华中科技大学

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

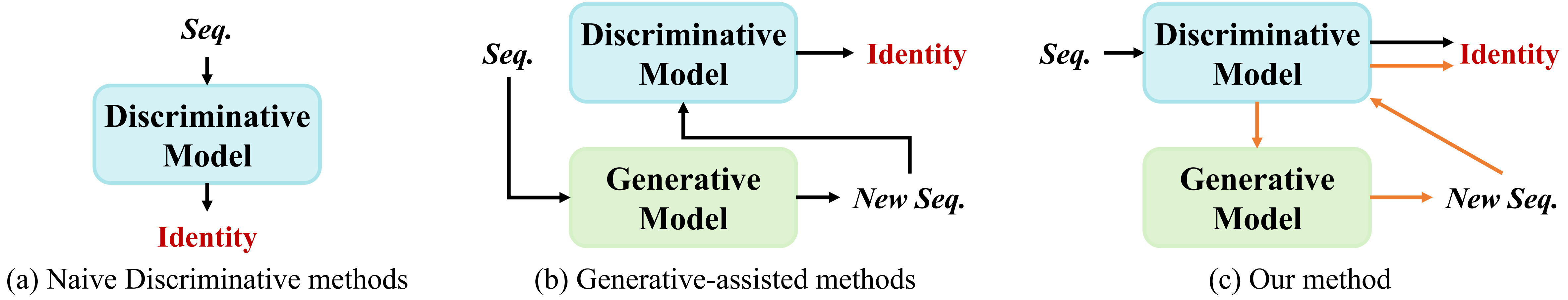
Gait Recognition via Collaborating Discriminative and Generative Diffusion Models

Talker: Haijun Xiong (熊海军), second-year Ph.D. student

Email: xionghj@hust.edu.cn

Supervisor: Prof. Bin Feng (冯镔)

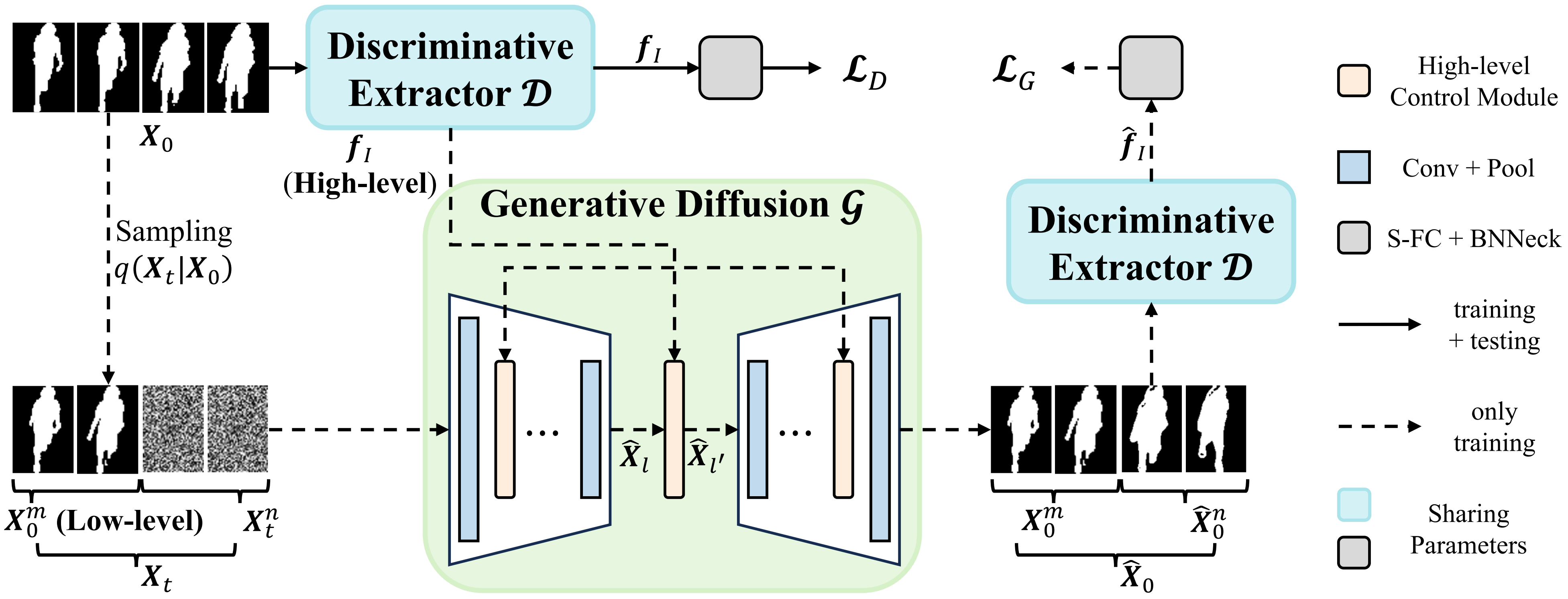
Motivation: Architecture Limitations



- (a) Learns discriminative features but **fails to model the underlying data distribution**.
- (b) Treats generation as a **detached processing module**, which may **not be optimized for the end task**.
- (c) Ours: A **collaborative discriminative–generative framework, jointly and end-to-end optimized** for gait recognition.

From isolated modules to end-to-end collaboration.

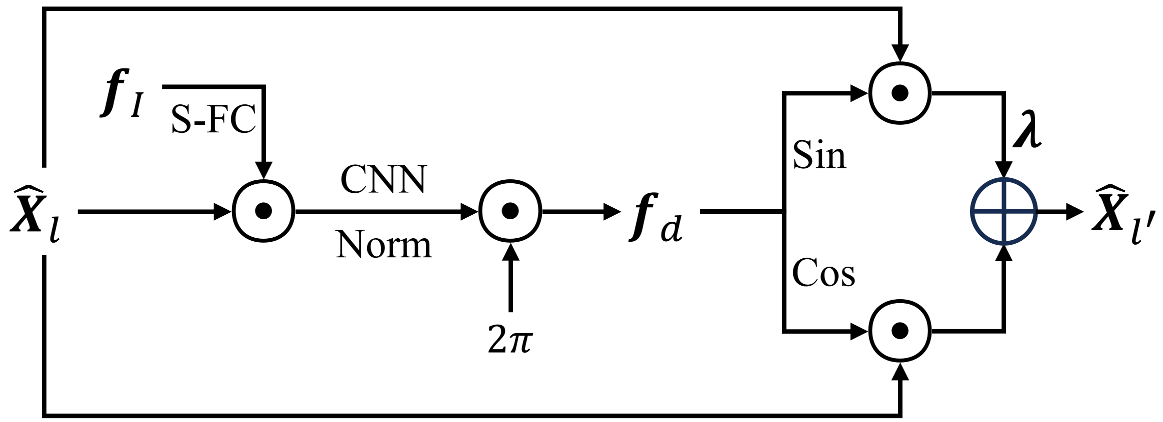
Method: CoD^2



? How to move beyond passive discrimination in gait recognition?

- Discriminative Path: Extract identity representations from gait sequences.
- Generative Path: Model the underlying gait distribution via diffusion-based recovery.
- Collaborative Learning: Enable generation to actively refine discriminative inference.

High-level Control Module



$$f_d = 2\pi \cdot \mathrm{Norm}\big(\mathrm{Conv}\big(\widehat{X}_l \cdot \mathrm{S-Fc}(f_I)\big)\big)$$

$$\widehat{X}_{l'} = \widehat{X}_l \cdot \cos(f_d) + \lambda \cdot \widehat{X}_l \cdot \sin(f_d)$$

Why High-level Control?

- Coarse condition fusion (e.g., element-wise addition)
- Lacks fine-grained control
- Degrades identity consistency

What Does It Enable?

- Identity-aware generation guidance
- Spatially adaptive semantic control
- Smooth and stable diffusion modulation

High-level identity semantics are injected as phase modulation, enabling fine-grained and stable identity control in diffusion-based gait generation.

Experiments

Modality	Method	Venue	Probe Sequence (R-1)								Overall	
			NM	BG	CL	CR	UB	UN	OC	NT	R-1	R-5
Silhouette	GaitSet	AAAI19	69.1	68.2	37.4	65.0	63.1	61.0	67.2	23.0	65.0	84.8
	GaitPart	CVPR19	62.2	62.8	33.1	59.5	57.2	54.8	57.2	21.7	59.2	80.8
	GaitGL	ICCV21	67.1	66.2	35.9	63.3	61.6	58.1	66.6	17.9	63.1	82.8
	GaitBase	CVPR23	81.5	77.5	49.6	75.8	75.5	76.7	81.4	25.9	76.1	89.4
	DeepGaitV2	TPAMI25	83.5	79.5	46.3	76.8	79.1	78.5	81.1	27.3	77.4	90.2
Silhouette + Skeleton	BiFusion	MTAP24	69.8	62.3	45.4	60.9	54.3	63.5	77.8	33.7	62.1	83.4
	SkeletonGait++	AAAI24	85.1	82.9	46.6	81.9	80.8	82.5	86.2	**47.5**	81.3	95.5
Silhouette	**Ours**	-	**87.9**	**84.5**	**55.4**	**82.8**	**87.2**	**85.1**	**88.7**	38.6	**83.8**	**95.8**

Performance comparisons on SUSTech1K.

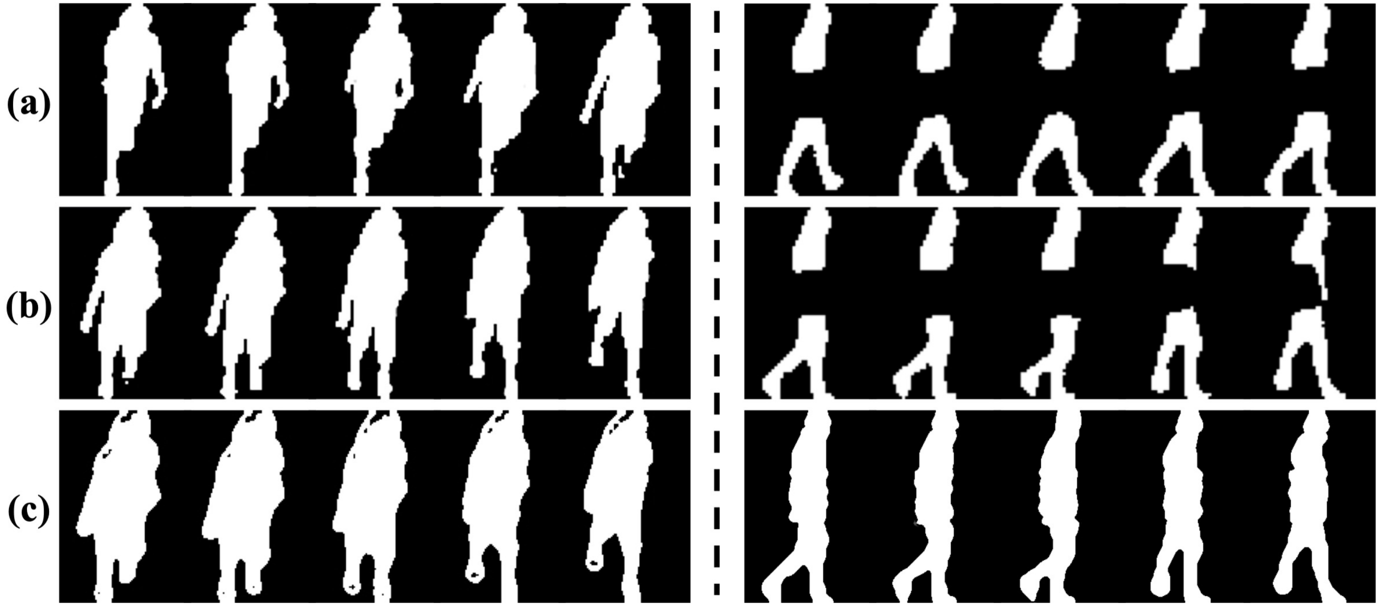
Method	Venue	Gait Evaluation Protocol				
		CL	UP	DN	BG	Mean
GaitSet	AAAI19	60.2	65.2	65.1	68.5	64.8
GaitPart	CVPR20	64.3	67.8	68.6	71.7	68.1
GaitBase	CVPR23	71.6	75.0	76.8	78.6	75.5
DeepGaitV2	TPAMI25	78.6	84.8	80.7	89.2	83.3
Ours	-	**80.1**	**86.9**	**81.6**	**90.9**	**84.8**

Performance comparisons on CCPG.

Method	Venue	GREW		Gait3D	
		Rank-1	Rank-5	Rank-1	mAP
GaitSet	AAAI19	46.3	63.6	36.7	30.0
GaitPart	CVPR19	44.0	60.7	28.2	21.6
GaitGL	ICCV21	47.3	63.6	29.7	22.3
SMPLGait	CVPR22	-	-	46.3	37.2
DANet	CVPR23	-	-	48.0	-
GaitBase	CVPR23	60.1	-	64.6	-
GaitGCI	CVPR23	68.5	80.8	50.3	39.5
HSTL	ICCV23	62.7	76.6	61.3	55.5
DyGait	ICCV23	71.4	83.2	66.3	56.4
QAGait	AAAI24	59.1	74.0	67.0	56.5
VPNet	CVPR24	80.0	89.4	75.4	-
CLTD	ECCV24	78.0	87.8	69.7	-
WaveLoss	AAAI25	-	-	75.6	66.5
DeepGaitV2	TPAMI25	77.7	87.9	74.4	65.8
Ours	-	**81.2**	**90.8**	**78.3**	**71.2**

Performance comparisons on GREW and Gait3D.

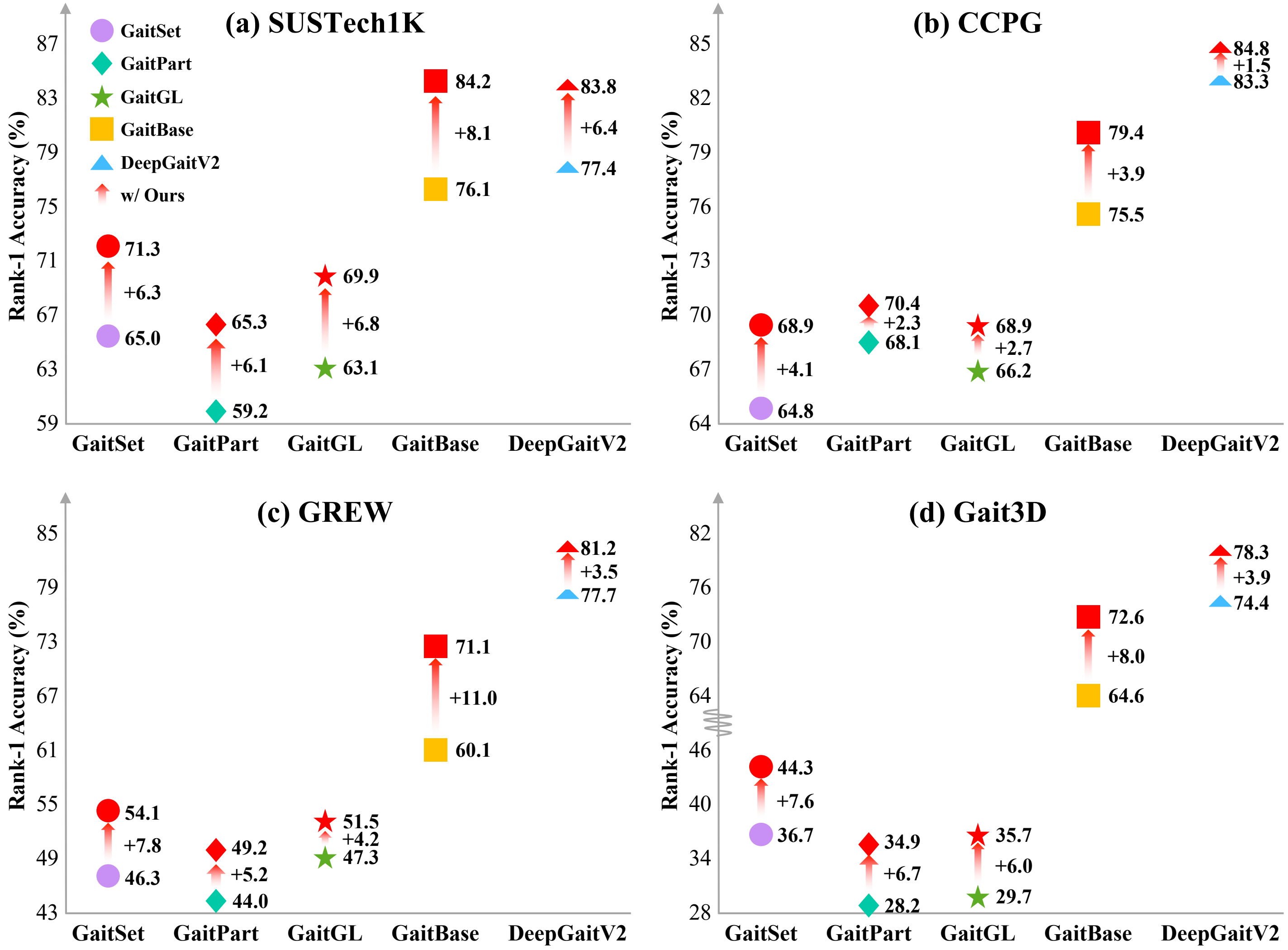
Visualization



(a) Reference sequence (b) Generate sequence GT (c) Model predicts sequence

The model learns to not only copy gait data, but to amplify its most identifiable features for recognition.

Ablation Study



High-level	Low-level	SUSTech1K	CCPG	GREW	Gait3D
✗	✗	77.4	83.3	77.7	74.4
✓	✗	81.9	84.0	80.4	77.4
✗	✓	81.3	83.7	79.9	77.2
✓	✓	**83.8**	**84.8**	**81.2**	**78.3**

Ablation study of Multi-level Conditional Control strategy.

Method	SUSTech1K	CCPG	GREW	Gait3D
Baseline	81.3	83.7	79.9	77.2
w/ addition	83.0	84.4	80.7	77.6
w/ Ours	**83.8**	**84.8**	**81.2**	**78.3**

The ablation study of High-level Control Module.

λ	SUSTech1K	CCPG	GREW	Gait3D
1	83.2	84.3	80.3	77.4
learnable scalar	83.5	84.4	80.8	77.8
learnable vector	**83.8**	**84.8**	**81.2**	**78.3**

The ablation study of the learnable vector λ.

Thanks for Attention!

Talker: Haijun Xiong (熊海军), second-year Ph.D. student

Email: xionghj@hust.edu.cn

Supervisor: Prof. Bin Feng (冯镔)