

Structure and Dynamics of Deep Neural Networks

A Perspective from Geometry and Physics

Docent Lecture

Jan E. Gerken

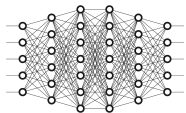


UNIVERSITY OF
GOTHENBURG

WASPI | WALLENBERG AI
AUTONOMOUS SYSTEMS
AND SOFTWARE PROGRAM

Machine learning with neural networks

- Neural networks

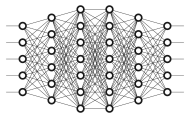


$$\Rightarrow \mathcal{N}_{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}^n, \quad \theta \in \mathbb{R}^N: \text{parameters}$$

$$\text{e.g. } \mathcal{N}_{\theta} : \text{picture} \mapsto P(\text{cat})$$

Machine learning with neural networks

- Neural networks



$$\Rightarrow \mathcal{N}_{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}^n, \quad \theta \in \mathbb{R}^N: \text{parameters}$$

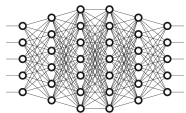
$$\text{e.g. } \mathcal{N}_{\theta} : \text{picture} \mapsto P(\text{cat})$$

- Training data
examples $x \mapsto y$ of target function, e.g.

$$\mathcal{D} = \left\{ \begin{array}{cc} \text{Grumpy Cat} & \mapsto 1.0, \\ \text{Dog with sunglasses} & \mapsto 0.0, \quad \dots \end{array} \right\}$$

Machine learning with neural networks

- Neural networks



$$\Rightarrow \mathcal{N}_{\theta}: \mathbb{R}^m \rightarrow \mathbb{R}^n, \quad \theta \in \mathbb{R}^N: \text{parameters}$$

$$\text{e.g. } \mathcal{N}_{\theta}: \text{picture} \mapsto P(\text{cat})$$

- Training data
examples $x \mapsto y$ of target function, e.g.

$$\mathcal{D} = \left\{ \begin{array}{cc} \text{Grumpy Cat} & \mapsto 1.0, \\ \text{Dog with sunglasses} & \mapsto 0.0, \quad \dots \end{array} \right\}$$

- Optimize θ so that \mathcal{N}_{θ} matches the target function

Neural networks

Neural networks

- Consider a very simple function

$$f_{W,b} : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad f_{W,b}(x) = Wx + b$$

Neural networks

- Consider a very simple function

$$f_{W,b} : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad f_{W,b}(x) = Wx + b$$

- Compose many of these


$$\mathcal{N}_\theta = f_{W^{(L)},b^{(L)}} \circ f_{W^{(L-1)},b^{(L-1)}} \circ \cdots \circ f_{W^{(1)},b^{(1)}}$$

Neural networks

- Consider a very simple function

$$f_{W,b} : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad f_{W,b}(x) = \sigma(Wx + b)$$

$\sigma(x) = \max(0, x)$




- Compose many of these

$$\mathcal{N}_\theta = f_{W^{(L)}, b^{(L)}} \circ f_{W^{(L-1)}, b^{(L-1)}} \circ \cdots \circ f_{W^{(1)}, b^{(1)}}$$

Neural networks

- Consider a very simple function

$$f_{W,b} : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad f_{W,b}(x) = \sigma(Wx + b)$$


$\sigma(x) = \max(0, x)$

- Compose many of these


$$\mathcal{N}_\theta = f_{W^{(L)}, b^{(L)}} \circ f_{W^{(L-1)}, b^{(L-1)}} \circ \cdots \circ f_{W^{(1)}, b^{(1)}}$$

- Optimize all parameters

$$\theta = \{(W^{(\ell)}, b^{(\ell)}) \mid \ell = 1, \dots, L\}$$

Neural networks

- Consider a very simple function

$$f_{W,b} : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad f_{W,b}(x) = \sigma(Wx + b)$$


$\sigma(x) = \max(0, x)$

- Compose many of these

$$\mathcal{N}_\theta = f_{W^{(L)}, b^{(L)}} \circ f_{W^{(L-1)}, b^{(L-1)}} \circ \cdots \circ f_{W^{(1)}, b^{(1)}}$$

- Optimize all parameters

$$\theta = \{(W^{(\ell)}, b^{(\ell)}) \mid \ell = 1, \dots, L\}$$

What can we say about \mathcal{N}_θ mathematically?

Surprising properties of neural networks

Neural networks can approximate any function

Surprising properties of neural networks

Neural networks can approximate any function

Consider a network $\mathcal{N}_\theta : \mathbb{R} \rightarrow \mathbb{R}$ with one hidden layer of width 4

$$f_i^{(1)}(x) = \sigma\left(W_i^{(1)}x + b_i^{(1)}\right), \quad i = 1, 2, 3, 4$$

$$\mathcal{N}_\theta(x) = \sum_{i=1}^4 W_i^{(2)} f_i^{(1)}(x) + b^{(2)}$$

Surprising properties of neural networks

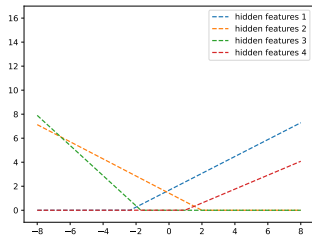
Neural networks can approximate any function

Consider a network $\mathcal{N}_\theta : \mathbb{R} \rightarrow \mathbb{R}$ with one hidden layer of width 4

$$f_i^{(1)}(x) = \sigma\left(W_i^{(1)}x + b_i^{(1)}\right), \quad i = 1, 2, 3, 4$$

$$\mathcal{N}_\theta(x) = \sum_{i=1}^4 W_i^{(2)} f_i^{(1)}(x) + b^{(2)}$$

For random parameters, we get



Surprising properties of neural networks

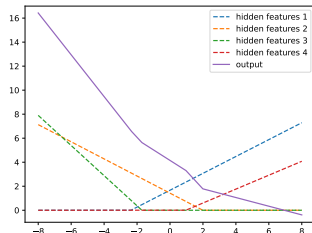
Neural networks can approximate any function

Consider a network $\mathcal{N}_\theta : \mathbb{R} \rightarrow \mathbb{R}$ with one hidden layer of width 4

$$f_i^{(1)}(x) = \sigma\left(W_i^{(1)}x + b_i^{(1)}\right), \quad i = 1, 2, 3, 4$$

$$\mathcal{N}_\theta(x) = \sum_{i=1}^4 W_i^{(2)} f_i^{(1)}(x) + b^{(2)}$$

For random parameters, we get



Surprising properties of neural networks

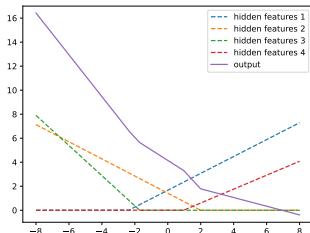
Neural networks can approximate any function

Consider a network $\mathcal{N}_\theta : \mathbb{R} \rightarrow \mathbb{R}$ with one hidden layer of width 4

$$f_i^{(1)}(x) = \sigma\left(W_i^{(1)}x + b_i^{(1)}\right), \quad i = 1, 2, 3, 4$$

$$\mathcal{N}_\theta(x) = \sum_{i=1}^4 W_i^{(2)} f_i^{(1)}(x) + b^{(2)}$$

For random parameters, we get



number of linear pieces
exponential in depth

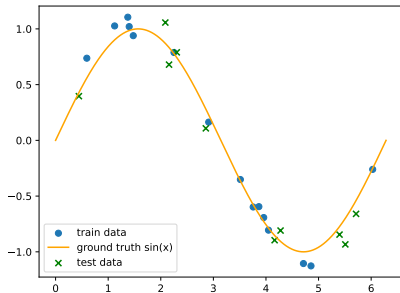
Surprising properties of neural networks

Neural networks work very well!

Surprising properties of neural networks

Neural networks work very well!

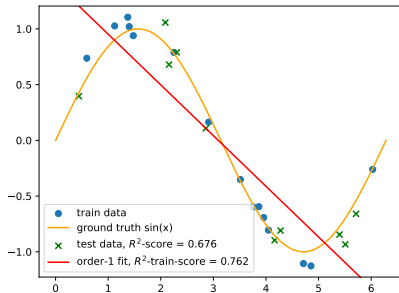
Consider a simple machine-learning problem



Surprising properties of neural networks

Neural networks work very well!

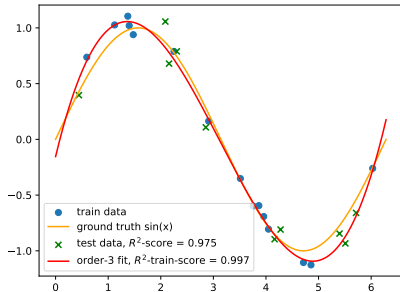
Consider a simple machine-learning problem



Surprising properties of neural networks

Neural networks work very well!

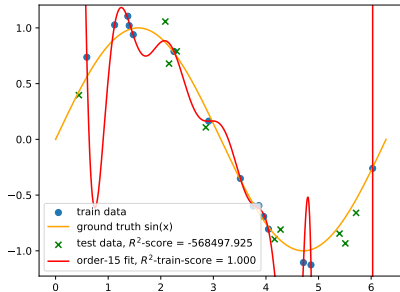
Consider a simple machine-learning problem



Surprising properties of neural networks

Neural networks work very well!

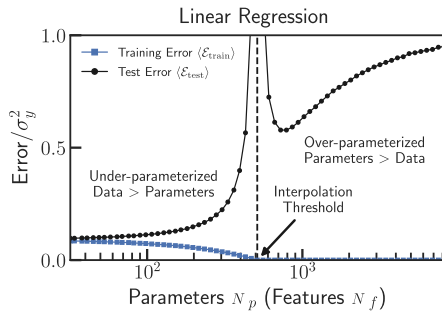
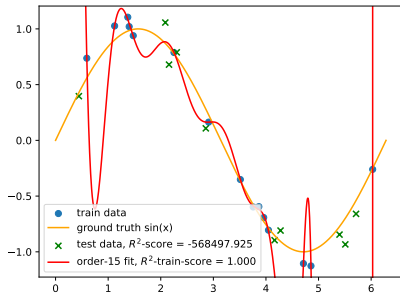
Consider a simple machine-learning problem



Surprising properties of neural networks

Neural networks work very well!

Consider a simple machine-learning problem

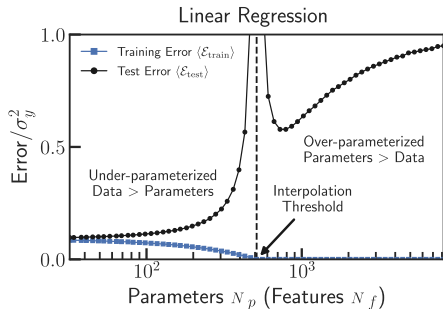


[Rocks, Mehta 2020]

Surprising properties of neural networks

Neural networks work very well!

Neural networks show double decent

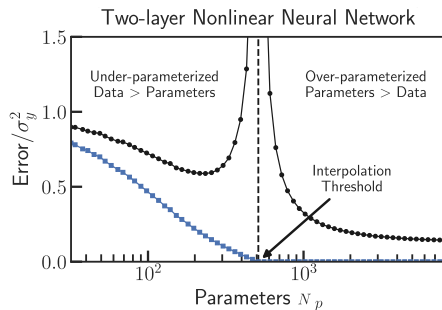
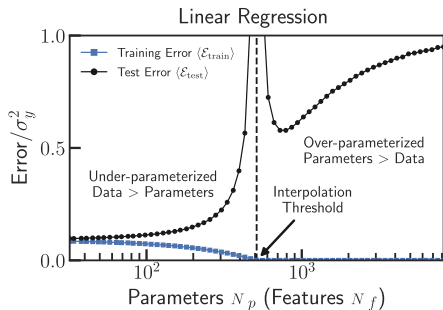


[Rocks, Mehta 2020]

Surprising properties of neural networks

Neural networks work very well!

Neural networks show double decent



[Rocks, Mehta 2020]

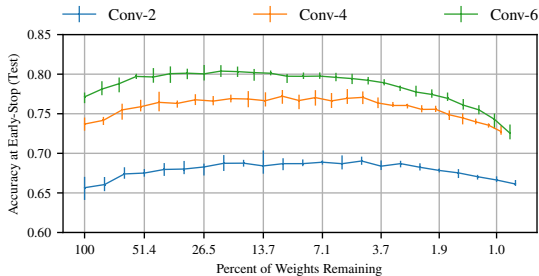
Surprising properties of neural networks

Only a small subnetwork matters

Surprising properties of neural networks

Only a small subnetwork matters

Possible to find tiny subnetworks which have almost the same performance as the full network



[Frankle, Carbin 2019]

Hyperparameter tuning

Hyperparameter tuning

- We use some algorithm to optimize the parameters θ

Hyperparameter tuning

- We use some algorithm to optimize the parameters θ
- **Hyperparameters** are not optimized
 - Parameters of the optimization algorithm
 - Sizes of intermediate vector spaces
 - How many layers, which non-linear function...

Hyperparameter tuning

- We use some algorithm to optimize the parameters θ
- **Hyperparameters** are not optimized
 - Parameters of the optimization algorithm
 - Sizes of intermediate vector spaces
 - How many layers, which non-linear function...
- In practice, huge amount of compute spent on trial and error

Geometry and Physics in Neural Networks

Geometry and Physics in Neural Networks

- Can we exploit the geometry of the data distribution?
- Can we exploit symmetries of the target function?
- Can we use methods from theoretical physics to understand the learning process?

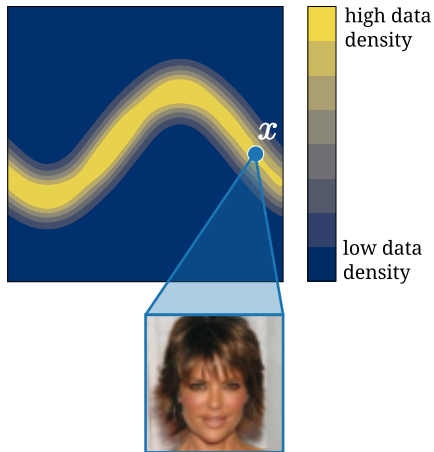
Geometry and Physics in Neural Networks

Geometric Deep Learning

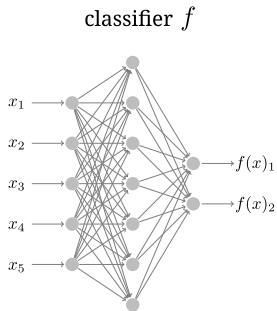
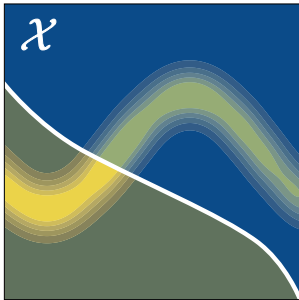
- Can we exploit the geometry of the data distribution?
 - Can we exploit symmetries of the target function?
-
- Can we use methods from theoretical physics to understand the learning process?

Geometry of the Data Distribution

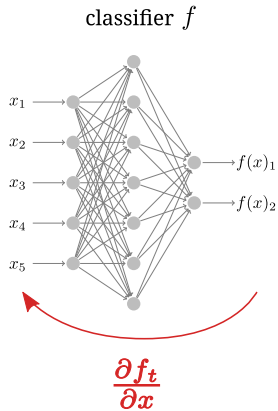
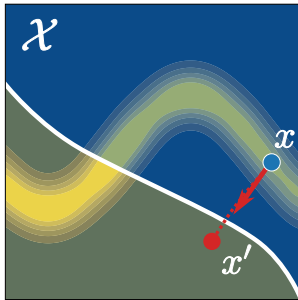
Manifold Hypothesis



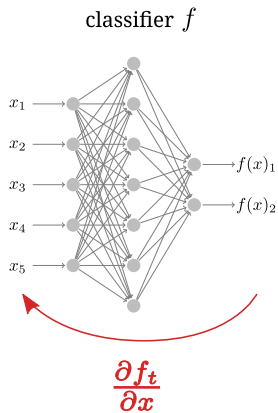
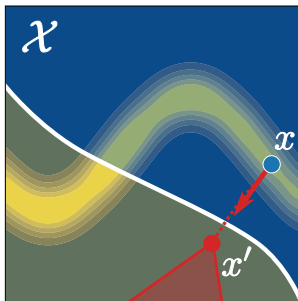
Adversarial Examples



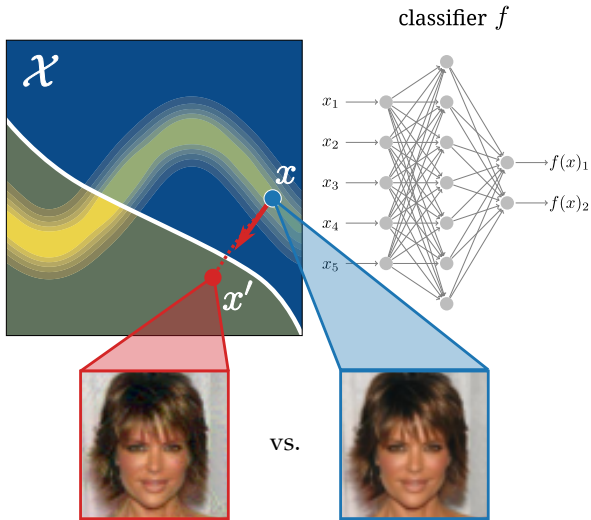
Adversarial Examples



Adversarial Examples



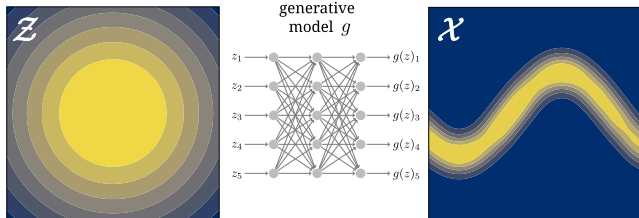
Adversarial Examples



Geometry of the Data Manifold

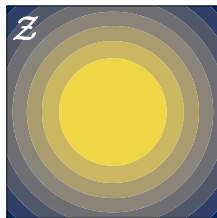
Geometry of the Data Manifold

Use model to learn diffeomorphism to normal distribution

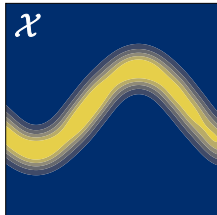
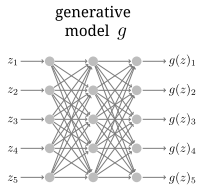


Geometry of the Data Manifold

Use model to learn diffeomorphism to normal distribution



Euclidean metric δ

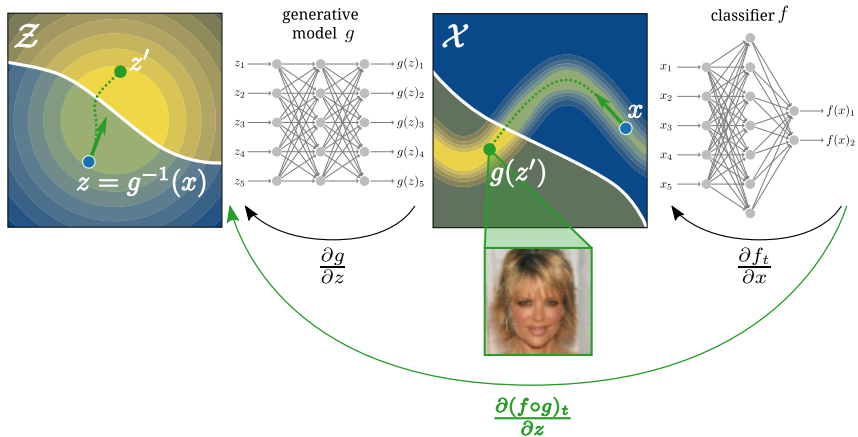


induced metric γ

Optimize along the Data Manifold

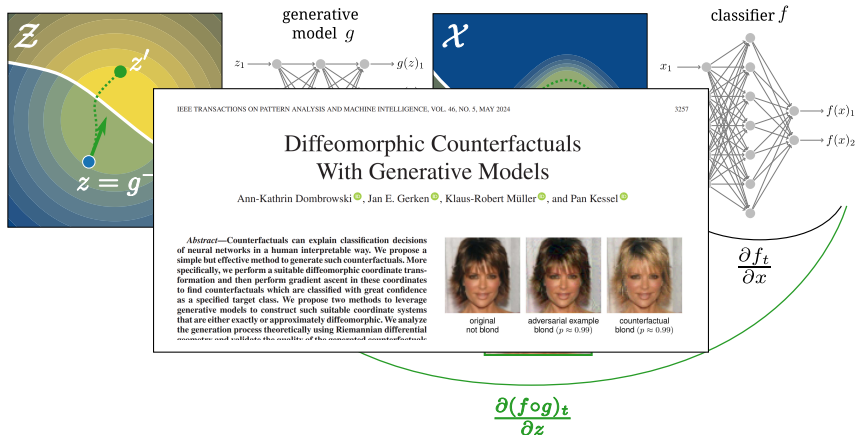
Optimize along the Data Manifold

Follow the gradient in the normalized coordinates



Optimize along the Data Manifold

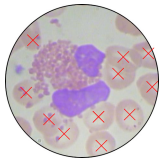
Follow the gradient in the normalized coordinates



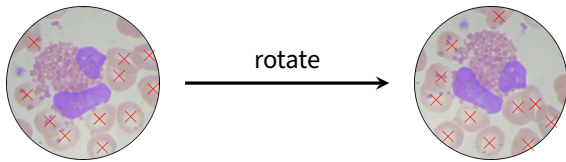
Symmetries

Symmetry of the Data

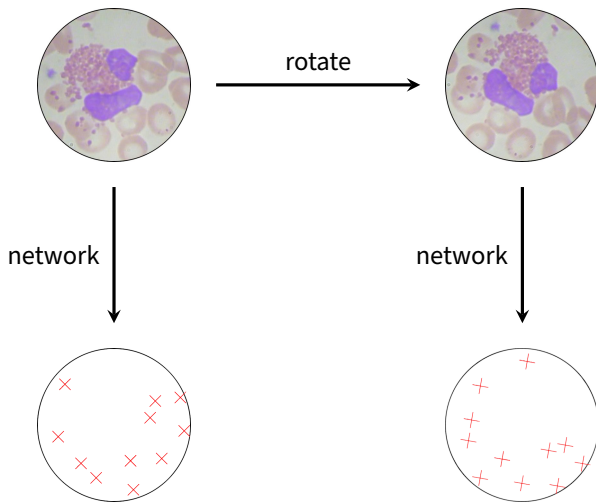
Symmetry of the Data



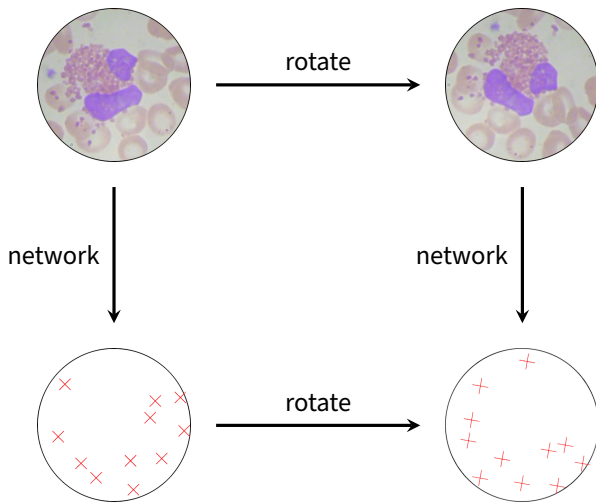
Symmetry of the Data



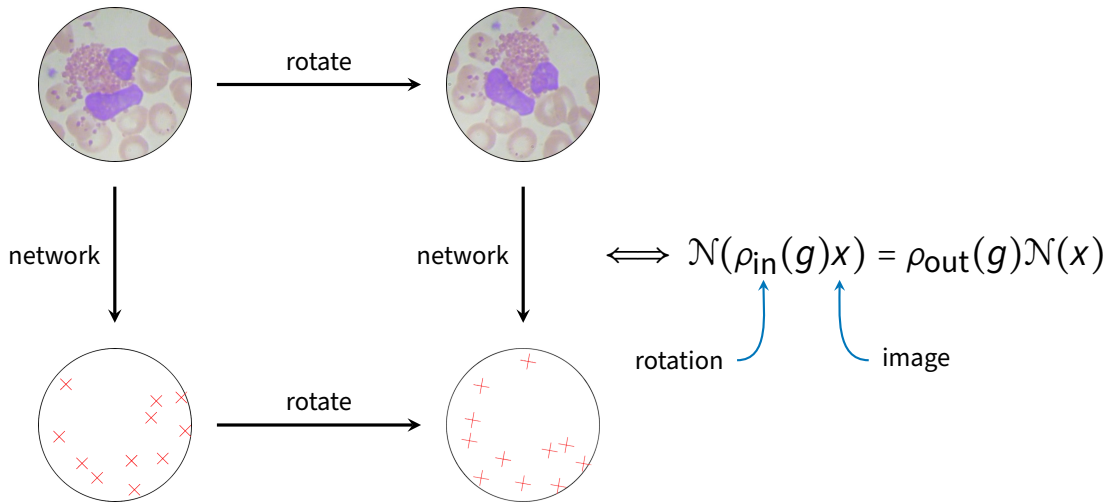
Symmetry of the Data



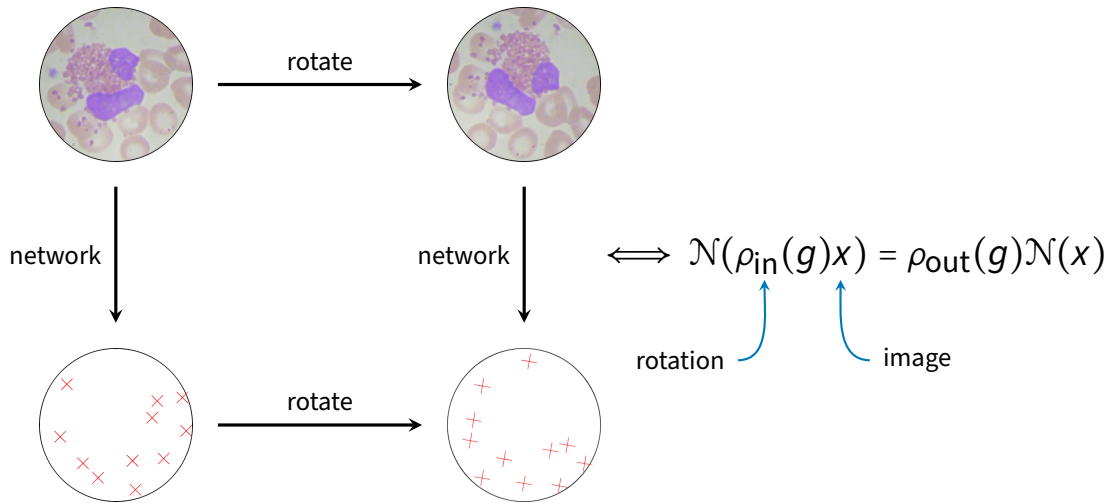
Symmetry of the Data



Symmetry of the Data



Symmetry of the Data \Rightarrow Equivariant Networks



Equivariant neural networks

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen

University of Amsterdam

Max Welling

University of Amsterdam

University of California Irvine

Canadian Institute for Advanced Research

T.S.COHEN@UVA.NL

M.WELLING@UVA.NL

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
University of Amsterdam
Max Welling
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

T.S.COHEN@UVA.NL

M.WELLING@UVA.NL

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 Tai¹ Peter Bailis¹ Gregory Valiant¹

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 & Veličković, 2017; Cohen et al., 2018) has explored new CNN architectures that are *invariant to certain modifiable to particular transforms*.

Equivariant neural networks

Group Equivariant Convolutional Networks

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

T.S.COHN@UVA.NL

M.WELLING@UVA.NL

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 Tai¹ Peter Bailis¹ Gregory Valiant¹

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 & Veličković, 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 Bucci,^{1,5} Michael Ragone,^{1,6} Patrick J. Coles,¹ Frédéric Sauvage,¹ Martin Larooca,^{1,7} and M. Cerezo¹

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
³Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
⁴Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
⁵Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy
⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA
⁷Center 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 symmetries. In this work, we present these ideas in the quantum realm by

Equivariant neural networks

Group Equivariant Convolutional Networks

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

T.S.COHEM@UVA.NL

M.WELLING@UVA.NL

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 Tai¹ Peter Bailis¹ Gregory Valiant¹

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 & Veličković, 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 Borella,^{1,5} Michael Ragone,^{1,6} Patrick J. Cules,¹ Frédéric Sauvage,¹ Martin Laroche,^{1,7} and M. Cerezo¹

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
³Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
⁴Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
⁵Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy
⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA
⁷Center 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 symmetries. In this work, we present these ideas in the quantum realm by

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,b,c,1} Qi Meng^b Jue Zhang^b Hulin Qu^c Congqiao Li² Sitian Qian^d Weitao Du^a Zhi-Ming Ma^a Tie-Yan Liu^b

^aAcademy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China

^bMicrosoft Research Asia, Danling Street, Beijing 100080, China

^cCERN, EP Department, CH-1211 Geneva 23, Switzerland

^dSchool of Physics, Peking University, Chenmefu Road, Beizina 100871, China

Equivariant neural networks

Group Equivariant Convolutional Networks

Taco S. Cohen
University of Amsterdam

T.S.COHEM@UVA.NL

Max Welling
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

M.WELLING@UVA.NL

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

Kal Sheng Tai¹ Peter Bailis¹ Gregory Valiant¹

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 & Veličković, 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. Cules,¹ Frédéric Sauvage,¹ Martin Larooca,^{1,7} and M. Cerezo¹

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
³Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
⁴Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
⁵Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy
⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA
⁷Center 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 symmetries. In this work, we present these ideas in the quantum realm by

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,b,1} Qi Meng^b Jue Zhang^b Hulin Qu^c Congqiao Li² Sitian Qian^d Weitao Du^a Zhi-Ming Ma^a Tie-Yan Liu^b

^aAcademy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China

^bMicrosoft Research Asia, Danling Street, Beijing 100080, China

^cCERN, EP Department, CH-1211 Geneva 23, Switzerland

^dSchool of Physics, Peking University, Chenhefa Road, Beizhen 100871, China

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

Simon Batzner^{a,1} Albert Musselian,¹ Litxin Sun,¹ Mario Geiger,² Jonathan P. Mailon,³ Mordechai Korbath,² Nicola Molinari,¹ Tess E. Smidt,^{4,5} and Boris Kozinsky^{a,1,2}

¹John A. Pashon School of Engineering and Applied Sciences,

Harvard University, Cambridge, MA 02138, USA

²École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

³Robert Bosch Research and Technology Center, Cambridge, MA 02139, USA

⁴Computational Research Division and Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁵Massachusetts 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 informative-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
University of Amsterdam

T.S.COHEN@UVA.NL

Max Welling
University of Amsterdam
University of California Irvine
Canadian Institute for Advanced Research

M.WELLING@UVA.NL

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

Kal Sheng Tai¹ Peter Bailis¹ Gregory Valiant¹

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 & Veličković, 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. Cules,¹ Frédéric Sauvage,¹ Martin Lucocca,^{1,7} and M. Cerezo¹

¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
³Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
⁴Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
⁵Dipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy
⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA
⁷Center 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 symmetries. In this work, we present these ideas in the quantum realm by

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,b,1} Qi Meng^a Jue Zhang^a Hulin Qu^a Gonggao Li^a Sitian Qian^d Weitao Du^a Zhi-Ming Ma^a Tie-Yan Liu^b

^aAcademy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhongguancun East Road, Beijing 100190, China

^bMicrosoft Research Asia, Danling Street, Beijing 100080, China

^cCERN, EP Department, CH-1211 Geneva 23, Switzerland

^dSchool of Physics, Peking University, Chenehe Road, Beihua 100871, China

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

Simon Batzner^{a,1} Albert Muscelian,¹ Lixin Sun,¹ Mario Geiger,² Jonathan P. Mailon,³ Mordechai Korbath,² Nicola Molinari,¹ Tess E. Smidt,^{4,5} and Boris Kozinsky^{a,1,2}

¹John A. Paulson School of Engineering and Applied Sciences,

Harvard University, Cambridge, MA 02138, USA

²École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

³Robert Bosch Research and Technology Center, Cambridge, MA 02139, USA

⁴Computational Research Division and Center for Advanced Mathematics for Energy Research Applications,

Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁵Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA 02139, 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^a
Department of Applied Physics
Stanford University
seismann@stanford.edu

Raphael J.L. Townsend^b
Department of Computer Science
Stanford University
raphael@cs.stanford.edu

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

Milind Jagota
Department of Electrical Engineering
Stanford University
mijagota@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

What is the general mathematical formulation of equivariant neural networks?

Geometric deep learning and equivariant neural networks



Jan E. Gerken^{1,2,3} · Jimmy Aronsson¹ · Oscar Carlsson¹ · Hampus Linander⁴ · Fredrik Ohlsson⁵ · Christoffer Petersson^{1,6} · Daniel Persson¹

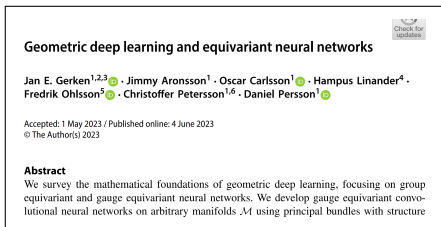
Accepted: 1 May 2023 / Published online: 4 June 2023
© The Author(s) 2023

Abstract

We survey the mathematical foundations of geometric deep learning, focusing on group equivariant and gauge equivariant neural networks. We develop gauge equivariant convolutional neural networks on arbitrary manifolds \mathcal{M} using principal bundles with structure

Equivariant Neural Networks

What is the general mathematical formulation of equivariant neural networks?



Fiber bundles

Representation theory

Spherical harmonics,
Wigner matrices...

Equivariant Neural Networks

Use gauge equivariant network to learn topological invariants

Learning Chern Numbers of Multiband Topological Insulators with Gauge Equivariant Neural Networks

Longde Huang

Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg
longde@chalmers.se

Oleksandr Balabanov

Department of Physics
Stockholm University

Hampus Linander

VERSES AI
Los Angeles, CA, USA
hampus.linander@verses.ai

Mats Granath

Department of Physics
University of Gothenburg
mats.granath@physics.gu.se

Daniel Persson

Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg
daniel.persson@chalmers.se

Jan E. Gerken

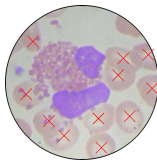
Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg
gerken@chalmers.se

Abstract

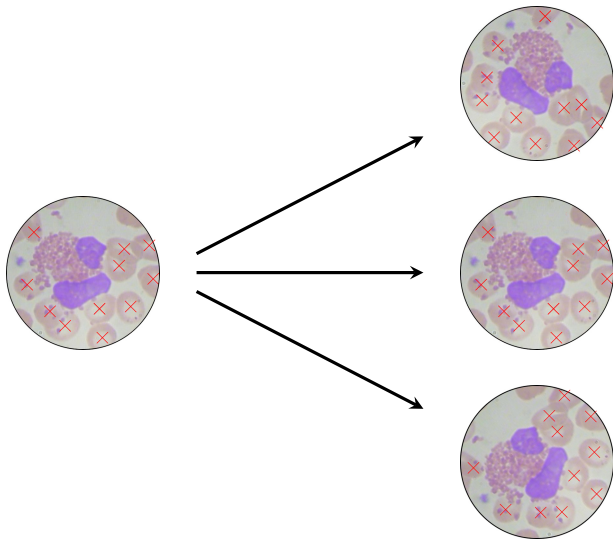
Equivariant network architectures are a well-established tool for predicting invariant or equivariant quantities. However, almost all learning problems considered in this context feature a global symmetry, i.e. each point of the underlying space is transformed with the same group element, as opposed to a local *gauge* symmetry, where each point is transformed with a different group element, exponentially enlarging the size of the symmetry group. We use gauge equivariant networks to

Learning symmetries

Learning symmetries



Learning symmetries



Impose symmetries or learn them?

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold 3

<https://doi.org/10.1038/s41586-024-03847-y>

Received: 18 December 2022

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

 Check for updates

Josh Abramson^{1,2}, Anna Adler^{1,2}, Jack Dongus^{1,2}, Richard Evans^{1,2}, Tim Green^{1,2}, Alexander Pritzel^{1,2}, Olaf Ronneberger^{1,2}, Lindsay Wilkerson^{1,2}, Andrew J. Ballard¹, Joshua Bonbrink¹, Sebastian W. Bowers^{1,2}, David A. Evans¹, Chao-Chun Hung¹, Michael O'Neill¹, David Reiss¹, Kathryn Tarnopolska^{1,2}, Zachary Wu¹, Abhishek Baghel^{1,2}, Evin Arvaniti¹, Charles Beattie¹, Ottavia Bertolli¹, Alan Bridgland¹, Alexey Chumakov¹, Miles Congreve¹, Alexander I. Cowen-Rivers¹, Andrew Cowie¹, Michael Figurnov¹, Fabian D. Fuchs¹, Harsh Gaudin¹, Rohan Jain¹, Yusuf A. Khan^{1,2}, Caroline M. K. Lee¹, Kuba Peltin¹, Anna Potapenko¹, Pascal Savay¹, Sukhdeep Singh¹, Adrian Stecula¹, Ashok Thillaisundaram¹, Catherine Tong¹, Sergei Yelensky¹, Ellen D. Zhang^{1,2}, Michal Zaidman¹, Augustin Eden¹, Victor Bapst^{1,2}, Pauliusius Kubilius^{1,2}, Max Jaderberg^{1,2}, Demis Hassabis^{1,2,3} & John M. Jumper^{1,2}

The introduction of AlphaFold 2¹ has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein modelling and design^{2–5}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantially improved accuracy

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold 3

<https://doi.org/10.1038/s41586-024-03867-w>

Received: 18 December 2022

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

 Check for updates

Josh Abramson^{1,2}, John Jumper^{1,2}, Richard Evans^{1,2}, Tim Green^{1,2}, Alexander Pritzel^{1,2}, Amy Williams^{1,2}, Andrew J. Ballard¹, Joshua Bernstein¹, David A. Evans¹, Chao-Chun Hung¹, Michael O'Neil¹, Tom H. P. Tang¹, Tamas Szepesvari¹, Zachary Wu¹, Abhishek Sanyal^{1,3,4}, Ermiyu Anyanwu¹, David Bertolli¹, Alan Bridgland¹, Alexey Chumakov¹, Allen Ewens¹, Andrew Goss¹, Michael Figurnov¹, Filippos Flokas^{1,5,6}, Yusuf A. Khan^{1,7}, Caroline M. R. Lee¹, Iain L. McQuibban¹, Pascal Savy¹, Subhojit Singh¹, Adrian Stecula¹, Catherine Tong¹, Sergei Yelensky¹, Ellen D. Zhang^{1,8}, Martin Zidek¹, Victor Zepher^{1,9}, Paulius Velička^{1,9}, Max Jaderberg^{10,11}, & John M. Jumper^{1,12}

The introduction of AlphaFold 2¹ has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein design and drug discovery^{2–5}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantially improved accuracy

No Equivariance!

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold3

<https://doi.org/10.1038/s41586-024-03847-4>
Received: 18 December 2023
Accepted: 28 April 2024
Published online: 8 May 2024
Open access
Check for updates

Josh Abramson^{1,2}, John Jumper^{1,2}, Richard Evans^{1,2}, Tim Green^{1,2}, Alexander Pritzel^{1,2}, Amy Williams^{1,2}, Andrew J. Ballard¹, Joshua Bernstein¹, David A. Evans¹, Chao-Chun Huang¹, Michael O'Neil¹, Daniel Tunyasuvunakool¹, Zachary Wu¹, Abhishek Bhargava¹, Emma Anyan¹, David Bertolli¹, Alan Bridgland¹, Alessio Chignone¹, Allen Edwards¹, Andrew Ganev¹, Michael Figurnov¹, Filippos Flokas¹, Heliuk Jais¹, Yusuf A. Khan¹, Caroline M. R. Lee¹, Mark Little¹, Pawan Kumar¹, Pauline R. Smith¹, Adrian Stecula¹, Catherine Tong¹, Sergei Yelensky¹, Ellen D. Zhang¹, Martin Zeller¹, Victor Zepeda¹, Paulius Velička¹, Max Jaderberg^{1,3}, & John M. Jumper^{1,2}

The introduction of AlphaFold 2 has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein design and discovery^{1–4}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantial improved accuracy

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

Aditi S. Krishnapriyan
UC Berkeley, LBNL
aditik1@berkeley.edu

Abstract

Scaling has been a critical factor in improving model performance and generalization across various fields of machine learning. It involves how a model's performance changes with increases in model size or input data, as well as how efficiently computational resources are utilized to support this growth. Despite successes in scaling other types of machine learning models, the study of scaling in Neural Network Interatomic Potentials (NNIPs) remains limited. NNIPs act as

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold3

<https://doi.org/10.1038/s41586-024-03847-9>

Received: 18 December 2022

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

 Check for updates

Josh Abramson^{1,2}, John Jumper^{1,2}, David Evans^{1,2}, Tim Green^{1,2}, Alexander Pritzel^{1,2}, Amy Williams^{1,2}, Andrew J. Ballard¹, Joshua Bernstein¹, David M. A. Evans¹, Chao-Chun Huang¹, Michael O'Byrne¹, Tom Tarran¹, Zachary Wu¹, Abella Zengyel^{1,3}, Evin Ayo¹, David Bertolli¹, Alan Bridgland¹, Alessio Chianteramo¹, Allen Ewens¹, Andrew Gossie¹, Michael Fagan¹, Fajun Gao¹, Richard Jiao¹, Yousaf A. Khan¹, Caroline M. K. Lee¹, Kyohei Ma¹, Pascal Savy¹, Sukhdeep Singh¹, Adrian Stecula¹, Catherine Tong¹, Sergei Yalovoy¹, Ellen D. Zhang¹, Martin Zidek¹, Victor Zepher¹, Pauline Zito¹, Max Jaderberg^{1,2}, & John M. Jumper^{1,2}

Structure prediction of AlphaFold 2 has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein design and drug discovery^{1–5}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantial accuracy

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

Aditi S. Krishnagiriyan
UC Berkeley, LBNL
aditik1@berkeley.edu

Abstract

Scaling has been a critical factor in improving model performance and generalization across various fields of machine learning. It involves how a model's performance changes with increases in model size or input data, as well as how efficiently computational resources are utilized to support this growth. Despite successes in scaling other types of machine learning models, the study of scaling in Neural Network Interatomic Potentials (NNIPs) remains limited. NNIPs act as

Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yiyang Wang¹, Ahmed A. Elhag^{1,2}, Navdeep Jaitly¹, Joshua M. Susskind¹, Miguel Ángel Bautista¹

Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state-of-the-art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (e.g. modeling torsional angles) we are able to radically simplify structure prediction and enable us to predict conformers that

is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Aschrod & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules. Systematic methods like OMFGA (Hawkins et al., 2010)

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold3

<https://doi.org/10.1038/s41586-024-0387-9>

Received: 18 December 2023

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

Check for updates

Josh Abramson^{1,2}, Jonas Bieganski^{1,2}, Alexander Pritzel^{1,2}, David A. Evans^{1,2}, Tim O’Neil^{1,2}, Amy Williams^{1,2}, Andrew J. Ballard^{1,2}, Joshua Bonhoeffer^{1,2}, David A. Evans^{1,2}, Chao-Chun Huang^{1,2}, Michael J. J. G. T. Turner^{1,2}, Zachary Wu^{1,2}, Abhishek Sanyal^{1,2}, Eric A. Bevilacqua^{1,2}, Alex Bridgland^{1,2}, Alessio Caporaso^{1,2}, Adam Evans^{1,2}, Andrew Ganev^{1,2}, Michael H. Fagan^{1,2}, Miles H. Fagan^{1,2}, Patrick Hain^{1,2}, Yousang A. Kim^{1,2}, Caroline M. Lee^{1,2}, Adam P. Koh^{1,2}, Pascal Savary^{1,2}, Subhojit Singh^{1,2}, Adrian Stecula^{1,2}, John D. Swinney^{1,2}, Victor B. Vogel^{1,2}, Paulius Velička^{1,2}, Max Jaderberg^{1,2}, & John M. Jumper^{1,2}

Structure prediction of AlphaFold 2 has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein design and drug discovery^{1–5}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantial improved accuracy

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

Aditi S. Krishnamurthy
UC Berkeley, LBNL
aditik1@berkeley.edu

Abstract

Scaling has been a critical factor in improving model performance and generalization across various fields of machine learning. It involves how a model’s performance changes with increases in model size or input data, as well as how efficiently computational resources are utilized to support this growth. Despite successes in scaling other types of machine learning models, the study of scaling in Neural Network Interatomic Potentials (NNIPs) remains limited. NNIPs act as

Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yiyang Wang¹, Ahmed A. Elhag^{1,2}, Navdeep Jaitly¹, Joshua M. Susskind¹, Miguel Ángel Bautista¹

Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state-of-the-art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (e.g. modeling torsional angles) we are able to radically simplify structure prediction and make it suitable for scale.

is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Aschrod & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules. Systematic methods like OMEGA (Hawkins et al., 2010)

Probing the effects of broken symmetries in machine learning

Marcel F. Langer¹, Sergey N. Poddanyakov¹, and Michele Ceriotti¹

Laboratory of Computational Science and Modeling and National Centre for Computational Design and Discovery of Novel Materials MARVEL, Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

* Author to whom any correspondence should be addressed.

E-mail: michele.ceriotti@epfl.ch

Keywords: machine learning, symmetry-constrained models, atomistic modeling, molecular simulation

Supplementary material for this article is available online

Abstract

Symmetry is one of the most central concepts in physics, and it is no surprise that it has also been widely adopted as an inductive bias for machine-learning models applied to the physical sciences. This is especially true for models targeting the properties of matter at the atomic scale. Both established and state-of-the-art approaches, with almost no exceptions, are built to be exactly equivariant to translations, permutations, and rotations of the atoms. Incorporating symmetries—rotations in particular—constrains the model design space and implies more complicated architectures that are often also computationally demanding. There are indications

Article

https://doi.org/10.1008/b9789810238103_034

Received: 15 December 2023

Accepted: 26 April 2014

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 111–117

Downloaded from <http://ajphaphysocpharm.sagepub.com>

 Check for updates

No Equivalence!

Introduction of AlphaFold 2¹ has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein modelling and design²⁻⁴. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold 3 model demonstrates substantially improved accuracy

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

Aditi S. Krishnapriyan
UC Berkeley, LBNL
aditik1@berkeley.edu

Abstract

Scaling has been a critical factor in improving model performance and generalization across various fields of machine learning. It involves how a model's performance changes with increases in model size or input data, as well as how efficiently computational resources are utilized to support this growth. Despite successes in scaling other types of machine learning models, the study of scaling in Neural Network Interatomic Potentials (NNIPs) remains limited. NNIPs act as

Two for One: Diffusion Models and Force Fields for Coarse-Grained Molecular Dynamics

Marloes Arts,^{a,†,‡,§} Victor García Satorras,^{a,*} Chin-Wei Huang,[†] Daniel Züchner,[§] Marco Federici,^{†,||} Cecilia Clementi,^{§,‡} Frank Noé,[§] Robert Pinsler,[§] and Rianne van den Berg[†]

‡University of Copenhagen, Department of Computer Science, Universitetsparken 1,
Copenhagen, 2100, Denmark.

¶AI4Science, Microsoft Research, Evert van de Beekstraat 354, Amsterdam, 1118 CZ, The Netherlands.

§Al4Science, Microsoft Research, Karl-Liebknecht-Strasse 32, Berlin, 10178, Germany.
 University of Amsterdam, Informatics Institute, Science Park 904, Amsterdam, 1098 XH

The Netherlands.
 Freie Universität Berlin, Department of Physics, Arnimallee 12, Berlin, 14195, Germany

#AI4Science, Microsoft Research, 21 Station Road, Cambridge, CB1 2FB, United Kingdom.

© Equal contribution

* E-mail: ma@di.ku.dk; victorzar@microsoft.com

Abstract

Coarse-grained (CG) molecular dynamics enables the study of biological processes at temporal and spatial scales that would be intractable at an atomistic resolution. However, accurately learning a CG force field remains a challenge. In this work, we leverage connections between score-based generative models, force fields and molecular

Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yuryang Wang¹ Ahmed A. Elhag^{1,2} Navdeep Jaitly¹ Joshua M. Susskind¹ Miguel Ángel Bautista

Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state of the art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (e.g. modeling torsional angles), we are able to radically simplify structure generation, and enable it to scale up to molecules

Is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Axelrod & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules.

Systematic methods, like OMEGA (Hyskins et al., 2010)

Probing the effects of broken symmetries in machine learning

Marcel F Langer , Sergey N Pozdnyakov and Michele Ceriotti*

Laboratory of Computational Science and Modeling and National Centre for Computational Design and Discovery of Novel Materials
MARVEL, Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

* Author to whom any correspondence should be addressed.

E-mail: nichole.gerlatti@epfl.ch

Keywords: machine learning, symmetry-constrained models, atomistic modeling, molecular simulation

Supplementary material for this article is available [online](#).

Abstract

Symmetry is one of the most central concepts in physics, and it is no surprise that it has also been widely adopted as an inductive bias for machine-learning models applied to the physical sciences. This is especially true for models targeting the properties of matter at the atomic scale. Both established and state-of-the-art approaches, with almost no exceptions, are built to be exactly equivariant to translations, permutations, and rotations of the atoms. Incorporating symmetries—rotations in particular—constrains the model design space and implies more complicated architectures that are often also computationally demanding. There are indications

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold3

<https://doi.org/10.1038/s41586-024-03847-w>

Received: 18 December 2023

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

Check for updates

Josh Abramson^{1,2}, Alexander Pitter^{1,2}, Richard Evans^{1,2}, Tim Ovchinnikov^{1,2}, Alexander Pritikin^{1,2}, Amy Wilkins^{1,2}, Andrew J. Ballard^{1,2}, Joshua Bonhoeffer^{1,2}, David A. Evans^{1,2}, Chao-Chun Huang^{1,2}, Michael J. Owen^{1,2}, Tatyana Potapova^{1,2}, Zachary Wu^{1,2}, Abhishek Baghel^{1,2}, Evan Wang^{1,2}, David Berntz^{1,2}, Alex Bridgland^{1,2}, Alessio Caporaso^{1,2}, Allen Rivers^{1,2}, Andrew Cowie^{1,2}, Michael Figarov^{1,2}, James F. Hester^{1,2}, Patrick Jain^{1,2}, Yousaf A. Khan^{1,2}, Caroline M. M. Lee^{1,2}, Akhil P. Patel^{1,2}, Pascal Savary^{1,2}, Subhojit Singh^{1,2}, Adrian Stecula^{1,2}, Catherine Tong^{1,2}, Sergei Yalovsky^{1,2}, Ellen D. Zhang^{1,2}, Justin Drenth^{1,2}, Victor Bapst^{1,2}, Pauline Kish^{1,2}, Max Jochberg^{1,2}, & John M. Jumper^{1,2}

Structure prediction of AlphaFold 2 has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein engineering and design^{1,2}. Here we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture that is capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues. The new AlphaFold model demonstrates substantial accuracy

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

Aditi S. Krishnamurthy
UC Berkeley, LBNL
aditik@berkeley.edu

Abstract

Scaling has been a critical factor in improving model performance and generalization across various fields of machine learning. It involves how a model's performance changes with increases in model size or input data, as well as how efficiently computational resources are utilized to support this growth. Despite successes in scaling other types of machine learning models, the study of scaling in Neural Network Interatomic Potentials (NNIPs) remains limited. NNIPs act as

Two for One: Diffusion Models and Force Fields for Coarse-Grained Molecular Dynamics

Marloes Arts,^{1,2,3,4} Victor Garcia Satorras,^{1,4,5} Chin-Wei Huang,¹ Daniel Zügner,¹ Marco Federici,^{1,2} Cecilia Clementi,^{1,2} Frank Noé,¹ Robert Pinsler,⁶ and Rianne van den Berg¹

¹Work done during an internship at Microsoft Research (Amsterdam).
²University of Copenhagen, Department of Computer Science, Universitetsparken 1, Copenhagen, 2100, Denmark.
³AI4Science, Microsoft Research, Evert van de Beekstraat 354, Amsterdam, 1118 CZ, The Netherlands.
⁴AI4Science, Microsoft Research, Karl-Liebknecht-Straße 32, Berlin, 10178, Germany.
⁵University of Amsterdam, Informatics Institute, Science Park 904, Amsterdam, 1098 XH, The Netherlands.
⁶Freie Universität Berlin, Department of Physics, Arnimallee 19, Berlin, 14195, Germany.
⁷AI4Science, Microsoft Research, 21 Station Road, Cambridge, CB1 3FB, United Kingdom.
⁸Equal contribution.

*E-mail: ma@cs.ku.dk; victorga@microsoft.com

Abstract

Coarse-grained (CG) molecular dynamics enables the study of biological processes at temporal and spatial scales that would be intractable at an atomistic resolution. However, accurately learning a CG force field remains a challenge. In this work, we leverage connections between score-based generative models, force fields and molecular

Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yiyang Wang¹, Ahmed A. Elhag^{1,2}, Navdeep Jaitly¹, Joshua M. Susskind¹, Miguel Ángel Bautista¹

Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state-of-the-art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (e.g. modeling torsional angles) we are able to radically simplify structure prediction and enable us to make new

is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Aschard & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules. Scalability methods like OMFGA (Hawkins et al., 2010)

Probing the effects of broken symmetries in machine learning

Marcel F. Langer¹, Sergey N. Rudnitskiy², and Michele Corietti¹

¹Laboratory of Computational Science and Modeling and National Centre for Computational Design and Discovery of Novel Materials MARVEL, Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
²Author to whom any correspondence should be addressed.

E-mail: michele.corietti@epfl.ch

Keywords: machine learning, symmetry-constrained models, atomistic modeling, molecular simulation
Supplementary material for this article is available online

Abstract

Symmetry is one of the most central concepts in physics, and it is no surprise that it has also been widely adopted as an inductive bias for machine-learning models applied to the physical sciences. This is especially true for models targeting the properties of matter at the atomic scale. Both established and state-of-the-art approaches, with almost no exceptions, are built to be exactly equivariant to translations, permutations, and rotations of the atoms. Incorporating symmetries—rotations in particular—constrains the model design space and implies more complicated architectures that are often also computationally demanding. There are indications

DOES EQUIVARIANCE MATTER AT SCALE?

Johann Brechmer¹, Sönke Behrendts¹, Pin de Haan¹, Theo Cohen¹
Qualcomm AI Research
naill@johannbrechmer.de

ABSTRACT

Given large data sets and sufficient compute, is it beneficial to design neural architectures for the structure and symmetries of each problem? Or is it more efficient to learn them from data? We study empirically how equivariant and non-equivariant networks scale with compute and training samples. Focusing on a benchmark problem of rigid-body interactions and on general-purpose transformer architectures, we perform a series of experiments, varying the model size, training steps, and dataset size. We find evidence for three conclusions. First, equivariance improves data efficiency, but training non-equivariant models with data augmentation can close this gap given sufficient epochs. Second, scaling with compute follows a power law, with equivariant models outperforming non-equivariant ones at each tested compute budget. Finally, the optimal allocation of a compute budget onto model size and training duration differs between equivariant and non-equivariant models.

Impose symmetries or learn them?

Article

Accurate structure prediction of biomolecular interactions with AlphaFold3

<https://doi.org/10.1038/s41586-024-0387-9>

Received: 18 December 2023

Accepted: 28 April 2024

Published online: 8 May 2024

Open access

Check for updates

No Equivalence!

Introduction of AlphaFold² has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in biology and design^{1–5}. Here we describe our AlphaFold³ model with an updated diffusion-based architecture that is capable of predicting the structure of complexes including proteins, nucleic acids, small molecules, ions and residues. The new AlphaFold model demonstrates substantial improve-

Josh Abramson^{1,2}, Jonas Alnæs^{1,2}, Richard Evans^{1,2}, Tim Green^{1,2}, Alexander Pritzel^{1,2}, Amy Allshire^{1,2}, Andrew J. Ballard^{1,2}, David A. Evans¹, Chao-Chun Huang¹, Michael J. Jurek¹, Tamas Szepesvári¹, Zachary Wu¹, Anna Zarka-Segal^{1,2}, Eric A. Anyan¹, David Berntz¹, Alex Bridgland¹, Alessio Cimpanaru^{1,2}, Adam Evans¹, Andrew Gower¹, Michael Hargrave¹, Peter H. Jones¹, Richard Jiao¹, Yousuik Kim^{1,2}, Caroline M. Lee¹, Kyohei Nozaki¹, Pascal Savary¹, Subhojit Sengupta¹, Adrian Stead¹, Catherine Tong¹, Sergei Yelensky¹, Ellen D. Zhang¹, Martin Zidek¹, Victor Zepeda¹, Pauliusius Zukas¹, Max Jader¹ & John M. Jumper^{1,2}

Probing the effects of broken symmetries in machine learning

Marcel F. Langer¹, Sergey N. Poudyalov¹ & Michele Corietti¹

Laboratory of Computational Science and Modeling and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

* Author to whom any correspondence should be addressed.

E-mail: michele.corietti@epfl.ch

Keywords: machine learning, symmetry-constrained models, atomistic modeling, molecular simulation

Supplementary material for this article is available online

Abstract

Symmetry is one of the most central concepts in physics, and it is no surprise that it has also been widely adopted as an inductive bias for machine-learning models applied to the physical sciences. This is especially true for models targeting the properties of matter at the atomic scale. Both established and state-of-the-art approaches, with almost no exceptions, are built to be exactly equivariant to translations, permutations, and rotations of the atoms. Incorporating symmetries—rotations in particular—constrains the model design space and implies more complicated architectures that are often also computationally demanding. There are indications

The Importance of Being Scalable: Improving the Speed and Accuracy of Neural Network Interatomic Potentials Across Chemical Domains

Eric Qu
UC Berkeley
ericqu@berkeley.edu

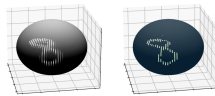
Aditi S. Krishnamoorti
UC Berkeley, LBNL
aditik1@berkeley.edu

Equivariance versus Augmentation for Spherical Images

Jan E. Gerken^{1,2,3}, Oscar Carlsson¹, Hampus Linander¹, Fredrik Ohlsson², Christoffer Petersson^{6,1}, Daniel Persson¹

Abstract

We analyze the role of rotational equivariance in convolutional neural networks (CNNs) applied to spherical images. We compare the performance of the group equivariant networks known as S2CNNs and standard non-equivariant CNNs trained with an increasing amount of data augmentation. The chosen architectures can be consid-



Netherlands.

¹AIQScience, Microsoft Research, Karl-Luecke-Strasse 32, Berlin, 10178, Germany.

²University of Amsterdam, Informatics Institute, Science Park 904, Amsterdam, 1098 XH, The Netherlands.

³Freie Universität Berlin, Department of Physics, Arnimallee 19, Berlin, 14195, Germany.

⁴AIQScience, Microsoft Research, 21 Station Road, Cambridge, CB1 3FB, United Kingdom.

⁵Equal contribution.

* E-mail: ma@ku.dk; victorga@microsoft.com

Abstract

Coarse-grained (CG) molecular dynamics enables the study of biological processes at temporal and spatial scales that would be intractable at an atomistic resolution. However, accurately learning a CG force field remains a challenge. In this work, we leverage connections between atom-based generative models, force fields and molecular

Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yiyang Wang¹, Ahmed A. Elhag^{1,2}, Navdeep Jaitly¹, Joshua M. Susskind¹, Miguel Ángel Bautista¹

Abstract

We present a novel way to predict molecular conformations through a simple formulation that many of the heuristics of prior works pay state of the art results by using the as of scale. By training a diffusion model directly on 3D atomic positions making assumptions about the explicit of molecules (e.g. modeling torsional we are able to radically simplify structure prediction in molecular conformer generation.

is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Ashford & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules.

Scalable methods like OMPGA (Hawkins et al., 2019)

DOES EQUIVARIANCE MATTER AT SCALE?

Johann Behrmer¹, Sönke Behrns¹, Pin de Haan¹, Theo Cohen¹
Qualcomm AI Research
nail1@johannbehrmer.de

ABSTRACT

Given large data sets and sufficient compute, is it beneficial to design neural architectures for the structure and symmetries of each problem? Or is it more efficient to learn them from data? We study empirically how equivariant and non-equivariant networks scale with compute and training samples. Focusing on a benchmark problem of rigid-body interactions and on general-purpose transformer architectures, we perform a series of experiments, varying the model size, training steps, and dataset size. We find evidence for three conclusions. First, equivariance improves data efficiency, but training non-equivariant models with data augmentation can close this gap given sufficient epochs. Second, scaling with compute follows a power law, with equivariant models outperforming non-equivariant ones at each tested compute budget. Finally, the optimal allocation of a compute budget onto model size and training duration differs between equivariant and non-equivariant models.

Practitioners like data augmentation

👍 Easy to implement

👍 No specialized architecture necessary

Practitioners like data augmentation

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

Practitioners like data augmentation

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

Can we understand data augmentation theoretically?

Infinite-Width Networks

Empirical NTK

Training dynamics under continuous gradient descent:

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

learning rate η

loss L

training sample x_i

Empirical NTK

Training dynamics under continuous gradient descent:

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

learning rate η points to the learning rate term in the equation.

loss L points to the loss term in the equation.

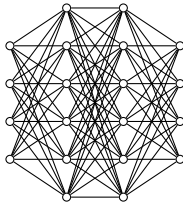
training sample x_i points to the training sample term in the equation.

with the **empirical neural tangent kernel (NTK)**

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

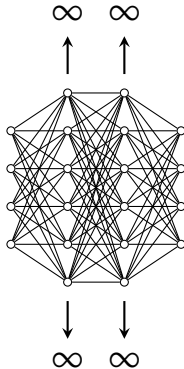
Infinite width limit

[Jacot et al. 2018]



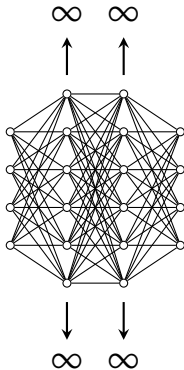
Infinite width limit

[Jacot et al. 2018]



Infinite width limit

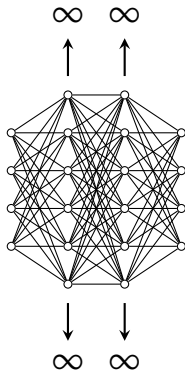
[Jacot et al. 2018]



👍 NTK becomes independent of initialization

Infinite width limit

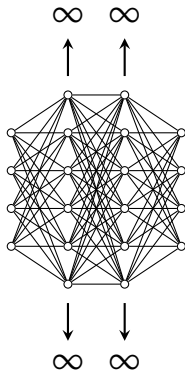
[Jacot et al. 2018]



- 👍 NTK becomes independent of initialization
- 👍 NTK becomes constant in training

Infinite width limit

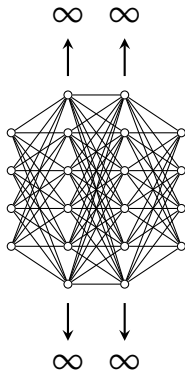
[Jacot et al. 2018]



- 👍 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]


① At infinite width, the mean prediction is given by

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

Mean prediction from NTK

[Jacot et al. 2018]

① At infinite width, the mean prediction is given by



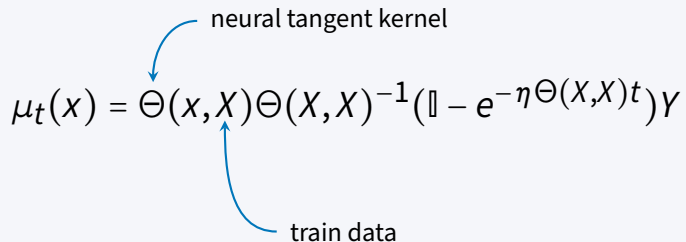
neural tangent kernel

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

Mean prediction from NTK

[Jacot et al. 2018]

① At infinite width, the mean prediction is given by



The diagram shows the equation $\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X) t}) \gamma$. A blue curved arrow points from the text "neural tangent kernel" to the $\Theta(x, X)$ term. Another blue curved arrow points from the text "train data" to the X in the $\Theta(X, X)$ term.

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

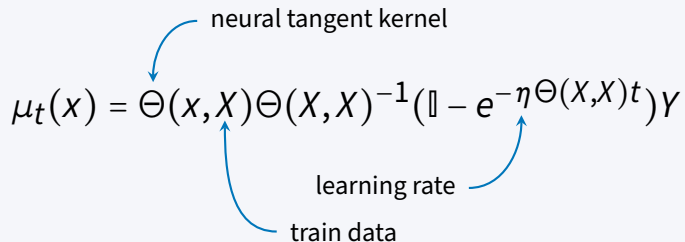
neural tangent kernel

train data

Mean prediction from NTK

[Jacot et al. 2018]

① At infinite width, the mean prediction is given by



The diagram shows the formula $\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X) t}) \gamma$. Three blue arrows point from text labels to parts of the formula: 'neural tangent kernel' points to $\Theta(x, X)$, 'train data' points to X in $\Theta(X, X)$, and 'learning rate' points to η in the exponential term.

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

neural tangent kernel

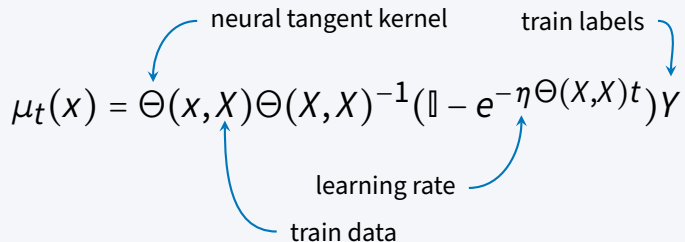
learning rate

train data

Mean prediction from NTK

[Jacot et al. 2018]

① At infinite width, the mean prediction is given by



The diagram shows the formula $\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X) t}) Y$ with four blue arrows pointing to its components: 'neural tangent kernel' points to $\Theta(x, X)$, 'train labels' points to Y , 'learning rate' points to η , and 'train data' points to X in the $\Theta(X, X)$ term.

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

neural tangent kernel

train labels

learning rate

train data

Data augmentation at infinite width

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

Data augmentation at infinite width

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

The diagram illustrates the components of the equation $\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X) t}) \gamma$. Blue arrows originate from the text labels below and point to specific parts of the equation:

- The label "augmented data" (in red) has three arrows pointing to the arguments x , X , and X within the first two Θ functions.
- The label "augmented labels" (in red) has two arrows pointing to the identity matrix \mathbb{I} and the vector γ .

Data augmentation at infinite width

group transformation

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

augmented data

augmented labels

The diagram illustrates the equation for data augmentation at infinite width. A blue arrow points from the text 'group transformation' to the term $\rho(g)$ in the equation. Below the equation, the text 'augmented data' has four blue arrows pointing to the terms $\rho(g)x$, X , X , and X in the expression $\Theta(\rho(g)x, X) \Theta(X, X)^{-1}$. The text 'augmented labels' has two blue arrows pointing to the terms X and X in the expression $\Theta(X, X)^{-1}$. A final blue arrow points from 'augmented labels' to the term Y at the end of the equation.

Data augmentation at infinite width

group transformation for augmented data

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

augmented data augmented labels

The diagram illustrates the equation for data augmentation at infinite width. The equation is $\mu_t(\rho(g)x) = \Theta(\rho(g)x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X)t}) Y$. Annotations include: 'group transformation' pointing to $\rho(g)$; 'for augmented data' pointing to the entire equation; 'augmented data' pointing to $\rho(g)x$; 'augmented labels' pointing to Y ; and a curved arrow labeled 'for augmented data' connecting the input $\rho(g)x$ to the output Y . Additionally, three arrows point from the 'augmented data' label to the arguments of the first Θ function, and two arrows point from the 'augmented labels' label to the arguments of the second Θ function.

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

The diagram illustrates the equation for data augmentation at infinite width. The equation is $\mu_t(\rho(g)x) = \Theta(x, X)\Theta(X, X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X, X)t})\rho(g)Y$. A blue arrow points from the text 'group transformation' to the $\rho(g)$ term. Another blue arrow points from the text 'augmented data' to the x term. A third blue arrow points from the text 'augmented labels' to the Y term. There are also blue arrows pointing from the 'augmented data' label to the $\Theta(x, X)$ and $\Theta(X, X)^{-1}$ terms, and from the 'augmented labels' label to the $\Theta(X, X)$ term in the exponent of the exponential function.

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

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

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

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

Mean prediction

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

Equivariant Ensembles

Ensembles of networks become
exactly equivariant under data augmentation

Emergent Equivariance in Deep Ensembles

Jan E. Gerken^{*1} Pan Kessel^{*2}

Abstract

We show that deep ensembles become equivariant for all inputs and at all training times by simply using full data augmentation. Crucially, equivariance holds off-manifold and for any architecture

emergent: while the prediction of the ensemble is equivariant, the predictions of its members are not. In particular, the ensemble members are not required to have an equivariant architecture.

We rigorously derive this surprising emergent equivariance

Equivariant Ensembles

Ensembles of networks become exactly equivariant under data augmentation

Emergent Equivariance in Deep Ensembles

Jan E. Gerken^{*1} Pan Kessel^{*2}

Abstract

We show that deep ensembles become equivariant for all inputs and at all training times by simply using full data augmentation. Crucially, equivariance holds off-manifold and for any architecture

emergent: while the prediction of the ensemble is equivariant, the predictions of its members are not. In particular, the ensemble members are not required to have an equivariant architecture.

We rigorously derive this surprising emergent equivariance

Also true for finite-width ensembles: [\[Nordenfors, Flinth 2024\]](#)

Equivariant NTKs

Extend NTK framework to equivariant models

Equivariant Neural Tangent Kernels

Philipp Misof^a

Pan Kessel^b

Jan E. Gerken^a

Abstract

Little is known about the training dynamics of equivariant neural networks, in particular how it compares to data augmented training of their non-equivariant counterparts. Recently, neural tangent kernels (NTKs) have emerged as a powerful tool to analytically study the training dynamics of wide neural networks. In this work, we take an important step towards a theoretical understanding of

Could show that an ensemble of augmented MLPs corresponds to an ensemble of GCNNs.

Methods from Physics

Non-Gaussian Corrections from Physics

- 👉 Gaussians are limiting
- 👉 Taylor-expand in $1/\text{width}$
- 👉 Use techniques from quantum field theory

Non-Gaussian Corrections from Physics

- 👎 Gaussians are limiting
- 👍 Taylor-expand in $1/\text{width}$
- 👍 Use techniques from quantum field theory

Neural Networks	Quantum Field Theory
infinite width	no interactions
Gaussian distribution	free fields
finite-width	interactions

Feynman diagrams for Neural Networks

$$\begin{aligned}
 & \frac{1}{n_\ell} \Theta_{12}^{\{1\}(\ell+1)} = \frac{1}{n_{\ell-1}} \Theta^{\{1\}(\ell)} + \frac{1}{n_{\ell-1}} K^{\{1\}(\ell)} + \frac{1}{n_{\ell-1}} V_4^{(\ell)} + \frac{1}{n_{\ell-1}} D_4^{(\ell)} + \frac{1}{n_{\ell-1}} F_4^{(\ell)} \\
 & \text{Diagram 1: } \text{Diagram 2: } \text{Diagram 3: } \text{Diagram 4: } \text{Diagram 5: }
 \end{aligned}$$

Finite-Width Neural Tangent Kernels from Feynman Diagrams

Max Guillen^{*a}

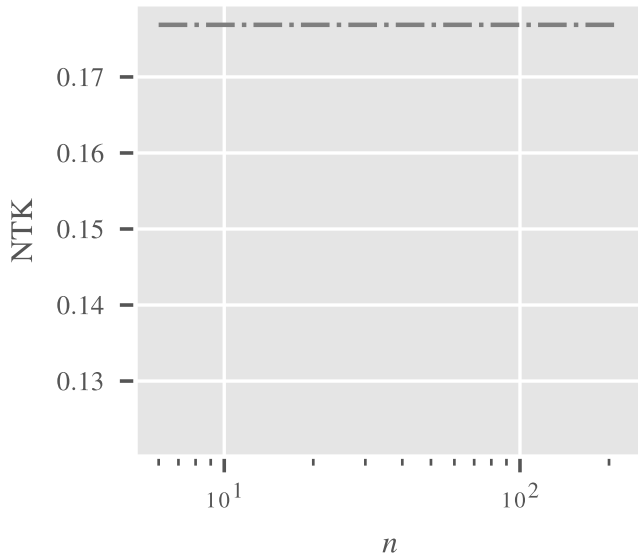
Philipp Misof^{*a}

Jan E. Gerken^a

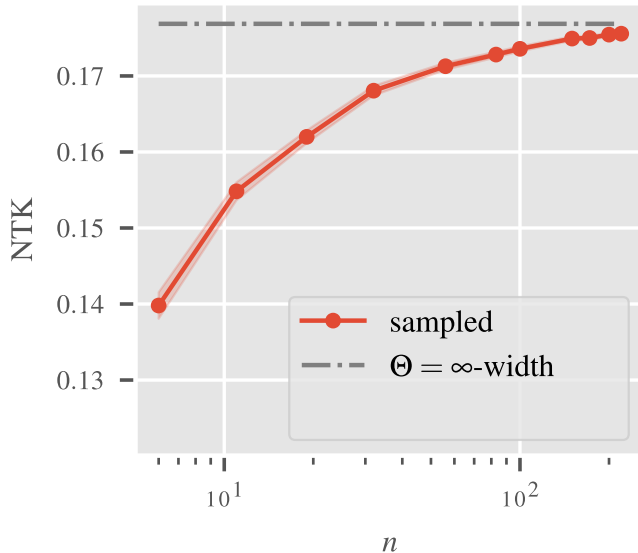
Abstract

Neural tangent kernels (NTKs) are a powerful tool for analyzing deep, non-linear neural networks. In the infinite-width limit, NTKs can easily be computed for most common architectures, yielding full analytic control over the training dynamics. However, at infinite width, important properties of training such as NTK evolution or feature learning are absent. Nevertheless, finite width effects can be included by computing corrections to the Gaussian statistics at infinite width. We introduce Feynman diagrams for computing finite-width corrections to NTK statistics. These dramatically simplify the necessary algebraic manipulations and enable the computation of layer-wise recursive relations for arbitrary statistics involving preactivations, NTKs and certain higher-derivative tensors (dNTK and

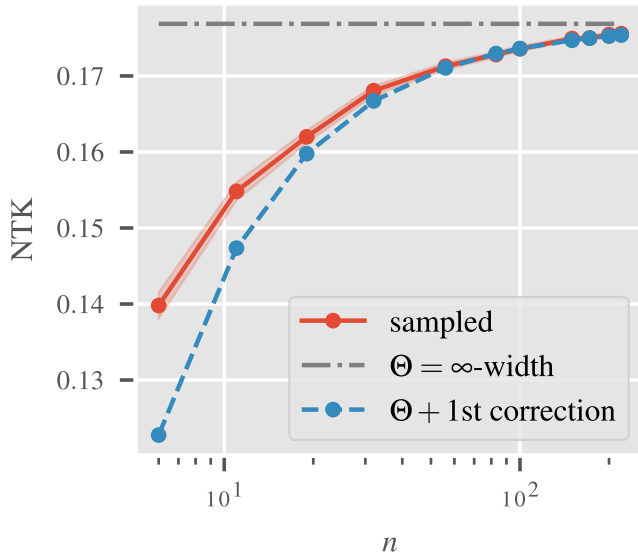
Finite-width corrections



Finite-width corrections



Finite-width corrections



Outlook

- Understanding training dynamics of neural networks remains challenging

Outlook

- Understanding training dynamics of neural networks remains challenging
- ⇒ Use symmetric case as simplified approach to study training

Outlook

- Understanding training dynamics of neural networks remains challenging
- ⇒ Use symmetric case as simplified approach to study training
- How does data augmentation affect the training dynamics?

Outlook

- Understanding training dynamics of neural networks remains challenging
- ⇒ Use symmetric case as simplified approach to study training
- How does data augmentation affect the training dynamics?
 - How do the training dynamics of equivariant models differ from those of non-equivariant models?

Outlook

- Understanding training dynamics of neural networks remains challenging
- ⇒ Use symmetric case as simplified approach to study training
- How does data augmentation affect the training dynamics?
- How do the training dynamics of equivariant models differ from those of non-equivariant models?
- Use insights from physics to make progress

Website

GAPinDNNs

Home

Members

Research

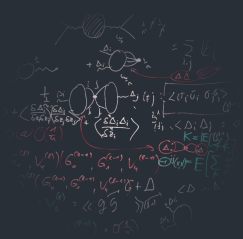
Output

Seminar

Teaching

Positions

Geometry, Algebra and Physics in
Deep Neural Networks



The research group on Geometry, Algebra and Physics in Deep Neural Networks (GAPinDNNs) is based at the Department for Mathematical Sciences at Chalmers University of Technology and the University of Gothenburg. Our vision is to develop a mathematical foundation for deep learning which elevates the field into a theoretically well-grounded science.

News

Paper accepted in NeurIPS 2025

22 Sep 2025

Our paper on *Learning Chern Numbers of Topological Insulators with Gauge Equivariant Neural Networks* has been accepted for a poster at NeurIPS 2025! In this paper, we combine lattice gauge equivariant networks with a novel training mechanism to learn topological invariants (Chern numbers) of topological insulators. This paper combines several beautiful topics in machine learning, physics and mathematics.

First author is our new PhD student [Longde Huang](#). Congratulations to his first publication! From our group, [Hampus Linde](#), [Daniel Persson](#) and [Jan Gertken](#) were also involved. Thanks to our physics-collaborators [Oleksandr Babichenov](#) (then at Stockholm University) and [Mats Granath](#) (University of Gothenburg) for their expertise and a fun collaboration!



Thank you!