Multimedia Tools and Applications (2024) 83:893–918
https://doi.org/10.1007/s11042-023-15807-x

Unmasking the potential: evaluating image inpainting techniques for masked face reconstruction

Chandni Agarwal[1] · **Charul Bhatnagar**[1]

Received: 23 June 2021 / Revised: 16 April 2023 / Accepted: 10 May 2023 /
Published online: 29 May 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
The performance of most Face Recognizers tends to degrade when dealing with masked faces, making face recognition challenging. Image inpainting, a technique traditionally used for restoring old or damaged images, removing objects, or retouching photos, could potentially aid in reconstructing masked faces. In this paper, we compared three state-of-the-art image inpainting models—PatchMatch, a traditional algorithm, and two deep learning GAN-based models, Edge Connect and Free form image inpainting—to assess their performance in regenerating masked faces. The evaluation was conducted using own created synthetic datasets MaskedFace-CelebA and MaskedFace-CelebA-HQ, along with a synthetic masked dataset created for paired comparisons of masked images with ground truth for face verification. The computed results for Image Quality Assessment (IQA) between ground truth and reconstructed facial images indicated that the Gated Convolution model performed better than the other two models. To further validate the results, the reconstructed and ground truth images were also subject to VGG16 classifier, a widely used benchmark model for image recognition. The classifier outcomes supported the quantitative and qualitative assessment based on IQA.

Keywords Generative Adversarial Network · Face reconstruction · Image Inpainting · Face inpainting · Deep learning · Masked Face Recognition · Face Recognition

1 Introduction

Face recognition (FR) systems have been widely used for identifying individuals based on their facial features, and their performance has been proven to be highly accurate. However, the emergence of the Covid-19 pandemic and the widespread adoption of face masks has posed a significant challenge to these systems. As a result, there has been a surge of interest in research on facial recognition with masked faces, as evident from the increasing number

✉ Chandni Agarwal
chandni.officialid@gmail.com

Charul Bhatnagar
charul@gla.ac.in

[1] GLA University, Mathura, India

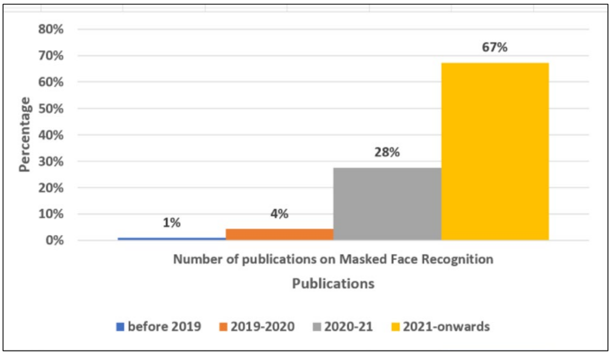

Fig. 1 Percentage of Publications on Masked Faces Recognition

of publications in this area (Fig. 1). While occluded face recognition was not extensively studied prior to the pandemic, the focus has now shifted towards developing methods to detect, validate, and ensure proper mask-wearing in public areas. Various deep learning models, such as transfer learning of VGG16 [37], InceptionV3 [23], Retinamask [21], FacemaskNet [19], and small CNN models [34, 35], have been used to develop automated face mask detection systems but at the same time recognizing face covered with mask is indeed challenging for the face recognition system. This in fact, decreases the accuracy of the deep learning models used for face recognition systems, leading to increased false positives. A suggested solution is the masked face reconstruction to create an unmasked face that reveals facial features similar to the ground truth image. This improves the accuracy of feature extraction and recognition tasks. This approach is certainly expected to reduce false positives and enhance the reliability of the face recognition systems where subjects are occluded due to Covid-type face mask.

In our study, we presented our work in two phases: face reconstruction using non-learning and learning-based image inpainting techniques in the first phase, and utilizing a VGG16 classifier to classify reconstructed faces and ground truth images, with decreased accuracy in similarity between the reconstructed and original faces in the second phase.

Image inpainting techniques involve synthesizing visually realistic and semantically plausible pixels in missing regions of images, which may be distorted due to factors such as occlusion, scratches etc. in an image (Fig. 2). These techniques can also be extended to tasks such as image un-cropping, rotation, stitching, and super-resolution. In our research, we aim to evaluate the effectiveness of various image inpainting techniques for reconstructing masked faces, considering the latest advancements in this field. Through a

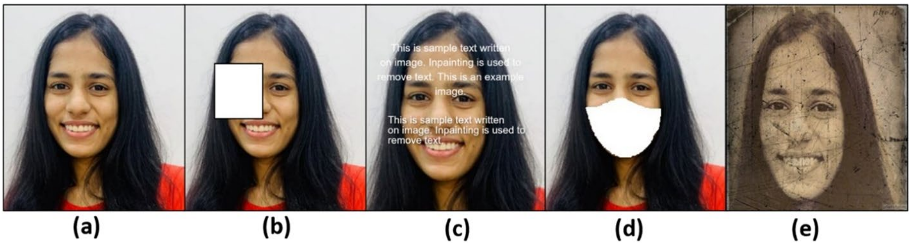
This is sample text written on image. Inpainting is used to remove text. This is an example image.
This is sample text written on image. Inpainting is used to remove text.
(a)
(b)
(c)
(d)
(e)

Fig. 2 Applications of Inpainting (a) Original Image (b) Block Distortion (c) Distortion due to Text (d) Distortion due to Occlusion (e) Distortion due to Scratches

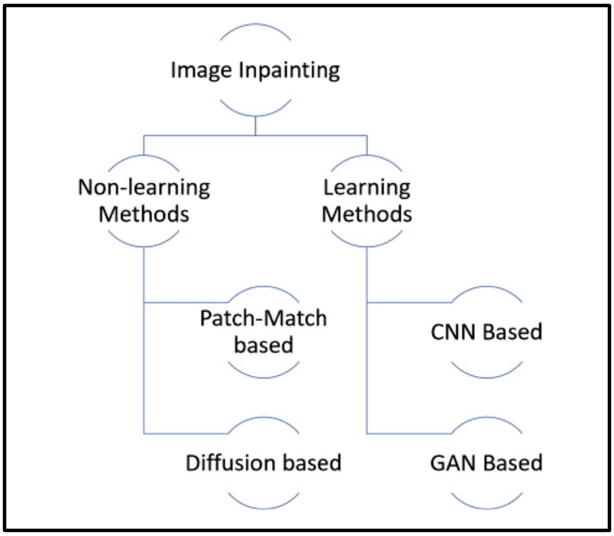

Fig. 3 Hierarchical representation of image inpainting techniques

comprehensive review of existing literature on masked face recognition and image inpainting, we assess the performance of state-of-the-art methods in this specific context. Masked face recognition has emerged as a cutting-edge research topic in computer vision [3, 14, 31], leading to the adoption of deep learning models in many face recognition systems. Additionally, automatic image inpainting has gained significance as a challenging research area, thanks to advancements in image processing tools and the flexibility of digital image editing. One of the primary challenges in image inpainting is accurately recovering large missing areas in an image, as it can be difficult to reconstruct missing regions without complete information about what is missing or hidden. In this paper, our evaluation focuses on non-learning and learning-based image inpainting techniques as a means to unmask the face by reconstructing the occluded area. We assess the effectiveness of these techniques in recovering missing regions and revealing the obscured facial features, aiming to contribute to the understanding of their potential in the context of masked face recognition. As shown in Fig. 3, Non-deep learning-based or traditional image inpainting techniques can be broadly classified into two categories: patch-based and diffusion-based methods. These approaches make use of low-level image features and prior knowledge, such as patch offset statistics and low rank estimation, in order to reconstruct the missing regions [5]. However, accurately restoring complex details that are specific to the missing area poses challenges for these techniques. Some earlier attempts have sought to overcome these limitations by employing strategies such as matching and replicating background patches into the holes or propagating from hole boundaries, akin to texture synthesis [12]. Although these methods demonstrate efficacy in background inpainting tasks, they may encounter difficulties in scenarios where the missing regions involve intricate and non-repetitive structures, and are unable to capture high-level semantics accurately, making them less effective in such cases.

On the other hand, deep learning models, as shown in Fig. 3, are further divided into CNN-based and GAN-based image inpainting techniques. These deep learning-based methods have formulated inpainting as a conditional image generation problem. These techniques have shown remarkable success in generating new content in highly structured

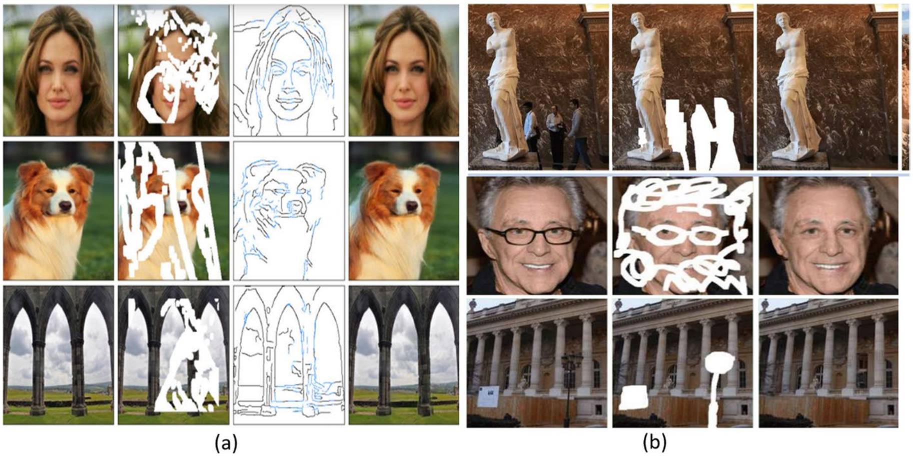

Fig. 4 Image Inpainting Models: (a) Edge Connect—Generative Image Inpainting with adversarial edge learning [29] (b) Free-form Image Inpainting with Gated Convolution (Source: Original paper)

images such as faces, objects, and natural scenes, and have overcome challenges such as boundary distortions, deformed structures, and hazy textures that are common in traditional CNN approaches. By utilizing learning from data distribution, these methods are capable of generating coherent structures [18] in the missing region, which was challenging for non-learning-based techniques.

Deep learning techniques, particularly generative adversarial networks (GANs), have shown remarkable success in generating visually convincing pixels for missing parts in complex images such as faces, objects, and natural scenes, making them suitable for challenging image inpainting tasks. Traditional non-learning methods struggle with accurately reconstructing such complex structures. Several studies have demonstrated the effectiveness of GANs[45,50] in face completion through image inpainting, particularly for restoring faces after removing facial occlusions [31, 38, 40, 43]. Figure 4 depicts the implementation of Edge Connect Generative Image Inpainting with adversarial edge learning [29] and Free-form Image Inpainting with Gated Convolution [44] models on various natural images, including facial and non-facial scene-based images.

As mentioned previously, these cutting-edge models are primarily designed for image inpainting to complete missing areas, particularly for facial and non-facial images. In our experiment, we aim to unmask masked faces by reconstructing the face using these state-of-the-art models. However, evaluating the performance of these models requires a large dataset with paired masked and unmasked images as ground truth, which is often scarce. To overcome this challenge, we created two synthetic masked face datasets, namely MaskedFace-CelebA and MaskedFace-CelebA-HQ. These datasets were generated by applying masks on original images from well-known benchmark face datasets, CelebA [28] and CelebA-HQ [24], resulting in a total of 21,844 masked faces from 25,000 CelebA images and 26,027 masked faces from 30,000 CelebA-HQ images. The performance of Edge Connect [29] and Free-Form Image Inpainting with Gated Convolution [44] models was then evaluated on these created masked datasets, which included male and female celebrity faces with variations in pose, illumination, age, and other factors. The MaskTheFace [2] script was utilized to detect key facial features, such as the forehead, eyes, nose, mouth, jawline, and chin, necessary for applying masks in the creation of these datasets. This approach enables us to overcome the scarcity of ground truth images and conduct comprehensive evaluations of the performance of these models in reconstructing masked faces.

The goal of this research is to use advanced image inpainting techniques to reconstruct the obscured portions of faces in masked images, with the aim of improving the performance of face recognition systems in masked scenarios. We will evaluate state-of-the-art image inpainting models, leveraging the VGG16[37] convolutional neural network (CNN) architecture, a popular deep learning model for image classification tasks. Our findings may have practical implications in enhancing the accuracy and reliability of face recognition systems in real-world scenarios where masks are commonly worn.

2 Major contributions

As outlined in the introduction section, this study provides an in-depth review of existing methods while also introducing our own datasets. The noteworthy contributions of this research can be summarized as follows:

(i) Creation of two synthetic masked face datasets, MaskedFace-CelebA and MaskedFace-CelebA-HQ, using benchmark face datasets CelebA [28] and CelebA-HQ [24]. These datasets contain 21,844 and 26,027 masked face images, respectively, along with paired ground truth images for face reconstruction. These datasets were used for training and evaluating the proposed model for face reconstruction and recognition.

(ii) Conducting a comparative study of three state-of-the-art image inpainting methods, namely PatchMatch [5], EdgeConnect [29], and Free-Form Image Inpainting with Gated Convolution [44]. Qualitative and quantitative evaluations were performed to assess the effectiveness of traditional and deep learning-based image inpainting methods for face reconstruction.

(iii) Providing a comprehensive review of the performance of image inpainting models for reconstructed faces in face recognition, considering the implications and potential applications of the reconstructed faces in real-world scenarios.

The rest of the paper is organized into several sections. Section 3 discusses the research scope and related work in the field of face reconstruction and recognition. Section 4 summarizes the benchmark datasets used in the study and describes the creation of synthetic masked face datasets for training and evaluation. Section 5 demonstrate the models used for masked face reconstruction and provides details about the experimental setup. Section 6 presents the comparative study of qualitative and quantitative results to evaluate the performance of the generated output(s) from the adopted models. Section 7 discusses the implications of reconstructed faces in face recognition. Finally, Sect. 8 highlights the future scope of research and provides a conclusion summarizing the findings and contributions of the study.

3 Related work

This section offers a comprehensive overview of the current research on image inpainting techniques and their diverse applications across various domains, including face reconstruction, as illustrated in the organization chart (Fig. 3) of this study. It provides a detailed

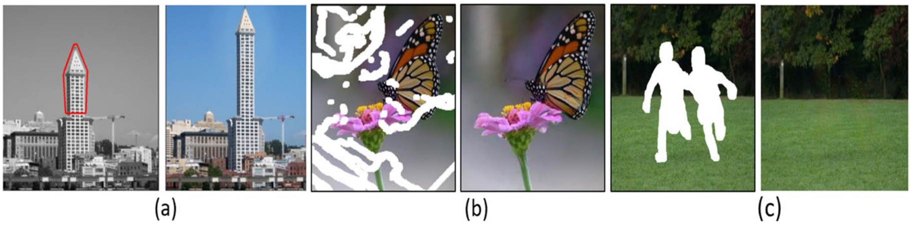

Fig. 5 Examples of Inpainting (a) Scaling up (b) removing scratches (c) Filling the missing block

summary of the existing literature in this field, highlighting the key findings and contributions of previous research in the area of image inpainting, with a specific focus on its applications for face reconstruction.

3.1 Image inpainting

Image inpainting, also known as image completion or image hole-filling, is a computational technique used to fill in missing regions of an image with visually coherent and semantically meaningful contents. It has applications in computer vision, graphics, and image processing, including tasks such as object removal, image synthesis, image restoration, and video frame interpolation, among others (Fig. 5).

Traditional approaches to image inpainting [62,63] rely on diffusion-based or patch-based methods that use low-level features for patch matching and filling in missing regions. However, these methods may struggle with complex inpainting scenarios or generating semantically meaningful contents. In recent years, deep learning-based approaches, particularly using convolutional neural networks (CNNs), have gained prominence for image inpainting. These approaches leverage the power of CNNs to learn complex patterns and structures from large training datasets, enabling them to generate realistic and high-quality inpainted images.

The use of image inpainting has been extended to various applications, including object removal in images, image synthesis for graphics and virtual reality, image restoration in medical imaging, video processing, multimedia editing, and more. The ability to generate visually coherent and semantically meaningful contents in missing regions of images has opened up new possibilities for image manipulation, editing, and enhancement in diverse domains.

3.2 Non-learning based image inpainting methods

Traditional image inpainting methods rely on searching for similar image patches either from the image itself or a large dataset and pasting them into the missing regions [13]. This search process can be time-consuming and typically involves hand-crafted distance measure metrics [11]. Traditional methods can be categorized into diffusion-based and patch-based methods.

Diffusion-based methods propagate neighbouring information into the missing regions using differential operators, but they are limited to locally available information and may struggle to recover meaningful structures in large missing regions [4, 7]. On the other hand, patch-based approaches, initially introduced by Criminisi et al. [11] in 2004, used texture generation techniques to fill in large missing areas by copying and pasting nearby patches from a source image into the destination image [13]. Patch similarity computation for each target-source pair can be computationally intensive, but fast algorithms like PatchMatch proposed by Barnes et al. [5] in 2009 have been developed to address this issue and shown practical values in various image editing applications, including inpainting.

To minimize discontinuities, blending of source and target regions has been proposed by Huang et al. in 2014 [42]. However, these methods can be computationally expensive due to the computation of similarity scores for each target-source pair, which limits their practical applications. Patch-based algorithms have been used for tasks such as image stitching, image denoising, super-resolution, object detection, and tracking [6, 8, 16, 32]. These methods are most effective in natural scene images that require image retargeting, completion, and shuffling. However, they may struggle to fill in holes with semantic or novel content as they rely solely on low-level features for patch matching.

Overall, traditional image inpainting methods have limitations in handling complex inpainting scenarios and generating semantically meaningful contents. Deep learning-based approaches, such as convolutional neural networks (CNNs), have emerged as a promising paradigm for high-quality image completion, as they can learn complex patterns and structures from large training datasets, enabling them to generate realistic and visually appealing inpainted images.

3.3 Deep learning based image inpainting

Deep learning-based techniques for image inpainting have gained popularity due to their ability to generate visually plausible results with good global consistency and local fine textures. Jain V.et.al. [20] initially designed architecture for image denoising using CNN and XIe et.al [39]. used sparse coding and deep neural network as denoising auto-encoder extended to image inpainting using CNN on their inhouse images dataset.

Recent researches show the popularity of GAN based techniques to achieve realistic inpainting results.One early approach was the Context Encoder proposed by Pathak et al. [30] in 2016, which used an encoder-decoder network trained to handle 64×64-sized holes. However, the output images often had over-smoothed or blurry regions due to the information bottleneck in the fully connected layer. To address this, Yang et al.[40] proposed an improved version of Context Encoder in 2017 that used a multi-scale neural patch synthesis technique to enhance texture details. However, this approach had limitations in filling missing parts in complicated scenes and required higher training time for real-time performance. Iizuka et al.[18] introduced global and local discriminators as adversarial losses in 2017, along with dilated convolution layers to replace the fully connected layer and handle input images of various sizes. This approach showed promising results, but the use of large dilation factors resulted in increased training time. Zeng et al.[46] developed a controlled image inpainting system by integrating a deep generative model with global matching based on closest neighbours, but this approach had limitations in generalizing to masks of any shape, size, or position.

Liu et al.[26] proposed partial convolution in 2018, where convolution weights were normalized by the mask area of the window to prevent capturing too many zeros in incomplete regions. This approach was the first to handle irregular holes and showed improved results. Wang et al. [26] suggested an image inpainting method based on the attention mechanism and partial convolution to achieve more realistic outcomes. Zheng et al.[45] introduced a pluralistic image inpainting approach to achieve various inpainting results.

To capture long-range spatial dependencies, Yu et al. [43] proposed a contextual attention module in 2018 and integrated it into networks to borrow information from distant spatial locations. They also used WGAN adversarial loss and weighted L1 loss to improve training stability. However, their model mainly trained on large rectangular masks and did not generalize well on free-form masks. For free-form image inpainting, Yu et al. [44] extended their work in 2019 with DeepFill v2, a generative image inpainting system based on gated convolutions and a patch-based GAN loss.

Jiang et al.[22] proposed a novel image inpainting approach based on Wasserstein GAN with skip-connections and autoencoders in 2020. The skip-connections enhanced the prediction ability of the generator and prevented gradient vanishing, resulting in improved image quality. Cai et al.[9] proposed PiiGAN in the same year, which added a new style extractor in the GAN to generate diverse results with plausible content for a single input image with missing regions. This model showed efficient results in art restoration, facial micro shaping, and image augmentation, but had limitations in handling large irregular missing areas. Liu et al.[27] proposed PD-GAN in 2020, a probabilistic diverse GAN for image inpainting. This model generates multiple inpainting results with diverse and visually realistic content for a given input image with arbitrary hole regions. Spatially probabilistic diversity normalization (SPDNorm) is proposed inside the modulation to control the diversity of generated images. Further research and advancements are reported in order to give improved image quality using deep learning algorithms.

Yang et al. [41] presented a contextual feature constrained DCGAN with paired discriminator for face completion. They used a pre-trained VGG network to extract features and introduced a paired feature matching loss to stabilize training. Experimental results showed promising outcomes with improved texture and semantic consistency. The proposed work by chen et.al. [10] introduces a two-stage framework for face image inpainting using latent feature reconstruction and mask awareness. It includes a pre-trained StyleGAN generator for preliminary restoration with a latent cosine similarity loss, and a hierarchical attention mechanism between the encoder and decoder.A recent work proposed by the Yao. et.al. [42] introduces a second order generative image inpainting model that combines edge and feature self-arrangement modules. The edge repair network effectively reconstructs structural information in the broken area, while the image inpainting network uses the generated edge map as prior conditions for decoding. A feature self-arrangement module is also incorporated to fill the broken area with effective information at the feature level. The proposed model demonstrates the ability to generate content with similar semantics, valid structure, and clear texture features as the original image. Overall, deep learning-based image inpainting techniques have shown significant progress in recent years, but there are still challenges in handling complex scenes, irregular holes, and large missing areas.

4 Dataset Generation

A masked face dataset is constructed using a computer vision script MaskTheFace[2] on the benchmark face datasets CelebA[28] and CelebA-HQ[24] to assess the suggested state-of-the-art image inpainting techniques for face reconstruction. This tool applies the user-selected mask by identifying variations such as face tilt, angle, and position, among other things, and employs a dlib based facial landmarks detector to determine the face tilt, which is one of the six critical aspects of the face required for precisely fitting the mask on the face (Fig. 6).

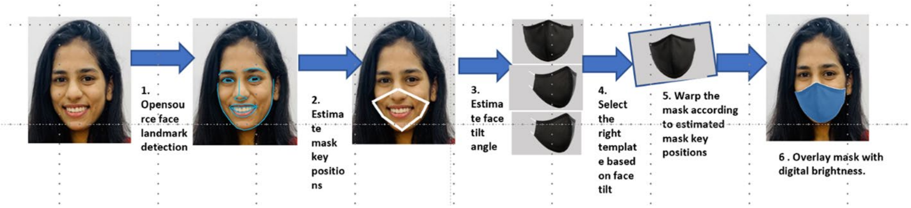
1. Opensou rce face landmark detection
2. Estima te mask key positio ns
3. Estima te face tilt angle
4. Select the right templat e based on face tilt
5. Warp the mask according to estimated mask key positions
6 . Overlay mask with digital brightness.

Fig. 6 Masking the face using Mask-the-face-tool

Due to the blending of the face boundary with the background, the tool failed to detect facial points for applying the correct mask location in some circumstances. Figures 7 and 8 show the failure of tool where faces are with extreme facial tilt, wide open mouth, or heavy face occlusion.

The MasktheFace tool supports multiple mask types, mask variations, supports both single & multi-face images, face angle coverage and bulk masking on the huge dataset. MaskTheFace is used to convert the Faces dataset into a masked face dataset such as MaskedFace-CelebA and MaskedFace-CelebA-HQ in our case. A mask was chosen at random from cloth, surgical-green, surgical-blue, and N95 for each image in the collection. In addition, the original unmasked image was included in the collection with masked images. This was done to ensure that the trained network performs equally well on masked and unmasked images.

5 Experimental setup—image inpainting models

In this section we will discuss the image inpainting models used here for reconstructing the face in detail and the process performed for the evaluation of models. As mentioned in Sect. 4 we have created our own synthetic masked datasets namely MaskedFace-CelebA and MaskedFace-CelebA-HQ based on applying face mask on the face images of benchmark datasets CelebA[28] and CelebA-HQ[24]. Before applying the face masks all the images were resized to 256 × 256 as per the requirement for input to the models. In this study we have performed experiment with Patch Match[5] as non-learning model, EdgeConnect[29] and Gated Convolution[44] as deep learning models. These three models accept masked face and segmented binary map as input and gives reconstructed face as output. For this purpose we have also created a binary map for the occluded area of all the images which are covered with synthetic face mask.

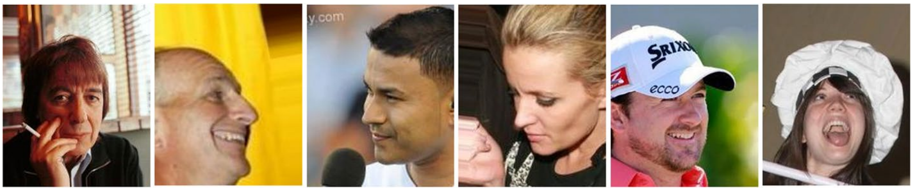

Fig. 7 Failure cases of mask application (MaskedFace-CelebA dataset)

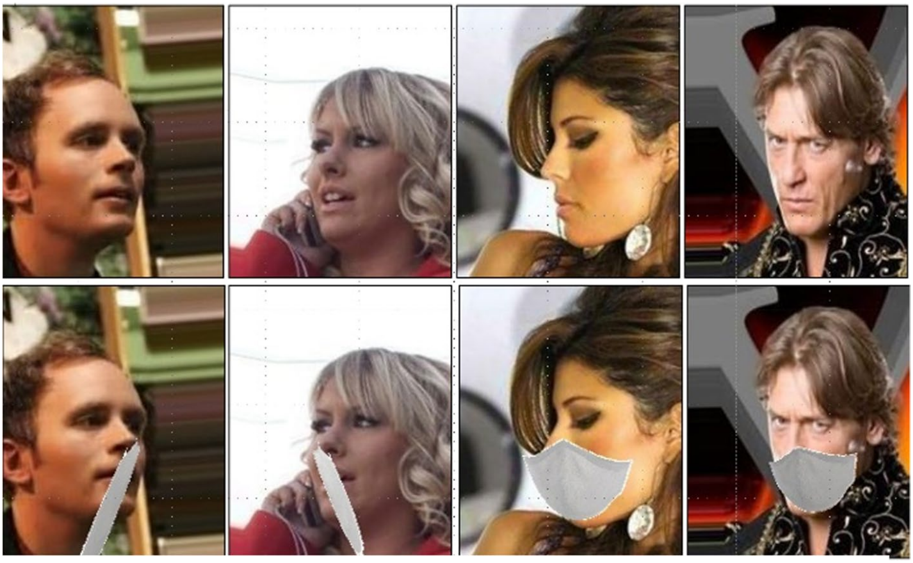

Fig. 8 Failure cases of mask application (MaskedFace-CelebA-HQ dataset)

5.1 Face reconstruction

Any missing parts of the face should be estimated and reconstructed after unmasking to conduct the identity matching procedure and make the recognition decision, i.e., recognized or unrecognized identity. Deep learning methods have addressed such challenges in order to recover the missing part in the facial image. In this section we are discussing the experimental result of three image inpainting models: Patch Match (Non-learning based), Edge Connect [29] method and free form with gated convolution[44] which are deep learning based models.

5.1.1 Patch match model

Patch-based methods, such as PatchMatch [5], employ a copy-paste approach to fill in missing regions by copying information from similar regions within the same image or a collection of images. PatchMatch is commonly used for structural image editing, utilizing a nearest neighbor field (NNF) framework. The algorithm relies on the assumption that good matches can be found through random sampling, and color coherence allows for quick propagation of matching patches to surrounding regions.Despite its success in tasks such as retargeting, hole completion, and content reshuffling, PatchMatch may not be suitable for reconstructing masked facial features, as demonstrated in Fig. 9 of the paper. Specifically, PatchMatch's reliance on picking matching pixels from the NNF for filling the missing region may result in inaccurate reconstruction of facial features, particularly in complex and detailed areas like faces.Notably, the experimental results discussed in the paper pertain specifically to the MaskedFace-CelebA-HQ dataset, and PatchMatch may still perform effectively in other image editing tasks or datasets. However, the limitations of PatchMatch in reconstructing masked facial features highlight the potential need for more advanced,

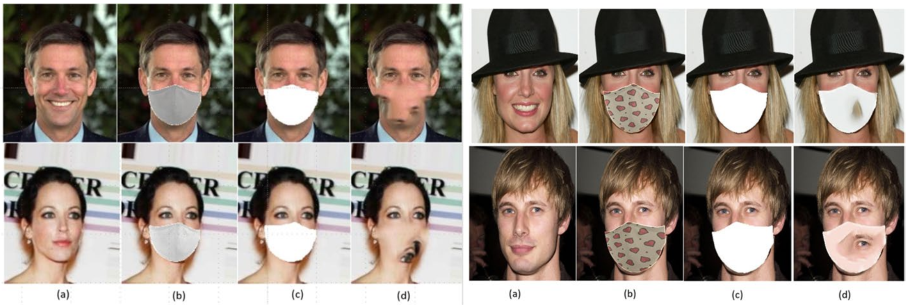

Fig.9 Reconstructed face using PatchMatch [5] model on MakedFace-CelebA (left) and MaskedFace-CelebA-HQ (right) dataset (a) Ground Truth (b) Masked Image (c)Image with Mask Overlay (d) Reconstructed Face using Patchmatch[5]

learning-based methods that can better capture the nuances of facial structures and textures for improved accuracy and visual quality.

5.1.2 Edge connect

Edge-Connect technique proposed by Yu, J. et.al [43]. works on the concept of 'Lines first, color next', is inspired by artists' work. The major contribution of the work is an edge generator that is capable of hallucinating edges in missing regions given edges and grayscale pixel intensities of the rest of the image. An image completion network is also proposed that combines edges in the missing regions with color and texture information to fill the missing region. The original work of authors of this model is to fill in missing regions demonstrating fine details an end-to-end trainable network is proposed that combines edge generation and image completion. We have used the same architecture for the purpose of reconstructing the face using the pre-trained celebA dataset which is trained using irregular form of mask.

The detailed architecture is shown in Fig. 10 which demonstrates an edge-to-image two-stage network i.e. two generators and two discriminators. To obtain the overall image structure, the first generator is in task of predicting the edges of the missing regions. The predicted edge map is a binary map that depicts an image's skeleton. The second generator is based on the predicted edge map and is in responsibility of filling in the missing texture details in the missing regions.

The first generator G1 produces a predicted edge map using the mask image, masked edge image, and masked grayscale image as input. The conventional adversarial loss and the feature matching loss are used to train this generator. The second generator G2 takes the expected edge map and the masked RGB image as input and generates a finished RGB image. The style loss, perceptual loss, L1 reconstruction loss, and conventional adversarial loss are all used to train this generator.

As per the requirement of network architecture an 8-bit binary mask and 24-bit overlayed mask image of 256×256 resolution is given as input and the completed image is given as output by the system. As obvious the model shown better result on MaskedFace-CelebA dataset as compare to MaskedFace-CelebA-HQ face dataset as the model is trained on CelebA face datasets (Fig. 11).

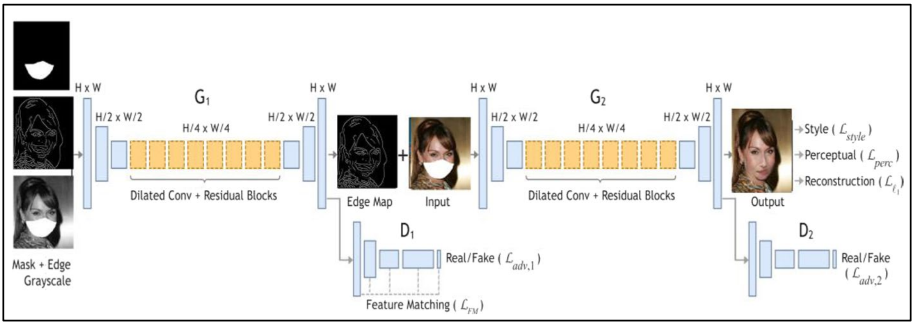
Mask + Edge
Grayscale
H x W
H/2 x W/2
G_1
H/4 x W/4
H/2 x W/2
H x W
Dilated Conv + Residual Blocks
Edge Map
+
Input
H x W
H/2 x W/2
G_2
H/4 x W/4
H/2 x W/2
H x W
Dilated Conv + Residual Blocks
Output
Style ($\mathcal{L}_{style}$)
Perceptual ($\mathcal{L}_{perc}$)
Reconstruction ($\mathcal{L}_{\ell_1}$)
D_1
Real/Fake ($\mathcal{L}_{adv,1}$)
Feature Matching ($\mathcal{L}_{FM}$)
D_2
Real/Fake
($\mathcal{L}_{adv,2}$)

Fig. 10 Network Architecture—Edge Connect. Incomplete grayscale image and edge map, and mask are the inputs of G1 to inputs of G1 to predict the full edge map. Predicted edge map and incomplete color image are passed to G2 to perform the inpainting task. (Architecture from original paper, Image from own experiment)

As per results mentioned in Table 1 in terms of PSNR,SSIM, FID and MAE, our work shows better values for image quality metrics except FID in comparison to results taken from original EdgeConnect[29] paper.

5.1.3 Free form image inpainting with gated convolution

Yu. Jiahui et.al [44]. proposed a generative image inpainting system to complete images with a free-form mask and guidance. It is a blending of the Contextual Attention (CA) layer proposed [43] in DeepFill v1, the concept of user guidance (optional user sketch input) introduced in EdgeConnect, and Partial Convolution (PConv) modified to Gated Convolution (GConv), in which rule-based mask update is formulated as a learnable gating to the next convolution layer as shown in Fig. 12. The model uses a coarse-to-fine network structure in two stages. The coarse reconstruction is managed by the first generator network, while the refinement of the coarse filled picture

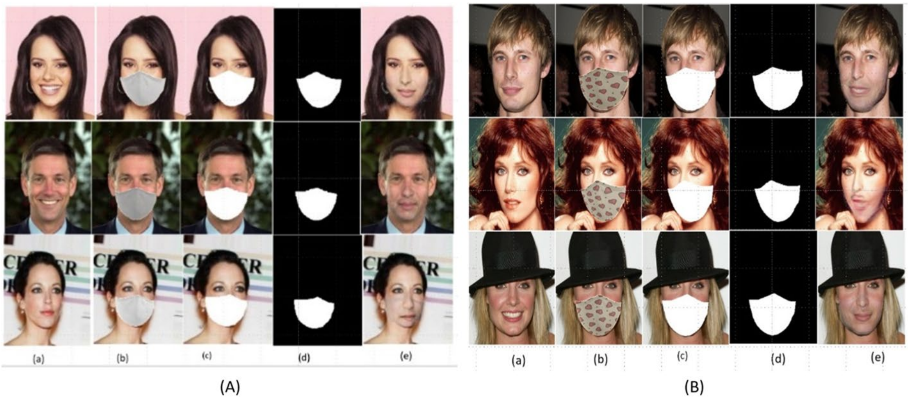

Fig. 11 Face reconstruction (A) Edge Connect[29] model evaluated on CelebA masked-face dataset (a) Ground Truth (b) Masked Image (c) Image with Mask Overlay (d) Image Mask (e) Reconstructed Face (B) Gated Convolution [44] model evaluated on CelebA-HQ masked-face dataset (a) Ground Truth (b) Masked Image (c) Image with Mask Overlay (d) Image Mask (e) Reconstructed Face

Table 1 Comparison of Image Quality metrics with state-of-the-art EdgeConenct[29] with our work

Metrics	Original Work EC[29]	Our Work using EC[29]
PSNR	25.28	**27.68**
SSIM	0.846	**0.962**
FID	**2.82**	4.32
MAE	0.846	**0.0298**

is performed by the second generator network. The network is trained using only the two most common loss terms, the L1 loss and the GAN loss. This is one of the paper's assertions, as other state-of-the-art inpainting papers train their networks with up to 6 loss terms.

For a coarse reconstruction of the missing regions, the coarse generator uses the masked image, mask image, and an optional user sketch image as input. A standard convolutional layer followed by a sigmoid activation function is used for updating the mask instead of a rule-based mask update in Partial Convolution and learning for Gated Convolution. The same architecture as shown in Fig. 9 is used for reconstructing the faces after removing the mask from face. The model is evaluated on the masked-face dataset of 9621 images (MaskedFace-CelebA faces) and 9883 images (MaskedFace-CelebA-HQ faces) and the result shows better performance on MaskedFace-CelebA-HQ dataset (Fig. 13,14) as the pre-trained model trained on CelebA-HQ is used for experiment.

6 Results and discussion

We evaluate the models discussed above on the masked-face datasets MaskedFace-CelebA, and MaskedFace-CelebA-HQ. The result is reported in terms of qualitative and quantitative results in sub-sections of this section.

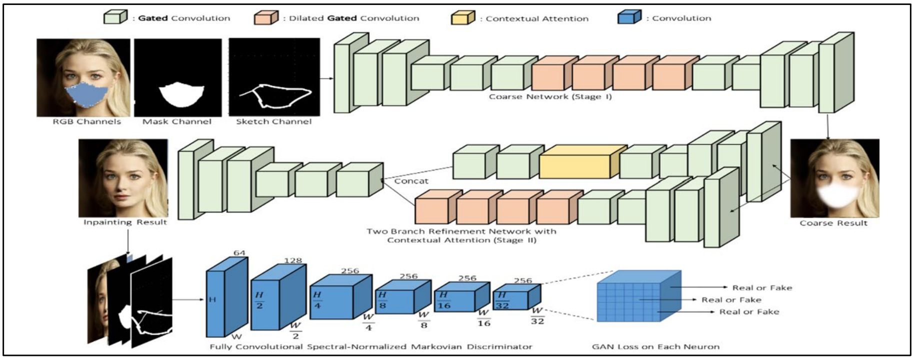

Fig. 12 Network Architecture: Free form Image Inpainting with Gated Convolution (As per Original literature)

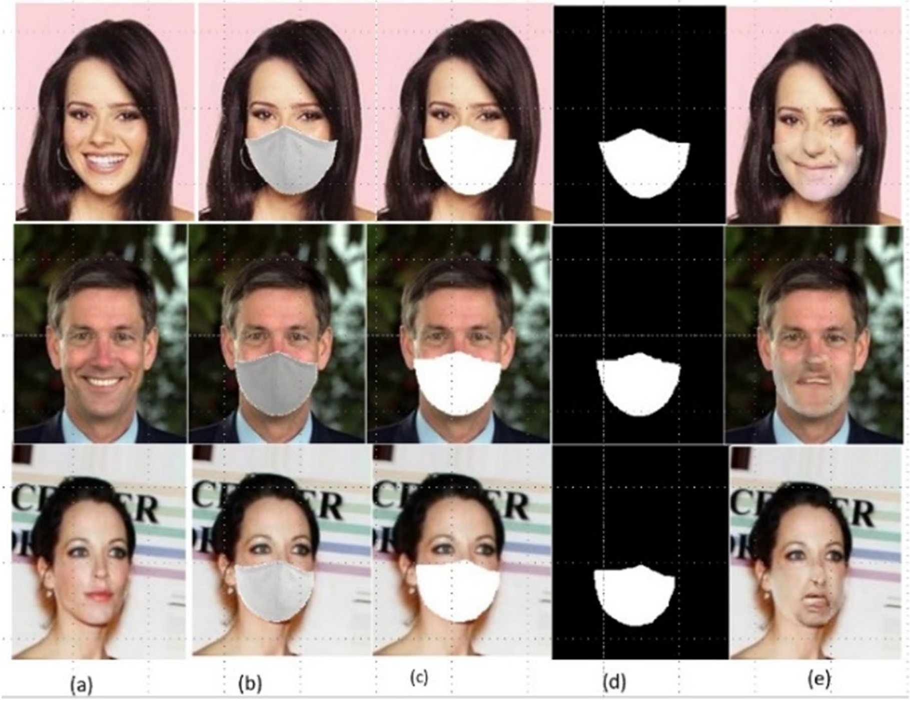

Fig. 13 Face reconstruction using Gated Convolution[44] model evaluated on MaskedFace-CelebA

6.1 Qualitative results

The masked face datasets were evaluated using three different methods: PatchMatch [5] as a traditional or non-learning based method, and two deep learning based methods, Edge Connect [29] and Gated Convolution [44], with the same architecture and pre-trained models for reconstructing the face. The results are shown in Figs. 15 and 16 for the MaskedFace-CelebA masked face dataset, and in Fig. 17 for the MaskedFace-CelebA-HQ masked face dataset.

The results indicate that PatchMatch [5] fails to accurately fill in the missing pixels with predicted facial features, whereas the learning-based methods show improved results. PatchMatch [5] also did not perform well on the MaskedFace-CelebA-HQ dataset compared to the MaskedFace-CelebA dataset. It is worth noting that image inpainting techniques may not yield promising results when used for face generation after removing a large occlusion such as a mask. The experiments demonstrate that PatchMatch [5] may not be suitable for the purpose of regenerating masked facial features, and the deep learning-based methods Edge Connect [29] and Gated Convolution [44] show better performance in this task.Fig. 19Failure cases: Reconstructed faces from Edge Connect Model (MaskedFace-CelebA-HQ Dataset)Fig. 19Failure cases: Reconstructed faces from Edge Connect Model (MaskedFace-CelebA-HQ Dataset)Fig. 19Failure cases: Reconstructed faces from Edge Connect Model (MaskedFace-CelebA-HQ Dataset)

The performance of the generated faces is generally better on the model for which they are trained, resulting in images that are close to the ground truth. However, there are still

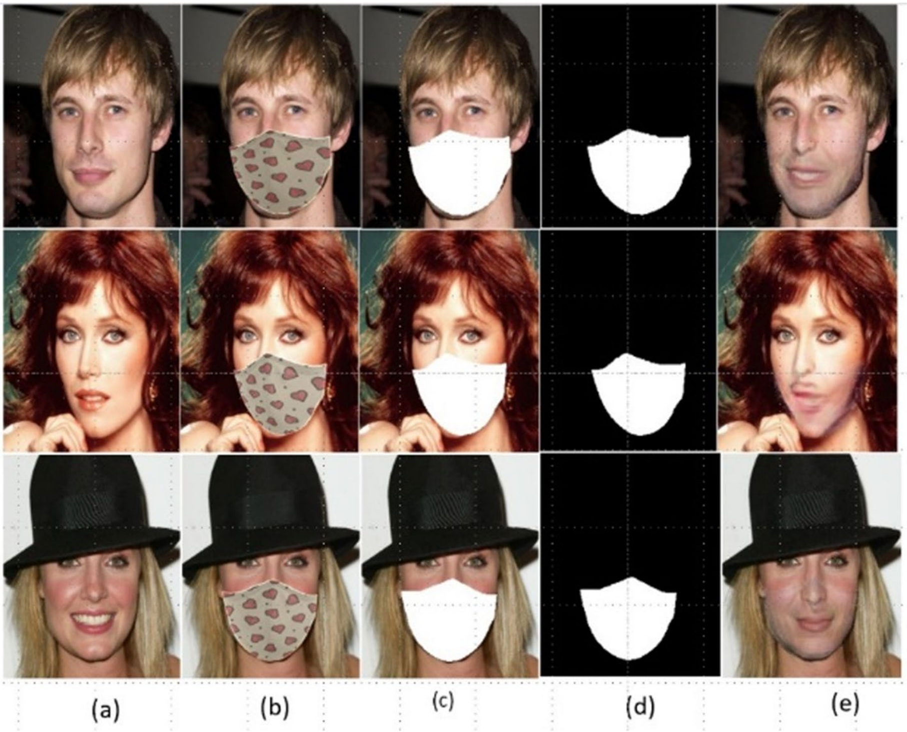

Fig. 14 Gated Convolution[44] model evaluated on MaskedFace- CelebA-HQ masked-face dataset

cases where the reconstructed faces may not be well-generated, especially in scenarios such as illumination changes, improper face detection, or merging of the face with the background. These issues are illustrated in Figs. 17, 18, 19 and 20, showing examples of images where the generated faces may not be accurately reconstructed.

6.2 Quantitative results

We compare the results of the experiments performed over three models: Patch Match [5], Edge Connect [29] and Gated Convolution [44] in terms of reconstructed face image quality. The quantitative result of these models is reported in terms of image quality assessment metrics as given below:

1. **Peak signal-to-noise ratio (PSNR)**[36]**:** PSNR is a full reference metrics and measures the quality of the generated image by comparing it to the ground truth image in terms of the signal-to-noise ratio. It quantifies the amount of noise or distortion present in the generated image compared to the original image. Higher PSNR values indicate lower distortion and better image quality, indicating a closer resemblance to the ground truth. The original image matrix and the degraded image matrix must have the same dimensions. The following is a definition:

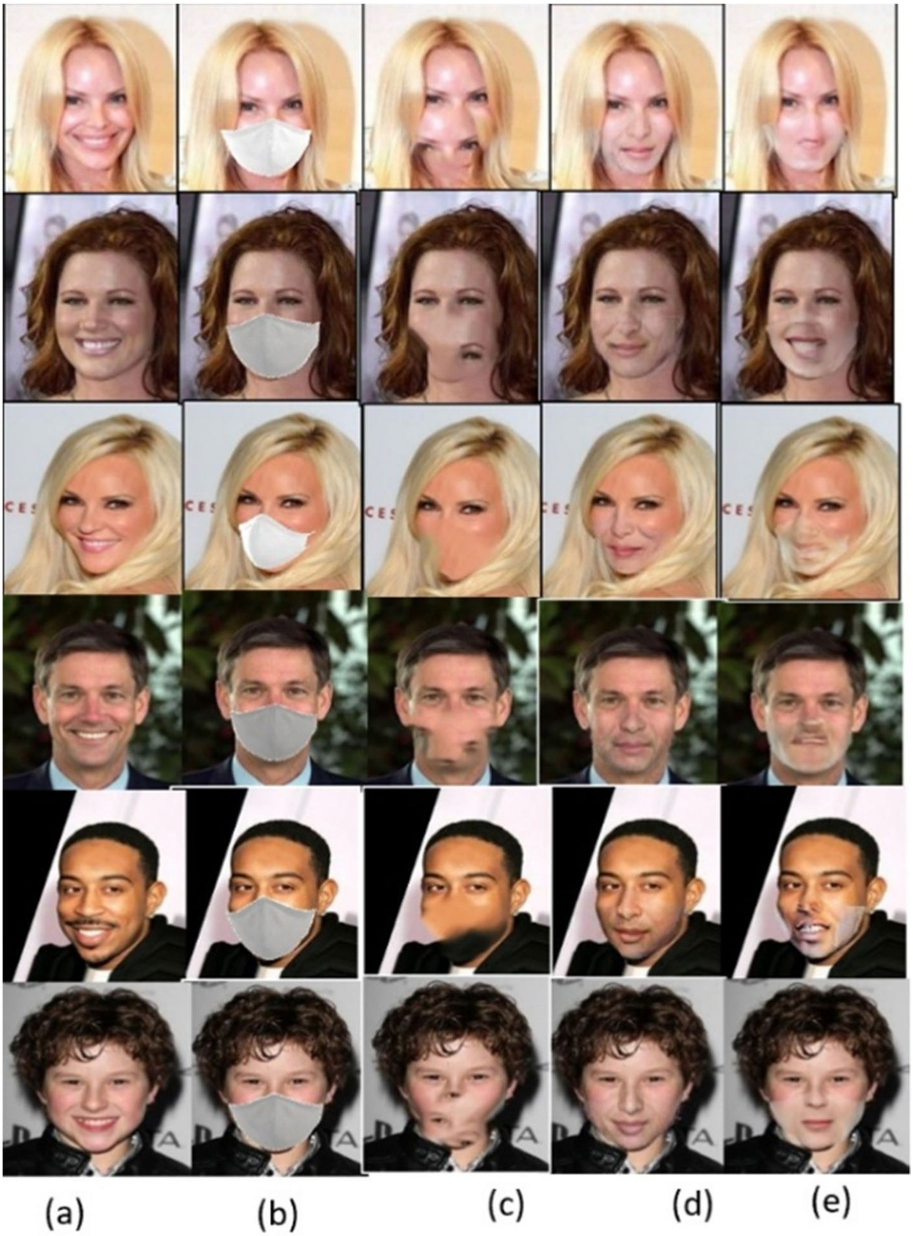

Fig. 15 Comparison of qualitative results with existing models on CelebA dataset (a) Ground Truth Image from CelebA Dataset. (b) Image with mask (c) PatchMatch [5] (d) Edge Connect [29]. (e) Gated Convolution [44]

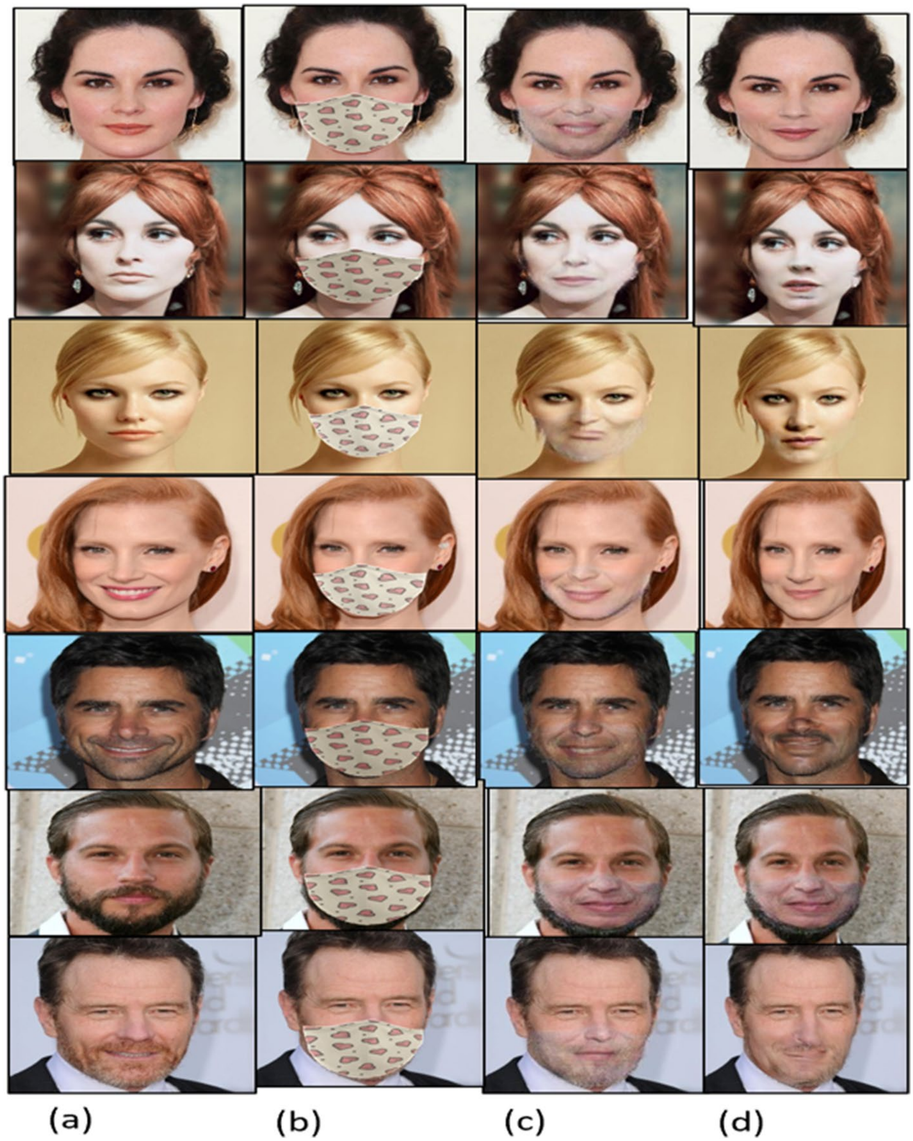

Fig. 16 Comparison of qualitative results with existing models on CelebA-HQ dataset (a) Ground Truth Image from CelebA-HQ Dataset. (b) Image with mask (c) Edge Connect [29]. (d) Gated Convolution [44]

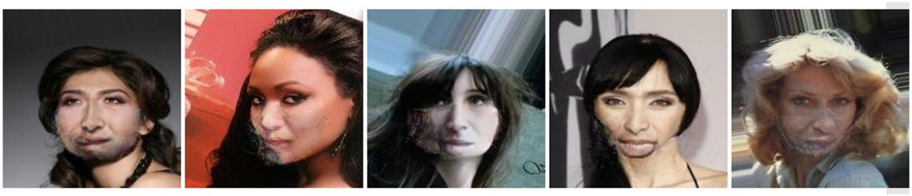

Fig. 17 Failure cases: Reconstructed faces from Edge Connect Model (MaskedFace-CelebA Dataset)

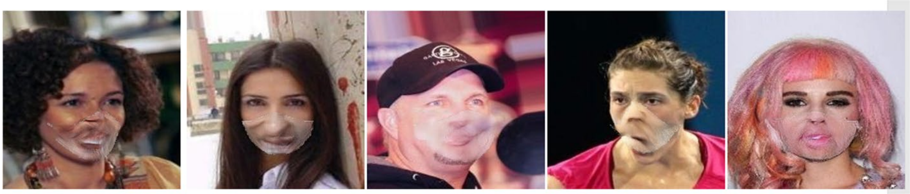

Fig. 18 Failure cases: Reconstructed faces from Gated Convolution Model (MaskedFace-CelebA Dataset)

$$\mathrm{PSNR} = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \tag{1}$$

where MAXf is the maximum signal value that exists in the original image, and mean squared error (MSE) is calculated as follows:

$$\mathrm{MSE} = \frac{1}{mn}\sum\nolimits_0^{m-1}\sum\nolimits_0^{n-1}\|f(i,j) - g(i,j)\|^2 \tag{2}$$

where f represents the matrix data of the original image, g represents the matrix data of the degraded image, m denotes the numbers of pixel rows of the image, i denotes the index of each row, n represents the number of pixel columns of the image, and j represents the index of each column.

2. **Structural SIMilarity (SSIM)**[1]**:** SSIM is a full reference metric and a measure of how similar the structure of the generated image is to the ground truth image. It considers the luminance, contrast, and structural similarities between the two images. Higher SSIM values indicate a higher degree of similarity, indicating better quality of the generated image in terms of structural consistency with the ground truth. It can be defined as follows:where μ denotes the mean value of a given image and σ is the standard deviation of the image; x and y represent the two images being compared; c1 and c2 are constants to guarantee stability when the divisor becomes 0.

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{3}$$

3. **Mean Absolute Error (MAE):** MAE measures the average absolute pixel-wise difference between the generated image and the ground truth image. It provides a quantitative

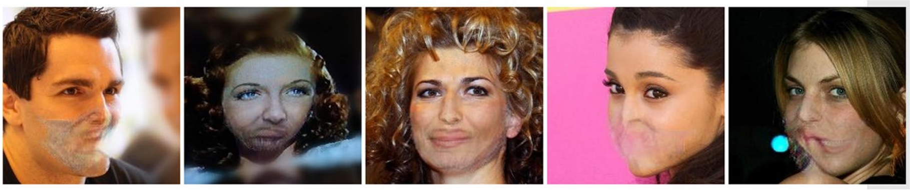

Fig. 19 Failure cases: Reconstructed faces from Edge Connect Model (MaskedFace-CelebA-HQ Dataset)

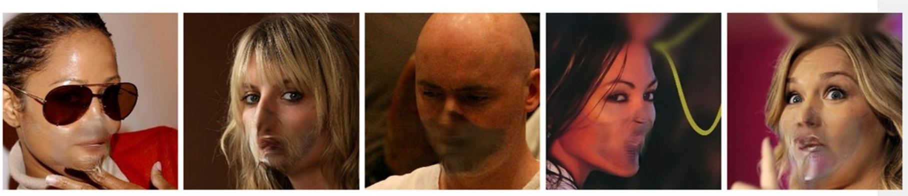

Fig. 20 Failure cases: Reconstructed faces from Gated Convolution Model (MaskedFace-CelebA-HQ Dataset)

measure of the overall pixel-level accuracy of the generated image. Lower MAE values indicate less discrepancy between the generated image and the ground truth, indicating higher image quality and better reconstruction.where r and g are the real and fake embeddings, and μr and μg are the magnitudes of the vectors r and g. Tr is the trace of the matrix, and ∑r and ∑g represent the covariance matrix of vectors [72].

$$MAE = \left(\frac{1}{N}\right) * \sum |GT - GENERATED| \tag{4}$$

4. **Fréchet Inception Distance (FID)**[17]: FID is a measure of the dissimilarity between the distribution of feature representations from the generated image and the ground truth image, as extracted from a pre-trained Inception-v3 model[23]. Lower FID values indicate a smaller distance between the distributions, indicating a higher level of similarity between the generated image and the ground truth, and hence better image quality in terms of distributional similarity.The following is a definition:where r and g are the real and fake embeddings, and μr and μg are the magnitudes of the vectors r and g. Tr is the trace of the matrix, and ∑r and ∑g represent the covariance matrix of vectors [72].

$$\text{FID} = \|\mu_r - \mu_g\|^2 + T_r\left(\sum_r + \sum_g -2(\sum_r \sum_g)^{1/2}\right) \tag{5}$$

Note that the Table 2 report our evaluation in terms of Peak signal-to-noise ratio PSNR, structural similarity index (SSIM),Mean absolute error MAE and FID on two masked-face datasets namely MaskedFace-CelebA and MaskedFace-CelebA-HQ separately using three models Patch Match [5], Edge Connect [29] and Gated Convolution [44]. Recent researches [19, 34, 35] have shown that metrics based on deep features are closer to those based on human perception.

The results from the Table 2 indicate that the GC[44] model generally performs better in terms of face reconstruction accuracy compared to the PM[5] and EC[29] models when evaluated on the MaskedFace-CelebA and MaskedFace-CelebA-HQ masked-face datasets. GC[44] has higher PSNR and SSIM values, which indicate better image quality and structural similarity, and lower MAE and FID values, which indicate lower error and better similarity to ground truth images. These results suggest that GC[44] may be a promising choice for face reconstruction tasks, as it demonstrates superior performance in multiple evaluation metrics. However, this is only a preliminary study which needs to be corroborated with further investigations. Therefore, these datasets are also subject to face recognition to match the reconstructed face with its ground truth using efficient face recognition algorithm. Which is reported in Sect. 6.

Table 2 Quantitative results over MaskedFace-CelebA and MaskedFace-CelebA-HQ with models: Patch Match (PM) [5], Free-Form Image Inpainting with Gated Convolution (GC) [44], Edge Connect—Generative Image Inpainting with Adversarial Edge Learning (EC) [29]. The best result of each boldfaced Lower is better. *Higher is better

Models	MaskedFace-CelebA dataset			MaskedFace-CelebA-HQ dataset		
	PM[5]	EC[29]	GC[44]	PM[5]	EC[29]	GC[44]
PSNR*	25.76	**27.68**	26.77	18.02	26.53	**28.1843**
SSIM*	0.90	0.950	**0.971**	0.7891	0.943	**0.9390**
Mean absolute error (MAE)	**0.0197**	0.0298	0.0198	0.0559	0.049	**0.0168**
Fréchet Inception Distance (FID)	18.7186	**4.3297**	**10.3297**	10.0352	8.012	**2.012**

7 Face recognition

In this sub-section, we report our work in terms of face recognition, where we classify the faces as ground truth or reconstructed images. For this task, we utilized the VGG16[37] convolutional neural network (CNN) architecture pretrained on Imagenet dataset[33], which is a popular deep learning model for image classification tasks. We trained the classifier model using 80% of the input data and tested it on the remaining 20% of the input data, following an 80–20% split for both the Masked-Face-CelebA and Masked-Face-CelebA-HQ datasets. We provide a brief introduction to the VGG16 classifier model, which is used in our work to carry out classification between the ground truth images and the reconstructed image sets.

7.1 The VGG16 model

We employed the VGG16 convolutional neural network (CNN) architecture, which is a widely-used deep learning model for image classification tasks. The VGG16 model comprises 16 weight layers, including 13 convolutional layers and 3 fully connected layers. It is known for its deep architecture, which allows it to capture complex features from images. To adapt the VGG16 model for our specific classification task, we removed the last fully connected layers of the original model and replaced them with custom layers. The modified VGG16 model was then fine-tuned on our dataset, which consisted of ground truth and reconstructed face images.

The input images were resized to the required input shape of the VGG16 model, typically 224×224 pixels, and were passed through the convolutional layers to extract high-level features. The ReLU activation function was applied after each convolutional layer to introduce non-linearity. Global average pooling was then applied to reduce the spatial dimensions of the features, resulting in a fixed-size vector. This vector was fed into custom fully connected layers with appropriate activation functions and output units for the final classification.

During training, we used a batch size of 100 and performed 30 epochs of training. We employed the Adam optimizer with a learning rate of 0.001 for gradient descent. To enhance the model's robustness to different variations in the input images, we applied data augmentation techniques such as rotation, horizontal flip, and zoom. The trained

Table 3 Face recognition VGG16 classifier for Ground Truth & Reconstructed Images

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
MaskedFace-CelebA	EC[29]	56.56	56.56	1.0	72.25
MaskedFace-CelebA	GC[44]	48.82	48.82	1.0	65.61
MaskedFace-CelebA-HQ	EC[29]	49.75	49.45	92.6	66.15
MaskedFace-CelebA-HQ	GC[44]	47.25	47.25	1.0	64.18

VGG16 model achieved an accuracy as shown in Table 3 on our validation set for the tested models Edge Connect and Gated Convolution on MaskedFace-CelebA and MaskedFace-CelebA-HQ datasets, respectively. These results demonstrate the effectiveness of our VGG16-based approach in classifying ground truth and reconstructed face images in our specific context.

As with any classifier model, the performance of the VGG16 classifier in the present paper's face recognition system was evaluated using several commonly used performance metrics, including Accuracy, Precision, Recall, and F1 score and their brief description is given below:

Accuracy Accuracy is a widely used metric for evaluating the performance of classifier models. It gauges the ability of the classifier to correctly identify both positive and negative samples in the dataset. The accuracy of the model is calculated as the ratio of the number of correct predictions (i.e., true positives and true negatives) to the total number of samples in the dataset. The formula for calculating accuracy is:

This metric provides an overall assessment of the model's performance in accurately classifying ground truth and reconstructed face images in the context of face recognition. It is an important evaluation metric that reflects the VGG16 classifier's ability to accurately classify samples in the dataset.

Precision Precision is a useful metric to consider when the consequences of false positives are significant. It evaluates the accuracy of the classifier in avoiding mislabelling negative samples as positive. Precision measures how precise the model is in terms of the proportion of predicted positives that are truly positive. It is calculated as:

A higher precision value indicates that the classifier is making fewer false positive predictions, which can be crucial in situations where misclassification costs are high. This metric provides insights into the classifier's ability to minimize false positives, where the model incorrectly identifies a negative sample as positive.

Recall Recall, also known as sensitivity or true positive rate, measures the ability of the classifier to correctly capture actual positives by labelling them as positive (true positives). Recall is a relevant metric to consider when the cost of false negatives is high, as it reflects the classifier's capacity to identify all positive samples. Recall is calculated as:

A higher recall value indicates that the classifier is able to capture more true positives, which can be crucial in situations where missing positive samples is costly. This metric provides insights into the classifier's ability to minimize false negatives, where the model incorrectly identifies a positive sample as negative.

The F1-Score Also known as the F-Score, is a metric used to evaluate the accuracy of a model on a binary classification dataset. It measures how well the model performs in classifying examples into 'positive' and 'negative' categories. The F-Score is calculated as the harmonic mean of the model's Precision and Recall, and it provides a way to combine both Precision and Recall into a single value. A perfect F1-Score is 1.0, indicating perfect Precision and Recall, while a value of 0 indicates that either Precision or Recall is zero.

The Accuracy, Precision, Recall, and F1-Score are comprehensive performance metrics for evaluating face recognition systems. They are used to correctly identify reconstructed faces as distinct from the ground truth faces. In our experiments, the set of parameters used for image quality assessment (IQA), including PSNR, SSIM, FID, and MAE, may behave inversely with respect to the four classifier parameters. Table 4 presents the compiled values generated by the VGG16 classifier, with results separately reported for EdgeConnect [29] and Gated Convolution [44] models. It's important to note that Patch Match [5] was found inadequate in the initial part of our work, and thus the VGG16 classifier was not tested on Patch Match results.

$$\text{Accuracy} = \frac{(True\ Positive + True\ Negative)}{(True\ positive + False\ Positive + True\ Negative + False\ Negative)} \tag{6}$$

$$\begin{aligned}\text{Precision} &= \frac{True\ Positive}{True\ Positive + False\ Positive}\\ &= \frac{True\ Positive}{True\ Predicted\ Positive}\end{aligned} \tag{7}$$

$$\begin{aligned}\text{Recall} &= \frac{True\ Positive}{True\ Positive + False\ Negative}\\ &= \frac{True\ Positive}{True\ Actual\ Positive}\end{aligned} \tag{8}$$

$$\text{F1 Score} = 2\text{x}\frac{Precision * Recall}{Precision + Recall} \tag{9}$$

VGG16 is widely used as a powerful image classifier in applications such as object recognition, scene classification, emotion recognition, and disease diagnosis. Its deep architecture and ability to learn complex features from images make it highly effective in accurately categorizing images. It finds applications in healthcare for classifying medical images, automotive industry for object recognition in autonomous vehicles, and entertainment industry for scene classification in video processing. VGG16's versatility as a classifier has made it a popular choice among researchers and practitioners in various fields where accurate image classification is essential. The results mentioned in Table 3 are clearly contrary to its performance as stated above. For MaskedFace-CelebA dataset gives 56.56% and 48.62% recognition accuracy for EC [29] and GC [44] respectively whereas the accuracy recorded for MaskedFace-CelebA-HQ is 49.75% and 47.25% for EC [29] and GC [44] respectively. The lower accuracy of the VGG16 classifier in classifying between ground truth and reconstructed facial images suggests that the distinction between these two sets may not be clear. This could be due to the fact that when ground truth and reconstructed faces are visually similar, classifiers may struggle to differentiate between them. This implies that the reconstructed facial images generated by EC [29] and GC [40] models may closely resemble their ground truth counterparts, as the classifier is unable to accurately distinguish between them. This is because, the recognition accuracy is poor in both these cases. Therefore, the computed quantitative metrics – PSNR (dB) and SSIM outcomes must also corroborate with classifier outcomes. So far as PSNR (dB) is concerned, the computed values are in the vicinity of 25–28 dB. On the other hand, the SSIM computed value is 0.971 & 0.974 respectively.

The high SSIM value indicates that the ground truth and reconstructed facial images are visually similar, as SSIM measures image quality or correlation with values ranging from 0 to 1, where 1 represents higher image quality. On the other hand, PSNR does not indicate the same, as it is dependent on MSE and amplifies errors between images. The computed PSNR values in the present work are within the range of 25–28 dB, which suggests that the ground truth and reconstructed images are distinct. However, PSNR is not considered as a robust Image Quality Assessment (IQA) metric due to its dependence on MSE and has been debated in the research community. SSIM, being a structural property-based metric, is considered more robust for IQA.

Based on the analysis, it is concluded that the computed outcomes of SSIM, FID, and VGG16 classifier accuracy align with each other, indicating similar results. The lower face recognition accuracy of the VGG16 classifier coinciding with higher SSIM values suggests better resemblance between the ground truth and reconstructed images.

The analysis suggests that out of the three models tested in the present work—Patch Match[5], EdgeConnect Model[29], and Gated Convolution Model [44], Patch Match is not found to be suitable for facial image reconstruction, while EdgeConnect model performs better on the Masked Face CelebA dataset and Gated Convolution model performs better on the MaskedFace-CelebA-HQ dataset.

8 Conclusion & Future work

In this study, a comparative analysis of image inpainting models was conducted using synthetic masked datasets, namely MaskedFace-CelebA and MaskedFace-CelebA-HQ. The traditional PatchMatch model yielded unsatisfactory results for face reconstruction, as it is primarily used for restoring images or removing blocked patches from images based on neighbourhood information. However, deep learning-based models such as EdgeConnect and Gated Convolution produced plausible images and outperformed traditional patch-based methods in quantitative comparisons.

Despite the promising results on synthetic datasets, the deep learning-based models faced challenges when dealing with real masked faces, such as lack of proper face visibility and illumination. Face recognition results showed synchronized performance in terms of differentiating ground truth images from reconstructed ones, as indicated by IQA and FID metrics.

It's important to note that the models used in this study were originally designed for removing face occlusions or reconstructing natural scenes, and were repurposed for masked face reconstruction. Further improvements can be made by training on new datasets, fine-tuning hyperparameters, and exploring the use of generative diverse image inpainting techniques for real masked face recognition in future work.

Acknowledgements We extend our sincere appreciation to Professor Anurag Mishra , Deen Dayal Upadhyay College, Delhi University, for his invaluable guidance in overseeing and meticulously reviewing our study proposal, as well as providing technical editing support for our manuscript. We are grateful for his contributions to our work.

Data availability The MaskedFace-CelebA and MaskedFace-CelebA-HQ synthetic masked datasets generated during and/or analysed during the current study are available in the ***Synthetic masked dataset*** repository, link: Google Drive[1]. The folder includes Masked face dataset and respective binary map of mask for MaskedFace-CelebA and MaskedFace-CelebA-HQ datasets in separate folders. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request and were derived from the CelebA[28] and CelebA-HQ[24] public datasets.

Declarations

Funding and/or conflicts of interests/competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors did not receive support from any organization for the submitted work.

References

1. All about Structural Similarity Index (SSIM): Theory + Code in PyTorch. Available online: https://medium.com/srm-mic/allabout-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e .
2. Anwar Aqeel, and Arijit Raychowdhury (2020)"Masked Face Recognition for Secure Authentication."&nbsp;arXiv preprint arXiv:2008.11104. https://github.com/aqeelanwar/MaskTheFace
3. Arya KV and Bhadoria RS eds (2019). The Biometric Computing: Recognition and Registration. CRC Press
4. Ballester C, Bertalmio M, Caselles V, Sapiro G, Verdera J (2001) Filling-in by joint interpolation of vector fields and gray levels. IEEE Trans Image Process 10(8):1200–1211. https://doi.org/10.1109/83.935036
5. Barnes C, Shechtman E, Goldman DB, Finkelstein A (2010) The generalized patchmatch correspondence algorithm. In European Conference on Computer Vision 29–43. Springer, Berlin, Heidelberg
6. Barnes C, Shechtman E, Finkelstein A, Goldman DB (2009) PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans Graph 28(3):24
7. Bertalmio M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 417–424). https://doi.org/10.1145/344779.344972
8. Buades, A., Coll, B., & Morel, J. M. (2005, June). A non-local algorithm for image denoising. In&nbsp;2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)&nbsp;(Vol. 2, pp. 60-65). IEEE. https://doi.org/10.1109/CVPR.2005.38
9. Cai W, Wei Z (2020).&nbsp;PiiGAN: Generative Adversarial Networks for Pluralistic Image Inpainting. IEEE Access, 8, 48451–48463. https://arxiv.org/abs/1912.01834v2
10. Chen F, Zhang T, Liu H (2022) Face image inpainting via latent features reconstruction and mask awareness. Comput Electr Eng 103:108282
11. Criminisi A, Pérez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13(9):1200–1212

[1] For Dataset access write to chandni.officialid@gmail.com.

12. Efros AA, Freeman WT (2001) Image quilting for texture synthesis and transfer. In&nbsp;Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 341–346. https://doi.org/10.1145/383259.383296
13. Efros AA, Leung TK (1999). Texture synthesis by non-parametric sampling. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1033–1038). IEEE
14. Elharrouss O, Almaadeed N, Al-Maadeed S, Akbari Y (2020) Image inpainting: A review. Neural Process Lett 51:2007–2028
15. Elharrouss O, Almaadeed N, Al-Maadeed S, & Akbari Y (2020). Image inpainting: A review.&nbsp;Neural Process. Lett,&nbsp;51(2), 2007–2028. https://arxiv.org/abs/1909.06399v1
16. Freeman WT, Jones TR, Pasztor EC (2002) Example-based super-resolution. IEEE Comput Graphics Appl 22(2):56–65
17. How to Evaluate GANs Using Frechet Inception Distance (FID). Available online: https://wandb.ai/ayush-thakur/ganevaluation/reports/How-to-Evaluate-GANs-using-Frechet-Inception-Distance-FID---Vmlldzo0MTAxOTI .
18. Iizuka S, Simo-Serra E, Ishikawa H (2017) Globally and locally consistent image completion. ACM Transactions on Graphics (ToG) 36(4):1–14. https://doi.org/10.1145/3072959.3073659
19. Inamdar M, Mehendale N (2020) Real-time face mask identification using facemasknet deep learning network. Available at SSRN 3663305.
20. Jain V, Seung S (2008) Natural image denoising with convolutional networks. Adv Neural Inf Process Syst, 21
21. Jiang M, Fan X, Yan H (2020) Retinamask: A face mask detector
22. Jiang Y, Xu J, Yang B, Xu J, Zhu J (2020) Image Inpainting Based on Generative Adversarial Networks. IEEE Access 8:22884–22892. https://doi.org/10.1109/ACCESS.2020.2970169
23. Jignesh Chowdary G, Punn NS, Sonbhadra SK, Agarwal S (2020) Face mask detection using transfer learning of inceptionv3. In Big Data Analytics: 8th International Conference, BDA 2020, Sonepat, India, December 15–18, 2020, Proceedings 8 (pp. 81–90). Springer International Publishing
24. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality, stability, and variation.&nbsp;https://arxiv.org/abs/1710.10196v3
25. Li Y, Liu S, Yang J, Yang MH (2017). Generative face completion. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3911–3919)
26. Liu G, Reda FA, Shih KJ, Wang TC, Tao A, Catanzaro B (2018) Image inpainting for irregular holes using partial convolutions. In&nbsp;Proceedings of the European Conference on Computer Vision (ECCV)&nbsp;(pp. 85–100). https://arxiv.org/abs/1804.07723v2
27. Liu H, Wan Z, Huang W, Song Y, Han X, Liao J (2021). PD-GAN: Probabilistic Diverse GAN for Image Inpainting.&nbsp;arXiv preprint arXiv:2105.02201
28. Liu Z, Luo P, Wang X, Tang X (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision 3730–3738
29. Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M (2019) Edgeconnect: Generative image inpainting with adversarial edge learning. https://arxiv.org/abs/1901.00212v3
30. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA (2016) Context encoders: Feature learning by inpainting. In&nbsp;Proceedings of the IEEE conference on computer vision and pattern recognition&nbsp;(pp. 2536–2544). https://arxiv.org/abs/1604.07379v2
31. Qin Z, Zeng Q, Zong Y, Xu F (2021) Image inpainting based on deep learning: A review. Displays 69:102028
32. Rother C, Bordeaux L, Hamadi Y, Blake A (2006). Autocollage.&nbsp;ACM transactions on graphics (TOG),&nbsp;25(3), 847-852. https://doi.org/10.1145/1141911.1141965
33. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Fei-Fei L (2015) Imagenet large scale visual recognition challenge. Int J Comput. Vis. 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y
34. Sengar N, Singh A, Yadav S, Dutta MK (2022). Automated System for Face-Mask Detection Using Convolutional Neural Network. In Proceedings of the Seventh International Conference on Mathematics and Computing: ICMC 2021 (pp. 373–380). Singapore: Springer Singapore
35. Sethy PK, Bag S, Panigrahi M, Behera SK, Rath AK (2022) Face Mask Detection in Public Places Using Small CNN Models. In Intelligent and Cloud Computing: Proceedings of ICICC 2021 (pp. 317–325). Singapore: Springer Nature Singapore
36. Signal-to-Noise Ratio as an Image Quality Metric. Available online: https://www.ni.com/en-in/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html.
37. Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

38. Wang N, Ma S, Li J, Zhang Y, Zhang L (2020) Multistage attention network for image inpainting. Pattern Recogn 106:107448
39. Xie J, Xu L, Chen E (2012). Image denoising and inpainting with deep neural networks. Adv Neural Inf Process Sys, 25
40. Yang C, Lu X, Lin Z, Shechtman E, Wang O, Li H (2017) High-resolution image inpainting using multi-scale neural patch synthesis. In&nbsp;Proceedings of the IEEE conference on computer vision and pattern recognition&nbsp;(pp. 6721–6729). https://arxiv.org/abs/1611.09969v2
41. Yang X, Xu P, Xue Y, Jin H (2021) Contextual feature constrained semantic face completion with paired discriminator. IEEE Access 9:42100–42110. https://doi.org/10.1109/ACCESS.2021.3065661
42. Yao F, Chu Y (2022) A Generative Image Inpainting Model Based on Edge and Feature Self-Arrangement Constraints. Computational Intelligence and Neuroscience, 2022
43. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2018) Generative image inpainting with contextual attention. In&nbsp;Proceedings of the IEEE conference on computer vision and pattern recognition&nbsp;(pp. 5505–5514). https://arxiv.org/abs/1801.07892v2
44. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2019) Free-form image inpainting with gated convolution. In&nbsp;Proceedings of the IEEE/CVF Int. J. Comput. Vis.&nbsp;(pp. 4471–4480). https://arxiv.org/abs/1806.03589v2
45. Zheng C, Cham TJ, Cai J (2019) Pluralistic image completion. In&nbsp;Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition&nbsp;(pp. 1438–1447). https://arxiv.org/abs/1903.04227v2
46. Zeng Y, Gong Y, Zeng X (2020) Controllable digital restoration of ancient paintings using convolutional neural network and nearest neighbor. Pattern Recogn Lett 133:158–164

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.