Feature Transfer Learning for Face Recognition with Under-Represented Data

Xi Yin[1*], Xiang Yu[2], Kihyuk Sohn[2], Xiaoming Liu[1], and Manmohan Chandraker[2,3]

[1]Department of Computer Science and Engineering, Michigan State University

[2]Department of Media Analytics, NEC Laboratories America, Inc.

[3]Department of Computer Science and Engineering, University of California, San Diego

NEC Laboratories America

Relentless passion for innovation

Highlights

Problem Statement:

Real world face recognition benchmarks are imbalanced, where a large portion of subjects are with insufficient samples and thus are under-represented (*UR*), which leads to a biased classifier and an inferior feature representation during the training process. We propose a Feature Transfer Learning (FTL) algorithm to solve this problem.

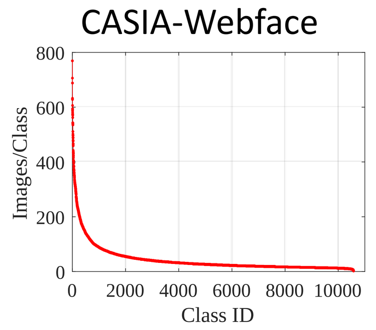
Insights and Contributions:

- ✧ Center-based feature transfer: enrich *UR* class distribution, disentanglement.
- ✧ m-L_2 regularization: simple yet effective, applicable to other frameworks.
- ✧ Two-stage alternative training: correct classifier bias, learn discriminative features.
- ✧ Empirical analysis: ablation study, general / one-shot recognition, visualization.

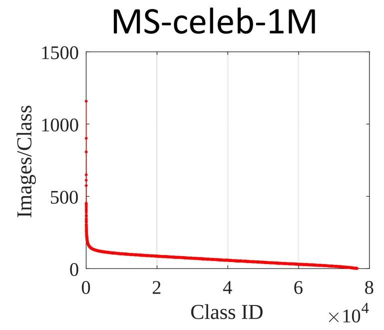
Motivation

(1) Imbalanced Distribution

CASIA-Webface



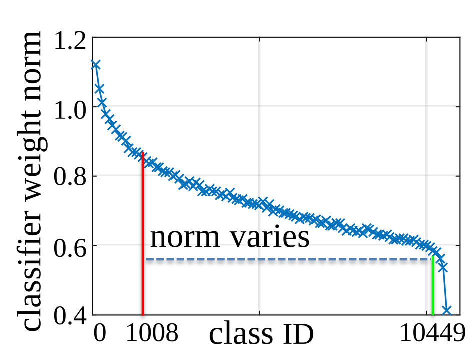
MS-celeb-1M

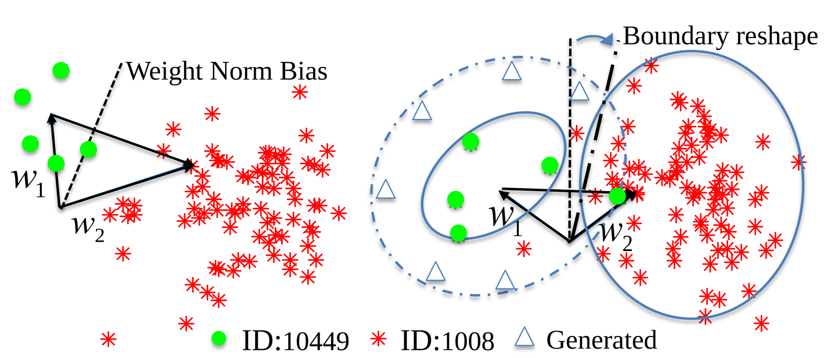


	CASIA-Webface		MS-celeb-1M	
	#Classes	#Images	#Classes	#Images
Full	10.6K	455.6K	76.5K	4,753.3K
Regular	6.5K	393.1K	67.0K	4,638.4K
UR	4.0K	62.5K	9.5K	114.9K

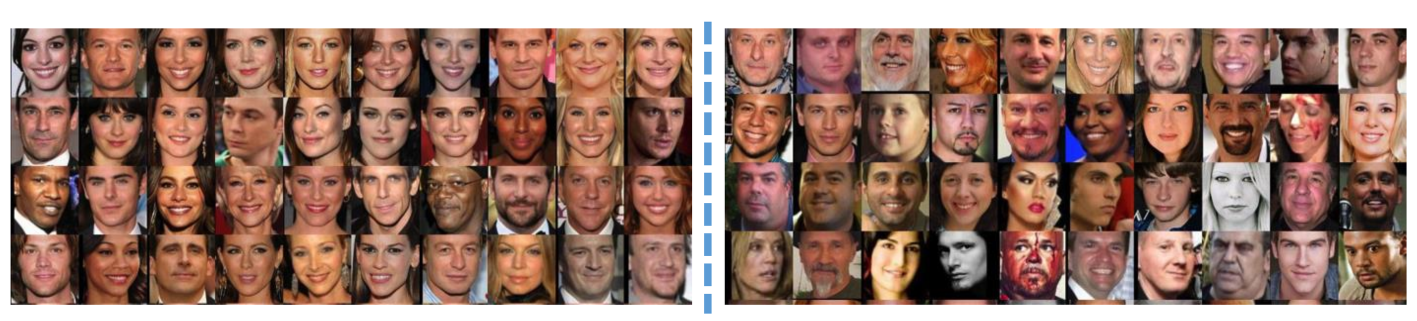
(2) Limitations of Training with *UR* Classes

weight norm bias and intra-class variance bias:





(3) Face Model

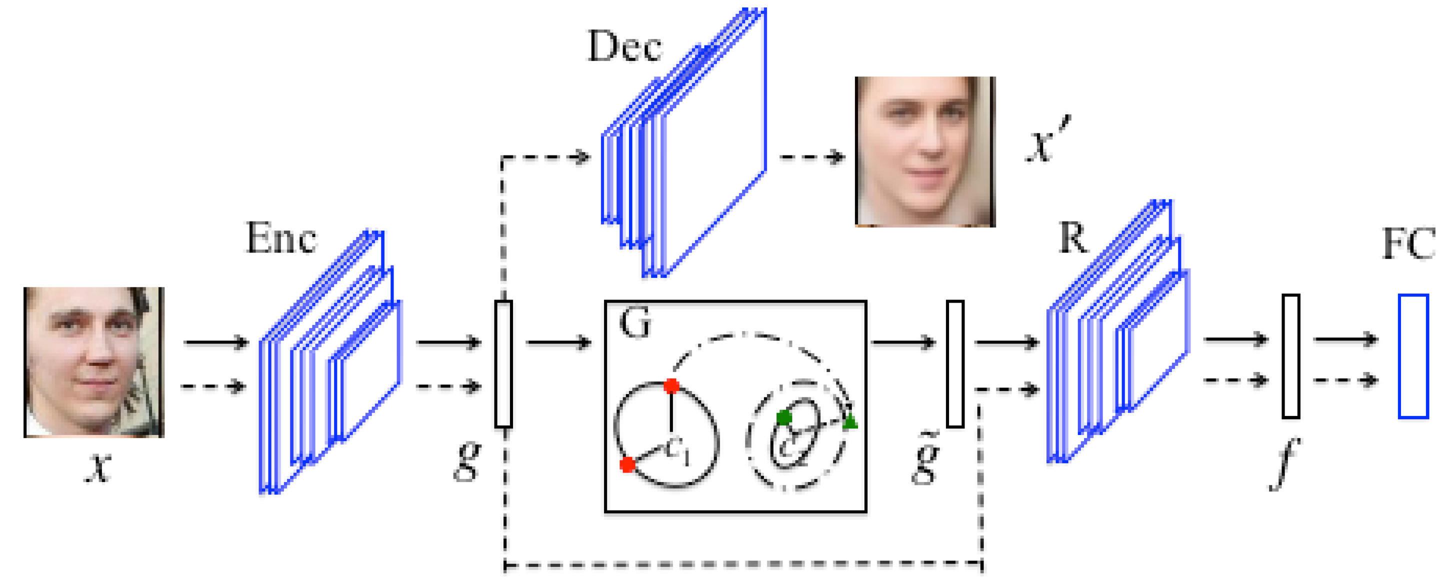


most images least images

$$c_i = \frac{1}{2|\Omega_i|}\sum_{k\in\Omega_i}(g_{ik}+\bar{g}_{ik}), \quad \Omega_i = \{j \mid \|p_{ik} - \bar{p}_{ik}\|_2 < t\}$$

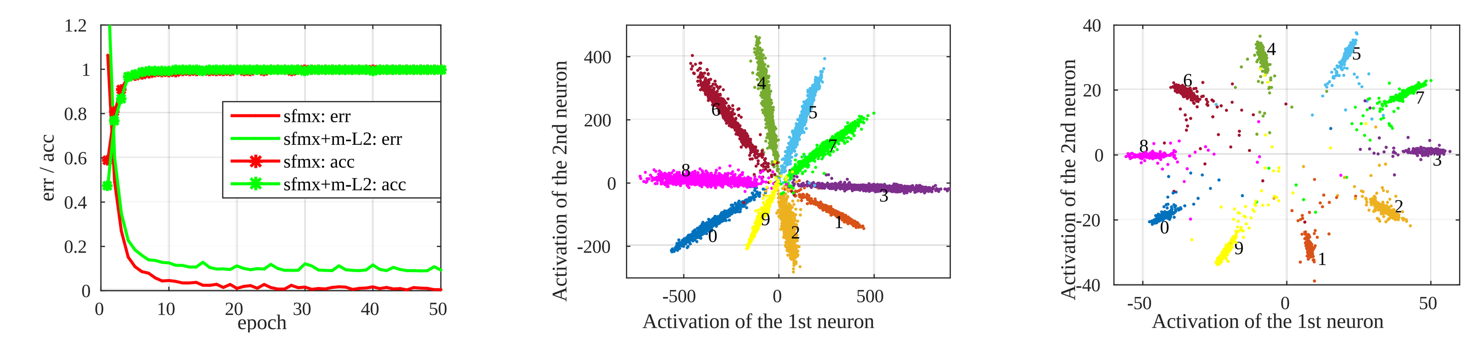
Linear face model: $g_i = c_i + e$

Framework



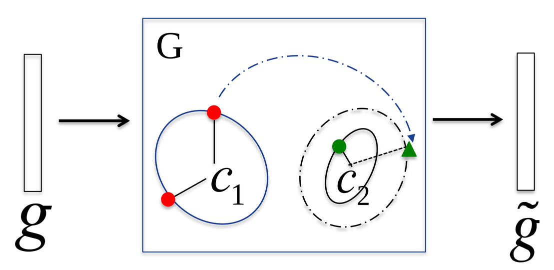
$$L_{recon} = \|x' - x\|_2^2 \quad L_{sfmx} = -\log\frac{\exp(w_i^T f_i)}{\sum_j \exp(w_j^T f_i)} \quad L_{reg} = \|W^T f\|_2^2$$

m-L_2 regularization on MNIST:



Alternative Transfer Learning

(1) Feature Transfer



$$\tilde{g}_{ik} = \boxed{c_i} + QQ^T\boxed{(g_{jk} - c_j)}$$

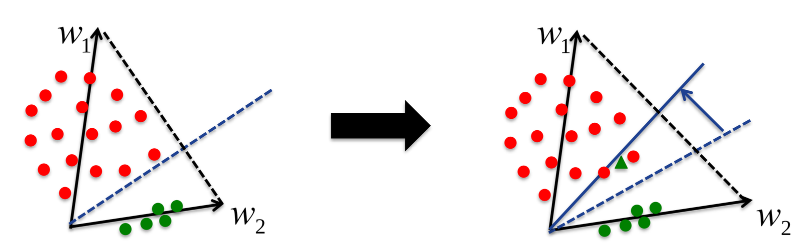
UR center intra-class variance

(2) Alternative Training

	Freeze	Training
Stage 1	Enc, Dec	R, FC
Stage 2	FC	Enc, Dec, R

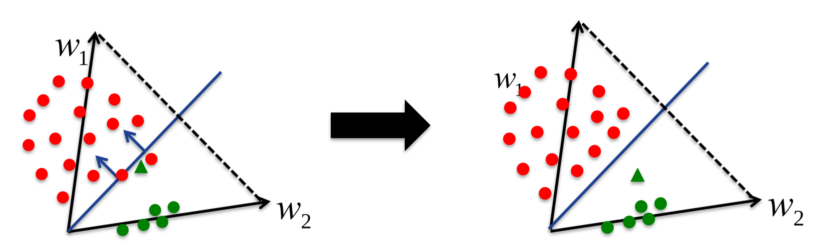
Stage 1: ⟶

Decision boundary reshape



Stage 2: - - →

Compact feature learning



Experimental Results

(1) Ablation Study

Train	Method	LFW g	LFW f	IJB-A: Verification FAR .01	IJB-A: Verification FAR .001	IJB-A: Identification Rank-1	IJB-A: Identification Rank-5	MS1M:NN Reg.	MS1M:NN *UR*
10K0K	sfmx	97.2	97.5	69.4	33.0	81.6	90.4	87.2	82.5
	sfmx+m-L_2	97.0	97.9	73.0	44.8	83.8	91.5	90.2	84.7
10K10K	sfmx+m-L_2	97.1	97.9	74.1	46.3	83.7	91.7	89.5	84.1
	FTL	96.7	98.3	80.3	55.0	85.9	92.8	92.3	88.2
10K30K	sfmx+m-L_2	97.1	98.1	76.9	47.2	84.8	91.9	90.6	86.4
	FTL	96.9	98.4	81.8	61.0	86.1	92.6	91.8	88.7
10K50K	sfmx+m-L_2	97.3	98.1	78.5	53.4	85.0	92.2	90.2	87.1
	FTL	97.0	98.5	82.6	62.6	86.5	93.1	92.1	89.4
60K0K	sfmx	97.5	98.3	82.8	62.3	87.1	93.8	90.4	89.6
	sfmx+m-L_2	**97.9**	**98.9**	**86.4**	**74.4**	**89.3**	**94.7**	**93.7**	**93.5**

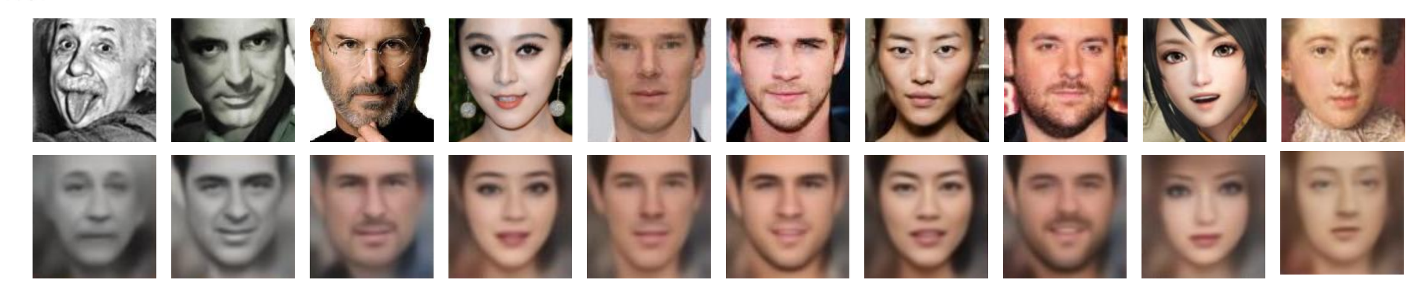
(2) Large-Scale Face Recognition

Method	LFW
VGGFace	98.95
CenterLoss	99.28
SphereFace	99.42
RangeLoss	**99.53**
FaceNet	**99.63**
sfmx	98.60
sfmx+L_2	98.53
sfmx+m-L_2 (Ours)	99.18
FTL (Ours)	99.37

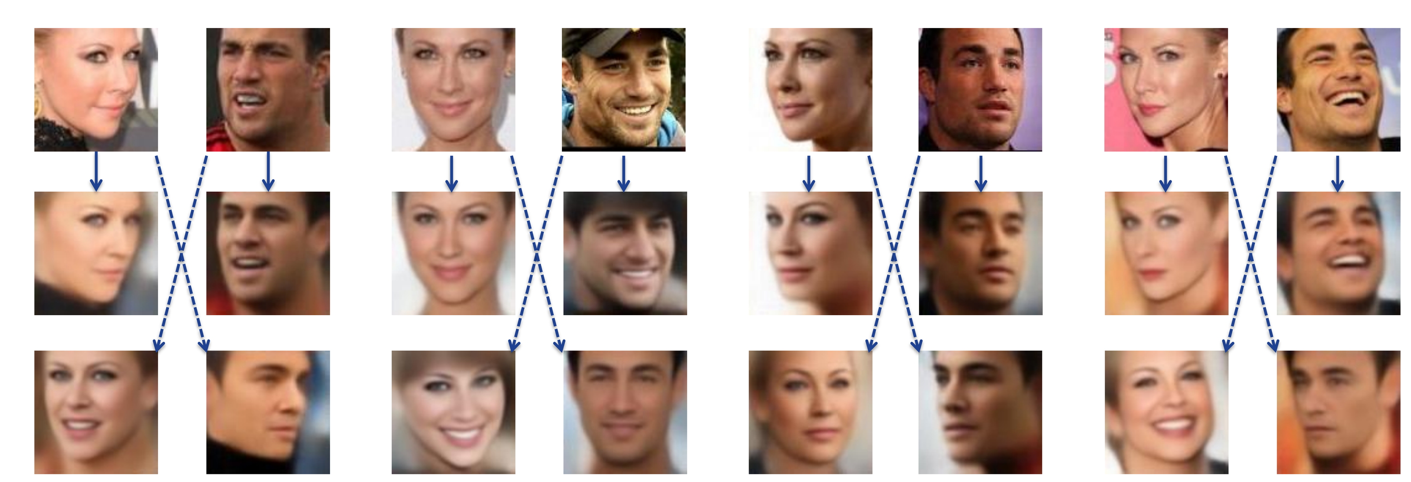
Method	IJBA: Verification FAR .01	IJBA: Verification FAR .001	IJBA: Identification Rank-1	IJBA: Identification Rank-5	IJBA: Identification Rank-10
TPE	90.0	81.3	86.3	93.2	97.7
TA	93.9	-	92.8	-	98.6
NAN	94.1	88.1	95.8	98.0	98.6
sfmx	91.5	77.4	92.4	96.4	97.3
Sfmx+m-L_2 (Ours)	92.5	80.2	93.9	97.2	97.9
FTL (Ours)	93.5	82.9	94.8	97.8	98.3
FTL+MP (Ours)	94.3	85.1	95.1	97.8	98.4
FTL+MP+TA (Ours)	**95.3**	**91.2**	**96.0**	**98.3**	**98.7**

(3) Feature Visualization

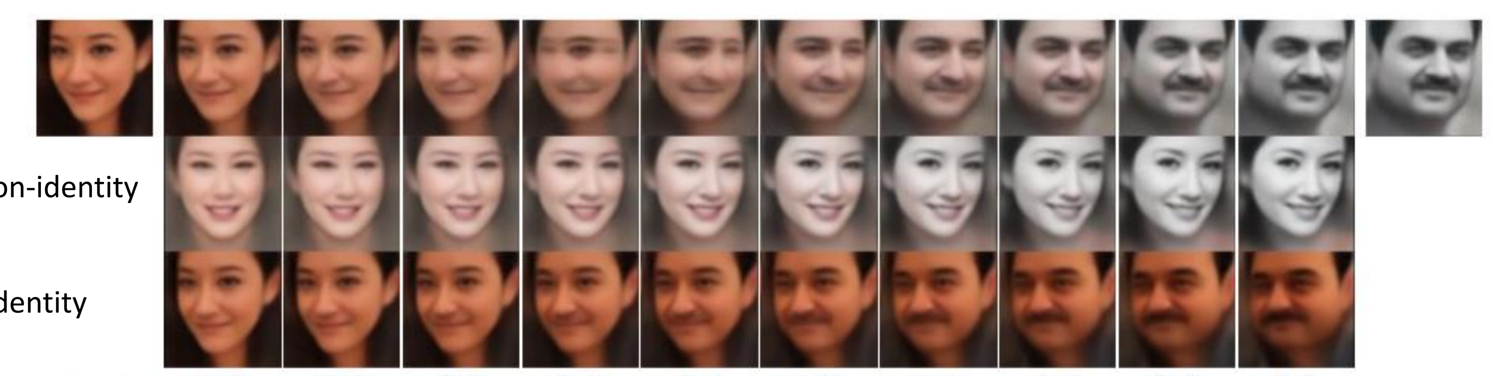
Feature Center



Feature Transfer



Feature Interpolation



• Main part of this work was done when Xi was an intern at NEC Laboratories America. Xi is now an Applied Scientist at Microsoft Cloud and AI. More about this work can be found: http://cvlab.cse.msu.edu/project-feature-transfer.html