A Characterization Theorem for Equivariant Networks with Point-wise Activations

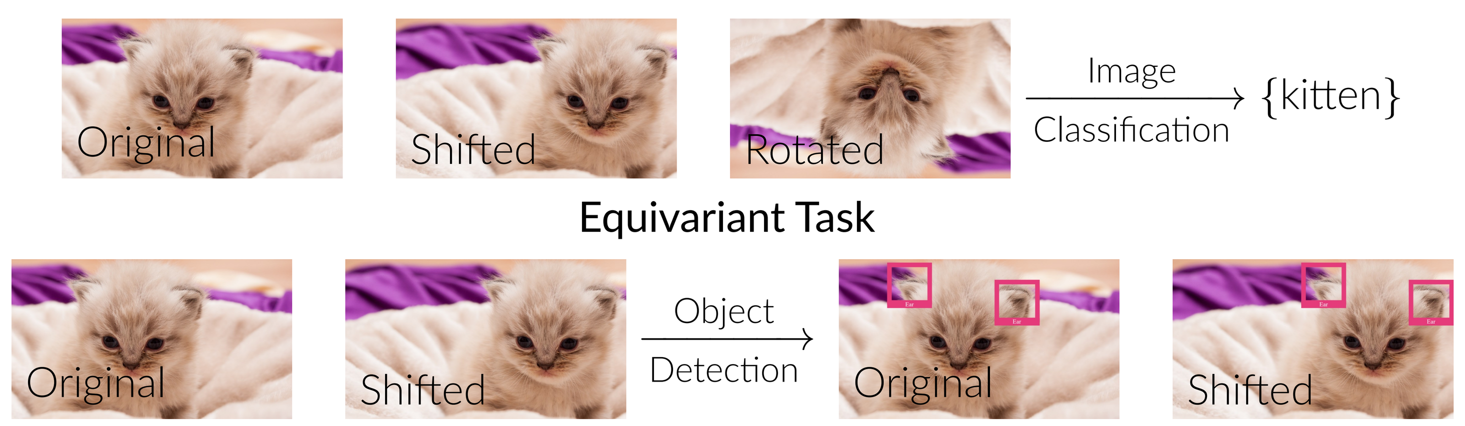
Marco Pacini [1, 2], Xiaowen Dong [3], Bruno Lepri [2], Gabriele Santin [4]

[1] University of Trento, [2] Fondazione Bruno Kessler, [3] University of Oxford, [4] University of Venice

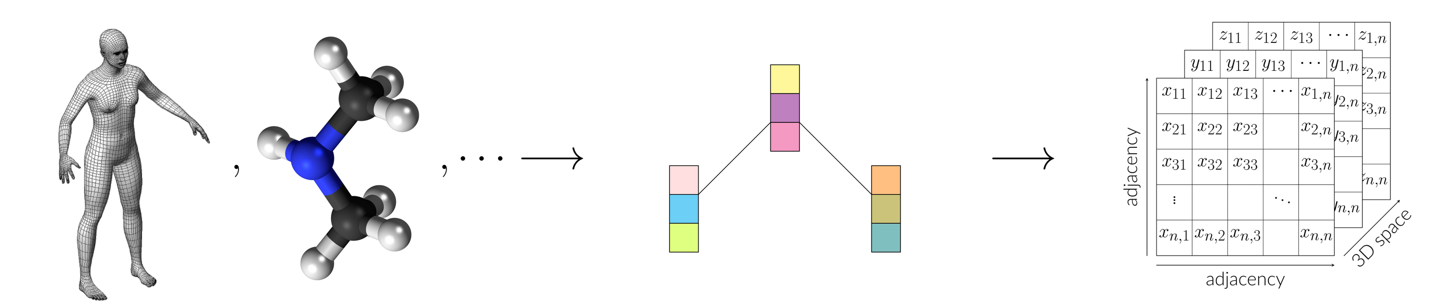
↓ paper link ↓

Basics of Geometric Deep Learning

Symmetries & Machine Learning

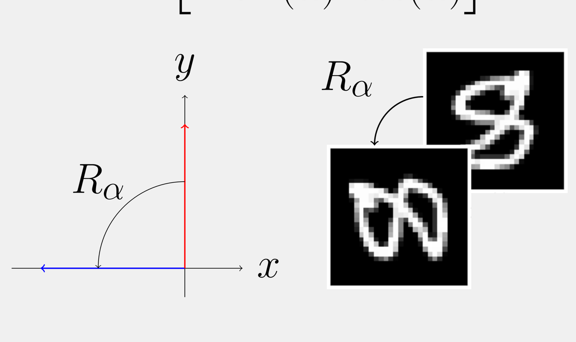


Representations of Geometric Data & Symmetries

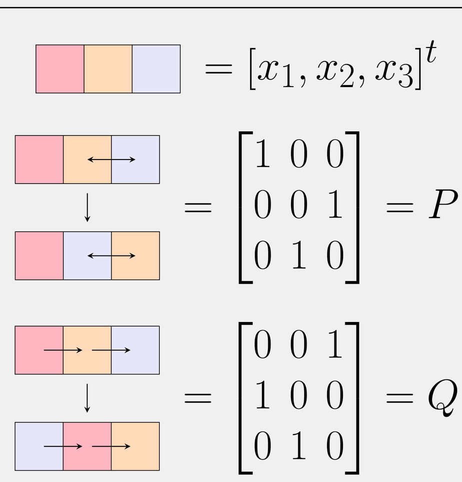


2D Rotations

$$R_\alpha = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

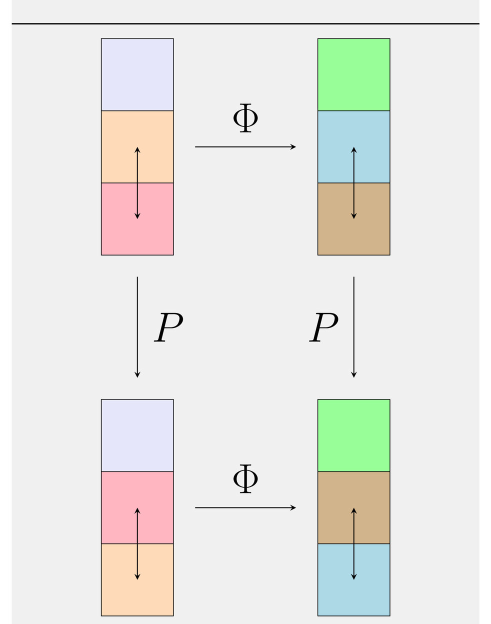


Permutations



Equivariant Neural Networks

Idea



Equivariance

A feature map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is **equivariant** with respect to P if

$$P \circ \Phi = \Phi \circ P$$

Equivariant Neural Networks

Composition

$$\Phi = \phi_m \circ \tilde{f}_{m-1} \circ \phi_{m-1} \circ \cdots \circ \tilde{f}_1 \circ \phi_0$$

of equivariant **linearities** ϕ_i and equivariant **activations** $\tilde{f}_i$

Breaking Symmetry

Point-wise Activations

$$\tilde{f}(x_1, \ldots, x_n) = (f(x_1), \ldots, f(x_n))$$

e.g. $f = \mathrm{ReLU}, \mathrm{sigmoid}, \tanh, \ldots$

ReLU **is equivariant** with respect to **permutation** representations

ReLU **is not equivariant** with respect to **rotations**. Indeed ↓

E.g.:

$$\mathrm{ReLU} \circ R_{\frac{\pi}{2}}\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \mathrm{ReLU}\left(\begin{bmatrix}-1\\0\end{bmatrix}\right) = \begin{bmatrix}0\\0\end{bmatrix}$$

$$=\qquad\qquad\qquad\qquad \neq$$

$$R_{\frac{\pi}{2}} \circ \mathrm{ReLU}\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = R_{\frac{\pi}{2}}\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}-1\\0\end{bmatrix}$$

Research Question

- Given **specific symmetries**, need to guess **viable activations** to design **equivariant layers**?
- Or manage to **completely describe** the following correspondence?

[?] (Symmetries P, Functions f) $\longleftrightarrow$ Equivariant Activations $\tilde{f}$ [?]

Main Theorem

T-Monomial Matrices

A T-**monomial matrix** has only one non-zero element for each column and row and it lies in T. Where $T \subseteq \mathbb{R}$ is a multiplicative group.

E.g.:

T = $\langle 1 \rangle$	T = $\langle \pm 1 \rangle$	T = $\mathbb{R}_{>0}$
$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}$	$\begin{bmatrix}0&0&1\\1&0&0\\0&-1&0\end{bmatrix}$	$\begin{bmatrix}0&0&1\\\sqrt{2}&0&0\\0&\frac{1}{3}&0\end{bmatrix}$

T-Equivariant Functions

A function $f : \mathbb{R} \to \mathbb{R}$ is T-**equivariant** if

$$f(tx) = tf(x)$$

for each $x \in \mathbb{R}$ and $t \in T$.

E.g.:

$T = \langle 1 \rangle$	$T = \langle \pm 1 \rangle$	$T = \mathbb{R}_{>0}$
Each f	Each odd f	ReLU LeakyReLU ...

Claim 1 - *Informal*

Assume f to be **non-affine** and **continuous**. Then

(T-monomial P, T-equivariant f) $\Longleftrightarrow$ Equivariant Activations $\tilde{f}$

Claim 2 - *Informal*

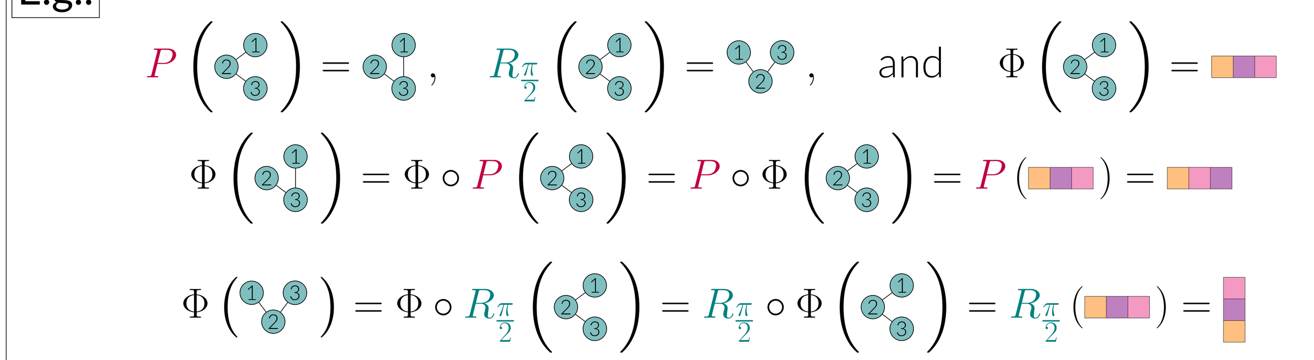
Let P belong to a **compact** set of symmetries. Then, up to coordinate change, P is a **signed-permutation matrix**. If f is not odd, P is a **permutation matrix**.

Relevant Implications

Geometric Invariant Graph Networks (IGNs)

A **Geometric IGN** is a neural network $\Phi : (\mathbb{R}^n)^{\otimes 2} \otimes \mathbb{R}^3 \to \mathcal{Y}$ simultaneously equivariant with respect to **permutations** P and **rotations** R_α.

E.g.:



Relevant Corollaries

Non-existence Results

Geometric IGNs with non-affine point-wise activations are **trivial**.

A common mitigation strategy is to require equivariance on only a finite subset of symmetries. This approach leads to networks which are

- **approximately equivariant** with respect to the entire group of symmetries
- **non-trivial** and completely described by the following **Corollary** of Claim 2 ↓

Finite Groups of Symmetries

Let G be a **finite** set of symmetries and suppose f non-odd non-affine. Then each hidden representation is of the form

$$\mathbb{R}^{G/H_1} \times \cdots \times \mathbb{R}^{G/H_m}$$

for a finite number of symmetry subgroups $H_i < G$.

Future Directions

- **Robustness of approximate equivariance**, i.e., satisfying equivariance on a finite subset of symmetries implies bounds on $\|P \circ \Phi - \Phi \circ P\|$. Experimental results are positive, no theoretical results are known to the authors.
- Apply the proposed approaches to models with activations **beyond real point-wise** ones. **E.g.: (i)** complex neural networks with complex point-wise activations similar to Harmonic Networks [2] or **(ii)** the more general Steerable CNNs [1].

References

[1] T. Cohen and M. Welling. Steerable cnns. *ICLR*, 2017.

[2] D. E. Worrall et al. Harmonic networks: Deep translation and rotation equivariance. *CVPR*, 2017.

[3] H. Maron et al. Invariant and equivariant graph networks. *ICLR*, 2019.

[4] J. Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete applied mathematics*, 1996.

X @MarcoPacini4 | marco-pacini.github.io | mpacini@fbk.eu