applied sciences

MDPI

Article

Effective Attention-Based Feature Decomposition for Cross-age Face Recognition

Suli Li [1,2] and Hyo Jong Lee [1,*]

[1] Department of Computer Science and Engineering, Jeonbuk National University, Jeonju 54896, South Korea;
[2] Department of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061000, China; sli88@foxmail.com;
* Correspondence: Hyo Jong Lee (hlee@jbnu.ac.kr);

Abstract: Deep-learning based cross-age face recognition has improved significantly in recent years. However, when using the discriminative method, it is still challenging to extract robust age-invariant features that can reduce the interference caused by age. In this paper, we propose a novel, effective, attention-based feature decomposition model, the Age-invariant Features Extraction Network, that can learn more discriminative feature representations and reduce the disturbance caused by aging. Our method uses an efficient channel attention block–based feature decomposition module to extract age-independent identity features from facial representations. In addition, we propose a direct sum loss function to reduce the interference of age-related features. Experimental results on both cross-age datasets (CACD-VS, AgeDB, CALFW, and FG-NET) and a general dataset (LFW) indicate the effectiveness and superiority of the proposed method.

Keywords: Convolutional neural networks, cross-age face recognition, deep learning, attention mechanism

Citation: Li, S.; Lee, H.J. Effective Attention-Based Feature Decomposition for Cross-age Face Recognition. *Appl. Sci.* **2022**, *12*, x. https://doi.org/10.3390/xxxxx

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Face recognition (FR) is a biometric identification technology that is convenient, friendly, contactless, non-invasive, and easy to integrate. It has played an important role in identity authentication and is widely used in many application areas, such as law enforcement [1], identity verification processes, and security [2]. FR technology has become more mature in recent decades. Many models [3-9] based on deep networks have been proposed to address FR tasks, such as Deepface [5], VGGFace [6], FaceNet [7], and light CNN. State-of-the-art FR approaches [10,11] using ResNet architecture [12] have even surpassed human performance in several scenarios and correctly identify faces in many real-world applications.

However, the general FR models might not be robust enough to identify faces with a wide range of ages. Cross-age face recognition (CAFR) has been an increasing research interest due to its potential in real-world applications. For example, finding missing children after many years or identifying criminals who have absconded years later usually involves recognizing the same face at different ages [13]. CAFR focuses on identifying a person from images taken at different ages. However, as seen in Fig. 1, a single person's facial shape and texture can change dramatically over time, making CAFR an extremely challenging task [14]. In particular, when the age span is large, the intra-class variations between face images from a single person are also large, making it difficult to learn age-invariant patterns.

Existing methods for CAFR can be divided into two categories: generative approaches and discriminative approaches. The generative approaches [14-19] synthesize a desired image into the target age group to assist with face recognition. The downside of

the generative models is their high computational costs caused by the high complexity involved in modeling aged faces. The discriminative approaches [20,21] address CAFR tasks by extracting age-invariant representations. The development of deep convolutional neural networks (CNNs) has enabled a breakthrough in discriminative approaches in recent years. In contrast to the older handcrafted features, CNN-based methods [22-25] learn how to extract more discriminative features from large-scale datasets. Though CNN-based approaches have shown superior performance, there remains the limitation that age-invariant features are still accompanied by age-related information. Therefore, it is still a challenge to extract robust age-invariant features to reduce the interference caused by age.

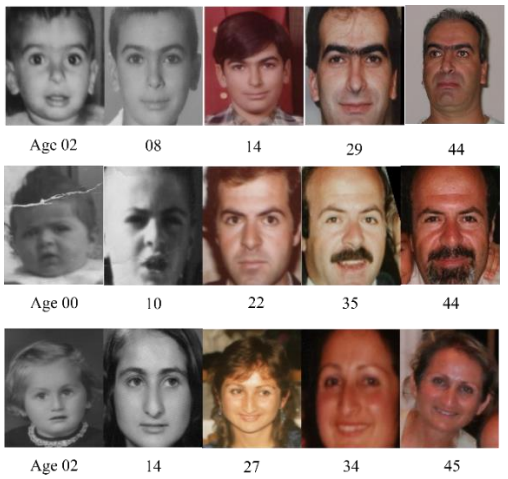


Figure 1. Example images from the FG-NET dataset showing the same person at different ages, illustrating the significant changes caused by facial aging.

Recently, visual attention mechanisms have been widely used to solve vision problems such as image classification. Attention mechanism–based approaches use channel attention to capture essential features. Therefore, we can extract age-related features in facial representations using channel attention. In this paper, we propose a novel, effective, attention-based feature decomposition model for CAFR that can learn many discriminative feature representations and reduce the disturbance caused by aging. We propose an efficient channel attention (ECA) block-based feature decomposition module (EFDM) to decompose mixed features into identity-specific and age-specific features. We also propose a novel loss function based on a direct sum to sufficiently separate age information and identity information from facial features. Through that direct sum loss function, we achieve a significant separation between identity-specific and age-specific features, which allows us to notably reduce the age information in the identity features.

The main contributions of this paper are three-fold:

- We propose an efficient end-to-end training model called the Age-invariant Features Extraction Network (AFEN) to learn age-invariant features for CAFR. Our end-to-end model greatly reduces the number of training parameters and training complexity, compared with existing multi-stage training methods.
- Based on the attention mechanism, we propose a novel EFDM to extract deep age-invariant features, as well as a direct sum loss function to sufficiently facilitate the decomposition of age-specific and identity-specific features.
- We report the results of extensive experiments conducted to compare our proposed approach with state-of-the-art models. Analyses using the developed model are conducted on popular public datasets (CACD-VS [20], AgeDB [26], CALFW [27], FG-NET [28], and LFW [29]) to demonstrate the robustness of the proposed method.

The rest of this paper is organized as follows. Related works on the generative and discriminative approaches are introduced in Section 2. The proposed method is described in Section 3. The experimental results and analyses of four public CAFR databases and one general face database are reported in Section 4, and conclusions are given in Section 5.

2. RELATED WORKS

2.1. GENERATIVE APPROACHES

Deep generative model–based networks are intensively applied in synthesis schemes. For instance, Zhang et al. [15] proposed a conditional adversarial autoencoder that learns a face manifold to achieve face age regression and progression. A pyramid architecture of generative adversarial networks (GANs) was proposed in an age progression model [16] to ensure that the generated faces have the desired aging effects while keeping the personalized properties stable. Wang et al. [17] proposed an identity-preserved conditional GAN for facial aging. A conditional GAN generates a realistic face at the target age, and an identity-preserved module preserves identity information. Zhao et al. [14] proposed a deep age-invariant model that jointly performs cross-age face synthesis and recognition. Huang et al. [19] proposed a unified, multi-task learning framework (MTLFace) for CAFR that simultaneously achieves age-invariant identity-related representation and face synthesis. The face synthesis approach improves the results. However, those methods still suffer ghosting artifacts on the synthesized face. In addition, those models carry high computational costs because of the high complexity of modeling an aged face, and they fail to achieve stable performance.

2.2. DISCRIMINATIVE APPROACHES

The cross-age discriminative models focus on extracting age-invariant representations. Chen et al. [20] introduced a novel coding method called cross-age reference coding that encodes an image on a cross-age reference to obtain an age-invariant feature representation. Du et al. [30] proposed a cycle age-adversarial model that extracts age-invariant features and only uses age labels for training. Huang et al. [31] proposed the Age-Puzzle FaceNet (APFN) based on an adversarial training mechanism to address the CAFR task. Huang et al. [32] later updated the APFN to make it more compact and robust to age variation.

Some recently proposed methods assume that whole-face features are composed of age-related factors and age-invariant factors. Those methods focus on decomposing the aging and identity components separately [4,21,25]. For example, Gong et al. [21] separated identity-related factors and age-related factors using a hidden factor analysis (HFA). Wen et al. [22] developed a latent identity analysis layer to separate the two components. The Age-Estimation Guided Convolutional Neural Network [23] uses the age estimation task to guide the separation of age features from the identity feature layer. Wang et al. [24] presented feature decomposition in an orthogonal embedding CNN (OE-CNN) and adapted SphereFace loss [33] to deal with the CAFR task. Wang et al. [25] proposed the decorrelated adversarial learning (DAL) algorithm to achieve feature decomposition in an adversarial manner. However, the age-invariant features are still accompanied by age-related information. Moreover, the performance of a discriminative network trained from scratch depends heavily on the number of images in the special cross-age dataset.

In mathematics, the direct sum of two abelian groups of identify features and aging features forms another abelian group consisting of ordered pairs of the two features. Therefore, the entire set of facial features extracted from a well pre-trained general FR network can be decomposed into an identity-specific feature and an age-specific feature. We introduce an attention-based module and direct sum loss to facilitate this characteristic in our proposed method.

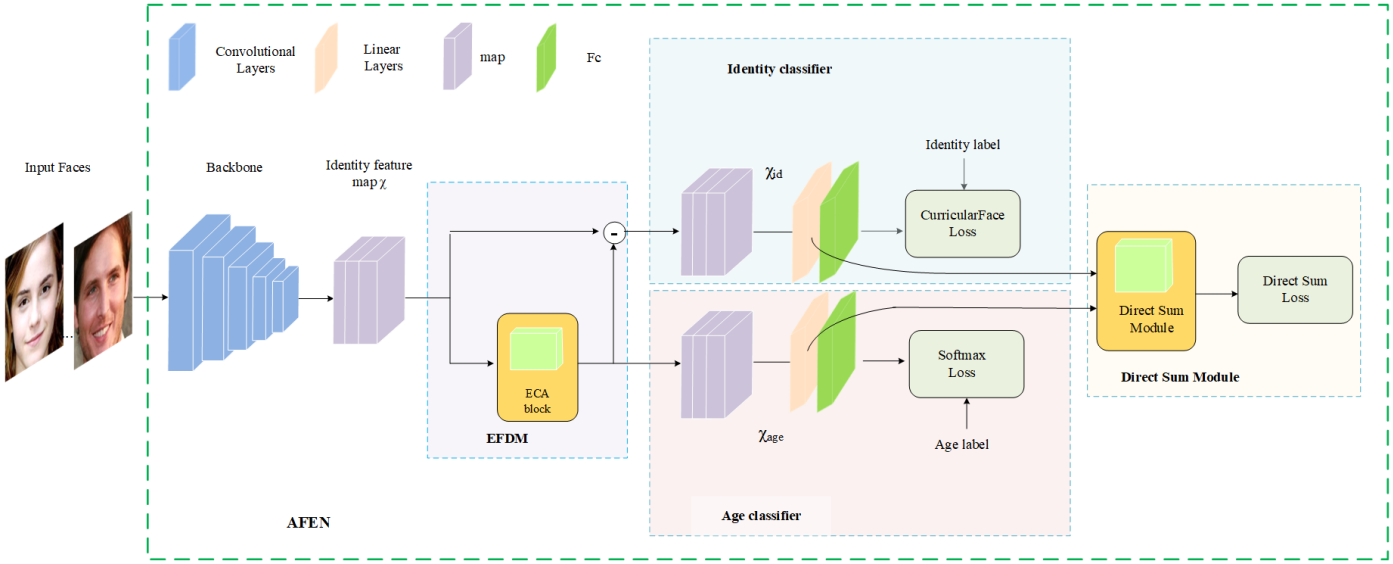


Figure 2. The overall framework of the proposed AFEN and its training process. After inputting a face image, an identity-specific feature map is extracted through the EFDM. The age-invariant features are then extracted through the identity classifier, age classifier, and direct sum module.

3. PROPOSED METHOD

This section provides a detailed description of the proposed method. The AFEN framework is shown in Fig. 2 and consists of five parts: a well-trained CNN as the backbone, the EFDM, an age classifier, an identity classifier, and a direct sum module. The five parts jointly perform end-to-end age-invariant facial feature decomposition.

3.1. FEATURE DECOMPOSITION MODULE

Because the facial features extracted from the backbone are severely entangled with age-related information such as texture changes, it is difficult to recognize two face images of the same person with a large age gap [19]. A CNN extracts the feature vector $x \in \mathbb{R}^d$ from an input image $I \in \mathbb{R}^{3\times H\times W}$. The linear factorization can be defined as:

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

where x_{id} and x_{age} denote the identity-dependent factor and age-dependent factor, respectively. The identity-related factor, x_{id}, can work as the age-invariant feature for CAFR. However, it has a drawback: this decomposition acts on a one-dimensional feature vector. The final identity-related factor lacks semantic feature information about the aging face, such as beards and wrinkles. To resolve that issue, we propose a feature decomposition module (the EFDM) that uses ECA to decompose the feature map instead of a feature vector.

Channel attention is a crucial component for improving the generalization capabilities of a deep CNN architecture. The channels are the result of convolutional filters that derive different features from the input, and they might not all have the same representative importance. Because some channels are more important than others, it makes sense to apply a weight to the channels based on their importance before they propagate to the next layer. In this paper, we adopt channel-wise attention to highlight age-related information at the channel level. Because the parameter and FLOPs overhead of the ECA block [34] are much smaller than the squeeze and excitation block [35], we use the ECA block to compute and apply attention weights to the channels of the input feature map.

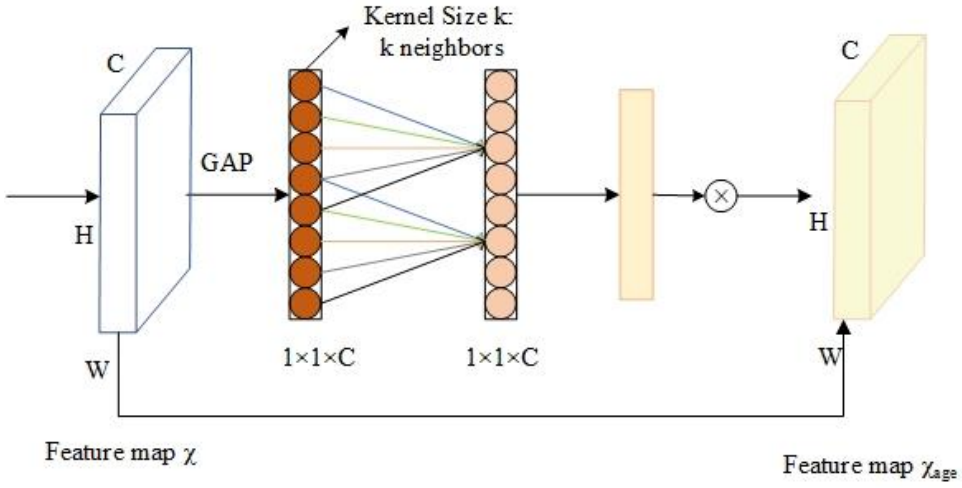


Figure 3. Diagram of the ECA block. The ECA block first uses global average pooling (GAP) for each channel independently to aggregate the feature map $\boldsymbol{\chi} \epsilon \mathbb{R}^{C \times H \times W}$. Then, the ECA generates channel weights by performing a fast 1D convolution of size k and a sigmoid function, where k represents how many neighbors participate in the attention prediction for one channel. Next, the age-specific feature map, $\boldsymbol{\chi}^{age} \in \mathbb{R}^{C \times H \times W}$, is obtained by channel-wise multiplication between the channel weights and the feature map $\boldsymbol{\chi}$.

The ECA block is an extremely lightweight channel attention module for deep CNNs. As illustrated in Fig. 3, after channel-wise global average pooling without dimensionality reduction, the ECA module efficiently captures cross-channel interactions by considering every channel and its k neighbors. The ECA generates channel weights by performing a fast 1D convolution of size k, where kernel size k represents the coverage of the local cross-channel interaction, i.e., how many neighbors participate in the attention prediction for one channel.

In Fig. 3, we use the backbone to extract a facial feature map, $\chi \epsilon \mathbb{R}^{C \times H \times W}$, from the input image $I \in \mathbb{R}^{3 \times H \times W}$. In the decomposition module, we use the ECA block to transform the facial feature map, χ, into an age-specific feature map, χ_{age}. The decomposition process can be defined as follows:

$$\chi_{id} = \chi - \chi_{age}. \quad (2)$$

By subtracting the age-specific feature map from the facial feature map χ, we obtain the identity-specific feature map χ_{id}.

We built the feature decomposition module based on the ECA block to obtain the identity-specific feature and age-specific feature. To achieve a significant separation of the identity-specific and age-specific features, we propose the direct sum loss, which is introduced in the following section.

3.2. DIRECT SUM LOSS

The direct sum of subspaces is a relationship between two linear spaces in higher algebra as a special case of the sum of subspaces.

Definition 3.1: Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two subspaces of linear space $\mathcal{V}$. And let $(\alpha_1, \alpha_2 \ldots, \alpha_s)$ be basis vectors of $\mathcal{W}_1$, and $(\beta_1, \beta_2 \ldots, \beta_t)$ be basis vectors of $\mathcal{W}_2$. If $(\alpha_1, \alpha_2 \ldots, \alpha_s, \beta_1, \beta_2 \ldots, \beta_t)$ are basis vectors of $\mathcal{V}$, then we say $\mathcal{W}_1 + \mathcal{W}_2$ meets the direct sum condition.

If two subspaces meet the direct sum condition, the redundant components between the two subspaces can be effectively removed. Therefore, we designed the direct sum loss

constraint to reduce the redundant components between the identity-related features and age-related features. The details on how to implement direct sum loss are as follows.

Let x_{id} denote the identity-related feature and x_{age} denote the age-related feature. The identity-related feature space and age-related feature space are marked as V_I and V_A, respectively.

We obtain the basis vectors $(v_{id1}, v_{id2} \ldots, v_{idK})$ from a fully connected layer. The space formed by the basis vectors $(v_{id1}, v_{id2} \ldots, v_{idK})$ is marked as V_{II}. In the same way, we obtain the space V_{AA} formed by the basis vectors $(v_{age1}, v_{age2} \ldots, v_{ageK})$. To make the space $V_{II} + V_{AA}$ meet the direct sum condition, $(v_{id1}, v_{id2} \ldots, v_{idK}, v_{age1}, v_{age2} \ldots, v_{ageK})$ must be linearly independent. Therefore, the direct sum loss is represented as follows:

$$L_{direct\_sum} = \frac{1}{K}\sum_{i=1}^{K} | cos(v_{id_i}, v_{age_i}) |. \quad (3)$$

The cosine similarity is as close to 0 as possible, making the two basis vectors linearly independent.

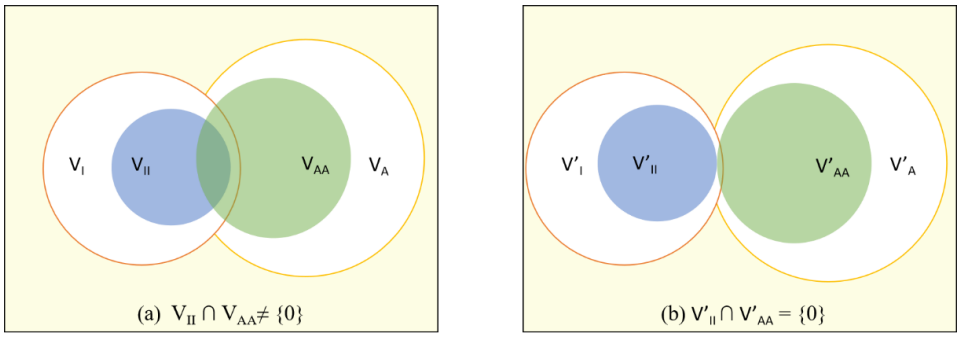


Figure 4. Direct sum. (a) Before applying the subspace direct sum constraint. (b) After applying the subspace direct sum constraint.

In the training process, the identity-related feature space V_I and age-related feature space V_A and the subspaces V_{II} and V_{AA} are updated continually. As illustrated in Fig. 4, the updated spaces are marked as V_I', V_A', V_{II}', *and* V_{AA}'. By applying the direct sum constraint to the two feature subspaces V_{II} and V_{AA}, the redundant components between the identity feature space V_I and the age feature space V_A are optimally separated. Making the identity-related features and age-related features linearly independent ultimately allows the facial identity features to be extracted. The process is summarized in Algorithm 1.

Algorithm 1 Procedure for making the V_{II} and V_{AA} subspaces meet the direct sum constraint

Input: Cross-age training dataset T={I_i | i=1,2…,N};
the number of basis vectors marked as K

1: P→0
2: Initialize the parameters of the model
3: **repeat**
4: Obtain $x_{id}^{(P)}, x_{age}^{(P)}$
5: Fully connected layer to obtain the basis vectors $(v_{id1}, v_{id2} \ldots, v_{idK})$ marked as V_{II}; Another FC layer to obtain $(v_{age1}, v_{age2} \ldots, v_{ageK})$ marked as V_{AA}
6: Calculate the loss
7: Update the parameters of the model
8: P→P+1
9: **until** convergence

3.3. END-TO-END OPTIMIZATION OF THE NETWORKS

Identity classification task. Through the feature decomposition module, we obtain the identity-specific feature map χ_{id}. Then, the χ_{id} is translated into an identity-specific feature, x_{id}, at the output layer for use as the age-invariant feature in the final cross-age face verification. To enhance the discriminative identity power, CurricularFace loss [11], which has been successfully applied to boost face recognition performance, is used to supervise the learning of x_{id} and to ensure identity-preserving information. The CurricularFace loss is represented as follows:

$$L_{Curr\_Face} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{s(cos(\theta_{yi}+m))}}{e^{s(cos(\theta_{yi}+m))}+\sum_{j=1,j\neq yi}^{n} e^{sI(t,cos\,\theta_j)}}, \quad (4)$$

$$\mathrm{I}(t, cos\,\theta_j) = \begin{cases} cos\,\theta_j, & cos(\theta_{yi}+m) - cos\,\theta_j \geq 0 \\ cos\,\theta_j(t + cos\,\theta_j), & cos(\theta_{yi}+m) - cos\,\theta_j < 0 \end{cases} \quad (5)$$

where N is the number of training samples, x_i is the ith feature vector corresponding to the ground-truth class of y_i, θ_j is the angle between the normalized features of the prototype corresponding to the jth identity and x_i, the hyperparameter s determines the radius of the mapped hypersphere, and m controls the cosine margin. The value of t indicates the model training stage. CurricularFace loss explores the discrepancy between the real identity and the predicted identity from the identity classification task.

Age classification task. Following previous work [14], we divide faces of different ages into seven groups: ages 0–20, 21–25, 26–30, 31–40, 41–50, 51–60, and 61 or older. We use an auxiliary age discriminator to guide the decomposition procedure and find intrinsic clues for age information. Softmax loss is widely used in existing face recognition for the ground truth identity. We use a softmax function with cross-entropy loss as a loss function in the age classification task to guide the predicted age approach to the actual age, which is represented as:

$$L_{age} = -\frac{1}{N}\sum_{i=1}^{N} log p_i, \quad (6)$$

where p_i is the predicted probability that input x_i belongs to the correct age group.

Complete loss and training algorithm. These three losses (direct sum, CurricularFace, and age) are combined into a multi-task loss for joint optimization, as given below:

$$L_{total} = \mathrm{L}_{\mathrm{Curr\_Face}} + \lambda_1 L_{direct\_sum} + \lambda_2 L_{age}, \quad (7)$$

where λ_1 and λ_2 are scalar hyperparameters to balance the three losses. We set both weights λ_1 and λ_2 to 0.01 after the experimental analysis. The model training process is summarized in Algorithm 2.

Algorithm 2 Model training for AFEN

Input: Cross-age training data x_i with both age label t and identity label y_i

Output: The age-invariant feature

1: Initialize the parameters of the backbone using the pre-trained model.

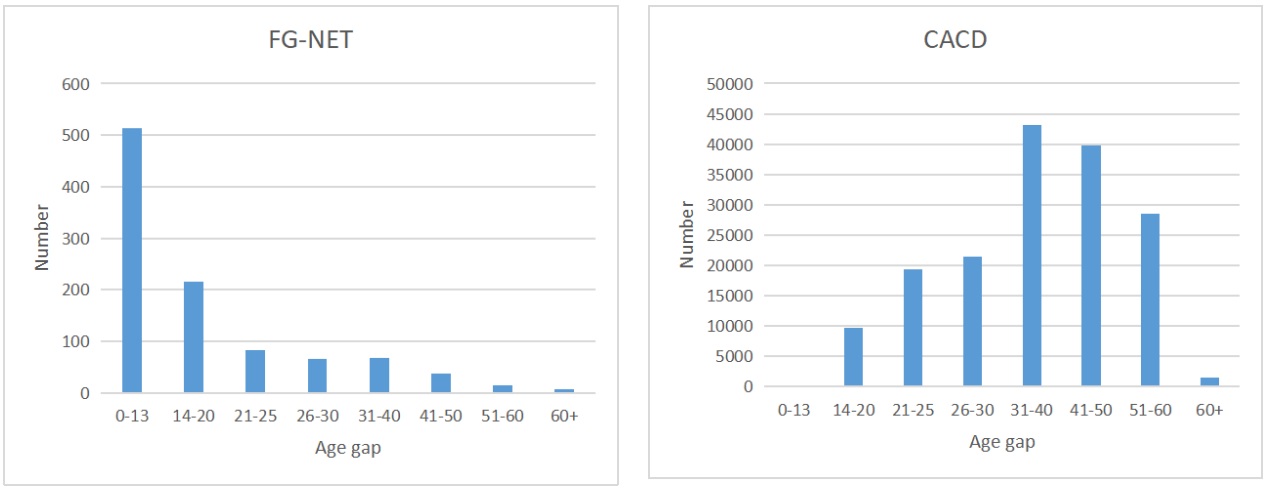
2: Train the EFDM, identity classifier, age classifier, and direct sum module with training data x_i and identity label y_i, and use the total loss function in (6) to obtain age-invariant feature x_{id}.

3: Train until convergence.

4. EXPERIMENTS

4.1. IMPLEMENTATION DETAILS

Datasets. Several public cross-age datasets were used for model training and evaluation: CACD, CACD Verification Subset (CACD-VS), AgeDB, CALFW, and FG-NET. We used some CACD dataset to train our model, and we used the rest of them for evaluation. The distribution of ages in the CACD and FG-NET datasets is shown in Fig. 5. The CACD dataset is used as a public benchmark for CAFR, and it is composed of 163,446 images of 2,000 celebrities. The images reflect various shooting conditions, such as illumination variations, pose variations, age variations, makeup, and practical scenarios. FG-NET contains 1,002 images of 82 people with an age range from 0 to 69; FG-NET has larger age gaps than CACD, but it contains only a few images of a small number of people. The CACD dataset can effectively reflect the robustness of our CAFR algorithm.



(**a**) (**b**)

Figure 5. The distribution of ages in two of the datasets: (a) FG-NET. (b) CACD.

Therefore, we choose the CACD dataset as the training dataset. We randomly selected 80% of its images as the training data and used the remaining 20% for validation. However, the CACD dataset contains some incorrectly labeled samples and duplicate images. In particular, the age labels do not match the real age. We used the DEX [36] method to produce age labels as the ground truth. To get better training results, we also manually removed duplicate images.

We conducted experiments on commonly used, public, cross-age datasets: CACD-VS [20], CALFW [27], AgeDB [26], and FG-NET [28]. We extracted only the identity-specific feature as the final facial feature representation for identity recognition in the testing process. The cosine similarity of these representations was then used to conduct face verification and identification.

Data preprocessing. The CACD dataset was used to fine-tune our network in the experiments. We used the multi-task cascaded convolutional network [37] to detect face areas and facial landmarks in the training images. After detecting the eye position, we applied an affine transformation to the data to align the face images based on the detected eye coordinates. All faces were globally cropped to 112×112 based on five facial landmarks (two mouth corners, nose center, and two eyes) and a similarity transformation. Fig. 6 shows some original and preprocessed face images from the CACD dataset.

Training protocols. Because the CNN architecture of the ResNet module has been proved to be an effective mapping function, we used IResNet-101, which is pre-trained on the general face dataset MS-Celeb-1M, as the backbone to capture the most prominent features for identity discrimination. MS-Celeb-1M is a dataset that contains 5.8 million images of 8,500 subjects across pose and age. The pre-trained model can classify tens of thousands of identities and extract multilevel, high-resolution features.

Figure 6. Examples of data used. The top row shows the original images, and the bottom row shows the aligned and normalized images.

We initialized the shared model with the pre-trained model and then trained the feature to decompose module on the CACD dataset with a batch size of 512 on four Nvidia Titan X Pascal GPUs. The models were trained with the SGD algorithm and a momentum of 0.9 and weight decay of 5e-4. We selected the hyperparameters by trial and error. The training process was finished at 30 epochs of 9.57K iterations. We used Adam as an optimizer and set the initial learning rate to 0.01. We followed the common setting as given in [10] to set the scale factor and multiplicative margin of CurricularFace loss to 64 and 0.5, respectively. Fig. 7 shows the variation trend for training loss with the optimal parameters.

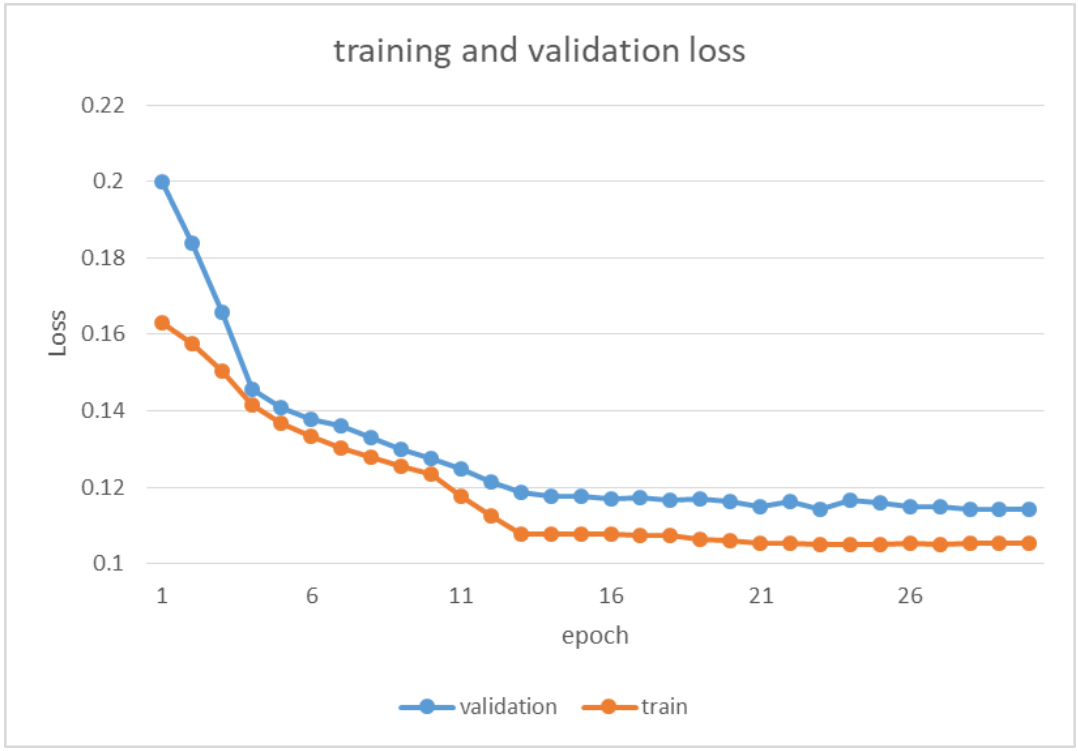


Figure 7. Trend of training and validation loss.

4.2. ABLATION STUDIES

In this subsection, we describe an experiment performed to investigate the efficacy of the proposed model with the CACD-VS, CALFW, and AgeDB-30 datasets. Then, we analyze the effect of taking different values for the hyperparameters, λ_1 and λ_2 in Eq. (7),

and K in Algorithm 1. In the end, we compare the time complexity with that of state-of-the-art methods.

Efficacy of the proposed method. To investigate the efficacy of the proposed decomposition module and direct sum loss in our method, we considered the following variants of our method for ablative comparison based on three benchmark datasets for CAFR: (1) Baseline: the baseline model was pre-trained only on IResNet-101; (2) Baseline +EFDM: the model was trained by the EFDM; (3) Baseline +EFDM +direct sum: our proposed model, which was trained simultaneously by the EFDM and direct sum module. As reported in Table 1, our model had the best performance on CALFW, CACD-VS, and AgeDB, demonstrating the efficacy of the proposed method.

Table 1. Evaluation results (%) by different methods.

EFDM	Direct sum	CALFW	CACD-VS	AgeDB-30
		96.03	99.33	98.37
√		96.06	99.40	98.38
√	√	96.10	99.63	98.38

Settings of the hyperparameters. As mentioned in our description of the whole loss function, we use hyperparameters λ_1 and λ_2 to balance the three losses. We conducted experiments to observe the effects of λ_1 and λ_2. We respectively set λ_1 as {1, 0.1, 0.01, 0.001} while λ_2=0.01, and then set λ_2 as {1, 0.1, 0.01, 0.001} while λ_1=0.01 to test the face verification accuracy with the CACD-VS dataset. The verification rates under different values of λ_1 and λ_2 are shown in Fig. 8, which indicates that the best performance was obtained when λ_1=0.01 and λ_2=0.01.

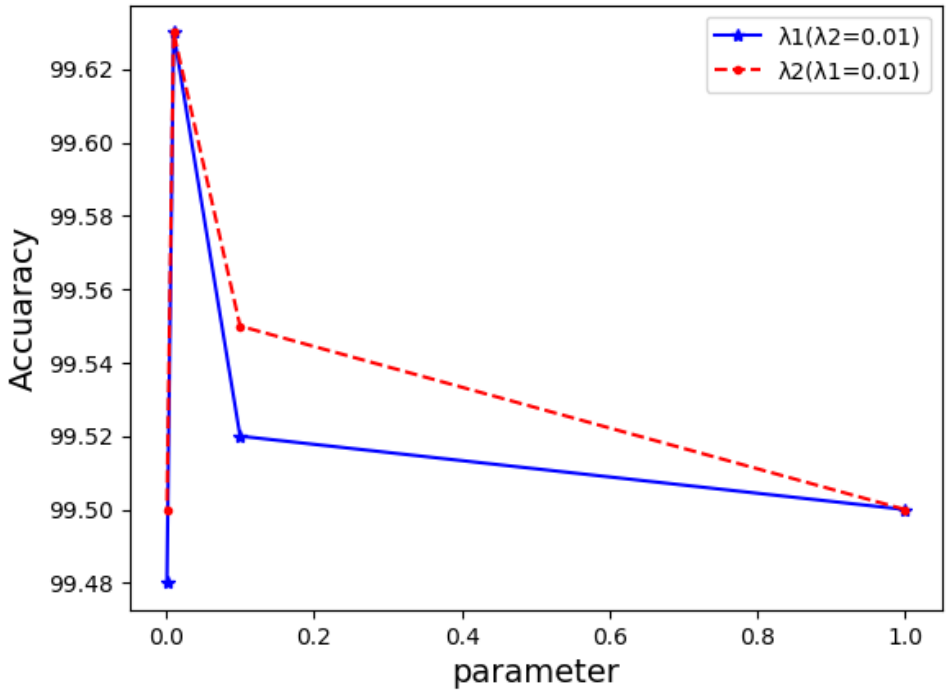


Figure 8. Face verification accuracy on the CACD-VS dataset with various values for $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$.

Parameter K is the number of basis vectors, as explained above in the direct sum loss section. We constructed face verification experiments using four values (K = 25, 50, 75, and 100). The evaluation results are tabulated in Table 2 and show that increasing K can improve face verification accuracy. That is understandable because a larger K leads to a more

powerful nonlinear transformation. The best performance was obtained when K=75. Further increases in K led to more noise vectors, which produced a drop in accuracy. Therefore, we set parameter K to 75.

Table 2. Evaluation results (%) on the CACD-VS dataset.

K (value)	25	50	75	100
Accuracy	99.60	99.60	**99.63**	98.55

Exploration of identity loss. The Arcface loss is also an effective method for generating discriminative identity features. Therefore, we compared the performance of Arcface and CurricularFace loss on the CACD-VS dataset by replacing the CurricularFace loss in Eq. (7) with the Arcface loss. Table 3 shows the results of CurricularFace and Arcface on CACD-VS. CurricularFace loss performed better than Arcface loss.

Table 3. Evaluation results (%) on CACD-VS using different losses.

Loss function	Arcface	CurricularFace
Accuracy	99.57	**99.63**

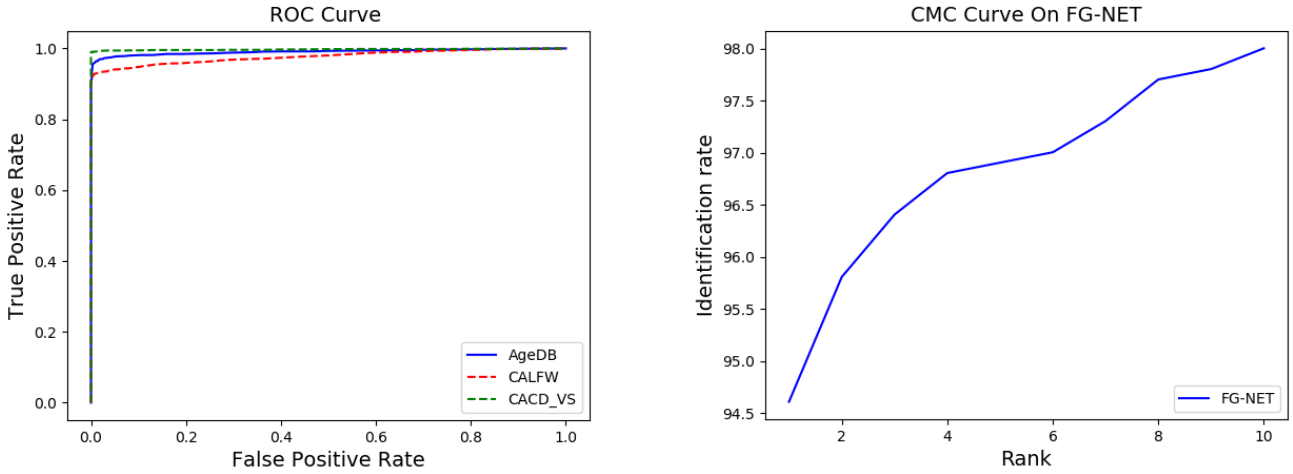
Time complexity. We used a well pre-trained general FR network to save resources and reduce the time required to train the network from scratch. With fewer training images and training parameters (in Table 4) than MTLFace [19] and DAL [25], the proposed method achieves a comparable and stable performance. Compared with the age-invariant representations learning method under the same environment and batch size, DAL [25] cost 0.963s for each iteration on the Nvidia Titan X Pascal GPUs, whereas our method cost only 0.416s. Therefore, our method reduces the complexity and computational power required for CAFR.

Table 4. Comparisons of the training parameters and images required by different methods.

Method	**Training parameters of networks (millions)**	**Training no. of images**
DAL [25]	54.73	313,986
MTLFace [19]	72.82	1,700,000
Ours	13.42	163,446

4.3. EVALUATIONS ON MULTIPLE BENCHMARK DATASETS

We evaluated our method on several benchmark cross-age datasets, CACD-VS [20], CALFW [27], AgeDB [26], and FG-NET [28], and a general face dataset, LFW [29]. Note that MORPH [38] was excluded because it was prepared for commercial use only. To evaluate the performance of our proposed method and compare it with other state-of-the-art CAFR methods, we chose the verification accuracy and rank-1 identification rate as evaluation metrics. We used the receiver operating characteristic curve (ROC curve), which expresses the quality of the 1:1 matcher, to evaluate the verification accuracy. As shown in Fig. 9(a), we created the ROC curve on the cross-age datasets (CACD-VS, CALFW, and AgeDB) by plotting the true positive rate against the false positive rate. To evaluate the identification accuracy with the FG-NET dataset, we used the cumulative match curve (CMC), to measure 1: k identification system performance. The evaluation schemes for the different datasets are described next.



(**a**) (**b**)

Figure 9. Face recognition performance of our model (**a**) ROC curves on CACD-VS, CALFW, AgeDB-30. (**b**) CMC curve on FG-NET.

CACD-VS. The CACD-VS consists of 4,000 face image pairs for face verification, 2,000 positive and 2,000 negative pairs. The age difference between most image pairs from the same person is less than nine years. We followed the pipeline suggested by Chen et al. to calculate the similarity score of all the sample pairs [20]. We strictly followed the cross-validation rule [24] to calculate the similarity score for all sample pairs. We first divided the dataset into ten folds, with each fold containing 400 image pairs (200 positive pairs and 200 negative pairs) from 200 celebrities. We used nine of those ten folds to compute the threshold references and then used the best threshold to evaluate the last one fold. We repeated those experiments ten times for each of the ten folds and finally calculated the average accuracy. The evaluation results with the different methods on CACD-VS are summarized in Table 5.

As shown in Table 5, the proposed method outperformed all the tested methods in a large-scale dataset, achieving an accuracy of 99.63%, which indicates the effectiveness of the proposed method. Note that the MTLFace method requires a large-scale dataset of almost 1.7 million faces for training, whereas ours used only 0.16 million faces.

Table 5. Evaluation results of the different methods on CACD-VS.

Method	Accuracy (%)
LFCNN [22] (2016)	98.5
OE-CNN [24] (2018)	99.20
AIM [14] (2020)	99.38
DAL [25] (2019)	99.40
AA_CNN [32] (2021)	99.20
PMADN [39] (2021)	99.20
MTLFace [19] (2021)	99.55
IEFP [40] (2022)	99.57
Ours	**99.63**

CALFW. To demonstrate the effectiveness of our method for face recognition with a larger age span, we implemented a face verification experiment on the cross-age labeled faces in the wild (CALFW) dataset. The CALFW dataset is an extension of the LFW dataset designed for unconstrained face verification with larger age gaps. First, 3,000 positive face

pairs with age gaps are selected from LFW, in which the age gaps of most positive pairs are larger than ten years. The average age gap is about 20 years. Then 3,000 negative pairs with the same gender and race are selected to reduce the influence of different attributes [32].

Table 6. Evaluation results of the different methods on CALFW.

Method	Accuracy (%)
Center Loss [41] (2016)	85.48
SphereFace [33] (2017)	90.30
VGGFace2 [6] (2018)	90.57
Arcface [10] (2019)	94.45
AA_CNN [32] (2021)	90.7
PMADN [39] (2021)	91.20
MTLFace [19] (2021)	95.62
IEFP [40] (2022)	95.82
Ours	**96.10**

We followed the same protocol as the LFW, in which the dataset is divided into ten separate folds using the same identities contained in the ten folds of the LFW. Each fold contains 300 positive pairs and 300 negative pairs. We evaluated our method on CALFW, and the results are shown in Table 6. The results show that our proposed method is robust and reliable, even with larger age spans.

AgeDB. AgeDB [26] is a face dataset in the wild with large variations in pose, age, illumination, and expression. It contains 16,488 face images of 568 distinct subjects. Every image is annotated manually to achieve noise-free annotation of the age labels. It provides four protocols for age-invariant face verification protocols, wherein the age difference between each pair of faces is fixed to a predefined value, i.e., 5, 10, 20, or 30 years. The evaluation experiments on AgeDB-30 might best demonstrate our model's superiority for large age-span face verification since the 30-year age gap is the most challenging. Similar to the LFW [29], we strictly followed the protocol on AgeDB-30 to perform the 10-fold cross-validation, compute the face verification rate, and compare our results with other state-of-the-art CAFR methods. Table 7 shows the evaluation results from various methods on AgeDB-30. Most methods achieved performance higher than 90%, but the proposed model outperformed the other state-of-the-art CAFR methods.

Table 7. Evaluation results of the different methods on AgeDB-30.

Method	Accuracy (%)
RJIVE [42] (2017)	55.20
VGG Face2 [6] (2018)	89.89
Center Loss [41] (2016)	93.72
SphereFace [33] (2017)	91.70
CosFace [43] (2018)	94.56
Arcface [10] (2019)	95.15
DAAE [44] (2020)	95.30
MTLFace [19] (2021)	96.23
Ours	**98.38**

Figure 10. Samples of false identifications on FG-NET.

FG-NET. FG-NET contains 1,002 face images from 82 subjects with ages ranging from 0 to 69. We experimented with the leave-one-out evaluation scheme adopted by HFA [21] and Li et al. [45] to separate the training and testing data. We selected one image as the testing datum and fine-tuned the model on the other 1,001 face images, repeating that process 1,002 times. Considering that every subject in the dataset has multiple face images of different ages, that evaluation tactic can well reflect the performance of a face-recognition model. Table 8 and Fig. 9(b) show the rank-1 recognition rate comparisons on the FG-NET dataset. Our method achieved good results (94.91%) and outperformed all the other state-of-the-art methods except for IEFP. Unlike our end-to-end model, the IEFP framework also trains an age estimation model, which increases the training time. We visualized some the false identification results on FG-NET dataset in Fig. 10. Note that the false identifications are mainly infants and children from 0–13 years old. As shown in Fig. 5, 51.2% of the images in the benchmark FG-NET are from 0–13 years. Meanwhile, the CACD dataset used to train our model does not include images from that age period, which is disadvantageous for a data-driven-based method trying to learn the latent distributions of that particular age group.

Table 8. Face recognition performance comparison on FG-NET.

Method	Accuracy (%)
Park et al. [46] (2011)	37.4
Li et al. [45] (2011)	47.5
HFA [21] (2013)	69.0
MEFA [47] (2015)	76.2
CAN [48] (2017)	86.5
LFCNN [22] (2016)	88.10
AIM [14] (2020)	93.20
DAL[25] (2019)	94.50
AA_CNN [32] (2021)	89.34
MTLFace [19] (2021)	94.78
IEFP [40] (2022)	**96.21**
Ours	94.91

LFW. The LFW [40] dataset has various images in the wild that vary in age, pose, occlusion, lighting, focus, makeup, resolution, facial expression, gender, race, accessories, background, and photographic quality. We conducted an evaluation experiment on the LFW to validate the generalization ability of our method for general face recognition (GFR). LFW is a standard face verification testing dataset for GFR. It contains 13,233 face

images from 5,749 subjects. We strictly followed the standard protocol of unrestricted labeling of outside data, as in [24,25]. We tested our model on 6,000 face pairs. Table 9 reports the verification rate on the LFW and compares it with other state-of-the-art CAFR methods. Our method outperformed the other state-of-the-art methods by a large margin, demonstrating the strong generalizability of our method.

Table 9. Evaluation results of the different methods on LFW.

Method	Accuracy (%)
CosFace [43] (2018)	99.33
SphereFace [33] (2017)	99.42
OE-CNN[24] (2018)	99.35
DAL [25] (2019)	99.47
MTLFace [19] (2021)	99.52
Ours	**99.82**

5. Conclusions

We have here proposed a new framework for cross-age face recognition called the Age-invariant Features Extraction Network. Because aging seriously degrades the accuracy of face recognition seriously, we introduced a block-based feature decomposition module to obtain discriminative and robust age-invariant identity-related features. In addition, we designed a loss function called the direct sum loss to reduce the redundant components between identity-related features and age-related features. The experiments on publicly available cross-age datasets demonstrate the superiority of our method over that of the existing CAFR approaches. In the future, we will explore a CAFR method that integrates a learning age-invariant identity-related representation task with the face generative method. The generated visual faces could better assist the police in finding missing children and identifying criminals.

Author Contributions: Conceptualization, S.L. and HJ L.; methodology, S.L.; software, S.L.; validation, S.L, formal analysis, S.L.; writing—original draft preparation, S.L.; writing—review and editing, S.L. and HJ.L.; supervision, HJ.L.; project administration, HJ.L.; funding acquisition, HJ.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (GR 2019R1D1A3A03103736).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lochner, S.A. Saving Face: Regulating Law Enforcement's Use of Mobile Facial Recognition Technology & Iris Scans. *Ariz. L. Rev.* **2013**, *55*, 201.
2. Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2014; pp. 1701-1708.
3. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. *Advances in neural information processing systems* **2014**, *27*.

4. Jain, A.K.; Nandakumar, K.; Ross, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern recognition letters* **2016**, *79*, 80-105.
5. Sun, Y.; Liang, D.; Wang, X.; Tang, X. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873* **2015**.
6. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. **2015**.
7. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp. 815-823.
8. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. Vggface2: A dataset for recognising faces across pose and age. In Proceedings of the 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), 2018; pp. 67-74.
9. Wu, X.; He, R.; Sun, Z.; Tan, T. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security* **2018**, *13*, 2884-2896.
10. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019; pp. 4690-4699.
11. Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; Huang, F. Curricularface: adaptive curriculum learning loss for deep face recognition. In Proceedings of the proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020; pp. 5901-5910.
12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp. 770-778.
13. Albert, A.M.; Ricanek Jr, K.; Patterson, E. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic science international* **2007**, *172*, 1-9.
14. Zhao, J.; Cheng, Y.; Cheng, Y.; Yang, Y.; Zhao, F.; Li, J.; Liu, H.; Yan, S.; Feng, J. Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2019; pp. 9251-9258.
15. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 5810-5818.
16. Yang, H.; Huang, D.; Wang, Y.; Jain, A.K. Learning face age progression: A pyramid architecture of gans. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 31-39.
17. Wang, Z.; Tang, X.; Luo, W.; Gao, S. Face aging with identity-preserved conditional generative adversarial networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 7939-7947.
18. Geng, X.; Zhou, Z.-H.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence* **2007**, *29*, 2234-2240.
19. Huang, Z.; Zhang, J.; Shan, H. When age-invariant face recognition meets face age synthesis: A multi-task learning framework. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; pp. 7282-7291.
20. Chen, B.-C.; Chen, C.-S.; Hsu, W.H. Cross-age reference coding for age-invariant face recognition and retrieval. In Proceedings of the European conference on computer vision, 2014; pp. 768-783.
21. Gong, D.; Li, Z.; Lin, D.; Liu, J.; Tang, X. Hidden factor analysis for age invariant face recognition. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2013; pp. 2872-2879.
22. Wen, Y.; Li, Z.; Qiao, Y. Latent factor guided convolutional neural networks for age-invariant face recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp. 4893-4901.

23. Zheng, T.; Deng, W.; Hu, J. Age estimation guided convolutional neural network for age-invariant face recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017; pp. 1-9.
24. Wang, Y.; Gong, D.; Zhou, Z.; Ji, X.; Wang, H.; Li, Z.; Liu, W.; Zhang, T. Orthogonal deep features decomposition for age-invariant face recognition. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018; pp. 738-753.
25. Wang, H.; Gong, D.; Li, Z.; Liu, W. Decorrelated adversarial learning for age-invariant face recognition. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; pp. 3527-3536.
26. Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. Agedb: the first manually collected, in-the-wild age database. In Proceedings of the proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017; pp. 51-59.
27. Zheng, T.; Deng, W.; Hu, J. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197* **2017**.
28. The FG-NET Aging Database. Available online: https://fipa.cs.kit.edu/433_451.php (accessed on Nov. 19. 2021).
29. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Proceedings of the Workshop on faces in'Real-Life'Images: detection, alignment, and recognition, 2008.
30. Du, L.; Hu, H.; Wu, Y. Cycle age-adversarial model based on identity preserving network and transfer learning for cross-age face recognition. *IEEE Transactions on Information Forensics and Security* **2019**, *15*, 2241-2252.
31. Huang, Y.; Chen, W.; Hu, H. Age-puzzle facenet for cross-age face recognition. In Proceedings of the Asian Conference on Computer Vision, 2018; pp. 603-619.
32. Huang, Y.; Hu, H. A parallel architecture of age adversarial convolutional neural network for cross-age face recognition. *IEEE Transactions on Circuits and Systems for Video Technology* **2020**, *31*, 148-159.
33. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 212-220.
34. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13-19 June 2020, 2020; pp. 11531-11539.
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 7132-7141.
36. Rothe, R.; Timofte, R.; Van Gool, L. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision* **2018**, *126*, 144-157.
37. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters* **2016**, *23*, 1499-1503.
38. Ricanek, K.; Tesafaye, T. MORPH: a longitudinal image database of normal adult age-progression. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 10-12 April 2006, 2006; pp. 341-345.
39. Wu, Y.; Du, L.; Hu, H. Parallel multi-path age distinguish network for cross-age face recognition. *IEEE Transactions on Circuits and Systems for Video Technology* **2020**, *31*, 3482-3492.
40. Xie, J.-C.; Pun, C.-M.; Lam, K.-M. Implicit and Explicit Feature Purification for Age-invariant Facial Representation Learning. *IEEE Transactions on Information Forensics and Security* **2022**.
41. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the European conference on computer vision, 2016; pp. 499-515.

42. Zafeiriou, S. Recovering joint and individual components in facial data.
43. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 5265-5274.
44. Li, P.; Huang, H.; Hu, Y.; Wu, X.; He, R.; Sun, Z. Hierarchical face aging through disentangled latent characteristics. In Proceedings of the European Conference on Computer Vision, 2020; pp. 86-101.
45. Li, Z.; Park, U.; Jain, A.K. A discriminative model for age invariant face recognition. *IEEE transactions on information forensics and security* **2011**, *6*, 1028-1037.
46. Park, U.; Tong, Y.; Jain, A.K. Age-invariant face recognition. *IEEE transactions on pattern analysis and machine intelligence* **2010**, *32*, 947-954.
47. Gong, D.; Li, Z.; Tao, D.; Liu, J.; Li, X. A maximum entropy feature descriptor for age invariant face recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp. 5289-5297.
48. Xu, C.; Liu, Q.; Ye, M. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing* **2017**, *222*, 62-71.