

# Global and Local Enhancement Networks For Paired and Unpaired Image Enhancement -Supplement Material-

Han-Ul Kim<sup>1</sup>[0000-0001-7450-6600], Young Jun Koh<sup>2</sup>[0000-0003-1805-2960], and Chang-Su Kim<sup>1</sup>[0000-0002-4276-1831]

<sup>1</sup> Korea University, Seoul, Korea  
hanulkim@mcl.korea.ac.kr, changsukim@korea.ac.kr  
<sup>2</sup> Chungnam National University, Daejeon, Korea  
yjkoh@cnu.ac.kr

## S-1 More Experiments on Paired Learning

### S-1.1 Comparison of the proposed GEN with the baseline network

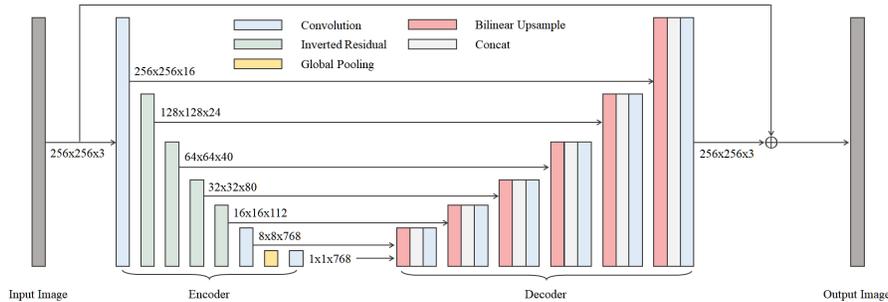


Fig.S-1: Architecture of the baseline network.

Fig. S-1 shows the architecture of the baseline network structure in Section 4.1 in the main paper, which has the encoder-decoder structure to obtain the pixel-wise color prediction. The encoder in Fig. S-1 is same to the architecture of GEN in Table 1. Table S-2 specifies the decoder, including six upsample blocks, each of which performs bilinear upsampling, concatenation, and convolution filtering, subsequently. All “Conv3x3” operations except the last one include convolution filters, batch normalization, and swish activation. The last “Conv3x3” operation only includes convolution filters. To this end, the baseline network produces an enhanced image by adding a residual image to an input image.

Table S-2 provides the number of parameters of GEN and the baseline network. The baseline network uses a large number of parameters, since it requires

Table S-1: Specification for the decoder architecture of the baseline network.

Stage	Operator	Output Resolution	Channels
1	Bilinear & Concat & Conv3x3	$8 \times 8$	112
2	Bilinear & Concat & Conv3x3	$16 \times 16$	80
3	Bilinear & Concat & Conv3x3	$32 \times 32$	40
4	Bilinear & Concat & Conv3x3	$64 \times 64$	24
5	Bilinear & Concat & Conv3x3	$128 \times 128$	16
6	Bilinear & Concat & Conv3x3	$256 \times 256$	3

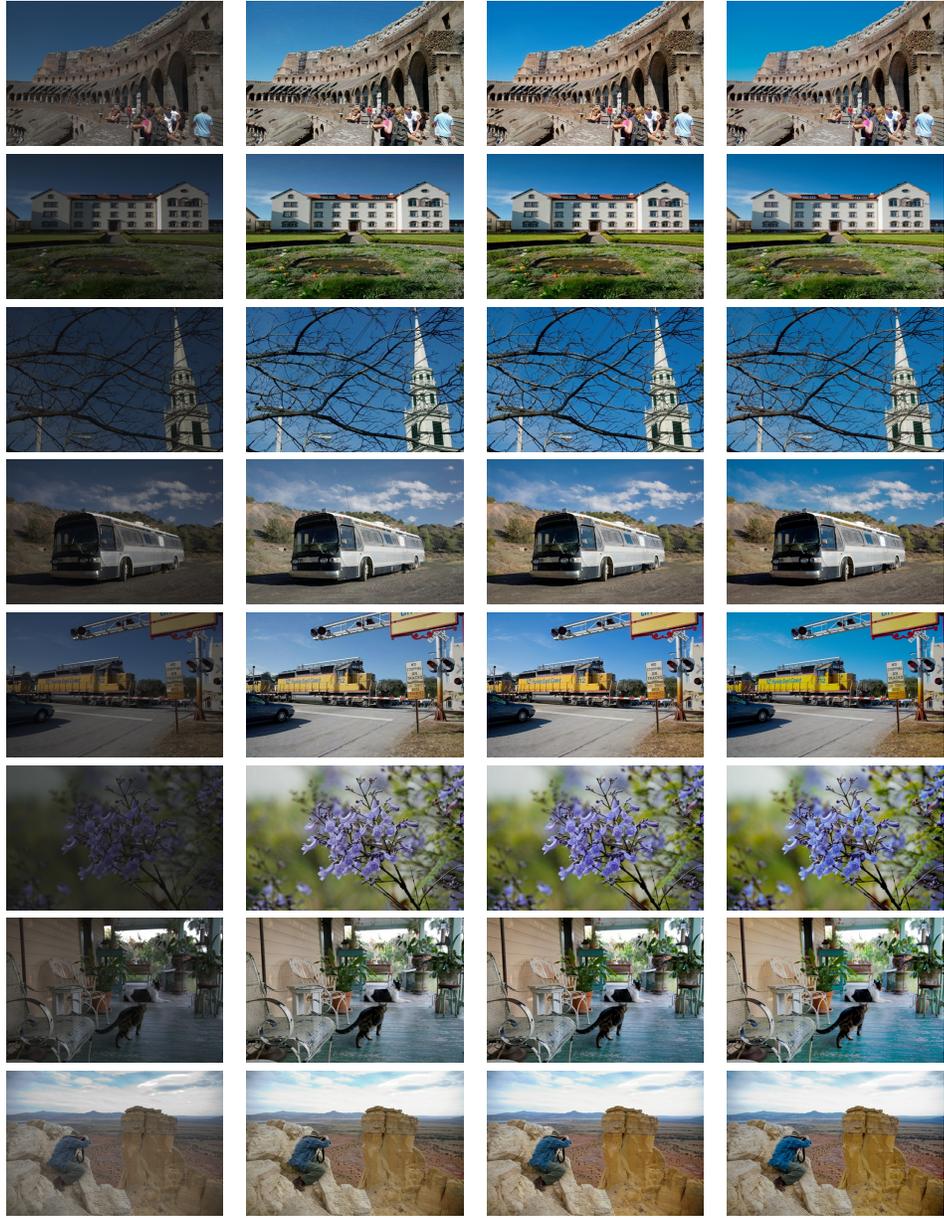
Table S-2: Comparison with the baseline network.

Method	No. Parameters	256px×256px		Long-512px	
		PSNR	SSIM	PSNR	SSIM
Baseline	1,366,144	25.08	0.886	24.52	0.884
GEN	868,240	25.53	0.919	25.47	0.917

the decoder part. Also, Table S-2 lists the PSNR and SSIM scores according to the spatial size of test images. Specifically, “256px×256px” denotes that the spatial size of test images are  $256 \times 256$ , while “Long-512px” denotes that long side of each test image is fixed to 512. We see that both networks yield lower scores on “Long-512px” than “256px×256px,” since both networks are trained using  $256 \times 256$  images. Nevertheless, GEN experiences only 0.06dB degradation in terms of PSNR, which is significantly lower than 0.56dB degradation in the baseline network. This indicates that the proposed GEN is robust to the image size by adopting the channel-wise intensity transformation.

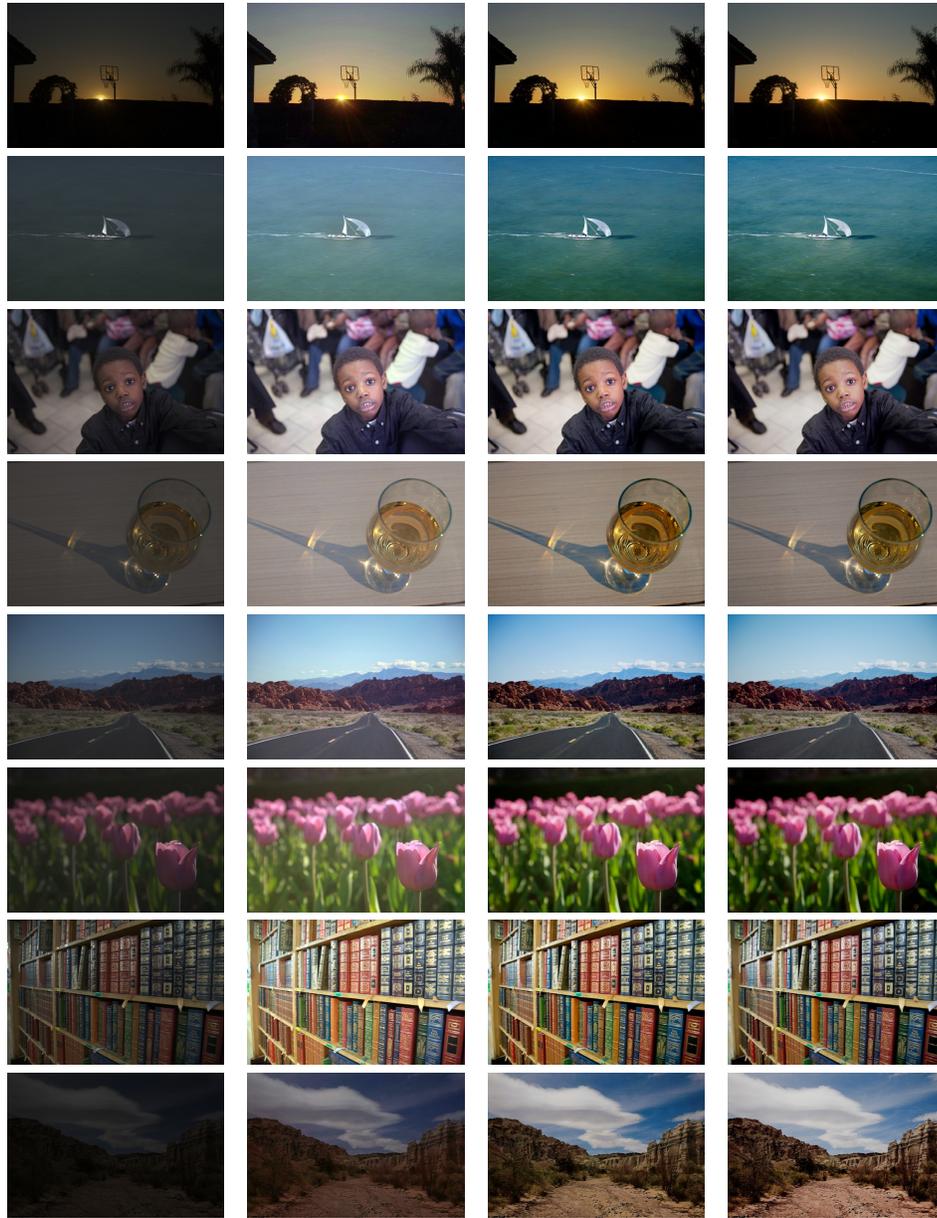
### S-1.2 More qualitative results

Fig. S-2 provides qualitative enhancement results of GEN and GEN & LEN on the MIT Adobe 5K dataset. All examples show that the proposed LEN improves GEN results by alleviating the one-to-many mapping problem. In Fig. S-3, we compare qualitative enhancement results of the proposed method with those of DUPE [27], which is the state-of-the-art as in Table 3 in the main paper. We see that the proposed method successfully reproduces color tones of images, which are similar to photographer C’s retouched images, while DUPE fails to yield faithful results.



(a) Input (b) GEN (c) GEN & LEN (d) Photographer C

Fig. S-2: Qualitative comparison between GEN and GEN & LEN methods.



(a) Input (b) DUPE [27] (c) GEN & LEN (d) Photographer C

Fig. S-3: Qualitative comparison of the proposed algorithm with DUPE [27].

## S-2 More Experiments on Unpaired Learning

### S-2.1 Analysis on the proposed unpaired learning scheme

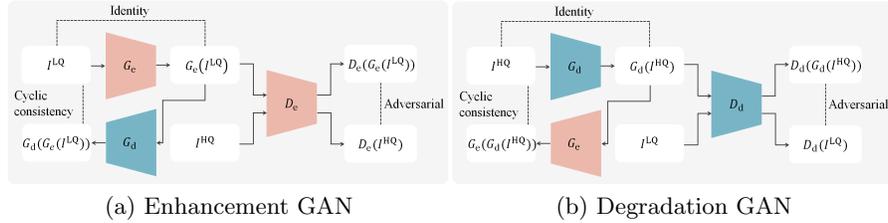


Fig. S-4: The network architectures of (a) enhancement GAN and (b) degradation GAN for the “CWGAN-GP” training scheme in Table 5 in the main paper.

Fig. S-4 shows the network architectures for the CWGAN-GP training scheme in [33]. The “CWGAN-GP” scheme employs the cyclic consistency loss to enforce that enhanced images should be reverted to input images through the degradation generator. In contrast, the proposed training scheme in Fig. 4 utilizes the cyclic color loss, which directly constrains the degradation generator to produce low-quality images from real high-quality images.

Table S-3: Performance of degraded images according to training schemes.

Training	PSNR	SSIM
CWGAN-GP	26.93	0.906
Proposed	29.10	0.920

In Table S-3, we compare the performance of two degradation generators based on the proposed and “CWGAN-GP” schemes. For this purpose, we measure the PSNR and SSIM scores between degraded images  $G_d(I^{HQ})$  and real low-quality images  $I^{LQ}$ . In Table S-3, we observe that the proposed cyclic loss yields more accurate low-quality images than the cyclic consistency loss in [33] does. To this end, the proposed degradation generator leads to notable performance improvement of the proposed enhancement generator as in Table 5.

### S-2.2 More qualitative results

Fig. S-5 provides more qualitative result of degraded images, real low-quality images, global enhancement results from degraded images, and global enhancement results from low-quality images. We observe that the degraded images are well

imitated with real images. Also, it is shown that global enhanced images from the degraded images and those from low-quality images are similar to each other. Fig. S-6 qualitatively compares the proposed method with FRL [18], which is the state-of-the-art conventional method in Table 4 in the main paper. In these examples, the proposed method provides more similar images to retouched images by Photographer C, compared with FRL does.

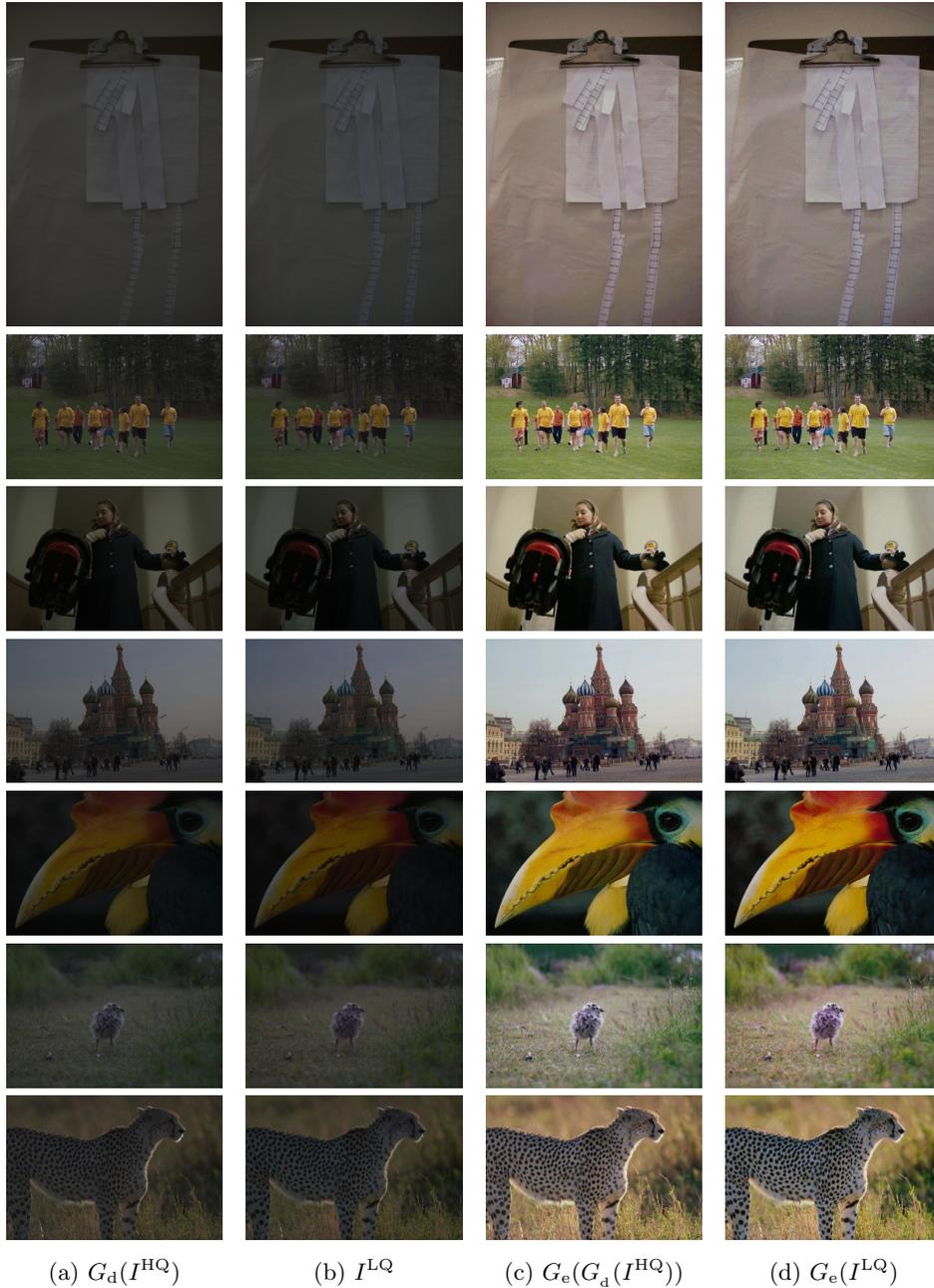
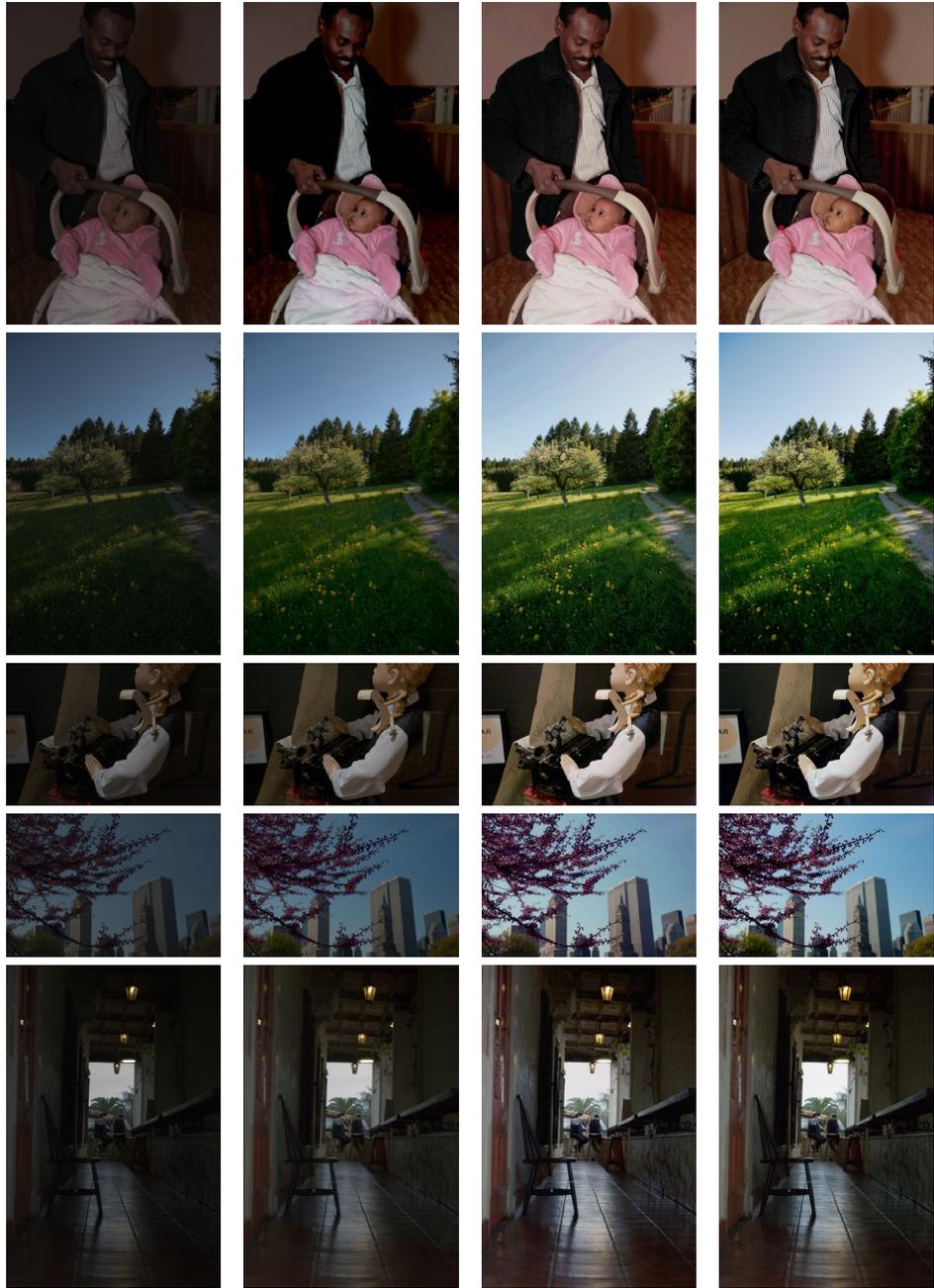


Fig. S-5: Examples of (a) degraded images, (b) real low-quality images, (c) global enhancement results from degraded images, and (d) global enhancement results from low-quality images.



(a) Input

(b) FRL [18]

(c) GEN & LEN

(d) Photographer C

Fig. S-6: Qualitative comparison of the proposed algorithm with FRL [18].