Visibility Constrained Wide-band Illumination Spectrum Design for Seeing-in-the-Dark

Muyao Niu, Zhuoxiao Li, Zhihang Zhong, Yinqiang Zheng*
The University of Tokyo
muyao.niu@gmail.com, {lizhuoxiao@g.ecc, zhong@is.s, yqzheng@ai}.u-tokyo.ac.jp

Abstract

*Seeing-in-the-dark is one of the most important and challenging computer vision tasks due to its wide applications and extreme complexities of in-the-wild scenarios. Existing arts can be mainly divided into two threads: 1) RGB-dependent methods restore information using degraded RGB inputs only (*e.g.*, low-light enhancement), 2) RGB-independent methods translate images captured under auxiliary near-infrared (NIR) illuminants into RGB domain (*e.g.*, NIR2RGB translation). The latter is very attractive since it works in complete darkness and the illuminants are visually friendly to naked eyes, but tends to be unstable due to its intrinsic ambiguities. In this paper, we try to robustify NIR2RGB translation by designing the optimal spectrum of auxiliary illumination in the wide-band VIS-NIR range, while keeping visual friendliness. Our core idea is to quantify the visibility constraint implied by the human vision system and incorporate it into the design pipeline. By modeling the formation process of images in the VIS-NIR range, the optimal multiplexing of a wide range of LEDs is automatically designed in a fully differentiable manner, within the feasible region defined by the visibility constraint. We also collect a substantially expanded VIS-NIR hyperspectral image dataset for experiments by using a customized 50-band filter wheel. Experimental results show that the task can be significantly improved by using the optimized wide-band illumination than using NIR only. Codes Available: https://github.com/MyNiuuu/VCSD.*

1. Introduction

Seeing-in-the-dark is critical for modern industries, because of its promising applications in nighttime photography and visual surveillance. However, it remains challenging due to complex degradation mechanisms and dynamics of in-the-wild environments.

To achieve this task, a number of methods have been proposed, which can be roughly divided into two threads. The first thread features RGB-dependent methods [3, 4, 29, 42, 44, 45, 52] that aim to fully exploit the RGB input, even with severe degradations. These methods have gained great success through directly learning the mapping from low-light input to normal-light output, in the presence of complex noises and color discrepancies. However, even state-of-the-art methods along this thread may struggle with in-the-wild data captured under nearly complete darkness.

In contrast, the second thread features RGB-independent methods [24, 28, 33, 38, 40] for non-interfering surveillance that try to recover RGB information from images of invisible ranges, without requiring any RGB input. The most attractive characteristics lie in its applicability to complete darkness and the visual friendliness of auxiliary illumination to naked eyes. NIR2RGB is one of the representative tasks of this thread, which aims to translate near-infrared images to RGB images.

As for auxiliary illumination in the NIR range, the industry practice is to use NIR LEDs, usually centered at 850 nm or 940 nm. However, the captured images are almost monochromatic and lack visual color and texture, which makes NIR2RGB translation ambiguous. The fundamental reasons for the ambiguities are two folds: 1) The spectral sensitivities of commodity RGB cameras almost overlap around both 850 nm and 940 nm, making it hard to recover three-channel color from a single intensity observation. 2) Reflectance spectra of many materials become almost indistinguishable beyond 850 nm, which leads to obvious structure gaps from RGB images. As a result, existing studies that tried to directly convert such NIR images to VIS images, even with the most advanced deep learning techniques, can hardly provide satisfying results due to these fundamental restrictions. In [24], Liu *et al.* proposed to properly multiplex different NIR LEDs, ranging from 700 nm to 1000 nm, to robustify the NIR2RGB task, and achieves apparently better results than using traditional 850 nm or 940 nm LEDs. However, structure gaps still exist due to the restriction of wavelengths in the NIR range, making the results far from satisfying.

[*] Corresponding author

The basic motivation of these methods arises from the invisibility of human naked eyes to NIR lights, so as to reduce visual interference and light pollution. However, up to now, none of these works have explicitly formulated the visibility of certain illumination. Liu *et al.* [24] empirically picked up the NIR range beyond $700\,\mathrm{nm}$, and there is a clear tendency that LEDs closer to this prescribed boundary are preferred according to their results. A natural question is: Is there an exact boundary between visible and invisible? This is important since it determines how much information in the VIS range can be utilized to help RGB recovery.

Inspired by the aforementioned methods, we propose to quantify and incorporate the human vision system into our model, which enables us to significantly robustify this task via illumination spectrum design in the wide-band spectral range from $420\,\mathrm{nm}$ to $890\,\mathrm{nm}$. Similar to [24], we directly optimize the spectral curve by training an image enhancement model on hyperspectral datasets. Specifically, based on the human vision system, we establish a Visibility Constrained Spectrum Design (VCSD) model to quantify the visibility of certain spectra, and to assure the prescribed visibility level will not be violated. To achieve this, a visibility threshold $\hat{\Psi}$ is introduced, which serves as the visibility upper bound during the spectrum design process. In practice, this threshold can be changed according to the desired level of visibility, without destroying the validity of our method. According to the upper bounded visibility level, the model scales down the designed LED spectrum (if necessary) to assure that the new spectrum is friendly to naked eyes. After that, we design a physic-based Imaging Process Simulation (IPS) model which synthesizes images using the corresponding LED spectrum, camera spectral sensitivity, and the reflectance spectrum of the scene. The IPS model also contains a noise model to consider the noise effect during the realistic imaging process. Since we consider the spectrum from $420\,\mathrm{nm}$ to $890\,\mathrm{nm}$, we synthesize one VIS image with lights shorter than $700\,\mathrm{nm}$ and one VIS-NIR image with the full spectrum. Through deep learning, we directly minimize the reconstruction loss and finally get the optimal LED spectral curve that can be physically realized by driving LEDs with appropriate voltage and current.

We evaluate the effectiveness of our model and designed curve on hyperspectral datasets including our proposed and previous [32] datasets. Compared to existing methods, our model clearly achieves superior results, demonstrating the powerfulness of wide-band illumination spectrum design under visibility constraints.

The main highlights of this work are:

- For the first time, we propose a paradigm that quantifies and incorporates the human vision system for seeing-in-the-dark, which enables us to significantly improve the task via illumination spectrum design in a wide-band coverage from $420\,\mathrm{nm}$ to $890\,\mathrm{nm}$.
- A novel Visibility Constrained Spectrum Design (VCSD) model is proposed to formulate and assure the visibility level of certain spectra to human naked eyes during the optimization process. The visibility threshold can be changed according to the desired level of visibility, without destroying the validity of the model.
- We design a physic-based Imaging Process Simulation (IPS) module which synthesizes the input images based on the imaging process and the noise model.
- We contribute a VIS-NIR wide-band hyperspectral image dataset to supplement existing ones in terms of quality and quantity.

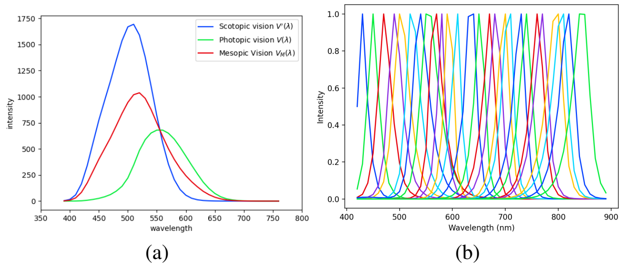
2. Related Work

Image Enhancement. Low-light image enhancement in the visible range is a critical and challenging task. Traditional image enhancement methods were mostly based on histogram manipulation [1, 5, 15, 23, 37] or Retinex theory [10, 13, 19, 22, 27, 43]. In recent years, many learning-based methods have been proposed and attracted increasingly wide interest [12, 17, 29, 42, 45]. Wei *et al.* [45] combined traditional Retinex theory with deep neural networks and provided an end-to-end framework for low-light enhancement. Supervised learning has been extensively exploited for enhancing low-light RAW images [4, 44] and videos [3, 16, 41, 52]. Very recently, enhancement methods with additional spectral image assistance have been proposed and gained great success. Xiong *et al.* [48] introduced a new flash technique for low-light imaging which uses deep-red light for assistance. They utilize the sensitivity of silicon sensors to $660\,\mathrm{nm}$ deep-red light and design a camera prototype together with a fusion network to reconstruct extra-dim scene images. However, the model suffers from color distortion and is unsuitable for wider application scenarios because deep-red flash is visible to human eyes and can be annoying or even harmful. Instead of using $660\,\mathrm{nm}$ deep-red light as guidance, Jin *et al.* [18] used $850\,\mathrm{nm}$ near-infrared images to guide the enhancement process. Compared to deep-red images, $850\,\mathrm{nm}$ NIR images suffer from structural discrepancy from corresponding RGB images under certain circumstances (*e.g.*, shadows and dyes). To overcome this issue, they proposed the Deep Inconsistency Prior (DIP) to adaptively leverage the structure inconsistency to guide the fusion of RGB-NIR.

NIR-to-RGB Translation. NIR-to-RGB Translation aims to colorize a NIR image into an RGB image. Limmer *et al.* [28] first trained a deep multi-scale convolutional neural network that performs direct and integrated transfer between NIR and RGB pixels. Suárez *et al.* [38] learned each color channel independently for NIR colorization based on the usage of a triplet model to pursue fast convergence and greater similarities. Wang *et al.* [40] proposed a multi-task framework that employs additional supervision, such as semantic loss, to aid in the NIR colorization process. To deal with the unpaired data, Nyberg *et al.* [33] and Mehri *et al.* [31] learned the mapping with an unsupervised Generative Adversarial Network (GAN) [11] based on CycleGAN [54]. However, the outputs of these methods suffer from extreme blurring as well as texture and chrominance mismatching due to the poor associations between outputs and ground truth images. Wu *et al.* [47] proposed a supervised learning-based method for NIR2RGB video translation, yet the training data were captured in the daytime, and the gap of illumination distribution between artificial LEDs and natural illuminants still exists.

NIR and RGB Image Fusion. Traditional image fusion methods were often based on spatial transformation techniques such as wavelet transform [25], contourlet transform [7], and edge-preserving filter-based transform [30]. Yan *et al.* [51] proposed a novel fusion method based on the spectral graph wavelet transform (SGWT) and the bilateral filter. Hu *et al.* [14] used the cumulative distribution of gray levels and entropy to adaptively retain infrared-hot targets and visible textures while fusing infrared and visible videos. Due to the rapid development of deep learning in recent years, many learning-based methods have attracted great attention and gained great success in this field [26, 49, 50]. Li *et al.* [26] proposed an end-to-end deep architecture with dense blocks and fusion layers to fuse infrared and visible images in one forward pass. DDcGAN [49] was proposed with a special dual-discriminator design to generate relatively realistic visible images of different resolutions. U2Fusion [50] was proposed to automatically estimate the importance of corresponding source images with adaptive information preservation degrees.

Hyperspectral Image Datasets. Various hyperspectral dataset has been proposed in order to analyze the characteristics of different wavelengths. Arad *et al.* [2] proposed an ICVL dataset that contains hyperspectral data of 201 different scenes. The dataset was taken in sufficient light using a Specim PS Kappa DX4 hyperspectral camera and a rotary stage for spatial scanning, and most of them are captured outdoors. Monno *et al.* [32] proposed the TokyoTech dataset containing 59-band visible-NIR hyperspectral images from $420\,\mathrm{nm}$ to $1000\,\mathrm{nm}$ at $10\,\mathrm{nm}$ intervals. The images were captured using a monochrome camera and two VariSpec tunable filters, VIS for $420\,\mathrm{nm}$-$650\,\mathrm{nm}$ and SNIR for $650\,\mathrm{nm}$-$1000\,\mathrm{nm}$, for capturing each hyperspectral image. Liu *et al.* [24] built a complex imaging system and contributed an IDH dataset containing hyperspectral images from $420\,\mathrm{nm}$ to $1000\,\mathrm{nm}$ at $10\,\mathrm{nm}$ intervals. The UI-3860CP grayscale camera together with the Kurios-XE2 tunable filter is used to record spectral images from $650\,\mathrm{nm}$ to $1000\,\mathrm{nm}$, while the 15S5C camera is used to record the RGB image. In this paper, we contribute a new hyperspectral image dataset to supplement the existing HSI datasets in terms of quality and quantity.



(a) (b)

Figure 1. (a) Photopic, scotopic, and mesopic intensity functions. (b) The spectra of 26 narrow band LEDs used in our experiments.

3. Method

In this section, we present our Visibility Constrained Spectrum Design (VCSD) method. We first introduce the human vision system in Sec.3.1 as the prerequisite of the VCSD model, which will be described in Sec.3.2. After that, we introduce the physic-based Imaging Process Simulation (IPS) model in Sec.3.3. We then describe our Image Restoration model in Sec.3.4. Finally, we sum up the training procedure in Sec. 3.5.

3.1. Human Vision System

The human eye is sensitive to wavelengths roughly between $400\,\mathrm{nm}$ and $700\,\mathrm{nm}$. Wavelengths shorter than $400\,\mathrm{nm}$ or longer than $700\,\mathrm{nm}$ are almost invisible. For wavelengths between $400\,\mathrm{nm}$ and $700\,\mathrm{nm}$, human eyes behave differently in high or low light conditions. In relatively high light conditions, the vision is mainly relevant to the center of the retina whose maximum sensitivity is at $555\,\mathrm{nm}$ (in the green region). This type of vision is called photopic vision. While in extremely low light conditions for human eyes, which is our case, the vision is done by the peripheral region of the retina whose maximum sensitivity is at $507\,\mathrm{nm}$ (in the blue-green region). This type of vision is called scotopic vision. At intermediate light levels, both rods and cones are active, which is called mesopic vision. As shown in Fig.1(a), given the photopic and scotopic luminosity functions as $V(\lambda)$ and $V'(\lambda)$, the mesopic luminosity function $V_M(\lambda)$ can be approximated as:

$$V_M(\lambda) = (1 - x)V'(\lambda) + xV(\lambda), \qquad (1)$$

where x is determined by photopic illuminance and the composition of light source [6, 8, 39, 48].

3.2. Visibility Constrained Spectrum Design

To consider different wavelengths, we propose to find an optimal LED spectral multiplexing based on K LED bases of different wavelengths:

$$\Phi(\lambda) = \sum_{k=0}^{K-1} \sigma^k \cdot \Phi^k(\lambda), \tag{2}$$

where $\boldsymbol{\sigma} = [\sigma^0, \sigma^1, ..., \sigma^{K-1}], \sigma^k \in (0, 1)$ is the parameter that determine the weight of corresponding LED base $\Phi^k(\lambda)$, and can be optimized during the training process.

To consider the human vision system, the key issue is to find a way to quantify the visibility of certain LED spectral curves. Given the scotopic intensity functions $V'(\lambda)$ and the LED spectral curve $\Phi(\lambda)$, The perceived power P_m of light by human naked eyes is proportional to the inner product of $V'(\lambda)$ with $\Phi(\lambda)$ [48]:

$$P_m \propto \int V'(\lambda)\Phi(\lambda)d\lambda = \Psi. \tag{3}$$

Since the relationship is proportional, not equal, it is hard to get the real value of P_m given certain $V'(\lambda)$ and $\Phi(\lambda)$. However, it is indeed possible to *determine a threshold $\hat{\Psi}$ through user studies, which represents 'just' invisible to the human eye.*

Therefore, given the threshold $\hat{\Psi}$ and a mutiplexed LED spectrum $\Phi(\lambda) = \sum_{k=0}^{K-1} \sigma^k \cdot \Phi^k(\lambda)$ that is visible to human eyes (*i.e.*, $\int V'(\lambda)\Phi(\lambda)d\lambda > \hat{\Psi}$), we calculate a scale factor $\xi \in (0, 1)$ so that when $\hat{\sigma}^k = \xi \cdot \sigma^k, (k = 0, 1, ..., K-1)$, $\hat{\Phi}(\lambda) = \sum_{k=0}^{K-1} \hat{\sigma}^k \cdot \Phi^k(\lambda)$ becomes just invisible to human eyes, *i.e.*, the perceived power of light by human scotopic vision equals to $\hat{\Psi}$:

$$\int V'(\lambda)\hat{\Phi}(\lambda)d\lambda = \hat{\Psi}. \tag{4}$$

$$\therefore\ \xi = \frac{\hat{\Psi}}{\int V'(\lambda)\Phi(\lambda)d\lambda}. \tag{5}$$

$$\Rightarrow\ \hat{\sigma}^k = \frac{\hat{\Psi}}{\int V'(\lambda)\Phi(\lambda)d\lambda + \epsilon} \cdot \sigma^k, k = 0, ..., K-1, \tag{6}$$

where ϵ is a small constant to avoid numerical issues. Therefore, given an LED spectrum $\Phi(\lambda)$ that is visible to human eyes, we scale it by ξ calculated from Eq.6 to make it just invisible to human naked eyes. Note that for LED spectrum that is already 'invisible' to human naked eyes (*i.e.*, $\int V'(\lambda)\Phi(\lambda)d\lambda < \hat{\Psi}$), we just let $\xi = 1$ since there is no need to adjust the spectral curve intensity of these LEDs.

3.3. Imaging Process Simulation Model

In computational photography, the formation of images depends on three factors: the reflectance spectrum $\mathcal{T}$, the illumination spectrum Φ, and the camera spectral sensitivity $\mathcal{C}$. Given these three factors, the process of acquiring light intensity for each pixel can be formulated as:

$$\mathcal{I}_{c,i,j} = \int \mathcal{T}_{i,j}(\lambda) \cdot \Phi(\lambda) \cdot \mathcal{C}_c(\lambda)\, d\lambda, \tag{7}$$

where $c \in \{R, G, B\}$ represents the color channel. $i \in \{1, 2, ..., W\}, j \in \{1, 2, ..., H\}$, W and H is the width and height of the image. $\mathcal{I}_{c,i,j}$ denotes the RGB intensity in channel c at position (i, j). $\mathcal{T}_{i,j}(\lambda)$ is the reflectance spectrum in position (i, j). $\Phi(\lambda)$ is the LED spectrum. $\mathcal{C}_c(\lambda)$ is the camera spectral sensitivity in channel c. The image can be obtained according to $\mathcal{I}_{c,i,j}$ in each position.

Based on this physical process, we design an Imaging Process Simulation (IPS) model to generate assistance images. The IPS module takes assistance spectral curve $\Phi \in \mathbb{R}^L$, camera spectral sensitivity $\mathcal{C} \in \mathbb{R}^{3\times L}$, and reflectance spectrogram $\mathcal{T} \in \mathbb{R}^{L\times W\times H}$ as inputs, and outputs the synthesized assistance images $\mathcal{I} \in \mathbb{R}^{3\times W\times H}$. We set the range of camera spectral sensitivity to $420\,\mathrm{nm}$-$890\,\mathrm{nm}$, with a $10\,\mathrm{nm}$ interval, so the process can be formulated as:

$$\mathcal{I}_{c,i,j} = \sum_{n=0}^{47} \mathcal{T}_{i,j}(n) \cdot \Phi(n) \cdot \mathcal{C}_c(n), \tag{8}$$

where $n = 0, 1, ..., 47$ represents 48 different wavelengths covered by the camera spectral sensitivity. Note that this process is fully differentiable under Equ.8, allowing us to optimize the parameter $\boldsymbol{\sigma}$ that determines the weight of each LED base.

Noise Simulation. In low-light environments, assistance images are usually free of obvious noise interference due to enough illumination provided by the LEDs, but noise still exists under these conditions. Also, according to our Visibility Model, the intensity of LED may become very small in order to become invisible to the human naked eyes, which makes the assistance images suffer from obvious noise interference. Since Equ. 8 can't model the real camera noise widely existing during the image formulation, we additionally introduce a noise model to consider the noise effects for assistance images.

Poisson Distribution has been widely considered to model the noise distribution [9, 46, 53]. Here we choose to combine Poisson Distribution with noise sampling from a real camera sensor to realize our noise model:

$$\hat{\mathcal{I}} = \kappa \cdot \mathcal{P}(\frac{\mathcal{I} \cdot \xi}{\kappa}) + \mathcal{N}, \tag{9}$$

where κ is the gain of the target camera, ξ is the scale-down factor calculated in the Visibility Constrained Spectrum Model, and $\mathcal{N}$ is the real noise pattern sampled from the target camera.

3.4. Image Restoration Model

Network Architecture. Our fusion network takes a VIS image $\hat{\mathcal{I}}^{vis}$ and an NIR-VIS image $\hat{\mathcal{I}}^{nir}$ as input and generates the result $\mathcal{X}$. We choose the same UNet [35] structure

Algorithm 1 Training

```
Input: Visibility Threshold $\hat{\Psi}$, LED bases $[\Phi^k]_{k=0}^{K-1}$, Camera gain $\kappa$, Camera Spectral Sensitivity $\mathcal{C}$, Hyperspectral Dataset $\mathcal{D}_{tr}$, and hyperparameters $\boldsymbol{\sigma}, \boldsymbol{\theta}$.
Output: Optimal wide-band spectral curve.
Initialize $\boldsymbol{\sigma} \leftarrow \boldsymbol{\sigma}_t = [\sigma_t^0, \sigma_t^1, ..., \sigma_t^{K-1}]$.
while not converged do
    Randomly sample $\mathcal{T}_t, \mathcal{Y}_t, \mathcal{N}_t^{vis}, \mathcal{N}_t^{nir}$ from $\mathcal{D}_{tr}$.
    $\Phi_t(\lambda) \leftarrow \sum_{k=0}^{K-1} \sigma_t^k \cdot \Phi^k(\lambda)$
    $\mathcal{T}_t^{vis} \leftarrow \mathcal{T}_t^{420:700nm}$, $\Phi_t^{vis} \leftarrow \Phi_t^{420:700nm}$
    $\xi_t \leftarrow \min(\frac{\hat{\Psi}}{\int V'(\lambda)\Phi_t(\lambda)d\lambda+\epsilon}, 1)$
    for $k \leftarrow 0$ to $K-1$ do
        if $\int V'(\lambda)\Phi^k(\lambda)d\lambda > 0$ then
            $\hat{\sigma}_t^k \leftarrow \xi_t \cdot \sigma_t^k$
        else
            $\hat{\sigma}_t^k \leftarrow \sigma_t^k$
        end if
    end for
    $\hat{\Phi}_t(\lambda) = \sum_{k=0}^{K-1} \hat{\sigma}_t^k \cdot \Phi^k(\lambda)$
    $\hat{\Phi}_t^{vis} \leftarrow \hat{\Phi}_t^{420:700nm}$
    $\xi_t^{vis} \leftarrow \frac{\int \hat{\Phi}_t^{vis}(\lambda)d\lambda}{\int \Phi_t^{vis}(\lambda)d\lambda+\epsilon}$, $\xi_t^{nir} \leftarrow \frac{\int \hat{\Phi}_t(\lambda)d\lambda}{\int \Phi_t(\lambda)d\lambda+\epsilon}$
    $\mathcal{I}_{c,i,j,t}^{vis} \leftarrow \int \mathcal{T}_{i,j,t}^{vis}(\lambda) \cdot \hat{\Phi}_t^{vis}(\lambda) \cdot \mathcal{C}_c(\lambda)d\lambda$
    $\mathcal{I}_{c,i,j,t}^{nir} \leftarrow \int \mathcal{T}_{i,j,t}(\lambda) \cdot \hat{\Phi}_t(\lambda) \cdot \mathcal{C}_c(\lambda)d\lambda$
    $\hat{\mathcal{I}}_t^{vis} \leftarrow \kappa \cdot \mathcal{P}(\frac{\mathcal{I}_t^{vis} \cdot \xi_t^{vis}}{\kappa}) + \mathcal{N}_t^{vis}$
    $\hat{\mathcal{I}}_t^{nir} \leftarrow \kappa \cdot \mathcal{P}(\frac{\mathcal{I}_t^{nir} \cdot \xi_t^{nir}}{\kappa}) + \mathcal{N}_t^{nir}$
    $\mathcal{X}_t \leftarrow G(\hat{\mathcal{I}}_t^{vis}, \hat{\mathcal{I}}_t^{nir}; \boldsymbol{\theta})$
    Take gradient descent step on
        $\nabla_{\boldsymbol{\theta},\boldsymbol{\sigma}} \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t)$
end while
return Optimized spectral curve $\boldsymbol{\sigma}^*$
```

as [24] during the curve design process, except for the number of input channels, which is 3 in [24], and 6 in our work.

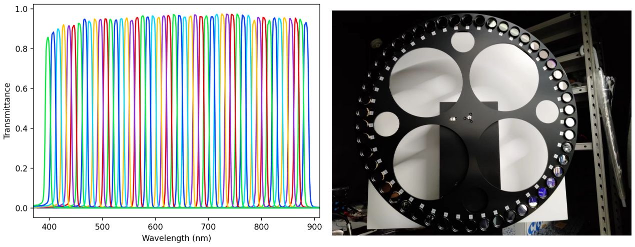
Loss Function. The perceptual loss [20] has been widely used in image reconstruction tasks due to its ability to recover details and preclude over-smooth results compared to pixel-wise losses:

$$\mathcal{L} = \sum_{i=1}^{I} \left\| \psi_i(\mathcal{X}) - \psi_i(\mathcal{Y}) \right\|_1, \tag{10}$$

where $\mathcal{X}$ is the output of G, and $\mathcal{Y}$ is the corresponding ground truth. ψ_i denotes the activation map at the i-th layer of the pre-trained VGG-19 network [36]. Particularly, we chose 5 layers including $relu_{1-1}$, $relu_{2-1}$, $relu_{3-1}$, $relu_{4-1}$, and $relu_{5-1}$ from the VGG-19 network.

3.5. Training Procedure

Algorithm 1 displays the complete training procedure. We first initialize the parameter $\boldsymbol{\sigma}$ that determines the weight of each LED base, which will be optimized by the gradient. In each iteration, scale factor ξ_t is first calculated and used to make the designed spectrum Φ_t invisible. Note that since there exist several NIR LED bases whose spectrum curves have no intersection with scotopic intensity functions (*i.e.*, $\int V'(\lambda)\Phi^k(\lambda)d\lambda = 0$), we may not want to scale down the corresponding coefficient σ_t^k since it provides no improvements for visibility but causes information loss. As a result, we choose to only scale down the coefficients of LED bases whose spectrum have intersections with scotopic intensity functions, and obtain the new curve $\hat{\Phi}_t(\lambda)$ that also fulfills the visibility limitations:



(a) (b)

Figure 2. (a) The transmittance curves of 50 band pass filters. (b) Camera system to capture the dataset.

Table 1. Comparison between different Hyperspectral Datasets.

Datasets	ICVL	TokyoTech	IDH	Ours
Resolution	1392×1300	512×512	256×256	1936 × 1096
Scenes	201	16	112	74
Range/nm	400-1000	420-1000	650-1000	400-890
Interval/nm	1.25	10	10	10

$$\xi_t = \min(\frac{\hat{\Psi}}{\int V'(\lambda)\Phi_t(\lambda)d\lambda + \epsilon}, 1) \tag{11}$$

$$\hat{\sigma}_t^k = \begin{cases} \xi_t \cdot \sigma_t^k, \int V'(\lambda)\Phi^k(\lambda)d\lambda > 0 \\ \sigma_t^k, \int V'(\lambda)\Phi^k(\lambda)d\lambda = 0 \end{cases} \tag{12}$$

$$\hat{\Phi}_t(\lambda) = \sum_{k=0}^{K-1} \hat{\sigma}_t^k \cdot \Phi^k(\lambda). \tag{13}$$

We then calculate the scale-down factor for VIS images and NIR images as:

$$\xi_t^{vis} = \frac{\int \hat{\Phi}_t^{vis}(\lambda)d\lambda}{\int \Phi_t^{vis}(\lambda)d\lambda + \epsilon}, \quad \xi_t^{nir} = \frac{\int \hat{\Phi}_t(\lambda)d\lambda}{\int \Phi_t(\lambda)d\lambda + \epsilon}, \tag{14}$$

where Φ_t^{vis} and $\hat{\Phi}_t^{vis}$ are the 420 nm-700 nm part of Φ_t and $\hat{\Phi}_t$, respectively. Based on $\hat{\Phi}_t(\lambda)$, input image $\hat{\mathcal{I}}_t^{nir}$ and $\hat{\mathcal{I}}_t^{vis}$ are simulated via the physic-based IPS module which considers noise effects related to ξ_t^{vis} and ξ_t^{nir}. After that, the output $\mathcal{X}_t$ is obtained through our image enhancement network G which takes auxiliary images $\hat{\mathcal{I}}_t^{nir}$ and $\hat{\mathcal{I}}_t^{vis}$ as input. Finally, gradient descent steps are taken based on the loss function $\mathcal{L}$.

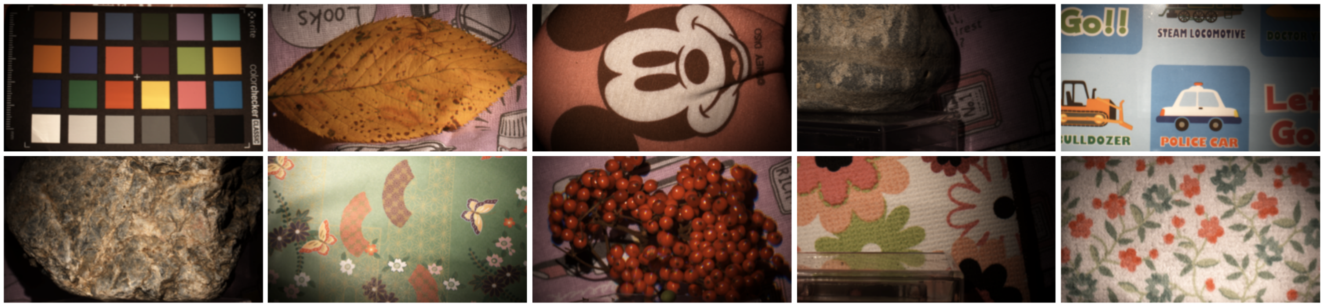


Figure 3. Data samples from our proposed dataset. The images are synthesized using the spectrum of white-light LED.

3.6. Implementation Details

We implement the training part of our model with Pytorch [34]. During the spectrum optimization process, we set the batch size to 16 and the learning rate to 1e-3. The total training iteration is 50,000, and the learning rate is multiplied by 0.1 every 20,000 iterations. We use Adam optimizer [21] with $\beta_1 = 0.5, \beta_2 = 0.999$, and randomly crop the input images to 256×256. We set the number K of LEDs to 26, covering the wide-band VIS-NIR range from $420\,\mathrm{nm}$ to $890\,\mathrm{nm}$. The spectrum of these LED bases is shown in Fig. 1(b). We use the camera GS3-U3-15S5C for both image synthesis and real image capture. We choose a normalized $660\,\mathrm{nm}$ LED spectrum to obtain the visibility threshold $\hat{\Psi} = 10$ for our main experiment, following the claim in [48]. We also further discuss the effect of different visibility thresholds on our model in the experiment part.

After obtaining the optimal spectrum, we train an image restoration network using synthesized input images. To train the restoration network on our proposed dataset, we set the batch size to 16 and the learning rate to 1e-4. The total training iteration is 10,000 iterations. We use Adam optimizer [21] with $\beta_1 = 0.5, \beta_2 = 0.999$, and randomly crop the input images to 256×256. To train the restoration model on TokyoTech [32] dataset, we set the batch size to 8 and keep the rest of the settings the same as training on our dataset.

4. Experiments

4.1. Settings

Datasets. To train our model, we need datasets that contain hyperspectral images of multiple scenes that cover from VIS range to NIR range. Up to now, some hyperspectral datasets have been proposed, including ICVL [2], TokyoTech [32], and IDH [24]. ICVL contains hyperspectral data of 201 scenes, with a spatial resolution of 1392 × 1300 and 519 spectral bands ($400\,\mathrm{nm}$ to $1000\,\mathrm{nm}$ at roughly $1.25\,\mathrm{nm}$ increments). TokyoTech contains 59-band hyperspectral images from $420\,\mathrm{nm}$ to $1000\,\mathrm{nm}$ at $10\,\mathrm{nm}$ intervals. The image resolution is 512 × 512, and only 16 scenes are publicly available. IDH dataset contains a total of 112 hyperspectral images from $650\,\mathrm{nm}$ to $1000\,\mathrm{nm}$ at $10\,\mathrm{nm}$ intervals, and the spatial resolution is only 256 × 256. In this paper, we contribute a new hyperspectral image dataset to supplement existing HSI datasets in terms of quality and quantity. The wavelength covers from $400\,\mathrm{nm}$ to $890\,\mathrm{nm}$ with $10\,\mathrm{nm}$ intervals. There are 74 scenes in the dataset, with a resolution of 1936 × 1096. The ground truth VIS image is synthesized via the white-light LED spectral curve. A comparison between different hyperspectral image datasets is shown in Tab. 1.

We test the effect of our designed spectrum on two datasets: our proposed dataset and the TokyoTech dataset. To test the result, we first obtain the two auxiliary images according to the designed curve, then train an image enhancement network on the training set. We then test the effect of this network on the test set. The synthetic dataset comes from our collected hyperspectral dataset. Two auxiliary images and the corresponding ground truth are obtained through the physic-based imaging process described in Equ. 8. The experimental results will be introduced in Sec. 4.2.

Methods. Based on existing popular solutions for low-light imaging, we conduct experiments on three different settings of our method: 1) *VIS+*NIR: Optimal spectrum design for both VIS and NIR (our main method), 2) *VIS+$850\,\mathrm{nm}$: Optimal spectrum design for VIS + fixed $850\,\mathrm{nm}$ Auxiliary Illumination (based on [18]), and 3) *VIS+$660\,\mathrm{nm}$: Optimal spectrum design for VIS + fixed $660\,\mathrm{nm}$ Auxiliary Illumination (based on [48]). All the VIS images are synthesized based on the $420\,\mathrm{nm}$-$700\,\mathrm{nm}$ spectrum. We then compare these three settings with [24], which achieve superior results than traditional RGB-independent methods by retrieving the optimal NIR spectrum whose wavelength is larger than $700\,\mathrm{nm}$.

4.2. Main Results

In this section, we test different methods on synthetic datasets including our proposed dataset and TokyoTech [32]. Quantitative results on our dataset and the TokyoTech dataset are reported in Tab. 2. During the evaluation process, SSIM (Structural Similarity), Peak Signal-to-Noise Ratio (PSNR), and Learned Perceptual Image Patch

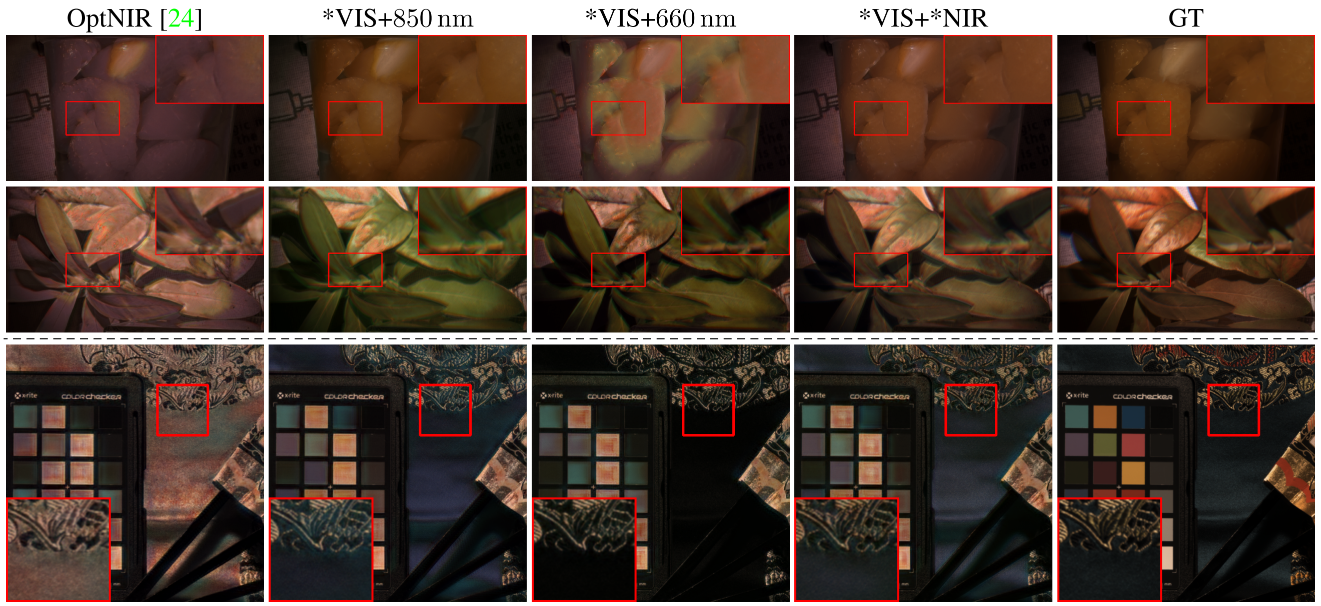


Figure 4. Qualitative results on our dataset (above dash) and the TokyoTech dataset [32] (below dash). Please zoom in for a clear view.

Table 2. Quantitative results on different synthetic datasets. The best results are in **red** whereas the second best are in blue.

Datasets	Methods	SSIM ↑	PSNR ↑	LPIPS ↓
Ours	OptNIR [24]	0.7688	22.67	0.1590
	*VIS+850 nm	0.8305	24.07	0.1276
	*VIS+660 nm	0.8316	23.72	0.1232
	*VIS+*NIR	**0.8383**	**24.12**	**0.1129**
TokyoTech	OptNIR [24]	0.7197	19.65	0.1841
	*VIS+850 nm	0.7902	21.78	**0.1365**
	*VIS+660 nm	0.7628	21.45	0.1383
	*VIS+*NIR	**0.7938**	**22.08**	0.1378

Similarity (LPIPS) are utilized to quantify the difference between restored images and ground truth. Visual results on our dataset and the TokyoTech dataset are shown in Fig. 4. We can see that three different settings of our method all perform significantly better than [24] in three metrics. Instead of empirically picking up the NIR range beyond 700 nm, we find the theoretically optimal curve under the visibility constraint $\hat{\Psi}$ by incorporating the human vision intensity function into the optimization process. As a result, we significantly robustify the task and achieve superior results on two datasets. Furthermore, if we use the combination of optimal VIS + empirically fixed NIR spectrum (*e.g.*, 660 nm, 850 nm), the results may get worse since these combinations are not optimal under the visibility constraint, which also proves the effectiveness of our visibility constrained spectrum design model.

Designed Curve. The red line in Fig. 5 shows the optimal curve for $\hat{\Psi}$=10 desinged by our model. To further demonstrate the practicality of our model, we implemented the designed spectrum under $\hat{\Psi}$=10 by properly multiplexing different LED bases. As shown in Fig. 5, we approximately fit the designed curve by controlling the voltage of six LED bases, demonstrating the practicality of our designed curve.

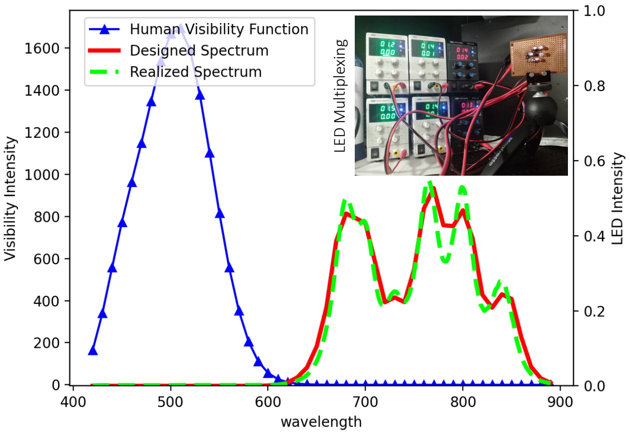


Figure 5. Realization for our designed LED spectrum.

4.3. Impact of Visibility Threshold

As has been introduced in Sec. 3.2, our model accepts a visibility upper-bound threshold $\hat{\Psi}$ during the spectrum optimization process. This threshold can be changed according to the desired level of visibility and largely affects the final results, without destroying the validity of our method. In this section, we further discuss the impact of this visibility threshold on our model. Specifically, we set different values for the visibility threshold $\hat{\Psi}$, designing the optimal spectrum under each value, and compare their restoration results. The visual and numerical results are reported

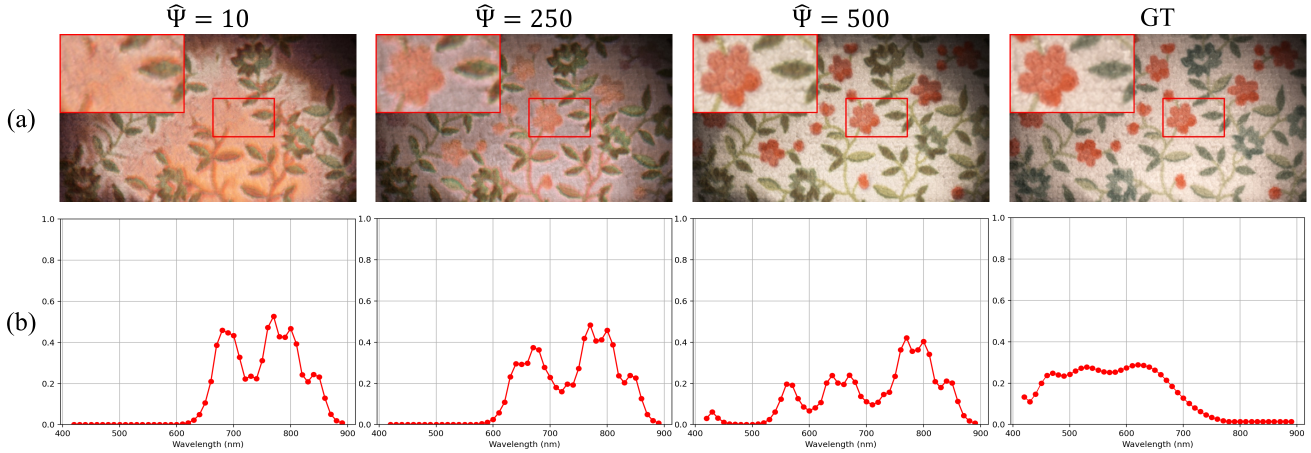


Figure 6. Qualitative results for the impact of different visibility thresholds. (a) Visual results on our proposed dataset, (b) designed curve under the corresponding threshold $\hat{\Psi}$ (For GT, we show the curve of white-light LED, which is used to obtain GT RGB images).

Table 3. Quantitative results on different values for visibility threshold $\hat{\Psi}$. The best results are in **red** whereas the second best are in blue.

Datasets	$\hat{\Psi}$	SSIM ↑	PSNR ↑	LPIPS ↓
Ours	10	0.8383	24.12	0.1129
	250	0.8779	25.64	0.0919
	500	**0.9326**	**29.07**	**0.0351**
TokyoTech	10	0.7938	22.08	0.1378
	250	0.8495	23.77	0.0920
	500	**0.9375**	**31.85**	**0.0355**

in Fig. 6 and Tab. 3. We can see that as the value of $\hat{\Psi}$ grows, the restoration results of our model become better since more VIS information is covered. The results also imply that we can trade visibility friendliness for restoration performance by setting different value for $\hat{\Psi}$, making our model applicable to a wider range of application scenarios.

Shape of the optimal curves. As shown in Fig. 5, the scotopic visibility function roughly covers 400 nm-600 nm, with a peak value of 1700. High visibility wavelengths have a higher 'cost' per intensity than wavelengths with lower visibility. As a result, the model chooses to approach from 'sides' with low visibility to the 'center' (500 nm) of the visibility curve during the design process. From Fig. 6, we can see that the model still tends not to use around 500 nm even when $\hat{\Psi}$=500 since the 'cost' is too high. An intuitive thought is that the designed curve should distribute low intensities that fulfill the visibility constraint to a wide VIS range to provide information of different wavelengths. This is, however, not always feasible because of the *noise interference*, especially under strict visibility constraints (*e.g.*, $\hat{\Psi}$=10). Specifically, when the intensity of light becomes very low to cover the high visibility range, the structures and colors may suffer from severe degradation due to *low signal-to-noise ratio*. This makes the model prefer wavelengths that can achieve relatively high intensity under strict visibility constraints, instead of distributing low intensities to a wide VIS range.

5. Conclusion

In this paper, we proposed a visibility-constrained wideband illumination spectrum design (VCSD) model for Seeing-in-the-Dark. Our key insight is to incorporate the quantified visibility constraint implied by the human vision system into the optimization process. By modeling the image formation process in the VIS-NIR range, the optimal multiplexing of a wide range of LEDs is designed in a fully automatic manner, while fulfilling the visibility constraint. We also collected a substantially expanded VIS-NIR hyperspectral image dataset for experiments by using a customized 50-band filter wheel. Experimental results show that the task can be significantly improved by using the optimized wide-band illumination than using NIR only. Further analysis also proved the generality and flexibility of our model to deal with different visibility thresholds.

Although narrow band LEDs are cost effective, they might not be the most appropriate choice when the purpose is to recover high-fidelity visible color, due to the scale-down operation implied by the visibility constraint. Our future work is to allow more flexible illumination design by using wide band fluorescent dyes or customizing thin-film interference filters.

Acknowledgement

This research was supported in part by JSPS KAKENHI Grant Numbers 22H00529, 20H05951, and ROIS NII Open Collaborative Research 2023-23S1201.

References

[1] 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. 2

[2] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 3, 6

[3] Chen Chen, Qifeng Chen, Minh N Do, and Vladlen Koltun. Seeing motion in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3185–3194, 2019. 1, 2

[4] 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. 1, 2

[5] Dinu Coltuc, Philippe Bolon, and J-M Chassery. Exact histogram specification. *IEEE Transactions on Image processing*, 15(5):1143–1152, 2006. 2

[6] BH Crawford. The scotopic visibility function. *Proceedings of the Physical Society. Section B*, 62(5):321, 1949. 3

[7] Arthur L Da Cunha, Jianping Zhou, and Minh N Do. The nonsubsampled contourlet transform: theory, design, and applications. *IEEE transactions on image processing*, 15(10):3089–3101, 2006. 3

[8] Marjukka Eloholma and Liisa Halonen. New model for mesopic photometry and its application to road lighting. *Leukos*, 2(4):263–293, 2006. 3

[9] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1436–1444, 2022. 4

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

[11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014. 3

[12] 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)*, pages 1780–1789, June 2020. 2

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

[14] Hai-Miao Hu, Jiawei Wu, Bo Li, Qiang Guo, and Jin Zheng. An adaptive fusion algorithm for visible and infrared videos based on entropy and the cumulative distribution of gray levels. *IEEE Transactions on Multimedia*, 19(12):2706–2719, 2017. 3

[15] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 2

[16] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7324–7333, 2019. 2

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

[18] Shuangping Jin, Bingbing Yu, Minhao Jing, Yi Zhou, Jiajun Liang, and Renhe Ji. Darkvisionnet: Low-light imaging via rgb-nir fusion with deep inconsistency prior. *AAAI*, 2022. 2, 6

[19] 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, 1997. 2

[20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 5

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

[22] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 2

[23] 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. 2

[24] Liu Lei, Chen Yuze, Yan Junchi, and Zheng Yinqiang. Optimal led spectral multiplexing for nir2rgb translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 3, 5, 6, 7

[25] John J Lewis, Robert J O'Callaghan, Stavri G Nikolov, David R Bull, and Nishan Canagarajah. Pixel-and region-based image fusion with complex wavelets. *Information fusion*, 8(2):119–130, 2007. 3

[26] Hui Li and Xiao-Jun Wu. Densefuse: A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5):2614–2623, 2019. 3

[27] 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. 2

[28] Matthias Limmer and Hendrik PA Lensch. Infrared colorization using deep convolutional neural networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 61–68. IEEE, 2016. 1, 2

[29] 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. 1, 2

[30] Jinlei Ma, Zhiqiang Zhou, Bo Wang, and Hua Zong. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Physics & Technology*, 82:8–17, 2017. 3

[31] Armin Mehri and Angel D Sappa. Colorizing near infrared images through a cyclic adversarial approach of unpaired samples. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 971–979. IEEE, 2019. 3

[32] Yusuke Monno, Hayato Teranaka, Kazunori Yoshizaki, Masayuki Tanaka, and Masatoshi Okutomi. Single-sensor rgb-nir imaging: High-quality system design and prototype implementation. *IEEE Sensors Journal*, 19(2):497–507, 2018. 2, 3, 6, 7

[33] Adam Nyberg, Abdelrahman Eldesokey, David Bergström, and David Gustafsson. Unpaired thermal to visible spectrum transfer using adversarial training. In *European Conference on Computer Vision Workshops*, pages 657–669. Springer, 2018. 1, 3

[34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, , et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[35] 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. 4

[36] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[37] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on image processing*, 9(5):889–896, 2000. 2

[38] Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla. Infrared image colorization based on a triplet dcgan architecture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 18–23, 2017. 1, 2

[39] George Wald. Human vision and the spectrum. *Science*, 101(2635):653–658, 1945. 3

[40] Fengqiao Wang, Lu Liu, and Cheolkon Jung. Deep near infrared colorization with semantic segmentation and transfer learning. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 455–458, 2020. 1, 2

[41] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9700–9709, 2021. 2

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

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

[44] Yuzhi Wang, Haibin Huang, Qin Xu, Jiaming Liu, Yiqun Liu, and Jue Wang. Practical deep raw image denoising on mobile devices. In *European Conference on Computer Vision*, pages 1–16. Springer, 2020. 1, 2

[45] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2

[46] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 4

[47] Guangming Wu, Yinqiang Zheng, Zhiling Guo, Zekun Cai, Xiaodan Shi, Xin Ding, Yifei Huang, Yimin Guo, and Ryosuke Shibasaki. Learn to recover visible color for video surveillance in a day. In *European Conference on Computer Vision*, pages 495–511. Springer, 2020. 3

[48] Jinhui Xiong, Jian Wang, Wolfgang Heidrich, and Shree Nayar. Seeing in extra darkness using a deep-red flash. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10000–10009, 2021. 2, 3, 4, 6

[49] Han Xu, Pengwei Liang, Wei Yu, Junjun Jiang, and Jiayi Ma. Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators. In *IJCAI*, pages 3954–3960, 2019. 3

[50] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3

[51] Xiang Yan, Hanlin Qin, Jia Li, Huixin Zhou, and Jing-guo Zong. Infrared and visible image fusion with spectral graph wavelet transform. *JOSA A*, 32(9):1643–1652, 2015. 3

[52] Huanjing Yue, Cong Cao, Lei Liao, Ronghe Chu, and Jingyu Yang. Supervised raw video denoising with a benchmark dataset on dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2301–2310, 2020. 1, 2

[53] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4593–4601, 2021. 4

[54] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 3