From Under- to Overexposure: Single Image Contrast Enhancement via Semi-Supervised Learning

Ziwen Li
D201980722@hust.edu.cn
Yuehuan Wang*
yuehwang@hust.edu.cn
Ruowang Chang
changruowang@hust.edu.cn

School of Artificial Intelligence and Automation,Huazhong University of Science and Technology,
Wuhan, China

Abstract

Poor lighting conditions result in photographs with low contrast. Most existing methods focus on low-light enhancement and perceptual quality improvement, and have not taken both under- and overexposure into consideration. In this paper, we propose a novel semi-supervised learning method for single image contrast enhancement. The supervised branch is trained using paired data under the constraint of supervised losses. While in the unsupervised branch, we explore content consistency and illumination prior as loss functions to train the network. The advantages of the proposed approach are two folds. First, guided by ground truth images, the supervised branch learns well to preserve image details and suppress noise. Second, the unsupervised branch learns to adapt to more illumination intensities and diverse illumination environments, which bridges the gap between various lighting conditions. With the help of the semi-supervised strategy, our method uses a single model to enhance both underexposed and overexposed images, and generalizes well to various lighting conditions. Experimental results show that the proposed method outperforms the state-of-the-art methods quantitatively and qualitatively.

1 Introduction

With the popularity of mobile camera devices, many people take photos to record their lives. The exposure settings for photos involve exposure time, aperture, ISO sensitivity, etc., which will directly affect the overall brightness of the rendered image. When expertise is lacking, images taken under poor lighting conditions can suffer from many degradation problems, the most common of which is low contrast due to under- and overexposure. Exposure errors and low contrast can lead to loss of detail and color distortion, which greatly degrade the visual quality of an image and affect the performance on high-level vision tasks (e.g., object detection and recognition).

Many conventional algorithms have been proposed for single image contrast enhancement. Histogram equalization and S-curve adjustment are simple and effective methods to

* indicates the corresponding author.

enhance image contrast. However, such methods perform pixel-level spatial transformations through region tone mapping functions, which tend to lose image details, amplify noise and produce undesirable results. Methods based on the Retinex model can produce results with better visual effects, including SSR [13], SMRCR [14], but still face some problems such as difficulties in estimating the illumination map and producing underexposed or overexposed local detail results.

Deep learning techniques have developed rapidly and achieving good results in low-level vision tasks, such as super-resolution [4], image denoising [21], depth estimation [5, 6]. In the field of image enhancement, many excellent deep learning algorithms [1, 8, 12, 18, 19, 25, 28, 30, 31, 36] have also been proposed, and these algorithms have achieved good enhancement results. Both over- and underexposed images are common in realistic photography, however, most algorithms focus only on the enhancement of underexposed images.

In general, when making paired datasets, the exposure values are usually set to fixed values such as $\{\pm 0.5, \pm 1.0, \pm 2.0\}$ for convenience. However, actual degraded images may have more exposure levels such as $\{\pm 0.8, \pm 1.5, \pm 1.8, \pm 2.2\}$, and more complex lighting environments such as backlit environments, and collecting a large number of ground truth images for these data is very time-consuming labor. Moreover, there are domain gaps in the data under different lighting conditions, and using supervised learning only would be limited to specific data and would not generalize well to other unseen images.

Considering the above issues, in this work, we propose a semi-supervised learning method for single image contrast enhancement. The proposed framework is divided into a supervised branch and an unsupervised branch. In the supervised branch, the learning of the network is driven by the supervised loss functions using ground truth images as references, such as pixel loss, perceptual loss, and adversarial loss. In the unsupervised branch, we design the gradient perceptual loss based on content consistency and the illumination control loss based on the illumination prior, and integrate the total variation loss function to constrain the network. The supervised and unsupervised branches share the network structure and model weights and are trained in an end-to-end iterative manner. Our approach embraces the advantages of both supervised and unsupervised methods. On the one hand, the supervised branch uses ground truth images to provide the accurate guidance for the learning of the network, which helps to reconstruct the structural details of the images and remove the noise. On the other hand, the unsupervised branch introduces degraded images with more illumination intensities and diverse illumination environments, and the knowledge from the unsupervised branch bridges the gap between the various lighting conditions. With the help of the semi-supervised learning strategy, our method enhances image contrast with a single model and performs well against the state-of-the-art methods under various lighting conditions.

The main contributions of this work are summarized as follows.

- We propose a semi-supervised learning method for single image contrast enhancement, where various lighting conditions can be handled well using only a single model. To the best of our knowledge, it is the first attempt to apply a semi-supervised learning strategy to correct underexposed and overexposed images (not only low-light image enhancement).

- We develop the gradient perceptual loss based on content consistency and the illumination control loss based on illumination prior to constrain the learning of unsupervised networks.

- We experimentally demonstrate that the proposed method outperforms the state-of-the-art methods both qualitatively and quantitatively. Furthermore, our method produces visually pleasing results and has better generalization performance.

2 Related Work

2.1 Conventional Methods

The HE-based method enhances contrast by adjusting the image luminance distribution. CLAHE [39] constructs mapping functions based on local histograms to avoid over enhancement. The S-curve adjustment uses the tone mapping function for contrast enhancement. Yuan *et al*. [34] splits the image into several different sub-regions and performs S-curve transformations on each sub-region separately. However, these methods do not consider the relationship of adjacent pixels and are prone to produce unnatural results.

The Retinex model decomposes the image into an illumination map and a reflection map, and adjusts them to obtain enhanced results. Early Retinex methods [13, 14] directly used reflection maps as enhanced results. Wang *et al*. [29] improved the Retinex model for non-uniform illumination to make the enhanced results more natural and realistic. Li *et al*. [17] proposed a robust Retinex model with noise components. Guo *et al*. [9, 10] used the maximum pixel value as the initial illumination map, which was further optimized using the structure-aware prior. However, the recovery effect of the Retinex model depends mainly on the quality of the evaluated illumination maps, which can easily yield underexposed or overexposed results in local regions.

2.2 Deep Learning-based Methods

Supervised Learning Methods. Supervised learning methods require paired data with ground truth image for training. Lore *et al*. [18] uses an autoencoder to learn patch-level low-light enhancement and denoising tasks. Based on the Retinex model, Wei *et al*. [30] and Zhang *et al*. [36, 37] use CNN to decompose the image into an illumination map and a reflection map, and then adjust the illumination map to recover the final result. Lv *et al*. [19] learns low-light image/video enhancement via the multi-branch fusion network. Wang *et al*. [28] extracts global and local features to obtain an illumination map of multiple channels to recover the image, and uses a bilateral grid to obtain real-time performance. Moran *et al*. [23] performs local enhancement of images by predicting the parameters of filters. Xu *et al*. [31] combines frequency decomposition to extract the structural information and image details, and is capable of removing noise simultaneously. Ren *et al*. [25] proposes an improved RNN to extract edge features to better restore the structural details of the images. Zhu *et al*. [38] proposed the fusion network to fuse multi-exposure images in combination with an edge enhancement module to recover extremely bright and dark areas of the image. Afifi *et al*. [1] proposes a coarse-to-fine deep learning model to correct both underexposed and overexposed photos from both color and detail perspectives, and provides a new dataset with multiple exposure images and correctly exposed reference images.

Unsupervised Learning Methods. The unsupervised methods mainly utilize image prior or generative adversarial networks to enhance image contrast. Guo *et al*. [8, 16] gradually adjust the image by predicting multiple high-order curves. Zhang *et al*. [35] learns image

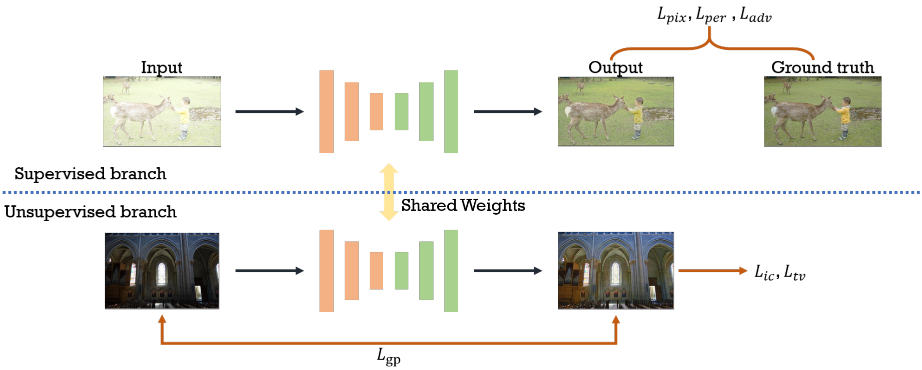


Figure 1: The proposed semi-supervised framework containing a supervised branch and an unsupervised branch for single image contrast enhancement.

enhancement using S-shaped curves in the zero-shot learning setting. Jiang *et al.* [12] and Chen *et al.* [3] employ generative adversarial networks to learn effective image enhancement in an unpaired manner.

Semi-Supervised Learning Methods. The semi-supervised methods uses both paired images and images without ground truth for learning. Yang *et al.* [32, 33] uses semi-supervised learning for low-light enhancement, supervised learning to learn a recursive bandwidth representation of images from coarse to fine, and unsupervised learning to improve the quality of bandwidth reconstructed images using generative adversarial networks.

In contrast to existing methods, our method uses a semi-supervised strategy to learn contrast enhancement. Specifically, most methods focus on low-light enhancement and generic image enhancement, while our method can enhance both underexposed and overexposed images. Our semi-supervised strategy is different from DRBN [32, 33]. Firstly, DRBN uses a two-stage design, while our method is end-to-end training and prediction. Second, DRBN utilizes a semi-supervised strategy to bridge the gap between fidelity and perceptual quality, while our approach bridges the gap between various lighting conditions.

3 Proposed Method

The proposed method learns single image contrast enhancement via semi-supervised learning, as shown in Fig.1. In this section, we describe the design of the semi-supervised framework, the network architecture and the loss function in detail.

3.1 Semi-Supervised Framework

In semi-supervised learning framework, a labeled paired dataset $D_s = \{(x_i, y_i), i \in (1, 2, ..., M)\}$ and an unlabeled dataset $D_u = \{(z_i), i \in (1, 2, ..., N)\}$ are provided, where M and N denote the size of the dataset, x_i and z_i denotes the input image, and y_i denotes the ground truth image corresponding to x_i, respectively.

Our semi-supervised learning framework can be divided into a supervised branch and an unsupervised branch as shown in Fig1. In the supervised branch, given a CNN model $f(\cdot)$ parameterized by θ_s, we can obtain the enhanced image $f(x_i|\theta_s)$ from the image x_i in the unpaired dataset D_s. Since ground truth images are available in the labeled dataset, we get the optimal parameter θ_s by minimizing the loss function between the input image and the ground truth image, such as L1 loss, MSE loss and perceptual loss. The supervised loss function can be expressed as:

$$\sum_i \mathcal{L}\left(f\left(x_i|\theta_s\right), y_i\right) \tag{1}$$

In the unsupervised branch, given a CNN model $f(\cdot)$ with parameters θ_u, feeding an image z_i from an unlabeled dataset D_u into the network yields an enhanced result $f(z_i|\theta_u)$. Since there is no ground truth image, we can explore the image prior as the loss functions to optimize the network parameters, such as content, illumination, smoothness, sparsity, etc. The unsupervised loss function can be expressed as:

$$\sum_i \mathcal{L}\left(f\left(z_i|\theta_u\right)\right) \tag{2}$$

Under the supervision of ground truth images, the supervised branch is able to learn knowledge about image detail recovery and noise removal. However, using only the supervised branch leads to restriction to specific data, and to avoid this, the unsupervised branch is used to acquire knowledge about how to adapt to more illumination intensities and diverse illumination environments. The goal is to integrate the supervised and unsupervised knowledge in a single model, so we share the network architecture and model weights, denoted as $\theta_s = \theta_u$. This allows the benefits of both supervised and unsupervised to be exploited, ultimately improving the generalization performance of the model under various lighting conditions.

3.2 Network Architecture

Enhancement Network. We use a Unet-like structure as our enhancement network. The encoder contains three downsampling modules for reducing the resolution of the feature maps. The downsampling module on each scale consists of one convolutional layer and three residual blocks. The downsampling operation is implemented by the `Conv` layer with the stride size of 2, instead of the pooling layer which loses information. We use the residual block from Nah *et al.* [24], which removes all normalization and pooling layers. Symmetrically with the encoder, the decoder contains three upsampling modules that improve the resolution of the feature maps. The upsampling module at each scale stacks three residual blocks and a `ConvTranspose` layer with the stride size of 2. The skip connection fuses shallow and deep features between the downsampling and upsampling modules at the same scale, which helps to preserve the shallow structural features. Finally, we also use a residual learning strategy to learn delta images instead of the enhanced results to retain the detailed information in the input images.

Discrimiator Network. The discrimination task is relatively easy and we directly use PatchGAN [11] as the discriminator network. PatchGAN stacks 5 convolutional layers, each followed by LeakyReLU as the activation function. The discriminator determines whether the image is a generated image or a real image.

3.3 Loss Function

Given low contrast images I_l as input, the network can generate enhanced images I_e. In the supervised branch, we use the ground truth images I_{gt} to compute the supervised loss functions, including pixel loss L_{pix}, perceptual loss L_{per}, and adversarial loss L_{adv}. In the unsupervised branch, we calculate the unsupervised loss functions, including illumination control loss L_{ic}, gradient perceptual loss L_{gp}, and total variation loss L_{tv}.

The total loss function can be expressed as

$$\begin{aligned} L_{total} =& L_{super} + L_{unsuper} \\ =& \lambda_1 L_{pix} + \lambda_2 L_{per} + \lambda_3 L_{adv} + \beta_1 L_{ic} + \beta_2 L_{gp} + \beta_3 L_{tv} \end{aligned} \tag{3}$$

where the parameter λ_1, λ_2, λ_3,β_1,β_2 and β_3 denote the weights of each loss.

Supervised Loss. In low-level vision tasks, the pixel loss is often used as an objective function to optimize the network. The pixel loss improves the overall similarity by reducing the pixel differences between the enhanced result and the ground truth image. However, the pixel loss does not take into account structural details and tends to blur the edges. Here we use the mean squared error as the pixel loss:

$$L_{pix} = \mathbb{E}||I_e - I_{gt}||_2 \tag{4}$$

The perceptual loss function focuses on the perceptual quality of the enhanced image. We use a pre-trained VGG-19 network to extract the semantic features of I_e and I_{gt}, and ϕ_i denotes the features extracted at the ith layer. Then the distance between features is calculated using L2 distance as the perceptual loss:

$$L_{per} = \mathbb{E}||\phi_i(I_e) - \phi_i(I_{gt})||_2 \tag{5}$$

The training of the generative adversarial network is a game between the generator and the discriminator, where the generator learns to generate images to deceive the discriminator, and the discriminator learns to distinguish the generated images from the real ones as much as possible. The adversarial loss is used to improve the detail of the image and produce impressive visual effects, and the adversarial loss is defined as:

$$L_{adv} = \mathbb{E}_{I_e}[\log(1 - D(I_e))] + \mathbb{E}_{I_{gt}}[\log D(I_{gt})] \tag{6}$$

Unsupervised Loss. Gradient can extract the spatial structure of an image and is often considered to represent the content of an image. According to the observation of Enlighten-GAN [12], the classification results of VGG network are not very sensitive to the brightness change of the image. Based on the characteristics of VGG networks and image gradients, we propose the gradient perceptual loss to ensure that the content before and after image enhancement remains as consistent as possible. We first extract the gradient maps of the input image and the generated image separately, and then use the pre-trained VGG-19 network to extract the perceptual features of the image gradient maps and calculate the distance between the features using the mean squared error. Unlike the perceptual loss, the gradient perceptual loss is a measure of the similarity between the enhanced result and the original image, rather than the ground truth image. We use ϕ_i to denote the features of the ith layer extracted by VGG19 and ∇ to denote the gradient operator, then the gradient perceptual loss can be defined as:

$$L_{gp} = \mathbb{E}||\phi_i(\nabla I_e) - \phi_i(\nabla I_l)||_2 \tag{7}$$

A badly exposed image will have an overall shift in brightness distribution relative to a normal image. To force the illumination distribution of an enhanced image to be close to normal, we propose illumination control loss. To be specific, we first calculate the illumination map of the image, which can be expressed as the average of the three RGB channels. Instead of using the overall brightness of the image, we first chunk the illumination map by 16*16 and then calculate the average value of each patch to obtain the average illumination map X. Finally, we make the expected value of the illumination of each patch close to the normal illumination value K to constrain the overexposed or underexposed area. The value of K is empirically set to 0.6. According to the above description, the illumination control loss can be expressed as:

$$L_{ic} = \mathbb{E}||X - K||_1 \tag{8}$$

We also use the total variation loss to ensure spatial smoothing. The total variation loss is a common regular term used to reduce the difference between adjacent pixels. The total variation loss function can be defined as:

$$L_{tv} = \mathbb{E}||\nabla_x(I_e) + \nabla_y(I_e)||_1 \tag{9}$$

where ∇_x and ∇_y denote the gradient operator in the horizontal and vertical directions, respectively.

4 Experiments

4.1 Implementation Details

All experiments are implemented using Pytorch on an Nvidia Titan V GPU. We randomly cropped the training image to 256×256 pixels. The Adam optimizer with default parameters was used to optimize the enhancement network and the discriminator network. We trained the network for 60 epochs using the fixed learning rate of 0.0001. The hyperparameters λ_1, λ_2, λ_3, β_1, β_2 and β_3 are set to 1000, 10, 10, 1, 1, $1e^{-4}$ as trade-offs between losses.

We use the SICE [2] dataset for the supervised training and evaluation. The SICE dataset contains image sequences taken at different fixed exposure levels, such as outdoor scenes with exposure values $\{\pm 0.5, \pm 0.7, \pm 1.0, \pm 2.0, \pm 3.0\}$. We randomly select 345 image sequences from the Part1 of SICE as the training dataset and the remaining 15 image sequences as the test dataset. We use the DICM [15], VV[1] and Fusion [27] as unsupervised training datasets, which contain various lighting conditions after data augmentation. We further evaluate the generalization ability on the MEF [20], NPE [29] and LIME [9] datasets, which are not visible during training.

4.2 Evaluation

Quantitative Comparison. We use PSNR and SSIM metrics to compare quantitatively with several state-of-the-art methods. The compared methods include conventional methods: CLAHE [39], LIME [9], WVM [7], and deep learning methods: RetinexNet [30],

[1] https://sites.google.com/site/vonikakis/datasets

Table 1: Quantitative comparison using PSNR and SSIM metrics on the SICE dataset. The best results are highlighted in red. Higher PSNR and SSIM values indicate better results.

Method	Underexposed PSNR	Underexposed SSIM	All PSNR	All SSIM
CLAHE [39]	13.51	0.48	13.37	0.51
WVM [7]	14.32	0.60	13.90	0.64
LIME [9]	16.33	0.70	13.50	0.65
RetinexNet [30]	15.53	0.70	13.82	0.69
MBLLEN [19]	15.34	0.63	15.52	0.65
EnlightenGAN [12]	15.88	0.67	13.54	0.65
Zero-DCE [8]	16.83	0.66	14.19	0.62
Ours	17.32	0.70	17.47	0.72

MBLLEN [19], EnlightenGAN [12], ZeroDCE [8]. Because some methods are designed for underexposure, we divided the experiments into two groups, (1) for underexposed images only, and (2) for all images (combining underexposed and overexposed images).

Table 1 shows the quantitative comparison with the state-of-the-art methods. For underexposed images, our method performs on par with the state-of-the-art methods. For all images (combining underexposed and overexposed images), our method achieves the highest scores in PSNR and SSIM metrics, indicating that our method can effectively handle contrast enhancement under various lighting conditions.

Qualitive Comparison. We qualitatively compare our method with other state-of-the-art methods. Fig.2 and Fig.3 show the enhanced results for overexposed and underexposed images, respectively. As shown in Fig.2, CLAHE loses image detail, LIME and RetinexNet produce over-enhanced results, WVM, MBLLEN and ZeroDCE still have underexposed areas, and EnlightenGAN produces color distortion. As shown in Fig.3, MBLLEN still suffers from over-saturated results, while the other methods are unable to recover overexposed images well. Unlike these methods, our method can cope with various exposure errors simultaneously and produce visually pleasing results.

Generalization. We use BRISQUE [22] and PIQE [26] as no-reference quality accessment metrics for unlabeled images, where BRISQUE is based on the predictable statistical properties of natural images and PIQE is related to human visual perception. The lower BRISQUE and PIQE values are better. We test the generalization ability on the NPE, MEF and LIME datasets. Note that the trained model has not seen these images. As seen in Table 2, our method achieves overall top ranking results on both BRISQUE and PIQE metrics, indicating that our method has better perceptual quality and better generalization performance.

Ablation Study. We perform several ablation studies on MEF dataset to demonstrate the effectiveness of the **unsupervised branch** and the **unsupervised loss**. We remove the unsupervised branch and place the unsupervised losses on the supervised branch, denoted by w/o UB. Each loss is removed separately and denoted by w/o L_{ic}, w/o L_{gp}, and w/o L_{tv}, respectively. We replace the gradient perceptual loss in the full model with the perceptual loss, denoted as $L_{gp} \rightarrow L_{per}$. The quantitative and qualitative ablative results are presented in Table 3 and Fig.4, respectively.

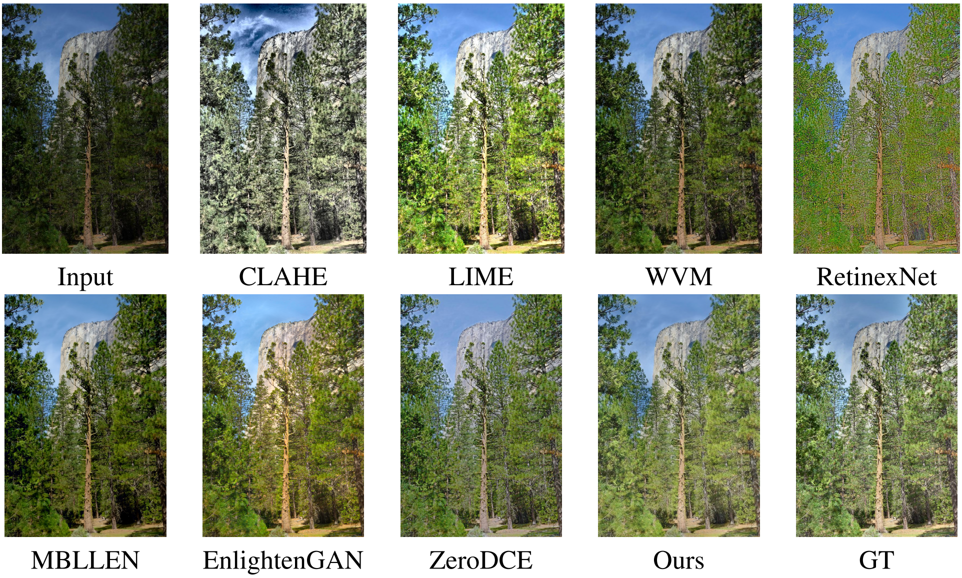


Figure 2: Qualitative comparison of underexposed images

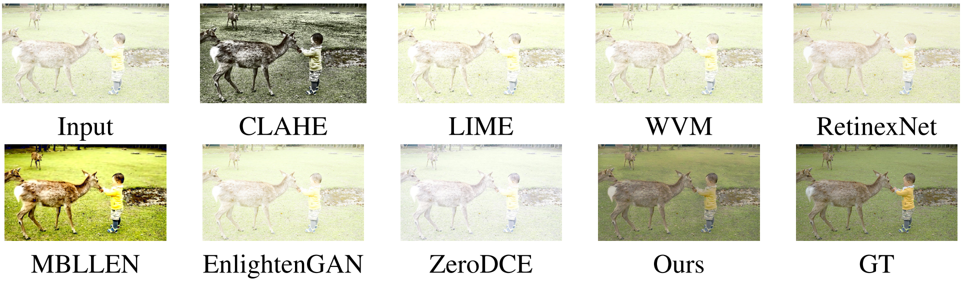


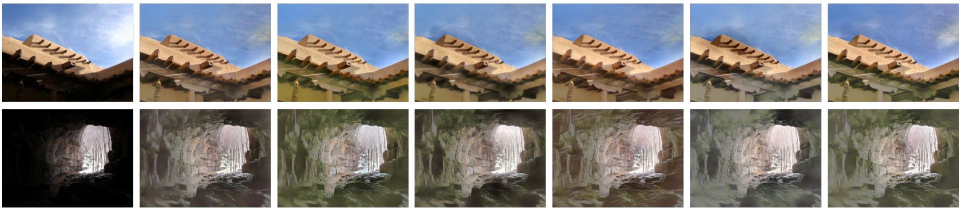
Figure 3: Qualitative comparison of overexposed images

Table 2: Generalization evaluation using BRISQUE and PIQE metrics on unlabeled datasets (LIME, MEF, NPE). Red color indicates the best result. Lower BRISQUE and PIQE values indicate better results.

Method	LIME	MEF	NPE	AVG
CLAHE [39]	34.46/49.52	35.38/55.26	29.16/40.83	33.00/48.54
WVM [7]	24.18/34.80	26.76/40.06	27.25/37.34	26.06/37.40
LIME [9]	23.57/38.20	29.62/43.32	28.60/40.20	27.26/40.57
RetinexNet [30]	26.10/42.77	29.28/40.91	29.05/38.97	28.14/40.88
MBLLEN [19]	30.39/52.28	30.33/47.07	28.61/44.91	29.78/48.09
EnlightenGAN [12]	20.61/33.72	25.71/31.34	24.87/32.81	23.73/32.62
ZeroDCE [8]	23.33/35.87	29.68/36.59	29.96/37.57	27.66/36.68
Ours	22.93/23.21	21.42/22.86	24.96/26.62	23.10/24.23

Table 3: Quantitative ablation studies of the unsupervised branch and the unsupervised loss on MEF datset. The best results are highlighted in red. Lower BRISQUE and PIQE values indicate better results.

Model	BRISQUE	PIQE
Ours	21.42	22.86
w/o UB	24.66	25.03
w/o L_{ic}	23.44	28.18
w/o L_{gp}	27.83	27.39
w/o L_{tv}	22.85	25.66
$L_{gp} \rightarrow L_{per}$	28.35	25.67



Input Ours w/o UB w/o L_{ic} w/o L_{gp} w/o L_{tv} $L_{gp} \rightarrow L_{per}$

Figure 4: Visualization results of ablation studies for the unsupervised branch and the unsupervised loss.

As shown in Table 3, the best results were obtained by the full model, which proved the effectiveness of the unsupervised branch and each unsupervised loss. Fig.4 shows the visualization results of the ablation study. The results of removing the unsupervised branch appear color distorted compared to the full model. This shows that the unsupervised branch can avoid limiting to specific data and learn to adapt to more lighting conditions. When the illumination control loss L_{ic} is removed, the enhanced brightness may not be uniform enough. Without gradient perceptual loss L_{gp}, the image will appear jagged textures, suggesting that L_{gp} helps to retain the content of input image. Removing the total variation loss L_{tv}, the enhanced result loses the relative relationships of the neighborhoods and introduces undesired artifacts. Compared with perceptual loss L_{per}, gradient perceptual loss L_{gp} can better preserve the image content under different illumination conditions and avoid blurred results.

5 Conclusion

In this paper, we propose a semi-supervised method for single image contrast enhancement. The supervised branch is trained using supervised losses under the guidance of ground truth images. In the unsupervised branch, we use gradient perceptual loss, illumination control loss and total variation loss to constrain the network. Experimental results demonstrate that the proposed method outperforms the state-of-the-art methods.

References

[1] Mahmoud Afifi, Konstantinos G. Derpanis, Bjorn Ommer, and Michael S. Brown. Learning multi-scale photo exposure correction. In *CVPR*, 2021.

[2] 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.

[3] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. In *CVPR*, 2018.

[4] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014.

[5] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *ICCV*, 2015.

[6] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NeurIPS*, 2014.

[7] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, 2016.

[8] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, 2020.

[9] Xiaojie Guo. LIME: a method for low-light image enhancement. In *ACM MM*, 2016.

[10] 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, 2017.

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

[12] 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. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[13] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, 6(3):451–462, 1997.

[14] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997.

[15] 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.

[16] Chongyi Li, Chunle Guo, and Change Loy Chen. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[17] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018.

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

[19] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. MBLLEN: Low-light image/video enhancement using CNNs. In *BMVC*, 2018.

[20] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

[21] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *NeurIPS*, 2016.

[22] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21 (12):4695–4708, 2012.

[23] Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. DeepLPF: Deep local parametric filters for image enhancement. In *CVPR*, 2020.

[24] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017.

[25] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019.

[26] N Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *NCC*, 2015.

[27] Qiuhong Wang, Xueyang Fu, Xiao-Ping Zhang, and Xinghao Ding. A fusion-based method for single backlit image enhancement. In *ICIP*, 2016.

[28] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, 2019.

[29] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[30] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018.

[31] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2020.

[32] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*, 2020.

[33] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Transactions on Image Processing*, 30:3461–3473, 2021.

[34] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In *ECCV*, 2012.

[35] Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. In *ACM MM*, 2019.

[36] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 2019.

[37] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021.

[38] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *AAAI*, 2020.

[39] Karel Zuiderveld. Contrast limited adaptive histograph equalization. *Graphics Gems IV*, pages 474–485, 1994.