LIVENet: A novel network for real-world low-light image denoising and enhancement

Dhruv Makwana
Independent Researcher
dmakwana503@gmail.com

Gayatri Deshmukh
Independent Researcher
dgayatri9850@gmail.com

Onkar Susladkar
Independent Researcher
onkarsus13@gmail.com

Sparsh Mittal
IIT Roorkee
sparsh.mittal@mfs.iitr.ac.in

Sai Chandra Teja R
Green PMU Semi Pvt Ltd
saichandrateja@gmail.com

Abstract

Low-light image enhancement (LLIE) is the process of improving the quality of images taken in low-light conditions while striking a balance between enhancing image illumination and maintaining their natural appearance. This involves reducing noise, enhancing details, and correcting colors, all while avoiding artifacts such as halo effects or color distortions. We propose LIVENet, a novel deep neural network that jointly performs noise reduction on low-light images and enhances illumination and texture details. LIVENet has two stages: the image enhancement stage and the refinement stage. For the image enhancement stage, we propose a Latent Subspace Denoising Block (LSDB) that uses a low-rank representation of low-light features to suppress the noise and predict a noise-free grayscale image. We propose enhancing an RGB image by eliminating noise. This is done by converting it into YCbCr color space and replacing the noisy luminance (Y) channel with the predicted noise-free grayscale image. LIVENet also predicts the transmission map and atmospheric light in the image enhancement stage. LIVENet produces an enhanced image with rich color and illumination by feeding them to an atmospheric scattering model. In the refinement stage, the texture information from the grayscale image is incorporated into the improved image using a Spatial Feature Transform (SFT) layer. Experiments on different datasets demonstrate that LIVENet's enhanced images consistently outperform previous techniques across various quality metrics. The source code can be obtained from https://github.com/CandleLabAI/LiveNet.

Sparsh is the corresponding author. Dhruv, Gayatri and Onkar contributed to this paper while working as an intern at IIT Roorkee. Support for this work was provided by Science and Engineering Research Board, India under the project CRG/2022/003821.

Figure 1. Traditional techniques show a significant color difference from the ground truth. LIVENet restores realistic colors while reducing noise. The values in parenthesis are (PSNR/SSIM).

1. Introduction

Images taken in low-light conditions, such as at night, are becoming increasingly common and important due to evening events, night surveillance systems, and self-driving cars. However, processing such Low-Light Images (LLIs) presents crucial challenges. LLIs are often affected by excessive amounts of noise [40], which may be challenging to remove without reducing image quality or losing crucial features. Some Low-Light Image Enhancement (LLIE) techniques amplify this noise, leading to speckles and distortion in the final images. LLIs suffer from low contrast, uneven illumination, and artifacts such as color shifts. These factors hamper human perception and hinder downstream vision tasks such as face identification and semantic segmentation. These challenges have motivated researchers to propose several LLIE techniques.

The traditional LLIE techniques may be broadly divided into two categories: Histogram Equalization (HE)-based and retinex-based (refer to the supplementary material for more details). The HE-based techniques [1, 2, 12, 25, 33] change the image histogram to enhance the contrast. How-

ever, these methods lose image details due to the over-enhancement of some image areas. This can lead to a loss of texture and fine details. The retinex-based techniques [3,4,9,14,15,23,28,34] decompose the image into reflectance and illumination and then restore reflectance from illumination-adjusted input using ground-truth enhanced images. However, these techniques require extensive parameter tuning, leading to unrealistic colors and noise.

Recent years have seen the introduction of deep-learning-based LLIE techniques [13, 18, 19, 31, 40]. While these techniques outperform conventional techniques, they still suffer from crucial limitations: (i) Low-light images usually contain a greater proportion of noise compared to normal-light images. Previous LLIE techniques generally involve increasing the image's brightness and adjusting the contrast to make details more visible. However, these enhancements may amplify the noise in the image. (ii) In low-light conditions, the available light is limited, and the color information can be distorted or inaccurate. This can result in color shifts, where the colors in the image may appear different from their actual appearance in real life. (iii) LLIs have a limited dynamic range, meaning there is a slight difference between the brightest and darkest parts of the image. This limits the amount of texture information that can be recovered by LLIE techniques, particularly in very bright or very dark image regions. To address the challenges mentioned above, we propose LIVENet, a coarse-to-fine network for low-light image visibility enhancement. Our major contributions are:

1. We propose two novel denoising strategies. (i) We implement GSIP (Gray Scale Image Prediction) and LSDB (Latent Subspace Denoising Block) modules, which leverage the low-rank representation of the noisy dark image to predict the noise-free grayscale counterparts. (ii) We denoise RGB images by transforming the image into YCbCr color space. We replace the noisy luminance (Y) channel with a predicted noise-free grayscale image. To our knowledge, this is the first work to propose these strategies in LLIE operation.

2. To recover the illumination and realistic colors in the enhanced image, we introduce an Atmospheric Scattering Model (ASM). The ASM generates the enhanced image based on the transmission map and atmospheric light predicted in the image enhancement stage.

3. To overcome the limitation of blurriness and lack of texture information in enhanced image, we design the refinement stage, which leverages the "Spatial Feature Transform" (SFT) layers to adopt the features from the noise-free grayscale image into the enhanced image.

Experimental results on LOL-v1 and LOL-v2-real datasets show that LIVENet outperforms previous networks on all the metrics (PSNR, SSIM, MAE, and LPIPS). A sample result is shown in Figure 1. Results on additional datasets and ablation results provide further insights.

2. Related Work

We now summarize deep-learning-based LLIE techniques, categorized into supervised methods that use paired low/normal light data for training and unsupervised methods that use unpaired low-light data for training.

Supervised Methods: LLNet [18] proposes a deep neural network for image enhancement and denoising. It uses a stack sparse denoising autoencoder to correlate low-light images with the corresponding enhanced images. LLNet generates training paired data by adding synthetic noise and randomly performing Gamma correction. Nevertheless, the link between real-world lighting and noise is not adequately addressed, resulting in residual noise and over-smoothing issues. Bread [7] uses YCbCr color space and suppresses noise in the brightened luma (Y) component under the guidance of illumination. The enhanced luma then guides the chroma mapper to produce realistic colors. Wang et al. [37] utilized additional input of segmentation masks of interpreted LR and spatial feature transformation (SFT) layer to handle them. LLFlow [38] proposes a normalizing flow model to capture the relationship between low-light and normal images. Normalizing flow learns the distribution of normal images into a Gaussian distribution.

KinD [50], and KinD++ [49] utilize three sub-networks for illumination and reflection learning. They use illumination as the guidance for the restoration task and apply denoising on only the reflection component. DCC-Net [52] separates a color image into grayscale and color histogram. The former is used to recover textures and structures, while the latter is used for maintaining color consistency. R2RNet [10] uses a UNet-based network for denoising and an FFT-based approach to extract frequency features. By enhancing high-frequency signals, which represent content or noise, their technique improves image contrast and preserves fine details. LIME [8] works by estimating the illumination of every pixel and refining it according to a structure prior based on the retinex theory. LIME effectively restores brightness and contrast; however, it fails to reduce the noise and creates a color gap between ground truth and generated enhanced image.

SIRE [5] tries to estimate reflectance and illumination using a retinex-based weighted variational model. Its output images are rich in colors but are blurry and have smooth structures around borders due to a lack of denoising. SNR [42] uses long-range operations (transformer) in image regions with low signal-to-noise (SNR) ratio and short-range operations (convolution) in image regions with high SNR. The SNR values across the image guide attention in the transformer and the final fusion of features of the two branches. MSRNet [29] extends the single scale retinex by fusing the results of multiple Gaussian blur functions

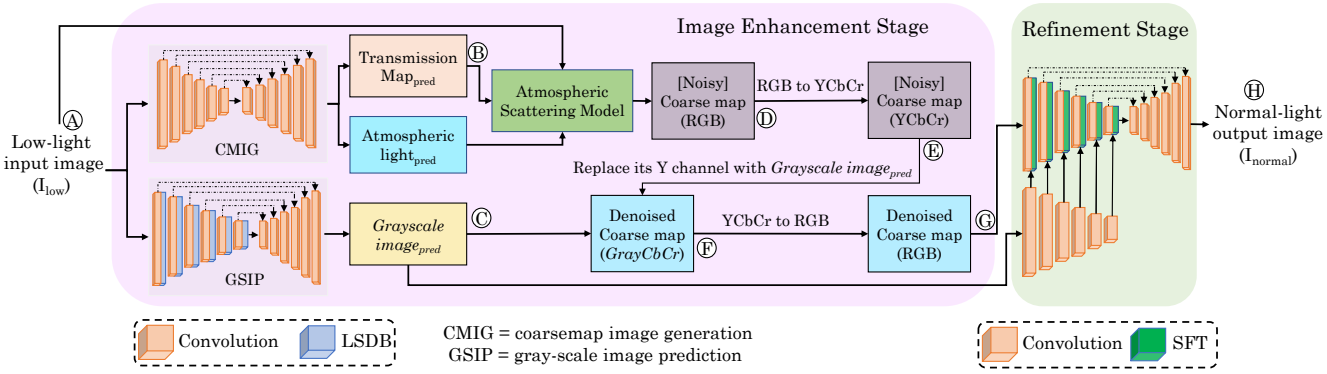


Figure 2. Architecture diagram of LIVENet

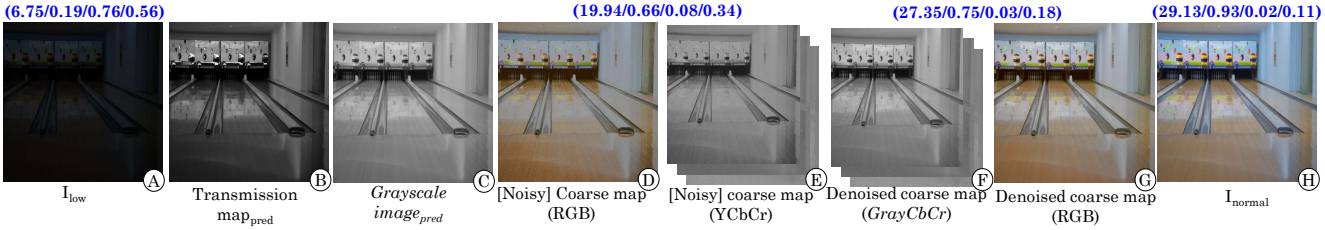


Figure 3. Outputs of various modules in LIVENet. The blue values are (PSNR/SSIM/MAE/LPIPS) metrics. A transmission map and a grayscale image are single-channel; hence, these metrics are not shown. The improvement in PSNR and SSIM from [Noisy] coarse map (19.94/0.66) to the Denoised coarse map (27.35/0.75) demonstrates the usefulness of the GSIP and LSDB modules. The improvement in SSIM and LPIPS from the denoised coarse map (0.75/0.18) to I_{normal} (0.93/0.11) shows the efficacy of the refinement stage and SFT layers.

with different variances. However, it suffers from over-enhancement. The GLAD technique [36] first predicts the global illumination and then reconstructs the detailed structure. Although it learns better texture information from the second stage, it suffers from color distortion due to global feature extraction in the first stage.

Unsupervised Methods: EnlightenGAN [13] uses a dual global-local discriminator to handle spatially-varying light conditions and perceptual loss to learn better features. It primarily focuses on the light factor and hence, cannot remedy other defects. DRBN [43] proposes a deep recursive band network, which extracts coarse-to-fine linear band representations for denoising and image enhancement. ZeroDCE [6] and ZeroDCE++ [17] use a deep network to estimate image-specific pixel-wise curves and uses them for dynamic range adjustment. SCI [21] proposes a lightweight self-calibrated module for illumination learning. It uses multiple cascaded blocks that share weights.

3. Proposed Method

Overall architecture: Figure 2 shows the architecture of LIVENet, and Figure 3 shows the output of each module. LIVENet is a coarse-to-fine network and is comprised of two stages: the image enhancement stage and the refinement stage. Conventional retinex-based and deep-learning-based methods concentrate on improving lighting while neglecting texture and color improvement. Hence, their output images suffer from a structure and color gap. To resolve these issues, our proposed image enhancement stage comprises two modules: GrayScale Image Prediction (GSIP) and Coarse Map Image Generation (CMIG). GSIP predicts a noise-free grayscale image, which is utilized in the image enhancement stage to remove noise from the coarse map and in the refinement stage to learn texture information from the enhanced image. CMIG predicts the transmission map and atmospheric light, which is used to generate a coarse map image. The coarse map image captures the coarse information such as lighting and colors; however, it has poor structure and some noise. The refinement stage refines this coarse map image to predict the final normal-light enhanced image. We now describe these modules.

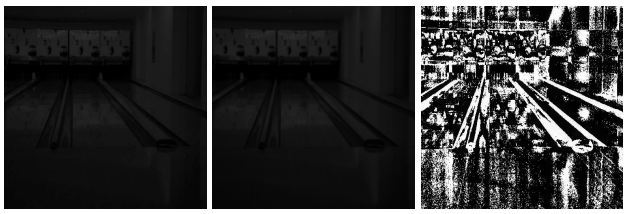
3.1. Stage one: grayscale image prediction (GSIP)

For a given noisy low-light image (Ⓐ in Figure 3), GSIP predicts the grayscale image (Ⓒ) of the normal-light version. This grayscale image is a good representation of the structural information in the image and is less affected by noise and other artifacts in the low-light image. Thus, by training GSIP, the network learns to focus on the crucial

structural features and suppress the noise.

GSIP architecture: Inspired by the efficacy of encoder-decoder networks for dehazing [30], matting [41], denoising [22], and inpainting [24], we construct GSIP with encoder-decoder pipeline (Figure 2). The encoder is created using six residual blocks with squeeze and excitation [11] module and a proposed "latent subspace denoising block" (LSDB). The residual blocks perform the feature extraction, and the LSDB suppresses the noise from extracted features. Also, to keep a flow of information throughout the network, the encoder shares skip connections with the decoder. The ground truth of the grayscale image is generated by converting the enhanced image into a grayscale image. GSIP uses $L1$ loss and $SSIM$ loss for grayscale image reconstruction.

LSDB block: As shown in Figure 4, an image can be decomposed into a low-rank matrix and a high-rank matrix using matrix factorization techniques. Here, the low-rank matrix captures the image's underlying structure and fine details, and the high-rank matrix contains the noise. To leverage this decomposition for denoising, we discard the high-rank matrix and use the low-rank matrix to estimate the original image. Since the low-rank matrix is a smooth approximation of the original image, it is less sensitive to noise and can be used for effective denoising.



(a) Low-light image (b) Low-rank component (c) High-rank component

Figure 4. Image decomposition into a low-rank component containing the texture information and a high-rank component containing the noise.

Previous methods [26, 27, 51] have used low-rank representation for image denoising; however, the optimization process of the low-rank image is time-consuming and non-differentiable. In this paper, we propose a differentiable LSDB inspired by [47] for LLI denoising. Figure 5 shows the architecture of LSDB, where F_1 and F_7 correspond to input LLI and low-rank component (respectively) of Figure 4. LSDB uses 3×3 convolution for feature extraction and 1×1 convolution for subspace decomposition. The feature map F_1 is first fed into 3×3 convolution to extract shallow features for subspace decomposition. These features are fed to two 1×1 convolution + GELU to obtain a low-rank representation of LLI. We use matrix multiplication between feature maps to obtain the latent space decomposition of LLI. We add shallow feature output and output of matrix multiplication (F_6) to obtain the denoised feature map F_7.

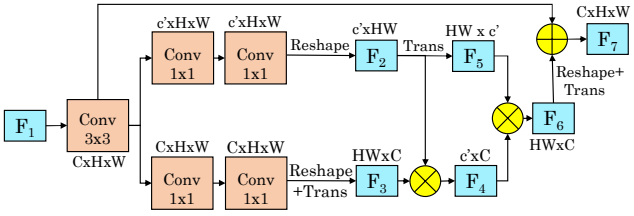


Figure 5. Proposed LSDB (inspired from [47])

3.2. Stage one: coarse map image generation (CMIG)

Existing LLIE techniques [5, 45, 46] face challenges in restoring fine detail and brightness due to the low SNR inherent in low-light images, as well as the loss of image contrast and color saturation. To mitigate this challenge, we propose a CMIG module (Figure 2), which aims to improve illumination and reduce color distortion in a low-light image I_{low} (Ⓐ in Figure 3). CMIG uses an encoder-decoder network where the encoder consists of six residual convolution blocks with squeeze and excitation for feature extraction and two parallel branches in the decoder for prediction of transmission map (t) and atmospheric light ($\mathcal{A}$). CMIG uses $L1$ loss to predict the transmission map and atmospheric light.

Atmospheric scattering model (ASM): The transmission map (Ⓑ) represents the ratio of original scene radiance to the incident light that has not been scattered by the atmosphere. The atmospheric light is the global illumination that has been scattered by the atmosphere and reaches the camera, and it is a global constant for a given image. These two components are essential for accurately estimating the scene's radiance and generating an image with improved brightness and color balance. Atmospheric scattering occurs when light interacts with particles in the atmosphere, leading to changes in the appearance of images taken under certain conditions, such as low-light environments. ASM is a mathematical algorithm that simulates how light behaves as it interacts with the atmosphere, accounting for scattering, absorption, and reflection factors. We use ASM to obtain an image that preserves important features and has enhanced illumination and visibility. ASM uses the equation 1. Here $J(x)$ represents the enhanced image (Ⓓ). The detailed steps for generating the coarse map are provided in the supplementary material.

$$J(x) = \frac{I_{low}(x) - \mathcal{A}}{t(x)} + \mathcal{A} \tag{1}$$

Removing noise from ASM output: A downside of ASM is that it can increase the noise while enhancing illumination. This unequal noise distribution produces various problems in the LLIE process, including color bleeding. Excessive blue channel denoising results in color bleeding and aberrations. Thus, the noise in the coarse map image

can propagate and get amplified during the subsequent image enhancement process, leading to further degradation of the image quality.

The ASM output is a coarse map in RGB space. The noise is typically distributed equally across all three color channels in the RGB color space. So, it is difficult to remove noise from the RGB color space. It is well known that different color spaces show different noise distributions in constituent channels. Hence, we propose addressing the above issue by changing the color space from RGB to one that separates luminance and chrominance, such as CIELAB, CIELUV, and YCbCr. Of these, we prefer YCbCr since it is a linear color space. The YCbCr color space divides the spectrum into a luminance channel (Y) and two chrominance channels (Cr and Cb). We note that (1) the noise distribution is contained in the luminance (Y) channel only (2) the human eye is more sensitive to changes in luminance than color changes, so noise in the luminance channel has a higher effect on the perceived quality of the image. Therefore, we focus on reducing noise in the luminance channel while preserving the color information in the chrominance channels. To realize this, we first convert the ASM-generated coarse map image from RGB color space to YCbCr color space (Ⓔ). Then, our key idea is to replace its Y channel with the noise-free grayscale image predicted by the GSIP module (Ⓒ). We call the resultant image a denoised coarse map in *GrayCbCr* space (Ⓕ). We convert this image back to RGB color space (Ⓖ) and feed it to the refinement stage.

3.3. Stage two: Refinement stage

Motivation: By using a combination of a dark channel prior and a bright channel prior to create the transmission map, stage one can understand the amount of enhancement required by different image regions. The denoised coarse map image (Ⓖ) is free from color distortion and has fine illumination. Nevertheless, it suffers from two limitations: (1) It fails to recover texture information due to LLI's limited dynamic range of pixel values. (2) Using residual convolution to extract features from the dark image results in blurriness and smoothness while denoising. To mitigate these limitations, we propose the refinement stage, which seeks to learn the missing structure and improve color and lighting in the coarse map image.

Architecture: We feed the coarse map image (Ⓖ) and the noise-free grayscale image (Ⓒ) generated by GSIP into the refinement stage. The refinement stage consists of two encoders: a fully convolutional encoder for learning structural information from the grayscale image and an encoder with SFT layers [37] for enhancing the illumination of the coarse map image. The lateral connections between the two encoders allow the network to incorporate both global and local information from the grayscale image while preserving the detailed features in the denoised coarse map image. The refinement stage uses $L1$ loss and $SSIM$ loss for normal-light image reconstruction.

SFT layer: The SFT layer was originally proposed for image super-resolution tasks. It was used for learning texture information with the help of segmentation probability maps as a prior condition. We adopt the SFT layer for the LLIE task. The SFT layer adjusts the distribution of the denoised coarse map based on features extracted from the grayscale image.

Based on the prior condition Ψ, the SFT layer seeks to train a mapping function $\mathcal{M}$ that outputs modulation parameters (γ, β). Here, prior condition Ψ can be represented as probability maps from grayscale image features, and the mapping function maps the coarse map to an enhanced image, given prior condition Ψ. To perform this mapping, SFT performs an affine transformation on the coarse map by scaling and shifting the input features by γ and β, respectively. γ and β parameters assist in adjusting structural features in the enhanced image. SFT can be stated mathematically as equation 2. Here, $\odot$ denotes element-wise multiplication. The SFT layer preserves the spatial dimensions; hence, it can perform both feature-wise and spatial-wise transformations. This enables it to recover structural information.

$$SFT(F|\gamma, \beta) = \gamma \odot F + \beta \tag{2}$$

4. Experimental Results and Analysis

For details on the experimental platform and metrics, refer to supplementary. To ensure uniformity, we have conducted experiments with all previous techniques ourselves. We use the models trained on the same dataset for LOL-v1 [40] and LOL-v2-real [44] datasets since the ground-truth images are available. For showing results on the unpaired datasets, viz., DICM [16], LIME [8], MEF [20] NPE [35], and VV [32], we have used the model trained on the LOL-v1 dataset.

Quantitative Results: Table 1 summarizes the quantitative results. We omit a few networks, viz., LIME [8], SIRE [5], MSRNet [29], GLAD [36], EnlightenGAN [13], and LLNet [18] from the comparison because the recent networks that we compare against, viz., DCCNet, LLFlow and SNR, have already outperformed them.

Our proposed LIVENet achieves the highest PSNR and the lowest MAE on both datasets, indicating that its results are closest to the ground truth. LIVENet can recover structures (shown by higher SSIM [39]), suppress noise, and retrieve color (shown by higher PSNR). Its output images are more realistic (shown by lower MAE) and best aligned with the human vision (shown by low LPIPS [48]).

LPIPS is a human perception metric. LIVENet significantly improves the LPIPS metric over most previous

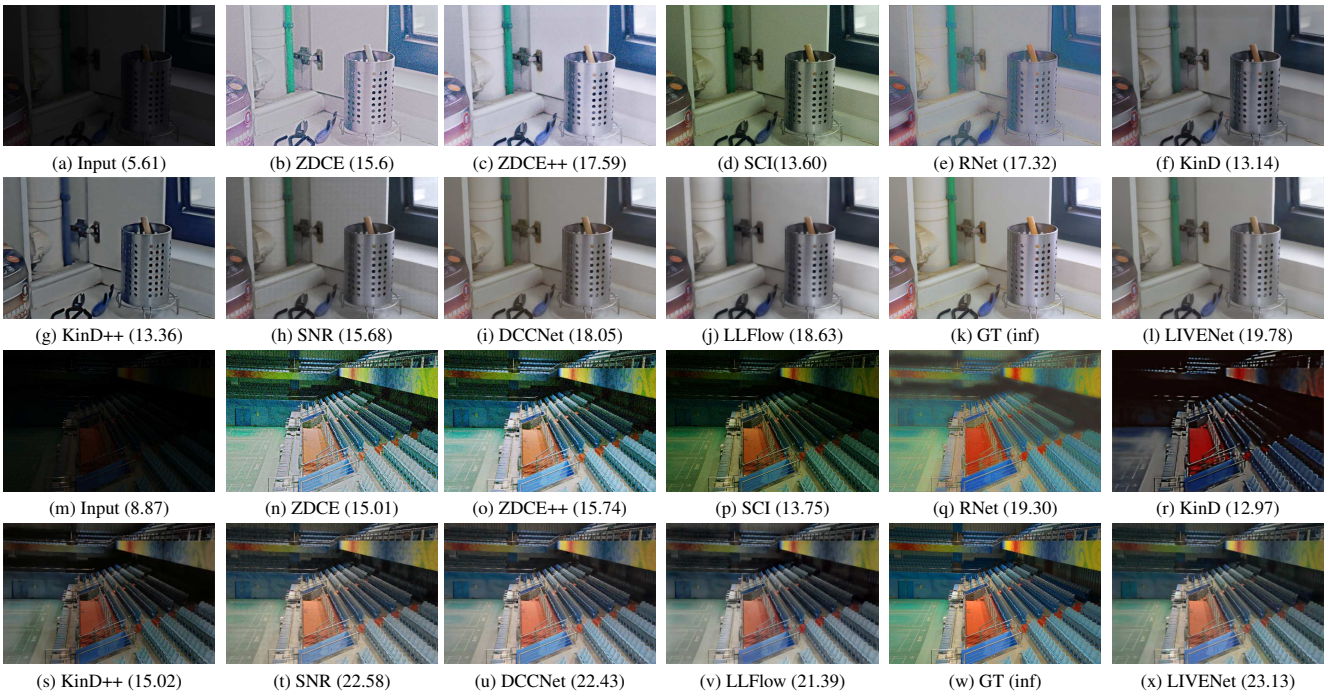


Figure 6. Results on LOL-v1 dataset. The number in parenthesis shows image PSNR. GT=ground-truth

Table 1. Results on the LOL-v1 and LOL-v2-real dataset. The best and second-best results are shown in bold and underlined, respectively.

Model	LOL-v1				LOL-v2-real			
	PSNR	SSIM	MAE	LPIPS	PSNR	SSIM	MAE	LPIPS
ZeroDCE (CVPR-20) [6]	14.36	0.49	0.18	0.45	12.37	0.44	0.24	0.50
ZeroDCE++ (TPAMI-21) [17]	15.34	0.57	0.23	0.33	12.66	0.45	0.24	0.45
SCI (CVPR-22) [21]	16.44	0.54	0.19	0.39	16.7	0.54	0.27	0.34
RetinexNet (BMVC-18) [40]	17.23	0.48	0.14	0.46	15.98	0.72	0.17	0.41
KinD (MM-19) [50]	19.36	0.84	0.11	0.18	18.11	0.81	0.13	0.27
KinD++ (IJCV-21) [49]	21.32	0.85	0.08	0.23	18.85	0.81	0.12	0.24
SNR (CVPR-22) [42]	22.75	0.80	0.09	0.19	20.15	0.83	0.12	0.17
DCCNet (CVPR-22) [52]	22.97	0.85	0.08	**0.14**	20.47	0.75	0.13	**0.11**
LLFlow (AAAI-22) [38]	23.84	0.86	**0.07**	**0.14**	22.58	0.82	0.08	0.16
LIVENet (Ours)	**24.68**	**0.87**	**0.07**	**0.14**	**23.19**	**0.84**	**0.07**	0.12

methods, suggesting that LIVENet is best aligned with human vision. The unsupervised methods (ZeroDCE, ZeroDCE++, and SCI) are inferior to the supervised approaches (KinD, KinD++, RetinexNet, SNR, DCCNet, LLFlow, and LIVENet) in maintaining the quality of enhanced images. The unsupervised methods cannot learn noise suppression from reference images, leading to poor PSNR. ZeroDCE was designed for low-light images with poor contrast and saturation but not for noisy LLIs. Hence, it produces inferior results (high LPIPS) on LOL-v1 and LOL-v2-real datasets, which have a noticeable amount of noise. SCI uses a self-calibrated module for illumination learning, which enhances exposure stability but disregards noise suppression.

RetinexNet employs a denoising tool (BM3D) to clean the reflectance component. After the illumination factor is extracted, the dark regions of the image contain much higher noise than the brighter regions. In such a case, using a denoiser that performs uniform denoising on all the image areas is no longer appropriate. Due to this, RetinexNet produces an image with higher noise (low PSNR of 17.23) and unrealistic colors (high LPIPS of 0.46).

Table 2 shows the naturalness image quality evaluator (NIQE) score for unpaired datasets. LIVENet achieves the best NIQE scores on all the datasets except the NPE dataset, where it is the second best. Table 2 also shows a detailed parameter comparison between LIVENet and the baseline model. LIVENet has 16.66M params.

Table 2. NIQE metric (lower is better) results on the five unpaired datasets: DICM, LIME, MEF, NPE, and VV.

Method	DICM	LIME	MEF	NPE	VV	Parameters (M)
ZeroDCE	4.7602	5.4574	4.7800	4.9092	4.8553	0.070
ZeroDCE++	4.4424	5.3330	4.7028	4.9658	5.0044	0.010
SCI	4.4700	5.1141	4.5076	4.9733	5.0498	**0.005**
RetinexNet	4.8895	5.9931	4.9269	5.5785	5.6219	0.450
KinD	4.0515	4.9392	4.9287	4.7428	4.5471	8.110
kinD++	4.4110	4.8626	4.2931	4.3225	4.8295	8.280
SNR	3.6947	4.5159	4.2671	4.6744	5.9617	40.01
DCCNet	3.6815	4.5188	4.0105	**3.6023**	4.6076	13.18
LLFlow	3.8409	4.6777	4.0863	4.3470	4.0940	1.700
Ours	**3.6024**	**4.2188**	**3.8937**	3.8931	**3.4433**	16.66

Qualitative results on LOL-v1 dataset (Figure 6): ZeroDCE, ZeroDCE++, and RetinexNet have been abbreviated as ZDCE, ZDCE++, and RNet, respectively. LIVENet significantly enhances the results by suppressing huge noise

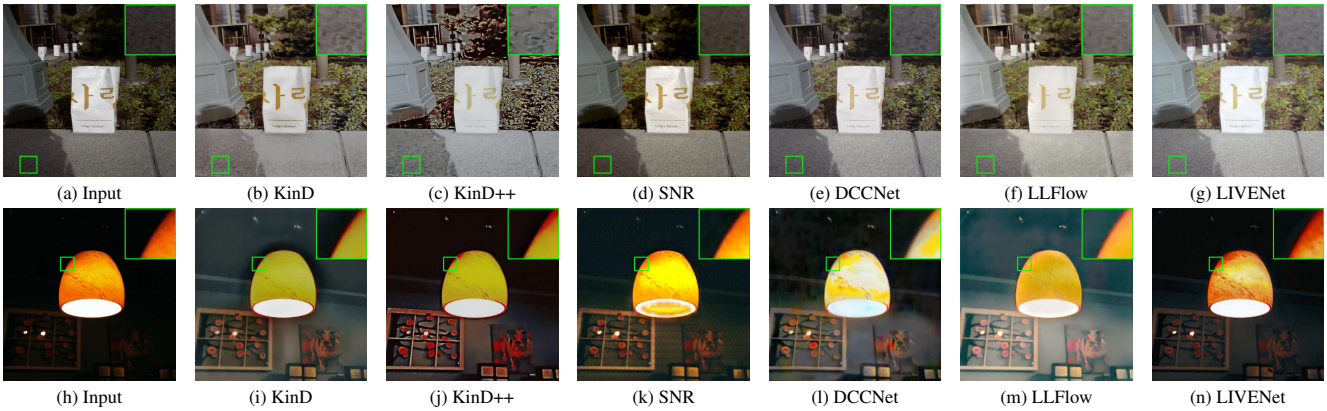


Figure 7. Qualitative results on the DICM dataset (Row 1) and the LIME dataset (Row 2)

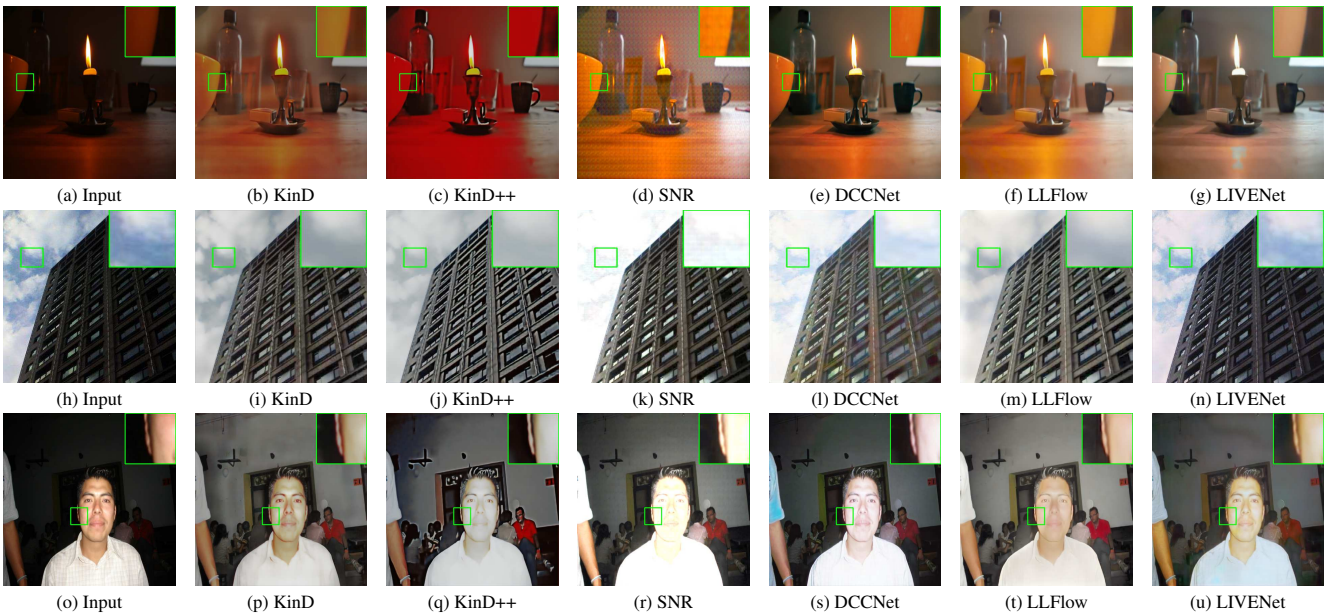


Figure 8. Qualitative results on the MEF (row 1), NPE (row 2), and VV (row 3) datasets

and revealing more realistic colors. ZeroDCE and ZeroDCE++ have increased the brightness, as evident from the first (b, c) and third row(n, o). SCI applies a zero-shot learning strategy without a reference image. This strategy cannot produce satisfying results in challenging cases and suffers from low contrast, as shown in the first row (d) and third row (p). RetinexNet improves illumination but fails to produce natural colors due to the smoothness gained from the denoiser. Similarly, KinD and KinD++ images suffer from darkness and do not adjust the illumination correctly.

Qualitative results on DICM and LIME datasets (Figure 7): For comparison, we choose the best five methods, viz., KinD, KinD++, SNR, DCCNet, and LLFlow. We have used the network trained with the LOL-v1 dataset to obtain these results. The green rectangle magnifies the small area of the image to highlight finer details. For the image in the first row (DICM dataset), KinD and KinD++ perform excessive enhancement, leading to the loss of some of the original colors. The region inside the green rectangle shows that LIVENet (by virtue of using LSDB) has significantly reduced the noise in the dark shadow image (first row). DCCNet decouples the RGB image into a grayscale image and a color histogram. It improves illumination in the bright region but fails to eliminate noise in the dark region, as shown in the person's shadow. The SNR technique applies the "signal-to-noise ratio" prior with a transformer. This transformer distributes its focus evenly across all patches, regardless of illumination. This effect can be seen from the lamp region (second row), which is overly saturated with surrounding pixels.

Our proposed CMIG uses a bright channel to create the transmission map ground truth, and it also uses the dark channel prior as a supplement to correct the over-enhanced transmission map for the bright object area. Its usefulness can be observed in the image in the second row, where LIVENet properly enhances the image without over-enhancing the source of light (lamp) or the surrounding area of the light.

Qualitative results on MEF, NPE, and VV datasets (Figure 8): Notice that the images produced by KinD and KinD++ show the wrong color information. For the candle image in the first row, SNR and DCCNet-generated images show the color of the left bowl as orange due to the presence of the candle. By contrast, LIVENet produces an image with a more realistic bowl color. LIVENet can effectively improve the quality of indoor objects and landscapes (rows 1 and 2), leading to an image with more realistic face color and reduced noise (row 3). Overall, LIVENet generalizes well on a wide variety of low-light images.

A major challenge in LLIE is reflection, which can lead to overexposure and saturation in the image. When a bright light source is present in the scene, it can reflect off surfaces such as windows, as shown in the landscape image (second row), creating an overexposed area. Due to this, DCCNet and SNR create artifacts around the reflection area while enhancing the overexposed region, as seen from the green color in the window. LIVENet handles overexposed regions much better and creates the least amount of artifacts.

5. Ablation study

Table 3 summarizes the ablation results.

1. Architecture-related ablation: On removing the LSDB module, the output images become more blurry and noisy. For obtaining results "w/o SFT", we replace the SFT block with a normal residual connection and feed the concatenated grayscale and coarse map as input to the network. The decrease in SSIM indicates a lack of texture information in the final result. Removal of the GSIP module leads to increased noise and erroneous colors. In the "w/o GSIP module" version, we generate the coarse map without replacing the Y color channel in the image enhancement stage and refine the coarse map without the SFT layer in the refinement stage. There is a higher amount of noise on not replacing the Y color channel of the YCbCr coarse map with a noise-free grayscale image. In summary, out of all the constituents in the image enhancement stage, the GSIP module has the highest contribution towards final quality. On skipping the refinement module, the output lacks refined structural information and realistic illumination.

2. Loss-related ablation: An entry such as "with L1 loss" indicates that only L1 loss, and no other loss, was used. On removing the SSIM loss function, the SSIM metric degrades, and on removing the L1 loss, the PSNR metric degrades. Thus, the SSIM loss function helps predict the grayscale image and refined outcomes; and L1 loss helps remove noise. The perceptual loss helps in learning higher-level features but fails to remove noise from the image. Even using L1 + perceptual loss, the results are inferior to those with L1 +SSIM loss. Charbonnier loss is less sensitive to outliers and provides a more robust optimization. However, it penalizes large and small errors equally, leading to a smoothing effect that blurs the sharp edges or details.

Table 3. Ablation results on LOL-v1 dataset

Approach	PSNR	SSIM	MAE	LPIPS
Full Network	24.68	0.87	0.07	0.14
1. Architecture related ablation				
W/o LSDB	22.65	0.84	0.09	0.16
W/o SFT	22.91	0.85	0.08	0.15
W/o GSIP module	22.01	0.84	0.08	0.22
W/o replacing Y channel	22.88	0.85	0.09	0.15
W/o Refinement Module	21.78	0.74	0.09	0.26
2. Loss related ablation				
With L1 loss	22.31	0.81	0.09	0.24
With SSIM loss	23.47	0.87	0.07	0.15
With perceptual loss	23.09	0.83	0.08	0.16
With L1 + perceptual loss	23.36	0.86	0.08	0.14
With charbonnier loss	22.32	0.81	0.09	0.24
3. Training methodolody related ablation				
WGAN-GP	23.44	0.85	0.08	0.16
Pix2Pix	23.37	0.86	0.08	0.14
4. Image size related ablation				
256×256	24.86	0.87	0.05	0.15
128×128	25.2	0.89	0.05	0.11

3. Training methodology related ablation: Utilizing WGAN+GP training results in higher quality outcomes compared to Pix2Pix, demonstrating that the gradient penalty improves the network's convergence. The discriminative loss functions used in GANs are meant to improve perceptual quality, which is not the best measure for enhancing low-light images. LLIE usually requires specialized metrics, such as the measures of image illumination, noise, and texture, to achieve optimal results. Nevertheless, both these methodologies are inferior to our non-GAN-based training approaches since LIVENet does not use any GAN component.

4. Image size-related ablation: Our technique remains effective with different image sizes.

6. Conclusion

We propose LIVENet for low-light image enhancement. We introduce techniques for retaining texture and color. Our novel idea is to remove noise from the coarse map by replacing the Y component of the YCbCr color space with a grayscale image. We propose a differentiable latent subspace denoising block by representing low-light features in low-rank structures. The refinement stage refines the global color and local texture information. Detailed experiments validate the effectiveness of our technique. Our future work involves extending LIVENet to handle low-light videos by considering both temporal and spatial information.

References

[1] M. Abdullah-Al-Wadud, Md. Hasanul Kabir, M. Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2):593–600, 2007. 1

[2] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 1

[3] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing*, 24(12):4965–4977, 2015. 2

[4] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 2

[5] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 2, 4, 5

[6] 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*, pages 1780–1789, 2020. 3, 6

[7] Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, pages 1–19, 2022. 2

[8] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 2, 5

[9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 2

[10] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90:103712, 2023. 2

[11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 4

[12] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 1

[13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021. 2, 3, 5

[14] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. 2

[15] D.J. Jobson, Z. Rahman, and G.A. Woodell. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, 6(3):451–462, 1997. 2

[16] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *2012 19th IEEE international conference on image processing*, pages 965–968. IEEE, 2012. 5

[17] 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*, 44(8):4225–4238, 2021. 3, 6

[18] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2, 5

[19] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018. 2

[20] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 5

[21] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 3, 6

[22] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems*, 29, 2016. 4

[23] Dr Parihar, Anil Singh, and Kavinder Singh. Illumination estimation for nature preserving low-light image enhancement. 2020. 2

[24] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 4

[25] S.M. Pizer, R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, and K.E. Muller. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pages 337–345, 1990. 1

[26] Jiahuan Ren, Zhao Zhang, Jicong Fan, Haijun Zhang, Mingliang Xu, and Meng Wang. Robust low-rank deep feature recovery in cnns: Toward low information loss and fast convergence. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 529–538. IEEE, 2021. 4

[27] Jiahuan Ren, Zhao Zhang, Richang Hong, Mingliang Xu, Haijun Zhang, Mingbo Zhao, and Meng Wang. Robust low-rank convolution network for image denoising. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6211–6219, 2022. 4

[28] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank

regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 2

[29] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 2, 5

[30] Onkar Susladkar, Gayatri Deshmukh, Subhrajit Nag, Ananya Mantravadi, Dhruv Makwana, Sujitha Ravichandran, Gajanan H Chavhan, C Krishna Mohan, Sparsh Mittal, et al. Clarifynet: A high-pass and low-pass filtering based cnn for single image dehazing. *Journal of Systems Architecture*, 132:102736, 2022. 4

[31] Li Tao, Chuang Zhu, Guoqing Xiang, Yuan Li, Huizhu Jia, and Xiaodong Xie. Llcnn: A convolutional neural network for low-light image enhancement. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2017. 2

[32] Vassilios Vonikakis, Dimitris Chrysostomou, Rigas Kouskouridas, and Antonios Gasteratos. Improving the robustness in feature detection by local contrast enhancement. In *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*, pages 158–163. IEEE, 2012. 5

[33] Qing Wang and Rabab K. Ward. Fast image/video contrast enhancement based on weighted thresholded histogram equalization. *IEEE Transactions on Consumer Electronics*, 53(2):757–764, 2007. 1

[34] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 2

[35] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013. 5

[36] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. Gladnet: Low-light enhancement network with global awareness. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 751–755. IEEE, 2018. 3, 5

[37] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. 2, 5

[38] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2604–2612, 2022. 2, 6

[39] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[40] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *British Machine Vision Conference*, 2018. 1, 2, 5, 6

[41] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2970–2979, 2017. 4

[42] 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. 2, 6

[43] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. 3

[44] 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. 5

[45] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*, 2017. 4

[46] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 3015–3022, 2017. 4

[47] Hongyan Zhang, Hongyu Chen, Guangyi Yang, and Liangpei Zhang. Lr-net: Low-rank spatial-spectral network for hyperspectral image denoising. *IEEE Transactions on Image Processing*, 30:8743–8758, 2021. 4

[48] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 5

[49] 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. 2, 6

[50] 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. 2, 6

[51] Zhao Zhang, Jiahuan Ren, Zheng Zhang, and Guangcan Liu. Deep latent low-rank fusion network for progressive subspace discovery. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2762–2768, 2021. 4

[52] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1899–1908, 2022. 2, 6