Group Equivariant Deep Learning

Lecture 3 - Equivariant graph neural networks

Lecture 3.1 - Motivation for SE(3) equivariant graph NNs

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/

Group Equivariant Deep Learning

Lecture 3 - Equivariant graph neural networks

Lecture 3.1 - Motivation for SE(3) equivariant graph NNs

Lecture 3.2 - Equivariant message passing as non-linear convolution

Lecture 3.3 - Tensor products as conditional linear layers (and MLPs)

A motivation for attributed conditioned message passing using bilinear layers

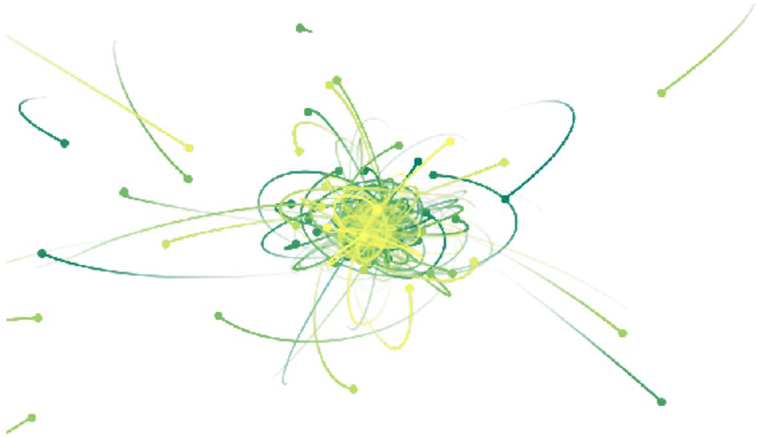
Lecture 3.4 - Group Theory | SO(3) irreps (Wigner-D matrices), Clebsch-Gordan TP

Preliminaries for 3D steerable g-convs

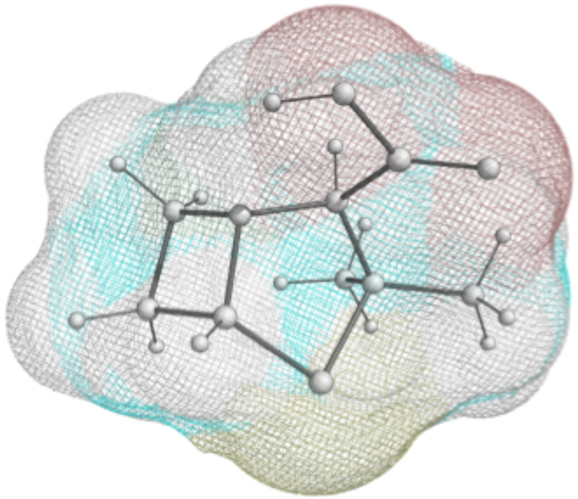
Lecture 3.5 - 3D Steerable (graph) convolutions

Lecture 3.6 - Regular (as opposed to steerable) equivariant graph NNs

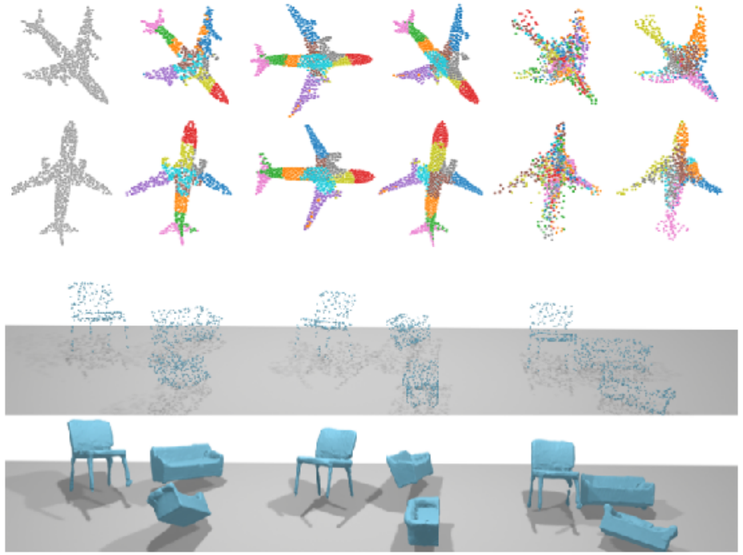
Lecture 3.7 - Gauge equivariant graph NNs



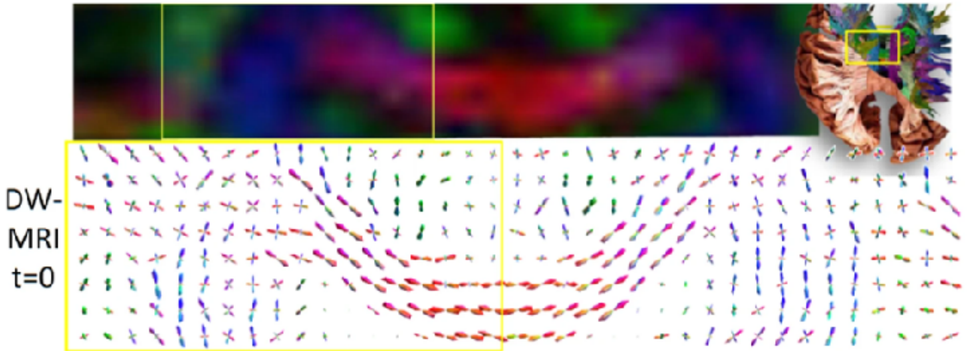
Computational Physics [1,*]



Computational Chemistry [2]



3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

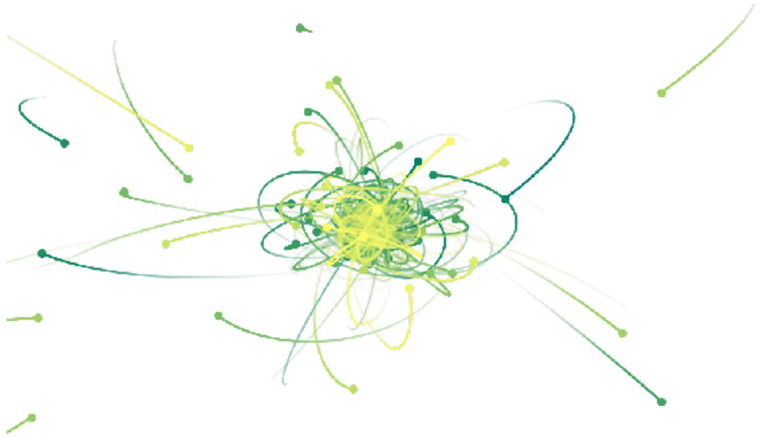
[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

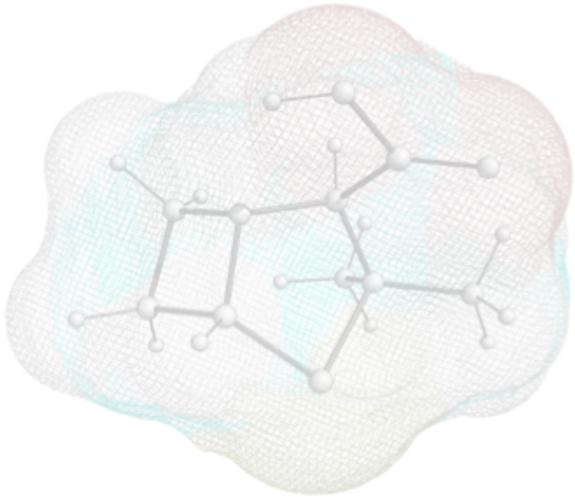
[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

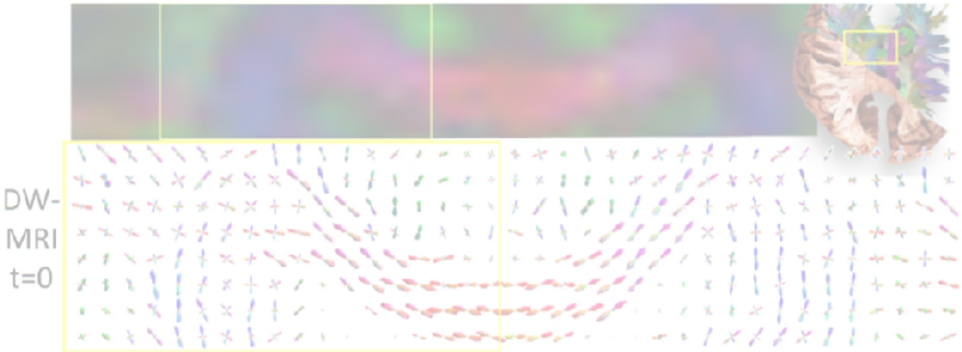


Computational Physics [1,*]



Computational Chemistry [2]

3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

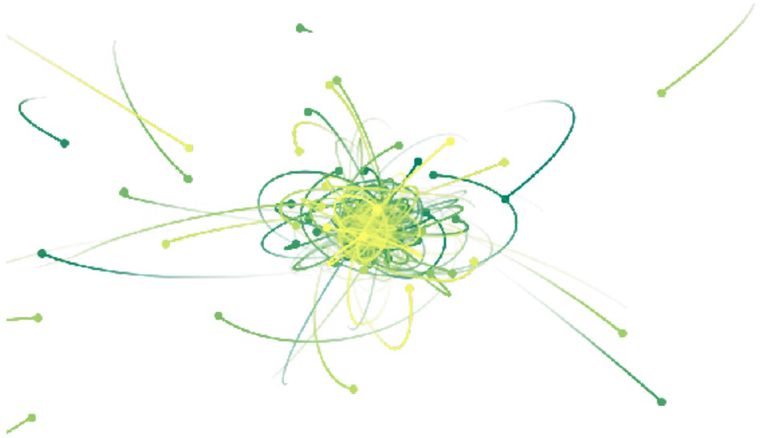
[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.



Computational Physics [1,*]

puter Vision [3,4,*]

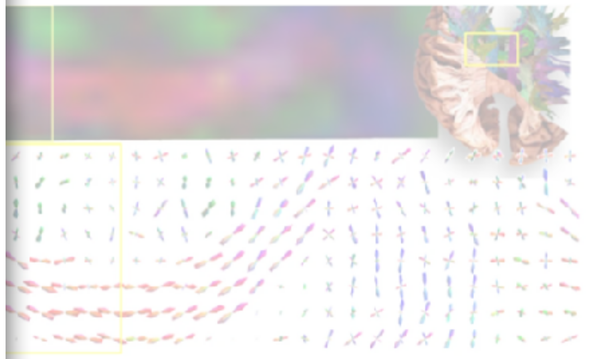


Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Ge Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecula**

[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (**canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-E in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2 **Image Processing via a Cartan Connection**. Handbook of Mathematical Models a Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais **Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

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

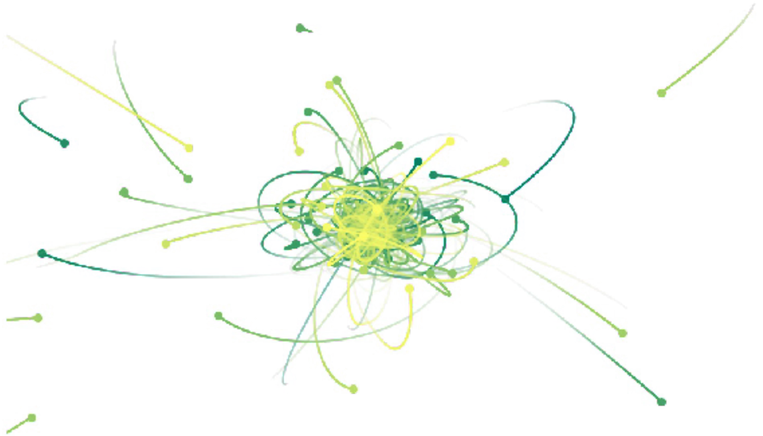
ABSTRACT

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, hereby 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.



Computational Physics [1,*]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Ge Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecula**

[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (**canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-E in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (**Image Processing via a Cartan Connection**. Handbook of Mathematical Models a Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais **Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

Published as a conference paper at ICLR 2022

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

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

Rob Hesselink*
University of Ams
r.d.hesselink

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
UvA-
Unive
m.we

ABSTRACT

Including covariant information, such as position, force, velocity or
tant in many tasks in computational physics and chemistry. We introd
E(3) Equivariant Graph Neural Networks (SEGNNs) that generali
graph networks, such that node and edge attributes are not restricte
scalars, but can contain covariant information, such as vectors or
model, composed of steerable MLPs, is able to incorporate geometri
information in both the message and update functions. Through th
steerable node attributes, the MLPs provide a new class of activat
for general use with steerable feature fields. We discuss ours and
through the lens of *equivariant non-linear convolutions*, which furth
pin-point the successful components of SEGNNs: *non-linear* messag
improves upon classic *linear* (steerable) point convolutions; *steera*
improve upon recent equivariant graph networks that send invariant
demonstrate the effectiveness of our method on several tasks in c
physics and chemistry and provide extensive ablation studies.

1 INTRODUCTION

The success of Convolutional Neural Networks (CNNs) (LeCun et al., 199
2015; Krizhevsky et al., 2012) is a key factor for the rise of deep learning, attri
of exploiting translation symmetries, hereby introducing a strong inductive
shown that designing CNNs to exploit additional symmetries via group convo
increased their performance (Cohen & Welling, 2016; 2017; Worrall et a
2018; Kondor & Trivedi, 2018; Weiler et al., 2018; Bekkers et al., 2018; Be
Cesa, 2019). Graph neural networks (GNNs) and CNNs are closely related
aggregation of local information. More precisely, CNNs can be formulat
layers (Gilmer et al., 2017) based on a sum aggregation of messages that a
position-dependent *linear* transformations of neighbouring node features.
passing layers is, however, that node features are transformed and propagated
manner. Equivariant GNNs have been proposed before as either PointConv
Kristof et al., 2017) implementations of steerable (Thomas et al., 2018; Ande
et al., 2020) or regular group convolutions (Finzi et al., 2020). The most in
these methods are the convolution layers. Although powerful, such layers
transform the graphs and non-linearity is only obtained via point-wise activa

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

Submitted to the Proceedings of the US Community Study
on the Future of Particle Physics (Snowmass 2021)

Symmetry Group Equivariant Architectures for Physics
A Snowmass 2022 White Paper

Alexander Bogatskiy[1], Sanmay Ganguly[2], Thomas Kipf[3], Risi Kondor[4], David W. Miller*[4], Daniel Murnane[5], Jan T. Offermann[4], Mariel Pettee†[5], Phiala Shanahan[6], Chase Shimmin[7], and Savannah Thais[8]

[1]*Flatiron Institute*
[2]*ICEPP, University of Tokyo*
[3]*Google Research*
[4]*University of Chicago*
[5]*Lawrence Berkeley National Laboratory*
[6]*Massachusetts Institute of Technology*
[7]*Yale University*
[8]*Princeton University*

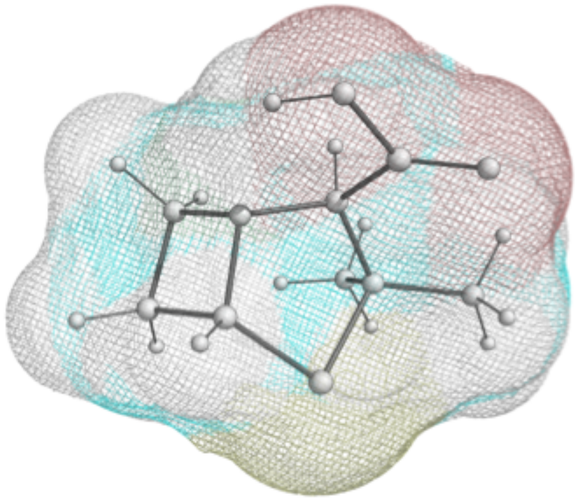
March 14, 2022

Abstract

Physical theories grounded in mathematical symmetries are an essential component of our understanding of a wide range of properties of the universe. Similarly, in the domain of machine learning, an awareness of symmetries such as rotation or permutation invariance has driven impressive performance breakthroughs in computer vision, natural language processing, and other important applications. In this report, we argue that both the physics community and the broader machine learning community have much to understand and potentially to gain from a deeper investment in research concerning symmetry group equivariant machine learning architectures. For some applications, the introduction of symmetries into the fundamental structural design can yield models that are more economical (i.e. contain fewer, but more expressive, learned parameters), interpretable (i.e. more explainable or directly mappable to physical quantities), and/or trainable (i.e. more efficient in both data and computational requirements). We discuss various figures of merit for evaluating these models as well as some potential benefits and limitations of these methods for a variety of physics applications. Research and investment into these approaches will lay the foundation for future architectures that are potentially more robust under new computational paradigms and will provide a richer description of the physical systems to which they are applied.

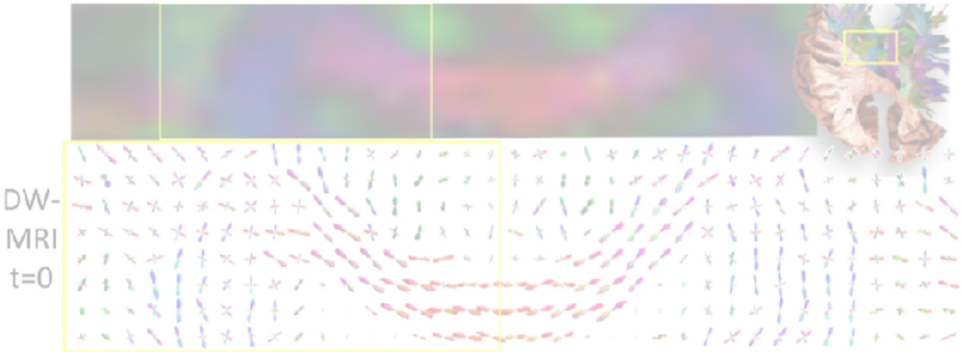
*Contact Editor, David.W.Miller@uchicago.edu
†Contact Editor, mpettee@lbl.gov

Computational Physics [1,*]



Computational Chemistry [2]

3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

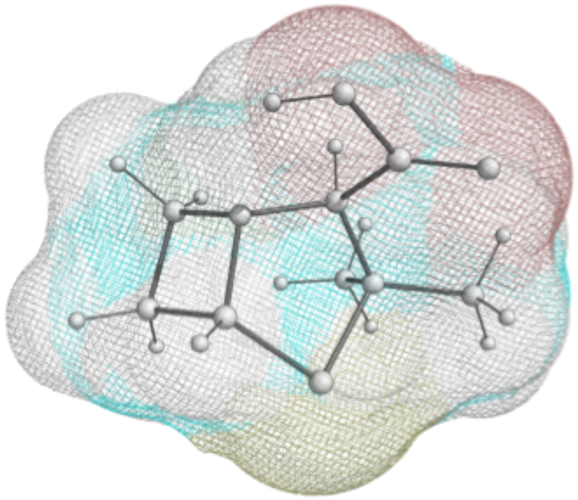
[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

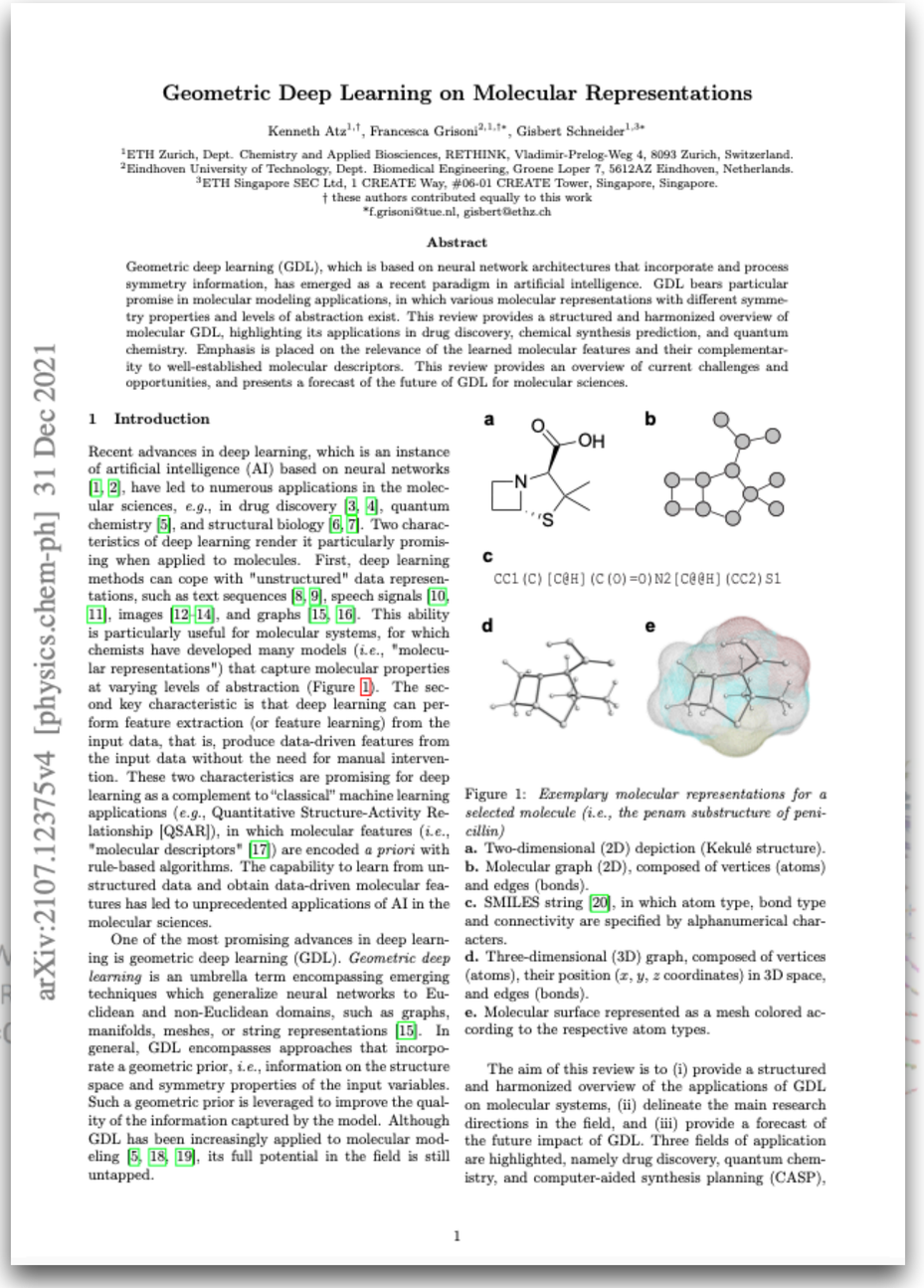
Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.



Computational Physics [1,*]

Computational Chemistry [2]



Geometric Deep Learning on Molecular Representations

Kenneth Atz[1,†], Francesca Grisoni[2,1,†*], Gisbert Schneider[1,3*]

[1]ETH Zurich, Dept. Chemistry and Applied Biosciences, RETHINK, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland.
[2]Eindhoven University of Technology, Dept. Biomedical Engineering, Groene Loper 7, 5612AZ Eindhoven, Netherlands.
[3]ETH Singapore SEC Ltd, 1 CREATE Way, #06-01 CREATE Tower, Singapore, Singapore.
† these authors contributed equally to this work
*f.grisoni@tue.nl, gisbert@ethz.ch

Abstract

Geometric deep learning (GDL), which is based on neural network architectures that incorporate and process symmetry information, has emerged as a recent paradigm in artificial intelligence. GDL bears particular promise in molecular modeling applications, in which various molecular representations with different symmetry properties and levels of abstraction exist. This review provides a structured and harmonized overview of molecular GDL, highlighting its applications in drug discovery, chemical synthesis prediction, and quantum chemistry. Emphasis is placed on the relevance of the learned molecular features and their complementarity to well-established molecular descriptors. This review provides an overview of current challenges and opportunities, and presents a forecast of the future of GDL for molecular sciences.

1 Introduction

Recent advances in deep learning, which is an instance of artificial intelligence (AI) based on neural networks [1, 2], have led to numerous applications in the molecular sciences, *e.g.*, in drug discovery [3, 4], quantum chemistry [5], and structural biology [6, 7]. Two characteristics of deep learning render it particularly promising when applied to molecules. First, deep learning methods can cope with "unstructured" data representations, such as text sequences [8, 9], speech signals [10, 11], images [12–14], and graphs [15, 16]. This ability is particularly useful for molecular systems, for which chemists have developed many models (*i.e.*, "molecular representations") that capture molecular properties at varying levels of abstraction (Figure 1). The second key characteristic is that deep learning can perform feature extraction (or feature learning) from the input data, that is, produce data-driven features from the input data without the need for manual intervention. These two characteristics are promising for deep learning as a complement to "classical" machine learning applications (*e.g.*, Quantitative Structure-Activity Relationship [QSAR]), in which molecular features (*i.e.*, "molecular descriptors" [17]) are encoded *a priori* with rule-based algorithms. The capability to learn from unstructured data and obtain data-driven molecular features has led to unprecedented applications of AI in the molecular sciences.

One of the most promising advances in deep learning is geometric deep learning (GDL). *Geometric deep learning* is an umbrella term encompassing emerging techniques which generalize neural networks to Euclidean and non-Euclidean domains, such as graphs, manifolds, meshes, or string representations [15]. In general, GDL encompasses approaches that incorporate a geometric prior, *i.e.*, information on the structure space and symmetry properties of the input variables. Such a geometric prior is leveraged to improve the quality of the information captured by the model. Although GDL has been increasingly applied to molecular modeling [5, 18, 19], its full potential in the field is still untapped.

Figure 1: *Exemplary molecular representations for a selected molecule (i.e., the penam substructure of penicillin)*
a. Two-dimensional (2D) depiction (Kekulé structure).
b. Molecular graph (2D), composed of vertices (atoms) and edges (bonds).
c. SMILES string [20], in which atom type, bond type and connectivity are specified by alphanumerical characters.
d. Three-dimensional (3D) graph, composed of vertices (atoms), their position (x, y, z coordinates) in 3D space, and edges (bonds).
e. Molecular surface represented as a mesh colored according to the respective atom types.

The aim of this review is to (i) provide a structured and harmonized overview of the applications of GDL on molecular systems, (ii) delineate the main research directions in the field, and (iii) provide a forecast of the future impact of GDL. Three fields of application are highlighted, namely drug discovery, quantum chemistry, and computer-aided synthesis planning (CASP),

1

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

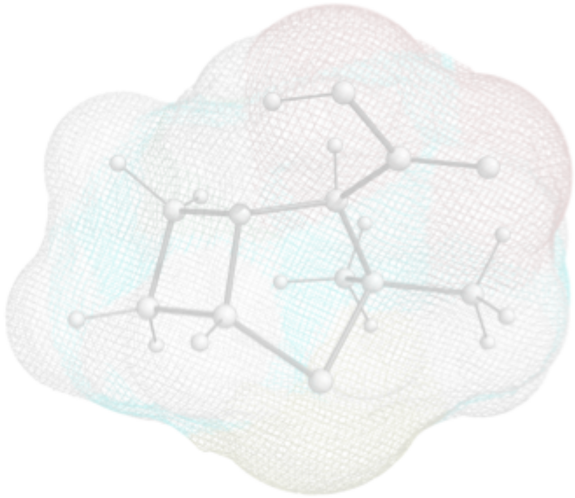
[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

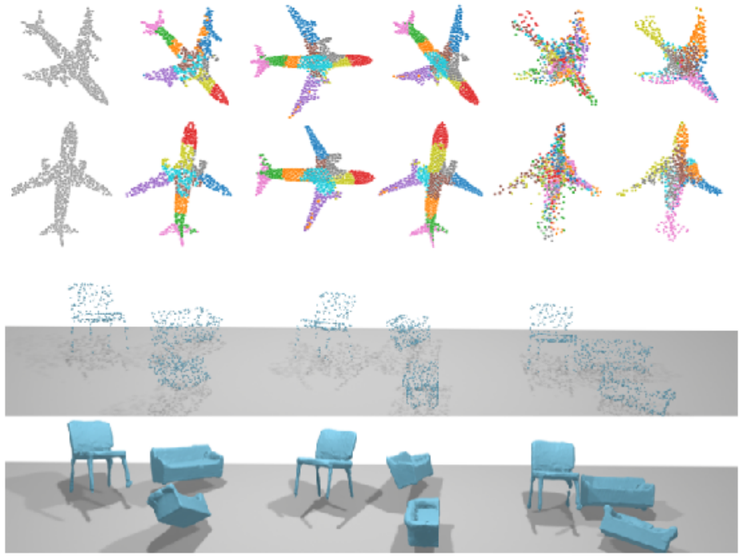
Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

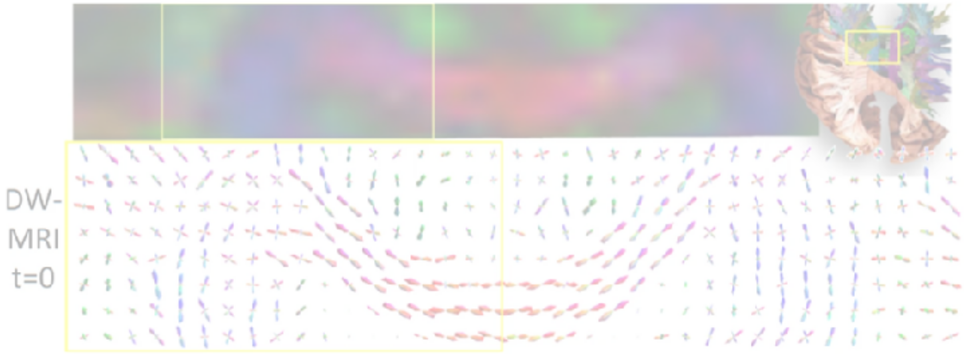
Computational Physics [1,*]



Computational Chemistry [2]



3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

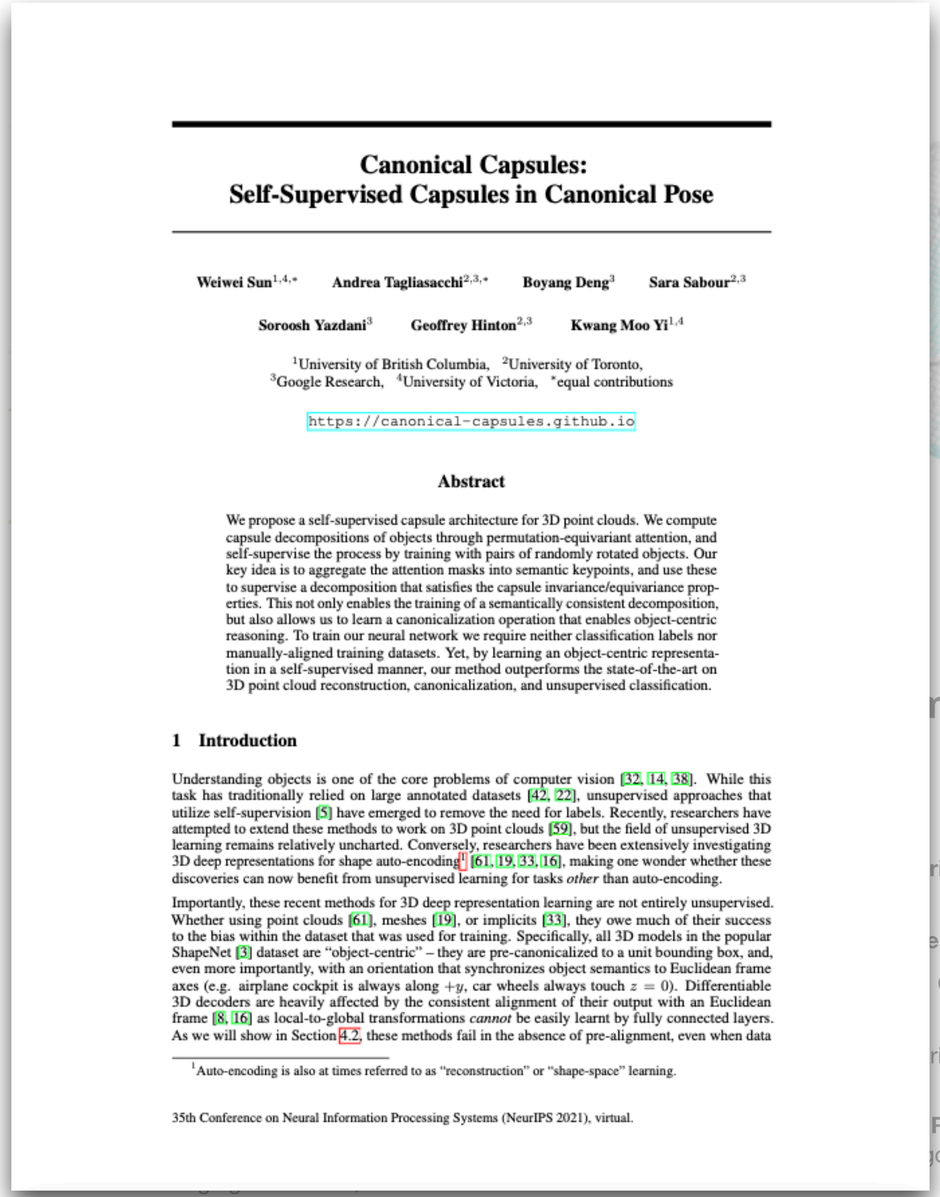
[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.



Canonical Capsules:
Self-Supervised Capsules in Canonical Pose

Weiwei Sun[1,4,*] Andrea Tagliasacchi[2,3,*] Boyang Deng[3] Sara Sabour[2,3]

Soroosh Yazdani[3] Geoffrey Hinton[2,3] Kwang Moo Yi[1,4]

[1]University of British Columbia, [2]University of Toronto,
[3]Google Research, [4]University of Victoria, *equal contributions

https://canonical-capsules.github.io

Abstract

We propose a self-supervised capsule architecture for 3D point clouds. We compute capsule decompositions of objects through permutation-equivariant attention, and self-supervise the process by training with pairs of randomly rotated objects. Our key idea is to aggregate the attention masks into semantic keypoints, and use these to supervise a decomposition that satisfies the capsule invariance/equivariance properties. This not only enables the training of a semantically consistent decomposition, but also allows us to learn a canonicalization operation that enables object-centric reasoning. To train our neural network we require neither classification labels nor manually-aligned training datasets. Yet, by learning an object-centric representation in a self-supervised manner, our method outperforms the state-of-the-art on 3D point cloud reconstruction, canonicalization, and unsupervised classification.

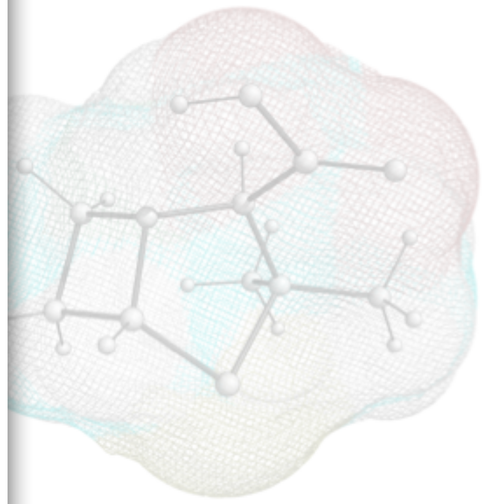
1 Introduction

Understanding objects is one of the core problems of computer vision [32, 14, 38]. While this task has traditionally relied on large annotated datasets [42, 22], unsupervised approaches that utilize self-supervision [5] have emerged to remove the need for labels. Recently, researchers have attempted to extend these methods to work on 3D point clouds [59], but the field of unsupervised 3D learning remains relatively uncharted. Conversely, researchers have been extensively investigating 3D deep representations for shape auto-encoding[1] [61, 19, 33, 16], making one wonder whether these discoveries can now benefit from unsupervised learning for tasks *other* than auto-encoding.

Importantly, these recent methods for 3D deep representation learning are not entirely unsupervised. Whether using point clouds [61], meshes [19], or implicits [33], they owe much of their success to the bias within the dataset that was used for training. Specifically, all 3D models in the popular ShapeNet [3] dataset are "object-centric" – they are pre-canonicalized to a unit bounding box, and, even more importantly, with an orientation that synchronizes object semantics to Euclidean frame axes (e.g. airplane cockpit is always along $+y$, car wheels always touch $z = 0$). Differentiable 3D decoders are heavily affected by the consistent alignment of their output with an Euclidean frame [8, 16] as local-to-global transformations *cannot* be easily learnt by fully connected layers. As we will show in Section 4.2, these methods fail in the absence of pre-alignment, even when data

[1]Auto-encoding is also at times referred to as "reconstruction" or "shape-space" learning.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), virtual.



nputational Chemistry [2]

ric and Physical Quantities improve E (3)

esentations. Nature Machine Intelligence, 1-10.

Canonical capsules: Unsupervised capsules in

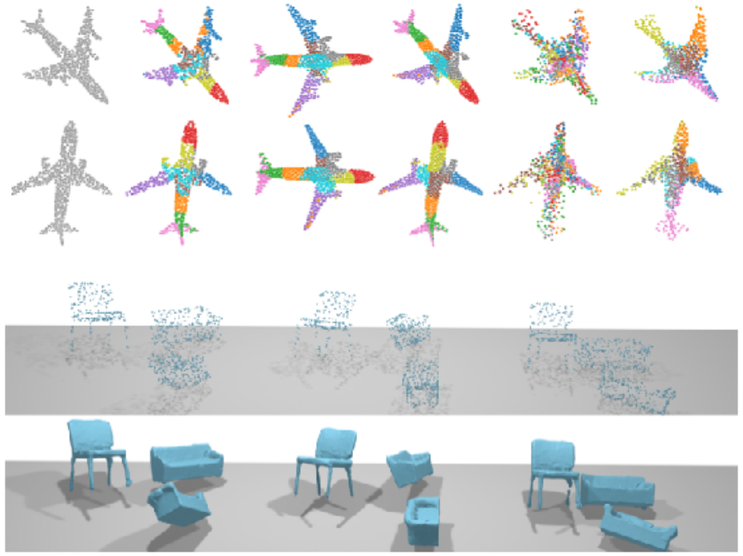
riant Attention Networks for Shape Reconstruction

Recent Geometric Flows in Multi-orientation
gorithms in Computer Vision and Imaging:

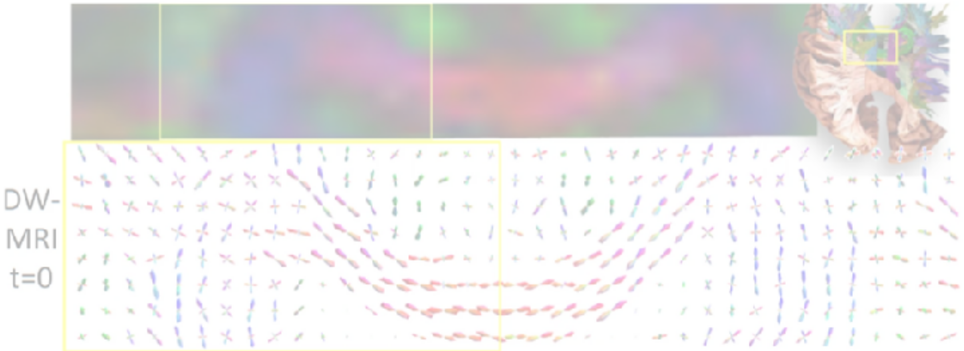
Also see:

*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

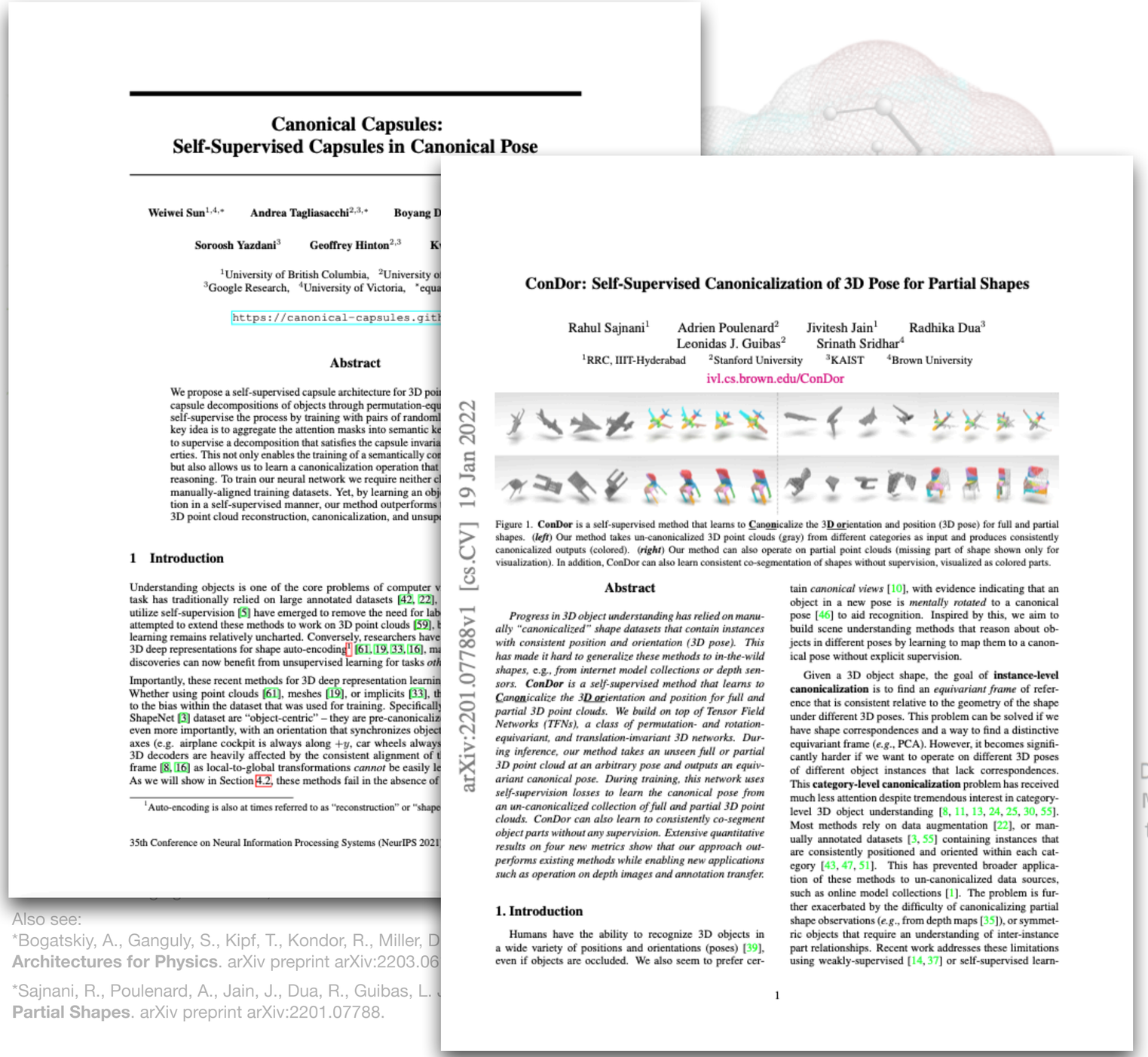
*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.



3D Computer Vision [3,4,*]



Medical Image Analysis [5]



Canonical Capsules:
Self-Supervised Capsules in Canonical Pose

Weiwei Sun[1,4,*] Andrea Tagliasacchi[2,3,*] Boyang D

Soroosh Yazdani[3] Geoffrey Hinton[2,3] K

[1]University of British Columbia, [2]University o
[3]Google Research, [4]University of Victoria, *equa

https://canonical-capsules.gith

Abstract

We propose a self-supervised capsule architecture for 3D poi
capsule decompositions of objects through permutation-equ
self-supervise the process by training with pairs of randoml
key idea is to aggregate the attention masks into semantic ke
to supervise a decomposition that satisfies the capsule invaria
erties. This not only enables the training of a semantically cor
but also allows us to learn a canonicalization operation that
reasoning. To train our neural network we require neither cl
manually-aligned training datasets. Yet, by learning an obj
tion in a self-supervised manner, our method outperforms
3D point cloud reconstruction, canonicalization, and unsup

1 Introduction

Understanding objects is one of the core problems of computer v
task has traditionally relied on large annotated datasets [42, 22],
utilize self-supervision [5] have emerged to remove the need for lab
attempted to extend these methods to work on 3D point clouds [59],
learning remains relatively uncharted. Conversely, researchers have
3D deep representations for shape auto-encoding[1] [61, 19, 33, 16], ma
discoveries can now benefit from unsupervised learning for tasks *oth*

Importantly, these recent methods for 3D deep representation learnin
Whether using point clouds [61], meshes [19], or implicits [33], th
to the bias within the dataset that was used for training. Specifically
ShapeNet [3] dataset are "object-centric" – they are pre-canonicaliz
even more importantly, with an orientation that synchronizes object
axes (e.g. airplane cockpit is always along $+y$, car wheels always
3D decoders are heavily affected by the consistent alignment of t
frame [8, 16] as local-to-global transformations *cannot* be easily le
As we will show in Section 4.2, these methods fail in the absence of

[1]Auto-encoding is also at times referred to as "reconstruction" or "shape

35th Conference on Neural Information Processing Systems (NeurIPS 2021

ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes

Rahul Sajnani[1] Adrien Poulenard[2] Jivitesh Jain[1] Radhika Dua[3]
Leonidas J. Guibas[2] Srinath Sridhar[4]
[1]RRC, IIIT-Hyderabad [2]Stanford University [3]KAIST [4]Brown University
ivl.cs.brown.edu/ConDor

Figure 1. **ConDor** is a self-supervised method that learns to Canonicalize the 3D orientation and position (3D pose) for full and partial shapes. (*left*) Our method takes un-canonicalized 3D point clouds (gray) from different categories as input and produces consistently canonicalized outputs (colored). (***right***) Our method can also operate on partial point clouds (missing part of shape shown only for visualization). In addition, ConDor can also learn consistent co-segmentation of shapes without supervision, visualized as colored parts.

Abstract

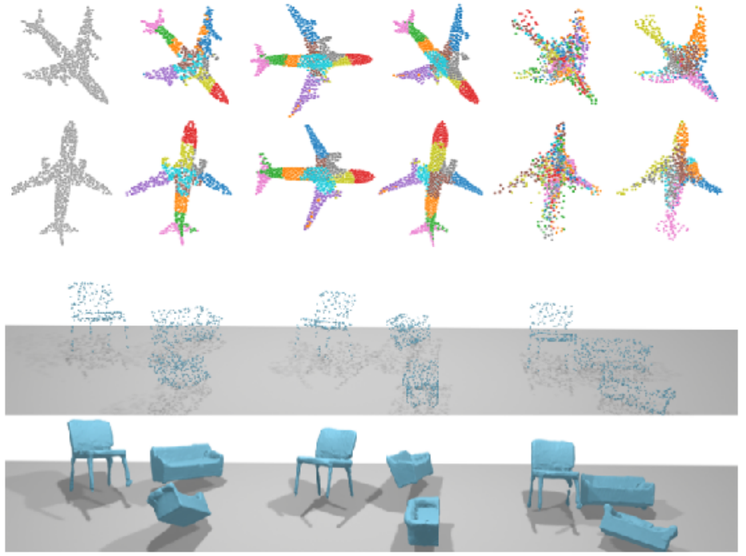
Progress in 3D object understanding has relied on manually "canonicalized" shape datasets that contain instances with consistent position and orientation (3D pose). This has made it hard to generalize these methods to in-the-wild shapes, e.g., from internet model collections or depth sensors. ***ConDor*** *is a self-supervised method that learns to Canonicalize the 3D orientation and position for full and partial 3D point clouds. We build on top of Tensor Field Networks (TFNs), a class of permutation- and rotation-equivariant, and translation-invariant 3D networks. During inference, our method takes an unseen full or partial 3D point cloud at an arbitrary pose and outputs an equivariant canonical pose. During training, this network uses self-supervision losses to learn the canonical pose from an un-canonicalized collection of full and partial 3D point clouds. ConDor can also learn to consistently co-segment object parts without any supervision. Extensive quantitative results on four new metrics show that our approach outperforms existing methods while enabling new applications such as operation on depth images and annotation transfer.*

1. Introduction

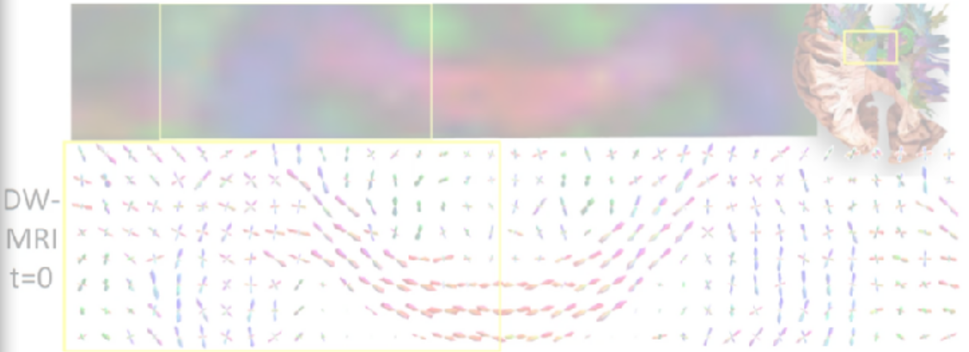
Humans have the ability to recognize 3D objects in a wide variety of positions and orientations (poses) [39], even if objects are occluded. We also seem to prefer certain *canonical views* [10], with evidence indicating that an object in a new pose is *mentally rotated* to a canonical pose [46] to aid recognition. Inspired by this, we aim to build scene understanding methods that reason about objects in different poses by learning to map them to a canonical pose without explicit supervision.

Given a 3D object shape, the goal of **instance-level canonicalization** is to find an *equivariant frame* of reference that is consistent relative to the geometry of the shape under different 3D poses. This problem can be solved if we have shape correspondences and a way to find a distinctive equivariant frame (*e.g.*, PCA). However, it becomes significantly harder if we want to operate on different 3D poses of different object instances that lack correspondences. This **category-level canonicalization** problem has received much less attention despite tremendous interest in category-level 3D object understanding [8, 11, 13, 24, 25, 30, 55]. Most methods rely on data augmentation [22], or manually annotated datasets [3, 55] containing instances that are consistently positioned and oriented within each category [43, 47, 51]. This has prevented broader application of these methods to un-canonicalized data sources, such as online model collections [1]. The problem is further exacerbated by the difficulty of canonicalizing partial shape observations (*e.g.*, from depth maps [35]), or symmetric objects that require an understanding of inter-instance part relationships. Recent work addresses these limitations using weakly-supervised [14, 37] or self-supervised learn-

1



3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Also see:

*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D
Architectures for Physics. arXiv preprint arXiv:2203.06

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L.
Partial Shapes. arXiv preprint arXiv:2201.07788.

Canonical Capsules: Self-Supervised Capsules in Canonical Pose

Weiwei Sun[1,4,*] Andrea Tagliasacchi[2,3,*] Boyang D
Soroosh Yazdani[3] Geoffrey Hinton[2,3] K
[1]University of British Columbia, [2]University o
[3]Google Research, [4]University of Victoria, *equa

https://canonical-capsules.gith

Abstract

We propose a self-supervised capsule architecture for 3D poi
capsule decompositions of objects through permutation-equ
self-supervise the process by training with pairs of randoml
key idea is to aggregate the attention masks into semantic ke
to supervise a decomposition that satisfies the capsule invaria
erties. This not only enables the training of a semantically cor
but also allows us to learn a canonicalization operation that
reasoning. To train our neural network we require neither cl
manually-aligned training datasets. Yet, by learning an obj
tion in a self-supervised manner, our method outperforms
3D point cloud reconstruction, canonicalization, and unsup

1 Introduction

Understanding objects is one of the core problems of computer v
task has traditionally relied on large annotated datasets [42, 22],
utilize self-supervision [5] have emerged to remove the need for lab
attempted to extend these methods to work on 3D point clouds [59],
learning remains relatively uncharted. Conversely, researchers have
3D deep representations for shape auto-encoding[1] [61, 19, 33, 16], ma
discoveries can now benefit from unsupervised learning for tasks *oth*

Importantly, these recent methods for 3D deep representation learnin
Whether using point clouds [61], meshes [19], or implicits [33], th
to the bias within the dataset that was used for training. Specifically
ShapeNet [3] dataset are "object-centric" – they are pre-canonicaliz
even more importantly, with an orientation that synchronizes object
axes (e.g. airplane cockpit is always along $+y$, car wheels always
3D decoders are heavily affected by the consistent alignment of t
frame [8, 16] as local-to-global transformations *cannot* be easily le
As we will show in Section 4.2, these methods fail in the absence of

[1]Auto-encoding is also at times referred to as "reconstruction" or "shape

35th Conference on Neural Information Processing Systems (NeurIPS 2021

ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes

Rahul Sajnani[1] Adrien Poulenard[2] Jivitesh Jain[1] Radhika Dua[3]
Leonidas J. Guibas[2] Srinath Sridhar[4]
[1]RRC, IIIT-Hyderabad [2]Stanford University [3]KAIST [4]Brown University

ivl.cs.brown.edu/ConDor

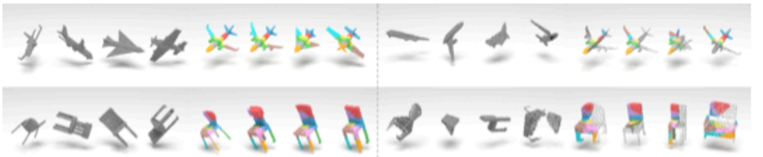


Figure 1. **ConDor** is a self-supervised method that learns to Canonicalize the 3D orientation and position (3D pose) for full and partial shapes. (*left*) Our method takes un-canonicalized 3D point clouds (gray) from different categories as input and produces consistently canonicalized outputs (colored). (*right*) Our method can also operate on partial point clouds (missing part of shape shown only for visualization). In addition, ConDor can also learn consistent co-segmentation of shapes without supervision, visualized as colored parts.

Abstract

*Progress in 3D object understanding has relied on manually "canonicalized" shape datasets that contain instances with consistent position and orientation (3D pose). This has made it hard to generalize these methods to in-the-wild shapes, e.g., from internet model collections or depth sensors. **ConDor** is a self-supervised method that learns to Canonicalize the 3D orientation and position for full and partial 3D point clouds. We build on top of Tensor Field Networks (TFNs), a class of permutation- and rotation-equivariant, and translation-invariant 3D networks. During inference, our method takes an unseen full or partial 3D point cloud at an arbitrary pose and outputs an equivariant canonical pose. During training, this network uses self-supervision losses to learn the canonical pose from an un-canonicalized collection of full and partial 3D point clouds. ConDor can also learn to consistently co-segment object parts without any supervision. Extensive quantitative results on four new metrics show that our approach outperforms existing methods while enabling new applications such as operation on depth images and annotation transfer.*

1. Introduction

Humans have the ability to recognize 3D objects in a wide variety of positions and orientations (poses) [39], even if objects are occluded. We also seem to prefer certain *canonical views* [10], with evidence indicating that an object in a new pose is *mentally rotated* to a canonical pose [46] to aid recognition. Inspired by this, we aim to build scene understanding methods that reason about objects in different poses by learning to map them to a canonical pose without explicit supervision.

Given a 3D object shape, the goal of **instance-level canonicalization** is to find an *equivariant frame* of reference that is consistent relative to the geometry of the shape under different 3D poses. This problem can be solved if we have shape correspondences and a way to find a distinctive equivariant frame (*e.g.*, PCA). However, it becomes significantly harder if we want to operate on different 3D poses of different object instances that lack correspondences. This **category-level canonicalization** problem has received much less attention despite tremendous interest in category-level 3D object understanding [8, 11, 13, 24, 25, 30, 55]. Most methods rely on data augmentation [22], or manually annotated datasets [3, 55] containing instances that are consistently positioned and oriented within each category [43, 47, 51]. This has prevented broader application of these methods to un-canonicalized data sources, such as online model collections [1]. The problem is further exacerbated by the difficulty of canonicalizing partial shape observations (*e.g.*, from depth maps [35]), or symmetric objects that require an understanding of inter-instance part relationships. Recent work addresses these limitations using weakly-supervised [14, 37] or self-supervised learn-

SE(3)-Equivariant Attention Networks for Shape Reconstruction in Function Space

Evangelos Chatzipantazis*[1], Stefanos Pertigkiozoglou*[1], Edgar Dobriban[2], and Kostas Daniilidis[1]

[1] Department of Computer and Information Science, University of Pennsylvania
[2] Department of Statistics and Data Science, University of Pennsylvania

vaghat@seas.upenn.edu, pstefano@seas.upenn.edu,
dobriban@wharton.upenn.edu, kostas@cis.upenn.edu

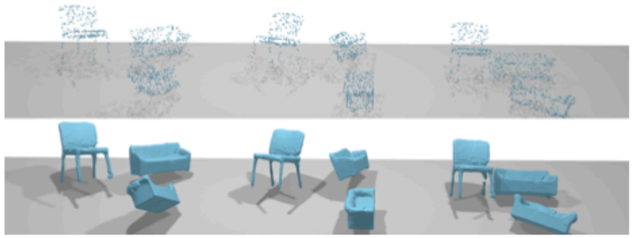


Fig. 1: (Above): Sparse point clouds individually SE(3)-transformed to form a single scene of nine objects. (Below): Our equivariant reconstruction. The network is agnostic to the number, position and orientation of the objects and is trained only on single objects in canonical pose.

Abstract. We propose the first SE(3)-equivariant coordinate-based network for learning occupancy fields from point clouds. In contrast to previous shape reconstruction methods that align the input to a regular grid, we operate directly on the irregular, unoriented point cloud. We leverage attention mechanisms in order to preserve the set structure (permutation equivariance and variable length) of the input. At the same time, attention layers enable local shape modelling, a crucial property for scalability to large scenes. In contrast to architectures that create a global signature for the shape, we operate on local tokens. Given an unoriented, sparse, noisy point cloud as input, we produce equivariant features for each point. These serve as keys and values for the subsequent equivariant cross-attention blocks that parametrize the occupancy field. By querying an arbitrary point in space, we predict its occupancy score. We show

* Equal contribution

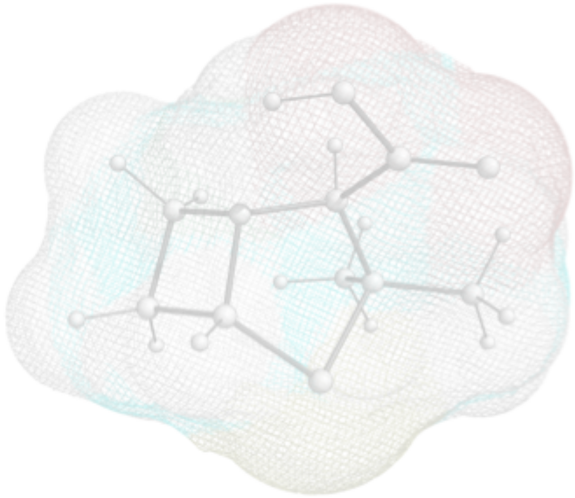
DW-MRI t=0

Also see:

*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D
Architectures for Physics. arXiv preprint arXiv:2203.06

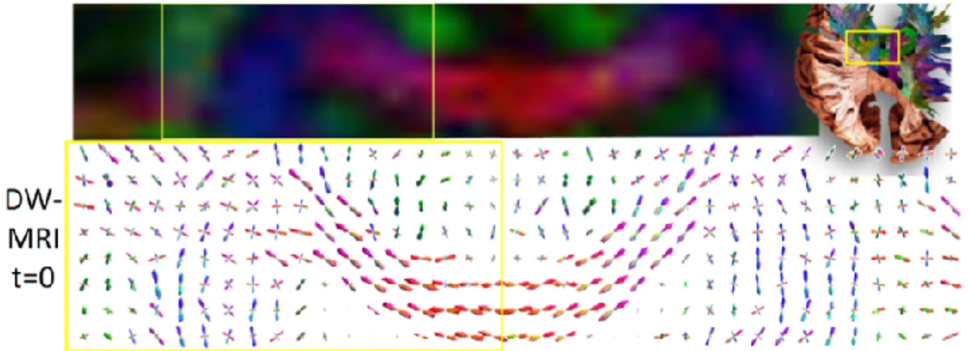
*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L.
Partial Shapes. arXiv preprint arXiv:2201.07788.

Computational Physics [1,*]



Computational Chemistry [2]

3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

Comput

Chemistry [2]

3D Computer Vision [3,4,*]

Figure sources:

[1]Brandstetter, J., Hesseli
Equivariant Message Pa

[2]Atz, K., Grisoni, F., & Sc

[3]Sun, W., Tagliasacchi, A
canonical pose. NeurIPS

[4]Chatzipantazis, E., Perti
in Function Space. arXiv

[5]Duits, R., Smets, B. M. N
Image Processing via a
Mathematical Imaging an

Also see:
*Bogatskiy, A., Ganguly, S
Architectures for Physic

*Sajnani, R., Poulenard, A
Partial Shapes. arXiv pre

ntities improve E (3)

lachine Intelligence, 1-10.

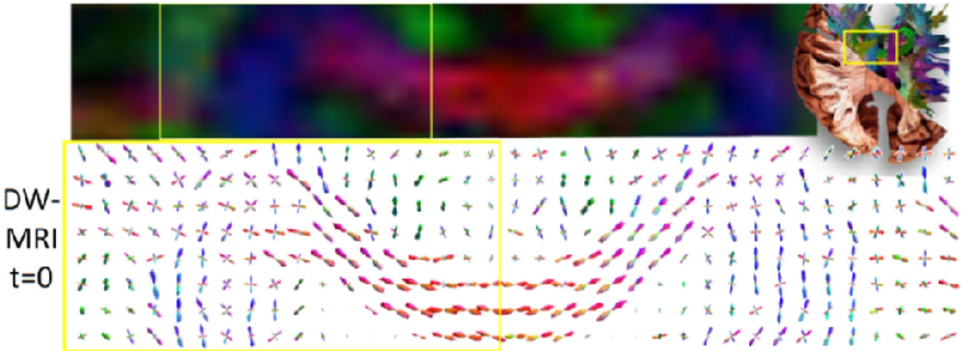
Unsupervised capsules in

ks for Shape Reconstruction

ows in Multi-orientation
Vision and Imaging:

p Equivariant

nicalization of 3D Pose for



Medical Image Analysis [5]

Check for updates

Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection

R. Duits, B. M. N. Smets, A. J. Wemmenhove, J. W. Portegies, and E. J. Bekkers

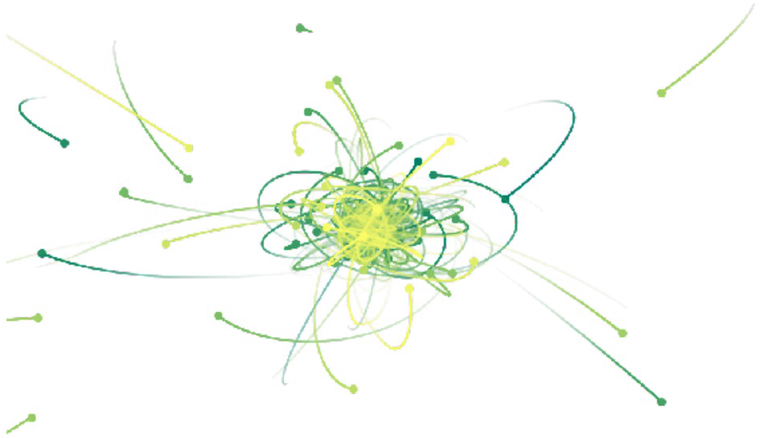
Contents

R. Duits (✉) · B. Smets · J. Wemmenhove
Applied Differential Geometry, Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands
e-mail: R.Duits@tue.nl; B.M.N.Smets@tue.nl; a.j.wemmenhove@tue.nl

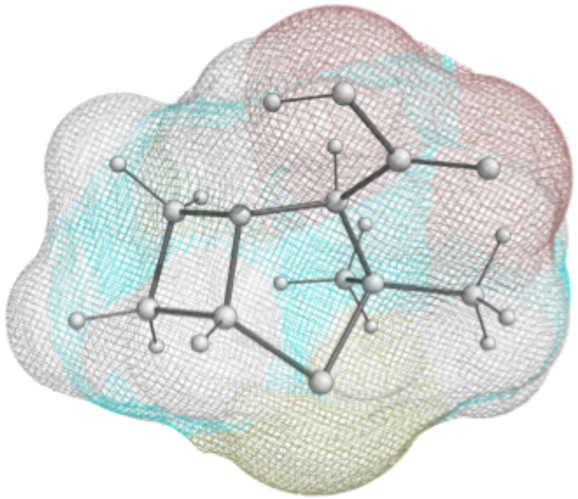
E. Bekkers
Amsterdam Machine Learning Lab, University of Amsterdam, Amsterdam, Netherlands
e-mail: E.J.Bekkers@uva.nl

J. Portegies
Center for Analysis, Scientific Computing and Applications, Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands
e-mail: J.W.Portegies@tue.nl

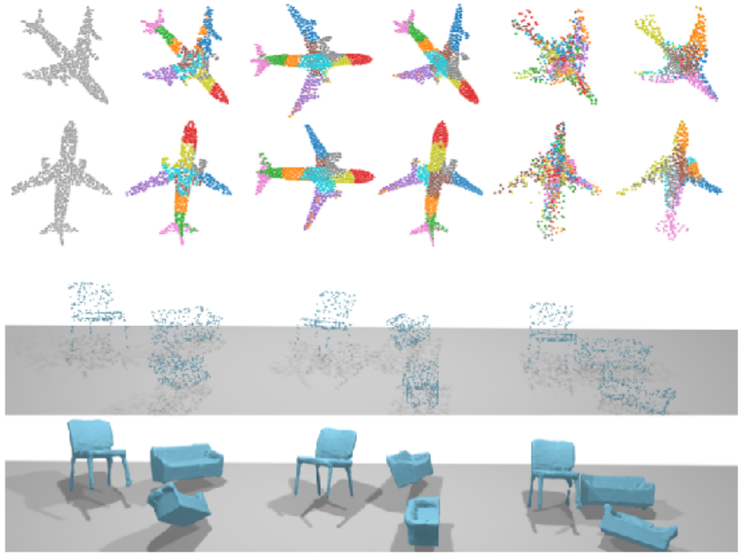
K. Chen et al. (eds.), *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging*, https://doi.org/10.1007/978-3-030-03009-4_101-1



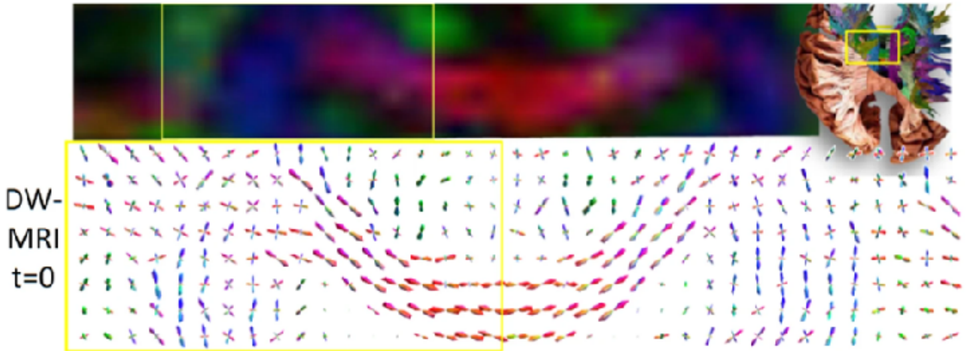
Computational Physics [1,*]



Computational Chemistry [2]



3D Computer Vision [3,4,*]



Medical Image Analysis [5]

Figure sources:

[1]Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E., & Welling, M. (2021). **Geometric and Physical Quantities improve E (3) Equivariant Message Passing.** In ICLR 2022

[2]Atz, K., Grisoni, F., & Schneider, G. (2021). **Geometric deep learning on molecular representations**. Nature Machine Intelligence, 1-10.

[3]Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G., & Yi, K. M. (2020). **Canonical capsules: Unsupervised capsules in canonical pose**. NeurIPS 2021

[4]Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., & Daniilidis, K. (2022). SE (3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. arXiv preprint arXiv:2204.02394.

[5]Duits, R., Smets, B. M. N., Wemmenhove, A. J., Portegies, J. W., & Bekkers, E. J. (2021). **Recent Geometric Flows in Multi-orientation Image Processing via a Cartan Connection**. Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, 1-60.

Also see:
*Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., ... & Thais, S. (2022). **Symmetry Group Equivariant Architectures for Physics**. arXiv preprint arXiv:2203.06153.

*Sajnani, R., Poulenard, A., Jain, J., Dua, R., Guibas, L. J., & Sridhar, S. (2022). **ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes**. arXiv preprint arXiv:2201.07788.

Steerable G-CNNs as Clebsch-Gordan networks

$$\hat{\mathcal{K}}(\hat{f})(\mathbf{x}) = \int_{\mathbb{R}^d} \hat{f}(\mathbf{x}') \otimes_{CG}^{\hat{w}(\|\mathbf{x}'-\mathbf{x}\|)} Y(\alpha_{\mathbf{x}'-\mathbf{x}})\, \mathrm{d}\mathbf{x}'$$

11

Figure sources:

[1]Brandstetter, J.
Equivariant Me

[2]Atz, K., Grisoni,

[3]Sun, W., Taglias
canonical pose

[4]Chatzipantazis,
in Function Spac

[5]Duits, R., Smet
Image Processi
Mathematical Im

Also see:
*Bogatskiy, A., G
Architectures f

*Sajnani, R., Pou
Partial Shapes.