Roto-translated Local Coordinate Frames For Interacting Dynamical Systems

Miltiadis Kofinas[1], Naveen Shankar Nagaraja[2], Efstratios Gavves[1]

[1]VIS Lab
University of Amsterdam
Amsterdam, Netherlands

[2]Department of Autonomous Driving
BMW Group
Munich, Germany

Geometric Deep Learning Study Visit, 2 June 2022

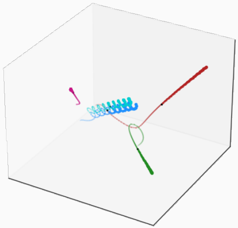
UNIVERSITY OF AMSTERDAM

BMW GROUP THE NEXT 100 YEARS

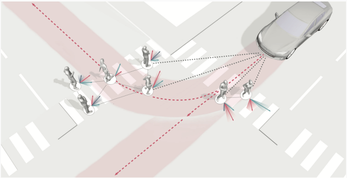
Rolls-Royce Motor Cars Limited

Interacting systems are everywhere

- Colliding particles
- N-body systems
- Molecules
- Motion capture
- Traffic scenes



3D Charged particles [6]



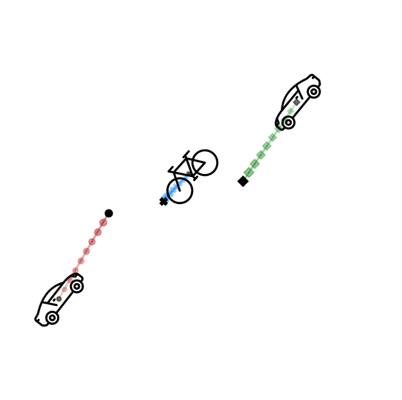
Traffic scene [8]

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

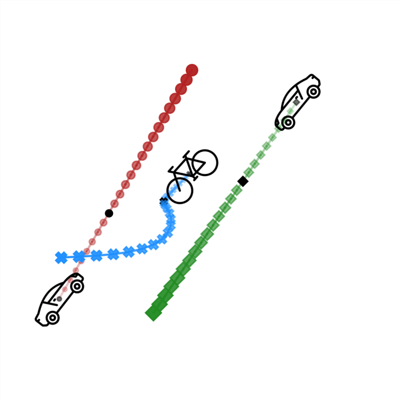
[8] Tim Salzmann†, Boris Ivanovic†, et al. "Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data". In: *ECCV*. 2020

Future forecasting

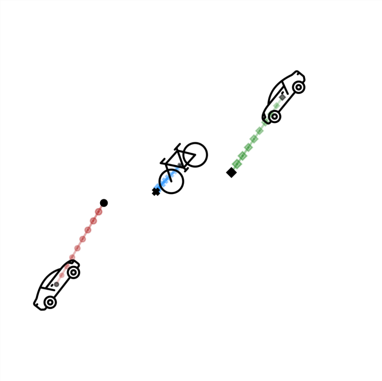
Past

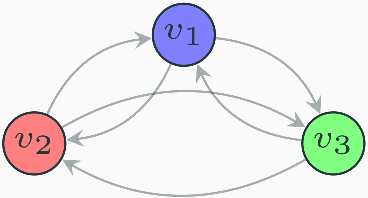


Future



Geometric graphs





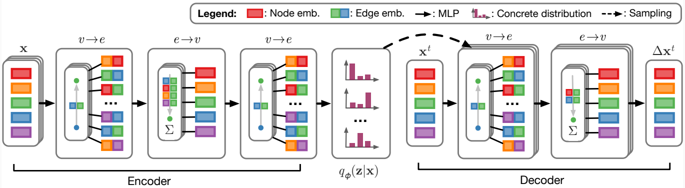
Geometric graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$$

$$\mathcal{V} = \{v_i\}_{i=1}^{N}, \quad \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$$

$$\mathbf{X} = \left\{ \begin{pmatrix} \mathbf{p}_i, & \text{position} \\ \mathbf{u}_i, & \text{velocity} \end{pmatrix} \right\}_{i=1}^{N}$$

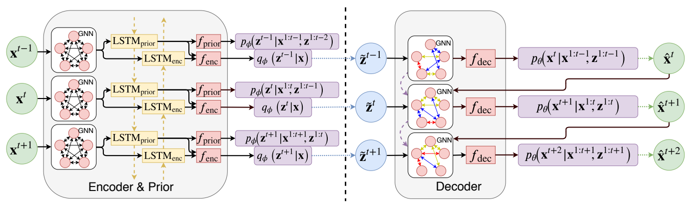
Related work - Neural Relational Inference [6]



- Explicitly infer graph structure over latent edge types
- Simultaneously learn the dynamical system

[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014
[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

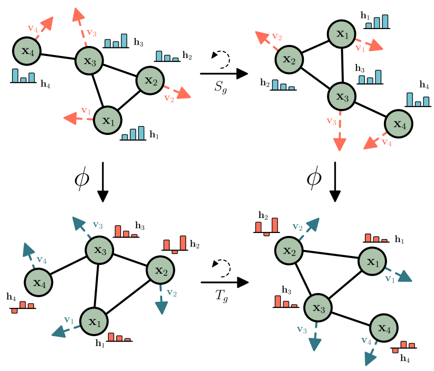
Related work - Dynamic Neural Relational Inference [4]



- Dynamic relations through time
- Sequential approximate posterior based on past states

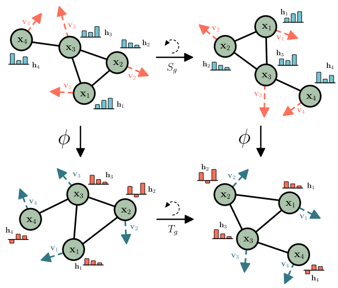
[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

Related work - E(n) Equivariant Graph Networks [9]



- Leverage rotation equivariant relative positions and invariant euclidean distances

[9] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. "E(n) Equivariant Graph Neural Networks". In: *ICML*. 2021



$$\mathbf{m}_{j,i} = \phi_e\left(\mathbf{h}_i^l, \mathbf{h}_j^l, \boxed{\left\|\mathbf{p}_j^l - \mathbf{p}_i^l\right\|_2^2}\right)$$

$$\mathbf{p}_i^{l+1} = \mathbf{p}_i^l + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \boxed{\left(\mathbf{p}_j^l - \mathbf{p}_i^l\right)} \cdot \phi_x(\mathbf{m}_{j,i})$$

$$\mathbf{m}_i = \sum_{j\in\mathcal{N}(i)} \mathbf{m}_{j,i}$$

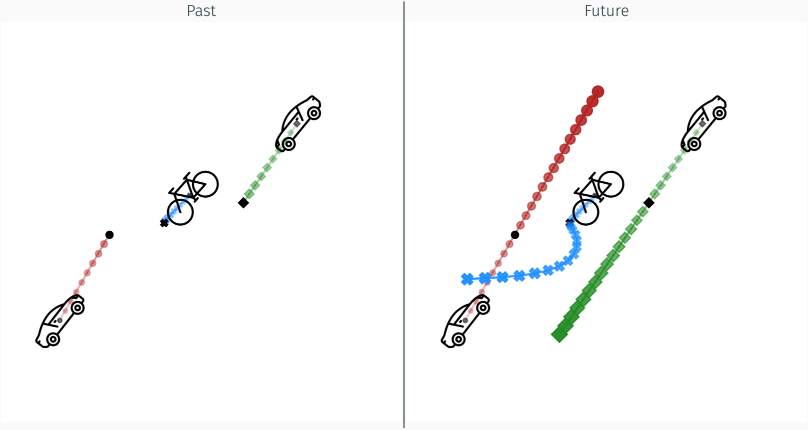
$$\mathbf{h}_i^{l+1} = \phi_h\left(\mathbf{h}_i^l, \mathbf{m}_i\right)$$

- Leverage rotation equivariant relative positions and invariant euclidean distances

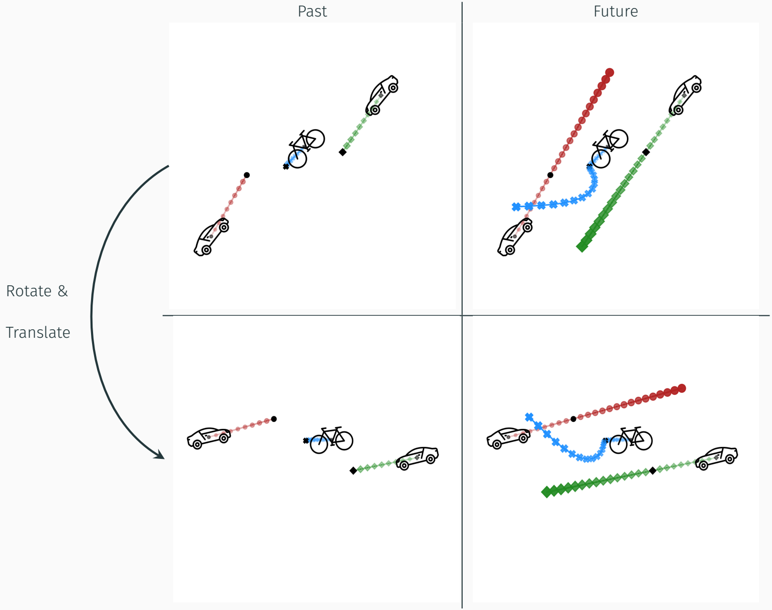
[9] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. “E(n) Equivariant Graph Neural Networks”. In: *ICML*. 2021

Motivation

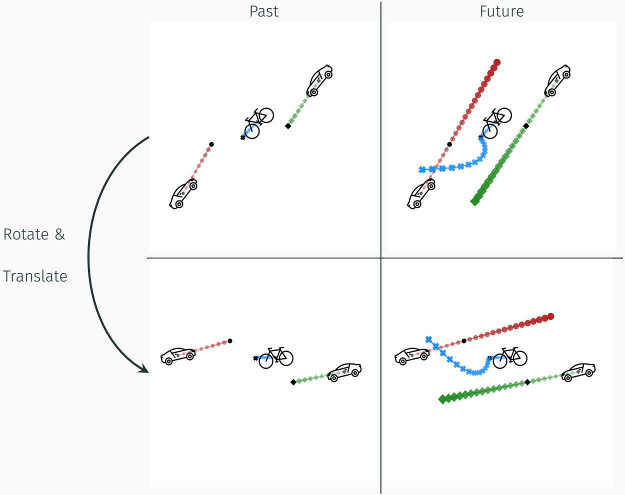
What happens when we rotate/translate the inputs?



Motivation



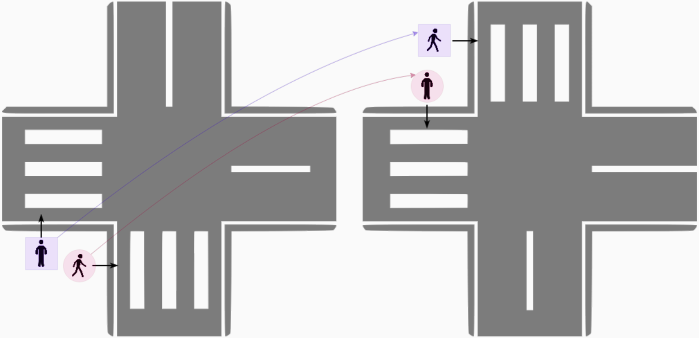
Motivation



Roto-translation equivariance

Dynamics do not change under rotations and translations

Motivation



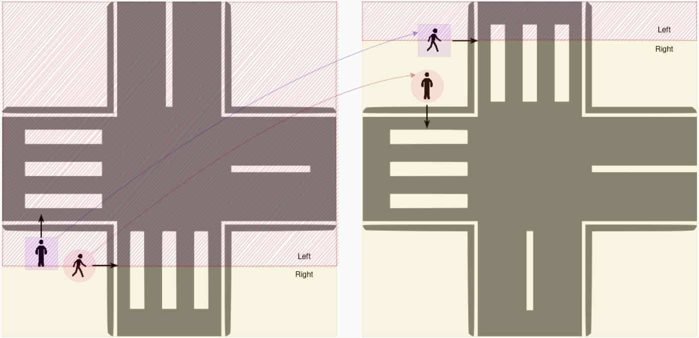
Ego-centric perspective

Objects operate in ego-centric and asymmetric views of the world

Global coordinate frames

Graphs embedded in arbitrary global coordinate frames

Motivation



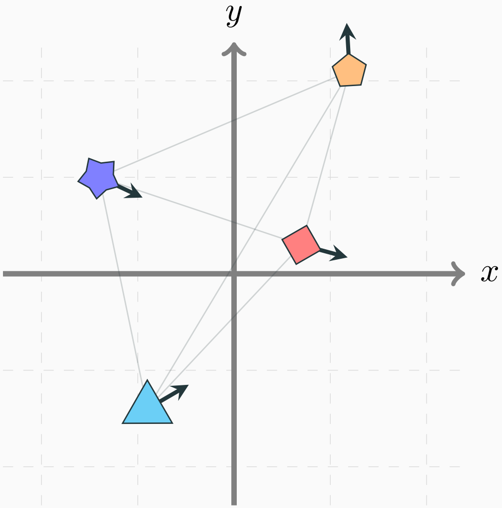
Ego-centric perspective

Objects operate in ego-centric and asymmetric views of the world

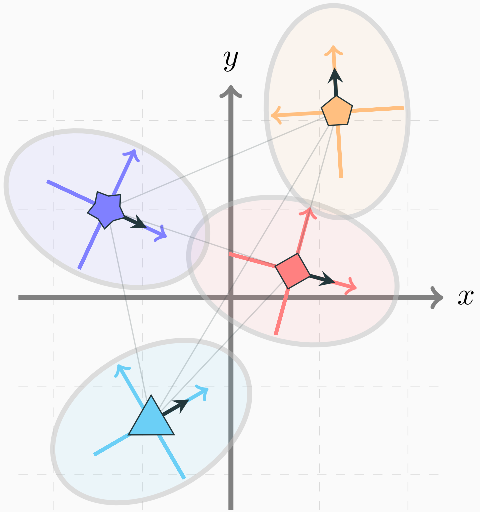
Global coordinate frames

Graphs embedded in arbitrary global coordinate frames

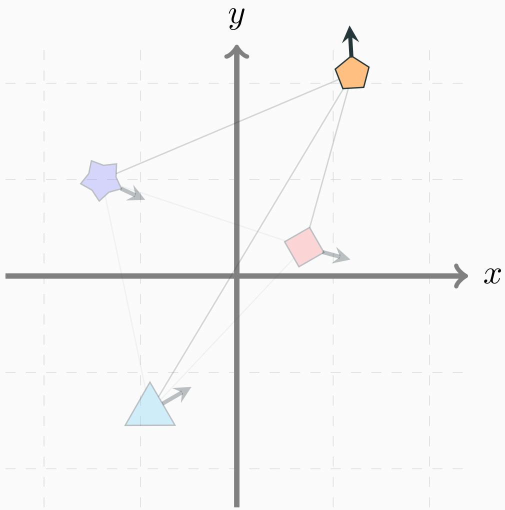
Local coordinate frames (LoCS)



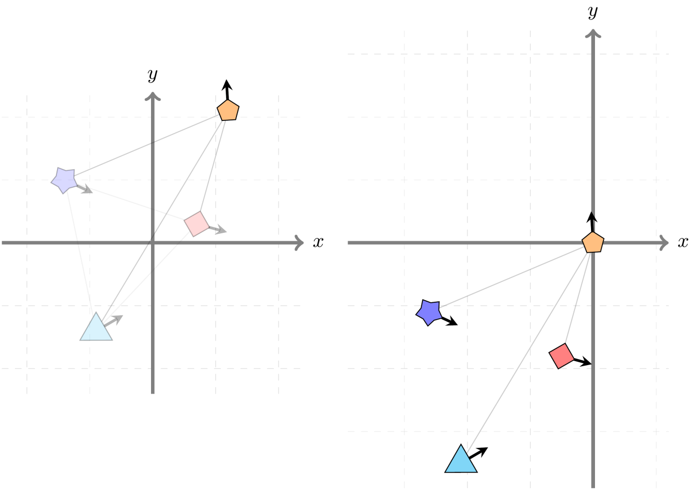
Local coordinate frames (LoCS)



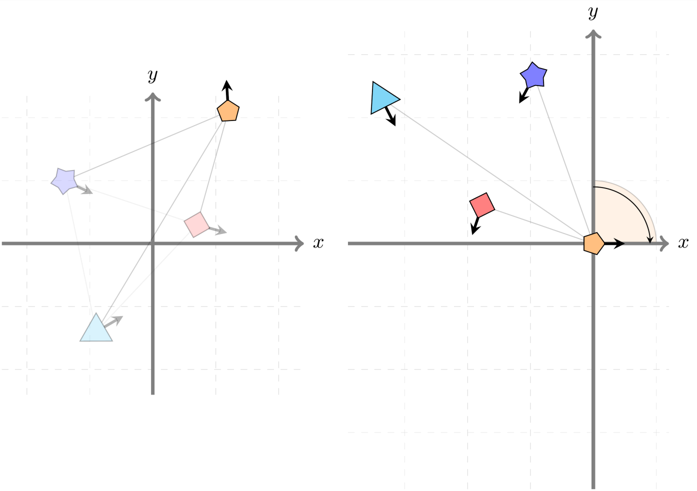
Local coordinate frames (LoCS)



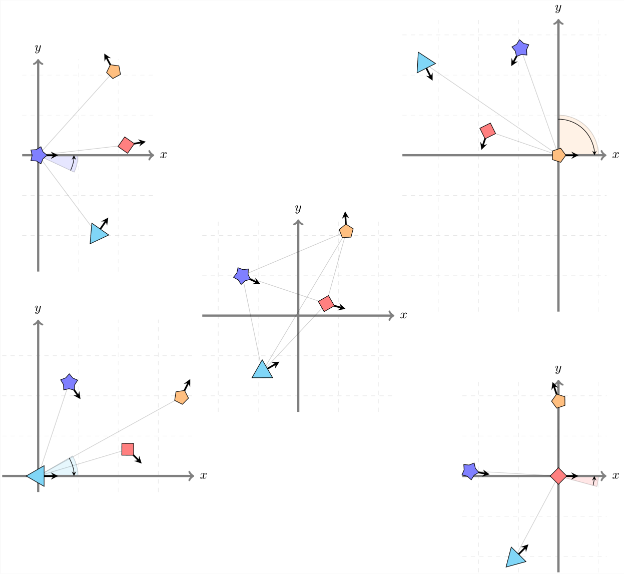
Local coordinate frames (LoCS)



Local coordinate frames (LoCS)



Local coordinate frames (LoCS)



Local coordinate frames (LoCS)

Node states

$$\mathbf{x}_j^t = \left\{ \begin{array}{ll} \mathbf{p}_j^t, & \text{position} \\ \mathbf{u}_j^t, & \text{velocity} \end{array} \right\}$$

$$\mathbf{v}_j^t = \left\{ \begin{array}{lr} \mathbf{p}_j^t, & \text{position} \\ \mathbf{u}_j^t, & \text{velocity} \\ \boldsymbol{\omega}_j^t, & \text{orientation} \end{array} \right\}$$

Relative positions $\qquad \mathbf{r}_{j,i}^t = \mathbf{p}_j^t - \mathbf{p}_i^t$

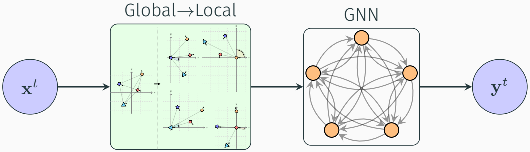
Rotation matrix $\qquad \mathbf{Q}_i^t = \mathbf{Q}\left(\boldsymbol{\omega}_i^t\right)$

Local state $\qquad \mathbf{v}_{j|i}^t = \begin{pmatrix} \mathbf{Q}_i^{t\top} \cdot \mathbf{r}_{j,i}^t \\ \mathbf{Q}_i^{t\top} \cdot \boldsymbol{\omega}_j^t \\ \mathbf{Q}_i^{t\top} \cdot \mathbf{u}_j^t \end{pmatrix}$

Graph Networks in local coordinate frames

Node states

$$\mathbf{x}_j^t = \left(\mathbf{p}_j^t, \mathbf{u}_j^t\right) \qquad \mathbf{v}_j^t = \left(\mathbf{p}_j^t, \mathbf{u}_j^t, \boldsymbol{\omega}_j^t\right)$$



Invariant GNN

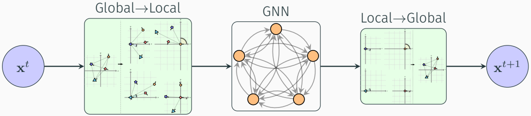
$$\mathbf{v}_{j|i}^t = \textsc{Global2Local}\left(\mathbf{v}_j^t, \mathbf{v}_i^t\right)$$

$$\mathbf{y}_{i|i}^t = \mathrm{GNN}\left(\mathbf{v}_{i|i}^t, \left\{\mathbf{v}_{j|i}^t\right\}_{j \in \mathcal{N}(i)}\right)$$

Graph Networks in local coordinate frames

Node states

$$\mathbf{x}_j^t = \left(\mathbf{p}_j^t, \mathbf{u}_j^t\right) \qquad \mathbf{v}_j^t = \left(\mathbf{p}_j^t, \mathbf{u}_j^t, \boldsymbol{\omega}_j^t\right)$$



Equivariant GNN

$$\mathbf{v}_{j|i}^t = \textsc{Global2Local}\left(\mathbf{v}_j^t, \mathbf{v}_i^t\right)$$

$$\boldsymbol{\Delta}\mathbf{x}_{i|i}^{t+1} = \mathrm{GNN}\left(\mathbf{v}_{i|i}^t, \left\{\mathbf{v}_{j|i}^t\right\}_{j\in\mathcal{N}(i)}\right)$$

$$\boldsymbol{\Delta}\mathbf{x}_i^{t+1} = \textsc{Local2Global}\left(\boldsymbol{\Delta}\mathbf{x}_{i|i}^{t+1}\right)$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \boldsymbol{\Delta}\mathbf{x}_i^{t+1}$$

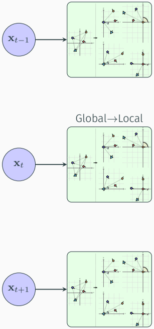
Graph network formulation

$$
\begin{aligned}
\mathbf{h}_{j,i} &= f_e\big(\big[\mathbf{v}_{j|i}, \mathbf{v}_{i|i}\big]\big) \\
\mathbf{h}_i &= f_v\left(g_v\big(\mathbf{v}_{i|i}\big) + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{h}_{j,i}\right)
\end{aligned}
$$

Local coordinate frames as Neural relational inference models

[5] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *ICLR*. 2014
[6] Thomas Kipf†, Ethan Fetaya†, et al. “Neural relational inference for interacting systems”. In: *ICML*. 2018
[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

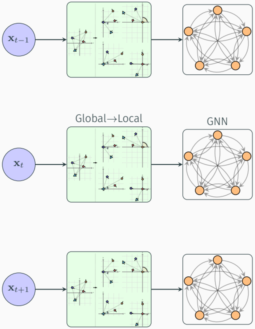


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

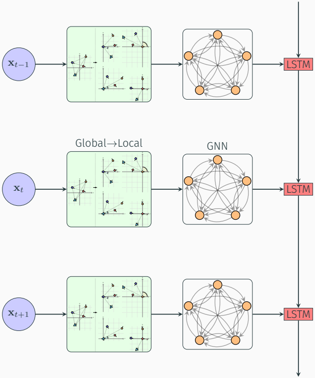


[5] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. “Neural relational inference for interacting systems”. In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

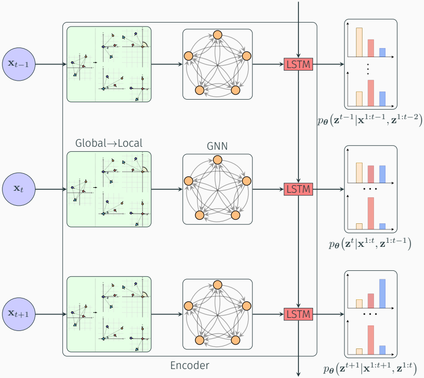


[5] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. “Neural relational inference for interacting systems”. In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

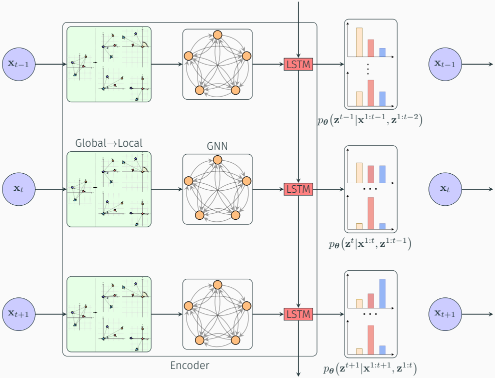


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

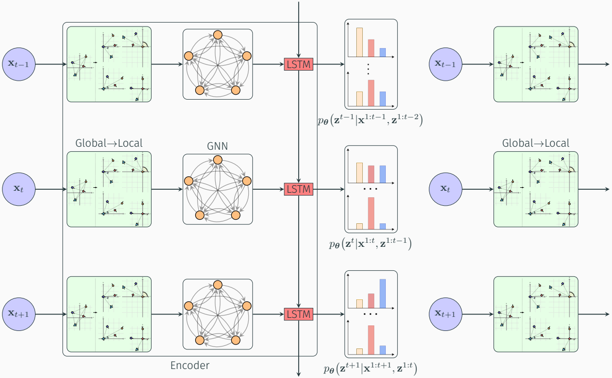


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

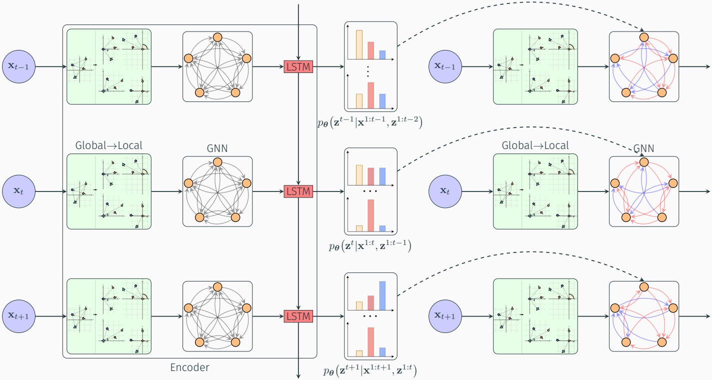


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

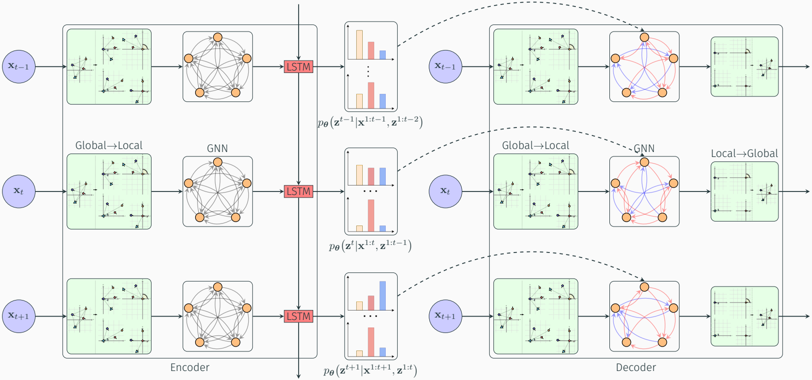


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models

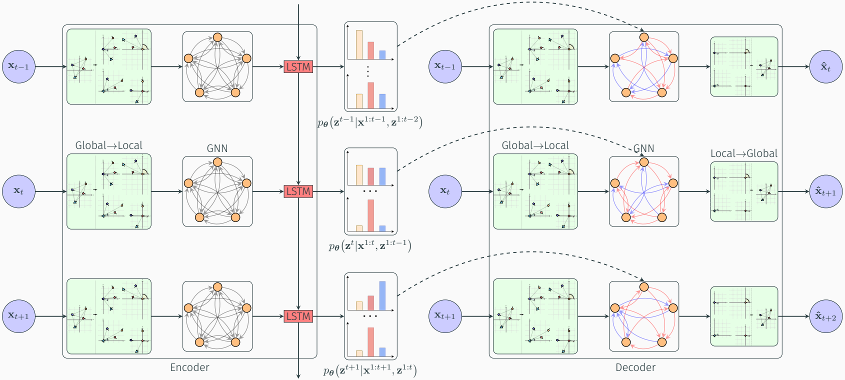


[5] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Local coordinate frames as Neural relational inference models



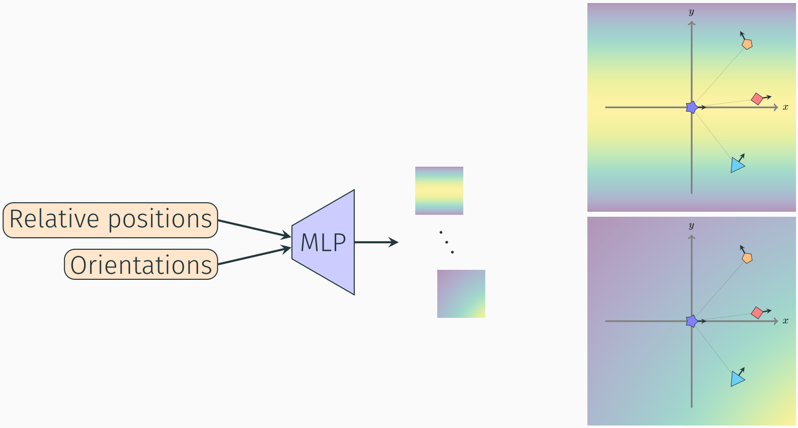
[5] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *ICLR*. 2014

[6] Thomas Kipf†, Ethan Fetaya†, et al. “Neural relational inference for interacting systems”. In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

Anisotropic continuous filtering in local coordinate frames

Directionality in graphs $\Longrightarrow$ Anisotropic filtering



Experiments

Synthetic [1]

- 2D, repulsive forces

InD [1]

- Traffic scenes, 2D, social interactions

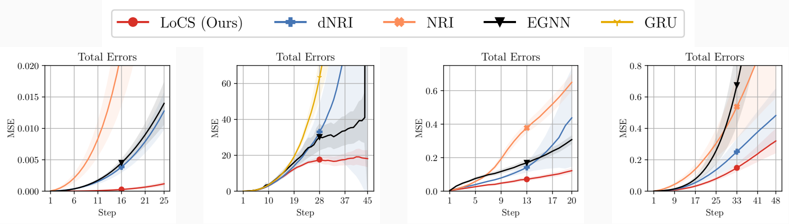
Charged particles [6]

- 3D, electrostatic forces

CMU Motion capture [3]

- 3D, subject #35, walking trajectories

Results



(a) Synthetic [4] (b) Traffic scenes, inD [1] (c) Charged particles [6] (d) Motion Capture [3]

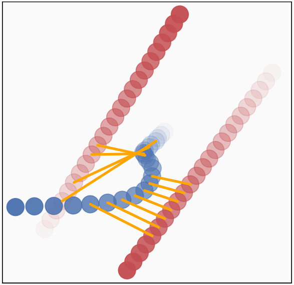
[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

[1] Julian Bock et al. "The inD dataset: A drone dataset of naturalistic road user trajectories at german intersections". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020

[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018

[3] CMU. *Carnegie-Mellon Motion Capture Database*. 2003. URL: http://mocap.cs.cmu.edu

Results - Synthetic Dataset



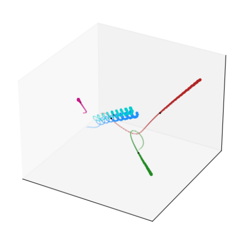
Relation prediction F1 score on synthetic dataset

Method	NRI	dNRI	LoCS
F1	26.5	60.8	88.9

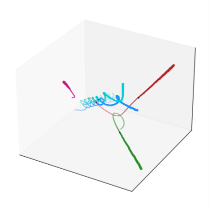
[6] Thomas Kipf†, Ethan Fetaya†, et al. "Neural relational inference for interacting systems". In: *ICML*. 2018
[4] Colin Graber and Alexander G Schwing. "Dynamic Neural Relational Inference". In: *CVPR*. 2020

Qualitative results - charged particles

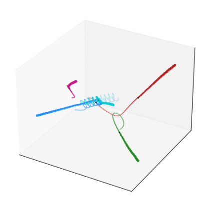
Groundtruth



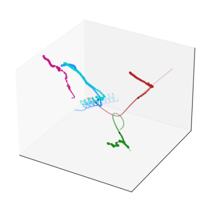
LoCS (Ours)



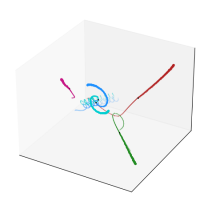
dNRI [4]



NRI [6]



EGNN [9]



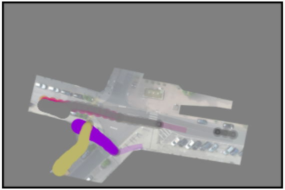
[6] Thomas Kipf†, Ethan Fetaya†, et al. “Neural relational inference for interacting systems”. In: *ICML*. 2018

[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

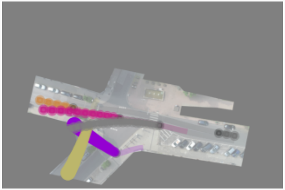
[9] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. “E(n) Equivariant Graph Neural Networks”. In: *ICML*. 2021

Qualitative results - inD

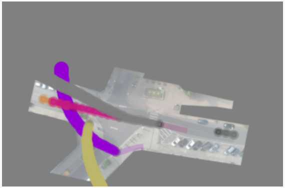
Groundtruth



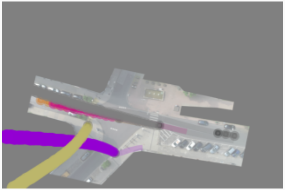
LoCS (Ours)



dNRI [4]



EGNN [9]



[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020

[9] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. “E(n) Equivariant Graph Neural Networks”. In: *ICML*. 2021

Conclusion

- Local coordinate frames for all objects
- Invariance/equivariance to global roto-translations
- Anisotropic continuous filters in local coordinate frames
- Demonstrate effectiveness on a range of 2D/3D settings

- Paper: https://arxiv.org/abs/2110.14961
- Source code: https://github.com/mkofinas/locs

References i

[1] Julian Bock et al. “The inD dataset: A drone dataset of naturalistic road user trajectories at german intersections”. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020.

[2] Kyunghyun Cho et al. “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014.

[3] CMU. *Carnegie-Mellon Motion Capture Database*. 2003. URL: `http://mocap.cs.cmu.edu`.

[4] Colin Graber and Alexander G Schwing. “Dynamic Neural Relational Inference”. In: *CVPR*. 2020.

[5] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *ICLR*. 2014.

[6] Thomas Kipf†, Ethan Fetaya†, Kuan-Chieh Wang, Max Welling, and Richard Zemel. “Neural relational inference for interacting systems”. In: *ICML*. 2018.

[7] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. “Stochastic backpropagation and approximate inference in deep generative models”. In: *ICML*. 2014.

[8] Tim Salzmann†, Boris Ivanovic†, Punarjay Chakravarty, and Marco Pavone. “Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data”. In: *ECCV*. 2020.

[9] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. "E(n) Equivariant Graph Neural Networks". In: *ICML*. 2021.

[10] Martin Simonovsky and Nikos Komodakis. "Dynamic edge-conditioned filters in convolutional neural networks on graphs". In: *CVPR*. 2017.