2024 Winter School on Biometrics

MICHIGAN STATE UNIVERSITY

Biometric Recognition at a distance

Dr. Xiaoming Liu

Computer Vision Lab

http://cvlab.cse.msu.edu

Michigan State University

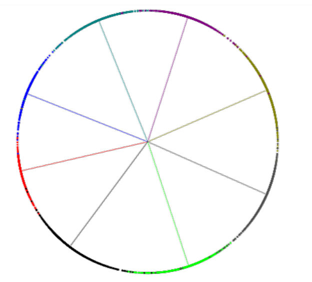
This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2022-21102100004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

Tremendous Research Progress

DR-GAN

Age synthesis

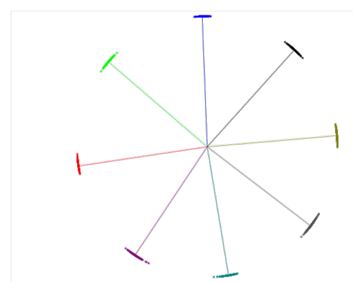
OpenGait



Softmax

SphereFace

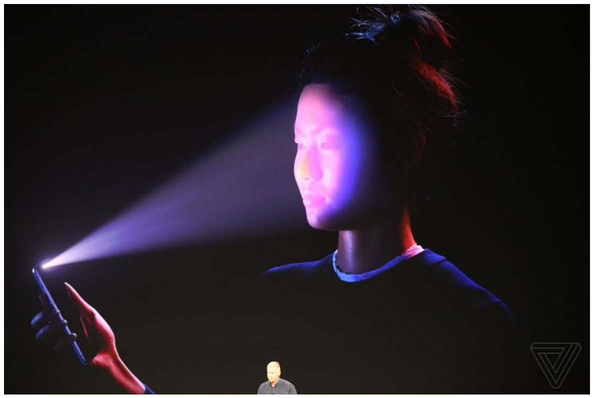
CosFace



ArcFace

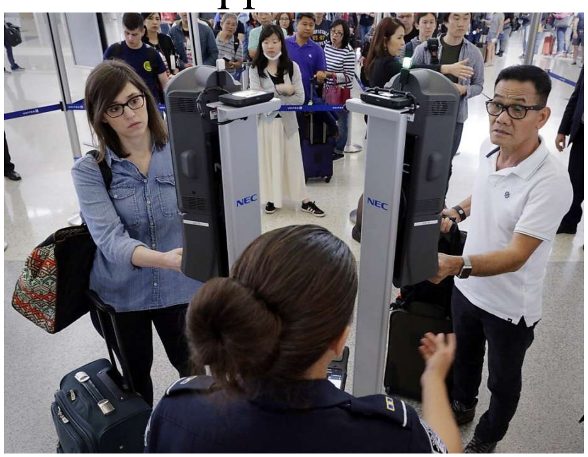
L. Tran et. al., Representation Learning by Rotating Your Faces, PAMI, 2018
H. Yang et. al., Learning Continuous Face Age Progression: A Pyramid of GANs, PAMI, 2019.
Fan et. al., OpenGait: Revisiting Gait Recognition Toward Better Practicality, CVPR, 2023.
J. Deng et. al., ArcFace: Additive Angular Margin Loss for Deep Face Recognition, CVPR, 2019.

Successful Applications



Apple

Alipay



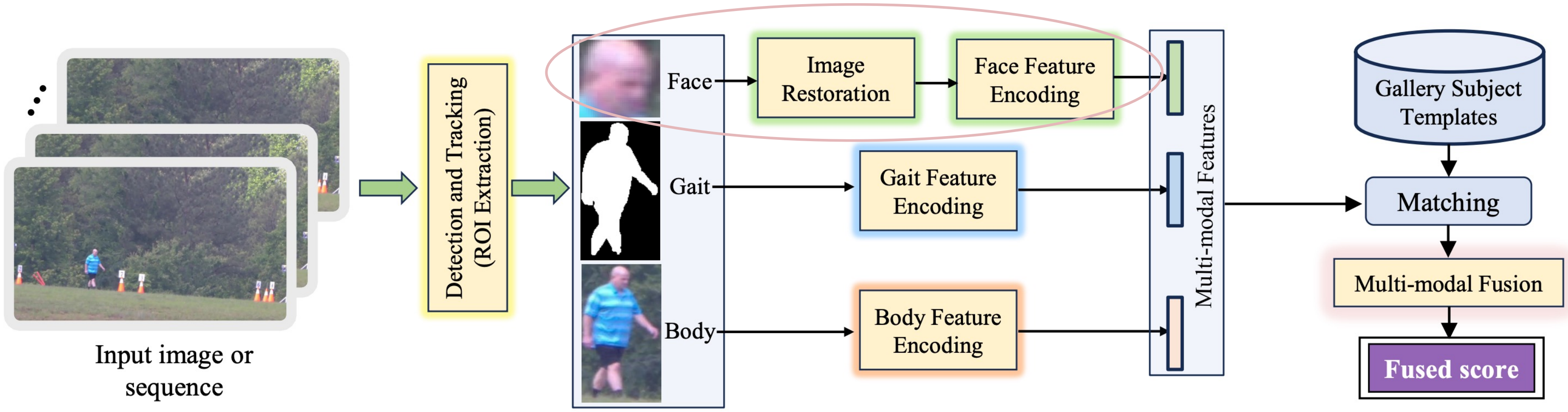
Boarding in Airports

Amazon One Palmprint

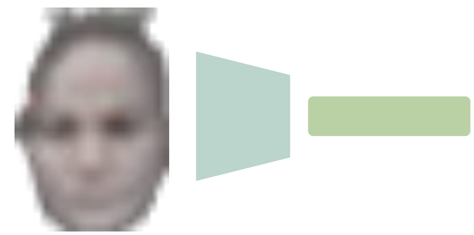
Identification at a Distance

BRIAR: The subject in the figure consented to publication.

Person Identification at a (far) distance



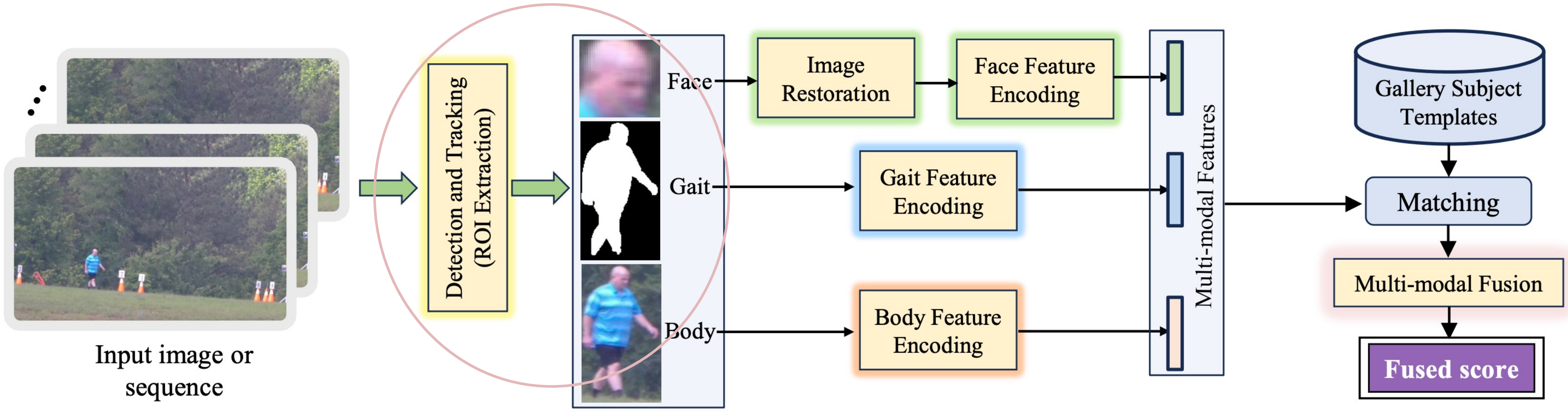
- **1. Generic matcher:** AdaFace (CVPR’22)



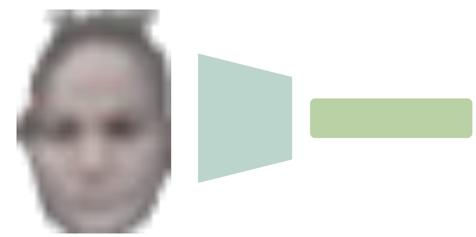
MICHIGAN STATE UNIVERSITY

The subject (ID: G01003) consented to publication.

Person Identification at a (far) distance



- **1. Generic matcher:** AdaFace (CVPR’22)

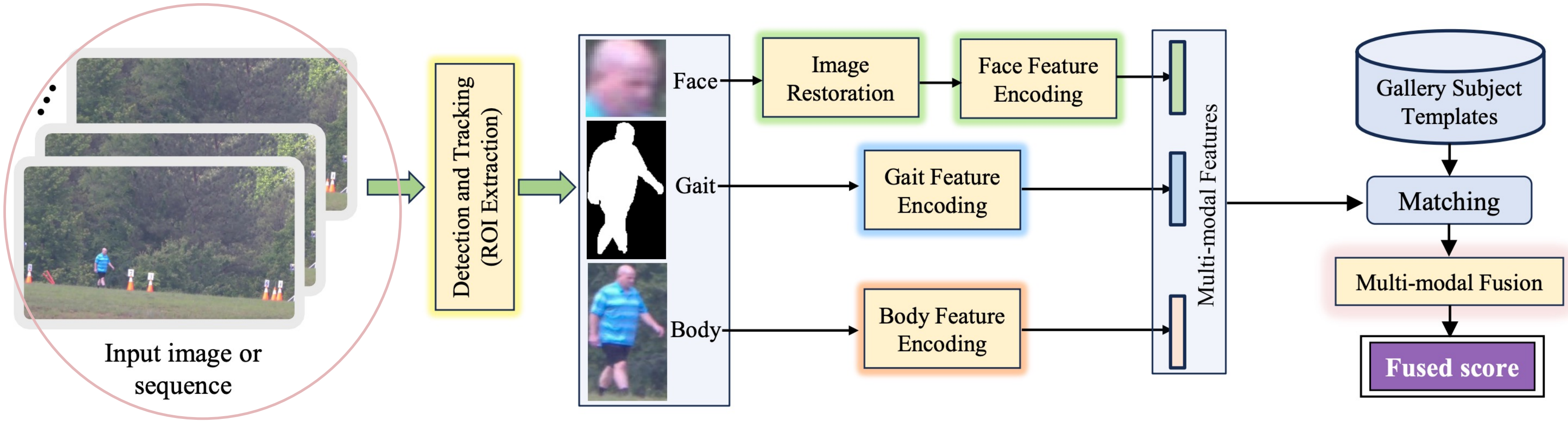


- **2. Domain adaption:** CFSM (ECCV’22)

MICHIGAN STATE UNIVERSITY

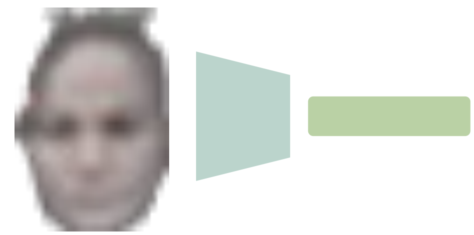
The subject (ID: G01003) consented to publication.

Person Identification at a (far) distance



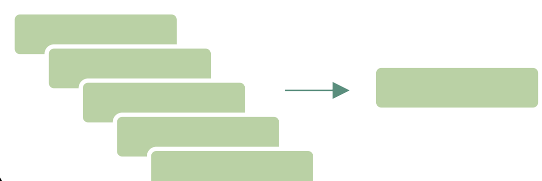
The subject (ID: G01003) consented to publication.

- **1. Generic matcher:**
  AdaFace (CVPR'22)

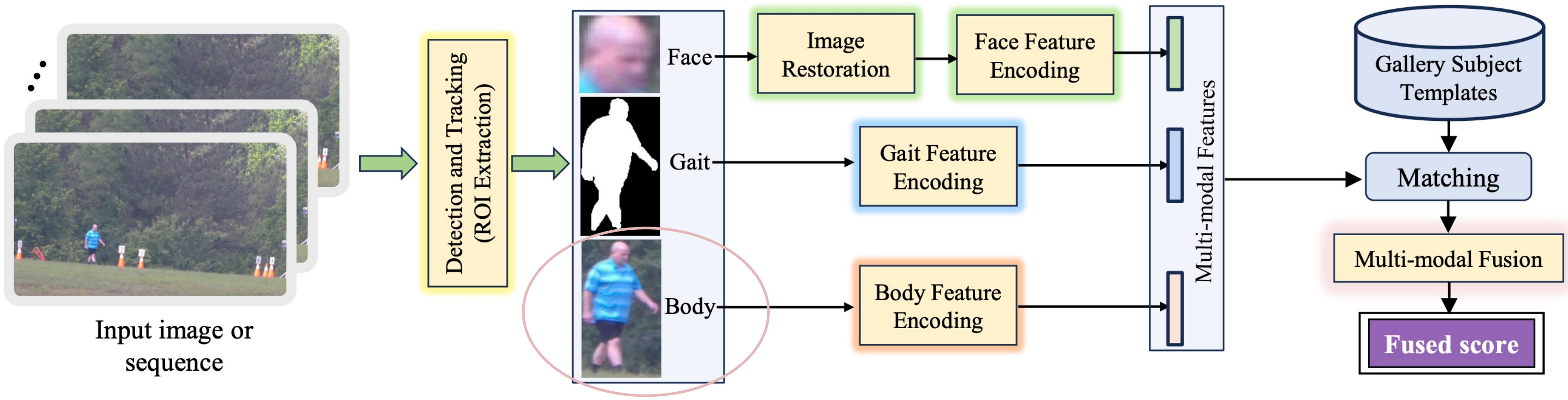


- **2. Domain adaption:**
  CFSM (ECCV'22)

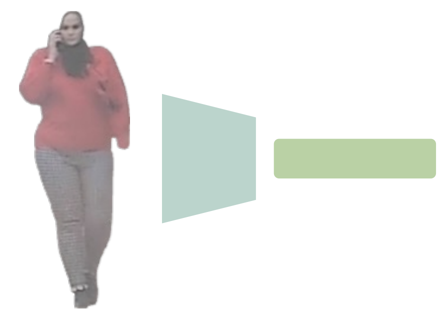
- **3. Video-based recognition:**
  CAFace (NeurIPS'22)



Person Identification at a (far) distance

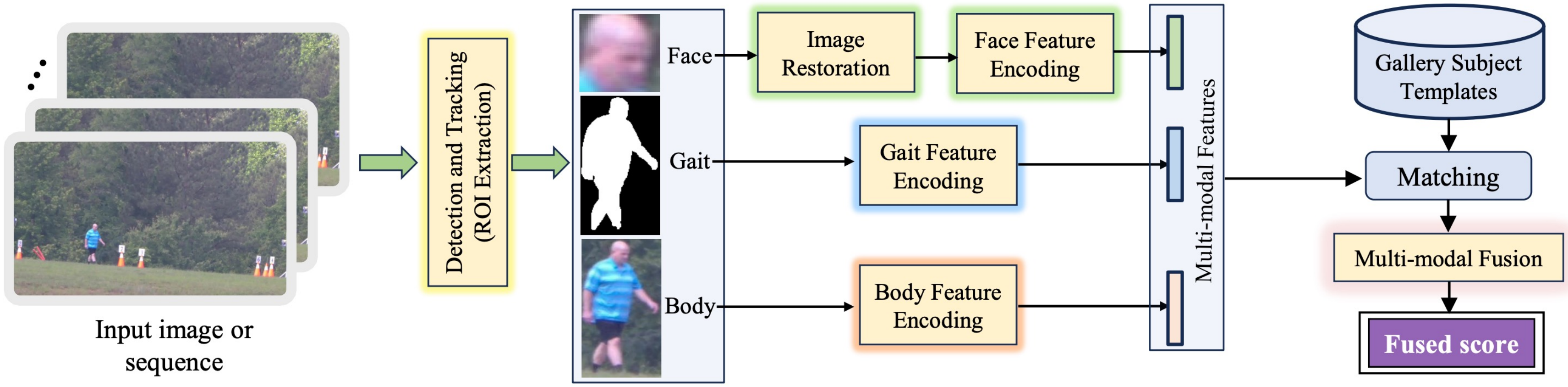


- **4. 3D body matching** (ICCV'23)

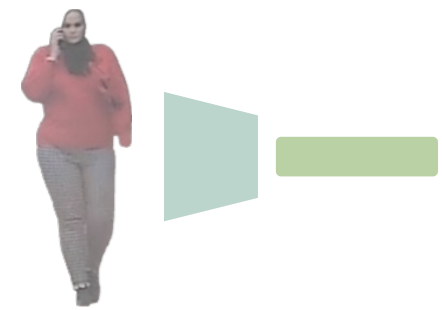


The subject (ID: G01003) consented to publication.

Person Identification at a (far) distance



- **4. 3D body matching** (ICCV'23)

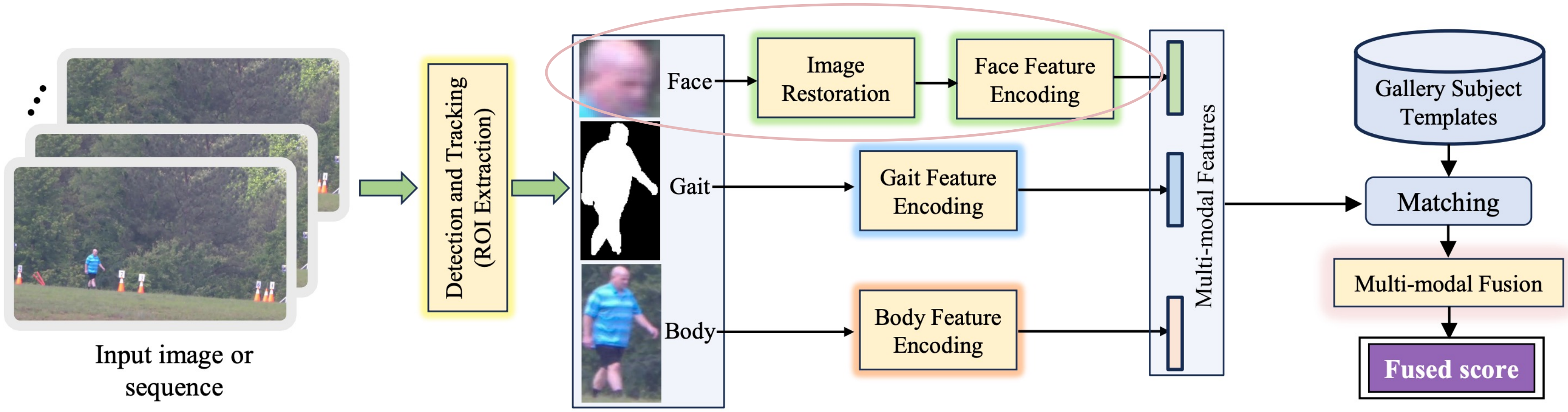


- **5. Synthetic training dataset** (CVPR'23)

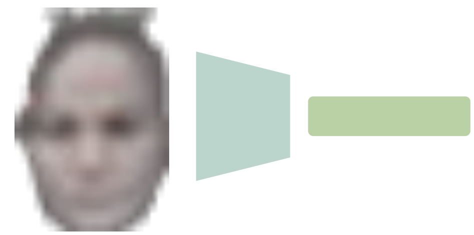
MICHIGAN STATE UNIVERSITY

The subject (ID: G01003) consented to publication.

Person Identification at a (far) distance



- **1. Generic matcher:** AdaFace (CVPR'22)

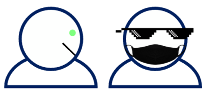


MICHIGAN STATE UNIVERSITY

The subject (ID: G01003) consented to publication.

Problem Definition

Training Datasets have Varying **Qualities**

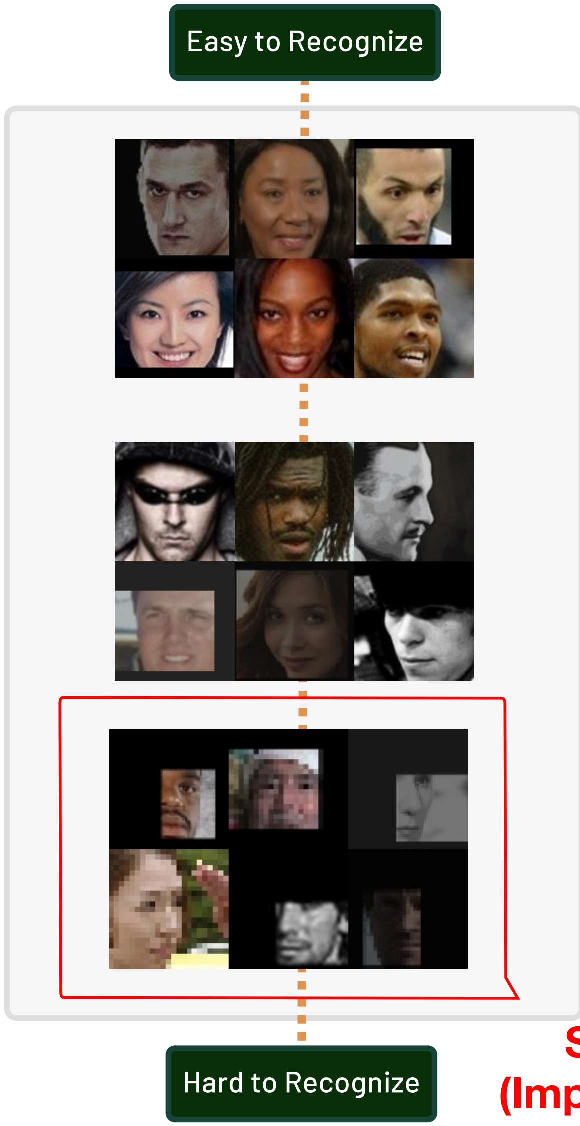


Pose and Occlusion

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

Landmark & Key points

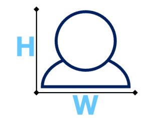
Images with visible and detectable facial landmarks are identifiable.



Source of Problem (Impossible to recognize)

Blur

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



Size

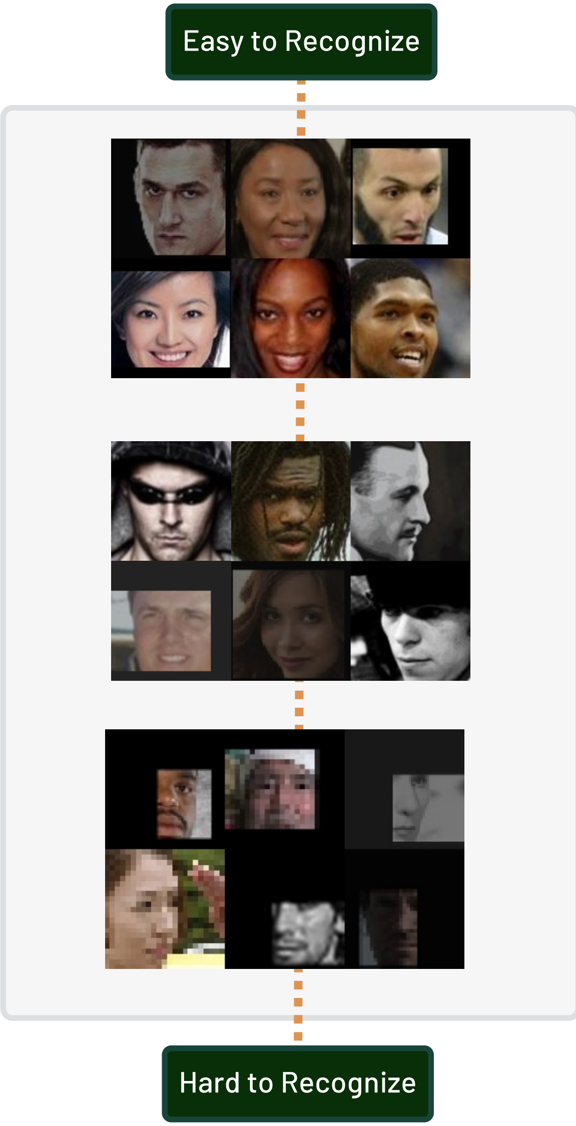
Too low image resolution causes the subject to be unidentifiable.

Illumination

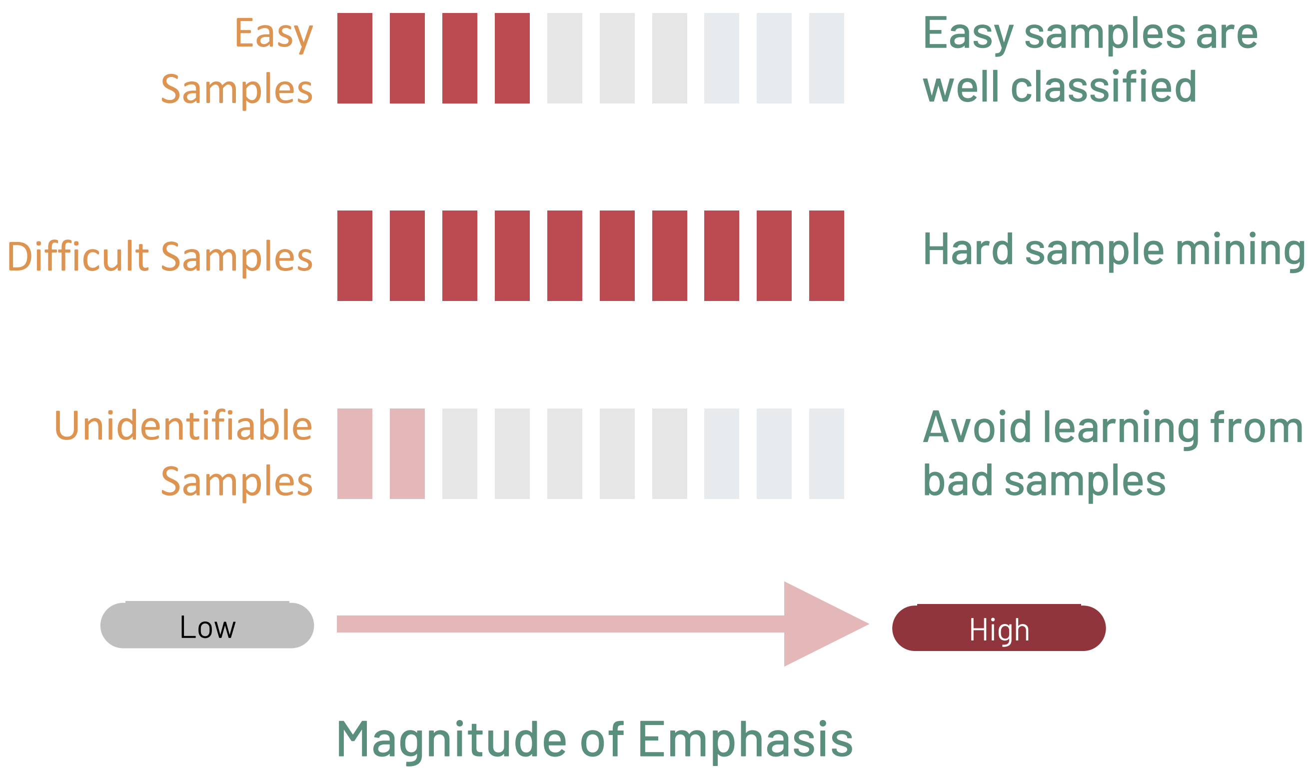
Too dark or too bright images cause the subject to be unidentifiable.

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

Motivation



Adaptive Sample Emphasis During Training



Motivation

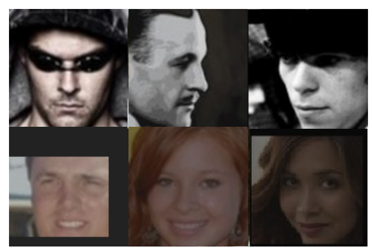
One More Way to Look at an Image

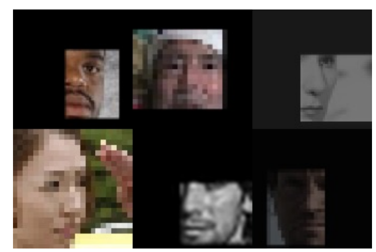
High Quality

Quality

Low Quality

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





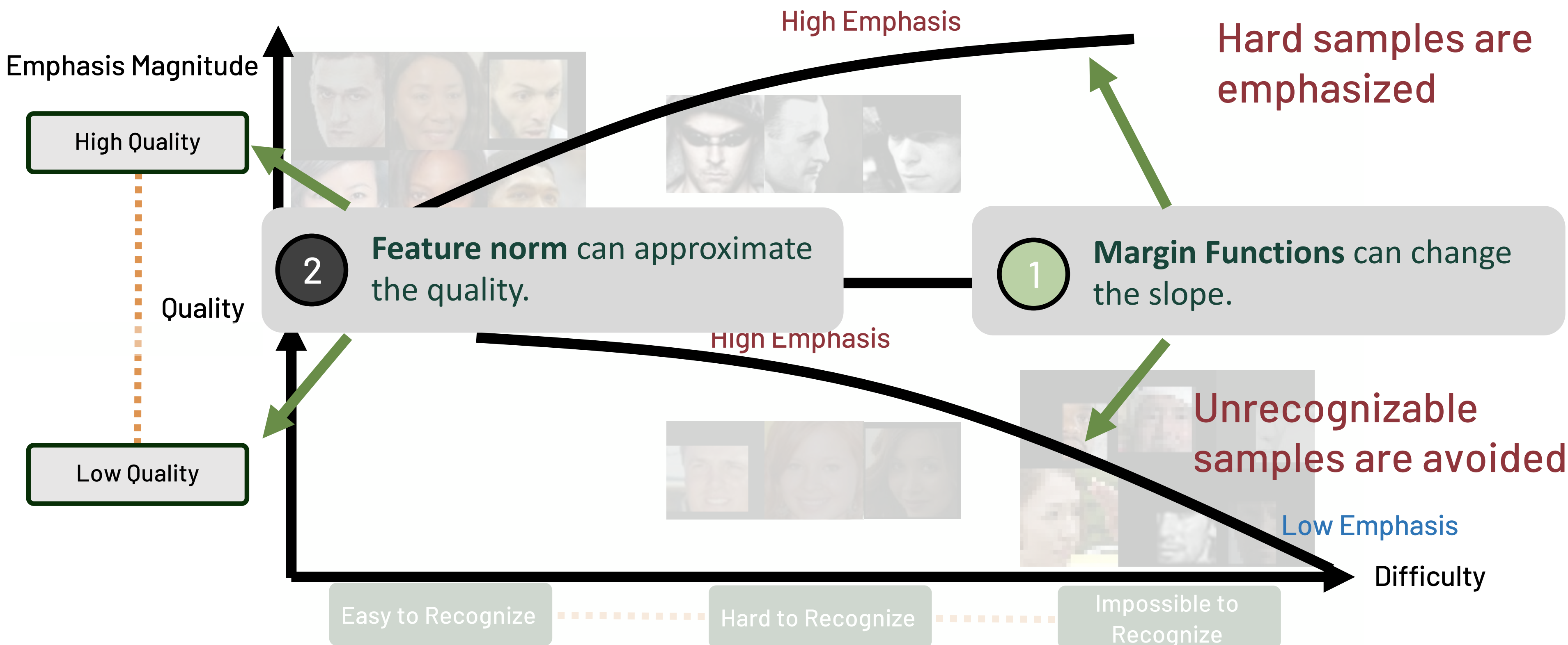
Easy to Recognize

Hard to Recognize

Impossible to Recognize

Difficulty

Our Findings and Methods



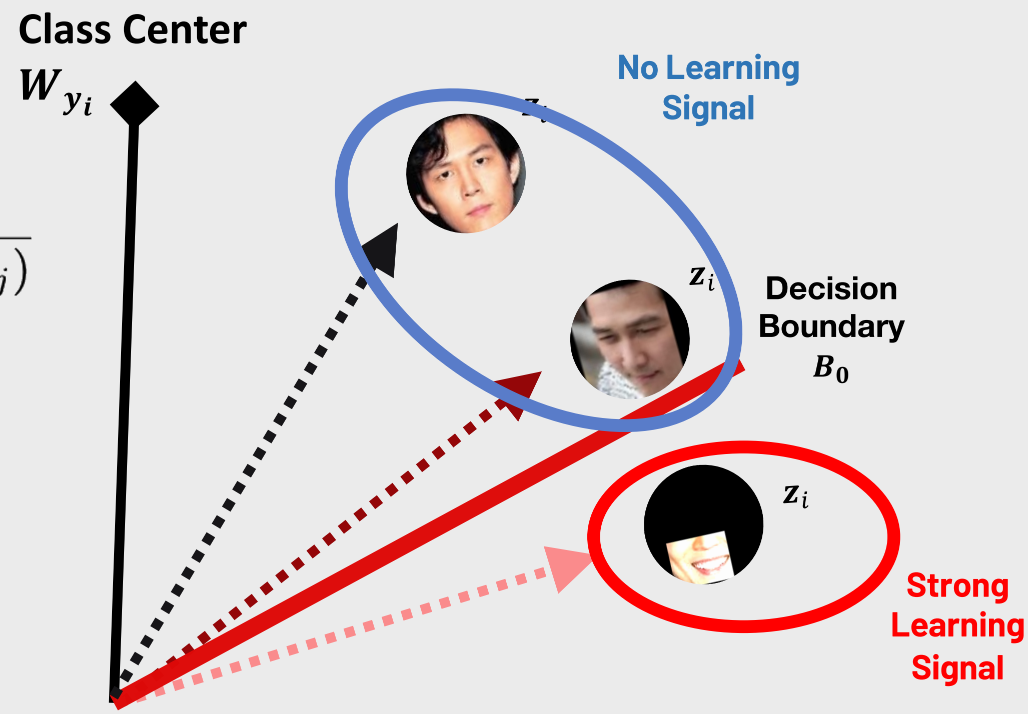
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

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

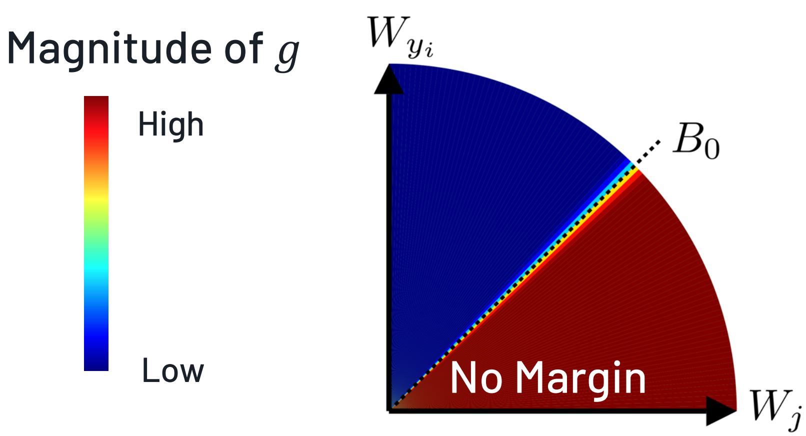
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos\theta_{y_i}) - m) & j = y_i \\ s \cos\theta_{y_i} & j \neq y_i \end{cases}$$

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



Plot of Gradient Scaling Term

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



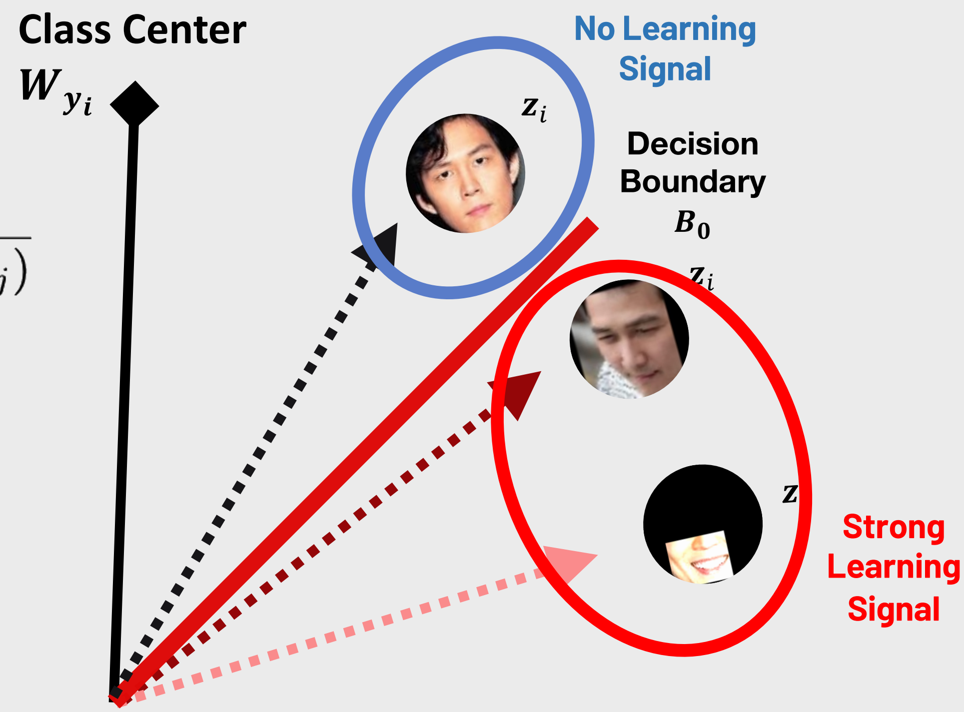
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

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

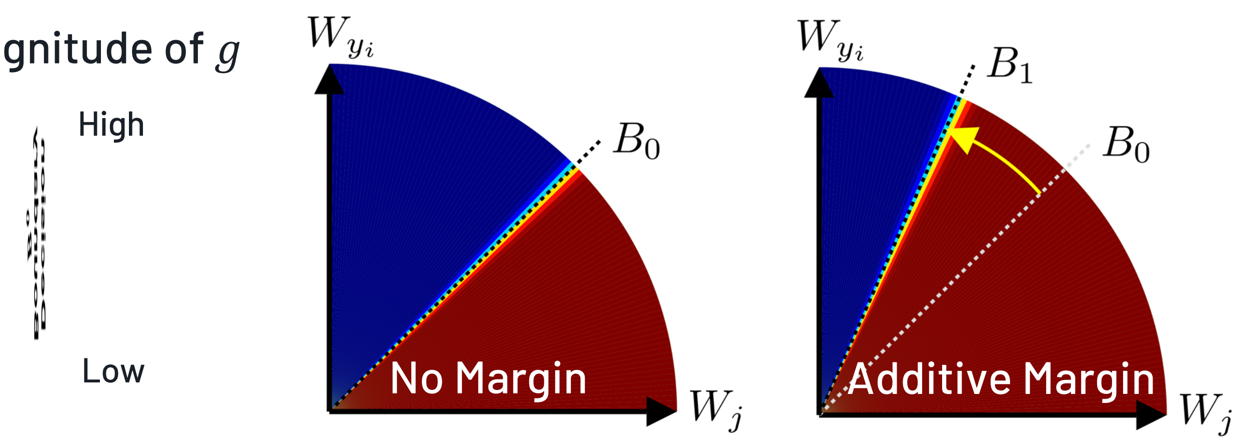
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos \theta_{y_i}) - m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$

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



Plot of Gradient Scaling Term

Magnitude of g



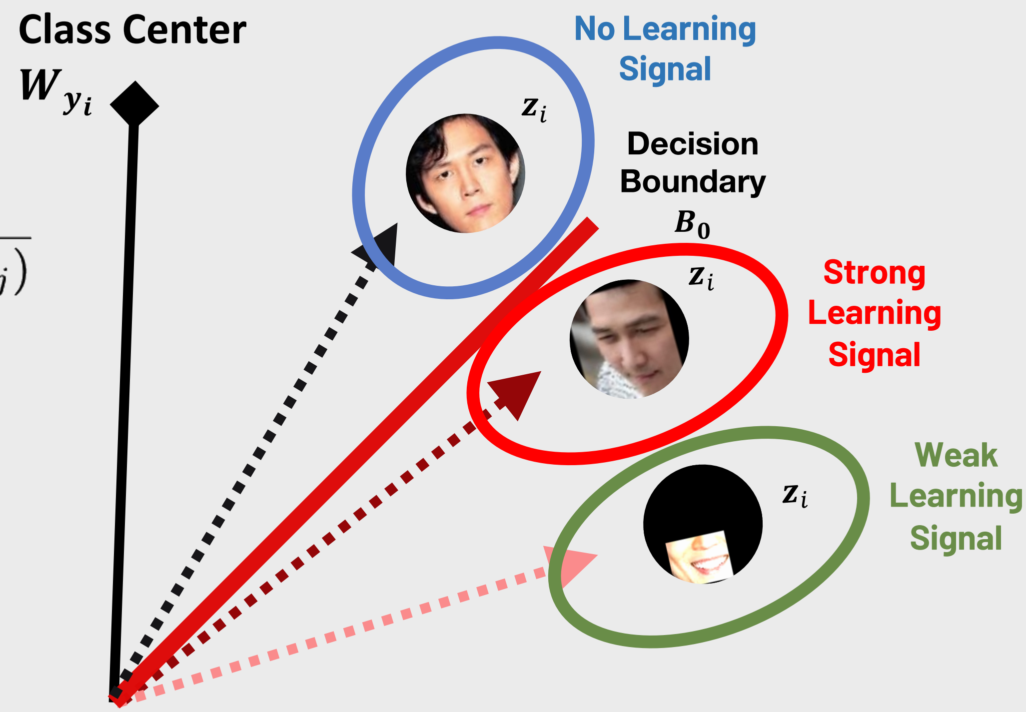
Effect of Margin on Sample Emphasis

Margin-based SoftMax Loss

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

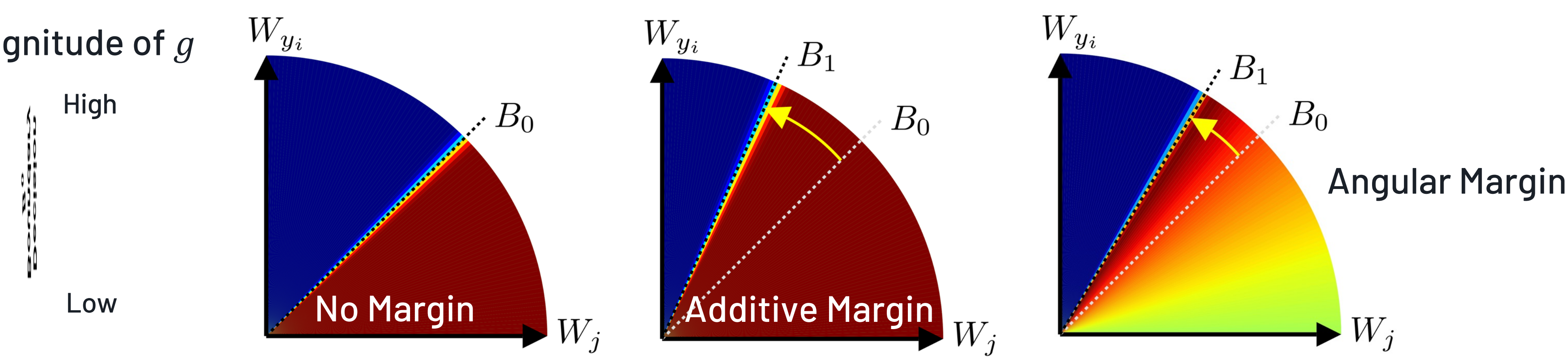
$$f(\cdot)_{\text{Additive}} = \begin{cases} s((\cos \theta_{y_i}) - m) & j = y_i \\ s \cos \theta_{y_i} & j \neq y_i \end{cases}$$

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



Plot of Gradient Scaling Term

Magnitude of g

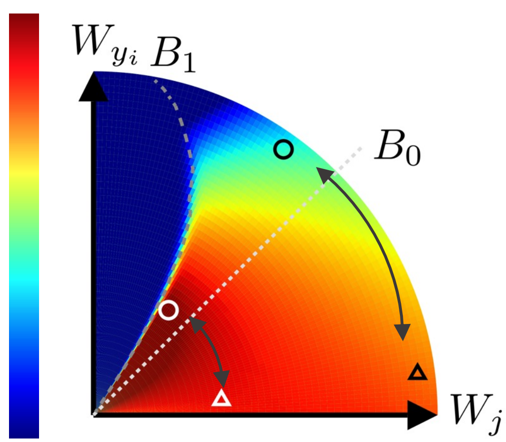


Method

How do we emphasize different samples?

AdaFace Objective

Magnitude of g



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

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

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

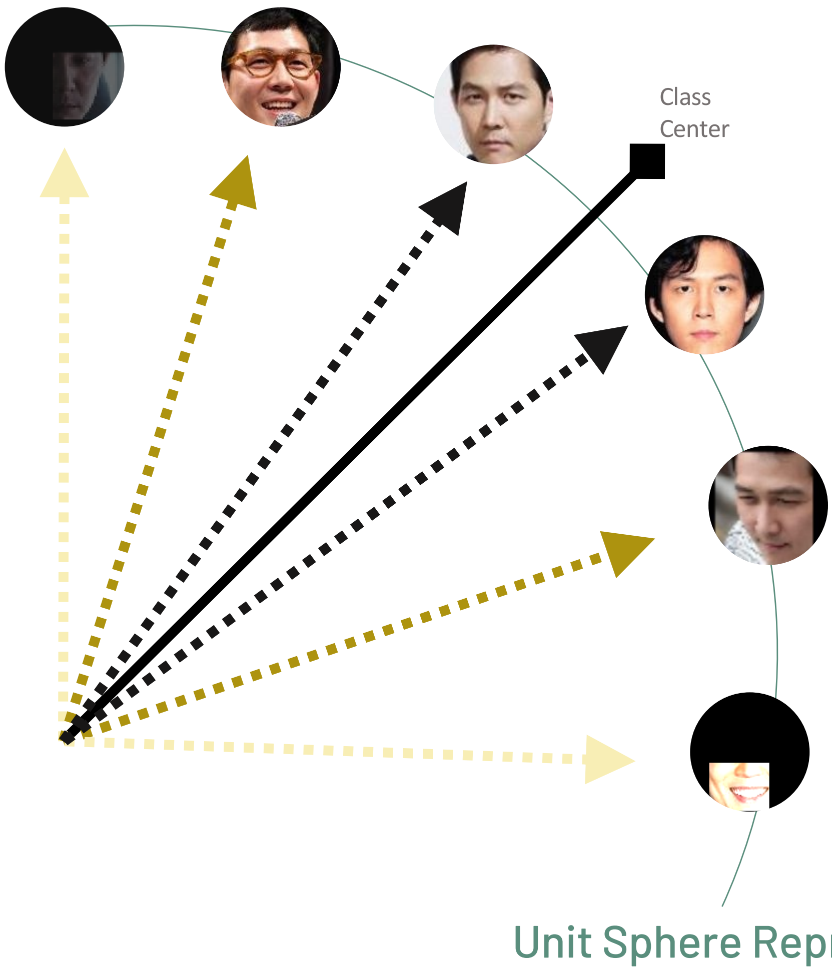
○ Easier Sample, Low Norm
△ Harder Sample, Low Norm

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

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

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

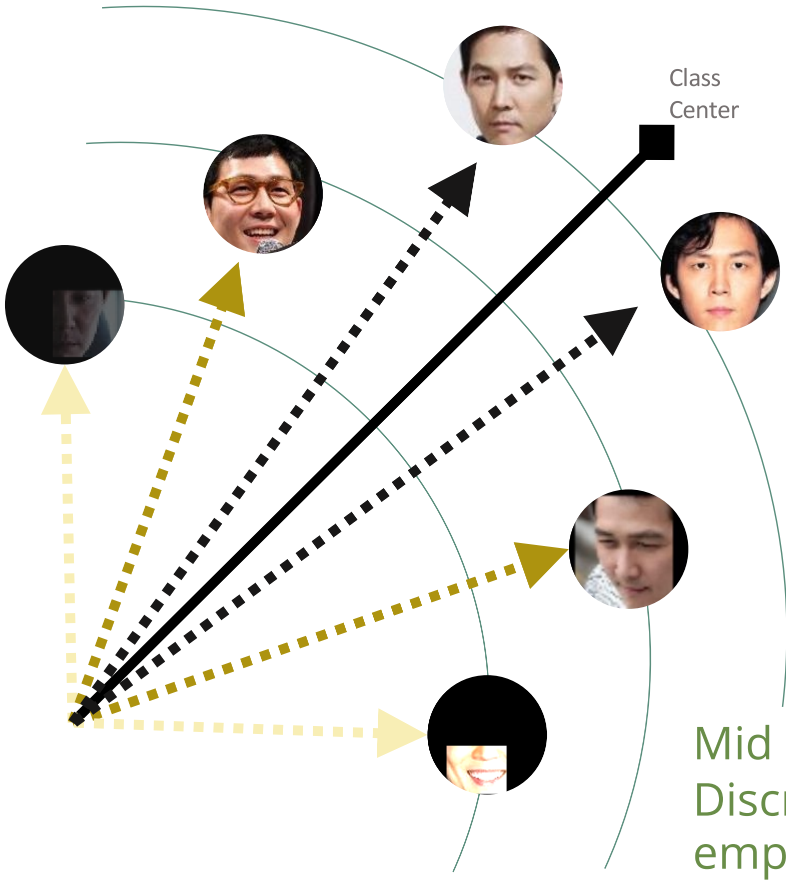
Method



Feature Space

Previous works apply
same margin for all samples

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

Performance

Performance in High Quality Datasets

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)

Best of:
- CosFace CVPR2018
- ArcFace CVPR2019
- CurricularFace CVPR2020
- MagFace CVPR2021
- SCF CVPR2021

Metric: 1:1 Verification Accuracy

Sample Images in high quality datasets

Performance

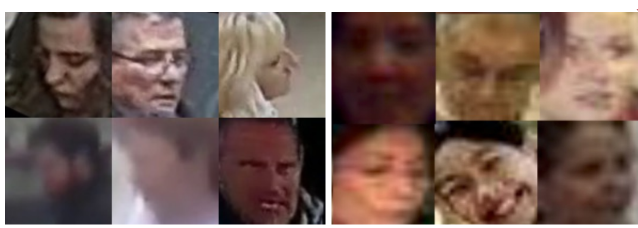
Performance in Low Quality Datasets

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)

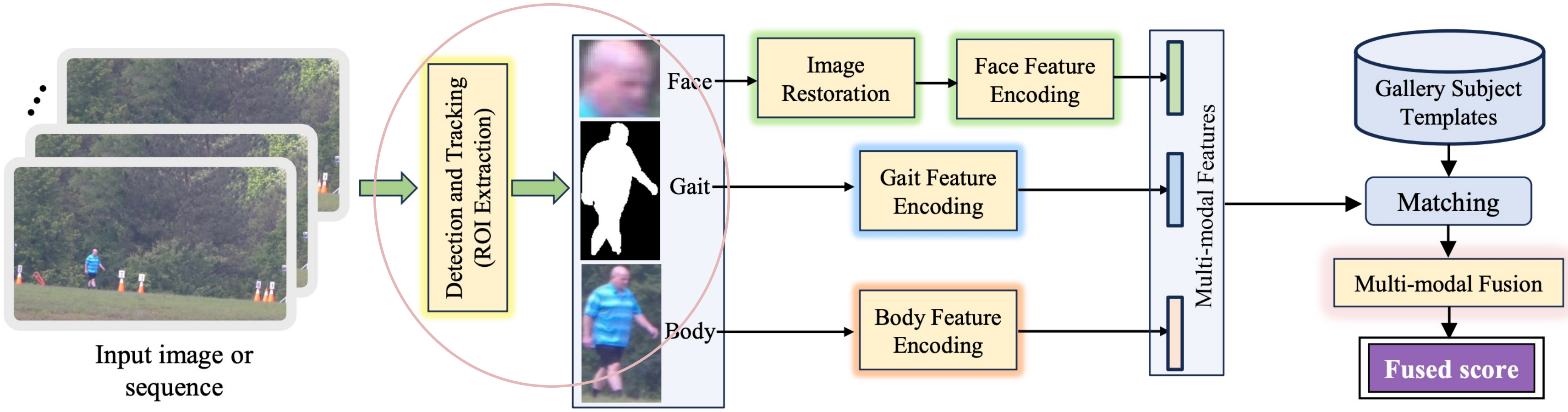
Metric: TPIR@FPIR=0.01%



Sample Images in low quality datasets

https://github.com/mk-minchul/AdaFace

Person Identification at a (far) distance

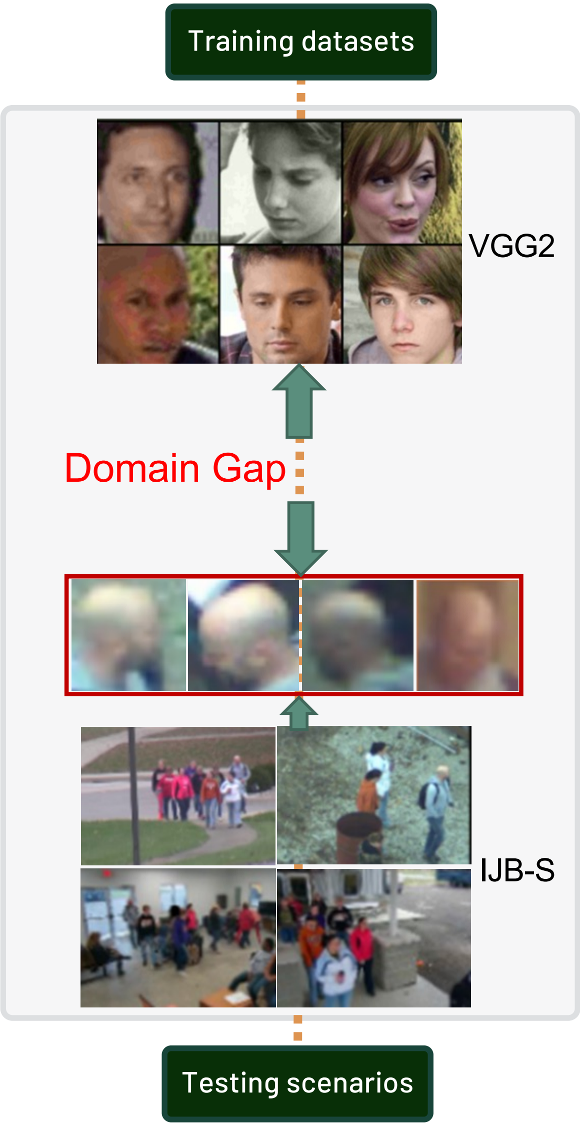


The subject (ID: G01003) consented to publication.

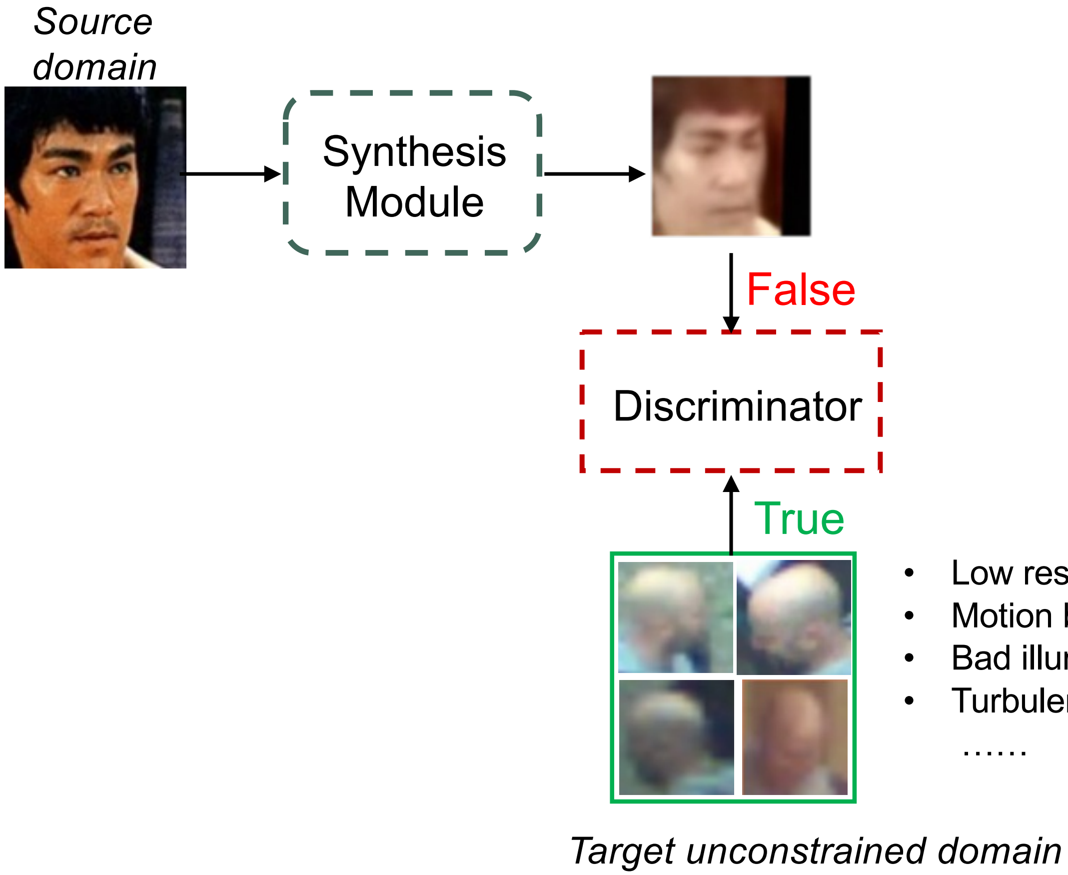
➢ **2. Domain adaption:** CFSM (ECCV'22)

vs

Motivation

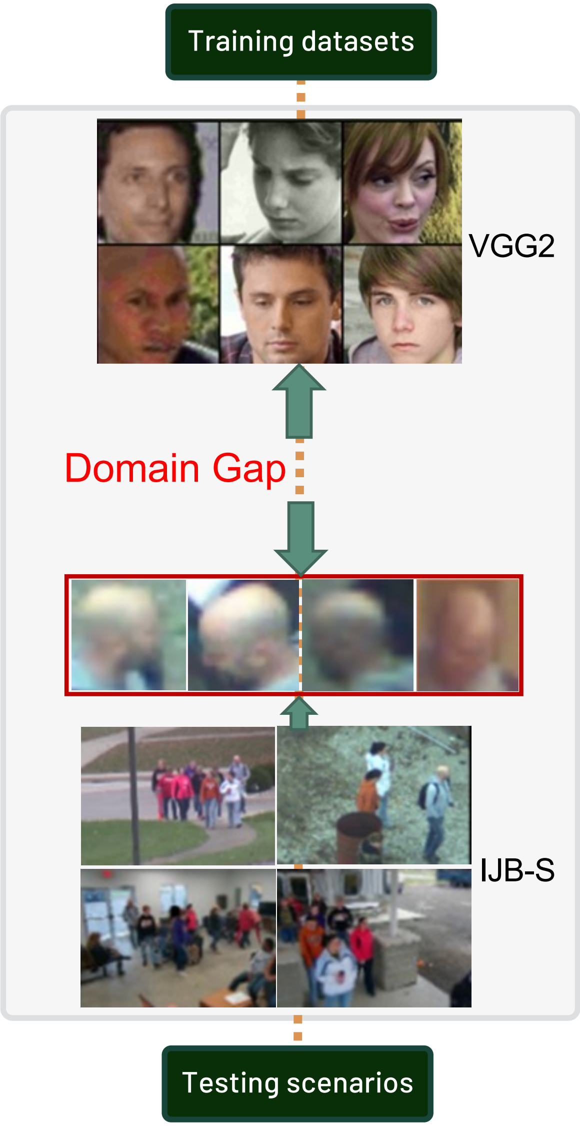


Feedback-based face synthesis

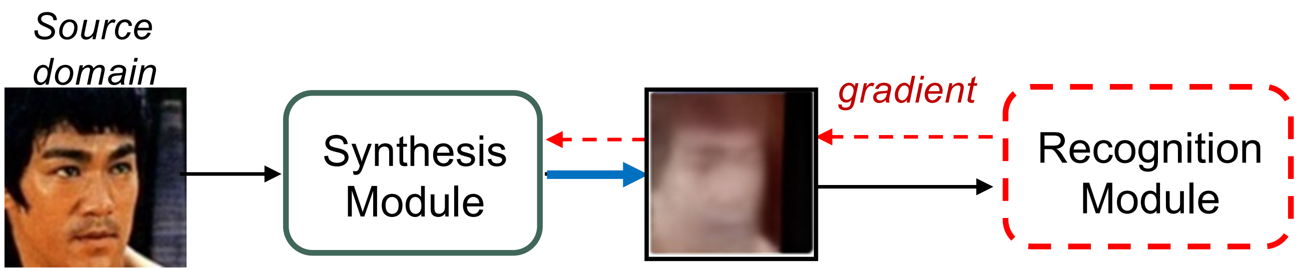


- Low resolution
- Motion blurring
- Bad illumination
- Turbulence effect

……

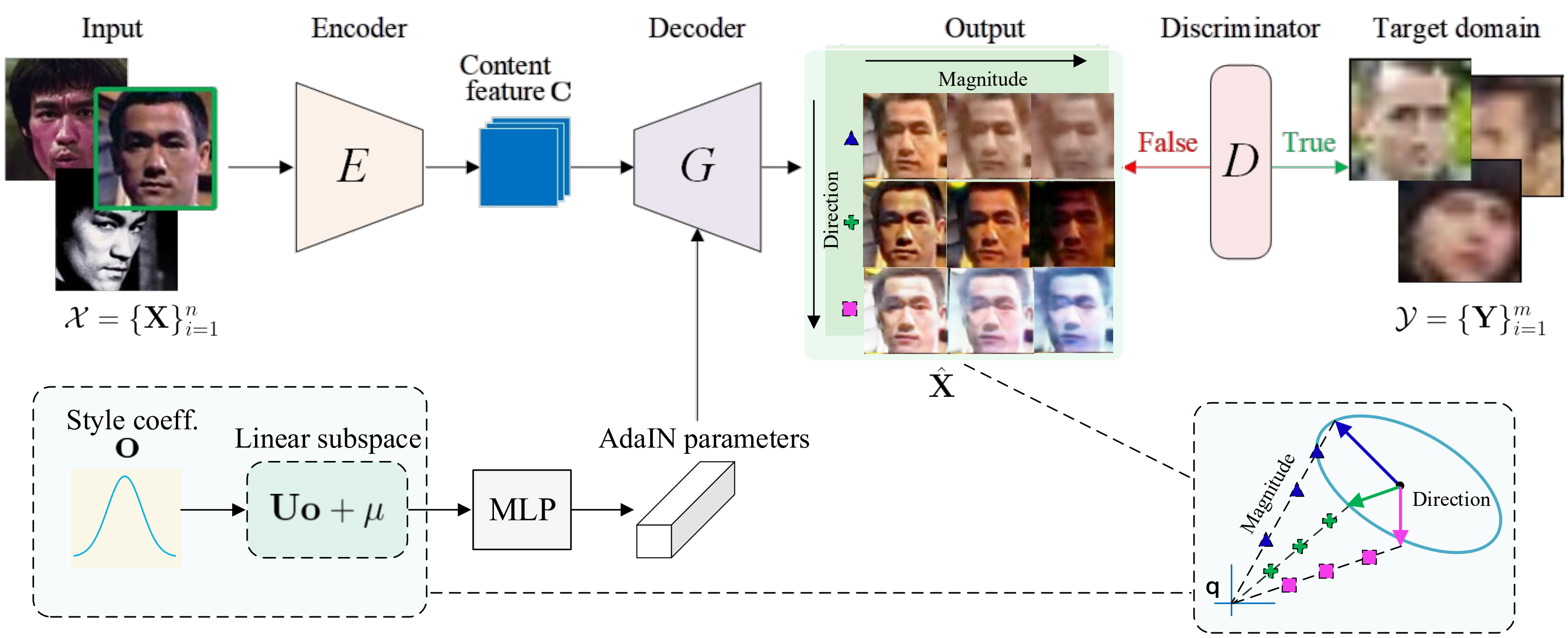


Feedback-based face synthesis

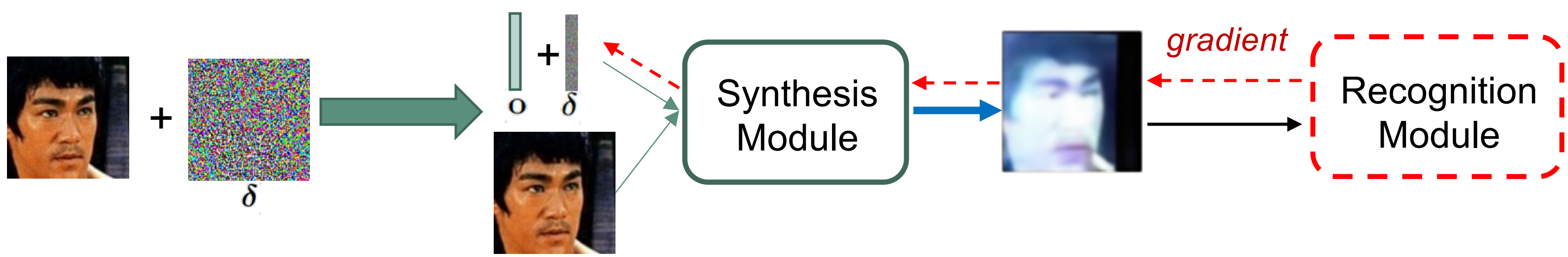


Controllable Face Synthesis Model (CFSM)

Precisely-controllable in the style latent space, in both diversity and degree

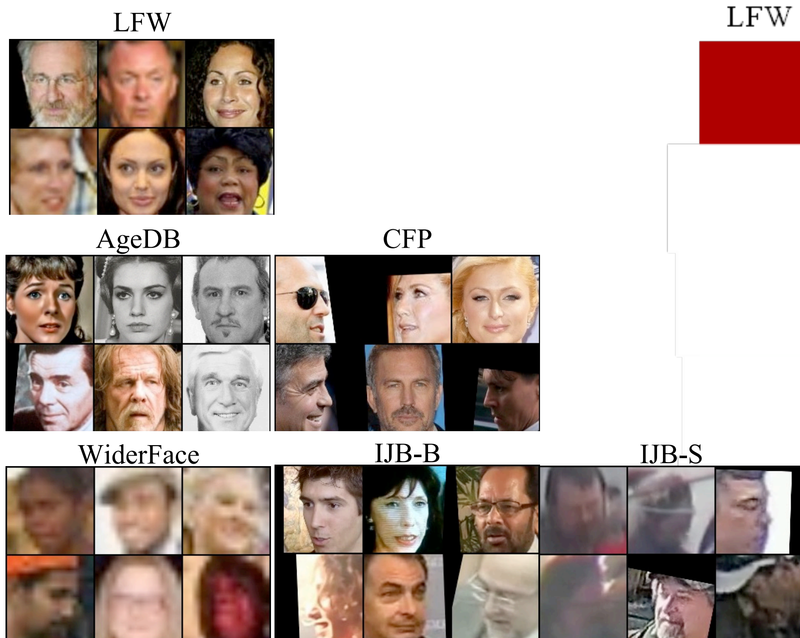


Introduce an adversarial regularization (style latent) to guide the face synthesis



Dataset Distribution Similarity Measurement

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



	LFW	AgeDB	CFP	WF	IJB-B	IJB-S
LFW		0.53	0.57	0.67	0.72	0.62
AgeDB			0.39	0.53	0.58	0.51
CFP				0.55	0.66	0.69
WF					0.72	0.72
IJB-B						0.77
IJB-S						

Face Recognition on Low Quality Datasets

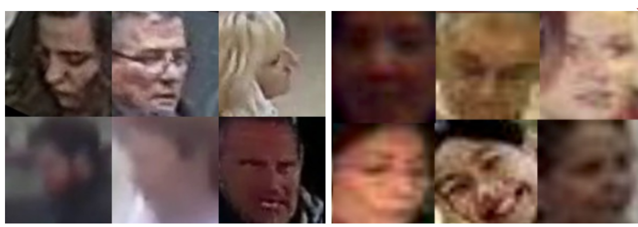
CFSM could be plugged into any SoTA FR model, e.g., AdaFace

Best of

- PFE CVPR2018
- ArcFace CVPR2019
- URL CVPR2020
- AdaFace CVPR2022

Best of	Dataset Names	AdaFace+ CFSM
71.35%	IJB-S Surveillance-to-Single	72.54% (+1.19)
71.93%	IJB-S Surveillance-to-Booking	72.65% (+0.72)
36.71%	IJB-S Surveillance-to-Surveillance	39.14% (**+2.43**)
72.29%	TinyFace	73.87% (+1.58)

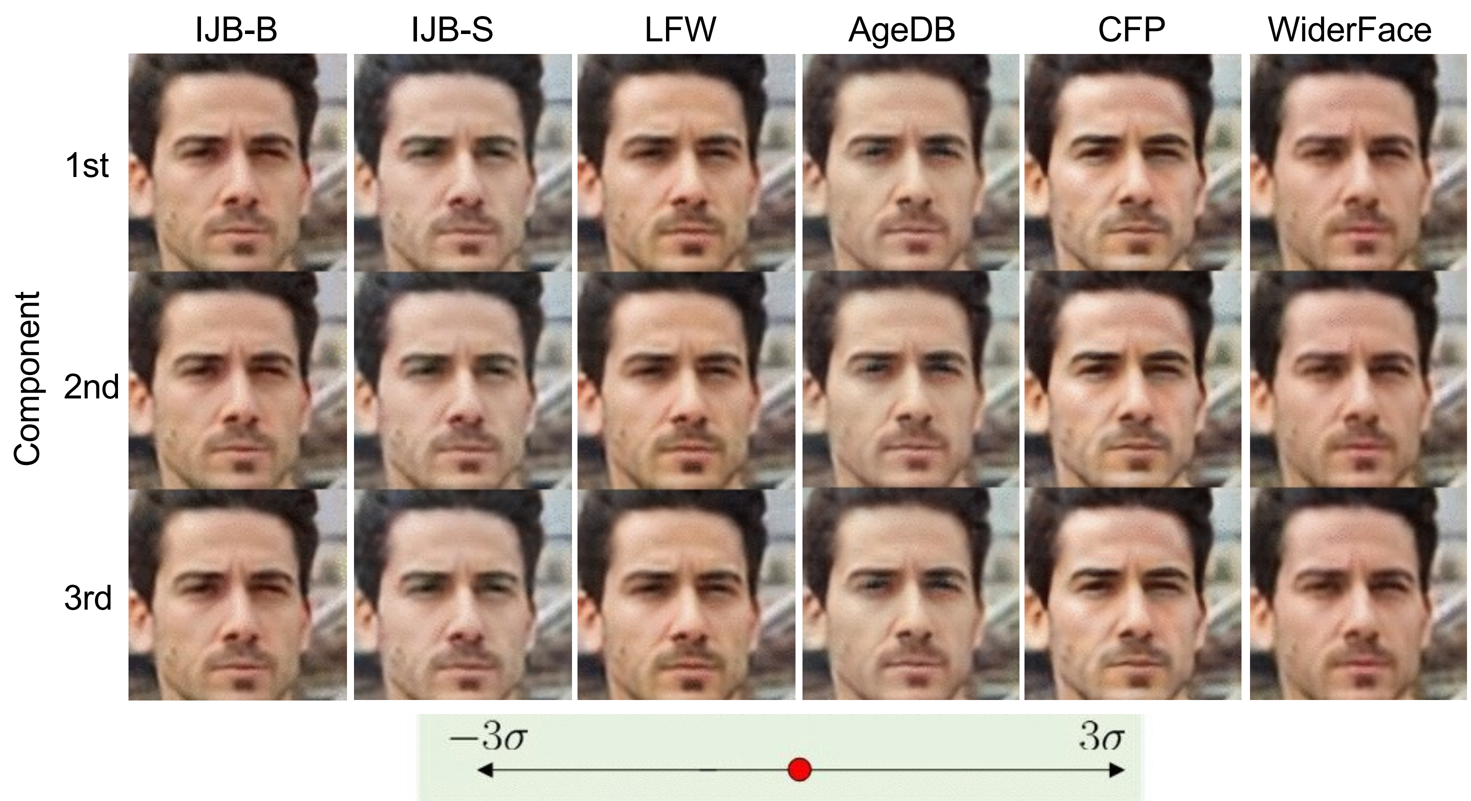
Metric: Rank1



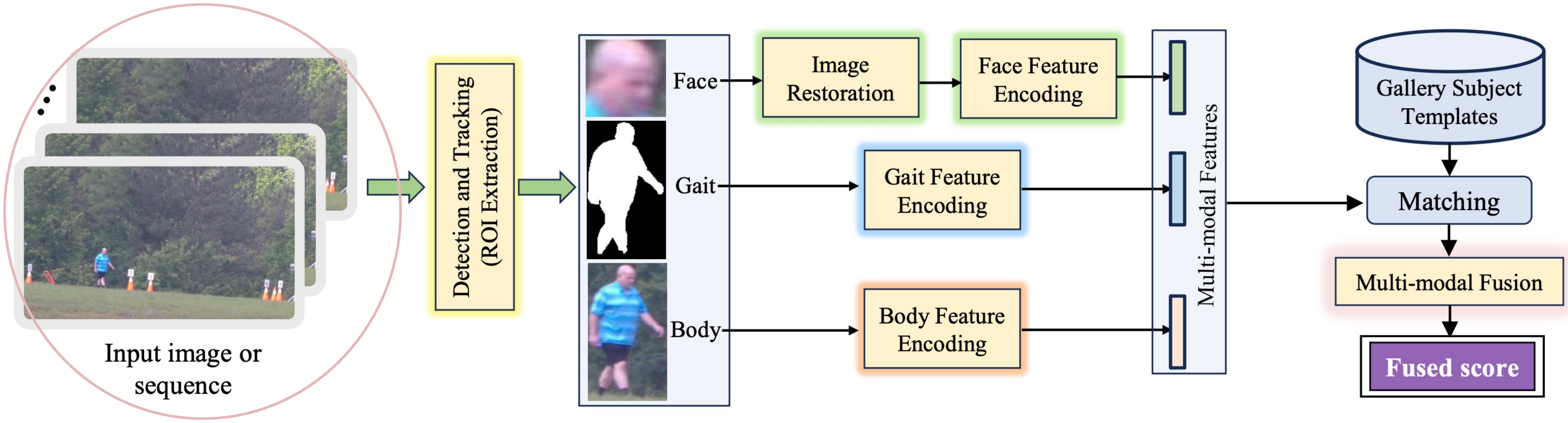
Sample Images in low quality datasets

Visualizations of the Face Synthesis Model

Interpretable magnitude of the style coefficient

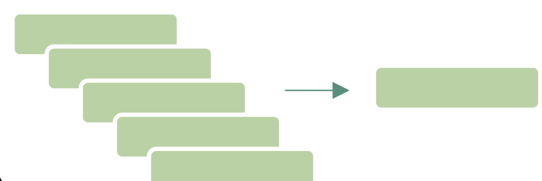


Person Identification at a (far) distance



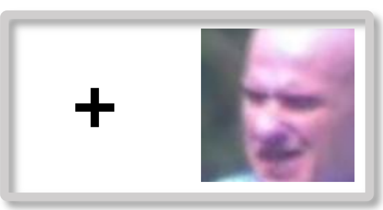
The subject (ID: G01003) consented to publication.

- **3. Video-based recognition:** CAFace (NeurIPS'22)

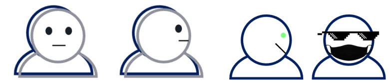


Problem Definition

Traits of Face Recognition with Videos



1



Varied Identifiability

Some images are more identifiable than others

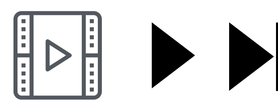
2

10 ~ 1,000,000

Varied Number of Images

Number of images are not fixed.

3

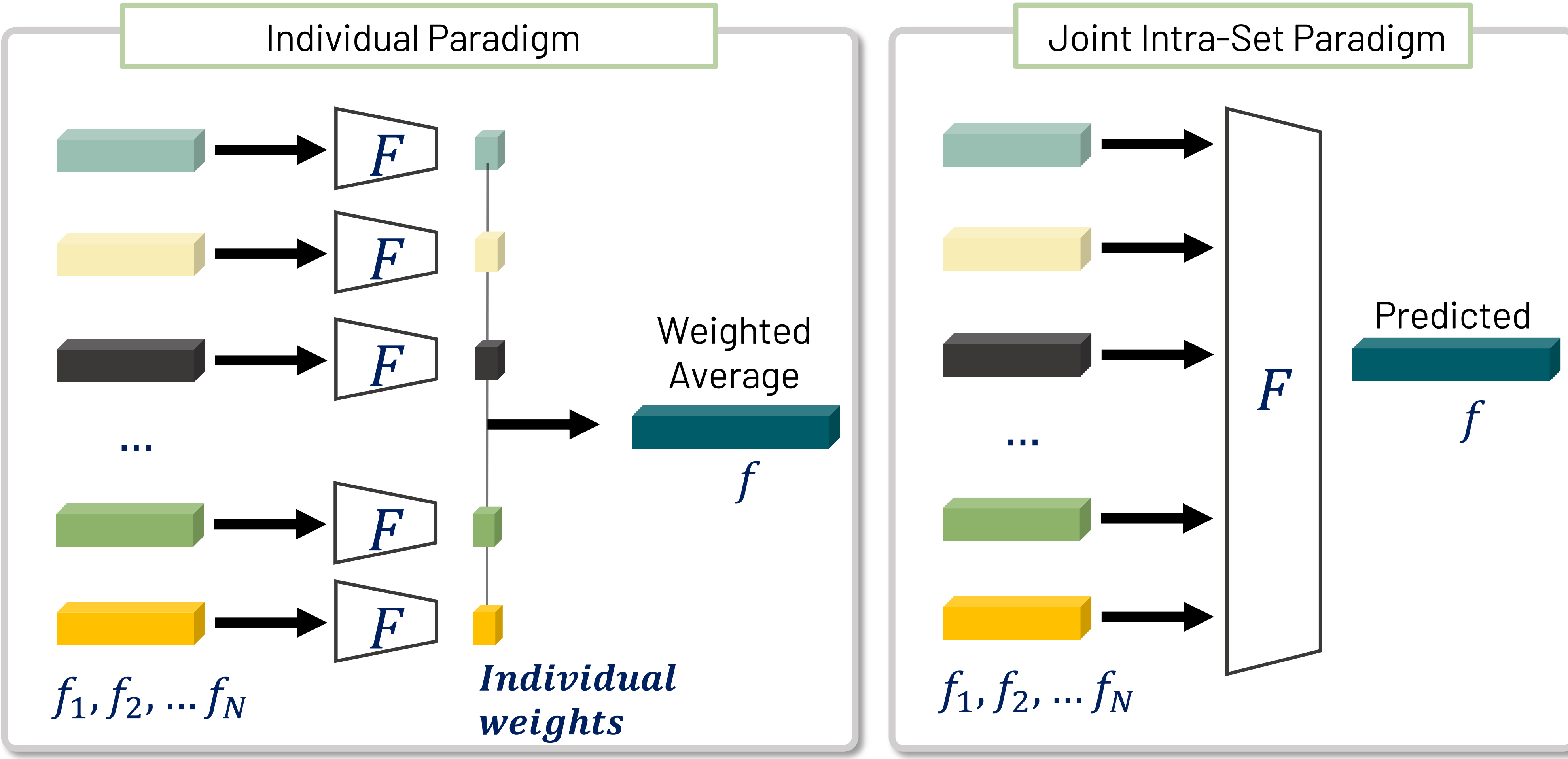


Sequential Inputs

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

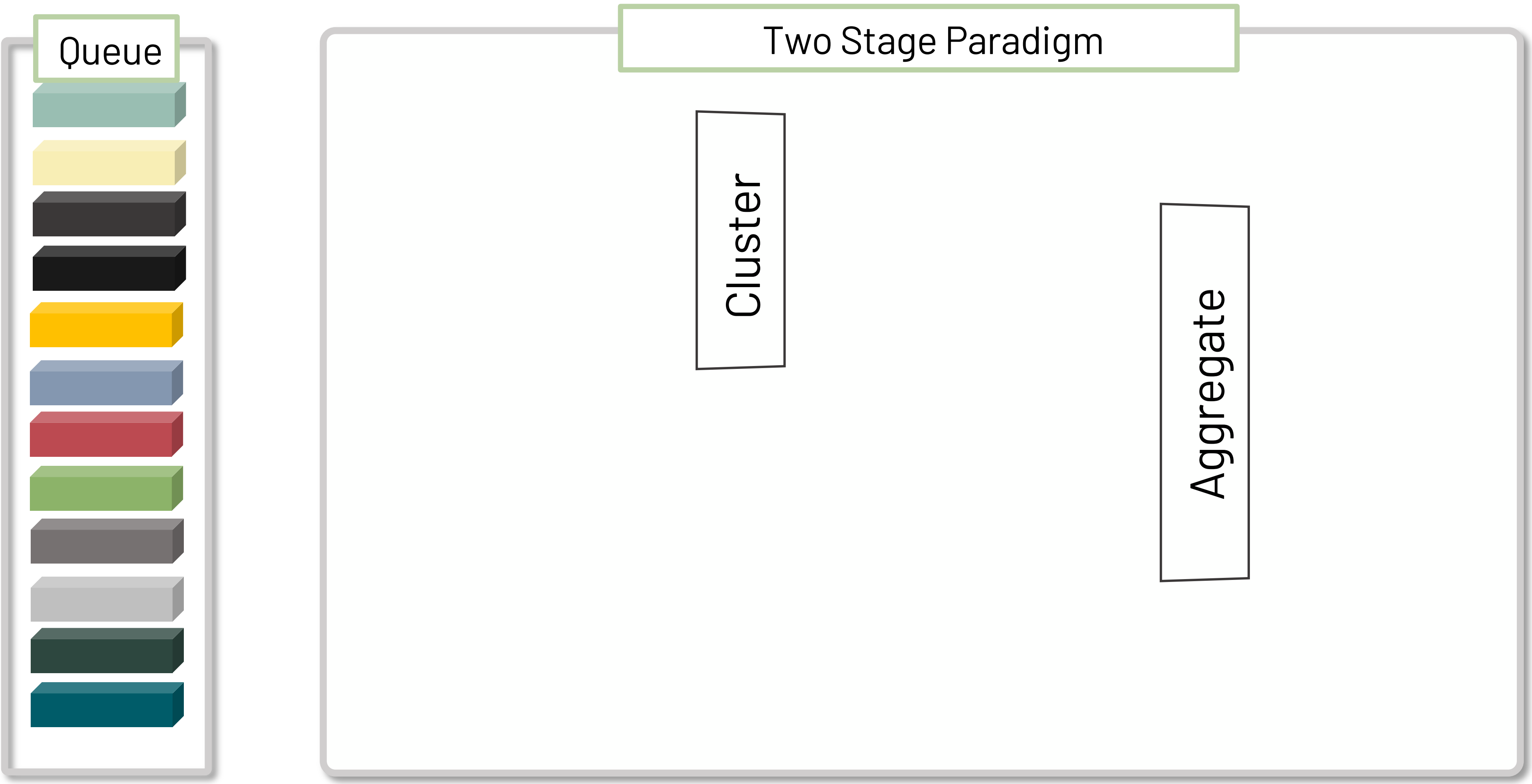


(No Intra—Set Relationship)

(Cannot handle large N)

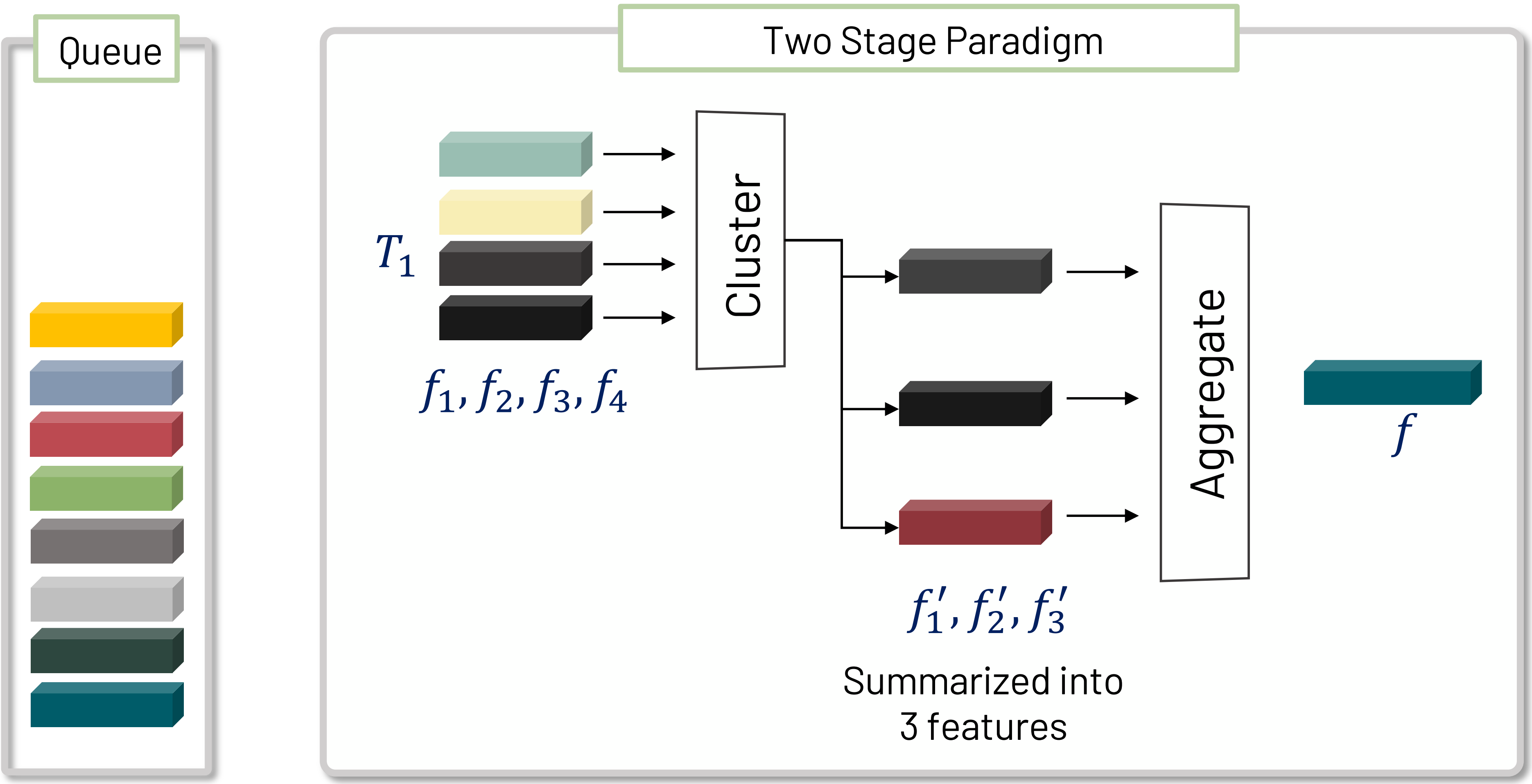
Motivation

Large N / Sequential Scenario



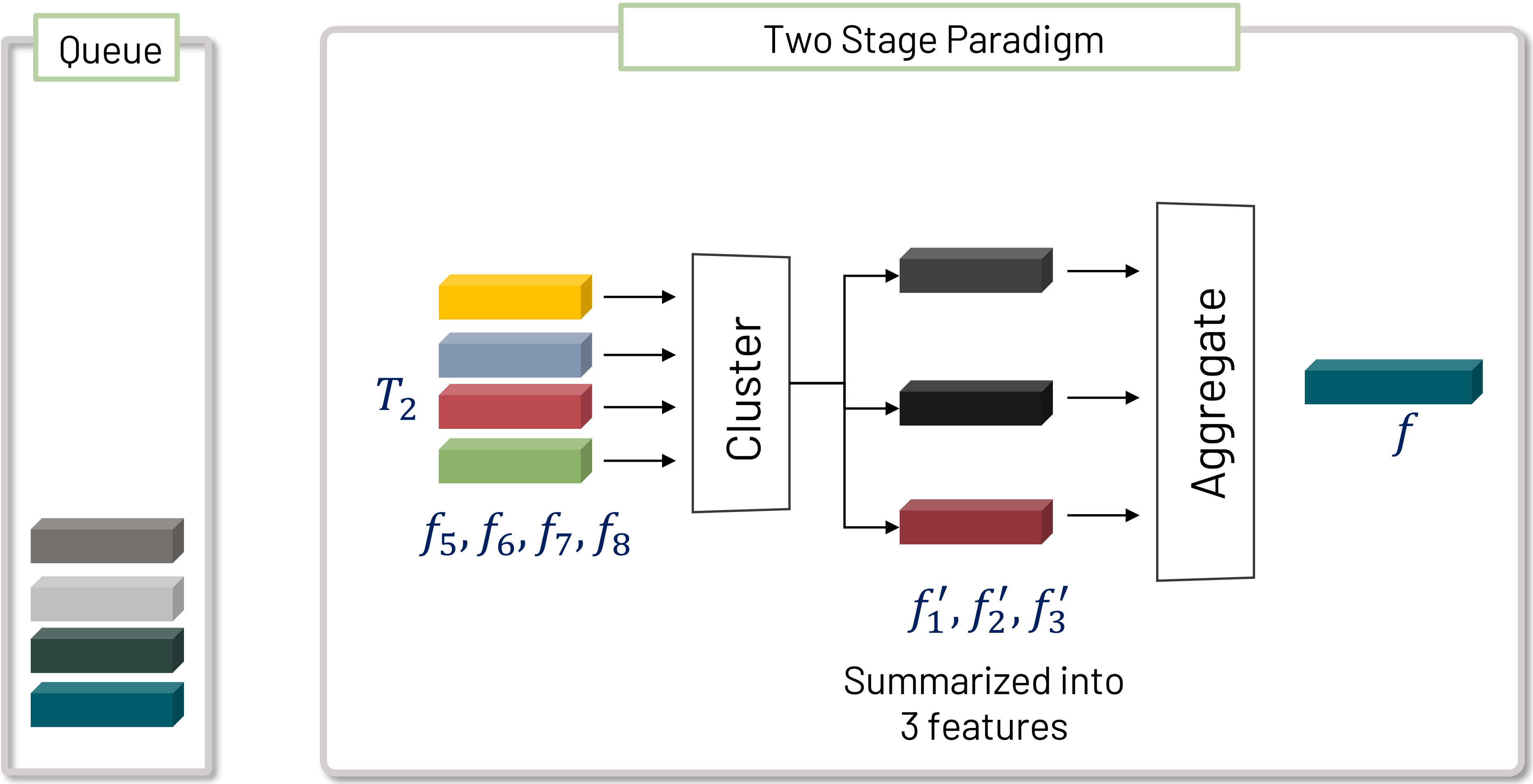
Motivation

Large N / Sequential Scenario



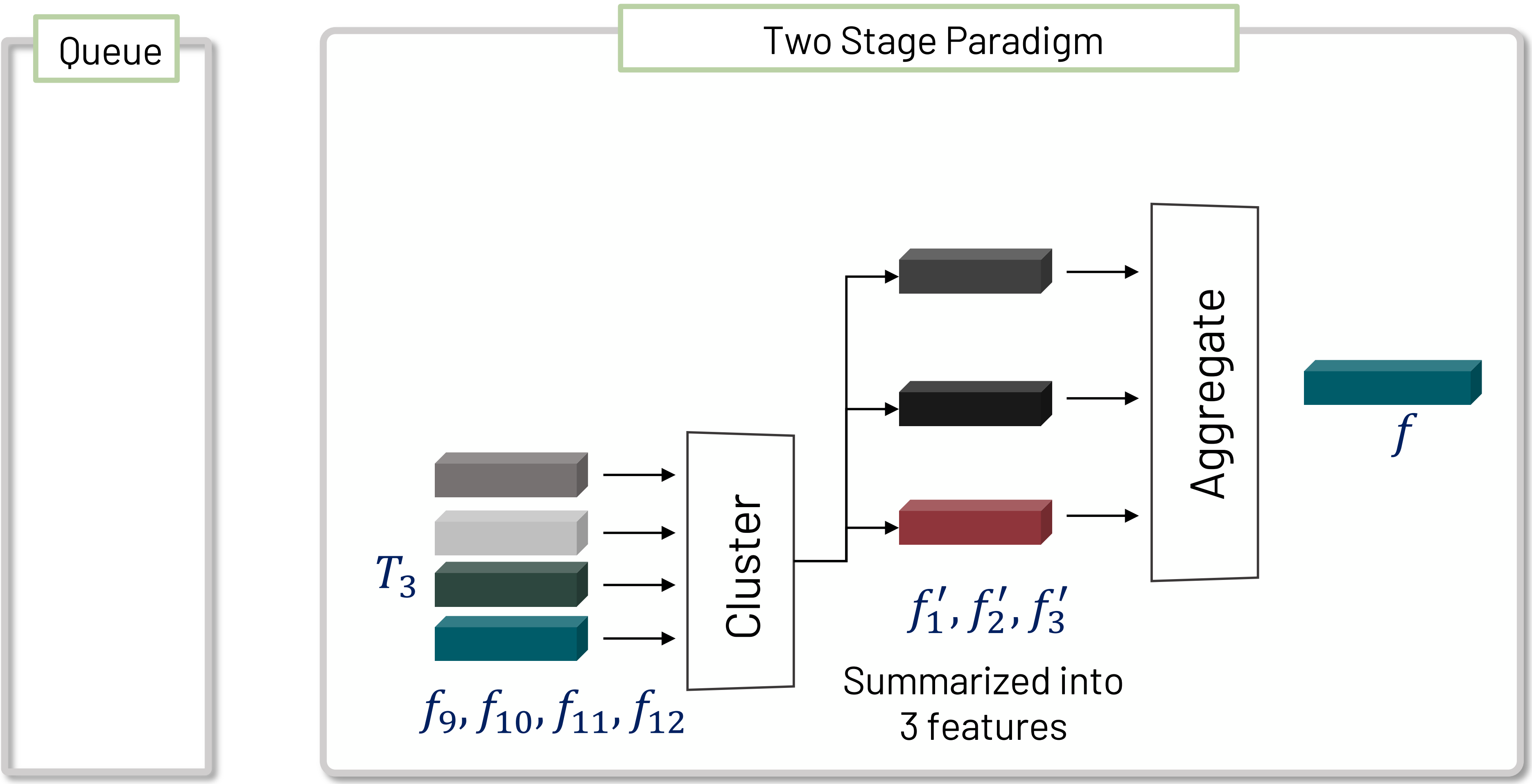
Motivation

Large N / Sequential Scenario



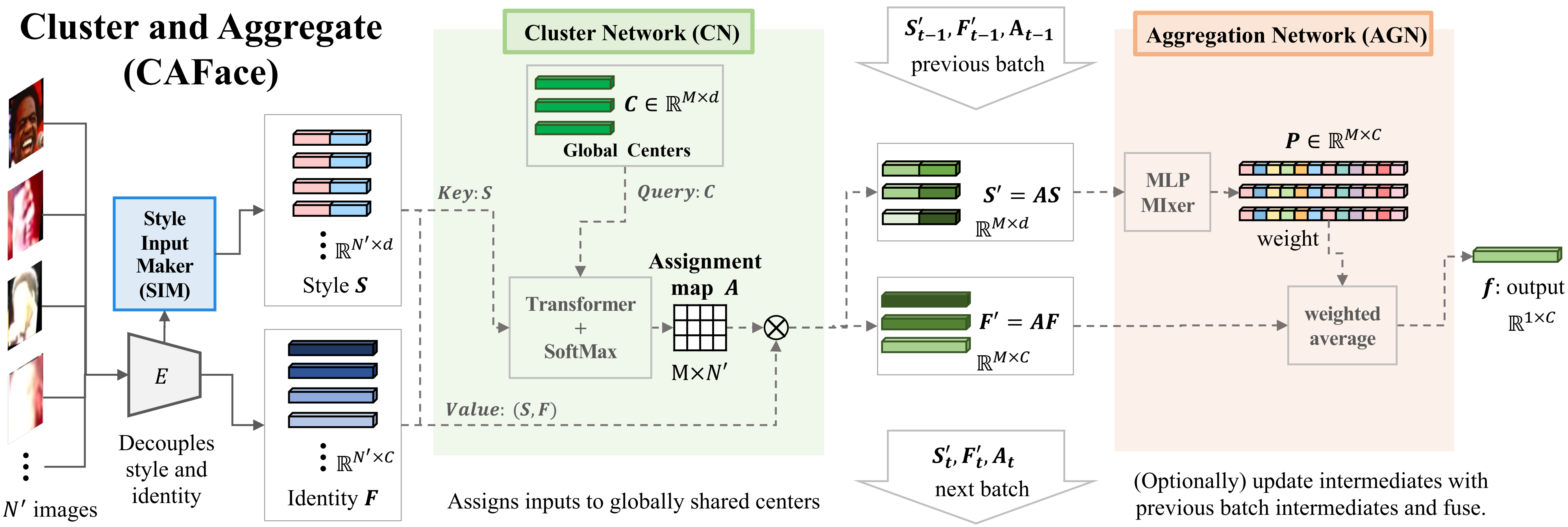
Motivation

Large N / Sequential Scenario



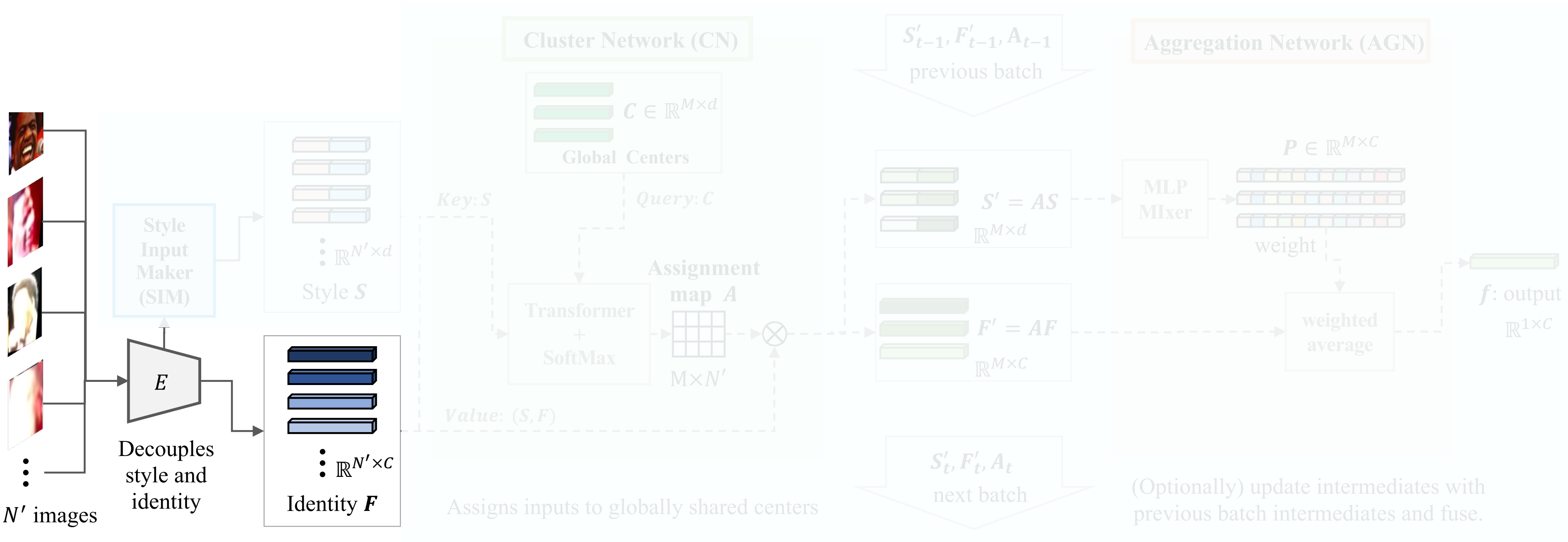
Method

Architecture



Overall Architectures
3 components (SIM, CN, AGN)

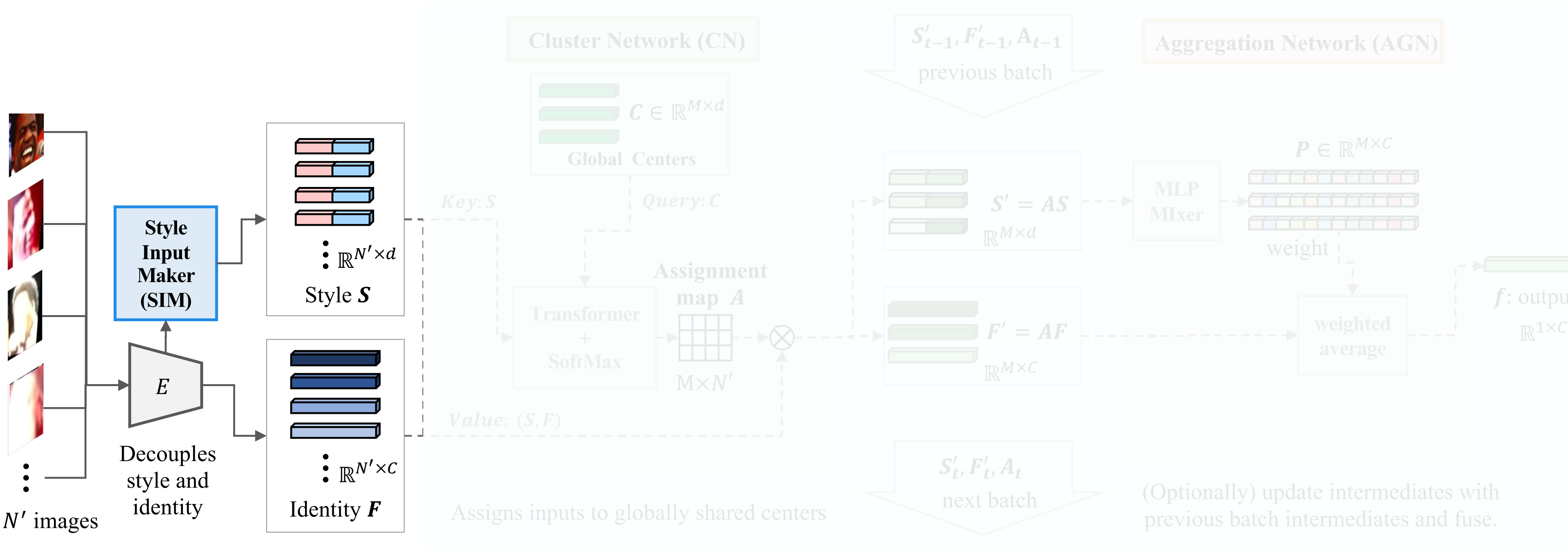
Architecture



Input images fed into the fixed feature extractor.

Method

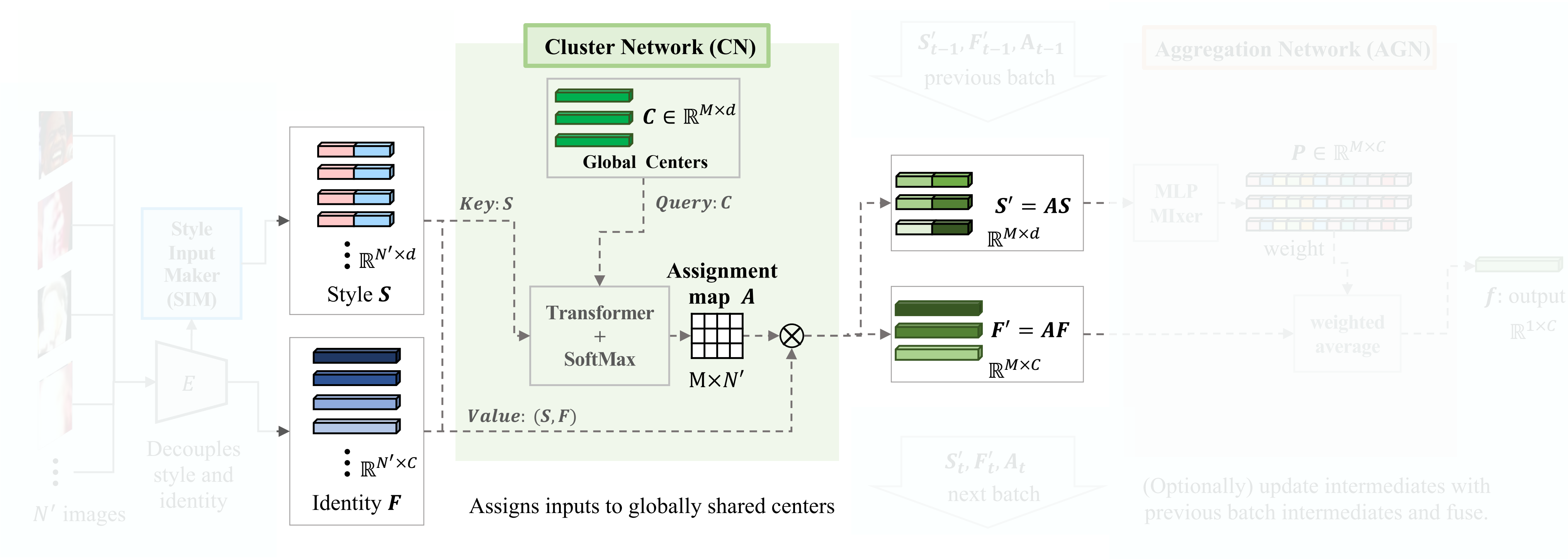
Architecture



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

Method

Architecture

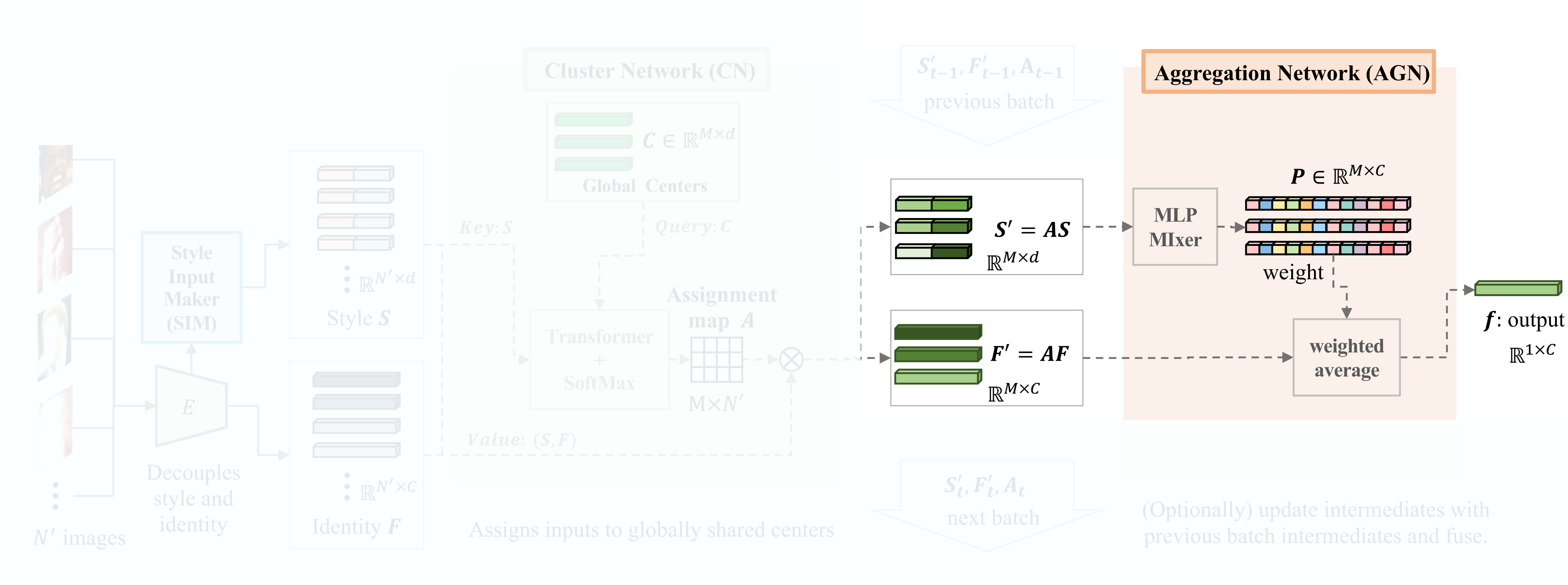


CN uses learned centers $\{c_j\}^M$ and $\{s_i\}^N$ to create assignment map A.

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

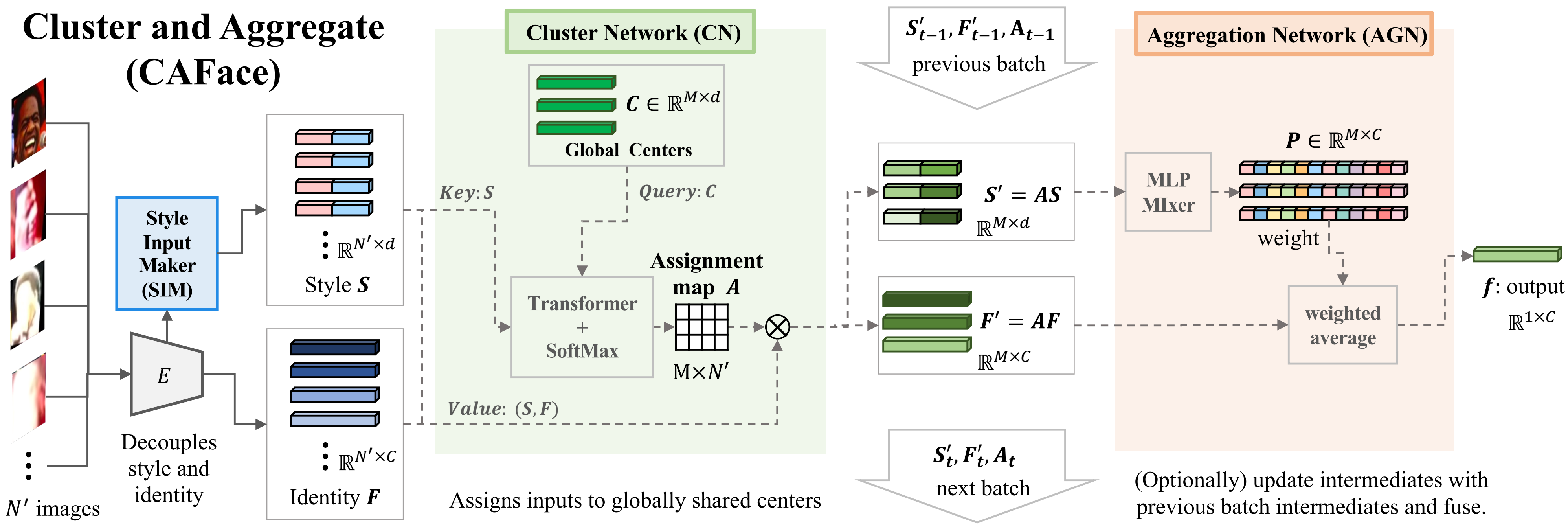
Method

Architecture



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

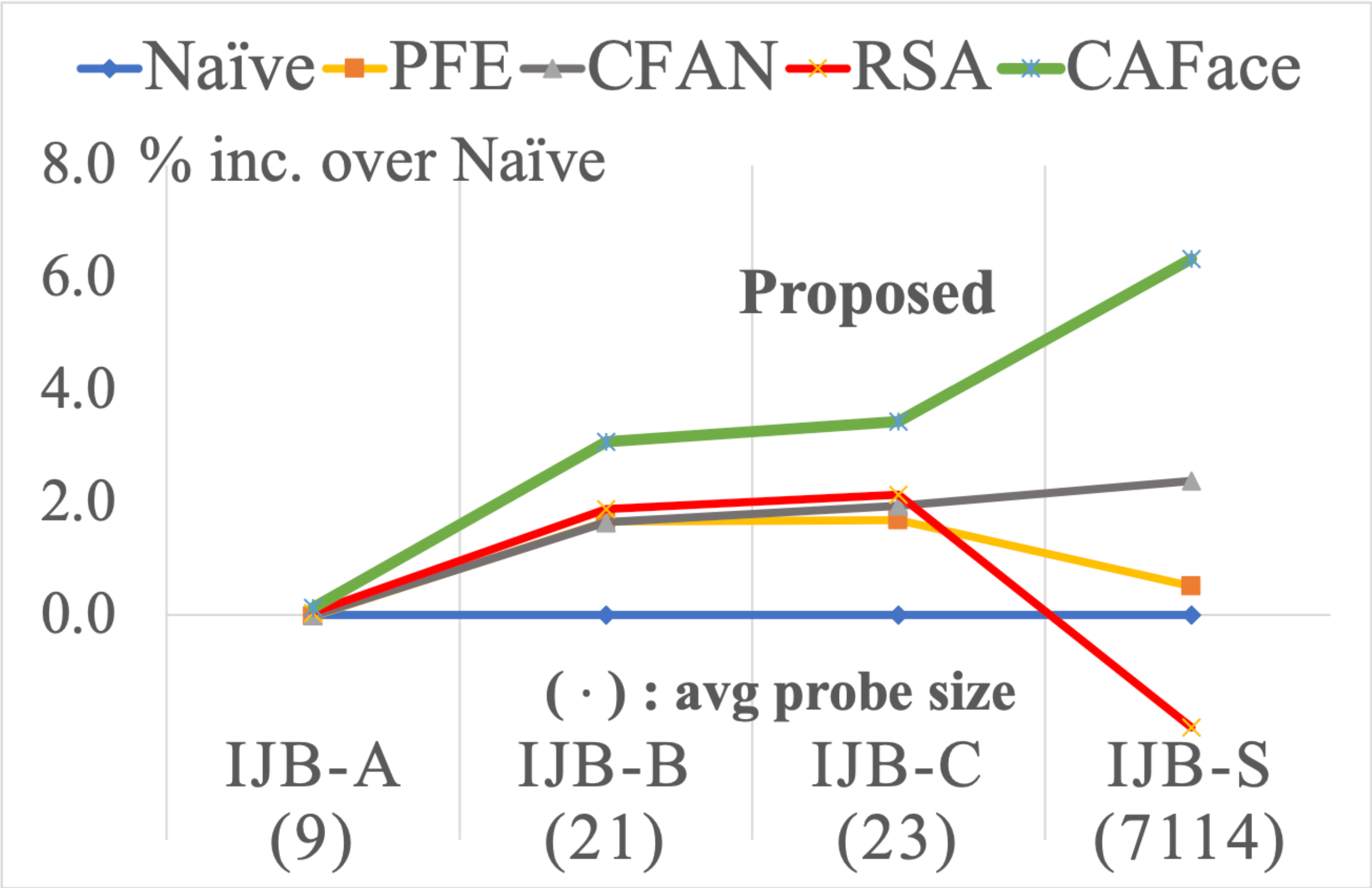
Architecture



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

Experiments

Performance Gain over simple average using feature fusion methods.



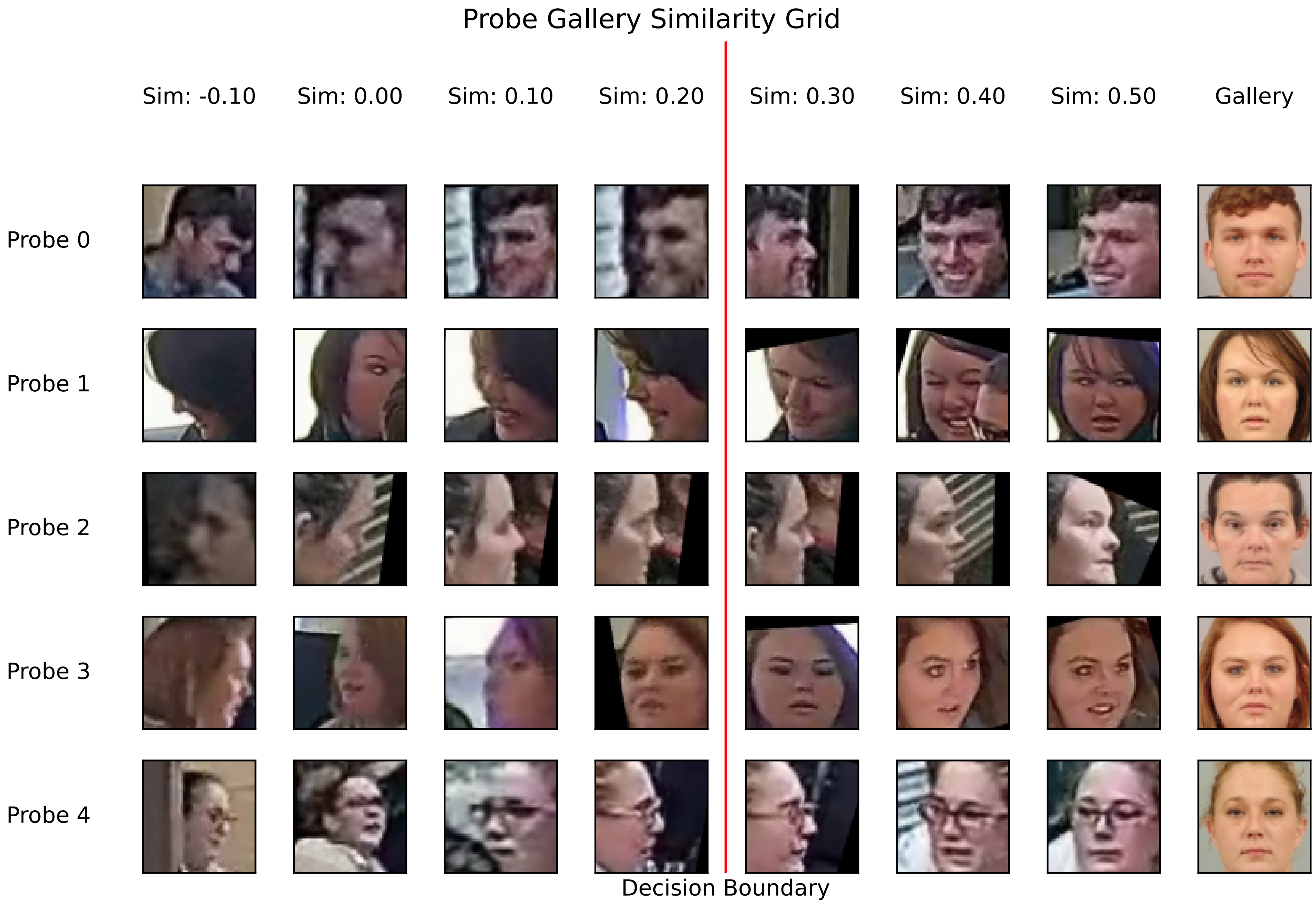
Naïve: Simple Average
PFE, CFAN: single image weight estimation
RSA: Attention Mechanism

Largest Probe size, Largest Perf. gain

CAFace Demo

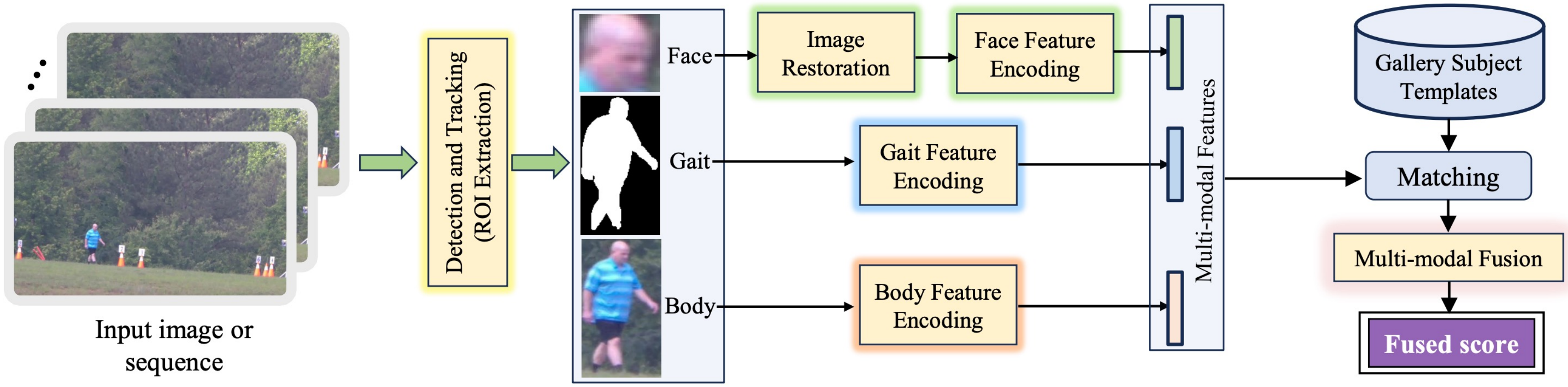
Video Feed - Frame Count: 1
Detected Face Frame
Top 5 CAFace Match
Top 5 Naive Match
0.13
0.12
0.10
0.10
0.10
0.13
0.12
0.10
0.10
0.10
Tok 5 Most influential Images in CAFace
Cosine Similarity To Gallery
Gallery Image
CAFace
Naive
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Recognition Examples

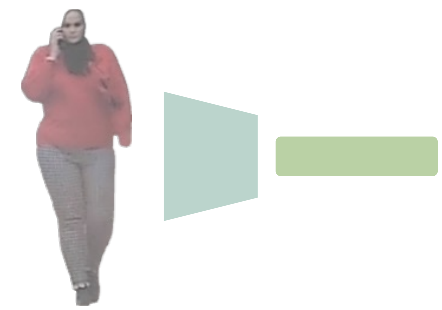


Recognition Examples

Person Identification at a (far) distance



- ➢ **4. 3D body matching (ICCV’23)**

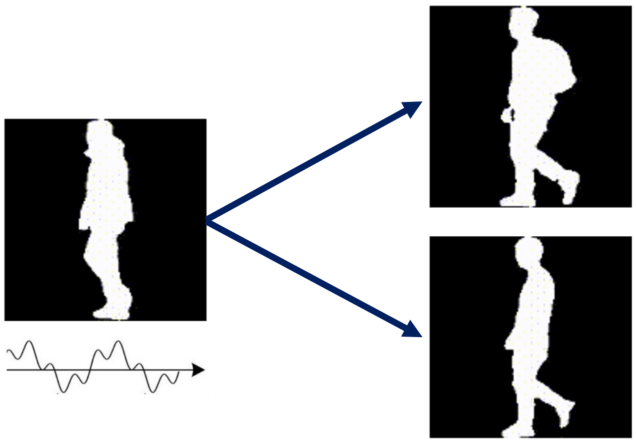


MICHIGAN STATE UNIVERSITY

The subject (ID: G01003) consented to publication.

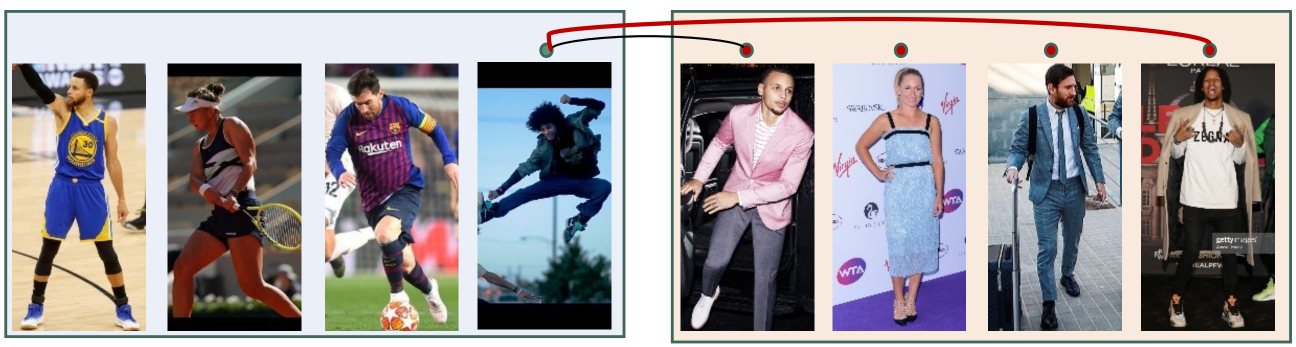
Problem Definition

People matching



Gait recognition

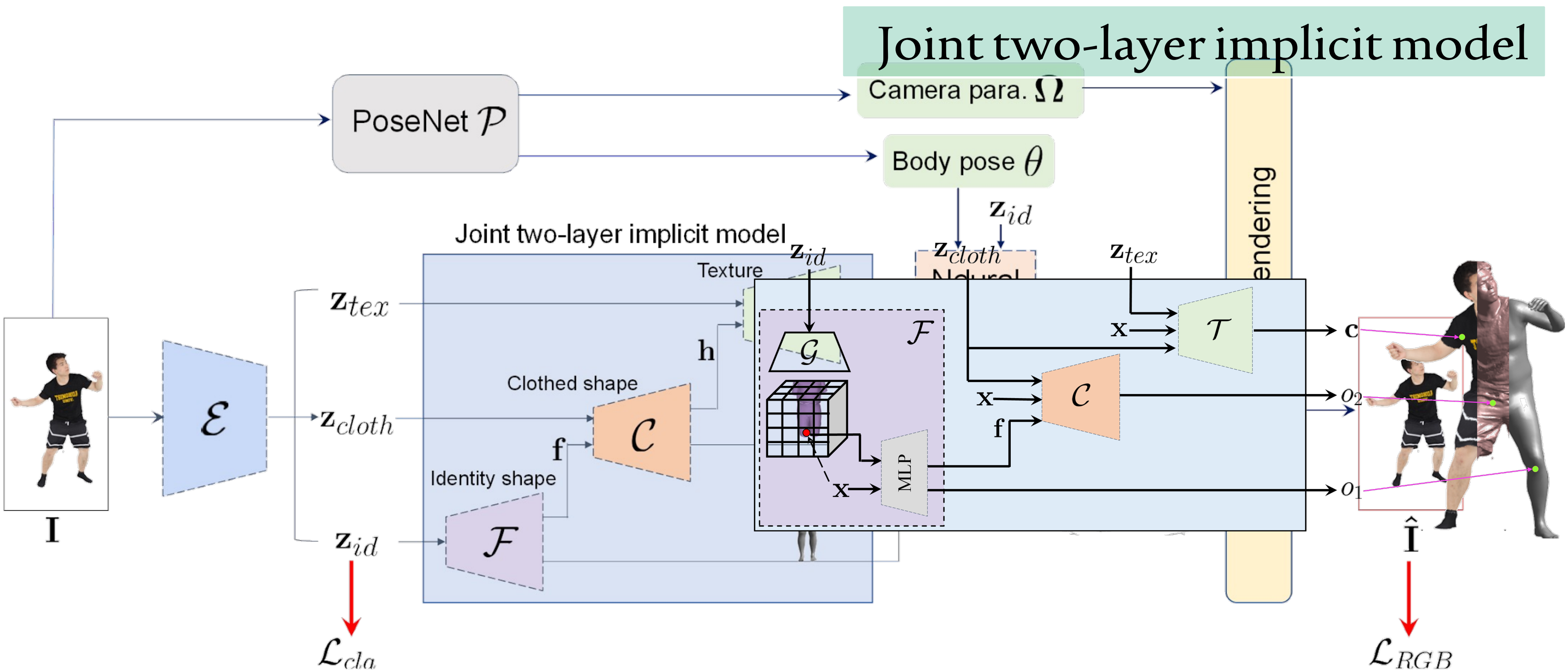
Person re-identification



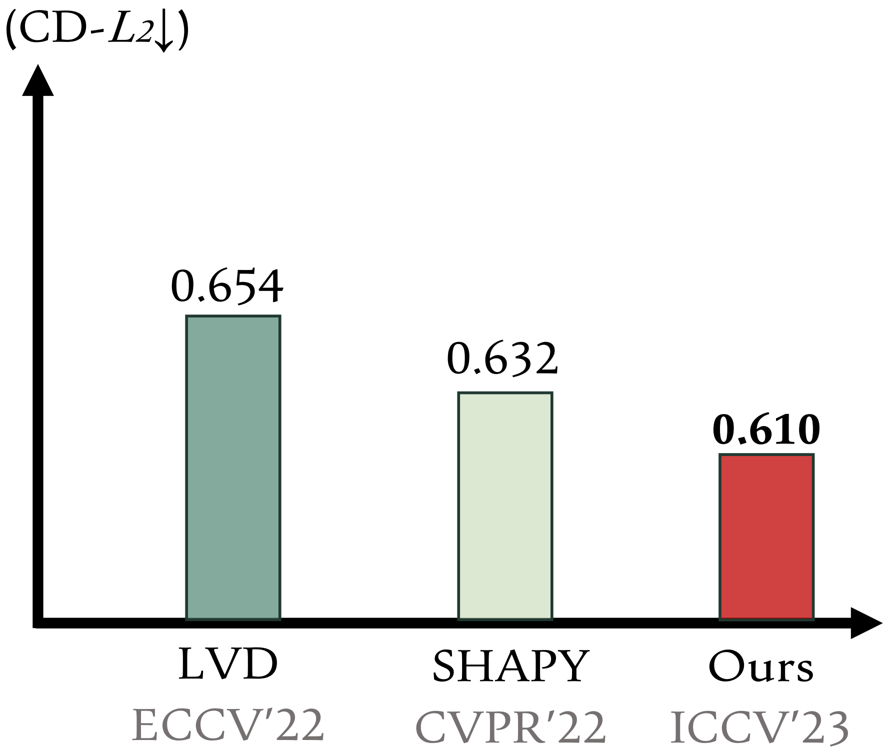
People matching

Feng Liu, et al., Learning Clothing and Pose Invariant 3D Shape Representation for Long-Term Person Re-Identification, ICCV 2023

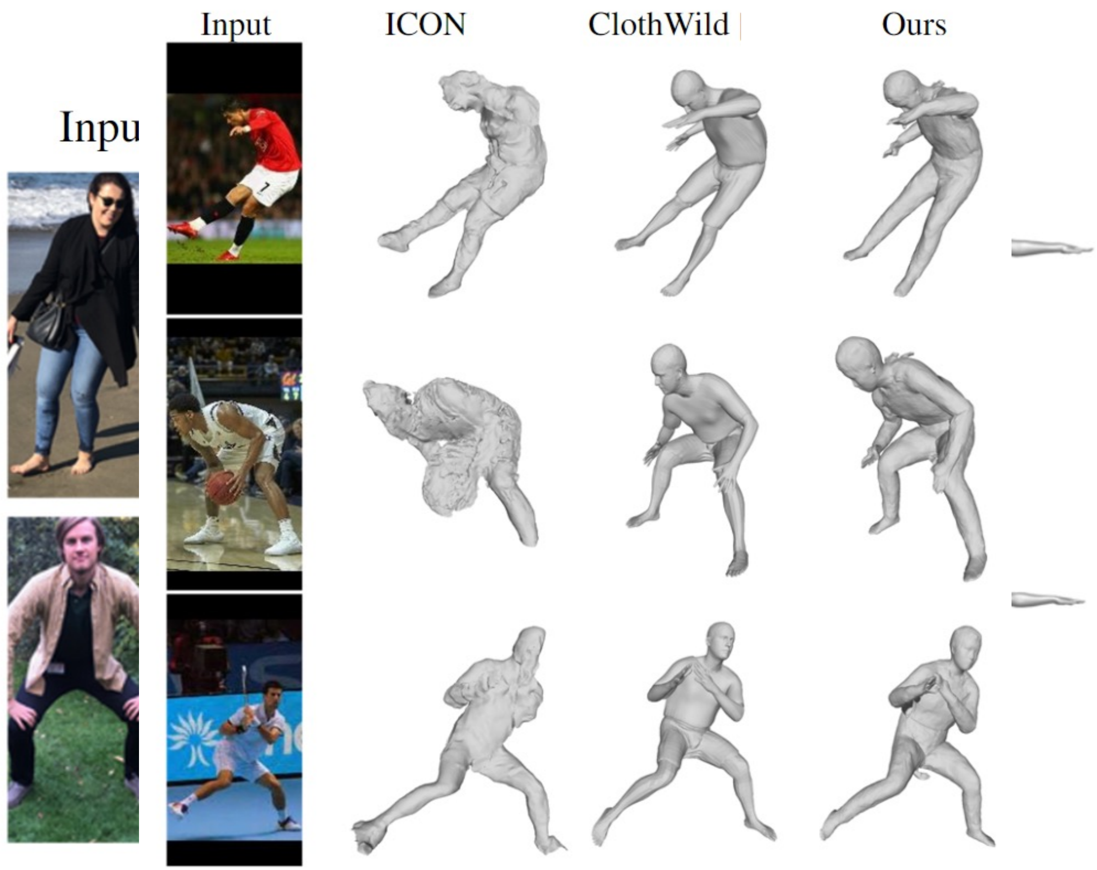
Discriminative 3D Human Shape Recon



3D Reconstruction Performance



Reconstruction on
HBW dataset

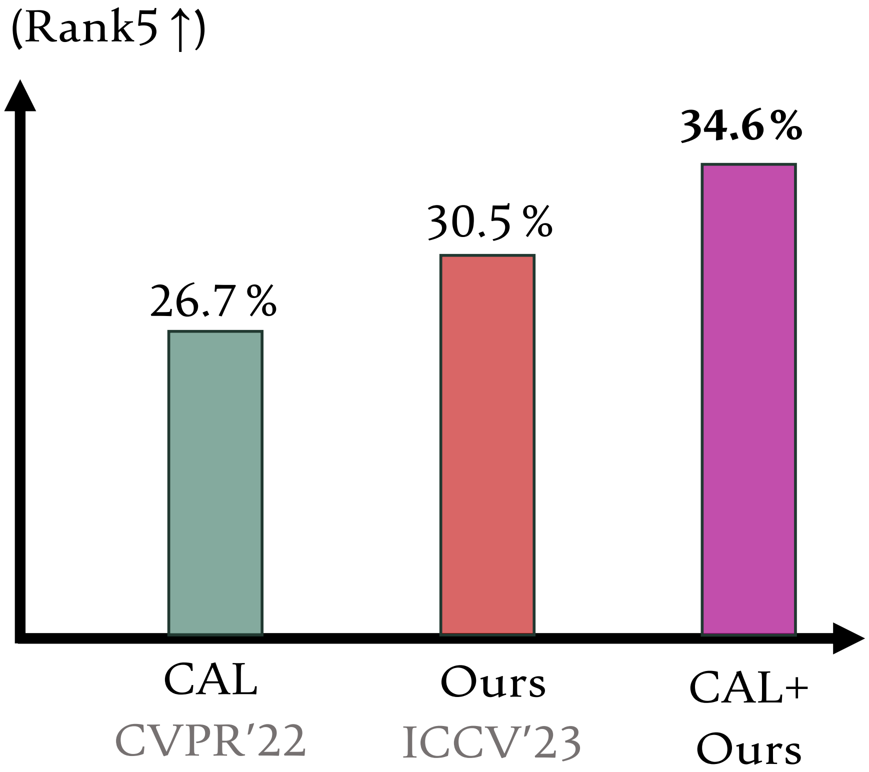


2-layer 3D Reconstruction

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

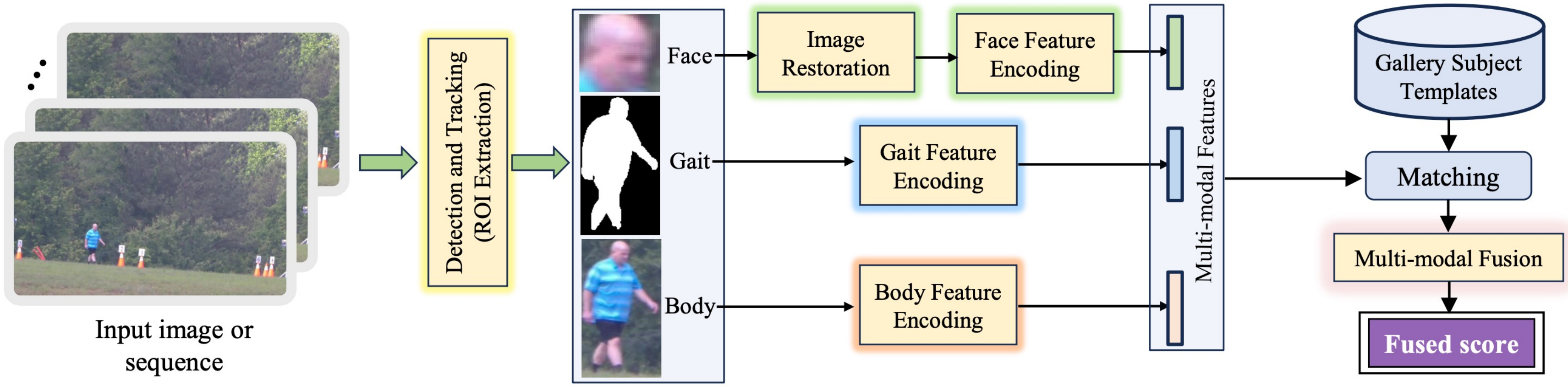
Body Matching Accuracy

A new Cloth-Changing and Diverse Activities (CCDA) dataset
1,555 images of 100 subjects



Body matching outperform gait!

Person Identification at a (far) distance

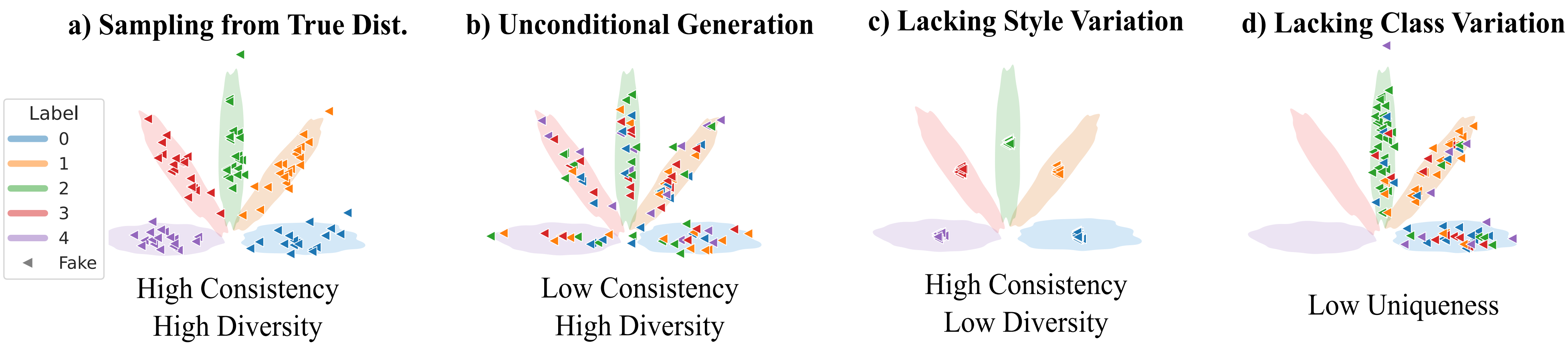


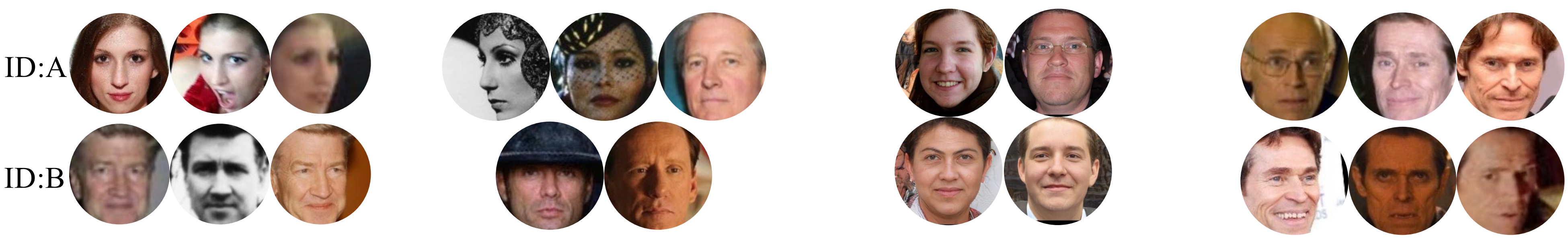
- **5. Synthetic training dataset** (CVPR’23)

MICHIGAN STATE UNIVERSITY

The subject (ID: G01003) consented to publication.

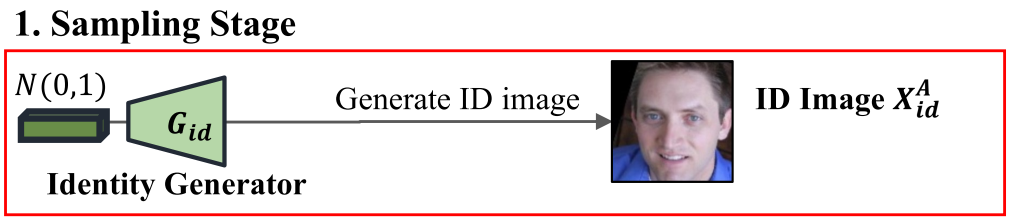
Characteristics of Labeled Faces





Good Dataset | No ID control | No Style Diversity | Not many Subjects

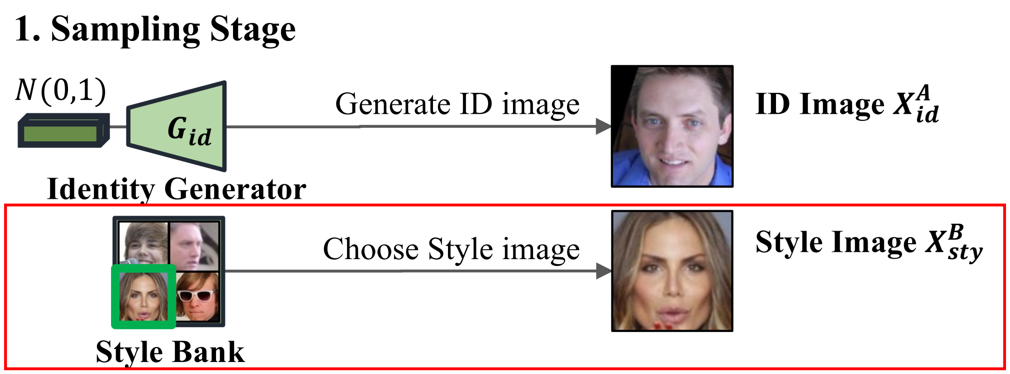
DCFace: Face dataset generation pipeline



1. Generate a facial image with unconditional DDPM

Method

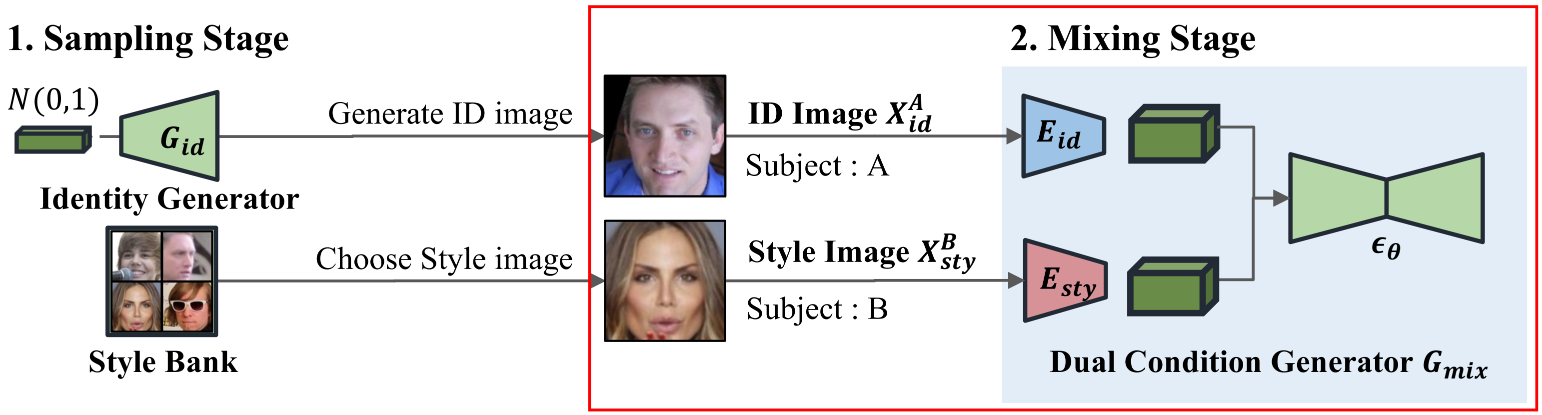
DCFace: Face dataset generation pipeline



2. Choose a real image that will be used for style information.
Style bank is an arbitrary set of real facial images.

Method

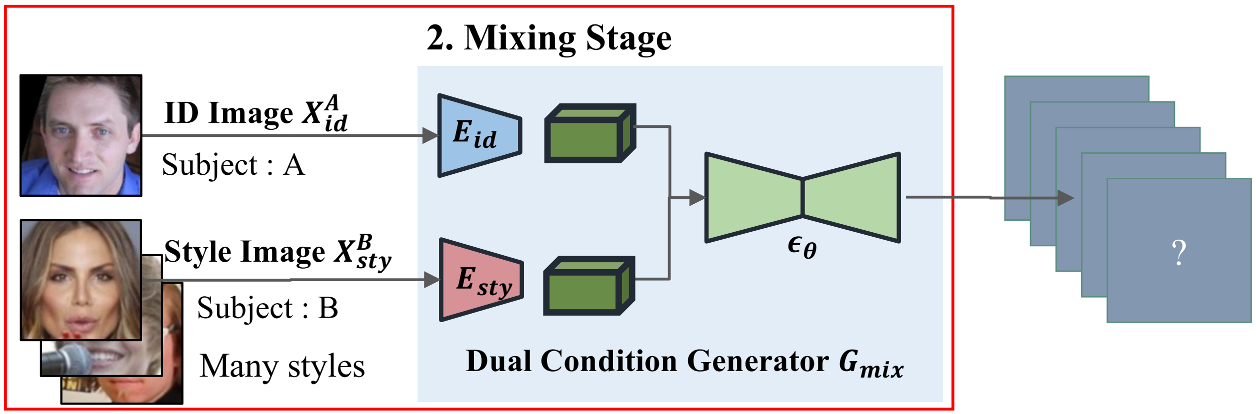
DCFace: Face dataset generation pipeline



3. Mix the ID image X_{id}^{A} and Style image X_{sty}^{B} using DCFace

Method

DCFace: Face dataset generation pipeline

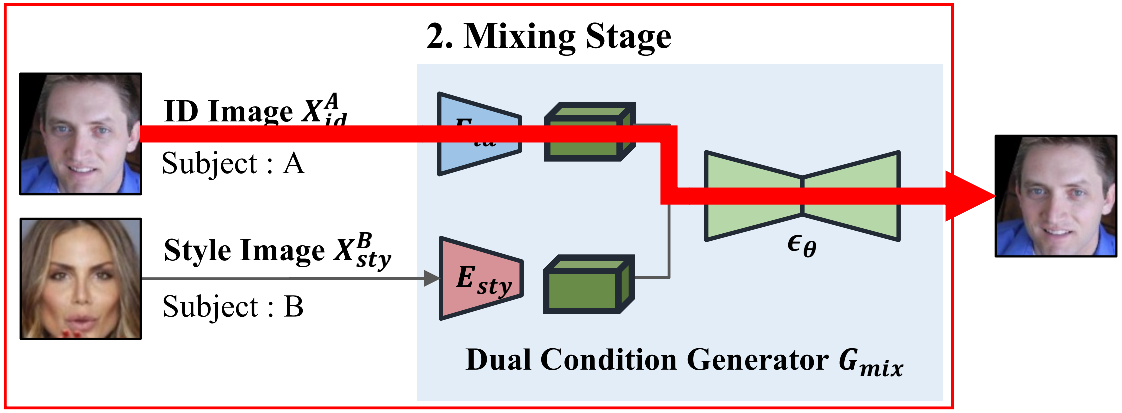


How do we ensure
1. ID consistency
2. Style Diversity
of generated samples?

3. Mix the ID image $\boldsymbol{X_{id}^{A}}$ and Style image $\boldsymbol{X_{sty}^{B}}$ using DCFace

Method

DCFace: Face dataset generation pipeline

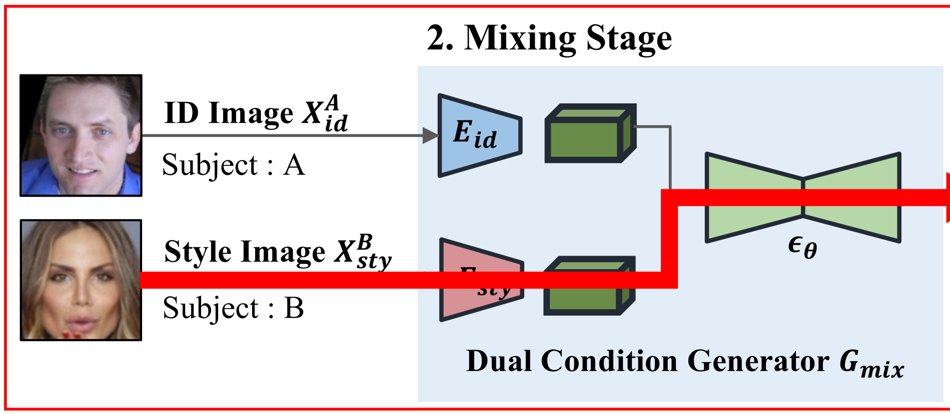


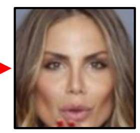
Complete dependence on X^A_{id} leads to no style diversity.

3. Mix the ID image X^A_{id} and Style image X^B_{sty} using DCFace

Method

DCFace: Face dataset generation pipeline



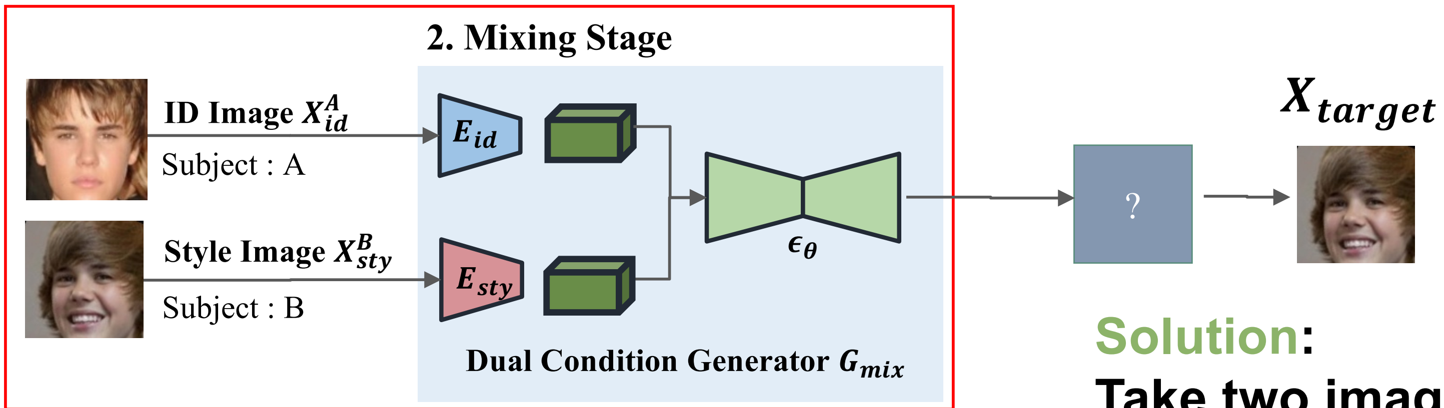


Complete dependence on X_{sty}^{B} leads to incorrect subject appearance.

3. Mix the ID image X_{id}^{A} and Style image X_{sty}^{B} using DCFace

Method

DCFace: Face dataset generation pipeline



Solution:

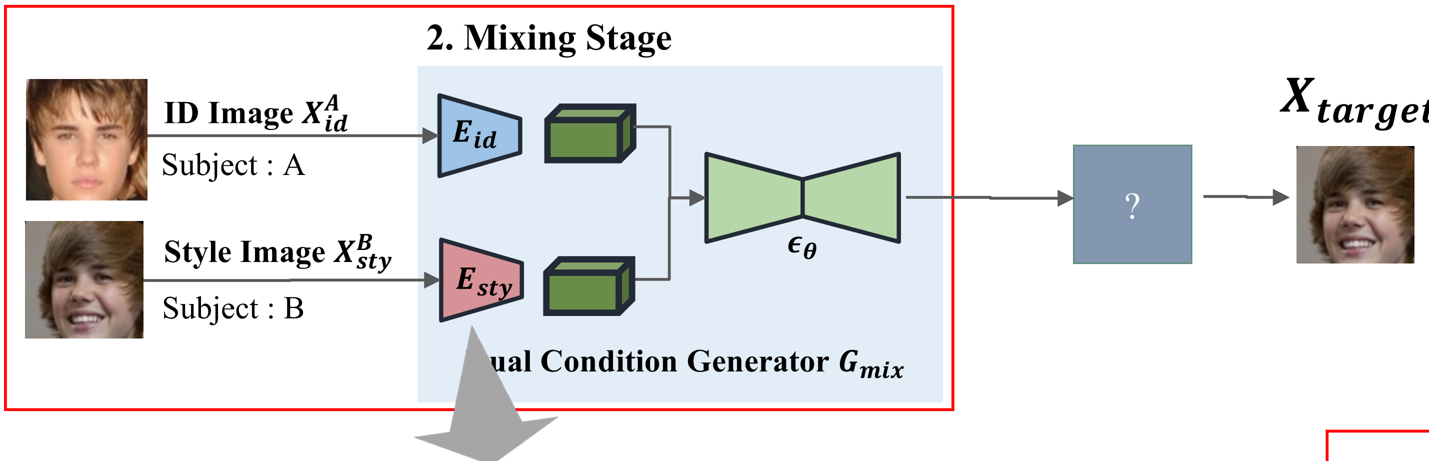
Take two images of one subject.

Let X_{id}^{A} provide fine details of ID.

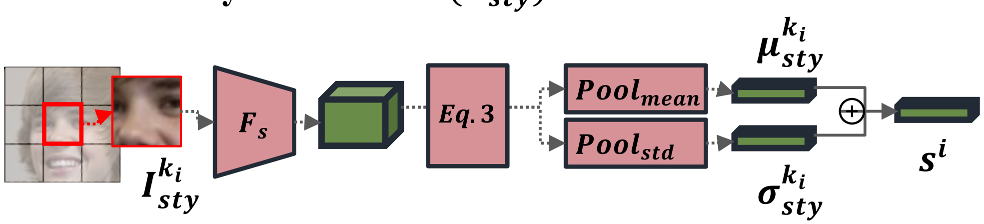
Let X_{sty}^{B} provide low frequency style.

Method

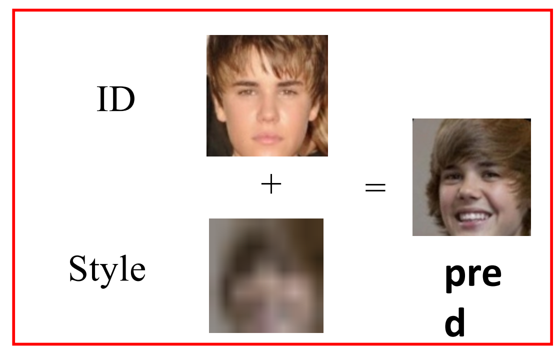
DCFace: Face dataset generation pipeline



Patch-wise Style Extractor (E_{sty})



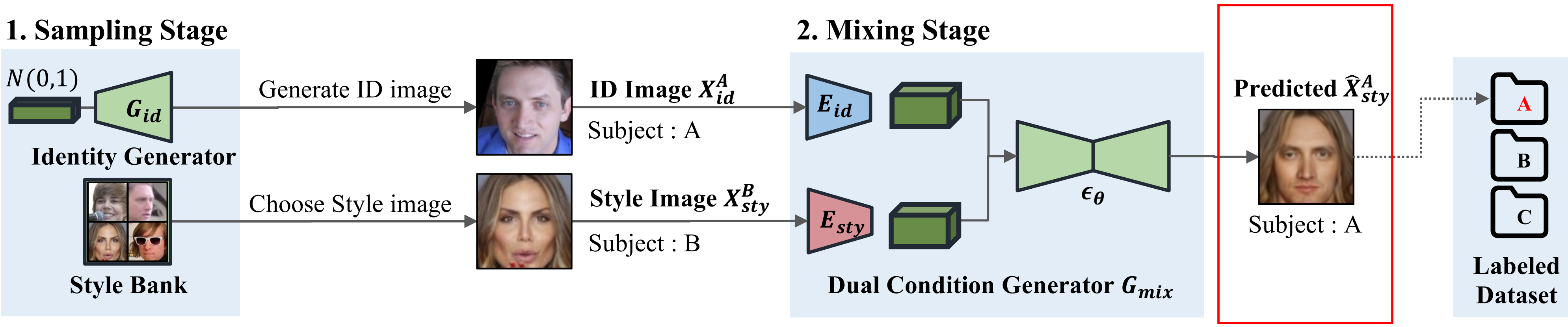
Conceptually



Patch-wise spatial mean+variance creates low frequency style information.

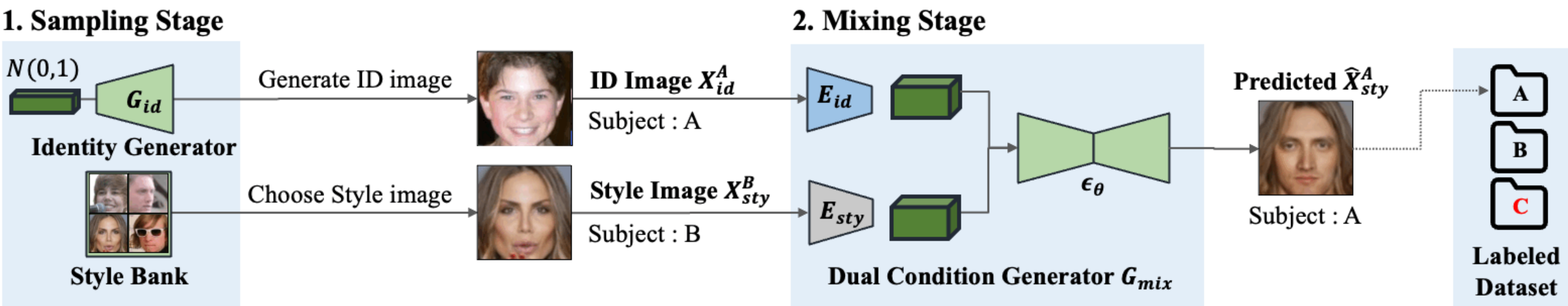
Method

DCFace: Face dataset generation pipeline



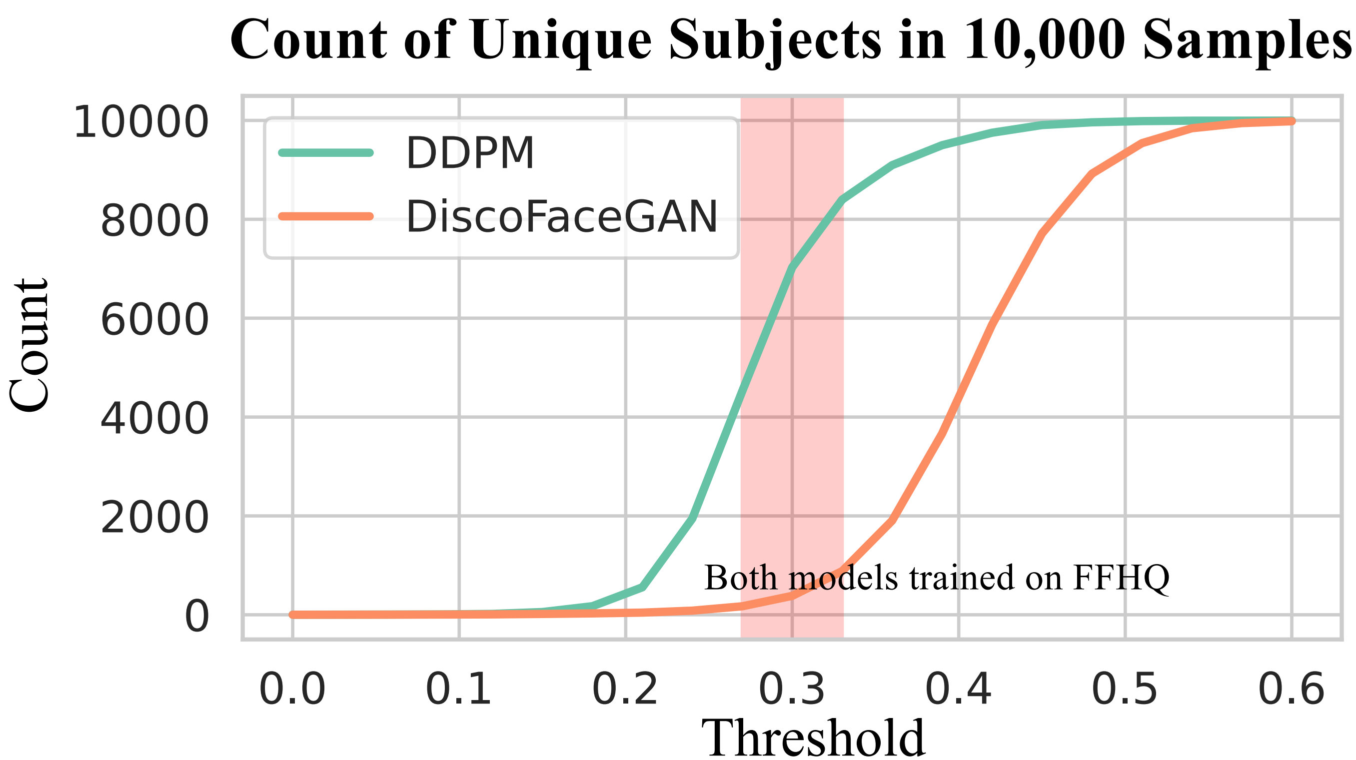
4. Predicted image has the ID of X_{id}^{A} while taking the style of X_{sty}^{B}.

Method



5. Repeat this procedure to generate a dataset.

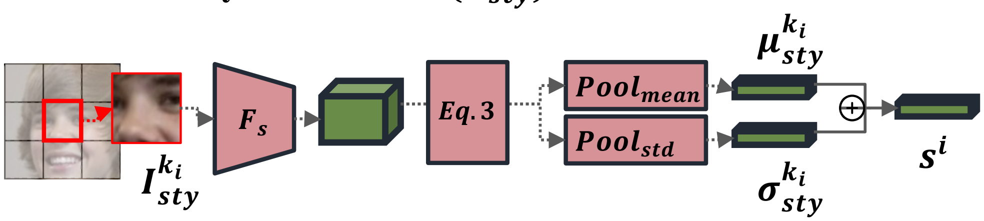
ID Image Generation



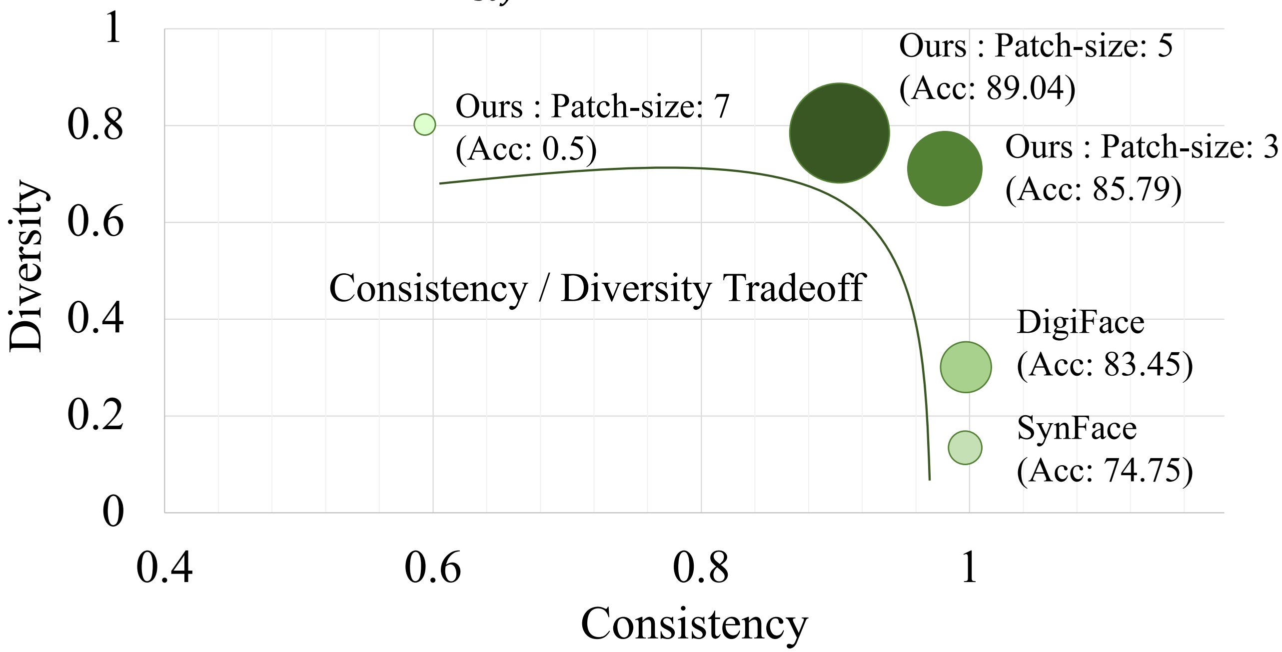
We take **Unconditional DDPM** to generate novel subjects.
Uniqueness determined by a pretrained face recognition model with varying threshold.

Style Extraction

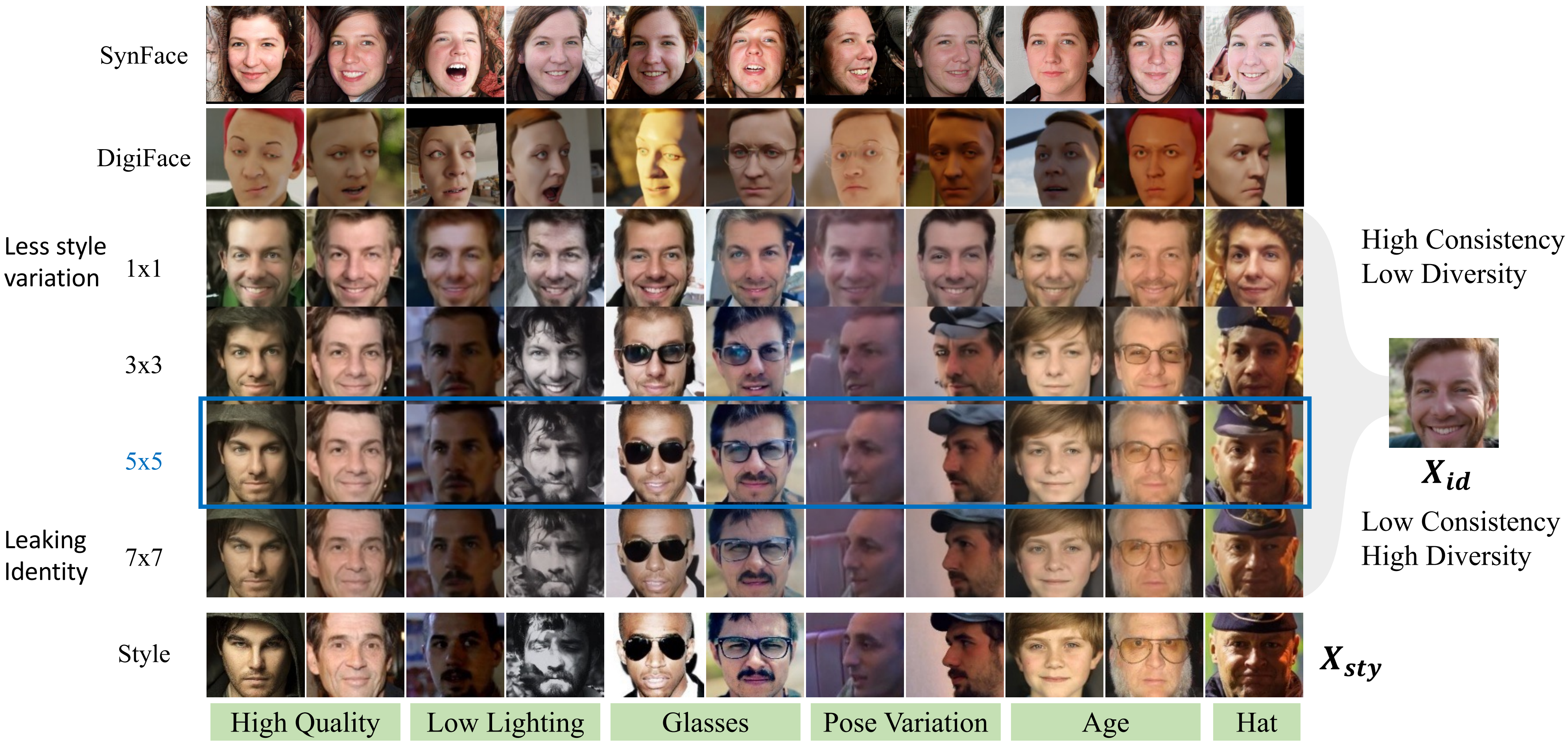
Patch-wise Style Extractor (E_{sty})



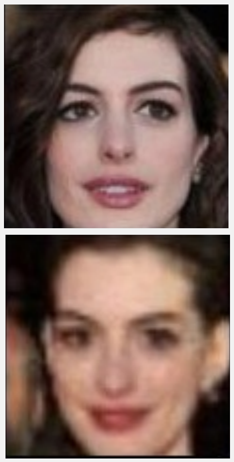
Patch-size controls the diversity/consistency trade-off



Effect of Patch Size

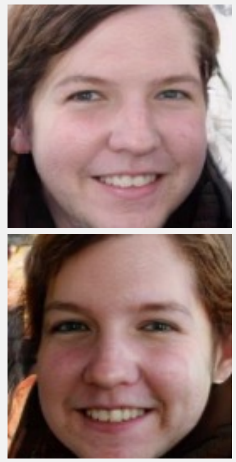


Synthetic Dataset Performance



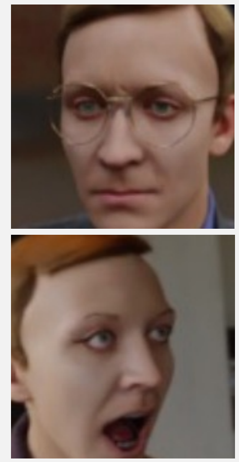
CASIA-WebFace

Real Dataset



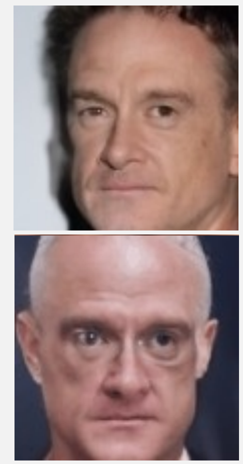
SynFace ICCV21

GAN-Based Synthesis (ID Consistent)



DigiFace WACV23

3DMM-Based Synthesis (Style Diverse)



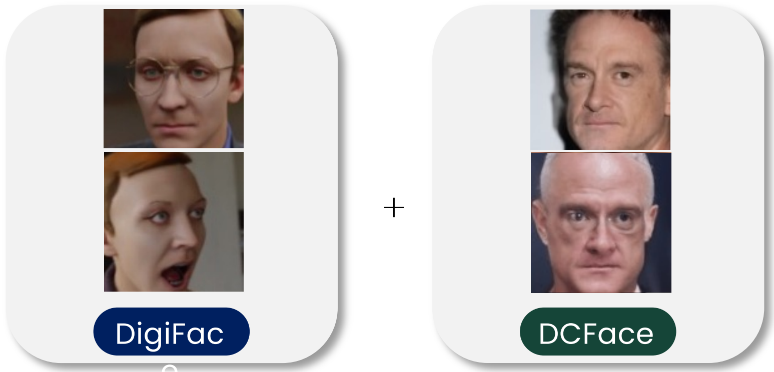
DCFace CVPR23

Dual Condition DDPM (Consistent + Diverse)

Face Recognition Performance

	Methods	Venue	# images (# IDs× # imgs/ID)	LFW	CFP-FP	CPLFW	AgeDB	CALFW	Avg	Gap to Real
	CASIA-WebFace (Real)		0.49M (approx. 10.5K × 47)	99.42	96.56	89.73	94.08	93.32	94.62	0.0
Same Amount as Real	SynFace	ICCV21	0.5M (10K × 50)	91.93	75.03	70.43	61.63	74.73	74.75	21.00
Same Amount as Real	DigiFace	WACV23	0.5M (10K × 50)	95.4	**87.4**	78.87	76.97	78.62	83.45	11.81
Same Amount as Real	DCFace (Ours)	-	0.5M (10K × 50)	**98.55**	85.33	**82.62**	**89.70**	**91.60**	**89.56**	**5.35**
More Samples	DigiFace	WACV23	1.2M (10K × 72 + 100K × 5)	96.17	**89.81**	82.23	81.10	82.55	86.37	8.72
More Samples	DCFace (Ours)	-	1.0M (20K × 50)	**98.83**	88.4	84.22	90.45	92.38	90.86	3.98
More Samples	DCFace (Ours)	-	1.2M (20K × 50 + 40K × 5)	98.58	88.61	**85.07**	**90.97**	**92.82**	**91.21**	**3.61**

Combining Multiple Synthetic Datasets

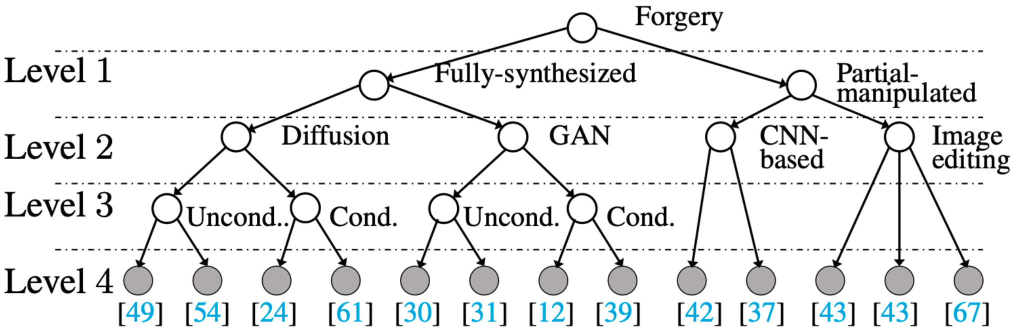


Other types of synthetic datasets are complementary

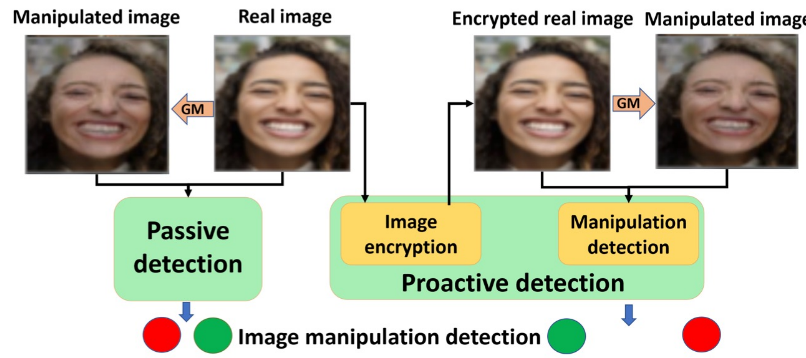
Face Recognition Performance

	# Synthetic Imgs	# Real Imgs	LFW	CFPFP	CPLFW	AGEDB	CALFW	AVG	Gap to Real
DigiFace	1.2M (10K×72+100K×5)	0	96.17	89.81	82.23	81.10	82.55	86.37	8.72
DigiFace	1.2M (10K×72+100K×5)	2K×20	99.17	94.63	88.1	90.5	90.97	92.67	2.06
DCFace	1.2M (20K×50+40K×5)	0	98.58	88.61	85.07	90.97	92.82	91.21	3.61
DCFace	1.2M (20K×50+40K×5)	2K×20	98.97	94.01	86.78	91.80	92.95	92.90	1.82
DCFace+DigiFace (2.4M)		0	99.20	93.63	87.25	92.25	92.95	93.06	**1.65**
CASIA	0	0.5M	99.42	96.56	89.73	94.08	93.32	94.62	0

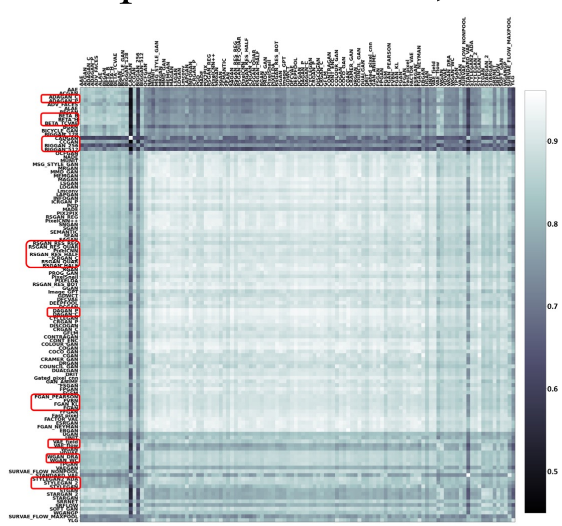
Trustworthiness



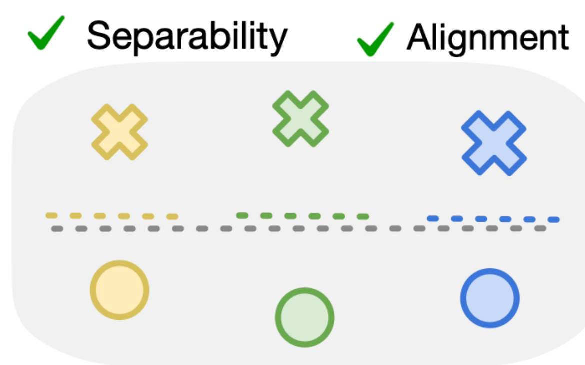
Deepfake detection, CVPR'23



Proactive CV, CVPR'22, 23



Model parsing, PAMI'23



Anti-Spoofing, CVPR'23

Future Directions

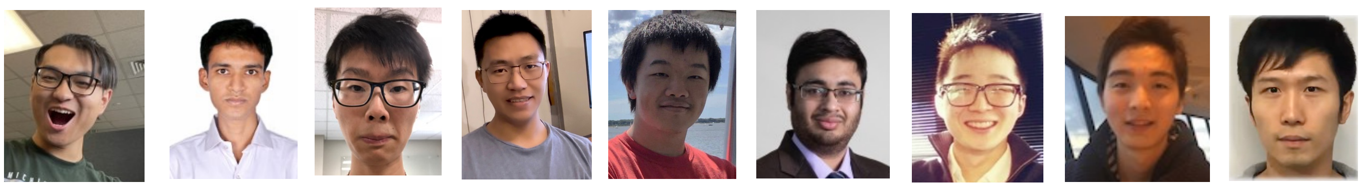
- New backbone for face recognition: ViT
- Move from close-set to open-set
- Fusion between face, body, and gait
- Advance AIGC to push “gap to real” to zero
- Explainable recognition systems
- 3D cloth modeling for body biometrics
- Leverage LLM for fine-grained recognition

Conclusions

- There are many research opportunities in person identification at a distance.
- Body biometrics is just at the beginning and there is a great potential for further development.
- Classic topics such as face recognition could benefit from the latest AI development, such as AIGC or LLM.

Thanks

Dr. Joseph Roth, Dr. Jamal Afridi, Dr. Morteza Safdarnejad, Dr. Yousef Atoum, Dr. Xi Yin, Dr. Amin Jourabloo, Dr. Luan Tran, Dr. Yaojie Liu, Dr. Garrick Brazil



Zhiyuan Ren, Abhinav Kumar, Shengjie Zhu, Dr. Feng Liu, Andrew Hou, Vishal Asnani, Minchul Kim, Yiyang Su, Xiao Guo

Sponsors:

Questions?

http://cvlab.cse.msu.edu