

RETINEXUTV: ROBUST RETINEX MODEL WITH UNFOLDING TOTAL VARIATION

Anonymous authors

Paper under double-blind review

ABSTRACT

Digital images are underexposed due to poor scene lighting or hardware limitations, reducing visibility and level of detail in the image, which will affect subsequent high-level tasks and image aesthetics. Therefore, it is of great practical significance to enhance low-light images. Among existing low-light image enhancement techniques, Retinex-based methods are the focus today. However, most Retinex methods either ignore or poorly handle noise during enhancement, which can produce unpleasant visual effects in low-light image enhancement and affect high-level tasks. In this paper, we propose a robust low-light image enhancement method RetinexUTV, which aims to enhance low-light images well while suppressing noise. In RetinexUTV, we propose an adaptive illumination estimation unfolded total variational network, which approximates the noise level of the real low-light image by learning the balance parameter of the total variation regularization term of the model, obtains the noise level map and the smooth noise-free sub-map of the image. The initial illumination map is then estimated by obtaining the illumination information of the smooth sub-map. The initial reflection map is obtained through the initial illumination map and original image. Under the guidance of the noise level map, the noise of the reflection map is suppressed, and finally it is multiplied by the adjusted illumination map to obtain the final enhancement result. We test our method on real low-light datasets LOL, VELOL, and experiments demonstrate that our method outperforms state-of-the-art methods.

1 INTRODUCTION

Recording people’s lives by taking pictures or videos is becoming more and more popular. However, since most users lack professional shooting skills, many photos are captured in less than ideal lighting conditions such as night time and backlight. Such images have low contrast, strong noise, and unclear details, which not only affect the human visual experience, but also limit the application of many computer vision algorithms, such as object recognition and object detection.

There are several approaches to enhance low-light images, including histogram equalization (Abdullah-Al-Wadud et al., 2007; Pizer, 1990), inverse domain operations (Li et al., 2015; Zhang et al., 2016), Retinex decomposition (Xiao & Shi, 2013; Jobson et al., 1997a;b; Herscovitz & Yadid-Pecht, 2004) and deep learning (Yang et al., 2016; Lore et al., 2017). Histogram equalization-based methods flatten the histogram and stretch the dynamic range of intensity, thereby amplifying the illumination of low-light images. If noise is not specifically considered, noise and artifacts will be amplified in its results. Some researchers noticed the similarity between haze images and inverted low-light images. Therefore, these inverse domain-based methods apply de-overlapping methods to enhance low-light images. To jointly adjust illumination and suppress noise, a method based on Retinex theory is proposed. Methods based on Retinex decomposition treat the scene in the human eye as the product of the reflection layer and the illumination layer. Enhanced results are produced by adjusting the corresponding layers. The earliest methods directly regard the decomposed reflection layer as the enhancement result (Jobson et al., 1997b;a; Xiao & Shi, 2013; Herscovitz & Yadid-Pecht, 2004). Single-scale Retinex (SSR)(Jobson et al., 1997b) and Multi-scale Retinex (MSR) (Jobson et al., 1997a) utilize Gaussian filters to build Retinex representations. In (Xiao & Shi, 2013), a bilateral filter is used to remove halo artifacts. Later methods adjust the illumination and reflection layers and reconstruct the enhanced result by combining them. In (Kimmel et al.,

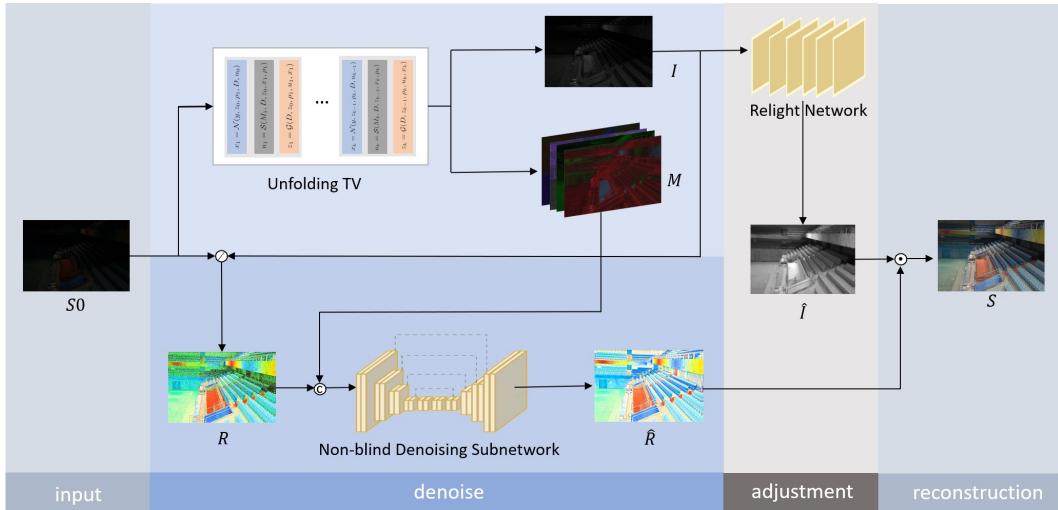


Figure 1: The proposed framework. Our sequential processing is divided into two stages. In the first stage, the illumination map and noise level map are obtained by using the unfolding total variation. In the second stage, the reflection map is obtained by calculating the illumination map and the original image. Input the reflection map and the noise level map into the non-blind denoising subnetwork to suppress the noise to obtain reflection map without noise, then adjusted with the illumination map to obtain an enhanced denoised image.

2003; Ng & Wang, 2011; Fu et al., 2014), the variational model estimates the piecewise continuous reflection layer and smooth illumination layer of the Retinex model. With the development of deep learning, many deep learning based methods have been proposed, Lore et al. (2017) used a deep autoencoder named Low Light Net (LLNet) to perform contrast enhancement and denoising.

Due to the interpretability of the Retinex theory and the ease of modeling, more and more methods have been proposed to combine deep learning with Retinex. The first one to combine deep learning with Retinex is RetinexNet(Wei et al., 2018), which uses a decomposition network and a brightness adjustment network. The two sub-networks complete the enhancement of the dark light image, and the subsequent methods are improved on this basis, such as adding a denoising network to the decomposed reflection image, and adding a brightness correction network to the decomposed lighting network, such as KinD(Zhang et al., 2019). These methods are all trained under the expected noise model, but still lack robustness in real dark-light environments because the noise exhibits different noise levels.

In our proposed method, we use the unfolding total variational model to estimate the noise level map and the illumination map, then the reflection map is obtained through the illumination map, and the reflection map is denoised through the guidance of the noise level map. In the Retinex theory, the anticipate illumination map needs to be smooth enough in space, while the reflection map needs edge details to represent the essence of the object. In this paper, we consider noise as a non-negligible factor in Retinex-based decomposition. The proposed model and method are noise-aware throughout the process, rather than in the form of individual ad-hoc operations. Compared with previous methods that only consider light noise modeling, this paper also aims to model and remove strong noise in low-light images(Wang et al., 2020). Firstly, we build a unfolding total variational model to estimate the noise level map and noise-free smooth map, and then obtain the illumination information of the noise-free smooth map to obtain the original illumination map. Then according to the retinex theory, the original illumination map and the original image are calculated to obtain the original reflection map, and the reflection map is denoised by the guidance of the noise level map. The illumination map is adjusted by the light adjustment network to obtain the adjusted illumination map. Finally, multiply them to get the enhanced image. Specific process is shown in Figure 1. The contributions of this paper are mainly reflected in the following aspects:

- We propose a robust low-light enhancement method RetinexUTV for joint denoising. Previous Retinex decompositions often ignored noise as a pre- or post-processing term, which compromised the overall visual quality. The variational model we built estimates both the noise level map and the illumination map, resulting in a noise-free enhancement result.
- We introduce noise levels into Retinex enhancements. Most of the existing enhancement algorithms only focus on low visibility problems or suppress noise under assumed noise levels, resulting in a lack of robustness. We learn a noise level map by learning a balance parameter in a model-based total variation regularization denoising method to approximate the noise level of real low-light images.
- Our proposed method achieves excellent results on real captured low-light images with various noise levels.

2 RELATED WORK

The traditional Retinex model(Land & McCann, 1971) regards the image S as the physical product of the reflection layer R and the illumination layer I , and models it as: $S = I \cdot R$, where the reflection R describes the intrinsic properties of the captured object, is considered consistent under any brightness condition, and is full of structural details. Illumination I represents various brightnesses on the object. It is segment-wise continuous and preserves dominant edges without small gradients. We know that low light may introduce a lot of noise to the image, and when the image is enhanced, it will inevitably exacerbate the noise, so we consider a robust Retinex model (Li et al., 2018), which contains an additional noise term N , As follows:

$$S = I \cdot R + N. \quad (1)$$

Many methods focus on the illumination component I . For example, in RetinexDIP(Zhao et al., 2021), two DIP(Ulyanov et al., 2018) networks are used to decompose I and R . Since DIP iterations take many times to learn enough details, R needs a lot of iterations, while I can learn approximate contours and illumination distributions in a few iterations, Therefore, the I and R decompositions cannot be well controlled. In the end simply take $R = S/I$ as the obtained reflectance, which actually keeps $R = R+N/I$. Therefore, these methods always lead to noisy results and are sensitive to noise, usually requiring an additional denoising process. Finally using R as the final enhancement result will lead to over-enhancement and affect the look and feel of the image. In this case, we try to assume full noise awareness. Noise has a negative impact on the visual quality of enhancement results, and is also a factor that cannot be ignored in low-light enhancement. Most previous methods suppress noise through preprocessing/postprocessing, which can easily lead to residual noise or over-smoothed details in the results. Therefore, an ideal low-light enhancement method should fully understand the noise and adaptively handle the noise throughout the enhancement process.

In RetinexNet(Wei et al., 2018), a decomposition network is used to decompose the illumination and reflection components, and then the reflection and illumination components are processed. For example, use BM3D(Burger et al., 2012) to denoise images. BM3D is one of the most classic algorithms for image denoising. By matching with adjacent image blocks, several similar blocks are integrated into a three-dimensional matrix, filtered in three-dimensional space, and then The resulting inverse transform is fused to 2D to form a denoised image. With the development of convolutional neural networks, more and more efficient and fast algorithms have been proposed. DnCNN(Zhang et al., 2017) denoises the image by learning the image residual, that is, the difference between the noisy image and the noiseless image, and estimating the noise by the residual learning. FFDNet(Zhang et al., 2018a) highlights the importance of the noise level map in balancing noise reduction and detail preservation. Additional noise level maps are added as input, which can handle different noise levels and can also handle spatially correlated noise. CBDNet(Guo et al., 2019) is divided into two parts, the first part is a five-layer fully convolutional network for noise estimation noise level map, and the second part is different from FFDNet, which is UNet with residual for noise reduction. On the basis of the above, Zheng et al. (2021) proposed the unfolding total variation model, and expanded the sub-problems of the fidelity term and the regularization term, using the network to predict the balance parameters, and approximate the noise level to estimate the noise level map for denoising. Here, we estimate the noise level map of the image through a variational model. At the same time, the illumination of the image is estimated. The illumination component of the image is piecewise smooth and noise-free, so the noise is forced to exist in another decomposed

component, namely $S = I \cdot (R + N)$. Finally, R is denoised under the guidance of the estimated noise level map.

3 PROPOSED METHOD

In this section, we build a total variational model to estimate the Retinex model’s illumination and noise level maps for images. Then, a sequential solution for robust low-light image enhancement is proposed. Figure 1 shows the framework of our method.

3.1 UNFOLDING TOTAL VARIATION MODEL

Most model-based denoising methods can be expressed as:

$$\hat{x} = \arg \min_x \frac{1}{2\sigma^2} \|x - y\|^2 + \lambda \Phi(x). \quad (2)$$

where $\frac{1}{2\sigma^2} \|x - y\|^2$ is the fidelity term, σ is the noise level, λ is the balance parameter, $\Phi(x)$ is the regularization term, \hat{x} is the solution to the problem, and x and y are the clean and observed images, respectively. In previous work FFDnet [32], when λ is combined with σ , setting the noise level also achieves the effect of setting λ , that is, controlling the trade-off between noise reduction and detail preservation. Inspired by this relationship, for some regularization terms, if we combine λ and σ as a balance parameter, we can learn such a balance parameter to approximate the noise level through the unrolling architecture.

The operator D is defined as $D = [D_x^T D_y^T]^T$, where D_x, D_y are the first-order forward finite difference operators along the horizontal and vertical directions, respectively. Therefore, the anisotropic total variation regularization term can be written as $\|D_x\|_1$. At the same time, the balance parameter λ can be incorporated into the regularization term, written as $\|\lambda D_x\|_1$. Since the noise of real scenes exhibits different patterns, λ can be extended into a noise level map $\hat{M} = [M_x^T M_y^T]^T$. So the above denoising model can be written as:

$$\underset{x}{\text{minimize}} \frac{1}{2} \|x - y\|^2 + \|\hat{M} \cdot (D_x)\|_1. \quad (3)$$

We set $M_x = M_y = M$ as the noise level map we need to approximate in the network. By introducing the intermediate variable $u = D_x$, we obtain the augmented Lagrangian function of the equation:

$$L(x, u, z) = \frac{1}{2} \|x - y\|^2 + \|\hat{M} \cdot u\| - z^T (u - D_x) + \frac{\rho_r}{2} \|u - D_x\| \quad (4)$$

where z is the Lagrange multiplier and r is the regularization parameter. Iteratively updates x , u , and z through ADMM. The subproblem for x and the subproblem for u can be solved using the Fast Fourier transform (Fast Fourier transform FFT) and the shrink function, respectively. Therefore, unrolled inferences can be obtained by solving the following subproblems (specific derivation can refer to our previous work (Zheng et al., 2021)):

$$\begin{cases} x_k = \mathcal{N}(y, z_{k-1}, \rho_k, D, u_{k-1}), \\ u_k = \mathcal{S}(M_k, D, z_{k-1}, x_k, \rho_k), \\ z_k = \mathcal{G}(D, z_{k-1}, \rho_k, u_k, x_k) \end{cases} \quad (5)$$

where the function $\mathcal{N}(\cdot)$ is designed to solve the fidelity sub-problem and guarantees the similarity of the smooth and noise-free layer y_s and the original image y . The function $\mathcal{G}(\cdot)$ is associated with the constraint $u = D_x$, the Lagrangian multiplier z should be updated as the iteration progresses. The function $\mathcal{S}(\cdot)$ contains the approximate noise level map M . This function can be seen as a special smoothing constraint for the low-frequency layer y_s , which smoothes details and noise according to the magnitude of the value corresponding to each pixel in the noise level map M . If M is too small, there will still be a lot of noise in the output sub-image after iteration. If M is too large, the output sub-image will be an overly smooth sub-image. From this, smooth noise-free y_s and noise level M are obtained.

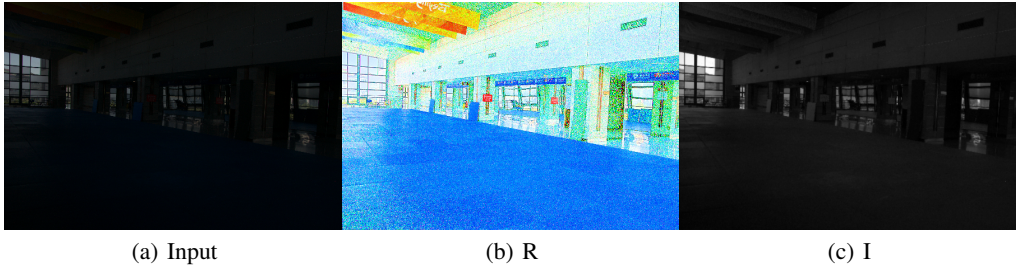


Figure 2: Illumination image I and reflection image R . We obtain illumination I by unfolding total variation, and then divide the input image by I to obtain the initial reflection image R .

3.2 ILLUMINATION ESTIMATE

Assumed in Retinex theory. 1) The illustration I is spatially smooth; 2) The value of reflection R ranges from 0 to 1, indicating that $I \leq S$, S represents the original image; 3) The reflection R contains high frequency parts, i.e. edge and texture information. Similar to LIME(Guo et al., 2016) and RetinexDIP(Zhao et al., 2021), perform maxRGB on the initial image to obtain the maximum value of the three color channels as the initial illumination I_0 , and then obtain the illumination through the network model or optimization. Due to the unfolding total variational model, we denoise and smooth the original image to obtain a noise-free and smooth layer y_s . Similarly, we maximize its three color channels to obtain our initial illumination map I , as above As shown in Figure 2(c), we then divide the input low-light image by the illustration map I to obtain our desired reflection image R , as shown in Figure 2(b). After obtaining the illumination map I , we put it into the light adjustment network. In the previous optimization method and RetinexDIP, the illumination map is usually processed by gamma transformation. Here we use a six-layer convolutional neural network to The illumination is processed to obtain an illumination image \hat{I} after illumination adjustment. As shown on the right in Figure 1.

3.3 REFLECTION DENOISING

After the appeal operation, we divide the input image by the estimated illumination image to obtain the reflection image R , we can see that the reflection image contains a lot of noise, as shown in Figure 2(b). In the noise suppression sub-network, as pointed out in CBDNet(Guo et al., 2019) and FFDnet(Zhang et al., 2018a), taking the noise image and noise level map as input helps to enhance the generalization ability of the model and improve the performance for blind denoising. We adopt a method that takes as input the reflection image R and an approximate noise level map M to perform noise suppression and detail recovery on the reflection image. We adopt the U-Net architecture with four scales, and the number of channels in the convolutional layer at each scale is 32, 64, 128, and 256, respectively. Average pooling and fully connected layers are used to process global features and connect them with previous layers. 2×2 convolution and transposed convolution are used to downscale and upscale feature layers, respectively. For all convolutional layers except the last one, we use LeakyReLU as activation function.

3.4 END-TO-END TRAINING

The TV minimization problem is solved when the TV module is unfolded to approximate the noise level, and the noise level map and the illumination map are obtained. The noise level map guides the reflection map for denoising, and finally the reflection map and the adjusted illumination map are multiplied to obtain Enhance image. Therefore, the network can be trained end-to-end. We optimize the weights and biases by minimizing the L_2 loss \mathcal{L}_{l_2} , perceptual loss \mathcal{L}_{per} and structural similarity loss \mathcal{L}_{ssim} to 1, 0.12 and 0.82, respectively.

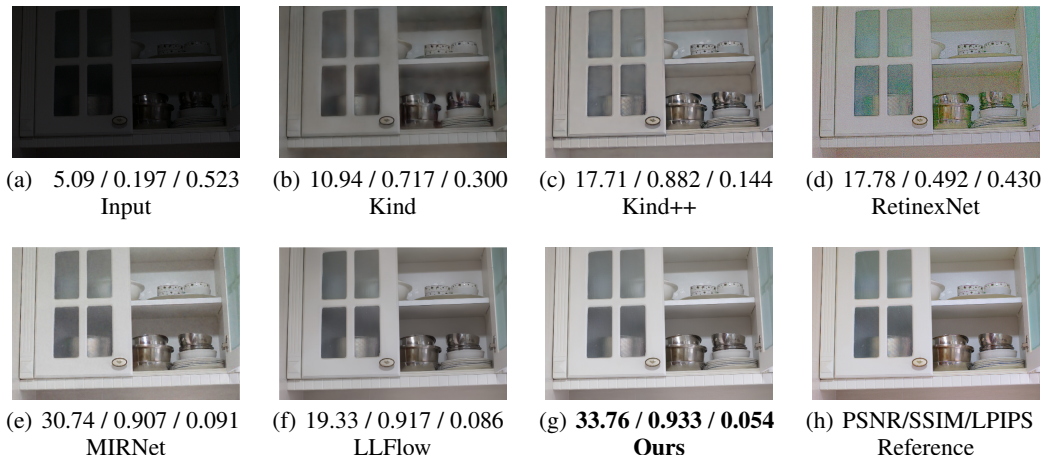


Figure 3: Visual Comparison with state-of-the-art low-light image enhancement methods on the LOL dataset. Please zoom in for a better review.

Table 1: Quantitative comparison on the LOL dataset (Wei et al., 2018) in terms of PSNR, SSIM and LPIPS. \uparrow (\downarrow) denotes that, larger (smaller) values lead to better quality. The best and second best scores are **highlighted** and underlined, respectively.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Zero-DCE (Guo et al. 2020)	14.86	0.54	0.33
LIME (Guo, Li, and Ling 2016)	16.76	0.56	0.35
EnlightenGAN (Jiang et al. 2021)	17.48	0.65	0.32
RetinexNet (Wei et al. 2018)	16.77	0.56	0.47
RUAS (Risheng et al. 2021)	18.23	0.72	0.35
DRBN (Yang et al. 2020)	20.13	0.83	0.16
KinD (Zhang et al. 2019)	20.87	0.80	0.17
KinD++ (Zhang et al. 2021)	21.30	0.82	0.16
MIRNet (Zamir et al. 2020)	24.14	0.83	0.13
LLFlow (Wang et al. 2022)	25.12	0.93	<u>0.11</u>
RetinexUTV (Ours)	<u>24.47</u>	<u>0.86</u>	0.09

4 EXPERIMENTS

4.1 IMPLEMENTATION DETAILS

The proposed RetinexUTV is implemented in the PyTorch framework, and before training, we initialize the parameters of RetinexUTV. First, the parameters of RetinexUTV are optimized, using ADAM optimizer for 2×10^3 epochs, the patch size is set to 256×256 and the batch size is set to 2. The initial learning rate is 5×10^{-4} and decrease to 1×10^{-6} by the cosine annealing strategy. All experiments are performed on an NVIDIA TITAN Xp GPU.

4.2 COMPARISON WITH STATE-OF-THE-ARTS ON THE REAL DATASETS

For a fair comparison, we use the published code for these methods without any modifications. Since Zero-DCE(Guo et al., 2020) and EnlightenGAN(Jiang et al., 2021) are trained with unpaired data, we use their published pretrained models for comparison. The LOL dataset(Wei et al., 2018) captures 500 pairs of real low/normal light images by varying the exposure time and ISO of the camera, which includes 485 training images and 15 testing images. VE-LOL(Liu et al., 2021) contains two subsets: paired VE-LOL-L is used to train and evaluate the LLIE method, and non-paired VE-LOL-H is used to evaluate the effect of the LLIE method on face detection. Here we use VELOL-L. There are 500 real scene images in VE-LOL-L, of which 400 are used for training and 100 are used for testing. Three metrics were used for quantitative comparison, including peak signal-to-

Table 2: Quantitative comparison on the VE-LOL dataset(Liu et al., 2021) in terms of PSNR, SSIM and LPIPS. The models are trained on the training set of LOL. \uparrow (\downarrow) denotes that, larger (smaller) values lead to better quality.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
RetinexNet (Wei et al. 2018)	14.68	0.525	0.642
BIMEF(Ying et al. 2017)	15.95	0.639	0.457
DeepUPE(Ying et al. 2019a)	13.19	0.490	0.463
JED(Ren et al. 2018)	16.73	0.682	0.390
LIME (Guo, Li, and Ling 2016)	14.07	0.527	0.402
LLNet (RLore et al. 2017)	17.57	0.739	0.402
KinD (Zhang et al. 2019)	18.42	0.766	0.288
KinD++ (Zhang et al. 2021)	17.63	0.799	0.226
Zero-DCE (Guo et al. 2020)	21.12	0.771	0.248
EnlightenGAN (Jiang et al. 2021)	20.43	0.792	0.242
LLFlow (Wang et al. 2022)	23.85	0.899	0.146
RetinexUTV (Ours)	29.71	0.915	0.056

noise ratio(PSNR), structural similarity(SSIM)(Wang et al., 2004), and learned perceptual image patch similarity(LPIPS)(Zhang et al., 2018b). We train our model on the LOL dataset and test it on the LOL dataset and the VE-LOL dataset, and the numerical results between different methods are shown in Table I and Table II. From Table I, we can find that our method roughly outperforms other competitors, and in Table II, we can see that our method outperforms all other methods. Higher PSNR values indicate that our method is better at suppressing artifacts and recovering color information. Better SSIM values indicate that our method better preserves the structural information of high-frequency details. Our method also achieves the best performance on LPIPS, a metric designed for human perception, which shows that our method is better aligned with human perception. The qualitative results on the LOL dataset are shown in Figure 3, and the qualitative comparison of the real images on the VE-LOL dataset is shown in Figure 4. Our method produces results with less noise and better color saturation.

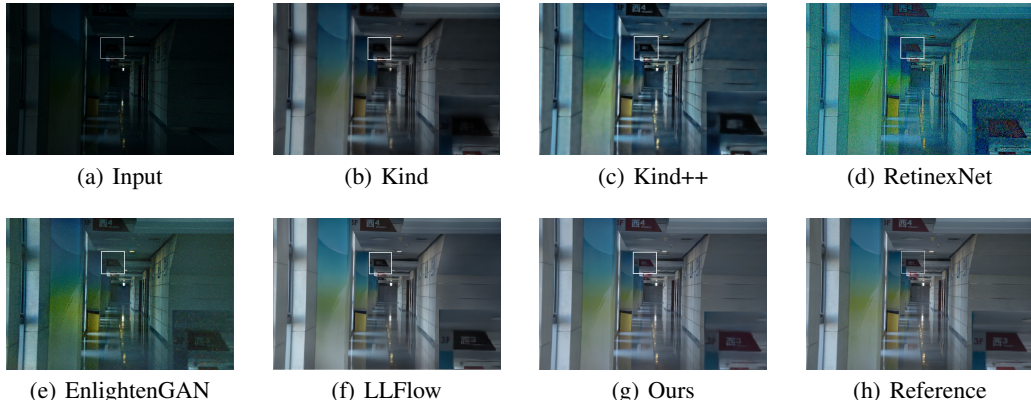


Figure 4: Visual comparison with state-of-the-art low-light image enhancement methods on the real-captured set of VE-LOL dataset. Please zoom in for a better review.

4.3 ABLATION STUDY

In this section, we quantitatively evaluate the effectiveness of the loss function setting in our model based on the LOL dataset. The results are shown in Table III.

Loss function setting: To explore the effectiveness of the loss function setting, we convert the MSE loss to L1 loss, remove the perceptual loss and remove the structural similarity loss for experiments respectively. Using MSE loss over L1 loss gives PSNR of 1.28dB (=24.47-23.19), SSIM of 0.007 (=0.860-0.853), and LPIPS of 0.008 (=0.098-0.090). Removing perceptual loss and structural sim-

Table 3: Ablation study. This table reports the performance under each condition based on the LOL dataset. In this table, "w/o" means without.

Conditions	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
default	24.47	0.860	0.090
MSE \rightarrow L1	23.19	0.853	0.098
w/o perceptual loss	22.15	0.842	0.142
w/o SSIM loss	23.71	0.836	0.093

ilarity loss leads to performance degradation. After removing the perceptual loss, PSNR drops by 2.32db ($=24.47-22.15$), SSIM drops by 0.018 ($0.860-0.842$) and LPIPS drops by 0.052 ($=0.142-0.090$). After removing the structural similarity loss, PSNR drops by 0.76db ($=24.47-23.71$), SSIM drops by 0.024 ($0.860-0.842$), and LPIPS drops by 0.003 ($=0.093-0.090$). The experimental results verify the rationality of our loss function settings.

5 CONCLUSION

In this study, we propose a new true low-to-true normal network for low-light image enhancement based on Retinex theory, RetinexUTV, which consists of three sub-networks: unfolding total variational network, denoising network and relighting network. The enhanced results obtained by our method have better visual quality. Results On publicly available datasets, our method can appropriately improve image contrast and suppress noise, and achieve the highest PSNR and SSIM scores, outperforming existing methods by a large margin.

In the future, since supervised learning-based methods face some challenges: 1) it is difficult to collect large-scale paired datasets covering various real-world low-light conditions, 2) synthetic low-light images cannot accurately represent real-world illumination conditions, and 3) training deep models on paired data may lead to overfitting and limited generalization to real-world images with different lighting properties(Li et al., 2021). Our previous work, RetinexDIP(Zhao et al., 2021), is an unsupervised method that can enhance various low-light conditions, but cannot deal with noise. We can see in Figure 5 that the supervised based methods do not perform well for augmentation compared to unsupervised RetinexDIP and will introduce halos. Therefore, our next work will go in the unsupervised direction, which is also the focus of low-light enhancement.

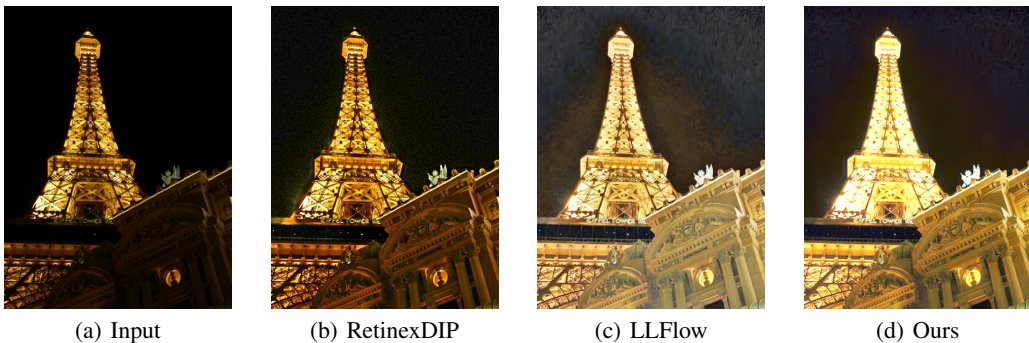


Figure 5: Visual comparison with state-of-the-art low-light image enhancement methods on the real-captured set of DICM(Lee et al., 2013). Please zoom in for a better review.

REFERENCES

Mohammad 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.

- Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pp. 2392–2399. IEEE, 2012.
- Xueyang Fu, Ye Sun, Minghui LiWang, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A novel retinex based approach for image enhancement with illumination adjustment. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1190–1194. IEEE, 2014.
- Chunle Guo 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 conference on computer vision and pattern recognition (CVPR)*, pp. 1780–1789, June 2020.
- Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1712–1722, 2019.
- 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.
- Marius Herscovitz and Orly Yadid-Pecht. A modified multi scale retinex algorithm with an improved global impression of brightness for wide dynamic range pictures. *Machine Vision and Applications*, 15(4):220–228, 2004.
- 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.
- Daniel J Jobson, Zia-ur Rahman, and Glenn 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, 1997a.
- Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997b.
- Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003.
- Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971.
- Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013.
- Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (01):1–1, 2021.
- Lin Li, Ronggang Wang, Wenmin Wang, and Wen Gao. A low-light image enhancement method for both denoising and contrast enlarging. In *2015 IEEE international conference on image processing (ICIP)*, pp. 3730–3734. IEEE, 2015.
- Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018.
- Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021.
- 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.

- Michael K Ng and Wei Wang. A total variation model for retinex. *SIAM Journal on Imaging Sciences*, 4(1):345–365, 2011.
- Stephen M Pizer. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In *Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia*, volume 337, pp. 1, 1990.
- Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9446–9454, 2018.
- Wencheng Wang, Xiaojin Wu, Xiaohui Yuan, and Zairui Gao. An experiment-based review of low-light image enhancement methods. *Ieee Access*, 8:87884–87917, 2020.
- 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.
- Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.
- Chuanmin Xiao and Zelin Shi. Adaptive bilateral filtering and its application in retinex image enhancement. In *2013 Seventh International Conference on Image and Graphics*, pp. 45–49. IEEE, 2013.
- Jie Yang, Xinwei Jiang, Chunhong Pan, and Cheng-Lin Liu. Enhancement of low light level images with coupled dictionary learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 751–756. IEEE, 2016.
- Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.
- Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018a.
- Liang Zhang, Peiyi Shen, Xilu Peng, Guangming Zhu, Juan Song, Wei Wei, and Houbing Song. Simultaneous enhancement and noise reduction of a single low-light image. *IET Image Processing*, 10(11):840–847, 2016.
- 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*, pp. 586–595, 2018b.
- 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*, pp. 1632–1640, 2019.
- Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1076–1088, 2021.
- Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4439–4448, 2021.