GAN-based Face Reconstruction for Masked-Face

Farnaz Farahanipad
farnaz.farahanipad@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Mohammad Rezaei
The University of Texas at Arlington
Arlington, Texas, USA

Mohammadsadegh Nasr
The University of Texas at Arlington
Arlington, Texas, USA

Farhad Kamangar
The University of Texas at Arlington
Arlington, Texas, USA

Vassilis Athitsos
The University of Texas at Arlington
Arlington, Texas, USA

ABSTRACT

Facial recognition and identification which play an important role in human-computer interaction, secure authentication and criminal face recognition, are impeded by the advent of face masks due to COVID-19 pandemic. This is a challenging problem due to the following reasons: (i) masks cover quite a large part of the face even below the chin, (ii) it is not possible to collect and prepare a real paired-face images with and without mask object, (iii) face alterations and the presence of different masks is even more challenging. In this work, we propose a general framework that can be used to reconstruct the hidden part of face concealed by mask. We have employed GAN-based unpaired domain translation technique to translate masked face images from the source to the unmasked images in the destination domain. To this end, we also create a paired datasets of real face images and synthesized correspondence's with face-masks and use it towards training of our proposed GAN-based facial reconstruction system which can be used for facial identification and secure authentication in human-computer interaction. The obtained results demonstrate that our model outperforms other representative state-of-the-art face completion approaches both qualitatively and quantitatively.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Face Reconstruction, Masked Face Recognition, Human-Computer Interaction, Object Removal, Image Inpainting, Generative adversarial network

ACM Reference Format:
Farnaz Farahanipad, Mohammad Rezaei, Mohammadsadegh Nasr, Farhad Kamangar, and Vassilis Athitsos. 2022. GAN-based Face Reconstruction for Masked-Face. In *The15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3529190.3534774

PETRA 2022, June 29-July 1, 2022, Corfu, Greece

ACM ISBN 978-1-4503-9631-8/22/06.
https://doi.org/10.1145/3529190.3534774

1 INTRODUCTION

In modern technology, face recognition is becoming a new trend for the security authentication systems and human-computer interaction (HCI)[8, 11]. However, with the recent world-wide COVID-19 pandemic, the use of these face masks has raised a serious question on the accuracy of the facial recognition system. Many HCI applications based on face recognition techniques, such as face access control, and face authentication based mobile payment, have nearly failed to effectively recognize the masked faces. Moreover, touch-less verification which allows individuals to perform photo ID checks with their mask on has become extremely important in public places due to the impact of the coronavirus.

Despite the rapid growth in the amount of research works in face identification, the problem of occluded face images, including masks, has not been completely addressed due to the lack of the masked face dataset, large size and complex nature of the mask, and variation in face.

Therefore, recognizing and authenticating people wearing masks will be a long-established research area, and more efficient methods are needed for real-time face recognition. In this work, we are going to attempt to tackle the problem of getting rid of face-masks in facial image by using Generative Adversarial Networks (GANs). The problem we are trying to solve can be viewed as image-to-image translation, which is generally considered to be the process of translation of images from the source to the destination domain. In other words, given a masked face image, we apply unpaired image-to-image translation [23], to remove the mask and synthesizes the affected region with fine details while retaining the global coherency of face structure. More details of the proposed model are discussed in the section 3.

The main contributions of this work are:

- Leveraged by GANs, we propose a novel approach that automatically removes mask object from face and reconstruct the affected region with delicate details.
- To overcome the data scarcity problem, we have collected a 10249 real face images of 12 people and add synthetic mask on the real faces in order to create a paired dataset of with and without mask faces.

The rest of this paper is organized as follows. Section 2 presents related studies. The proposed model is detailed in Section 3. Sections 4 and 5 describe experimental setup and results, respectively.

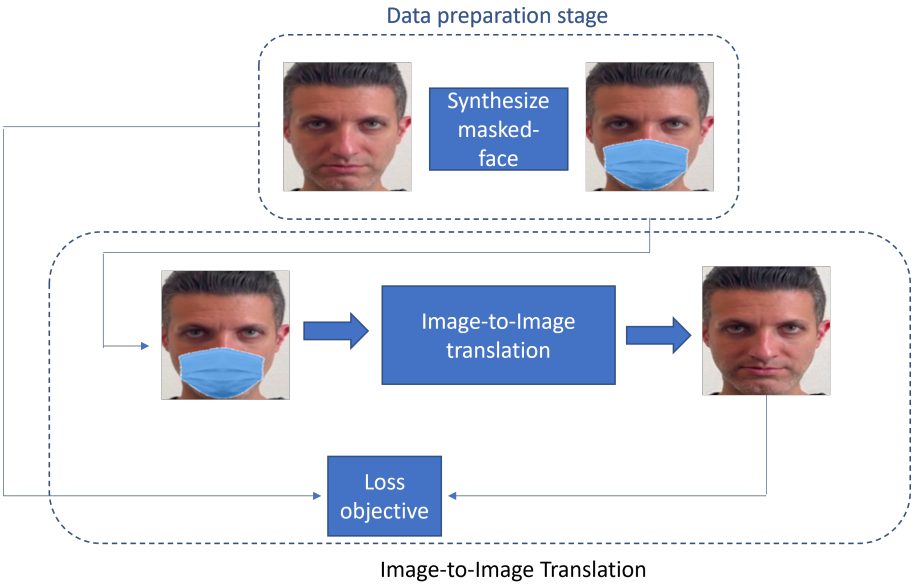

Figure 1: Overview of our proposed model.

2 RELATED WORK

In order to remove undesired object in the images two main problem should be tackled: a) object detection and removal, b) image completion. There has been a considerable amount of learning and non-learning based object removal algorithm to tackle object removal in an image.

Recently, due to the GAN's nature of unsupervised learning, ability to generate high-quality, natural and realistic images, and the power of adversarial training, deep learning GAN-based approaches have merged as a promising paradigm for variety of application such as data augmentation [22], pose estimation [7], and image inpainting [16]. However, due to the plethora of related literature, we only review some representative works related to undesired object such as sunglasses, microphone, hand, and face masks.

Non-learning based object removal algorithms tried to solve the problem by removing the undesired object such as sunglasses, and random objects and synthesize the missing content by matching similar patches from other part of the image [3, 4]. In [17], they introduced a regularized factor to adjust the path priority function in computing function to remove eyeglasses from facial images. However, these methods can only handle relatively small holes, where the color and texture variance are small.

On the other hand, learning-based method mainly describe image inpainting with the main application of object removal and outperform the traditional methods both quantitatively and qualitatively. In [10], Iizuka et al. proposed a GAN-based model, that removes an object and reconstruct the damage part. Their proposed method, leveraged two discriminators (global and local) to ensure local and global realism of the reconstructed image. They also apply Poisson blending as a post-processing. Poisson blending technique is an image processing operator that allows the user to insert one image into another, without introducing any visually unappealing seams. Despite the ability to complete the random damaged part, this meth, is not capable to complete high resolutions images and it is resulting artifacts when damaged part is around the margin of the image. Yu et al. [21], presented a two-stage image inpainting network. First, stage includes a dilated convolutional network which is trained with reconstruction loss to rough out the missing parts. The contextual attention is integrated in the second stage to encourage spatial coherency of attention. In another study by Khan et al., [14], a coarse-to-fine GAN-based approach to remove object from facial images was introduced. For mask removal, Boutros et al. [2] introduced an embedding unmasking model which takes a feature embedding extracted from the masked face as input and generates a new feature embedding similar to an embedding of the unmasked face. Moreover, Din et al. [5, 6] used GAN-based image inpainting for image completion through an image-to-image translation approach to automatically remove face masks.

Due to the great success of learning methods to recover missing part of facial image, we proposed a novel framework which aims to automatically reconstruct hidden part of the masked-face through Image-to-Image translation, and it is able to remove masks regardless of facial angle or underlying facial expression.

3 PROPOSED METHOD

In this section, we provide the details of our proposed GAN-based framework that automatically removes mask and completes the missing hole through image-to-image translation so that the completed face not only looks natural and realistic but also has consistency with the rest of the image. The overall structure of our framework is illustrated in Figure 1.

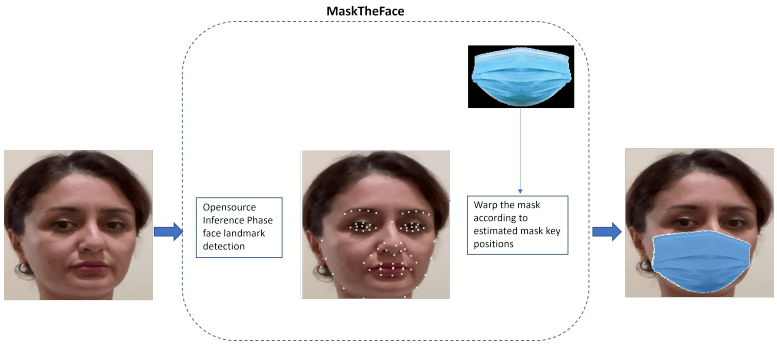

Figure 2: Dataset preparation: To create paired-face dataset with and without mask, "MaskTheFace" [1], tool warps the mask template based on the key face landmark positions of the face.

3.1 Translation using a Cycle-consistency Constraint

Unsupervised Image-to-Image Translation (UI2I) [23], composed of 2 GANs which uses two large but unpaired sets of training images to convert images from one representation to another and vice versa. The data distributions are denoted as $a \sim pdata(a)$ and $b \sim pdata(b)$. More specifically, given an input masked face image, the unpaired image-to-image translation model aims to generate a complete image without the mask using unpaired collections of facial images with and without mask. CycleGAN loss primarily consists of adversarial loss [9] and cycle-consistency loss. Adversarial loss, in GAN, enforce the generated image to be indistinguishable from real photos. While generator, G, tries to find the mapping $G : A \rightarrow B$, its discriminator's D_b objective function is defined as:

$$\mathcal{L}_{GAN}(G, D_B, A, B) = \mathbb{E}_{b \sim pdata(b)}[logD_B(b)] + \mathbb{E}_{a \sim pdata(a)}[log(1 - D_B(G(a))], \quad (1)$$

where, G generates images $G(a)$ that appears like images from field B and D_b observes between translated samples $G(a)$ and original samples b. A similar adversarial loss is postulated for the second generator, $F : B \rightarrow A$ and its discriminator D_a.

However, the adversarial loss alone is not sufficient to produce good images, as it leaves the model under-constrained. Adversarial loss, enforces the generated output be of the appropriate domain but does not enforce that the input and output are recognizably the same. The cycle consistency loss addresses this issue. It relies on the expectation that if you convert an image to the other domain and back again, by successively feeding it through both generators, you should get back something similar to what you put in. In other words, it compares the reconstructed image and input image using L1-norm distance and enforces that $F(G(a)) \approx a$ and $G(F(b)) \approx b$.

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{a \sim p_{data}(a)}[||F(G(a)) - a||_1] + \mathbb{E}_{b \sim p_{data}(b)}[||G(F(b)) - b||_1] \quad (2)$$

The full loss with cycling parameter *lambda* is:
$\mathcal{L} = \mathcal{L}_{GAN}(G, D_B, A, B) + \mathcal{L}_{GAN}(F, D_A, B, A) + \lambda \mathcal{L}_{cyc}(G, F)$ which is used for training the model with an Adam optimizer [15]. *lambda* controls the relative importance of the two objectives.

4 EXPERIMENTAL DETAILS

4.1 Data Preparation

To address the lack of the masked face dataset, this paper firstly contributes a new dataset of high-quality, paired face images with and without mask simultaneously which in real world is not possible. To this end, we propose to superimpose artificial face masks onto real face images based on key face landmarks position, through an open-source masking tool "MaskTheFace" [1], as shown in Figure 2. To this end, first, we collected 10249 high quality face images from 12 people (10 user for training and 2 users for testing) and all faces in the images were detected by YOLO algorithm[18]. Then, we used the dlib library-based face landmarks detector to identify the face tilt and six key features (eyes, nose, lips, face edges etc.,) of the face necessary for applying mask. The template mask is then transformed based on the six key features to fit perfectly on the face. This results in creation of a large, paired dataset of face images with and without masks which can be used as real world masked-face test data and the ground truth data without mask.

4.2 Implementation Details

A CycleGAN model was trained to unmask the masked-face. We adapt the architecture for our generative networks from [23], it utilizes two parts Generators and Discriminators. Each generator composed of three initial convolutional, nine 64-channel convolutional ResNET block, two fractionally strided convolutions, and a final convolution to reduce the output's channel. Furthermore, each discriminator is a 70x70 Patch GAN which penalizes images at the level of individual patches as opposed to per-pixel or per-image basis. We trained the model for 150 epochs with 8131 unpaired facial images of size 256x256 with and without mask with learning rate of 0.0002 and lambda value of 10 to calculate cycle loss. Once the model is trained, we evaluate it using 2118 masked face test images of from our created dataset.

4.3 Evaluation Metric

We compared the results between our method and the other method using the Structural SIMilarity (SSIM) quantitative metrics [19]. However, as reported by many other works [12] and [20], we argue

that quantitative analysis may not be the most effective measure of the image editing task.

Figure 3: Output examples generated by our model for test samples of our created dataset. First column, masked face image in source domain, second column, generated unmasked face in target domain and, third column, ground truth unmasked face in target domain.

5 COMPLETION RESULTS

We now discuss the qualitative and quantitative performance of our method and its comparison with other previous state-of-the-art image manipulation methods on real world images with mask.

5.1 Qualitative comparison

Figure 3 shows the sample generated by our model for masked face test images. We also present a qualitative comparison with Iizuka et al. [10] and Yu et al. [21] on real world test images as can be seen in Figure 4. Although, proposed model by Iizuka et al. [10], completes the image for random damaged region in facial images, it is limited to relatively low resolutions (178x218) and produces artifacts when damaged part is at the margins of an image. Moreover, Yu et al. [21] reduces the artifacts at margins but is unable to recover a complex face structure. Moreover, although in each test image, almost half of key facial semantics are covered by face mask, our model offers significantly improved results for real world data than the other previous state-of-the-art image manipulation methods and successfully removes the mask object and generates natural looking outputs with structural consistency.

Table 1: Performance comparison in term of Structural SIMilarity (SSIM).

Methods	SSIM
Yu et al.[21]	0.86
Ours	**0.89**

5.2 Quantitative comparison

To have a fair comparison, we have created a synthetic dataset of 6446 images using the publicly available, CelebA-HQ [13], celebrity face dataset and trained Yu et al. [21] and our model. Then, we evaluate model performance and training effectiveness by Structural SIMilarity (SSIM) [19]. It is a full reference metric that requires two images from the same image capture and it measures the perceptual difference between two similar images. Since real images with mask do not have corresponding ground truth without mask object, we have evaluated SSIM on 2459 Synthetic test data created using CelebA-HQ. Table 1 provides a quantitative comparison with Yu et al. [21].

Figure 4: Visual comparison of our proposed method with representative image completion methods on real world images. From left to right: Input image, [10], [21], and ours. Note: There is no ground truth since all samples are real world images collected from the Internet.

6 CONCLUSION

Partially concealed faces by mask in situation like pandemics, or air pollution has exerted dramatic influences and reduce the performance of existing security and authentication systems due to the absence of large-scale training data and the presence of large intra-class variation between masked faces and full faces. This imposes the demand to tackle such authentication concerns using more robust and reliable facial recognition systems under different settings. To this end, we proposed a novel method for interaction-free mask removal from facial images. The hidden parts of the face are regenerated in the most realistic way by GAN-based image-to-image translation. Our proposed pipeline could be involved in various areas such as criminal face recognition, and secure authentication. Moreover, due to the lack of public datasets containing real masked face images, we create a high-quality paired dataset of real faces along with their simulated masked one by placing synthetic masks over the real face images for training. To the best of our knowledge, this is the first effort to create high-quality, well-established face benchmarks paired dataset of face images with and without masks. The proposed dataset is part of an ongoing effort to gather a larger scale database with realistic variations of masks and will be available upon request. Moreover, both qualitative and quantitative comparison show that our model demonstrates superior performance for large facial object (face mask) as compared to the state-of-the art.

Several future steps could be taken to improve the results as well as put the trained model into practical usage. First, we plan to collect and expand our masked-face dataset to improve our face reconstruction model. We also plan to develop a user-friendly interface for unmasking the masked-face. Furthermore, future research will continue to leverage these reconstructed face images in state-of-the-art face recognition models in automobile security, secure authentication, and access control.

REFERENCES

[1] Aqeel Anwar and Arijit Raychowdhury. 2020. Masked face recognition for secure authentication. *arXiv preprint arXiv:2008.11104* (2020).

[2] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. 2021. Unmasking face embeddings by self-restrained triplet loss for accurate masked face recognition. *arXiv preprint arXiv:2103.01716* (2021).

[3] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing* 13, 9 (2004), 1200–1212.

[4] Soheil Darabi, Eli Shechtman, Connelly Barnes, Dan B Goldman, and Pradeep Sen. 2012. Image melding: Combining inconsistent images using patch-based synthesis. *ACM Transactions on graphics (TOG)* 31, 4 (2012), 1–10.

[5] Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi. 2020. Effective removal of user-selected foreground object from facial images using a novel GAN-based network. *IEEE Access* 8 (2020), 109648–109661.

[6] Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi. 2020. A novel GAN-based network for unmasking of masked face. *IEEE Access* 8 (2020), 44276–44287.

[7] Farnaz Farahanipad, Mohammad Rezaei, Alex Dillhoff, Farhad Kamangar, and Vassilis Athitsos. 2021. A pipeline for hand 2-D keypoint localization using unpaired image to image translation. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. 226–233.

[8] Felix Ferdinand Goldau, Tejas Kumar Shastha, Maria Kyrarini, and Axel Gräser. 2019. Autonomous multi-sensory robotic assistant for a drinking task. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 210–216.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[10] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2017. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1–14.

[11] Ayush Jain, Deepanshu Arora, Raman Bali, and Deependra Sinha. 2021. Secure Authentication for Banking Using Face Recognition. *Journal of Informatics Electrical and Electronics Engineering* 2, 02 (2021), 1–8.

[12] Kamran Javed, Nizam Ud Din, Seho Bae, Rahul S Maharjan, Donghwan Seo, and Juneho Yi. 2019. UMGAN: Generative adversarial network for image unmosaicing using perceptual loss. In *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 1–5.

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

[14] Muhammad Kamran Javed Khan, Nizam Ud Din, Seho Bae, and Juneho Yi. 2019. Interactive removal of microphone object in facial images. *Electronics* 8, 10 (2019), 1115.

[15] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[16] Avisek Lahiri, Arnav Kumar Jain, Sanskar Agrawal, Pabitra Mitra, and Prabir Kumar Biswas. 2020. Prior guided gan based semantic inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13696–13705.

[17] Jeong-Seon Park, You Hwa Oh, Sang Chul Ahn, and Seong-Whan Lee. 2005. Glasses removal from facial image using recursive error compensation. *IEEE transactions on pattern analysis and machine intelligence* 27, 5 (2005), 805–811.

[18] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[19] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[20] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. 2017. High-resolution image inpainting using multi-scale neural patch synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6721–6729.

[21] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2018. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5505–5514.

[22] Mohammad Zaki Zadeh, Ashwin Ramesh Babu, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. 2021. Self-supervised human activity recognition by augmenting generative adversarial networks. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. 171–176.

[23] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.