Extended Abstract Track

Towards architectural optimization of equivariant neural networks over subgroups

Editors: List of editors' names

Abstract

Incorporating equivariance to symmetry groups in artificial neural networks (ANNs) can improve performance on tasks exhibiting those symmetries, but such symmetries are often only approximate and not explicitly known. This motivates algorithmically optimizing the architectural constraints imposed by equivariance. We propose the equivariance relaxation morphism, which preserves functionality while reparameterizing a group equivariant layer to operate with equivariance constraints on a subgroup, and the $[G]$-mixed equivariant layer, which mixes operations constrained to equivariance to different groups to enable within-layer equivariance optimization. These two architectural tools can be used within neural architecture search (NAS) algorithms for equivariance-aware architectural optimization.

Keywords: equivariance, neural architecture search, geometric deep learning

1. Introduction

Incorporating constraints of symmetry group equivariance into neural networks can improve their task performance, efficiency, and generalization capabilities (Bronstein et al., 2021), as shown by translation-equivariant convolutional neural networks for image-based tasks (Fukushima and Miyake, 1982; LeCun et al., 1989). Seminal works have developed general theories and architectures for equivariance in neural networks, providing a blueprint for equivariant operations on complex structured data (Cohen and Welling, 2016; Ravanbakhsh et al., 2017; Kondor and Trivedi, 2018; Weiler et al., 2021). However, these works design model constraints based on an explicit equivariance property. Furthermore, the architectural assumption of full equivariance may be overly constraining; e.g., in handwritten digit recognition, full equivariance to 180° rotation may lead to misclassifying samples of "6" and "9". Weiler and Cesa (2019) found that local equivariance from a final subgroup convolutional layer improves performance over full equivariance. If appropriate equivariance constraints were instead learned, the benefits of equivariance could extend to applications where the data may have unknown or imperfect symmetries.

Learning approximate equivariance has been recently approached through novel layer operations. Wang et al. (2022) relaxes equivariance through low-rank partial expansion. Finzi et al. (2021) uses an equivariant layer in parallel to a non-equivariant layer with weighted regularization. Zhou et al. (2020) and Yeh et al. (2022) represent symmetry-inducing weight sharing through learnable matrices. Separately, the field of Neural Architecture Search (NAS) aims to optimize full neural network architectures. Existing NAS methods have not yet been developed for optimizing equivariance, although approaches like Romero and Lohit (2022) and van der Ouderaa et al. (2022), which learn partial or soft equivariances for each layer, do allow for custom equivariant architectures. An important aspect of NAS is network morphisms: function-preserving architectural changes (Wei et al.,

© 2022 .

2016) which can be used during training to change the loss landscape and gradient descent trajectory while immediately maintaining the current functionality and loss value (Maile et al., 2022). Agrawal and Ostrowski (2022) presents theoretical contributions on subgroup-based network morphisms for group-invariant shallow neural networks. Our work focuses on layer-wise equivariance for deep convolutional architectures. Developing tools for searching over a space of architectural representations of equivariance would allow for existing NAS algorithms to be applied towards architectural optimization of equivariance.

We present two contributions aimed to enable the search for appropriate equivariance. The first is the *equivariance relaxation morphism* for group convolutional layers that partially expands the parameters of the layer to enable less constrained learning with a prior on symmetry. Secondly, we define the $[G]$-*mixed equivariant layer* that parameterizes a layer as a weighted sum of layers equivariant to different groups, allowing for the learning of architectural weighting parameters. We conclude by proposing extensions of these theoretical results in equivariance-aware neural architecture search for future work.

2. Towards Architectural Optimization over Subgroups

We propose two mechanisms to enable search over subgroups: the equivariance relaxation morphism and a $[G]$-mixed equivariant layer. The proposed morphism changes the equivariance constraint from one group to another subgroup while preserving the learned weights of the initial group convolutional operator; this is useful for architecture search methods like evolutionary NAS that navigate between discrete architectural choices (Elsken et al., 2017; Lu et al., 2019). The second contribution, the $[G]$-mixed equivariant layer, allows for a single layer to represent equivariance to multiple subgroups through a weighted sum. Learning the weights of such a hypernetwork is common in differentiable NAS approaches, such as DARTS (Liu et al., 2018). We first present a preliminary background on group convolutions and then the theoretical foundations of the two proposed mechanisms.

2.1. Group Convolution Background

We assume familiarity with group theory (Herstein, 2006). Let G be a finite symmetry group. The lth G-equivariant group convolutional layer (Cohen and Welling, 2016) of a group convolutional neural network convolves the feature map $f\colon G \to \mathbb{R}^{C_{l-1}}$ output from the previous layer with a filter with kernel size k represented as learnable parameters $\psi\colon G \to \mathbb{R}^{C_l \times C_{l-1}}$. For each output channel $d \in [C_l]$ where $[C] := \{1, \ldots, C\}$, and group element $g \in G$, the layer's output is defined via the correlation operator:

$$[f \star_G \psi]_d(g) = \sum_{h \in G} \sum_{c=1}^{C_{l-1}} f_c(h) \psi_{d,c}(g^{-1}h). \tag{1}$$

The first layer is a special case: in the case of image data, the input is $x\colon \mathbb{Z}^2 \to \mathbb{R}^{C_0}$, so the layer is instead a *lifting* correlation:

$$[x \star_G \psi]_d(g) = \sum_{y \in \mathbb{Z}^2} \sum_{c=1}^{C_0} x_c(y) \psi_{d,c}(g^{-1}y). \tag{2}$$

Extended Abstract Track

We present our contributions in the group convolutional layer case, although similar claims apply for the lifting convolutional layer case.

2.2. Equivariance Relaxation Morphism

The *equivariance relaxation morphism* reparameterizes a G-equivariant group (or lifting) convolutional layer to operate over a subgroup of G, partially removing weight-sharing constraints from the parameter space while maintaining the functionality of the layer.

Let $G' \leq G$ be a subgroup of G. Let R be a system of representatives of the left quotient (including the neutral element), so that:

$$G' \setminus G = \left\{ G'r \mid r \in R \right\}, \tag{3}$$

$$G'r := \left\{ g'r \mid g' \in G' \right\}. \tag{4}$$

Given a G-equivariant group convolutional layer with input feature map f and kernel ψ, we define the relaxed feature map $\tilde{f} \colon G' \to \mathbb{R}^{(C_{l-1} \times |R|)}$ and relaxed kernel $\tilde{\psi} \colon G' \to \mathbb{R}^{(C_l \times |R|) \times (C_{l-1} \times |R|)}$ as follows. For $c \in [C_{l-1}]$, $s, t \in R$, $d \in [C_l]$:

$$\tilde{f}_{(c,s)}(g') := f_c(g's), \tag{5}$$

$$\tilde{\psi}_{(d,t),(c,s)}(g') := \psi_{d,c}(t^{-1}g's). \tag{6}$$

We define the *equivariance relaxation morphism* from G to G' as the reparameterization of ψ as $\tilde{\psi}$ (Eq. 6) and reshaping of f as $\tilde{f}$ (Eq. 5). The new output of the layer, $[\tilde{f} \star_{G'} \tilde{\psi}]_{(d,t)}(g')$, is equivalent down to reshaping, as shown in the following proof, and thus is function-preserving for the network.

$$t \in R, g \in G \implies G'g = G't \in G' \setminus G \tag{7}$$

$$\implies gt^{-1} \in G' \tag{8}$$

$$\implies \exists g' \in G' \colon g = g't. \tag{9}$$

The mapping $G' \times R \to G$, $(g', t) \mapsto g't$ is bijective. Similarly, $s \in R, h \in G \implies \exists h' \in G' : h = h's$. By such decompositions, we get:

$$[f \star_G \psi]_d(g't) = [f \star_G \psi]_d(g) \tag{10}$$

$$= \sum_{h \in G} \sum_{c=1}^{C_{l-1}} f_c(h) \psi_{d,c}(g^{-1}h), \tag{11}$$

$$= \sum_{h' \in G'} \sum_{s \in R} \sum_{c=1}^{C_{l-1}} f_c(h's) \psi_{d,c}(t^{-1}g'^{-1}h's), \tag{12}$$

$$= \sum_{h' \in G'} \sum_{c=1}^{C_{l-1}} \sum_{s \in R} \tilde{f}_{(c,s)}(h') \tilde{\psi}_{(d,t),(c,s)}(g'^{-1}h'), \tag{13}$$

$$= \left[\tilde{f} \star_{G'} \tilde{\psi} \right]_{(d,t)} (g'). \tag{14}$$

Thus, the correlation of $\tilde{f}$ with $\tilde{\psi}$ is equivariant to G but parametrized as a G'-equivariant group convolutional layer, where the representatives are expanded into independent channels. It can be viewed as an initializing a G'-equivariant layer with a pre-trained prior of equivariance to G, maintaining any previous training.

Standard convolutional layers are a special case of group equivariant layers, where the group is translational symmetry over pixel space. Group convolutional layers are often implemented by relaxation to the translational symmetry group by expanding the kernel via the group actions then using a standard convolutional layer implementation. The equivariance relaxation morphism generalizes this concept to any subgroup.

2.3. $[G]$-Mixed Equivariant Layer

Towards learning equivariance, we additionally propose partial equivariance through a mixture of layers, each constrained to equivariance to different groups and applied in parallel to the same input then all combined via a weighted sum. The equivariance relaxation morphism provides a mapping of group elements between pairs of groups where one is a subgroup of the other. For a set of finite groups $[G]$ where each group is a subgroup or supergroup of all other groups within the set, we define a $[G]$-*mixed equivariant layer* as:

$$\left[f \hat{\star}_{[G]} [\psi]\right]_{(d,t)}(g) = \sum_{G \in [G]} z_G \left[f \star_{G'} \widetilde{\psi^G}\right]_{(d,t)}(g) \tag{15}$$

$$= \left[f \star_{G'} \sum_{G \in [G]} z_G \widetilde{\psi^G}\right]_{(d,t)}(g), \tag{16}$$

where each element z_G of $[z] \coloneqq \{z_G | G \in [G]\}$ is an architectural weighting parameter such that $\sum_{G \in [G]} z_G = 1$, G' is a subgroup of all groups in $[G]$, each element ψ^G of $[\psi]$ is a kernel with a domain of G, and $\widetilde{\psi^G}$ is the transformation of ψ^G from a domain of G to G' as in Equation 6. Thus, the layer is parametrized by $[\psi]$ and $[z]$ and computes a weighted sum of operations each equivariant to a group in $[G]$. The layer may be equivalently computed by convolving the input with the weighted sum of transformed kernels, shown in Equation 16.

3. Conclusion

Towards architectural optimization of partial equivariance based on subgroup decomposition, we propose the equivariance relaxation morphism and the $[G]$-mixed equivariant layer. The equivariance relaxation morphism can be utilized in a simple evolutionary NAS algorithm, such as NASH (Elsken et al., 2017), while the $[G]$-mixed equivariant layer can be optimized with a differentiable NAS algorithm, such as DARTS (Liu et al., 2018). For the latter case, note that architectural discretization and parameter retraining is not necessary, as the mixture of kernels can be combined into a single kernel (Eq. 16). As such, the $[G]$-mixed equivariant layer could be directly used as a layer in standard network training, beyond NAS algorithms; the equivariance relaxation morphism could also be extended to applications beyond architecture search such as fine-tuning and distillation. By demonstrating that groups can be relaxed or mixed in a single layer, we aim to facilitate the learning of networks which automatically find optimal partial equivariance.

Extended Abstract Track

References

Devanshu Agrawal and James Ostrowski. A classification of g-invariant shallow neural networks. *Advances in Neural Information Processing Systems*, 35, 2022.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999. PMLR, 2016.

Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Simple and efficient architecture search for convolutional neural networks. In *NIPS Workshop on Meta-Learning*, 2017.

Marc Finzi, Gregory Benton, and Andrew G Wilson. Residual pathway priors for soft equivariance constraints. *Advances in Neural Information Processing Systems*, 34:30037–30049, 2021.

Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets*, pages 267–285. Springer, 1982.

Israel N Herstein. *Topics in algebra*. John Wiley & Sons, 2006.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pages 2747–2755. PMLR, 2018.

Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

Hanxiao Liu, Karen Simonyan, and Yiming Yang. DARTS: differentiable architecture search. In *International Conference on Learning Representations*, 2018.

Zhichao Lu, Ian Whalen, Vishnu Boddeti, Yashesh Dhebar, Kalyanmoy Deb, Erik Goodman, and Wolfgang Banzhaf. NSGA-Net: neural architecture search using multi-objective genetic algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 419–427, 2019.

Kaitlin Maile, Hervé Luga, and Dennis George Wilson. Structural learning in artificial neural networks: A neural operator perspective. *Transactions on Machine Learning Research*, 2022.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. In *International Conference on Machine Learning*, pages 2892–2901. PMLR, 2017.

David W Romero and Suhas Lohit. Learning partial equivariances from data. *Advances in Neural Information Processing Systems*, 35, 2022.

Tycho FA van der Ouderaa, David W Romero, and Mark van der Wilk. Relaxing equivariance constraints with non-stationary continuous filters. *Advances in Neural Information Processing Systems*, 35, 2022.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *International Conference on Machine Learning*. PMLR, 2022.

Tao Wei, Changhu Wang, Yong Rui, and Chang Wen Chen. Network morphism. In *International Conference on Machine Learning*, pages 564–572. PMLR, 2016.

Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. *Advances in Neural Information Processing Systems*, 32, 2019.

Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. Coordinate Independent Convolutional Networks - Isometry and Gauge Equivariant Convolutions on Riemannian Manifolds. *arXiv preprint arXiv:2106.06020*, 2021.

Raymond A Yeh, Yuan-Ting Hu, Mark Hasegawa-Johnson, and Alexander Schwing. Equivariance discovery by learned parameter-sharing. In *International Conference on Artificial Intelligence and Statistics*, pages 1527–1545. PMLR, 2022.

Allan Zhou, Tom Knowles, and Chelsea Finn. Meta-learning symmetries by reparameterization. In *International Conference on Learning Representations*, 2020.