Extended Abstract Track

Equivariant Representations for Non-Free Group Actions

Editors: List of editors' names

Abstract

We introduce a method for learning representations that are equivariant with respect to general group actions over data. Differently from existing equivariant representation learners, our method is suitable for actions that are not free i.e., that stabilize data via nontrivial symmetries. Our method is grounded in the orbit-stabilizer theorem from group theory, which guarantees that an ideal learner infers an isomorphic representation. Finally, we provide an empirical investigation on image datasets with rotational symmetries and show that taking stabilizers into account improves the quality of the representations.

Keywords: Representation Learning, Symmetries, Group Theory

1. Introduction

The problem of incorporating symmetries into representations defines a fundamental challenge and has been considered in a number of recent works (Quessard et al. (2020); Higgins et al. (2022); Cohen and Welling (2014); Tonnaer et al. (2022); Ahuja et al. (2021)). The overall aim is to design representations which preserve symmetries – a property known as *equivariance*. This is because preservation of symmetries leads to the extraction of geometric and semantic structure in data, which can be exploited for reasoning, efficiency and generalization (Bengio et al. (2013)). As an example, the challenge of *disentangling* semantic factors of variations has been rephrased in terms of equivariant representations (Higgins et al. (2018); Caselles-Dupré et al. (2019)).

The majority of literature relies on the assumption that the group of symmetries acts *freely* on data (Marchetti et al. (2022)) i.e., that no datapoint is stabilized by (nontrivial) symmetries. This avoids the need to model stabilizers, which are unknown subgroups of the symmetry group considered. However, non-free group actions arise in several practical scenarios. This happens for example when considering images of objects acted upon by the rotation group via change of orientation. Such objects might be symmetrical, resulting in rotations leaving the image (almost) identical and consequently ambiguous in its orientation (see Figure 1).

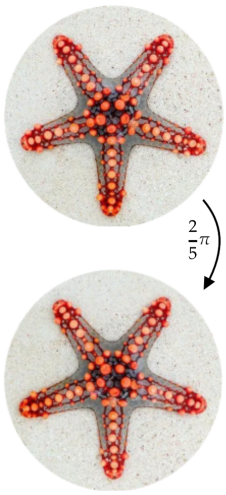

Figure 1

In this work we propose a method for learning equivariant representation for general and potentially non-free group actions. Based on the orbit-stabilizer theorem from group theory, we design a model that outputs cosets of the stabilizer subgroup. The representation learner optimizes an equivariance loss based on supervision from symmetries alone. The above-mentioned theoretical results guarantee that a learner infers representations that are isomorphic to the original dataset.

© 2022 .

2. Group Theory Background

Let G be the group of symmetries with multiplication denoted by $(g, h) \to gh$ and identity denoted by $1 \in G$. Suppose that G acts on a set $\mathcal{X}$ via $(g, x) \to g \cdot x$. The action defines a set of *orbits* $\mathcal{X}/G$ given by the equivalence classes of the relation $x \sim y$ iff $y = g \cdot x$ for some $g \in G$. For each $x \in \mathcal{X}$, the *stabilizer* subgroup is defined as $G_x = \{g \in G \mid g \cdot x = x\}$. Stabilizers are conjugate as x varies in its orbit, and by abuse of notation we refer to the conjugacy class G_O for $O \in \mathcal{X}/G$. The action is said to be *free* if $G_O = \{1\}$ for every O.

Recall that a map $\varphi : \mathcal{X} \to \mathcal{Z}$ between sets acted upon by G is said to be *equivariant* if $\varphi(g \cdot x) = g \cdot \varphi(x)$ for every $x \in \mathcal{X}$ and $g \in G$. An equivariant bijection is referred to as *isomorphism*. The following is the fundamental result on group actions (Rotman (2012)).

Theorem 1 (Orbit-Stabilizer) *Each orbit O is isomorphic to the set of (left)* cosets $G/G_O = \{gG_O \mid g \in G\}$ *(on which G acts by left multiplication). In other words, there is an isomorphism:*

$$\mathcal{X} \simeq \coprod_{O \in \mathcal{X}/G} G/G_O \quad \subseteq 2^G \times \mathcal{X}/G \tag{1}$$

where 2^G denotes the power-set of G on which G acts by left multiplication $g \cdot A = \{ga \mid a \in A\}$. Moreover, any equivariant map $\varphi : \mathcal{X} \to \coprod_{O \in \mathcal{X}/G} G/G_O$ which induces a bijection on orbits is an isomorphism.

3. Equivariant Representation Learning

Our goal is to design an equivariant representation learner based on Theorem 1. We aim to train a model $\varphi : \mathcal{X} \to \mathcal{Z}$ with a latent space $\mathcal{Z}$ on a loss encouraging equivariance. We assume that G is known a priori but its action on $\mathcal{X}$ is not. Theoretically the ideal choice for $\mathcal{Z}$ is given by the right-hand side of the isomorphism in Equation 1. In other words φ has to output cosets of stabilizers. However, the core challenge is that the stabilizers are unknown a priori since they depend on the group action. In order to circumvent the modeling of stabilizers and their cosets, we appeal to the following simple result (a proof is provided in the Appendix):

Proposition 2 *Let $\varphi : \mathcal{X} \to 2^G$ be an equivariant map. Then for each $x \in \mathcal{X}$ from an orbit O, $\varphi(x)$ contains a coset of (a conjugate of) G_O.*

Proposition 2 enables φ to output subsets of G instead of cosets of stabilizers. As long as those subsets are *minimal* w.r.t. to inclusion, they will coincide with the desired cosets. Based on this, we define the latent space as $\mathcal{Z} = \mathcal{Z}_G \times \mathcal{Z}_O$ and implement the map φ as a pair of neural networks $\varphi_G : \mathcal{X} \to \mathcal{Z}_G$, $\varphi_O : \mathcal{X} \to \mathcal{Z}_O$. The component $\mathcal{Z}_G$ represents cosets of stabilizers while $\mathcal{Z}_O$ represents orbits. Since the output space of a neural network is a finite-dimensional vector space, we assume that G is a linear Lie group ,i.e. G is a manifold of matrices, and that the stabilizers of the action are finite. The model φ_G first outputs N elements $(\varphi_G^1(x), \cdots, \varphi_G^N(x)) = \varphi_G(x)$ in the matrix Lie algebra $\mathfrak{g}$ that are converted to G by the exponential map $\exp : \mathfrak{g} \to G$. The hyper-parameter N ideally should be chosen to be larger than the cardinality of the stabilizers. On the other hand, the output of φ_O consists of a vector of arbitrary dimensionality. The only requirement is that the output space of φ_O should have enough capacity to contain $\mathcal{X}/G$.

Extended Abstract Track

Our dataset $\mathcal{D}$ consists of samples from the group action, meaning that datapoints are triplets $(x, g, y) \in \mathcal{X} \times G \times \mathcal{X}$ with $y = g \cdot x$. Given a datapoint $(x, g, y) \in \mathcal{D}$ the learner φ_G optimizes the equivariance loss over its parameters:

$$\mathcal{L}_G(x, g, y) = d(g \cdot \varphi_G(x),\ \varphi_G(y)) \tag{2}$$

where d is a (semi) metric for sets. We opt for the (asymmetric) *Chamfer distance* $d(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} d_G(a, b)$ because of its differentiability properties. Here d_G is a metric on G and is typically set as the (squared) Euclidean one for $G = \mathbb{R}^n$ and as the (squared) Frobenius one for $G = \mathrm{SO}(n)$. As previously discussed, we wish $\varphi_G(x)$, when seen as a set, to be minimal in cardinality. To this end we add the following regularization term which measures the discrete entropy discouraging discrete entropy:

$$\widetilde{\mathcal{L}}_G(x) = \frac{\lambda}{N^2} \sum_{1 \leq i,j \leq N} d_G(\varphi_G^i(x), \varphi_G^j(x)) \tag{3}$$

where λ is a small weight set canonically to 0.001. On the other hand, since orbits are invariant to the group action φ_O optimizes a *contrastive loss*. We opt for the popular InfoNCE loss from the literature (Chen et al. (2020)) i.e.,

$$\mathcal{L}_O(x, y) = d_O(\varphi_O(x), \varphi_O(y)) + \log \mathbb{E}_{x'} \left[e^{-d_O(\varphi_O(x'), \varphi_O(x))} \right] \tag{4}$$

where x' is marginalized from $\mathcal{D}$. As customary for the InfoNCE loss, we normalize the output of φ_O and set $d_O(a, b) = -\cos(\angle ab) = -a \cdot b$. The denominator of Equation 4 encourages injectivity of φ_O as required by the second part of Theorem 1. This together with Proposition 2 guarantees that an ideal learner infers an isomorphic representation.

4. Experiments

We test an implementation of the neural networks φ_G and φ_O on the following two datasets consisting of images with a resolution of 64×64 pixels subject to non-free group actions:

Rotating Arrows: each image depicts a radial configuration of $\nu \in \{4, 5, 6\}$ arrows pointing outside. The group of two-dimensional rotations $G = \mathrm{SO}(2)$ acts by rotating the configuration. The number of arrows ν determines the orbit and the stabilizer. The latter is the cyclic group $C_\nu \subseteq G$ of order ν.

Double Arrows: each image consists of two radial configurations of arrows with 2 and 3 arrows respectively placed in two corners of the image. The group of symmetries is a torus $G = \mathrm{SO}(2) \times \mathrm{SO}(2)$. The action produces a single orbit, i.e. it is *transitive*, and the stabilizer is a product of cyclic groups $C_2 \times C_3$.

Tetrahedron: this dataset consists of images of a monochromatic tetrahedron in different orientations (Murphy et al. (2021)). The group of symmetries is $G = \mathrm{SO}(3)$ with a single orbit and the stabilizer is the alternating group A_4 of order 12.

We compare our model with the *baseline* in which the model φ_G produces a single output i.e., $N = 1$ and $\varphi_G(x) = \varphi_G^1(x) \in G$. In other words, the latent space is $\mathcal{Z} = G \times \mathcal{Z}_O$, on which G acts freely. This baseline is similar to what has been proposed in previous work assuming free group actions (Quessard et al., 2020; Caselles-Dupré et al., 2019; Marchetti et al., 2022; Tonnaer et al., 2022).

The models are compared based on two evaluation metrics. The first is the equivariance loss, see Equation 3, on a test set. The second is the reconstruction loss on a test set of a decoder $\psi : \mathcal{Z} \to \mathcal{X}$ trained jointly with φ. The loss is given by $\mathbb{E}_x[\, d_{\mathrm{bc}}(x, \psi(\varphi(x)) \,]$ where d_{bc} is the pixel-wise binary cross-entropy.

Quantitative results are presented in Table 1 while qualitative visualizations are shown in Figure 2. As can be seen, our model achieves equivariance and reconstruction errors that are significantly smaller than the baseline. The latter is not able to capture the stabilizers in its latent space, leading to representations of poor quality and loss of information.

Table 1: Mean and std over 3 runs of the evaluation metrics for our model and the baseline.

Dataset	Model	N	Equivariance	Reconstruction
Rotating Arrows	Baseline	1	$1.985_{\pm 0.027}$	$0.551_{\pm 0.051}$
	Ours	10	$0.011_{\pm 0.002}$	$0.372_{\pm 0.040}$
Double Arrows	Baseline	1	$4.016_{\pm 0.027}$	$0.219_{\pm 0.002}$
	Ours	6	$0.009_{\pm 0.006}$	$0.152_{\pm 0.011}$
Tetrahedron	Baseline	1	$6.025_{\pm 0.063}$	$0.470_{\pm 0.045}$
	Ours	20	$0.032_{\pm 0.007}$	$0.286_{\pm 0.029}$

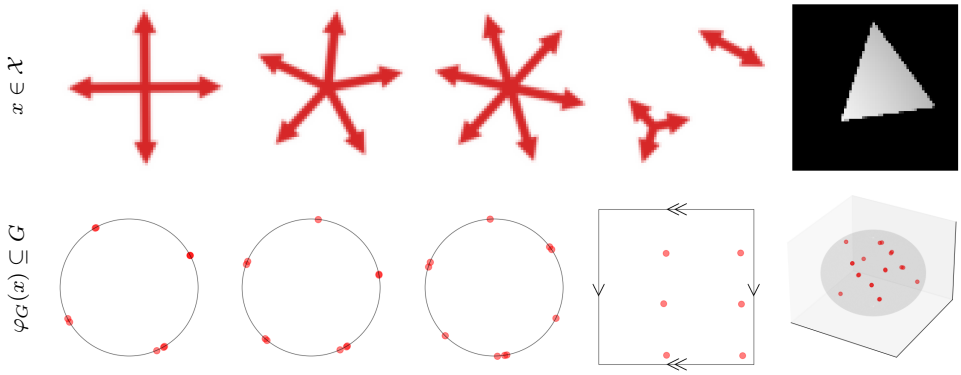

Figure 2: Visualization of the data x and the predicted stabilizer $\varphi_G(x)$, some components of which get overlapped as expected. In the double arrows, the stabilizer $C_2 \times C_3 \subseteq \mathrm{SO}(2) \times \mathrm{SO}(2)$ is visualized on a torus seen as an identified a square. For the tetrahedron, the stabilizer A_4 is visualized as rotational vectors in the projective space $\mathbb{RP}^3 \simeq \mathrm{SO}(3)$.

5. Conclusion and Future Work

In this work we introduced a method for learning equivariant representations for general and potentially non-free group actions. We discussed the theoretical foundations and empirically investigated the method on images with rotational symmetries.

Our model relies on the assumptions that the stabilizers are finite. However, non-discrete stabilizer subgroups sometimes occur, for example in the case of symmetrical objects such as a cone or a cylinder. An interesting future direction is designing an equivariant representation learner suitable for group actions with non-discrete stabilizers.

References

Kartik Ahuja, Jason Hartford, and Yoshua Bengio. Properties from Mechanisms: An Equivariance Perspective on Identifiable Representation Learning. 2021.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. ISSN 01628828. doi: 10.1109/TPAMI.2013.50.

Hugo Caselles-Dupré, Michael Garcia-Ortiz, and David Filliat. Symmetry-Based Disentangled Representation Learning Requires Interaction with Environments. *Advances in Neural Information Processing Systems*, 32(NeurIPS):1–10, 2019. ISSN 10495258.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

Taco Cohen and Max Welling. Learning the Irreducible Representations of Commutative Lie Groups. *International Conference on Machine Learning*, 5:3757–3770, 2014.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 37, pages 770–778. IEEE, 6 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. URL https://onlinelibrary.wiley.com/doi/10.1002/chin.200650130http://ieeexplore.ieee.org/document/7780459/.

Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a Definition of Disentangled Representations. pages 1–29, 2018. URL http://arxiv.org/abs/1812.02230.

Irina Higgins, Sébastien Racanière, and Danilo Rezende. Symmetry-Based Representations for Artificial and Biological General Intelligence. *Frontiers in Computational Neuroscience*, 16, 4 2022. ISSN 1662-5188. doi: 10.3389/fncom.2022.836498. URL https://www.frontiersin.org/articles/10.3389/fncom.2022.836498/full.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2 2019.

Giovanni Luca Marchetti, Gustaf Tegnér, Anastasiia Varava, and Danica Kragic. Equivariant Representation Learning via Class-Pose Decomposition. 7 2022. doi: 10.48550/arxiv.2207.03116. URL https://arxiv.org/abs/2207.03116v2.

Kieran Murphy, Carlos Esteves, Varun Jampani, Srikumar Ramalingam, and Ameesh Makadia. Implicit representation of probability distributions on the rotation manifold. In *International Conference on Machine Learning*, 2021.

Robin Quessard, Thomas D Barrett, and William R Clements. Learning Disentangled Representations and Group Structure of Dynamical Environments. In *34th Conference on Neural Information Processing Systems*, 2020.

Joseph J Rotman. *An introduction to the theory of groups*, volume 148. Springer Science & Business Media, 2012.

Loek Tonnaer, Luis Armando Perez Rey, Vlado Menkovski, Mike Holenderski, and Jim Portegies. Quantifying and Learning Linear Symmetry-Based Disentanglement. In *Proceedings of the 39th International Conference on Machine Learning*, pages 21584–21608. PMLR, 6 2022. URL https://proceedings.mlr.press/v162/tonnaer22a.html.

Extended Abstract Track

6. Appendix

6.1. Proofs of Theoretical Results

Proposition 6.1:
Let $\varphi : \mathcal{X} \to 2^G$ be an equivariant map. Then for each $x \in \mathcal{X}$ from an orbit O, $\varphi(x)$ contains a coset of (a conjugate of) G_O.

Proof Pick $x \in \mathcal{X}$. Then for every $g \in G_x$ it holds that $\varphi(x) = \varphi(g \cdot x) = g \cdot \varphi(x)$. In other words $G_x h = hh^{-1}G_x h \subseteq \varphi(x)$ for each $h \in \varphi(x)$. Since $h^{-1}G_x h$ is conjugate to G_x the thesis follows. ■

6.2. Training Details

We implement the neural networks φ_G and φ_O with a backbone ResNet18 (He et al. (2016)). For a datapoint $x \in \mathcal{X}$, the network implements multiple heads to produce embeddings $\left(\varphi_G^1(x), \cdots, \varphi_G^N(x)\right)$ with $\varphi_G^i(x) \in G$. The output dimension of φ_O is set to 3. We train each experiment for 50 epochs using the AdamW optimizer (Loshchilov and Hutter (2019)) with a learning rate of 10^{-4} and batches of 16 triplets $(x, g, y) \in \mathcal{D}$.

The rotating arrows dataset consists of 5000 datapoints per orbit while the double arrows and the tetrahedron datasets consist of 20000 datapoints. For all datasets the test set consists of a random 10% split.

6.3. Reconstructions

We present in Figure 3 some examples of images reconstructed by a decoder ψ trained jointly with the encoder φ. The baseline model is not capable of clearly reconstructing the images compared to our model.

As seen from the quantitative results in Table 1, the baseline model is not capable of optimizing the equivariance loss from Equation 2. This provides a hint that the encoder φ might not be converging to a stable representation. Consequently, the decoder is incapable of consistently reconstructing the data which results in a higher reconstruction loss.

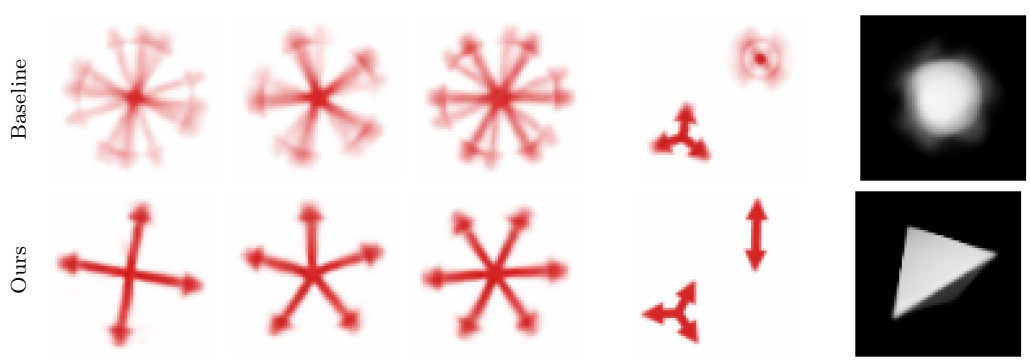

Figure 3: Reconstruction examples produced by a decoder trained on the latent variables produced by our model and the baseline.