2020.07 하계방학 Part.2

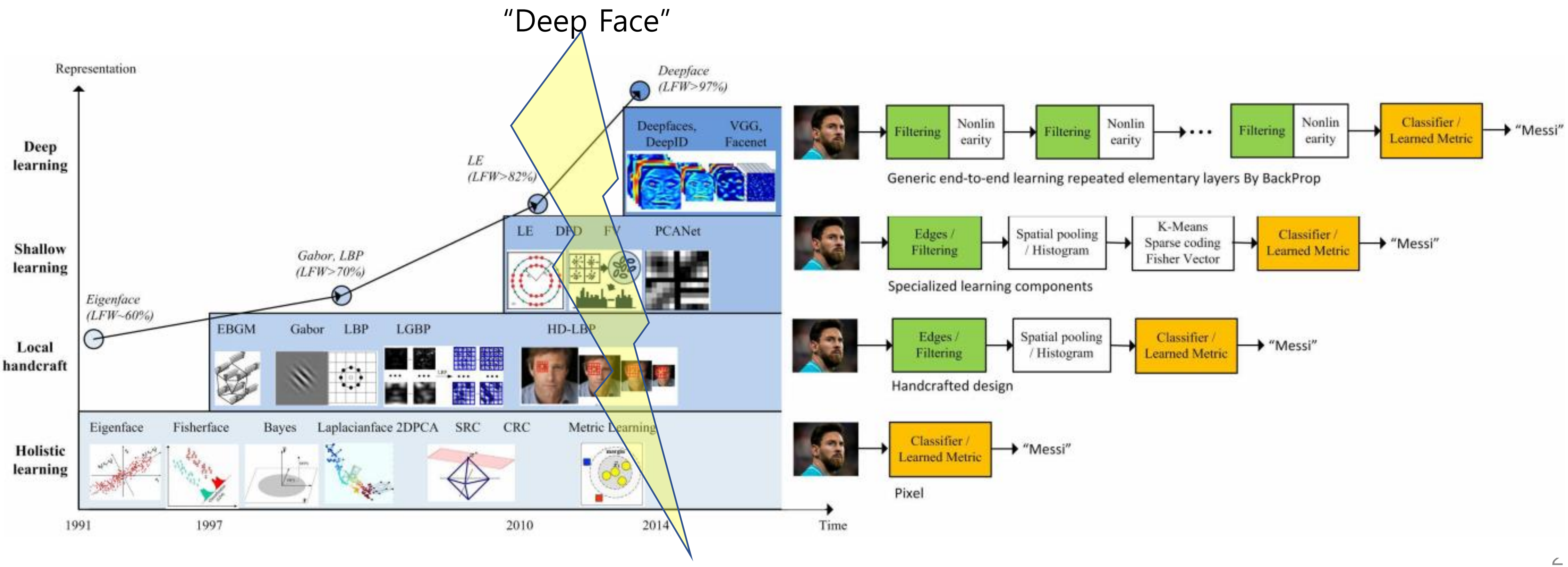
Deep Face Recognition: Survey

Mei Wang, Weihong Deng School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China.

20151739 공대현

• Face Recognition(FR)의 발전

-2014년을 기점으로 DL 사용 -> 비약적인 정확도 상승



Main Point of Face Recognition

- Face Processing
- Loss Function

- **Prior Knowledge**

  - Gallery : Subjects enrolled in the system (Training Data)

  - Probe : New Subjects (Test Data)

[273]

Prior Knowledge

- Face Verification

- 두 사람(Gallery-Probe)이 같은 사람인지 비교

$$r(x_1, x_2) = log\frac{P\left(x_1, x_2|H_I\right)}{P\left(x_1, x_2|H_E\right)}$$

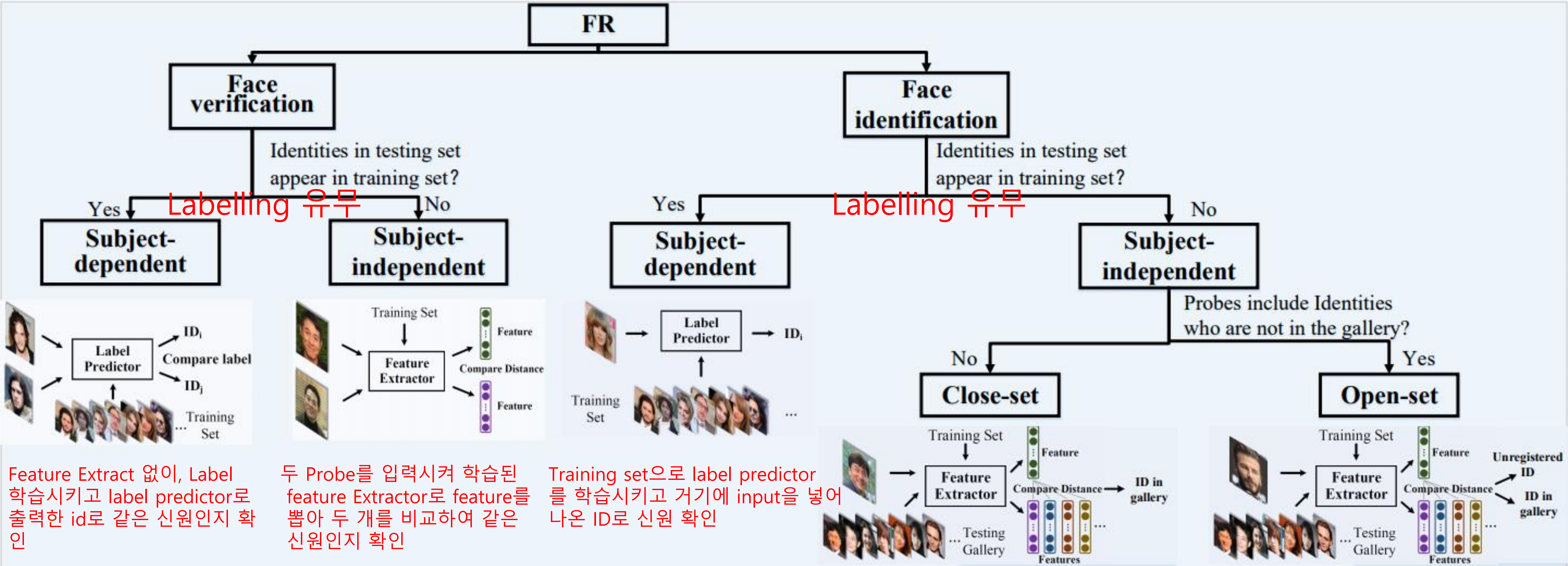
- Face Identification

- 여러 사람 중 가장 비슷한 사람 찾아냄(Gallery에서 Probe와 가장 비슷한 사람 찾아냄)

- Both use Feature Similarity (Metric)

[273]

- # Prior Knowledge



Labelling 유무

Labelling 유무

Feature Extract 없이, Label 학습시키고 label predictor로 출력한 id로 같은 신원인지 확인

두 Probe를 입력시켜 학습된 feature Extractor로 feature를 뽑아 두 개를 비교하여 같은 신원인지 확인

Training set으로 label predictor를 학습시키고 거기에 input을 넣어 나온 ID로 신원 확인

Training set으로 Feature Extractor를 학습시키고 거기에 input을 넣어 나온 feature와 학습된 feature를 비교하여 신원확인

Training set으로 label predictor를 학습시키고 거기에 input을 넣어 나온 ID로 신원 확인하는데, input이 gallaery에 없는 신원일 수 있음

• Prior Knowledge

Dataset

TABLE VI
THE COMMONLY USED FR DATASETS FOR TRAINING

Datasets	Publish Time	#photos	#subjects	# of photos per subject [1]	Key Features
MS-Celeb-1M (Challenge 1)[69]	2016	10M 3.8M(clean)	100,000 85K(clean)	100	breadth; central part of long tail; celebrity; knowledge base
MS-Celeb-1M (Challenge 2)[69]	2016	1.5M(base set) 1K(novel set)	20K(base set) 1K(novel set)	1/-/100	low-shot learning; tailed data; celebrity
MS-Celeb-1M (Challenge 3) [2]	2018	4M(MSv1c) 2.8M(Asian-Celeb)	80K(MSv1c) 100K(Asian-Celeb)	-	breadth;central part of long tail; celebrity
MegaFace [105], [145]	2016	4.7M	672,057	3/7/2469	breadth; the whole long tail;commonalty
VGGFace2 [22]	2017	3.31M	9,131	87/362.6/843	depth; head part of long tail; cross pose, age and ethnicity; celebrity
CASIA WebFace [243]	2014	494,414	10,575	2/46.8/804	celebrity
UMDFaces-Videos [10]	2017	22,075	3,107	–	video
VGGFace [149]	2015	2.6M	2,622	1,000	depth; celebrity; annotation with bounding boxes and coarse pose
CelebFaces+ [187]	2014	202,599	10,177	19.9	private
Google [176]	2015	>500M	>10M	50	private
Facebook [195]	2014	4.4M	4K	800/1100/1200	private

[1] The min/average/max numbers of photos or frames per subject

[273]

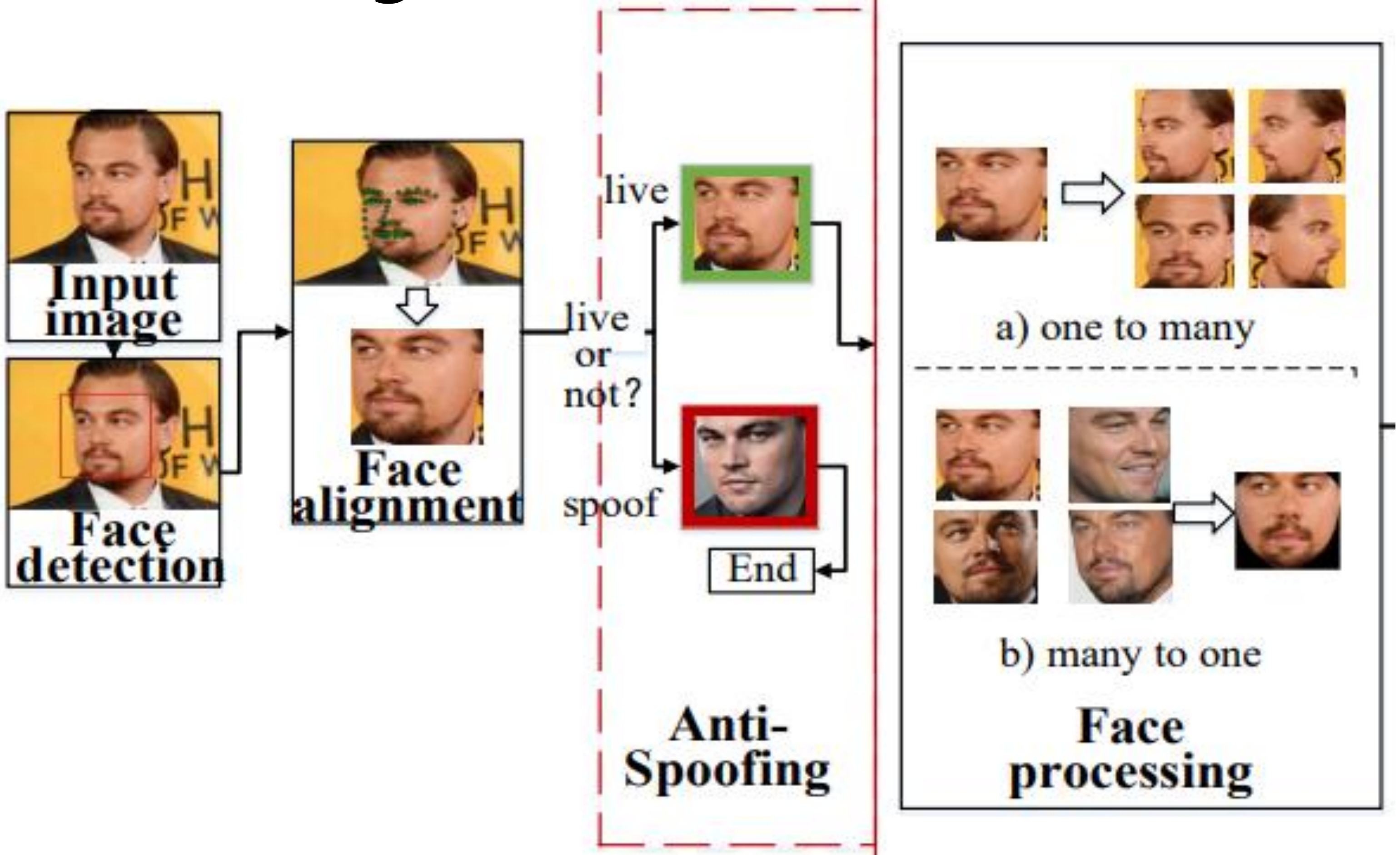
• Face Processing

-Conditions(**Poeses**, illumination, expressions, occlusion …) affect Performance of FR -> need to be data-processed

- “One-to-many augmentation”: single image -> images of the pose variability (ex: 정면 -> 여러각도의 얼굴 생성)

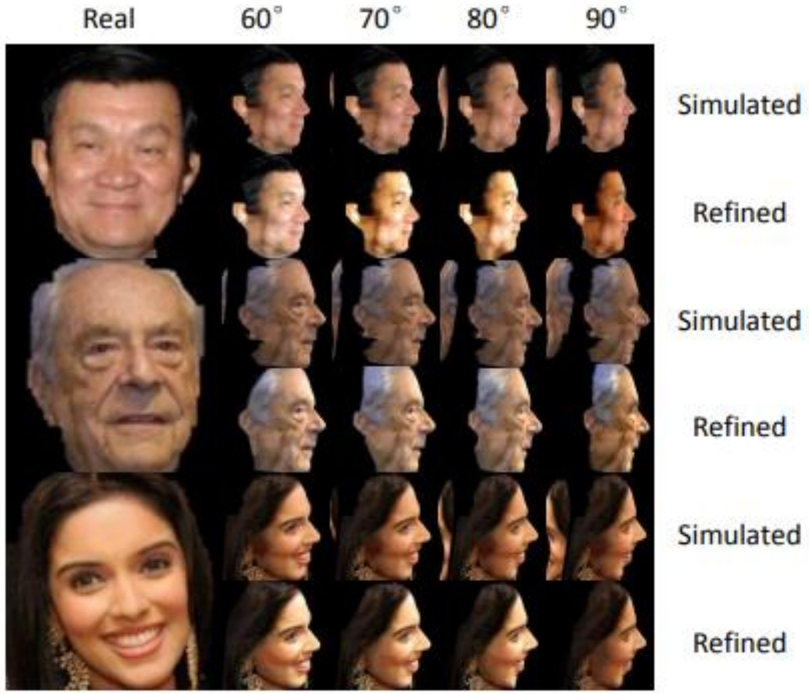
- “Many-to-one normalization” : recovering the canonical view of face images from one or many images of a nonfrontal view (ex : 여러각도의 얼굴 -> 정면으로 normalize)

- # Face Processing

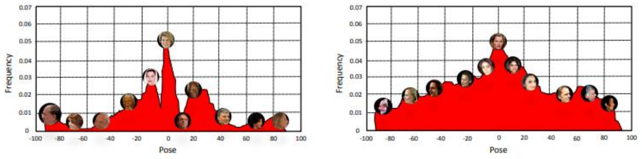


- # Face Processing

  - DA GAN (Dual Agent GAN)

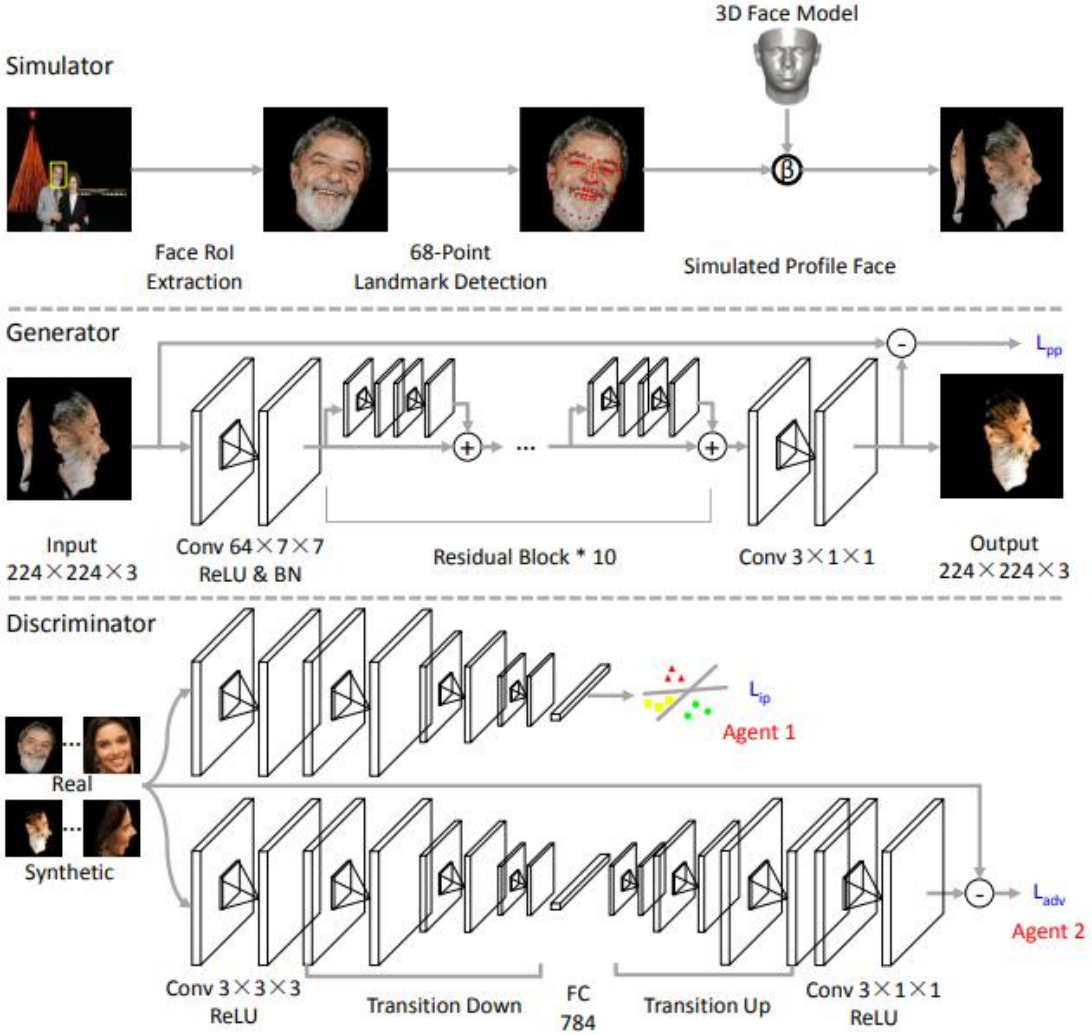


(a) Refined results of DA-GAN.



(a) Extremely unbalanced pose distribution. (b) Well balanced pose distribution with DA-GAN.

Figure 1: Comparison of pose distribution in the IJB-A (15) dataset w/o and w/ DA-GAN.



• Face Processing

- CG GAN (Dual Agent GAN)

use the first 200 subjects under 9 poses (−60◦ ~ +60◦) and 20 illuminations for training,
and use the remaining 137 subjects under 9 poses and 20 illuminations for testing

Table I
RECOGNITION RATES (%) COMPARING CROSS-GENERATING, FRONTAL REPRESENTATION AND REMOTE CODING

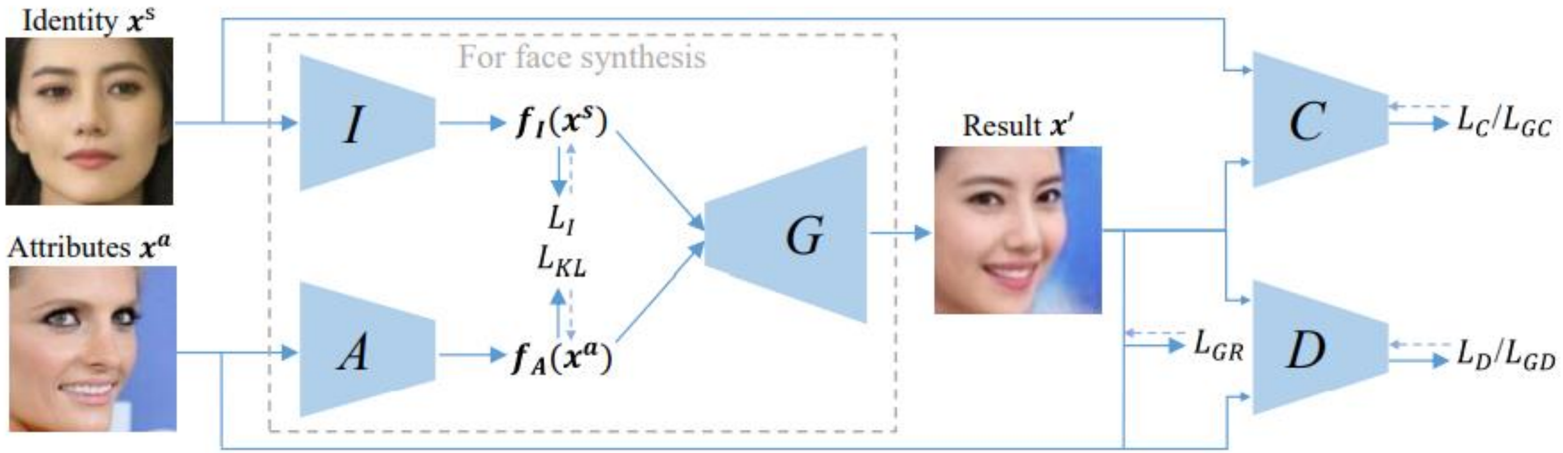
FIR Method	0°	15°	30°	45°	60°	Average
Cross-genrating	**99.7**	**99.1**	**97.2**	**93.9**	**85.3**	**94.5**
Frontal representation	99.1	97.9	93.8	91.2	82.4	92.1
Remote coding	98.7	96.2	93.0	89.8	80.5	90.9

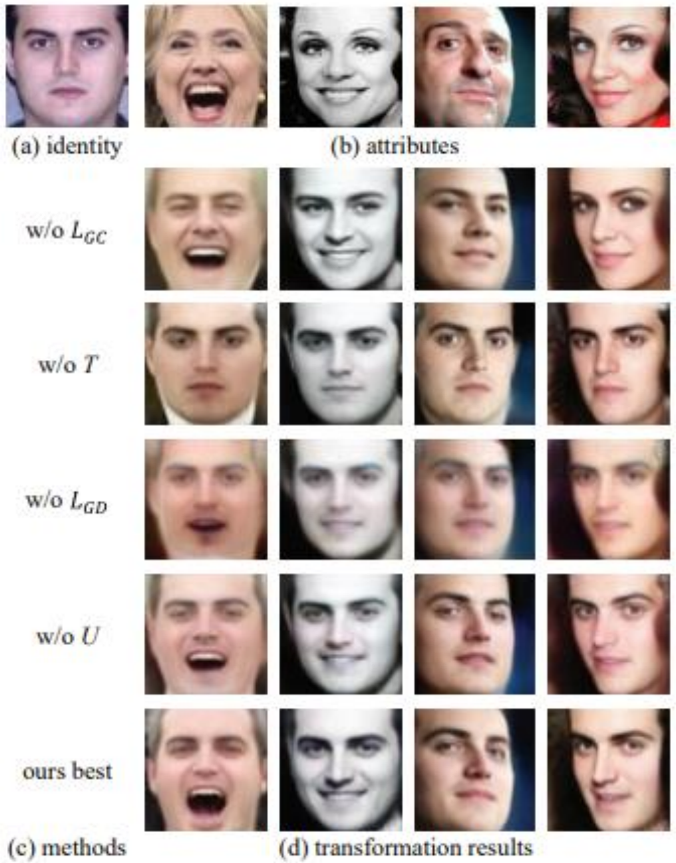
Table III
BENCHMARK COMPARISION OF RECOGNITION RATES (%) ON MULTI-PIE

Method	0°	15°	30°	45°	60°	Average
Zhu et al. [3]	94.3	90.7	80.7	64.1	45.9	72.9
Yim et al. [7]	99.5	95.0	88.5	79.9	61.9	83.3
Tran et al. [2]	97.0	94.0	90.1	86.2	83.2	89.2
CG-GAN	**99.7**	**99.1**	**97.2**	**93.9**	**85.3**	**94.5**

- # Face Processing

Towards Open-Set Identity Preserving Face Synthesis





• Face Processing

- Different Data Processing Approaches

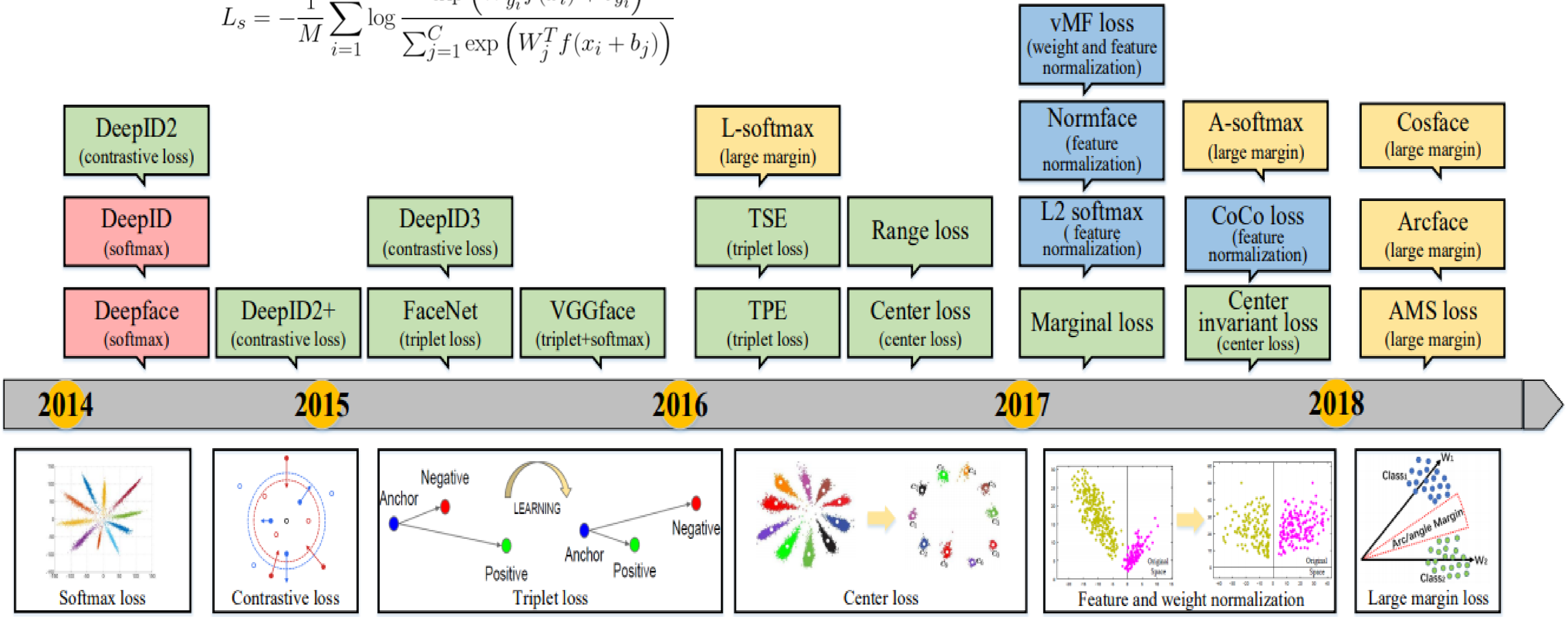
TABLE I
DIFFERENT DATA PREPROCESSING APPROACHES

Data processing	Brief Description	Subsettings
one to many	generating many patches or images of the pose variability from a single image	3D model [139], [137], [165], [166], [53] [67], [197], [196]
		2D deep model [279], [267], [182]
		data augmentation [124], [276], [51] [222], [187], [188], [192], [202]
many to one	recovering the canonical view of face images from one or many images of nonfrontal view	SAE [101], [264], [240]
		CNN [278], [280], [89], [37], [246]
		GAN [91], [198], [41], [249]

• Loss Function

- **Softmax Loss : if intra-variation > inter-variation, not good**

$$L_s = -\frac{1}{M}\sum_{i=1}^{M}\log\frac{\exp\left(W_{y_i}^T f(x_i) + b_{y_i}\right)}{\sum_{j=1}^{C}\exp\left(W_j^T f(x_i + b_j)\right)}$$



• Loss Function

• Euclidean-distance-based loss : metric learning method

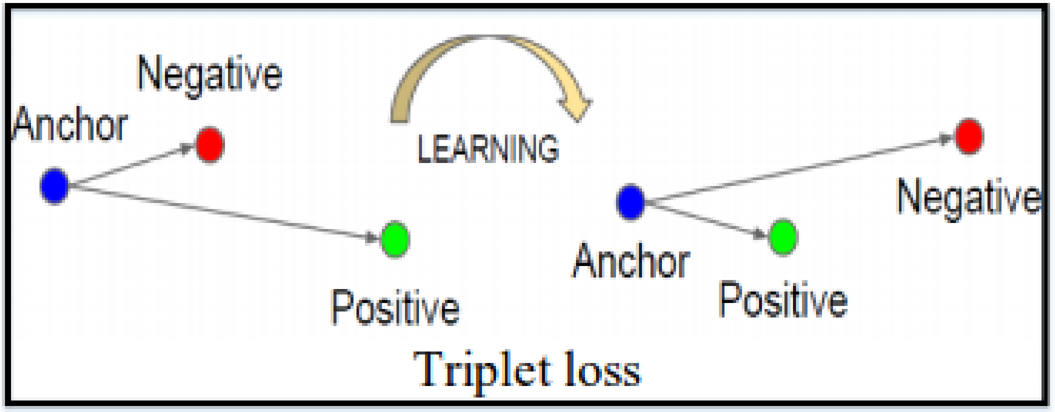
$$\mathcal{L} = y_{ij} max\left(0, \|f(x_i) - f(x_j)\|_2 - \epsilon^+\right) + (1 - y_{ij}) max\left(0, \epsilon^- - \|f(x_i) - f(x_j)\|_2\right)$$

If $x_i = x_j$, y_{ij} = 1 → verification
If $x_i \neq x_j$, y_{ij} = -1 → identification

FaceNet : Triplet Loss $\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < -\|f(x_i^a) - f(x_i^n)\|_2^2$

Minimizes the distance between an anchor and a positive sample of the same identity and maximizes the distance between the anchor and a negative sample of a different identity

단점: Training instability due to
the selection of effective training samples



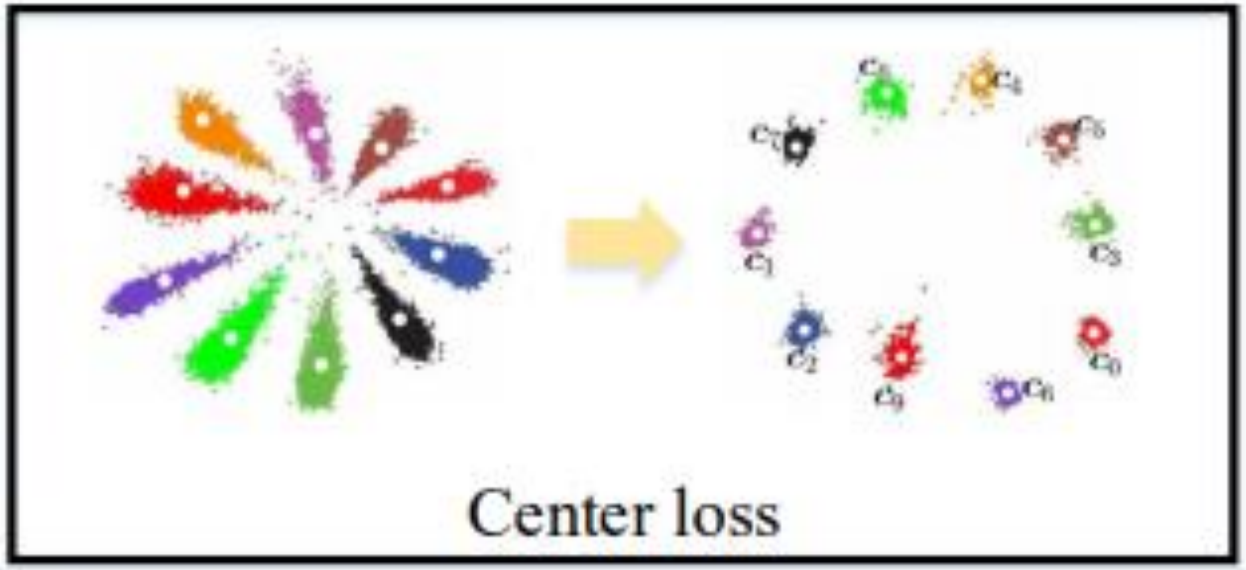
• Loss Function

• Euclidean-distance-based loss : metric learning method

$$\mathcal{L}_C = \frac{1}{2}\sum_{i=1}^{m} \|x_i - c_{y_i}\|_2^2$$

단점 : Massive GPU memory consumption on the classification layer

Centor loss:



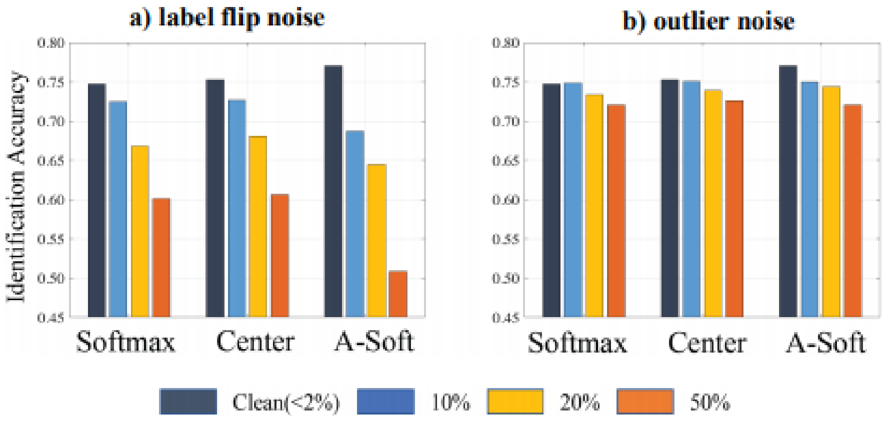
• Loss Function

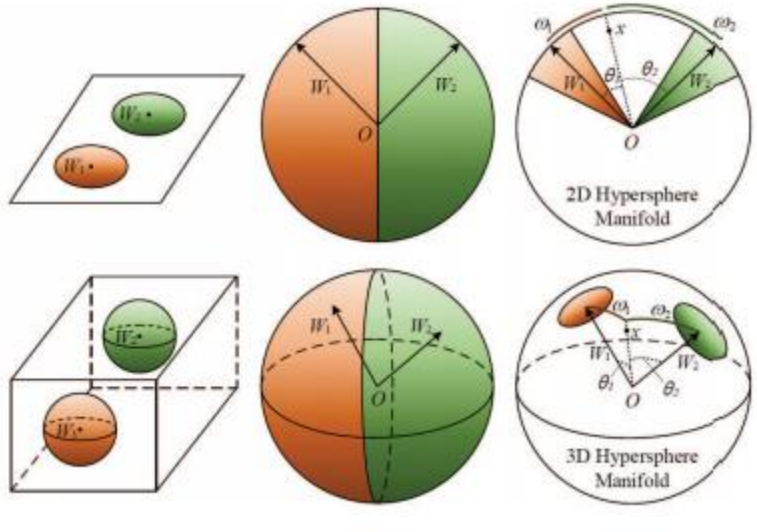
-Angular/cosine-margin-based Loss

Hypersphere Manifold with an angular margin

$$\mathcal{L}_i = -log\left(\frac{e^{\|W_{yi}\|\|x_i\|\varphi(\theta_{yi})}}{e^{\|W_{yi}\|\|x_i\|\varphi(\theta_{yi})}+\sum_{j\neq y_i} e^{\|W_{yi}\|\|x_i\|cos(\theta_j)}}\right)$$

단점: clean dataset이 아니면 성능이
Centor loss나 Softmax보다 안좋다





A-Softmax

• Loss Function

-Angular/cosine-margin-based Loss

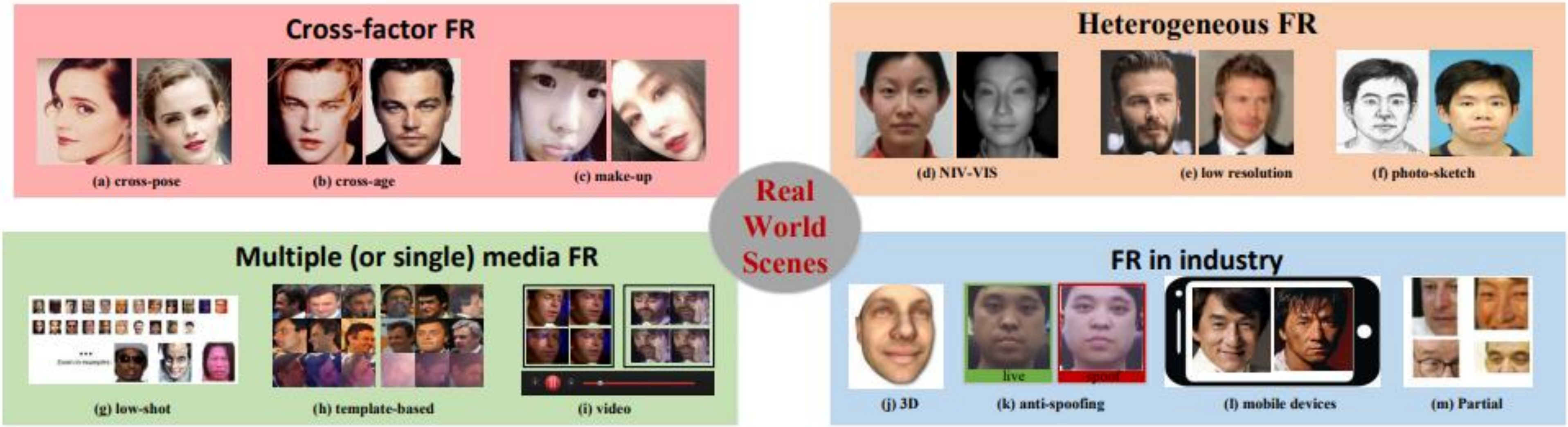
TABLE V
DECISION BOUNDARIES FOR CLASS 1 UNDER BINARY CLASSIFICATION CASE, WHERE $\hat{x}$ IS THE NORMALIZED FEATURE. [42]

Loss Functions	Decision Boundaries
Softmax	$(W_1 - W_2) x + b_1 - b_2 = 0$
L-Softmax [126]	$\|x\| (\|W_1\| cos(m\theta_1) - \|W_2\| cos(\theta_2)) > 0$
A-Softmax [125]	$\|x\| (cosm\theta_1 - cos\theta_2) = 0$
CosineFace [205]	$\hat{x} (cos\theta_1 - m - cos\theta_2) = 0$
ArcFace [42]	$\hat{x} (cos (\theta_1 + m) - cos\theta_2) = 0$

	Megaface challenge1			
	FaceScrub		FGNet	
Method	Rank1 @10^6	TPR @10^{-6}FPR	Rank1 @10^6	TPR @10^{-6}FPR
Arcface [42]	0.9836	0.9848	-	-
Cosface [205]	0.9833	0.9841	-	-
A-softmax [125]	0.9743	0.9766	-	-
Marginal loss [43]	0.8028	0.9264	0.6643	0.4370

Challenges

- Data noise
- Data bias (Age, Race, Pose…)
-



- **Reference**

https://arxiv.org/abs/1804.06655