

SIAM: Synchronous Interaction Attention for Human Mesh Recovery

Niaz Ahmad^{*}, Saif Ullah⁺, Youngmoon Lee⁺, Guanghui (Richard) Wang^{*}

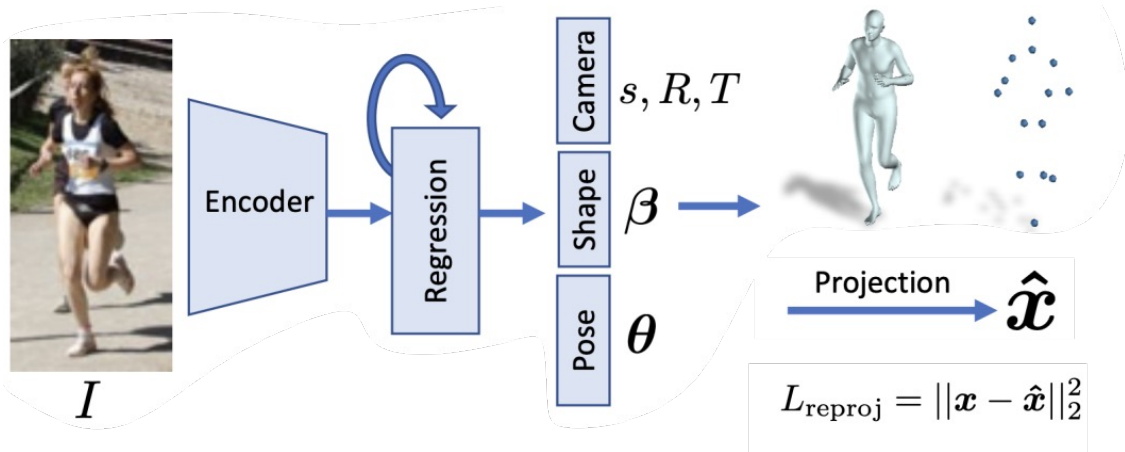
^{*}**CVIS Lab Toronto Metropolitan University**

⁺Hanyang University South Korea

Motivation

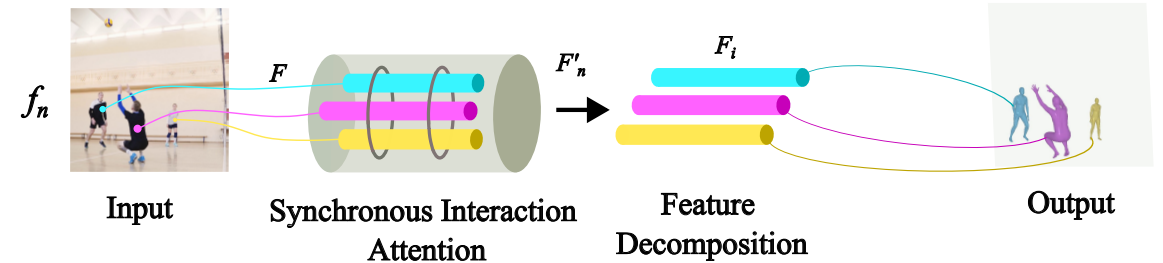
Existing methods

Conventional 3D body mesh reconstruction methods often use decoupling strategies that isolate individual features for separate representation, lacking relational cues among entities.



Our motivation

- (1) Explicit modeling of spatial relational cues among multiple human instances in live scenes.
- (2) Extracts enriched instance-specific features from the share attributes captured in the scene.



Proposed Method

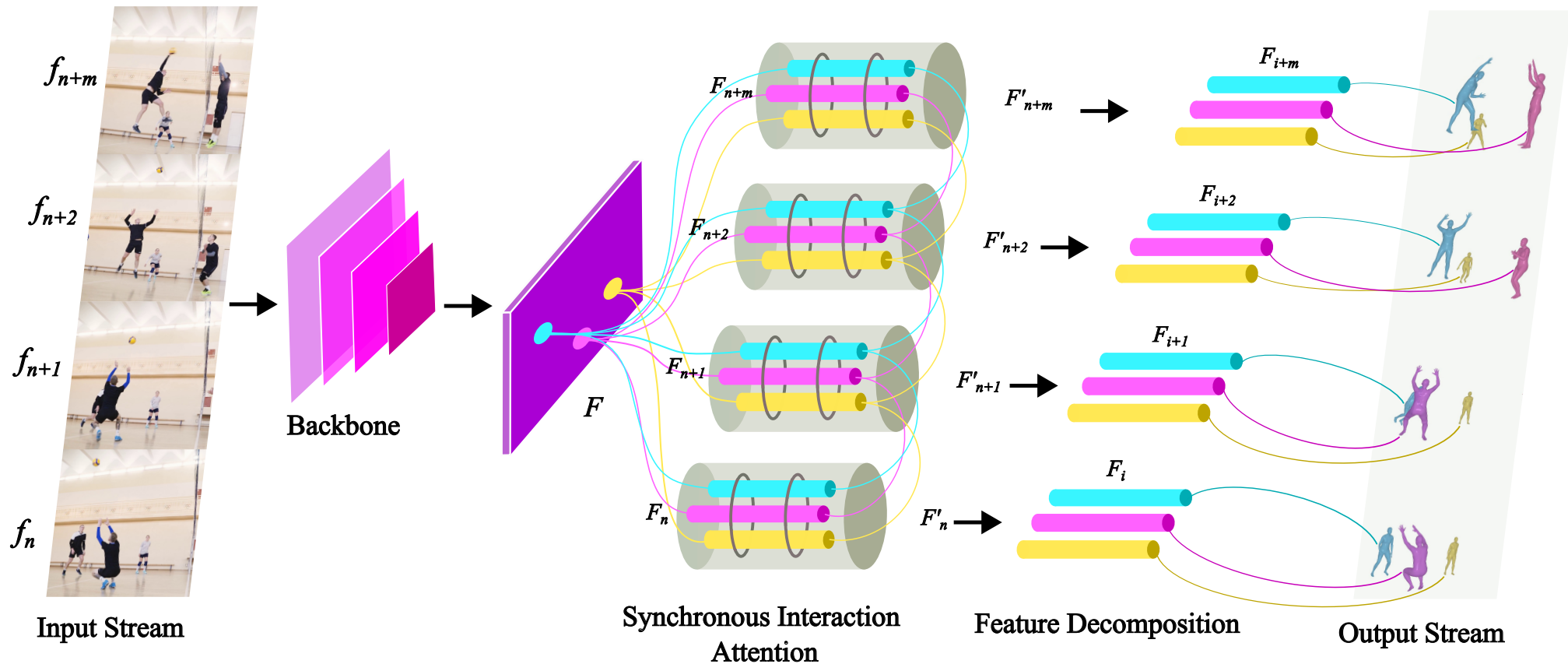


Figure 1. Overview of the proposed SIAM framework. Given a stream of video frames $\{f_n, f_{n+1}, \dots, f_{n+m}\}$, spatially rich multi-person features F are extracted using a high-resolution backbone. The Synchronous Interaction Attention (SIA) module captures cross-instance relational cues across both spatial and temporal dimensions to produce interaction-aware representations $\{F'_n, F'_{n+1}, \dots, F'_{n+m}\}$. These are then passed through the Feature Decomposition (FD) module to isolate individual-specific representations, which are finally used to reconstruct accurate 3D human mesh predictions in the output stream.

Synchronous Interaction Attention

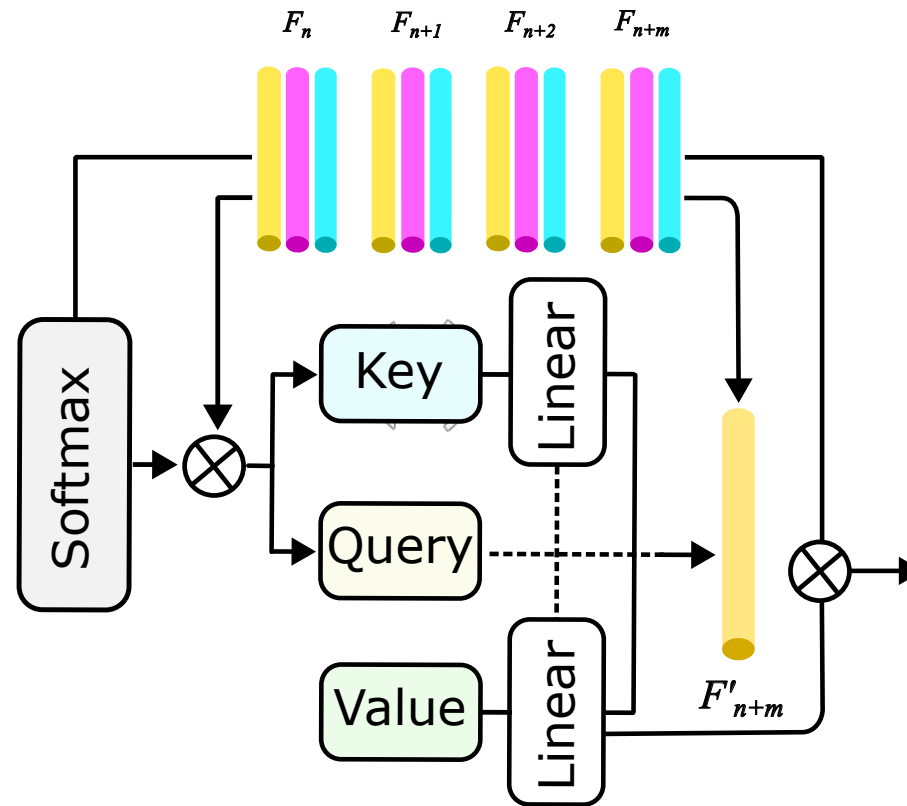


Figure 2: Architecture of the Attention Block. Instance-level features are transformed into query, key, and value representations through linear projections.

Real-time Capabilities

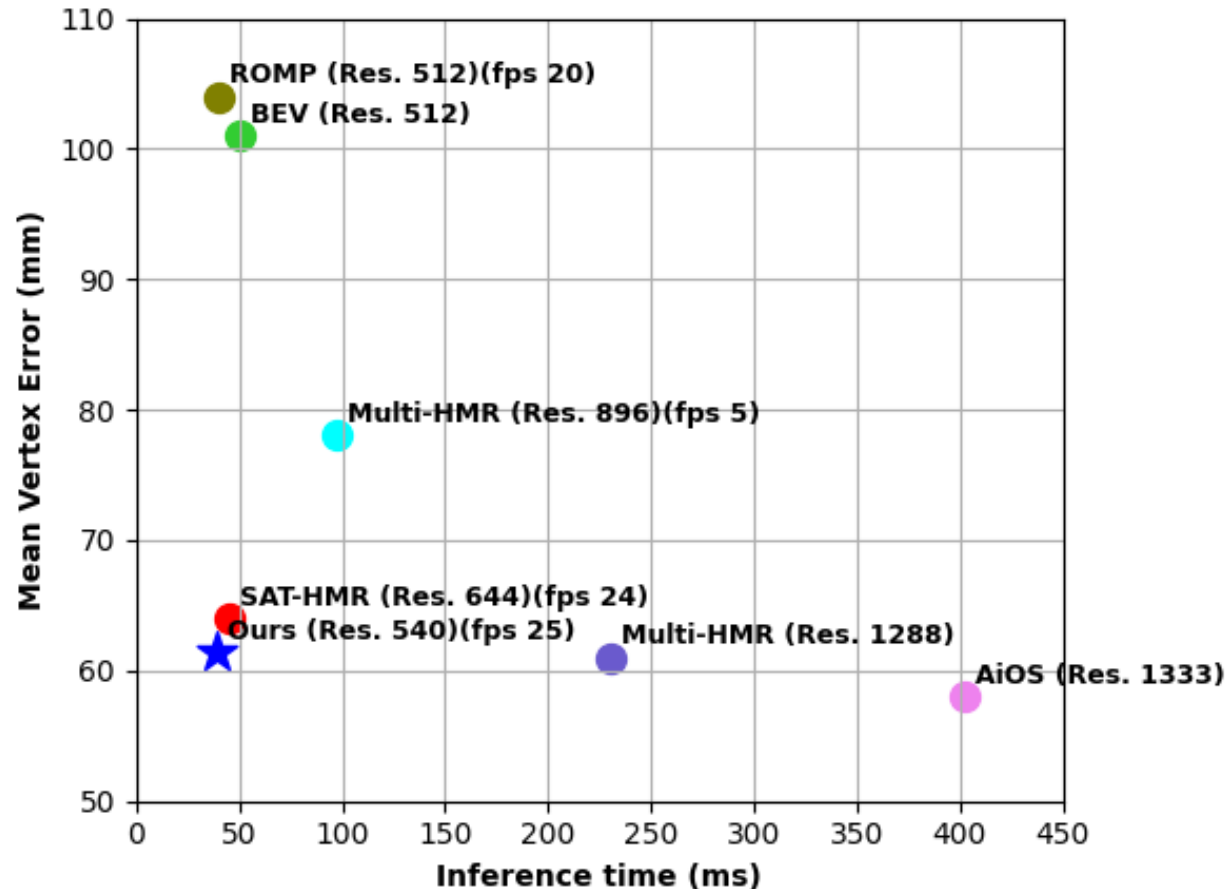
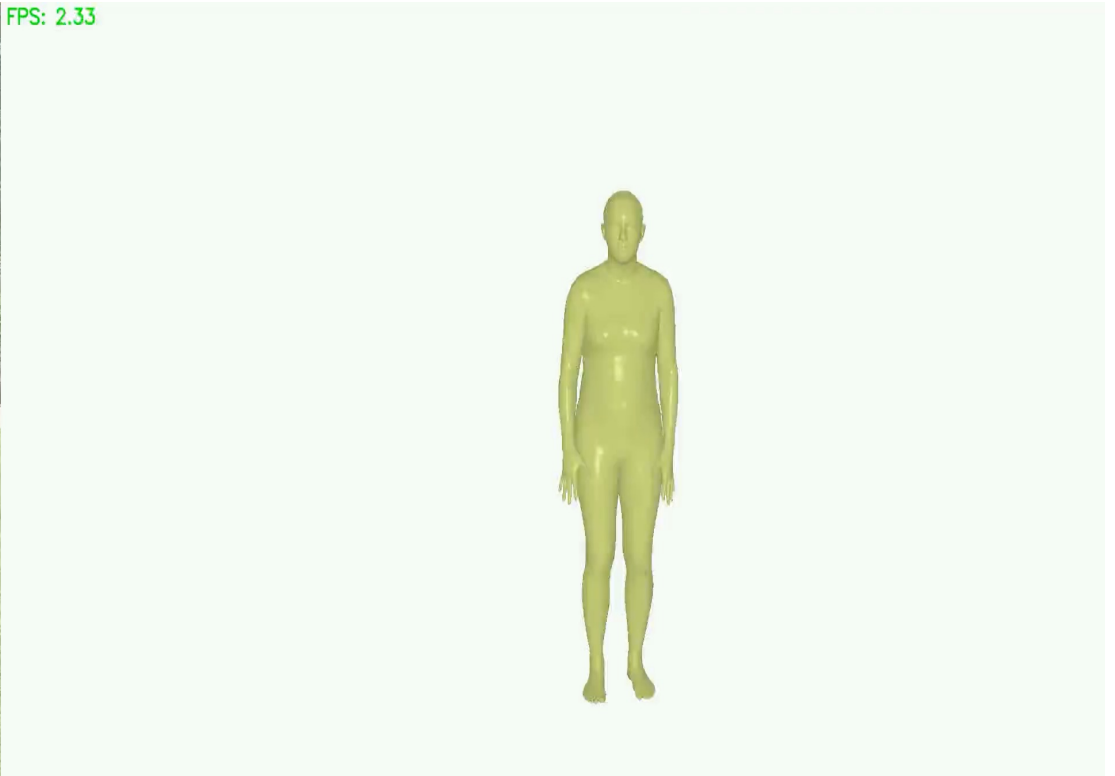


Figure 3. A comparative analysis of error and inference time, with corresponding input resolutions and frame rates (FPS) indicated in parentheses. The proposed method, by utilizing an optimal input resolution, demonstrates comparative performance with SOTA approaches on the AGORA test set while achieving real-time inference capability on a single RTX 3090.

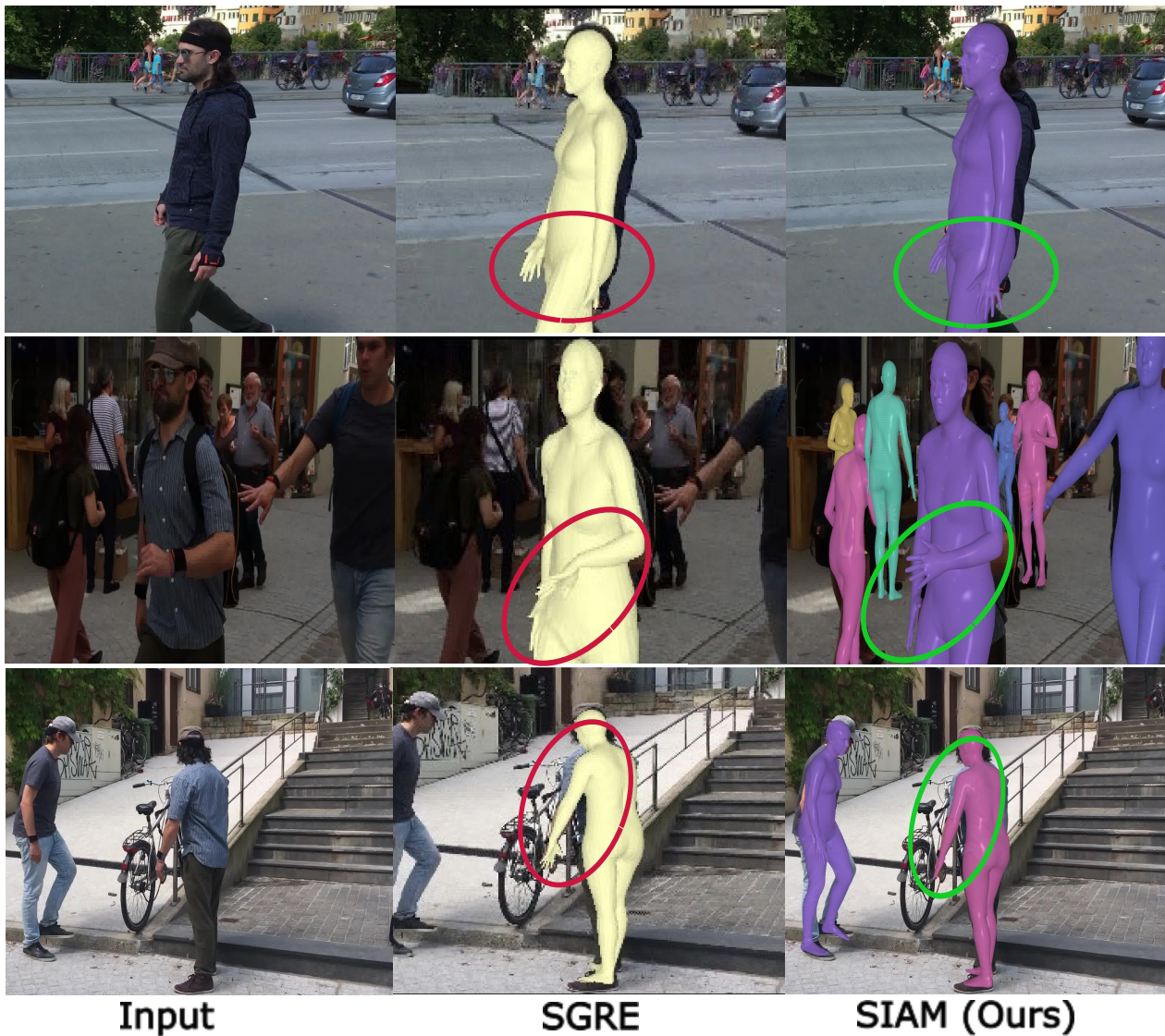
Qualitative analysis



FPS: 2.33



Comparison with SGRE



Error

MPJPE ↓ 18 mm

PA-MPJPE ↓ 15 mm

PVE ↓ 24 mm

MPJPE ↓ 17 mm

PA-MPJPE ↓ 14 mm

PVE ↓ 19 mm

Ablations 3DPW

Set.	Base-32	<i>SIA</i>	<i>FD</i>	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Frame	✓			74.7	49.4	90.1
	✓	✓		69.2	45.8	85.3
	✓		✓	71.0	47.5	87.2
	✓	✓	✓	64.6	40.4	80.7
Video	✓			70.9	47.1	87.3
	✓	✓		66.8	43.3	83.0
	✓		✓	68.5	45.1	84.7
	✓	✓	✓	59.2	35.1	75.9

Set.	Base-48	<i>SIA</i>	<i>FD</i>	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Frame	✓			71.4	47.3	86.8
	✓	✓		66.3	43.6	82.5
	✓		✓	68.0	45.1	84.0
	✓	✓	✓	61.9	38.6	77.3
Video	✓			67.5	45.6	83.2
	✓	✓		63.5	40.7	79.8
	✓		✓	65.1	42.3	81.1
	✓	✓	✓	56.9	32.8	72.2

Table 1. Validation of the SIA and FD modules under both frame-based and video-based settings. Base-32 and 48 refers to our base model using HRNet-32 and 48.

Qualitative CMU-Panoptic

Method	Haggling ↓	Mafia ↓	Ultimatum ↓	Pizza ↓	Avg. ↓
Zanfir <i>et al.</i> [51]	140.0	156.9	150.7	156.0	150.9
CRMH [15]	129.6	133.5	153.0	156.7	143.2
SPIN [22]	124.3	132.4	150.4	153.5	140.1
REMIPS [11]	121.6	137.1	146.4	148.0	138.3
OCHMR [19]	115.5	123.7	142.6	150.6	133.1
ROMP [42]	110.8	122.8	141.6	137.6	128.2
3DCrowdNet [8]	109.6	135.9	129.8	135.6	127.7
JOTR [25]	99.9	113.5	115.7	123.6	113.1
BEV [43]	90.7	103.7	113.1	125.2	108.1
PSVT [35]	88.7	97.9	115.2	121.2	105.7
SAT-HMR [38]	67.9	78.5	95.8	94.6	84.2
SIAM (W32)	65.7	79.3	92.3	93.4	82.7
SIAM (W48)	63.9	77.7	90.5	90.4	80.6

Table 2. Comparison with SOTA methods on the CMU-Panoptic test set. The results are reported from four activities, including their averages. The numbers denote the MPJPE. W32 indicates HRNet-W32 and W48 indicates HRNet-W48.

Qualitative AGORA test set

Method	F1-Score \uparrow	MPJPE \downarrow	MVE \downarrow	NMJE \downarrow	NMVE \downarrow
ROMP [42]	0.91	108.1	103.4	118.8	113.6
BEV [43]	0.93	105.3	100.7	113.2	108.3
PSVT [35]	0.93	97.7	94.1	105.1	101.2
AiOS [40]	0.94	63.9	57.5	68.0	61.2
Mlti-HMR [4]	0.93	82.8	77.6	89.0	83.4
SAT-HMR [38]	0.95	67.9	63.3	71.5	66.6
SIAM (W32)	0.95	55.5	64.3	58.4	72.9
SIAM (W48)	0.94	53.2	61.4	56.0	69.9

Table 3. Comparison of previous methods on the AGORA test set. W32 indicates HRNet-W32 and W48 indicates HRNet-W48.

End

Thank You



Site



Code