中国科学院深圳先进技术研究院
SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

Multi-Level Graph Encoding with Structural-Collaborative Relation Learning for Skeleton-Based Person Re-Identification

Haocong Rao[1,2] Shihao Xu[1,2,3] Xiping Hu[1,2,3] Jun Cheng[1,2] Bin Hu[4,3]

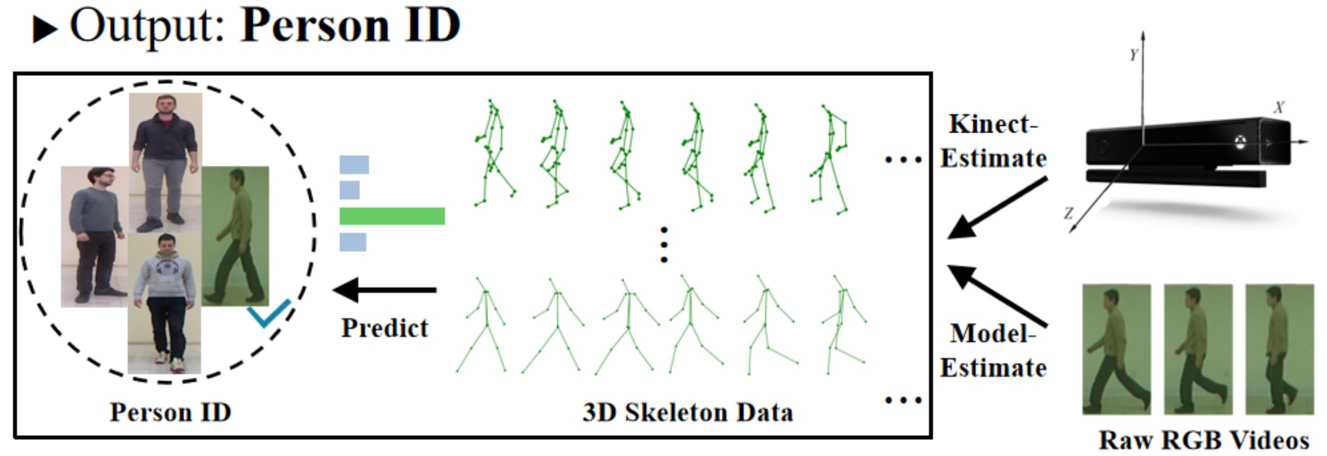
[1] SIAT, CAS [2] CUHK, HK [3] LZU [4] BIT

Paper, Code and Data:
https://github.com/Kali-Hac/MG-SCR

Goal

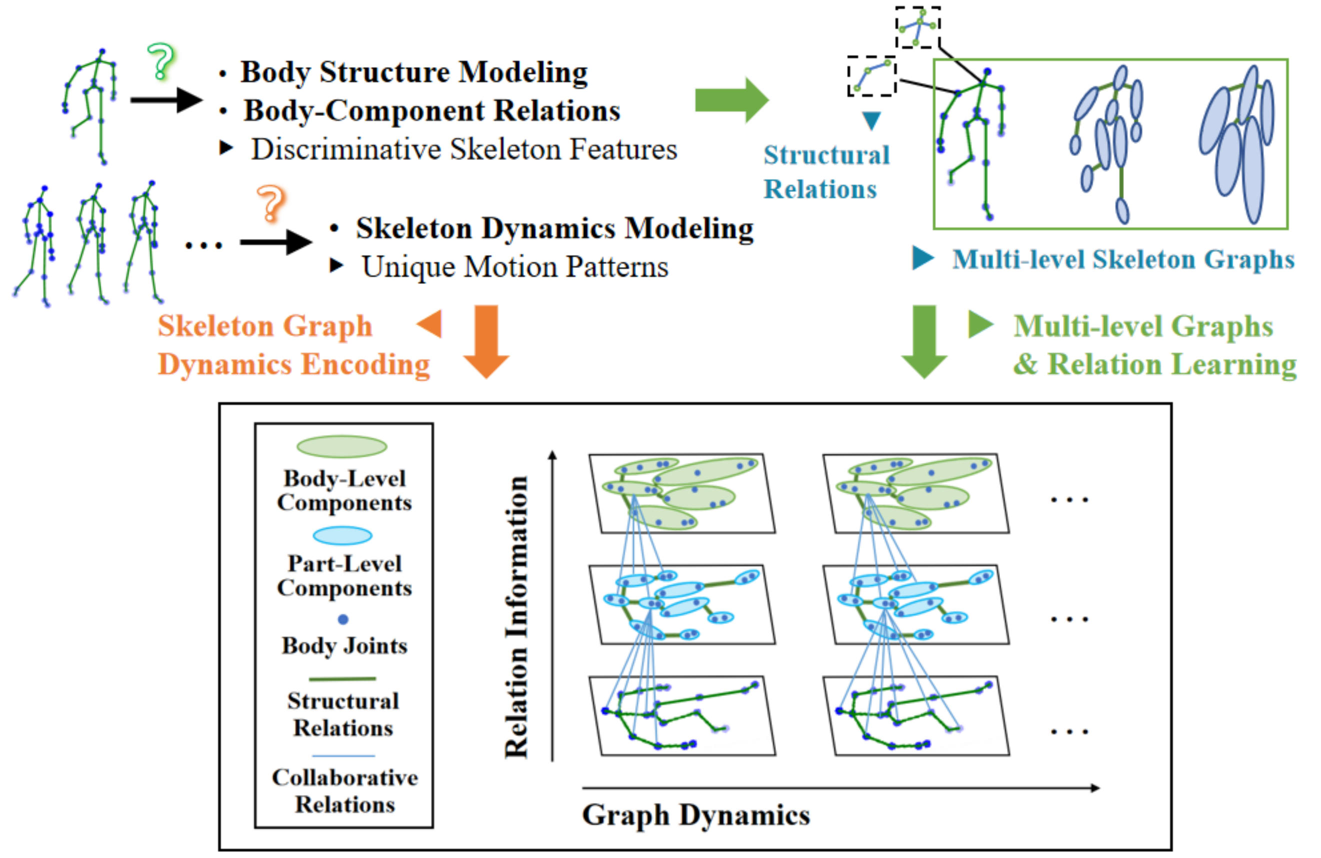
Skeleton-based person re-identification (Re-ID) aims to re-identify a specific person via 3D skeletons in different views or scenes.

- Input: **3D Skeletons** (3D coordinates of body joints)
- ▶ Output: **Person ID**



3D Skeleton-based Re-ID Task

Motivation

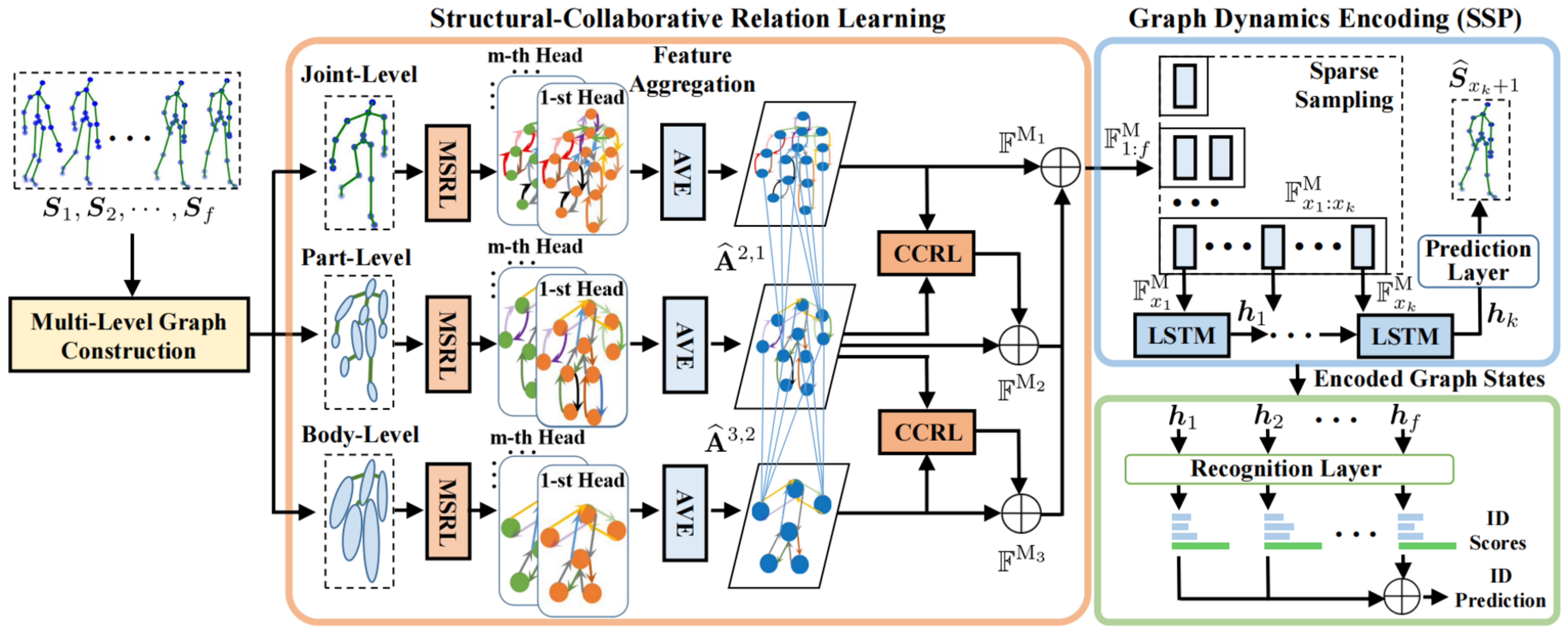


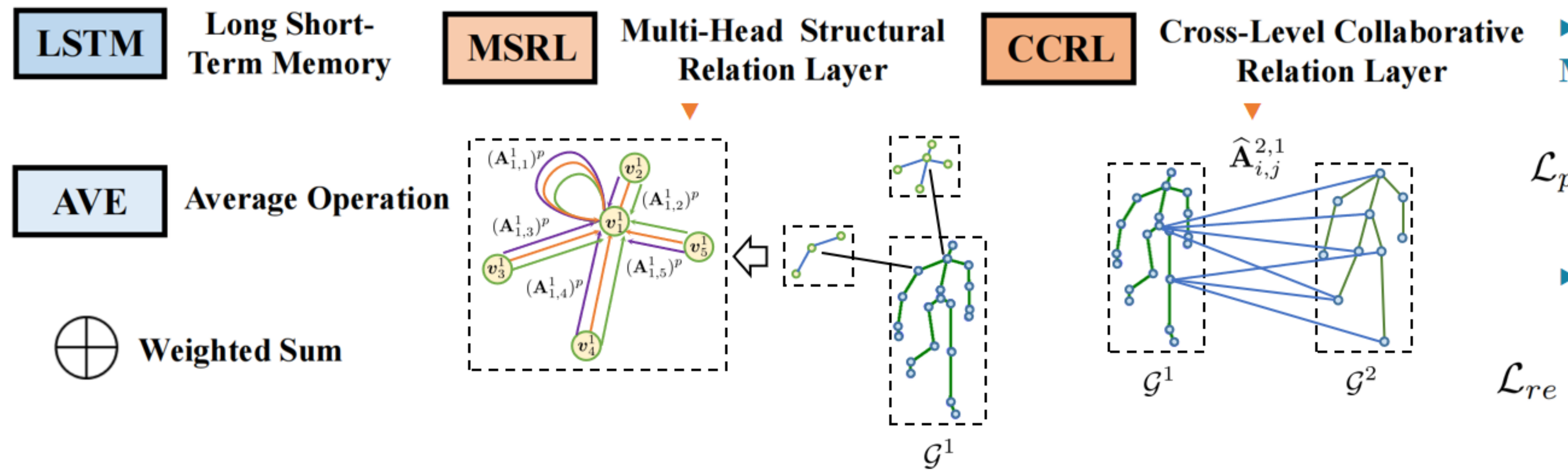
Overview of Multi-level Graph Encoding Model

Contributions

Multi-level Graphs ▸ Fully model body structure
MSRL ▸ Learn structural correlations of adjacent joints
CCRL ▸ Learn collaboration between body components
SSP ▸ Encode graph dynamics and capture high-level semantics with a self-supervised prediction task

Method





▸ Sparse Sequential Prediction (SSP) - MSE Loss between Prediction and Truth:

$$\mathcal{L}_{pred} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{f-1}\|\boldsymbol{S}_{l_k+1}^{(i)} - \widehat{\boldsymbol{S}}_{l_k+1}^{(i)}\|_2^2$$

▸ Person Re-ID - Cross-Entropy Loss:

$$\mathcal{L}_{re} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{i,j}\log\hat{y}_{i,j} + \beta\|\Theta\|_2^2$$

Experiments on Skeleton Re-ID Benchmarks

▸ **Outperform state-of-the-art skeleton-based methods**

▸ **Superior to many depth-based and multi-modal methods**

			BIWI		IAS-A		IAS-B		KGBD		KS20	
	Id	Methods	Rank-1	nAUC	Rank-1	nAUC	Rank-1	nAUC	Rank-1	nAUC	Rank-1	nAUC
Depth-Based Methods	1	Gait Energy Image [2010]	21.4	73.2	25.6	72.1	15.9	66.0	—	—	—	—
	2	3D CNN + Average Pooling [2010]	27.8	84.0	33.4	81.4	39.1	82.8	—	—	—	—
	3	Gait Energy Volume [2011]	25.7	83.2	20.4	66.2	13.7	64.8	—	—	—	—
	4	3D LSTM [2016]	27.0	83.3	31.0	77.6	33.8	78.0	—	—	—	—
Multi-Modal Methods	5	PCM + Skeleton [2014a]	42.9	—	27.3	—	81.8	—	—	—	—	—
	6	Size-Shape descriptors + SVM [2016]	20.5	87.2	—	—	—	—	—	—	—	—
	7	Size-Shape descriptors + LDA [2016]	22.1	88.5	—	—	—	—	—	—	—	—
	8	DVCov + SKL [2017]	21.4	—	46.6	—	45.9	—	—	—	—	—
	9	ED + SKL [2017]	30.0	—	52.3	—	63.3	—	—	—	—	—
	10	CNN-LSTM with RTA [2018]	50.0	—	—	—	—	—	—	—	—	—
Skeleton-Based Methods	11	D^{13} descriptors + KNN [2014b]	39.3	64.3	33.8	63.6	40.5	71.1	46.9	90.0	58.3	78.0
	12	Single-layer LSTM [2016]	15.8	65.8	20.0	65.9	19.1	68.4	39.8	87.2	80.9	92.3
	13	Multi-layer LSTM [2019]	36.1	75.6	34.4	72.1	30.9	71.9	46.2	89.8	81.6	94.2
	14	D^{16} descriptors + Adaboost [2019]	41.8	74.1	27.4	65.5	39.2	78.2	69.9	90.6	59.8	78.8
	15	PostGait [2020]	33.3	81.8	41.4	79.9	37.1	74.8	90.6	97.8	70.5	94.0
	16	Attention Gait Encodings [2020]	59.1	86.5	56.1	81.7	58.2	85.3	87.7	96.3	86.5	94.7
	17	**MG-SCR (Ours)**	**61.6**	**91.9**	**56.5**	**87.0**	**65.9**	**93.1**	**96.3**	**99.9**	**87.3**	**95.5**

Evaluation on Model-Estimated Skeletons (from CASIA B)

▸ **Outperform state-of-the-art supervised method under cross-view evaluation**

Methods	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
PoseGait	10.7	37.4	52.5	28.3	24.3	18.9	23.5	17.2	23.6	18.8	4.3
Ours	**20.0**	**63.1**	**60.3**	**52.0**	**54.0**	**80.4**	**75.1**	**74.3**	**65.6**	**39.1**	**25.5**

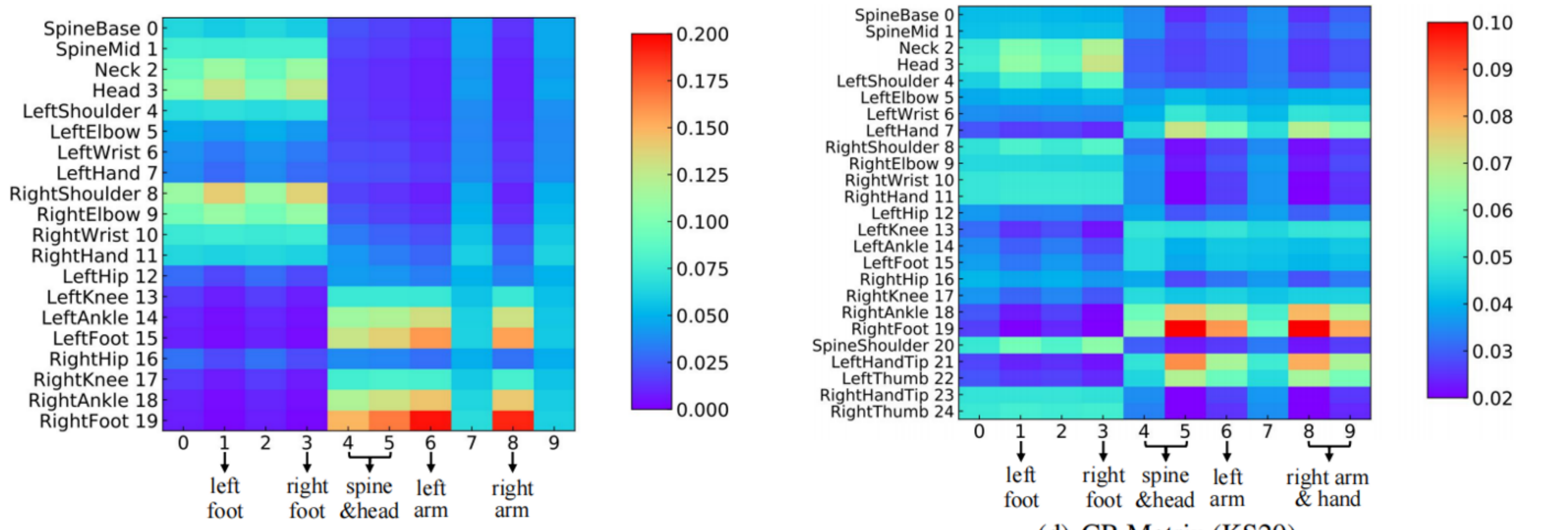
Component Analysis and Relation Visualization

MSRL		CCRL	SSP	h_f	AP	Rank-1	nAUC
Single-Level	Multi-Level						
✓					✓	56.8	89.1
	✓				✓	57.3	89.9
✓			✓	✓		56.9	89.6
	✓		✓	✓		57.6	90.2
✓			✓		✓	57.2	89.2
	✓		✓		✓	59.3	91.0
	✓	✓		✓		58.4	89.4
	✓	✓			✓	59.1	90.6
	✓	✓	✓	✓		59.7	91.0
	✓	✓	✓		✓	**61.6**	**91.9**

Table 2: Performance of our model with different components (MSRL, CCRL, SSP). "Single-Level" denotes using only joint-level graph. "AP" indicates exploiting average prediction of encoded graph states $\boldsymbol{h}_1, \cdots, \boldsymbol{h}_f$ rather than final state $\boldsymbol{h}_f$ for person Re-ID.

- ▸ **Multi-level graphs** improve Re-ID performance compared with joint-level graphs
- ▸ **Combining MSRL and CCRL** can produce higher performance gain
- ▸ **SSP pre-training task** enhances graph dynamics encoding for better person Re-ID
- ▸ **Average prediction (AP)** encourages better sequence-level predictions

(a) CR Visualization (BIWI) (b) CR Visualization (KS20)



(c) CR Matrix (BIWI) (d) CR Matrix (KS20)

▸ **Our model can capture global/local collaborative relations between body components.**

Related Works

▸ **3D Skeleton-Based Person Re-ID:**
[1] [T-PAMI 2021] A Self-Supervised Gait Encoding Approach with Locality-Awareness for 3D Skeleton Based Person Re-Identification
[2] [IJCAI 2020] Self-Supervised Gait Encoding with Locality-Aware Attention for Person Re-Identification
▸ **Multi-Scale Skeleton Graphs:**
[3] SM-SGE: A Self-Supervised Multi-Scale Skeleton Graph Encoding Framework for Person Re-Identification