

Global and Local Enhancement Networks for Paired and Unpaired Image Enhancement

Han-Ul Kim¹[0000-0001-7450-6600], Young Jun Koh²[0000-0003-1805-2960], and Chang-Su Kim¹[0000-0002-4276-1831]

¹ Korea University, Seoul, Korea
hanulkim@ncl.korea.ac.kr, changsukim@korea.ac.kr

² Chungnam National University, Daejeon, Korea
yjkoh@cnu.ac.kr

Abstract. A novel approach for paired and unpaired image enhancement is proposed in this work. First, we develop global enhancement network (GEN) and local enhancement network (LEN), which can faithfully enhance images. The proposed GEN performs the channel-wise intensity transforms that can be trained easier than the pixel-wise prediction. The proposed LEN refines GEN results based on spatial filtering. Second, we propose different training schemes for paired learning and unpaired learning to train GEN and LEN. Especially, we propose a two-stage training scheme based on generative adversarial networks for unpaired learning. Experimental results demonstrate that the proposed algorithm outperforms the state-of-the-arts in paired and unpaired image enhancement. Notably, the proposed unpaired image enhancement algorithm provides better results than recent state-of-the-art paired image enhancement algorithms. The source codes and trained models are available at <https://github.com/hukim1124/GleNet>.

Keywords: Image enhancement, unpaired learning, generative adversarial network

1 Introduction

Nowadays, many people take photographs to record everyday life as well as important events. However, uncontrolled environments often make photographs have low dynamic ranges or distorted color tones. Therefore, image enhancement becomes more popular that edits photographs to improve their aesthetic quality. Image enhancement methods can be categorized into global and local approaches. The former derives a transformation function that maps input color to output color. On the other hand, the latter performs spatial filtering to determine a pixel color according to local neighborhood information. Professional software applications such as Photoshop provide various global and local enhancement tools to support manual image enhancement. However, the manual process is time-consuming. Moreover, its results highly depend on users' skills and experience.

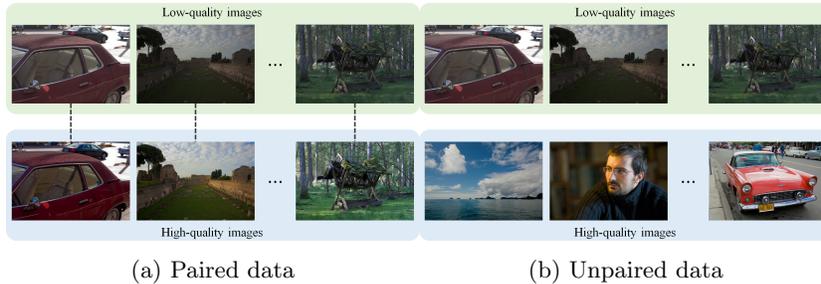


Fig. 1: Examples of paired and unpaired data

For automatic image enhancement, many studies have been proposed. Most early studies focus on the global approach since it is more stable and requires less computational complexity than the local approach. However, using only a single transformation function may be insufficient to produce satisfying enhanced images. In contrast, recent studies [8, 25, 30, 33] based on deep learning mainly take the local approach. These methods learn a robust pixel-wise mapping from lots of paired data, which consists of inputs and ground-truth enhanced images, and provide promising enhanced images. However, they require many image pairs of low-quality and high-quality images as in Fig. 1(a). To overcome this problem, unpaired image enhancement, which does not require the image pairs, has drawn much attention to many researches [4, 5, 12, 19, 26, 34]. Especially, generative adversarial networks (GANs) [4, 5, 19] or reinforcement learning [12, 26, 34] are employed to achieve unpaired image enhancement using unpaired data in Fig. 1(b). However, despite some progress by existing studies, their results are not satisfying when compared with existing paired image enhancement methods.

In this paper, we propose two networks, global enhancement network (GEN) and local enhancement network (LEN), to achieve both paired and unpaired image enhancement. GEN performs the channel-wise intensity transform, which can be trained much easier than the pixel-wise prediction based on U-Net architecture [27]. LEN conducts spatial filtering to refine GEN results. We then develop two training schemes for paired learning and unpaired learning. Especially, we propose a two-stage training scheme for unpaired learning based on generative adversarial networks. Experiments on the MIT-Adobe 5K dataset [2] demonstrate that the proposed method outperforms the state-of-the-arts in both paired and unpaired image enhancement. Moreover, it is shown that the proposed unpaired method yields better enhanced results than conventional paired methods.

To summarize, this work has three main contributions:

- We propose GEN and LEN for both paired and unpaired image enhancement.
- We propose the two-stage training scheme for unpaired image enhancement.
- The proposed method shows outstanding performance on the MIT-Adobe 5K dataset.

2 Related Work

Early studies on image enhancement mainly focus on improving the global contrast of an input image [16, 24]. They often derive a transformation function that maps input pixel values to output pixel values. The global contrast technique uses a single mapping function for all pixels in an entire image. For instance, power-law (gamma) and log transformations [9] are well-known global methods. Histogram equalization [9] and its variants [1, 17, 21–23, 29, 32] modify the histogram of an image to stretch its limited dynamic range. Retinex methods [3, 6, 7, 11, 13, 14, 31, 35] decompose an image into reflectance and illumination [20], and modify the illumination to enhance a poorly lit image. However, these methods may not emulate the complex mapping function between an image and its professionally enhanced version.

Recent studies on image enhancement take data-driven approaches that learn the mapping between input and enhanced images using a large dataset. For this purpose, Bychkovsky *et al.* [2] introduced the MIT-Adobe 5K dataset, which contains 5,000 input images and enhanced images retouched by 5 different photographers. This dataset is widely adopted to train deep neural networks. Yan *et al.* [33] predicts a pixel-wise color mapping using image descriptors from a deep neural network. Lore *et al.* [25] first adopt an autoencoder approach to enhance low-light images. Gharbi *et al.* [8] achieved real-time image enhancement by developing deep bilateral learning, which predicts local affine transforms. Based on the retinex theory, Wang *et al.* [30] proposed a deep network to estimate an image-to-illumination mapping function. These deep learning methods [8, 25, 30, 33] yield promising enhancement performances, but they are limited in that they demand many pairs of input and enhanced images to train their networks.

Collecting pairs of input and manually enhanced images is a labor-intensive task. To overcome this problem, unpaired learning methods [4, 5, 12, 19, 26, 34], which do not require paired data, have been proposed. Park *et al.* [26] adopted deep reinforcement learning to mimic step-by-step human retouching processes. Also, they proposed a distort-and-recover training scheme, which distorts a high-quality image to generate a pseudo input and trains networks to enhance the generated pseudo input to be similar to the corresponding high-quality image. Deng *et al.* [5] employed a generative adversarial network (GAN) to develop an aesthetic-driven image enhancement method. Chen *et al.* [4] proposed an adaptive weighting scheme for stable training of two-way GANs. Hu *et al.* [12] integrated an adversarial loss into reinforcement learning to generate a sequence of enhancement operations. Yu *et al.* [34] trained local exposures with deep reinforcement adversarial learning, which divides an image into sub-images and enhances them with different policies. Recently, Kosugi and Toshihiko [19] combined reinforcement learning and adversarial learning to control tools in professional image editing software. However, these unpaired learning methods provide relatively poor results than paired learning methods.

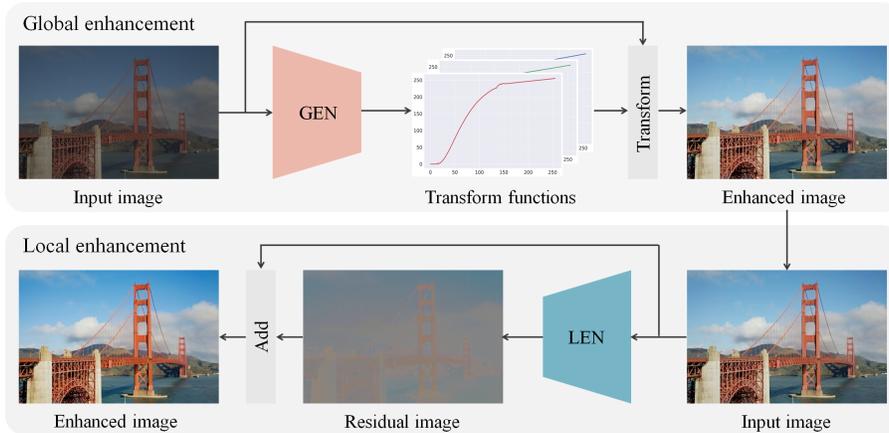


Fig. 2: Overview of the proposed global enhancement network (GEN) and local enhancement network (LEN).

3 Proposed Algorithm

3.1 Model

Fig. 2 shows an overview of the proposed image enhancement framework. First, we develop GEN that produce channel-wise intensity transform functions to achieve global image enhancement. Second, we learn LEN, which performs spatial filtering, to refine global enhanced images. Let us describe each network subsequently.

Global Enhancement Network: Let $\mathbf{I}(p) = (I_r(p), I_g(p), I_b(p))^T$ denote 8-bit intensity for red, green, and blue channels at pixel position p . Also, let $\mathbf{w}_c = [w_{c,0}, \dots, w_{c,255}]^T$ denote the transformation function for the channel $c \in \{r, g, b\}$, whose k th element $w_{c,k}$ maps intensity k in I_c to intensity $w_{c,k}$ in the output intensity \tilde{I}_c . Thus, the transformed intensity for the channel c at pixel p is defined as

$$\tilde{I}_c(p) = \mathbf{v}_p^T \mathbf{w}_c \quad (1)$$

where \mathbf{v}_p^T denotes a 256-dimensional one-hot vector, whose $I_c(p)$ th element is 1 and the others are 0.

Given an RGB image of size 256×256 , GEN produces a 768-dimensional vector $\mathbf{w} = \mathbf{w}_r \parallel \mathbf{w}_g \parallel \mathbf{w}_b$, which is a concatenated vector of three transformation functions $\mathbf{w}_r, \mathbf{w}_g, \mathbf{w}_b$. Table 1 specifies the detailed architecture of GEN. We employ the inverted residual block in MobileNetV3 [4] to reduce the number of network parameters. All “Conv” operations except the last one include convolution filters, batch normalization, and swish activation. The last “Conv” only contains convolution filters. Finally, we perform the channel-wise intensity

Table 1: Specification for global enhancement network.

Stage	Operator	Output Resolution	Channels
1	Conv5x5	128×128	16
2	Inverted Residual, 5x5	64×64	24
3	Inverted Residual, 5x5	32×32	40
4	Inverted Residual, 5x5	16×16	80
5	Inverted Residual, 5x5	8×8	112
6	Conv1x1 & Pool 8x8	1×1	768
7	Conv1x1	1×1	768

transformation to obtain global enhanced images by sequentially applying output functions \mathbf{w}_r , \mathbf{w}_g , and \mathbf{w}_b to (1). Note that GEN can be trained in an end-to-end manner, since the intensity transformation is differentiable operation.

The proposed GEN has advantages when compared with the conventional image enhancement network in [4], which contains a decoder to produce pixel-wise color predictions. First, GEN can enhance an image regardless of its resolution scale by performing the channel-wise intensity transformation unlike the pixel-wise color prediction [4]. In other words, the channel-wise intensity transformation can produce enhanced images without any image resize process, while the pixel-wise color prediction often requires the resize process according to the spatial size of input images. Second, GEN can save the memory for network parameters, since it does not require a decoder part to restore the spatial resolution of enhanced images. Third, training GEN is much easier than the networks that have the encoder-decoder architecture. In Section 4.1, we will clarify that GEN requires less training steps for the convergence than the encoder-decoder architecture does.

Different from early global enhancement methods, we do not suppose that the three color intensity transformation functions should be a monotonic function. Most existing global methods focus on enhancing the gray intensity instead of color intensities and suppose the monotonic constraint to prevent annoying artifacts due to the reservation of the gray intensity ordering. However, the monotonic constraint does not work in the channel-wise intensity transformation. Fig. 3 shows examples of pairs of input and retouched images and their channel-wise intensity transformation functions. In these example, we see that there are many non-monotonic functions between low-quality and high-quality images.

Local Enhancement Network Despite many strengths in GEN, it is limited in that GEN considers only one-to-one mapping. However, as in Fig. 3, there are many one-to-many mappings, which are delineated by shading in the channel-wise transformation functions, between low-quality and high-quality images. Moreover, GEN may experience difficulty on removing noises and blur in an input image through the channel-wise intensity transformation. Therefore, we

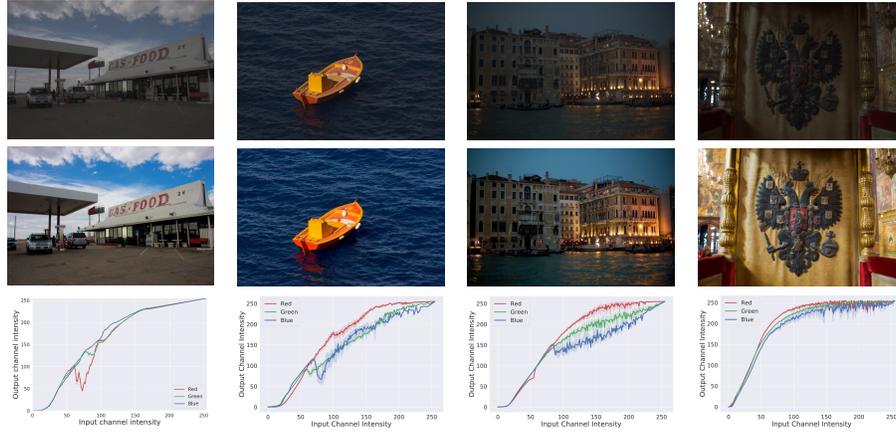


Fig. 3: From top to bottom, input images, retouched images, and channel-wise intensity transformation functions.

develop LEN, which performs spatial filtering for local enhancement, to overcome these limitations of GEN.

Table 2 provides the specification of the LEN architecture. LEN has an encoder-decoder structure. The encoder takes an enhanced images $\tilde{I}_{\text{global}}$ of GEN. The encoder reduces the spatial resolution of I_{global} to exploit larger receptive fields for spatial filtering, while the decoder performs up-sampling to restore the spatial resolution. LEN uses the inverted residual block to decrease the number of network parameters. In Table 2, “Upsample” denotes the bilinear interpolation to increase the size of the feature map with scale factor 2. “Concat” layers in the 6th, 8th, and 10th stages concatenates the previous stage results with the outputs of the 3rd, 2nd, and 1st stages, respectively. The last convolution layer yields the residual image $\Delta\tilde{I}_{\text{local}}$ for enhancing local regions of $\tilde{I}_{\text{global}}$. Finally, the enhanced image \tilde{I} is obtained by

$$\tilde{I} = \tilde{I}_{\text{global}} + \Delta\tilde{I}_{\text{local}}. \quad (2)$$

3.2 Learning

We describe training schemes for unpaired learning and paired learning. First, we train GEN and LEN using pairs of low-quality and high-quality images for paired learning. Second, we propose the two-stage training scheme to learn GEN and LEN in unpaired learning. Let us explain each training scheme subsequently.

Paired Learning: Suppose the set of image pairs $\{(I_i^{\text{LQ}}, I_i^{\text{HQ}})\}_{i=1}^N$ are available, where I_i^{LQ} and I_i^{HQ} are the low-quality image and its high-quality image, respectively. We train GEN and LEN simultaneously to minimize the color loss

Table 2: Specification for local enhancement network.

Stage	Operator	Output Resolution	Channels
1	Conv5x5	H × W	16
2	Inverted Residual, 5x5	H/2 × W/2	24
3	Inverted Residual, 5x5	H/4 × W/4	40
4	Inverted Residual, 5x5	H/8 × W/8	80
5	Inverted Residual, 5x5	H/8 × W/8	40
6	Upsample & Concat	H/4 × W/4	80
7	Inverted Residual, 5x5	H/4 × W/4	24
8	Upsample & Concat	H/2 × W/2	48
9	Inverted Residual, 5x5	H/2 × W/2	16
10	Upsample & Concat	H × W	32
11	Conv5x5 & Add	H × W	3

and the perceptual loss between the estimated image \tilde{I}_i^{HQ} and the ground-truth high-quality image I_i^{HQ} . The total loss is defined as

$$\mathcal{L}_p = \|\tilde{I}_i^{\text{HQ}} - I_i^{\text{HQ}}\|_1 + \lambda_p \sum_{k=2,4,6} \|\phi^k(\tilde{I}_i^{\text{HQ}}) - \phi^k(I_i^{\text{HQ}})\|_1. \quad (3)$$

The color loss in the first term penalizes the mean absolute error between the predicted and ground-truth high-quality images. On the other hand, the second term is the perceptual loss [15] to encourage the enhanced image and the ground-truth image to have similar features on the pre-trained embedding space. Thus, we employ VGG-16 [28] pre-trained on ImageNet to extract features. In (3), $\phi^k(\cdot)$ denotes the feature, which is extracted from the k th VGG-16 layer. The hyper parameter λ_p balances two loss components.

Unpaired Learning: Let $\{I_i^{\text{LQ}}\}_{i=1}^M \in \mathbf{I}^{\text{LQ}}$ and $\{I_j^{\text{HQ}}\}_{j=1}^N \in \mathbf{I}^{\text{HQ}}$ be the sets of low-quality images and high-quality images, respectively. Our goal in unpaired learning is to learn GEN and LEN using unpaired training samples \mathbf{I}^{LQ} and \mathbf{I}^{HQ} . First, we adopt the adversarial learning framework to train GEN. We regard GEN as a generator. Also, the architecture of a discriminator is the same as the generator except the last convolution layer to produce a scalar output that discriminates between generated samples and real samples. Then, we design two types of GANs, where the one enhances low-quality to high-quality images (Fig. 4(a)), and the other degrades high-quality to low-quality images (Fig. 4(b)). Let $G_e : \mathbf{I}^{\text{LQ}} \rightarrow \mathbf{I}^{\text{HQ}}$ and $G_d : \mathbf{I}^{\text{HQ}} \rightarrow \mathbf{I}^{\text{LQ}}$ denote mapping functions for generators to enhance and degrade input images, respectively. Also, let D_e and D_d denote discriminators to discriminate between high-quality images $\{I^{\text{HQ}}\}$ and enhanced images $\{G_e(I^{\text{LQ}})\}$ and between low-quality images $\{I^{\text{LQ}}\}$ and degraded images $\{G_d(I^{\text{HQ}})\}$, respectively.

We employ the Wasserstein GAN with gradient penalty (WGAN-GP) [10] to define objective functions of generators and discriminators. The discriminator

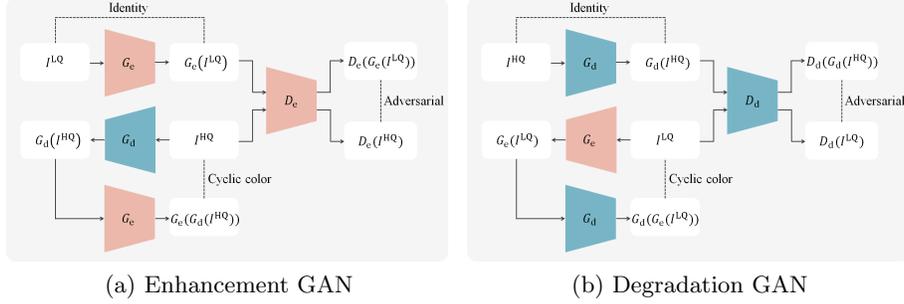


Fig. 4: The network architectures of (a) enhancement GAN and (b) degradation GAN.

losses \mathcal{L}_{D_e} and \mathcal{L}_{D_d} for D_e and D_d are defined as

$$\begin{aligned}\mathcal{L}_{D_e} &= D_e(G_e(I^{LQ})) - D_e(I^{HQ}) + \lambda_{gp}(\|\nabla_{\tilde{I}^{HQ}} D_e(\tilde{I}^{HQ})\|_2 - 1)^2 \\ \mathcal{L}_{D_d} &= D_d(G_d(I^{HQ})) - D_d(I^{LQ}) + \lambda_{gp}(\|\nabla_{\tilde{I}^{LQ}} D_d(\tilde{I}^{LQ})\|_2 - 1)^2.\end{aligned}\quad (4)$$

In both discriminator losses \mathcal{L}_{D_e} and \mathcal{L}_{D_d} , the first two terms are adversarial losses. The last terms are gradient penalty to satisfy the Lipschitz constraint in WGAN-GP. The image \tilde{I}^{HQ} is obtained by interpolating $G_e(I^{LQ})$ and I^{HQ} with random weights. Similarly, \tilde{I}^{LQ} is computed using $G_d(I^{HQ})$ and I^{LQ} . The hyper-parameter λ_{gp} is a weight for the gradient penalty.

Also, for training generators G_e and G_d , we define the loss functions as

$$\begin{aligned}\mathcal{L}_{G_e} &= -D_e(G_e(I^{LQ})) + \lambda_i \|G_e(I^{LQ}) - I^{LQ}\|_1 + \lambda_c \|G_e(G_d(I^{HQ})) - I^{HQ}\|_1 \\ \mathcal{L}_{G_d} &= -D_d(G_d(I^{HQ})) + \lambda_i \|G_d(I^{HQ}) - I^{HQ}\|_1 + \lambda_c \|G_d(G_e(I^{LQ})) - I^{LQ}\|_1\end{aligned}\quad (5)$$

which is composed of adversarial, identity, and cyclic color losses. The adversarial loss (first term) penalizes the Wasserstein distance between generated images and real images. The identity loss (second term) prevents generated images from becoming too different from input images. Note that the identity supports stable training by reducing the space of possible mapping functions. Also, we design the cyclic color loss (third term), which enforces that the reconstructed image should be similar to its origin. For instance, the cyclic color loss minimizes the mean absolute error between $G_e(G_d(I^{HQ}))$ and I^{HQ} to train the enhancement GAN in Fig. 4(a). To this end, we can learn the generator G_e to yield enhanced images that are similar to the high-quality images in \mathbf{I}^{HQ} . Note that the cyclic color loss is different from the cyclic consistency loss in [36], which argues that I^{LQ} and $G_d(G_e(I^{LQ}))$ should be similar for training G_e . In Section 4.2, we will verify the effectiveness of the proposed cyclic color loss as compared with the cyclic consistency loss in [36].

Next, we train LEN using the trained GEN. Notice that training LEN is more difficult than training GEN since LEN is designed to produce the pixel-

wise prediction, which requires the more complicated mapping function than the channel-wise intensity transform in GEN. Therefore, we take a different approach to train LEN. More specifically, we degrade a high-quality image I^{HQ} in the training samples using the generator G_d to obtain a pseudo pair of low-quality and high-quality images, $(G_d(I^{\text{HQ}}), I^{\text{HQ}})$. Then, we enhance the degraded image using the generator G_e . To this end, we can obtain a pseudo pair of global enhanced image and high-quality images, $(G_e(G_d(I^{\text{HQ}})), I^{\text{HQ}})$. Finally, we train LEN using this paired data by minimizing the loss in (3).

4 Experiments

Experiments are organized as follows. In Section 4.1, we verify the effectiveness of GEN and LEN when pairs of low-quality and high-quality images are available. We compare the performance of the proposed GEN and LEN with state-of-the-art algorithms based on paired learning. In Section 4.2, we train GEN and LEN using the proposed unpaired learning and perform the comparison with state-of-the-art methods in unpaired image enhancement.

For all experiments, we use the MIT-Adobe 5K dataset [2] that contains 5,000 input images, each of which was manually retouched by five different photographers (A/B/C/D/E). Thus, there are five sets of 5,000 pairs of input and retouched images, one set for each photographer. Among these sets, we use high-quality images retouched by photographer C only for training and test as done in most existing image enhancement algorithms. We split the 5,000 images into 500 and 4,500 images, which are used for the training and test sets, respectively. We use all 4,500 image pairs in training set for paired learning. In contrast, for unpaired learning, the 4,500 image pairs divided into two groups, each of which has 2,250 image pairs. Then, 2,250 input images in the first group are included in the low-quality image set, while 2,250 retouched images in the second group are used for the high-quality image set. Notice that images in the low-quality set and the high-quality set are not overlapped.

For quantitative assessment, we employ PSNR and SSIM, which measure, respectively, color and structural similarity between predicted and ground-truth high-quality images.

4.1 Paired Learning

For paired learning, we use 4,500 training image pairs to train GEN and LEN. We minimize the loss in (3) using the Adam optimizer [18] with an learning rate of 1.0×10^{-4} . The training is iterated for 25,000 mini-batches. The mini-batch size is 16. For data augmentation, we randomly rotate images by multiples of 90 degrees. The parameter λ_p in (3) is fixed to 0.04.

First, we verify the effectiveness of the proposed GEN by comparing the channel-wise intensity transform in GEN with the pixel-wise color prediction. For this comparison, we design a baseline network, which produces pixel-wise enhanced results. More specifically, the baseline network has the encoder-decoder

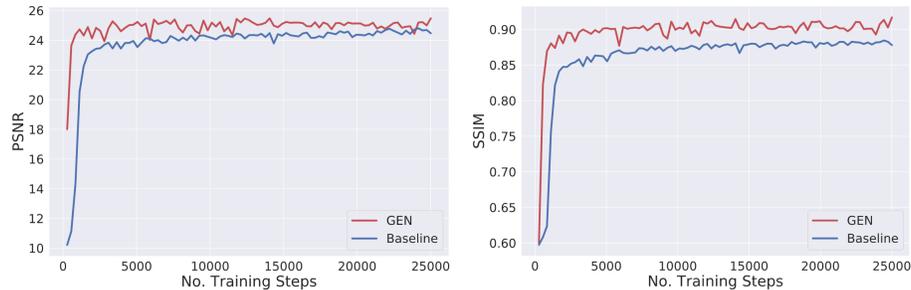


Fig. 5: PSNR and SSIM scores of GEN and the baseline network according to the number of training iterations.

Table 3: Quantitative comparison of the proposed algorithm with state-of-the-art methods based on paired learning. The best results are boldfaced.

Method	HDRNet [8]	DPE [4]	DUPE [30]	GEN	GEN & LEN
PSNR	23.44	22.34	23.61	25.47	25.88
SSIM	0.882	0.873	0.887	0.917	0.925

architecture, where the encoder has the same structure as the encoder in GEN, and the decoder consists of 6 up-sample blocks to perform bilinear interpolation, concatenation, and convolution filtering, subsequently. The detailed architecture of the baseline network can be found in the supplementary material. Fig. 5 shows PSNR and SSIM scores of GEN and the baseline network according to training steps. We observe that GEN achieves faster training than the baseline through the channel-wise intensity transform. This is because the space of possible functions for the intensity transform is much smaller than that of the pixel-wise color transform. Notably, the proposed GEN surpasses the best performance of the baseline networks within 5,000 iterations in both metrics.

Next, we compare the proposed GEN and LEN with recent state-of-the-art algorithms [4, 8, 30]. For comparison, we obtain the results of conventional algorithms using the source codes and settings provided by respective authors. Table 3 reports PSNR and SSIM scores. The proposed GEN significantly outperforms all conventional algorithms. For instance, it convinces margins of 1.86dB and 0.030 against DUPE [30] in terms of PSNR and SSIM. Also, LEN overcomes the one-to-many mapping problems of GEN by exploiting local neighbor information. Note that LEN further improves results of GEN, and thus joint GEN and LEN (GEN & LEN) achieves the best performance in both metrics.

Fig. 6 illustrates the efficacy of the proposed LEN. In Fig. 6(b), GEN yields slightly different color tones in the sky, water, and a tractor compared to photographer C’s retouched images in Fig. 6(d). This is because GEN fails to deal with one-to-many transformation. For instance, since sky and ground regions in

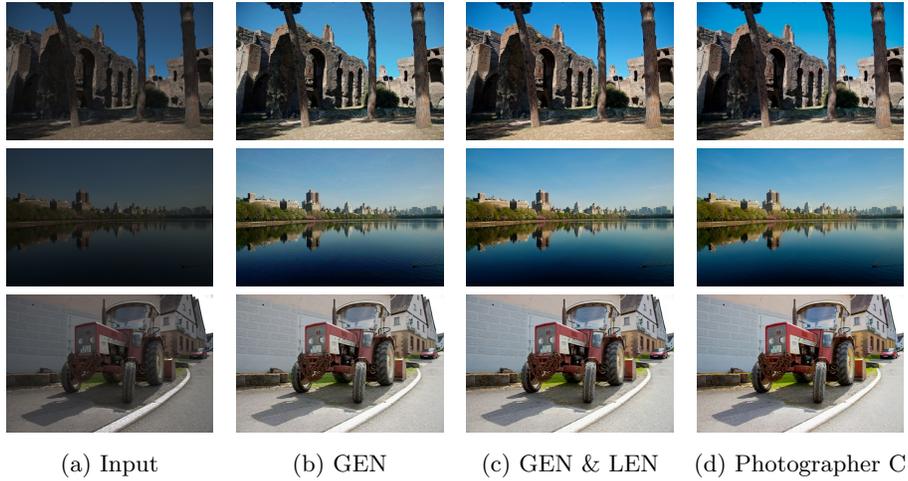


Fig. 6: Qualitative comparison between GEN and GEN & LEN methods.

the first row in Fig. 6(a) have similar intensities in the blue channel. Then, GEN produces similar blue intensities between the sky and ground regions. Therefore, as in Fig. 6(b), blue intensities in the sky region are not sufficiently enhanced since GEN is tailored to enhance the ground region. LEN overcomes this problem through effective spatial filtering, as in Fig. 6(c). Compare to GEN, GEN & LEN yields more visually pleasing results, which have similar color tones to the manually retouched images in Fig. 6(d).

Fig. 7 compares the proposed algorithm with DUPE [30] qualitatively. In Fig. 7(b), DUPE [30] fails to express similar color tones and brightness to photographer C’s retouched images in Fig. 7(d). Also, the results of DUPE have limited contrast. On the other hand, the proposed algorithm successfully yields high-quality images with vivid color tones, which are similar to photographer C’s retouched images.

4.2 Unpaired Learning

We perform the two-stage training for unpaired image enhancement. Specifically, we train GEN and LEN for 5,000 and 25,000 mini-batches, respectively, where the size of mini-batch is fixed to 8. The Adam optimizer [18] is employed again. We set the initial learning rate to 1.0×10^{-4} , and reduce it by a factor of 0.5 every 10,000 mini-batches. Hyper parameters λ_{gp} , λ_i , λ_c , and λ_p are set to 10, 5, 50, and 0.04, respectively. For data augmentation, we randomly rotate images by multiples of 90 degrees.

In Table 4, we compare the proposed algorithm with the conventional unpaired image enhancement algorithms [4, 19, 26] using the MIT-Adobe 5K dataset. The proposed GEN outperforms all conventional algorithms since it can be eas-



Fig. 7: Qualitative comparison of the proposed algorithm with DUPE [30].

Table 4: Quantitative comparison of the proposed algorithm with state-of-the-art methods based on unpaired learning. The best results are boldfaced.

Method	D&R [26]	DPE [4]	FRL [19]	GEN	GEN & LEN
PSNR	21.60	21.86	22.27	23.74	23.82
SSIM	0.875	0.880	0.881	0.885	0.889

ily trained with unpaired data. This indicates that the channel-wise intensity transform in GEN is suitable for unpaired learning. Also, we see that GEN & LEN improves both PSNR and SSIM scores, as compared with GEN, and yields the best results in all metrics. It is worth pointing that GEN & LEN outperforms all conventional paired image enhancement algorithms in Table 3, even though only unpaired data is used for training.

Fig. 8 qualitatively compares the proposed algorithm with FRL [19]. The proposed GEN & LEN model provides more faithful images than FRL. For instance, FRL fails to increase brightness sufficiently, as in images in Fig. 8(b). In contrast, the proposed algorithm successfully enhances low-quality images to be similar to high-quality images retouched by Photographer C.

The proposed training scheme in unpaired learning generates pseudo paired data to train LEN. For the generation of pseudo paired data, each high-quality image is first degraded ($I^{\text{HQ}} \rightarrow G_d(I^{\text{HQ}})$) and then the degraded image is enhanced to mimic global image enhancement ($G_d(I^{\text{HQ}}) \rightarrow G_e(G_d(I^{\text{HQ}}))$). We qualitatively analyze the accuracy of the pseudo pair generation. Fig. 9(a) and (b) show degraded images and real low-quality images, respectively. We observe that the degraded images $G_d(I^{\text{HQ}})$ are well imitated with real images I^{LQ} . Also,

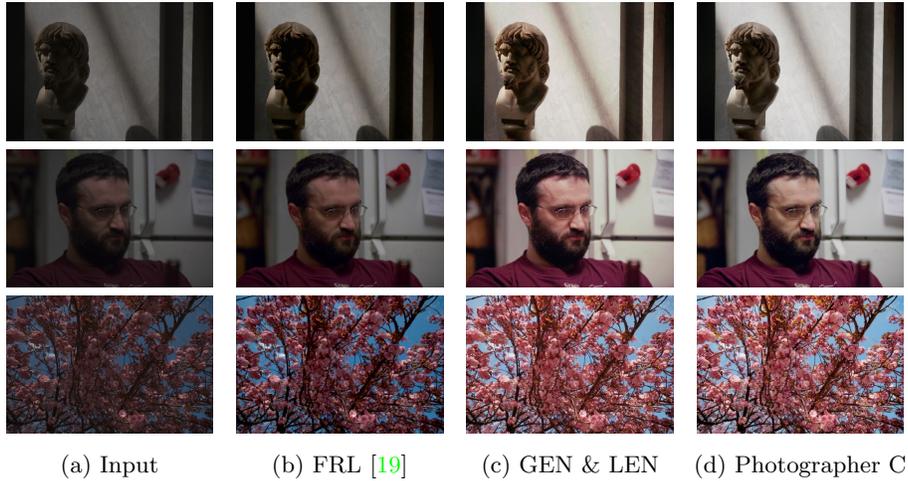


Fig. 8: Qualitative comparison of the proposed algorithm with FRL [19].

Table 5: PSNR and SSIM scores in different training schemes.

Model	Training	PSNR	SSIM
GEN	WGAN-GP	23.05	0.868
GEN & LEN*	WGAN-GP	21.78	0.847
GEN	CWGAN-GP	23.11	0.869
GEN & LEN	CWGAN-GP	23.18	0.874
GEN	Proposed	23.74	0.885
GEN & LEN	Proposed	23.82	0.889

it is worth pointing out that global enhanced images from the degraded images and low-quality images are similar to each other, as in Fig.(c) and (d).

In Table 5, we analyze the efficacy of the proposed training schemes. “WGAN-GP” denotes the training scheme that only adopts the adversarial loss in (4) and (5). In other words, it does not utilize the degradation GAN. Notice that pseudo paired data cannot be obtained without the degradation GAN. Therefore, joint GEN and LEN (GEN & LEN*) in “WGAN-GP” training scheme is learned using the adversarial loss only. Low PSNR and SSIM scores in GEN & LEN* indicate that pseudo paired data is essential to train LEN. On the other hand, GEN in “WGAN-GP” yields reasonable performance as compared with GEN & LEN*. Because GEN based on the intensity transform is more suitable for unpaired learning than LEN. “CWGAN-GP” training scheme substitutes the cyclic color loss in (5) with the cyclic consistency loss in [36]. We can see that the cyclic color loss is more effective than the cyclic consistency loss in [36].

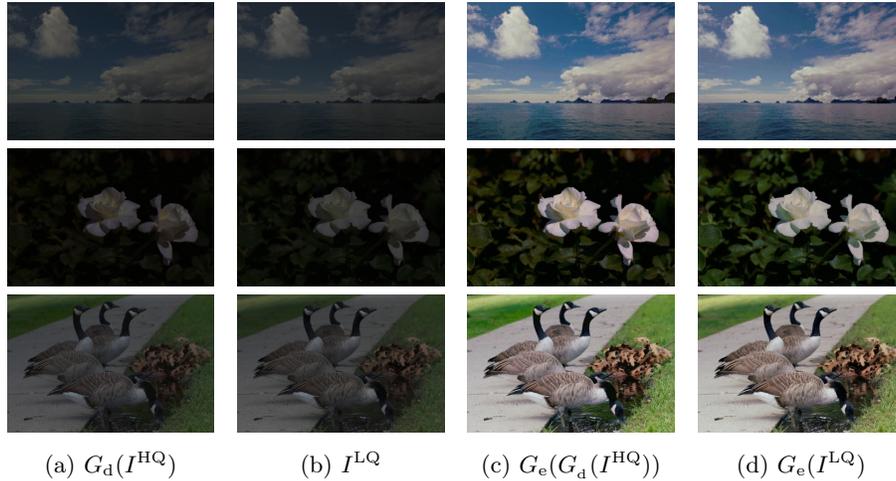


Fig. 9: Examples of (a) degraded images, (b) real low-quality images, (c) global enhancement results from degraded images, and (d) global enhancement results from low-quality images.

5 Conclusions

In this paper, we proposed a novel algorithm to achieve both paired and unpaired image enhancement. The proposed GEN performs the channel-wise intensity transformations and LEN improves the global enhanced images from GEN. For training GEN and LEN, we developed paired learning and unpaired learning methods. For unpaired learning, we proposed the two-stage training scheme based on GANs to exploit the strengths of GEN that can be trained easily. Experimental results demonstrated that the proposed algorithm outperforms the state-of-the-art algorithms on the MIT-Adobe 5K dataset. Remarkably, GEN and LEN, which are trained by the proposed unpaired learning, outperforms the conventional paired image enhancement algorithms.

Acknowledgements

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00464) supervised by the IITP (Institute for Information & communications Technology Promotion), in part by the National Research Foundation of Korea (NRF) through the Korea Government (MSIP) under Grant NRF-2018R1A2B3003896, and in part by the research fund of Chungnam National University.

References

1. Arici, T., Dikbas, S., Altunbasak, Y.: A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. Image Process.* **18**(9), 1921–1935 (2009) [3](#)
2. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: *CVPR* (2011) [2](#), [3](#), [9](#)
3. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: *ICCV* (2017) [3](#)
4. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. In: *CVPR* (2018) [2](#), [3](#), [4](#), [5](#), [10](#), [11](#), [12](#)
5. Deng, Y., Loy, C.C., Tang, X.: Aesthetic-driven image enhancement by adversarial learning. In: *ACM MM* (2018) [2](#), [3](#)
6. Fu, X., Liao, Y., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Trans. Image Process.* **24**(12), 4965–4977 (2015) [3](#)
7. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: *CVPR* (2016) [3](#)
8. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. *ACM Trans. Graphics* **36**(4), 118 (2017) [2](#), [3](#), [10](#)
9. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing* (4th Edition). Pearson (2018) [3](#)
10. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: *NeurIPS*. pp. 5767–5777 (2017) [7](#)
11. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2016) [3](#)
12. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. *ACM Trans. Graphics* **37**(2), 26 (2018) [2](#), [3](#)
13. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **6**(7), 965–976 (1997) [3](#)
14. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **6**(3), 451–462 (1997) [3](#)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *ECCV* (2016) [7](#)
16. Kim, J.H., Jang, W.D., Sim, J.Y., Kim, C.S.: Optimized contrast enhancement for real-time image and video dehazing. *J. Vis. Commun. Image. Represent.* **24**, 410–425 (Apr 2013) [3](#)
17. Kim, Y.T.: Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Trans. Consum. Electro.* **43**(1), 1–8 (1997) [3](#)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *ICLR* (2014) [9](#), [11](#)
19. Kosugi, S., Yamasaki, T.: Unpaired image enhancement featuring reinforcement-learning-controlled image editing software. In: *AAAI* (2020) [2](#), [3](#), [11](#), [12](#), [13](#)
20. Land, E.H.: The retinex theory of color vision. *Scientific American* **237**(6), 108–129 (1977) [3](#)

21. Lee, C., Kim, J.H., Lee, C., Kim, C.S.: Optimized brightness compensation and contrast enhancement for transmissive liquid crystal displays. *IEEE Trans. Circuits Syst. Video Technol.* **24**, 576–590 (Apr 2014) [3](#)
22. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans. Image Process.* **22**(12), 5372–5384 (2013) [3](#)
23. Lee, C., Lee, C., Lee, Y.Y., Kim, C.S.: Power-constrained contrast enhancement for emissive displays based on histogram equalization. *IEEE Trans. Image Process.* **21**(1), 80–93 (2011) [3](#)
24. Lim, J., Heo, M., Lee, C., Kim, C.S.: Contrast enhancement of noisy low-light images based on structure-texture-noise decomposition. *J. Vis. Commun. Image Represent.* **45**, 107–121 (May 2017) [3](#)
25. Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: a deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **61**, 650–662 (2017) [2, 3](#)
26. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: *CVPR* (2018) [2, 3, 11, 12](#)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI* (2015) [2](#)
28. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *ICLR* (2015) [7](#)
29. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **9**(5), 889–896 (2000) [3](#)
30. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: *CVPR* (2019) [2, 3, 10, 11, 12](#)
31. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* **22**(9), 3538–3548 (2013) [3](#)
32. Wang, Y., Chen, Q., Zhang, B.: Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Trans. Consum. Electro.* **45**(1), 68–75 (1999) [3](#)
33. Yan, Z., Zhang, H., Wang, B., Paris, S., Yu, Y.: Automatic photo adjustment using deep neural networks. *ACM Trans. Graphics* **35**(2), 11 (2016) [2, 3](#)
34. Yu, R., Liu, W., Zhang, Y., Qu, Z., Zhao, D., Zhang, B.: Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In: *NeurIPS* (2018) [2, 3](#)
35. Yue, H., Yang, J., Sun, X., Wu, F., Hou, C.: Contrast enhancement based on intrinsic image decomposition. *IEEE Trans. Image Process.* **26**(8), 3981–3994 (2017) [3](#)
36. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *ICCV* (2017) [8, 13](#)