On Universality Classes of Equivariant Networks

Marco Pacini [1, 2], Gabriele Santin [3], Bruno Lepri [2], Shubhendu Trivedi

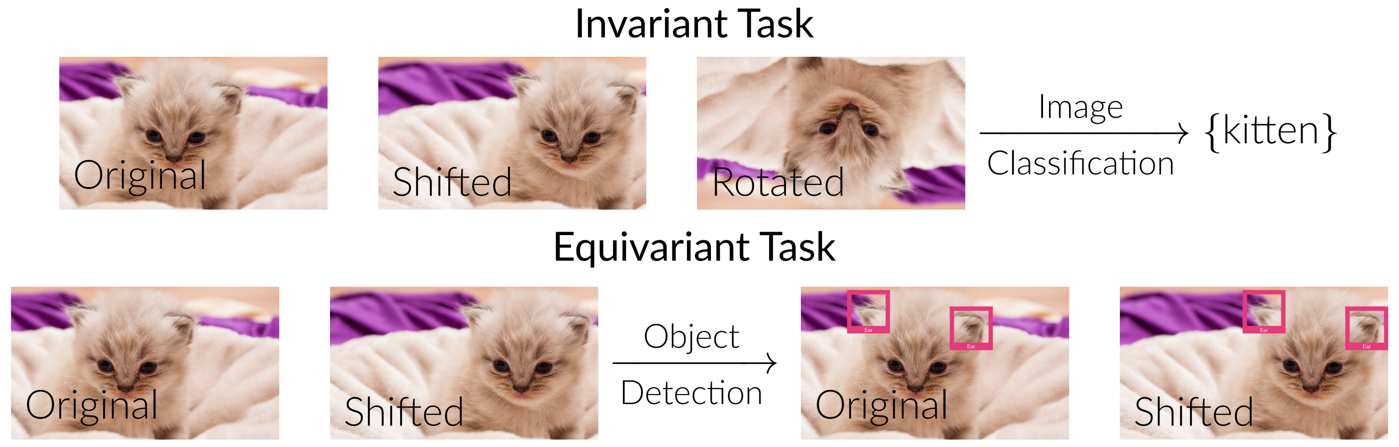
[1] University of Trento, [2] Fondazione Bruno Kessler, [3] University of Venice

NEURAL INFORMATION PROCESSING SYSTEMS

↓ paper link ↓

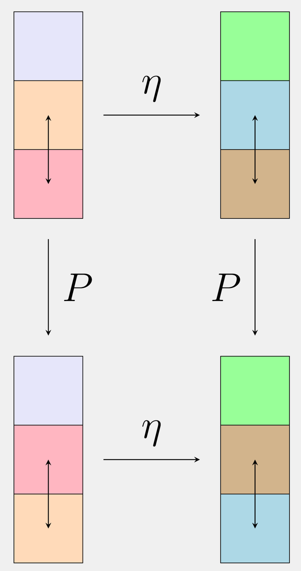
Basics of Geometric Deep Learning

Symmetries & Machine Learning

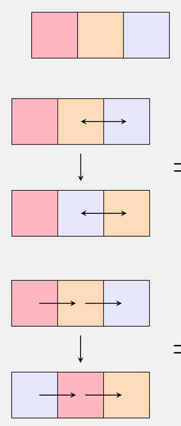


Permutations, Symmetries & Equivariance

Idea



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

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

Point-wise Activations

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

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

Equivariant Neural Networks

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

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

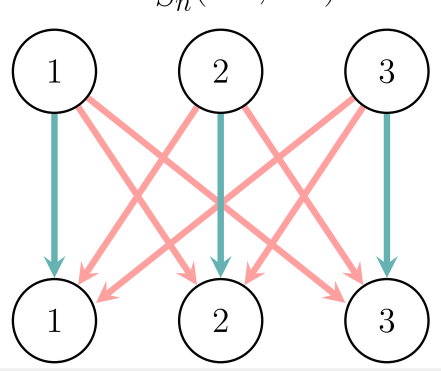
Spaces of Equivariant Neural Networks

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

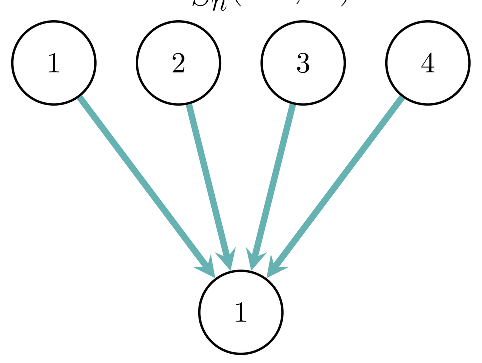
Example — PointNet

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

$\mathrm{Aff}_{S_n}(\mathbb{R}^n, \mathbb{R}^n)$



$\mathrm{Aff}_{S_n}(\mathbb{R}^n, \mathbb{R})$

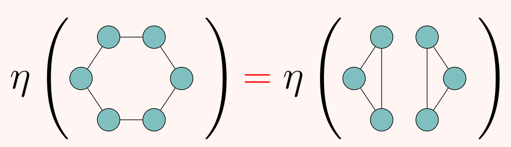


Issue: Separation constraints Universality

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

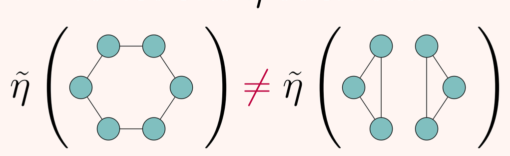
For k-WL theory, if $\mathcal{N} = \{1-\mathrm{WL} \sim \mathrm{GNNs}\}$:

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



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

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



Research Question – Informal

Is separation a complete proxy for universality?

Separation

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

- Separation relation: $\rho = \rho(\mathcal{N}) := \{(x, y) \in V \times V \mid \eta(x) = \eta(y) \ \forall \eta \in \mathcal{N}\}$
- Separation-constrained functions: $\mathcal{C}_\rho(V) := \{f \in \mathcal{C}(V) \mid f(x) = f(y) \ \forall (x, y) \in \rho\}$
- Research Question — formal: $\overline{\mathcal{N}} \overset{?}{=} \mathcal{C}_\rho(V)$

Examples of 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 to $\mathbb{R}^m, \mathbb{R}, \mathbb{R}$ 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

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

We define the universality class:

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

Then,

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

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; Mircea Petrache; B. Lepri; S. Trivedi; R. Walters. On Universality Classes of Equivariant Networks. *NeurIPS*, 2025.

Results

(?) Universality for Invariant Shallow PointNets (?)

$$\overline{\bigcup_{h \in \mathbb{N}} \mathcal{N}(\mathbb{R}^n, \mathbb{R}^{n \times h}, \mathbb{R})} \overset{?}{=} \mathcal{C}_{S_n}(\mathbb{R}^n)$$

Definition — Universality Class

The universality class associated to $\mathbb{R}^X, \mathbb{R}^Y$ and $\mathbb{R}^Z$ is

$$\mathcal{U}(\mathbb{R}^X, \mathbb{R}^Y, \mathbb{R}^Z) := \overline{\bigcup_{h \in \mathbb{N}} \mathcal{N}_\sigma(\mathbb{R}^X, \mathbb{R}^{h \times Y}, \mathbb{R}^Z)}$$

Universality Classes with same Separation Power

Prop: The following universality classes have the same separation power:

$$\rho(\mathcal{U}(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R})) = \rho(\mathcal{U}(\mathbb{R}^n, \mathbb{R}^{S_n}, \mathbb{R})).$$

Failure of Separation-constrained Universality

Prop: The following spaces differ in their approximation capabilities when $n > 2$:

$$\mathcal{U}(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R}) \subsetneq \mathcal{U}(\mathbb{R}^n, \mathbb{R}^{S_n}, \mathbb{R}).$$

Example

$x_1 \cdots x_n \in \mathcal{U}(\mathbb{R}^n, \mathbb{R}^{S_n}, \mathbb{R})$ but $x_1 \cdots x_n \notin \mathcal{U}(\mathbb{R}^n, \mathbb{R}^n, \mathbb{R})$.

In this case, separation ***is not*** *a complete proxy for universality*

Characterization

Theorem: Let f be an invariant function, then

$$f \in \mathcal{U}(\mathbb{R}^X, \mathbb{R}^Y, \mathbb{R})$$
$$\iff$$
$$P(\partial_1, \dots, \partial_d) f = 0 \quad \forall P \text{ such that } P(\phi_1) = \dots = P(\phi_m) = 0,$$

where $\phi_1, \dots, \phi_m \in \mathbb{R}^d$ depending on X, Y.

Examples of Separation-Constrained Universality

Normal subgroup

A subgroup $H \subseteq G$ is *normal* if $ghg^{-1} \in H$ for each $h \in H, g \in G$.

Let V and Z be permutation representations of a finite group G, and let H be a normal subgroup of G. Therefore, $\mathcal{U}(V, \mathbb{R}^{G/H}, Z) = \mathcal{C}_\rho(V, Z)$.

Future Directions

- **Depth:** Characterization of universality in the deep case?
- **Equivariance:** What does it changes with equivariant readouts?

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