Group Equivariant Deep Learning

Lecture 3 - Equivariant graph neural networks

Lecture 3.5 - 3D Steerable (graph) convolutions

Erik Bekkers, Amsterdam Machine Learning Lab, University of Amsterdam

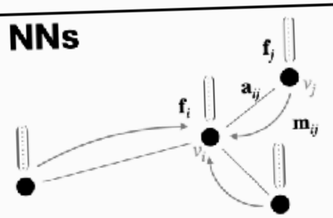
This mini-course serves as a module with the UvA Master AI course Deep Learning 2 https://uvadl2c.github.io/

Analysis of literature in terms of non-linear group convolutions

Lecture 2.2

Linear vs non-linear (group) convolutions

Message passing NNs



Compute messages: $\mathbf{m}_{ij} = \phi_m(\mathbf{f}_i, \mathbf{f}_j, \mathbf{a}_{ij})$

Aggregate and update: $\mathbf{f}'_i = \phi_f\left(\mathbf{f}_i, \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}\right)$

Classic point convolutions

(Lecture 1.7: regular g-convs on homogeneous spaces)

$$\mathbf{m}_{ij} = \mathbf{W}(\|\mathbf{x}_j - \mathbf{x}_i\|)\mathbf{f}_j$$

$$\mathbf{W}(g_i^{-1}g_j)\mathbf{f}_j$$

Linear convolution

Steerable G-CNNs

(Lecture 2: steerable g-convs)

$\mathcal{F}_H$

$$\mathbf{m}_{ij} = \mathbf{W}_{\hat{\mathbf{a}}_{ij}}(\|\mathbf{x}_j - \mathbf{x}_i\|)\hat{\mathbf{f}}_j$$

$$:= \hat{\mathbf{f}}_j \otimes_{cg}^{\mathbf{W}(\|\mathbf{x}_j - \mathbf{x}_i\|)} \hat{\mathbf{a}}_{ij}$$

Invariant Message **Passing NNs**

(Lecture 3)

$$\mathbf{m}_{ij} = \text{MLP}(\mathbf{f}_i, \mathbf{f}_j, \|\mathbf{x}_j - \mathbf{x}_i\|)$$

Non-linear "convolution"

Equivariant (Steerable) Message **Passing**

(Lecture 3)

$$\hat{\mathbf{m}}_{ij} = \widehat{\text{MLP}}\,(\hat{\mathbf{f}}_i, \hat{\mathbf{f}}_j, \mathbf{x}_j - \mathbf{x}_i)$$

With steerable MLP:

$$\widehat{\text{MLP}}_{\hat{\mathbf{a}}_{ij}}(\hat{\mathbf{f}}_i, \hat{\mathbf{f}}_j, \|\mathbf{x}_j - \mathbf{x}_i\|) := \sigma(\mathbf{W}^{(n)}_{\hat{\mathbf{a}}_{ij}}(\ldots(\sigma(\mathbf{W}^{(1)}_{\hat{\mathbf{a}}_{ij}}\hat{\mathbf{h}}_i))))$$

29

GEOMETRIC AND PHYSICAL QUANTITIES IMPROVE E(3) EQUIVARIANT MESSAGE PASSING

Johannes Brandstetter*
University of Amsterdam
Johannes Kepler University Linz
brandstetter@ml.jku.at

Rob Hesselink*
University of Amsterdam
r.d.hesselink@uva.nl

Elise van der Pol
UvA-Bosch DeltaLab
University of Amsterdam
e.e.vanderpol@uva.nl

Erik J Bekkers
University of Amsterdam
e.j.bekkers@uva.nl

Max Welling
UvA-Bosch DeltaLab
University of Amsterdam
m.welling@uva.nl

ABSTRACT

Including covariant information, such as position, force, velocity or spin is important in many tasks in computational physics and chemistry. We introduce Steerable E(3) Equivariant Graph Neural Networks (SEGNNs) that generalise equivariant graph networks, such that node and edge attributes are not restricted to invariant scalars, but can contain covariant information, such as vectors or tensors. This model, composed of steerable MLPs, is able to incorporate geometric and physical information in both the message and update functions. Through the definition of steerable node attributes, the MLPs provide a new class of activation functions for general use with steerable feature fields. We discuss ours and related work through the lens of *equivariant non-linear convolutions*, which further allows us to pin-point the successful components of SEGNNs: *non-linear* message aggregation improves upon classic *linear* (steerable) point convolutions; *steerable messages* improve upon recent equivariant graph networks that send invariant messages. We demonstrate the effectiveness of our method on several tasks in computational physics and chemistry and provide extensive ablation studies.

1 INTRODUCTION

The success of Convolutional Neural Networks (CNNs) (LeCun et al., 1998; 2015; Schmidhuber, 2015; Krizhevsky et al., 2012) is a key factor for the rise of deep learning, attributed to their capability of exploiting translation symmetries, thereby introducing a strong inductive bias. Recent work has shown that designing CNNs to exploit additional symmetries via group convolutions has even further increased their performance (Cohen & Welling, 2016; 2017; Worrall et al., 2017; Cohen et. al., 2018; Kondor & Trivedi, 2018; Weiler et al., 2018; Bekkers et al., 2018; Bekkers, 2019; Weiler & Cesa, 2019). Graph neural networks (GNNs) and CNNs are closely related to each other via their aggregation of local information. More precisely, CNNs can be formulated as message passing layers (Gilmer et al., 2017) based on a sum aggregation of messages that are obtained by relative position-dependent *linear* transformations of neighbouring node features. The power of message passing layers is, however, that node features are transformed and propagated in a highly *non-linear* manner. Equivariant GNNs have been proposed before as either PointConv-type (Wu et al., 2019; Kristof et al., 2017) implementations of steerable (Thomas et al., 2018; Anderson et al., 2019; Fuchs et al., 2020) or regular group convolutions (Finzi et al., 2020). The most important component in these methods are the convolution layers. Although powerful, such layers only (pseudo[1]) linearly transform the graphs and non-linearity is only obtained via point-wise activations.

[1]Methods such as SE(3)-transformers (Fuchs et al., 2020) and Cormorant (Anderson et al., 2019) include an input-dependent attention component that augments the convolutions.

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** between
feature maps on homogeneous spaces
is a **group convolution**"

Table 2: Performance comparison on the QM9 dataset. N
Error (MAE) between model predictions and ground truth

			Task	α	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	μ	C_ν
			Units	bohr^3	meV	meV	meV	D	cal/mol
non-linear		no geometry	NMP	.092	69	43	38	.030	.040
	regular	$\mathbb{R}^3$	SchNet *	.235	63	41	34	.033	.033
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	34	38	.038	.026
	steerable	$SE(3)$	L1Net	.088	68	46	35	.043	.031
	regular	G	LieConv	.084	49	30	25	.032	.038
	steerable	$SE(3)$	TFN	.223	58	40	38	.064	.101
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	35	33	.051	.054
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	24	19	.029	.023
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	**23**	**18**	.026	**.021**
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	27	20	**.012**	.024
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	29	25	.029	.031
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	24	21	.023	.031

Table 2: Comparison on QM9.

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** bet
feature maps on homogeneous sp
is a **group convolution**"

Table 2: Performance comparis
Error (MAE) between model p

			Task	α	$\Delta\varepsilon$	ε_{HO}
			Units	bohr^3	meV	m
non-linear		no geometry	NMP	.092	69	
	regular	$\mathbb{R}^3$	SchNet *	.235	63	
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	
	steerable	$SE(3)$	L1Net	.088	68	
	regular	G	LieConv	.084	49	
	steerable	$SE(3)$	TFN	.223	58	
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	

Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds

Nathaniel Thomas*
Stanford University
Stanford, California, USA
ncthomas@stanford.edu

Tess Smidt*
University of California, Berkeley
Berkeley, California, USA
Lawrence Berkeley National Lab
Berkeley, California, USA
tsmidt@berkeley.edu

Steven Kearnes
Google
Mountain View, California, USA
kearnes@google.com

Lusann Yang
Google
Mountain View, California, USA
lusann@google.com

Li Li
Google
Mountain View, California, USA
leeley@google.com

Kai Kohlhoff
Google
Mountain View, California, USA
kohlhoff@google.com

Patrick Riley
Google
Mountain View, California, USA
pfr@google.com

Abstract

We introduce tensor field neural networks, which are locally equivariant to 3D rotations, translations, and permutations of points at every layer. 3D rotation equivariance removes the need for data augmentation to identify features in arbitrary orientations. Our network uses filters built from spherical harmonics; due to the mathematical consequences of this filter choice, each layer accepts as input (and guarantees as output) scalars, vectors, and higher-order tensors, in the geometric sense of these terms. We demonstrate the capabilities of tensor field networks with tasks in geometry, physics, and chemistry.

1 Motivation

Convolutional neural networks are translation-equivariant, which means that features can be identified anywhere in a given input. This capability has contributed significantly to their widespread success.

In this paper, we present a family of networks that enjoy richer equivariance: the symmetries of 3D Euclidean space. This includes 3D rotation equivariance (the ability to identify a feature in any 3D rotation and its orientation) and 3D translation equivariance.

*Equal contribution.

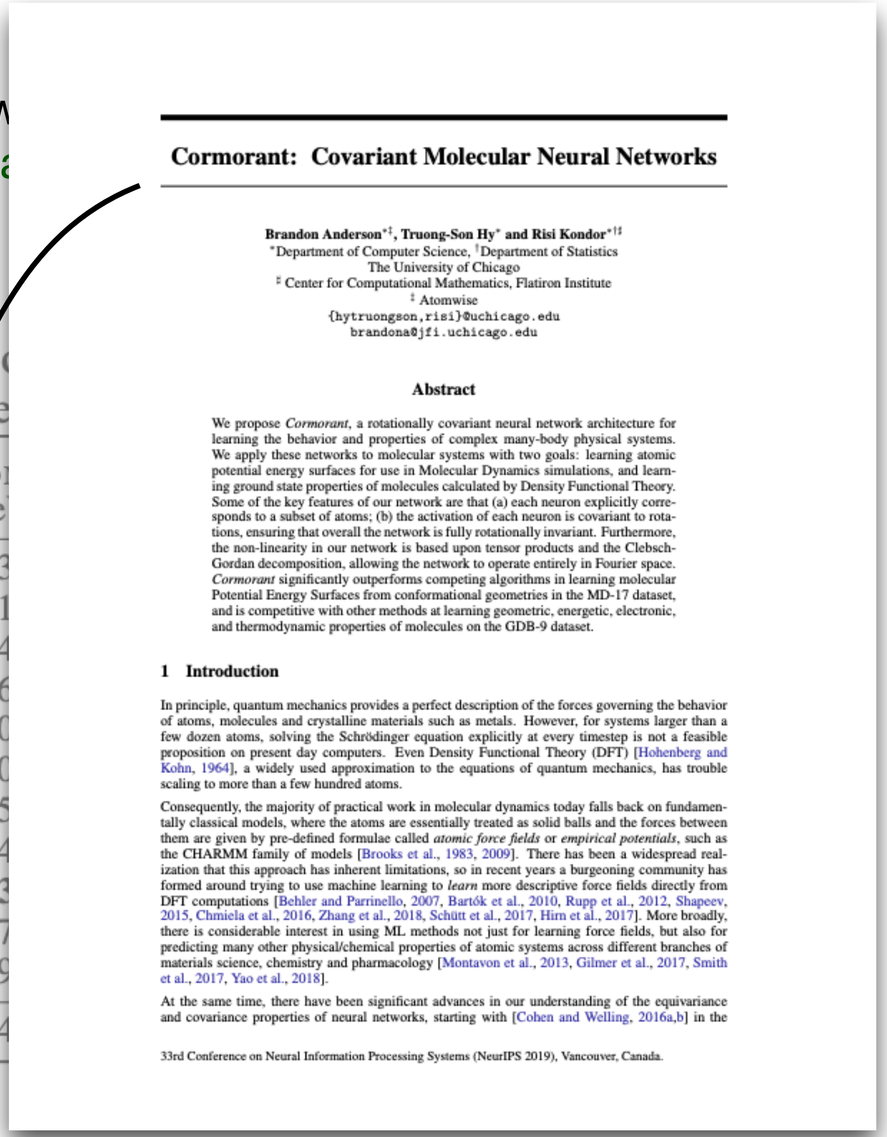
Preprint. Work in progress.

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** betw
feature maps on homogeneous spa
is a **group convolution**"

Table 2: Performance comparis
Error (MAE) between model pre

			Task	α	$\Delta\varepsilon$	ε_{HO}
			Units	bohr^3	meV	me
non-linear		no geometry	NMP	.092	69	43
	regular	$\mathbb{R}^3$	SchNet *	.235	63	41
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	34
	steerable	$SE(3)$	L1Net	.088	68	46
	regular	G	LieConv	.084	49	30
	steerable	$SE(3)$	TFN	.223	58	40
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	35
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	24
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	23
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	27
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	29
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	24



Cormorant: Covariant Molecular Neural Networks

Brandon Anderson[*‡], **Truong-Son Hy**[*] **and Risi Kondor**[*†‡]
[*]Department of Computer Science, [†]Department of Statistics
The University of Chicago
[‡] Center for Computational Mathematics, Flatiron Institute
[‡] Atomwise
{hytruongson,risi}@uchicago.edu
brandona@jfi.uchicago.edu

Abstract

We propose *Cormorant*, a rotationally covariant neural network architecture for learning the behavior and properties of complex many-body physical systems. We apply these networks to molecular systems with two goals: learning atomic potential energy surfaces for use in Molecular Dynamics simulations, and learning ground state properties of molecules calculated by Density Functional Theory. Some of the key features of our network are that (a) each neuron explicitly corresponds to a subset of atoms; (b) the activation of each neuron is covariant to rotations, ensuring that overall the network is fully rotationally invariant. Furthermore, the non-linearity in our network is based upon tensor products and the Clebsch-Gordan decomposition, allowing the network to operate entirely in Fourier space. *Cormorant* significantly outperforms competing algorithms in learning molecular Potential Energy Surfaces from conformational geometries in the MD-17 dataset, and is competitive with other methods at learning geometric, energetic, electronic, and thermodynamic properties of molecules on the GDB-9 dataset.

1 Introduction

In principle, quantum mechanics provides a perfect description of the forces governing the behavior of atoms, molecules and crystalline materials such as metals. However, for systems larger than a few dozen atoms, solving the Schrödinger equation explicitly at every timestep is not a feasible proposition on present day computers. Even Density Functional Theory (DFT) [Hohenberg and Kohn, 1964], a widely used approximation to the equations of quantum mechanics, has trouble scaling to more than a few hundred atoms.

Consequently, the majority of practical work in molecular dynamics today falls back on fundamentally classical models, where the atoms are essentially treated as solid balls and the forces between them are given by pre-defined formulae called *atomic force fields* or *empirical potentials*, such as the CHARMM family of models [Brooks et al., 1983, 2009]. There has been a widespread realization that this approach has inherent limitations, so in recent years a burgeoning community has formed around trying to use machine learning to *learn* more descriptive force fields directly from DFT computations [Behler and Parrinello, 2007, Bartók et al., 2010, Rupp et al., 2012, Shapeev, 2015, Chmiela et al., 2016, Zhang et al., 2018, Schütt et al., 2017, Hirn et al., 2017]. More broadly, there is considerable interest in using ML methods not just for learning force fields, but also for predicting many other physical/chemical properties of atomic systems across different branches of materials science, chemistry and pharmacology [Montavon et al., 2013, Gilmer et al., 2017, Smith et al., 2017, Yao et al., 2018].

At the same time, there have been significant advances in our understanding of the equivariance and covariance properties of neural networks, starting with [Cohen and Welling, 2016a,b] in the

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** betw
feature maps on homogeneous spa
is a **group convolution**"

Table 2: Performance compari
Error (MAE) between model pre

			Task	α	$\Delta\varepsilon$	ε_{HO}
			Units	bohr^3	meV	me
non-linear		no geometry	NMP	.092	69	43
	regular	$\mathbb{R}^3$	SchNet *	.235	63	41
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	34
	steerable	$SE(3)$	L1Net	.088	68	46
	regular	G	LieConv	.084	49	30
	steerable	$SE(3)$	TFN	.223	58	40
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	35
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	24
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	**23**
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	27
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	29
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	24

arXiv:2006.10503v3 [cs.LG] 24 Nov 2020

SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks

Fabian B. Fuchs[*†]
Bosch Center for Artificial Intelligence
A2I Lab, Oxford University
fabian@robots.ox.ac.uk

Daniel E. Worrall[*]
Amsterdam Machine Learning Lab, Philips Lab
University of Amsterdam
d.e.worrall@uva.nl

Volker Fischer
Bosch Center for Artificial Intelligence
volker.fischer@de.bosch.com

Max Welling
Amsterdam Machine Learning Lab
University of Amsterdam
m.welling@uva.nl

Abstract

We introduce the SE(3)-Transformer, a variant of the self-attention module for 3D point clouds and graphs, which is *equivariant* under continuous 3D roto-translations. Equivariance is important to ensure stable and predictable performance in the presence of nuisance transformations of the data input. A positive corollary of equivariance is increased weight-tying within the model. The SE(3)-Transformer leverages the benefits of self-attention to operate on large point clouds and graphs with varying number of points, while guaranteeing SE(3)-equivariance for robustness. We evaluate our model on a toy N-body particle simulation dataset, showcasing the robustness of the predictions under rotations of the input. We further achieve competitive performance on two real-world datasets, ScanObjectNN and QM9. In all cases, our model outperforms a strong, non-equivariant attention baseline and an equivariant model without attention.

1 Introduction

Self-attention mechanisms [31] have enjoyed a sharp rise in popularity in recent years. Their relative implementational simplicity coupled with high efficacy on a wide range of tasks such as language modeling [31], image recognition [18], or graph-based problems [32], make them an attractive component to use. However, their generality of application means that for specific tasks, knowledge of existing underlying structure is unused. In this paper, we propose the *SE(3)-Transformer* shown in Fig. 1, a self-attention mechanism specifically for 3D point cloud and graph data, which adheres to *equivariance constraints*, improving robustness to nuisance transformations and general performance.

Point cloud data is ubiquitous across many fields, presenting itself in diverse forms such as 3D object scans [29], 3D molecular structures [21], or N-body particle simulations [14]. Finding neural structures which can adapt to the varying number of points in an input, while respecting the irregular sampling of point positions, is challenging. Furthermore, an important property is that these structures should be invariant to global changes in overall input pose; that is, 3D translations and rotations of the input point cloud should not affect the output. In this paper, we find that the explicit imposition of equivariance constraints on the self-attention mechanism addresses these challenges. The SE(3)-Transformer uses the self-attention mechanism as a data-dependent filter particularly suited for sparse, non-voxelised point cloud data, while respecting and leveraging the symmetries of the task at hand.

[*]equal contribution
[†]work done while at the Bosch Center for Artificial Intelligence

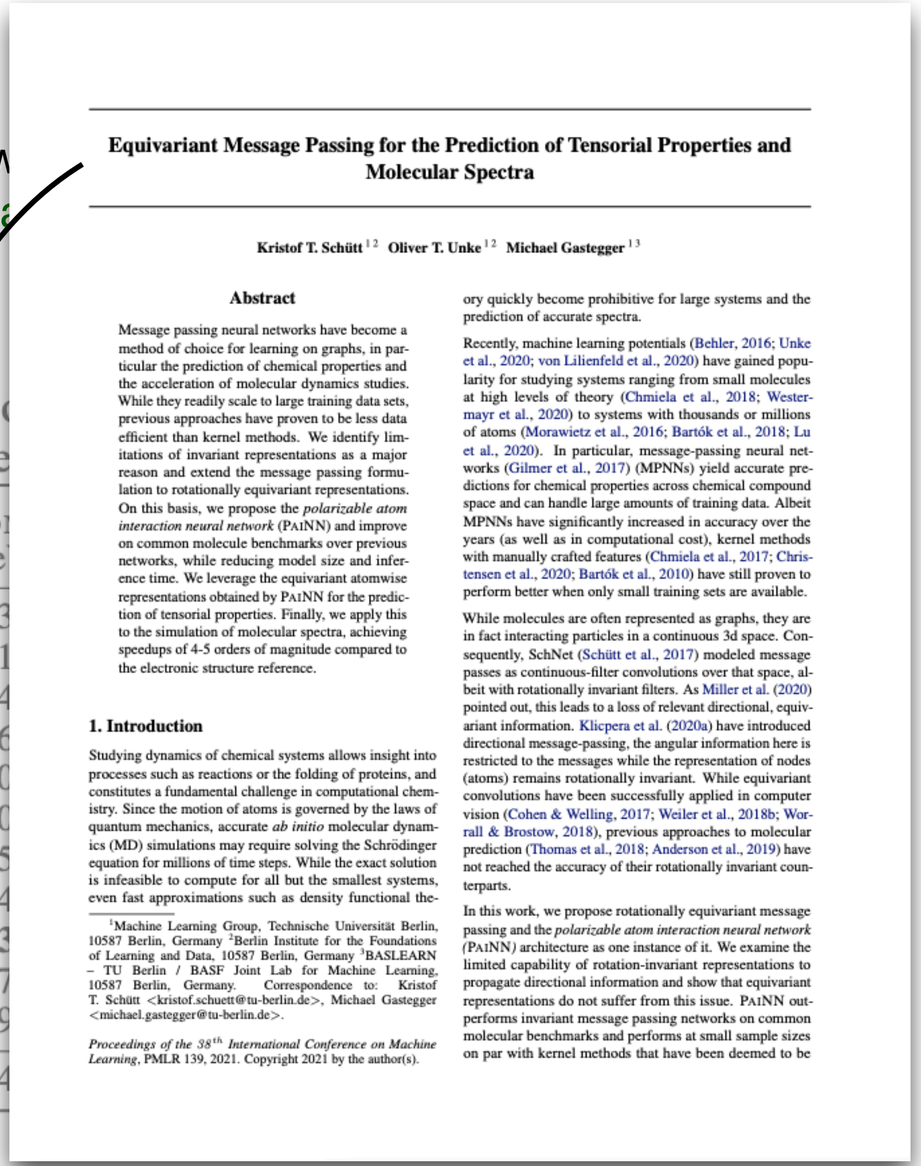
34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** betw
feature maps on homogeneous spa
is a **group convolution**"

Table 2: Performance comparis
Error (MAE) between model pre

			Task	α	$\Delta\varepsilon$	ε_{HO}
		no geometry	Units	bohr^3	meV	me
non-linear			NMP	.092	69	43
	regular	$\mathbb{R}^3$	SchNet *	.235	63	41
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	34
	steerable	$SE(3)$	L1Net	.088	68	46
	regular	G	LieConv	.084	49	30
	steerable	$SE(3)$	TFN	.223	58	40
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	35
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	24
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	23
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	27
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	29
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	24



Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra

Kristof T. Schütt [1 2] Oliver T. Unke [1 2] Michael Gastegger [1 3]

Abstract

Message passing neural networks have become a method of choice for learning on graphs, in particular the prediction of chemical properties and the acceleration of molecular dynamics studies. While they readily scale to large training data sets, previous approaches have proven to be less data efficient than kernel methods. We identify limitations of invariant representations as a major reason and extend the message passing formulation to rotationally equivariant representations. On this basis, we propose the *polarizable atom interaction neural network* (PaiNN) and improve on common molecule benchmarks over previous networks, while reducing model size and inference time. We leverage the equivariant atomwise representations obtained by PaiNN for the prediction of tensorial properties. Finally, we apply this to the simulation of molecular spectra, achieving speedups of 4-5 orders of magnitude compared to the electronic structure reference.

1. Introduction

Studying dynamics of chemical systems allows insight into processes such as reactions or the folding of proteins, and constitutes a fundamental challenge in computational chemistry. Since the motion of atoms is governed by the laws of quantum mechanics, accurate *ab initio* molecular dynamics (MD) simulations may require solving the Schrödinger equation for millions of time steps. While the exact solution is infeasible to compute for all but the smallest systems, even fast approximations such as density functional theory quickly become prohibitive for large systems and the prediction of accurate spectra.

Recently, machine learning potentials (Behler, 2016; Unke et al., 2020; von Lilienfeld et al., 2020) have gained popularity for studying systems ranging from small molecules at high levels of theory (Chmiela et al., 2018; Westermayr et al., 2020) to systems with thousands or millions of atoms (Morawietz et al., 2016; Bartók et al., 2018; Lu et al., 2020). In particular, message-passing neural networks (Gilmer et al., 2017) (MPNNs) yield accurate predictions for chemical properties across chemical compound space and can handle large amounts of training data. Albeit MPNNs have significantly increased in accuracy over the years (as well as in computational cost), kernel methods with manually crafted features (Chmiela et al., 2017; Christensen et al., 2020; Bartók et al., 2010) have still proven to perform better when only small training sets are available.

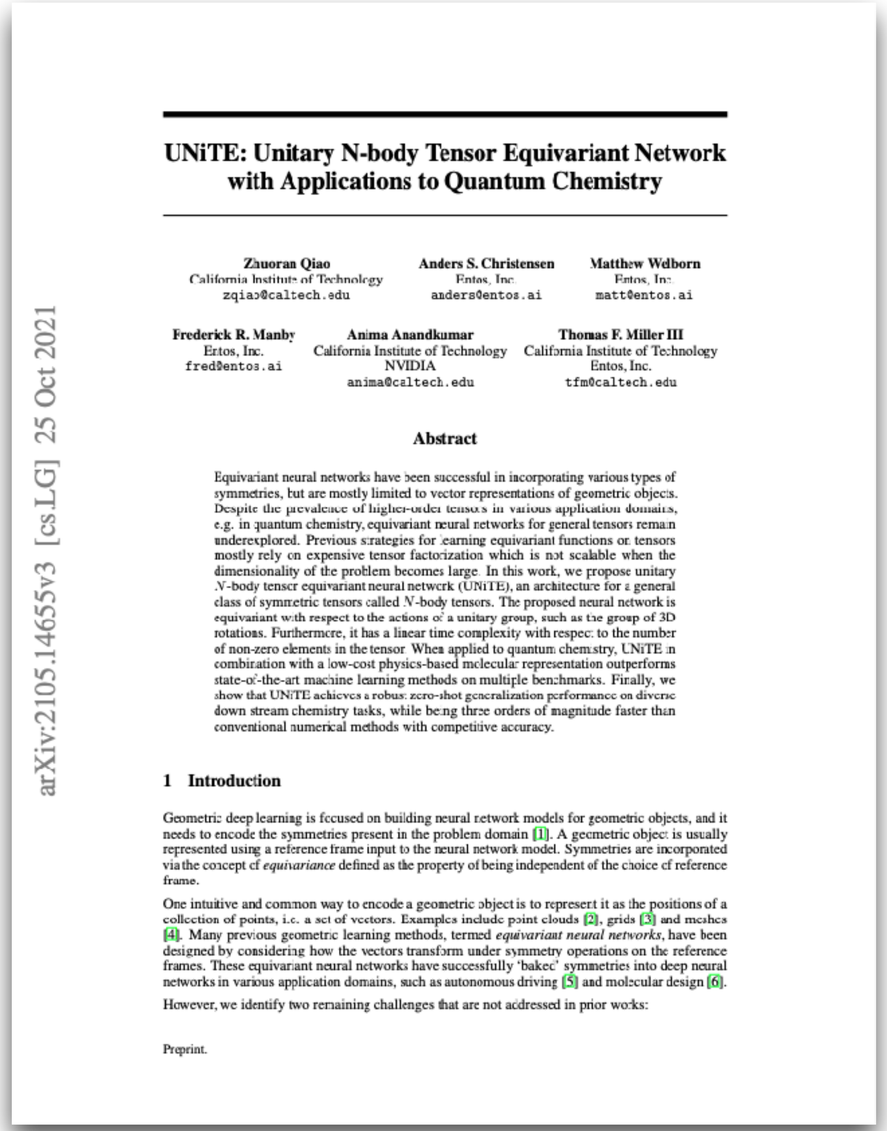
While molecules are often represented as graphs, they are in fact interacting particles in a continuous 3d space. Consequently, SchNet (Schütt et al., 2017) modeled message passes as continuous-filter convolutions over that space, albeit with rotationally invariant filters. As Miller et al. (2020) pointed out, this leads to a loss of relevant directional, equivariant information. Klicpera et al. (2020a) have introduced directional message-passing, the angular information here is restricted to the messages while the representation of nodes (atoms) remains rotationally invariant. While equivariant convolutions have been successfully applied in computer vision (Cohen & Welling, 2017; Weiler et al., 2018b; Worrall & Brostow, 2018), previous approaches to molecular prediction (Thomas et al., 2018; Anderson et al., 2019) have not reached the accuracy of their rotationally invariant counterparts.

In this work, we propose rotationally equivariant message passing and the *polarizable atom interaction neural network* (PaiNN) architecture as one instance of it. We examine the limited capability of rotation-invariant representations to propagate directional information and show that equivariant representations do not suffer from this issue. PaiNN outperforms invariant message passing networks on common molecular benchmarks and performs at small sample sizes on par with kernel methods that have been deemed to be

[1]Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany [2]Berlin Institute for the Foundations of Learning and Data, 10587 Berlin, Germany [3]BASLEARN – TU Berlin / BASF Joint Lab for Machine Learning, 10587 Berlin, Germany. Correspondence to: Kristof T. Schütt <kristof.schuett@tu-berlin.de>, Michael Gastegger <michael.gastegger@tu-berlin.de>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

Linear vs Non-linear & Regular vs Steerable



UNiTE: Unitary N-body Tensor Equivariant Network with Applications to Quantum Chemistry

Zhuoran Qiao
California Institute of Technology
zqiao@caltech.edu

Anders S. Christensen
Entos, Inc.
anders@entos.ai

Matthew Welborn
Entos, Inc.
matt@entos.ai

Frederick R. Manby
Entos, Inc.
fred@entos.ai

Anima Anandkumar
California Institute of Technology
NVIDIA
anima@caltech.edu

Thomas F. Miller III
California Institute of Technology
Entos, Inc.
tfm@caltech.edu

Abstract

Equivariant neural networks have been successful in incorporating various types of symmetries, but are mostly limited to vector representations of geometric objects. Despite the prevalence of higher-order tensors in various application domains, e.g. in quantum chemistry, equivariant neural networks for general tensors remain underexplored. Previous strategies for learning equivariant functions on tensors mostly rely on expensive tensor factorization which is not scalable when the dimensionality of the problem becomes large. In this work, we propose unitary N-body tensor equivariant neural network (UNiTE), an architecture for a general class of symmetric tensors called N-body tensors. The proposed neural network is equivariant with respect to the actions of a unitary group, such as the group of 3D rotations. Furthermore, it has a linear time complexity with respect to the number of non-zero elements in the tensor. When applied to quantum chemistry, UNiTE in combination with a low-cost physics-based molecular representation outperforms state-of-the-art machine learning methods on multiple benchmarks. Finally, we show that UNiTE achieves a robust zero-shot generalization performance on diverse down stream chemistry tasks, while being three orders of magnitude faster than conventional numerical methods with competitive accuracy.

1 Introduction

Geometric deep learning is focused on building neural network models for geometric objects, and it needs to encode the symmetries present in the problem domain [1]. A geometric object is usually represented using a reference frame input to the neural network model. Symmetries are incorporated via the concept of *equivariance* defined as the property of being independent of the choice of reference frame.

One intuitive and common way to encode a geometric object is to represent it as the positions of a collection of points, i.e. a set of vectors. Examples include point clouds [2], grids [3] and meshes [4]. Many previous geometric learning methods, termed *equivariant neural networks*, have been designed by considering how the vectors transform under symmetry operations on the reference frames. These equivariant neural networks have successfully 'baked' symmetries into deep neural networks in various application domains, such as autonomous driving [5] and molecular design [6].

However, we identify two remaining challenges that are not addressed in prior works:

Preprint.

Linear vs Non-linear & Regular vs Steerable

UNiTE: Unitary N-body Tensor Equivariant Network with Applications to Quantum Chemistry

Zhuoran Qiao
California Institute of Technology
zqiao@caltech.edu

Anders S. Christensen
Entos, Inc.
anders@entos.ai

Matthew Welborn
Entos, Inc.
matt@entos.ai

Frederick R. Manby
Entos, Inc.
fred@entos.ai

Anima Anandkumar
California Institute of Technology
NVIDIA
anima@caltech.edu

Thomas F. Miller III
California Institute of Technology
Entos, Inc.
tfm@caltech.edu

Abstract

Equivariant neural networks have been successful in incorporating various types of symmetries, but are mostly limited to vector representations of geometric objects. Despite the prevalence of higher-order tensors in various application domains, e.g. in quantum chemistry, equivariant neural networks for general tensors remain underexplored. Previous strategies for learning equivariant functions on tensors mostly rely on expensive tensor factorization which is not scalable when the dimensionality of the problem becomes large. In this work, we propose unitary N-body tensor equivariant neural network (UNiTE), an architecture for a general class of symmetric tensors called N-body tensors. The proposed neural network is equivariant with respect to the actions of a unitary group, such as the group of 3D rotations. Furthermore, it has a linear time complexity with respect to the number of non-zero elements in the tensor. When applied to quantum chemistry, UNiTE in combination with a low-cost physics-based molecular representation outperforms state-of-the-art machine learning methods on multiple benchmarks. Finally, we show that UNiTE achieves a robust zero-shot generalization performance on diverse down stream chemistry tasks, while being three orders of magnitude faster than conventional numerical methods with competitive accuracy.

1 Introduction

Geometric deep learning is focused on building neural network models for geometric objects, and it needs to encode the symmetries present in the problem domain [1]. A geometric object is usually represented using a reference frame input to the neural network model. Symmetries are incorporated via the concept of *equivariance* defined as the property of being independent of the choice of reference frame.

One intuitive and common way to encode a geometric object is to represent it as the positions of a collection of points, i.e. a set of vectors. Examples include point clouds [2], grids [3] and meshes [4]. Many previous geometric learning methods, termed *equivariant neural networks*, have been designed by considering how the vectors transform under symmetry operations on the reference frames. These equivariant neural networks have successfully 'baked' symmetries into deep neural networks in various application domains, such as autonomous driving [5] and molecular design [6].

However, we identify two remaining challenges that are not addressed in prior works:

Preprint.

E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials

Simon Batzner*,[1] Albert Musaelian,[1] Lixin Sun,[1] Mario Geiger,[2] Jonathan P. Mailoa,[3] Mordechai Kornbluth,[3] Nicola Molinari,[1] Tess E. Smidt,[4,5] and Boris Kozinsky*[1,3]

[1]*John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA*
[2]*École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
[3]*Robert Bosch Research and Technology Center, Cambridge, MA 02139, USA*
[4]*Computational Research Division and Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[5]*Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA 02142, USA*

This work presents Neural Equivariant Interatomic Potentials (NequIP), an E(3)-equivariant neural network approach for learning interatomic potentials from *ab-initio* calculations for molecular dynamics simulations. While most contemporary symmetry-aware models use invariant convolutions and only act on scalars, NequIP employs E(3)-equivariant convolutions for interactions of geometric tensors, resulting in a more information-rich and faithful representation of atomic environments. The method achieves state-of-the-art accuracy on a challenging and diverse set of molecules and materials while exhibiting remarkable data efficiency. NequIP outperforms existing models with up to three orders of magnitude fewer training data, challenging the widely held belief that deep neural networks require massive training sets. The high data efficiency of the method allows for the construction of accurate potentials using high-order quantum chemical level of theory as reference and enables high-fidelity molecular dynamics simulations over long time scales.

INTRODUCTION

Molecular dynamics (MD) simulations are an indispensable tool for computational discovery in fields as diverse as energy storage, catalysis, and biological processes [1–3]. While the atomic forces required to integrate Newton's equations of motion can in principle be obtained with high fidelity from quantum-mechanical calculations such as density functional theory (DFT), in practice the unfavorable computational scaling of first-principles methods limits simulations to short time scales and small numbers of atoms. This prohibits the study of many interesting physical phenomena beyond the time and length scales that are currently accessible, even on the largest supercomputers. Owing to their simple functional form, classical models for the atomic potential energy can typically be evaluated orders of magnitude faster than first-principles methods, thereby enabling the study of large numbers of atoms over long time scales. However, due to their limited mathematical form, classical interatomic potentials, or force fields, are inherently limited in their predictive accuracy which has historically led to a fundamental trade-off between obtaining high computational efficiency while also predicting faithful dynamics of the system under study.

The construction of flexible models of the interatomic potential energy based on Machine Learning (ML-IP), and in particular Neural Networks (NN-IP), has shown great promise in providing a way to move past this dilemma, promising to learn high-fidelity potentials from *ab-initio* reference calculations while retaining favorable computational efficiency [4–13]. One of the limiting factors of NN-IPs is that they typically require collection of large training sets of *ab-initio* calculations, often including thousands or even millions of reference structures [4, 9, 10, 14–16]. This computationally expensive process of training data collection has severely limited the adoption of NN-IPs as it quickly becomes a bottleneck in the development of force-fields for new systems.

In this work, we present the Neural Equivariant Interatomic Potential (NequIP), a highly data-efficient deep learning approach for learning interatomic potentials from reference first-principles calculations. We show that the proposed method obtains high accuracy compared to existing ML-IP methods across a wide variety of systems, including small molecules, water in different phases, an amorphous solid, a reaction at a solid/gas interface, and a Lithium superionic conductor. Furthermore, we find that NequIP exhibits exceptional data efficiency, enabling the construction of accurate interatomic potentials from limited data sets of fewer than 1,000 or even as little as 100 reference *ab-initio* calculations, where other methods require

* **Corresponding authors**
B.K., E-mail: bkoz@seas.harvard
S.B., E-mail: batzner@g.harvard.edu

Linear vs Non-linear & Regular vs Steerable

UNiTE: Unitary N-body Tensor Equivariant Network with Applications to Quantum Chemistry

Zhuoran Qiao
California Institute of Technology
zqiao@caltech.edu

Anders S. Christensen
Entos, Inc.
anders@entos.ai

Matthew Welborn
Entos, Inc.
matt@entos.ai

Frederick R. Manby
Entos, Inc.
fred@entos.ai

Anima Anandkumar
California Institute of Technology
NVIDIA
anima@caltech.edu

Thomas F. Miller III
California Institute of Technology
Entos, Inc.
tfm@caltech.edu

Abstract

Equivariant neural networks have been successful in incorporating various types of symmetries, but are mostly limited to vector representations of geometric objects. Despite the prevalence of higher-order tensors in various application domains, e.g. in quantum chemistry, equivariant neural networks for general tensors remain underexplored. Previous strategies for learning equivariant functions on tensors mostly rely on expensive tensor factorization which is not scalable when the dimensionality of the problem becomes large. In this work, we propose unitary N-body tensor equivariant neural network (UNiTE), an architecture for a general class of symmetric tensors called N-body tensors. The proposed neural network is equivariant with respect to the actions of a unitary group, such as the group of 3D rotations. Furthermore, it has a linear time complexity with respect to the number of non-zero elements in the tensor. When applied to quantum chemistry, UNiTE in combination with a low-cost physics-based molecular representation outperforms state-of-the-art machine learning methods on multiple benchmarks. Finally, we show that UNiTE achieves a robust zero-shot generalization performance on diverse down stream chemistry tasks, while being three orders of magnitude faster than conventional numerical methods with competitive accuracy.

1 Introduction

Geometric deep learning is focused on building neural network models for geometric objects, and it needs to encode the symmetries present in the problem domain [1]. A geometric object is usually represented using a reference frame input to the neural network model. Symmetries are incorporated via the concept of *equivariance* defined as the property of being independent of the choice of reference frame.

One intuitive and common way to encode a geometric object is to represent it as the positions of a collection of points, i.e. a set of vectors. Examples include point clouds [2], grids [3] and meshes [4]. Many previous geometric learning methods, termed *equivariant neural networks*, have been designed by considering how the vectors transform under symmetry operations on the reference frames. These equivariant neural networks have successfully 'baked' symmetries into deep neural networks in various application domains, such as autonomous driving [5] and molecular design [6].

However, we identify two remaining challenges that are not addressed in prior works:

Preprint.

E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials

Simon Batzner*,[1] Albert Musaelian,[1] Lixin Sun,[1] Mario Geiger,[2] Jonathan P. Mailoa,[3]
Mordechai Kornbluth,[3] Nicola
[1]*John A. Paulson*
Harvard Un
[2]*École Polytechnique F*
[3]*Robert Bosch Research a*
[4]*Computational Research Division and C*
Lawrence Berkeley N
[5]*Massachusetts Inst*
Engineering and Co

This work presents Neural Equi
neural network approach for learning
dynamics simulations. While most c
and only act on scalars, NequIP emp
tensors, resulting in a more informa
The method achieves state-of-the-a
materials while exhibiting remarkal
up to three orders of magnitude fe
neural networks require massive trai
construction of accurate potentials
and enables high-fidelity molecular

INTRODUCTION

Molecular dynamics (MD) simulations
indispensable tool for computational discover
as diverse as energy storage, catalysis, and
processes [1–3]. While the atomic forces r
integrate Newton's equations of motion can i
be obtained with high fidelity from quantum-m
calculations such as density functional theo
in practice the unfavorable computational
first-principles methods limits simulations to
scales and small numbers of atoms. This pr
study of many interesting physical phenome
the time and length scales that are currently
even on the largest supercomputers. Owin
simple functional form, classical models for t
potential energy can typically be evaluated
magnitude faster than first-principles method
enabling the study of large numbers of atoms
time scales. However, due to their limited ma
form, classical interatomic potentials, or force
inherently limited in their predictive accur
has historically led to a fundamental trade-o
obtaining high computational efficiency
predicting faithful dynamics of the system un

* Corresponding authors
B.K., E-mail: bkoz@seas.harvard
S.B., E-mail: batzner@g.harvard.edu

Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics

Albert Musaelian*,[1] Simon Batzner*,†,[1]
Anders Johansson,[1] Lixin Sun,[1] Cameron J. Owen,[1] Mordechai Kornbluth,[2] and Boris Kozinsky†,[1,2]
[1]*Harvard University*
[2]*Robert Bosch LLC Research and Technology Center*

A simultaneously accurate and computationally efficient parametrization of the energy and atomic forces of molecules and materials is a long-standing goal in the natural sciences. In pursuit of this goal, neural message passing has lead to a paradigm shift by describing many-body correlations of atoms through iteratively passing messages along an atomistic graph. This propagation of information, however, makes parallel computation difficult and limits the length scales that can be studied. Strictly local descriptor-based methods, on the other hand, can scale to large systems but do not currently match the high accuracy observed with message passing approaches. This work introduces Allegro, a strictly local equivariant deep learning interatomic potential that simultaneously exhibits excellent accuracy and scalability of parallel computation. Allegro learns many-body functions of atomic coordinates using a series of tensor products of learned equivariant representations, but without relying on message passing. Allegro obtains improvements over state-of-the-art methods on the QM9 and revised MD-17 data sets. A single tensor product layer is shown to outperform existing deep message passing neural networks and transformers on the QM9 benchmark. Furthermore, Allegro displays remarkable generalization to out of distribution data. Molecular dynamics simulators based on Allegro recover structural and kinetic properties of an amorphous phosphate electrolyte in excellent agreement with first principles calculations. Finally, we demonstrate the parallel scaling of Allegro with a dynamics simulation of 100 million atoms.

INTRODUCTION

Molecular dynamics (MD) and Monte-Carlo (MC) simulation methods for the study of properties of molecules and materials are a core pillar of computational chemistry, materials science, and biology. Common to a diverse set of applications ranging from energy materials [1] to protein folding [2] is the requirement that predictions of the potential energy and the atomic forces must be both accurate and computationally efficient to faithfully describe the evolution of complex systems over long time scales. While first-principles methods such as density functional theory (DFT), which explicitly treat the electrons of the system, provide an accurate and transferable description of the system, they exhibit poor scaling with system size and thus limit practical applications to small systems and short simulation times. Classical force-fields based on simple functions of atomic coordinates are able to scale to large systems and long time scales but are inherently limited in their fidelity and can yield unfaithful dynamics. Descriptions of the potential energy surface (PES) using machine learning (ML) have emerged as a promising approach to move past this trade-off [3–24]. Machine learning interatomic potentials (MLIPs) aim to approximate a set of high-fidelity energy and force labels at improved computational efficiency that scales linearly with the number of atoms. A variety of different approaches have been proposed, from shallow neural networks and kernel-based approaches [3–6] to more recent methods based on deep learning [14, 15, 20, 25, 26]. In particular, a class of MLIPs based on message passing neural networks (MPNNs) has shown remarkable accuracy [9, 11, 14, 15, 26, 27]. In interatomic potentials based on MPNNs, an atomistic graph is induced by connecting with edges each atom (node) to all neighboring atoms inside a finite cutoff sphere surrounding the central atom. Information is then iteratively propagated along this graph, allowing MPNNs to learn many-body correlations and access non local information outside of the local cutoff. This iterated propagation, however, leads to large receptive fields with many effective neighbors for each atom, which slows down parallel computation and limits the length scales accessible to message passing MLIPs. MLIPs using *strictly local* descriptors such as Behler Parrinello neural networks [5], GAP [6], SNAP [7], DeepMD [20], Moment Tensor Potentials [8], or ACE [12] do not suffer from this obstacle due to their strict locality. As a result they can easily be parallelized across devices and have successfully been scaled to extremely large system sizes [28–31]. Approaches based on local atom-density based descriptors, however, have so far fallen behind in accuracy compared to state-of-the-art equivariant message passing interatomic potentials [15].

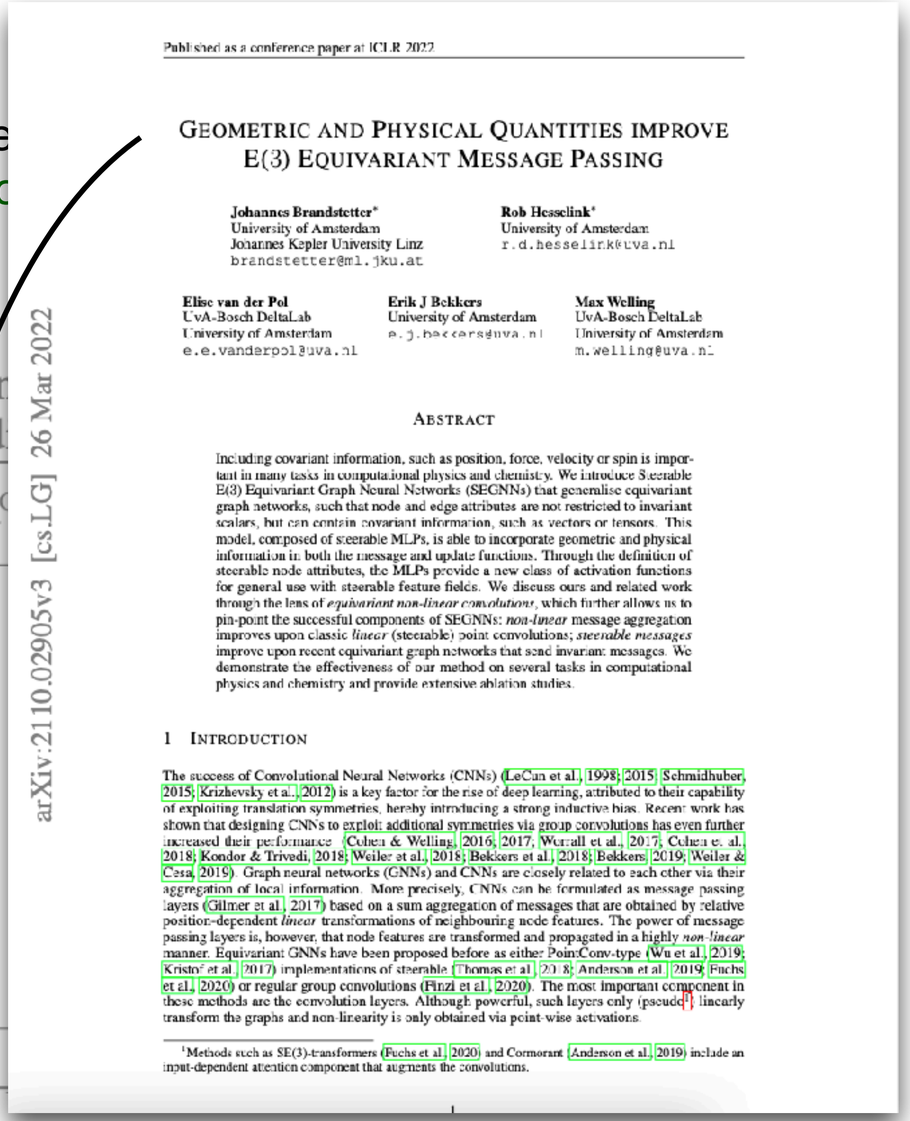
* Equal Contribution. Order is random.
† Corresponding author
S.B., E-mail: batzner@g.harvard.edu
B.K., E-mail: bkoz@seas.harvard.edu

Linear vs Non-linear & Regular vs Steerable

Recall lecture 1.7:
"Any **equivariant linear layer** betwe
feature maps on homogeneous spac
is a **group convolution**"

Table 2: Performance comparison
Error (MAE) between model pred

			Task	α	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$
			Units	bohr^3	meV	meV
non-linear		no geometry	NMP	.092	69	43
	regular	$\mathbb{R}^3$	SchNet *	.235	63	41
pseudo-linear	steerable	$SE(3)$	Cormorant	.085	61	34
	steerable	$SE(3)$	L1Net	.088	68	46
	regular	G	LieConv	.084	49	30
	steerable	$SE(3)$	TFN	.223	58	40
pseudo-linear	steerable	$SE(3)$	SE(3)-Tr.	.142	53	35
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	DimeNet++ *	**.043**	**32**	24
non-linear	regular	$\mathbb{R}^3 \times S^2 \times \mathbb{R}^+$	SphereNet *	.046	**32**	**23**
non-linear	reguleerable?	$SE(3)$	PaiNN *	.045	45	27
non-linear	regular	$\mathbb{R}^3$	EGNN	.071	48	29
non-linear	steerable	$SE(3)$	SEGNN (Ours)	.060	42	24



Published as a conference paper at ICLR 2022

GEOMETRIC AND PHYSICAL QUANTITIES IMPROVE E(3) EQUIVARIANT MESSAGE PASSING

Johannes Brandstetter*
University of Amsterdam
Johannes Kepler University Linz
brandstetter@ml.jku.at

Rob Hesselink*
University of Amsterdam
r.d.hesselink@uva.nl

Elise van der Pol
UvA-Bosch DeltaLab
University of Amsterdam
e.e.vanderpol@uva.nl

Erik J Bekkers
University of Amsterdam
e.j.bekkers@uva.nl

Max Welling
UvA-Bosch DeltaLab
University of Amsterdam
m.welling@uva.nl

ABSTRACT

Including covariant information, such as position, force, velocity or spin is important in many tasks in computational physics and chemistry. We introduce Steerable E(3) Equivariant Graph Neural Networks (SEGNNs) that generalise equivariant graph networks, such that node and edge attributes are not restricted to invariant scalars, but can contain covariant information, such as vectors or tensors. This model, composed of steerable MLPs, is able to incorporate geometric and physical information in both the message and update functions. Through the definition of steerable node attributes, the MLPs provide a new class of activation functions for general use with steerable feature fields. We discuss ours and related work through the lens of *equivariant non-linear convolutions*, which further allows us to pin-point the successful components of SEGNNs: *non-linear* message aggregation improves upon classic *linear* (steerable) point convolutions; *steerable messages* improve upon recent equivariant graph networks that send invariant messages. We demonstrate the effectiveness of our method on several tasks in computational physics and chemistry and provide extensive ablation studies.

1 INTRODUCTION

The success of Convolutional Neural Networks (CNNs) (LeCun et al., 1998; 2015; Schmidhuber, 2015; Krizhevsky et al., 2012) is a key factor for the rise of deep learning, attributed to their capability of exploiting translation symmetries, thereby introducing a strong inductive bias. Recent work has shown that designing CNNs to exploit additional symmetries via group convolutions has even further increased their performance (Cohen & Welling, 2016; 2017; Worrall et al., 2017; Cohen et al., 2018; Kondor & Trivedi, 2018; Weiler et al., 2018; Bekkers et al., 2018; Bekkers, 2019; Weiler & Cesa, 2019). Graph neural networks (GNNs) and CNNs are closely related to each other via their aggregation of local information. More precisely, CNNs can be formulated as message passing layers (Gilmer et al., 2017) based on a sum aggregation of messages that are obtained by relative position-dependent *linear* transformations of neighbouring node features. The power of message passing layers is, however, that node features are transformed and propagated in a highly *non-linear* manner. Equivariant GNNs have been proposed before as either PointConv-type (Wu et al., 2019; Kipf et al., 2017) implementations of steerable (Thomas et al., 2018; Anderson et al., 2019; Fuchs et al., 2020) or regular group convolutions (Finzi et al., 2020). The most important component in these methods are the convolution layers. Although powerful, such layers only (pseudo[1]) linearly transform the graphs and non-linearity is only obtained via point-wise activations.

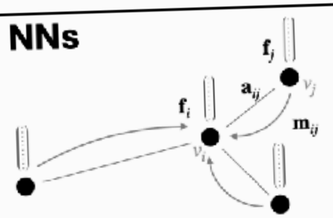
[1]Methods such as SE(3)-transformers (Fuchs et al., 2020) and Cormorant (Anderson et al., 2019) include an input-dependent attention component that augments the convolutions.

Analysis of literature in terms of non-linear group convolutions

Lecture 2.2

Linear vs non-linear (group) convolutions

Message passing NNs



Compute messages: $\mathbf{m}_{ij} = \phi_m(\mathbf{f}_i, \mathbf{f}_j, \mathbf{a}_{ij})$

Aggregate and update: $\mathbf{f}'_i = \phi_f\left(\mathbf{f}_i, \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}\right)$

Classic point convolutions

(Lecture 1.7: regular g-convs on homogeneous spaces)

$$\mathbf{m}_{ij} = \mathbf{W}(\|\mathbf{x}_j - \mathbf{x}_i\|)\mathbf{f}_j$$

$$\mathbf{W}(g_i^{-1}g_j)\mathbf{f}_j$$

Linear convolution

Steerable G-CNNs

(Lecture 2: steerable g-convs)

$\mathcal{F}_H$

$$\mathbf{m}_{ij} = \mathbf{W}_{\hat{\mathbf{a}}_{ij}}(\|\mathbf{x}_j - \mathbf{x}_i\|)\hat{\mathbf{f}}_j$$

$$:= \hat{\mathbf{f}}_j \otimes_{cg}^{\mathbf{W}(\|\mathbf{x}_j - \mathbf{x}_i\|)} \hat{\mathbf{a}}_{ij}$$

Invariant Message **Passing NNs**

(Lecture 3)

$$\mathbf{m}_{ij} = \text{MLP}(\mathbf{f}_i, \mathbf{f}_j, \|\mathbf{x}_j - \mathbf{x}_i\|)$$

Non-linear "convolution"

Equivariant (Steerable) Message **Passing**

(Lecture 3)

$$\hat{\mathbf{m}}_{ij} = \widehat{\text{MLP}}\,(\hat{\mathbf{f}}_i, \hat{\mathbf{f}}_j, \mathbf{x}_j - \mathbf{x}_i)$$

With steerable MLP:

$$\widehat{\text{MLP}}_{\hat{\mathbf{a}}_{ij}}(\hat{\mathbf{f}}_i, \hat{\mathbf{f}}_j, \|\mathbf{x}_j - \mathbf{x}_i\|) := \sigma(\mathbf{W}^{(n)}_{\hat{\mathbf{a}}_{ij}}(\ldots(\sigma(\mathbf{W}^{(1)}_{\hat{\mathbf{a}}_{ij}}\hat{\mathbf{h}}_i))))$$

29

Published as a conference paper at ICLR 2022

GEOMETRIC AND PHYSICAL QUANTITIES IMPROVE E(3) EQUIVARIANT MESSAGE PASSING

Johannes Brandstetter*
University of Amsterdam
Johannes Kepler University Linz
brandstetter@ml.jku.at

Rob Hesselink*
University of Amsterdam
r.d.hesselink@uva.nl

Elise van der Pol
UvA-Bosch DeltaLab
University of Amsterdam
e.e.vanderpol@uva.nl

Erik J Bekkers
University of Amsterdam
e.j.bekkers@uva.nl

Max Welling
UvA-Bosch DeltaLab
University of Amsterdam
m.welling@uva.nl

ABSTRACT

Including covariant information, such as position, force, velocity or spin is important in many tasks in computational physics and chemistry. We introduce Steerable E(3) Equivariant Graph Neural Networks (SEGNNs) that generalise equivariant graph networks, such that node and edge attributes are not restricted to invariant scalars, but can contain covariant information, such as vectors or tensors. This model, composed of steerable MLPs, is able to incorporate geometric and physical information in both the message and update functions. Through the definition of steerable node attributes, the MLPs provide a new class of activation functions for general use with steerable feature fields. We discuss ours and related work through the lens of *equivariant non-linear convolutions*, which further allows us to pin-point the successful components of SEGNNs: *non-linear* message aggregation improves upon classic *linear* (steerable) point convolutions; *steerable messages* improve upon recent equivariant graph networks that send invariant messages. We demonstrate the effectiveness of our method on several tasks in computational physics and chemistry and provide extensive ablation studies.

1 INTRODUCTION

The success of Convolutional Neural Networks (CNNs) (LeCun et al., 1998; 2015; Schmidhuber, 2015; Krizhevsky et al., 2012) is a key factor for the rise of deep learning, attributed to their capability of exploiting translation symmetries, thereby introducing a strong inductive bias. Recent work has shown that designing CNNs to exploit additional symmetries via group convolutions has even further increased their performance (Cohen & Welling, 2016; 2017; Worrall et al., 2017; Cohen et. al., 2018; Kondor & Trivedi, 2018; Weiler et al., 2018; Bekkers et al., 2018; Bekkers, 2019; Weiler & Cesa, 2019). Graph neural networks (GNNs) and CNNs are closely related to each other via their aggregation of local information. More precisely, CNNs can be formulated as message passing layers (Gilmer et al., 2017) based on a sum aggregation of messages that are obtained by relative position-dependent *linear* transformations of neighbouring node features. The power of message passing layers is, however, that node features are transformed and propagated in a highly *non-linear* manner. Equivariant GNNs have been proposed before as either PointConv-type (Wu et al., 2019; Kristof et al., 2017) implementations of steerable (Thomas et al., 2018; Anderson et al., 2019; Fuchs et al., 2020) or regular group convolutions (Finzi et al., 2020). The most important component in these methods are the convolution layers. Although powerful, such layers only (pseudo[1]) linearly transform the graphs and non-linearity is only obtained via point-wise activations.

[1]Methods such as SE(3)-transformers (Fuchs et al., 2020) and Cormorant (Anderson et al., 2019) include an input-dependent attention component that augments the convolutions.