Deep-FlexISP: A Three-Stage Framework for Night Photography Rendering

Shuai Liu Chaoyu Feng Xiaotao Wang Hao Wang Ran Zhu Yongqiang Li Lei Lei
Xiaomi Inc., China
{liushuai21, fengchaoyu, wangxiaotao, wanghao35, zhuran, liyongqiang, leilei1}@xiaomi.com

Abstract

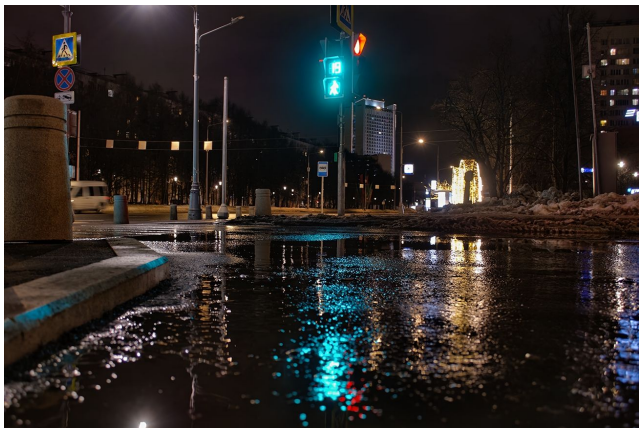
Night photography rendering is challenging due to images' high noise level, less vivid color, and low dynamic range. In this work, we propose a three-stage cascade framework named Deep-FlexISP, which decomposes the ISP into three weakly correlated sub-tasks: raw image denoising, white balance, and Bayer to sRGB mapping, for the following considerations. First, task decomposition can enhance the learning ability of the framework and make it easier to converge. Second, weak correlation sub-tasks do not influence each other too much, so the framework has a high degree of freedom. Finally, noise, color, and brightness are essential for night photographs. Our framework can flexibly adjust different styles according to personal preferences with the vital learning ability and the degree of freedom. Compared with the other Deep-ISP methods, our proposed Deep-FlexISP shows state-of-the-art performance and achieves first place in people's choice and photographer's choice in NTIRE 2022 Night Photography Render Challenge.

1. Introduction

Night photography is a challenging task due to several reasons. First, the low light condition will cause high-level noise in the raw image. Second, it is hard to estimate the accurate white balance in a night scene where multiple illuminants are visible. Third, most night scenes require specific tone curves and photo-finishing strategies to recover the high dynamic range environment.

Image signal processing (ISP) is designed to render raw sensor images to the final image encoded in a standard color space, such as sRGB. ISP is a complex system composed of hand-crafted modules, each of which handles a specific task, such as denoising, white balance (WB), demosaicing, tone mapping, etc. Each module in the traditional ISP contains many parameters that must be manually tuned. Moreover, traditional ISP cannot handle all the scenarios completely, especially for the complex night photography rendering task.

(a) Baseline



(b) Our Deep-FlexISP

Figure 1. Rendering results from the baseline of competition [1] and our proposed Deep-FlexISP.

With the development of deep learning [16], many studies have shown that CNNs have strong competitiveness in many low-level visual tasks. Some of them are closely correlated to ISP modules, including denoising, white balance, color enhancement, etc. An intuitive way is to use individual networks to learn each module in ISP and then concatenate them as a whole framework. However, it may still have the same problem as traditional ISP, the accumulating error [24]. In addition, it is expensive and complex to gen-

erate ground truth for each network.

Recently, Deep-ISP methods [21, 31] are proposed, which train an end-to-end single-stage network to directly accomplish the entire ISP tasks from the raw to sRGB image. It has been shown that Deep-ISP is more straightforward and more effective than the traditional ISP. However, it also has some problems. First, some modules in the ISP are weakly correlated, such as the denoising module handles noise, and the WB module handles color. It will limit the learning capabilities of the network if weakly correlated tasks are mixed. Second, the parameters are fixed after the model is trained. Thus, the only way to adjust output styles is to retrain the whole network with different settings. Especially for night photography rendering, ISP needs to complete a mixed task of subjective enhancement (color, brightness, contrast, etc.) and objective reconstruction (noise, detail, etc.). Thus the single-stage design cannot handle such a complex task well.

This paper proposes a novel three-stage cascaded framework named Deep-FlexISP, which decomposes the complex ISP system into three weakly correlated sub-tasks: raw image denoising, white balance, and Bayer to sRGB mapping. First, task decomposition can increase the learning ability of the whole framework and make it easier to converge to the global optimum. Second, weakly correlated subtasks reduce the coupling between each network and increase the global degrees of freedom. So the individual adjustment of each network does not affect downstream tasks. Furthermore, noise, color, and brightness are the main factors that affect the subjective perception of night photographs. So we can flexibly adjust different network weights, structures, parameters, etc., to get different styles.

Specifically, our contributions include:

- A novel framework for night photography rendering is proposed, named Deep-FlexISP, which decomposes the ISP system into three weakly correlated sub-tasks: raw image denoising, white balance, and Bayer to sRGB mapping.
- Experimental results show that the proposed method outperforms the SOTA Deep-ISP methods. Using the proposed Deep-FlexISP, we achieved the first place in both the people's choice and the photographer's choice in NTIRE22 Night Photography Rendering Challenge.

2. Related work

ISP is a particular system designed for rendering a pleasant and accurate image of the world. The modern ISP systems usually contain many modules [23, 29], including denoising, white balance, demosaicing, tone mapping, etc. The denoising module can increase the signal-to-noise ratio and reduce the noise level of the image. The white balance module ensures that the colors in the image correlate to the field's light sources. The demosaicing module reconstructs a full-color image from the incomplete color samples, which are output from the camera sensor. The tone mapping module is used to map high-dynamic-range images to medium-range images.

Deep learning has been widely used in low-level vision processing tasks, including image restoration [9, 17, 25, 32, 34, 35], enhancement [3, 6, 7, 15], etc. Moreover, it is shown that the deep learning-based methods outperform traditional methods in many aspects. Olaf et al. propose a network with a U-like structure [30], which increases the receptive field while reducing the amount of computation. Zhang et al. propose a self-attention network [36] for super-resolution tasks. Gharbi et al. tackle photographic style enhancement tasks by learning to estimate per-pixel affine mappings in a bilateral grid data structure [15]. In FC4 [17], Hu et al. propose to use a fully convolutional neural network to predict the RGB gain and then apply it to the image to achieve the effect of white balance. Galbi et al. propose a feed-forward network for jointly finishing the denoising and demosaicing tasks [14]. Zhou et al. propose a residual network for jointly demosaicing and super-resolution [37].

In addition, there are some studies on how to learn the entire ISP. In PyNet [21], Ignatov et al. propose to use a single network to simulate the entire ISP and uses Huawei P20 and SLR camera to take paired raw-RGB data for learning, and more researchers improved it later [20]. In [31], Eli et al. use a two-layer structure to process local and global features separately. In CameraNet [24], Liang et al. propose a two-stage framework to handle the restoration and enhancement tasks. Researchers also simplify the model structure to make it easier to deploy to mobile devices [18].

3. Proposed method

In this work, we propose a three-stage cascade framework to solve the night photography rendering task called Deep-FlexISP, which consists of denoising, white balance (WB), and Bayer to sRGB (bayer2rgb) networks. First, task decomposition can increase each network's learning ability and make the whole framework easier to converge. Second, the three parts are decoupled to some extent, so local network adjustment will not affect downstream tasks. This can increase the flexibility of the whole framework. Finally, noise, color, and brightness are the main factors that affect the subjective perception of night photographs. Based on the solid learning ability and flexibility, the proposed framework can better handle the above aspects.

In this section, we first explain the workflow of the entire framework. Next, we discuss why do we do decomposition and why do we decompose such three tasks. Finally, we provide three efficient yet straightforward structures for each network.

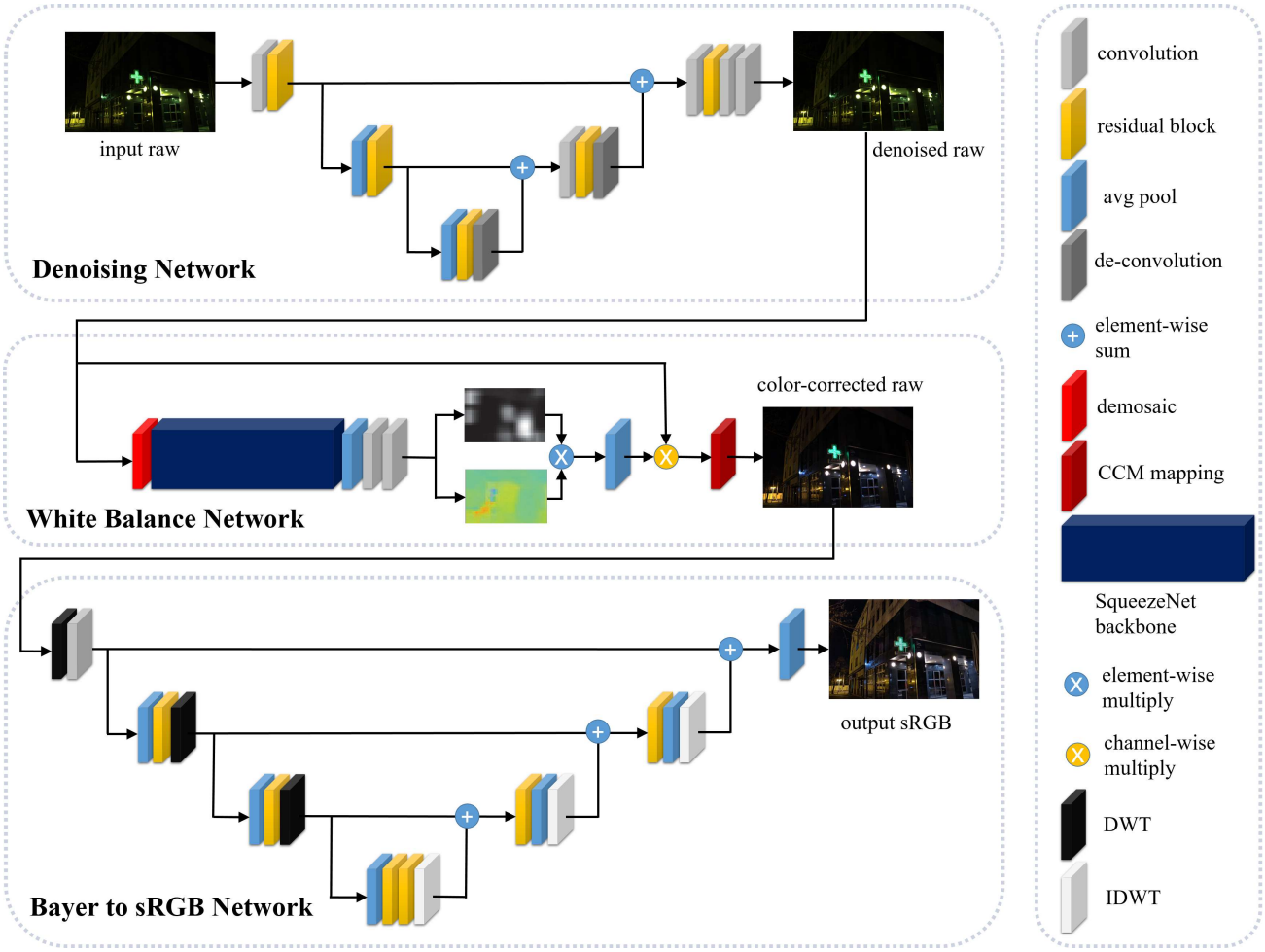


Figure 2. The overall framework of our proposed Deep-FlexISP. It includes three networks: denoising network, white balance network, and Bayer to sRGB network. The input is the raw data captured by the camera, and the output is the sRGB image.

3.1. Overall Workflow

Many studies [2, 4, 5] shown that the denoising performance in the raw domain is better than the RGB domain. Denoising in the raw domain can remove noise better and retain more details. So we first put the original raw image into the denoising network to get a noise-free raw image. Second, the denoised raw image will go through the white balance (WB) network to estimate the RGB gain parameters. After the color is corrected, the Bayer to sRGB (bayer2rgb) network will map the raw image to the final output sRGB image. Bayer to sRGB mapping includes demosaic, tone-mapping, etc.

Our overall framework can be represented by the following formula:

$$I_{output}^{sRGB} = M_{bayer2rgb}(M_{WB}(M_{denoising}(I_{input}^{Raw})))$$

where $M_{denoising}$, M_{WB}, $M_{bayer2rgb}$ denotes denoising network, WB network and bayer2rgb network respectively, and I_{input}^{Raw}, I_{output}^{sRGB} denotes input raw image, output sRGB image respectively.

3.2. Why do we do decomposition

Some modules in the ISP system are weakly correlated. The denoising module controls the noise intensity, the white balance module controls the color, and the tone mapping module controls the global and local brightness. A single-stage network is hard to train to fit the complex ISP, especially night photographs.

The three-stage framework of task decomposition can solve the above problems well. First, the three networks have individual tasks, so each can quickly converge to its own global optimum. Second, the three tasks are weakly correlated, so the assembled framework is closer or more accessible to converge to the global optimum. Our proposed task decomposition framework can better handle complex night photography rendering tasks than single-stage networks.

We also performed corresponding ablation experiments to verify the effectiveness. For details, please refer to Sec-

tion 5.1.

3.3. Why do we decompose such three tasks

The low light conditions of the night will cause high-level noise in the raw image. High-level noise will disturb people's recognition of detailed areas (i.e., buildings, trees) and flat areas (i.e., sky). Noise and details are tradeoffs, requiring a flexible adjustment in some scenarios. So we use an independent network for the denoising task.

Due to the complex lighting environment at night, it is not easy to estimate the color accurately. Moreover, color is also an aspect that influences subjective feeling. The color cast will cause the image to be unreal, deviating from human cognition. Moreover, the range of maneuver in color rending is broad. So we use an independent network for the white balance task.

The night scene image has no sunlight or vital light source, so the overall brightness is relatively dark. Brightness is also one of the aspects that affect people's subjective feelings. An area that is too dark or too light will cause a loss of details, making it impossible for people to obtain the corresponding information. The remaining photo-finishing tasks, including demosaic, tone mapping, etc., are simple interpolation and mapping, so we combine them into one task and name it Bayer to sRGB (bayer2rgb). This can simplify the framework structure.

The raw image denoising, the white balance, and the tone mapping modules in ISP influence the final output image's noise, color, and brightness. In general, these three aspects are weakly correlated to each other. We can obtain night images with different denoising levels, color styles, local contrast, etc., by replacing different network structures, weights, and parameters. As the weak correlation of the three networks, replacing one network will not affect the performance of other networks. It also demonstrates the flexibility of our framework.

We also performed corresponding ablation experiments to verify the flexibility. For details, please refer to Section 5.2.

3.4. Network structure

For the structure of each network, we consider a simple and efficient one, but there could be many possible better designs. The overall architecture of the three networks is shown in Figure 2. For the denoising network, we use 3-level U-Net [30] with residual block, average pooling layer, and additive. We follow studies [25, 28] to remove the bias and batch normalization layers. For the WB network, we use FC4 [17] with an additional demosaic layer and CCM mapping layer. For the bayer2rgb network, we use the MW-ISP [20, 26] network without the upsampling layer.

4. Experiments

4.1. Dataset

Our framework is divided into three networks, each of which is trained separately in supervised form. So for the training phase, different networks use different datasets. The details are as follows.

Denoising network. Since there are only input raw images in the NTIRE night photography rendering challenge dataset [10], supervised training cannot be performed only based on this. We estimate the noise distribution of the input data to construct our own paired images for supervised learning. The noise-free images are collected by ourselves. There are about 150 sets of training data with the different noise levels.

White balance network. We use the public Color Checker Dataset [13] and the NUS 8-Camera Dataset [8] for training. These datasets contain 568 and 1736 raw images, respectively.

Bayer to sRGB mapping network. First, we select some training data from the challenge and the night scene data from cube++ [11] as our training dataset. Second, this dataset is processed by the trained denoising network and WB network, and then it is used as input data of the bayer2rgb network. Third, based on the photographers' opinions [12], we hand-tune some of our ISP modules, including demosaicking, RGB space conversion, tone-mapping, etc. The hand-tuned ISP modules are used to process the dataset to get the corresponding ground truth.

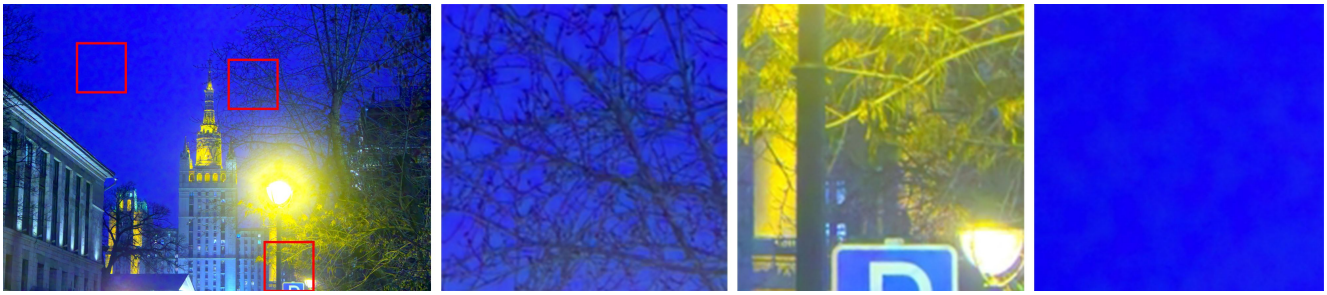
For the testing phase, we use the 100 test data provided by the competition.

4.2. Training settings

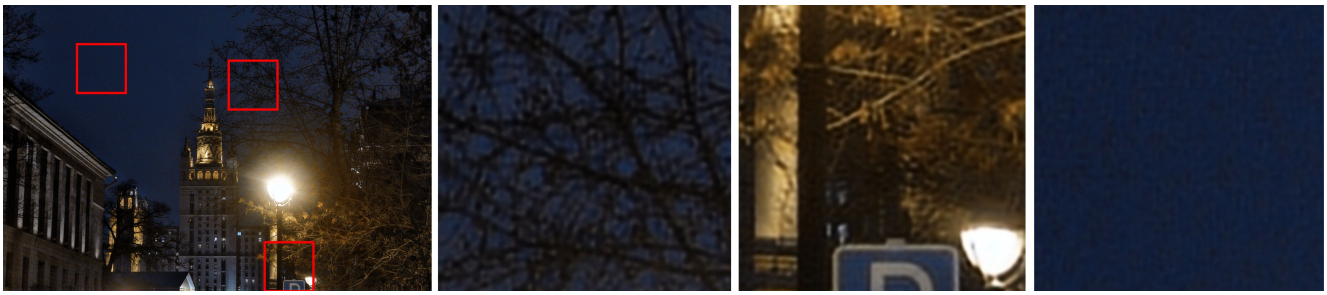
Based on the three-stage framework, We use a two-step training scheme [24]. In the first step, the three networks are trained independently, while in the second step, the three networks are jointly fine-tuned. The special settings are as follows. In the first step, the denoising network is trained 500k iterations with L1 loss. The batch size is set to 1, and the patch size is set to 512. For the white balance network, we use the same configuration as FC4 [17] and add the perceptual loss [22] for training. The bayer2rgb network is trained 200k iterations with L1 loss, SSIM loss [33], and color loss [19]. The batch size is set to 4, and the patch size is set to 1300×866. In the fine-tuning step, we use the entire training data from the challenge for training 50k iterations.

4.3. Qualitative evaluation

We compare the proposed Deep-FlexISP with some open-source SOTA Deep-ISP methods, including the simple ISP baseline from the challenge [1], the PyNet [21], and the HERN [27]. We retrain them using competition data to



(a) Baseline



(b) PyNet

(c) HERN

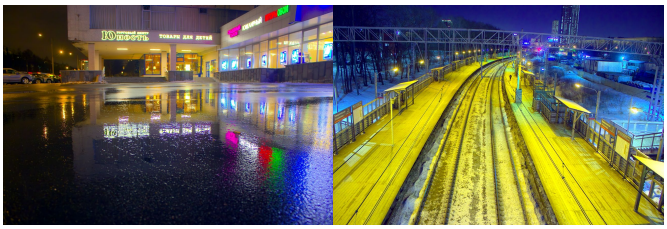
(d) Ours Deep-FlexISP

Figure 3. Comparison of the rendering results from the SOTA Deep-ISP methods and our Deep-FlexISP. Best zoomed on screen.

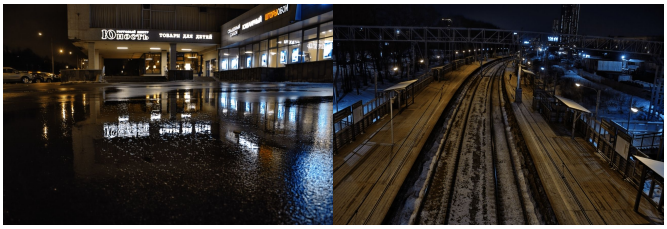
ensure fairness. We provide some comparisons in terms of noise, color, and detail, as shown in Figure 3. Since night scene rendering is a subjective task, we also provide a comparison of artistic renderings, such as color saturation and contrast, as shown in Figure 4.

As shown by the branch area in Figure 3, our framework can reconstruct more details with almost no residual noise. As shown in the street lights and the sign area, we can assume a priori that the sign's color is dark blue, and among all the comparison methods, only our method restores the actual color. As shown in the sky area, our result is flat with no residual noise, and while the baseline result also has no residual noise, its colors are heavily distorted. The above comparison of subjective effects proves the effectiveness of our proposed Deep-FlexISP.

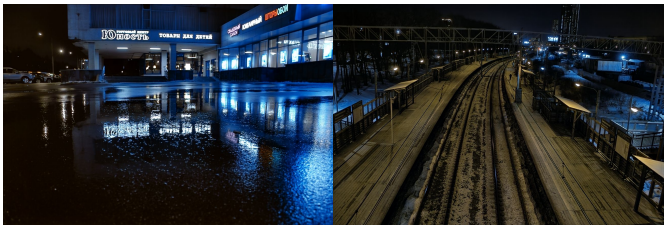
Night scene rendering has much leeway in terms of color saturation and contrast [12], so it is a very subjective task. The two scenes shown in Figure 4 are complex night scenes with multiple light sources. In the left set of the figure, the contrast and saturation of the baseline are too high. The PyNet has low contrast and color cast. The HDRN is rendered in blue overall, which is clearly distorted. However,



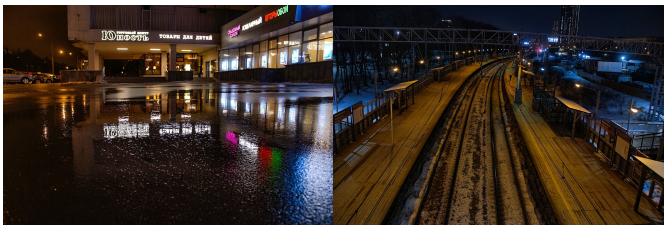
(a) Baseline



(b) PyNet



(c) HERN



(d) Ours Deep-FlexISP

Figure 4. Comparison of the rendering results from the SOTA Deep-ISP methods and our Deep-FlexISP. Best zoomed on screen.

our results are with batter contrast and saturation. The right set of the figure shows almost the same result.

5. Ablation studies

In this section, we construct some ablation experiments to demonstrate the effectiveness and feasibility of our proposed Deep-FlexISP.

5.1. Effectiveness

First, we verify the effectiveness of our Deep-FlexISP. For detailed discussion, please refer to Section 3.2.

We perform ablation experiments on the three networks, denoising, white balance, and bayer2rgb. For fairness, we retrain the networks and change the number of network channels in each experiment so that the comparison networks have the same amount of parameters. The detail settings are shown in Table 1, and the results are shown in Figure 5.

	denoising	WB	bayer2rgb	visual in Figure 5
setting 1	✗	✗	✓	(a)
setting 2	✗	✓	✓	(b)
setting 3	✓	✗	✓	(c)
default setting	✓	✓	✓	(d)

Table 1. Settings of ablation experiments on effectiveness.

We first construct a single-stage structure that only contains the bayer2rgb network to learn the entire ISP process (setting 1). As can be seen from Figure 5 (a). Although the single-stage structure has a certain rendering ability, there will be a serious color cast and residual noise. The above problems arise because the single-stage network cannot take into account the two tasks of noise and color simultaneously. Then, We add an independent white balance network (setting 2). The result is shown in Figure 5 (b). The color cast problem has been solved with the independent white balance network, but the noise still remains. Similarly, we add an independent denoising network (setting 3), as shown in Figure 5 (c), the noise is removed, the flat areas are smooth, and the details are not lost much, but the overall color is slight deviation compared with the setting 2. Finally, we show the result of the default setting. As shown in Figure 5 (d), the three-stage network removes the noise batter and estimates the color more accurately. The framework can take into account the two tasks of noise and color simultaneously. The above experiments demonstrate the effectiveness of our design for task decomposition.

5.2. Flexibility

We also construct ablation experiments to demonstrate the flexibility of our proposed framework. For detailed discussion, please refer to Section 3.3.

To clarify this experiment, we use a simple strategy to adjust networks. We use three parameters to control the strength of denoising, color case, and overall brightness. That is,

$$I_d = M_d(I_{in}) * \alpha + I_{in} * (1 - \alpha)$$

$$I_w = M_{wb1}(I_{denoised_out}) * \beta + M_{wb2}(I_d) * (1 - \beta)$$

$$I_{out} = M_{b2r}(I_w) * \gamma$$

where M_d, M_{wb1}, M_{wb2}, M_{b2r} denotes denoising network, cool style WB network, warm style WB network, and bayer2rgb network respectively, and I_{in}, I_d, I_w, I_{out} denotes input raw image, denoised image, color-corrected image, output sRGB image respectively. The α, β, and γ are the parameters that control noise level, color cast, and overall brightness.

The flexibility of denoising. We get the resulting subfigures with different denoising strengths by controlling α

(a) Setting 1

(b) Setting 2

(c) Setting 3

(d) Default setting

Figure 5. Comparison of rendering results between the different settings. Setting 1 only contains the bayer2rgb network. Setting 2 contains the WB and bayer2rgb network without an independent denoising network. Setting 3 contains the denoising and bayer2rgb network without an independent WB network. The default setting contains denoising, WB, and bayer2rgb network.

and keeping others constant. As shown in Figure 6 (a, b, c), the α of the three subfigures is 0.5, 0.75, 1, the β is 0.5, and the γ is 1. It can be seen that although the residual noise levels are different, the color estimates are relatively accurate, and the brightness is almost the same. This can demonstrate that different denoising strengths do not impact downstream tasks. The noise residue is obvious when the denoising strength is low (the flat area in Figure 6 (a)). However, there is no obvious noise residue when the denoising strength is high, and some details are also erased (weak texture areas on the ground in Figure 6 (c)). This proves that our denoising network can adjust the details and noise flexibly.

The flexibility of white balance. We select two WB networks with entirely different styles as the base (cold and warm colors) and obtain the results of different color shifts by controlling the β. As shown in the Figure 6 (d, e, f), the β of the three subfigures are 0, 0.5, 1, the α is 1, and the γ is 1. It can be seen that the WB networks with different tendencies only change the color, while the brightness and details are almost the same. This proves that different color cases do not affect the subsequent bayer2rgb network. In Figure 6 (d, e, f), the street lamp emits a light source that illuminates the ground, and there are many types of street lamps in the real world, such as yellow chrome lamps ($\beta = 0$) and white nickel lamps ($\beta = 1$). Therefore, different people have different inclinations for color temperature, and the above results are all applicable to this scene. This proves that our white balance network can adjust the color flexibly.

The flexibility of brightness. We get images of different brightness by controlling the γ. Figure 6 (g, h, i) shows that the γ are 0.8, 1, 1.2, and the α and β are 1, 0.5, respectively. It can be seen that the different results differ only in the brightness, which shows that our framework can control the brightness flexibly.

The above experiments prove that simply adjusting each network through parameters will not affect downstream tasks, and multiple aspects can be flexibly adjusted according to personal preferences. Further, more results can be achieved by substituting different network weights or structures. This also illustrates the effectiveness and flexibility of our proposed framework.

6. Night Photography Rendering Challenge

New Trends in Image Restoration and Enhancement (NTIRE), in conjunction with CVPR 2022, has challenges with night photography rendering [10]. Using the proposed Deep-FlexISP, we achieve the first place in both the people's choice and the photographer's choice. In particular, in the people's choice, we are ahead by a considerable margin, which proves the effectiveness of our method. As shown in Table 2, where "Votes" represent the number of subjective votes, the maximum value is 3250.

(a) $\alpha = 0.5, \beta = 0.5, \gamma = 1$ (b) $\alpha = 0.75, \beta = 0.5, \gamma = 1$ (c) $\alpha = 1, \beta = 0.5, \gamma = 1$

(d) $\alpha = 1, \beta = 0, \gamma = 1$ (e) $\alpha = 1, \beta = 0.5, \gamma = 1$ (f) $\alpha = 1, \beta = 1, \gamma = 1$

(g) $\alpha = 1, \beta = 0.5, \gamma = 0.8$ (h) $\alpha = 1, \beta = 0.5, \gamma = 1$ (i) $\alpha = 1, \beta = 0.5, \gamma = 1.2$

Figure 6. Different result of our proposed Deep-FlexISP. The α, β, and γ are the parameters that control noise level, color cast, and overall brightness.

Rank	Method	Votes	Score
1st	**Our Deep-FlexISP**	**2603**	**0.8009**
2nd	-	2047	0.6298
3rd	-	1979	0.6089
4th	-	1964	0.6045
5th	-	1935	0.5955

Table 2. Ranking results of NTIRE 2022 Night Photography Rendering Challenge, people's choice

7. Conclusion

In this paper, we propose a new Deep-ISP method, named Deep-FlexISP, to handle the night photography rendering task. Our method includes three sub-tasks, raw image denoising, white balance, and Bayer to sRGB mapping. Compared with other SOTA methods, our method excels in the aspect of denoising, color correction, brightness correction, contrast, saturation, etc.

References

[1] https : / / github . com / createcolor / nightimaging. 1, 4

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 3

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 2

[4] 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. 3

[5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE con-*

ference on computer vision and pattern recognition, pages 3291–3300, 2018. 3

[6] Qifeng Chen, Jia Xu, and Vladlen Koltun. Fast image processing with fully-convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2497–2506, 2017. 2

[7] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018. 2

[8] 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. 4

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 2

[10] Ershov Egor, Savchik Alex, Shepelev Denis, Banić Nikola, Brown Michael, and Timofte Radu. Ntire 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 4, 7

[11] 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. 4

[12] Michael Freeman. Photographer's opinion. https://nightimaging.org/expert-opinion-second-validation.html. 4, 5

[13] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 4

[14] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 2

[15] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 2

[16] 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. 1

[17] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017. 2, 4

[18] Andrey Ignatov, Cheng-Ming Chiang, Hsien-Kai Kuo, Anastasia Sycheva, and Radu Timofte. Learned smartphone isp on mobile npus with deep learning, mobile ai 2021 challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2503–2514, 2021. 2

[19] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3277–3285, 2017. 4

[20] Andrey Ignatov, Radu Timofte, Zhilu Zhang, Ming Liu, Haolin Wang, Wangmeng Zuo, Jiawei Zhang, Ruimao Zhang, Zhanglin Peng, Sijie Ren, et al. Aim 2020 challenge on learned image signal processing pipeline. In *European Conference on Computer Vision*, pages 152–170. Springer, 2020. 2, 4

[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. 2, 4

[22] 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. 4

[23] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *European Conference on Computer Vision*, pages 429–444. Springer, 2016. 2

[24] Zhetong Liang, Jianrui Cai, Zisheng Cao, and Lei Zhang. Cameranet: A two-stage framework for effective camera isp learning. *IEEE Transactions on Image Processing*, 30:2248–2262, 2021. 1, 2, 4

[25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2, 4

[26] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 773–782, 2018. 4

[27] 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. 4

[28] Sreyas Mohan, Zahra Kadkhodaie, Eero P Simoncelli, and Carlos Fernandez-Granda. Robust and interpretable blind image denoising via bias-free convolutional neural networks. *arXiv preprint arXiv:1906.05478*, 2019. 4

[29] Rajeev Ramanath, Wesley E Snyder, Youngjun Yoo, and Mark S Drew. Color image processing pipeline. *IEEE Signal Processing Magazine*, 22(1):34–43, 2005. 2

[30] 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, 4

[31] Eli Schwartz, Raja Giryes, and Alex M Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline.

IEEE Transactions on Image Processing, 28(2):912–923, 2018. 2

[32] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2

[33] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4

[34] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2

[35] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 2

[36] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 2

[37] Ruofan Zhou, Radhakrishna Achanta, and Sabine Süsstrunk. Deep residual network for joint demosaicing and super-resolution. In *Color and imaging conference*, volume 2018, pages 75–80. Society for Imaging Science and Technology, 2018. 2