We show **improvements** of **relative representations** through **invariance to symmetries** in parameter space, and **topological regularization** of the latent spaces

Relative Representations: Topological and Geometric Perspectives

Alejandro García-Castellanos, Giovanni Luca Marchetti, Danica Kragic, Martina Scolamiero

1) Representational Universality: *"isometries up to scale"*

Let $\varphi\colon \mathcal{X} \to \mathcal{Z}$ be the feature extractor of the network, and let $\mathcal{A} = \{a_1, \dots, a_k\} \subset \mathcal{Z}$ be a set of elements called *anchors*, and let $\mathrm{sim}\colon \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a similarity function.

The ***Relative Representation*** of $z \in \mathcal{Z}$ w.r.t. $\mathcal{A}$ is

$$T^{\mathcal{A}}_{\mathrm{rel}}(z) = (\mathrm{sim}(z, a_1), \dots, \mathrm{sim}(z, a_k)) \in \mathbb{R}^k$$

sim = cosine sim → **Invariant to** ***isometries + isotropic rescalings***

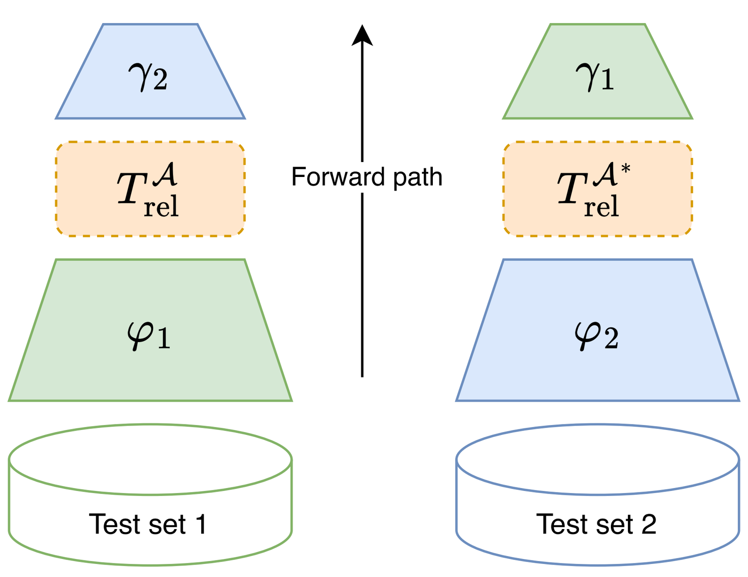
Good performing models have "similar" latent representations

⇓

Empirical evidence of being isometric up to scale

⇓

Relative Representations enable zero-shot Model Stitching



2) Symmetry Groups of Activation Functions

The **intertwiner group** of the activation function $\sigma\colon \mathbb{R}^n \to \mathbb{R}^n$ is the set G^n_σ of invertible linear transformations that exhibit equivalent transformations before and after σ, i.e.,

$$G^n_\sigma = \{A \in GL_n \mid \exists B \in GL_n\colon \sigma \circ A = B \circ \sigma\}$$

For common activation functions *(e.g. GELU, ReLU, sigmoid)*, the elements of G^n_σ are the product of a ***permutation and a diagonal matrix***

Theoretical explanation for the emergence of structurally-similar representations in networks

3) Invariance trading: Robust Relative Representation

Robust Relative Representation: We apply *Gaussian normalization* with respect to a batch $\mathcal{B}$ of data, i.e., a simple form of *batch normalization* (without learnable parameters), *before computing the cosine sim*

We are now invariant to shifts + **intertwiner group** actions

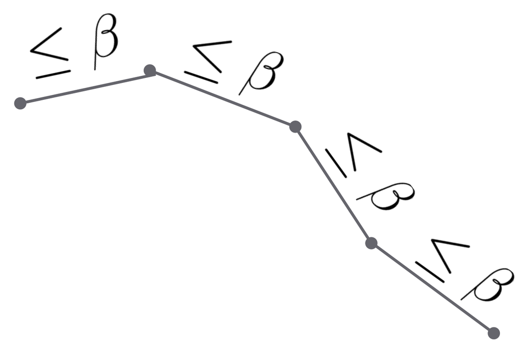
We trade off invariance to isometries other than permutations with more general non-isotropic rescalings → **Good trade in high dimensional latent spaces:**

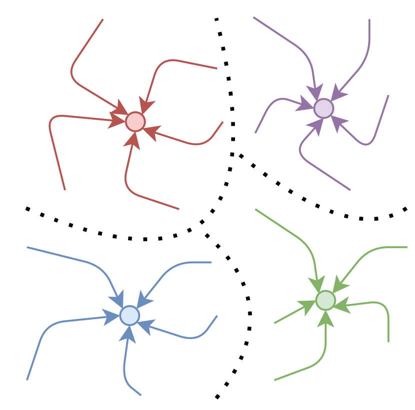
Performance comparison on zero-shot model stitching.

		Absolute			Relative Vanilla			Relative Robust		
γ	φ	Acc (↑)	F_1 (↑)	MAE (↓)	Acc (↑)	F_1 (↑)	MAE (↓)	Acc (↑)	F_1 (↑)	MAE (↓)
en	en	$59.26_{\pm 0.66}$	$58.27_{\pm 0.83}$	$49.52_{\pm 0.89}$	$38.84_{\pm 1.23}$	$23.50_{\pm 2.77}$	$84.95_{\pm 9.48}$	$60.84_{\pm 0.64}$	$60.30_{\pm 0.72}$	$45.35_{\pm 0.74}$
	fr	$24.28_{\pm 10.11}$	$22.27_{\pm 8.86}$	$139.27_{\pm 35.32}$	$40.96_{\pm 2.40}$	$31.15_{\pm 3.29}$	$73.09_{\pm 5.18}$	$49.92_{\pm 1.51}$	$50.13_{\pm 1.60}$	$57.56_{\pm 1.60}$
fr	en	$24.96_{\pm 9.27}$	$23.19_{\pm 8.12}$	$132.35_{\pm 24.01}$	$35.42_{\pm 1.16}$	$20.86_{\pm 1.09}$	$79.68_{\pm 11.68}$	$60.74_{\pm 0.88}$	$60.18_{\pm 1.14}$	$45.19_{\pm 1.16}$
	fr	$49.26_{\pm 1.04}$	$48.74_{\pm 0.73}$	$63.89_{\pm 1.50}$	$41.99_{\pm 3.18}$	$35.33_{\pm 4.55}$	$67.77_{\pm 2.24}$	$50.31_{\pm 0.88}$	$50.95_{\pm 0.82}$	$57.08_{\pm 1.22}$

4) Constrained clusters: Topological Densification

High likelihood of β-connected → Mass attract mass



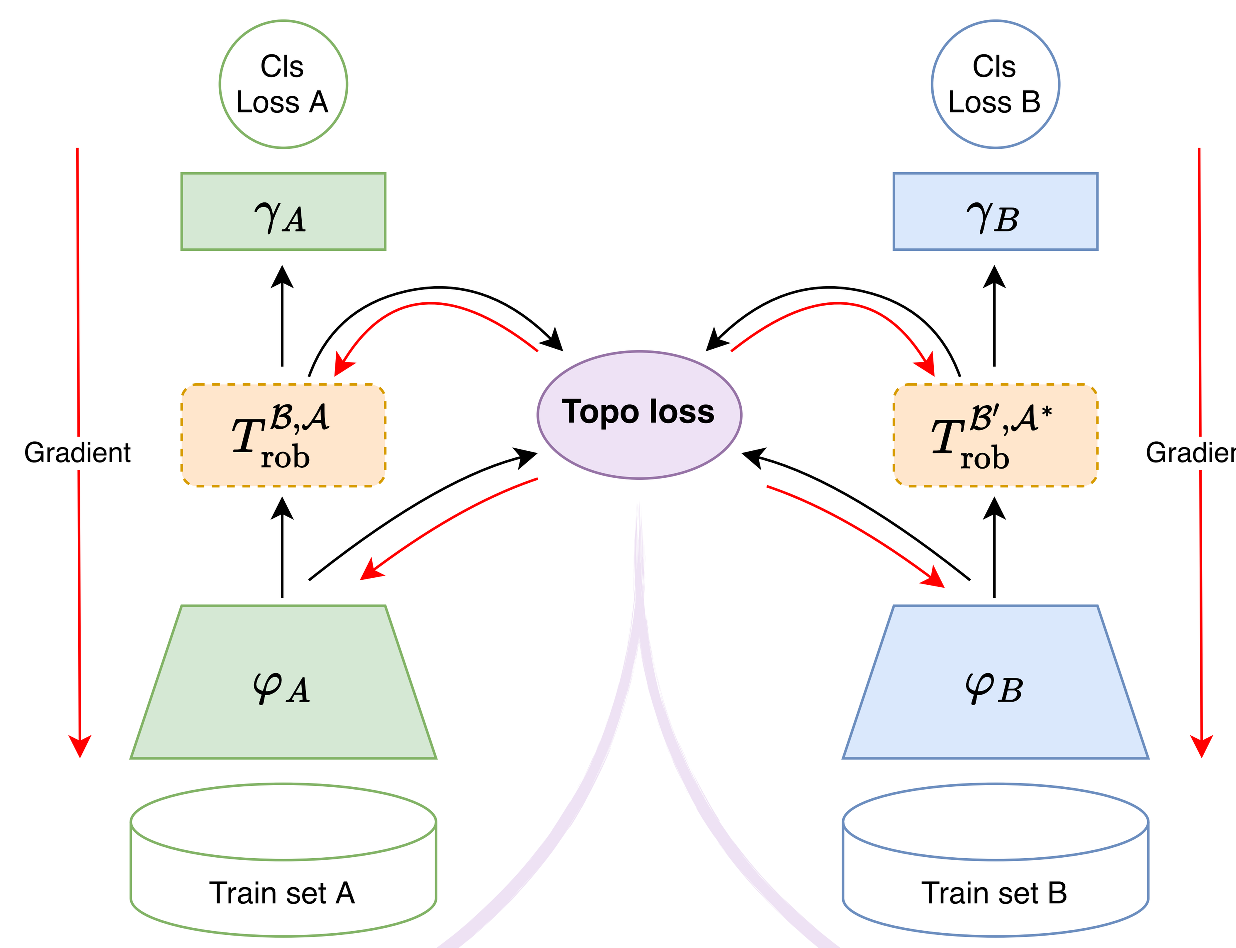


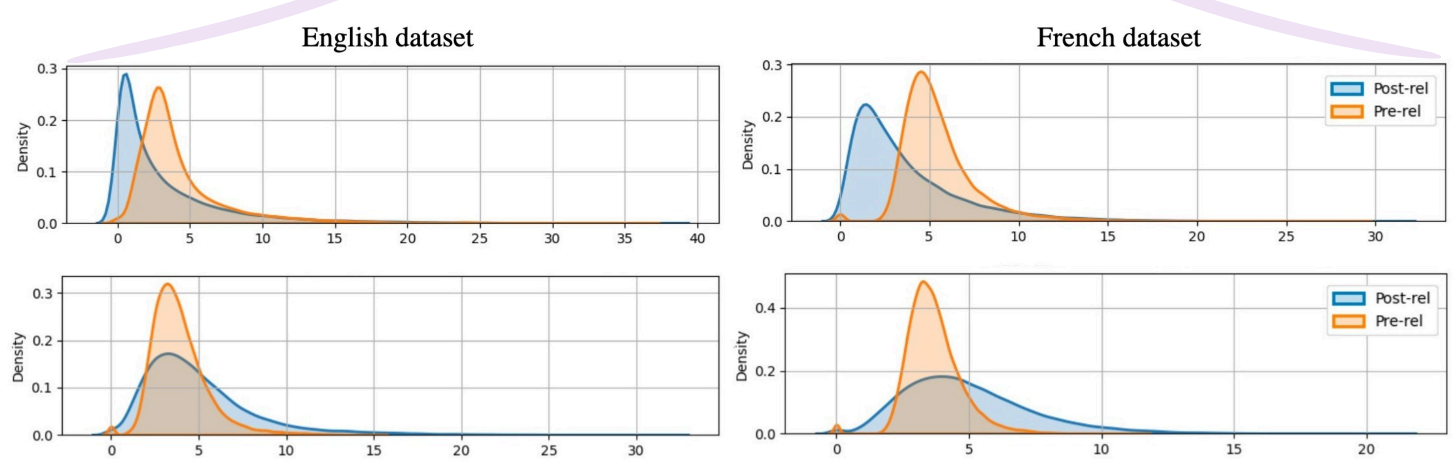
- Equal to having all **0-dimensional persistent homology** death-times of the Vietoris-Rips complex in $(0, \beta)$
- Can be enforced with **regularization**

- **Condensate**, for each class, its push-forward **distributions inside their decision boundary**
- Reduce generalization error

5) Topologically regularized relative representation

We apply the consistent topological densification before and after the *(robust)* relative transformation in all of our models during the fine-tuning phase





Distribution of death times on the English (left) and French (right) datasets. **Top:** without topological densification. **Bottom:** with a combination of pre-relative and post-relative topological densification.

Performance with topological densification.

		Relative Robust		
γ	φ	Acc (↑)	F_1 (↑)	MAE (↓)
en	en	$61.16_{\pm 0.42}$	$61.26_{\pm 0.18}$	$44.63_{\pm 0.26}$
	fr	$50.48_{\pm 1.04}$	$50.85_{\pm 1.25}$	$57.70_{\pm 0.73}$
fr	en	$60.93_{\pm 0.56}$	$61.23_{\pm 0.46}$	$44.54_{\pm 0.51}$
	fr	$50.63_{\pm 0.79}$	$50.97_{\pm 0.85}$	$57.76_{\pm 0.71}$

Hybrid Intelligence

KTH Royal Institute of Technology, Sweden
University of Amsterdam, Netherlands