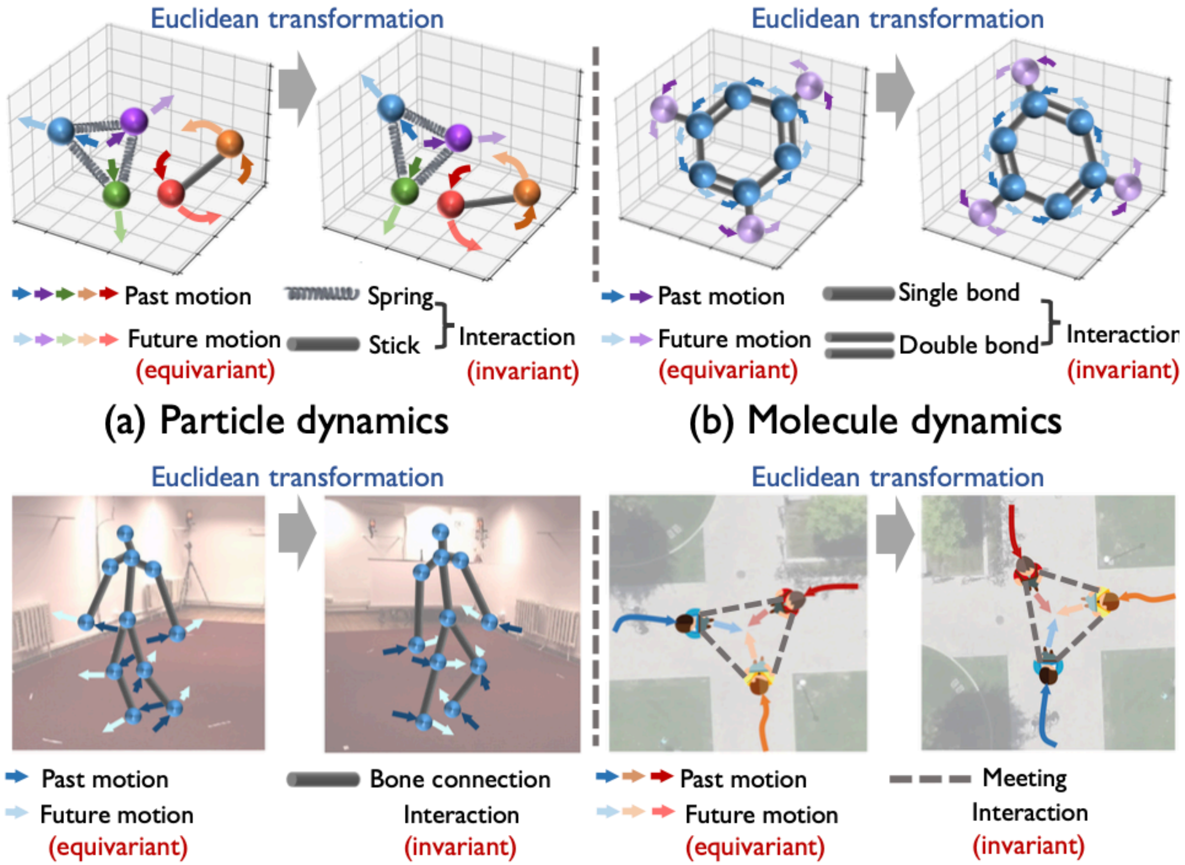
EqMotion: Equivariant Multi-agent Motion Prediction with Invariant Interaction Reasoning

Chenxin Xu[1,2], Robby T. Tan[2], Yuhong Tan[1], Siheng Chen[1,3], Yu Guang Wang[1] , Xinchao Wang[2] , Yanfeng Wang[3,1]

[1]Shanghai Jiao Tong University, [2]National University of Singapore, [3]Shanghai AI Laboratory

Summary

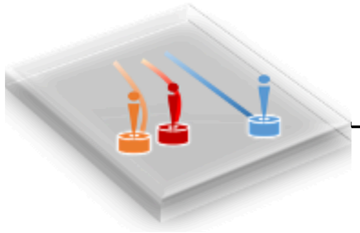
- We propose EqMotion, the first motion prediction model that theoretically ensures sequence-to-sequence motion equivariance.
- We propose a novel invariant interaction reasoning module, in which the captured interactions between agents are invariant to the input motion.
- We conduct experiments on four types of scenarios and find that EqMotion is applicabl to all these different tasks, and importantly outperforms existing state-of-the-art methods on all the tasks.



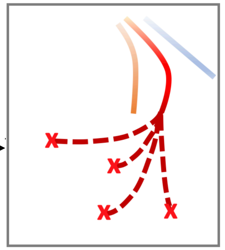
Introduction

- Multi-Agent Trajectory Prediction

Given the past trajectories, predict the future trajectories for multiple interactive agents



Predictor



$$\mathbb{X}_{\mathcal{N}} = [\mathbf{X}_{\mathcal{N}_1}, \mathbf{X}_{\mathcal{N}_2}, \cdots, \mathbf{X}_{\mathcal{N}_N}]$$

Previous/Observed Trajectories

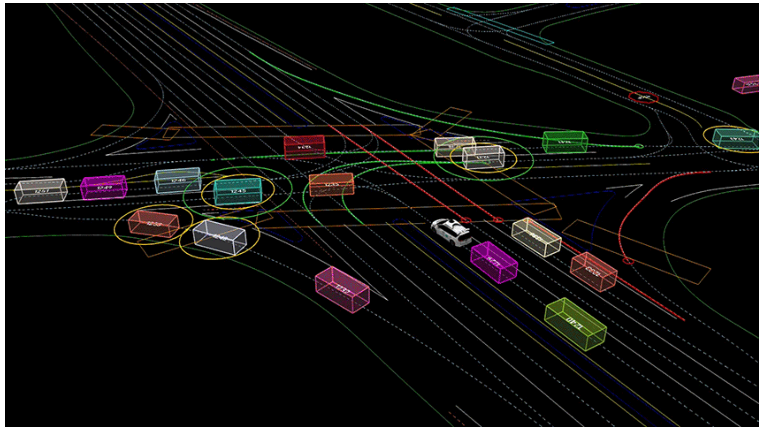
$$\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^{T_\mathrm{f}}]$$

Future/Predicted Trajectories

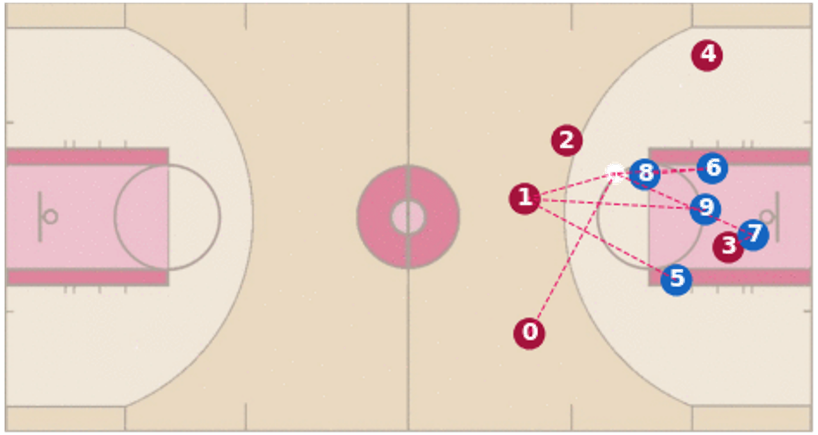
Introduction

- Multi-Agent Trajectory Prediction

Given the past trajectories, predict the future trajectories for multiple interactive agents



Autonomous driving



Sports

Tracking and Surveillance

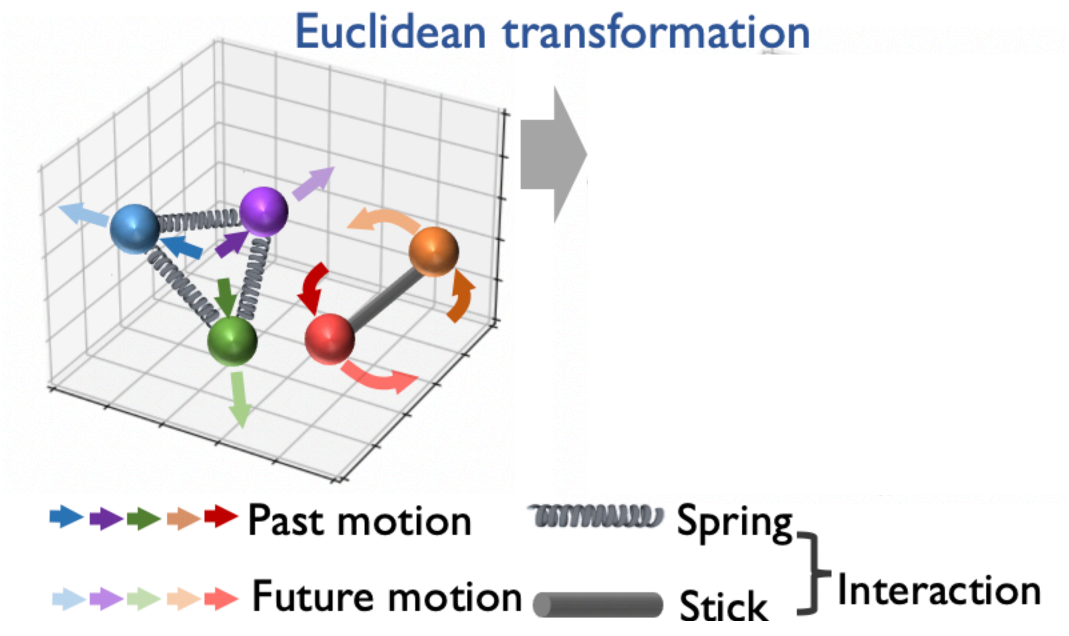
Motivation

- An often-overlooked yet fundamental principle
  - Predicted trajectory — Equivariant to Euclidean transformations
  - Inferred Relationship — Invariant to Euclidean transformations

Motivation

- An often-overlooked yet fundamental principle
  - Predicted trajectory — Equivariant to Euclidean transformations
  - Inferred Relationship — Invariant to Euclidean transformations

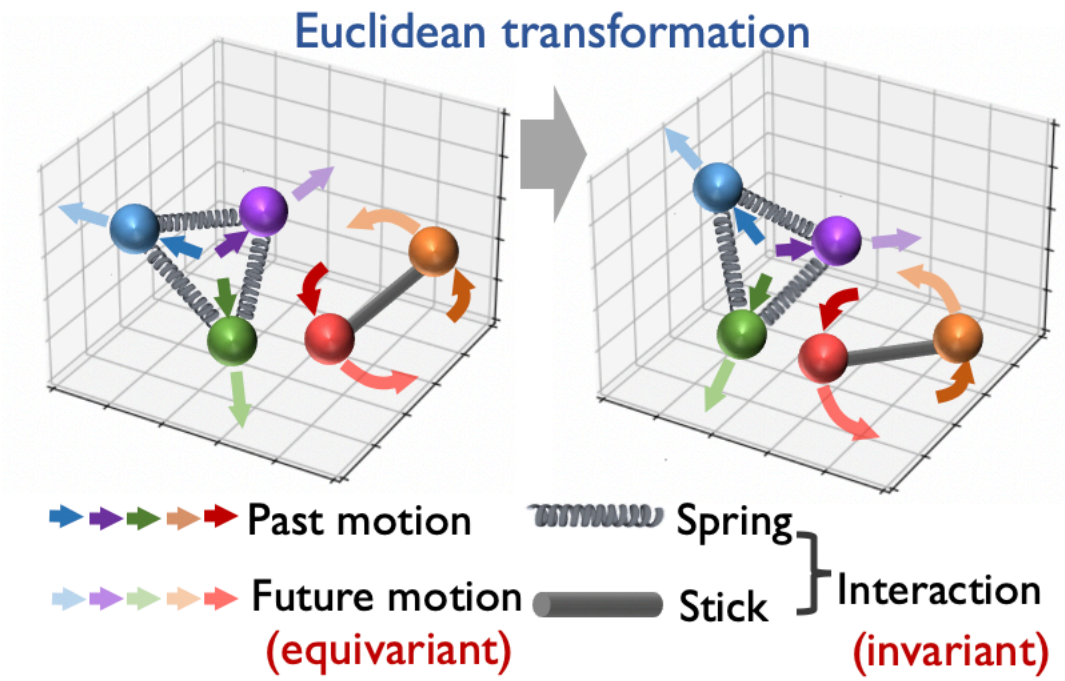
Particle dynamics



Motivation

- An often-overlooked yet fundamental principle
  - Predicted trajectory — Equivariant to Euclidean transformations
  - Inferred Relationship — Invariant to Euclidean transformations

Particle dynamics



Motivation

- How to employing this principle into a network ?

Previous methods — Normalization or data augmentation

- Unable to guarantee the equivariance property
- Bringing more learning burden

Motivation

- How to employing this principle into a network ?

  Previous methods — Normalization or data augmentation

  - Unable to guarantee the equivariance property
  - Bringing more learning burden

  Our EqMotion — Embed this principle directly into the network structure!

Motivation

- How to employing this principle into a network ?

  Previous methods — Normalization or data augmentation

  - Unable to guarantee the equivariance property
  - Bringing more learning burden

  Our EqMotion — Embed this principle directly into the network structure!

  - Theoretically robust to arbitrary Euclidean transformations
  - Reducing the network's learning burden

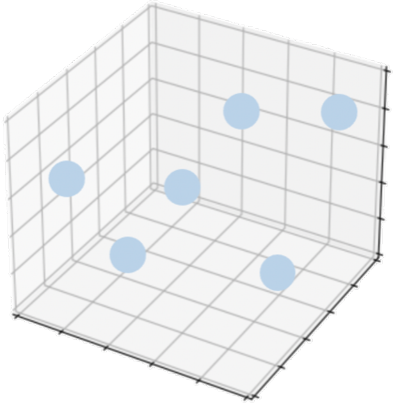
Methodology - EqMotion

- Feature Description

Traditional vector feature
-> cannot be operated by Euclidean transformation
-> network cannot maintain equivariance

Methodology - EqMotion

- Feature Description



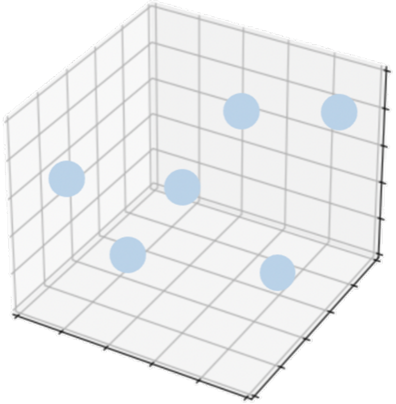
Geometric features $\{\mathbf{G}_i^{(0)}\}$

C×n matrix

- Preserves equivariant property
- Preserve motion attributes that are sensitive to Euclidean geometric transforms

Methodology - EqMotion

- Feature Description



Geometric features $\{\mathbf{G}_i^{(0)}\}$

C×n matrix

- Preserves equivariant property
- Preserve motion attributes that are sensitive to Euclidean geometric transforms

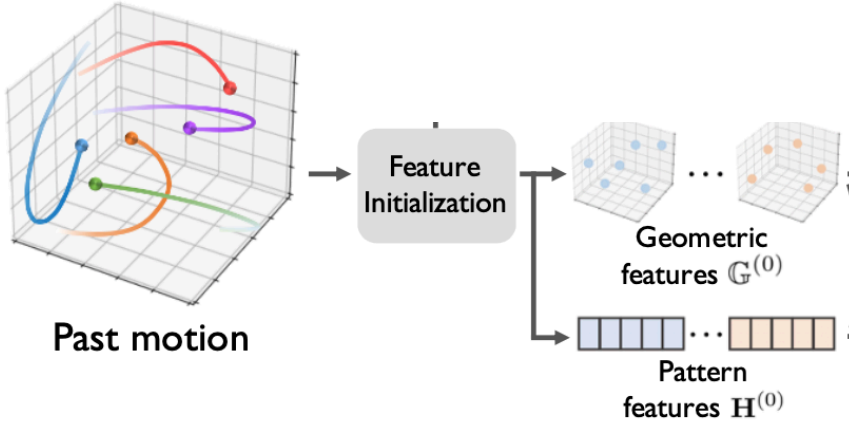
Pattern features $\{\mathbf{h}_i^{(0)}\}$

D - Vector

- Preserves invariant property
- Preserve motion attributes that are independent to Euclidean geometric transforms

Methodology - EqMotion

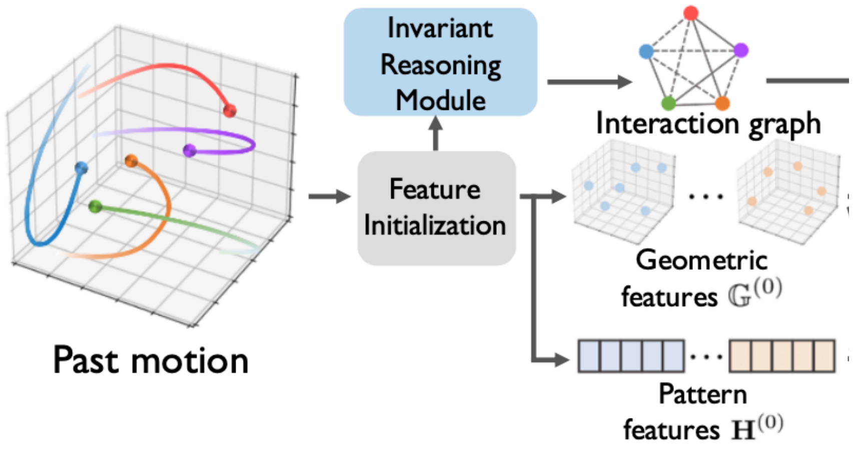
- Overview



1. Feature initialization

Methodology - EqMotion

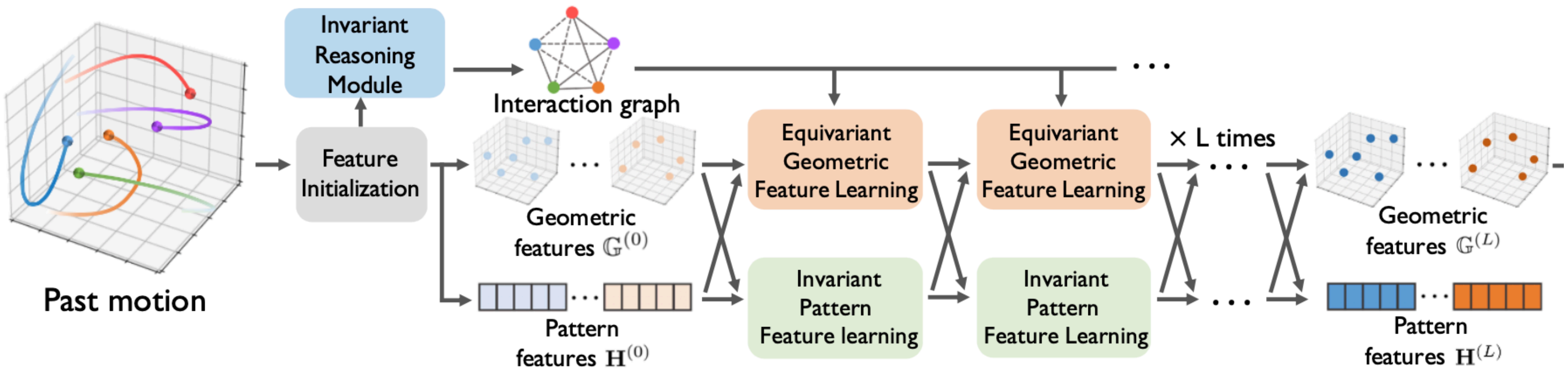
- Overview



1. Feature initialization
2. Invariant reasoning

Methodology - EqMotion

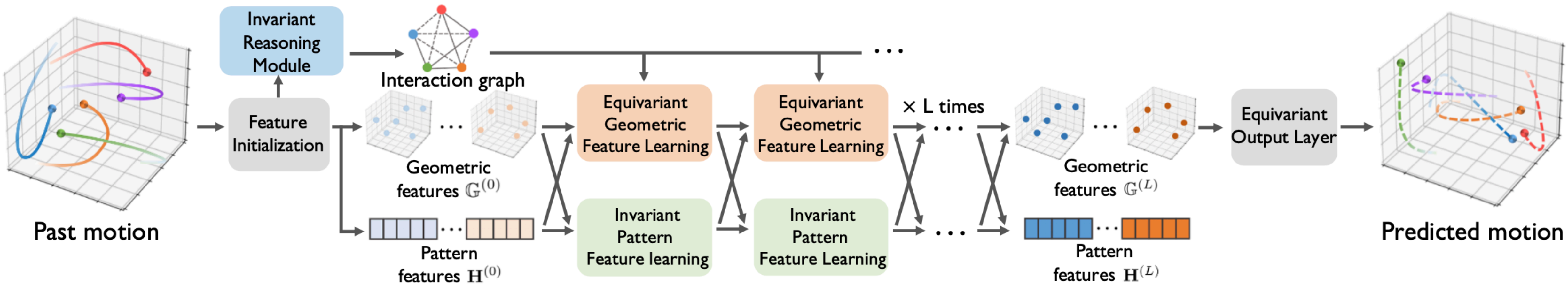
- Overview



1. Feature initialization
2. Invariant reasoning
3. Equivariant geometric feature learning
4. Invariant pattern feature learning

Methodology - EqMotion

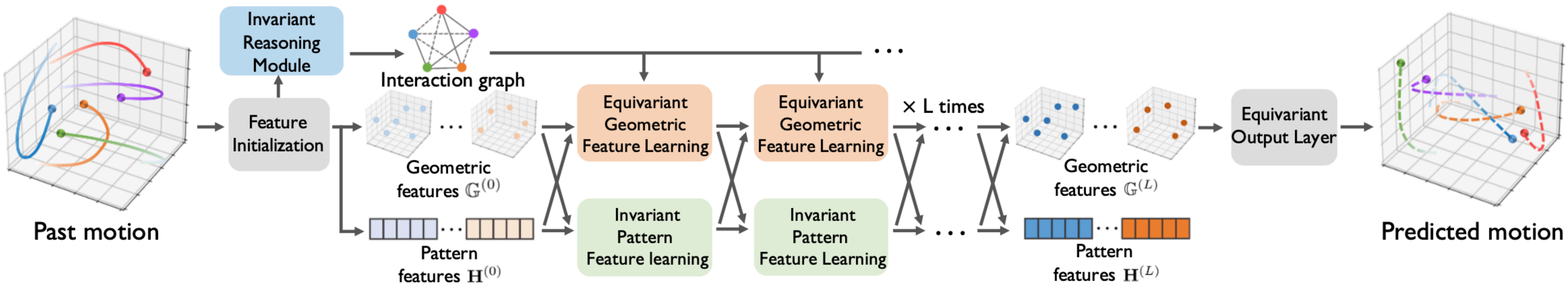
- Overview



1. Feature initialization
2. Invariant reasoning
3. Equivariant geometric feature learning
4. Invariant pattern feature learning
5. Equivariant output layer

Methodology - EqMotion

- Overview

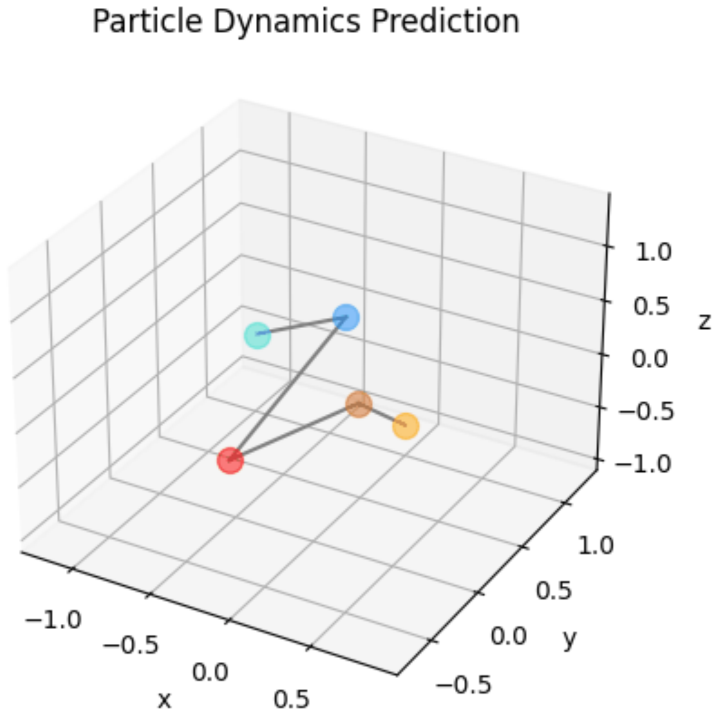


All network operation satisfy:

- Geometric feature – Equivariant!
- Pattern feature – Invariant!

Experiment

- Scenario 1: Particle Dynamic Prediction



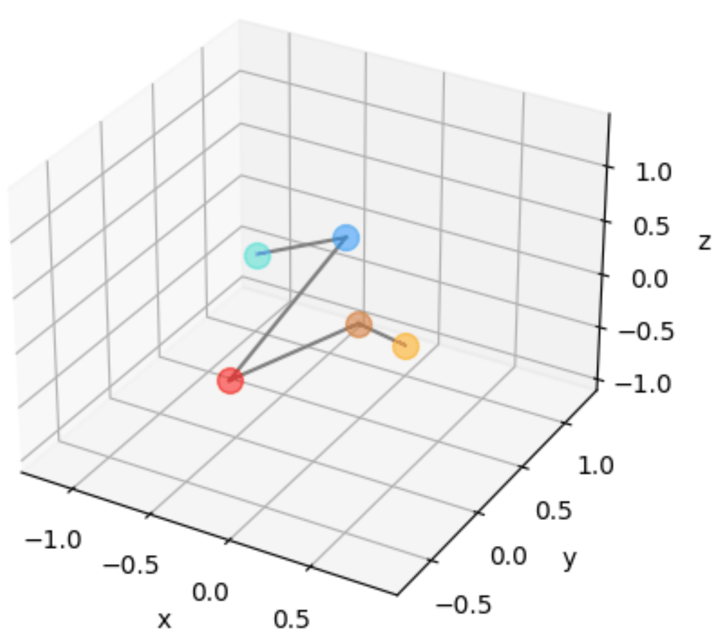
On reasoning:

Table 1. Interaction recognition accuracy and consistency (mean ± std in % in 5 independent runs) on the physical simulation.

Model	Springs		Charged	
	Accuracy	Consistency	Accuracy	Consistency
Corr.(path) [28]	58.1 ± 0.0	99.8 ± 0.1	57.5 ± 0.1	87.9 ± 0.1
Corr.(LSTM) [28]	53.5 ± 0.5	92.4 ± 2.1	57.2 ± 0.4	91.7 ± 1.1
EGNN [58]	61.0 ± 1.3	**100.0** ± 0.0	58.2 ± 1.4	**100.0** ± 0.0
NRI [28]	93.0 ± 1.1	93.7 ± 1.2	70.0 ± 0.6	88.5 ± 1.3
dNRI [17]	93.3 ± 2.0	89.6 ± 2.0	70.4 ± 1.7	83.6 ± 1.8
Ours	**97.6** ± 1.1	**100.0** ± 0.0	**80.9** ± 3.4	**100.0** ± 0.0
Supervised	98.7 ± 0.2	100.0 ± 0.0	97.4 ± 0.2	100.0 ± 0.0

Experiment

- Scenario 1: Particle Dynamic Prediction



On reasoning:

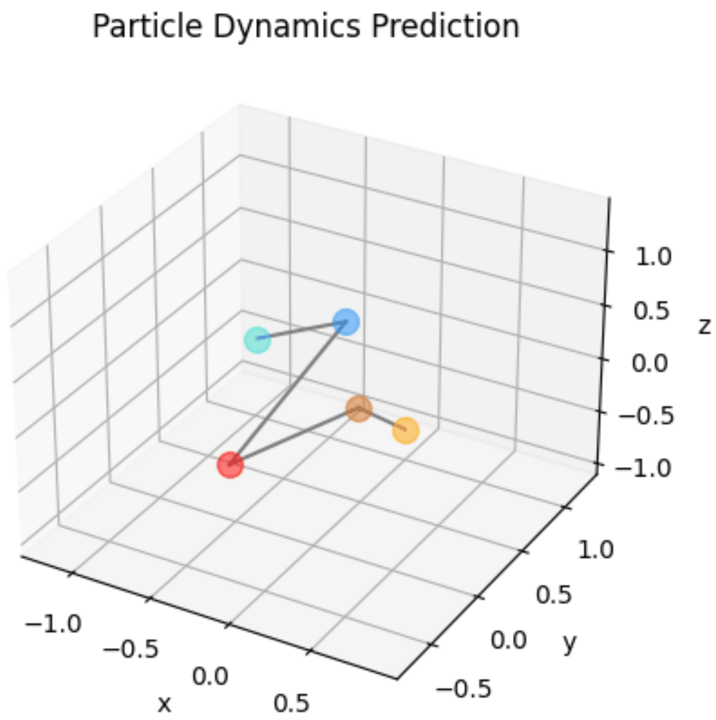
Table 1. Interaction recognition accuracy and consistency (mean ± std in % in 5 independent runs) on the physical simulation.

Model	**Springs**		**Charged**	
	Accuracy	Consistency	Accuracy	Consistency
Corr.(path) [28]	58.1 ± 0.0	99.8 ± 0.1	57.5 ± 0.1	87.9 ± 0.1
Corr.(LSTM) [28]	53.5 ± 0.5	92.4 ± 2.1	57.2 ± 0.4	91.7 ± 1.1
EGNN [58]	61.0 ± 1.3	$\mathbf{100.0} \pm 0.0$	58.2 ± 1.4	$\mathbf{100.0} \pm 0.0$
NRI [28]	93.0 ± 1.1	93.7 ± 1.2	70.0 ± 0.6	88.5 ± 1.3
dNRI [17]	93.3 ± 2.0	89.6 ± 2.0	70.4 ± 1.7	83.6 ± 1.8
Ours	$\mathbf{97.6} \pm 1.1$	$\mathbf{100.0} \pm 0.0$	$\mathbf{80.9} \pm 3.4$	$\mathbf{100.0} \pm 0.0$
Supervised	98.7 ± 0.2	100.0 ± 0.0	97.4 ± 0.2	100.0 ± 0.0

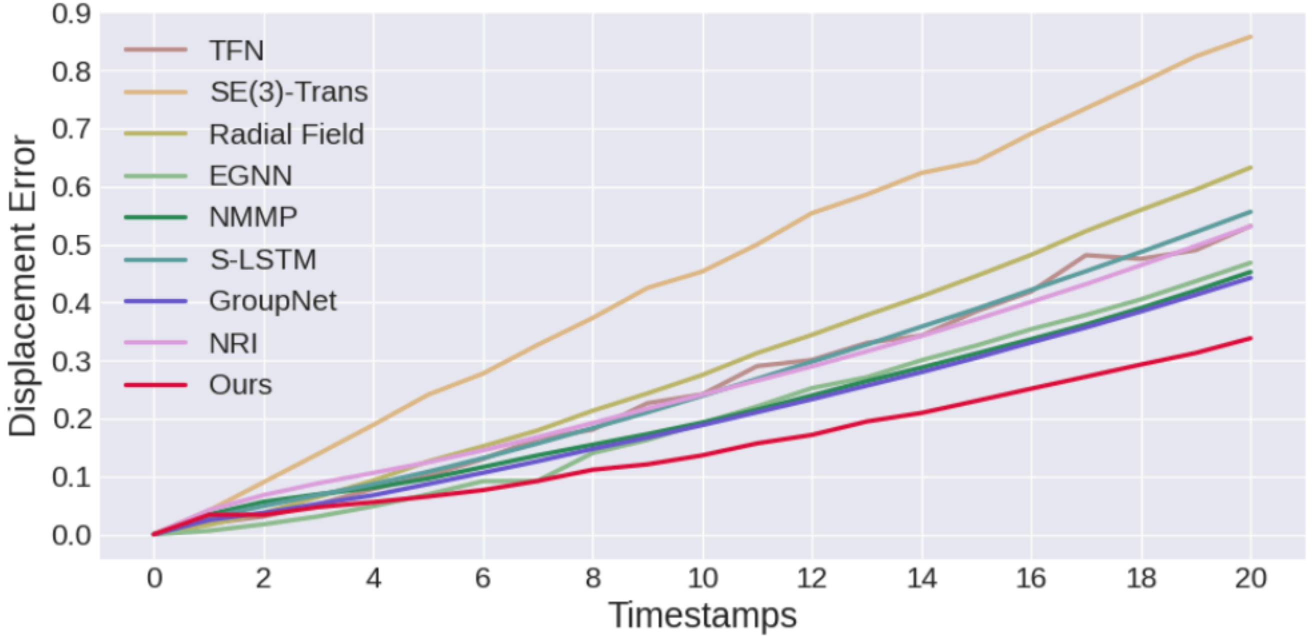
Accurate and consistent reasoning!

Experiment

- Scenario 1: Particle Dynamic Prediction



On prediction:



Accurate prediction!

Experiment

- Scenario II: Molecule Dynamic Prediction

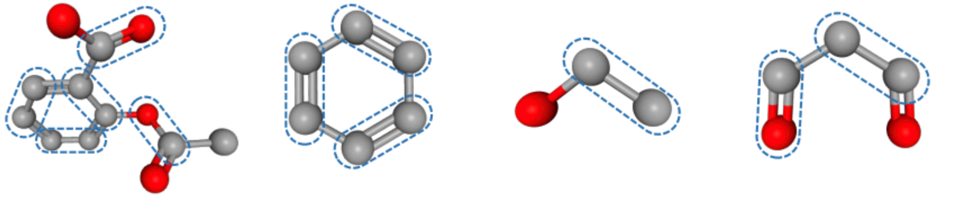


Table 4. Prediction ADE/FDE ($\times 10^{-2}$) on the MD17 dataset.

	Aspirin	Benzene	Ethanol	Malonaldehyde
Radial Field [31]	17.98/26.20	7.73/12.47	8.10/10.61	16.53/25.10
TFN [61]	15.02/21.35	7.55/12.30	8.05/10.57	15.21/24.32
SE(3)-Trans [14]	15.70/22.39	7.62/12.50	8.05/10.86	15.44/24.47
EGNN [58]	14.61/20.65	7.50/12.16	8.01/10.22	15.21/24.00
LSTM	17.59/24.79	6.06/9.46	7.73/9.88	15.14/22.90
S-LSTM [1]	13.12/18.14	3.06/3.52	7.23/9.85	11.93/18.43
NRI [28]	12.60/18.50	1.89/2.58	6.69/8.78	12.79/19.86
NMMP [22]	10.41/14.67	2.21/3.33	6.17/7.86	9.50/14.89
GroupNet [69]	10.62/14.00	2.02/2.95	6.00/7.88	7.99/12.49
EqMotion(Ours)	**5.95/8.38**	**1.18/1.73**	**5.05/7.02**	**5.85/9.02**

Experiment

- Scenario II: Molecule Dynamic Prediction

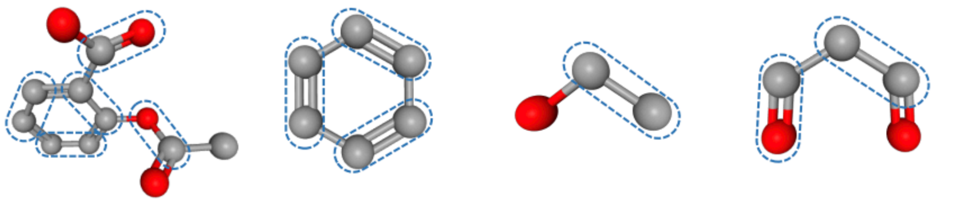


Table 4. Prediction ADE/FDE ($\times 10^{-2}$) on the MD17 dataset.

	Aspirin	Benzene	Ethanol	Malonaldehyde
Radial Field [31]	17.98/26.20	7.73/12.47	8.10/10.61	16.53/25.10
TFN [61]	15.02/21.35	7.55/12.30	8.05/10.57	15.21/24.32
SE(3)-Trans [14]	15.70/22.39	7.62/12.50	8.05/10.86	15.44/24.47
EGNN [58]	14.61/20.65	7.50/12.16	8.01/10.22	15.21/24.00
LSTM	17.59/24.79	6.06/9.46	7.73/9.88	15.14/22.90
S-LSTM [1]	13.12/18.14	3.06/3.52	7.23/9.85	11.93/18.43
NRI [28]	12.60/18.50	1.89/2.58	6.69/8.78	12.79/19.86
NMMP [22]	10.41/14.67	2.21/3.33	6.17/7.86	9.50/14.89
GroupNet [69]	10.62/14.00	2.02/2.95	6.00/7.88	7.99/12.49
EqMotion(Ours)	**5.95/8.38**	**1.18/1.73**	**5.05/7.02**	**5.85/9.02**

Accurate prediction on all kinds of molecules!

Experiment

- Scenario III: Human Skeleton Motion Prediction

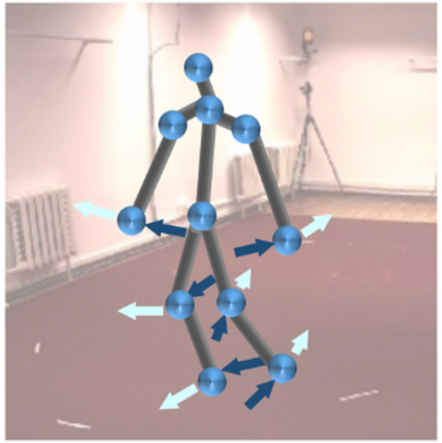


Table 2. Comparisons of short-term skeleton motion prediction on 11 representative actions and average results across all actions on H3.6M.

Motion	Walking				Eating				Smoking				Discussion				Directions				Phoning			
millisecond	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup. (CVPR'17)	29.4	50.8	76.0	81.5	16.8	30.6	56.9	68.7	23.0	42.6	70.1	82.7	32.9	61.2	90.9	96.2	35.4	57.3	76.3	87.7	38.0	69.3	115.0	126.7
Traj-GCN (ICCV'19)	12.3	23.0	39.8	46.1	8.4	16.9	33.2	40.7	7.9	16.2	31.9	38.9	12.5	27.4	58.5	71.7	9.0	19.9	43.4	53.7	10.2	21.0	42.5	52.3
DMGNN (CVPR'20)	17.3	30.7	54.6	65.2	11.0	21.4	36.2	43.9	9.0	17.6	32.1	40.3	17.3	34.8	61.0	69.8	13.1	24.6	64.7	81.9	12.5	25.8	48.1	58.3
MSRGCN (ICCV'21)	12.2	22.7	38.6	45.2	8.4	17.1	33.0	40.4	8.0	16.3	31.3	38.2	12.0	26.8	57.1	69.7	8.6	19.7	43.3	53.8	10.1	20.7	41.5	51.3
PGBIG (CVPR'22)	10.2	19.8	34.5	40.3	7.0	15.1	30.6	38.1	6.6	14.1	28.2	34.7	10.0	23.8	53.6	66.7	7.2	17.6	40.9	51.5	8.3	18.3	38.7	48.4
SPGSN (ECCV'22)	10.1	19.4	34.8	41.5	7.1	14.9	30.5	37.9	6.7	13.8	28.0	34.6	10.4	23.8	53.6	67.1	7.4	17.2	39.8	50.3	8.7	18.3	38.7	48.5
EqMotion (Ours)	**9.0**	**17.5**	**32.6**	**39.2**	**6.3**	**13.6**	**28.9**	**36.5**	**5.5**	**11.3**	**23.0**	**29.3**	**8.2**	**18.9**	**42.1**	**53.9**	**6.3**	**15.8**	**38.9**	**50.1**	**7.4**	**16.7**	**36.9**	**47.0**

Motion	Posing				Sitting				Sitting Down				Waiting				Walking Together				Average			
millisecond	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup. (CVPR'17)	36.1	69.1	130.5	157.1	42.6	81.4	134.7	151.8	47.3	86.0	145.8	168.9	30.6	57.8	106.2	121.5	26.8	50.1	80.2	92.2	34.7	62.0	101.1	115.5
Traj-GCN (ICCV'19)	13.7	29.9	66.6	84.1	10.6	21.9	46.3	57.9	16.1	31.1	61.5	75.5	11.4	24.0	50.1	61.5	10.5	21.0	38.5	45.2	12.7	26.1	52.3	63.5
DMGNN (CVPR'20)	15.3	29.3	71.5	96.7	11.9	25.1	44.6	50.2	15.0	32.9	77.1	93.0	12.2	24.2	59.6	77.5	14.3	26.7	50.1	63.2	17.0	33.6	65.9	79.7
MSRGCN (ICCV'21)	12.8	29.4	67.0	85.0	10.5	22.0	46.3	57.8	16.1	31.6	62.5	76.8	10.7	23.1	48.3	59.2	10.6	20.9	37.4	43.9	12.1	25.6	51.6	62.9
PGBIG (CVPR'22)	10.7	25.7	60.0	76.6	8.8	19.2	42.4	53.8	13.9	27.9	57.4	71.5	8.9	20.1	43.6	54.3	8.7	18.6	34.4	41.0	10.3	22.7	47.4	58.5
SPGSN (ECCV'22)	10.7	25.3	59.9	76.5	9.3	19.4	42.3	53.6	14.2	27.7	56.8	**70.7**	9.2	19.8	43.1	54.1	8.9	18.2	33.8	40.9	10.4	22.3	47.1	58.3
EqMotion (Ours)	**8.2**	**18.9**	**43.4**	**57.5**	**8.1**	**18.0**	**41.2**	**52.9**	**13.0**	**26.5**	**56.2**	**70.7**	**7.6**	**17.4**	**39.9**	**51.1**	**7.8**	**16.1**	**30.6**	**37.1**	**9.1**	**20.1**	**43.7**	**55.0**

Table 3. Comparisons of long-term skeleton motion prediction on 8 representative actions and average results across all actions on H3.6M.

Motion	Walking		Eating		Smoking		Discussion		Greeting		Phoning		Posing		Walking Together		Average	
millisecond	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms
Res-Sup. [50]	81.7	100.7	79.9	100.2	94.8	137.4	121.3	161.7	156.3	184.3	143.9	186.8	165.4	236.8	173.6	202.3	129.2	165.0
Traj-GCN [47]	54.1	59.8	53.4	77.8	50.7	72.6	91.6	121.5	115.4	148.8	69.2	103.1	114.5	173.0	55.0	65.6	81.6	114.3
DMGNN [41]	71.4	85.8	58.1	86.7	50.9	72.2	81.9	138.3	144.5	170.5	71.3	108.4	125.5	188.2	70.5	86.9	93.6	127.6
MSRGCN [9]	52.7	63.0	52.5	77.1	49.5	71.6	88.6	117.6	116.3	147.2	68.3	104.4	116.3	174.3	52.9	65.9	81.1	114.2
PGBIG [44]	48.1	56.4	51.1	76.0	46.5	69.5	87.1	118.2	110.2	143.5	65.9	102.7	106.1	164.8	51.9	64.3	76.9	110.3
SPGSN [40]	46.9	53.6	49.8	73.4	46.7	68.6	89.7	118.6	111.0	143.2	66.7	102.5	110.3	165.4	49.8	60.9	77.4	109.6
EqMotion (Ours)	**43.4**	**52.8**	**48.4**	**73.0**	**41.0**	**63.4**	**75.3**	**105.6**	**108.7**	**142.0**	**64.7**	**101.0**	**84.9**	**139.4**	**44.5**	**56.0**	**73.4**	**106.9**

Experiment

- Scenario III: Human Skeleton Motion Prediction

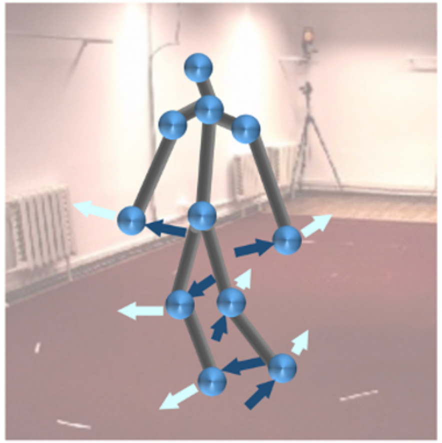


Table 2. Comparisons of short-term skeleton motion prediction on 11 representative actions and average results across all actions on H3.6M.

Motion	Walking				Eating				Smoking				Discussion				Directions				Phoning			
millisecond	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup. (CVPR'17)	29.4	50.8	76.0	81.5	16.8	30.6	56.9	68.7	23.0	42.6	70.1	82.7	32.9	61.2	90.9	96.2	35.4	57.3	76.3	87.7	38.0	69.3	115.0	126.7
Traj-GCN (ICCV'19)	12.3	23.0	39.8	46.1	8.4	16.9	33.2	40.7	7.9	16.2	31.9	38.9	12.5	27.4	58.5	71.7	9.0	19.9	43.4	53.7	10.2	21.0	42.5	52.3
DMGNN (CVPR'20)	17.3	30.7	54.6	65.2	11.0	21.4	36.2	43.9	9.0	17.6	32.1	40.3	17.3	34.8	61.0	69.8	13.1	24.6	64.7	81.9	12.5	25.8	48.1	58.3
MSRGCN (ICCV'21)	12.2	22.7	38.6	45.2	8.4	17.1	33.0	40.4	8.0	16.3	31.3	38.2	12.0	26.8	57.1	69.7	8.6	19.7	43.3	53.8	10.1	20.7	41.5	51.3
PGBIG (CVPR'22)	10.2	19.8	34.5	40.3	7.0	15.1	30.6	38.1	6.6	14.1	28.2	34.7	10.0	23.8	53.6	66.7	7.2	17.6	40.9	51.5	8.3	18.3	38.7	48.4
SPGSN (ECCV'22)	10.1	19.4	34.8	41.5	7.1	14.9	30.5	37.9	6.7	13.8	28.0	34.6	10.4	23.8	53.6	67.1	7.4	17.2	39.8	50.3	8.7	18.3	38.7	48.5
EqMotion (Ours)	**9.0**	**17.5**	**32.6**	**39.2**	**6.3**	**13.6**	**28.9**	**36.5**	**5.5**	**11.3**	**23.0**	**29.3**	**8.2**	**18.9**	**42.1**	**53.9**	**6.3**	**15.8**	**38.9**	**50.1**	**7.4**	**16.7**	**36.9**	**47.0**

Motion	Posing				Sitting				Sitting Down				Waiting				Walking Together				Average			
millisecond	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup. (CVPR'17)	36.1	69.1	130.5	157.1	42.6	81.4	134.7	151.8	47.3	86.0	145.8	168.9	30.6	57.8	106.2	121.5	26.8	50.1	80.2	92.2	34.7	62.0	101.1	115.5
Traj-GCN (ICCV'19)	13.7	29.9	66.6	84.1	10.6	21.9	46.3	57.9	16.1	31.1	61.5	75.5	11.4	24.0	50.1	61.5	10.5	21.0	38.5	45.2	12.7	26.1	52.3	63.5
DMGNN (CVPR'20)	15.3	29.3	71.5	96.7	11.9	25.1	44.6	50.2	15.0	32.9	77.1	93.0	12.2	24.2	59.6	77.5	14.3	26.7	50.1	63.2	17.0	33.6	65.9	79.7
MSRGCN (ICCV'21)	12.8	29.4	67.0	85.0	10.5	22.0	46.3	57.8	16.1	31.6	62.5	76.8	10.7	23.1	48.3	59.2	10.6	20.9	37.4	43.9	12.1	25.6	51.6	62.9
PGBIG (CVPR'22)	10.7	25.7	60.0	76.6	8.8	19.2	42.4	53.8	13.9	27.9	57.4	71.5	8.9	20.1	43.6	54.3	8.7	18.6	34.4	41.0	10.3	22.7	47.4	58.5
SPGSN (ECCV'22)	10.7	25.3	59.9	76.5	9.3	19.4	42.3	53.6	14.2	27.7	56.8	**70.7**	9.2	19.8	43.1	54.1	8.9	18.2	33.8	40.9	10.4	22.3	47.1	58.3
EqMotion (Ours)	**8.2**	**18.9**	**43.4**	**57.5**	**8.1**	**18.0**	**41.2**	**52.9**	**13.0**	**26.5**	**56.2**	**70.7**	**7.6**	**17.4**	**39.9**	**51.1**	**7.8**	**16.1**	**30.6**	**37.1**	**9.1**	**20.1**	**43.7**	**55.0**

Table 3. Comparisons of long-term skeleton motion prediction on 8 representative actions and average results across all actions on H3.6M.

Motion	Walking		Eating		Smoking		Discussion		Greeting		Phoning		Posing		Walking Together		Average	
millisecond	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms	560ms	1000ms
Res-Sup. [50]	81.7	100.7	79.9	100.2	94.8	137.4	121.3	161.7	156.3	184.3	143.9	186.8	165.4	236.8	173.6	202.3	129.2	165.0
Traj-GCN [47]	54.1	59.8	53.4	77.8	50.7	72.6	91.6	121.5	115.4	148.8	69.2	103.1	114.5	173.0	55.0	65.6	81.6	114.3
DMGNN [41]	71.4	85.8	58.1	86.7	50.9	72.2	81.9	138.3	144.5	170.5	71.3	108.4	125.5	188.2	70.5	86.9	93.6	127.6
MSRGCN [9]	52.7	63.0	52.5	77.1	49.5	71.6	88.6	117.6	116.3	147.2	68.3	104.4	116.3	174.3	52.9	65.9	81.1	114.2
PGBIG [44]	48.1	56.4	51.1	76.0	46.5	69.5	87.1	118.2	110.2	143.5	65.9	102.7	106.1	164.8	51.9	64.3	76.9	110.3
SPGSN [40]	46.9	53.6	49.8	73.4	46.7	68.6	89.7	118.6	111.0	143.2	66.7	102.5	110.3	165.4	49.8	60.9	77.4	109.6
EqMotion (Ours)	**43.4**	**52.8**	**48.4**	**73.0**	**41.0**	**63.4**	**75.3**	**105.6**	**108.7**	**142.0**	**64.7**	**101.0**	**84.9**	**139.4**	**44.5**	**56.0**	**73.4**	**106.9**

Far more accurate prediction without any specific design for human skeleton!

Experiment

- Scenario IV: Pedestrian Trajectory Prediction

Table 5. Prediction performance on the ETH-UCY dataset. The **bold**/underline font denotes the best/second best result.

	Performance (ADE/FDE)					
Deterministic	ETH	Hotel	Univ	Zara1	Zara2	Average
S-LSTM [1]	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
SGAN-ind [20]	1.13/2.21	1.01/2.18	0.60/1.28	0.42/0.91	0.52/1.11	0.74/1.54
Traj++ [55]	1.02/2.00	0.33/0.62	0.53/1.19	0.44/0.99	0.32/0.73	0.53/1.11
TransF [16]	1.03/2.10	0.36/0.71	0.53/1.32	0.44/1.00	0.34/0.76	0.54/1.17
MemoNet [70]	1.00/2.08	0.35/0.67	0.55/1.19	0.46/1.00	0.37/0.82	0.55/1.15
EqMotion(Ours)	**0.96/1.92**	**0.30/0.58**	**0.50/1.10**	**0.39/0.86**	**0.30/0.68**	**0.49/1.03**
Multi-prediction	ETH	Hotel	Univ	Zara1	Zara2	Average
SGAN [20]	0.87/1.62	0.67/1.37	0.76/0.52	0.35/0.68	0.42/0.84	0.61/1.21
NMMP [22]	0.61/1.08	0.33/0.63	0.52/1.11	0.32/0.66	0.43/0.85	0.41/0.82
Traj++ [55]	0.61/1.02	0.19/0.28	0.30/0.54	0.24/0.42	0.18/0.31	0.30/0.51
PECNet [45]	0.54/0.87	0.18/0.24	0.35/0.60	0.22/0.39	0.17/0.30	0.29/0.48
Agentformer [76]	0.45/0.75	0.14/0.22	0.25/0.45	0.18/**0.30**	0.14/0.24	0.23/0.39
GroupNet [69]	0.46/0.73	0.15/0.25	0.26/0.49	0.21/0.39	0.17/0.33	0.25/0.44
MID [18]	**0.39**/0.66	0.13/0.22	**0.22**/0.45	**0.17/0.30**	**0.13**/0.27	**0.21**/0.38
GP-Graph [2]	0.43/0.63	0.18/0.30	0.24/0.42	**0.17**/0.31	0.15/0.29	0.23/0.39
EqMotion(Ours)	0.40/**0.61**	**0.12/0.18**	0.23/**0.43**	0.18/0.32	**0.13/0.23**	**0.21/0.35**

Experiment

- Data efficiency

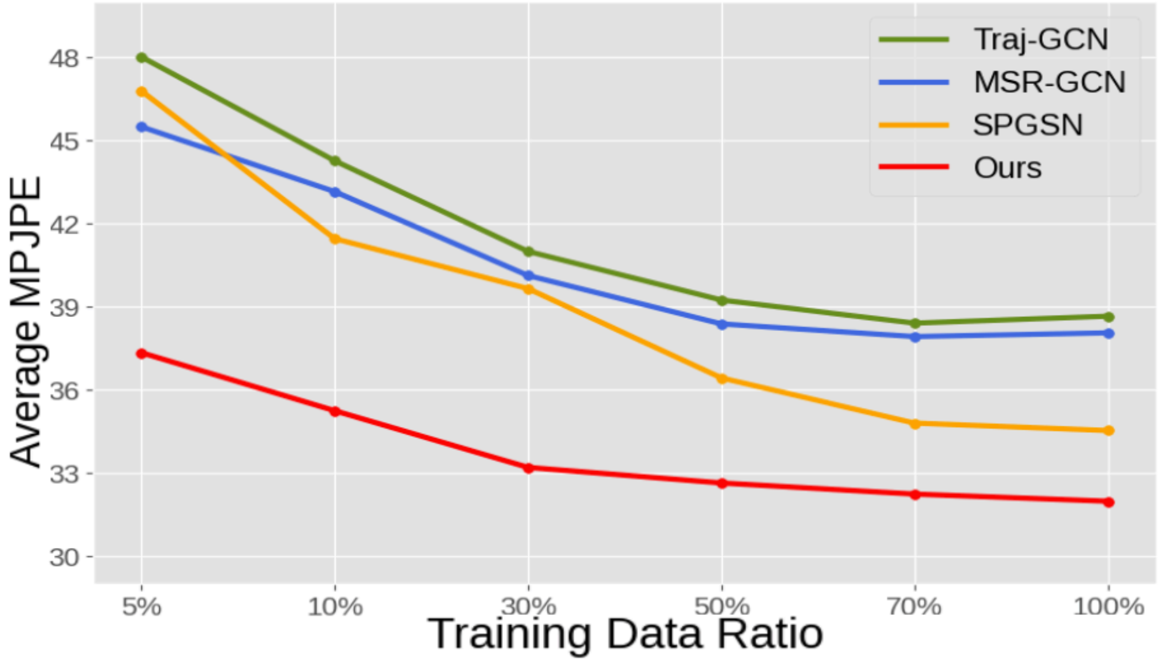


Figure 5. Comparison of model performance on different amounts of data in short-term prediction on H3.6M dataset.

Experiment

- Data efficiency

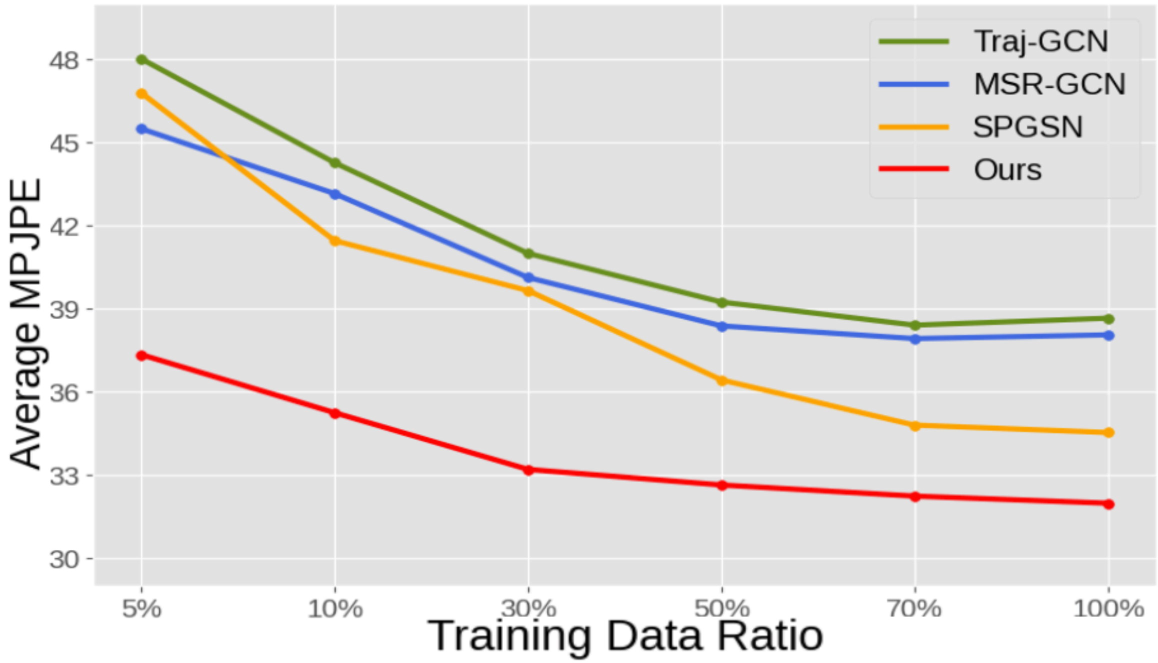


Figure 5. Comparison of model performance on different amounts of data in short-term prediction on H3.6M dataset.

Outperform SOTA by only 30% data!

Experiment

- Model Size

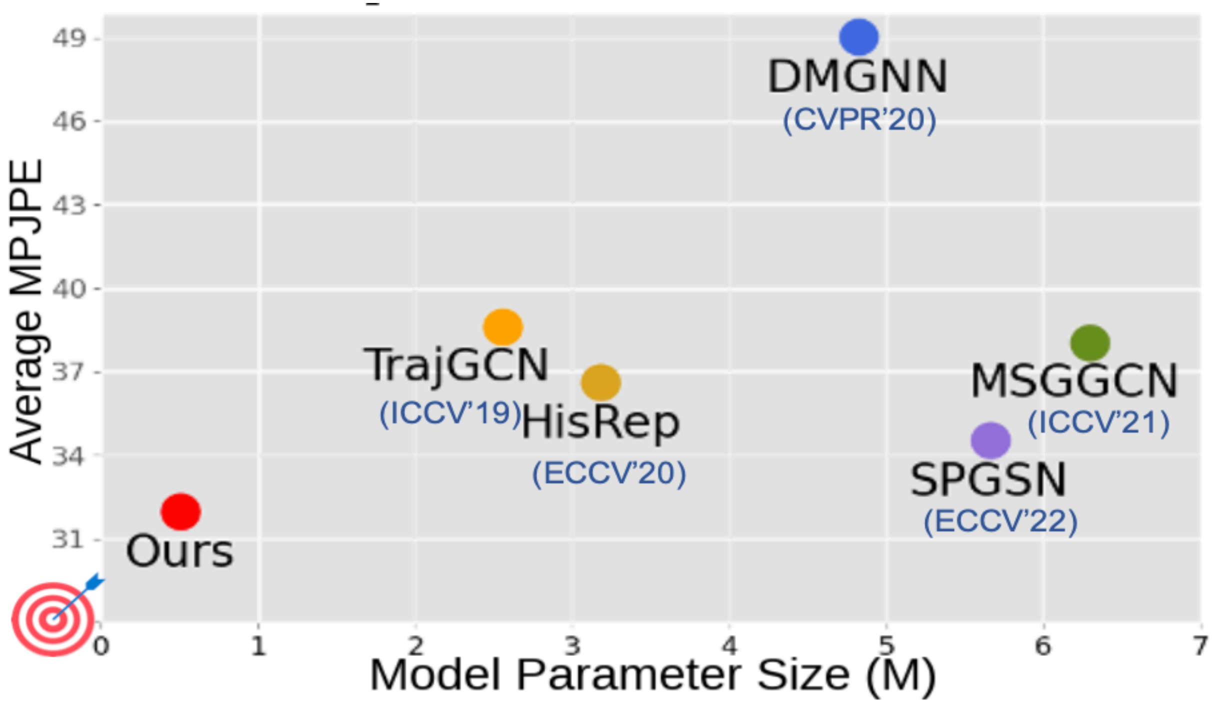


Figure 6. Comparison of model size and MPJPE in short-term prediction on H3.6M dataset. The target means the ideal model.

Experiment

- Model Size

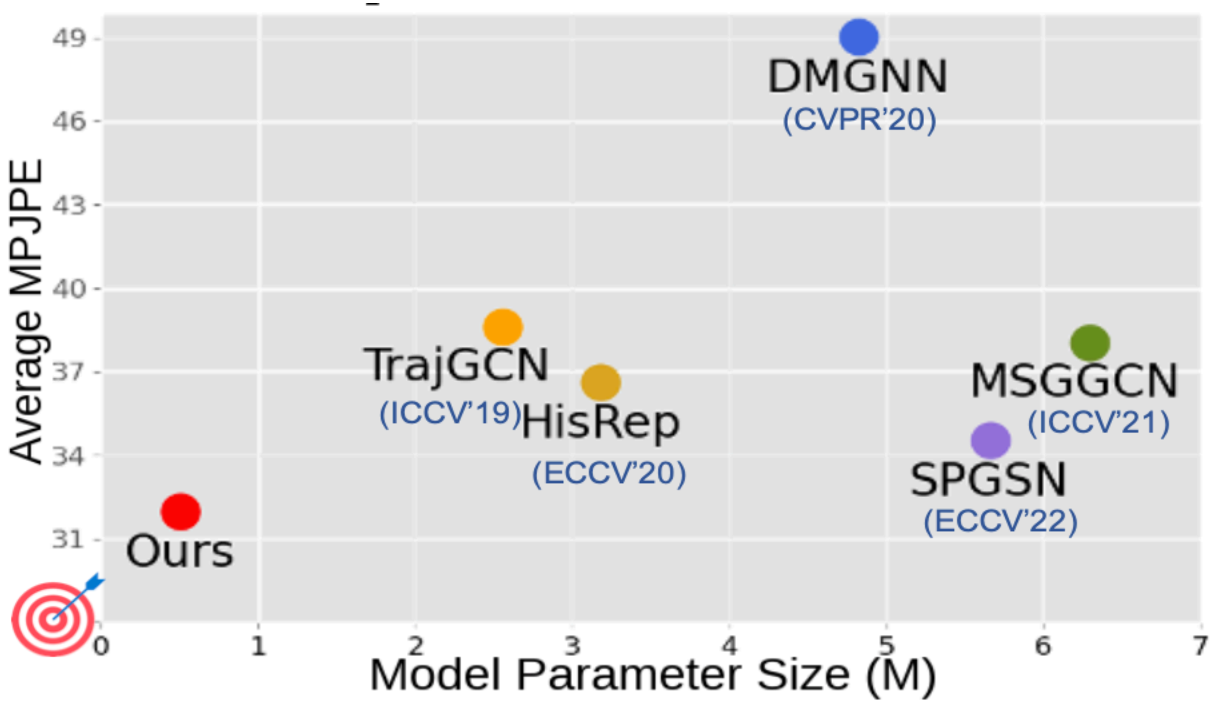


Figure 6. Comparison of model size and MPJPE in short-term prediction on H3.6M dataset. The target means the ideal model.

Less than 30% of other SOTA models' sizes !

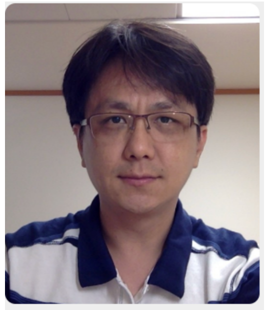
Thanks for your listening!

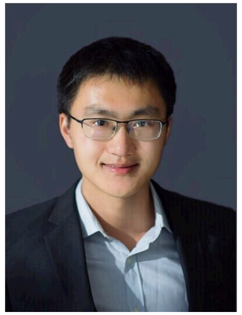
Question/comments: xcxwakaka@sjtu.edu.cn

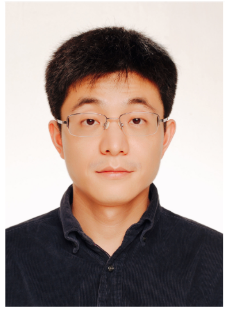
Code: https://github.com/MediaBrain-SJTU/EqMotion

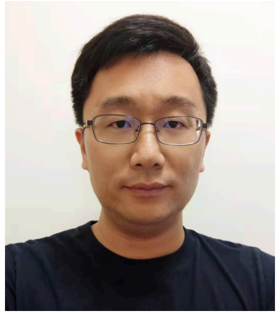
Our Team


Chenxin Xu


Robby. T. Tan


Siheng Chen


Yu Guang Wang


Xinchao Wang


Yanfeng Wang

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

NUS National University of Singapore

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory