Progress and Trends of Facial Recognition

소프트웨어학과 창경현

아주대학교
AJOU UNIVERSITY

Contents

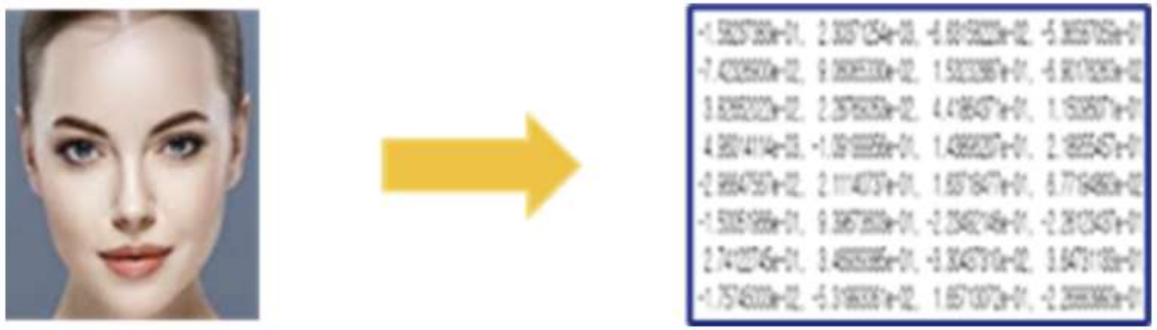
아주대학교
AJOU UNIVERSITY

1. Facial Recognition Technology Process

1. Facial Recognition Technology Process

Image Preprocessing

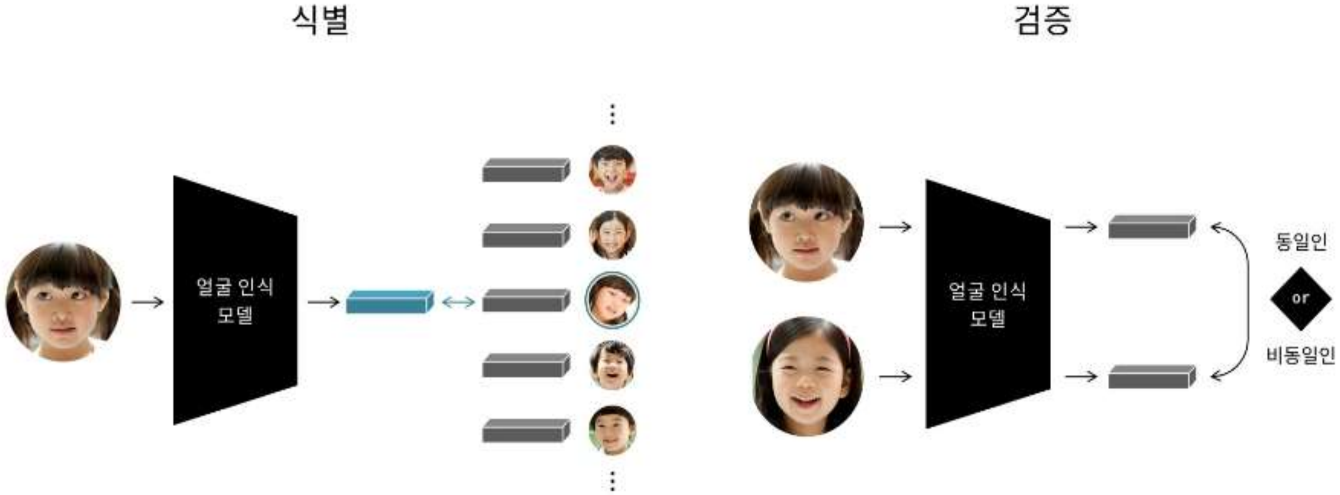
Extract facial features



Ref) https://kakaoenterprise.github.io/deepdive/200723

1. Facial Recognition Technology Process

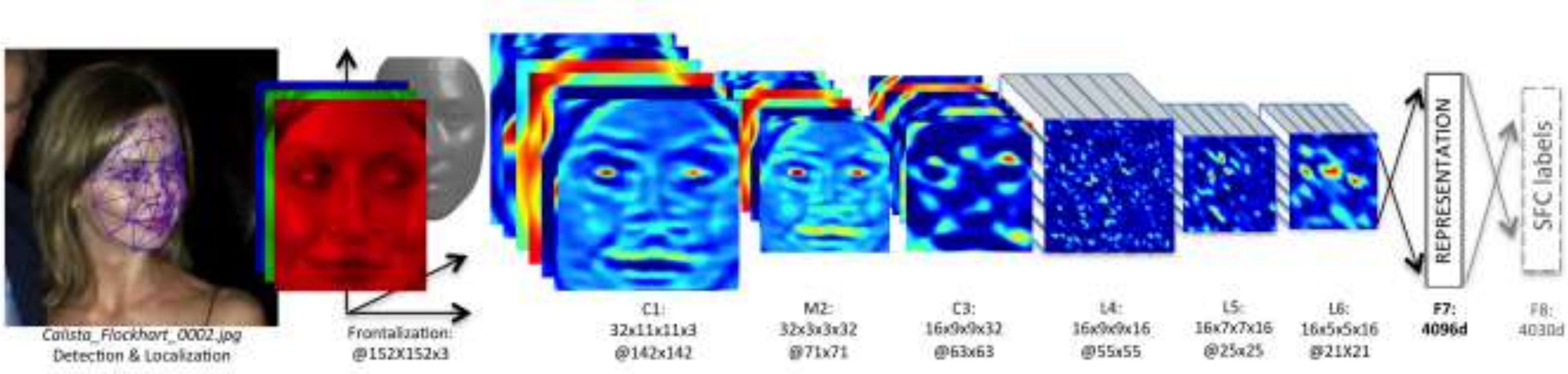
Extract facial features



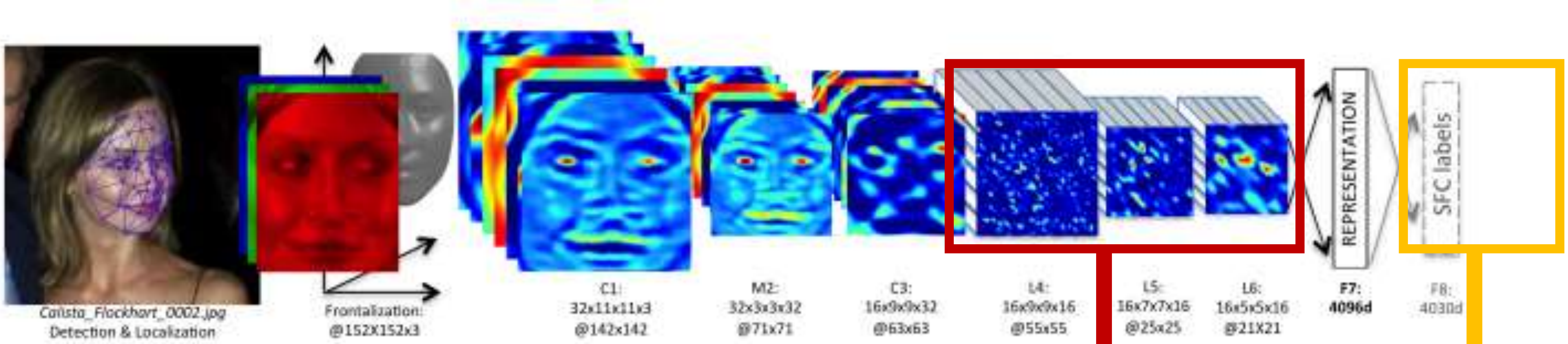
Ref) https://kakaoenterprise.github.io/deepdive/200723

2. Transition of facial recognition algorithms

2. Transition of facial recognition algorithms: DeepFace



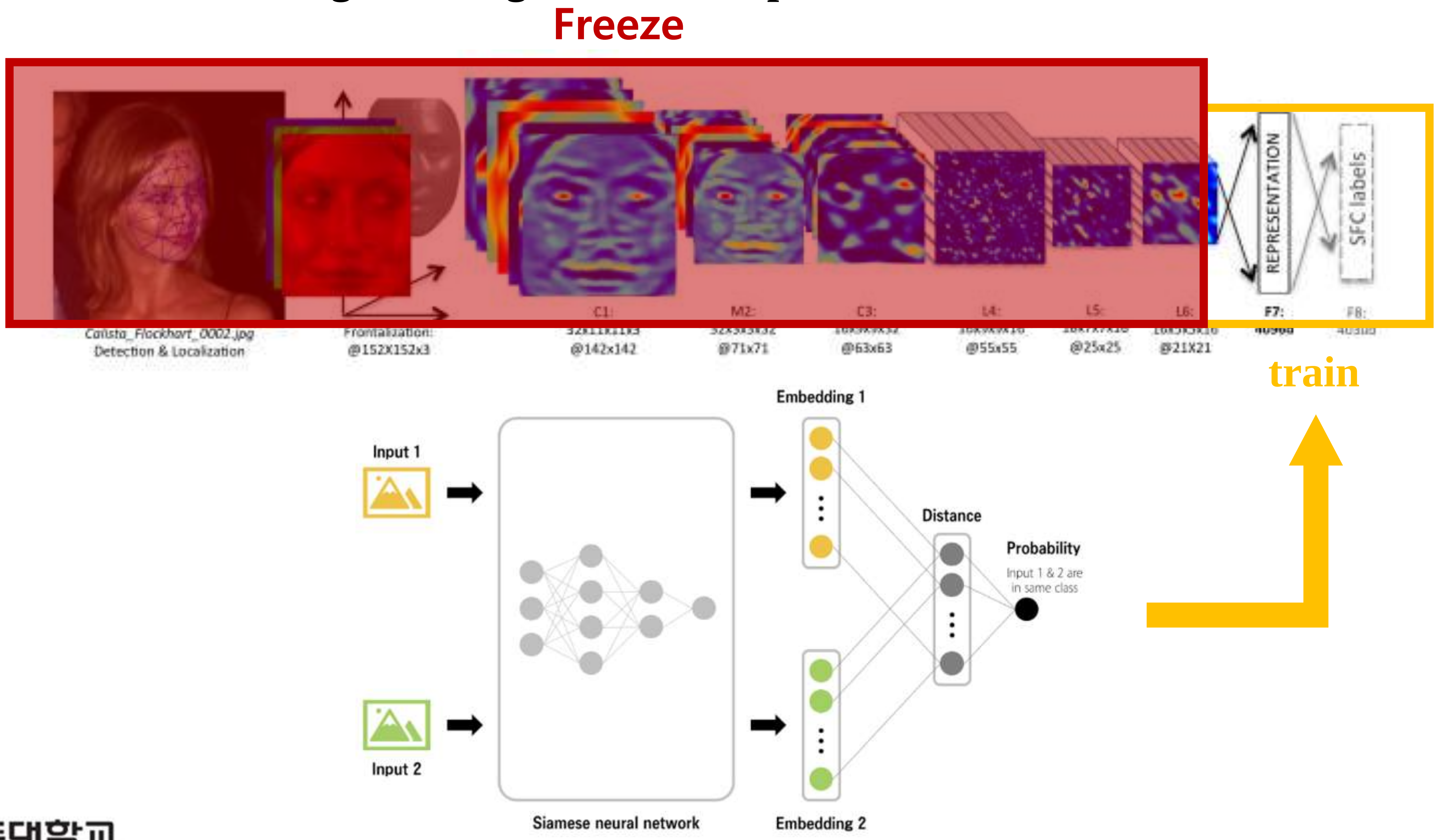
2. Transition of facial recognition algorithms: DeepFace



Local convolution

Softmax Function
classification purpose

2. Transition of facial recognition algorithms: DeepFace



Ref) https://tyami.github.io/deep learning/Siamese-neural-networks/

2. Transition of facial recognition algorithms: FaceNet

Triplet loss

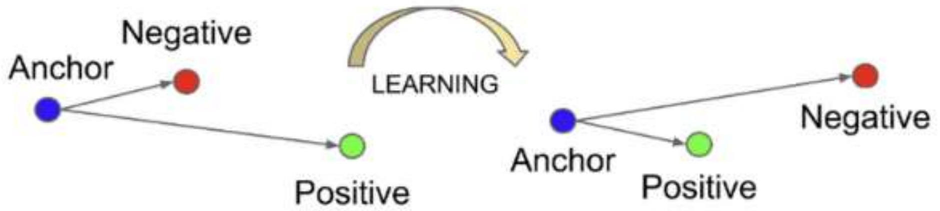


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 , \quad (1)$$

$$\forall \left(f(x_i^a), f(x_i^p), f(x_i^n)\right) \in \mathcal{T} . \quad (2)$$

2. Transition of facial recognition algorithms: FaceNet

Triplet loss

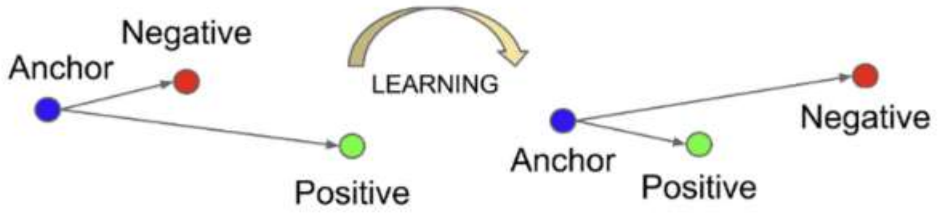


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \,, \quad (1)$$

$$\forall \left(f(x_i^a), f(x_i^p), f(x_i^n)\right) \in \mathcal{T} \,. \quad (2)$$

Triplet selection

$$\operatorname{argmax}_{x_i^p} \|f(x_i^a) - f(x_i^p)\|_2^2$$

$$\operatorname{argmin}_{x_i^n} \|f(x_i^a) - f(x_i^n)\|_2^2$$

$$\|f(x_i^a) - f(x_i^p)\|_2^2 < \|f(x_i^a) - f(x_i^n)\|_2^2$$

Angular margin based loss function
Introduction of "angle"

- **Problems with softmax loss**

1. Not enough discernment in the open-set face recognition problem where a new case may emerge

- **Problem with triplet loss**

1. In large-scale dataset, the number of triplet combinations increases explosively.
2. Semi-hard sample mining is quite difficult (it is difficult to process hard positive and negative)

2. Transition of facial recognition algorithms: SphereFace

$$p_1 = \frac{\exp(W_1^T x + b_1)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

$$p_2 = \frac{\exp(W_2^T x + b_2)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

Equation 1

$$\boldsymbol{P_1} > \boldsymbol{P_2} \rightarrow \textbf{class 1}$$

$$\boldsymbol{P_1} < \boldsymbol{P_2} \rightarrow \textbf{class 2}$$

$$\left(\mathbf{W}_1^{\mathbf{T}} - \mathbf{W}_2^{\mathbf{T}}\right)\boldsymbol{x} + (\mathbf{b}_1 - \mathbf{b}_2) = \mathbf{0}$$

2. Transition of facial recognition algorithms: SphereFace

$$p_1 = \frac{\exp(W_1^T x + b_1)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

$$p_2 = \frac{\exp(W_2^T x + b_2)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

Equation 1

$$\boldsymbol{P_1} > \boldsymbol{P_2} \rightarrow \textbf{class 1}$$

$$\boldsymbol{P_1} < \boldsymbol{P_2} \rightarrow \textbf{class 2}$$

$$\left(\mathbf{W}_1^{\mathrm{T}} - \mathbf{W}_2^{\mathrm{T}}\right)\boldsymbol{x} + (\mathbf{b}_1 - \mathbf{b}_2) = 0$$

$$\begin{aligned} L_i &= -\log\left(\frac{e^{\boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_j e^{\boldsymbol{W}_j^T \boldsymbol{x}_i + b_j}}\right) \\ &= -\log\left(\frac{e^{\|\boldsymbol{W}_{y_i}\|\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i}) + b_{y_i}}}{\sum_j e^{\|\boldsymbol{W}_j\|\|\boldsymbol{x}_i\|\cos(\theta_{j,i}) + b_j}}\right) \end{aligned}$$

$$W_{\mathrm{i}}^{\mathrm{T}} x = \| W_{\mathrm{i}}^{\mathrm{T}} \| \| x \| \cos(\theta_{\mathrm{i}})$$

$$\| \boldsymbol{x} \| \left(\boldsymbol{cos}(\boldsymbol{\theta_1}) - \boldsymbol{cos}(\boldsymbol{\theta_2})\right) = 0$$

2. Transition of facial recognition algorithms: SphereFace

$$p_1 = \frac{\exp(W_1^T x + b_1)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

$$p_2 = \frac{\exp(W_2^T x + b_2)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}$$

Equation 1

$$\boldsymbol{P_1 > P_2} \rightarrow \textbf{class 1}$$

$$\boldsymbol{P_1 < P_2} \rightarrow \textbf{class 2}$$

$$\boldsymbol{(\mathrm{W}_1^{\mathrm{T}} - \mathrm{W}_2^{\mathrm{T}})x + (\mathrm{b}_1 - \mathrm{b}_2) = 0}$$

$$L_i = -\log\left(\frac{e^{\boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_j e^{\boldsymbol{W}_j^T \boldsymbol{x}_i + b_j}}\right)$$
$$= -\log\left(\frac{e^{\|\boldsymbol{W}_{y_i}\| \|\boldsymbol{x}_i\| \cos(\theta_{y_i,i}) + b_{y_i}}}{\sum_j e^{\|\boldsymbol{W}_j\| \|\boldsymbol{x}_i\| \cos(\theta_{j,i}) + b_j}}\right)$$

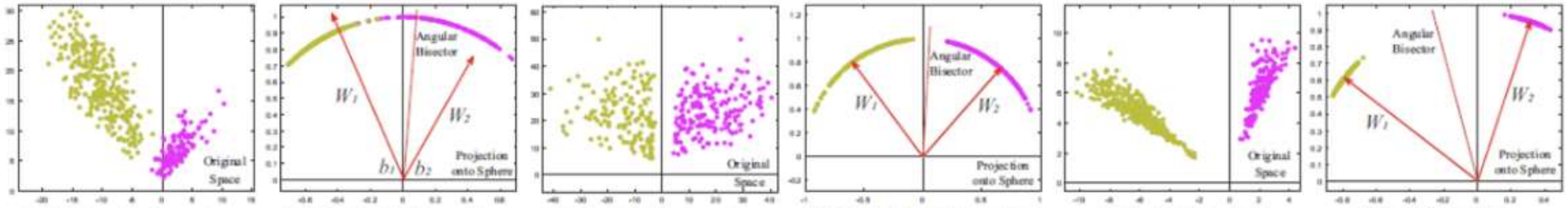
$$W_{\mathrm{i}}^{\mathrm{T}} x = \| W_{\mathrm{i}}^{\mathrm{T}} \| \| x \| \cos(\theta i)$$

$$L_{\text{modified}} = \frac{1}{N} \sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\| \cos(\theta_{y_i,i})}}{\sum_j e^{\|\boldsymbol{x}_i\| \cos(\theta_{j,i})}}\right)$$

$$L_{\text{ang}} = \frac{1}{N} \sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\| \cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\| \cos(m\theta_{y_i,i})} + \sum_{j \neq y_i} e^{\|\boldsymbol{x}_i\| \cos(\theta_{j,i})}}\right)$$

2. Transition of facial recognition algorithms: SphereFace

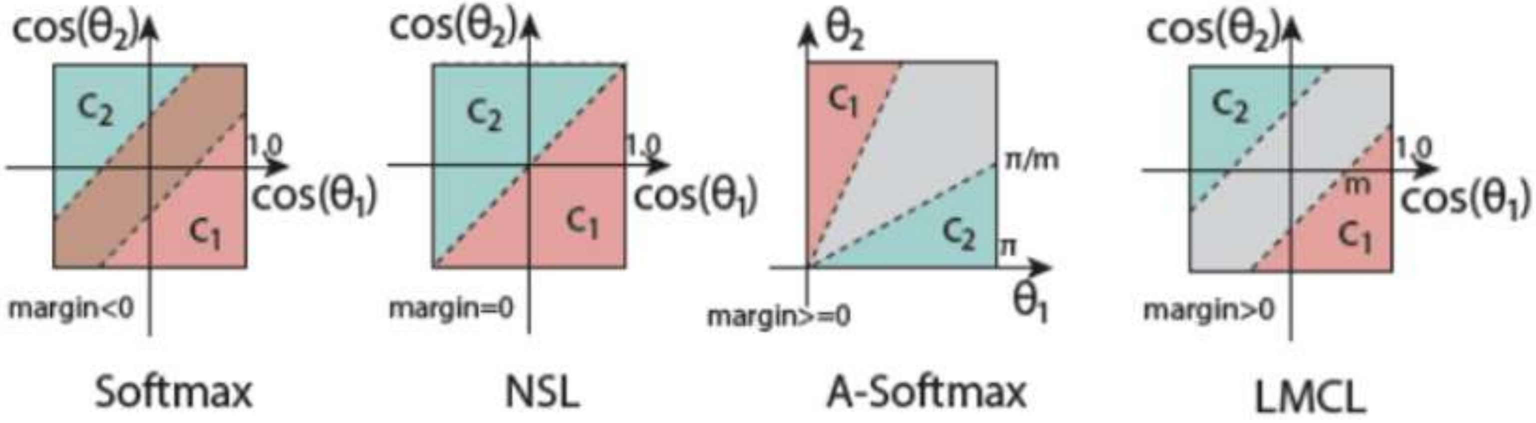
Loss Function	Decision Boundary
Softmax Loss	$(\boldsymbol{W}_1 - \boldsymbol{W}_2)\boldsymbol{x} + b_1 - b_2 = 0$
Modified Softmax Loss	$\|\boldsymbol{x}\|(\cos\theta_1 - \cos\theta_2) = 0$
A-Softmax Loss	$\|\boldsymbol{x}\|(\cos m\theta_1 - \cos\theta_2) = 0$ for class 1 $\|\boldsymbol{x}\|(\cos\theta_1 - \cos m\theta_2) = 0$ for class 2



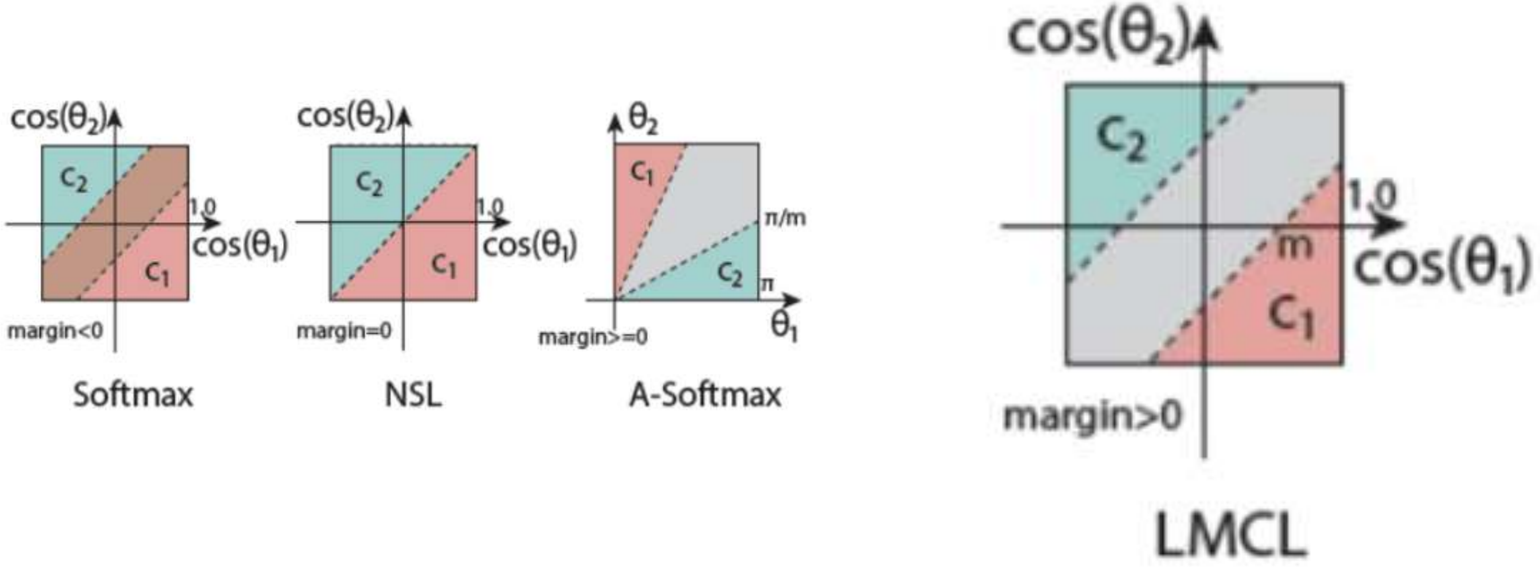
(a) Original Softmax Loss (b) Original Softmax Loss (c) Modified Softmax Loss (d) Modified Softmax Loss (e) A-Softmax Loss (f) A-Softmax Loss

Figure 2: Comparison among softmax loss, modified softmax loss and A-Softmax loss. In this toy experiment, we construct a CNN to learn 2-D features on a subset of the CASIA face dataset. In specific, we set the output dimension of FC1 layer as 2 and visualize the learned features. Yellow dots represent the first class face features, while purple dots represent the second class face features. One can see that features learned by the original softmax loss can not be classified simply via angles, while modified softmax loss can. Our A-Softmax loss can further increase the angular margin of learned features.

2. Transition of facial recognition algorithms: CosFace



2. Transition of facial recognition algorithms: CosFace

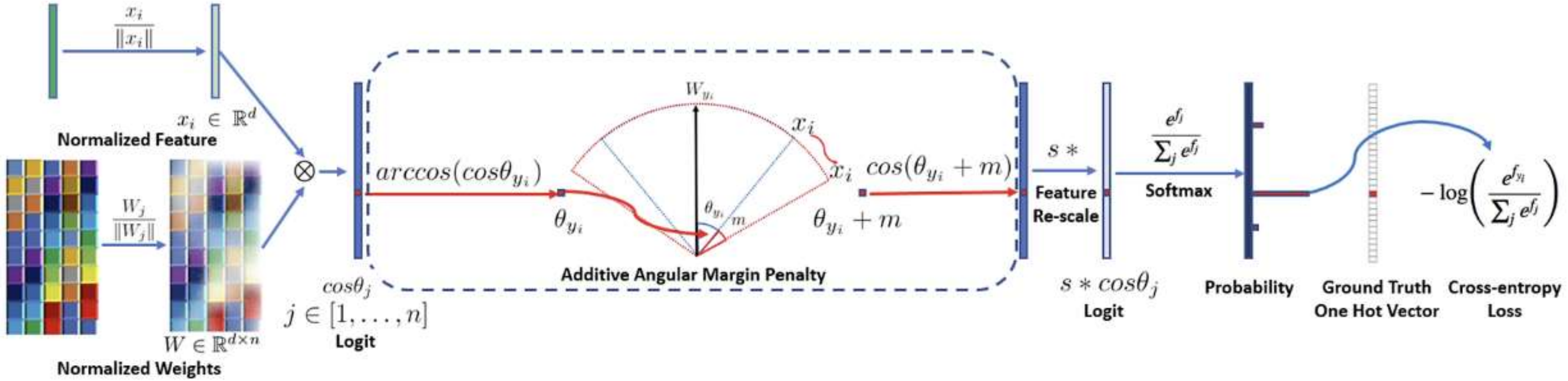


$$L_{lmc} = \frac{1}{N}\sum_{i} -\log \frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)} + \sum_{j\neq y_i} e^{s\cos(\theta_{j,i})}},$$

Equation 3

2. Transition of facial recognition algorithms: ArcFace

$$\cos(\theta) \rightarrow \cos(\theta+m)$$



Method	LFW	CALFW	CPLFW
HUMAN-Individual	97.27	82.32	81.21
HUMAN-Fusion	99.85	86.50	85.24
Center Loss [36]	98.75	85.48	77.48
SphereFace [15]	99.27	90.30	81.40
VGGFace2 [3]	99.43	90.57	84.00
MS1MV2, R100, ArcFace	**99.82**	**95.45**	**92.08**

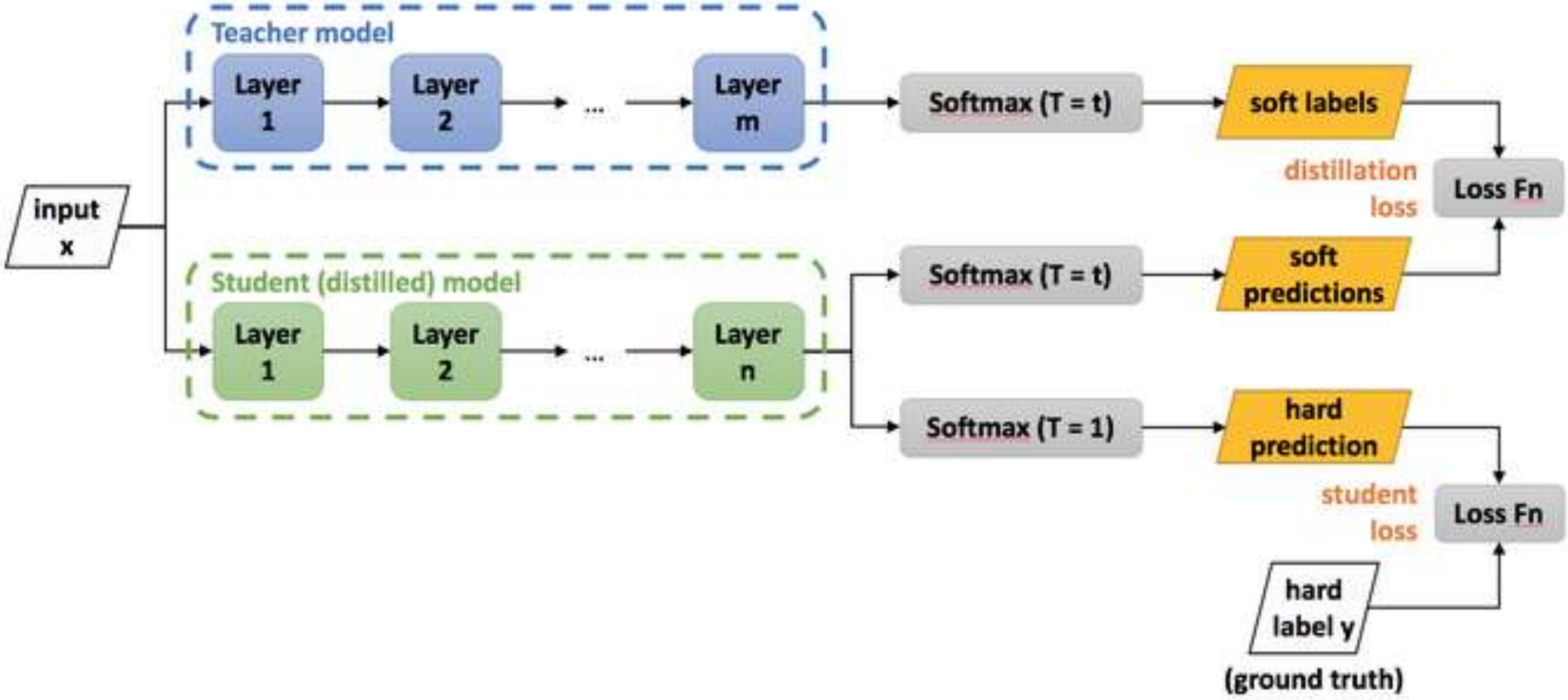
Table 5. Verification performance (%) of open-sourced face recognition models on LFW, CALFW and CPLFW.

3. Latest facial recognition trends

Evaluation-oriented Knowledge Distillation for Deep Face Recognition

3. EKD for Face Recognition

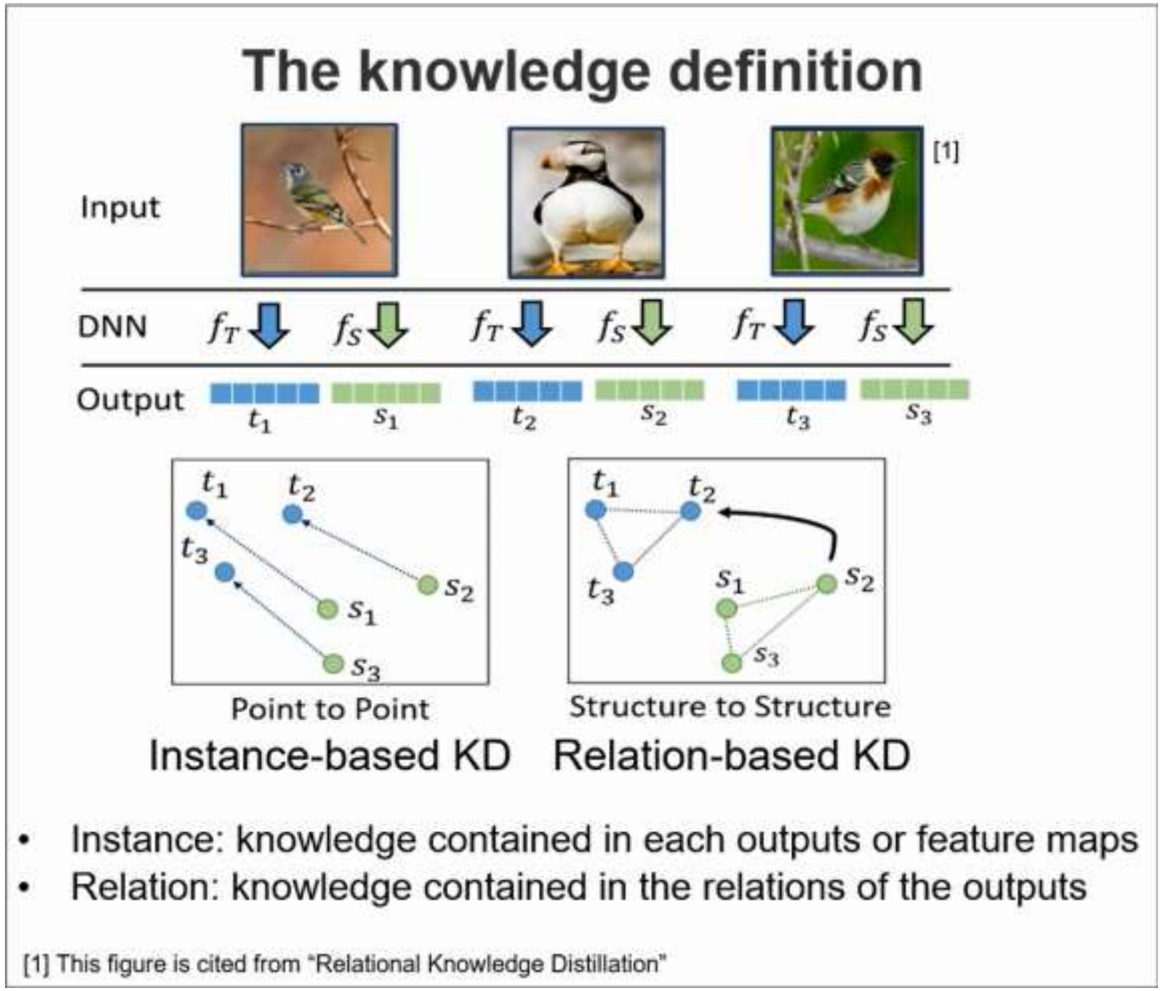
What is Knowledge distillation?



아주대학교
AJOU UNIVERSITY

Ref) https://nervanasystems.github.io/distiller/knowledge_distillation.html

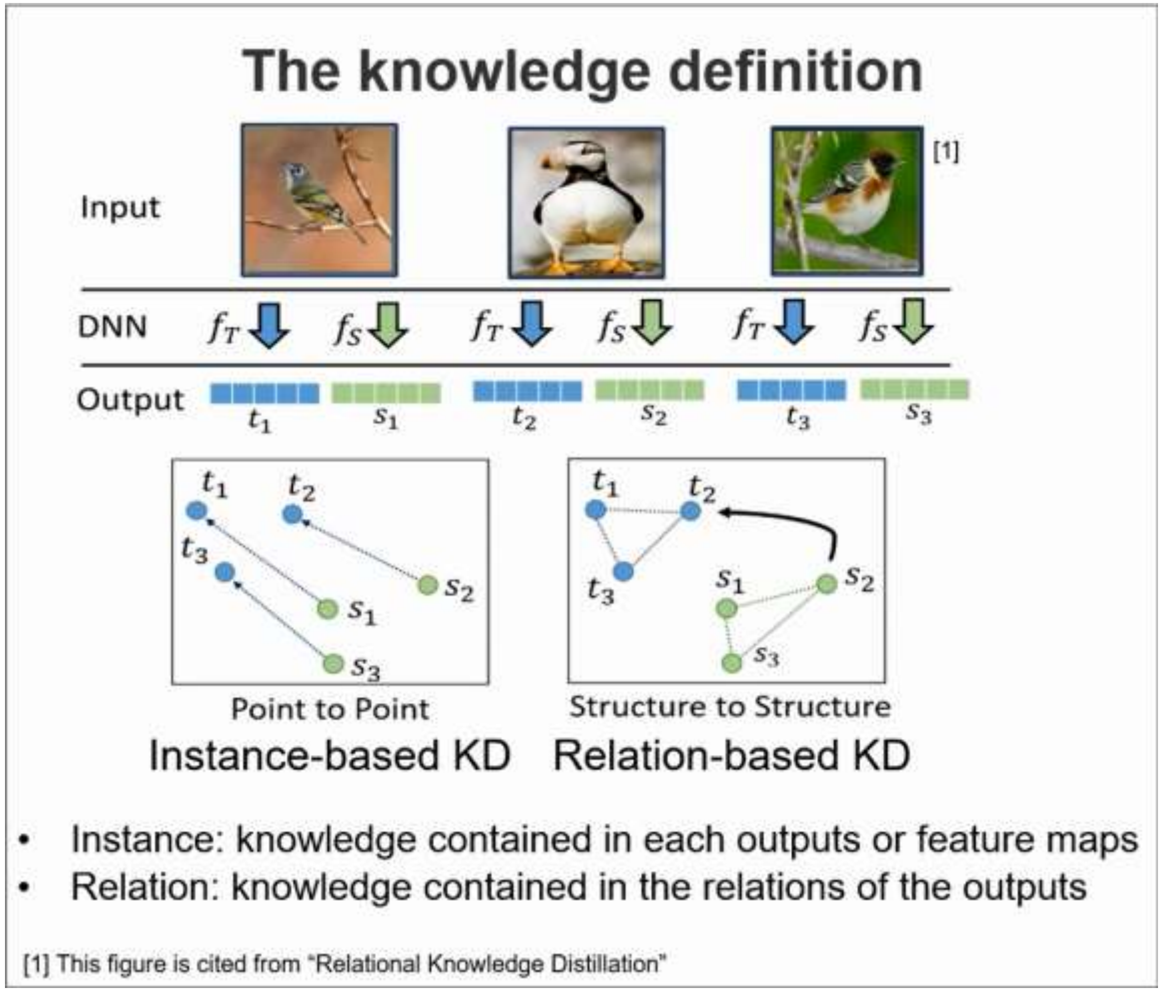
3. EKD for Face Recognition

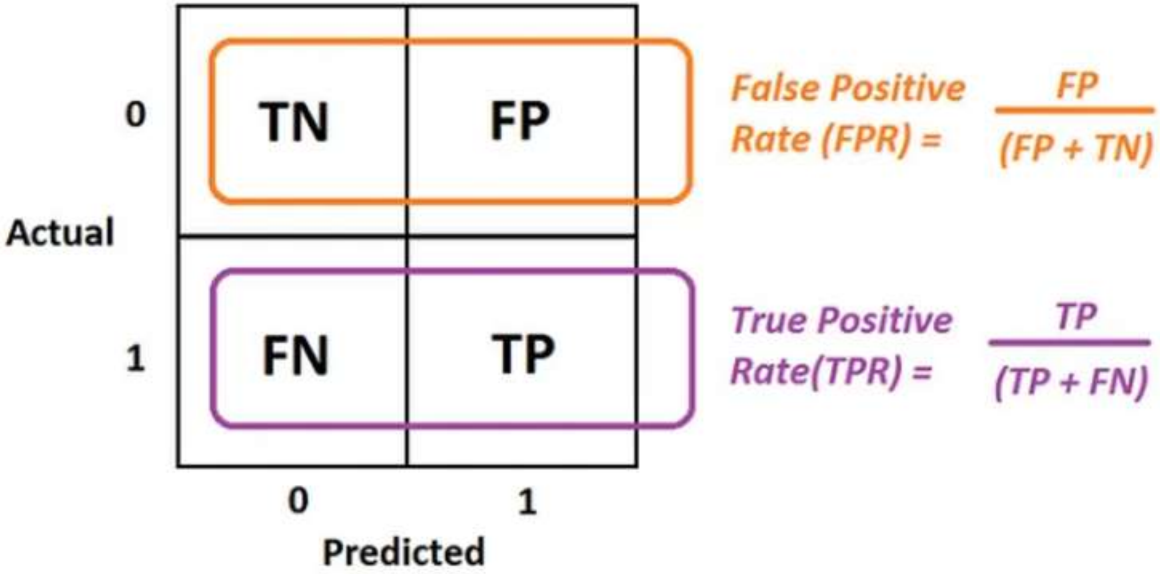


아주대학교 AJOU UNIVERSITY

Ref) https://www.youtube.com/watch?v=Vwh5HNx3uaA

3. EKD for Face Recognition



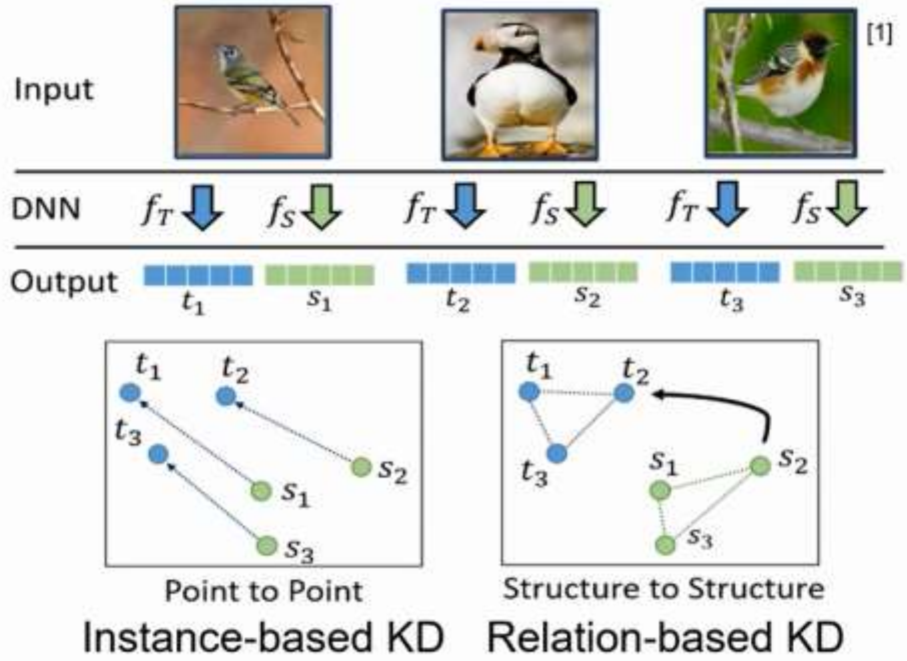


아주대학교
AJOU UNIVERSITY

Ref) https://www.youtube.com/watch?v=Vwh5HNx3uaA

3. EKD for Face Recognition

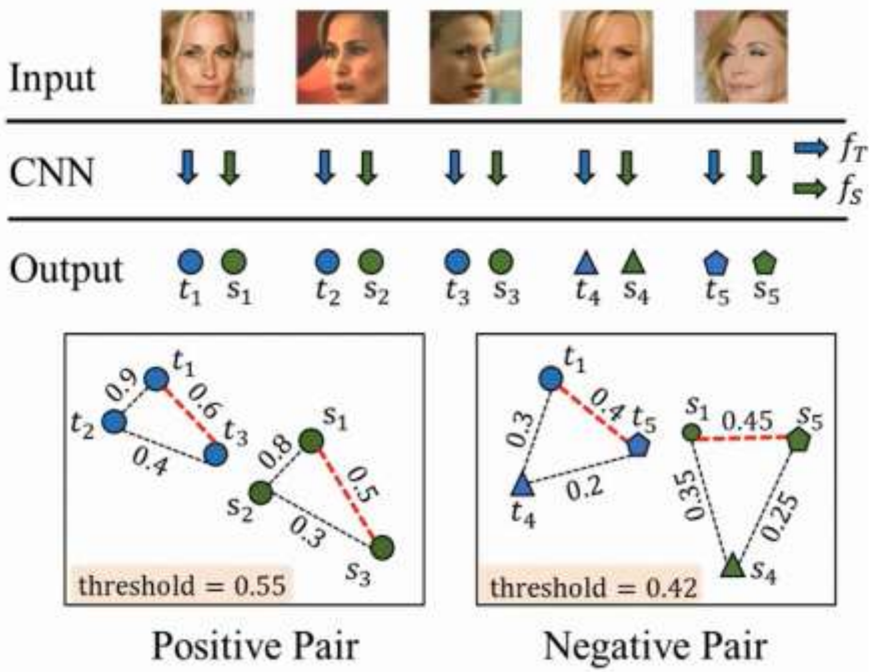
The knowledge definition



- Instance: knowledge contained in each outputs or feature maps
- Relation: knowledge contained in the relations of the outputs

[1] This figure is cited from "Relational Knowledge Distillation"

The knowledge in FR



- Based on FPR/TPR evaluation metrics, the key is **the relation between the pair similarity and a chosen threshold**, not the similarity itself.

Ref) https://www.youtube.com/watch?v=Vwh5HNx3uaA

3. EKD for Face Recognition

Experiments

Table 2. Verification comparison with SOTA methods on LFW, two pose benchmarks: CFP-FP and CPLFW, and two age benchmarks: AgeDB and CALFW.

Methods (%)	LFW	CFP-FP	CPLFW	AgeDB	CALFW
ResNet50	99.80	97.63	92.50	97.92	96.05
MobileFaceNet	99.52	91.66	87.93	95.82	95.12
FitNet (arxiv'14)	99.47	91.30	88.30	96.18	95.12
KD (NIPSW'14)	99.50	91.71	87.85	95.93	95.03
DarkRank (AAAI'18)	99.55	91.84	87.77	95.60	95.07
SP (ICCV' 19)	99.53	92.33	88.45	96.17	95.07
CCKD (ICCV' 19)	99.47	91.90	88.48	95.83	95.22
RKD (CVPR'19)	99.58	92.13	87.97	96.18	95.25
ShrinkTeaNet (arxiv'19)	99.47	91.97	88.52	96.00	94.98
TripletDistillation (ICIP'20)	99.55	93.14	88.03	95.53	94.97
MarginDistillation (arxiv'20)	**99.61**	92.01	88.03	**96.55**	95.13
EKD (Ours)	99.60	**94.33**	**89.35**	96.48	**95.37**

Teacher: Resnet50, trained by ArcFace

Student: MobileFaceNet or Resnet18

How will the model develop in the future?

3. Reference

[1] Yaniv Taigman, Ming Yang and Marc'Aurelio Ranzato, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, *In IEEE Conf. on CVPR*, 2014

[2] Florian Schroff, Dmitry Kalenichenko and James Philbin, FaceNet: A Unified Embedding for Face Recognition and Clustering, *CVPR*, 2015

[3] Weiyang Liu 외 5인, SphereFace: Deep Hypersphere Embedding for Face Recognition, *CVPR*, 2017

[4] Hao Wang 외 7인, CosFace: Large Margin Cosine Loss for Deep Face Recognition, *CVPR*, 2018

[5] Jiankang Deng 외 5인, ArcFace: Additive Angular Margin Loss for Deep Face Recognition, *CVPR*, 2018

[6] Yuge Huang 외 3인, Evaluation-oriented Knowledge Distillation for Deep Face Recognition, *CVPR*, 2022

Thank you

아주대학교
AJOU UNIVERSITY