On Universality of Deep Equivariant Networks

Marco Pacini [1, 2], Mircea Petrache [3, 4], Bruno Lepri [2], Shubhendu Trivedi, Robin Walters [5]

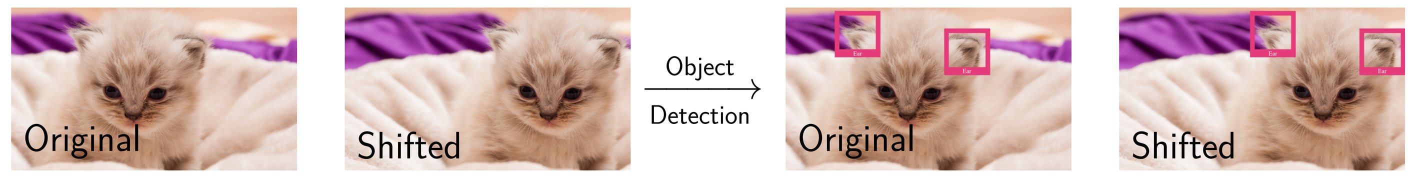
[1] University of Trento, [2] Fondazione Bruno Kessler, [3] PUC Chile, [4] CenIA, [5] Northeastern University

Basics of Geometric Deep Learning

Symmetries & Machine Learning

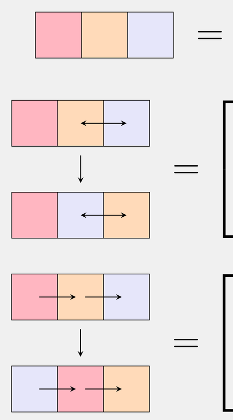
Invariant Task

Equivariant Task



Permutations, Symmetries & Equivariance

Permutations



$= [x_1, x_2, x_3]^t$

$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = P$

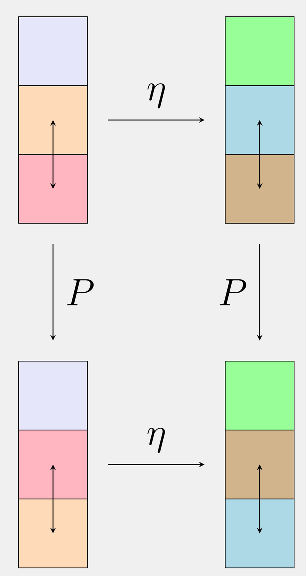
$= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = Q$

Equivariance

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

$$P \circ \eta = \eta \circ P$$

Diagram



Point-wise Activations

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

E.g.: $\sigma = \mathrm{ReLU}, \mathrm{sigmoid}, \tanh, \ldots$

Equivariant Neural Networks

Equivariant **linearities** ϕ_i and equivariant **activation** $\tilde{\sigma}$:

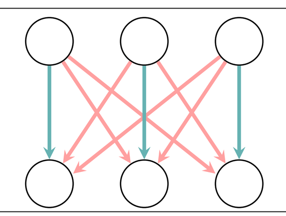
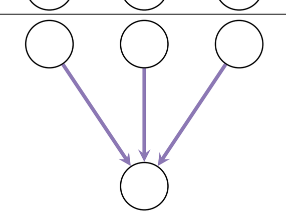
$$\eta := \phi_m \circ \tilde{\sigma} \circ \phi_{m-1} \circ \cdots \circ \tilde{\sigma} \circ \phi_0$$

Spaces of Equivariant Neural Networks

$$\mathcal{N} = \mathcal{N}_\sigma(\mathbb{R}^{X_0} \ldots, \mathbb{R}^{X_d}) := \left\{ \phi_m \circ \tilde{\sigma} \circ \phi_{m-1} \circ \cdots \circ \tilde{\sigma} \circ \phi_0 \mid \phi_i \in \mathrm{Aff}_G(\mathbb{R}^{X_{i-1}}, \mathbb{R}^{X_i}) \right\}$$

Example — PointNet

Consider $S_n \curvearrowright \mathbb{R}^{[n]} \cong \mathbb{R}^n$. The space of shallow PointNets is $\phi \circ \tilde{\sigma} \circ \psi \in \mathcal{N}_\sigma(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R})$.

Layer spaces	Matrices	Weight-sharing Schemes
$\psi \in \mathrm{Aff}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)$	$\begin{bmatrix} \lambda & \mu & \cdots & \mu \\ \mu & \lambda & \cdots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \cdots & \lambda \end{bmatrix}$	
$\phi \in \mathrm{Aff}_{S_n}(\mathbb{R}^n, \mathbb{R})$	$\begin{bmatrix} \nu & \nu & \cdots & \nu \end{bmatrix}$	

Issue: Separation Constraints Universality

GNNs, or more generally **IGNs**, **fail** to approximate **continuous equivariant functions**.

For k-WL theory, if $\mathcal{N} = \{1 - \text{WL} \sim \text{GNN}s\}$:

$\forall \eta \in \mathcal{N},\ \eta(\ ) = \eta(\ )$

Thus, $\mathcal{N}$ **fails to approximate** equivariant functions $\tilde{\eta}$ such that:

$\tilde{\eta}(\ ) \neq \tilde{\eta}(\ )$

Research Question — Informal

Is $\mathcal{N}$ universal in continuous functions with its separation constraint?

Towards a Formalization — Separation

Let $\mathcal{N} \subseteq \mathcal{C}(V)$, define:

- **Separation relation:** $\rho = \rho(\mathcal{N}) := \left\{ (x, y) \in V \times V \mid \eta(x) = \eta(y)\ \forall \eta \in \mathcal{N} \right\}$
- **Separation-constrained functions:** $\mathcal{C}_\rho(V) := \left\{ f \in \mathcal{C}(V) \mid f(x) = f(y)\ \forall (x, y) \in \rho \right\}$

Research Question — formal: $\overline{\mathcal{N}} \stackrel{?}{=} \mathcal{C}_\rho(V)$

On Universality

Standard Neural Networks (Pinkus, 1999)

The set of shallow neural networks with variable width can be written as

$$\mathcal{N} := \bigcup_{h \in \mathbb{N}} \mathcal{N}_\sigma(\mathbb{R}^m, \mathbb{R}^h, \mathbb{R}) \quad \text{and} \quad \rho(\mathcal{N}) = \{(x, y) \in V^2 \mid x = y\}.$$

We define the **universality class** associated with $\mathbb{R}^m, \mathbb{R}, \mathbb{R}$ as

$$\mathcal{U}_\sigma(\mathbb{R}^m, \mathbb{R}, \mathbb{R}) := \overline{\mathcal{N}} \subseteq \mathcal{C}(\mathbb{R}^n) = \mathcal{C}_\rho(\mathbb{R}^n).$$

Then,

$$\sigma \text{ is non-polynomial} \iff \mathcal{U}(\mathbb{R}^m, \mathbb{R}, \mathbb{R}) = \mathcal{C}(\mathbb{R}^m).$$

Invariant Symmetrization (Ravanbakhsh, 2020)

The set of shallow neural networks with variable width can be written as

$$\mathcal{N} := \bigcup_{h \in \mathbb{N}} \mathcal{N}_\sigma(V, \mathbb{R}^{h \times G}, \mathbb{R}) \quad \text{and} \quad \rho(\mathcal{N}) = \{(x, y) \in V^2 \mid \mathrm{Orb}_G(x) = \mathrm{Orb}_G(y)\}.$$

We define the universality class:

$$\mathcal{U}(V, \mathbb{R}^G, \mathbb{R}) := \overline{\mathcal{N}} \subseteq \mathcal{C}_G(V, \mathbb{R}).$$

Then,

$$\sigma \text{ is non-polynomial} \iff \mathcal{U}(V, \mathbb{R}^G, Z) = \mathcal{C}_G(V, Z).$$

Definition — Universality Class

The universality class associated with $\mathbb{R}^{X_0}, \mathbb{R}^{X_1}, \ldots, \mathbb{R}^{X_d}$ is

$$\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \mathbb{R}^{X_1}, \ldots, \mathbb{R}^{X_d}) := \overline{\bigcup_{h_1, \ldots, h_{d-1} \in \mathbb{N}} \mathcal{N}_\sigma(\mathbb{R}^{X_0}, \mathbb{R}^{h_1 \times X_1}, \ldots, \mathbb{R}^{h_{d-1} \times X_{d-1}}, \mathbb{R}^{X_d})}.$$

References

[1] C. K. Joshi et al. On the Expressive Power of Geometric Graph Neural Networks. *ICLR*, 2023.

[2] H. Maron; H. Ben-Hamu; H. Serviansky; Y. Lipman. Provably Powerful Graph Networks. *ICLR*, 2019.

[3] M. Pacini; X. Dong; B. Lepri; G. Santin. A Characterization Theorem for Equivariant Networks with Point-wise Activations. *ICLR*, 2024.

[4] M. Pacini; X. Dong; B. Lepri; G. Santin. Separation Power of Equivariant Neural Networks. *ICLR*, 2025.

[5] M. Pacini; G. Santin; B. Lepri; S. Trivedi. On Universality Classes of Equivariant Networks. *NeurIPS*, 2025.

Results

Theorem (Invariant Universality)

Set $\rho = \rho\big(\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_d}, \mathbb{R})\big)$. *Then*

$$\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_d}, \mathbb{R}, \mathbb{R}) = \mathcal{C}_\rho(\mathbb{R}^{X_0}).$$

Example — Universality of Invariant PointNet

Three layer invariant PointNets are universal, namely, $\mathcal{U}(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R}^n, \mathbb{R}) = \mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R})$.

Entry-wise Separation

Let $X = \{x_1, \ldots, x_\ell\}$, $\mathcal{N} \subseteq \mathcal{C}(V, \mathbb{R}^X)$, for each $x_i \in X$, let

$$\pi_{x_i} : \begin{array}{c} \mathbb{R}^X \to \mathbb{R} \\ (a_{x_1}, \ldots, a_{x_\ell}) \mapsto a_{x_i} \end{array} \quad \text{and} \quad \pi_{i*}\mathcal{N} := \{\pi_i \circ f \mid f \in \mathcal{N}\},$$

We define the **entry-wise separation relation** $\bar{\rho}(\mathcal{N}) := \{\rho(\pi_{i*}\mathcal{N})\}_{i \in X}$.

Example

Consider the set of all $f \in \mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)$. We can rewrite each of them as

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n)),$$

where $f_i = \pi_{i*} f$ for each $i = 1, \ldots, n$. For each permutation σ such that $\sigma(1) = 1$,

$$\sigma f(x_1, \ldots, x_n) = f(\sigma(x_1, \ldots, x_n)) = f\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right) = f\left(x_1, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\right).$$

Hence f_1 is $\mathrm{Stab}_{S_n}(1)$-invariant. Namely,

$$(x, y) \in \rho(\pi_{1*}\mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)) \iff \exists \sigma \in \mathrm{Stab}_{S_n}(1) \quad x = \sigma(y).$$

Repeating this argument for $i = 2, \ldots, n$, we obtain

$$\bar{\rho}\left(\mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)\right) = \{\rho(\pi_{i*}\mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R}^n))\}_{i = 1, \ldots, n}.$$

Depth Stabilization

There exist M such that for each $N \geq M$:

$$\bar{\rho}(\mathcal{N}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_{h-1}}, \underbrace{\mathbb{R}^{X_h}, \ldots, \mathbb{R}^{X_h}}_{N \text{ times}}, \mathbb{R}^{X_{h+1}}, \ldots, \mathbb{R}^{X_d})) =$$

$$= \bar{\rho}(\mathcal{N}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_{h-1}}, \underbrace{\mathbb{R}^{X_h}, \ldots, \mathbb{R}^{X_h}}_{M \text{ times}}, \mathbb{R}^{X_{h+1}}, \ldots, \mathbb{R}^{X_d})).$$

Theorem (Equivariant Universality)

Let d *be such that*

$$\bar{\rho} := \bar{\rho}\big(\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_f}, \underbrace{\mathbb{R}^X, \ldots, \mathbb{R}^X}_{d \text{ times}})\big) = \bar{\rho}\big(\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_f}, \underbrace{\mathbb{R}^X, \ldots, \mathbb{R}^X}_{d+1 \text{ times}})\big).$$

Then,

$$\mathcal{U}_\sigma(\mathbb{R}^{X_0}, \ldots, \mathbb{R}^{X_f}, \underbrace{\mathbb{R}^X, \ldots, \mathbb{R}^X}_{d+1 \text{ times}}) = \mathcal{C}_{\bar{\rho}}(\mathbb{R}^{X_0}, \mathbb{R}^X).$$

Example — Universality of Equivariant PointNet

Three layer equivariant PointNets are universal, namely, $\mathcal{U}_\sigma(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R}^n, \mathbb{R}^n) = \mathcal{C}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)$.

Future Directions

- **Approximation rates:** How fast can these functions be approximated?
- **Learning dynamics:** Can layer-wise constraints also obstruct the dynamics?
- **Generalization and scaling laws:** How do these models generalize with respect to dataset size?