

Shadow Removal from Images Using Conditional GANs

Amrit Acharya¹, Ramesh Thapa¹¹Department of Electronics and Computer Engineering, IOE, Pashchimanchal Campus, Tribhuvan University, Nepal

(Manuscript Received:2023/08/31; Revised:2023/00/00; Accepted: 2023/00/00)

Abstract

Shadow removal has many applications in computer vision and shadow-free images have better visual quality. In recent studies, deep learning-based CNN models have shown better performance than traditional approaches to shadow removal. GAN takes the advantage of two independent neural networks. This study about shadow removal is implemented using GAN. Shadow removal is divided into two tasks: detection and removal. The two sub-networks stacked upon each other are based on conditional GAN. The input shadow image 256*256 is fed to the first generator network to produce a shadow mask, which is input to the second generator network along with a shadow image to obtain a shadow-free image.

Keywords: CGAN; GAN; Shadow detection; shadow removal; U-net

1. Introduction

Shadows are formed when the path of light is obstructed by an opaque object, creating an area of reduced illumination on the surface opposite to the source of light. The shadow of an object in the image can distort the appearance, shape, color and texture of the other object in the same image. Removing shadows in an image is of great interest in computer vision tasks such as image segmentation, object detection and object tracking. Shadow can affect the accurate analysis of images and the performance of algorithms that depend on consistent lighting conditions.

The traditional method of detecting shadows used color, texture and geometric properties to differentiate between shadow and non-shadow regions. Deep learning methods have revolutionized the field of computer vision as they can automatically learn relevant features from the data, allowing them to capture complex patterns and relationships.

Generative Adversarial Networks have emerged as a powerful tool for image-to-image translation tasks. A GAN consists of two networks: a generator that produces images resembling real data and a discriminator that distinguishes real from fake images; both networks improve iteratively as the generator learns to create more realistic images, while the discriminator becomes better at distinguishing real images from the generated ones in a continuous cycle of mutual training. Conditional GAN is a type of GAN where both the generator network and the discriminator network

receive additional Conditioning variables. The generator network in this work is based on U-net architecture, which excels in image segmentation tasks, and the discriminator network is based on a convolutional neural network. Two tasks are to be done in this proposed work: shadow detection and shadow removal. Each task is carried out by stacking two different GANs, one for each task. The details of the architecture are explained later in the methodology section.

In the domain of image analysis, segmentation plays a vital role in different applications. One of the pioneering contributions to the segmentation task is the introduction of U-net [10]. The U-Net architecture has a unique structure, combining contracting and expansive paths with skip connections, demonstrating impressive image segmentation capabilities by effectively capturing local details and global context, thereby addressing complex structure segmentation challenges. Different modifications and architectures have been used based on deep-learning neural networks for shadow detection and removal from the image. One of the extensions is the use of multiple stacked GANs. [2] Uses conditional GAN-based two different for shadow detection and removal. The first GAN aims to identify shadow region and the second GAN focuses on shadow removal, utilizing the result from the first stage to guide the removal process. The study [1] uses GAN to generate realistic image shadow patterns. Thus, generated shadow images are used as supplementary data to train the removal process. This model uses different sub-networks: shadow generation, shadow removal, and enhancement, which are interconnected to each other. The study in [3] highlights the

*Corresponding author. Tel.: +977-9846266519,
E-mail address: amrit.manutd@gmail.com

relationship between shadow detection and removal by integrating both tasks in the GAN framework. It uses multiple generators and discriminators to demonstrate better performance. The first stage uses GAN to detect shadow, and the generator aims to produce shadow-free images. and the result is used to guide the second stage for producing shadow-free images. The [5] uses conditional GAN for accurate shadow detection. The use of conditional information has produced better performance. The study in [6] presents a novel approach to remove shadow from images using unpaired data. The framework consists of two components: a generator network to remove the shadow from input images and a mask predictor that predicts the shadow mask in the image, which ultimately guides the removal process. The collaboration between the two components produced an effective result. The study in [11] extended the U-net architecture for better performance. It uses multiple U-net sequentially connected for improved results.

Thus, stacking U-net models can increase the network's overall capacity, allowing it to capture the intricate patterns and details in the image. Each U-net model used in each stage helps to specialize in different aspects of the task. This adaptability helps to deal with a wide range of objects, sizes, and textures. The model gains access to diverse contextual information by combining predictions from different stack levels.

2. Materials and Method

2.1 Block Diagram

The architecture comprises two sub-networks based on conditional GAN for shadow detection and removal. The block diagram is illustrated below:

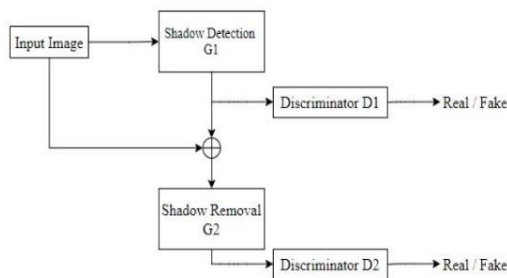


Figure 1: Block diagram of the model

2.2 Network Architecture

2.2.1 The Generator

The generator network on both GANs is based on U-Net Architecture. Both the generator networks G1 and G2 have the same U-net architecture except for the

number of inputs and outputs. The first generator network takes an input image of dimension $256 \times 256 \times 3$. Each network comprises convolutional blocks followed by Leaky ReLU activation in the contracting path and ReLU activation in the expanding path of U-net architecture. The U-net architecture of G1 is shown below, with the number of output channels and input channels in each layer.

The first generator network is trained to produce the shadow mask. Thus, it takes 3 channel inputs and provides 1 channel output. The input for generator 2 is the combination of the shadow image and the output from the first generator. Thus, the number of input channels for the second generator network increases to 4. It produces 3 channel outputs and the architecture for G2 is also the same as in Figure 2, except for the number of input channels.

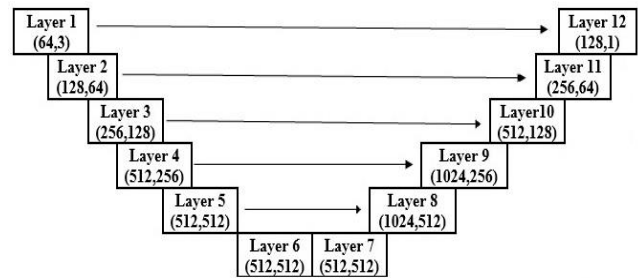


Figure 2: U-net based Generator (G1)

2.2.2 The Discriminator

The discriminator network for both networks is based on the Convolutional Neural Network. The discriminator network uses multiple convolutional layers to distinguish between real and generated images. The discriminator network consists of four convolutional layers activated with the Leaky ReLU activation function. Both discriminators D1 and D2 have the same architecture except for the number of input channels for the discriminator network. The architecture for discriminator D1 is shown below:

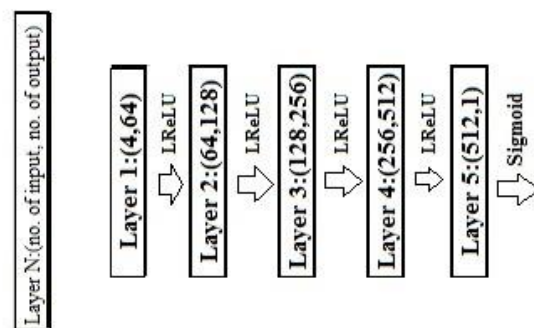


Figure 3: Architecture for Discriminator D1

The Discriminator D1 differentiates between the ground truth shadow mask and the shadow mask generated by the generator 1. Thus, it takes 4 channel inputs. The second discriminator network takes an additional 3-channel input of shadow-free image along with the first discriminator network, increasing dimensionality to 7-channel input. D2 compares the output from the stacked network with the real shadow-free image.

2.3 Dataset and Preprocessing:

This study uses the publicly available dataset Image Shadow Triplets Dataset (ISTD), which contains image triplets of shadow and shadow-free image and shadow masks. The image size is 286*286. A random flip is done for data augmentation.

2.4 Training Details

The results are obtained by implementing the training parameters as follows:

- Image size: 286*286
- Input image size: 256 * 256 * 3
- Learning rate: 2 e-4
- Optimizer: Adam Optimizer
- Discriminator and Adversarial loss: Binary cross entropy with logit loss

3. Results and Discussion

The output image obtained in the different training steps for different images is shown below. The first row of images is the shadow images (input), the second is the ground-truth shadow images, and the third is the shadow-free images generated by our model. The corresponding shadow mask and generated shadow mask are also shown below:

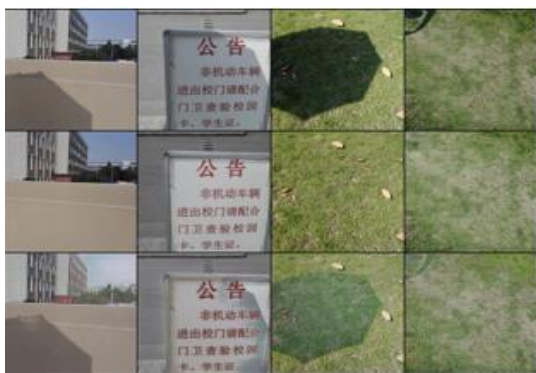


Figure 4: 100th training step

Initially, the generated shadow-free image looks similar to the shadow image and the shadow is only partially removed in Figure 4.

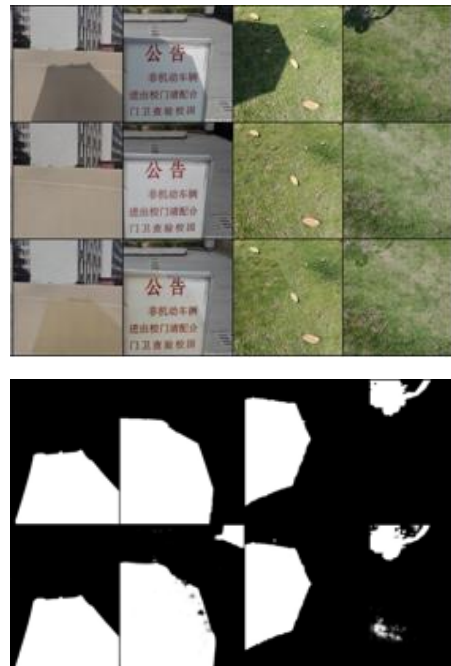


Figure 5: 500th training steps

With the progression of the training step, the shape of the detected shadow looks similar to the ground truth shadow mask. The amount of shadow in the generated image is significantly removed.





Figure 6: 1450th training step

In the fig above, the detected shadow looks similar to the real shadow mask, and the image's shadow is completely removed.

The similarity index between the real shadow-free image and the ground truth image is shown below:

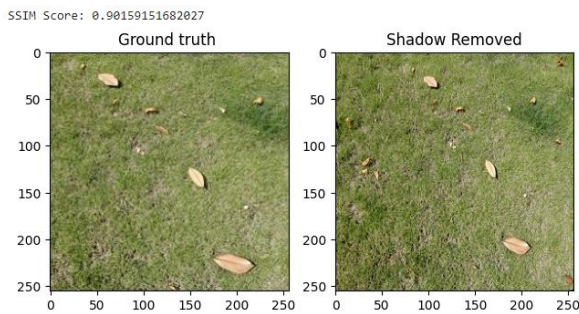


Figure 7: SSIM between ground truth and shadow removed image

The Discriminator vs. Generator loss for the first 1450 training steps is shown below:

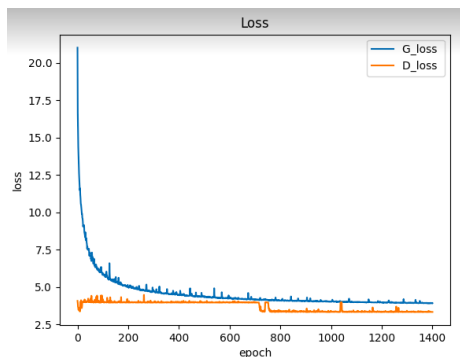


Figure 8: generator and discriminator loss with respect to epochs

The generator and discriminator loss are shown in Figure 8. In the beginning, the generator loss is high, but with the progression of iteration, it decreases. As the generator is not properly trained initially, the discriminator loss is low as it quickly learns to differentiate between real and fake samples. The generator and discriminator loss have variances as they compete against

each other.

4. Conclusions

The proposed approach of using conditional GAN with interconnected sub-networks demonstrated promising results. Shadows are significantly removed, revealing the true colors and details previously obscured. The use of U-net architecture ensures the fine details of the image are preserved. The generated images seem realistic in appearance.

Despite promising results, challenges like complex scenes, such as the shadow overlapping objects with similar colors, have led to occasional artifacts. The training procedure with multiple U-net generators has computational complexity.

In conclusion, using conditional GAN with multiple U-net architectures in the generator effectively removes shadow from the image. The inclusion of conditional information in the GAN framework provided guidance for shadow removal. Reducing computational complexity is the potential area for further refinement.

References

- [1] Liu, Z., Yin, H., Wu, X., Wu, Z., Mi, Y., & Wang, S. From shadow generation to shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021) 4927-4936.
- [2] Wang, J., Li, X., & Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018) 1788-1797.
- [3] Nagae, T., Abiko, R., Yamaguchi, T., & Ikehara, M. Shadow detection and removal using GAN. In *2020 28th European Signal Processing Conference (EUSIPCO), IEEE, (2021)* 630-634.
- [4] Hu, X., Fu, C. W., Zhu, L., Qin, J., & Heng, P. A. Direction-aware spatial context features for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence*, 42(11) (2019) 2795-2808.
- [5] Nguyen, V., Yago Vicente, T. F., Zhao, M., Hoai, M., & Samaras, D. Shadow detection with conditional generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, (2017) 4510-4518.
- [6] Hu, X., Jiang, Y., Fu, C. W., & Heng, P. A. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF international conference on computer vision*, (2019) 2472-2481.
- [7] Liu, Z., Yin, H., Mi, Y., Pu, M., & Wang, S. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing*, 30 (2021) 1853-1865.
- [8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B.,

Warde-Farley, D., Ozair, S., ... & Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*, 27 (2014).

[9] Vicente, T. F. Y., Hou, L., Yu, C. P., Hoai, M., & Samaras, D. Large-scale training of shadow detectors with noisily annotated shadow examples. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part VI 14* (2016) 816-832. Springer International Publishing.

[10] Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, Proceedings, Part III 18* (2015) 234-241. Springer International Publishing.

[11] Shah, S., Ghosh, P., Davis, L. S., & Goldstein, T. Stacked U-Nets: a no-frills approach to natural image segmentation. (2018). *arXiv preprint arXiv:1804.10343*.