

Improved Sketch-to-Photo Generation Using Filter Aided Generative Adversarial Network

Jyoti S. Raghatwan¹

Research Scholar, SKN College of Engineering & Asst.Prof., RMD
Sinhgad School of Engineering
Pune, India
jyotiraghatwan2@gmail.com,
<https://orcid.org/0000-0002-8091-9695>

Dr. Sandhya Arora²

Professor,
MKSSS's Cummins College of Engineering for Women,
Pune, India
sandhya.arora@cumminscollege.in
<https://orcid.org/0000-0002-6412-7831>

Abstract— Generating a photographic face image from given input sketch is most challenging task in computer vision. Mainly the sketches drawn by sketch artist used in human identification. Sketch to photo synthesis is very important applications in law enforcement as well as character design, educational training. In recent years Generative Adversarial Network (GAN) shows excellent performance on sketch to photo synthesis problem. Quality of hand drawn sketches affects the quality generated photo. It might be possible that while handling the hand drawn sketches, accidentally by touching the user hand on pencil sketch or similar activities causes noise in given sketch. Likewise different styles like shading, darkness of pencil used by sketch artist may cause unnecessary noise in sketches. In recent year many sketches to photo synthesis methods are proposed, but they are mainly focused on network architecture to get better performance. In this paper we proposed Filter-aided GAN framework to remove such noise while synthesizing photo images from hand drawn sketches. Here we implement and compare different filtering methods with GAN. Quantitative and qualitative result shows that proposed Filter-aided GAN generate the photo images which are visually pleasant and closer to ground truth image.

Keywords—Generative adversarial network, sketch-photo synthesis, filters, image to image translation.

I. INTRODUCTION

Main challenge in face sketch to photo synthesis is learning of high-level feature of target image. Sketch image generally contain simple information but not detail information about picture such as attribute like color, texture, shape etc. So, it is more difficult to generate exact attribute from sketch to generated image. Traditional methods [1] for photo/sketch synthesis used linear combination of similar training sketch/photo patches which include searching and weight computation techniques [2][3][4][5][6]. These techniques require more time. Model driven methods [7][8] require more effort to find out learning techniques. Researcher also explore deep learning approach for photo/sketch synthesis [9][10]. But traditional CNN methods fail to give expected result. It gives blurred output. Recently Generative Adversarial Network (GAN) [11] achieved good performance in image-to-image translation task. GAN consist of strong generator as well as discriminator network due to which it outperforms well in image-to-image translation task. Main aim of generator is generating the sample which look like as a real sample. And the job of discriminator is to identify generated samples are real or fake samples. Generator learn to generate more real sample during training phase to fool

the discriminator. Discriminator train directly on real and fake sample while generator is trained through discriminator.

In this paper we proposed filter aided GAN framework. Main idea behind this proposed work is to generate accurate and good quality images. As we know noisy image affect the performance of any image processing algorithm. To overcome that here we apply different filtering techniques on sketch before passing it to GAN. We analyze the performance of each filtering techniques like mean, median and bilateral filter. Contextual loss and pixel loss is used to improve performance of our GAN model.

II. RELATED WORK

In recent years different GAN architectures are proposed by researcher for sketch to photo synthesis. In [12] Author proposed contextual GAN which used joint image completion approach for completing incomplete sketches while generating photo images. Contextual GAN to learn the joint distribution of sketch and corresponding photo while training. They use two loss functions contextual loss and perceptual Loss to improve the result. But limitation of this method is there is no guarantee of identity preserving face generation and some attribute from input images are missing in output images. In [13] Author proposed framework where sketch with attributes is input to generated adversarial network. It improves the identity of generated faces.

In network architecture to reduce the no. of layers they used approach of skip connections and retained the network performance. Separate subnetwork is used to for attribute vector and sketch images to extract low level and high-level features information. They have used Adversarial loss and Reconstruction loss to calculate total loss. In [14] author proposed framework which generate multiple photo images of single sketch having unique attribute. Hybrid discriminator approach is used to predicts and distinguishes real and synthesized photos according to the set of desired attributes. They calculated adversarial loss and reconstruction loss, attribute classification loss, contrastive loss, content loss to improve the performance. In [15] Authors proposed a method which use facial composition information to help the synthesis of face sketch/photo. They propose framework called composition-aided generative adversarial network (CA-GAN) where sketch with pixel-wise face labels is input to generated adversarial network. During training phase hard-generated components and delicate facial structures are focused. Perceptual loss function is used to calculate perceptual similarity between synthesized image and generated image. In [16] Author Proposed high-fidelity face sketch-photo synthesis method using Generation Adversarial Network (GAN). It uses a deep residual U-Net as generator and a Patch-GAN with residual blocks as discriminator. They design effective loss functions. Low-level, high-level and edge feature restrictions are imposed using pixel loss, high level loss and edge feature loss. In [17] Authors implements skip connections in cycle-consistent generative adversarial network framework and proposed feature Encoder Guided Generative Adversarial Network (EGGAN) framework. Feature encoder basically used to improve quality of generated images by guiding in training phase. Feature loss and feature consistency loss is used to maintain identity information. In [18] Authors used CycleGAN framework and explore it to generate high resolution photo images using multi-adversarial networks. They focus on hidden layer of generator as well as additionally used Cycle consistency loss to generate better quality images. In [19] author proposed conditional cycle GAN, where condition is given on facial attributes such as skin color. This framework does not require paired data i.e., sketch and its ground truth photo for training. This framework retains the face style though some attributes get edited while synthesizing photo. In [20] author proposed Identity-Aware CycleGAN. Main focus of this proposed model is to improve identity recognition of generated faces by considering key attribute of faces like nose eyes. They used perceptual loss function to improve recognition accuracy. S. L. Bangare et al. [24, 25] worked in Machine learning, Image processing and IoT domain. S. D. Pande et al., [26, 27] presented the analysis and implementation of capsule network for various image processing applications and medicinal leaf classification. P. S. Bangare et al. [28] have proposed object detection work. N. Shelke et al. [29] worked with LRA-DNN approach. S. Gupta

et al. [30] has shown extraction related work. G. Awate et al. [31] applied the CNN methods.

III. PROPOSED SYSTEM

As shown in figure1 proposed system architecture first we pass sketch image to image filter. By experimental studies it is find out that when we directly passed sketch image to GAN, the sketch images that containing noise, generate low quality photo images. So, to overcome that as shown in above architecture before passing face sketch to GAN we apply different filtering techniques to reduce unwanted noise.



Figure1 Proposed System Architecture

We applied median filter, mean filter and bilateral filter and analyzed the performance of each filter by calculating Structural Similarity Index (SSIM) [11] of generated face photo from filtered sketch.

A. Median Filter

This is a nonlinear filter. In this technique each pixel value is replaced by median value of its neighbourhood pixels. In our experiment we used 5*5 kernel size for median filtering. salt and pepper noise, random noise is effectively removed by median filter. Noise pixel is effectively removed median filtering technique as they are very far from median. Median filter is very good at edge preserving but due to median concept fine details of images like line is removed. We demonstrate here working of median filter using 3*3 kernel size as below.

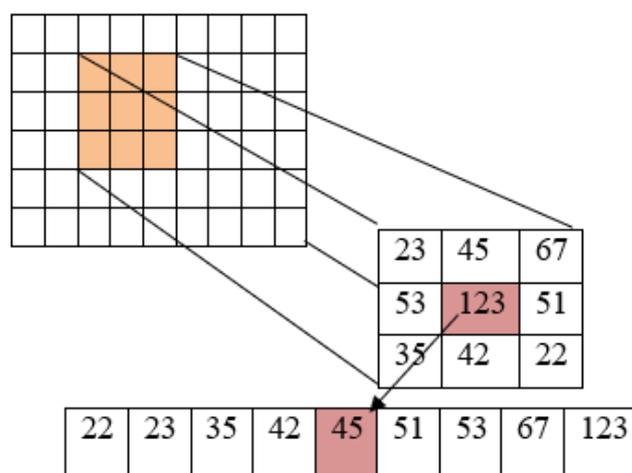
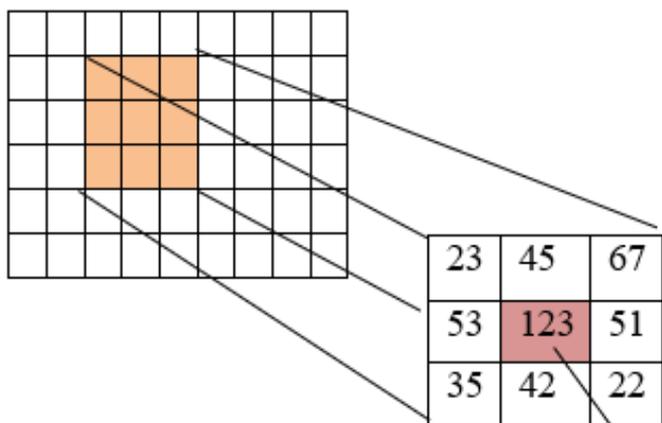


Figure 2 Median Filter

B. Mean Filter

This is a linear filter. In this technique each pixel value is replaced by mean value of its neighbourhood pixels. Similar to median filter, we used 5*5 kernel size for mean filtering. Mean filter reduces the random noise and main advantage of mean

filter is it retain sharpest step response. We demonstrate here working of mean filter using 3*3 kernel size as below.



$$23+45+67+53+123+51+35+42+22= 461/9 = 51$$

Figure 4 Mean Filter

C. Bilateral Filter

This is a nonlinear filter. Bilateral filtering done by replacing each pixel by average weighted intensity values of its neighbourhood pixels. Weights are depended upon radiometric differences as well as Euclidean distance of pixels. Sharp edges are preserves using bilateral filter.

The bilateral filter is represented as

$$BF(I)_x = \frac{1}{W_p} \sum_{y \in S} G_{\sigma_s}(\|x - y\|) G_{\sigma_r}(|I_x - I_y|) I_y \quad (1)$$

W_p is normalization term,

$$W_p = \sum_{y \in S} G_{\sigma_s}(\|x - y\|) G_{\sigma_r}(|I_x - I_y|) \quad (2)$$

Where:

x is current pixel of image to be filtered

S is x centered window

y is another pixel belonging to S

G_{σ_s} is spatial kernel, decreases coordinates difference

G_{σ_r} is range kernel, decreases intensities difference

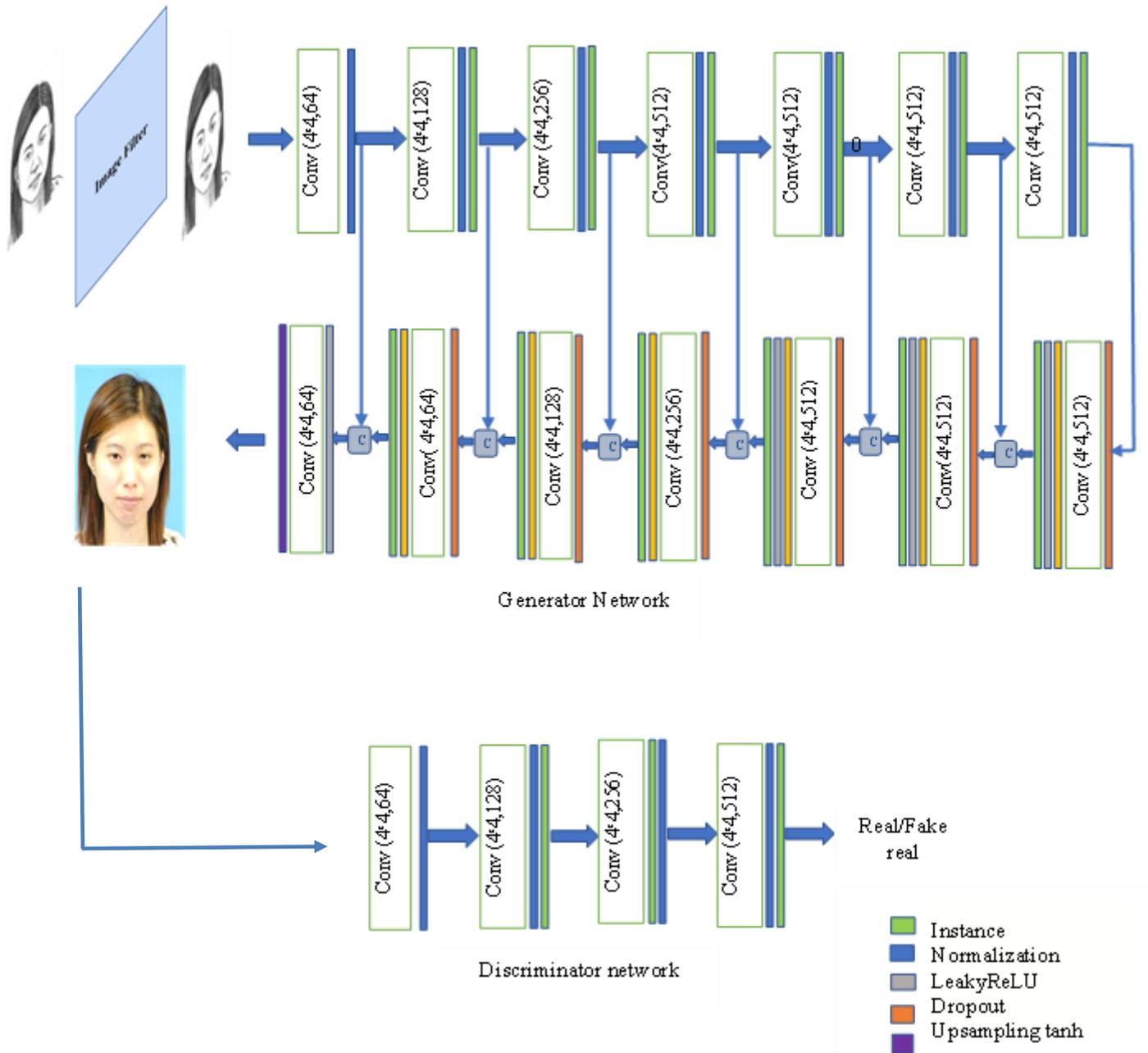


Figure 2 Detailed System Network Architecture

D. Network Architecture

Standard Generative Adversarial network, which consist of generator network and discriminative network as shown in figure2 is used.

Generator

Input to generative adversarial network is filtered sketch image of size $256 \times 256 \times 3$. Generator network consist of 7 convolution layers and 7 deconvolution layers which is described as C64-C128-C256-C512-C512-C512-DC512-DC512-DC256-DC128-DC64-C3 where Ck means C represent 4×4 convolution layer and k is no. of filters used and stride 2. Similarly, Dck means DC represent 4×4 deconvolution layer and

k represent no. of filters used and stride 2. LeakyReLU activation function is used after each convolution layer and ReLU activation function is used after each deconvolution layer except output layer where tanh activation function is used.

Discriminator

Discriminator network is used to distinguish between real and fake images. We train our discriminator on ground truth sketch image with their corresponding photo image. Discriminator network is build using 4 convolution layers which is described as C64-C128-C256-C512. Output of discriminator is either true or false. It predicts whether the image generated by generator is real or fake.

Loss Functions

Statistical distribution of training data is learned by GAN framework. In Our framework generator model generate photo image z for given input sketch image x . Generator model trained to generate the photo image which is no distinguishable from real photo images y and try to fool the discriminator. Discriminator learn to identify fake images generated by generator. The objective function is given as:

$$\mathcal{L}_{adv}(G) = E_{x,y \sim p_{data}}[\log D(x,y)] + E_{x,z \sim p_z}[\log(1 - D(x,G(x,z)))] \quad (3)$$

We calculate Pixel loss to retain low level features and contextual loss to retain low level features. Pixel loss is calculated as follows:

$$\mathcal{L}_{pixel}(G) = \frac{1}{N} \|\sum_{i=1}^N y - G(z)\| \quad (4)$$

Contextual loss is used to measure similarity between the features. Similar semantic region is compared using this function hence it overcome the need of aligned images. Contextual loss is calculated as follows:

$$\mathcal{L}_{contextual} = -\log(CX(y, G(z))) \quad (5)$$

Where CX represent feature similarity

The total generator loss is calculated by considering GAN loss and pixel losses and context loss as follows:

$$\mathcal{L}(G) = \lambda_1 \mathcal{L}_{adv}(G) + \lambda_2 \mathcal{L}_{pixel}(G) + \lambda_3 \mathcal{L}_{contextual}(G) \quad (6)$$

where we used $\lambda_1=1$, $\lambda_2=0.2$ and $\lambda_3=0.8$

IV. EXPERIMENTAL RESULT

Here we implement and compare different filtering technique to check performance of our proposed model.

A. Dataset and Training details

We used CUHK Face Sketch Database (CUFS) [21]. CUHK student data set consist of total 188 photo-sketch pairs. Among which testing set consist of 100 photo-sketch pairs and training set consist of 88 photo-sketch pairs. As size of training set is small, we perform data augmentation by applying different transformation techniques. We increase our training dataset upto 400 photo-sketch pairs which consist cropped faces, shifted and

rotated faces at different angles. Each input image of size 256×256 is provided during model training. For both discriminator and generator Adam optimizer [22] is used. Learning rate is set to 0.0002 and beta is 0.5. Model is trained using 200 epochs with batch size 64.

B. Qualitative and Quantitative Evaluation

Testing is performed on 100 images. We used quantitative measure structural similarity metric (SSIM) [23] to evaluate our network performance. Initially we have directly passed test sketches to our GAN model without applying any filtering techniques and generates corresponding photo images. Without filtering we achieved average Structural similarity metric (SSIM) for our model is 0.78. We identify some sample sketches for which SSIM is below 0.50 and apply median, bilateral and mean filtering technique on them and then passed filtered sample to GAN for photo synthesis. We analyze the Performance of each technique which is summarized in table 1. By applying Median filter, average SSIM is increased by 1%, hence average SSIM of proposed system is 0.79. Bilateral filter, increases average SSIM by 0.8% i.e., average SSIM of proposed system is 0.788. and by applying Mean filter, average SSIM increased by 1.7% i.e., average SSIM of proposed system is 0.797. Qualitative results of different filtering techniques with GAN is shown in figure

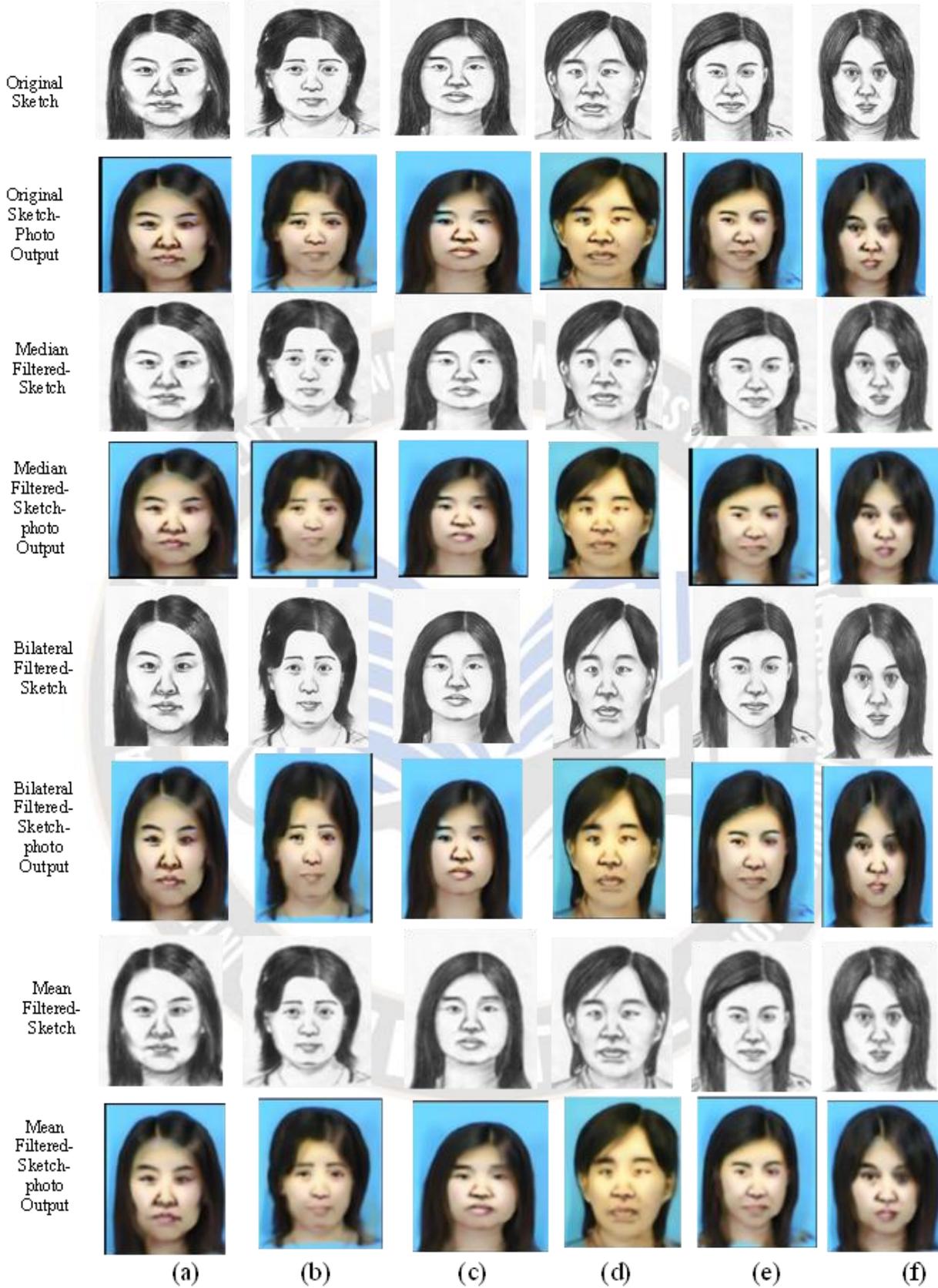


Figure 5 Qualitative Result of synthesized photos with different filtering methods

Table 1. Quantitative Result of Different Samples with Different Filtering Methods

Sample	SSIM Without Filter	SSIM With Median Filter	SSIM With Bilateral Filter	SSIM With Mean Filter
a	0.5912	0.61	0.6016	0.6074
b	0.5736	0.5849	0.5797	0.6074
c	0.5797	0.5935	0.5896	0.5998
d	0.6015	0.6135	0.6127	0.6085
e	0.5084	0.5202	0.5112	0.5283
f	0.5986	0.6058	0.6058	0.6077

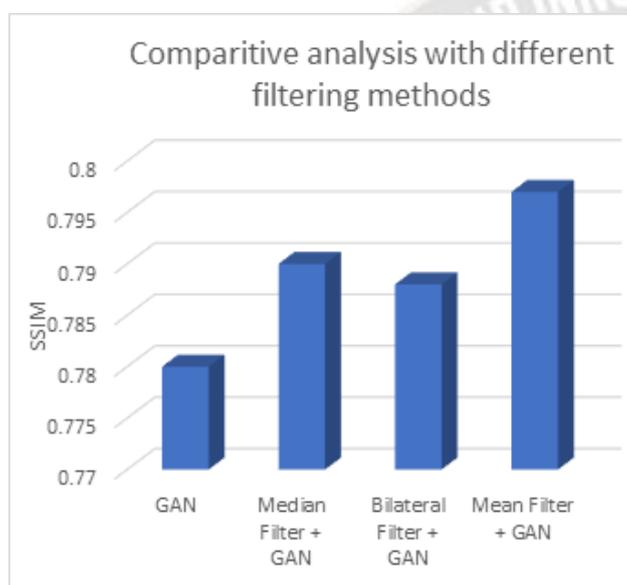


Figure 6 Performance Filter-aided GAN

V. CONCLUSION

In proposed filter-aided GAN framework we implement and analyze different filtering techniques like median filtering, bilateral filtering and mean filtering. Experimental studies show mean filter with GAN increased the overall performance of our model. In future work we try to explore more filtering methods as well as network architecture to achieve better performance.

REFERENCES

[1] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, "Data-driven vs. model-driven: Fast face sketch synthesis," *Neurocomputing*, vol. 257, pp. 214–221, 2017

[2] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition", In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009

[3] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch photo synthesis and retrieval using sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1213–1226, 2012.

[4] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. "Sketch2photo: internet image montage" In: *ACM Transactions on Graphics*, 2009.

[5] Y. Song, L. Bao, Q. Yang, and M. H. Yang, "Real-time exemplar-based face sketch synthesis," in *European Conference on Computer Vision*, 2014, pp. 800–813.

[6] Y. Song, L. Bao, S. He, Q. Yang, and M.-H. Yang, "Stylizing face images via multiple exemplars," *Computer Vision and Image Understanding*, vol. 162, pp. 135–145, 2017.

[7] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face sketch synthesis via sparse representation-based greedy search." *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2466–77, 2015

[8] Ahmed Cherif Megri, Sameer Hamoush, Ismail Zayd Megri, Yao Yu. (2021). *Advanced Manufacturing Online STEM Education Pipeline for Early-College and High School Students*. *Journal of Online Engineering Education*, 12(2), 01–06. Retrieved from <http://onlineengineeringeducation.com/index.php/joe/article/view/47>

[9] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketchphoto synthesis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1364–1376, 2013.

[10] L. Zhang, L. Lin, X. Wu, S. Ding, L. Zhang, End-to-end photo-sketch generation via fully convolutional representation learning, in: *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, in: *ICMR '15*, ACM, New York, NY, USA, 2015, pp. 627–634.

[11] D. Zhang, L. Lin, T. Chen, X. Wu, W. Tan, and E. Izquierdo, "Contentadaptive sketch portrait generation by decompositional representation learning," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 328–339, 2016.

[12] Monowar, M. M. (2022). TRW-MAC: A thermal-aware receiver-driven wake-up radio enabled duty cycle MAC protocol for multi-hop implantable wireless body area networks in Internet of Things. *International Journal of Communication Networks and Information Security (IJCNIS)*, 14(1).

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, DavidWarde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. *Generative adversarial nets*. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[14] Lu Y., Wu S., Tai YW., Tang CK. (2018) *Image Generation from Sketch Constraint Using Contextual GAN*. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) *Computer Vision – ECCV 2018*. ECCV 2018. *Lecture Notes in Computer Science*, vol 11220. Springer, Cham.

[15] Mohammed, M. M., Viktor, Z., Yurii, K., Lyubov, B., Konstantyn, P., & Turovsky, O. L. (2022). *Methods for assessing the impact of bandwidth of control channels on the quality of telecommunications networks in the*

- transmission of data packets of different types. *International Journal of Communication Networks and Information Security (IJCNIS)*, 13(2).
- [16] J. Zhao, X. Xie, L. Wang, M. Cao and M. Zhang, "Generating Photographic Faces from the Sketch Guided by Attribute Using GAN," in *IEEE Access*, vol. 7, pp. 23844-23851, 2019
- [17] Uche Osahor, Hadi Kazemi, Ali Dabouei, Nasser Nasrabadi,(2020), "Quality Guided Sketch-to-Photo Image Synthesis", In: *Computer Vision And Pattern Recognition (CVPR)*.
- [18] Yu et al., "Toward Realistic Face Photo-Sketch Synthesis via Composition-Aided GANs," in *IEEE Transactions on Cybernetics*, doi: 10.1109/TCYB.2020
- [19] Chao, Wentao et al. "High-Fidelity Face Sketch-To-Photo Synthesis Using Generative Adversarial Network," 2019 *IEEE International Conference on Image Processing (ICIP) (2019)*: 4699-4703
- [20] Lidan Wang et al. "High-Quality Facial Photo-Sketch Synthesis Using Multi-Adversarial Networks", 2018 13th *IEEE International Conference on Automatic Face & Gesture Recognition*
- [21] Wang, V. Sindagi and V. Patel, "High-Quality Facial Photo-Sketch Synthesis Using Multi-Adversarial Networks," 2018 13th *IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 2018, pp. 83-90.
- [22] Hadi Kazemi et al. "Facial Attributes Guided Deep Sketch-to-Photo Synthesis", 2018 *IEEE Winter Conference on Applications of Computer Vision Workshops*
- [23] Joy, P., Thanka, R., & Edwin, B. (2022). Smart Self-Pollination for Future Agricultural-A Computational Structure for Micro Air Vehicles with Man-Made and Artificial Intelligence. *International Journal of Intelligent Systems and Applications in Engineering*, 10(2), 170–174. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/1743>
- [24] Yuke Fang Yuke Fang "Identity-aware CycleGAN for face photo-sketch synthesis and recognition," *Pattern Recognition*, Volume 102, 2020, 107249, ISSN 0031-3203.
- [25] Xiaogang Wang and Xiaoou Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.
- [26] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [27] Kumar, S., Gornale, S. S., Siddalingappa, R., & Mane, A. (2022). Gender Classification Based on Online Signature Features using Machine Learning Techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 10(2), 260–268. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2020>
- [28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.
- [29] S. L. Bangare, "Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images", *Neuroscience Informatics*, Volume 2, Issue 3, September 2022, 100019, ISSN 2772-5286, <https://doi.org/10.1016/j.neuri.2021.100019>.
- [30] S. L. Bangare, G. Pradeepini, S. T. Patil, "Implementation for brain tumor detection and three dimensional visualization model development for reconstruction", *ARNP Journal of Engineering and Applied Sciences (ARNP JEAS)*, Vol.13, Issue.2, ISSN 1819-6608, pp.467-473. 20/1/2018 http://www.arnpjournals.org/jeas/research_papers/rp_2018/jeas_0118_6691.pdf
- [31] Sandeep Pande and Manna Sheela Rani Chetty, "Bezier Curve Based Medicinal Leaf Classification using Capsule Network", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, No. 6, pp. 2735-2742, 2019.
- [32] Pande S.D., Chetty M.S.R. (2021) Fast Medicinal Leaf Retrieval Using CapsNet. In: Bhattacharyya S., Nayak J., Prakash K.B., Naik B., Abraham A. (eds) *International Conference on Intelligent and Smart Computing in Data Analytics. Advances in Intelligent Systems and Computing*, vol 1312.
- [33] P. S. Bangare, S. L. Bangare, R. U. Yawle and S. T. Patil, "Detection of human feature in abandoned object with modern security alert system using Android Application," 2017 *International Conference on Emerging Trends & Innovation in ICT (ICEI)*, 2017, pp. 139-144, doi: 10.1109/ETIICT.2017.7977025
- [34] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare et al. "An efficient way of text-based emotion analysis from social media using LRA-DNN", *Neuroscience Informatics*, Volume 2, Issue 3, September 2022, 100048, ISSN 2772-5286, <https://doi.org/10.1016/j.neuri.2022.100048>.
- [35] Suneet Gupta, Sumit Kumar, Sunil L. Bangare, Shibili Nuhmani, Arnold C. Alguno, Issah Abubakari Samori, "Homogeneous Decision Community Extraction Based on End-User Mental Behavior on Social Media", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 3490860, 9 pages, 2022. <https://doi.org/10.1155/2022/3490860>.
- [36] Gururaj Awate, S. L. Bangare, G. Pradeepini and S. T. Patil, "Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow", *arXiv*, <https://doi.org/10.48550/arXiv.1806.10170>