This WACV 2022 workshop paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the accepted version; the final published version of the proceedings is available on IEEE Xplore.

# Synthesizing Face Images from Match Scores

Thomas Swearingen and Arun Ross Michigan State University

{swearin3, rossarun}@cse.msu.edu

## Abstract

In this paper, we consider the problem of generating a face image based on its match scores with other face images. Such an exercise is not only useful in understanding the relationship between face images, but it can also be used to understand the degree of privacy associated with match scores. We address the problem using two approaches. The first mixes face images to deduce the appearance of a missing face image and the second uses a convolutional autoencoder to further enhance the mixed face image. Experiments suggest the potential of the proposed approaches in generating a missing face image in a database by utilizing its relationship with other images in the database.

## 1. Introduction

Biometrics is the process of recognizing people using biological or behavioral attributes like face, fingerprint, iris, voice, or gait [12]. A typical biometric system includes a feature extraction stage and a comparison stage. In the feature extraction stage, a biometric sample (e.g., face image) is analyzed by a machine to extract salient features. In the comparison stage, the features corresponding to a pair of biometric samples are compared. The result of the comparison is a *match score* which is a single number representing the degree of similarity (if a similarity score) or dissimilarity (if a distance score) between the biometric samples. In the verification mode, two biometric samples are compared and the match score is used to render a decision (match or no-match). In the identification mode, an input (probe) biometric sample is compared against several biometric samples in a database (gallery) and the resulting match scores are used to deduce the identity of the probe.

With the rise in deployment of biometric systems [8], so too has the interest in the security and privacy of these systems. For example, it has now been established that the feature set stored in a database (known as the template) can be used to reconstruct the original biometric sample in some cases [14, 19, 6] and thereby compromising the privacy of subjects. Further, the templates may also reveal de-



Figure 1: An illustration depicting face images and match scores, with one face image missing (indicated by a question mark). The aim of this work is to deduce the missing face image using information from the remaining face images and the match scores between the missing image and the other images. Face images are from the LFW dataset [11].

mographic attributes of an individual such as age, gender, or race [20].

While previous work has focused on the security and privacy of the extracted features, one other entity has received relatively less attention – the match scores. Match scores can reveal information about an unknown or missing face image. For example, if it is known that the missing face image has a similarity match score of 0.2 with Alice and 0.6 with Bob, then we can infer that the missing face will look more like Bob than Alice. Availability of additional match scores pertaining to several identities can potentially aid in reconstructing a missing face image with more certainty. In 2007, Mohanty *et al.* [16] were successfully able to generate face images from match scores by modeling a matcher using a "break in" set. The generated face images were successfully able to fool simple face matchers (*e.g.*, Eigenfaces).

In this paper, we explore the effectiveness of using only relationship information (*i.e.*, match scores) to learn a concept (*i.e.*, a face image). Consider a dataset consisting of Nfaces where the match scores between every pair of faces is known (resulting in  $\binom{N}{2}$  match scores). If some of the face images were missing, can we use the match scores to synthesize a missing face in a single shot?

The first benefit for answering this question is related to privacy. Match scores are not generally considered personal identification information. As biometric databases from governments and companies grow in size, then understanding the risk of compromise is paramount. If some match scores can be used to synthesize a subject's appearance, then the match scores require additional privacy considerations.

A second use-case is related to missing data. In an operational setting, such a technique could be useful when secure data goes missing due to accidentally deleted data, hardware failures leading to data loss, or adversaries preventing data access by installing ransomware. If all data is lost, then re-acquiring the data is the only option. However, if there is partial data loss, then the relationship of the known data to the missing data may assist with its recovery.

In this work, we consider a set of face images and match scores. The match scores between every pair of images is assumed to be known including the match scores between the known face images and the missing face image. We seek to determine if this information can be used to successfully generate the missing face image in a single shot (rather than multiple shots like in a hill-climbing attack). Figure 1 illustrates this problem.

Our paper is organized as follows. Section 2 reviews autoencoders. Section 3 formulates the problem in detail. Section 4 and Section 5 presents the proposed approaches. Section 6 describes the data, and Section 7 reports the experiments and results. Section 8 discusses the results, and Section 9 concludes the work.

### 2. Autoencoders

Autoencoders provide a method to condense and then regenerate data [10]. At a high-level, an autoencoder is a neural network which copies the input to the output. An autoencoder consists of an encoder network,  $E(\cdot)$ , and a decoder network,  $D(\cdot)$ . The encoder network takes input,  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ , and outputs a corresponding latent code,  $Z = \{z_1, z_2, ..., z_n\}$ , where

$$\boldsymbol{z}_i = E(\boldsymbol{x}_i)$$
.

The decoder network takes the latent code,  $z_i$ , and outputs a reconstructed input,  $\tilde{x}_i$ , *i.e.*,

$$\tilde{\boldsymbol{x}}_i = D(\boldsymbol{z}_i)\,,$$

such that  $\tilde{x}_i$  is approximately the same as  $x_i$ . Originally, autoencoders were used to reconstruct the input faithfully so that the latent code could be used as a compressed representation or feature. However, subsequently, autoencoders were developed with alternate goals such as to generate or modify images (e.g., in medicine [9], astronomy [13], or biometrics [23]).

The use of autoencoders with face images is particularly relevant to our work. Autoencoders have been extensively used in the context of face recognition and analysis. Examples include altering facial appearance [22, 26]; simulating facial aging [24]; imparting facial privacy [15]; reconstructing occluded or obscured faces [2, 25]; face frontalization [1]; and cross-spectral face recognition [5]. Autoencoders have also been directly incorporated into face recognition networks [7, 21].

## 3. Problem Formulation

Suppose we are given N face images,  $x_1, x_2, \ldots, x_N$ . We use a matcher,  $m(\cdot, \cdot)$ , which takes two face images as input and produces a face similarity score, s. For two images  $x_i$  and  $x_j$ , a match score between them is given by  $s_{i,j} = m(\boldsymbol{x}_i, \boldsymbol{x}_j)$ . We have a set of face images,  $\mathcal{X} = \{ \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N \}$ , and a set of face similarity scores,  $\mathcal{S} = \{s_{1,2}, s_{1,3}, \dots, s_{1,N}, s_{2,3}, \dots, s_{N-1,N}\}.^{1}$  Now, suppose one of the face images  $x_t$  is no longer available. Can we deduce the missing face image  $x_t$ ? That is, can we synthesize an image  $\tilde{x}_t$  from  $\{\mathcal{X} \setminus x_t, \mathcal{S}\}$  such that  $\tilde{x}_t$  and  $x_t$ are similar?

### 4. Image Mixing Approach

A simple approach for deducing the missing image is by mixing other images that are similar to it. Such an approach has been previously used to enhance the privacy of face images [17]. Figure 2 illustrates the approach. To create a mixed image  $(\tilde{x}_t)$ , the input images from the set  $\mathcal{X} \setminus x_t$  are considered.<sup>2</sup> We record the indices of images which satisfy the condition  $s_{i,t} \geq \tau$ , *i.e.*, images whose match scores with the missing image are greater than a threshold. Thus the indices of the mixing set for a target image  $x_t$  are given by:

$$\mathcal{M}_t = \{i \mid s_{i,t} \ge \tau\}.$$

If  $|\mathcal{M}_t| > K$ , we only retain the K indices corresponding to the K highest match scores.

To create a single image,  $\tilde{x}_t$ , to reconstruct the target image  $x_t$ , we use a weighted average. The mixed image

<sup>&</sup>lt;sup>1</sup>To simplify notation, and since the scores are assumed to be symmetric  $(s_{i,j} = s_{j,i})$ , we will use  $s_{i,j}$  and  $s_{j,i}$  interchangeably. <sup>2</sup>The notation  $A \setminus B$  indicates set difference, *i.e.*, elements in A that

are not in B.



Figure 2: An overview of the two approaches to reconstruct face images from match scores. The first, a face mixing approach, simply uses a weighted average of the input images to create a mixed face. The second takes as input the mixed face image and modifies it using a convolutional autoencoder to a produce a synthetic version of the target image. Face images (with the eye region obscured to preserve privacy) in this figure are from the BIOMDATA dataset [3].

is a weighted average of the unaligned input face images. That is,

$$ilde{oldsymbol{x}}_t = \sum_{i \in \mathcal{M}_t} rac{s_{i,t}}{\sum_{j \in \mathcal{M}_t} s_{j,t}} oldsymbol{x}_i$$

We evaluate the quality of the mixed image,  $\tilde{x}_t$ , by determining its biometric similarity with  $x_t$ .

## 5. Convolutional Autoencoder Approach

The goal of our work is to infer the face image of a person based on their relationship to other face images of different individuals. To carry out this task, we need to define the function,  $f(\cdot)$ , which will further refine the mixed face image  $(\tilde{x}_t)$ . That is,

$$\hat{\boldsymbol{x}}_t = f(\tilde{\boldsymbol{x}}_t)\,,$$

where,  $\tilde{x}_t$  is the mixed face image calculated as described in Section 4 and  $\hat{x}_t$  is the synthetic image (which should reconstruct  $x_t$  as faithfully as possible). Next, we use an autoencoder neural network to further enhance the mixed images generated in Section 4. The mixed images result from a simple average of input face images without alignment, *i.e.*, face landmarks like the mouth, nose, or eyes are not aligned. This gives the mixed face a "blurry" appearance (see examples in Figure 4). Here, we train the autoencoder to synthesize a face image which looks even more like the target image. Figure 2 illustrates the approach.

### 5.1. Convolutional Autoencoder

The Convolutional Autoencoder (Conv. AE) takes a single channel  $d \times d$  face image as input and outputs a single channel  $d \times d$  face image,  $\hat{x}_t$ . The autoencoder consists of two parts: an encoder and a decoder. The encoder,  $E(\cdot)$ , takes the mixed face image as input and outputs a latent code. The decoder,  $D(\cdot)$ , takes the latent code as input and outputs a single face image.

As our datasets are small, the encoder and decoder consist of only 3 convolution layers and 3 deconvolution layers, respectively. All layers of the encoder use Leaky ReLU activation. The first two layers of the decoder use Leaky ReLU activation while the third layer layer uses TanH activation. See Table 1 for details about the architecture.

### 5.2. Loss Function

The loss function for our network consists of two components:  $L_1$  loss and identity loss. The first term in the loss function is a per-pixel loss. This loss can be defined as

$$\mathcal{L}_{\text{PP}}(E, D, \tilde{\boldsymbol{x}}_t, \boldsymbol{x}_t) = \|D(E(\tilde{\boldsymbol{x}}_t)) - \boldsymbol{x}_t\|_{1}$$

This term guides the training to learn a mapping where the output is visually similar to the ground-truth image  $x_t$ . The second term is an identity loss. This loss can be defined as

$$\mathcal{L}_{\mathrm{I}}(E, D, \tilde{\boldsymbol{x}}_{t}, \boldsymbol{x}_{t}) = \left\| \Phi\left( D(E\left(\tilde{\boldsymbol{x}}_{t}\right)) \right) - \Phi\left(\boldsymbol{x}_{t}\right) \right\|_{1},$$

where,  $\Phi(\cdot)$  returns the feature representation from a VGG-19 network [18] pre-trained on a face dataset. This term guides the training to learn a mapping where the synthesized face image shares similar biometric content as the target ground truth image. The overall loss is given as follows:

$$\begin{aligned} \mathcal{L}\left(E, D, \tilde{\boldsymbol{x}}_{t}, \boldsymbol{x}_{t}\right) &= \lambda_{\text{PP}} \mathcal{L}_{\text{PP}}\left(E, D, \tilde{\boldsymbol{x}}_{t}, \boldsymbol{x}_{t}\right) \\ &+ \lambda_{\text{I}} \mathcal{L}_{\text{I}}\left(E, D, \tilde{\boldsymbol{x}}_{t}, \boldsymbol{x}_{t}\right). \end{aligned}$$

Module	Layer	Input Size	Output Size	Activation
	$2 \times 2$ Conv. @32	112x112x1	56x56x32	Leaky ReLU
Encoder	$2 \times 2$ Conv. @64	56x56x32	28x28x64	Leaky ReLU
	$2 \times 2$ Conv. @128	28x28x64	14x14x128	Leaky ReLU
Decoder	$2 \times 2$ DeConv. @64	14x14x128	28x28x64	Leaky ReLU
	$2 \times 2$ DeConv. @32	28x28x64	56x56x32	Leaky ReLU
	$2 \times 2$ DeConv. @1	56x56x32	112x112x1	TanH

Table 1: Architecture of the Convolutional Autoencoder (Conv. AE).



Figure 3: Examples of face images (with the eye region obscured to preserve privacy) in the BIOMDATA dataset.

where,  $\lambda_{PP}$  and  $\lambda_{I}$  denote the regularization parameters.

## 6. Datasets

In this work we use two datasets, the BIOMDATA dataset and a proprietary (PROP) dataset. These two datasets contain multiple modalities, but we use only the face modality in this work. These two datasets were selected as they contain frontal face images. In this preliminary work, the goal is to evaluate the possibility of generating a missing face image from match scores. "In-the-wild" face images have a myriad of pose and illumination variations that can detract from the task of face reconstruction by introducing too many uncontrolled variables.

Table 2: Summary of the BIOMDATA and PROP datasets used in this work.

Dataset	Partition	# Images	# Subjects
	Fold 1	320	48
	Fold 2	320	48
BIOMDATA	Fold 3	320	48
	Fold 4	320	48
	Fold 5	318	48
	Fold 1	442	220
	Fold 2	442	220
PROP	Fold 3	441	220
	Fold 4	441	220
	Fold 5	440	219

#### 6.1. BIOMDATA

The WVU Multimodal Biometric Dataset Collection (BIOMDATA) Release 1 consists of biometric data collected in a controlled setting [3]. The available modalities include iris, face, voice, fingerprint, hand geometry, and palm print. In this work, we use only the face modality data from BIOMDATA. The face images were collected at 5 different yaw angles  $(-90^{\circ}, -45^{\circ}, 0^{\circ}, 45^{\circ}, 90^{\circ})$ , but we use only the  $0^{\circ}$  yaw angle face images (*i.e.*, those with subjects looking directly at the camera). There are 1,598 faces images of 240 subjects with an average of 6.7 images per subject. The data is split into 5-folds in a subject-disjoint manner. Table 2 summarizes the data and Figure 3 shows examples of images from the dataset.

## 6.2. PROP

We also use a proprietary dataset (referred to as PROP) consisting of biometric data collected in a controlled setting. The available modalities include iris and face. In this work, we use only the face modality data from PROP. There are 2,206 faces images of 1,099 subjects with an average of 2.0073 images per subject. The data is split into 5-folds in a subject-disjoint manner. Table 2 summarizes the data.

### 7. Experiments

Our experiments evaluate the face mixing approach and the convolutional autoencoder approach. We refer to images generated by the face mixing approach as *mixed* images  $(\tilde{x}_t)$  and images generated by the convolutional autoencoder as *synthetic* images  $(\hat{x}_t)$ . We use the term *generated* image to mean either a mixed image or a synthetic image.

All experiments are 5-fold cross-validated. In each run, four of the folds are combined into a single training set and the remaining fifth fold is used as the test set. In each run, a different fold is used as the test set.

We use two face matchers in this work: (1) the Arc-Face matcher [4] (specifically the "buffalo\_l" model pack<sup>3</sup>)

<sup>&</sup>lt;sup>3</sup>https://github.com/deepinsight/insightface/

tree/master/python-package

<sup>(</sup>commit ID: 41bf106fa0f9a3c998dd717f2abb5f3c2fada6da)

Table 3: Statistics indicating the (a) percentage of "missing" images for which input images were available (*i.e.*, impostor scores greater than  $\tau$ ); and (b) the percentage of images for which a template could be extracted using the COTS matcher. Statistics are reported as mean and standard deviation across the five folds.

Dataset	Method	Successfully Generated	Successfully Extracted
BIOMDATA	Mixing ( $\tau = 1.1$ )	$100.0\% \pm 0.0\%$	$99.9\% \pm 0.3\%$
	Conv. AE ( $\tau = 1.1$ )	$100.0\% \pm 0.0\%$	$99.8\%\pm0.3\%$
	Mixing ( $\tau = 1.2$ )	$40.1\% \pm 7.4\%$	$49.1\% \pm 7.4\%$
	Conv. AE ( $\tau = 1.2$ )	49.170 ± 7.470	$49.1\% \pm 7.4\%$
PROP	Mixing ( $\tau = 1.1$ )	$100.0\% \pm 0.0\%$	$99.5\% \pm 0.4\%$
	Conv. AE ( $\tau = 1.1$ )	$100.0\% \pm 0.0\%$	$99.5\% \pm 0.4\%$
	Mixing ( $\tau = 1.2$ )	60.4% + 3.2%	$69.1\% \pm 3.1\%$
	Conv. AE ( $\tau = 1.2$ )	$09.4\% \pm 3.2\%$	$69.0\% \pm 3.2\%$

and (2) a commercial-off-the-shelf (COTS) matcher. The ArcFace matcher is used to compute match scores between images in a dataset which are then used for selecting impostor images and generating the mixed image. The COTS matcher is used to compare the generated image with the original ground-truth image. For the ArcFace matcher, the resulting match score is in the range 0 to 2 (*i.e.*,  $m(\boldsymbol{x}_i, \boldsymbol{x}_j) \in [0, 2]$ ). For the COTS matcher, the resulting match score is in the range 0 to 1 (*i.e.*,  $m(\boldsymbol{x}_i, \boldsymbol{x}_j) \in [0, 1]$ ).

To evaluate the quality of the mixed image and the synthetic image, we require a threshold to render a match or non-match decision. We calculate the threshold corresponding to a False Match Rate (FMR) of 1% for the COTS matcher. This threshold is 0.405 for the BIOMDATA dataset and 0.250 for the PROP dataset.

### 7.1. Face Mixing

For the face mixing approach, we use K = 10. Thus, we construct a generated target image from the 10 imposter images with the highest match score (provided they are above the threshold  $\tau$ ). We also experiment with two values for  $\tau$ : (1)  $\tau = 1.1$  and (2)  $\tau = 1.2$ . These  $\tau$  thresholds correspond to score values near the 80th and 98th percentile of all imposter scores. Table 3 shows the percentage of images that are generated and the percentage of templates that are extracted by the COTS matcher. Figure 4 shows examples of generated images for the BIOMDATA dataset with  $\tau = 1.1$  and  $\tau = 1.2$ . Figure 5 shows the matching results for  $\tau = 1.2$ . Table 4 shows the percentage of mixed faces that are correctly matched at the threshold corresponding to the 1% FMR.

#### 7.2. Convolutional Autoencoder Approach

For the Convolutional Autoencoder approach, we first train the network from scratch using the training set of each cross-validation run. For this approach, the mixed image is used as input to the Convolutional Autoencoder. This experiment is repeated twice, once with  $\tau = 1.1$  and once with  $\tau = 1.2$ . The training proceeds for 1000 epochs with a learning rate of 0.00001, a weight decay of 0.00001,  $\lambda_{PP} = 0.5$ , and  $\lambda_I = 0.5$ . Table 3 shows the percentage of images that are generated and the percentage of templates that are extracted by the COTS matcher. Figure 4 shows examples of generated images for the BIOMDATA dataset with  $\tau = 1.1$  and  $\tau = 1.2$ . The match scores between the synthetic image ( $\hat{x}_t$ ) and the real image ( $x_t$ ) are presented in Figure 7 for  $\tau = 1.1$  and Figure 8 for  $\tau = 1.2$  Table 4 shows the percentage of synthetic faces that are accepted at the threshold corresponding to the 1% FMR.

### 8. Discussion

The mixing approach was observed to be more successful than the autoencoder approach as far as matching with the original image is concerned. However, the autoencoder approach generated images from which face templates were more successfully extracted by the COTS matcher.

The face mixing approach successfully generates the missing target image when the minimum imposter score threshold  $(\tau)$  is high. However, there is a clear trade-off between the number of successfully generated images and the number of mixed images accepted at the 1% FMR threshold. If  $\tau$  is low, then the number of successful generations is high (see Table 3) but the percentage of generations which would be classified as a match at the 1% FMR threshold is low (see Table 4). But if  $\tau$  is high, then the number of successful generation is low (see Table 3) but the percentage of generations which would be classified as a match at the 1% FMR threshold is high (see Table 4). A potential way forward is to use a larger database of face images. The number of input images (K) is fixed regardless of database size, so increasing the number of imposters will also increase the chances of observing imposters above the threshold  $\tau$ . It makes intuitive sense that higher values of  $\tau$  produce better reconstructions of the target image. The higher the imposter score the more similar a missing image is with that impos-



Figure 4: Examples of generated BIOMDATA dataset face images using the mixing approach and the Convolutional Autoencoder (Conv. AE) approach with  $\tau = 1.1$  and  $\tau = 1.2$ .

Table 4: Results of comparing the generated image and the ground-truth target image using a COTS matcher. The acceptance	ce
rates are given as the mean percentage across all test folds for the 5 runs, plus or minus the standard deviation.	

Dataset	1% FMR Threshold	Method	Generated Test Images Accepted at 1% FMR
BIOMDATA	0.405	Mixing ( $\tau = 1.1$ )	$24.6\% \pm 7.1\%$
		Mixing ( $\tau = 1.2$ )	$48.5\% \pm 6.9\%$
		Conv. AE ( $\tau = 1.1$ )	$11.1\% \pm 5.1\%$
		Conv. AE ( $\tau = 1.2$ )	$21.4\%\pm6.9\%$
PROP	0.250	Mixing ( $\tau = 1.1$ )	$58.5\% \pm 3.0\%$
		Mixing ( $\tau = 1.2$ )	$70.0\% \pm 5.4\%$
		Conv. AE ( $\tau = 1.1$ )	$40.8\% \pm 2.4\%$
		Conv. AE ( $\tau = 1.2$ )	$42.9\%\pm2.8\%$

tor image. Thus, better information is available to generate the missing biometric content.

A limitation of the convolutional autoencoder approach is it requires a fixed number of input channels. However, if the number of input channels could vary dynamically, we can input each individual image used to create the mixed image as a separate channel into the convolutional autoencoder (rather than input a single mixed image). Larger datasets might also allow the network to be deeper facilitating a better reconstruction of the target image.

As stated earlier, the ArcFace matcher was used to select imposter images used in the mixing process, while the COTS matcher was used to compare the generated image to the ground-truth image. These matchers, presumably, use different techniques to output a match score. This demonstrates that missing face images generated based on scores from one matcher can produce a face image that can be matched to the original image using an entirely different matcher.

One other interesting observation is that the PROP dataset has much higher generation capability compared to the BIOMDATA dataset. Both contain frontal face images captured in controlled conditions, so it is unlikely that images in one dataset exhibit wider variations in pose or illumination compared to the other. The threshold corresponding to the 1% FMR is quite different between the datasets, with the PROP threshold being lower than the BIOMDATA threshold (see Table 4). This could indicate that this particu-



Figure 5: Score histogram when comparing the mixed face image  $(\tilde{x}_t)$  with  $\tau = 1.1$  and the ground-truth target image  $(x_t)$ . The colored bar shows the mean probability across all test folds for the 5 runs, while the error bar shows the standard deviation.



Figure 6: Score histogram when comparing the mixed face image  $(\tilde{x}_t)$  with  $\tau = 1.2$  and the ground-truth target image  $(x_t)$ . The colored bar shows the mean probability across all test folds for the 5 runs, while the error bar shows the standard deviation.

lar COTS matcher is better able to represent the faces in the PROP dataset, as the imposter scores more frequently occur near the bottom of the score range (*i.e.*, overall, the imposter scores are closer to 0). However, the threshold  $\tau$  is fixed at a value shared for both datasets and is not changed based on the 1% FMR threshold. The PROP dataset has a 1% FMR threshold of 0.250 while the BIOMDATA has a 1% FMR



Figure 7: Score histogram when comparing the synthetic face image  $(\hat{x}_t)$  with  $\tau = 1.1$  and the ground-truth target image  $(x_t)$ . The colored bar shows the mean probability across all test folds for the 5 runs, while the error bar shows the standard deviation.



Figure 8: Score histogram when comparing the synthetic face image  $(\hat{x}_t)$  with  $\tau = 1.2$  and the ground-truth target image  $(x_t)$ . The colored bar shows the mean probability across all test folds for the 5 runs, while the error bar shows the standard deviation.

threshold of 0.405. Thus, when  $\tau = 1.1$ , more number of PROP input images will be above the 1% FMR threshold and, when combined, will likely be classified as a match at the 1% FMR threshold.

## 9. Summary and Future Work

In this preliminary work, we evaluated the use of match scores to deduce a missing face image. We introduced two approaches, a face mixing approach and a convolutional autoencoder approach. The face mixing approach resulted in some modest success at generating missing images. We also observed a clear relationship between the minimum threshold and the quality of the generated image. Possible future directions include improvements to the convolutional autoencoder by including the match scores directly as input or by more explicitly ingesting relationship information in the form a graph. In such a network, the relationship information could reveal additional cues on how the imposter images relate to the target image. In summary, this work suggests that match scores can potentially be used to reveal face images.

### References

- Jie Cao, Yibo Hu, Hongwen Zhang, Ran He, and Zhenan Sun. Towards high fidelity face frontalization in the wild. *International Journal of Computer Vision*, 128(5):1485–1504, 2020.
- [2] Lele Cheng, Jinjun Wang, Yihong Gong, and Qiqi Hou. Robust deep auto-encoder for occluded face recognition. In *Proceedings of the ACM International Conference on Multimedia*, pages 1099–1102, 2015.
- [3] S Crihalmeanu, Arun Ross, Stephanie Schuckers, and L Hornak. A protocol for multibiometric data acquisition, storage and dissemination. In *Technical Report*, WVU, Lane Department of Computer Science and Electrical Engineering. 2007.
- [4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference* on Computer Vision and Pattern Recognition, pages 4690– 4699, 2019.
- [5] Chaoyou Fu, Xiang Wu, Yibo Hu, Huaibo Huang, and Ran He. Dual variational generation for low shot heterogeneous face recognition. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- [6] Javier Galbally, Arun Ross, Marta Gomez-Barrero, Julian Fierrez, and Javier Ortega-Garcia. Iris image reconstruction from binary templates: An efficient probabilistic approach based on genetic algorithms. *Computer Vision and Image Understanding*, 117(10):1512–1525, 2013.
- [7] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang. Single sample face recognition via learning deep supervised autoencoders. *IEEE Transactions on Information Forensics and Security*, 10(10):2108–2118, 2015.
- [8] Rachel German and K. Suzanne Barber. Current biometric adoption and trends. Technical report, University of Texas at Austin CID Report #18-02, 2017.
- [9] L. Gondara. Medical image denoising using convolutional denoising autoencoders. In *Proceeding of the IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 241–246, 2016.

- [10] Geoffrey E Hinton and Richard S Zemel. Autoencoders, minimum description length, and helmholtz free energy. Advances in neural information processing systems, 6:3–10, 1994.
- [11] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [12] Anil K Jain, Arun A Ross, and Karthik Nandakumar. Introduction to Biometrics. Springer, 2011.
- [13] A. Karmakar, D. Mishra, and A. Tej. Stellar cluster detection using gmm with deep variational autoencoder. In Proceedings of Conference on Recent Advances in Intelligent Computational Systems (RAICS), pages 122–126, 2018.
- [14] Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, 41(5):1188–1202, 2018.
- [15] Vahid Mirjalili, Sebastian Raschka, Anoop Namboodiri, and Arun Ross. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In *Proceedings of the International Conference on Biometrics (ICB)*, pages 82–89, 2018.
- [16] Pranab Mohanty, Sudeep Sarkar, and Rangachar Kasturi. From scores to face templates: a model-based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2065–2078, 2007.
- [17] Asem Othman and Arun Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *European Conference on Computer Vision*, pages 682–696, 2014.
- [18] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.
- [19] Arun Ross, Jidnya Shah, and Anil K Jain. From template to image: Reconstructing fingerprints from minutiae points. *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, 29(4):544–560, 2007.
- [20] Philipp Terhörst, Daniel Fährmann, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Beyond identity: What information is stored in biometric face templates? In *Proceedings of the International Joint Conference on Biometrics* (*IJCB*), pages 1–10, 2020.
- [21] Chenfei Xu, Qihe Liu, and Mao Ye. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing*, 222:62–71, 2017.
- [22] Raymond Yeh, Ziwei Liu, Dan B Goldman, and Aseem Agarwala. Semantic facial expression editing using autoencoded flow. arXiv preprint arXiv:1611.09961, 2016.
- [23] Xin Yu and Fatih Porikli. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In *Proceedings of the IEEE Conference* on Computer Vision and Pattern Recognition (CVPR), July 2017.
- [24] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

- [25] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan. Robust lstm-autoencoders for face de-occlusion in the wild. *IEEE Transactions on Image Processing*, 27(2):778–790, 2018.
- [26] Y. Zhou and B. E. Shi. Photorealistic facial expression syn-

thesis by the conditional difference adversarial autoencoder. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 370–376, 2017.