ISSN 2464-4617 (print)
ISSN 2464-4625 (online)
Computer Science Research Notes - CSRN
http://www.wscg.eu
WSCG 2025 Proceedings

PeDesCar: A Large-Scale Dataset for Simulated Car-Pedestrian Collisions to Advance Safety Research

Edgar Medina
QualityMinds GmbH
edgar.medina@qualityminds.de

Leyong Loh*
QualityMinds GmbH
leyong.loh@tum.de

ABSTRACT

Car-pedestrian collisions are a daily occurrence worldwide, yet there is a notable absence of public datasets in this domain. Research in this area is crucial, as it directly impacts pedestrian safety and serves as a basis for validating autonomous driving systems. Although finite element simulations are used, they are computationally intensive and yield insufficient data for deep learning applications. In this work, we present the PeDesCar dataset for safe autonomous driving, which spans around 15 days of simulated time and encompasses over 1 million collision events, each constrained within a temporal window of up to 2 seconds per event. The dataset is generated using MuJoCo as a physics simulator, proving its effectiveness in sim2real robotics research. We use PeDesCar to train and assess state-of-the-art models in human motion prediction and validate the realism of the simulation against realistic high-fidelity finite element simulations. Our results validate that PeDesCar is sufficient for preliminary car-pedestrian collision research. The visualization code and dataset are accessible on the project website `https://github.com/QualityMinds/PeDesCar`.

1 INTRODUCTION

Recent strides in the field of human motion prediction have brought forth many research endeavors showcasing notable outcomes Lyu et al. [2022]. Currently, several publicly available datasets are available to forecast future human movements. These datasets cover indoor and outdoor scenarios, capturing data from 18 to 22 joints. They serve as pivotal references in numerous researches within this field. Despite the advances made by these dataset benchmarks in human motion with diverse actions and environments, there persists a need for datasets tailored specifically to manufacturing or safety applications. These include tasks such as injury prediction or assembly processes, which require datasets with specialized characteristics and annotations Lyu et al. [2022], Niranjan et al. [2023]. In autonomous vehicles, novel concepts have emerged in the real-time injury prediction sustained by unprotected road users (URVs). The prospective ability to predict injuries facilitates the implementation of automated vehicle risk mitigation strategies, thus ensuring safe and efficient traffic flow Niranjan et al. [2023]. However, simulation tools have improved efficiency, accuracy, and execution time. We used MuJoCo Mujoco to simulate various environments featuring different humanoid agents sourced from Loco-Mujoco Al-Hafez et al. [2023], along with two CAD car models. Spanning 'walking' and 'running' actions in these agents which were trained via imitation learning.

We acknowledge the inherent disparities between synthetic and real data, as well as the potential inaccuracies in physical simulations that may result in deviations from reality. While several datasets exist for 3D human motion prediction, outlined in the following section, it is important to recognize that two commonly employed methods, MoCap (motion capture) and generative models, face significant challenges when applied to injury cases in autonomous driving scenarios. MoCap, while widely used, is impractical for injury cases due to its inability to accurately capture or replicate complex injury dynamics, and is unreasonable to test directly in humans. Similarly, generative models present hurdles in accurately representing the latent distribution of multiple injury scenarios. In other words, they primarily focus on generating 3D poses for forward steps without incorporating the intricate physics involved in human motion dynamics. These constraints curtail their utility for injury prediction and prevention within autonomous driving contexts. This area has garnered increasing attention in the literature, yet it persists as an open challenge lacking a definitive resolution. Hence, our work intends to provide a benchmark dataset, bridging training models for injury prediction and human motion forecasting. To address this gap in research, our dataset encompasses a wide array of initial conditions for both subjects and cars. We summarize our contributions as follows.

- Delivering a new dataset for impact collisions with different humanoid models.
- Presenting a benchmark in this dataset using state-of-the-art architectures.
- Releasing the open-source code and dataset.

2 LITERATURE REVIEW

2.1 Human Motion Prediction

In recent years, we observed an increment of works in the direction of predicting the future movement in 3D

*Work done during a research internship at QualityMinds GmbH

Dataset	Year	subjects	joints	frames	FPS	capture system
CMU Ionescu et al. [2014]	2003	8	38	1.5M	25	Marker-less MoCap
H36M Ionescu et al. [2014]	2014	11	32	3.6M	25	Marker-based MoCap
MPI-INF-3DHP Mehta et al. [2018]	2017	8	8	1.3M	25	Marker-less MoCap
3DPW von Marcard et al. [2018]	2018	7	17	51K	30	Cameras with IMUs
AMASS Mahmood et al. [2019]	2019	300	52	9M	25	Marker-based MoCap
NBA2K Zhu et al. [2020]	2020	27	14	27K	30	NBA2K19 game engine
GTA-IM Cao et al. [2020]	2020	20	21	1M	30	GTA game engine
PeDesCar	2025	4	18-22-23	176.7M	200	Mujoco Engine

Table 1: Comparison of our dataset with the most relevant datasets, including some comparable attributes.

Model	Dataset Train	Dataset Test	Time (ms) 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	AMASS	AMASS	11.2	20.6	36.5	43.1	52.5	68.7
	AMASS	MujAmass	46.5	69.6	123.7	149.4	191.9	267.4
	MujAmass	AMASS	27.4	46.7	96.4	121.6	167.6	245.1
	MujAmass	**MujAmass**	**12.2**	**21.9**	**42.5**	**52.0**	**69.6**	**114.9**
	H36M	H36M	11.4	24.8	51.2	62.6	81.1	113.8
	H36M	MujH36m	30.2	61.5	112.9	139.7	188.0	278.7
	MujH36m	H36M	36.2	51.4	96.9	120.8	164.9	256.9
	MujH36m	**MujH36m**	**14.8**	**24.5**	**47.4**	**57.9**	**77.4**	**126.5**
MotionMixer Bouazizi et al. [2022]	AMASS	AMASS	10.1	18.4	32.7	38.9	48.3	64.2
	AMASS	MujAmass	33.8	64.1	116.4	139.4	181.0	275.6
	MujAmass	AMASS	19.0	40.0	83.3	103.1	137.1	178.3
	MujAmass	**MujAmass**	**10.7**	**19.4**	**35.3**	**42.9**	**58.3**	**102.4**
	H36M	H36M	11.0	23.6	47.8	59.3	-	111.0
	H36M	MujH36m	24.6	51.8	106.7	132.0	176.4	269.1
	MujH36m	H36M	20.8	42.3	83.3	101.2	131.8	184.3
	MujH36m	**MujH36m**	**13.1**	**25.2**	**47.9**	**58.6**	**79.4**	**136.2**
CIST-GCN Medina et al. [2024]	AMASS	AMASS	9.8	18.6	33.6	39.8	49.2	63.6
	AMASS	MujAmass	18.2	47.1	112.2	143.6	194.6	291.2
	MujAmass	AMASS	20.3	46.0	106.7	134.8	188.2	282.5
	MujAmass	**MujAmass**	**9.4**	**19.7**	**39.8**	**49.0**	**66.0**	**111.6**
	H36M	H36M	10.5	23.2	47.9	59.0	77.2	110.3
	H36M	MujH36m	17.7	41.4	93.3	118.4	162.9	259.7
	MujH36m	H36M	19.5	45.1	104.7	133.4	185.3	279.3
	MujH36m	**MujH36m**	**10.4**	**22.2**	**45.8**	**56.7**	**77.5**	**129.7**

Table 2: MPJPE evaluation of models trained on AMASS, H36M, and pro1A from PeDesCar. We use the MuJoCo extension, which provides skeletons with the same joint distribution as H36M and AMASS.

poses. Given an input 3D pose sequence a neural network predicts the subsequent 3D pose sequence output. Initial research started with RNN, GCN, and CNN architectures Li et al. [2018], Sofianos et al. [2021], Lyu et al. [2022]. Later, more sophisticated approaches intended to predict and produce sample-based interpretability Fu et al. [2023], Zhong et al. [2022], Medina et al. [2024], used MLP-based Bouazizi et al. [2022] architectures, or multiple-stages strategies Dang et al. [2021], Mao et al. [2021], Ma et al. [2022]. New approaches using transformers have recently achieved cutting-edge performance in this task Chen et al. [2022], Nargund and Sra [2023], Xu et al. [2023].

2.2 3D Pose datasets

The Human3.6M dataset Ionescu et al. [2014] contains 3.6 million accurate 3D human poses and corresponding images. It features data from 11 subjects (5 females, 6 males) across 17 scenarios, recorded from 4 viewpoints using a high-speed motion capture system with 4 high-resolution cameras. The dataset includes 3D joint positions, video, 24 body part labels, and 3D laser scans. Standard experimental setups downsample the frame rate to 25Hz, using 22 joints, with subjects 1, 6, 7, 8, and 9 for training, subject 11 for validation, and subject 5 for testing Li et al. [2018], Sofianos et al. [2021], Dang et al. [2021], Mao et al. [2021], Fu et al. [2023], Bouazizi et al. [2022], Ma et al. [2022], Zhong et al. [2022], Medina et al. [2024]. Another benchmark is AMASS Mahmood et al. [2019], which unifies 15 motion capture datasets using the MoSH++ algorithm. It standardizes diverse body parameterizations via the SMPL (Skinned Multi-Person Linear) model Loper et al. [2015], resulting in over 40 hours of motion data from more than 300 subjects and 11,000+ motions. Experiments typically downsample to 25Hz and use 18 joints, with 8 subdatasets for training, 4 for validation, and BMLrub for testing Bouazizi et al. [2022], Zhong et al. [2022], Medina et al. [2024]. The 3DPW dataset von Marcard et al. [2018], used to test the generalization of AMASS-trained models, combines IMUs and handheld camera footage. It includes 60 video sequences, 2D annotations, and 3D poses derived from IMUs, covering both indoor and outdoor activities. It contains 51,000 frames at 30Hz, with the same joint structure as AMASS. In ExPI Wen et al. [2022], Participants are outfitted with IMUs while undergoing concurrent capture through handheld cameras. The ExPI dataset comprises 115 sequences

encompassing approximately 30,000 frames, depicting two professional couples executing 16 distinct dance actions. This dataset was meticulously annotated with 3D body poses and shapes. ExPI serves as a valuable resource for collaborative motion prediction. All recordings are conducted within a multiview motion capture studio, operating at a frame rate of 25 frames per second (fps). The authors in Wen et al. [2022] opted for an analysis involving 18 joints per dancer and conducted experiments utilizing two distinct benchmarks, referred to as 'common' and 'unseen' action splits.

In Wdowicz and Ptak [2023] also reference several other datasets relevant in the literature that are used for benchmarking analysis, such as NBA2K Zhu et al. [2020] with 27,000 frames, GTA-IM Cao et al. [2020] with 1 million frames, and Occlusion-Person with 73,000 frames Zhang et al. [2021]. All were generated in game engines. Simultaneously, FORCE Zhang et al. [2024] with 192,000 frames in 450 motion sequences in pervasive interactions of carrying, pushing, and pulling objects. More recently, Wagener et al. [2022] introduced MoCapAct, a dataset of expert policies trained via reinforcement learning to track motion-capture data, along with their corresponding observations and actions. In Guo et al. [2022], a dual-stage methodology comprising text2length and text2motion modules was employed for sampling and generation tasks. Text2length entails sampling from a trained distribution function of motion lengths, conditioned on input text. Subsequently, the text2motion module utilizes a temporal variational autoencoder to synthesize human motions corresponding to the sampled lengths. To handle text and motion in this task, the authors used HumanML3D, composed by HumanAct12 Guo et al. [2020] and AMASS datasets, and KIT-ML Plappert et al. [2016]. HumanML3D, is constructed, consisting of 14,616 motion clips and 44,970 text descriptions while KIT-ML consists only of 3,911 motions and 6,278 texts.

While numerous methodologies exist for predicting 3D pose datasets, none have yet been devised, whether through MoCap techniques or generative modeling, to accurately replicate injury scenarios pertinent to autonomous driving. Also, acquiring such datasets via conventional methods remains exceedingly unfeasible. Some of these datasets are presented in Tab. 1.

2.3 Physics Simulation

Recently, simulations have played an increasingly important role in robotics and reinforcement learning (RL) research. While various methods exist for data generation, such as game engines Cao et al. [2020] which may lack fidelity in terms of physics reliability, and finite element method (FEM) simulation software known for its high cost Trube et al. [2023], Niranjan et al. [2023], we have chosen a middle-ground approach. In this pedestrian collision study, we employ a 'light' physics simulator, such as MuJoCo, which balances realism and computational efficiency. For comparison purposes, Wagener et al. [2022] evaluate the most commonly used RL simulation environments in robotics research, such as MuJoCo, PyBullet, Gazebo, IsaacSim, and Webots. MuJoCo has proven its effectiveness as a simulator, especially for sim-to-real transfer Wang et al. [2022], Byravan et al. [2022], Haarnoja et al. [2023], Zhao et al. [2020]. The authors in Al-Hafez et al. [2023] introduce the LocoMuJoCo imitation learning benchmark for locomotion, which provides physically realistic humanoid models. We utilized these models for our experiments. Due to the ease of experimentation offered by the benchmark and its robust foundation in sim-to-real transfer, we selected MuJoCo as our backend simulator for dataset generation.

3 METHODOLOGY

3.1 Preliminary

As our focus lies within the realm of 3D human pose estimation, we define the input sequence as $X = \{x_0, x_1, ..., x_{t_1-1}\} \in \mathbb{R}^{t_{in} \times J \times D}$, which is fed into a model to obtain the output sequence $Y = \{y_0, y_1, ..., y_{t_1-1}\} \in \mathbb{R}^{t_{out} \times J \times D}$, where t_{in} and t_{out} typically correspond to 400 ms and 1000 ms, respectively, with a sampling rate of 40 ms. Nevertheless, some studies have explored extended input durations ranging from 400 ms to 1000 ms while maintaining equivalent output shapes. Benchmark datasets such as AMASS Mahmood et al. [2019], 3DPW von Marcard et al. [2018], and H36M Ionescu et al. [2014] are commonly evaluated using either 18 or 22 joints, represented in Euclidean coordinates or joint angles. Here, J denotes the number of joints, typically 18 or 22, and D represents the number of dimensions, which ranges from 3 to 10 in this paper. Each pose x or y corresponds to a single frame with shape $J \times D$.

3.2 Human Motion Prediction

We have chosen state-of-the-art (SOTA) models in human motion prediction for our evaluation. Some SOTA models in the field encompass MotionMixer Bouazizi et al. [2022], siMLPe Guo et al. [2023], STS-GCN Sofianos et al. [2021], and CIST-GCN Medina et al. [2024]. MotionMixer and siMLPe are both MLP-based models that borrow the idea of Mixer architecture Tolstikhin et al. [2021] and apply it to the domain of human pose forecasting. STS-GCN and CIST-GCN are founded on GCNs. STS-GCN utilizes two successive GCNs to sequentially encode temporal and spatial pose data, while CIST-GCN incorporates specific layers to aid model interpretability. In contrast to other SOTA models, MotionMixer adopts pose displacements as its input representation, whereas siMLPe utilizes DCT encoding for the input pose data.

Protocol	joints format	collision frame	train	validation	test	total	FPS
pro10	18-22-23	0	492052	82028	245909	819989	200
pro14	18-22-23	5	475483	79274	237902	792659	200
pro19	18-22-23	9	410862	68501	205682	685045	200
pro1A	18-22-23	0-9	540160	90051	270053	900264	200
pro40	18-22-23	0	184096	30581	91872	306549	50
pro44	18-22-23	5	177757	29570	88785	296112	50
pro49	18-22-23	9	154949	25792	77346	258087	50
pro4A	18-22-23	0-9	540160	90051	270053	900264	50

Table 3: Comparison of our dataset with the most relevant datasets was performed, focusing on comparable attributes. All units in the table after 'joints format' are expressed in frames.

Protocol	H1-walk	H2-walk	H3-walk	H4-walk	H1-run	H2-run	H3-run	H4-run	Total
pro10	22647	35899	32987	31745	23809	32843	33003	32976	245909
pro14	27165	33850	30304	29164	26347	30608	30680	29784	237902
pro19	28635	28366	24274	23601	25602	25793	25284	24127	205682
pro1A	37032	35991	33017	31753	33265	32988	33009	32998	270053
pro40	1970	9391	14971	22243	1647	6144	13887	21619	91872
pro44	2045	9468	14630	20807	1691	6345	13824	19975	88785
pro49	1900	8880	12599	17367	1579	6194	12091	16736	77346
pro4A	37032	35991	33017	31753	33265	32988	33009	32998	270053

Table 4: Number of samples per class considering different experiment protocols. Samples contain 35 frames (10 inputs and 25 outputs).

3.3 Environment Definition

In our problem formulation, we establish our environment in the MuJoCo engine and adapt LocoMujoco humanoids Al-Hafez et al. [2023] and cars from open-source models H. [2020], Robotics [2017]. These environments allow us to achieve a flexible number of joints (18 or 22) within the same simulation. thereby affording significant flexibility for comparative analysis across multiple datasets. As our simulations involve car impacts, we have access to pertinent car attributes, including position, velocity, and other relevant parameters, similar to those in real-world autonomous driving problems.

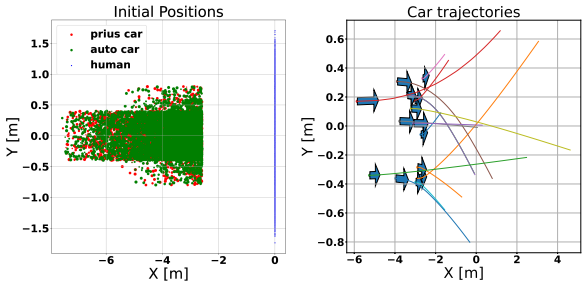
First, we will introduce the car models. We have adapted freely available open-source models to assess the adequacy of physics engines in supporting collisions effectively to acquire the 3D pose sequences. They are the AutoCarRos H. [2020] and the Prius Robotics [2017]. We control the speed of the car directly at the chassis level to mitigate significant fluctuations in collision velocity. Additionally, the front and rear wheels, both tangential speed and orientation, can be configured in the initial settings. In essence, we have control over all pertinent elements of the car to execute forward movements and rotations, thereby simulating real-world conditions. We recorded all these values as joints for ablation studies.

The car rotations are in a range from 55° to -55° for the front wheels while the car speed goes from 5 km/h to 65 km/h. In Fig. 1a, we show the sampling of the car's initial positions, and trajectories. As we can observe, we used uniform distributions for all the parameters we can control. Some challenges we found during the collision generation are described in Section 5.

The humanoids employed in our simulations were trained via Variational Adversarial Imitation Learning (VAIL) Peng et al. [2019] for 200 epochs, with each epoch consisting of 1 million steps. Model updates occurred every 10,000 steps. We used a learning rate of 1×10^{-5} with a maximum Kullback-Leibler (KL) set to 5×10^{-3} and an initial standard deviation set to 0.5. We applied the same configuration for 'walking' and 'running' actions. Following the training phase, we positioned the humanoids within a range of 0 to +/-2 along the X-axis and subsequently moved them perpendicular to the car direction to facilitate car-pedestrian collisions. In Fig. 2, a small set of simulations depicting the trajectory of the humanoid pelvis is showcased. The simulations are delineated into X-Y (Fig. 2a) and X-Z (Fig. 2b) planes, facilitating the observation of parabolic motion patterns after the collisions, with a focus on discerning variations contingent upon the vehicle velocity. Additionally, it can be observed that, following the collision, the humanoids often come to rest on top of the vehicle, remaining in position without detaching or falling until the final frame. This behavior occurs under various conditions, depending on the initial positions, orientations, and speeds.

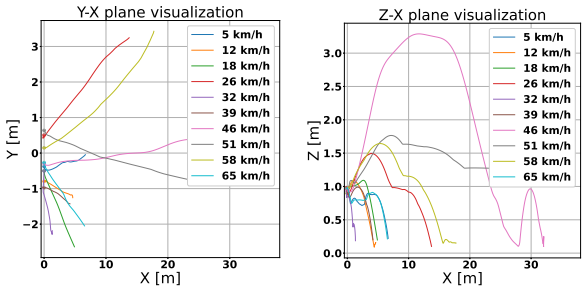
3.4 Implementation Details

We introduced the trained humanoid models and started simulations using seeds to ensure reproducibility. Every simulation generates a json file containing the relevant physics parameters and joints. To use them in a machine learning pipeline, a post-processing step is required to convert them into faster-loading files. Given the versatility of the MuJoCo framework, it is possible to experiment in different components, such as longer periods, period sampling, and different angle collisions between the car and the humanoid. Our experiments found that



(a) Initial positions. (b) Car Trajetories.

Figure 1: Subset of the initial positions and trajectories.



(a) X-Y plane visualization (b) X-Z plane visualization

Figure 2: Simulations illustrating the behavior of the humanoid pelvis joint at varying car velocities.

running them below the $0.5 \times e - 4$ becomes stable for diverse scenarios.

3.5 Dataset Structure

We conducted internal evaluations of various conditions and benchmarks to establish a standardized dataset with a consistent frame rate. Our findings indicate that the optimal simulation time step for MuJoCo is 1×10^{-5} seconds. However, kinematic angles and 3D positions are recorded at intervals of 0.005 seconds (200 FPS). Recording at a higher sampling rate does not yield additional benefits or capture significant movements. Moreover, current models are incapable of predicting movements with a precision finer than 1 millisecond, rendering higher sampling rates unnecessary. We have an overall of 176.7M frames contained in around 1.02M simulations. It is noteworthy that every simulation comprises one car-pedestrian collision only. Yet, training models utilizing sequences devoid of collision events remains feasible.

We established protocols within the PeDesCar database to ensure comparability with benchmark datasets. Table 3 outlines these protocols and specifies the attributes relevant for model training. In particular, we focus on sequences involving car-pedestrian collisions, consisting of 10 input frames and 25 output frames. Collisions are required to occur within the first 10 input frames, as events occurring later could lead to unpredictable subsequent movements. The protocols follow the format pro-collision frameskip frames; for example, pro14 indicates

Figure 3: Visualization of collision simulations (rows). All subjects were represented with both actions, utilizing a single-car model for consistency. (left) Initial frame (right) Last frame

a frame step of 1, with the collision occurring at frame 4 (with input frames indexed from 0). Furthermore, we divided the dataset into the following classes: ‘H1-walk’, ‘H2-walk’, ‘H3-walk’, ‘H4-walk’, ‘H1-run’, ‘H2-run’, ‘H3-run’, ‘H4-run’. These classes represent both the 4 Ages subjects and actions from the LocoMujoco settings, for example, ‘4Ages1-run’ and ‘4Ages3-walk’ are expressed as ‘H1-run’ and ‘H3-walk’ respectively. The number of samples per class for each protocol is detailed in Tab. 4. This table presents samples using only 10 input and 25 output frames, summing a total of 35 frames per sample, a common approach in numerous studies Lyu et al. [2022]. Notably, our dataset, generated using MuJoCo, includes more joints than other benchmark datasets. To ensure comparability, we aligned the joint configurations with those of the H36M and AMASS datasets, using only 22 and 18 joints respectively, and adhered to the typical frame counts used in several studies Lyu et al. [2022]. This alignment allows for the comparison of models trained on one of these datasets and evaluated on our dataset, as shown in Tab. 2. The matched versions of H36M and AMASS are called MujH36m and MujAmass. More details about the dataset are provided in Appendix 9.

A visualization code based on MuJoCo was employed to reproduce each simulation utilizing forward kinematics exclusively, meaning that each simulation step was performed without involving extensive physics computation because the simulations were recorded and replayed. Visualization samples are illustrated in Fig. 3 for qualitative evaluation and elaborated upon in Appendix 10. As demonstrated, it is possible to alter the camera view and zoom, which is similar to the standard MuJoCo engine. For the sake of consistency, we utilized a single car and one of the four distinct subjects.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	15.0	26.7	54.6	68.2	95.1	168.1	11.6	21.1	41.3	50.8	68.2	108.3	11.9	21.6	40.5	48.5	62.4	97.2
MotionMixer Bouazizi et al. [2022]	14.3	28.1	56.1	70.3	99.8	182.7	12.7	23.5	44.0	54.1	74.2	127.7	13.9	26.6	48.1	57.2	74.3	123.1
CIST-GCN Medina et al. [2024]	12.1	28.0	63.2	80.4	114.2	202.7	10.7	3.1	49.6	62.3	85.9	137.6	10.8	22.8	45.3	55.0	72.5	112.0
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	9.5	17.3	31.5	37.4	47.7	77.4	14.8	25.6	52.8	66.7	93.7	165.3	12.7	22.8	44.7	55.2	74.5	118.6
MotionMixer Bouazizi et al. [2022]	11.4	22.4	41.7	49.9	64.2	104.2	13.6	27.1	56.5	72.0	103.8	187.8	13.5	25.1	46.8	57.3	78.2	133.4
CIST-GCN Medina et al. [2024]	8.4	16.0	28.0	32.8	41.2	70.4	11.2	26.0	59.6	76.6	110.8	199.1	11.0	23.8	50.4	62.9	86.2	137.8
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	12.0	21.5	39.7	47.5	61.5	97.0	10.5	18.9	34.9	41.7	54.0	87.1	12.2	21.9	42.5	52.0	69.6	114.9
MotionMixer Bouazizi et al. [2022]	13.4	25.1	45.6	54.5	70.7	115.8	11.8	23.5	44.7	53.9	70.4	115.2	13.1	25.2	47.9	58.6	79.4	136.2
CIST-GCN Medina et al. [2024]	10.2	20.7	39.5	47.4	62.1	97.9	8.7	17.0	30.8	36.6	46.8	80.0	10.4	22.2	45.8	56.7	77.5	129.7

Table 5: Performance comparison using MPJPE across all action-subject cases in the MujH36m dataset under protocol pro1A.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	17.0	29.2	60.7	76.3	105.6	181.4	15.2	24.3	46.9	57.6	77.4	122.0	14.7	25.0	46.9	56.0	72.3	111.5
MotionMixer Bouazizi et al. [2022]	13.2	25.3	49.3	61.6	87.2	162.0	11.3	19.7	35.6	43.7	60.1	101.4	10.5	17.8	30.0	35.3	46.0	79.2
CIST-GCN Medina et al. [2024]	10.7	24.7	54.9	68.8	96.2	174.7	9.7	20.3	41.7	52.0	70.8	115.4	9.5	19.6	38.0	46.1	59.7	91.9
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	12.0	20.5	37.1	43.7	55.7	90.5	16.6	26.8	56.2	71.5	101.1	178.6	15.6	25.2	48.5	59.5	79.9	125.7
MotionMixer Bouazizi et al. [2022]	7.9	13.8	23.3	26.9	33.4	55.4	12.5	24.3	49.1	62.0	89.0	165.5	11.7	20.9	38.5	47.3	65.0	111.0
CIST-GCN Medina et al. [2024]	7.3	13.8	23.9	27.7	34.5	59.2	10.3	23.3	52.3	66.2	93.7	169.7	10.4	21.4	44.1	55.2	75.5	123.2
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	14.4	23.6	42.9	51.2	66.0	103.4	12.6	21.5	39.7	47.3	61.1	98.9	14.8	24.5	47.4	57.9	77.4	126.5
MotionMixer Bouazizi et al. [2022]	10.3	17.4	29.2	34.5	45.1	77.5	8.6	15.6	27.3	32.1	40.6	67.4	10.7	19.4	35.3	42.9	58.3	102.4
CIST-GCN Medina et al. [2024]	9.6	19.0	36.2	43.7	56.9	89.5	8.0	15.2	27.1	32.0	40.7	69.3	9.4	19.7	39.78	49.0	66.0	111.6

Table 6: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro1A.

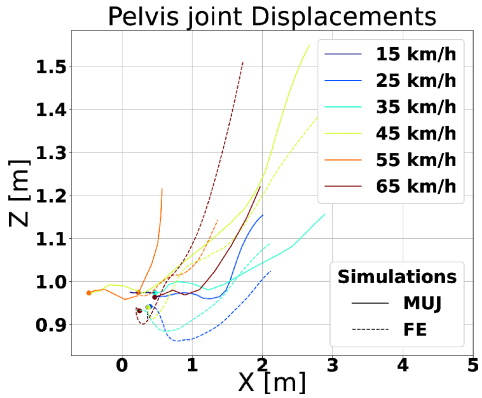


Figure 4: Comparison of displacement comparison focusing on the pelvis joint in the X-Z plane for multiple car speed scenarios in both cases MuJoCo (lines) and FE (dashed lines).

4 EXPERIMENTAL ANALYSIS

4.1 Evaluation and Measurements

We use the metric Mean Per Joint Position Error (MPJPE) as our evaluation metric, which is widely adopted and employed by the SOTA Bouazizi et al. [2022], Guo et al. [2023], Sofianos et al. [2021], Tolstikhin et al. [2021], Medina et al. [2024] to compare two pose sequences and is described in Eq. 1.

$$\mathscr{L}_{\text{MPJPE}} = \frac{1}{J \times T} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\| \hat{x}_{j,t} - x_{j,t} \right\|_2 \qquad (1)$$

For initial experimentation, we employed a pre-trained model, initially trained on the AMASS and H36M datasets, and applied it to our dataset. Additionally, we trained the model using our dataset and tested its performance on the AMASS and H36M datasets. The results of these experiments are presented in Tab. 2. Our empirical analysis revealed that the distribution of our dataset differs from standard benchmarks. Models trained on H36M or AMASS consistently underperformed when evaluated on our dataset. Conversely, models trained on our dataset showed significant improvements in terms of MPJPE when evaluated on the original datasets. Specifically, MotionMixer and STS-GCN exhibited notable improvements, while CIST-GCN maintained consistent performance. This indicates that our dataset is notably complex and challenging, thereby enhancing the gener-

alization capabilities of models when assessed against current benchmarks. We extended our analysis to examine the difficulty levels of various subject and action classes. The results for H36M and AMASS are presented in Tab. 5 and Tab. 6. Our findings indicate that, in terms of MPJPE, larger subjects are easier to predict in motion than smaller subjects (e.g., children). Also, there is no significant difference in performance between the 'walking' and 'running' classes in either table. This short discrepancy is likely attributable to the greater displacements inherent in the 'running' motion. The analysis revealed that the average error rates on the H36M were consistently higher compared to those on the AMASS across various timeframes when using the CIST-GCN and MotionMixer models. However, the STS-GCN model manifested the opposite trend, where the error rates for H36M were lower.

5 DISCUSSION

5.1 Challenges on sequence simulations

Simulations inherently differ from real-world conditions, necessitating a comparison of our simulations. FE simulations are known for their high fidelity but come with significant computational costs Trube et al. [2023], Niranjan et al. [2023]. MuJoCo, on the other hand, offers a reliable and significantly faster alternative, albeit at the expense of fidelity. Additionally, minor discrepancies in skeletal models play a crucial role in influencing the simulation results.

We validated our dataset using two approaches. First, we employed a small dataset of FE simulations involving pedestrian-car interactions. We adjusted the MuJoCo agents to match the positions of FE subjects and compared the initial and final positions using standard metrics. The MPJPE between FE and MuJoCo simulations, under similar initial conditions, was 69.5 and 766.0 for the initial and final frames (with a 10ms framerate). Despite these efforts, minor discrepancies in humanoid shapes resulted in differences, such as higher acceleration peaks, increased cumulative errors, and different simulation lengths. Second, we observed similarities in displacement trajectories, particularly in their gradients. However, differences in acceleration between FE and MuJoCo explained the deviation in the trajectory endpoints. Overall, the similar trends in the trajectories suggest the physical plausibility of MuJoCo simulations, as illustrated in Fig. 4. Following the methodology in Trube et al. [2023], Niranjan et al. [2023], this comparison is sufficient for validating physics simulations, particularly in car-pedestrian collision scenarios. These methods collectively support the reliability of MuJoCo for simulating complex interactions, as demonstrated by its consistent physics profiles with FE simulations.

Other potential issues of our approach are associated with the performance of agents trained using reinforcement learning, the absence of muscular activity, and the approximation of the car CAD models. If the agents do not accurately replicate human walking and running, the physical plausibility of the simulation scenario may be compromised. However, we inspected qualitatively several simulations and removed outliers and broken simulations. Some simulations encountered failures due to glitches or the generation of large and inconsistent movements by the subject. For instance, significant and inconsistent accelerations and force collisions were observed. Another example involved large values along the vertical axis, indicating a substantial jump perpendicular to the floor. This is physically implausible, as the car exerts a horizontal force on the subject after the collision.

5.2 Challenges in simulating Human behavior

It is exciting to recognize the unique complexity of human behavior, which currently exceeds the capabilities of reinforcement learning algorithms. For instance, when a car approaches a person, the individual's instinctive reactions - such as raising their hands for protection or jumping - introduce variables that can alter the dynamics of the simulation. This complexity enriches the scenarios, making them more intricate and offering rich opportunities for future exploration, though it extends beyond the scope of this paper.

5.3 Case Study: Incorporating Car Information

Expanding the scope of the dataset, we included the car joints in our analysis. We conducted additional experiments by incorporating only the car's central location as input to the model, allowing the model to process the car's speed and acceleration. This modification aimed to enhance the model's performance by accounting for car dynamics. We trained the model with this additional input and compared the results, as shown in Tabs. 7 and 8. As observed in Tab. 7, including the car joint reduces the MPJPE metric. On average, this adjustment results in an approximate 700% improvement in MPJPE across all models for both the MujAmass and MujH36m datasets. Additionally, Tab. 8 shows that models incorporating the car joint outperform those that do not. Specifically, CIST-GCN achieved a better MPJPE for short-term movements, while MotionMixer achieved a better MPJPE for long-term movements. This improvement is likely attributed to the model's enhanced capability to determine the rotation of the subject body following a collision, facilitated by the car's directional information. This data is also available in real-world applications and could be leveraged to enhance motion prediction tasks in collision scenarios. Additional experiments using different protocols are detailed in the supplementary materials.

Model	Dataset Train	Car	Time (ms) 80	160	320	400	560	1000	Avg.
STS-GCN Sofianos et al. [2021]	MujAmass		37.0	67.3	109.7	125.8	155.0	206.6	134.0
	MujAmass	X	2.9	5.2	8.0	9.1	11.0	16.2	10.0
	MujH36m		40.9	71.9	118.2	133.7	158.4	209.1	137.9
	MujH36m	X	2.9	4.7	7.4	8.4	10.0	**14.8**	9.2
MotionMixer Bouazizi et al. [2022]	MujAmass		35.1	68.7	109.9	122.3	144.4	190.3	130.7
	MujAmass	X	**2.0**	**4.1**	6.8	7.8	**9.7**	14.9	8.8
	MujH36m		45.2	87.8	140.2	156.6	183.2	237.3	159.2
	MujH36m	X	3.4	6.1	9.6	10.9	13.1	17.8	11.7
CIST-GCN Medina et al. [2024]	MujAmass		35.4	69.0	113.9	130.6	161.1	213.5	138.6
	MujAmass	X	2.2	**4.1**	7.0	8.3	10.8	16.3	9.6
	MujH36m		36.8	70.6	115.7	134.3	165.3	218.8	142.2
	MujH36m	X	**2.0**	3.9	**6.4**	**7.6**	9.8	15.0	**8.8**

Table 7: Evaluation of different models trained on MujAmass and MujH36m including the car joint. The datasets use protocol Pro44. 'X' indicates that the car joint is incorporated into the model. But, it was not used to compute MPJPE. 'Bold' means best MPJPE in each column.

Time (ms)	Car	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN		83.2	161.1	232.6	251.5	282.1	296.5	17.7	30.6	70.1	93.1	138.4	191.7	19.0	30.0	56.8	69.2	92.9	159.6
STS-GCN	X	5.2	9.7	12.6	13.5	15.0	19.0	1.7	2.7	5.2	**6.4**	**8.7**	**14.5**	2.1	3.4	6.0	**7.2**	**9.4**	17.3
MotionMixer		83.5	168.7	250.1	263.9	279.0	298.9	15.8	33.0	70.1	88.8	124.9	169.7	15.1	29.1	53.6	63.6	83.5	137.2
MotionMixer	X	4.0	7.8	11.5	12.3	**13.4**	**17.9**	**1.3**	**2.5**	**5.1**	6.5	9.3	15.8	1.7	3.2	6.3	7.8	10.3	**16.8**
CIST-GCN		84.2	171.7	255.0	266.7	280.7	312.6	16.4	35.2	88.4	121.2	187.4	239.0	15.2	27.3	53.2	67.3	96.6	164.4
CIST-GCN	X	**3.6**	**7.1**	**10.6**	**12.0**	14.0	18.9	**1.3**	2.6	5.6	7.4	11.4	18.1	**1.6**	**3.0**	**5.9**	7.6	10.9	18.5

Time (ms)	Car	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN		23.0	38.8	66.9	78.0	98.1	156.0	87.6	166.4	237.1	252.2	278.3	315.9	19.4	34.8	77.9	101.4	144.1	202.9
STS-GCN	X	3.3	5.5	9.0	10.4	13.0	21.3	5.2	9.4	12.1	13.1	14.8	18.9	1.6	2.7	5.1	6.3	8.4	14.5
MotionMixer		19.5	35.5	56.6	64.2	78.9	129.3	90.0	172.6	243.3	251.7	264.0	304.2	18.3	38.2	81.1	103.2	144.6	121.8
MotionMixer	X	3.0	5.5	9.6	11.2	13.9	20.7	4.0	7.5	10.7	**11.4**	**12.5**	**17.0**	**1.2**	**2.3**	**4.7**	**5.9**	**8.2**	**14.4**
CIST-GCN		20.0	34.6	55.6	63.8	78.0	132.4	91.7	177.8	248.6	261.5	280.0	318.3	17.2	36.6	90.5	121.8	179.5	238.9
CIST-GCN	X	**2.5**	**4.5**	**7.8**	**9.3**	**12.0**	**20.3**	**3.6**	**6.9**	**10.1**	**11.4**	13.4	18.0	**1.2**	2.4	5.1	6.6	9.5	15.8

Time (ms)	Car	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN		20.1	32.6	60.3	72.5	96.1	158.5	25.8	44.4	76.0	88.2	109.9	166.4	37.0	67.3	109.7	125.8	155.0	206.6
STS-GCN	X	2.1	3.5	6.0	7.1	9.1	15.9	3.4	5.8	9.5	10.9	13.5	21.2	3.0	5.3	8.2	9.4	11.5	17.8
MotionMixer		15.8	30.9	56.8	66.7	86.5	135.7	22.6	41.8	67.2	76.1	93.6	144.0	35.1	68.7	109.9	122.3	144.4	190.3
MotionMixer	X	1.6	2.9	5.4	6.4	8.4	**14.0**	3.0	5.4	8.9	10.2	12.5	**11.4**	2.4	4.6	7.8	9.0	**11.1**	**17.0**
CIST-GCN		16.4	29.4	55.5	69.0	96.2	159.2	22.3	39.5	64.2	73.5	90.2	143.6	35.4	69.0	113.9	130.6	161.1	213.5
CIST-GCN	X	**1.5**	**2.8**	**5.0**	**6.1**	**8.2**	14.5	**2.6**	**4.7**	**7.8**	**9.0**	**11.2**	18.6	**2.2**	**4.2**	**7.2**	**8.7**	11.3	17.9

Table 8: Performance comparison for motion prediction using MPJPE across all action-subject cases in the MujAmass dataset. We use the protocol pro44 in these experiments. 'X' indicates that the car joint is incorporated into the model, but it was not used to compute MPJPE.

6 CONCLUSIONS

We introduced the first dataset specifically for Car-Pedestrian Collisions, along with a framework for visualizing simulation samples. This dataset was utilized to train SOTA models, which were subsequently validated using both our dataset and commonly used benchmarks in human motion prediction. Additionally, we have made our dataset, models, and code publicly available under permissive licenses.

While we acknowledge that assessing the reliability of the data presents significant challenges, which may even be insurmountable, the dataset remains adequate for preliminary research purposes. Our framework enables rapid data generation compared to finite element methods. Despite certain factors that may raise concerns about the physical realism of our simulations, we validated the dataset by comparing it with high-fidelity finite element simulations.

By providing this dataset and framework, we aim to foster research in safe automated driving and injury risk prediction. The availability of abundant data also enables the use of deep learning techniques, offering a practical alternative to computationally intensive physics-based algorithms.

7 ACKNOWLEDGEMENT

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "ATTENTION – Artificial Intelligence for realtime injury prediction". The authors appreciate the consortium's successful collaboration and extend special thanks to Niels Heller for his contributions to data management.

8 REFERENCES

F. Al-Hafez, G. Zhao, J. Peters, and D. Tateo. LocoMuJoCo: A Comprehensive Imitation Learning Benchmark for Locomotion. nov 2023. URL http://arxiv.org/abs/2311.02496.

A. Bouazizi, A. Holzbock, U. Kressel, K. Dietmayer, and V. Belagiannis. MotionMixer: MLP-based 3D Human Body Pose Forecasting. *IJCAI International Joint Conference on Artificial Intelligence*, pages 791–798, jul 2022. ISSN 10450823. doi: 10.24963/ijcai.2022/111.

A. Byravan, J. Humplik, L. Hasenclever, A. Brussee, F. Nori, T. Haarnoja, B. Moran, S. Bohez, F. Sadeghi, B. Vujatovic, and N. Heess. NeRF2Real: Sim2real Transfer of Vision-guided Bipedal Motion Skills using Neural Radiance Fields. oct 2022.

Z. Cao, H. Gao, K. Mangalam, Q. Z. Cai, M. Vo, and J. Malik. Long-Term Human Motion Prediction with Scene Context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12346 LNCS:387–404, jul 2020. ISSN 16113349. doi: 10.1007/978-3-030-58452-8_23.

L. Chen, R. Liu, X. Yang, D. Zhou, Q. Zhang, and X. Wei. STTG-net: a Spatio-temporal network for human motion prediction based on transformer and graph convolution network. *Visual Computing for Industry, Biomedicine, and Art*, 5(1):19, dec 2022. ISSN 25244442. doi: 10.1186/s42492-022-00112-5.

L. Dang, Y. Nie, C. Long, Q. Zhang, and G. Li. MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction. *Proceedings of the IEEE International Conference on Computer Vision*, pages 11447–11456, aug 2021. ISSN 15505499. doi: 10.1109/ICCV48922.2021.01127.

J. Fu, F. Yang, Y. Dang, X. Liu, and J. Yin. Learning Constrained Dynamic Correlations in Spatiotemporal Graphs for Motion Prediction. *IEEE Transactions on Neural Networks and Learning Systems*, apr 2023. ISSN 21622388. doi: 10.1109/TNNLS.2023.3277476.

C. Guo, X. Zuo, S. Wang, S. Zou, Q. Sun, A. Deng, M. Gong, and L. Cheng. Action2Motion: Conditioned Generation of 3D Human Motions. In *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, pages 2021–2029, New York, NY, USA, oct 2020. ACM. ISBN 9781450379885. doi: 10.1145/3394171.3413635.

C. Guo, S. Zou, X. Zuo, S. Wang, W. Ji, X. Li, and L. Cheng. Generating Diverse and Natural 3D Human Motions from Text. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, pages 5142–5151, 2022. ISBN 9781665469463. doi: 10.1109/CVPR52688.2022.00509.

W. Guo, Y. Du, X. Shen, V. Lepetit, X. Alameda-Pineda, and F. Moreno-Noguer. Back to MLP: A Simple Baseline for Human Motion Prediction. *Proceedings - 2023 IEEE Winter Conference on Applications of Computer Vision, WACV 2023*, pages 4798–4808, jul 2023. doi: 10.1109/WACV56688.2023.00479.

W. H. AutoCarROS. https://github.com/winstxnhdw/AutoCarROS, 2020.

T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, J. Humplik, M. Wulfmeier, S. Tunyasuvunakool, N. Y. Siegel, R. Hafner, M. Bloesch, K. Hartikainen, A. Byravan, L. Hasenclever, Y. Tassa, F. Sadeghi, N. Batchelor, F. Casarini, S. Saliceti, C. Game, N. Sreendra, K. Patel, M. Gwira, A. Huber, N. Hurley, F. Nori, R. Hadsell, and N. Heess. Learning Agile Soccer Skills for a Bipedal Robot with Deep Reinforcement Learning. apr 2023. doi: 10.1126/scirobotics.adi8022.

C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, jul 2014. ISSN 01628828. doi: 10.1109/TPAMI.2013.248.

C. Li, Z. Zhang, W. S. Lee, and G. H. Lee. Convolutional Sequence to Sequence Model for Human Dynamics. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 5226–5234, may 2018. ISSN 10636919. doi: 10.1109/CVPR.2018.00548.

M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL. *ACM Transactions on Graphics*, 34(6):1–16, nov 2015. ISSN 0730-0301. doi: 10.1145/2816795.2818013.

K. Lyu, H. Chen, Z. Liu, B. Zhang, and R. Wang. 3D human motion prediction: A survey. volume 489, pages 345–365, mar 2022. doi: 10.1016/j.neucom.2022.02.045.

T. Ma, Y. Nie, C. Long, Q. Zhang, and G. Li. Progressively Generating Better Initial Guesses Towards Next Stages for High-Quality Human Motion Prediction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June:6427–6436, mar 2022. ISSN 10636919. doi: 10.1109/CVPR52688.2022.00633.

N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. Black. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-Octob, pages 5441–5450. IEEE, oct 2019. ISBN 9781728148038. doi: 10.1109/ICCV.2019.00554.

W. Mao, M. Liu, M. Salzmann, and H. Li. Multi-level Motion Attention for Human Motion Prediction. *International Journal of Computer Vision*, 129(9):2513–2535, sep 2021. ISSN 15731405. doi: 10.1007/s11263-021-01483-7.

E. Medina, L. Loh, N. Gurung, K. H. Oh, and N. Heller. Context-Based Interpretable Spatio-Temporal Graph Convolutional Network for Human Motion Forecasting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3232–3241. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024.

D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt. Monocular 3D human pose estimation in the wild using improved CNN supervision. *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, pages 506–516, nov 2018. doi: 10.1109/3DV.2017.00064.

Mujoco. Mujoco Environments. URL https://www.gymlibrary.dev/environments/mujoco/.

A. A. Nargund and M. Sra. SPOTR: Spatio-temporal Pose

Transformers for Human Motion Prediction. mar 2023.

B. Niranjan, S. Thomas, D. Michael, Niclas, Trube, F. D. Dirk, Gesellschaft, F. Ingenieurdienstleistungen, and mbH Germany. A Method For Efficient Generation And Optimization Of Simulation-Based Training Data For Data-Driven Injury Prediction In Vru-Vehicle Accident Scenarios. 2023.

X. B. Peng, A. Kanazawa, S. Toyer, P. Abbeel, and S. Levine. Variational discriminator bottleneck: Improving imitation learning, inverse RL, and GANs by constraining information flow. *7th International Conference on Learning Representations, ICLR 2019*, oct 2019.

M. Plappert, C. Mandery, and T. Asfour. The KIT Motion-Language Dataset. *Big Data*, 4(4):236–252, jul 2016. ISSN 2167647X. doi: 10.1089/big.2016.0028.

O. Robotics. Prius Car. https://github.com/osrf/car_demo, 2017.

T. Sofianos, A. Sampieri, L. Franco, and F. Galasso. Space-Time-Separable Graph Convolutional Network for Pose Forecasting. *Proceedings of the IEEE International Conference on Computer Vision*, pages 11189–11198, oct 2021. ISSN 15505499. doi: 10.1109/ICCV48922.2021.01102.

I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy. MLP-Mixer: An all-MLP Architecture for Vision. *Advances in Neural Information Processing Systems*, 29:24261–24272, may 2021. ISSN 10495258.

N. Trube, P. Matt, M. Jenerowicz, N. Ballal, T. Soot, D. Fressmann, N. Lazarov, J. Moennich, T. Lich, P. Lerge, and Others. Plausibility Assessment of Numerical Cyclist to Vehicle Collision Simulations based on Accident Data. In *Proceedings of the IRCOBI Conference*, pages 113–135, 2023.

T. von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll. Recovering accurate 3D human pose in the wild using IMUs and a moving camera. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11214 LNCS, pages 614–631. 2018. ISBN 9783030012489. doi: 10.1007/978-3-030-01249-6_37.

N. Wagener, A. Kolobov, F. V. Frujeri, R. Loynd, C.-A. Cheng, and M. Hausknecht. {MoCapAct}: A Multi-Task Dataset for Simulated Humanoid Control. In *Advances in Neural Information Processing Systems*, volume 35, pages 35418–35431, 2022.

H. Wang, V. Caggiano, G. Durandau, M. Sartori, and V. Kumar. MyoSim: Fast and physiologically realistic MuJoCo models for musculoskeletal and exoskeletal studies. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8104–8111. IEEE, may 2022. ISBN 978-1-7281-9681-7. doi: 10.1109/ICRA46639.2022.9811684.

D. Wdowicz and M. Ptak. Numerical Approaches to Pedestrian Impact Simulation with Human Body Models: A Review. *Archives of Computational Methods in Engineering*, 30(8):4687–4709, jun 2023. ISSN 18861784. doi: 10.1007/s11831-023-09949-2.

G. Wen, B. Xiaoyu, A.-P. Xavier, and M.-N. Francesc. Multi-Person Extreme Motion Prediction. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Q. Xu, W. Mao, J. Gong, C. Xu, S. Chen, W. Xie, Y. Zhang, and Y. Wang. Joint-Relation Transformer for Multi-Person Motion Prediction. *Proceedings of the IEEE International Conference on Computer Vision*, pages 9782–9792, aug 2023. ISSN 15505499. doi: 10.1109/ICCV51070.2023.00900.

X. Zhang, B. L. Bhatnagar, S. Starke, I. Petrov, V. Guzov, H. Dhamo, E. Pérez-Pellitero, and G. Pons-Moll. FORCE: Dataset and Method for Intuitive Physics Guided Human-object Interaction. mar 2024.

Z. Zhang, C. Wang, W. Qiu, W. Qin, and W. Zeng. AdaFuse: Adaptive Multiview Fusion for Accurate Human Pose Estimation in the Wild. *International Journal of Computer Vision*, 129(3):703–718, oct 2021. ISSN 15731405. doi: 10.1007/s11263-020-01398-9.

W. Zhao, J. P. Queralta, and T. Westerlund. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. sep 2020. doi: 10.1109/SSCI47803.2020.9308468.

C. Zhong, L. Hu, Z. Zhang, Y. Ye, and S. Xia. Spatio-Temporal Gating-Adjacency GCN for Human Motion Prediction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June:6437–6446, mar 2022. ISSN 10636919. doi: 10.1109/CVPR52688.2022.00634.

L. Zhu, K. Rematas, B. Curless, S. M. Seitz, and I. Kemelmacher-Shlizerman. Reconstructing NBA Players. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12350 LNCS:177–194, jul 2020. ISSN 16113349. doi: 10.1007/978-3-030-58558-7_11.

ISSN 2464-4617 (print)
ISSN 2464-4625 (online)
Computer Science Research Notes - CSRN
http://www.wscg.eu
WSCG 2025 Proceedings

PeDesCar: A Large-Scale Dataset for Simulated Car-Pedestrian Collisions to Advance Safety Research

Supplementary Material

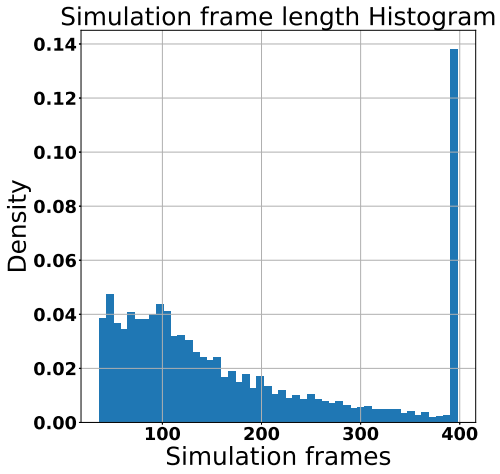


Figure 5: Histogram of the simulation lengths.

9 DATASET STRUCTURE

9.1 Simulation lengths

Other statistics from the PeDesCar dataset can be explained by the number of samples and the simulation lengths. Fig. 5 illustrates the density distribution of simulation durations, with the majority of simulations concluding within around 400 frames. This analysis excludes simulations in which the pedestrian either did not collide with the vehicle or remained standing on top of the car for extended periods, as these were capped at a maximum duration of 4 seconds. In other words, excluding the largest bin in the histogram, the dataset exhibits a diverse range of simulation lengths, reflecting various initial conditions. Longer simulations were restricted due to practical considerations, including computational limitations and concerns related to feasibility and environmental impact. Simulations where pedestrian-car collisions did not occur, as well as those extending longer times beyond the collision event, fall outside the scope of our study and were therefore excluded from the analysis. Further analysis could be conducted to perform binary classification of collision occurrences ('yes' and 'no'). This would involve assessing whether collisions occur within specific sequences. However, this analysis is beyond the scope of the current study.

9.2 Metrics in other protocols

To demonstrate comparability across protocols, we conducted a diverse set of experiments to complement the main results. These metrics may provide researchers with the ability to select the most relevant experiments for their specific needs and perform the comparison.

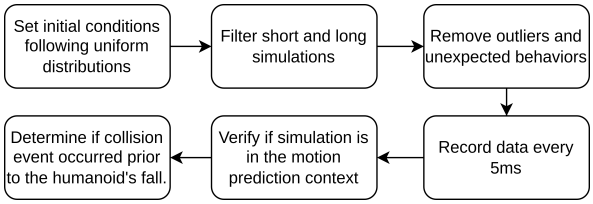


Figure 6: Dataset generation workflow.

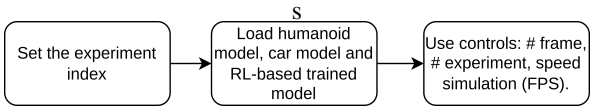


Figure 7: Histogram of the simulation lengths.

Specifically, the 3 models we selected for this paper were applied for the protocols pro4A (Tab. 9), pro40 (Tab. 10), pro44 (Tab. 11), pro49 (Tab. 12), pro10 (Tab. 13), pro19 (Tab. 14), pro14 (Tab. 15). Additionally, we extended the experiments to include the car joint, and we present the in Tab. 16. For the sake of simplicity, we used only MotionMixer Bouazizi et al. [2022] for these experiments.

10 PROCESS WORKFLOW

In Fig. 6, we illustrate the generation workflow for sequence generations. Initially, the workflow sets the physics parameters required for generating a sequence. If the sequence exceeds 10,000 frames, it is truncated and filtered. Subsequently, we examine the sequence for any anomalies, such as excessive acceleration peaks, high force impacts, or violations of physics constraints. Following this, a subsampling is applied to the sequence, with timestamps recorded every 5 milliseconds. We then verify whether the sequence meets the minimum operational criteria for other benchmarks. Specifically, the sequence must contain at least 35 frames, and a collision must occur within the first 10 frames. Lastly, we assess whether the collision occurred before the humanoid falls or during the humanoid's descent towards the floor. This step ensures that the collision reflects more realistic 'walking' and 'running' scenarios.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	19.1	29.9	58.8	72.7	99.5	173.3	15.0	24.0	46.6	57.1	76.6	121.4	14.5	24.9	46.6	55.8	72.1	111.8
MotionMixer Bouazizi et al. [2022]	12.1	23.9	47.5	59.9	86.2	163.5	10.2	18.7	35.2	43.8	61.0	104.4	10.1	18.3	31.4	36.9	47.5	82.7
CIST-GCN Medina et al. [2024]	14.3	23.9	45.4	58.5	78.4	120.1	10.3	20.9	42.4	52.5	71.4	116.1	10.1	20.5	39.5	47.6	61.4	95.6
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	11.9	20.4	37.0	43.6	55.3	90.2	18.4	28.0	56.0	70.0	97.3	171.2	15.3	24.8	47.8	58.5	78.1	123.4
MotionMixer Bouazizi et al. [2022]	7.9	14.6	25.3	29.3	36.1	60.0	11.8	23.7	48.3	61.4	89.1	168.8	11.0	20.3	38.2	47.3	65.6	113.6
CIST-GCN Medina et al. [2024]	11.7	17.5	27.2	33.2	40.3	69.5	14.6	16.2	23.0	33.0	37.0	69.7	13.3	12.1	22.1	31.4	38.5	67.2
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	14.2	23.6	43.0	51.2	65.8	103.1	12.6	21.5	39.8	47.4	61.1	99.0	15.1	24.7	46.9	57.0	75.7	124.2
MotionMixer Bouazizi et al. [2022]	10.1	17.9	30.4	36.0	46.7	80.8	8.6	16.4	29.4	34.6	43.6	72.1	10.2	19.2	35.7	43.6	59.5	105.7
CIST-GCN Medina et al. [2024]	11.0	22.3	45.1	56.0	76.0	123.5	10.1	19.7	37.3	45.3	59.4	53.9	12.15	19.2	35.25	44.7	57.8	89.4

Table 9: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro4A.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	95.4	173.8	240.9	259.7	283.0	302.2	17.9	28.9	64.9	87.0	126.3	174.1	18.0	28.5	54.3	66.5	89.8	156.8
MotionMixer Bouazizi et al. [2022]	82.2	167.9	256.8	269.8	283.8	300.3	14.4	31.3	72.7	94.4	127.8	177.7	13.6	27.4	54.5	65.4	86.5	144.5
CIST-GCN Medina et al. [2024]	92.1	169.5	233.7	255.3	279.9	301.9	15.9	34.9	90.1	123.5	185.8	230.1	15.2	27.4	55.1	70.7	102.5	172.3
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	22.3	37.5	65.4	77.0	98.0	155.9	104.4	180.1	244.3	262.5	288.7	319.6	19.4	32.6	71.8	93.3	130.4	196.5
MotionMixer Bouazizi et al. [2022]	18.2	33.8	56.3	64.3	80.0	133.1	87.5	173.1	252.6	262.9	275.4	309.4	16.5	35.4	78.8	100.5	135.2	202.5
CIST-GCN Medina et al. [2024]	19.6	33.4	54.8	63.4	78.9	134.8	100.4	180.2	243.7	263.5	286.2	317.2	17.4	36.6	91.2	121.3	174.7	234.7
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	19.1	31.1	58.9	71.3	94.4	156.8	24.9	42.9	74.8	88.1	111.1	167.9	40.2	69.4	109.4	125.7	152.7	203.7
MotionMixer Bouazizi et al. [2022]	14.1	28.5	56.4	67.8	88.9	144.2	20.7	39.7	67.0	76.8	95.2	148.1	33.4	67.1	111.9	125.2	146.6	195.0
CIST-GCN Medina et al. [2024]	16.0	29.0	55.5	69.8	97.8	161.9	21.9	38.8	63.6	73.7	91.8	147.1	37.3	68.7	111.0	130.1	162.2	212.5

Table 10: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro40.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	83.2	161.1	232.6	251.5	282.1	296.5	17.7	30.6	70.1	93.1	138.4	191.7	19.0	30.0	56.8	69.2	92.9	159.6
MotionMixer Bouazizi et al. [2022]	83.5	168.7	250.1	263.9	279.0	298.9	15.8	33.0	70.1	88.8	124.9	169.7	15.1	29.1	53.6	63.6	83.5	137.2
CIST-GCN Medina et al. [2024]	84.2	171.7	255.0	266.7	280.7	312.6	16.4	35.2	88.4	121.2	187.4	239.0	15.2	27.3	53.2	67.3	96.6	164.4
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	23.0	38.8	66.9	78.0	98.1	156.0	87.6	166.4	237.1	252.2	278.3	315.9	19.4	34.8	77.9	101.4	144.1	208.0
MotionMixer Bouazizi et al. [2022]	19.5	35.5	56.6	64.2	78.9	129.3	90.0	172.6	243.3	251.7	264.0	304.2	18.3	38.2	81.1	103.2	144.6	203.8
CIST-GCN Medina et al. [2024]	20.0	34.6	55.6	63.8	78.0	132.4	91.7	177.8	248.6	261.5	280.0	318.7	17.2	36.6	90.5	121.8	179.5	238.9
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	20.1	32.6	60.3	72.5	96.1	158.5	25.8	44.4	76.0	88.2	109.9	166.4	37.0	67.3	109.7	125.8	155.0	206.6
MotionMixer Bouazizi et al. [2022]	15.8	30.9	56.8	66.7	86.5	135.7	22.6	41.8	67.2	76.1	93.6	144.0	35.1	68.7	109.9	122.3	144.4	190.3
CIST-GCN Medina et al. [2024]	16.4	29.4	55.5	69.0	96.2	159.2	22.3	39.5	64.2	73.5	90.2	143.6	35.4	69.0	113.9	130.6	161.1	213.5

Table 11: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro44.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	73.8	157.8	235.2	251.0	274.3	300.2	19.8	32.5	68.5	85.4	120.8	181.7	20.6	32.1	55.7	65.4	84.9	150.2
MotionMixer Bouazizi et al. [2022]	87.8	168.4	231.3	239.4	266.8	305.5	19.2	37.6	72.6	88.1	120.1	171.8	18.1	32.3	54.7	64.0	81.6	132.0
CIST-GCN Medina et al. [2024]	75.9	142.7	212.7	224.9	245.6	288.6	15.9	32.7	72.7	94.9	145.6	208.6	16.8	28.8	50.2	59.9	80.5	144.0
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	24.8	41.1	67.9	77.8	95.7	153.2	79.7	160.0	230.4	246.7	274.5	310.8	21.8	36.0	74.7	94.2	131.0	199.3
MotionMixer Bouazizi et al. [2022]	21.8	37.0	54.8	61.7	74.9	124.5	89.8	169.0	231.3	240.8	257.0	300.5	22.3	43.4	86.6	108.6	151.4	218.3
CIST-GCN Medina et al. [2024]	21.5	37.1	57.8	64.9	78.1	129.8	77.2	164.1	211.8	242.1	254.4	309.4	17.6	35.1	75.8	99.3	148.0	217.3
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	22.0	35.1	60.6	70.7	90.1	152.1	28.4	47.5	78.0	89.2	109.4	164.6	36.4	67.8	108.9	122.5	147.6	201.5
MotionMixer Bouazizi et al. [2022]	18.7	33.3	57.0	66.8	84.7	133.4	26.0	44.5	66.2	74.5	90.2	141.5	38.0	70.7	106.8	118.0	140.8	190.9
CIST-GCN Medina et al. [2024]	18.2	31.4	54.7	64.6	85.5	145.6	24.5	42.5	67.4	75.5	91.1	143.3	33.45	64.3	100.4	115.8	141.1	198.3

Table 12: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro49.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	11.1	15.6	32.0	42.1	64.7	134.7	8.8	13.5	26.0	32.8	46.9	86.3	9.5	15.5	28.9	35.3	47.0	78.4
MotionMixer Bouazizi et al. [2022]	5.9	13.3	32.6	43.7	68.0	139.2	5.6	11.7	25.8	33.2	48.6	87.6	6.5	12.2	23.5	28.8	39.1	68.7
CIST-GCN Medina et al. [2024]	6.2	13.7	32.3	43.3	67.5	138.9	5.3	11.5	25.6	33.1	48.7	92.0	6.1	12.3	24.4	30.0	40.6	72.1
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	8.8	14.9	28.0	33.6	43.8	71.9	11.2	15.9	32.5	42.6	64.7	133.2	9.4	14.3	28.0	35.4	50.5	92.3
MotionMixer Bouazizi et al. [2022]	5.9	11.0	19.8	23.3	29.5	50.0	5.9	13.6	33.9	45.6	70.8	142.6	5.8	12.1	27.1	35.1	51.5	95.0
CIST-GCN Medina et al. [2024]	6.0	11.6	20.8	24.6	31.4	55.3	6.3	13.8	32.5	43.5	66.9	137.1	5.6	12.1	27.2	35.1	51.2	95.3
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	9.7	15.6	29.3	35.7	47.8	80.7	9.6	16.2	31.0	37.8	50.0	81.9	9.8	15.2	29.5	36.9	51.9	94.9
MotionMixer Bouazizi et al. [2022]	6.5	12.1	23.2	28.2	38.1	67.7	6.5	12.7	23.9	28.4	36.6	62.0	6.1	12.3	26.2	33.3	47.8	89.1
CIST-GCN Medina et al. [2024]	6.3	12.5	25.0	30.6	41.5	78.9	6.5	12.8	24.1	28.8	37.3	65.1	6.0	12.5	26.5	33.6	48.1	91.9

Table 13: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro10.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	13.4	21.9	42.6	53.7	76.4	139.7	12.2	21.9	41.5	50.7	68.1	108.9	11.8	22.5	42.0	50.2	64.6	98.9
MotionMixer Bouazizi et al. [2022]	9.2	18.3	38.9	50.5	75.3	145.8	10.7	19.1	35.3	43.3	59.3	98.0	9.9	17.3	30.1	35.5	46.1	77.1
CIST-GCN Medina et al. [2024]	13.1	24.7	45.8	57.8	77.3	155.0	10.1	20.4	41.4	51.7	70.1	114.0	9.7	19.7	38.7	47.2	60.9	94.1
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	9.4	17.2	31.3	37.2	47.6	77.4	13.7	22.3	42.9	54.0	76.6	139.4	13.0	23.3	44.7	54.9	74.1	118.3
MotionMixer Bouazizi et al. [2022]	7.5	13.4	22.9	26.6	33.1	55.0	9.8	19.4	41.1	52.9	77.9	149.8	11.4	20.9	39.1	48.0	65.6	109.5
CIST-GCN Medina et al. [2024]	7.2	13.6	23.3	27.2	33.9	58.5	14.4	23.9	46.1	59.5	78.5	150.1	12.1	26.1	48.0	47.5	65.3	105.5
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	12.2	21.9	40.1	47.8	61.8	97.3	10.5	18.8	34.5	41.2	53.3	86.5	12.0	21.2	40.0	48.7	65.3	108.3
MotionMixer Bouazizi et al. [2022]	9.8	17.0	29.1	34.5	45.0	75.9	8.2	15.2	26.7	31.4	40.0	66.7	9.6	17.6	32.9	40.3	55.3	97.2
CIST-GCN Medina et al. [2024]	9.5	18.6	35.7	43.5	56.5	89.7	7.9	15.0	26.5	31.2	39.5	67.7	10.5	20.3	38.2	45.7	60.3	104.3

Table 14: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro19.

Time (ms)	H1-walk 80	160	320	400	560	1000	H2-walk 80	160	320	400	560	1000	H3-walk 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	39.1	83.9	162.6	198.2	262.2	386.2	30.9	69.4	127.4	149.2	183.3	242.2	20.9	46.9	90.9	108.1	135.4	186.6
MotionMixer Bouazizi et al. [2022]	31.3	68.2	123.2	146.8	190.7	289.2	22.3	45.9	75.6	86.0	103.3	145.8	14.5	28.8	50.1	58.4	72.1	108.2
CIST-GCN Medina et al. [2024]	35.1	77.9	155.5	178.9	244.3	369.9	27.5	69.7	115.3	166.3	167.3	244.1	17.4	44.1	89.0	106.4	135.1	186.4
Time (ms)	H4-walk 80	160	320	400	560	1000	H1-run 80	160	320	400	560	1000	H2-run 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	15.4	29.3	56.6	69.1	90.8	141.3	39.0	83.5	163.4	198.3	259.1	373.2	30.7	67.1	122.5	143.8	178.4	241.7
MotionMixer Bouazizi et al. [2022]	11.4	23.2	43.9	53.1	68.4	108.5	31.1	68.4	124.3	146.9	186.4	273.5	22.8	49.3	83.4	95.8	116.7	167.1
CIST-GCN Medina et al. [2024]	14.3	26.3	52.4	64.6	88.9	135.8	34.7	80.1	145.9	190.6	243.5	366.9	26.5	61.3	121.2	145.8	170.1	230.9
Time (ms)	H3-run 80	160	320	400	560	1000	H4-run 80	160	320	400	560	1000	Average 80	160	320	400	560	1000
STS-GCN Sofianos et al. [2021]	23.2	45.5	82.8	97.9	123.4	178.6	17.5	31.9	59.8	72.1	93.5	144.2	27.1	57.2	108.3	129.6	165.8	236.8
MotionMixer Bouazizi et al. [2022]	16.4	30.7	51.0	58.9	72.8	111.3	13.1	26.1	49.1	59.0	75.9	119.6	20.4	42.5	75.1	88.1	110.8	165.4
CIST-GCN Medina et al. [2024]	19.3	42.8	81.3	96.5	122.9	175.9	13.2	27.0	51.6	62.3	80.9	128.4	23.5	53.6	101.5	126.4	156.6	229.8

Table 15: Performance comparison using MPJPE across all action-subject cases in the MujAmass dataset under protocol pro14.

Protocol	Dataset Train	Car	Time (ms) 80	160	320	400	560	1000	Avg.
Pro40	MujAmass		33.4	67.1	111.9	125.2	146.6	195.0	127.7
	MujAmass	X	2.3	4.3	7.3	8.4	10.3	15.3	9.4
	MujH36m		43.3	84.1	138.0	156.3	182.9	239.3	158.5
	MujH36m	X	3.2	5.8	9.4	10.8	12.8	17.7	11.5
Pro44	MujAmass		35.1	68.7	109.9	122.3	144.4	190.3	130.7
	MujAmass	X	2.0	4.1	6.8	7.8	9.7	14.9	8.8
	MujH36m		45.2	87.8	140.2	156.6	183.2	237.3	159.2
	MujH36m	X	3.4	6.1	9.6	10.9	13.1	17.8	11.7
Pro49	MujAmass		38.0	70.7	106.8	118.0	140.8	190.9	124.9
	MujAmass	X	2.5	4.7	7.6	8.6	10.9	15.7	9.8
	MujH36m		45.5	88.6	140.0	155.6	182.6	236.5	158.7
	MujH36m	X	3.3	6.0	9.7	11.0	13.4	18.5	12.0
Pro4A	MujAmass		9.1	17.3	31.5	37.7	48.8	73.4	43.3
	MujAmass	X	8.8	16.8	30.4	36.3	46.8	69.5	41.8
	MujH36m		13.1	25.2	47.9	58.6	79.4	136.2	73.2
	MujH36m	X	8.0	15.1	27.3	32.5	41.8	61.6	37.3
Pro10	MujAmass		6.1	12.3	26.2	33.3	47.8	89.1	44.8
	MujAmass	X	1.3	2.6	5.1	6.2	8.4	14.8	7.8
	MujH36m		9.4	18.3	36.8	46.5	66.1	119.6	61.6
	MujH36m	X	1.8	3.4	6.8	8.5	11.7	20.7	10.9
Pro14	MujAmass		9.6	17.6	32.9	40.3	55.3	97.2	51.4
	MujAmass	X	1.9	3.7	6.6	7.9	10.5	18.0	9.8
	MujH36m		12.9	24.7	46.9	57.5	78.0	132.4	71.6
	MujH36m	X	2.7	5.2	9.7	11.7	15.4	25.4	14.1
Pro19	MujAmass		20.4	42.5	75.1	88.1	110.8	165.4	98.3
	MujAmass	X	4.0	7.4	12.5	14.6	18.5	28.7	16.9
	MujH36m		26.9	56.6	97.5	114.1	143.3	209.1	126.5
	MujH36m	X	5.1	10.0	17.1	20.0	25.3	38.2	22.9
Pro1A	MujAmass		10.7	19.4	35.3	42.9	58.3	102.4	54.4
	MujAmass	X	8.8	16.8	30.4	36.3	46.8	69.5	41.8
	MujH36m		13.1	25.2	47.9	58.6	79.4	136.2	73.2
	MujH36m	X	8.0	15.1	27.3	32.5	41.8	61.6	37.3

Table 16: Evaluation of MotionMixer Bouazizi et al. [2022] models trained on MujAmass and MujH36m including the car joint. 'X' indicates that the car joint is incorporated into the model. But it was not used to compute the MPJPE metric.