Supplementary Material for Virtual Fully-Connected Layer

Pengyu Li[1], BiaoWang[1], Lei Zhang[1,2]
[1] Artificial Intelligence Center, DAMO Academy, Alibaba Group
[2]Department of Computing, The Hong Kong Polytechnic University
lipengyu007@gmail.com, wangbiao225@foxmail.com, cslzhang@comp.polyu.edu.hk

1. Summary

The Supplementary Material will include:

1) The analysis of the corresponding anchor generation methods in Section 2. One is the estimation with a two-layer Multi-layer Perceptron (MLP) based on the attention mechanism. The other one set the $\{\alpha_{i,l}\}$ to be a constant value and make $anchor_{corr,l}$ be the centroid.

2) Report the performance on CALFW [9], CPLFW [8], SLLFW [1], and YTF [6] datasets in Section 3. Besides, the TopK accuracy and TAR under varying FAR on IJB-A [3], IJB-B [5], IJB-C [4] datasets are reported on this section.

3) Table 4 in the submitted paper did not report the performance on MegaFace [2] because of the paper width limitation. We report it in Section 4 of this material.

2. Analysis of the Corresponding Anchor Generation methods

We propose the corresponding anchor estimated by a weighted average function, as Equation 2 in the submitted paper shows. We copy the equation in the following:

$$anchor_{corr,l} = \frac{\sum_{i=1}^{K} \alpha_{i,l} f_{i,l}}{\sum_{i}^{K} \alpha_{i,l}} \tag{1}$$

$f_{i,l}$ is the feature of i-th image which belongs to the group l. Because we hypothesize there is no anchor conflict in this section, the $\{f_{i,l}\}(i = 1, 2, ..., K)$ belongs to a single identity. $\alpha_{i,l}$ is the attention estimation to weight the $f_{i,l}$. $\{\alpha_{i,l}\}$ can be estimated by the attention mechanism or set to be a constant value. If $\{\alpha_{i,l}\}(i = 1, 2, ..., k)$ equal to a constant value, the $anchor_{corr,l}$ is the centroid of $\{f_{i,l}\}$.

In this section, we analyze two methods of the $\{\alpha_{i,l}\}$ generation: 1) Constant value. 2) Prediction of an attention module. We build the attention module as a two-layer Multi-layer Perceptron (MLP). The structure of MLP is inspired by the MetaCleaner [7] and illustrated in Figure 1. The comparison of their performance is shown in Table 1. Besides, we sample 10,000 images from the training dataset randomly to get the histograms of $\{\alpha_{i,l}\}$. The histograms are shown in the Figure 2.

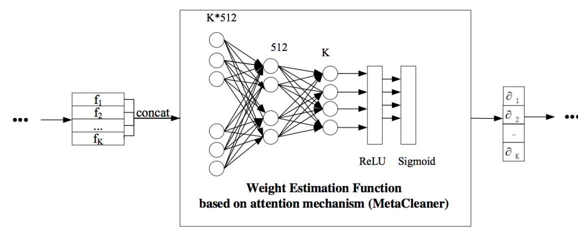

Figure 1. Weight Estimation Function based on Attention mechanism (MetaCleaner). K is the number of sampling images per identity in a mini-batch.

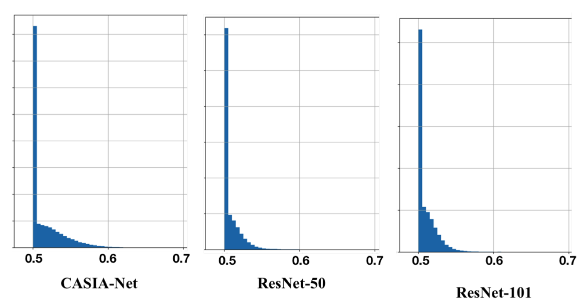

Figure 2. Histogram of $\{\alpha_{i,l}\}$

Table 1 shows that the performance of two corresponding anchor generation methods is comparable in all three backbones. The histograms in Figure 2 proves that the $\{\alpha_{i,l}\}$ estimated by an attention module tend to be a constant value (0.5). Based on the discovery, we use the centroid directly in our submitted paper. One of the advantages of using the centroid is that the centroid calculation is flexible because it does not require a fixed K.

3. Performance on Several Evaluation Dataset

In this section, we report:

1) The performance on CALFW [9], CPLFW [8], SLLFW [1], and YTF [6] datasets in Table 2

2) The TopK accuracy and TAR under varying FAR on

Network	Strategy	LFW	CFP-FF	CFP-FP	IJB-A	IJB-B	IJB-C	MegaFace
CASIA-Net	Average (Centroid)	98.75±0.27	99.07±0.30	93.04±1.84	91.77	90.78	**92.21**	81.44
	Weighted Average (Attention Mechanism)	**98.77±0.75**	**99.26±0.39**	**93.19±1.63**	**92.37**	**91.82**	91.82	**82.72**
ResNet-50	Average (Centroid)	99.32±0.27	99.73±0.33	**95.77±1.11**	**92.83**	93.21	94.53	93.18
	Weighted Average (Attention Mechanism)	**98.33±0.60**	**99.74±0.29**	95.66±1.15	92.79	**93.46**	**94.77**	**93.47**
ResNet-101	Average (Centroid)	**99.38±0.38**	99.61±0.31	95.55±1.42	**93.69**	**94.05**	**95.30**	**94.04**
	Weighted Average (Attention Mechanism)	98.70±0.95	**99.66±0.30**	**95.91±1.22**	93.11	93.85	95.19	93.45

Table 1. Performance Comparison of Corresponding Anchor Generation Methods.

IJB-A, IJB-B, IJB-C datasets in Table 3, Table 4, and Table 5

We can get all the conclusions that are derived in the submitted paper, again: 1) Our Virtual FC surpasses the lower boundary and all other candidate solutions consistently and significantly. It also achieves comparable performance to the upper boundary with 1% computational resource of the FC layer. 2) The superiority of our Virtual FC is more significant in complex neural network structure (e.g., ResNet50 and ResNet101) than in a simple one (CASIA-Net).

Furthermore, we find the superiority of our Virtual FC is more significant in the tough evaluation datasets/protocols (e.g., CALFW, CPLFW, SLLFW) than the simple ones (e.g., YTF).

		Evaluation Dataset			
		CALFW	CPLFW	SLLFW	YTF
CASIA-Net	lower boundary	86.77	73.60	92.97	93.34
	upper boundary	90.78	77.9	96.57	94.02
	N-pair	88.50	74.48	94.00	92.84
	Multi-similarity	88.08	74.25	94.02	93.00
	TCP	88.9 0	76.32	94.47	93.50
	ours:VFC	**89.35**	**76.78**	**94.48**	**93.92**
ResNet-50	lower boundary	87.43	75.45	93.52	93.78
	upper boundary	94.46	85.03	98.72	95.88
	N-pair	87.32	72.80	92.28	92.62
	Multi-similarity	85.4 0	73.60	91.03	92.76
	TCP	88.05	76.00	93.23	93.92
	ours:VFC	**91.93**	**79.00**	**96.23**	**95.08**
ResNet-101	lower boundary	88.78	76.72	94.00	93.86
	upper boundary	94.88	86.23	98.97	96.16
	N-pair	85.27	74.22	90.97	93.02
	Multi-similarity	85.52	73.07	90.33	92.66
	TCP	91.45	78.33	95.55	95.14
	ours:VFC	**92.27**	**79.03**	**96.67**	**95.32**

Table 2. Performance On CALFW, CPLFW, SLLFW, YTF

4. performance of MegaFace on submitted Table 4

IJB-A		10^{-1}	10^{-2}	10^{-3}	10^{-4}
CASIA-Net	lower boundary	96.82	89.90	77.61	67.61
	upper boundary	98.04	93.89	83.71	66.17
	N-pair	95.47	85.65	54.54	17.04
	Multi-similarity	96.33	85.53	47.18	13.50
	TCP	97.35	89.75	72.22	46.35
	ours:VFC	**97.69**	**91.77**	**79.41**	**61.65**
ResNet-50	lower boundary	97.11	90.59	78.13	61.53
	upper boundary	98.89	97.33	94.57	90.65
	N-pair	95.95	85.32	70.30	55.03
	Multi-similarity	96.36	83.49	62.81	47.3
	TCP	96.97	85.53	62.71	37.43
	ours:VFC	**97.84**	**92.83**	**78.72**	**64.43**
ResNet-101	lower boundary	97.24	90.33	78.38	64.05
	upper boundary	99.00	97.81	95.95	93.29
	N-pair	96.72	83.55	62.32	48.06
	Multi-similarity	95.92	82.02	57.53	36.84
	TCP	97.77	89.23	66.23	40.76
	ours:VFC	**98.35**	**93.69**	**84.33**	**73.37**

Table 3. Performance On IJB-A

IJB-B		1 vs 1				1 vs N		
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	top 1	top 5	top 10
CASIA-Net	lower boundary	96.61	88.40	76.98	55.41	82.3	88.7	90.86
	upper boundary	98.21	93.54	84.25	59.95	85.16	90.45	92.50
	N-pair	97.00	87.25	62.48	15.02	76.71	81.84	83.69
	Multi-similarity	96.95	85.21	55.14	9.33	74.19	80.19	82.96
	TCP	97.41	90.32	76.67	42.5	82.85	88.32	90.41
	ours:VFC	**97.68**	**90.78**	**78.47**	**56.55**	**83.80**	**89.88**	**91.77**
ResNet-50	lower boundary	97.43	91.26	80.75	59.17	83.91	90.40	92.32
	upper boundary	98.78	96.71	92.16	82.41	90.49	94.04	95.32
	N-pair	97.24	88.46	73.93	55.46	78.01	85.57	88.20
	Multi-similarity	97.47	88.30	72.32	52.94	75.65	84.53	87.75
	TCP	97.95	90.35	68.17	36.71	82.55	89.51	91.83
	ours:VFC	**98.29**	**93.21**	**80.92**	**64.78**	**86.79**	**91.95**	**93.95**
ResNet-101	lower boundary	97.32	91.19	80.7	65.56	85.51	91.26	92.97
	upper boundary	98.82	97.01	93.22	83.74	91.91	95.03	96.10
	N-pair	97.47	86.57	69.05	43.65	77.44	86.75	89.57
	Multi-similarity	96.56	85.65	60.50	32.79	79.37	87.51	89.89
	TCP	98.44	92.67	74.52	46.68	86.06	92.57	94.22
	ours:VFC	**98.57**	**94.05**	**83.26**	**67.44**	**86.94**	**92.82**	**94.52**

Table 4. Performance On IJB-B

IJB-C		1 vs 1				1 vs N		
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	top 1	top 5	top 10
CASIA-Net	lower boundary	97.07	89.94	79.13	56.74	83.64	88.72	90.65
	upper boundary	98.41	94.37	85.59	64.03	86.05	90.4	91.94
	N-pair	97.19	88.69	64.42	14.62	76.71	81.84	83.69
	Multi-similarity	97.11	85.90	53.06	10.01	74.29	79.72	81.48
	TCP	97.75	91.86	79.15	48.61	83.53	88.21	89.87
	ours:VFC	**97.79**	**92.21**	**80.51**	**60.69**	**84.58**	**89.66**	**91.37**
ResNet-50	lower boundary	97.89	92.78	83.4	65.19	84.66	90.20	92.33
	upper boundary	98.94	97.19	93.25	83.94	91.43	94.07	94.98
	N-pair	97.44	90.08	77.54	61.75	83.60	88.87	90.85
	Multi-similarity	97.73	90.06	75.97	57.82	75.74	83.39	86.32
	TCP	98.16	92.08	74.86	43.58	83.03	89.02	90.91
	ours:VFC	**98.48**	**94.53**	**84.81**	**70.12**	**87.38**	**91.7**	**93.39**
ResNet-101	lower boundary	97.75	92.38	83.79	71.02	86.51	91.31	92.90
	upper boundary	98.98	97.61	94.25	85.60	92.90	95.02	95.80
	N-pair	97.75	88.72	74.71	51.13	77.75	86.10	88.85
	Multi-similarity	97.09	87.50	65.95	38.87	80.27	87.53	89.83
	TCP	98.76	94.22	79.66	52.71	86.87	91.92	93.72
	ours:VFC	**98.84**	**95.30**	**86.44**	**71.47**	**87.84**	**92.3**	**94.27**

Table 5. Performance On IJB-C

Network	Strategy	MegaFace
CASIA-Net	Sampling	78.15
	Re-grouping	**81.44**
ResNet-50	Sampling	84.12
	Re-grouping	**93.18**
ResNet-101	Sampling	74.94
	Re-grouping	**94.04**

Table 6. The MegaFace Performance which is ignored in Table 4 of the submitted paper.

References

[1] Weihong Deng, Jiani Hu, Nanhai Zhang, Binghui Chen, and Jun Guo. Fine-grained face verification: Fglfw database, baselines, and human-dcmn partnership. *Pattern Recognition*, 66:63–73, 2017. 4321

[2] 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 (CVPR)*, pages 4873–4882, 2016. 4321

[3] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015. 4321

[4] 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. 4321

[5] 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. 4321

[6] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534. IEEE, 2011. 4321

[7] Weihe Zhang, Yali Wang, and Yu Qiao. Metacleaner: Learning to hallucinate clean representations for noisy-labeled visual recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7373–7382, 2019. 4321

[8] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5, 2018. 4321

[9] 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. 4321