Enhanced Shadow Removal for Surveillance Systems*

Jishnu P[1] and Rajathilagam B[2]

[1] Department of Computer Science Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
cb.en.p2aid19017@cb.students.amrita.edu
[2] Department of Computer Science Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
b_rajathilagam@cb.amrita.edu

Abstract. The presence of shadow is unavoidable while dealing with outdoor images in a variety of computer vision applications. In order to unveil the information occluded by shadow, it is essential to remove the shadow. This is a two-step process which involves shadow detection and shadow removal. In this paper, shadow-less image is generated using a modified conditional GAN (cGAN) model and using shadow image and the original image as the inputs. The proposed novel method uses a discriminator that judges the local patches of the images. The model not only use the residual generator to produce high-quality images, but also use combined loss, which is the weighted sum of reconstruction loss and GAN loss for training stability. Proposed model evaluated on the benchmark dataset,i.e.ISTD, and achieved significant improvements in the shadow removal task compared to the state of the art models..

Keywords: Shadow Removal· cGAN · Combined loss · ISTD

1 Introduction

Removing shadows from the images has been considered as a challenging task in the field of computer vision. The presence of opaque objects in the path of sunlight leads to the formation of shadows and depend on different factors such as the altitude of the sun and location of the object. For example, consider a bike and bus in traffic such that the bike is standing left side of the bus. There are chances for the shadow of the bus to cover the bike if the sunlight is from the right side of the bus. Shadow of different shapes distorts two different objects into a single object. This is called as the occlusion. This is a difficult situation in which we can't efficiently detect different objects. In this example, it will be difficult for us to distinguish between bike and bus. Probably the Bus and its shadow will merge together and form another shape which will be far different from the shape of a bus.

* Supported by Amrita Vishwa Vidyapeetham

Fig. 1. Expected outcome of Shadow removal.

Fig. 1 illustrates the expected outcome of the shadow removal process. In traditional approaches, a common method to remove shadow consist of detecting shadows and using detected shadow masks as a clue for removing the shadows.The field of shadow detection predicts the location of the shadowed region in the image and separates the shadowed and non-shadowed region of the original image in pixels. This has been considered a challenging task to classify shadows in an image because shadows have various properties. Depending on the degrees of occlusion by the object, the brightness of shadow varies such as umbra and penumbra. The dark part of the shadow is called the umbra, and the part of a shadow that's a little lighter is called the penumbra.

After the introduction of Generative Adversarial Networks (GAN) in 2014[1], the computer vision domain has taken leap at various tasks. Shadow removal is an important task which can be considered as an invaluable preprocessing step for higher level Computer Vision tasks in surveillance systems like road accident identification and severity determination from CCTV surveillance[2] and plant leaf recognition using machine learning techniques[3].

The challenging task is to ensure the higher quality in shadow-less image by using efficient evaluation metric and enhanced architecture. Shadow removal is also a difficult task because we have to remove shadows and restore the information in that region according to the degree of occlusion.

2 Background

Shadow removal considered as the complex process compared to the shadow detection phase due to the difficulty while reconstructing the pixels in the detected shadow region. Jifeng Wang[4] introduced the ISTD dataset as part of the work titled "Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal" and which considered as one of the benchmark dataset for the shadow removal process. From this view it is clear that the research in the shadow detection domain almost saturated and the focus is on the enhancements in the shadow removal phase. They proposed an architecture which contains two conditional GAN(cGAN) stacked together and performs shadow detection and removal tasks simultaneously. With this model RMSE value of 7.47 achieved on the ISTD dataset. Lack of considering the context information in the shadow removal phase was the drawback of this model.

In order to consider more details of the shadowed image like the illumination information, Ling Zhang[5] proposed a GAN based model which contains 4 generators and 3 discriminators in the work entitled “RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal”. This model achievedan RMSE value of 6.97 which is better than the model proposed by JifengWang[4]. The model trained on two different benchmark datasets for shadow removal which are SRD and ISTD datasets. RMSE value of 6.78 achieved for the the SRD dataset.

The context information was still missing in these models. The model proposed by L. Qu[6] use the multi-context information for removing the shadow and reconstructing the shadowed region. Their work titled as “DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal ” uses only deep convolutional neural networks(DCNN) for shadow detection and shadow removal. Acquisition of multi-context information achieved by training three different networks named G-Net(Global Net), A-Net(Appearance net), S-Net(Semantic Net). Global context, appearance information and semantic context of the shadowed region captured by these networks. All these three networks together called as the DeshadowNet. The RMSE value scored on the SRD dataset is 6.64. Blurred effect was present in the shadowed region of the output images.

Generative Adversarial Network(GAN) base models produce high quality output images compared to the traditional model. Sidorov[7] proposed AngularGAN architecture to improve the color constancy on the images. Their work entitled “Conditional GANs for Multi-Illuminant Color Constancy: Revolution or Yet Another Approach?” used the GAN based model along with the angular information of the incident light on the object. The model trained on the synthetic dataset called GTAV dataset. Peak Signal to Noise Ratio(PSNR) used to evaluate the performance of the model. PSNR value for the GTAV dataset was 21dB. ”Shadow Detection and Removal for Illumination Consistency on the Road” by Wang[8] showed promising results for shadow removal using non-linear SVM classifier. Proposed model was exclusively for traffic images and the model was not good for the large sized shadows in the scene. The model performs better for small sized shadows in the traffic images. Accuracy is used as the performance evaluation measure and achieved an accuracy of 83 % on UCF dataset. They improved the accuracy of the previous work entitled ”Shadow detection and removal for illumination consistency on the road”[9] by introducing adaptive variable scale regional compensation operator to remove the shadows.

Introduction of Generative Adversarial Networks[1] made drastic changes in the computer vision researches. Already implemented models migrated to GAN architectures as extension to their previous works to improve the results. Yun [10] Introduced GAN for shadow removal as an extension to the previous work entitled "Shadow Detection and Removal From Photo-Realistic Synthetic Urban Image Using Deep Learning" and improved the performance. However, traditional modeled methods have limited ability to remove shadows when irregular illumination or objects with various colors are present.

In order to address the color inconsistencies in the shadowed region, Xiaodong Cun[11] proposed a novel network structure called dual hierarchically aggregation network (DHAN) which contains a series of convolutions as the backbone without any down-sampling and hierarchically aggregate multi-context features for attention and prediction for successful shadow removal. The model proposed in the work entitled " Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN" shows RMSE = 5.76 , PSNR = 34.48 dB for ISTD dataset and RMSE = 5.46, PSNR = 33.72 for SRD dataset respectively.

The shadow removal phase is open to enhancements and not yet saturated. GAN based models show significant improvements in generating shadow-less images. Conditional Generative Adversarial Networks(cGAN)[12] are very helpful for narrowing down the generated image space of the generator and thereby reducing the time for training the model. We can modify GAN architecture according to our needs and based on the inputs to the generator and the discriminator, the behaviour of the GAN changes significantly and produces good results.

3 Methodology

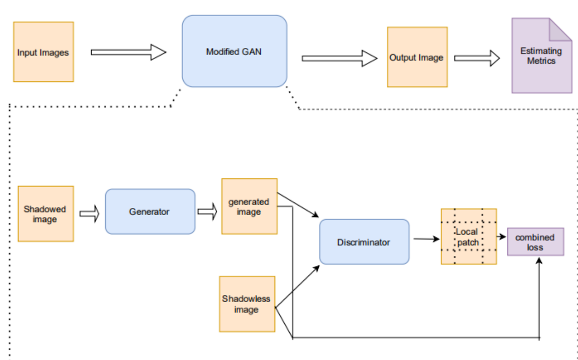

Fig. 2. Block diagram of proposed approach.

Fig. 2 Illustrates the proposed architecture for the shadow removal task. Shadowed image given to the generator module and produces the shadow-less (generated image)version of the input image. The discriminator module takes the paired image containing the generated shadow-less image and the real shadow-less image. The duty of the discriminator is to check whether the paired image is real or fake. The generator module trained in such a way that to minimize the loss between the expected target image and generated image and fool the discriminator module.

3.1 Dataset Description

Since we are more focused on the shadow removal task, ISTD[4] shadow removal bench mark dataset is used. Dataset contains shadowed images, shadow mask, shadow-less images. Train data - 1330 images (640 * 480 pixels) Test data - 540 images(640 * 480 pixels)

Fig. 3. Architecture of Generator.

During the data preprocessing phase, the images are loaded and re-scaled to (256*256) pixels for processing convenience and converted to numpy array

3.2 Architecture of Generator

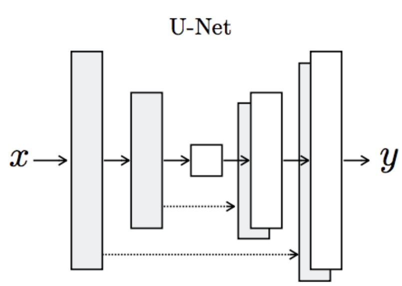

Fig. 4. Sample ISTD dataset .

Fig. 4 Shows the overall abstract architecture of the Generator. The generator is an encoder-decoder model using a U-Net architecture[13]. The encoder and decoder of the generator contain convolutional, batch normalization, dropout, and activation layers. The encoder encodes the information in the given input image and the context information generated in the bottleneck is used for reconstructing the image using the decoder block. Skip connections are used between corresponding encoder and decoder layers for improving the quality of the image generated by the generator.

3.3 Architecture of Discriminator

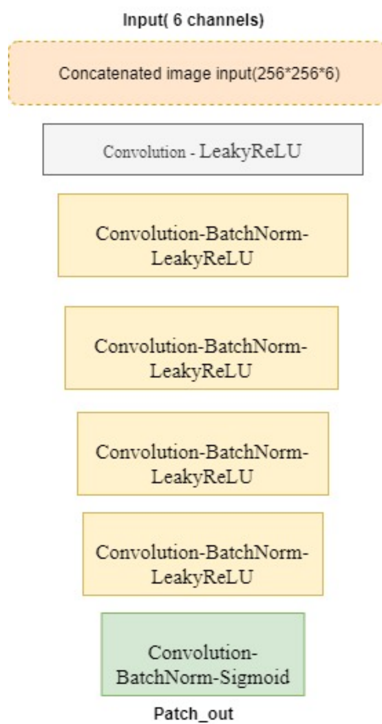

Fig. 5. Architecture of Discriminator.

Fig. 5 Illustrates the architecture of local patch discriminator. The discriminator is a deep convolutional neural network that performs image classification. Both the source image(generated image) and target image(shadow-less image) given as the input to the discriminator and check the likelihood of whether the shadow-less image is real or a translated version of generated image. The discriminator model is trained in the same way as a traditional GAN model.

Adversarial training is used for training the discriminator model. The training of generator is too slow compared to the discriminator. This will lead to the issues in the GAN such as Vanishing gradient, Mode collapse, Non - convergence. The obvious solution is to balance generator & discriminator training to avoid overfitting.

Combined Loss(1) introduced to balance the training of Generator and Discriminator. Giving more importance to the reconstruction loss(2) rather than the adversarial loss(3) during the Generator training will reduce the effect of the fast training of Discriminator on the Generator.

$$Combinedloss = \lambda 1 * reconstructionloss + \lambda 2 * adversarialloss \tag{1}$$

Where,
$\lambda 1, \lambda 2 = 100, 1$
Loss function between the generated fake image and the shadow-less image is called as the reconstruction loss.

$$Lreconstruction = \sum_{i=1}^{n} |Ytrue - Ypred| \tag{2}$$

Where,
Ypred = The predicted value for the ith pixel.
Ytrue = The observed(actual) value for the ith pixel
n = Total number of pixels.
Adversarial loss (Binary Cross Entropy loss)

$$Ladversarial(G, D) = E_{xy}[logD(x, y] + E_{xy}[log(1 - D(x, G(x, y)))] \tag{3}$$

Where,
G - Generator , D - Discriminator
x - shadow-less image , y - shadowed image

4 Results and Discussion

Table 1. Comparing with the existing models

Sno.	**Model.**	**RMSE**	**PSNR(dB)**
1	ST-GANl [4]	7.47	....
2	RIS-GAN[5]	6.67	....
3	Our Model	1.997	53.38

In this project, ISTD dataset is used to train our model and the metrics RMSE, PSNR are calculated. The state of the art models RIS-GAN[5], Stacked

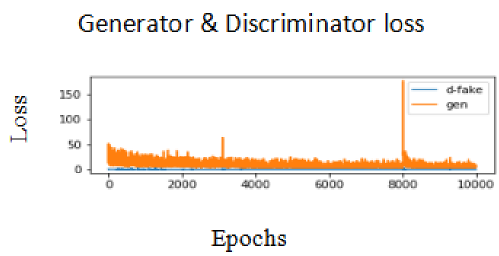

Fig. 6. Architecture of Discriminator.

Conditional GAN(ST-GAN)[4] used as the base models for comparing the performance of our model(Table 1). The training graph based on the generator loss and discriminator loss is shown in figure 6.

From the training graph, it is clear that the generator and discriminator training is like a min-max game. The generator tries to fool the discriminator and at the same time, the discriminator try not to fool by the generator's fake image. The loss never become same for the generator and discriminator for a good GAN[1] model.

Fig. 7. sample output. source, generated, expected images are arranged row wise.

From Fig.7,we can see that our model performs well for the indoor images and the outdoor images. First and third column corresponds to the outdoor images. RMSE = 0.045 and PSNR = 78.76dB for these two images. Second column corresponds to the indoor image. It is an image of a black board in a classroom. The shadow successfully removed by our model. RMSE = 0.08 and PSNR = 74.58db for this particular indoor shadowed image sample.

Fig.8 illustrates the sample output of the proposed model on the sample images which are entirely different from the training dataset. First row corresponds to an indoor shadowed image and corresponding output of our model.

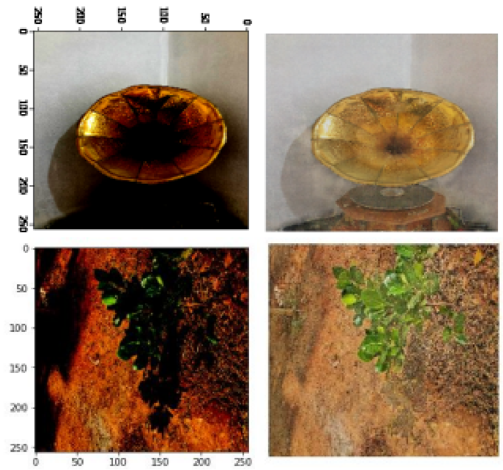

Fig. 8. output of the proposed model on real-world sample images.

Second row corresponds to a outdoor shadowed image from a real life instance and corresponding shadow-less image produced by our model.

It is clear that, proposed model performs well in the shadowed images outside the training dataset and it is evident in the shadow-less images produced by the proposed model.

5 Conclusion

In this project, we proposed a GAN based shadow removal model for generating enhanced shadow-less images. Initially, we use basic conditional GAN model on ISTD dataset and analyzed the areas of improvements. Secondly, we modified the architecture and introduced a combined loss for training the model. Tuned the parameters by conducting repeated experiments and identified the appropriate set of parameters for the model. From the experiments, it was seen that the proposed model performs better than the existing model. The model showed promising results on the shadowed(outdoor and indoor) real-time images collected by us.

As a future enhancement, shadow-less image generated by the proposed model can be improved further by using image super-resolution techniques. Parameter tuning is a time-consuming task and also requires more domain knowledge. In addition to future enhancement, Neuroevolution techniques[14] can be used to tune the hyper-parameters to identify the best set of parameters for the model and thereby improving the performance of our model. The enhanced model can be used for improving the efficiency of the object detection and tracking applications, especially for the applications like Wild animal detection and recognition from aerial videos using computer vision technique[15] in which the presence of shadow is unavoidable.

References

1. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, International Conference on Neural Information Processing Systems, 2014, https://arxiv.org/abs/1406.2661
2. S Veni, Anand, R., and Santosh, B., “Road Accident Detection and Severity Determination from CCTV Surveillance”, in Advances in Distributed Computing and Machine Learning, Singapore, 2021.
3. Sujee R. and Thangavel, S. Kumar, “Plant Leaf Recognition Using Machine Learning Techniques”, New Trends in Computational Vision and Bio-inspired Computing: Selected works presented at the ICCVBIC 2018, Coimbatore, India. Springer International Publishing, Cham, pp. 1433–1444, 2020.
4. Jifeng Wang, Xiang Li2, Jian Yang, “Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal”,CVPR 2017, https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00192
5. Ling Zhang,1 Chengjiang Long,2 Xiaolong Zhang,1 Chunxia Xiao3, “RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal”, AAAI 2020,https://doi.org/10.1609/aaai.v34i07.6979
6. L. Qu, J. Tian, S. He, Y. Tang and R. W. H. Lau, ”DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017,https://ieeexplore.ieee.org/document/8099731
7. Sidorov, Oleksii, “Conditional GANs for Multi-Illuminant Color Constancy: Revolution or Yet Another Approach?”,The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018,https://arxiv.org/abs/1811.06604
8. Wang, H. Xu, Z. Zhou, L. Deng and M. Yang, ”Shadow Detection and Removal for Illumination Consistency on the Road,” in IEEE Transactions on Intelligent Vehicles, 2020, https://ieeexplore.ieee.org/document/9068460
9. Wang, L. Deng, Z. Zhou, M. Yang and B. Wang, ”Shadow detection and removal for illumination consistency on the road, ”https://ieeexplore.ieee.org/document/8304275
10. Yun, Heejin Kang Jik, Kim Chun, Jun-Chul. “Shadow Detection and Removal From Photo-Realistic Synthetic Urban Image Using Deep Learning”, Computers, Materials Continua2019, https://www.techscience.com/cmc/v62n1/38123
11. Xiaodong Cun, Chi-Man Pun, Cheng Shi, “ Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN” AAAI 2020, https://arxiv.org/abs/1911.08718
12. M. Mirza, S. Osindero, ”Conditional Generative Adversarial Nets”, In Arxiv 2014. http://arxiv.org/abs/1411.1784
13. Olaf Ronneberger, Philipp Fischer, Thomas Brox,”U-Net: Convolutional Networks for Biomedical Image Segmentation”MICCAI 2015, https://arxiv.org/abs/1505.04597
14. K. Sree and Dr. Jeyakumar G., “An Evolutionary Computing Approach to Solve Object Identification Problem for Fall Detection in Computer Vision-Based Video Surveillance Applications”, Journal of Computational and Theoretical Nanoscience, vol. 17, no. 1, pp. 1-18, 2020. https://doi.org/10.1166/jctn.2020.8687
15. M. K. and Dr. Padmavathi S., “Wild animal detection and recognition from aerial videos using computer vision technique”, International Journal of Emerging Trends in Engineering Research, vol. 7, no. 5, pp. 21-24, 2019.