MICHIGAN STATE UNIVERSITY

Biometric Recognition in the Era of Foundation Models

Dr. Xiaoming Liu

Anil K. and Nandita Jain Endowed Professor of Engineering

MSU Foundation Professor

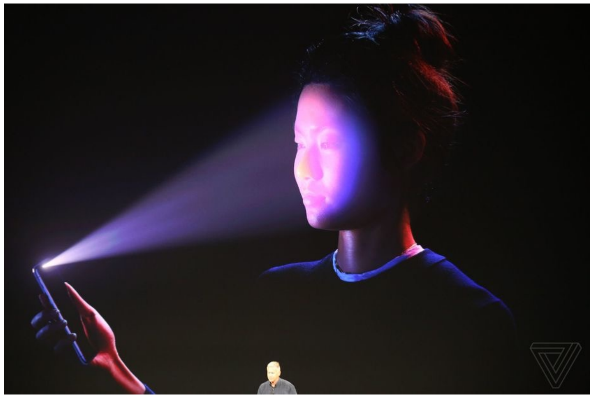
Department of Computer Science and Engineering

Michigan State University

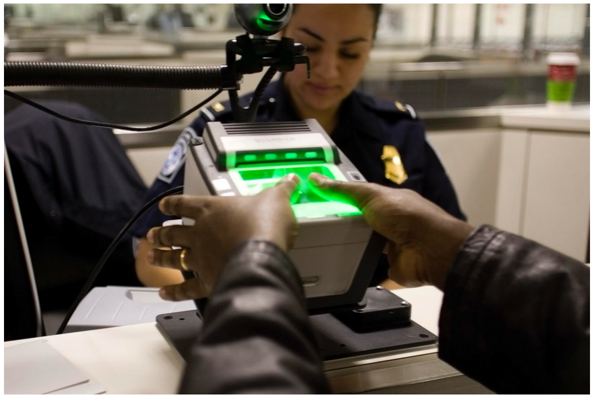
Joint work with Anil Jain, Shiqi Yu, Feng Liu, Chao Fan

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.

Successful Applications



Apple



Fingerprint

Boarding in Airports

Amazon One Palmprint

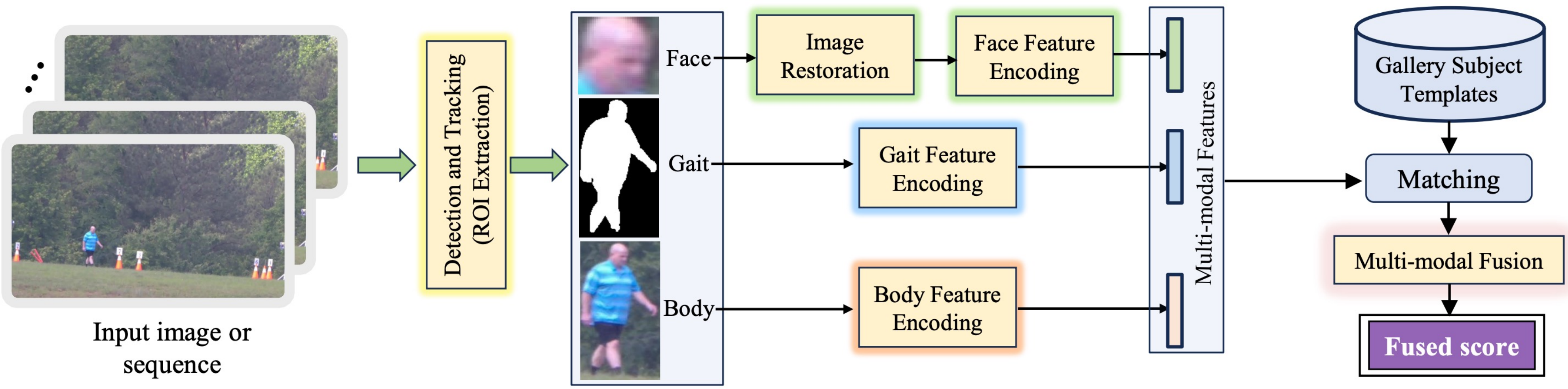
Identification at a Distance

BRIAR: The subject in the figure consented to publication.

BRIAR

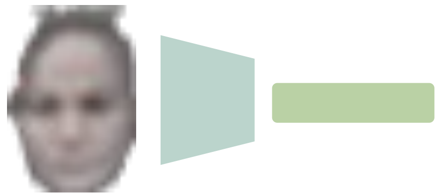
- One of the largest Biometrics projects in US
- Sponsored by IARPA
- 7 full teams and 2 partial teams in phase 1
- 5 teams in phase 2
- 2 teams in phase 3
- Advancing the SOTA in face, body, gait recognition, multi-modality fusion, AIGC, and image restoration

MSU BRIAR System



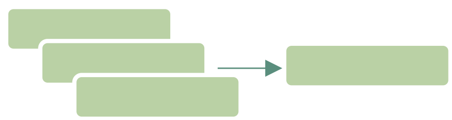
System Components

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

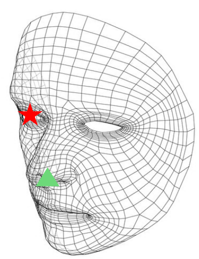


- **2. Domain adaption:** CFSM (ECCV'22)

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

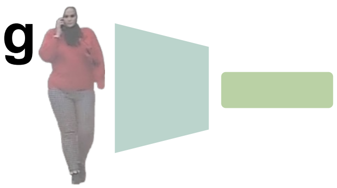


- **4. Landmark assisted reco** KP-RPE (CVPR'24)



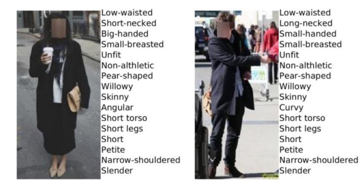
- **5. Synthetic training dataset** (CVPR'23)

- **6. 3D body matching** (ICCV'23)

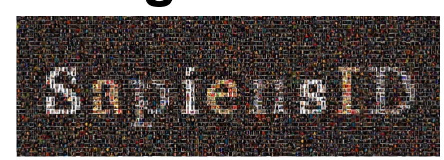


- **7. Large vision models** (CVPR'24)

- **8. CLIP 3D Re-ID** (CVPR'24)

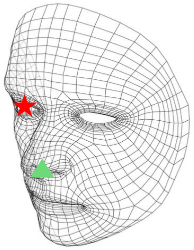


- **9. Unified human recognition** (under review)



Highlights

- **4. Landmark assisted reco**
  KP-RPE (CVPR'24)



- **7. Large vision models**
  (CVPR'24)

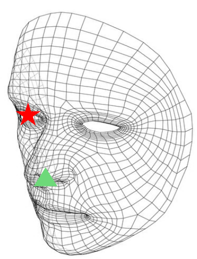
- **8. CLIP 3D Re-ID**
  (CVPR'24)

- **9. Unified human recognition**
  (under review)

Highlights

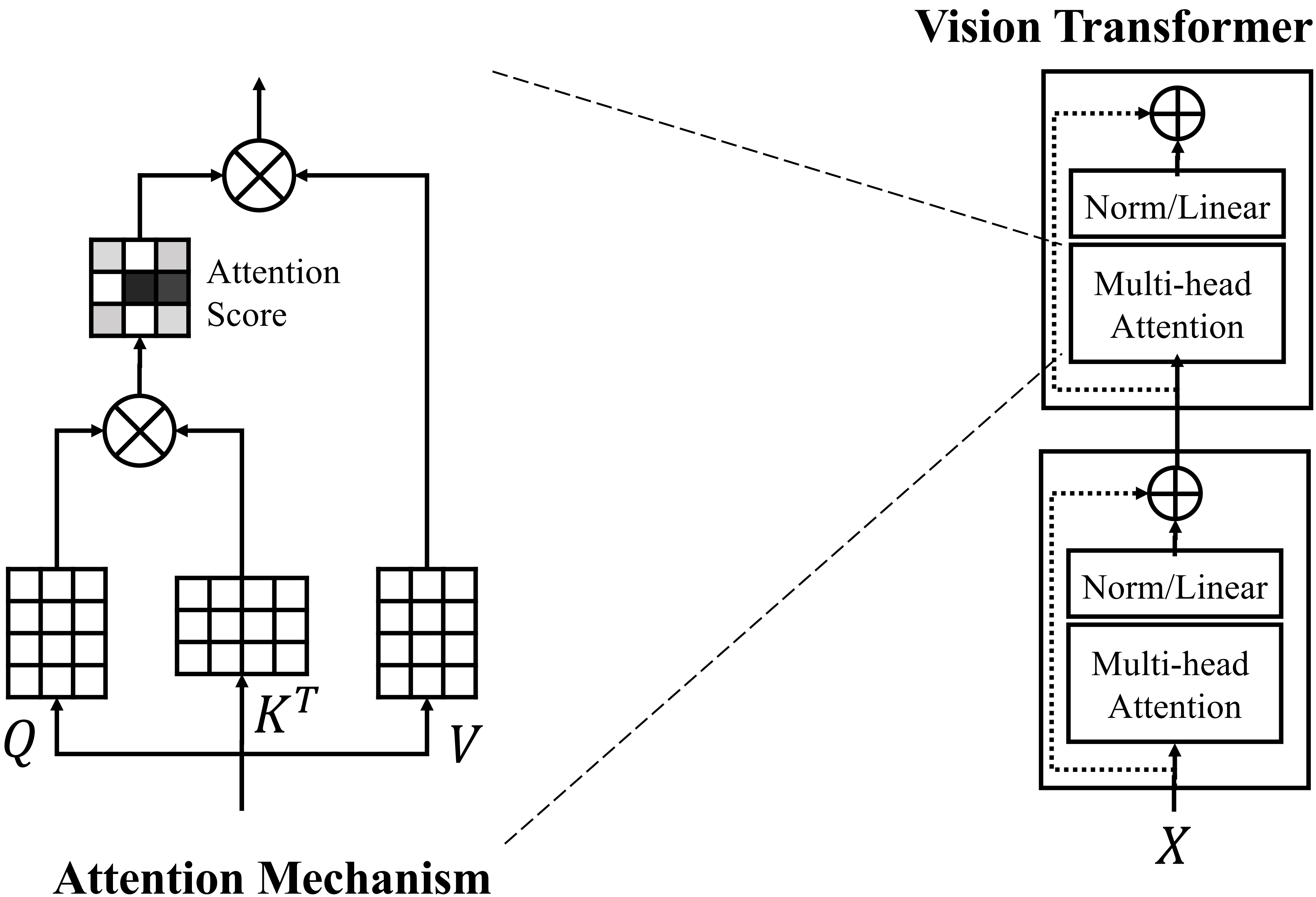
➢ **4. Landmark assisted reco**

KP-RPE (CVPR'24)



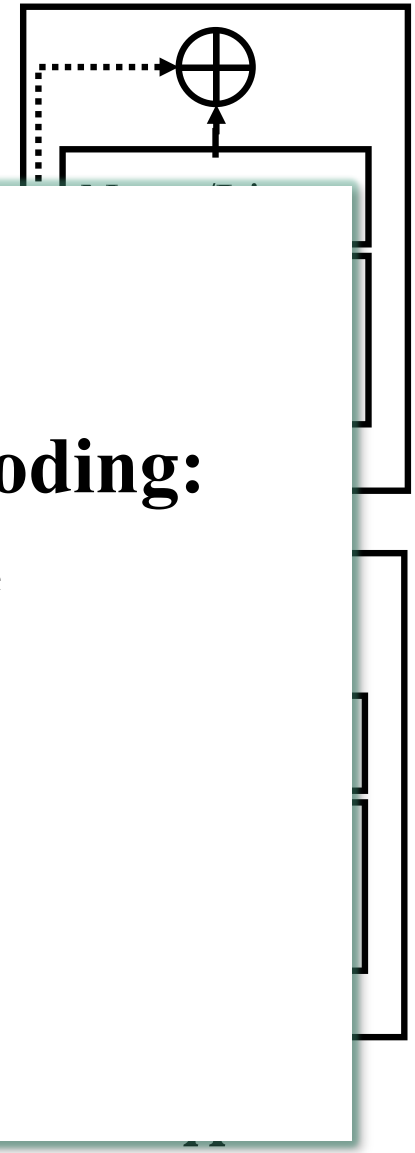
Minchul Kim, Feng Liu, Yiyang Su, Anil K. Jain, Xiaoming Liu, "KeyPoint Relative Position Encoding for Face Recognition," in CVPR 2024

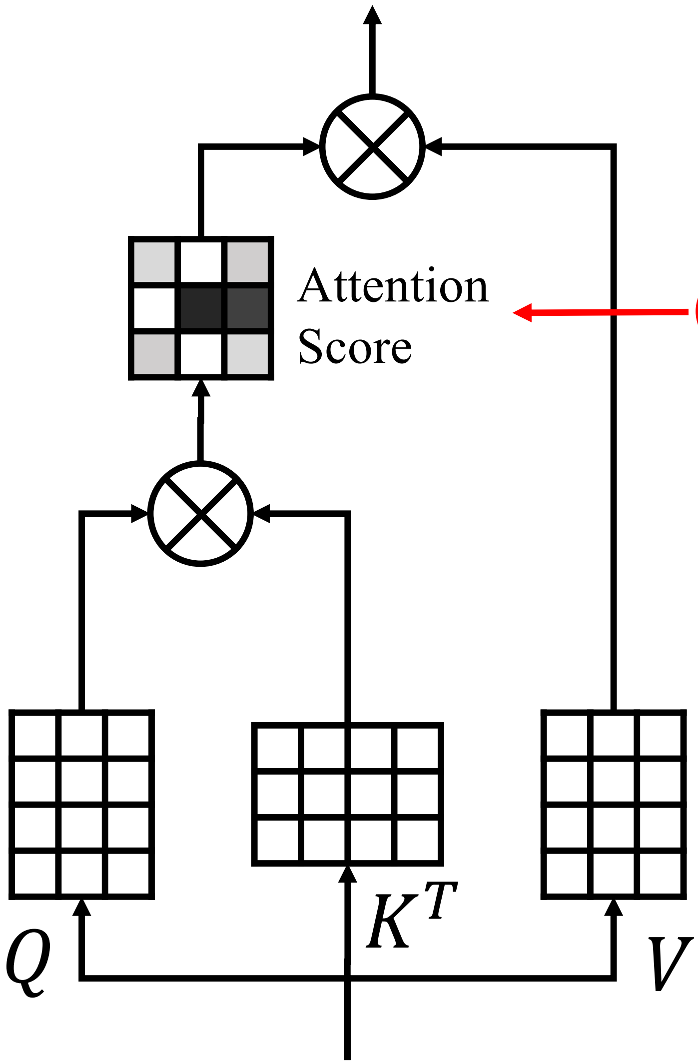
Problem Definition

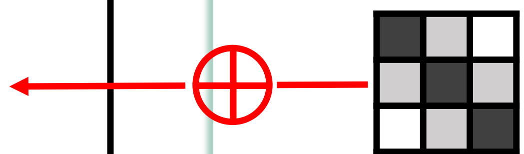


Problem Definition

Vision Transformer







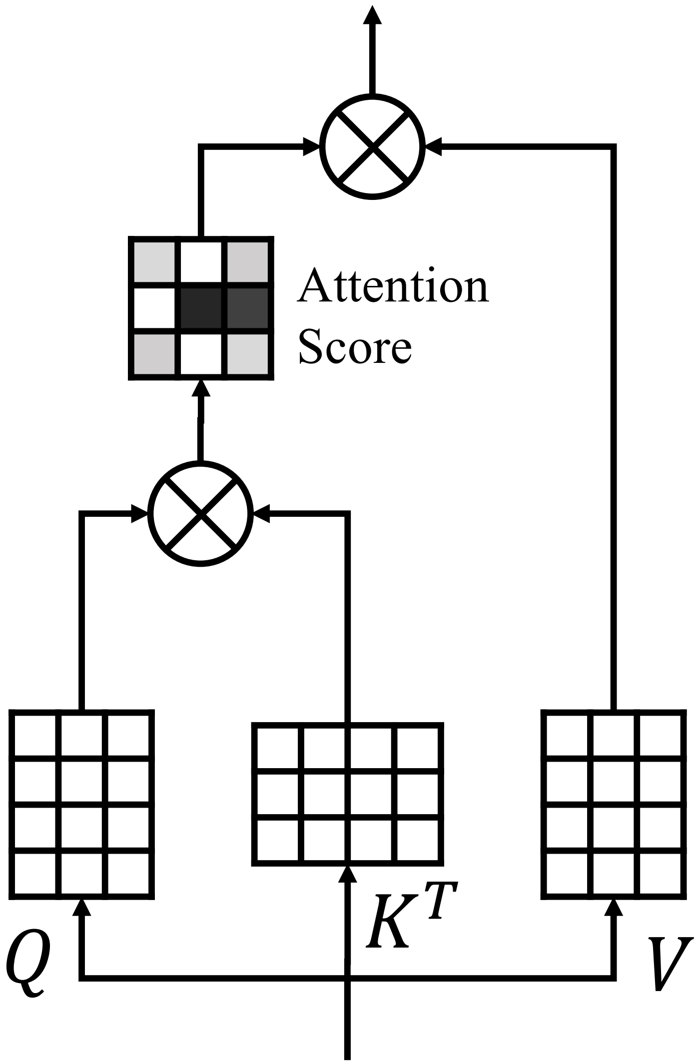
Relative Position Encoding:

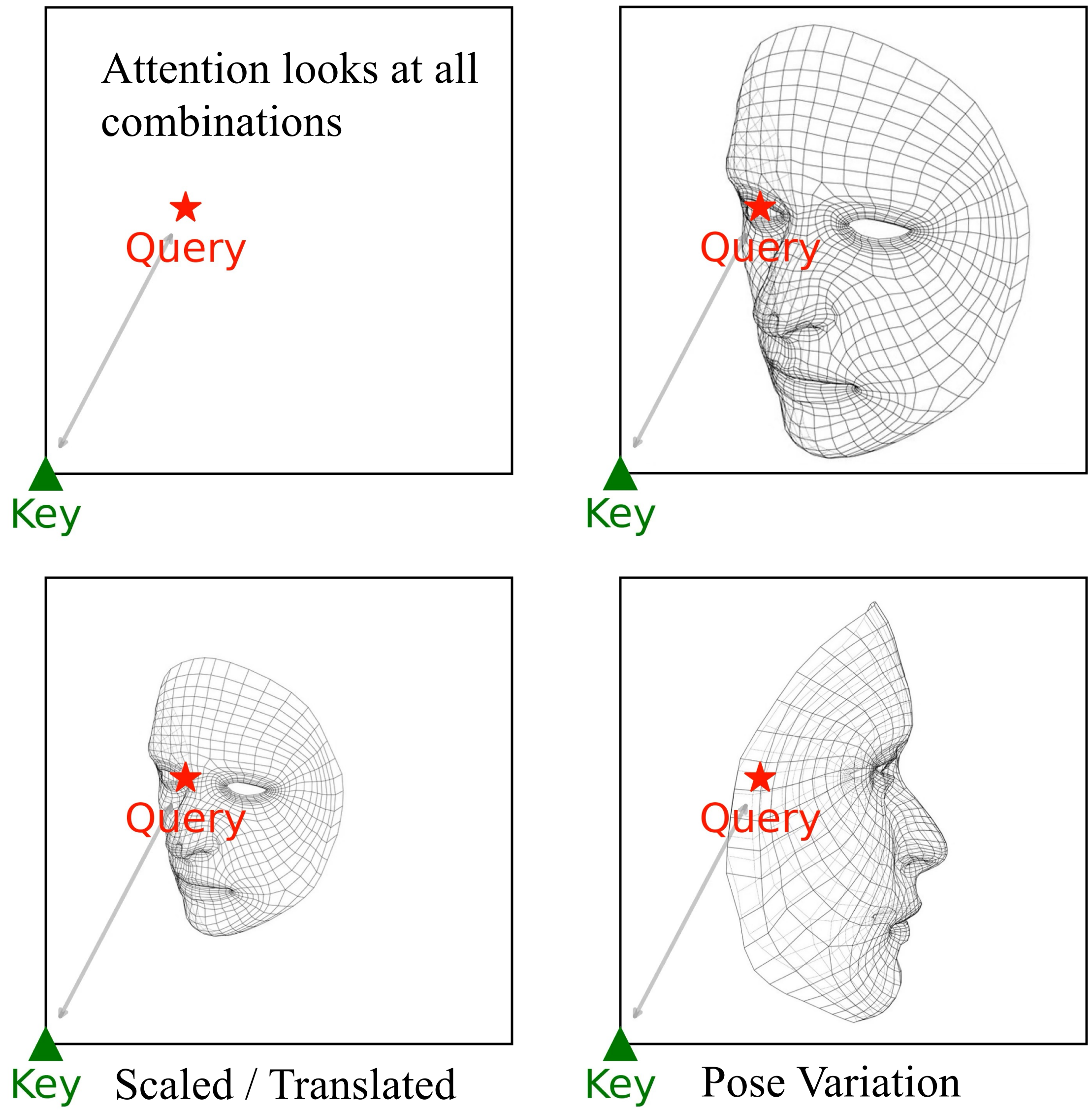
Injects priors in pairwise relationship.
***i.e.*) Nearby points are more important.**

Problem:
It is same for all images.

Attention Mechanism

Problem Definition

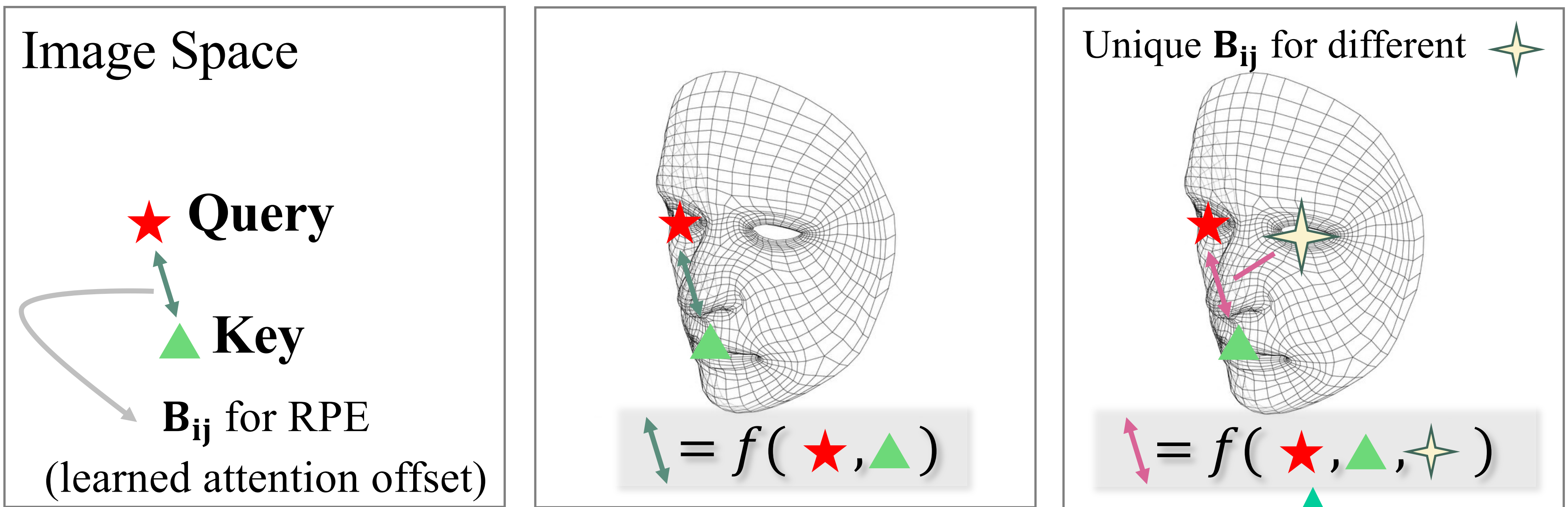




Same Query-Key Locations **does not** represent same semantics

Method

RPE[1] **KP-RPE (Ours)**



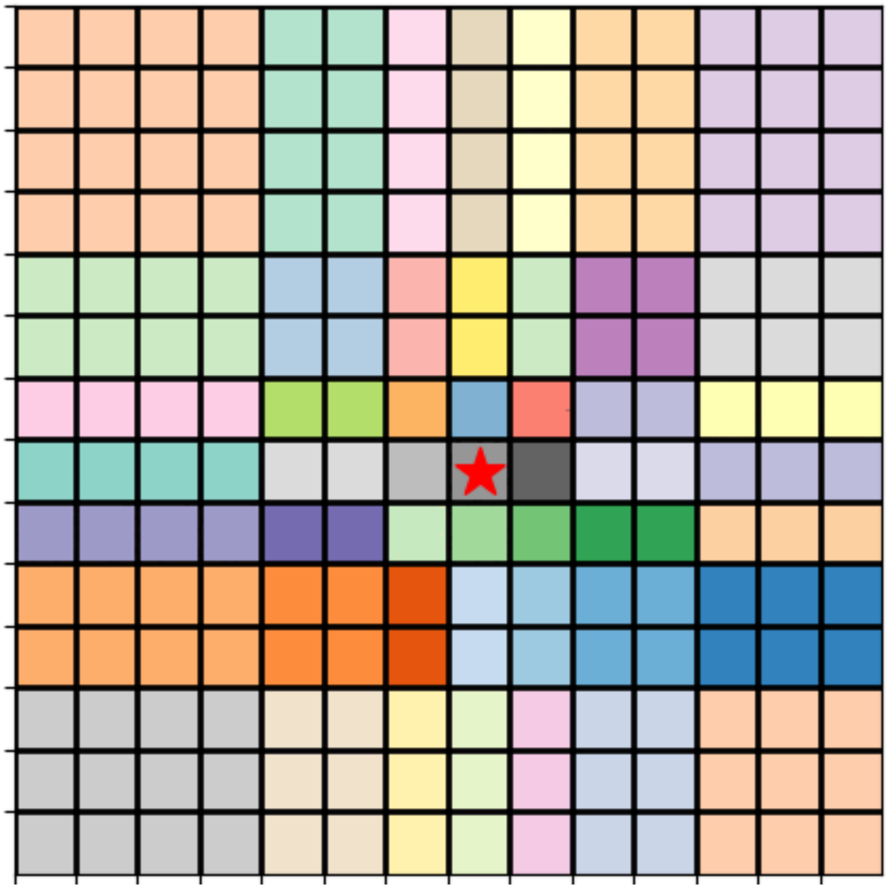
RPE becomes Keypoint ( ) Dependent

Attention Score: $e'_{ij} = \dfrac{\mathbf{Q}i\mathbf{K}j^T + \mathbf{B}_{ij}}{\sqrt{d_k}}$

[1] Shaw, Peter, Jakob Uszkoreit, and Ashish Vaswani. "Self-attention with relative position representations." arXiv preprint arXiv:1803.02155 (2018).

Method

KP-RPE for one query at ★

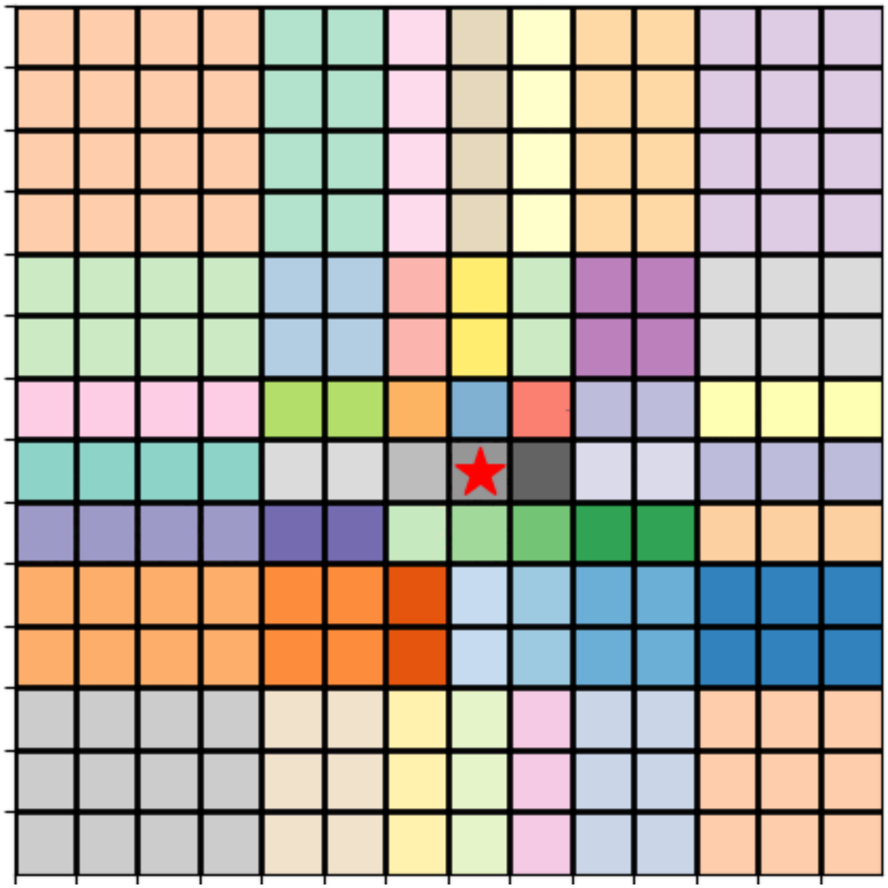


1. Compute distance to query.
(We use quantized distance)

* Same color implies same distance

Method

KP-RPE for one query at ★



1. Compute distance to query. (We use quantized distance)

2. Learn optimal values for each distance as a function of facial landmarks L

$= W_1(L - (q_x, q_y))$

$= W_2(L - (q_x, q_y))$

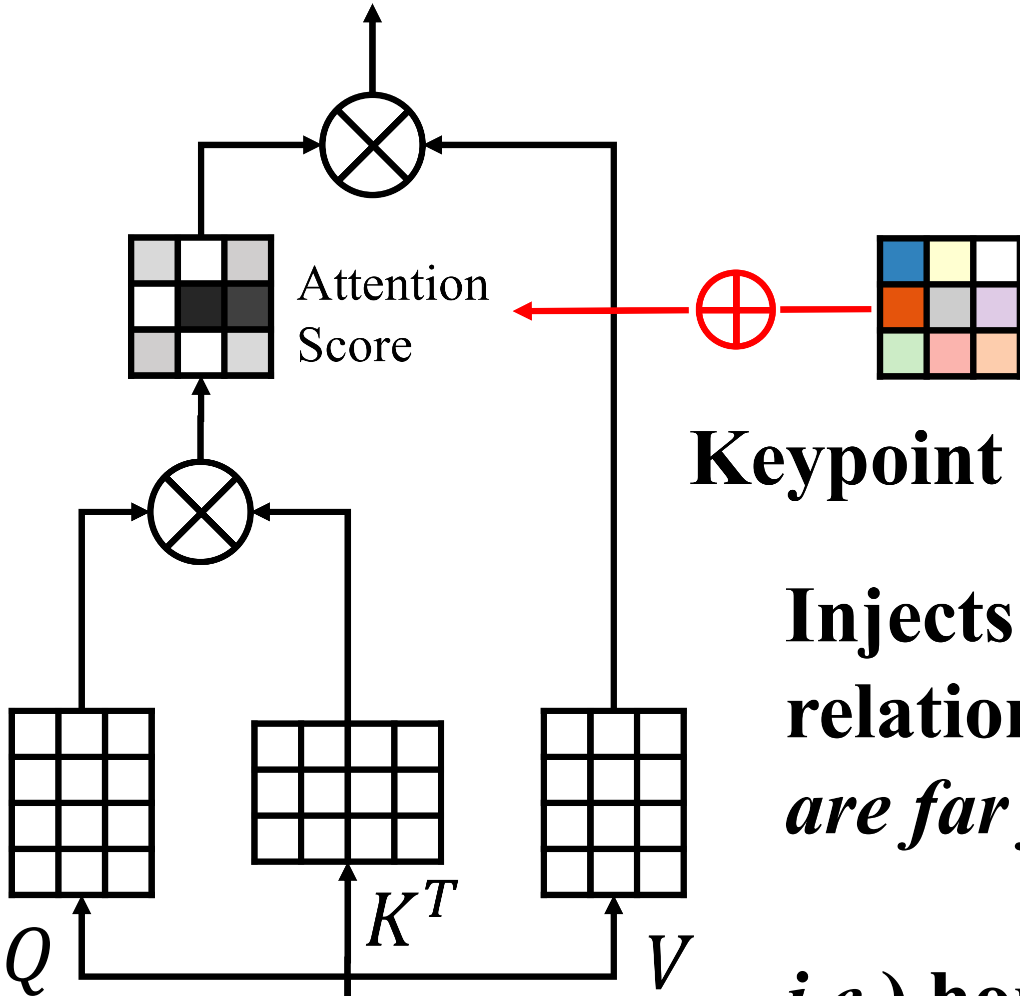
$\vdots$

$= W_1(L - (q_x, q_y))$

$= W_2(L - (q_x, q_y))$

* Same color implies same distance

Summary



Keypoint Relative Position Encoding

Attention Mechanism

Injects priors in pairwise relationship based *on how they are far from keypoints*.

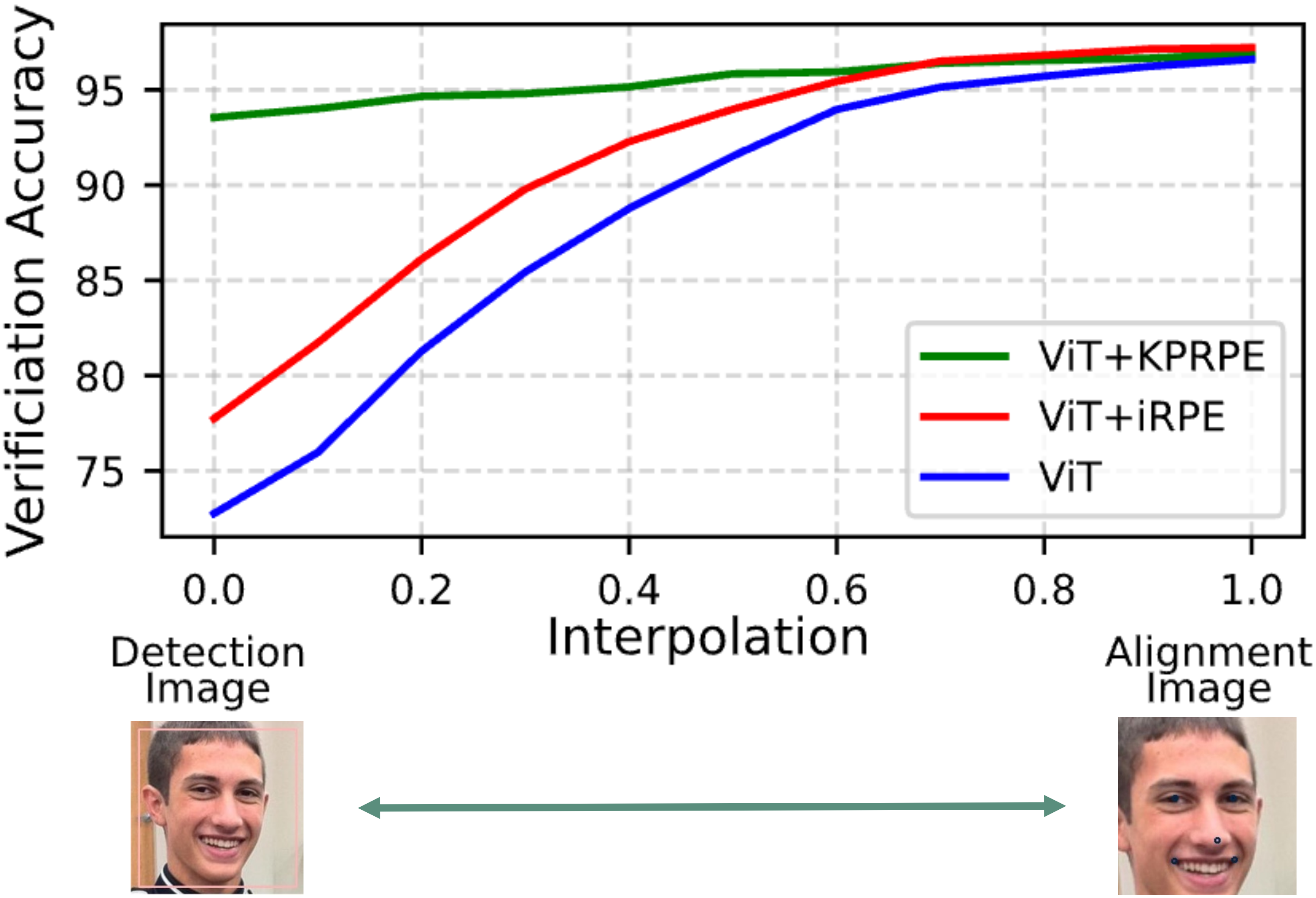
***i.e.*) horizontal points are important near the eyes.**

Experiments

Method	Backbone	Train Data	Low Quality Dataset			
			TinyFace [7]		IJB-S [29]	
			Rank-1	Rank-5	Rank-1	Rank-5
PFE [61]	CNN64	MS1MV2 [11]	-	-	50.16	58.33
ArcFace [11]	ResNet101	MS1MV2 [11]	-	-	57.35	64.42
URL [62]	ResNet101	MS1MV2 [11]	63.89	68.67	59.79	65.78
CurricularFace [27]	ResNet101	MS1MV2 [11]	63.68	67.65	62.43	68.68
AdaFace [11]	ResNet101	MS1MV2 [11]	68.21	71.54	65.26	70.53
AdaFace [11]	ResNet101	MS1MV3 [13]	67.81	70.98	67.12	72.67
AdaFace [30]	ViT	MS1MV3 [13]	72.05	74.84	65.95	71.64
AdaFace [30]	**ViT+KP-RPE**	MS1MV3 [13]	**73.50**	**76.39**	**67.62**	**73.25**
ArcFace [11]	ResNet101	WebFace4M [91]	71.11	74.38	69.26	74.31
AdaFace [30]	ResNet101	WebFace4M [91]	72.02	74.52	70.42	75.29
AdaFace [30]	ViT	WebFace4M [91]	74.81	77.58	71.90	77.09
AdaFace [30]	ViT+iRPE	WebFace4M [91]	74.92	77.98	71.93	77.14
AdaFace [30]	**ViT+KP-RPE**	WebFace4M [91]	**75.80**	**78.49**	**72.78**	**78.20**
AdaFace [30]	ResNet101	WebFace12M [91]	72.42	74.81	71.46	77.04
AdaFace [30]	**ViT+KP-RPE**	WebFace12M [91]	**76.18**	**78.97**	**72.94**	**77.46**

Large performance improvements in
Hard / Low quality Imagery Datasets.

Experiments



KP-RPE even shows robust to
image transformations unseen during training

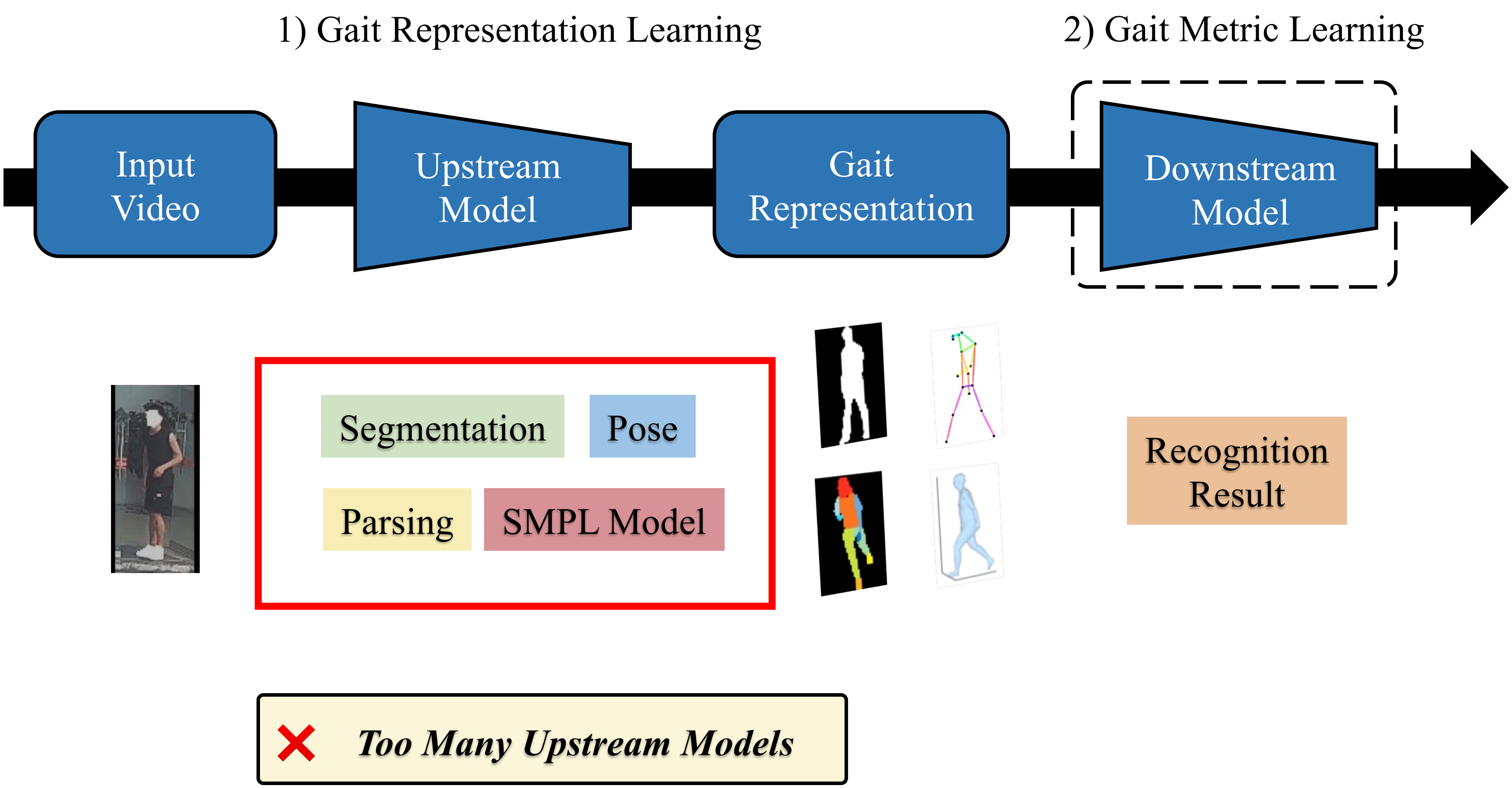
Highlights

- **7. Large vision models** (CVPR'24)

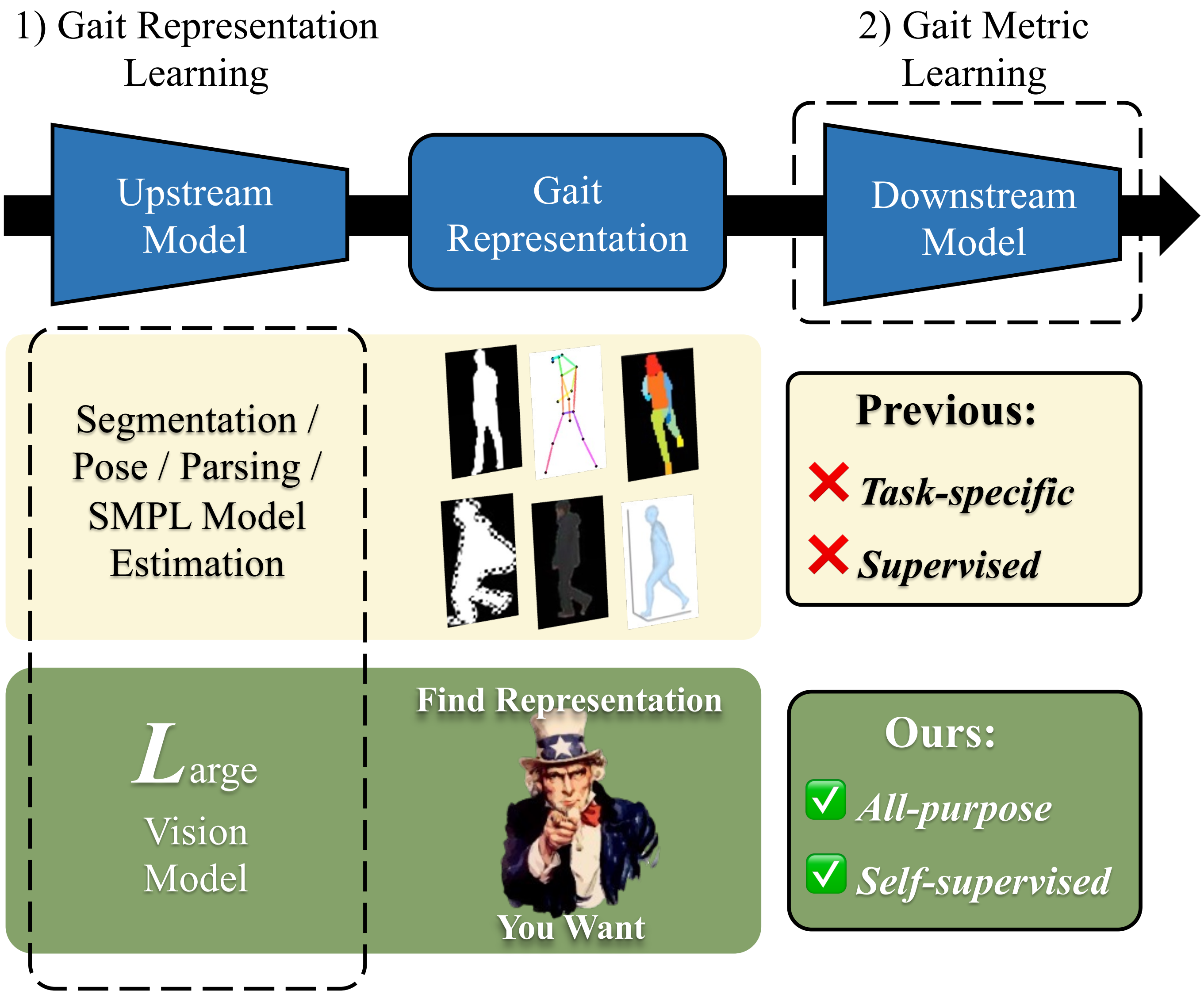
Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, Shiqi Yu, "BigGait: Learning Gait Representation You Want by Large Vision Models," in CVPR 2024

Problem of Previous Methods

Gait Recognition Pipeline



Motivation

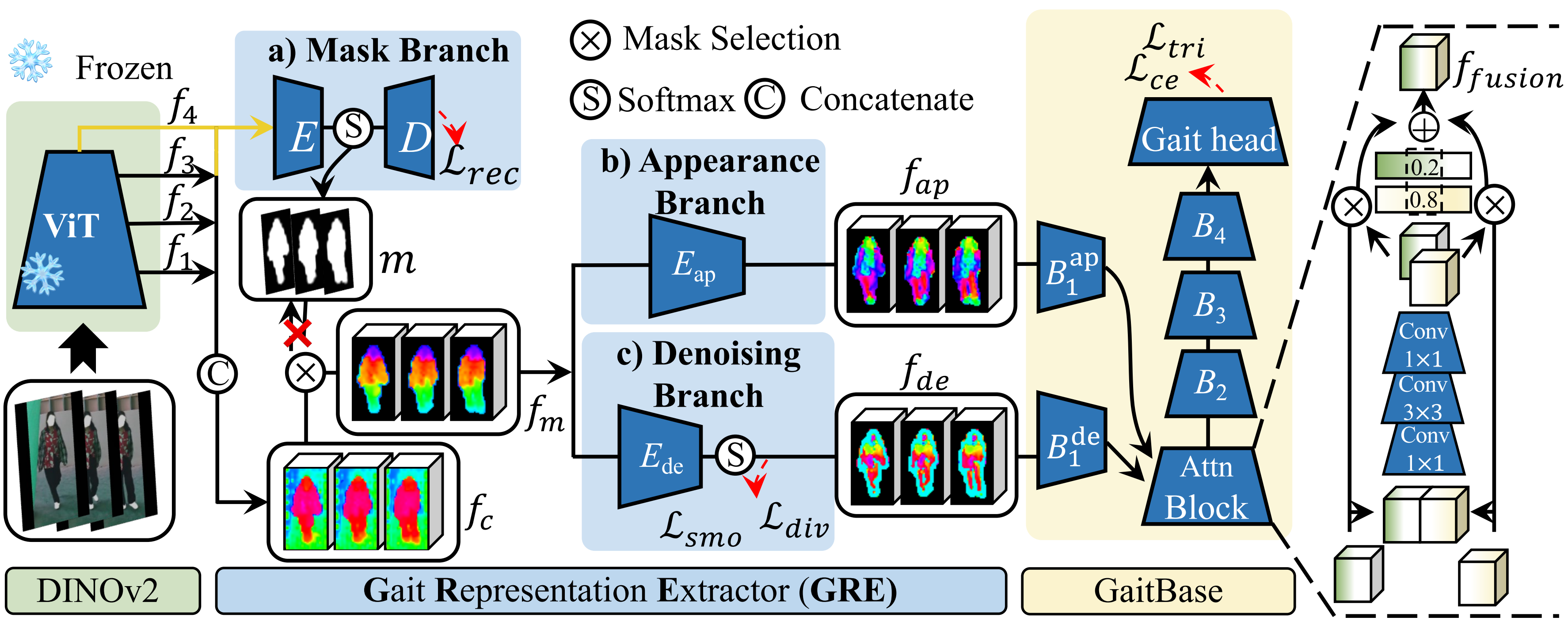


The idea of BigGait

One Large Vision Model = Multiple Specific Models

Method

Architecture

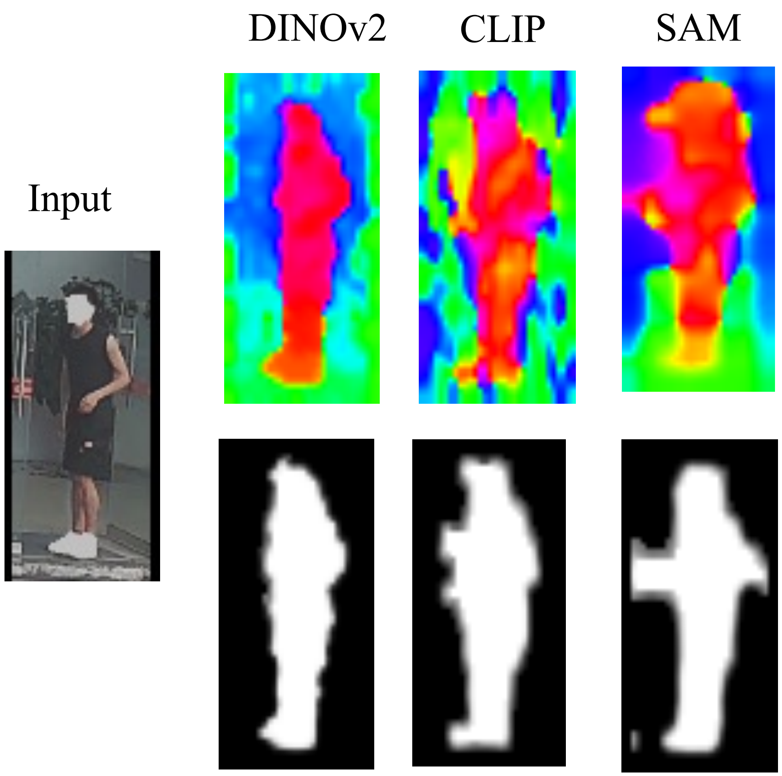


The architecture of BigGait

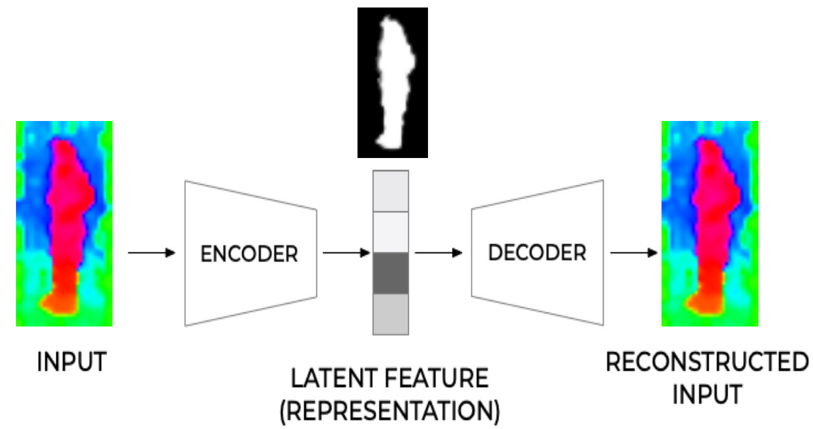
Ye et.al., BigGait: Learning Gait Representation You Want by Large Vision Models, CVPR, 2024.

Method

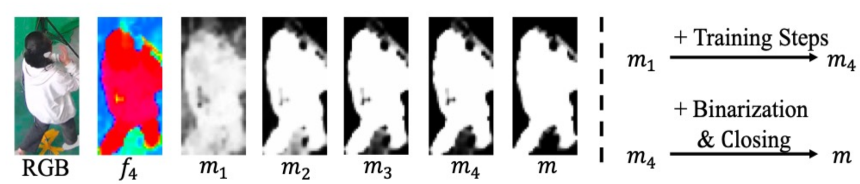
Architecture (Mask Branch)



PCA/Mask Visualization of feature from multi-LVMs



Mask Branch is An Unsupervised Auto-Encoder

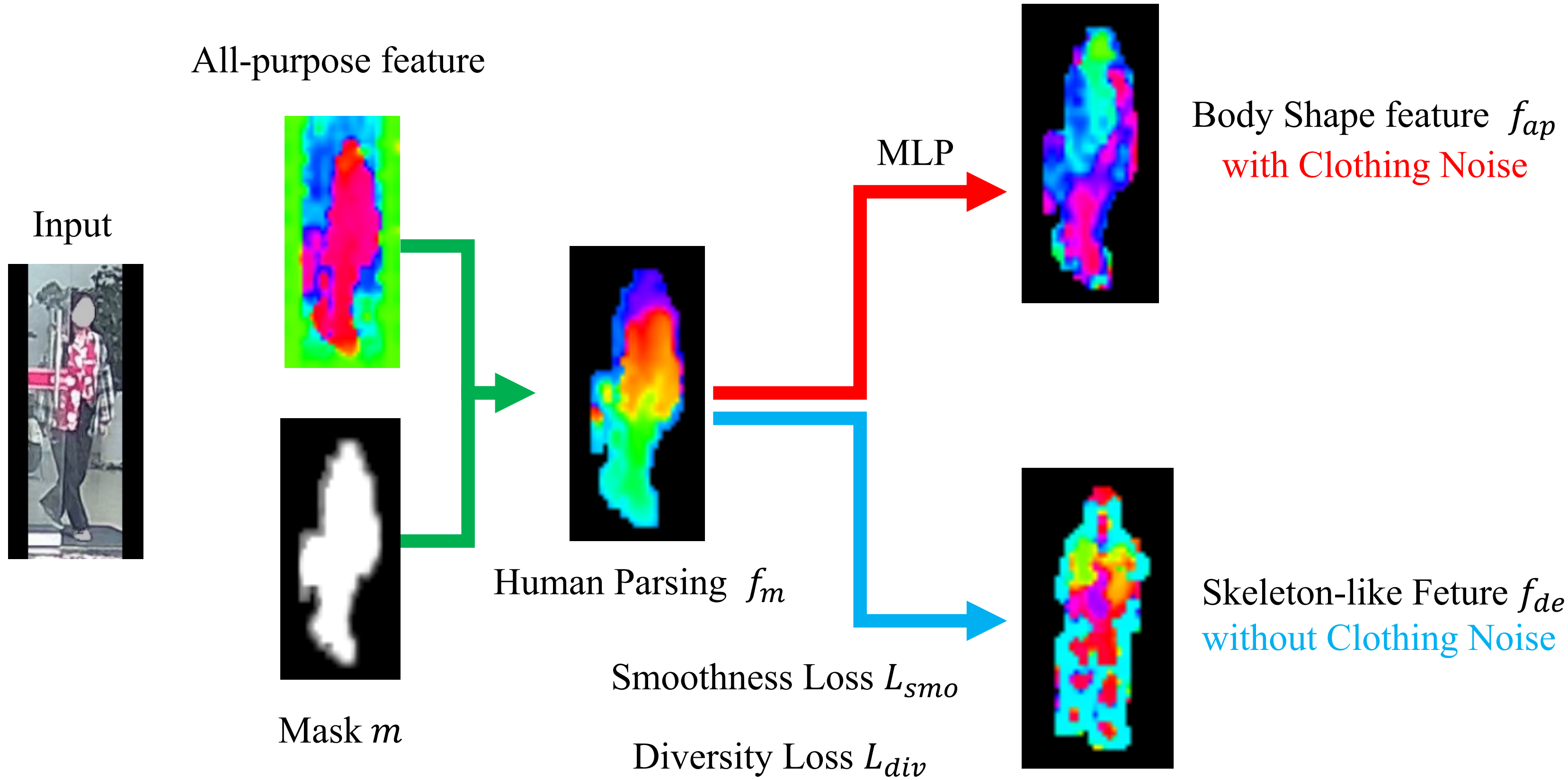


Red Region Extracting

Oquab, Maxime, et al. Dinov2: Learning robust visual features without supervision. Arxiv, 2023.
Radford, Alec, et al. Learning transferable visual models from natural language supervision. ICML, 2021.
Kirillov, Alexander, et al. Segment anything. ICCV, 2023.

Method

Architecture (Appearance/ Denoising Branch)

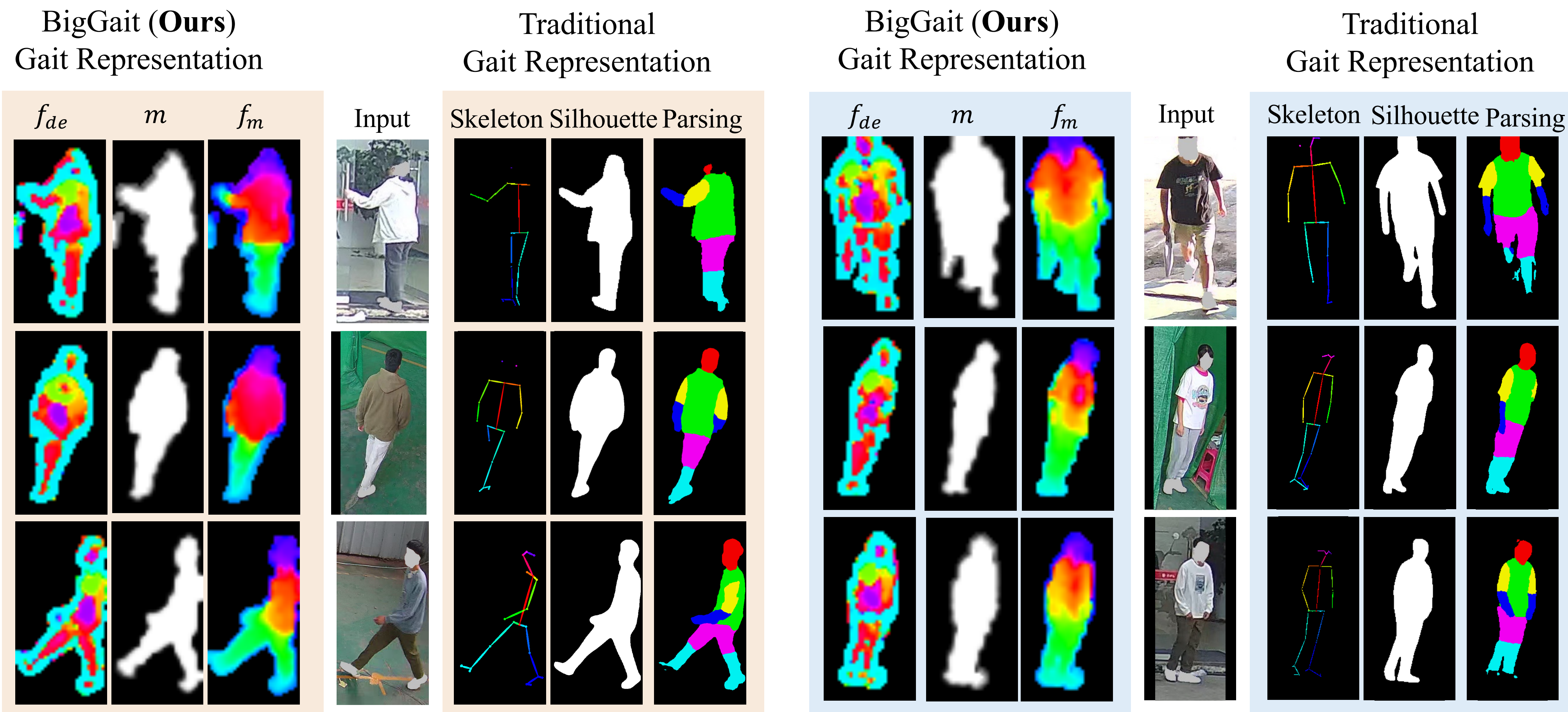


$$f_{de} = \mathrm{softmax}(E_{\mathrm{de}}(f_m))$$
$$L_{smo} = |\mathrm{sobel_x} * f_{de}| + |\mathrm{sobel_y} * f_{de}|,$$

$$p_i = \mathrm{sum}(f_{de}^i) / \sum_{i=1}^{C} \mathrm{sum}(f_{de}^i)$$

$$L_{div} = \log C + \sum_{i=1}^{C} p_i \log p_i,$$

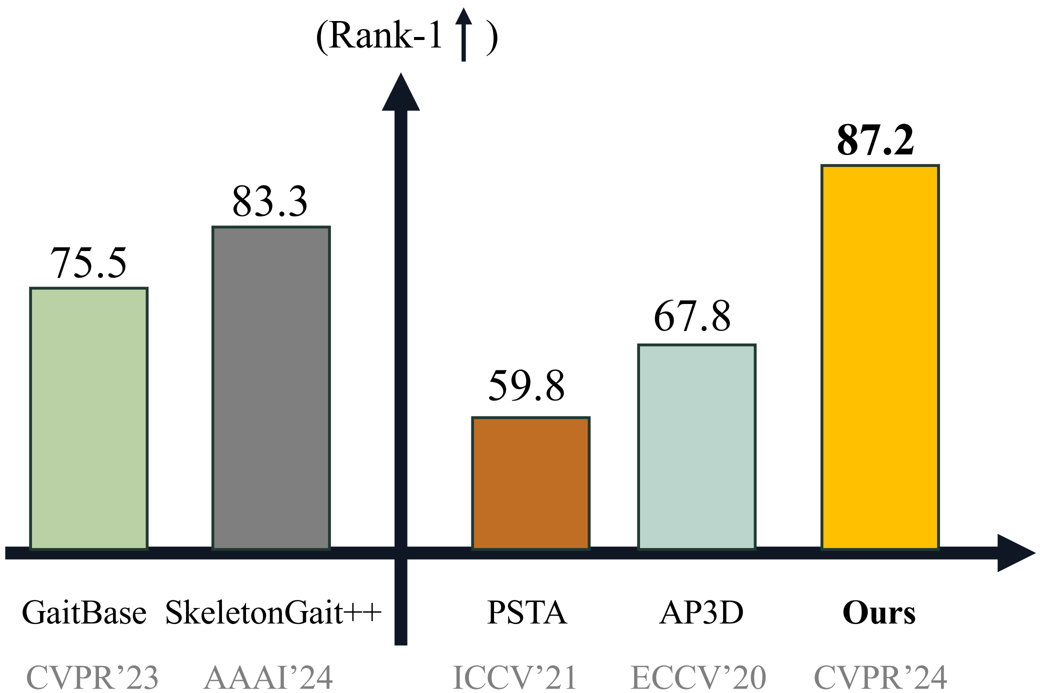
Visualization

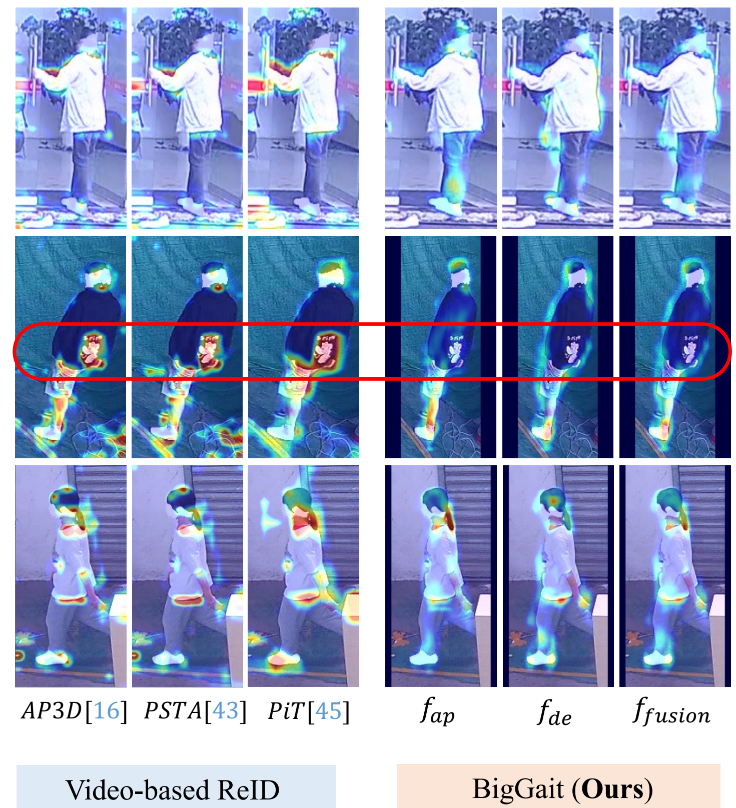


✅ ***One* Large Vision Model ≈ *Multiple* Traditional Gait Extraction Models**

Experiment

Performance on CCPG Dataset





Activation Map Visualization

Highlights

➢ **9. CLIP 3D Re-ID**
(CVPR'24)

Feng Liu, Minchul Kim, Zhiyuan Ren, Xiaoming Liu, "Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation," in CVPR 2024

Whole-Body Matching

- ➢Person re-identification (static, body characteristics)
- ➢Gait recognition (dynamic, walking patterns)

Person Re-ID

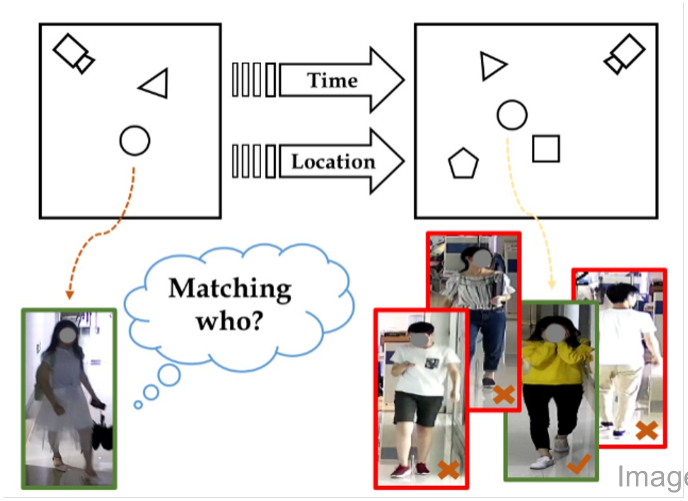


Image Courtesy [1]

Gait Recognition

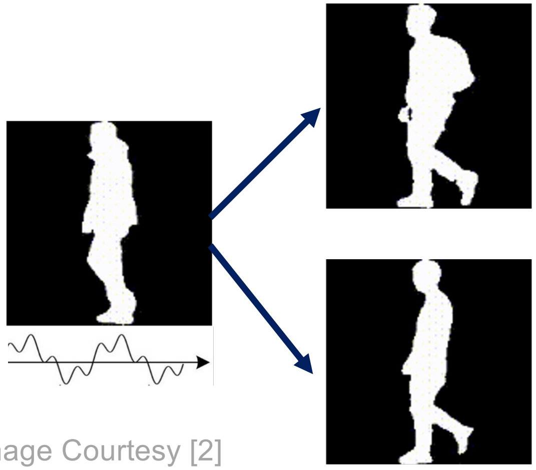


Image Courtesy [2]

[Long-Term Cloth-Changing Person Re-identification. ACCV 2020. [2] https://github.com/ShiqiYu/OpenGait

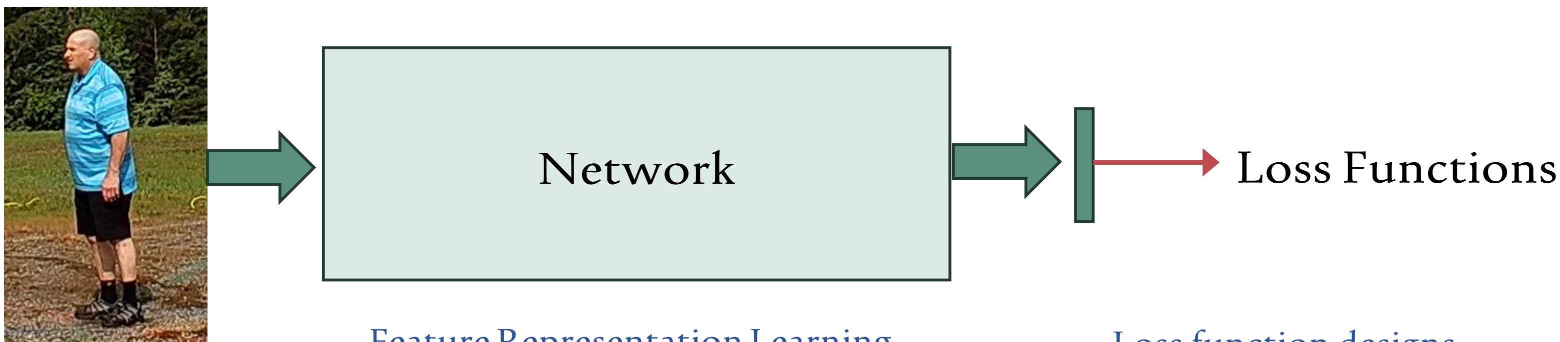
Person ReID Challenges

- Person ReID challenges lie in learning a discriminative, robust visual representation against diverse variations (view/pose and appearance)

All people love yellow shirt and short pants?

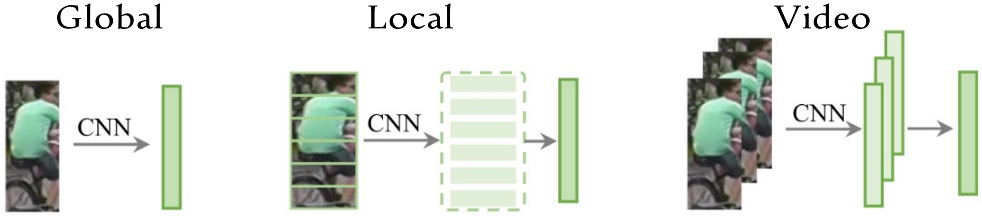
Zheng et al, Deep learning for person re-identification. https://www.zdzheng.xyz/seminar/

Standard Person Re-ID System



Feature Representation Learning

Loss function designs



- TransReID, ICCV '21
- HPM, AAAI '19
- MEVID, WACV '23

- Identity loss,
- Verification loss
- Triplet loss
- Clothes-based adversarial loss (CVPR '22)

Deep Learning for Person Re-Identification: A Survey and Outlook. TPAMI 2022

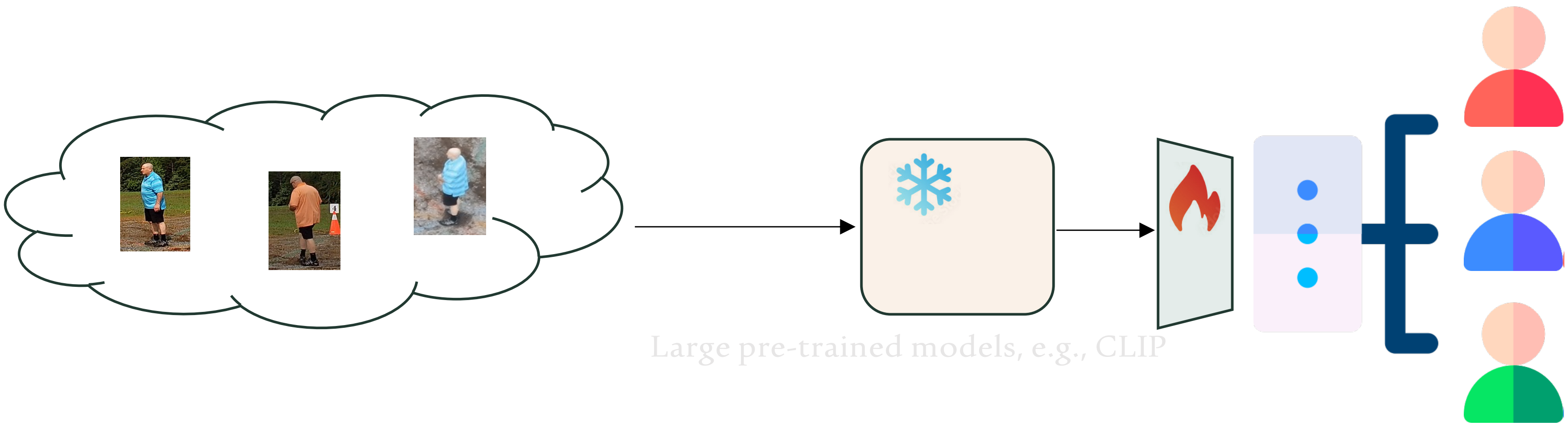
Bottleneck

➢ Existing datasets have limited identities and variations

Dataset	Year	# ID	# Sample
Celeb-reID	2019	1,052	34,186
PRCC	2019	221	33,698
LTCC	2020	152	17,138
COCAS	2020	5,266	62,832
VC-Clothes	2020	512	19,060
DeepChange*	2021	1,082	171,352
LaST*	2022	10,860	224,721
CCVID	2022	226	347,833
WebFace260M	2018	4M	260M

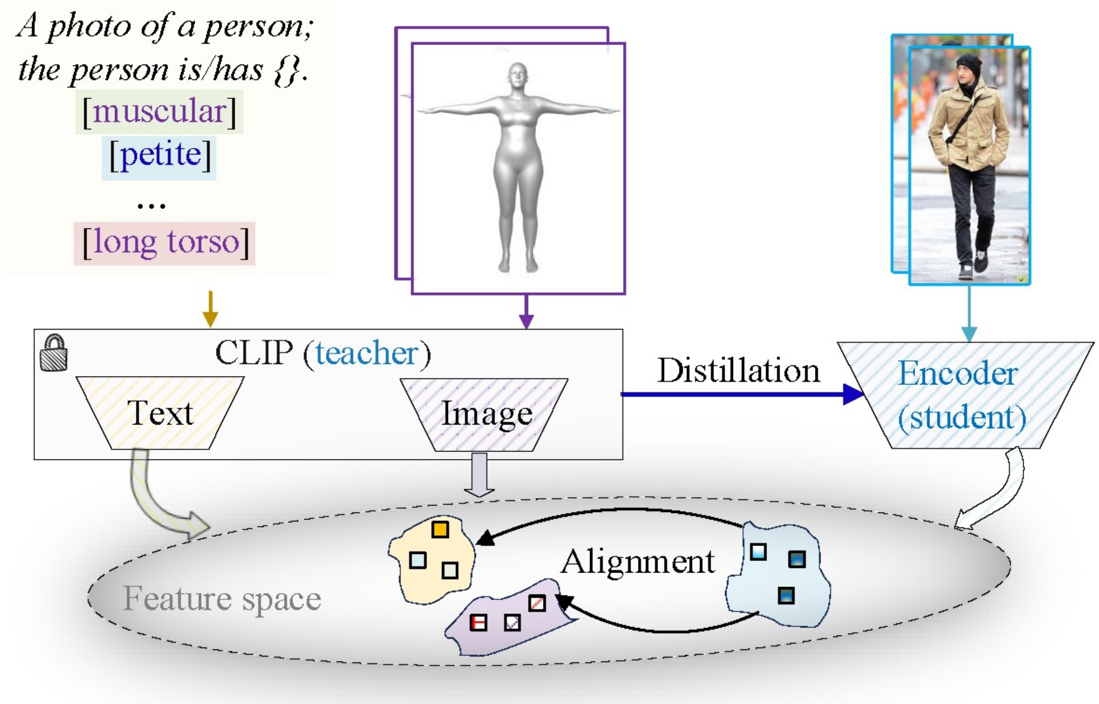
Motivation

➢Enhancing feature robustness and generalization by distilling knowledge from pre-trained large models



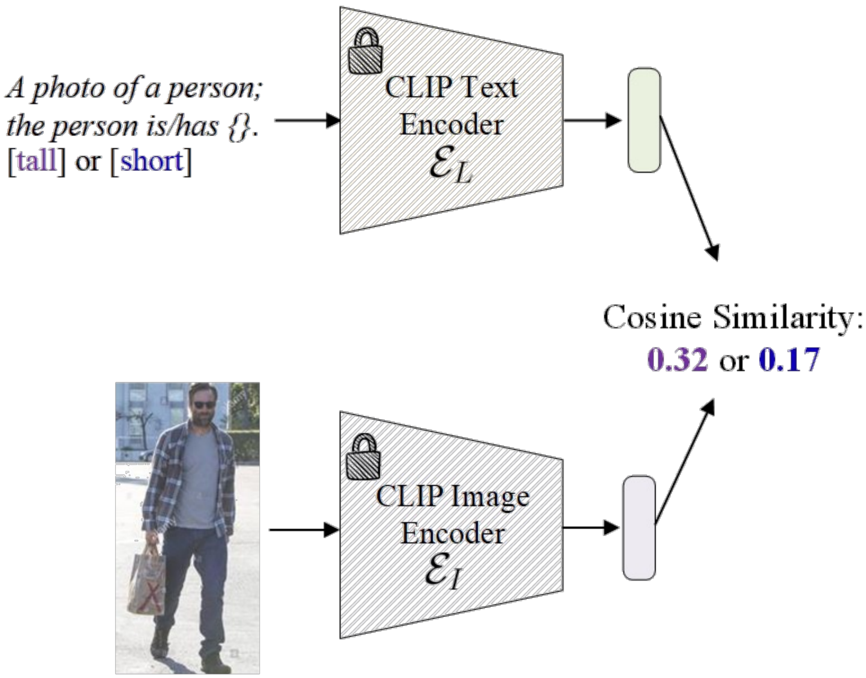
Motivation

➢ Enhancing feature robustness and generalization by distilling knowledge from pre-trained large models (e.g., CLIP)



Motivation

➢Distilling discriminative body shape representation from the CLIP model



	Phrase 1		Phrase 2
1	Muscular	↔	Slender
2	Broad-Shouldered	↔	Narrow-Shouldered
3	Heavyset	↔	Petite
4	Tall	↔	Short
5	Long Legs	↔	Short Legs
6	Long Torso	↔	Short Torso
7	Curvy	↔	Angular
8	Full-Figured	↔	Skinny
9	Stocky	↔	Willowy
10	Pear-Shaped	↔	Apple-Shaped
11	Athletic	↔	Non-Athletic
12	Fit	↔	Unfit
13	Large-Breasted	↔	Small-Breasted
14	Long-Armed	↔	Short-Armed
15	Long-Necked	↔	Short-Necked
16	High-Waisted	↔	Low-Waisted

Motivation

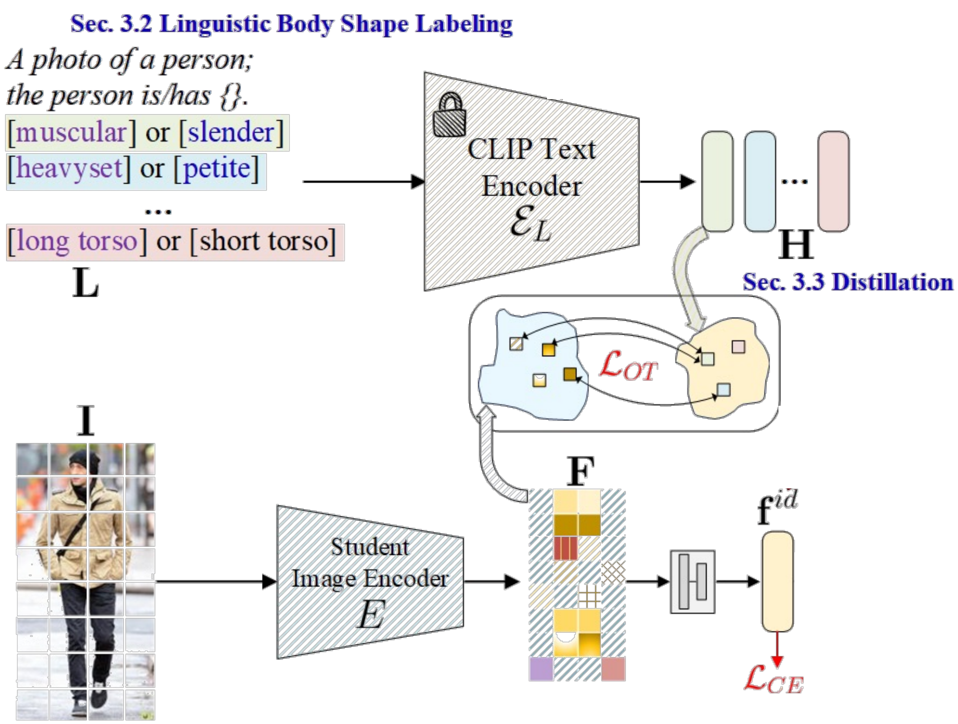
- Labeling linguistic body description

Low-waisted
Short-necked
Big-handed
Small-breasted
Unfit
Non-althletic
Pear-shaped
Willowy
Skinny
Angular
Short torso
Short legs
Short
Petite
Narrow-shouldered
Slender

Low-waisted
Long-necked
Small-handed
Small-breasted
Unfit
Non-althletic
Pear-shaped
Willowy
Skinny
Curvy
Short torso
Short legs
Short
Petite
Narrow-shouldered
Slender

Our Method

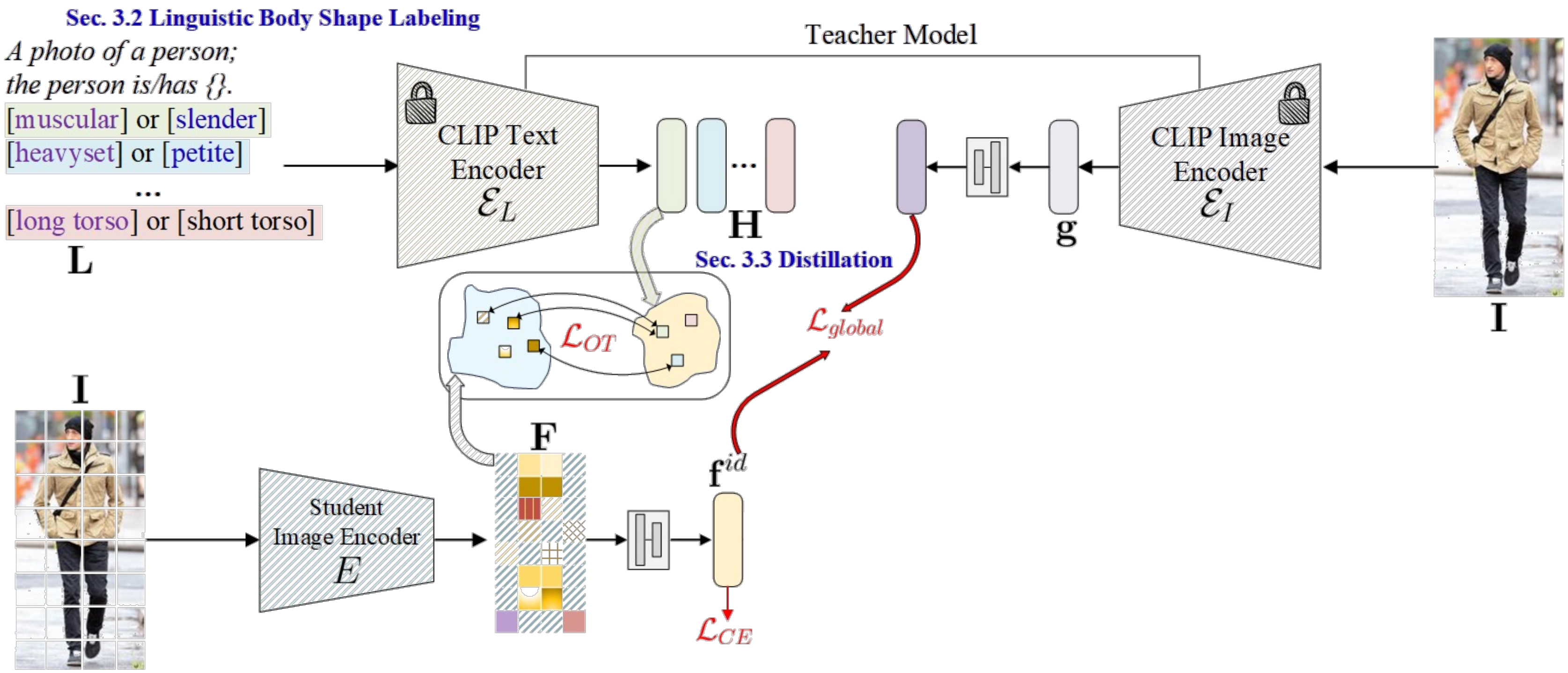
➢ Distilling discriminative body shape representation from the CLIP model



Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation, CVPR 2024

Our Method

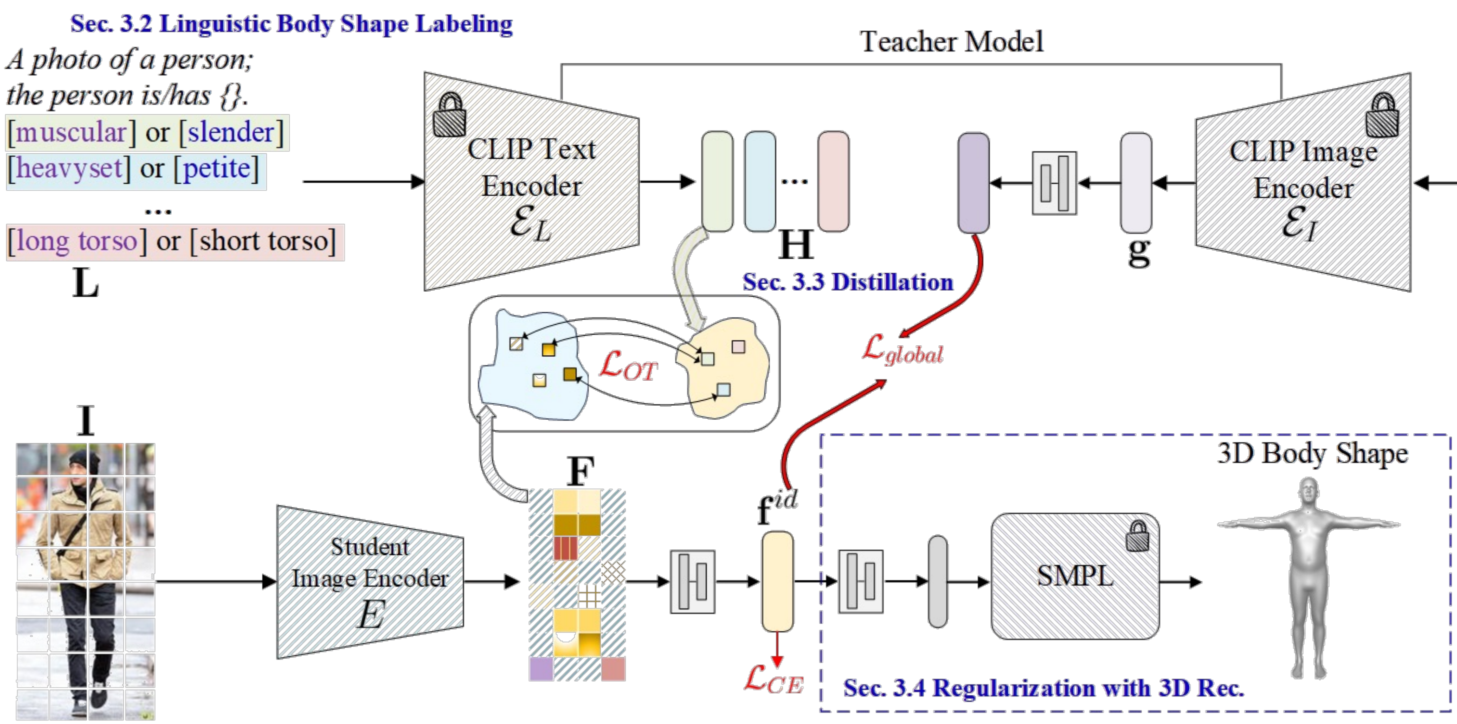
➢ Distilling discriminative body shape representation from the CLIP model



Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation, CVPR 2024

Our Method

➢Distilling discriminative body shape representation from the CLIP model



Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation, CVPR 2024

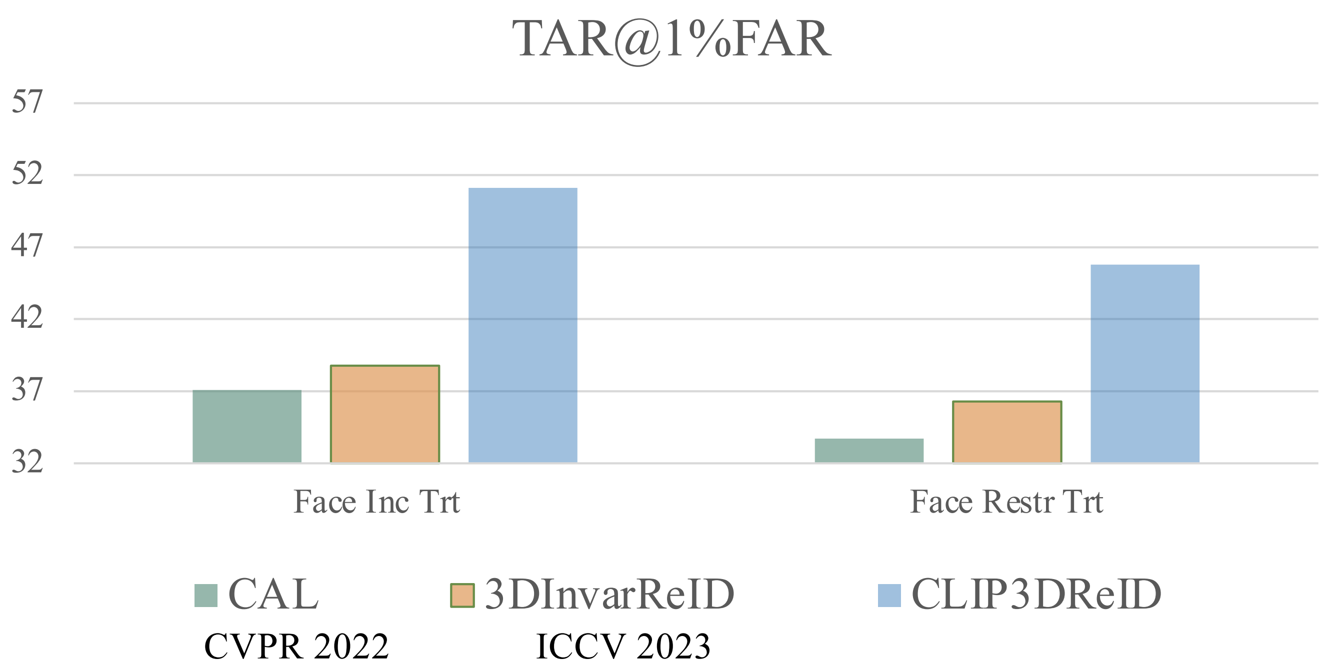
3D Body Reconstruction

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

Feng Liu, Minchul Kim, ZiAng Gu, Anil Jain, Xiaoming Liu, “Learning Clothing and Pose Invariant 3D Shape Representation for Long-Term Person Re-Identification,” in ICCV 2023

Results

➢Verification performance on BRIAR dataset



Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation, CVPR 2024

Highlights

➢ **9. Unified human recognition**
(under review)

Minchul Kim, Dingqiang Ye, Yiyang Su, Feng Liu, Xiaoming Liu, "SapiensID: Foundation for Human Recognition," under review in CVPR 2025.

Motivation

Can one model perform comparison across different body pose and visual area?

Minchul Kim, Dingqiang Ye, Yiyang Su, Feng Liu, Xiaoming Liu, SapiensID: Foundation for Human Recognition, under review in CVPR 2025

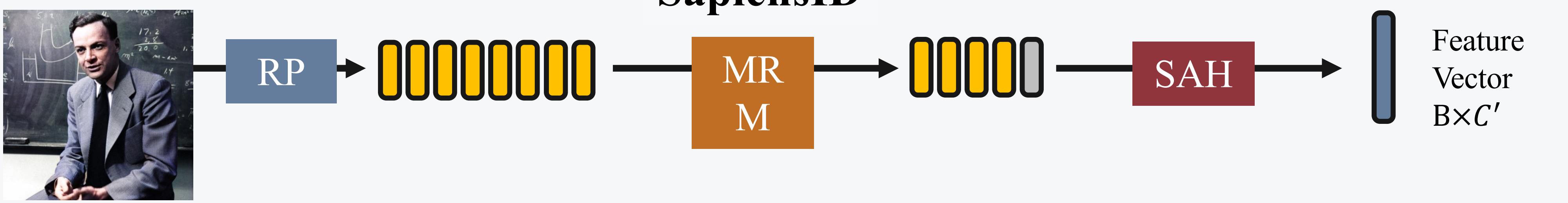
WebBody Dataset

4 Million Labeled Images   263,920 Subjects   Large Pose–Scale Variation

Overview

Previous Methods: Cannot handle large pose-scale variation

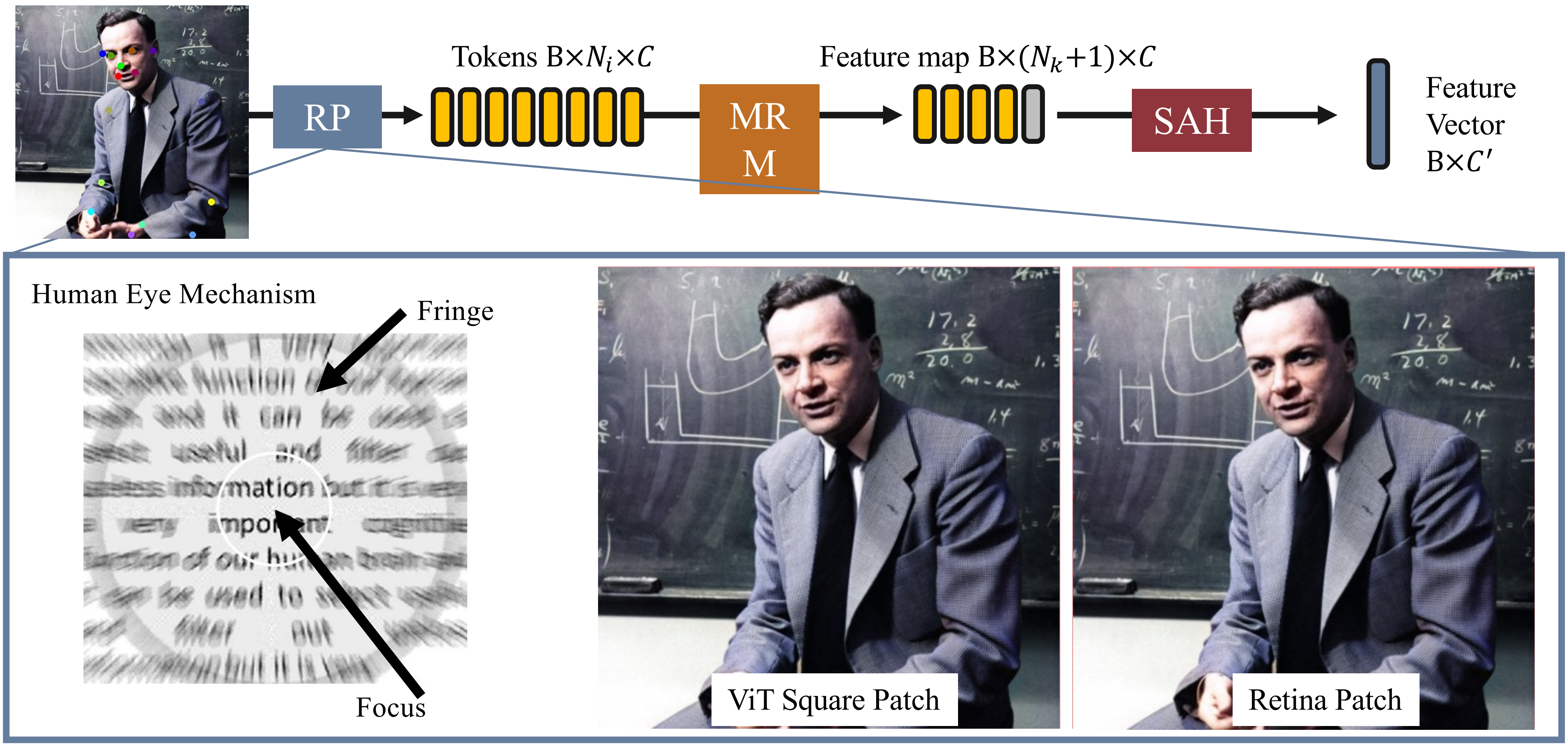
SapiensID



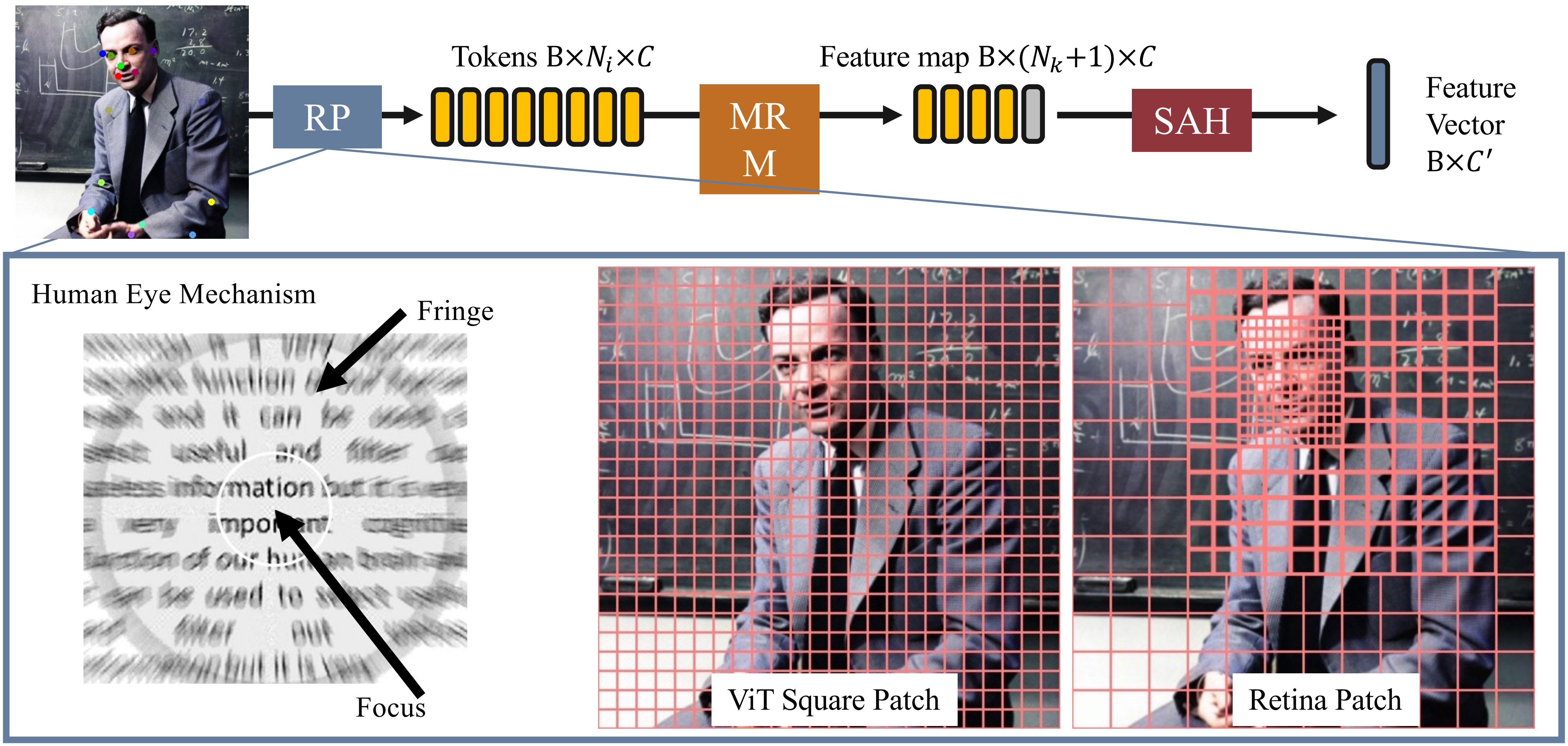
SapiensID proposes 3 things to **handle large pose and scale variation.**

RP: Retina Patch MRM: Masked Recognition Model SAH: Semantic Attention Pooling

1. Retina Patch



1. Retina Patch



1. Retina Patch

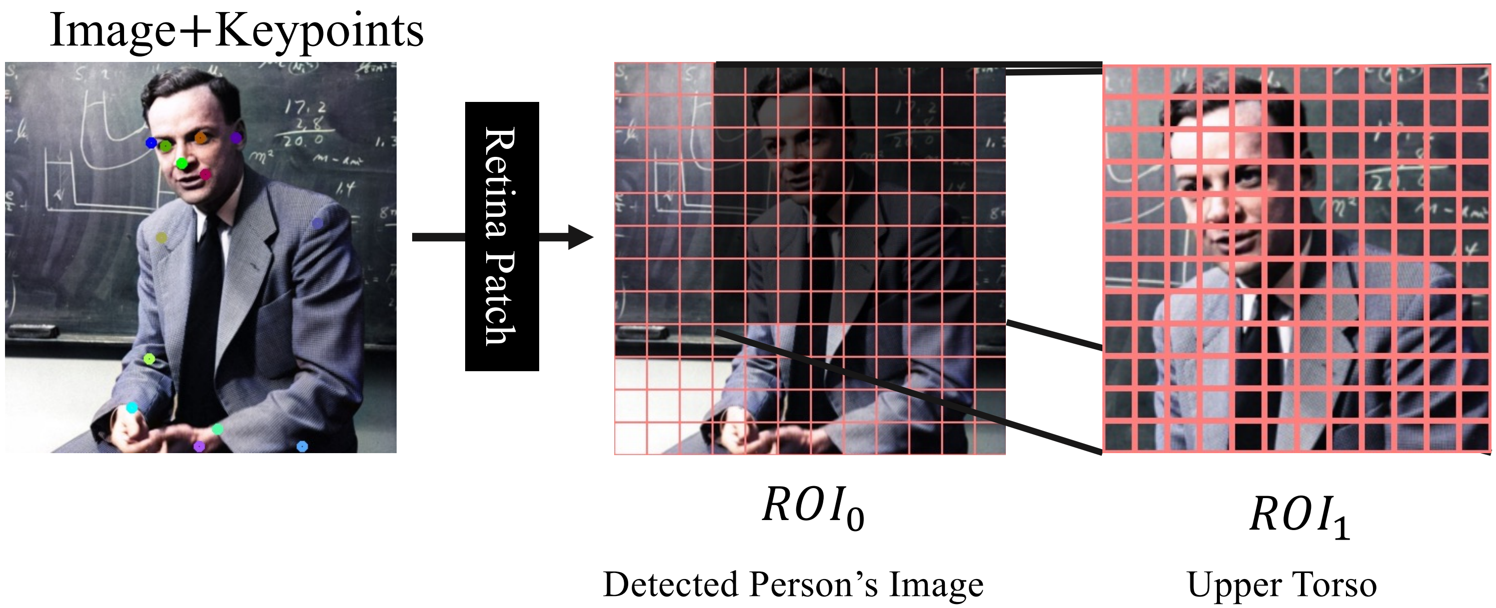
Image+Keypoints

Retina Patch

ROI_0

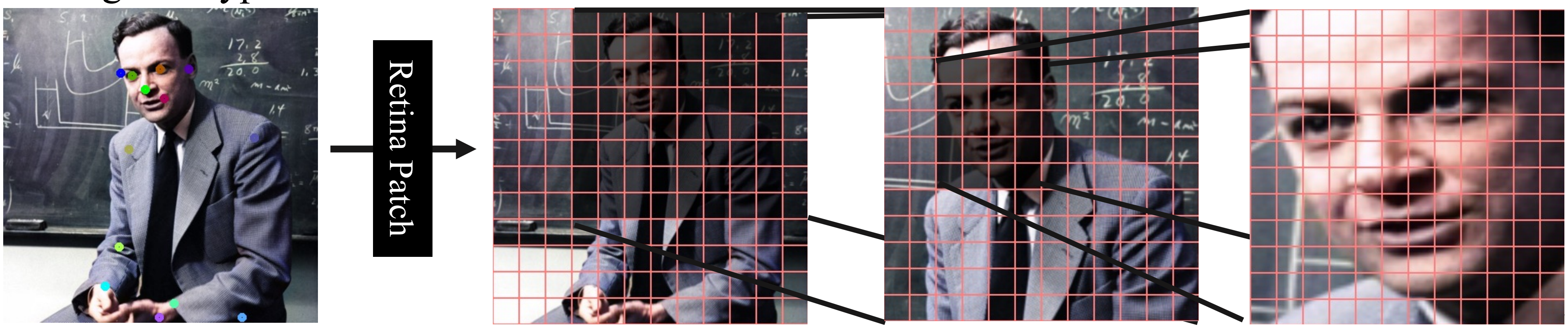
Detected Person's Image

1. Retina Patch



1. Retina Patch

Image+Keypoints



ROI_0

Detected Person's Image

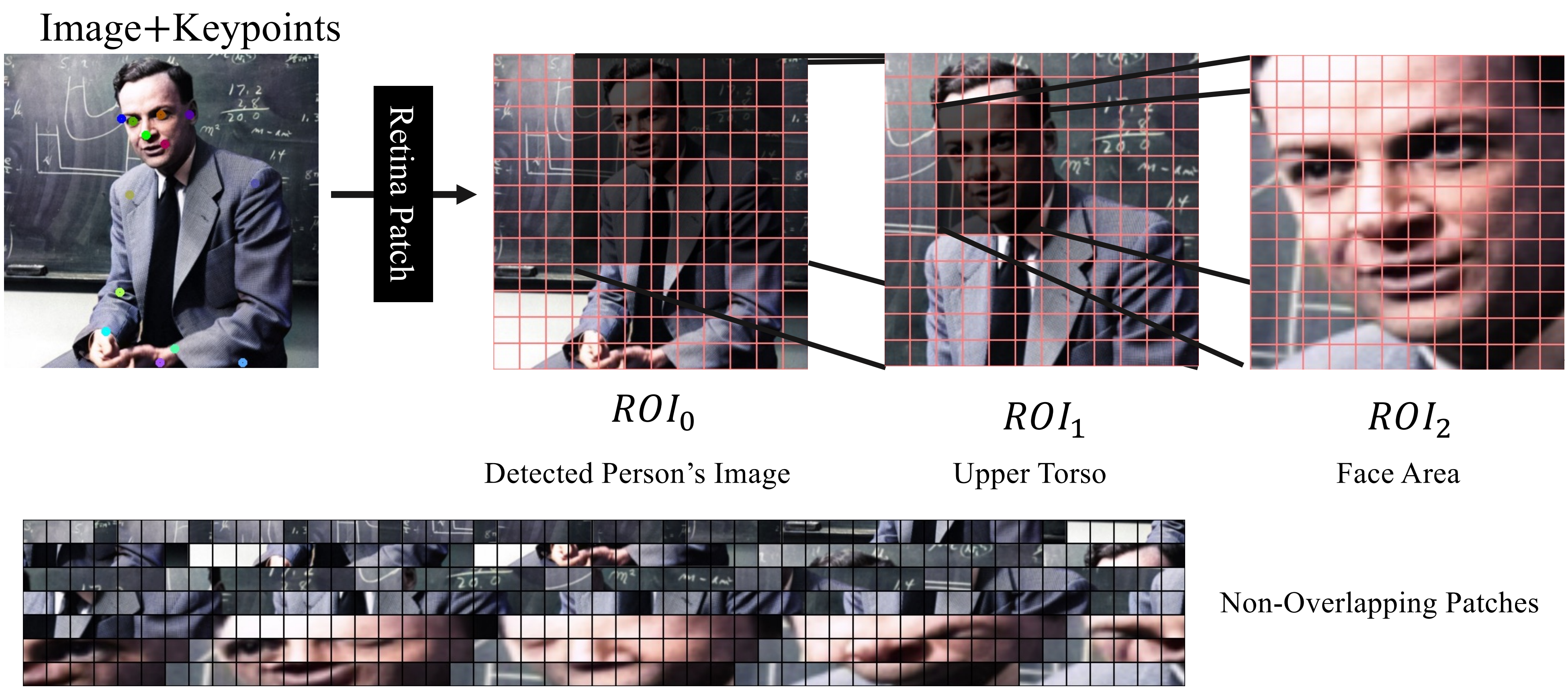
ROI_1

Upper Torso

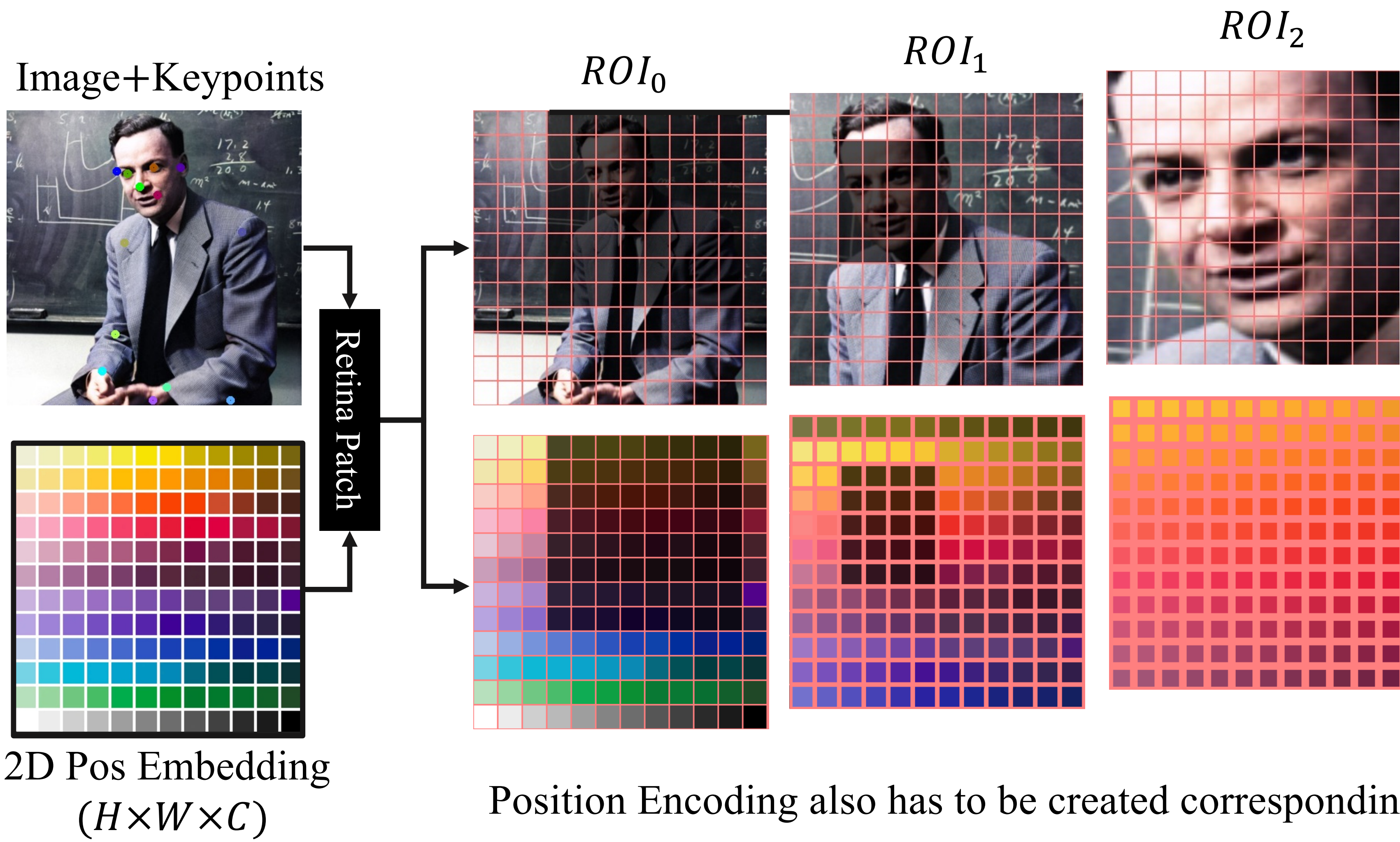
ROI_2

Face Area

1. Retina Patch

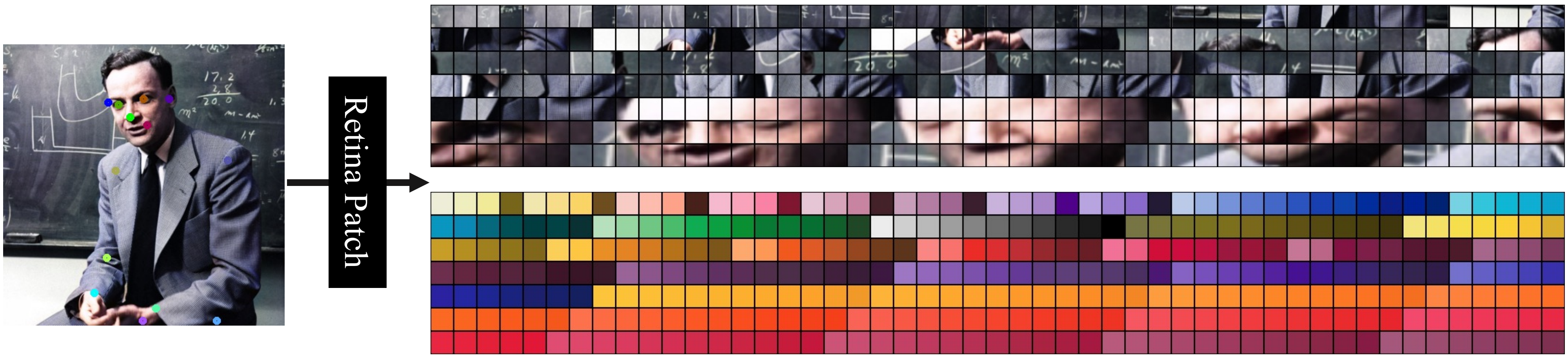


1. Retina Patch

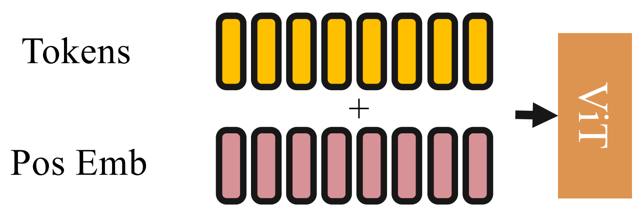


Position Encoding also has to be created correspondingly.

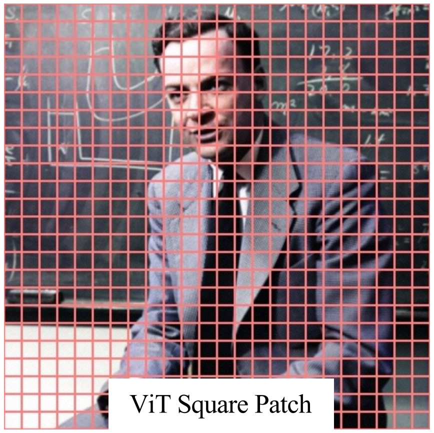
1. Retina Patch



Different sizes of patches are resized to a same sized and projected to tokens.



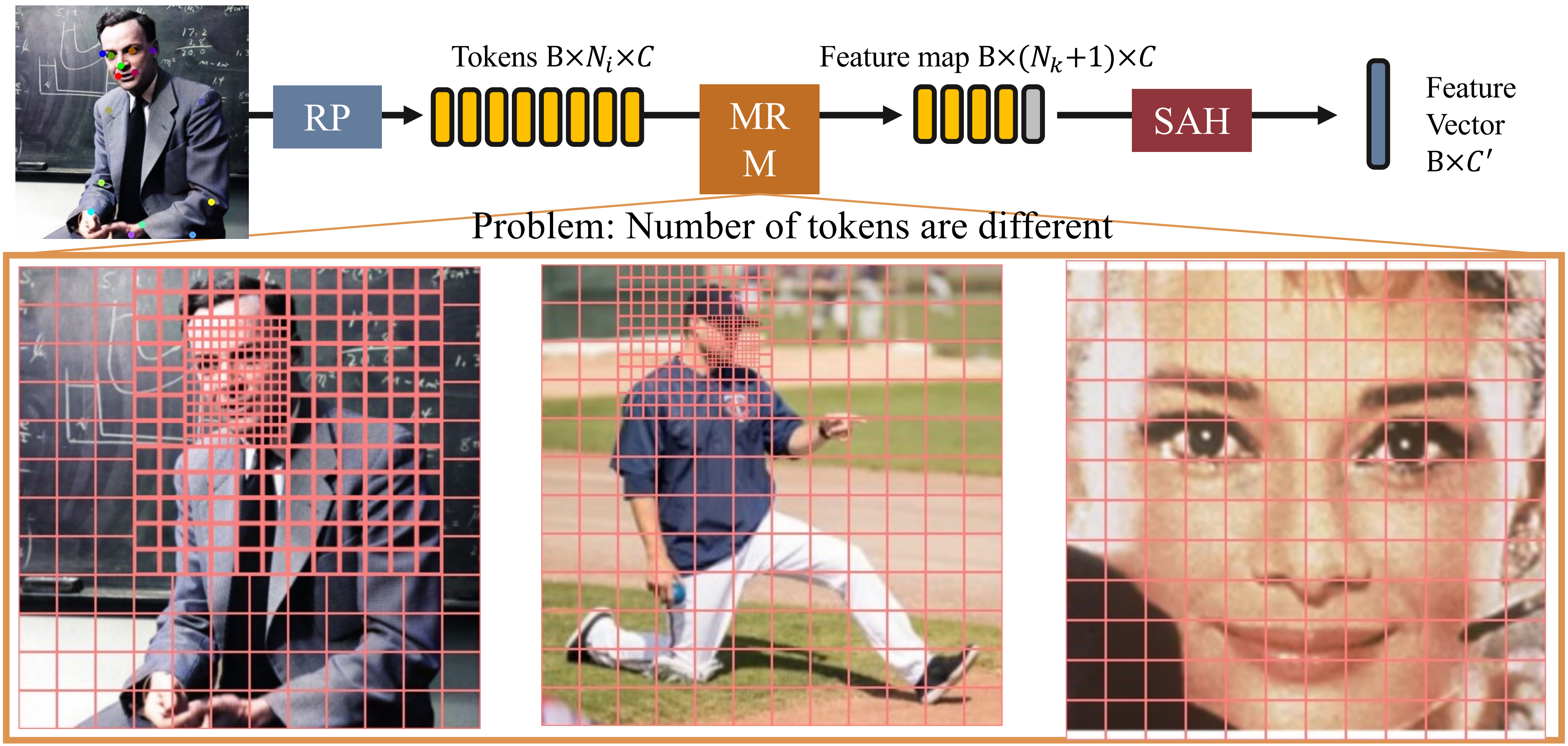
1. Retina Patch



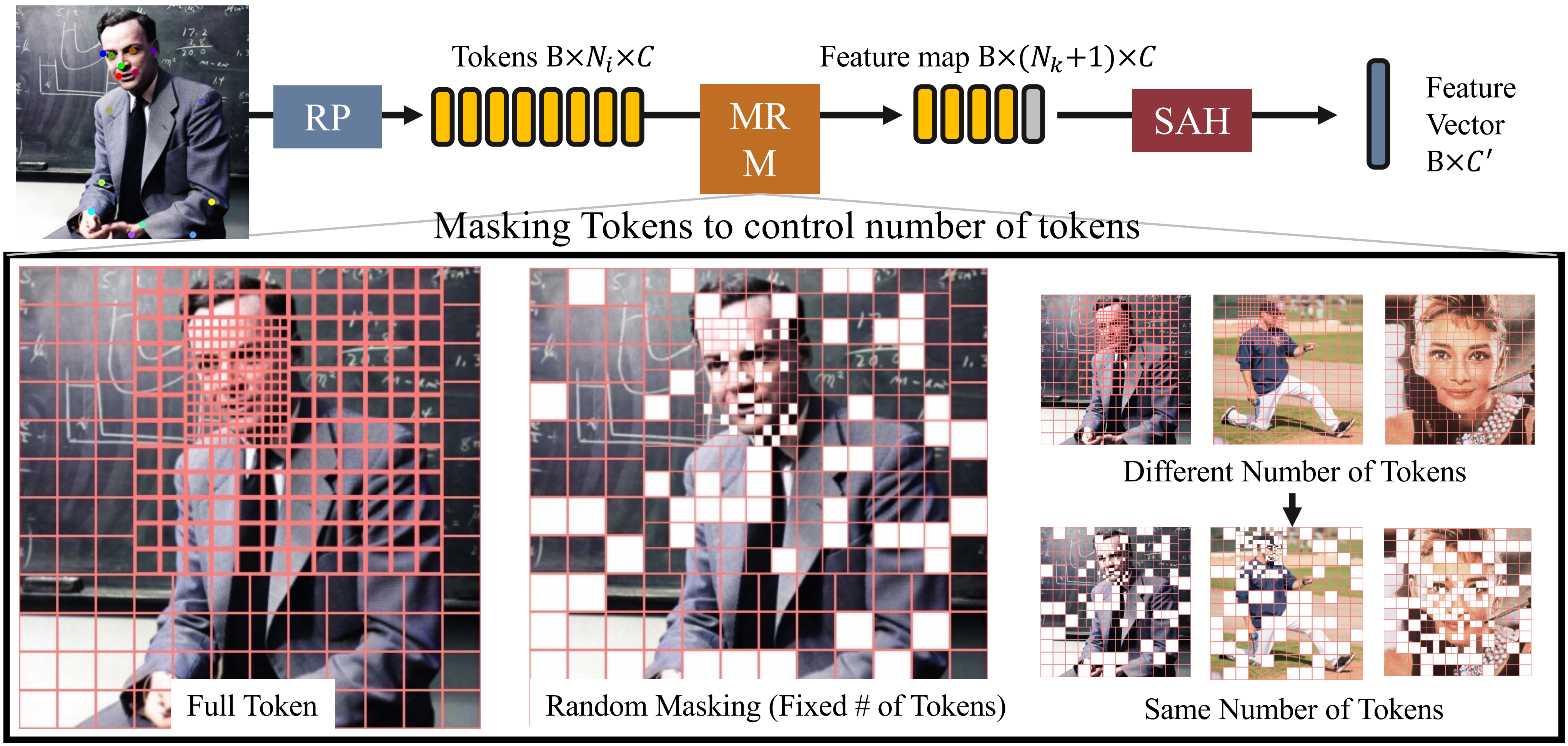
	All	Face	Whole Body ReID Short	Whole Body ReID Long
SoTA Face and Body Models	67.97	**97.63**	61.49	44.90
(1) ViT	59.54	90.63	56.17	31.81
(2) ViT+RP	66.35	92.93	59.16	46.95
(3) ViT+SAH	71.67	95.84	72.63	46.55
(4) ViT+RP+SAH (SapiensID)	**78.67**	**97.31**	**73.05**	**66.30**
(4) without Random Mask Ratio	74.39	95.95	69.58	57.64

Performance Impact of using Retina Patch vs Square Patching

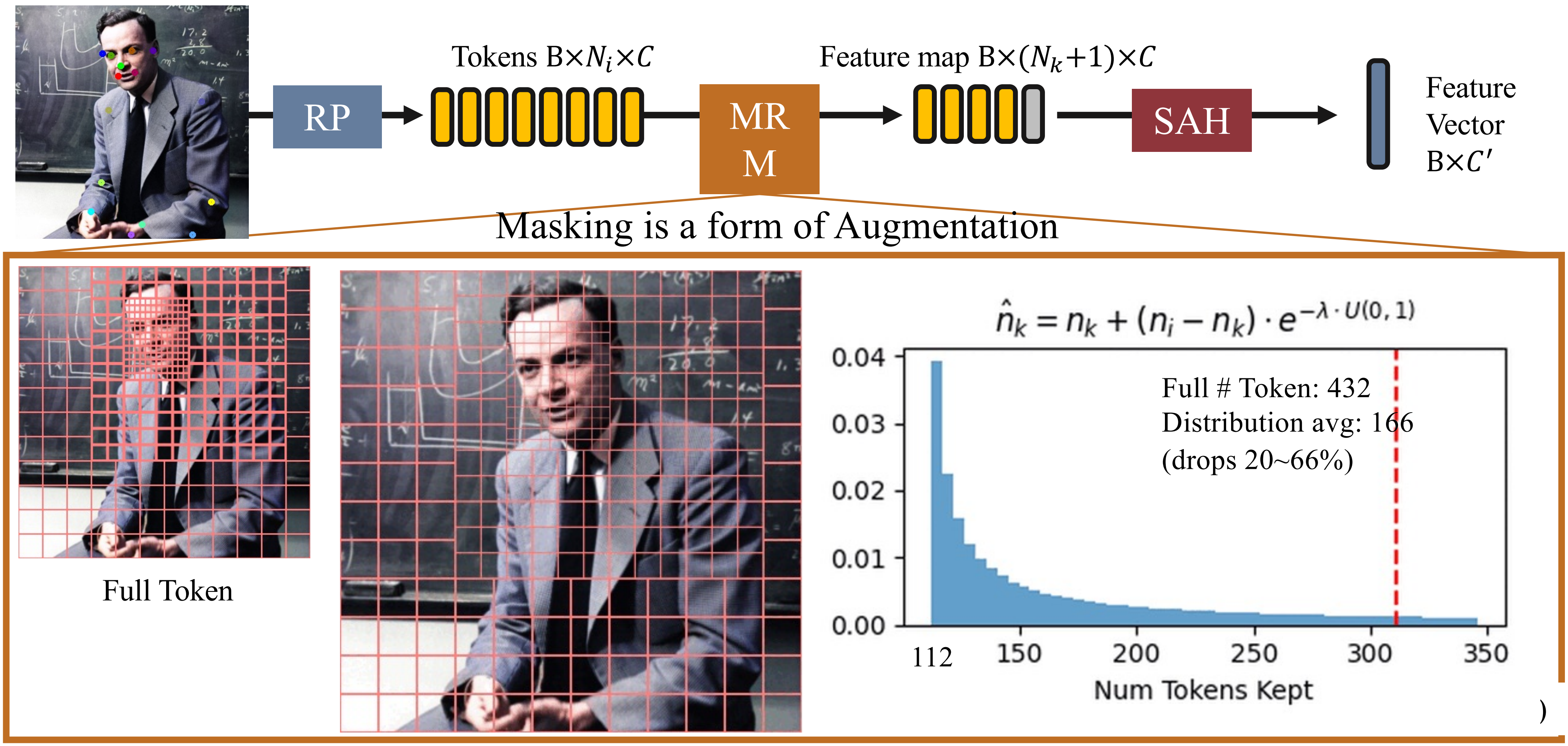
1. Retina Patch



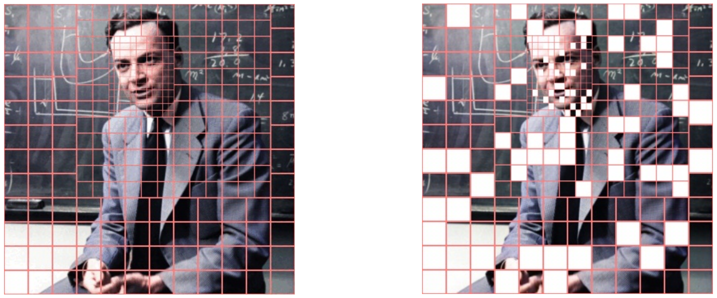
2. Masked Recognition Model



2. Masked Recognition Model



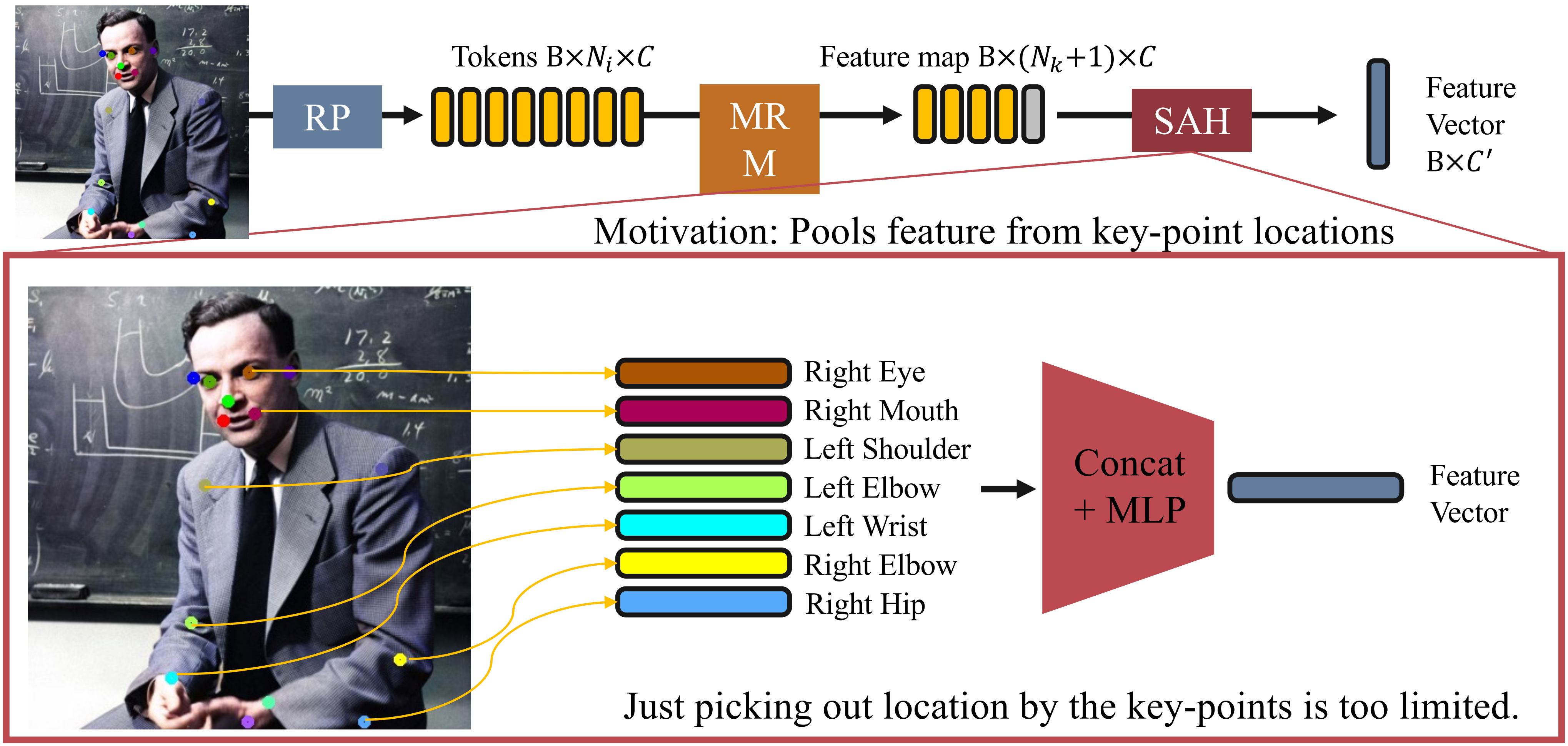
2. Masked Recognition Model



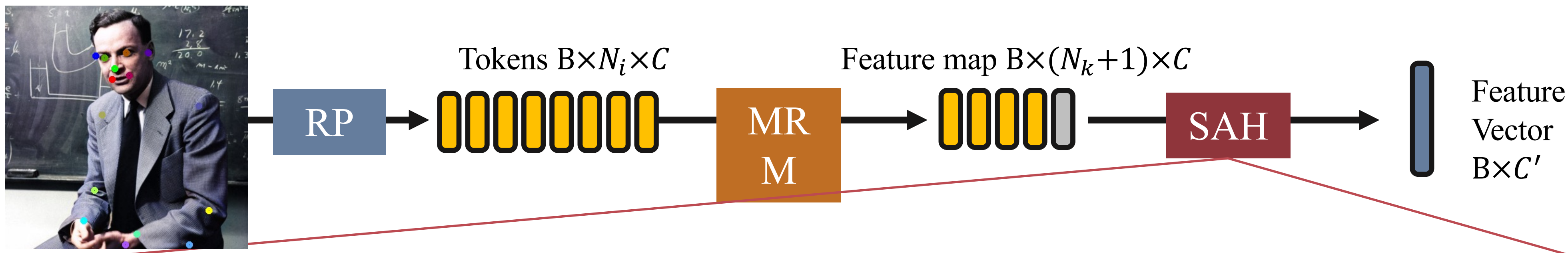
	All	Face	Whole Body ReID Short	Whole Body ReID Long
SoTA Face and Body Models	67.97	**97.63**	61.49	44.90
(1) ViT	59.54	90.63	56.17	31.81
(2) ViT+RP	66.35	92.93	59.16	46.95
(3) ViT+SAH	71.67	95.84	72.63	46.55
(4) ViT+RP+SAH (SapiensID)	**78.67**	**97.31**	**73.05**	**66.30**
(4) without Random Mask Ratio	74.39	95.95	69.58	57.64

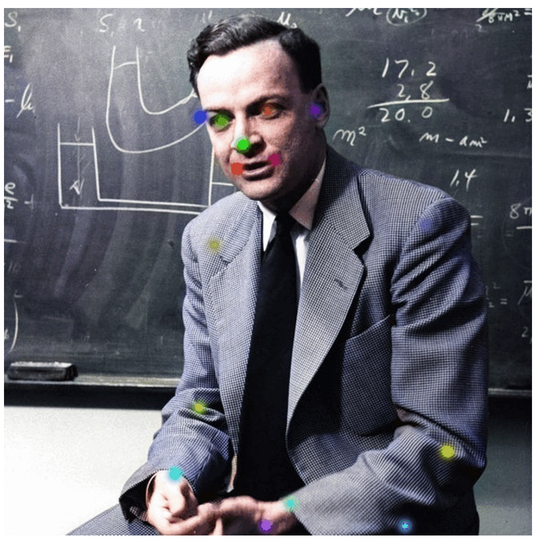
Performance Impact of not using Random Masking Ratio

3. Semantic Attention Pooling

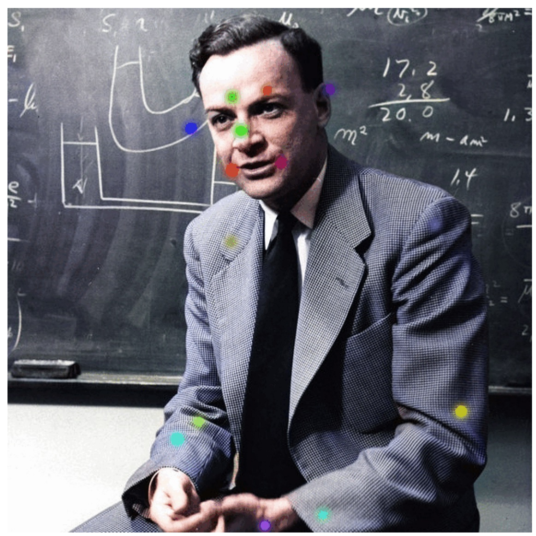


3. Semantic Attention Pooling





Learns Attention Size

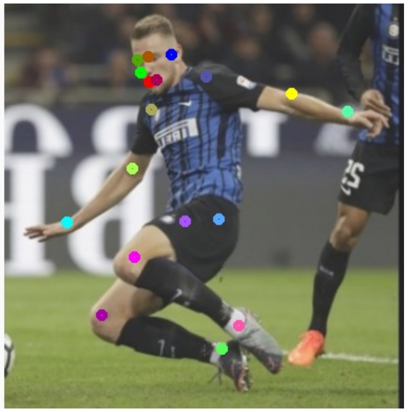


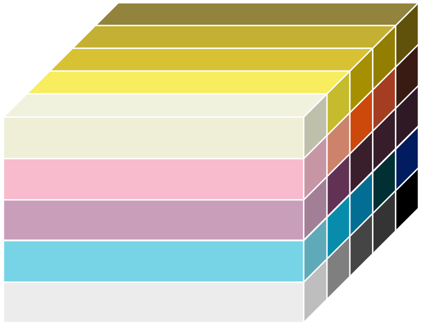
Learns Attention Offset Location

Attention for learning the appropriate size and offset locations from **keypoints.**

3. Semantic Attention Pooling

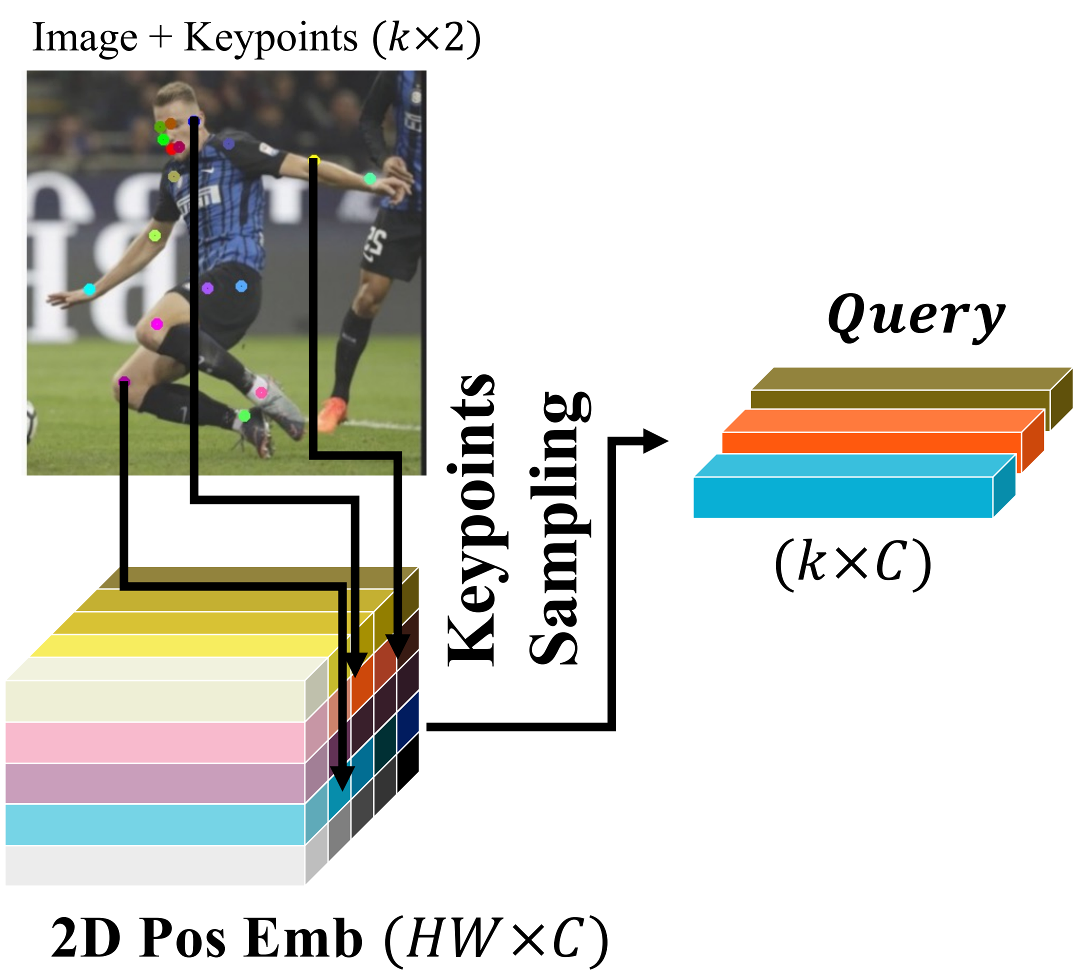
Image + Keypoints ($k \times 2$)



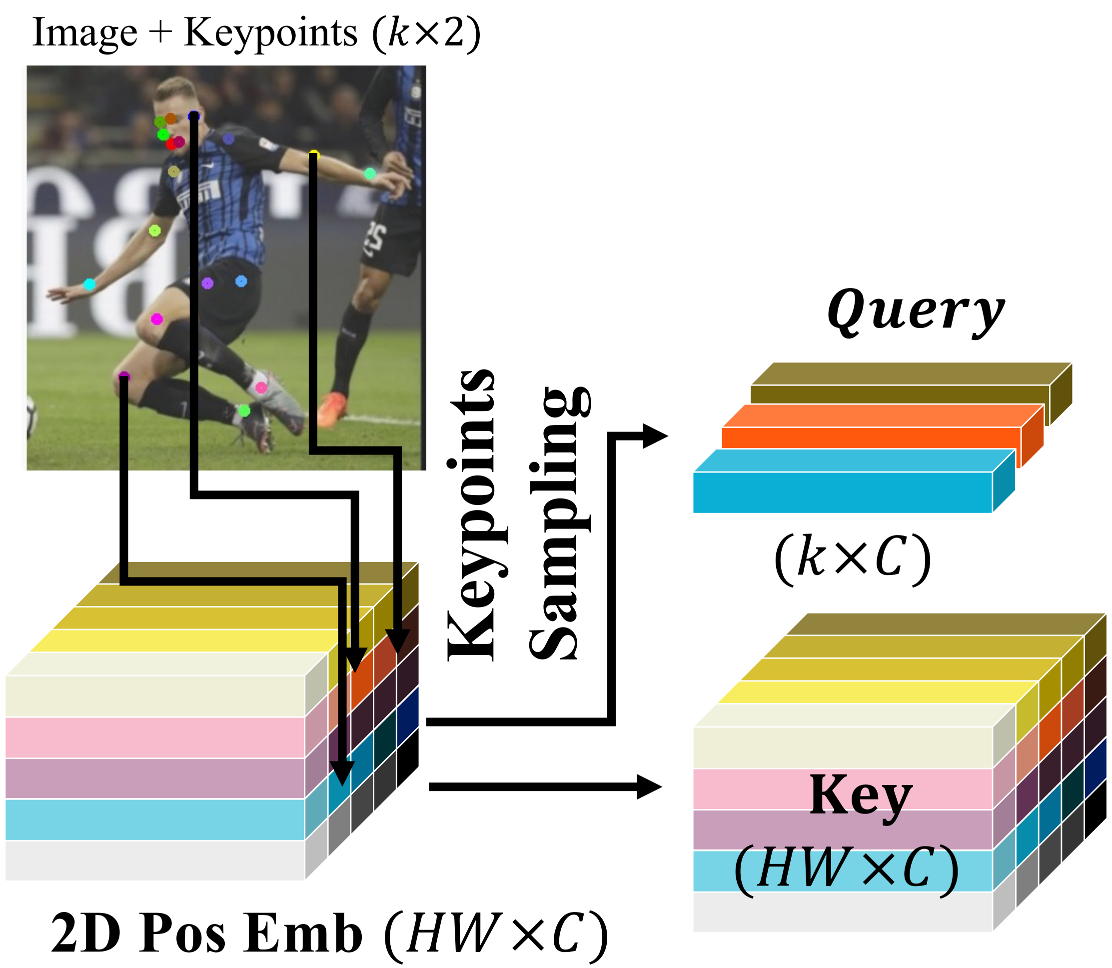


2D Pos Emb ($HW \times C$)

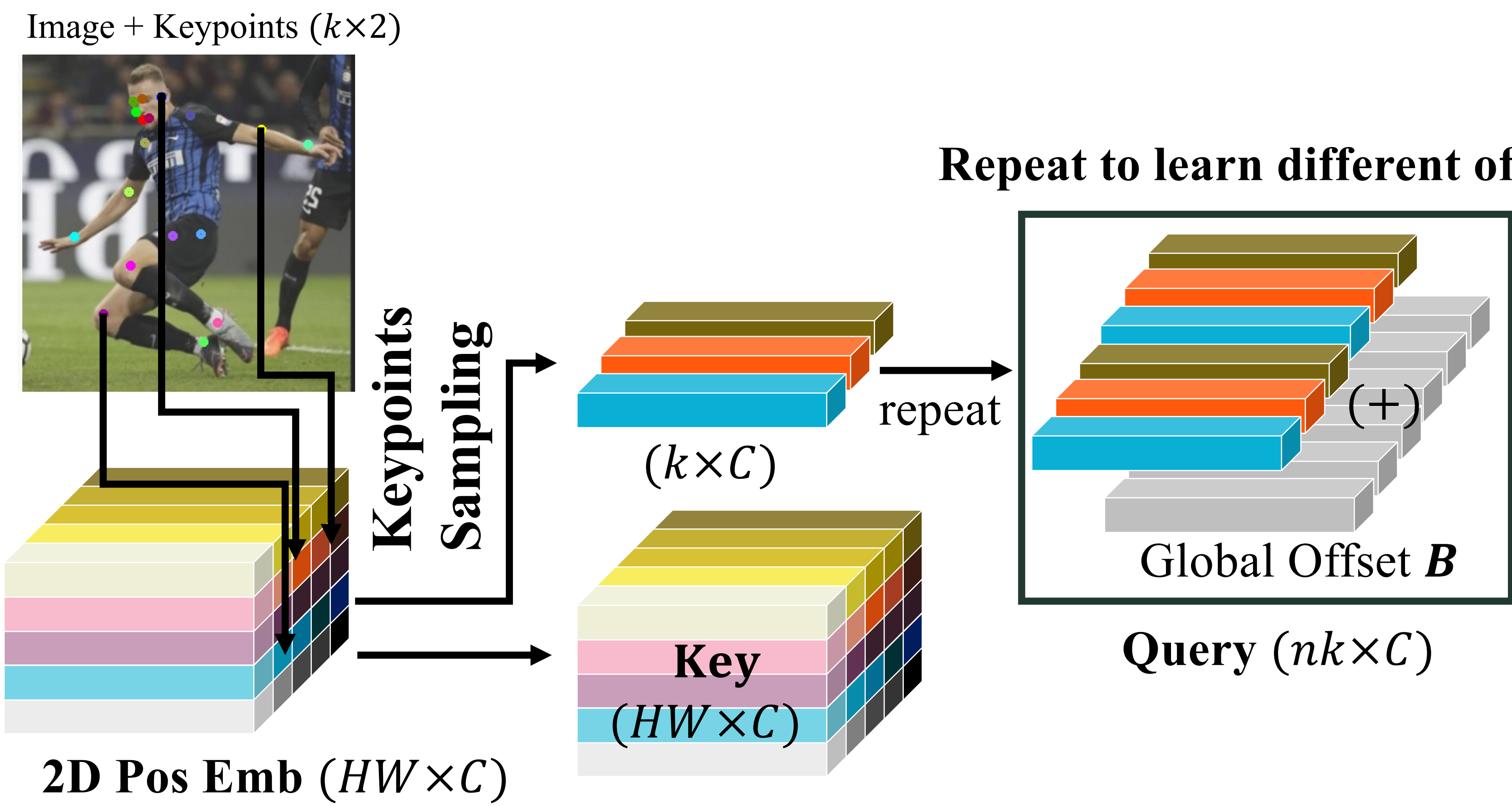
3. Semantic Attention Pooling



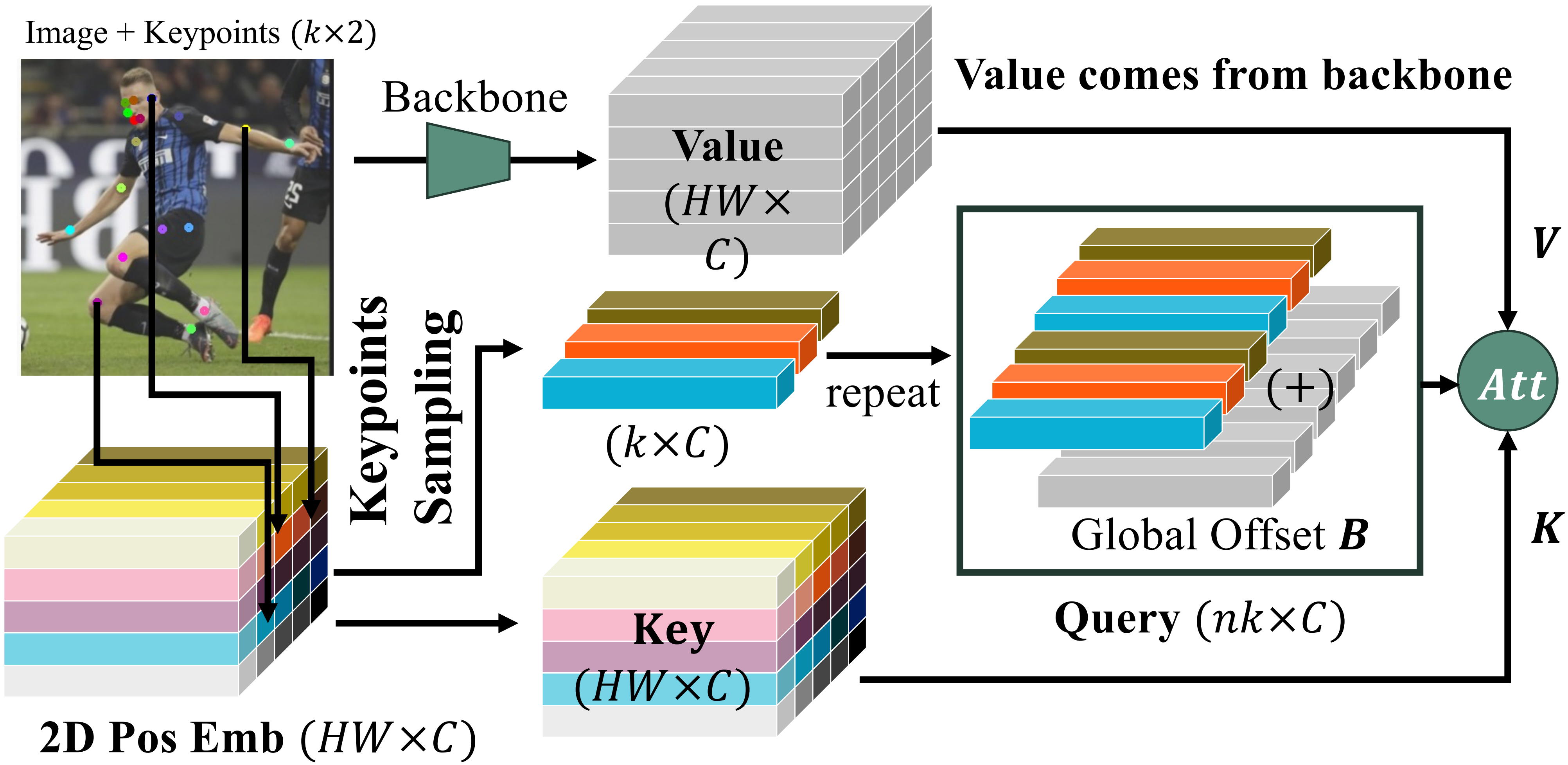
3. Semantic Attention Pooling



3. Semantic Attention Pooling



3. Semantic Attention Pooling



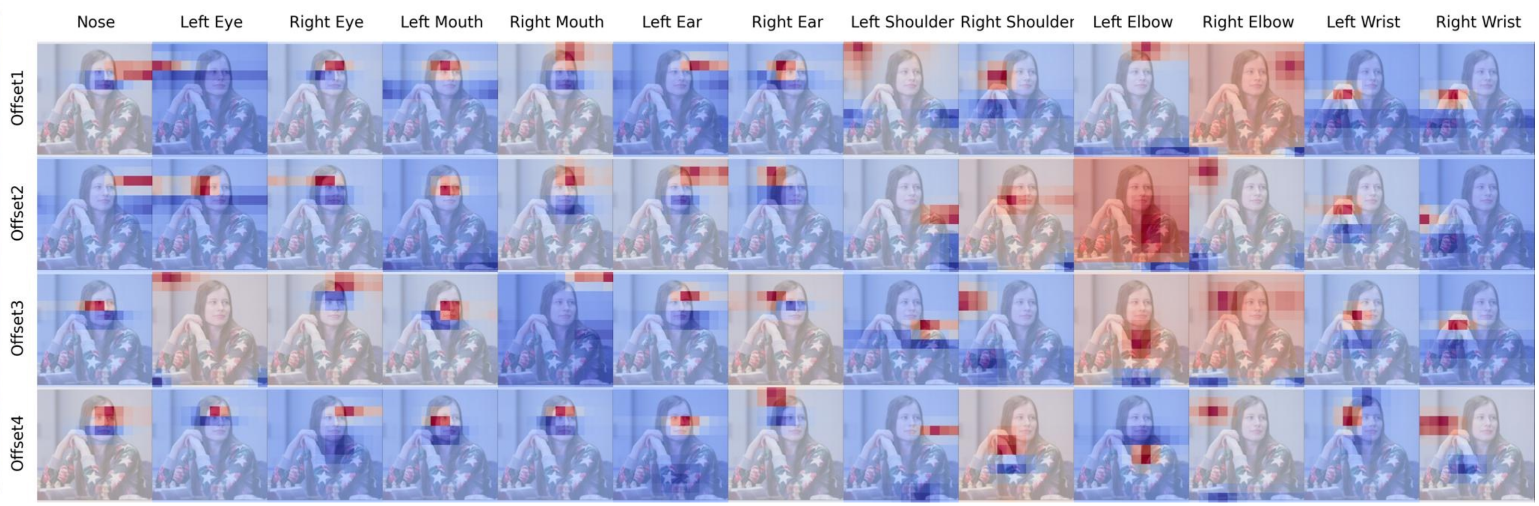
3. Semantic Attention Pooling

Query: $Q = \mathrm{GridSample}(PosEnc, keypoints) + B$

Key: $K = PosEnc$

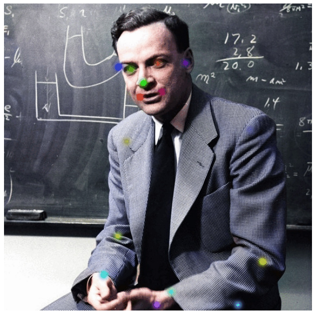
Visualizing

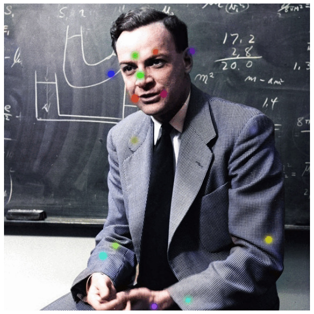
$$\mathrm{softmax}\left(\frac{\mathbf{W}_q\mathbf{Q}\mathbf{W}_k\mathbf{K}^\top}{\sqrt{d}}\right)$$



Actual Learned Attention's Visualization
It learns different scales and offsets as intended.

3. Semantic Attention Pooling

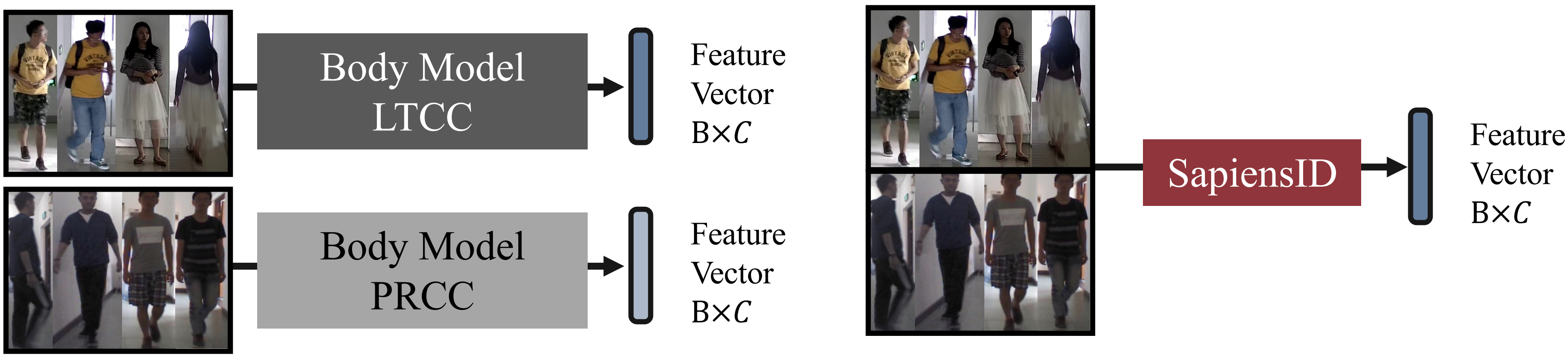




	All	Face	Whole Body ReID Short	Whole Body ReID Long
SoTA Face and Body Models	67.97	**97.63**	61.49	44.90
(1) ViT	59.54	90.63	56.17	31.81
(2) ViT+RP	66.35	92.93	59.16	46.95
(3) ViT+SAH	71.67	95.84	72.63	46.55
(4) ViT+RP+SAH (SapiensID)	**78.67**	**97.31**	**73.05**	**66.30**
(4) without Random Mask Ratio	74.39	95.95	69.58	57.64

Performance Impact of Using Semantic Attention Pooling

Performance



Method	Train Data	Avg	LTCC Top1	PRCC Top1	CCVID Top1	CCDA Top1	Celeb-ReID Top1
CAL [19]	LTCC	38.26	38.01	37.00	74.97	3.91	37.42
CAL [19]	PRCC	32.02	6.38	55.69	71.61	2.85	23.59
CAL [19]	LTCC+PRCC	38.65	33.16	45.39	73.89	3.74	37.11
CLIP3DReID [46]	LTCC	40.11	41.84	40.81	76.28	4.31	37.31
CLIP3DReID [46]	PRCC	33.06	6.63	62.40	69.32	3.17	23.82
HAP [74]	LU4M+LTCC	33.07	25.00	26.14	41.64	4.56	30.28
HAP [74]	LU4M+PRCC	31.16	29.08	38.05	45.73	5.13	37.79
HAP [74]	WebBody4M (Ours)	52.11	22.70	54.93	88.34	28.80	65.78
SapiensID (Ours)	WebBody4M (Ours)	**72.89**	**42.35**	**78.75**	**88.72**	**61.84**	**92.80**

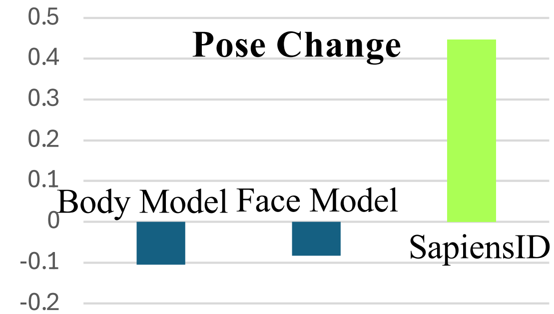
Performance

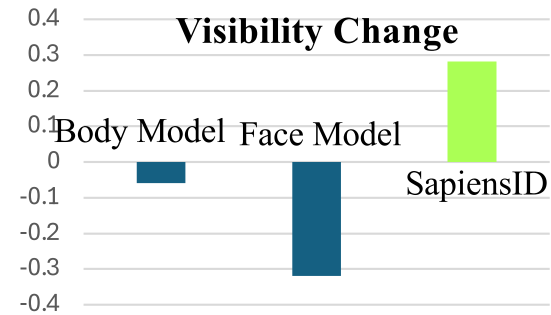
Method	AdaFace-ViT [32]	SapiensID (Ours)
Train Data	WebBody4M-FaceCrop	WebBody4M
LFW [24]	99.82	99.82
CPLFW [79]	95.12	94.85
CFPFP [54]	99.19	98.74
CALFW [80]	96.07	95.78
AGEDB [52]	97.97	97.33
Face Avg	**97.63**	97.31
LTCC [55]	21.70	72.01
Market1501 [77]	7.81	88.18
Body Avg	14.76	**80.10**
Combined Avg	56.19	**89.80**

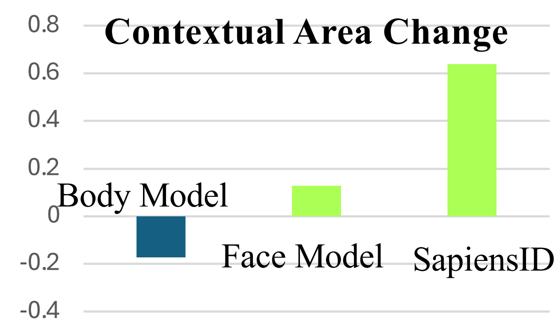
Analysis

Probe

Galler
y







Y-axis: Similarity – Threshold

Predict match

Predict no match

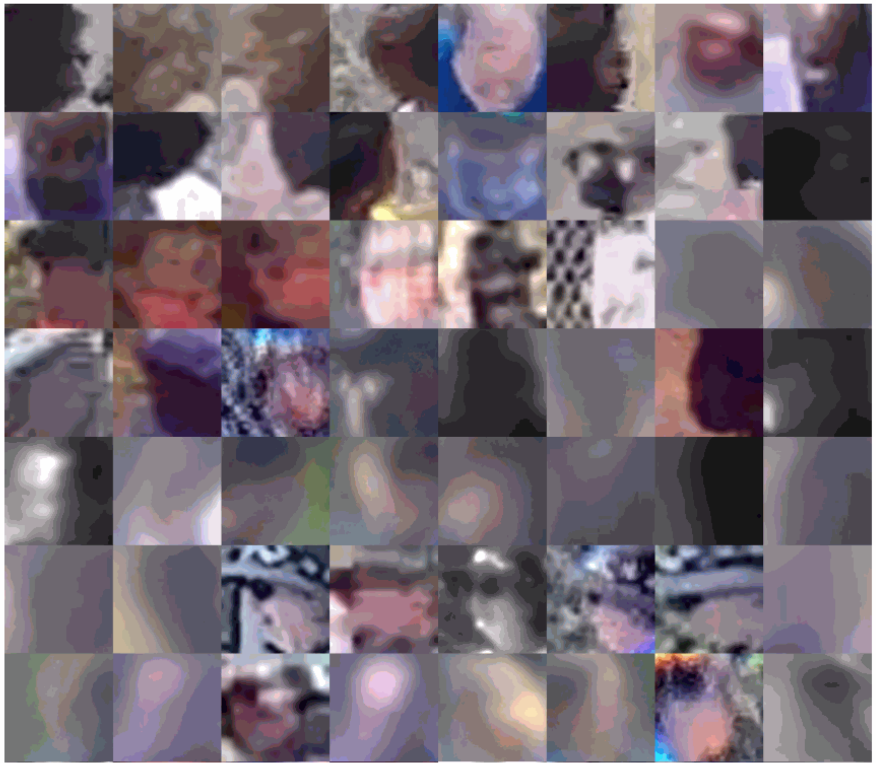
Success and Failure Cases

Case Analysis

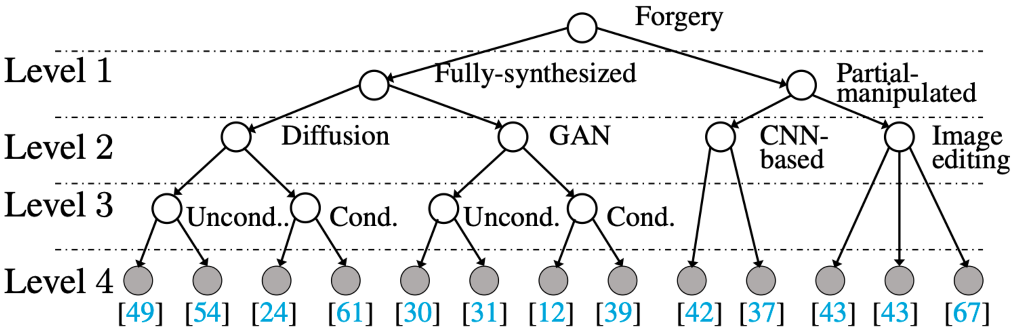
[08/5/24] FR2.2 (Rank20: 82.64)

wrong $\longmapsto$ correct

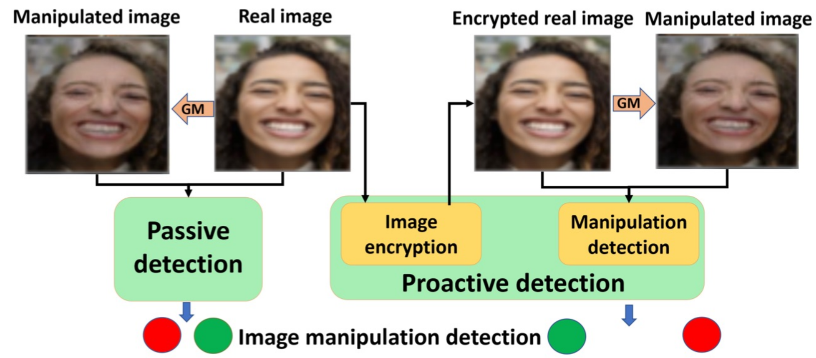
wrong $\longmapsto$ wrong



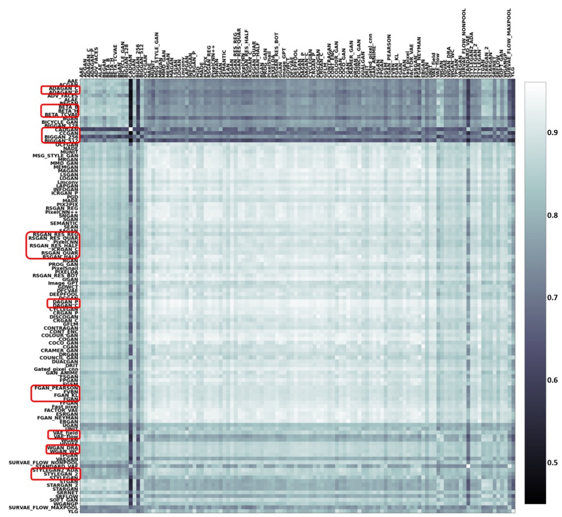
Trustworthy Biometrics



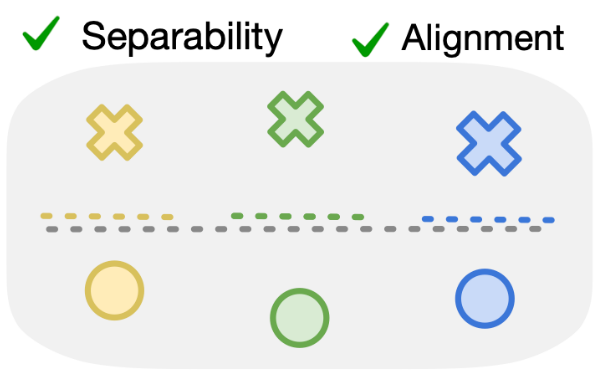
Deepfake detection, CVPR'23



Proactive CV, CVPR'22,23,24, NeurIPS'23



Model parsing, PAMI'23



Anti-Spoofing, CVPR'23 & earlier

Future Directions

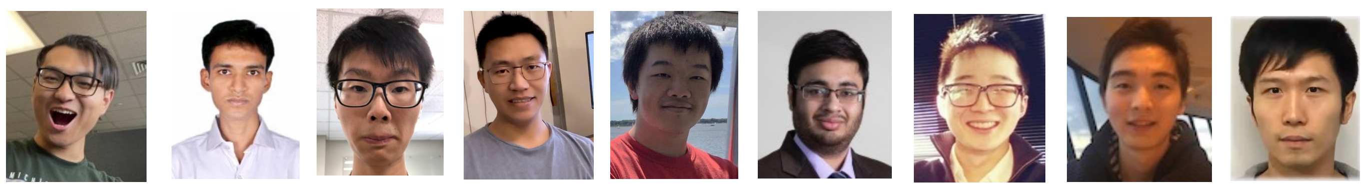
- Move from close-set to open-set
- Fusion of face, body, and gait
- Advance AIGC to push “gap to real” to zero
- Explainable recognition systems
- Build foundation models for biometrics

Conclusions

- There are many new research opportunities in person identification.
- Pre-trained foundation models could be enhanced for biometrics.
- Building a unified model for periocular/face/body/gait leads to a foundation model for biometrics.

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

MICHIGAN STATE UNIVERSITY