Supplementary Material for
Global-Local GCN: Large-Scale Label Noise Cleansing for Face Recognition

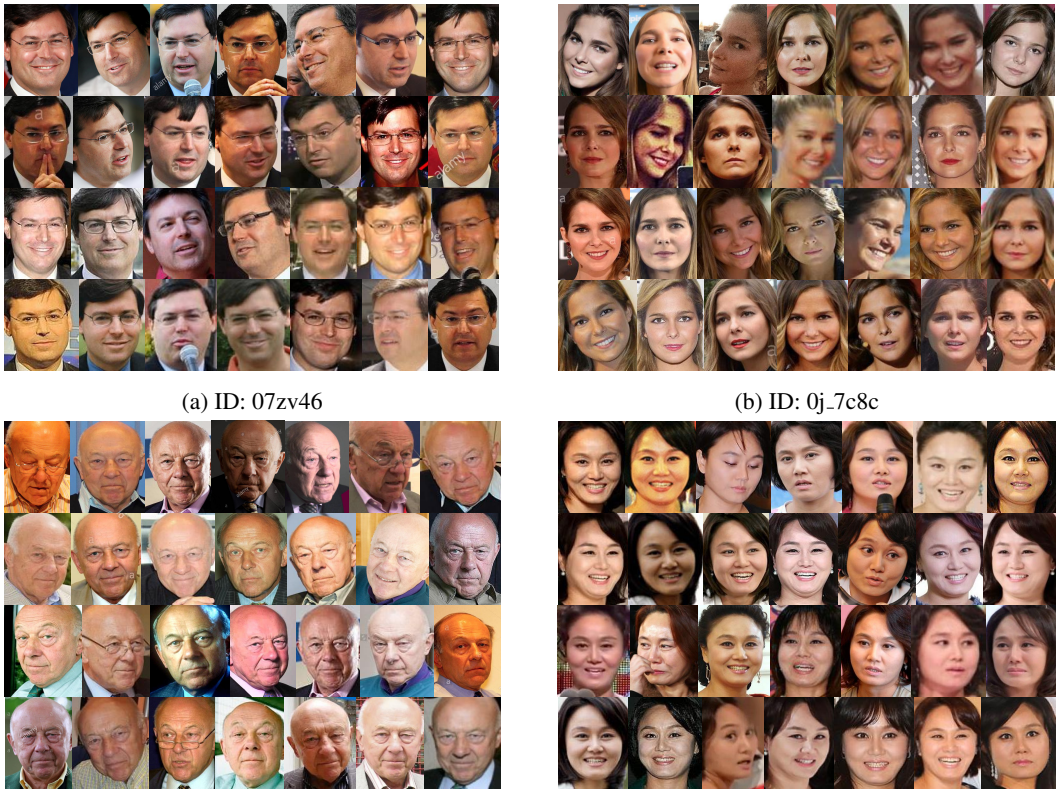
Abstract

In this supplementary material, we present fully detailed information on 1) the proposed MillionCelebs dataset; 2) the Cooperative Learning algorithm; 3) wrong case analysis; 4) a comparison with noisy label learning methods.

S1. The MillionCelebs Dataset

To promote state-of-the-art face recognition performance and facilitate the study on large-scale deep learning, we collect the MillionCelebs dataset, which contains 87.0M images of 1M celebrities originally, and 18.8M images of 636K celebrities after cleansed by FaceGraph.

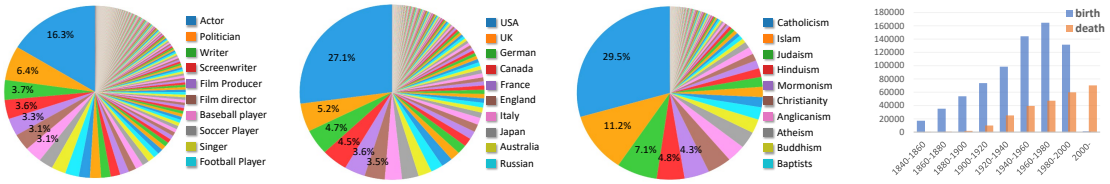
With a name list of 1M celebrities from Freebase [1] provided by Guo *et al*. [5], we download 50-100 images for each identity from Internet Image Search Engine in three months. Since the original images take up too much space, MTCNN [10] is used to synchronously detect faces, and only the cropped face warps are stored. Following the image saving protocol of VGGFace2 [2], we save the face warps within 1.3 times the bounding boxes. For training, the faces are aligned with similarity transformation, resized to the shape 112×112, and normalized by subtracting 127.5



(a) ID: 07zv46 (b) ID: 0j_7c8c

(c) ID: 0bvpk2 (d) ID: 0jy0sy5

Figure S1: Examples of MillionCelebs cleansed by FaceGraph of four identities.



(a) Profession (b) Nationality (c) Religion (d) Date of birth/death

Figure S2: Detailed demography statistics of MillionCelebs cleansed by FaceGraph.

Figure S3: **From left to right**: Left/Removed images in one class.

and divided by 128. Figure S1 shows example images of four identities after cleansed by FaceGraph. As shown in the Figure, MillionCelebs provides in-the-wild face images of high quality and cleanliness, and also contains a big variety for one person. In Figure S3, we visualize the result of cleansing ID "05f5ck7" to intuitively show the performance of FaceGraph. Faces in the left block remain in the dataset, and faces in the right block are removed. It is observed that the search engine indeed returns many incorrect images, and the incorrect people usually have *entity relationships* with the correct one. For example, searching for an actor can also get his partner, and searching for a coach can also get his teammates. FaceGraph performs well at distinguishing wanted faces in a noisy environment.

The paper describes brief information about MillionCelebs. Figure S2 shows more detailed demography statistics. Different from many celebrity datasets in which most identities are actors, MillionCelebs contains a big range of professions. Celebrities in MillionCelebs are a subset of the large collaborative knowledge base, Freebase [1], in where we can extract personal information such as gender, ethnicity, profession, nationality, religion, and date of birth and death. With the abundant statistical information, we can easily select a subset of MillionCelebs to meet the research needs, for instance, the race or gender bias problem [8] in deep face recognition.

S2. Cooperative Learning

We present detailed training procedures. Algorithm 1 separately trains GGN and LGN. Algorithm 2 trains FaceGraph with Cooperative Learning (CL). The end-to-end CL algorithm effectively unifies global and local scales so that GGN and LGN can promote each other during training.

Algorithm 1 FaceGraph - GGN + LGN

Input: Global Graph Net $\mathbf{G}_\theta$, Local Graph Net $\mathbf{L}_\phi$, training set $\mathbb{S} = \left\{\left(\mathcal{G}, X, \hat{Y}\right)\right\}$, number of GGN iterations I_G, number of LGN iterations I_L, batch size N

Output: optimal parameters θ, ϕ

- Initial θ and ϕ.

for i = $1, \cdots, I_G$ **do**

- Randomly select N samples from set $\mathbb{S}$ to get the input mini-batch M.
- Update θ by the GGN loss $\mathcal{L}_G$.

end for

for i = $1, \cdots, I_L$ **do**

- Randomly select N samples from set $\mathbb{S}$ to get the input mini-batch M.
- Forward propagate $\mathbf{G}_\theta$ with M to get input graphs and features $\left\{\left(\mathcal{G}_L, X_L, \hat{Y}_L\right)\right\}$ for $\mathbf{L}_\phi$.
- Update ϕ by the LGN loss $\mathcal{L}_L$.

end for

Algorithm 2 FaceGraph - CL

Input: Global Graph Net $\mathbf{G}_\theta$, Local Graph Net $\mathbf{L}_\phi$, training set $\mathbb{S} = \left\{\left(\mathcal{G}, X, \hat{Y}\right)\right\}$, number of iterations I, batch size N, scaling factor α

Output: optimal parameters θ, ϕ

- Initial θ and ϕ.

for i = $1, \cdots, I$ **do**

- Randomly select N samples from set $\mathbb{S}$ to get the input mini-batch M.
- Update θ by the GGN loss $\mathcal{L}_G$.
- Forward propagate $\mathbf{G}_\theta$ with M to get input graphs and features $\left\{\left(\mathcal{G}_L, X_L, \hat{Y}_L\right)\right\}$ for $\mathbf{L}_\phi$.
- Update ϕ by the LGN loss $\mathcal{L}_L$.
- Update θ by $\alpha \times \mathcal{L}_L$.

end for

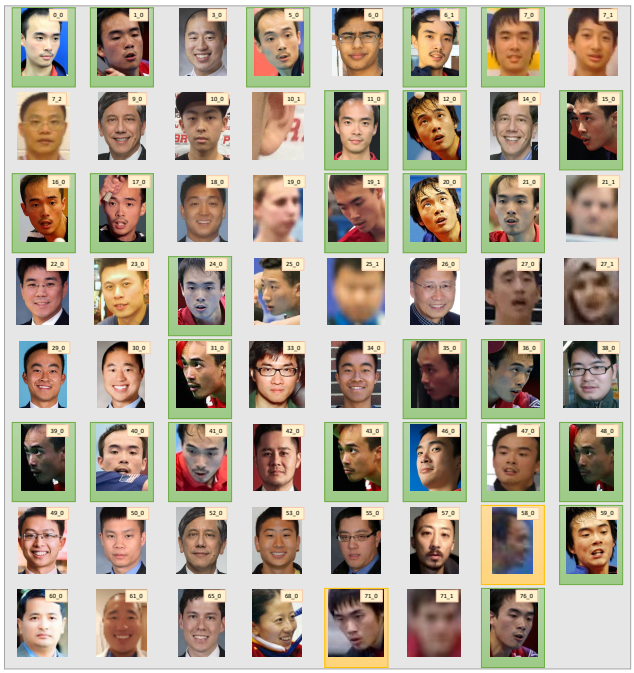


Figure S4: Cleansing one identity. Green rectangles: the true positives. Yellow rectangles: the false negatives.

S3. Wrong Case Analysis

Figure S4 shows the result of cleansing ID "0k8rzzq". There are 63 face images downloaded from the search engine with tags in the upper right corner. The positive samples are marked with green or yellow rectangles. The noise rate is as high as 55%. The green rectangles mark all true positives, and the yellow rectangles mark all false negatives. It is observed that no negative images are accepted, but two positive images are removed by mistake, resulting in 100% precision and 92.8% recall. "58_0" is removed because of low resolution and large pose, and "71_0" is removed because of the large age span. Therefore, although FaceGraph achieves remarkable cleansing results in general, how to distinguish such difficult cases is still worth further study.

S4. Noisy Label Learning

There are usually two methods to effectively address the label noise problem: data cleansing and noisy label learning. The data cleansing methods attempt to remove the label noise directly to obtain better training data. The proposed FaceGraph is a novel large-scale data cleansing algorithm based on GCN, which can achieve state-of-the-art cleansing performance on the face recognition datasets. On the other hand, the noisy label learning methods deploy all noisy data for training and design a filtering algorithm to reduce the impact of noisy data on the training process, that is, to achieve end-to-end cleansing and training. If the filtering algorithm is designed properly, the noisy label learning methods can make up for the loss caused by the wrongly cleansed data in the data cleansing methods. For example, state-of-the-art Co-Mining [9] deploys two peer networks to detect noisy labels with the loss values, then exchanges the high-confidence clean faces to alleviate the errors accumulated issue and re-weights the predicted clean faces to make them dominant to learn discriminative features.

Method	CALFW	CPLFW	AgeDB	CFP	Avg.
Co-Mining	91.06	87.31	**94.05**	**95.87**	92.07
FaceGraph	**91.52**	**88.85**	93.98	95.69	**92.51**

Table S1: Results (%) of noisy label learning and data cleansing methods training on VGGFace2 [2] dataset.

Method	CALFW	CPLFW	AgeDB	CFP	Avg.
Co-Mining	93.28	85.70	95.80	93.32	92.02
FaceGraph	**94.23**	**87.42**	**95.85**	**94.99**	**93.12**

Table S2: Results (%) of noisy label learning and data cleansing methods training on MS1M [5] dataset.

This raises an intuitive question: Which of the data cleansing and noisy label learning is more effective when processing a face recognition dataset? The performance comparison between FaceGraph and Co-Mining [9] is reported in Table S1 and Table S2. For a fair comparison, we follow the same experimental setup as in Co-Mining [9]. For example, MobileFaceNet [3] with 512-dimension output features is trained from scratch with batch size 512. m and s in ArcFace loss [4] are set 0.5 and 32, respectively. CALFW [12], CPLFW [11], AgeDB [6] and CFP [7] are used for evaluation. It is observed that FaceGraph outperforms Co-Mining [9] on processing noisy MS1M [5] and VGGFace2 [2]. For the less noisy VGGFace2, FaceGraph performs better on CALFW [12] and CPLFW [11] in the four evaluation sets. The model trained by MS1M that is cleansed by FaceGraph comprehensively surpasses Co-Mining on the four evaluation sets to improve the average accuracy by 1.10%. This shows that state-of-the-art cleansing method FaceGraph performs better than state-of-the-art noisy label learning method Co-Mining, especially in the case of big noise. This is as expected because most noisy label learning methods like Co-Mining [9] are hard to converge from scratch and hard to distinguish signals from large noise. As illustrated in the paper, the proposed FaceGraph aims to cleanse large-scale severely noisy data like collected data from the web. Unfortunately, noisy label learning approaches are less effective in this case.

References

[1] Freebase data dump. www.freebase.com.

[2] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 67–74. IEEE, 2018.

[3] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Chinese Conference on Biometric Recognition*, pages 428–438. Springer, 2018.

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

[5] 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.

[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 *Computer Vision and Pattern Recognition Workshops*, pages 1997–2005, 2017.

[7] C.D. Castillo V.M. Patel R. Chellappa D.W. Jacobs S. Sengupta, J.C. Cheng. Frontal to profile face verification in the wild. In *IEEE Conference on Applications of Computer Vision*, February 2016.

[8] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 692–702, 2019.

[9] Xiaobo Wang, Shuo Wang, Jun Wang, Hailin Shi, and Tao Mei. Co-mining: Deep face recognition with noisy labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9358–9367, 2019.

[10] Jia Xiang and Gengming Zhu. Joint face detection and facial expression recognition with mtcnn. In *Information Science and Control Engineering (ICISCE), 2017 4th International Conference on*, pages 424–427. IEEE, 2017.

[11] T. Zheng and W. 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.

[12] 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.