

Article

Throwaway Shadows Using Parallel Encoders Generative Adversarial Network

Kamran Javed ^{1,*}, Nizam Ud Din ², Ghulam Hussain ² and Tahir Farooq ³

¹ National Centre of Artificial Intelligence (NCAI), Saudi Data and Artificial Intelligence Authority (SDAIA), Riyadh 11543, Saudi Arabia

² College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, Korea; nizam@skku.edu (N.U.D.); hussain@skku.edu (G.H.)

³ Leverify LLC, 16301 NE 8th Street Suite 206, Bellevue, WA 98008, USA; Tahir@leverify.com

* Correspondence: kamran.javed@giki.edu.pk

Abstract: Face photographs taken on a bright sunny day or in floodlight contain unnecessary shadows of objects on the face. Most previous works deal with removing shadow from scene images and struggle with doing so for facial images. Faces have a complex semantic structure, due to which shadow removal is challenging. The aim of this research is to remove the shadow of an object in facial images. We propose a novel generative adversarial network (GAN) based image-to-image translation approach for shadow removal in face images. The first stage of our model automatically produces a binary segmentation mask for the shadow region. Then, the second stage, which is a GAN-based network, removes the object shadow and synthesizes the effected region. The generator network of our GAN has two parallel encoders—one is standard convolution path and the other is a partial convolution. We find that this combination in the generator results not only in learning an incorporated semantic structure but also in disentangling visual discrepancies problems under the shadow area. In addition to GAN loss, we exploit low level L1, structural level SSIM and perceptual loss from a pre-trained loss network for better texture and perceptual quality, respectively. Since there is no paired dataset for the shadow removal problem, we created a synthetic shadow dataset for training our network in a supervised manner. The proposed approach effectively removes shadows from real and synthetic test samples, while retaining complex facial semantics. Experimental evaluations consistently show the advantages of the proposed method over several representative state-of-the-art approaches.

Keywords: shadow removal; image restoration; image reconstruction; partial convolution



Citation: Javed, K.; Ud Din, N.; Hussain, G.; Farooq, T. Throwaway Shadows Using Parallel Encoders Generative Adversarial Network. *Appl. Sci.* **2022**, *12*, 824. <https://doi.org/10.3390/app12020824>

Academic Editor: Pietro Zanuttigh

Received: 11 October 2021

Accepted: 10 January 2022

Published: 14 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Facial images have become one of the most popular sources of images captured daily, transmitted through electronic media and/or shared on the social networks. In the real world, these images are often corrupted by some image conditions, especially the shadows of different objects. This not only degrades image quality but also affects the visual appearance of the image. The main objective of this research is to automatically detect and remove the shadow of an object from the facial images and produce a shadow free image. Most of the previous shadow removal works deal with removing shadow from the scene images and to the best of our knowledge there is no previous work for shadow removal from the facial images. Since faces have a complex semantic structure, shadow removal from facial images is an extremely challenging problem in computer vision.

Traditional shadow removal works [1,2] by using a physical model. These non-trivial methods take lot of processing time and suffer for shadow removal in facial images. On the other-hand, learning-based methods [3–6] outperformed non-learning based methods for the shadow removal task. Although they produce good results as compared to the

traditional algorithms for removing shadow from scene images, they are unable to remove shadows from facial images due to the complex nature of face semantics.

In this work, instead of improving or modifying the previous deep learning based shadow removal model for removing shadow from face images, we took a totally different approach by using an image inpainting approach as a shadow removal model. Image inpainting is the method used to reconstruct lost or damaged parts of an image. The current state-of-the-art deep learning based image inpainting methods [7–11] reconstruct the damaged region by providing the mask of the damaged part. Refs. [7,8] fill the missing pixel by copying similar patches from the surrounding region. Ref. [10] used some guidance information to reconstruct the corrupt part of the image while [12] use two discriminators to enforce global coherency. Some models [9,11,13] use two-stage networks by generating coarser results in the first stage and refine it in the second stage.

All of the above mentioned deep learning works only use standard convolution as the backbone operation of a neural network. The standard convolution employs the same filter weights all over the image, nevertheless pixels are valid or affected. Eventually, it generates a well-incorporated structure under the affected area but fails to remove visual artifacts, particularly at the boundary of the affected and valid area as mentioned in [14,15]. In the shadow removal problem, this issue becomes more severe because most of the time the shadow area is large and an irregular shape. To overcome these issues, many researchers use extensive post processing steps and/or additional refinement stages as in [8,11]. To incorporate with irregular shape recovery and limitations of standard convolution, improved convolution—called partial convolution [14]—is proposed.

In partial convolution, convolution is only employed on a valid pixel area and is re-normalized. A segmentation mask is used to locate a valid pixel area [14]. The valid pixel mask is updated after each iteration to compute new valid pixels. Additionally, our approach does not have any post processing or refinement stages. In this paper, we are considering the shadow part of an image as a damaged or corrupted area. The first segmentation mask of a shadow is generated by a simple convolution auto-encoder network and then use that mask of the shadow part along with input image, to reconstruct the shadow part of the image. We propose a GAN based deep network that takes an input image along with the binary mask of the shadow region and produces a shadow free image that is consistent both visually and structurally. The main contributions of this work are summarized as follows:

- We propose a novel GAN-based image inpainting approach to remove the shadows of objects from facial images;
- Our method generates a well-incorporated semantic structure and disentangles the visual discrepancies issue under the shadow region by employing a combined parallel operation of standard and partial convolution in a single generator model;
- To train our shadow removal network in a supervised manner, we create a paired synthetic shadow dataset using facial images from the CelebA dataset;
- Our model removes the shadow and creates perceptually better outputs with fine details in challenging facial images.

The remaining parts of the paper are organized as follows. Section 2 covers related works. The architecture of the shadow removal network is described in Section 3. Section 4 covers the experimental setting. Section 5 details the results and discussion.

2. Related Work

Generative Adversarial Network (GAN): GANs have shown a promising ability for image generation problems [16]. GAN is a two network model; one is a generator network and other is a discriminator network. The purpose of the generator network is to learn a given data distribution, where the intention of the discriminator network is to estimate the probability that a given sample is real or fake, that is, generated from generator network. GAN uses adversarial training, where the generator and the discriminator networks train alternatively. One popular improvement is to use multiple stages of GAN. Zhang et al.

proposed Stacked Generative Adversarial Networks (StackGAN) [17]. It is a two-stage GAN network to produce a high resolution output from the text description. The first stage GAN network generates low resolution results from the given text description and the output image of Stage-I GAN along with the input text. Then the second stage is fed with the results of the first stage and produce high resolution photo realistic images with fine details. GAN has a proven powerful solution to generate natural looking results [18]. Due to its success for various tasks, GANs are widely used for problems such as domain translation [19–21], texture synthesis [22,23], image inpainting [7,8,11,12,21,24–26] and shadow removal [6,27–30].

Image inpainting: The goal of inpainting is to recover the missing part in an image. There are countless applications of inpainting, from removing undesired objects and restoring corrupted regions to adding specific objects. Traditional inpainting approaches propagate the information to fill-in the corrupted area from neighboring pixels [31,32]. All of these methods can only fill-in a small area with stationary texture and failed to inpaint areas where texture and color variance is large. To overcome the texture issue, patch based methods are introduced, which copy similar patches from the input image and paste it into the target image [33,34]. However, this approach works well for non-stationary texture inpainting. Patch based methods have a computational cost because they search in an iterative manner, which is inefficient for real-time implementation.

A pioneer deep-learning based image inpainting method is proposed in [7], which can inpaint the large missing region conditioned on its neighbouring information. The combination of pixel-wise loss and adversarial loss is used for training. However, high frequency details are missing and sometimes generate artifacts in the output images. For better perceptual results, structural inpainting [8], which is based on [7], used perceptual reconstruction loss in addition to existing loss. Structural inpainting can inpaint the complex structures. Khan et al. [13] proposed two stage GAN to de-pixelate the mosaic face image. Their network first removes the mosaic part in the image and then generates face semantics. It works in the coarse-to-fine manner. For better perceptual results, Ref. [35] proposed UMGAN with perceptual loss from the pre-trained network. Refs. [13,35] are limited to square-shape corrupted areas only, but shadows can be an irregular shape. Similar to [7,8,24], we exploit both low-level (l_1) loss and high-level (*SSIM*) loss in terms of reconstruction loss to inpaint the region under the shadow.

A two-stage network called EdgeConnect [10] is proposed to inpaint the corrupted image by employing the hallucinated edge information of the corrupted region. Since the results of EdgeConnect rely on the quality of the edge map produced by the edge generator network, so the output suffers when the edge generator network failed to produce a right edge map. New convolution schemes, such as partial convolution [14] and gated convolution [15], were developed to overcome the limitations of the aforementioned methods. These methods produce better results in terms of color correspondence and incorporated semantics.

Object removal: The exemplar-based method for texture synthesis is proposed by Criminisi et al. [36]. It inpaints the missing area with plausible texture but fails to generate reasonable results for the regions which do not have similar patches in the image. Improved exemplar based inpainting methods are described in [37] to remove an object from a single image. Normalized cross correlation along with the summation of squared differences is used to find a matching patch in the image. However, it removes the object accurately in simple scenes but the boundary of the removed region has some artifacts. Kamran et al. [9] proposed a two-stage GAN based neural network to remove a microphone object in facial images. It can efficiently remove small objects like a microphone and recover semantics under that but struggle to recover a large area. Recently, Din et al. [38,39] proposed a GAN-based network to effectively remove a large occluded object from facial images.

Shadow removal: Shadow removal is one of the popular topics in the computer vision field nowadays, where the goal is to remove shadows from photographs which were taken on a sunny day. Ding et al. [29] proposed a robust attentive recurrent GAN based network

to detect and remove shadows. Their approach is able to remove shadows from complex scene images. The model is very flexible to incorporate sufficient unsupervised shadow images to train a powerful model. As compared to conventional approaches, which uses an illumination model to remove shadows, Ref. [27] proposed a deep neural network, which accurately and automatically estimates the parameters for the model from a single image.

Mask-Shadow GAN, presented in [30] uses un-paired images to remove shadows from scene images. Instead of shadow-free to shadow translation, Mask-Shadow GAN is a deterministic image translation technique, which uses shadow masks as a guidance, which are automatically learned from the real world images with shadow. RIS-GAN is proposed in [28], and exploits the residual and illumination. They explored the correlation between residual, illumination and the shadow by using a unified end-to-end framework. One recent work [6] proposed a method, which first aggregates with context using an aggregation model and then hierarchically aggregates the attentions and features. A shadow matting generative network is trained to generate the shadow images from the corresponding shadow-free images and masks. The shadow matting generative network not only enlarges the scenes in the shadow database but also reduces the color discrepancies.

3. Our Method

This section describes the network architecture of the proposed shadow removal method and the details of the objective function we used for training. Our network consists of two stages; in the first stage, we used a convolution auto-encoder to detect the shadow of an object. In the second stage, we used a GAN based image-to-image translation method, which effectively removes shadows in facial images and produces fine details. Figure 1 shows the overall shadow removal architecture.

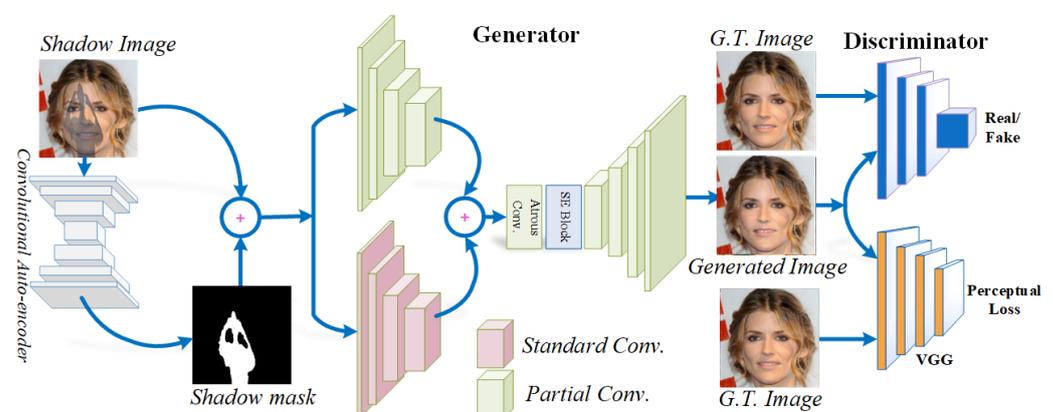


Figure 1. Proposed Network Architecture for Shadow Removal.

3.1. Network Architecture

The first stage of our network (Convolutional Auto-encoder) consists of a CNN-based encoder and decoder architecture. The encoder consists of five layers where each layer consists of a convolution layer followed by an activation function (*Lrelu*) and an instance normalization layer, except the first layer. The decoder architecture is a mirror copy of the encoder architecture except that convolution is replaced by a deconvolution layer. The convolutional auto-encoder takes the input shadow image and produces a binary segmentation mask for the object's shadow. We used a cross-entropy loss as an objective function between the predicted binary segmentation mask and the corresponding target segmentation map.

Since the second stage (shadow removal) of our network utilizes a GAN based model, it has generator and discriminator networks. The generator network has two parallel encoders; one is a standard convolution path and the other is a partial convolution path. This combination in the generator results not only in learning incorporated semantic structures but also disentangling the visual discrepancies problem under the shadow

area. We start with a UNET-like architecture [40], which has skip connections between the standard convolution encoder and decoder. The purpose of the skip connections is to provide super highways for a gradient during back propagation and avoid the vanishing gradient problem.

Additionally, we used an atrous convolution [41] layer and a squeeze and excitation block [42] between the encoders and the decoder networks of the generator. The intention of atrous convolution is to capture a large field of view for a semantically coherent output and reduce the trainable parameters. To increase the representational power of our architecture, we used a squeeze and excitation block followed by atrous convolution. The purpose of the squeeze and excitation block is to perform dynamic channel-wise feature re-calibration. Moreover, a decoder network is similar to the standard convolution encoder, except transpose convolution is used instead of convolution. Each convolution layer consists of relu+conv+instant norm operation. Our discriminator is a Patch-GAN based architecture, which penalizes the patches instead of each pixel.

3.2. Objective Function

To enforce the generator to remove shadow and produce realistic and perceptually correct content under the shadow, we used a joint objective function, which is a combination of four different loss terms. The overall training objective function can be written as follows:

$$\mathcal{L}_{obj} = \alpha \cdot (\mathcal{L}_{l_1} + \mathcal{L}_{ssim}) + \mathcal{L}_{adv} + \beta \mathcal{L}_{perc}, \quad (1)$$

where \mathcal{L}_{l_1} is a pixel level l_1 penalty, \mathcal{L}_{ssim} is a structural penalty, \mathcal{L}_{adv} is a cross entropy adversarial loss and \mathcal{L}_{perc} is a perceptual penalty, which we calculated by measuring the distance between feature map values of the loss network for the generator output and the corresponding ground truth. Particularly, we used a pre-trained VGG-19 [43] as a loss network. The weight of loss terms can be adjusted with respective constants α and β .

4. Experimental Setup

In this section, we present the experimental setting of our shadow removal network. First, we created a synthetic shadow database and then trained our shadow removal network on it. For fair comparison, we retrained state-of-the-art works such as Edge-Connect [10], Partial Convolution [14], Gated Convolution [15] and Ghost-free Shadow removal [6] on a new synthetic database. At evaluation time, we also show results on real world shadow images, collected from the internet.

Database: We trained our shadow removal network in a supervised manner. We started with 20,000 randomly selected images from the CelebA Face dataset [44] and created a synthetic shadow database. CelebA face images contain various celebrity images with wild backgrounds and were taken in diverse conditions. We used OpenFace dlib [45] to align the faces using facial landmark positions. This alignment helps the model to generate the face semantics (e.g., eyes) at the right place on the face. Finally, we generate synthetic images by placing the shadow of various objects using Adobe Photoshop. We consider shadows of various objects of different sizes and scales and placed them at various positions in the face image. Corresponding shadow mask images were also created to train the shadow detection stage of our network. Compared to the shadow dataset created by [46,47], we focus on creating shadow images that contain shadows of various objects instead of producing a relit image with hard cast shadows.

Training setting: The convolutional auto-encoder network is fed with an input shadow image and generates a binary map of the object's shadow in the input image. The generator of our shadow removal network then takes the pair of input shadow image and its corresponding mask generated by the convolutional auto-encoder and produces an output image without shadow. While the job of discriminator is to differentiate between the generated and ground truth images without shadows. We trained our network with a joint objective function as in Equation (1). The generator network produces an output face image without the shadow. The data split is 70% for training and 30% for testing. There

is no subject overlap between the training and testing sets; both have distinct images. We used the Adam optimizer [48] with learning rate 2×10^{-4} and momentum 0.5 to train the shadow removal network. Random crop and random flip techniques were used for data augmentation. We fed 10 images together in a batch. At the start, the discriminator trained quickly and the generator became weak. We first trained the generator network for one hundred epochs and then both the generator network and the discriminate network were trained for five hundred epochs to avoid this problem. We implemented our network on python using the TensorFlow platform [49]. Our training took around three days on an NVIDIA GeForce 1080Ti graphic card. The code with the pre-trained model will be published on GitHub after acceptance of this manuscript.

5. Comparison and Discussion

This section presents the quantitative and qualitative comparison of our shadow removal method with the state-of-the-art works on both real world shadow images and synthetic shadow images.

5.1. Visual Comparison for Facial Images

Figure 2 shows a comparison of our model results with current state-of-the-art representative methods such as EdgeConnect [10], Partial Convolution [14], Gated Convolution [15] and Ghost-free Shadow removal [6]. To make the comparison fair, we train these methods on our synthetic shadow database. Examples in the first couple of rows are the real test images (no ground truth), while the other two rows show the result for the synthetic test sample.

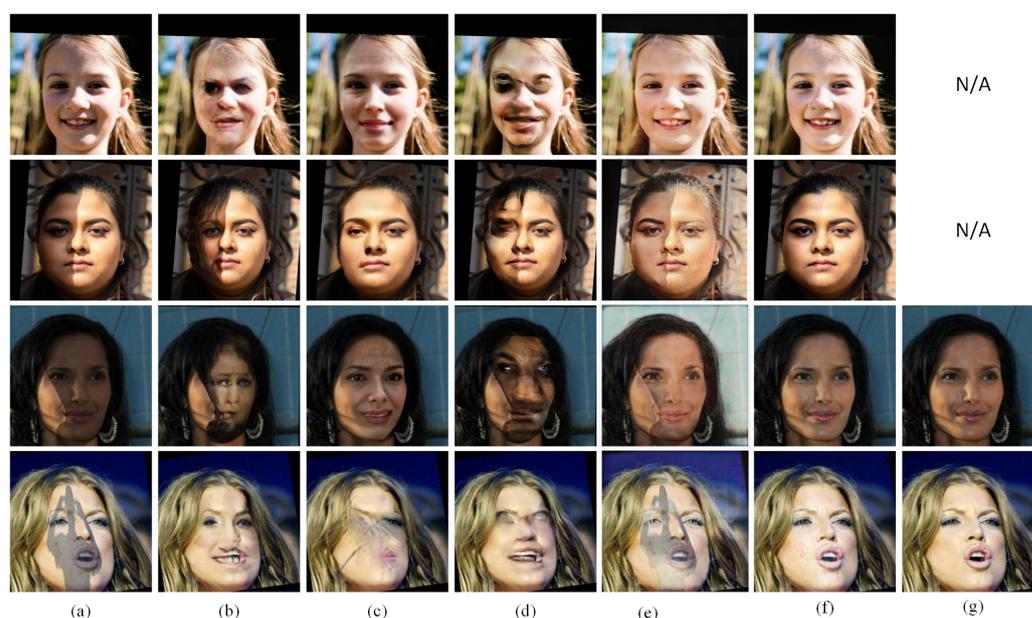


Figure 2. Visual comparison of shadow removal. (a) Input image, (b) EdgeConnect [10], (c) Partial Convolution, [14], (d) Gated Convolution [15], (e) Ghost-free Shadow removal [6], (f) Ours, (g) Ground truth. **Note:** There is no ground truth for the first couple of rows since these samples are real world shadow images collected from the Internet. The last two samples are from our synthetic database.

As can be seen in Figure 2, our technique plausibly removes shadows from the facial images for both complex real and synthetic test samples. On the other hand, all other representative methods struggle to produce reasonable results. EdgeConnect [10] struggles to produce a proper edge map for a large damaged region of the face resulting in artifacts. Partial Convolution [14] produces sharp results as compared to EdgeConnect and Gated Convolution but still shows artifacts especially at the borders of damaged and undamaged

regions. Ghost-free shadow removal [6] plausibly removes the shadow but is unable to produce natural looking face semantics due to the complex nature of the face semantics.

On the other hand, our model combines the benefits of vanilla and partial convolution encoders. This helps our model in removing the shadow and learning the well-incorporated and artifact-free semantics of the face under the shadow.

Figure 3 shows additional qualitative results of our model for complex and large size shadow samples in our synthetic database. The first column shows the input image, second and third columns show the segmentation map of the object shadow and shadow free image generated by our model, respectively. Last column represents the ground truth for input images. The results show that our model effectively removes different types of complex, large and challenging shadow occlusions from the face images. The last input sample contains shadows created by a lightning effect (not by an occluded object); thus, our model is unable to produce the accurate segmentation mask of the shadow region and a plausible shadow free output.

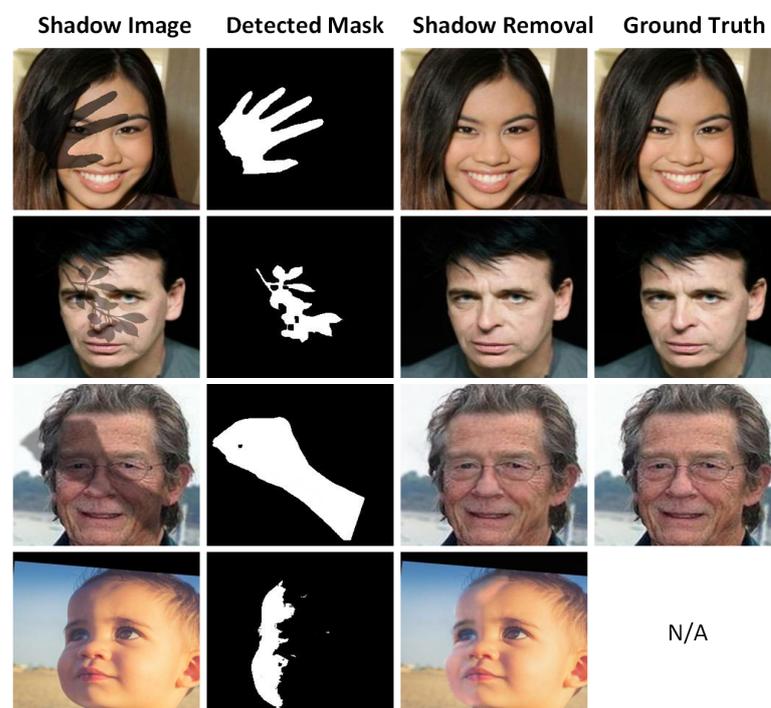


Figure 3. Additional qualitative results of our model for complex and large size shadow samples in our synthetic database.

5.2. Quantitative Evaluation

In this section, we describe a quantitative comparison of the proposed method with previous state-of-the-art methods such as EdgeConnect [10], Partial convolution [14], Gated convolution [15] and Ghost-free Shadow removal [6] in terms of Root Mean Square Error (RMSE), Naturalness Image Quality Evaluator (NIQE) [50] and Blind Referenceless Image Spatial Quality Evaluator (BRISQUE) [51]. NIQE and BRISQUE measure the naturalness of an image without any reference. Smaller NIQE and BRISQUE scores are better. To measure NIQE and BRISQUE, we used only generated images without providing corresponding ground truths. We have evaluated RMSE on the test images from our synthetic database, which has corresponding ground truths. Table 1 provides a quantitative comparison with previous methods such as EdgeConnect [10], Partial convolution [14], Gated convolution [15] and Ghost-free Shadow removal [6]. The table shows that for the shadow removal problem, the results of our shadow removal method are better than or comparable to those of the state-of-the-art methods.

Table 1. Quantitative comparisons of shadow removal in terms of Root Mean Square Error (RMSE), Naturalness Image Quality Evaluator (NIQE), and Blind Referenceless Image Spatial Quality Evaluator (BRISQUE).

Methods	RMSE	NIQE	BRISQUE
EdgeConnect [10]	24.73	4.429	37.01
Partial Conv [14]	22.41	4.248	38.60
Gated Conv [15]	19.00	4.614	36.44
Ghost-free Shadow removal [6]	29.44	4.190	41.30
Ours	13.91	4.005	37.93

5.3. Results for Scene Images

To check the effectiveness of our model for removing shadows in scene images, we trained our proposed model on the publicly available ISTD dataset [4] for removing shadows in scene images. The ISTD dataset consists of 1870 training samples from 135 scenes and 540 test samples from 135 scenes. Our model effectively removed shadow not only in the facial images but also in the scene images as shown in Figure 4. It has potential applications in outdoor photography and surveillance by removing undesired shadows from the images.

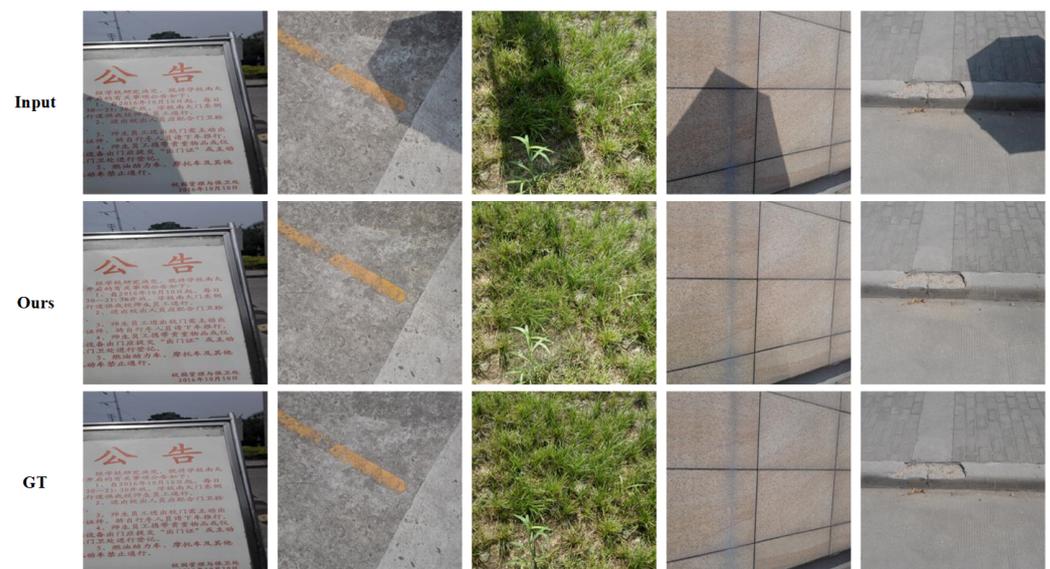


Figure 4. Shadow removal results of our proposed method on the scene images from ISTD dataset [4].

6. Conclusions

Our shadow removal approach is a GAN based image-to-image translation, which effectively removes shadow in facial images. In this work, we advocate a novel technique for automatically detecting and removing object shadows in facial images. To train our model in a supervised manner, we have created a paired synthetic shadow database. Our method not only generates well-incorporated semantic structures but also disentangles the visual discrepancies problem under the shadow area by employing combined parallel encoders of standard and partial convolution in a single generator model. The performance of our shadow removal method on real world shadow images is adequate although we trained the model using our synthetic shadow database. In the future, we are planning to expend our shadow removal work to automatically detect and remove shadows from lighting effects and occluded objects.

Author Contributions: K.J. developed the method; N.U.D. performed the experiments; K.J., N.U.D. did the analysis; and N.U.D., G.H., T.F. and K.J. wrote the paper. K.J., T.F. and N.U.D. proof read the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The APC for this research was funded by National Centre of Artificial Intelligence (NCAI) at Saudi Data and Artificial Intelligence Authority (SDAIA), Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Not applicable.

Abbreviations

We used following abbreviations in our manuscript:

GAN	Generative Adversarial Networks
SSIM	StructuralSIMilarity
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator
RMSE	Root Mean Square Error
NIQE	Naturalness Image Quality Evaluator
SE	Squeeze and Excitation block

References

1. Barrow, H.; Tenenbaum, J.; Hanson, A.; Riseman, E. Recovering intrinsic scene characteristics. *Comput. Vis. Syst.* **1978**, *2*, 2.
2. Finlayson, G.D.; Hordley, S.D.; Lu, C.; Drew, M.S. On the removal of shadows from images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *28*, 59–68. [[CrossRef](#)]
3. Qu, L.; Tian, J.; He, S.; Tang, Y.; Lau, R.W. Deshadownet: A multi-context embedding deep network for shadow removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4067–4075.
4. 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, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1788–1797.
5. 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 Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 816–832.
6. Cun, X.; Pun, C.M.; Shi, C. Towards Ghost-Free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 10680–10687.
7. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
8. Vo, H.V.; Duong, N.Q.; Pérez, P. Structural inpainting. In Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference, Seoul, Korea, 22–26 October 2018; ACM: New York, NY, USA, 2018; pp. 1948–1956.
9. Khan, M.K.J.; Ud Din, N.; Bae, S.; Yi, J. Interactive removal of microphone object in facial images. *Electronics* **2019**, *8*, 1115. [[CrossRef](#)]
10. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.; Ebrahimi, M. EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. *arXiv* **2019**, arXiv:1901.00212.
11. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 28–23 June 2018.
12. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph. (ToG)* **2017**, *36*, 1–14. [[CrossRef](#)]
13. Javed, K.; Din, N.U.; Bae, S.; Yi, J. Image unmosaicing without location information using stacked GAN. *IET Comput. Vis.* **2019**, *13*, 588–594. [[CrossRef](#)]
14. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
15. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 4471–4480.
16. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; Curran Associates Inc.: Montreal, QC, Canada, 2014; pp. 2672–2680.

17. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Huang, X.; Wang, X.; Metaxas, D. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5907–5915.
18. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
19. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976.
20. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
21. Bae, S.; Din, N.U.; Javed, K.; Yi, J. Efficient Generation of Multiple Sketch Styles Using a Single Network. *IEEE Access* **2019**, *7*, 100666–100674. [[CrossRef](#)]
22. Li, C.; Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 702–716.
23. Jetchev, N.; Bergmann, U.; Vollgraf, R. Texture synthesis with spatial generative adversarial networks. *arXiv* **2016**, arXiv:1611.08207.
24. Liu, P.; Qi, X.; He, P.; Li, Y.; Lyu, M.R.; King, I. Semantically Consistent Image Completion with Fine-grained Details. *arXiv* **2017**, arXiv:1711.09345.
25. Yang, C.; Lu, X.; Lin, Z.; Shechtman, E.; Wang, O.; Li, H. High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
26. Yeh, R.A.; Chen, C.; Yian Lim, T.; Schwing, A.G.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with deep generative models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5485–5493.
27. Le, H.; Samaras, D. Shadow removal via shadow image decomposition. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 8578–8587.
28. Zhang, L.; Long, C.; Zhang, X.; Xiao, C. Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. *arXiv* **2019**, arXiv:1911.09178.
29. Ding, B.; Long, C.; Zhang, L.; Xiao, C. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 10213–10222.
30. Hu, X.; Jiang, Y.; Fu, C.W.; Heng, P.A. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 2472–2481.
31. Bertalmio, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*; ACM Press/Addison-Wesley Publishing Co.: New York, NY, USA, 2000; pp. 417–424.
32. Ballester, C.; Bertalmio, M.; Caselles, V.; Sapiro, G.; Verdera, J. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.* **2001**, *10*, 1200–1211. [[CrossRef](#)] [[PubMed](#)]
33. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*; ACM: New York, NY, USA, 2001; pp. 341–346.
34. Kwatra, V.; Essa, I.; Bobick, A.; Kwatra, N. Texture optimization for example-based synthesis. *ACM Trans. Graph. (ToG)* **2005**, *24*, 795–802. [[CrossRef](#)]
35. Javed, K.; Din, N.U.; Bae, S.; Maharjan, R.S.; Seo, D.; Yi, J. UMGAN: Generative adversarial network for image unmosaicing using perceptual loss. In Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 27–31 May 2019; pp. 1–5.
36. Criminisi, A.; Pérez, P.; Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **2004**, *13*, 1200–1212. [[CrossRef](#)]
37. Wang, J.; Lu, K.; Pan, D.; He, N.; Bao, B.K. Robust object removal with an exemplar-based image inpainting approach. *Neurocomputing* **2014**, *123*, 150–155. [[CrossRef](#)]
38. Din, N.U.; Javed, K.; Bae, S.; Yi, J. A novel GAN-based network for unmasking of masked face. *IEEE Access* **2020**, *8*, 44276–44287. [[CrossRef](#)]
39. Din, N.U.; Javed, K.; Bae, S.; Yi, J. Effective Removal of User-Selected Foreground Object From Facial Images Using a Novel GAN-Based Network. *IEEE Access* **2020**, *8*, 109648–109661. [[CrossRef](#)]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
41. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.

42. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation Networks. In Proceedings of the IEEE CVPR, Salt Lake City, UT, USA, 18–23 June 2018.
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the ICLR, San Diego, CA, USA, 7– May 2015.
44. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3730–3738.
45. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications*; Technical Report; CMU-CS-16-118; CMU School of Computer Science: Pittsburgh, PA, USA, 2016.
46. Hou, A.; Zhang, Z.; Sarkis, M.; Bi, N.; Tong, Y.; Liu, X. Towards High Fidelity Face Relighting with Realistic Shadows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 14719–14728.
47. Zhou, H.; Hadap, S.; Sunkavalli, K.; Jacobs, D.W. Deep single-image portrait relighting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 7194–7202.
48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. *OSDI* **2016**, *16*, 265–283.
50. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)]
51. Mittal, A.; Moorthy, A.K.; Bovik, A.C. Blind/referenceless image spatial quality evaluator. In Proceedings of the 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 6–9 November 2011; pp. 723–727.