Computer Vision and Image Understanding

journal homepage: www.elsevier.com

BacklitNet: A dataset and network for backlit image enhancement

Xiaoqian Lv[a], Shengping Zhang[a,**], Qinglin Liu[a], Haozhe Xie[b], Bineng Zhong[c], Huiyu Zhou[d]

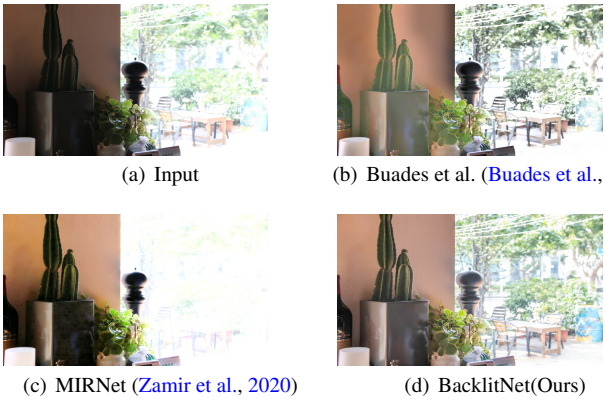
[a]*Harbin Institute of Technology, Weihai 264209, China.*
[b]*Harbin Institute of Technology, Harbin 150001, China.*
[c]*Guangxi Normal University, Guilin 541004, China.*
[d]*University of Leicester, Leicester LE1 7RH, United Kingdom.*

ABSTRACT

Backlit images are usually taken when the light source is opposite to the camera. The uneven exposure (e.g., underexposure on the foreground and overexposure on the background) makes the backlit images more challenging than general image enhancement tasks that only need to increase or decrease the exposure on the whole images. Compared to traditional approaches, Convolutional Neural Networks perform well in enhancing images due to the abilities of exploiting contextual features. However, the lack of large benchmark datasets and specially designed models impedes the development of backlit image enhancement. In this paper, we build the first large-scale BAcklit Image Dataset (BAID), which contains 3000 backlit images and the corresponding ground truth manually adjusted by trained photographers. It covers a broad range of categories under different backlit conditions in both indoor and outdoor scenes. Furthermore, we propose a saliency guided backlit image enhancement network, namely BacklitNet, for robust and natural restoration of backlit images. In particular, our model innovatively combines a nested U-structure with bilateral grids, which enables fully extracting multi-scale saliency information and rapidly enhancing arbitrary resolution images. Moreover, a carefully designed loss function based on prior knowledge of brightness distribution of backlit images is proposed to enforce the network to focus more on backlit regions during the training phase. We evaluate the proposed method on the BAID dataset and two public small-scale backlit image datasets. Experimental results demonstrate that our method performs favorably against the state-of-the-art approaches.

1. Introduction

Backlit images are usually taken when the light source is opposite to the camera. These images are visually unpleasant because of extremely varied exposure between bright backgrounds and dark foregrounds, poor contrast and barely-visible details as shown in Figure 1(a). Furthermore, they also degenerate the performance of the high-level computer vision task, including image classification (Ciocca et al., 2018; He et al., 2016; Simonyan and Zisserman, 2015), object detection (Car-



(a) Input (b) Buades et al. (Buades et al., 2020) (c) MIRNet (Zamir et al., 2020) (d) BacklitNet(Ours)

Fig. 1. (a) A challenging backlit image. (b) enhanced by a backlit image enhancement method (Buades et al., 2020). (c) enhanced by a general image enhancement method (Zamir et al., 2020). (d) enhanced by our method.

**Corresponding author.
e-mail: s.zhang@hit.edu.cn (Shengping Zhang)

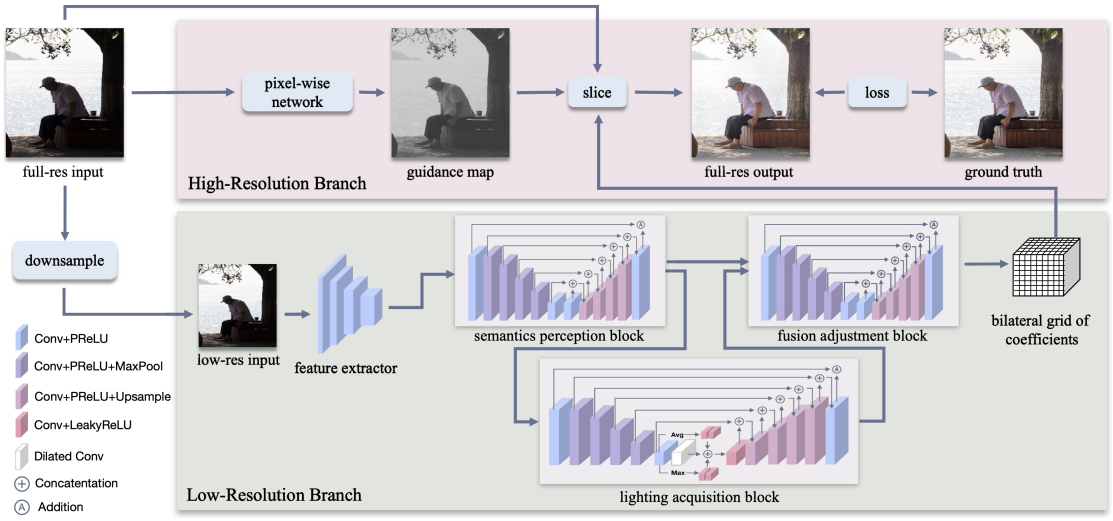


Fig. 2. The architecture of our BacklitNet network. First, we downsample the full-resolution input to low-resolution input and extract its low-level features. Next, the feature maps pass through a two-level nested U-structure to obtain global brightness and local backlit information, which is stored in a bilateral grid. Finally, we apply the bilateral grid to the original image and use a guidance map to guide the process of image upsampling to obtain the final result.

ion et al., 2020; Vaca-Castano et al., 2017; Zhang et al., 2017), and object segmentation (He et al., 2017; Le et al., 2019; Xie et al., 2020). Therefore, there is a great demand for an effective backlit image enhancement method.

Because of the extremely uneven exposure, enhancing backlit images is a more challenging task than other general image enhancement tasks. Although most modern imaging sensors allow to automatically adjust hardware parameters according to lighting conditions, they still cannot eliminate the adverse influence of backlit conditions. As a post-processing technology, High Dynamic Range (HDR) (Debevec and Malik, 2008) can correct wrong exposure. However, it requires multiple images of the same scene and is prone to produce artifacts (Yan et al., 2019). Besides, several image editing softwares (*e.g.* Photoshop, Lightroom and GIMP) help photographers obtain visually appealing images. Nevertheless, they usually require complex operations and professional skills, which are difficult to use for common users.

In computer vision, image enhancement has been attracting increasing interest (Chai et al., 2020; Gharbi et al., 2017; Sun et al., 2021; Wang et al., 2019a; Zhang et al., 2019b; Zeng et al., 2020; Zamir et al., 2020). However, most existing methods (Afifi et al., 2020; Chen et al., 2018; Wang et al., 2019a; Zhang et al., 2019b) focus on enhancing either under-exposed or over-exposed images, which yields unpromising results on backlit images as shown in Figure 1. Existing methods for backlit images enhancement can be roughly divided into three categories: segmentation-based methods (Li et al., 2015; Li and Wu, 2018; Trongtirakul et al., 2020; Vazquez-Corral et al., 2018), fusion-based methods (Buades et al., 2020; Ueda et al., 2020; Wang et al., 2016) and learning-based methods (Zhang et al., 2019a). Segmentation-based methods attempt to segment an image into backlit and frontlit regions and use different tone mapping functions to enhance different regions separately. Nevertheless, the enhanced results are unstable in complex scenes because they heavily rely on the performance of segmentation. Fusion-based methods usually process a backlit image in different function spaces and then fuse them. However, the manually designed constraints and parameters limits these methods to be widely used in variable scenes. Due to the absence of large-scale backlit image dataset, there are few learning-based studies. The only learning-based method, ExCNet, cannot get satisfactory results and speed in extremely backlit cases. Existing methods do not make full use of the feature representation abilities of CNNs. Therefore, there is still a lot of room for improvement in backlit image enhancement.

(a) Indoor (b) Outdoor

Fig. 3. Several examples of the proposed BAID dataset. Top: backlit images. Bottom: normal light images (ground truth).

In this paper, we propose a novel end-to-end learning-based method to process backlit images in various scenes. Firstly, we construct the first large-scale BAcklit Images Dataset (BAID) which contains 3000 backlit images and the corresponding ground truth. It covers a broad range of scenes, subjects, and lighting conditions. Furthermore, we propose a saliency guided backlit image enhancement network, which we refer to as BacklitNet. Specifically, we use a two-level nested U-structure to extract saliency information from different receptive fields. Such a multi-scale nested structure prompts our model to fully consider the brightness of entire images and specific objects in backlit areas. Then, we introduce a bilateral grid to estimate the transformation from input to output. Unlike the direct processing of the output from U-structure, transformation based on the bilateral grid has the advantages of resolution-independent, rapidity and no artifact. Finally, through the histogram statistics of backlit images and normal images, we obtain prior knowledge of the brightness distribution and design a loss function based on it.

The main contributions can be summarized as follows:

- We build a large-scale backlit image dataset (BAID) which contains 3000 backlit images and the corresponding high-quality ground truth. To the best of our knowledge, it is the first large-scale dataset for backlit image enhancement. The constructed dataset makes end-to-end learning of robust backlit enhancers possible and promotes the application of neural networks in backlit image enhancement.
- We design a novel backlit image enhancement framework, which efficiently restores ill-exposed regions of backlit images based on subject salience and brightness distribution prior knowledge without damaging the well-exposed regions.
- We evaluate the proposed method on the BAID dataset and two public small-scale backlit image datasets. The results show that our model achieves state-of-the-art performance both on visual effects and commonly-used metrics.

2. Related Work

Segmentation-based Method. Segmentation-based methods often segment an image into backlit and frontlit regions and process them separately. Li et al. (2015) design a two-component Gaussian mixture model to obtain underexposed and overexposed regions and perform different tone-mappings on them, respectively. Li and Wu (2018) further introduce an object-guided segmentation method followed by spatially adaptive tone mapping. Vazquez-Corral et al. (2018) propose a variational region split model to obtain a set of weight maps, which can divide the image into darker and lighter areas. Then, they compute as many tone-cures as weight maps and apply them to the original image. Trongtirakul et al. (2020) present an unsupervised single backlit image enhancement method, which stretches luminance in piece-wise regions and uses a logarithmic weighted luminance function to enhance the stretched images. However, these methods are time-consuming and not suitable for high-resolution images. In addition, the accuracy of segmentation is vulnerable to the complexity of the scene, which can lead to unstable enhancement results.

Fusion-based Method. Fusion-based methods address this problem by processing backlit images in different function spaces and fusing them using a specific fusion algorithm. Wang et al. (2016) introduce a multi-scale fusion-based method,

which can simultaneously improve luminance and contrast. At the same time, the method selects exposedness as the weight map to measure and extract more details of the input image. Buades et al. (2020) adopt gamma and logarithmic tone mapping functions to improve the contrast of backlit images in each color channel of the RGB space. Then, they fuse images based on a modified Mertens algorithm (Mertens et al., 2007) and refine the fused images by sharpening details and correcting chrominance. Ueda et al. (2020) use the triangular-shaped unimodal histogram to improve the bimodal distribution of the intensity histogram of a backlit image. However, these methods are not robust to changeable scenes due to many fixed parameters. Meanwhile, the enhanced result is prone to objectionable color distortion and unnatural illumination when images contain quite dark regions.

Learning-based Method. Unlike the great success of deep learning in other vision tasks, few researches have been done on learning-based backlit image enhancement methods due to the absence of a large-scale backlit image dataset. At present, to our knowledge, there is only one learning-based method for backlit image enhancement, namely ExCNet (Zhang et al., 2019a). It is a zero-shot scheme that utilizes a block-based loss function to guide the restoration progress to estimate "S-curve" for image enhancement. But this method requires iterative training for each test image to obtain an image-specific "S-curve", which makes the model non-real time. Recently, learning-based methods in the field of general image enhancement have developed rapidly. Chai et al. (2020) estimate the coefficient of parameterized color mapping through CNN and apply it to original images of any resolution. Zeng et al. (2020) learn multiple image-adaptive 3-dimensional tables in couple with a lightweight CNN weight predictor to manipulate the color and tone of images in real-time. Zamir et al. (2020) combine contextual information and spatial details by parallel multi-resolution convolution streams and attention based on multi-scale feature extraction and aggregation. Although these methods demonstrate their effectiveness in image enhancement benchmark datasets (Bychkovsky et al., 2011; Hasinoff et al., 2016; Wei et al., 2018), they do not consider the specific characteristics of the backlit images. Our pipeline is complementary to existing methods in two ways: First, we collect the first large-scale backlit image dataset, which further facilitates the relevant learning-based research. On the other hand, we develop an effective and efficient model that can restore underexposed backlit regions while preserving the harmony of overall brightness without image degradation.

3. The BAcklit Images Dataset

Large-scale datasets are indispensable for training a robust backlit image enhancement model. However, due to the diversity of lighting conditions, the complexity of scenes and the large cost of manual editing, there is no large-scale backlit image benchmark dataset currently. Existing backlit image enhancement researches usually use Li and Wu dataset (Li and Wu, 2018) and Vonikakis et al. dataset (Vonikakis et al., 2018) to verify the performance of their methods. However, the numbers of images in the two datasets are only 23 and 38 respectively, which is far from sufficient to train neural networks. Besides, the lack of ground truth makes these two datasets impossible for end-to-end model training. In order to train an end-to-end network and promote the application of CNNs in backlit image enhancement, we contribute a large-scale publicly available backlit image enhancement dataset with high-resolution, named BAcklit Images Dataset (BAID), containing 3000 backlit images and the corresponding ground truth.

To ensure obtaining a general and robust backlit image enhancer, the dataset should cover a broad range of scenes, subjects, weather and lighting conditions with commonly used capture devices. Therefore, our dataset, BAID, contains various representative real-world scenes (e.g., libraries, parks and streets) and diverse categories (e.g., people, buildings and plants). Table 1 reports the number of different categories and Figure 3 shows several samples of the dataset. Meanwhile, aiming to simulate the situations we encounter in daily life as much as possible, we set AUTO mode in camera to capture images in the resolution of 5472×3648 and 3648×5472 using Canon

Table 1. The number of images in different categories of the proposed BAID dataset.

Category	Outdoor					Indoor			
	people	vehicle	building	plant	sculpture	people	plant	furniture	ornament
#Images	975	186	325	226	177	390	195	262	264

Table 2. The number of images taken by different cameras of the proposed BAID dataset.

Camera Brand	Canon EOS 5D Mark III	Canon EOS 6D	Canon EOS 70D	Nikon D750
#Images	704	732	754	810

EOS 5D Mark III, Canon EOS 6D, Canon EOS 70D and Nikon D750. As shown in Table 2, the number of images taken by different cameras is roughly balanced. It is worth noting that we provide different images formats containing RAW, PNG and JPEG formats.

Moreover, high-quality ground truth is crucial for training a supervised neural network. Therefore, we recruit five professionally trained photographers to retouch the captured backlit images using Adobe Lightroom. Then, we invite 20 volunteers to rate the images among the five results of each image to obtain the best one as the ground truth. By this means, we can get relatively objective reference images as shown in the bottom row of Figure 3. Finally, we randomly split the dataset into two subsets: 2600 images for training and 400 for testing.

4. The Proposed Method

Backlit images are different from low-light images in that they have a higher dynamic range and contain over-exposed, well-exposed and under-exposed regions simultaneously. Thus, when processing backlit images, we need to pay more attention to the local backlit region while ensuring the naturalness of the whole images. Existing methods typically separate the foreground and background and process them separately. However, there are several disadvantages: low efficiency of multi-stage, inaccurate segmentation in complex scenes and unnatural edges of the fused result. To address these problems, we propose an end-to-end backlit image enhancement network BacklitNet.

4.1. Network Architecture

BacklitNet consists of a dual-resolution framework and a powerful feature extractor. Given an input image I, the enhanced result $\hat{I}$ is formulated as

$$\hat{I} = U(A(X), g, I) \tag{1}$$

where $X = f(I)$ is the features of the input, which is obtained from our feature extractor. $A(\cdot)$ is the bilateral grid storing the transform coefficients of X. g is the guidance map transformed from I. $U(\cdot)$ is a bilateral grid upsampling function. The detailed network is provided below.

Dual-resolution Framework. Figure 2 illustrates the framework of BacklitNet. In order to process high-resolution images quickly, the model is divided into two branches. One is a high-resolution branch (the pink area in Figure 2), the other is a low-resolution branch (the green area in Figure 2). In the low-resolution branch, the downsampled image passes through a low-level feature extractor and a two-level nested U-structure to learn the transform coefficients from input to output and store it in a 3D bilateral grid. Calculation at low-resolution makes this process efficient. In the high-resolution branch, we perform the nonlinear transformation on the original input to obtain a full-resolution edge-aware guidance map which guides the upsampling of bilateral grid. Then, utilizing the principle of Bilateral Guided Upsampling (BGU) (Chen et al., 2016) and the slice operation mentioned in Gharbi et al. (2017), we upsample the 3D bilateral grid to obtain 2D full-resolution coeffi-

cient maps. Specifically, under the guidance of a full-resolution guidance map g, the transform coefficients in the bilateral grid A are sampled by tri-linearly interpolation in the spatial domain and intensity domain, which can be written as

$$\bar{A}_m[x,y] = \sum_{i,j,k} \tau(s_x x - i)\,\tau\left(s_y y - j\right)\tau(d \cdot g[x,y] - k) A_m[i,j,k] \tag{2}$$

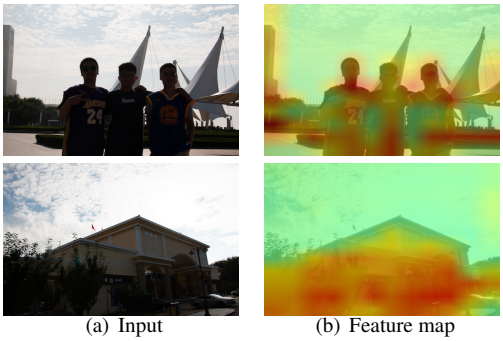
where $\tau(\cdot) = max(1 - |\cdot|, 0)$ is a linear interpolation kernel, s_x and s_y are the width and height ratios of the grid's dimensions to the input's dimensions. The depth d is set to 8. By doing this, we obtain a set of full-resolution coefficient maps $\bar{A}$ containing pixel-wise transformation information. Then, we apply $\bar{A}$ to input I ($n_I = 3$) to obtain each channel of output $\hat{I}_c$.

$$\hat{\mathrm{I}}_c[x,y] = \bar{A}_{n_I+(n_I+1)c}[x,y] + \sum_{c'=0}^{n_I-1} \bar{A}_{c'+(n_I+1)c}[x,y] I_{c'}[x,y] \tag{3}$$

We upsample the transformation coefficients instead of image pixels, which reduce the loss of detail and prevent artifacts. Through the framework, BacklitNet can obtain high quality full-resolution results with low computational cost.

Feature Extractor. How to extract features of the backlit area is crucial in the task of backlit image enhancement. However, the existing methods, like ExCNet (Zhang et al., 2019a), do not design a particular model to extract the features of the backlit area. Meanwhile, preserving the harmony and naturalness of the overall brightness of images is equally essential. Under these circumstances, we treat the backlit area as a salient area and incorporate the idea of salient object detection (SOD) (Liu et al., 2019a,b; Qin et al., 2019) into backlit image enhancement. SOD-based design empowers the network to focus more on the backlit region to extract semantic and lighting features.

Specifically, we use a two-level nested U-structure to learn enriched semantic and spatial information. Overall, its top level is a big U-structure consisting of three blocks: semantics perception block (SP), lighting acquisition block (LA) and fusion adjustment block (FA). Its bottom level is a residual U-structure whose detailed configurations are presented in the supplementary material. Many previous researches (Huang et al., 2020; Kohl et al., 2018; Ronneberger et al., 2015; Wang et al., 2019b) have demonstrated the effectiveness of U-structure to extract



(a) Input (b) Feature map

Fig. 4. Visualization of input images and their feature maps obtained from the lighting acquisition block of BacklitNet.

global and local features. Concretely, SP aims to perceive the semantic information of foreground and background and LA is to obtain global lighting conditions information. Then, FA fuses semantic and lighting information and learns the intensity value of different areas. The structure of SP, LA, and FA are similar, including an input convolution layer, a symmetric encoder-decoder structure and a residual connection. The difference is that, in the last layer of the encoder in LA, we add max pooling and average pooling to further enlarge receptive field to capture global lighting features. Utilizing the nested U-structure, our model enables to extract and aggregate multi-scale and multi-level contextual and exposure features in different receptive fields.

Figure 4 gives the visualization of input backlit images and corresponding feature maps obtained from the LA block. As Figure 4(b) shows, backlit foreground accompany with larger activation value, which means the exposure of these areas will be improved significantly. On the contrary, the well-exposed background is given a relatively small activation value, changes in exposure will decrease accordingly. The results prove that the proposed feature extractor can fully exploit lighting information and effectively extract global saliency information.

4.2. Loss Function

We introduce a simple but effective backlit image brightness distribution prior knowledge. The prior is a kind of statistics of the backlit and normal light images. Specifically, counting the intensity histograms (as shown in Figure 5) of a large

number of backlit image pairs, we obtain the following prior knowledge: Compared with the intensity distribution of normal images, backlit images have a denser distribution of pixels in the low-intensity part, namely, the underexposed foreground region. Considering the change in distribution, we integrate the prior into the loss function to guide the network to restore the tone and details information of the darker foreground regions. The constraint $L_{backlit}$ is formulated as

$$L_{backlit} = \frac{1}{N}\sum_{i=0}^{N}(\hat{I}_i - G_i)^2 \cdot \exp(\lambda \cdot |I_i - G_i|) \tag{4}$$

where I and $\hat{I}$ are the input and output images of the proposed network, respectively. G is the ground truth image retouched manually. i is the index of each pixel in the image, N is the number of pixels in the image. The term $exp(\lambda \cdot |I_i - G_i|)$ can strengthen the supervision of under-exposed foreground regions by using the brightness distribution prior of the corresponding image pairs, which can be controlled by λ. In such a way, the model can pay more attention to the backlit region in the training process. In addition, we introduce the perceptual loss to improve the authenticity and naturalness of enhanced results

$$L_{perceptual} = \|\varphi(G) - \varphi(\hat{I})\|_2^2 \tag{5}$$

where φ is the feature map extracted from a VGG-16 model pre-trained on ImageNet. Therefore, the loss function L of BacklitNet is denote as

$$L = L_{backlit} + L_{perceptual} \tag{6}$$

5. Experiment

5.1. Datasets

We evaluate our model on the proposed BAID dataset and two public small-scale backlit image datasets: Li and Wu dataset (Li and Wu, 2018) and Vonikakis et al. dataset (Vonikakis et al., 2018), which contains 38 and 23 images, respectively. Additionally, in order to verify the stability and generalization of our model, we also conduct experiments on the low-light benchmark dataset LOL (Wei et al., 2018), including 500 real low-light image pairs and 1000 synthetic images pairs.

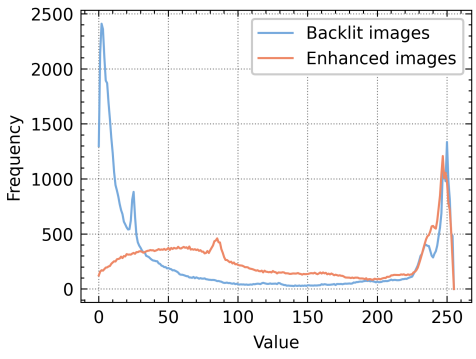


Fig. 5. Intensity histogram of backlit images and normal light images.

5.2. Evaluation Metrics

To comprehensively evaluate our model, four commonly-used metrics are used to measure the enhanced results including PSNR, SSIM (Wang et al., 2004), ΔE^* (Backhaus et al., 2011) and NIQE (Mittal et al., 2013). ΔE^* is a color difference metric defined in the CIELAB color space, which reflects human perception. NIQE is a blind image quality assessment metric using measurable deviations from statistical regularities observed in an image. Contrary to PSNR and SSIM, a smaller ΔE^* and NIQE means better performance.

5.3. Implementation Details

We build our model on PyTorch and train it for 200 epochs with a mini-batch size of 32 on an NVIDIA 2080Ti GPU. The input images with arbitrary resolution are downsampled to 384×384 in the low-resolution branch of BacklitNet. To prevent over-fitting, we use random flipping and rotation for data augmentation. We set λ =2 during the period of training, which is determined by a large number of experiments. The entire network is optimized by Adam optimizer. The initial learning rate is set to 0.001 for the first 100 epochs and decreases by half every 10 epochs.

(a) Input (b) Wang et al. (Wang et al., 2016) (c) Li and Wu (Li and Wu, 2018) (d) ExCNet (Zhang et al., 2019a) (e) Buades et al. (Buades et al., 2020)

(f) RetinexNet (Wei et al., 2018) (g) KinD (Zhang et al., 2019b) (h) DeepUPE (Wang et al., 2019a) (i) HDRNet (Gharbi et al., 2017) (j) DPED (Ignatov et al., 2017)

(k) Zeng et al. Zeng et al. (2020) (l) MIRNet (Zamir et al., 2020) (m) Chai et al. (Chai et al., 2020) (n) Ours (o) GT

Fig. 6. Qualitative comparison with state-of-the-art image enhancement methods on the proposed BAID dataset.

(a) Input (b) Wang et al. (Wang et al., 2016) (c) Li and Wu (Li and Wu, 2018) (d) ExCNet (Zhang et al., 2019a) (e) Buades et al. (Buades et al., 2020)

(f) RetinexNet (Wei et al., 2018) (g) KinD (Zhang et al., 2019b) (h) DeepUPE (Wang et al., 2019a) (i) HDRNet (Gharbi et al., 2017) (j) DPED (Ignatov et al., 2017)

(k) Zeng et al. (Zeng et al., 2020) (l) MIRNet (Zamir et al., 2020) (m) Chai et al. (Chai et al., 2020) (n) Ours

Fig. 7. Qualitative comparison with state-of-the-art image enhancement methods on the Li and Wu dataset whose ground truth is not provided.

5.4. Comparison with State-of-the-art

Quantitative Comparison. We compare our method quantitatively with four backlit image enhancement methods (Buades et al., 2020; Li and Wu, 2018; Wang et al., 2016; Zhang et al., 2019a), three excellent low-light image enhancement methods (Wei et al., 2018; Wang et al., 2019a; Zhang et al., 2019b) and five state-of-the-art general image enhancement methods (Chai et al., 2020; Gharbi et al., 2017; Ignatov et al., 2017; Zeng et al., 2020; Zamir et al., 2020). We retrained these aforementioned methods on the same dataset using the public source codes provided by the authors. For a fair comparison, we tried different hyperparameters and found that the recommended hyperparameters perform best. Therefore, we use the recommended hyperparameters. In the training phase, we visualized the loss curve of each method to ensure convergence. Table 3 reports the quantitative results on the BAID dataset. It is evident that our method achieves the best results far surpassing existing backlit image enhancement methods in terms of PSNR, SSIM, ΔE^* and NIQE. Because there is no ground truth in Li and Wu dataset and Vonikakis et al. dataset, we use the no-reference metric NIQE to perform quantitative comparisons. As shown in Table 4, the proposed method performs favorably against the state-of-the-art methods. BacklitNet can not only enhance backlit images, but also achieve satisfactory results on low-light image enhancement task. Table 5 reports the quantitative results of our method and state-of-the-art methods on the LOL dataset. As we can see, our method obtains the best SSIM and comparable PSNR that is slightly lower than MIRNet (Zamir et al., 2020). The results further confirm the effectiveness of our method.

Qualitative Comparison. To give an intuitive understanding of the promising performance of BacklitNet, we illustrate the sample results of BacklitNet and several state-of-the-art methods. Except for the BAID dataset, we also conduct experiments on two widely-used testing datasets without corresponding ground truth: Li and Wu dataset (Li and Wu, 2018) and Vonikakis et al. dataset (Vonikakis et al., 2018). More qualitative results can be seen in the supplementary materials. Figures 6 and 7 show the results of the compared methods on different test images of

Table 3. Quantitative comparison on the proposed BAID dataset with state-of-the-art methods, including (i) backlit image enhancement methods, (ii) low-light image enhancement methods and (iii) general image enhancement methods. The best and the second-best scores are shown in bold and underlined, respectively.

Method	PSNR	SSIM	ΔE^*	NIQE
Wang et al. (Wang et al., 2016)	17.96	0.86	13.35	3.60
Li and Wu (Li and Wu, 2018)	17.16	0.82	13.10	4.86
ExCNet (Zhang et al., 2019a)	19.31	0.90	11.41	2.94
Buades er al. (Buades et al., 2020)	17.47	0.89	13.46	3.67
RetinexNet (Wei et al., 2018)	21.26	0.90	9.67	3.48
KinD (Zhang et al., 2019b)	22.69	0.91	7.76	3.06
DeepUPE (Wang et al., 2019a)	21.05	0.90	9.73	2.97
HDRNet (Gharbi et al., 2017)	23.78	<u>0.95</u>	7.54	<u>2.91</u>
DPED (Ignatov et al., 2017)	22.97	0.93	8.24	3.04
Zeng et al. (Zeng et al., 2020)	23.21	0.93	7.62	3.16
MIRNet (Zamir et al., 2020)	<u>24.21</u>	0.94	<u>7.00</u>	4.69
Chai et al. (Chai et al., 2020)	21.39	0.88	9.49	4.72
BacklitNet (Ours)	**25.06**	**0.96**	**6.45**	**2.81**

Table 4. Quantitative comparison on the Li and Wu dataset and Vonikakis et al. dataset in terms of NIQE with state-of-the-art methods, including (i) backlit image enhancement methods, (ii) low-light image enhancement methods and (iii) general image enhancement methods. The best and the second-best scores are shown in bold and underlined, respectively.

Method	Li and Wu dataset	Vonikakis et al. dataset
Wang et al. (Wang et al., 2016)	3.41	2.68
Li and Wu (Li and Wu, 2018)	3.31	3.19
ExCNet (Zhang et al., 2019a)	3.21	2.08
Buades et al. (Buades et al., 2020)	3.41	2.14
RetinexNet (Wei et al., 2018)	3.79	2.48
KinD (Zhang et al., 2019b)	3.14	2.59
DeepUPE (Wang et al., 2019a)	3.15	2.05
HDRNet (Gharbi et al., 2017)	3.07	<u>1.99</u>
DPED (Ignatov et al., 2017)	3.10	2.66
Zeng et al. (Zeng et al., 2020)	3.32	2.54
MIRNet (Zamir et al., 2020)	<u>3.06</u>	3.76
Chai et al. (Chai et al., 2020)	3.41	3.19
BacklitNet (Ours)	**2.88**	**1.96**

Table 5. Comparison with state-of-the-art methods on the LOL dataset. The best and the second-best scores are shown in bold and underlined, respectively.

Method	HDRNet (Gharbi et al., 2017)	DPED (Ignatov et al., 2017)	RetinexNet (Wei et al., 2018)	KinD (Zhang et al., 2019b)	DeepUPE (Wang et al., 2019a)	Chai et al. (Chai et al., 2020)	Zeng et al. (Zeng et al., 2020)	MIRNet (Zamir et al., 2020)	BacklitNet (ours)
PSNR	18.75	19.71	16.77	20.86	19.56	16.56	19.97	**24.14**	<u>22.79</u>
SSIM	0.80	0.80	0.55	0.80	0.74	0.79	0.81	<u>0.83</u>	**0.85**
Time (ms)	6.78	58.6	94.8	33.2	7.04	11.3	**2.50**	710	<u>3.86</u>

these datasets. As we can see, our method achieve appealing results in various scenes. Specifically, in terms of visibility of the images, Wang et al. (Wang et al., 2016), DeepUPE (Wang et al., 2019a) and Chai et al. (Chai et al., 2020) cannot produce satisfactory results such a fact can be observed through example shown in Figures 6(b), 6(h), 6(m), 7(b), 7(h) and 7(m). The enhanced results by Li and Wu (Li and Wu, 2018) and MIRNet (Zamir et al., 2020) are prone to produce halo artifacts in Figures 6(c) and 6(i). Additionally, as shown in Figures 6(d) and 6(e), the foreground edges generated by ExCNet (Zhang et al., 2019a) and Buades et al. (Buades et al., 2020) are unnatural. For low-light enhancement methods, the enhanced results by RetinexNet (Wei et al., 2018) and KinD (Zhang et al., 2019b) suffer from severe color deviation and artifacts in backlit scenes, which are demonstrated in Figures 6(f), 6(g), 7(f) and 7(g). DPED (Ignatov et al., 2017) and Zeng et al. (Zeng et al., 2020) decrease the contrast of images in Figures 6(j), 6(k), 7(j) and 7(k). As shown in Figures 6(i) and 7(i), HDRNet (Gharbi et al., 2017) work well for restoring backlit images. However, both of them do not brighten the dark regions specifically. In contrast, BacklitNet is able to enhance backlit images with satisfactory visual effect, harmonized brightness, natural contrast and no objectionable artifacts under diverse light conditions.

User Study. Moreover, we conduct a user study with 50 participants (25 males and 25 females) to evaluate the subjective perception of different methods. For a fair comparison, the user study is conducted in the same environment (room, display and light). Specifically, we randomly select 50 testing images from the BAID dataset, Li and Wu dataset and Vonikakis et al. dataset. Then, we perform a pairwise comparison between the enhanced results of all the methods. In order to avoid subjective bias, the group of images and the order of method pairs are randomized. For each pairwise comparison, there are three options for the participant to choose: "left is better", "right is better" and "no preference". Finally, we use the Bradley-Terry model to estimate the subjective score and rank the evaluation results. As shown in Table 6, results generated by our method are more preferred by human subjects.

Table 6. Psychophysical analysis of competing methods using the Bradley-Terry model. The best and the second-best scores are shown in bold and underlined, respectively.

Method	Bradley-Terry score	Rank
Wang et al. (Wang et al., 2016)	-1.13	10
Li and Wu (Li and Wu, 2018)	-0.15	9
ExCNet (Zhang et al., 2019a)	0.42	6
Buades et al. (Buades et al., 2020)	-1.86	11
RetinexNet (Wei et al., 2018)	-2.44	12
KinD (Zhang et al., 2019b)	1.19	4
DeepUPE (Wang et al., 2019a)	0.39	7
HDRNet (Gharbi et al., 2017)	<u>1.98</u>	<u>2</u>
DPED (Ignatov et al., 2017)	0.88	5
Zeng et al. (Zeng et al., 2020)	-0.06	8
MIRNet (Zamir et al., 2020)	1.69	3
Chai et al. (Chai et al., 2020)	-3.16	13
BacklitNet (Ours)	**2.25**	**1**

Running time. Apart from the quantitative comparison of image quality, we also evaluate the running time of the compared methods in different resolutions. The average time of all methods on the BAID dataset measured by milliseconds is reported in Table 7, which indicates the proposed method significantly advances the existing backlit image enhancement methods (Buades et al., 2020; Li and Wu, 2018; Wang et al., 2016; Zhang et al., 2019a). Compared to high-speed general image

Table 7. Running time (ms) comparison at different resolutions on the proposed BAID dataset with state-of-the-art methods, including (i) backlit image enhancement methods, (ii) low-light image enhancement methods and (iii) general image enhancement methods. The best and the second-best scores are shown in bold and underlined, respectively. "N.A." means that the result is not available.

Method	512×512	1920×1080	5472×3648
Wang et al. (Wang et al., 2016)	102	767	7.6e3
Li and Wu (Li and Wu, 2018)	3.5e4	2.6e6	N.A.
ExCNet (Zhang et al., 2019a)	9.5e3	1.2e4	2.9e4
Buades et al. (Buades et al., 2020)	N.A.	N.A.	N.A.
RetinexNet (Wei et al., 2018)	3.84	283	N.A.
KinD (Zhang et al., 2019b)	3.76	261	N.A.
DeepUPE (Wang et al., 2019a)	7.4	41.2	621
HDRNet (Gharbi et al., 2017)	7.15	39.3	592
DPED (Ignatov et al., 2017)	61.9	482	N.A.
Zeng et al. (Zeng et al., 2020)	**2.45**	**2.55**	**3.17**
MIRNet (Zamir et al., 2020)	762	N.A.	N.A.
Chai et al. (Chai et al., 2020)	7.00	N.A.	N.A.
BacklitNet (Ours)	<u>3.71</u>	<u>4.81</u>	<u>4.92</u>

Table 8. Results of ablation study on different values of λ.

	$\lambda=0$	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$
PSNR	24.34	24.65	**25.06**	24.45	24.19
SSIM	0.955	0.958	**0.959**	0.957	0.955
ΔE^*	7.04	6.72	**6.45**	6.92	7.06
NIQE	8.52	8.48	**8.45**	8.49	8.51

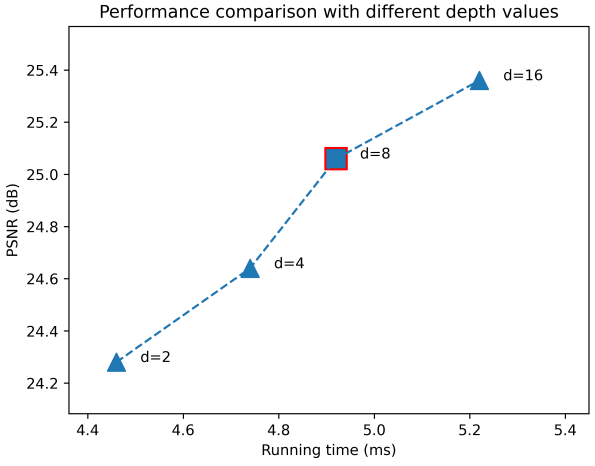


Fig. 8. Results of ablation study on different values of depth d.

Table 9. Results of ablation study on different blocks.

SP	LA	FA	PSNR	SSIM	ΔE^*	NIQE
–	–	–	22.31	0.939	8.53	8.77
✓	–	–	24.18	0.948	6.89	8.64
✓	✓	–	24.86	0.956	6.60	8.53
✓	✓	✓	25.06	0.959	6.45	8.48

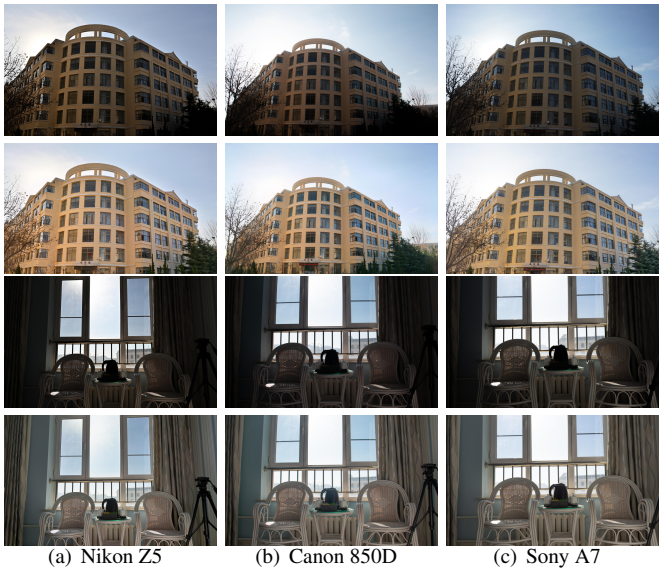
enhancement methods (Chai et al., 2020; Gharbi et al., 2017; Wei et al., 2018; Wang et al., 2019a; Zhang et al., 2019b; Zeng et al., 2020), we obtain a comparable result, which satisfies the real-time requirement of practical application. As the resolution of the image increases, our speed advantage becomes more obvious. Table 5 shows the running time of all methods on the LOL dataset. It is obvious that BacklitNet can achieve a balance of efficiency and performance.

5.5. *Ablation study*

In this section, we explore the influence of different components in our method. Firstly, we conduct an ablation study to explore the influence of the weight factor λ in Equation 4. All the experiments follow the same implementation setup. The quantitative results are reported in Table 8. One can see that, λ=2 outperforms other choices, and exceed λ=0 in PSNR, SSIM, ΔE^* and NIQE by 0.72, 0.004, 0.59 and 0.07, respectively. It is worth noting that even if we set λ=0, our qualitative results are better than the existing methods, which further demonstrates the effectiveness of the designed network architecture. More qualitative results are shown in the supplementary material, from which we see that the visual effect is more appealing when λ=2. Consequently, we select λ=2 for all the experiments.

Furthermore, we conduct experiments to investigate the value of the depth parameter d in Equation 2. We explore the effects of different depth values on both the quality and speed of our method. As shown in Figure 8, the network with a larger depth performs better but becomes more time-consuming. To achieve a trade-off between efficiency and quality, we set the value of depth d to 8.

In addition, we evaluate the effect of semantics perception



(a) Nikon Z5 (b) Canon 850D (c) Sony A7

Fig. 9. Qualitative comparison on different camera brands. The odd and even rows are input and output images, respectively.

block (SP), lighting acquisition block (LA) and fusion adjustment block (FA). Table 9 shows that each block contributes to generating better results.

5.6. Generalization

To explore the performance of BacklitNet on different camera brands, we capture backlit images with Nikon Z5, Canon 850D and Sony A7, which are not used in the BAID dataset. The captured images contain 7 indoor scenes and 8 outdoor scenes and each scene contains three images taken by three different cameras. Although there are slight differences in each set of images due to the difference in camera lenses and sensors, we ensure that the three images in each set are taken under the same scene and lighting condition. Then, we use the model trained on the BAID dataset to enhance these backlit images. As shown in Figure 9, BacklitNet performs well on different camera brands in each scene. Specifically, BacklitNet restores the ill-exposed backlit regions while preserving the harmony of overall brightness without image degradation. The quantitative results on 15 scenes shown in Table 10 further demonstrate that the performance of BacklitNet does not depend on the camera brand.

Table 10. Quantitative comparison on different camera brands in terms of NIQE.

Camera Brand	Nikon Z5	Canon 850D	Sony A7
NIQE	2.94	2.87	2.83

6. Conclusions

In this paper, we propose a novel learning-based network BacklitNet for backlit image enhancement. The main innovations are as follows: 1) we build, to the best of our knowledge, the first large-scale backlit image dataset which contains 3000 backlit images with different backlit degrees and corresponding high-quality references; 2) we introduce the idea of salient object detection into our network to learn enriched semantic and spatial information; 3) we present the prior of backlit image brightness distribution and integrate it into the loss function, which can pay more attention to the local backlit region while preserving the harmony of overall brightness. Through these strategies, we can not only restore the underexposed foreground in various scenes, but also recover harmonious illumination, natural contrast and clear details in backlit images. Consistent achievement of state-of-the-art results on four datasets corroborates the effectiveness and stability of the proposed method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61872112 and 62072141).

References

Afifi, M., Derpanis, K.G., Ommer, B., Brown, M.S., 2020. Learning to correct overexposed and underexposed photos. arXiv 2003.11596.

Backhaus, W.G., Kliegl, R., Werner, J.S., 2011. Color vision: Perspectives from different disciplines. Walter de Gruyter.

Buades, A., Lisani, J.L., Petro, A.B., Sbert, C., 2020. Backlit images enhancement using global tone mappings and image fusion. IET Image Process. 14, 211–219.

Bychkovsky, V., Paris, S., Chan, E., Durand, F., 2011. Learning photographic global tonal adjustment with a database of input/output image pairs, in: CVPR, pp. 97–104.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers, in: ECCV.

Chai, Y., Giryes, R., Wolf, L., 2020. Supervised and unsupervised learning of parameterized color enhancement, in: WACV, pp. 992–1000.

Chen, C., Chen, Q., Xu, J., Koltun, V., 2018. Learning to see in the dark, in: CVPR, pp. 3291–3300.

Chen, J., Adams, A., Wadhwa, N., Hasinoff, S.W., 2016. Bilateral guided upsampling. ACM Transactions on Graphics (TOG) 35, 1–8.

Ciocca, G., Napoletano, P., Schettini, R., 2018. Cnn-based features for retrieval and classification of food images. Computer Vision and Image Understanding 176, 70–77.

Debevec, P.E., Malik, J., 2008. Recovering high dynamic range radiance maps from photographs, in: SIGGRAPH, pp. 1–10.

Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F., 2017. Deep bilateral learning for real-time image enhancement. ACM Transactions on Graphics (TOG) 36, 1–12.

Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M., 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (TOG) 35, 1–12.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN, in: ICCV, pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: CVPR, pp. 770–778.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y., Wu, J., 2020. Unet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP, pp. 1055–1059.

Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Gool, L.V., 2017. Dslr-quality photos on mobile devices with deep convolutional networks, in: ICCV, pp. 3277–3285.

Kohl, S., Romera-Paredes, B., Meyer, C., Fauw, J.D., Ledsam, J.R., Maier-Hein, K.H., Eslami, S.M.A., Rezende, D.J., Ronneberger, O., 2018. A probabilistic u-net for segmentation of ambiguous images, in: NIPS, pp. 6965–6975.

Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A., 2019. Anabranch network for camouflaged object segmentation. Computer Vision and Image Understanding 184, 45–56.

Li, Z., Cheng, K., Wu, X., 2015. Soft binary segmentation-based backlit image enhancement, in: MMSP, pp. 1–5.

Li, Z., Wu, X., 2018. Learning-based restoration of backlit images. TIP 27, 976–986.

Liu, C., Chen, L., Schroff, F., Adam, H., Hua, W., Yuille, A.L., Li, F., 2019a. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation, in: CVPR, pp. 82–92.

Liu, J., Hou, Q., Cheng, M., Feng, J., Jiang, J., 2019b. A simple pooling-based design for real-time salient object detection, in: CVPR, pp. 3917–3926.

Mertens, T., Kautz, J., Reeth, F.V., 2007. Exposure fusion, in: PG, pp. 382–390.

Mittal, A., Soundararajan, R., Bovik, A.C., 2013. Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. 20, 209–212.

Qin, X., Zhang, Z.V., Huang, C., Gao, C., Dehghan, M., Jägersand, M., 2019. Basnet: Boundary-aware salient object detection, in: CVPR, pp. 7479–7489.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, pp. 234–241.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: ICLR.

Sun, Z., Zhang, Y., Bao, F., Shao, K., Liu, X., Zhang, C., 2021. Icyclegan: Single image dehazing based on iterative dehazing model and cyclegan. Computer Vision and Image Understanding 203, 103133.

Trongtirakul, T., Chiracharit, W., Agaian, S.S., 2020. Single backlit image enhancement. IEEE Access 8, 71940–71950.

Ueda, Y., Moriyama, D., Koga, T., Suetake, N., 2020. Histogram specification-based image enhancement for backlit image, in: ICIP, pp. 958–962.

Vaca-Castano, G., Das, S., Sousa, J.P., Lobo, N.D., Shah, M., 2017. Improved scene identification and object detection on egocentric vision of daily activities. Computer Vision and Image Understanding 156, 92–103.

Vazquez-Corral, J., Cyriac, P., Bertalmío, M., 2018. Perceptually-based restoration of backlit images, in: Color and Imaging Conference, pp. 32–37.

Vonikakis, V., Kouskouridas, R., Gasteratos, A., 2018. On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications 77, 9211–9231.

Wang, Q., Fu, X., Zhang, X.S., Ding, X., 2016. A fusion-based method for single backlit image enhancement, in: ICIP, pp. 4077–4081.

Wang, R., Zhang, Q., Fu, C., Shen, X., Zheng, W., Jia, J., 2019a. Underexposed photo enhancement using deep illumination estimation, in: CVPR, pp. 6849–6857.

Wang, W., Yu, K., Hugonot, J., Fua, P., Salzmann, M., 2019b. Recurrent u-net for resource-constrained segmentation, in: ICCV, pp. 2142–2151.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. TIP 13, 600–612.

Wei, C., Wang, W., Yang, W., Liu, J., 2018. Deep retinex decomposition for low-light enhancement, in: BMVC.

Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., Shen, C., Luo, P., 2020. Polarmask: Single shot instance segmentation with polar representation, in: CVPR, pp. 12193–12202.

Yan, Q., Gong, D., Shi, Q., van den Hengel, A., Shen, C., Reid, I.D., Zhang, Y., 2019. Attention-guided network for ghost-free high dynamic range imaging, in: CVPR, pp. 1751–1760.

Zamir, S.W., Arora, A., Khan, S.H., Hayat, M., Khan, F.S., Yang, M., Shao, L., 2020. Learning enriched features for real image restoration and enhancement, in: ECCV.

Zeng, H., Cai, J., Li, L., Cao, Z., Zhang, L., 2020. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. TPAMI .

Zhang, L., Zhang, L., Liu, X., Shen, Y., Zhang, S., Zhao, S., 2019a. Zero-shot restoration of back-lit images using deep internal learning, in: MM.

Zhang, Q., Liu, Y., Zhu, S., Han, J., 2017. Salient object detection based on

super-pixel clustering and unified low-rank representation. Computer Vision and Image Understanding 161, 51–64.

Zhang, Y., Zhang, J., Guo, X., 2019b. Kindling the darkness: A practical low-light image enhancer, in: MM.