NTIRE 2023 Challenge on Night Photography Rendering

Alina Shutova Egor Ershov Georgy Perevozchikov Ivan Ermakov Nikola Banić
Radu Timofte Richard Collins Maria Efimova Arseniy Terekhin Simone Zini
Claudio Rota Marco Buzzelli Simone Bianco Raimondo Schettini Chunxia Lei
Tingniao Wang Song Wang Shuai Liu Chaoyu Feng Guangqi Shao Hao Wang
Xiaotao Wang Lei Lei Lu Xu Chao Zhang Yasi Wang Jin Guo
Yangfan Sun Tianli Liu Dejun Hao Furkan Kınlı Barış Özcan Furkan Kıraç
Hyerin Chung Nakyung Lee Sung Keun Kwak Marcos Conde Tim Seizinger
Florin Vasluianu Omar Elezabi Chia-Hsuan Hsieh Wei-Ting Chen
Hao-Hsiang Yang Zhi-Kai Huang Hua-En Chang I-Hsiang Chen Yi-Chung Chen
Yuan-Chun Chiang

Abstract

This paper presents a review of the NTIRE 2023 challenge on night photography rendering. The goal of the challenge was to find solutions that process raw camera images taken in nighttime conditions conditions, and thereby produce a photo-quality output images in the standard RGB (sRGB) space. Unlike the previous year's competition, participants were not provided with a large training dataset for the target sensor. Instead, this time they were given images of a color checker illuminated by a known light source. To evaluate the results, a sufficient number of viewers were asked to assess the visual quality of the proposed solutions, considering the subjective nature of the task. The highest ranking solutions were further ranked by Richard Collins, a renowned photographer. The top ranking participants' solutions effectively represent the state-of-the-art in nighttime photography rendering. More results can be found at https://nightimaging.org/

1. Introduction

In-camera processing is widely used to process raw images obtained directly from the sensor into photographies encoded in a standard color space, such as sRGB. The main objective of this processing is to produce images that are visually pleasing and that simultaneously realistically represent the captured scene. However, nighttime photography presents unique challenges that are not typically encountered in daytime photography. For example, while a single illuminant can often be assumed for daytime images, there are typically multiple illuminants present in nighttime scenes and these can be significantly different. This makes it difficult to determine which illuminant(s) should be primarily taken into account during scene color correction. Moreover, common photo-finishing strategies used for daytime images may not be appropriate for night images due to differences in lighting conditions. Additionally, commonly used image metrics such as SSIM [45], LPIPS [50] or MetaQA [53]) do not appropriately assess the quality of night images. Furthermore, there is a dearth of published research focused specifically on image processing for night photography, resulting in fewer established "best practices" than for daytime photography. Having all this in mind, the main objective of this challenge is, similarly to the previous one, to further encourage research into image processing techniques for night photography. The following sections provide a detailed description of the NTIRE challenge and the solutions proposed by the participating teams.

The Challenge on Night Photography Rendering is one of the NTIRE 2023 Workshop [1] series of challenges on: HR depth from images of specular and transparent surfaces [47], image denoising [29], video colorization [26], shadow removal [42], quality assessment of video enhancement [32], stereo super-resolution [43], light field image super-resolution [44], image super-resolution (×4) [51], 360° omnidirectional image and video super-resolution [9], lens-to-lens bokeh effect transformation [13], real-time 4K super-resolution [16], HR nonhomogenous dehazing [3], efficient super-resolution [28].

[1] https://cvlai.net/ntire/2023/

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

2. Challenge

In the challenge, the participating teams were required to develop automated solutions capable of producing visually appealing images. To this end, the teams were provided with several raw images of ColorChecker SG, captured under lab conditions by a Canon 7D camera. Illumination spectra were also provided to the participants. The teams' objective was to submit the corresponding rendered sRGB images. Given the subjective nature of this task, the submissions were evaluated using mean opinion scores assigned by observers who were presented with pairs of two different renderings of the same scene and who then had to choose the rendering that they deemed visually more appealing.

2.1. Challenge Data

The raw images of night scenes were captured using Canon 7D 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. A baseline code was provided 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 with a question: "Which image is more pleasant?". The answer options were: "left", "right" or "they are the same". To ensure basic quality control, all Toloka users who chose "left" of "right" for a pair of same images have been banned, while all their previous answers have been declined. It is worth noting that in our setup Toloka mainly relied on labelers from Eastern Europe to perform the image ranking. As a result, there may be a cultural bias in terms of the preferred image aesthetics by the labelers. All solutions have been anonymized to guarantee unbiased results.

During each validation checkpoint, 50 new test images were provided, and each participating team could submit up to two distinct solution image sets, each consisting of exactly 50 images. The purpose of having two validation sets was to allow participants to test different solutions' behavior and receive feedback on their solution's quality.

For the final submission, only one solution was allowed, and 50 test images were provided. Additionally, 50 hidden images were used for the final evaluation. The user study images were generated using the code provided by the participants by means of Docker. Only open and reproducible results were accepted. The top 10 solutions according to Toloka proceeded to the professional judgment stage, where Richard Collins provided his selection of the final winners.

3. Results

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

3.1. People Choice and Discussion

Table 1 provides the ranking of the mean opinion reported by Toloka users for the different teams' final submissions.

Before this challenge was started, there were concerns regarding three issues that were suspected to potentially negatively impact the results. First, the observation conditions such as operating system and environment lighting could not be controlled. Second, there were concerns regarding the quality and variations in Tolokers. Despite the fact that limits on age, nationality, and language could be imposed in advance, there was no guarantee that the tasks would be carried out exactly by the person registered on Toloka. To control this factor, only those Tolokers who had a top 10% rating were chosen. Third, there were concerns that technical difficulties might make it hard to provide checkpoint results and a timely final evaluation. This is one of the reasons why the solutions had to be provided 1300x866 for landscape orientation and 866x1300 for portrait orientation. Namely, these resolutions were found to be suitable for quick download. An option was also given to 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, the people's choice evaluation worked for both checkpoints and the final judging smoothly and satisfactory.

Rank	Team	Mean Score
1	IVLTeam	0.67
2	DH_ImageAlgo	0.645
3	MiAlgo	0.626
4	BSSC	0.606
5	DH-AISP	0.583
6	Manual image enhancement	0.491
7	OzUVGL	0.453
8	The Majestic Mavericks	0.444
9	JMUCVLAB	0.439
10	NTU607	0.376
11	Baseline ISP	0.345

Table 1. People's choice ranking results.

This year's competitors have presented a diverse range of solutions that produce visually appealing images. All of the submitted solutions have surpassed the ISP baseline (the winner has a two times higher MOS than the baseline), with only half of them being outperformed by non-professional photographers.

3.2. Professional Choice and Discussion

Table 2 provides the ranking provided by Richard Collins. The following text describes several factors used to make the final evaluation briefly summarized in Fig. 1.

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. Colored illuminants, including some with restricted spectrum.
4. Localized 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 colors (canonical colors) 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 216°.
3. Over-saturation of small, bright colored 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 color balance with colorful small elements. If there is any color 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 colorful.
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 (colorfulness 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 colorfulness, 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	DH_ImageAlgo
3	IVLTeam
4	The Majestic Mavericks
5	BSSC
6	NTU607
7	DH-AISP
8	Manual image enhancement
9	OzUVGL
10	JMUCVLAB

Table 2. Professional choice ranking results.

This year Spearman correlation coefficient between peoples and professional choices is 0.67 (against 0.82 in previous year). Apparently, it is reasonably to argue that this is due to the personal aesthetic criteria of the professional photographers. In part, this effect can also be associated with the inefficiency of mass pairwise comparisons when comparing aesthetically close solutions.

3.3. Teams' solutions

3.3.1 Baselines

In this year's challenge, two baseline methods were given to the participants to use: a simple classic ISP and manual image enhancement. The simple classic pipeline involved debayering using linear interpolation, white balancing with

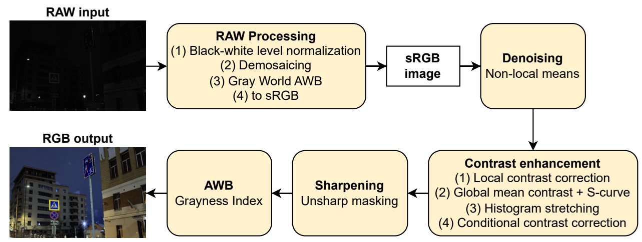

Figure 2. IVLTeam pipeline schema.

Gray World [7], a single matrix for CST, and a standard transform from XYZ to sRGB. This pipeline was also provided as a baseline for participants[2].

To enhance the images manually, we employed the Adobe Camera RAW application and invited non-professional photographers to participate. Each image was corrected individually within a short span of 3 to 5 minutes. The corrections comprised of adjusting the temperature to cool down the image, adding a violet tint, increasing brightness via exposure adjustment, enhancing contrast, reducing highlights, brightening shadows, and reducing whites. Finally, the built-in noise reduction and color mixer were used to correct the hue and intensity of red, orange, and yellow (and sometimes blue and purple).

3.3.2 IVLTeam

Our solution [55] is illustrated in Figure 2. It relies on conventional image processing techniques and consists of five stages, each one addressing different aspects of the image. It is based on our previous work [54], where we improved some critical aspects related to contrast enhancement and color management.

The **first stage** works in the raw domain and consists of four steps: black and white levels image normalization, raw demosaicing operation, Automatic White Balancing (AWB) using the GrayWorld algorithm [7], and conversion from the camera-sensor color space to the sRGB color space.

The **second stage** consists of a denoising operation using the Non-local means algorithm [6]. The intensity of denoising is proportional to the noise standard deviation estimated in the image using the method in [17]. Stronger denoising is applied to the color channels than to the luma channel to effectively remove color noise while preserving image details and edges.

The **third stage** is made of several algorithms that enhance image contrast by manipulating the histogram distribution. First, the Local Contrast Correction (LCC) algorithm in [37] is applied. As this process tends to decrease the overall contrast and saturation, the next step consists of a contrast and saturation enhancement using the approach

[2] available at https://github.com/createcolor/nightimaging23

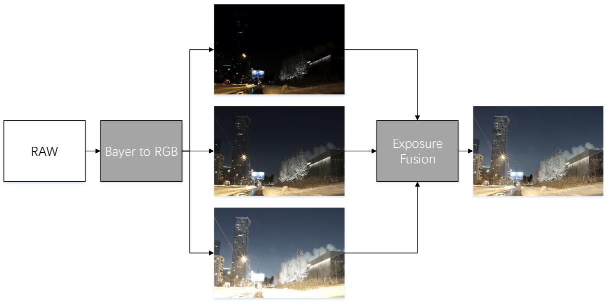

Figure 3. DH_ImageAlgo's algorithm pipeline scheme.

proposed in [41]. Then, three steps to improve the image color appearance are applied. The first one adjusts the contrast by stretching the pixel values by a given factor around their mean. The second one is the application of the S-curve defined in [25], where the center of the curve is set at zero. The third one consists of a histogram stretching operation that increases the dynamic range and improves the overall contrast. An extra conditional contrast correction operation, consisting of an additional S-curve or gamma correction, is applied depending on the mean value of the histogram to improve visibility for very dark images and restore the mood of nighttime scenes when images are too bright.

The **fourth and fifth stages** perform sharpening and AWB, respectively. Unsharp masking is used to sharpen image details, which may have been flattened by the denoising operation in stage two. AWB is performed using the Grayness Index algorithm [39] to reduce color casts in certain scenarios where the Gray World approach may have failed.

3.3.3 DH_ImageAlgo

Our algorithm proposes to aggregate the advantages of the traditional and deep learning approaches to achieve the enhancement effect of night image. The main modules are the raw2rgb module and exposure fusion module. The former is used to learn the projection from raw to RGB with the deep CNN while the latter adopts the traditional fusion strategy to adjust the light distribution, see Figure 3.

Raw2rgb module. It is known that the image captured at night has lower brightness value. To capture more details with better visual effect, a higher gain is commonly used for the sensor, but simple gain multiplication causes noise degradation. To solve this problem, we used a modification of end-to-end U-Net++ model [52] to learn the projection relationship between raw data and GT. Compared with U-Net [40], U-Net++ enjoys better capability to aggregate the multi-scale features to reconstruct more robust results. Meanwhile, the joint constraints of *L1 and adversarial loss* are used to encourage more realistic contents generation. Then, to get different exposure data in the exposure fusion

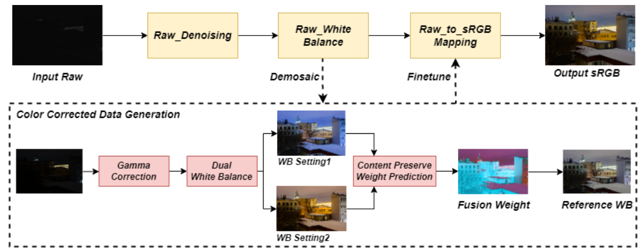

Figure 4. BSSC team pipeline scheme.

algorithm, we adjust the gain of the input data to obtain the underexposed, normal exposed and overexposed images respectively.

Exposure fusion module based on traditional methods [36]. To guarantee the fusion effect, the fusion weight is calculated by considering the contrast, saturation, and brightness of the images with different exposures. More specifically, we generate different exposures Laplace pyramids from the original images, while the Gaussian pyramid is decomposed from the corresponding weight map. To produce the high-quality output, the corresponding components are fused at each scale.

3.3.4 BSSC

The BSSC team proposed a three-stage framework as shown in Figure 4, including raw Image Denoising, raw Image White Balance, and raw to sRGB Mapping with Color Correction. For raw Image Denoising and raw Image White Balance, pretrained models are used [1, 31]. After obtaining the initial color corrected raw images, we process the images with standard ISP modules and adjust the results based on photographer's opinion. The adjusted ISP results are used as ground truth for raw to sRGB Mapping training [24].

The raw to sRGB Mapping model is further fine-tuned to remove color casts caused by scene illumination. Color corrected data is achieved by fusing dual white balance results. Based on the color distribution of the night image dataset, we render the captured scene using two predefined white balance settings as shown in Figure 4. A content-preserve weight prediction module is proposed, which takes two rendered images as input and predicts fusion weight map with the same spatial size.

3.3.5 MiAlgo

We made improvements based on Deep-FlexISP [31], and the overall pipeline is shown in the Figure 5. We keep the denoising and white balance modules unchanged, and split the implementation of the Bayer to sRGB module into several sub-modules, as follows.

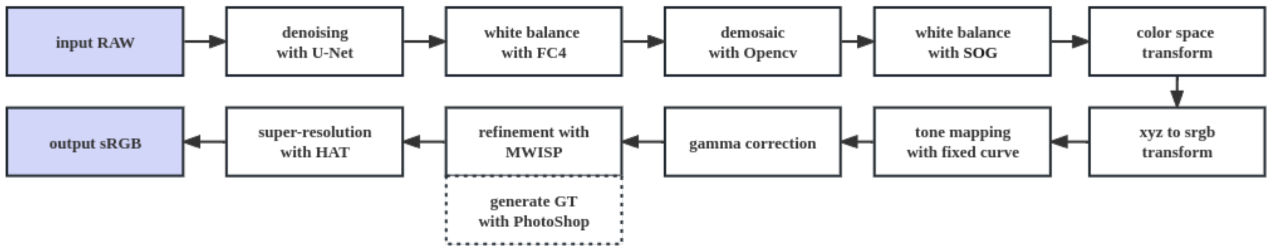

Figure 5. MiAlgo team pipeline schema.

Figure 6. DH-AISP pipeline illustration.

The input image is first processed by the Bayer domain denoising module, which utilizes a simplistic U-Net [40]. Following this, the FC4 [20] network is employed to predict white balance parameters and adjust the Bayer image accordingly. The Bayer image is then subjected to a demosaic operation utilizing the OpenCV [38] library's built-in functions. Since white balance is conducted in the Bayer domain during the second stage, the white balance parameters are predicted and corrected using the Shades of Gray [11] in the RGB domain. The image is then subjected to color space conversion and at this point, it has been converted from Bayer to sRGB. Using a fixed curve, tone mapping is carried out on the sRGB image, and gamma correction with a value of $\gamma = 3.2$ is applied. MWISPNet [24, 30] is used to enhance the image by adjusting brightness, contrast, saturation, and other parameters. The ground truth used in the MWISPNet training is obtained by manual post-processing in PhotoShop [2]. Training high-resolution images is a challenging task, and GPU memory limitations are a significant constraint. As such, we downsample the image, enhance it, and perform super-resolution through Hybrid Attention Transformer [12] before obtaining the output sRGB image.

3.3.6 DH-AISP

Our main goal is to develop a technology for creating realistic and visually pleasing photographs of night scenes. By considering the data quality and modality, we construct the NISP-net, outlined in Figure 6. It contains three parts, involving the data preprocessing module, raw to RGB module, and color enhancement module, which are elaborated in the following parts.

Data preprocessing. This module contains of two steps: the black and white level normalization; automatic white balancing performed using the simple Gray World algorithm [21]. With these two modules, we can obtain a corrected raw data, which is a necessary step to achieve normal and better results.

Raw to RGB. We propose an end-to-end solution for the joint demosaicing, denoising, brightness adjustment, and tone mapping. Specifically, since the low and high-frequency components suffer from the heterogenous degradation, we propose to individually refine them with an elaborated dual-branch U-Net [40], so that the brightness and details of the image can be better processed respectively.

Brightness adjustment and color enhancement. Night images often have multiple illuminants, and light distribution of contents varies with the locations. Consequently, learning a unified light distribution in the existing methods produces undesired results. In this way, we propose a controllable parameter to adjust the output of raw2rgb module, which allows us to produce the underexposed and overexposed candidates with different light distribution. Following that, an exposure fusion model based on U-Net is designed to adaptively learn the fusion weights, and to generate an image with satisfactory overall brightness. Finally, the CCM algorithm is introduced to further optimize the color distribution of the final output.

3.3.7 VGL OzU

The quality of night photography rendering mostly depends on how to handle the multiple illuminants in the scenes and the noise accumulated by consecutive image processing operations. In our ISP pipeline, we mainly concentrated on more advanced white-balance (WB) correction and post-process denoising strategies. We use recently proposed style-based WB correction [27], which can model the mixed illuminant scenarios as the style factor to reverse their effects back to the white-balanced version. For post-process denoising, we include Restormer [48] to our proposed pipeline, which is a Transformer-based efficient restoration architecture that can operate on high-resolution images. Also, we introduce a new auto-contrast strategy for night photography, which dynamically adjusts the min-

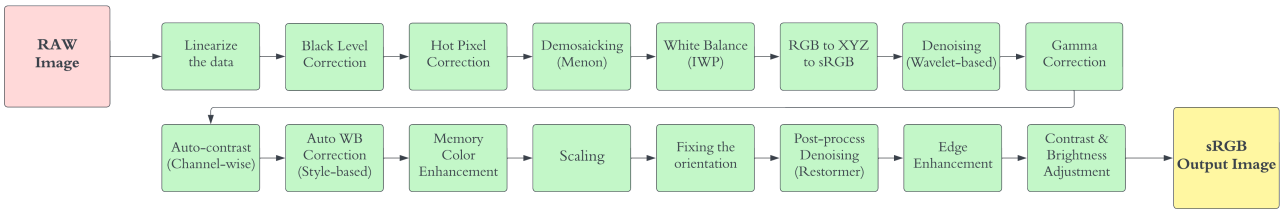

Figure 7. Illustration of our proposed ISP framework.

imum cut-off values by the outliers in the histogram of the input image. The proposed ISP pipeline for this challenge is presented in Figure 7.

In the scope of this challenge, we assume that given raw images are pre-linearized by the camera pipeline since the linearization tables for all inputs are null in the given metadata. Therefore, we did not apply any linearization step to the given data. The 16-bit PNG images are given as the data, and we normalized them to correct the black level by the values provided in the metadata. We then applied hot/bad pixel correction to the data to transform the possible defective pixels by their neighboring pixels. Instead of default CFA interpolation, the pipeline follows the Directional Filtering algorithm [35], also known as *Menon*, for demosaicing the corrected pixel data. We included the random subsampling-based White Patch algorithm [4] to our pipeline in order to initially attempt to mitigate the illumination effects in the raw-RGB domain. The next step is to transform the image from the raw-RGB domain to the sRGB domain. To achieve this, we employ Color Component Transfer Function (CCTF) Encoding. Before applying the Gamma correction method given by the standard pipeline, we applied wavelet-based denoising [10] to the encoded data, where adaptive noise thresholding is used for computing different thresholds for each wavelet subband. To increase the quality of rendering lit areas in darker scenes, we adapt the default auto-contrast strategy to the one that dynamically adjusts the minimum normalization cut-off by considering the outliers in the lower part of the histogram. Next, we included a style-based WB correction algorithm [27] to further mitigate the effects of different illuminants in a scene that frequently occurs in night images. To neutralize the colors of sky, grass, or spot colors, the memory color enhancement algorithm [5] is applied to the WB-corrected images in our proposed pipeline. Before post-process denoising, we fixed the orientation of the images as given in the metadata, and then rescaled the images to the expected image size for the challenge output (*i.e.* 5202×3464). Transformer-based efficient restoration model [48], namely *Restormer*, was used for removing the noise that arose during the previous operations applied. Due to the computational complexity of Restomer in our resolution space, we have to split the images into grids and apply denoising to these grids one by one. To obtain the final sRGB output, we applied unsharp masking to sharpen the edges and adjusted the contrast and brightness values to enhance the quality of the images.

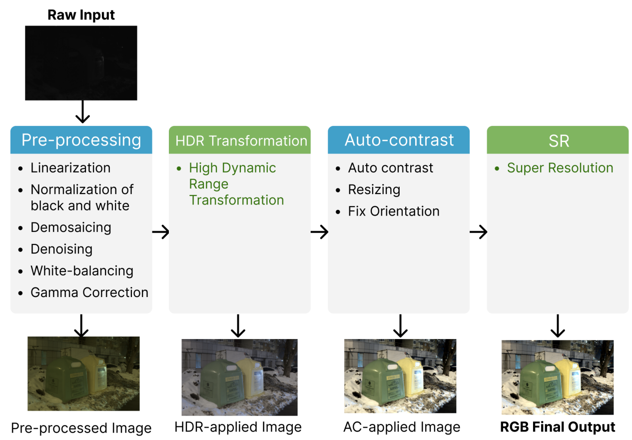

Figure 8. Pipeline scheme of The Majestic Mavericks.

3.3.8 The Majestic Mavericks

Since night images are typically captured under low-light conditions, they have low contrast and high noise. Therefore, we pay special attention to enhance contrasts of colors and to minimize noise in images.

As depicted in Figure 8, our pipeline is largely composed of four modules: Pre-processing, HDR (High Dynamic Range) transformation, Auto-contrast, and SR (Super Resolution).

First, the **pre-processing module** align with those in the baseline model. It involves the use of OpenCV to implement linearization of raw images, normalization of black and white, demosaicing, denoising, white-balancing, and gamma correction.

Next, the **HDR (High Dynamic Range) transformation module** spots the luminosity of the brightest and darkest area in the photo. For instance, in the example shown in Figure 8, the color difference between the yellow and green trash bins is emphasized, making them more distinguishable. Furthermore, it can expand the range of colors in areas with large uniform colors, such as roads and snow, thereby improving the ability to distinguish shadows in those regions. Our implementation employs ExpandNet [34] for the conversion of images to HDR. In this way we are able to work on the contrast which is vital for the reconstruction of images. Moreover, we are able to find the pixels which can give great contrast to image.

Auto-contrast module calculates an image's histogram and then adjusts the pixel values to set the darkest pixels to black and the brightest pixels to white, while disregarding a specified percentage of the extreme pixel values. It additionally carries out operations to resize and correct orientation.

Lastly, we incorporated a deep learning-based **super resolution technique** into our approach, which boosts the resolution but also reduces noise and optimizes the tonal balance. We used the BSRGAN [49], which was trained on a degraded image dataset that closely resembles real-world image degradation.

3.3.9 JMU-CVLAB

We focused on creating an efficient rendering pipeline able to run on the mobile devices, starting from a model-based ISP [14, 18]. We achieve this by simple knowledge distillation between a teacher network and student network. As a teacher, we use Deep-FlexISP [31] the winner of NTIRE 2022 Challenge on Night Photography Rendering [18], a multi-stage network that includes custom denoising, white balance and learned ISP. For the student network we select MicroISP [22], which is the current state of the art for efficient end-to-end ISP networks, designed to process high-resolution raw images on mobile devices.

Knowledge Distillation We create the ground truth images by processing the raw images using Deep-FlexISP [31].

We process 150 raw images, and we split the dataset using a holdout 80/20 split. Next, we pre-process the full resolution images into batches of size 256×256 for training our student network MicroISP [22].

The student ISP network is trained to process raw images and render pleasant sRGB night images. Note that we used MicroISP network pre-trained of MAI Learned Smartphone ISP challenge [23] and fine-tuned the model using our curated dataset. For fine-tuning the student network we combine: MSE, SSIM, and Perceptual VGG losses. We fine-tuned the student model for 500 epochs with a learning rate of 1e-5 using a NVIDIA RTX 3090 Ti ($\approx$ 4 hours of training).

Because we train the network using patches, we realized about some limitations such as vignetting effect, and an overall dimmer image than the teacher network — this was also pointed out by the professional photographer judge. Finally, our student network has $\approx 100\times$ less parameters than Deep-FlexISP [31], and can emulate its results with a reconstruction of 24.36dB PSNR (on our 20% testset). Additional experiments for low-light image enhancement [15, 19] did not produce pleasant results.

3.3.10 NTU607

Our solution is based on the deep learning solution Zero-Reference Deep Curve Estimation (Zero-DCE) [19]. It is known as the self-supervised learning framework for brightness curve estimation.

First, we use the guideline from organizers to transform the raw image to sRGB images. Resulting sRGB images were used as inputs. We use the following quadratic equations to estimate pixel-wise brightness:

$$J(x) = I(x) + \alpha I(x)(1 - I(x)), \tag{1}$$

where $J(x)$ and $I(x)$ are output and input images, while α is the curve parameters that should be learned. To make the model handle challenging low-light conditions, we consider iterative curve estimation:

$$J_n(x) = J_{n-1}(x) + \alpha J(x)_{n-1}(1 - J(x)_{n-1}) \tag{2}$$

where n is the number of iterations (in our solution $n = 8$). To get the α, we apply a plain CNN containing seven convolutional layers and Each layer consists of 32 convolutional kernels of size 3×3 and stride 1 followed by the ReLU activation function. The last convolutional layer is followed by the Tanh activation function. We use *spatial consistency loss, exposure control loss, color constancy loss, and illumination smoothness loss* to optimize the network. The spatial consistency loss encourages spatial coherence of the enhanced image by preserving the difference of neighboring regions between the input image and its enhanced version. The exposure control loss measures the distance between the average intensity value of a local region to the well-exposedness level controls the exposure level. The color constancy loss is designed to correct the potential color deviations in the enhanced image and also build relations among the three adjusted channels. The illumination smoothness loss preserves the monotonicity relations between neighboring pixels.

Implementation details. Besides the dataset provided by organizers, we also use the DARK FACE dataset [46], and underexposed images from the SICE dataset [8]. A batch size of 8 is applied. The filter weights of each layer are initialized with standard zero mean and 0.02 standard deviation normal noize. Bias is initialized as a constant. AdamW [33] is used as an optimization algorithm with a mini-batch size of 8. We set the initial learning rate to

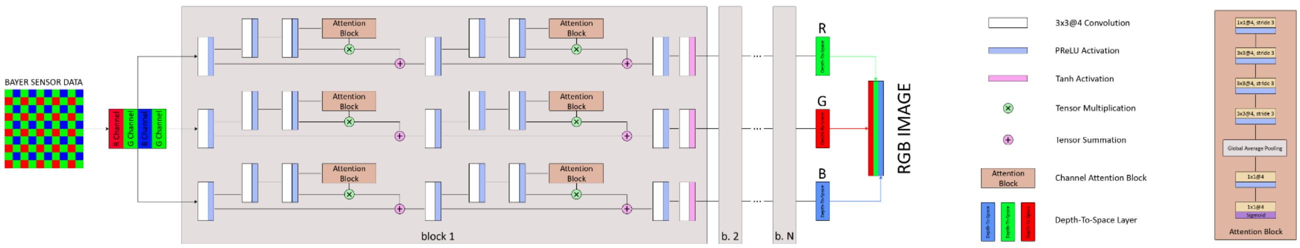

Figure 9. MicroISP Architecture [22].

0.0001 and it is unchanged. The models are trained for 200 iterations. The overall framework is implemented with PyTorch on an NVIDIA V100 GPU.

4. Teams and Affiliations

Team:
Organizers
Members:
Alina Shutova[1], Egor Ershov[1], Georgy Perevozchikov[1], Ivan Ermakov[1], Nikola Banić[2], Radu Timofte[3], Richard Collins, Maria Efimova[1], Arseniy Terekhin[1]
Affiliations:
[1]: Institute for Information Transmission Problems (Kharkevich institute), Russia
[2]: Gideon Brothers, Croatia
[3]: Swiss Federal Institute of Technology in Zurich, Switzerland and University of Würzburg, Germany

Team:
IVLTeam
Members:
Simone Zini[1], Claudio Rota[1], Marco Buzzelli[1], Simone Bianco[1], Raimondo Schettini[1]
Affiliations:
[1]: University of Milano - Bicocca, Italy

Team:
DH_ImageAlgo
Members:
Chunxia Lei[1], Tingniao Wang[1], Song Wang[1]
Affiliations:
[1]: Zhejiang Dahua Technology, China

Team:
MiAlgo
Members:
Shuai Liu[1], Chaoyu Feng[1], Guangqi Shao[1], Hao Wang[1], Xiaotao Wang[1], Lei Lei[1]
Affiliations:
[1]: Xiaomi Inc., China

Team:
SCBC
Members:
Lu Xu[1], Chao Zhang[1], Yasi Wang[1]
Affiliations:
[1]: Beijing Samsung Teleconmmunication R&D Center, China

Team:
DH-AISP
Members:
Jin Guo[1], Yangfan Sun[1], Tianli Liu[1], Dejun Hao[1]
Affiliations:
[1]: Zhejiang Dahua Technology, China

Team:
VGL OzU
Members:
Furkan Kınlı[1,2,3], Barış Özcan[1,2], Furkan Kıraç[1,2]
Affiliations:
[1]: Vision and Graphics Lab, Özyeğin University, Türkiye
[2]: Fishency Innovation, Norway
[3]: T-Fashion Inc., Türkiye & Canada

Team:
The Majestic Mavericks
Members:
Hyerin Chung[1], Nakyung Lee[1], Sungkeun Kwak[1]
Affiliations:
[1]: CJ OliveNetworks AI Research, Republic of Korea

Team:
JMU-CVLAB
Members:
Omar R. Elezabi[1], Marcos V. Conde*, Radu Timofte[1]
Affiliations:
[1]: University of Wuerzburg, Computer Vision Lab, CAIDAS, Germany
**** Team leader & Corresponding author:***
* marcos.conde@uni-wuerzburg.de
Code and Results:
`https://github.com/mv-lab/AISP`

Team:
NTU607
Members:
Chia-Hsuan Hsieh[4], Wei-Ting Chen[1], Hao-Hsiang Yang[2], Zhi-Kai Huang[2], Hua-En Chang[2], I-Hsiang Chen[2], Yi-Chung Chen[3], Yuan-Chun Chiang[2], Sy-Yen Kuo[2]
Affiliations:
[1]: Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2]: Department of Electrical Engineering, National Taiwan University, Taiwan
[3]: Graduate Institute of Communication Engineering, National Taiwan University, Taiwan
[4]: ServiceNow, USA

5. Acknowledgements

This work was supported by Toloka Research Grants program [3]. We thank the NTIRE 2023 sponsors: Sony Interactive Entertainment, Meta Reality Labs, ModelScope, ETH Zürich (Computer Vision Lab) and University of Würzburg (Computer Vision Lab). Authors also want to thank Artem Pogadaev for his helping with report preparation.

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, and Michael S Brown. NTIRE 2020 challenge on real image denoising: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 5

[2] Adobe Systems Incorporated. Photoshop. https://www.adobe.com/products/photoshop.html, 2021. Accessed: 2021-11-01. 6

[3] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2023 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[4] Nikola Banić and Sven Lončarić. Improving the white patch method by subsampling. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014. 7

[5] Karen M Braun. Memory color enhancement algorithm. In *International Congress of Imaging Science*, 2006. 7

[6] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2. IEEE, 2005. 4

[7] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, 1980. 4

[8] 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. 8

[9] Mingdeng Cao, Chong Mou, Fanghua Yu, Xintao Wang, Yinqiang Zheng, Jian Zhang, Chao Dong, Ying Shan, Gen Li, Radu Timofte, et al. NTIRE 2023 challenge on 360° omnidirectional image and video super-resolution: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[10] S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000. 7

[11] Kun Chen, John Snyder, Jiayi Wang, and James Davis. Shades of gray: A pipeline for software shading on commodity 3d hardware. *ACM Transactions on Graphics (TOG)*, 27(3):1–9, 2008. 6

[12] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating more pixels in image super-resolution transformer. *arXiv preprint arXiv:2205.04437*, 2022. 6

[13] Marcos V Conde, Manuel Kolmet, Tim Seizinger, Thomas E. Bishop, Radu Timofte, et al. Lens-to-lens bokeh effect transformation. NTIRE 2023 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[14] Marcos V Conde, Steven McDonagh, Matteo Maggioni, Ales Leonardis, and Eduardo Pérez-Pellitero. Model-based image signal processors via learnable dictionaries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2022. 8

[15] Marcos V Conde, Florin Vasluianu, Javier Vazquez-Corral, and Radu Timofte. Perceptual Image Enhancement for Smartphone Real-Time Applications. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1848–1858, 2023. 8

[16] Marcos V Conde, Eduard Zamfir, Radu Timofte, et al. Efficient deep models for real-time 4k image super-resolution. NTIRE 2023 benchmark and report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[17] David L Donoho and Iain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. 4

[18] Egor Ershov et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 8

[19] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 8

[20] 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*, 2017. 6

[21] Robert William Gainer Hunt. *The reproduction of colour*, volume 4. Wiley Online Library, 1995. 6

[3] https://toloka.ai/grants/

[22] Andrey Ignatov, Anastasia Sycheva, Radu Timofte, Yu Tseng, Yu-Syuan Xu, Po-Hsiang Yu, Cheng-Ming Chiang, Hsien-Kai Kuo, Min-Hung Chen, Chia-Ming Cheng, et al. MicroISP: processing 32mp photos on mobile devices with deep learning. In *Proceedings of the European Conference on Computer Vision (ECCV): Tel Aviv, Israel, October 23–27, 2022, Part II*. Springer, 2023. 8, 9

[23] Andrey Ignatov, Radu Timofte, Shuai Liu, Chaoyu Feng, Furui Bai, Xiaotao Wang, Lei Lei, Ziyao Yi, Yan Xiang, Zibin Liu, et al. Learned smartphone ISP on mobile gpus with deep learning, mobile AI & AIM 2022 challenge: report. In *Proceedings of the European Conference on Computer Vision (ECCV) 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Part III*. Springer, 2023. 8

[24] 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*. Springer, 2020. 5, 6

[25] Sing Bing Kang, Ashish Kapoor, and Dani Lischinski. Personalization of image enhancement. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010. 5

[26] Xiaoyang Kang, Xianhui Lin, Kai Zhang, Zheng Hui, Wangmeng Xiang, Jun-Yan He, Xiaoming Li, Peiran Ren, Xuansong Xie, Radu Timofte, et al. NTIRE 2023 video colorization challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[27] Furkan Kınlı, Doğa Yılmaz, Barış Özcan, and Furkan Kıraç. Modeling the lighting in scenes as style for auto white-balance correction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 6, 7

[28] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[29] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on image denoising: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[30] 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*, 2018. 6

[31] Shuai Liu, Chaoyu Feng, Xiaotao Wang, Hao Wang, Ran Zhu, Yongqiang Li, and Lei Lei. Deep-FlexISP: A three-stage framewArk for night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5, 8

[32] Xiaohong Liu, Xiongkuo Min, Wei Sun, Yulun Zhang, Kai Zhang, Radu Timofte, Guangtao Zhai, Yixuan Gao, Yuqin Cao, Tengchuan Kou, Yunlong Dong, Ziheng Jia, et al. NTIRE 2023 quality assessment of video enhancement challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 8

[34] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. *Computer Graphics Forum*, 37, 03 2018. 8

[35] Daniele Menon, Stefano Andriani, and Giancarlo Calvagno. Demosaicing with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing*, 16:132–141, 2007. 7

[36] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*. IEEE, 2007. 5

[37] Nathan Moroney. Local color correction using non-linear masking. In *Color and Imaging Conference*, volume 2000.1. Society for Imaging Science and Technology, 2000. 4

[38] OpenCV. OpenCV library. https://opencv.org/, 2021. Accessed: 2021-10-30. 6

[39] 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*, 2019. 5

[40] 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*. Springer, 2015. 5, 6

[41] 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. 5

[42] Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, et al. NTIRE 2023 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[43] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2023 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[44] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2023 challenge on light field image super-resolution: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[45] 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

[46] Ye Yuan, Wenhan Yang, Wenqi Ren, Jiaying Liu, Walter J Scheirer, and Zhangyang Wang. UG^{2+} track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments. *arXiv preprint arXiv:1904.04474*, 2019. 8

[47] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, et al. NTIRE 2023 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2023. 1

[48] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 6, 7

[49] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *IEEE International Conference on Computer Vision*, 2021. 8

[50] 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*, 2018. 1

[51] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, et al. NTIRE 2023 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 1

[52] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNET++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2019. 5

[53] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. MetaIQA: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 1

[54] Simone Zini, Claudio Rota, Marco Buzzelli, Simone Bianco, and Raimondo Schettini. Shallow camera pipeline for night photography rendering. *arXiv preprint arXiv:2204.08972*, 2022. 4

[55] Simone Zini, Claudio Rota, Marco Buzzelli, Simone Bianco, and Raimondo Schettini. Back to the future: a night photography rendering ISP without deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 4