



Face Completion Using Generative Adversarial Network with Pretrained Face Landmark Generator

Reda Ghanem^{1*} Mohamed Loey²

¹*Mathematics Department, Faculty of Science, Benha University, Egypt*

²*Computer Science Department, Faculty of Computer and Artificial Intelligence, Benha University, Egypt*

* Corresponding author's Email: reda.ghanem@fsc.bu.edu.eg

Abstract: This paper, present a novel database of coloured and grey, plausible face images and proposes an improvement method for facial completion. The database contains 389 images of 79 Arab celebrities with automatically generated landmarks acquired from the web in wild-life which is a set of 68 landmark points was defined to provide information about the human face. Face detection applied using Caffe-Model with open cv to extract faces from images then store it in 256×256 pixels images. A good inpainting algorithm should produce a realistic face image. The current image face completion methods, recover the damaged areas of face images with low texture, there are problems such as low accuracy of face image recognition after inpainting. Therefore, this paper proposes an improvement method in facial inpainting using Generative Adversarial Network (GAN) with predicted landmarks to provide the structural information about damaged face to help the inpainter in generating plausible face image. Finally, evaluation for proposed model done over the available datasets CelebA, CelebA-HQ and our Novel Landmarked Face Database for Arab Celebrities. From the quantitative results, our proposed method achieves the maximum score of 34.97, 0.989 and 1.82 on PSNR (Peak Signal to Noise Ratio), SSIM (Structure Similarity Index Measure) and FID (Fréchet Inception Distance) metrics, respectively.

Keywords: Face inpainting, Generative adversarial network, Deep learning, Face database.

1. Introduction

For many years image completion, or the art of repairing old and deteriorated images has been rounded, it has recently gained even more popularity, because of the recent development in image processing techniques. With the improvement of image processing tools and the flexibility of digital image editing, automatic Image inpainting (a.k.a. image completion) has found important applications in computer vision and has also become an important and challenging topic of research in image processing.

Face inpainting is a sub-task of image inpainting designed to repair broken or occluded incomplete portraits. Due to the high complexity of face image details, inpainting on the face is more difficult. At present, face-related tasks often draw on excellent methods from face recognition and face detection, using multitasking to boost its effect. It's a

challenging problem in computer vision and has many practical applications. Typically, the face completion process semantically fills missing pixels in accordance with the available visual and predicted data.

At present, there are many databases available for research purposes. A detailed review of them was presented by Gross. Most of them consist of low-resolution images (as indicated in Table 1), which result from a more prohibitive computational cost of various recognition algorithms applied to high-resolution images at the time of creating the databases. However, most of the current database images not contain Arab faces, this causes a problem in applying some algorithms and making sure the model is efficient in working with most ethnicity of different faces and this is our first problem. This notion, motivate us to create a database containing plausible

Table 1. Face databases: This table shows some of the face databases available at the time of writing

Database	No. people	No. Images	Resolution
BioID Face [1]	23	1521	384 × 286
CelebA-HQ [2]	-	30K	1024 x 1024
CelebA [3]	≈ 10K	≈ 202K	178 x 218
Georgia Tech Face [4]	50	750	150 × 150
Faces94 [5]	153	153	180 × 200
The PUT Face [6]	100	10K	2048 × 1536
VALID [7]	106	530	360 × 288
Yale Face [8]	15	165	320 × 243

one size images of 79 Arab celebrities with automatically generated landmarks acquired from the web in wild-life in order to build an inpainting system with high accuracy in complete courted Arab face images.

Researchers have the access to many databases for face recognition applications; Table 1 is a non-exhaustive list. The databases vary in terms of purpose, size, and scope. Face recognition is the main purpose of these databases' photographs, collected by small teams of researchers. Acquisition of a face database over a short time and in a particular location has significant upper hand for certain types of research. Such an acquisition gives the experimenter direct control over the parameters of variability in the database.

Face inpainting has improved in the last few years. Much of this success can be attributed to the development of deep learning techniques such as Convolutional Neural Networks (CNNs) [9] and Generative Adversarial Network (GANs) [10]. The technology of face image inpainting is a crucial research discipline in image recovery, which is a good candidate for various applications such as Photoshop as an intelligent photo editing tool, Retouch Pilot as a tool for restoring old photos and improving the accuracy of face recognition systems (e.g. mobile face unlock).

It is challenging to inpaint face images in the wild, due to the large variation of appearance, such as expressions and ethnicity. Even the best commercial facial recognition systems are wrong up to 50 PERCENT of the time when trying to identify masked faces as proved in [11].

By using freshly developed algorithms, image completion can restore coherently both structure components and texture of the image. The obtained results demonstrate that these methods can remove

Table 2. Summarization of sequential-based methods evaluations

Image type	Distortion type	Method	PSNR	Images source
RGB	Random	Duan et al. [12]	27.40	Internet
	Objects	Muddala et al. [13]	28.52	Internet
	Scratch	Ružic' et al. [14]	23.89	Internet
Grayscale	Text	Geo et al. [15]	23.79	Lena
	Random	Lu et al. [16]	28.61	Barbara
	Objects	Fan et al. [17]	22.63	Lena

Table 3. Performance of CNN-based methods

Method	Distortion type	Dataset	PSNR
Yan et al. [20]	Random	Places, Paris StreetView	26.51
Sidorov and Hardeberg [21]	Vertical lines	Indian Pines	37.54
Pathak et al. [22]	Random	Paris StreetView	17.59
Guo et al. [23]	Masks (10,20%)	CelebA	33.66
	Masks (20,30%)	CelebA	29.01

unwanted objects from the images without leaving bad effects like noise and blur. Until now, few methods are proposed for blind image completion regarding the huge number of acceptable works with different techniques like sequential-based, CNN-based and GAN-based methods.

The evaluation metrics for image inpainting algorithms differ according to the using technique. In order to evaluate the efficiency of the proposed methods, researchers use some evaluation measures including Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) [18] and Fréchet inception distance (FID) [19]. Summarization of sequential-based methods evaluations use PSNR metrics shown in Table 2.

Recently, the strong potential of deep convolutional networks (CNNs) is being exhibited in all computer vision tasks, especially in image inpainting. CNNs are used specifically in order to improve the expected results in this field using large-scale training data. The sequential based methods succeed in some parts of image inpainting like filling texture details with promising results, yet the problem of capturing the global structure is still a challenging task [20].

Many of CNN-based methods for image inpainting and their performance evaluation and the

Table 4. GAN-based performance results

Method	Distortion type	Dataset	PSNR
Chen and Hu [24]	Blocks	CelebA	21.45
Victoria et al. [25]	Random	CelebA	23.06
Li et al. [26]	Blocks	CelebA	25.56
	Blocks	Places2	21.02
Yuan et al. [27]	Blocks	CelebA	22.09

datasets used, the type of distortion, evaluation metrics and the resolution of the images used in training represented in Table 3.

In 2014 [10], GAN-Based methods were introduced as a technique for image generation. Thanks to its high-quality image inpainting provided by a coarse-to-fine network and contextual attention, it is the most used technique in all digital visual applications nowadays. Table 4 represents some of GAN-Based methods for image inpainting and their performance evaluation, the distortion type, the used datasets, metrics used for evaluation and the images' resolution used in training.

Two issues arise when facial inpainting is performed using GAN, based on research that has been done before. The first issue is that the inpainting outcomes are not realistic if the input images are Arab face and non-aligned face images. The second issue is that the effect of the inpainting is inconsistent with the colours of its surrounding pixels. Such issues may be described as a problem of spatial correlation/consistency and the previous two issues considered as our second problem.

To overcome the first problem, we will create a Novel Landmarked Face Database for Arab Celebrities (NLFDAC for short) and to solve the second problem, we propose an enhancement technique for facial inpainting using GAN with pretrained face landmark MobileNet-V2 network generator, proposed method comprises of a facial landmark predicting subnet and an image inpainting subnet based on the (U-Net) structure that employs the predicted landmarks as guidance to preserve its deep-features, which relate to its perceptual quality. Overcome problems fully illustrated in (section 3)

Training process done on different face datasets, including CelebA-HQ, CelebA [2, 3] and (NLFDAC) face dataset. difference among the methods. Then, we compare our proposed method results using (PSNR), (SSIM) [18] and (FID) [19]. Qualitative analyses, demonstrate that our enhanced algorithm is capable of generating a good resolution satisfactory face completion.

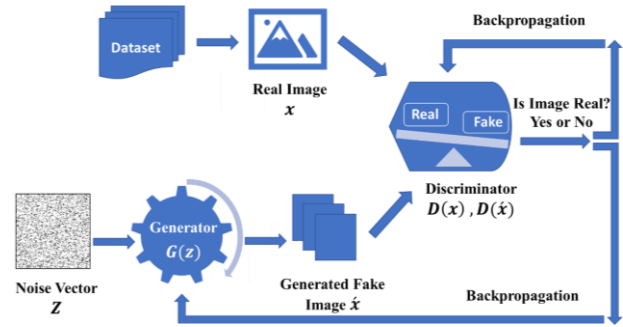


Figure. 1 Structure of GANs

This paper, organized as follows. In Section 2, we discuss the basic of generative adversarial network and the concept of face completion. In Section 3, presents the materials and method as first how we create the database in details, give additional data about landmarks and second our proposed face completion. The experiments are listed in Section 4 and some current methods are compared with our proposed method. Finally, In Section 5, we provide the conclusion and future work.

2. Literature review

2.1 Generative adversarial network (GAN)

(GANs) [10] were proposed by Goodfellow et al. in 2014. It's one of deep learning techniques which contain two networks, a generator G and a discriminator D . While the generator produces plausible fake images, the discriminator compares the fake images produced by the former with the real images (as shown in Fig. 1). Formally, the generator G maps samples of noise vector (z) in the latent space to an image, the generator is defined as $G(z) \rightarrow x'$, and the discriminator is defined as $D(x') \rightarrow [0,1]$, which classifies an image as a real image (close to 1) or as a fake image (close to 0).

This adversarial principle resembles the concept of a two-player mini-max game and the objective function of GAN can be stated as in the Eq. (1)

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

2.2 Concept of face completion

Face inpainting problem can be defined as art of repairing the old and deteriorated face image (C) with corrupted regions masked by (M), and the

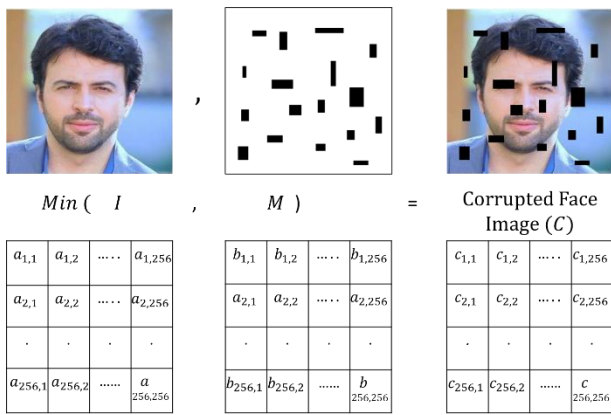


Figure. 2 How to create corrupted face image

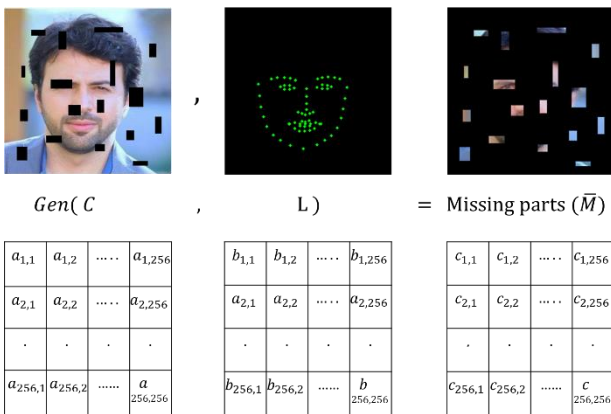


Figure. 3 Find missing parts of face image

original complete face image is denoted as (I). Let (\bar{M}) designate the complement of (M) which mean missing parts after recovering, (L) denote to predicted landmark, $Min(I, M)$ function to add mask on the original completed face image by taking minimum pixel values between both input images, and $Gen(C, L)$ function to generate and predict the missing parts of face image using advanced neural networks. The goal is to fill the target part with plausible and visually continuous information to the observed part. In other words, the completed face image (N) result will be as in the Eq. (2):

$$N := \bar{M} + C \quad (2)$$

where $\bar{M} := Gen(C, L), C := Min(I, M)$

So, the complete face image (N) should preserve the topological structure among face components such as eyes, nose and mouth, and the attribute consistency on like pose, gender, ethnicity and expression.

For more clarification example of inpainting face process introduced as following in Fig. 2 which show how to create corrupted face image by adding mask on the original face image, Fig. 3 show process of find the missing parts of face image using predicted

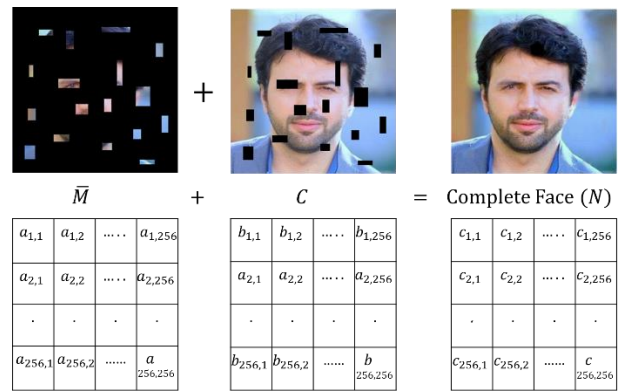


Figure. 4 Adding missing parts to the corrupted face image to get complete face (N)

landmark and finally Fig. 4 show getting a new completed face image by adding generated missing parts to the corrupted face image to get complete face image (N).

3. Materials and method

3.1 Create a novel arab face dataset

The specific details of how the database was constructed are given in this section, the process of building the dataset can be broken into the following steps:

a) Download images from web

As a starting point, images for Arab Celebrities in wild-life downloaded from web and remove duplicated images, we make sure that database will be variant and contain faces for male and female.

b) Convert images to .jpg format and labelling

Convert all images to .jpg format using (format factory) software then labelling each image with the real name of the detected face.

c) Face detection

Applying face detection using Caffe-Model [28] with open cv to extract faces from images then store it in 256×256 pixels images, Fig. 5 show process of face detects.

d) Face alignment

After face detection process some faces cropped in corner of image not centered, so we do face alignment manually for this image as in Fig. 6.

e) Face Landmark

The image containing face and eyes were automatically marked using neural network known as Face Alignment Network (FAN) [29] to generate landmarks as ground truth landmarks on each image in order to provide precise face alignment required by many face recognition methods. Coordinates and dimensions of those regions were stored in files in the txt format.



Figure. 5 Detect and crop face image



Figure. 6 Detect, crop and alignment face image

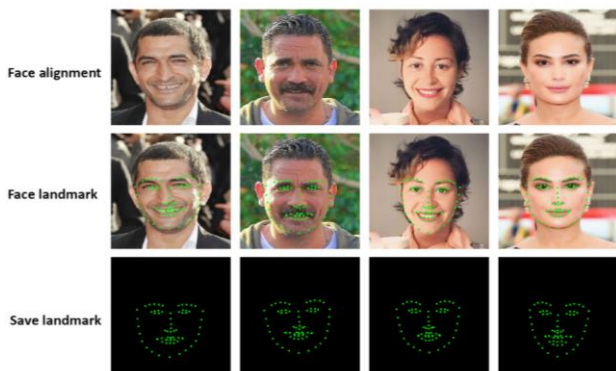


Figure. 7 Detect and crop face image

A set of 68 landmark points was defined to provide information on accurate position of facial features (e.g. the eyes corners, the mouth corners as in Fig. 7).

Each image in the database was supplied with a file containing automatically marked locations of all visible landmarks. Additionally, those points form polylines representing following contours:

- face outline (17 points),
- nose outline (9 points),
- eyes outlines (12 points),
- eyebrows outlines (10 points),
- mouth outlines (inner and outer - 20 points).

Some potential applications of the presented database have been presented throughout this paper. To summarize the following activities could benefit from using our Novel Landmarked Face Database for Arab Celebrities such as: evaluation face recognition to pose variations, evaluation of face and facial features localization algorithms, evaluation of the performance of face inpainting algorithms.

These of course are not the only possible applications of the presented database. The availability of such an extensive Arab face database of color, plausible resolution face images can contribute to the development of new algorithms related to the face images processing. Already existing techniques can also profit from applying to this database.

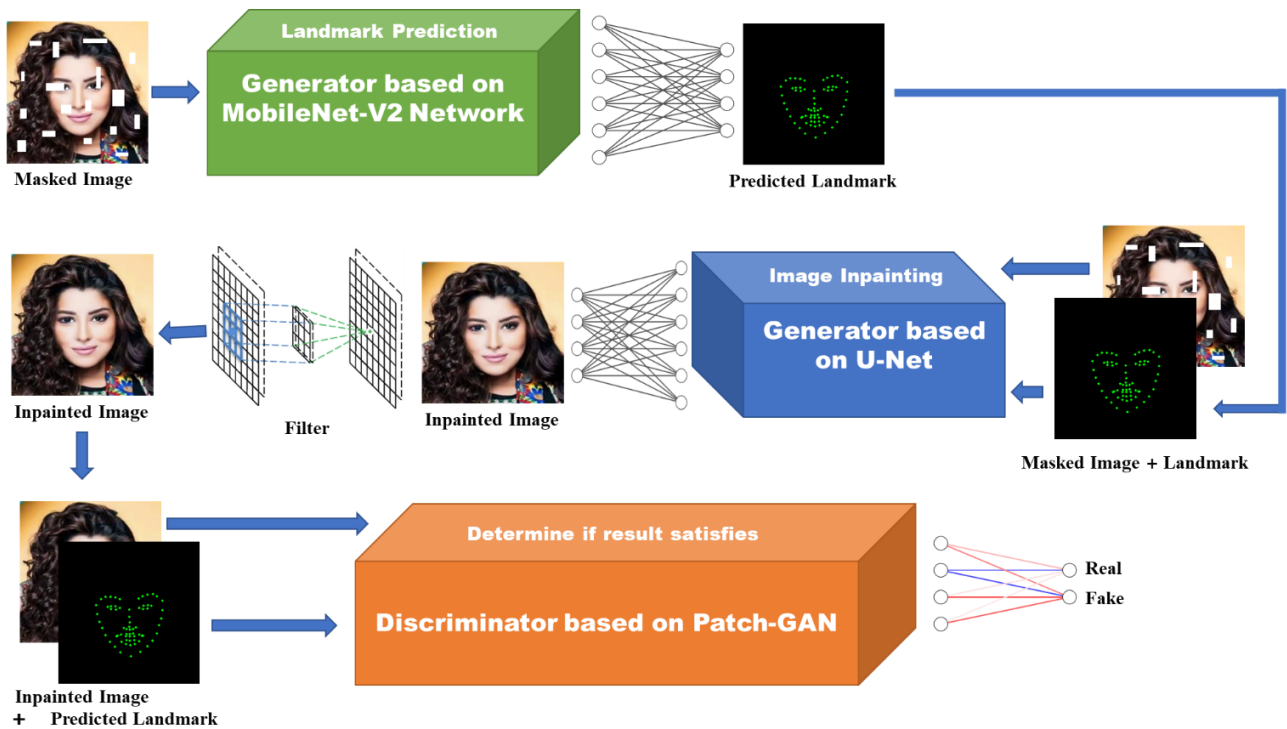


Figure. 8 Proposed architecture for our model

3.2 Proposed face completion

As stated previously, it's a challenging for completing face images in wild. A high-quality face inpainting algorithm must factor in the following two concerns to guarantee the realism of output:

- The strong structure of the faces, the fact that the facial features are topologically related and organized. The image output must primarily satisfy this topology harmonic structure
- Consistency is a must-be result of the inpainted regions with respect to gender and ethnicity, apart from the pose and facial expression determined by the landmarks.

Our inpainting system depending on generative adversarial network with pretrained face landmark generator, which comprises of a facial landmark predicting subnet and an image inpainting subnet, for solving the face inpainting problem. The main contributions for our enhanced deep network can be summarized as following:

- Facial landmarks act as the indicator for face inpainting. We used a module built upon the MobileNet-V2 model proposed in [30] for predicting landmarks on incomplete faces, which reflect the topological structure of our target face.
- To inpaint the faces by generate a new face, we used an inpainting subnet based on the (U-Net) structure that employs the predicted landmarks as guidance.
- After model inpaint the image it's goes through two image processing filter (Correcting improperly white-balanced images [31], Brightness preserving dynamic fuzzy histogram equalization [32]) to enhance the quality of image.

In Fig. 9 we will schematically illustrate the architecture of our proposed model as following: For a corrupted image, the landmarks are first estimated by the landmark prediction module. Then the inpaint module applies the landmarks as prior to inpaint the image then applies two image processing filters on inpainted image (Correcting improperly white-balanced images [31] and Brightness preserving dynamic fuzzy histogram equalization [32]) to enhance the quality of image. Finally, discriminator constrict the inpainted image as a real or fake in order to send feedback to enhance the generator processing. So fundamentally, the network is composed of two subnets first one for predicting landmarks, and second one for generating new pixels values conditioned on the predicted landmarks. In the subsequent subsections, we explore the network in detail.

3.3 Landmark prediction module

The landmark prediction module represented as (G_L) work to recover a set of ($n = 68$ in this work) landmarks (L) from a corrupted face image (C), with θ_L the trainable parameters as in the Eq. (3).

$$\begin{aligned} C &:= \text{Min}(I, M) \\ L &:= G_L(C; \theta_L) \end{aligned} \quad (3)$$

The G_L is built upon the MobileNet-V2 model which focuses on feature extraction this model proposed in [30]. The fully connecting fused feature maps predicts the final landmark at different rear stages as in Fig. 8.

Landmark predictor is based on the MobileNet-V2 [30]. A series of bottlenecks are employed to extract the features and speed up the network. Feature maps at different stages of fusion layers are fully connected to achieve the final landmark prediction. The detailed architecture is shown in Table 5, each line represents a sequence of identical layers, repeating (n) times. For layers in the same sequence, they have the same number (c) of output channels. Each sequence's first convolution layer has a stride (s). Expansion factor (t) is recruited to input size in bottleneck layers.

3.4 Image inpainting module

The inpainting network Gen_p tries to integrate missing regions of the faces by taking the deteriorated image (C) and their (predicted or ground-truth) landmarks L (L or L_{gt}) as input, with θ_p the network trainable parameters to get the completed face image (N) as in Eq. (4).

$$N := Gen_p(I^M, L; \theta_p) \quad (4)$$

Image inpainting module is a subnet comprises of a generator and a discriminator we will illustrate them in next sections.

3.4.1. Generator of the Inpaintor

U-Net is the structure upon which the generator of the inpaintor is based. Down-sampling is carried out by three encoding blocks, then seven residual blocks with dilated convolutions are applied to enlarge the capacity of receptive fields. The long-short term attention block links the following: a feature from last residual block and another from the last down-sampling block. This, in turn, widen the range for the features to be employed.

Table 5. The network architecture of the landmark predictor

Input	Operator	t	c	n	s
$256^2 \times 3$	conv2d	-	32	1	2
$128^2 \times 32$	bottleneck	1	16	1	1
$128^2 \times 16$	bottleneck	6	24	2	2
$64^2 \times 24$	bottleneck	6	32	3	2
$32^2 \times 32$	bottleneck	6	64	4	2
$16^2 \times 64$	bottleneck	6	96	3	1
$16^2 \times 96$	bottleneck	6	160	3	2
$8^2 \times 160$	bottleneck	6	320	1	1
(C1) $8^2 \times 320$	conv2d 1x1	-	1280	1	1
(C2) $8^2 \times 1280$	avgpool 8x8	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	64	-	-
(S3) $1 \times 1 \times 64$	-	-	64	-	-
C1	conv2d 1x1	-	128	1	1
$8^2 \times 128$	avgpool 8x8	-	128	1	-
(S1) $1 \times 1 \times 128$	-	-	128	1	-
C2	conv2d 1x1	-	128	1	1
$8^2 \times 128$	avgpool 8x8	-	128	1	-
(S2) $1 \times 1 \times 128$	-	-	128	1	-
S1, S2, S3	FC	-	136	1	-

The shortcuts are added among the corresponding encoders and decoders. To modify the weights of features from shortcut and last layer, we integrated the 1x1 convolutions as channel attention. Table 6 shows the architecture in detail. Each line represents a sequence of listed layers or a hole block. The kernel size, output channels, stride and padding of convolution or deconvolution layers are denoted as (k) , (c) , (s) and (p) respectively. Reflection paddings are applied in the first convolution layer and last deconvolution layer while others apply zero-padding.

3.4.2. Discriminator of the Inpaintor

The Patch-GAN is the architecture upon which the discriminator is based. We employed spectral normalization on the convolution layers to stabilize the training process. The attention block is placed in the discriminator to adaptively treat the features. The detailed architecture is given in Table 7, each line

Table 6. The generator of the inpaintor network architecture

Input	Operator	k	c	s	p	Out
$256^2 \times 4$	Conv-IN-ReLU	7	64	1	3	E1
$256^2 \times 64$	Conv-IN-ReLU	4	128	2	1	E2
$128^2 \times 128$	Conv-IN-ReLU	4	256	2	1	E3
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	-
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	-
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	-
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	-
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	-
$64^2 \times 256$	Dilated Residual Block	-	256	-	-	R7
E3, R7 $64^2 \times 512$	Short + Long Term Attention	-	256	-	-	-
$64^2 \times 256$	Deconv-IN-ReLU	4	128	2	1	D1
E2, D1 $128^2 \times 256$	Conv-IN-ReLU	1	256	1	0	-
$128^2 \times 256$	Deconv-IN-ReLU	4	64	2	1	D2
E1, D2 $256^2 \times 128$	Conv-IN-ReLU	1	128	1	0	-
$256^2 \times 128$	Conv-IN-tanh	7	3	1	3	-
$256^2 \times 3$	-	-	3	-	-	-

Table 7. The discriminator of the inpaintor network architecture

Input	Operator	k	c	s	p
$256^2 \times 4$	Conv-SN-LReLU	4	64	2	1
$128^2 \times 64$	Conv-SN-LReLU	4	128	2	1
$128^2 \times 128$	Attention	-	128	-	-
$64^2 \times 128$	Conv-SN-LReLU	4	256	2	1
$32^2 \times 256$	Conv-SN-LReLU	4	512	1	1
$31^2 \times 512$	Conv-SN-Sigmoid	4	1	1	1
$30^2 \times 1$	-	-	1	-	-

represents a sequence of listed layers or a hole block. The SN refers to spectral normalization and the LReLU means leaky relu with the slope set to (0:2).

4. Results

We evaluate the intelligent face inpainting system performance on CelebA-HQ, CelebA face dataset [2, 3] and our (NLFDAC). Random mask databases provide us with the masks used for training also, additional block masks are manually generated in a random way. The following methods are the ones that are incorporated into comparison: (1) Contextual Attention (CA) [33], (2) Pluralistic Image Completion (PIC) [34], (3) EdgeConnect (EC) [35], (4) Context Encoder (CE) [22] and (5) Generative Face Completion (GFC) [36].

For a good and fair measuring the performance difference among the methods, we employ PSNR, SSIM and FID, as metrics. Higher values indicate better performance for PSNR and SSIM, while the lower is the better for FID.

PSNR is most easily defined via the mean squared error (MSE). Given a noise-free ($m \times n$) image (I) and its noisy approximation (K), Then the MSE is defined as in Eq. (5).

$$MSE = \frac{1}{m * n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (5)$$

SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric; the SSIM index is calculated on various windows of an image. The measure between two windows x and y of common size $N \times N$ is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (6)$$

where \rightarrow

- μ_x, μ_y the average of x, y
- σ_x^2, σ_y^2 the variance of x, y
- σ_{xy} the covariance of x, y
- $c_1 = (k_1L)^2, c_2 = (k_2L)^2$ two variables to stabilize the division with weak denominator.
- L the dynamic range of the pixel-values (typically this is $2^{(\text{number of bits per pixel})} - 1$)
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

FID is a metric that calculates the distance between feature vectors calculated for real and generated images. The FID score is calculated using the following Eq. (7).

$$FID = \|\mu_R - \mu_G\|^2 + Tr(\sigma_R + \sigma_G - 2\sqrt{(\sigma_R * \sigma_G)}) \quad (7)$$

Where (μ_R, μ_G) refer to the feature-wise mean of the real and generated images. The (σ_R, σ_G) are the covariance matrix for the real and generated feature vectors, often referred to as sigma. The ($\|\mu_R - \mu_G\|^2$) refers to the sum squared difference between the two mean vectors. The (Tr) refers to the trace linear algebra operation which is the sum of the elements along the main diagonal of the square matrix.

Table 8. Quantitative comparison on the celebA-HQ dataset in terms of PSNR, SSIM and FID using random and center masks

	Mask	CA	PIC	EC	Ours
PSNR	10-20%	27.51	30.33	30.73	34.97
	20-30%	24.42	27.05	27.56	29.12
	Center	24.13	24.22	24.82	26.02
SSIM	10-20%	0.942	0.968	0.971	0.989
	20-30%	0.892	0.936	0.942	0.969
	Center	0.864	0.870	0.874	0.967
FID	10-20%	7.29	2.72	2.33	1.82
	20-30%	14.62	4.49	3.89	2.91
	Center	7.39	4.98	8.26	5.88

We employed FAN [29] (as reference information for training our landmark predictor) on the results since the ground-truth landmarks are not available for the dataset CelebA-HQ [2].

Result comparison in Table 8 which reports the performance of CA, EC, PIC and our approach with different types and sizes of mask. It can be concluded from the numbers in Table 8 that EC is almost always superior over PIC and CA since it recruits the edge information to help inpainting. But overall, our approach outperforms the others by large margins in terms of all PSNR, SSIM and FID, except for the case of center falling behind PIC in terms of FID (5.88) vs. (4:98) In the situation of center mask. This comparison proves that the strength of landmarks is more superior and rather more reliable than edges for the task of inpainting.

Table 9 indicates more quantitative comparisons with EC, CE and GFC under center random blocks masks on CelebA [3]. Finally, our new results using random blocks masks on our new (NLFDAC) face dataset are (37.81, 0.989 and 1.82) on (PSNR, SSIM and FID) metrics, respectively. Obviously, the result show how our enhanced model produces a good result on Arab faces dataset.

Table 9. Quantitative comparison on the CelebA dataset in PSNR, SSIM and FID using center random blocks masks.

Metric	CE	GFC	EC	Ours
PSNR	25.46	21.04	25.83	27.12
SSIM	0.909	0.766	0.899	0.947
FID	1.731	14.958	3.519	3.506

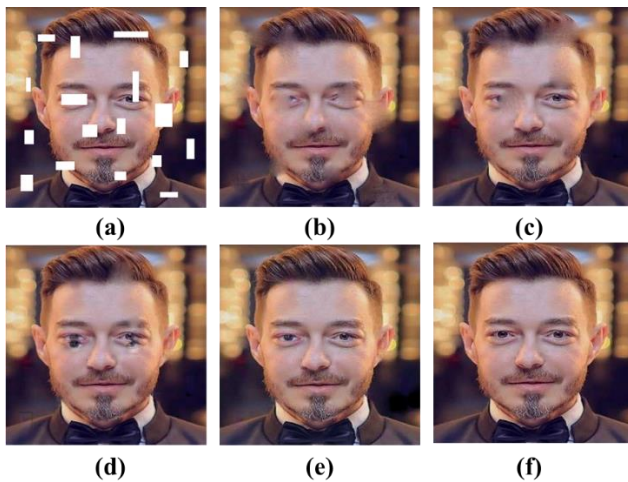


Figure. 9 Qualitative comparison with other state-of-the-art techniques on our NLFDAC face dataset: (a) input masked image, (b)-(e) are the results obtained by CA, EC, PIC and our Approach, respectively, and (f) ground-truth images

Visual comparisons depict in Fig. 9 among CA, EC, PIC and our Approach, from which, we can see that our enhanced face inpainting system can generate more natural-looking and visually striking results even on the cases with large poses and extreme occlusions.

Also, some results after testing our inpainting system on our (NLFDAC) dataset are shown in Fig. 10 starting from input masked face image to predicted face landmark then get two enhanced results for face image.

5. Conclusion and future work

We have introduced a novel database of colored and gray, plausible face images which contains 389 images of 79 Arab celebrities with automatically generated landmarks acquired from the web in wild-life. Many different adaptations, tests, and experiments have been left for the future to make the database contain more images for more individuals of Arab faces and increase resolutions of face image. The database is publicly available for the research purposes only to receive the database please contact one of the authors.

In this study, we have enhanced a system uses generative adversarial network with pretrained face landmark generator for completing face images. The network is composed of a subnet for predicting landmarks, and another one for generating new pixels conditioned on the predicted landmarks. First the suggested technique predicts the landmarks and then performs the inpainting based on the predicted landmarks. Our concept is that the landmarks are clean, adequate and robust to serve as guidance for

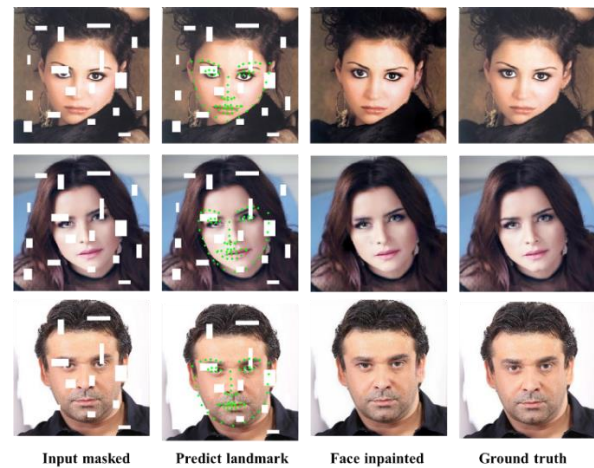


Figure. 10 Some results after testing our inpainting system on our NLFDAC dataset

the face in painting module to provide structural details. Extensive results were compared to validate our statements, expose the feasibility of our design and demonstrate its advancements both qualitatively and quantitatively over state-of-the-art alternatives. The enhanced system trained on two dataset CelebA and our (NLFDAC). We plan to improve the implementation by Improve the performance of inpainting system and train the model on more other datasets.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper methodology, conceptualization, validation, collecting-database, formal-analysis, resources, investigation, writing-original draft preparation, writing-review and editing, visualization, have been done by the 1st author. The supervision, and project administration, have been done by 2nd author.

References

- [1] O. Jesorsky, K. Kirchberg, and R. Frischholz, "Robust face detection using the hausdorff distance", In: *Proc. of International Conf. on Audio-and Video-Based Biometric Person Authentication*, pp. 90-95, 2001. doi: 10.1007/3-540-45344-x_14.
- [2] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation", *arXiv:1710.10196*, pp. 1710-10196, 2017.
- [3] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild", In: *Proc. of*

- the IEEE International Conf. on Computer Vision*, pp. 3730-3738, 2015. doi: 10.1109/iccv.2015.425.
- [4] A. Nefian and M. Hayes, "Maximum likelihood training of the embedded HMM for face detection and recognition", In: *Proc. of 2000 International Conf. on Image Processing (Cat. No. 00CH37101)*, Vol. 1, pp. 33-36, 2000. doi: 10.1109/icip.2000.900885.
- [5] L. Spacek, "Collection of facial images: Faces94", [Online]. Available: cmp.felk.cvut.cz/~spacelib/faces/faces94, 2007.
- [6] A. Kasinski, A. Florek, and A. Schmidt, "The PUT face database", *Image Processing and Communications*, Vol. 13, No. 3-4, pp. 59-64, 2008.
- [7] N. Fox, B. O'Mullane, and R. Reilly, "The realistic multi-modal VALID database and visual speaker identification comparison experiments", In: *Proc. of 5th International Conf. on Audio-and Video-Based Biometric Person Authentication*, 2005.
- [8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, 1997. doi: 10.1109/34.598228.
- [9] N. Cai, Z. Su, Z. Lin, H. Wang, Z. Yang, and B. W.-K. Ling, "Blind inpainting using the fully convolutional neural network", *The Visual Computer*, Vol. 33, No. 2, pp. 249-261, 2015. doi: 10.1007/s00371-015-1190-z.
- [10] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets", In: *Proc. of Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014. doi: arXiv:1406.2661.
- [11] M. Ngan, P. Grother, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms", *National Institute of Standards and Technology (U.S.)*, 2020. doi: 10.6028/nist.Ir.8311.
- [12] J. Duan, Z. Pan, and B. Zhang, "Fast algorithm for color texture image inpainting using the non-local CTV model", *Journal of Global Optimization*, Vol. 62, No. 4, pp. 853-876, 2015. doi: 10.1007/s10898-015-0290-7
- [13] S. Muddala, R. Olsson, and M. Sjöström, "Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting", *EURASIP Journal on Image Video Processing*, Vol. 2016, No. 1, p. 9, 2016. doi: 10.1186/s13640-016-0109-6
- [14] T. Ružić and A. Pižurica, "Context-aware patch-based image inpainting using Markov random field modeling", *IEEE transactions on image processing*, Vol. 24, No. 1, pp. 444-456, 2014. doi: 10.1109/tip.2014.2372479
- [15] J. Gao, J. Zhu, and K. Nie, "An image inpainting method for interleaved 3D stacked image sensor", *IEEE Sensors Journal*, Vol. 19, No. 24, pp. 12253-12260, 2019. doi: 10.1109/jsen.2019.2939357
- [16] H. Lu, Q. Liu, and M. Zhang, "Gradient-based low rank method and its application in image inpainting", *Multimedia Tools Applications*, Vol. 77, No. 5, pp. 5969-5993, 2018. doi: 10.1007/s11042-017-4509-0
- [17] Q. Fan and L. Zhang, "A novel patch matching algorithm for exemplar-based image inpainting", *Multimedia Tools Applications*, Vol. 77, No. 9, pp. 10807-10821, 2018. doi: 10.1007/s11042-017-5077-z
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600-612, 2004. doi: 10.1109/tip.2003.819861
- [19] A. Obukhov and M. Krasnyanskiy, "Quality Assessment Method for GAN Based on Modified Metrics Inception Score and Fréchet Inception Distance", In: *Proc. of the Computational Methods in Systems and Software*, pp. 102-114, 2020.
- [20] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: Image inpainting via deep feature rearrangement", In: *Proc. of the European Conf. on Computer Vision (ECCV)*, pp. 1-17, 2018. doi: 10.1007/978-3-030-01264-9_1.
- [21] O. Sidorov and J. Hardeberg, "Deep hyperspectral prior: denoising, inpainting, super-resolution", *CoRR*, arXiv:1902.00301, 2019.
- [22] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting", In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2536-2544, 2016. doi: 10.1109/cvpr.2016.278
- [23] Z. Guo, Z. Chen, T. Yu, J. Chen, and S. Liu, "Progressive image inpainting with full-resolution residual network", In: *Proc. of 27th ACM International Conf. on Multimedia*, pp. 2496-2504, 2019. doi: 10.1145/3343031.3351022
- [24] Y. Chen and H. Hu, "An improved method for semantic image inpainting with gans:

- Progressive inpainting”, *Neural Processing Letters*, Vol. 49, No. 3, pp. 1355-1367, 2019. doi: 10.1007/s11063-018-9877-6
- [25] P. Vitoria, J. Sintes, and C. Ballester, “Semantic image inpainting through improved wasserstein generative adversarial networks”, In: *Proc. of 14th International Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 7, pp. 249-260, 2019. doi: 10.5220/0007367902490260
- [26] A. Li, J. Qi, R. Zhang, X. Ma, and K. Ramamohanarao, “Generative image inpainting with submanifold alignment”, In: *Proc. of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019. doi: 10.24963/ijcai.2019/114
- [27] L. Yuan, C. Ruan, H. Hu, and D. Chen, “Image inpainting based on patch-GANs”, *IEEE Access*, Vol. 7, pp. 46411-46421, 2019. doi: 10.1109/access.2019.2909553.
- [28] Vaishali and S. Singh, “Real-Time Object Detection System using Caffe Model”, *International Research Journal of Engineering and Technology (IRJET)*, Vol. 6, No. 5, pp. 5727-5732, 2019.
- [29] M. Kowalski, J. Naruniec, and T. Trzcinski, “Deep alignment network: A convolutional neural network for robust face alignment”, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 88-97, 2017. doi: 10.1109/cvprw.2017.254
- [30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks”, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018. doi: 10.1109/cvpr.2018.00474
- [31] M. Afifi, B. Price, S. Cohen, and M. Brown, “When color constancy goes wrong: Correcting improperly white-balanced images”, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1535-1544, 2019. doi: 10.1109/cvpr.2019.00163
- [32] D. Sheet, H. Garud, A. Suveer, M. Mahadevappa, and J. Chatterjee, “Brightness preserving dynamic fuzzy histogram equalization”, *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 4, pp. 2475-2480, 2010. doi: 10.1109/tce.2010.5681130
- [33] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, “Generative image inpainting with contextual attention”, In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5505-5514, 2018. doi: 10.1109/cvpr.2018.00577
- [34] C. Zheng, T.-J. Cham, and J. Cai, “Pluralistic image completion”, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1438-1447, 2019. doi: 10.1109/cvpr.2019.00153
- [35] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, “Edgeconnect: Generative image inpainting with adversarial edge learning”, *arXiv:1901.00212*, 2019.
- [36] Y. Li, S. Liu, J. Yang, and M.-H. Yang, “Generative face completion”, In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3911-3919, 2017. doi: 10.1109/cvpr.2017.624