Signal, Image and Video Processing (2023) 17:3749–3757
https://doi.org/10.1007/s11760-023-02602-2

ORIGINAL PAPER

SAM C-GAN: a method for removal of face masks from masked faces

Akhil Kumar[1] · Manisha Kaushal[2] · Akashdeep Sharma[3]

Received: 25 July 2022 / Revised: 8 March 2023 / Accepted: 14 April 2023 / Published online: 26 May 2023

Abstract
The past years of COVID-19 have attracted researchers to carry out benchmark work in face mask detection. However, the existing work does not focus on the problem of reconstructing the face area behind the mask and completing the face that can be used for face recognition. In order to address this problem, in this work we have proposed a spatial attention module-based conditional generative adversarial network method that can generate plausible images of faces without masks by removing the face masks from the face region. The method proposed in this work utilizes a self-created dataset consisting of faces with three types of face masks for training and testing purposes. With the proposed method, an SSIM value of 0.91231 which is 3.89% higher and a PSNR value of 30.9879 which is 3.17% higher has been obtained as compared to the vanilla C-GAN method.

Keywords Image translation · Object removal · Image editing · C-GAN · Spatial attention module

1 Introduction

The COVID-19 pandemic has opened a new realm in computer vision for the detection of the face with masks which has gained the attention of researchers across the globe. Researchers across the globe proposed several methods for the classification and detection of persons wearing face masks using deep learning-based methods. Initially, when the data for persons wearing face masks were scarce, researchers utilized the method of data augmentation to enhance the size of the dataset. The limitation of data augmentation is that it can only create new images from existing images by changing the viewpoints and orientations. However, the other way to enhance the dataset is by utilizing generative adversarial network (GAN)-based methods which are capable of producing synthetic and plausible images from real images. The other problem that aroused during the COVID-19 pandemic is the identification of persons behind face masks as face masks aid in camouflaging the significant facial attributes of the face which are essential for the identification of a person behind a face mask. As an advantage of this, the crime rate increased in many places of the world [1–3]. A certain mischievous section of the society preferred face masks as an aid to hide their facial identities and committed crimes such as thefts, robberies, etc. To propose a solution to generate synthetic images for persons without face masks, in this work we have developed an improved conditional generative adversarial network (C-GAN)-based method to create plausible images of faces without masks. The proposed method generates plausible images for faces without masks by reconstructing the facial area hidden behind the masks. The generated images with the proposed method can be a helpful tool in generating new samples of faces without masks and help crime investigation agencies to reveal the identities of criminals committing crimes by hiding their identities behind a face mask.

The objective of this work is to develop and propose an interaction-free face mask removal technique for images having faces with masks. The objective of this work is achieved by utilizing a deep learning-based conditional generative adversarial network (C-GAN) [4] method. The deep learning-based C-GAN method is used for generating images

✉ Akashdeep Sharma
akashdeep@pu.ac.in

Akhil Kumar
akhil.hpucs@gmail.com

Manisha Kaushal
manisha.kaushal@thapar.edu

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

[2] CSED, Derabassi Campus, Thapar Institute of Engineering and Technology, Punjab, India

[3] CSE, UIET, Panjab University, Chandigarh, India

by performing the image-to-image translation. The C-GAN method is based on a single generator and discriminator where the generator generates the plausible images that are identical to the real images, whereas the discriminator classifies the generated images as real or fake. In recent years, C-GAN has shown fascinating results in image-to-image translation focusing on the task of data augmentation. The authors in [5] employed a C-GAN method for augmenting the fruit quality and defect classification dataset. With the C-GAN method, the authors compressed the size of original images by 50% and achieved a classification accuracy of 81.16% for fruit quality and defect. Another work [6] utilized the C-GAN method for creating synthetic samples for speech adversarial examples and classification. The authors in [7] utilized the C-GAN method for augmenting the scalogram images of respiratory signals for COVID-19 and further classified the images generated by the C-GAN method using deep learning-based classifiers. Furthermore, along with the usage of C-GAN for data augmentation, generative adversarial network (GAN) has proven its ability in image editing tasks. The GAN-based methods empower image editing tasks by learning from large-scale datasets. The authors in [8] used a GAN architecture consisting of a single generator and two discriminators to remove unwanted objects from an image and fill the damaged regions with synthetic content. The GAN architecture-based EdgeConnect [9] and SPG-Net [10] used two-stage adversarial method to remove unwanted objects from the images.

In the last two years, several works using computer vision and deep learning for the detection of faces with masks are proposed; however, the other important aspect of identification of the person behind the face mask is yet to be addressed. To address this important problem, in this work we have proposed a novel solution that can unmask the faces with masks and reconstruct the facial area hidden behind the masks. In order to develop the solution, firstly we have created an edited image dataset consisting of faces with White, Anti-COVID, and 3 M masks. Further, we performed face mask detection on the created dataset to extract the face mask regions as key points. Our solution employs an improved conditional generative adversarial network (C-GAN) consisting of a generator and a discriminator. To improve the classification accuracy of the C-GAN method, we have applied the spatial attention module (SAM) with the discriminator which produces an attention map and aid the discriminator to classify between real images for masked faces and synthetic images for masked faces with high accuracy.

Following are the highlights and contributions of this work:

- Proposal of spatial attention module-based conditional generative adversarial network (SAM C-GAN)-based method for unmasking masked faces. We created an edited image dataset for faces with White, Anti-COVID, and 3 M mask consisting of 11,262 images for each face mask type.
- We performed face mask detection on a publicly available dataset consisting of 52,535 images to extract face mask regions and used extracted face mask regions as key points passed to the developed SAM C-GAN-based method to classify and localize face masks of varying sizes and viewpoints.
- The proposed SAM C-GAN method is evaluated for SSIM and PSNR metrics. In comparison to C-GAN, the proposed method achieved a 3.89% higher value for SSIM and a 3.17% higher value for PSNR. Further, the proposed SAM C-GAN method achieved 0.49–4.77% higher value for SSIM and 1.88–4.71% higher value for PSNR as compared to the related work in literature.

This work is composed of the following sections: Sect. 2 presents the related work in the domain of image editing and face mask removal; Sect. 3 presents the materials and methods describing the created dataset and developed method; Sect. 4 presents the experimental evaluations and results; and Sect. 5 presents the conclusion.

2 Related work

R. Shetty et al. [11] proposed an end-to-end deep learning method for object removal. In the proposed method, the authors utilized deep learning as a tool to find and remove objects automatically from scene images. To carry out this work, authors utilized MS COCO [12] dataset. S. Iizuka et al. [8] carried out work for image editing using deep learning. In the proposed work, the authors utilized GANs as a method to learn global coherent and corrupted region completion jointly with global and local discriminators. The proposed work was carried out on Places2 [13] and CelebA [15] datasets. J. Yu et al. [15] proposed a GAN-based image editing method that utilized a coarse-to-fine GAN-based approach with a contextual attention module for image inpainting. To carry out this work, the authors utilized ImageNet [16], Places2, and CelebA datasets. K. Nazeri et al. [9] proposed a two-stage adversarial network consisting of an edge generator followed by an image editing module for an image editing task. The authors tested their network on CelebA, Places2, and Paris StreetView [18] dataset. M. Khan et al. [19] utilized a coarse-to-fine generative adversarial network as an inpainting technique to remove the microphone from facial images. To carry out this work, the author's utilized synthetic images created using the CelebA dataset. Dong et al. [20] utilized a conditional GAN to synthesize high-quality results for filling the damaged regions in radar data. K. Javed et al. [17] proposed a stacked GAN which generates image information in the first step and subsequently edits the image. The proposed method utilizes a GAN model with one discriminator for image information generation and image editing.

In faces with and without mask inpainting and reconstruction, a few works have been carried out in recent years using GAN-based models. Ud Din et al. 2020 [21] utilized U-NET [22] architecture to binarize the images and further utilized a single generator as the editing module and two discriminators to create the synthetic images for faces without a mask. The authors trained and tested their method on CelebA dataset. Jiang et al. [27] proposed a dual GAN-based model to generate the missing parts of the face covered under the mask area. The authors first created a synthetic dataset by applying face masks on the face area and further generated the plausible samples of faces without masks. Farahanipad et al. [28] employed a GAN-based model that performs image-to-image translation and uncovers the facial area hidden behind the mask. To carry out this task, the authors generated a synthetic dataset by placing masks on the face area [14] and generated plausible samples using the Cycle-GAN.

The literature on object removal, image editing, and face mask removal and reconstruction indicates that there is a scarcity of methods that are efficient in the detection of masks on the face area, further removing the face mask and reconstructing the facial area behind the mask. Considering the gaps in the literature, in this work, we have paved efforts to come up with a solution that can remove face masks from the face region and generate the synthetic image of a real complete face.

3 Materials and methods

The approach followed to execute this work combines an edited dataset and further training of the developed SAM C-GAN-based method to create the plausible images from the dataset and removal of the face masks from the face area by reconstructing the face region behind the face masks. The elaborative details of the dataset and method are presented in subsequent subsections.

3.1 Dataset

To carry out this work, we have created a custom dataset by placing the face masks on the face region on the images extracted from the Bollywood Celebrity Faces dataset [25]. The dataset consists of two sets of images, i.e., faces without masks and faces with masks. The original extracted dataset consists of real 11,262 images of persons without face masks. To create plausible images of faces with masks, firstly we edited the original dataset by placing face masks on the face area. In an effort to generate realistic images, we have used three types of face masks that are generally used by the people, i.e., White mask, Anti-COVID mask, and 3 M mask. In total, the edited dataset consists of 11,262 images of faces without masks, 11,262 images for faces with white face masks, 11,262 images for faces with Anti-COVID face masks, and 11,262 images for faces with 3 M face masks. However, the method exploited for removing face masks from the face area utilizes both the images for faces without masks and faces with masks. To further state, the identity of each face is preserved by using the same images as of the original dataset while creating the synthetic dataset having faces with mask. The dataset in detail is presented in Fig. 1.

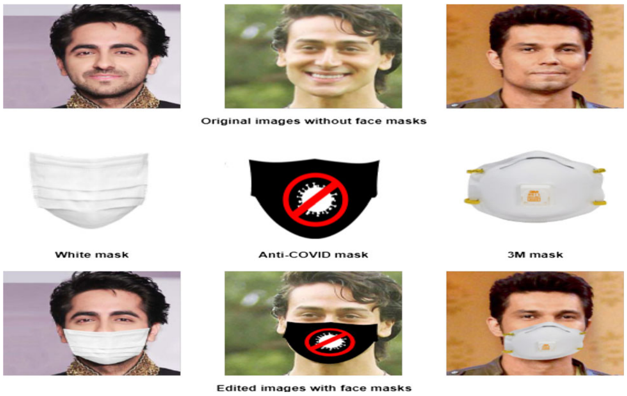
Original images without face masks
White mask
Anti-COVID mask
3M mask
Edited images with face masks

Fig. 1 Dataset description

For accurate prediction of the face mask region by the proposed method, we have trained the tiny YOLO v4-SPP face mask detector [24] and performed detection of the face mask area on a publicly available face mask detection dataset [23] consisting of 52,635 images of faces with and without masks. The detected face mask regions by the tiny YOLO v4-SPP face mask detector are cropped and extracted from the images detected with the presence of a face mask. The detected face mask regions by the tiny YOLO v4-SPP face mask detector are used as black rectangle boxes highlighting the face mask regions and applied to the edited Bollywood Celebrity dataset consisting of faces with three types of face masks. We have placed highlighted mask regions on the edited images so that face masks of varying types and complexities can be accurately recognized by the proposed method for removal of the face masks from masked faces. The details of mask area detection by tiny YOLO v4-SPP face mask detector and placing highlighted mask regions on the edited images are illustrated in Figs. 2 and 3.

3.2 Approach

To carry out this work and to perform the task of image-to-image translation for the removal of face masks from the masked faces, we have used conditional generative adversarial network (C-GAN). The C-GAN is composed of a

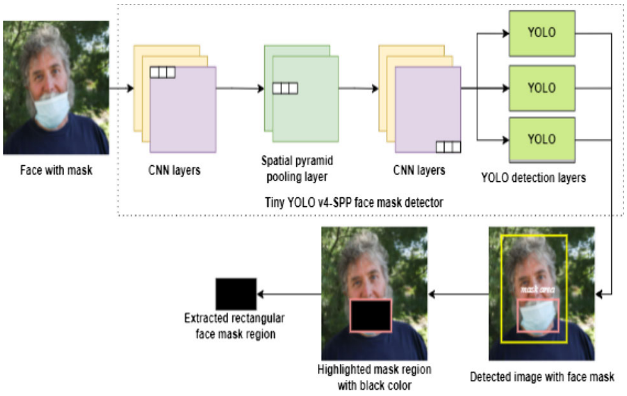
Face with mask
CNN layers
Spatial pyramid pooling layer
CNN layers
YOLO
YOLO
YOLO
YOLO detection layers
Tiny YOLO v4-SPP face mask detector
Extracted rectangular face mask region
Highlighted mask region with black color
Detected image with face mask

Fig. 2 Face mask area detection using tiny YOLO v4-SPP

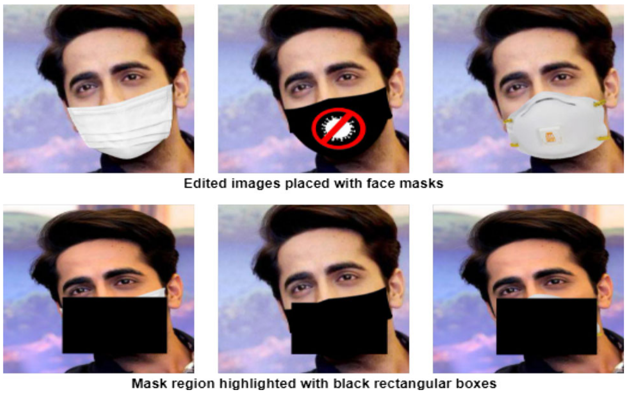
Edited images placed with face masks
Mask region highlighted with black rectangular boxes

Fig. 3 (Colour figure online) Covering face masks with black rectangular boxes

generator and discriminator module where the generator consists of an encoder and a decoder. The role of the generator is to create plausible images for the input images, whereas the discriminator checks whether the image is real or fake. In the present work, we have utilized the generator of the C-GAN method without any modifications. However, to make the discriminator of the C-GAN method more efficient, we have added the spatial attention module (SAM) to it. The spatial attention module (SAM) in the discriminator aided in the classification of images as real or fake along with highlighting the most discriminative regions in the images by generating an attention map. The detailed working of the proposed SAM C-GAN-based method is as follows:

3.2.1 Generator module

The generator module of the proposed SAM C-GAN method is composed of an encoder and a decoder which is made up of convolutional layers. The encoder and decoder of the generator module act as an image editor where it is conditioned by an input, i.e., images from the dataset. In the present work, the generator is fed by the images of faces with masks and black rectangular boxes as the key points. To encourage the generator to create synthetic images in the target domain, i.e., synthetic images of faces with masks and key points, it is trained via adversarial loss. Furthermore, the generator is also updated via the L1 loss which is measured between the generated image and the expected output image. The L1 loss in the generator module also aids in creating plausible transformations of the source domain. The generator module is made up of fourteen convolutional layers where seven convolutional layers are utilized by the encoder and seven convolutional layers are utilized by the decoder. The input size of the generator module is 256×256. The encoder of the generator module is applied with the Leaky ReLU activation function to prevent overfitting, whereas the decoder is applied with the TanH activation function to aid the generator module to perform backpropagation. However, no batch normalization has been applied with the encoder and decoder networks. The generator module of the proposed SAM C-GAN-based method takes input as images of faces with masks and black rectangular boxes, and by utilizing the convolutional layers of the encoder and decoder it generates synthetic images for faces without masks.

3.2.2 Discriminator module

The discriminator module of the proposed SAM C-GAN method is composed of convolutional layers and a spatial attention module (SAM) [26]. The discriminator module is composed of six convolutional layers and takes the output of the generator module as the input along with the real images of faces without masks. The discriminator module compares the synthetic images of faces without masks generated by the generator module with real images of faces without masks. The comparison is performed to achieve pixel to pixel translation and to determine the feature match representation. The discriminator can take an input image of arbitrary size. To improve the discriminator module, we have added a spatial attention module (SAM) to it. The addition of the spatial attention module aids the discriminator module to classify between real and fake images with high accuracy by highlighting the discriminative features present in each image. The spatial attention module (SAM) is composed of residual blocks and spatial attentive blocks. The residual blocks in the spatial attention module (SAM) act as skip connections and the spatial attentive block pass the features between the convolutional layers of the discriminator and residual blocks. The spatial attention module (SAM) generates an attention map that highlights the most discriminative features present in the images and helps the discriminator to classify the images as real or fake. The detailed working of the proposed SAM C-GAN method consisting of generator and spatial attentive discriminator module is presented in Fig. 4.

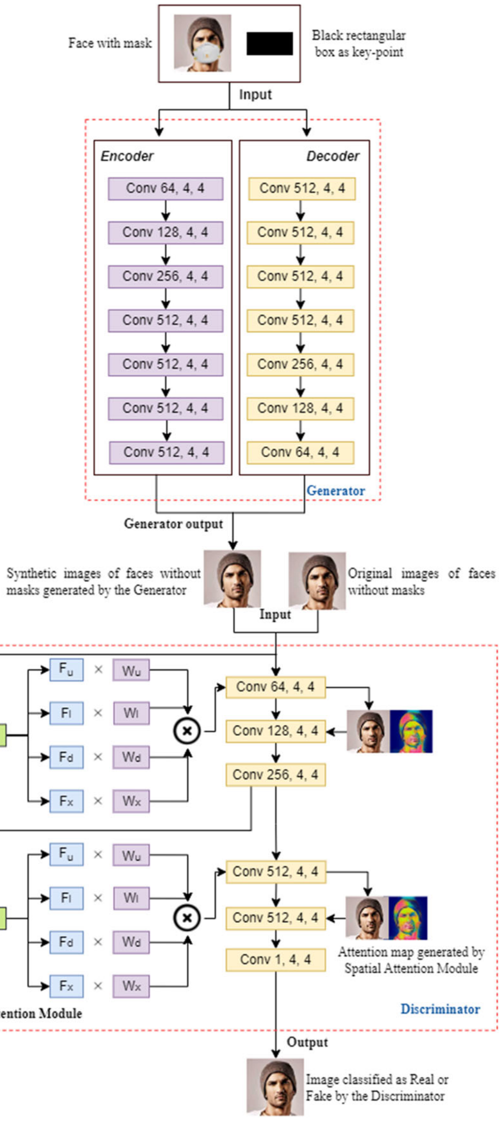

Fig. 4 Proposed SAM C-GAN method

The spatial attention module (SAM) of the discriminator module takes an input image of size x and produces a spatial attention map $A_{\text{DX}}(x)$ which is of the size of the dimension of the input image fed to the discriminator. The spatial attention map generated by the spatial attention module (SAM) highlights the most discriminative regions in the input images that are to be classified by the discriminator module. The spatial attention map $A_{\text{DX}}(x)$ can be defined as the sum of the absolute values of activation maps in each spatial location of each layer of the convolutional network of the discriminator across the channels of the input images. The $A_{\text{DX}}(x)$ can be defined using Eq. (1).

$$A_D = \sum_{i=1}^{C} |F_i| \tag{1}$$

In the above equation, F_i represents the ith feature plane of a discriminator layer for a given input and C represents the number of channels. The A_D in the above equation indicates the significance of the hidden units at each spatial location in classifying the image as real or fake. Furthermore, the spatial attention module W represents the weight assigned to each feature of the input image. The resultant of the spatial attention module is a spatial attention map which is a concatenation of features and weights $(F * W)$ for each input image.

In the proposed SAM C-GAN method, both generator and discriminator are trained simultaneously in an adversarial manner. The approach followed by the method updates the discriminator module directly, and the generator module is updated via the discriminator module. The adversarial training aids the SAM C-GAN-based method to produce the plausible images by the generator and discriminator to identify the counterfeit images.

3.2.3 Loss function

The adversarial loss for the C-GAN can be mapped as: $G : X \rightarrow Y$, and its discriminator D_Y can be stated by Eq. (2).

$$\begin{aligned} L^{x}_{\text{GAN}}(G,\ D_Y,\ \text{X},\ \text{Y}) = {} & \text{E}_{y\sim P\text{data}(y)}[log D_Y(y)] \\ & + \text{E}_{x\sim P\text{data}(x)}[1 - log D_Y(G(x))] \end{aligned} \tag{2}$$

The inverse mapping for $F : Y \rightarrow X$ can be stated using adversarial loss as shown in Eq. (3).

$$\begin{aligned} L^{y}_{\text{GAN}}(F,\ D_X,\ \text{Y},\ \text{X}) = {} & \text{E}_{x\sim P\text{data}(x)}[log D_X(x)] \\ & + \text{E}_{y\sim P\text{data}(y)}[1 - log D_X(F(y))] \end{aligned} \tag{3}$$

In Eqs. (2) and (3), G and F are the mapping functions that aim to minimize the loss against the adversary discriminators D_X and D_Y to maximize the loss. Since the C-GAN is based on min–max problem, the objective loss for the proposed SAM C-GAN method can be stated using Eq. (4).

$$G^*,\ F^*,\ D_X^*,\ D_Y^* = \arg \frac{\min}{G,\ F} \frac{\max}{D_X,\ D_Y} L(G,\ F,\ D_X,\ D_Y) \tag{4}$$

This work combines the above-specified generator, discriminator, and loss function to remove the face masks from the masked faces. The input is the images fed from the employed dataset. The generator takes images of masked faces and black rectangular boxes as key points and produces synthetic images of faces without masks. The discriminator takes the output produced by the generator and real faces without masks as input and compares the images using a spatial attention module (SAM) and convolutional layers to generate a spatial attention map and classify the image as real or fake. Like other GAN-based methods, the limitation of the proposed method is that it works with paired images.

4 Experiments and results

The proposed SAM C-GAN method for the unmasking of masked faces is implemented using TensorFlow and Keras on a system with configuration: Intel Xeon ® W-2245 CPU @ 3.90 GHz × 16 with RAM 64 GB and NVIDIA RTX 3090 24 GB GPU. For training and testing the proposed SAM C-GAN method, we divided the created synthetic dataset consisting of 33,786 images with three type of face masks into a ratio of 80:20. To generate plausible images of faces without masks using the generator module, we have used 80% of images, and for classifying the output of the generator module by the discriminator module as real or fake, we have used the output of the generator module and real 11,262 images of faces without mask. For testing the method, we have utilized 20% images, i.e., 6,757 images which are separated from the training set consisting of images of faces with masks. We trained the SAM C-GAN method for 25,000 iterations with a batch size of 10 and the Adam optimizer. Furthermore, with the generator module of the SAM C-GAN method, we have used binary cross entropy (BCE) loss, whereas for the discriminator module we have used Poisson loss and mean absolute error (MAE) loss.

4.1 Evaluation metrics

The present work is based on image-to-image translation; therefore, the structural similarity of the two images and the signal-to-noise ratio are of prime importance. To find out the structural similarity, we have used the SSIM metric, whereas for signal-to-noise ratio, we have used the PSNR metric.

The SSIM metric refers to the structural similarity index. It utilizes luminance intensity, contrast, and structural information to measure the similarity between two images. The SSIM metric can be stated using Eq. (5).

$$\mathrm{SSIM}(x, y) = l(x, y) * c(x, y) * s(x, y) \tag{5}$$

In the above equation, x and y are two images; l signifies the luminance; c is the measure of contrast; and s represents the structural information of the image.

The other metric PSNR used in this work signifies the peak signal-to-noise ratio. It is employed to measure the quality estimation for the loss of quality of different codecs and image compression. The PSNR metric considers the real image as the signal and the noise as the error that occurred by compressing the image. The PSNR value is computed by finding the mean squared error (MSE). For a given noise-free mxn monochrome image I and its noise approximation K, the MSE can be stated using Eq. (6).

Table 1 Evaluation results for SAM C-GAN method

Method	SSIM	PSNR
C-GAN	0.87341	27.8145
Proposed SAM C-GAN	**0.91231**	**30.9879**

Bold values represent better results with the proposed method

$$\mathrm{MSE} = \frac{1}{\mathrm{mn}} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \tag{6}$$

Furthermore, the PSNR can be stated using Eq. (7).

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{7}$$

In Eq. (7), MAX_I^2 represents the maximum valid value for a pixel.

4.2 Evaluation results

In order to evaluate the performance of the proposed SAM C-GAN method for the unmasking of masked faces, we trained and tested the original C-GAN and SAM C-GAN method on the employed edited dataset and computed the evaluation metrics. The performance results of the SAM C-GAN method in comparison to C-GAN for SSIM and PSNR metrics are presented in Table 1.

As shown in Table 1, the proposed SAM C-GAN method outperformed the performance of the C-GAN method by 3.89% for SSIM and 3.17% for PSNR metrics. The proposed SAM C-GAN utilizes the same generator and discriminator module as utilized by the C-GAN method for the image-to-image translation task. However, as an improvement, we have added a spatial attention module (SAM) to the discriminator module of the C-GAN method. The addition of the spatial attention module (SAM) to the discriminator module of the C-GAN method aided the discriminator to generate an attention map for the fed images and highlighted the most discriminative features in the images. The attention map generated by the spatial attention module (SAM) allowed the discriminator to classify the images accurately by comparing the intensity, contrast and structural similarity, and signal-to-noise ratio of the fed images and attention map, and classify the unmasked faces as real or fake. To provide more intuitive results, pictorial illustrations generated by the SAM C-GAN method for unmasking the masked faces are presented in Fig. 5.

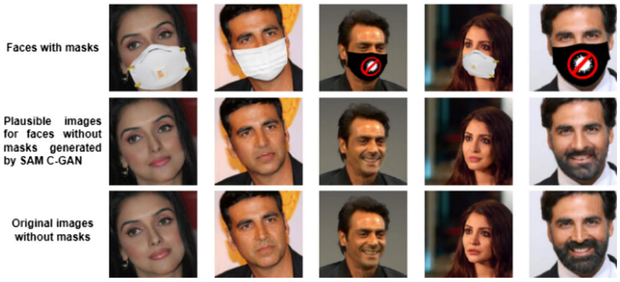
Faces with masks
Plausible images for faces without masks generated by SAM C-GAN
Original images without masks

Fig. 5 Faces without masks generated by SAM C-GAN

The pictorial results as shown in Fig. 5 consist of real images of faces without masks, edited images with face masks placed on the face area, and plausible images of faces without masks generated by the SAM C-GAN method. The pictorial results show that the SAM C-GAN method is capable of generating plausible images of faces without masks with high accuracy by maintaining illuminance, contrast, structural similarity, and identity preservation. Furthermore, the SAM C-GAN method is capable of removing the face masks from the face area from the edited images and reconstructing the face area behind the face mask region with high accuracy with inconsequential distortion.

To gauge the efficacy of the proposed method in generating images similar to the real images of faces, we performed experiments for face recognition using the deep learning-based classifiers, namely ResNet-101 [29] and EfficientNetV2 [30]. For this task, we used paired images, i.e., real images of the faces, and synthetic images of faces generated by the proposed SAM C-GAN method. To train and test the deep learning-based classifiers, the dataset has been prepared by selecting 500 images of real faces and 500 images of synthetic faces. Further, from the dataset 80% images are used for training and 20% images are used for testing the classifiers. The performance of deep learning classifiers has been evaluated using metrics, namely accuracy, precision, recall, and F1 score as mentioned in Eqs. (8–11). The performance results for face recognition with deep learning classifiers are provided in Table 2.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{False Positive}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (8)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (9)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (10)$$

$$\text{F} - 1\ \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

For the classification of real and synthetic faces, the ResNet-101 classifier achieved an accuracy of 98.10% and EfficientNetV2 achieved an accuracy of 97.60%. Moreover, for other performance metrics, significant values are achieved that prove that the synthetic faces generated by the proposed method have features that make them distinguishable from real faces. The results as shown in Table 2 indicate that the images generated by the proposed SAM C-GAN are accurate and can be used for face recognition to identify real or fake faces.

4.3 Comparison with related work

For the problem of removal of face masks from the face area, very few works are available in the literature. The existing work solves the problem by using GAN-based methods. In an effort to find the efficacy of the proposed SAM C-GAN-based face mask removal method, we compared its performance with the work of Ud Din et al. [21], Jiang et al. [27], and Farahanipad et al. [28]. We trained and tested the methods proposed in the related work on the dataset proposed in this work and evaluated them for SSIM and PSNR metrics. The results of the proposed SAM C-GAN method in comparison to the related work in the literature are presented in Table 3.

As shown in Table 3, the proposed SAM C-GAN method achieved a significant higher value for SSIM and PSNR metrics as compared to the work in the literature. The better results with the proposed method in terms of SSIM were 0.49–4.77% and PSNR were 1.88–4.71% higher as compared to the existing work [21][27][28]. The results also indicate that the proposed SAM C-GAN method with a single generator and single discriminator with a spatial attention module (SAM) achieves better performance results as compared to the method proposed by Ud Din et al. [21] which is based on a single generator and two discriminators and state-of-the art Cycle-GAN-based method [27]. The results of the method of Jiang et al. [27] are closest to this work as it is similar to the method proposed in this work which is based on single generator and single discriminator. The higher values for SSIM and PSNR with the SAM C-GAN method are due to the addition of the spatial attention module (SAM) in the discriminator module. The spatial attention module (SAM) generates attention maps and provides discriminative features to the discriminator to classify between real and fake images, thus generating the plausible images for faces with masks by the proposed SAM C-GAN method.

5 Conclusion

This work proposed a novel spatial attention module and conditional generative adversarial network (SAM C-GAN) method for interaction-free face mask removal from facial

Table 2 Face recognition using deep learning classifiers

Classifier	Class	Precision	Recall	F-1 Score	Accuracy
ResNet-101	Real	98%	98%	98%	98.10%
	Synthetic	97%	98%	98%	
EfficientNetV2	Real	96%	97%	97%	97.60%
	Synthetic	97%	96%	97%	

Table 3 Comparison with related work

Work	Method	SSIM	PSNR
Ud Din et al. [21]	GAN with single generator and two discriminators	0.86348	26.2741
Jiang et al. [27]	GAN with single generator and single discriminator	0.90631	29.1057
Farahanipad et al. [28]	Cycle-GAN	0.86825	26.4957
Ours	**SAM C-GAN**	**0.91231**	**30.9879**

Bold values represent better results with the proposed method

images. For face mask removal, we have employed C-GAN with a generator to generate synthetic images of faces without masks which are fed to the discriminator having a spatial attention module (SAM). The spatial attention module (SAM) of the discriminator model highlights the discriminative features of the input images as an attention map. The discriminator compares the synthetic input images as fed from the generator, attention maps, and real images of faces without a mask to classify them as real or fake. The proposed SAM C-GAN method acts as an image editing and reconstruction model to produce fake images by removing face masks from the images containing masks on the face area. The quantitative comparison of the proposed SAM C-GAN method with the C-GAN method shows a 3.89% higher value for SSIM and a 3.17% higher value for PSNR metrics. In comparison the works in this domain, the proposed SAM C-GAN method achieved a 0.49–4.77% higher value for SSIM and 1.88–4.71% higher value for PSNR metrics. The proposed method can be used as a tool by the security and law enforcement agencies to unmask the faces of criminals and law offenders who commit crimes by hiding their faces behind face masks and determining their identities. The future work in this domain can be extended to use of identity preserved C-GAN to check identities, age, and other facial features of faces with mask.

Acknowledgements Not applicable.

Author contributions AK contributed to the study design, implementation, results analysis and preparation of the final draft. AS contributed to the design, conceptualization and preparation of the dataset. MK contributed to the preparation of the dataset. All authors reviewed the manuscript.

Funding This work is supported and funded by All India Council of Technical Education, India under Research Promotion Scheme vide Grant Number: File no. 8–108/FDC/RPS (POLICY-1/2019–20).

Availability of data and materials The created dataset is available on reasonable request to the corresponding author.

Declarations

Conflict of interest The authors have no competing interests to disclose.

References

1. Babwin, D., Dazio, S.: Coronavirus masks a boon for crooks who hide their faces. CTV News. https://www.ctvnews.ca/health/coronavirus/coronavirus-masks-a-boon-for-crooks-who-hide-their-faces-1.4942454 (2020)
2. Gaiss, K.: Masks make it more difficult for police to identify suspects. WCAX3. https://www.wcax.com/2021/04/05/masks-make-it-more-difficult-for-police-to-identify-suspects/ (2021)
3. Southall, A., Van Syckle, K.: Coronavirus bandits? 2 armed men in surgical masks rob racetrack. The New York times. https://www.nytimes.com/2020/03/08/nyregion/aqueduct-racetrack-robbery.html (2020)
4. Mirza, M., Osindero, S.: Conditional generative adversarial nets. https://arxiv.org/abs/1411.1784 (2014)
5. Bird, J.J., Barnes, C.M., Manso, L.J., Ekárt, A., Faria, D.R.: Fruit quality and defect image classification with conditional GAN data augmentation. Sci. Hortic. **293**, 110684 (2022). https://doi.org/10.1016/j.scienta.2021.110684
6. Wang, D., Dong, L., Wang, R., Yan, D.: Fast speech adversarial example generation for keyword spotting system with conditional GAN. Comput. Commun. **179**, 145–156 (2021). https://doi.org/10.1016/j.comcom.2021.08.010
7. Jayalakshmy, S., Sudha, G.F.: Conditional GAN based augmentation for predictive modeling of respiratory signals. Comput. Biol. Med. **138**, 104930 (2021). https://doi.org/10.1016/j.compbiomed.2021.104930
8. Izuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (2017). https://doi.org/10.1145/3072959.3073659
9. Nazeri, K., NG, E., Joseph, T., Qureshi, F.Z., Ebrahimi, M.: EdgeConnect: generative image inpainting with adversarial edge learning. https://arxiv.org/abs/1901.00212 (2019)

10. Song, Y., Yang, C., Shen, Y., Wang, P., Huang, Q., Jay Kuo, C.C.: SPG-Net: segmentation prediction and guidance network for image inpainting. https://arxiv.org/abs/1805.03356 (2018)
11. Shetty, R.R., Fritz, M., Schiele, B.: Adversarial scene editing: automatic object removal from weak supervision. https://arxiv.org/abs/1806.01911 (2018)
12. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: Microsoft COCO: common objects in context. https://arxiv.org/abs/1405.0312 (2014)
13. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: a 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(6), 1452–1464 (2018). https://doi.org/10.1109/TPAMI.2017.2723009
14. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. https://arxiv.org/abs/1411.7766 (2014)
15. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. https://arxiv.org/abs/1801.07892 (2018)
16. Deng, J., Dong, W., Socher, R., Li, L-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Miami, FL, USA, (2009). https://doi.org/10.1109/CVPR.2009.5206848
17. Javed, K., Ud, D.N., Bae, S., Yi, J.: Image unmosaicing without location information using stacked GAN. IET Comput. Vis. **13**(6), 588–594 (2019). https://doi.org/10.1049/iet-cvi.2018.5623
18. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What makes Paris look like Paris? Commun. ACM. (2015). https://doi.org/10.1145/2830541
19. Khan, M.K.J., Ud, D.N., Bae, S., Yi, J.: Interactive removal of microphone object in facial images. Electronics **8**(10), 1115 (2019). https://doi.org/10.3390/electronics8101115
20. Dong, G., Huang, W., Smith, W.A.P., Ren, P.: A shadow constrained conditional generative adversarial net for SRTM data restoration. Remote Sens. Environ. **237**, 111602 (2020). https://doi.org/10.1016/j.rse.2019.111602
21. Ud, D.N., Javed, K., Bae, S., Yi, J.: A novel GAN-based network for unmasking of masked face. IEEE Access **8**, 44276–44287 (2020). https://doi.org/10.1109/ACCESS.2020.2977386
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. https://arxiv.org/abs/1505.04597 (2015)
23. Kumar, A., Kalia, A., Verma, K., Sharma, A., Kaushal, M.: Scaling up face masks detection with YOLO on a novel dataset. Optik **239**, 166744 (2021). https://doi.org/10.1016/j.ijleo.2021.166744
24. Kumar, A., Kalia, A., Sharma, A., Kaushal, M.: A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system. J. Ambient Intell. Hum. Comput. (2021). https://doi.org/10.1007/s12652-021-03541-x
25. Bollywood Celebrity Faces Dataset. https://www.kaggle.com/datasets/havingfun/100-bollywood-celebrity-faces (2020)
26. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. https://arxiv.org/abs/1805.08318 (2018)
27. Jiang, Y., Yang, F., Bian, Z., Lu, C., Xia, S.: Mask removal: face inpainting via attributes. Multimed. Tools Appl. **81**, 29785–29797 (2022). https://doi.org/10.1007/s11042-022-12912-1
28. Farahanipad, F., Rezaei, M., Nasr, M., Kamangar, F., Athitsos, V.: GAN-based face reconstruction for masked-face. In: Proceedings of the 15th international conference on PErvasive technologies related to assistive environments (PETRA), Corfu, Greece, (2022). https://doi.org/10.1145/3529190.3534774
29. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. https://arxiv.org/abs/1512.03385 (2015)
30. Tan, M., Le, Q.V.: EfficientNetV2: smaller models and faster training. https://arxiv.org/abs/2104.00298 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.