ENHANCEMENT OF WEAKLY ILLUMINATED IMAGES BY DEEP FUSION NETWORKS

Yu Cheng[⋆] *Jia Yan*[⋆] *Zhou Wang*[†]

[⋆] Electronic Information School, Wuhan University, Wuhan, China
[†] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada

ABSTRACT

We propose an end-to-end deep fusion-based approach to enhance the quality of images acquired in weak illumination environment. The proposed deep fusion network (DFN), without estimating illumination explicitly, uses a convolutional neural network (CNN) to generate confidence maps as spatial weighting factors to fuse images created by multiple base image enhancement techniques that complement each other in a content-dependent manner. Our tests on both synthetic and real weakly illuminated images show that the proposed DFN approach delivers superior performance in terms of both subjective visual perception and objective quality assessment.

***Index Terms*—** weakly illuminated image enhancement, Retinex, image fusion, convolution neural network

1. INTRODUCTION

High image quality is an essential requirement in many computer vision applications such as object detection and image classification, etc. Poor-quality images captured under weakly illumination conditions may largely affect the accuracy of further tasks. Many approaches have been proposed to enhance the quality of weakly illuminated images. Existing methods fall roughly into three categories: histogram-based, Retinex-based and learning-based.

Histogram-based methods [1] [2] [3] [4] improve image contrast and brightness by remapping pixel values in image intensity histograms. The Retinex model, proposed by Land and McCann [5], describes an image as the product of illumination and reflectance and enhance the degraded images by lightening the illumination component. There have been many variations of the Retinex approach, mainly differing in the illumination estimation methods. Wang et al. [6] estimates illumination by a bright-pass filter. Guo et al. [7] extracts the bright channel of RGB representation as initial coarse illumination and refine it based on a structure prior. Fu et al. [8] smooth the bright channel coarse illumination with a guided filter [9] and then fuse it with its two varieties by hand-crafted weights. Essentially, estimating illumination is an ill-posed problem. Any error in estimation may affect the quality of the restored images. Besides, it requires adjusting image intensity manually when lightening the estimated illumination. Many learning-based methods were proposed in recent years. Lore et al. [10] designed a stacked antoencoder named LLNet. Li et al. [11] utilize convolutional neural network (CNN) to estimate coarse illumination and smooth it using a guided filter. Though learning-based, the learned CNN produces illumination estimation that is subsequently employed under the Retinex framework. Thus this method belongs to both categories.

In this work, we propose a deep fusion network (DFN) approach for image enhancement. DFN is an end-to-end trainable network that combines images created by multiple base enhancement methods, without performing illumination estimation explicitly. Synthetic image pairs are generated to train DFN. Experiments on both synthetic and real weakly illuminated images show that DFN enhances image brightness and contrast while better preserving structural details than state-of-the-art image enhancement algorithms. DFN enhanced images offer better performance in terms of both perceptual visual quality and objective image quality assessment.

2. PROPOSED METHOD

Images captured under weak illumination conditions have defects of low contrast, dark brightness and flat color saturation. Existing image enhancement methods may be successful in improving the quality of such images, but they are often unrobust, producing inconsistent results in many corner case. To make the best use of the advantages of existing methods and produce robust results, we opt to train a network that automatically fuses the results produced by multiple base enhancement algorithms in a content-adaptive manner. The architecture of the proposed method is shown in Fig. 1 and will be elaborated in detail in the following sections.

2.1. Base Image Enhancer

CLAHE. Contrast limited adaptively histogram equalization (CLAHE) [4] is an image enhancement technique to lighten brightness and enhance contrast, the contrast limitation of which can decrease over-enhancement and avoid noise amplification. We denote CLAHE enahcned image as I_{ch}.

Log Correction. Improvement in brightness is insufficient in CLAHE derived image when compared with the significant enhancement of contrast. Log correction is a simple non-linear method to adjust image brightness. With appropriate parameters settings, it can lighten dark regions while keeping bright regions from over-exposure. Thus, we utilize log correction to create the second base enhancement image I_{log}:

$$I_{log} = \log_{11}(1 + 10I) \tag{1}$$

where I is the original weakly illuminated image.

Bright Channel Enhancement. To make image look vivid, an approach which improves color saturation and illumination is desired. The Retinex model assumes that illuminations of the RGB channels are identical and what causes the difference in color is the nonidentical reflectance. Motivated by this assumption, we employ three nonidentical enhanced illuminations of RGB channels to lift color saturation.

More specifically, the bright channel of the original image is extracted as illumination and a three-channel reflectance is yielded according to Retinex model; then, the difference between RGB channels of the original image is enlarged by gamma correction to produce three nonidentical enhanced illuminations; finally, the bright

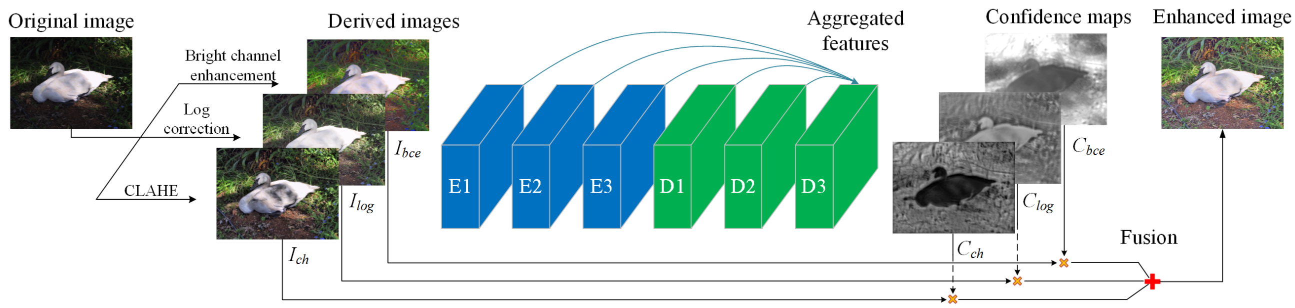


Fig. 1. The architecture of DFN-based image enhancement method.

(a) I (b) I_{ch} (c) I_{log} (d) I_{bce}

Fig. 2. Sample images created by base enhancement methods.

channel enhancement derived image I_{bce} is obtained by multiplying the reflectance with three nonidentical enhanced illuminations:

$$R = \frac{I}{\max(I(r,g,b))} \tag{2}$$

$$I_{bce} = I^{0.5} \cdot R \tag{3}$$

As shown in Fig. 2, color saturation and brightness are significantly improved in I_{bce}.

2.2. Network Architecture

Fig. 1 shows the architecture of DFN-based image enhancement method. The base enhancer generated images are fed as the input to the DFN, which generates three content-dependent confidence maps. Weighted by these confidence maps, element-wise fusion is implemented to combine the significant features of the derived images as follows:

$$I_{en} = I_{ch} \cdot C_{ch} + I_{log} \cdot C_{log} + I_{bce} \cdot C_{bce} \tag{4}$$

where C_{ch}, C_{log} and C_{bce} represent the confidence maps corresponding to derived images I_{ch}, I_{log} and I_{bce}, respectively, and I_{en} is the final enhanced image.

Our DFN is designed based on Encoder-Decoder structure, which has been demonstrated to produce excellent results in many generative tasks [12] [13] [14]. DFN contains 3 Encoders and 3 Decoders. An Encoder consists of 3 convolution layers, and a Decoder is the same except that the first convolution layer is replaced by a deconvolution layer. Each convolution or deconvolution layer is followed by a Leaky rectification layer with 0.1 slopes in the negative range. The weight dimension of each layer is the same: $32\times32\times3\times3$ (output channels $\times$ input channels $\times$ filter height$\times$ filter width), except that the first convolution layer and the last deconvolution layer are $32\times12\times7\times7$ and $3\times192\times3\times3$, respectively. In addition, we add skip connections to DFN. Skip connections effectively accelerate training convergence and contribute to generating clearer results due to repeated use of feature maps [12]. The feature maps produced by all Encoders and shallow Decoders are delivered into the last Decoder, so the number of input channels of the last layer is 192.

(a) Normal image (b) $\alpha = 0.98, \gamma = 1.9$ (c) $\alpha = 0.95, \gamma = 2.6$

Fig. 3. Synthetic weakly illuminated image examples.

2.3. Loss Function and Training Details

A recent study [15] indicates that combinational loss function performs better than single mean square error (MSE). We train DFN with the loss function formed by a combination of L1 and MSE losses:

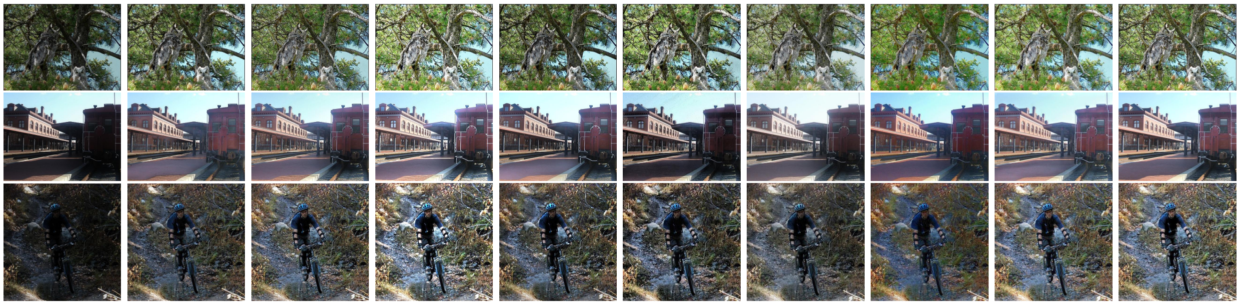
$$\begin{aligned} L(w) = \frac{1}{N}\sum_{i=1}^{N}(0.3\left|\left|F(I_i;w) - J_i\right|\right|_1 \\ +0.7\left|\left|F(I_i;w) - J_i\right|\right|_2^2) \end{aligned} \tag{5}$$

Where N is the number of training pairs, I_i is the weakly illuminated image and J_i is the ground truth, and w and F represent the parameters and the activation of DFN, respectively. The ratio of L1 loss to MSE loss is selected by grid search in set {(0,1), (0.3,0.7), (0.5,0.5), (0.7,0.3), (1,0)}. Better results may be further produced if more ratios were explored.

During training, the patch size is 128×128 and the batch size is 10. ADAM [16] with default parameters is selected as the optimizer. The initial learning rate is 0.00001 which decreases by 70% every 20000 iterations. The total iterations are 80000, which takes 12 hours on a computer with Nvidia TITAN XP GPU.

2.4. Training Data

One major challenge in training-based methods is that a significant number of image pairs under normal and weak illumination conditions are required in the training process. Unfortunately, there is no such dataset readily available. In the research area of image quality assessment [17] [18] and image enhancement [10] [19], gamma correction is widely used to simulate low contrast and weak illumination. Hence, we resort to synthesize training data by gamma correction. First, 600 normal illuminated images are collected from existing image quality assessment datasets [17] [20] and regarded as the ground truth; then, gamma correction is performed on the V channel of the HSV representation of the images to avoid color bias.



(a) Inputs (b) NPEA (c) MF (d) LIME (e) LNET (f) CLAHE (g) LC (h) BCE (i) DFN (j) Ground truth

Fig. 4. Sample results on the synthetic test set. From up to down, images are named *owl*, *station*, and *ride*, respectively.

Specifically, the weakly illuminated images are given by:

$$V_{dark} = \alpha V_{normal}^{\gamma}, \alpha \in (0.8, 1), \gamma \in (1.8, 3.4) \tag{6}$$

Fig. 3 shows examples of synthesized images. Following the assumption that image content is independent of illumination, we randomly sample 7 pairs of parameters for each normal illuminated image. As a result, 4200 image pairs are synthesized in total, of which 3500 pairs are used for training and the rest compose the validation and test sets.

3. EXPERIMENT RESULTS

3.1. Test on Synthetic Images

We first compare our method with state-of-the-art methods NPEA [6], MF [8], LIME [7] and LNET [11], and the three base image enhancers CLAHE, Log Correction (LC) and Bright Channel Enhancement (BCE) used in DFN. The comparison is conducted on a synthetic test set which contains 50 image pairs of various scenes and diverse illumination conditions. None of the 50 test image pairs appear in our training dataset.

Fig. 4 shows sample experiment results from the synthetic test set. NPEA, MF and LNET perform well on back-light image *station*, but generate poor results on images of low light such as *owl* and *ride*. The results of LIME exhibit the highest brightness and contrast, but tend to over-enhance certain regions such as the color distortion of the gravel road in image *ride*. CLAHE enhancer produces non-uniform illuminated results. The results of LC enhancer show flat color. BCE enhancer generates images with over-saturated color. The results produced by the proposed DFN method show more natural color and are the closest to the ground truth. As shown in Table 1, the quantitative evaluations using MSE, PSNR and SSIM metrics confirm the superior performance of the proposed method.

Table 1. Average MSE, PSNR and SSIM values between enhanced images and the ground truth of the synthetic test set

Metric	NPEA	MF	LIME	LNET	CLAHE	LC	BCE	DFN
MSE	1134	845	491	971	2170	1169	1464	**288**
PSNR	18.17	19.74	21.53	19.04	15.30	18.05	17.01	**24.12**
SSIM	0.832	0.863	0.873	0.834	0.742	0.808	0.757	**0.896**

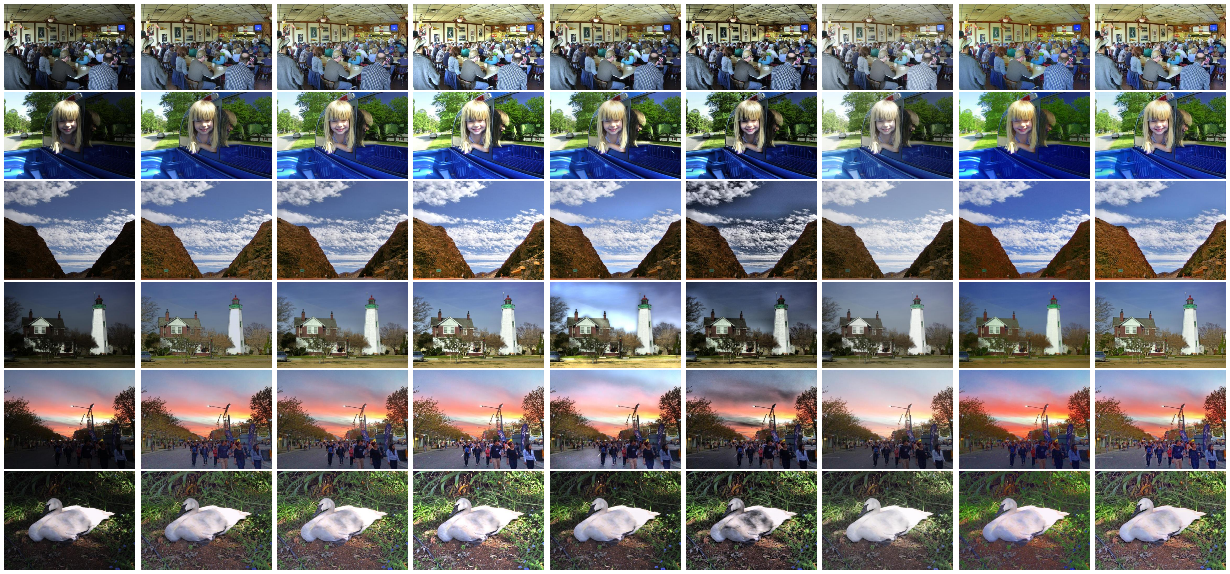
3.2. Test on Real Weakly Illuminated Images

To further evaluate our method, we carry out experiments on real weakly illuminated images collected from existing databases [6] [7] [8]. Fig. 5 shows the visual comparison of different methods. Similar results as for synthetic images are observed. NPEA and MF often generate lower contrast images. MF may also produce halo, for example, near the tree region in the left part of image *tower*. On the other hand, LIME often produces over-enhanced images and may lose the fine details in the over-exposed brightest regions. LNET generates unrealistic artifacts such as the sky in image *tower* and the ground in image *avenue*. The results of CLAHE enhancer show illumination distortion caused from over-enhanced contrast in some bright regions. LC enhancer generates bright but colorless images. BCE enhancer generates excessive color saturation. The proposed method produces more reliable and visually more pleasing results, and the details of scenes and objects are well restored. There is no ground truth corresponding to real weakly illuminated images. Alternatively, to quantitatively evaluate these methods, we use a blind image quality assessment metric NIQE [21], which is based on statistical regularities derived from natural and undistorted images.

As shown in Table 2, NIQE scores of the proposed DFN are ranked first in two images, second in one and second in average. Although LNET achieves first in average, the visual artifacts exist in some result images of LNET. For example, the NIQE score of LNET is the lowest in image *duck*, but the ground is still too dark in the result image as shown in Fig. 5(e). The NIQE results indicate that the proposed method shows advantage in performance on real weakly illuminated images in terms of image quality and naturalness.

Table 2. NIQE evaluation on real weakly illuminated images

Image	NPEA	MF	LIME	LNET	CLAHE	LC	BCE	DFN
meeting	3.22	3.45	3.83	**3.08**	3.43	**3.01**	3.15	3.11
girl	2.07	2.13	1.88	1.70	2.00	**1.56**	1.80	**1.53**
mountain	2.02	2.08	**1.98**	2.01	2.48	2.04	2.30	**1.95**
tower	**3.00**	3.37	3.09	3.09	3.51	3.37	3.42	**3.07**
avenue	2.52	2.40	2.32	**2.28**	**2.27**	2.41	2.38	2.30
duck	2.81	2.82	2.59	**2.25**	2.64	**2.28**	2.56	2.51
average	2.61	2.71	2.61	**2.40**	2.72	2.45	2.60	**2.41**



(a) Inputs (b) NPEA (c) MF (d) LIME (e) LNET (f) CLAHE (g) LC (h) BCE (i) DFN

Fig. 5. Sample real weakly illuminated images. From up to down, images are named *meeting*, *girl*, *mountain*, *tower*, *avenue* and *duck*, respectively.

(a) Patch of image *meeting* (b) Generated by DFN (c) Generated by DEN

Fig. 6. Visual comparisons between DFN and DEN enhanced image.

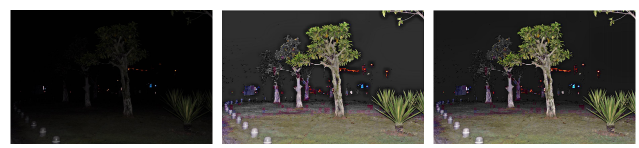
3.3. Effectiveness of Fusion Design

It should be noted that the outputs of DFN are confidence maps instead of the final enhanced image. To better understand the effectiveness of the fusion design in DFN, we train a network DEN (Direct Enhancement Network) to directly generate an enhanced image without the fusion operation. The architecture of DEN is the same as DFN except that the outputs of DEN are regarded as RGB channels of the final enhanced image instead of the confidence maps for fusion.

Fig. 6 shows the visual comparisons of the enhanced image generated by DFN and DEN respectively. As shown in the yellow box, more structural details are kept in DFN enhanced images. By contrast, DEN directly generates enhanced results, and as a result image details are more likely to be distorted by the convolution operation and the nonlinear activation within the network.

3.4. Artificial Edges

Due to the non-edge-preserving sliding windows convolutions in network, smoothing edges appear on confidence maps, resulting artificial edges exist in final fused images, especially when the intensity difference between the two sides of the edge is large. To some extend, this case can be alleviated by manipulating guided filtering [9] on confidence maps. As shown in Fig.7, the artificial edges are faded when the confidence maps are guided filtered using the log corrected image as guidance image. However, the limitation is that the guided filtering post-process is fully independent of network training, which make the trained network parameters a sub-optimal solution, so how to embed the edge-preserving operation into our network needs to be explored in future works.



(a) Weakly illuminated image (b) Enhanced by DFN without guided filtering post-process (c) Enhanced by DFN with guided filtering post-process

Fig. 7. Artificial edges and guided filtering post-process.

4. CONCLUSION

We propose an end-to-end deep fusion network based approach for the enhancement of weakly illuminated images without estimating illumination explicitly. By training on synthetic images, the proposed DFN learned to generate confidence maps to adaptively fuse three derived images created by base image enhancement techniques. Experiments on both synthetic and real weakly illuminated images demonstrate that the proposed DFN approach can produce enhanced results of better subjective and objective quality than state-of-the-art methods.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (61701351).

5. REFERENCES

[1] Robert Hummel, "Image enhancement by histogram transformation.," *Computer Graphics and Image Processing*, vol. 6, no. 2, pp. 184–195, 1977.

[2] Chulwoo Lee, Chul Lee, and Chang-Su Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[3] Tarik Arici, Salih Dikbas, and Yucel Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transactions on image processing*, vol. 18, no. 9, pp. 1921–1935, 2009.

[4] Karel Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994.

[5] Edwin H Land, "The retinex," *American Scientist*, vol. 52, no. 2, pp. 247–264, 1964.

[6] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[7] Xiaojie Guo, Yu Li, and Haibin Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[8] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[9] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European conference on computer vision*. Springer, 2010, pp. 1–14.

[10] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[11] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognition Letters*, vol. 104, pp. 15–22, 2018.

[12] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, 2016, pp. 2802–2810.

[13] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang, "Deep image harmonization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 2.

[14] Wei Xiong, Wenhan Luo, Lin Ma, Wei Liu, and Jiebo Luo, "Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2364–2373.

[15] Yu Liu, Guanlong Zhao, Boyuan Gong, Yang Li, Ritu Raj, Niraj Goel, Satya Kesav, Sandeep Gottimukkala, Zhangyang Wang, Wenqi Ren, et al., "Improved techniques for learning to dehaze and beyond: A collective study," *arXiv preprint arXiv:1807.00202*, 2018.

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

[17] KGGZ Gu, Guangtao Zhai, Weisi Lin, and Min Liu, "The analysis of image contrast: From quality assessment to automatic enhancement.," *IEEE Trans. Cybernetics*, vol. 46, no. 1, pp. 284–297, 2016.

[18] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[19] Seonhee Park, Soohwan Yu, Minseo Kim, Kwanwoo Park, and Joonki Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, 2018.

[20] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017.

[21] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a completely blind image quality analyzer.," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.