J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr


# GAN-based shadow removal using context information☆

Hee-jin Yoon[1] Kang-jik Kim[1] Jun-chul Chun[1*]

ABSTRACT

When dealing with outdoor images in a variety of computer vision applications, the presence of shadow degrades performance. In order to understand the information occluded by shadow, it is essential to remove the shadow. To solve this problem, in many studies, involves a two-step process of shadow detection and removal. However, the field of shadow detection based on CNN has greatly improved, but the field of shadow removal has been difficult because it needs to be restored after removing the shadow. In this paper, it is assumed that shadow is detected, and shadow-less image is generated by using original image and shadow mask. In previous methods, based on CGAN, the image created by the generator was learned from only the aspect of the image patch in the adversarial learning through the discriminator. In the contrast, we propose a novel method using a discriminator that judges both the whole image and the local patch at the same time. We not only use the residual generator to produce high quality images, but we also use joint loss, which combines reconstruction loss and GAN loss for training stability. To evaluate our approach, we used an ISTD datasets consisting of a single image. The images generated by our approach show sharp and restored detailed information compared to previous methods.

☞ keyword : Shadow Removal, Generative Adversarial Network, Deep-Learning

## 1. Introduction

Removing shadows has been considered a challenge in the field of computer vision. When there are opaque objects in the path of sunlight, shadows arise and depend on various factors such as the location of the object and the altitude of the sun. Shadows of various shapes distort two different objects into a single object or occlude details information of the object. Traditional researches, a common approach to removing shadows consists of detecting shadows and using the detected mask as a clue to remove shadows.

The field of shadow detection predicts the location of shadows within an image and separates the shadow and non-shadow region of the original image in pixels. This has been considered challenging to classify shadows because shadows have various properties. For example, if there is an object with a black texture in the image, it can be misclassified as a shadow region. In addition, depending on the degree of occlusion by the object, the brightness of the shadows varies such as umbra, penumbra.


(Figure 1) The result of shadow removal using different discriminator. Left : only local patch, Right : global & local patch

Shadow removal is also a difficult task because we have to remove the shadows and restore the color and image information according to the degree of occlusion.

1 Div. of Computer Science and Engineering, Kyonggi University, 154-42 Kwangkyosan-ro Youngtong-gu Suwon-si Gyeonggi-do, 16227, Republic of Korea
* Corresponding author (jcchun@kgu.ac.kr)
[Received 20 June 2019, Reviewed 23 July 2019(R2 10 September 2019), Accepted 15 October 2019]
☆ Apreliminary version of this paper was presented at ICONI 2018.
☆ This research was supported by Basic Science Research program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No 2018R1D1A1B07042498).
☆ This work was also supported by Kyonggi University's Graduate Research Assistantship 2019.

Recently, as hardware has evolved, it has become possible to apply deep learning in many fields. It is also applied to shadow detection and removal field and has high performance. Many shadow detection and removal researches can be divided into CNN(Convolution Neural Network) and GAN (Generative Adversarial Network). In this paper, we remove the shadow based on GAN which generates realistic image. Using the detected shadow mask as a condition, we tried to improve the shadow region in the original image. However, traditional U-net based generators have problems such as large size of model and loss of color information and detailed information. Also, previous researches have made it difficult to generate realistic images because the discriminator works adversarial learning with only local patch information[1][2]. To solve these problems, We propose a generator network consisting of encoder, decoder and residual block[3]. We also added extra layers that extract global context information by using traditional patch discriminator as shared layers. One is to understand the whole scene of the generated image, and the other layer is to understand the details of the image patches as shown Figure 1. The loss function uses joint loss considering the stability of training. Weighted MSE loss and GAN loss improve the realism of fake image[4].

During the training progressed, the shadow was removed from the original image, and the color and detailed information of the occluded object were restored. We evaluated using the publicly available shadow removal dataset, ISTD[5]. We show the results of the traditional local patch, global, and ours. When we use both contextual information simultaneously, the details of the generated image are restored and show improvement in performance measurement.

Our contribution in detail is as follows.

- First, We design a residual based generator optimized to remove shadows and restore various information.
- Second, We propose a new discriminator that takes into account global and local patch information of the generated image.
- Third, We evaluated the ISTD benchmark dataset and compared the shadow and non-shadow areas separately.

# 2. Related work

## 2.1 Shadow detection

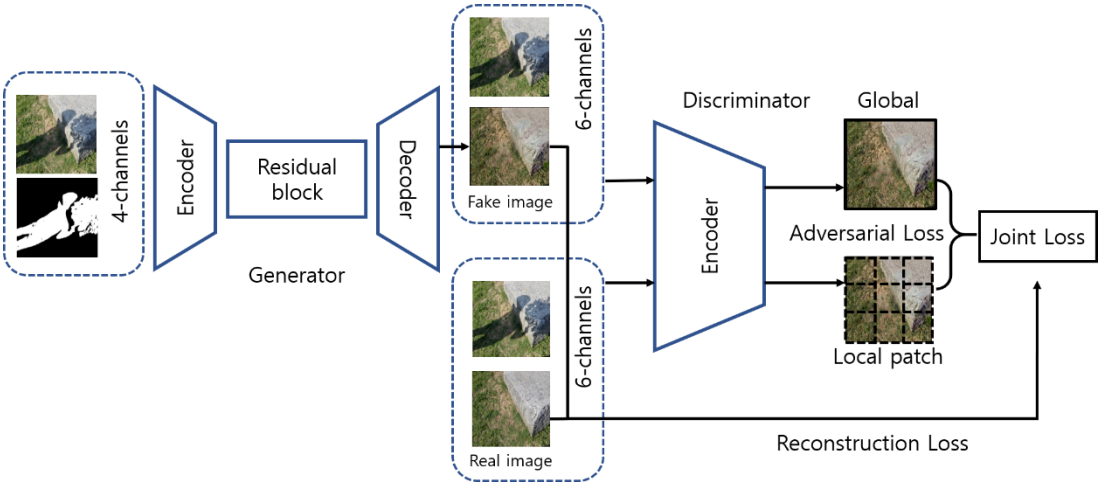
### 2.1.1 Deep learning based shadow detection

Many researches have been conducted to detect shadow regions that affect shadow removal quality using global context information. Zhu et al. [6] proposed CNN of the bidirectional feature pyramid architecture, They predicts the map of the state by repeatedly combine and refinement the context information. Qu et al. [7] detected shadows using a network that extracts the features of the three contexts of the image. They use the appearance, semantic, and global features of the input image to predict the final mask. They also proposed multi-context detection method. V. Nguyen et al. [8] adjusted the amount of shadow mask pixels generated through sensitivity parameters based on CGAN and detected shadow regions and non-shadow regions with different weights. They train by manipulating weights on image patches at various locations.

## 2.2 Shadow removal

### 2.2.1 Physical based shadow removal

In the initial shadow removal researches, shadows were detected and removed using physical features such as color, illumination, and shape. Guo et al. [9] performed region-based inference using graph cut method and estimated shaded area considering surface material properties. The detected shadow area reconstructs the pixel by the modeled reconstruction formula. Yang et al. [10] obtained a 2d-intrisic image by using a color-coded camera without a shadow detection process, and applied a bilateral filtering method to remove a shadow by estimating a jointed 3d intrisic image. Gong et al. [11] is a method to remove shadows by interpolating and extrapolating input shadow area information roughly drawn by user interaction. However, traditional modeled methods have limited ability to remove shadows when irregular illumination or objects with various colors are present.




(Figure 2) Overview of our architecture

#### 2.2.2 Deep learning based shadow removal

Hu et al. [12] used the Euclidean loss function to remove shadows as well as to detect shadows by formulating weights according to four directions in the RNN-based spatial directional module through the attention mechanism based on CNN. Wang et al. [5] is a multi-task method that detects and removes shadow in an end-to-end approach. It consists of two CGANs based on U-net. One network module detects shadows to predict shadow masks, and the other removes shadow. Since these methods remove shadow using either global or local context feature, there is a limitation in restoring more realistic images.

### 2.3 Generative Adversarial Network

#### 2.3.1 Adversarial training

The basic GAN consists of a generator that generates an image and a discriminator that maps to approximate the real image distribution[13]. The two modules are competitively trained and consist of a pair. Regarding the commonly known deceiver and police concept, a deceiver tries to make a plausible fake money to deceive the police, and the police try to classify it as truth. By the same concept the generator acts as a deceiver, the discriminator acts as a police officer, and the police try to classify it as truth. By the same concept the generator acts as a deceiver, the discriminator acts as a police officer, and gives feedback to each other. If training goes well, gradually the distribution of the fake data generated by the generator becomes approximate to the distribution of the real data, making it difficult to distinguish between real and fake. However, it is difficult to achieve an optimized result by balancing the competitive learning of the two modules[14].

#### 2.3.2 Conditional GAN

Mirza et al. [15] proposed a framework for generating desired data by providing a condition y for each generator and discriminator in the existing GAN. The form of the condition y can be a multimodal vector such as a single one hot vector in MNIST label or a user tag in an image. The network has three components: an encoder that extracts feature vectors of the image, a decoder that restores the compressed vector back to the original image, and an encoder and skip connection of the corresponding decoder. In the discriminator, image patchGAN [2] was used to judge image patch information instead of whole image.

## 3. Proposed method

The overall flow of the shadow removal network is shown in Figure 2. Our network consists of two modules, generator and discriminator based on CGAN which makes the data

generated by the generator similarly to the real data through adversarial learning. We applied a residual-based generator and context discriminator. The main contribution of our proposed method is to make adversarial learning by judging the fake data generated by the generator through the global and local context discriminator from the aspect of contexts.

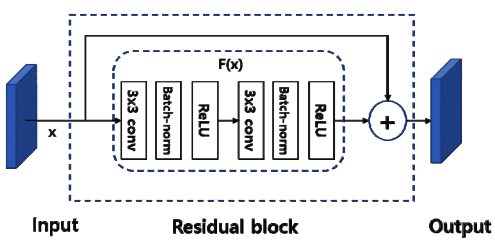
## 3.1 Network architecture

### 3.1.1 Adversarial training

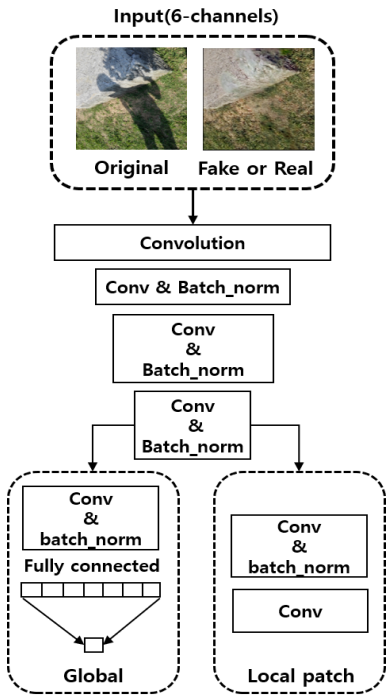
As shown in Figure 2, it consists of a generator and a discriminator. In the generator, it is optimized for shadow removal by residual based network. The generator consists of three parts: an encoder, nine residual blocks, and a decoder. The encoder part consists of four convolution blocks, and images are down-sampled to extract features. In the first convolution block, padding – 7x7 convolution – batch-norm – ReLU, and the others are consists of 3x3 convolution-batch norm- ReLU. The central part of the network consists of nine residual blocks. As shown in Figure 3, the 3x3 convolution - ReLU - 3x3 convolution block has solved many problems caused by deepening the layer through the shortcut connection.

$$F(x) := H(x) - x \qquad (1)$$

Considering H (x) as a network output, F (x) composed of convolution blocks approximates subtracting input image x from H (x). Therefore, it shows excellent results in terms of speed and performance[3]. In the decoder, the compressed vector is up-sampled into the original image and is paired with the encoder part. The first three layers consist of convTranposed2d - batch_norm - ReLU, and the last layer consists of ReflectionPad2d - 7x7 convolution – Tanh




(Figure 3) Residual block




(Figure 4) Global & local patch discriminator

activation function. In the discriminator, we concatenate the fake image generated in the generator with the original image or concatenate shadow-less GT image with original image. After that, input the discriminator alternately to judge the real or fake. The architecture of the discriminator module is as shown in Figure 4, in which a structure used in the traditional patchGAN [2] is used as a shared layer, and a layer for extracting global context information is added. The global discriminator judges the entire image through a 4x4 convolution - batch_norm - Fully Connected (FC) beyond the shared layer. In contrast, the local discriminator uses the 4x4 convolution - batch_norm - 4x4 convolution to apply the loss objective function in image local patch information. This discriminator has a problem of generating a blurry image although the speed is fast and the size of the model is small.

The proposed method is capable of adversarial learning on global and patch information, and the image detail and color information can be restored realistically compared with traditional methods.

### 3.2 Training and implementation

We implemented it based on the deep learning framework Pytorch, and the gpu used 1080Ti. The loss function is a joint of reconstruction loss and adversarial loss as shown Equation 2.

$$L = \lambda_1 reconstruction + \lambda_2 global + \lambda_3 local \quad (2)$$

The reconstruction loss is the L1 loss function between the generated fake image and the shadow-less GT image, as shown in Equation 3. This loss function computes the difference between the predicted image and the GT image in the model and applies the absolute value to help generate a plausible fake.

$$L_{reconstruction} = \sum_{i=1}^{n} |y_{true} - y_{predicted}| \quad (3)$$

In adversarial loss, Binary Cross Entropy loss is used as an objective function to classify real data and fake data, as shown Equation 4. x is the original image, and y is the mask data, which is the conditional data.

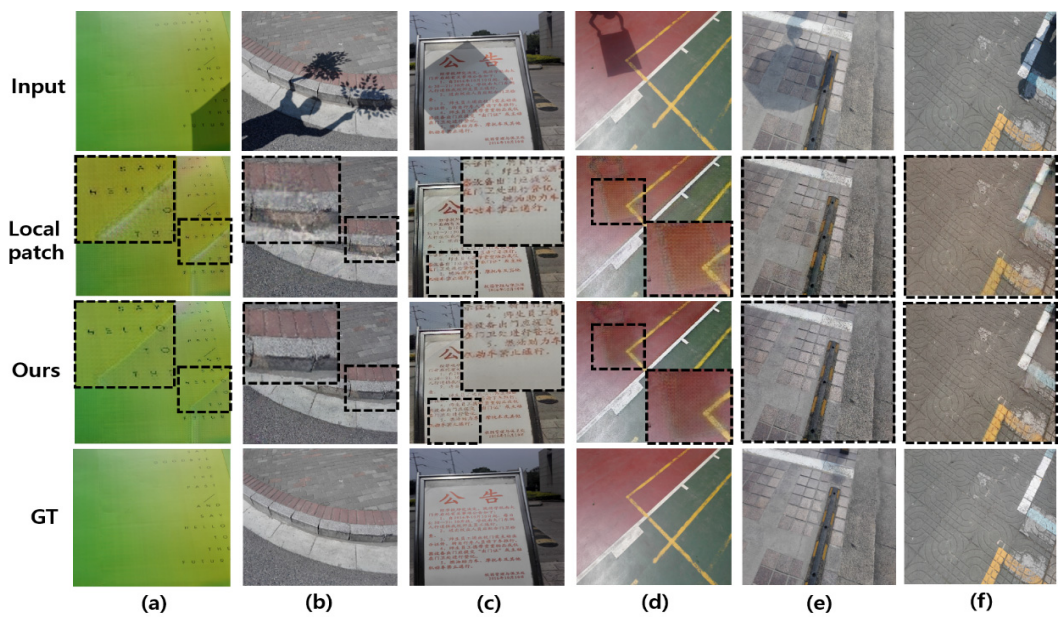
$$L_{adversarial}(G,D) = E_{x,y}[\log D(x,y)] + E_{x,y}[\log(1 - D(x, G(x,y)))] \quad (4)$$

The parameters are used $\lambda_1$= 0.995, $\lambda_2$= 0.0025, $\lambda_3$= 0.0025, and the learning rate is 0.0002[16]. The optimizer, which improves learning speed and stability, uses a adam optimizer for generator and discrimator.

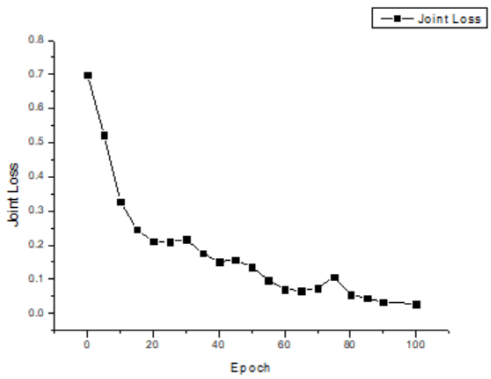
## 4. Experimental result

In this paper, dataset for shadow removal consists of 1870 triplet shadow images, shadow masks, and shadow-less images with ISTD[5]. In addition, the training data is 1,330 and is used as 540 test data. The scene in the image consists of 135 different types of texture with single image.

The training graph based on the joint loss function is shown in Figure 6. The loss function decreases and the network is optimized. We performed the test in the LAB




(Figure 5) The result shows the difference according to the discriminator in the same generator. (a ~ d) is the difference of the restored details information, and (e ~ f) is the difference in the whole image color.




(Figure 6) Training graph for our method

color space and used the Root Mean Square Error method to calculate the per pixel error as shown Equation 5.

$$RMSE=(\theta_1,\theta_2)=\sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_{1,i}-x_{2,i})^2} \quad (5)$$

This formula means the difference between the predicted image and GT image, where n is amount of image pixel.

(Table 1) Performance comparison of Shadow Removal

| Method | RMSE | | |
|---|---|---|---|
| | Shadow | Non shadow | All |
| Guo[9] | 18.95 | 7.46 | 9.3 |
| Yang[10] | 19.82 | 14.82 | 15.63 |
| Gong[11] | 14.98 | 7.29 | 8.53 |
| DSC+[12] | - | - | 6.67 |
| Local patch[2] | 2.44 | 3.58 | 5.15 |
| Ours | 2.02 | 3.62 | 4.93 |

In addition to lab color space is known as adevice-independent color space, so difference between the two images can be calculated more accurately than in the RGB color space[17]. We evaluated the performance by dividing it into shadows, non-shadows and the whole image region, and the results are shown in Table 1. Especially Compared with the traditional methods, it showed an improvement over local patch method [2] based on GAN in the shadow region. Regarding Figure 5, (a~d) shows the difference in the information of the restored image according to the discriminator, and (e~f) also shows the difference in the whole image color.

## 5. Conclusion

In this paper, We propose a novel context discriminator that can simultaneously judges global and local information. Based on the proposed approach, the shadow is removed using the detected mask and the original image. Using global and local patch information together, we can generate a more natural and improved image than when using only local patch information. In addition, we used adversarial loss and reconstruction loss as joint loss for stability in training. Our method can be used as a preprocessing process in various computer vision fields and can help interpret natural images with shadows. In the future, we will extend the end-to-end method to add an auxiliary model to resolve shadow detection and removal. In addition, the model is optimized for real-time processing and is expected to be useful in various real-time processing fields such as video surveillance and automobile lane detection.

## Reference

[1] O. Ronneberger, P. Fischer, and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, In: Medical Image Computing and Computer Assisted Intervention – MICCAI, Springer, pp.234-241, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[2] P. Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros “Image-to-Image Translation with Conditional Adversarial Networks.”, In CVPR, pp.1125-1134, 2017, https://10.1109/CVPR.2017.632

[3] K. He, X. Zhang, Shaoqing Ren, and J. Sun. “Deep residual learning for image recognition.”, In CVPR, pp. 770-778, 2016. https://doi.org/10.1109/CVPR.2016.90

[4] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, Alexei A. Efros, “Context Encoders: Feature Learning by Inpainting.”, In CVPR, 2016.

https://doi.org/10.1109/CVPR.2016.278

[5] J. Wang, X. Li, and J. Yang. “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal.” In CVPR, 2018.
https://doi.org/10.1109/cvpr.2018.00192

[6] L. Zhu, Z. Deng, et al, “Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection.”, In ECCV, 2018.
https://doi.org/10.1007/978-3-030- 01231-1_8

[7] L. Qu, J. Tian et al. “DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal.”, In CVPR, 2017.
https://doi.org/10.1109/CVPR.2017.248

[8] V. Nguyen, Tomás F. Yago Vicente et al. “Shadow Detection with Conditional Generative Adversarial Networks”, In ICCV, 2017.
https://doi.org/10.1109/ICCV.2017.483

[9] R. Guo, Q. Dai, D. Hoiem, “Paired Regions for Shadow Detection and Removal”, In IEEE, Vol.35, pp.2956-2967, 2012.
https://doi.org/10.1109/TPAMI.2012.214

[10] Q. Yang, K.-H. Tan, N. Ahuja. “Shadow removal using bilateral filtering.” IEEE Transactions on Image Processing(TIP), Vol. 21, No. 10, pp.4361 – 4368, 2012.
https://doi.org/10.1109/TIP.2012.2208976

[11] H. Gong, D. Cosker. “Interactive shadow removal and ground truth for variable scene categories.” In British Machine Vision Conference (BMVC). 2014.
https://doi.org/10.5244/C.28.36

[12] X. Hu, C.W. Fu et al, “Direction-aware Spatial Context Features for Shadow Detection and Removal.” In IEEE, 2019.
https://doi.org/10.1109/TPAMI.2019.2919616

[13] R. Mu, X. Zeng, ″A Review of Deep Learning Research″, In TIIS, Vol. 13, No.4, 2019,
10.3837/tiis.2019.04.001

[14] L. Metz, B. Poole, D. Pfau and Jascha Sohl-Dickstein, ″Unrolled Generative Adversarial Networks.″, In Arxiv, 2016.
http://arxiv.org/abs/1611.02163

[15] M. Mirza, S. Osindero, ″Conditional Generative Adversarial Nets″, In Arxiv 2014.
http://arxiv.org/abs/1411.1784

[16] S. Iizuka, E. Simo-Serra, H. Ishikawa “Globally and Locally consistent Image Completion.”, In ACM Transactions on Graphics, 2017.
https://doi.org/0.1145/3072959.3073659

[17] S. Y. Kahu, K. M. Bhurchandi, “JPEG-based Variable Block-Size Image Compression using CIE La*b* Color Space.”, In TIIS, Vol. 12, No.10, 2018.
https://doi.org/10.3837/tiis.2018.10.023

## ◐ 저 자 소 개 ◑

**윤 희 진(Heejin Yoon)**
2018 B.S in Computer Science, Kyonggi University, Suwon, South Korea
2019~Present, M.S. Computer Science, Kyonggi University, Suwon, South Korea
Research Interests : Computer Vision, Deep Learning.
E-mail : skek0624@naver.com

**김 강 직(Kangjik Kim)**
2019 B.S in Computer Science, Kyonggi University, Suwon, South Korea
2019~Present, M.S. Computer Science, Kyonggi University, Suwon, South Korea
Research Interests : Computer Vision, Deep Learning.
E-mail : Kangjik94@naver.com

**전 준 철(Junchul Chun)**
1984 B.S. in Computer Science, Chung-Ang University, Seoul, South Korea
1986 M.S. in Computer Science(Software Engineering), Chung-Ang University, Seoul, South Korea
1992 M.S. in Computer Science and Engineering(Computer Graphics), The Univ. of Connecticut, USA
1995 Ph.D. in Computer Science and Engineering(Computer Graphics), The Univ. of Connecticut, USA
2001.02~2002.02 Visiting Scholar, Michigan State Univ. Pattern Recognition and Image Processing Lab.
2009.02~2010.02 Visiting Scholar, Univ. of Colorado, Wellness Innovation and Interaction Lab.
Research Interests : Augmented Reality, Computer Vision, Human Computer Interaction
E-mail : jcchun@kgu.ac.kr