GAN 신경망을 통한 자각적 사진 향상

궈월*, 이효종*, ㅇ
*전북대학교 컴퓨터공학부
e-mail: yque86@gmail.com, hlee@chonbuk.ac.kr

Perceptual Photo Enhancement with Generative Adversarial Networks

Yue Que*, Hyo Jong Lee*, ㅇ
*Division of Computer Science and Engineering, CAIIT, Chonbuk National University

ABSTRACT

In spite of a rapid development in the quality of built-in mobile cameras, their some physical restrictions hinder them to achieve the satisfactory results of digital single lens reflex (DSLR) cameras. In this work we propose an end-to-end deep learning method to translate ordinary images by mobile cameras into DSLR-quality photos. The method is based on the framework of generative adversarial networks (GANs) with several improvements. First, we combined the U-Net with DenseNet and connected dense block (DB) in terms of U-Net. The Dense U-Net acts as the generator in our GAN model. Then, we improved the perceptual loss by using the VGG features and pixel-wise content, which could provide stronger supervision for contrast enhancement and texture recovery.

1. Introduction

With the rapid development of mobile phone cameras quality, users enjoy taking photographs even more. However, current cameras have physical limitations. Low-end devices must rely on advanced software and hardware post-processing tools to take decent photos under appropriate lighting conditions. They have to reconstruct a high-quality image from a set imperfect samples of the scene. However, mobile cameras still fall behind their digital single lens reflex (DSLR) counterparts in photography. DSLR camera can take high-quality photos because it has larger sensors and high-aperture optics yield better photo resolution, color rendition and less noise. These advantages can be a problem with small mobile cameras because of their small sensors and compact lenses [1].

Image enhancement techniques can address those issues with color rendition and image sharpness in some extent. There are a number of photographer tools for this purpose. They are usually focused on adjusting only global parameters such as histogram equalization, sharpening and contrast, without improving texture quality. And it often takes plenty of time to obtain satisfactory retouching results. Therefore, the primary method to image post-processing is still based on manual image correction using specialized retouching tool. In addition, they could be weak and lead to bad effects [2].

This paper proposes a method for image enhancement by learn the transformation that modifies photos taken by a given camera to DSLR quality ones. We treat the image enhancement problem as an image-to-image translation problem in which an input image is transformed into an enhanced image with the DSLR quality. Since the

Figure 1. iPhone photo enhanced to DSLR-quality by our method.

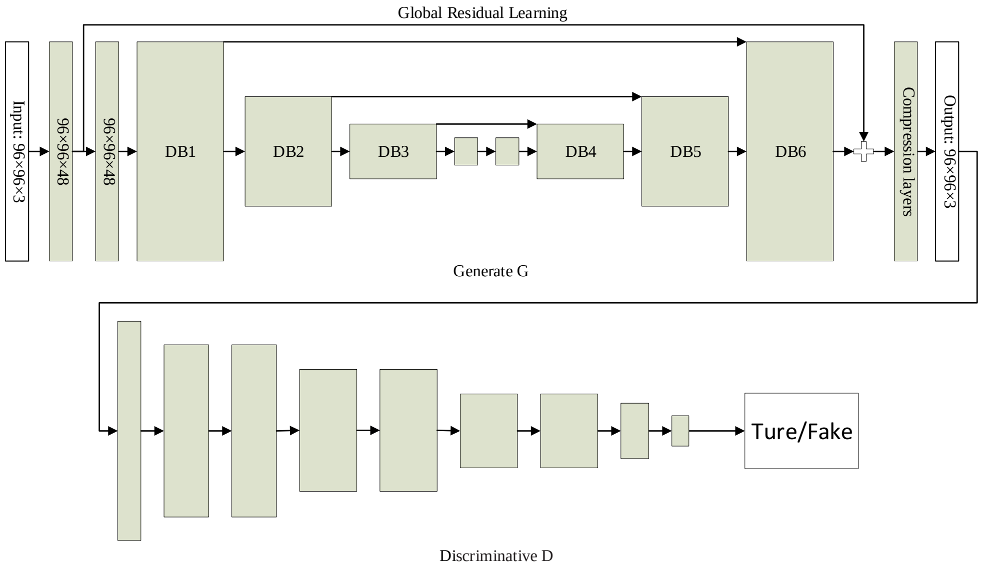

Figure 2. Image enhanced to DSLR-quality by our method. It consists of an image generation network G and a discriminative network D.

convolution neural network (CNN) has been great successful in computer vision and image processing tasks, deep learning based approaches have brought prosperous development. Various network architecture designs and training strategies have continuously improved the image-to-image translation performance. In addition, generative adversarial networks (GANs) [3] have been proved powerful in a variety of applications. Thus, we tackle the problem with a framework based on GANs with several improvements. First, for the design of the generator, we combine the U-Net [4] and Dense net [5] with global feature learning. The global learning adaptively preserves the hierarchical features in a global way. They are helpful for extract multi-level features and preserve spatial feature for better intra context exploration. Second, we propose an improved perceptual loss by using the VGG features and multi-scale structural similarity index (MS-SSIM) [6] instead of pure pixel space. We find that the adjusted perceptual loss provides sharper edges and more visually pleasing results.

2. Proposed Model

Our main goal is to improve the overall perceptual quality for image enhancement. In this section, we first describe our proposed network architecture and then discuss the improvements from the perceptual loss.

In order to further improve the enhanced image quality of standard GANs, we propose a novel architecture of generator G. We build an end-to-end network to fully make use of all of the hierarchical information from the source image. It combines the advantages of dense connections and U-Net connections together. We connect dense block (DB) in terms of U-Net as depicted in Fig. 2. DenseNet allows for direct connections between any two layers for each dense block. Hence, their feature maps, which are obtained by each layer in the encoding network, are cascaded as the input of the next layers. The dense connections can preserve as much information between layers as possible. In the decoder, we use skip-connections to transfer each level information of the encoder to the corresponding level on the decoder part.

The discriminative network D is introduced to compute the discrepancy between the generated images and the ground-truth images. Specifically, given an input image, the discriminative network D extracts high-level features using a series of Convolution, BatchNorm and LeakyReLU layers. Finally, the last convolution layer is flattened and then fed into a single sigmoid output. The output of the discriminative network D estimates the probability that the input image comes from the ground-truth dataset rather than from the image generation G.

Pixel-wise losses and perceptual losses are widely used in existing works for generating images towards the ground truth. The pixel-wise losses penalize the discrepancy occurred in the pixel space, but often produce blurry results. The perceptual losses explore the discrepancy between high-dimensional representations of images extracted from a well-trained classifier, e.g., the VGG net trained on the ImageNet dataset. We develop a more effective loss that combine the perceptual loss, pixel-wise loss and an adversarial loss component. The perceptual loss L_{p} by extract features using VGG19. It is previously defined on the activation layers of a pre-trained deep network, where the distance between two activated features is minimized. Wang et al. [7] proposed to use features before the activation layers, which will overcome some drawbacks of the original design. Therefore, the total loss for the generator is:

$$L_G = L_p + \lambda L_{GAN} + \eta(1 - L_{\mathrm{ssim}}) \quad (1)$$

where L_{ssim} evaluate the MS-SSIM between recovered image and the ground-truth, and λ, η are the coefficients to balance different loss terms.

3. Experiments

In order to make the end-to-end learning of single image enhancement possible, obtaining an enough large set of well-structured training image is a key challenge. Ignatov *et al.* [1] collect a large-scale real-world dataset, namely the “DSLR

(a) input (b) ours (c) NPEA (d) CLHE

Figure 3. Comparisons of our models with CLHE and NPEA.

Photo Enhancement Dataset". It consists of photos taken in the wild synchronously by three smartphones and one DSLR camera. Training CNN on the aligned high-resolution images is infeasible, patches of size 96×96 were extracted from these photos. The total number of cropped patches is 160471. Our method is implemented with Pytorch framework and updated with Adam optimizer. All of the experiments are implemented on a PC with GTX Titan X GPU and 64GB RAM. For test, we use our proposed method on a set of test images taken by mobile devices and compare how close the results are to the DSRL shots.

Figure 3 compares our models with several methods including CLHE [8] (Figure 3(c)) and NPEA [9] (Figure 3(d)). The result of our method shows the best result among all compared methods. It looks sharp and natural with good color rendition. Note that the inputs of the first row of Figure 3 were taken by mobile phones.

4. Conclusion

We proposed a photo enhancement method to effectively transform images from common smartphones into high quality DSLR images. Our end-to-end deep learning approach uses a composite perceptual loss function that combines vgg19 net based feature loss and MS-SSIM based pixel-wise loss. Our qualitative assessments reveal that the enhanced images demonstrate a quality comparable to DSLR taken photos, and the method itself can be applied to cameras of various quality levels.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2016R1D1A3B03931911). The work was also supported by the China Scholarship Council.

References

[1] Ignatov A, Kobyshev N, Timofte R, et al. DSLR-quality photos on mobile devices with deep convolutional networks. Proceedings of the IEEE International Conference on Computer Vision. 2017: 3277-3285.

[2] Chen Y S, Wang Y C, Kao M H, et al. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6306-6314.

[3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Advances in neural information processing systems. 2014: 2672-2680.

[4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[5] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[6] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment. The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2: 1398-1402.

[7] Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks. European Conference on Computer Vision. Springer, Cham, 2018: 63-79.

[8] Wang S, Cho W, Jang J, et al. Contrast-dependent saturation adjustment for outdoor image enhancement. JOSA A, 2017, 34(1): 7-17.

[9] Wang S, Zheng J, Hu H M, et al. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 2013, 22(9): 3538-3548.