

## Masked GANs for Face Completion: A Novel Deep Learning Approach

Anshuman Sharma<sup>1,\*</sup>, Biswaroop Nath<sup>1</sup>, Tejaswini Kar<sup>1</sup> and D. Khasim<sup>2</sup>

<sup>1</sup>KIIT Deemed to be University, Bhubaneswer, Odisha

<sup>2</sup>VIT-AP University, Amravati, Andhra Pradesh, India

### Abstract

**INTRODUCTION:** Recent deep learning based image editing methods have achieved promising results for removing object in an image but fail to generate appreciable performance for removing large objects of complex nature, especially mask from facial images. Towards this goal the objective of this work is to remove mask objects in facial images. In this study, authors propose a novel approach for face completion using Generative Adversarial Networks (GANs) that utilize masked data. This technology can help in image restoration and preservation, thus enabling us to cherish those memories that are held dear to our hearts.

**OBJECTIVES:** Train a GAN to learn the mapping from incomplete to complete face images by utilizing a masked input image.

**METHODS:** The discriminator is trained to distinguish between face images and full ground truth images. Our results indicate that our technique generates high-quality, realistic facial images that are visually comparable to the ground truth and that it can generalise to new faces that were not encountered during training.

**RESULTS:** Our findings indicate that GANs with masked inputs are a good approach for generating whole face images from partial or masked data.

**CONCLUSION:** Our experimental findings show that our method produces facial images of great quality and realism that are visually equivalent to the actual thing. Our proposed approach can also be applied to fresh faces that weren't seen. The performance can still be improved further using larger dataset. Also, further investigation into adversarial attacks may help in improving performance. This technology can be further utilized for developing realtime mask removal software as well.

**Keywords:** Generative Adversarial Networks, masked data, face images

Received on 01 November 2023, accepted on 05 January 2024, published on 15 January 2024

Copyright © 2024 A. Sharma *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ectpht.9.4850

### 1. Introduction

The COVID-19 pandemic [1] has affected our lives in countless ways, including the way the world interacts, and the way people commemorate our essential moments. Wearing masks became an absolute necessity and slowly incorporated into our daily social norms. While face mask has become a fashion for some, it was a nuisance for others.

Especially for many, the major regrets might be that their special photos taken during the pandemic era were taken with face masks. While masks were necessary to protect ourselves and others from the spread of COVID-19, they made it difficult to capture the essence and emotions of those special moments.

The purpose of the paper is to remove masks from these photos, as realistically as possible, thereby restoring the visual memories of the special moments. Our aim of this paper is to use the Generative Deep Learning model [2] to

\*Corresponding author. Email: [anshumansharma179@gmail.com](mailto:anshumansharma179@gmail.com)

train on data to automatically remove face masks from photos which can be further used to develop image editing software that can remove face masks with precision and ease. By removing the face masks from our special photos, one can preserve their moments as they were meant to be remembered, without the visual barrier of a face mask. This would help many to look back at their past memories during the pandemic with more clarity and appreciation for the moments that they shared with their loved ones, even in midst of a challenging time.

To develop a solution to our problem of removing masks from photos during the pandemic, authors have used normal faces dataset, with multiple images of both men and women with their faces not covered by a mask. To simulate the effect of a face mask on a real image, they have employed a mask generator, which applied various types of masks over the real faces, thus mimicking the appearance of face masks in photographs. The mask generator has applied various types of masks with varying degrees of coverage and thickness to suit the face, representing different types of masks that were commonly worn during the pandemic. This is an effective way of generating masked photos while also having an original photo of the respective faces. The deep learning model that had used is Generative Adversarial Network (GAN) [3], which is just another neural network, commonly used for generating new data, such as images. The primary concept of GAN is behind one concept it has two parts the “generator” which creates the data samples and the “discriminator” which distinguished between the generated image and the real image for the training set of images. The two networks are trained in such a process usually known as adversarial training, which constitutes a generator trying to create realistic data that can fool the discriminator, while the discriminator tries to correctly identify which data is real and which is fake. As the two networks keep on competing with each other, the generator keeps on getting better and creates more realistic data, while the discriminator becomes better at identifying real and fake data. Over time, the generator can produce data that is indistinguishable from real data, creating new realistic data. GANs hence, have a wide range of applications from creating realistic images to generating music and speech. GANs also have the advantage of using less training data to create realistic, high-quality data in an unsupervised fashion that can support a variety of data types.

## 2. Literature Review

In recent years deep learning-based approaches are gaining popularity in different fields [4–6]. In this paper, we primarily focus on the unmasking of masked faces using GAN. In 2020, a similar work was published in which authors used GAN-Based Network for the Unmasking of Masked Faces [7] where the authors of the study employed a network based on GAN (Generative Adversarial Network) with two discriminators. The first discriminator was used to learn the overall structure of the face, while the second

discriminator was utilized to focus on learning the intricate details of the missing regions. Similarly, in 2014, a study [8] was conducted where the authors introduced a modified version of generative adversarial nets, called the conditional version. This model could be created by feeding the desired data,  $y$ , to both the generator and discriminator. The authors demonstrated that this model was capable of generating MNIST digit based on class labels. Furthermore, they explained how this model could be utilized to learn a multi-modal model. Moreover, in 2017 authors proposed a new framework called Seq-GAN [9]. In this study, the author proposes a novel method for sequence generation using a combination of Generative Adversarial Networks (GANs) and reinforcement learning techniques. The authors introduce a framework called Seq-GAN that uses a generator network to generate sequences, a discriminator network to evaluate the quality of the generated sequences, and a reinforcement learning algorithm to improve the generator’s performance. Additionally, in 2017 [10] a work was published in which the authors introduce a framework called Dual Discriminator Generative Adversarial Nets (D2GAN) that uses two discriminators to evaluate the generated samples. One discriminator focuses on global features while the other one focuses on local features. The generator is trained to fool both discriminators simultaneously. Furthermore, in 2017 [11] a new approach was proposed in which authors used a new approach to train the Generative Adversarial Nets (GANs) with a mixture of generators the main intuition is to employ multiple generators, instead of using a single one as in the original GAN. Additionally, in 2021 [12] a new approach was presented two-stage Facial Structure Guided Generative Adversarial Network (FSG-GAN). Besides, in 2023 [13] a paper was published in which the survey covers various applications of GANs for face generation, such as face image synthesis, style transfer, facial attribute manipulation, and face aging. The authors also discuss the ethical considerations related to the use of GANs for face generation, such as privacy concerns and potential biases in the generated images. In 2019, a research paper [14] was presented that introduced a method for successful GAN (Generative Adversarial Network) training with limited data by randomly masking certain image information. In another research of GANs, [15] the authors found that the hierarchical convolutional nature of typical GANs generally depends upon the absolute pixel coordinates, resulting in less detailed image generation. They made small architectural changes in the model that guaranteed that unwanted information cannot leak into the synthesis process, thus resulting in better image generation. Again, in [16], the authors have suggested using deep convolutional generative adversarial networks for generating image-like containers using Steganalysis, which is the exact opposite process of collection of methods to hide secret information. In another research of GANs [17], the authors have introduced Slimmable GANs or SlimGANs, which can automatically and flexibly switch the width of the generator to accommodate various trade-offs for quality-efficiency at runtime. The authors have

maintained the consistency between generators of different widths, by using an in-place distillation technique that enables the narrow generators to learn more from the wide ones. Also there had been an interesting research [18], where the authors have used a hybrid network architecture, by decoupling feature generation and neural rendering, to produce high quality 3D synthesis with FFHQ and AFHQ Cats, along with other experiments. The authors of the study created two distinct methods of masking that were used along different dimensions of the training images. One of the methods involved spatial masking that randomly masked certain portions of the images based on random shifts. The other method was balanced spectral masking, which masked certain image spectral bands using self-adaptive probabilities. There is a plethora of GANs research, but the authors could only cover so much in this paper.

### 3. Materials and methods

The following flowchart Fig. 1 illustrates the step-by-step process of this research. The first step involves problem identification, in which the authors have identified the problem of this research. The following step includes a collection of the dataset that was related to the problem. After collection, a Data pipeline was formed in which first data was loaded afterward the authors resized the images according to model requirements after resizing cropping was done on images then eventually data was normalized. Next in this paper, the authors have built Generator, the generator is in charge of producing fresh data that closely resembles the distribution of the training data. It takes a random noise vector as input and creates a sample that is meant to be comparable to the original data. Later on, the discriminator was formed. The discriminator, on the other hand, is in charge of discriminating between genuine and fake data. It takes a sample as input and guesses whether it is accurate or not. In most cases, the discriminator is a neural network with learnable parameters that have been trained to distinguish genuine and fraudulent data successfully. After creating all the necessary requirements, they trained and tested the model for 200000 steps.

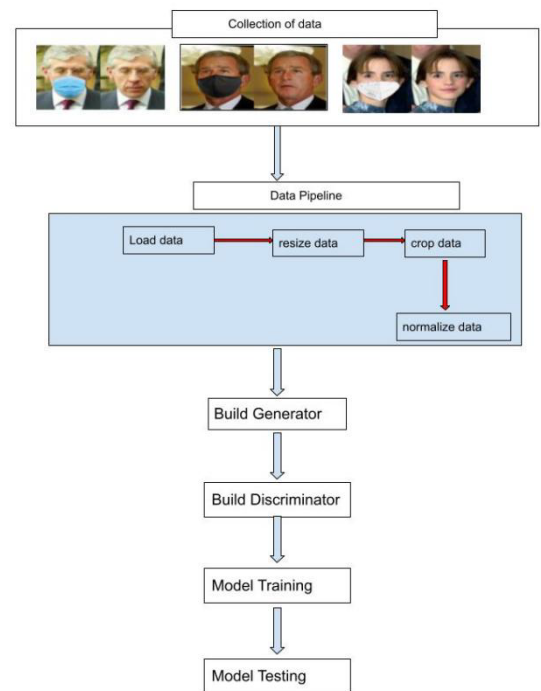


Fig. 1: Workflow

### 4. Dataset

The authors broke the problem into two parts. The first stage of the model automatically produced binary segmentation for the mask region. The second stage removed the mask and synthesizes the affected region with fine details while retaining the global coherency of the face structure. The data contains 7030 training data and 3030 testing data. The image in the data contains input image and ground truth using these parameters. Our purpose of the model is to produce images, the image size used in this study is 256 x 256. Fig. 2 and Fig. 3 show a sample dataset that contains both input image and ground truth.

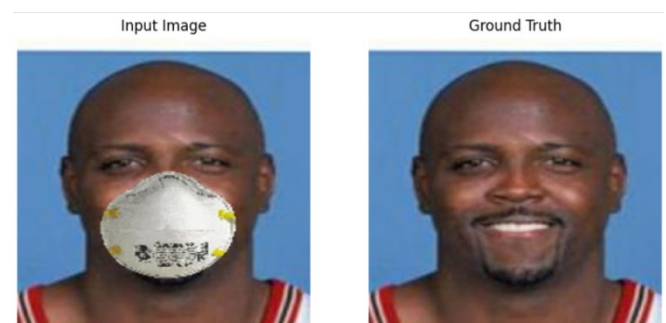


Fig. 2: Data Sample 1



Fig. 3: Data Sample 2

### 5. Model

In this segment, the results will be discussed. After loading and preprocessing data, they worked on the model architecture of GANs, in which authors started by building the generator, and then after the generator, the discriminator was built. The generator network Fig. 4 takes an input picture and produces an output image. To create the output picture, the input image is first down sampled via a number of convolutional layers to extract low-level information, then up sampled through a sequence of transposed convolutional layers. The discriminator Fig. 5 network receives as input pairs of actual and produced pictures and returns a scalar value expressing the likelihood that the input pair is real. To extract features, the input pairs are first concatenated along the channel dimension, then down sampled via a sequence of convolutional layers. To generate the scalar output, the resultant feature map is processed through a final convolutional layer. The generator and discriminator networks are trained alternately throughout training. In each iteration of the training process, the generator network is instructed to produce images that resemble the real images more closely. On the other hand, the discriminator network is trained to become more adept at differentiating between real images and images generated by the generator network.

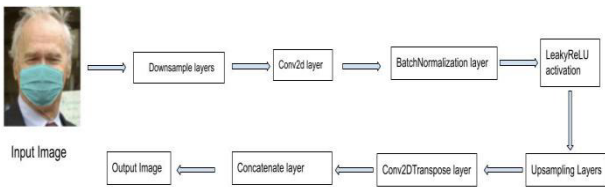


Fig. 4: Generator

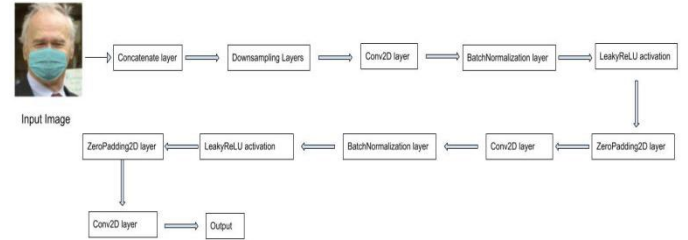


Fig. 5: Discriminator

### 6. Results

In this section, the results produced by this research will be discussed. The authors evaluate the proposed approach to the dataset which contains images of people wearing face masks. They trained their model for 2,00,000 iterations. Time take for every 1000 steps was 49.73 seconds. They randomly split the dataset into training and test sets with a 70/30 split. They used the training set to train our GAN model and evaluated the performance of our approach on the test set. Figure 1 shows some qualitative results of our approach to the test set. One can see that their GAN model is able to generate high-quality, realistic images of people without masks, even when the input image is partial or masked. The generated images closely resemble the ground truth images, with detailed facial features and realistic skin tones. They evaluated the performance of our approach using two common metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

#### 6.1. Peak Signal-to-Noise Ratio

PSNR is a measure of image quality defined as the ratio between the maximum possible value of a signal and the noise in the signal. It is commonly used to compare the quality of a compressed image to the original or uncompressed version. PSNR can be defined.

$$PSNR = 10 \log_{10} \frac{MAX^2}{MSE} \quad (1)$$

MAX: the maximum pixel value of the image (e.g., 255 for 8-bit grayscale or RGB images)

MSE: mean squared error between the original and compressed images.



## 6.2. Structural Similarity Index

SSIM is a measure of image quality that takes into account structural information, such as luminance, contrast, and structure. It is also used to evaluate the similarity between two images.

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)}$$

- x, y: the two images being compared
- $\mu_x, \mu_y$  : the means of x and y, respectively
- $\sigma_x, \sigma_y$  : the standard deviations of x and y, respectively
- $\sigma_{xy}$  : the covariance between x and y
- c1, c2: small constants to avoid instability (often set to  $(k1L)^2$  and  $(k2L)^2$ , respectively, where L is the dynamic range of the pixel values and k1 and k2 are constants)

Fig. 6, Fig. 7 and fig 8 show us the results provided by this research which contain input image, ground truth, and predicted image, which was predicted using the GAN model, it is clearly visible that the predicted image is similar to the ground truth value. Fig. 9, Fig. 10, and fig 11 provide information about parameters that were used to evaluate our model.

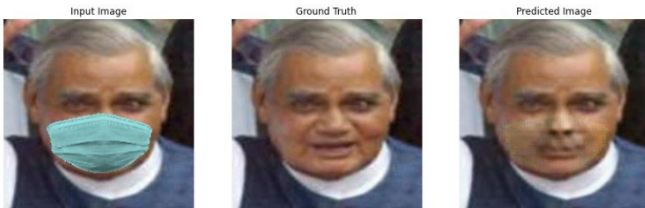


Fig. 6: Sample Result 1



Fig. 7: Sample Result 2



Fig. 8: Sample Result 3



Fig. 9: Sample Result [PSNR+SSIM] 1



Fig. 10: Sample Result [PSNR+SSIM] 2



Fig. 11: Sample Result [PSNR+SSIM] 3

## 6. Conclusion

This study's primary goal is to provide a novel technique for face completion that makes use of Generative Adversarial Networks (GANs) and masked data. The authors suggested a method of feeding a masked input image into a GAN in order to train it to understand the map of the mask on facial images. The discriminator is taught to distinguish between full ground truth images and face images. Our experimental findings show that our method produces facial images of great quality and realism that are visually equivalent to the actual thing. Our approach can also be applied to fresh faces that weren't seen throughout training. Overall, our research indicates that using GANs with masked inputs to create entire face images from partial or masked data is a promising method.

## 8. Future Scope

For future work, more complex GANs with multimodal completion can be tested and improved. Future work can also include further investigation of adversarial attacks and the robustness of the model and try to reduce them. A real-time-based face mask removal software can also be implemented in the future.

## References

[1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, S. Bernardini, The covid-19 pandemic, Critical Reviews in Clinical Laboratory Sciences 57 (6)(2020)365–

- 388,pMID:32645276.arXiv:<https://doi.org/10.1080/10408363.2020.1783198>, doi:10.1080/10408363.2020.1783198.
- [2] R. Salakhutdinov, Learning deep generative models, *Annual Review of Statistics and Its Application* 2 (1) (2015)361–385 arXiv:<https://doi.org/10.1146/annurev-statistics-010814-020120>, doi:10.1146/annurev-statistics-010814-020120
- [3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks (2014). arXiv:1406.2661.
- [4] N. Ud Din, K. Javed, S. Bae, J. Yi, A novel gan-based network for unmasking of masked face, *IEEE Access* 8 (2020) 44276–44287. doi:10.1109/ACCESS.2020.2977386.
- [5] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784 (2014).
- [6] L. Yu, W. Zhang, J. Wang, Y. Yu, Seqgan: Sequence generative adversarial nets with policy gradient 31 (Feb. 2017).doi:10.1609/aaai.v31i1.10804.URL:<https://ojs.aaai.org/index.php/AAAI/article/view/10804>
- [7] T. D. Nguyen, T. Le, H. Vu, D. Phung, Dual discriminator generative adversarial nets, in: *Advances in Neural Information Processing Systems*, 2017, pp. 2670–2680.
- [8] Q. Hoang, T. Nguyen, T. Le, D. Phung, Multi-generator generative adversarial nets (CoRR ,08-2017).
- [9] Y.-M. Cheung, M. Li, R. Zou, Facial structure guided gan for identity-preserved face image de-occlusion, in: *Proceedings of the 2021 International Conference on Multimedia Retrieval, ICMR '21*, Association for Computing Machinery, New York, NY, USA, 2021, p. 46–54.doi:10.1145/3460426.3463642.URL:<https://doi.org/10.1145/3460426.3463642>
- [10] A. Kammoun, R. Slama, H. Tabia, T. Ouni, M. Abid, Generative adversarial networks for face generation: A survey,*ACMComput.Surv.*55(5)(dec2022).doi:10.1145/3527850.URL:<https://doi.org/10.1145/3527850>
- [11] J. Huang, K. Cui, D. Guan, A. Xiao, F. Zhan, S. Lu, S. Liao, E. Xing, Masked generative adversarial networks are data-efficient generation learners, in: S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh (Eds.), *Advances in Neural Information Processing Systems*, Vol. 35, Curran Associates, Inc., 2022, pp. 2154–2167. URL [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/0efcb1885b8534109f95ca82a5319d25-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/0efcb1885b8534109f95ca82a5319d25-Paper-Conference.pdf)
- [12] T. Karras, M. Aittala, S. Laine, E. H“ark”onen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), *Advances in Neural Information Processing Systems*, Vol. 34, Curran Associates, Inc., 2021, pp.852–863.URL: [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/076ccd93ad68be51f23707988e934906-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/076ccd93ad68be51f23707988e934906-Paper.pdf)
- [13] D. Volkhonskiy, I. Nazarov, E. Burnaev, Steganographic generative adversarial networks, in: W. Osten, D. P. Nikolaev (Eds.), *Twelfth International Conference on Machine Vision (ICMV 2019)*, Vol. 11433, International Society for Optics and Photonics, SPIE, 2020, p. 114333M. doi:10.1117/12.2559429.URL:<https://doi.org/10.1117/12.2559429>
- [14] L. Hou, Z. Yuan, L. Huang, H. Shen, X. Cheng, C. Wang, Slimmable generative adversarial networks, *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (9) (2021) 7746–7753. doi:10.1609/aaai.v35i9.16946. URL <https://ojs.aaai.org/index.php/AAAI/article/view/16946>
- [15] E. R. Chan, C. Z. Lin, M. A. Chan, K. Nagano, B. Pan, S. De Mello, O. Gallo, L. J. Guibas, J. Tremblay, S. Khamis, T. Karras, G. Wetzstein, Efficient geometry-aware 3d generative adversarial networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16123–16133.