VirFace: Enhancing Face Recognition via Unlabeled Shallow Data

Wenyu Li[1,*], Tianchu Guo[*], Pengyu Li, Binghui Chen, Biao Wang, Wangmeng Zuo[1(✉)], Lei Zhang[2]
[1]School of Computer Science and Technology, Harbin Institute of Technology, China
[2]Hong Kong Polytechnic University, Hong Kong, China
liwenyu27@gmail.com, antares_tcguo@163.com, lipengyu007@gmail.com,
chenbinghui@bupt.edu.cn, wangbiao225@foxmail.com, wmzuo@hit.edu.cn, cslzhang@comp.polyu.edu.hk

Methods	**MegaFace**	
	Veri.	**Id.**
ResNet50 Baseline	62.15	58.42
Cluster (#sample>1)	45.57	44.23
UIR [5]	64.36	65.16
N-pair [3]	63.86	59.75
VirClass	**70.72**	**67.22**
VirFace	**75.40**	**72.25**

Table 1. Face identification and verification results on MegaFace Challenge1 using FaceScrub as the probe set. "Veri." refers to face verification TAR at 1e-6 FAR, and "Id." refers to face identification rank1 accuracy with 1M distractors. Our methods are shown in bold, and the best results are shown in red.

1. Additional Experiment Results

In this section, we present evaluation results on MegaFace [1] and ablation study results on IJB-B [4] and IJB-C [2]. In these test datasets, we present both face verification TAR rate and face identification rank1 accuracy.

1.1. Evaluation Results

Table 1 shows the MegaFace results of our proposed method and the conventional semi-supervised methods. The face identification and verification results show that our proposed method can significantly improve the performance and outperform the conventional semi-supervised methods.

1.2. VirClass

Influence of Shallow Rate and Identity Number. As introduced in the paper Section 4.3.1, the scale of the unlabeled training dataset is fixed to 80,068, and evaluate different combinations of shallow rate and identity number. Table 2 shows the verification and identification results on IJB-B and IJB-C. From the results, we find that the performance has little change for different combinations which can demonstrate the completion in Section 4.3.1 of the paper.

Methods	**IJB-B**		**IJB-C**	
	Veri.	**Id.**	**Veri.**	**Id.**
ResNet50 Baseline	57.84	72.14	61.06	71.80
shallow rate = 1 80,068 ids	60.73	74.32	64.60	74.36
shallow rate = 2 40,034 ids	61.38	74.39	65.21	74.43
shallow rate = 5 16,014 ids	61.57	73.97	64.40	74.02

Table 2. Result of different combination of shallow rate and identity number when fixing the scale of the unlabeled training set. "Veri." refers to face verification rate at 1e-4 FAR, and "Id." refers to face identification rank1 accuracy.

Methods	**IJB-B**		**IJB-C**	
	Veri.	**Id.**	**Veri.**	**Id.**
ResNet50 Baseline	57.84	72.14	61.06	71.80
80,068 samples shallow rate = 1	60.73	74.32	64.60	74.36
160,136 samples shallow rate = 2	62.03	75.37	65.46	75.31
400,222 samples shallow rate = 5	62.37	75.93	66.26	76.25

Table 3. Results of different scales of the unlabeled training set. "Veri." refers to face verification rate at 1e-4 FAR, and "Id." refers to face identification rank1 accuracy.

Influence of the Scale of the Unlabeled Training Set. In this part, the identity number is fixed to 80,068 in order to guarantee the diversity. The influence of different the scale of the unlabeled training set is evaluated in IJB-B and IJB-C. The results are shown in Table 3. The IJB-B and IJB-C results support the conclusion in the paper Section 4.3.1 that the scale of the unlabeled training set is the most important factor on the improvement of VirClass.

[*] Equal contribution.

Methods	IJB-B		IJB-C	
	Veri.	**Id.**	**Veri.**	**Id.**
VirClass Baseline	60.73	74.32	64.60	74.36
VirFace (Sampling number = 2)	63.56	75.36	67.03	75.49
VirFace (Sampling number = 5)	64.34	76.23	67.67	76.31
VirFace (Sampling number = 10)	63.60	75.86	66.92	75.70
VirFace (Sampling number = 20)	63.54	75.13	66.81	75.30

Table 4. Results of different sampling number on distribution generator. “Veri.” refers to face verification rate at 1e-4 FAR, and “Id.” refers to face identification rank1 accuracy.

Methods	IJB-B		IJB-C	
	Veri.	**Id.**	**Veri.**	**Id.**
VirClass Baseline	60.73	74.32	64.60	74.36
VirClass + DataAug (Generation number = 5)	60.84	73.67	65.55	73.42
VirFace (Sampling number = 5)	64.34	76.23	67.67	76.31

Table 5. Comparison with data augmentation method. “Veri.” refers to face verification rate at 1e-4 FAR, and “Id.” refers to face identification rank1 accuracy.

1.3. VirInstance

Sampling Number of Distribution Generator. We study the effect of different sampling number of distribution generator on the 1-shallow rate VirClass model. Table. 4 is an addition to Table 6 in the paper. The performance gets better as the sampling number increases. When sampling number is 5, our VirFace method achieves the best performance. The overall performance drops a little and tends to be stable when sampling number exceeds 5.

Comparison with Data Augmentation. In this part, we compare distribution generator of VirInstance method with traditional data augmentation method on IJB-B and IJB-C datasets as an addition of Table 8 in the paper. The data augmentation generates different instances of images. We implement the random combination of blur and color jitters as data augmentation. We generate 5 instances for both methods. The results are shown in Table. 5. Our method outperforms on both IJB-B and IJB-C.

Analysis of VirInstance on the Feature Space. In this part, we generate instances through our proposed distribution generator on both the labeled and the unlabeled datasets. We define the classification accuracy as the rate that the generated features can be correctly classified to the identity which the original feature belongs to. Figure 1(a) shows the classification accuracy on both the labeled dataset denoting as labeled and the unlabeled dataset denoting as unlabeled. From this figure, the unlabeled accuracy maintains at a high level which means the generated features do not introduce many noises. The labeled accuracy starts from a low level which is about 75%, and quickly increases to above 95% at epoch 8. This indicates that the generated features prefer to be hard samples which can bring enough variance, and as training progresses, these hard sample information can be learned.

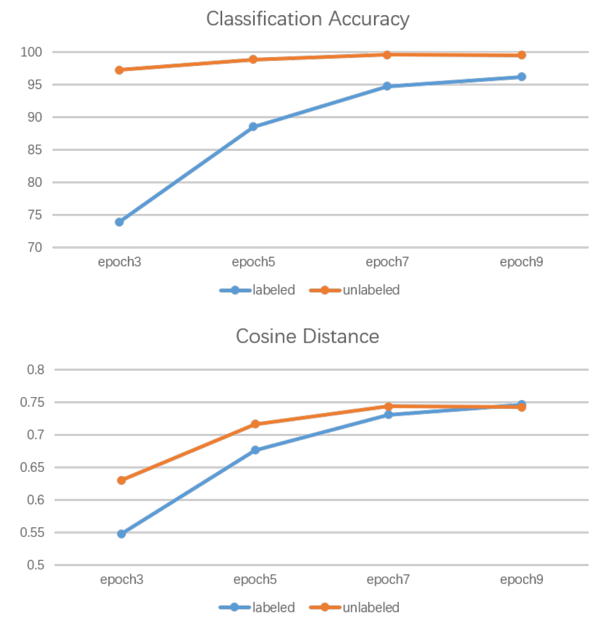

Figure 1. Classification accuracy and cosine distance on the labeled and unlabeled data.

We also calculate the cosine distance between the generated features and their corresponding centroids. For the labeled features, their centroids are the weights of the last FC layer, while for the unlabeled features, the centroids are the original unlabeled feature as described in VirClass section in Methods. Figure 1(b) presents the results on both the labeled and unlabeled data. From this figure, the cosine distance increases as the training progresses both on the labeled and on the unlabeled data. This means our proposed VirInstance method can compact the intra-class distance.

2. Details of the VAE Network

We use a VAE network to predict the feature distribution and generate virtual instances. The detailed architecture is shown below.

The sampling layer is implemented through the equation below:

$$z = \boldsymbol{\alpha} + e^{\boldsymbol{\beta}/2} * \boldsymbol{\gamma} \tag{1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is the output of the encoder, $\boldsymbol{\gamma}$ is a random matrix following the normal distribution.

We use the encoder to predict the feature distribution which is presented by the output of the encoder denoting

Input	Input Feature, 512-D	
Encoder	FC layer, ReLU, 256-D	
	FC layer, 128-D	FC layer, 128-D
Sampling layer		
Decoder	FC layer, ReLU, 256-D	
	FC layer, 512-D	
Output	Output Feature, 512-D	

Table 6. Architecture of distribution generator.

as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Then, the sampling layer randomly samples virtual instances from the feature distribution and the decoder reconstructs the sampled virtual instances to 512-D features. In the loss function, we use the common KL divergence loss in training:

$$L_{KL} = \frac{1}{2}\sum_{i=1}^{d}(\boldsymbol{\alpha}^2 + e^{\boldsymbol{\beta}} - \boldsymbol{\beta} + 1) \tag{2}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the outputs of the encoder which represent the distribution.

References

[1] 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.

[2] 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 (ICB)*, pages 158–165. IEEE, 2018.

[3] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in neural information processing systems*, pages 1857–1865, 2016.

[4] 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.

[5] Haiming Yu, Yin Fan, Keyu Chen, He Yan, Xiangju Lu, Junhui Liu, and Danming Xie. Unknown identity rejection loss: Utilizing unlabeled data for face recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.