E2F-GAN: Eyes-to-Face Inpainting via Edge-Aware Coarse-to-Fine GANs

AHMAD HASSANPOUR[1], **AMIR ETEFAGHI DARYANI**[2], **MAHDIEH MIRMAHDI**[3], **KIRAN RAJA**[1], (Senior Member, IEEE), **BIAN YANG**[1], **CHRISTOPH BUSCH**[1], (Senior Member, IEEE), **AND JULIAN FIERREZ**[4], (Member, IEEE)

[1]Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway
[2]Department of Electrical Engineering, Amirkabir University of Technology, Tehran 15875-4413, Iran
[3]Department of Computer Engineering, University of Isfahan, Isfahan 8174673461, Iran
[4]School of Engineering, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Corresponding author: Ahmad Hassanpour (ahmad.hassanpour@ntnu.no)

ABSTRACT Face inpainting is a challenging task aiming to fill the damaged or masked regions in face images with plausibly synthesized contents. Based on the given information, the reconstructed regions should look realistic and more importantly preserve the demographic and biometric properties of the individual. The aim of this paper is to reconstruct the face based on the periocular region (eyes-to-face). To do this, we proposed a novel GAN-based deep learning model called Eyes-to-Face GAN (E2F-GAN) which includes two main modules: a coarse module and a refinement module. The coarse module along with an edge predictor module attempts to extract all required features from a periocular region and to generate a coarse output which will be refined by a refinement module. Additionally, a dataset of eyes-to-face synthesis has been generated based on the public face dataset called CelebA-HQ for training and testing. Thus, we perform both qualitative and quantitative evaluations on the generated dataset. Experimental results demonstrate that our method outperforms previous learning-based face inpainting methods and generates realistic and semantically plausible images. We also provide the implementation of the proposed approach to support reproducible research via (https://github.com/amiretefaghi/E2F-GAN).

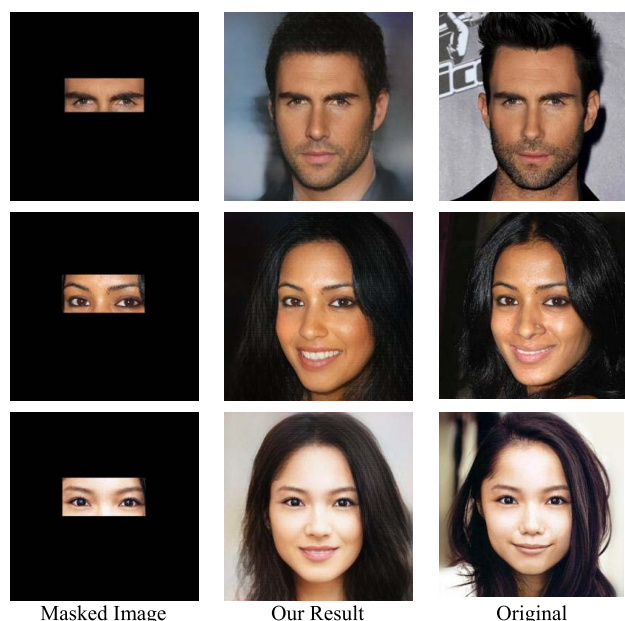**INDEX TERMS** Face inpainting, generative adversarial networks, image inpainting.

I. INTRODUCTION

Image inpainting is used to complete missing information or substituting undesired regions of pictures with conceivable and fine-grained content. It encompasses a wide extend of applications in fields of restoring harmed photos, editing pictures, removing objects, etc. [1], [2]. Many conventional methods typically use low-level and hand-crafted features from the corrupted input image and utilize the priors or additional data. By propagating the extracted features from visible and well-structured parts to the missing regions or by filling missed small areas by looking and melding comparative patches from the same or other images. In spite of the fact that these strategies have great effects in the completion of replicating structures, they are restricted by the accessible

The associate editor coordinating the review of this manuscript and approving it for publication was Khoa Luu.

regions in an image and cannot create novel image substance. In recent years, learning-based strategies have been proposed to overcome these confinements by utilizing huge volumes of training data [3], [4]. Notably, despite of great achievements of learning-based methods in this task, they are limited by at least three challenges: the inpainted area should be C1) semantically filled based on overall scene, C2) continuously structured with unmasked regions, and C3) visually realistic.

Recently, deep convolutional neural networks (CNNs) and generative adversarial networks (GANs), known as learning-based methods, have been widely used for various image inpainting tasks including removing objects, noises, texts, and masks. Based on convolutional neural networks (CNNs) and using encoder-decoder network structure several works have been proposed for image inpainting [5]–[8]. For instance, Sidorov and Hardeberg [6] proposed an encoder-decoder network for denoising, inpainting

FIGURE 1. Example completion results of our proposed method in comparison with original images.

and super-resolution for noised, inpainted and low-resolution images. Zhu *et al.* [5] proposed a patch-based inpainting method for various deep learning (DL) modules that have been proposed recently.

Coarse-to-fine based methods exploit one [9]–[11] or two [12]–[14]-stage architecture to complete content formation and texture refinement. A one-stage architecture (also termed coarse-and-fine architecture) consists of two parallel branches, coarse and fine, that extract two kinds of information simultaneously, coarse and fine information. The missed region, then, can be constructed from the extracted information. Alternatively, a two-stage architecture generates an intermediate coarse image after recovering structures in the first stage, and then feeds it to the second stage for improving the texture. Additionally, another category called structural guidance-based methods uses an assistance algorithm to provide more information for the main inpainting method. An edge and a contour generator have been used within a two-stage architecture in [15] and [16] respectively.

Although, it is worthy to mention that in face inpainting, besides the above-mentioned challenges (i.e., C1-C3), we are facing further requirements. Notably, a facial representation can be considered for the purpose of biometric recognition due to the special topology of different facial elements (i.e., forehead, eyes, eyebrows, nose, mouth, jaw, chin, cheek) and their distinctive characteristics [42]. Thus, revealing the hidden parts of a face by using other elements such that the topological face elements along with consistency in face attributes (e.g., demographic and other biometric information [43]) are preserved is a challenging task, yet it will have a strong impact on the feasibility of biometric recognition conducted by human experts (i.e. in forensic investigation [44]) or by machine learning [45] or hand-crafted algorithms [46]. Therefore, the requirements of face inpainting are as follows:

R1) the face topological structure should be reconstructed so that all elements are placed in the right position semantically and continuously. For this, first, the shape of the face (oval shape, square shape, round shape, etc.) should be predicted. Then all other elements should be placed proportionally within the predicted frame. Additionally, to look more realistic, the head pose should be naturally aligned and integrated with other elements. These requirements are the main challenges (i.e., C1-C3) of every inpainting method modified for face inpainting solutions. Since the aim of this paper is a special case of face inpainting where a large region of the face except eyes is hidden, besides R1, two other requirements which make the inpainting task more challenging should be considered. R2) Researchers have found that the area of skin around the eyes is useful to determine soft biometric information such as age or gender [17], [37]. The proposed inpainting model should utilize the color, texture, and size of eyes and eyebrows to estimate this kind of demographic attributes and inpaint other face elements according to the estimated features. R3) The proposed solution should preserve the identity-related biometric properties present in the eyes regions [18], [38] when generating the full face [39]. Noteworthy, this eye region is demonstrated to encode a large part of the identity information present in the face [44] enabling both person recognition and fake face detection [40].

Additionally, it is worth to mention that the hidden portion of the image can directly affect the performance of proposed solutions, and clearly large masks make meeting the referred requirements (i.e., R1-R3) more difficult. Considering this issue, the aim of this paper is to complete the face based on the eyes region (periocular region), our used mask type will cover most parts of the face.

In this paper, a novel DL-based architecture has been proposed such that it complies with the referred requirements (i.e., R1-R3, see Fig. 1). Therefore, our contributions and novelties can be summarized as follows:

- In this work, an effective end-to-end solution for reconstructing the face based on just the eyes region has been proposed. This innovative GAN-based architecture called E2F-GAN benefits from the advantages of coarse-to-fine, coarse-and-fine, and structural guidance-based architectures. The code for our proposed method is available in GitHub.[1]
- By using various loss functions during the training process [41], not only the quality of inpainted regions but also demographic and biometric features have been preserved and measured by several quantitative and qualitative evaluation metrics.
- A new dataset of masked faces called E2Fdb has been generated and made publicly available (same GitHub indicated before).
- In terms of selecting the most informative guidance-based method, we experimentally show that edges

---

[1]https://github.com/amiretefaghi/E2F-GAN

provide more structural and contextual information compared to landmarks.

II. RELATED WORKS

In eyes-to-face inpainting, a face (a raw image indicated by $I^{H \times W \times N}$ hereafter) is corrupted by a binary image mask ($M^{H \times W \times N}$), where H, W, and N show the height, width, and number of channels of the image respectively, and the corrupted image will be shown by I_m ($I_m = I \odot M$, where $\odot$ is the element-wise production). The inpainting model H takes I_m and M as input, and its output, reconstructed face, should fulfill the **R1**, **R2**, and **R3** ($I \cong \hat{I}$). The proposed inpainting methods use different architectures and various types of masks. In this section, we review recent face inpainting methods based on DL architectures and widely used mask types.

A. FACE INPAINTING METHODS

Apart from traditional methods which utilize low-level features extracted from the same image or a group of images, the learning-based strategy is the main focus of recent proposed methods due to using high-level features that enable them to inpaint the damaged regions semantically. In the following, we review several learning-based existing works that attempted to inpaint corrupted faces, similar to the aim of this paper.

The coarse-to-fine structure has been used in recent face inpainting tasks. Li *et al.* [19] proposed a generative-based coarse-to-fine structure that benefits from an attention layer to capture long dependency between features to generate more realistic images. Yu *et al.* [13] uses a coarse-to-fine structure to inpaint free-form masks. In the same context, Liu *et al.* [12] proposed a coarse-to-fine architecture with a novel attention layer. Chen *et al.* [19] proposed a coarse-and-fine structure including a coarse network for extracting global semantic information and a fine network to extract multi-level local features. Besides the coarse-to-fine based strategies, another category so-called structural guidance uses additional information to assist the main inpainting module. Nazari *et al.* [15] leverage an edge generator first to recover the edges, and the corrupted image is fed to the image inpainting network along with predicted edges. Chen and Liu [16] use a dual branch network including texture and edge branches to extract features and recover structures and textures of missed regions. Some works estimate facial landmarks to assist the main inpainting network [20], [21]. In this paper, we will take the advantages of different architectures, i.e., coarse-to-fine, coarse-and-fine, and structural guidance.

The above-mentioned methods produce a unique result per each input. On the other hand, some approaches inpaint the corrupted regions differently per each execution for each specific input. Zheng *et al.* proposed a Variational Auto-Encoders (VAEs)-based [22] dual pipeline including a reconstructive path that uses the ground truth to learn the prior distribution of missing regions and a generative path for which the conditional prior is connected to the distribution

obtained in the reconstructive path. An unsupervised conditional framework based on generative adversarial networks for varied image inpainting that can learn conditional completion distribution has been proposed by Zhao *et al.* [23]. A similar approach using GANs to restore low quality face images was recently proposed in [47]. It should be noted that, in E2F-GAN, we need a unique output for each input even after several executions to fulfill the requirements R2 and R3.

B. MASK COVERAGE

The used masks in face inpainting scenarios can be classified into two categories called free and fixed-form masks. In widely used free-form masks [8], [10], [15], [21], [22], [24], [26], there are irregular shapes randomly placed on the images (Fig. 2a). Instead, in the fixed-form masks [13], [21], [24], [25], regular shapes cover part of the images which are located on the images randomly or purposefully (Fig. 2b) [24], [25]. Since the aim of this paper is to complete the face based on eyes, our used mask type is in the latter category with a large-size mask ($\approx 75\%$ of the face).

III. PROPOSED METHOD

The overall network architecture of our proposed method, which is based on a coarse-to-fine architecture and includes two main modules called coarse and refinement, is shown in Fig. 3. Different from others [2], [3], [7], [13], [19], both modules (i.e., coarse and refinement) are GAN-based networks, therefore, each of which includes a generator and a discriminator. The coarse module, which comprises a generator called coarse generator (C), has a dual encoder that follows the coarse-and-fine structure to capture global semantic features and extract multi-level features from the eyes region. Besides this module, a GAN-based refinement module which consists of a refinement generator (F) and a discriminator (D_2) has been utilized to improve the coarse outputs. Intuitively, the refinement network sees a more completed scene than the masked images, so its network can learn better feature representations than the coarse network. Therefore, our end-to-end method includes two GAN-based modules which are training to generate the final result. In the following subsections, each module is described in detail.

Notably, facial landmarks [21] or edges [15] are usually the most widely-used structural guidance in image inpainting tasks. In our proposed E2F-GAN, where the used mask covers most parts of the face, predicting both landmarks and edges is a challenging problem. As a consequence, our proposed method utilizes both landmarks and edges during our experiments, in an effort to use the most effective structure (e.g., landmarks or edges). For facial landmarks, we used the landmark prediction method proposed in [27] and for predicting edges, we used the edge predictor proposed by Nazari *et al.* [15]. Both methods have been trained again on our generated dataset that contains specific eye masks. As we will see in the experiments, our quantitative and qualitative metrics will show that the edge structural guidance provides more effective information for our coarse generator.
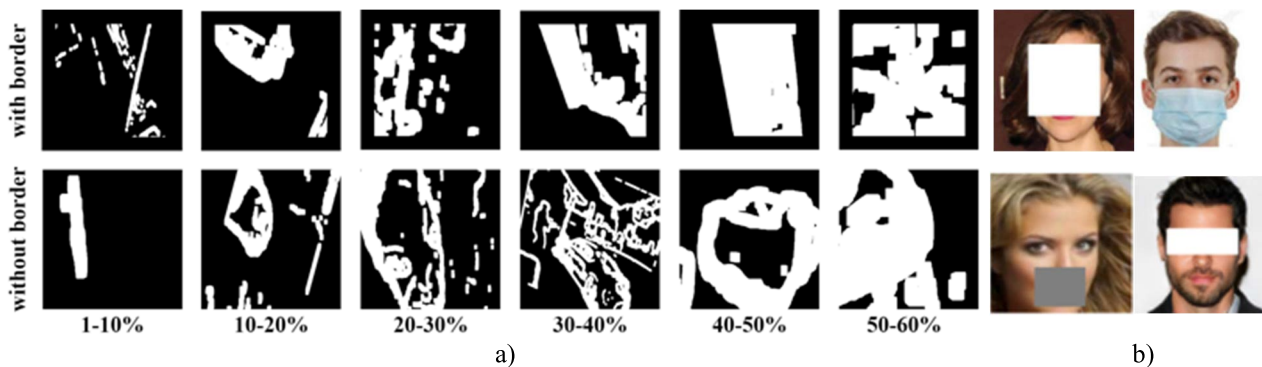
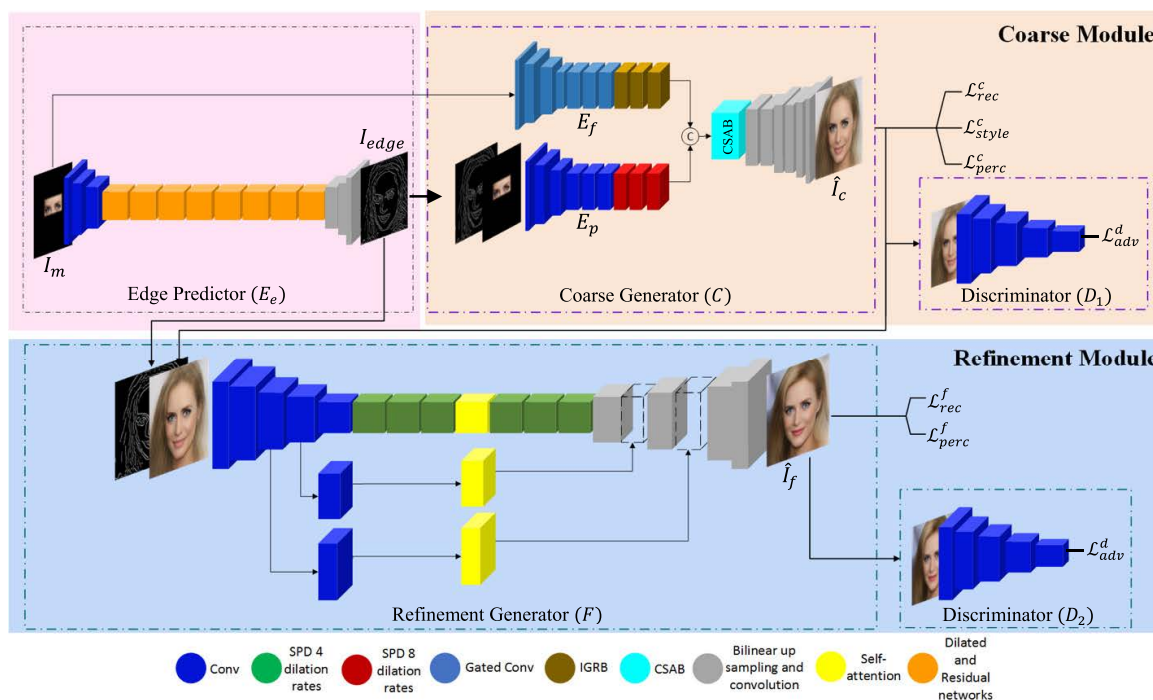**FIGURE 2.** Examples of two types of widely used masks called free-form [26] (a) and fixed-form [24], [25] (b).

FIGURE 3. Overview of our architecture with three main modules including Edge Predictor (E_e), Coarse Module (C), and Refinement Generator (F).

Therefore, in our final setup we use edges generated by an edge predictor (E_e) as structural guidance for C.

A. COARSE MODULE

The proposed GAN-based coarse module is responsible for extracting the required features from the masked image and generating the first coarse result. To do this, we designed the module with three submodules including edge predictor (E_e), coarse generator (C), and discriminator (D_1). In the following, we explain the role of each network, its architecture, and the used loss functions.

1) COARSE GENERATOR

The coarse generator has the main responsibility for meeting the three requirements (i.e., **R1**-**R3**). Not only the biometric and demographic feature should be extracted from the periocular region, but also the initial coarse prediction should look realistic, and semantically and continuously structured. This is achieved using three networks: two encoders so-called fine encoder (E_f) and pose encoder (E_p), and a decoder. The encoder E_f deals with the finest features of I_m and E_p deals with the predicted structure of faces obtained from E_e. Therefore, first I_m is fed to E_e to predict edges of visible and hidden regions (I_{edge}) and then I_{edge} is concatenated with I_m to fed E_p. This assists to predict the pose of different elements of the face. Additionally, I_m will be fed to E_f with the aim of extracting identity attributes. Finally, the decoder will predict and inpaint the hidden regions based on the two feature maps received from E_f and E_p. In the following, we describe each of these networks and their roles in our scheme.

a: FINE ENCODER

The aim of using this encoder is mainly to extract demographic (e.g., age, gender) and biometric properties (e.g.,

identity, skin color) from I_m. Therefore, the skin color around the eyes, wrinkles, the size of eyes and eyebrows, the distance between two eyes, and other possible properties should be considered. On the other hand, it should be noted that due to the high coverage ratio of I_m, E_f is fed with a lot of unusable information (the black region). To prevent deteriorating the quality of output and filter out these pixels, the first seven blocks of E_f are configured as with gated convolutions (GC) [14]. These blocks contain parallel convolution layers with different sorts of activation functions which assist to extract an appropriate feature map and eliminate extracted features from the masked region. Then, three interleaved gated residual blocks (IGRB) [19] have been placed after GC blocks to extract multi-level features.

b: POSE ENCODER

For extracting coarse structure and global semantics features, and consequently preserving the quality as well as the structure of the predicted face, an encoder called pose encoder (E_p) has been placed in the Coarse Module (C). It has been fed by concatenation of I_{edge} and I_m. Doing this, a receptive field for recognizing face structures will be available for E_p. However, the inputs I_{edge} and I_m are both sparse. To extract a meaningful feature map, similar to [29], we used three spatial pyramid dilation blocks (SPD) after six convolution layers. Notably, SPD blocks contain parallel convolution layers with various dilation rates to extract a large receptive field from the given input image.

c: DECODER

To inpaint the coarse output based on features extracted by E_p and E_f, a decoder including seven layers (one attention layer and six upsampling convolution layers) has been used. In common encoder-decoder approaches, the decoder receives features directly from the encoder but in our proposed method, the decoder receives two types of features including low-level features extracted by large receptive fields that may lack detailed information (i.e., the output of E_p), and high-level detailed features with a small receptive field (i.e., the output of E_f). Thus, we use a CSAB as the first layer of the decoder to discriminate the more effective features from others by assigning more weights.

Channel and Spatial Attention Block (CSAB): According to the outputs of E_p and E_f, the input to the attention block contains two types of features: a) large receptive field that may lack detailed information and b) output of E_f, i.e., high-level detailed features with small receptive fields. We adopt the concatenating operation to aggregate these two types of features. On the other hand, we may achieve redundant information about multi-level contextual information and this situation will not be efficient for our goals. Thereby, as shown in Fig.3, we adopt a specific attention block called channel and spatial attention block (CSAB) [19] to assign more weight to important features [48] and alleviate the interference of redundant features by channel and spatial attention. Hence, attention block composes of two main attentions which we

will introduce. Convolution operation leads to local contextual information. Discriminative features representation is essential for inpainting. We leverage the attention mechanism to fulfill this desire. The channel attention emphasizes interdependent feature maps by exploiting the dependencies between channels. Meanwhile, the spatial attention encodes a wide range of contextual dependency within each channel, thereby improving the overall representation capability by gaining mutual for similar features.

B. REFINEMENT MODULE

The coarse module's output ($\hat{I}_c$) consists of face coarse structure including placed face elements, stated face pose, specified color skin, etc., suffering from fine details. To add more details to the $\hat{I}_c$, we propose a GAN-based refinement module.

1) REFINEMENT GENERATOR

Inspired by the U-Net architecture [28] and the refinement network proposed by [29], we proposed a more effective architecture by replacing some DL blocks with SPD and self-attention (SA) blocks which receive the concatenation of $\hat{I}_c$ and I_{edge} as its input. We have adopted SPD blocks with four dilation rates in the middle of our architecture to extract features with various receptive fields from input images and then used SA blocks between middle layers. SA benefits from the concept of self-similarity, which is useful for reclaiming the reconstructed pattern based on the remaining ground truth in a masked image. As mentioned before, the duty of this stage is that it should improve fine details of images, hence, we use reconstruction and perceptual losses to adjust the fine details.

C. DISCRIMINATOR

To inpaint and generate more realistic high-quality faces, both coarse and refinement modules have been designed based on GAN structures, thus, two discriminators have the responsibility of evaluating the output of C and F. The coarse module's discriminator (D_1) receives $\hat{I}_c$ and consequently the refinement module's discriminator (D_2) has been fed by $\hat{I}_f$. We have combined the concept of SN-GAN [30] and Patch-GAN [31] for these discriminators to distinguish real or fake images. Besides this combination, we have used the hinge adversarial loss function for our discriminators. These combinations and loss functions help us to train our discriminators faster and more stable, distinguishing real or fake images efficiently.

D. E2F-GAN END-TO-END TRAINING

The E2F-GAN model is trained in a supervised and end-to-end manner. We have defined four groups of loss functions [41] for various parts of our proposed method to achieve considerable results. To train C, we have utilized four specific loss functions including reconstruction loss, perceptual loss, style loss, and adversarial loss; and just reconstruction and perceptual losses have been used for training F. With the aim

TABLE 1. Quantitative results over EtoF dataset for EtoFGAN and other compared methods (PIC, LaFIn, EC). The best result of each column is boldfaced. ↑ indicates that the higher the number the better is the model and ↓ indicates the lower the number the better is the model.

Method	FID ↓	SSIM ↑	PSNR ↑	TV ↓	ℓ_1Loss ↓
PIC	57.02	0.41	11.19	8.50	50.37
LaFIn	63.16	0.47	13.18	6.89	**40.94**
EC	70.63	0.42	12.67	5.27	121.08
E2F-GAN (ours)	**46.39**	**0.51**	**13.66**	**0.02**	41.54

of having an end-to-end training process, we define the total loss $\mathcal{L}$ which consists of four groups of component losses as below:

$$\mathcal{L} = \lambda_{rec}\left(\mathcal{L}_{rec}^c + \mathcal{L}_{rec}^f\right) + \lambda_{perc}\left(\mathcal{L}_{perc}^c + \mathcal{L}_{prec}^f\right)$$
$$+ \lambda_{style}\mathcal{L}_{style}^c + \left(\lambda_{adv}^g \mathcal{L}_{adv}^g + \lambda_{adv}^d \mathcal{L}_{adv}^d\right). \quad (1)$$

In the following, the formulation of the used losses and the notion behind each loss is described. The reconstruction loss ($\mathcal{L}_{rec}$) or per-pixel loss measures the pixel-wise difference between the synthesized image and the ground truth image. This loss is essential for maintaining texture information. It is calculated as the L1-norm between $\hat{I}_z$ and the corresponding ground truth I_g. $\mathcal{L}_{rec}$ is defined as follows:

$$\mathcal{L}_{rec}^z = \frac{1}{H \times W \times N} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| \hat{I}_z(i,j) - I_g(i,j) \right| \quad (2)$$

where z is replaced with c or f depending on the $\mathcal{L}_{rec}$ is used for C or F, respectively.

It is worth to mention that, an element-wise loss cannot consider high-level semantics. Accordingly, recent research [19], [21], [22] suggests using perceptual distances based on a pre-trained network, VGG19 which was trained on the ImageNet. The perceptual loss ($\mathcal{L}_{perc}$) measures the difference between features extracted from the various layers of the VGG19 network for $\hat{I}$ and its corresponding ground truth.

$$\mathcal{L}_{perc}^z = \sum_{l=1}^{L} \frac{\left| \hat{\varphi}_z - \varphi_g \right|}{N_l \times H_l \times W_l} \quad (3)$$

where $\hat{\varphi}_z$ and φ_g are extracted features from $\hat{I}$ and I_g respectively, and z is replaced with c or f depending on the $\mathcal{L}_{perc}$ is used for coarse or refinement, respectively. We extract features from L layers of the pre-trained network. $relu1\_1$, $relu2\_1$, $relu3\_1$, $relu4\_1$, and $relu5\_1$ of the VGG19 utilized to calculate $\mathcal{L}_{perc}$ as well as $\mathcal{L}_{style}$ described below.

In order to provide richer texture, we also employ style loss ($\mathcal{L}_{style}$). In style loss, a Gram matrix calculates the correlation between channels in a feature map. The style loss then calculates on the features map produced by the pre-trained VGG19 network.

$$\mathcal{L}_{style}^c = \sum_{l=1}^{L} \frac{1}{N_l \times N_l} \left\| \frac{G_l\left(\hat{I}_c\right) - G_l\left(I_g\right)}{H_l \times W_l \times N_l} \right\|_1 \quad (4)$$
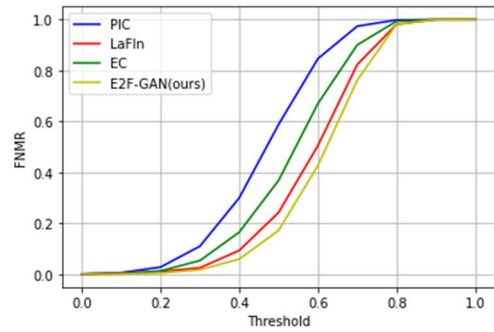
FIGURE 4. FNMR curve for our proposed method (E2F-GAN) and other compared methods (PIC, LaFIn, EC).

where $G_l(.) = \varphi_l(.)^T \varphi_l(.)$ stands for the Gram matrix corresponding to $\varphi(.)$.

For generative adversarial learning, our discriminators are trained to distinguish between generated images and ground truth images. on the other hand, the generators strive to cheat the discriminators by hardening that classification. We employ hinge loss to train our model, L_{adv}^d and L_{adv}^d computed as follows:

$$\mathcal{L}_{adv}^g = -E\left[D_1\left(C\left(I_m, I_{edge}\right)\right)\right] - E\left[D_2\left(F\left(\hat{I}_c, I_{edge}\right)\right)\right] \quad (5)$$

$$\mathcal{L}_{adv}^d = E\left[Relu\left(1 + D_1\left(C\left(I_m, I_{edge}\right)\right)\right)\right]$$
$$+ E\left[Relu\left(1 + D_2\left(F\left(\hat{I}_c, I_{edge}\right)\right)\right)\right]$$
$$+ E\left[Relu\left(1 - D_1\left(I_g\right)\right)\right] + E\left[Relu\left(1 - D_2\left(I_g\right)\right)\right] \quad (6)$$

As mentioned before, we combine the used loss functions with appropriate weights as follows: $\lambda_{rec} = 1, \lambda_{perc} = 0.1, \lambda_{style} = 250, \lambda_{adv}^g = 0.1, \lambda_{adv}^d = 1$.

IV. EXPERIMENTS AND DISCUSSION
In this section, we evaluate the E2F-GAN performance on a new generated face dataset (E2Fdb) based on CelebA-HQ. We compared our results with three other methods called EdgeConnect (EC) [15], Pluralistic Image Completion (PIC) [22], LaFIn [21]. To have fair comparison, the three methods have been trained using the E2Fdb. For quantitatively measuring the performance difference among the methods, we employ several statistical metrics. Moreover, to measure the amount of preservation of demographic and biometric features, we calculate False Non-Match Rate between original and inpainted faces. Using a competitive face biometric matcher [49] based on ArcFace [36].

A. DATASETS
We conduct all experiments on our generated dataset called E2Fdb (available on project's GitHub page)[2] extracted from the well-known CelebA-HQ dataset [32], [49]. To extract the periocular region from each face image, the images are

---

[2]https://github.com/amiretefaghi/E2F-GAN

TABLE 2. The effect of various parts of E2F-GAN on final results.

Edge Predictor	Attention Block	Refinement module	FID ↓	SSIM ↑	PSNR ↑	TV ↓	ℓ_1Loss ↓
✗	✓	✓	64.89	0.34	12.18	12.91	45.99
✓	✗	✓	50.12	0.48	13.24	6.33	43.97
✓	✓	✗	75.21	0.46	13.22	5.49	42.13
✓	✓	✓	**46.39**	**0.52**	**13.66**	**0.02**	**41.54**

Masked Image	PIC	EC	LaFIn	E2F-GAN(Ours)	Original

FIGURE 5. Quality comparison among PIC, EC, LaFIn, and our proposed method.

reshaped to size 256 × 256 and then by utilizing a landmark detector [27], eyes are detected, similar to [50]. Doing this, *M* and I_m are produced for each image. Moreover, we removed misleading samples including those eyes covered by sunglasses or faces that have more than 45 degrees in one angle (roll, pitch, yaw) leading to hiding one of the eyes by using WHENet [33] algorithms. Finally, the total number of samples is 24,554 among which 22,879 will be used for the training process and the rest, which is 1,685 images, for the test.

B. EVALUATION METRICS
We evaluate the image inpainting performance of the proposed model using quantitative and qualitative comparisons.
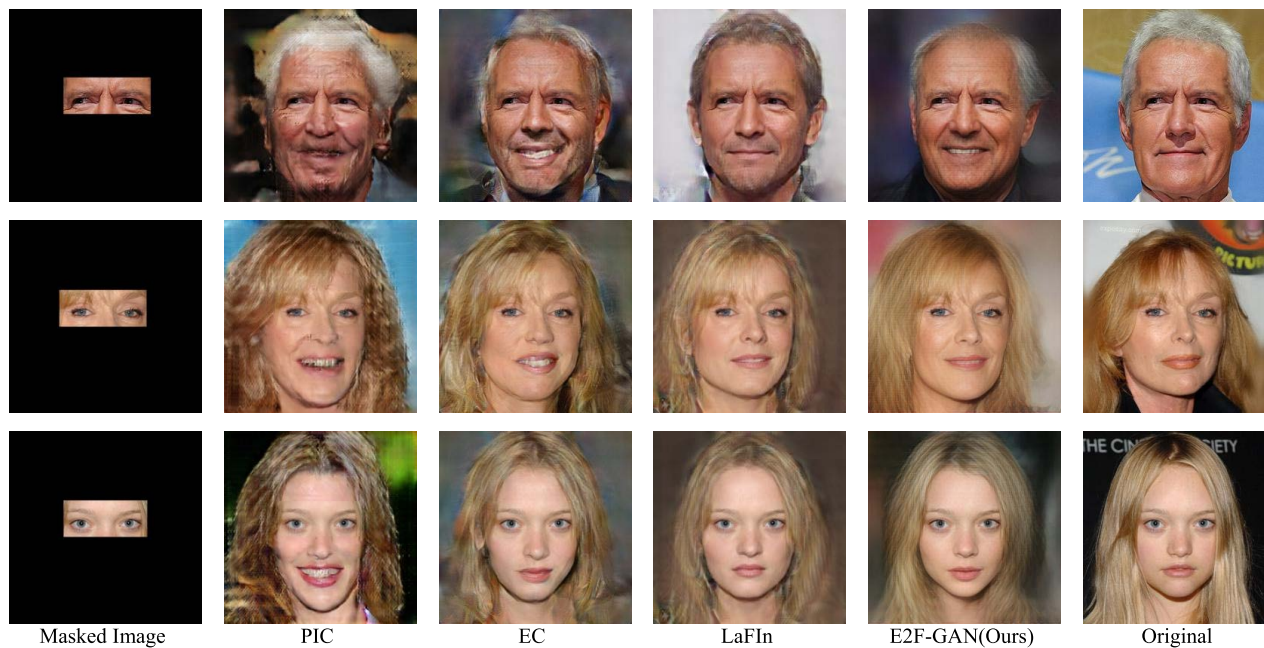
FIGURE 6. Illustration of image reconstruction at different age of the subject among our proposed method, PIC, EC, LaFIn, and original image.

For quantitative comparison, two types of metrics called statistical and identity metrics have been measured. In the following, we describe each category and its corresponding metrics briefly.

1) STATISTICAL METRICS

We use five statistical metrics: ℓ_1 loss, Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) [34], Frenchet Inception Distance (FID) [35], and Total Variation (TV). Notably, the ℓ_1 loss shows the model's reconstruction ability for images. PSNR measures the visibility of errors between the ground truth I_g and image inpainting $\hat{I}$ to evaluate the image quality. SSIM aims at estimating the perceptual changes in the structural information, which shows human's subjective feelings more accurately than PSNR. FID is a widely used metric in the image generation field to measure the visual quality. TV assists to measure the amount of noise in the image by calculating the sum of the absolute differences for neighboring pixels.

2) IDENTITY METRICS

To measure the amount of preservation of demographic and biometrics characteristics after completing inpainting process, we calculate the False Non-Match Rate (FNMR). FNMR is the rate at which a biometric algorithm miss-categorizes two captures from the same individual as being from different individuals. Here, we assumed that $\hat{I}$ and I are two faces for the same individual and using Arc-Face [36], we calculate the corresponding embedding vectors for each face, and finally calculate the cosine similarity between each pair. Finally, the FNMR for different thresholds is shown.

C. COMPARISON WITH EXISTING WORK

By using the above-mentioned metrics and presenting some outputs, the results of our proposed method have been qualitatively and quantitatively compared against three state-of-the-art approaches, named PIC, EC, and LaFIn. We trained the three methods over our generated dataset (i.e., E2Fdb) according to the best configurations of each method mentioned in the corresponding paper. In the following subsections, we present the results.

1) QUANTITATIVE COMPARISONS

The results of the statistical metrics calculated on the validation set of E2Fdb including 1,675 samples are reported in Table 1. As can be seen from the numbers in Table 1, E2F-GAN is superior over PIC, LaFIn, and EC in most metrics, except for the ℓ_1 loss for which LaFIn works slightly better. Overall, our E2F-GAN outperforms the others by large margins in terms of FID, SSIM, PSNR, and TV metrics. More specifically, our large margins in FID and TV metrics demonstrate that our method can inpaint the masked image with much higher quality compared to other methods. Moreover, FNMR has been measured for E2F-GAN and other three compared methods as shown in Fig. 4. For different thresholds, E2F-GAN has lower false non-match rate which shows the ability of our algorithm extracting identity information from the periocular region and transferring it to the reconstructed face. Notably, since the PIC method generates different outputs for a specific input, we executed this method five times and the best results have been reported.

2) QUALITATIVE COMPARISONS

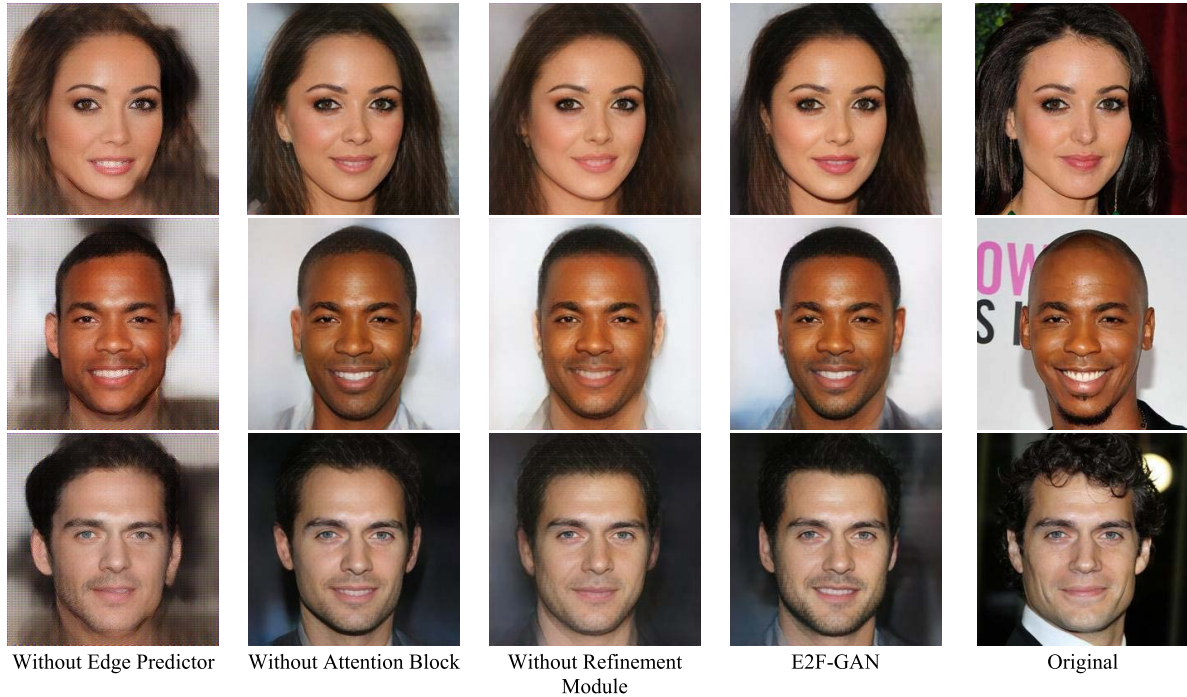Fig. 5 shows some faces generated by our model, PIC, EC, and LaFIn. Our model is able to generate high quality results

Without Edge Predictor	Without Attention Block	Without Refinement Module	E2F-GAN	Original

FIGURE 7. Illustrative comparison of the effect of various parts on final output.

TABLE 3. The effect of landmark or edge guidance on final results.

	FID ↓	SSIM ↑	PSNR ↑	TV ↓	ℓ_1Loss ↓	Landmark Loss ↓
Landmark Guidance	48.56	0.44	**13.33**	7.31	42.48	**15.37**
Edge Guidance	**46.39**	51.9	13.66	**0.02**	**41.54**	15.42

FIGURE 8. The impact of various parts of E2F-GAN on FNMR ratio.

and a large fraction of face structures including face shape, nose, mouth, forehead, etc. are appropriately placed with a plausible size. Moreover, to compare the quality of results in terms of gender and skin color, we present different faces in Figs. 5 and 6. As it can be observed, the quality of PIC and EC is really low compared to our and LaFIn results. Therefore, although like EC we used edge predictor in our scheme, there is a large margin between our outcomes.

Additionally, with aim of further investigation of the models' outputs regarding age and gender prediction based on the periocular region, we presented some challenging examples in Fig. 6. That figure shows three faces including two elders (a man and a woman) and a young woman. As seen in those examples, E2F-GAN can assess the age based on periocular region and reconstruct the face with a reasonable quality.

V. ABLATION STUDY
In this section, firstly we qualitatively and quantitatively analyze the effect of three main components of our proposed model including the edge predictor, the refinement module, and the attention block. Table 2 and Fig. 8 report statistical and identity metrics indicating the degree of effectiveness each of the three components in the performance
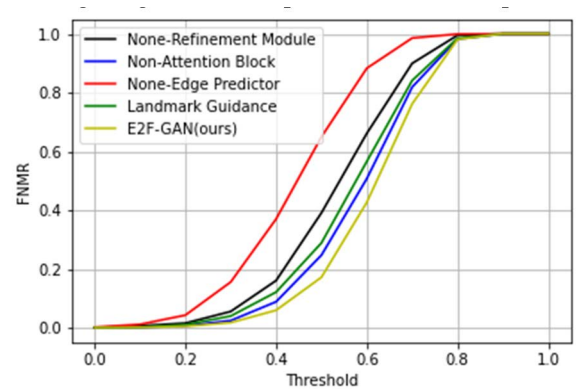
of E2F-GAN. Specifically, the refinement network is the most conspicuous one which benefits the model by providing conformity and consistency among face components and skin texture around the eyes, such as wrinkles and skin color. The edge guidance contributes to ensuring that the structure of the face is well-preserved (see Fig. 7). Visually, the effectiveness of the attention block may not seem tangible. However, the quantitative results demonstrate the advantages of attention block. We also compared the effect of edge and landmark predictors. As shown in Table 3, the edge guidance provides better values in most quantitative metrics specially for SSIM metric.

Finally, Fig. 9 shows a few challenging examples for preserving the gender of the person based on the periocular region. Our observations show that E2F-GAN can preserve the gender of subjects with a high accuracy.
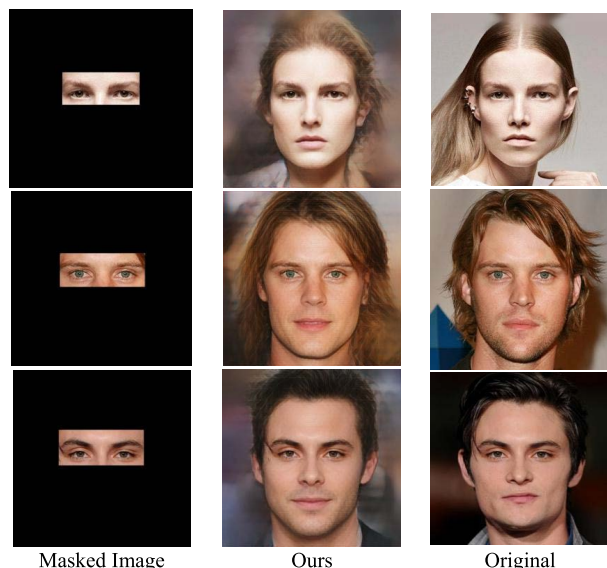
Masked Image	Ours	Original

FIGURE 9. Illustration of gender preserve in our proposed method.

VI. CONCLUSION

The aim of this paper is a particular case of face inpainting where we try to reconstruct the face based on just using the periocular region. To do this, we presented E2F-GAN, a GAN-based architecture that benefits from the advantages of coarse-to-fine, coarse-and-fine, and structural guidance-based architectures for face inpainting. It includes three main modules for extracting face's edges (edge predictor), coarse prediction of face elements (coarse generator) and refining the coarse predicted image (refinement generator). We analyzed E2F-GAN and compared it with other well-known face inpainting methods to measure the efficiency and quality performance. For doing this, we modified a widely used face inpainting dataset called CelebA-HQ such that the whole face except the periocular region is masked and used for E2F-GAN input, calling the resulting dataset E2Fdb. Our proposed inpainting algorithm E2F-GAN and the used dataset E2Fdb are both available in the project GitHub.[3]

Several qualitative and quantitative metrics have been measured during our experiments to show the performance of E2F-GAN in terms of preserving identity and non-identity features of each face after inpainting. Experimental results show that our method outperforms previous learning-based face inpainting methods and E2F-GAN can generate realistic and semantically plausible images.

Future work includes analyzing biometric quality aspects of the resulting faces using recent objective measures [51], [52]; analyzing [49] and reducing [53] undesired biases in the face generation process; and combining multiple face generation approaches for better outputs [48].

[3]https://github.com/amiretefaghi/E2F-GAN

REFERENCES

[1] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[2] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," *Neural Process. Lett.*, vol. 51, no. 2, pp. 2007–2028, 2019.

[3] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.

[4] H. Yamauchi, J. Haber, and H.-P. Seidel, "Image restoration using multiresolution texture synthesis and image inpainting," in *Proc. Comput. Graph. Int.*, Jul. 2003, pp. 120–125.

[5] X. Zhu, Y. Qian, X. Zhao, B. Sun, and Y. Sun, "A deep learning approach to patch-based image inpainting forensics," *Signal Process., Image Commun.*, vol. 67, pp. 90–99, Sep. 2018.

[6] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3844–3851.

[7] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1486–1494.

[8] H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4170–4179.

[9] Z. Guo, Z. Chen, T. Yu, J. Chen, and S. Liu, "Progressive image inpainting with full-resolution residual network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2496–2504.

[10] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, "Recurrent feature reasoning for image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7760–7768.

[11] H. Zhang, Z. Hu, C. Luo, W. Zuo, and M. Wang, "Semantic image inpainting with progressive generative networks," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1939–1947.

[12] H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4170–4179.

[13] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[14] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4471–4480.

[15] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edge-Connect: Structure guided image inpainting using edge prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3265–3274.

[16] M. Chen and Z. Liu, "EDBGAN: Image inpainting via an edge-aware dual branch generative adversarial network," *IEEE Signal Process. Lett.*, vol. 28, pp. 842–846, 2021.

[17] E. Bobrov, A. Georgievskaya, K. Kiselev, A. Sevastopolsky, A. Zhavoronkov, S. Gurov, K. Rudakov, M. D. P. B. Tobar, S. Jaspers, and S. Clemann, "PhotoAgeClock: Deep learning algorithms for development of non-invasive visual biomarkers of aging," *Aging (Albany NY)*, vol. 10, no. 11, p. 3249, 2018.

[18] C. Rathgeb and C. Busch, *Handbook Iris and Periocular Biometric Recognition*. London, U.K.: Institution of Engineering and Technology (IET), 2017.

[19] M. Chen, Z. Liu, L. Ye, and Y. Wang, "Attentional coarse-and-fine generative adversarial networks for image inpainting," *Neurocomputing*, vol. 405, pp. 259–269, Sep. 2020.

[20] L. Song, J. Cao, L. Song, Y. Hu, and R. He, "Geometry-aware face completion and editing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, Jul. 2019, pp. 2506–2513.

[21] Y. Yang and X. Guo, "Generative landmark guided face inpainting," in *Proc. Pattern Recognit. Comput. Vis.*, Oct. 2020, pp. 14–26.

[22] C. Zheng, T.-J. Cham, and J. Cai, "Pluralistic image completion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1438–1447.

[23] L. Zhao, Q. Mo, S. Lin, Z. Wang, Z. Zuo, H. Chen, W. Xing, and D. Lu, "UCTGAN: Diverse image inpainting based on unsupervised cross-space translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5741–5750.

[24] N. U. Din, K. Javed, S. Bae, and J. Yi, "A novel GAN-based network for unmasking of masked face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020.

[25] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1486–1494.

[26] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 85–100.

[27] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1021–1030.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241.

[29] C. T. Li, W. C. Siu, Z. S. Liu, L. W. Wang, and D. P. K. Lun, "Deep-GIN: Deep generative inpainting network for extreme image inpainting," in *Proc. Comput. Vis. ECCV 2020 workshops*, Glasgow, U.K., 2020, pp. 5–22.

[30] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. ICLR*, Vancouver, BC, Canada, 2018, pp. 1–29.

[31] K. Lata, M. Dave, and K. N. Nishanth, "Image-to-image translation using generative adversarial network," in *Proc. 3rd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Jun. 2019, pp. 186–189.

[32] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability and variation," in *Proc. ICLR*, Vancouver, BC, Canada, 2018.

[33] Y. Zhou and J. Gregson, "WHENet: Real-time fine-grained estimation for wide range head pose," in *Proc. 31st Brit. Mach. Vis. Conf. (BMVC)*, 2020.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[35] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637.

[36] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4690–4699.

[37] F. Alonso-Fernandez, K. Hernandez-Diaz, S. Ramis, F. J. Perales, and J. Bigun, "Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images," *IET Biometrics*, vol. 10, no. 5, pp. 562–580, Sep. 2021.

[38] F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, J. Bigun, R. Vera-Rodriguez, and J. Fierrez, "Cross-sensor periocular biometrics for partial face recognition in a global pandemic: Comparative benchmark and novel multialgorithmic approach," 2019, *arXiv:1902.08123*.

[39] F. Alonso-Fernandez, R. A. Farrugia, J. Fierrez, and J. Bigun, "Super-resolution for selfie biometrics: Introduction and application to face and iris," in *Selfie Biometrics*. Cham, Switzerland: Springer, 2019, pp. 105–128.

[40] R. Tolosana, S. Romero-Tapiador, R. Vera-Rodriguez, E. Gonzalez-Sosa, and J. Fierrez, "DeepFakes detection across generations: Analysis of facial regions, fusion, and performance evaluation," *Eng. Appl. Artif. Intell.*, vol. 110, Apr. 2022, Art. no. 104673.

[41] A. Morales, J. Fierrez, A. Acien, R. Tolosana, and I. Serna, "SetMargin loss applied to deep keystroke biometrics with circle packing interpretation," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108283.

[42] P. Tome, J. Fierrez, R. Vera-Rodriguez, and D. Ramos, "Identification using face regions: Application and assessment in forensic scenarios," *Forensic Sci. Int.*, vol. 233, nos. 1–3, pp. 75–83, Dec. 2013.

[43] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, "Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 8, pp. 2001–2014, Aug. 2018.

[44] P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia, "Combination of face regions in forensic scenarios," *J. Forensic Sci.*, vol. 60, no. 4, pp. 1046–1051, Jul. 2015.

[45] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J. C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "Deep learning for understanding faces: Machines May be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.

[46] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Facial soft biometric features for forensic face recognition," *Forensic Sci. Int.*, vol. 257, pp. 171–284, Dec. 2015.

[47] J. Hernandez-Ortega, J. Fierrez, I. Serna, and A. Morales, "FaceQgen: Semi-supervised deep learning for face image quality assessment," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.

[48] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "Multiple classifiers in biometrics. Part 2: Trends and challenges," *Inf. Fusion*, vol. 44, pp. 103–112, Nov. 2018.

[49] I. Serna, A. Morales, J. Fierrez, and N. Obradovich, "Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning," *Artif. Intell.*, vol. 305, Apr. 2022, Art. no. 103682.

[50] R. Daza, D. DeAlcala, A. Morales, R. Tolosana, R. Cobos, and J. Fierrez, "ALEBk: Feasibility study of attention level estimation via blink detection applied to e-learning," in *Proc. AAAI Workshop Artif. Intell. Educ. (AI4EDU)*, 2022.

[51] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "FaceQNet: Quality assessment for face recognition based on deep learning," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[52] J. Hernandez-Ortega, J. Fierrez, L. F. Gomez, A. Morales, J. L. Gonzalez-de-Suso, and F. Zamora-Martinez, "FaceQvec: Vector quality assessment for face biometrics based on ISO compliance," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 84–92.

[53] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning agnostic representations with application to face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2158–2164, Jun. 2021.

AHMAD HASSANPOUR received the M.Sc. degree in computer engineering from the Shiraz University of Technology. He is currently pursuing the Ph.D. degree with the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He is also a Marie Skłodowska-Curie Fellow and part of an EU project called Privacy Matters. His research interests include deep learning, computer vision, and privacy.

AMIR ETEFAGHI DARYANI received the B.Sc. degree in electrical engineering from the University of Guilan, in 2019. He is currently pursuing the M.Sc. degree in electrical engineering with the Amirkabir University of Technology.

He was a Research Assistant at the Digital Systems Laboratory, Amirkabir University of Technology, which was directed by Prof. Saeed Sharifian to pursue his M.Sc. degree. His research interests include the development of decision-making systems based on the reinforcement learning approaches and image processing by using deep learning-based techniques to build autonomous systems that act like humans.

MAHDIEH MIRMAHDI received the B.Sc. degree in computer science and the M.Sc. degree in computer engineering-artificial intelligence and robotics from the University of Isfahan, in 2018 and 2021, respectively. During her master's degree, she focused on computer vision, specifically the deep learning, and worked on video semantic segmentation under the supervision of Prof. Amirhassan Monadjemi. Her research interests include in the intersection of computer vision, machine learning, and 3D visual computing. She was awarded as an exceptional talent in computer science at the University of Isfahan, in 2018.

KIRAN RAJA (Senior Member, IEEE) received the Ph.D. degree in computer science from the Norwegian University of Science and Technology, Norway, in 2016. He is currently a Faculty Member of the Department of Computer Science, Norwegian University of Science and Technology. He was/is participating in EU projects SOTAMD and iMARS, and other national projects. His main research interests include statistical pattern recognition, image processing, and machine learning with applications to biometrics, security, and privacy protection. He is a member of the European Association of Biometrics (EAB) and the Chair of the Academic Special Interest Group at EAB. He is also a member of the editorial board for various journals. He serves as a reviewer for a number of journals and conferences.

BIAN YANG received the B.S., M.S., and Ph.D. degrees in electronic engineering and information security from the Harbin Institute of Technology, in 2000, 2002, and 2006, respectively. He visited Fraunhofer IGD, Darmstadt, from 2003 to 2005, for research on data hiding and media security, worked as a Lecturer at the School of Computer Science and Technology, Harbin Institute of Technology, from 2005 to 2007, and as a Research Engineer at Thomson Corporate Research, Beijing, from 2007 to 2008. He joined the Norwegian Information Security Laboratory (NISlab), Gjøvik University College, working on privacy-preserving biometrics, from 2008 to 2015, and founded in 2016 and has since been coordinating the eHealth and Welfare Security (eHWS) Group, Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU). His research interests include cybersecurity and privacy for e-health and welfare technologies and services, privacy modeling and enhancing technologies, security biometrics and identity management, and security practice and human factors.

CHRISTOPH BUSCH (Senior Member, IEEE) received the Ph.D. degree in computer graphics, in 1997.

He is a member of the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with the Computer Science Faculty, Hochschule Darmstadt (HDA), Germany. Further, he has been lecturing on biometric systems at the Technical University of Denmark (DTU), since 2007. In 1997, he joined the Fraunhofer Institute for Computer Graphics (Fraunhofer IGD) as the Head of the Department of Security Technology. He has been responsible for the acquisition, the management, and the control of various applied research and development projects. On behalf of the German Federal Office for Information Security (BSI), he has been the responsible Project Coordinator for the project series BioIS, BioFace and BioFinger, BioKeyS Pilot-DB, and KBEinweg—all projects dealing with biometric applications in general and performance and security testing in detail. In the European research program, he was the Initiator of multiple projects (e.g., 3D-Face, FIDELITY, and iMARS). Further, he was/is the partner in the projects TURBINE, BEST Network, ORIGINS, INGRESS, PIDaaS, SOTAMD, RESPECT, and TReS-PAsS. He is also the Principal Investigator at the National Research Center for Applied Cybersecurity (ATHENE). He has coauthored more than 600 technical papers and has been a speaker at international conferences. He has served for various program committees (NIST IBPC, ICB, BTAS, IJCB, ISBA, BSI-Congress, and EUROGRAPHICS) and a Reviewer of journals, such as *ACM-SIGGRAPH*, *ACM-TISSEC*, IEEE Computer Graphics and Applications, IEEE Transactions on Signal Processing, IEEE Transactions on Information Forensics and Security, IEEE Transactions on Pattern Analysis and Machine Intelligence, and the Elsevier journal *Computers and Security*. He is also an appointed member of the editorial board of the *IET Biometrics* and has served previously for the IEEE Transactions on Information Forensics and Security journal.

JULIAN FIERREZ (Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Universidad Politécnica de Madrid, Spain, in 2001 and 2006, respectively. Since 2004, he has been with the Universidad Autónoma de Madrid, where he is currently an Associate Professor. From 2007 to 2009, he was a Visiting Researcher with Michigan State University, USA, under a Marie Curie Postdoctoral. His research interests include signal and image processing, HCI, responsible AI, and biometrics for security and human behavior analysis. He is actively involved in large EU projects in these topics (e.g., TABULA RASA and BEAT in the past and currently IDEA-FAST and TRESPASS-ETN), and has attracted notable impact for his research. He is a member of the ELLIS Society. He was a recipient of a number of distinctions, including the EAB Industry Award 2006, the EURASIP Best Ph.D. Award 2012, and the 2017 IAPR Young Biometrics Investigator Award. He has received best paper awards at ICB and ICPR. He is an Associate Editor of the IEEE Transactions on Information Forensics and Security and the IEEE Transactions on Image Processing.

. . .