Pedestrian-Centric 3D Pre-collision Pose and Shape Estimation from Dashcam Perspective

Meijun Wang[1], Yu Meng*[1], Zhongwei Qiu[2], Chao Zheng[1], Yan Xu[1], Xiaorui Peng[1], Jian Gao[3]

[1]*University of Science and Technology Beijing*， [2]*Alibaba DAMO Academy*， [3]*Northwest University*

Motivation and Main Contribution

Motivation and Main Contribution

Motivation

- Pedestrian pre-collision pose is a key factor in determining collision injury.
- Lack of real pedestrian collision pose dataset.
- Robustness of human pose estimation algorithm.

Daily human pose

VS

Pedestrian pre-collision pose

The contributions are as follows:

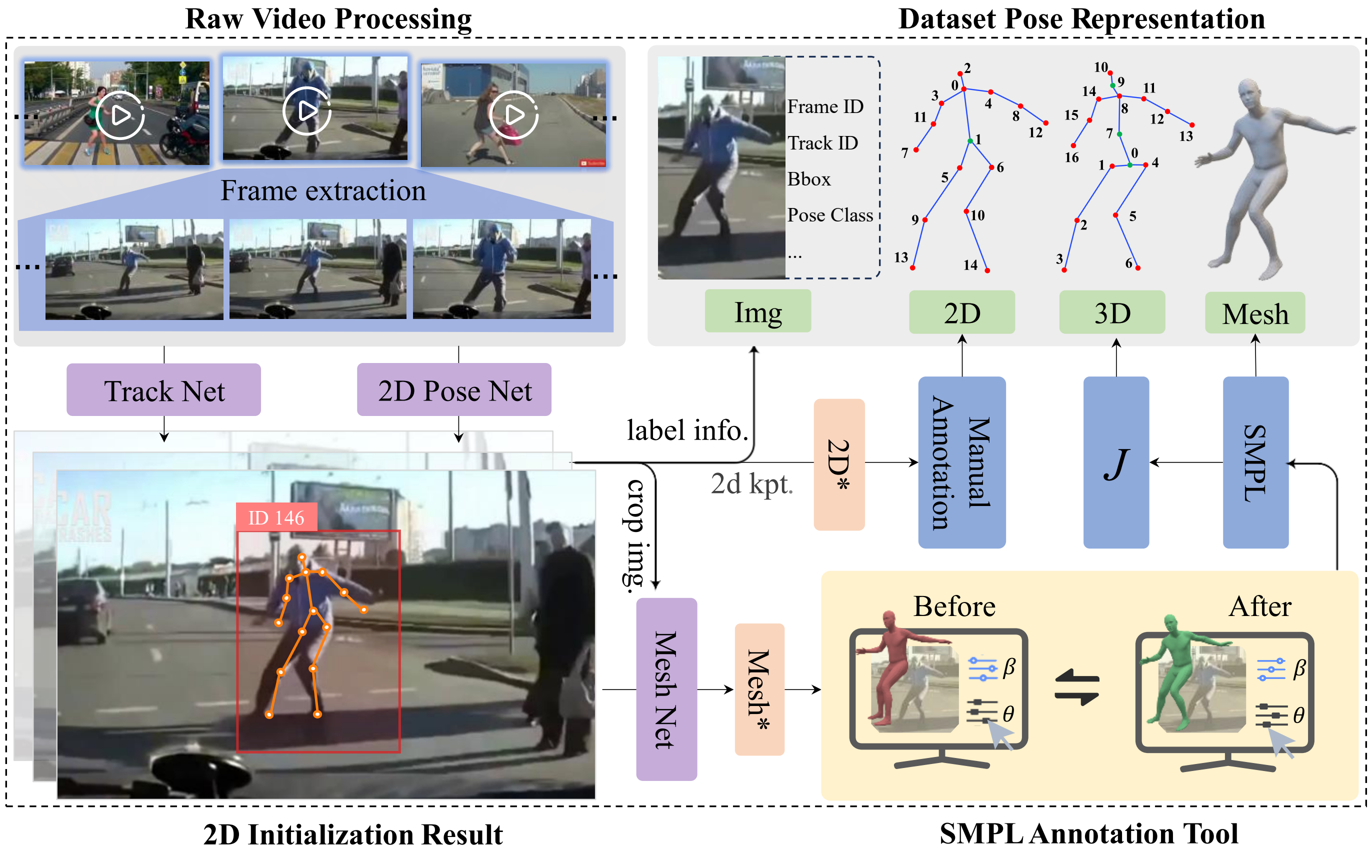
- **PVCP**, a **P**edestrian-**V**ehicle pre-**C**ollision **P**ose dataset.
- **PPSE**, a Pedestrian **P**re-collision **P**ose and **S**hape **E**stimation network.
- Both data and algorithmic support for active safety protection for pedestrians.

PVCP Dataset Pipeline

PVCP Dataset Pipeline

NEURAL INFORMATION PROCESSING SYSTEMS

- Semi-automatic data set annotation process.
- Dashcam Perspective of a real pedestrian-vehicle collision.
- Algorithm initialization annotation and manual annotation tool correction.
- Multiple representation data annotation results (Bbox, ID,2D kpt, 3D kpt and mesh).



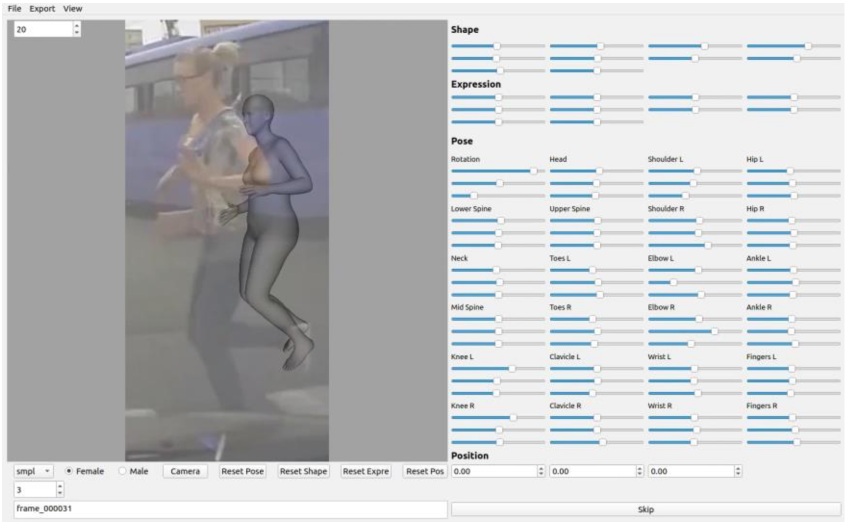
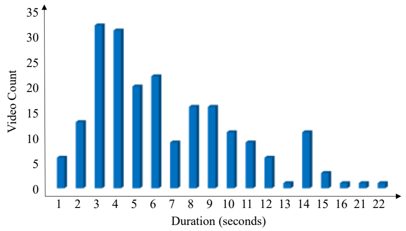


Diagram of the SMPL Annotation Tool

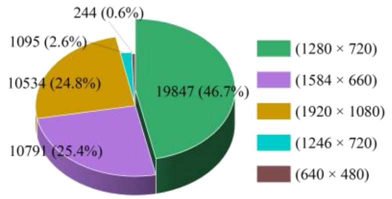
PVCP Dataset Pipeline

Table 1: Comparison of datasets on *Accident Warning*, *Traffic Scene* and *Pedestrian Pose*. 'V' represents the vehicle perspective, 'M' represents the monitoring perspective, 'D' represents a dynamic background and 'S' represents a static background.

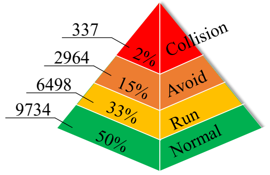
Type	Dataset	Year	Perspective	Background	Detection	Track	Depth	Pose	Shape	Class	Frame
Accident Warning	DAD[22]	2016	V	D	✓(2D Bbox)	✓	×	×	×	×	>62k
	ShanghaiTech[46]	2017	M	S	✓(Mask)	✓	×	×	×	×	>300k
	A3D[23]	2019	V	D	✓(2D Bbox)	✓	×	×	×	×	>128k
	DADA[47]	2019	V	D	✓(3D Bbox)	×	×	×	×	×	>650k
	CCD[24]	2020	V	D	✓(2D Bbox)	✓	×	×	×	×	>75k
Traffic Scene	KITTI[19]	2012	V	D	✓(3D Bbox)	✓	✓	×	×	×	>30K
	Cityscapes[48]	2015	V	D	✓(Mask)	×	×	×	×	×	>5k
	CityPersons[49]	2016	V	D	✓(2D Bbox)	×	×	×	×	×	>5k
	MOT[50]	2012-2017	V/M	D/S	✓(2D Bbox)	✓	×	×	×	×	–
	Nuscenes[18]	2019	V	D	✓(3D Bbox)	✓	✓	×	×	×	>35k
Pedestrian Pose	MSCOCO[20]	2014-2017	Daily scene	S	✓(2D Bbox)	×	×	✓(2D)	×	×	>1000k
	Human3.6M[16]	2014	M	S	✓(2D Bbox)	✓	✓	✓(2D/3D)	×	×	>500k
	PW3D[21]	2018	hand-held camera	D	×	✓	×	✓(3D)	×	×	>50k
	Accident Video[15]	2020	V/M	D/S	×	✓	×	×	×	–	
	PedX[14]	2018	M	S	✓(Mask)	✓	✓	✓(2D/3D)	✓	×	>10k
Ours	**PVCP**	**2024**	**V(Dashcam)**	**D/S**	**✓(2D Bbox)**	✓	✓	**✓(2D/3D)**	✓	✓	**>40k**



Video duration

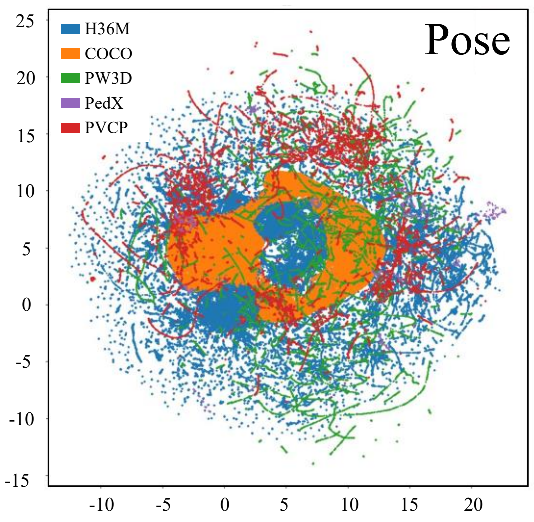


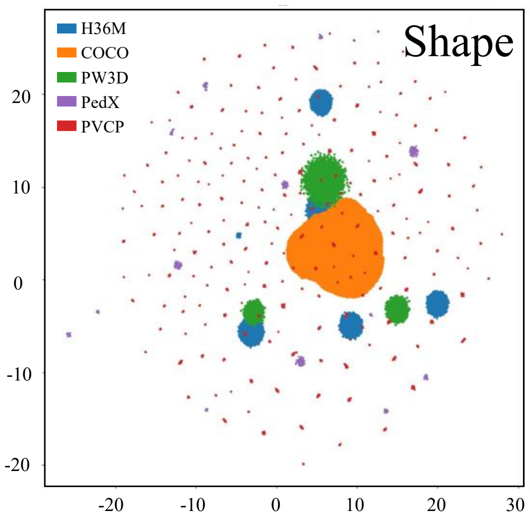
Frame resolution



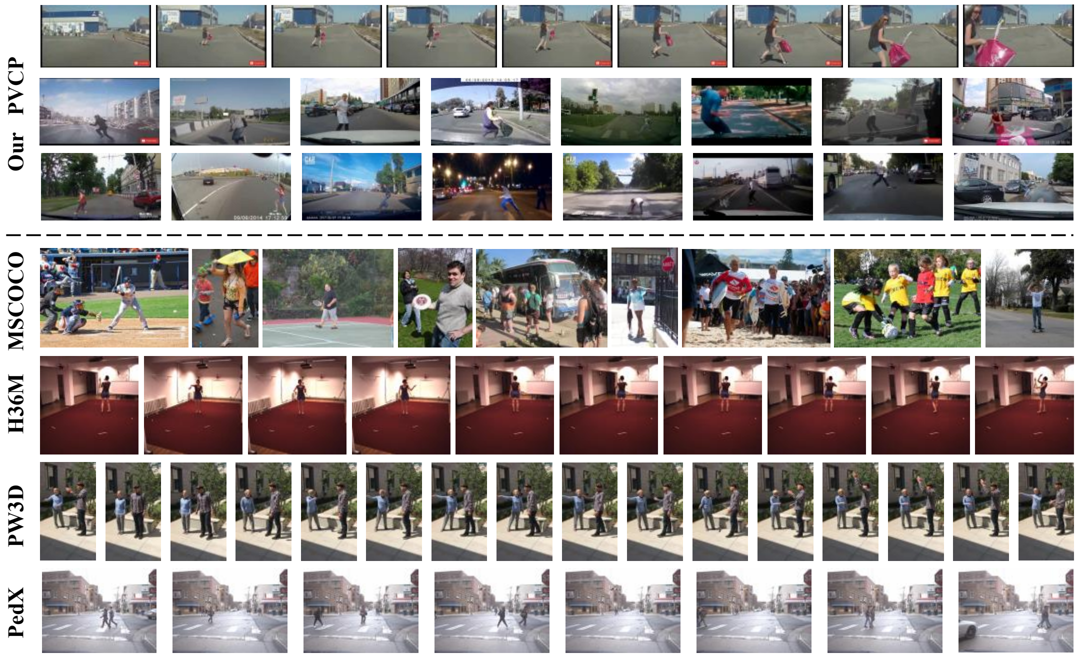
Pose class count

PVCP dataset statistics





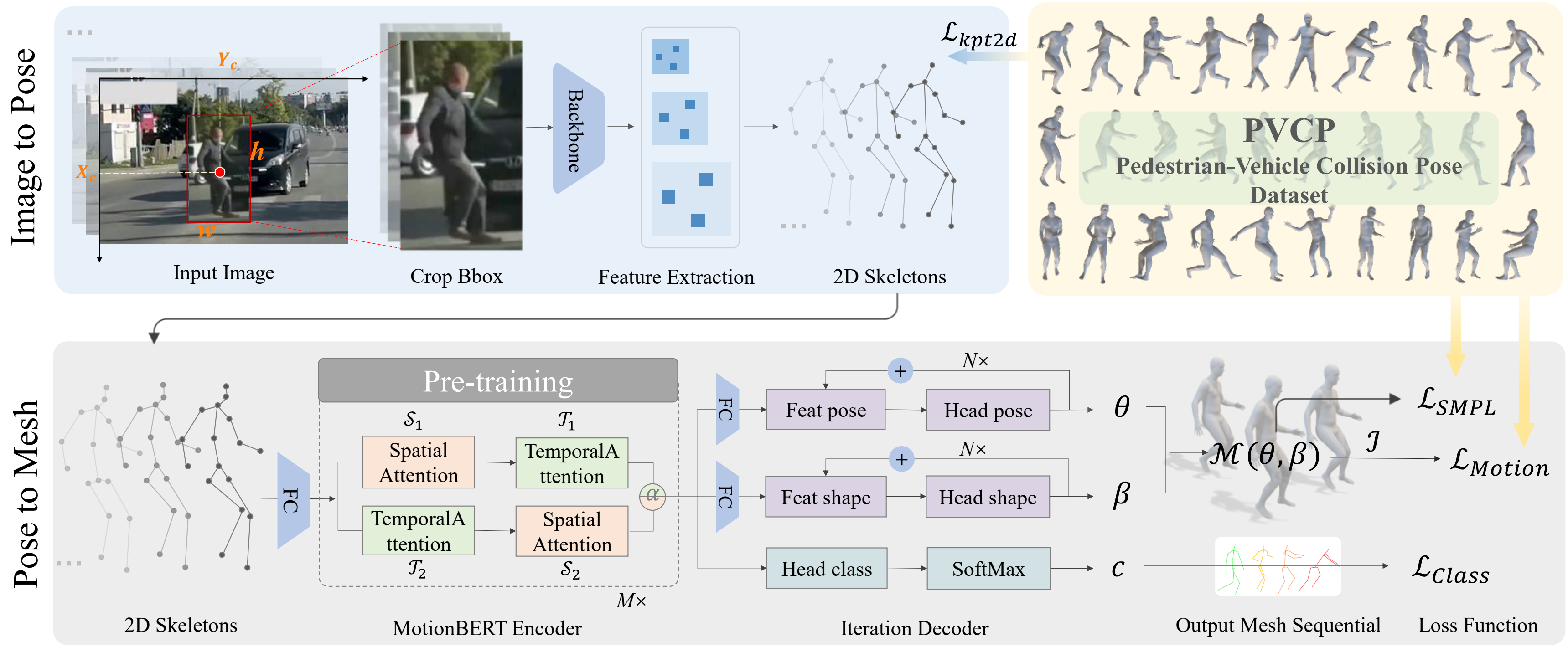
Pose and Shape parameters distribution



Visualization comparison of PVCP with other pose datasets

PPSE Network Architecture

PPSE Network Architecture



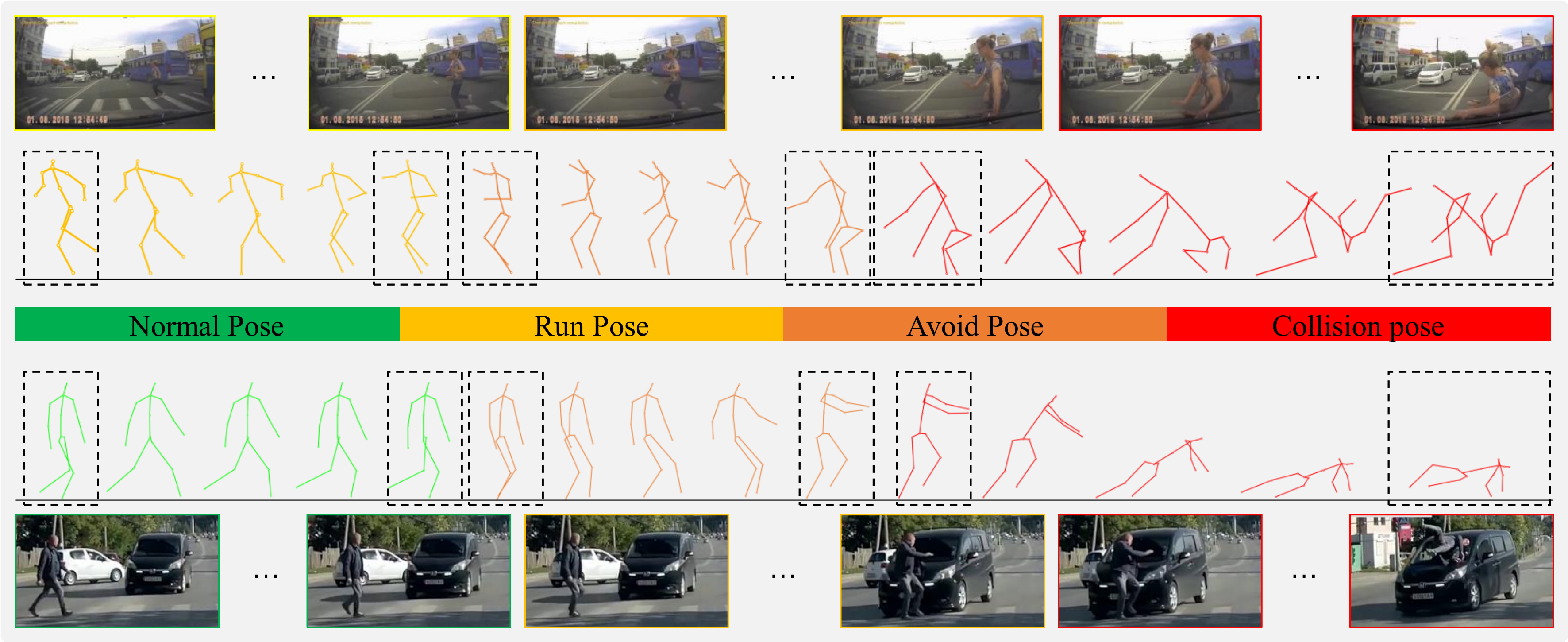
ITP (Image to Pose)

- *Input*: Accident frames and pre-selected pedestrian collision targets $Bbox$.
- *Output*: Pedestrian 2D pre-collision pose $P_{2d} \in \mathbb{R}^{15\times 2}$.

PTM (Pose to Mesh)

- *Input*: Pedestrian 2D pre-collision pose sequence $P_{2d}^{L} \in \mathbb{R}^{T\times J\times C}$.
- *Output*: Pedestrian 3D mesh pre-collision pose sequence $\mathcal{M}(\theta,\beta) \in \mathbb{R}^{T\times N\times C}$.

PPSE Network Architecture



- **Pre-trained model**

$$F^i = \alpha_{ST}^i \circ \mathcal{T}_1^i(\mathcal{S}_1^i(F^{i-1})) + \alpha_{TS}^i \circ \mathcal{S}_2^i(\mathcal{T}_2^i(F^{i-1}))$$

$$\alpha_{ST}^i, \alpha_{TS}^i = softmax(\mathcal{W}_f(\mathcal{T}_1^i(\mathcal{S}_1^i(F^{i-1})) \oplus \mathcal{S}_2^i(\mathcal{T}_2^i(F^{i-1}))))$$

- **Iterative regression**

$$\theta^k = W_\theta^k(F_\theta) + \theta^{k-1}$$

$$\beta^k = W_\beta^k(F_\beta) + \beta^{k-1}$$

$$c = softmax(W_c(F_c))$$

- **Introduce pose category loss**

$$\mathcal{L}_{Class} = \lambda_c \mathcal{L}_{Cross\ Entropy}(\hat{C}, C)$$

$$\mathcal{L} = \mathcal{L}_{PTM} + \mathcal{L}_{PTM}$$

$$= \mathcal{L}_2 + \mathcal{L}_{SMPL} + \mathcal{L}_{Motion} + \mathcal{L}_{Class}$$

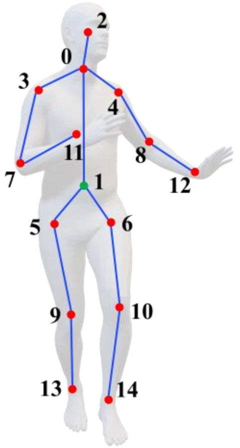
Experimental and Results

Experimental and Results

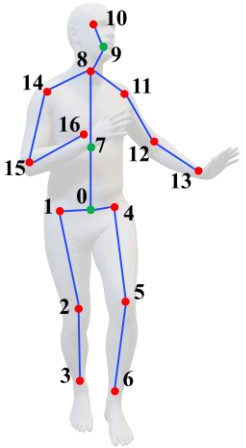
Evaluation Metric

- (**P**rocrustes-**A**ligned) **M**ean **P**er-**V**ertex **E**rror
- (**P**rocrustes-**A**ligned) **M**ean **P**er **J**oint **P**osition **E**rror
- ***X_14j*** (14 common keypoints, Red keypoints)
- ***X_17j*** (Representation of the Human3.6M)

ID	Joint name
0	neck
1	belly
2	head
3	right_shoulder
4	left_shoulder
5	right_hip
6	left_hip
7	right_elbow
8	left_elbow
9	right_knee
10	left_knee
11	right_wrist
12	left_wrist
13	right_ankle
14	left_ankle



(a) 15 keypoints of JHMDB



ID	Joint name
0	pelvis
1	right_hip
2	right_knee
3	right_ankle
4	left_hip
5	left_knee
6	left_ankle
7	spine
8	neck
9	nose
10	head
11	left_shoulder
12	left_elbow
13	left_wrist
14	right_shoulder
15	right_elbow
16	right_wrist

(a) 17 keypoints of Human3.6M

Table 2: Effects of Dataset and Pre-training. Top use detected 2D pose sequences. Bottom use GT 2D pose sequences.

Input	Train Set	testset	Pose class	MPVE	PAMPVE	MPJPE_14j	PAMPJPE_14j	MPJPE_17j	PAMPJPE_17j
2D Det	PVCP	PVCP	Normal	315.94	160.25	272.18	130.72	246.42	121.30
			Run	318.29	189.84	274.78	160.35	246.95	145.07
			Avoid	305.01	159.19	260.31	121.42	232.56	113.21
			Collision	347.53	171.82	311.88	145.64	281.35	139.46
			All	315.64	168.11	271.91	137.75	245.35	126.92
	Pretrain	PVCP	Normal	347.10	190.17	312.21	154.85	285.62	145.55
			Run	309.19	183.27	277.01	152.19	251.53	141.11
			Avoid	330.18	189.69	293.76	155.54	264.89	144.38
			Collision	334.14	164.32	301.52	133.26	275.28	128.19
			All	335.11	188.09	300.80	154.06	274.27	144.12
	Pretrain + PVCP	PVCP	Normal	294.73	170.10	253.80	137.39	232.74	128.24
			Run	253.16	149.99	219.06	124.01	200.19	115.27
			Avoid	286.85	159.69	246.94	124.86	222.02	114.96
			Collision	250.58	161.25	222.47	127.37	200.38	120.47
			All	**282.50**	**163.58**	**243.59**	**132.43**	**222.70**	**123.33**
2D GT	PVCP	PVCP	Normal	304.65	167.56	260.68	138.49	233.70	126.83
			Run	296.75	192.00	254.58	163.80	226.49	146.66
			Avoid	277.51	157.48	234.30	123.02	206.55	113.44
			Collision	354.76	178.22	319.56	154.95	287.38	146.83
			All	300.04	173.09	256.69	143.73	229.30	130.85
	Pretrain	PVCP	Normal	175.24	111.72	152.10	87.68	138.87	82.11
			Run	153.45	107.26	131.93	84.02	118.99	77.76
			Avoid	143.32	93.61	122.91	73.45	111.33	68.89
			Collision	151.18	91.20	133.60	77.66	124.71	71.06
			All	165.90	108.48	143.52	85.14	130.56	79.48
	Pretrain + PVCP	PVCP	Normal	156.06	103.16	132.74	80.59	120.35	74.92
			Run	129.49	89.31	109.93	70.91	100.19	65.70
			Avoid	127.04	85.36	108.30	65.35	96.74	60.44
			Collision	135.89	89.71	127.11	70.86	112.50	64.94
			All	**145.77**	**97.50**	**124.04**	**76.34**	**112.43**	**70.87**

Experimental and Results

Table 3: Component of system. Top use detected 2D pose sequences. Bottom use GT 2D pose sequences.

Input	Pretrain	Iter	Class Loss	Pose class	MPVE	PAMPVE	MPJPE_14j	PAMPJPE_14j	MPJPE_17j	PAMPJPE_17j
2D Det	✓			All	282.50	163.58	243.59	132.43	222.70	123.33
	✓	3		All	266.20	146.88	225.38	116.99	204.98	108.63
	✓		✓	All	259.05	**143.52**	<u>220.39</u>	<u>115.47</u>	<u>200.16</u>	<u>107.03</u>
	✓	3	✓	All	**257.75**	<u>144.19</u>	**218.61**	**114.50**	**198.16**	**105.86**
2D GT	✓			All	145.77	97.50	124.04	76.34	112.43	70.87
	✓	3		All	145.75	96.69	123.16	75.13	111.90	69.89
	✓		✓	All	<u>141.28</u>	**92.78**	<u>120.16</u>	**72.43**	<u>108.90</u>	**67.58**
	✓	3	✓	All	**140.43**	<u>96.43</u>	**118.80**	<u>75.13</u>	**107.47**	<u>69.56</u>

Table 4: Comparison of 2D GT input in different iterations number.

Iter	Pose class	MPVE	PAMPVE	MPJPE_14j	PAMPJPE_14j	MPJPE_17j	PAMPJPE_17j
2	All	141.95	97.43	120.04	75.45	108.63	69.85
3	All	<u>140.43</u>	**96.43**	<u>118.80</u>	**75.13**	<u>107.47</u>	**69.56**
4	All	**139.96**	<u>96.92</u>	**118.46**	<u>75.19</u>	**107.16**	<u>69.62</u>
5	All	140.01	97.10	118.54	75.40	107.27	69.83
6	All	140.41	97.42	118.89	75.70	107.68	70.14

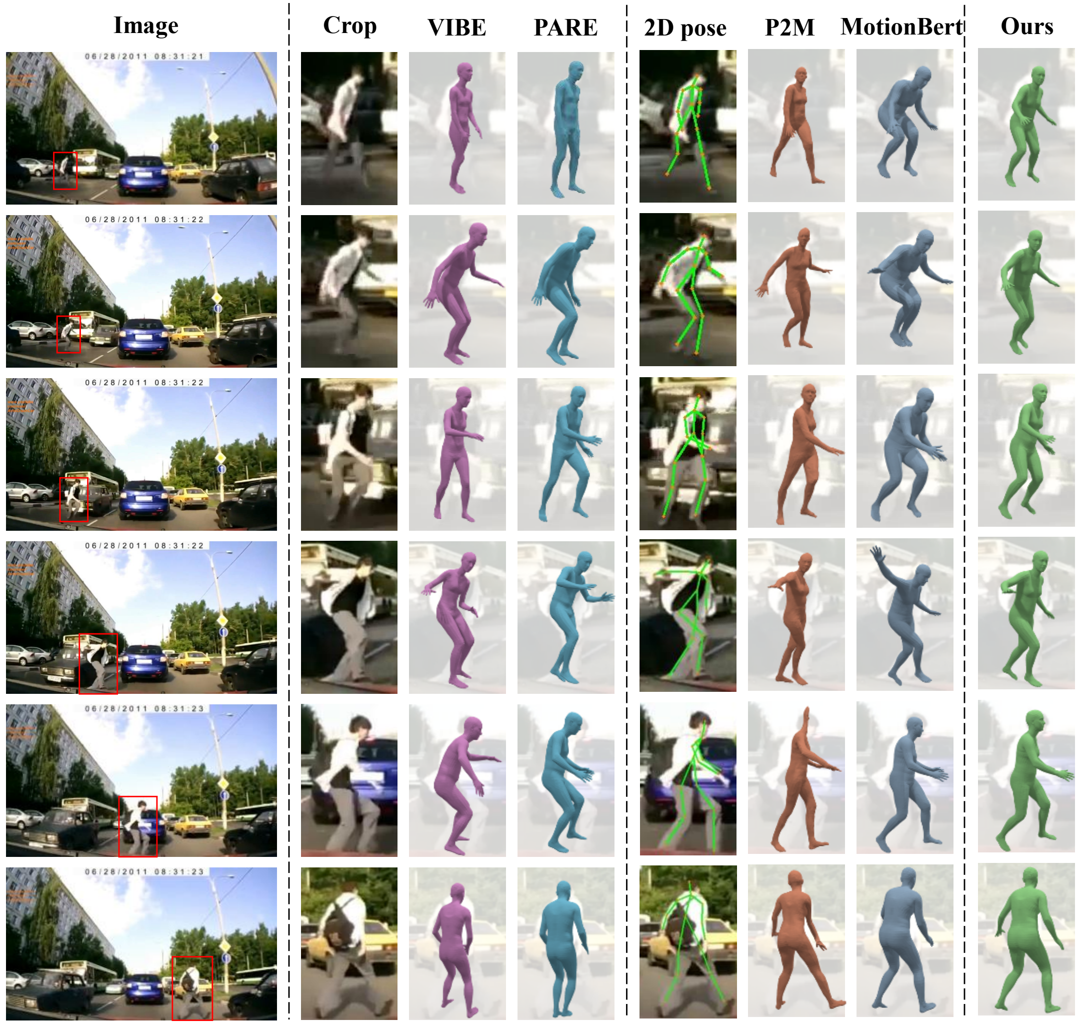
Experimental and Results

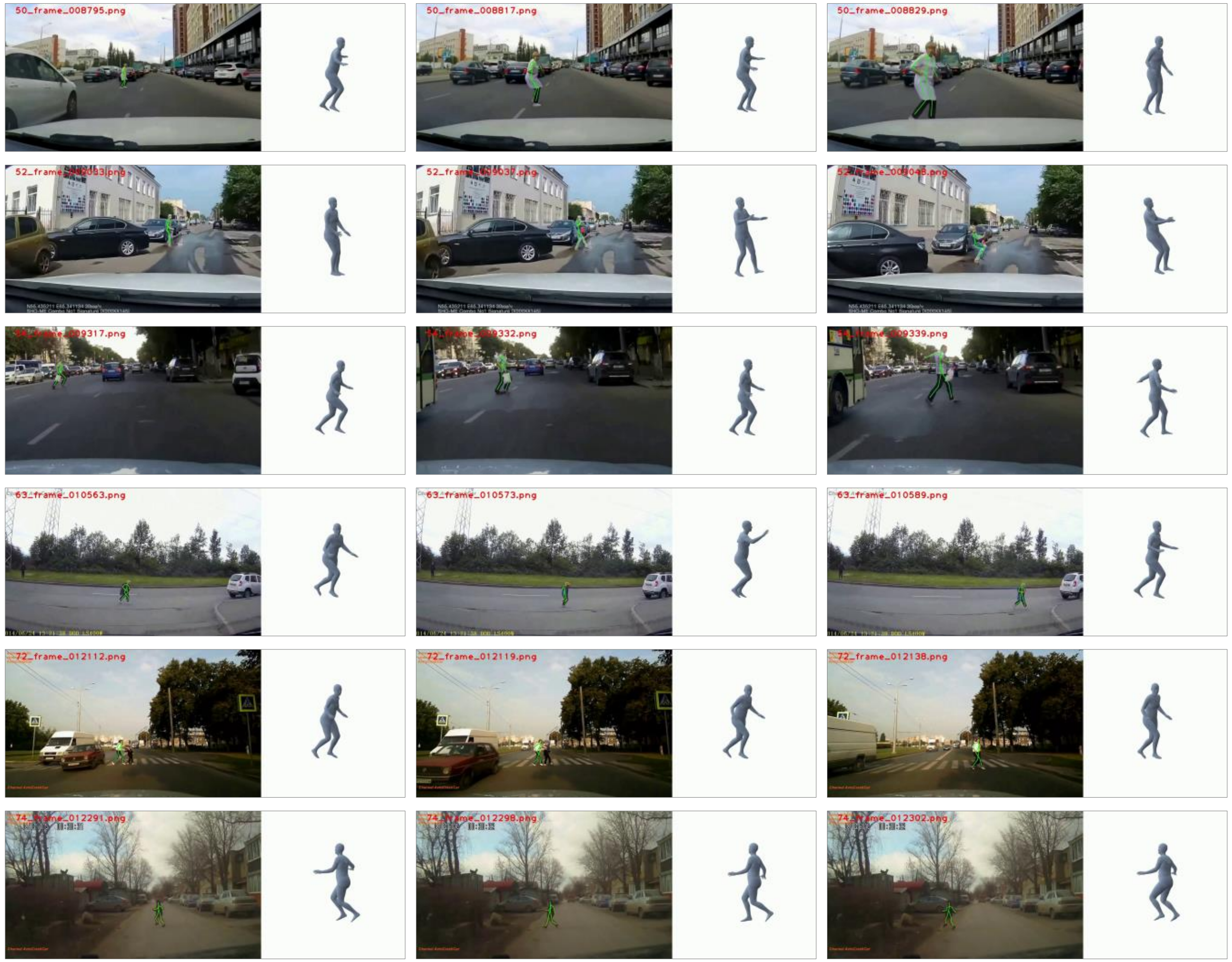
Table 5: Comparison of state-of-the-art methods on the PVCP testset. † denotes that the training weights provided by the official are used, and * denotes the model weights trained together with the PVCP trainset.

Paradigm	Method	Pose class	MPVE	PAMPVE	MPJPE_14j	PAMPJPE_14j	MPJPE_17j	PAMPJPE_17j
One Stage	†VIBE(66)	Normal	856.87	234.47	731.90	217.35	–	–
		Run	856.10	232.67	732.33	226.45	–	–
		Avoid	777.92	227.16	664.25	216.72	–	–
		Collision	950.47	212.21	869.86	202.01	–	–
		All	849.09	233.08	725.92	219.55	–	–
	†PARE(67)	Normal	225.99	147.04	193.62	114.35	–	–
		Run	235.99	180.98	193.40	137.08	–	–
		Avoid	210.02	143.88	176.76	109.10	–	–
		Collision	247.18	167.62	225.96	132.89	–	–
		All	**226.98**	<u>155.72</u>	**191.97**	<u>119.85</u>	–	–
Two Stage	†Pose2Mesh(68)	Normal	247.24	148.87	222.34	122.42	–	–
		Run	255.26	181.16	222.33	145.14	–	–
		Avoid	217.97	141.43	191.38	112.35	–	–
		Collision	231.65	174.44	210.44	145.54	–	–
		All	245.88	156.69	218.71	127.41	–	–
	*MotionBERT(12)	Normal	294.73	170.10	253.80	137.39	232.74	128.24
		Run	253.16	149.99	219.06	124.01	200.19	115.27
		Avoid	286.85	159.69	246.94	124.86	222.02	114.96
		Collision	250.58	161.25	222.47	127.37	200.38	120.47
		All	282.50	163.58	243.59	132.43	222.70	123.33
	*PPSET(Ours)	Normal	272.79	149.02	230.49	117.47	209.99	109.04
		Run	226.22	133.45	193.75	109.50	174.47	100.73
		Avoid	251.60	143.52	212.75	109.75	190.00	100.09
		Collision	217.68	134.95	201.15	113.10	174.57	105.94
		All	257.75	**144.19**	<u>218.61</u>	**114.50**	**198.16**	**105.86**

Experimental and Results

NEURAL INFORMATION
PROCESSING SYSTEMS





Limitations and Future Work

Limitations and Future Work

Integrity of the dataset

- ✓ Due to the difficulty of collecting the dataset, the dataset is *small in size* and lacks real *camera parameters*, *vehicle speed* information, *global position* and *direction* of pedestrians.

Real-time performance of the model

- ✓ Our method is not real-time at present, because our input is *Image* and *pre-selected Bbox sequence* of collision pedestrian targets.

Thanks for your listening.

NEURAL INFORMATION PROCESSING SYSTEMS

Pedestrian-Centric 3D Pre-collision Pose and Shape Estimation from Dashcam Perspective

GitHub

https://github.com/wmj142326/PVCP

北京科技大学
University of Science and Technology Beijing

達摩院
DAMO ACADEMY

西北大学
NORTHWEST UNIVERSITY