Equivariant Spatio-Temporal Attentive Graph Networks to Simulate Physical Dynamics

Liming Wu[1*], Zhichao Hou[2*], Jirui Yuan[3], Yu Rong[4], Wenbing Huang[1]

GSAI, Renmin University of China[1], North Carolina State University[2], Tsinghua University[3], Tencent AI Lab[4]

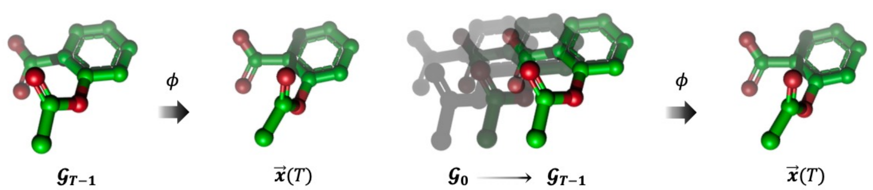
Equal Contributions[*], Corresponding Author

中国人民大學 RENMIN UNIVERSITY OF CHINA

Code Lab Wechat

Introduction

- TL; DR: We design a graph neural network to capture both spatial and temporal dependencies while respecting the underlying symmetries of the simulated systems.
- Learning to represent and simulate the dynamics of physical systems is a crucial yet challenging task.
- Frame-to-frame formulation of the task overlooks the non-Markov property. We reformulate dynamics simulation as a spatio-temporal prediction task, by employing the trajectory in the past period.



(a) Previous Setting (b) Our Setting

Equivariant Spatio-Temporal Attentive GNN

Existing Challenges

1. Frame-to-Frame Prediction

$$\boxed{(H(T-1), \vec{X}(T-1), A) \rightarrow \vec{X}(T)}$$

✔ E(3)-Equivariance ✘ Unobserved Influence

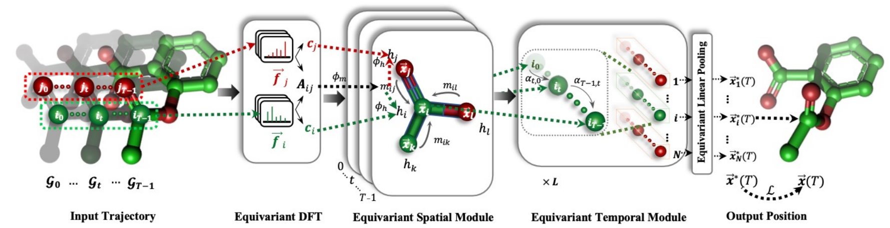
✔ Generalization ✘ Expressivity

2. Spatio-Temporal Prediction (2D setting)

$$\boxed{\{(H(T-1), A)\}_{t=0}^{T-1} \rightarrow H(T)}$$

✔ Temporal Dependency ✘ Geometric Feature

The overall architecture of ESTAG



1. Equivariant Discrete Fourier Transform (EDFT)

From time domain to frequency domain

$$\vec{f}_i(k) = \sum_{t=0}^{T-1} e^{-i'\frac{2\pi}{T}kt}\left(\vec{x}_i(t) - \overline{\vec{x}(t)}\right),$$

Compute cross-correlation and amplitude

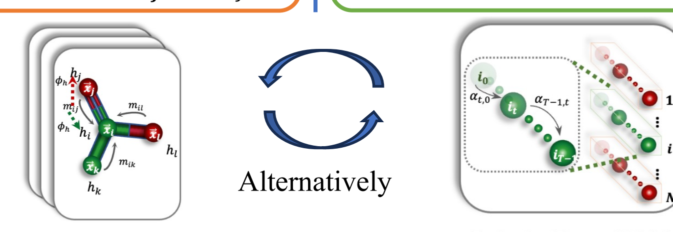
$$A_{ij}(k) = w_k(h_i)w_k(h_j)\left|\langle \vec{f}_i(k), \vec{f}_j(k)\rangle\right|,$$
$$c_i(k) = w_k(h_i)\|\vec{f}_i(k)\|^2.$$

2. Equivariant Message Passing (ESM & ETM)

We conduct inner-graph message passing and inter-graph message passing alternatively.

$$m_{ij} = \phi_m(h_i, d_{ij}, A_{ij})$$
$$h_i = h_i + \phi_h(h_i, c_i, \sum m_{ij})$$
$$x_i = x_i + \sum \boldsymbol{x_{ij}}\phi_x(m_{ij})$$

$$\alpha_{ts} = \text{attention}(Q, K, V)$$
$$h_i(t) = h_i(t) + \sum \alpha_{ts} v_i(s)$$
$$x_i(t) = x_i(t) + \sum \alpha_{ts}\boldsymbol{x_i(ts)}\phi_x(v_i(s))$$



3. Equivariant Temporal Pooling (ETP)

The trajectory is finally aggregated by a learned weight w, as the ultimate output for the prediction.

$$\vec{x}_i^{*}(T) = \hat{X}_i w + \vec{x}_i^{(L)}(T-1), w \in R^{T-1} \rightarrow L = \sum \|\vec{x}_i(T) - \vec{x}_i^{*}(T)\|_2^2$$

Where, $\hat{X}_i = [\vec{x}_i^{(L)}(0) - \vec{x}_i^{(L)}(T-1), \cdots, \vec{x}_i^{(L)}(T-2) - \vec{x}_i^{(L)}(T-1)]$

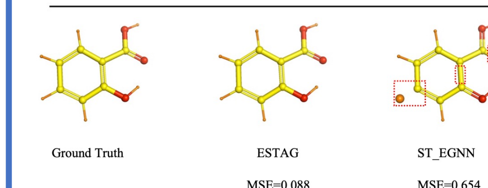
Experiments

We evaluate the efficacy of ESTAG in three scenarios.

1. Molecular Dataset: MD17

Table 1: Prediction error ($\times 10^{-3}$) on MD17 dataset. Results averaged across 3 runs. We do not display the standard deviation due to its small value.

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
Pt-s	15.579	4.457	4.332	13.206	8.958	12.256	6.818	10.269
Pt-m	9.058	2.536	2.688	6.749	6.918	8.122	5.622	7.257
Pt-t	0.715	0.114	0.456	0.596	0.737	0.688	0.688	0.674
EGNN-s	12.056	3.290	2.354	10.635	4.871	8.733	3.154	6.815
EGNN-m	6.237	1.882	1.532	4.842	3.791	4.623	2.516	3.606
EGNN-t	0.625	0.112	0.416	0.513	0.614	0.598	0.577	0.568
ST_TFN	0.719	0.122	0.432	0.569	0.688	0.684	0.628	0.669
ST_GNN	1.014	0.210	0.487	0.664	0.769	0.789	0.713	0.680
ST_SE(3)TR	<u>0.669</u>	0.119	0.428	0.550	0.625	0.630	0.591	0.597
ST_EGNN	0.735	0.163	<u>0.245</u>	<u>0.427</u>	0.745	0.687	<u>0.553</u>	<u>0.445</u>
EqMotion	0.721	0.156	0.476	0.600	0.747	0.697	0.691	0.681
STGCN	0.715	<u>0.106</u>	0.456	0.596	0.736	0.682	0.687	0.673
AGL-STAN	0.719	<u>0.106</u>	0.459	0.596	<u>0.601</u>	<u>0.452</u>	0.683	0.515
ESTAG	**0.063**	**0.003**	**0.099**	**0.101**	**0.068**	**0.047**	**0.079**	**0.066**



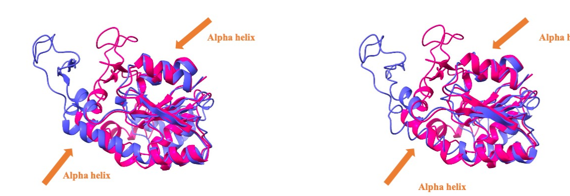
ESTAG achieves the lowest prediction error and almost reconstructs the ground truths.

2. Protein Dynamics: AdK

Table 2: Prediction error and training time on Protein dataset. Results averaged across 3 runs.

Method	MSE	Time(s)
Pt-s	3.260	-
Pt-m	3.302	-
Pt-t	2.022	-
EGNN-s	3.254	1.062
EGNN-m	3.278	1.088
EGNN-t	1.983	1.069
ST_GNN	1.871	2.769
ST_GMN	<u>1.526</u>	4.705
ST_EGNN	1.543	4.705
STGCN	1.578	1.840
AGL-STAN	1.671	1.478
ESTAG	**1.471**	6.876

ESTAG also performs better, particularly for the prediction of alpha helix parts.



3. Human Motion: CMU Motion Capture

Table 3: Prediction error ($\times 10^{-1}$) on Motion dataset. Results averaged across 3 runs.

Method	Walk	Basketball
Pt-s	329.474	886.023
Pt-m	127.152	413.306
Pt-t	3.831	15.878
EGNN-s	63.540	749.486
EGNN-m	32.016	335.002
EGNN-t	0.786	12.492
ST_GNN	0.441	15.336
ST_TFN	0.597	13.709
ST_SE(3)TR	0.236	13.851
ST_EGNN	0.538	13.199
EqMotion	1.011	<u>4.893</u>
STGCN	0.062	4.919
AGL-STAN	**0.037**	5.734
ESTAG	<u>0.040</u>	**0.746**

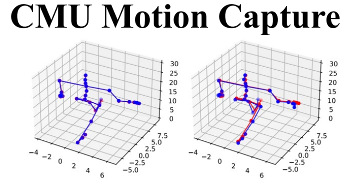


Figure 10: Comparison on Motion walk subject between ESTAG (left, MSE=0.0048) and ST_EGNN (right, MSE=0.0811). The ground truths are in red while the predicted states are in blue.

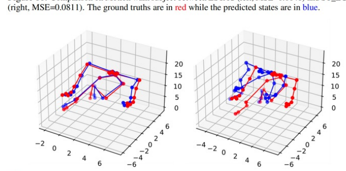


Figure 11: Comparison on Motion basketball subject between ESTAG (left, MSE=0.0749) and ST_EGNN (right, MSE=2.6380). The ground truths are in red while the predicted states are in blue.

The motions predicted by ESTAG are closer to the ground truths.