# Semi-Siamese Training for Shallow Face Learning Supplementary Material

Hang Du[1,2⋆], Hailin Shi[2⋆], Yuchi Liu[2], Jun Wang[2], Zhen Lei[3], Dan Zeng[1⊠], and Tao Mei[2]

[1] Shanghai University, Shanghai, China.
{duhang, dzeng}@shu.edu.cn
[2] JD AI Research, Beijing, China.
{shihailin, wangjun492, tmei}@jd.com, u6009551@anu.edu.au
[3] Institute of Automation, Chinese Academy of Science, Beijing, China
zlei@nlpr.ia.ac.cn

## 1 Additional Experiments and Analysis

### 1.1 SST on Deep Data Learning

First, we provide more details about utilizing SST on deep data learning. In each iteration of deep data training, a batch of ID is randomly sampled, and for each ID, two images are randomly sampled. The arbitrary one acts as gallery, and the other one acts as probe. So, every image has 50% chance to play the role of gallery or probe. Besides, as a supplementary experiment for Section 4.5 of the main paper, we evaluate SST with loss functions of DP-softmax [11], Contrastive [8], Triplet [5] and N-pairs [7]. For evaluation, we use seven test benchmarks, including LFW [1], BLUFR [3], AgeDB [4], CFP [6], CALFW [10], CPLFW [9], MegaFace [2]. From the results, we can find all the loss functions can achieve better performance on various benchmarks after employing SST. Moreover, we can observe the original embedding loss functions (*i.e.* Contrastive, Triplet and N-pairs) have poor performance in strict FAR ranges (such as BLUFR and MegaFace); after integrated with SST, they obtain significant improvement on these benchmarks.

### 1.2 Ablation Study

In ablation study (the Section 4.2 of the main paper), we can see all the combination of "gallery queue" and "semi-siamese" (whatever network constraint or momentum) leads to the most significant boost for each training loss function. Besides, AM-softmax gains larger benefit than Arc-softmax from SST. We assume that the angular margin by Arc-softmax provides stronger supervision than AM-softmax, and such strong supervision distorts the feature space to some extent because the margin penalty performs on feature-feature pairs instead of features-FC pairs (especially training from scratch).

⋆ Equal contribution. This work was performed at JD AI Research.

Table 1: Comparison of Semi-Siamese Training and the conventional training on deep data. In MegaFace, "Id." refers to face identification rank1 accuracy with 1M distractors, and "Veri." refers to face verification rate at 1e-6 FAR.

| Method | LFW | AgeDB | CFP | CALFW | CPLFW | BLUFR | MegaFace | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Id. | Veri. |
| DP-softmax | 99.63 | 95.68 | 91.74 | 93.03 | 83.88 | 92.37 | 89.27 | 90.94 |
| Contrastive | 99.50 | 92.91 | 91.76 | 87.56 | 80.13 | 74.72 | 60.14 | 63.59 |
| Triplet | 99.47 | 93.32 | 94.49 | 89.32 | 82.25 | 79.10 | 65.65 | 69.18 |
| N-pairs | 99.53 | 94.58 | 93.43 | 92.10 | 83.15 | 85.19 | 76.87 | 78.28 |
| SST (DP-softmax) | **99.68** | **96.24** | 94.56 | 93.78 | **86.04** | **94.78** | **92.08** | **93.57** |
| SST (Contrastive) | 99.56 | 93.14 | 92.71 | 92.13 | 81.78 | 87.95 | 77.59 | 82.44 |
| SST (Triplet) | 99.50 | 94.30 | 93.30 | 92.05 | 82.67 | 89.72 | 81.76 | 83.29 |
| SST (N-pairs) | 99.65 | 96.12 | **94.86** | **94.32** | 84.74 | 94.17 | 91.72 | 93.48 |

### 1.3 Pretrain and Finetune

In pretrain and finetune experiment (the Section 4.6 of the main paper), we can find the different improvement for softmax-based methods (softmax, AM-softmax, Arc-softmax) and pair/triplet-based methods (contrastive, triplet, N-pairs). We argue the heavy parameters of original softmax-based methods in classification FC layer brings the sub-optimal results in this experiment. After integrating with SST, the FC layer is replaced by an updating feature queue, which can significantly alleviate the optimization issue. Meanwhile, the pair/triplet-based methods adopt features rather than FC layer in the original version. So, the benefit brought by SST for softmax-based methods is larger than for pair/triplet-based methods in the finetuning stage.

## References


1. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007) 1
2. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4873–4882 (2016) 1
3. Liao, S., Lei, Z., Yi, D., Li, S.Z.: A benchmark study of large-scale unconstrained face recognition. In: IEEE international joint conference on biometrics. pp. 1–8 (2014) 1
4. 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. pp. 51–59 (2017) 1
5. 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. pp. 815–823 (2015) 1
6. Sengupta, S., Chen, J.C., Castillo, C., Patel, V.M., Chellappa, R., Jacobs, D.W.: Frontal to profile face verification in the wild. In: 2016 IEEE Winter Conference on Applications of Computer Vision. pp. 1–9. IEEE (2016) 1
7. Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective. In: Advances in Neural Information Processing Systems. pp. 1857–1865 (2016) 1
8. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: Advances in neural information processing systems. pp. 1988–1996 (2014) 1
9. Zheng, T., Deng, W.: Cross-pose lfw: A database for studying crosspose face recognition in unconstrained environments. Beijing University of Posts and Telecommunications, Tech. Rep pp. 18–01 (2018) 1
10. Zheng, T., Deng, W., Hu, J.: Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197 (2017) 1
11. Zhu, X., Liu, H., Lei, Z., Shi, H., Yang, F., Yi, D., Qi, G., Li, S.Z.: Large-scale bisample learning on id versus spot face recognition. International Journal of Computer Vision **127**(6-7), 684–700 (2019) 1