Unsupervised Night Image Enhancement: When Layer Decomposition Meets Light-Effects Suppression

Yeying Jin[1][0000−0001−7818−9534], Wenhan Yang[2][0000−0002−1692−0069], and Robby T. Tan[1,3][0000−0001−5440−9678]

[1] National University of Singapore
[2] Nanyang Technological University
[3] Yale-NUS College
jinyeying@u.nus.edu, wenhan.yang@ntu.edu.sg, robby.tan@{nus,yale-nus}.edu.sg

Abstract. Night images suffer not only from low light, but also from uneven distributions of light. Most existing night visibility enhancement methods focus mainly on enhancing low-light regions. This inevitably leads to over enhancement and saturation in bright regions, such as those regions affected by light effects (glare, floodlight, etc). To address this problem, we need to suppress the light effects in bright regions while, at the same time, boosting the intensity of dark regions. With this idea in mind, we introduce an unsupervised method that integrates a layer decomposition network and a light-effects suppression network. Given a single night image as input, our decomposition network learns to decompose shading, reflectance and light-effects layers, guided by unsupervised layer-specific prior losses. Our light-effects suppression network further suppresses the light effects and, at the same time, enhances the illumination in dark regions. This light-effects suppression network exploits the estimated light-effects layer as the guidance to focus on the light-effects regions. To recover the background details and reduce hallucination/artefacts, we propose structure and high-frequency consistency losses. Our quantitative and qualitative evaluations on real images show that our method outperforms state-of-the-art methods in suppressing night light effects and boosting the intensity of dark regions. [4]
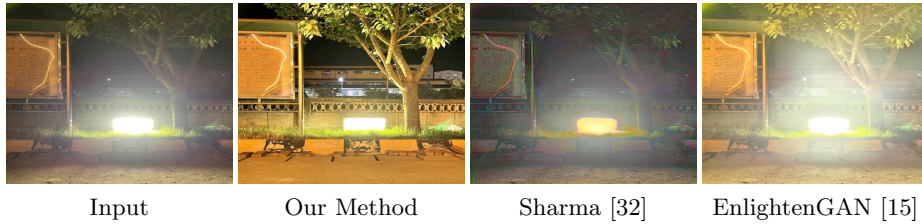
Keywords: Night image enhancement, low-light image, light-effects suppression

1 Introduction

Night images can contain uneven light distributions, as shown in Fig. 1, where some regions are dark and some are significantly brighter, due to the presence of light effects[5]. Most existing nighttime visibility enhancement methods focus mainly on boosting the intensity of low-light regions, e.g., [14, 7, 33, 13, 15].

---

[4] Our data and code is available at: https://github.com/jinyeying/night-enhancement
[5] Following [32], light effects in this paper refer to glare, floodlight, etc.

| Input | Our Method | Sharma [32] | EnlightenGAN [15] |

Fig. 1. An existing night light-effects suppression method [32] suffers from hallucination/artefacts and generates improper light effects, while an image enhancement method [15] is not designed to handle night light effects and incorrectly intensifies it. In contrast, our method jointly suppresses light effects and enhances dark regions.

Hence, when these methods are applied to night images that contain light effects, they inevitably amplify the light effects, and impair the visibility of the images even further. Unlike these methods, our goal in this paper is to suppress the light effects while, at the same time, boosting the intensity of dark regions.

Fully-supervised learning methods could be a possible solution to achieving our goal. However, these methods would require a diverse and large collection of paired night images taken with and without light effects, which is intractable to obtain. Another possible solution would be the use of synthetic night images with rendered light effects. However, the effectiveness of methods trained on synthetic night data depends on the quality of the light-effects rendering model. To our knowledge, rendering physically correct night light effects with various background scenes and lighting conditions is still challenging [36].

In this paper, we introduce an unsupervised learning approach that integrates a decomposition network and a light-effects suppression network in a single unified framework. Our decomposition network is derived from an image-layer model and guided by our layer-specific prior losses to decompose the input image into shading, reflectance and light-effects layers (Fig. 3 shows the examples of these three layers). Subsequently, our light-effects suppression network, which is trained on unpaired images with and without light effects, provides additional unsupervised constraints. This network not only strengthens the light effects decomposition but also enhances the intensity in dark regions. The two networks, the decomposition and light-effect suppression networks, are connected.

To recover the background details behind light-effects regions, we introduce structure and high-frequency (HF) features consistency losses. We employ the structure consistency based on the VGG network and utilize the guided filter to obtain HF features. The structure and HF-features consistency losses can also reduce hallucination. In summary, our main contributions are as follows:

– To enhance the visibility of night images that suffer from low light and light effects simultaneously, we introduce a network architecture that integrates layer decomposition and light-effects suppression in one unified framework.

- To distinguish light effects from background regions, particularly when the color of the light effects is white or achromatic, we propose utilizing the estimated light-effects layer as guidance for our unsupervised light-effects suppression network.
- To restore the background details, we introduce novel unsupervised losses based on the structure and HF-features consistency. Our perceptual structure information and HF texture information are less affected by light effects. Thus, they can be employed to preserve background details, and, importantly, to suppress unwanted artefacts.

Our experiments and evaluations show that our method is effective in suppressing light-effects regions and enhancing dark regions, outperforming state-of-the-art methods both quantitatively and qualitatively.

2   Related Work

Sharma and Tan [32] introduce a method based on camera response function (CRF) estimation and HDR imaging to suppress light effects. The method is the first method that can suppress light effects and improve the dynamic range for night images. However, it suffers from artefacts and missing details as shown in Fig. 1, particularly for white (or achromatic) lights.

In the field of night image dehazing, a few methods have been proposed to suppress glow due to haze/fog particles. Li et al. [23] address glow removal on foggy nights using layer separation. Zhang et al. [44] use maximum reflectance prior for haze and glow removal. Ancuti et al. [2, 3] use a fusion process and the Laplace operator to deglow and dehaze. Yan et al. [38, 39] propose a semi-supervised method [37] employing a grayscale guided network. However, all these methods are designed for glow suppression in haze or foggy night, and not for removing light effects in clear night images. Moreover, unlike our method, they are also not designed for enhancing dark regions.

A number of methods have been developed to boost the brightness of low-light images without considering the presence of night light effects. A few methods are based on histogram equalization [28], inversion and dehazing [8], the retinex model (e.g. [9, 21]), while more recent methods are based on deep networks [20]. Most deep-learning-based methods (e.g. [1, 7, 33]) adopt supervised learning to train their model and thus require a large number of pairs of low/normal-light images. A few unsupervised methods (e.g. [15]) rely on adversarial training using unpaired low/normal-light images. Semi-supervised methods (e.g. [40, 41]) recompose coarse-to-fine representations towards perceptually pleasing images with the help of unpaired high-quality images. Recently, zero-shot learning methods (e.g. [19, 13]) have been proposed for low-light enhancement. Most of these night image enhancement methods, however, are not designed to suppress night light effects and enhance low light regions simultaneously; therein lies the main difference with our work.
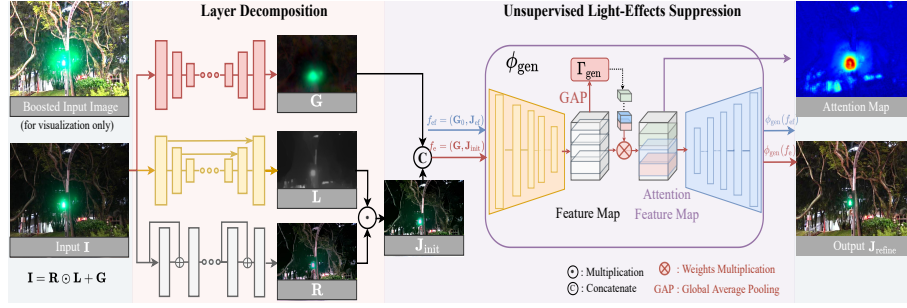
Fig. 2. The overall architecture of our proposed method. We integrate decomposition and light-effects suppression networks in one unified unsupervised framework. Given the input night image, we suppress light effects through the layer decomposition network, in which light-effects, shading, and reflectance layers are obtained (see Fig. 3). The light-effects suppression is guided by the decomposed light-effects layer $\mathbf{G}$ and based on unpaired learning (see Fig. 4) to further suppress light effects and boost dark regions.

3 Proposed Method

To suppress light effects and, at the same time, boost the intensity of dark regions, we propose an unsupervised framework by integrating a decomposition network and a light-effects suppression network. Our decomposition network is based on an image-layer model and produces three separate layers: shading, reflectance, and light-effects layers. We input these layers into our light-effects suppression network to obtain our final output, where light effects are suppressed and dark regions are boosted. This network learns from unpaired data and is guided by our estimated light-effects layer.

3.1 Model-Based Layer Decomposition Network

Our decomposition is based on the following image-layer model:

$$\mathbf{I} = \mathbf{R} \odot \mathbf{L} + \mathbf{G}, \tag{1}$$

where $\mathbf{I}$ represents the input night image, $\mathbf{G}$ represents the light-effects layer, $\mathbf{R}$ and $\mathbf{L}$ are the reflectance and shading layers, respectively. The notation $\odot$ represents element-wise multiplication. In this equation, we assume a linear gamma function. However, we do not use this equation explicitly in our method. Instead, we use it only to guide the design of our network in Fig. 2 (i.e., the layer decomposition network). When non-linear images with non-linear gamma functions are used in training, the background scenes are approximations of the physically correct values. Our decomposition goal is to obtain a background scene that is free from light effects, i.e., we want to estimate the background scene, $\mathbf{J}_{\mathrm{init}} = \mathbf{R} \odot \mathbf{L}$. Hence, even when non-linear images are used in training, applications that are

| (a) Input $\mathbf{I}$ | (b) $\mathbf{G}$ | (c) $\mathbf{L}$ | (d) $\mathbf{R}$ |
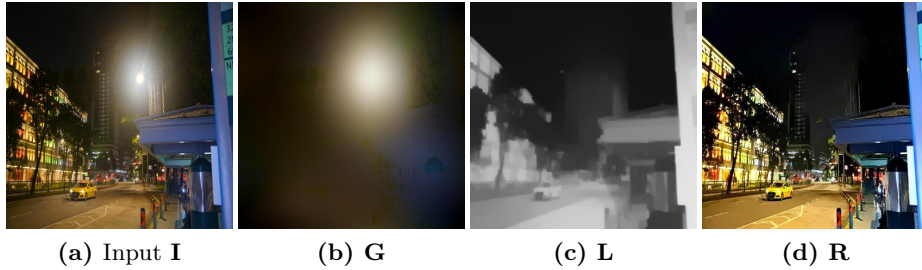
Fig. 3. Results of our model-based layer decomposition. (a) Input. (b) Light-effects layer $\mathbf{G}$. (c) Shading layer $\mathbf{L}$ and (d) Reflectance layer $\mathbf{R}$.

less concerned about physically correct intensity values but suffer from light effects can benefit from our method. Our model differs from the widely used intrinsic model [11, 4], as the latter does not incorporate the light-effects layer.

Fig. 2 shows our pipeline. The decomposition network is based on our image-layer model in Eq. (1). Given the input image ($\mathbf{I}$), we first perform image decomposition. We use three separate networks and our novel unsupervised losses to obtain the light effects ($\mathbf{G}$), shading ($\mathbf{L}$), and reflectance ($\mathbf{R}$) layers.

Learning Light Effects, Shading and Reflectance Layers To obtain the light effects ($\mathbf{G}$), shading ($\mathbf{L}$), and reflectance ($\mathbf{R}$) layers, we use three networks respectively: Light-Effects-Net (ϕ_G), Shading-Net (ϕ_L) and Reflectance-Net (ϕ_R), where $\mathbf{G} = \phi_\mathrm{G}(\mathbf{I})$, $\mathbf{L} = \phi_\mathrm{L}(\mathbf{I})$, and $\mathbf{R} = \phi_\mathrm{R}(\mathbf{I})$. The three networks are trained using unsupervised losses, which will be discussed in the subsequent paragraphs. Fig. 3 shows examples of these three layers.

Light-Effects and Shading Initialization To resolve the decomposition ambiguity problem, it is important to provide proper initial estimates of the layers. For the shading layer, we employ a shading map $\mathbf{L}_\mathrm{i}$ obtained by taking the maximum value of the three color channels, for each pixel [14]. For the light-effects layer, we use a light-effects map $\mathbf{G}_\mathrm{i}$, computed using the relative smoothness technique [22]. This is extracted using the second-order Laplacian filter from the input image, since light effects are smooth variations. We define the loss function for the initialization step as:

$$\mathcal{L}_\mathrm{init} = |\mathbf{G} - \mathbf{G}_\mathrm{i}|_1 + |\mathbf{L} - \mathbf{L}_\mathrm{i}|_1. \tag{2}$$

Gradient Exclusion Loss The gradients of the light effects layer have a short tail distribution, similar to that of 'glow' [23]. In contrast, the gradients of the background image have a long tail distribution [22]. Hence, we employ a gradient exclusion loss to recover the uncorrelated layers $\{\mathbf{G}, \mathbf{J}_\mathrm{init}\}$, where the goal is to separate the two layers as far as possible in the gradient space. The definition of
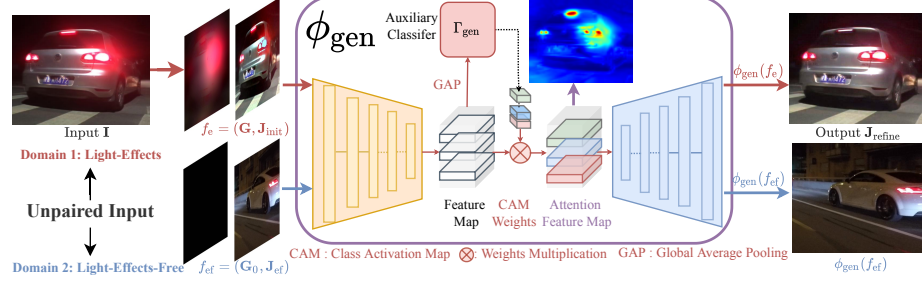
Fig. 4. Overview of our unsupervised light-effects suppression network. The network comprises a generator ϕ_{gen} and a classifier Γ_{gen}. The encoder block of our generator extracts feature maps from the input image layers. Our classifier Γ_{gen} is trained to learn the weights [49] of the feature maps. Γ_{gen} performs domain classification based on two domains, i.e., the light-effects domain $f_{\text{e}} = (\mathbf{G}, \mathbf{J}_{\text{init}})$ and the unpaired light-effects-free domain $f_{\text{ef}} = (\mathbf{G}_0, \mathbf{J}_{\text{ef}})$. Averaging the weighted feature maps generates an attention map that shows the network is focusing on the light-effects regions. As a result, the light effects are significantly suppressed in our output $\mathbf{J}_{\text{refine}}$.

the loss follows [10, 46]:

$$\mathcal{L}_{\text{excl}} = \sum_{n=1}^{3} \left\| \tanh(\lambda_{\mathbf{G}^{\downarrow n}} |\nabla \mathbf{G}^{\downarrow n}|) \circ \tanh(\lambda_{\mathbf{J}_{\text{init}}^{\downarrow n}} |\nabla \mathbf{J}_{\text{init}}^{\downarrow n}|) \right\|_F, \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{G}^{\downarrow n}$ and $\mathbf{J}_{\text{init}}^{\downarrow n}$ represent $\mathbf{G}$ and $\mathbf{J}_{\text{init}}$ downsampled using the bilinear interpolation, and the parameters $\lambda_{\mathbf{G}^{\downarrow n}}$ and $\lambda_{\mathbf{J}_{\text{init}}^{\downarrow n}}$ are normalization factors.

Color Constancy Loss To minimize any color shift in our decomposition output, inspired by the Gray World assumption [5, 13, 32], we use a color-constancy prior, which encourages the range of the intensity values of the three color channels in the background image $\mathbf{J}_{\text{init}}$ to be balanced:

$$\mathcal{L}_{\text{cc}} = \sum_{(c1,c2)} \left(|\mathbf{J}_{\text{init}}^{c1} - \mathbf{J}_{\text{init}}^{c2}|_1 \right), \tag{4}$$

where $(c1, c2) \in \{(r, g), (r, b), (g, b)\}$ denotes a combination of two color channels.

Reconstruction Loss For our decomposition task, recombining the estimated layers should give us back the original input image. Hence, we define our reconstruction loss as:

$$\mathcal{L}_{\text{recon}} = |\mathbf{I} - (\mathbf{R} \odot \mathbf{L} + \mathbf{G})|_1. \tag{5}$$

We multiply each unsupervised loss with its respective weight, where we set λ_{init}, λ_{excl} all set to 1 since they are in the same scale. We empirically set $\lambda_{\text{recon}} = 0.1$ and employ the weight $\lambda_{\text{cc}} = 0.5$ from [13] to balance the decomposition process.
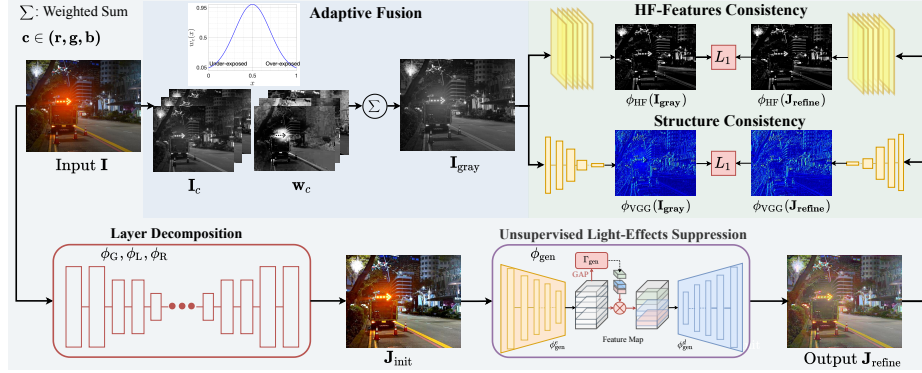
Fig. 5. Overview of our structure and HF-features consistency losses. We first use our adaptive fusion scheme to obtain a fused grayscale image $\mathbf{I}_{\mathrm{gray}}$. Then, from $\mathbf{I}_{\mathrm{gray}}$, we compute VGG features $\phi_{\mathrm{VGG}}(\mathbf{I}_{\mathrm{gray}})$ that are less affected by light effects, and HF-features $\phi_{\mathrm{HF}}(\mathbf{I}_{\mathrm{gray}})$ that are more robust to light effects and contain background details.

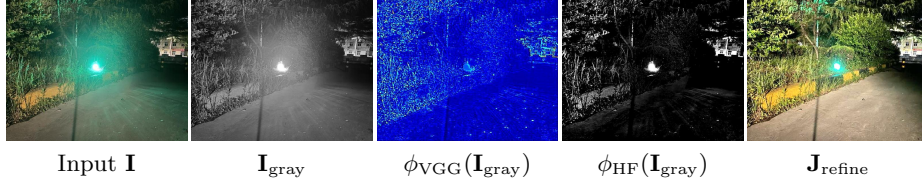### 3.2   Light-Effects Suppression Network

To better suppress light effects, we integrate our decomposition network with an unpaired light-effects suppression network. We design this network to suppress light effects by using the guidance of our estimated light-effects layer, enforcing the network to focus on light-effects regions. As shown in Fig. 2, our network comprises a generator ϕ_{gen} and a classifier Γ_{gen}. It refines the initially estimated background scene ($\mathbf{J}_{\mathrm{init}}$), and generates the final light-effects-free output ($\mathbf{J}_{\mathrm{refine}}$). The details are as follows.

Light-Effects Layer Guidance We employ the estimated light-effects layer $\mathbf{G}$ to guide our training process, as shown in Fig. 4. The light-effects layer is taken as part of the input of our encoder-decoder network, and is modulated with the feature maps of the network at different scales. Specifically, we concatenate $\mathbf{J}_{\mathrm{init}}$ with the light-effects layer $\mathbf{G}$, and then we input them to our network ϕ_{gen}.

By resizing the light-effects layer, $\mathbf{G}$, to fit the size of each feature map, and multiplying it with all the intermediate feature maps, our light-effects layer can guide our network to focus more on light-effects regions. Fig. 3b and Fig. 12 show some results of our light-effects layers, demonstrating that our method can successfully separate white and multi-color light effects.

Light-Effects Suppression Besides the light-effects layer, our suppression network is also guided by an attention mechanism [15, 18, 16]. The basic idea is that, we input the light-effects and light-effects-free unpaired images into our encoder-decoder network. We then, use a domain classifier to judge whether the encoded

| Input $\mathbf{I}$ | $\mathbf{I}_{\text{gray}}$ | $\phi_{\text{VGG}}(\mathbf{I}_{\text{gray}})$ | $\phi_{\text{HF}}(\mathbf{I}_{\text{gray}})$ | $\mathbf{J}_{\text{refine}}$ |

Fig. 6. Examples of feature map from VGG for $\mathbf{I}_{\text{gray}}$, and a HF feature map for $\mathbf{I}_{\text{gray}}$. As one can observe, these features are less affected by light effects.

features come from a certain domain, i.e., to judge whether the input is light-effects or light-effects-free. Using this domain classification, the activated feature regions can form an attention map [49] that is useful when guiding our network in suppressing light effects.

More specifically, as shown in Fig. 4, our network ϕ_{gen} contains an auxiliary classifier Γ_{gen}. One of the inputs of the network is the concatenation of $\mathbf{J}_{\text{init}}$ and $\mathbf{G}$. Another input is a light-effects-free reference image, $\mathbf{J}_{\text{ef}}$, concatenated with a dummy all zero map $\mathbf{G}_0$, which of course has no light effects. Our classifier, Γ_{gen}, then performs domain classification based on the encoded features from $f_{\text{e}} = (\mathbf{G}, \mathbf{J}_{\text{init}})$ or $f_{\text{ef}} = (\mathbf{G}_0, \mathbf{J}_{\text{ef}})$. To train the auxiliary classifier Γ_{gen}, we use the following attention loss:

$$\mathcal{L}_{\text{atten}} = -\big(\mathbb{E}\big[\log(\Gamma_{\text{gen}}(f_{\text{e}}))\big] + \mathbb{E}\big[\log(1 - \Gamma_{\text{gen}}(f_{\text{ef}})))\big]\big). \qquad (6)$$

Structure and HF-Features Consistency Losses To address hallucination/artefacts [31], and also to preserve background details, we employ two constraints: structure consistency, based on features obtained from the VGG network [17]; and HF-features consistency, based on the HF features obtained from the guided filter [35].

As shown in Fig. 5, to obtain the structure information and HF-features that are more robust to light effects, we adaptively fuse the RGB color channels of the input night image by applying: $I_{\text{gray}}(\text{x}) = \sum_c \frac{1}{3}(w_c(\text{x})I_c(\text{x}))$ where $c \in (r, g, b)$ is a color channel, x is a pixel location, and the input image $\mathbf{I} = \{I_r, I_g, I_b\}$. The weight map for each color channel of the night image $I_c(\text{x})$ is computed by $w_c(\text{x}) = \exp\left(\frac{-(I_c(\text{x})-0.5)^2}{2\sigma^2}\right)$. Note that the range of $I_c(\text{x})$ is [0,1], thus 0.5 is the median of the intensity range. Our weight has a low value if a pixel in a color channel is either low (under-exposed) or high (e.g., a light-effects pixel). We define $\sigma = 0.2$, which measures how well-exposed a pixel is. This makes the resulting grayscale image I_{gray} less affected by light effects, as can be observed in Fig. 5 and Fig. 6.

Having obtained I_{gray}, we define our loss as follows:

$$\mathcal{L}_{\text{gray-feat}} = \big|\phi_{\text{HF}}(\mathbf{J}_{\text{refine}}) - \phi_{\text{HF}}(\mathbf{I}_{\text{gray}})\big|_1 + \big|\phi^l_{\text{VGG}}(\mathbf{J}_{\text{refine}}) - \phi^l_{\text{VGG}}(\mathbf{I}_{\text{gray}})\big|_1,$$

where $\mathbf{I}_{\text{gray}} = \{I_{\text{gray}}, I_{\text{gray}}, I_{\text{gray}}\}$. $\phi^l_{\text{VGG}}(.)$ represents the feature maps extracted from the lth layer of the VGG16 network (we set $l = 15$ in our experiments).

Table 1. User study evaluation on the real night data, our method obtained the highest mean and lowest standard deviation (the max score is 7), showing our method is realistic, light-effects (L.E.) suppressed, and has good visibility.

Three Aspects	EG [15]	Afifi [1]	Yan [38]	Zhang [44]	Li [23]	Sharma [32]	Ours
1.Realism↑	3.3 ± 1.5	5.5 ± 1.3	3.7 ± 2.0	3.5 ± 1.6	3.1 ± 1.8	2.8 ± 1.5	$\mathbf{6.1 \pm 0.8}$
2.L.E. Supp.↑	1.7 ± 0.8	3.1 ± 1.3	4.6 ± 1.4	3.9 ± 1.1	5.2 ± 1.2	3.0 ± 1.5	$\mathbf{6.6 \pm 0.7}$
3.Visibility↑	3.1 ± 1.6	4.2 ± 1.5	4.7 ± 1.5	3.7 ± 1.1	3.8 ± 1.5	3.0 ± 1.4	$\mathbf{6.4 \pm 0.7}$

$\phi_{\mathrm{HF}}(.)$ represent the high-frequency feature maps obtained from the guided filter. We concatenate these HF layers to get $\phi_{\mathrm{HF}}(\mathbf{I}_{\mathrm{gray}})$. We use these features to better preserve the HF information in the generated refined background image $\mathbf{J}_{\mathrm{refine}}$. Fig. 5 shows our adaptive fusion scheme to obtain $\mathbf{I}_{\mathrm{gray}}$ from which we compute HF-features and VGG-features. Fig. 6 shows that with our loss in place, the VGG and HF features of $\mathbf{I}_{\mathrm{gray}}$ preserve the structural information.

Adversarial and Identity Losses Our adversarial loss for the generator and discriminator ϕ_{dis} uses its standard definition [12, 26]:

$$\mathcal{L}_{\mathrm{adv}} = \mathbb{E}\big[\log\big(\phi_{\mathrm{dis}}(\mathbf{J}_{\mathrm{ef}})\big)\big] + \mathbb{E}\big[\log\big(1 - \phi_{\mathrm{dis}}(\mathbf{J}_{\mathrm{refine}})\big)\big]. \tag{7}$$

While our light-effects suppression network is designed to refine $\mathbf{J}_{\mathrm{init}}$ by suppressing any remaining light effects, we also encourage it to output the same light-effects-free image when the input has no light-effects $\mathbf{J}_{\mathrm{ef}}$. We achieve this by using the following identity loss function [51]:

$$\mathcal{L}_{\mathrm{iden}} = \mathbb{E}\big[\|\phi_{\mathrm{gen}}(\mathbf{J}_{\mathrm{ef}}) - \mathbf{J}_{\mathrm{ef}}\|_1\big]. \tag{8}$$

We multiply each loss function with its respective weight, we adjust $\lambda_{\mathrm{gray\text{-}feat}} = 1$, $\lambda_{\mathrm{atten}} = 0.5$ with the same scale, and employ the weights of $\lambda_{\mathrm{adv}} = 1$ and $\lambda_{\mathrm{iden}} = 5$ from [51]. The HF layers use smoothing kernels K, with size given by $k = 2^i$, $i = 2, 3, 4, ...$, the regularization $\epsilon = 0.04, 0.08$.

4 Experimental Results

Light-Effects Suppression on Night Data The real night images used in our experiment are downloaded from the Internet and collected by ourselves. We use these images for our unpaired training since collecting the corresponding light-effects-free ground truth images is difficult.

For the user study, we randomly selected 210 outputs (30 per method, seven methods) and presented them to the 12 participants in random order. We asked them to rank these methods from unrealistic (1) to realistic (7); light effects still present (1) to suppressed (7); poor visibility (1) to good visibility (7). Table 1 shows the user study results. Table 2 shows the quantitative results on the night data, where our method has the highest PSNR and SSIM scores.

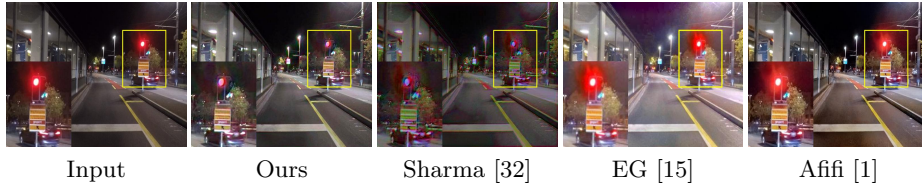| Input | Ours | Sharma [32] | EG [15] | Afifi [1] |

Fig. 7. Comparing light-effects suppression and dark regions enhancement results on the real night images.

Table 2. Quantitative light-effects suppression comparison on the night data. In the table, UL = unsupervised learning, SL = supervised learning, SSL = semi-supervised learning, ZSL = zero-shot learning, Opti = optimization method.

Learning	-	UL	ZSL	SL	SL	SSL	Opti	Opti	SSL	UL
Datasets	Metrics	EG [15]	ZD+ [19]	RN [7]	Afifi [1]	Yan [38]	Zhang [44]	Li [23]	Sharma [32]	Ours
GTA5 [38]	PSNR↑	10.94	21.13	7.79	15.47	26.99	20.92	21.02	8.14	**29.79**
	SSIM↑	0.31	0.68	0.23	0.53	0.85	0.65	0.64	0.29	**0.88**
Syn-light-effects [27]	PSNR↑	7.38	7.84	6.39	11.31	14.88	16.30	14.66	14.00	**16.95**
	SSIM↑	0.17	0.20	0.16	0.35	0.23	0.38	0.37	0.37	**0.39**

Fig. 7 shows the qualitative results on real night images, which demonstrate the superiority of our results compared to the baseline methods. Fig. 8 shows the evaluation on the Dark Zurich [30] dataset. As can be observed, the light-effects suppression baseline [32] suffers from hallucination/artefacts and cannot handle white light effects. In the supplementary material, we show the results of night dehazing baselines [38, 44, 23], which are too dark since they are not designed to enhance dark regions; while low-light image enhancement baselines [15, 1, 19, 7] wrongly intensify light effects, and thus degrade the visibility of the images.

Input   Ours   Sharma [32]   EG [15]   Afifi [1]

Fig. 8. Comparing light-effects suppression and dark regions enhancement results on the real night image from Dark Zurich [30] dataset.

Table 3. Quantitative comparisons on the LOL-test dataset [7].

Learning	Method	LOL-test			
		MSE($\times 10^3$)↓	PSNR↑	SSIM↑	LPIPS↓
Opti	LIME [14]	-	16.760	0.560	0.350
SL	RetinexNet [7]	1.651	16.774	0.462	0.474
	KinD++ [47]	1.298	17.752	0.760	**0.198**
	Afifi [1]	4.520	15.300	0.560	0.392
	RUAS [24]	3.920	18.230	0.720	0.350
ZSL	ZeroDCE [13]	3.282	14.861	0.589	0.335
SSL	DRBN [40]	2.359	15.125	0.472	0.316
UL	EnlightenGAN [15]	1.998	17.483	0.677	0.322
SSL	Sharma [32]	3.350	16.880	0.670	0.315
UL	Ours	**1.070**	**21.521**	**0.763**	0.235

Table 4. Quantitative comparisons on the *LOL-Real* dataset [42].

Learning	NA	Opti	Opti	Opti	ZSL	ZSL	ZSL	ZSL	SL
Method	Input	JED [29]	RRM [21]	SRIE [9]	RDIP [48]	MIRNet [43]	RRDNet [50]	ZD [13]	RUAS [24]
PSNR↑	9.72	17.33	17.34	17.34	11.43	12.67	14.85	20.54	15.33
SSIM↑	0.18	0.66	0.68	0.68	0.36	0.41	0.56	0.78	0.52
Learning	SL	SL	SL	SL	SL	SSL	UL	SSL	UL
Method	LLNet [25]	RN [7]	DUPE [34]	SICE [6]	Afifi [1]	DRBN [41]	EG [15]	Sharma [32]	Ours
PSNR↑	17.56	15.47	13.27	19.40	16.38	19.66	18.23	18.34	**25.53**
SSIM↑	0.54	0.56	0.45	0.69	0.53	0.76	0.61	0.64	**0.88**

Low-Light Enhancement Besides night light-effects suppression, our method can boost the brightness of low light images with no light effects, by simply setting the light-effects layer to $\mathbf{G}_0$, which has no light-effects. For a fair comparison, we compare low-light boosting with image enhancement methods without considering the presence of light effects.

We adopt the LOL dataset [7][6], 485 training and 15 testing images, respectively. Table 3 shows quantitative results, where our method achieves better performance compared with the baseline methods in terms of PSNR, SSIM, Mean Square Error (MSE) and Learned Perceptual Image Patch Similarity (LPIPS) [45]. We evaluate on *LOL-Real* [42][7], 100 testing images with more diversified scenes. We train our method on the LOL dataset and test on the

---

[6] The LOL dataset link: https://daooshee.github.io/BMVC2018website/
[7] *LOL-Real* dataset link: https://github.com/flyywh/CVPR-2020-Semi-Low-Light/

| Input | Ground Truth | Ours | Sharma [32] | EG [15] |
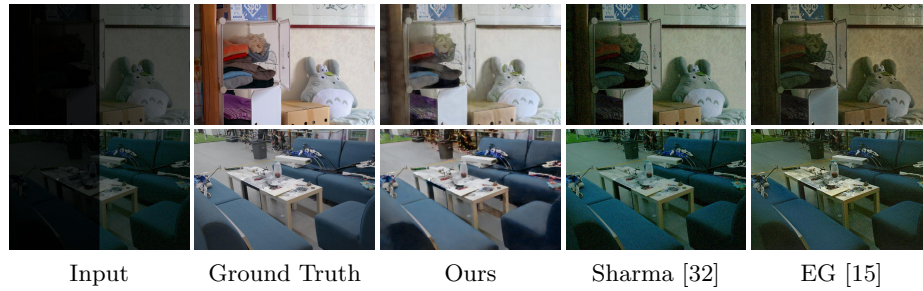
Fig. 9. Low-light enhancement results on the LOL-test [7], *LOL-Real* [42] datasets.

Table 5. Summary of comparisons between our method and existing night image enhancement methods. Our method can suppress light effects (including white light effects), preserve light source (L.S.) details, and boost dark regions simultaneously.

Learning	Methods	Light Effects (L.E.) Suppression			Dark Regions Boosting
		Normal L.E.	White L.E.	Details in L.S.	
UL	Ours	✓	✓	✓	✓
SSL	Sharma and Tan [32]	✓	✕	✕	✓
Opti;SSL	Night Dehazing [23, 44, 38]	✓	✓	✕	✕
SL;ZSL;UL	Low-light Enhancement [1, 19, 15]		✕		✓

LOL-Real test-split. The results are shown in Table 4 and Fig. 9, showing the generality of our method. Our method achieves better performance compared with the baseline methods in terms of PSNR, SSIM.

Baselines As shown in Table 5, there is only one algorithm, i.e., Sharma and Tan [32] that suppresses night light effects and boosts the dark regions simultaneously. Yet, the method cannot handle white light effects and suffers from hallucination/artefacts. Night dehazing methods can suppress glow, but are suboptimal to enhancing low-light regions. Low-light image enhancement methods are not designed to suppress night light effects and enhance low light regions simultaneously.

Nevertheless for comprehensive comparisons, besides comparing with [32], we also compare our method with the state-of-the-art single-image low-light image enhancement methods: EnlightenGAN [15], Afifi et al. [1], etc. and the night dehazing methods: Yan et al. [38], Zhang et al. [44], Li et al. [23], etc. The codes of all the baseline methods are obtained from the authors. More baseline results are provided in the supplementary material.

Joint Light-Effects Suppression and Dark Region Boosting As shown in Fig. 10, jointly suppressing light-effects and then boosting dark regions are more effective than any other possibilities (namely, (b) the light-effects suppression alone, (c) light-effects suppression followed by boosting without jointly training them, (d) boosting alone, and (e) boosting followed by light effects suppression without joint training). If we suppress light effects first, then boost the intensity

| (a) | (b) | (c) | (d) | (e) | (f) |
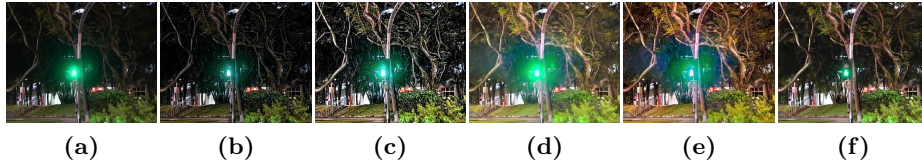
Fig. 10. Experiments on the effectiveness of joint light-effects suppression and dark regions boosting: (a) input, (b) light-effects suppression, (c) light-effects suppression followed by boosting without joint training, (d) boosting, (e) boosting followed by light-effects suppression without joint training, (f) our joint training light-effects suppression and boosting.

| Input | w/o Trans. | w/o Decomp. | Output |
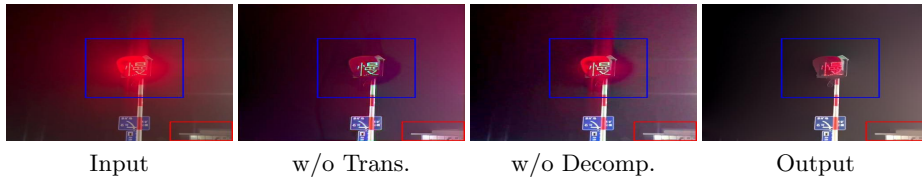
Fig. 11. Ablation studies on the framework. 'w/o Trans' denotes our method without light-effects suppression. 'w/o Decomp.' denotes our method without layer decomposition. We can observe that our framework is important for night image enhancement.

without the joint training, as shown in Fig. 10c, artefacts and remaining light effects are also enhanced. If we boost the intensity first, then suppress light effects without joint training, as shown in Fig. 10e, light effects cannot be effectively suppressed since the amplified light effects cause information and detail loss.

Ablation Studies Fig. 11, Fig. 12 and Fig. 13 show the effectiveness of our framework, light-effects layer guidance and structure and HF-features consistency losses used in our method, which clearly show that all the components are important for better performance.

Decomposition + Suppression To show the effectiveness of our model-based unsupervised decomposition, we train our network without the decomposition module. We directly input the night images to the light-effects suppression network, thus there is no light-effects layer guidance and initial background results. Similarly, to show the effectiveness of our unsupervised light-effects suppression, we assume the initial background image $\mathbf{J}_{\mathrm{init}}$ generated by the decomposition part is the final result without any refinement. Our final results are more effective in suppressing light effects and more natural in recovering the background.

Light-Effects Layer Guidance We compare the results by our method with and without light-effects layer guidance. Instead of input $(\mathbf{G}, \mathbf{J}_{\mathrm{init}})$, we input $(\mathbf{G}_0, \mathbf{J}_{\mathrm{init}})$ to the light-effects suppression network. That means there is no light-effects layer $\mathbf{G}$, we concatenate the initially estimated background scene with all

| | | | |
| Input | Light-Effects Layer | w/o **G** guidance | w/ **G** guidance |

Fig. 12. Ablation studies on light-effects layer guidance, with light-effects layer **G** guidance, our light-effects suppression network can focus on light-effects regions, separate light effects more properly.

| Input | w/o $\mathcal{L}_{\text{gray-feat}}$ | w/ $\mathcal{L}_{\text{gray-feat}}$ | Input | w/o $\mathcal{L}_{\text{gray-feat}}$ | w/ $\mathcal{L}_{\text{gray-feat}}$ |

Fig. 13. Ablation studies on the structure and HF-features consistency losses $\mathcal{L}_{\text{gray-feat}}$, with $\mathcal{L}_{\text{gray-feat}}$, our method suppresses artefacts, and preserves details.

zero maps $\mathbf{G}_0$. Figs. 2-4 show the results of the light-effects layer. Fig. 12 shows with light-effects layer guidance, our method can distinguish light-effects regions from background regions, focus on light-effects regions and properly suppress light effects (including white and multi-color light effects).

Structure and HF-features Consistency Structure and HF-features consistency losses can suppress artefacts and restore missing details. Fig. 13 compares the results by our method with and without this loss.

5   Conclusion

In this paper, we have proposed a method to suppress light effects, and at the same time, boost the intensity of dark regions, from a single night image. To achieve our goal, we cast the problem of light-effects suppression as an unsupervised decomposition problem. We proposed an integrated network consisting of layer decomposition and light-effects suppression networks. Our experiments show that our method outperforms the state-of-the-art visibility enhancement and light effects suppression methods.

Acknowledgment

References

1. Afifi, M., Derpanis, K.G., Ommer, B., Brown, M.S.: Learning multi-scale photo exposure correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9157–9167 (2021)
2. Ancuti, C., Ancuti, C.O., De Vleeschouwer, C., Bovik, A.C.: Night-time dehazing by fusion. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 2256–2260. IEEE (2016)
3. Ancuti, C., Ancuti, C.O., De Vleeschouwer, C., Bovik, A.C.: Day and night-time dehazing by local airlight estimation. IEEE Transactions on Image Processing **29**, 6264–6275 (2020)
4. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Transactions on Graphics (TOG) **33**(4), 1–12 (2014)
5. Buchsbaum, G.: A spatial processor model for object colour perception. Journal of the Franklin institute **310**(1), 1–26 (1980)
6. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing **27**(4), 2049–2062 (2018)
7. Chen Wei, Wenjing Wang, W.Y.J.L.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference. British Machine Vision Association (2018)
8. Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y.: Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE International Conference on Multimedia and Expo. pp. 1–6. IEEE (2011)
9. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2782–2790 (2016)
10. Gandelsman, Y., Shocher, A., Irani, M.: " double-dip": Unsupervised image decomposition via coupled deep-image-priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11026–11035 (2019)
11. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: Advances in Neural Information Processing Systems 24. pp. 765–773. Curran Associates, Inc., Red Hook, NY, USA (2011)
12. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. arXiv preprint arXiv:1406.2661 (2014)
13. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1780–1789 (2020)
14. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing **26**(2), 982–993 (2016)
15. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing **30**, 2340–2349 (2021)
16. Jin, Y., Sharma, A., Tan, R.T.: Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5027–5036 (2021)

17. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
18. Kim, J., Kim, M., Kang, H., Lee, K.: U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. arXiv preprint arXiv:1907.10830 (2019)
19. Li, C., Guo, C., Chen, C.L.: Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
20. Li, C., Guo, C., Han, L.H., Jiang, J., Cheng, M.M., Gu, J., Loy, C.C.: Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis & Machine Intelligence (01), 1–1 (2021)
21. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing **27**(6), 2828–2841 (2018)
22. Li, Y., Brown, M.S.: Single image layer separation using relative smoothness. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2752–2759 (2014)
23. Li, Y., Tan, R.T., Brown, M.S.: Nighttime haze removal with glow and multiple light colors. In: Proceedings of the IEEE international conference on computer vision. pp. 226–234 (2015)
24. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10561–10570 (2021)
25. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61**, 650–662 (2017)
26. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2794–2802 (2017)
27. Metari, S., Deschenes, F.: A new convolution kernel for atmospheric point spread function applied to computer vision. In: 2007 IEEE 11th international conference on computer vision. pp. 1–8. IEEE (2007)
28. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing **39**(3), 355–368 (1987)
29. Ren, X., Li, M., Cheng, W.H., Liu, J.: Joint enhancement and denoising method via sequential decomposition. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–5. IEEE (2018)
30. Sakaridis, C., Dai, D., Gool, L.V.: Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7374–7383 (2019)
31. Sharma, A., Cheong, L.F., Heng, L., Tan, R.T.: Nighttime stereo depth estimation using joint translation-stereo learning: Light effects and uninformative regions. In: 2020 International Conference on 3D Vision (3DV). pp. 23–31. IEEE (2020)
32. Sharma, A., Tan, R.T.: Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11977–11986 (2021)

33. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6849–6857 (2019)
34. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
35. Wu, H., Zheng, S., Zhang, J., Huang, K.: Fast end-to-end trainable guided filter. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1838–1847 (2018)
36. Wu, Y., He, Q., Xue, T., Garg, R., Chen, J., Veeraraghavan, A., Barron, J.T.: How to train neural networks for flare removal. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2239–2247 (2021)
37. Yan, W., Sharma, A., Tan, R.T.: Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13259–13268 (2020)
38. Yan, W., Tan, R.T., Dai, D.: Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In: European Conference on Computer Vision. pp. 473–488. Springer (2020)
39. Yan, W., Tan, R.T., Yang, W., Dai, D.: Self-aligned video deraining with transmission-depth consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11966–11976 (2021)
40. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: 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. pp. 3063–3072 (2020)
41. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. IEEE Transactions on Image Processing **30**, 3461–3473 (2021)
42. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021)
43. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV (2020)
44. Zhang, J., Cao, Y., Fang, S., Kang, Y., Wen Chen, C.: Fast haze removal for nighttime image using maximum reflectance prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7418–7426 (2017)
45. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: 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 (2018)
46. Zhang, X., Ng, R., Chen, Q.: Single image reflection separation with perceptual losses. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4786–4794 (2018)
47. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. International Journal of Computer Vision **129**(4), 1013–1037 (2021)
48. Zhao, Z., Xiong, B., Wang, L., Ou, Q., Yu, L., Kuang, F.: Retinexdip: A unified deep framework for low-light image enhancement. IEEE Transactions on Circuits and Systems for Video Technology (2021)

49. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)
50. Zhu, A., Zhang, L., Shen, Y., Ma, Y., Zhao, S., Zhou, Y.: Zero-shot restoration of underexposed images via robust retinex decomposition. In: 2020 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2020)
51. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)