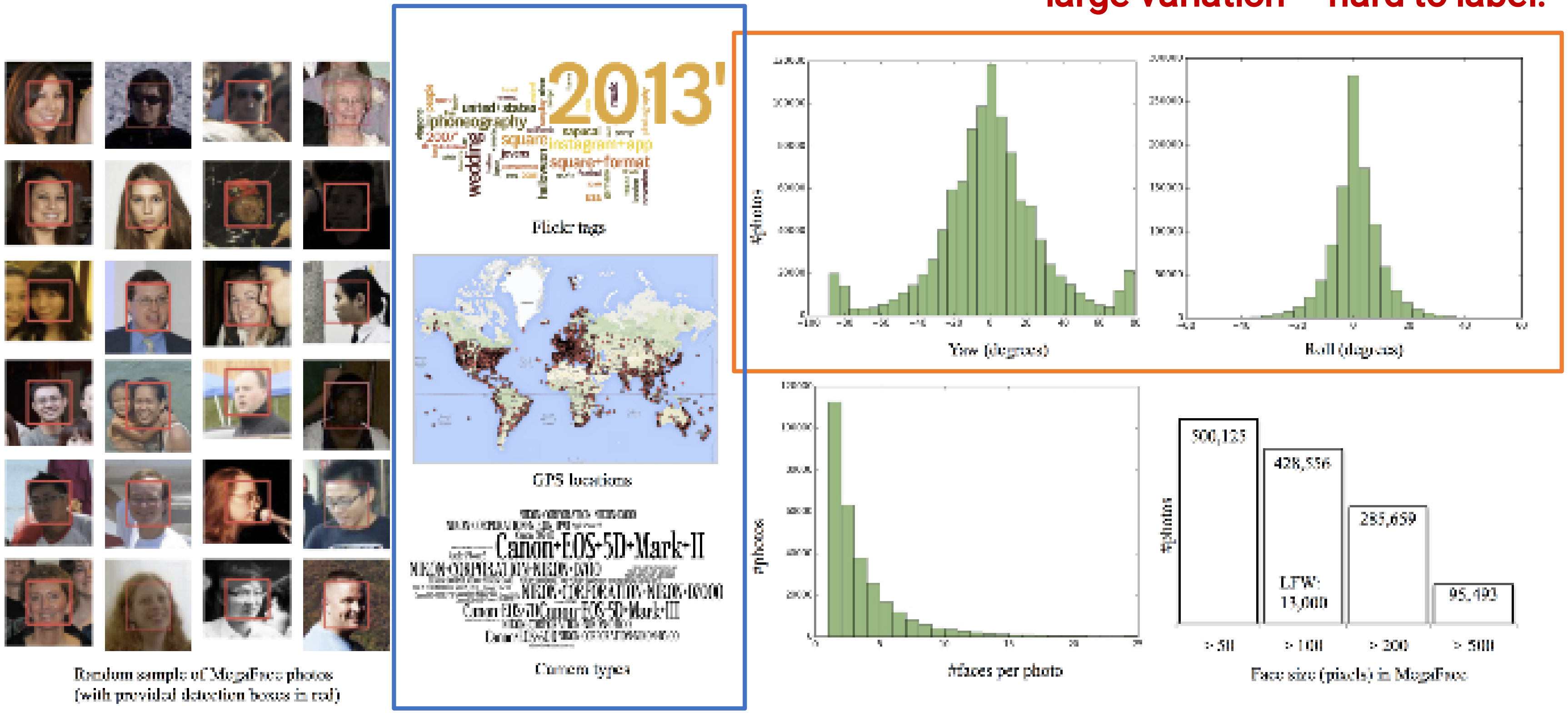
Consensus-Driven Propagation in Massive Unlabeled Data for Face Recognition

Hongwei Fan

Lab of Pattern Recognition and Intelligent Systems, BUPT

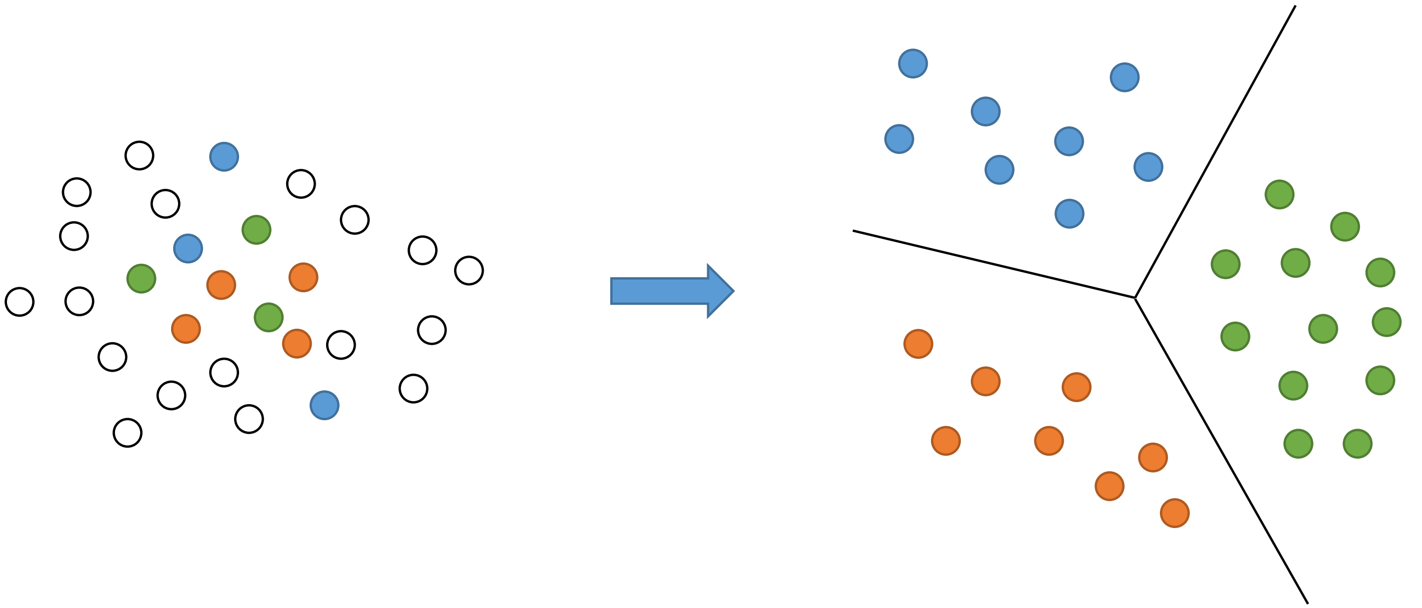
Motivation: Unlabeled Data with Large Variation

large variation – hard to label!



Example: Statistic Data of MegaFace^2 Benchmark Set

Motivation: Semi-supervised Data Collection



Example: Semi-supervised Data Collection

Methodology: Overall Structure

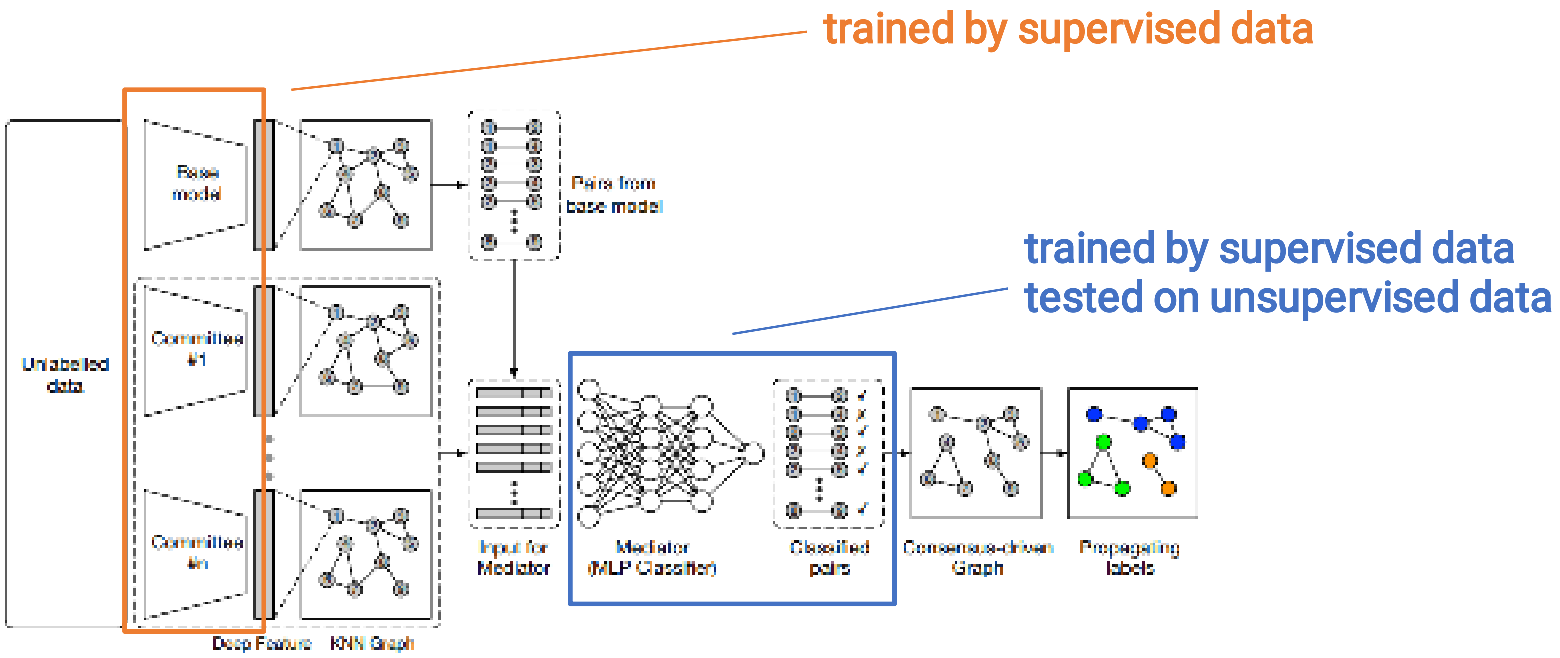
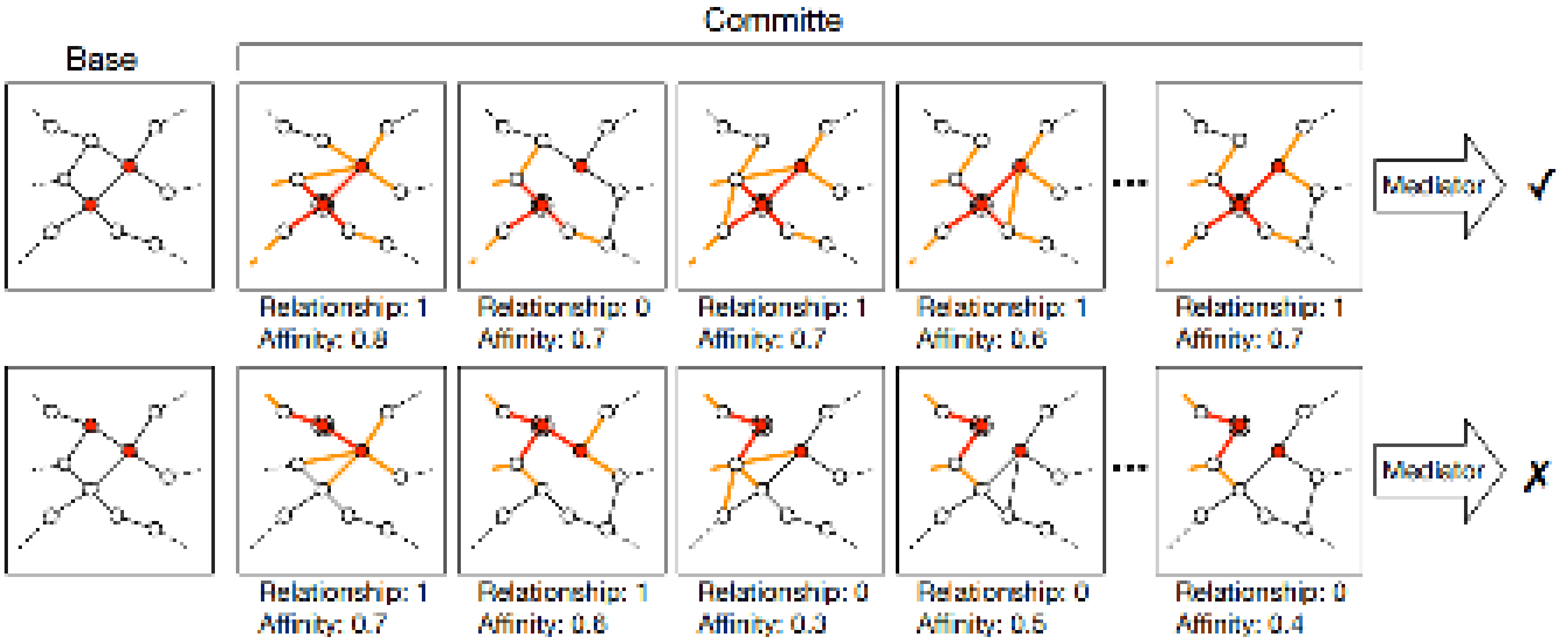


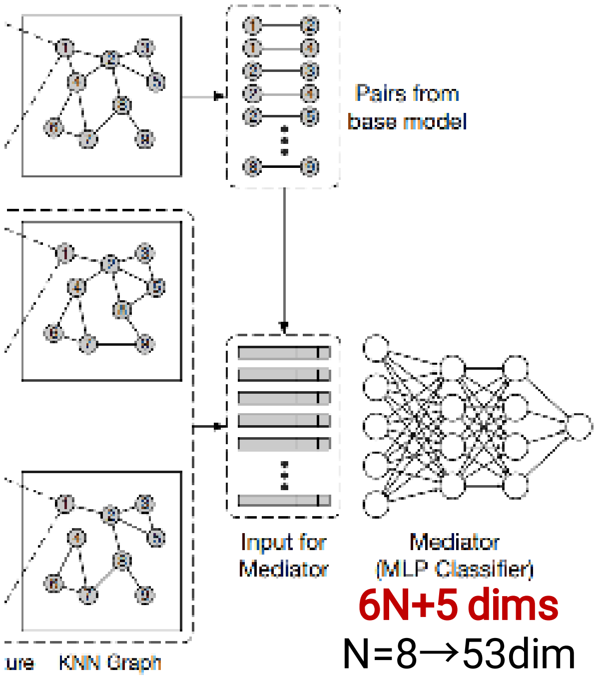
Fig. 1: **Consensus-Driven Propagation.** We use a base model and committee models to extract features from unlabeled data and create k-NN graphs. The input to the mediator is constructed by various local statistics of the k-NN graphs of the base model and committee. Pairs that are selected by the mediator compose the "consensus-driven graph". Finally, we propagate labels in the graph, and the propagation for each category ends by recursively eliminating low-confidence edges.

Methodology: k-NN Graph



Vertex: embedding
Edge: B is in the knn of A or not
Asymmetrical

Methodology: Input of Mediator



6N+5 dims

N=8→53dims

$$R_{C_i}^{(n_0,n_1)} = \begin{cases} 1 & \text{if } (n_0, n_1) \in \mathcal{E}\left(\mathcal{G}_{C_i}\right) \\ 0 & \text{otherwise.} \end{cases}, \quad i = 1, 2, \ldots, N,$$

$$A_{C_i}^{(n_0,n_1)} = \cos\left(\left(\mathcal{F}_{C_i}(n_0), \mathcal{F}_{C_i}(n_1)\right)\right), \quad i = 1, 2, \ldots, N.$$

$$D_{C_i}^{x} = \left\{\cos\left(\left(\mathcal{F}_{C_i}(x), \mathcal{F}_{C_i}(x_k)\right)\right), k = 1, 2, \ldots, K\right\}, \quad i = 1, 2, \ldots, N.$$

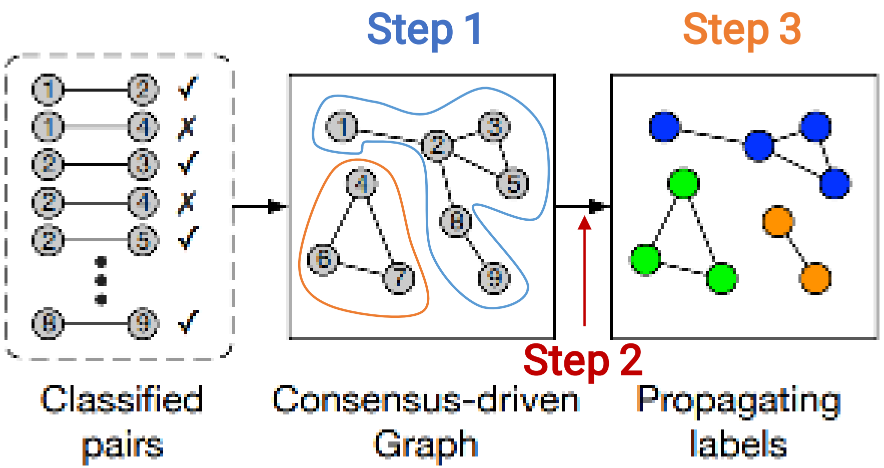
N dims $I_R \in \mathbb{R}^N$: $I_R = \left(\ldots R_{C_i}^{(n_0,n_1)} \ldots\right), i = 1, 2, \ldots, N$

N+1 dims $I_A \in \mathbb{R}^{N+1}$: $I_A = \left(\ldots A_{C_i}^{(n_0,n_1)} \ldots\right), i = 0, 1, 2, \ldots, N$

2N+2 dims $I_{D_{mean}} = \left(\ldots E\left(D_{C_i}^{n_0}\right) \ldots, \ \ldots E\left(D_{C_i}^{n_1}\right) \ldots\right), i = 0, 1, 2, \ldots, N,$

2N+2 dims $I_{D_{var}} = \left(\ldots \sigma\left(D_{C_i}^{n_0}\right) \ldots, \ \ldots \sigma\left(D_{C_i}^{n_1}\right) \ldots\right), i = 0, 1, 2, \ldots, N,$

Methodology: Label Propagation



Step 1: Find connected components based on the current edges in the graph and add it to a queue
Step 2: For each identified component, if its node number is larger than a pre-defined value, we eliminate low-score edges in the component, find connected components from it, and add the new disjoint components to the queue.
Step 3: If the node number of a component is below the pre-defined value, we annotate all nodes in the component with a new pseudo label.

Methodology: Joint Training

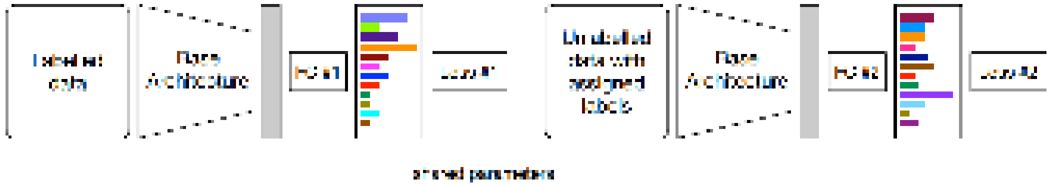
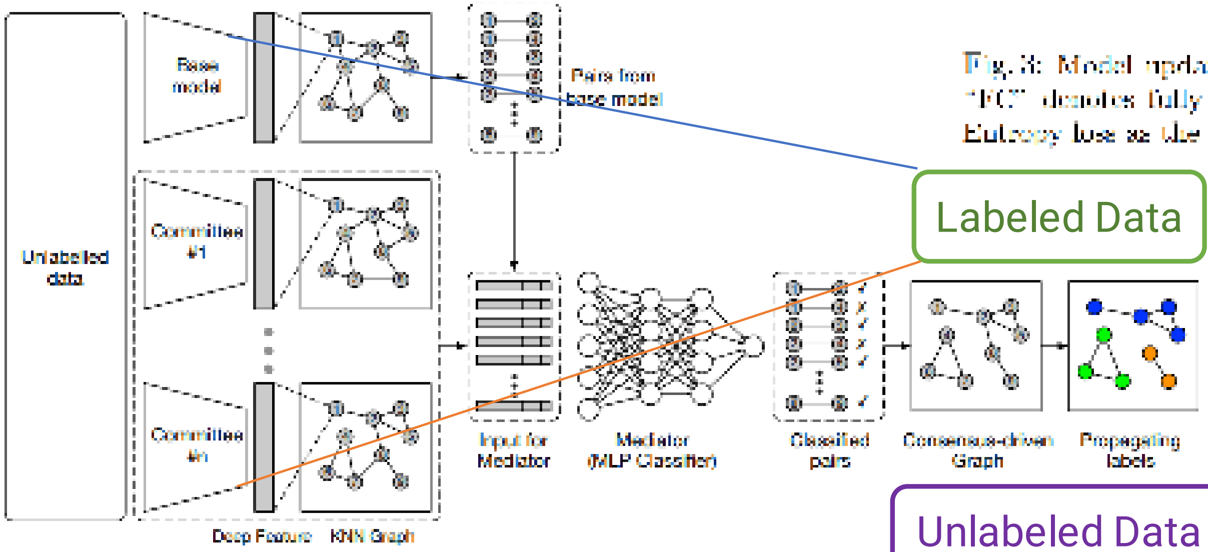


Fig. 3: Model updating in multi-task fashion. The weights of two CNNs are shared. "FC" denotes fully connected classifier. In our experiments we use weighted Cross-Entropy loss as the objective.



Labeled Data

$$\mathcal{L} = \lambda \sum_{x_l, y_l} \ell(x_l, y_l) + (1 - \lambda) \sum_{x_u, \tilde{y}_u} \ell(x_u, \tilde{y}_u)$$

Unlabeled Data

Fig. 1: **Consensus-Driven Propagation.** We use a base model and committee models to extract features from unlabeled data and create k-NN graphs. The input to the mediator is constructed by various local statistics of the k-NN graphs of the base model and committee. Pairs that are selected by the mediator compose the "consensus-driven graph". Finally, we propagate labels in the graph, and the propagation for each category ends by recursively eliminating low-confidence edges.

Experiments: Details

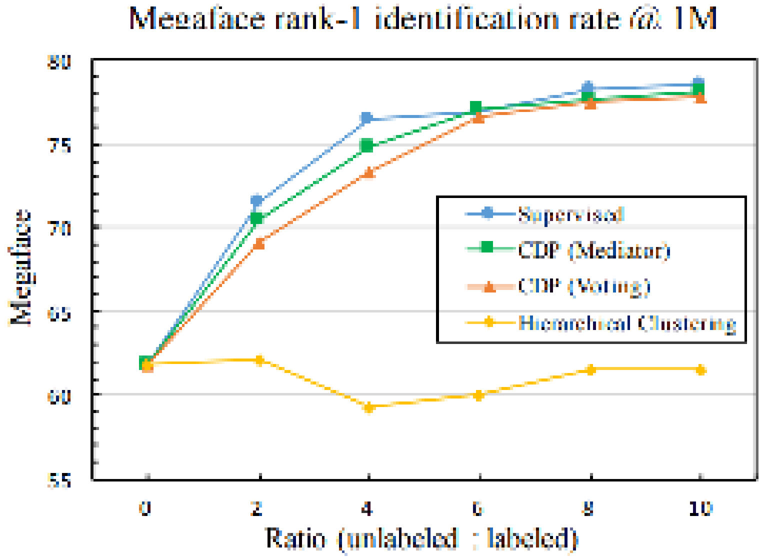
- Training Set: MS1M, **11 folds, 1 as labeled, 10 as unlabeled(1 of them is used to validate)**
- Testing Set: MegaFace(FaceScrub, Rank-1, 1e6), IJB-A(fpr=0.001)
- Committee Setting: ResNet-18, ResNet-34, ResNet-50, ResNet-101, DenseNet-121, VGG-16, Inception-V3, Inception-ResNetV2, **Tiny NASNet-A(base model)**
- Mediator Setting: MLP with 2 hidden layers, 50 nodes each, ReLU activation
- **similarity_threshold=0.96**

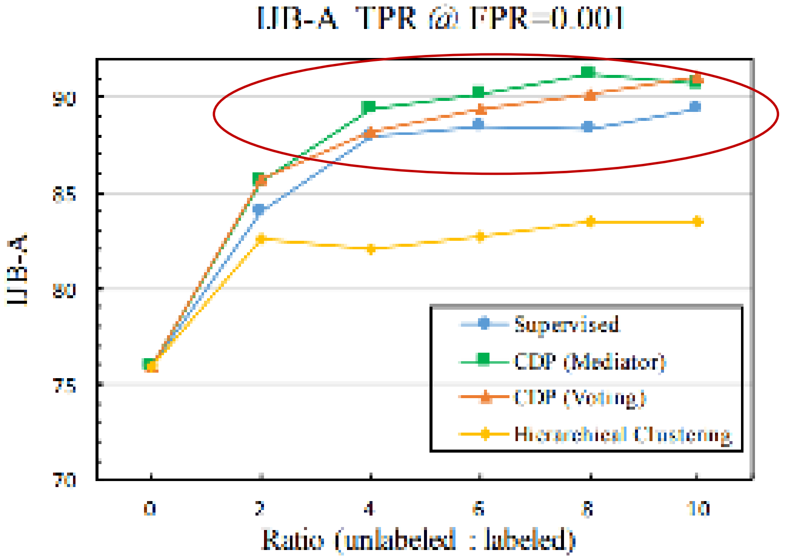
Table 1: Performance and the number of parameters of the base model and the committee members.

	Architecture	MegaFace	IJB-A	Parameters
Base	Tiny NASNet-A	**61.78**	**75.87**	$20.1M$
Committee	VGG16	50.22	70.75	$75.6M$
	ResNet18	51.48	69.23	$23.5M$
	ResNet34	52.44	72.52	$33.6M$
	Inception V3	52.82	75.53	$33.0M$
	ResNet50	56.16	73.21	$36.3M$
	ResNet101	57.87	74.52	$55.3M$
	Inception-ResNet V2	58.68	75.13	$66.1M$
	DesNet121	60.77	69.78	$28.9M$
Ensemble	(multiple)	69.86	76.97	-

Experiments: Fraction of Unlabeled Data

The proposed method is more robust to noise





Supervised: the real labels of unlabeled data is recovered
Mediator: the method proposed
Voting: a pair is selected if this pair is voted by all the committee members
Hierarchical Clustering: a naïve clustering method

Experiments: Ablation Study

Pairs generated by mediator   Pairs finally assigned

Methods	Committee number	Mediator inputs	Pair selection: pair number	Pair selection: recall	Pair selection: precision	Assigned labels: pairwise recall	Assigned labels: pairwise precision
Clustering	-	-	-	-	-	0.558	0.950
Voting	0	-	$1.4M$	0.313	0.966	0.680	0.829
	2	-	$1.4M$	0.313	0.986	0.783	0.849
	4	-	$1.4M$	0.313	**0.987**	0.791	0.862
	6	-	$1.4M$	0.313	0.984	0.801	**0.877**
	8	-	$1.4M$	0.313	0.979	**0.807**	0.876
Mediator	8	I_R	$1.4M$	0.318	0.975	0.825	0.822
		I_R+I_A	$2.5M$	**0.561**	0.982	**0.832**	0.888
		$I_R+I_A+I_D$	$2.4M$	0.527	**0.983**	0.825	**0.912**

Keep same to compare

Experiments: Backbone

Best Performance

Base	ResNet18		ResNet50		Tiny NASNet-A		Inception-ResNet V2	
	MegaFace	IJB-A	MegaFace	IJB-A	MegaFace	IJB-A	MegaFace	IJB-A
Lower Bound	51.48	69.23	56.16	73.12	61.78	75.87	58.68	75.13
CDP	72.75	86.23	75.66	88.34	78.18	90.64	81.88	92.07
Supervised	73.88	85.08	77.13	87.92	78.52	89.40	84.74	91.90

Experiments: Nums of Neighbors

k	Pair selection			Assigned labels	
	pair number	recall	precision	pairwise recall	pairwise precision
10	1.61M	**0.601**	**0.985**	0.810	**0.940**
20	2.54M	0.527	0.983	0.825	0.912
30	2.96M	0.507	0.982	0.834	0.886
40	**3.17M**	0.464	0.982	**0.837**	0.874

Need a trade-off

High k makes no impact
except more pairs

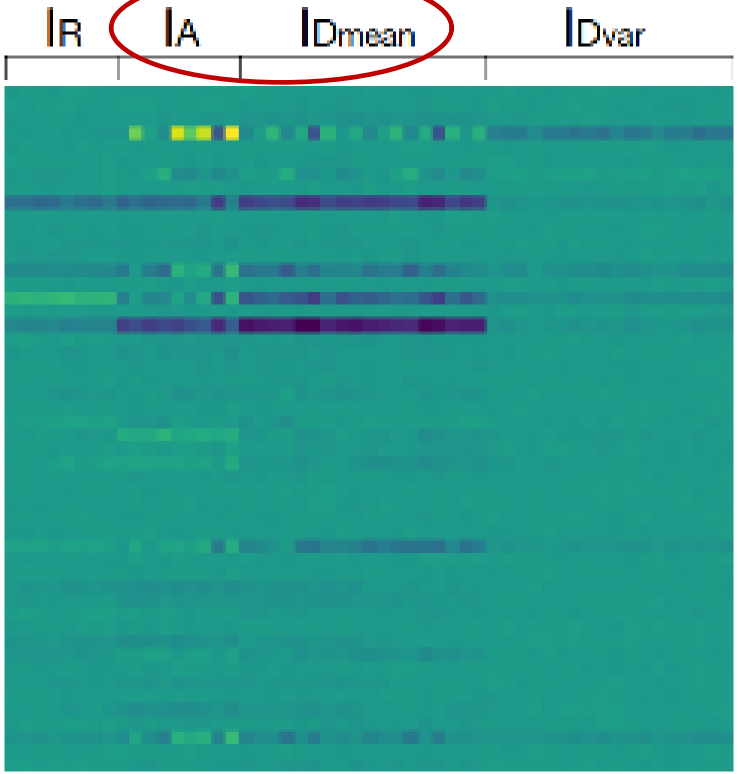
Experiments: Homogeneous/Heterogeneous

Committee	Methods	Pair selection			Assigned labels	
		pair number	recall	precision	pairwise recall	pairwise precision
Homogeneous	voting	1.93M	0.368	0.648	0.746	0.681
	mediator	2.46M	0.508	0.853	0.798	0.831
Heterogeneous	voting	1.41M	0.313	0.979	0.807	0.876
	mediator	2.54M	**0.527**	**0.983**	**0.825**	**0.912**

Heterogeneous is better

Experiments: Importance of inputs of mediator

Contribute more



Green≈0
Blue<0
Yellow>0

Experiments: Import ArcFace

- m=0.5
- Output setting E(BN-Dropout-FC-BN)
- Cleaner training set for higher baseline
- Test on MegaFace

	Softmax	ArcFace [7]
baseline	61.78%	76.93%
CDP (Ratio = 2)	70.51%	83.08%

Experiments: IJB-A Example

Wrong Annotated

Low Quality

Cartoon

Discussion

- Contribution
  - Hard pairs mining: pipeline(Filtering) and benchmark(Precision-Recall)
  - Committee mechanism: heterogeneous is better
  - Overlapped: multi-task learning
- Further Direction
  - Committee: supervised feature extractor to unsupervised methods?
  - K-NN Graph: Graph Convolution Network[3]

References

- [1] X. Zhan, Z. Liu, J. Yan, D. Lin, and C. Change Loy. Consensus-driven propagation in massive unlabeled data for face recognition. In ECCV 2018.
- [2] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The MegaFace Benchmark: 1 Million Faces for Recognition at Scale. In CVPR 2016.
- [3] Zhongdao Wang, Liang Zheng, Yali Li, Shengjin Wang. Linkage Based Face Clustering via Graph Convolution Network. CVPR 2019.

Thank you for listening

Hongwei Fan

Lab of Pattern Recognition and Intelligent Systems, BUPT