Symmetries and Neural Tangent Kernels

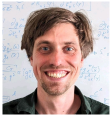
Jan E. Gerken

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

WASP
WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

in collaboration with



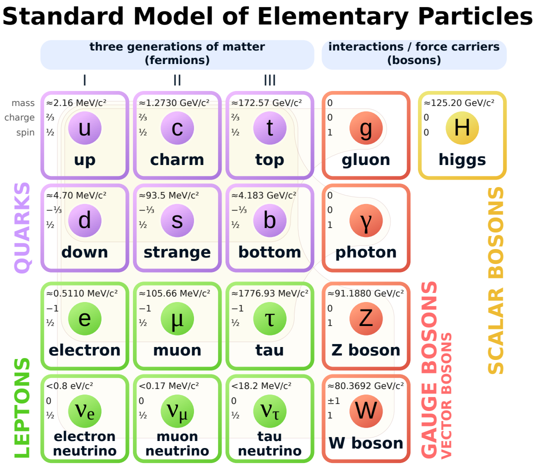
Pan Kessel

from Prescient Design
A Genentech Accelerator

IAIFI Summer Workshop 2024
August 12–August 16

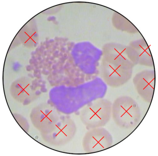
Symmetries in physics

$SU(2) \times SU(3) \times U(1)$ ⟷

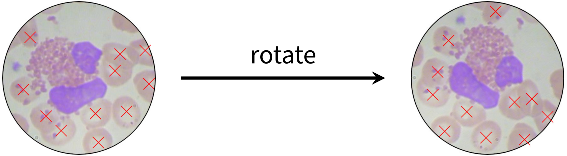


Symmetries in deep learning

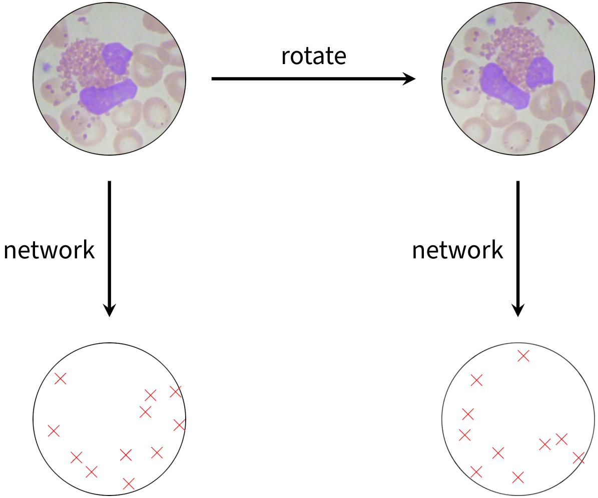
Symmetries in deep learning



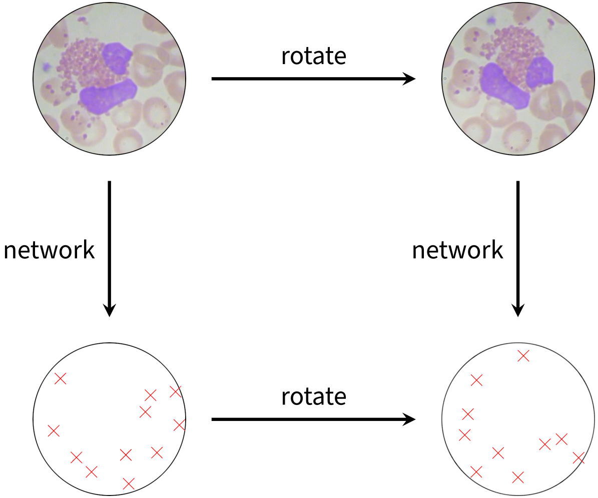
Symmetries in deep learning



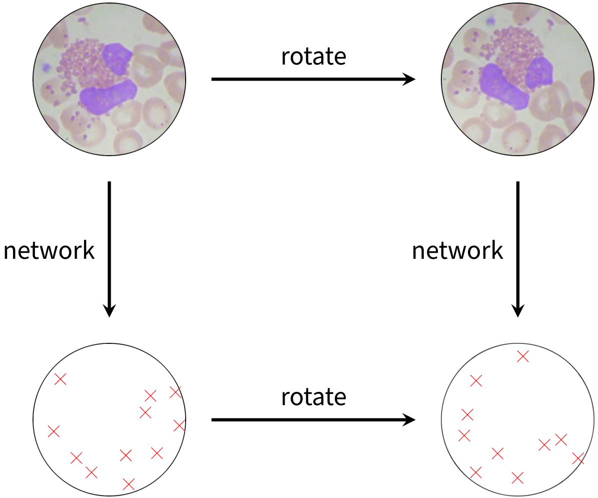
Symmetries in deep learning



Symmetries in deep learning

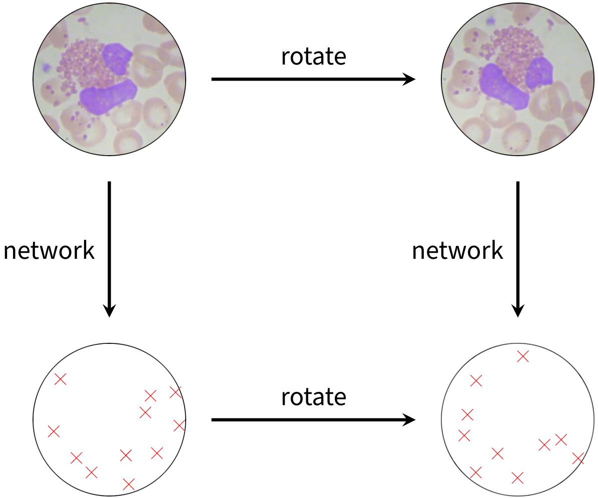


Symmetries in deep learning



$$\iff \mathcal{N}(\rho_{\text{in}}(g)x) = \rho_{\text{out}}(g)\mathcal{N}(x)$$

Symmetries in deep learning



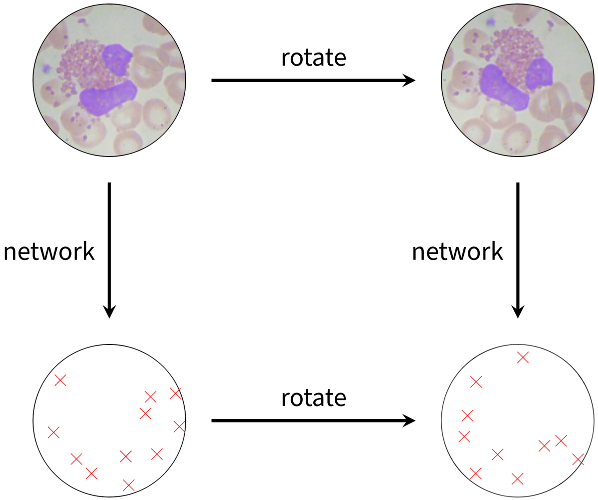
network

$$\Longleftrightarrow \quad \mathcal{N}(\rho_{\text{in}}(g)x) = \rho_{\text{out}}(g)\mathcal{N}(x)$$

rotation

image

Equivariance



network

$$\Longleftrightarrow \quad \mathcal{N}(\rho_{\text{in}}(g)x) = \rho_{\text{out}}(g)\mathcal{N}(x)$$

rotation

image

Equivariant neural networks

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant Transformer Networks

Kai Sheng Tai [1] **Peter Bailis** [1] **Gregory Valiant** [1]

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canonical coordinate systems, ETs incorporate functions that

scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are guaranteed to respond predictably to particular transforma-

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant Transformer Networks

Kai Sheng Tai [1] **Peter Bailis** [1] **Gregory Valiant** [1]

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canonical coordinate systems, ETs incorporate functions that

scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are guaranteed to respond predictably to particular transforma-

Theory for Equivariant Quantum Neural Networks

Quynh T. Nguyen,[1,2] Louis Schatzki,[3,4] Paolo Braccia,[1,5] Michael Ragone,[1,6] Patrick J. Coles,[1] Frédéric Sauvage,[1] Martín Larocca,[1,7] and M. Cerezo[3]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[5] *Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019 , Italy*
[6] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action com-

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant Transformer Networks

Kai Sheng Tai [1] **Peter Bailis** [1] **Gregory Valiant** [1]

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canonical coordinate systems, ETs incorporate functions that

scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are guaranteed to respond predictably to particular transforma-

Theory for Equivariant Quantum Neural Networks

Quynh T. Nguyen,[1,2] Louis Schatzki,[3,4] Paolo Braccia,[1,5] Michael Ragone,[1,6]
Patrick J. Coles,[1] Frédéric Sauvage,[1] Martín Larocca,[1,7] and M. Cerezo[3]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[5] *Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019 , Italy*
[6] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action commutes with that of the symmetry. In this work, we import these ideas to the quantum realm by

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong[a,e,1] **Qi Meng**[b] **Jue Zhang**[b] **Huilin Qu**[c] **Congqiao Li**[d] **Sitian Qian**[d] **Weitao Du**[a] **Zhi-Ming Ma**[a] **Tie-Yan Liu**[b]

[a] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China*
[b] *Microsoft Research Asia, Danling Street, Beijing 100080, China*
[c] *CERN, EP Department, CH-1211 Geneva 23, Switzerland*
[d] *School of Physics, Peking University, Chengfu Road, Beijing 100871, China*

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant Transformer Networks

Kai Sheng Tai [1] **Peter Bailis** [1] **Gregory Valiant** [1]

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canonical co-

scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are

Theory for Equivariant Quantum Neural Networks

Quynh T. Nguyen,[1,2] Louis Schatzki,[3,4] Paolo Braccia,[1,5] Michael Ragone,[1,6] Patrick J. Coles,[1] Frédéric Sauvage,[1] Martín Larocca,[1,7] and M. Cerezo[3]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[5] *Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019 , Italy*
[6] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action com-

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong[a,e,1] Qi Meng[b] Jue Zhang[b] Huilin Qu[c] Congqiao Li[d] Sitian Qian[d] Weitao Du[a] Zhi-Ming Ma[a] Tie-Yan Liu[b]

[a] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China*
[b] *Microsoft Research Asia, Danling Street, Beijing 100080, China*
[c] *CERN, EP Department, CH-1211 Geneva 23, Switzerland*
[d] *School of Physics, Peking University, Chengfu Road, Beijing 100871, China*

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

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen T.S.COHEN@UVA.NL
University of Amsterdam

Max Welling M.WELLING@UVA.NL
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can be used effectively in a *deep* network because all the layers in such a network are *translation equivariant*: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature maps (at least up to edge-effects). In

Equivariant Transformer Networks

Kai Sheng Tai [1] **Peter Bailis** [1] **Gregory Valiant** [1]

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canonical coordinate systems, ETs incorporate functions that

scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are guaranteed to respond predictably to particular transforma-

Theory for Equivariant Quantum Neural Networks

Quynh T. Nguyen,[1,2] Louis Schatzki,[3,4] Paolo Braccia,[1,5] Michael Ragone,[1,6]
Patrick J. Coles,[1] Frédéric Sauvage,[1] Martín Larocca,[1,7] and M. Cerezo[3]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[5] *Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019 , Italy*
[6] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action commutes with that of the symmetry. In this work, we import these ideas to the quantum realm by

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong[a,e,1] **Qi Meng**[b] **Jue Zhang**[b] **Huilin Qu**[c] **Congqiao Li**[d] **Sitian Qian**[d] **Weitao Du**[a] **Zhi-Ming Ma**[a] **Tie-Yan Liu**[b]

[a] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China*
[b] *Microsoft Research Asia, Danling Street, Beijing 100080, China*
[c] *CERN, EP Department, CH-1211 Geneva 23, Switzerland*
[d] *School of Physics, Peking University, Chengfu Road, Beijing 100871, China*

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

HIERARCHICAL, ROTATION-EQUIVARIANT NEURAL NETWORKS TO SELECT STRUCTURAL MODELS OF PROTEIN COMPLEXES

Stephan Eismann*
Department of Applied Physics
Stanford University
seismann@stanford.edu

Raphael J.L. Townshend*
Department of Computer Science
Stanford University
raphael@cs.stanford.edu

Nathaniel Thomas*
Department of Physics
Stanford University
nthomas103@gmail.com

Milind Jagota
Department of Electrical Engineering
Stanford University
mjagota@stanford.edu

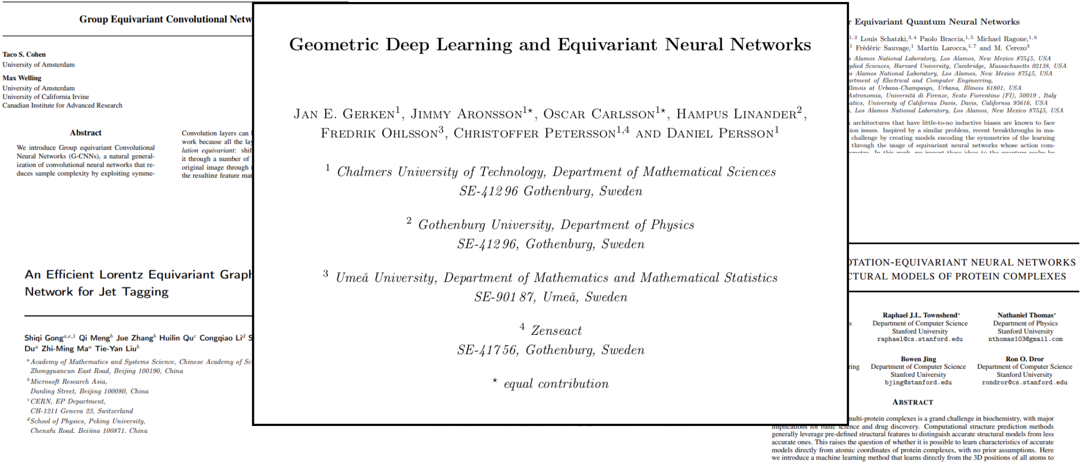
Bowen Jing
Department of Computer Science
Stanford University
bjing@stanford.edu

Ron O. Dror
Department of Computer Science
Stanford University
rondror@cs.stanford.edu

ABSTRACT

Predicting the structure of multi-protein complexes is a grand challenge in biochemistry, with major implications for basic science and drug discovery. Computational structure prediction methods generally leverage pre-defined structural features to distinguish accurate structural models from less accurate ones. This raises the question of whether it is possible to learn characteristics of accurate models directly from atomic coordinates of protein complexes, with no prior assumptions. Here we introduce a machine learning method that learns directly from the 3D positions of all atoms to

Equivariant neural networks



Group Equivariant Convolutional Netw

Taco S. Cohen
University of Amsterdam
Max Welling
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symme-

Convolution layers can
work because all the la
lation equivariant: shi
it through a number of
original image through
the resulting feature ma

An Efficient Lorentz Equivariant Graph
Network for Jet Tagging

Shiqi Gong[a,e,1] Qi Meng[b] Jue Zhang[b] Huilin Qu[c] Congqiao Li[d] S
Du[a] Zhi-Ming Ma[a] Tie-Yan Liu[b]

[a]*Academy of Mathematics and Systems Science, Chinese Academy of Sc*
Zhongguancun East Road, Beijing 100190, China
[b]*Microsoft Research Asia,*
Danling Street, Beijing 100080, China
[c]*CERN, EP Department,*
CH-1211 Geneva 23, Switzerland
[d]*School of Physics, Peking University,*
Chengfu Road, Beijing 100871, China

Geometric Deep Learning and Equivariant Neural Networks

Jan E. Gerken[1], Jimmy Aronsson[1⋆], Oscar Carlsson[1⋆], Hampus Linander[2],
Fredrik Ohlsson[3], Christoffer Petersson[1,4] and Daniel Persson[1]

[1] *Chalmers University of Technology, Department of Mathematical Sciences*
SE-412 96 Gothenburg, Sweden

[2] *Gothenburg University, Department of Physics*
SE-412 96, Gothenburg, Sweden

[3] *Umeå University, Department of Mathematics and Mathematical Statistics*
SE-901 87, Umeå, Sweden

[4] *Zenseact*
SE-417 56, Gothenburg, Sweden

⋆ *equal contribution*

Equivariant Quantum Neural Networks

Louis Schatzki,[3,4] Paolo Braccia,[1,5] Michael Ragone,[1,6]
Frédéric Sauvage,[1] Martín Larocca,[1,7] and M. Cerezo[3]

Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
plied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
rtment of Electrical and Computer Engineering,
llinois at Urbana-Champaign, Urbana, Illinois 61801, USA
Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019 , Italy
atics, University of California Davis, Davis, California 95616, USA
Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

architectures that have little-to-no inductive biases are known to face
ion issues. Inspired by a similar problem, recent breakthroughs in ma-
challenge by creating models encoding the symmetries of the learning
through the usage of equivariant neural networks whose action com-

OTATION-EQUIVARIANT NEURAL NETWORKS
CTURAL MODELS OF PROTEIN COMPLEXES

Raphael J.L. Townshend*
Department of Computer Science
Stanford University
raphael@cs.stanford.edu

Nathaniel Thomas*
Department of Physics
Stanford University
nthomas103@gmail.com

Bowen Jing
Department of Computer Science
Stanford University
bjing@stanford.edu

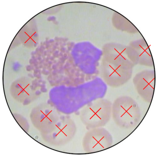
Ron O. Dror
Department of Computer Science
Stanford University
rondror@cs.stanford.edu

Abstract

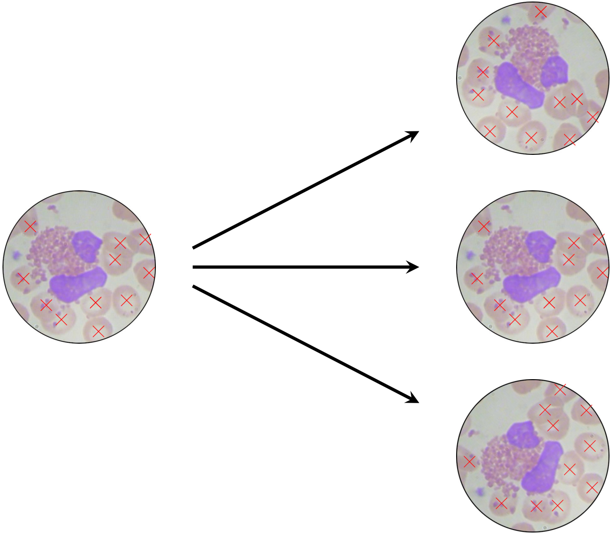
multi-protein complexes is a grand challenge in biochemistry, with major
implications for basic science and drug discovery. Computational structure prediction methods
generally leverage pre-defined structural features to distinguish accurate structural models from less
accurate ones. This raises the question of whether it is possible to learn characteristics of accurate
models directly from atomic coordinates of protein complexes, with no prior assumptions. Here
we introduce a machine learning method that learns directly from the 3D positions of all atoms to

Data augmentation

Data augmentation



Data augmentation



Data augmentation

- 👍 Easy to implement
- 👍 No specialized architecture necessary

Data augmentation

- 👍 Easy to implement
- 👍 No specialized architecture necessary
- 👎 No exact equivariance

Data augmentation

- Easy to implement
- No specialized architecture necessary
- No exact equivariance

Can we understand data augmentation theoretically?

Neural tangent kernel

Empirical NTK

Training dynamics under continuous gradient descent:

$$\frac{\mathrm{d}\mathcal{N}_\theta(x)}{\mathrm{d}t} = -\frac{\eta}{N}\sum_{i=1}^{N} \Theta_\theta(x, x_i) \frac{\partial L}{\partial \mathcal{N}(x_i)}$$

learning rate (η)

loss (L)

training sample (x_i)

Empirical NTK

Training dynamics under continuous gradient descent:

learning rate → η; loss → ∂L; training sample → x_i

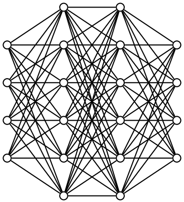
$$\frac{\mathrm{d}\mathcal{N}_\theta(x)}{\mathrm{d}t} = -\frac{\eta}{N}\sum_{i=1}^{N} \Theta_\theta(x, x_i) \frac{\partial L}{\partial \mathcal{N}(x_i)}$$

with the empirical neural tangent kernel (NTK)

$$\Theta_\theta(x, x') = \sum_\mu \frac{\partial \mathcal{N}(x)}{\partial \theta_\mu} \frac{\partial \mathcal{N}(x')}{\partial \theta_\mu}$$

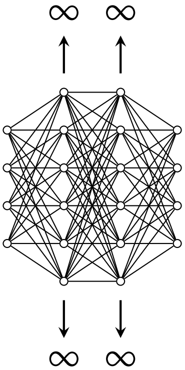
Infinite width limit

[Jacot et al. 2018]



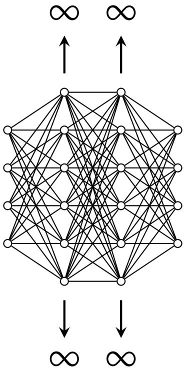
Infinite width limit

[Jacot et al. 2018]



Infinite width limit

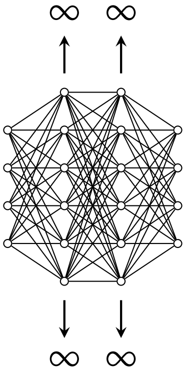
[Jacot et al. 2018]



👍 NTK becomes independent of initialization

Infinite width limit

[Jacot et al. 2018]

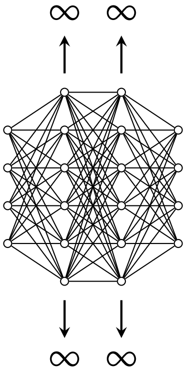


👍 NTK becomes independent of initialization

👍 NTK becomes constant in training

Infinite width limit

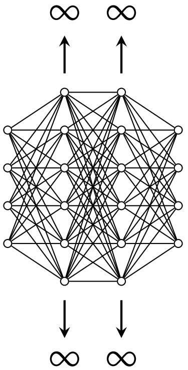
[Jacot et al. 2018]



- NTK becomes independent of initialization
- NTK becomes constant in training
- NTK can be computed for most networks

Infinite width limit

[Jacot et al. 2018]



- 👍 NTK becomes independent of initialization
- 👍 NTK becomes constant in training
- 👍 NTK can be computed for most networks
- ✓ Training dynamics can be solved

Mean prediction from NTK

[Jacot et al. 2018]

At infinite width, the mean prediction is given by

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

Mean prediction from NTK

[Jacot et al. 2018]

ⓘ At infinite width, the mean prediction is given by

neural tangent kernel

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

Mean prediction from NTK

[Jacot et al. 2018]

ⓘ At infinite width, the mean prediction is given by

neural tangent kernel

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

train data

Mean prediction from NTK

[Jacot et al. 2018]

At infinite width, the mean prediction is given by

neural tangent kernel

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

learning rate

train data

Mean prediction from NTK

[Jacot et al. 2018]

ⓘ At infinite width, the mean prediction is given by

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

neural tangent kernel → Θ

train labels → Y

train data → X

learning rate → η

Data augmentation

Data augmentation at infinite width

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

Data augmentation at infinite width

$$\mu_t(x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

augmented data

augmented labels

Data augmentation at infinite width

group transformation

$$\mu_t(\rho(g)x) = \Theta(\rho(g)x, X)\Theta(X, X)^{-1}(\mathbb{1} - e^{-\eta\Theta(X,X)t})Y$$

augmented data

augmented labels

Data augmentation at infinite width

group transformation

for augmented data

$$\mu_t(\rho(g)x) = \Theta(\rho(g)x, X)\,\Theta(X,X)^{-1}\left(\mathbb{I} - e^{-\eta\Theta(X,X)t}\right)Y$$

augmented data

augmented labels

Data augmentation at infinite width

group transformation

$$\mu_t(\rho(g)x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})\rho(g)Y$$

augmented data

augmented labels

Data augmentation at infinite width

group transformation

augmented labels

$$\mu_t(\rho(g)x) = \Theta(x,X)\Theta(X,X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})\underbrace{\rho(g)Y}_{=Y}$$

for invariance

Data augmentation at infinite width

group transformation

$$\mu_t(\rho(g)x) = \Theta(x,X)\Theta(X,X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})\underbrace{\rho(g)Y}_{=Y}$$

$$= \mu_t(x)$$

for invariance

Mean prediction

$\mu_t(x)$

Mean prediction

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}}[\mathcal{N}_{\theta_t}(x)]$$

Mean prediction

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}}[\mathcal{N}_{\theta_t}(x)] = \lim_{n \to \infty} \frac{1}{n} \sum_{\theta_0 = \text{init}_1}^{\text{init}_n} \mathcal{N}_{\theta_t}(x)$$

Mean prediction

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}}[\mathcal{N}_{\theta_t}(x)] = \lim_{n \to \infty} \underbrace{\frac{1}{n} \sum_{\theta_0 = \text{init}_1}^{\text{init}_n} \mathcal{N}_{\theta_t}(x)}_{\text{mean prediction of deep ensemble}}$$

Main conclusion

Deep ensembles trained with data augmentation are equivariant.

Main conclusion

Deep ensembles trained with data augmentation are equivariant.

- ✓ Proof of exact equivariance for
  - full data augmentation
  - infinite ensembles
  - at infinite width

Main conclusion

Deep ensembles trained with data augmentation are equivariant.

- ✓ Proof of exact equivariance for
  - full data augmentation
  - infinite ensembles
  - at infinite width
- ✓ Equivariance holds for all training times

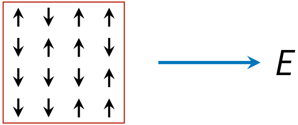
Main conclusion

Deep ensembles trained with data augmentation are equivariant.

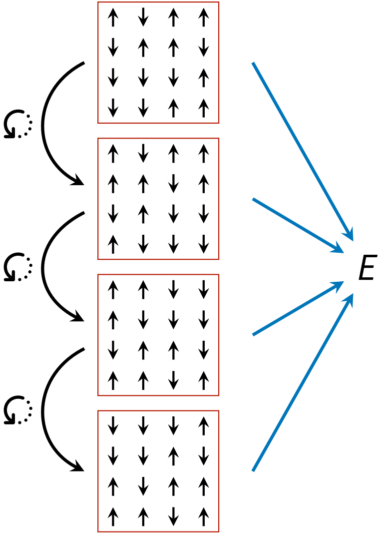
- ✓ Proof of exact equivariance for
  - full data augmentation
  - infinite ensembles
  - at infinite width
- ✓ Equivariance holds for all training times
- ✓ Equivariance holds away from the training data

Experiments

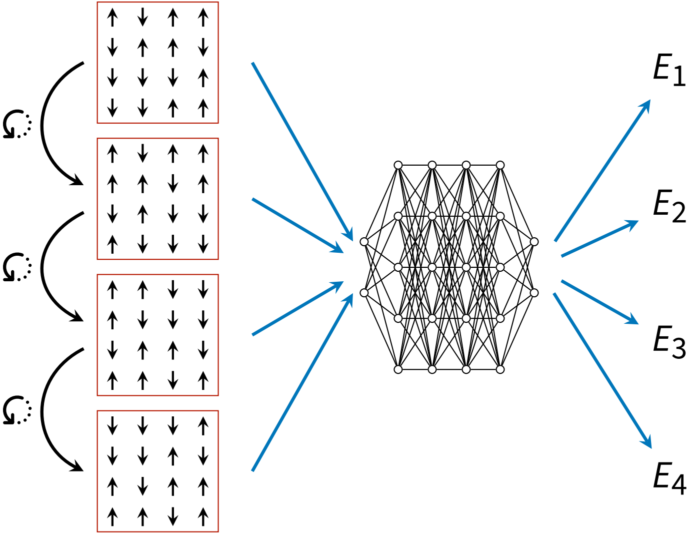
Ising model



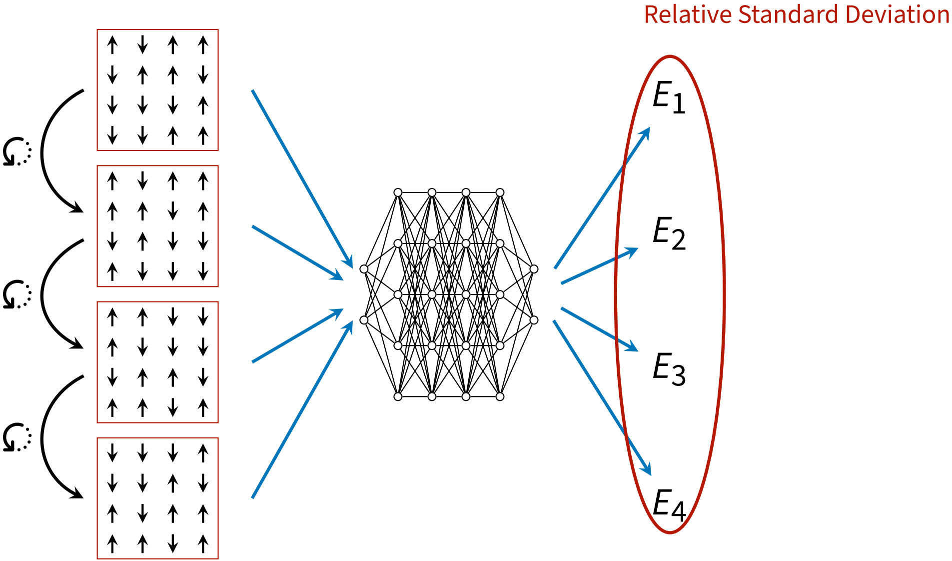
Ising model

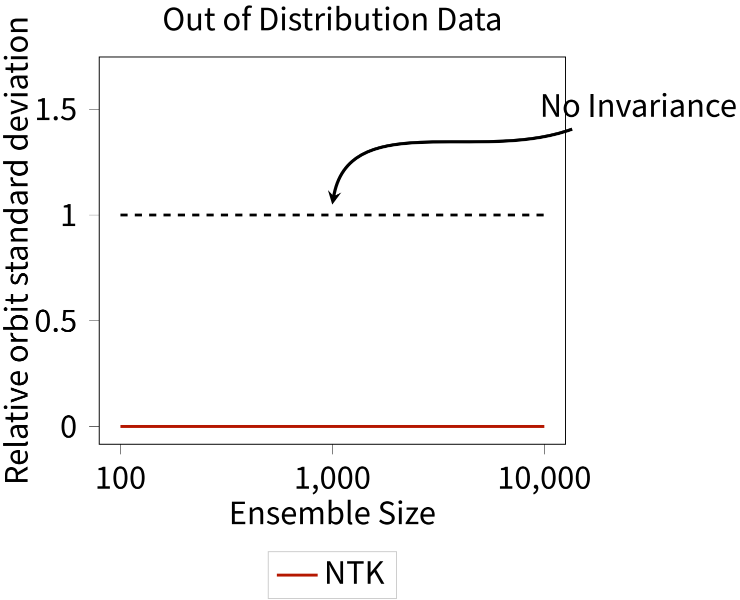


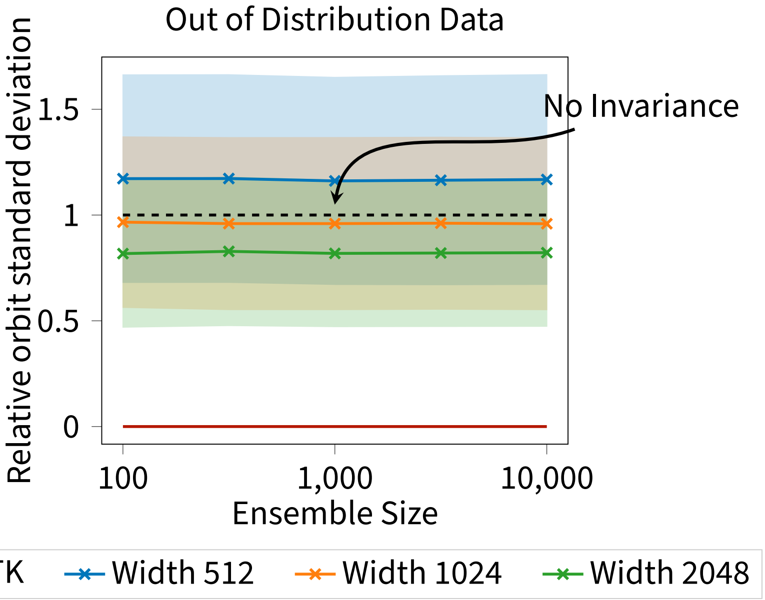
Ising model



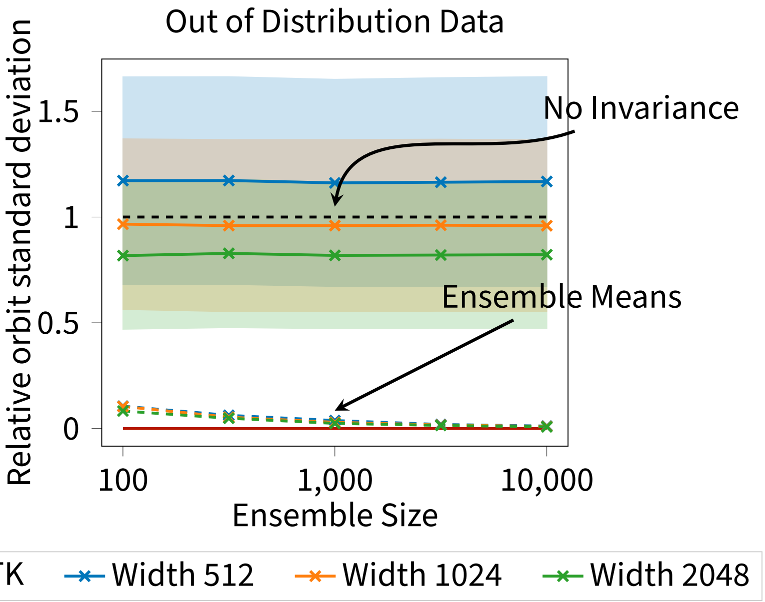
Ising model




Out of Distribution Data
No Invariance
Relative orbit standard deviation
1.5
1
0.5
0
100
1,000
10,000
Ensemble Size
NTK



Out of Distribution Data
No Invariance
Relative orbit standard deviation
1.5
1
0.5
0
100
1,000
10,000
Ensemble Size
NTK
Width 512
Width 1024
Width 2048


Out of Distribution Data
Relative orbit standard deviation
1.5
1
0.5
0
No Invariance
Ensemble Means
100
1,000
10,000
Ensemble Size
NTK
Width 512
Width 1024
Width 2048

Key takeaways

- 👍 Deep ensembles trained with data augmentation are equivariant

Key takeaways

- Deep ensembles trained with data augmentation are equivariant
- Using physics approaches in deep learning can be very fruitful

Key takeaways

- 👍 Deep ensembles trained with data augmentation are equivariant
- 👍 Using physics approaches in deep learning can be very fruitful
- 👍 Neural tangent kernels provide a powerful theoretical handle

Papers

- *Emergent Equivariance in Deep Ensembles*
  Jan E. Gerken*, Pan Kessel*
  ICML 2024 (Oral)

  * Equal contribution

Papers

- *Emergent Equivariance in Deep Ensembles*
  Jan E. Gerken*, Pan Kessel*
  ICML 2024 (Oral)

  * Equal contribution

- *Equivariant Neural Tangent Kernels*
  Philipp Misof, Pan Kessel, Jan E. Gerken
  Preprint

Papers

- *Emergent Equivariance in Deep Ensembles*
  Jan E. Gerken[*], Pan Kessel[*]
  ICML 2024 (Oral)

  [*] Equal contribution

- *Equivariant Neural Tangent Kernels*
  Philipp Misof, Pan Kessel, Jan E. Gerken
  Preprint

Group Website