Supplementary Material: Attentional Feature-Pair Relation Networks for Accurate Face Recognition

Bong-Nam Kang[1,3], Yonghyun Kim[2,3], Bongjin Jun[1], Daijin Kim[3]
[1]StradVision, Inc. [2]Kakao Corp. [3]POSTECH
{bongnam.kang, bongjin.jun}@stradvision.com, aiden.kyh@kakaocorp.com, dkim@postech.ac.kr

A. More Experiments

A.1. Evaluation on the CALFW, CPLFW, CFP, and AgeDB datasets

We conduct experiments to demonstrate the effects of the proposed method on the Cross-Age LFW (CALFW) [16], Cross-Pose LFW (CPLFW) [15], Celebrities in Frontal-Profile in the Wild (CFP) [10], and AgeDB [8] datasets.

CALFW. The CALFW is constructed by reorganizing the LFW [4, 5] verification pairs with apparent age gaps as large as possible to form the positive pairs and then selecting negative pairs using individuals with the same race and gender. The CALFW is more challenging than LFW. Similar to LFW, CALFW evaluation consists of verifying 6,000 pairs of images in 10 folds and report the average accuracy.

CPLFW. The CPLFW is also constructed by reorganizing the LFW verification pairs by searching and selecting of 3,000 positive pairs with pose difference to add pose variation to intra-class variance. Negative pairs are also reorganized to reduce the influence of attribute differences between positive and negative pairs. Therefore, the CPLFW is more focused on cross-pose face recognition, and is more challenging than the LFW.

CFP. The CFP consists of 500 subjects each with 10 frontal and 4 profile images. The evaluation protocol includes frontal-frontal (FF) and frontal-profile (FP) face verification, and each protocol has 10 folders with 350 positive pairs with same identity and 350 negative pairs with different identities.

AgeDB. The AgeDB is a dataset for age invariant face recognition in the wild with in pose, expression, illumination, and age. The AgeDB contains 12,240 images of 440 unique subjects. The minimum and maximum ages are 3 and 101 years old, respectively. The test set is divided into four groups with different year gaps such as 5, 10, 20, and 30 years. Each group has ten split of face images, and each split includes 300 psotive examples and 300 negative examples.

Table 1. Performances of the proposed face recognition method on the CALFW, CPLFW, CFP, and AgeDB datasets.

Method	CALFW	CPLFW	CFP	AgeDB
CenterFace [12]	85.48	77.48	-	-
SphereFace [6]	90.30	81.40	94.38	91.70
VGGFace2 [1]	90.57	84.00	-	-
ArcFace [2]	95.45	92.08	95.56	95.15
model B (AFRN w/o pair selection)	94.57	91.17	93.30	93.40
model C (AFRN w/ pair selection)	**96.30**	**93.48**	**95.56**	**95.35**

Evaluation Results. In image-based recognition on the CALFW, CPLFW, CFP, and AgeDB, we use a squared L_2 distance threshold to determine the classification of same and different. Table 1 shows that our proposed AFRN with pair selection (**model C**) itself provides better accuracy than the AFRN without pair selection (**model B**). Finally, the **model C** acheives the outperformed accuracy and the *state-of-the-art* results on the CALFW, CPLFW, CFP, AgeDB, respectively.

A.2. Evaluation on the IJB-C dataset

We also conduct experiments to demonstrate the effects of the proposed AFRN on the IJB-C [7] datasets. The IJB-C is an extenstion of the IJB-B, which contains a total of 31,334 still images with 3,531 unique subjects, and 117,542 video frames in unconstrained environments. It has an average of up to 6 imagew per subject, an average of up to 33 frames per subject and 3 videos per subject. Since the IJB-C contains two set of galleries 1 and 2, we report the average performance of both the gallery sets.

Three models (**model A**, **model B**, and **model C**) are trained on the roughly 2.8M refined VGGFace2 training set, with no people overlapping with subjects in the IJB-C dataset. For 1:1 face verification, we report the test results by using true accept rate (TAR) *vs.* false accept rate (FAR) (Table 2). For 1:N face identification, we report the

Table 2. Comparison of performances of the proposed AFRN method with the *state-of-the-art* on the IJB-C dataset. For verification, TAR *vs.* FAR are reported. For identification, TPIR *vs.* FPIR and the Rank-N accuracies are presented.

Method	1:1 Verification TAR				1:N Identification TPIR				
	FAR=0.00001	FAR=0.0001	FAR=0.001	FAR=0.01	FPIR=0.01	FPIR=0.1	Rank-1	Rank-5	Rank-10
VGGFace2 [1]	0.747	0.840	0.910	0.960	0.746 ± 0.018	0.842 ± 0.022	0.912 ± 0.017	0.949 ± 0.010	0.962 ± 0.007
VGGFace2_ft [1]	0.768	0.862	0.927	0.967	0.763 ± 0.018	0.865 ± 0.018	0.914 ± 0.020	0.951 ± 0.013	0.961 ± 0.010
CenterFace [12]	0.781	0.853	0.912	0.953	0.772 ± 0.026	0.853 ± 0.015	0.907 ± 0.013	0.941 ± 0.007	0.952 ± 0.004
Comparator Net [14]	-	0.885	0.947	0.983	-	-	-	-	-
ArcFace [2]	0.883	0.924	0.956	0.977	-	-	-	-	-
Rajeev *et. al* [9]	0.869	0.925	0.959	0.979	0.873 ± 0.032	0.925 ± 0.017	0.949 ± 0.018	0.969 ± 0.010	0.975 ± 0.009
model A (baseline)	$\mathbf{0.794}$	$\mathbf{0.865}$	$\mathbf{0.921}$	$\mathbf{0.958}$	$\mathbf{0.785 \pm 0.022}$	$\mathbf{0.870 \pm 0.021}$	$\mathbf{0.918 \pm 0.017}$	$\mathbf{0.949 \pm 0.013}$	$\mathbf{0.958 \pm 0.010}$
model B (AFRN w/o pair selection)	$\mathbf{0.851}$	$\mathbf{0.903}$	$\mathbf{0.951}$	$\mathbf{0.977}$	$\mathbf{0.853 \pm 0.018}$	$\mathbf{0.905 \pm 0.018}$	$\mathbf{0.931 \pm 0.022}$	$\mathbf{0.956 \pm 0.010}$	$\mathbf{0.964 \pm 0.009}$
model C (AFRN w/ pair selection)	$\mathbf{0.883}$	$\mathbf{0.930}$	$\mathbf{0.963}$	$\mathbf{0.987}$	$\mathbf{0.884 \pm 0.017}$	$\mathbf{0.931 \pm 0.013}$	$\mathbf{0.957 \pm 0.015}$	$\mathbf{0.976 \pm 0.017}$	$\mathbf{0.977 \pm 0.007}$

results by using the true positive identification rate (TPIR) *vs.* false positive identification rate (FPIR) and Rank-N (Table 2). We average all the $1,024$ dimensional output vectors of the last fully connected layer of $\mathcal{F}_{\boldsymbol{\theta}}$ for a media in the template, then we average these media-averaged features to get the final template feature as face representation. Similarity to evaluation on the IJB-A and IJB-B, all performance evaluations are based on the squared L_2 distance threshold.

Table 2 shows that the proposed **model C** shows a consistently higher accuracy than **model B** by the improvement of 1.0-3.2% TAR at FAR = 0.00001-0.01 in the verification task, 2.6-3.1% TPIR at FPIR = 0.01-0.1 in the identification open set task, and 2.6% for rank-1 in the identification close set task. Although **model C** is trained from scratch, it outperformed the state-of-the-art method. This validates the effectiveness of the proposed AFRN with the pair selection on the large-scale and challenging unconstrained face recognition.

From the experimental results (Table 2), we have the following observations. First, compared to **model A**, **model B** achieves a consistently superior accuracies (TAR and TPIR) by 1.9-5.7% for TAR at FAR = 0.00001-0.01 in the verification task, 3.5-6.8% for TPIR at FPIR = 0.01 and 0.1 in the identification open set task, and 1.3% for Rank-1 in the identification close set task. Second, **model C** shows a consistently higher accuracy than **model A** by the improvement of 2.9-8.9% TAR at FAR = 0.00001-0.01 in the verification task, 6.1-9.9% TPIR at FPIR = 0.01-0.1 in the identification open set task, and 3.9% Rank-1 in the identification close set task. Third, **model C** shows a consistently higher accuracy than **model B** by the improvement of 1.0-3.2% TAR at FAR = 0.00001-0.01 in the verification set task, 2.6-3.1% TPIR at FPIR = 0.01-0.1 in the identification open set task, and 2.6% for Rank-1 in the identification close set task. Last, although **model C** is trained from scratch, it outperformed the state-of-the-art method (Rajeev *et. al* [9]) by 0.4-1.4% at FAR = 0.00001-0.01 in verification task, 0.6-1.1% TPIR at FPIR = 0.01-0.1 in the identification open set task, and 0.8% Rank-1 of identification close set task on the IJB-C dataset.

The method proposed by Rajeev *et. al* [9] is a fusion of ResNet-101 [3] and Inception ResNet-v2 [11] models. The Inception ResNet-v2 network has 224 *conv.* layers, which are considerably more complex than our proposed AFRN method, and they used the training set with 5.6M images of 58,000 identities whereas we have a smaller number of subjects with 2.8M images of 8,900 identities. In order to obtain the comparable or better performance, it is considered that the proposed attention module and pair selection is effective because it obtains high performance even if a lesser amount of training images is used. This validates the effectiveness of the proposed AFRN with the pair selection on the large-scale and challenging unconstrained face recognition.

References

[1] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. *CoRR*, abs/1710.08092, 2017.

[2] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *ArXiv e-prints*, Jan 2018.

[3] 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)*, pages 770–778, June 2016.

[4] 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.

[5] Gary B. Huang Erik Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.

[6] 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)*, pages 6738–6746, 2017.

[7] Brianna Maze, Jocelyn Adams, James A. Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K. Jain, W. Tyler Niggel, Janet Anderson, Jordan Cheney, and Patrick Grother. Iarpa janus benchmark - c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165, Feb 2018.

[8] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: The first manually collected, in-the-wild age database. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1997–2005, July 2017.

[9] Rajeev Ranjan, Ankan Bansal, Jingxiao Zheng, Hongyu Xu, Joshua Gleason, Boyu Lu, Anirudh Nanduri, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa. A fast and accurate system for face detection, identification, and verification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(2):82–96, April 2019.

[10] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, March 2016.

[11] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 4278–4284. AAAI Press, 2017.

[12] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *Computer Vision – ECCV 2016*, pages 499–515. Springer International Publishing, 2016.

[13] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K. Jain, James A. Duncan, Kristen Allen, Jordan Cheney, and Patrick Grother. Iarpa janus benchmark-b face dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 592–600, 2017.

[14] Weidi Xie, Li Shen, and Andrew Zisserman. Comparator networks. In *European Conference on Computer Vision (ECCV 2018)*, September 2018.

[15] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. Technical Report 18-01, Beijing University of Posts and Telecommunications, February 2018.

[16] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *CoRR*, abs/1708.08197, 2017.