Rendering Nighttime Image Via Cascaded Color and Brightness Compensation

Zhihao Li, Si Yi, and Zhan Ma*

Abstract

Image signal processing (ISP) is crucial for camera imaging, and neural networks (NN) solutions are extensively deployed for daytime scenes. The lack of sufficient nighttime image dataset and insights on nighttime illumination characteristics poses a great challenge for high-quality rendering using existing NN ISPs. To tackle it, we first built a high-resolution nighttime RAW-RGB (NR2R) dataset with white balance and tone mapping annotated by expert professionals. Meanwhile, to best capture the characteristics of nighttime illumination light sources, we develop the CBUnet, a two-stage NN ISP to cascade the compensation of color and brightness attributes. Experiments show that our method has better visual quality compared to traditional ISP pipeline, and is ranked at the second place in the NTIRE 2022 Night Photography Rendering Challenge [13] for two tracks by respective People's and Professional Photographer's choices. The code and relevant materials are avaiable on our website[1].

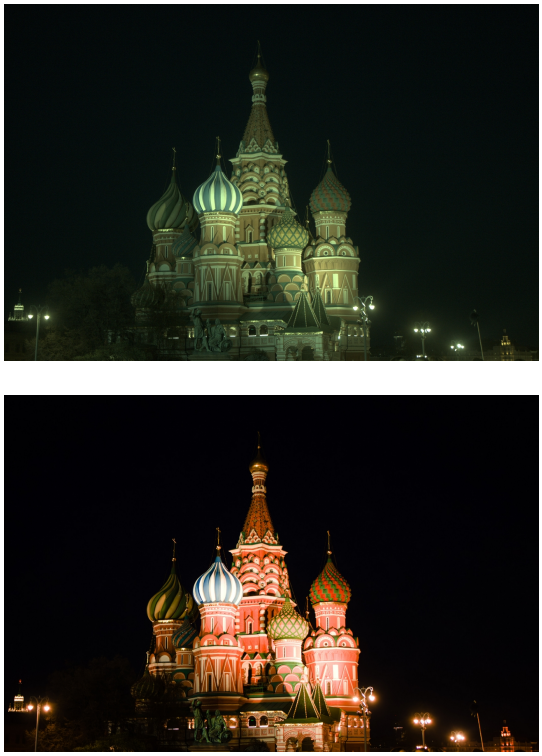

Figure 1. Nighttime RAW image (visualized) and the corresponding RGB image processed rened by the proposed CBUnet.

1. Introduction

As for camera imaging, photons first converge on the sensor chip to generate the raw electrical image (RAW) reflecting acquired scene; And then the image signal processor (ISP) is usually devised to transform camera RAW image to corresponding RGB representation pleasantly perceivable to human visual system (HVS). Therefore, the efficiency of ISP is of great importance for the generation of high-quality RGB images used in vast applications.

Traditional ISP systems usually include demosaicing, automatic white balance (AWB), color space conversion, tone mapping, denoising, compression, etc. Among them, most steps aim at alleviating or eliminating the inherent defects incurred by the image sensor and environment to render high-quality images in terms of the color, brightness, texture sharpness, dynamic range, etc. for visually pleasant presentation to HVS perception. Oftentimes, modular subsystems in ISP exemplified previously are developed separately, usually taking into account the characteristics of underlying sensor and optics for systematic optimization.

In practice, the HVS can efficiently render dynamic scenes because of the priors in brain memory [25, 38] (e.g., background illumination, environment understanding), which however is difficult for traditional ISPs to capture due to the lack of understanding of scene. Therefore, some recent studies [5, 6, 28] have tried to use neural networks (NN) to replace some modules in ISP to dynamically adjust the output RGB image with scene priors accordingly. For example, Bianco *et al*. [6] presented a three-stage CNN model method to do illuminant estimation of RAW images. Panetta *et al*. [28] proposed a deep learning-based tone mapping operator (TMO-Net), which offered a generalized, efficient and parameter-free way across a wider

*Z. Li (lizhihao6@outlook.com) and Z. Ma (mazhan@nju.edu.cn) are with Nanjing University, Jiangsu 210093, China; And S. Yi (1811326@mail.nankai.edu.cn) is with Nankai University, Tianjin, China.

[1]https://njuvision.github.io/CBUnet

Team No.	#Votes	Mean Score	Rank People	Rank Pro.
Team1	2603	0.8009	1	1
Ours	**2047**	**0.6298**	**2**	**2**
Team3	1979	0.6089	3	3
Team4	1964	0.6045	4	4
Team5	1935	0.5955	5	6
Team6	1866	0.5742	6	7
Team7	1559	0.4798	7	8
Team8	1505	0.4631	8	9
Team9	1433	0.4411	9	10
Team10	1288	0.3965	10	5

Table 1. Result Illustration of Number of Votes (#Votes), Mean Score, Rank of respective People and Professional Photographer's choice for NTIRE 2022 Night Photography Rendering Challenge.

spectrum of HDR (high dynamic range) content. Moreover, there had been other works [18–20] that attempted to replace the entire ISP system with an end-to-end trainable neural network.

Training a robust NN ISP, e.g., either modular function or end-to-end system, required a great amount of paired data, and then significant manpower and resources were devoted to build proper datasets. However, most exiting datasets were collected at daytime, making it not be applicable for applications in the night because of the great illumination variations between daytime and nighttime. Taking the AWB as an example, daytime light sources are primarily strong sunlight, while nighttime light sources are more complex, including a variety of artificial light sources that have never appeared in existing daytime datasets.

To address the lack of paired RAW-RGB data captured at night time, we labeled a high-resolution nighttime RAW to RGB (NR2R) dataset. Specifically, we selected and denoised 150 RAW images with a resolution of 3464×5202 from the training and validation set provided by the night image rendering challenge [13]. Then, we first converted the RAW sample to its corresponding RGB format by a simple ISP for extracting the white patch manually. The white patch of each converted RGB image was used to estimate the ground truth illumination for subsequent AWB. Later then, the denoised RAW image was fed through a serial operations including bilinear demosaicing, auto white balance with label white balance, color space correction with camera inner color correction matrix (CCM) and neural network based local tone mapping to get a 16-bit intermediate RGB image. Finally we import this 16-bit intermediate RGB image into the Lightroom and manually perform adjustments of global exposure, shadows, highlights and contrast accordingly to best reflect the visual preference for deriving the final high-quality 8-bit RGB image.

Unlike daytime images mostly with high illumination, nighttime images were typically acquired under the illumination conditions with complex light sources. We proposed a two-stage CBUnet to cascade the processing of the color and brightness compensation where we used a Unet [32] with channel-attention for color correction at the first stage, and in the second stage we applied a histogram-aware Unet for tone mapping by leveraging statistical brightness information from nighttime images.

Our main contributions can be summarized as follows:

- A high-resolution image dataset with nighttime RAW-RGB (NR2R) pairs that are rendered and annotated by experts with white balance and tone mapping is provided.
- A novel two-stage *CBUnet* to compensate color and brightness attributes consecutively is developed to render acquired RAW images.
- As shown in Table 1, our CBUnet achieved the second best performance in both people's and photographer choice of IEEE CVPR NTIRE 2022 Night Photography Rendering Challenge.

2. Related Work

Over past decades, modular subsystems of a camera ISP system had been extensively examined like the demosaicing [11, 17, 24], denoising [7, 9, 14], white balancing [8, 15, 36], tone mapping [22, 33, 36], etc. Most ISP subsystems had been successfully emulated and enhanced using deep learning techniques by carefully modeling associated function as the image-to-image translation problems.

Demosaicing refers to restoring the color information of the remaining two channels of each pixel through interpolation. Park *et al*. applied residual learning and densely connected convolutional neural network to do color filter array demosaicking, where the proposed model did not require any initial interpolation step for mosaicked input images [29].

Denoising aims to remove the noise and recover the latent observation from the given noisy image. Zhang *et al*. proposed a depth image denoising and enhancement framework using a lightweight convolutional network [40] where a three-layer network model was applied for high dimension projection, missing data completion and image reconstruction. Zhou *et al*. proposed a novel Bi-channel Convolutional Neural Network (Bi-channel CNN) [41] for the same purpose.

White balance follows the color constancy of HVS to eliminate the influence of the color attributes of the light source on scenes acquired by underlying image sensor. Bianco *et al*. used a CNN model to accurately predict the

scene illumination [6] where unlike handcrafted features explored previously, this CNN model accepted spatial image patches for illumination estimation directly. The CCC [3], a.k.a., convolutional color constancy, and its follow-up refinement FFCC [4], a.k.a., Fast fourier color constancy, proposed by Barron *et al*., reformulated the color constancy problem as a 2D spatial localization task in log chromaticity space, and thereby applied object detection and structured prediction techniques to solve it.

Tone mapping operator converts High Dynamic Range (HDR) image to its Low Dynamic Range (LDR) representation for the rendering on normal LDR displays. To train HDR-LDR mapping, a popular approach was to use a tone-mapping operator (TMO) to generate a set of image labels and narrow it down using an image quality index to obtain final training examples [30, 31]. By incorporating the manual supervision, Zhang *et al*. proposed a tone mapping network (TMNet) in Hue-Saturation-Value (HSV) color space to obtain better luminance and color mapping results [39]. Montulet *et al*. introduced a new end-to-end tone mapping approach based on Deep Convolutional Adversarial Networks (DCGANs) along with a data augmentation technique, which reportedly showed the state-of-the-art results on benchmark datasets [27].

As a matter of fact, most works mentioned above mainly paid their attention on daytime image optimization. In contrast, this work deals with the nighttime image rendering by building a nighttime RAW-RGB dataset and developing a two-stage CBUnet to reconstruct "visually more pleasing" images from RAW scenes acquired during the night.

3. Method

3.1. NR2R Dataset

RAW Samples. To form the collection of nighttime RAW samples, we first selected a total of 150 images with the spatial resolution at 3464×5202 from the training and validation sets provided by the night image challenge [13]. And then these RAW images are pre-processed to best produce noise-free samples using a notable CNN based denoiser [1]. This is because nighttime imaging experiences a very challenging situation with heavy noises incurred by high ISO setting under poor illumination condition (e.g., underexposure).

Paired RGB Derivation. We applied a two-stage process to derive the corresponding RGB image of each RAW input.

As shown in Fig. 2, we first used a simple ISP that was comprised of linear demosaicing, gray-world white balance, color correction, and gamma correction to convert each denoised RAW input to its RGB format for groundtruth illumination estimation. To this aim, we mark the "White Patch" from each converted RGB, where the patch is presented in neutral gray, and its RGB channels are approximately the same. Since the gray surface presumably reflects all incoming light radiation, it can be used to represent the ground truth illumination of the RAW image accordingly.

We then perform the 2-stage labeling using the illumination from the 1-stage. Specifically, first we get the correct color image by a serial operations including linear demosaicing, white balance using the label white balance and color correction with the camera inner color correction matrix (CCM). The brightness adjustment consists of local and global tone mapping jointly. Since local tone mapping requires fine-grained adjustment of each small patch in the scene, it is difficult to annotate it manually. Therefore, we use a pre-trained local tone mapping model in [37] to fulfill the task. Since the pre-trained tone mapping network was trained using daytime image, it is good for local adjustment, but fails to control the global brightness. We save the model output using a 16-bit intermediate format in PNG, and then import it into the Lightroom app to adjust the global exposure, brightness, shadows and contrast manually for final high-quality RGB image rendering, with which we emulate the image rendering knowledge from Professional Photographers.

Thereafter, we successfully obtain a high-resolution nighttime RAW-RGB image dataset for the training of CBUnet in next sections.

3.2. CBUnet Architecture

Figure 3 illustrates the architecture of the proposed CBUnet. It consists of two stages, which are cascaded for color correction and brightness adjustment, respectively. The first stage takes an demosaiced noise free RAW image I_{raw} as input. We modified an encoder-decoder based Unet [3] as our primary backbone to estimate the illumination color. The network's deep features are usually too generic without special emphasis even though it is clear that some channel features are more significant than others in different scenes. In order to make the Unet pay more attention to these crucial channels, we employ channel attention-convolutions (CA-Convs) blocks inherited from W-Net [34] to specify the various scenes' response. As shown in Figure 4, CA-Convs block first uses global pooling to extract spatial information from convolutional features, and then transforms them via fully connected layers (FC), ReLU, and sigmoid. At last, it multiplies the convolutional features with sigmoid's output, a.k.a., the weights of the channel attention. All activation functions are implemented with 0.2 negative slope's parametric rectified linear units (PReLU) [16]. We apply global average pooling to get the predict illumination color $\hat{R}$ at the output layer. The input I_{raw} is then corrected based on the $\hat{R}$, and the color space is further converted to CIE-XYZ based on the in-camera CCM. The whole process of stage 1 can be writ-

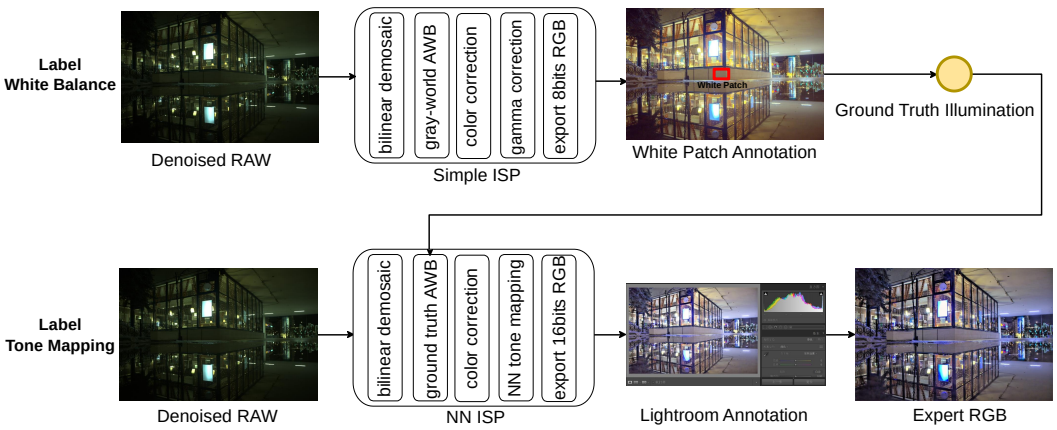

Figure 2. Labeling pipeline. We first perform the white balance annotation to derive the ground truth illumination; then input denoised RAW with estimated ground truth illumination into the NN ISP module (including bilinear demosaic, AWB, color correction, tone mapping) to obtain the brightness-adjusted RGB image, and finally use Lightroom to adjust global exposure, brightness, shadows and contrasts.

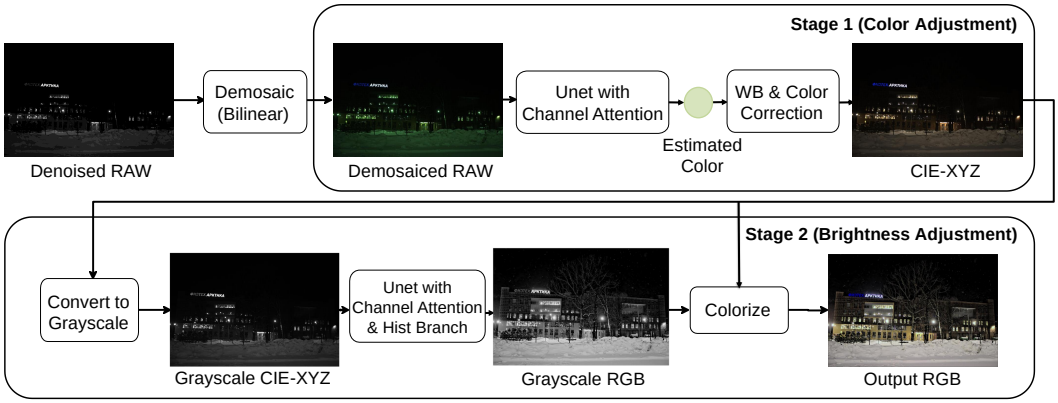

Figure 3. The architecture of the CBUnet. It consist of two stages, where the first stage is used to correct the color and the second one adjusts the brightness for a high-quality output.

ten as:

$$\hat{I}_{cie\text{-}xyz} = CCM \cdot f\left(I_{raw}\right) \cdot I_{raw}, \tag{1}$$

where $f\left(\cdot\right)$ is the illumination color estimation neural network.

The second stage is mainly used to adjust the brightness, so we do not change the color properties of the image at this stage. We first extract the grayscale component of $\hat{I}_{cie\text{-}xyz}$ and then feed it into the brightness prediction network $g\left(\cdot\right)$ to get the target brightness map, and then colorize it according to the color information in $\hat{I}_{cie\text{-}xyz}$. The structure of the $g\left(\cdot\right)$ is similar to the $f\left(\cdot\right)$ in stage 1, but an additional histogram extraction branch is added to obtain the global brightness distribution. Histogram extraction branch uses a 256 bits histogram of the $\hat{I}_{cie\text{-}xyz}$ as input, passes through two fully connected layers with ReLU activation functions, and finally expands to the same size as Unet's bottom feature and sums directly. Thus, the stage 2 could be formulated as:

$$\hat{I}_{rgb} = \frac{g\left(\mathbf{G}\left(\hat{I}_{cie\text{-}xyz}\right)\right)}{\mathbf{G}\left(\hat{I}_{cie\text{-}xyz}\right)} \cdot \hat{I}_{cie\text{-}xyz}, \tag{2}$$

where $\mathbf{G}\left(\cdot\right)$ is the grayscale function.

4. Experiment

4.1. Loss Function

Angular Loss In stage 1, we use angular loss as an evaluation criterion between prediction illumination color $\hat{R}$ [5, 6] and ground truth illumination R:

$$L_{angular} = \frac{180}{\pi} arccos(\hat{R} \cdot R). \tag{3}$$

Pixel Loss In stage 2, we first use L1 Loss to ensure the accuracy of local tone mapping, which is defined as:

$$L_1 = \|\mathbf{G}(\hat{I}_{rgb}) - \mathbf{G}(I_{rgb})\|. \tag{4}$$

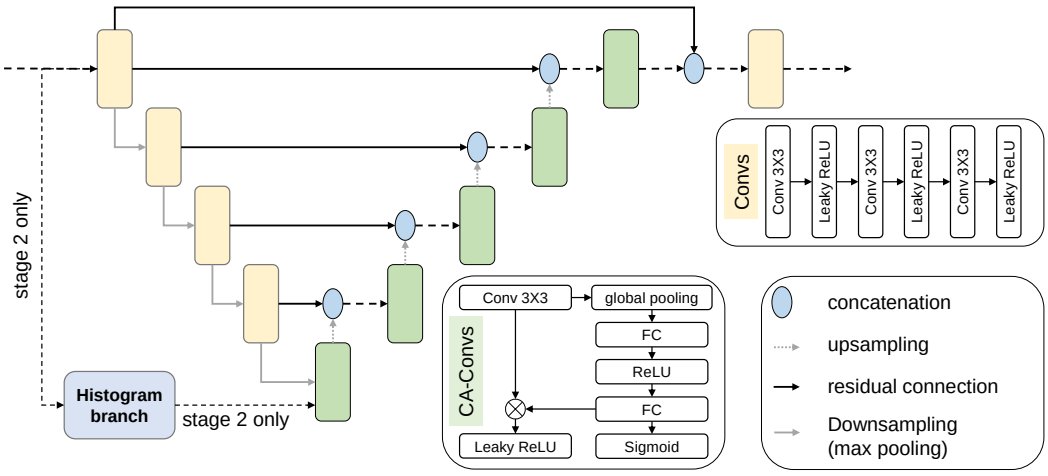

Figure 4. The architecture of our modified Unet. The histogram branch is only used in stage 2.

Histogram Loss However, it is difficult to constrain the global brightness distribution by using L_1 only, so we use the histogram loss to make the histogram of the generated images conform to the statistical distribution of the night-time images. The histogram loss could be formulated as:

$$L_{hist} = \|\mathbf{H}(\hat{I}_{rgb}) - \mathbf{H}(I_{rgb})\|, \tag{5}$$

where $\mathbf{H}$ is differentiable histogram function from [35].

Finally, we define our loss function by the sum of the aforementioned losses as follows:

$$L_{total} = L_{angular} + L_1 + L_{hist}. \tag{6}$$

4.2. Training Details

The model was implemented in Pytorch and was trained on a single Nvidia Tesla 3090 GPU with a batch size 16. We devide the NR2R dataset into two parts: 120 images are used for training and the rest 30 images are used for testing. The stage 1 was first pretrained on Cube++ dataset [12] which is captured within the same camera of the NR2R.Then we use Adam optimizer [23] with 5e-5 learning rate. Then the stage 2 was trained for 300 epochs with the same learning rate while the parameters of stage 1 was frozen. Finally, the stage 1 and stage 2 were joint finetuned for 10 epochs.

4.3. Results

We compare our CBUnet with previous learnt approaches like PyNET [21], HERN [26] and AWNet [10]. As shown in Table 2, we list the reconstruction PSNR (averaged), parameters and floating-point operation (FLOPs). Our CBUnet offers the state-of-the-art efficiency for all metrics. More importantly, our method provides 1.14dB gains over AWNNet [10], but its FLOPs and parameters only present a very small fraction percentage (e.g., <26%). In the mean time, as shown in Figure 5, our method have demonstrated the superior performance to other solution in synthesing the local details, color saturation and contrast.

Method	PSNR	Parameter (M)	FLOPs (G)
PyNET [21]	20.67	47.55	1370.79
HERN [26]	19.57	56.18	466.74
AWNet [10]	21.16	46.99	1532.46
Ours	**22.29**	**23.64**	**391.58**

Table 2. Comparative studies of reconstruction PSNR, parameter size and FLOPs, where FLOPs is calculated when the input size is 1024×1024.

CA	Two Stage	Hist Branch	PSNR
			19.85
√			20.53
√	√		22.01
√	√	√	**22.29**

Table 3. Ablation studies on network architectures, where CA and Hist Branch mean channel attention and histogram branch, respectively.

4.4. Ablation Study

Network Architecture First, we demonstrated the effectiveness of our CBUnet. As shown in Table 3, our most primitive network is Unet. The channel attention brought a gain of 0.68 dB. The two stage design is the most important part of our network performance improvement. The addition of two stage brought 1.48 dB performance gain. Finally, the addition of histogram branch to extract global information results in a gain of 0.28 dB.

(a) RAW (visulized) (b) PyNET (PSNR: 17.25dB) (c) HERN (PSNR: 19.42dB)

(d) AWNet (PSNR: 17.09dB) **(e) Ours (PSNR: 20.39dB)** (f) Ground Truth

Figure 5. Visual comparison of reconstructions from four different methods including PyNET [21], HERN [26], AWNet [10] and our proposed CBUnet.

L_1	$L_{angular}$	L_{hist}	**PSNR**
√			21.98
√	√		22.14
√	√	√	**22.29**

Table 4. Ablation studies on loss functions.

Loss Function Then, we verified the efficiency of our loss function which consists of pixel loss L_1, angular loss $L_{angular}$ and histogram loss L_{hist}. The experimental results are shown in Table 4. When $L_{angular}$ is added, 0.16 dB PSNR increase was achieved. Finally, the addition of L_{hist} also improved the results in a gain of 0.15 dB.

4.5. Generalization to Other Camera Sensors

Although our proposed CBUnet achieved the best performance with minimal computation in the NR2R dataset, but the NR2R dataset was captured by a single DSLR camera. Thus, we test our method on a mobile phone camera to verify the generalization of our method. We took a set of nighttime RAW images with a resolution of 4032×3024 using the main camera of iPhone 8 Plus which has a completely different CMOS and optical lens. Figure 6 shows that our CBUnet has better color correction compared to the iPhone's internal ISP, e.g. zebra lines are correctly corrected to white and the sky is blue. Also, our model renders images with better retention of details in the shadow and better suppression of highlights such as haloes.

4.6. NTIRE 2022 Night Photography Rendering Challenge

As for the NTIRE 2022 Night Photography Rendering Challenge [13], We submitted 100 night images rendered with this method. Rendering results are shown in Figure 7. The baseline ISP provided by challenge organizers consists

Figure 6. The results of the proposed CBUnet on RAW images from the iPhone 8 Plus smartphone. From left to right: the original visualized RAW image, the same photo obtained with iPhone's built-in ISP system and our reconstructed RGB image.

Figure 7. Rendering Results of ours CBUnet in the NITRE 2022 Photography Rendering Challenge [13].

of bilinear demosaicing, bilateral denosing, AWB using gray world assumption, color space conversion with cam-

era inner CCM and flash tone mapping [2]. Compared to the baseline ISP, our renderings have a more accurate white balance and a brightness distribution that is more consistent with nighttime scene characteristics.

5. Conclusion

Night photography is challenging due to the lack of sufficient nighttime image dataset and comprehensive understanding of complex light illumination in the night. This work therefore built a NR2R dataset with dedicated expert annotations as the ground-truth for NN ISP training, and developed a CBUnet for rendering image in the night. The proposed CBUnet showed high-performance and consistent imaging capacity voted by both experts and amateurs, reporting the second place in the NTIRE 2022 Night Photography Rendering Challenge for both People's and Photographer's choices.

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 496–497, 2020. 3

[2] Nikola Banic and Sven Loncaric. Flash and storm: Fast and highly practical tone mapping based on naka-rushton equation. In *VISIGRAPP (4: VISAPP)*, pages 47–53, 2018. 8

[3] Jonathan T Barron. Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015. 3

[4] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 886–894, 2017. 3

[5] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12212–12221, 2019. 1, 4

[6] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017. 1, 3, 4

[7] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005. 2

[8] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 2

[9] Laurent Condat. A simple, fast and efficient approach to denoisaicking: Joint demosaicking and denoising. In *2010 IEEE International Conference on Image Processing*, pages 905–908. IEEE, 2010. 2

[10] Linhui Dai, Xiaohong Liu, Chengqi Li, and Jun Chen. Awnet: Attentive wavelet network for image isp. In *European Conference on Computer Vision*, pages 185–201. Springer, 2020. 5, 6

[11] Eric Dubois. Filter design for adaptive frequency-domain bayer demosaicking. In *2006 International Conference on Image Processing*, pages 2705–2708. IEEE, 2006. 2

[12] Egor Ershov, Alexey Savchik, Illya Semenkov, Nikola Banić, Alexander Belokopytov, Daria Senshina, Karlo Koščević, Marko Subašić, and Sven Lončarić. The cube++ illumination estimation dataset. *IEEE Access*, 8:227511–227527, 2020. 5

[13] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banic, Michael S Brown, Radu Timofte, et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1, 2, 3, 6, 7

[14] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008. 2

[15] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2011. 2

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 3

[17] Keigo Hirakawa and Thomas W Parks. Adaptive homogeneity-directed demosaicing algorithm. *Ieee transactions on image processing*, 14(3):360–369, 2005. 2

[18] Jie Huang, Pengfei Zhu, Mingrui Geng, Jiewen Ran, Xingguang Zhou, Chen Xing, Pengfei Wan, and Xiangyang Ji. Range scaling global u-net for perceptual image enhancement on mobile devices. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 2

[19] Zheng Hui, Xiumei Wang, Lirui Deng, and Xinbo Gao. Perception-preserving convolutional networks for image enhancement on smartphones. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 2

[20] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 536–537, 2020. 2

[21] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 536–537, 2020. 5, 6

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

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[24] Xin Li, Bahadir Gunturk, and Lei Zhang. Image demosaicing: A systematic survey. In *Visual Communications and Image Processing 2008*, volume 6822, pages 489–503. SPIE, 2008. 2

[25] Laurence T Maloney. Physics-based approaches to modeling surface color perception. *Color vision: From genes to perception*, pages 387–416, 1999. 1

[26] Kangfu Mei, Juncheng Li, Jiajie Zhang, Haoyu Wu, Jie Li, and Rui Huang. Higher-resolution network for image demosaicing and enhancing. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3441–3448. IEEE, 2019. 5, 6

[27] Rico Montulet, Alexia Briassouli, and N Maastricht. Deep learning for robust end-to-end tone mapping. In *BMVC*, volume 2, page 4, 2019. 3

[28] Karen Panetta, Landry Kezebou, Victor Oludare, Sos Agaian, and Zehua Xia. Tmo-net: A parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale hdr dataset. *IEEE Access*, 9:39500–39517, 2021. 1

[29] Bumjun Park and Jechang Jeong. Color filter array demosaicking using densely connected residual network. *IEEE Access*, 7:128076–128085, 2019. 2

[30] Vaibhav Amit Patel, Purvik Shah, and Shanmuganathan Raman. A generative adversarial network for tone mapping hdr images. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pages 220–231. Springer, 2017. 3

[31] Aakanksha Rana, Praveer Singh, Giuseppe Valenzise, Frederic Dufaux, Nikos Komodakis, and Aljosa Smolic. Deep tone mapping operator for high dynamic range images. *IEEE Transactions on Image Processing*, 29:1285–1298, 2019. 3

[32] 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. 2

[33] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 13(6):42–48, 1993. 2

[34] Kwang-Hyun Uhm, Seung-Wook Kim, Seo-Won Ji, Sung-Jin Cho, Jun-Pyo Hong, and Sung-Jea Ko. W-net: Two-stage u-net with misaligned data for raw-to-rgb mapping. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3636–3642. IEEE, 2019. 3

[35] Evgeniya Ustinova and Victor Lempitsky. Learning deep embeddings with histogram loss. *Advances in Neural Information Processing Systems*, 29, 2016. 5

[36] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007. 2

[37] Yael Vinker, Inbar Huberman-Spiegelglas, and Raanan Fattal. Unpaired learning for high dynamic range image tone mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14657–14666, 2021. 3

[38] Akiko Yoshida, Volker Blanz, Karol Myszkowski, and Hans-Peter Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision and Electronic Imaging X*, volume 5666, pages 192–203. International Society for Optics and Photonics, 2005. 1

[39] Ning Zhang, Chao Wang, Yang Zhao, and Ronggang Wang. Deep tone mapping network in hsv color space. In *2019 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2019. 3

[40] Xin Zhang and Ruiyuan Wu. Fast depth image denoising and enhancement using a deep convolutional network. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2499–2503, 2016. 2

[41] Erjin Zhou, Haoqiang Fan, Zhimin Cao, Yuning Jiang, and Qi Yin. Learning face hallucination in the wild. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015. 2