Improving Face Recognition with Large Age Gaps by Learning to Distinguish Children

Jungsoo Lee*[1]
bebeto@kaist.ac.kr

Jooyeol Yun*[1]
blizzard072@kaist.ac.kr

Sunghyun Park[1]
psh01087@kaist.ac.kr

Yonggyu Kim[2]
dydrb2535@kaist.ac.kr

Jaegul Choo[1]
jchoo@kaist.ac.kr

[1] KAIST AI,
South Korea

[2] Korea University,
South Korea

Abstract

Despite the unprecedented improvement of face recognition, existing face recognition models still show considerably low performances in determining whether a pair of child and adult images belong to the same identity. Previous approaches mainly focused on increasing the similarity between child and adult images of a given identity to overcome the discrepancy of facial appearances due to aging. However, we observe that reducing the similarity between child images of different identities is crucial for learning distinct features among children and thus improving face recognition performance in child-adult pairs. Based on this intuition, we propose a novel loss function called the Inter-Prototype loss which minimizes the similarity between child images. Unlike the previous studies, the Inter-Prototype loss does not require additional child images or training additional learnable parameters. Our extensive experiments and in-depth analyses show that our approach outperforms existing baselines in face recognition with child-adult pairs. Our code and newly-constructed test sets of child-adult pairs are available at this link [1].

1 Introduction

Although recent face recognition models achieve remarkable improvements [4, 10, 13, 17, 18, 19, 21], they still show low performance in deciding whether child-adult pairs match to the same identity. Performing face recognition with child-adult pairs requires models to consider the changes of facial appearances of a given identity (*e.g.* added wrinkles or changes of chin shapes) which occurs due to the aging process. Such issue needs to be resolved when considering its application such as finding missing children which requires face recognition models to match childhood and grown-up images of a given identity.

This task remains challenging due to the fact that the child images are excessively scarce in publicly available datasets (*e.g.* CASIA-WebFace [23] and MS1MV2 [2]), as shown in

---

[1]https://github.com/leebebeto/Inter-Prototype
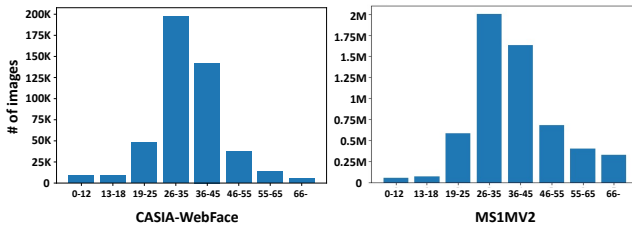[2]MS1MV2 is a refined version of the MS-Celeb-1M dataset. [6]

Figure 1: Pseudo-labeled age distribution of the CASIA-WebFace dataset [23] and MS1MV2. The small number of child images in the dataset limits the model in learning the representation of child images.

Fig. 1. It requires a non-trivial amount of time to collect both childhood and adulthood images of an identity when building a new dataset. Even worse, we empirically find that child images generally share appearances such as big eyes, rounded shape of chins, and short hair, which inhibits the face recognition models from extracting the distinct features among child images [20]. These reasons altogether make it challenging for models to learn the representation of children and verify pairs of child and adult images. [1].

Recent face recognition models utilize various loss functions [4, 10, 12, 19, 21] in order to increase the similarity of features extracted from images within an identity (*i.e.*, intra-class similarity). Through an in-depth analysis, we observe that existing loss functions encourage the models to increase the intra-class similarity to a reasonable level. However, we find that the similarities between different classes (*i.e.*, inter-class similarity) among child images are substantially higher than those of adult images, which hinders the model in learning the discriminative features of a child image (Section. 3).

In this work, we propose a novel loss function called the Inter-Prototype loss which minimizes the inter-class similarity between identities including child images. With the reduced inter-class similarity, our approach enables the models to learn the discriminative features of a child image, which improves the face recognition performance in child-adult pairs. Compared to the previous age-invariant face recognition models [11, 20, 22, 24, 25] which require a substantial amount of the child images in the training dataset, our approach shows superior performances over existing baselines even with the limited number of child images in the training dataset. Our contributions are as follows:

- We provide an in-depth analysis on existing approaches and reveal that reducing the inter-class similarity between children is crucial for improving face recognition including child-adult pairs.

- We propose a novel loss function called the Inter-Prototype loss which minimizes the inter-class similarity of the identities including child images in a *simple yet effective* manner.

- Our approach outperforms the existing baselines on face verification/identification of child-adult pairs with large age gaps. We also release the test sets of child-adult pairs.

2 Related Work

Recent advancements of face recognition models [4, 13, 19] follow the line of work where models learn the feature embeddings through training a classifier. These studies interpret the weight of the last fully-connected layer of a classifier as the representative vectors for each identity in the train set, which we term as *prototype vectors*. They utilize the margin-based loss functions [4, 13, 19] that assign an additional penalty to the angles between extracted
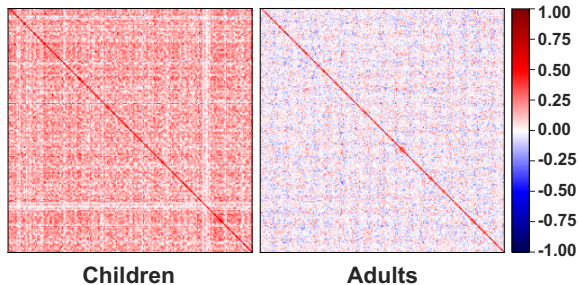
Figure 2: Intra- and inter-class similarity of child and adult images. The diagonal and off-diagonal cells indicate the intra- and inter-class similarity, respectively.

features and their corresponding prototype vectors. Such margin penalty using the prototype vectors increases the intra-class similarity. Recent studies [10, 21] further improve the margin-based loss functions by assigning higher weights on the loss for the misclassified samples. While these methods improve the face recognition performance in general datasets [9, 16], they still show considerably low performance in verifying child-adult pairs.

Several age-invariant face recognition models [11, 20, 22, 24, 25] try to tackle such an issue. AE-CNN [25] and OE-CNN [22] add an auxiliary age classifier to an identity classifier in order to eliminate the age-related information from the feature vectors. Wang *et al*. present DAL [20] to disentangle the feature vectors into age and identity vectors through an adversarial learning approach. While these approaches focus on obtaining the age-invariant identity features by modifying the architecture of classification models, Huang *et al*. propose MTLFace [11] which leverages generative models and uses the representation obtained from the generative models for the face recognition. Although the methods mentioned above improve the performance in age-invariant face recognition, they bear several limitations. First, they heavily rely on the additionally collected cross-age images. This can not be a general solution since their performance is only guaranteed with the newly built train sets, which are currently not publicly available. Second, introducing an additional generative model increases the training time and the computational cost [11]. In this regard, we propose a *simple yet effective method* which does not require additional child images or learnable parameters.

3 Analysis on the Inter-class Similarity

In order to examine the representation of child images, we analyze the intra-class similarity and the inter-class similarity of child and adult images in the CASIA-WebFace dataset [23] using a pretrained ArcFace [4] model. Since the CASIA-WebFace dataset does not include age labels for each image, we utilize the pseudo labels provided by Huang *et al*. [11] who annotated the labels using the public Azure Facial API [2]. Inspired by Wang *et al*. [20], we group the pseudo age labels into 8 ranges: 0-12, 13-18, 19-25, 26-35, 36-45, 46-55, 56-65, and ≥ 66. We empirically observe the images and define the images in the age group of 0-12 as child images considering their facial appearances. The distribution of the pseudo-labeled ages in CASIA-Webface dataset [23] is presented in Fig. 1.

Given the number of identities n and the dimension of a feature vector d, the weight matrix of the last fully-connected layer in a classification model $\mathbf{W} \in \mathbb{R}^{d \times n}$ can be interpreted as a set of vectors that represent each class (*i.e.*, an identity) [13, 19]. As aforementioned in Section. 2, we term such representative vector of the i-th identity in the last fully-connected layer $\mathbf{W}$ as *prototype vector*, formulated as $\mathbf{W}_i \in \mathbb{R}^{d \times 1}$. In this paper, the *intra-class similarity* of an identity is defined as the averaged cosine similarity between feature vectors

extracted from images and their corresponding prototype vectors, which is formulated as

$$\mathbf{S}_i^{\text{intra}} = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{W}_i^{\mathsf{T}} f\left(\mathbf{x}_{i_k}\right), \tag{1}$$

where N_i is the number of training images in the i-th identity, $\mathbf{x}_{i_k}$ is the k-th input image of the i-th identity, and $f\left(\mathbf{x}_{i_k}\right)$ indicates the extracted feature vectors of $\mathbf{x}_{i_k}$. Also, *Inter-class similarity* between two identities refers to the averaged cosine similarity between feature vectors extracted from images of different classes, which is described as

$$\mathbf{S}_{ij}^{\text{inter}} = \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} f(\mathbf{x}_{i_k})^{\mathsf{T}} f(\mathbf{x}_{j_l}), \text{ where } i \neq j. \tag{2}$$

Fig. 2 visualizes the averaged the intra- and inter-class similarity using child and adult images, respectively. Each row and column indicate the index of 200 randomly sampled identities, respectively. The diagonal cells indicate the intra-class similarity. Each off-diagonal cell indicates the inter-class similarity of all image pairs between two given identities (*i.e.*, each from the row and the column, respectively).

There are two observations from the visualization. We observe that the intra-class similarity is both high in child and adult images (red-colored diagonals). This indicates that the ArcFace [4] alone can properly increase the intra-class similarity during the training phase. On the other hand, the inter-class similarities between child images are substantially higher (colored as red) than those of adult images (colored as white or blue). Such visualization indicates that the model did not fully learn to distinguish children. This makes it challenging for the model to extract child representation, which results low performance in verifying child-adult pairs. The feature vectors used for the visualization are obtained from the training dataset, so the inter-class similarities should have been minimized since they are explicitly supervised from the ArcFace loss. However, we observe that the inter-class similarities between child images are still high. Based on the two observations, despite the high intra-class similarity achieved by the model, we conclude that the poor performance on child-adult pairs is derived from the lack of discriminative features obtained from child images due to the high inter-class similarity. In this regard, we propose a novel loss function called Inter-Prototype loss to mitigate this issue.

4 Proposed Approach

4.1 Margin-based loss

Margin-based loss aims to increase the cosine similarity between feature vectors and their corresponding prototype vectors by utilizing margin terms, which penalizes the angles between them. The margin-based loss function we use throughout our experiment is the ArcFace [4], which is written as follows,

$$\mathcal{L}_m = \frac{1}{N} \sum_{i=1}^{N} -\log \frac{e^{s\left(\cos(\theta_{y_i,i}+m)\right)}}{e^{s\left(\cos(\theta_{y_i,i}+m)\right)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_{j,i}}}, \tag{3}$$

$$\cos\theta_{ji} = \frac{\mathbf{W}_j^{\mathsf{T}} f(\mathbf{x}_i)}{\|\mathbf{W}_j\|\|f(\mathbf{x}_i)\|}, \tag{4}$$

where N is the batch size, n is the number of classes, y_i is the ground-truth class, $\cos\theta_{j,i}$ is the cosine similarity between $\mathbf{W}_j$ and x_i, s is the scale factor, and m is the margin.
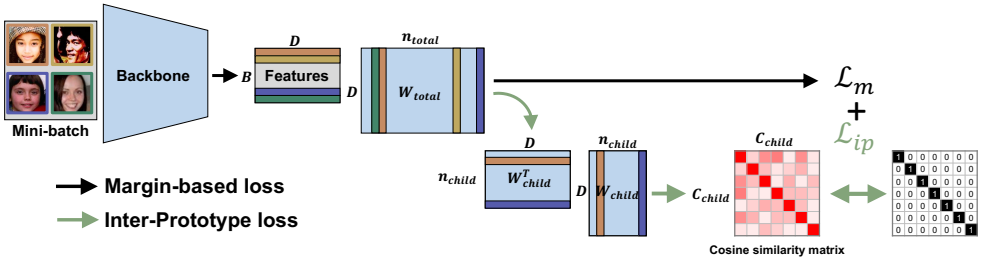
Figure 3: Overview of our method. We add our Inter-Prototype loss to an existing margin-based loss function [4]. Among all identities, we obtain the identities which include child images (n_{child}). Then, we compute a cosine similarity matrix of the prototype vectors using those identities. We enforce the cosine similarity matrix of the prototype vectors to become an identity matrix.

4.2 Inter-Prototype loss

The Inter-Prototype loss is not imposed on the feature vectors extracted from the backbone model, but rather enforced on the prototype vectors. Fig. 3 illustrates the workflow of our method. We collect the prototype vectors of classes that contain the child images denoted as $\mathbf{W}_{child} \in \mathbb{R}^{d \times n_{child}}$, where n_{child} is the number of classes containing the child images.

After normalizing the prototype vectors of $\mathbf{W}_{child}$, we obtain the cosine similarity matrix $\mathbf{C} \in \mathbb{R}^{n_{child} \times n_{child}}$ using the prototype vectors. The Inter-Prototype loss focuses on minimizing the inter-class similarity between identities including child images which is formulated as,

$$\mathcal{L}_{ip} = \sum_{i=1}^{n_{child}} \sum_{j \neq i} \mathbf{C}_{ij}^2, \tag{5}$$

$$\mathbf{C}_{ij} = \frac{\mathbf{W}_{child_i}^{\mathsf{T}} \mathbf{W}_{child_j}}{\|\mathbf{W}_{child_i}\| \|\mathbf{W}_{child_j}\|}, \mathbf{W}_{child_i} \in \mathbb{R}^{d \times 1}. \tag{6}$$

Since we add the Inter-Prototype loss on the margin-based loss function [4], our final objective function is described as

$$\mathcal{L}_{total} = \mathcal{L}_m + \lambda_{ip} \mathcal{L}_{ip}, \tag{7}$$

where $\mathcal{L}_m$ refers to a margin-based loss function [4, 13, 19] and λ_{ip} is a hyper-parameter to balance between the two loss functions which is set to 1. One possible solution for reducing the inter-class similarity between the child identities would be directly sampling the child image pairs and reducing their similarities in the embedding space. However, this approach requires sampling additional child image pairs in a single batch, which is sensitive to the sampling. Even without such additional sampling of child images, adding the Inter-Prototype effectively minimizes the inter-class similarities between child images.

Algorithm 1 explains the overall training procedure of our approach with the pseudo code. The Inter-Prototype loss improves the face recognition accuracy on child-adult pairs in two aspects. First, we can improve the capability of our backbone network to distinguish the difference between children. The high inter-class similarity between different child images indicates that the model is not able to discriminate between two different children, implying that the representation of child images is not formed to a reasonable level. The Inter-Prototype loss addresses such an issue by explicitly minimizing the inter-class similarity between identities including the child images, improving the representation of child

images overall. Second, adding the Inter-Prototype loss also minimizes the inter-class similarity between *child and adult* images of different identities. The prototype vector used for the Inter-Prototype loss is a representative vector for both child and adult images of a given identity. Since we supervise such prototype vectors to be distinct from each other, we are also minimizing the inter-class similarity between child and adult pairs of different identities, as shown in Fig. 4. Additional to such intuition, Inter-Prototype loss is applicable to existing margin-based models as long as the last fully-connected layer can be interpreted as a set of prototype vectors. Moreover, it is insensitive to the number of child images in a mini-batch and does not require additional learnable parameters.

5 Experiments

5.1 Datasets

For the training dataset, we separately conduct experiments using the CASIA-WebFace dataset [23] and the MS1MV2 dataset in order to evaluate our proposed method both in the small and the large dataset. The CASIA-WebFace dataset consists of 0.5M images for 10K individuals (*i.e.*, small evaluation protocol) while the MS1MV2 dataset includes 5M images for 85K individuals (*i.e.*, large evaluation protocol). As aforementioned, we utilize the age labels annotated from MTLFace [11] for both datasets to specify child images. We could not train our model with the newly built datasets with additional child images, as done in the previous age-invariant face recognition studies [11, 20, 22], since they are not publicly available.

Face verification is the task of determining whether a pair of facial images belong to the same identity or not. For the face verification with child-adult pairs, we build the child-adult pairs using the original cross-age datasets, FG-NET [6] and AgeDB [14]. We pair the images of children (*i.e.*, under age 13) and adults with an age gap of more than 20 years and 30 years. We name these datasets as FG-NET-C20, AgeDB-C20, FG-NET-C30, and AgeDB-C30 in our work. The LAG dataset [3], a publicly available test dataset which is composed of images with large age gaps, is also used for the evaluation. Since the LAG dataset does not have age labels, we do not consider the age gaps for pairing the images. All test sets have the equal number of positive and negative pairs. We also compare our approach with previous age-invariant face recognition models [11, 20, 22] using the existing general

---

Algorithm 1: Training with Inter-Prototype Loss

Input: image x, label y, iteration i, learning rate α, margin-based loss $\mathcal{L}_m$, Inter-Prototype loss $\mathcal{L}_{ip}$, identity matrix $\mathbf{I}$, backbone network $f(\cdot)$, backbone network parameters Θ, and prototype vectors $\mathbf{W}$

while *not converged* **do**

    Extract feature vectors $f(x_i)$

    Calculate the cosine similarity matrix $\mathbf{C} = \frac{\mathbf{W}_{child}^{\mathsf{T}}\mathbf{W}_{child}}{\|\mathbf{W}_{child}\|\|\mathbf{W}_{child}\|}$

    Compute the loss $\mathcal{L}_{total} = \mathcal{L}_m(f(x_i),y) + \lambda_{ip}\mathcal{L}_{ip}(\mathbf{C},\mathbf{I})$ (Eq. 7)

    Update Θ and $\mathbf{W}$ by:

        $\mathbf{W}^{i+1} = \mathbf{W}^i - \alpha \frac{\partial \mathcal{L}_{total}}{\partial \mathbf{W}^i}$ where $\frac{\partial \mathcal{L}_{total}}{\partial \mathbf{W}^i} = \frac{\partial \mathcal{L}_m}{\partial \mathbf{W}^i} + \lambda_{ip}\frac{\partial \mathcal{L}_{ip}}{\partial \mathbf{W}^i}$,

        $\Theta^{i+1} = \Theta^i - \alpha \frac{\partial \mathcal{L}_{total}}{\partial \Theta^i}$

end

Method	AgeDB-C20	FG-NET-C20	AgeDB-C30	FG-NET-C30	LAG [6]
CurricularFace [11]	82.11±0.36	81.56±0.56	81.49±0.42	78.85±0.16	90.35±0.06
MV-AM [27]	81.01±0.17	81.37±0.90	80.45±0.78	79.48±1.02	89.38±0.15
BroadFace [13]	82.93±0.20	80.4±0.25	81.85±0.48	79.03±0.54	89.63±0.26
CosFace [19]	81.65±0.66	81.27±0.57	81.02±0.79	79.03±1.42	89.58±0.06
ArcFace [4]	82.65±0.31	80.37±1.05	81.91±0.55	79.03±0.94	89.80±0.25
OECNN [22]	78.87±0.62	78.95±0.70	78.24±0.86	77.78±0.82	87.33±0.49
DAL [20]	77.28±1.41	77.18±1.19	75.90±1.01	76.53±2.50	86.13±0.51
MTL [1]	82.18±0.47	80.42±0.29	80.64±1.17	80.02±0.68	88.17±0.32
Ours	**83.55**±0.20	**82.27**±1.03	**82.97**±0.76	**81.00**±1.27	**90.49**±0.17

Table 1: Face verification accuracy (%) trained on CASIA-WebFace [23] and evaluated on test sets with child-adult pairs.

Method	AgeDB-C20	FG-NET-C20	AgeDB-C30	FG-NET-C30	LAG [6]
CurricularFace [11]	85.76±1.34	83.03±1.70	84.85±1.60	80.46±1.66	91.47±0.86
MV-AM [27]	84.70±0.91	81.72±1.27	83.43±0.51	79.12±1.09	90.81±0.68
CosFace [19]	85.82±0.03	81.46±0.62	85.07±0.50	78.05±0.86	91.37±0.23
ArcFace [4]	86.21±0.67	83.88±0.91	85.85±0.60	80.91±1.17	91.78±0.34
OECNN [22]	83.94±0.42	84.88±1.12	82.81±0.69	83.96±1.12	90.33±1.03
DAL [20]	83.58±0.10	82.11±1.17	82.59±0.36	82.62±1.35	90.31±0.46
Ours	**87.36**±0.08	**85.81**±0.24	**87.01**±0.15	**85.04**±0.56	**93.12**±0.23

Table 2: Face verification accuracy (%) trained on MS1MV2 and evaluated on test sets with child-adult pairs.

Method	AgeDB-C20	AgeDB-C30	LAG [6]	FLOPs	# of params.
OECNN [22]	31.75±4.77	24.19±3.23	42.62±1.69	6.32B	49.21M
DAL [20]	27.51±4.57	22.58±5.81	38.77±1.53	6.32B	50.26M
MTL [1]	33.86±2.42	31.73±4.93	48.76±0.75	12.51B	134.05M
Ours	**41.80**±0.92	**34.94**±0.93	**53.70**±0.42	**6.32B**	**43.79M**

Table 3: Face identification accuracy (Rank-1) (%) on test sets with along with computational costs (FLOPs, number of parameters) between age-invariant face recognition models. B and M refer to billion and million, respectively.

and cross-age face recognition test sets. Note that the difference between our test datasets and the existing cross-age test datasets is that we **intentionally include a child image in each pair** while the latter does not necessarily include them.

Face identification is the task of predicting the identity of a given facial image (*i.e.*, probe image) among the candidate images (*i.e.*, gallery images). To be more specific, we use the child image as the probe image and use the adult images as gallery images. We evaluate models with the rank-1 accuracy. We select child-adult pairs for each identity from the AgeDB-20C, AgeDB-30C dataset, and LAG dataset to build an identification set. We did not use FG-NET since the number of child-adult pairs is insufficient for the evaluation. Further details of the test datasets are included in our supplementary material.

5.2 Implementation Details

For the data processing, we leverage RetinaFace [5] to detect faces and facial landmarks on both train and test datasets. After the face detection and alignment, we perform similarity transformation according to the five facial key points (two eyes, nose, and two mouth corners) and obtain images of 112×112 pixels. We utilize ResNet50 [8] as the backbone network and obtain 512-dimensional embedded feature vectors for the evaluation. We use ArcFace[4], with a margin of 0.5 and a scale factor of 64, for $\mathcal{L}_m$ in our method. We train our

Method	LFW	CFP-FP	CPLFW	AgeDB-30	CALFW
ArcFace [4]	**99.53***	**95.56***	85.79$_{\pm0.48}$	**95.15***	92.89$_{\pm0.29}$
OECNN [22]	98.65$_{\pm0.20}$	91.81$_{\pm0.85}$	84.31$_{\pm0.56}$	91.82$_{\pm0.38}$	91.28$_{\pm0.27}$
DAL [20]	98.37$_{\pm0.29}$	90.55$_{\pm1.12}$	84.27$_{\pm0.22}$	90.24$_{\pm1.06}$	90.39$_{\pm0.48}$
MTL [1]	99.04$_{\pm0.17}$	92.67$_{\pm0.79}$	85.23$_{\pm0.55}$	92.95$_{\pm0.26}$	92.34$_{\pm0.16}$
Ours	99.24$_{\pm0.15}$	94.71$_{\pm0.37}$	**86.62**$_{\pm0.86}$	94.19$_{\pm0.31}$	**93.33**$_{\pm0.15}$

Table 4: Face verification accuracy (%) with general/cross-age datasets. * indicates that the results are from the original paper of ArcFace [4]. Note that the original paper only reports the best accuracy while we report the averaged accuracy of three independent trials.

model for 50 epochs using a batch size of 64. We use the SGD optimizer [15] with a weight decay of 0.0005 and the learning rate is initially set to 0.1, divided by 10 at the epoch of 28, 36, 46. For each experiment, we report the averaged accuracy and its standard deviation over three independent trials. For the Inter-Prototype loss, we obtain (n_{child}, n_{total})=(229, 10,575) and $(1,148, 85,742)$ in the CASIA-WebFace dataset [23] and MS1MV2, respectively.

5.3 Quantitative Evaluation

Face verification/identification with child-adult pairs Table. 1 and Table. 2 show that our approach outperforms all baseline approaches on every test set in the face verification with child-adult pairs when trained with the small and large dataset, respectively. Our proposed approach shows its strength as the age gap between verification pairs increases from at least 20 to 30. Additionally, recent age-invariant face recognition models, OECNN [22], DAL [20], and MTL [1], generally show low verification accuracy across all test sets. We believe this result demonstrates that previous studies heavily rely on the additionally collected child images in the training dataset. Table. 3 also shows that our approach outperforms the existing work of age-invariant face recognition on the face identification using three test sets with child-adult pairs. Along with the outstanding face recognition performances, we also report the FLOPs (*i.e.*, the number of floating-point operations) and the required number of parameters for each method. Note that MTL [1] requires a considerable amount of computational cost since it utilizes an image generation model. This shows that using image generation models requires lengthy training and a substantial amount of parameters. This result again demonstrates that our method is *simple yet effective*.

Face verification with general/cross-age test sets Table. 4 shows the face recognition performance on both general and cross-age test sets. Since both our approach and the previous age-invariant face recognition models focus on the pairs with age gaps, there may exist performance degradation on the general test sets. To be specific, compared to Arcface [4], the previous age-invariant face recognition models show large performance degradation in general and cross-age test sets. On the other hand, our approach shows reasonable performance in general and cross-age test sets even with superior performances in the child-adult pairs. This result demonstrates that the Inter-Prototype loss does not bring negative effect on the general face recognition performance compared to existing age-invariant face recognition models.

5.4 Qualitative Analysis

Reduced inter-class similarity Fig. 4 (a) and Fig. 4 (b) visualize the inter-class similarity using the child-child pairs and child-adult pairs on ArcFace [4] and our approach, respectively. We use the identities that include child images for the visualization. The inter-class

Method	AgeDB-C20	FG-NET-C20	AgeDB-C30	FG-NET-C30	LAG [3]
ArcFace [4]	$82.65_{\pm 0.31}$	$80.37_{\pm 1.05}$	$81.91_{\pm 0.55}$	$79.03_{\pm 0.94}$	$89.80_{\pm 0.25}$
Ours + Full	$83.29_{\pm 0.39}$	$82.11_{\pm 0.44}$	$82.41_{\pm 0.40}$	$80.56_{\pm 0.67}$	$90.44_{\pm 0.08}$
Ours + Child	$\mathbf{83.55}_{\pm 0.20}$	$\mathbf{82.27}_{\pm 1.03}$	$\mathbf{82.97}_{\pm 0.76}$	$\mathbf{81.00}_{\pm 1.27}$	$\mathbf{90.49}_{\pm 0.17}$

Table 5: Face verification accuracy (%) on the different number of prototype vectors. Ours + Full indicates applying the Inter-Prototype loss on all identities. Similarly, Ours + Child refers to applying the Inter-Prototype loss only on identities including the child images.

Method	AgeDB-C20	AgeDB-C30	LAG [3]
CosFace [19]	$81.65_{\pm 0.66}$	$81.02_{\pm 0.79}$	$89.58_{\pm 0.06}$
Ours + CosFace	$\mathbf{82.41}_{\pm 0.57}$	$\mathbf{81.52}_{\pm 0.57}$	$\mathbf{90.14}_{\pm 0.56}$
ArcFace [4]	$82.65_{\pm 0.31}$	$81.91_{\pm 0.55}$	$89.80_{\pm 0.25}$
Ours + ArcFace	$\mathbf{83.55}_{\pm 0.20}$	$\mathbf{82.97}_{\pm 0.76}$	$\mathbf{90.49}_{\pm 0.17}$

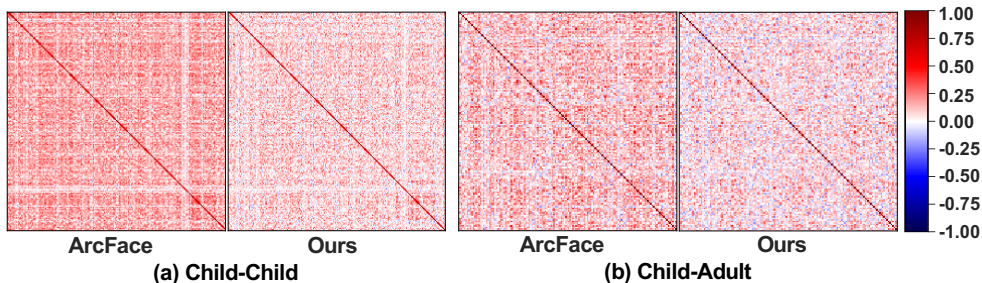Table 6: Face verification accuracy (%) of adding Inter-Prototype loss on both CosFace and Arcface.

Figure 4: Comparison of inter-class similarity between ArcFace and our approach using (a) child-child pairs and (b) child-adult pairs. Each row and column indicate the identities used for the visualization, respectively.

similarity between child-child pairs is generally high in ArcFace, overall colored red. After adding the Inter-Prototype loss, the inter-class similarity decreases substantially, showing reduced portions of red, which demonstrates the effectiveness of our proposed approach. The Inter-Prototype loss also decreases the inter-class similarity between child-adult pairs. As aforementioned in Section 3, the Inter-Prototype loss also enforces the child-adult pairs of different identities to be distinct. Due to this fact, we observe that the off-diagonals of our approach in Fig. 4 (b) are colored blue or white compared to the ArcFace.

2D projection of representation Fig. 5 visualizes the 2D projection of prototype vectors after training with ArcFace and our approach. The red and blue dots indicate the embedded prototype vectors that include child images and the ones that do not, respectively. The 2D projection of prototype vectors trained with existing face recognition models show that the identities including child images are embedded closely, indicating that the model does not completely distinguish the child images. We observe that adding the Inter-Prototype loss mitigates this issue by inspecting that the red dots are evenly distributed in the embedding space.

6 Discussion

Applicability on margin-based loss functions Table. 6 demonstrates that the proposed Inter-Prototype loss improves the face verification accuracy of both CosFace and ArcFace
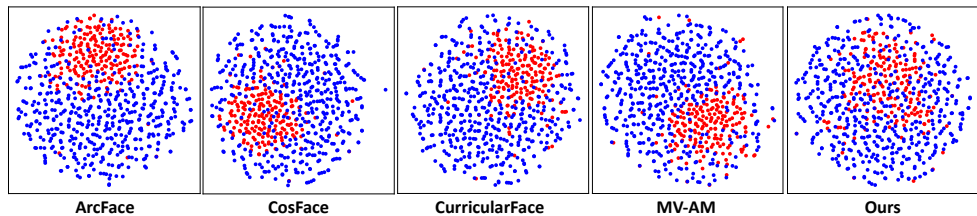
Figure 5: 2D projection of prototype vectors. Red and blue indicate the prototype vectors that include the child images and the ones without them, respectively.

on AgeDB-C20, AgeDB-C30, and LAG. As described in Eq. 7, we add the Inter-Prototype loss to the margin-based loss functions (*e.g.* CosFace and ArcFace), respectively, and show that adding it improves the face verification accuracy in each model. Such a result demonstrates that the proposed Inter-Prototype loss can be widely applied to existing margin-based loss functions as long as the last fully-connected layer can be interpreted as a set of prototype vectors.

Extending the Inter-Prototype loss on all identities Table 5 compares the face verification performance of applying the Inter-Prototype loss on all identities and only the identities including child images. Applying the Inter-Prototype loss on all identities improves the performance of ArcFace. This indicates that even without using the age labels, minimizing the inter-class similarity between identities improves the representation of child images. However, we observe that using only the identities including child images shows further improvement. We conjecture that applying the Inter-Prototype loss on all identities gives less emphasis on the loss backpropagated on the prototype vectors including child images. Therefore, this experiment shows that the Inter-Prototype loss effectively minimizes the inter-class similarity when given a specific set of identities.

7 Conclusion

In this work, we first reveal that reducing the inter-class similarity among child images is crucial for verifying child-adult pairs. Based on our observation, we propose a novel loss function called the Inter-Prototype loss to minimize the inter-class similarity between child images using the prototype vectors. Our loss function does not require additional child images or learnable parameters which were necessary for previous age-invariant face recognition models. Through extensive experiments, we demonstrate that our method clearly outperforms the existing face recognition models on test sets with child-adult pairs. Additionally, the various visualization and ablation studies verify the effectiveness of our approach. We hope our simple yet effective method inspires the future researchers who plan to delve into this area.

Acknowledgement

References

[1] Vítor Albiero and Kevin W. Bowyer. Is face recognition sexist? no, gendered hairstyles and biology are. In *31st British Machine Vision Conference 2020, BMVC 2020, Virtual Event, UK, September 7-10, 2020*. BMVA Press, 2020.

[2] Microsoft Azure. Microsoft azure cognitive services facial recognition., 2021. URL https://azure.microsoft.com/en-us/services/cognitive-services/face/.

[3] Simone Bianco. Large age-gap face verification by feature injection in deep networks. *Pattern Recognit. Lett.*, 90:36–42, 2017.

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019.

[5] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 5202–5211. IEEE, 2020.

[6] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Yuan Yao, and Shaogang Gong. Interestingness prediction by robust learning to rank. In *ECCV*, 2014.

[7] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition, 2016.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognitionn, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. doi: 10.1109/CVPR.2016.90.

[9] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[10] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. IEEE, 2020.

[11] Zhizhong Huang, Junping Zhang, and Hongming Shan. When age-invariant face recognition meets face age synthesis: A multi-task learning framework. In *CVPR*, 2021.

[12] Yonghyun Kim, Wonpyo Park, and Jongju Shin. Broadface: Looking at tens of thousands of people at once for face recognition. In *European Conference on Computer Vision*, pages 536–552, 2020.

[13] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6738–6746. IEEE Computer Society, 2017.

[14] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, page 5, 2017.

[15] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[16] S. Sengupta, J.C. Cheng, C.D. Castillo, V.M. Patel, R. Chellappa, and D.W. Jacobs. Frontal to profile face verification in the wild. In *IEEE Conference on Applications of Computer Vision*, 2016.

[17] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

[18] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014. doi: 10.1109/ CVPR.2014.220.

[19] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 5265–5274. IEEE Computer Society, 2018.

[20] Hao Wang, Dihong Gong, Zhifeng Li, and Wei Liu. Decorrelated adversarial learning for age-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 2019.

[21] Xiaobo Wang, Shifeng Zhang, Shuo Wang, Tianyu Fu, Hailin Shi, and Tao Mei. Misclassified vector guided softmax loss for face recognition. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020.

[22] Yitong Wang, Dihong Gong, Zheng Zhou, Xing Ji, Hao Wang, Zhifeng Li, Wei Liu, and Tong Zhang. Orthogonal deep features decomposition for age-invariant face recognition. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XV*, 2018.

[23] Dong Yi, Zhen Lei, S. Liao, and S. Li. Learning face representation from scratch. *ArXiv*, abs/1411.7923, 2014.

[24] Jian Zhao, Yu Cheng, Yi Cheng, Yang Yang, Fang Zhao, Jianshu Li, Hengzhu Liu, Shuicheng Yan, and Jiashi Feng. 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*, 2019.

[25] T. Zheng, W. Deng, and J. Hu. Age estimation guided convolutional neural network for age-invariant face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.