Patching up the Outdoor Images by Enlarging the Dataset

Jung Hwan Kim and Dong Seog Han*
[1]School of Electronic and Electrical Engineering, Kyungpook National University, Republic of Korea
jkim267@knu.ac.kr, dshan@knu.ac.kr*

Abstract

Degraded and damaged images happen due to bad weather conditions after taking a photo with a camera in outdoor. The denoising technique is to restore the degraded areas of the original images. To complete the building of the denoising images, a few photos can be used to inspect the progress of building the system foundation. However, a hundred samples for training and testing cannot produce any satisfying performance to restore images, and performance limitations exist. To solve the problem, we increased the number of samples to train and test and constantly monitored the experimental results. After testing pix2pix architecture and increasing up to 50,000 paired samples, we increased the structural similarity index measure (SSIM) by 2.64%.

I. Introduction

Degraded images happen due to poor weather conditions during the outdoor photoshoots [1] [2]. The denoising technique assists in restoring the original photos without altering them. However, such a technique requires building the essential foundation for training and testing. We separated them into 100 paired images for building the foundation and 50,000 paired samples for actual training and testing after we obtained Snow100K from the Desnownet [3]. After training the pictures with 100 paired samples, the pix2pix [4] neural architecture produced the generated outputs. Still, the generated images are irrelevant to either ground truth or noisy images. In our approach, we produced improved pictures from the photos by increasing the number to 50,000 paired images and inspecting the performance.

II. Proposed Approaches

From Fig. 1, we obtained paired data samples called Snow100K from the DesnowNet. The dataset consists of ground truths and synthetics and has 50,000 images in pairs. The image size is generally $640 \times 480 \times 3$. During the training, we set the size of an image as $256 \times 256 \times 3$ pixels.

Fig. 1 The obtained data paired samples from DesnowNet.

From Fig.2, the pix2pix architecture is a generative adversary network with a unique generator model's design for encoder, decoder, and U-Net designs. The discriminator model's design is more straightforward and shallower than the generative model's. The pix2pix architecture has generative and discriminator models and produces realistic images. The generative model of the pix2pix architecture produces fake outputs. At the same time, the discriminator considers them counterfeit. The discriminator provides feedback to the generative model until it considers the fake outputs from the generative model to be real images.

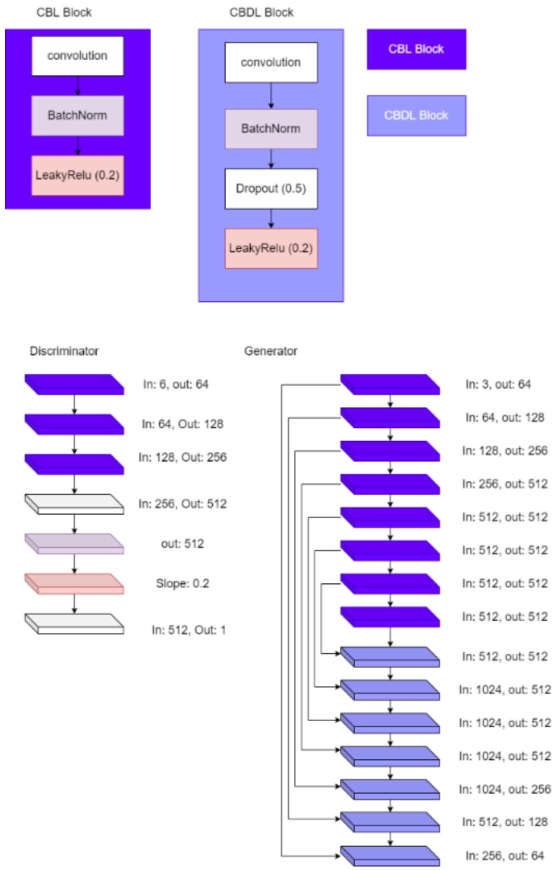

Fig. 2 The Structural diagram of the pix2pix architecture.

While building the system's foundation for training and testing, 50,000 paired samples were enormous, and it took a lot of time to inspect if the system's foundation was fully functional. Thus, 100 paired samples out of the 50,000 paired samples were brought to the pix2pix architecture to examine if the foundation created outputs. After the system's foundation was fully operational, we trained the pix2pix architecture with 50,000 paired examples and inspected the pre-trained pix2pix model's performance.

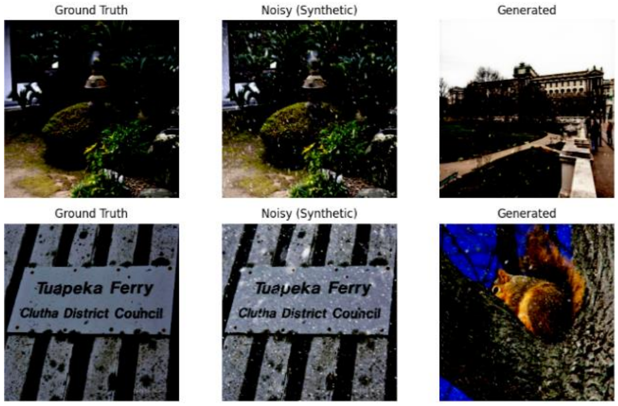

Fig. 3 Experimental results after training 100 paired images.

Fig. 4 Experimental results after training 50,000 paired images.

III. Experimental Results

Our hardware specifications were intel® Core™ i5-10600KCPU @ 4.10Ghz(12 CPUs), 4.1 Ghz, 32GB of RAM, and NVIDIA GeForce RTX 2070 Super. Before conducting training on the pix2pix architecture, we set the epoch as 150, 256 × 256 × 3 pixels as input and output images, the initial learning rate as 0.001, the mini-batch size as 4, and the split ratio for training and testing as 70 to 30. The inputs are noisy (synthetic) images, and the generated outputs must be similar to the ground truth images and evaluate the structural similarity index measure (SSIM) [5] between the ground truth images and generated outputs.

From Fig. 3, we conducted 70 paired images for training and 30 paired images for testing. The generated outputs were irrelevant to the ground truth images and unrecognizable to any shape in the picture. The SSIM was 26.07% between the ground truth and generated outputs.

From Fig 4, we observed the generated results from 35,000 paired images as a training group and 15,000 paired for testing. The shapes in the photos are recognizable, yet the outputs are still irrelevant to the ground truth. The SSIM was 28.71% since we are more recognizable than Fig. 3 shows. Therefore, 2.64% increased after training pix2pix architecture with 50,000 paired examples.

IV. Conclusion

In this paper, we improved the SSIM score by 2.64% as we enlarged the number of the trainable dataset. Expanding the dataset may be a considerable remedy for the poor neural network's performance, yet a few datasets may be used to construct the fundamental system to be operated.

ACKNOWLEDGMENT

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) Funded by the Ministry of Education under Grant 2021R1A6A1A03043144.

References

[1] F. Ding, K. Yu, Z. Gu, X. Li, and Y. Shi, "Perceptual enhancement for autonomous vehicles: Restoring visually degraded images for context prediction via adversarial training," IEEE Transactions on Intelligent Transportation Systems, vol. 23, pp. 9430-9441, 2021.

[2] Y.-C. Zeng, M.-F. Chueh, and C.-H. Tsai, "Automatic recognition of frame quality degradation for inspection of surveillance camera," in Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, 2012, pp. 1-8.

[3] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "Desnownet: Context-aware deep network for snow removal," IEEE Transactions on Image Processing, vol. 27, pp. 3064-3073, 2018.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125-1134.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, pp. 600-612, 2004.