Continuous Exposure for Extreme Low-Light Imaging

Evgeny Hershkovitch Neiterman*  Michael Klyuchka*  Gil Ben-Artzi
neiterman@ariel.ac.il  michael.klyucha@msmail.ariel.ac.il  http://www.gil-ba.com

Computer Science Department, Ariel University, Israel

Abstract

We consider the problem of enhancing an underexposed dark image captured in a very low-light environment where details cannot be detected. Existing methods learn to adjust the input image’s exposure to a predetermined value. In practice, however, the optimal enhanced exposure varies from one input image to another, and as a result, the enhanced images may contain visual artifacts such as low-contrast or dark areas. We address this limitation by introducing a deep learning model that allows the user to continuously adjust the enhanced exposure level during runtime in order to optimize the output based on his preferences. We present a dataset of 1500 raw images captured in both outdoor and indoor scenes in extreme low-light conditions, with five different exposure levels and various camera parameters, as a key contribution. We demonstrate that, when compared to previous methods, our method can significantly improve the enhancement quality of images captured in extreme low-light conditions under a variety of conditions.

1. Introduction

Images captured in low light are characterized by low photon counts, which results in a low signal-to-noise ratio (SNR). Setting the exposure level while capturing an image can be done by the user in manual mode, or automatically by the camera in auto exposure (AE) mode. In manual mode, the user can adjust the ISO, f-number, and exposure time. In auto exposure (AE) mode, the camera measures the incoming light based on through-the-lens (TTL) metering and adjusts the exposure values (EVs), which refers to configurations of the above parameters.

We consider the problem of enhancing a dark image captured in an extremely low-light environment, based on a single image [8]. In a dark environment, adjusting the parameters to increase the SNR has its own limitations. For example, high ISO increases the noise as well, and lengthening the exposure time might introduce blur. Various approaches have been proposed as post-processing enhancements in low-light image processing [7, 15, 16, 20, 34, 35]. In extreme low light conditions, such methods often fail to produce satisfactory results. Recent works [8, 17, 24, 27, 31] introduce data-driven approaches to replace the traditional image signal processing pipeline and learn a direct mapping from low-exposure input images to well-lit output images. Such models are trained to set the exposure level of the output image to a single fixed value, which can result in less visually appealing images when the optimal exposure is unknown, as often is the case.

Our post-processing technique overcomes this limitation by allowing the user to set the most appropriate exposure value of the enhanced image from a continuous range. Figure 1 shows such an example. The input is a dark image (top left). The result of [8] is shown in the bottom left, where the exposure time of the enhanced image is always set to a fixed value regardless of the input image. As can be seen (the enlarged rectangle), the output image includes artifacts. In our approach (bottom right), the user sets the exposure time to the best value, which makes the image more visually appealing.

*These authors contributed equally to this work
We show that our approach yields enhanced well-lit images with a significant reduction in artifacts. Previous approaches [13, 19] considered only signal-independent noise using synthetic noise samples. In order to present an accurate model that can enhance both signal-dependent and signal-independent noises, synthetic noise samples are not sufficient, and we need multiple short and long real-world exposure images of the same scene in a variety of environments. To the best of our knowledge, there is no such public dataset. As a main contribution and to support further research in the field, we have collected 1500 raw images captured with five different exposure levels in extreme low-light conditions, in both indoor and outdoor environments, and under various camera parameters. Using the dataset, we train our model, showing that it can successfully control the desired exposure level at runtime.

Unlike previous approaches, which aimed to map multiple degradation levels in the input into the same single output, our approach can conditionally map a single input exposure, e.g., 0.1s, to multiple output exposures (1s-10s). We show that the choice of the network’s architecture is critical, and employ a U-Net [25] to operate directly on the raw data rather than on the sRGB data [13, 19]. Our model includes two input parameters. The first one controls the increase in the brightness of the output image by simple multiplication. Such an increase results in an inevitable amplification of the noise and other artifacts as well. The second parameter continuously tunes the operation of the image signal processing (ISP) unit to enhance these degradations accordingly, conditioned on the first one. To effectively enhance the image, we operate on the raw data from the sensors. Figure 2 illustrates the interaction with our model.

This paper makes the following contributions:

- A model that is tailored for extreme low-light imaging and can continuously enhance during inference the exposure of images captured in extreme low-light conditions with both signal-dependent and signal-independent noise sources.
- A new dataset of 1,500 raw images captured in extreme low-light conditions, with five different exposure levels from indoor and outdoor scenes.
- Qualitative and quantitative evaluations indicate a significant improvement in the enhancement quality of images captured in extreme low-light conditions when the exact optimal exposure of the ground truth image is unknown.

2. Related Work

Low-light Image Enhancement. Widely used enhancement methods are histogram equalization, which globally balances the histogram of the image; and gamma correction, which increases the brightness of dark pixels. More advanced methods include illumination map estimation [11], semantic map enhancement [32], bilateral learning [10], multi-exposure [3, 6, 33], Retinex model [5, 9, 30, 36] and unpaired enhancement [18]. In contrast to these methods, we consider an extreme low-light environment with very low SNR, where the scene is barely visible to the human eye. Chen [8] has introduced an approach to extreme low-light imaging by replacing the traditional image processing pipeline with a deep learning model based on raw sensor data. Wang [27] introduced a neural network for enhancing underexposed photos by incorporating an illumination map into their model, while Xu [31] presented a model for low-light image enhancement based on frequency-based decomposition. These methods are optimized to output an enhanced image with a fixed exposure. In cases where the user requires a change in the exposure of the output image, these methods require retraining the models, typically on additional sets of images. In contrast, we introduce an approach that enables continuous setting of the desired exposure at inference time.

Continuous Restoration Networks. Recently, there has been a growing interest in constructing networks that can be continuously tuned at inference time. These can
Figure 3: The multi-exposure dataset. The top two rows are images from outdoor scenes, and the bottom two rows are images of indoor scenes. From left to right are exposure times of 0.1s, 0.5s, 1s, 5s, and 10s.

| Property         | Value                |
|------------------|----------------------|
| Filter Array     | Bayer                |
| Exposure Times   | 0.1s, 0.5s, 1s, 5s, 10s |
| Exposure Ratios  | x2, x5, x10, x50, x100 |
| Resolution       | 6000 × 4000          |

Table 2: The dataset includes images with multiple exposure times in dark environments. In each scene, we used a tripod, and the camera parameters for the image with the longest exposure time were optimized to obtain the best perceptual quality. Subsequent images of the same scene were captured with shorter exposure times without touching the camera using a smartphone application.

broadly be categorized as models that allow tuning different objectives at runtime [14] or different restoration levels of the same objective. A typical approach is to train a network on different restoration levels and apply interpolation between the resulting weights. Dynamic-Net [26] adds specialized blocks directly after convolution layers, which are optimized during the training for an additional objective. CFSNet [28] uses branches, each one targeted for a different objective. AdaFM [13] adds modulation filters after each convolution layer. Deep Network Interpolation (DNI) [29] trains the same network architecture on different objectives and interpolates all parameters. Son [21] extends the approach of AdaFM with a Filter Transition Network (FTN), allowing better non-linear interpolation. These methods are optimized for denoising, super-resolution, style transfer, and compression artifact reduction. We focus on the complementary task of continuous exposure for images captured in extreme low-light conditions.

Datasets Darmstadt Noise Dataset (DND) [23] contains pairs of real images with low and high ISO to address noise and light effects. The images were mostly captured under normal lighting conditions and cannot be used in our setting. The RENOIR dataset [4] aims to propose a benchmark for noisy images but includes spatial misalignment. The Smartphone Image Denoising Dataset (SIDD) [2] introduces a large collection of ground truth for real noisy images. Their dataset does not offer outdoor scenes with extremely low-light and corresponding ground truth and mostly includes a fixed combination of parameters suited for denoising. The Google HDR+ dataset [12] uses bursts of images to increase dynamic range and reduce noise but has not been captured in low-light settings. [30] introduced a low-light paired dataset for underexposure enhancement. The most relevant dataset was introduced by Chen [8].
which includes raw sensor data to address extreme low-light imaging. [31] prepared a real noisy low-light and ground-truth dataset based on [8] for sRGB images. Unlike existing datasets, we introduce a long-exposure reference image with multiple exposure times for each extreme low-light scene, in both indoor and outdoor scenes, and directly operate on the raw sensor data.

3. Our Approach

3.1. Our Model

Modern digital cameras are designed to provide a pleasing and accurate image of the world in the same way that humans perceive it. The raw sensor data of a camera contains noisy measurements, and the image signal processing unit (ISP) converts the noisy linear intensities into a clean final image. Here, we describe how each stage of the pipeline, from sensor measurement to final image, is modeled in order to restore realistic extreme low light raw data.

The noise in raw sensor data results from two main sources: signal-dependent noise and signal-independent. The first one is referred to as shot noise, which is related to the uncertainty that is a property of the underlying signal itself, describing the photon arrival statistics. The second one is read-noise, which is the result of uncertainty generated by the electronics of the camera when the charge stored is read out. The shot noise is a Poisson random variable, whose mean is the expected number of photons per unit time interval, describing the true light intensity. The read noise is a Gaussian random variable with zero mean whose variance is fixed.

The heteroscedastic Gaussian model is a more widely acknowledged alternative to the Poisson-Gaussian model, which substitutes the Poisson component with a Gaussian distribution whose variance is signal-dependent:

\[ y \sim N(x, \beta_{\text{shot}} + \beta_{\text{read}} x), \]  

where \( y \in \mathcal{Y} \) is the observed (raw) intensity at a pixel in the raw data space \( \mathcal{Y} \), \( x \) is the original (unknown) signal, \( \beta_{\text{shot}} \) is proportional to the analog gain \((g_a)\) and digital gain \((g_d)\) and \( \beta_{\text{read}} \) is proportional to the sensor readout variance \((\sigma_r^2)\) and digital gain:

\[ \beta_{\text{read}} = g_d^2 \sigma_r^2, \quad \beta_{\text{shot}} = g_d g_a. \]  

Each pixel corresponds to a single sensor. The color filter array (CFA) in our camera is arranged in a Bayer pattern, R-G-B-G, resulting in a raw image that contains the linear intensities. The Raw-to-sRGB pipeline converts the raw to sRGB format and includes demosaicing, white balance, digital gain, denoising, color correction, gamma compression, quantization, and tone mapping.

The goal of existing extreme low light approaches is to find a function that can map at inference time a data point from raw data space to a single data point in the sRGB space, denoted as \( f: \mathcal{Y} \rightarrow \mathcal{Y}_{\text{rgb}} \), where \( \mathcal{Y}_{\text{rgb}} \) is the sRGB space. This approach leads to inaccurate results and may result in noticeable artifacts. We take an alternative approach and use the modulation module in [13] to enable the mapping of a data point from raw data space to multiple points in sRGB according to the user preference, each with a different exposure level. Our Raw-to-sRGB pipeline is formulated as a function \( f: \mathcal{Y} \times \mathcal{R} \times \mathcal{R} \rightarrow \mathcal{Y}_{\text{rgb}} : \)

\[ y_{\text{rgb}} = f(y, \alpha_1, \alpha_2; \theta), \]  

where \( \alpha_1 \) is a scalar that sets the mean of the signal Eq. (1) to the desired level by multiplication of the raw data, \( \alpha_2 \) control the enhancement level of the Raw-to-sRGB pipeline, \( \theta \) represents the parameters of \( f(\cdot) \) and \( y_{\text{rgb}} \) is the signal of the sRGB image.

The function \( f \) is realized by a deep network with modulation layers. To obtain \( \theta \), we train our network in two steps. First, the base model is trained to fit the enhanced image with an initial exposure level, without any additional modifications to the existing architecture. Then we freeze the weights of the base model, and each modulation layer \((g)\) is inserted after each existing convolutional kernel:

\[ g(w, b) \circ X, \]  

where \( X \) as the output feature map of existing convolutional kernels in the base network and \( w, b \) are weights and bias of the modulation layer’s convolutional filter kernel (Fig. 5). The network is then fine-tuned to fit the enhanced image with a final exposure level by learning the weights of the additional convolutional kernels. Thus, in our formulation, \( \theta \) includes the parameters of both the base network and the modulation layers.

During runtime, the weights of filter \( w \) and the bias \( b \) in each modulation layer are set to:

\[ g(\alpha_2; w, b) = \alpha_2 w + (1 - \alpha_2) \ast I, b = \alpha_2 b, \]  

for the given scalar \( 0 \leq \alpha_2 \leq 1 \) representing the enhancement parameter. Note that as the base network weights are fixed, by setting the filters in the modulation modules to identity and the biases to zero, the network output is the same as the trained base network.

As stated in [2], operating in sRGB domain for noisy images limits the representation power of the architecture. To alleviate this problem, we operate in the raw domain and employ a U-Net [25] as our base architecture \((f)\). It replaces the entire image signal processing (ISP) pipeline [8]. The input is a short exposure raw image from Bayer sensor data and the output is an sRGB image. The raw Bayer sensor data is packed into four channels, the spatial resolution is reduced by a factor of two in each dimension; and the black
Figure 4: The architecture of our network. There are two input parameters, \( \alpha_1 \) (brightness) and \( \alpha_2 \) (enhancement). \( \alpha_1 \) controls the brightness of the raw input data. \( \alpha_2 \) modulates the weights of the filters and tunes the network, which operates as an Image Signal Processing (ISP) unit. We train the model for an initial and final exposure level, where for each value of \( \alpha_1 \) there is a single value of \( \alpha_2 \). At inference time, each parameter can be set independently of the other (see Fig. 10).

Figure 5: The dashed red rectangle is the modulation module. The enhancement parameter \( \alpha_2 \) represents a weighted sum between the feature map of the initial and final exposure levels. The blue dashed line is to emphasize that the operation of the modulation module is also affected by the \( \alpha_1 \) parameters which control the brightness of the image.

Figure 6: The input to the network is a Bayer sensor packed into four channels, each half the height and width of the original color filter array. The preprocessing includes dark level subtraction. The output is a 12-channel image processed to recover the original resolution of the input image. The overall architecture of our network is presented in Fig. 4.

3.2. Training and Inference

We train the model using L1 loss and the Adam optimizer. The inputs are random 512x512 patches which are cropped, rotated and flipped. The learning rate is \( 10^{-4} \) for 1000 epochs and then \( 10^{-5} \) for an additional 1000 epochs, a total of 2000 epochs for the training phase. Fine-tuning the model for the final exposure level requires an additional 1000 epochs. The model is implemented using PyTorch [22].

During training, the brightness and enhancement parameters are adjusted to the corresponding ground-truth image. The input arrays’ values are multiplied by the brightness parameter, which represents the ratio between the input image’s exposure time and the required output image’s exposure time, effectively determining the brightness of the output. In practice, the maximum exposure ratio is truncated by 100 and the output is clipped to \([0,1]\). The enhancement parameter is set between zero and one, where zero corresponds to the ground-truth image with the lower exposure level before adding the modulation blocks, and one corresponds to the ground-truth image with the higher exposure level and hence less noisy.

For testing, we set the brightness and the enhancement parameters of the network to the desired exposure as follows. The input image is multiplied according to the brightness parameter, resulting in a noisy, brighter image. The weights of the filter and bias in the modulation module after the fine-tuning phase are adjusted according to the enhancement parameter (Eq. (5)).

3.3. Multi-Exposure Extreme Low-Light Dataset

We collected a total of 1500 images. In order to capture a variety of realistic low-light conditions and cover a broad range of scenes with extreme low-light conditions, the images were captured in both indoor and outdoor scenes. The images were captured over different days in multiple locations. We captured five different exposures for each of the scenes - 0.1s, 0.5s, 1s, 5s, and 10s. The outdoor images were captured late at night under moonlight or street lighting. The indoor images were captured in closed rooms with indirect illumination. Generally, the lowest exposure image
in both indoor and outdoor scenes is completely dark and no details of the scene can be observed.

All the scenes in the dataset are static to accommodate the long exposure. For each scene, the settings of the camera such as ISO and f-number were adjusted to optimize the longest-exposure image. We used a tripod and a mirrorless camera to capture the exact same scene without any misalignment. At each scene, after the long exposure image was optimally captured, we used a smartphone application to decrease the exposure and capture the images without touching the camera or changing the camera’s parameters. After capturing the images, we manually verified that the images are aligned and the long-exposure reference images are of high perceptual quality.

The images were captured using a Sony α5100 with a Bayer sensor. The resolution of the images is 6000 × 4000. Fig. 3 shows samples from our dataset. Tables 1,2 summarize its properties.

4. Experiments

Baselines. We compare our results with recent popular methods representing the various approaches available in the literature. Using our dataset, we train them in accordance with their authors’ instructions. The SID [8] model enhances raw image data (like ours) and can adjust the brightness of the output image using an input parameter at runtime. For additional comparisons, we include RGB-based low-light enhancement methods Retinex [30], AdaFM [13] and CResMD [19]. Our approach is Raw-to-RGB, which has been proven to be essential for enhancing the quality of images captured in very low-light conditions [8]. The inputs of the compared models were modified to operate on raw images in order to ensure fair comparisons.

Evaluation Metrics. We use 70%, 10%, and 20% of the images for training, validation, and testing, respectively, with uniform sampling and equal representation for indoor and outdoor scenes in each set. The ground truth images are the corresponding long-exposure images processed by LibRaw [1] to sRGB format. Following common practice, for quality metrics we rely on SSIM and PSNR to capture perceptual notions.

4.1. Quantitative Comparisons

Table 3 presents the PSNR and SSIM metrics for various experiments designed to evaluate the different approaches. Each section (A-D) represents a different experiment. The left column shows the different methods and their training protocols. The input for both training and testing is a dark image with an exposure time of 0.1s. For each method, the ground truth exposure times that were used for training (1s/5s/10s) are shown with each model (using the ⇒).

In Table 3.A we train the SID model for every single input and output exposure independently. By testing the model on the same exposure as trained, we obtain the optimal achievable restoration accuracy as the model is specialized on a single exposure time. Testing on other exposures (e.g., training on 5s and testing on 10s) shows that the resulting enhanced image quality is significantly reduced, which is the key limitation of single-output methods. The goal of our approach is to overcome this and achieve high restoration quality of the continuous range of possible exposure times with a single model.

Table 3.B evaluates the ability to train single-output approaches ( [8, 30]) to generalize to multiple output exposures. We train the models based on all possible output exposures and evaluate their ability to enhance specific exposure times within the trainable range. It can be seen that using multiple ground-truth exposures with models that are designed to output only a single one reduces the restoration quality for all the possible outputs.

We compare our approach with previous continuous methods. Table 3.C presents the results for one of the most common use-cases: where the optimal exposure time is within the trainable range in runtime. We train the models to enhance input images with an exposure of 0.1s and a ground truth exposure range of [1s,10s]. At inference time, the models can enhance an image to a range of exposures and the specific one is selected. We evaluate the models with input images of 0.1s and optimal output exposure of 5s. It can be seen that our approach outperforms all other methods.

In real-world scenarios, the actual optimal exposure time of the enhanced image can be outside the trained range. We experiment with such cases, training the models for optimal exposure times of [1s,5s], and testing with input images such that the ground truth exposure time is 10s. The results are presented in Table 3.D. As before, our approach achieves the best restoration accuracy.

4.2. Qualitative Comparisons

Fig. 7 shows the effect of adjusting the exposure time for a value within the trained range, 5s. The model is trained using input images with an exposure time of 0.1s and ground truth images with exposure times of 1s and 10s. SID was trained on all possible output exposure times. The enhanced images after adjusting the brightness and enhancement parameters are shown. The left column shows the effect of multiplying the intensity of the input images by 50, which is the ratio between the ground truth exposure of the input (0.1s) and the ground truth output (5s). It can be seen that our model successfully removes the artifacts presented by the other approaches.

Figure 8 shows the enhancement results when the optimal exposure time falls outside the trained range, demon-
| Train/Test | 1s PSNR SSIM | 5s PSNR SSIM | 10s PSNR SSIM |
|------------|-------------|-------------|-------------|
| **A**      |             |             |             |
| SID [8] 0.1 ⇒ 1 | 38.17 0.95  | 30.7 0.87  | 27.7 0.84   |
| SID [8] 0.1 ⇒ 5  | 36.82 0.94  | **33.35 0.91** | 28 0.86    |
| SID [8] 0.1 ⇒ 10 | 34.88 0.9   | 30.52 0.88  | **30 0.88** |
| **B**      |             |             |             |
| SID [8] ⇒ 1,5,10 | 35.77 0.92  | 29.55 0.86  | 26.25 0.82  |
| Retinex [30] ⇒ 1,5,10 | 16.29 0.08  | 15.15 0.12  | 13.67 0.16  |
| **C**      |             |             |             |
| AdaFM [13] ⇒ 1,10 | 37.86 0.85  | 30.51 0.73  | 26.95 0.72  |
| CResMD [19] ⇒ 1,10 | 36.37 0.8   | 21.63 0.46  | 26.52 0.64  |
| Ours ⇒ 1,10   | **38.17 0.95** | **32.35 0.89** | **29.67 0.87** |
| **D**      |             |             |             |
| AdaFM [13] ⇒ 1,5  | 37.86 0.85  | 31.12 0.76  | 25.98 0.7   |
| CResMD [19] ⇒ 1,5  | 34.97 0.73  | 23.73 0.59  | 16.17 0.17  |
| Ours ⇒ 1,5    | **38.17 0.95** | **31.78 0.89** | **28.65 0.86** |

Table 3: Comparisons with previous works. For all methods, the input exposure for both training and testing is 0.1s. ⇒ denotes the ground-truth images used for training. The bold are the two best results. As can be seen, our model outperforms all other methods. See text for more details.

| Filter Size | PSNR | SSIM |
|-------------|------|------|
| 1 × 1       | 31.87|      |
| 3 × 3       | 32.35|      |
| 5 × 5       | 32.39|      |
| 7 × 7       | 32.48|      |

Table 4: Filter size comparisons. The model is trained from 0.1s to 1s and fine-tuned to 10s, and tested for an unseen exposure level of 5s.

strating its robustness. We train our model for exposure times of [1s, 5s] the SID model for 5s. The left column shows the results of SID, the middle column our results, and the right one the ground truth. As can be seen, using a single-output approach results in significant distortions, whereas in our model, the parameters can be adjusted beyond the trainable range to output better restoration results.

In everyday situations, the exposure of the input image can be different from what is expected. Figure 9 shows the results of such a case. Our model and SID are trained as above to map images from 0.1s to 5s in the case of SID, and [1s, 5s] in the case of our model. However, the actual illumination conditions of the input images differ from those of the training set and the average intensity of the images is approximately 10%. Due to the fact that our model can be continuously adjusted to different brightness and enhancement, we can find the optimal parameters to output clear images.

Fig. 10 presents the effect of the brightness and enhancement parameters, where each is adjusted independently of the other. The model is trained using input images with exposure times of 0.1s (not shown) and ground-truth images with exposure times of 1s (top left) and 10s (top right). During training, the values of the parameters are (10, 0) for 1s and (100, 1) for 10s. Other values were not realized during training, and their effects on output during runtime are shown.

### 4.3. Ablation Study

**Filter Size.** We evaluate the sizes of different filters in the modulation module. We consider filter sizes of – 1×1, 3×3, 5×5, and 7×7. We train our base model with an exposure of 0.1s and an output of 1s, then fine-tune it to an output of 10s. The test images are with an exposure of 5s. Table 4 shows our comparisons. It can be seen that the most significant gain is achieved when using a filter size of 3 × 3.

**Tuning Direction.** We evaluate the optimal direction for the tuning. We compare two models. The first one is trained from 0.1s to 1s and fine-tuned for 10s. The second one is trained from 0.1s to 10s and fine-tuned for 1s. We compare the results with respect to unseen output images with an exposure time of 5s. The forward direction from 0.1s to 10s achieved better results than the backward one, with a PSNR of 32.35 vs. 28.2.
5. Conclusion

Extreme low-light imaging is challenging and has recently gained growing interest. Current methods allow enhancement of dark images, assuming the input exposure and the optimal output exposure are known at inference time, which prevents its adaptation in practical scenarios. We present an approach that enables continuously controlling of the optimal output exposure levels of the images at runtime, without the need to retrain the model. We collected a dataset of 1500 images with multiple exposure levels for extreme low-light imaging and showed that our model presents promising results on a wide range of both indoor and outdoor images. We believe that our dataset as well as our model will support further research in the field of extreme low-light imaging, making a step forward towards its widespread adoption.

References

[1] LibRaw Project Goals and Objectives, 2018.

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[3] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9157–9167, 2021.

[4] Josue Anaya and Adrian Barbu. Renoir—a dataset for real low-light image noise reduction. Journal of Visual Communication and Image Representation, 51:144–154, 2018.

[5] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In Proceedings of the IEEE Interna-
Figure 8: The restoration effect of enhancing images to exposure levels outside the trained range. Our model provides more visually appealing results.

[6] 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.

[7] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011.

[8] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[9] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinlong Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016.

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

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

[12] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiwen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016.

[13] Jingwen He, Chao Dong, and Yu Qiao. Modulating image restoration with continual levels via adaptive feature modification layers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11056–11064, 2019.

[14] Jingwen He, Chao Dong, and Yu Qiao. Multi-dimension modulation for image restoration with dynamic controllable residual learning. *arXiv preprint arXiv:1912.05293*, 2019.

[15] Zhe Hu, Sunghyun Cho, Jue Wang, and Ming-Hsuan Yang. Deblurring low-light images with light streaks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3382–3389, 2014.
Figure 9: The restoration effect of enhancing images with different ambient illumination. In our model, we can adjust the parameters to compensate for the different mismatches that arise in real-world scenarios.

[16] Sung Ju Hwang, Ashish Kapoor, and Sing Bing Kang. Context-based automatic local image enhancement. In *European conference on computer vision*, pages 569–582. Springer, 2012.

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

[18] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision, 2021.

[19] He Jingwen, Dong Chao, and Qiao Yu. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. In *European Conference on Computer Vision*, pages 53–68. Springer, 2020.

[20] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013.

[21] Hyeongmin Lee, Taeoh Kim, Hanbin Son, Sangwook Baek, Minsu Cheon, and Sangyoun Lee. Smoother network tuning and interpolation for continuous-level image processing. *arXiv preprint arXiv:2010.02270*, 2020.

[22] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[23] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017.

[24] Tal Remez, Or Litany, Raja Giryes, and Alex M Bronstein. Deep convolutional denoising of low-light images. *arXiv preprint arXiv:1701.01687*, 2017.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical*
Figure 10: The effect of the enhancement and brightness parameters. The input is a dark image, top left, and top right are the ground truth images for training. All other images are acquired during runtime by continuously setting the parameters of the model. See text for further details.

[26] Alon Shoshan, Roey Mechrez, and Lihi Zelnik-Manor. Dynamic-net: Tuning the objective without re-training for synthesis tasks. In Proceedings of the IEEE International Conference on Computer Vision, pages 3215–3223, 2019. 3

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

[28] Wei Wang, Ruiming Guo, Yapeng Tian, and Wenming Yang. Cfsnet: Toward a controllable feature space for image restoration. In Proceedings of the IEEE International Conference on Computer Vision, pages 4140–4149, 2019. 3

[29] Xintao Wang, Ke Yu, Chao Dong, Xiaou Tang, and Chen Change Loy. Deep network interpolation for continuous imagery effect transition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1692–1701, 2019. 3

[30] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560, 2018. 2, 3, 6, 7

[31] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2281–2290, 2020. 1, 2, 4

[32] Zhiheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. Automatic photo adjustment using deep neural networks. ACM Transactions on Graphics (TOG), 35(2):1–15, 2016. 2

[33] 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. 2

[34] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In European Conference on Computer Vision, pages 771–785. Springer, 2012. 1

[35] Xiangdong Zhang, Peiyi Shen, Lingli Luo, Liang Zhang, and Juan Song. Enhancement and noise reduction of very low light level images. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pages 2034–2037. Ieee, 2012. 1

[36] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, pages 1632–1640, 2019. 2