NeurIPS 2023 poster paper

Equivariant Spatio-Temporal Attentive Graph Networks to Simulate Physical Dynamics

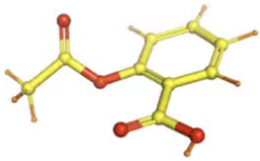
Liming Wu*, Zhichao Hou*, Jirui Yuan, Yu Rong, Wenbing Huang[1]

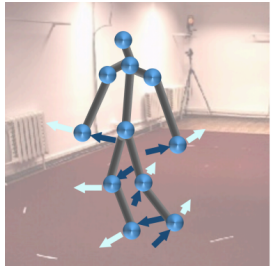
Gaoling School of Artificial Intelligence,
Renmin University of China

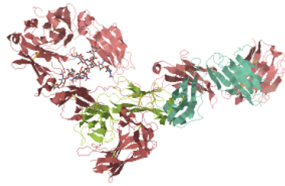
December 1, 2023

Outline

Physical Dynamics







- Target: Predict future states according to the historical trajectories.
- Symmetry, such as E(3)-equivariance, is significant in 3D scenarios.

[1]. Xu, Chenxin, et al. "EqMotion: Equivariant Multi-agent Motion Prediction with Invariant Interaction Reasoning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

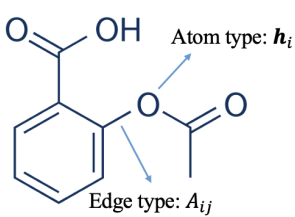
Challenges

- Is markovian assumption reasonable?
  - It needs to consider **unobserved** dynamics in the environment.
- How to fully leverage the past states?
  - RNN? Gradient vanishing...
  - **Attention** is all you need!
- How to preserve the symmetry?
  - At present, the majority of spatio-temporal graph models are only applicable to 2D settings and struggle to perform effectively in 3D environments.
  - **Equivariant** GNN can make sense!

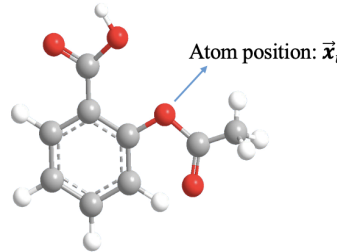
ESTAG can capture both spatial and temporal dependencies while respecting the underlying symmetries of dynamics simulation problems.

What's the Specialty of Geometric Graph

Topological Graph
$\mathcal{G}(\boldsymbol{H}, \boldsymbol{A})$

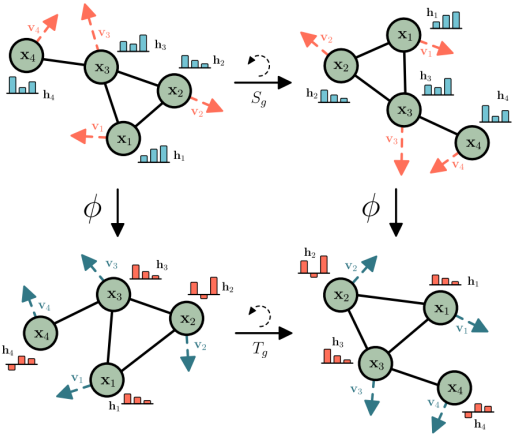


Geometric Graph
$\mathcal{G}(\vec{\boldsymbol{X}}, \boldsymbol{H}, \boldsymbol{A})$



- Scalars: $\boldsymbol{H} = [\boldsymbol{h}_0, \cdots, \boldsymbol{h}_N]$, $\boldsymbol{A} = \left(A_{ij}\right)_{i=1,j=1}^{N}$
- Vectors: $\vec{\boldsymbol{X}} = [\vec{\boldsymbol{x}}_0, \cdots, \vec{\boldsymbol{x}}_N]$

Equivariance and Invariance



$$\vec{\boldsymbol{x}}_{t+1}, \boldsymbol{h}_{t+1} = \phi(\vec{\boldsymbol{x}}_t, \boldsymbol{h}_t)$$

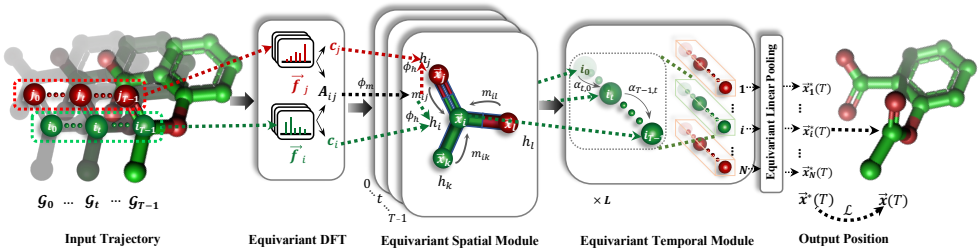
$$\Downarrow$$

$$R\vec{\boldsymbol{x}}_{t+1} + \vec{\boldsymbol{t}}, \boldsymbol{h}_{t+1} = \phi(R\vec{\boldsymbol{x}}_t + \vec{\boldsymbol{t}}, \boldsymbol{h}_t)$$

[1]. Satorras, Vıctor Garcia, Emiel Hoogeboom, and Max Welling. "E (n) equivariant graph neural networks." International conference on machine learning. PMLR, 2021.

Outline

Overview of ESTAG



- Equivariant Discrete Fourier Transform (EDFT).
- Equivariant **Spatial** Module (ESM).
- Equivariant **Temporal** Attention Module (ETM).
- Equivariant Temporal Pooling (ETP).

EDFT: Extract Periodic Patterns from Input Trajectory

- Obtain frequency domain representations. It first translates the node trajectories to ensure **translation invariant** frequency features.

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

- Calculate the cross-correlation in frequency domain, as **edge features** in later modules.

$$\boldsymbol{A}_{ij}(k) = w_k(\boldsymbol{h}_i) w_k(\boldsymbol{h}_j) |\langle \vec{\boldsymbol{f}}_i(k), \vec{\boldsymbol{f}}_j(k) \rangle|,$$

- Compute the frequency amplitude of each node trajectory, which captures the **node-wise** temporal dynamics. This is used to update node representations.

$$\boldsymbol{c}_i(k) = w_k(\boldsymbol{h}_i) ||\vec{\boldsymbol{f}}_i(k)||^2.$$

Message Passing Alternates between ESM and ETM

Spatial Interaction

$$\boldsymbol{m}_{ij} = \phi_m\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{h}_j^{(l)}(t), ||\vec{\boldsymbol{x}}_{ij}^{(l)}(t)||^2, \boldsymbol{A}_{ij}\right),$$

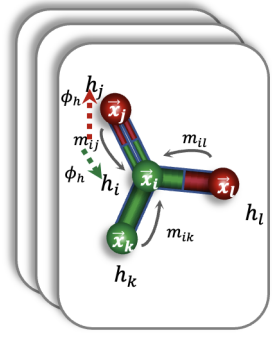
$$\boldsymbol{h}_i^{(l+1)}(t) = \boldsymbol{h}_i^{(l)}(t) + \phi_h\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{c}_i, \sum_{j\neq i}\boldsymbol{m}_{ij}\right),$$

$$\vec{\mathbf{a}}_i(t) = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)}\vec{\boldsymbol{x}}_{ij}^{(l)}(t)\phi_x(\boldsymbol{m}_{ij}),$$

$$\vec{\boldsymbol{x}}_i^{(l+1)}(t) = \vec{\boldsymbol{x}}_i^{(l)}(t) + \vec{\mathbf{a}}_i(t),$$

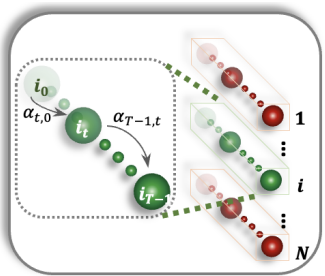
where, $\vec{\boldsymbol{x}}_{ij}(t) = \vec{\boldsymbol{x}}_i(t) - \vec{\boldsymbol{x}}_j(t)$.

$\vec{\boldsymbol{x}}$ is E(3)-equivariant and $\boldsymbol{h}$ is E(3)-invariant.



Equivariant Spatial Module

Message Passing Alternates between ESM and ETM



Equivariant Temporal Module

Temporal Dependency

$$\alpha_i^{(l)}(ts) = \frac{\exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))}{\sum_{s=0}^{t} \exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))},$$

$$\boldsymbol{h}_i^{(l+1)}(t) = \boldsymbol{h}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts) \boldsymbol{v}_i^{(l)}(s),$$

$$\vec{\boldsymbol{x}}_i^{(l+1)}(t) = \vec{\boldsymbol{x}}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts) \vec{\boldsymbol{x}}_i^{(l)}(ts) \phi_x(\boldsymbol{v}_i^{(l)}(s)),$$

where, $\vec{\boldsymbol{x}}_i(ts) = \vec{\boldsymbol{x}}_i(t) - \vec{\boldsymbol{x}}_i(s), \boldsymbol{q}_i^{(l)}(t) = \phi_q\left(\boldsymbol{h}_i^{(l)}(t)\right).$

$\vec{\boldsymbol{x}}$ is E(3)-equivariant and $\boldsymbol{h}$ is E(3)-invariant.

Equivariant Temporal Pooling Layer for the Final Prediction

- This layer applies a learned weighted combination of past node coordinates, where the translation ensures the pooling is **equivariant**.

$$\vec{\boldsymbol{x}}_i^*(T) = \hat{\boldsymbol{X}}_i \mathbf{w} + \vec{\boldsymbol{x}}_i^{(L)}(T-1), \mathbf{w} \in \mathbb{R}^{T-1},$$

Where,

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

- ESTAG is trained via the mean squared error (MSE) loss:

$$\mathcal{L} = \sum_{i=1}^{N} ||\vec{\boldsymbol{x}}_i(T) - \vec{\boldsymbol{x}}_i^*(T)||_2^2.$$

Outline

Molecular Level: ESTAG Achieves the Lowest Prediction Error

Table: 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	0.669	0.119	0.428	0.550	0.625	0.630	0.591	0.597
ST_EGNN	0.735	0.163	0.245	0.427	0.745	0.687	0.553	0.445
EqMotion	0.721	0.156	0.476	0.600	0.747	0.697	0.691	0.681
STGCN	0.715	0.106	0.456	0.596	0.736	0.682	0.687	0.673
AGL-STAN	0.719	0.106	0.459	0.596	0.601	0.452	0.683	0.515
ESTAG	**0.063**	**0.003**	**0.099**	**0.101**	**0.068**	**0.047**	**0.079**	**0.066**

Protein & Macro Levels: ESTAG Also Performs Well

Table: 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

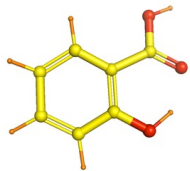
Table: Prediction error ($\times 10^{-1}$) on Motion dataset. Results averaged across 3 runs.

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**

Visualization–MD17(SALICYLIC)

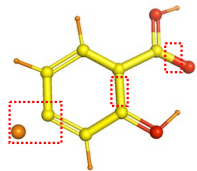
We highlight the main difference in red rectangles, such as the absence and incorrect types of chemical bonds.

Ground Truth



ESTAG

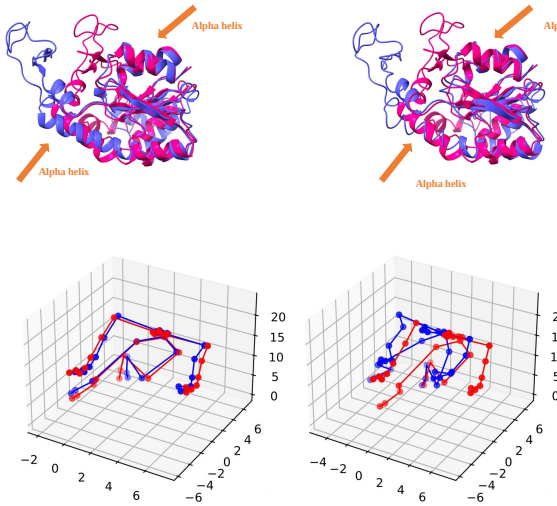
MSE=0.088



ST_EGNN

MSE=0.654

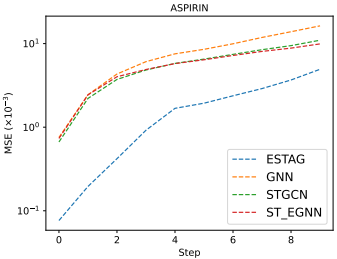
Visualization–Protein & Motion

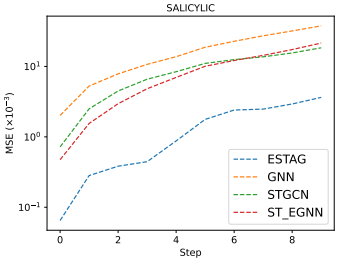


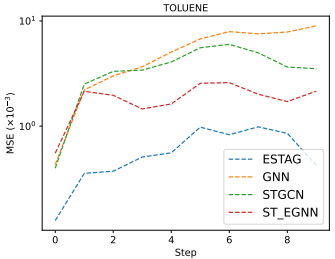
Comparison between ESTAG and ST_EGNN. The ground truths are in red while the predicted states are in blue.

Visualization–Rollout

Here we recurrently predict the future frames in a rollout manner, where the currently-predicted frame will be used as the input for the next frame prediction.







Outline

Future Improvements

- Performance enhancements are required when dealing with **more challenging systems**.
  - The superiority of ESTAG on protein dataset is not as obvious as that on molecule and motion datasets, owing to various kinds of physical interaction between different amino acids, let along each amino acid compose of a certain number of atoms

- It is also promising to extend our model to multi-scale GNN, which is useful particularly for industrial-level applications involving **huge graphs**.

The End

Thanks!!