Face Recognition at a Distance

Prof. Xiaoming Liu

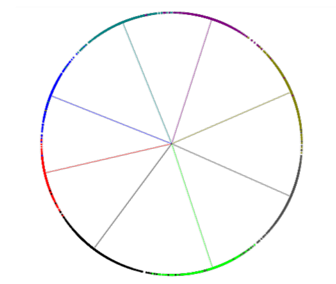
Computer Vision Lab
Department of Computer Science and Engineering
Michigan State University

Tremendous Research Progress

DR-GAN

Age synthesis

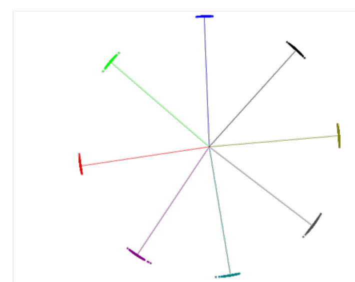
Disguised faces



Softmax

SphereFace

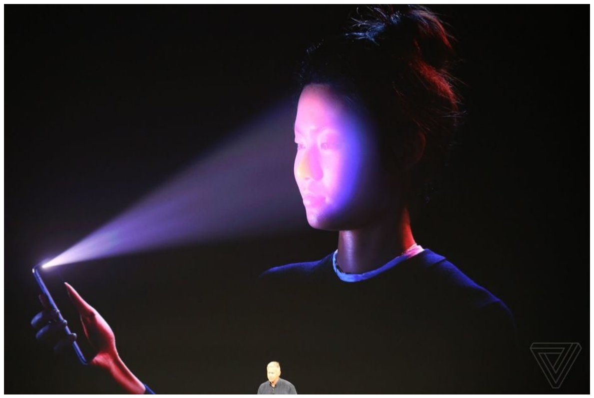
CosFace



ArcFace

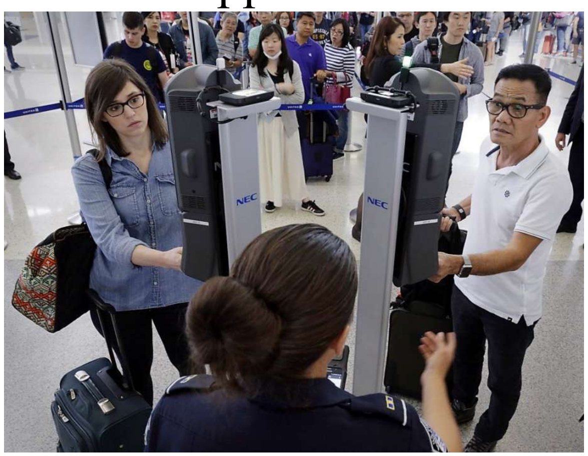
L. Tran et. al., Representation Learning by Rotating Your Faces, PAMI, 2018
H. Yang et. al., Learning Continuous Face Age Progression: A Pyramid of GANs, PAMI, 2019.
M. Singh et. al., Recognizing disguised faces in the wild. *IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019.*
J. Deng et. al., ArcFace: Additive Angular Margin Loss for Deep Face Recognition, CVPR, 2019.

Successful Applications



Apple

Alipay



Boarding in Airports

Entrance to Beijing University

Overview

Objective: Recognize individuals from a video stream captured at a distance and altitude.

Modality: Face, gait and body

Biometrics

Outline:

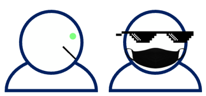
- Generic matcher: AdaFace (CVPR 2022)
- Domain adaption: CFSM (ECCV 2022)
- Video-based recognition: CAFace (NeurIPS 2022)
- 3D body matching (Under review)

AdaFace: Quality Adaptive Margin for Face Recognition

Minchul Kim, Anil K. Jain, Xiaoming Liu
CVPR 2022

Problem Definition

Training Datasets have Varying **Qualities**

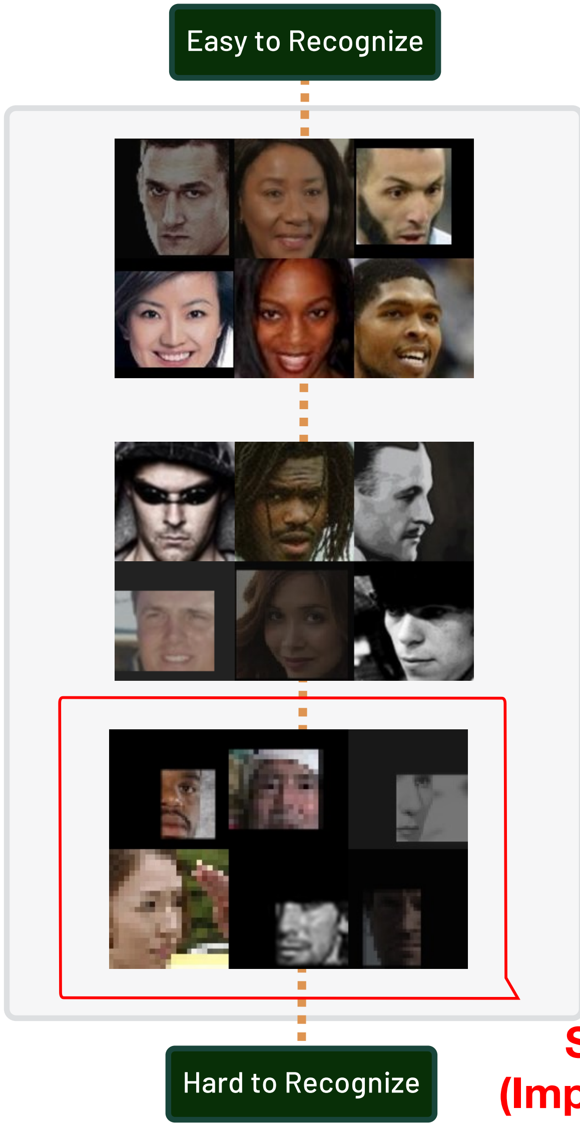


Pose and Occlusion

Faces that are front facing and free of occlusions such as hands or sunglasses are identifiable.

Landmark & Key points

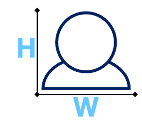
Images with visible and detectable facial landmarks are identifiable.



Source of Problem (Impossible to recognize)

Blur

Subject's distance, camera setting and other environmental factors cause the image to be blurred.



Size

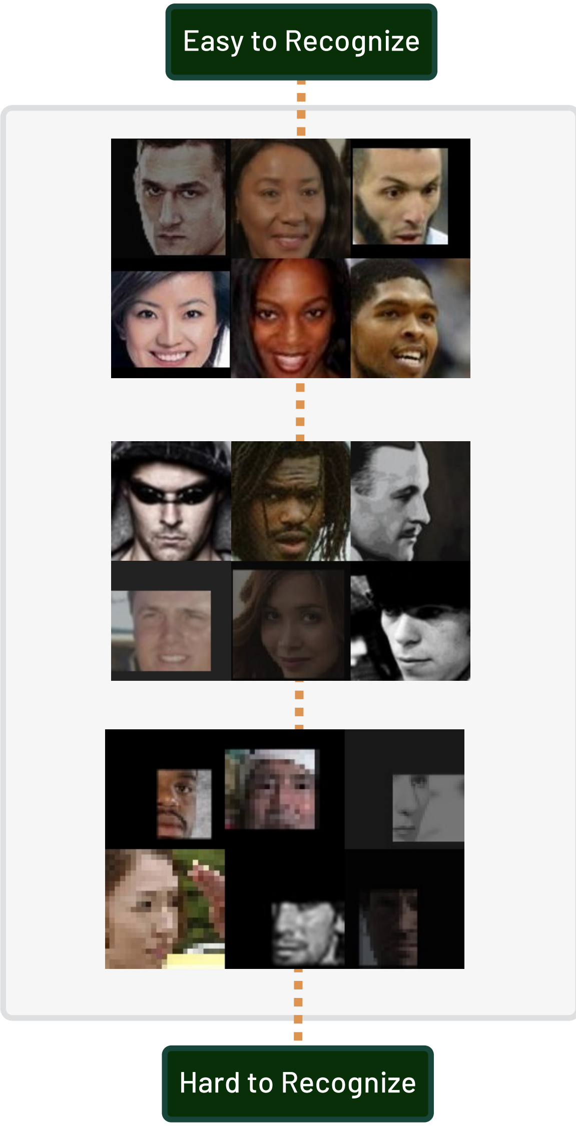
Too low image resolution causes the subject to be unidentifiable.

Illumination

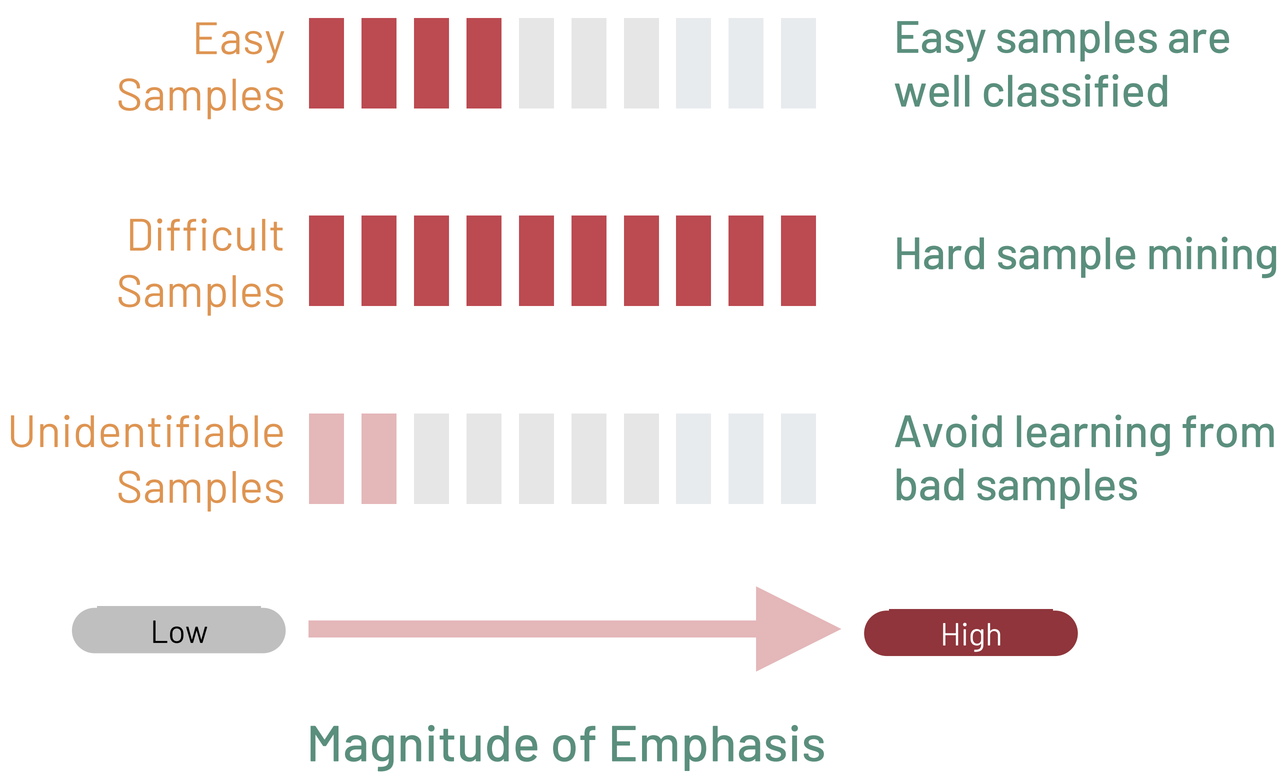
Too dark or too bright images cause the subject to be unidentifiable.

Training dataset **without identifiable traits** can be equivalent to **noisy label samples**

Motivation



Adaptive Sample Emphasis During Training



Motivation

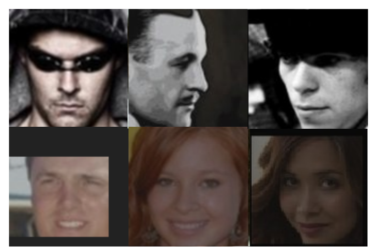
One More Way to Look at an Image

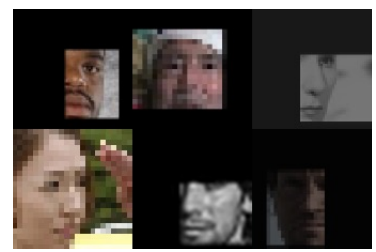
High Quality

Quality

Low Quality

Hypothesis:
Difficulty distribution will be different based on image quality.





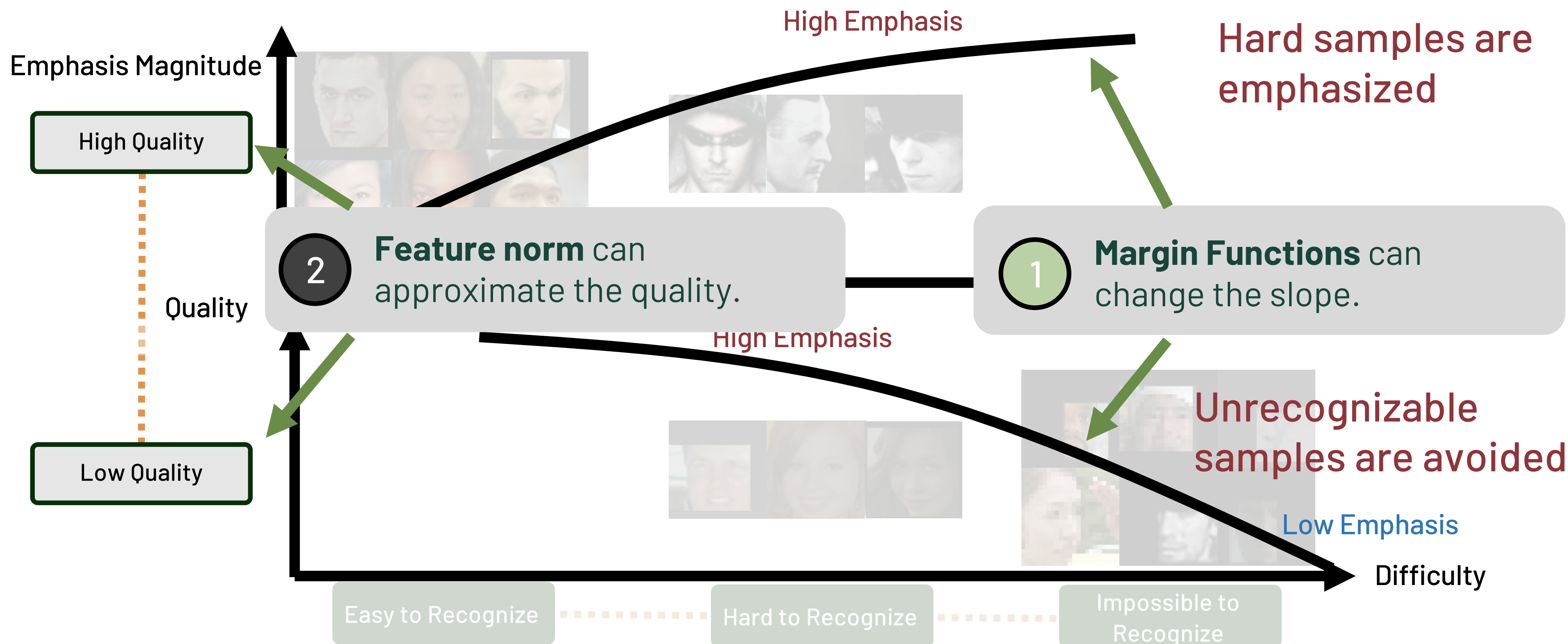
Easy to Recognize

Hard to Recognize

Impossible to Recognize

Difficulty

Our Findings and Methods



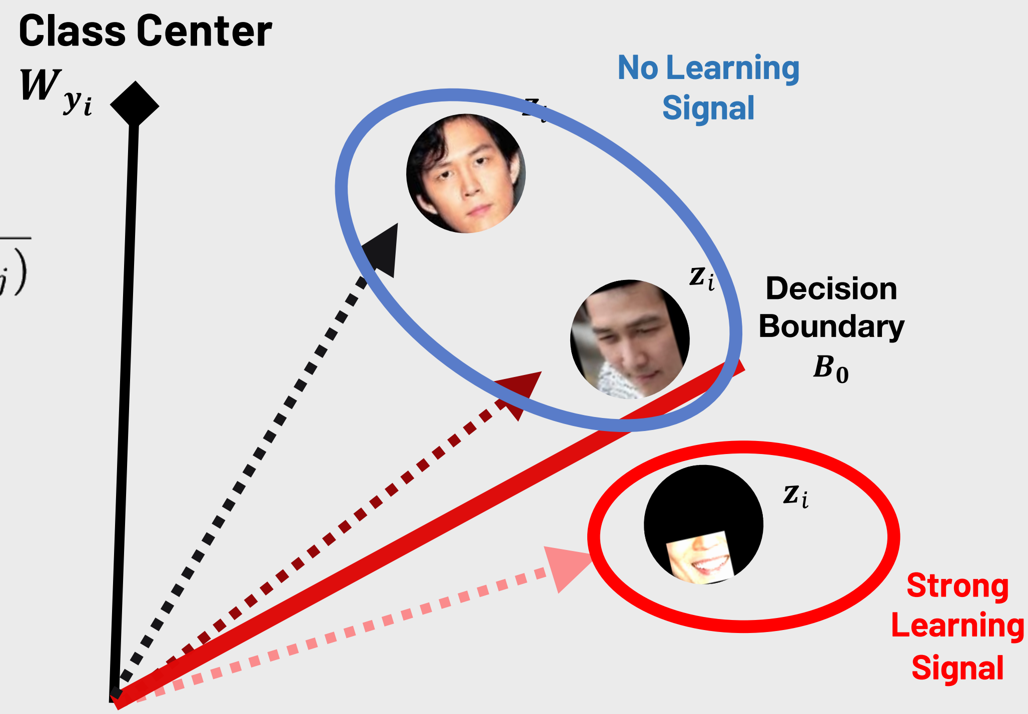
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

$$\mathcal{L} = -\log \frac{\exp(f(\theta_{y_i}, m))}{\exp(f(\theta_{y_i}, m)) + \sum_{j \neq y_i}^{n} \exp(s \cos \theta_j)}$$

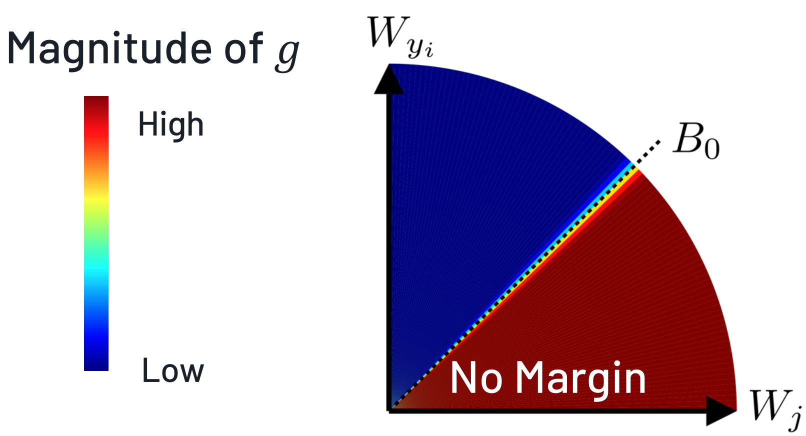
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos \theta_{y_i}) - m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$

$$f(\cdot)_{\text{Angular}} = \begin{cases} s \cos(\theta_{y_i} + m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$



Plot of Gradient Scaling Term

$$\frac{\partial \mathcal{L}_{\mathrm{CE}}}{\partial \boldsymbol{x}_i} = \sum_{k=1}^{C} \left(P_k^{(i)} - \mathbb{1}(y_i = k) \right) \frac{\partial f(\cos \theta_k)}{\partial \cos \theta_k} \frac{\partial \cos \theta_k}{\partial \boldsymbol{x}_i}.$$



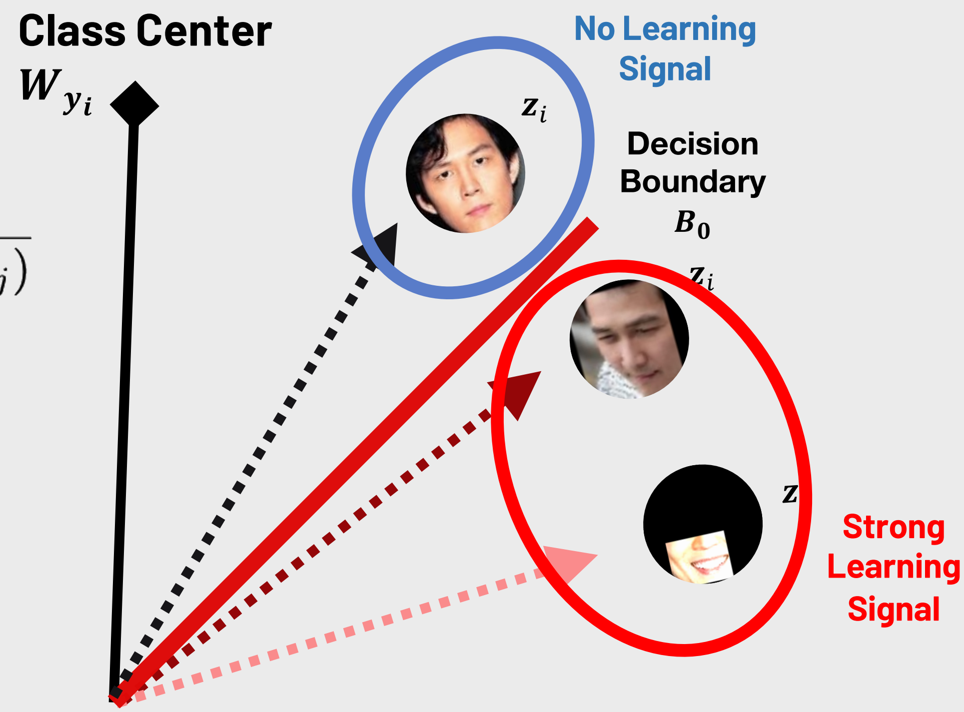
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

$$\mathcal{L} = -\log \frac{\exp(f(\theta_{y_i}, m))}{\exp(f(\theta_{y_i}, m)) + \sum_{j \neq y_i}^{n} \exp(s \cos \theta_j)}$$

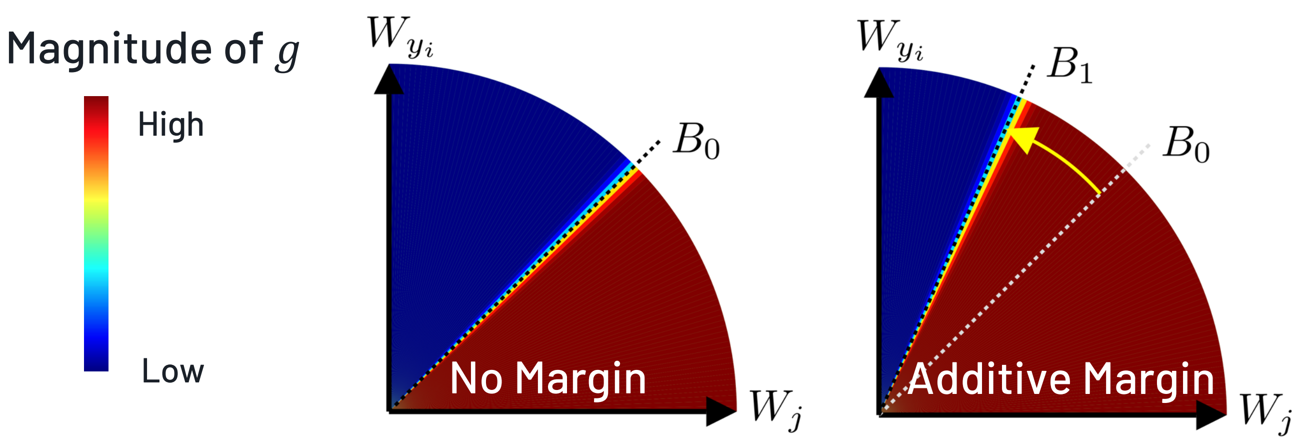
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos \theta_{y_i}) - m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$

$$f(\cdot)_{\text{Angular}} = \begin{cases} s \cos(\theta_{y_i} + m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$



Plot of Gradient Scaling Term

$$\frac{\partial \mathcal{L}_{\text{CE}}}{\partial \boldsymbol{x}_i} = \sum_{k=1}^{C} \left(P_k^{(i)} - \mathbb{1}(y_i = k) \right) \frac{\partial f(\cos \theta_k)}{\partial \cos \theta_k} \frac{\partial \cos \theta_k}{\partial \boldsymbol{x}_i}.$$



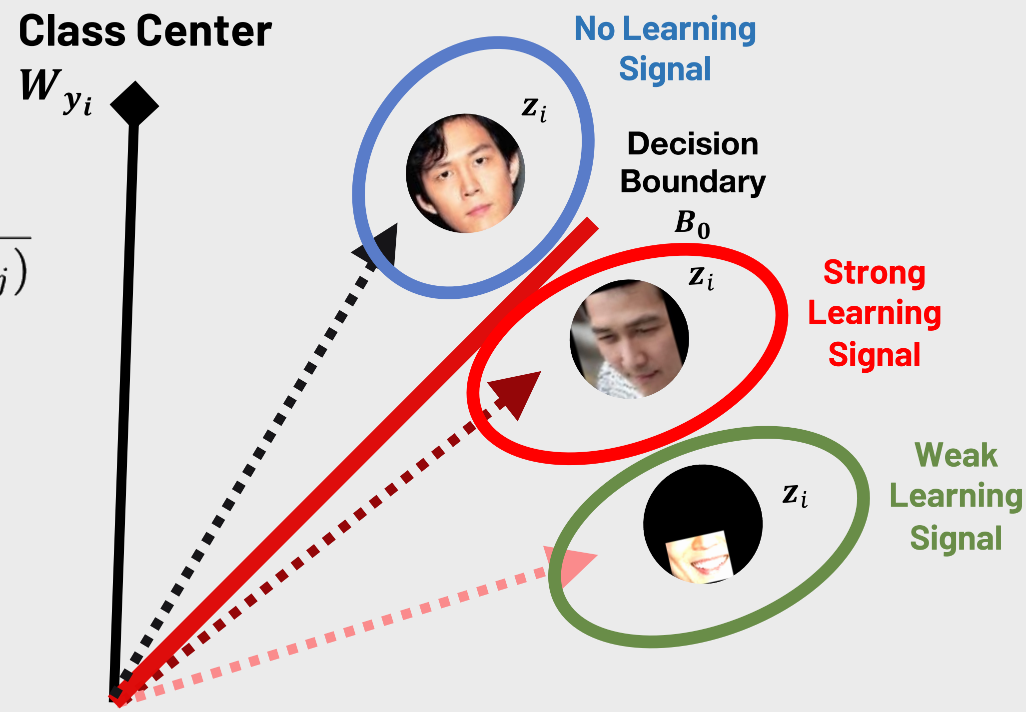
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

$$\mathcal{L} = -\log \frac{\exp(f(\theta_{y_i}, m))}{\exp(f(\theta_{y_i}, m)) + \sum_{j \neq y_i}^{n} \exp(s \cos \theta_j)}$$

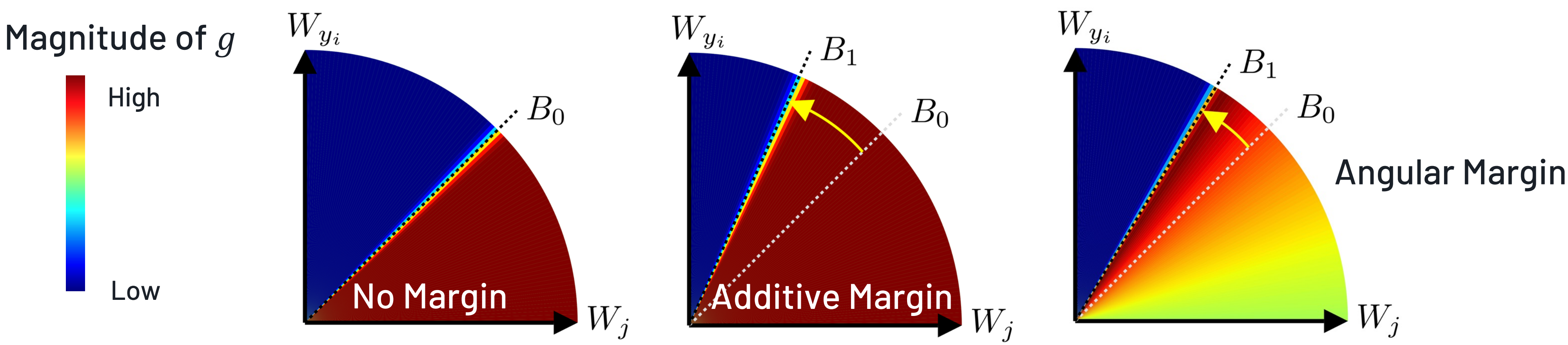
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos \theta_{y_i}) - m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$

$$f(\cdot)_{\text{Angular}} = \begin{cases} s \cos(\theta_{y_i} + m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$



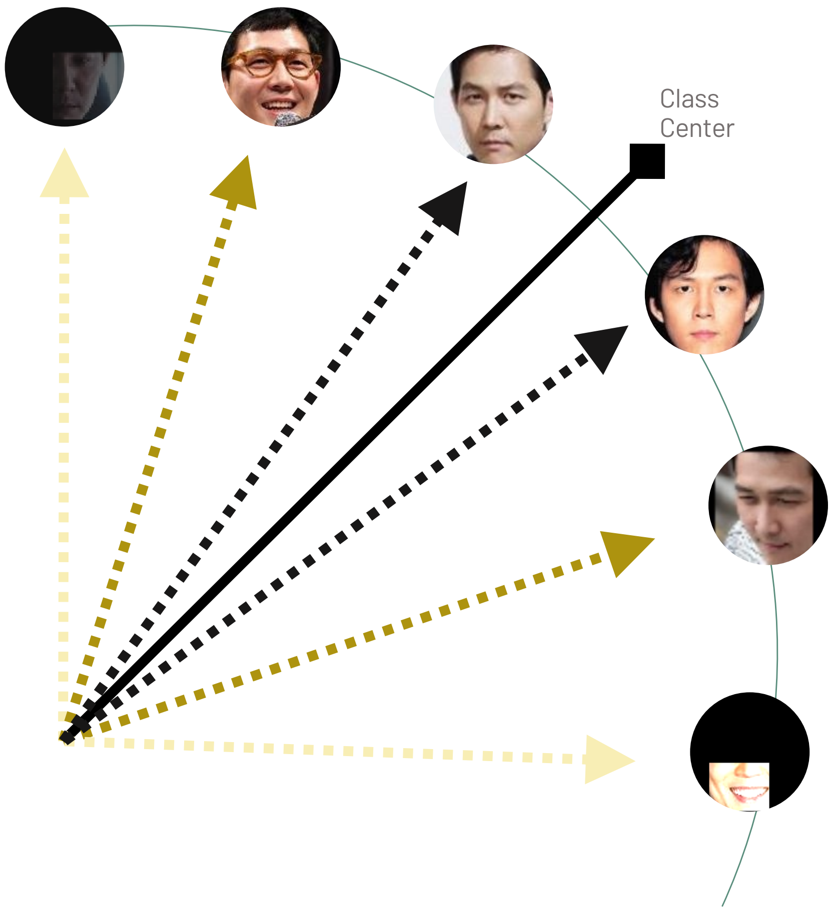
Plot of Gradient Scaling Term

$$\frac{\partial \mathcal{L}_{\text{CE}}}{\partial \boldsymbol{x}_i} = \sum_{k=1}^{C} \left(P_k^{(i)} - \mathbb{1}(y_i = k) \right) \frac{\partial f(\cos \theta_k)}{\partial \cos \theta_k} \frac{\partial \cos \theta_k}{\partial \boldsymbol{x}_i}.$$



Method

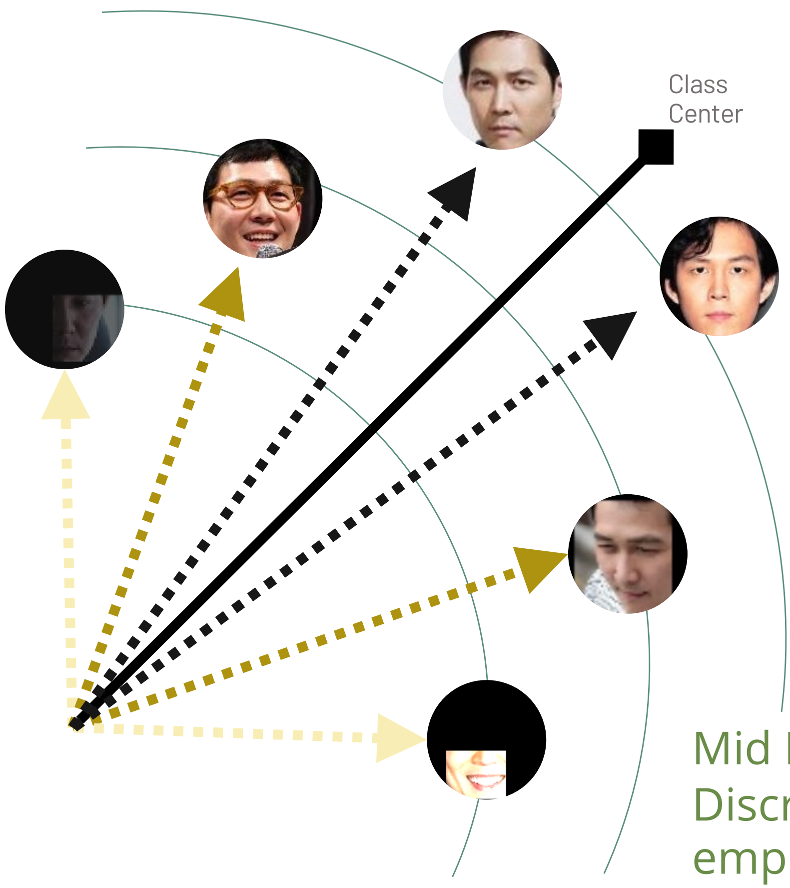
Feature Space



Previous works apply
same margin for all samples

Unit Sphere Representation

Method



AdaFace
Adaptive Margin

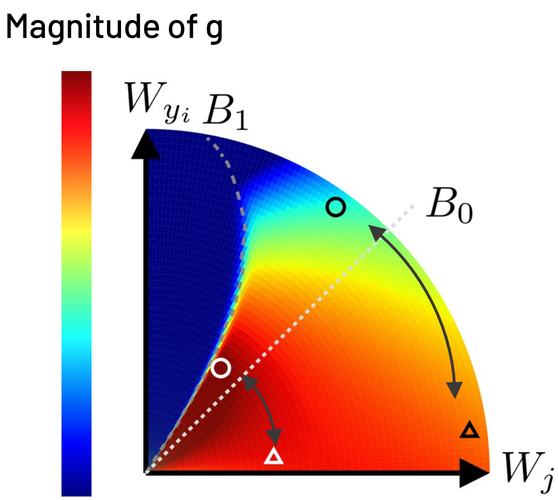
High Norm = Negative Angular Margin
De-emphasize trivial samples

Mid Norm = Additive Margin
Discriminative feature, equal emphasis.

Low Norm = Positive Angular Margin
De-emphasize unrecognizable images

How do we emphasize different samples?

AdaFace Objective



$$\mathcal{L} = -\log \frac{\exp(f(\theta_{y_i}, m))}{\exp(f(\theta_{y_i}, m)) + \sum_{j \neq y_i}^{n} \exp(s \cos \theta_j)}$$

$$f(\theta_j, m)_{\text{AdaFace}} = \begin{cases} s \cos(\theta_j + g_{\text{angle}}) - g_{\text{add}} & j = y_i \\ s \cos \theta_j & j \neq y_i \end{cases}$$

$$g_{\text{angle}} = -m \cdot \widehat{\|z_i\|}, \quad g_{\text{add}} = m \cdot \widehat{\|z_i\|} + m.$$

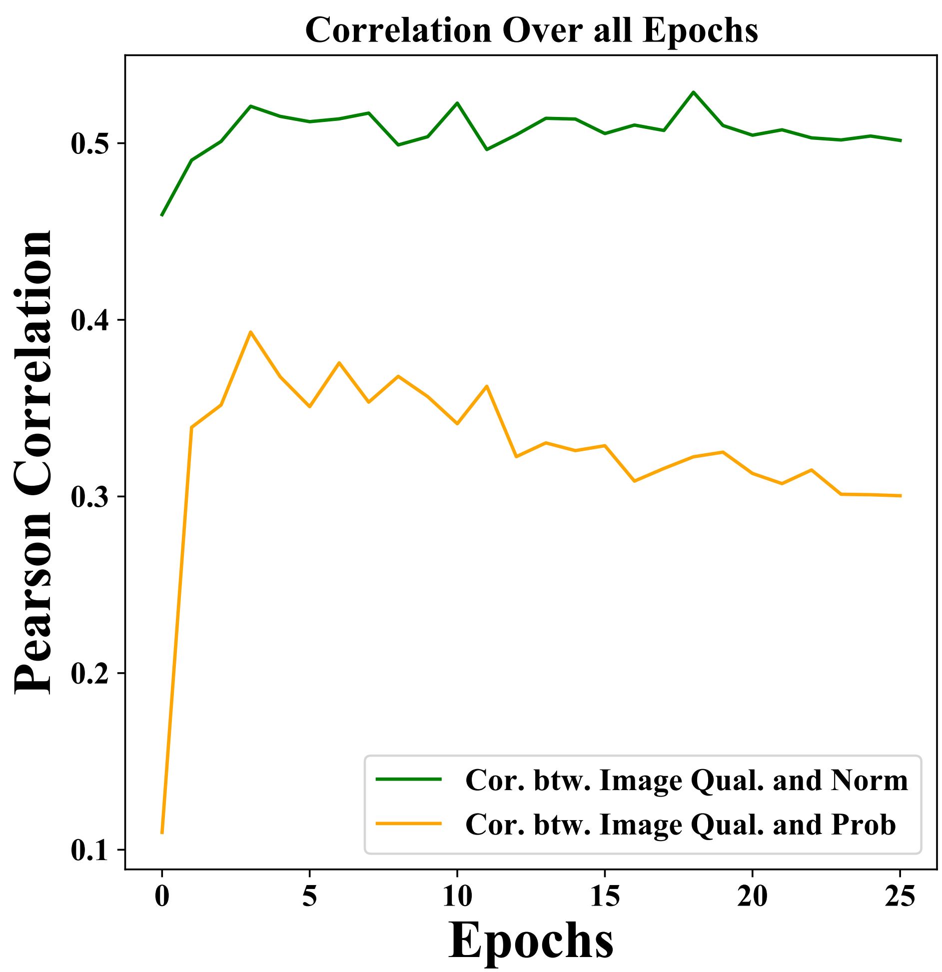
$$\widehat{\|z_i\|} = \left\lfloor \frac{\|z_i\| - \mu_z}{\sigma_z / h} \right\rceil_{-1}^{1}$$

○ Easier Sample, Low Norm
△ Harder Sample, Low Norm

○ Easier Sample, High Norm
△ Harder Sample, High Norm

Combine different margin functions adaptively
to emphasize samples of different difficulty based on the image quality.

Relationship between IQ and Feature Norm



- Image Quality: calculated with BRISQUE algorithm.
- The correlation between feature norm and the image quality exists from the early stage of training

We use feature norm as a proxy for the quality and use it to change the margin.

Performance in High Quality Datasets

Best of: CosFace (CVPR2018), ArcFace (CVPR2019), CurricularFace (CVPR2020), MagFace (CVPR2021), SCF (CVPR2021)

Best of	Dataset Names	AdaFace
99.83%	LFW	99.82%
98.46%	CFP-FP	**98.49%**
93.16%	CPLFW	**93.53%**
98.32%	AgeDB	98.05%
96.20%	CALFW	96.08%
97.16%	AVG	**97.19% (+0.03)**

Metric: 1:1 Verification Accuracy

Sample Images in high quality datasets

Performance in Mixed Quality Datasets

Dataset Names

Best of	Dataset	AdaFace
94.80%	IJB-B	95.67% (+0.87)
96.37%	IJB-C	96.89% (+0.52)

Metric: TAR@FAR=0.01%

Best of

- CosFace CVPR2018
- ArcFace CVPR2019
- CurricularFace CVPR2020
- MagFace CVPR2021
- SCF CVPR2021

AdaFace

Sample Images in mixed quality datasets

Performance

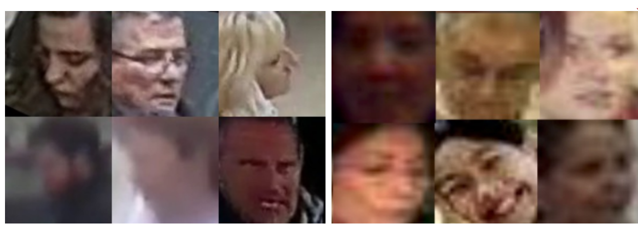
Performance in Low Quality Datasets

Best of

- PFE CVPR2018
- ArcFace CVPR2019
- CurricularFace CVPR2020
- URL CVPR2020

Best of	Dataset Names	AdaFace
62.34%	IJB-S Surveillance-to-Single	65.26% (+2.83)
63.81%	IJB-S Surveillance-to-Booking	66.27% (+2.46)
19.54%	IJB-S Surveillance-to-Surveillance	23.74% (+4.20)
63.89%	TinyFace	68.21% (+4.32)

Metric: TPIR@FPIR=0.01%



Sample Images in low quality datasets

Controllable and Guided Face Synthesis for Unconstrained Face Recognition

Feng Liu, Minchul Kim, Anil Jain, and Xiaoming Liu

ECCV 2022

Unconstrained Face Recognition

- Domain gap between the semi-constrained training datasets and unconstrained testing scenarios.

Large-scale Training datasets

VGG2. WebFace

Semi-constrained

Faces collected from the web.

Testing scenarios

LFW

Semi-constrained

Accuracy:99.83%

IJB-S

Unconstrained

Accuracy<70%

- Potential solution

Source domain

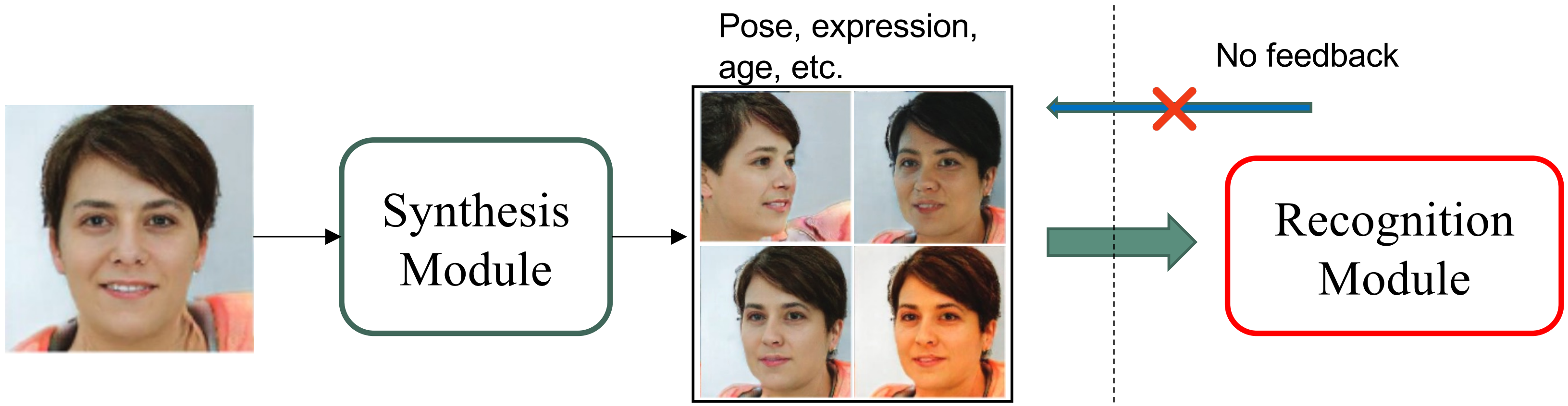
Target unconstrained domain

- Low resolution
- Motion blurring
- Bad illumination
- Turbulence effect

……

Motivation

- Previous synthesis models: ***limited face properties***; offline and ***blind*** data augmentation.

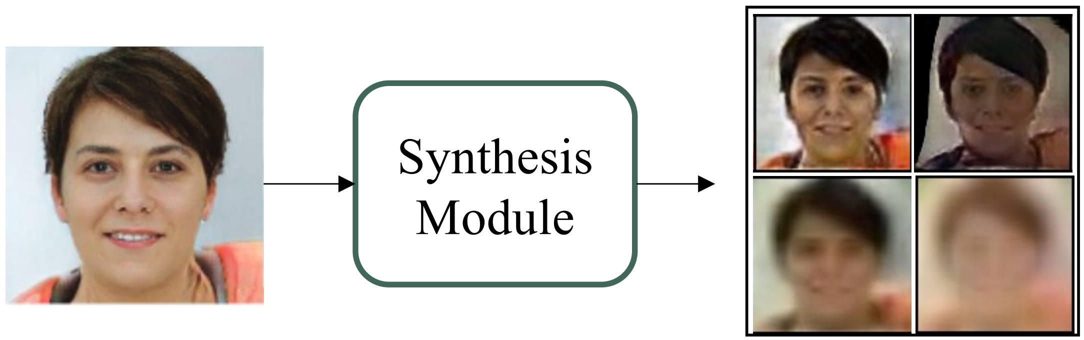


Motivation

- Previous synthesis models: ***limited face properties***; offline and ***blind*** data augmentation.
- Facial properties should be generalizable to the challenging unconstrained testing scenarios.

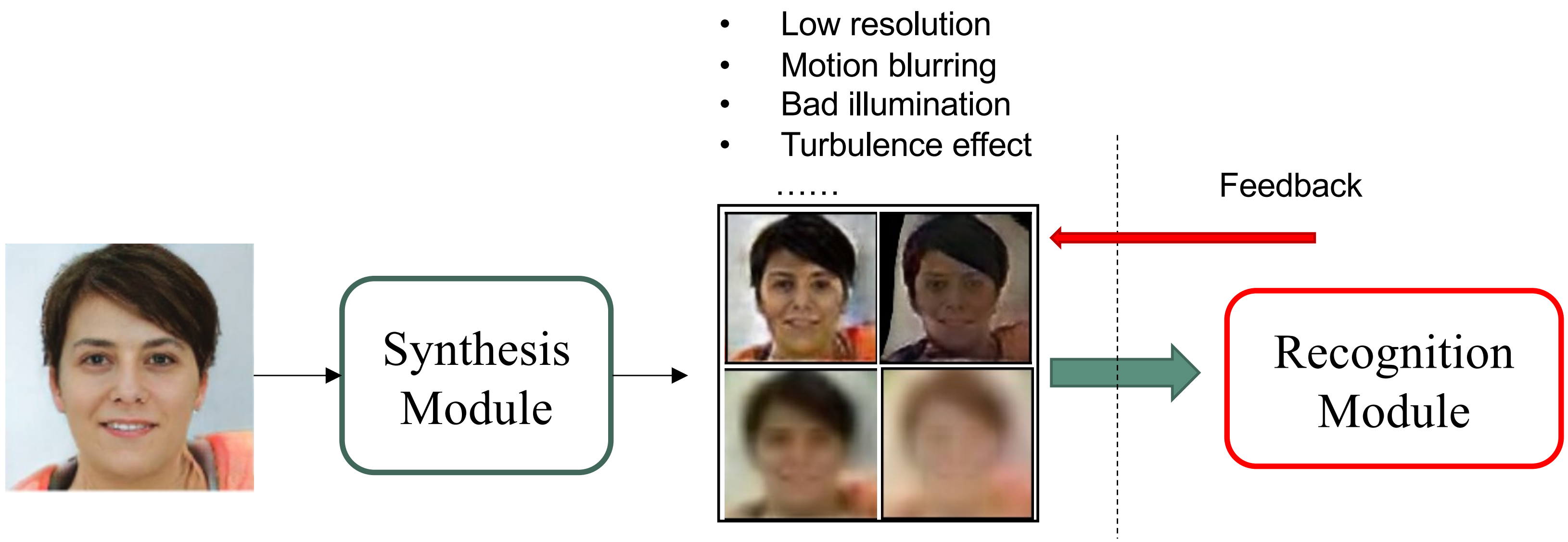
- Low resolution
- Motion blurring
- Bad illumination
- Turbulence effect

......



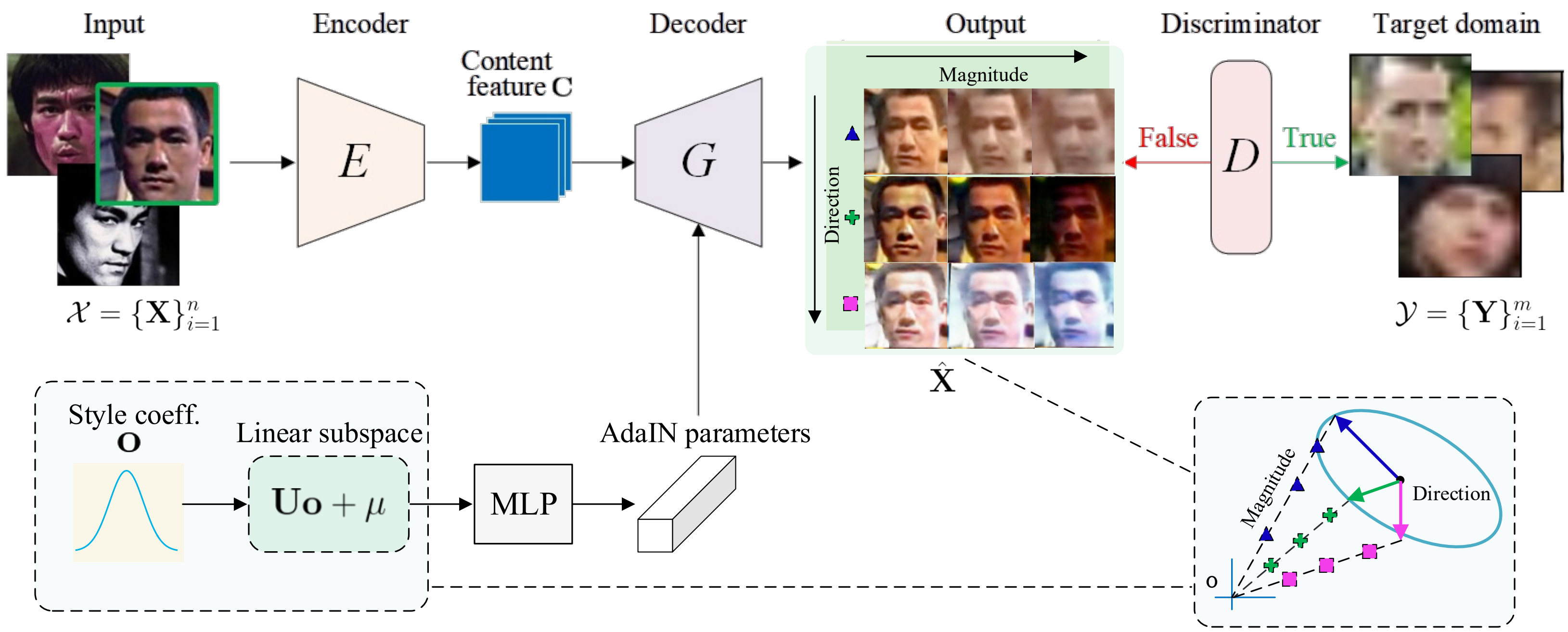
Motivation

- Previous synthesis models: ***limited face properties***; offline and ***blind*** data augmentation.
- Facial properties should be generalizable to the challenging unconstrained testing scenarios.
- Feedback-based face synthesis is more beneficial to FR models.



Controllable Face Synthesis

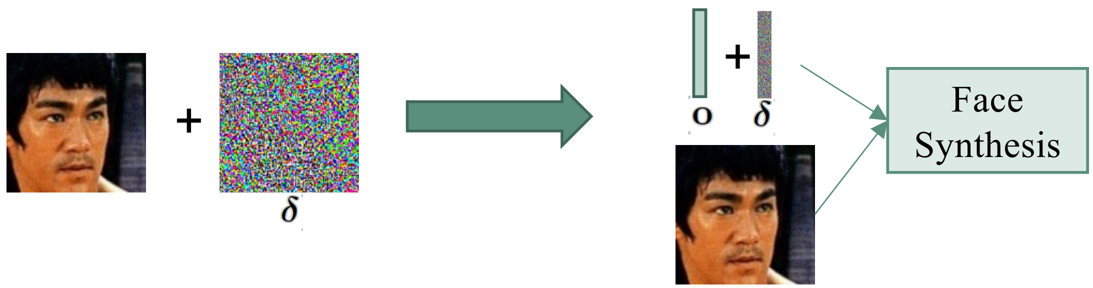
- The synthesis model can discover the **styles** in the target unconstrained data.
- The synthesis model is precisely-controllable in the style latent space, in both ***diversity*** and ***degree***.



Feng Liu, Minchul Kim, Anil Jain, and Xiaoming Liu. Controllable and Guided Face Synthesis for Unconstrained Face Recognition. ECCV 2022

Guided Face Synthesis for Face Recognition

- The FR model feedback signal is incorporated into the face generation using ***adversarial perturbation***.
- The manipulation of the low-dimensional style space renders this feedback ***meaningful*** and ***efficient***.



- Style latent perturbations to maximize the classifier loss

$$\boldsymbol{\delta}^* = \underset{||\boldsymbol{\delta}||<\epsilon}{\arg\max}\, \mathcal{L}_{cla}\left(\mathcal{F}(\mathbf{X}^*), l\right), \text{where } \mathbf{X}^* = G(E(\mathbf{X}), \text{MLP}(\mathbf{U}(\mathbf{o}+\boldsymbol{\delta})+\boldsymbol{\mu}))$$

- Optimize the face embedding model

$$\min_{\theta} \mathcal{L}_{cla}([\mathbf{X}^*, \mathbf{X}], l)$$

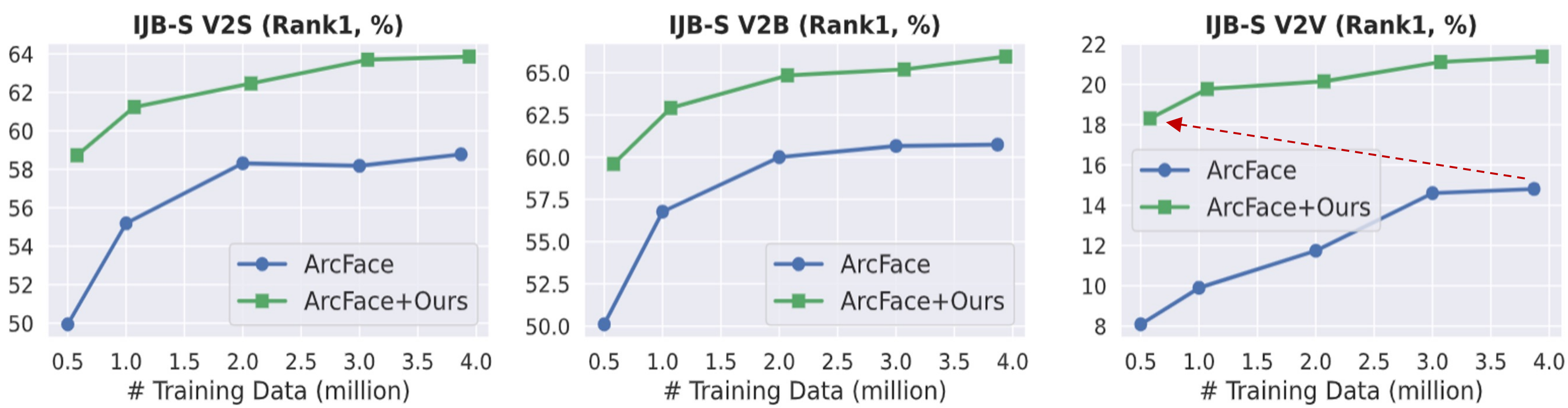
Feng Liu, Minchul Kim, Anil Jain, and Xiaoming Liu. Controllable and Guided Face Synthesis for Unconstrained Face Recognition. ECCV 2022

Face Recognition Results on IJB-S and TinyFace

- Our synthesis models could be plugged into any SoTA FR model and improve its performance.

Method	Labeled Train Data	Backbone	IJB-S V2S				IJB-S V2B				IJB-S V2V				TinyFace	
			Rank1	Rank5	1%	10%	Rank1	Rank5	1%	10%	Rank1	Rank5	1%	10%	Rank1	Rank5
ArcFace[1]	MS1MV2-*	ResNet-50	58.78	66.40	40.99	50.45	60.66	67.43	43.12	51.38	14.81	26.72	2.51	5.72	62.21	66.85
ArcFace+Ours*	MS1MV2-*	ResNet-50	61.69	68.33	43.99	53.34	62.20	69.50	44.38	53.49	18.14	31.34	2.09	4.51	62.39	67.36
ArcFace+Ours	MS1MV2-*	ResNet-50	**63.86**	**69.95**	**47.86**	**56.44**	**65.95**	**71.16**	**47.28**	**57.24**	**21.38**	**35.11**	**2.96**	**7.41**	**63.01**	**68.21**
AdaFace[2]	WebFace12M	IResNet-100	71.35	76.24	59.40	**66.34**	71.93	76.56	59.37	**66.68**	36.71	50.03	4.62	11.84	72.29	74.97
AdaFace+Ours	WebFace12M	IResNet-100	**72.54**	**77.59**	**60.94**	66.02	**72.65**	**78.18**	**60.26**	65.88	**39.14**	**50.91**	**5.05**	**13.17**	**73.87**	**76.77**

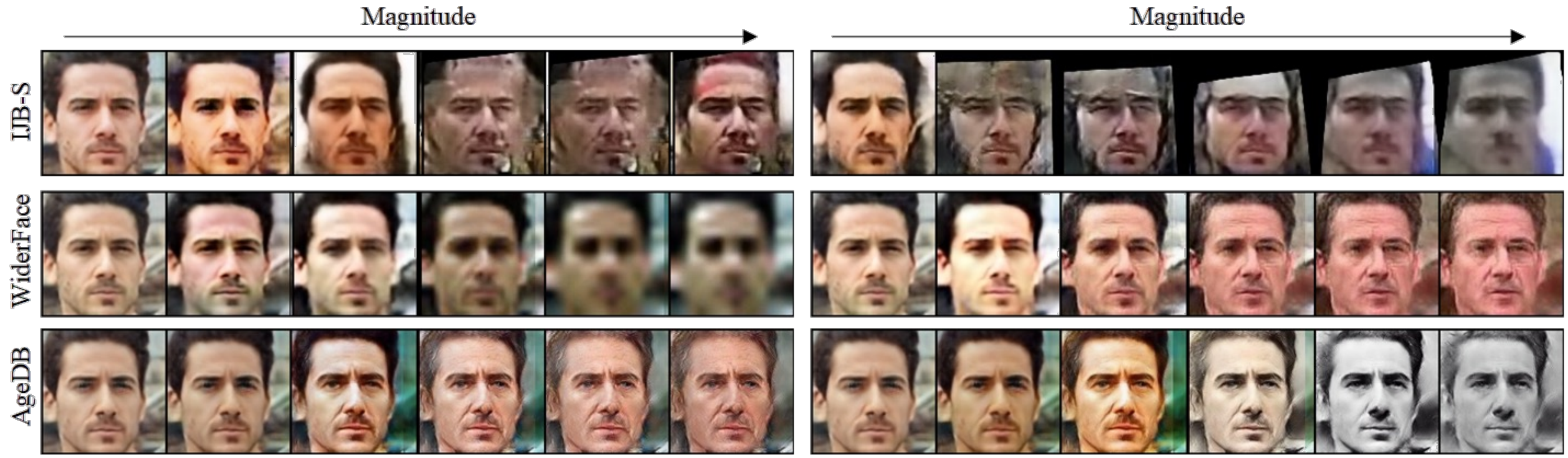
- Our synthesis models can boost FR performance even with less labelled training samples.



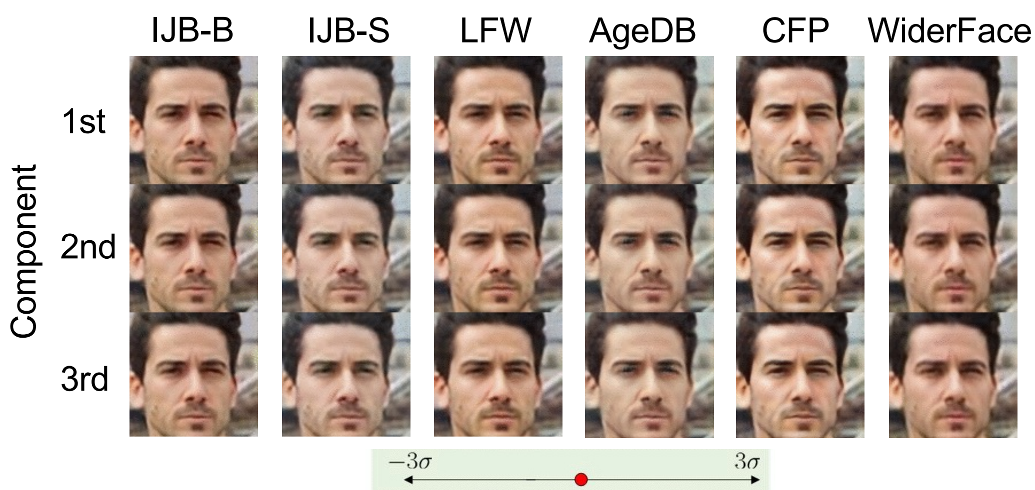
[1] Arcface: Additive angular margin loss for deep face recognition. CVPR 2019.
[2] AdaFace: Quality Adaptive Margin for Face Recognition. CVPR 2022.

Visualizations of the Face Synthesis Model

- Interpretable magnitude of the style coefficient.



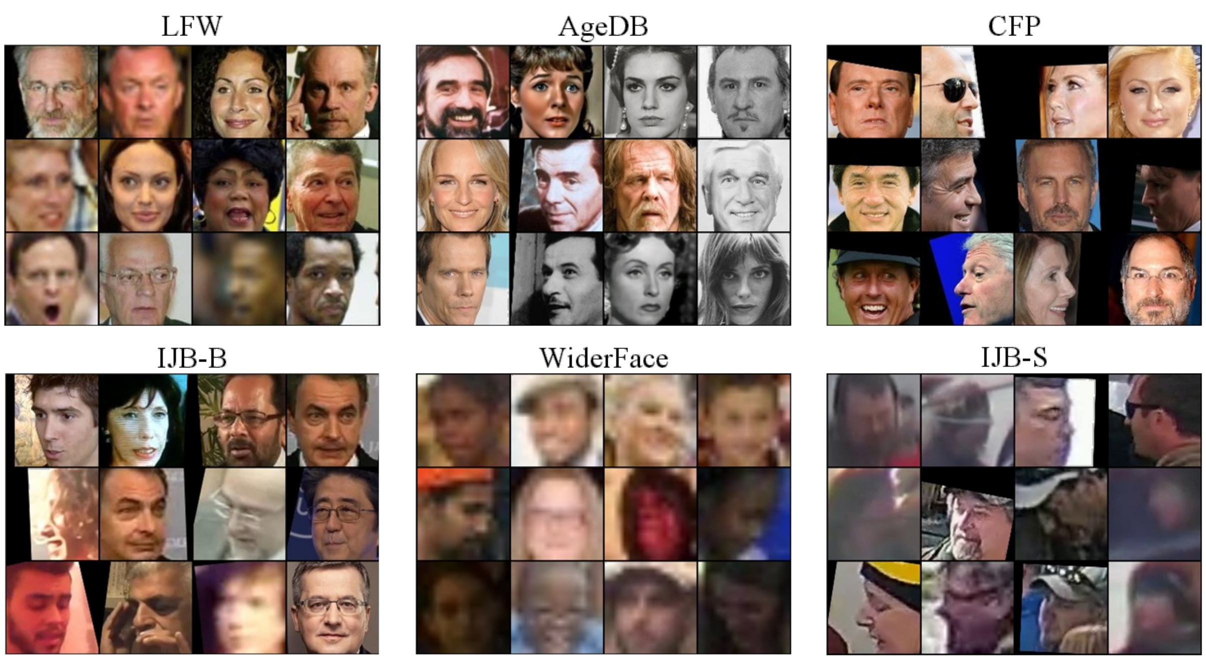
- Learned the orthonormal basis of the subspace.

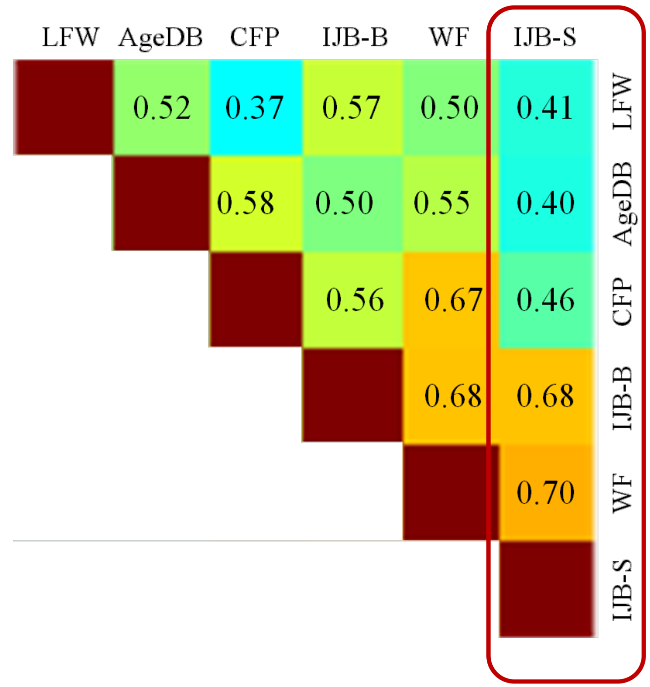


By-product: Dataset Distribution Similarity Measurement

- The distribution similarity between datasets A and B

$$\mathcal{S}(A,B) = \frac{1}{q}\left(\sum_{i}^{q} S_C(\mathbf{u}_A^i + \boldsymbol{\mu}_A, \mathbf{u}_B^i + \boldsymbol{\mu}_B)\right)$$





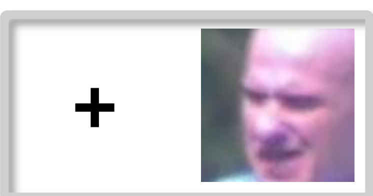
Cluster and Aggregate: Face Recognition with Large Probe Set

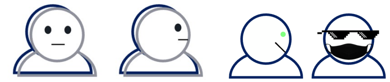
Minchul Kim, Feng Liu, Anil K. Jain, Xiaoming Liu

NeurIPS 2022

Problem Definition

Traits of Face Recognition with Videos





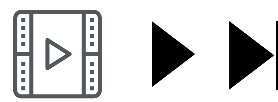
1 Varied Identifiability

Some images are more identifiable than others

2 10 ~ 1,000,000

Varied Number of Images

Number of images are not fixed.

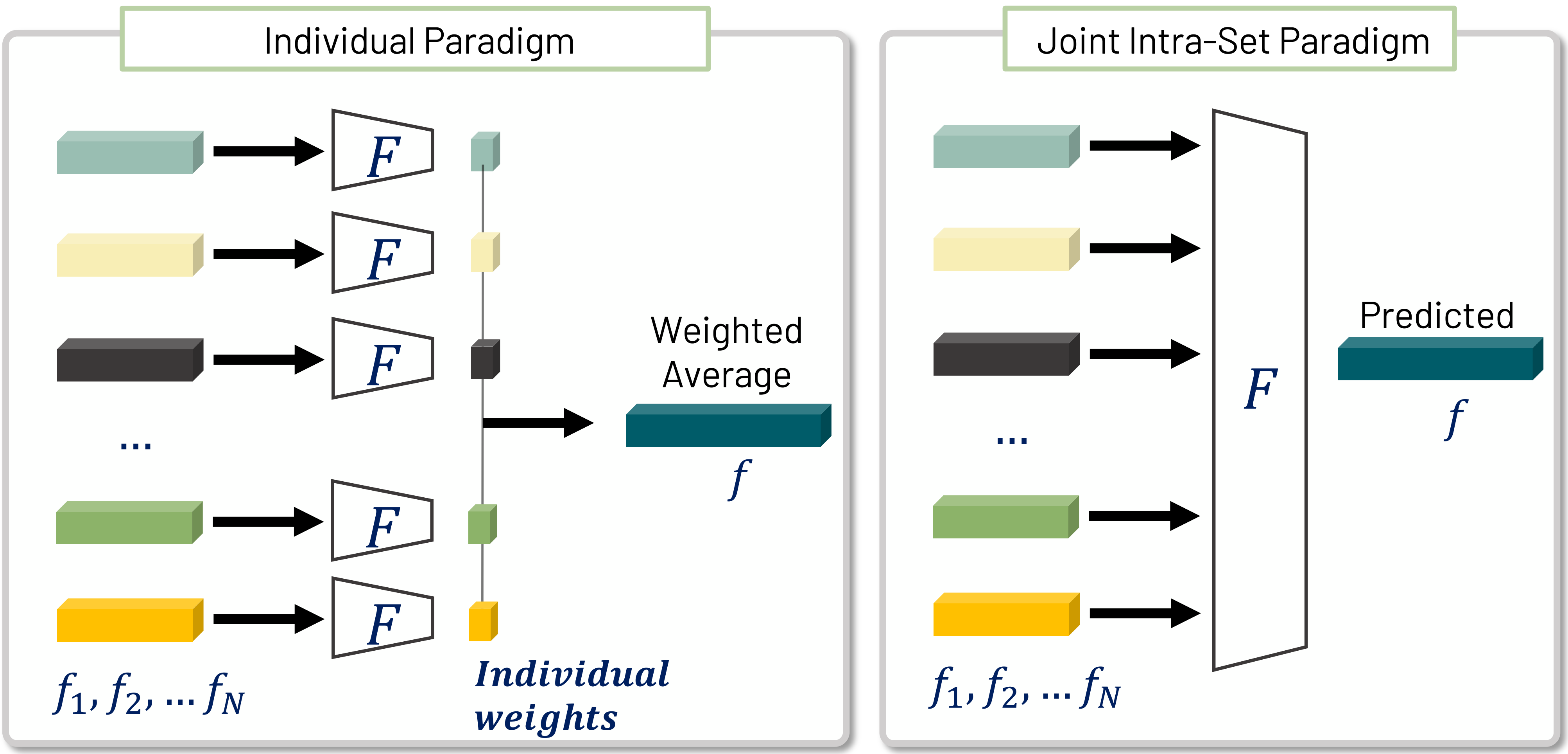


3 Sequential Inputs

Videos come in sequentially. We use what we have up-to the current timeframe.

Subject in the slide consented to publication

Problem of Previous Methods

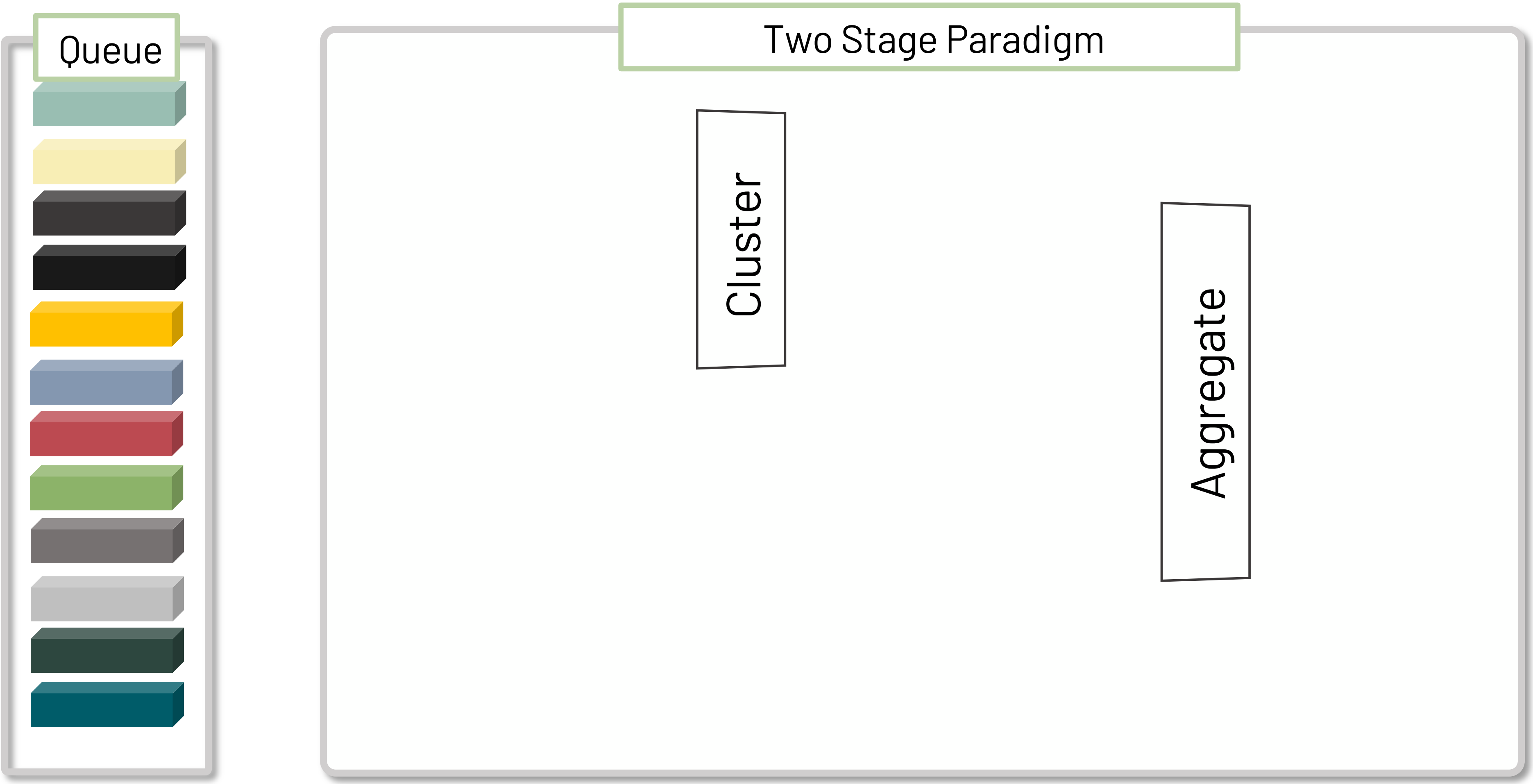


(No Intra—Set Relationship)

(Cannot handle large N)

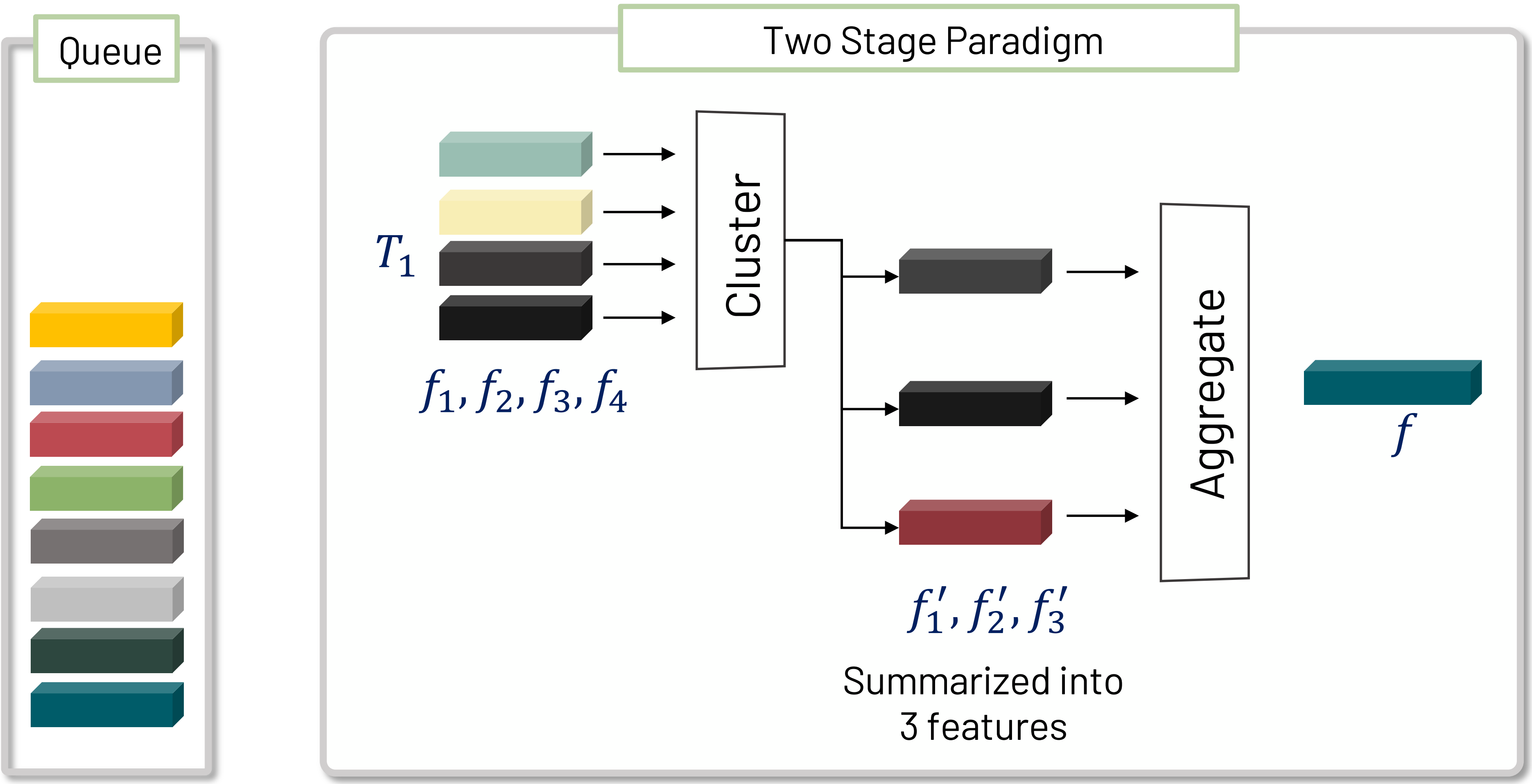
Motivation

Large N / Sequential Scenario



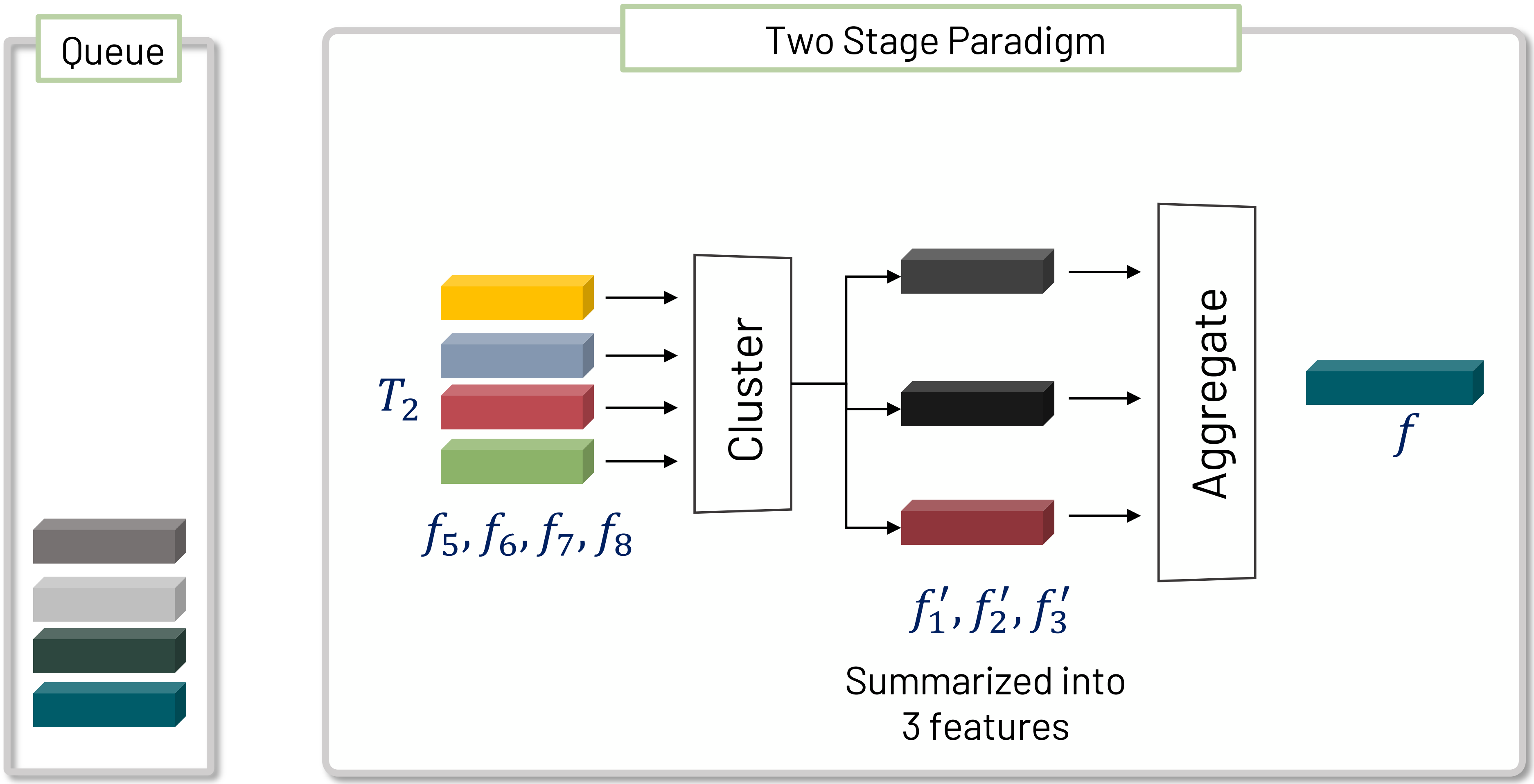
Motivation

Large N / Sequential Scenario



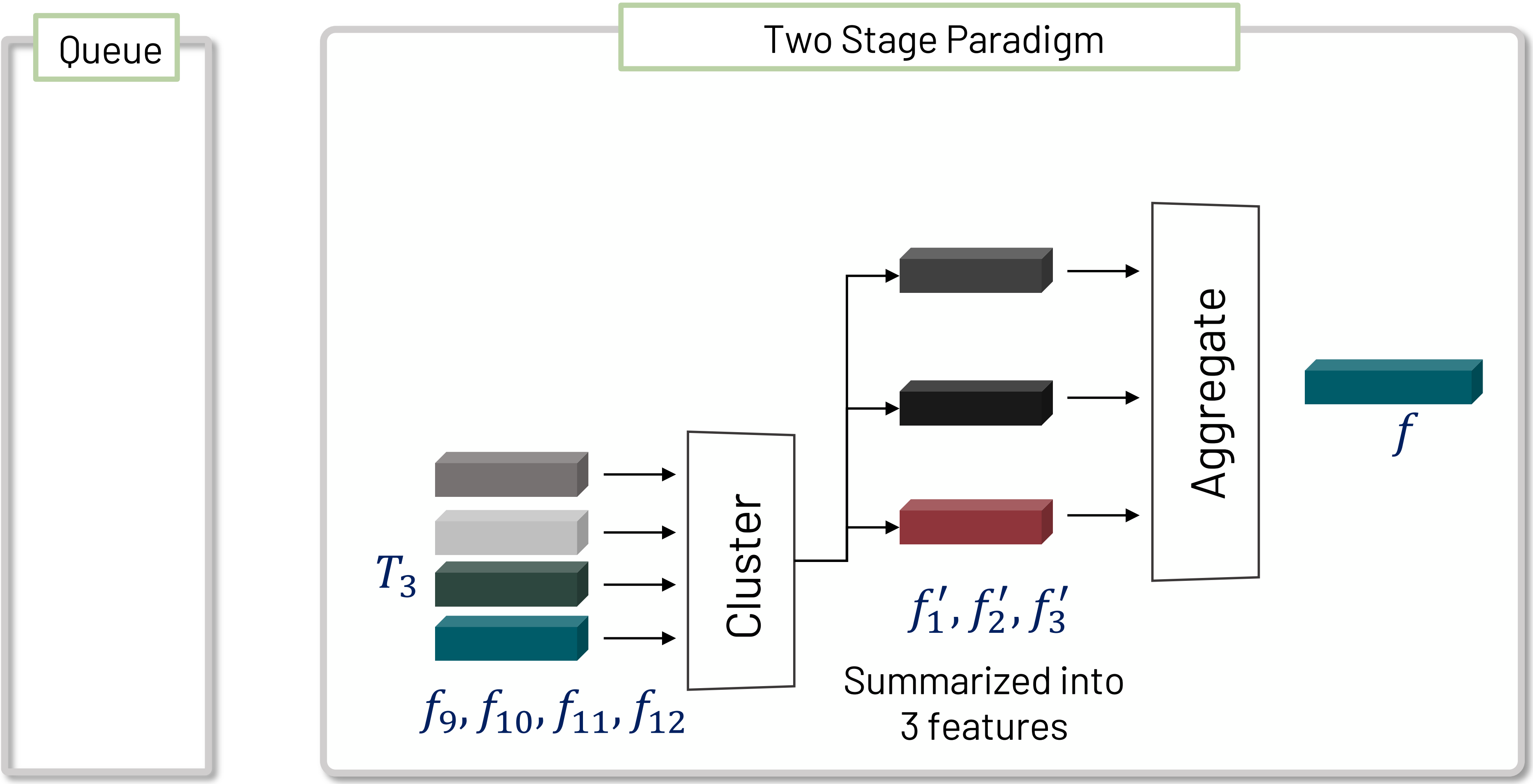
Motivation

Large N / Sequential Scenario



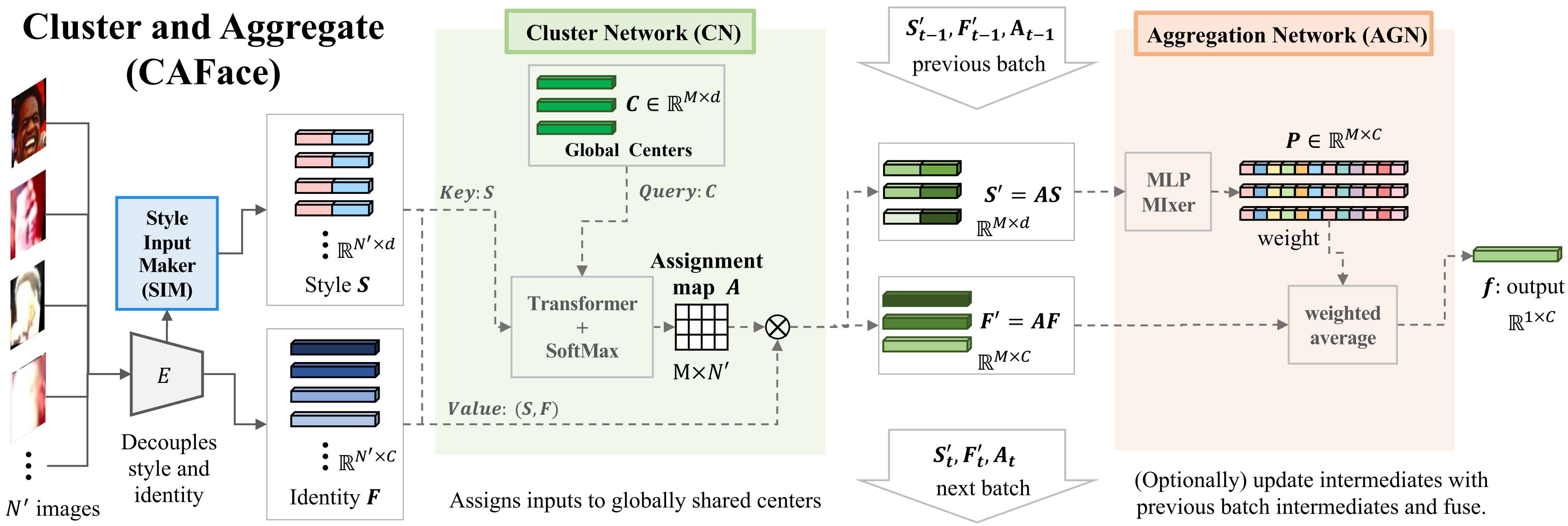
Motivation

Large N / Sequential Scenario



Method

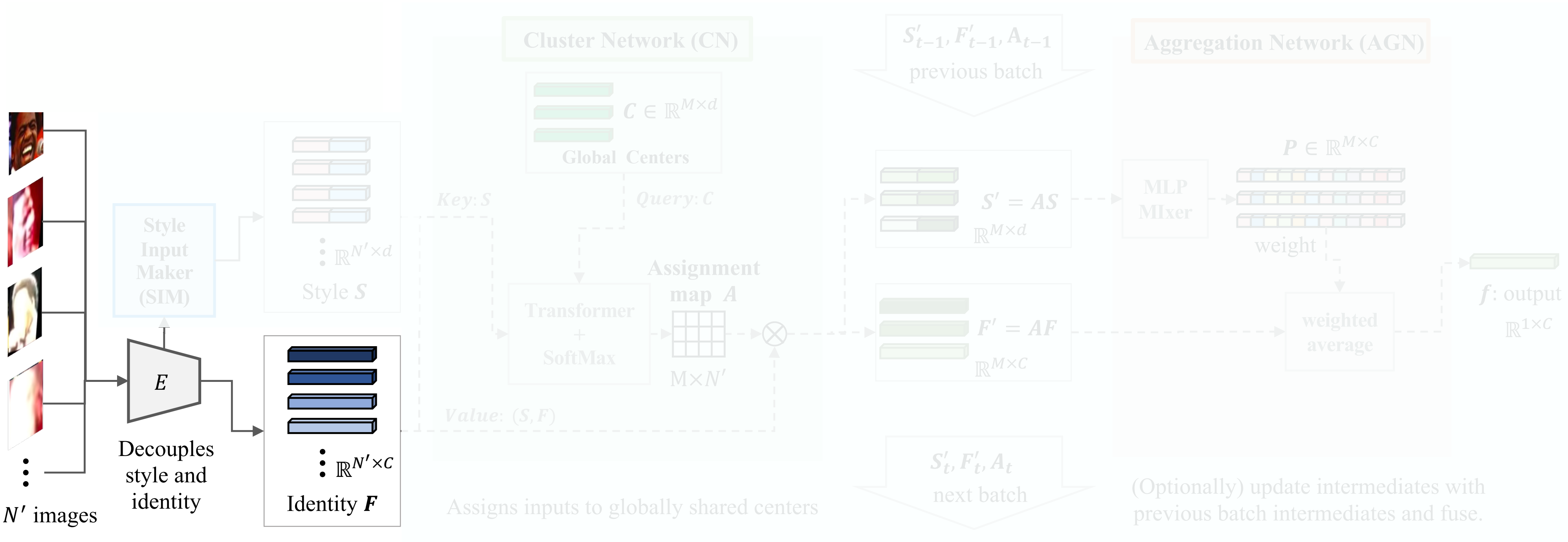
Architecture



Overall Architectures
3 components (SIM, CN, AGN)

Method

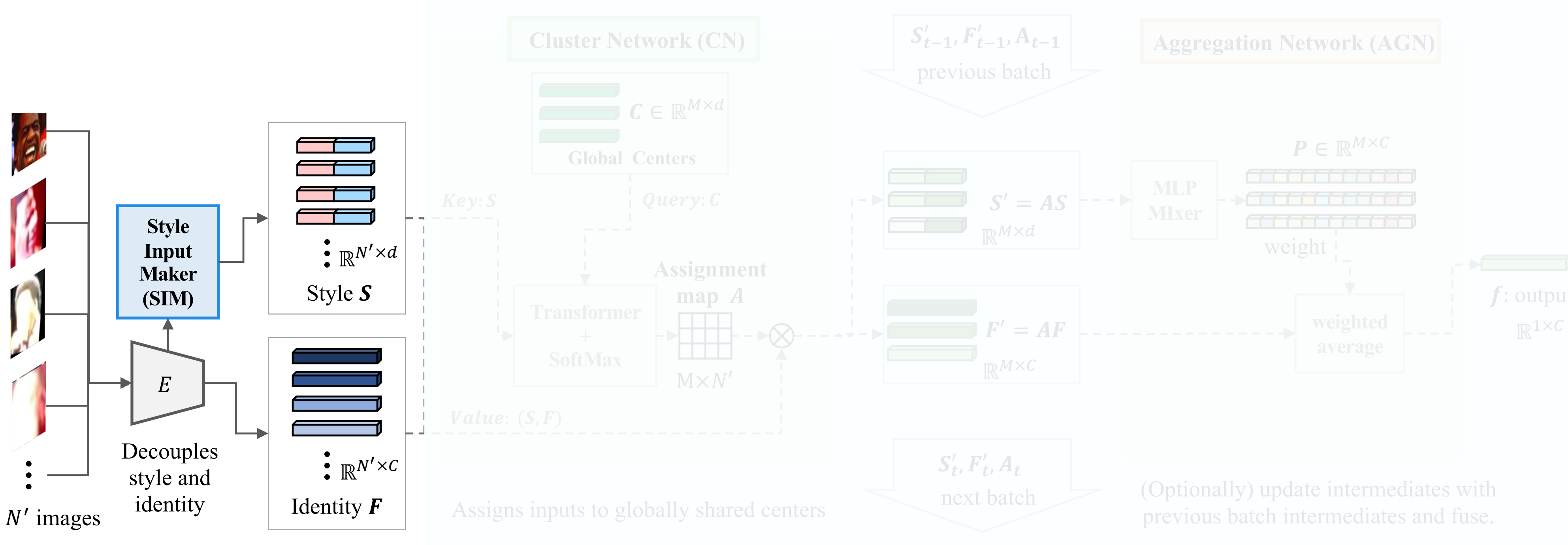
Architecture



Input images fed into the fixed feature extractor.

Method

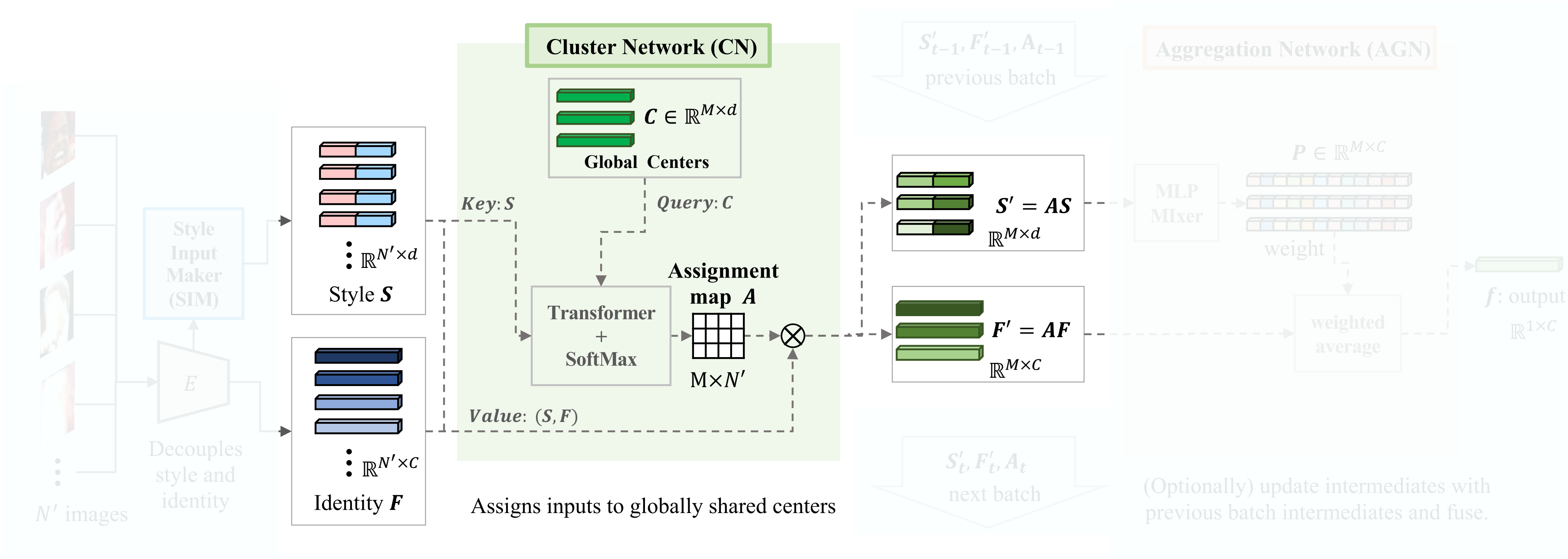
Architecture



Extract 1) style $\{\boldsymbol{s}_i\}^N$ and 2) identity $\{\boldsymbol{f}_i\}^N$
using the fixed feature extractor

Method

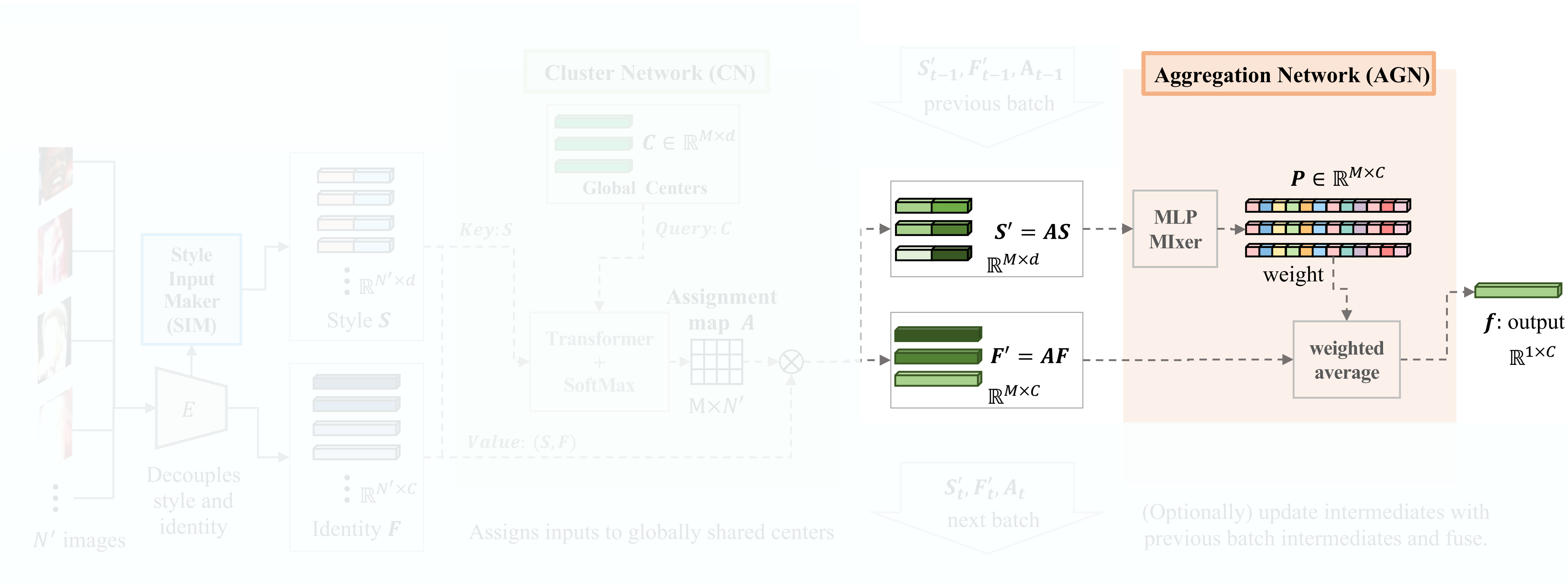
Architecture



CN uses learned centers $\{c_j\}^M$ and $\{s_i\}^N$ to create assignment map A.

A is used to map $\{f_i\}^N \rightarrow \{f_j\}^M$ and $\{s_i\}^N \rightarrow \{s_j\}^M$

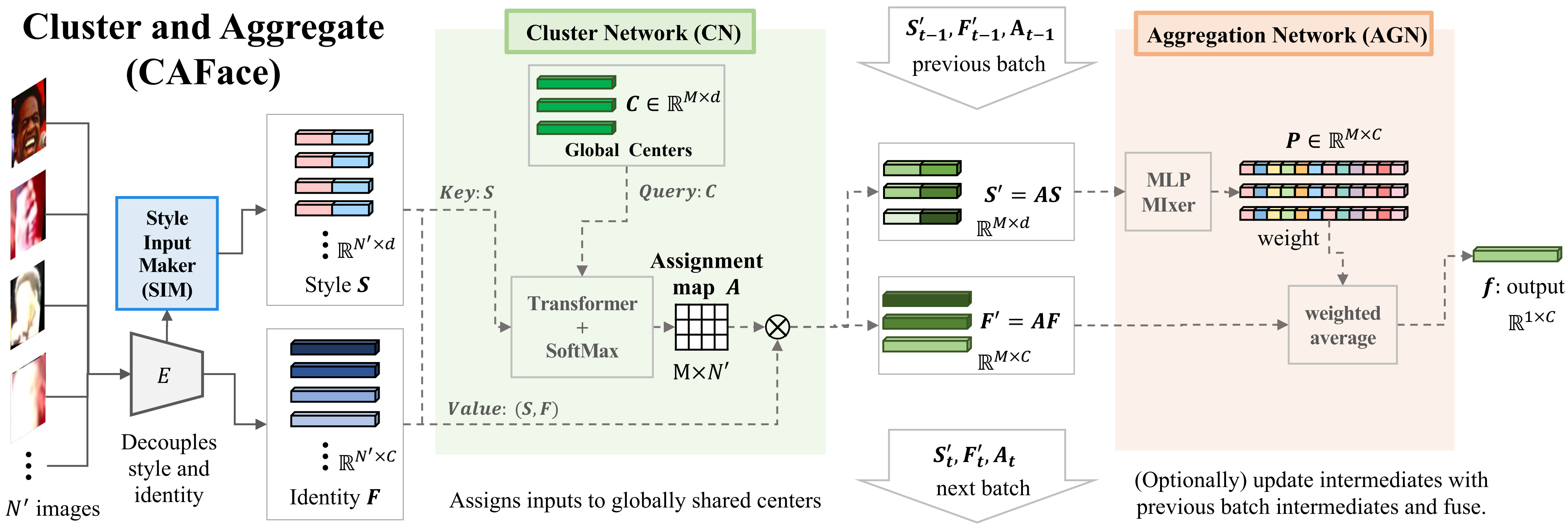
Architecture



AGN maps $\{f_j\}^M, \{s_j\}^M \rightarrow f$ with intra-set relationship.

Method

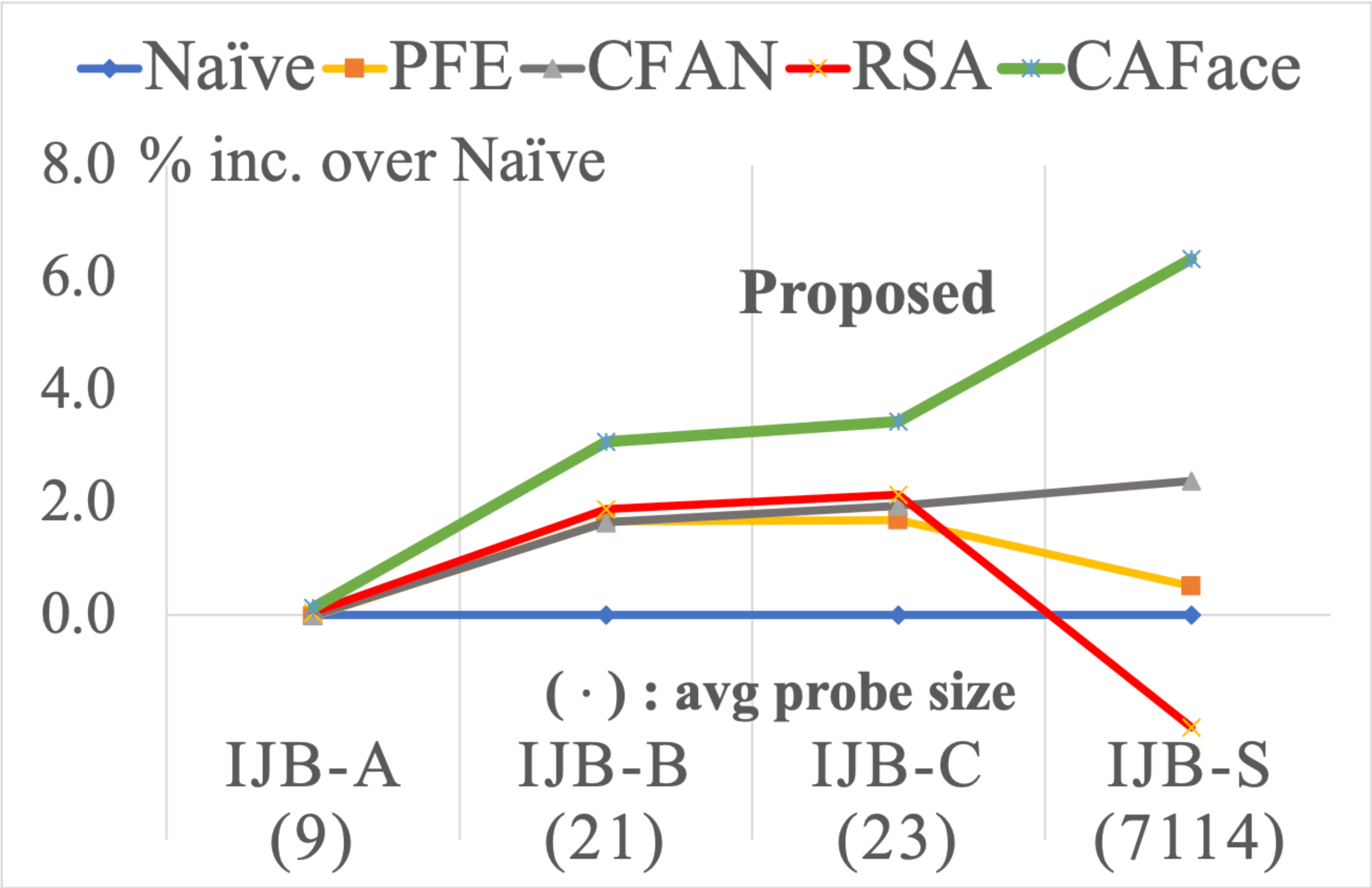
Architecture



Intermediate features $\{f_j\}^M$ and $\{s_j\}^M$ are updated in sequential setting.

Experiments

Performance Gain over simple average using feature fusion methods.



Naïve: Simple Average
PFE, CFAN: single image weight estimation
RSA: Attention Mechanism

Largest Probe size, Largest Perf. gain

Experiments

Table 3: A performance comparison of recent methods on the IJB-S [24] dataset.

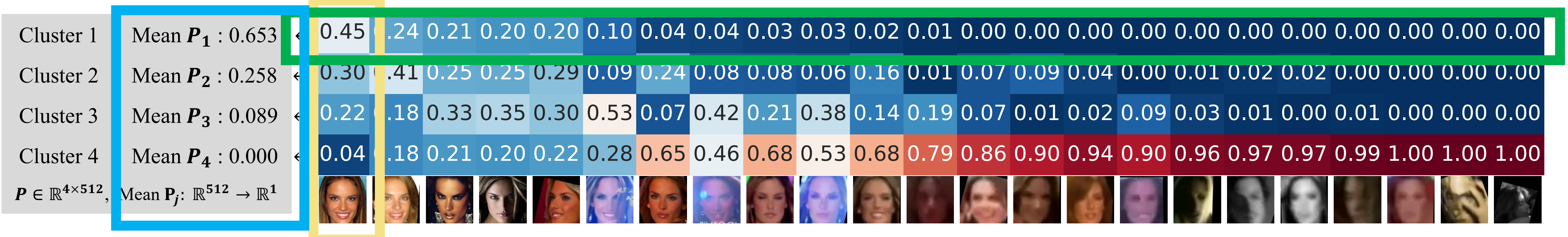
Method	Surveillance-to-Single			Surveillance-to-Booking			Surveillance-to-Surveillance		
	Rank-1	Rank-5	1%	Rank-1	Rank-5	1%	Rank-1	Rank-5	1%
Naive Average	69.26	74.31	57.06	70.32	75.16	56.89	32.13	46.67	5.32
PFE [46]	69.50	74.39	57.51	70.53	75.29	57.98	32.27	46.70	5.41
CFAN [15]	70.00	74.58	57.93	70.90	75.58	58.09	31.66	45.59	5.79
RSA [31]	63.04	67.33	51.62	63.54	68.23	51.89	16.82	31.80	0.75
CAFace	**71.61**	**76.43**	**62.21**	**72.72**	**77.41**	**62.68**	**36.51**	**49.59**	**8.78**
CAFace (Random Order)	71.65 $\pm$0.05	76.37 $\pm$0.04	62.27 $\pm$0.11	72.77 $\pm$0.04	77.37 $\pm$0.03	62.70 $\pm$0.06	36.43 $\pm$0.08	49.40 $\pm$0.05	8.89 $\pm$0.03

Changing the order of probe sequence does not affect the performance.

Visualization of Assignment Map

Assignment Map $\boldsymbol{A}$ Visualization

Each cluster is formed by weighted averaging each row.

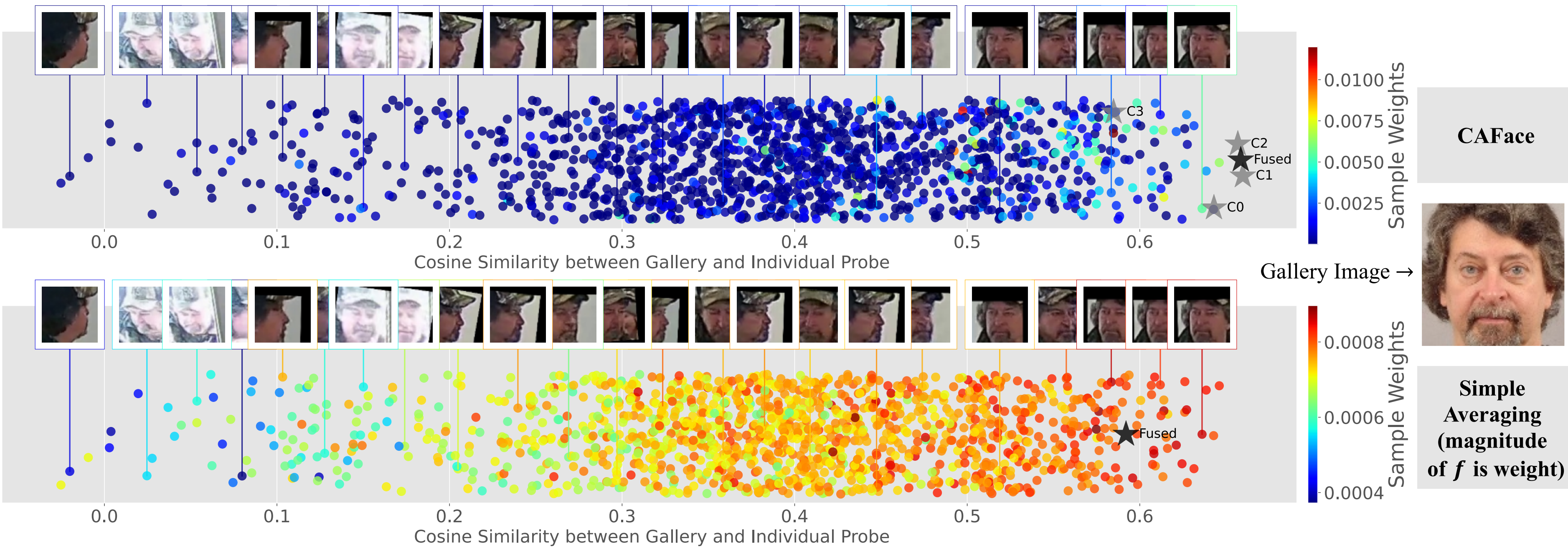


Importance of each cluster during aggregation.

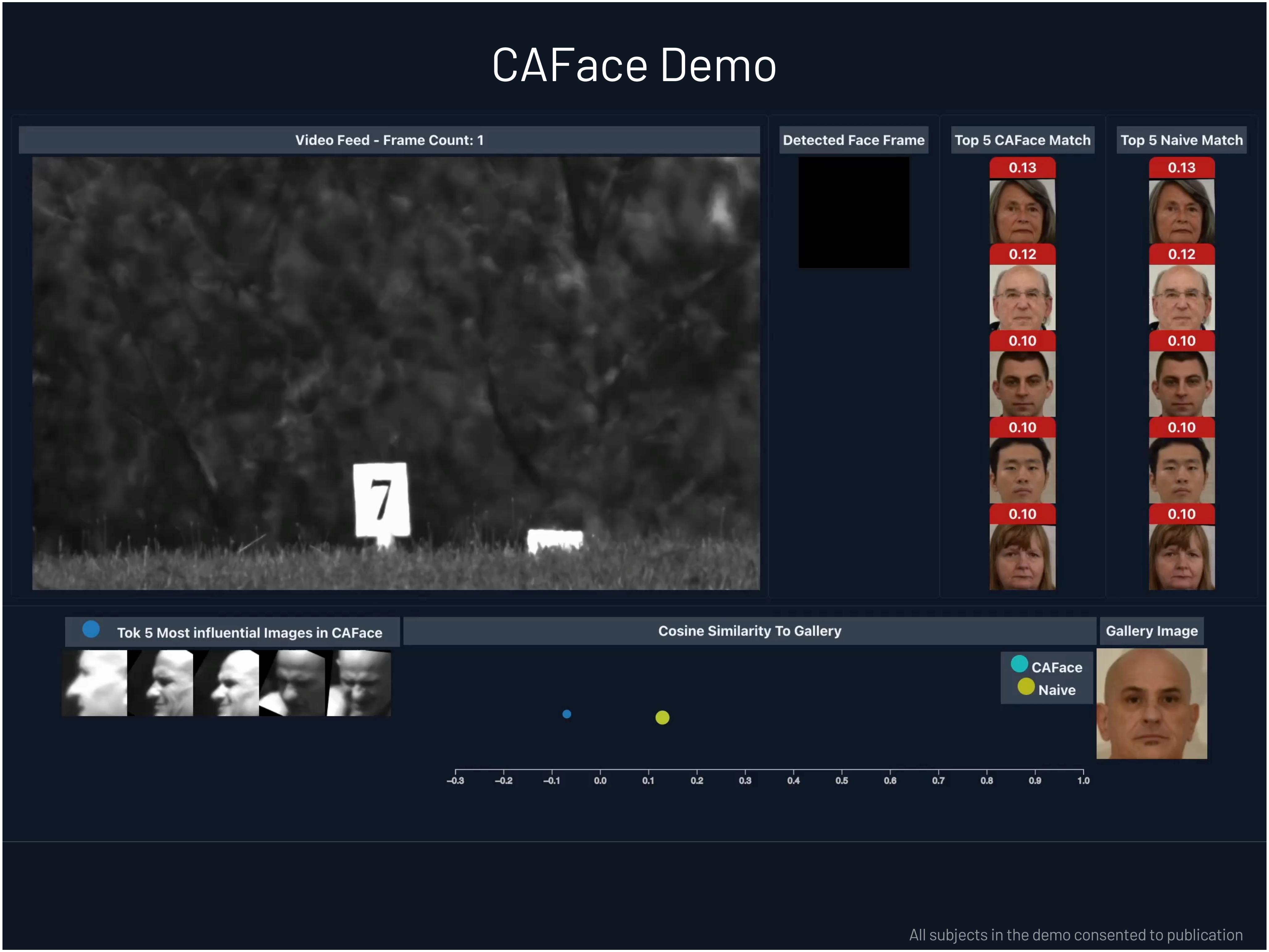
Each Column sums up to 1. They are soft assigned to cluster centers.

Weight Visualizations

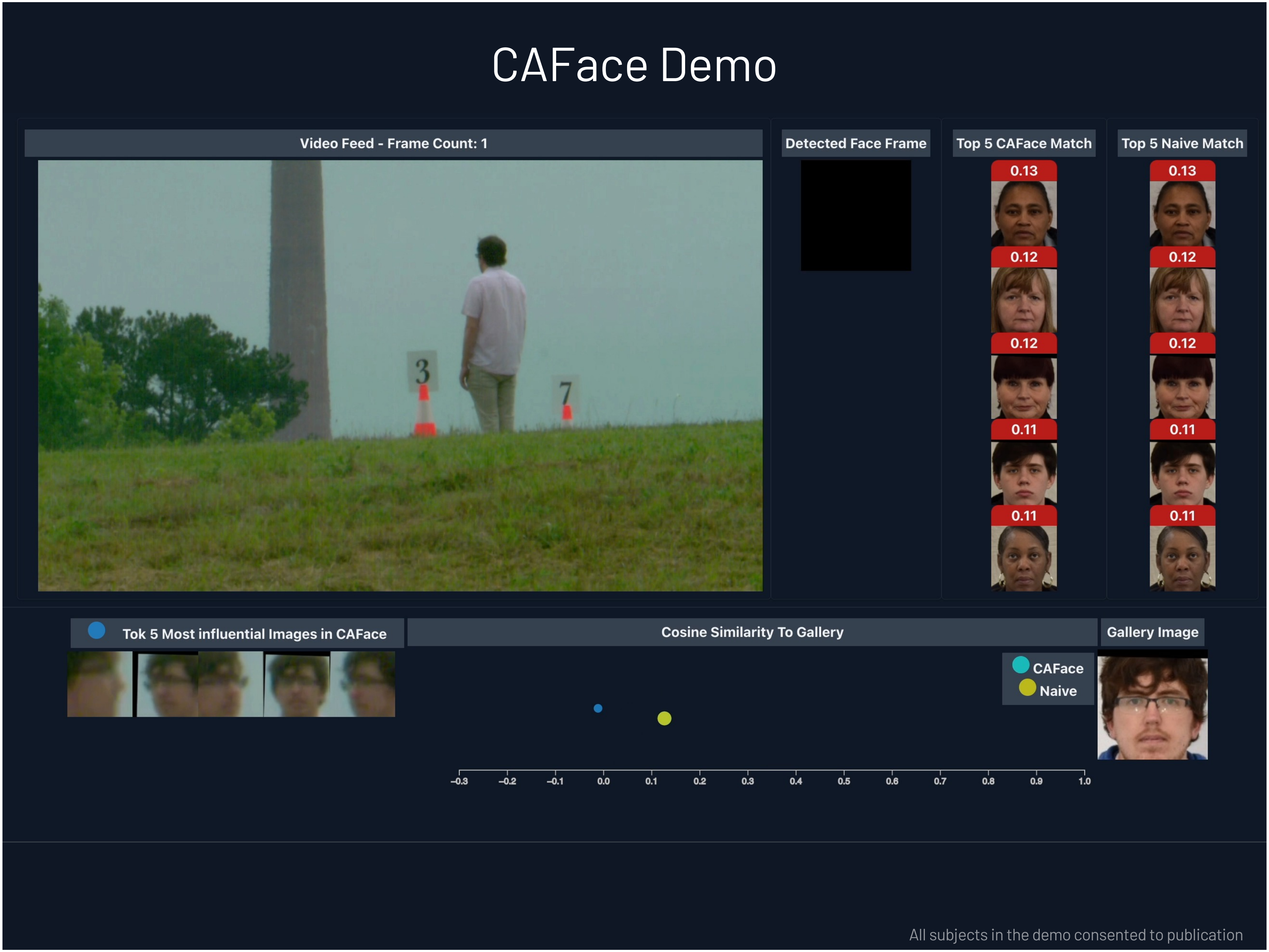
IJBS Probes' similarity to Gallery Visualization



Point colors indicate the weight during fusion.



CAFace Demo
Video Feed - Frame Count: 1
7
Detected Face Frame
Top 5 CAFace Match
0.13
0.12
0.10
0.10
0.10
Top 5 Naive Match
0.13
0.12
0.10
0.10
0.10
Tok 5 Most influential Images in CAFace
Cosine Similarity To Gallery
CAFace
Naive
−0.3 −0.2 −0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Gallery Image
All subjects in the demo consented to publication


CAFace Demo
Video Feed - Frame Count: 1
3
7
Detected Face Frame
Top 5 CAFace Match
0.13
0.12
0.12
0.11
0.11
Top 5 Naive Match
0.13
0.12
0.12
0.11
0.11
Tok 5 Most influential Images in CAFace
Cosine Similarity To Gallery
CAFace
Naive
−0.3 −0.2 −0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Gallery Image
All subjects in the demo consented to publication

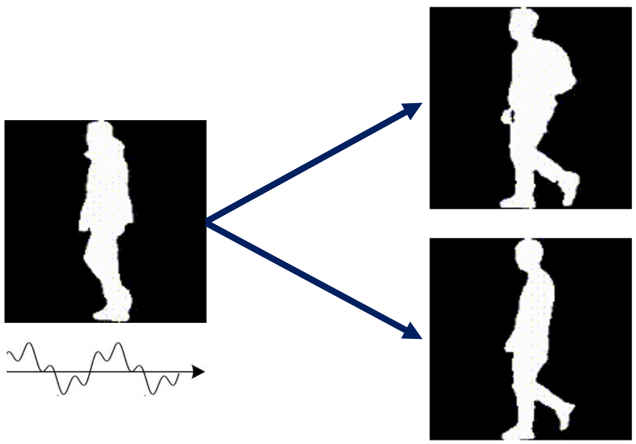
People Matching: Learning Clothing Invariant 3D Shape Representation

Feng Liu, Minchul Kim, ZiAng Gu, Anil Jain, and Xiaoming Liu

Under review

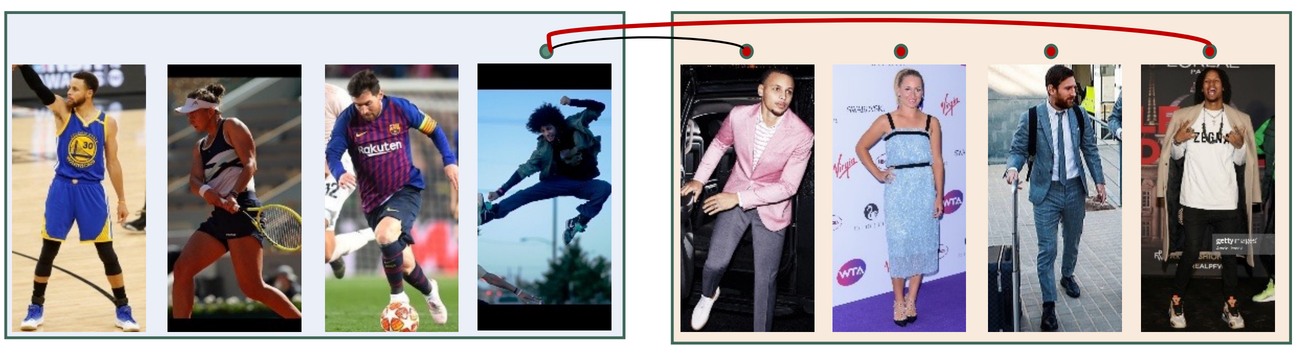
Problem Definition

People matching. Two main characteristics: diverse human activities and clothing changes



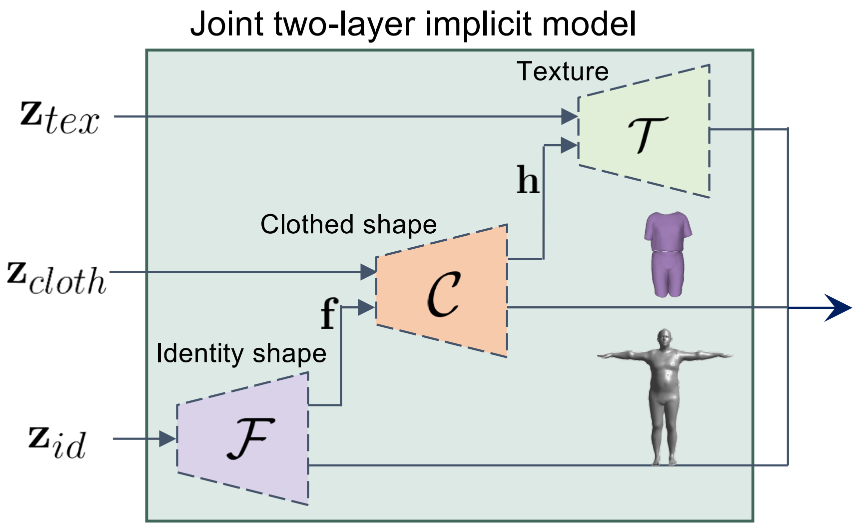
Gait recognition

Person re-identification

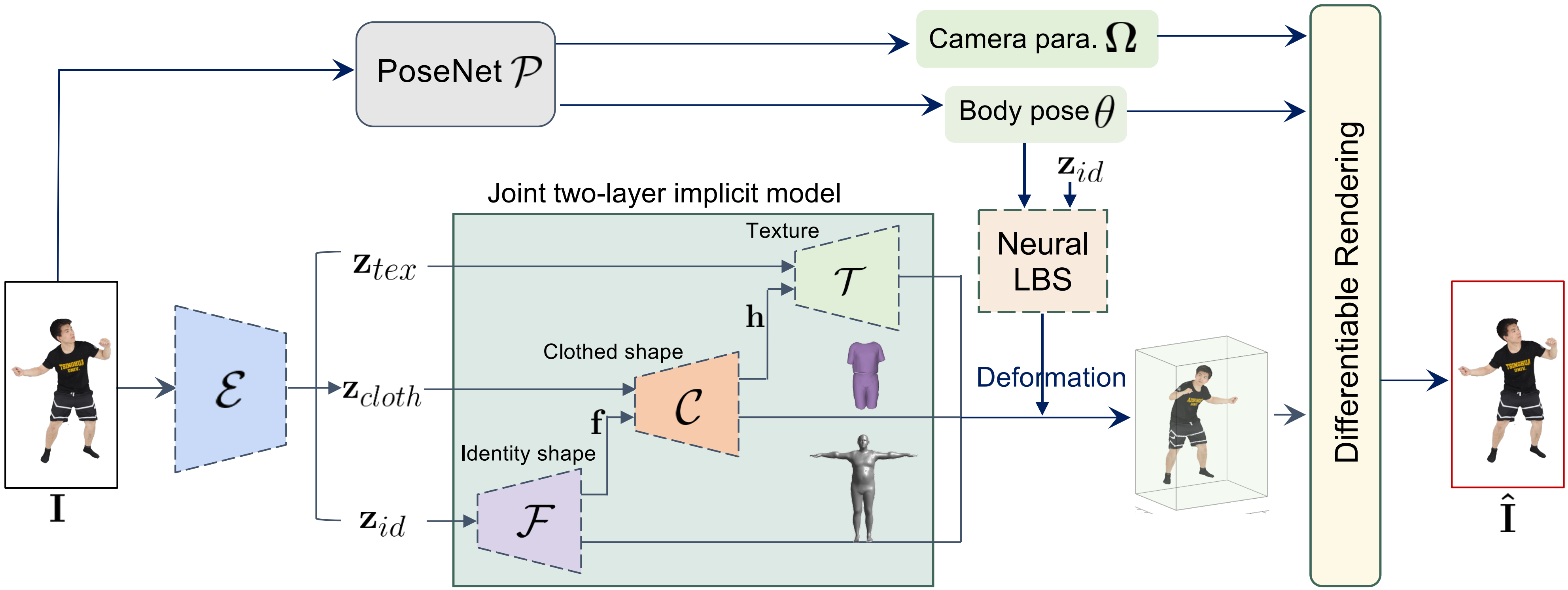


People matching

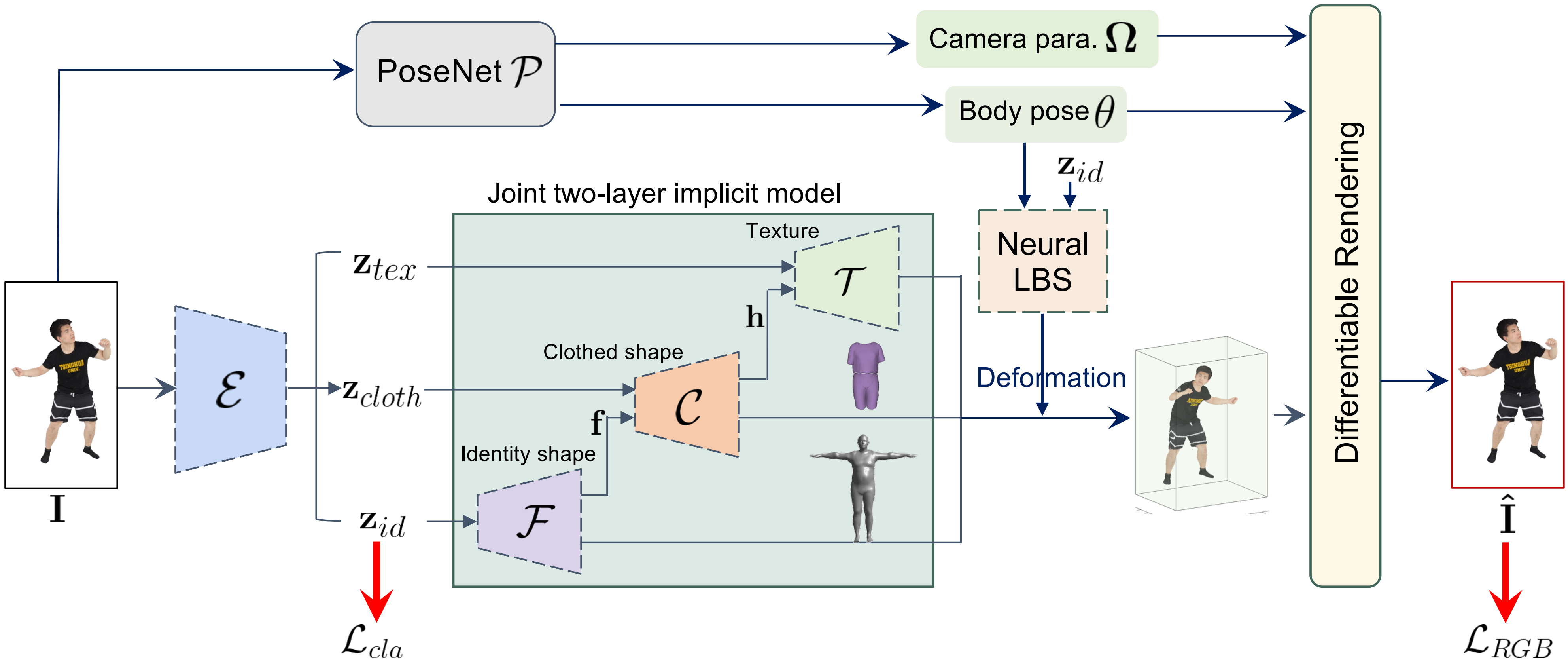
Disentangle identity and non-identity features in 3D body shape space



Disentangle identity and non-identity features in 3D body shape space



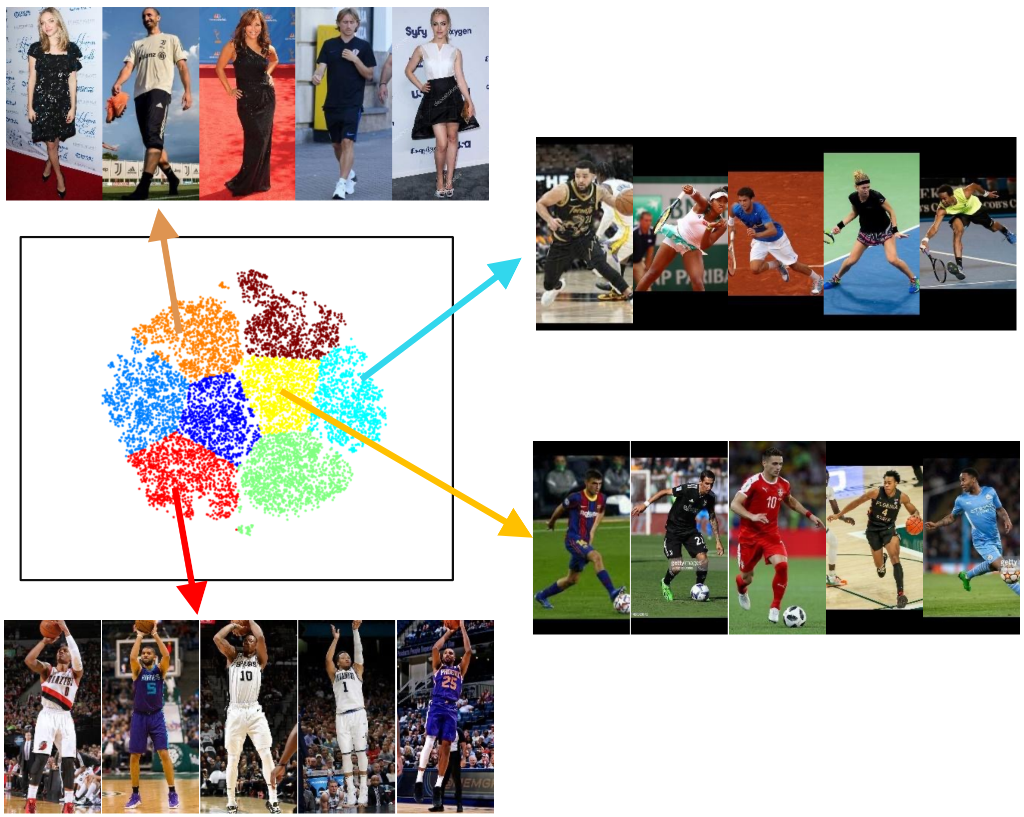
Disentangle identity and non-identity features in 3D body shape space



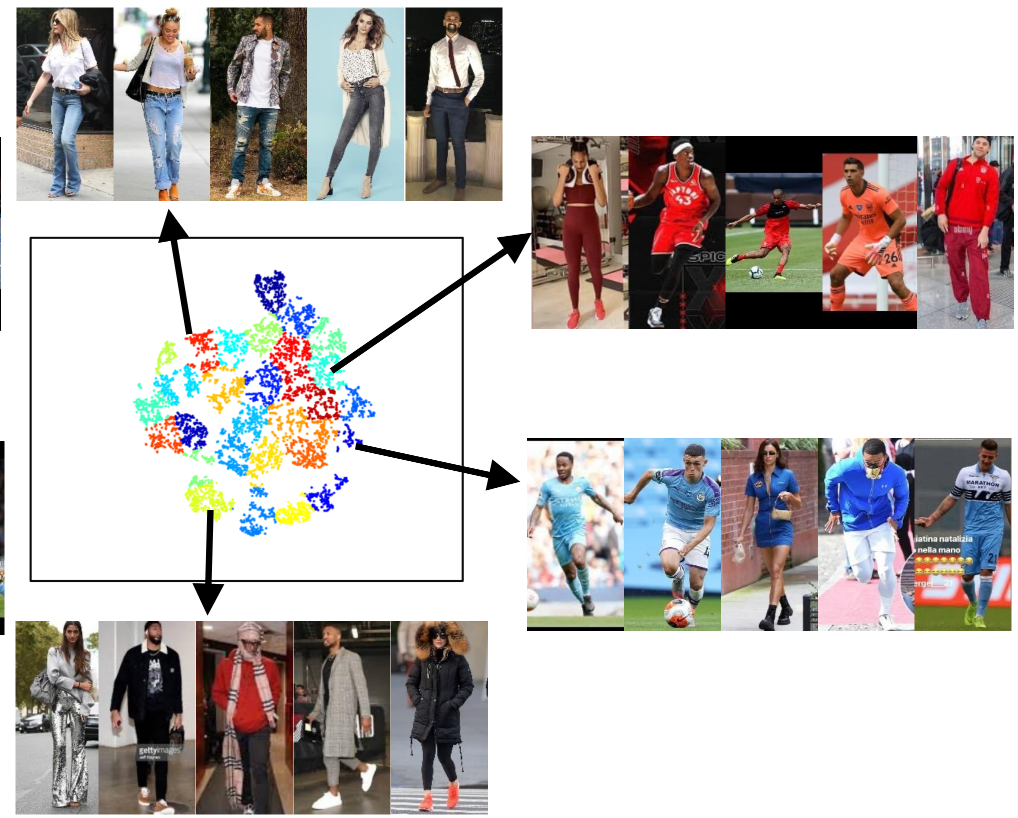
Diverse People Matching Dataset (DPMD)

87,821 images of 536 subjects

Examples of diverse poses



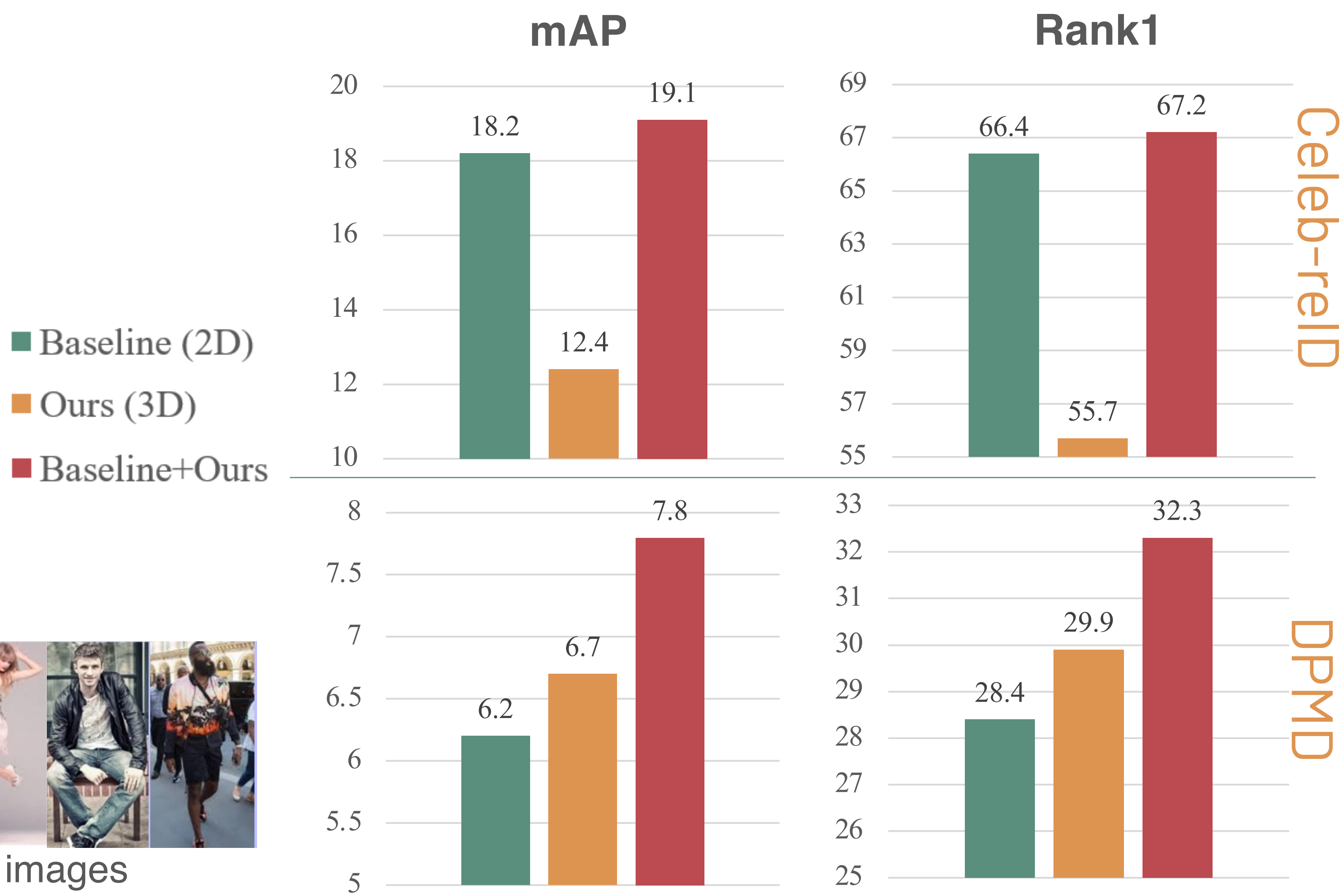
Examples of diverse clothes



People Matching Results

Best of baselines

- ReIDCaps
  TCSVT2019
- DG-Net
  CVPR2019
- Li et al.
  WACV2021
- CAL
  CVPR2022



Sample images

3D Reconstruction Results

Naked body | Clothed body | Rec. image | Naked body | Clothed body | Rec. image

Conclusions

- There are many research questions for low-quality recognition

- Even for conventional FR problems, there are research opportunities such as explainability, new architecture, etc.

- Body biometrics is just at the beginning and there is a great potential for further development.

Questions?