

Supplementary for 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 Park¹
psh01087@kaist.ac.kr
Yonggyu Kim²
dydrb2535@kaist.ac.kr
Jaegul Choo¹
jchoo@kaist.ac.kr

¹ KAIST AI,
South Korea
² Korea University,
South Korea

This supplementary presents 1) how we build the test sets with child-adult pairs, 2) further implementation details, and 3) comparison with re-weighting and oversampling methods. Our codes and test sets are available at <https://github.com/leebebeto/Inter-Prototype>.

1 Test sets with child-adult pairs

Since our work focuses on face recognition with child-adult pairs, we build new test sets using the existing cross-age test sets including AgeDB [1] and FG-NET [2]. As mentioned in the main paper, these cross-age test sets do not necessarily include child images in each pair since they did not focus on image pairs with *child-adult pairs*. Due to this fact, we intentionally include the child images and make pairs with a certain age gap. For the face verification, we pair the child and the adult images of the same identity (*i.e.*, positive pairs) and different identities (*i.e.*, negative pairs) with a certain age gap. Then, we randomly sample the equal number of positive and negative pairs.

For the face identification, we select identities which have child and adult images with a certain age gap similar to the face verification. Since face identification is the task of matching a given image (*i.e.*, probe image) to one of the candidate images (*i.e.*, gallery set), we use child images as the probe image and adult images as the gallery set. The adult images in the gallery set need to have a certain age gap with child images of its identity and other identities. After such pre-processing, FG-NET-C20 and FG-NET-C30 only include 38 and 22 identities for the gallery set, respectively. Due to the small number of identities in the

* indicates equal contribution.

Test dataset	# of pairs	# of gallery
AgeDB-C20	3786	63
FG-NET-C20	1034	-
AgeDB-C30	2778	62
FG-NET-C30	372	-
LAG [10]	3400	1005

Table 1: Test datasets with child-adult pairs. The second and third column indicates the number of pairs and the number of gallery images used for the face verification and face identification, respectively. We did not use FG-NET for face identification since the number of gallery images was insufficient.

gallery set, we *did not* use these two test sets for the face identification. Table 1 shows the test datasets used in our work.

2 Further Implementation Details

Our backbone model for extracting features vectors from face images is based on the ResNet model [10]. We adopt a variant of the ResNet used in the ArcFace paper [10] with a different Residual Block, called the Improved Residual Block, which has a different ordering of layers in the bottleneck layers. As shown in Fig. 1, the Improved Residual Block uses an additional convolution filter followed by Batch Normalization in the residual connection. The Improved Residual Block also adds a Squeeze-and-Excitation module [10] before residual connections. Note that we use the same backbone network for all experiments including our baseline models.

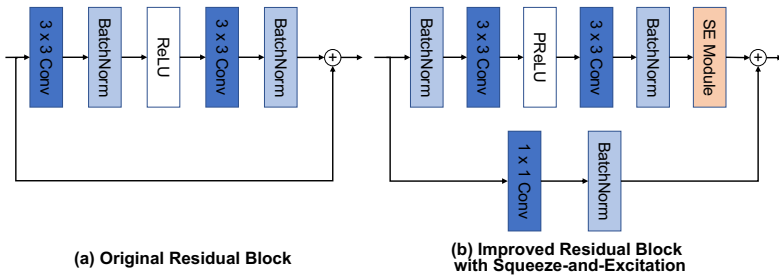


Figure 1: Architecture of the original Residual Block and the Improved Residual Block.

3 Extended Results on Comparison with Re-weighting/Oversampling

This section provides the comparisons between our approach and straight-forward alternatives for reducing the inter-class similarity between child images. Those approaches are 1) re-weighting the margin-based loss function for child images with a scale of w , 2) assigning higher margins than the default value ($m=0.5$), and 3) oversampling child images with the ratio of ρ . ρ indicates the ratio between the number of child images and the number of adult images in a mini-batch. Table 2 compares our approach with such approaches on the test sets

Method	AgeDB-C20	FG-NET-C20	AgeDB-C30	FG-NET-C30	LAG [■]
ArcFace [■]	82.65±0.31	80.37±1.05	81.91±0.55	79.03±0.94	89.80±0.25
Re-weighting ($w=2, 5, 10, 50$)	0.5	0.5	0.5	0.5	0.5
Margin ($m=0.55$)	82.73±0.64	79.75±0.93	81.94±0.68	77.96±0.47	89.94±0.19
Margin ($m=0.60$)	82.60±0.13	79.85±1.02	81.99±0.80	78.05±1.21	89.56±0.40
Margin ($m=0.65$)	82.83±0.09	80.88±0.39	81.88±0.63	77.78±1.55	89.90±0.11
Margin ($m=0.70$)	83.12±0.10	80.24±0.85	81.66±0.33	77.77±1.24	89.79±0.22
Margin ($m=0.75$)	82.87±0.54	81.04±0.93	82.07±0.33	79.30±0.71	89.38±0.37
Margin ($m=0.80$)	83.19±0.47	80.27±0.67	82.69±0.42	76.88±0.97	90.11±0.38
Margin ($m=0.85$)	83.01±0.25	80.75±0.70	82.02±0.70	78.94±0.94	90.13±0.54
Margin ($m=0.90$)	82.88±0.37	80.21±0.39	82.22±0.38	78.32±1.38	90.15±0.33
Oversampling ($\rho=0.25$)	80.84±0.95	80.92±1.29	80.06±1.40	79.48±1.71	88.37±0.66
Oversampling ($\rho=0.5$)	80.83±0.36	80.14±0.78	80.13±0.38	79.03±1.50	88.89±0.32
Oversampling ($\rho=0.75$)	80.02±1.04	79.14±0.15	79.72±0.39	77.24±0.95	88.55±0.21
Oversampling ($\rho=1.00$)	80.06±0.97	79.08±1.39	79.22±1.01	78.23±2.39	88.23±0.54
Ours	83.55±0.20	82.27±1.03	82.97±0.76	81.00±1.27	90.49±0.17

Table 2: Comparison of our approach with re-weighting, adding high margin, and oversampling on face verification accuracy (%).

with child-adult pairs. We conducted extensive experiments to find the best hyper-parameter for each approach. Even with an extensive hyper-parameter search, none of the alternative methods outperform our proposed method.

References

- [1] Simone Bianco. Large age-gap face verification by feature injection in deep networks. *Pattern Recognit. Lett.*, 90:36–42, 2017.
- [2] 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.
- [3] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Yuan Yao, and Shaogang Gong. Interestingness prediction by robust learning to rank. In *ECCV*, 2014.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. doi: 10.1109/CVPR.2016.90.
- [5] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [6] 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.