Decorrelated Adversarial Learning for Age-Invariant Face Recognition

Hao Wang, Dihong Gong, Zhifeng Li, Wei Liu
Tencent AI Lab
hawelwang@tencent.com, gongdihong@gmail.com, michaelzfli@tencent.com, wl2223@columbia.edu

Abstract

There has been an increasing research interest in age-invariant face recognition. However, matching faces with big age gaps remains a challenging problem, primarily due to the significant discrepancy of face appearance caused by aging. To reduce such discrepancy, in this paper we present a novel algorithm to remove age-related components from features mixed with both identity and age information. Specifically, we factorize a mixed face feature into two uncorrelated components: identity-dependent component and age-dependent component, where the identity-dependent component contains information that is useful for face recognition. To implement this idea, we propose the Decorrelated Adversarial Learning (DAL) algorithm, where a Canonical Mapping Module (CMM) is introduced to find maximum correlation of the paired features generated by the backbone network, while the backbone network and the factorization module are trained to generate features reducing the correlation. Thus, the proposed model learns the decomposed features of age and identity whose correlation is significantly reduced. Simultaneously, the identity-dependent feature and the age-dependent feature are supervised by ID and age preserving signals respectively to ensure they contain the correct information. Extensive experiments have been conducted on the popular public-domain face aging datasets (FG-NET, MORPH Album 2, and CACD-VS) to demonstrate the effectiveness of the proposed approach.

1. Introduction

With the impressive advancement driven by deep learning [22, 36, 17, 47], current face recognition methods[37, 39, 38, 46, 42, 40, 28, 10] have achieved excellent performance. Many of these models are even more accurate than humans in various scenarios. However, identifying faces across a wide range of ages remains under-exploring.

Recently, modern advances [41, 46, 28, 42, 40, 10] introduce the margin-based metrics and normalization mechanism to train the models in order to improve the face recognition performance. However, most of these methods usually lack the discriminating power for face identification in the scenario of Age Invariant Face Recognition (AIFR). The crucial challenge for AIFR is subject to the significant discrepancy resulting from the aging process. Figure 1 shows an example that face images have great variations within the same identity across different ages, while those of different identities share similar age-related information. As a result, those faces with big age gaps serve as hard examples that the current face recognition systems cannot identify correctly. In particular, the intra-identity distance is increasing larger if there are more faces of the child and the elderly.

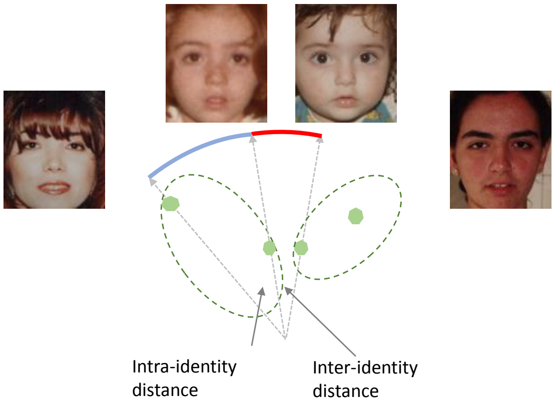

Figure 1. We show a typical example for AIFR, where the intra-identity distance is greater than the inter-identity distance due to the large age variations. As a result, many current face recognition systems fail to identify faces across big age gaps.

In the meanwhile, increasing research attentions have been attracted to the age-invariant face recognition (AIFR). Recent research studies on AIFR mainly focus on the design of either generative models or discriminative models. The generative methods [12, 23, 32] propose to synthesize face images of different ages to assist the face recognition. Very recently, several studies [53, 2, 11] aim at improving the quality of generated aging faces by utilizing the powerful GAN-based models. However, accurately modeling the aging process is difficult and complicated. The unstable ar-

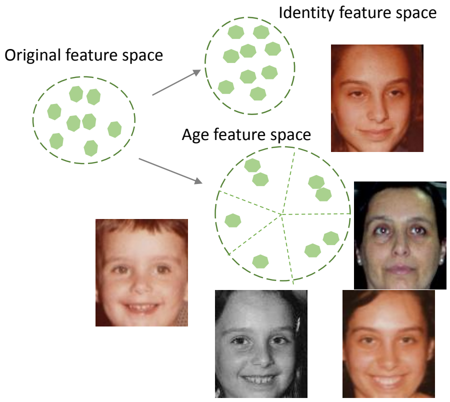

Figure 2. The face features are decomposed into the identity-dependent component and the age-dependent component. Only the identity features participate the testing of face recognition.

tifacts in the synthesized faces can significantly affect the performance of face recognition. In contrast, discriminative methods draw increasing interest in recent studies. For example , the [13] separates the identity-related information and the age-related information through the hidden factor analysis (HFA). The [45] is based on similar analysis and extends the HFA to the deep learning framework. More recently, the OE-CNN [43] presents the orthogonal feature decomposition to solve the AIFR. According to all these studies, feature decomposition plays a key role in invariant feature learning under the assumption that facial information can be perfectly modeled by the decomposed components. However, the decomposed components practically have latent relationship with each other and the identity-dependent component may still contain age information.

In this paper, we introduce a deep feature factorization learning framework that factorizes the mixed face features into two uncorrelated components: identity-dependent component ($\mathbf{x_{id}}$) and age-dependent component ($\mathbf{x_{age}}$). Figure 2 illustrates our feature factorization schema. We implement such factorization through a residual mapping module inspired by [4]. This means that, the age-dependent embeddings are encoded through a residual mapping function $\mathbf{x_{age}} = \mathcal{R}(\mathbf{x})$. We have the following formulation: $\mathbf{x} = \mathbf{x_{id}} + \mathbf{R}(\mathbf{x})$, where $\mathbf{x}$ is the initial face feature, and $\mathbf{x_{id}}$ is the identity-dependent feature.

To reduce the mutual variations in the decomposed components, we propose a novel Decorrelated Adversarial Learning (DAL) algorithm that adversarially minimizes the correlation between $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$. Specifically, a Canonical Mapping Module is introduced to find maximum correlations between $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$, while the backbone and factorization module aims to reduce the correlation. In the meanwhile, $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ are learned by the identity and age classification signals respectively. Through the adversarial training, we wish the $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ will be sufficiently uncorrelated, and the age information in $\mathbf{x_{id}}$ can be significantly reduced.

Our major contributions are summarized as follows:

1. We propose a novel Decorrelated Adversarial Learning (DAL) algorithm based on the linear feature factorization, in order to regularize the learning of decomposed features. In this way, we wish to capture the ID-preserving while age invariant features for AIFR. To the best of our knowledge, this is the first work to introduce decorrelated adversarial feature learning to AIFR.

2. We present the Batch Canonical Correlation Analysis (BCCA), an extension of CCA in the fashion of stochastic gradient decent optimization. The proposed BCCA can be integrated to the deep neural networks for correlation regularization.

3. The proposed method has significantly improved the state-of-the-art performance on the AIFR datasets including MORPH Album2[34], FG-NET[1] and CACD-VS[5], which strongly demonstrates its effectiveness.

2. Related Work

Age-Invariant Face Feature Learning. Many prior studies[15, 24, 26, 7, 25, 27, 5, 6, 13] in the literature extracted hand-craft features with heuristic methods. For example, the [25] developed a multi-feature discriminant analysis method with local feature descriptions. The [13] proposed the hidden factor analysis (HFA) to model the feature factorization and reduce the age variations in identity-related features. The [15] introduced an effective maximum entropy feature descriptor and a robust identity matching framework for AIFR. Several recent methods [45, 55, 43] are mainly based on deep neural networks. The [45] developed the Latent Factor guided Convolutional Neural Network (LF-CNN) to improve the HFA. The [55] introduced the Age Estimation guided CNN (AE-CNN) method for AIFR. The OE-CNN [43] proposed the orthogonal embedding decomposition such that the identity information is encoded in the angular space while the age information is represented in the radial direction. Our work presents a DAL algorithm with the linear residual decomposition.

Canonical Correlation Analysis. Canonical Correlation Analysis (CCA) [18] is a well-known algorithm to measure the linear relationship between two multidimensional variables. Some previous works have introduced this method to face recognition in various scenarios. For example, the [49] proposed a 2D-3D face matching method using the CCA. The [14] developed a multi-feature CCA method for face-sketch recognition. Compared to these typical CCA based methods, our work presents the extension of CCA to deep neural network as a regularization method for AIFR.

Adversarial Approaches. Generative adversarial networks (GAN) [16] have shown effective in various generative tasks, such as face aging [53, 2, 11], face super-resolution [51, 8], etc. Besides, the adversarial networks has also been explored to the improve the discriminative models. For example, the [3] utilized GAN to generate high-resolution of small faces in order to improve face detection. The [9] developed an adversarial UV completion framework (UV-GAN) to solve the pose invariant face recognition problem. The [29] proposed to learn the identity-distilled features and the identity-dispelled features in an adversarial autoencoder framework. The [54] proposed an adversarial network to generate hard triplet feature examples. In this work, we propose a decorrelated adversarial learning method to significantly minimize the correlation between the decoupled components of identity and age, thus the identity-dependent features are age invariant.

3. Method

3.1. Feature Factorization

As faces contain intrinsic identity information and age information, they can be jointly represented by the identity-dependent features and the age-dependent features. Motivated by this, we design a linear factorization module that decomposes the initial features into these two unrelated components. Formally, given an initial feature vector $\mathbf{x} \in \mathbb{R}^d$ that extracted from an input image $\mathbf{p}$ by a backbone CNN $\mathcal{F}$ (i.e, $\mathbf{x} = \mathcal{F}(\mathbf{p})$), we define the linear factorization as follows:

$$\mathbf{x} = \mathbf{x_{id}} + \mathbf{x_{age}}, \tag{1}$$

where $\mathbf{x_{id}}$ denotes the identity-dependent component, and $\mathbf{x_{age}}$ denotes the age-dependent component. We design a deep residual mapping module similar to [4] to implement this. Specifically, we obtain the age-dependent feature through a mapping function $\mathcal{R}$, and the residual part is regarded as the identity-dependent feature. We refer to this as Residual Factorization Module (RFM), which is formulated as:

$$\begin{aligned} \mathbf{x_{age}} &= \mathcal{R}\left(\mathbf{x}\right), \\ \mathbf{x_{id}} &= \mathbf{x} - \mathcal{R}\left(\mathbf{x}\right). \end{aligned} \tag{2}$$

At testing stage, only the identity-dependent features are used for face recognition. It is desirable that $\mathbf{x_{id}}$ encodes the identity information while $\mathbf{x_{age}}$ draws the age variations. We simultaneously put the identity discriminating signal and the age discriminating signal onto these two decoupled features to respectively supervise the multi-task learning of these two components. Figure 3 shows the overall framework of our work. The resnet-like backbone extracts the initial features, upon which we build the residual module for feature factorization. Based on such factorization, we propose the Decorrelated Adversarial Learning, which is introduced in the following section.

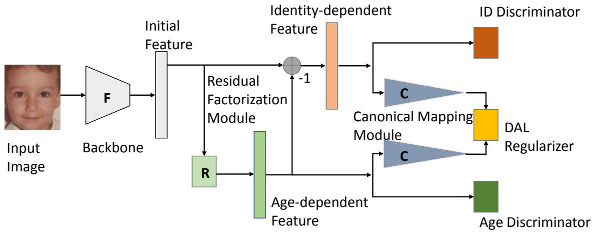

Figure 3. An overview of the proposed method. The initial features are extracted by backbone net, followed by the residual factorization module. The two factorized components x_{id} and x_{age} are then used for classification and DAL regularization.

3.2. Decorrelated Adversarial Learning

Through feature factorization, it is crucial for AIFR that the $\mathbf{x_{id}}$ should be identity preserving and necessarily age-invariant. Unfortunately, the $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ practically have latent relationship with each other. For example, $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ may have high linear correlation with each other. Thus, the $\mathbf{x_{id}}$ may partially involve the age variation, which leads to negative effect on face recognition. On the other hand, the $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ should be mutually uncorrelated to force the non-trivial learning such that they both improve themselves.

To this end, we design a regularization algorithm that is helpful to reduce the correlation between the decomposed features, namely Decorrelated Adversarial Learning (DAL). The DAL basically calculates the canonical correlation between the paired features of the decomposed components.

Formally, given paired features $\mathbf{x_{id}}, \mathbf{x_{age}}$, we design a linear Canonical Mapping Module (CMM) that maps $\mathbf{x_{id}}, \mathbf{x_{age}}$ to the canonical variables $\mathbf{v_{id}}, \mathbf{v_{age}}$:

$$\forall t \in \{id, age\} : \mathbf{v_t} = \mathcal{C}(\mathbf{x_t}) = \mathbf{w_t^T} \mathbf{x_t}, \tag{3}$$

where the $\mathbf{w_{id}}, \mathbf{w_{age}}$ are the learning parameters for canonical mapping. After that, we define the canonical correlation as:

$$\rho = \frac{\mathbf{Cov}(\mathbf{v_{id}}, \mathbf{v_{age}})}{\sqrt{\mathbf{Var}(\mathbf{v_{id}})\mathbf{Var}(\mathbf{v_{id}})}}. \tag{4}$$

Based on such definition, we first find maximum of $|\rho|$ by updating CMM with respect to $\mathbf{w_{id}}, \mathbf{w_{age}}$, and then try to reduce the correlation by training the backbone and RFM. That is, on the one hand, we freeze $\mathcal{F}, \mathcal{R}$ and train $\mathcal{C}$ in the canonical correlation maximizing process. On the other hand, we update $\mathcal{F}, \mathcal{R}$ with $\mathcal{C}$ fixed in the feature correlation minimizing process. Obviously, they compete with each other playing a two-player min-max game during the adversarial training procedure. In this way, our goal is to minimize the correlation between $\mathbf{x_{id}}, \mathbf{x_{age}}$ by always decreasing their maximum canonical correlation. In other words, the optimal feature projections having maximum correlation act as the primary target to be decorrelated. Thus, $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ learns continuously to have small correlation and finally they are significantly uncorrelated.

Overall, the objective function for DAL is formulated as:

$$\mathcal{L}_{DAL} = \min_{\mathcal{F},\mathcal{R}} \max_{\mathcal{C}} (|\rho(\mathcal{C}(\mathcal{F}(\mathbf{p}) - \mathcal{R}\ (\mathcal{F}(\mathbf{p})), \mathcal{C}(\mathcal{R}\ (\mathcal{F}(\mathbf{p})))|) \tag{5}$$

We believe the strong decorrelation enhanced by DAL will encourage the $\mathbf{x_{id}}$ and $\mathbf{x_{age}}$ to be sufficiently invariant with each other. Importantly, this will improve robustness of x_{id} for age-invariant face recognition.

3.3. Batch Canonical Correlation Analysis

In contrast to the typical canonical correlation analysis (CCA) methods, our work introduces the canonical correlation Analysis (BCCA) based on stochastic gradient decent (SGD) optimization. Since the correlation statistics on the entire dataset is practically impossible, we follow similar strategy of batch normalization [20] to compute the correlation statistics based on mini-batches. Thus, it naturally suits the deep learning framework.

Given a mini-batch size of m, we have two sets of the decomposed features: $B_{id} = \{x_{id}^{1,\ldots,m}\}$ and $B_{age} = \{x_{age}^{1,\ldots,m}\}$. Thus, the canonical correlation can be written as:

$$\rho = \frac{\mathbf{\frac{1}{m}\Sigma_{i=1}^{m}(v_{id}^{i} - \mu_{id})(v_{age}^{i} - \mu_{age})}}{\sqrt{\sigma_{id}^2 + \epsilon}\sqrt{\sigma_{age}^2 + \epsilon}}. \tag{6}$$

Here, the $\mu_{\mathbf{id}}$ and σ_{id}^2 are the mean and variance of $\mathbf{v_{id}}$ respectively, and similarly for μ_{age} and σ_{age}^2. The ϵ is a constant parameter for numerical stability.

Equation 6 serves as the objective function for BCCA and we leverage the SGD based algorithm to optimize it. Note that the canonical correlation $|\rho|$ is demanded to be necessarily maximized when updating the $\mathcal{C}$., while being minimized when training the $\mathcal{F}, \mathcal{R}$. The derivation of gradients are:

$$\frac{\partial \rho}{\partial \mathbf{v_{id}^{i}}} = \frac{1}{m}(\frac{\mathbf{v_{age}^{i}} - \mu_{\mathbf{age}}}{\sqrt{\sigma_{id}^2 + \epsilon}\sqrt{\sigma_{age}^2 + \epsilon}} - \frac{(\mathbf{v_{id}^{i}} - \mu_{\mathbf{id}}) \cdot \rho}{\sigma_{\mathbf{id}}^{\mathbf{2}} + \epsilon}),$$

$$\frac{\partial \rho}{\partial \mathbf{v_{age}^{i}}} = \frac{1}{m}(\frac{\mathbf{v_{id}^{i}} - \mu_{\mathbf{id}}}{\sqrt{\sigma_{id}^2 + \epsilon}\sqrt{\sigma_{age}^2 + \epsilon}} - \frac{(\mathbf{v_{age}^{i}} - \mu_{\mathbf{age}}) \cdot \rho}{\sigma_{\mathbf{age}}^{\mathbf{2}} + \epsilon}). \tag{7}$$

Thus, the optimization consists of a forward propagation that outputs the ρ, and a backward propagation that calculate the gradients for updating. The detailed learning algorithm of BCCA is described in Algorithm 1.

Algorithm 1 Learning algorithm of BCCA for each iteration.

```
Input: $B_{id} = \{x_{id}^{1,\ldots,m}\}$; $B_{age} = \{x_{age}^{1,\ldots,m}\}$;
Output: the canonical correlation $\rho$ for forward pass; the
    gradients for backward pass.
for each $t \in \{id, age\}$ do
    CMM forward: $v_t^i = w_t^T x_t^i$ for $i = 1 \ldots m$;
    Compute means: $\mu_t = \frac{1}{m}\Sigma_{i=1}^{m} v_t^i$;
    Compute variances: $\sigma_t^2 = \frac{1}{m}\Sigma_{i=1}^{m}(v_t^i - \mu_t)^2$;
end for
Forward propagation: Compute $\rho$ with Equation 6.
for each $t \in \{id, age\}$ do
    Compute $\frac{\partial \rho}{\partial v_t}$ with Equation 7;
    CMM backward: $\frac{\partial L}{\partial x_t^i} = w_t^i \frac{\partial L}{\partial v_t^i}$; for $i = 1 \ldots m$;
    CMM backward: $\frac{\partial L}{\partial w_t^i} = x_t^i \frac{\partial L}{\partial v_t^i}$; for $i = 1 \ldots m$;
end for
```

3.4. Multi-task Training

In this section, we describe the multi-task training strategy to supervise the learning of the decomposed features. As shown in Figure 3, there are three basic supervision modules: age discriminator, identity discriminator and DAL regularizer.

Age Discriminator. For the learning of age information, we feed $\mathbf{x_{age}}$ into an age discriminator to ensure the age discriminating information. Since age labels are rough with uncertain noises in practice, we follow [13, 45] and perform classifications on ages by dividing them into different groups. We use the softmax layer with cross-entropy loss for the age classification.

Identity Discriminator. Following the recent [42, 40], we utilize the CosFace loss to supervise the learning of $\mathbf{x_{id}}$ and ensure the identity-preserving information. The CosFace loss is formulated as:

$$\mathcal{L}_{ID} = \frac{1}{N}\sum_{i} -\log \frac{e^{s(\cos(\theta_{\mathbf{y_i},\mathbf{i}}) - m)}}{e^{s(\cos(\theta_{\mathbf{y_i},\mathbf{i}}) - m)} + \sum_{j \neq y_i} e^{s\cos(\theta_{\mathbf{j},\mathbf{i}})}}, \tag{8}$$

where N is the number of identities, y_i is the corresponding identity label, $\mathbf{cos}(\theta_{\mathbf{j}}, \mathbf{i}) = \frac{\mathbf{W_j^T}}{\|\mathbf{W_j}\|} \cdot \frac{\mathbf{x_{id}^i}}{\|\mathbf{x_{id}^i}\|}$ is the cosine of angle between the i-th feature $\mathbf{x_{id}^i}$ and the j-th weight vector $\mathbf{W_j}$ of the classifier. The m a constant margin term controlling the cosine margin and the s is a constant scaling factor s. The CosFace loss aims to introduce much strict constraints to the identity classification such that the learned features are encouraged to be separated by a margin between different identities. A properly large m will encourage powerful discriminating information in the learned features for face recognition.

DAL Regularizer. The proposed DAL regularization also participants the joint supervision to guide the feature learning such that the correlations between the paired decomposed features can be significantly reduced. Through the joint supervision, the model simultaneously learns to encourage both the discriminating power of x_{id}, x_{age}, and decorrelation information between of these two decomposed components.

In summary, the training is supervised by the following combined multi-task loss:

$$\mathcal{L} = \mathcal{L}_{ID}(\mathbf{x_{id}}) + \lambda_1 \mathcal{L}_{SM}(\mathbf{x_{age}}) + \lambda_2 \mathcal{L}_{DAL}(\mathbf{x_{id}}, \mathbf{x_{age}}), \tag{9}$$

where L_{ID} denotes the CosFace loss, L_{SM} denotes the softmax with cross-entropy loss, λ_1 and λ_2 are scalar hyper-parameters to balance these three losses. In the testing phase, we extract the identity-dependent features x_{id} for AIFR evaluations.

3.5. Discussion

The proposed method has the following advantages. First, the DAL regularization on features is helpful to encourage the uncorrelated and co-invariant information between the decomposed components. Related works such as HFA[13], LF-CNN[45] and OE-CNN[43] have neglected the underlying correlation. Instead, we aim to minimize the classification error as well as the correlation effect simultaneously. Second, the BCCA provides an extension of CCA that is inserted to the deep learning framework such that the entire model can be trained in an end-to-end process. Finally, our method can be easily generalized to other components factorization model, such as pose, illumination, emotion, etc. To the best of our knowledge, we are the first to develop the decorrelated adversarial regularization framework to AIFR.

4. Experiments

4.1. Implementation Details

Network Architecture. (1) Backbone: our backbone network is a 64-layer CNN similar to [43] . It consists of 4 stages with respectively 3, 4, 10, 3 stacked residual blocks. Every residual block has 3 stacked units of “3x3 Conv + BN + ReLU”. Finally, a FC layer outputs the initial face features of 512 dimension. (2)Residual factorization Module (RFM): the initial face features are mapped to form the age-dependent feature through 2 “FC +ReLU”, and the residual part is regarded as the identity-dependent feature. (3) Age discriminator: we stack 3 “FC +ReLU” upon $\mathbf{x_{age}}$, and perform age classification. (4) Identity discriminator: we directly use $\mathbf{x_{id}}$ for identification by CosFace loss. (5) DAL regularizer: we feed the $\mathbf{x_{age}}$ and $\mathbf{x_{id}}$ into the FC layers respectively and output their linear combinations, which are then used for the BCCA calculation and optimization.

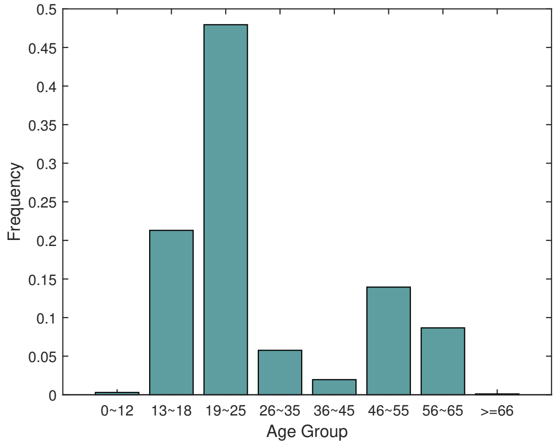

Figure 4. The age distribution of our small training dataset. It contains 0.5M face images covering large age variations.

Data Preprocessing. We use MTCNN [52] to detect face areas and facial landmarks on both the training and testing sets. Then, similarity transformation is performed according to the 5 facial key points (two eyes, nose and two mouth corners) in order to crop the face patch to 112×96 . Finally, each pixel ([0,255]) of the cropped face patch is normalized by subtracting 127.5 then divided by 128.

Training Details. Our training data includes the Cross-Age Face (CAF) dataset provided by [43] and other common face datasets such as CASIA-WebFace [50], VGG Face [33] and celebrity+ [30]. It totally contains about 1.7M images from 19.9k individuals, which is similar to [43]. Meanwhile, we build a subset containing about 0.5M images from 12k individuals following [43] in order to conduct fair experimental comparisons. We refer to this subset as *small training dataset* and our whole training dataset as *large training dataset* for clarify. We adopt the pre-trained age estimation model [35] to generate predicted age labels for the face images of the entire training set. Note that only those predicted ages with relatively high confidence (i.e. more likely to be true label) are considered valid and will participate the age-classification. After that, the predicted ages are divided into 8 groups: 0-12, 13-18, 19-25, 26-35, 36-45, 46-55, 56-65, $\geq$ 66. The grouped age labels are then used for the age-classification training. The joint supervision in Equation 9 guides the DALtraining process in an adversarial manner. More specifically, in an adversarial loop, we alternately run the canonical correlation maximizing process for 20 iterations and then change to feature correlation minimizing process for 50 iterations. The empirically setting of hyper-parameters λ_1 and λ_2 in Equation 9 are: $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $m = 0.35$, $s = 64$. All our experiment models are trained through stochastic gradient descent (SGD), with batch size of 512. The whole training procedure is about 40-th epochs and the learning rate is

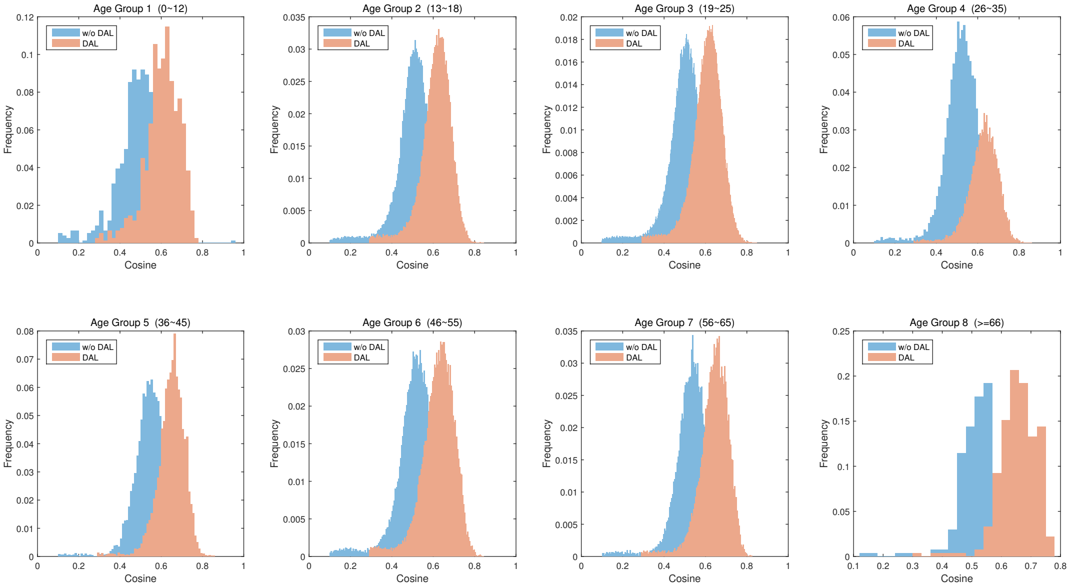

Figure 5. The distribution of the cosine similarity between features and their class center at different age groups. Our DAL model consistently increases the cosine similarity compared against the baseline model without DAL across all the age groups, which demonstrates the effectiveness of our method to encourage less intra-identity variations. Best viewed in colors.

Model	FG-NET (MF1)	FG-NET (MF2)	FG-NET (leave-one-out)	MORPH Album 2	CACD -VS
Baseline	55.86%	58.85%	93.4%	98.21%	99.07%
+Age	55.84%	58.64%	93.6%	98.11%	99.05%
+Age+DAL	**57.92%**	**60.01%**	**94.5%**	**98.93%**	**99.40%**

Table 1. Comparison of our method against the baseline models. The evaluation results are rank-1 face identification rate on FG-NET, under protocols of MF1, MF2 and leave-one-out.

initially set to 0.1 and reduced by a factor of 0.1 at 22-th, 33-th, 38-th epoch.

Testing Details. We conduct evaluation experiments on the well-known public AIFR face datasets: FG-NET[1], MORPH Album 2[34] and CACD-VS[5]. In the testing process, we extract the identity-dependent features and concatenate features of the original image and the flipped image to form the final representation. The cosine similarity of these representations are then used to conduct face verification and identification.

4.2. Ablation Study

In this subsection, we study the different variants of the proposed models to show the effectiveness of our method.

Visualization of Cosine Similarity. For a better understanding of the DAL and its ability to improve the identity-preserving information, we conduct an experiment to visualize the cosine similarities across different age groups. Given the learned identity-dependent features $\mathbf{x_{id}}$, we first calculate their class centers by clustering every identity in the identity feature space, and then compute the cosine similarity between each sample and its class center. After that, we plot the distribution of cosine similarity across different age groups. In this study, we conduct such visualization analysis on the small training dataset which contains 0.5M face images covering various age differences. Figure 4 shows the age distribution of this dataset. We present a comparison between the "w/o DAL" model (trained by the joint supervision signals of age and identity but without DAL) and our proposed DAL model. As shown in Figure 5, compared against the "w/o DAL" model, the DAL model consistently increases the cosine similarity between $\mathbf{x_{id}}$ and its class center across all the age groups. This observation proves that our method encourages features to have small intra-identity variations and thus the samples of the same identity but different ages are pulled together in the feature space. Thus, the discriminating power of the learned identity features can be effectively improved by the proposed DAL method.

Quantitative Evaluation. To show the impact of the joint learning framework with our proposed DAL method, we conduct the ablative evaluations on several public AIFR datasets including FG-NET, MORPH Album 2 and CACD-VS. Moreover, we also test our models on FG-NET following the protocols of Megaface challenge 1 (MF1) [21] and Megaface challenge 2 (MF2) [31]. Both the MF1 and the

Method	#Test Subjects	Rank-1
HFA [13]	10,000	91.14%
CARC [5]	10,000	92.80%
MEFA [15]	10,000	93.80%
MEFA+SIFT+MLBP [15]	10,000	94.59%
LPS+HFA [24]	10,000	94.87%
LF-CNNs [45]	10,000	97.51%
OE-CNNs	10,000	98.55%
Ours	10,000	**98.93%**
GSM [26]	3,000	94.40%
AE-CNNs [55]	3,000	98.13%
OE-CNNs [43]	3,000	98.67%
Ours	3,000	**98.97%**

Table 2. Evaluation results on the MORPH Album 2 dataset.

MF2 include an additional distractor set respectively that contains 1 million face distractors, making the benchmarks much more difficult. The MF2 provides a training dataset such that all the evaluation methods should be trained on the same dataset and without any additional training data. We consider the following models for ablative comparison in this study: (1) Baseline: the baseline model is trained by the identification loss only and without any extra age supervision. (2) +Age: this model is trained by the joint supervision of the identification signal and the age classification signal. (3) +Age+DAL: our proposed model that is trained simultaneously by the DAL regularization and the joint supervision signals. As reported in Table 1, without DAL the joint supervision model achieves comparable results with the baseline model. On the contrary, our "+Age+DAL" model improves the performance of FG-NET on all the schemes. The improvement on FG-NET with the scheme of MF2 is relatively limited compared with that of MF1 and 'leave-one-out', mainly due to the less aging variations of MF2 training dataset. Nevertheless, the consistently performance improvement demonstrates the effectiveness of our method. Moreover, our method improves the baseline models by more than 0.7% on MORPH Album 2, and more than 0.3% on CACD-VS, which are remarkable improvements at the high accuracy level above 98% and 99%.

4.3. Experiments on the MORPH Album 2 Dataset

The MORPH Album 2 dataset consists of 78,000 face images of 20,000 individuals across different ages. For fair comparison, we follows [43] and conduct evaluations under two benchmark schemes where the testing set consists of 10,000 subjects and 3,000 subjects respectively. In the testing sets, two face images of each subjects with the largest age gaps are selected to compose the probe set and the gallery set. We train the model with our proposed DAL on the large training dataset(1.7M images). Note that we have not conducted any training or finetuning on the MORPH Album 2.

Method	Acc.	AUC.
High-Dimensional LBP [7]	81.6%	88.8%
HFA [13]	84.4%	91.7%
CARC [5]	87.6%	94.2%
LF-CNNs [45]	98.5%	99.3%
Human, Average [6]	85.7%	94.6%
Human, Voting [6]	94.2%	99.0%
Softmax	98.4%	99.4%
A-Softmax	98.7%	99.5%
OE-CNNs [43]	99.2%	99.5%
Ours	**99.4%**	**99.6%**

Table 3. Evaluation results on the CACD-VS dataset.

Method	Rank-1
Park et al. [32] (2010)	37.4%
Li et al. [25] (2011)	47.5%
HFA [13] (2013)	69.0%
MEFA [15] (2015)	76.2%
CAN [48]	86.5%
LFCNNs [11]	88.1%
Ours	**94.5%**

Table 4. Evaluation results on the FG-NET dataset under the protocol of leave-one-out.

In this experiment, we compare our DAL model against the recently AIFR algorithms in the literature. As shown in Table 2, the proposed method has effectively improved the rank-1 identification performance. Particularly, our method outperforms the recent top-performing AIFR methods by a clear margin, setting new state-of-the-art on the MORPH Album 2 database.

4.4. Experiments on the CACD-VS Dataset

As a public released dataset for AIFR, the CACD dataset is composed of 163,446 images from 2,000 celebrities with age variations. The collected face images also include different illumination, various poses and makeup. The subset CACD-VS consists of 4000 face image pairs for face verification, and the face pairs are divided into 2,000 positive pairs and 2,000 negative pairs. In our experiment, we strictly follow [5, 43] to perform the 10-fold cross-validation for fair comparisons. We use the same trained models in Sec 4.3 to evaluate the performance on the CACD-VS Dataset. Table 3 shows the verification accuracy of our models compared against the other state-of-the-art AIFR methods. Not surprisingly, the proposed DAL model obtains consistent improvement over the prior methods, demonstrating the superiority of our method again.

Method	Protocol	Rank-1
FUDAN-CS_SDS [44]	Small	25.56%
SphereFace [28]	Small	47.55%
TNVP [11]	Small	47.72%
Softmax	Small	35.11%
A-Softmax	Small	46.77%
OE-CNNs [43]	Small	52.67%
Ours	small	**57.92%**

Table 5. Evaluation results on the FG-NET dataset under the protocol of MF1.

Method	Protocol	Rank-1
GRCCV	Large	21.04%
NEC	Large	29.29%
3DiVi	Large	35.79%
GT-CMU-SYSU	Large	38.21%
OE-CNNs [43]	Large	53.26%
Ours	Large	**60.01%**

Table 6. Evaluation results on the FG-NET dataset under the protocol of MF2

4.5. Experiments on the FG-NET Dataset

Compared to MORPH Album 2 and CACD-VS, the FG-NET dataset is much more challenging containing a wide covering of ages from 0 to 69. It has 1002 face images from 82 individuals.The dataset includes lots of face images at the age phase of the child and the elderly. We conducted experiments under three different evaluation schemes for overall fair benchmark comparison: leave-one-out, MegaFace challenge 1 (MF1) and MegaFace challenge 2 (MF2).

Evaluation with leave-one-out. We directly use the DAL model trained on the small training set (0.5M images) and test on the FG-NET dataset. The evaluation is conducted by leave-one-out. It is noticeable that we have not used any data of FG-NET for training or finetuning. The performance comparisons are given in Table . We can see that our method has improved the priors [13] by a significant margin.

Evaluation with MF1. The MF1 [21] contains 1 million distractor images from 690K different individuals. According to [21], evaluations are conducted under the two protocols: large or small training set. The training set less than 0.5M is considered small. We strictly follow the protocol of small training set to train the model and conduct evaluations on FG-NET. The experimental results are reported in Table 5. The performance improvement over the other methods strongly demonstrates the effectiveness of the proposed DAL method.

Evaluation with MF2. We also conducte experiments on the MF2 [31], which has 1 million distractors as well. But the distractors of MF1 and MF2 are totally different. Unlike the MF1, the MF2 requires that all the models should be trained on the same training set, thus yields very fair comparisons. The training set provided by MF2 contains 4.7 million faces from 672K identities. Following this protocol, we train our models and conduct evaluations on the MF2. Table 6 shows the performance comparisons between ours and the previous methods. Again, our DAL method significantly improves the identification accuracy and set new state-of-the-art on the MF2 dataset.

Method	LFW	MF1-Facescrub
SphereFace[28]	99.42%	72.73%
CosFace[42]	99.33%	77.11%
OE-CNNs[43]	99.35%	N/A
Ours	**99.47%**	**77.58%**

Table 7. Evaluation results on LFW and MF1-Facescrub dataste. The reported results are verification rate for LFW, and rank-1 identification rate for MF1-Facescrub.

4.6. Experiments on the General Face Recognition Datasets

To compare against the state-of-the-art methods in General Face Recognition(GFR), we further conduct experimental evaluations on the LFW and the MegaFace Challenge 1 Facescrub (MF1-Facescrub) datasets. The LFW [19] is a public benchmark for GFR that has 13,233 face images from 5,749 subjects. The MF1-Facescrub [21] includes the Facescrub (containing 106,863 face images from 530 celebrities) as a probe set and contains a million distractors in the gallery set. We strictly follow the same training and evaluation procedure in OE-CNNs [43]. That is, our training data contains 0.5M images that are the same as OE-CNNs [43]. Table 7 reports the verification rate on LFW and the rank-1 identification rate in MF1-Facescrub. Our model outperforms the [43] as well as the state-of-the-art General Face Recognition (GFR) models [28, 42] on both datasets, which demonstrates the strong generalization ability of our proposed approach.

5. Conclusion

In this paper, we propose the decorrelated adversarial learning method for AIFR. Our model learns to minimize the correlation between the paired decomposed features of identity and age in an adversarial process. We present the BCCA algorithm as an extension of CCA in deep learning. Besides the DAL, we simultaneously train the model with the joint supervision of identification and age classification. In the testing, only the identity features are used for face recognition. Evaluations conducted on the AIFR benchmarks demonstrate the superiority of our method.

References

[1] *FG-NET Aging Database,http://www.fgnet.rsunit.com/*. 2, 6

[2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face Aging With Conditional Generative Adversarial Networks. *IEEE International Conference on Image Processing (ICIP)*, 2017. 1, 3

[3] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem. Finding tiny faces in the wild with generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*, 2018. 3

[4] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy. Pose-robust face recognition via deep residual equivariant mapping. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*, 2018. 2, 3

[5] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European Conference on Computer Vision (ECCV)*, 2014. 2, 6, 7

[6] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6):804–815, 2015. 2, 7

[7] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3025–3032, 2013. 2, 7

[8] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang. FSRNet: End-to-end learning face super-resolution with facial priors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*, 2018. 3

[9] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou. Uv-gan: Adversarial facial uv map completion for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*, 2018. 3

[10] J. Deng, J. Guo, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *arXiv:1801.07698*, 2018. 1

[11] C. N. Duong, K. G. Quach, K. Luu, M. Savvides, et al. Temporal Non-Volume Preserving Approach to Facial Age-Progression and Age-Invariant Face Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 3, 7, 8

[12] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence (TPAMI)*, 2007. 1

[13] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden factor analysis for age invariant face recognition. In *International Conference on Computer Vision (ICCV)*, 2013. 2, 4, 5, 7, 8

[14] D. Gong, Z. Li, J. Liu, and Y. Qiao. Multi-feature Canonical Correlation Analysis for Face Photo-Sketch Image Retrieval. In *Proceedings of ACM international conference on Multimedia*, pages 617–620, 2013. 2

[15] D. Gong, Z. Li, D. Tao, J. Liu, and X. Li. A maximum entropy feature descriptor for age invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5289–5297, 2015. 2, 7

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, , and Y. Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, 2014. 2

[17] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[18] H. Hotelling. Relations between two sets of variates. *Biometrika,*, 1936. 2

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Technical Report 07-49, University of Massachusetts, Amherst*, 2007. 8

[20] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of International Conference on Machine Learning*, 2015. 4

[21] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 6, 8

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2012. 1

[23] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2002. 1

[24] Z. Li, D. Gong, X. Li, and D. Tao. Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing (TIP)*, 25(5):2146–2154, 2016. 2, 7

[25] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE transactions on Information Forensics and Security (TIFS)*, 2011. 2, 7

[26] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(6):1089–1102, 2017. 2, 7

[27] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. Face verification across age progression using discriminative methods. *IEEE transactions on Information Forensics and Security (TIFS)*, 2010. 2

[28] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. SphereFace: Deep Hypersphere Embedding for Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 8

[29] Y. Liu, F. Wei, J. Shao, L. Sheng, J. Yan, and X. Wang. Exploring Disentangled Feature Representation Beyond Face Identification. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[30] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *International Conference on Computer Vision (ICCV)*, 2015. 5

[31] A. Nech and I. Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 6, 8

[32] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010. 1, 7

[33] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference (BMVC)*, 2015. 5

[34] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *International Conference on Automatic Face and Gesture Recognition*, 2006. 2, 6

[35] R. Rothe, R. Timofte, and L. V. Gool. Dex: Deep expectation of apparent age from a single image. In *International Conference on Computer Vision Workshops (ICCVW)*, December 2015. 5

[36] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 1

[37] Y. Sun, Y. Chen, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. In *Advances in Neural Information Processing Systems (NIPS)*, 2014. 1

[38] Y. Sun, D. Liang, X. Wang, and X. Tang. Deepid3: Face recognition with very deep neural networks. In *arXiv preprint arXiv:1502.00873*, 2015. 1

[39] Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1

[40] F. Wang, J. Cheng, W. Liu, and H. Liu. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25:926–930, 2018. 1, 4

[41] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille. NormFace: L_2 Hypersphere Embedding for Face Verification. In *Proceedings of the 2017 ACM on Multimedia Conference (ACM MM)*, 2017. 1

[42] H. Wang, Y. Wang, Z. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, and W. Liu. Cosface: Large margin cosine loss for deep face recognition. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 4, 8

[43] Y. Wang, D. Gong, Z. Zhou, X. Ji, , H. Wang, Z. Li, W. Liu, and T. Zhang. Orthogonal Deep Features Decomposition for Age-Invariant Face Recognition. In *European Conference on Computer Vision (ECCV)*, 2018. 2, 5, 7, 8

[44] Z. Wang, K. He, Y. Fu, R. Feng, Y.-G. Jiang, and X. Xue. Multi-task Deep Neural Network for Joint Face Recognition and Facial Attribute Prediction. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR)*, 2017. 8

[45] Y. Wen, Z. Li, and Y. Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2, 4, 5, 7

[46] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision (ECCV)*, pages 499–515, 2016. 1

[47] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. *arXiv preprint arXiv:1611.05431*, 2016. 1

[48] C. Xu, Q. Liu, and M. Ye. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing*, 222:62–71, 2017. 7

[49] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li. 2D-3D Face Matching using CCA. 2

[50] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. In *arXiv preprint arXiv:1411.7923*, 2014. 5

[51] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang. Super-Identity Convolutional Neural Network for Face Hallucination. In *European Conference on Computer Vision (ECCV)*, 2018. 3

[52] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *Signal Processing Letters*, 23(10):1499–1503, 2016. 5

[53] Z. Zhang, Y. Song, and H. Qi. Age Progression/Regression by Conditional Adversarial Autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 3

[54] Y. Zhao, Z. Jin, G. jun Qi, H. Lu, and X. sheng Hua. An adversarial approach to hard triplet generation. In *European Conference on Computer Vision (ECCV)*, 2018. 3

[55] T. Zheng, W. Deng, and J. Hu. Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. 2, 7