Unified Negative Pair Generation toward Well-discriminative Feature Space for Face Recognition (Supplementary Material)

Junuk Jung
rnans33@koreatech.ac.kr
Seonhoon Lee
seonhoon1002@koreatech.ac.kr
Heung-Seon Oh
ohhs@koreatech.ac.kr
Yongjun Park
qkr2938@koreatech.ac.kr
Joochan Park
green669@koreatech.ac.kr
Sungbin Son
sbson0621@koreatech.ac.kr

School of Computer Science and Engineering
Korea University of Technology and Education (KOREATECH)

In a methodology section, we explain the geometric interpretation of a feature space to help understand our UNPG. Subsequently, we demonstrate the backward propagation of unified loss with UNPG. The experiments section delivers the implementation details and the evaluation results with further analysis.

A Methodology

Geometrical Interpretation of Feature Space. We interpret the role of UNPG by associating a feature space, as shown in Fig. 1. To form WDFS satisfying $\inf \mathcal{S}^p > \sup \mathcal{S}^n$, a loss function should assign a large loss in the feature space with low discriminability, whereas it should assign a small loss, and vice versa. Many loss functions fail to form WDFS because of the mismatch between similarity sets of the sampled pairs and all pairs. Fig. 1 (a) depicts the ideal behavior of a loss function that assigns a large loss in the feature space with low discriminability for $\inf \mathcal{S}^p \gg \sup \mathcal{S}^n$. In contrast, in Fig. 1 (b), a small loss is assigned in the feature space with low discriminability because sampled $\hat{\mathcal{S}}^p$ and $\hat{\mathcal{S}}^n$ are well-separated. This problem is alleviated using a similarity score with a margin, as shown in Fig. 1 (c). This makes $\hat{\mathcal{S}}^p$ informative and worthy of training, as the interval of $\hat{\mathcal{S}}^p$ shifts to the left. However, it is difficult to generate similarities of extremely hard negative pairs such as $\cos\theta^n_{min}$ of Fig. 1 (a) owing to the convergence of the class weight vectors to their center. So, small loss is assigned because they still have the gap between $\cos\theta^n_{min}$ of Fig. 1 (a) and $\cos\hat{\theta}^n_{min}$ of Fig. 1 (c). Fig. 1 (d) shows the effect of UNPG. The interval of $\hat{\mathcal{S}}^n$ becomes wider as more negative pairs are included in $\hat{\mathcal{S}}^n$. They also have the chance to sample the extremely hard negative

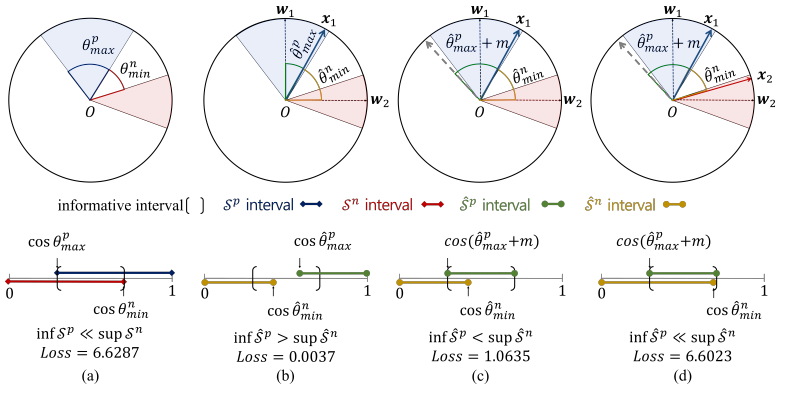

Figure 1: Geometrical interpretation of feature space associated with similarity space. (a) As ideal behavior of the loss function, it imposes a large loss in feature space with low discriminability. A shading area in the same color represents the target region of the same class. θ^p_{max} and θ^n_{min} are the respective angles of max positive and min negative pairs in the feature space. $\mathcal{S}^p$ and $\mathcal{S}^n$ represent similarity sets. (b) In spite of being equally low discriminative, a very small loss is given by vanilla loss (e.g., norm-softmax). $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are the normalized weight vectors of classes 1 and 2, while $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the normalized feature vector. $\hat{\theta}^p_{max}$ and $\hat{\theta}^n_{min}$ represent the angle of max positive and min negative pairs in $\hat{\mathcal{S}}^p \subset \mathcal{S}^p$ and $\hat{\mathcal{S}}^n \subset \mathcal{S}^n$, respectively. (c) Mismatch between $\mathcal{S}^p$ and $\hat{\mathcal{S}}^p$ is reduced by using a marinal classification loss (e.g., ArcFace). However, still a small loss is given because of a mismatch between $\mathcal{S}^n$ and $\hat{\mathcal{S}}^n$. (d) Marginal classification loss with UNPG behaves closest to ideal by alleviating mismatch between $\mathcal{S}^n$ and $\hat{\mathcal{S}}^n$.

pairs. Consequently, a large loss is assigned in the feature space with low discriminability, similar to Fig. 1 (a).

Backward Propagation. The gradients of unified loss equipped with UNPG about $\boldsymbol{x}_i$ and $\boldsymbol{w}_c$ are derived as follows:

$$\begin{aligned}\frac{\partial \mathcal{L}_i^{uni}}{\partial \boldsymbol{x}_i} &= \frac{\partial \mathcal{L}_i^{uni}}{\partial \bar{\boldsymbol{x}}_i}\frac{\partial \bar{\boldsymbol{x}}_i}{\partial \boldsymbol{x}_i} = \gamma(P_i - y_i)(\mathbb{1}\cdot \bar{\boldsymbol{w}}_c + \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\tilde{\mathcal{N}}^{ml}} \bar{\boldsymbol{x}}_j)\frac{\partial \bar{\boldsymbol{x}}_i}{\partial \boldsymbol{x}_i} \\ \frac{\partial \mathcal{L}_i^{uni}}{\partial \boldsymbol{w}_c} &= \frac{\partial \mathcal{L}_i^{uni}}{\partial \bar{\boldsymbol{w}}_c}\frac{\partial \bar{\boldsymbol{w}}_c}{\partial \boldsymbol{w}_c} = \gamma(P_i - y_i)^\top \bar{\boldsymbol{x}}_i \frac{\partial \bar{\boldsymbol{w}}_c}{\partial \boldsymbol{w}_c}\end{aligned} \tag{1}$$

where $\bar{\boldsymbol{x}}_i = \boldsymbol{x}_i/\|\boldsymbol{x}_i\|$ and $\bar{\boldsymbol{w}}_c = \boldsymbol{w}_c/\|\boldsymbol{w}_c\|$. $\mathbb{1}$ denotes $(\boldsymbol{x}_i, \boldsymbol{w}_c) \in \mathcal{N}_i^{cl}$ appears. Note that the above back-propagation does not take into account injecting a margin.

B Experiments

B.1 Implementation Details

Datasets. For training, MS1M-V2[4] and K-FACE:T4[11] datasets were employed. For testing, several benchmark datasets (IJB-B[36], IJBC[16], MegaFace[12], LFW[10], CFP-FP[23], AgeDB-30[18], CALFW[40], and K-FACE:Q1-Q4[11]) were used to evaluate FR

models. Table 1 summarizes the datasets used in our experiments.

Train	# Identities	# Images
MS1M-V2[2]	85K	5.8M
K-FACE:T4[1]	370	3.8M
Test	**# Identities**	**# Images**
IJB-B[16]	1,845	76.8K
IJB-C[9]	3,531	148.8K
MegaFace (P)[6]	530	100K
MegaFace (G)[6]	690K	1M
LFW[3]	5,749	13,233
CFPFP[12]	500	7,000
AgeDB-30[10]	568	16,488
CALFW[18]	5,749	12,174
Test[1]	**# Pairs**	**# Variance**
K-FACE:Q1	1,000	Very Low
K-FACE:Q2	10K	Low
K-FACE:Q3	10K	Middle
K-FACE:Q4	10K	High

Table 1: A brief overview of FR datasets. (P) and (G) refer to the probe set the gallery set on MegaFace, respectively.

Training. For preprocessing, face images were resized to 112×112 and normalized using the mean (0.485, 0.456, 0.406) and standard deviations (0.229, 0.224, 0.225). For data augmentation, a horizontal flip was applied with a 50% of chance. All experiments were performed using two NVIDIA-RTX A6000 GPUs with a mini-batch size of 512. ResNet-34 (R34) and ResNet-100 (R100) were used as backbone models. We re-implemented the state-of-the-art models: CosFace[14], ArcFace[2], and MagFace[9].

The hyper-parameters used in our experiments were as follows: In ArcFace and CosFace, scale factor $\gamma = 64$ and margin $m = 0.5$ were set. In MagFace, $\gamma = 64, l_a = 10, u_a = 110, l_m = 0.4, l_m = 0.8, \lambda_g = 35$ were used. For K-FACE, SN-pair[5] and circle-loss[13] employed $\gamma = 64$ and $\gamma = 32, m = 0.25$, respectively. In MixFace[5], $\varepsilon = 1e-22$ and $m = 0.5$ were set. In MS-loss[15], $\alpha = 2, \gamma = 0.5, \beta = 50$ were used. Triplet loss employed $m = 0.5$. In contrastive loss, positive and negative margins were set to 0 and 1, respectively. Finally, in UNPG, the wisker size $r = 1.0$ was used with ResNet-34, whereas $r = 1.5$ or $r = 2.0$ were used ResNet-100. The stochastic gradient descent (SGD) optimizer was utilized in conjunction with a cosine annealing scheduler[7] to control the learning rate, which started from 0.1. The momentum, weight decay, and warm-up epochs were set to 0.9, 0.0005, and 3, respectively. The maximum number of training epochs was set to 20 for all models, except that it was set to 25 with MagFace for a fair comparison. The size of the deep feature space extracted from the backbone model was set to 512.

Test. Cosine similarity was used as a similarity score. Different evaluation metrics were applied depending on the FR tasks. In the verification task (1:1), verification accuracy using the best threshold was exploited for a dataset that has a small number of test images with the same ratio between positive and negative pairs, such as LFW, CFP-FP, AgeDB-30, CALFW, and CPLFW. Otherwise, TAR@FAR was used on IJB-B, IJB-C, and K-FACE. In the identification task (1:N) on MegaFace, rank-1accuracy was utilized.

Method	Loss Type	LFW	CFP-FP	Age-DB	IJBB	IJBC
Contrastive	Metric Loss	98.79	81.04	92.01	72.18	76.16
Triplet	Metric Loss	98.35	90.79	88.36	32.65	36.86
ArcFace	Classification Loss	99.81	97.10	98.06	93.38	95.08
Arc+Contrastive	Multi-Objective Loss	99.80	96.78	97.73	93.03	94.86
Arc+Triplet	Multi-Objective Loss	99.81	96.79	97.89	93.22	94.79
MixFace (Arc+SN-pair)	Multi-Objective Loss	99.53	96.32	95.56	93.22	94.79
Arc+UNPG	Unified Loss	**99.83**	**97.15**	**98.08**	**93.66**	**95.33**

Table 2: Verification accuracy on LFW, CFP-FP, AgeDB-30, IJBB (1e-4), and IJBC (1e-4) with ResNet-34 backbone.

Method	LFW	CFP-FP	Age-DB	IJB-B		IJB-C	
				1e-5	1e-4	1e-5	1e-4
ArcFace	99.81	97.10	98.06	86.28	93.38	92.21	95.08
Arc+UNPG, semi-hard	99.80	**97.35**	98	87.25	93.58	92.59	95.20
Arc+UNPG, $r = 1.0$	**99.83**	97.15	**98.08**	**88.05**	**93.66**	**93.02**	**95.33**

Table 3: Verification accuracy on LFW, CFP-FP, Age-DB, IJBB, and IJBC with ResNet-34 backbone.

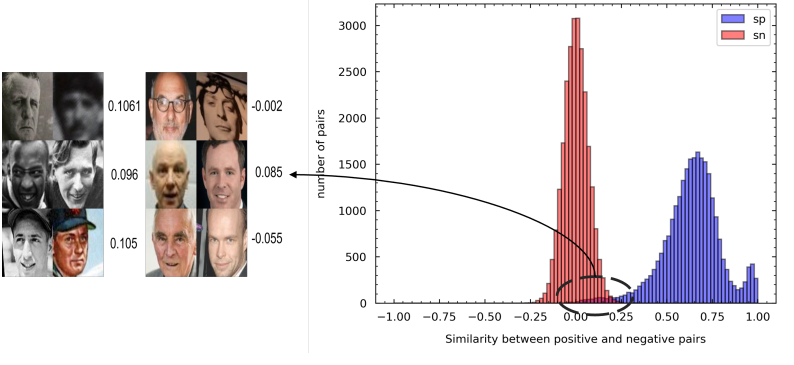

Figure 2: Mislabeled positive pair (left) and similarity distribution for 25,784 randomly generated positive and negative pairs (right). Similarity distribution and the sample of positive pairs were constructed using Arc+UNPG with ResNet-34 on MS1MV2.

B.2 Analysis

Multi-Objective Loss vs. Unified Loss. The unified loss with UNPG, which combines two types of pair generation strategies (MLPG and CLPG), is different from the multi-objective loss, which combines two losses with a mixture weight. As shown in Table 2, the multi-objective losses such as Arc+Contrastive, Arc+Triplet, and MixFace performed worse than ArcFace. The unified loss with UNPG achieved the best performance compared to others.

Semi Hard Negative Mining vs. Noise Negative Pair Filtering. Some research[11, 17] addressed the importance of semi-hard negative mining due to the divergence problem caused by extremely difficult pairs in pair optimization tasks such as face recognition and image retrieval. We modified semi-hard mining methods [11, 17] to apply to our unified loss function and performed experiments using ArcFace with UNPG on LFW, CFP-FP, AgeDB, IJB-B and IJB-C.

$$\tilde{\mathcal{N}}^{ml}_{semi-hard} = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) | (y_i \neq y_j) \wedge (s^n_j < \inf \hat{\mathcal{S}}^p)\}$$

Table 3 shows that our noise negative pair filtering performed better than the semi-hard mining.

Why do not use positive pair of metric learning? MS1MV2[2] has erroneous labels because it is a semi-auto-labeled version of MS-Celeb-1M[3]. As shown in Fig 2, the variance of the similarity distribution of $\hat{\mathcal{S}}^p$ is very large compared to that of $\mathcal{S}^n$. Also, some of the existing pairs in the overlap of $\hat{\mathcal{S}}^p$ and $\mathcal{S}^n$ have wrong labels. Due to this, we observed that adopting positive ML pairs lead to a divergence of a loss even using different wisker sizes. In contrast, a model can tolerate erroneous positive CL pairs by adjusting the class weight vector $\boldsymbol{w}_c$. Such data characteristics may be a reason for the poor performance of ML in FR.

References

[1] Yeji Choi, Hyunjung Park, Gi Pyo Nam, Haksub Kim, Heeseung Choi, Junghyun Cho, and Ig-Jae Kim. K-face: A large-scale kist face database in consideration with unconstrained environments. *arXiv preprint arXiv:2103.02211*, 2021.

[2] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[3] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.

[4] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: Detection, Alignment, and Recognition*, 2008.

[5] Junuk Jung, Sungbin Son, Joochan Park, Yongjun Park, Seonhoon Lee, and Heung-Seon Oh. Mixface: Improving face verification focusing on fine-grained conditions. *arXiv preprint arXiv:2111.01717*, 2021.

[6] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, 2016.

[7] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

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

[9] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14225–14234, 2021.

[10] 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 workshops*, pages 51–59, 2017.

[11] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[12] 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 *IEEE Winter Conference on Applications of Computer Vision*, pages 1–9. IEEE, 2016.

[13] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6398–6407, 2020.

[14] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[15] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R Scott. Multi-similarity loss with general pair weighting for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5022–5030, 2019.

[16] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 90–98, 2017.

[17] Hong Xuan, Abby Stylianou, and Robert Pless. Improved embeddings with easy positive triplet mining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2474–2482, 2020.

[18] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.