Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Naturalness-Preserving Image Tone Enhancement Using Generative Adversarial Networks

Hyeongseok Son, Gunhee Lee, Sunghyun Cho, and Seungyong Lee

POSTECH

(a) Input (b) Son *et al.* [SLKL14] (d) Ours

Figure 1: *Image tone enhancement. Our method is trained to mimic the behavior of Son et al.'s method [SLKL14] while preserving the naturalness for tone enhancement. While Son et al.'s method may produce unnatural-looking results because of their goal to maximize tone and detail, our method produces natural-looking tone enhancement results with appropriately boosted local contrasts and details.*

Abstract

This paper proposes a deep learning-based image tone enhancement approach that can maximally enhance the tone of an image while preserving the naturalness. Our approach does not require carefully generated ground-truth images by human experts for training. Instead, we train a deep neural network to mimic the behavior of a previous classical filtering method that produces drastic but possibly unnatural-looking tone enhancement results. To preserve the naturalness, we adopt the generative adversarial network (GAN) framework as a regularizer for the naturalness. To suppress artifacts caused by the generative nature of the GAN framework, we also propose an imbalanced cycle-consistency loss. Experimental results show that our approach can effectively enhance the tone and contrast of an image while preserving the naturalness compared to previous state-of-the-art approaches.

CCS Concepts

• ***Computing methodologies*** → ***Image processing;***

1. Introduction

Photographic tone is one of the most important components that make pictures aesthetically appealing. For tone adjustment, many image filtering techniques have been developed such as classical contrast enhancement methods using intensity histograms [GW06], two-scale tone management [BPD06], local Laplacian filters [PHK11], and art-photographic detail enhancement [SLKL14]. However, most of these methods require the involvement of a user to adjust the parameters to maximize the enhancement effects while preserving the naturalness, as simply maximizing filtering effects often leads to unnaturally-looking results as shown in Fig. 1(b).

Recently, many deep learning-based approaches have been introduced to automatically enhance images without manual intervention [YZW*15, IKT*17, CWKC18, IKT*18, HHX*18, PLYK18]. These methods can produce high-quality results without manual parameter controls. However, they require carefully designed ground-truth labels, i.e., ideally adjusted results for input images, which have to be manually prepared by human experts, as done for the MIT-Adobe FiveK dataset [BPCD11]. Otherwise, the networks may learn to produce unnatural-looking results.

In this paper, we propose a deep learning-based approach to photographic tone enhancement (Fig. 1(c)). Our approach trains a deep neural network (DNN) to mimic the behavior of previous classi-

cal filtering while preserving the naturalness. Our approach does not require any manually generated ground-truth labels, but directly learns from pairs of an input image and its automatically-generated filtering result that could look unnatural due to inappropriate parameters. To preserve the naturalness while mimicking the effect of a classical filter, we adopt the generative adversarial network (GAN) framework as a regularizer for the naturalness.

Simply adopting GAN to preserve the naturalness, however, may produce artifacts in the results, such as synthetically generated structures that are not present in the input images due to the generative nature of GAN. Besides, there may be remaining artifacts that occur in mimicking the classical filtering results. To resolve this limitation, we propose an *imbalanced cycle-consistency loss* to suppress the generator from producing artifacts. Experimental results show that our approach can effectively enhance the tone and contrast of an image while preserving the naturalness compared to previous state-of-the-art approaches.

Our contributions can be summarized as follows:

- We propose a novel deep learning-based approach that can automatically adjust tone and contrast while preserving the naturalness.
- We propose a new training strategy that can train a deep neural network without ground-truth images that have been manually adjusted by experts.
- We propose a GAN based regularization approach to preserving the naturalness.
- We propose an imbalanced cycle-consistency loss to effectively remove artifacts remaining after the regularization.

2. Related Work

Manipulating the tone and contrast of an image has been widely investigated with different goals such as tone mapping, detail enhancement, inverse tone mapping, tone transfer, and artistic tone and contrast enhancement. Durand *et al.* [DD02] propose a fast bilateral filter, and use it for tone mapping. They use the fast bilateral filter to decompose a high-dynamic-range (HDR) image into base and detail layers, which contain large-scale structures and small-scale details, respectively. Only the base layer has its contrast reduced, and then recomposition with the detail layer produces a low-dynamic-range (LDR) image while preserving details. Bae *et al.* [BPD06] propose a tone transfer method that transfers the tone of a reference image to a target image. Fattal *et al.* [FAR07] use multi-scale bilateral decomposition to enhance surface details. Paris *et al.* [PHK11] present local Laplacian filters to decompose an image into multiple scale details, and transform the details for edge-preserving smoothing, detail enhancement, tone mapping, and inverse tone mapping. Aubry *et al.* [APH*14] speed up the local Laplacian filters by simplifying pixel neighborhood calculation. Son *et al.* [SLKL14] propose a detail enhancement method inspired by art photography. Their method extremely exaggerates fine-scale details of an image to make a hyper-realistic HDR-looking image.

Recently, deep learning-based approaches have also been introduced. Bychkovsky *et al.* [BPCD11] propose a global color tone adjustment method by learning photographers' retouching styles. Gharbi *et al.* [GCB*17] introduce a DNN architecture to predict local affine transforms in a bilateral grid [CPD07], which is a 3D array that combines the two-dimensional spatial domain with a one-dimensional range dimension, for real-time color adjustment. Ignatov *et al.* [IKT*17, IKT*18] propose learning a translation function from photos taken by smartphone cameras to DSLR-quality images. Talebi and Milanfar [TM18] propose a deep learning based image assessment approach and use it for automatic parameter selection of existing enhancement methods. Chen *et al.* [CWKC18] use the CycleGAN [ZPIE17] to learn color enhancement from unpaired data. Recently, reinforcement learning based approaches [HHX*18, PLYK18] have also been proposed to enhance photos step-by-step using simple adjustment operations.

While all these methods show excellent results, they require either careful parameter tuning or carefully generated training data, such as the MIT-Adobe FiveK dataset [BPCD11] and the DPED dataset [IKT*17], to retain the naturalness in the results. CycleGAN-based methods such as Chen *et al.*'s [CWKC18] do not need manually generated training data as they can directly learn from unpaired data. However, learning from unpaired data is difficult as unpaired data provide only indirect guidance on image enhancement. On the other hand, in our framework, we have specific target images generated by a classical image filtering method, so our network can be trained more effectively and produce better image enhancement results as will be shown in Sec. 4.

3. Photographic Tone Enhancement Framework

In this work, we aim to develop a photographic tone enhancement method to maximize tone and detail while retaining the naturalness. We also aim to avoid manual parameter tuning and carefully generated training data. To achieve these goals, we design our framework as shown in Fig. 2. Our framework consists of one tone enhancement network and three different losses for training the network. The tone enhancement network takes a photograph as input and produces its enhanced version. The color similarity loss trains the enhancement network to mimic the behavior of a classical image filter with automatically generated training data. The naturalness loss trains the network to retain the naturalness based on the GAN framework. Finally, the artifact suppression loss trains the network to remove artifacts in enhancement results based on an imbalanced cycle-consistency. In the following, we describe each part of the framework in detail. Detailed architectures of our networks can be found in the supplementary material.

3.1. Enhancement Network

The enhancement network G adopts an encoder-decoder structure, which consists of an encoder, residual blocks, and a decoder as shown in Fig. 2. It is well-known that a small CNN structure [XRY*15] effectively learns classical image filtering such as bilateral filtering [TM98] and L_0 smoothing [XLXJ11]. However, our problem requires region-specific tone adjustment, which needs large receptive fields. To effectively enlarge the receptive fields, we use encoder-decoder network with symmetric skip-connections and additionally place 16 residual blocks with two convolution layers between the encoder and decoder. The upsampling in the decoder is implemented by a resize-convolution layer using nearest neighbor

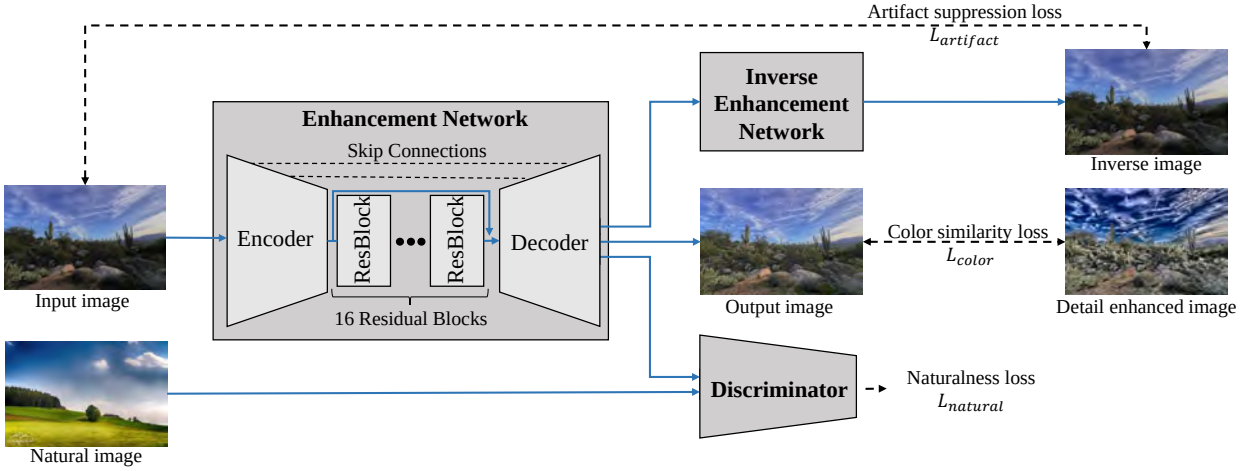


Figure 2: *Our framework consists of an enhancement network and three loss functions for training the network.*

(a) Input (b) Results of [SLKL14]

Figure 3: *Examples of Son et al.'s method [SLKL14]. The method maximizes tone and details of an input image and the result may look unnatural.*

upsampling. In our preliminary test, we used deconvolution layers instead of resize-convolution layers, but it caused severe blocky artifacts in our task. We use batch normalization after each convolution layer except the final layer. In our preliminary test, we tested batch normalization, instance normalization, and no normalization, and found that batch normalization provided the most stable training. We use a long skip-connection to add the input image to the output of the network for residual learning.

3.2. Color Similarity Loss

The color similarity loss enforces the enhancement network G to learn the behavior of classical image filtering. Specifically, in our work, we adopt an enhancement filter of Son *et al.* [SLKL14], which increases local contrast as much as possible. For generating training data, we sampled 2,000 input images before retouching of human experts in the MIT-Adobe FiveK dataset [BPCD11]. We applied Son *et al.*'s method to the sampled images with the default parameters in a fully automatic way, and obtained their filtered results, which we call *guide images*. The generated guide images have more contrast and details, but may look unnatural when the contrast and details are overly enhanced as shown in Fig. 3.

During the training phase, we feed the sampled input images to the enhancement network G, and update the network parameters to minimize the loss between the network outputs and the guided images. The color similarity loss L_{color} is defined as:

$$L_{color} = \mathbb{E}_{x\sim p_{data}(x)}[MSE(G(x), I_D)], \quad (1)$$

where $\mathbb{E}_x$ is an expectation over the distribution $p_{data}(x)$ of x, MSE is the mean squared error, and I_D is a guide image corresponding to an input image x. By minimizing L_{color}, the network G is trained to mimic the behavior of Son *et al.*'s method, which produces tone-enhanced but possibly unnatural-looking images.

3.3. Naturalness Loss

To retain the naturalness, we adopt an adversarial loss based on the GAN framework. Specifically, we define a discriminator network D that discriminates synthetically enhanced images against natural images. Using the discriminator, we define the naturalness loss $L_{natural}$ based on least-square GAN [MLX*17]:

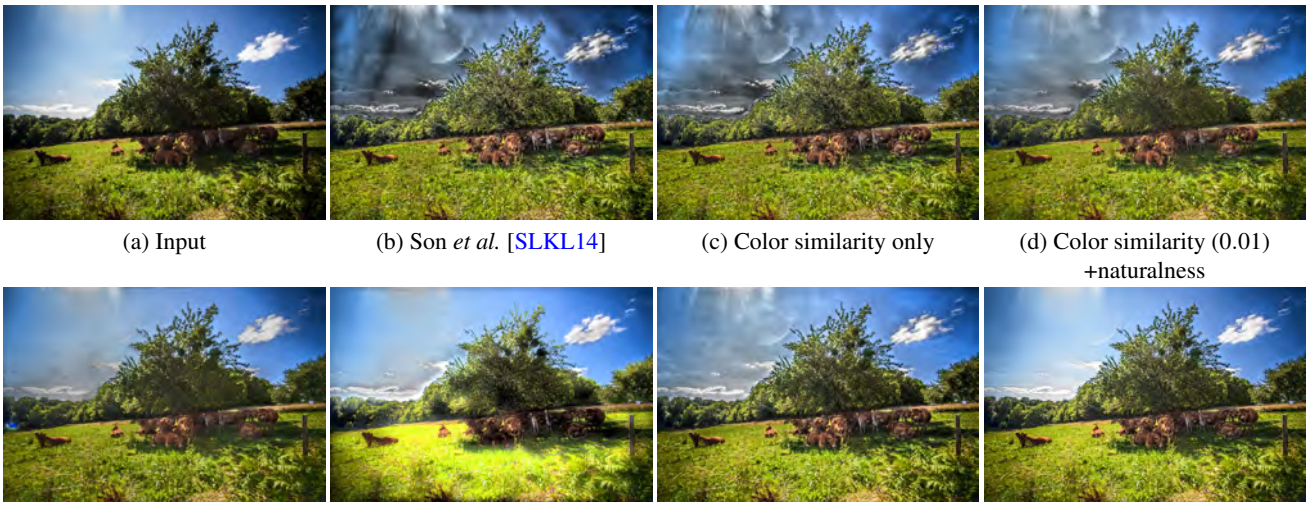
$$L_{natural} = \mathbb{E}_{x\sim p_{data}(x)}[(D(G(x)) - 1)^2]. \quad (2)$$

By minimizing L_{color} and $L_{natural}$ together, the enhancement network is trained to produce images that are similar to enhanced results of Son *et al.*'s method and natural-looking.

To train the discriminator, we define the discriminator loss L_{dis} as:

$$L_{dis} = \mathbb{E}_{y\sim p_{data}(y)}[(D(y) - 1)^2] + \mathbb{E}_{x\sim p_{data}(x)}[(D(G(x))^2], \quad (3)$$

where $p_{data}(y)$ is the distribution of natural-looking images, and y is a sample from the distribution. For the natural-looking images, we collected 3,200 high-resolution images larger than 1000×1000 that are labeled as 'HDR' from Flicker. During the training phase,



(a) Input (b) Son *et al.* [SLKL14] (c) Color similarity only (d) Color similarity (0.01) +naturalness

(e) Color similarity (0.001) + naturalness (f) Color similarity (0.0001) + naturalness (g) Color similarity + naturalness + artifact (big) (h) Color similarity + naturalness + artifact (small)

Figure 4: *Ablation study. v in 'Color similarity (v)' is the weight for the color similarity loss function. The default value for the weight used in (c), (g), and (h) is* 0.02*. Please zoom for better comparison.*

we train the enhancement network G and the discriminator D alternatingly as done in the conventional training process of the GAN framework [LTH*17, PSC*18]. The choice of the GAN method is flexible in our framework. A more recent GAN method such as Wasserstein GAN [GAA*17] can be readily applied to improve the overall quality.

3.4. Artifact Suppression Loss

As the loss $L_{natural}$ follows the conventional GAN formulation, minimizing it can cause artifacts in the results of the enhancement network G, such as synthetically generated image structures and details. While we may avoid such artifacts by giving more weight to L_{color}, this will guide the enhancement network G to simply reproduce Son *et al.*'s results without preserving the naturalness. To resolve this problem, we introduce an artifact suppression loss that measures only the structural difference between an input image and its enhanced result, disregarding tone enhancement effects. The unnatural artifacts in guide images affect image structures more than the tone enhancement effects so that minimizing this loss can help to remove artifacts while preserving tone enhancement effects.

Our artifact suppression loss is based on an *imbalanced cycle-consistency*, which is inspired by CycleGAN [ZPIE17]. We first introduce an inverse enhancement network C, which is similar to an inverse network in CycleGAN. The inverse enhancement network C takes an enhanced result from the enhancement network G as input, and predicts the original input image before the enhancement. For the inverse enhancement network, we use a small network architecture that consists of five convolution layers with a single skip-connection, so it cannot remove structural artifacts in the enhanced images. This is a sharp contrast to the CycleGAN where the inverse network tries to fully learn the inverse operation. If the inverse enhancement network C can fully revert the enhancement operation done by the enhancement network G, then it will also be able to revert any artifacts caused by the enhancement network G. As a result, the artifact suppression loss will not be able to remove any artifacts.

Using the inverse enhancement network C, we define the artifact suppression loss $L_{artifact}$ as:

$$L_{artifact} = \mathbb{E}_{x\sim p_{data}(x)}[MSE(C(G(x)),x)]. \quad (4)$$

By minimizing $L_{artifact}$, G is constrained to preserve structural details of the input image x in the tone enhancement result. We use MSE to measure the distance between an input image x and the inverse enhancement result $C(G(x))$, but other distance functions such as SSIM [ZBSS04] can be used instead of MSE. We refer the reader to our supplementary material for an additional experiment using SSIM.

3.5. Training

We trained our framework in three steps. First, we pre-trained the enhancement network G using L_{color} to learn tone enhancement. In the second step, we pre-trained our inverse enhancement network C with $L_{artifact}$ while fixing G. For pre-training G and C, we used a learning rate 10^{-4} and a batch size 3. We trained G and C for 70,000 and 35,000 iterations, respectively. We did not use pre-training for D as training a discriminator is usually easier than training a generator in the GAN framework as shown in [LTH*17, PSC*18]. Finally in the third step, we used all the three loss functions while alternatingly updating the networks G, C, and D. In this fine-tuning step, we used smaller learning rates to prevent a sudden change of pre-trained weights and to fine-control the weights. We used a learning rate 10^{-6} for D, and 10^{-7} for G and C with 35,000 iterations. The loss functions L_{color}, $L_{natural}$, and $L_{artifact}$ were weighted by 0.02, 0.003, and 0.1, respectively. We resized all input images to 530×530, and randomly cropped 512×512 regions were used for training.

4. Experiments

In this section, we first analyze the effect of each loss in our framework, and evaluate the performance of our method compared to previous state-of-the-art image enhancement methods. We also present our user study results.

4.1. Ablation Study

For analyzing the effect of each loss, we performed an ablation study. As the quality of image enhancement cannot be quantitatively measured, we qualitatively compare the results of different combinations of the loss functions in Fig. 4. The model using only the color similarity loss is simply trained to reproduce Son *et al.*'s method [SLKL14], so the result in Fig. 4(c) looks similar to Son *et al.*'s result in Fig. 4(b). This result also proves that our enhancement network can effectively reproduce Son *et al.*'s method. Figs. 4(d-f) are results of combining the color similarity and naturalness loss functions with different weights of L_{color}. Giving more weight to the naturalness loss (or equivalently giving less weight to the color similarity loss) results in more natural-looking results. However, we can see there are still remaining artifacts in those results.

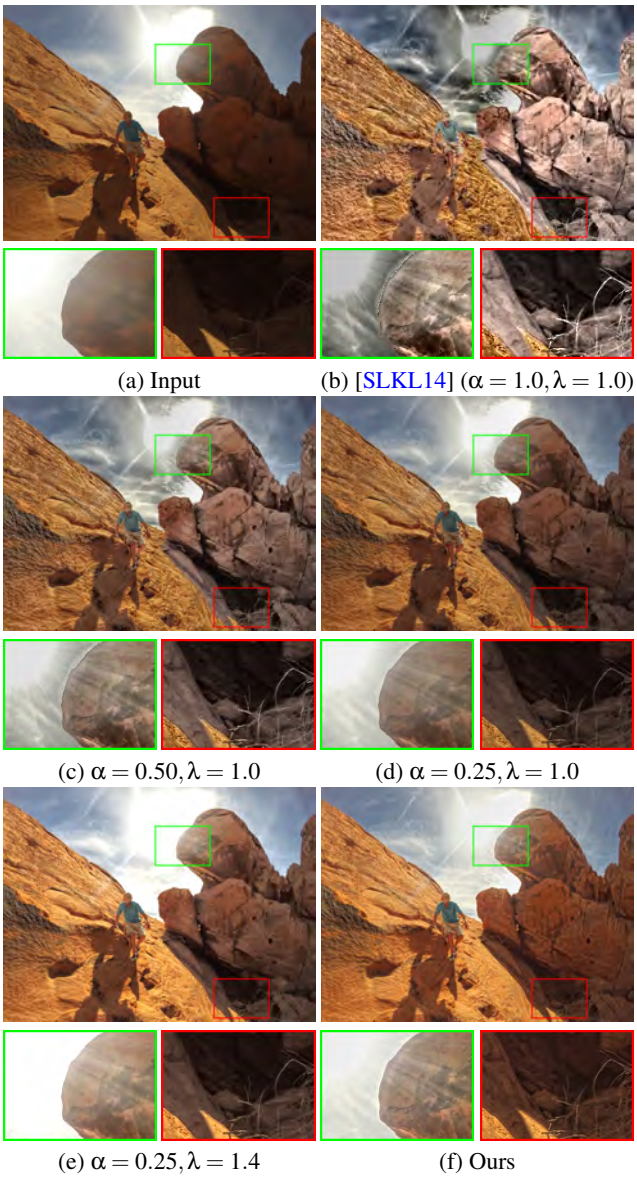
Figs. 4(g) and 4(h) show results of using all the loss functions together. In Fig. 4(g), we use a larger network structure for the inverse enhancement network C, which is the same structure as our enhancement network G. A large inverse enhancement network is able to learn even removing artifacts caused by the color similarity loss. Consequently, the result in Fig. 4(g) still has artifacts caused by the color similarity loss. For our final result in Fig. 4(h), we use a small network described in Sec. 3.4 for the inverse enhancement network. In that case, the inverse enhancement network can learn the inverse of the enhancement operation only, excluding the artifacts in the enhanced images that would need a larger capacity to handle. As a result, our enhancement network G can be trained to enhance the tone of an input image while avoiding artifacts in guide images, which could be reproduced by the color similarity loss.

4.2. Comparisons with Baseline Methods

One simple baseline method to control the trade-off between the naturalness and the tone enhancement effect would be a linear blending of an input image and its unnatural-looking tone enhancement result. Specifically, we can define a baseline method as:

$$I = \alpha I_{enhance} + \lambda(1-\alpha)I_{input}, \tag{5}$$

where I is a linear blending result, $I_{enhance}$ is the result of Son *et al.*'s method [SLKL14], I_{input} is the input image, and α is the blending weight. λ is a scale factor for the input image, which is included because we can also enhance the contrast of an input image while preserving the naturalness by setting λ larger than 1. Fig. 5 shows a comparison between our result and linear blending results with different parameters. As shown in the figure, linear blending cannot adaptively suppress artifacts in different image regions as it is uniformly and globally applied to images regardless of tone and details. For example, Fig. 5(c) with $\alpha = 0.50$ still contains artifacts from $I_{enhance}$, while Fig. 5(d) with $\alpha = 0.25$ shows less contrasts of image details than our result in Fig. 5(f), especially in dark regions. Using λ larger than 1 may enhance contrast in dark regions as shown in Fig. 5(e), but it causes saturated pixels in bright regions. On the contrary, our result in Fig. 5(f) shows enhanced tone and details in both bright and dark regions.

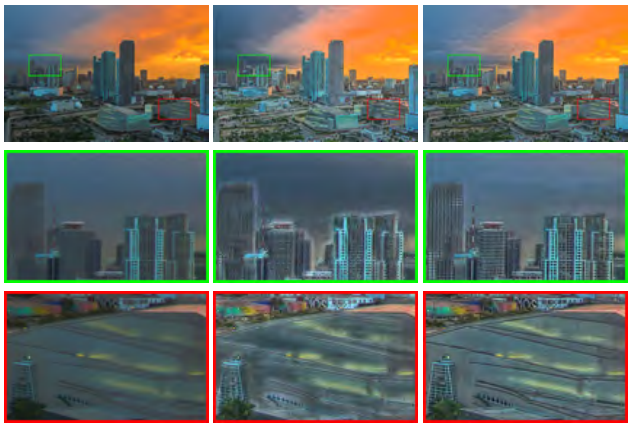

(a) Input (b) [SLKL14] ($\alpha = 1.0, \lambda = 1.0$)

(c) $\alpha = 0.50, \lambda = 1.0$ (d) $\alpha = 0.25, \lambda = 1.0$

(e) $\alpha = 0.25, \lambda = 1.4$ (f) Ours

Figure 5: *Comparison with linear combinations of an input image and the tone enhanced result by Son et al. [SLKL14].*

To further suppress artifacts, our system adopts the imbalanced cycle-consistency loss with a small-capacity inverse enhancement network, which is not capable of reverting structural artifacts. One question that naturally follows is *why don't we directly use a small-capacity network for the image enhancement network G?* In other words, we may adopt a small-capacity network for image enhancement instead of the imbalanced cycle-consistency loss in order to prevent artifacts. However, we found that such a small-capacity network cannot sufficiently mimic the classical tone enhancement method. As shown in Fig. 6(b), adopting a small-capacity network



(a) Input (b) Small network (c) Ours

Figure 6: *Comparison with directly using a small-capability enhancement network.*

cannot consistently adjust tones in large objects (red box in Fig. 6(b)) and produces artifacts such as halo (green box in Fig. 6(b)).

4.3. Qualitative Comparison

For qualitative evaluation, we compare our method with four state-of-the-art image enhancement methods, which include two classical image enhancement methods of Aubry *et al.* [APH*14] and Son *et al.* [SLKL14], and two deep learning-based image enhancement methods of Ignatov *et al.* [IKT*17] and Chen *et al.* [CWKC18]. Ignatov *et al.*'s method learns a mapping from iPhone3 camera images to DSLR-quality images. Chen *et al.*'s method adopts the CycleGAN framework and learns a mapping from an image to its enhanced version from unpaired data. For the comparison, we used the models trained by the authors for the deep learning-based methods, and used test images of [SLKL14] and other images collected from Flickr.

Fig. 7 shows a comparison. Both results of Aubry *et al.* and Son *et al.* are unnatural due to their exaggerated contrast and details. In the result of Ignatov *et al.*, the colors are brightened, and the details in the sky are lost. This is because Ignatov *et al.*'s method is specifically designed to handle images from iPhone3 camera images, which are usually darker than images captured by DSLR cameras. Chen *et al.*'s result shows more vivid color than the input image. However, it has less details and less local contrasts, so it looks rather foggy. On the other hand, our result shows properly boosted contrasts and details than the results of Ignatov *et al.*'s and Chen *et al.*'s. Moreover, our result looks more natural than the results of Aubry *et al.* and Son *et al.* thanks to our naturalness loss and artifact suppression loss. Another comparison is given in Fig. 8, which shows that our method works well for a challenging scene containing bright and dark regions together. It also shows that our method produces a sharp result with well-enhanced details for a high resolution image. Fig. 9 shows additional comparison results.

4.4. User Study

Finally, we present our user study result using Amazon Mechanical Turk. We collected 20 images from Flickr and used them for all 30 participants. For each image, we showed a participant five different enhancement results one by one, which have been obtained by our method and the four other methods. Then, we asked the participant two questions for each image: 1) "Does this image look natural? (score 1(no) - 5(yes))" and 2) "In terms of tone and detail, rate the quality of this image (score 1(bad) - 5(good))". The first question is to validate that our method can preserve the naturalness, and the second question is to evaluate the quality of our tone enhancement. 'Tone' and 'detail' in the second question are ambiguous terms that may not be evaluated objectively. Nevertheless, tone and detail were the most noticeable differences among the images compared in the user study, as we showed the participants the results of the *same* input image produced by different enhancement methods. Thus we expect that participants rated the images mostly in terms of tone and detail.

Table 1 reports the user study result. Our method achieved the second place in terms of naturalness, and the first place in terms of tone and detail enhancement. The methods of Ignatov *et al.* [IKT*17] and Chen *et al.* [CWKC18] marginally enhance details, so their results obtained high naturalness scores. While our method enhance details rather aggressively, it shows comparable performance with these methods in term of naturalness. This result implies that our naturalness loss and artifact suppression loss properly work in the way they are designed for. Fig. 10 shows some images used in our user study. We refer the readers to our supplementary material for more examples.

5. Conclusion

In this paper, we proposed a novel deep learning-based approach to photographic tone enhancement, which can enhance the tone and contrast of an image while preserving the naturalness. Our approach does not require carefully generated ground-truth images for training, but utilizes possibly unnatural-looking images that are automatically generated by a classical image filtering method. To preserve the naturalness, we proposed a GAN-based naturalness loss. To prevent artifacts, we also proposed artifact suppression loss using an imbalanced cycle-consistency. With our proposed loss functions, our method successfully produces appropriately tone enhanced and natural-looking results without ground-truth labels. We hope this approach would be useful for other image processing tasks as well, where it is hard to obtain ground-truth labels.

Limitations In our method, noise and compression artifacts could be boosted in dark regions as shown in Fig. 11. This limitation is inherited from our guide algorithm [SLKL14], which tends to magnify such artifacts when enhancing details, as shown in Fig. 11(b). While our naturalness preserving loss restrains boosting those artifacts to some extent, it cannot completely prevent the boosting as the artifacts already present in the input image are hard to distinguish from low-contrast image details.

Acknowledgements

This work was supported by the Ministry of Science and ICT, Korea, through IITP grant (IITP-2015-0-00174) and NRF grants (NRF-2017M3C4A7066316, NRF-2017M3C4A7066317).

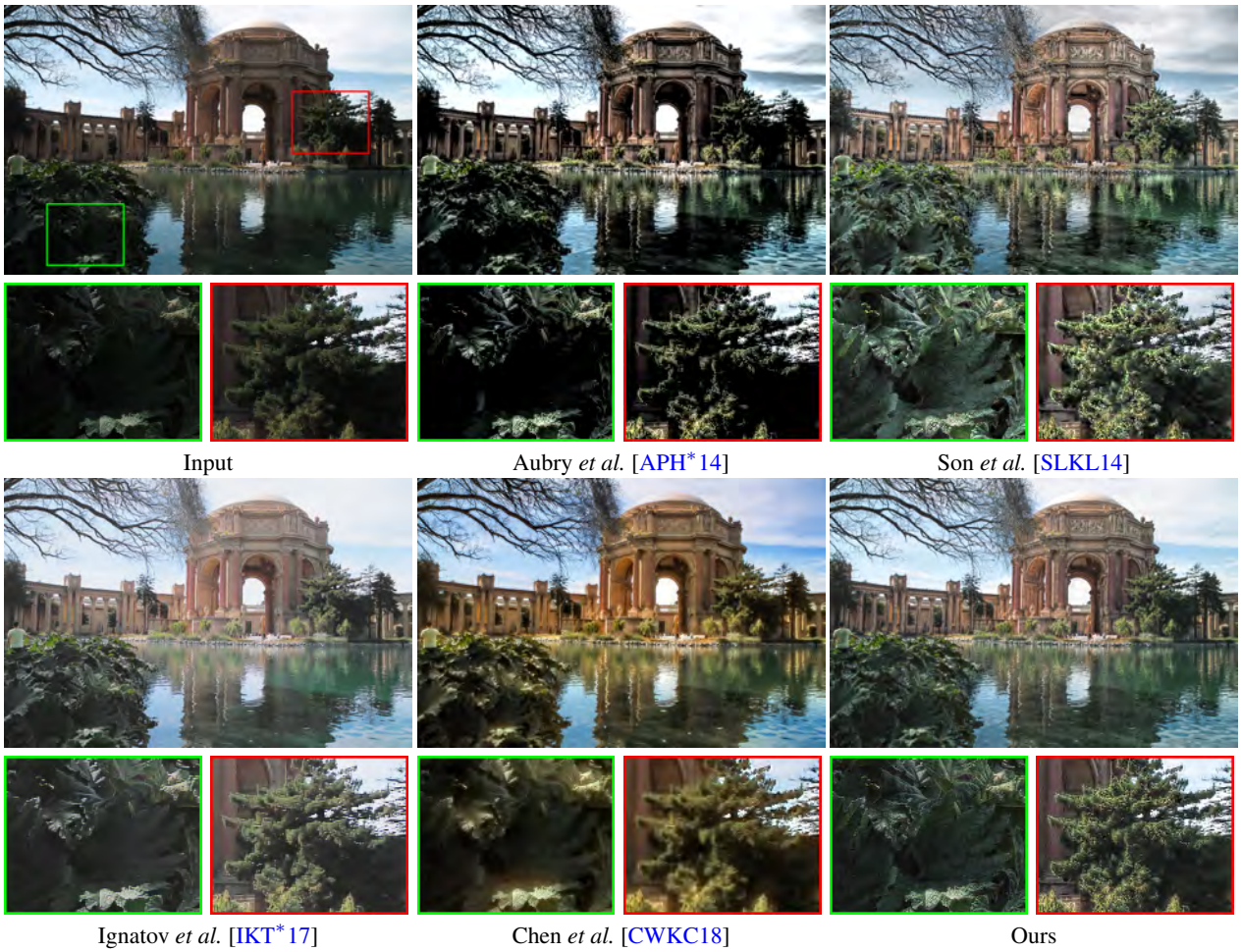


Figure 7: *Qualitative comparison against state-of-the-art image enhancement methods.*

Figure 8: *Additional comparison against state-of-the-art image enhancement methods for a high-resolution image (*1600×1200*).*

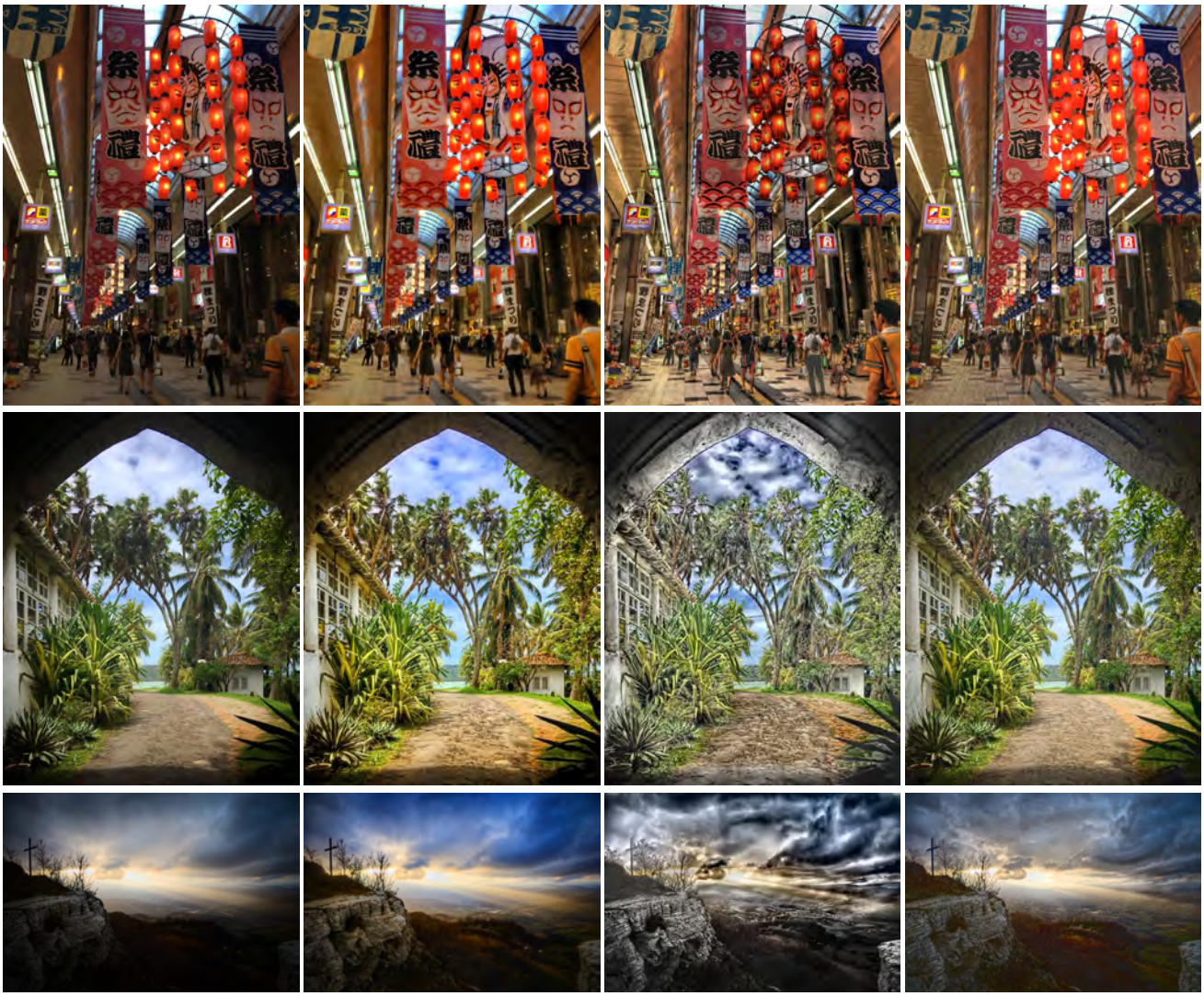
References

[APH*14] AUBRY M., PARIS S., HASINOFF S. W., KAUTZ J., DURAND F.: Fast local Laplacian filters: Theory and applications. *ACM TOG 33*, 5 (2014). 2, 6, 7, 8

[BPCD11] BYCHKOVSKY V., PARIS S., CHAN E., DURAND F.: Learning photographic global tonal adjustment with a database of input / output image pairs. In *Proc. CVPR* (2011). 1, 2, 3

[BPD06] BAE S., PARIS S., DURAND F.: Two-scale tone management for photographic look. *ACM TOG 25*, 3 (2006), 637–645. 1, 2

[CPD07] CHEN J., PARIS S., DURAND F.: Real-time edge-aware image



Input | Chen *et al.* [CWKC18] | Son *et al.* [SLKL14] | Ours

Figure 9: *Additional comparison results.*

Subjects	Aubry [APH*14]	Son [SLKL14]	Ignatov [IKT*17]	Chen [CWKC18]	Ours
naturalness	2.83 (±1.44)	2.47 (±1.46)	**3.83** (±1.00)	3.46 (±1.30)	3.77 (±1.09)
Tone + detail quality	2.80 (±1.14)	2.45 (±1.23)	3.30 (±1.03)	3.19 (±1.20)	**3.56** (±1.07)
Average	2.82	2.46	3.56	3.33	**3.67**

Table 1: *User study results. Values in parentheses denote standard deviations. Our method achieved runner-up performance in terms of naturalness while achieving the best performance in the quality of tone and detail enhancement.*

processing with the bilateral grid. *ACM TOG 26*, 3 (2007). 2

[CWKC18] 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 *Proc. CVPR* (2018). 1, 2, 6, 7, 8, 9

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM TOG 21*, 3 (2002), 257–266. 2

[FAR07] FATTAL R., AGRAWALA M., RUSINKIEWICZ S.: Multiscale shape and detail enhancement from multi-light image collections. *ACM TOG 26*, 3 (2007). 2

[GAA*17] GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A. C.: Improved training of wasserstein GANs. In *Proc. NIPS*. 2017. 4

[GCB*17] GHARBI M., CHEN J., BARRON J. T., HASINOFF S. W., DURAND F.: Deep bilateral learning for real-time image enhancement. *ACM TOG 36*, 4 (2017). 2

[GW06] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006. 1

[HHX*18] HU Y., HE H., XU C., WANG B., LIN S.: Exposure: A white-

Figure 10: *Images used in the user study. Values under each image are the average user scores for the naturalness and enhancement quality. (Please zoom in for best view)*

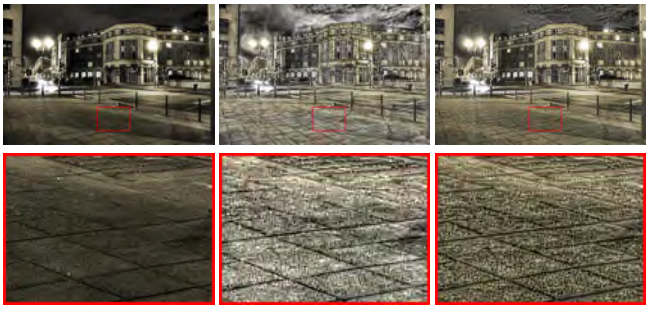


Figure 11: *Limitation example.*

box photo post-processing framework. *ACM TOG 37*, 2 (2018). 1, 2

[IKT*17] IGNATOV A., KOBYSHEV N., TIMOFTE R., VANHOEY K., GOOL L. V.: DSLR-quality photos on mobile devices with deep convolutional networks. In *Proc. ICCV* (2017). 1, 2, 6, 7, 8, 9

[IKT*18] IGNATOV A., KOBYSHEV N., TIMOFTE R., VANHOEY K., GOOL L. V.: WESPE: Weakly supervised photo enhancer for digital cameras. In *Proc. CVPR Workshop* (2018). 1, 2

[LTH*17] LEDIG C., THEIS L., HUSZAR F., CABALLERO J., CUNNINGHAM A., ACOSTA A., AITKEN A., TEJANI A., TOTZ J., WANG Z., SHI W.: Photo-realistic single image super-resolution using a generative adversarial network. In *Proc. CVPR* (2017). 4

[MLX*17] MAO X., LI Q., XIE H., LAU R. Y., WANG Z., SMOLLEY S. P.: Least squares generative adversarial networks. *Proc. ICCV* (2017). 3

[PHK11] PARIS S., HASINOFF S. W., KAUTZ J.: Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. *ACM TOG 30*, 4 (2011). 1, 2

[PLYK18] PARK J., LEE J.-Y., YOO D., KWEON I. S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In *Proc. CVPR* (2018). 1, 2

[PSC*18] PARK S.-J., SON H., CHO S., HONG K.-S., LEE S.: SRFeat: Single image super-resolution with feature discrimination. In *Proc. ECCV* (2018). 4

[SLKL14] SON M., LEE Y., KANG H., LEE S.: Art-photographic detail enhancement. *Computer Graphics Forum 33*, 2 (2014), 391–400. 1, 2, 3, 4, 5, 6, 7, 8, 9

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Proc. ICCV* (1998). 2

[TM18] TALEBI H., MILANFAR P.: NIMA: Neural image assessment. *IEEE TIP 27*, 8 (2018), 3998–4011. 2

[XLXJ11] XU L., LU C., XU Y., JIA J.: Image smoothing via L0 gradient minimization. *ACM TOG 30*, 6 (2011). 2

[XRY*15] XU L., REN J. S. J., YAN Q., LIAO R., JIA J.: Deep edge-aware filters. In *Proc. ICML* (2015). 2

[YZW*15] YAN Z., ZHANG H., WANG B., PARIS S., YU Y.: Automatic photo adjustment using deep neural networks. *ACM TOG 35*, 2 (2015). 1

[ZBSS04] ZHOU WANG, BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE TIP 13*, 4 (2004), 600–612. 4

[ZPIE17] ZHU J.-Y., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. ICCV* (2017). 2, 4