.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Fingerphoto Deblurring Using Attention-guided Multi-stage GAN

AMOL S. JOSHI, ALI DABOUEI, JEREMY DAWSON, AND NASSER M. NASRABADI, (Fellow, IEEE)
West Virginia University, Morgantown WV, USA

Corresponding author: Amol S. Joshi (e-mail: asj00003@mix.wvu.edu), and Nasser M. Nasrabadi (e-mail: nasser.nasrabadi@mail.wvu.edu).

ABSTRACT Using fingerphoto images acquired from mobile cameras, low-quality sensors, or crime scenes, it has become a challenge for automated identification systems to verify the identity due to various acquisition distortions. A significant type of photometric distortion that notably reduces the quality of a fingerphoto is the blurring of the image. This paper proposes a deep fingerphoto deblurring model to restore the ridge information degraded by the image blurring. As the core of our model, we utilize a conditional Generative Adversarial Network (cGAN) to learn the distribution of natural ridge patterns. We perform several modifications to enhance the quality of the reconstructed (deblurred) fingerphotos by our proposed model. First, we develop a multi-stage GAN to learn the ridge distribution in a coarse-to-fine framework. This framework enables the model to maintain the consistency of the ridge deblurring process at different resolutions. Second, we propose a guided attention module that helps the generator to focus mainly on blurred regions. Third, we incorporate a deep fingerphoto verifier as an auxiliary adaptive loss function to force the generator to preserve the ID information during the deblurring process. Finally, we evaluate the effectiveness of the proposed model through extensive experiments on multiple public fingerphoto datasets as well as real-world blurred fingerphotos. In particular, our method achieves 5.2 dB, 8.7%, and 7.6% improvement in PSNR, AUC, and EER, respectively, compared to a state-of-the-art deblurring method.

INDEX TERMS Biometrics, contactless fingerprints, fingerphoto deblurring, generative adversarial networks, guided attention, multi-stage generative architecture.

I. INTRODUCTION

IDENTIFYING humans with unique biometric features is a tremendous advantage to security and identification tasks. It offers immense benefits in cases such as crime-solving, evidence collection, and human verification and authorization. Recent advances in deep learning have demonstrated promising performance in biometric identification [1]. Zhao and Kumar [2] proposed a deep model to learn descriptive iris features for iris recognition. Taigman *et.al.* [3] used a deep learning approach to learn face representation and alignment such that a simple classifier could recognize the face ID. Even though real-world applications adopt traits like the iris and face in several biometric systems, they still have limitations. For instance, current iris acquisition systems require significant cooperation from the subject, whereas face images are prone to manipulation or pose variations. Compared to other biometric traits, fingerprints have several advantages. First of all, the fingerprints contain such unique features that a partial sample is sufficient for identification, and even different fingers of the same person have different unique patterns which, in combination, can provide abundant evidence for identification [4]. This phenomenon makes fingerprints one of the most robust biometric modalities for identification purposes.

In crime scene investigations, unlike other traits, fingerprints leave traces after the interaction of individuals with their surrounding objects that can be used to verify their presence in the area. It is a major advantage that allows law enforcement officials to identify potential suspects. The ease of collecting fingerprints using a wide range of capturing devices, such as mobile cameras, ink, sensors, *etc.*, is an exceptional feature of the fingerprint. Based on the collection techniques, the fingerprint samples are divided into contact-based, *i.e.*, collected using a sensor that requires physical interaction and contact with a surface, or contactless, referring to the process of capturing fingerprints with a digital

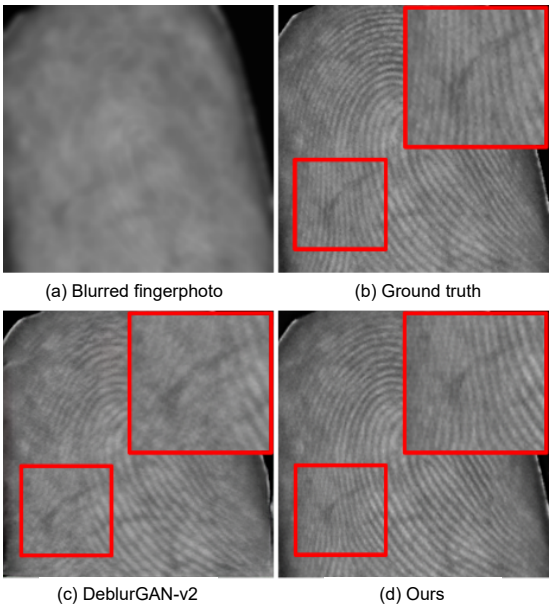

FIGURE 1. Comparison of deblurred images by DeblurGAN-v2 [5] and our method. Image (a) is a blurred sample. (b) Corresponding ground truth fingerphoto. (c) Deblurred fingerphoto from DeblurGAN-v2, and (d) is the deblurred fingerphoto by our proposed model. The highlighted regions in the red blocks show the superiority of our method in ridge reconstruction during deblurring.

camera or a smartphone. Hence, under many circumstances where a quick, hygienic, and reliable verification is required, fingerphotos are certainly ideal. These include on-the-go quick person recognition and all other cases where the legacy contact-based fingerprints were being used [6], [7].

Although they have several benefits compared to contact-based samples, the recognition systems for fingerphotos captured in the wild need to be more accurate due to the distortions which need to be addressed sufficiently [8]. Photos of latent impressions acquired from crime scenes or samples captured by law enforcement agents using a smartphone's camera may not exhibit the same quality as contact-based fingerprints. The photometric distortion often deteriorates the quality of fingerphotos. It is mainly caused by the non-ideal conditions of the capturing device, *i.e.*, out-of-focus lens, motion blur, perspective distortion, *etc* [9]. Blurring is a common type of photometric distortion that can be caused by several factors, such as human errors, trembling fingers, a slow frame rate of the capturing sensor [10], inappropriate focusing of the camera, or intentional blurring by malicious users to evade being identified. Recognition of such blurred samples is a cumbersome task that even state-of-the-art systems may fail to accomplish. Therefore, deblurring is an ineluctable step in the recognition of blurred fingerphotos.

Since the emergence of generative adversarial networks (GAN) [11], generative tasks like image restoration and translation have become far more efficient and reliable. With the advent of different approaches and recent developments, GAN training has become stable enough to produce realistic images under specific conditional settings. Hence, conditional GAN (cGAN) models [12] have become the principal approach for image restoration tasks, such as image deblurring [5], [13], [14], [15], [16], partial image reconstruction [17], [18], or natural image denoising and inpainting [19], [20], [21], [22]. Previous works on natural image deblurring have extensively adopted convolutional neural networks (CNN) for deblurring. However, the shortcoming of algorithms trained with natural images is that they may not be well-suited for fingerprint ridge patterns. The algorithms [15], [23], [24], [25] that address a particular type of blurring effect may not be helpful in real-world scenarios, where the blurring magnitude and kernels are entirely arbitrary.

The techniques based on the generative models and the fusion of the image characteristics perform well in deblurring. Zheng *et.al.* [14] used edge heuristics and a GAN model to remove the non-uniform blurring from dynamic scenes. This model deblurs the images precisely, but its performance may plunge if the underlying identity (ID) information needs to be preserved. Another approach uses the statistical and physical features of the blurring kernel to remove the blurring effect. The authors in [26] proposed a mathematical model-based technique to deblur Gaussian or motion blur. This method does not require vast amounts of blurred images to train and can still produce state-of-the-art deblurred images. Nevertheless, this method may not be as effective due to the abnormal randomness in blurring kernels of real-world blurred images.

Here, we seek to address the blurring problem in fingerphotos using a deep deblurring model. We treat this problem as a domain translation task in which the source and target domains are blurred and clean fingerphotos, respectively. We develop a deep model to learn the inverse mapping between the two domains given a large dataset of synthetically blurred fingerphoto samples with multiple types of blurring kernels such as Gaussian, motion, and defocusing. The proposed deep convolutional neural network model accepts a blurred fingerphoto and returns the deblurred version. Here, we use cGAN architecture consisting of a generator and discriminator subnetworks as the base model for the deblurring task and enhance its functionality by considering four modifications to the cGAN architecture. In the first modification, we extract the intermediate features from different layers of the generator and feed them to their corresponding discriminators, which allows the generator to capture ridge information at different scales, from coarse ridge patterns to fine ridge details. In the second modification, we use a deep fingerphoto verifier to force the generator to preserve the ID of the deblurred fingerphotos during the deblurring process.

Contrary to other image deblurring tasks, in fingerphoto deblurring, the ridges encompass unique structures and properties that represent the fingerphoto ID information. The deblurring process may alter this information by adding spurious ridge/valley patterns and removing the critical minutiae information. Hence, we use a pre-trained deep fingerphoto verifier as an additional loss term to ensure that the gen-

erator is preserving the fingerphoto's ID information while deblurring fingerphotos. In the third modification, we use a guided attention block to force the network to pay attention to the partially blurred regions in the fingerphotos. This way, the trained network is agnostic to blurred and sharp region imbalance in defocused images. Lastly, we introduce a multi-task learning approach to effectively augment the training supervision and help the model specialize in different types of blurring. To this aim, we design an additional task of blurring type classification. All the auxiliary modules, discriminators, and verifiers work simultaneously to enhance the deblurring process. A sample result of deblurring using the proposed model and a state-of-the-art method DeblurGAN-v2 [5] is shown in Figure 1. In summary, our contributions are as follows:

- We propose a deep multi-task, multi-stage generative model to solve the challenging task of ridge deblurring process in fingerphotos that has not been previously addressed by biometrics researchers.
- A deep fingerphoto verifier is incorporated to force the generator to preserve the ID and ridge information of fingerphotos during the deblurring process.
- The performance of deblurring is further enhanced by designing a multi-stage generative process based on coarse-to-fine supervision.
- We additionally use guided-attention and multi-task learning approaches to increase the robustness of the model towards non-uniform and partial blurring with complex blurring kernels.

Further, in Section II, we review previous work. Then, section III illustrates the proposed method, and later in the paper, Sections IV & V present the experimentation and an extensive ablation study, respectively. Finally, Section VI concludes the paper.

II. RELATED WORK

In this section, we first review the literature on fingerphoto data acquisition and preprocessing. Then, we discuss the recent advancements in image deblurring, especially the methods based on deep learning. Then, we discuss several techniques we built upon to improve the performance of the proposed deblurring model, including multi-task learning, attention mechanism, and the approach to preserve the identification information during the image generation process.

A. FINGERPHOTO DATA ACQUISITION AND PREPROCESSING

Automated fingerphoto identification systems consist of three major parts, data acquisition, preprocessing, and matching [27], [28], [9]. Multiple techniques have been developed for fingerphoto acquisition. Compared to expensive contact-based fingerprint-specific sensors, digital cameras are cheaper and readily available since they are embedded in most smartphones. Due to this ubiquity, smartphones have drawn notable attention in biometrics. The authors in [29], [30], and [31] use smartphone cameras to capture fingerphotos whereas, other researchers [32], [33], [34] use digital cameras to acquire fingerphotos. Stein *et al.* [35] explored the data acquisition and quality assurance methods necessary to realize a reliable fingerphoto recognition system using an edge density metric based on the sharpness of the captured image.

Samples collected from contactless sensors or smartphone cameras often suffer from degradation, such as low ridge-valley contrast and perspective distortion. Several preprocessing algorithms have been developed to enhance the quality of fingerphotos. Lee *et al.* [36] proposed a preprocessing algorithm for finger images captured using a mobile camera. The proposed method involves segmentation and orientation estimation of the image. The common preprocessing approaches usually involve enhancing the ridge patterns [37], [38], [39], [40], and correcting the perspective distortion [41], [42].

The last step in identification systems is feature extraction and matching fingerphotos. The performance of current algorithms for feature extraction and matching depends highly on the quality of the input samples. Many studies have been conducted to improve feature extraction [42], [43], [44]. In [34], the authors surveyed several touchless recognition technologies and studied non-idealities such as blur, defocusing, noise, and perspective distortion that arise during fingerphoto acquisition. However, the previous methods do not address the problem of motion blur or defocusing issues that affect the fingerphoto recognition performance.

B. IMAGE DEBLURRING

While working with low-quality fingerphotos, such as blurred or distorted images, additional preprocessing steps, such as deblurring, are inevitable. The deep learning literature presents numerous algorithms and techniques to undertake image deblurring. However, these techniques are implemented on natural images [5], [14], [15], [16] or other biometric modalities such as faces [45], [46]. To the best of our knowledge, the fingerphoto deblurring problem has yet to be addressed profoundly in the literature.

Based on prior information about the blurring kernel, deblurring can be classified into two types: blind and non-blind. In blind deblurring, the kernel of the blurring effect is unknown. Therefore, some algorithms have been proposed to estimate the blurring kernel and remove the blur from the image [23], [47], [24]. Approaches without kernel estimation use image information and learn to deblur the images [5], [15], [48]. On the other hand, in non-blind deblurring, the blurring kernel is known. One of the techniques in non-blind deblurring uses statistical information about the blurring kernel to deblur the image [49], [50], [51]. Several attempts have been made in the literature to deblur images with different blurring causes, such as motion blur, Gaussian blur, defocusing, etc. For instance, [5] used a feature pyramid network as a generator with a GAN-based learning model to remove motion blur from natural images. Using a

double-scale discriminator and lightweight backbones, they improved deblurring accuracy and efficiency. The work presented in [52], [53], and [54] used a coarse-to-fine strategy such that, at every scale, a sharp latent image is produced.

Existing methods for deblurring biometric images are primarily focused on other biometric traits such as iris [55], [56], face [45], [46], [57] and hand-based biometric traits such as palm prints [58] or finger wrinkles [13]. Cho *et al.* [13] proposed a GAN-based model for deblurring finger wrinkles for authentication. Despite the potential usefulness of contactless fingerprint recognition, fingerphoto deblurring has not received sufficient attention. Even though deep image deblurring models achieve much higher accuracy, a deblurred fingerphoto needs ID preservation in order to be accurately matched with the ground truth fingerphoto.

C. PRESERVING ID INFORMATION

Although generative models like GANs are applicable in synthesizing samples from a given dataset, it may be necessary to preserve the semantic information of the input images. Specifically, in the case of biometric data, identity information is crucial for further identification processes. A plethora of algorithms has been proposed to preserve ID information during generative processes. Algorithms based on natural image generation use some distance measure to minimize the distance between the synthetic and real image [5], [15], [52]. To further improve the performance of ID-specific data, algorithms were proposed to minimize the distance between the embedding of the real and generated images. In order to achieve this, [17], [31], [59], [60], [61] used the L_2 norm distance on the features extracted from some pre-trained network which is trained on the data of the corresponding domain. The efforts discussed in [62] & [63] used cross-entropy-based identity loss to preserve the ID during image generation. Authors in [64], [65], [66] proposed some manipulations in the GAN architecture along with a combination of some distance measures to preserve the identity better.

D. MULTI-TASK LEARNING

Multi-task learning has demonstrated its significance in several deep learning applications such as classification [67], [68], [69], restoration [70], [71], [72] and translation [73], [74]. Ranjan *et al.* [75] used a multi-task learning-based approach for face detection, landmark localization, pose estimation, and gender recognition. They show that the feature fusion and learning inter-related tasks helped the network to better generalize on each task. Dai *et al.* [76] used multi-task learning for a semantic segmentation task and simultaneously improved object detection accuracy. Although multi-task learning is not explored enough for the deblurring task, the algorithm in [74] learns to deblur the image and generate a motion field simultaneously. The significant improvement in deblurring performance shows its effectiveness in removing motion blur from dynamic scenes.

E. ATTENTION MECHANISM

After outstanding achievements in natural language processing, attention mechanisms have demonstrated their usefulness in the computer vision field for a long time. The authors in [77], [78], and [79] showed the phenomenal benefits of visual attention in deep neural networks. The work in [77] used the recurrent neural network and LSTM-based attention to generate image captions. In [78], the authors proposed a network to generate images from a given text. They applied the attention mechanism on multiple scales throughout the generator of the GAN model to achieve a fine-grained image reconstruction. Attention has also been used in domain translation tasks such as image deblurring. Suin *et al.* [80] used self-attention on a patch-hierarchical architecture to remove motion blurring in dynamic scenes. Shen *et al.* [81] proposed human-aware attention, which uses a supervised method to incorporate human awareness into the attention map.

III. PROPOSED METHOD

In this section, we describe our proposed framework consisting of a cGAN deblurring network with auxiliary subnetworks arranged to perform accurate deblurring of fingerphotos. Furthermore, in this section, we first go through the conditional GAN, then discuss the multi-stage deblurring scheme followed by the ID preservation technique, guided-attention mechanism, and multi-task learning approach. Lastly, we discuss the overall objective function and the proposed network architecture. The equations follow the following notations: bold lower case letters denote vectors, upper case letters denote functions, whereas calligraphic upper case letters represent sets.

A. CONDITIONAL GAN

When it comes to cross-domain transformation, GAN models [11], [12], [82] are the most popular generative networks. The original GAN maps the input sample $\boldsymbol{z}$ from a known random distribution $p_{\boldsymbol{z}}(\mathcal{Z})$ to the target domain such that $\boldsymbol{y} = G(\boldsymbol{z}, \boldsymbol{\theta_g}) : \boldsymbol{z} \rightarrow \boldsymbol{y}$, where $\boldsymbol{\theta_g}$ represents the trained parameters of the network. GANs generally consist of two networks, a generator, and a discriminator. The role of the generator is to produce accurate images of the target domain, whereas the discriminator distinguishes between the generated sample and the corresponding real sample. The feedback from the discriminator acts as an adaptive loss to guide the generator to do better in the transformation. In a nutshell, there is a min-max game going on between generator G and discriminator D. The objective function for a GAN is as follows:

$$\begin{aligned} O_{GAN}(G, D) = \mathbb{E}_{\boldsymbol{y} \sim \mathcal{P}_{data}(\boldsymbol{y})}[logD(\boldsymbol{y})] \\ + \mathbb{E}_{\boldsymbol{z} \sim \mathcal{P}_{z}(\boldsymbol{z})}[log(1 - D(G(\boldsymbol{z})))], \end{aligned} \tag{1}$$

where, given the ground truth as $\boldsymbol{y}$, the generator, G, tries to minimize the objective function, and the discriminator, D, maximizes it.

In [82], Isola *et al.* proposed a modification to the GAN model for training it in a constrained manner. In conventional

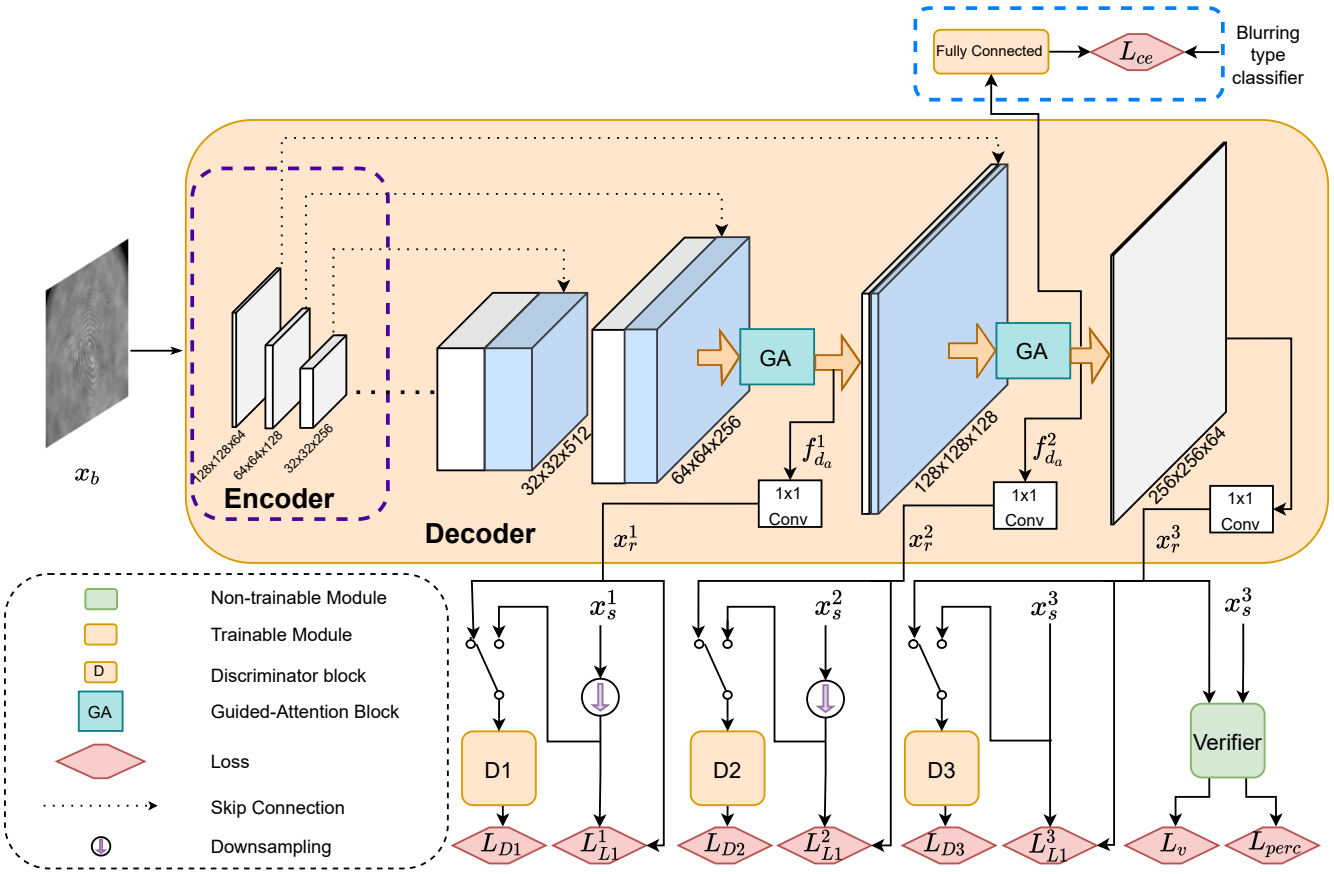

FIGURE 2. The architecture of the proposed model. Input to the generator is a blurred image $\boldsymbol{x}_b$ (256×256) and intermediate outputs are $\boldsymbol{x}_r^1$ (64×64) and $\boldsymbol{x}_r^2$ (128×128). The output of the generator is a 256×256 deblurred image $\boldsymbol{x}_r^3$. Each intermediate feature, along with the final output, has separate discriminators, D1, D2, and D3, respectively. $\boldsymbol{f}_{d_a}^1$ and $\boldsymbol{f}_{d_a}^2$ are the feature maps obtained from the guided-attention blocks (GAs), and $\boldsymbol{f}_{d_a}^2$ is used to predict the blurring type by a classifier. During the training, $\boldsymbol{x}_s^1$, $\boldsymbol{x}_s^2$, and $\boldsymbol{x}_s^3$ are used as the ground truth fingerphotos at resolutions 64×64, 128×128, and 256×256, respectively.

GANs, the input to the generator is from a predefined random distribution. Hence, the generator learns a mapping between the sample distribution and the unknown distribution of the target domain from which a set of training images are given. For cross-domain image translation, samples $\boldsymbol{x}$ from the source domain are added as the input to the network. The discriminator is also conditioned with the concatenated input of the target domain and the generated sample. After the modifications, the adversarial objective function becomes:

$$O_{cGAN}(G,D) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}_{data}}[logD(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{P}_{data}}[log(1-D(\boldsymbol{x},G(\boldsymbol{x})))]. \quad (2)$$

An additional L_2 or L_1 loss term is added to the objective function to calculate the error between the input and output such that it penalizes the generator for creating dissimilar outputs. The final optimization objective of the generative model is as follows:

$$G_{optimal} = \min_G \max_D O_{cGAN}(G,D) + \lambda L_{L_1}(\boldsymbol{y},G), \quad (3)$$

where λ is the Lagrangian coefficient to control the relative strength of the reconstruction loss and $\mathbf{y}$ is the ground truth fingerphoto. The L_1 distance is given by:

$$L_{L_1}(\boldsymbol{y},G) = \|\boldsymbol{y} - G(\boldsymbol{z})\|_1. \quad (4)$$

B. MULTI-STAGE DEBLURRING

Inspired by the multi-scale discriminator approach in [83], we develop an additional modification to the cGAN model. We introduce a multi-stage scheme to the deblurring network to leverage the underlying low-level information at different resolutions and avoid erroneous fingerphoto reconstruction. Fingerphotos naturally contain features at different scales (level-1 to level-3 features). Therefore, analyzing the quality of generated fingerphotos at different stages has the potential to provide more valuable information for training the network. To this aim, in addition to the final output of the generator G that has the spatial size of 256×256, we extract the intermediate feature maps from different layers of the decoder of the generator with spatial sizes 64×64 and 128×128 and force them to reconstruct the deblurred fingerphotos. Each of the three outputs at different resolutions is then passed to a dedicated discriminator to produce the adversarial loss. In this way, we create a multi-adversarial game between

the generator G and three different discriminators D_1, D_2, and D_3, as illustrated in Figure 2.

This multi-stage adversarial scheme provides additional supervision for the training of the generator by using multi-scale features and corresponding ground truth images. Note that the supervision for the intermediate stages can be conveniently obtained by downsampling the ground truth target fingerphotos. The intermediate layers convey features representing the deblurred fingerphotos at lower resolutions. The multi-stage approach allows the model to exploit this information and enhance the deblurring, even at early stages. Hence, the generator is guided more accurately toward estimating the deblurred fingerphoto. To construct a deblurred image from the intermediate feature maps, we use a 1×1 convolution that shrinks the depth information of feature maps to generate a deblurred fingerphoto image. Afterward, we compare each intermediate deblurred image against its down-sampled version of the ground truth fingerphoto. Given a blurred input fingerphoto, $\boldsymbol{x}_b$, the generator, G, reconstructs three deblurred fingerphotos $\boldsymbol{x_r^1}$, $\boldsymbol{x_r^2}$, and $\boldsymbol{x_r^3}$ at resolutions of 64×64, 128×128, and 256×256, respectively. Further, we downsample the original ground truth fingerphoto to 64×64, 128×128, given by $\boldsymbol{x}_s^1$ and $\boldsymbol{x}_s^2$, respectively. The original ground truth $\boldsymbol{x}_s^3$ has the resolution of 256×256 therefore, we use it without any downsampling and apply the L_1 reconstruction loss on all the three of them to directly supervise the generation process. Our modified objective function for training the model is:

$$\begin{aligned} O_{cGAN}(G, D_1, D_2, D_3) = \\ \mathbb{E}_{\boldsymbol{x}_s \sim \mathcal{P}_{data}} \big[\log \big(D_1(\boldsymbol{x}_b^1, \boldsymbol{x}_s^1) \big) \\ + \log \big(D_2(\boldsymbol{x}_b^2, \boldsymbol{x}_s^2) \big) + \log \big(D_3(\boldsymbol{x}_b^3, \boldsymbol{x}_s^3) \big) \big] \\ + \mathbb{E}_{\boldsymbol{x}_s \sim \mathcal{P}_{data}} \big[\log \big(1 - D_1(\boldsymbol{x}_b^1, \boldsymbol{x}_r^1) \big) \\ + \log \big(1 - D_2(\boldsymbol{x}_b^2, \boldsymbol{x}_r^2) \big) \\ + \log \big(1 - D_3(\boldsymbol{x}_b^3, \boldsymbol{x}_r^3) \big) \big], \end{aligned} \tag{5}$$

where, $\{\boldsymbol{x}_b^1, \boldsymbol{x}_b^2, \boldsymbol{x}_b^3\}$ are the blurred input fingerphotos at resolution of 64×64, 128×128, and 256×256, respectively. As given in Eq. 2, the conditional GAN uses the input image as a condition on the discriminator. In our case, these blurred fingerphotos work as the condition on the corresponding discriminator. Therefore each discriminator observes the blurred fingerphoto concatenated with the deblurred version and the blurred fingerphoto concatenated with the ground truth.

For the deblurring task, the training data for the source and target domains are available or can be generated synthetically. Therefore, we use the above model as the core network of our deblurring model. It accepts blurred fingerphotos and maps them to their corresponding deblurred images. Figure 3 shows the deblurred fingerphotos at different resolutions. However, unlike other natural imagery, biometric images contain identity-specific information that needs to be preserved during the deblurring. To this aim, we incorporate a deep fingerphoto verifier module to preserve the identity information [17]. This module also helps improve the deblurring performance through the loss function defined on the identification similarity of the generated and ground truth fingerphotos, which acts as a perceptual similarity loss [84]. The decomposition of the deblurring task using a multi-stage approach along with a verifier network showed promising results on the test datasets. Due to the multi-stage approach in the proposed method, unlike a traditional GAN, we have three L_1 reconstruction loss terms. They compute the distance between the intermediate deblurred fingerphotos and the ground truth fingerphotos such that, during the training, the error from the intermediate layers is minimized. The reconstruction loss is given by:

$$L_{L_1}(\boldsymbol{x}_s, G) = \sum_{i=1}^{3} \lambda_i \|\boldsymbol{x}_s^i - \boldsymbol{x}_r^i\|_1, \tag{6}$$

where i indicates the index of the intermediate layer in the generator used to extract the features and $\boldsymbol{x}_s^i$ is the ground truth at the resolution of the i^{th} layer. The scaling coefficients λ_1, λ_2, and λ_3 are selected using grid search and set to 0.1, 0.4, and 0.6, respectively. Therefore, from Eq. 3, the objective function becomes:

$$\begin{aligned} G_{optimal} = \min_G \max_D O_{cGAN}(G, D_1, D_2, D_3) \\ + L_{L1}(\boldsymbol{x}_s, G). \end{aligned} \tag{7}$$

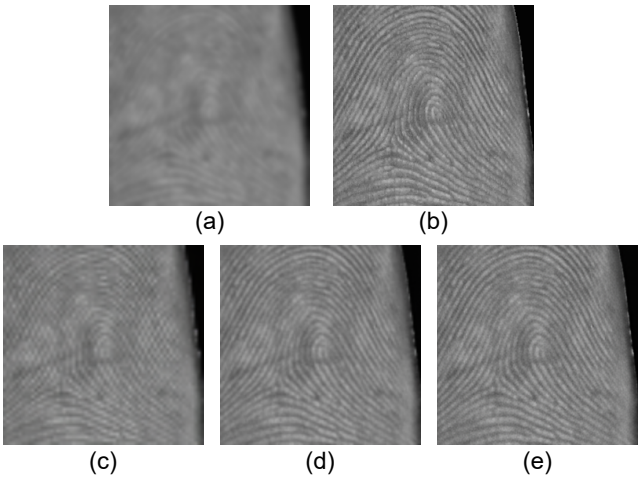

FIGURE 3. Shows the intermediate outputs along with input and ground truth fingerphotos. (a) The blurred input fingerphoto, (b) the ground truth fingerphoto, (c) the first intermediate deblurred output of size 64×64, (d) the second intermediate deblurred output of size 128×128, and (e) the final deblurred output of size 256×256. The low-resolution images are upscaled to 256×256 for better visualization.

C. ID PRESERVING NETWORK

While the multi-stage network deblurs the input fingerphotos using adversarial and reconstruction loss, the underlying ID information, which is crucial for further fingerphoto recognition, may not be preserved. The synthesized fingerphotos from the generator are the transformed representations of the source images; hence, they may look similar to the sources. However, there is no guarantee that the set of minutia points in the source and generated fingerphotos match. To deal with

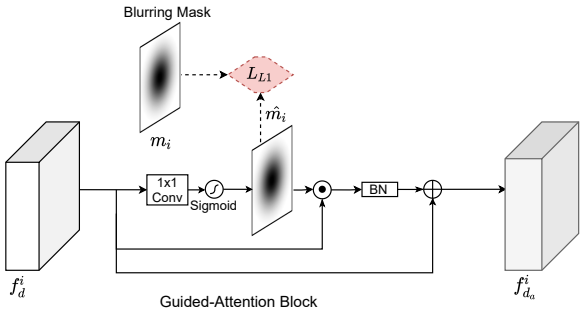

FIGURE 4. Guided-Attention (GA) block. $\hat{\boldsymbol{m}}_{\boldsymbol{i}}$ and $\boldsymbol{f}_{d_a}^i$ are the generated attention map and output feature maps after applying the attention, respectively. Here, dotted lines represent connections that are used only during the training. BN in the diagram represents 2D Batch Normalization.

this problem, we introduce a Siamese Res-Net18 [85] based verifier network, V, which takes the final output, $\boldsymbol{x}_r^3$, of the generator, G, and the ground truth fingerphoto, $\boldsymbol{x}_s^3$. It then returns a feature vector for each image that represents the ID information from each fingerphotos. We compute the Euclidean distance between the two to construct the ID preserving loss as:

$$L_v(V,G) = \|V(\boldsymbol{x}_s^3) - V(\boldsymbol{x}_r^3))\|_2^2, \tag{8}$$

where, $V(\boldsymbol{x}_s^3)$ and $V(\boldsymbol{x}_r^3)$ are the feature vectors for the ground truth fingerphoto, $\boldsymbol{x}_s^3$, and reconstructed fingerphoto, $\boldsymbol{x}_r^3$, respectively. This loss is minimized to make the feature representations as close as possible to one another. With this approach, we observe in Section V-C and Figure 17 that the deblurred fingerphotos have an almost identical set of minutia points as the corresponding sharp fingerphotos.

Even though the verifier loss L_v helps produce better results, we use the perceptual loss as another measure to leverage the verifier network fully. The authors in [84] have shown the merits of using perceptual loss for style transfer and image reconstruction tasks. To achieve a perceptual similarity between the generated and ground truth images, we compare them with the embeddings of a pre-trained model. A lower distance in the embedding of the pre-trained model implies that the two input fingerphotos are more similar to each other. Perceptual loss extracted from intermediate layers of VGG-19 [86] like CNN is widely adapted in image generation models [87], [88], [89]. This approach works well on natural images [5], [15], [90]. However, the major difference between natural images and fingerphotos is that the latter is more like a textured pattern than an object or a scene. Therefore, inspired by [17], we use a perceptual loss based on the ID of the fingerphoto, which is obtained from the same verifier network. Adding the L_2 loss computed on all the intermediate representations of the verifier balances the similarity of higher-level features, such as the coarse ridge formation, and lower-level features, such as pores and minutiae, between the generated fingerphoto and the ground truth fingerphoto. This perceptual loss is computed on three

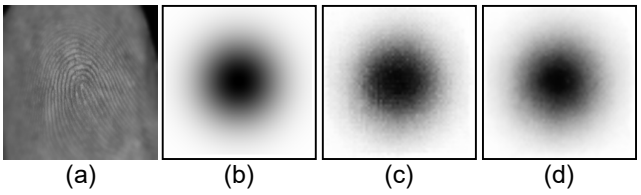

FIGURE 5. A blurred sample generated using a Gaussian blurring kernel and the corresponding attention maps. Figure (a) shows the blurred fingerphoto, (b) is the corresponding partial blurring mask, (c) shows the generated attention map at resolution 64×64, and (d) shows the generated attention map at resolution 128×128. In the blurring mask and attention maps, the white area represents the blurred region where attention is being paid. For the samples where the whole fingerphoto is blurred, the blurring mask is one everywhere.

intermediate layers and is defined as:

$$L_{perc}(V,G) = \sum_{i=1}^{3} \|V(\boldsymbol{x}_s^3)_i - V(\boldsymbol{x}_r^3)_i\|_2, \tag{9}$$

where i is the index of the residual block in the verifier network.

D. GUIDED-ATTENTION MECHANISM

The spatial attention mechanism is known to weigh the elements in the input feature maps such that the important elements are highlighted. In real-world scenarios, the blurring effect is often non-uniform, causing some parts of the finger to be blurred more intensively than other regions, referred to as local blurring. In such cases, some part of the image is still visible, which the network can easily exploit and generate random ridge patterns based on the visible ones. To mitigate this problem, we resort to the attention mechanism. The spatial attention block takes feature maps and produces the attention map, which is then applied to the same features. In this way, the deblurring model attends to the blurred region of the fingerphoto and refrains from emphasizing the visible (easy-to-reconstruct) regions. We use the last two layers of size 64×64 and 128×128 of the generator to add the guided attention (GA) block. The attention block is demonstrated in Figure 4.

Given the input features, $\boldsymbol{f}_d^i$, where i is the index of the corresponding layer, the GA block calculates an attention map, $\hat{\boldsymbol{m}}^i$, using a 1×1 convolutional layer followed by the Sigmoid function. Each attention map is a single-channel tensor with values in the range $[0, 1]$. Figure 5 shows the generated attention maps at different resolutions along with the blurring masks. At each resolution, it multiplies the predicted attention map with the corresponding input feature maps and applies batch normalization (BN). Finally, we add the input feature maps, $\boldsymbol{f}_d^i$, to the attended feature maps to generate the final output features, $\boldsymbol{f}_{d_a}^i$. In the case of fingerphotos that are locally blurred, the sharp regions in the fingerphoto may affect the attention map on the blurred regions. Therefore, during the training, we provide additional guidance to the attention map using a blurring mask, $\boldsymbol{m}^i$, generated while applying the blurring kernel. We force the predicted attention

map, $\hat{\boldsymbol{m}}^i$, to be similar to the ground truth blurring mask, which ensures that the attention is focused on the blurred region. The guided attention generation loss is given by:

$$L_{attn} = \|\hat{\boldsymbol{m}}^i - \boldsymbol{m}^i\|_1. \tag{10}$$

E. MULTI-TASK LEARNING MODEL

During the initial phase of training the network, we observed that the model was not robust to estimate the correct blurring type. It causes problems since if two blurring types are similar, the network may confuse them and generate degraded quality output. Hence, we added another task to explicitly force the model to identify the blurring type during the deblurring process. The multi-task learning (MTL) approach has shown that learning interrelated tasks can help each other and the primary task [91], [75], [74]. Therefore, we incorporate MTL into our network and show that learning relatively simple and small sub-tasks together with a primary task works well. In the entire network, we have a primary task of deblurring, a sub-task of predicting the blurring type, and a virtual task of generating the blurring mask. To implement this, we first branch out the features from the second-to-last layer of the decoder and pass it to a fully connected layer. It then produces two predictions of the blurring type (either Gaussian or motion blurring). Here, we minimize the cross-entropy loss to get an accurate prediction of the blurring type. It is given as follows:

$$L_{ce} = -(y \log(p) + (1-y)\log(1-p)), \tag{11}$$

where, y indicates ground truth blurring type and p indicates predicted blurring type.

As mentioned above, we learned the virtual task of generating a blurring mask. In the attention module described in Section III-D, we use the intermediate features and force them to be like the blurring mask. This reconstructed mask works as the attention which highlights the blurred area of a partially blurred fingerphoto. These two sub-tasks support the deblurring task of generating sharp fingerphotos. Later in Section V-E, we empirically validate that adding the MTL module improves deblurring performance.

F. OBJECTIVE FUNCTION

The total objective function of the generator is the addition of all the cost terms described before. Therefore, the updated cost function from Eq. 7 is given by:

$$\begin{aligned} G_{optimal} = \min_G \max_D & O_{cGAN}(G, D_1, D_2, D_3) \\ & + L_{L_1}(\boldsymbol{x}_s, G) \\ & + \lambda_v L_v(V, G) + \lambda_{perc} L_{perc}(V, G) \\ & + \lambda_{attn} L_{attn} + \lambda_{ce} L_{ce}. \end{aligned} \tag{12}$$

The scaling coefficients λ_v, λ_{perc}, λ_{attn}, λ_{ce}, and the λs used for the reconstruction loss in Eq. 6, were chosen empirically and are set to the best performance values. We evaluate the effectiveness of each loss term in the overall objective function further in Section V.

G. NETWORK ARCHITECTURE

As stated in Section I, the base of the proposed deblurring network is a cGAN model, which deblurs the input image. The generator G is a U-Net [82], [92] based network. The U-Net architecture has skip connections from the encoder to the decoder to preserve the residual information during the training. In the encoder section, we have eight blocks of 4×4 convolutional layers along with batch normalization and ReLU activation. Here, to reduce the spatial size of the image, we use a stride of two instead of a pooling operation in the convolutional layers. The decoder section also has eight blocks having transposed convolutional layers with batch normalization, ReLU activation, and a stride of two to upsample the image. The discriminator network is a three-layer PatchGAN discriminator [82]. It classifies the image as real or fake on every 70×70 patch of the output of the generator and the ground truth fingerphotos from the target domain. It has five 4×4 convolutional layers, each followed by batch normalization and LeakyReLU activation. Out of five, the first three convolutions have a stride of two, and the rest have a stride of 1. Figure 2 shows the details of our proposed architecture.

The ID preserving network *i.e.*, the verifier is a Siamese network [93] trained with contrastive loss [93] using a ResNet-18 [85] architecture. While training the proposed model, we froze the weights of the verifier and used it for feature extraction. This network gets the same input as the last discriminator, D3, in Figure 2 and compares the two fingerphoto samples. During this, we extract the features from conv1, conv2_x, and conv4_x of the ResNet-18 architecture in the verifier and compute the L_2 loss on each. Further, the guided attention block has a 1×1 convolutional layer followed by batch normalization and Sigmoid activation. Lastly, for predicting the blurring type, we use a fully connected layer that accepts 128 features and maps them to two outputs. These 128 features are obtained by applying global average pooling on the penultimate feature maps in the generator model.

IV. EXPERIMENTS

In this section, we elaborate on the training setup, experiments, and ablation studies performed to evaluate the proposed model. First, training details and datasets are discussed in Section IV-A and IV-B, respectively. Then, section IV-C illustrates the evaluation criteria and methods followed by Section IV-D, which goes through a comparison with the state-of-the-art deblurring method (i.e., DeblurGAN-v2). Lastly, in Section IV-E, we present the evaluation of our approach on real-world blurred fingerphotos.

A. TRAINING

We trained the proposed model for 200 epochs with two Nvidia Titan X GPUs. The Adam optimizer [94] is used as the optimizer with an initial learning rate of 2×10^{-4}. The momentum parameters used are: $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and the batch size of 16 was used during the training. The

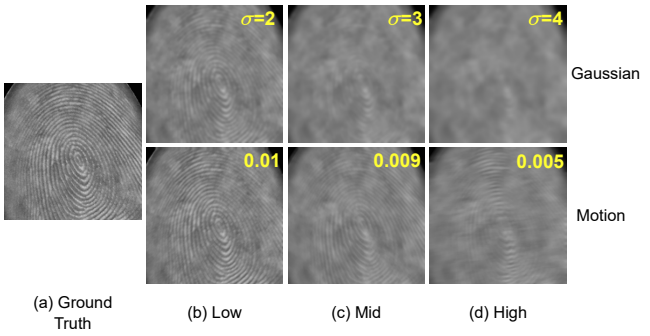

FIGURE 6. Sample blurred input fingerphotos with low to high parameter values for the blurring kernels. Column (a) shows the ground truth fingerphoto. The first row contains a sample blurred using the Gaussian blurring kernel with σ values mentioned at the top right corner of the image. The second row shows the sample blurred using motion blurring kernel using the parameter values shown at the top right corner of each image.

TABLE 1. Mean and standard deviation of the score distributions for the matching experiments on blurred fingerphotos using our deep verifier, the commercial VeriFinger SDK v10.0, and IDKit SDK v8.0 matchers. The distributions belong to genuine, impostor, and combined pairs. The scores from each matcher are normalized between 0 and 1.

Verifier	Combined		Genuine		Impostor	
	Mean	Std	Mean	Std	Mean	Std
Deep Verifier	0.5477	0.2323	0.4584	0.2421	0.6186	0.1931
VeriFinger SDK V10.0	0.024	0.092	0.0601	0.1519	0.0814	0.1856
IDKit SDK v8.0	0.0291	0.0494	0.053	0.0794	0.275	0.1446

PyTorch framework [95] was used for the development and related experimentation of the project.

B. DATASETS

For training and testing, we used the newly collected Multimodal Biometric Dataset[1,2] [96], which consists of 3,851 fingerphoto images belonging to around 600 subjects. This dataset was split into two parts, $3,542$ images were used for training, and the rest 309 was used for testing. The dataset was split so that the subjects in the training and testing set were mutually exclusive. During the initial study, we found that there is no publicly available blurred fingerphoto dataset. Therefore, we created a synthetic dataset using a Gaussian, partial, and motion blur function with arbitrary kernel parameters. The kernel size for the Gaussian blur function is given by: $k = 6\sigma - 1$, where σ values are randomly chosen from a range between 1.5 to 4. We used a non-uniform Gaussian blurred mask to generate partially blurred images and applied it to the fingerphoto to get the desired effect. Further, for generating the motion blur effect, we used the method proposed in [15] with the parameter values between 0.003 to 0.01, where lower values generate complex trajectories. Figure 6 shows examples of the blurred samples.

[1]This data was collected at WVU under IRB # H-24519 and 1310112238 with appropriate human subjects' approval.

[2]Dataset available upon request to the authors.

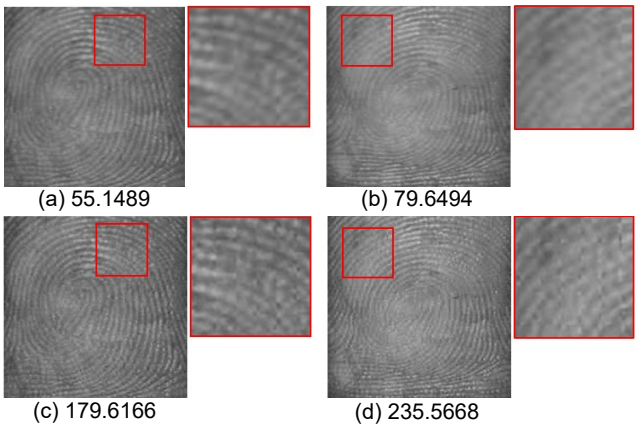

FIGURE 7. Fingerphoto samples from the Blurry set. Images (a) and (b) are blurry fingerphotos, whereas (c) and (d) are the corresponding deblurred fingerphotos. The red boxes in (a), (b), (c), and (d) are enlarged for visual comparison. The numbers at the bottom of each image are the measure of blurriness computed using the variance of Laplacian magnitude [97], where a smaller value means higher blur distortion.

We augmented the data using random blurring kernels, blurring types, and random flipping while training. As a result, the number of training samples totaled approximately $700,000$. For consistency and reproducibility of the test set across different models and training paradigms, we use σ of 3.0 for Gaussian blurring and 0.009 for motion blurring on all the fingerphotos. In addition to this, we create another test set with parameters not in the range of the training dataset. For this, σ of 1.5, 5.0, 7.0 for Gaussian blurring and 0.05, 0.005, 0.0005 for motion blurring was used. While generating blurring kernels, we observed that if a motion blurring kernel has a short trajectory or the magnitude of the σ in a Gaussian kernel is low, the fingerphotos can be matched without deblurring. On the other hand, with a higher magnitude of σ or motion-blurring kernels with complex trajectories, the fingerphotos lose the necessary information and become intractable. Therefore, we empirically selected a valid range for the parameters that resulted in good blurred samples. Figure 8 shows an example of a blurred and sharp ground truth sample from the WVU multimodal dataset [96].

Preparing the dataset for the deblurring network included minimal pre-processing on the original clean ground truth fingerphotos. Initially, the collected data contained noise and misaligned finger positions. Therefore, we first applied skin detection on the images to eliminate such problems from the training data, replaced the background with black color, and grayscaled them. Next, to get the correct finger region of interest (ROI), we performed finger ridge segmentation using Gabor responses [98]. Then, we extracted the finger core using the directional histogram of the fingerphoto [99]. After extracting the core point, we crop a 256x256 image around the core and use that as the input to the proposed model. Later in this section, we discuss the processing steps done on the evaluation datasets. For cross-database evaluation, we use two publicly available datasets, 'Touch and Touchless Finger-

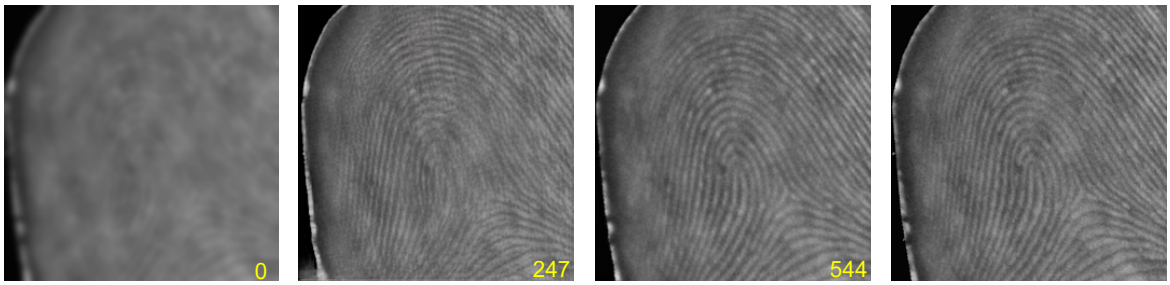

FIGURE 8. A sample from the WVU dataset deblurred using the DeblurGAN-v2 and our method. The number on the bottom right corner represents the matching score between the ground truth and the corresponding fingerphoto using VeriFinger v10.0.

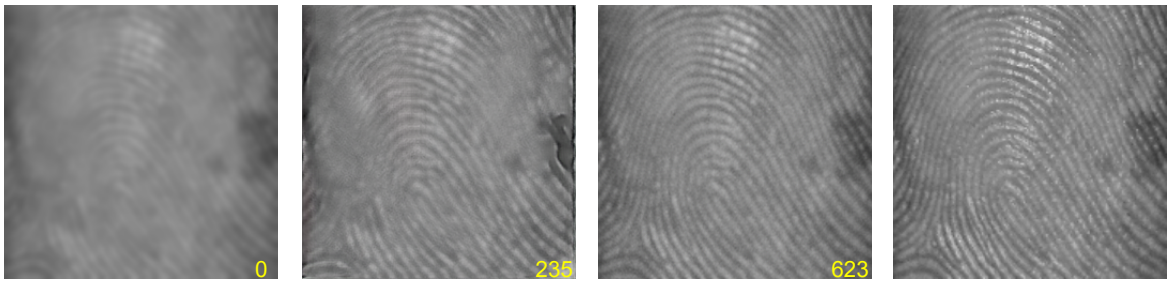

FIGURE 9. A sample from the IIT-B dataset deblurred using the DeblurGAN-v2 and our method. The number on the bottom right corner represents the matching score between the ground truth and the corresponding fingerphoto VeriFinger v10.0.

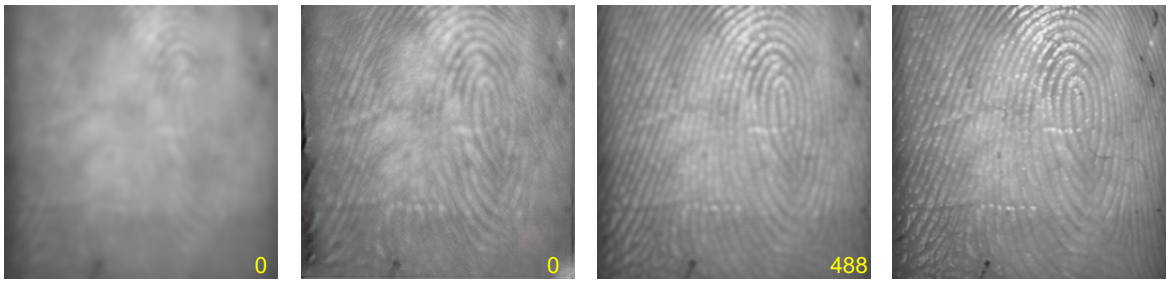

FIGURE 10. A sample from the PolyU dataset deblurred using the DeblurGAN-v2 and our method. The number on the bottom right corner represents the matching score between the ground truth and the corresponding fingerphoto VeriFinger v10.0.

print Dataset' by IIT-Bombay [100] and the 'Contactless Fingerprint Dataset' by the Hong Kong Polytechnical University [9] referred in the paper as IIT-B, PolyU dataset, respectively. The IIT-B dataset has 800 samples from 200 subjects, where all the samples are collected using a smartphone camera. The images in the dataset are segmented, so we skip the skin detection and apply the same preprocessing used for the WVU dataset. After the preprocessing, we obtain 687 samples for evaluation. The remaining samples are discarded since the preprocessing algorithm fails to extract enough information from them. Then we apply the blurring kernels to use for the evaluation task. The PolyU dataset contains 2,976 images from 336 subjects. The images in this dataset are sharper than the other two datasets. Therefore, we only crop and grayscale these images. Lastly, we apply the blurring kernels such that each blur effect among the Gaussian, motion, and partial is equally likely to be applied to the dataset. Then, we randomly select one image from each subject in the gallery for the matching experiment and add the rest of the images to the probe set. Samples from the IIT-B and PolyU datasets are depicted in Figures 9 and 10, respectively.

In a recognition scenario, the system needs to know whether the fingerphoto requires deblurring. This can be achieved by measuring the blurriness in an image. Our emphasis in this study is to analyze fingerphoto deblurring. Therefore, we assume the model receives blurry fingerphotos. Additionally, we evaluated our model on non-blurry fingerphotos to analyze the performance difference. A perfect

TABLE 2. Comparison of performance of the matching experiments on ground truth, blurred, and deblurred fingerphotos using the DeblurGAN-v2 and our method. Evaluations are conducted on all three datasets using three different verifiers.

Verifier	Model	WVU		IIT-B		PolyU	
		EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)
Deep Verifier	Ground Truth	1.6181	99.9131	8.9431	97.2123	14.4118	93.6531
	W/o Deblurring	27.8317	78.8733	20.7317	86.3315	28.5294	77.8380
	DeblurGAN-v2 [5]	15.8576	90.3908	21.1765	84.9239	26.6176	79.8478
	Ours	**5**.8252	**9**8.7704	**12**.0935	**9**4.1447	**18**.3824	**88**.0564
VeriFinger SDK v10.0	Ground Truth	0.6472	99.3580	4.2683	95.8693	5.5882	94.6063
	W/o Deblurring	50.5405	51.1065	49.7810	54.8966	49.7194	55.9611
	DeblurGAN-v2 [5]	17.7994	84.3356	10.9756	89.7495	12.0425	89.4693
	Ours	**1**.9417	**9**8.0622	**5**.8130	**9**4.6539	**6**.6176	**93**.8511
IDKit SDK v8.0	Ground Truth	0.0000	100.00	0.8130	99.9596	0.2941	99.8518
	W/o Deblurring	31.6641	75.3145	26.5032	81.8691	24.0397	83.8753
	DeblurGAN-v2 [5]	6.7422	98.3931	4.2683	99.0298	7.5597	97.1537
	Ours	**1**.2945	**99**.9576	**2**.3882	**99**.5255	**3**.2958	**99**.2087

AUC from the commercial verifiers in Table 5 suggests that the model does not affect the fingerphotos if they are non-blurry. For evaluating the model on real-world blurry data, we experimented on the WVU and IIT-B test sets since there is no readily available dataset. Despite high-quality fingerphotos in these datasets, we observed that they also contain blurry samples. Hence, through manual inspection, we gathered the blurry samples and deblurred them using our model. The resulting dataset contains 450 blurry samples and is referred to as the Blurry Set in this paper[3]. Figure 7 shows some blurry samples and the deblurred counterparts.

C. EVALUATION

To evaluate the performance, matching experiments were conducted between the ground truth, blurred, and deblurred fingerphotos using three different verifiers. One is the deep verifier mentioned in Sec. III-G, and the other two are the commercial VeriFinger SDK v10.0 [101], Innovatrics IDKit SDK v8.0 [102]. We fixed the genuine and impostor pairs for fair comparisons for all three datasets. For generating the pairs, we used sharp ground truth fingerphotos as the gallery and, depending on the experiment switched the probe set between sharp, blurry, and deblurred fingerphotos. Figure 11 presents the results for these evaluations. We can see that the matching performance on deblurred fingerphotos using the proposed model closely follows the curve of the clean ground truth fingerphotos. Note that the VeriFinger is a commercial matching software and is highly sensitive to incorrect minutiae to limit the false acceptance rate. As a result, a few wrong minutiae cause the score to drop significantly, which causes the sharp peak in low scores making the ROC less smooth compared to the curves from the other two verifiers. The mean and standard deviation of the score distribution for all three matchers are provided in Table 1.

Considering the matching experiment on blurred fingerphotos as a baseline, our model achieves a significant performance boost on all three test datasets in the deblurring task, suggesting that the reconstruction quality of the deblurred fingerphoto is improved. Table 2 shows the results of the verification experiment performed on the blurred data and the same experiment performed on the deblurred data. The lower EER demonstrates the ability of the proposed model to deblur the fingerphotos and preserve the necessary identity information. Furthermore, we performed the matching experiment on deblurred samples at different resolutions to highlight the progressive improvement of the result in our model. Table 3 presents the results for these evaluations. It can be observed that the matching performance is lower at low resolutions and boosts as the resolution increases. Deblurred visual samples from the WVU, IIT-B, and PolyU datasets are shown in Figures 8, 9, and 10, respectively. In addition, we compared the two methods in terms of another image similarity/quality metric, and the results are provided in Table 4.

TABLE 3. Comparison of performance of matching experiment at different scales of the deblurred fingerphotos.

Scale	EER(%)	AUC(%)
64×64	33.3333	72.8066
128×128	9.3851	96.6575
256×256	**5**.8252	**9**8.7704

We evaluated our model on fingerphotos that are blurred with unseen kernel parameters. The results of the matching experiments using three different matchers are provided in Table 5. As expected, matching performance on blurred and deblurred fingerphotos of unseen blur levels is slightly degraded. It is due to the more severe nature of the parameters used to create the blurring kernels than the kernels used for the training set. As a result, deblurring these fingerphotos is challenging for the model. Nonetheless, a high AUC by the three matchers demonstrates the generalizability of the network on unseen blurring kernels.

Furthermore, we calculate the quality scores of the images using the NIST NFIQ2 tool provided by NBIS software [103]. The NFIQ2 tool provides quality scores from 0 to 100

[3]The list of indexes will be provided upon request to the authors.

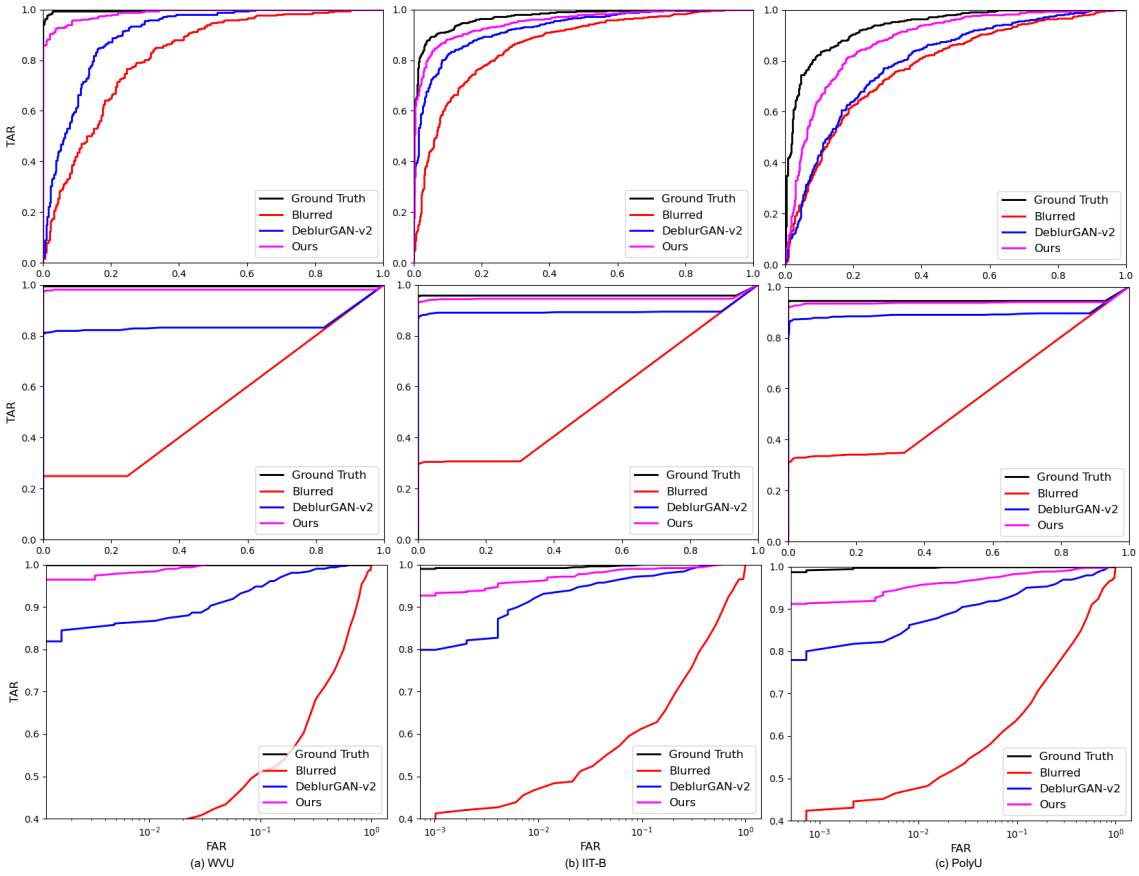

FIGURE 11. Evaluating the matching performance on different datasets and models. The curves in the first, second, and third rows are generated using the deep verifier, the VeriFinger SDK v10.0, and the IDKit SDK v8.0, respectively. Column (a) represents curves for the WVU dataset, column (b) represents curves for the IIT-B dataset, and column (c) represents curves for the PolyU dataset. Our method, represented by the magenta color, outperforms the DeblurGAN-v2 (blue curve). Detailed Area Under the Curve (AUC) and Equal Error Rate (EER) values are given in Table 2. The curves in the last row are log-scaled for better visualization.

TABLE 4. Comparison of performance based on the Structural Similarity Index (SSIM) and Peak-Signal-to-Noise-Ratio (PSNR) image quality metrics between ground truth and deblurred fingerphotos using the DeblurGAN-v2 and our method.

Model	WVU		IIT-B		PolyU	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
DeblurGAN-v2 [5]	0.8646	26.2646	0.7993	25.4103	0.8462	26.5482
Ours	**0**.9474	**30**.4601	**0**.9205	**30**.3644	**0**.9653	**33**.9360

for the quality of the fingerprints, where 0 means no utility value, and 100 is considered the highest utility value. Our model preserves and, in some cases, improves the quality of the ground truth fingerphotos during the deblurring process. The quality score plots for the three test sets are shown in Figure 12.

D. COMPARISON WITH STATE OF THE ARTS

Since there is no dedicated deblurring algorithm for fingerphotos in the literature, we reviewed several natural image deblurring methods. We tested and compared our method against a state-of-the-art deblurring algorithm i.e., DeblurGAN-v2 [5]. We used the trained weights provided by the authors and fine-tuned the DeblurGAN-v2 network on the same training set we used to train our algorithm. Table 2 and Figure 11 presents the comparison results on the WVU multimodal, IIT-B, and PolyU datasets. As expected, our model performs better in terms of ID preservation due to the verifier network that penalizes ID disparities between the generated and the ground truth fingerphoto. However, at the same time, from Table 7 (performance of MCGAN-A (see Section V)) and Table 2 (performance of DeblurGAN-v2 on WVU dataset), it can be observed, that our method, regardless of the verifier, performs better than the DeblurGAN-v2. It shows that even though state-of-the-art deblurring techniques work effectively on blurred natural images, they lack the ID preservation essential in biometric data.

E. EVALUATION ON REAL-WORLD DATA

In order to make a fair evaluation, it is imperative to verify the model performance on real-world blurred fingerphotos. Therefore, as mentioned in Section IV-B, we used the Blurry set for the evaluation. However, since no sharp ground truth is available for all the blurry fingerphotos, we used the corresponding contact-based fingerprint samples as the ground

TABLE 5. Performance evaluation on fingerphotos blurred using multiple Gaussian and motion kernels.

Verifier	Model	No Blur		Gaussian(1.5)		Gaussian(5)		Gaussian(7)		Motion(0.05)		Motion(0.005)		Motion(0.0005)	
		EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)	EER(%)	AUC(%)
Deep Verifier	W/o Deblurring	-	-	9.7087	96.8795	46.9256	53.8887	45.6311	55.0466	8.0906	97.0224	18.123	89.0476	28.1553	79.056
	DeblurGAN-v2 [5]	3.8835	98.9972	4.8544	98.9092	44.3366	57.8125	49.3528	52.0899	5.9871	98.3986	10.5178	95.8531	16.5049	90.0425
	Ours	3.5599	99.269	**2**.1036	**99**.7502	**26**.8608	**80**.9753	**36**.8932	**68**.7718	**3**.5599	**99**.4538	**6**.1489	**98**.7039	**9**.3851	**96**.4522
VeriFinger SDK v10.0	W/o Deblurring	-	-	1.4887	99.9102	49.7018	50.5498	50.0412	49.9162	1.6181	99.7568	11.4563	94.8008	28.073	81.361
	DeblurGAN-v2 [5]	0.0	100.0	0.0	100.0	47.4434	53.8432	49.1027	52.147	0.0	100.0	0.6472	99.983	9.2899	97.6487
	Ours	0.0	100.0	0.0	100.0	**21**.6154	**86**.3575	**36**.6482	**67**.3283	0.0	100.0	0.3236	99.9955	1.8608	99.6471
IDKit SDK v8.0	W/o Deblurring	-	-	4.8544	99.093	49.2294	51.76	49.5146	51.2929	3.5599	99.4062	17.8242	89.9967	32.0129	76.8729
	DeblurGAN-v2 [5]	0.0	100.0	0.3236	99.9976	43.7198	58.9761	47.1133	54.4087	0.2157	99.9992	1.6181	99.74	11.4347	94.9995
	Ours	0.0	100.0	**0**.0	**100**.0	**28**.3647	**97**.022	**39**.5581	**64**.8818	**0**.1079	**99**.9997	**1**.8031	**99**.8576	**6**.5085	**97**.5267

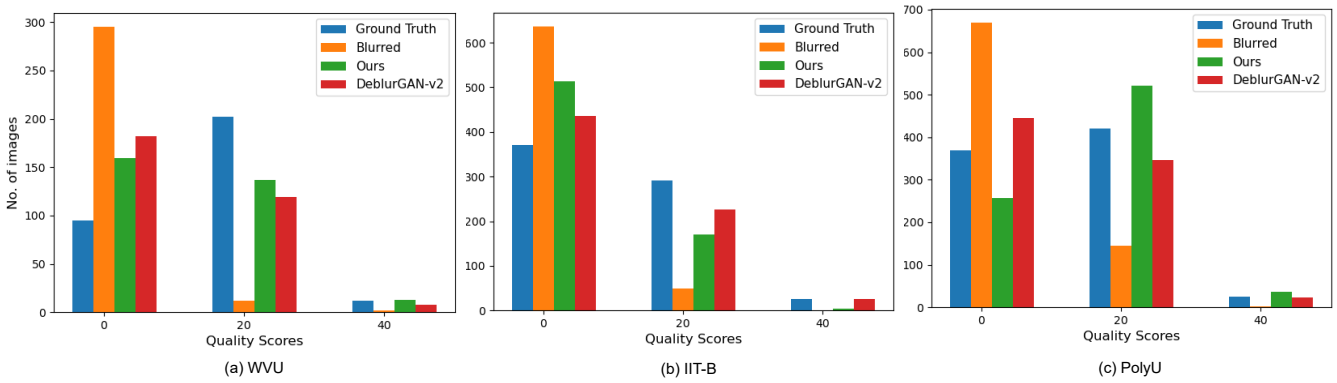

FIGURE 12. NFIQ2 quality score assessment of the ground truth, blurred, and deblurred fingerphotos from the DeblurGAN-v2 [5] and ours. Plots (a), (b), and (c) show the quality scores of samples from the WVU Multimodal dataset, IIT-B dataset, and PolyU dataset, respectively. Blue, orange, and green bars represent the ground truth fingerphotos, blurred fingerphotos, and deblurred fingerphotos using our method, respectively. The red bar represents the deblurred fingerphotos from the DeblurGAN-v2.

TABLE 6. Comparison of performance of the matching experiments on contact-based fingerprints against the blurry and deblurred fingerphotos from the Blurry set. Evaluations are conducted using three different verifiers.

Verifier	Model	EER(%)	AUC(%)
Deep Verifier	Blurry	46.1538	54.3305
	DeblurGAN-v2 [5]	31.375	74.2958
	Ours	**21**.0983	**85**.9346
VeriFinger SDK v10.0	Blurry	36.6917	69.3978
	DeblurGAN-v2 [5]	17.2674	89.8046
	Ours	**13**.5319	**95**.0257
IDKit SDK v8.0	Blurry	42.6805	62.4648
	DeblurGAN-v2 [5]	30.441	77.6312
	Ours	**21**.4262	**86**.5103

truth for the matching experiment. Using the samples from the Blurry set, we created 900 genuine and impostor pairs where the gallery set consists of fingerprints. Depending on the experiment, the probe set is blurry or deblurred fingerphotos. Figure 13 and Table 6 provide the performance of the matching experiment using three verifiers. From the curves in the figure and values in Table 6, it is evident that the model is able to deblur the real-world fingerphotos. Some blurry and deblurred samples are provided in Figure 7. We also calculated the amount of blurriness using the variance of Laplacian magnitude [97] before and after deblurring the fingerphoto. The values state that the amount of blur in the blurry image significantly reduces after deblurring.

To visualize the effects of deblurring on minutiae, we collected 12 samples from three subjects. This smaller dataset is referred to as the Lab collected dataset in the paper. During the collection, we fixed the finger position on a surface and adjusted the camera to get sharp and multiple blurry fingerphotos. As the purpose of the collection is only visualization, we do not consider factors such as scaling, shifting, rotation, *etc*. After acquiring the fingerphotos, we manually cropped the ROI and grayscaled the images to feed them to the network. For illustration, we extracted the minutiae of the fingerphotos using the commercial VeriFinger SDK v10.0 [101] and conducted a matching experiment among the ground truth, blurred, and deblurred fingerphotos. Figure 14 shows the samples and the matched minutiae.

A graph is plotted on images where the vertices are the matched minutiae between the gallery and probe. Noticeably, the minutiae points in the blurred images hardly match those in the ground truths due to the distortion. The verifier fails to obtain any minutia points from the blurry fingerphoto in the left-most sample in Figure 14. However, the deblurring process recovers approximately 60% of the minutiae suggesting that the deblurring method is recovering information in such detail that it improves the matching performance. The low matching scores between the ground truth and the blurred fingerphoto confirm the degraded quality of the blurred fingerphoto. On the other hand, the matcher produces high score values when the ground truth is compared with the deblurred

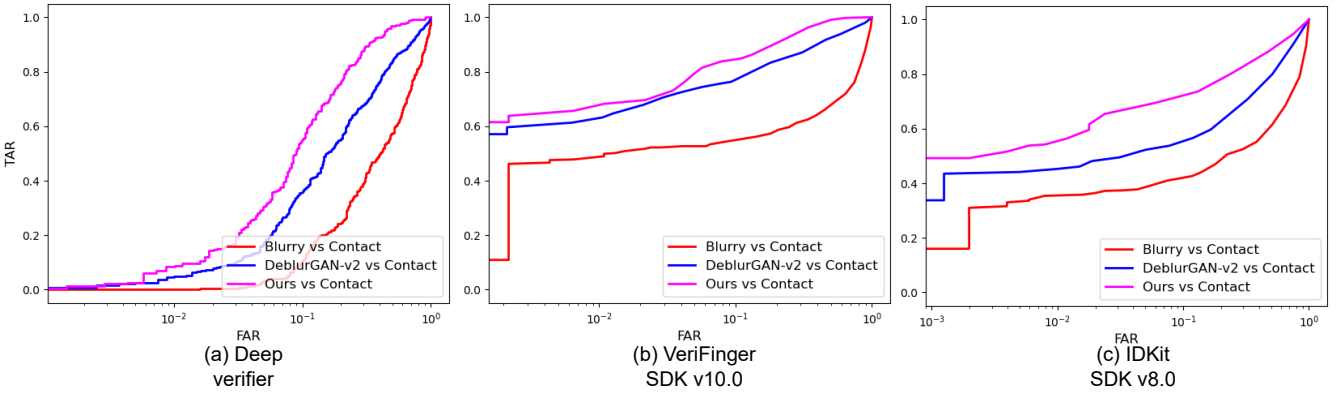

FIGURE 13. Evaluating the matching performance on real blurry, deblurred, and corresponding contact-based fingerprint samples from the Blurry set. Here, curves in the (a), (b), and (c) plots are generated using the deep verifier, the VeriFinger SDK v10.0, and the IDKit SDK v8.0, respectively. Our method, represented by magenta color, outperforms DeblurGAN-v2 [5] and helps improve the sample quality for verification. Detailed Area Under the Curve (AUC) and Equal Error Rate (EER) values are given in Table 6. The plots are log-scaled for better visualization.

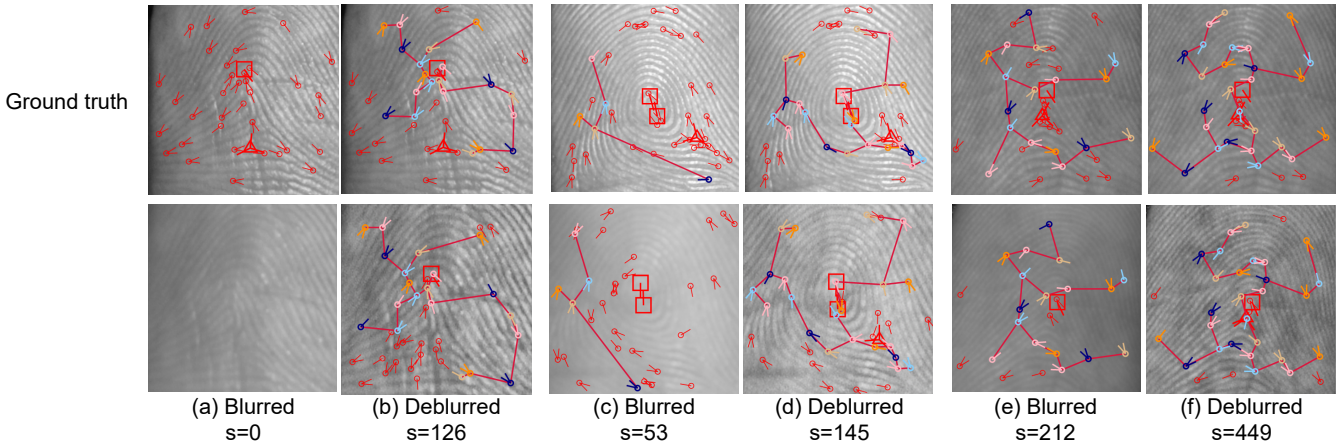

FIGURE 14. Blurred fingerphotos from the Lab collected set. Columns (a, c, e) represent the matching minutiae between the ground truth fingerphotos and blurry fingerphotos, whereas columns (b, d, f) represent the matching minutiae between the ground truth and deblurred fingerphotos. The matching score associated with each pair is represented by s. The matched minutiae on the ground truth and corresponding probe (blurred/ deblurred) are represented as a graph where the matched minutiae are the vertices of the graph. Additional details are provided in Section IV-E.

fingerphoto. These experiments validate the robustness of the model towards complex and random blurring kernels in a real-world scenario. It also shows the potential that, given a real-world blurred dataset, we can fine-tune the model on a subset and achieve state-of-the-art deblurring performance.

V. ABLATION STUDY

In this section, we explore the contribution of each modification in generator G towards the deblurring and matching performance. Here, the baseline model is a plain cGAN [82], and we make one modification at a time and evaluate the resulting model. First, we replace the cGAN architecture with a multi-stage cGAN framework (McGAN). Then, we add the attention mechanism to the McGAN model (McGAN-A). Next, we add the verifier network to the McGAN-A model (i.e., McGAN-AV), and finally, we add the classifier to the previous network to get the proposed model. In the following sections, we extensively discuss the contribution of each part of the model to the final performance.

A. DEBLURRING WITH PLAIN MODEL

First, to determine the performance of the cGAN, we trained the model without any additional constraints. With the advent of image-to-image translation [82], the deblurring task is primitive for the network, but preserving the necessary information is challenging. Table 7 shows the results of the plain cGAN model. We use this performance as a baseline for further experiments. Figure 15 (c) shows a deblurred sample from the cGAN model. As illustrated, the plain cGAN model can produce deblurred fingerphotos with good visual quality. However, the generated fingerphotos have erroneous ridge patterns and minutia points due to the lack of a proper constraint to preserve the ID. Such a shortcoming often alters the identity of the fingerphoto, deteriorating the recognition performance. Based on this observation, we add the multi-stage deblurring scheme to the plain cGAN network to force

(a) Blurred (b) Ground truth (c) Plain cGAN (d) McGAN (e) McGAN-A (f) McGAN-AV (g) Ours

FIGURE 15. Visualizing the impact of each module in the proposed model on the quality of the deblurred samples. (a) Input blurred fingerphoto. (b) Corresponding ground truth fingerphoto. The rest of the images show the deblurred output of (c) plain cGAN model, (d) multi-stage plain cGAN (McGAN), (e) McGAN with guided attention (McGAN-A), (f) McGAN-A with verifier (McGAN-AV), (g) McGAN-AV with multi-tasking (Ours). A more detailed discussion of the models is provided in Section V.

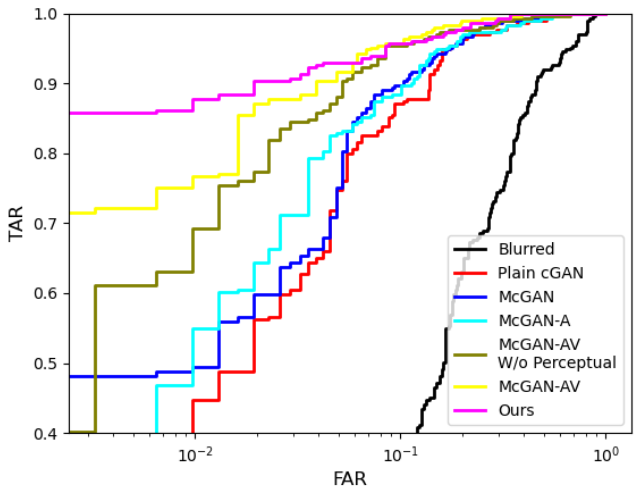

FIGURE 16. The log-scaled ROC curves of different models evaluated during the ablation study. Here, TAR and FAR are the True Acceptance Rate and False Acceptance Rate, respectively. The AUC and EER details are provided in Table 7.

TABLE 7. Results of each modified cGAN model tested for the ablation study. Additional details about each model are given in Section V.

Model	EER(%)	AUC(%)
Blurred	27.8317	78.8733
Plain cGAN	12.2977	95.3122
McGAN	10.3560	96.2673
McGAN-A	10.3560	96.3165
McGAN-AV W/o Perceptual	7.4434	97.7744
McGAN-AV	6.1489	98.5683
Ours	**5**.8252	**98**.7704

the generator to consider the ID of the fingerphotos during the generation process.

B. EFFECT OF MULTI-STAGE APPROACH

As discussed in Section III-B, we developed a coarse-to-fine multi-stage scheme to enhance deblurring performance. Our results in Table 7 and Figure 3 demonstrate that guiding the generator by forcing the intermediate features to mimic the coarse structure of the ground truth images indeed improves deblurring performance. Particularly, based on Table 7, employing multiple discriminators improves EER and AUC by 2% and 1%, respectively. As the results in Table 3 indicate, having more than three stages would degrade the performance. Fingerphotos at 32×32 resolution may not have enough useful low-level information to help a multi-stage scheme. On the other hand, the matching performance at 64×64 resolution has a moderate AUC of 72.81, so excluding this stage might worsen the final deblurring performance. Therefore, we incorporate the multi-stage scheme with three stages for the best deblurring performance.

C. EVALUATION WITHOUT VERIFIER

Further, we tested the proposed model without the verifier network and removed the L_v and L_{perc} loss terms from the overall network objective function in Eq. 12. In the case of fingerphoto deblurring, the best way to evaluate the model performance is by monitoring the identification performance. Figure 16 and Table 7 show the degraded performance without the verifier. The verifier network is one of the constraints on the generator, which restricts it from deblurring the images while keeping the minutiae points and the ridge patterns similar. The verifier plays a vital role in the training. It enforces the generator to preserve the ID information from the blurred fingerphoto. To support this claim, we extracted the minutiae points from the deblurred fingerphoto generated without including a verifier using the commercial VeriFinger SDK v10.0 [101] and compared them against the minutiae of the ground truth and deblurred fingerphoto from our proposed model. Figure 17 shows the comparison. The matching minutiae between the ground truth and deblurred fingerphoto increases after adding the verifier. As the verification cost term is added to the final objective function in Eq. 12, it resulted in a 2.5% increase in the accuracy of the identification task.

Moreover, we tested the network by excluding the loss of the intermediate features of the verifier, which we treat as a perceptual loss. As shown in Table 7, due to the generalized characteristics in the intermediate features, the network performs better in verifying the fingerphotos. According to Figure 16, the accuracy in the lower FAR region is low without the intermediate feature loss. However, adding the loss term improves the accuracy, specifically in the lower FAR region, which is the critical performance region in real-

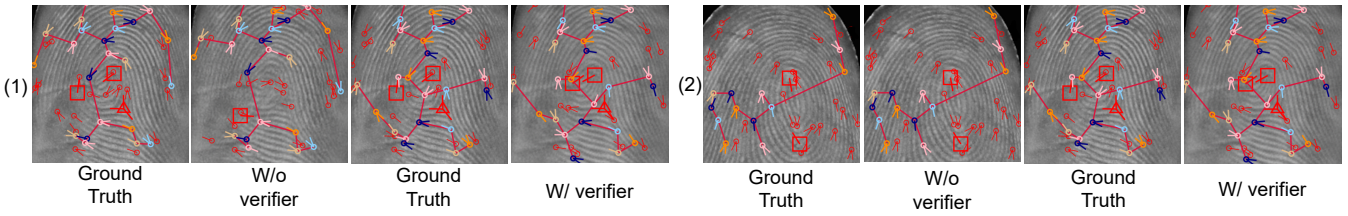

FIGURE 17. Visual comparison of matched minutiae between the ground truth and the deblurred fingerphoto by including and excluding the verifier. Matched minutiae between each pair are displayed using a graph of red edges. The vertices of the graph are matched minutiae in the probe.

world applications.

D. TRADITIONAL PERCEPTUAL LOSS VS. ID-BASED PERCEPTUAL LOSS

In this section, we explore the effectiveness of conventional perceptual loss against ID-based perceptual loss. The goal here is to generate fingerphotos that are perceptually similar to the ground truth sharp fingerphotos. To this aim, a perceptual loss is introduced to the overall loss function. Conventionally, a pre-trained network VGG-19 [86] network trained on the ImageNet dataset is used to extract general features from the intermediate convolutional layers. To compare the extracted features, MSE is calculated and added to the total loss in Eq. 12 to improve the perceptual similarity. However, the VGG-19 network is trained on natural images, while the patterns in fingerphotos are constructed by the ridges and valleys, mainly forming random textures with specific characteristics. Hence, there is a discrepancy between the distribution of features in the two domains. Thus, we observed that, in this situation, using the conventional perceptual loss is not helpful, and a pre-trained network trained on fingerphotos is a prudent choice. Hence, we extract features from our deep verifier, discussed in Section III-C, and use them to compute the alternative perceptual loss. In Table 8, we show the effect of using conventional perceptual loss vs. the ID preserving perceptual loss. These results confirm that in fingerphoto reconstruction tasks, using a model trained on similar domain results in a better performance than the classical method for computing the perceptual loss.

E. BALANCING CONSTRAINTS ON DEBLURRING

In the multi-constrained setting of the proposed network architecture, each constraint has a unique role, and therefore, they need to be balanced optimally to get better deblurring performance. Here, we emphasize the usefulness of other constraints, such as guided attention and multi-task learning. Adding the attention block to the network does not make much difference in terms of the AUC; however, it enhances the reconstruction capability of the network. In the case of a partially blurred fingerphoto, the network may copy the sharp region and neglect the blurred part during deblurring. However, due to the attention block, it focuses on the blurred region and neglects the sharp region in a fingerphoto. Subtle visual improvements can be observed in Figures 15 (d) and (e).

TABLE 8. Matching experiment performance on deblurred fingerphotos using Perceptual and ID preserving Perceptual loss term.

Model	EER(%)	AUC(%)
W/o Perceptual Loss	7.4434	97.7744
W Conventional Perceptual Loss	6.7961	98.1190
W ID Preserving Perceptual Loss	**5**.8252	**98**.7704

Additionally, we removed the blurring type classification module from the network to evaluate the usefulness of multi-task learning in the proposed network and checked the performance. Figures 15 (f) and (g) show the deblurred fingerphoto without and with the classifier module. During deblurring, the generator tries to estimate the blurring kernel and deblurs the input fingerphoto. Without the classification module, the network confuses between the Gaussian and motion blurring kernel affecting the matching performance. On the other hand, when the classifier is included, it estimates the blurring type and helps the generator to produce a better fingerphoto regardless of the blurring kernel.

VI. CONCLUSION

In this paper, we propose a novel method for deblurring fingerphotos by introducing several modifications to the conditional generative adversarial networks inspired by prior knowledge of the fingerprint recognition task. First, we develop a coarse-to-fine scheme using a multi-stage framework that deblurs the fingerphoto at a different resolution, enhancing the overall deblurring performance of the model. Second, we develop an ID preserving network to ensure that the reconstructed image has not lost the unique ID information, *i.e.*, the formation of minutiae in the fingerphoto. Third, we developed a novel attention block that helps the generator to focus more on the blurred region of the fingerphoto enabling the network to deblur partially blurred fingerphotos effectively. Additionally, the multi-task learning approach we used in the generator of the cGAN is a non-traditional method that estimates the blurring type to improve the accuracy of the fingerphoto deblurring task. We demonstrate the effectiveness of each modification through evaluation experiments and extensive ablation studies. Further, our comparison with state-of-the-art indicates the superiority of our deblurring method, developed specifically for fingerphotos.

Our work considers global, local, uniform, and non-

uniform blurring effects along with motion and Gaussian blurring types. Evaluation of the proposed deblurring model on naturally blurred fingerphotos shows a significant recovery of minutiae points. Moreover, consistently better performance using multiple matchers for matching experiments exhibits the deblurring ability of the model on not only synthetic but also real-world blurry fingerphotos. The NFIQ2.0 quality assessment also shows improvement in the quality of the fingerphotos. The proposed model is a novel approach with numerous real-world applications in crime scene investigation, border control, and forensic sciences.

REFERENCES

[1] K. Sundararajan and D. L. Woodard, "Deep learning for biometrics: A survey," *ACM Comput. Surv.*, vol. 51, no. 3, may 2018. [Online]. Available: https://doi.org/10.1145/3190618

[2] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3809–3818.

[3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.

[4] Y. Zhu, S. C. Dass, and A. K. Jain, "Statistical models for assessing the individuality of fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 391–401, 2007.

[5] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 8877–8886.

[6] C. Kauba, D. Söllinger, S. Kirchgasser, A. Weissenfeld, G. Fernández Domínguez, B. Strobl, and A. Uhl, "Towards using police officers' business smartphones for contactless fingerprint acquisition and enabling fingerprint comparison against contact-based datasets," *Sensors*, vol. 21, no. 7, p. 2248, 2021.

[7] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, and C. Busch, "Mobile contactless fingerprint recognition: implementation, performance and usability aspects," *Sensors*, vol. 22, no. 3, p. 792, 2022.

[8] J. Priesnitz, C. Rathgeb, N. Buchmann, C. Busch, and M. Margraf, "An overview of touchless 2D fingerprint recognition," *EURASIP Journal on Image and Video Processing*, vol. 2021, no. 1, pp. 1–28, 2021.

[9] C. Lin and A. Kumar, "Matching contactless and contact-based conventional fingerprint images for biometrics identification," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2008–2021, 2018.

[10] S. Prabhakar, A. Ivanisov, and A. Jain, "Biometric recognition: Sensor characteristics and image quality," *IEEE Instrumentation Measurement Magazine*, vol. 14, no. 3, pp. 10–16, 2011.

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *ArXiv*, vol. abs/1406.2661, 2014.

[12] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[13] N. S. Cho, C. S. Kim, C. Park, and K. R. Park, "GAN-Based blur restoration for finger wrinkle biometrics system," *IEEE Access*, vol. 8, pp. 49 857–49 872, 2020.

[14] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic GAN for non-uniform blind deblurring," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1546–1550, 2019.

[15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," *ArXiv e-prints*, 2017.

[16] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2737–2746.

[17] A. Dabouei, S. soleymani, H. Kazemi, S. M. Iranmanesh, J. Dawson, and N. M. Nasrabadi, "ID preserving generative adversarial network for partial latent fingerprint reconstruction," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–10.

[18] H. Walhazi, A. Maalej, and N. E. B. Amara, "Mask2LFP: Mask-constrained adversarial latent fingerprint synthesis," in *2020 International Conference on Cyberworlds (CW)*, 2020, pp. 265–271.

[19] M. A. Souibgui and Y. Kessentini, "DE-GAN: A conditional generative adversarial network for document enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[20] P. Sharma, P. Jain, and A. Sur, "Scale-aware conditional generative adversarial network for image dehazing," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 2355–2365.

[21] H. Zhang, V. Sindagi, and V. M. Patel, "Image De-raining using a conditional generative adversarial network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3943–3956, 2019.

[22] J. Gu, Y. Shen, and B. Zhou, "Image processing using multi-code GAN prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3012–3021.

[23] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *European Conference on Computer Vision (ECCV)*. Springer, 2010, pp. 157–170.

[24] J. Pan, R. Liu, Z. Su, and X. Gu, "Kernel estimation from salient structure for robust motion deblurring," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1156–1170, 2013.

[25] A. S. Joshi, A. Dabouei, J. Dawson, and N. M. Nasrabadi, "FDeblurGAN: Fingerprint deblurring using generative adversarial network," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, 2021, pp. 1–8.

[26] L. Song and E. Y. Lam, "MBD-GAN: Model-based image deblurring with a generative adversarial network," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 7306–7313.

[27] A. K. Jain, J. Feng, and K. Nandakumar, "Fingerprint matching," *Computer*, vol. 43, no. 2, pp. 36–44, 2010.

[28] A. Attrish, N. Bharat, V. Anand, and V. Kanhangad, "A contactless fingerprint recognition system," 2021.

[29] K. Tiwari and P. Gupta, "A touch-less fingerphoto recognition system for mobile hand-held devices," in *2015 International Conference on Biometrics (ICB)*, 2015, pp. 151–156.

[30] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh, "On smartphone camera based fingerphoto authentication," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2015, pp. 1–7.

[31] S. A. Grosz, J. J. Engelsma, E. Liu, and A. K. Jain, "C2CL: Contact to contactless fingerprint matching," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2021.

[32] B. Y. Hiew, A. B. Teoh, and Y.-H. Pang, "Touch-less fingerprint recognition system," in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 2007, pp. 24–29.

[33] A. Kumar and C. Kwong, "Towards contactless, low-cost and accurate 3d fingerprint identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3438–3443.

[34] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Touchless fingerprint biometrics: a survey on 2D and 3D technologies," *Journal of Internet Technology*, vol. 15, no. 3, pp. 325–332, 2014.

[35] C. Stein, C. Nickel, and C. Busch, "Fingerphoto recognition with smartphone cameras," *Biometrics Special Interest Group (BIOSIG)*, pp. 1–12, 01 2012.

[36] C. Lee, S. Lee, J. Kim, and S.-J. Kim, "Preprocessing of a fingerprint image captured with a mobile camera," in *Advances in Biometrics*, D. Zhang and A. K. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 348–355.

[37] B. Hiew, A. Teoh, and D. Ngo, "Automatic digital camera based fingerprint image preprocessing," in *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)*, 2006, pp. 182–189.

[38] X. Liu, M. Pedersen, C. Charrier, F. A. Cheikh, and P. Bours, "An improved 3-step contactless fingerprint image enhancement approach for minutiae detection," in *2016 6th European Workshop on Visual Information Processing (EUVIP)*, 2016, pp. 1–6.

[39] H. Ravi and S. K. Sivanath, "A novel method for touch-less finger print authentication," in *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, 2013, pp. 147–153.

[40] X. Yin, J. Hu, and J. Xu, "Contactless fingerprint enhancement via intrinsic image decomposition and guided image filtering," in *2016 IEEE*

11th Conference on Industrial Electronics and Applications (ICIEA), 2016, pp. 144–149.

[41] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Contactless fingerprint recognition: A neural approach for perspective and rotation effects reduction," in *2013 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, 2013, pp. 22–30.

[42] R. Raghavendra, C. Busch, and B. Yang, "Scaling-robust fingerprint verification with smartphone camera in real-life scenarios," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.

[43] K. Tiwari and P. Gupta, "A touch-less fingerphoto recognition system for mobile hand-held devices," *2015 International Conference on Biometrics (ICB)*, pp. 151–156, 2015.

[44] A. Dabouei, S. Soleymani, J. Dawson, and N. M. Nasrabadi, "Deep contactless fingerprint unwarping," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–8.

[45] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8260–8269.

[46] F. Alaoui, A. Ghlaifan, D. Vamara, and N. Abdelkarim, "Application of blind deblurring algorithm for face biometric," *International Journal of Computer Applications*, vol. 105, p. 0975 – 8887, 11 2014.

[47] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1628–1636.

[48] H. Zhang, D. Wipf, and Y. Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1051–1058.

[49] U. Schmidt, K. Schelten, and S. Roth, "Bayesian deblurring with integrated noise estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 2625–2632.

[50] U. Schmidt, C. Rother, S. Nowozin, J. Jancsary, and S. Roth, "Discriminative non-blind deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 604–611.

[51] S. Vasu, V. R. Maligireddy, and A. N. Rajagopalan, "Non-blind deblurring: Handling kernel uncertainty with cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[52] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[53] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 821–14 831.

[54] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 4641–4650.

[55] X. Huang, Liu Ren, and R. Yang, "Image deblurring for less intrusive iris capture," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1558–1565.

[56] J. Liu, Z. Sun, and T. Tan, "A novel image deblurring method to improve iris recognition accuracy," in *2011 IEEE International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–8.

[57] M. Jin, M. Hirsch, and P. Favaro, "Learning face deblurring fast and wide," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 745–753.

[58] M. S. Shakeel and W. Kang, "Efficient blind image deblurring method for palm print images," in *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, 2015, pp. 1–7.

[59] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun, "Pose-guided photorealistic face rotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8398–8406.

[60] P. Li, X. Wu, Y. Hu, R. He, and Z. Sun, "M2FPA: A multi-yaw multi-pitch high-quality dataset and benchmark for facial pose analysis," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 10 043–10 051.

[61] Y. Yin, S. Jiang, J. P. Robinson, and Y. Fu, "Dual-attention GAN for large-pose face frontalization," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020, pp. 249–256.

[62] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Towards large-pose face frontalization in the wild," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3990–3999.

[63] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing *et al.*, "Towards pose invariant face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2207–2216.

[64] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 994–1003.

[65] J. Liu, W. Li, H. Pei, Y. Wang, F. Qu, Y. Qu, and Y. Chen, "Identity preserving generative adversarial network for cross-domain person re-identification," *IEEE Access*, vol. 7, pp. 114 021–114 032, 2019.

[66] N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper, "D-id-net: Two-stage domain and identity learning for identity-preserving image generation from semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3677–3682.

[67] J. Yu and J. Jiang, "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2016, pp. 236–246. [Online]. Available: https://aclanthology.org/D16-1023

[68] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 94–108.

[69] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3994–4003.

[70] K. Yu, X. Wang, C. Dong, X. Tang, and C. C. Loy, "Path-restore: Learning network path selection for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[71] R. Gao and K. Grauman, "On-demand learning for deep image restoration," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1086–1095.

[72] S. Zhang, R. He, Z. Sun, and T. Tan, "Multi-task convnet for blind face inpainting with application to face verification," in *2016 International Conference on Biometrics (ICB)*. IEEE, 2016, pp. 1–8.

[73] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, "Rotating your face using multi-task deep neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 676–684.

[74] H. Jung, Y. Kim, H. Jang, N. Ha, and K. Sohn, "Multi-task learning framework for motion estimation and dynamic scene deblurring," *IEEE Transactions on Image Processing*, vol. PP, pp. 1–1, 09 2021.

[75] R. Ranjan, V. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, 03 2016.

[76] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3150–3158, 2016.

[77] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 2048–2057. [Online]. Available: https://proceedings.mlr.press/v37/xuc15.html

[78] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1316–1324, 2018.

[79] Z. Yang, X. He, J. Gao, L. Deng, and A. Smola, "Stacked attention networks for image question answering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21–29, 2016.

[80] M. Suin, K. Purohit, and A. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3606–3615.

[81] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5571–5580, 2019.
[82] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[83] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5908–5916.
[84] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 694–711.
[85] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[86] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.
[87] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *European Conference on Computer Vision Workshops*, 2018.
[88] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690.
[89] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2439–2448.
[90] Y. Wen, J. Chen, B. Sheng, Z. Chen, P. Li, P. Tan, and T.-Y. Lee, "Structure-aware motion deblurring using multi-adversarial optimized CycleGAN," *IEEE Transactions on Image Processing*, vol. 30, pp. 6142–6155, 2021.
[91] S. Ruder, "An overview of multi-task learning in deep neural networks," *ArXiv*, vol. abs/1706.05098, 2017.
[92] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241.
[93] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 539–546 vol. 1.
[94] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
[95] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
[96] Multimodal Dataset, *Biometrics and Identification Innovation Center, http://biic.wvu.edu/.*
[97] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martinez, and J. Fernández-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 3. IEEE, 2000, pp. 314–317.
[98] Lin Hong, Yifei Wan, and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 777–789, 1998.
[99] V. Srinivasan and N. Murthy, "Detection of singular points in fingerprint images," *Pattern Recognition*, vol. 25, no. 2, pp. 139–153, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0031320392900962
[100] P. Birajadar, M. Haria, P. Kulkarni, S. Gupta, P. Joshi, B. Singh, and V. M. Gadre, "Towards smartphone-based touchless fingerprint recognition," *Sādhanā*, vol. 44, pp. 1–15, 2019.
[101] Neurotechnology Inc., https://www.neurotechnology.com/verifinger.html.
[102] Innovatrics, https://www.innovatrics.com/innovatrics-abis/.
[103] C. Watson, M. Garris, E. Tabassi, C. Wilson, R. McCabe, S. Janet, and K. Ko, "NIST biometric image software," 2011.

AMOL S. JOSHI received a B.Sc. degree in computer science and an M.Sc. degree in information technology from India. He is currently pursuing a Ph.D. degree at the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. His research domains include deep learning, machine learning, and their application in computer vision and biometrics.

ALI DABOUEI received his master of science in electrical engineering from Sharif University of Technology, Tehran, Iran. He has been pursuing his doctoral studies since 2017 at West Virginia University, USA. His area of research includes machine learning, deep learning, and their applications in computer vision and biometrics. His research received best student paper awards at IEEE International Conference on Biometrics (BTAS), 2018, and IAPR International Conference on Biometrics (ICB), 2018. He has authored over 20 publications, including journals and peer-reviewed conferences. He has also served as a reviewer for prestigious venues such as IEEE Transactions on Neural Networks and Learning Systems (TNNLS), International Conference on Computer Vision and Pattern Recognition (CVPR), and International Conference on Computer Vision (ICCV).

JEREMY DAWSON ,Associate Professor, joined the Lane Department of Computer Science and Electrical Engineering in Fall of 2015. His background is in microelectronics and nanophotonics, and he has extensive experience in complex, multi-domain system integration and hardware system implementation. His current biometrics research efforts are focused on the creation of large-scale biometrics datasets that can be used to evaluate sensor operation and other human factors, as well as apply novel machine learning algorithms to solve human identification challenges. Dr. Dawson also has extensive experience in micro and nanophotonic sensor platforms. His research in biosensors led to the identification of a need for new signal processing methodologies for DNA systems, which resulted in the first molecular biometrics (DNA) project funded through the WVU Center for Identification Technology Research (CITeR), an NSF IUCRC.

NASSER M. NASRABADI (S'80 – M'84 – SM'92 – F'01) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from the Imperial College of Science and Technology, University of London, London, U.K., in 1980 and 1984, respectively. In 1984, he was with IBM, U.K., as a Senior Programmer. From 1985 to 1986, he was with the Philips Research Laboratory, New York, NY, USA, as a member of the Technical Staff. From 1986 to 1991, he was an Assistant Professor with the Department of Electrical Engineering, Worcester Polytechnic Institute, Worcester, MA, USA. From 1991 to 1996, he was an Associate Professor with the Department of Electrical and Computer Engineering, State University of New York at Buffalo, Buffalo, NY, USA. From 1996 to 2015, he was a Senior Research Scientist with the U.S. Army Research Laboratory. Since 2015, he has been a Professor at the Lane Department of Computer Science and Electrical Engineering. His current research interests are in image processing, computer vision, biometrics, statistical machine learning theory, sparsity, robotics, and neural network applications to image processing. He is a fellow of ARL and SPIE and has served as an Associate Editor for the IEEE Transactions on Image Processing, the IEEE Transactions on Circuits, Systems, and Video Technology, and the IEEE Transactions on Neural Networks.