Equivariant Neural Networks: From Training Dynamics to Topology

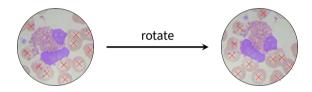
Jan E. Gerken

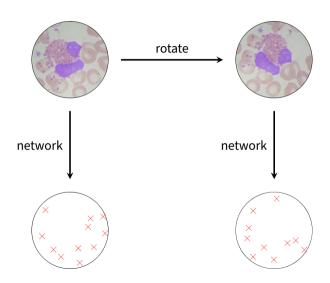


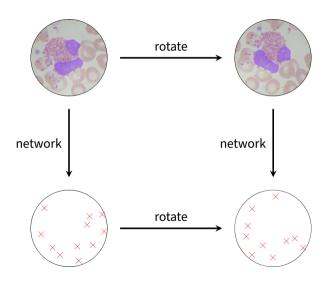


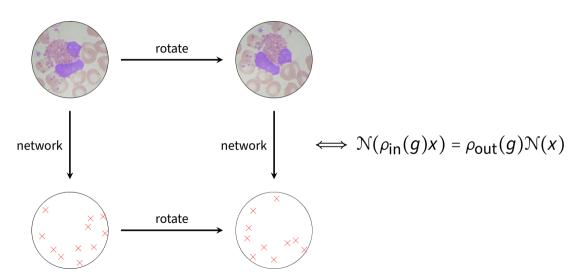


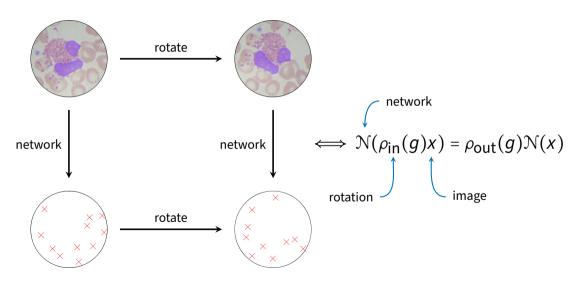




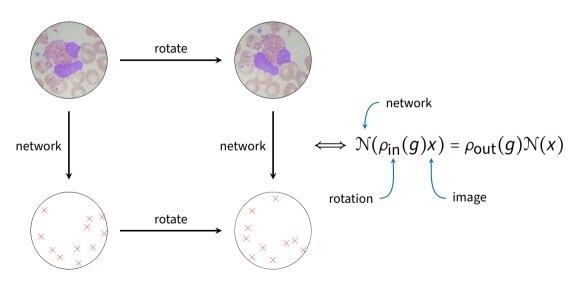








Equivariance



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

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeCorrolation layers can be used effectively in a deep network because all the layers in such an entwork nee transfactive equirators: shifting the image and then feeding it through a runther of layers is the same as feeding the ceiginal image through the same layers and then shifting the resultine feature mans (at least un to edoe-effects). In

Group Equivariant Convolutional Networks

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

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 symmeCorrodution layers can be used effectively in a deep network because all the layers in such a network are translative equitariater; shifting the image and then feeding it through a number of layers is the same no feeding the ceiginal image through the same layers and then shifting the resultine feature mans (at least un to edue-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation intratances of a domain be incorporated into the architecture of a neural network? We propose Equivation 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 cannels of conlines of the property of the property of the property of the statement of the property of the scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation imariance 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; Dielensan et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Henriques & Vodaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are automated for moment medicatable to automate by temporament.

Group Equivariant Convolutional Networks

Taco S. Cohen University of Amsterdam

Max Welling

T.S.COHEN@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 symmeCorrelation layers can be used effectively in a deep network because all the layers in such a network me translative equirarises: 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 resultine feature mans (at least un to edue-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation interatances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformer (ETX), a family of differentiable image-to-image mappings that improve the relocations of models towards pre-defined continuous transformation groups. Those the proposed of the proposed of the prolenges of the proposed of the prolenges of the proposed of the protocological pr 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) Marcos et al., 2017; Weerall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are extrastated to moreous predictable in nonricolar transference

Theory for Equivariant Quantum Neural Networks

Ogruh T. Nguya, ^{1,1} Lusis Schattal, ^{1,1} Paulo Braccia, ^{1,1} Medaer Bagonn, ^{1,1} Parick J. Codes, ¹ Pedrick awange, ¹ Marth Lacesca, ¹ and M. Cewaco, ¹ Marth M. Cewaco, ¹ Marth M. Cewaco, ¹ Marth M. Cewaco, ¹ Martine States Marine National Automation Lanca, New Marine 1924, USA ¹ Submersion Stimus, ¹ Leaders, ¹ Keman, Nicoland Laboratory, ¹ Leaders, ¹ Memory Marine 1924, USA ¹ Information Stimus, ¹ Leaders, ¹ Memory Marine 1924, USA ¹ Department of Harbert and Computer Supraerray, ¹ Memory Marine 1924, USA ¹ Memory Memory Marine 1924, USA ¹ Memory Memory

^b Dipartimento di Finica e Astronomia, Università di Firenzo, Seste Fiorentino (FI), 50019, Italy a Dipartimento di Finica e Astronomia, Università di Firenzo, Seste Fiorentino (FI), 50019, Italy a Department of Mathematica, University of California Davis, Davis, Coldornia 95046, USA Center for Neolinear Stadies, Los Alamas National Laboratory, Los Alamos, New Mexico 87545, USA Daustum purual network architectures that have little-to-no industrine biasse are known to fee

trainability and generalization issues. Impired by a similar problem, recent brackthroughs 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 consents with but of the neurantin. In this word, and insent those ideas to the constitution of the property of the constitution of the constitu

M.WELLING@UVA.NL

Group Equivariant Convolutional Networks

Taco S. Cohen University of Amsterdam

T.S.COHEN@UVA.NL

Max Welling University of Amsterdam University of California Irvine

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 symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are mustlative equinarious: shifting the image and thea feeding it through a runther of layers is the same as feeding the original image through the same layers and then shifting the resultine feature mans (at least us to coler-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation intrafasces of a domain be incorporated into the architecture of a neural network? We propose figures in Transformers (ETs), a fample of differentiable image-to-image mappings that improve the robustness of models towards pre-diffield confinements are foundation groups. Through the use of specially-derived carotical coconditions eventure. The incorporate function that scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no genantice 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) Marcos et al., 2017; Weerall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are extrastated to moreous predictable in nonricolar transference

Theory for Equivariant Quantum Neural Networks

Ogruh T. Nguya, ^{1,1} Lusis Schattal, ^{1,1} Paulo Braccia, ^{1,1} Medaer Bagonn, ^{1,1} Parick J. Codes, ¹ Pedrick awange, ¹ Marth Lacesca, ¹ and M. Cewaco, ¹ Marth M. Cewaco, ¹ Marth M. Cewaco, ¹ Marth M. Cewaco, ¹ Martine States Marine National Automation Lanca, New Marine 1924, USA ¹ Submersion Stimus, ¹ Leaders, ¹ Keman, Nicoland Laboratory, ¹ Leaders, ¹ Memory Marine 1924, USA ¹ Information Stimus, ¹ Leaders, ¹ Memory Marine 1924, USA ¹ Department of Harbert and Computer Supraerray, ¹ Memory Marine 1924, USA ¹ Memory Memory Marine 1924, USA ¹ Memory Memory

⁵ Dipartimento d'Internation de Creace-Compagni, creato, dimini, creato, dimini et 1891; USA
⁵ Dipartimento di Fiscia e Astronomia, Università di Firenze, Seato Fiorentine (FI), 50019 , Italy
⁶ Department of Mathematics, University of California Davis, Davis, California 5646, USA
⁷ Center for Nonlinear Studies, Los Alamas National Laboratory, Los Alamos, New Mexico 87845, USA
Control programme pural network architectures that have little-to-no industries biases are known to for

Quantum neural network arctitectures 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 equivariants around networks whose action conmute with that of the accurate. In this work we inspired these these to the accurate models the

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,c,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

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Thomasureum East Board Residence 199199. China

2honggueneun East Road, Berjing 10019t ^bMicrosoft Research Asia.

Danling Street, Beijing 100080, China **CERN, EP Department,

CH-1211 Geneva 23, Switzerland

 $^dSchool\ of\ Physics,\ Peking\ University,$

Chenafa Road. Beijina 100871. China

M.WELLING@UVA.NL

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

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are translative equiracts shifting the image and then feeding it through a rumber of layers in the same as feeding the original image through the same layers and then shifting the resultine feature mans (at least un to edue-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network! We propose Engineerin Transformers (ETs), a fainity of differentiable imagoto-image mappings that improve the robustness of models towards pre-diffield continuous transformation groups. Through the use of specially-derived canonical coconfining systems. PER internesses functions the scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guerantee 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 (Coben & 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 surrouted to removed mediciable in auxiliariable transforma-

Theory for Equivariant Quantum Neural Networks

Quyah T. Nguyan, ^{1,1} Lunis Schattal, ^{1,1} Paolo Braccia, ^{1,1} Medari Bagona, ^{1,1}
Patrick J. Codes, Prefeite Sensong, ¹ Martin Lances, ¹ and M. Coresco, ¹ Memorical Districts, ¹ Annex Roberts (Marco, New Mexico 1974), ¹ USA ¹ Information Science, Inc. Almons, Nova Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Information Science, ¹ Lee Almons, Nov. Mexico 1974, ¹ USA ¹ Lee Almons, ¹ L

Onterently of Illicons of Orlans-Champaign, Orlans, filmins 61801, CSA.
⁸ Dipartiment of Piscae Astronomic, Uniscentia of Pirente, State Forentine [F1], 50019, Buly a Department of Mathematics, Uniscensity of California Davis, Davis, California 9616, USA.
⁸ Center for Nonlinear Studies, Los Aismas National Laboratory, Los Aismas, New Mexico 87845, USA.
Onanatum noural network architectures that have Bitlato, no industrie biases are known to fore.

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 neutral networks whose action conmitted that of the consistent in this work is insent these ideas to the assessment such to

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

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

"Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Thomasumous Foot Road Resident 199199, China

^bMicrosoft Research Asia,

Danling Street, Beijing 100080, China **CERN, EP Department,

CH-1211 Geneva 23, Switzerland

School of Physics, Peking University, Chengla Road, Beijina 100871. China

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

Simon Batzner*, Albert Musoelian, Lixin Sun, Mario Geiger, Jonathan P. Mallon, Mordechal Kombluth, Nicola Mollman, Tess E. Smidt, A. and Borts Kozinsky*1,3 and Jordan A. Pankon Schul of Fanjaceria and Amilia Sciences.

¹ John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
² Ecole Polytechnique, Edding, & Louvene, 1015, Laurence, Science

Ecole Freguezzanque recereia no Lonassino, 1910 Lonassino, Commercia de Conspilation del Technologo Center, Cambridge, M. (2013), 1834.
*Computational Research Dissission and Center for Advanced Mathematics for Energy Research Applications, Learner Berkeley Medisonal Industry, Berkeley, CA 91270, USA.
*Massachusetts Institute of Technologo, Department of Electrical Environmental Commercia Vision. Commercia Med. 2012. USA.

This work presents Neural Enginearium Interatorate Potentials (NegalT), an [5], equivainnt, second services approach for learning interactions; potential reach suppose the for learning interesting potentials from a social exclusion for reduced dynamics simulations. While most contemporary symmetry-source models use invariant convolutions and only act on socials, NegalT employs, [5], equivalent convolutions for interactions of geometric tensors, resulting in a more information-rich and faithful representation of atomic environments. The method architects state-of-the-architects are convolutionally and relevance to of moleculum of the property of t

M.WELLING@UVA.NL

Group Equivariant Convolutional Networks

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

Max Welling University of Amsterdam University of California Irvine

Canadian Institute for Advanced Research

Abstract

We introduce Group assissment Completional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are trouvlation 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 many (at least up to odae-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

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 caponical coordinate systems. ETs incorporate functions that scaling to each training image). While data augmentation typically below 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 sugmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrell et al., 2017; Henriques & Visdaldi, 2017; Cohenet al., 2018) has explored new CNN architectures that are

Theory for Equivariant Quantum Neural Networks

Ouvub T. Neuven 3,2 Louis Schatzki 3,4 Paolo Braccia 3,5 Michael Rarone 3,6 Patrick J. Coles, Frédéric Sauvage, Martin Larocca, 1,7 and M. Cerezo ¹ Theoretical Division, Los Alamos National Laboratoru, Los Alamos, New Mexico 87515, USA ²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 62138, USA Information Sciences Les Alemas National Laboratory, Les Alemas New Merico 87515, 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 Department of Mathematics, University of California Davis, Davis, California 95616, USA Ovantum 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 let Tagging

Shini Gong^{4,c,1} Qi Meng⁵ Jue Zhang⁵ Huilin Qu^c Congniao Li^c Sitian Qian^d Weitao Dug Zhi-Ming Mag Tip-Yan Liub

- * Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Thomasuspeam Fort Board Reiting 100190 China
- b Microsoft Research Asia.
- Danling Street, Briting 100080, China
- CERN, EP Department, CH-1211 Geneva 23. Switzerland
- & Submitted District Public Plainments Chenefa Road. Beijina 100871. China

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

Simon Batzner* 1 Albert Muselian 1 Livin Sun 1 Mario Geiser 2 Jonathan P. Mailon 3 Mordechai Kornbluth. 3 Nicola Molinari. 1 Tess E. Smidt. 4,5 and Boris Kozinsky 1,3 John A. Postson School of Engineering and Amilial Sciences.

Harvard University, Cambridge, MA 02138, USA

*Robert Basch Research and Technology Center, Cambridge, MA 02139, USA ⁴Computational Research Division and Center for Advanced Mathematics for Energy Research Applications. Engineering and Computer Science, Cambridge, MA 09119, USA

This work presents Neural Equivariant Interatomic Potentials (NeguIP), an E(3)-equivariant dynamics simulations. While most contemporary symmetry-aware models use invariant convolutions and only act on scalars. NaouIP condess E(T) conjunions consolutions for interactions of prometric 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-FOULVARIANT NEURAL NETWORKS TO SELECT STRUCTURAL MODELS OF PROTEIN COMPLEXES

Stephan Elemann* Department of Applied Physics Stanford University naismannflatanford adu

Raphael LL. Townshend: Nathaniel Thomas* Department of Physics ranhanifor stanford adu

Milled Innets Bowen Box Department of Electrical Engineering Department of Computer Science Stanford University Stanford University miarota@stanford.edu hiine@atanford.edu

nthomas 103ffemail.com

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

ABSTRACT

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

Training Dynamics

✓ A lot of work on equivariant architectures

Training Dynamics

- ✓ A lot of work on equivariant architectures
- Training dynamics much less studied

Empirical NTK

Training dynamics under continuous gradient descent:

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

Е

Empirical NTK

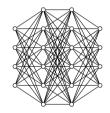
Training dynamics under continuous gradient descent:

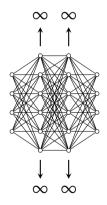
learning rate
$$\frac{d\mathcal{N}_{\theta}(x)}{dt} = -\frac{\eta}{N} \sum_{i=1}^{N} \Theta_{\theta}(x, x_{i}) \frac{\partial L}{\partial \mathcal{N}(x_{i})}$$
training sample

with the empirical neural tangent kernel (NTK)

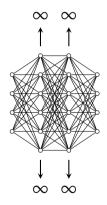
$$\Theta_{\theta}(x,x') = \sum_{\mu} \frac{\partial \mathcal{N}(x)}{\partial \theta_{\mu}} \frac{\partial \mathcal{N}(x')}{\partial \theta_{\mu}}$$

[Jacot et al. 2018]



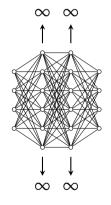


[Jacot et al. 2018]

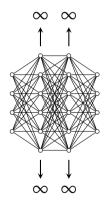


△ NTK becomes independent of initialization

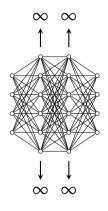
[Jacot et al. 2018]



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



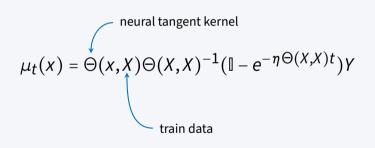
- △ NTK becomes independent of initialization
- 凸 NTK becomes constant in training
- △ NTK can be computed for most networks

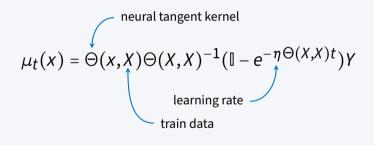


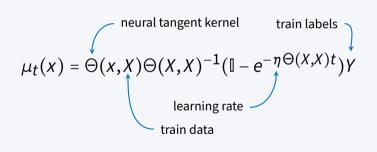
- △ NTK becomes independent of initialization
- **心** NTK can be computed for most networks
- ✓ Training dynamics can be solved

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

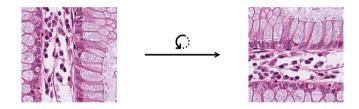
neural tangent kernel
$$\mu_t(x) = \Theta(x,X)\Theta(X,X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})Y$$

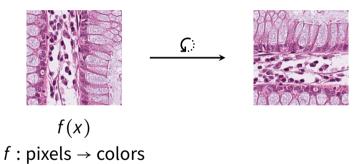


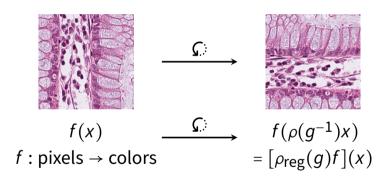




Equivariant Neural Tangent Kernels







[Cohen, Welling 2016]

[Cohen, Welling 2016]

Group conv's are the (unique) linear layers equivariant wrt $ho_{
m reg}$

Group conv's are the (unique) linear layers equivariant wrt $ho_{ m reg}$

Ordinary convolutions

$$f'(y) = \int_X dx \, \kappa(x - y) \, f(x)$$

Group conv's are the (unique) linear layers equivariant wrt ρ_{reg}

Ordinary convolutions

$$f'(y) = \int_X dx \, \kappa(x - y) \, f(x)$$

Group convolutions

$$f'(g) = \int_{X} dx \, \kappa(\rho(g^{-1})x) f(x)$$
 lifting

Group conv's are the (unique) linear layers equivariant wrt ρ_{reg}

Ordinary convolutions

$$f'(y) = \int_X dx \, \kappa(x - y) \, f(x)$$

Group convolutions

$$f'(g) = \int_X dx \, \kappa(\rho(g^{-1})x) \, f(x)$$
 lifting
$$f'(g) = \int_G dg \, \kappa(g^{-1}h) \, f(h)$$
 group convolution

group pooling

Group conv's are the (unique) linear layers equivariant wrt ρ_{reg}

Ordinary convolutions

$$f'(y) = \int_{X} dx \, \kappa(x-y) \, f(x)$$

Group convolutions

$$f'(g) = \int_X dx \, \kappa(\rho(g^{-1})x) \, f(x)$$
 lifting
$$f'(g) = \int_G dg \, \kappa(g^{-1}h) \, f(h)$$
 group convolution
$$f' = \frac{1}{\text{vol}(G)} \int_G dg \, f(g)$$
 group pooling

Stack GConv-layers to obtain an invariant network

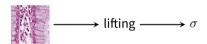
Stack GConv-layers to obtain an invariant network



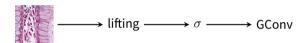
Stack GConv-layers to obtain an invariant network



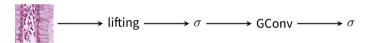
Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



For GCNN-layers, define the NNGP and NTK via

$$K_{\mathbf{g},\mathbf{g'}}^{(\ell)}(f,f') = \mathbb{E}\left[\left[\mathcal{N}^{(\ell)}(f)\right](\mathbf{g})\left(\left[\mathcal{N}^{(\ell)}(f')\right](\mathbf{g'})\right)^{\mathsf{T}}\right]$$

For GCNN-layers, define the NNGP and NTK via

$$K_{g,g'}^{(\ell)}(f,f') = \mathbb{E}\left[\left[\mathcal{N}^{(\ell)}(f)\right](g)\left(\left[\mathcal{N}^{(\ell)}(f')\right](g')\right)^{\mathsf{T}}\right] \\
\Theta_{g,g'}^{(\ell)}(f,f') = \mathbb{E}\left[\sum_{\ell'=1}^{\ell} \frac{\partial[\mathcal{N}^{(\ell)}(f)](g)}{\partial \theta^{(\ell')}}\left(\frac{\partial[\mathcal{N}^{(\ell)}(f')](g')}{\partial \theta^{(\ell')}}\right)^{\mathsf{T}}\right]$$

$$\left[\mathcal{N}^{(\ell)}(f)\right](g) = \int_{G} \mathrm{d}g \, \kappa(g^{-1}h) \left[\mathcal{N}^{(\ell-1)}(f)\right](h)$$

The layer-recursion for a GCNN-layer is given by

$$K_{g,g'}^{(\ell+1)}(f,f') = \frac{1}{|S_{\kappa}|} \int_{S_{\kappa}} dh \, K_{gh,g'h}^{(\ell)}(f,f')$$

$$\left[\mathbb{N}^{(\ell)}(f) \right] (g) = \int_{G} \mathrm{d}g \, \kappa(g^{-1}h) \left[\mathbb{N}^{(\ell-1)}(f) \right] (h)$$

The layer-recursion for a GCNN-layer is given by

$$K_{g,g'}^{(\ell+1)}(f,f') = \frac{1}{|S_{\kappa}|} \int_{S_{\kappa}} dh \, K_{gh,g'h}^{(\ell)}(f,f')$$

$$\Theta_{g,g'}^{(\ell+1)}(f,f') = K_{g,g'}^{(\ell+1)}(f,f') + \frac{1}{|S_{\kappa}|} \int_{S_{\kappa}} dh \, \Theta_{gh,g'h}^{(\ell)}(f,f')$$

Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



Stack GConv-layers to obtain an invariant network



Compute NTK with layer-wise recursion

Stack GConv-layers to obtain an invariant network



$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f')$$

Stack GConv-layers to obtain an invariant network



$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f') \longrightarrow \Theta_{g,g'}^{(2)}(f,f')$$

Stack GConv-layers to obtain an invariant network



$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f') \longrightarrow \Theta_{g,g'}^{(2)}(f,f') \longrightarrow \Theta_{g,g'}^{(3)}(f,f')$$

Stack GConv-layers to obtain an invariant network



$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f') \longrightarrow \Theta_{g,g'}^{(2)}(f,f') \longrightarrow \Theta_{g,g'}^{(3)}(f,f') \longrightarrow \Theta_{g,g'}^{(4)}(f,f')$$

Stack GConv-layers to obtain an invariant network



$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f') \longrightarrow \Theta_{g,g'}^{(2)}(f,f') \longrightarrow \Theta_{g,g'}^{(3)}(f,f') \longrightarrow \Theta_{g,g'}^{(4)}(f,f') \longrightarrow \Theta_{g,g'}^{(5)}(f,f')$$

Stack GConv-layers to obtain an invariant network

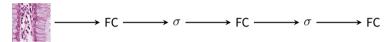


$$0 \longrightarrow \Theta_{g,g'}^{(1)}(f,f') \longrightarrow \Theta_{g,g'}^{(2)}(f,f') \longrightarrow \Theta_{g,g'}^{(3)}(f,f') \longrightarrow \Theta_{g,g'}^{(4)}(f,f') \longrightarrow \Theta_{g,g'}^{(5)}(f,f') \longrightarrow \Theta(f,f')$$

• Consider two neural networks

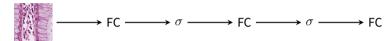
Consider two neural networks

An MLP



Consider two neural networks

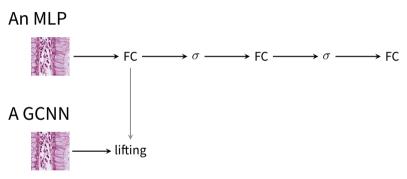
An MLP



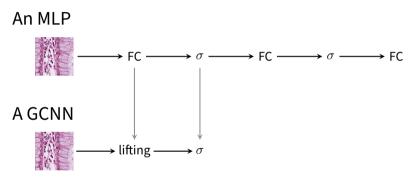
A GCNN



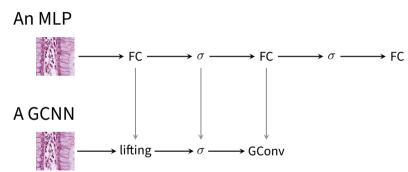
Consider two neural networks



Consider two neural networks



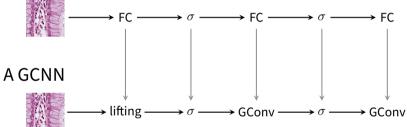
Consider two neural networks



Consider two neural networks

Consider two neural networks

An MLP



NTKs of MLPs and GCNNs

Consider two neural networks

An MLP

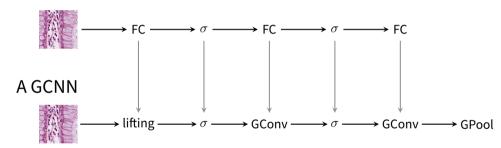
FC $\rightarrow \sigma \rightarrow FC \rightarrow \sigma \rightarrow FC$ A GCNN

lifting $\rightarrow \sigma \rightarrow GConv \rightarrow \sigma \rightarrow GConv \rightarrow GPool$

NTKs of MLPs and GCNNs

Consider two neural networks

An MLP



• Then

$$\Theta^{\mathsf{GCNN}}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\mathsf{MLP}}(f,\rho_{\mathsf{reg}}(g)f')$$

f trained without data augmentation f trained with data augmentation f $\Theta^{\mathsf{non-aug}}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\mathsf{aug}}(f,\rho_{\mathsf{reg}}(g)f')$

Then

$$\mu_t^{\text{non-aug}}(x) = \mu_t^{\text{aug}}(x)$$

at infinite width. ⇒ Predictions of infinite ensembles agree.

Then

$$\mu_t^{\text{non-aug}}(x) = \mu_t^{\text{aug}}(x) \quad \forall t$$

at infinite width. ⇒ Predictions of infinite ensembles agree.

Then

$$\mu_t^{\text{non-aug}}(x) = \mu_t^{\text{aug}}(x) \quad \forall t \quad \forall x$$

at infinite width. ⇒ Predictions of infinite ensembles agree.

$$\Theta^{\mathsf{GCNN}}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\mathsf{MLP}}(f,\rho_{\mathsf{reg}}(g)f')$$

before: non-aug
$$\Theta^{\mathsf{GCNN}}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\mathsf{MLP}}(f,\rho_{\mathsf{reg}}(g)f')$$

before: non-aug before: aug
$$\Theta^{\sf GCNN}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\sf MLP}(f,\rho_{\sf reg}(g)f')$$

before: non-aug before: aug
$$\Theta^{\sf GCNN}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\sf MLP}(f,\rho_{\sf reg}(g)f')$$

training the MLP on G-augmented data

before: non-aug before: aug
$$\Theta^{\sf GCNN}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\sf MLP}(f,\rho_{\sf reg}(g)f')$$

training the MLP on G-augmented data

training the GCNN on unaugmented data

before: non-aug before: aug
$$\Theta^{\sf GCNN}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\sf MLP}(f,\rho_{\sf reg}(g)f')$$

training the MLP on G-augmented data

training the GCNN on unaugmented data

in the ensemble mean

before: non-aug before: aug
$$\Theta^{\sf GCNN}(f,f') = \frac{1}{|G|} \sum_{g \in G} \Theta^{\sf MLP}(f,\rho_{\sf reg}(g)f')$$

training the MLP on G-augmented data

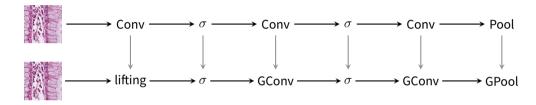
training the GCNN on unaugmented data

in the ensemble mean, $\forall t$, $\forall x$

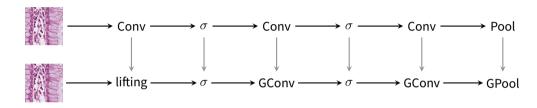
Consider a CNN



Consider a CNN and a GCNN invariant wrt. roto-translations



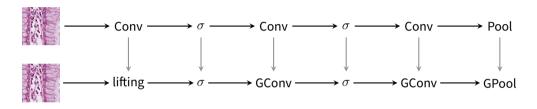
Consider a CNN and a GCNN invariant wrt. roto-translations



Then

$$\Theta^{\mathsf{GCNN}}(f, f') = \frac{1}{n} \sum_{r \in C_n} \Theta^{\mathsf{CNN}}(f, \rho_{\mathsf{reg}}(r)f')$$

Consider a CNN and a GCNN invariant wrt. roto-translations



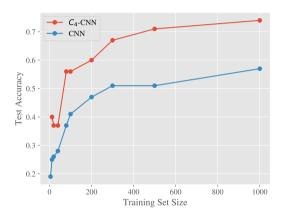
Then

$$\Theta^{\mathsf{GCNN}}(f,f') = \frac{1}{n} \sum_{r \in C_n} \Theta^{\mathsf{CNN}}(f, \rho_{\mathsf{reg}}(r)f')$$

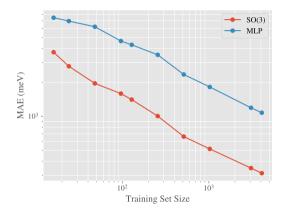
⇒ By training the CNN on rotated images, one obtains a roto-translation invariant GCNN

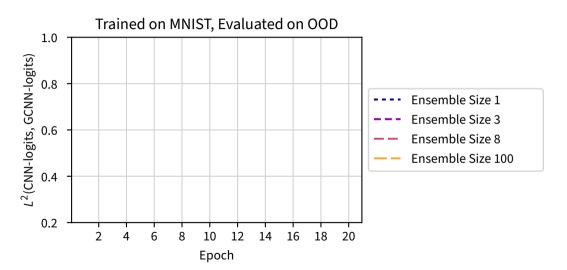
Experiments

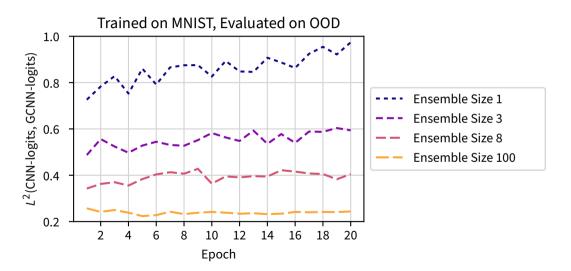
Equivariant NTKs for medical image classification

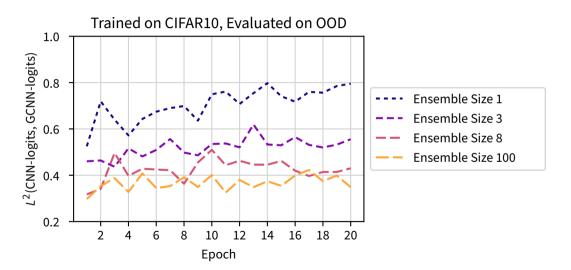


Equivariant NTKs for molecular property regression









Learning Topology with Gauge Equivariant Neural Networks

• So far, we discussed global symmetries:

$$[\rho_{\mathsf{reg}}(g)f](x) = f(\rho(g^{-1})x) \quad g \in G$$

• So far, we discussed global symmetries:

$$[\rho_{\text{reg}}(g)f](x) = f(\rho(g^{-1})x) \quad g \in G$$

• Now, extend to local (gauge) symmetries:

$$[\rho_{\text{gauge}}(g)f](x) = f(\rho(g_{\mathbf{X}}^{-1})x) \quad g_{\mathbf{X}} \in G$$

• So far, we discussed global symmetries:

$$[\rho_{\mathsf{reg}}(g)f](x) = f(\rho(g^{-1})x) \quad g \in G$$

Now, extend to local (gauge) symmetries:

$$[\rho_{\text{gauge}}(g)f](x) = f(\rho(g_{\mathbf{X}}^{-1})x) \quad g_{\mathbf{X}} \in G$$

For instance, choose basis in tangent space at each point

• So far, we discussed global symmetries:

$$[\rho_{\text{reg}}(g)f](x) = f(\rho(g^{-1})x) \quad g \in G$$

• Now, extend to local (gauge) symmetries:

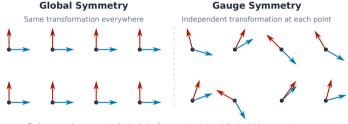
$$[\rho_{\mathsf{gauge}}(g)f](x) = f(\rho(g_{\mathsf{x}}^{-1})x) \quad g_{\mathsf{x}} \in G$$

- For instance, choose basis in tangent space at each point
- ① This symmetry is exponentially larger!

$$G_{\text{gauge}} = G^{|X|}$$

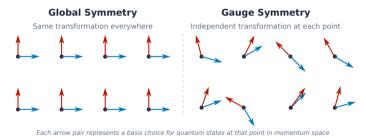
Topological materials

 Need to choose an arbitrary basis for quantum states at each point in momentum space



Topological materials

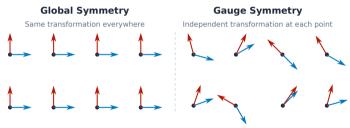
 Need to choose an arbitrary basis for quantum states at each point in momentum space



 $\Rightarrow U(N)$ gauge symmetry at each point

Topological materials

 Need to choose an arbitrary basis for quantum states at each point in momentum space



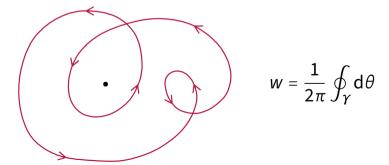
- Each arrow pair represents a basis choice for quantum states at that point in momentum space
- $\Rightarrow U(N)$ gauge symmetry at each point
- Topology of states determines physical properties and is independent of choice of basis

The Chern number

The topology is characterized by the Chern number

$$C = \frac{1}{2\pi i} \int \text{Tr}(F(k)) d^2k$$

Higher-dimensional analog of winding number



The Chern number

• On a discrete lattice in 2d, it is given by

$$C = \sum_{i,j} \operatorname{Im}(\operatorname{Tr}(\log W_{ij})) \in \mathbb{Z}$$

The Chern number

On a discrete lattice in 2d, it is given by

Wilson loop,
$$N \times N$$
 matrix
$$C = \sum_{i,j} \operatorname{Im}(\operatorname{Tr}(\log W_{ij})) \in \mathbb{Z}$$

The Chern number is invariant under gauge transformations

$$W_{ij} \to \Omega_{ij}^{\dagger} W_{ij} \Omega_{ij}$$
 with $\Omega_{ij} \in U(N)$

The Chern number

On a discrete lattice in 2d, it is given by

Wilson loop,
$$N \times N$$
 matrix
$$C = \sum_{i,j} \operatorname{Im}(\operatorname{Tr}(\log W_{ij})) \in \mathbb{Z}$$

The Chern number is invariant under gauge transformations

$$W_{ij} \to \Omega_{ij}^{\dagger} W_{ij} \Omega_{ij}$$
 with $\Omega_{ij} \in U(N)$

The Chern number

• On a discrete lattice in 2d, it is given by

$$C = \sum_{i,j} \operatorname{Im}(\operatorname{Tr}(\log W_{ij})) \in \mathbb{Z}$$

The Chern number is invariant under gauge transformations

$$W_{ij} \to \Omega_{ij}^{\dagger} W_{ij} \Omega_{ij}$$
 with $\Omega_{ij} \in U(N)$

Learning problem: Predict C given $\{W_{ij} | (i, j) \in \text{lattice}\}$

 Since Tr(log X) = log(det X), try learning the determinant as a toy problem

 Since Tr(log X) = log(det X), try learning the determinant as a toy problem

MLPs cannot learn determinants

• The determinant of a 4 × 4 matrix is

$$\det A = \sum_{i,j,k,l=1}^{4} \varepsilon^{ijkl} A_{1i} A_{2j} A_{3k} A_{4l}$$

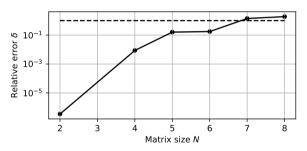
The determinant of a 4 × 4 matrix is

$$\det A = \sum_{i,j,k,l=1}^{4} \varepsilon^{ijkl} A_{1i} A_{2j} A_{3k} A_{4l}$$

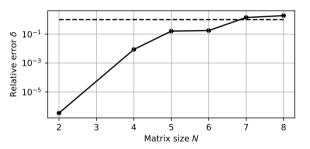
Build a ResNet with polynomial layers

$$A_{ij}^{\text{out}} = \sum_{k_1, \dots, k_{2R}} \theta_{ij}^{k_1 \dots k_{2R}} A_{k_1 k_2}^{\text{in}} \dots A_{k_{2R-1} k_{2R}}^{\text{in}}$$

 Even deep models with layers of order R ≤ 4 cannot learn determinants of larger matrices



① Even deep models with layers of order $R \le 4$ cannot learn determinants of larger matrices



⇒ Use gauge equivariant model to learn Chern number

Gauge Equivariant Network

[Favoni et al. 2022]

Gauge equivariant network for lattice QCD

[Favoni et al. 2022]

- Gauge equivariant network for lattice QCD

channel index

• Gauge Equivariant Bilinear Layer (GEBL) channel index
$$W_{ij}^{\prime \gamma} = \sum_{\mu\nu} \alpha_{\gamma\mu\nu} W_{ij}^{\mu} W_{ij}^{\nu}$$
trainable parameters

channel index

Gauge Equivariant Network

- Gauge equivariant network for lattice QCD
- Gauge Equivariant Bilinear Layer (GEBL)

$$W'_{ij}^{\gamma} = \sum_{\mu\nu} \alpha_{\gamma\mu\nu} W^{\mu}_{ij} W^{\nu}_{ij}$$
 trainable parameters

Gauge Equivariant Activation Layer (GEAct)

$$W'_{ij}^{\gamma} = \sigma(\text{Re}(\text{Tr}\,W_{ij}^{\gamma}))W_{ij}^{\gamma}$$

- Gauge equivariant network for lattice QCD
- Gauge Equivariant Bilinear Layer (GEBL) ____ channel index

$$W'_{ij}^{\gamma} = \sum_{\mu\nu} \alpha_{\gamma\mu\nu} W^{\mu}_{ij} W^{\nu}_{ij}$$
 trainable parameters

Gauge Equivariant Activation Layer (GEAct)

$$W'_{ij}^{\gamma} = \sigma(\operatorname{Re}(\operatorname{Tr} W_{ij}^{\gamma}))W_{ij}^{\gamma}$$
 N_{x}
 N_{ch}
 N

[Huang et al. 2025]

For a compact Lie group G and bounded and non-decreasing activation function σ , GEBLNets can approximate any gauge invariant function.

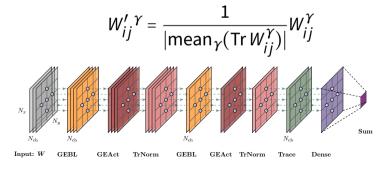
• Train on randomly sampled Wilson loops W_{ij} , generate labels from analytical expression

- Train on randomly sampled Wilson loops W_{ij} , generate labels from analytical expression
- Training is unstable due to bilinear layers

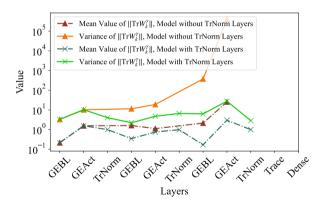
- Train on randomly sampled Wilson loops W_{ij} , generate labels from analytical expression
- Training is unstable due to bilinear layers
- Introduce gauge equivariant normalization layer (TrNorm)

$$W'_{ij}^{\gamma} = \frac{1}{|\text{mean}_{\gamma}(\text{Tr}\,W_{ij}^{\gamma})|}W_{ij}^{\gamma}$$

- Train on randomly sampled Wilson loops W_{ij} , generate labels from analytical expression
- Training is unstable due to bilinear layers
- Introduce gauge equivariant normalization layer (TrNorm)



The TrNorm layers stabilize the training statistics



Results

The normalized network can predict Chern numbers for higher bands

Bands	4	5	6	7	8
Accuracy	95.9%	94.0%	93.8%	91.7%	52.5%

• The entire network architecture before the final sum is per-site

- The entire network architecture before the final sum is per-site
- Can learn local features just from topologically trivial (C = 0) samples: No training labels necessary!

- The entire network architecture before the final sum is per-site
- Can learn local features just from topologically trivial (C = 0) samples: No training labels necessary!
- Network generalizes to topologically non-trivial (C > 0) samples

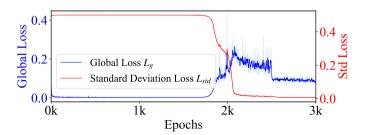
• Problem: Collapse to zero-solution

- Problem: Collapse to zero-solution
- Solution: Add loss *L*_{std} to encourage non-zero std across sites

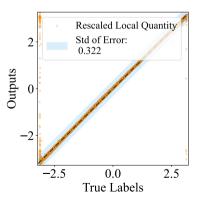
$$L_{std} = ||\min(\operatorname{std}(g(W_{ij})), \delta) - \delta||_1$$
per-site output of final dense layer

- Problem: Collapse to zero-solution
- Solution: Add loss *L*_{std} to encourage non-zero std across sites

$$L_{std} = ||\min(\operatorname{std}(g(W_{ij})), \delta) - \delta||_1$$
per-site output of final dense layer

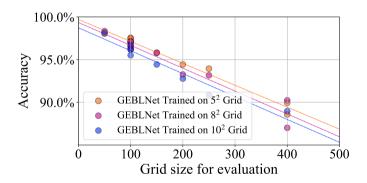


Network trained on trivial topology alone can predict non-trivial Chern numbers up to global factor



Generalization to larger grids

Since the learned features are local, network generalizes to larger grids



① Training dynamics of equivariant networks can be studied using equivariant NTKs

- Training dynamics of equivariant networks can be studied using equivariant NTKs
- ② In the infinite-width limit, MLPs trained with data augmentation are equivalent to GCNNs

- Training dynamics of equivariant networks can be studied using equivariant NTKs
- ② In the infinite-width limit, MLPs trained with data augmentation are equivalent to GCNNs
- ③ Gauge equivariant networks are required to learn Chern numbers of topological materials

- ① Training dynamics of equivariant networks can be studied using equivariant NTKs
- ② In the infinite-width limit, MLPs trained with data augmentation are equivalent to GCNNs
- ③ Gauge equivariant networks are required to learn Chern numbers of topological materials
- Networks trained on trivial topologies generalize to non-trivial topologies

Papers

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

 ICML 2025
- Learning Chern Numbers of Topological Insulators with Gauge Equivariant Neural Networks

Longde Huang, Oleksandr Balabanov, Hampus Linander, Mats Granath, Daniel Persson, Jan E. Gerken

NeurIPS 2025



Thank you