

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, CAIT, Jeonbuk National University, Jeonju 54896, Korea; sli88@foxmail.com

² Department of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061000, China

* Correspondence: 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, which 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. Our end-to-end framework learns the age-invariant features directly, which is more convenient and can greatly reduce training complexity compared with existing multi-stage training methods. In addition, we propose a direct sum loss function to reduce the interference of age-related features. Our method achieves a comparable and stable performance. Experimental results demonstrate superior performance on four benchmarked datasets over the state-of-the-art. We obtain the relative improvements of 0.06%, 0.2%, and 2.2% on the cross-age datasets CACD-VS, AgeDB, and CALFW, respectively, and a relative 0.03% improvement on a general dataset LFW.

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*, 4816. <https://doi.org/10.3390/app12104816>

Academic Editors: Erika Ottaviano, Jose Machado, Katarzyna Antosz, Dariusz Mazurkiewicz, Yi Ren, Pierluigi Rea, Rochdi El Abdi, Marina Ranga, Vijaya Kumar Manupati and Emilia Villani

Received: 8 April 2022

Accepted: 6 May 2022

Published: 10 May 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://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 identified 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 Figure 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.

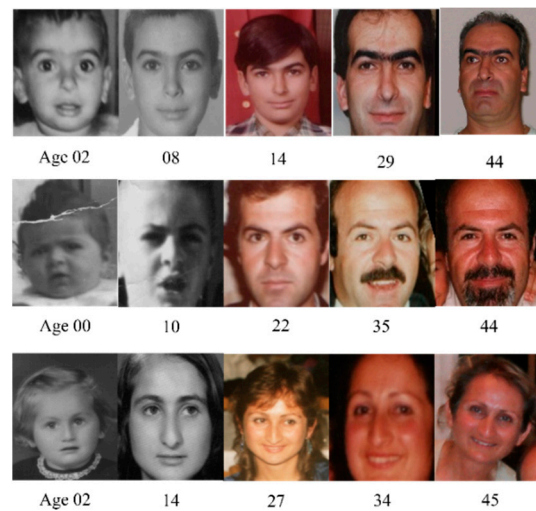


Figure 1. Example images from the FG-NET dataset showing the same person at different ages, illustrating the significant changes caused by facial aging.

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. In addition, various objective functions are required to limit the generator to just creating high-quality 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. The deep features extracted from the face images at different ages usually contain two types of information, which is related to the age and face identity, respectively. Therefore, the cross-age discriminative models focus on how to separate the identity-dependent components from the extracted facial features [19,22–31]. The main limitation of multi-stage training methods is the difficulty in accelerating model convergence. For example, Wang et al. [30] proposed the decorrelated adversarial learning (DAL) algorithm to achieve feature decomposition in an adversarial manner. The alternative training feature extraction and decomposition modules make it difficult to converge. Therefore, we propose an end-to-end training model, including an efficient feature decomposition module and a novel loss function to sufficiently separate age information and identity information from facial features. Another limitation is that existing methods to train a high-performance model usually require a large-scale cross-age face dataset with good age labels and a wide age gap, which most current public databases lack. Therefore, we adopt an FR network, which is pre-trained on the general face dataset MS-Celeb-1M, as a backbone to extract facial features.

Recently, visual attention mechanisms and residual learning have been widely used to solve vision problems such as image classification [32,33]. 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 framework learns the age-invariant features directly, which is more convenient and can greatly reduce training complexity compared with existing multi-stage training methods. As the well pre-trained backbone does not require training, we only train the feature decompose model, which greatly reduces the number of training parameters.
- Based on the attention mechanism, we propose a novel EFDM to separate the age- and identity-related feature on high-level features maps. Age classification and face recognition tasks are incorporated to supervise the decomposition process. On the one hand, by minimizing the direct sum loss between features from the two subnetworks, the face-recognition branch is forced to generate identity-specific facial representations, with the age information removed as much as possible.
- 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 to demonstrate the robustness of the proposed method, and it verifies that the proposed method obtains the relative improvements up to 2.2%.

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 caused by the high complexity of modeling an aged face, and they fail to achieve stable performance.

2.2. Discriminative Approaches

The deep features extracted from face images at different ages usually contain two types of information, which is related to the age and the face identity, respectively. Therefore, the cross-age discriminative models focus on how to separate the identity-dependent components from the extracted facial features. 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. [34] proposed a cycle age-adversarial model that extracts age-invariant features and only uses age labels for training. Huang et al. [35] proposed the Age-Puzzle FaceNet (APFN) based on an adversarial training mechanism to address the CAFR task. Huang et al. [36] 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 [4,21,30]. For example, Gong et al. [21] sepa-

rated identity-related factors and age-related factors using a hidden factor analysis (HFA). Wen et al. [27] developed a latent identity analysis layer to separate the two components. The age-estimation-guided convolutional neural network [28] uses the age estimation task to guide the separation of age features from the identity feature layer. Wang et al. [29] presented feature decomposition in an orthogonal embedding CNN (OE-CNN) and adapted SphereFace loss [37] to deal with the CAFR task. Wang et al. [30] proposed the DAL algorithm to achieve feature decomposition in an adversarial manner. Wu et al. [38] divided face features into groups and then recombined them to create identity-dependent feature representations that are resistant to age progression. Xie et al. [31] proposed a purification unit to remove the irrelevant age information and retain the identity information only.

In mathematics, the direct sum of two abelian groups of identity 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.

3. Proposed Method

This section provides a detailed description of the proposed method. The AFEN framework is shown in Figure 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 face feature decomposition. 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.

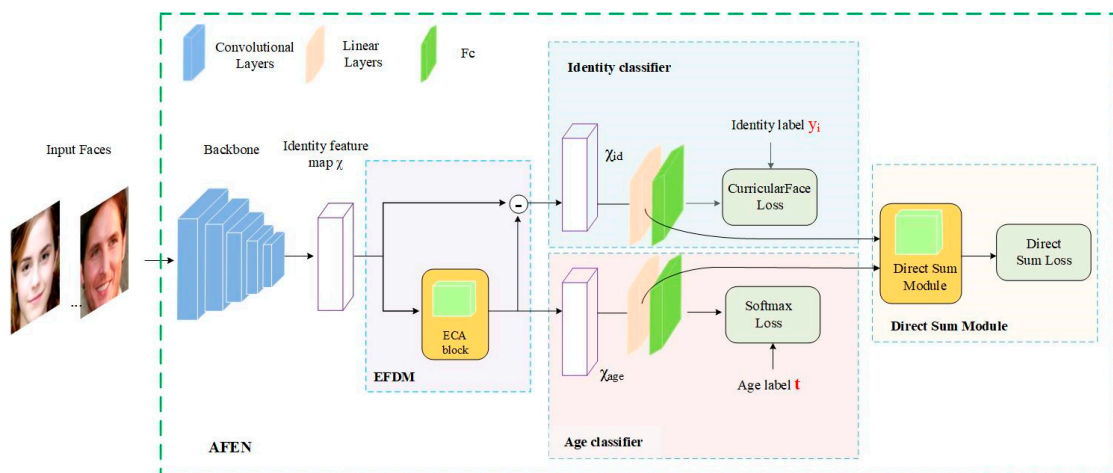


Figure 2. Overall framework of the proposed AFEN and its training process.

3.1. Feature Decomposition Module

As the face 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 from an input image. The linear factorization can be defined in Equation (1).

$$x = x_{id} + x_{age} \quad (1)$$

where x_{id} and x_{age} denote the identity-dependent factor and age-dependent factor of the facial feature vector, 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. As 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. As the parameter and FLOPs overhead of the ECA block [39] are much smaller than the squeeze and excitation block [40], we use the ECA block to compute and apply attention weights to the channels of the input feature map.

The ECA block is an extremely lightweight channel attention module for deep CNNs. The ECA block first uses global average pooling (GAP) for each channel independently to aggregate the feature map $\chi \in \mathbb{R}^{C \times H \times W}$, where C, H, W denote the channel, height, width, respectively. 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, $\chi^{age} \in \mathbb{R}^{C \times H \times W}$, is obtained by channel-wise multiplication between the channel weights and the feature map χ . As illustrated in Figure 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.

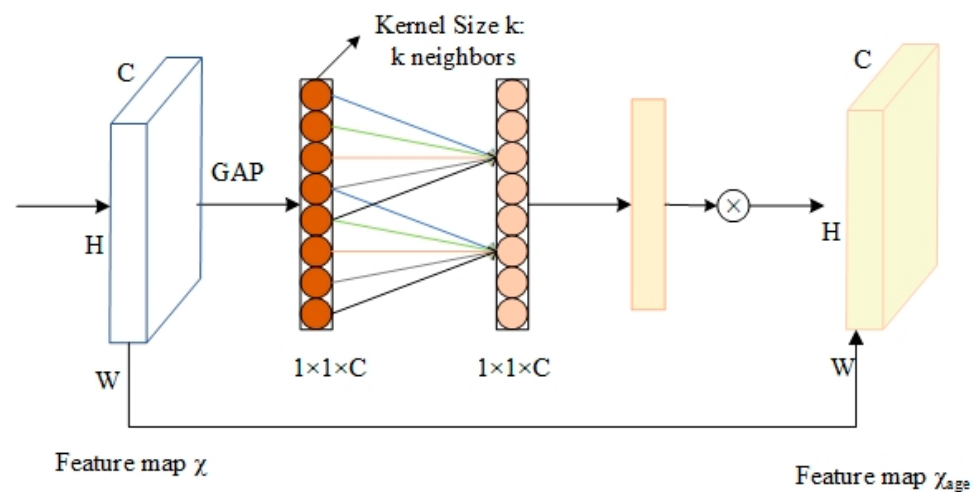


Figure 3. Diagram of the ECA block.

In Figure 3, the backbone is used to extract a face feature map, $\chi \in \mathbb{R}^{C \times H \times W}$, from the input image. In the decomposition module, we use the ECA block to transform the face feature map, χ , into an age-specific feature map, χ_{age} . The decomposition process can be defined in Equation (2).

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

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

The feature decomposition module based on the ECA block is built to obtain the identity-specific feature and age-specific feature. To achieve a significant separation of the identity-specific and age-specific features, the direct sum loss is proposed, 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 1. Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two subspaces of linear space $\mathcal{V}$. Let $(\alpha_1, \alpha_2, \dots, \alpha_s)$ be basis vectors of $\mathcal{W}_1$ and $(\beta_1, \beta_2, \dots, \beta_t)$ be basis vectors of $\mathcal{W}_2$. If $(\alpha_1, \alpha_2, \dots, \alpha_s, \beta_1, \beta_2, \dots, \beta_t)$ 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, the direct sum loss constraint is designed 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.

The basis vectors $(v_{id1}, v_{id2}, \dots, v_{idK})$ are obtained from a fully connected layer. The space formed by the basis vectors $(v_{id1}, v_{id2}, \dots, v_{idK})$ is marked as V_{II} . In the same way, the space V_{AA} is obtained by the basis vectors $(v_{age1}, v_{age2}, \dots, v_{ageK})$. To make the space $V_{II} + V_{AA}$ meet the direct sum condition, $(v_{id1}, v_{id2}, \dots, v_{idK}, v_{age1}, v_{age2}, \dots, v_{ageK})$ must be linearly independent. Therefore, the direct sum loss is represented in Equation (3).

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

where K denotes the number of basis vectors. The cosine similarity is as close to 0 as possible, making the two basis vectors linearly independent.

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 Figure 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 Figure 5.

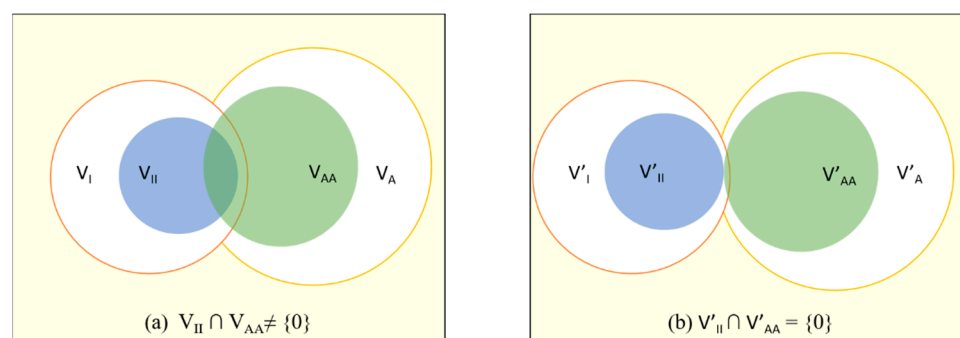


Figure 4. Direct sum. (a) Before applying the subspace direct sum constraint. (b) After applying the subspace direct sum constraint.

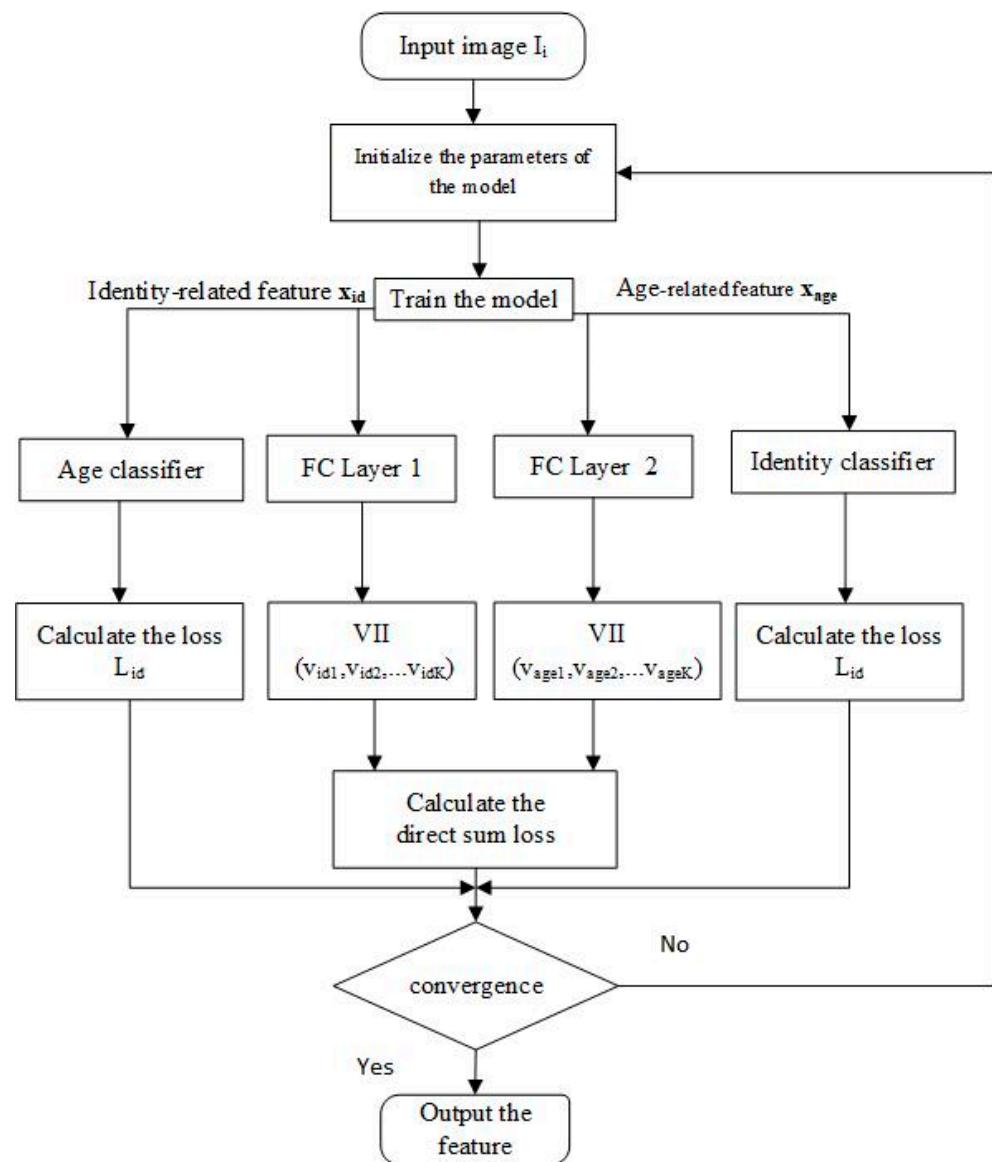


Figure 5. The flow chart of the training process. The whole loss is obtained through the identity classifier, age classifier, and direct sum modules.

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 in Equations (4) and (5).

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

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

where N is the number of training samples, x_i is the i th 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 j th 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], the faces of different ages are divided 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 in Equation (6).

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

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 in Equation (7).

$$L_{total} = L_{Curr_Face} + \lambda_1 L_{direct_sum} + \lambda_2 L_{age} \quad (7)$$

where λ_1 and λ_2 are scalar hyperparameters to balance the three losses. Both weights λ_1 and λ_2 are set to 0.01 after the experimental analysis. The model training process is summarized Figure 5.

4. Experiments

4.1. Implementation Details

Datasets. Several public cross-age datasets were used for model training and evaluation: cross-age celebrity dataset (CACD) [20], CACD verification subset (CACD-VS) [20], AgeDB [41], CALFW (cross-age labeled face in the wild) [42], and face and gesture recognition network (FG-NET) [43] dataset. We used some of the CACD dataset to train our model, and the rest of them were used for evaluation. The distribution of ages in the CACD and FG-NET datasets is shown in Figure 6. The CACD dataset is used as a public benchmark for CAFR, and it is composed of 163,446 images of 2000 celebrities. The images reflect various shooting conditions, such as illumination variations, pose variations, age variations, makeup, and practical scenarios. FG-NET contains 1002 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.

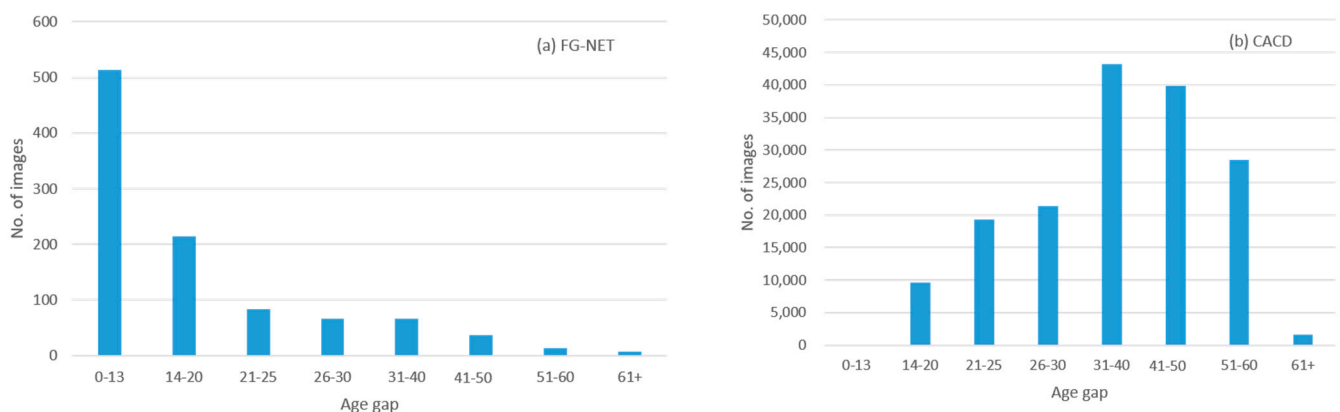


Figure 6. The distribution of ages in two of the datasets.

Therefore, we choose the CACD dataset as the training dataset. We randomly selected 80% of its images as the training data (130,757 images) and used the remaining 20% for validation (32,689 images). 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 [44] method to produce age labels as the ground truth. To obtain better training results, we also manually removed duplicate images.

We conducted experiments on commonly used, public, cross-age datasets: CACD-VS, CALFW, AgeDB, and FG-NET. We extracted only the identity-specific feature as the final face 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 [45] 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. Figure 7 shows some original and preprocessed face images from the CACD dataset.



Figure 7. Examples of the data used. The top row shows the original images, and the bottom row shows the aligned and normalized images.

Training protocols. As 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 8500 subjects across pose and age. The pre-trained model can classify tens of thousands of identities and extract multilevel, high-resolution features.

We initialized the shared model with the pre-trained model and then trained the feature to decompose the 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 5×10^{-4} . We selected the hyperparameters by trial and error. The training process was finished at 30 epochs of 9.57 K 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. Figure 8 shows the variation trend for training loss with the optimal parameters.

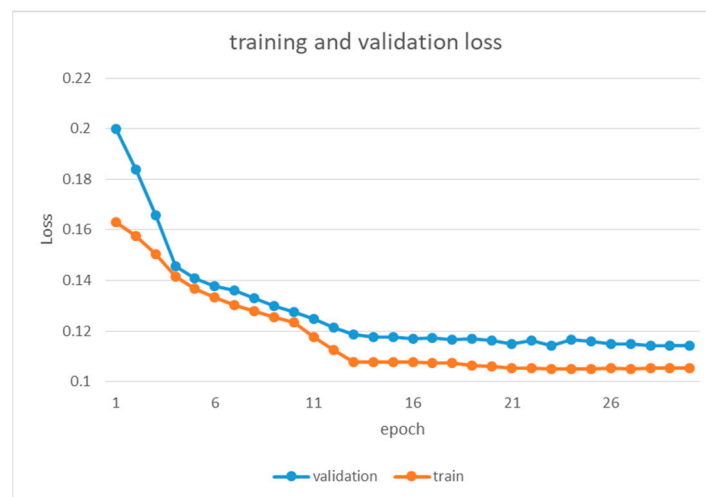


Figure 8. Trend of training and validation loss.

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 Equation (7), and K in Equation (3). 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 $\lambda_2 = 0.01$, and then set λ_2 as {1, 0.1, 0.01, 0.001} while $\lambda_1 = 0.01$ 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 Figure 9, which indicates that the best performance was obtained when $\lambda_1 = 0.01$ and $\lambda_2 = 0.01$.

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.

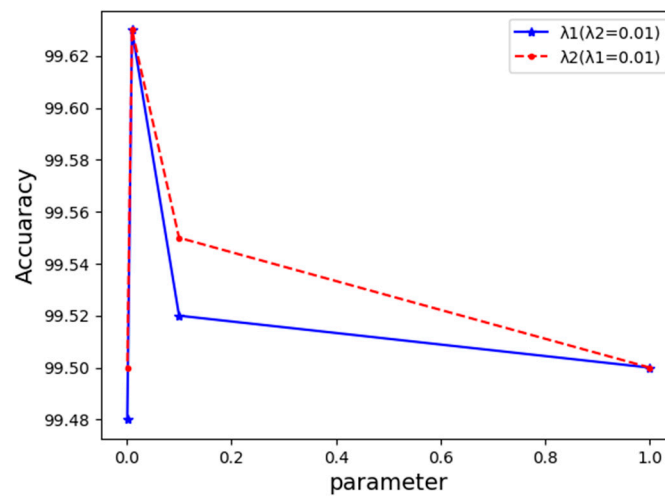


Figure 9. Face verification accuracy on the CACD-VS dataset with various values for λ_1 and λ_2 .

Table 2. Evaluation results (%) on the CACD-VS dataset. The bold represents the best value.

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 Equation (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	Curricular Face
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 [30], 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 [30] costs 0.963 s for each iteration on the Nvidia Titan X Pascal GPUs, whereas our method costs only 0.416 s. 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 [30]	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, CALFW, AgeDB, and FG-NET, and a general face dataset, LFW [46]. Note that MORPH [47] 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 Figure 10a, 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 the 1:k identification system performance. The evaluation schemes for the different datasets are described next.

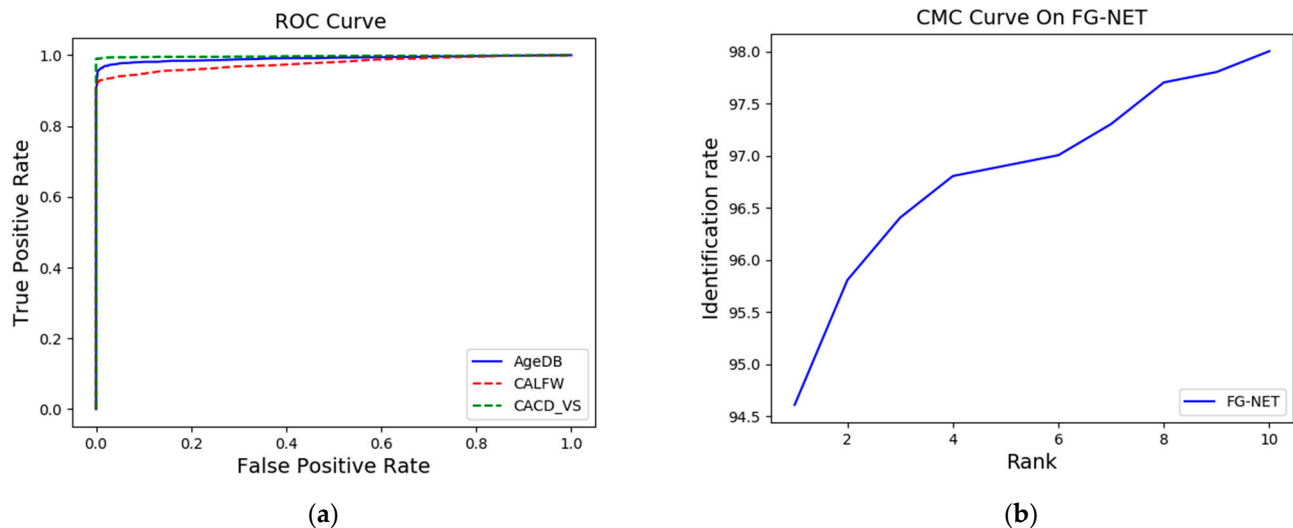


Figure 10. Face recognition performance of our model (a) ROC curves on CACD-VS, CALFW, AgeDB-30. (b) CMC curve on FG-NET.

CACD-VS. The CACD-VS consists of 4000 face image pairs for face verification, 2000 positive and 2000 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 [43]. We strictly followed the cross-validation rule [29] 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.

Table 5. Evaluation results of the different methods on CACD-VS. The bold represents the best value.

Method	Accuracy (%)
LFCNN [27] (2016)	98.5
OE-CNN [29] (2018)	99.20
AIM [14] (2020)	99.38
DAL [30] (2019)	99.40
AA_CNN [36] (2021)	99.20
PMADN [38] (2021)	99.20
MTLFace [19] (2021)	99.55
IEFP [31] (2022)	99.57
Ours	99.63

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.

CALFW. To demonstrate the effectiveness of our method for face recognition with a larger age span, we implemented a face verification experiment on the CALFW dataset. The CALFW dataset is an extension of the LFW dataset designed for unconstrained face verification with larger age gaps. First, 3000 positive face pairs with age gaps were 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 3000 negative pairs with the same gender and race were selected to reduce the influence of different attributes [36].

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.

Table 6. Evaluation results of the different methods on CALFW. The bold represents the best value.

Method	Accuracy (%)
Center Loss [48] (2016)	85.48
SphereFace [37] (2017)	90.30
VGGFace2 [6] (2018)	90.57
Arcface [10] (2019)	94.45
AA_CNN [36] (2021)	90.7
PMADN [38] (2021)	91.20
MTLFace [19] (2021)	95.62
IEFP [31] (2022)	95.82
Ours	96.10

AgeDB. AgeDB [41] 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 [46], 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 [49] (2017)	55.20
VGG Face2 [6] (2018)	89.89
Center Loss [48] (2016)	93.72
SphereFace [37] (2017)	91.70
CosFace [50] (2018)	94.56
Arcface [10] (2019)	95.15
DAAE [51] (2020)	95.30
MTLFace [19] (2021)	96.23
Ours	98.38

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 [43]

and Li et al. [52] to separate the training and testing data. We selected one image as the testing datum and fine-tuned the model on the other 1001 face images, repeating that process 1002 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 Figure 10b 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 IEF. Unlike our end-to-end model, the IEF framework also trains an age estimation model, which increases the training time. We visualized some of the false identification results on FG-NET dataset in Figure 11. Note that the false identifications are mainly infants and children from 0 to 13 years old. As shown in Figure 6, 51.2% of the images in the benchmark FG-NET are from 0 to 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. The bold represents the best value.

Method	Accuracy (%)
Park et al. [53] (2011)	37.4
Li et al. [52] (2011)	47.5
HFA [43] (2013)	69.0
MEFA [54] (2015)	76.2
CAN [55] (2017)	86.5
LFCNN [27] (2016)	88.10
AIM [14] (2020)	93.20
DAL [30] (2019)	94.50
AA_CNN [36] (2021)	89.34
MTLFace [19] (2021)	94.78
IEFP [31] (2022)	96.21
Ours	94.91

LFW. The LFW [18] 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 5749 subjects. We strictly followed the standard protocol of unrestricted labeling of outside data, as in [29,30]. We tested our model on 6000 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.

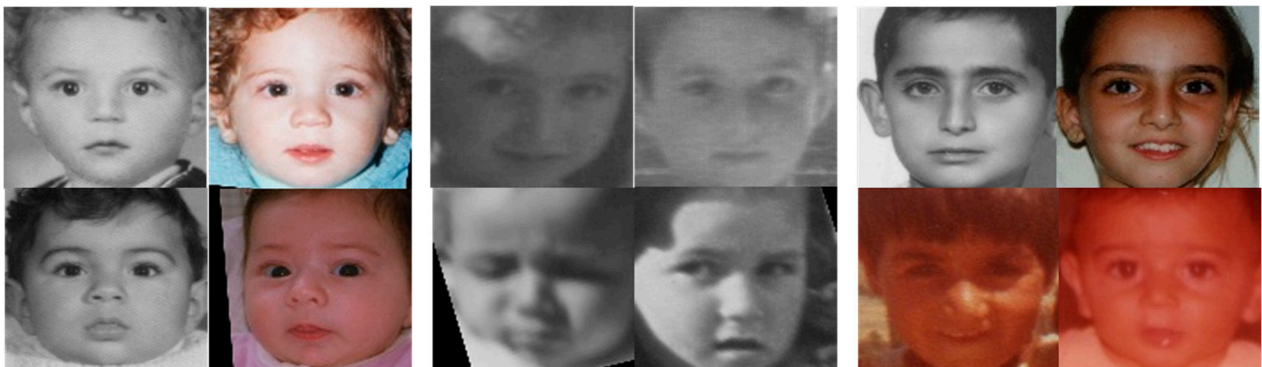


Figure 11. Samples of false identifications on FG-NET.

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

Method	Accuracy (%)
CosFace [50] (2018)	99.33
SphereFace [37] (2017)	99.42
OE-CNN [29] (2018)	99.35
DAL [30] (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. As 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. Extensive ablation studies have demonstrated that our method is more convenient and achieves performance improvements. We obtained the relative improvements of 0.06%, 0.2%, 2.2%, 0.03% on the datasets CACD-VS, AgeDB, CALFW and LFW, respectively. The experiments on publicly available cross-age datasets demonstrate the superiority of our method over the state-of-the-art methods. As in practical applications, the generated visual faces could directly assist the police in finding missing children and identifying criminals. In future works, we explore a CAFR method that integrates the proposed learning age-invariant identity-related representation task with the face generative method.

Author Contributions: Conceptualization, S.L. and H.J.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 H.J.L.; supervision, H.J.L.; project administration, H.J.L.; funding acquisition, H.J.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 conflict of interest.

References

- Lochner, S.A. Saving Face: Regulating Law Enforcement’s Use of Mobile Facial Recognition Technology & Iris Scans. *Ariz. L. Rev.* **2013**, *55*, 201.
- Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
- Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. [CrossRef]
- Jain, A.K.; Nandakumar, K.; Ross, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognit. Lett.* **2016**, *79*, 80–105. [CrossRef]
- Sun, Y.; Liang, D.; Wang, X.; Tang, X. Deepid3: Face recognition with very deep neural networks. *arXiv* **2015**, arXiv:1502.00873.
- Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.
- Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
- 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), Xi’an, China, 15–19 May 2018; pp. 67–74.

9. Wu, X.; He, R.; Sun, Z.; Tan, T. A light cnn for deep face representation with noisy labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [[CrossRef](#)]
10. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–18 June 2020; pp. 5901–5910.
12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
13. Albert, A.M.; Ricanek, K., Jr.; Patterson, E. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Sci. Int.* **2007**, *172*, 1–9. [[CrossRef](#)]
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 AAAI Conference on Artificial Intelligence, Online, 5 January 2019; pp. 9251–9258.
15. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 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 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 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 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7939–7947.
18. Geng, X.; Zhou, Z.-H.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240. [[CrossRef](#)]
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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 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 *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 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 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 2872–2879.
22. Du, L.; Hu, H.; Wu, Y. Age factor removal network based on transfer learning and adversarial learning for cross-age face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 2830–2842. [[CrossRef](#)]
23. Li, H.; Hu, H.; Yip, C. Age-related factor guided joint task modeling convolutional neural network for cross-age face recognition. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2383–2392. [[CrossRef](#)]
24. Du, L.; Hu, H. Cross-age identity difference analysis model based on image pairs for age invariant face verification. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 2675–2685. [[CrossRef](#)]
25. Meng, L.; Yan, C.; Li, J.; Yin, J.; Liu, W.; Xie, H.; Li, L. Multi-features fusion and decomposition for age-invariant face recognition. In Proceedings of the 28th ACM International Conference on Multimedia, Online, 12 October 2020; pp. 3146–3154.
26. Shakeel, M.S.; Lam, K.-M. Deep-feature encoding-based discriminative model for age-invariant face recognition. *Pattern Recognit.* **2019**, *93*, 442–457. [[CrossRef](#)]
27. Wen, Y.; Li, Z.; Qiao, Y. Latent factor guided convolutional neural networks for age-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4893–4901.
28. Zheng, T.; Deng, W.; Hu, J. Age estimation guided convolutional neural network for age-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1–9.
29. 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 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 738–753.
30. Wang, H.; Gong, D.; Li, Z.; Liu, W. Decorrelated adversarial learning for age-invariant face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3527–3536.
31. Xie, J.-C.; Pun, C.-M.; Lam, K.-M. Implicit and Explicit Feature Purification for Age-invariant Facial Representation Learning. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 399–412. [[CrossRef](#)]
32. Kwon, S. MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach. *Expert Syst. Appl.* **2021**, *167*, 114177.
33. Kwon, S. Att-Net: Enhanced emotion recognition system using lightweight self-attention module. *Appl. Soft Comput.* **2021**, *102*, 107101.
34. Du, L.; Hu, H.; Wu, Y. Cycle age-adversarial model based on identity preserving network and transfer learning for cross-age face recognition. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 2241–2252. [[CrossRef](#)]
35. Huang, Y.; Chen, W.; Hu, H. Age-puzzle facenet for cross-age face recognition. In *Asian Conference on Computer Vision*; Springer: Cham, Switzerland, 2018; pp. 603–619.

36. Huang, Y.; Hu, H. A parallel architecture of age adversarial convolutional neural network for cross-age face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 148–159. [[CrossRef](#)]
37. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 212–220.
38. Wu, Y.; Du, L.; Hu, H. Parallel multi-path age distinguish network for cross-age face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 3482–3492. [[CrossRef](#)]
39. 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), Virtual, 13–19 June 2020; pp. 11531–11539.
40. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
41. 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 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 51–59.
42. Zheng, T.; Deng, W.; Hu, J. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv* **2017**, arXiv:1708.08197.
43. The FG-NET Aging Database. Available online: https://fipa.cs.kit.edu/433_451.php (accessed on 19 November 2021).
44. Rothe, R.; Timofte, R.; Van Gool, L. Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [[CrossRef](#)]
45. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
46. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Proceedings of the Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition, Marseille, France, 12–18 October 2008.
47. 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), Southampton, UK, 10–12 April 2006; pp. 341–345.
48. 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, Online, 16 September 2016; pp. 499–515.
49. Zafeiriou, S. *Recovering Joint and Individual Components in Facial Data*; IEEE: Piscataway, NJ, USA, 2017.
50. 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 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5265–5274.
51. 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, Online, 3 December 2020; pp. 86–101.
52. Li, Z.; Park, U.; Jain, A.K. A discriminative model for age invariant face recognition. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 1028–1037. [[CrossRef](#)]
53. Park, U.; Tong, Y.; Jain, A.K. Age-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 947–954. [[CrossRef](#)] [[PubMed](#)]
54. Gong, D.; Li, Z.; Tao, D.; Liu, J.; Li, X. A maximum entropy feature descriptor for age invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5289–5297.
55. Xu, C.; Liu, Q.; Ye, M. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing* **2017**, *222*, 62–71. [[CrossRef](#)]