NTIRE 2022 Challenge on Night Photography Rendering

Egor Ershov Alex Savchik Denis Shepelev Nikola Banić Michael S. Brown
Radu Timofte Karlo Koščević Michael Freeman Vasily Tesalin Dmitry Bocharov
Illya Semenkov Marko Subašić Sven Lončarić Arseniy Terekhin Shuai Liu
Chaoyu Feng Hao Wang Ran Zhu Yongqiang Li Lei Lei Zhihao Li Si Yi
Ling-Hao Han Ruiqi Wu Xin Jin Chunle Guo Furkan Kınlı Sami Menteş
Barış Özcan Furkan Kıraç Simone Zini Claudio Rota Marco Buzzelli
Simone Bianco Raimondo Schettini Wei Li Yipeng Ma Tao Wang Ruikang Xu
Fenglong Song Wei-Ting Chen Hao-Hsiang Yang Zhi-Kai Huang Hua-En Chang
Sy-Yen Kuo Zhexin Liang Shangchen Zhou Ruicheng Feng Chongyi Li
Xiangyu Chen Binbin Song Shile Zhang Lin Liu Zhendong Wang
Dohoon Ryu Hyokyoung Bae Taesung Kwon Chaitra Desai Nikhil Akalwadi
Amogh Joshi Chinmayee Mandi Sampada Malagi Akash Uppin
Sai Sudheer Reddy Ramesh Ashok Tabib Ujwala Patil Uma Mudenagudi

Abstract

This paper reviews the NTIRE 2022 challenge on night photography rendering. The challenge solicited solutions that processed RAW camera images captured in night scenes to produce a photo-finished output image encoded in the standard RGB (sRGB) space. Given the subjective nature of this task, the proposed solutions were evaluated based on the mean opinions of viewers asked to judge the visual appearance of the results. Michael Freeman, a world-renowned photographer, further ranked the solutions with the highest mean opinion scores. A total of 13 teams competed in the final phase of the challenge. The proposed methods provided by the participating teams represent state-of-the-art performance in nighttime photography. Results from the various teams can be found here: https://nightimaging.org/

1. Introduction

Cameras apply onboard processing to render RAW sensor images to the final photo-finished images encoded in a standard color space (e.g., sRGB). The goal of in-camera processing is to produce visually appealing photographs. Images captured at night present unique challenges that are not typical in most daytime images. For example, in the scenes of daytime images, it is often sufficient to assume a single illuminant, while in the scenes of night images, there are often multiple illuminants present. The unique lighting environment present in night photography makes it unclear which of the illuminants should be taken into account during the correction of scene colors, see Figure 1. In addition, tone curves and similar photo-finishing strategies used to process daytime images may not be appropriate for night photography. Moreover, common image metrics (e.g., SSIM [53] and LPIPS [59]) may not be suitable for night images. Finally, there is significantly less published research focused on image processing for night photography [38]. As a result, there are fewer "best practices" regarding night photography than daytime photography. Because of that, the main motivation of this challenge was to encourage the research targeting night photography. The following sections describe the NTIRE challenge and solutions for the various teams.

This challenge is one of the NTIRE 2022 associated challenges: spectral recovery [6], spectral demosaicing [5], perceptual image quality assessment [26], inpainting [46], efficient super-resolution [35], learning the super-resolution space [41], super-resolution and quality enhancement of compressed video [55], high dynamic range [44], stereo super-resolution [52], burst super-resolution [8].

2. Challenge

Our challenge required teams to develop automated solutions that produce "visually pleasing" images. Teams were provided with a wide range of RAW night images. Teams had to submit corresponding rendered sRGB images. Given the subjective nature of this task, the submissions were

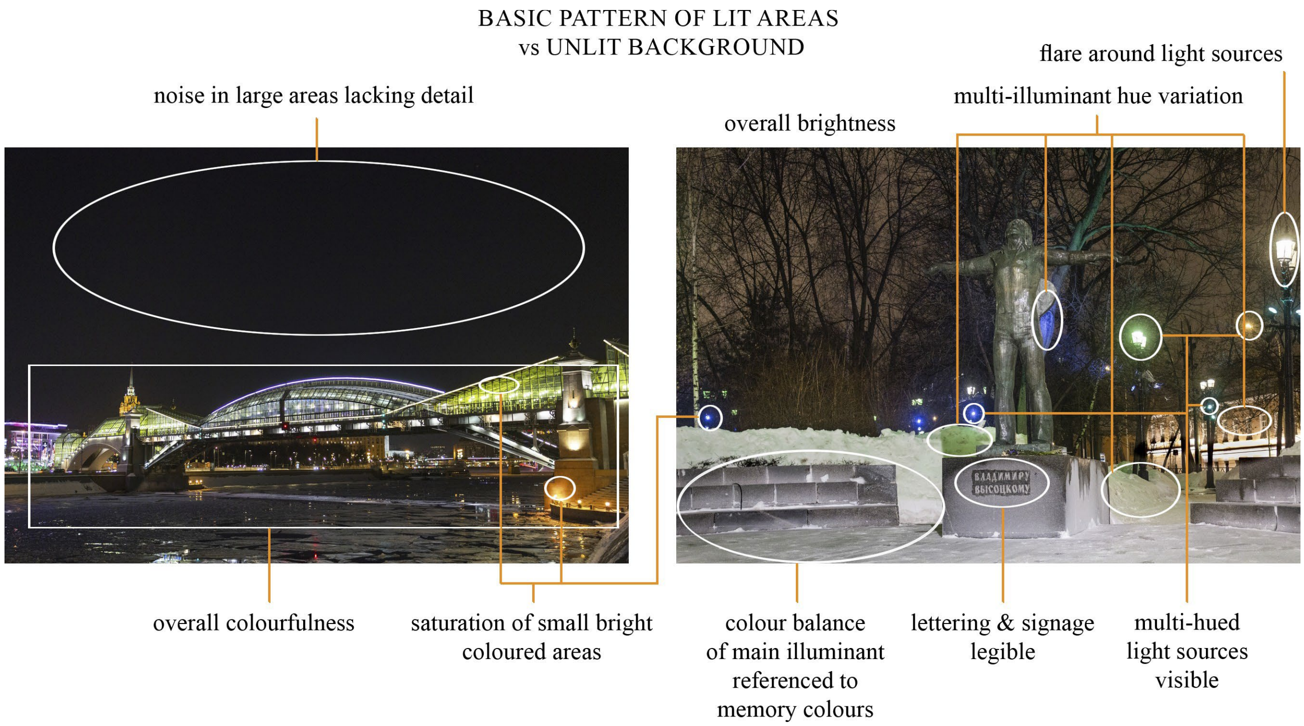

Figure 1. A graphical summary – prepared by Michael Freeman – of the main aesthetic issues in urban night scene photographs.

judged using mean opinion scores from observers who compared and ranked the submissions based on their visual appearance. We were honored to have a world-renowned photographer, Michael Freeman[1], who graciously volunteered his time to judge the top-ranking solutions further.

2.1. Challenge Data

The RAW images of night scenes were captured using the same sensor and encoded in 16-bit PNG files with additional meta-data provided in JSON files. The challenge started with an initial 50 images provided to participants for algorithm development and testing. Additional images were made available during the challenge. Participants were also recommended to use the Cube++ [21] dataset for extra data since it was collected using the same cameras used in this challenge. We provided a baseline code to emulate the basic in-camera rendering as a starting point.

2.2. Evaluation

The evaluation consisted of two validation checkpoints during the contest and a final evaluation to determine the winners. Mean opinion scores were obtained using Toloka (a service similar to Mechanical Turk) for the checkpoints and final evaluation. Toloka users ranked their preferred solutions in a forced-choice manner. It is worth noting that Toloka mainly relied on observers from Eastern Europe and Russia to perform the image ranking. As a result, there may be a cultural bias in terms of the preferred image aesthetics by the observers. Toloka users did not know the identity of the participants.

The results obtained during the validation checkpoints provided feedback to challenge teams on their solutions' quality. During each validation checkpoint, 50 new test images were given. Each participating team was able to send up to two distinct solution image sets, each ranked separately. Note that each of these solutions had to consist of exactly 50 images, namely one solution image per test image. Having two validation sets was intended to help the participants in testing the behavior of different solutions.

For the final submission, 100 test images were made available. Only a single solution image set for the 100 test images was allowed for the final submission. Among these 100 images, only 50 images were used for further evaluation. The indices of the selected images were the same for all participants and given in advance in an encrypted form, with the password provided only after the contest. Additionally, during the final evaluation, the submitted solutions ranked among the top-10 based on the Toloka scores–and that have results reproducible by the provided Docker code–proceeded to the professional judgment stage. In this final stage, Michael Freeman provided his selection of the final

[1] http://www.michaelfreemanphoto.com/

winners.

For the final solution submission, participants were given the option to make their Docker container public after the challenge. Providing a public Docker container was a prerequisite for winning a monetary prize. If the participants kept their Docker container closed, they were still eligible for winner certificates.

2.3. Submissions

As mentioned above, the participants were allowed to submit 50 images in JPEG format (high-quality compression) for each checkpoint evaluation. Each team could submit at most two solution sets. Submissions were available via Google form, which was sent to registered teams.

We expected images of size 1300x866 for landscape orientation and 866x1300 for portrait orientations; images of different sizes were rescaled.

For the final evaluation, a submission had to contain 100 processed images in JPEG format and a Docker container with a runnable solution that could reproduce the submitted results. Among these 100 images, only 50 images were used for further evaluation. The indices of the selected images were the same for all participants and given in advance in an encrypted form, with the password being given only after the contest.

3. Results

The section presents the ranking results obtained using the Toloka service, as well as the ranking performed by a professional photographer.

3.1. People Choice and Discussion

Table 1 provides the ranking of the mean opinion reported by Toloka users for the different Team's final submissions. Our people's choice experience showed that aesthetic evaluation using Toloka provides close to professional ranking, which makes it suitable for user preferences evaluation.

Before we started this challenge, we had concerns regarding three issues that we worried could impact the results. First, we could not control the observation conditions, such as operating system and environment lighting. To help reduce issues here, we selected only users with Windows 10 OS. Second, we had concerns regarding the quality and variations in Tolokers. Despite the fact that the age, nationality and language could be set, there was no guarantee that the tasks were carried out by the same person registered on Toloka. To control this factor, we chose only Tolokers, who had a top 10% rating. Third, we were concerned that technical difficulties might make it hard to provide checkpoint results and a timely final evaluation. This is one reason we asked the solutions to be provided 1300x866 for landscape orientation and 866x1300 for portrait orientation. We found these resolutions suitable for quick download. We also gave the option for Tolokers to flag any images that did not download correctly. Out of the several thousand images downloaded and observed, this happened only a handful of times.

In the end, we were happy with how smoothly the people's choice evaluation worked for both checkpoints and the final judging.

Rank	Team	Mean Score	Votes
1	MIALGO	0.8009	2603
2	Sorashiro	0.6298	2047
3	Feedback	0.6089	1979
4	OzU-VVGL	0.6045	1964
5	IVLTeam	0.5955	1935
6	NoahTCV	0.5742	1866
7	NTU607QCO	0.4798	1559
8	Winter	0.4631	1505
9	Sigma_WHU	0.4411	1433
10	Namecantbenull	0.3965	1288
11	BISPL	0.3683	1197
12	baseline	0.2734	888
13	Low Light Hypnotize	0.0182	59

Table 1. People's choice ranking results.

3.2. Professional Choice and Discussion

Table 2 provides the ranking provided by Michael Freeman. The following describes several factors used to make the final evaluation.

Characteristics of urban night scenes

Urban night scenes are now more significant in photography, not only because lighting itself and signage have become stronger, more varied and more colourful over the years, but because improvements in camera sensors and computational techniques allow easy capture without effort or tripods for many people. Unlike daytime scenes, however, there has been little or no evolution of perceptual experience as to how such scenes should look in a photograph. The principal characteristics are:

1. Large unlit areas.
2. Several-to-many point light sources and speculars.
3. Coloured illuminants, including some with restricted spectrum.
4. Localised high contrast from light pooling e.g. building floodlighting
5. Lighting may be dominated by a single-hue illuminant, or there may be dual-hue illuminants.

Likely rendering issues

1. Artefacting, in particular noise in large featureless areas such as sky, banding, and sharp clipping edges and colour banding around light sources.
2. Colour balance of lit areas. Here, memory colours (canonical colours) can help as references. In descending order of usefulness and reliability, for the night scenes here, they are:
    - Roads, pavements. Assumed to be neutral grey.
    - Concrete. Assumed to be neutral grey.
    - Snow. Assumed to be neutral white with light grey shadows.
    - Clouds, steam, smoke. Assumed to be neutral grey.
    - Clear sky. Assumed to be dark blue, with an HSB hue angle about 216º.
3. Over-saturation of small, bright coloured areas (see below).
4. Deciding overall brightness.
5. Deciding overall colourfulness.
6. Legibility of signage.
7. Suppression of flare around prominent light sources.

Aesthetic expectations

1. Artefact-free
2. Overall fairly neutral colour balance with colourful small elements. If there is any colour cast, blue is more acceptable, while greens (from cyan to yellow-green) are by tradition less acceptable
3. Full tonal range from 1 % above black to white.
4. Unlit and weakly lit areas dark.
5. No clipping except for point light sources and speculars.
6. Overall moderately colourful.
7. Saturation not to reach 100 %, which reads as unrealistic. This is particularly important for night scenes, featuring both light sources and illuminated small areas against an overall dark background, which enhances brightness. Both the Hunt effect (colourfulness increases with luminance) and Helmholtz-Kohlrausch effect (saturation increases with brightness) help exaggerate these.
8. Signage and any significant lettering legible.
9. With all the above in mind, the scene should not look like daytime. In most cases there should already be sufficient clues that it is night-time, but in some cases it might be desirable to lower the overall brightness.

The above represent a broad "envelope" of expectations, but within this there is an as yet undefined latitude for interpretation. This is likely to be in the overall brightness of lit areas, overall colourfulness, and in a multi-illuminant scene, the balance of hue between two (or possibly three) equally important but different illuminants. In the last case, the bias could be toward one of the illuminants or at some point in between.

Rank	Team
1	MIALGO
2	Sorashiro
3	Feedback
4	OzU-VVGL
5	Namecantbenull
6	IVLTeam
7	NoahTCV
8	NTU607QCO
9	Winter
10	Sigma_WHU

Table 2. Professional choice ranking results.

3.3. Teams' solutions

3.3.1 MIALGO team

MIALGO team proposes a three-stage cascaded framework, which includes raw image denoising, white balance processing, and bayer to RGB mapping, see figure 2.

Raw image denoising. Almost all raw images have noise, especially night images. Denoising does not change the brightness, color, and other information of the image, so we denoise in the first stage. Based on U-net [47], we propose a small denoising network, which is more suitable for denoising tasks, and use our own collected data for training. Although the noise distribution of training data and test data is different, the artifacts caused by this gap can be eliminated by the bayer to RGB mapping stage.

White balance processing. We first use a fully convolutional network to estimate the white balance parameters and apply them to the image. We use the Color Checker Dataset [24] and the NUS 8-Camera Dataset [15] for training. Then we multiply the image by a fixed color correction matrix to get a color-corrected image.

Bayer to RGB mapping. After getting the denoised and color-corrected raw image, we process it with some of our developed ISP modules to get an RGB ground truth image, including denoising, demosaic, RGB space conversion, tone mapping, etc. Then we construct a network to learn bayer to RGB mapping between the raw image and the ground truth, the network is modified from MW-ISPNet [31]. During the test, the color-corrected raw image is used as input, and the

Figure 2. MIALGO team pipeline schema.

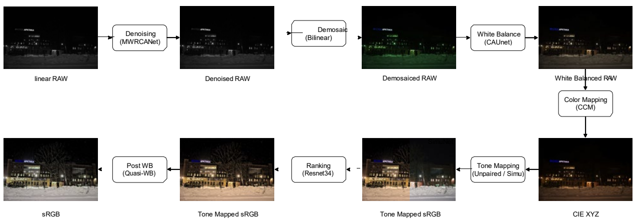

Figure 3. Sorashiro team pipeline schema.

final RGB image can be obtained after processing by the network.

3.3.2 Sorashiro team

Nighttime image processing faces many challenges because of the low light and complex light sources for nighttime photography. These mainly include the noise, the white balance problems brought about by the complexity of the light source, and the tone mapping challenges caused by the great difference between the dark and the light. For this purpose, we have designed a special ISP for nighttime image processing.

Firstly, as shown in the Figure 3, we take the denoising in the RAW domain directly. The denoising model used is the top ranked MWRCANet in NTIRE2020 RAW image denoising [3]. We then demosaicing the denoised image using simple bilinear interpolation. After this we trained a white balance network on the dataset [23] of the same camera to adjust the color contrast. For the white balance model we used CAUnet [36], which was ranked first in the 2nd illumination estimation challenge [23]. After that we convert the white balanced image into a CIE-XYZ domain image by using the color conversion matrix. We believe that tone mapping is a crucial step in nighttime image processing. For this we tried two approaches, the first one we used self-supervised Unpaired-HDR-TMO [51] and the second one we manually labeled the results of tone mapping on the training set and trained a SimuNet to simulate the human labeling process. Therefore, in the tone mappping step we get two outputs, and we select one of them for further processing by a image evaluation model using Resnet34 as backbone. Finally we found that there is a lot of common sense information in night images that can be used to optimize the white balance effect, such as white snow, white smoke, etc. So we use Quai-WB [9] to extract these common sense to further optimize the white balance result and get the final output sRGB.

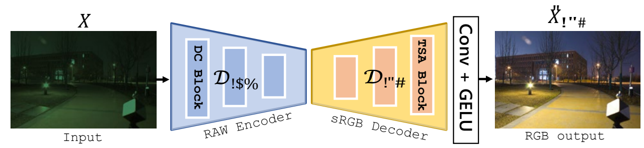

Figure 4. Feedback team pipeline schema. DC Block denotes Depth-wise Convolution Block. TSA Block denotes the Transposed Self-Attention Block.

3.3.3 Feedback team

We address Ultra-ISP (U-ISP). To exploit full advantage of the RAW data, we design our network after a traditional imaging signal processing (ISP) pipeline. Different denoising levels are applied between different channels of the image in traditional ISP pipeline. Followed by this idea, we propose using Depth-wise Convolution (DC) Block to perform the denoising independently with channels on features. The DC Block was inspired by Liu et al. [40]. Channel-wise Transformer [58] has made great progress in the field of image restoration. Our Transposed Self-Attention (TSA) Block was designed after Zamir et al.'s work [58].

The overall architecture (Figure 4) of our model is based on the U-Net [47] framework, which includes an encoder as well as a decoder. With mainly responsible for the denoising task, the encoder is composed with 4 stacked DC blocks at different scales. The decoder consists of 4 stacked TSA blocks, which handle the restorations task at different scales corresponding to the encoders settings.

Most modern dataset for low-light image enhancement in RAW domain need to be multiplied by a fixed enhancement factor, which will limit the network to handle multiple exposures of the same scene. We avoid this situation in two ways: 1) The MESony100 Dataset: Only the image in the SID [13] Sony dataset with fixed enhancement factor equal to 100. Randomly assigned an enhancement factor of 10 to 100 during training 2) The Canon Dataset captured by ourselves, including multiple scenes, with ISO settings from 100 to 6400. Notated that both the MESony100 and the Canon Dataset are retouched by us to meet standards. Then the network is trained end to end with the paired data in our dataset with only L_1 Loss apllied.

3.3.4 VVGL-OzU team

Night images often consist of the scenes with multiple illuminants and are prone to noise due to the inherent structure of sensing tools. To render high-quality night photography, the proposed pipeline blends the common processing stages with more advanced white-balance correction and denoising strategies. Particularly, VVGL-OzU team employs Mixed WB [4] for WB correction, specialized on the scenes with mixed lighting conditions. Also for denoising the sRGB images, we pick SwinIR [37] to include the proposed pipeline, which is a baseline image restoration model based on Swin Transformer [39].

Overall pipeline for VVGL-OzU team's framework is shown in Figure 5. In the scope of this challenge, the input is provided in the linearized 16-bit PNG format, so it is assumed that the camera pipeline initially takes a RAW image and applies the linearization as the first step. The 16-bit PNG image is normalized to correct the black level by provided metadata. Defective hot pixels are corrected with interpolated values based on neighboring pixels of the same color channel. After this operation, the pipeline follows Directional Filtering algorithm proposed by Menon *et al*. [42] for demosaicking the corrected pixel data, instead of default CFA interpolation. The next step in this pipeline is to apply a colorimetric conversion of raw-RGB images to sRGB images. Raw-RGB images are transformed into a standard perceptual color space (*i.e*. CIE XYZ), then converted to sRGB color space. To ensure better WB correction on mixed-illuminant scenes in night images, we include Mixed WB [4] that renders the sRGB images with a small set of predefined white-balance settings and blends the estimated weighting maps to apply correction. Memory color enhancement algorithm [10] is included to the proposed pipeline, which improves the colors of skin, sky, grass, or spot color by hue squeezing. Next, gamma correction is applied to the intermediate images where $\gamma = 0.8$. Flash [7] is selected as tone mapping operator, and applied to the images after gamma correction. Auto-contrast operator is applied to the images for normalizing the image contrast based on their histograms. Then, we include Transformer-based image restoration strategy, namely SwinIR [37], to our pipeline in order to handle the noise in night images. Note that we infer the noise level from the noise profile information in provided metadata. Due to the computational complexity of SwinIR, we have to reduce the size of images (*i.e*. sub-scaling) before denoising. In the next steps, we fix the orientation of the images according to the metadata and further reduce the size of images to the expected image size for the challenge output (*i.e*. 1300×866). Finally, unsharp masking is applied to the sRGB images to sharpen the edges of the final sRGB output.

3.3.5 IVLTeam

The solution proposed by the IVLTeam [60] is based on traditional image processing techniques and is depicted in Figure 6. The entire pipeline can be divided into two parts: the preliminary steps, which are the basic stages of a typical processing pipeline, and the low-light specific part.

The first part works in the RAW domain and is made of four steps: the black and white level image normalization, demosaicing operation, automatic white balancing performed using the simple Gray World algorithm [11], and color space transform from camera specific color space to sRGB color space. This first part corresponds to the original baseline pipeline provided by the organizers, with the denoising step removed.

The second part has been specifically designed to handle images taken by night in low light conditions. The very first step is the use of the Local Contrast Correction (LCC) algorithm from Moroney et al. [43]. Here a local correction is performed using a mask, obtained by blurring with a Gaussian filter the luminance channel of the image (Y channel in YCbCr color space) in order to brighten dark areas and to not clip pixels that are already bright. Since this operation tends to reduce the overall contrast and saturation, the second step consists in a contrast and saturation enhancement using the solution proposed by Schettini et al. [48]. After these two operations, a black point correction step is performed, since local contrast correction adjusts local statistics but produces an overall washed-out result. In order to restore the natural aesthetics of night images, the black stretch operation is performed by clipping to zero the pixels below the 20-th percentile value. A gamma correction is then applied using a gamma value empirically set to 1/1.4, followed by a sharpening operation using unsharp masking.

The image is then converted to uint8 format, resized according to the target resolution, and processed with the Block-Matching and 3-D Filtering (BM3D) denoising algorithm [17]. Here the strength of the denoising operation is controlled by a parameter empirically determined depending on the noise_profile value from the image metadata. The denoised version of the image is blended with the original noisy one using a mask generated from the luminance channel of a YCbCr version of the original noisy image to preserve part of the high frequency information in brighter areas.

Before the final orientation fixing operation, another automatic white balance step is performed using the Grayness Index [45] algorithm to reduce color casts in certain scenarios where the simple Gray World approach may have failed. The image is then rotated in relation to the information stored in the metadata and saved as JPEG image at quality 100.

3.3.6 NoahTCV team

Rendering RAW sensor images captured during night environments to visually pleasing photo-finished images is extremely challenging, because of the complex illuminants, high dynamic range and heavy noise. Conventional image signal processing (ISP) pipeline can not adaptively deal with such complicated scenes. In this work, we propose a two-stage night photography imaging pipeline that com-

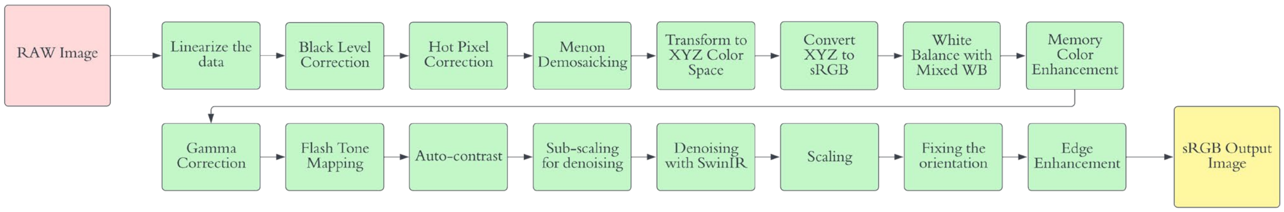

Figure 5. Overall pipeline proposed by VVGL-OzU Team.

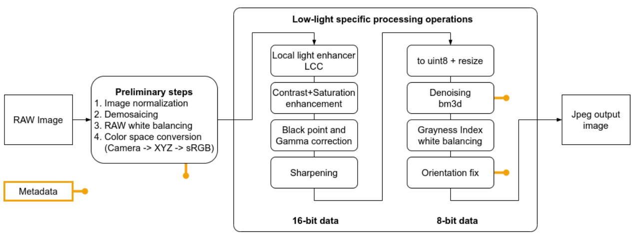

Figure 6. Overview of the complete IVLTeam pipeline. The entire pipeline can be divided in two parts: preliminary data preparation steps and low-light processing steps. Metadata extra information is exploited in the steps marked with the orange dot.

bines the classical ISP pipeline processing and advanced deep-learning-based image enhancement methods.

Specifically, we decompose our night photography imaging pipeline into following two stages, namely traditional preprocessing stage and tone enhancement stage based on a convolutional neural network (CNN). The input RAW data is sequentially processed by preprocessing and CNN enhancement to generate the final enhanced sRGB image for display. We firstly adopt classical ISP pipeline to transform the input RAW image from linear domain into sRGB domain. The modules we apply here are from the officially provided baseline code [1], including (i) black level correction (BLC), (ii) white level normalization (WLN), (iii) demosaicing using simple averaging the two green channels, (iv) white balance (WB) correction, (v) XYZ color space transformation, (vi) followed by sRGB color space transformation and (vii) gamma correction with parameter $\gamma = 2.2$.

We skip the tone-mapping procedure in previous pipeline because a simple tone-mapping function can not adaptively handle such complex environments. Instead, we develop a enhancement model driven by large-scale paired night imaging dataset. Specifically, after we acquire the basic sRGB images, we refine them by *Adobe Camera RAW [2]* to generate the ground-truth images, which are used for later supervised model training. The backbone of our enhancement neural network is RSGUnet [30], which is of a standard Unet structure with one global feature vector and a novel range scaling layer. Moreover, in order to achieve better visual sharpness and definition, we replace all residual blocks with Laplacian enhancement unit (LEU) [29].

We implement our neural network using Pytorch and Adam [33] is selected as the optimizer. All the images are cropped to patches with size 1024×1024 in order to guarantee essential global information during training. And we train the images with the batch size 8 by using the learning rate of 0.0001 on a single NVIDIA-V100 GPU for 10000 epochs. L_1 loss function is applied during training.

3.3.7 NTU607QCO team

Our solution is based on traditional image processing techniques. Our method consists of two main parts: the low-light rendering part and the noise suppression part.

For low-light rendering, first, we adopt the technique in [56] as the backbone which applies multiple stages to achieve effective night-time rendering results. Specifically, it consists of Exposure evaluator, Under-exposed recovery, Exposure optimization, and Exposure fusion.

Exposure Evaluator. The key of night-time image rendering is to improve the low contrast problem in under-exposed regions while maintaining the well-exposed regions. To this end, it is necessary to identify the exposure of each pixel. For the well-exposed regions, we need to assign large weights to maintain the raw image contents. For the under-exposed parts, the smaller weights are adopted since these regions require the image with better exposure to generate desired results. To compute the exposure of each pixel, given an image $\mathbf{I}$, we adopt scene illumination map $\mathbf{K}$. We set the initial value of $\mathbf{K}$ to the lightness component $\mathbf{M}$. $\mathbf{M}(x) = \max_{h\in\{r,g,b\}} \mathbf{I}^h(y)$, where x is the index of each pixel. Then, we further optimize the $\mathbf{M}$ to generate the scene illumination map $\mathbf{K}$ based on the optimization scheme [56].

Under-exposed Recovery. This stage aims to enhance the under-exposed regions by the Camera Response Function. We adopt the Beta-Gamma Correction Model [57] $\mathbf{U}$ which can be defined as: $\mathbf{U}(\mathbf{I}, p) = \exp^{i(1-p^j)} \mathbf{I}^{p^j}$, where i and j are two camera parameters and, p denotes the exposure rate.

Exposure Optimization. To achieve better low-light enhancement results, we need to select the appropriate exposure ratio p in our framework. Initially, we need to

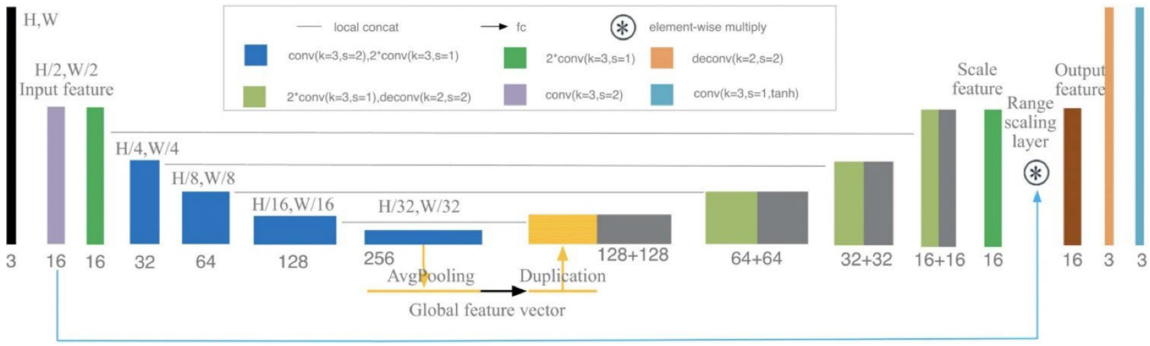

Figure 7. Overall Architecture of RSGUNet [30].

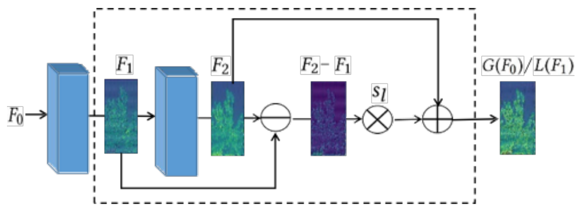

Figure 8. The structure of Laplacian Enhancing Unit (LEU) [29]

find the under-exposed regions. Specifically, we extract these regions by thresholding the estimated scene illumination map $\mathbf{K}$. The values in $\mathbf{K}$ which are smaller than 0.5 are identified as the under-exposed pixels $\mathbf{R}$. Then, we calculate the geometric mean for the enhanced result by the under-exposed recovery in three channels. $\mathbf{Q}^p = \sqrt[3]{\mathbf{U}(\mathbf{R}^r, p) \odot \mathbf{U}(\mathbf{R}^g, p) \odot \mathbf{U}(\mathbf{R}^b, p)}$, where $\odot$ presents the element-wise multiplication. $\mathbf{R}^r$, $\mathbf{R}^g$, and $\mathbf{R}^b$ denote the red, green, and blue channels of the thresholding results. $\mathbf{Q}^p$ is the under-exposed regions with under-exposed recovery in exposure ratio p. Then, we identify the best exposure ratio $p*$ which can generate the most desired results for the final fusion stage. $p* = \operatorname{argmin}_p \sum_k N^k(\mathbf{Q}^p) log(N^k(\mathbf{Q}^p))$, where $N^k(.)$ denotes the normalized histogram counts at value k (we fix $k \in [0, 255]$)

Exposure Fusion. Based on the operations in previous stages, we adopt the original image $\mathbf{I}$ and enhanced-exposed component $\mathbf{U}$ to fuse the final result $\mathbf{F}$. $\mathbf{F}(x) = \mathbf{W} \odot \mathbf{I} + (1 - \mathbf{W}) \odot \mathbf{U}$, where $\mathbf{W}$ denotes the weight map for exposure fusion and it can be computed by $\mathbf{W} = \mathbf{K}^\gamma$. γ presents the enhanced ratio.

Then, we adopt the denoised method called trilateral weighted sparse coding (TWSC) [54] to generate the denoised results.

3.3.8 Winter team

In general, we use a pretrained model to do denoising, white balance and tone mapping, use AdaIN layer to do further color correction and use some traditional method, like gamma correction and auto-contrast, to brighten and enhance the output images. The details of our solution are as follows.

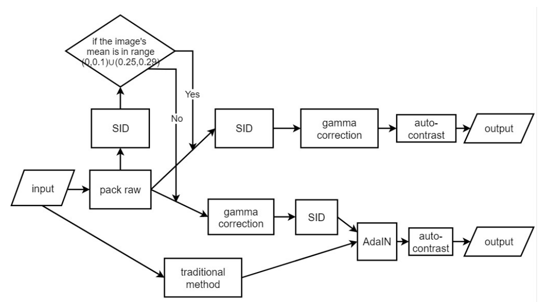

Figure 9. Overall pipeline proposed by Winter Team.

Firstly, we pack the 1-channel raw data into 4 channels (R,G,B,G) following the RGGB bayer pattern and then put the 4-channel data into the pretrained SID model [14] to get the RGB output and then calculate the mean of each image. Then we use some thresholds to divide the images into two groups based on their means. After that, we use different methods to render the images of different groups.

The reason why we do this is that the details of the outputs obtained by the pretrained model are different for the inputs of different means. For example, if the raw data with a larger mean value is highlighted first and then input, it will cause overexposure and lose some details.

The two methods we use are as follows:

SID + gamma correction. After packing the raw data into 4 channels, we let all the images with mean in the range (0,0.1) go through the pretrained SID network and then output. And let all the images with mean in the range (0.25,0.29) go through the pretrained SID network after packing the raw with (R,1.01G,1.01B) and then use gamma correction with coefficient 0.7 to light up the output images.

Gamma correction + SID + AdaIN. We use gamma correction with coefficient $\frac{1}{2.2}$ on the packed raw data first and then go through the pretrained SID and use baseline to correct the color with AdaIN layer. Specifically, AdaIn layer transfers the color of the image without color cast obtained by the traditional method to the denoised but color cast image obtained by SID by transferring the mean and variance of the reference image.

After using the above two methods, we rotate the images to the right direction and use auto-contrast and filter sharpening to further enhance images' quality.

3.3.9 Namecantbenull team

Following the basic camera ISP, our solution pipeline performs image demosaicing, denoising, white balance, and tone-mapping sequentially, as shown in Figure 10. We use traditional methods to realize the first two operations and use CNN-based methods for the latter two operations.

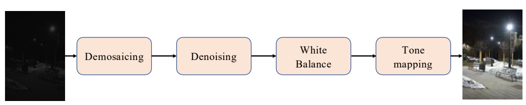

Figure 10. Overall pipeline proposed by Namecantbenull Team.

Image demosaicing. Under the basic RGGB Bayer Pattern, we reshape the RAW image to a three-channel map by copying the R, B component and averaging the two G components.

Image denoising. To perform image denoising, we employ bilateral filter [49], which is an edge-preserving and noise reducing filter. It averages pixels based on their spatial closeness and radiometric similarity.

Image white balance. This operation aims to discover the real white part of the image and discard the color incorrectness. White balance is more challenging at night than daylight. Thus we use a state-of-the-art model proposed in [16], which is a cascaded framework sharing backbone while learning attention specifically in each stage. Since there is no white-balance algorithm developed specifically for night scenes, we train the model on the SimpleCube++ dataset [22] (2234 images, mostly in daytime). To further improve the white balance effect for night scenes, we additionally utilize a statistical algorithm [12] after the network.

Image tone-mapping. In order to get paired data for training the tone mapping network, we use Adobe Photoshop to colorize white-balanced images through the above steps. Then, the white-balanced images are used as the input for the tone mapping network, and the colorized sRGB images are utilized as the ground-truth. We choose HDRNet [25] to train the tone-mapping model.

We implement our network by Pytorch 1.8.0 on GeForce RTX 3090 GPU. We train the white-balance network for 300 epochs with batch size of 16. The Learning rate is set to 3e-4. For the tone-mapping network, we train the model for 1000 epochs with batch size of 4.

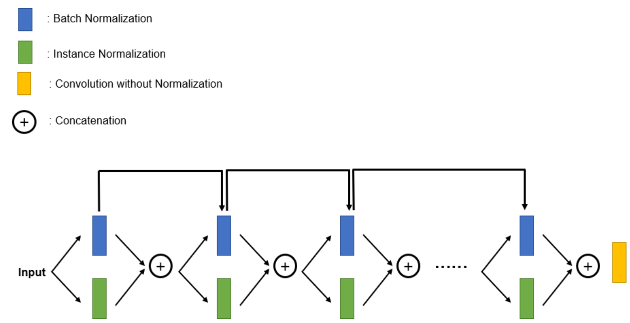

Figure 11. Architecture of BINResNet

3.3.10 BISPL team

A model propose by BISPL is BINResNET(Batch-Instance Normalization ResNet) which is depicted in Figure 11. To train the model, we have pre-processed Cube++ dataset [22], using ground truth illumination value, and obtained raw noisy and white balanced images. The images without mean ground truth values were abandoned. The training process was done by randomly adjusting contrast, saturation and exposure of the white balanced images which are used as target labels. Only the images from Cube++ dataset were used for training, and the images of night scenes were used for fine tuning after the training.

The combination of batch and instance normalization allows the model to capture structure and color of images well. As this task is a highly ill-posed problem, there is no absolute answer that the model should predict. In that reason, it was unable to capture color properly when trained with ResNet [27] generator with batch normalization [32] which is commonly used in many generation tasks. Especially, we could observe that the structure of the original images were well preserved, but the colors were almost lost in many regions. On the other hand, when replaced with instance normalization [50], the colors were preserved, but not the structure. So we have combined those two normalization techniques using repeatedly dividing and concatenating strategy, not to loss both information.

BINResNet receives the input images and start with directly passing them through batch and instance normalization residual block separately. The outputs of each block are concatenated and divided again to repeat the same process. The final outputs of each normalization layer are passed through one convolution layer without any normalization and reconstruct 3-channels RGB image.

As the model predicts the white balanced images with randomly contrast, saturation and exposure adjusted image, we manually modified those values using provided night scene images for fine-tuning after training was done. As a result, the model was able to render the images without using actual night scene images despite only images from

Cube++, which does not contain any data similar to the competition photos, were used. The training was done with learning rate of 0.0005 on a single GeForce GTX 1080 Ti for 100k iterations with the batch size of 8. The images were cropped into 256x256 when training as a loading time of raw image files was too long for efficient training.

3.3.11 Low Light Hypnotize

Images captured in low light conditions finds wide range of applications like night vision in autonomous driving systems, traffic surveillance, wild life photography, drone surveillance, underwater coral reef monitoring and protection. Typically, in underwater scenario intensity of light decreases gradually with depth. Based on this intensity of light and irradiance, Jerlov classified the water types into 10 classes. Authors in [20] [19] [18] [28] consider the classification of water-types as a clue to perform enhancement and restoration of underwater images accordingly. Classes 5C, 7C and 9C belong to the murkish water and intensity of light further decreases nearing to zero. Hence, enhancement and restoration of underwater images specifically belonging to class 5C, 7C and 9C is the need of the hour, towards monitoring and surveillance of aquatic flora and fauna.

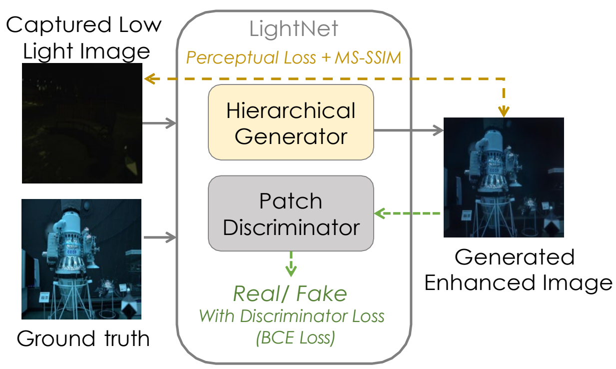

Figure 12. LightNet: Generative model for enhancement of images captured in low-light conditions.

In this work, we propose a generative model for enhancement of images captured in low-light conditions. Camera sensors are often sensitive to the light source during image or video capture. Images of low light conditions find challenges to capture details due to in-sufficient amount of light. Towards this, different deep learning algorithms aim to enhance poorly lit images to generate a high quality image. However, these algorithms capture global features ignoring the underlying local features and hence limiting the performance.

Towards this, we propose a generative model for enhancement of low lit images considering local and global information, and call it as LightNet as shown in Figure 12. The proposed architecture includes an encoder-decoder module to capture global information and a patch discriminator to capture local information as a key towards improving the quality of enhancement. The encoder-decoder module downsamples the input low light image into different scales, to facilitate learning at different levels. Learning at different scales helps to capture the local and global variance of features thereby suppressing the unwanted features (noise, blur). We demonstrate the results of proposed methodology on NTIRE 2022 challenge dataset. We show the enhancement results using different quality metrics.

Unlike the authors in [34], the proposed methodology includes hierarchical generator with corresponding patch discriminator ensuring the retention of local and global features as shown in Figure 12.

Figure 13. Results of proposed methodology (LightNet). 1^{st} row shows input images, 2^{nd} row depicts results of proposed methodology (PSNR: 24.2210 SSIM:0.5766, PSNR: 24.0243 SSIM: 0.7071, PSNR: 25.1962 SSIM:0.7363, PSNR: 26.4667 SSIM: 0.7860, PSNR: 27.1109 SSIM: 0.8134, PSNR: 22.4618 SSIM: 0.7393), 3^{rd} row shows the corresponding ground-truth images.

Results and Discussions The results of proposed LightNet are shown in Figure 13 with corresponding PSNR and SSIM scores.

4. Teams and Affiliations

NTIRE 2022 Team

Team:
Organizers
Members:

Egor Ershov[1], Alex Savchik[1], Denis Shepelev[1], Nikola Banić[2], Michael S. Brown[3], Radu Timofte[4], Karlo Koščević[5], Michael Freeman[6], Vasily Tesalin[1], Dmitry Bocharov[1], Illya Semenkov[1], Marko Subašić[5], Sven Lončarić[5], Arseniy Terekhin[1]
Affiliations:
[1]: Institute for Information Transmission Problems (Kharkevich institute), Russia
[2]: Gideon Brothers, Croatia
[3]: York University, Canada
[4]: Swiss Federal Institute of Technology in Zurich,

Switzerland and University of Würzburg, Germany
[5]: University of Zagreb Faculty of Electrical Engineering and Computing, Croatia
[6]: Michael Freeman Photography

Team:
MIALGO
Members:
Shuai Liu, Chaoyu Feng, Hao Wang, Ran Zhu, Yongqiang Li, Lei Lei
Affiliations:
Xiaomi Inc., China

Team:
Sorashiro
Members:
Zhihao Li, Si Yi
Affiliations:
Nanjing University, China

Team:
Feedback
Members:
Ling-Hao Han[1], Ruiqi Wu[2], Xin Jin[1], Chunle Guo[1]
Affiliations:
[1] Nankai University, China
[2] Wuhan University of Technology, China

Team:
VVGL-OzU
Members:
Furkan Kınlı, Sami Menteş, Barış Özcan, Furkan Kıraç
Affiliations:
Özyeğin University, Turkey

Team:
IVLTeam
Title:
Shallow camera pipeline for night photography rendering [60]
Members:
Simone Zini, Claudio Rota, Marco Buzzelli, Simone Bianco, Raimondo Schettini
Affiliations:
University of Milano-Bicocca, Italy

Team:
NoahTCV
Members:
Wei Li, Yipeng Ma, Tao Wang, Ruikang Xu, Fenglong Song
Affiliations:
Huawei Noah's Ark Lab

Team:
NTU607QCO
Members:
Wei-Ting Chen[1] , Hao-Hsiang Yang[2], Zhi-Kai Huang[2], Hua-En Chang[2] and Sy-Yen Kuo[2]
Affiliations:
[1]: Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2]: Department of Electrical Engineering, National Taiwan University, Taiwan

Team:
Winter
Members:
Zhexin Liang, Shangchen Zhou, Ruicheng Feng, Chongyi Li
Affiliations:
Nanyang Technological University, Singapore

Team:
Namecantbenull
Members:
Xiangyu Chen, Binbin Song, Shile Zhang, Lin Liu, Zhendong Wang
Affiliations:
University of Macau, University of Science and Technology of China

Team:
BISPL
Members:
Dohoon Ryu, Hyokyoung Bae, Taesung Kwon
Affiliations:
Korea Advanced Institute of Science and Technology, Republic of Korea

Team:
Low Light Hypnotize
Title:
LightNet: Generative Model for Enhancement of Low-Light Images
Members:
Chaitra Desai, Nikhil Akalwadi, Amogh Joshi, Chinmayee Mandi, Sampada Malagi, Akash Uppin, Sai Sudheer Reddy, Ramesh Ashok Tabib, Ujwala Patil, Uma Mudenagudi
Affiliations:
Center of Excellence in Visual Intelligence (CEVI), KLE Technological University, Hubballi, Karnataka, India

5. Acknowledgements

This work was supported in part by the Croatian Science Foundation under Grant DOK-01-2018. We thank the NTIRE 2022 sponsors: Huawei, Reality Labs, Bending Spoons, MediaTek, OPPO, Oddity, Voyage81, ETH Zürich (Computer Vision Lab) and University of Würzburg (CAIDAS). We thank our sponsor Huawei for providing funding for the monetary awards given to the winning teams in this challenge. We thank Anton Orlovskii, Irina Zhdanova for their contribution in data collection.

References

[1] https://github.com/createcolor/nightimaging/blob/main/demo/process_img.ipynb. 7

[2] https://www.adobe.com/products/photoshop.html. 7

[3] 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. 5

[4] Mahmoud Afifi, Marcus A. Brubaker, and Michael S. Brown. Auto white-balance correction for mixed-illuminant scenes. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2022. 5,6

[5] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[6] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[7] Nikola Banić and Sven Lončarić. Flash and storm: Fast and highly practical tone mapping based on naka-rushton equation. In *VISIGRAPP*, 2018. 6

[8] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 burst super-resolution challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[9] 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. 5

[10] Karen M Braun. Memory color enhancement algorithm. In *International Congress of Imaging Science*, pages 227–229, 2006. 6

[11] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 6

[12] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1), 1980. 9

[13] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 5

[14] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. *CoRR*, abs/1805.01934, 2018. 8

[15] 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

[16] Xiaodong Cun, Zhendong Wang, Chi-Man Pun, Jianzhuang Liu, Wengang Zhou, Xu Jia, and Houqiang Li. Learning enriched illuminants for cross and single sensor color constancy. *arXiv preprint arXiv:2203.11068*, 2022. 9

[17] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 6

[18] Chaitra Desai, Ramesh Ashok Tabib, Anisha Patil, Samanvitha Karanth, Adarsh Jamadandi, Ujwala Patil, and Uma Mudenagudi. Framework for underwater dataset generation and classification towards modeling restoration. 10

[19] Chaitra Desai, Ramesh Ashok Tabib, Sai Sudheer Reddy, Ujwala Patil, and Uma Mudenagudi. Rendering of synthetic underwater images towards restoration. In *SIGGRAPH Asia 2021 Posters*, SA '21 Posters, New York, NY, USA, 2021. Association for Computing Machinery. 10

[20] Chaitra Desai, Ramesh Ashok Tabib, Sai Sudheer Reddy, Ujwala Patil, and Uma Mudenagudi. Ruig: Realistic underwater image generation towards restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2181–2189, June 2021. 10

[21] 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. 2

[22] 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. 9

[23] Egor Ershov, Alex Savchik, Ilya Semenkov, Nikola Banić, Karlo Koscević, Marko Subašić, Alexander Belokopytov, Zhihao Li, Arseniy Terekhin, Daria Senshina, et al. Illumination estimation challenge: experience of past two years. *arXiv preprint arXiv:2012.15779*, 2020. 5

[24] 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

[25] 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*, 36:118, 2017. 9

[26] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy Ren, Radu Timofte, et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[27] 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 (CVPR)*, June 2016. 9

[28] Deepti Hegde, Chaitra Desai, Ramesh Tabib, Ujwala B. Patil, Uma Mudenagudi, and Prabin Kumar Bora. Adaptive cubic spline interpolation in cielab color space for underwater image enhancement. *Procedia Computer Science*, 171:52–61, 2020. Third International Conference on Computing and Network Communications (CoCoNet'19). 10

[29] Jie Huang, Zhiwei Xiong, Xueyang Fu, Dong Liu, and Zheng-Jun Zha. Hybrid image enhancement with progressive laplacian enhancing unit. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 1614–1622, New York, NY, USA, 2019. Association for Computing Machinery. 7, 8

[30] 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*, September 2018. 7, 8

[31] 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. 4

[32] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 9

[33] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[34] Mohit Lamba and Kaushik Mitra. Restoring extremely dark images in real time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3487–3497, 2021. 10

[35] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, et al. NTIRE 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[36] Zhihao Li and Zhan Ma. Robust white balance estimation using joint attention and angular loss optimization. In *Thirteenth International Conference on Machine Vision*, volume 11605, page 116051E. International Society for Optics and Photonics, 2021. 5

[37] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *IEEE International Conference on Computer Vision Workshops*, 2021. 6

[38] Chang Liu and Xiaolin Wu. Light pollution reduction in nighttime photography. *arXiv preprint arXiv:2106.10046*, 2021. 1

[39] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 6

[40] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *arXiv:2201.03545*, 2022. 5

[41] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 challenge on learning the super-resolution space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[42] Daniele Menon, Stefano Andriani, and Giancarlo Calvagno. Demosaicing with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing*, 16:132–141, 2007. 6

[43] Nathan Moroney. Local color correction using non-linear masking. In *Color and Imaging Conference*, volume 2000.1, pages 108–111. Society for Imaging Science and Technology, 2000. 6

[44] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[45] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8062–8070, 2019. 6

[46] Andres Romero, Angela Castillo, Jose M Abril-Nova, Radu Timofte, et al. NTIRE 2022 image inpainting challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[47] 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, 5

[48] Raimondo Schettini, Francesca Gasparini, Silvia Corchs, Fabrizio Marini, Alessandro Capra, and Alfio Castorina. Contrast image correction method. *Journal of Electronic Imaging*, 19(2):023005, 2010. 6

[49] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998. 9

[50] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 9

[51] 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. 5

[52] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2022 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[53] 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. 1

[54] Jun Xu, Lei Zhang, and David Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *Proceedings of the European conference on computer vision (ECCV)*, pages 20–36, 2018. 8

[55] Ren Yang, Radu Timofte, et al. NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[56] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*, 2017. 7

[57] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017. 7

[58] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv:2111.09881*, 2021. 5

[59] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 1

[60] Simone Zini, Claudio Rota, Marco Buzzelli, Simone Bianco, and Raimondo Schettini. Shallow camera pipeline for night photography rendering. In *Submitted to IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022. 6,11