Separation Power of Equivariant Neural Networks

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

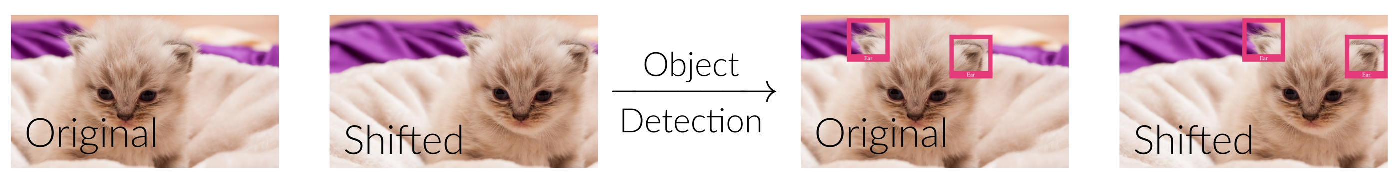
↓ paper link ↓

Basics of Geometric Deep Learning

Symmetries & Machine Learning

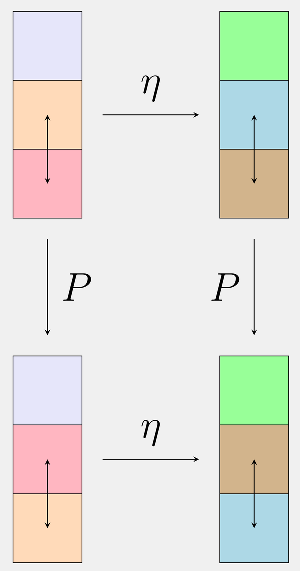
Invariant Task

Equivariant Task

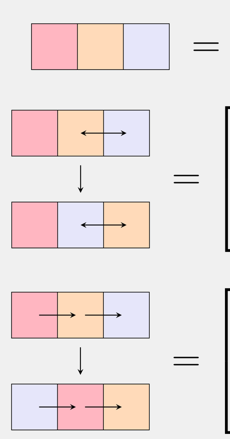


Permutations, Symmetries & Equivariance

Idea



Permutations



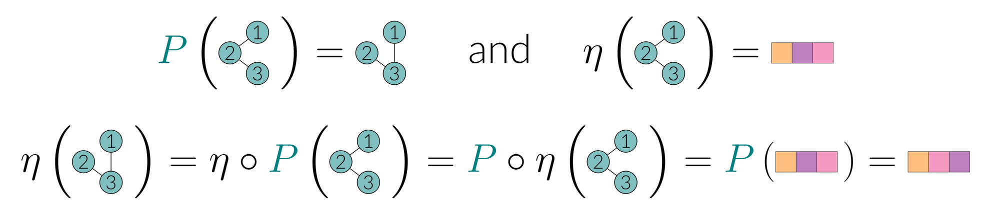
$$\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix} = P \qquad \begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix} = Q$$

Equivariance

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

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

E.g.:



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** $\tilde{\sigma}$:

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

Spaces of Equivariant Neural Networks

Given **layer spaces** as subspaces $M_i \subseteq \mathrm{Aff}(\mathbb{R}^{n_{i-1}}, \mathbb{R}^{n_i})$, define:

$$\mathcal{N} = \mathcal{N}_\sigma(M_1, \dots, M_d) := \{\phi_m \circ \tilde{\sigma} \circ \phi_{m-1} \circ \cdots \circ \tilde{\sigma} \circ \phi_0 \mid \phi_i \in M_i\}$$

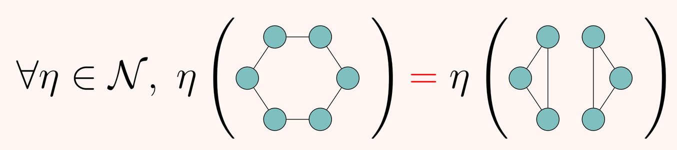
E.g., Shallow Invariant CNNs: $\mathcal{N} = \mathcal{N}_\sigma(M_1, M_2)$

- M_1 represents all the convolutions with fixed filter width and filter number
- M_2 represents invariant layers with fixed output dimension

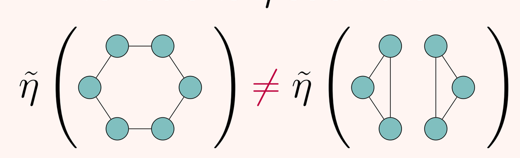
Issue: Separation constraints Universality

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

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



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



Workaround: Separation-constrained Universality

Can we approximate functions with **specific separation constraints**, as GNNs do?

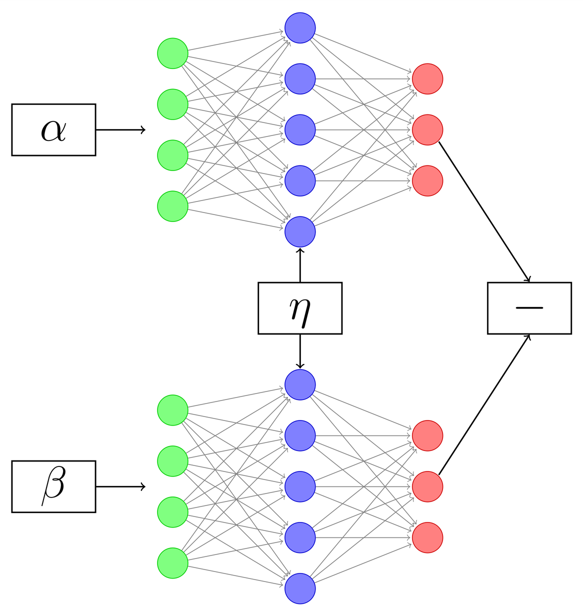
- First, we need to **describe** these constraints — but we do not have WL!
- Then, we need to understand how to **influence** these constraints in an actionable way!

Research Questions

1. How can we **compute** input pairs **identified** by $\mathcal{N}$? Formally:
$$\rho(\mathcal{N}) := \{(\alpha, \beta) \in X \times X \mid \eta(\alpha) = \eta(\beta), \forall \eta \in \mathcal{N}\}$$
2. How do **hyperparameters** and **architecture choices** influence $\rho(\mathcal{N})$?

1. Main Theorem - How to Compute $\rho(\mathcal{N})$

The Twin Network Trick



Separation ↔ Zero Locuses

- Define the **zero locus**:
$$\mathcal{I}(\mathcal{N}) := \{x \in X \mid \eta(x) = 0 \ \forall \eta \in \mathcal{N}\}$$
- The **space of twin networks**:
$$\Delta(\mathcal{N}) := \{(\alpha, \beta) \mapsto \eta(\alpha) - \eta(\beta) | f \in \mathcal{N}\}$$
- Then, we have the **equivalence**:
$$\rho(\mathcal{N}) = \mathcal{I}(\Delta(\mathcal{N}))$$

Claim - *Informal*

For **non-polynomial** activations σ, we have the following depth-recursive formula:

$$\mathcal{I}(\mathcal{N}_\sigma(M_1, \dots, M_d)) = \bigcap_{h,k} \bigcup_{\mathcal{Q} \in \Psi_{h,k}} \bigcap_{\substack{P \in \mathcal{Q} \\ i,j \in P}} \mathcal{I}(\mathcal{N}_\sigma(M_1, \dots, M_{d-2}, (M_{d-1})_{ij}))$$

2. Hyperparameters and Architecture Choices

The Role of Activations

Non-polynomial activations σ, τ yield **equivalent** separation power:

$$\rho(\mathcal{N}_\sigma(M_1, \dots, M_d)) = \rho(\mathcal{N}_\tau(M_1, \dots, M_d))$$

Furthermore, separation power can **only decrease** for **polynomial activations**

The Role of Depth

For compatible layer spaces M_i and $m \leq n$:

$$\rho(\mathcal{N}_\sigma(M_1, \dots, M_{h-1}, \underbrace{M_h, \dots, M_h}_{n \text{ times}}, M_{h+1}, \dots, M_d)) \subseteq$$
$$\subseteq \rho(\mathcal{N}_\sigma(M_1, \dots, M_{h-1}, \underbrace{M_h, \dots, M_h}_{m \text{ times}}, M_{h+1}, \dots, M_d))$$

but there is a **repetition threshold** beyond which separation power **stabilizes**

The Role of Multiple Features

Given $M_i = \mathrm{Aff}_G(\mathbb{R}^{n_{i-1}}, \mathbb{R}^{n_i})$ for each i, write

$$\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_d}) := \mathcal{N}_\sigma(M_1, \dots, M_d)$$

- For **arbitrary** $G \curvearrowright \mathbb{R}^{n_i}, \mathbb{R}^{n_{i'}}$:
$$\rho(\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_i} \oplus \mathbb{R}^{n_{i'}}, \dots, \mathbb{R}^{n_d})) =$$
$$\rho(\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_i}, \dots, \mathbb{R}^{n_d})) \ \cap \ \rho(\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_{i'}}, \dots, \mathbb{R}^{n_d}))$$
- For the **trivial** $G \curvearrowright \mathbb{R}^f$:
$$\rho(\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_i} \otimes \mathbb{R}^f, \dots, \mathbb{R}^{n_d})) = \rho(\tilde{\mathcal{N}}_\sigma(\mathbb{R}^{n_0}, \dots, \mathbb{R}^{n_i}, \dots, \mathbb{R}^{n_d}))$$

Corollary on IGNs

There exists $d > 0$ such that for **any** hidden feature dimensions $f_1, \dots, f_d > 0$, the space $\tilde{\mathcal{N}}_\sigma((\mathbb{R}^n)^{\otimes 2} \otimes \mathbb{R}^{f_0}, (\mathbb{R}^n)^{\otimes k} \otimes \mathbb{R}^{f_1}, \dots, (\mathbb{R}^n)^{\otimes k} \otimes \mathbb{R}^{f_d}, \mathbb{R})$ matches the separation power of k-WL

The Role of Representation Type

For $H \leq G$, define the **left cosets** of H and the relative vector space $\mathbb{R}^{G/H}$:

$$G/H := \{gH \mid g \in G\} \quad \text{and} \quad \mathbb{R}^{G/H} := \{f : G/H \to \mathbb{R}\}$$

For $K \leq H \leq G$, we have:

$$\rho(\tilde{\mathcal{N}}_\sigma(V, \dots, \mathbb{R}^{G/K}, \dots, W)) \subseteq \rho(\tilde{\mathcal{N}}_\sigma(V, \dots, \mathbb{R}^{G/H}, \dots, W))$$

Future Directions

- **Separation-constrained Approximation:** Can universality results be derived under the separation constraint?
- **Generalization Bounds:** How do models with the same separation power generalize?

References

[1] C. K. Joshi et al. On the Expressive Power of Geometric Graph Neural Networks. *ICLR*, 2023.
[2] H. Maron et al. Provably Powerful Graph Networks. *ICLR*, 2019.
[3] M. Pacini et al. A Characterization Theorem for Equivariant Networks with Point-wise Activations. *ICLR*, 2024.
[4] F. Geerts and J. L. Reutter. Expressiveness and Approximation Properties of Graph Neural Networks. *ICLR*, 2022.

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