables it to generate more natural and visually appealing results. Furthermore, our TFFormer surpasses its counterparts by producing cleaner and more plausible images, thanks to its LC Encoding and LC Guidance mechanisms. These results underline the significant contributions of the LSD dataset and TFFormer in tackling real-world SLLIE challenges and their practicability in

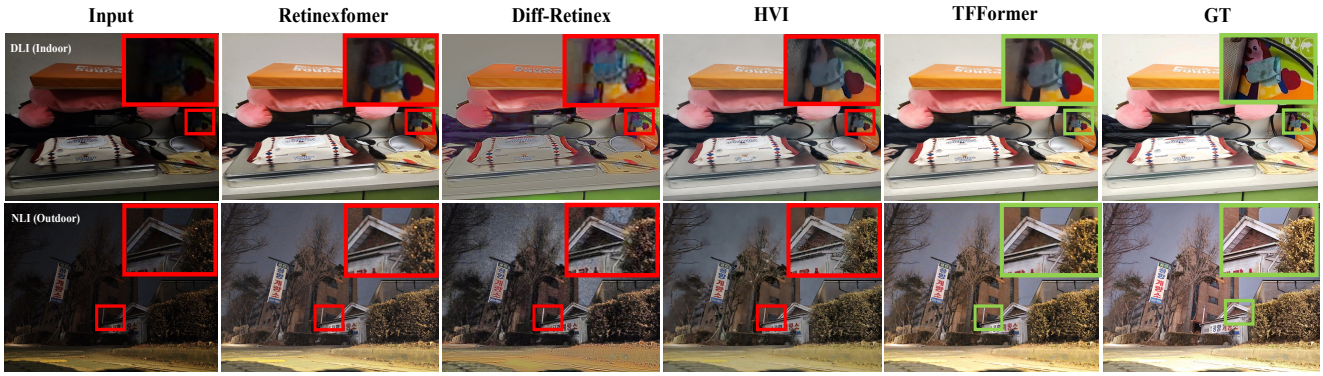


Figure 11. Real-world complex scene enhancement with low-light image enhancement methods. (a) input image capture with Samsung S22 Ultra. (b)-(c) Retinexformer[2] with existing datasets. (d) Retinexformer[2] with LSD. (e) Proposed (LSD-TFFormer)

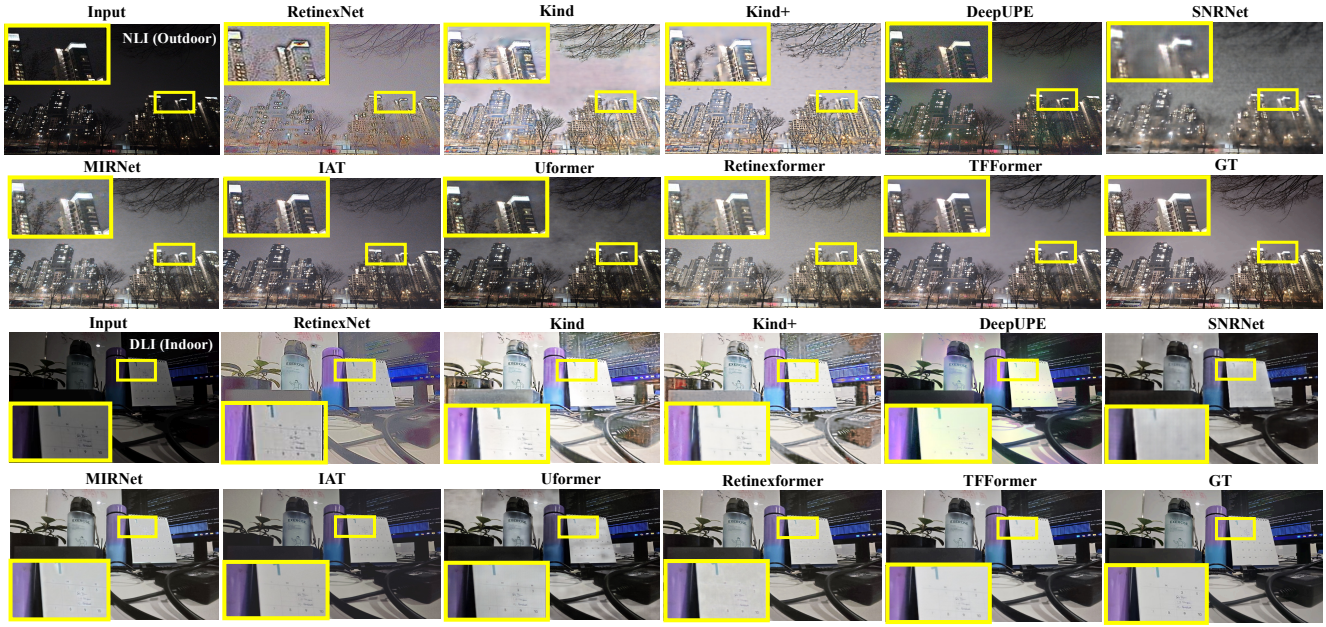


Figure 12. Qualitative comparison of existing SLLIE methods on the LSD dataset. The top scenes show performance on NLI, while the bottom examples illustrate performance on DLI scene types. TFFormer produces cleaner-plausible images, outperforming existing methods..

improving SLLIE methods.

4.4. LSD-TFFormer on Real-world Scenes

Visual Comparison. Fig. 14 illustrates more examples of real-world SLLIE obtained by the proposed LSD-TFFormer. Our method can handle diverse scenes obtained by numerous hardware and lighting conditions. Our LSD-TFFormer is also effective in enhancing images stored on social media (e.g., Facebook, Twitter, Instagram, etc.). These popular social platforms contain billions of user images captured with old camera hardware in low-light conditions. Our proposed method opens a new life to these images by enhancing and providing them with a modern touch.

User Study. We also performed a separate user study to verify the mass acceptance of our LSD-TFFormer in real-world scenarios. In this study, 35 participants aged 15 to 62 evaluated enhanced outputs across various lighting conditions and capture sources. Each participant was shown three randomly selected low-light versus enhanced image pairs from multiple SLLIE datasets and asked to choose the image they preferred, based solely on aesthetic appeal and without knowledge of the study's purpose.

As summarized in Table 8, over 82% of participants preferred the outputs generated by LSD-TFFormer. The preference was particularly strong in categories involving

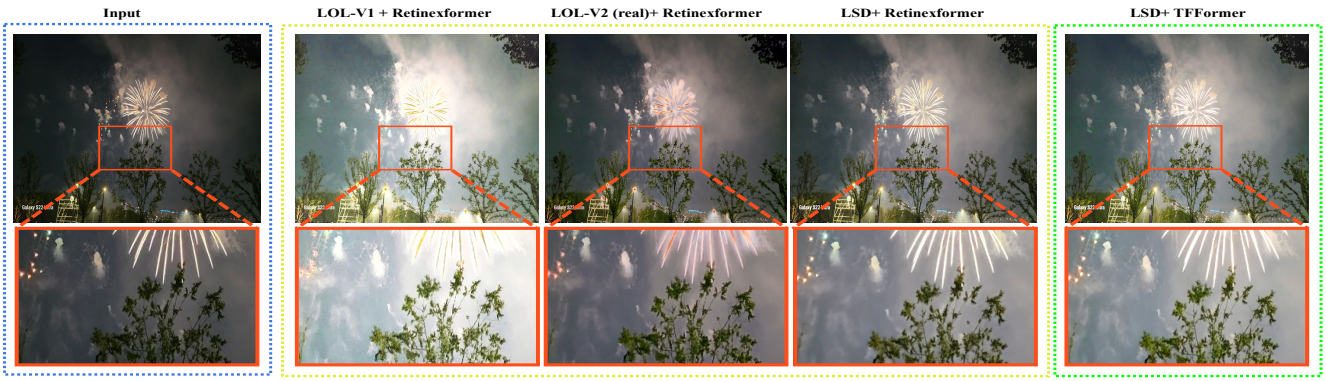


Figure 13. Real-world complex scene enhancement with low-light image enhancement methods. (a) input image capture with Samsung S22 Ultra. (b)-(c) Retinexformer[2] with existing datasets. (d) Retinexformer[2] with LSD. (e) Proposed (LSD-TFFormer)

older social media photos and noisy video frames, where traditional models tend to fail due to compression artifacts and low signal-to-noise ratios. This preference distribution supports the claim that our model generalizes beyond controlled lab settings and remains robust across highly diverse, real-world image sources.

Fig. 15 visually reinforces these findings by comparing LSD-trained models against those trained on existing datasets, such as LOL-V1/V2, LSRW, and NCLLIE. While models trained on traditional datasets often produce flat or over-smoothed results, LSD-trained TFFormer outputs exhibit more natural contrast, reduced artifacts, and better color fidelity. In many cases, the LSD-enhanced images preserved ambient lighting cues and semantic details that were lost in other reconstructions. This aligns with user preferences and demonstrates the practical utility of our dataset and model for consumer-grade image enhancement applications.

Device	Low-light ↑	LSD-TFFormer ↑
DSLR	0.0952	0.9048
Video (Frames)	0.3095	0.6905
iPhone	0.1905	0.8095
Android	0.2063	0.7937
Social Media	0.0714	0.9286
Average	0.1746	0.8254

Table 8. User study on LSD-TFFormer. In 82% of cases, user preferred LSD-TFFormer over low-light images.

4.5. Visual Improvement on Vision Task

Apart from aesthetical use cases, the proposed method can also accelerate everyday vision tasks. Fig. 16 Illustrates the performance gain achieved in two prominent vision tasks by incorporating our proposed method. Our approach significantly enhances the performance of widely utilized object detection (OD) [16] algorithms and facilitates improved matching of images [14] in low-light conditions. Notably, matching and registering low-light and well-lit images to make a paired dataset for SLLIE is extremely difficult. However, our method can support future studies by seamlessly matching collected low-light and well-lit images to enrich SLLIE research.

4.6. LSD-TFFormer vs Night mode

Night-mode photos are typically produced by combining multiple shots with different exposure settings. Comparing multi-shot approaches like night mode with single-shot methods such as LSD-TFFormer is unfair. However, to push the limit and study the feasibility of single-shot SLLIE in a broader aspect, we evaluated and compared our LSD-TFFormer with night mode photos. Thus, we mounter our smartphone on a fixed point and fixed its focal point. Later, we captured one scene in auto mode and another by enabling the dedicated night mode.

We enhanced the auto-mode photo (without night mode) with our LSD-TFFormer. Fig. 17 illustrates the comparison between LSD-TFFormer and night mode from Samsung Galaxy Z fold 5. We found that such dedicated night photography mode of smartphones is a trend to produce smooth images that lack salient details. On the other hand, our method can produce brighter images compared to the multi-shot processing while maintaining the salient details of the short-exposure images.

4.7. Failure Case

Despite promising results in numerous test cases, our TFFormer can produce visible noisy regions in NLL scenes captured under 1 lux. Fig. 18 illustrates such an example of a failure case. We planned to address such extremely challenging cases in future studies.

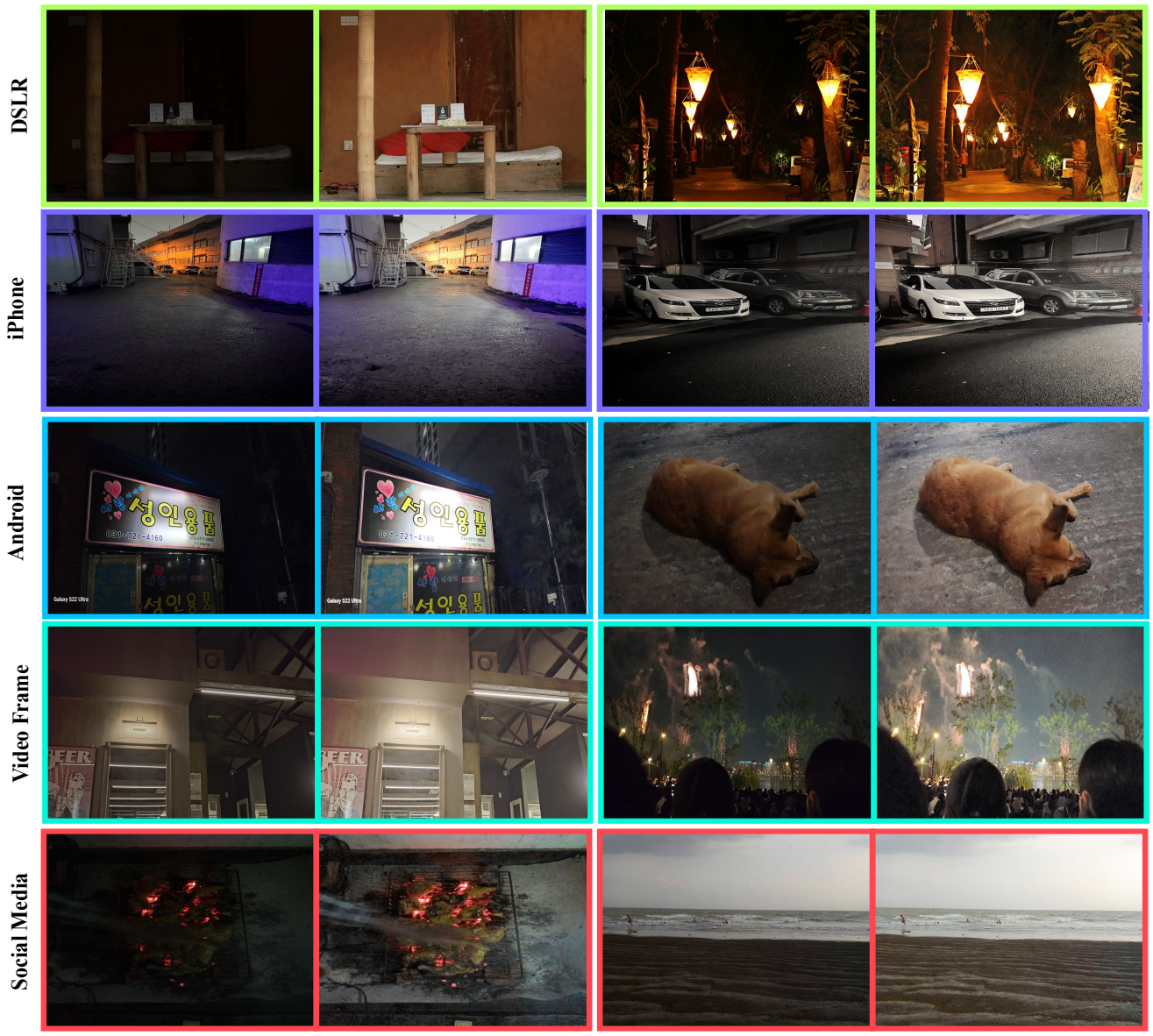


Figure 14. Few more examples of generic SLLIE obtained by LSD-TFFormer. In each pair, on the left is the input image, and on the right is the enhanced image by LSD-TFFormer.

References

[1] Alexandru Brateanu, Raul Balmez, Adrian Avram, Ciprian Orhei, and Cosmin Ancuti. Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement. *IEEE Signal Processing Letters*, 2025. 7

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. *arXiv preprint arXiv:2303.06705*, 2023. 6, 7, 8, 9, 10, 12

[3] PyTorch Contributors. Pytorch documentation - torch.optim.lr_scheduler.reducelronplateau, 2024. Accessed on 2024-03-02. 6

[4] PyTorch Contributors. Models and pre-trained weights, 2024. Accessed on 2024-03-10. 2

[5] 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. arxiv 2022. *arXiv preprint arXiv:2205.14871*, 238, 2022. 7

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[7] Google. Android camera api documentation, 2024. Accessed on 2024-03-02. 1

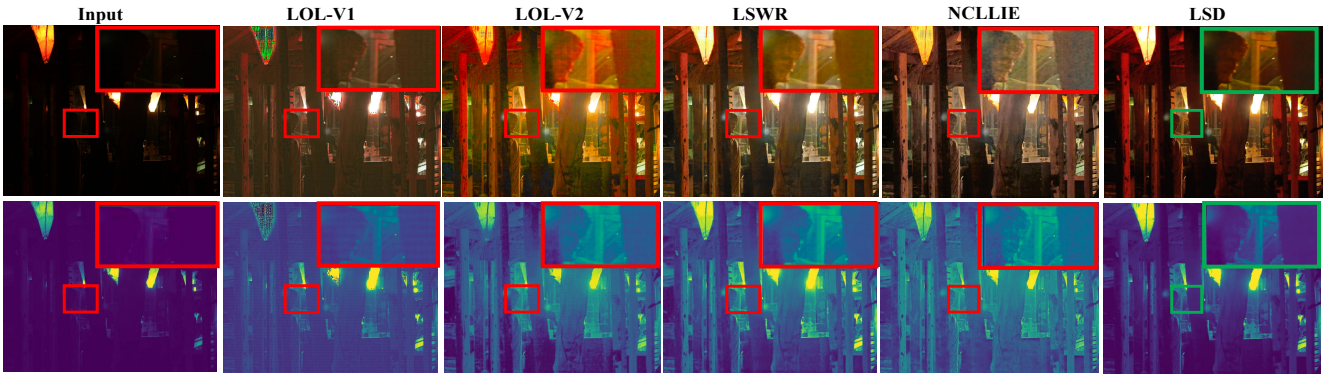


Figure 15. Real-world complex scene enhancement with low-light image enhancement methods. (a) input image capture with Samsung S22 Ultra. (b)-(c) Retinexformer[2] with existing datasets. (d) Retinexformer[2] with LSD. (e) Proposed (LSD-TFFormer)

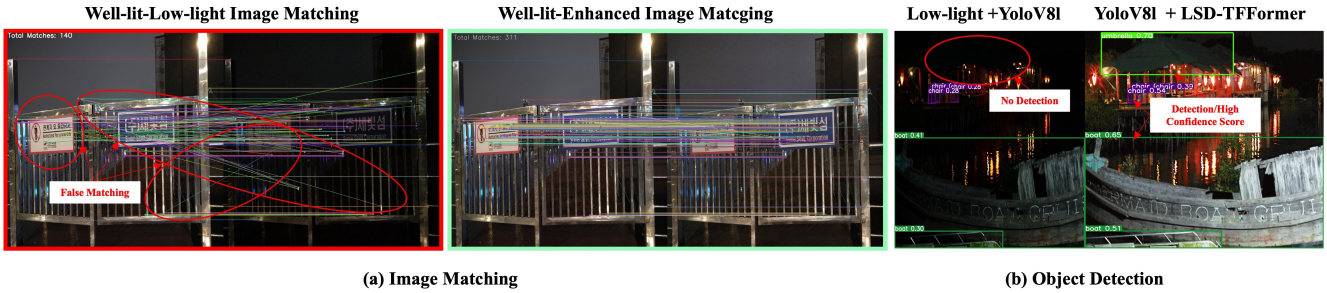


Figure 16. LSD-TFFormer usability in vision tasks. (a) Image matching. (b) Object Detection

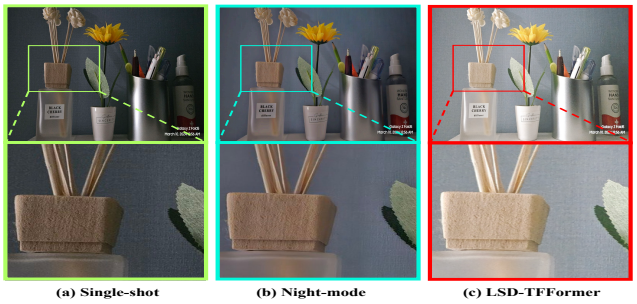


Figure 17. Comparison between LSD-TFFormer and night mode. Scenes capture under 20 lux with Samsung Galaxy Z fold 5. (a) Z Fold 5 without night mode. (b) Z Fold 5 with night mode. (c) Z Fold 5 without night mode scene enhanced by LSD-TFFormer.

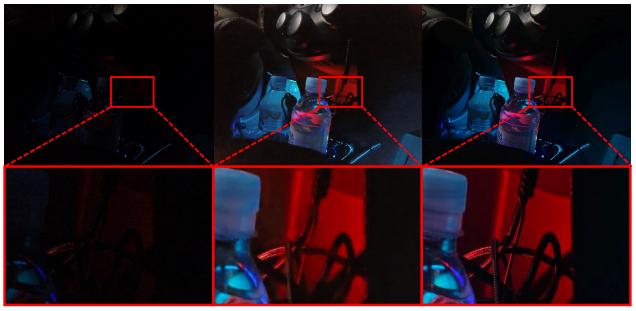


Figure 18. Example of failure case. For NLL scenes, TFFormer can illustrate visible noise in some extreme cases (e.g., under 1 lux). Left: NLL input (under 1 lux), right: LSD-TFFormer.

[8] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36:79734–79747, 2023. 7

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 2

[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[11] Lianqiang Li, Jie Zhu, and Ming-Ting Sun. Deep learning based method for pruning deep neural networks. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 312–317. IEEE, 2019. 6

[12] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian conference on pattern recognition (ACPR)*, pages 730–734. IEEE, 2015. 2

[13] Pytorch. PyTorch Framework code. https://pytorch.org/, 2016. Accessed: 2020-08-24. 5

[14] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary R. Bradski. Orb: An efficient alternative to sift or surf. In *2011 International Conference on Computer Vision (ICCV)*, pages 2564–2571, 2011. 10

[15] Vishnu Subramanian. *Deep Learning with PyTorch: A practical approach to building neural network models using PyTorch*. Packt Publishing Ltd, 2018. 2

[16] Ultralytics. Ultralytics github repository. https://github.com/ultralytics/ultralytics, 2024. Accessed on 2024-03-02. 10

[17] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 7

[18] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 7

[19] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6, 7

[20] Zifeng Wu, Chunhua Shen, and Anton Van Den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern recognition*, 90:119–133, 2019. 2

[21] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 7

[22] Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5678–5687, 2025. 6

[23] Jiwei Yang, Xu Shen, Jun Xing, Xinmei Tian, Houqiang Li, Bing Deng, Jianqiang Huang, and Xian-sheng Hua. Quantization networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7308–7316, 2019. 6

[24] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 6

[25] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023. 6

[26] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 7

[27] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 7

[28] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. 7