Uncertainty and Bias
in Face Recognition and Expression Analysis

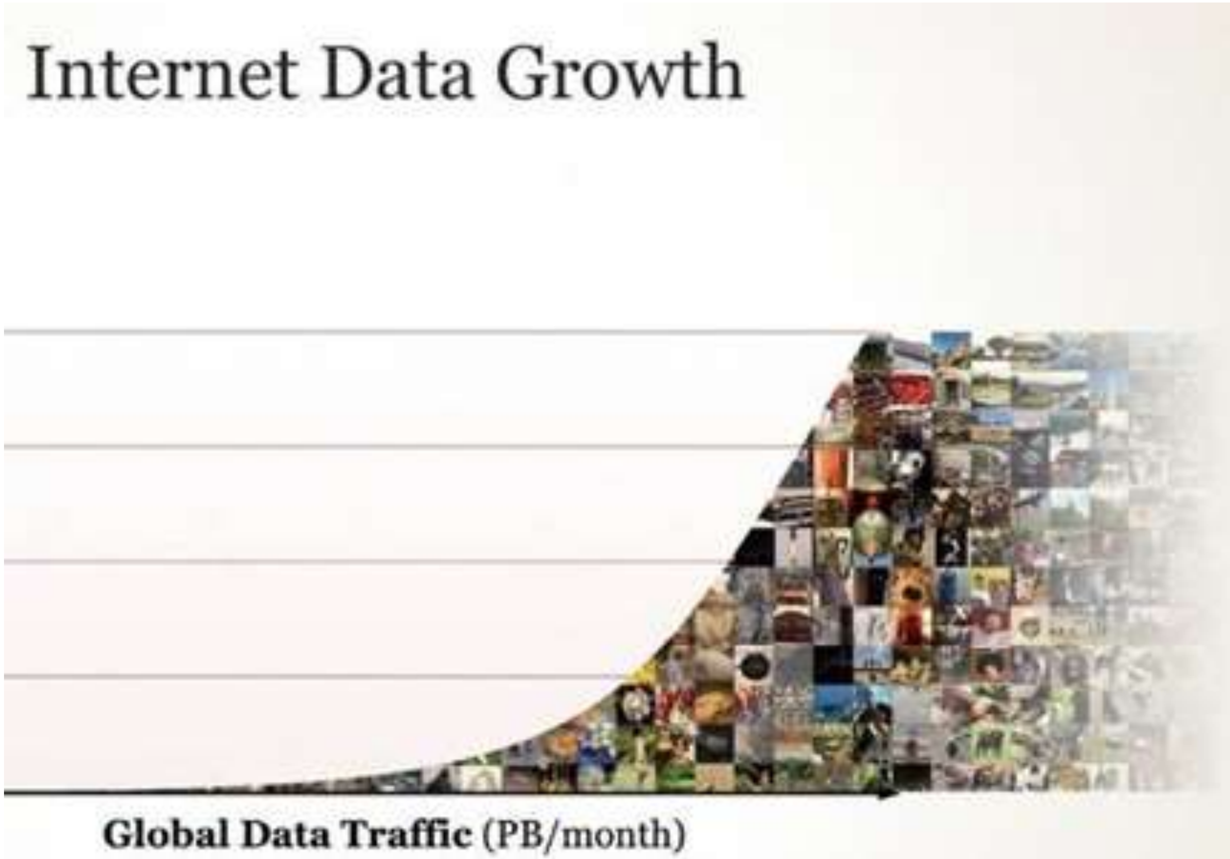
Weihong Deng

Beijing Univ. Post. & Telecom.

Computer vision: revolutionary technology

deliver sight to the blind

alarm us when a child is drowning



manage the massive photos and videos

••• •••

Courtesy of Prof. Fei-Fei Li

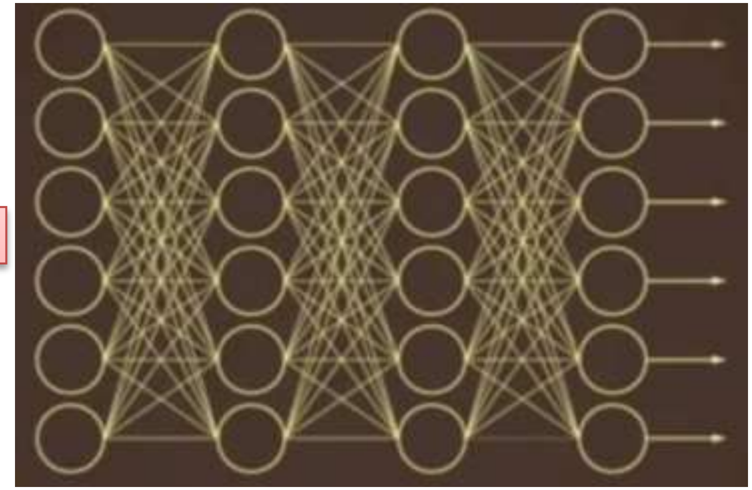
Visual recognition

naming objects

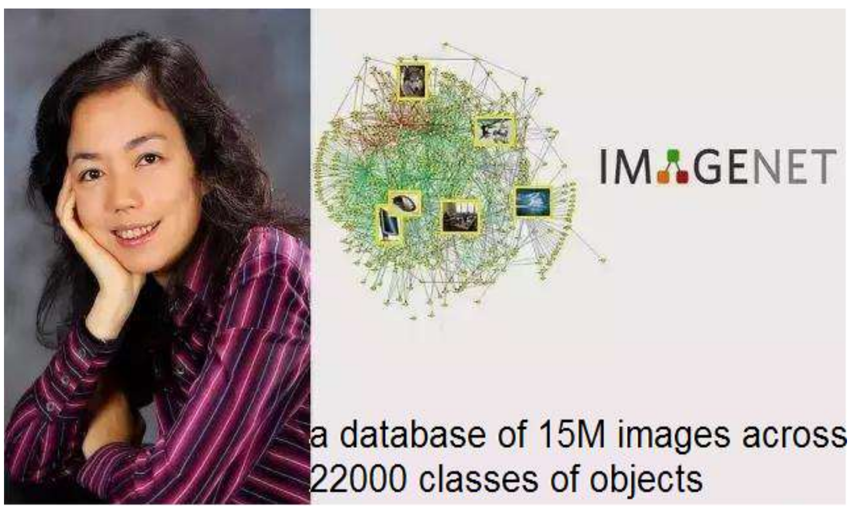
identifying people

understanding human emotions and behaviors

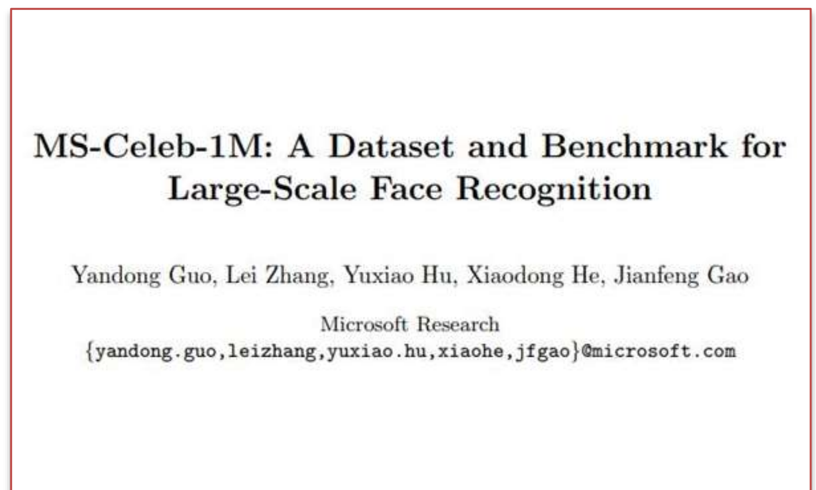
Visual recognition



Object



Face



MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition

Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, Jianfeng Gao

Microsoft Research
{yandong.guo,leizhang,yuxiao.hu,xiaohe,jfgao}@microsoft.com

Courtesy of Prof. Fei-Fei Li

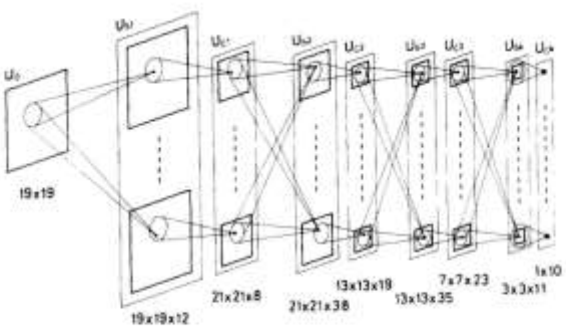
History of CNN

Kunihiko Fukushima

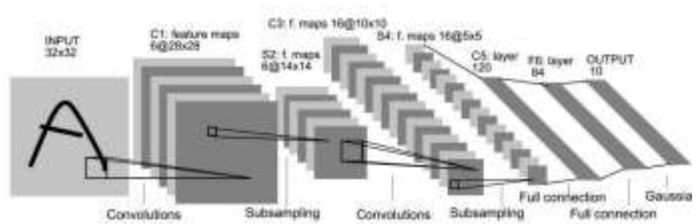
Yann LeCun

Geoff Hinton

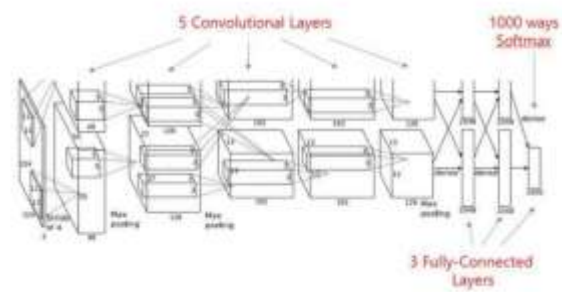
Neocognitron



LeNet



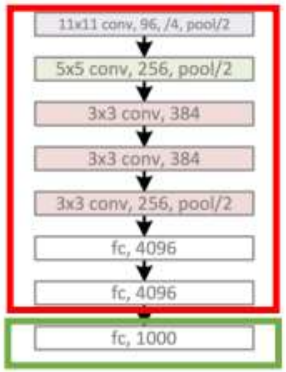
AlexNet



K Fukushima, Biological cybernetics, 1980 **Y LeCun, et al, Proceedings of the IEEE, 1998** **A Krizhevsky, I Sutskever, GE Hinton, NIPS 2012**

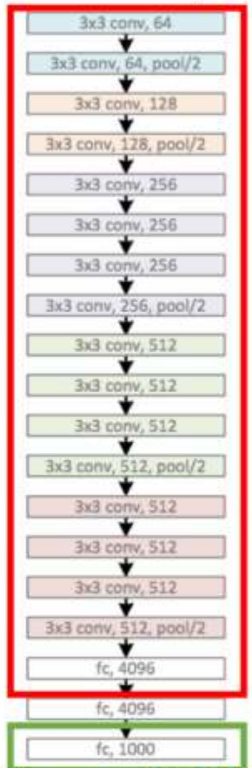
Blossom of CNN

2012: AlexNet
5 conv. layers



Error: 15.3%

2014: VGG
16 conv. layers

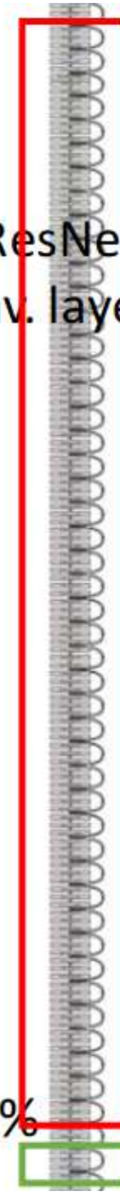


Error: 8.5%

2015: GoogLeNet
22 conv. layers

Error: 7.8%

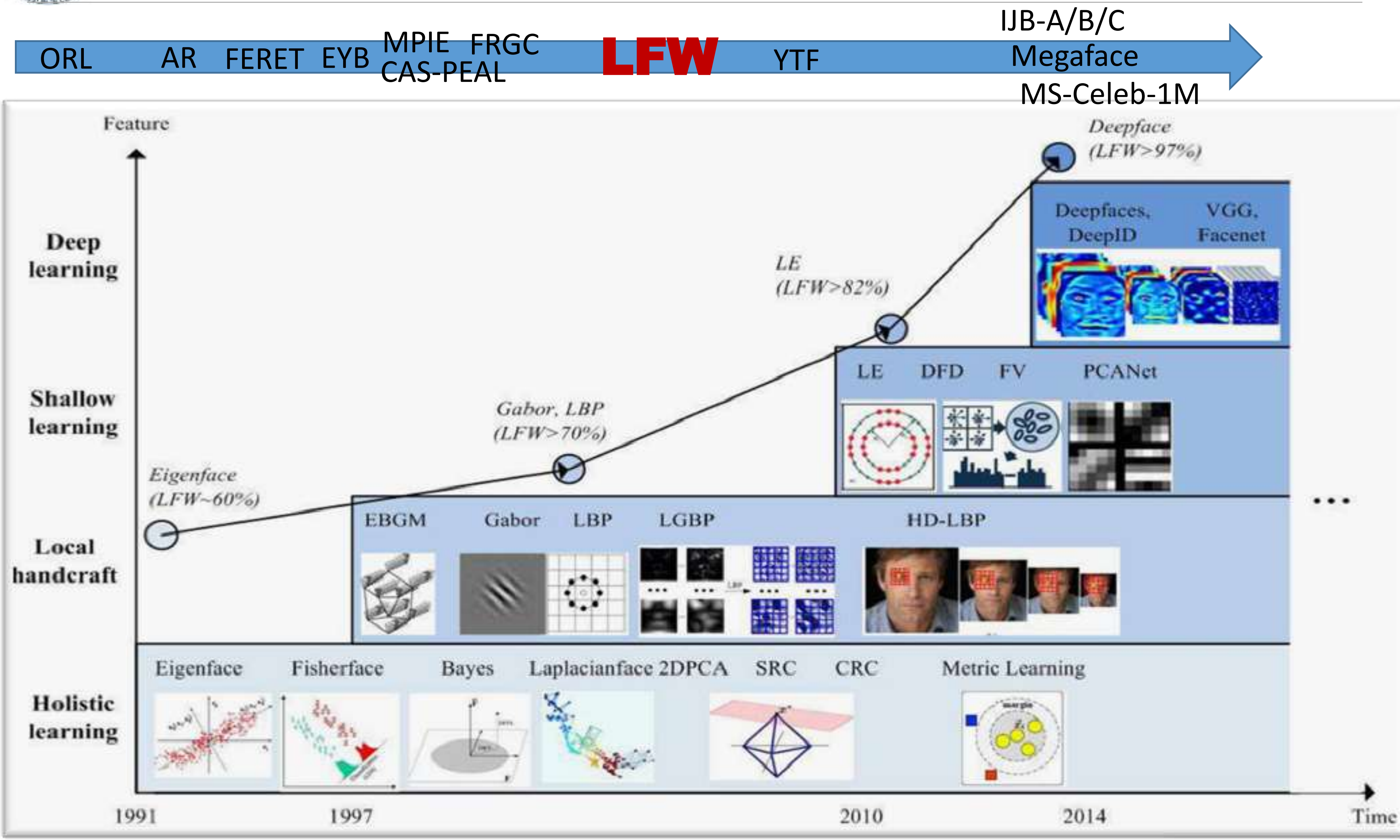
2016: ResNet
>100 conv. layers



Error: 4.4%

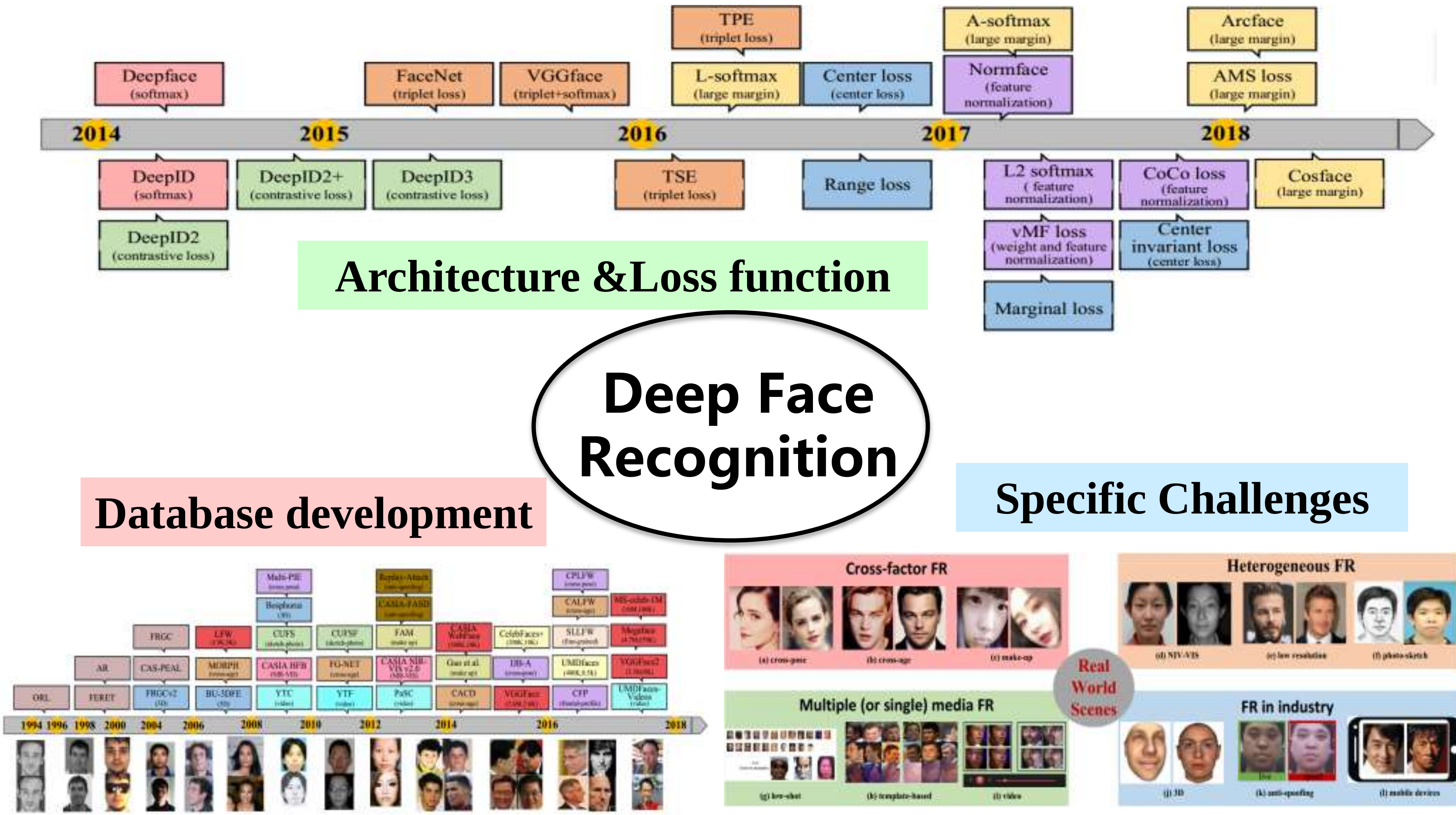
Courtesy of Dr. Bolei Zhou

Deep Face Recognition: A Survey



Mei Wang & Weihong Deng, Deep Face Recognition: A Survey, arXiv:1804.06655

Deep Face Recognition: A Survey



Same or Different People?

Angelababy

Angelababy

DCNN correct， Students wrong

The first 4 image pairs are from Similar-Looking LFW database

Weihong Deng, et al., Pattern Recognition , 2017

Why visual recognition is so hard?

Data uncertainty

- Poses viewpoints, expression, modality
- Make up, occlusion
- Intra-class >> inter-class
- Adversarial attack

Facial images of Lady Gaga

Increase uncertainty on LFW

Focus on face uncertainty: Deliberately selected pairs

Negative pairs

Similar-looking

Positive pairs

Cross-Aging

Cross-Poses

From LFW to SL/CA/CPLFW

Identical celebrities, scale, and protocols

Similar-looking

Aging

Poses

Similar-Looking

3K negative pairs

Similar-look face pairs selected by crowd-sourcing

Weihong Deng et al. Pattern Recognition 2017

Cross-Age

3K positive pairs

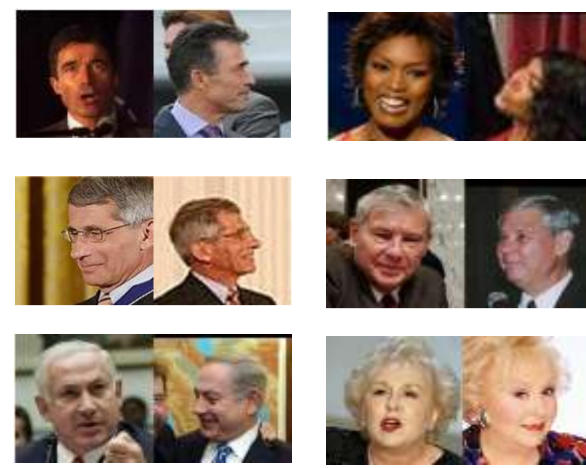
Cross-age face pairs selected by crowd-sourcing

Tianyue Zheng, Weihong Deng, arXiv:1708.08197

Cross-Pose

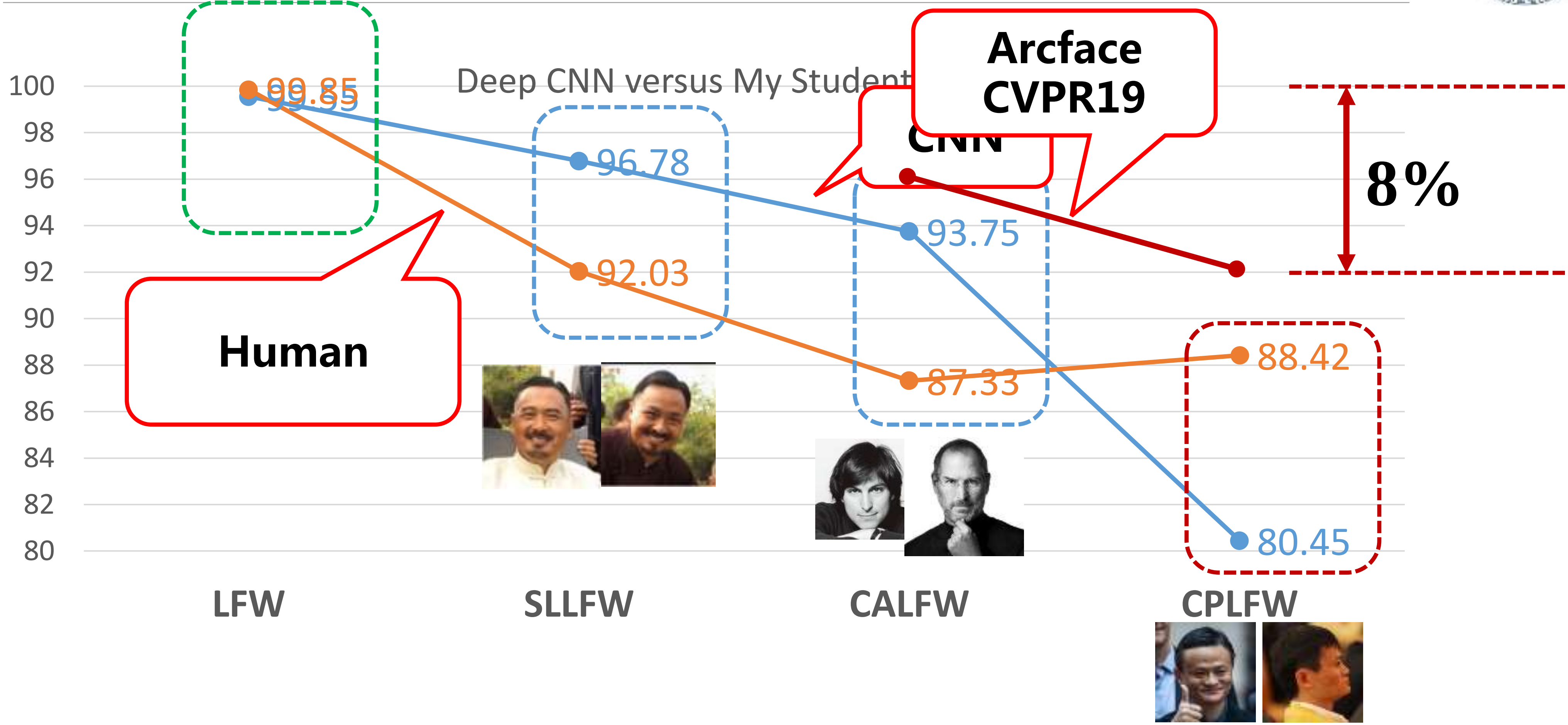
3K positive pairs

Cross-pose face pairs selected by crowd-sourcing



Tianyue Zheng, Weihong Deng, BUPT-TR 2018

Human-Machine Comparison

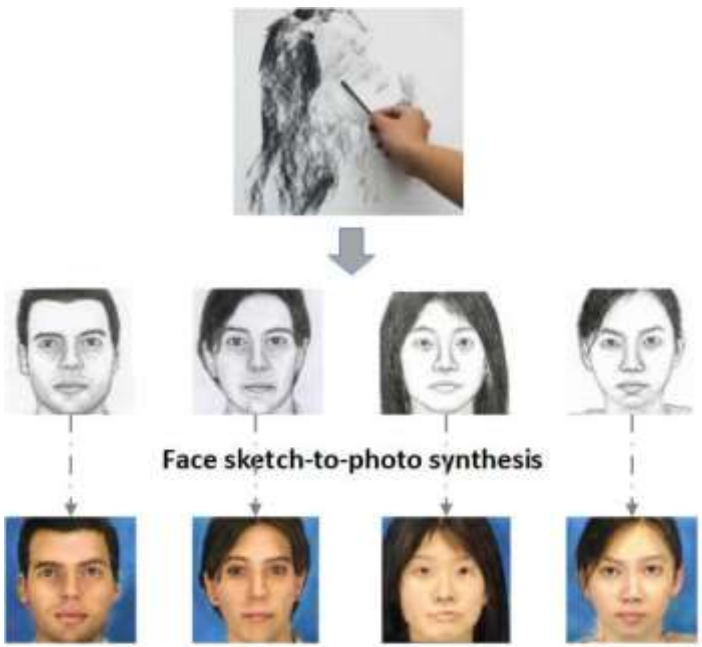


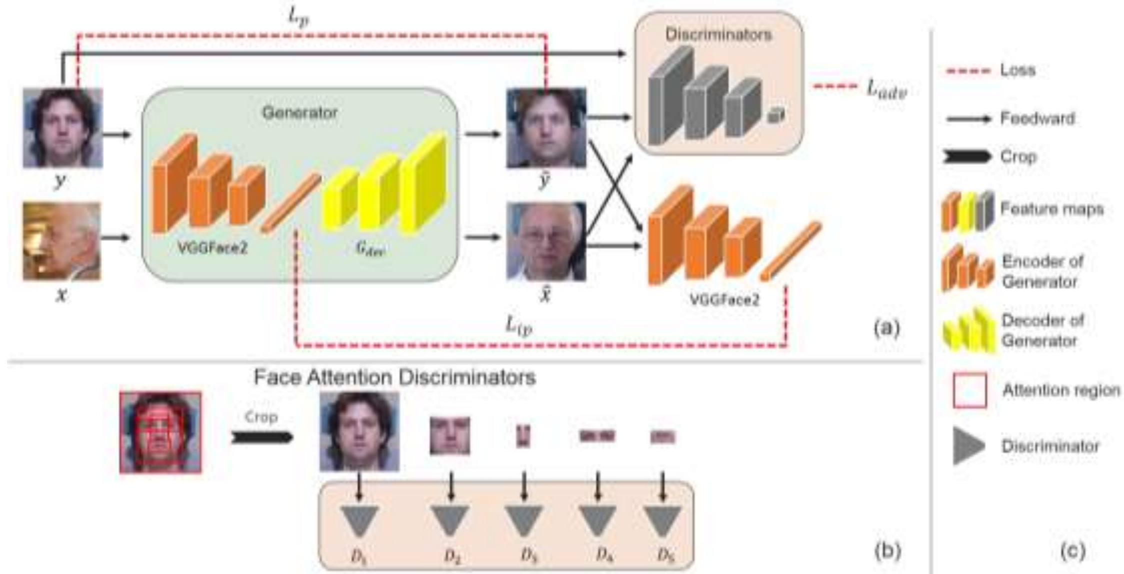
CNN ~ Human
~ 100%

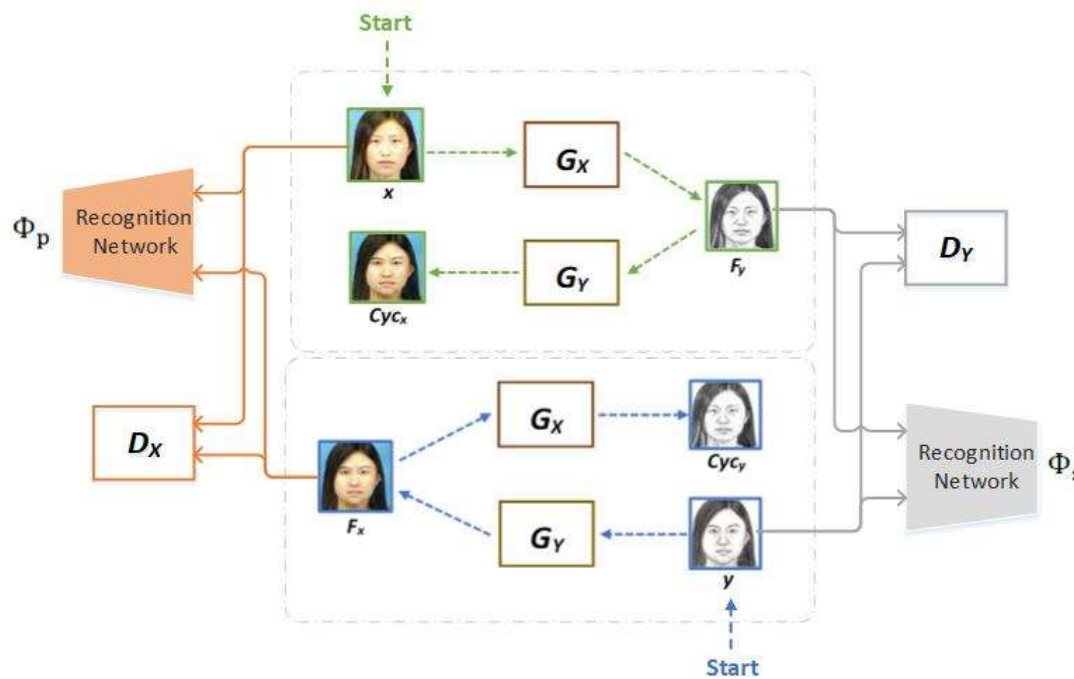
CNN >> Human

CNN > Human

Face Transfer to reduce variability







Face Normalization Results

Face normalization of the same person

Yichen Qian, Weihong Deng, et al., CVPR 2019.

Face Normalization Results

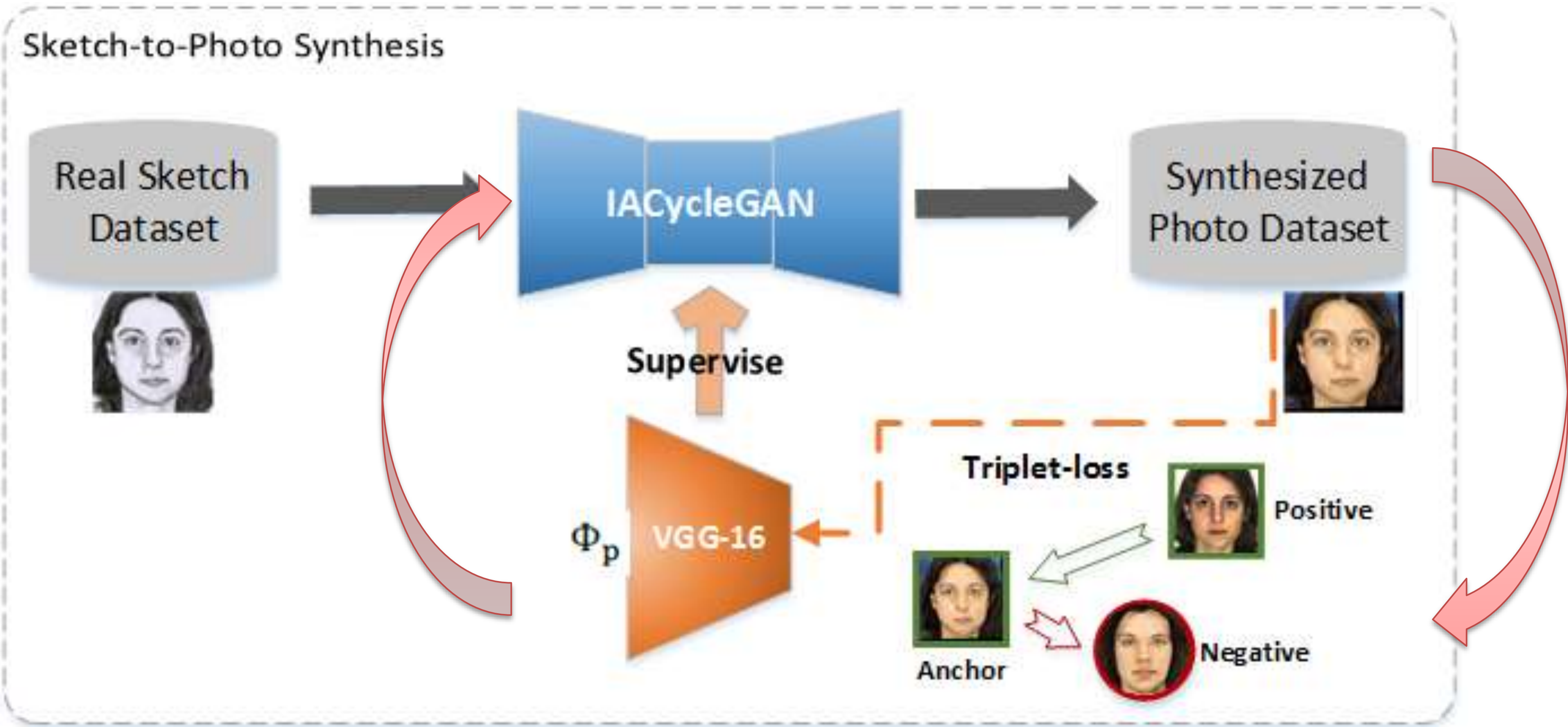
Face normalization results on IJB-A.

Yichen Qian, Weihong Deng, et al., CVPR 2019.

Face Sketch→Photo Results

Sketch→ Photo Synthesis on CUFS.

Identity-Aware CycleGAN (IACycleGAN)



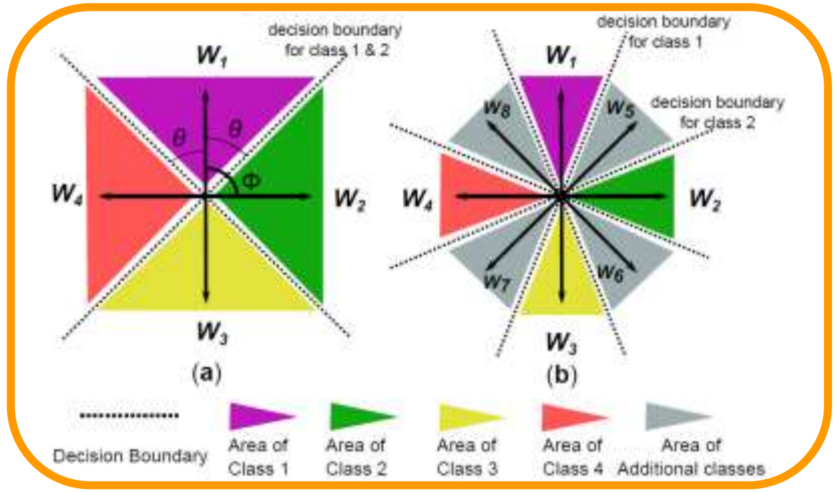
Virtual Softmax and Normalized metric learning

Virtual Softmax:

--Inject ***dynamic virtual negative*** class

$$L_i = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T X_i}}{\sum_{j=1}^{C} e^{W_j^T X_i} + e^{W_{virt}^T X_i}}$$

where $W_{virt} = \frac{\|W_{y_i}\| X_i}{\|X_i\|}$



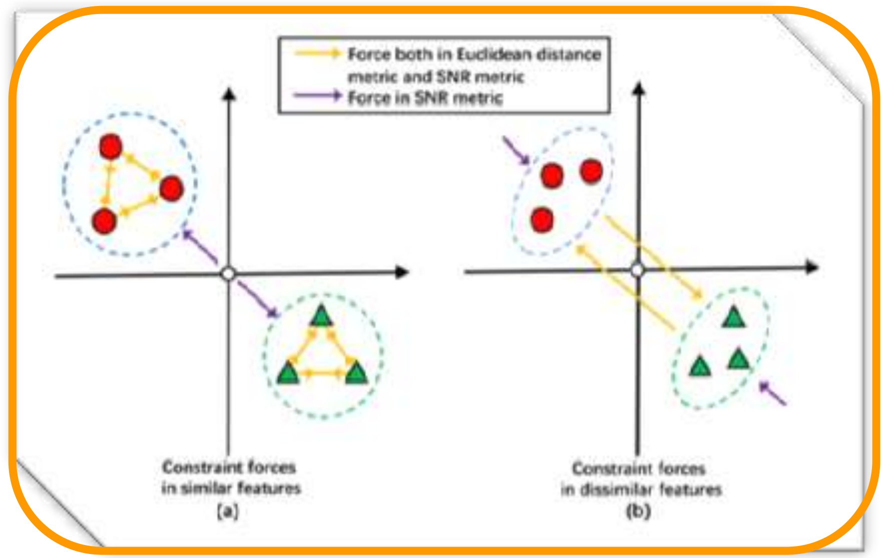
Binghui Chen and Weihong Deng, NIPS 2018

Signal-to-Noise Ratio Metric Learning:

Given two images x_i and x_j ,the learned features can be denoted as $h_i = f(:, x_i)$ and $h_j = f(:, x_j)$, we define the SNR between the anchor feature h_i and the compared feature h_j as:

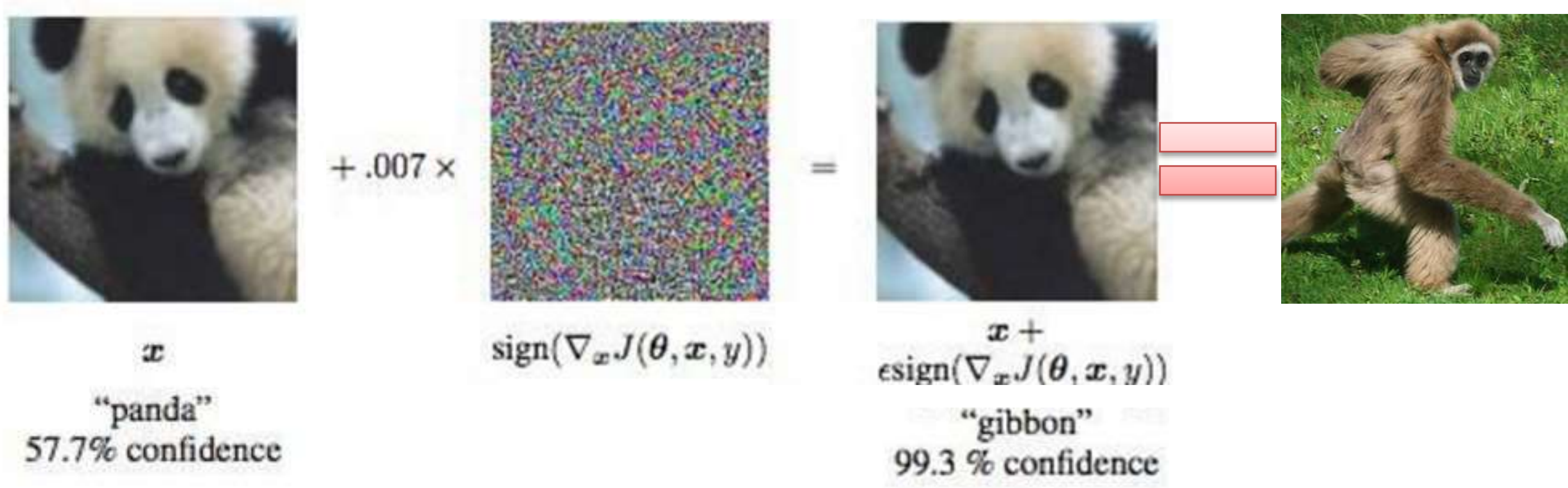
$$SNR_{ij} = \frac{var(h_i)}{var(h_j - h_i)}.$$

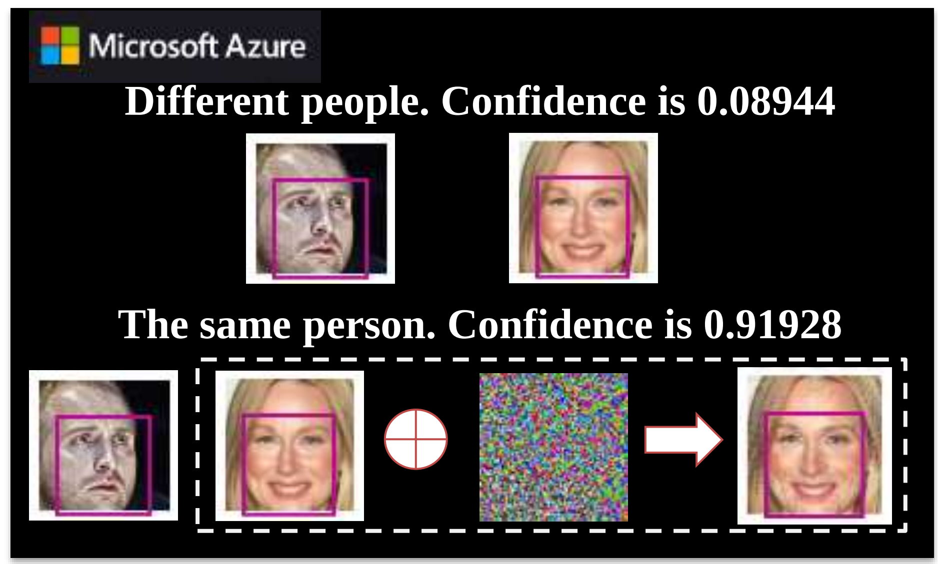
SNR distance is $d_S(h_i, h_j) = \frac{1}{SNR_{ij}} = \frac{var(h_j - h_i)}{var(h_i)}$



TongTong Yuan, Weihong Deng et al., CVPR 2019

Man-made Adversarial Uncertainty





Uncertainty in Visual recognition


Auto-driving Uber hit and killed a pedestrian

Autopolit Tesla crashed into a fire truck

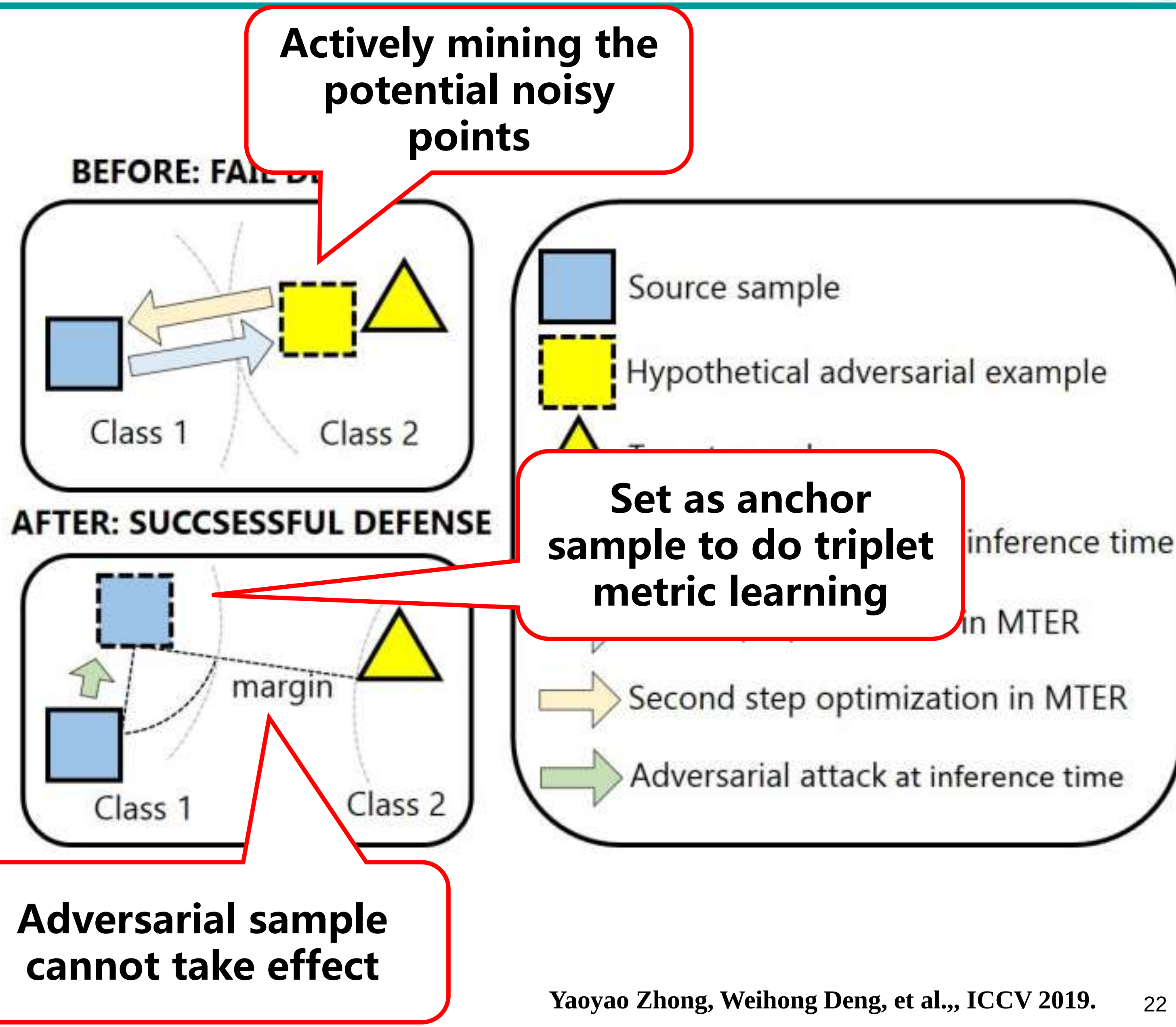

Autopolit Tesla crashed into a truck


Autopolit Tesla crashed to roadside fence

Adversarial Training Framework

First step: seek the potential adversarial examples by gradient vulnerability exploitation

Second step: conduct triplet metric learning based on the anchors of potential adversarial samples.



Yaoyao Zhong, Weihong Deng, et al.,, ICCV 2019.

Adversarial Learning MTER- Experiment

The experimental results on MNIST, CASIA-WebFace, VGGFace2 and MS-Celeb-1M reveal that our method increases the robustness of the network against adversarial attacks in simple object classification and deep face recognition.

Figure 1. Embedding space visualization of MNIST trained with Softmax and Softmax+MTER.

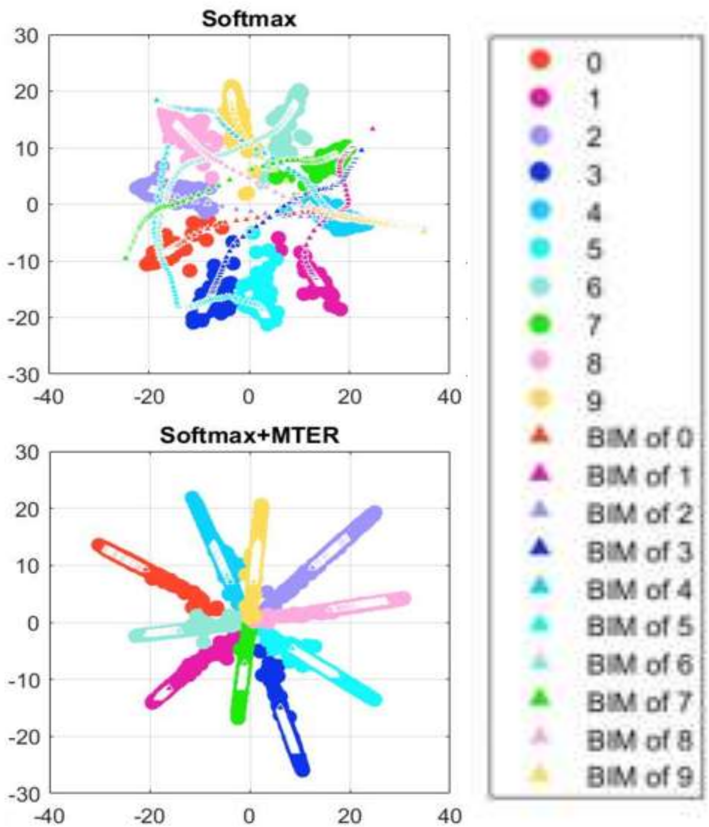
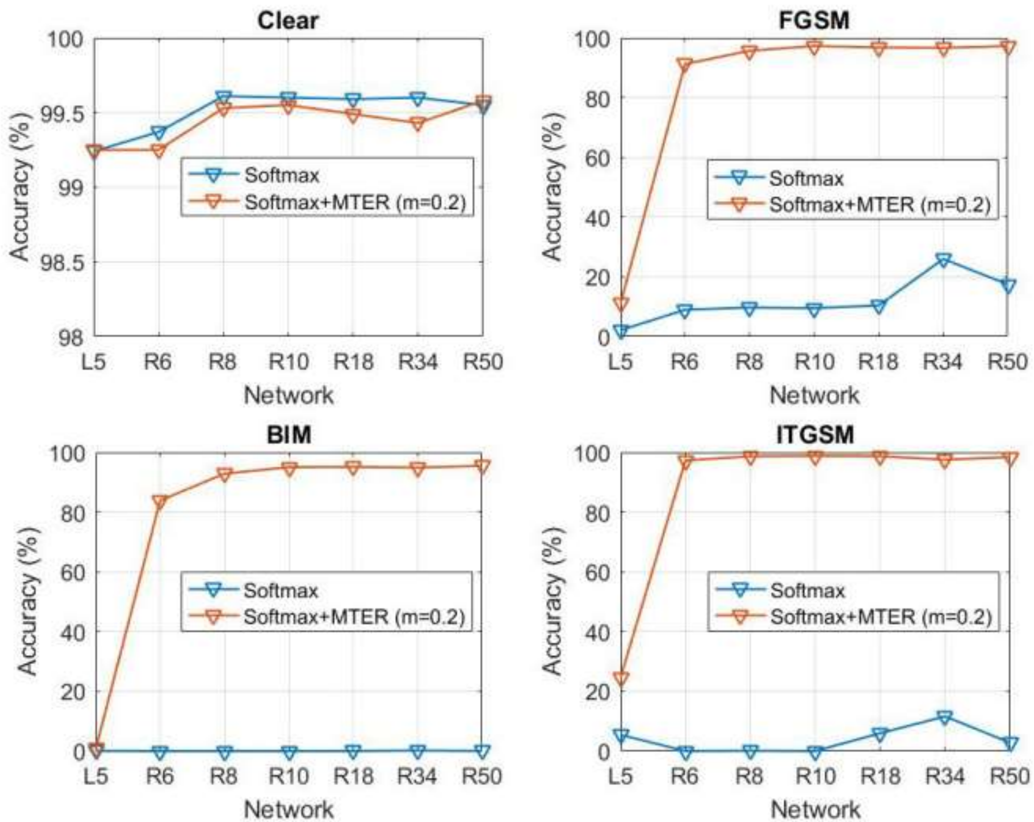


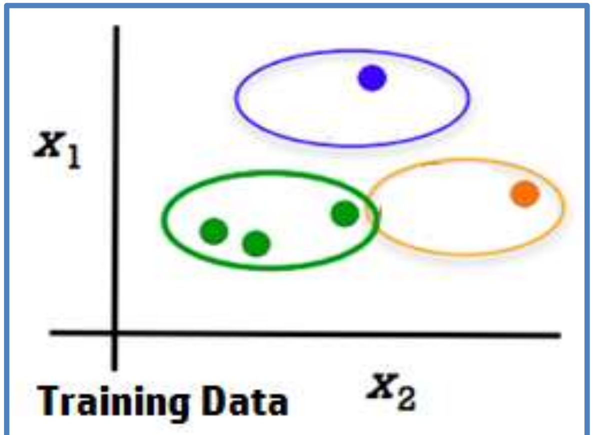
Figure 2. Accuracy on clean images, and adversarial examples on MNIST.



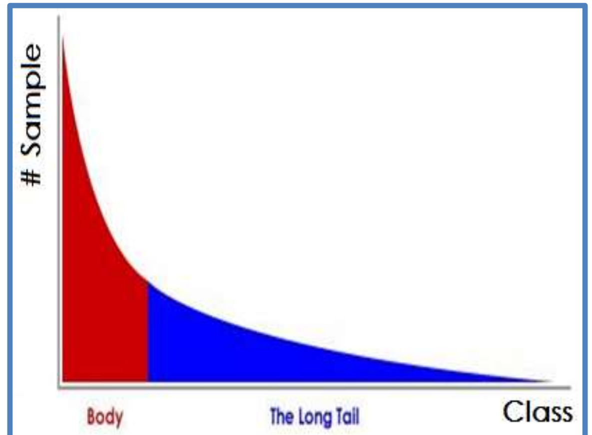
Uncertainty Challenges from Data Deficiency

Uncertainty

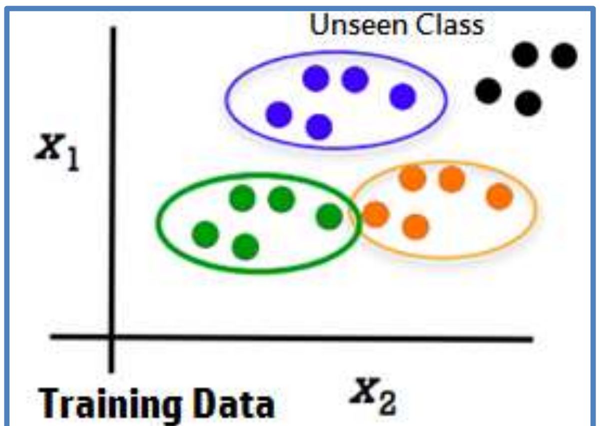
Bias



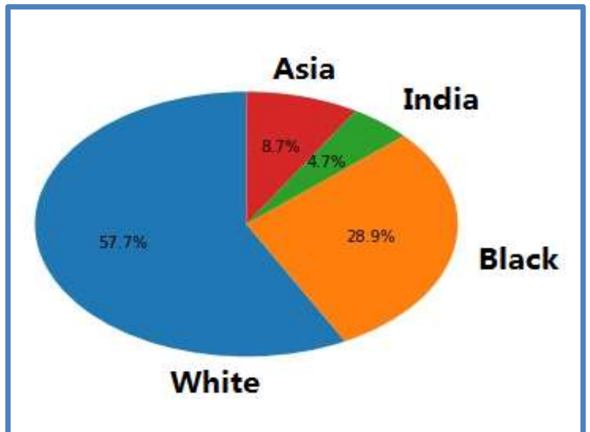
Low-shot training data



Long tailed training data

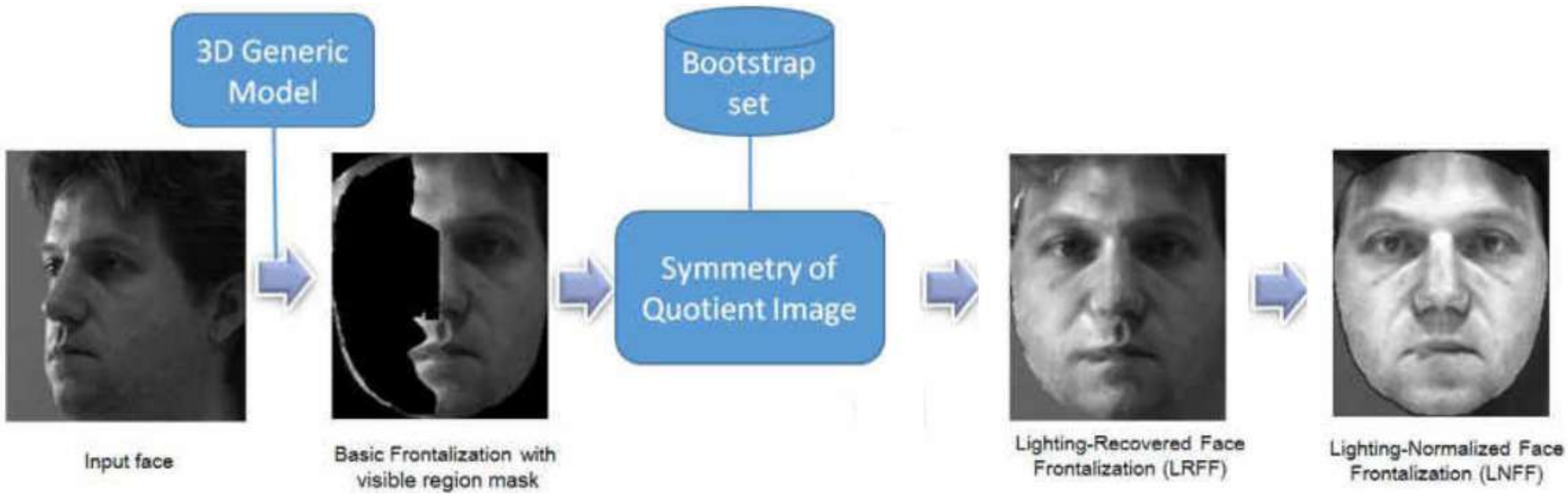


Zero-shot training data

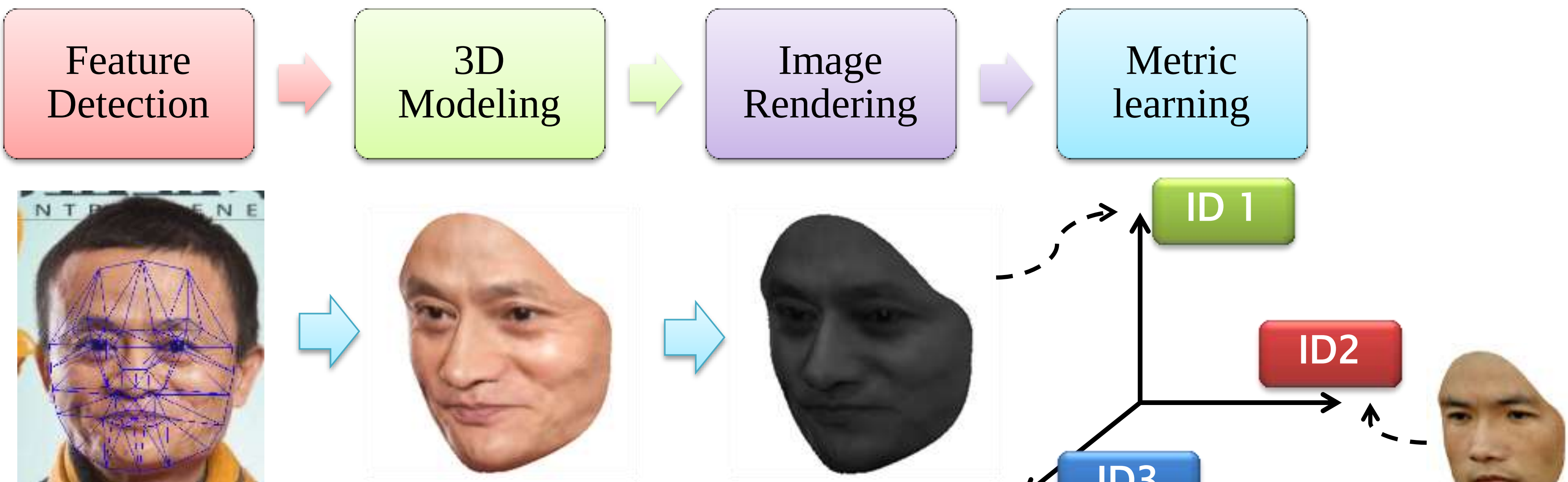


Biased training data

Normalization and Generation via 3D Prior Knowledge

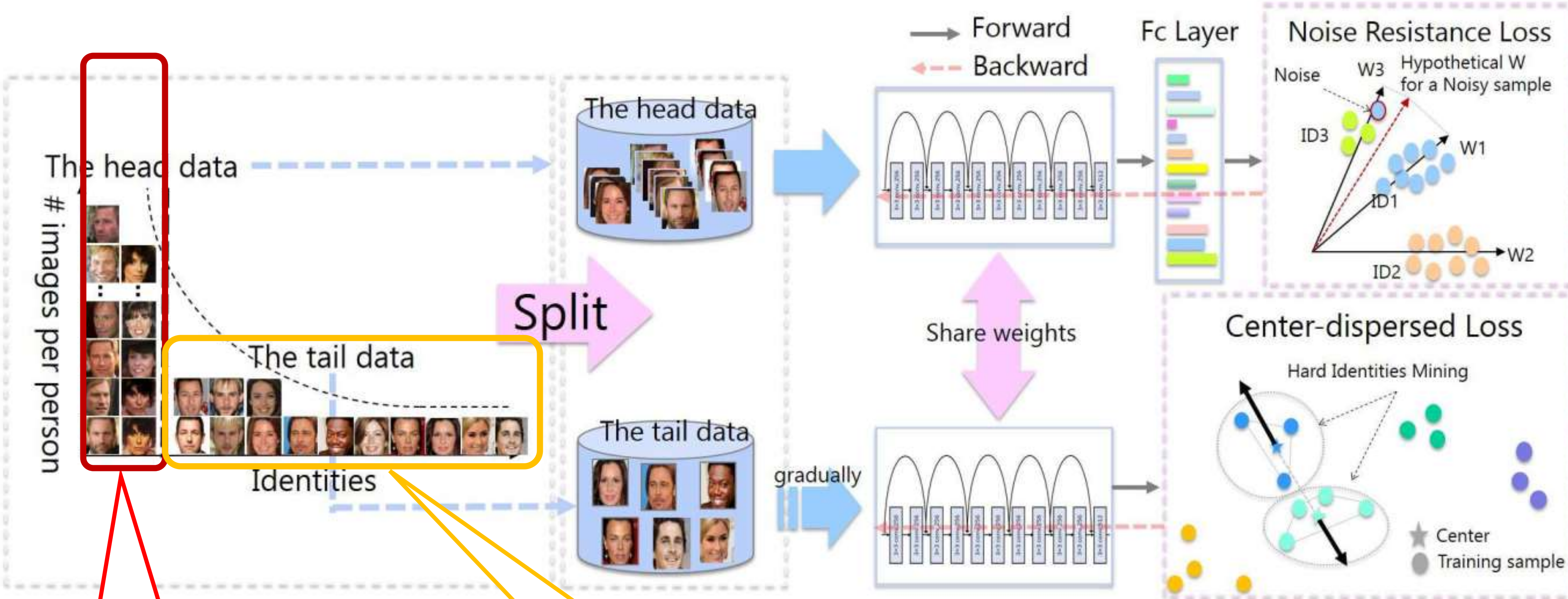


Weihong Deng, et al. Pattern Recognition , 2017



Weihong Deng, et al., Pattern Recognition , 2018

Unequal Training for Long-tailed learning

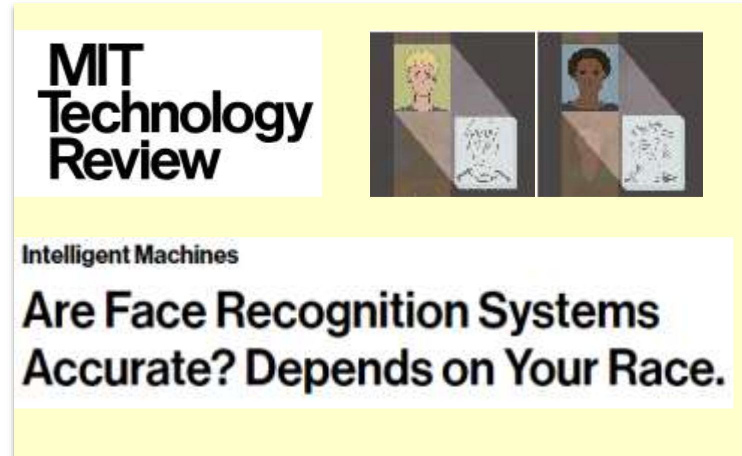


sufficient number of samples to model intra-class variability

Contain sufficient number of classes to model inter-class variability

Yaoyao Zhong, Weihong Deng, et al., CVPR 2019.

Racial Bias



Racial Faces in-the-Wild (RFW)

Mei Wang, Weihong Deng, et al., ICCV 2019.

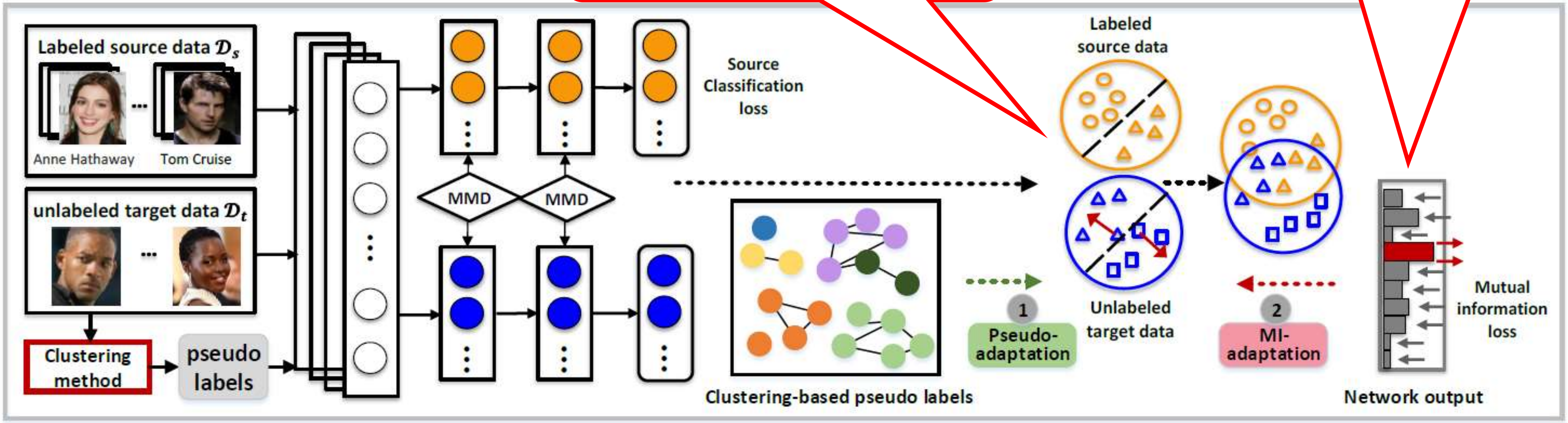
Existence of racial bias

	Model	RFW			
		Caucasian	Indian	Asian	African
SOTA Algorithms	Center-loss	87.18	81.92	79.32	78.00
	SphereFace	90.80	87.02	82.95	82.28
	ArcFace	92.15	88.00	83.98	84.93
	VGGFace2	89.90	86.13	84.93	83.38
	Mean	**90.01**	**85.77**	**82.80**	**82.15**
Commercial APIs	Face++	93.90	88.55	92.47	87.50
	Baidu	89.13	86.53	90.27	77.97
	Amazon	90.45	87.20	84.87	86.27
	Microsoft	87.60	82.83	79.67	75.83
	Mean	**90.27**	**86.28**	**86.82**	**81.89**

Deep information maximization adaptation network (IMAN)

Match global distribution at race-level

Learn discriminative distribution at cluster-level for color races



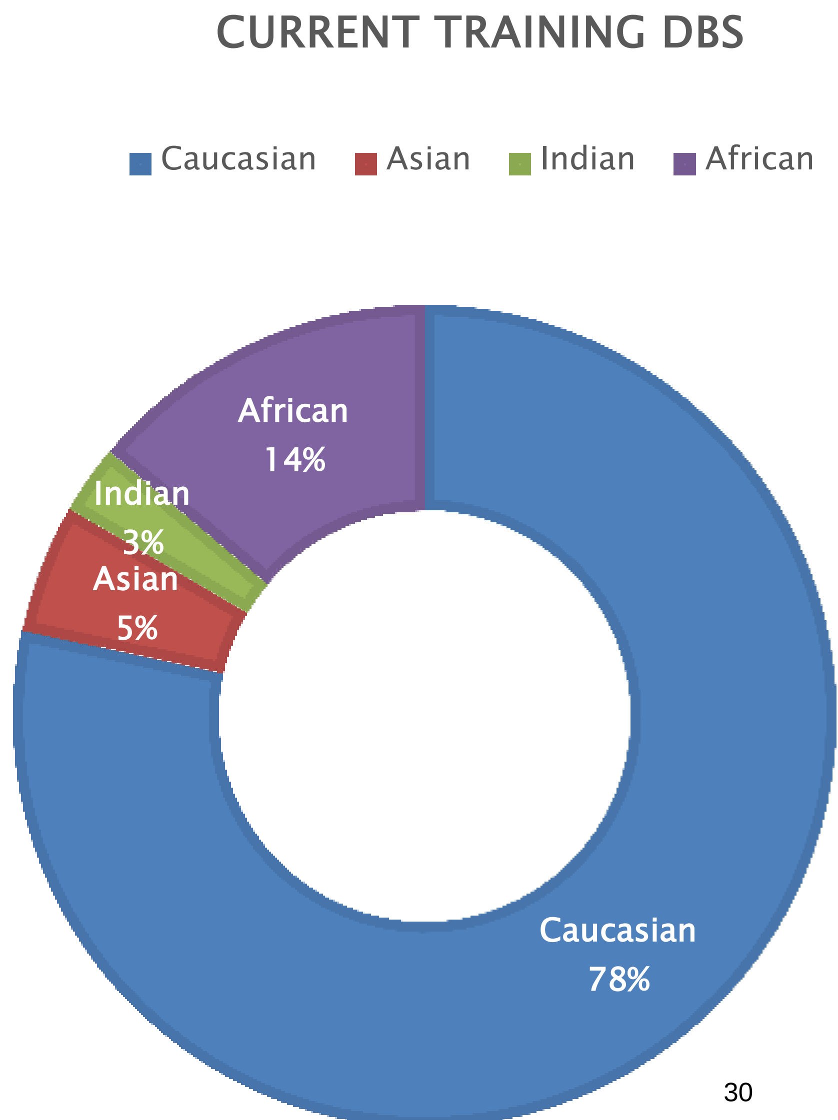
Methods	Caucasian	Indian	Asian	African
Softmax	94.12	88.33	84.60	83.47
DDC-S	-	90.53	86.32	84.95
DAN-S	-	89.98	85.53	84.10
IMAN-S (ours)	-	**91.08**	**89.88**	**89.13**

Recognition accuracy on color races is boosted

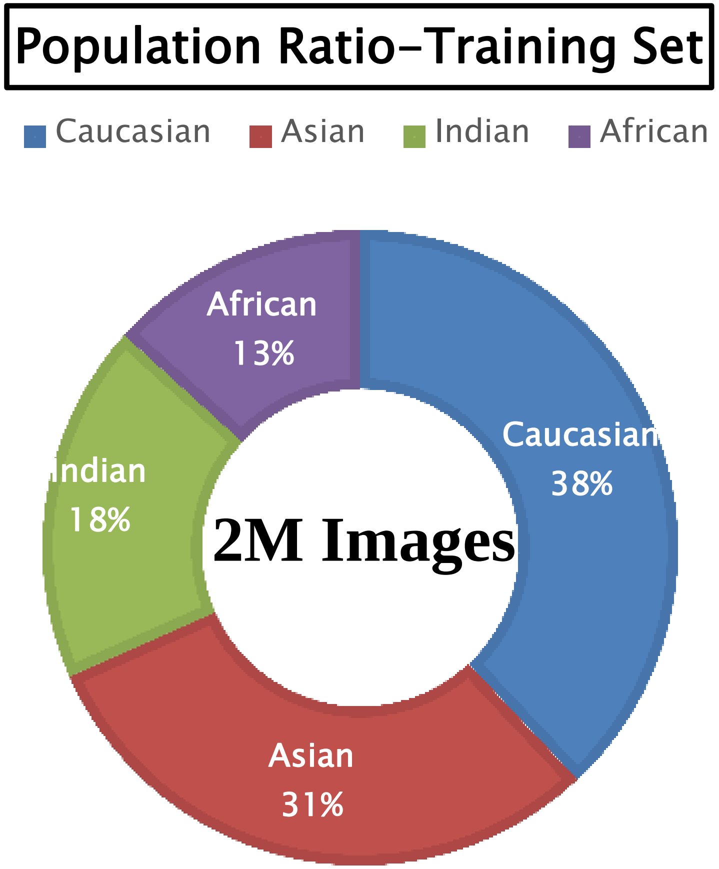
Mei Wang, Weihong Deng, et al., ICCV 2019.

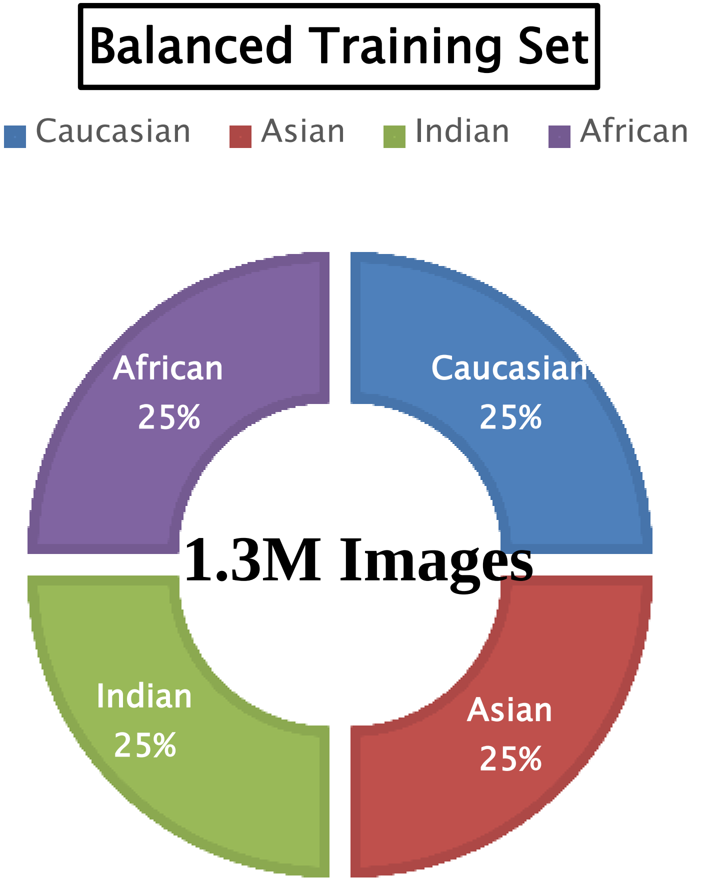
A major driver of bias in face recognition

Database	Racial distribution (%)			
	Caucasian	Asian	Indian	African
CASIA-WebFace	84.5	2.6	1.6	11.3
VGGFace2	74.2	6.0	4.0	15.8
MS-Celeb-1M	76.3	6.6	2.6	14.5
Average	**78.3**	**5.0**	**2.7**	**13.8**



Ethnicity Aware Training Sets for RFW

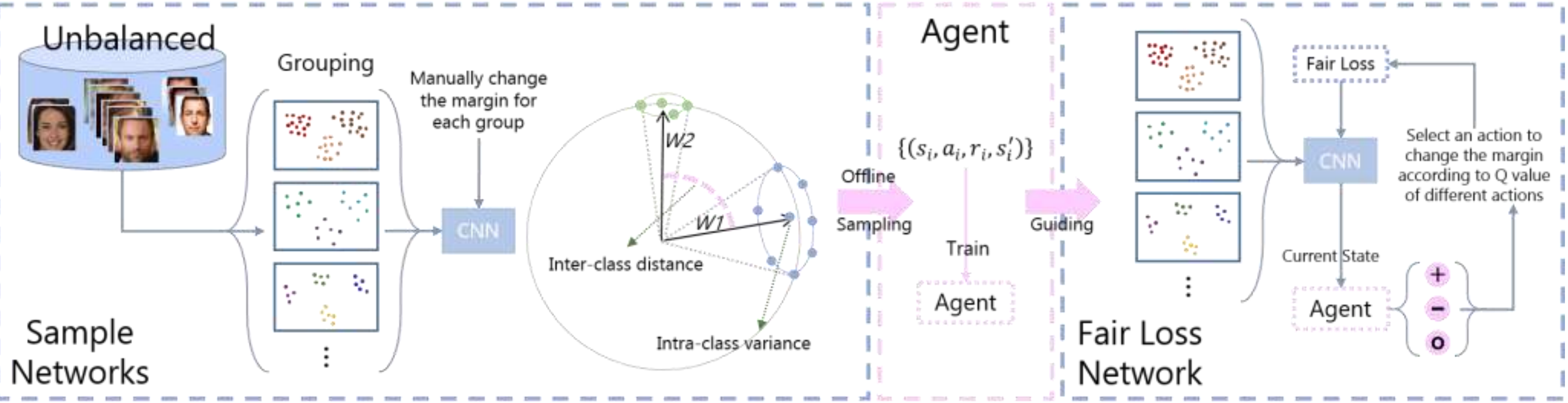




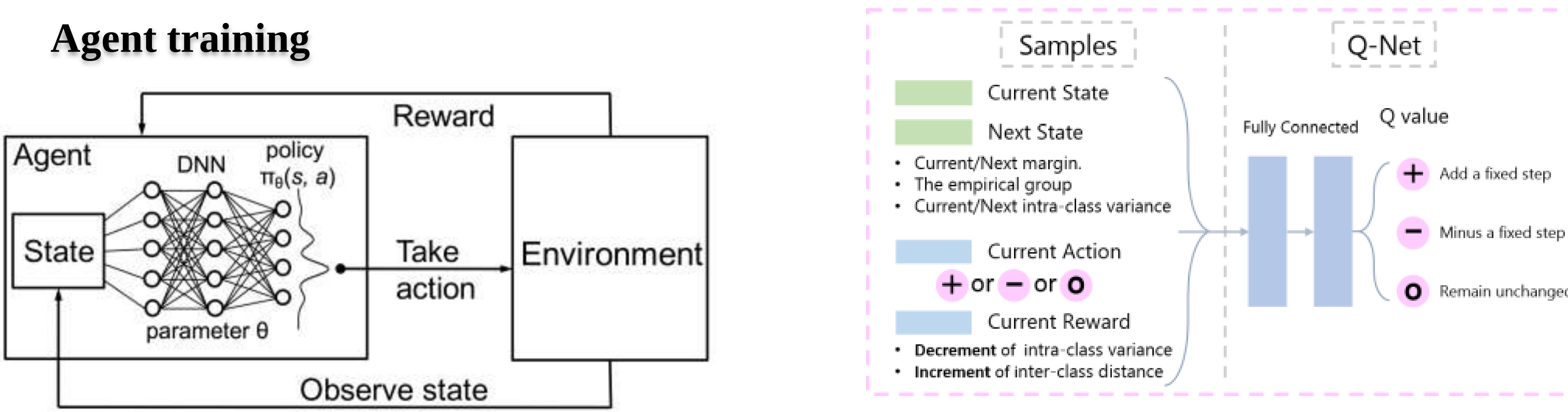
To release in Nov. 2019.

Fair loss for Unbiased Training

Overview



Agent training

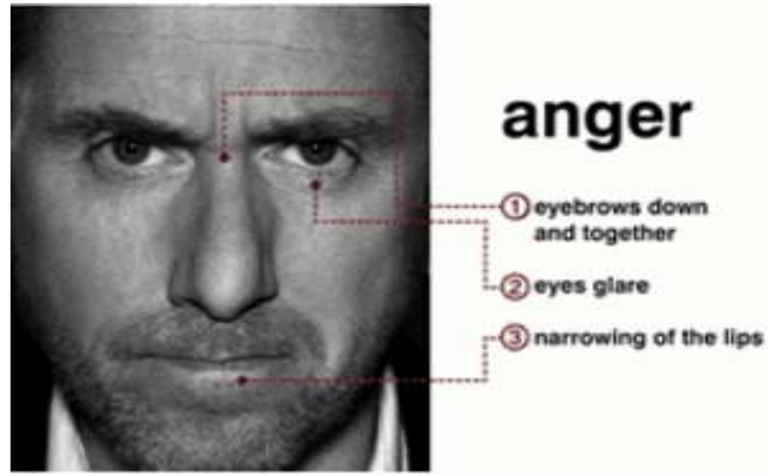


Bingyu Liu, Weihong Deng, et al.,, ICCV 2019.

Understanding Human Emotions

- **Charles Darwin theorized that emotional expression was evoluted by natural selection**

similar



- **Different emotions evolved at different times:**

Ancient parts of the brain

Primal Emotions
e.g., fear

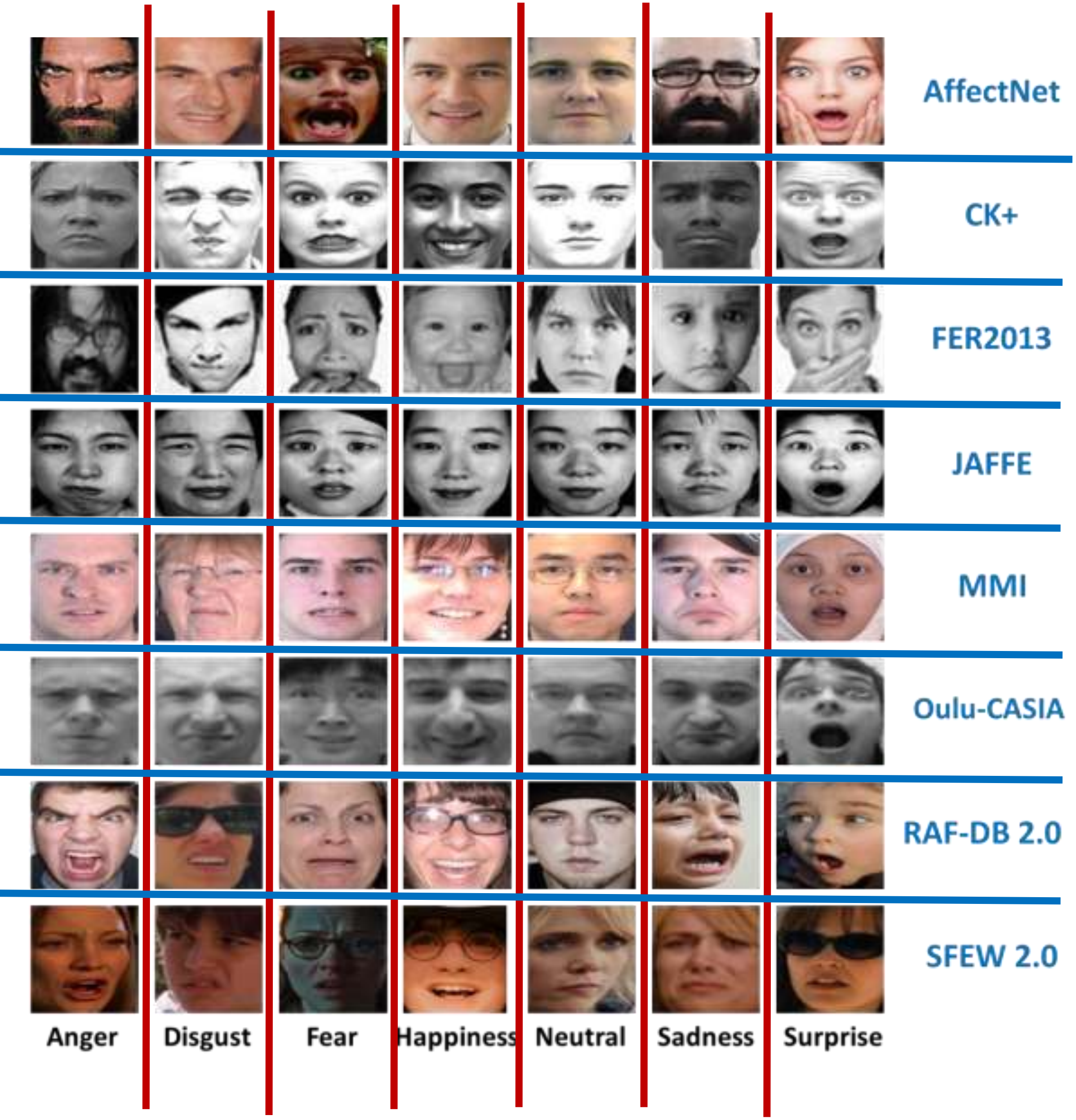
Early mammals

Filial Emotions
e.g., smile

Social Primates

Social Emotions
e.g., guilt

Deeper Look at Expression Dataset Bias

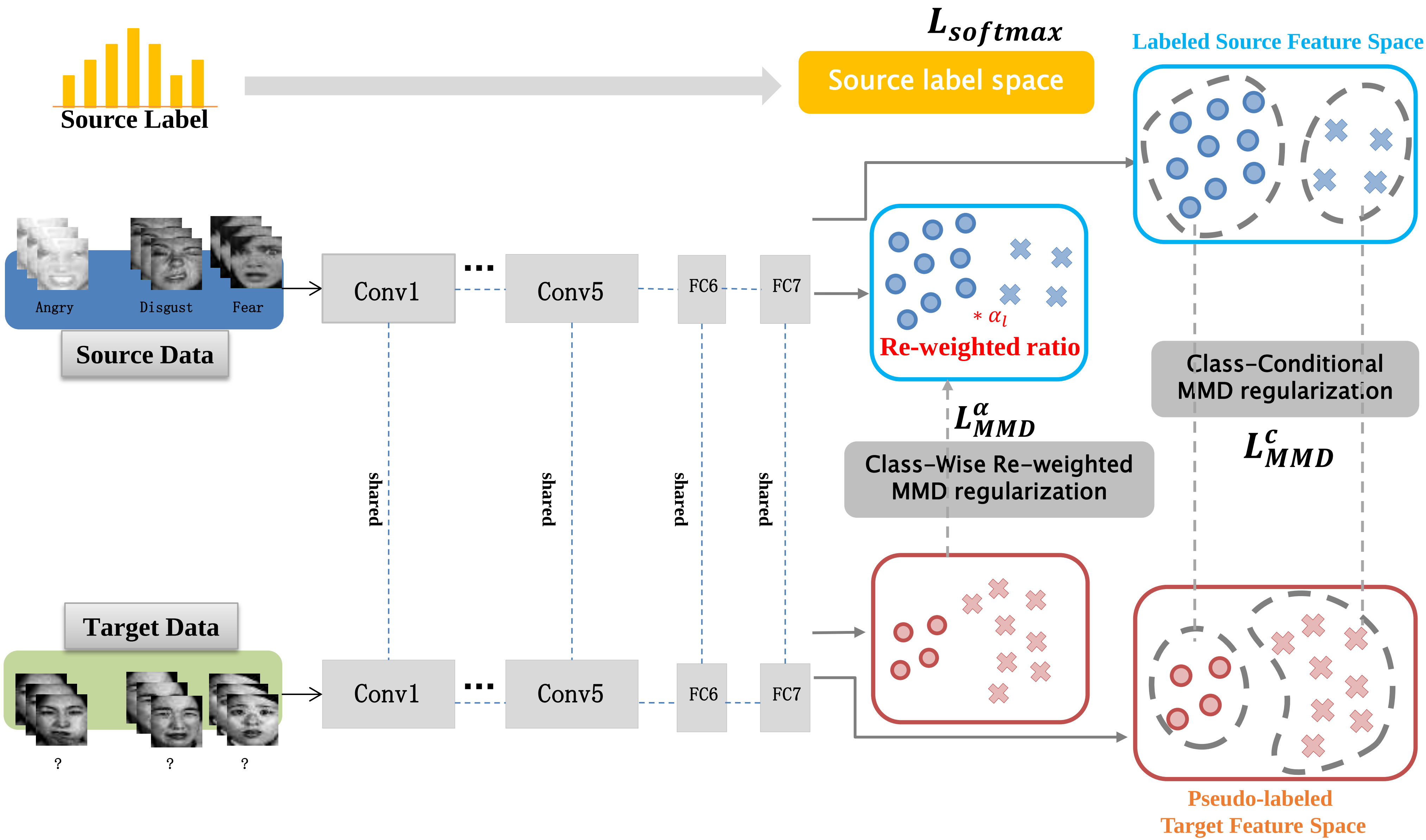


	AffectNet	CK+	FER2013	MMI	Oulu-CASIA	RAF-DB	SFEW 2.0
AffectNet	57.40	1.00	6.75	1.80	0.75	24.75	7.55
CK+	0.80	96.69	0.10	0.00	0.30	1.56	0.55
FER2013	8.05	0.50	78.05	0.50	1.05	9.55	2.30
MMI	9.86	0.15	1.10	76.16	2.14	10.35	0.25
Oulu-CASIA	0.47	0.00	0.78	0.05	96.30	2.34	0.05
RAF-DB	23.70	1.90	8.65	1.75	2.10	58.40	3.50
SFEW 2.0	3.60	0.00	0.80	0.05	0.15	1.70	93.70

Confusion matrix

Shan Li, Weihong Deng, arXiv:1904.11150, TAC 2020.

Emotion-Conditional Adaption Network (ECAN)



Shan Li, Weihong Deng, arXiv:1904.11150, TAC 2020.

Crowdsourcing: Select a single basic expression

Dataset Construction (RAF-ML)

Data collection and Annotation Process

1. Collection

Keywords
'smile' 'crying' 'OMG'… → flickr → XML ↓ parse URLs → download → 30K images
60,000 images downloaded

2. Annotation

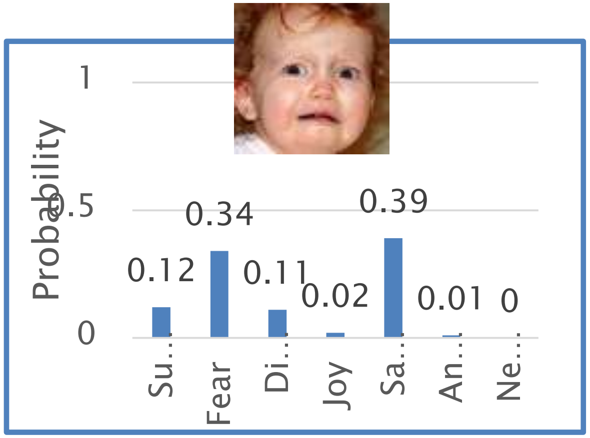
Crowd-sourcing
315 volunteers online
Each image labelled 40 times

Learning from labels → 1.2M labels

3. Reliability Estimation

An **EM** framework → Filter out unreliable labels → *Enhanced Reliability*

Results



Single label / Mutli-label

RAF-DB *RAF-ML*

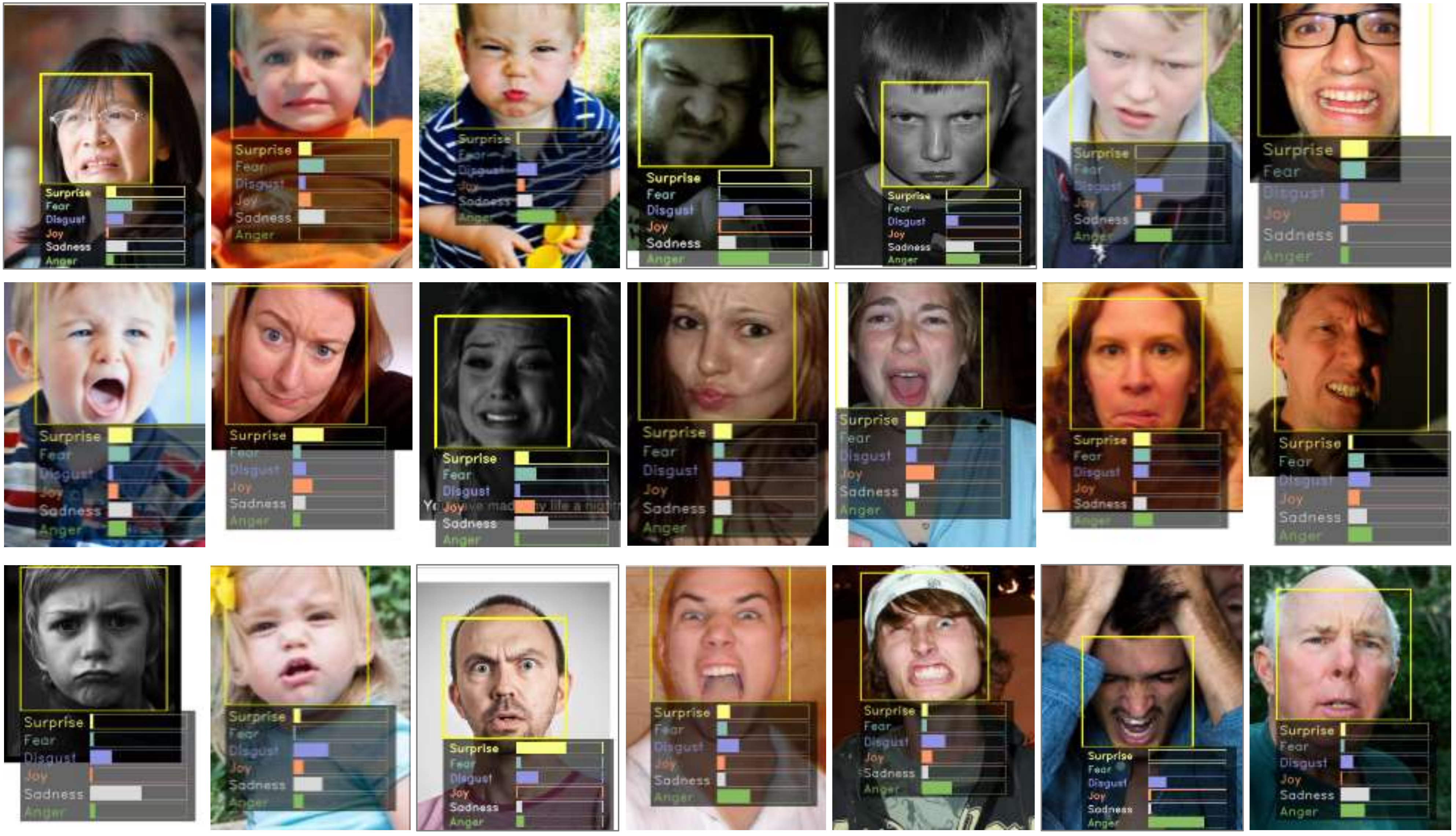
Dataset Construction (RAF-DB)

7 classes Basic Emotions

12 classes Compound Emotions

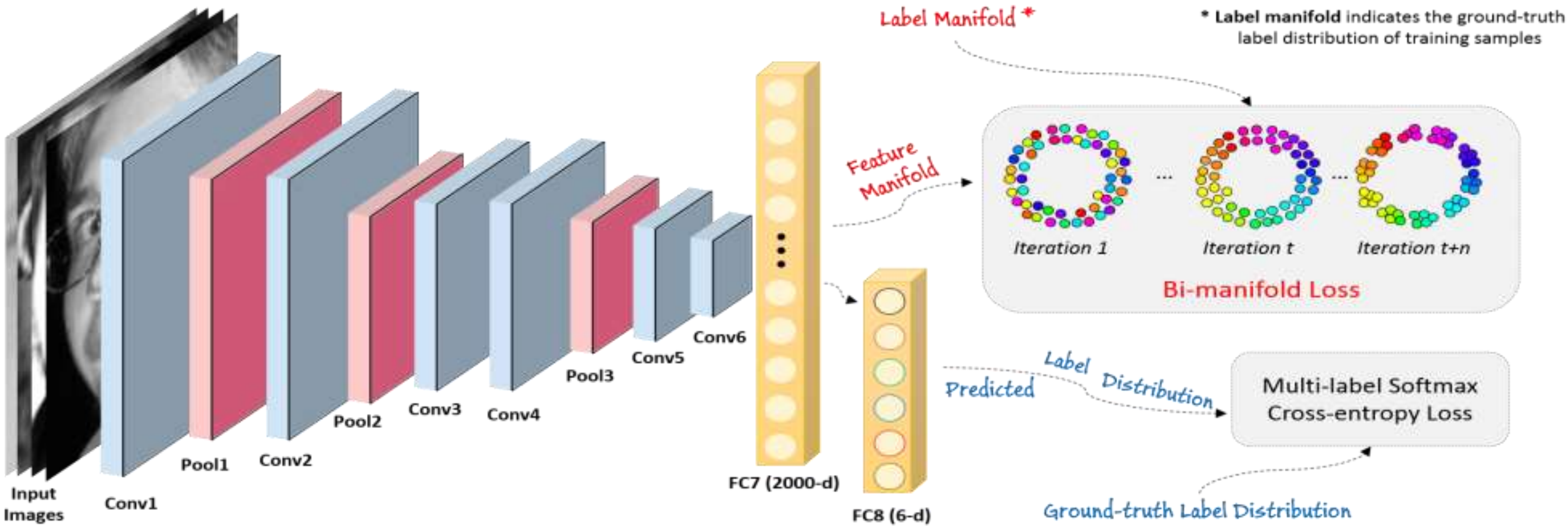
Shan Li, Weihong Deng, CVPR 2017, TIP 2019.

Dataset Construction (RAF-ML)



Shan Li, Weihong Deng, IJCV 2019.

Deep Bi-manifold CNN (DBM-CNN)



Our goal: x^f: samples in feature manifold
x^l: samples in label manifold

$$\min \sum_{i,j} (S_{ij}^f + S_{ij}^l) ||x_i^f - x_j^f||_2^2$$

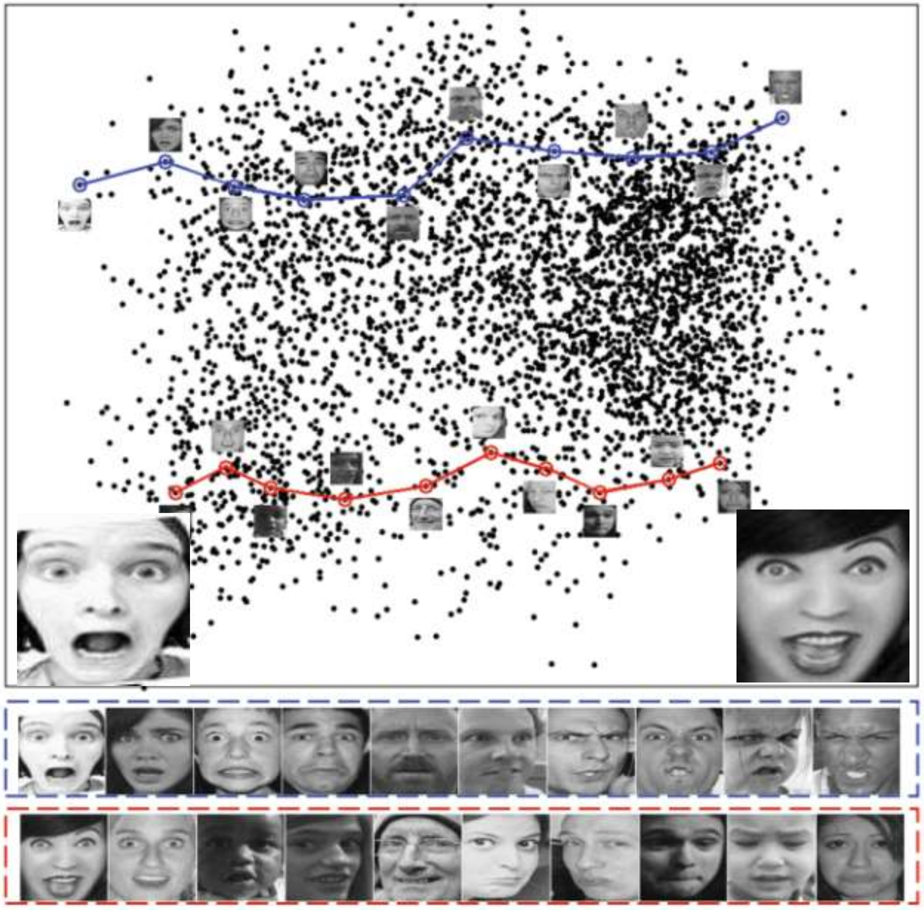
$$S_{ij}^* = \begin{cases} 1, x_j^* is\ knn\ of\ x_i^*, vice\ versa \\ 0, otherwise \end{cases}$$

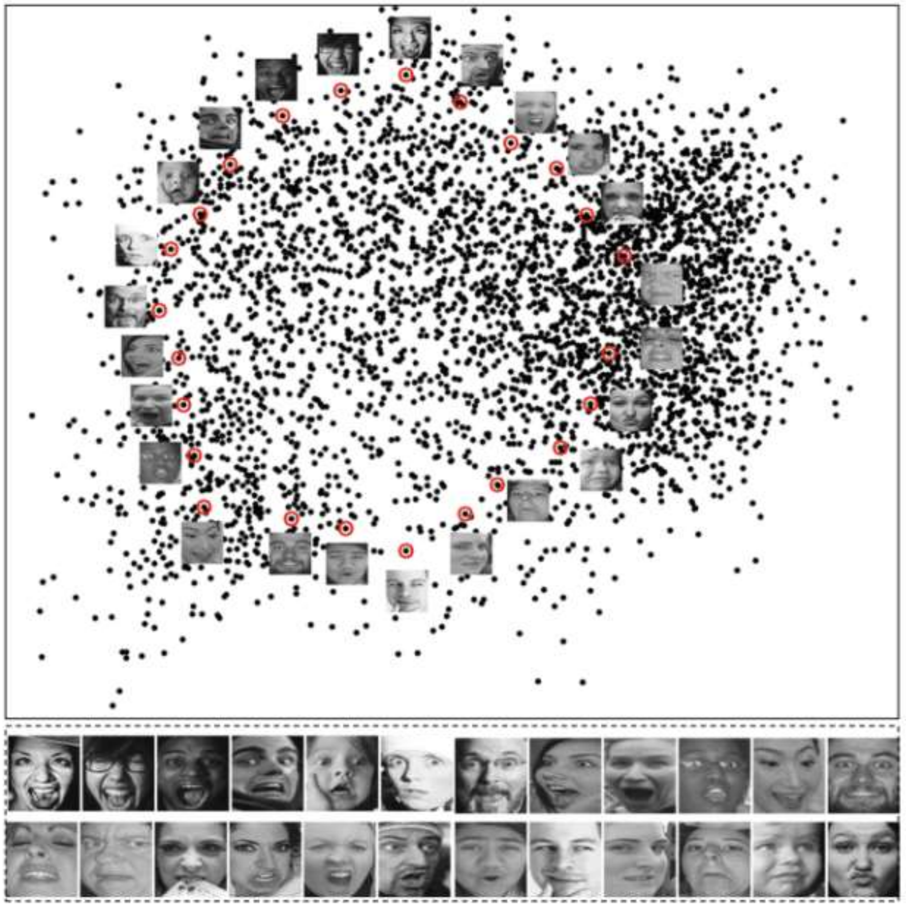
Bi-Manifold Loss:

$$L_{bm} = \frac{1}{2n} \sum_{i=1}^{n} \left\| 2x_i^f - \frac{1}{k} \sum_{x \in N_k^f\{x_i\}} x^f - \frac{1}{k} \sum_{x \in N_k^l\{x_i\}} x^f \right\|$$

Shan Li, Weihong Deng, IJCV 2019.

Deep Bi-manifold CNN (DBM-CNN)

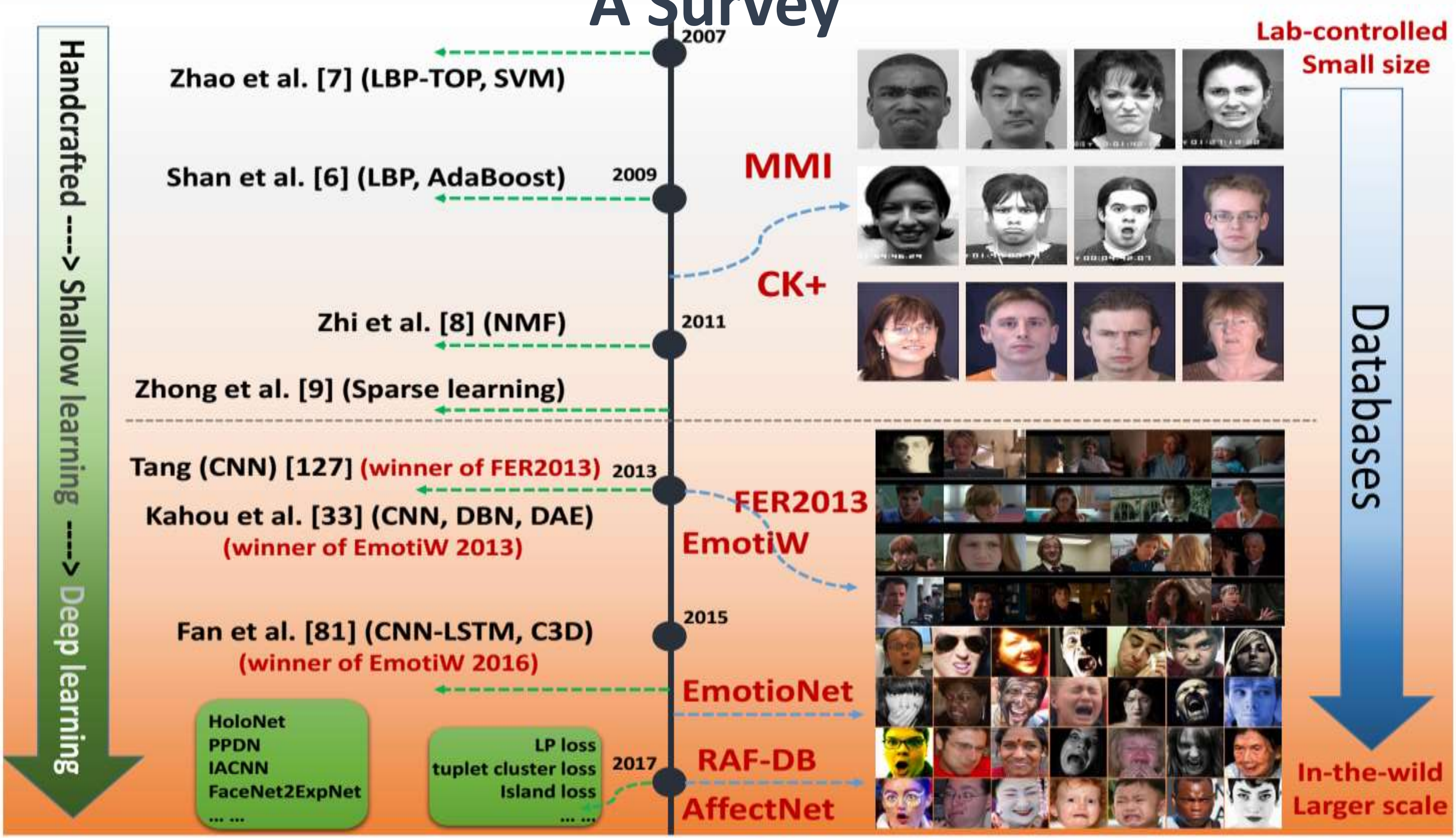




Smoothness in terms of both face appearance and emotion perception

Shan Li, Weihong Deng, IJCV 2019.

Deep Facial Expression Recognition: A Survey



Shan Li & W. Deng, Deep Facial Expression Recognition: A Survey, [(arXiv:1804.08348)](arXiv:1804.08348)

Summary: Technical contributions

- Major Contributions: Deep representation learning

Tiny interclass-difference

- Massive number of classes
- Adversarial images

Robustness

Separable→Large margin

- Virtual softmax (NIPS'18)
- Unequal learning, Fair loss (CVPR'19 ,ICCV'19)
- MTER for adversarial learning (ICCV'19)

Huge intra-class variations

- Variable intensity
- Spatial variations

Adaptiveness

Lobality→Locality

- DLP-CNN (TIP'19), DBM-CNN (IJCV'19)
- Hybird attention (CVPR'19)
- High-order attention (ICCV'19)

Global distribution shift

- zero-shot learning
- Data bias problem, non-iid

Generalization

Specificity→commonality

- Confusion regularization of metric learning (AAAI'18)
- Information maximum TL (ICCV'19)
- Reducing racial bias (ICCV'19)

Towards Visual Intelligence

Acknowledgements

For data, code on :

http://www.whdeng.cn

Thank you !

Major Sponsors

国家自然科学基金委员会
National Natural Science Foundation of China
NSFC

Canon

滴滴
滴滴一下 美好出行