Navigating the Effect of Parametrization for Dimensionality Reduction

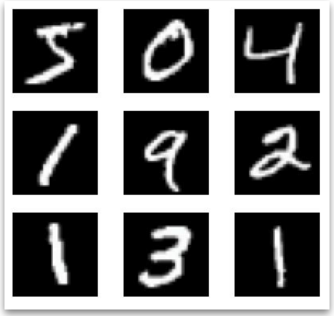
Haiyang Huang **Yingfan Wang** **Cynthia Rudin**
Duke University

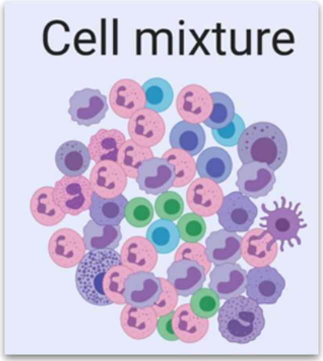
NeurIPS, Dec 2024

Duke DEPARTMENT *of* COMPUTER SCIENCE

DR Methods Preserve High-Dim Structure

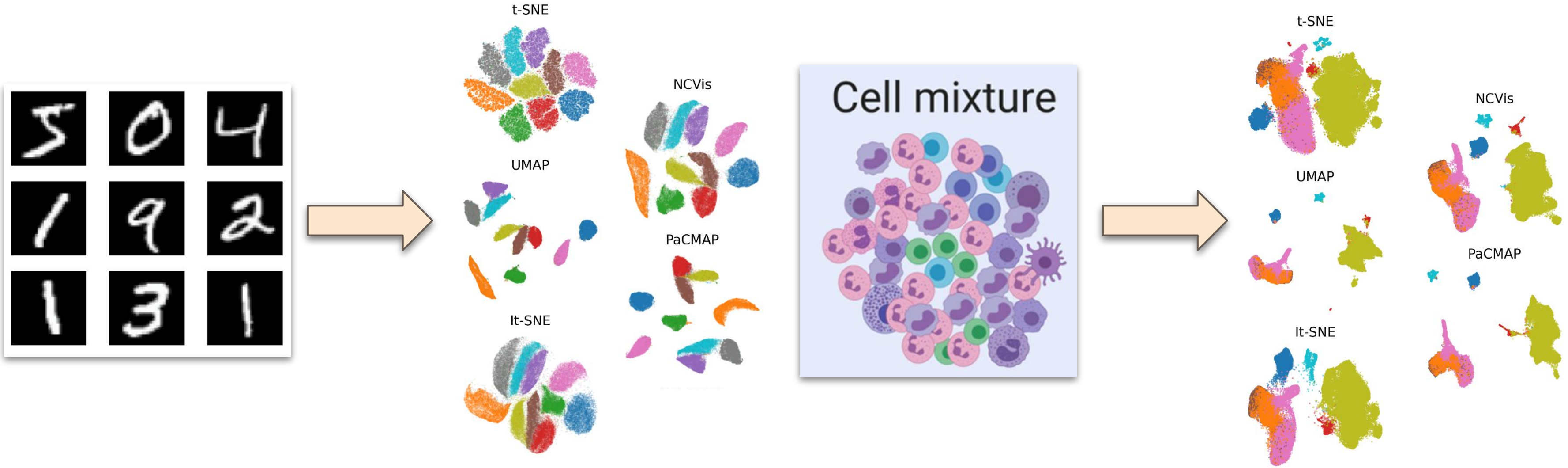
DR algorithm captures structure in high-dim space





DR Methods Preserve High-Dim Structure

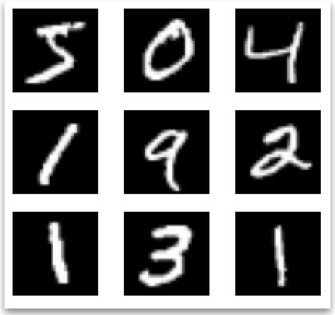
DR algorithm captures structure in high-dim space



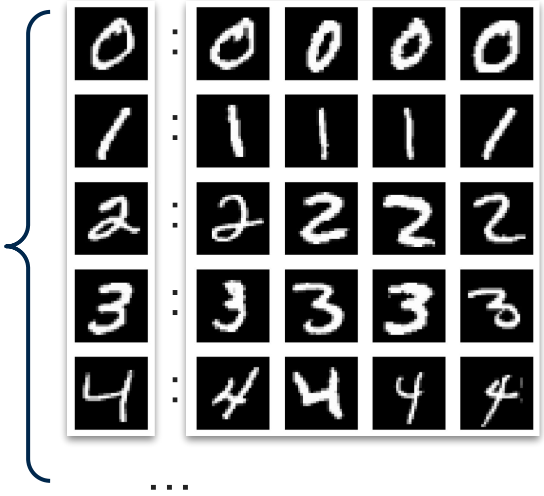
And render them in low-dim for visualization

DR Methods Fail for Continual Learning

Original Data



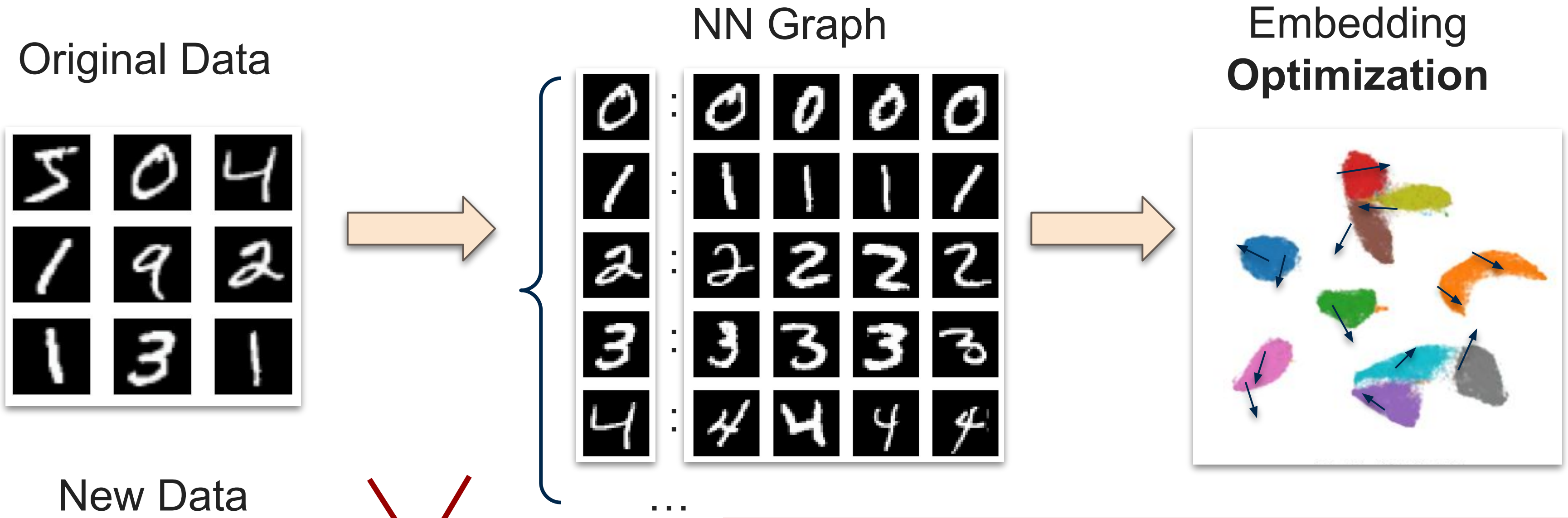
NN Graph

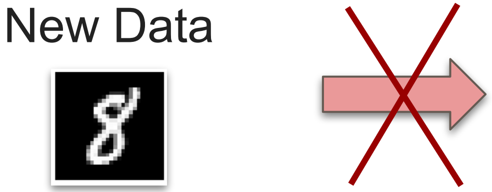


Embedding **Optimization**

They perform effectively in offline learning…

DR Methods Fail for Continual Learning

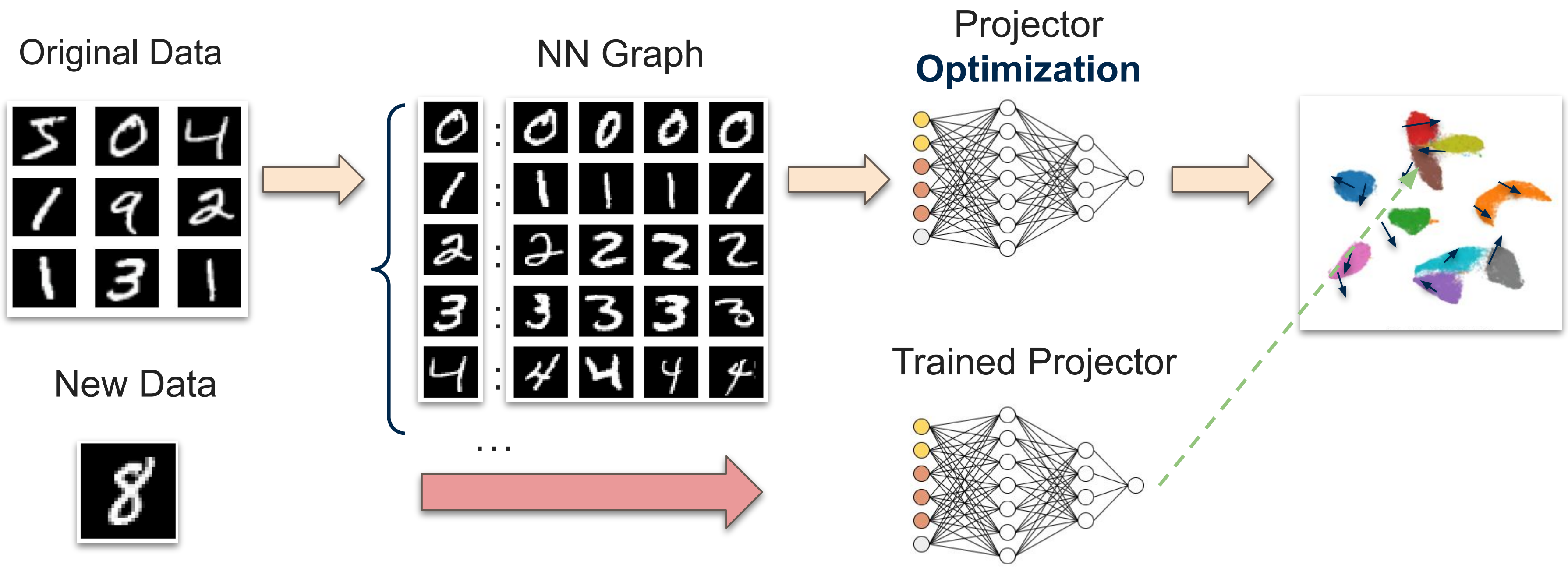




…but struggle to adapt to continual learning.

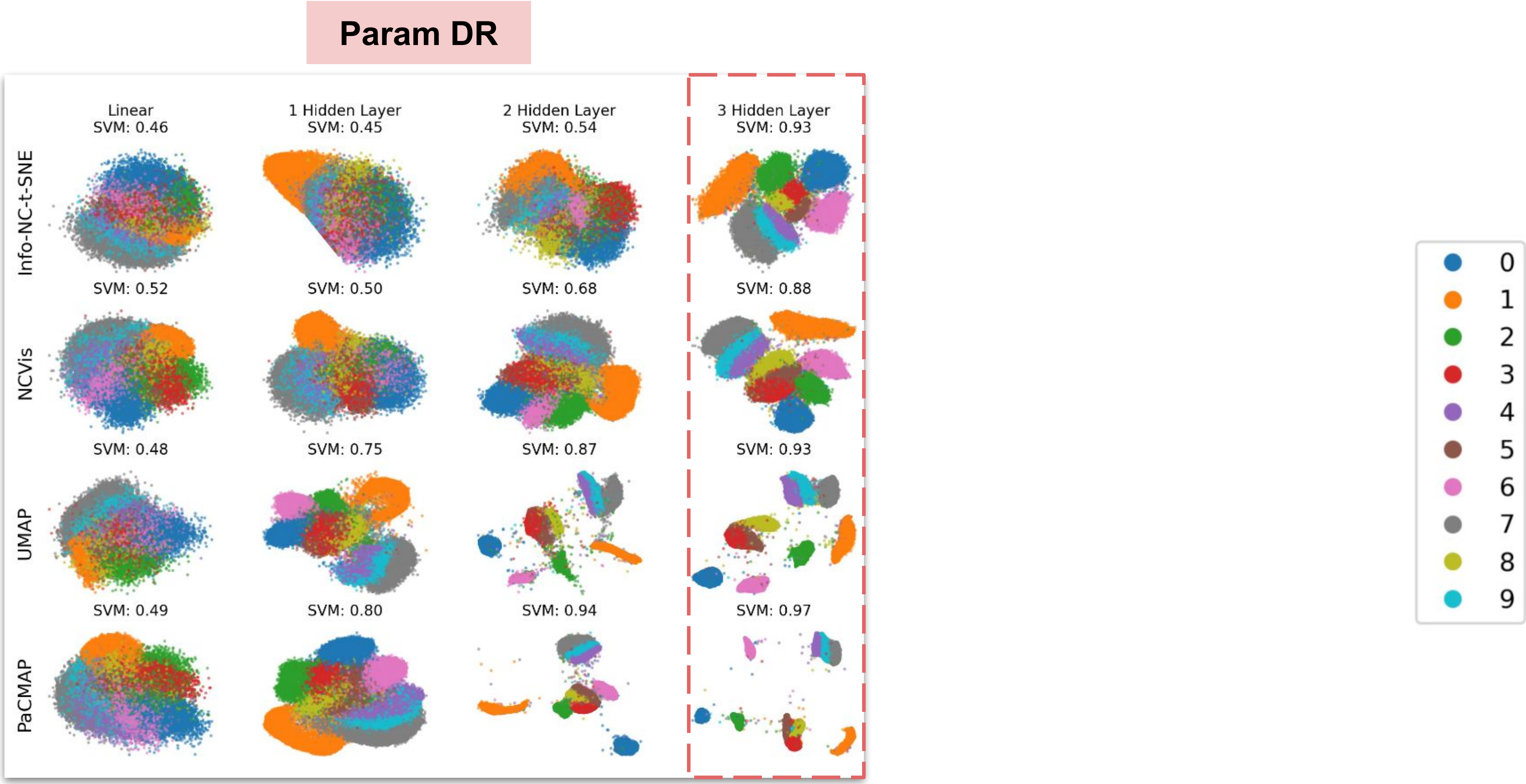
- Limits adaptability to new data
- Demands substantial time for large datasets

Param DR Embeds New Data in Existing Space



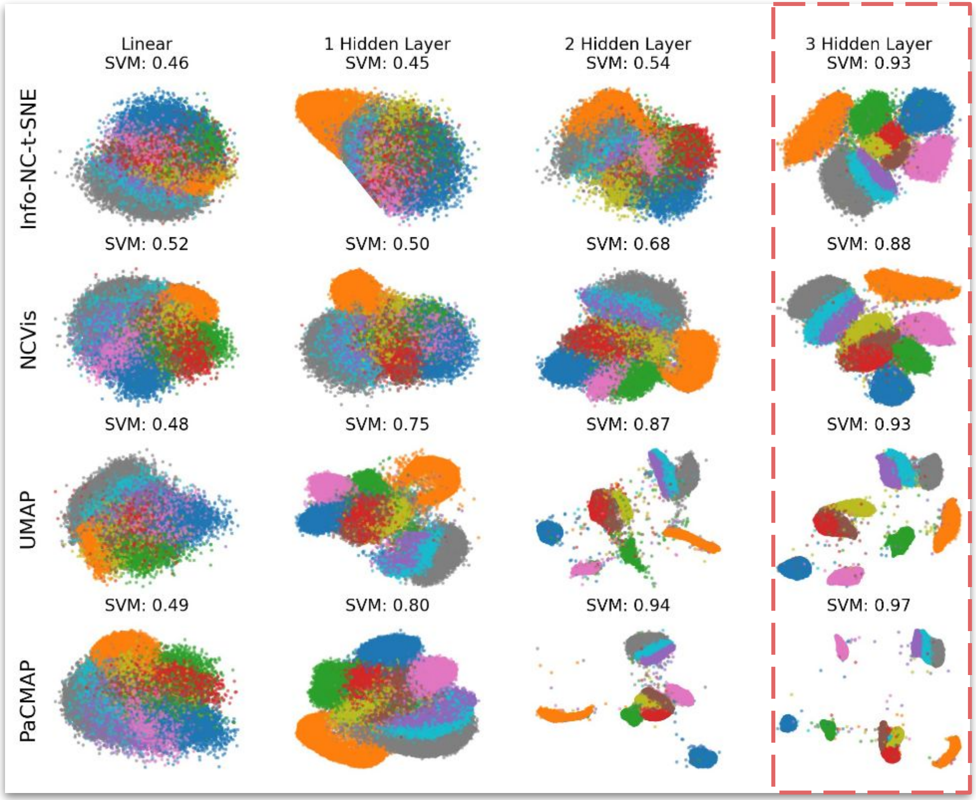
Parametric DR methods maintain continuity while efficiently managing new mappings

Param DR Fail to Keep Structure



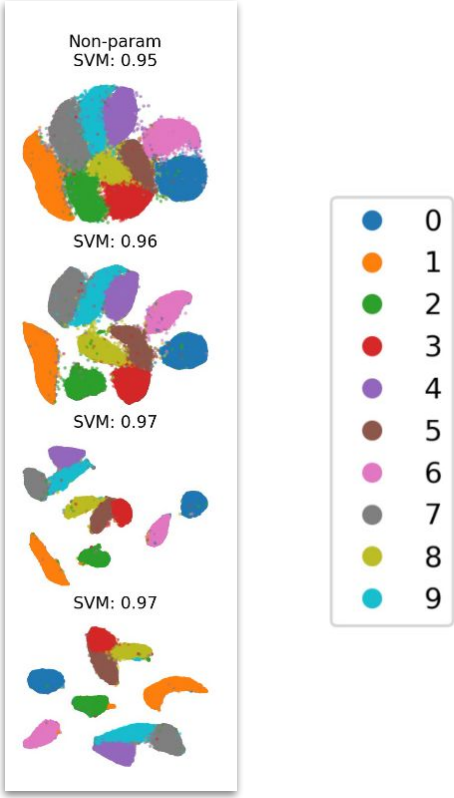
Param DR Fail to Keep Structure

Param DR

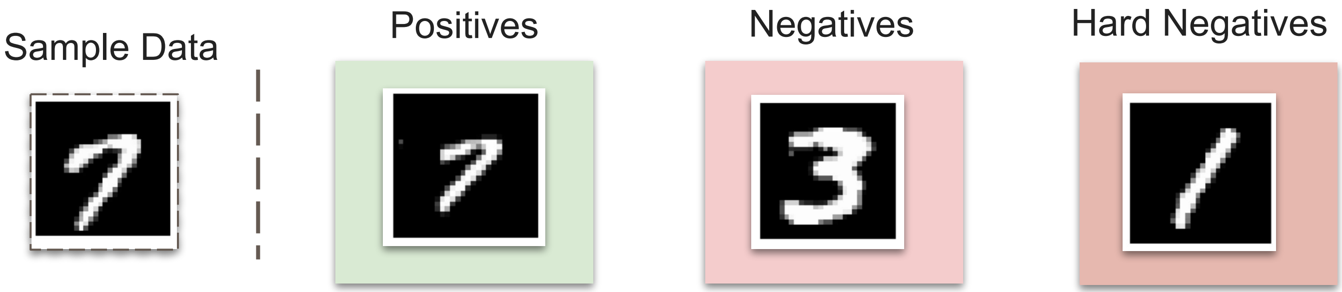


Clusters are overlapping

Non-Param DR



Hard Negatives and Insufficient Repulsion



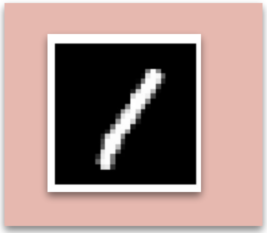
Hard Negatives and Insufficient Repulsion

Sample Data

Positives

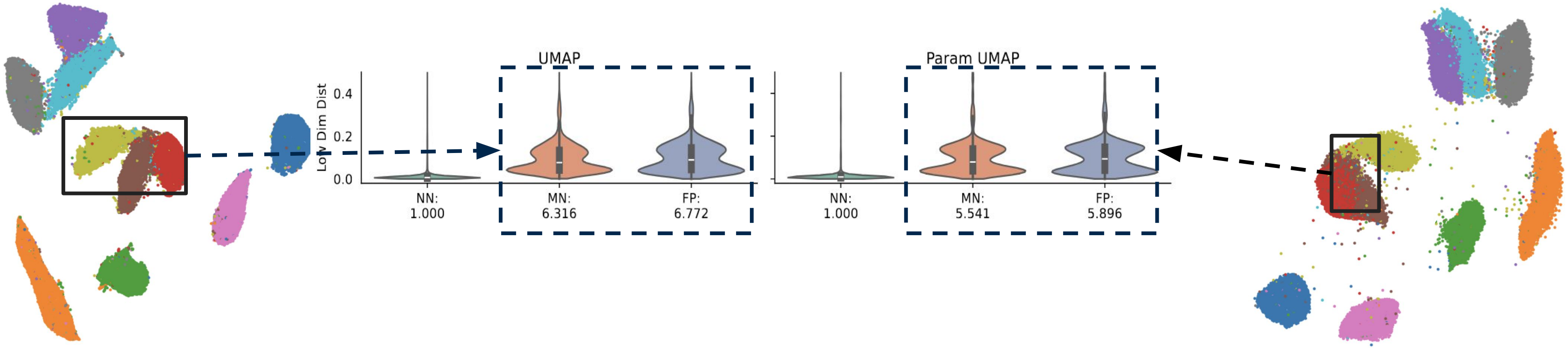
Negatives

Hard Negatives

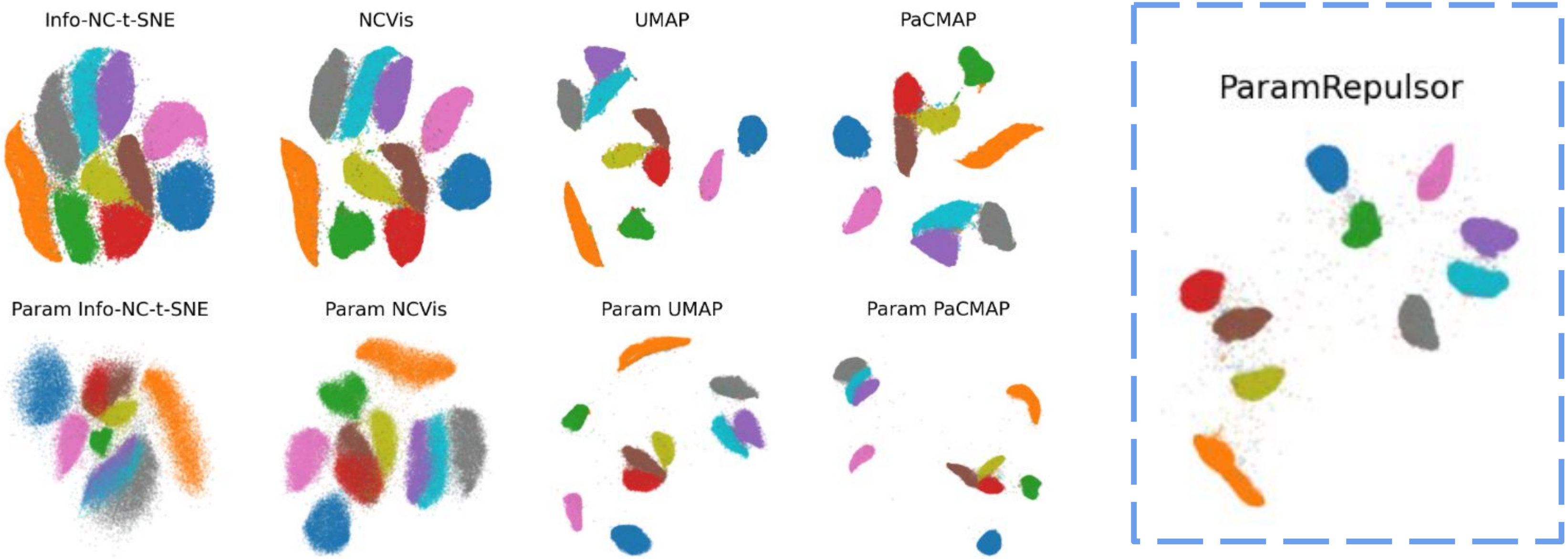


UMAP

P-UMAP

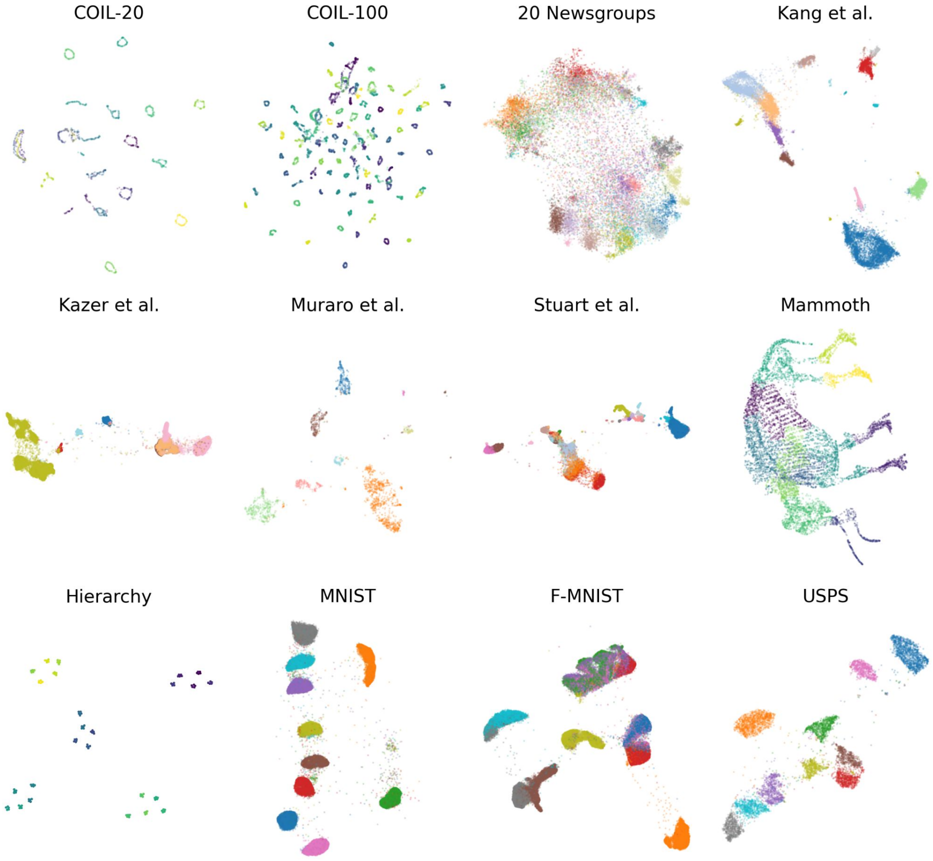


P-Repulsor Encourages Separation



ParamRepulsor ensures cluster separability with hard negatives

More visualizations



ParamRepulsor achieves state-of-the-art on multiple datasets

Check our paper and code for more information!

Paper

Code