Random Forest Autoencoders for Guided Representation Learning

The 39th Annual Conference on Neural Information Processing Systems

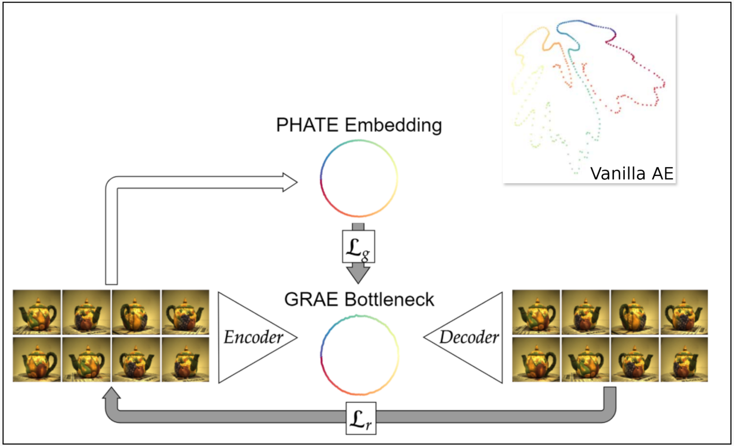
A. Aumon[1], S. Ni[1], M. Lizotte[1], G. Wolf[1], K. R. Moon[2], J. S. Rhodes[3]

[1]Université de Montréal; Mila [2]Utah State University [3]Brigham Young University

Background: Geometry regularized autoencoders

Unconstrained autoencoders (AEs) tend to deviate from the underlying data geometry

Example: Geometry Regularized Autoencoders (GRAE, Duque et al., 2022)



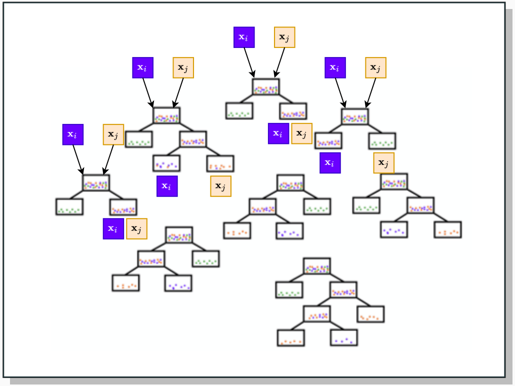
Background: RF-GAP proximities to filter out irrelevant features

Given a Random Forest (RF) trained on $\mathbf{D}_{\text{train}} = (X, Y)$, $\mathit{RFGAP}(\cdot, \cdot)$ (Rhodes et al., 2023) is a **supervised similarity** based on **RF leaf co-occurrence**:

- Captures supervised proximities by filtering out irrelevant features.
  $\implies$ Unified supervised manifold
- Induces N-dimensional supervised kernel representations for each $\mathbf{x}_i$:

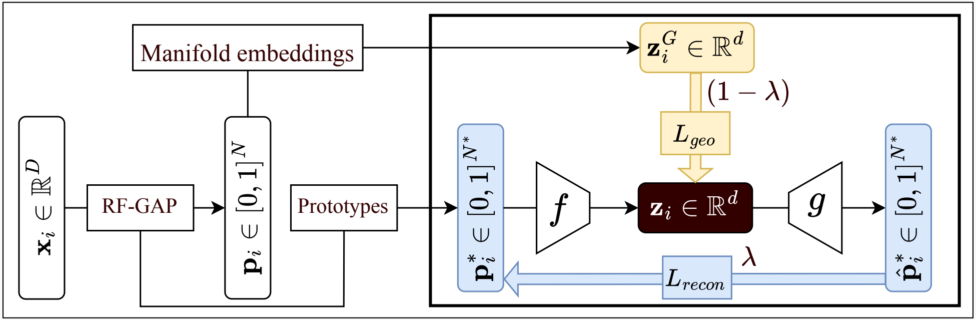
$$\mathbf{p}_i = \begin{bmatrix} \mathit{RFGAP}(\mathbf{x}_i, \mathbf{x}_1) & \cdots & \mathit{RFGAP}(\mathbf{x}_i, \mathbf{x}_N) \end{bmatrix}$$

- *Naturally extends to OOS points $x \notin X$*



RF-AE architecture: Encoding Random Forest-guided manifold geometry

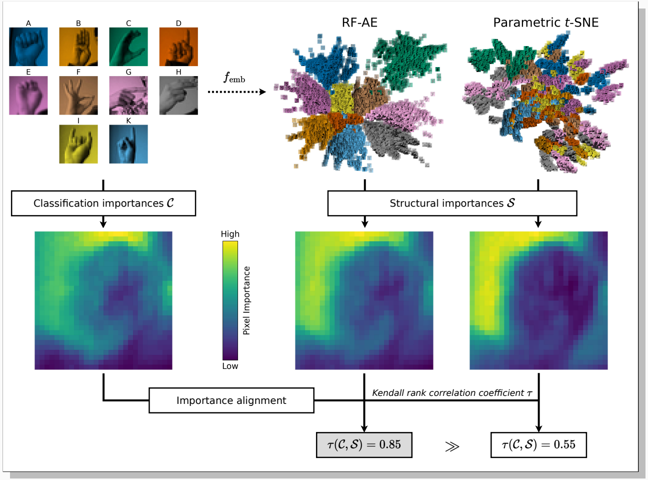
RF-AE = Prototype RF-GAP reconstruction + GRAE-like supervised geometric constraint to match RF-PHATE embeddings $\{z_i^G \in \mathbb{R}^d \mid 1, \ldots, N\}$ (Rhodes, Aumon et al., 2025)



$$\implies L_{RFAE}(f,g) = \frac{1}{N}\sum_{i=1}^{N}\left[\lambda L_{recon}(\mathbf{p}_i^*, \hat{\mathbf{p}}_i^*) + (1-\lambda) L_{geo}(\mathbf{z}_i, \mathbf{z}_i^G)\right]$$

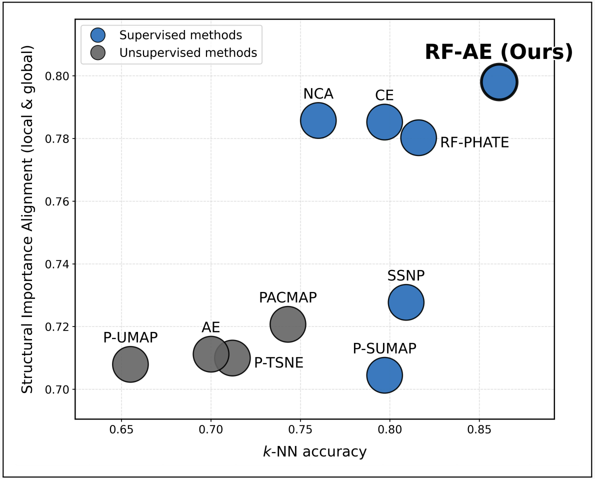
Structural Importance Alignment to quantify supervised structure preservation

Evaluation through classification accuracy is not enough... we need **representations that preserve domain-ware structure** $\longrightarrow$ we introduce *Structural Importance Alignment* (SIA)



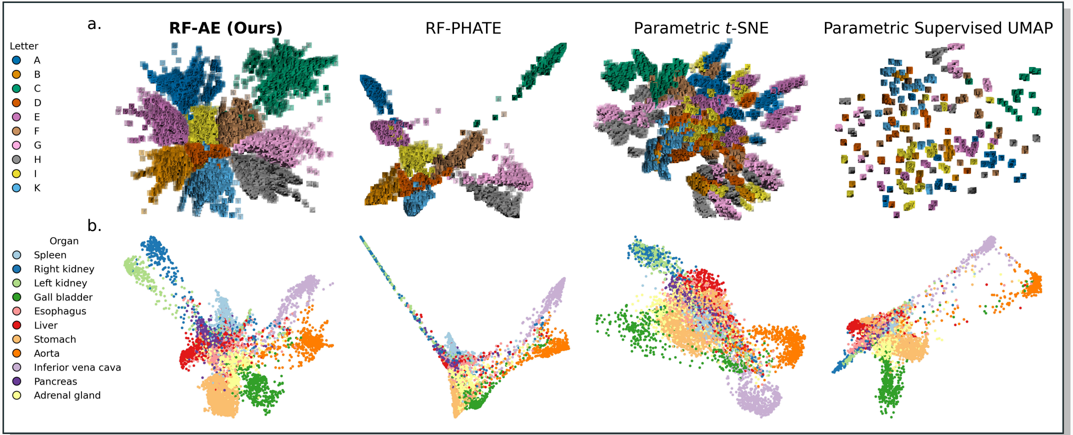
Quantitative comparison: accuracy & SIA

10 repetitions over 20 datasets spanning diverse domains (image, time-series, tabular...)



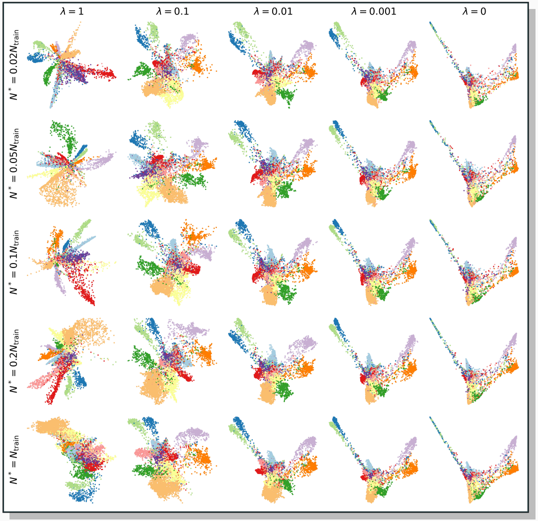
RF-AE achieves **impressive accuracy** while **aligning with domain-relevant structures**

Qualitative comparison (Sign MNIST & OrganC MNIST)

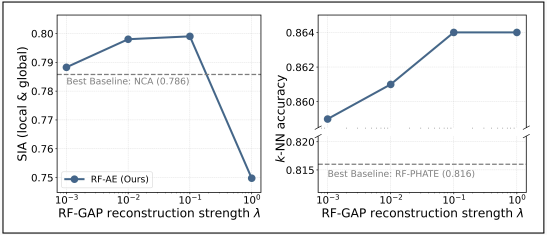


RF-AE **(a)** preserves denoised local and global sign variations, and **(b)** distinguishes **individual organs** while preserving their **anatomical/functional relationships**

Robustness analysis



RF-AE is **robust to hyperparameters**, but requires geometric regularization ($\lambda < 1$) to **avoid distortion** (SIA drop)



Conclusion

RF-AE is a very powerful and versatile tool for guided representation learning:

- **High classification performance**
- **Domain-relevant structure preservation**
- **OOS and semi-supervised support**
- **Highly robust** to hyperparameters
- **Insensitive to feature types and scales**
- Natively handles missing values (via the *`sklearn`* Random Forest implementation)

Link to RF-AE repository

Limitation: Faster than traditional kernel extensions, but still slower than fully parametric methods (see our GitHub for updates)