Two Indispensable Tools for Scientific Discovery

Cynthia Rudin

Gilbert, Louis, and Edward
Lehrman Distinguished
Professor of Computer Science

Duke University

Joint work with Yingfan Wang, Haiyang Huang. Alina Barnett, Stark Guo, Ed Browne, Chaofan Chen, and many others

Two Indispensable Tools for Scientific Discovery

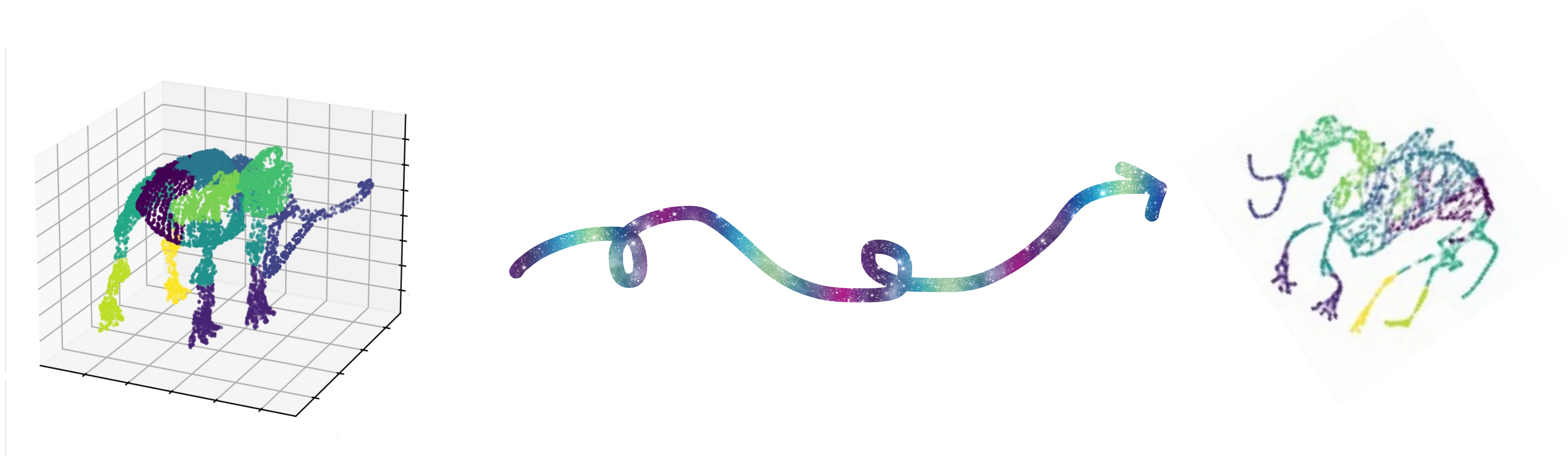
Dimension Reduction for Data Visualization

PaCMAP & Friends

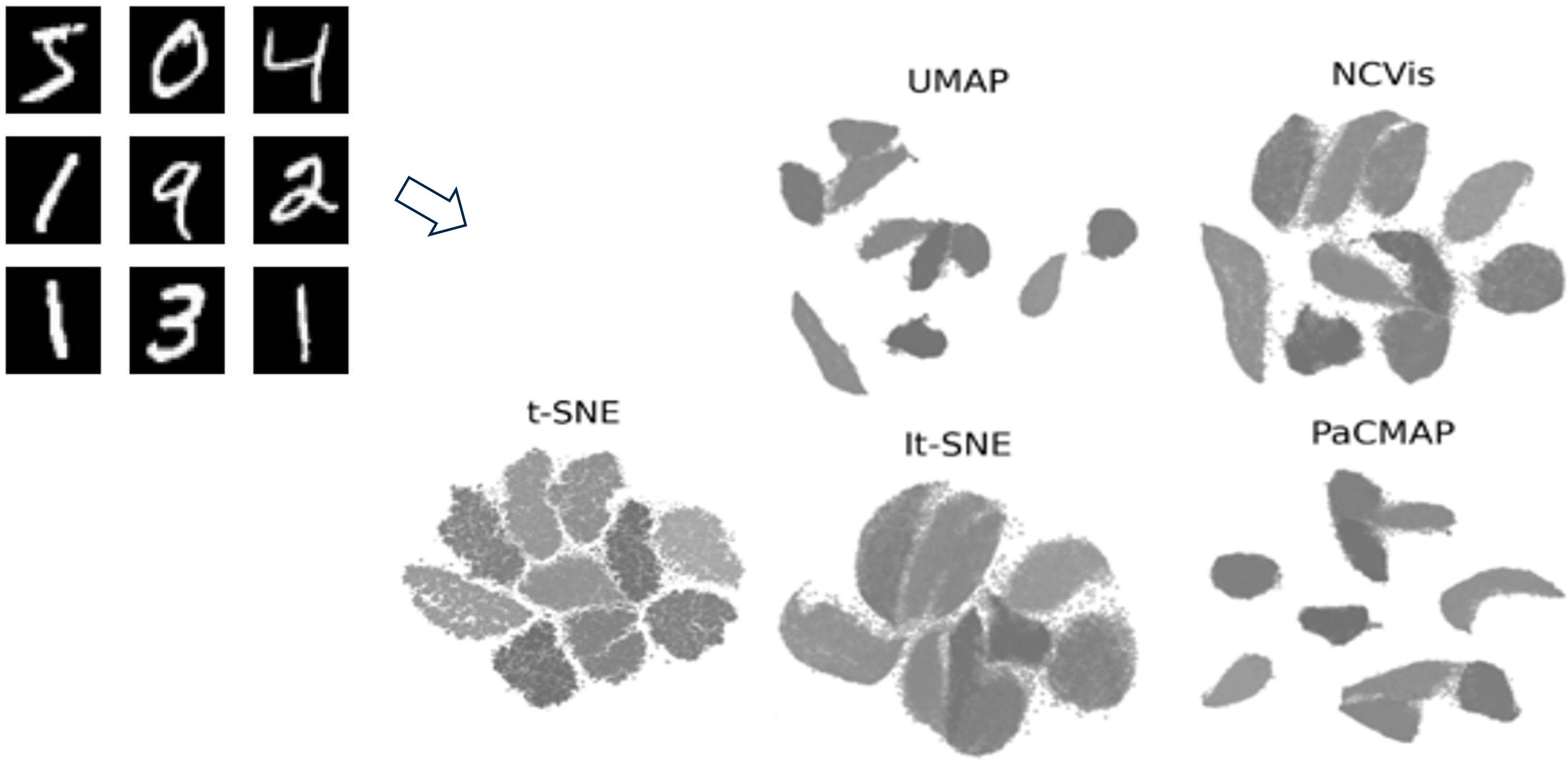
Interpretable Neural Networks

ProtoPNet & Friends

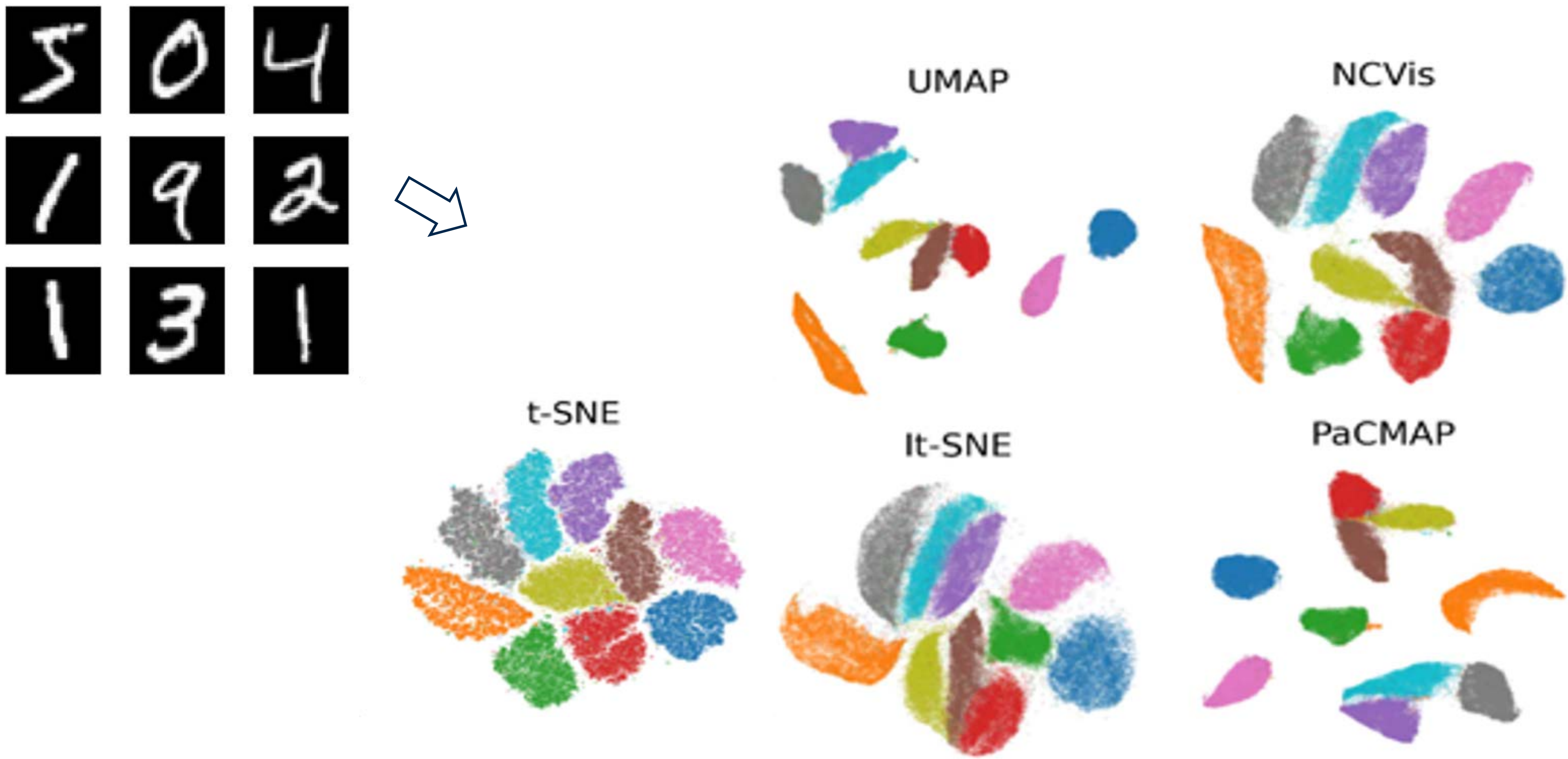
Dimension Reduction for Data Visualization



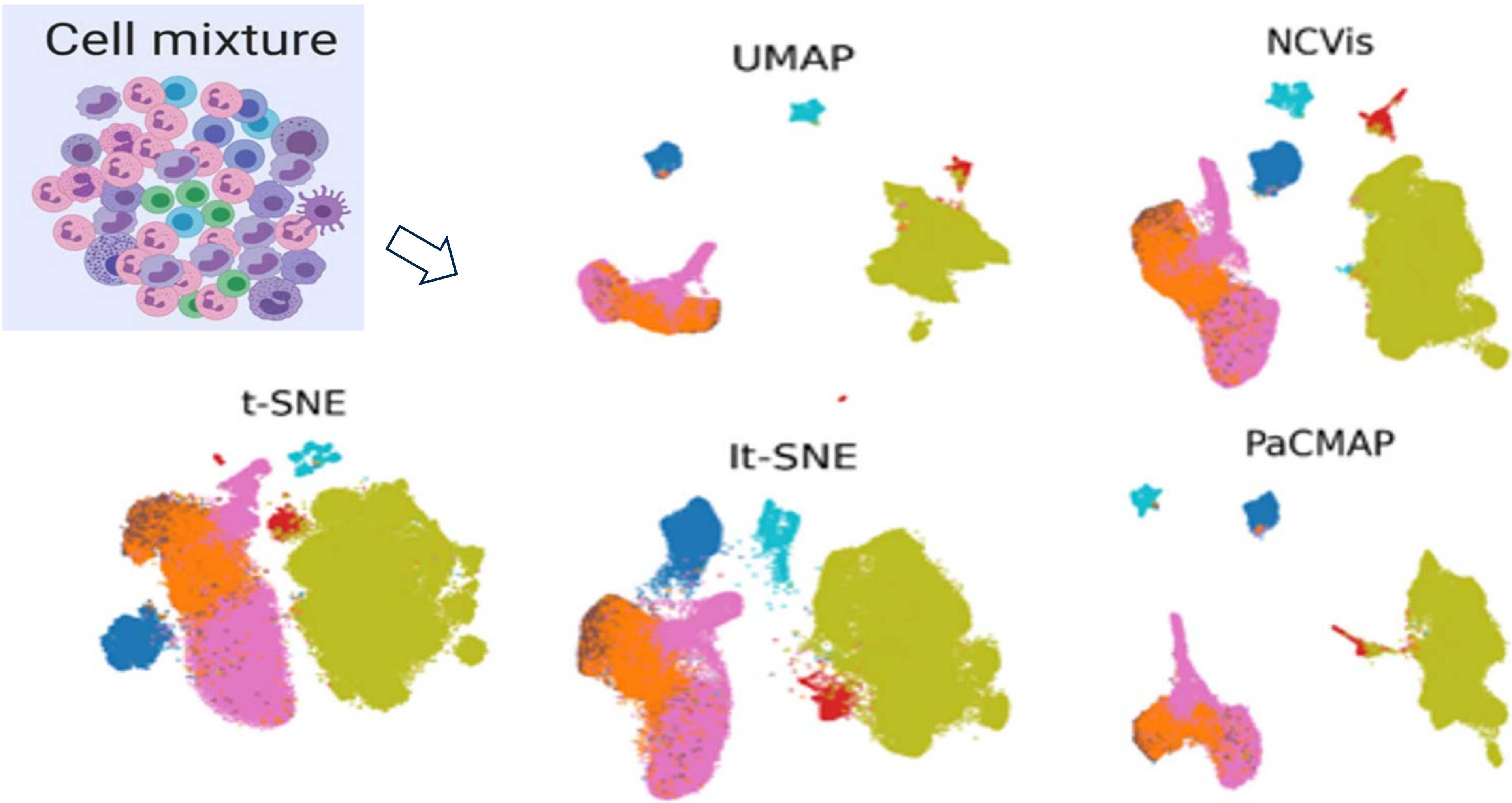
Joint work with Yingfan Wang, Haiyang Huang, Yiyang Sun, Edward Browne and Yaron Shaposhnik, Lesia Semenova, David Murdoch



UMAP
NCVis
t-SNE
It-SNE
PaCMAP


UMAP
NCVis
t-SNE
It-SNE
PaCMAP

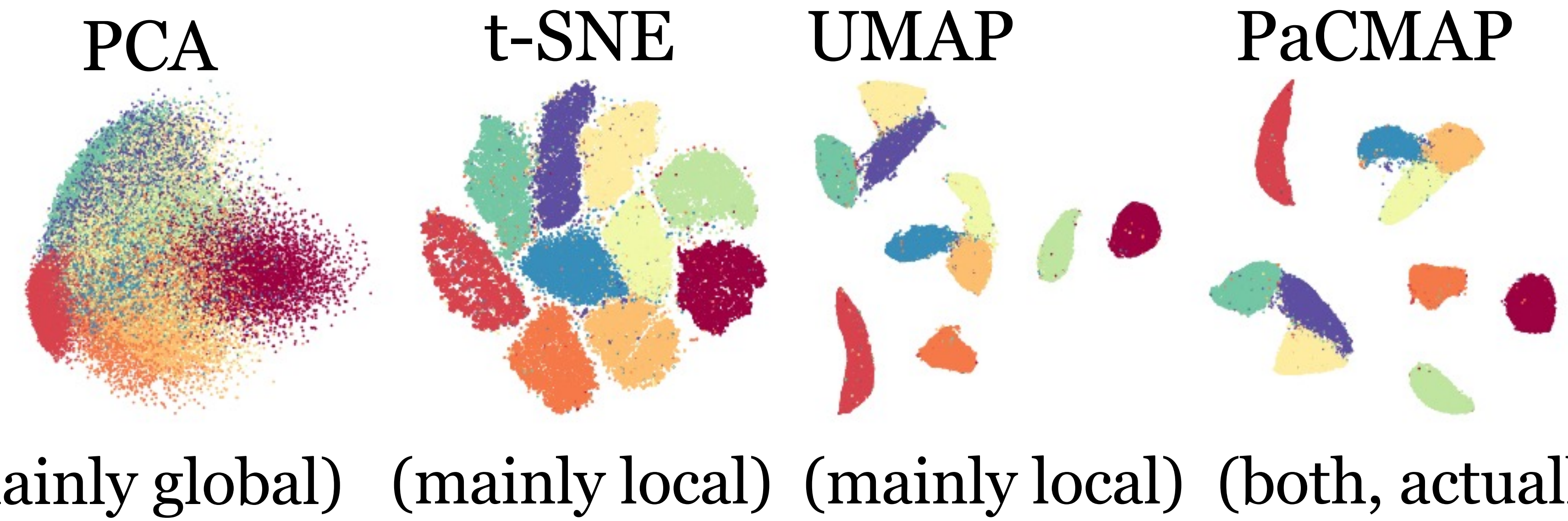
Kazer et. al dataset



Kazer et. al dataset

Local vs Global

- Local structure: local neighborhood graph, nearest neighbors
- Global structure: relationships between clusters, respect relative distances.



Global Methods

- PCA (Pearson, 1901)
- MDS (Torgerson, 1952)

:

Local Methods

- LLE (Roweis and Saul, 2000),
- Isomap (Tenenbaum et al., 2000)
- Hessian Local Linear Embedding (Donoho and Grimes, 2003)
- Laplacian Eigenmaps (Belkin and Niyogi, 2001)
- Stochastic Neighborhood Embedding (SNE) (Hinton and Roweis, 2003)
- t-SNE (van der Maaten and Hinton, 2008)
- LargeVis (Tang et al., 2016)
- UMAP (McInnes et al., 2018)

:

Preserve distances, not neighborhoods

Crowding problem

Preserve neighborhoods

Glol

• PC

• M

:

Local Methods

• LLE (Roweis and Saul, 200

• Isomap (Tenenbaum et al.,

Niyogi, 2001)

• UMAP (McInnes et al., 2

:

Article | Open Access | Published: 28 November 2019

The art of using t-SNE for single-cell transcriptomics

Dmitry Kobak & Philipp Berens

Nature Communications **10**, Article number: 5416 (2019) | Cite this article

How to Use t-SNE Effectively

MARTIN WATTENBERG Google Brain
FERNANDA VIÉGAS Google Brain
IAN JOHNSON Google Cloud

Oct. 13 2016

arXiv.org > cs > arXiv:1708.03229

Search...

Help | Advanced S

Computer Science > Artificial Intelligence

[Submitted on 10 Aug 2017]

Automatic Selection of t-SNE Perplexity

Yanshuai Cao, Luyu Wang

t-Distributed Stochastic Neighbor Embedding (t-SNE) is one of the m
dimensionality reduction methods for data visualization, but it has a
hyperparameter that requires manual selection. In practice, proper tu

Automated optimal parameters for T-distributed stochastic neighbor embedding improve visualization and allow analysis of large datasets

October 2018
DOI: 10.1101/451690
Project: Automated Analysis of Flow Cytometry Multidimensional Datasets

Authors:

Anna C Belkina Boston University
Christopher O. Ciccolella
Rina Anno
Josef Spidlen

How to Use t-SNE Effectively

MARTIN WATTENBERG
Google Brain

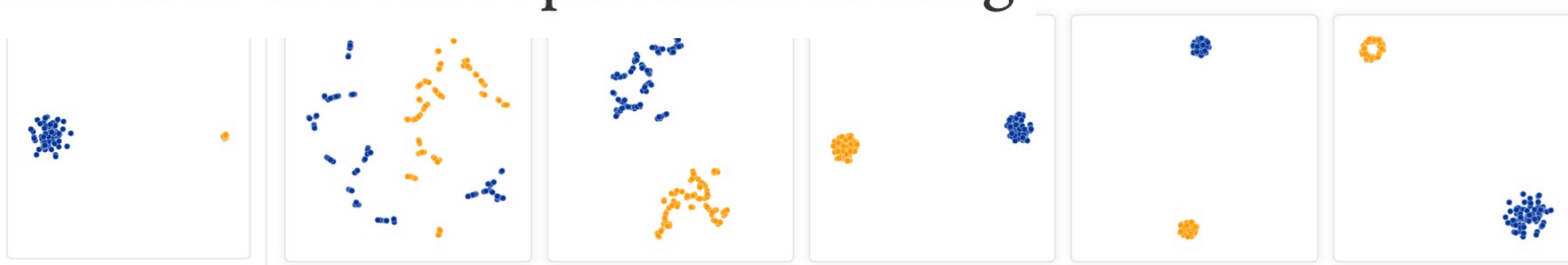
FERNANDA VIÉGAS
Google Brain

IAN JOHNSON
Google Cloud

Oct. 13
2016

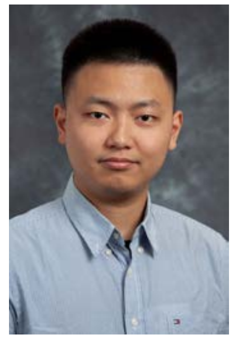
1. Those hyperparameters really matter

2. Cluster sizes in a t-SNE plot mean nothing



PaCMAP

Yingfan Wang
former PhD student, Duke



Haiyang Huang
former PhD student, Duke

Yaron Shaposhnik
Prof, U Rochester

JMLR

Home Page

Papers

Submissions

News

Editorial Board

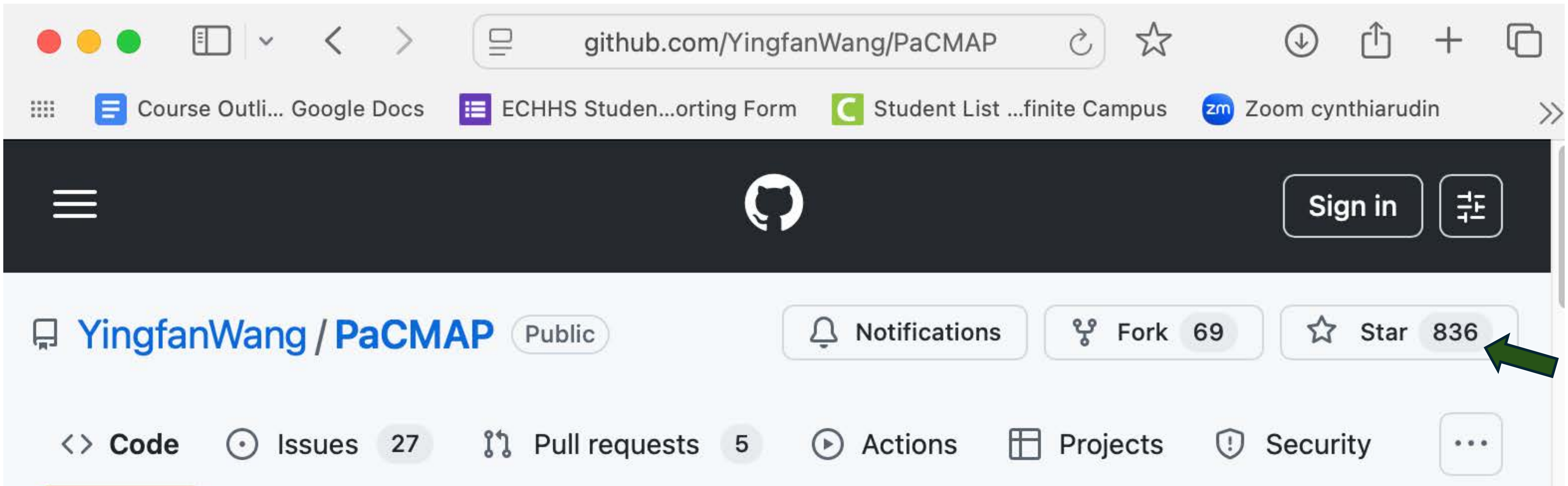
Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization

Yingfan Wang, Haiyang Huang, Cynthia Rudin, Yaron Shaposhnik; 22(201):1−73, 2021.

Abstract

Dimension reduction (DR) techniques such as t-SNE, UMAP, and TriMap have demonstrated impressive visualization performance on many real-world datasets. One tension that has always faced these methods is the trade-off between preservation of global structure and preservation of local structure: these methods can either handle one or the other, but not both. In this work, our main goal is to understand what aspects of DR methods are important for preserving both local and global structure: it is difficult to design

PaCMAP: https://github.com/YingfanWang/PaCMAP
Via pip: pip install pacmap / via conda: conda install pacmap -c conda-forge
Now supports R/Seurat integration



*Winner of the **2023 John M. Chambers Statistical Software Award** and the **2024 Award for Innovation in Statistical Programming and Analytics** from the American Statistical Association

Chapter 9
Machine Vision Applied to Entomology

Gabriel R. Palma, Conor P. Hackett, and Charles Markham

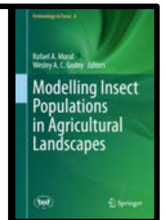
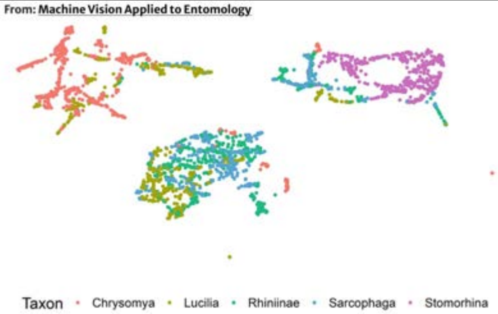


Fig. 9.4

From: Machine Vision Applied to Entomology

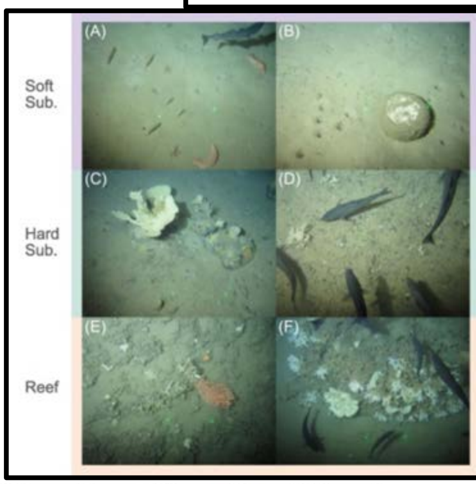


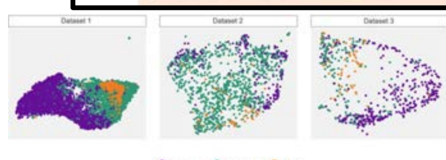
Application of PaCMAP to the medically and forensically important flies dataset based on colour features of the flies specimens using computer vision techniques. The colour represents each taxonomic group of flies presented in the dataset

Ecological Informatics
Volume 81, July 2024, 102619

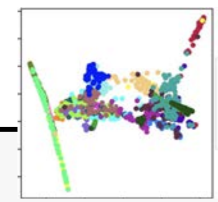
Machine learning for non-experts: A more accessible and simpler approach to automatic benthic habitat classification

Chloe A. Game [a b], Michael B. Thompson [c], Graham D. Finlayson [b]





Front. Mar. Sci., 08 August 2024
Sec. Marine Megafauna
Volume 11 - 2024 | https://doi.org/10.3389/fmars.2024.1416247



Unsupervised identification of Greater Caribbean manatees using Scattering Wavelet Transform and Hierarchical Density Clustering from underwater bioacoustics recordings

Fernando Merchan[1] Kenji Contreras[1] Héctor Poveda[1] Hector M. Guzman[2]
Javier E. Sanchez-Galan[3*]

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

Article | Open access | Published: 19 December 2023

Neoadjuvant durvalumab plus radiation versus durvalumab alone in stages I–III non-small cell lung cancer: survival outcomes and molecular correlates of a randomized phase II trial

Nasser K. Altorki, Zachary H. Walsh, Johannes C. Melms, Jeffery L. Port, Benjamin E. Lee, Abu Nasar, Cathy Spinelli, Lindsay Caprio, Meri Rogava, Patricia Ho, Paul J. Christos, Ashish Saxena, Olivier Elemento, Bhavneet Bhinder, Casey Ager, Amit Dipak Amin, Nicholas J. Sanfilippo, Vivek Mittal, Alain C. Borczuk, Silvia C. Formenti, Benjamin Izar & Timothy E. McGraw

Science

Current Issue First release papers Archive About ⌄ Submit mar

HOME > SCIENCE > VOL. 384, NO. 6700 > CELLULAR ARCHITECTURE SHAPES THE NAÏVE T CELL RESPONSE

RESEARCH ARTICLE | IMMUNOLOGY

Cellular architecture shapes the naïve T cell response

BENJAMIN D. HALE, YANNIK SEVERIN, FABIENNE GRAEBNITZ, DOMINIQUE STARK, DANIEL GUIGNARD, JULIEN MENA, YASMIN FESTL, SOHYON LEE, JACOB HANIMANN, NATHAN S. ZANGGER, MICHELLE MEIER, DAVID GOSLINGS, OLGA LAMPRECHT, BEAT M. FREY, ANNETTE OXENIUS, AND BEREND SNIJDER fewer Authors Info & Affiliations

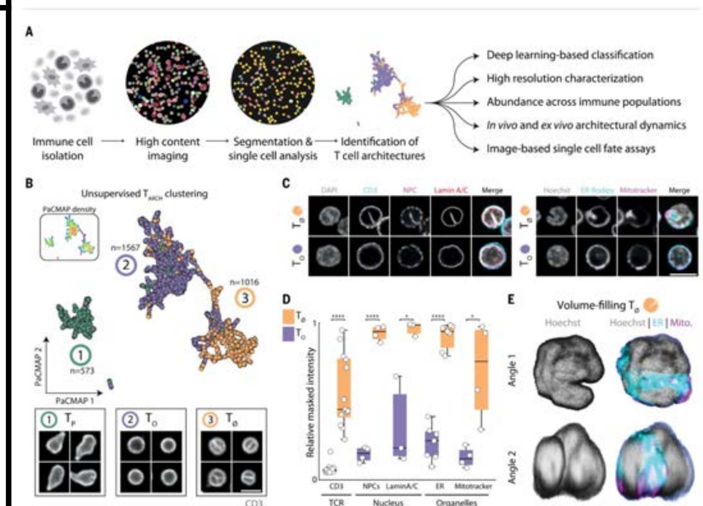


Fig. 1. Discovery and characterization of primary T_{ARCH} heterogeneity.
(**A**) Workflow for T_{ARCH} discovery and characterization. (**B**) Unsupervised clustering of human peripheral T cell phenotypes. TopOMetry embedding and subsequent PaCMAP dimensionality reduction of the 384-di-

nature biotechnology

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > nature biotechnology > articles > article

Article | Published: 23 May 2024

Mapping variant effects on anti-tumor hallmarks of primary human T cells with base-editing screens

Zachary H. Walsh, Parin Shah, Neeharika Kothapalli, Shivem B. Shah, Gergo Nikolenyi, D. Zack Brodtman, Giuseppe Leuzzi, Meri Rogava, Michael Mu, Patricia Ho, Sinan Abuzaid, Neil Vasan, Mohammed AlQuraishi, Joshua D. Milner, Alberto Ciccia, Johannes C. Melms & Benjamin Izar

many more…

Algorithm	Graph components and Loss function
t-SNE (van der Maaten & Hinton, 2008)	Graph components: Edges (i,j) $\text{Loss}_{i,j}^{\text{t-SNE}} = p_{ij} \log \frac{p_{ij}}{q_{ij}}$, where $q_{ij} = \frac{\left(1+\lVert\mathbf{y}_i-\mathbf{y}_j\rVert^2\right)^{-1}}{\sum_{k\neq l}(1+\lVert\mathbf{y}_k-\mathbf{y}_l\rVert^2)^{-1}}$ where p_{ij} is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and other $\mathbf{x}_\ell$'s.
UMAP (McInnes et al., 2018)	Graph components: Edges (i,j) $\text{Loss}_{i,j}^{\text{UMAP}} = \begin{cases} \bar{w}_{i,j} \log\left(1 + a\left(\lVert\mathbf{y}_i - \mathbf{y}_j\rVert_2^2\right)^b\right)^{-1} & i,j \text{ neighbors} \\ (1-\bar{w}_{i,j}) \log\left(1 - \left(1 + a\left(\lVert\mathbf{y}_i - \mathbf{y}_j\rVert_2^2\right)^b\right)^{-1}\right) & \text{otherwise,} \end{cases}$ where $\bar{w}_{i,j}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and nearby $\mathbf{x}_\ell$'s.
TriMAP (Amid & Warmuth, 2019)	Graph components: Triplets (i,j,k) where $\text{Distance}_{i,j} \leq \text{Distance}_{i,k}$ $\text{Loss}_{i,j,k}^{\text{TM}} = \omega_{i,j,k} \frac{s(\mathbf{y}_i,\mathbf{y}_k)}{s(\mathbf{y}_i,\mathbf{y}_j)+s(\mathbf{y}_i,\mathbf{y}_k)}$, where $s(\mathbf{y}_i,\mathbf{y}_j) = \left(1 + \lVert\mathbf{y}_i - \mathbf{y}_j\rVert^2\right)^{-1}$ and $\omega_{i,j,k}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ and nearby points.

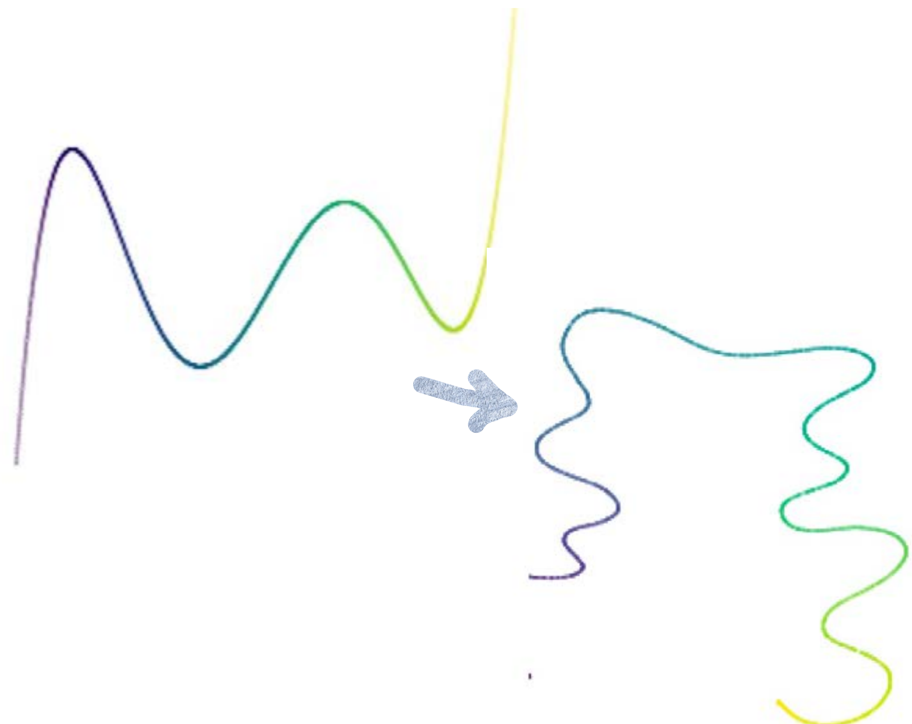
Hard to understand what's important here…

Start from the obvious:

- Attraction: high-dimensional neighbors should be attracted.
- Repulsion: points far in original space should be far in low-dim space.

But that's not enough...

$$\underbrace{\sum_{(i,j)\in\mathcal{T}_{\text{neighbors}}} l^{\text{attract}}(i,j)}_{\text{Attract neighbors}} + \underbrace{\sum_{(i,k)\in\mathcal{T}_{\text{further}}} l^{\text{repulse}}(i,k)}_{\text{Repulse far points}}$$



$$\sum_{\substack{\text{graph} \\ \text{components } \{i\}}} \text{Weight(component } i \text{ in high dim space)} \cdot \text{Loss(component } i \text{ in low dim space)}$$

PaCMAP's ideas :

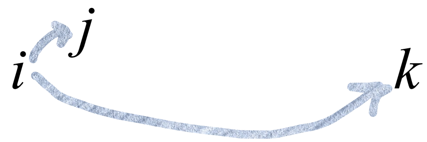
- Properties of the loss function determine local structure.
- The choice of graph components determines global structure.

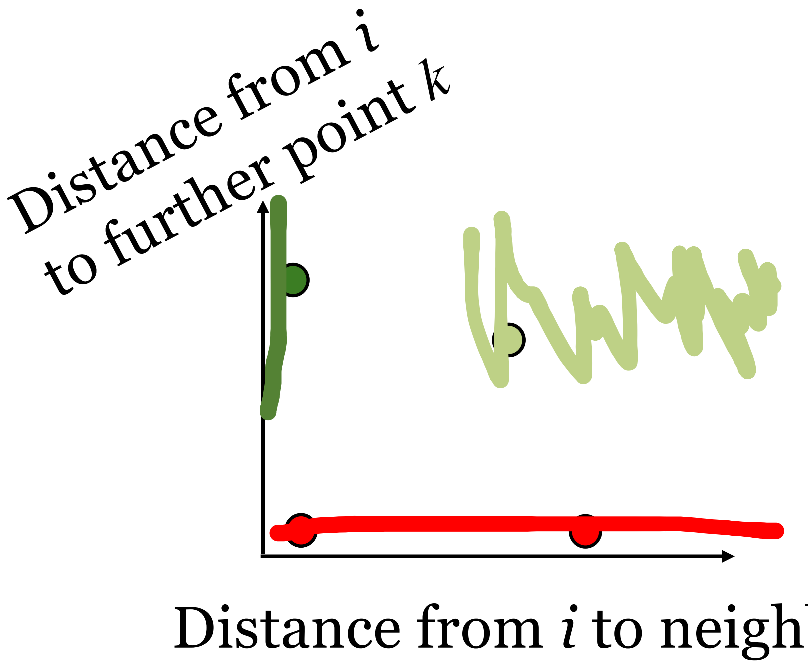
Algorithm	Graph components and Loss function
t-SNE (van der Maaten & Hinton, 2008)	Graph components: Edges (i,j) $\text{Loss}_{i,j}^{\text{t-SNE}} = p_{ij} \log \frac{p_{ij}}{q_{ij}}$, where $q_{ij} = \frac{(1+\|\mathbf{y}_i-\mathbf{y}_j\|^2)^{-1}}{\sum_{k\neq l}(1+\|\mathbf{y}_k-\mathbf{y}_l\|^2)^{-1}}$ where p_{ij} is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and other $\mathbf{x}_\ell$'s.
UMAP (McInnes et al., 2018)	Graph components: Edges (i,j) $\text{Loss}_{i,j}^{\text{UMAP}} = \begin{cases} \bar{w}_{i,j} \log \left(1 + a \left(\|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right)^b\right)^{-1} & i,j \text{ neighbors} \\ (1-\bar{w}_{i,j}) \log \left(1 - \left(1 + a \left(\|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right)^b\right)^{-1}\right) & \text{otherwise,} \end{cases}$ where $\bar{w}_{i,j}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and nearby $\mathbf{x}_\ell$'s.
TriMAP (Amid & Warmuth, 2019)	Graph components: Triplets (i,j,k) where $\text{Distance}_{i,j} \leq \text{Distance}_{i,k}$ $\text{Loss}_{i,j,k}^{\text{TM}} = \omega_{i,j,k} \frac{s(\mathbf{y}_i,\mathbf{y}_k)}{s(\mathbf{y}_i,\mathbf{y}_j)+s(\mathbf{y}_i,\mathbf{y}_k)}$, where $s(\mathbf{y}_i,\mathbf{y}_j) = (1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}$ and $\omega_{i,j,k}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ and nearby points.

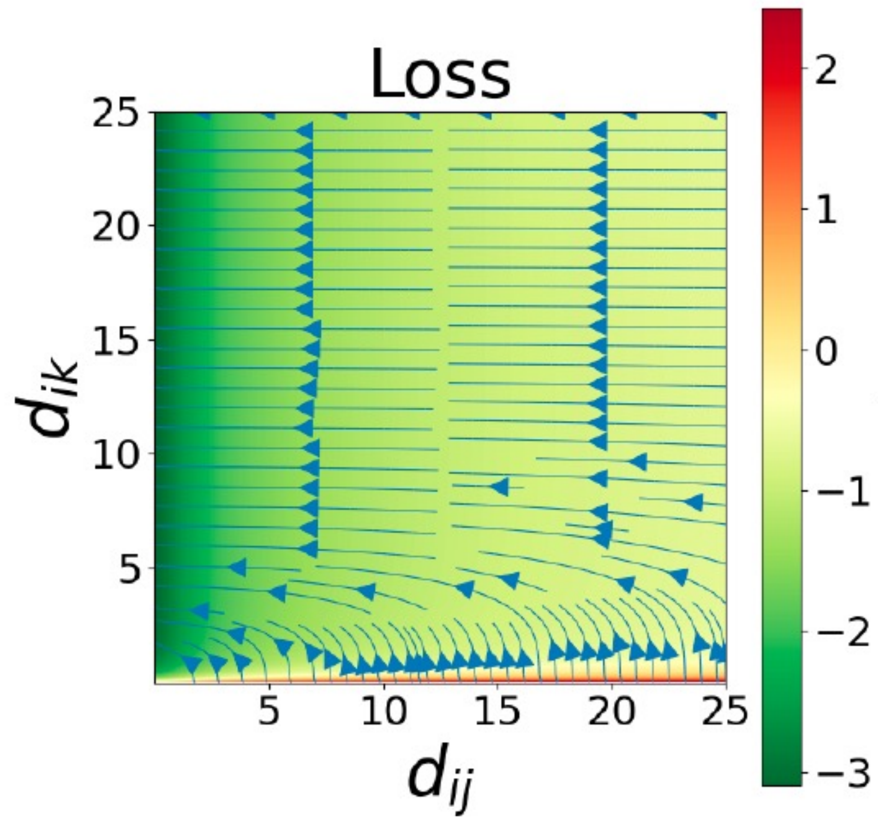
Hard to understand what's important here…

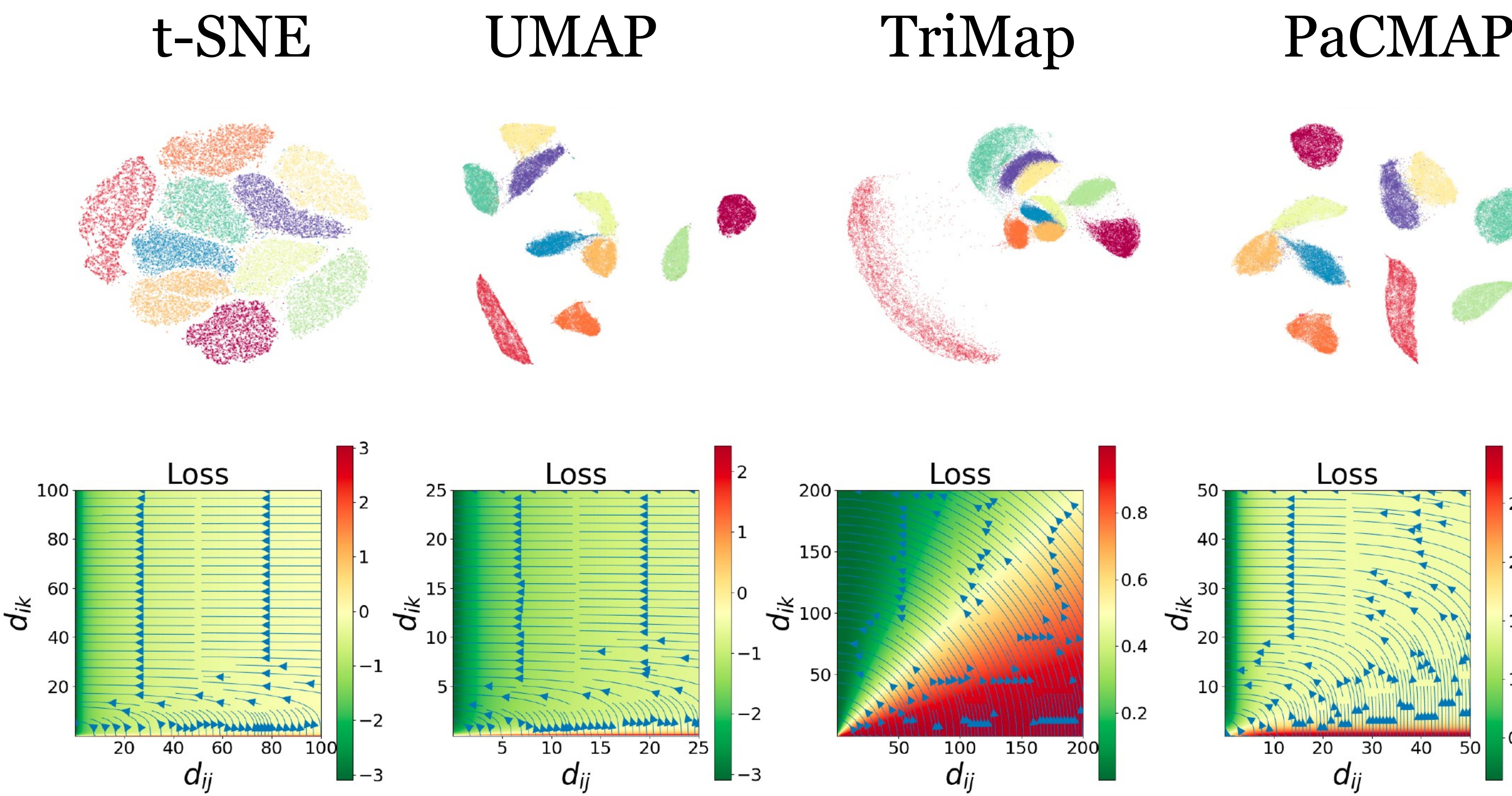
The “rainbow” plot

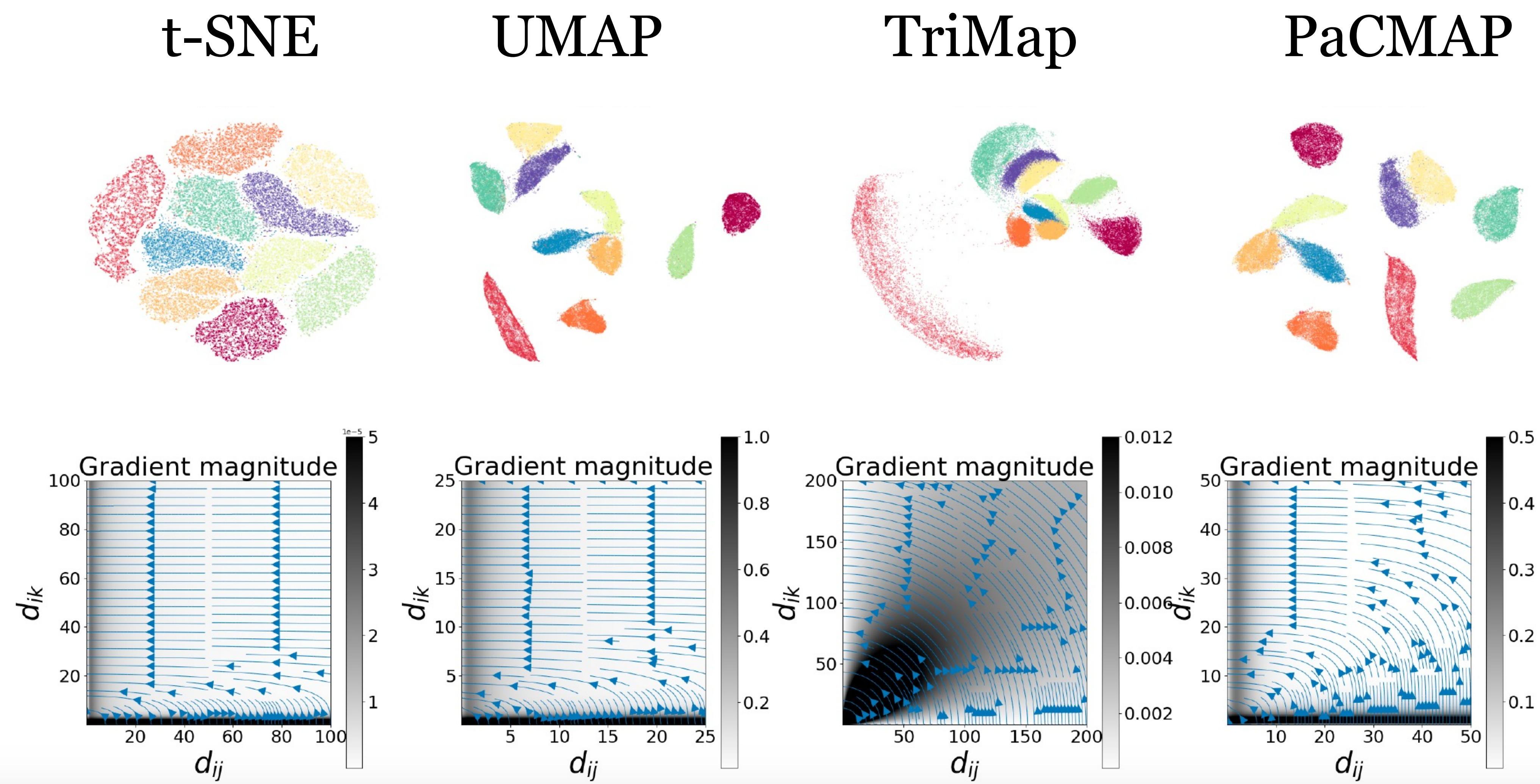
Triple i, j (neighbor), k (further)





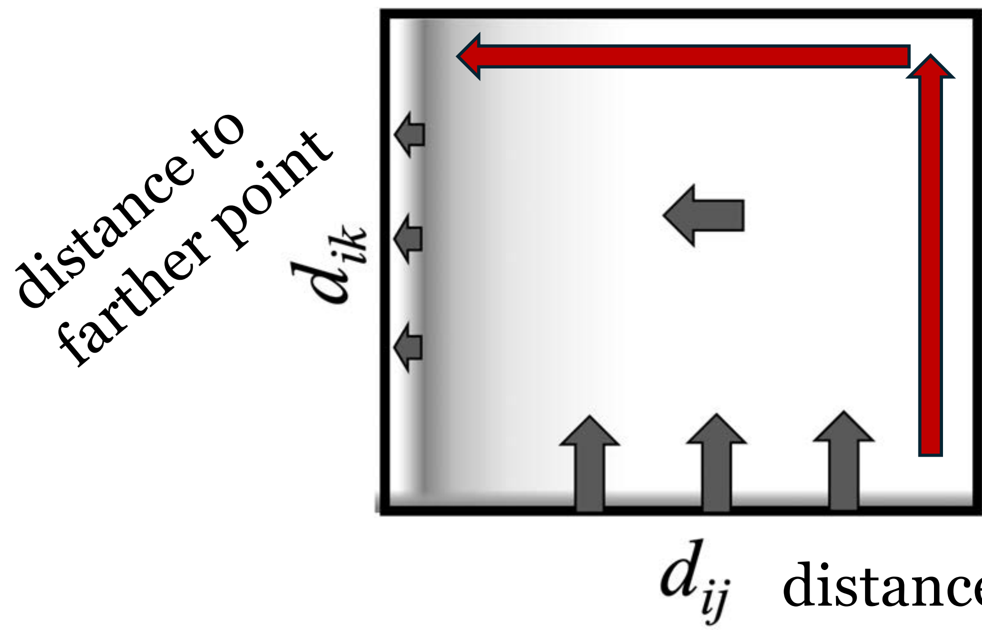



t-SNE
UMAP
TriMap
PaCMAP
Loss
Loss
Loss
Loss
d_{ik}
d_{ij}



t-SNE
UMAP
TriMap
PaCMAP
Gradient magnitude
Gradient magnitude
Gradient magnitude
Gradient magnitude
d_{ik}
d_{ij}

Our principles for a good loss function

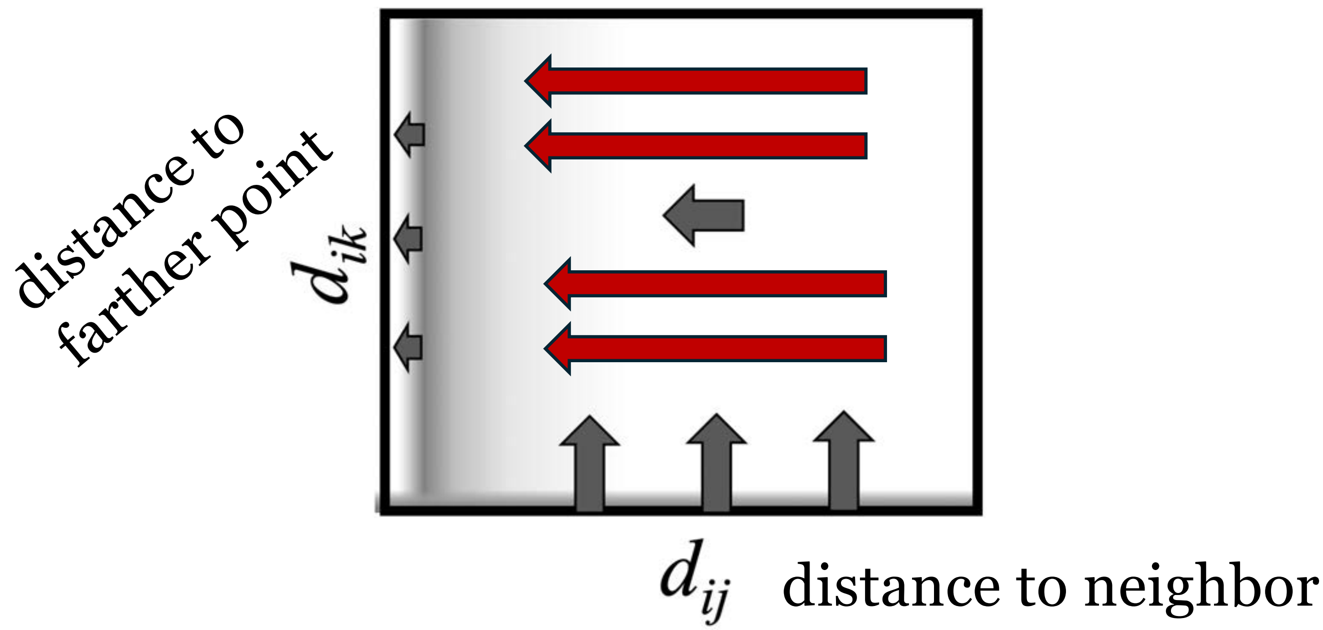


1. Monotonicity

Up, Left

$$\forall d_{ij} : \frac{\partial Loss}{\partial d_{ik}} \leq 0 \text{ and } \forall d_{ik} : \frac{\partial Loss}{\partial d_{ij}} \geq 0.$$

Our principles for a good loss function

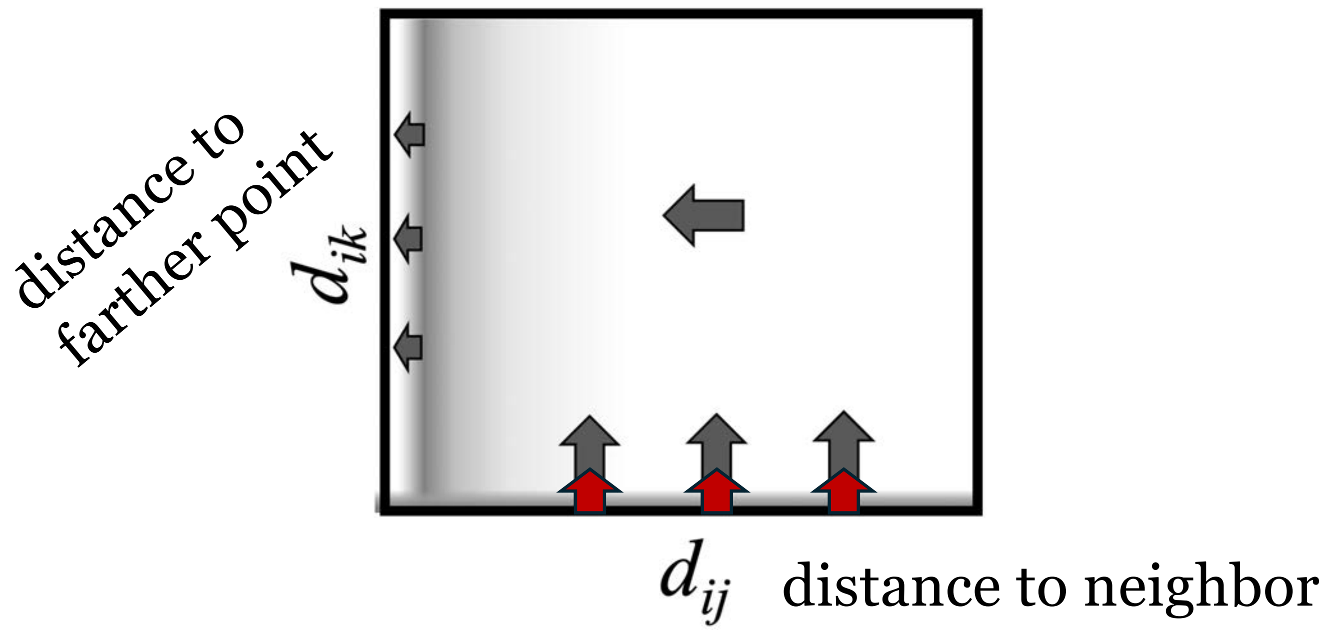


2. Except at bottom, gradient should go mainly to the left.
(if further point is sufficiently far, should focus on pulling neighbor closer.)

$$\forall d_{ij}, \forall \epsilon > 0,\ \exists \theta_{ik}^{\epsilon} : \forall d_{ik} > \theta_{ik}^{\epsilon} \text{ we have } \left| \frac{\partial Loss}{\partial d_{ik}} / \frac{\partial Loss}{\partial d_{ij}} \right| < \epsilon.$$

(if sufficiently far) (left gradient is bigger)

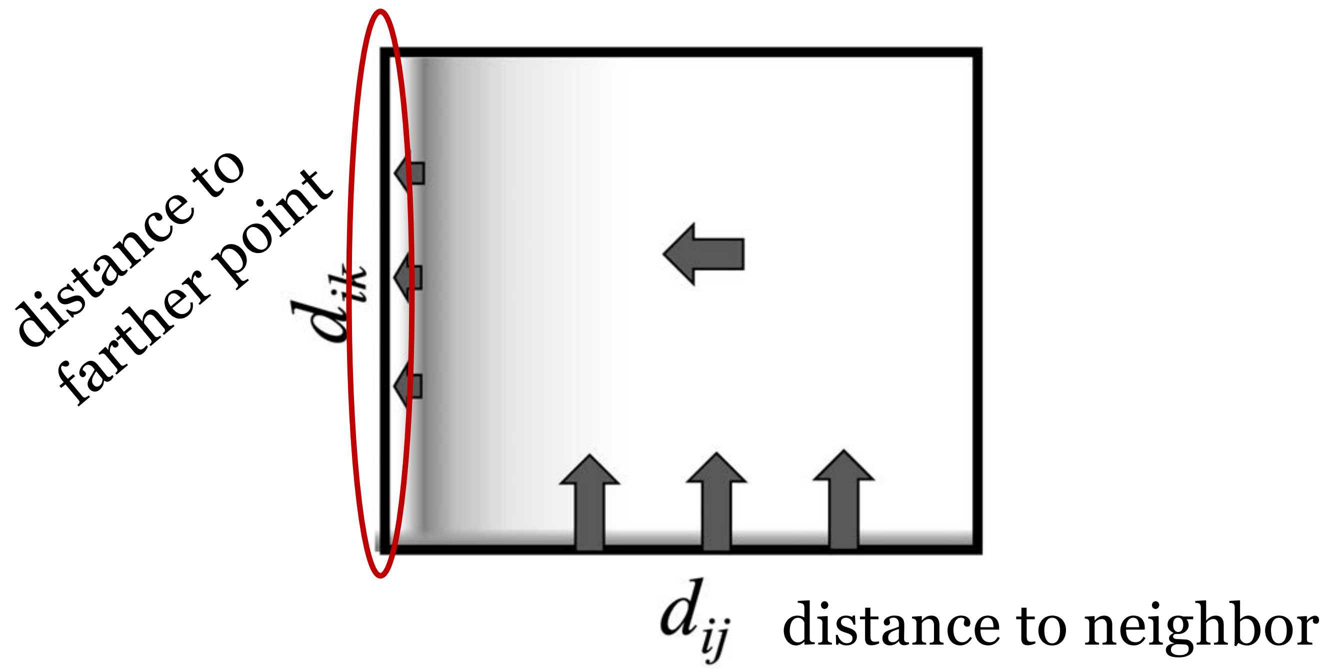
Our principles for a good loss function



3. At bottom, gradient goes up, not to the left (push further points away more than attracting neighbors.)

$$\forall d_{ik} > 0,\ \forall \epsilon > 0,\ \exists \theta_{ij}^{\epsilon} : \forall d_{ij} > \theta_{ij}^{\epsilon} \text{ we have } \left| \frac{\partial Loss}{\partial d_{ij}} / \frac{\partial Loss}{\partial d_{ik}} \right| < \epsilon.$$

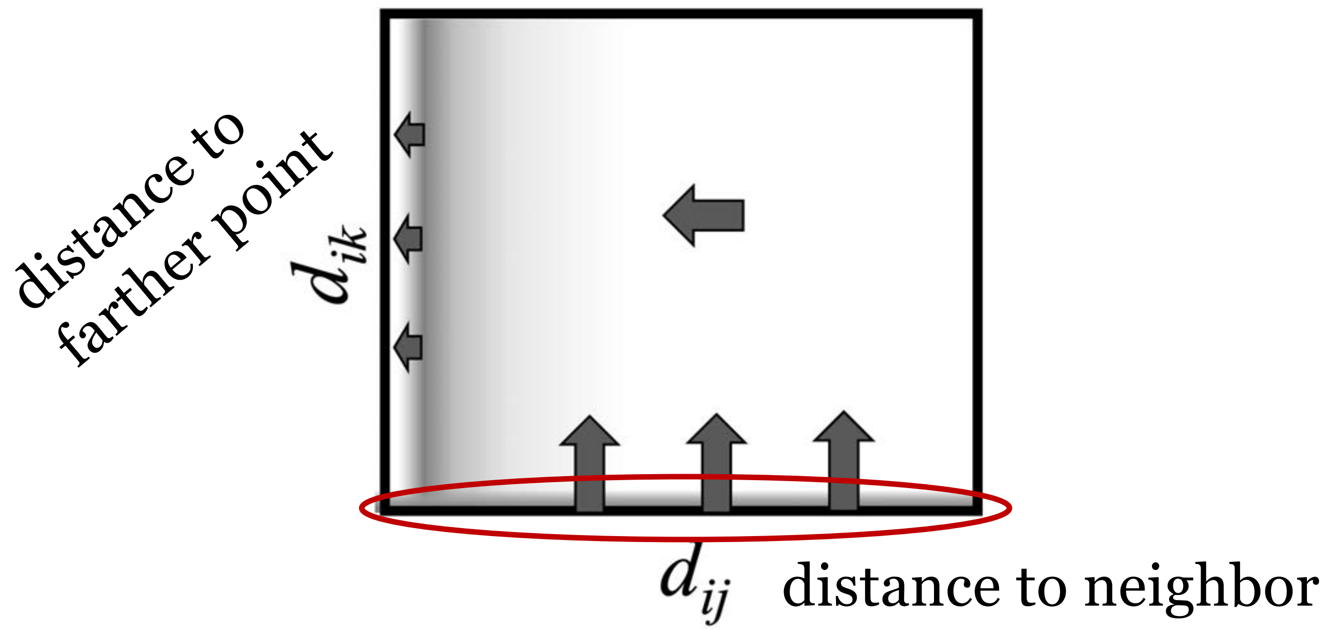
Our principles for a good loss function



4. At left, gradient has small magnitude (don't crowd).

$$\forall \epsilon > 0,\ \exists \theta_{ik}^{\epsilon} : \forall d_{ik} \geq \theta_{ik}^{\epsilon},\ \lim_{d_{ij} \to 0} \left| \frac{\partial Loss}{\partial d_{ik}} \right| < \epsilon, \text{ and } \lim_{d_{ij} \to 0} \left| \frac{\partial Loss}{\partial d_{ij}} \right| < \epsilon.$$

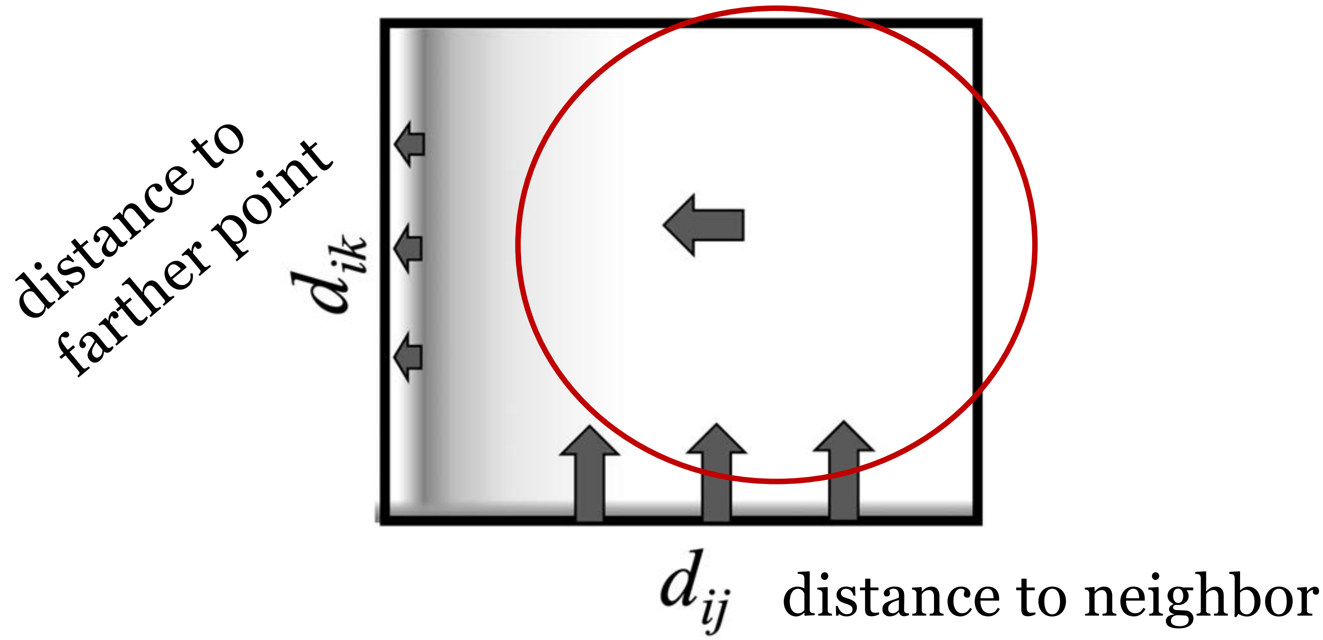
Our principles for a good loss function



5. At bottom, gradient has large magnitude (push farther point away)

$$\forall d_{ij},\ \exists \theta_{ik}\ \forall d_{ik} > \theta_{ik} : \left|\frac{\partial Loss}{\partial d_{ik}}\right|^2 + \left|\frac{\partial Loss}{\partial d_{ij}}\right|^2 \text{ is non-increasing in } d_{ik}.$$

Our principles for a good loss function



6. Gradient fades as neighbor gets farther away.
   (give up on neighbors when they are too far)

$$\forall \epsilon > 0,\ \exists \theta_{ik} : \forall d_{ik} \geq \theta_{ik},\ \lim_{d_{ij} \to \infty} \left| \frac{\partial Loss}{\partial d_{ik}} \right| < \epsilon, \ \text{and} \ \lim_{d_{ij} \to \infty} \left| \frac{\partial Loss}{\partial d_{ij}} \right| < \epsilon.$$

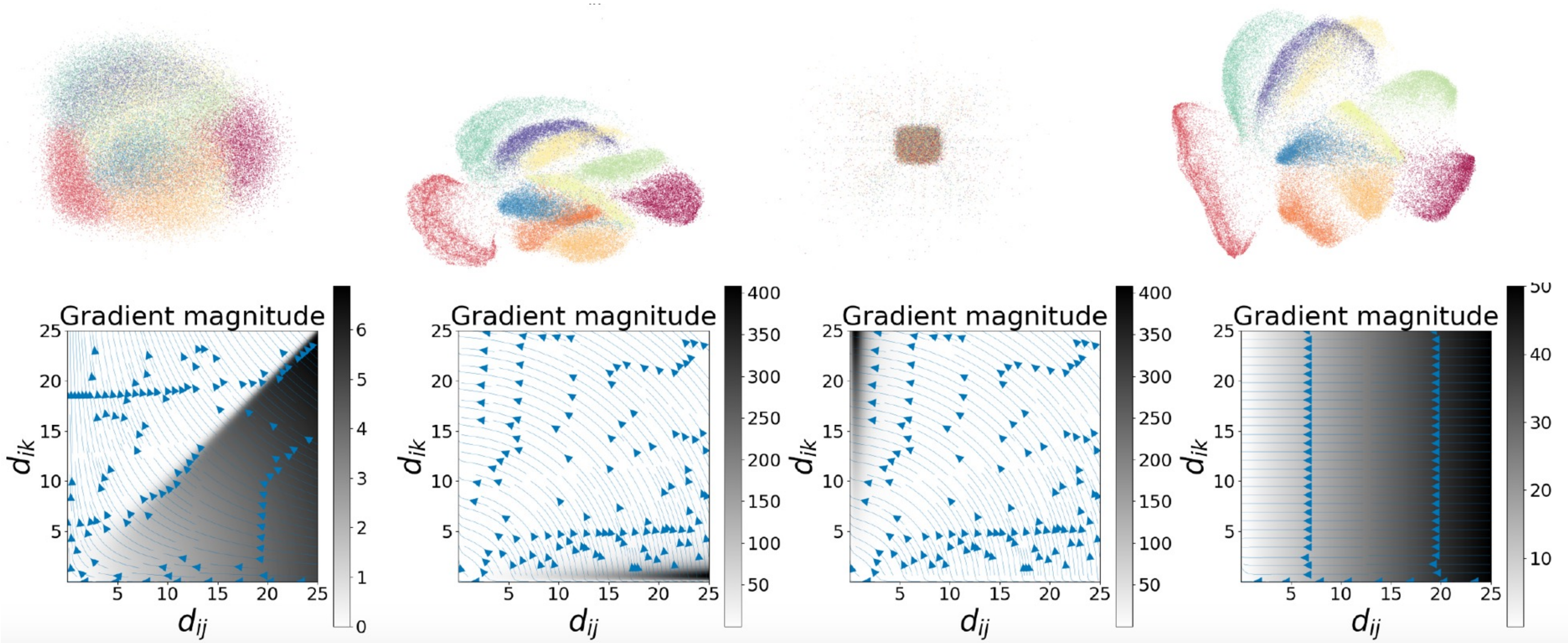
Bad loss functions

$$Loss = log\left(1 + exp\left(\frac{d_{ij}^2 - d_{ik}^2}{10}\right)\right)$$

$$Loss = \frac{d_{ij}^2 + 1}{d_{ik}^2 + 1}$$

$$Loss = -\frac{d_{ik}^2 + 1}{d_{ij}^2 + 1}$$

$$Loss = log\left(1 + exp(d_{ij}^2) + exp(-d_{ik}^2)\right)$$



Algorithm	Graph components and Loss function
t-SNE (van der Maaten & Hinton, 2008)	Graph components: Edges (i,j) $\mathrm{Loss}_{i,j}^{\text{t-SNE}} = p_{ij} \log \frac{p_{ij}}{q_{ij}}$, where $q_{ij} = \frac{(1+\|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k\neq l}(1+\|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}}$ where p_{ij} is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and other $\mathbf{x}_\ell$'s.
UMAP (McInnes et al., 2018)	Graph components: Edges (i,j) $\mathrm{Loss}_{i,j}^{\text{UMAP}} = \begin{cases} \bar{w}_{i,j} \log \left(1 + a \left(\|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right)^b\right)^{-1} & i,j \text{ neighbors} \\ (1-\bar{w}_{i,j}) \log \left(1 - \left(1 + a \left(\|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right)^b\right)^{-1}\right) & \text{otherwise,} \end{cases}$ where $\bar{w}_{i,j}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$ and nearby $\mathbf{x}_\ell$'s.
TriMAP (Amid & Warmuth, 2019)	Graph components: Triplets (i,j,k) where $\mathrm{Distance}_{i,j} \leq \mathrm{Distance}_{i,k}$ $\mathrm{Loss}_{i,j,k}^{\text{TM}} = \omega_{i,j,k} \frac{s(\mathbf{y}_i,\mathbf{y}_k)}{s(\mathbf{y}_i,\mathbf{y}_j)+s(\mathbf{y}_i,\mathbf{y}_k)}$, where $s(\mathbf{y}_i,\mathbf{y}_j) = \left(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2\right)^{-1}$ and $\omega_{i,j,k}$ is a function of $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ and nearby points.

Hard to understand what's important here…

PaCMAP's Loss

$$\text{Loss}^{\text{PaCMAP}} = w_{\text{neighbors}}\text{Loss}_{\text{neighbors}} + w_{MN}\text{Loss}_{MN} + w_{FP}\text{Loss}_{FP}$$

$$\text{distance}\,(i,j) := \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2 + 1$$

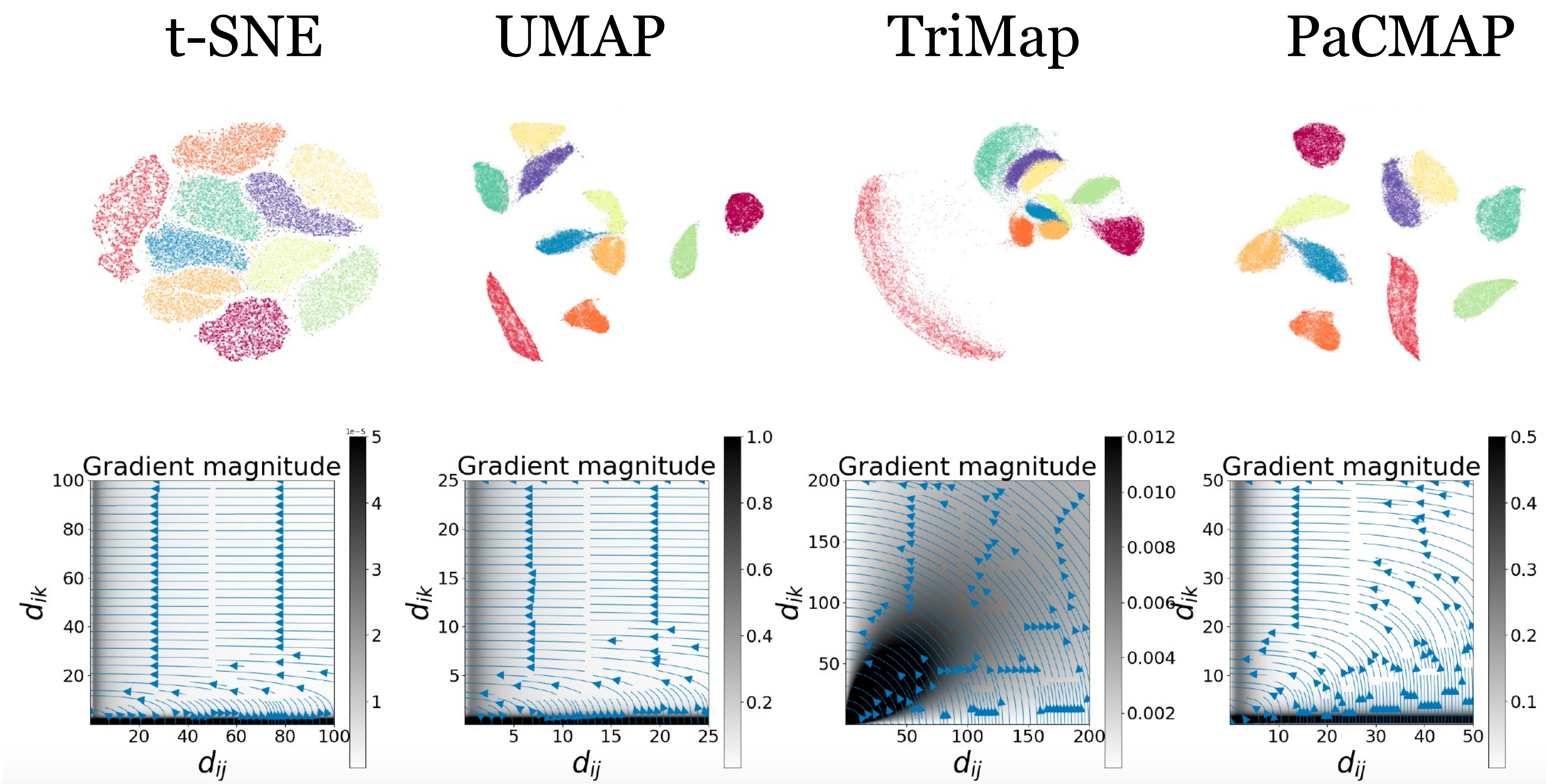
$$\text{Loss}_{\text{neighbors}} = \frac{\text{distance}\,(i,j)}{\text{distance}\,(i,j)+10} \quad \text{Loss}_{MN} = \frac{\text{distance}\,(i,l)}{\text{distance}\,(i,l)+10000} \quad \text{Loss}_{FP} = \frac{1}{\text{distance}\,(i,l)+1}$$

Neighbors: attractive

Mid-near pairs: mild attractive

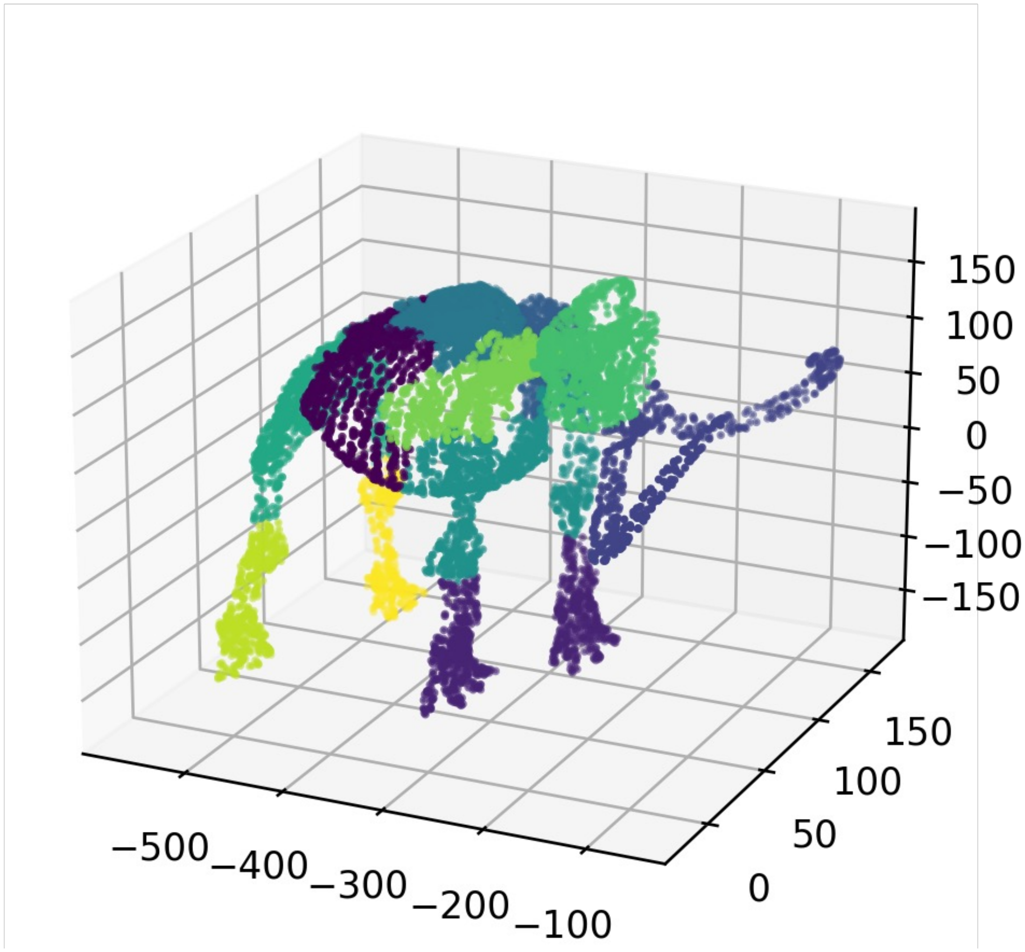
Further points: repulsive

Mid near pair for i : sample 6 points, choose the second closest, pair it with i.

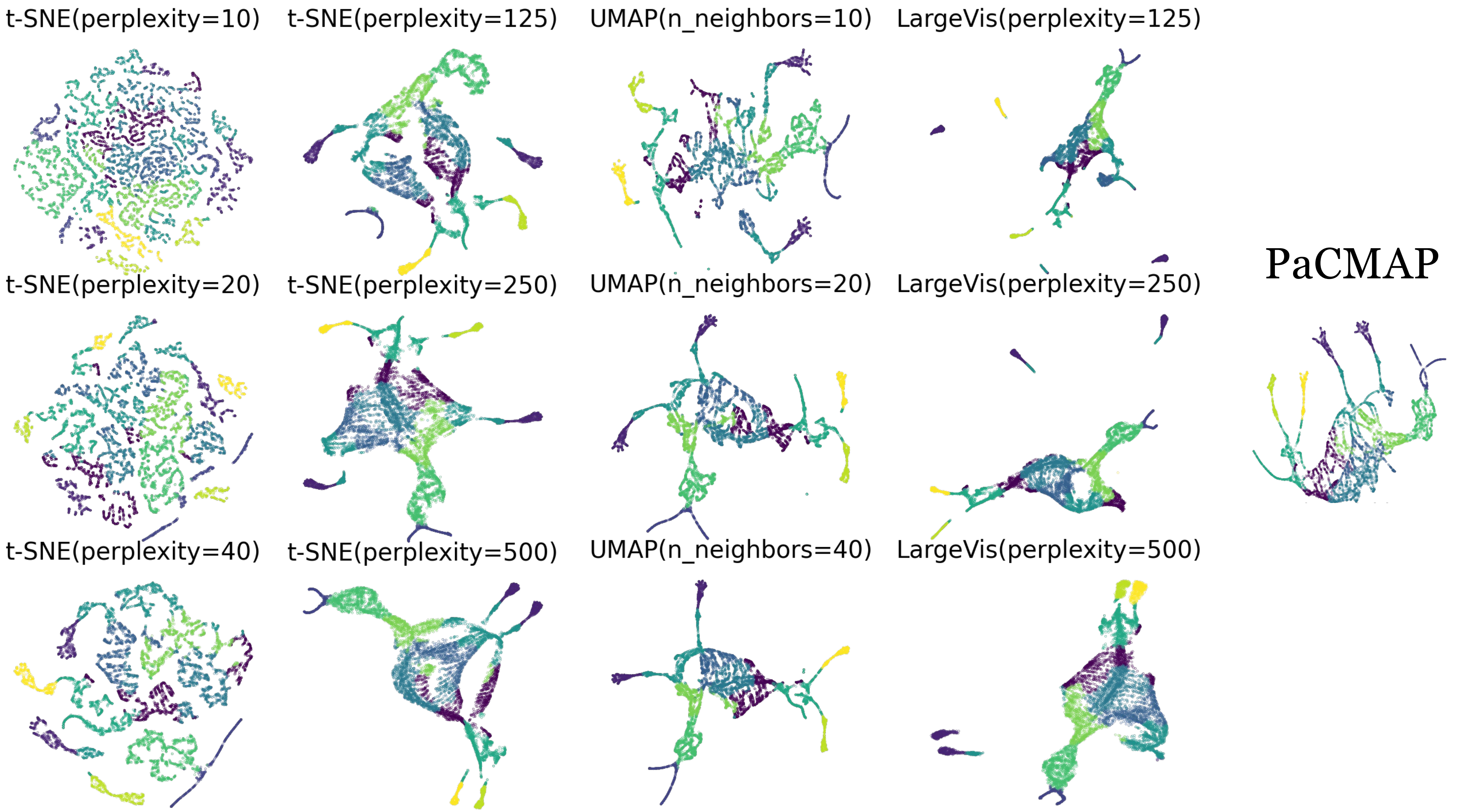


t-SNE
UMAP
TriMap
PaCMAP
Gradient magnitude
d_{ik}
d_{ij}

Original Mammoth



Task: 3d to 2d.


t-SNE(perplexity=10)
t-SNE(perplexity=125)
UMAP(n_neighbors=10)
LargeVis(perplexity=125)
t-SNE(perplexity=20)
t-SNE(perplexity=250)
UMAP(n_neighbors=20)
LargeVis(perplexity=250)
PaCMAP
t-SNE(perplexity=40)
t-SNE(perplexity=500)
UMAP(n_neighbors=40)
LargeVis(perplexity=500)

Studying Clusters of People with HIV

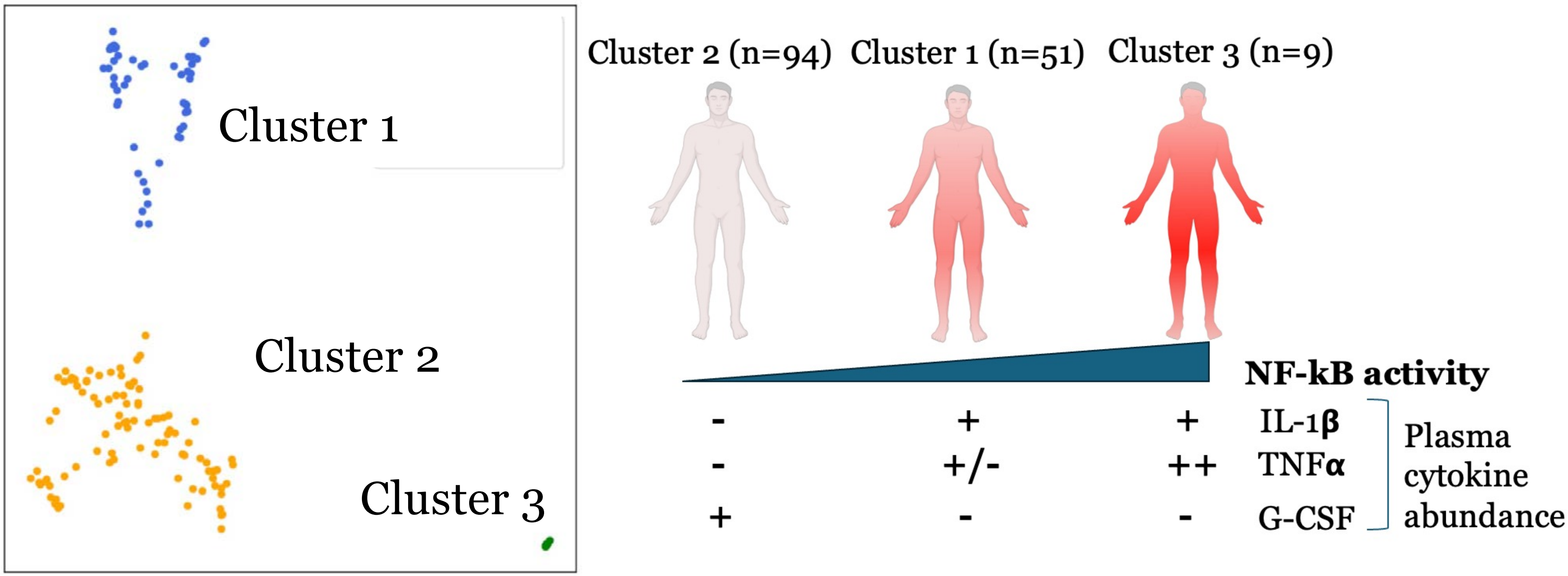
Yingfan Wang
AWS

Prof. Lesia Semenova
Rutgers

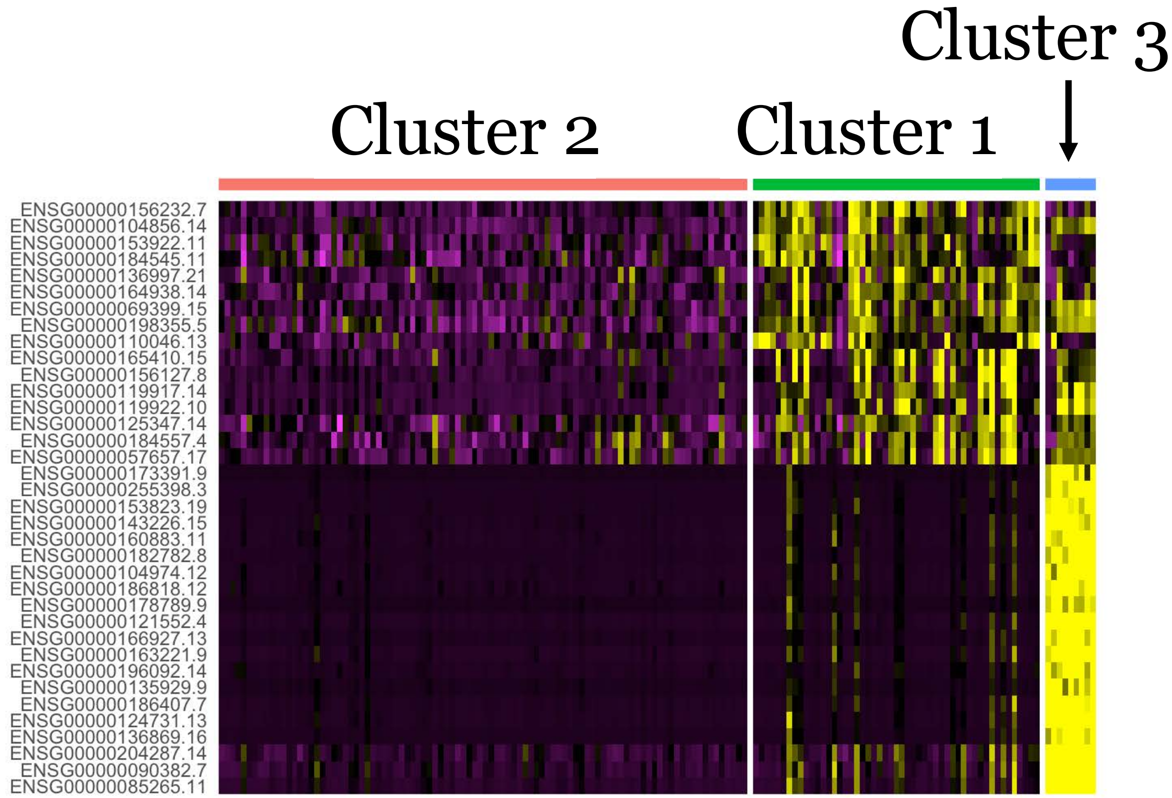
Prof. Edward Browne
University of North Carolina, Chapel Hill
Division of Infectious Diseases

Prof. David Murdoch
Pulmonary, Allergy, and Critical Care Medicine
Duke University

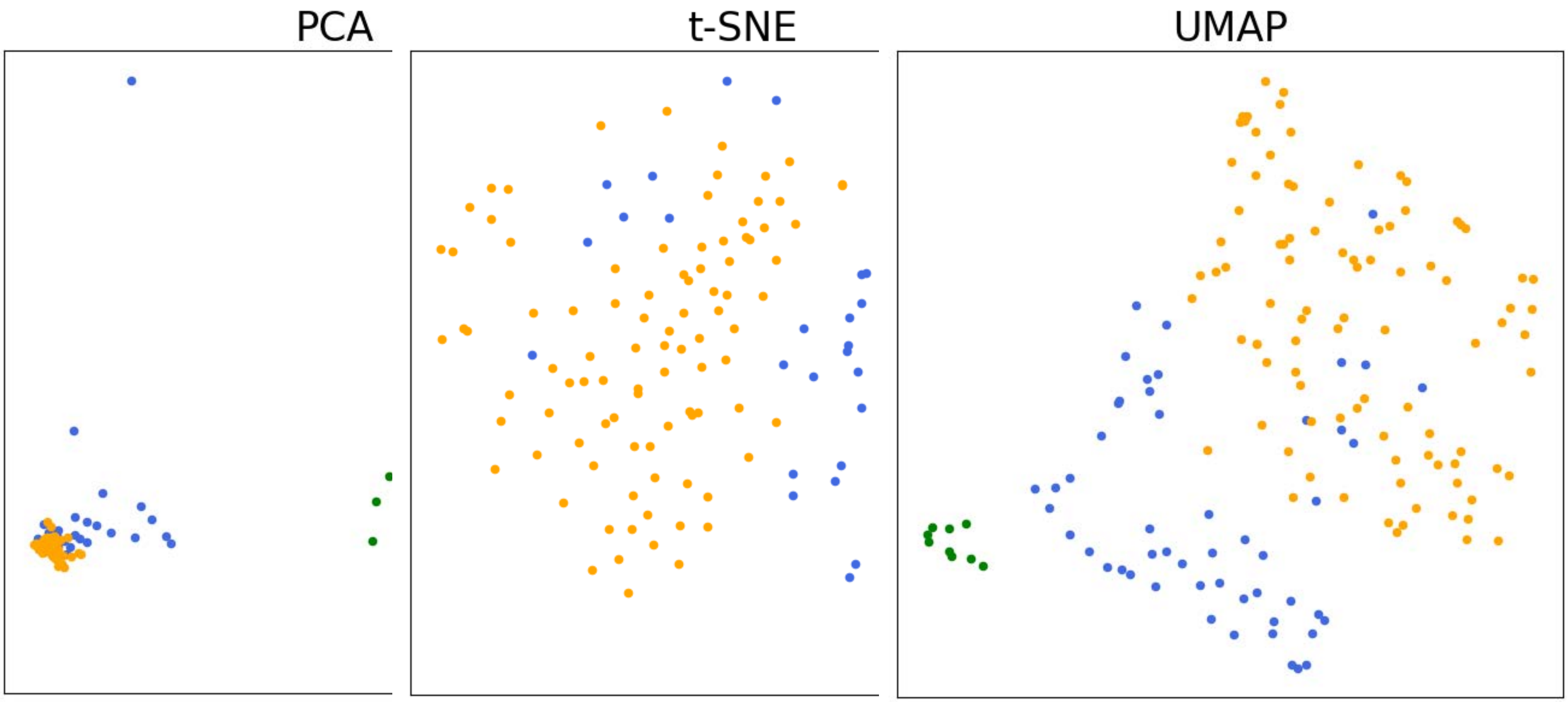
Clustering of people with HIV associated with differential expression of genes regulated by the transcription factor NF-κB



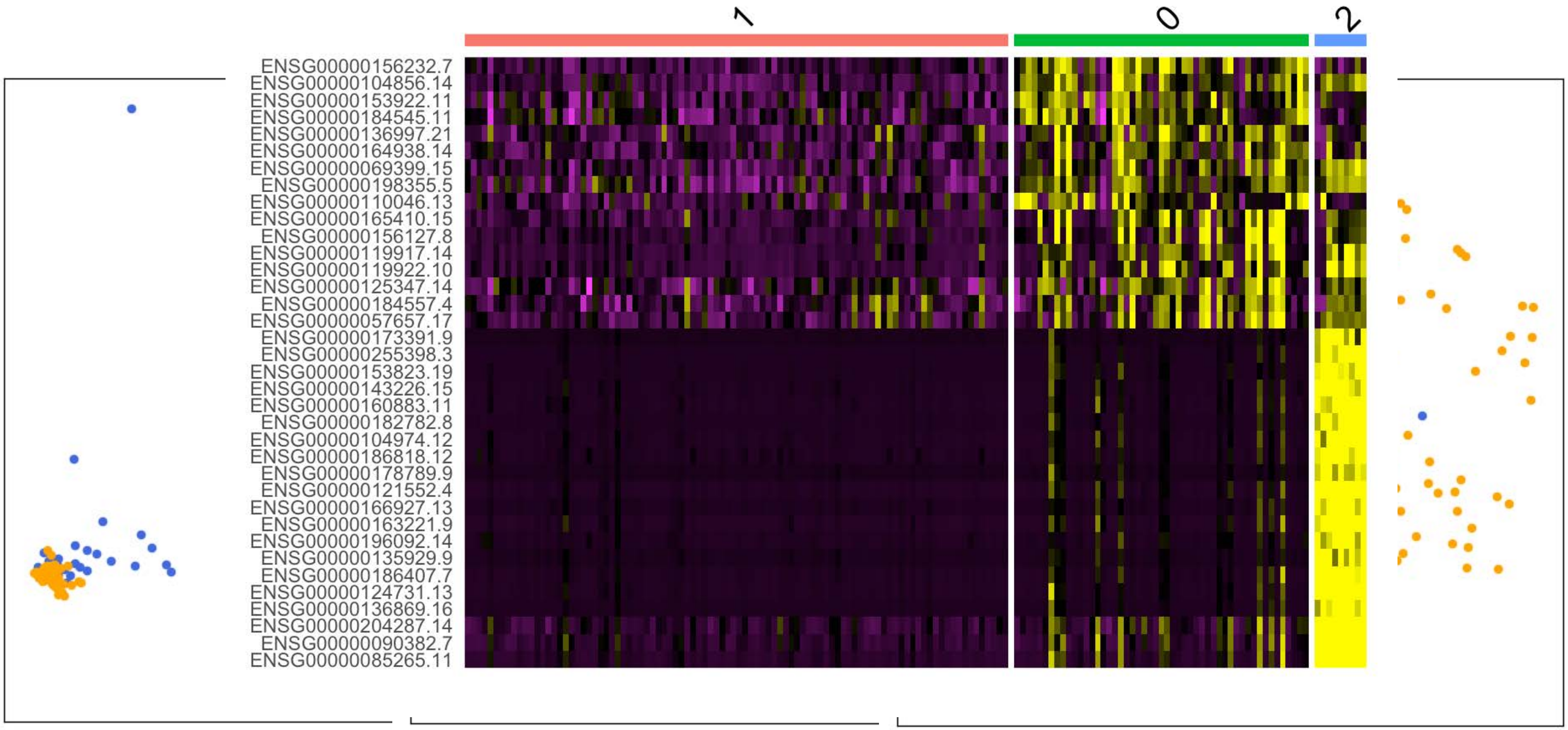
Differential Expression of Genes Between Clusters



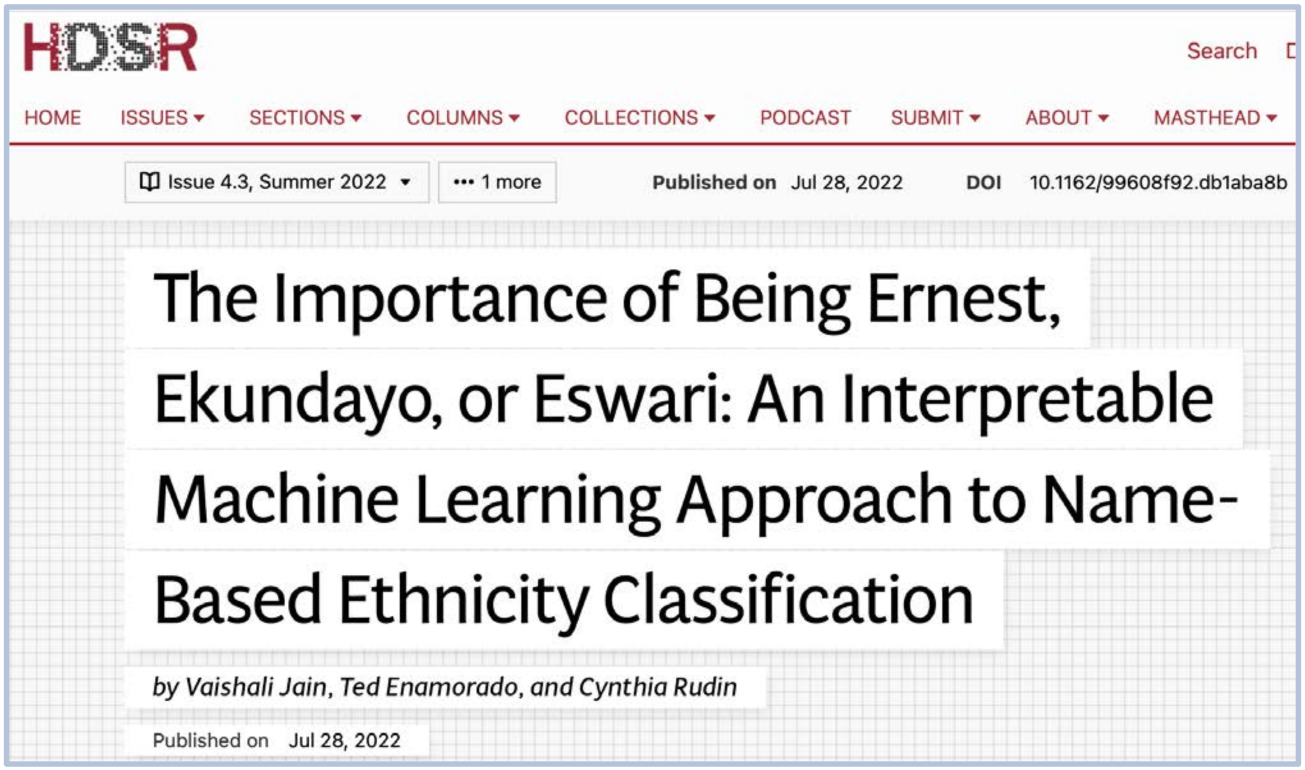
Other DR methods fail to identify the clusters



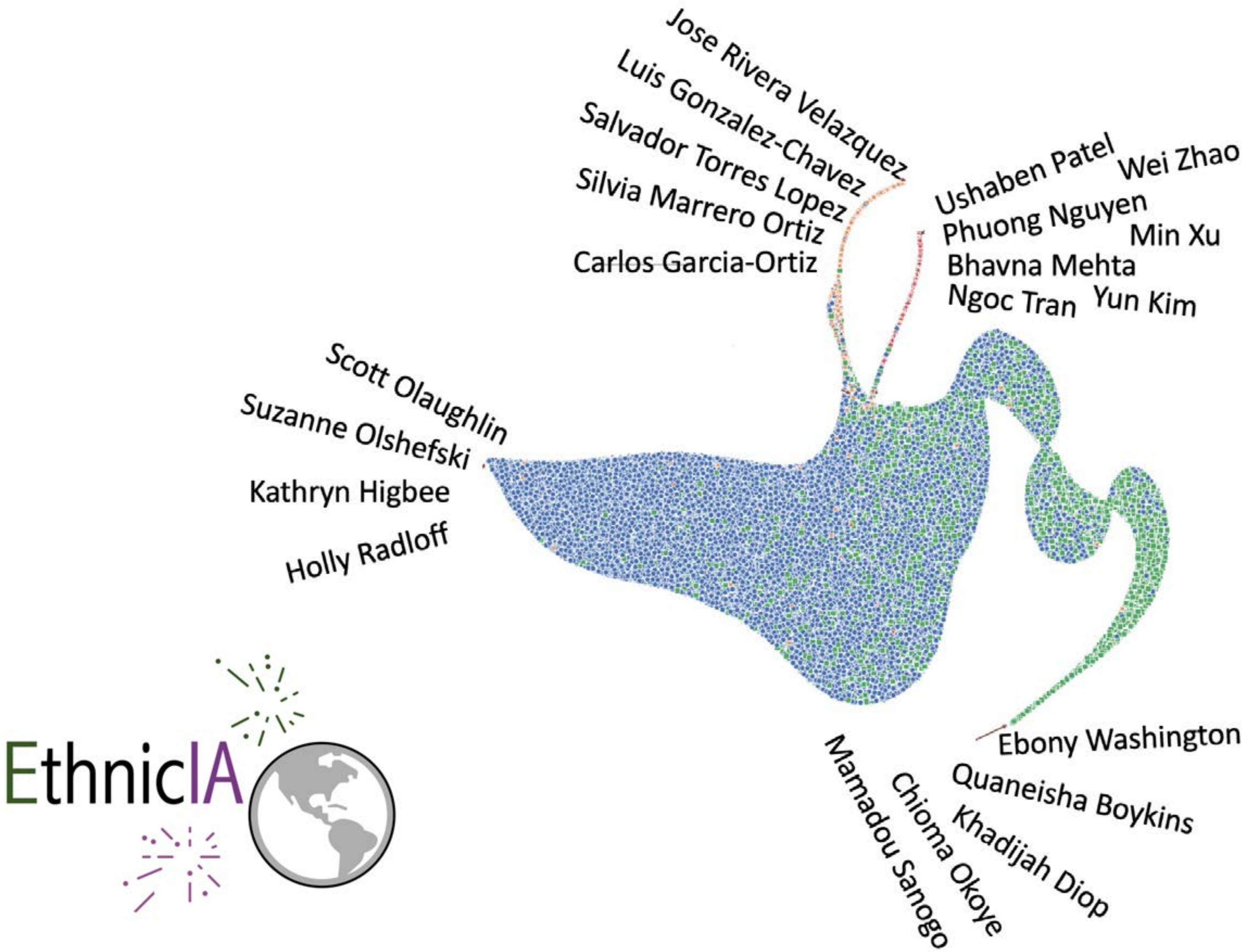
Other DR methods fail to identify the clusters



Name-Ethnicity Classification
(helpful for assessing fairness)



Harvard Data Science Review, 2022



Jose Rivera Velazquez
Luis Gonzalez-Chavez
Salvador Torres Lopez
Silvia Marrero Ortiz
Carlos Garcia-Ortiz
Ushaben Patel
Wei Zhao
Phuong Nguyen
Min Xu
Bhavna Mehta
Ngoc Tran
Yun Kim
Scott Olaughlin
Suzanne Olshefski
Kathryn Higbee
Holly Radloff
Ebony Washington
Quaneisha Boykins
Khadijah Diop
Chioma Okoye
Mamadou Sanogo
EthnicIA

Our latest DR papers

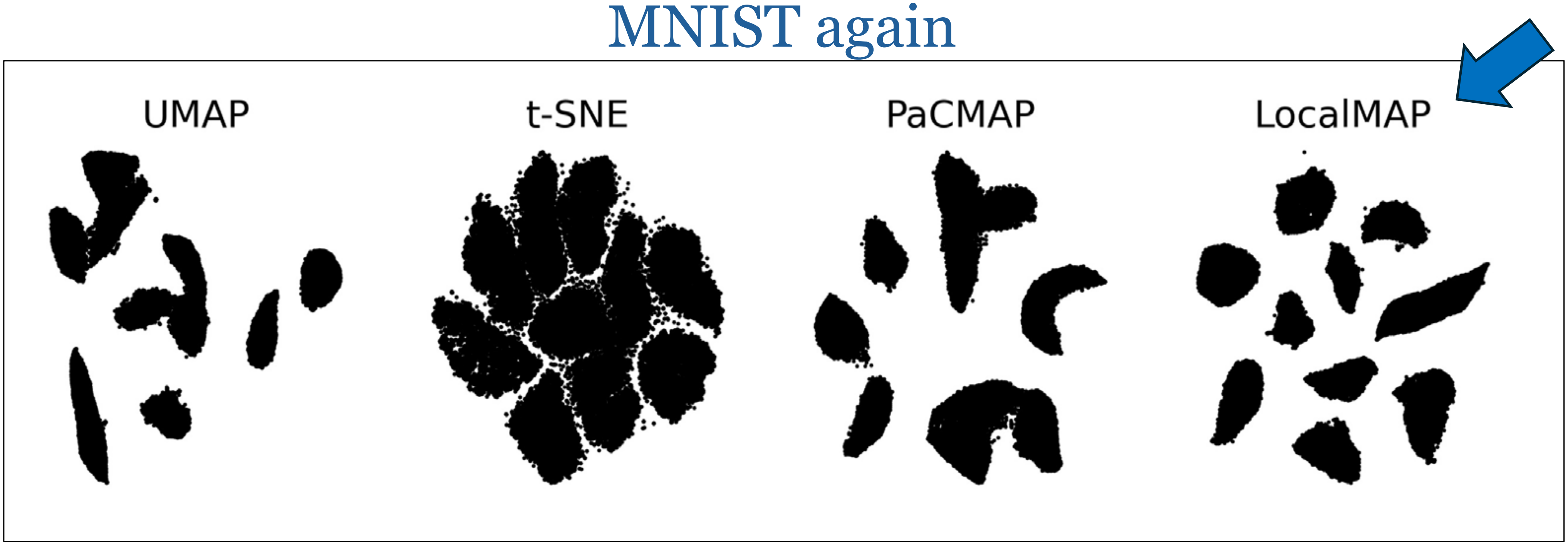
Dimension Reduction with Locally Adjusted Graphs

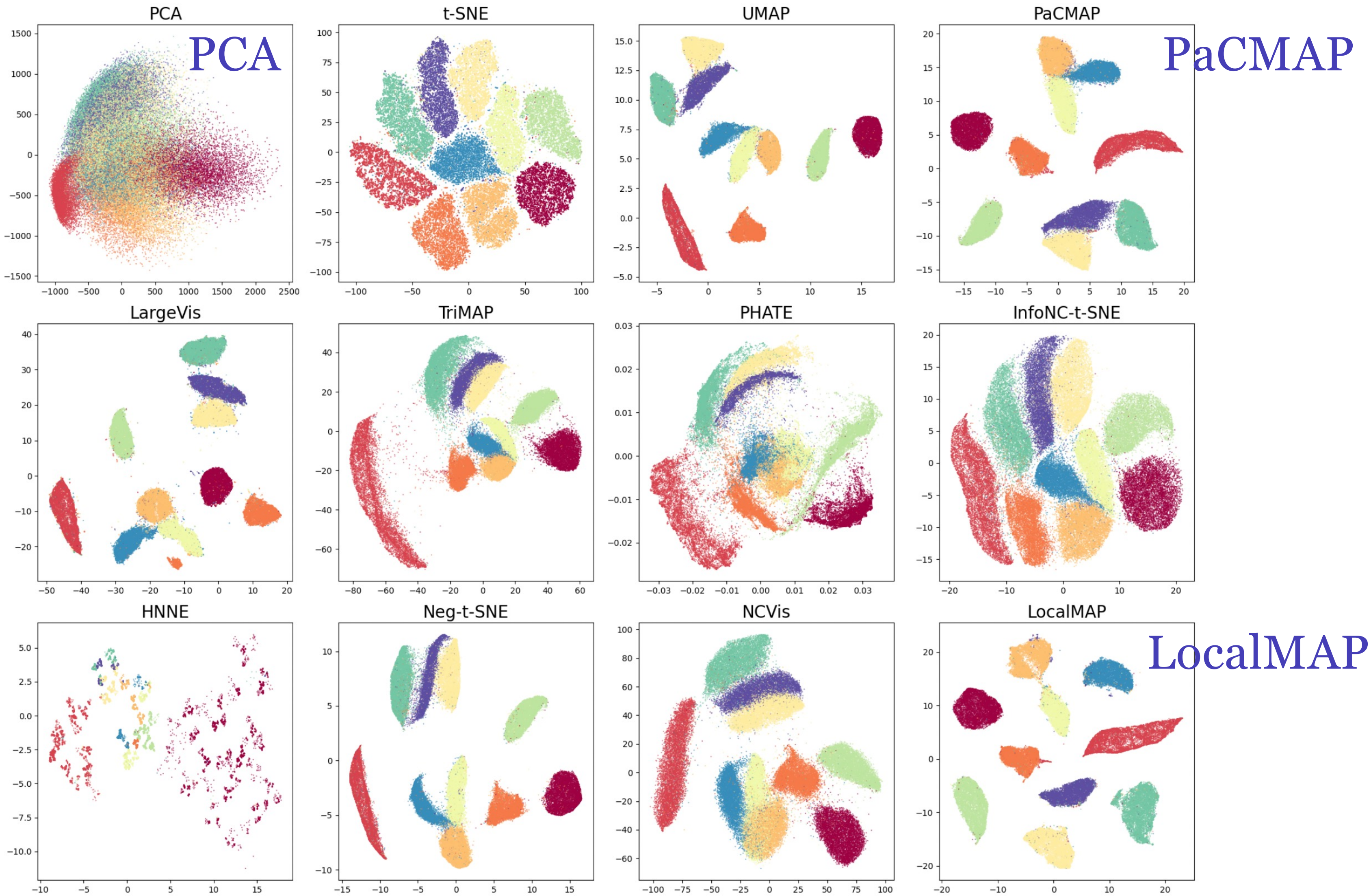
Yingfan Wang*, Yiyang Sun*, Haiyang Huang*, Cynthia Rudin

Duke University

AAAI, 2025

MNIST again

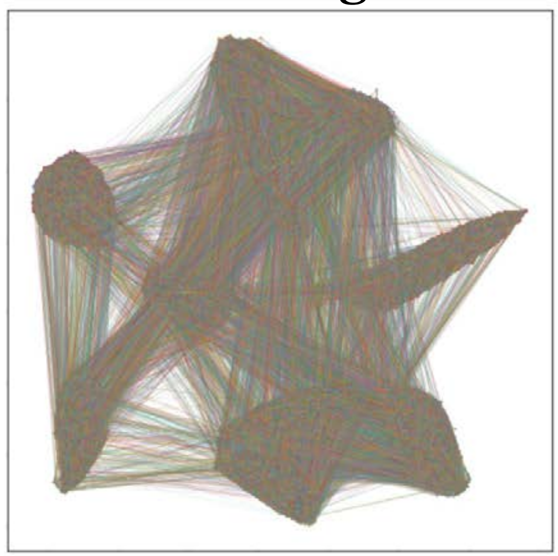



PCA
t-SNE
UMAP
PaCMAP
LargeVis
TriMAP
PHATE
InfoNC-t-SNE
HNNE
Neg-t-SNE
NCVis
LocalMAP
PCA
PaCMAP
LocalMAP

Standard DR methods (including PaCMAP) include data incorrectly!

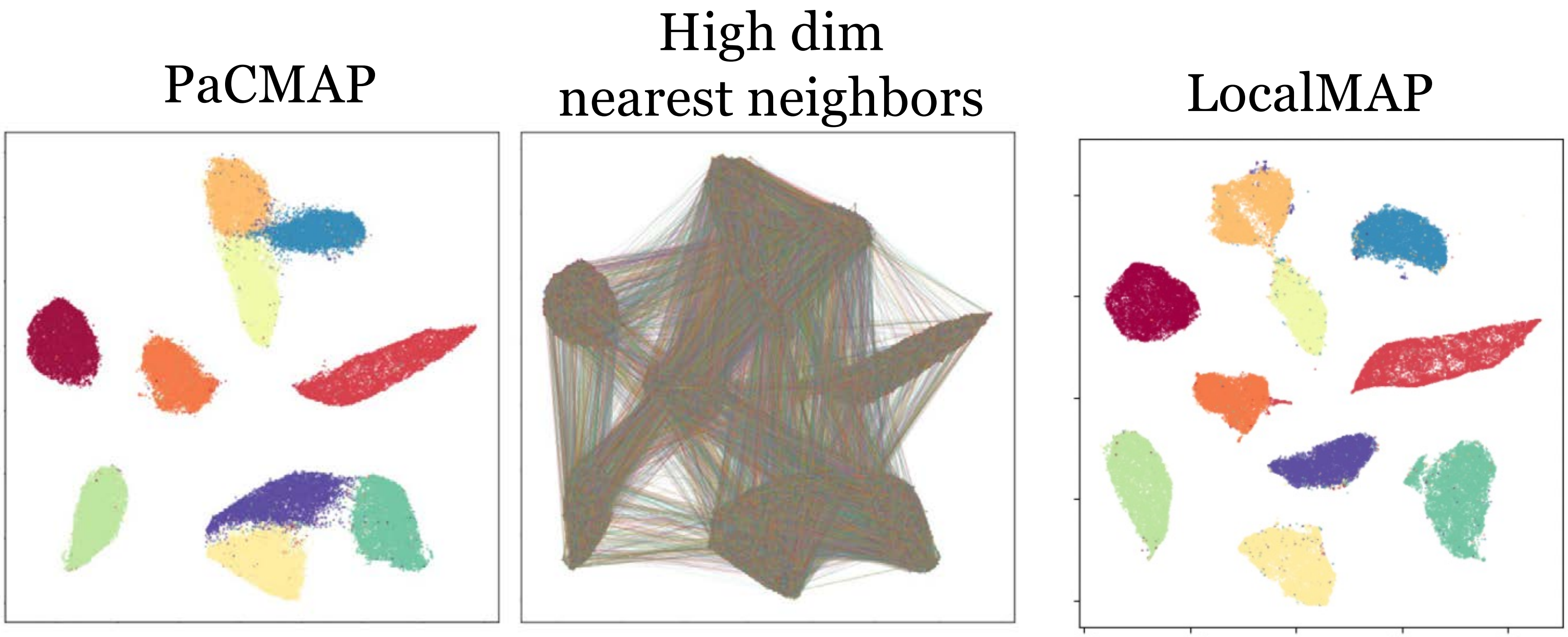
PaCMAP

High dim
nearest neighbors



LocalMAP dynamically resamples/reweights data to devalue incorrect edges.

Standard DR methods (including PaCMAP) include data incorrectly!

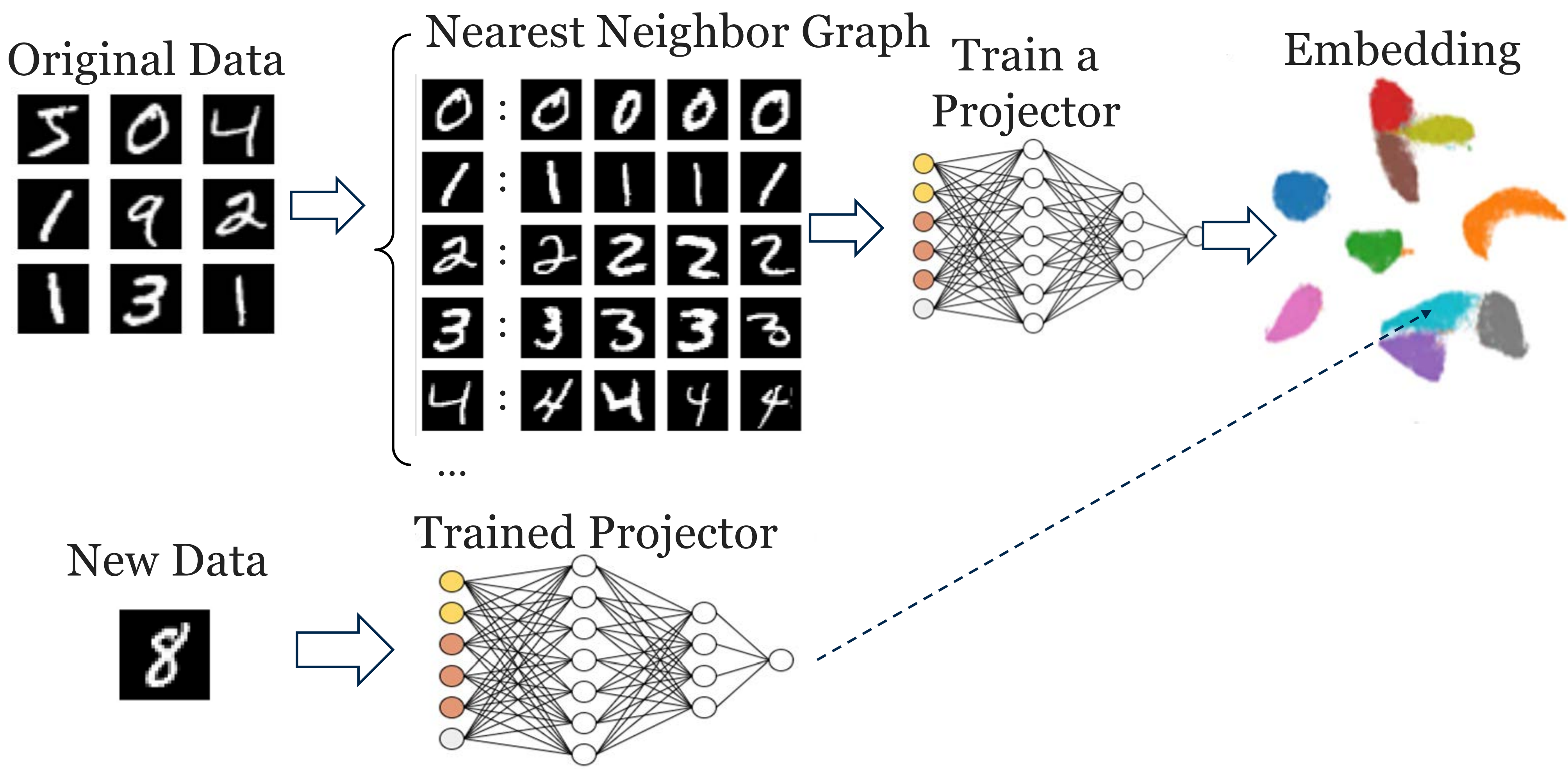


LocalMAP dynamically resamples/reweights data to devalue incorrect edges.

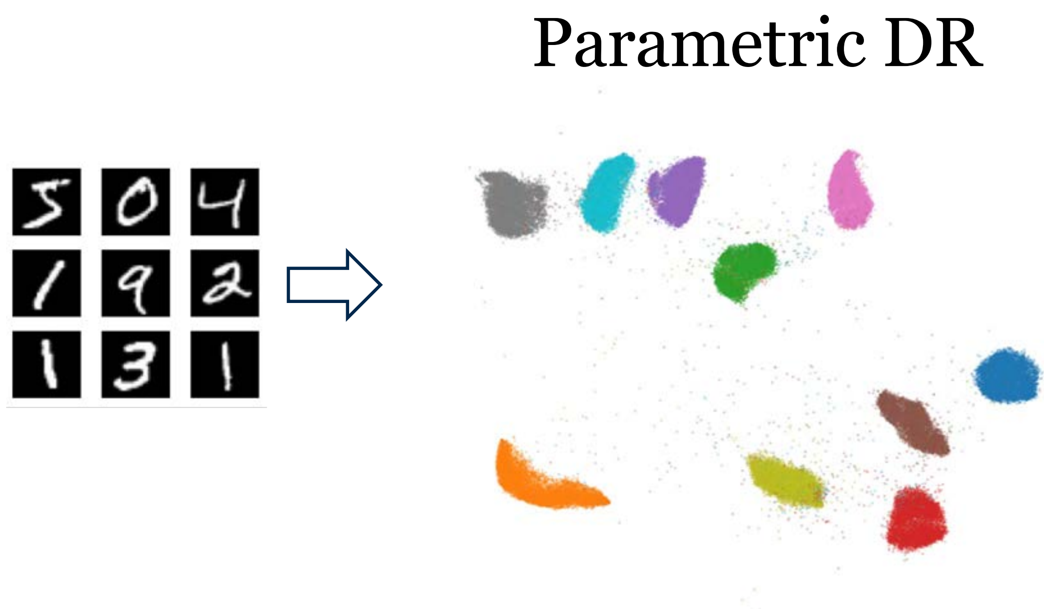
Navigating the Effect of Parametrization for Dimensionality Reduction

Haiyang Huang* **Yingfan Wang*** **Cynthia Rudin**
Duke University
{hyhuang, yw416, cynthia}@cs.duke.edu

Use a **parametrized** projector to map original data into embedding



Use a **parametrized** projector to map original data into embedding



Alina Barnett

Stark Guo

EEG Monitoring

NEJM AI

CURRENT ISSUE RECENTLY PUBLISHED PODCAST EVENTS AUTHOR CENTER ABOUT PUBLICATIONS

ORIGINAL ARTICLE

Improving Clinician Performance in Classifying EEG Patterns on the Ictal–Interictal Injury Continuum Using Interpretable Machine Learning

Authors: Alina Jade Barnett, Ph.D., Zhicheng Guo, Jin Jing, Ph.D., Wendong Ge, Ph.D., Peter W. Kaplan, M.B.B.S., Wan Yee Kong, M.D., Ioannis Karakis, M.D., Ph.D., +7, and M. Brandon Westover, M.D., Ph.D. Author Info & Affiliations

Published May 23, 2024 | NEJM AI 2024;1(6) | DOI: 10.1056/AIoa2300331 | VOL. 1 NO. 6 | Copyright © 2024

Jin Jing

Brandon Westover

Alina Barnett, Zhicheng (Stark) Guo, Jin Jing and Brandon Westover

EEG Monitoring

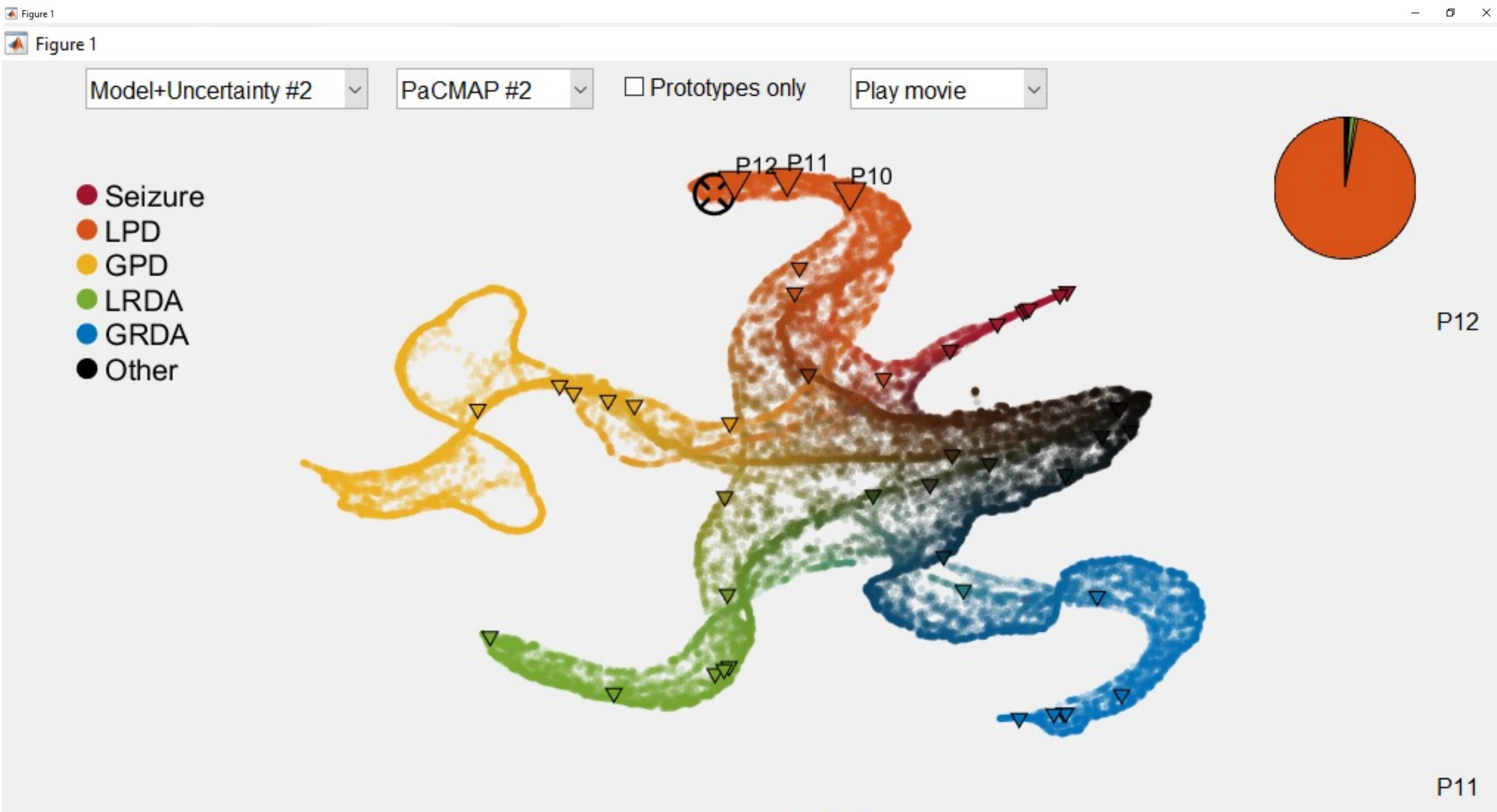
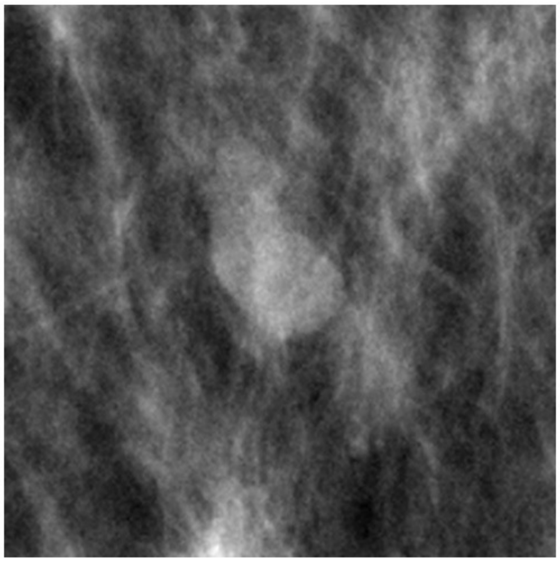


Figure 1
Figure 1
Model+Uncertainty #2
PaCMAP #2
Prototypes only
Play movie
Seizure
LPD
GPD
LRDA
GRDA
Other
P12
P11
P10
P12
P11

Interpretable neural networks

- ProtoPNet uses case-based reasoning

Should I biopsy this breast lesion?



Black box approach:

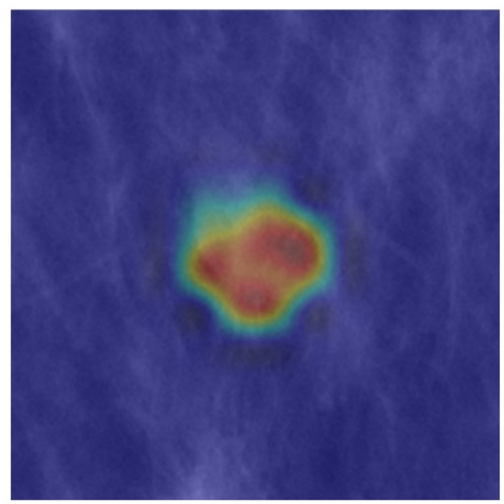
Probability of malignancy is low. Predict benign.

Reason: N/A

Saliency map approach:

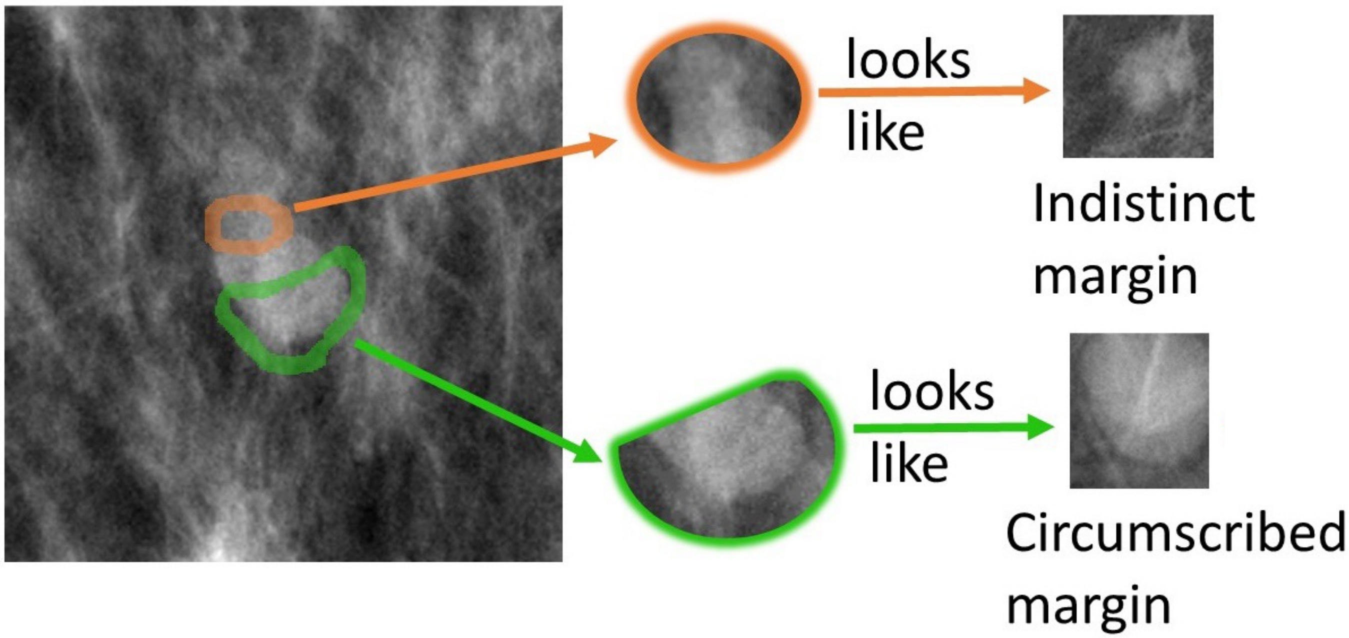
Probability of malignancy is low. Predict benign.

Reason: Here is where I am looking.



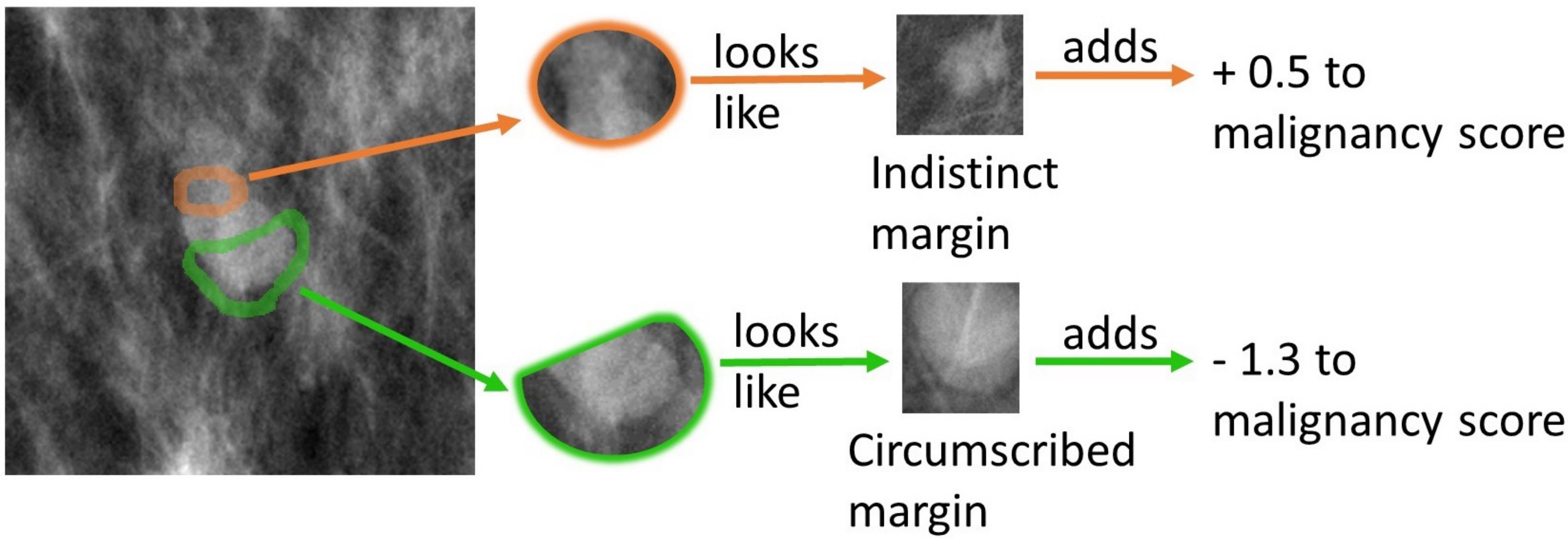
Should I biopsy this breast lesion?

Our approach (IAIA-BL)



Should I biopsy this breast lesion?

Our approach (IAIA-BL)

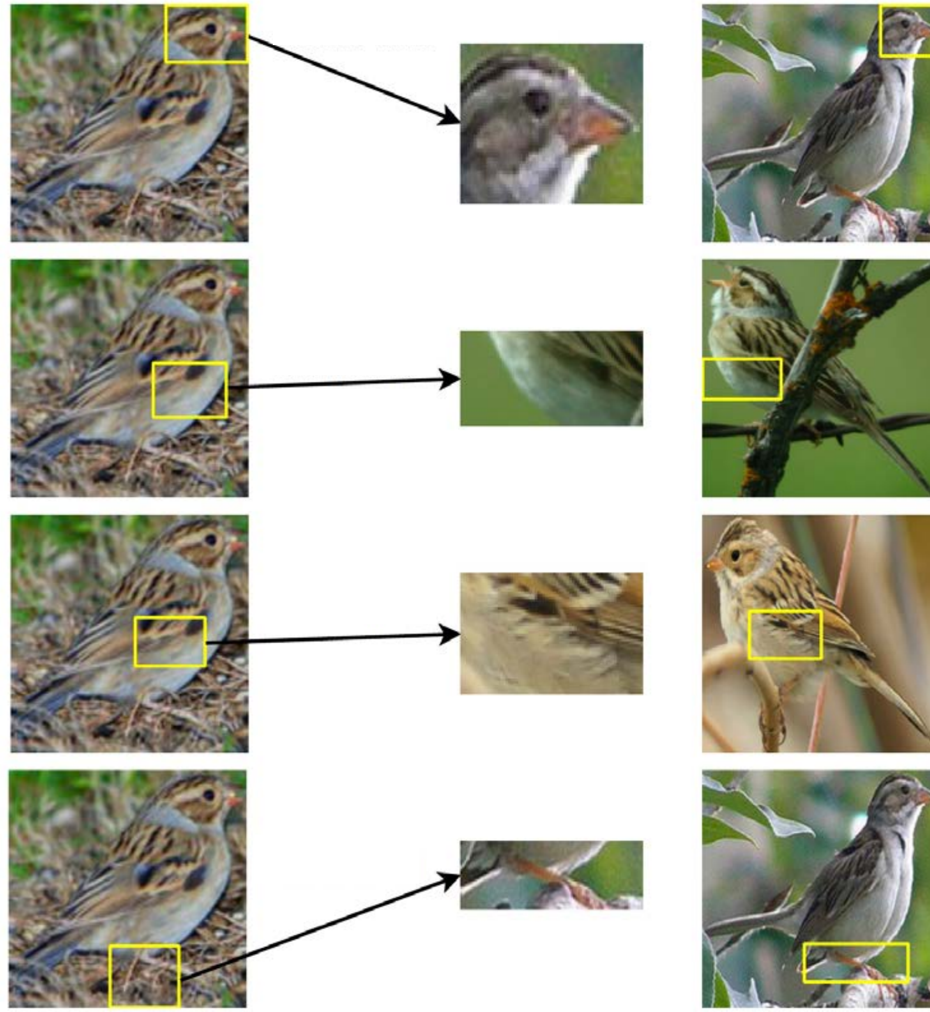


Probability of malignancy is low. Predict benign.

Reason: Mass primarily has circumscribed margin.

Why is this bird classified as a clay-colored sparrow?

Because this looks like | that part | of a prototypical clay-colored sparrow



***This* Looks Like *That*: Deep Learning for Interpretable Image Recognition**

NeurIPS 2019 (spotlight)
1.6K+ citations

ProtoPNet

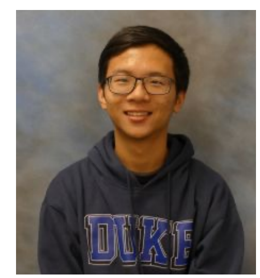
- Adds a "prototype" layer to a black box, forces the network to do case-based reasoning.

Chaofan Chen

Oscar Li

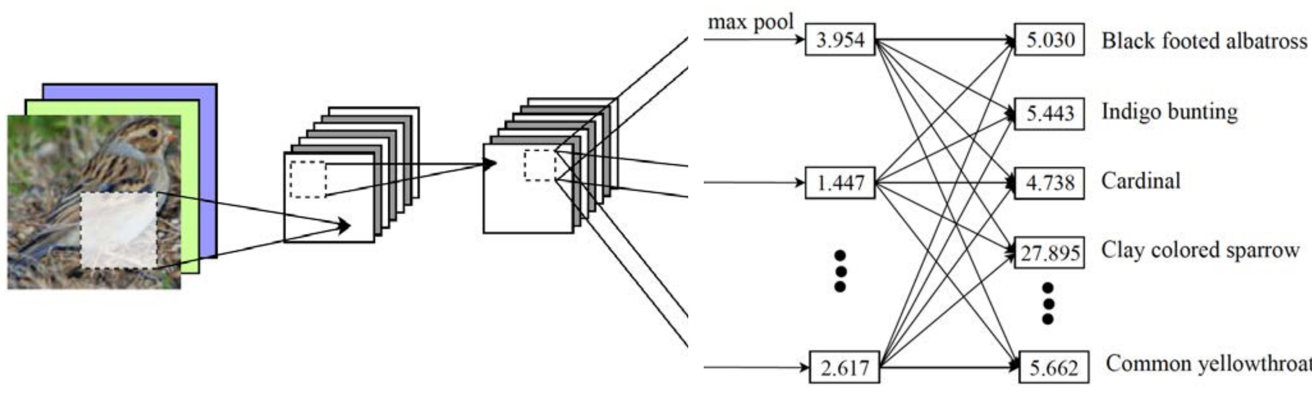
Alina Barnett

Jonathan Su



Daniel Tao

Take any “standard” black box CNN…

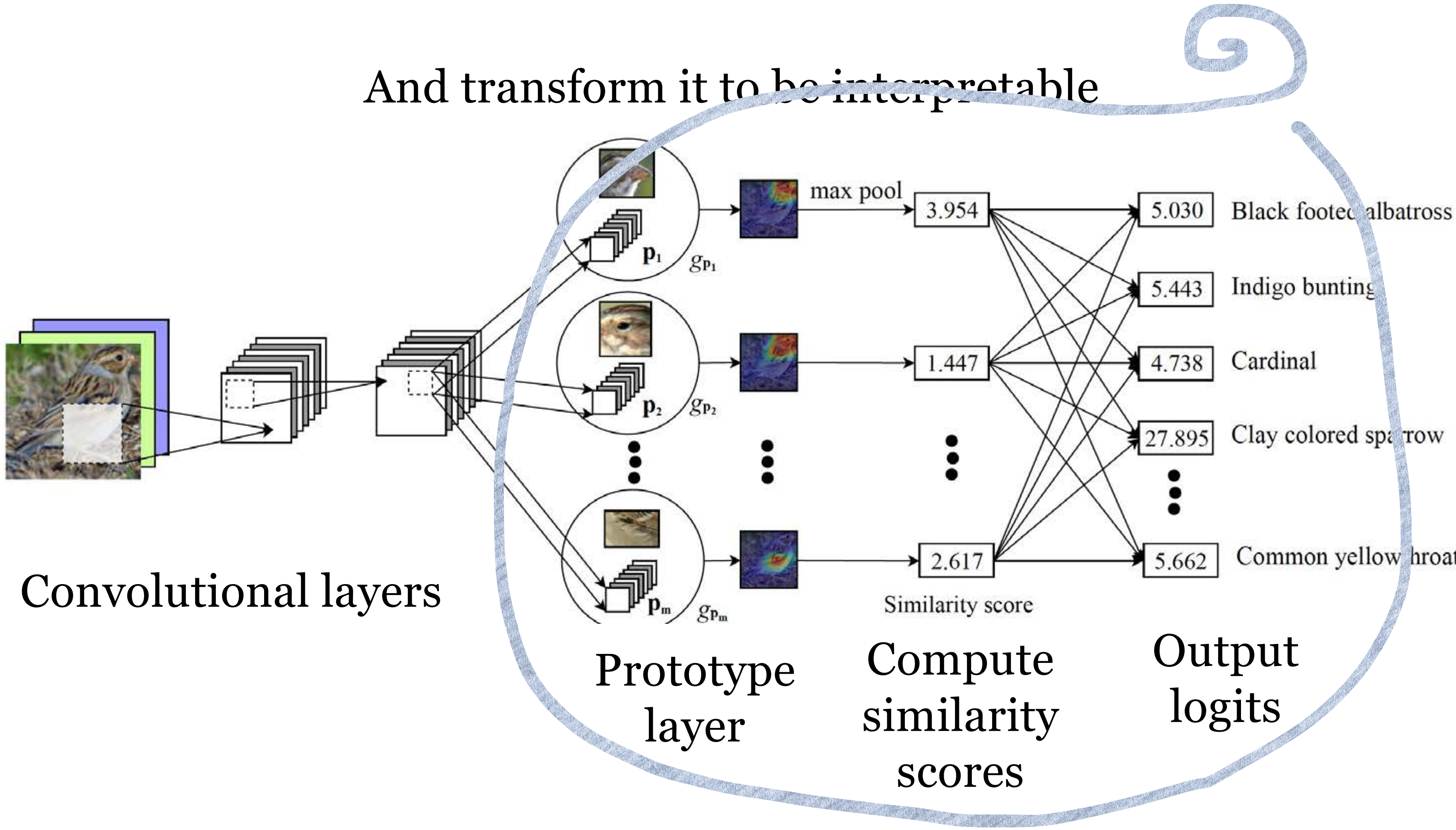


Convolutional layers

Fully connected layer

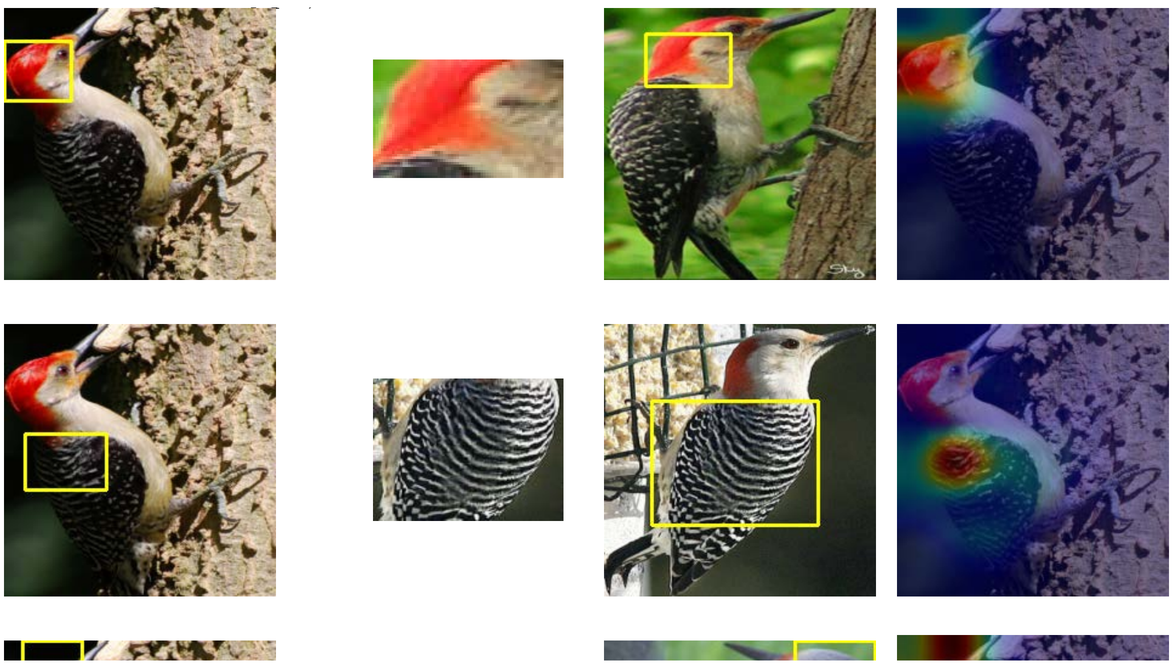
Output logits

And transform it to be interpretable



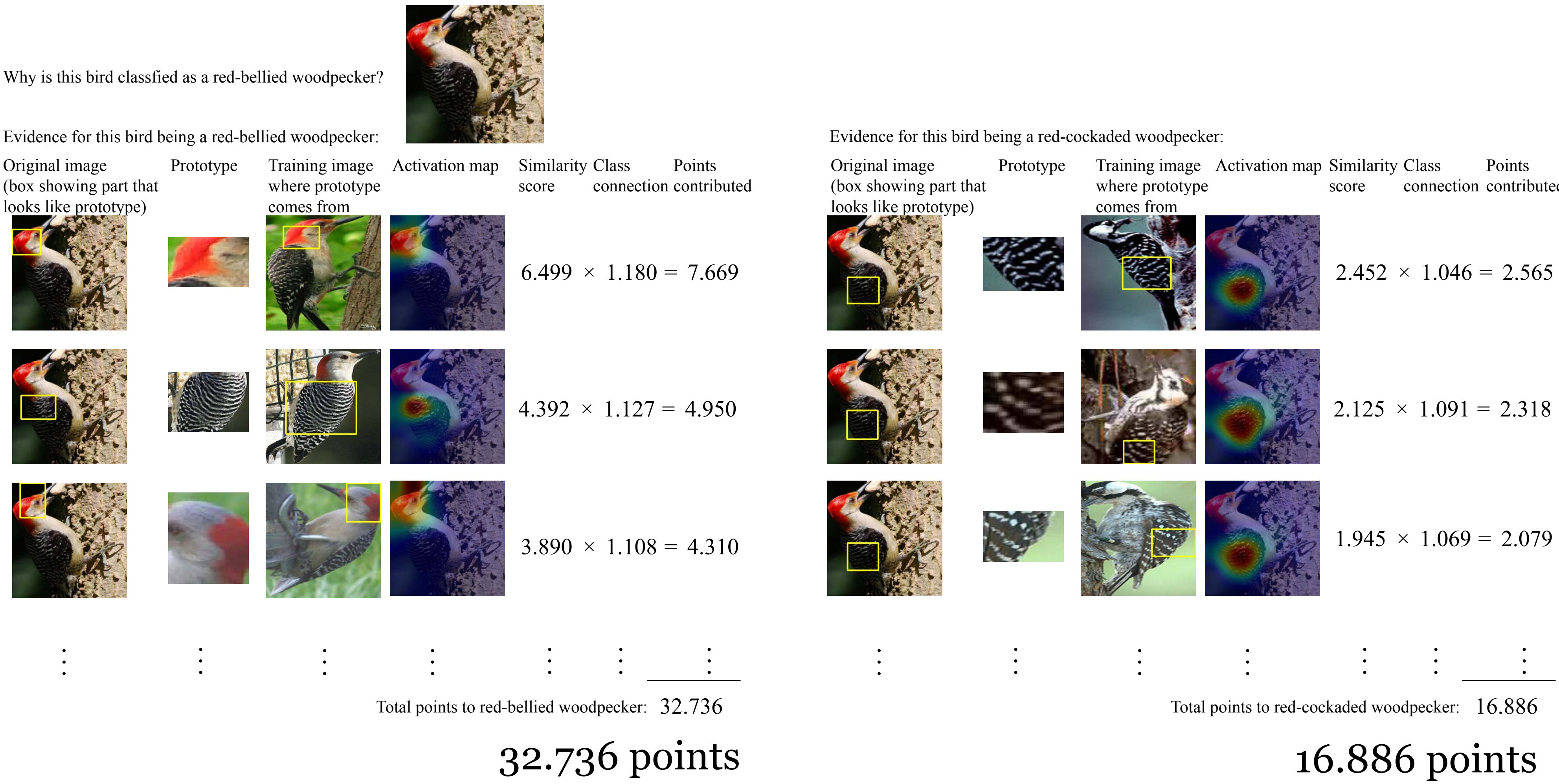
Why is this bird classified as a red-bellied woodpecker?

Original image | Prototype | Training image | Activation map | Similarity score | Class connection | Points contributed



$6.499 \times 1.180 = 7.669$

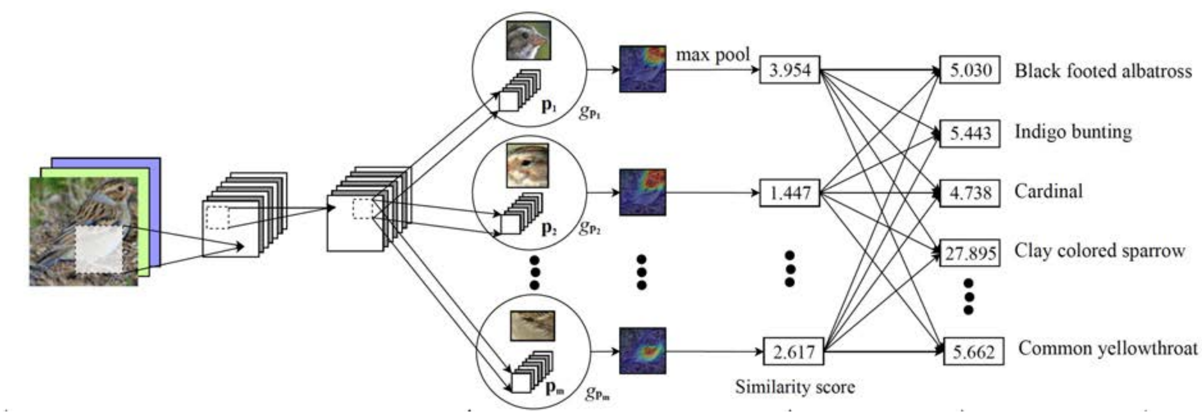
$4.392 \times 1.127 = 4.950$


Why is this bird classfied as a red-bellied woodpecker?
Evidence for this bird being a red-bellied woodpecker:
Original image (box showing part that looks like prototype)
Prototype
Training image where prototype comes from
Activation map
Similarity score
Class connection
Points contributed
6.499 × 1.180 = 7.669
4.392 × 1.127 = 4.950
3.890 × 1.108 = 4.310
Total points to red-bellied woodpecker: 32.736
32.736 points
Evidence for this bird being a red-cockaded woodpecker:
Original image (box showing part that looks like prototype)
Prototype
Training image where prototype comes from
Activation map
Similarity score
Class connection
Points contributed
2.452 × 1.046 = 2.565
2.125 × 1.091 = 2.318
1.945 × 1.069 = 2.079
Total points to red-cockaded woodpecker: 16.886
16.886 points

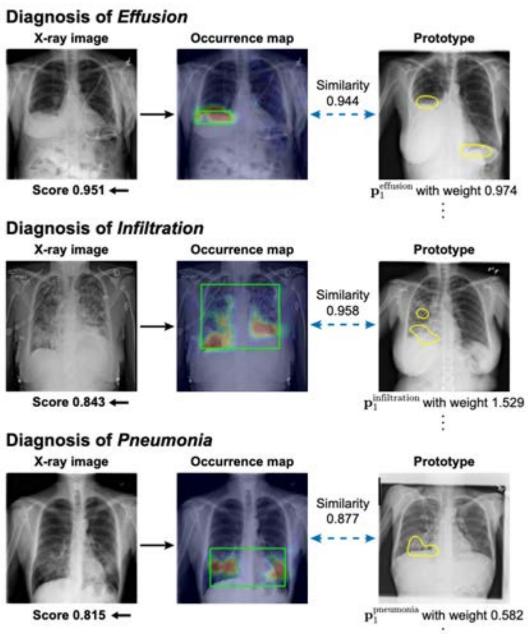
Training ProtoPNet:

Minimize
Weights, Prototypes

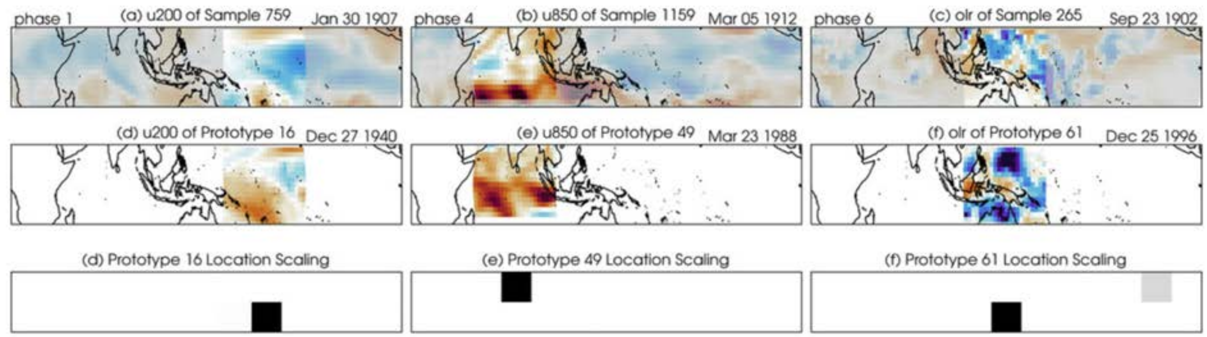
Cross-Entropy Loss between labels and predictions
+ Distance from prototype to nearest patch of correct class
- Distance from prototype to nearest patch of incorrect class



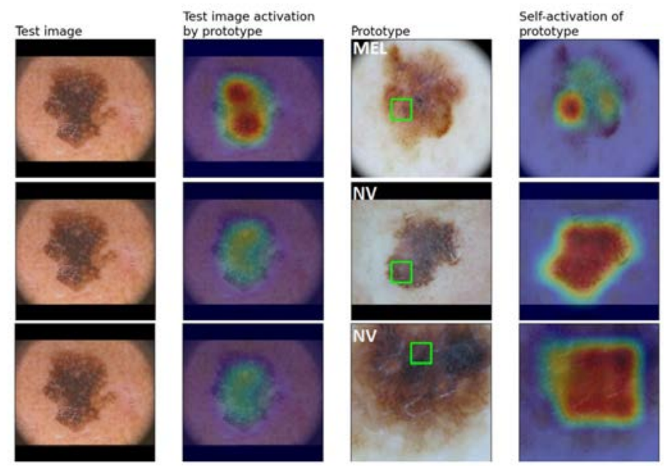
ProtoPNet Use Cases



Kim, Eunji, et al. "XProtoNet: diagnosis in chest radiography with global and local explanations." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.



Barnes, Elizabeth A., et al. "This Looks Like That There: Interpretable neural networks for image tasks when location matters." *Artificial Intelligence for the Earth Systems* 1.3 (2022): e220001.



Correia, Miguel, et al. "XAI for Skin Cancer Detection with Prototypes and Non-Expert Supervision." *arXiv preprint arXiv:2402.01410* (2024).

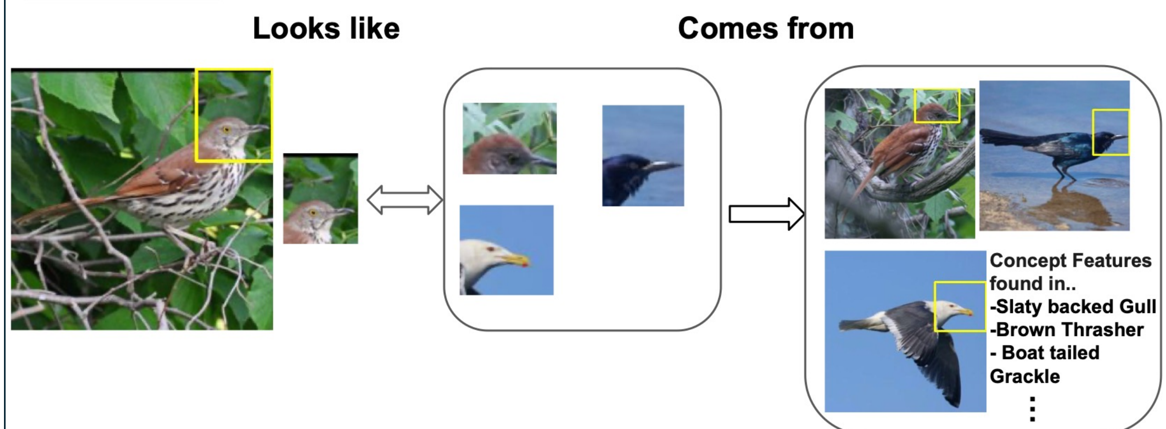
Our latest ProtoPNet-ish papers

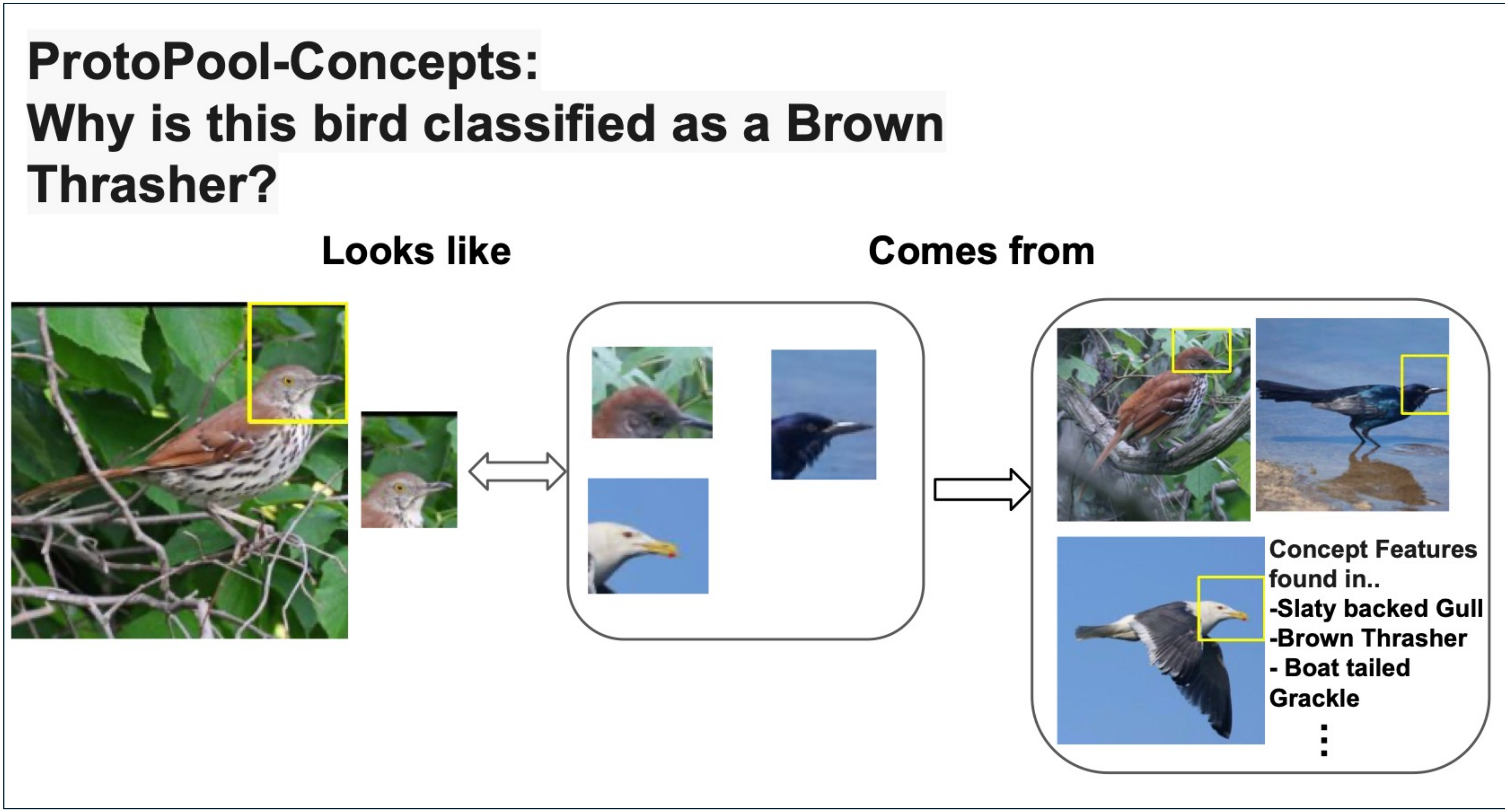
This Looks Like Those: Illuminating Prototypical Concepts Using Multiple Visualizations

Chiyu Ma* Dartmouth College chiyu.ma.gr@dartmouth.edu
Brandon Zhao* Caltech byzhao@caltech.edu
Chaofan Chen University of Maine chaofan.chen@maine.edu
Cynthia Rudin Duke University cynthia@cs.duke.edu

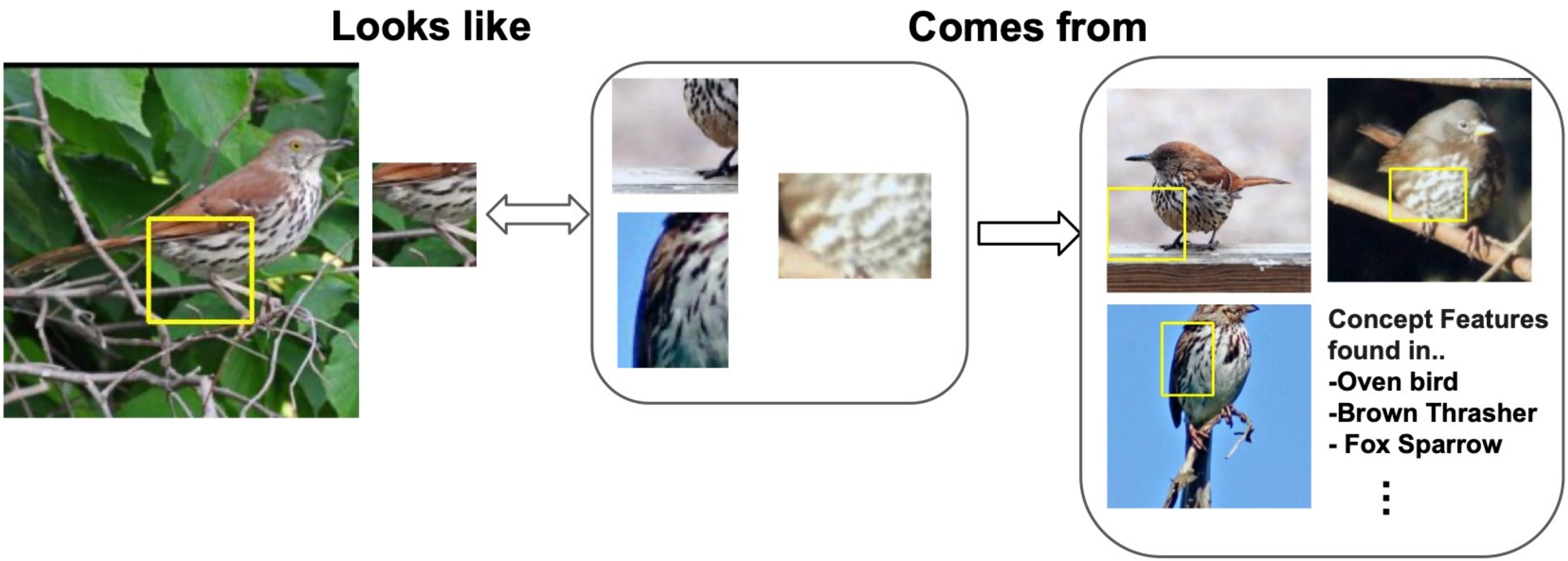
NeurIPS 2023

ProtoPool-Concepts:
Why is this bird classified as a Brown Thrasher?




ProtoPool-Concepts:
Why is this bird classified as a Brown Thrasher?
Looks like
Comes from
Concept Features found in..
-Slaty backed Gull
-Brown Thrasher
- Boat tailed Grackle
⋮

ProtoPool-Concepts:
Why is this bird classified as a Brown Thrasher?



Interpretable Image Classification with Adaptive Prototype-based Vision Transformers

Chiyu Ma
Dartmouth College
chiyu.ma.gr@dartmouth.edu

Jon Donnelly
Duke University
jon.donnelly@duke.edu

Wenjun Liu
Dartmouth College
wenjun.liu.gr@dartmouth.edu

Soroush Vosoughi
Dartmouth College
soroush.vosoughi@dartmouth.edu

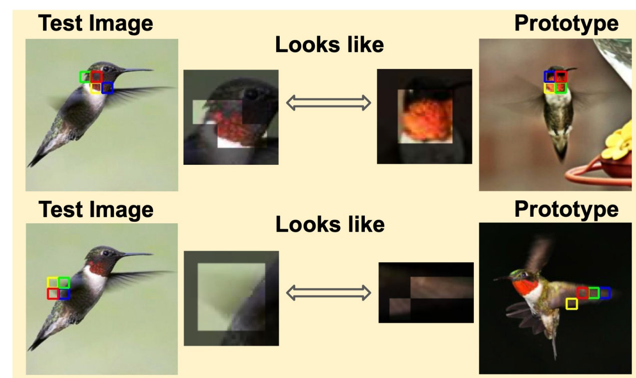
Cynthia Rudin
Duke University
cynthia@cs.duke.edu

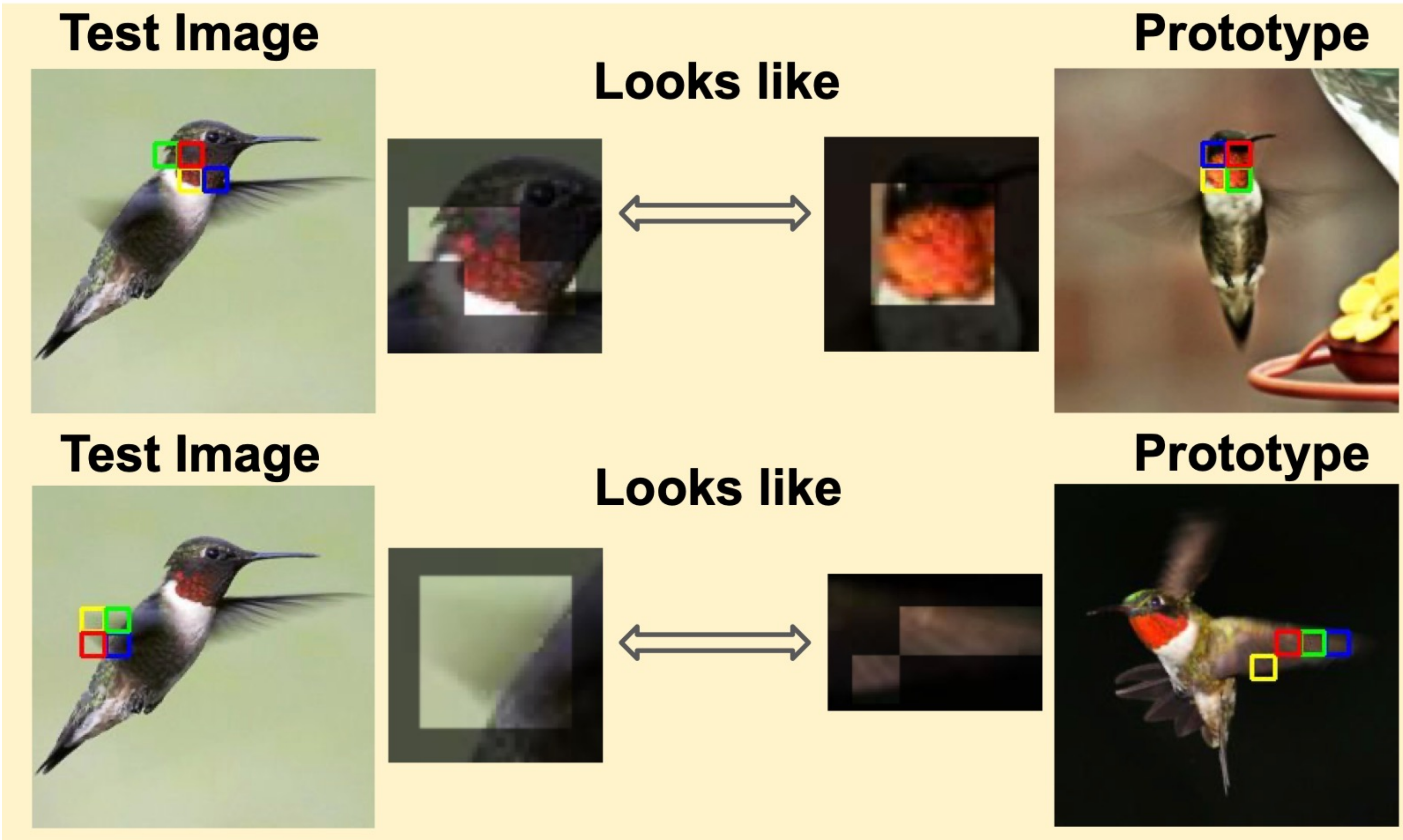
Chaofan Chen
University of Maine
chaofan.chen@maine.edu

NeurIPS 2024

ProtoViT

- uses prototype logic (thus, interpretable)
- as accurate as the black box vision transformers




Test Image
Looks like
Prototype
Test Image
Looks like
Prototype

EEG Monitoring

Alina Barnett

Stark Guo

NEJM AI

CURRENT ISSUE | RECENTLY PUBLISHED | PODCAST | EVENTS | AUTHOR CENTER | ABOUT | PUBLICATIONS

ORIGINAL ARTICLE

Improving Clinician Performance in Classifying EEG Patterns on the Ictal–Interictal Injury Continuum Using Interpretable Machine Learning

Authors: Alina Jade Barnett, Ph.D., Zhicheng Guo, Jin Jing, Ph.D., Wendong Ge, Ph.D., Peter W. Kaplan, M.B.B.S., Wan Yee Kong, M.D., Ioannis Karakis, M.D., Ph.D., +7, and M. Brandon Westover, M.D., Ph.D. Author Info & Affiliations

Published May 23, 2024 | NEJM AI 2024;1(6) | DOI: 10.1056/AIoa2300331 | VOL. 1 NO. 6 | Copyright © 2024

Jin Jing

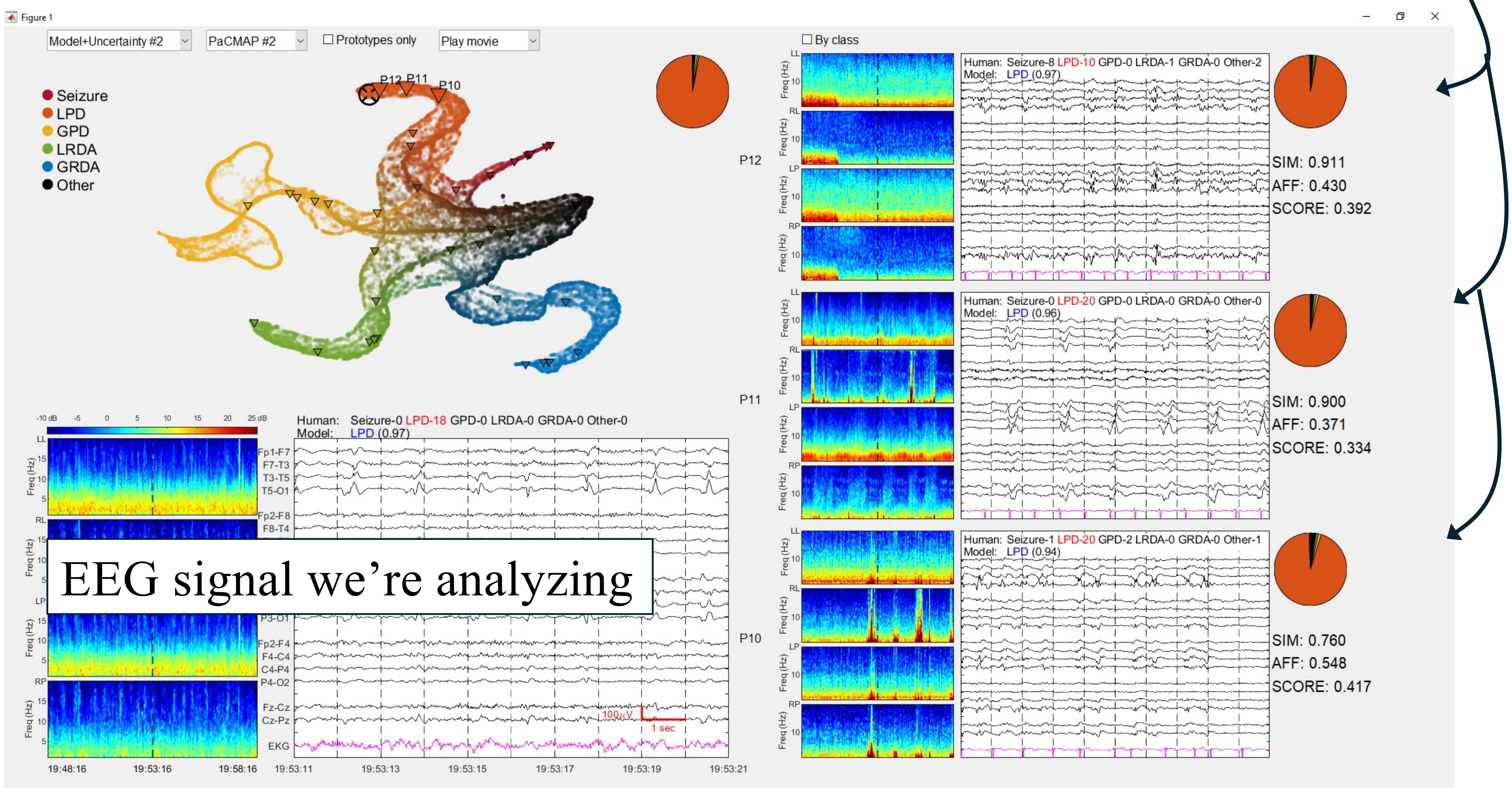
Brandon Westover

Alina Barnett, Zhicheng (Stark) Guo, Jin Jing and Brandon Westover

EEG Monitoring

Visualization of Last Layer (using PaCMAP)

Prototypes



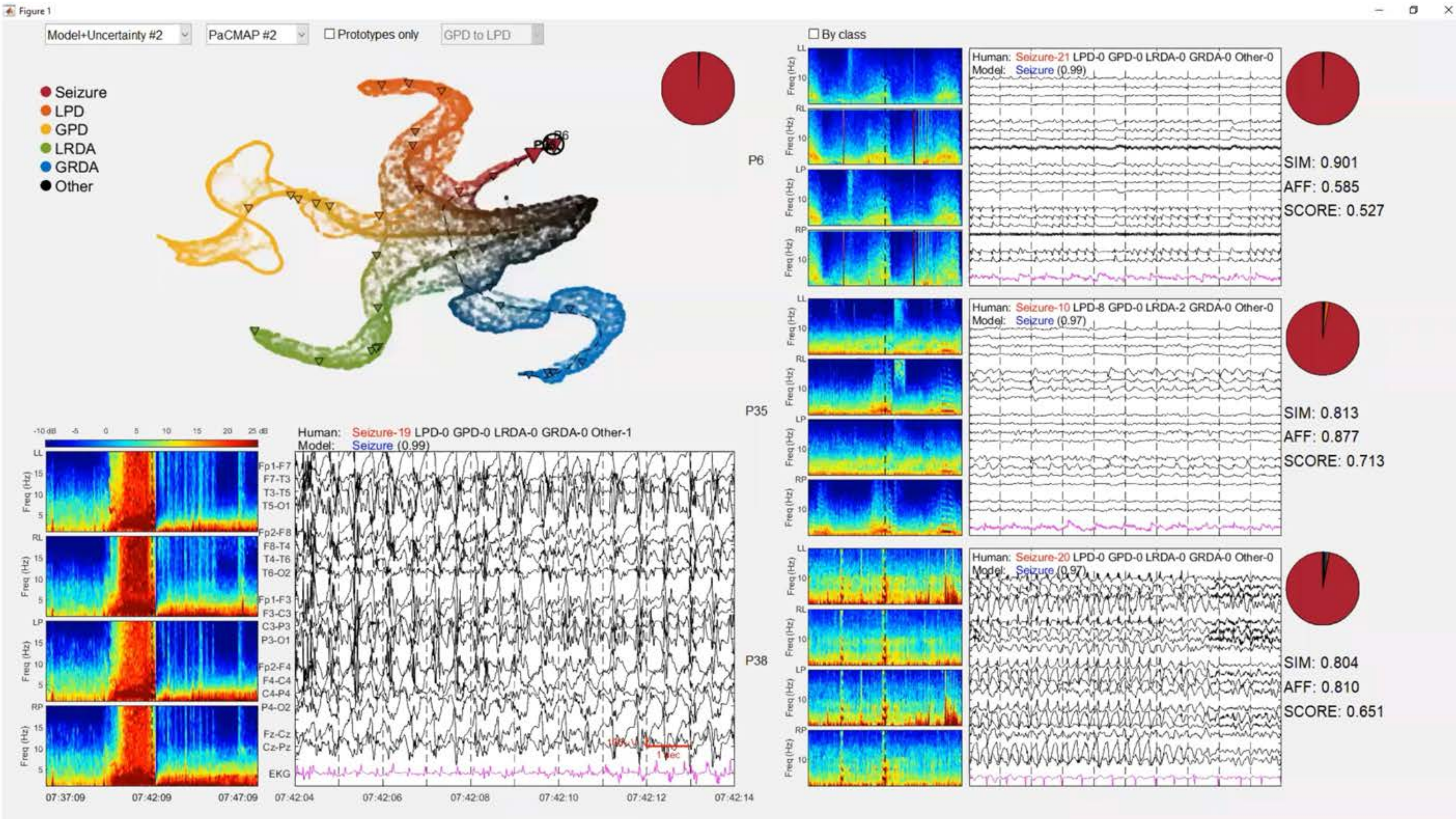


Figure 1
Model+Uncertainty #2
PaCMAP #2
Prototypes only
GPD to LPD
By class
Seizure
LPD
GPD
LRDA
GRDA
Other
P6
P35
P38
Human: Seizure-21 LPD-0 GPD-0 LRDA-0 GRDA-0 Other-0
Model: Seizure (0.99)
SIM: 0.901
AFF: 0.585
SCORE: 0.527
Human: Seizure-10 LPD-8 GPD-0 LRDA-2 GRDA-0 Other-0
Model: Seizure (0.97)
SIM: 0.813
AFF: 0.877
SCORE: 0.713
Human: Seizure-20 LPD-0 GPD-0 LRDA-0 GRDA-0 Other-0
Model: Seizure (0.97)
SIM: 0.804
AFF: 0.810
SCORE: 0.651
Human: Seizure-19 LPD-0 GPD-0 LRDA-0 GRDA-0 Other-1
Model: Seizure (0.99)
Fp1-F7
F7-T3
T3-T5
T5-O1
Fp2-F8
F8-T4
T4-T6
T6-O2
Fp1-F3
F3-C3
C3-P3
P3-O1
Fp2-F4
F4-C4
C4-P4
P4-O2
Fz-Cz
Cz-Pz
EKG
LL
RL
LP
RP
Freq (Hz)
07:37:09
07:42:09
07:47:09
07:42:04
07:42:06
07:42:08
07:42:10
07:42:12
07:42:14

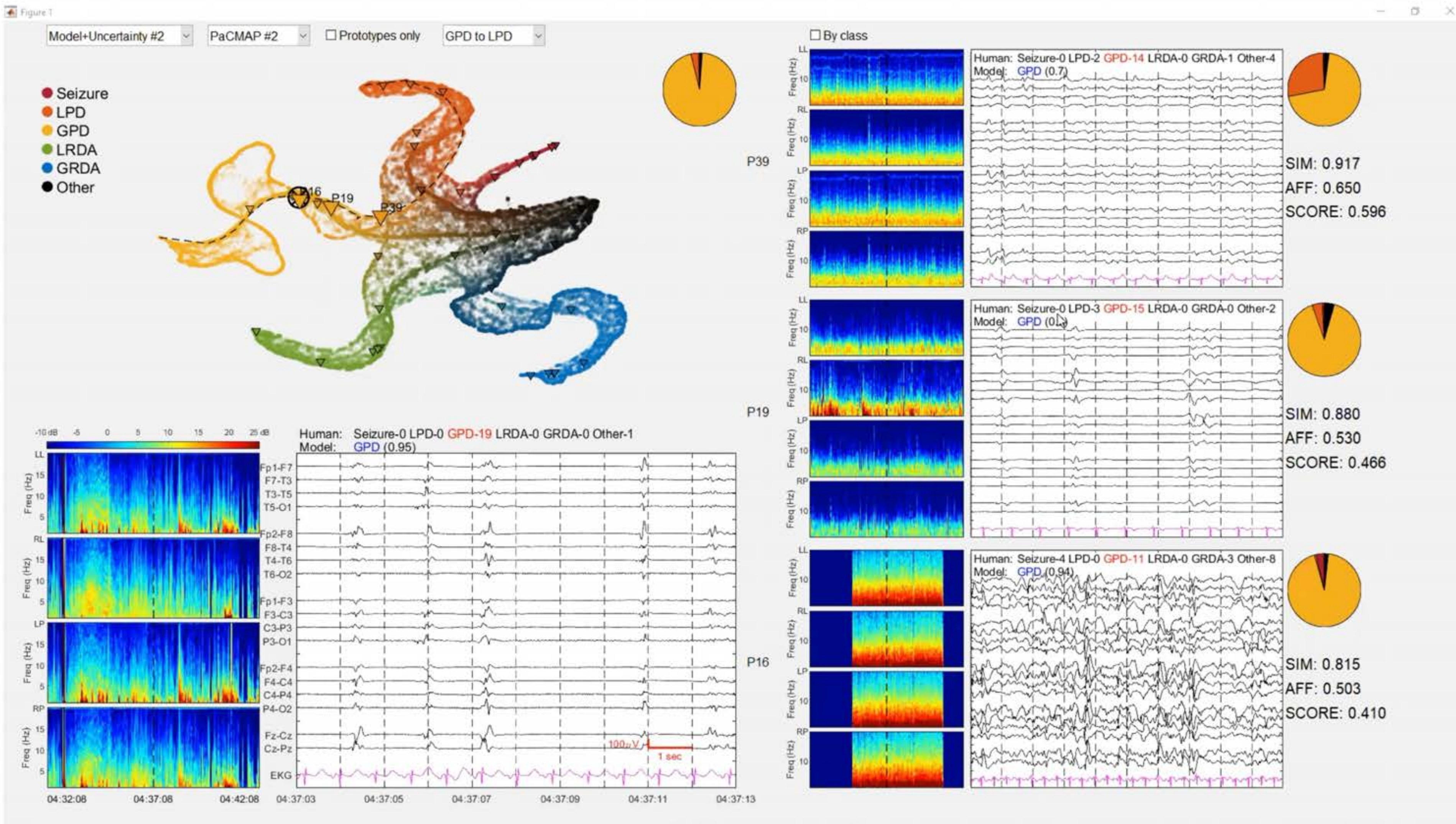


Figure 1
Model+Uncertainty #2
PaCMAP #2
Prototypes only
GPD to LPD
By class
Seizure
LPD
GPD
LRDA
GRDA
Other
P16
P19
P39
Human: Seizure-0 LPD-0 GPD-19 LRDA-0 GRDA-0 Other-1
Model: GPD (0.95)
Human: Seizure-0 LPD-2 GPD-14 LRDA-0 GRDA-1 Other-4
Model: GPD (0.7)
SIM: 0.917
AFF: 0.650
SCORE: 0.596
Human: Seizure-0 LPD-3 GPD-15 LRDA-0 GRDA-0 Other-2
SIM: 0.880
AFF: 0.530
SCORE: 0.466
Human: Seizure-4 LPD-0 GPD-11 LRDA-0 GRDA-3 Other-8
Model: GPD (0.94)
SIM: 0.815
AFF: 0.503
SCORE: 0.410
Fp1-F7
F7-T3
T3-T5
T5-O1
Fp2-F8
F8-T4
T4-T6
T6-O2
Fp1-F3
F3-C3
C3-P3
P3-O1
Fp2-F4
F4-C4
C4-P4
P4-O2
Fz-Cz
Cz-Pz
EKG
1 sec
04:32:08
04:37:08
04:42:08
04:37:03
04:37:05
04:37:07
04:37:09
04:37:11
04:37:13
Freq (Hz)
LL
RL
LP
RP

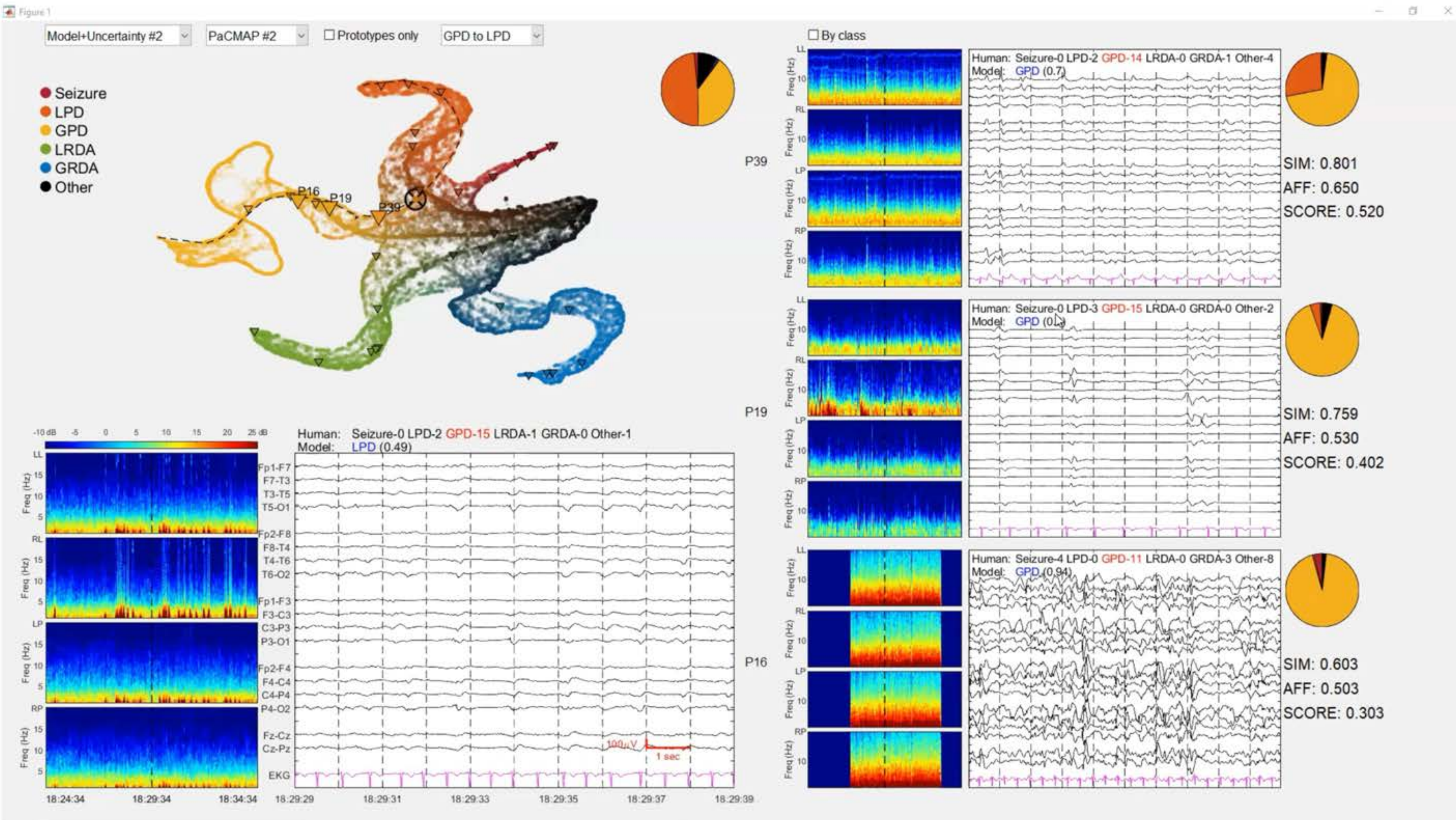


Figure 1
Model+Uncertainty #2
PaCMAP #2
Prototypes only
GPD to LPD
By class
Seizure
LPD
GPD
LRDA
GRDA
Other
P16
P19
P39
Human: Seizure-0 LPD-2 GPD-15 LRDA-1 GRDA-0 Other-1
Model: LPD (0.49)
Fp1-F7
F7-T3
T3-T5
T5-O1
Fp2-F8
F8-T4
T4-T6
T6-O2
Fp1-F3
F3-C3
C3-P3
P3-O1
Fp2-F4
F4-C4
C4-P4
P4-O2
Fz-Cz
Cz-Pz
EKG
100 µV
1 sec
18:24:34
18:29:34
18:34:34
18:29:29
18:29:31
18:29:33
18:29:35
18:29:37
18:29:39
Human: Seizure-0 LPD-2 GPD-14 LRDA-0 GRDA-1 Other-4
Model: GPD (0.7)
SIM: 0.801
AFF: 0.650
SCORE: 0.520
Human: Seizure-0 LPD-3 GPD-15 LRDA-0 GRDA-0 Other-2
SIM: 0.759
AFF: 0.530
SCORE: 0.402
Human: Seizure-4 LPD-0 GPD-11 LRDA-0 GRDA-3 Other-8
Model: GPD (0.94)
SIM: 0.603
AFF: 0.503
SCORE: 0.303

Two Indispensable Tools for Scientific Discovery

Dimension Reduction for Data Visualization

PaCMAP & Friends

Interpretable Neural Networks

ProtoPNet & Friends