To appear in International Conference On Pattern Recognition (ICPR), 2020

Identifying Missing Children: Face Age-Progression via Deep Feature Aging

Debayan Deb, Divyansh Aggarwal, Anil K. Jain
Michigan State University, East Lansing, MI, USA
{debdebay, aggarw49, jain}@cse.msu.edu

***Abstract*—Given a face image of a recovered child at age age_{probe}, we search a gallery of missing children with known identities and age $age_{gallery}$ at which they were either lost or stolen in an attempt to unite the recovered child with his family. We propose a feature aging module that can age-progress deep face features output by a face matcher to improve the recognition accuracy of age-separated child face images. In addition, the feature aging module guides age-progression in the image space such that synthesized aged gallery faces can be utilized to further enhance cross-age face matching accuracy of any commodity face matcher. For time lapses larger than 10 years (the missing child is recovered after 10 or more years), the proposed age-progression module improves the rank-1 open-set identification accuracy of CosFace from 22.91% to 25.04% on a child celebrity dataset, namely ITWCC. The proposed method also outperforms state-of-the-art approaches with a rank-1 identification rate of 95.91%, compared to 94.91%, on a public aging dataset, FG-NET, and 99.58%, compared to 99.50%, on CACD-VS. These results suggest that aging face features enhances the ability to identify young children who are possible victims of child trafficking or abduction.**

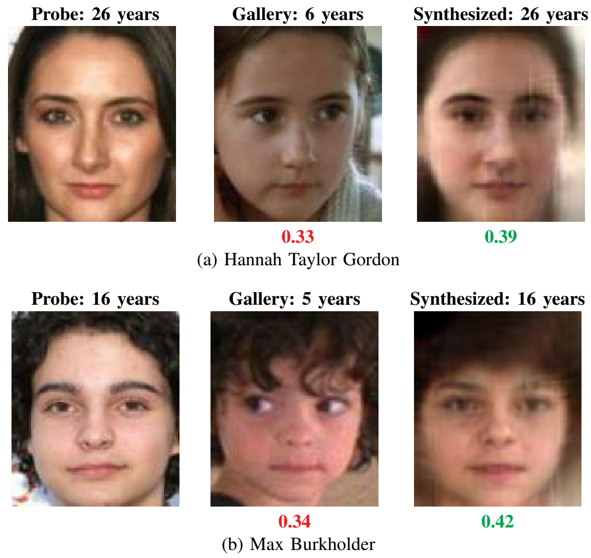


Fig. 1: A state-of-the-art face matcher, CosFace [4] fails to match gallery faces of two child celebrities to their corresponding older probes. With the proposed feature aging scheme, age-progressed faces successfully match the probes with their mates with higher similarity scores (cosine similarity scores given below each score range from $[-1, 1]$ and threshold for CosFace is 0.35 at 0.1% FAR).

I. Introduction

Human trafficking is one of the most adverse social issues currently faced by countries worldwide. According to the United Nations Children's Fund (UNICEF) and the Inter-Agency Coordination Group against Trafficking (ICAT), 28% of the identified victims of human trafficking globally are children[1] [2]. The actual number of missing children is much more than these official statistics as only a limited number of cases are reported because of the fear of traffickers, lack of information, and mistrust of authorities.

Face recognition is perhaps the most promising biometric technology for recovering missing children, since parents and relatives are more likely to have a lost child's face photograph than say fingerprint or iris[2]. While Automated Face Recognition (AFR) systems have been able to achieve high identification rates in several domains [4]–[6], their ability to recognize children as they age is still limited.

A human face undergoes various temporal changes, including skin texture, face morphology, facial hair, etc. (see Figure 1) [7], [8]. Several studies have analyzed the extent to which facial aging affects the performance of AFR. Two major conclusions can be drawn based on these studies: (i) performance decreases with an increase in time lapse between subsequent image acquisitions [9]–[11], and (ii) performance degrades more rapidly in the case of younger individuals than older individuals [11], [12]. Figure 1 illustrates that a state-of-the-art face matcher (CosFace) fails when it comes to matching child's image in the gallery of missing children with the corresponding probe over large time lapses. Thus, it is essential to enhance the cross-age face recognition performance of AFR systems, especially for age-separated child face images.

We propose a feature aging module that learns a projection in the deep feature space (see Figure 2). In addition, the proposed feature aging module can guide the aging process in the image space such that we can synthesize visually convincing face images for any specified target age (not age-groups). These aged images can be used by any face matcher for enhanced cross-age face recognition performance. Our empirical results on several cross-age face datasets show that the

[1]The United Nations Convention on the Rights of the Child defines a child as "a human being below the age of 18 years unless under the law applicable to the child, majority is attained earlier" [1]

[2]Indeed, face is certainly not the only biometric modality for identification of lost children. Sharbat Gula, first photographed in 1984 (age 12) in a refugee camp in Pakistan, was later recovered via iris recognition at the age of 30 from a remote part of Afghanistan in 2002 [3]. However, the iris image was extracted from face image

[2]The award-winning 2016 movie, Lion, is based on the true story of Saroo Brierley [13].

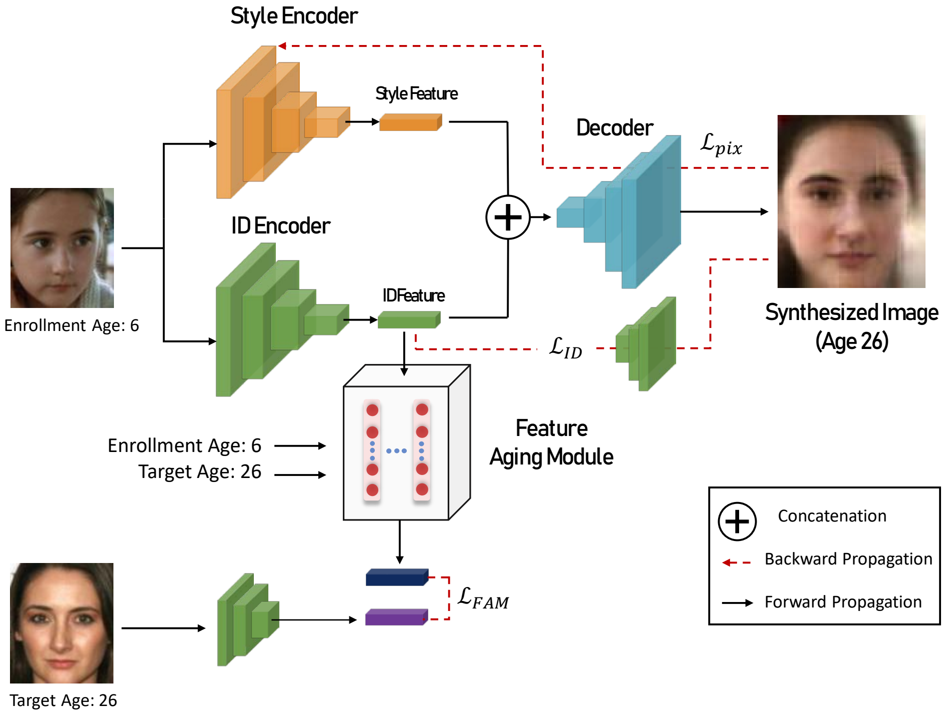


Fig. 2: Overview of training the proposed child face aging framework. The feature module guides the decoder to synthesize a face image to any specified target age while the style-encoder injects style from the input probe into the synthesized face. $\mathcal{L}_{ID}$ represents the Identity-Preservation loss, $\mathcal{L}_{pix}$ represents the pixel supervision loss and $\mathcal{L}_{FAM}$ represents the loss for the Feature Aging Module. For simplicity, we omit the total variation loss ($\mathcal{L}_{TV}$).

proposed feature aging module, along with the age-progressed[3] images, can improve the cross-age face recognition performance of three face matchers (FaceNet [5], CosFace [4], and a commercial-off-the-shelf (COTS) matcher) for matching children as they age.

The specific contributions of the paper are as follows:

- A feature aging module that learns to traverse the deep feature space to preserve subject identity while aging the face to enhance cross-age matching.
- An image aging scheme, guided by our feature aging module, to synthesize an age-progressed or age-regressed face image at any specified target age. The aged face images can also improve cross-age face recognition performance for any commodity face matcher.
- Improved identification (closed-set and open-set) accuracy compared to state-of-the-art face matchers, CosFace [4], FaceNet [5], and a COTS, on two different child face aging datasets: CFA and ITWCC [14]. In addition, the proposed module boosts accuracies on two public face aging benchmark datasets, FG-NET [15] and CACD-VS [16][4].

[3]Though our module is not strictly restricted to age-progression, we use the word progression largely because in the missing children scenario the gallery would generally be younger than the probe. Our module does both age-progression and age-regression when we benchmark our performance on public datasets.

[4]We follow the protocols provided with these datasets.

II. Related Work

A. Discriminative Approaches

Approaches prior to deep learning leveraged robust local descriptors [17]–[21] to tackle recognition performance degradation due to face aging. Recent approaches focus on age-invariant face recognition by attempting to discard age-related information from deep face features [22]–[26]. All these methods operate under two critical assumptions: (i) age and identity related features can be disentangled, and (ii) the identity-specific features are adequate for face recognition performance. On the other hand, several other studies show that age is indeed a major contributor to face recognition performance [27], [28]. Therefore, instead of completely discarding age factors, we exploit the age-related information to progress or regress the deep feature directly to the desired age.

B. Generative Approaches

Ongoing studies leverage Conditional Auto-encoders and Generative Adversarial Networks (GANs) to synthesize faces by learning aging patterns from face aging datasets [25], [29]–[34]. The primary objective of these methods is to synthesize visually realistic face images that appear to be age progressed. As a result, a majority of these studies do not report the face recognition performance on the synthesized faces.

[5]MORPH: https://bit.ly/31P6QMw, CACD: https://bit.ly/343CdVd, FG-NET: https://bit.ly/2MQPL0O, UTKFace: https://bit.ly/2JpvX2b

TABLE I: Face aging datasets. Datasets below solid line includes cross-age face images of children.

Dataset	No. of Subjects	No. of Images	No. Images / Subject	Age Range (years)	Avg. Age (years)	Public[5]
MORPH-II [35]	13,000	55,134	2-53 (avg. 4.2)	16-77	42	Yes
CACD [16]	2,000	163,446	22-139 (avg. 81.7)	16-62	31	Yes
FG-NET [15]	82	1,002	6-18 (avg. 12.2)	0-69	16	Yes
UTKFace [30]†	N/A	23,708	N/A	0-116	33	Yes
ITWCC [36]	745	7,990	3-37 (avg. 10.7)	0-32	13	No††
CLF [12]	919	3,682	2-6 (avg. 4.0)	2-18	8	No††
CFA	9,196	25,180	2-6 (avg. 2.7)	2-20	10	No††

† Dataset does not include subject labels; Only a collection of face images along with the corresponding ages.
†† Concerns about privacy issues are making it extremely difficult for researchers to place the child face images in public domain.

TABLE II: Related work on age-separated face recognition. Studies below bold line deal with children.

Study	Objective	Dataset	Age groups or range (years)
Yang *et al.* [31]*	Age progression of face images	MORPH-II, CACD	31-40, 41-50, 50+
Wang *et al.* [26]*	Decomposing age and identity	MORPH-II, FG-NET, CACD	0-12, 13-18, 19-25, 26-35, 36-45,46-55, 56-65, 65+
Best-Rowden *et al.* [37]	Model for change in genuine scores over time	PCSO, MSP	18-83
Ricanek *et al.* [36]	Face comparison of infants to adults	ITWCC	0-33
Deb *et al.* [12]	Feasibility study of AFR for children	CLF	2-18
This study	Aging face features for enhanced AFR for children	CFA, ITWCC, FG-NET, CACD	0-18

* Study uses cross-sectional model (ages are partitioned into age groups) and not the correct longitudinal model [10], [38].

C. Face Aging for Children

Best-Rowden *et al.* studied face recognition performance of newborns, infants, and toddlers (ages 0 to 4 years) on 314 subjects acquired over a time lapse of only one year [39]. Their results showed a True Accept Rate (TAR) of 47.93% at 0.1% False Accept Rate (FAR) for an age group of [0, 4] years for a commodity face matcher. Deb *et al.* fine-tuned FaceNet [5] to achieve a rank-1 identification accuracy of 77.86% for a time lapse between the gallery and probe image of 1 year. Srinivas *et al.* showed that the rank-1 performance of state of the art commercial face matchers on longitudinal face images from the In-the-Wild Child Celebrity (ITWCC) [14] dataset ranges from 42% to 78% under the youngest to older protocol. These studies primarily focused on evaluating the longitudinal face recognition performance of state-of-the-art face matchers rather than proposing a solution to improve face recognition performance on children as they age. Table I summarizes cross-age face datasets that include children and Table II shows related work in this area.

III. Proposed Approach

Suppose we are given a pair of face images (x_e, y_t) for a child acquired at ages e (enrollment) and t (target), respectively. Our goal is to enhance the ability of state-of-the-art face recognition systems to match x_e and y_t when $(e \ll t)$.

We propose a feature aging module that learns a projection of the deep features in a lower-dimensional space which can directly improve the accuracy of face recognition systems in identifying children over large time lapses. The feature aging module guides a generator to output aged face images that can be used by any commodity face matcher for enhancing cross-age face recognition performance.

A. Feature Aging Module (FAM)

It has been found that age-related components are highly coupled with identity-salient features in the latent space [27], [28]. That is, the age at which a face image is acquired can itself be an intrinsic feature in the latent space. Instead of disentangling age-related components from identity-salient features, we would like to automatically learn a projection within the face latent space.

Assume $x_e \in \mathcal{X}$ and $y_t \in \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are two face domains when images are acquired at ages e and t, respectively. Face manipulations via generative approaches [25], [29]–[34] shift images in the domain of $\mathcal{X}$ to $\mathcal{Y}$ via,

$$\hat{y}_t = \mathcal{F}(x_e, e, t) \tag{1}$$

where, $\hat{y}_t$ is the output image and $\mathcal{F}$ is the operator that changes x_e from $\mathcal{X}$ to $\mathcal{Y}$. Domains $\mathcal{X}$ and $\mathcal{Y}$ generally differ in factors other than aging, such as noise, quality and pose. Therefore, $\mathcal{F}$ can be highly complex. We can simplify $\mathcal{F}$ by modeling the transformation in the deep feature space by defining an operator $\mathcal{F}'$ and rewrite equation 1

$$\hat{y}_t = \mathcal{F}'(\psi(x_e), e, t) \tag{2}$$

where $\psi(x_e)$ is an encoded feature in the latent space. Here, $\mathcal{F}'$ learns a projection in the feature space that shifts an image x_e in $\mathcal{X}$ to $\mathcal{Y}$ and therefore, 'ages' a face feature from age e to age t. Since face representations lie in d-dimensional Euclidean space[6], the latent space $\mathcal{Z}$ is linear. That is, given any face image, face recognition systems encode the features in a linear space where operations, such as matching face features, are performed linearly. Since one of our primary goals is to improve the performance of the face recognition system in its latent space, we have to learn a linear transformation. Non-linearity will result in leaving the latent space completely and will fail to match. Therefore, $\mathcal{F}'$ is a linear shift in the deep space, that is,

$$\hat{y}_t = \mathcal{F}'(\psi(x_e), e, t) = W \times (\psi(x_e) \oplus e \oplus t) + b \tag{3}$$

[6]Assume these feature vectors are constrained to lie in a d-dimensional hypersphere, *i.e.*, $||\psi(x)||_2^2 = 1$.

where $W \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$ are learned parameters of $\mathcal{F}'$ and $\oplus$ is concatenation. Since some face features, such as eye color, do not change drastically during the aging process, the scale parameter, W allows for each feature to scale differently given the enrollment and target ages.

Upchurch *et al.* [40] explored a similar idea of linearly interpolating features in the deep feature space. However, they performed linear interpolation by computing the mean features of two age clusters and moving along that direction by a manually specified "scale" hyperparameter. In contrast, our model can directly model the age clusters without the need for a scale parameter. In other words, deep linear interpolation is a handcrafted version of our automatic feature aging module.

B. Image Generator

While FAM can directly output face embeddings that can be used for face recognition, FAM trained on the latent space of one matcher may not generalize to feature spaces of other matchers. Still, the feature aging module trained for one matcher can guide the aging process in the image space. In this manner, an aged image can be directly used to enhance cross-age face recognition performance for any commodity face matcher without requiring any re-training. The aged images should (i) enhance identification performance of any commodity face matcher under aging, and (ii) appear visually realistic to humans.

IV. Learning Face Aging

Our proposed framework for face aging comprises of three modules, (i) feature aging module (FAM), (ii) style-encoder, and (iii) decoder. For training these three modules, we utilize a fixed ID encoder ($\mathcal{E}_{ID}$) that outputs a deep face feature from an input face image. An overview of our proposed aging framework is given in Figure 2.

A. Feature Aging Module (FAM)

This module consists of a series of fully connected layers that learn the scale W and bias b parameters in Equation 3. The training set consists of a genuine pair of face features $(\psi(x_e), \psi(y_t))$ extracted from an identity encoder ($\mathcal{E}_{ID}$), where x_e and y_t are two images of the same person acquired at ages e and t, respectively. In order to ensure that the identity-salient features are preserved and the synthesized features are age-progressed to the desired age t, we train FAM via a mean squared error (MSE) loss which measures the quality of the predicted features:

$$\mathcal{L}_{FAM} = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} ||\mathcal{F}'(\psi(x_e), e, t) - \psi(y_t)||_2^2, \quad (4)$$

where $\mathcal{P}$ is the set of all genuine pairs. Globally, let S_e and S_t represent all the face images at ages e and t, respectively. Since there is no other systematic bias except for the difference in the ages of these two sets of face images, therefore, when Mean Squared Error is used for training, FAM learns a projection along the age direction keeping other covariates (e.g. expression, pose etc) intact. Our experiments show that $\mathcal{L}_{FAM}$ forces the FAM module to retain all other covariates in the predicted features from the input features. After the model is trained, FAM can progress a face feature to the desired age.

B. Style-Encoder

The ID encoder ($\mathcal{E}_{ID}$) encodes specific information pertaining to the identity of a person's face. However, a face image comprises of other pixel-level residual features that may not relate to a person's identity but are required for enhancing the visual quality of the synthesized face image, which we refer to as $style$. In fact, directly decoding a face feature vector into an image can severely affect the visual quality of the synthesized face. Therefore, we utilize a style-encoder ($\mathcal{E}_{style}$) that takes a face image, x, as an input and outputs a k-dimensional feature vector that encodes style information from the image.

C. Decoder

In order to synthesize a face image, we propose a decoder $\mathcal{D}$ that takes as input a style and an ID vector obtained from $\mathcal{E}_{style}$ and $\mathcal{E}_{ID}$, respectively, and outputs a synthesized image $\hat{y}$. For training $\mathcal{D}$, we do not require the aged features predicted by the feature aging module, since we enforce the predicted aged features to reside in the latent space of $\mathcal{E}_{ID}$.

a) Identity-Preservation: Our goal is to synthesize aged face images that can benefit face recognition systems. Therefore, we need to constrain the decoder $\mathcal{D}$ to output face images that can be matched to the input image. To this end, we adopt an identity-preservation loss to minimize the distance between the input and synthesized face images of the decoder in the $\mathcal{E}_{ID}$'s latent space,

$$\mathcal{L}_{ID} = \sum_{i=0}^{n} ||\mathcal{E}_{ID}(\mathcal{D}(\mathcal{E}_{style}(x_i), \mathcal{E}_{ID}(x_i))) - \mathcal{E}_{ID}(x_i)||_2^2 \quad (5)$$

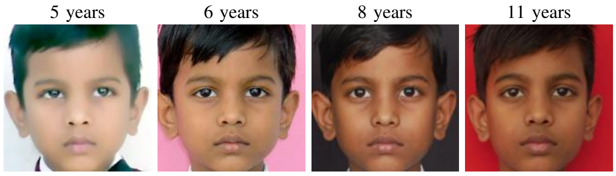
where x_i are samples in the training set. Note that the task of the decoder is just to take as input a style vector and an ID vector and synthesize the corresponding image. To synthesize the age progressed image, the ID vector can be replaced by the aged ID vector output from FAM at test time. Therefore, the identity preservation loss is maintained between the synthesized and the input image as opposed to between synthesized and target image.

b) Pixel-level Supervision: In addition to identity-preservation, a pixel-level loss is adopted to maintain the consistency of low-level image content between the input and output of the decoder,

$$\mathcal{L}_{pix} = \sum_{i=0}^{n} ||\mathcal{D}(\mathcal{E}_{style}(x_i), \mathcal{E}_{ID}(x_i)) - x_i||_1 \quad (6)$$

To synthesize a smooth image, devoid of sudden changes in high-frequency pixel intensities, we regularize the total variation in the synthesized image,

$$\mathcal{L}_{TV} = \sum_{i=0}^{n} \left[\sum_{r,c}^{H,W} \left[\left(x_{i_{r+1,c}} - x_{i_{r,c}}\right)^2 + \left(x_{i_{r,c+1}} - x_{i_{r,c}}\right)^2 \right] \right] \quad (7)$$



(a) CFA

(b) ITWCC [14]

Fig. 3: Examples of cross-age face images from (a) CFA and (b) ITWCC [14] datasets. Each row consists of images of one subject; age at image acquisition is given above each image

Our final training objective for the style-encoder and decoder is,

$$\mathcal{L}_{(\mathcal{E}_{style}, \mathcal{D})} = \lambda_{ID}\mathcal{L}_{ID} + \lambda_{pix}\mathcal{L}_{pix} + \lambda_{TV}\mathcal{L}_{TV} \quad (8)$$

where λ_{ID}, λ_{pix}, λ_{TV} are the hyper-parameters that control the relative importance of every term in the loss. We emperically set $\lambda_{ID} = 1.0$, $\lambda_{pix} = 10.0$, and $\lambda_{tv} = 1e-4$.

We train the Feature Aging Module and the Image Generator in an end-to-end manner by minimizing $\mathcal{L}_{(\mathcal{E}_{style}, \mathcal{D})}$ and $\mathcal{L}_{FAM}$.

V. Implementation Details

a) Feature Aging Module: For all the experiments, we stack two fully connected layers and set the output of each layer to be of the same dimensionality as the ID encoder's feature vector. We train the proposed framework for $200,000$ iterations with a batch size of 64 and a learning rate of 0.0002 using Adam optimizer with parameters $\beta_1 = 0.5, \beta_2 = 0.99$. In all our experiments, $k = 32$. Implementations are provided in the supplementary materials.

b) ID Encoder: For our experiments, we employ 3 pre-trained face matchers[7]. Two of them, FaceNet [5] and CosFace [4], are publicly available. FaceNet is trained on VGGFace2 dataset [41] using the Softmax+Center Loss [5]. CosFace is a 64-layer residual network [42] and is trained on MS-ArcFace dataset [43] using AM-Softmax loss function [43]. Both matchers extract a 512-dimensional feature vector. We also evaluate results on a commercial-off-the-shelf face matcher, COTS[8].

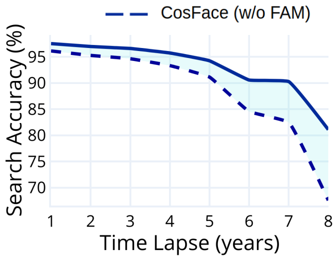
VI. Experiments

A. Cross-Age Face Recognition

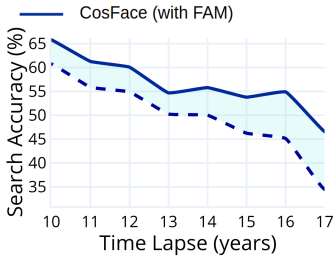
a) Evaluation Protocol: We divide the subjects in CFA and ITWCC datasets into 2 non-overlapping sets which are used for training and testing, respectively. Testing set of ITWCC consists of all subjects with image pairs separated by at least a 10-year time lapse. Since, the maximum time lapse in the CFA dataset is 8 years, testing set of CFA consists of all subjects with image pairs separated by at least 5-year time lapse. As mentioned earlier, locating missing children is akin to the identification scenario, we compute both the *closed-set* identification accuracy (recovered child is in the gallery) at rank-1 and the rank-1 *open-set* identification accuracy (recovered child may or may not be in the gallery) at 1.0% False Accept Rate. For open-set identification scenarios, we extend the gallery of CFA by adding all images in the testing of ITWCC and vice-versa. Remaining images of the subject form the distractor set in the gallery. For the search experiments, we age all the gallery features to the age of the probe and match the aged features (or aged images via our Image Generator) to the probe's face feature.

[7]Both the open-source matchers and the COTS matcher achieve 99% accuracy on LFW under its standard protocol.

[8]This particular COTS has been used for identifying children in prior studies [12], [14]. It is also one of the top performers in the NIST Ongoing Face Recognition Vendor Test (FRVT) [11]. This is a closed system so we do not have access to its feature vector and therefore only utilize COTS as a baseline.



(a) CFA



(b) ITWCC

Fig. 4: Rank-1 search accuracy for CosFace [4] on (a) CFA and (b) ITWCC datasets with and without the proposed Feature Aging Module (FAM).

b) Results: In Table III, we report the Rank-1 search accuracy of our proposed Feature Aging Module (FAM) as well as accuracy when we search via our synthesized age-progressed face images on CFA and ITWCC. We find that the age-progression scheme can improve the search accuracy of both FaceNet [5] and CosFace [4] matchers. Indeed, images aged via the FAM can also enhance the performance of these matchers which highlights the benefit of using both the feature aging scheme and the Image Generator. With the proposed feature aging module, an open-source face matcher CosFace [4] can outperform the COTS matcher[9].

We also investigate the Rank-1 identification rate with varying time lapses between the probe and its true mate in the gallery in Figures 4a and 4b. While our aging model improves matching across all time lapses, its contribution gets larger as the time lapse increases.

In Figure 8, we show some example cases where CosFace [4], without the proposed deep feature aging module, retrieves a wrong child from the gallery at rank-1. With the proposed method, we can correctly identify the true mates for the same probes at Rank-1.

In order to evaluate the generalizability of our module to adults, we train it on CFA and ITWCC [14] datasets

[9]CosFace [4] matcher takes about 1.56ms to search for a probe in a gallery of $10,000$ images of missing children. Our model takes approximately 27.45ms (on a GTX 1080 Ti) to search for a probe through the same gallery size.

TABLE III: Rank-1 identification accuracy on two child face datasets, CFA and ITWCC [14], when the time gap between a probe and its true mate in the gallery is larger than 5 years and 10 years, respectively. The proposed aging scheme (in both the feature space as well as the image space) improves the performance of FaceNet and CosFace on cross-age face matching. We also report the number of probes (P) and gallery sizes (G) for each experiment.

Method	CFA (Constrained)		ITWCC (Semi-Constrained) [14]	
	Closed-set	Open-set†	Closed-set	Open-set†
	Rank-1	Rank-1 @ 1% FAR	Rank-1	Rank-1 @ 1% FAR
	P: 642, G: 2213	P: 3290, G: 2213	P: 611, G: 2234	P: 2849, G: 2234
COTS [14]	91.74	91.58	53.35	16.20
FaceNet [5] (w/o FAM)	38.16	36.76	16.53	16.04
FaceNet (with FAM)	**55.30**	**53.58**	**21.44**	**19.96**
CosFace [4] (w/o FAM)	91.12	90.81	60.72	22.91
CosFace (with FAM)	**94.24**	**94.24**	**66.12**	**25.04**
CosFace (Image Aging)	**93.18**	**92.47**	**64.87**	**23.40**

† A probe is first claimed to be present in the gallery. We accept or reject this claim based on a pre-determined threshold @ 1.0% FAR (verification). If the probe is accepted, the ranked list of gallery images which match the probe with similarity scores above the threshold are returned as the candidate list (identification).

TABLE IV: Face recognition performance on FG-NET and CACD-VS.

Method	FG-NET [15]	CACD-VS [15]
	Rank-1 (%)	Accuracy (%)
HFA [17]	69.00%	84.40%
LF-CNN [22]	88.10%	98.50%
AIM [25]	93.20%	99.38%
Wang *et al.* [26]	94.50%	99.40%
COTS	93.61%	99.32%
CosFace [4] (w/o FAM)	94.91%	99.50%
CosFace (finetuned on children)	93.71%	96.78%
CosFace (with FAM)	**95.91%**	**99.58%**

and benchmark our performance on publicly available aging datasets, FG-NET [15] and CACD-VS [16][10] in Table IV. We follow the standard protocols [17], [20] for the two datasets and benchmark our proposed FAM against baselines. We find that the feature aging module enhances cross-age performance of CosFace [4]. We also fine-tuned the last layer of CosFace on the same training set as ours, however, the decrease in accuracy suggests that moving to a new latent space can negatively impact general face recognition performance. Our module can boost the performance while still operating in the same feature space as the original matcher. While the two datasets do contain children, the protocol does not explicitly test large time gap between probe and gallery images of children. Therefore, we also evaluated the performance of the proposed approach on a subset children in FG-NET with the true mate and the probe being separated by at least 10 years. The proposed method was able to boost the performance of Cosface from 12.74% to 15.09%. Fig. 8 shows examples where the proposed approach is able to retrieve the true mate correctly at Rank 1.

B. Qualitative Results

To demonstrate the benefit of synthesized aged images, we show aged images of 4 children in the ITWCC and CFA datasets in Figure 5. Unlike prior generative methods [30]–[33], [45], [46], we can synthesize an age-progressed or age-regressed image to *any* desired age without adversely affecting the identity information. The proposed Feature Aging Module significantly contributes to a continuous aging process in the image space where other covariates such as poses, quality, etc. remain unchanged from the input probes.

[10]Since CACD-VS does not have age labels, we use DEX [44] (a publicly available age estimator) to estimate the ages.

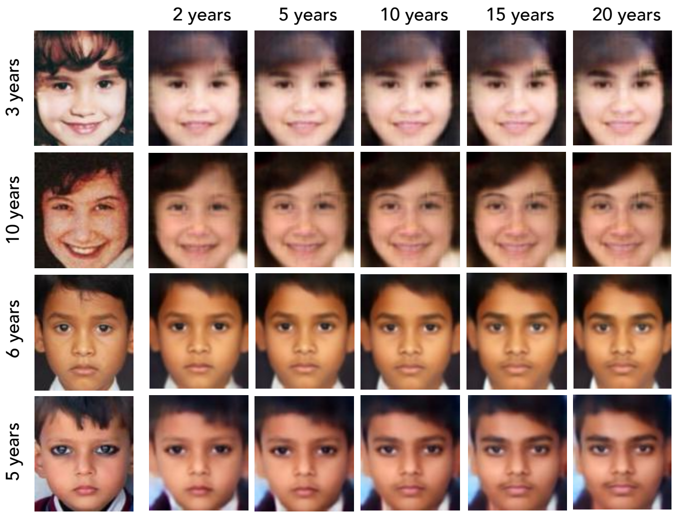


Fig. 5: Column 1 shows the probe images of 4 children. Rows 1 and 2 consist of two child celebrities from the ITWCC dataset [36] and rows 3 and 4 are children from the CFA dataset. Columns 2-6 show the corresponding synthesized aged images via the proposed Image Generator. Our approach can output an aged image at any desired target age while preserving the identity.

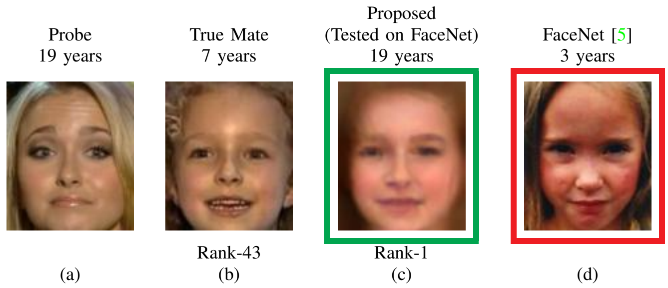
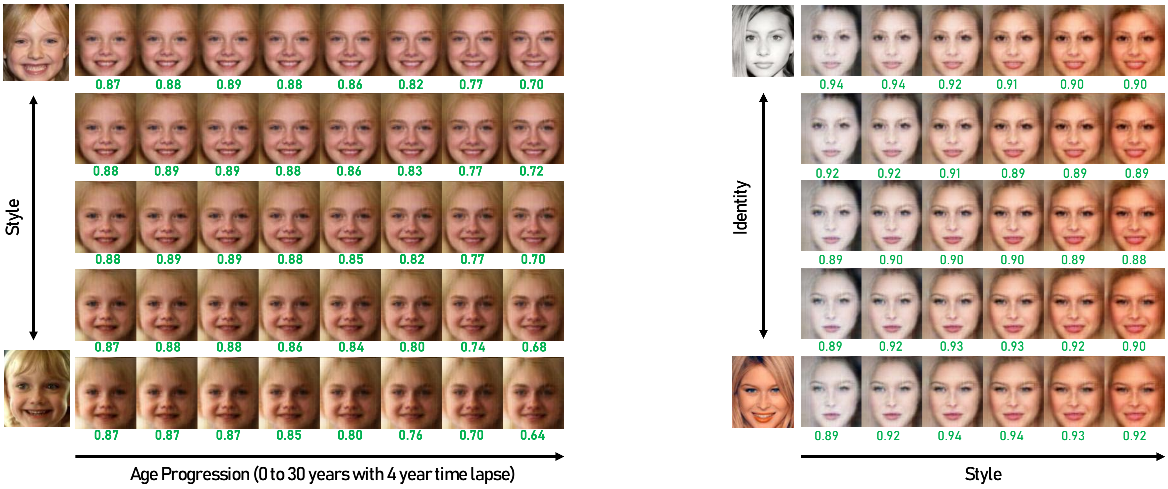


Fig. 6: (a) A probe image whose true mate is shown in (b). Without our aging module, FaceNet retrieves the true mate at Rank-43 while image in (c) is retrieved at Rank-1. (d) With our aged image *trained on CosFace*, FaceNet can retrieve the true mate at Rank-1.

C. Generalization Study

Prior studies in the adversarial machine learning domain have found that deep convolutional neural networks are highly transferable, that is, different networks extract similar deep features. Since our proposed feature aging module learns to age features in the latent space of one matcher, directly utilizing the aged features for matching via another matcher



(a) (b)

Fig. 7: Interpolation between style, age, and identity. (a) Aged features extracted via proposed FAM do not alter other covariates (quality and pose), while the style transitions due to the style-encoder. (b) Gradual transitioning between style and identity indicates disentangled identity and style features.

may not work. However, since all matchers extract features from face images, we tested the cross-age face recognition performance on FaceNet when our module is trained via CosFace. On the ITWCC dataset [14], FaceNet originally achieves 16.53% Rank-1 accuracy, whereas when we test the aged images, FaceNet achieves **21.11%** Rank-1 identification rate. We show an example where the proposed scheme trained on CosFace can aid FaceNet's cross-age face recognition performance in Figure 6.

D. Discussion

Given our framework, we can obtain the style vector from the style-encoder and the identity feature from the identity encoder. In Figure 7(a) we interpolate between the styles of two images both belonging to the same child acquired at age 7 along the y-axis. Along the x-axis, we synthesize the aged images by passing the interpolated style feature and the aged feature from the Feature Aging Module. We also provide the cosine similarity scores obtained from CosFace when the aged images are compared to the youngest synthesized image. We show that throughout the aging process, both the proposed FAM and the Image Generator can preserve the identity of the child. In addition, it can be seen that the FAM module does not affect other covariates such as pose or quality. In Figure 7(b), we take two images of two different individuals with vastly different styles. We then interpolate between the styles along the x-axis and the identity along the y-axis. We see a smooth transition between style and identity which verifies that the proposed style-encoder and decoder are effective in disentangling identity and non-identity (style) features.

E. Two Case Studies of Missing Children

Carlina White was abducted from the Harlem hospital center in New York City when she was just 19 days old. She was reunited with her parents 23 years later when she saw a photo resembling her as a baby on the National Center for Missing

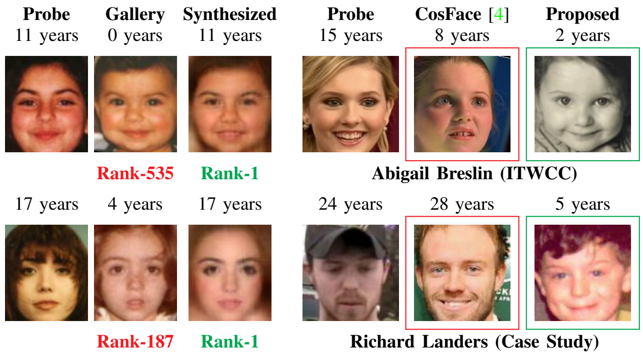


Fig. 8: Identities originally incorrectly retrieved at Rank-1 by CosFace [4] (highlighted in red). With the proposed method, CosFace can correctly retrieve the true mates in the gallery at rank-1.

and Exploited Children website[11]. We constructed a gallery of missing children consisting of $12,873$ face images in the age range 0 - 32 years from the UTKFace [30] dataset and Carlina's image as an infant when she went missing (19 days old). Her face image when she was later found (23 years old) was used as the probe. State-of-the-art face matchers, CosFace [4] and COTS, were able to retrieve probe's true mate at ranks $3,069$ and $1,242$, respectively. It is infeasible for a human operator to look through such a large number of retrieved images to ascertain the true mate. With the proposed FAM, CosFace is able to retrieve the true mate at **Rank-268**, which is a significant improvement in narrowing down the search.

In another missing child case, Richard Wayne Landers was abducted by his grandparents at age 5 in July 1994 in Indiana. In 2013, investigators identified Richard (then 24 years old) through a Social Security database search (see Figure 8). Akin to Carlina's case, adding Richard's 5 year old face image in the gallery and keeping his face image at age 24 as the probe, CosFace [4] was able to retrieve his younger image at Rank-

[11] http://www.missingkids.org

23. With the proposed feature aging, CosFace was able to retrieve his younger image at **Rank-1**.

These examples show the applicability of our feature aging module to real world missing children cases. By improving the search accuracy of any face matcher in children-to-adult matching, our model makes a significant contribution to the social good by reuniting missing children with their loved ones.

VII. Conclusion

We propose a new method that can age deep face features in order to enhance the cross-age face recognition performance in identifying missing children. The proposed method also guides the aging process in the image space such that the synthesized aged images can boost cross-age face recognition accuracy of any commodity face matcher. The proposed approach enhances the rank-1 open-set identification accuracies of FaceNet from 16.04% to 19.96% and CosFace from 22.91% to 25.04% on a child celebrity dataset, namely, ITWCC. In addition, the proposed method improves the rank-1 accuracy of CosFace on a public aging face dataset, FG-NET, from 94.91% to 95.91%. These results suggest that the proposed aging scheme can enhance the ability of commodity face matchers to locate and identify young children who are lost at a young age in order to reunite them back with their families. We plan to extend our work to unconstrained child face images which is typical in child trafficking cases.

References

[1] UN General Assembly, "Convention on the Rights of the Child," https://bit.ly/2Cl6pTh, 1989.
[2] "Children account for nearly one-third of identified trafficking victims globally," https://uni.cf/2OqsMIt, 2018.
[3] "A life revealed," https://on.natgeo.com/3iOzEyD, 2002.
[4] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," in *CVPR*, 2018.
[5] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015, pp. 815–823.
[6] P. J. Grother, G. W. Quinn, and P. J. Phillips, "Report on the evaluation of 2d still-image face recognition algorithms," *NIST Interagency Report*, vol. 7709, p. 106, 2010.
[7] S. R. Coleman and R. Grover, "The Anatomy of the Aging Face: Volume Loss and Changes in 3-Dimensional Topography," *Aesthetic Surgery Journal*, vol. 26, pp. S4–S9, 2006.
[8] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *CVPR*, 2006.
[9] B. Klare and A. K. Jain, "Face recognition across time lapse: On learning feature subspaces," in *IEEE IJCB*, 2011.
[10] D. Deb, L. Best-Rowden, and A. K. Jain, "Face recognition performance under aging," in *CVPRW*, 2017.
[11] P. J. Grother, M. Ngan, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT), Part 2: Identification," *NIST Interagency Report*, 2018.
[12] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in *IEEE ICB*, 2018.
[13] G. Davis, http://lionmovie.com/, 2016.
[14] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, and M. A. King, "Face Recognition Algorithm Bias: Performance Differences on Images of Children and Adults," in *CVPR Workshops*, 2019.
[15] "FG-NET dataset," https://bit.ly/2ZfrIyu, 2014.
[16] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *ECCV*, 2014.
[17] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang, "Hidden factor analysis for age invariant face recognition," in *CVPR*, 2013.
[18] D. Gong, Z. Li, D. Tao, J. Liu, and X. Li, "A maximum entropy feature descriptor for age invariant face recognition," in *CVPR*, 2015.
[19] Z. Li, D. Gong, X. Li, and D. Tao, "Aging face recognition: A hierarchical learning model based on local patterns selection," *IEEE TIP*, vol. 25, no. 5, pp. 2146–2154, 2016.
[20] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE TIFS*, vol. 6, no. 3, pp. 1028–1037, 2011.
[21] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE TIFS*, vol. 5, no. 1, pp. 82–91, 2009.
[22] C. Nhan Duong, K. Gia Quach, K. Luu, N. Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *ICCV*, 2017.
[23] T. Zheng, W. Deng, and J. Hu, "Age estimation guided convolutional neural network for age-invariant face recognition," in *CVPRW*, 2017.
[24] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, 2017.
[25] J. Zhao, Y. Cheng, Y. Cheng, Y. Yang, F. Zhao, J. Li, H. Liu, S. Yan, and J. Feng, "Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition," in *AAAI*, 2019.
[26] H. Wang, D. Gong, Z. Li, and W. Liu, "Decorrelated adversarial learning for age-invariant face recognition," in *CVPR*, 2019.
[27] M. Q. Hill, C. J. Parde, C. D. Castillo, Y. I. Colon, R. Ranjan, J.-C. Chen, V. Blanz, and A. J. O'Toole, "Deep convolutional neural networks in the face of caricature: Identity and image revealed," *arXiv preprint arXiv:1812.10902*, 2018.
[28] P. Dhar, A. Bansal, C. D. Castillo, J. Gleason, P. J. Phillips, and R. Chellappa, "How are attributes expressed in face DCNNs?" *arXiv preprint arXiv:1910.05657*, 2019.
[29] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent Face Aging," in *CVPR*, 2016.
[30] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *CVPR*, 2017.
[31] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," in *CVPR*, 2018.
[32] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *IEEE ICIP*, 2017.
[33] Z. Zhai and J. Zhai, "Identity-preserving conditional generative adversarial network," in *IEEE IJCNN*, 2018.
[34] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010.
[35] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *IEEE FG*, 2006.
[36] K. Ricanek, S. Bhardwaj, and M. Sodomsky, "A review of face recognition against longitudinal child faces," in *IEEE BIOSIG*, 2015.
[37] L. Best-Rowden and A. K. Jain, "Longitudinal study of automatic face recognition," *IEEE T-PAMI*, vol. 40, no. 1, pp. 148–162, 2017.
[38] S. Yoon and A. K. Jain, "Longitudinal study of fingerprint recognition," *PNAS*, vol. 112, no. 28, pp. 8555–8560, 2015.
[39] L. Best-Rowden, Y. Hoole, and A. Jain, "Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation," in *IEEE BIOSIG*, 2016, pp. 1–8.
[40] P. Upchurch, J. Gardner, G. Pleiss, R. Pless, N. Snavely, K. Bala, and K. Weinberger, "Deep feature interpolation for image content changes," in *CVPR*, 2017.
[41] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE FG*, 2018.
[42] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017.
[43] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019.
[44] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *ICCV Workshop*, 2015.
[45] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE T-PAMI*, vol. 29, no. 12, pp. 2234–2240, 2007.
[46] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE T-PAMI*, vol. 24, no. 4, pp. 442–455, 2002.