Nighttime Visibility Enhancement by Increasing the Dynamic Range and Suppression of Light Effects

Aashish Sharma[1], and Robby T. Tan[1,2]
[1]National University of Singapore, [2]Yale-NUS College
aashish.sharma@u.nus.edu, robby.tan@{nus,yale-nus}.edu.sg

Abstract

Most existing nighttime visibility enhancement methods focus on low light. Night images, however, do not only suffer from low light, but also from man-made light effects such as glow, glare, floodlight, etc. Hence, when the existing nighttime visibility enhancement methods are applied to these images, they intensify the effects, degrading the visibility even further. High dynamic range (HDR) imaging methods can address the low light and over-exposed regions, however they cannot remove the light effects, and thus cannot enhance the visibility in the affected regions. In this paper, given a single nighttime image as input, our goal is to enhance its visibility by increasing the dynamic range of the intensity, and thus can boost the intensity of the low light regions, and at the same time, suppress the light effects (glow, glare) simultaneously. First, we use a network to estimate the camera response function (CRF) from the input image to linearise the image. Second, we decompose the linearised image into low-frequency (LF) and high-frequency (HF) feature maps that are processed separately through two networks for light effects suppression and noise removal respectively. Third, we use a network to increase the dynamic range of the processed LF feature maps, which are then combined with the processed HF feature maps to generate the final output that has increased dynamic range and suppressed light effects. Our experiments show the effectiveness of our method in comparison with the state-of-the-art nighttime visibility enhancement methods.

1. Introduction

Due to varying illumination and multiple man-made light sources, night images not only contains low-light regions but also glow, glare, floodlight, etc., which can severely degrade the visibility of the images. This degradation poses challenges for many vision algorithms when applied to nighttime conditions. Hence, enhancing the visibility of night images by boosting the intensity of low light regions, yet at the same time suppressing the night light effects (glow, glare) is an important task.

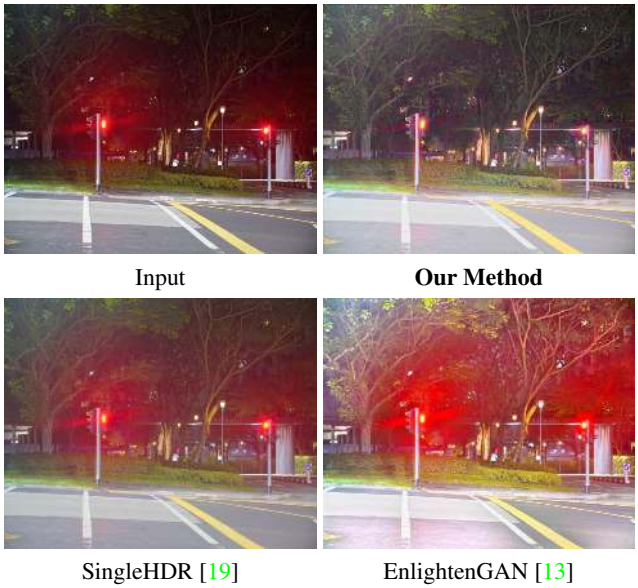


Figure 1. For the input nighttime image with glow/glare light effects, existing visibility enhancement [13] and HDR [19] imaging methods cannot handle the light effects and incorrectly intensify them. In contrast, our method suppresses the light effects and generates better visibility enhancement results.

Existing nighttime visibility enhancement methods [11, 3, 13, 26, 10] assume that the input image is under-exposed and has low-light regions. Hence, to improve the visibility, these methods perform intensity boosting and denoising. Night images, however, also suffer from over-exposed regions as well as glow, glare, etc. which get intensified upon intensity boosting, and this further degrades the visibility. Fig. 1 shows an example. As can be observed, the existing visibility enhancement method EnlightenGAN [13] incorrectly intensifies the glow/glare light effects.

HDR imaging methods [19, 6], to some extent, can improve the visibility of night images. They take a single

†This work is supported by MOE2019-T2-1-130.

image as input and generates an output image that has a higher dynamic range, which provides better visibility for the under-exposed and over-exposed regions of the image. Unfortunately, these methods also cannot suppress glow, glare and floodlight, and thus cannot enhance the visibility of the scenes behind these light effects (see Fig. 1).

In this paper, given a single night image as input, our goal is to improve its visibility by simultaneously increasing the dynamic range (to deal with low-light and over-exposed region) and suppressing the light effects (glow, glare, etc.). To achieve the goal, we propose a semi-supervised network. We use paired images (with HDR ground-truths) to train our network to increase the dynamic range, and unpaired images (without ground-truths) to train our network to suppress the light effects. We first estimate the inverse CRF of the input night image using a linearisation network. Unlike methods [16, 19], which use fully-supervised training based on synthesized data, our method uses semi-supervised training using both synthesized (with CRF ground-truths) and real data (without CRF ground-truths), which provides better generalization capability to our method.

Having obtained the linearised image, we decompose it into low-frequency (LF) and high-frequency (HF) feature maps. The LF feature maps are likely to contain the glow/glare light effects (since they are smooth [17]), while the HF feature maps will contain noise, textures, edges, etc. The LF and HF feature maps are processed separately using our two networks to suppress the light effects and remove the noise respectively. The processed LF features maps that contain suppressed light effects are passed to another network to increase their dynamic range. The resulting LF feature maps are then fused with the processed HF feature maps to generate the output image that has both increased dynamic range and suppressed light effects.

In summary, our contributions are as follows:

- We introduce a new method for single-image nighttime visibility enhancement such that the enhanced output image has both increased dynamic range and suppressed glow/glare light effects. To our knowledge, our method is the first method to address this problem.
- We train our method using semi-supervised learning. We use paired data for learning to increase the dynamic range; and, unpaired data for learning the light effect suppression. We use priors such as the smoothness prior for the light effects (glow, glare, floodlight) in designing our unsupervised losses. Our CRF estimation is also semi-supervised, where the unsupervised losses are designed based on the monotonicity constraint of CRFs and the linearisation constraint obeyed by edge-based pixels in the irradiance domain.

Our experiments show that our method outperforms the state-of-the-art single-image nighttime visibility enhancement and HDR imaging methods.

2. Related Work

Several methods are designed for single-image visibility enhancement [22, 2, 11, 3, 13, 26, 10]. The visibility enhancement is performed through intensity boosting and denoising, where denoising is either part of the methods [3, 13, 2] or carried out as a post-processing step [11]. Some methods even ignore the presence of CRF or assume it to be linear, as a result of which the physical properties of the image are not preserved during enhancement and the enhanced image have colour-distortions [11, 3, 13, 26, 10]. The method [2] designed for extreme low-light enhancement can preserve the physical properties during enhancement. However, it accepts a RAW image as input (not RGB image) which limits its practicality since RAW images are not available in many practical scenarios. All these methods are not well-suited for enhancing night images since night images contain over-exposed regions and man-made light effects (glow, glare, floodlight) which also get intensified upon intensity boosting degrading the visibility of the images even further.

Another approach to visibility enhancement is to increase the dynamic range of the intensity by using HDR imaging methods. Most commonly, HDR images are created by estimating the inverse CRF from a stack of aligned bracketed exposure images. The images are linearised using the estimated inverse CRF and fused to create the HDR image [5, 20]. Some recent methods have relaxed the requirement of using aligned images [14, 28]. To improve the practicality, several methods using a single image are proposed [19, 6]. The methods use fully-supervised learning to directly learn a mapping from the input image to the HDR output image. While the method [6] assumes a fixed CRF which is not applicable for different camera images, the method [19] propose to estimate the CRF in the HDR imaging process. However, all these methods also cannot handle the light effects, and thus cannot enhance the visibility in the affected regions.

In contrast to the aforementioned approaches, our method is designed to handle the light effects problem and increase the dynamic range simultaneously. Moreover, unlike the existing methods such as [19] which use fully-supervised training, our method uses semi-supervised training using both unlabelled and labelled data, which provides better generalization capability to our method.

3. Proposed Method

Fig. 2 shows our pipeline. Given an input image, our linearisation-Net estimates the CRF, which enables us to linearise the input image. We decompose the linearised image into low-frequency (LF) and high-frequency (HF) feature maps using the decomposition model [27]. The LF feature maps contain the light effects (glow, glare, floodlight)

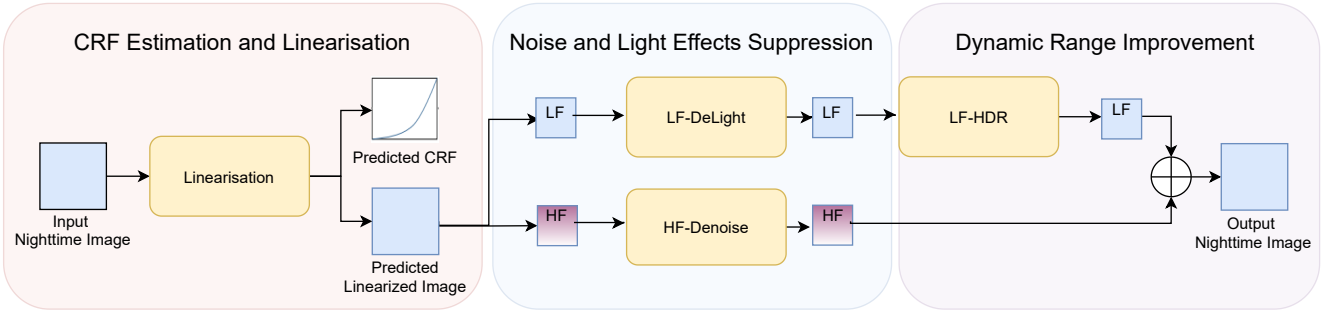


Figure 2. The overall architecture of our proposed method. From the input nighttime image, we estimate the inverse CRF using the Linearisation network. The inverse CRF is used to linearise the image. The linearised image is decomposed into low-frequency (LF) and high-frequency (HF) feature maps, which are processed through LF-DeLight and HF-Denoise networks for suppressing the light effects and noise respectively. Finally, we pass the processed LF feature maps to the LF-HDR network to increase the dynamic range, and the resulting feature maps are combined with the processed HF feature maps to generate the enhanced output nighttime image.

and the HF feature maps contain noise, textures, edges, etc. We process the LF feature maps using a network we call as LF-DeLight to suppress the light effects in the LF feature maps. We process the HF features maps using a network we call as HF-DeNoise to remove the noise from the HF features maps. The processed LF features maps that have suppressed light effects are passed to a network we call as LF-HDR to increase their dynamic range. The resulting LF feature maps are then combined with the processed HF feature maps to generate the output image that has increased dynamic range and suppressed light effects.

We assume that we have M nighttime images represented by $\mathbb{X} = \{\mathbf{X}_1, ..., \mathbf{X}_M\}$ taken with varying exposure levels, and also their corresponding HDR ground-truth images represented by $\mathbb{X}^{gt} = \{\mathbf{X}_1^{gt}, ..., \mathbf{X}_M^{gt}\}$. We further assume that we have N nighttime images that do not have the paired HDR ground-truths. We represent these images by $\mathbb{Y} = \{\mathbf{Y}_1, ..., \mathbf{Y}_N\}$. The images in $\mathbb{X}$ have negligible light effects, while the images in $\mathbb{Y}$ have prominent glow, glare, or floodlight light effects.

For all the images in $\mathbb{X} \cup \mathbb{Y}$, we do not have the corresponding CRF ground-truths. Hence, we take another set of P images (which include indoor, day and night images) represented by $\mathbb{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_P\}$. Samples drawn from the sets $\mathbb{X}, \mathbb{X}^{gt}, \mathbb{Y}, \mathbb{Z}$ are represented by $\mathbf{X}, \mathbf{X}^{gt}, \mathbf{Y}, \mathbf{Z}$ respectively. While the images in $\mathbb{X} \cup \mathbb{Y}$ are RGB (JPEG) images, the images in $\mathbb{Z}$ are RAW images which allows us to synthesize our own RGB images with known CRF ground-truths. Specifically, we take the 201 CRFs from the Database of Response Functions (DoRF) provided by [9], which we represent by $\mathbb{F} = \{\mathbf{f}_1^{gt}, ..., \mathbf{f}_{201}^{gt}\}$; and, a CRF sampled from $\mathbb{F}$ is represented by $\mathbf{f}^{gt}$. Using the camera imaging pipeline presented in [15], we synthesize new RGB images by: $\mathbf{Z}' = \mathbf{f}^{gt}(\mathbf{t}(\mathbf{Z}))$, where $\mathbf{t}$ represents the color transformation (e.g. white-balance) operator, which is read from the RAW metadata of $\mathbf{Z}$. We further add a small amount of random Gaussian noise and JPEG compression to $\mathbf{Z}'$ to make it more closely represent a real night image.

Our linearisation-Net, LF-DeLight, HF-Denoise and LF-HDR networks are represented by ϕ_{LIN}, $\phi_{\text{LF-DL}}$. $\phi_{\text{HF-DN}}$ and $\phi_{\text{LF-HDR}}$ respectively. We use the ResNet-18 [12] architecture as the backbone of the ϕ_{LIN} network, while for all the other networks, we adopt an encoder-decoder architecture with skip connections [23]. We train our method in two stages: (1) Main training, where we use paired data with supervised losses to train our method, and (2) Test-time training, where we use unsupervised losses to finetune our method on the test image taken from unpaired data.

3.1. Supervised Training

Learning CRF Estimation For training the linearisation-Net, ϕ_{LIN}, given $\mathbf{Z}'$ as input and $\mathbf{g}^{gt} = \mathbf{f}^{gt\,-1}$ as the ground-truth inverse CRF, we use the following loss:

$$\mathcal{L}_{\text{mse}} = \|\hat{\mathbf{g}} - \mathbf{g}^{gt}\|_2, \tag{1}$$

where $\hat{\mathbf{g}} = \mathbf{g_0} + \sum_{i=1}^{11} \mathbf{h}_i \mathbf{c}_i$ is the predicted inverse CRF using the basis-function model [9]. $\mathbf{g_0}$ and $\{\mathbf{h}_1, ..., \mathbf{h}_{11}\}$ are the base inverse CRF and set of 11 basis functions taken from the model, respectively. $\{\mathbf{c}_1, ..., \mathbf{c}_{11}\}$ are the 11 coefficients of the basis functions generated by the network ϕ_{LIN}, i.e. $\{\mathbf{c}_1, ..., \mathbf{c}_{11}\} = \phi_{\text{LIN}}(\mathbf{Z}')$. Since we also have the ground-truth linearised image, $\mathbf{g}^{gt}(\mathbf{Z}')$, we use the linearisation loss to train the network ϕ_{LIN}:

$$\mathcal{L}_{\text{lin}} = \|\hat{\mathbf{g}}(\mathbf{Z}') - \mathbf{g}^{gt}(\mathbf{Z}')\|_1. \tag{2}$$

Increasing Dynamic Range For training the networks $\phi_{\text{LF-HDR}}$ and $\phi_{\text{HF-HDR}}$, we use the images in $\mathbb{X}$ and their corresponding HDR ground-truth images in $\mathbb{X}^{gt}$. Given $\mathbf{X}$ as input and $\mathbf{X}^{gt}$ as the HDR ground-truth, we first estimate the inverse CRF from the image $\mathbf{X}$ using our linearisation-Net ϕ_{LIN}, and then linearise the image.

Let the linearised image be represented by $\mathbf{L_X}$. Using

the differentiable decomposition model [27], we obtain the low-frequency and high-frequency feature maps from $\mathbf{L_X}$, which we represent by $\mathbf{LF_X} = \{\mathbf{LF}_{\mathbf{X}1}, ..., \mathbf{LF}_{\mathbf{X}K}\}$ and $\mathbf{HF_X} = \{\mathbf{HF}_{\mathbf{X}1}, ..., \mathbf{HF}_{\mathbf{X}K}\}$ respectively. K being the total number of filters used for decomposition. We use the following HDR loss function from [14]:

$$\mathcal{L}_{\text{HDR}} = \left\| \frac{\log(1+\mu\hat{\mathbf{X}})}{\log(1+\mu)} - \frac{\log(1+\mu\mathbf{X}^{gt})}{\log(1+\mu)} \right\|_1, \quad (3)$$

where $\hat{\mathbf{X}} = \frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{HrLF}_{\mathbf{X}k} + \mathbf{DnHF}_{\mathbf{X}k}\right)$ is the predicted HDR image; $\mathbf{HrLF}_{\mathbf{X}k}$ is the output of the $\phi_{\text{LF-HDR}}$ network, i.e., $\mathbf{HrLF}_{\mathbf{X}k} = \phi_{\text{LF-HDR}}(\mathbf{LF}_{\mathbf{X}k})$ and $\mathbf{DnLF}_{\mathbf{X}k}$ is the output of the $\phi_{\text{HF-DN}}$ network, i.e., $\mathbf{DnHF}_{\mathbf{X}k} = \phi_{\text{HF-DN}}(\mathbf{HF}_{\mathbf{X}k})$. μ is a constant parameter set to 10. Note that, the loss $\mathcal{L}_{\text{HDR}}$ also back-propagates to ϕ_{LIN} and jointly trains ϕ_{LIN}, $\phi_{\text{LF-HDR}}$ and $\phi_{\text{HF-DN}}$ networks.

3.2. Unsupervised Test-Time Training

In this stage, we fine-tune our method using unsupervised losses. In the main supervised training stage, the network ϕ_{LIN} was trained using synthesized data, hence, it is important to fine-tune it on real data. The networks $\phi_{\text{LF-HDR}}$ and $\phi_{\text{HD-DN}}$, however, were trained on real data and are not needed to adapt any further. For this reason, we keep them frozen in this stage, and only fine-tune the networks ϕ_{LIN} and $\phi_{\text{LF-DL}}$. Also, different from the main supervised training stage where we use all the images in $\mathbb{X}$ and $\mathbb{Z}$ data sets, in this stage, we perform the unsupervised fine-tuning on the test input image $\mathbf{Y}$ sampled from the data set $\mathbb{Y}$.

Fine-tuning CRF Estimation Since the image $\mathbf{Y}$ has no CRF ground-truth available, we use unsupervised losses to fine-tune our linearisation-Net ϕ_{LIN} on the image $\mathbf{Y}$.

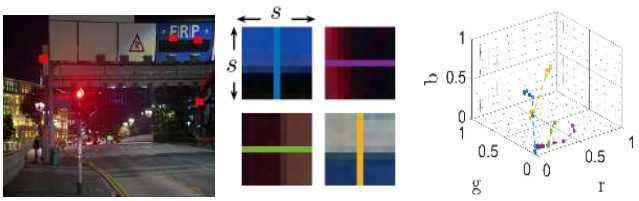
CRFs are known to be monotonically increasing [9], hence we use the monotonicity loss [25]:

$$\mathcal{L}_{\text{mon}} = \sum_{t=0}^{1} H\left(-\frac{\partial\hat{\mathbf{g}}(t)}{\partial t}\right), \quad (4)$$

where we impose that the derivative of predicted inverse CRF $\hat{\mathbf{g}}$, $\partial\hat{\mathbf{g}}/\partial t$, to be positive. $H(\cdot)$ is the Heaviside step function: $H(x) = 1$ when $x \geq 0$, and 0 otherwise, and t is a variable of 1024 equally spaced values in $[0, 1]$.

We also use an unsupervised loss based on the property that edge-based pixels in the irradiance (or linearised) image form linear distributions in the 3D RGB space [18]. Specifically, from the image $\mathbf{Y}$, we extract non-overlapping edge-patches represented by $\{\mathbf{p}_{\mathbf{Y}1}, ..., \mathbf{p}_{\mathbf{Y}E}\}$. E is the number of patches. Each patch is of resolution $S \times S$. From each patch, we obtain S pixel distributions either horizontally or vertically. Fig. 3 shows an example.

The entire set of pixel distributions obtained from all the patches is represented by $\{\mathbf{d}_{\mathbf{Y}11}, ..., \mathbf{d}_{\mathbf{Y}S1}, ..., \mathbf{d}_{\mathbf{Y}E1}, ..., \mathbf{d}_{\mathbf{Y}ES}\}$. For each pixel distribution $\mathbf{d}_{\mathbf{Y}es}$ where $e \in [1, E]$ and $s \in [1, S]$, we first linearise it using the predicted inverse CRF $\hat{\mathbf{g}}$, and then normalize it to have the range [0, 1]. The normalisation operation is important in order to avoid trivial solutions [24]. Representing the linearised-and-normalized distribution by $\mathbf{n}_{\mathbf{Y}es}$, and its minimum and maximum values by $\mathbf{n}^{\text{min}}_{\mathbf{Y}es}$ and $\mathbf{n}^{\text{max}}_{\mathbf{Y}es}$ respectively, we define our linearisation loss for pixel distributions as:



(a) Image **(b)** Patches **(c)** Pixel Distributions

Figure 3. Demonstration of our selection of pixel distributions from edge-based patches. (a) Sample image. (b) Edge-based patches extracted from the image. The coloured lines in the middle of the patches show the direction of the pixel distributions. (c) Pixel distributions taken from the midline in the patches plotted in the RGB space, which are non-linear due to the CRF. Note that, we select the distributions in the direction of maximum variance.

$$\mathcal{L}_{es} = \sum_{i=1}^{S}\left(\frac{|\mathbf{n}^{\text{min}}_{\mathbf{Y}es} - \mathbf{n}^{\text{max}}_{\mathbf{Y}es}| \times |\mathbf{n}^{\text{min}}_{\mathbf{Y}es} - \mathbf{n}^{\text{i}}_{\mathbf{Y}es}|}{|\mathbf{n}^{\text{min}}_{\mathbf{Y}es} - \mathbf{n}^{\text{max}}_{\mathbf{Y}es}|}\right), \quad (5)$$

$$\mathcal{L}_{\text{distlin}} = \sum_{e=1}^{E}\left(\sum_{s=1}^{S}(\mathcal{L}_{es})\right), \quad (6)$$

where $\mathbf{n}^{\text{i}}_{\mathbf{Y}es}$ is a pixel on the distribution $\mathbf{n}_{\mathbf{Y}es}$.

Suppressing Light Effects To suppress the light effects (glow, glare, floodlight), as there are no ground-truths to learn from, we use unsupervised losses to train our method. Following the same steps as mentioned before, we obtain the low-frequency feature maps of the input image $\mathbf{Y}$, which are represented by $\mathbf{LF_Y} = \{\mathbf{LF}_{\mathbf{Y}1}, ..., \mathbf{LF}_{\mathbf{Y}K}\}$. For each feature map $\mathbf{LF}_{\mathbf{Y}k}$, the network $\phi_{\text{LF-DL}}$ generates a map $\mathbf{G}_{\mathbf{Y}k}$ which contains the light effects, i.e. $\mathbf{G}_{\mathbf{Y}k} = \phi_{\text{LF-DL}}(\mathbf{LF}_{\mathbf{Y}k})$. Subtracting $\mathbf{G}_{\mathbf{Y}k}$ from $\mathbf{LF}_{\mathbf{Y}k}$, we obtain the LF feature map that has suppressed light effects.

We represent this feature map by $\mathbf{DeLF}_{\mathbf{Y}k}$. Since this feature map only differs from the input feature map in the regions containing the light affects, we add a reconstruction loss to ensure that the two are not very different:

$$\mathcal{L}_{\text{recon}} = \|\mathbf{DeLF}_{\mathbf{Y}k} - \mathbf{LF}_{\mathbf{Y}k}\|_1. \quad (7)$$

Also, since light effects are smooth [17], we use a smoothness loss to ensure $\mathbf{G}_{\mathbf{Y}k}$ is smooth:

$$\mathcal{L}_{\text{smooth}} = \left\| \left|\partial_x^2(\mathbf{G}_{\mathbf{Y}k})\right| + \left|\partial_y^2(\mathbf{G}_{\mathbf{Y}k})\right| \right\|_1, \quad (8)$$

(a) Input **(b)** LF Map **(c)** Light Effects (l. e.) **(d)** LF Map (w/o l. e.) **(e)** Output (w/o l. e.)

Figure 4. Examples showing the suppression of light effects from our method. (a) Input image. (b) Sample LF feature map obtained from the image. (c) Generated map showing glow/glare/floodlight light effects. (d) Sample LF feature map after subtracting the light effects map. (e) Output image with suppressed light effects. Note that, the output images shown are not with increased dynamic range.

(a) Input **(b)** Glow (prev) **(c)** Output (prev) **(d)** Glow (new) **(e)** Output (new)

Figure 5. (a) Input image with white glow. (b) Glow layer. (c) Output from our method. (d) Output from [19]. (e) Glow layer (new). (f) Output from our method (new). The results shown in (e) and (f) are obtained after adding an additional gradient exclusion constraint [8, 29] to better handle white glow/glare suppression.

where the functions ∂_x^2 and ∂_y^2 compute the 2$^{\text{nd}}$ order horizontal and vertical gradients respectively.

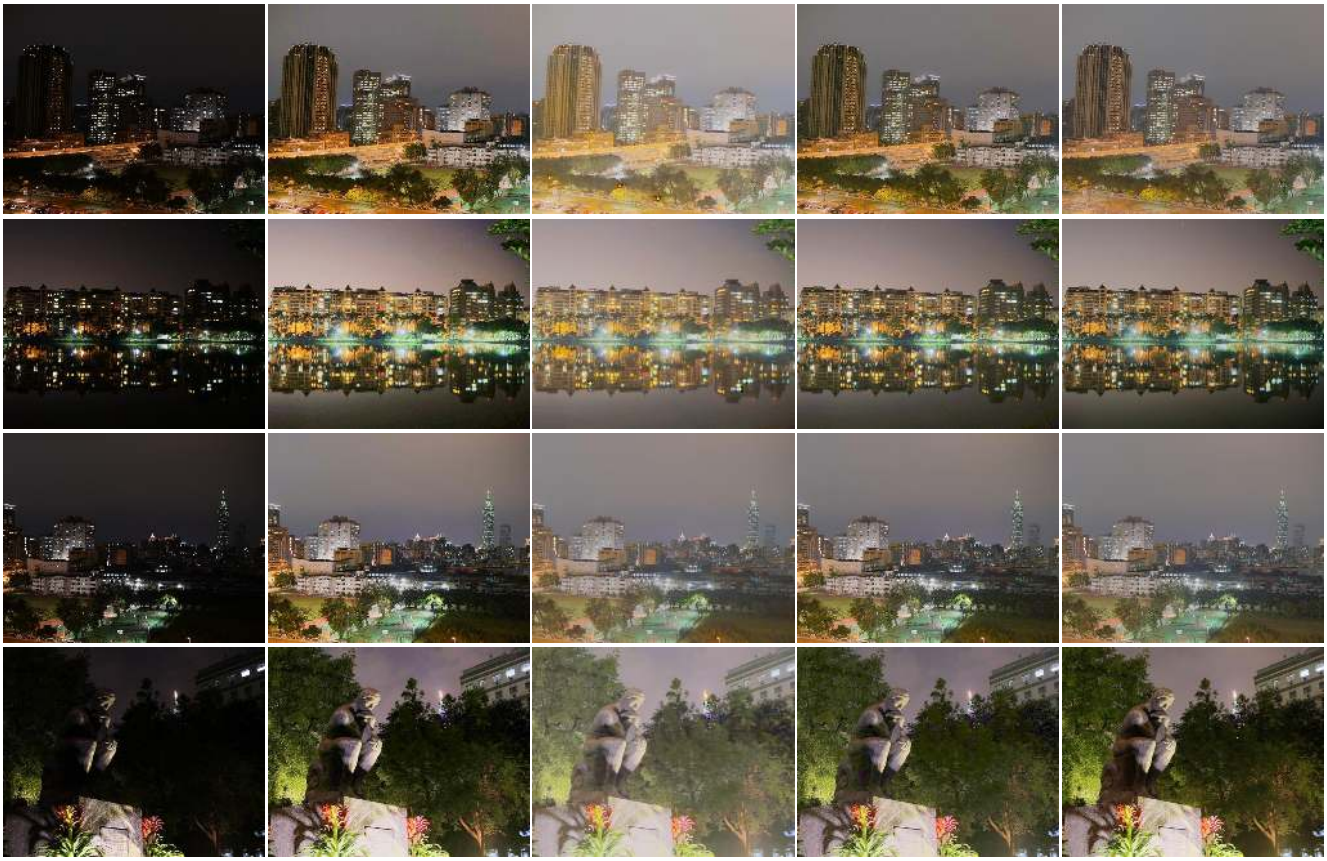
An additional constraint comes from the gray world assumption [1, 17], which encourages the range of the intensity values for the three color channels in $\mathbf{DeLF}_{\mathbf{Y}k}$ to be balanced. We define this loss by:

$$\begin{aligned}\mathcal{L}_{\text{gray}} = &\big(\|\mathbf{DeLF}_{\mathbf{Y}k}(\text{r}) - \mathbf{DeLF}_{\mathbf{Y}k}(\text{g})\|_1\big) + \\ &\big(\|\mathbf{DeLF}_{\mathbf{Y}k}(\text{r}) - \mathbf{DeLF}_{\mathbf{Y}k}(\text{b})\|_1\big) + \\ &\big(\|\mathbf{DeLF}_{\mathbf{Y}k}(\text{g}) - \mathbf{DeLF}_{\mathbf{Y}k}(\text{b})\|_1\big), \end{aligned} \tag{9}$$

where $\mathbf{DeLF}_{\mathbf{Y}k}(\text{r})$, $\mathbf{DeLF}_{\mathbf{Y}k}(\text{g})$ and $\mathbf{DeLF}_{\mathbf{Y}k}(\text{b})$ represent the red, green and blue colour channels of $\mathbf{DeLF}_{\mathbf{Y}k}$, respectively. Fig. 4 shows some examples of light effects suppression obtained from our method. Having obtained $\mathbf{DeLF}_{\mathbf{Y}k}$, we increase their dynamic range by passing it to the network $\phi_{\text{LF-HDR}}$. This gives us the $\mathbf{HrLF}_{\mathbf{Y}k}$, which we fuse with the denoised HF feature map, $\mathbf{DnHF}_{\mathbf{Y}k}$ to obtain the final output image $\hat{\mathbf{Y}}$, i.e. $\hat{\mathbf{Y}} = \frac{1}{K}\sum_{k=1}^{K}\big(\mathbf{HrLF}_{\mathbf{Y}k} + \mathbf{DnHF}_{\mathbf{Y}k}\big)$, which has both increased dynamic range and suppressed light effects.

4. Discussion

The results shown in Fig. 4 show that our unsupervised losses are effective in suppressing glow/glare light effects. However, for white glow/glare (or achormatic glow/glare), our light effects suppression can be improper since input night images with white glow/glare already satisfy constraints such as the gray-world assumption [1, 17]. Fig. 5 shows an example where we can observe that for the night



Input HDRCNN [6] SingleHDR [19] **Our Method** Ground-Truth

Figure 6. Qualitative results on the nighttime images from the HDR-Real [19] dataset. We can observe the better performance of our method compared to the baseline methods.

Table 1. Quantitative results on the nighttime images from the HDR-Real [19] dataset. The performance numbers are reported for 394 test images. 'Masked' denotes the evaluation where we do not evaluate for the uninformative pixels, i.e. pixels under dark noise (intensity < 10) and near saturation (intensity > 240).

	PSNR	SSIM
HDRCNN [6]	19.01	0.8037
SingleHDR [19]	23.01	0.8540
Our Method	**23.51**	**0.8622**
HDRCNN (Masked) [6]	19.36	0.8289
SingleHDR (Masked) [19]	24.19	0.8786
Our Method (Masked)	**24.66**	**0.8816**

image with white glow/glare (see Fig. 5a), our light effect suppression is not proper (see Figs. 5b and 5c).

To address the problem, we can add additional constraints in our method. From [17], we know that the gradient histogram of glow/glare images (irrespective of the glow/glare color) have a short-tail distribution with most values near zero. Hence, from [8, 29], we can add a gradient exclusion constraint that tries to maximize the distance between the output image and the glow layer in the gradient domain. In addition, we can also also constrain the gradients of the output image to be greater than the gradients of the input image. As shown in Fig. 5, adding these new constraints allows us to better deal with white glow/glare (see Figs. 5d and 5e). While the gradient exclusion constraint works properly to deal with achromatic glow/glare, we have not deeply investigated the effects of this constraint, particularly in the correlations with the other constraints. This investigation will be part of our future work.

5. Experiments

For training and evaluation of our dynamic range improvement, we use the HDR-Real dataset [19]. For training our CRF estimation using synthesized data, we take RAW images from the Color-Constancy [4] and SID (long exposure) [2] datasets. For qualitative evaluation of our method on nighttime images with light effects, we collect our own data with glow/glare/floodlight light effects. For all the results shown in the paper, the default weights of supervised

Input | **Our Method** | Zero-DCE [10] | EnlightenGAN [13] | SingleHDR [19]

Figure 7. Qualitative results on the nighttime images with glow/glare/floodlight light effects. We can observe the better performance of our method compared to the baseline methods. Our result not only have the light effects suppressed but also increased dynamic range (e.g., better visibility in the low-light regions) which shows the effectiveness of our method for nighttime visibility enhancement.

losses, $\{\mathcal{L}_{\text{mse}}, \mathcal{L}_{\text{lin}}, \mathcal{L}_{\text{HDR}}\}$, and unsupervised losses, $\{\mathcal{L}_{\text{mon}}, \mathcal{L}_{\text{distlin}}, \mathcal{L}_{\text{recon}}, \mathcal{L}_{\text{smooth}}, \mathcal{L}_{\text{gray}}\}$, are set $\{10.0, 1.0, 1.0\}$ and $\{1.0, 0.1, 1.0, 0.5, 1.0\}$ respectively.

5.1. Evaluation of Dynamic Range Improvement

For the baseline methods for evaluating dynamic range improvement, we take the state-of-the-art single-image HDR imaging methods: SingleHDR [19] and HDR-CNN [6]. Since the outputs of all the methods (including ours) and ground-truths are in HDR, we use tone-mapping [21] to convert them into LDR images for evaluation and visualisation. The qualitative results are shown in Fig. 6 and the quantitative results are shown in Table 1.

From the results, we can observe that our method performs better than the baseline methods, both quantitatively and qualitatively. The results from SingleHDR [19] appear to be washed-out, while the results from HDRCNN [6] are not properly exposed in many regions which could be because of its assumption of a constant CRF for every image.

5.2. Evaluation of Light Effects Suppression

Since we have no ground-truths to quantitatively evaluate our light effects suppression performance, we use qualitative comparisons. For the baseline methods, we take the state-of-the-art single-image visibility enhancement methods Zero-DCE [10] and EnlightenGAN [13]; and, Sin-

Table 2. Comparison of CRF estimation performance. The numbers represents Root Mean Square Error (RMSE) computed between the method's predicted CRFs and the ground-truth CRFs.

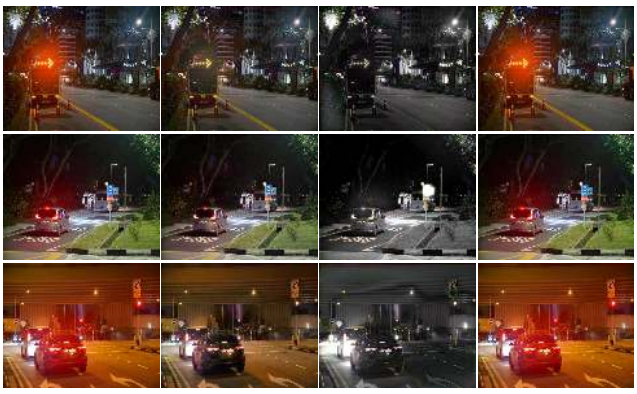
	Mean	Min	Max
HDRCNN [6]	0.1761	0.1461	0.1995
CRFNet [16]	0.1471	0.0177	0.2522
SingleHDR [19]	0.0628	0.0128	0.1369
Our Method	**0.0582**	**0.0107**	**0.1299**

gleHDR [19], which is the state of the art HDR imaging method. Our results are generated following the test-time training optimisation scheme presented in Sec. 3.2. The results are shown in Fig. 7.

From the qualitative results shown in Fig. 7, we can observe that compared to all the baseline visibility enhancement methods, our method performs visibility enhancement that suppresses the light effects and increases the dynamic range of the intensity at the same time. We can see that the low-light regions in the images are better exposed and the glow/glare light effects are suppressed. While the baseline methods also can achieve intensity boosting in the low-light regions, they wrongly intensify the glow/glare light effects, which degrade the visibility of the images even further in the affected regions of the input night images.

6. Ablation Study

Unsupervised CRF estimation To evaluate our CRF estimation performance on real images, we collect 50 nighttime images taken from Sonyα7S III, NikonD40 and NikonD80 cameras. We additionally take 50 well-exposed nighttime images from the long-exposure set of the SID [2] dataset which is created using Sonyα7S II and Fujifilm X-T2 cameras, giving us a total of 100 test images. To get the ground-truth CRFs, we use images containing Macbeth Colour-Checker [18, 9, 25]; and, we further validate the obtained CRFs using the RAW-JPEG pairs which are available for all the test images. We compare our CRF estimation performance with HDRCNN [6], SingleHDR [19] and CRFNet [16]. HDRCNN [6] assumes a constant inverse CRF: $\mathbf{g}(\mathbf{Y}) = \mathbf{Y}^2$, while SingleHDR [19] and CRFNet [16] use fully-supervised training using synthesized data for learning CRF estimation. In comparison, our method uses semi-supervised training using both supervised synthesized data and unsupervised real data for learning CRF estimation. The results are shown in Table 2. Since the baseline methods SingleHDR [19] and CRFNet [16] are trained on synthesized data, their performance is not optimum when tested on real data due to the domain gap that exists between synthesized and real data. In contrast, our method shows better results since our method gets finetuned on real images using unsupervised losses (Sec. 3.2), which provides better generalization capability to our method.



(a) Input **(b)** Output **(c)** w/o $\mathcal{L}_{\text{recon}}$ **(d)** w/o $\mathcal{L}_{\text{gray}}$

Figure 8. (a) Input image. (b) Output image with suppressed light effects. (c) Output image without using the reconstruction loss ($\mathcal{L}_{\text{recon}}$. Eq. (7)). (d) Output image without using the gray world loss ($\mathcal{L}_{\text{gray}}$, Eq. (9)). We can observe that all the unsupervised losses are important for suppressing the light effects properly.

Unsupervised Light Effects Suppression To suppress the light effects (glow, glare, floodlight), as there are no ground-truths to learn from, we use unsupervised losses to train our method (Sec. 3.2 in the main paper). Fig. 8 shows the importance of each unsupervised loss. For the input nighttime images shown in Fig. 8a, the output images using all our unsupervised losses for light effects suppression are shown in Fig. 8b. Fig. 8c shows the output images obtained without using the reconstruction loss (i.e. without $\mathcal{L}_{\text{recon}}$. Eq. (7)). While the images do have suppressed light effects, the colours in the images are not properly recovered and the images appear grayscale. Fig. 8d shows the output images without using the gray world loss (i.e. without $\mathcal{L}_{\text{gray}}$. Eq. (9)). As we can observe, without the gray world loss, the light effects suppression is not proper. Hence, all our unsupervised losses are important for suppressing the light effects effectively as shown in Fig. 8b.

7. Conclusion

We have introduced a single-image nighttime visibility enhancement method. Our idea is to increase the dynamic range of the intensity, and thus can boost the intensity of the low-light regions, and at the same time, suppress the light effects (glow, glare, floodlight). To our knowledge, our method is the first method to have addressed this problem. Our method is based on semi-supervised learning and uses paired data (with ground-truths) and unpaired data (without ground-truths) for learning dynamic range improvement and light effects suppression respectively. Our experiments have confirmed that our method outperforms the state-of-the-art methods qualitatively and quantitatively.

References

[1] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 5

[2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 2, 6, 8

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

[4] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 6

[5] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, pages 1–10. 2008. 2

[6] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017. 1, 2, 6, 7, 8

[7] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6):177–1, 2017.

[8] Yosef Gandelsman, Assaf Shocher, and Michal Irani. " double-dip": Unsupervised image decomposition via coupled deep-image-priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11026–11035, 2019. 5, 6

[9] Michael D Grossberg and Shree K Nayar. Modeling the space of camera response functions. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1272–1282, 2004. 3, 4, 8

[10] 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. 1, 2, 7, 8

[11] 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. 1, 2

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[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. *arXiv preprint arXiv:1906.06972*, 2019. 1, 2, 7, 8

[14] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 2, 4

[15] Seon Joo Kim, Jan-Michael Frahm, and Marc Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 3

[16] Han Li and Pieter Peers. Crf-net: Single image radiometric calibration using cnns. In *Proceedings of the 14th European Conference on Visual Media Production (CVMP 2017)*, pages 1–9, 2017. 2, 8

[17] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *Proceedings of the IEEE international conference on computer vision*, pages 226–234, 2015. 2, 4, 5, 6

[18] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. Radiometric calibration from a single image. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II. IEEE, 2004. 4, 8

[19] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1651–1660, 2020. 1, 2, 5, 6, 7, 8

[20] S Mann and R Picard. On being undigital with digital cameras: extending dynamic range by combining differently exposed pictures [a]. In *IS&Ts 48th Annual Conference*, pages 422–428. Washington, USA: IS&T, 1995. 2

[21] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 267–276, 2002. 7

[22] Yurui Ren, Zhenqiang Ying, Thomas H Li, and Ge Li. Lecarm: Low-light image enhancement using the camera response model. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(4):968–981, 2018. 2

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[24] Aashish Sharma, Robby T Tan, and Loong-Fah Cheong. Single-image camera response function using prediction consistency and gradual refinement. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 4

[25] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 4, 8

[26] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019. 1, 2

[27] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Fast end-to-end trainable guided filter. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1838–1847, 2018. 2, 4

[28] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 117–132, 2018. 2

[29] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4786–4794, 2018. 5, 6