Two Applications of Topological Methods for Neuronal Morphology Analysis

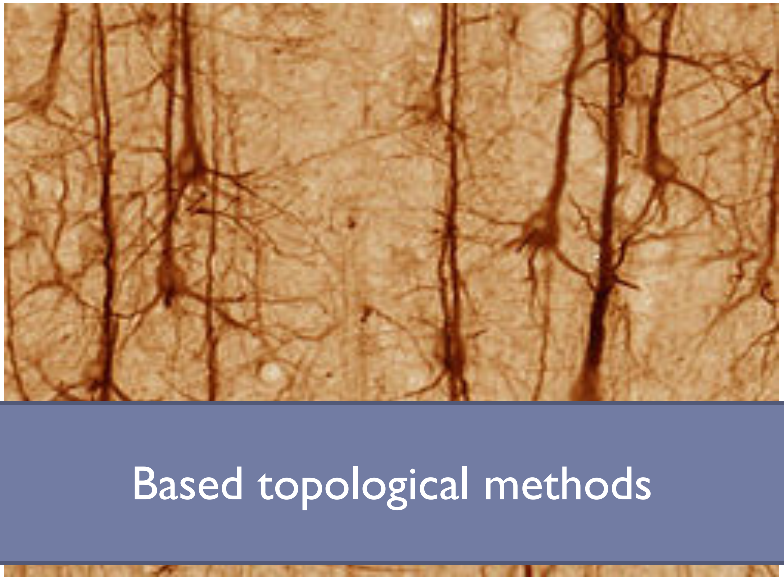
Yusu Wang
Computer Science and Engineering Dept.,
The Ohio State University

Joint work with

Suyi Wang, Yanjie Li (*Ohio State University*),
Partha Mitra (*Cold Spring Harbor Laboratory*)
Giorgio Ascoli (*Krasnow Institute for Advanced Study at George Mason University*)

Introduction

- Neurons essential to the functioning of life
- Neuronal morphology important in neuron functions
- Understanding 3D morphology of individual neurons
  - Reconstruction from 2D/3D images
  - Characterizing and comparing neuron structures



Based topological methods

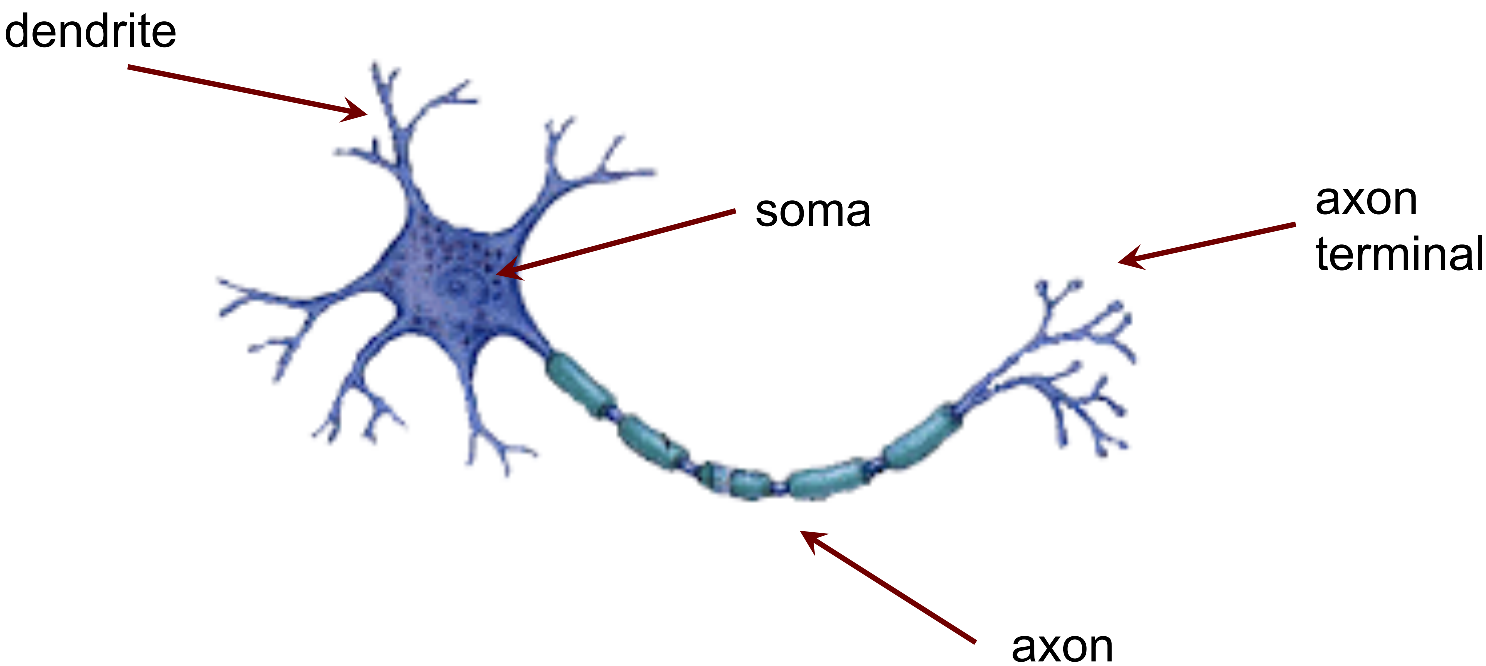
Image from https://en.wikipedia.org/wiki/Neuron

This Talk

Topological methods for:

- Part I:
  - Neuron structures comparison
- Part II:
  - Neuronal Morphology Reconstruction

Neuronal structure 101

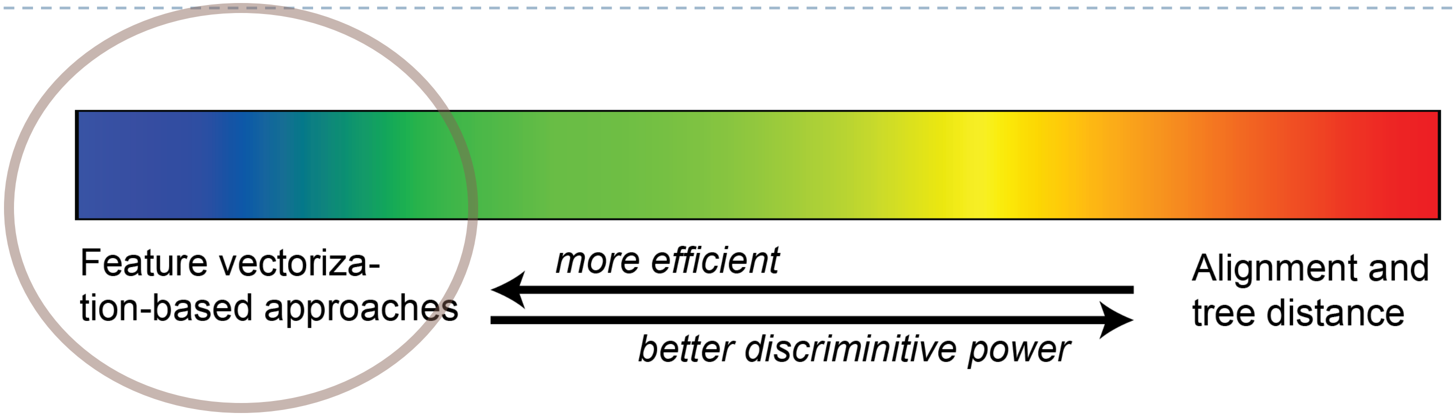


Can be considered as a tree structure with augmented information.

Neuron Structures Comparison

- Large number of neuroanatomical data publically available
  - e.g, FlyCircuit.org, NeuroMorpho.org
- Efficient algorithms to compare neuron structures
  - E.g, to organize / classify large collection of neurons, to understand variability within a cell type, or to identify features

Related Work



- L-measure tool
  - *[Scorcioni et al, 2008]*
- Sholl-like analysis
  - *[Sholl 1953]*
- Arbor density representation
  - *[Sümbül et al 2013]*
- NBLAST
  - *[Costa et al 2016]*

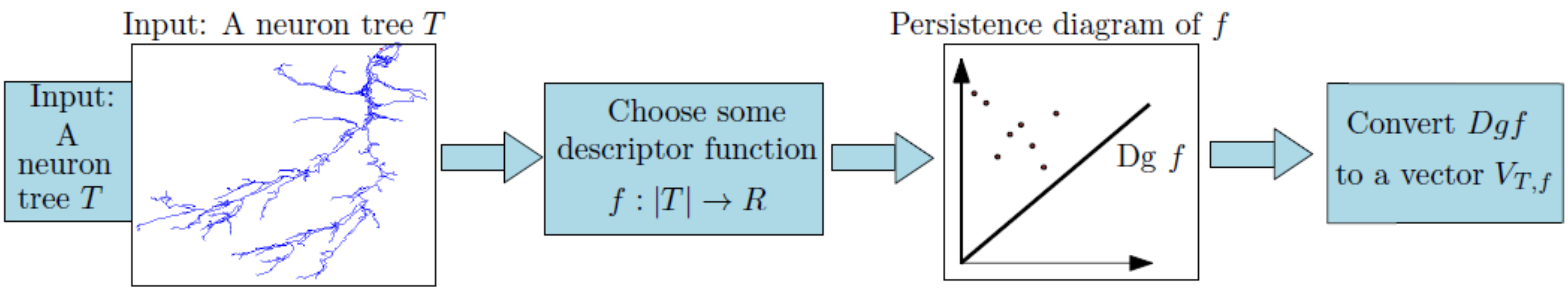
Our goal:

- Simple representation to facilitate efficient comparison,
- yet at the same time discriminative, capturing global tree structure

Develop a persistence-based feature-vectorization and comparison framework.

Vectorization Framework

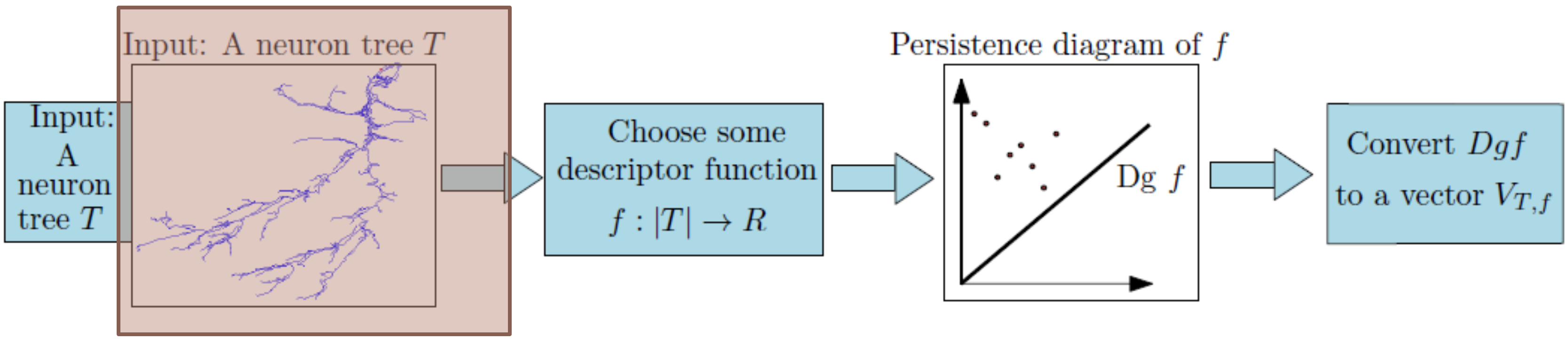
- Persistence-based feature vectorization framework



A similar persistence-based vectorization method was proposed independently in
[Kanari, Dlotko, Scolamiero, Levi, Shillcock, Hess, Markram, arXiv 2016]

Vectorization Framework

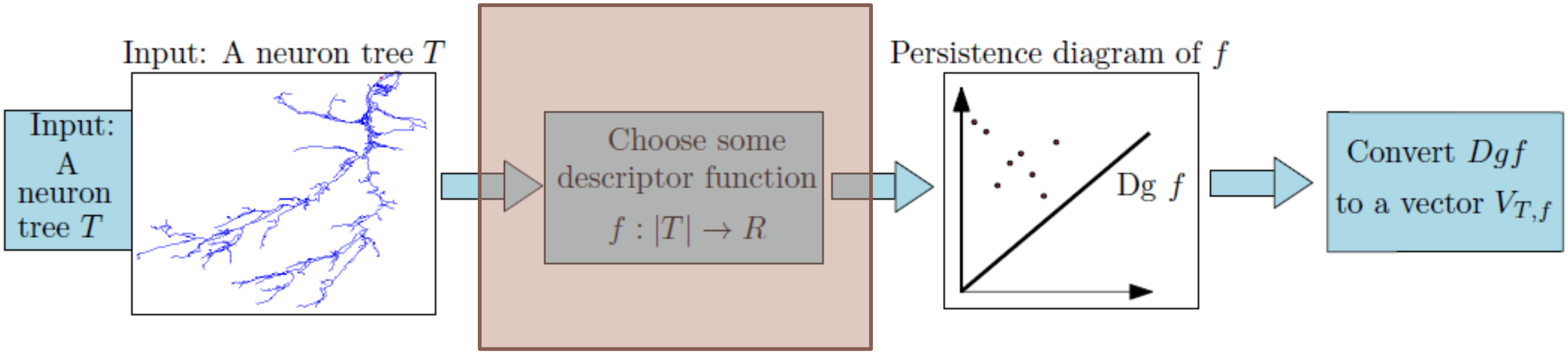
- Persistence-based feature vectorization framework



- Tree representation of neurons
  - A set of tree nodes and arcs, where each arc is modeled by a polygonal curve.
  - Often assume rooted tree with root r located at soma
  - Tree nodes / arc may be associated with other information

Vectorization Framework

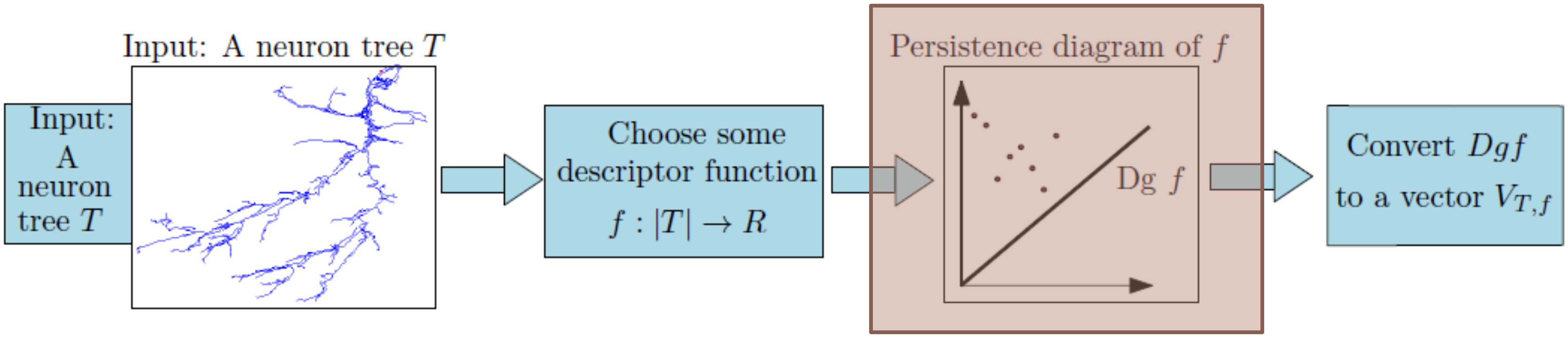
- Persistence-based feature vectorization framework



- Descriptor function(s) on T: $f: |T| \to R$
  - Euclidean distance
    - For any $x \in |T|$, $f(x) = ||x - r||$
  - Geodesic distance
  - L-measure based and other morphological descriptors
  - Electrophysiological measures

Vectorization Framework

- Persistence-based feature vectorization framework



- Given descriptor function $f: |T| \to R$
  - Compute the persistence diagram induced by the sub-level set and super-level set filtrations of f as its summary

Persistent Homology 101

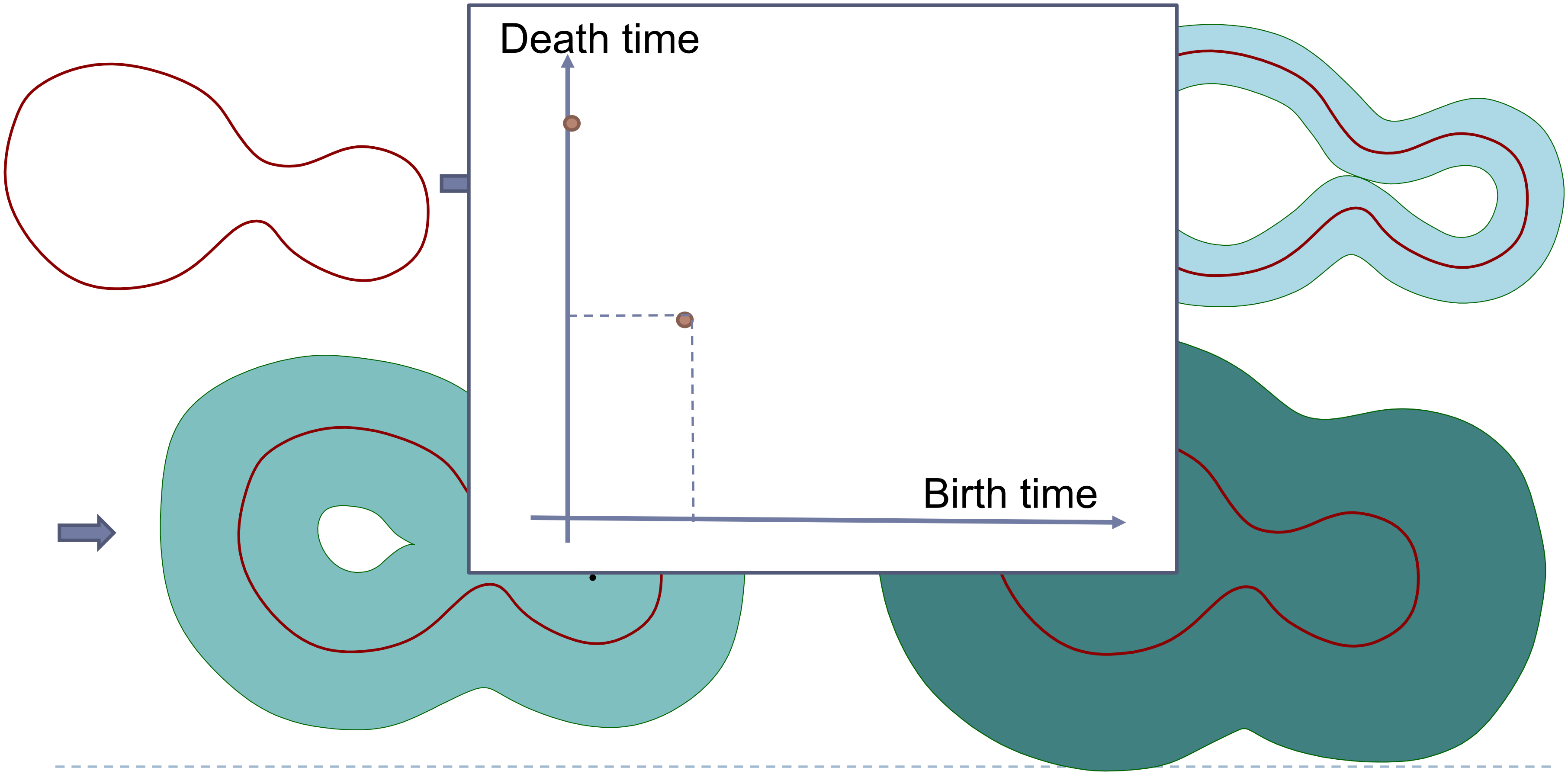
- *[Edelsbrunner, Letscher, Zomordian 2000], [Zomorodian and Carlsson 2005],* Earlier developments: *[Frosini 1990], [Robins 1999]*

- Given a filtration of a space X
  - $X_1 \subset X_2 \subset \cdots X_i \subset \cdots \subset X_j \subset \cdots X_n = X$
  - Consider this as a lens through which we inspect X

- Capture creation and death of ``features” by homology
  - $H_*(X_1) \to \cdots \to H_*(X_i) \to \cdots \to H_*\left(X_j\right) \to \cdots H_*(X_n) = H_*(X)$
  - Summarize the birth/death of homological features in the persistence diagram

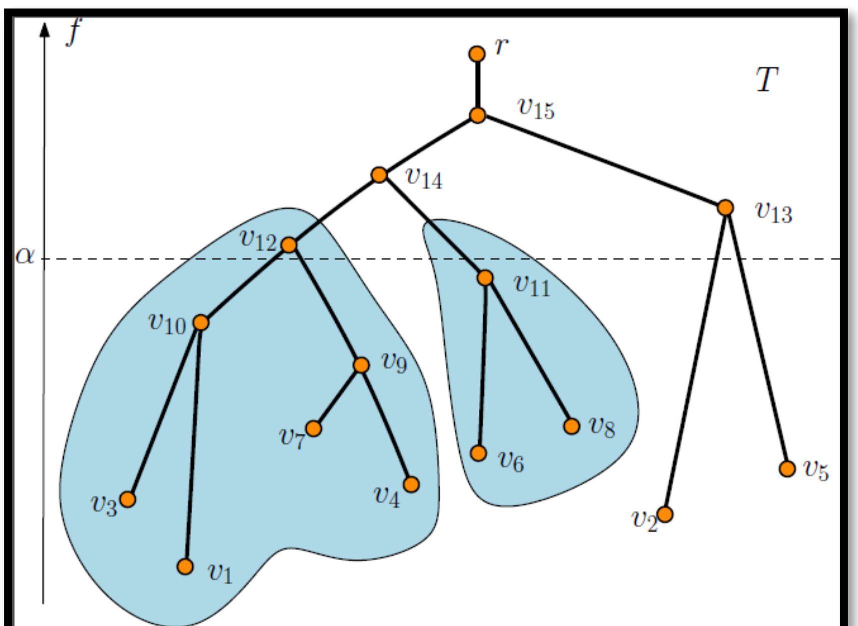
Distance Field Filtration Example

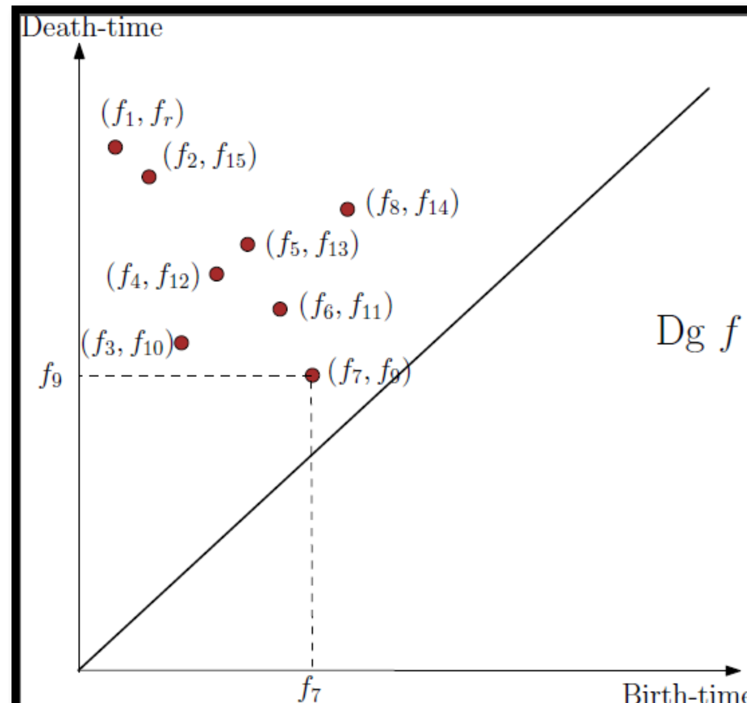
- A filtration induced by distance field.



In Neuron Setting

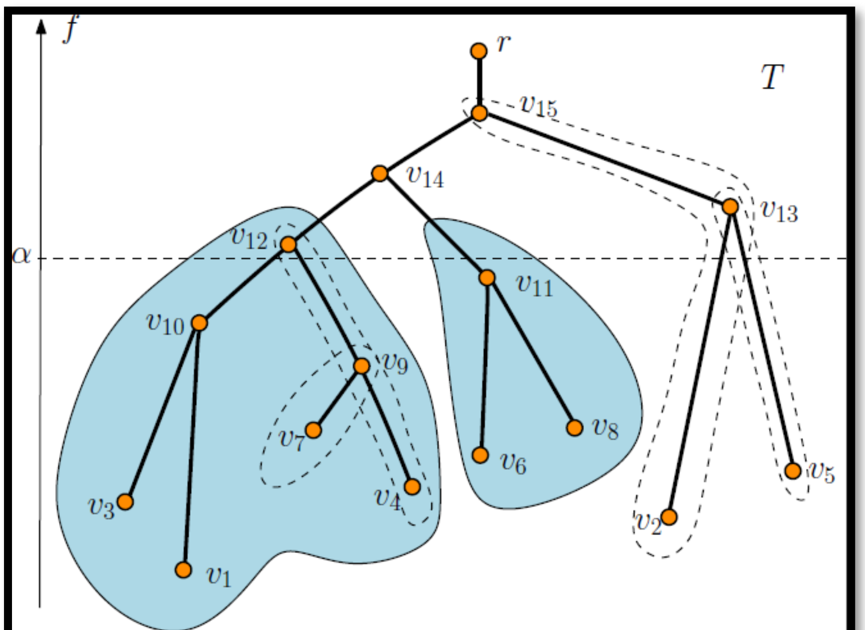
- Assume f is plotted as height function
- Filtration induced by the *sub-level set filtration*
  - $f^{-1}(-\infty, a_0) \subseteq f^{-1}(-\infty, a_1) \subseteq \cdots \subseteq f^{-1}(-\infty, a_n) = T$

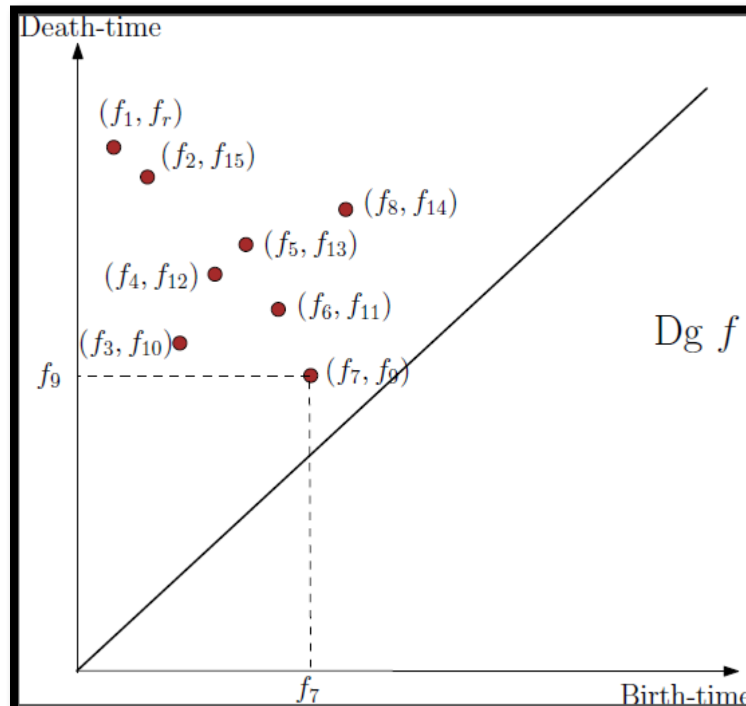




In Neuron Setting

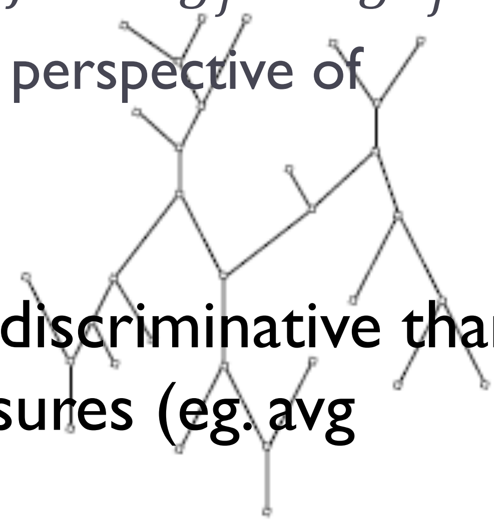
- Assume f is plotted as height function
- Filtration induced by the *sub-level set filtration*
  - $f^{-1}(-\infty, a_0) \subseteq f^{-1}(-\infty, a_1) \subseteq \cdots \subseteq f^{-1}(-\infty, a_n) = T$





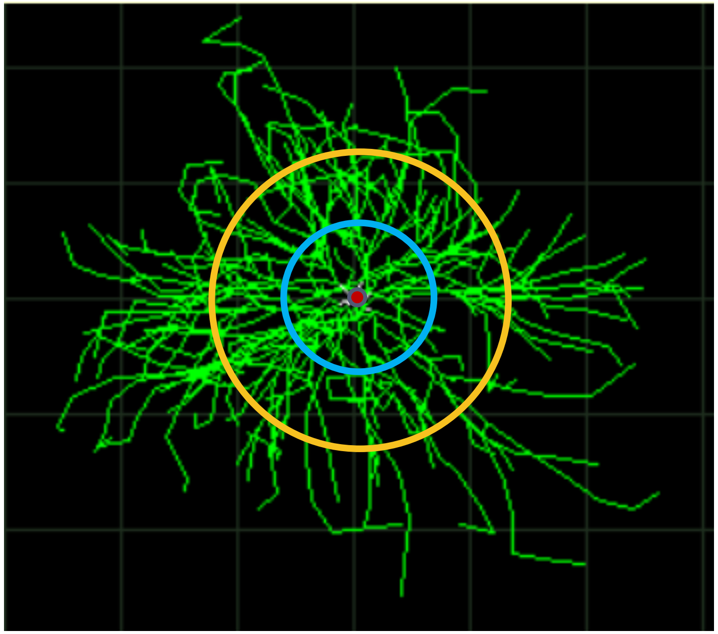
Remarks

- Depending on the descriptor function $f: |T| \to R$, a tree may have both down-forks and up-forks.
  - Also consider *super-level sets filtration*, and its induced persistence diagram Dg_{-f}
- Given a descriptor function f,
  - Obtain persistence diagram summary $Dgf = Dg_f \cup Dg_{-f}$
  - $Dg\ f$ serves as a summary of T from the perspective of descriptor function f
- Persistence-summary intuitively more discriminative than simply statistics of morphological measures (eg. avg branching angles)



Connection to Sholl-like Analysis

- Sholl function $N: R^+ \to R^+$
  - $N(\lambda) \coloneqq$ number of intersection of T with a circle (sphere) centered at the root r with radius λ

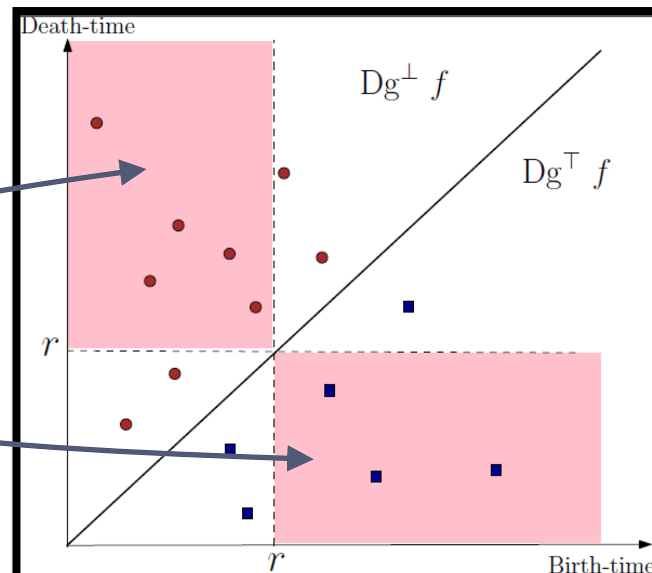


Connection to Sholl-like Analysis

- Sholl function $N: R^+ \rightarrow R^+$
  - $N(\lambda) :=$ number of intersection of T with a circle (sphere) centered at the root r with radius λ

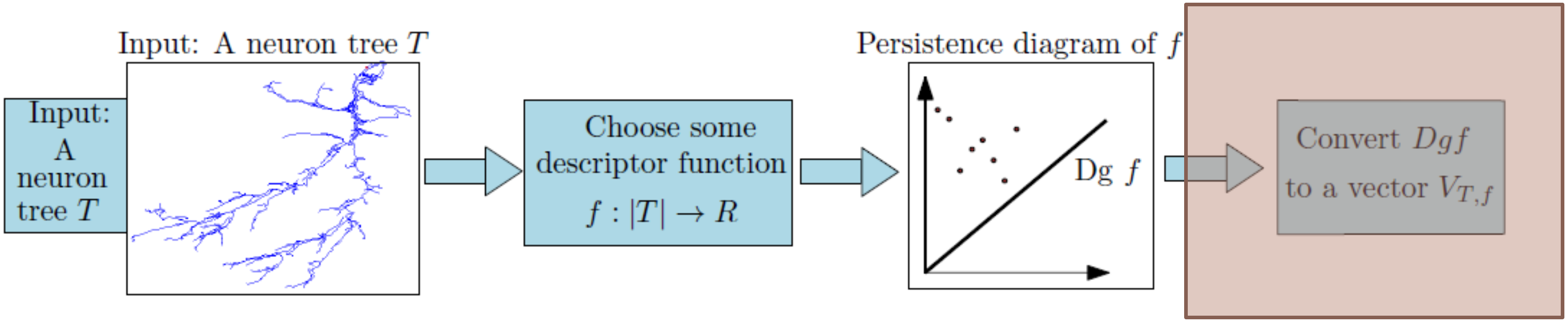
- One can recover full Sholl function from persistence diagrams induced by Euclidean distance function

$N(r) =$ total number of points in these two regions



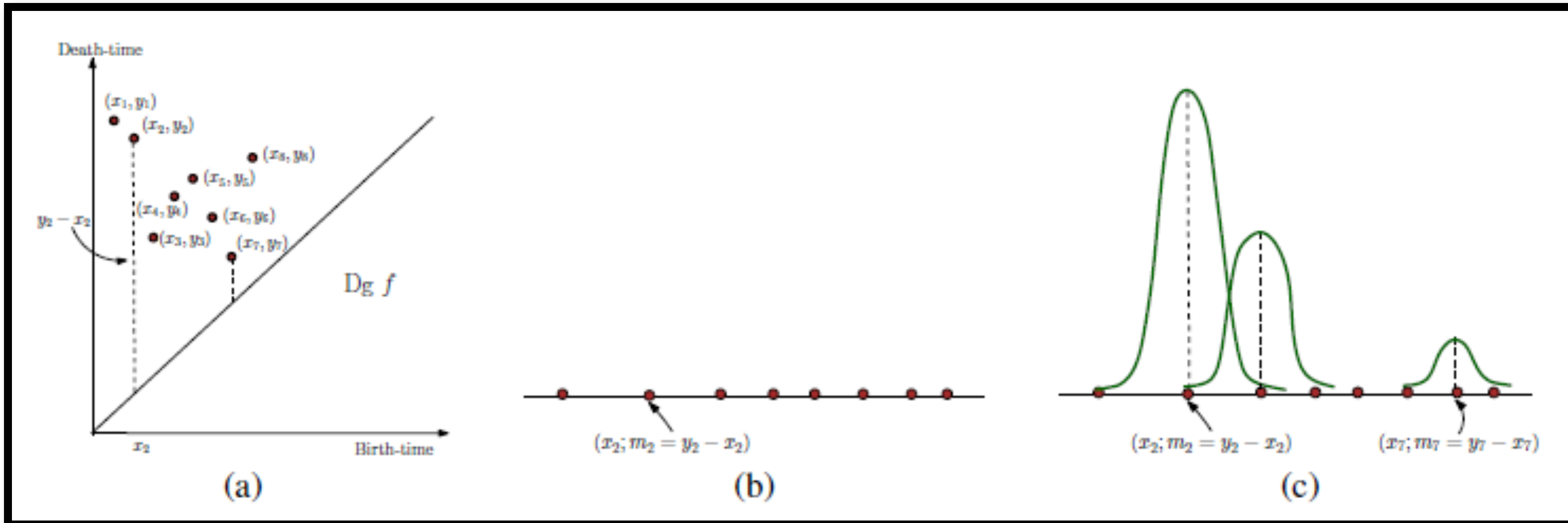
Vectorization Framework

- Persistence-based feature vectorization framework



- To facilitate efficient distance computation
  - Convert persistence diagram $Dg\ f$ to a featue vector $V_{T,f}$
  - *[Bubenik 2012], [Reininghaus et al 2015], [Adams et al 2015]*,...

Feature Vectorization

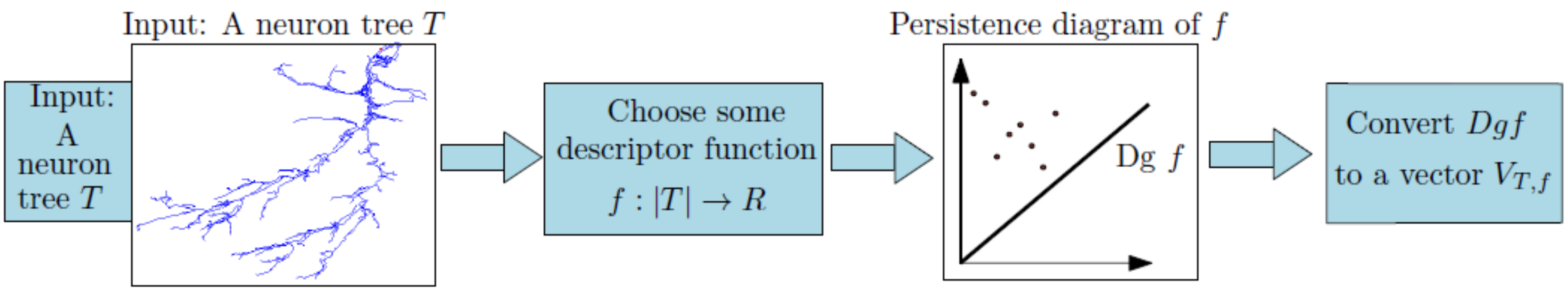


(a) (b) (c)

- Convert diagram D to a 1D density field
  - $\rho_D(x) := \sum_{i=1 \in k} m_i \cdot K_t(x, x_i)$, for any $x \in \mathbb{R}$,
- Discretize it to a m-vector
  - $\nu_D := [\rho_D(\mathsf{a} + \frac{I}{m}),\ \rho_D(\mathsf{a} + \frac{2I}{m}),\ \cdots, \rho_D(\mathsf{a} + \frac{mI}{m}) = \rho_D(\mathsf{b})].$

Vectorization Framework

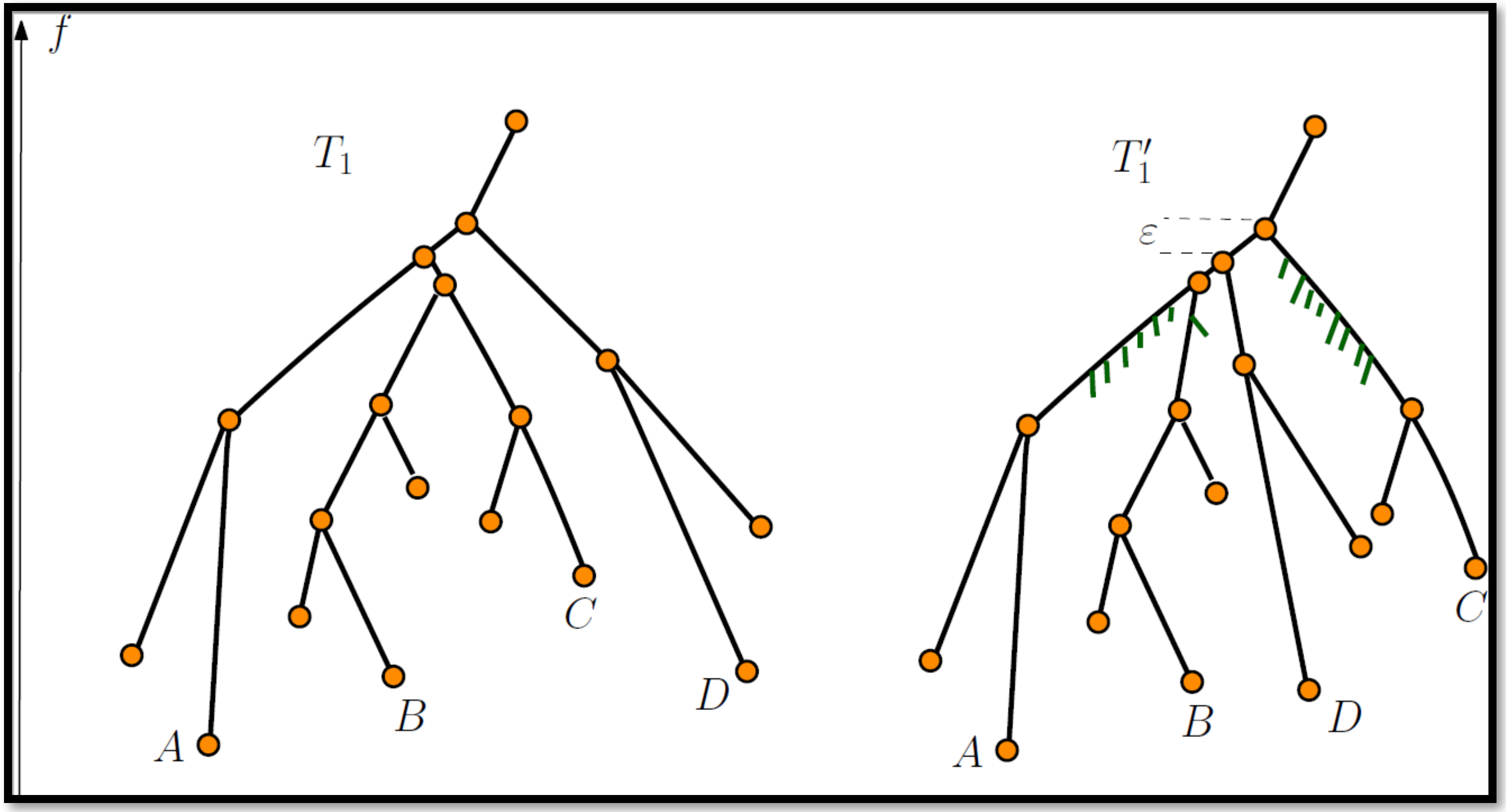
- Persistence-based feature vectorization framework



- If there are multiple descriptor functions
  - Concatenate their respective feature vectors
  - Perform dimensionality reduction to reduce dimension

Remarks

- Versatile framework
  - Can combine multiple type of information of neurons, morphological or electrophysiological measures
  - Easy to add new measurements
- Discreminative features
  - E.g, persistence features from Euclidean function contains more information than Sholl function
  - E.g, persistence features from geodesic function encodes global morphological information
- Have certain stability guarantees



f
T_1
A
B
C
D
T_1'
ε
A
B
C
D

Three Test Datasets

- Dataset 1:
  - 379 neurons taken from neuromorpho.org category Drosophila-Chklovskii, manually categorized into 89 types
  - *[Takemura et al, 2013]*
- Dataset 2:
  - 127 neurons from four families: *Purkinje, olivocerebellar neurons, Spinal motoneurons* and *hippocampal interneurons*, downloaded also from neuronmorpho.org
- Dataset 3:
  - 1268 neurons from Human Brian Project, downloaded from neuromorpho.org. Two primary cell classes: interneurons and principal cells, known for 1130 cells
  - *[Markram et al 2015]*

Preliminary Results

- Leave-one-out classification tests based on k-nearest neighbors

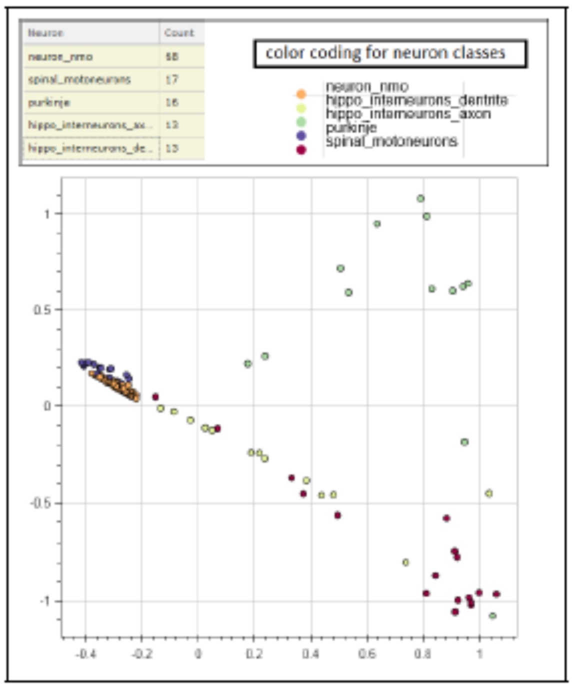
	# neurons for classified correctly out of 346 for all non-singleton classes Dataset 1		
# nearest neighbors	Persist-distance d_P	Persist-vec d_V	Sholl-distance d_S
1	190	164	104
2	221	199	142
3	235	226	160
4	250	235	170
5	262	239	180

	# neurons for classified correctly out of 127 neurons in Dataset 2		
# nearest neighbors	Persist-distance d_P	Persist-vec d_V	Sholl-distance d_S
1	117	111	92
2	121	118	108
3	122	120	113
4	122	120	117
5	123	121	124

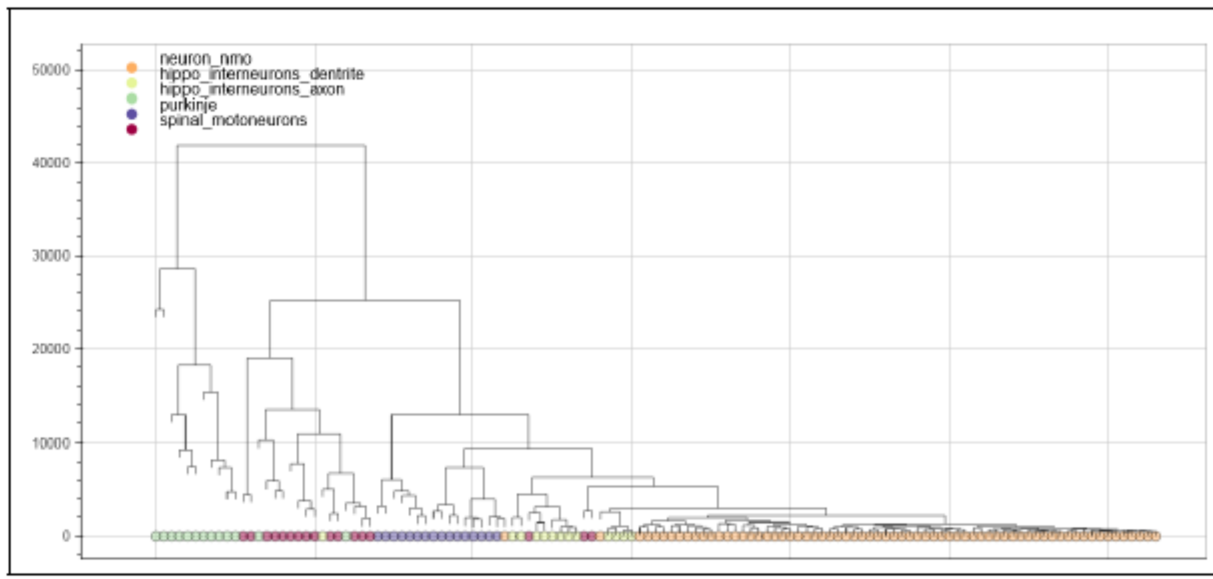
	# neurons for classified correctly out of 1130 in Dataset 3		
# nearest neighbors	Persist-distance d_P	Persist-vec d_V	Sholl-distance d_S
1	832	812	763
2	990	985	942
3	1065	1054	1019
4	1093	1083	1058
5	1107	1100	1081

Preliminary Results

- Clustering for Dataset 2



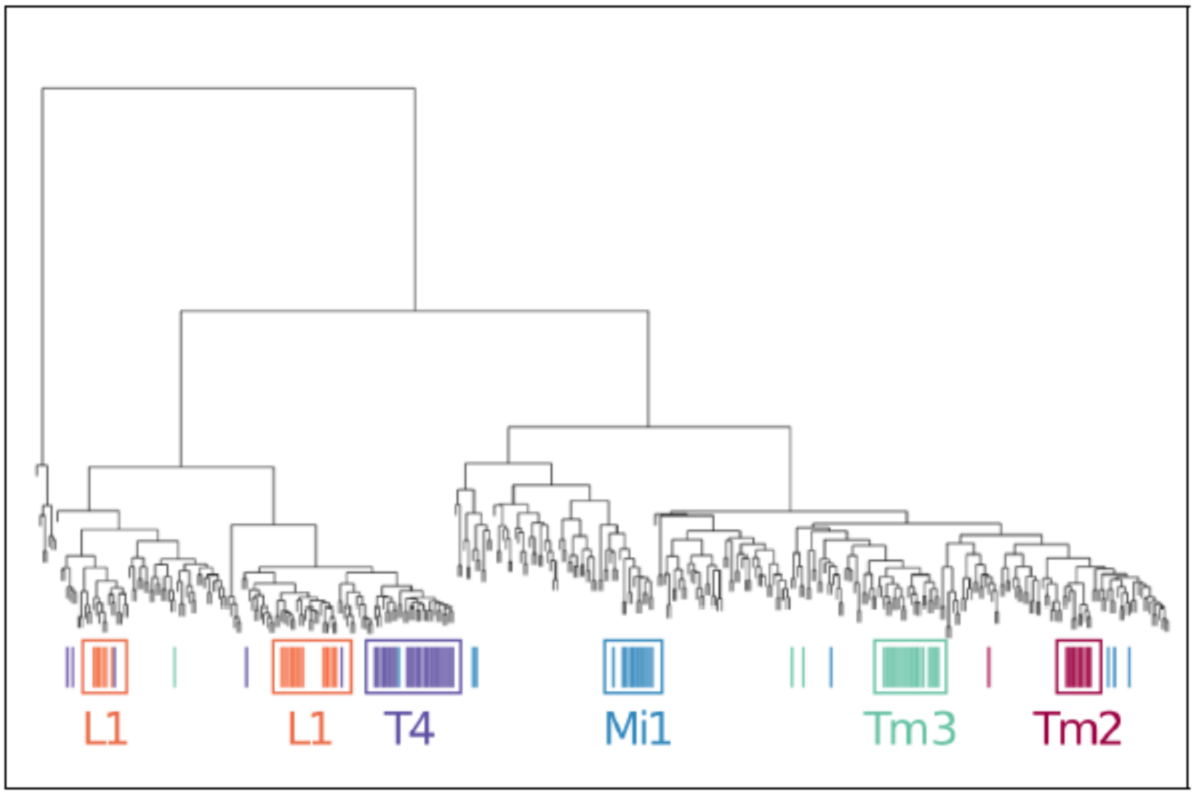
(a)



(b)

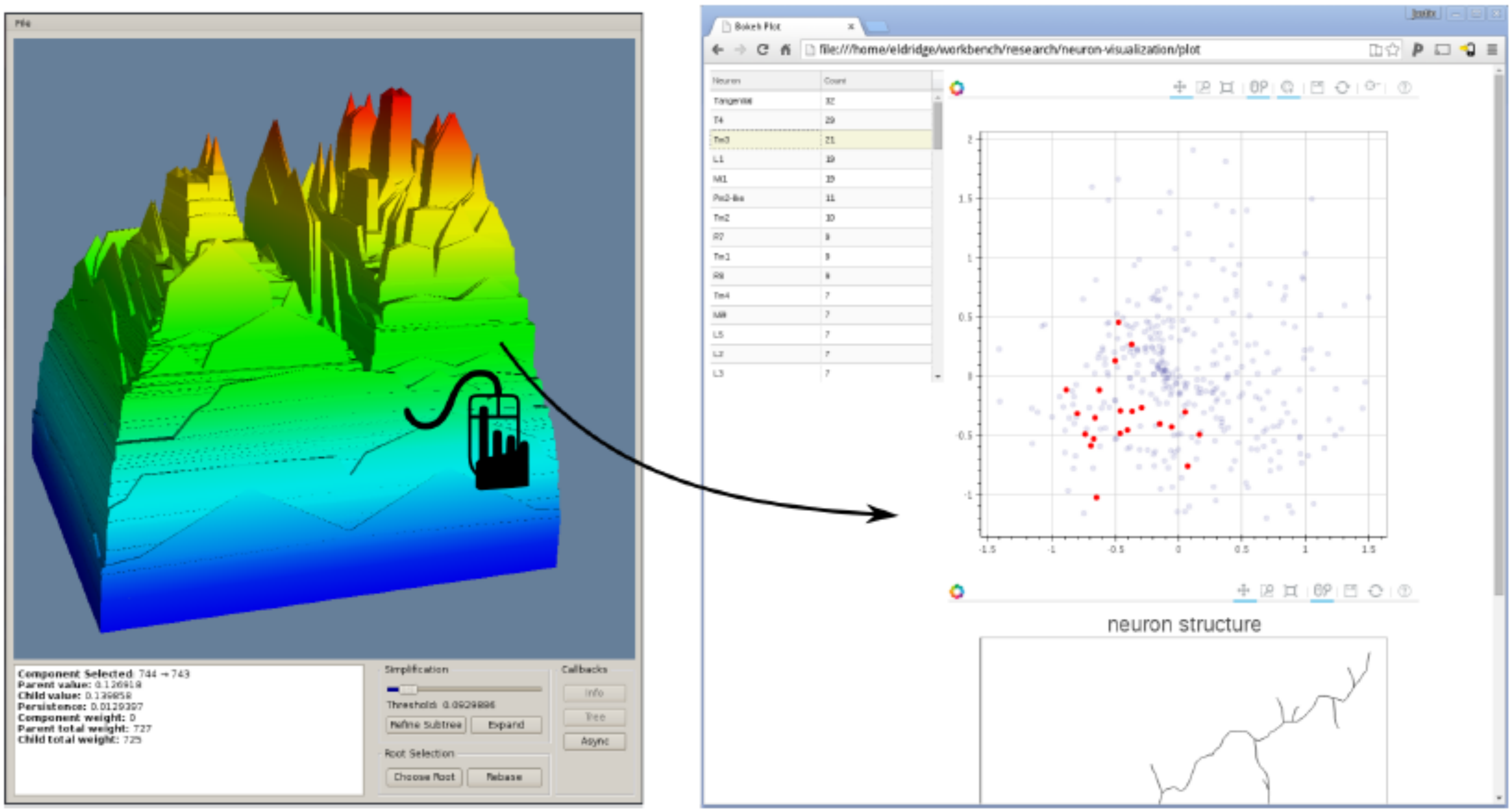
Preliminary Results

- Clustering for Dataset 1
  - Five largest families other than “Tangential”



Preliminary Results

- An interactive visualization tool

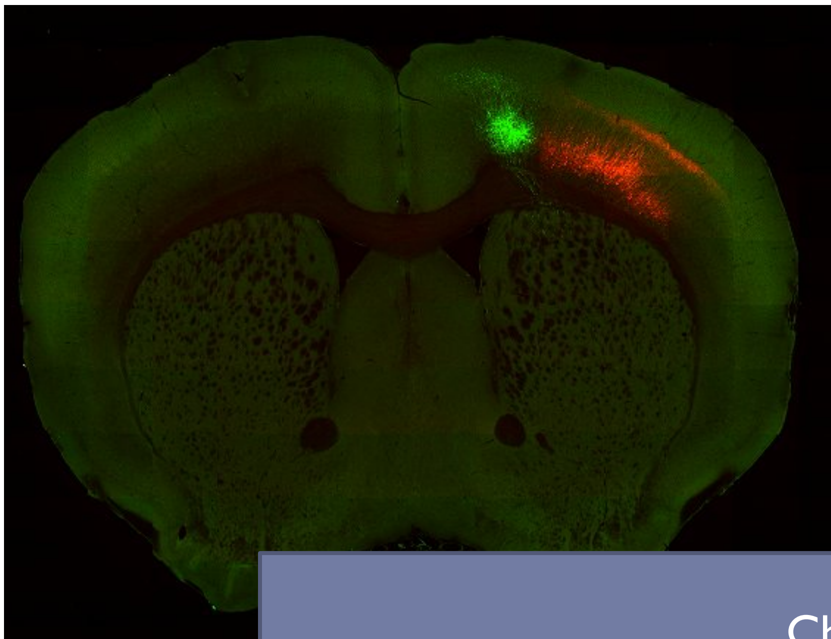


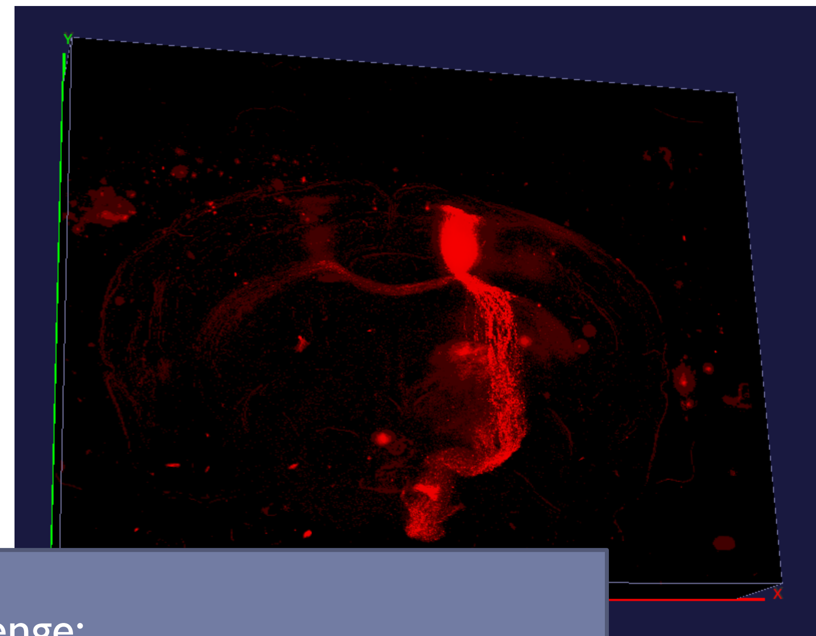
This Talk

- Part I:
  - Neuron structures comparison
- Part II:
  - Neuronal Morphology Reconstruction

Neuronal Morphology Reconstruction

- Various imaging techniques produce large number of 2D/3D images





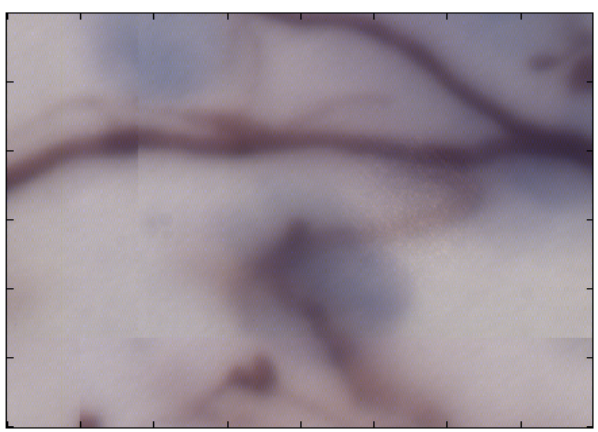
Challenge:
Automatic reconstruction of neuronal morphology from various imaging data.

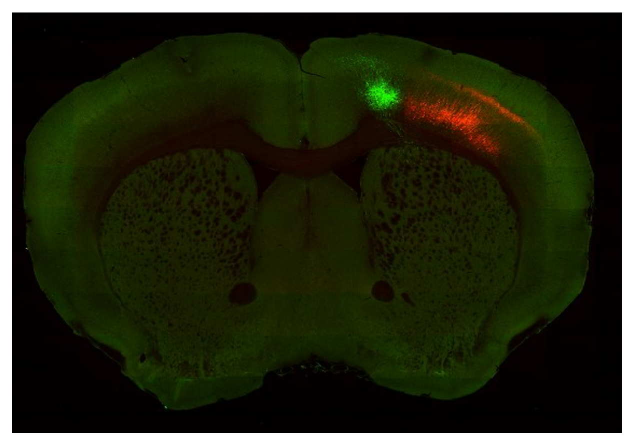
Related Work

- *DIADEM* challenge (2009—2010)
  - *Diginal Reconstruction of Axonal and Dendritic Morphology*
  - http://diademchallenge.org/
- *BigNeuron* (launched in 2015)
  - Large-scale 3D single neuron reconstruction
  - Sponsored by 14 neuroscience-related research organizations and international research groups
  - https://www.alleninstitute.org/bigneuron/about/
- Many algorithms already integrated into platform *Vaa3D*
  - *[Peng et al., 2010]* www.vaa3D.org.

The Problem

- On the high level:
  - Given a 2D / 3D image data, the goal is to extract one (or multiple) tree-like structure(s) from it.
- Some challenges:
  - Various types of background ``noise”
  - Non-homogeneous distribution of signal in raw data
  - Mixture of multiple neurons

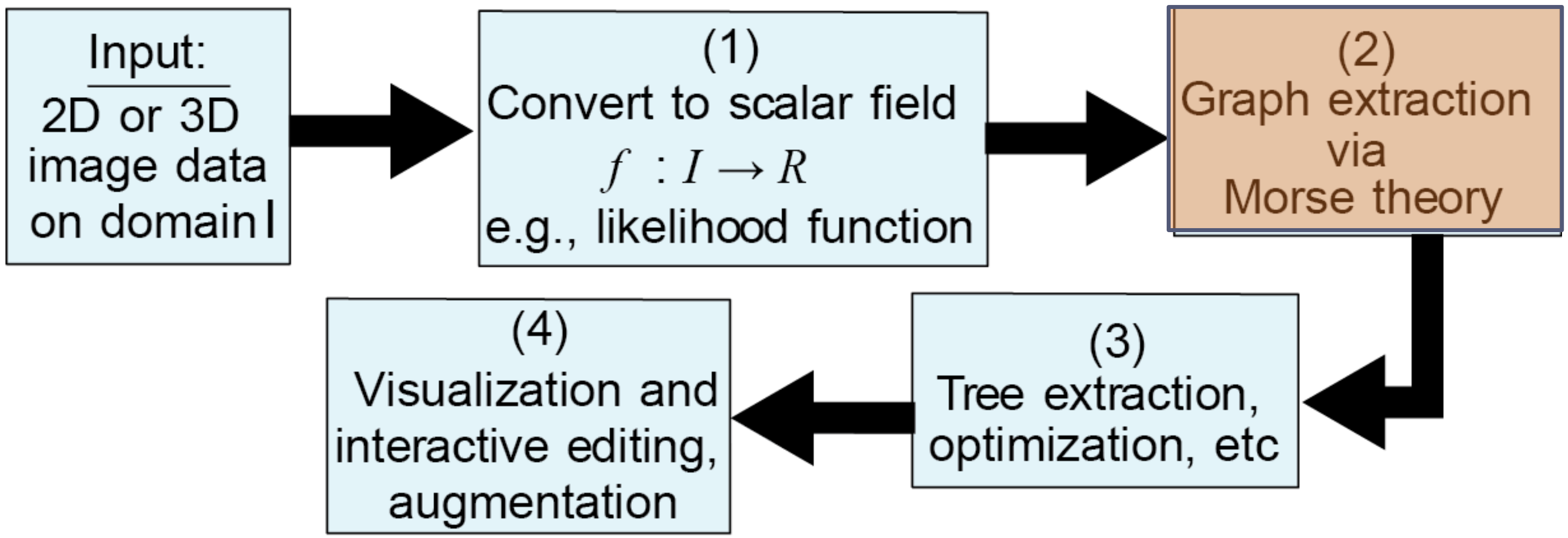




- On the high level:
  - Given a 2D / 3D image data, the goal is to extract one (or multiple) tree-like structure(s) from it.

- Previous methods:
  - Often rely on local information for decision making, sensitive to noise
  - Some thresholding involved, challenging in handling non-uniform signal distribution
  - Junction nodes identification challenging
    - E.g, ``growing" individual branches and ``gluing" them to obtain tree topology

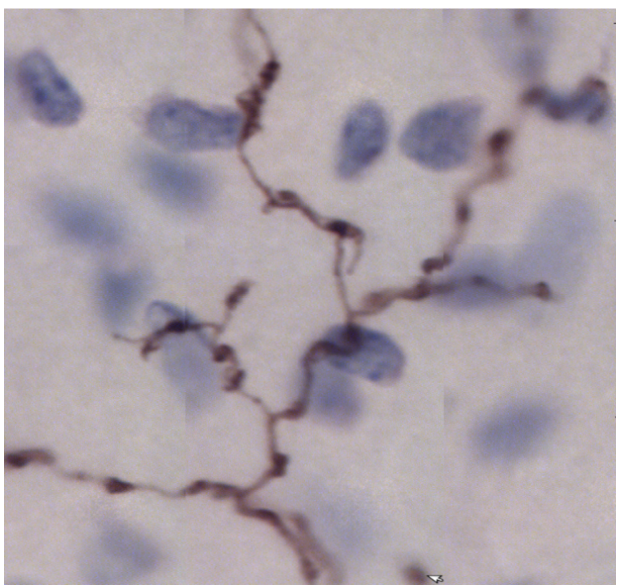
Morse-based Reconstruction Framework

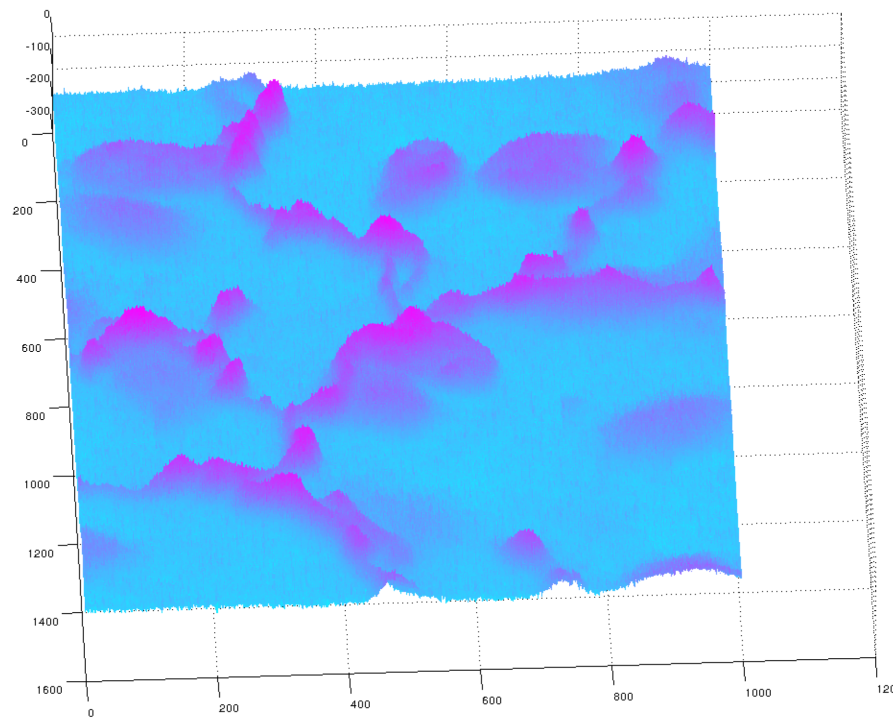


- Morse-based approach
  - uses global structure behind data
  - junction nodes identification reliable w/o special processing
  - robust against noise, small gaps, and non-uniformity in data
  - conceptually clean, helps reducing pre-processing of data

Main Idea

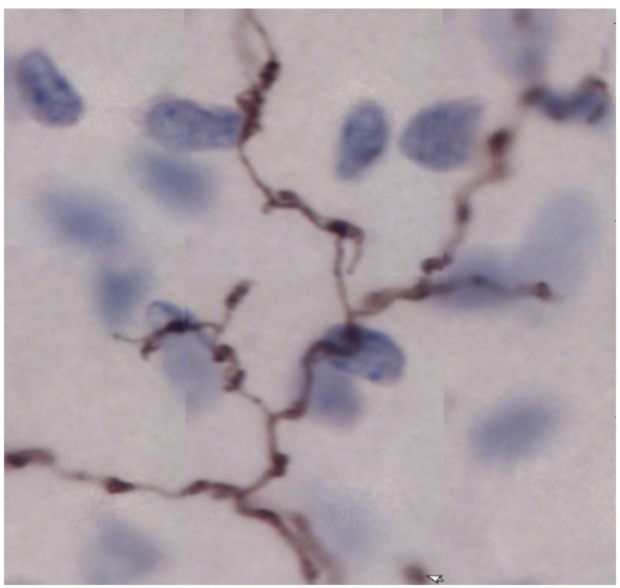
- Assume input is a scalar field
  - $f: I \to R$, where high value of f indicates high signal value
- Consider graph f as a terrain (mountain range) on $I \times R$
  - I can be $[0,1]^2 \subset R^2$ or $[0,1]^3 \subset R^3$

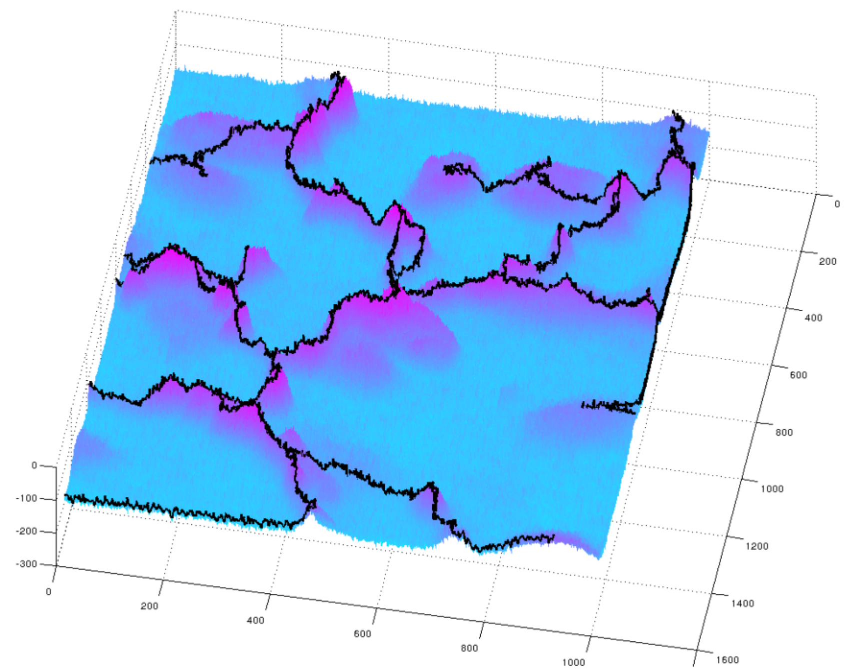




Main Idea

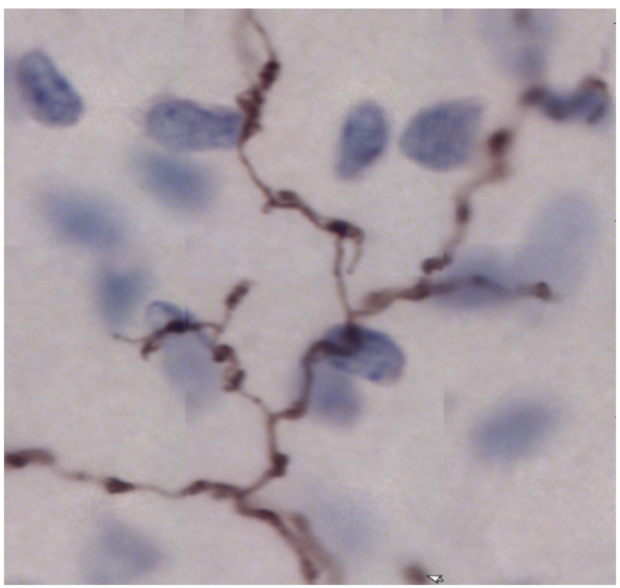
- Assume input is a scalar field
  - $f: I \to R$, where high value of f indicates high signal value
- Consider graph f as a terrain (mountain range) on $I \times R$
  - I can be $[0,1]^2 \subset R^2$ or $[0,1]^3 \subset R^3$

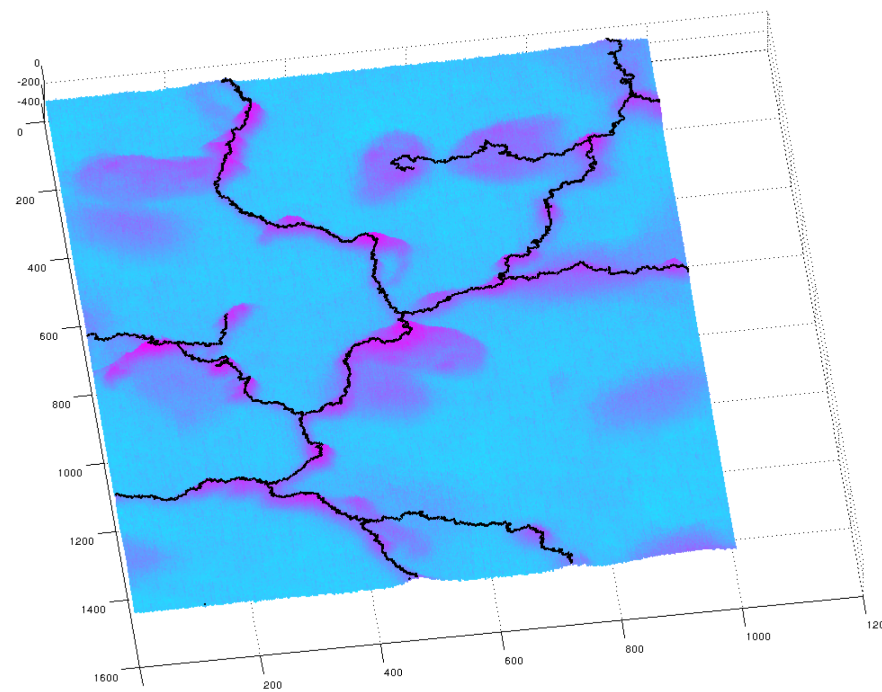




Main Idea

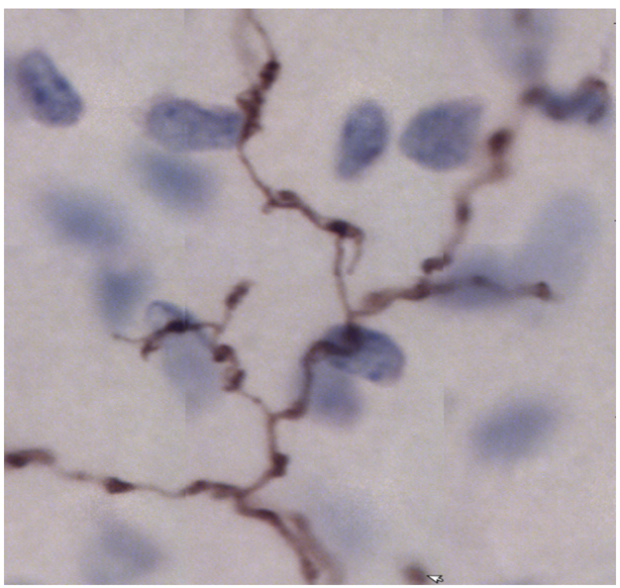
- Assume input is a scalar field
  - $f: I \to R$, where high value of f indicates high signal value
- Consider graph f as a terrain (mountain range) on $I \times R$
  - I can be $[0,1]^2 \subset R^2$ or $[0,1]^3 \subset R^3$

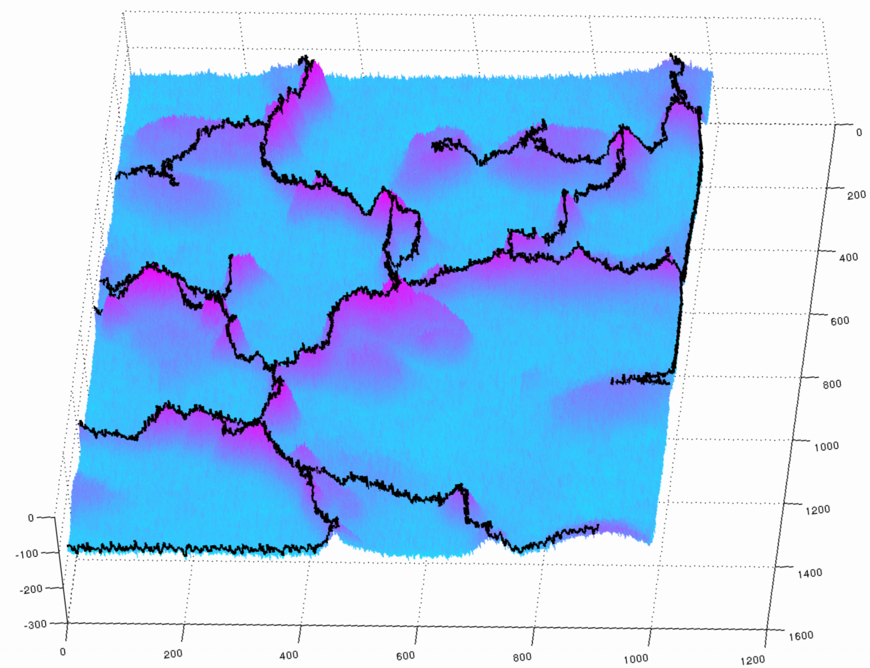




Main Idea

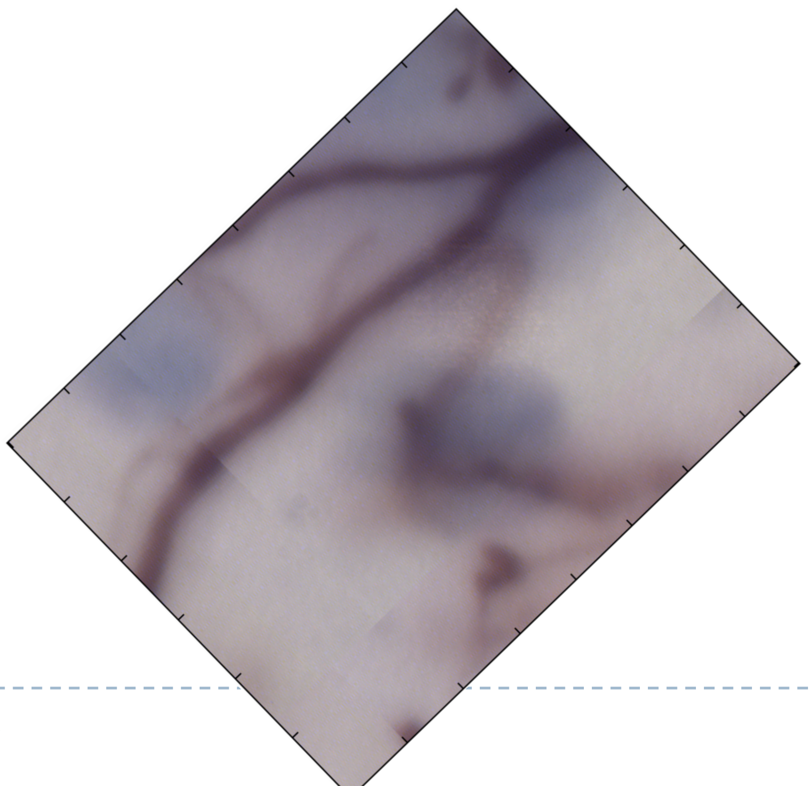
- Assume input is a scalar field
  - $f: I \rightarrow R$, where high value of f indicates high signal value
- Consider graph f as a terrain (mountain range) on $I \times R$
  - I can be $[0,1]^2 \subset R^2$ or $[0,1]^3 \subset R^3$

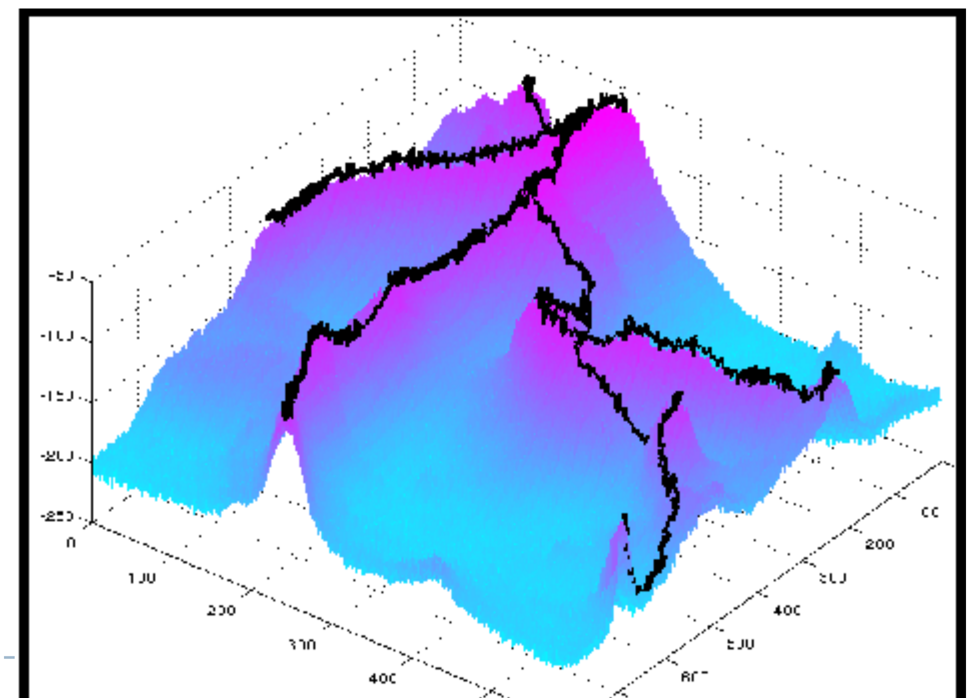


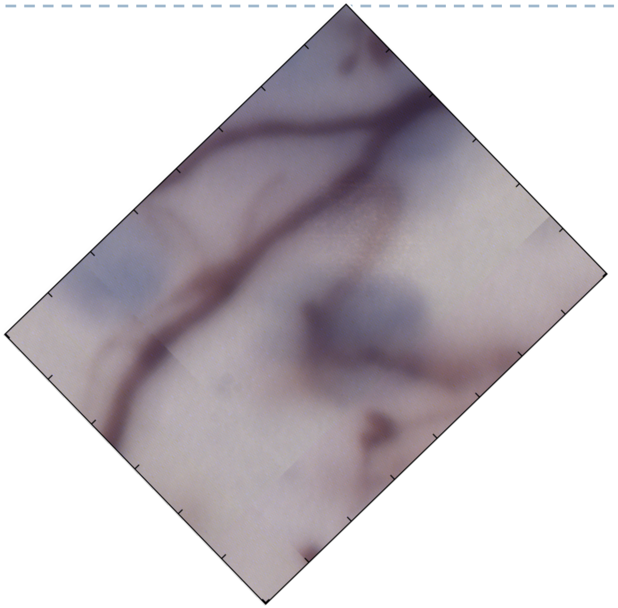


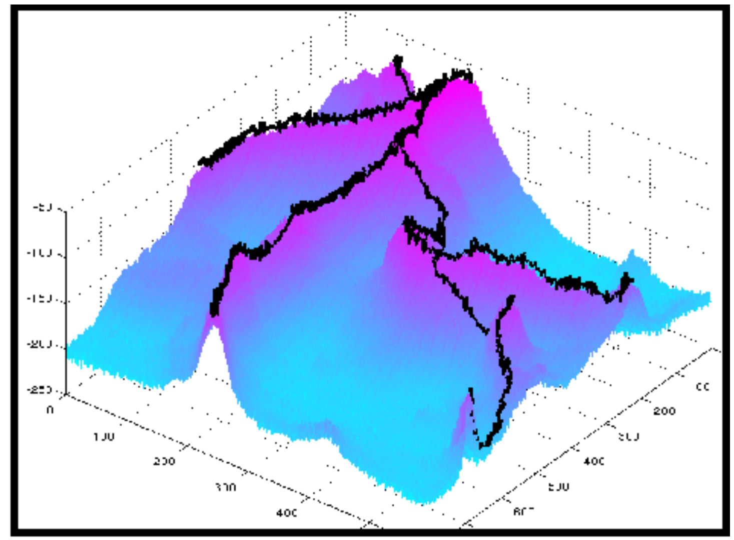
Main Idea

- Assume input is a scalar field
  - $f : I \to R$, where high value of f indicates high signal value
- Consider graph f as a terrain (mountain range) on $I \times R$
  - I can be $[0,1]^2 \subset R^2$ or $[0,1]^3 \subset R^3$







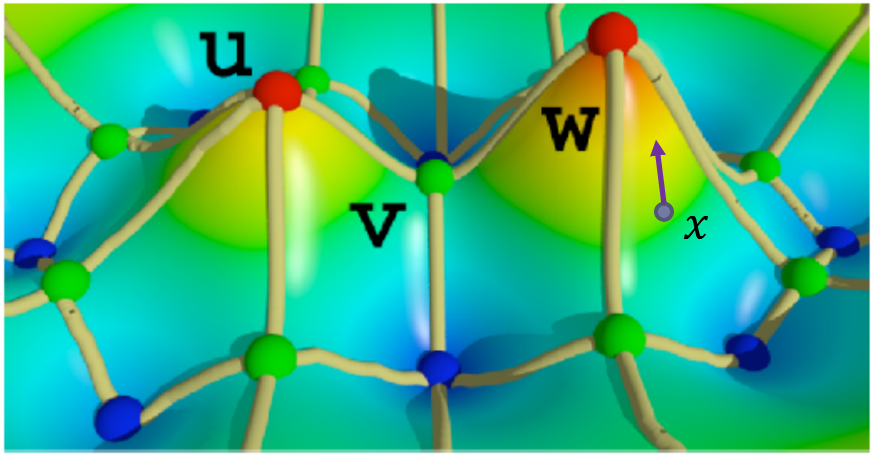


The neuron structure ``tends to'' correspond to ridges of this terrain.

Use Morse theory to help identify ``mountain ridges''.

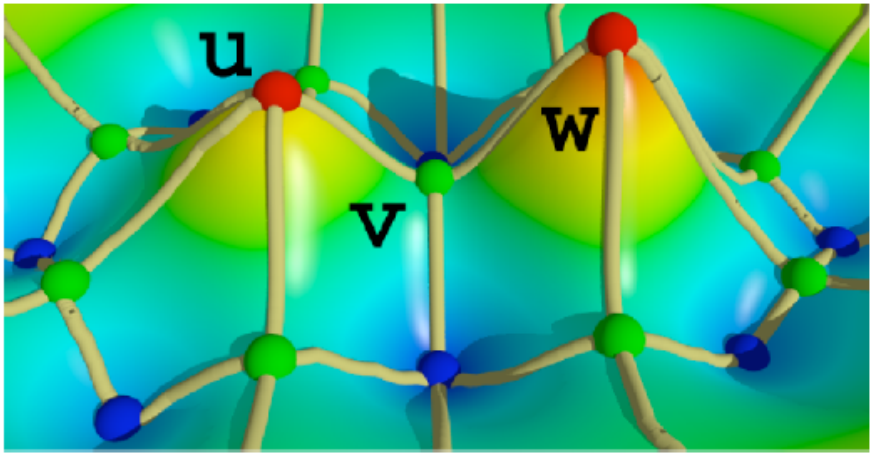
Morse Theory: Smooth Case

- Let $f: R^d \rightarrow R$ be a Morse function
- Gradient of f at x: $\nabla f(x) = \left[\frac{\partial f}{x_1}, \frac{\partial f}{x_2}, \dots, \frac{\partial f}{x_d}\right]^T$



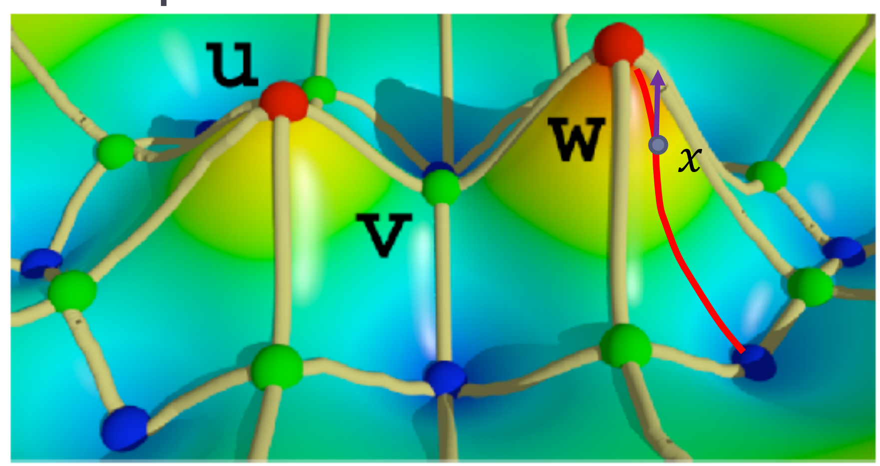
Morse Theory: Smooth Case

- Let $f: R^d \rightarrow R$ be a Morse function
- Gradient of f at x: $\nabla f(x) = \left[\frac{\partial f}{x_1}, \frac{\partial f}{x_2}, \ldots, \frac{\partial f}{x_d}\right]^T$
- Critical points of f:
  - $\{ x \in R^d \mid \nabla f(x) = 0 \}$



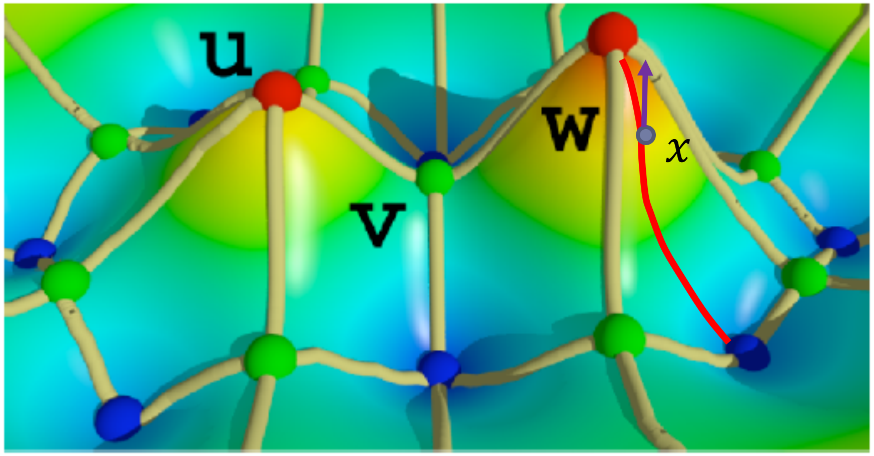
Morse Theory: Smooth Case

- Let $f: R^d \rightarrow R$ be a Morse function
- Gradient of f at x: $\nabla f(x) = \left[\frac{\partial f}{x_1}, \frac{\partial f}{x_2}, \ldots, \frac{\partial f}{x_d}\right]^T$
- Critical points of f: $\{ x \in R^d \mid \nabla f(x) = 0 \}$
- An integral line $L: (0, 1) \rightarrow R^d$:
  - a maximal path in R^d whose tangent vectors agree with gradient of f at every point of the path



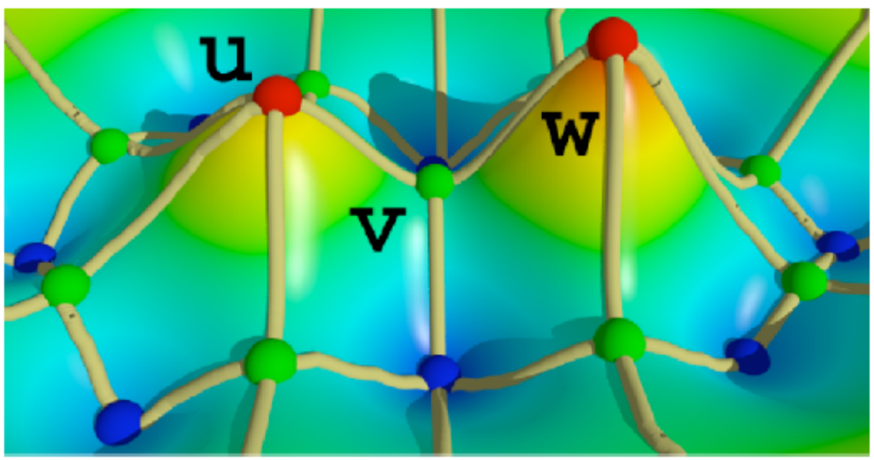
Morse Theory: Smooth Case

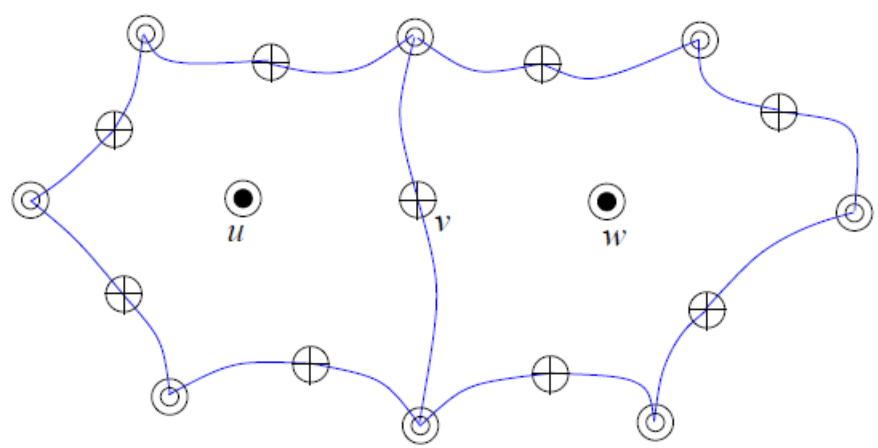
- Let $f: R^d \to R$ be a Morse function
- Gradient of f at x: $\nabla f(x) = \left[\frac{\partial f}{x_1}, \frac{\partial f}{x_2}, \dots, \frac{\partial f}{x_d}\right]^T$
- Critical points of f: $\{ x \in R^d \mid \nabla f(x) = 0 \}$
- An integral line $L: (0, 1) \to R^d$:
  - a maximal path in R^d whose tangent vectors agree with gradient of f at every point of the path
  - has origin and destination at critical points
    - $Dest(L) = \lim_{p \to 1} L(p)$
    - $Ori(L) = \lim_{p \to 0} L(p)$



Stable / Unstable Manifolds

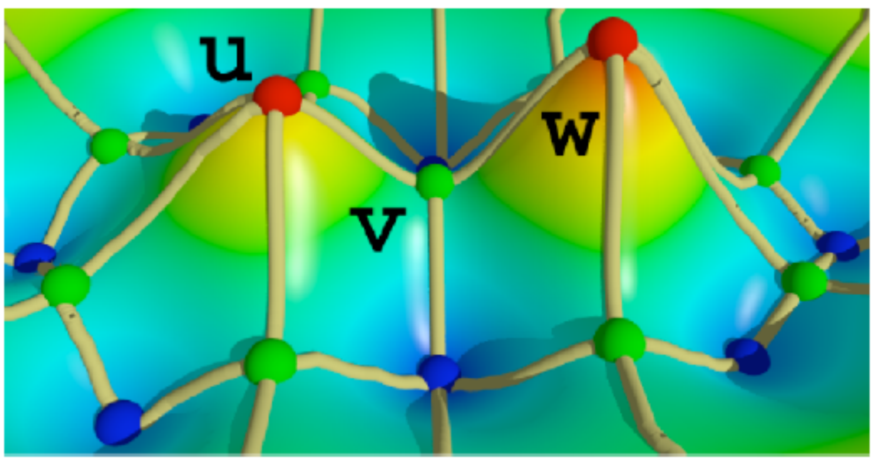
- Given a critical point x of f
  - Stable manifold $S(x) = \{ y \in R^d \mid dest(y) = x \}$
  - Unstable manifold $U(x) = \{ y \in R^d \mid ori(y) = x \}$
- Morse complex, Morse-Smale complex

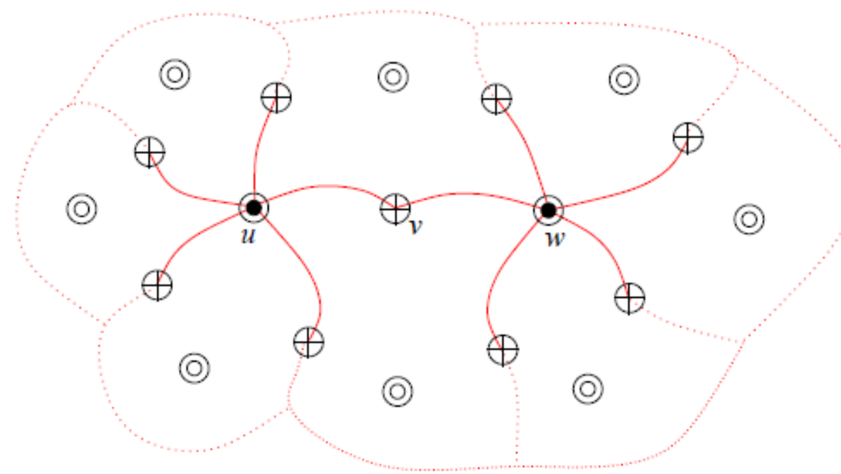




Stable / Unstable Manifolds

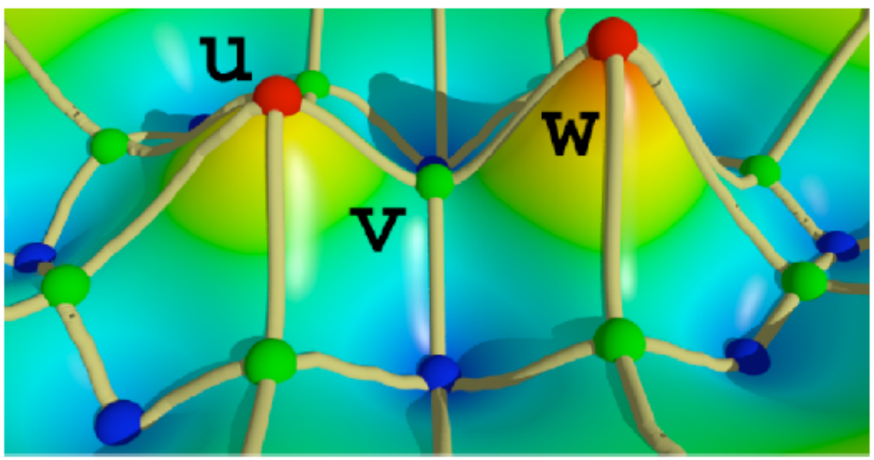
- Given a critical point x of f
  - Stable manifold $S(x) = \{\ y \in R^d \mid dest(y) = x\ \}$
  - Unstable manifold $U(x) = \{\ y \in R^d \mid ori(y) = x\ \}$
- Morse complex, Morse-Smale complex

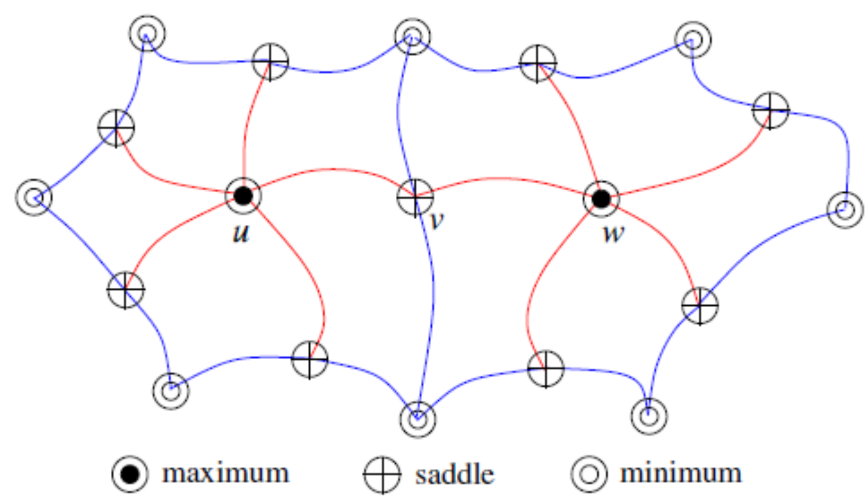




Stable / Unstable Manifolds

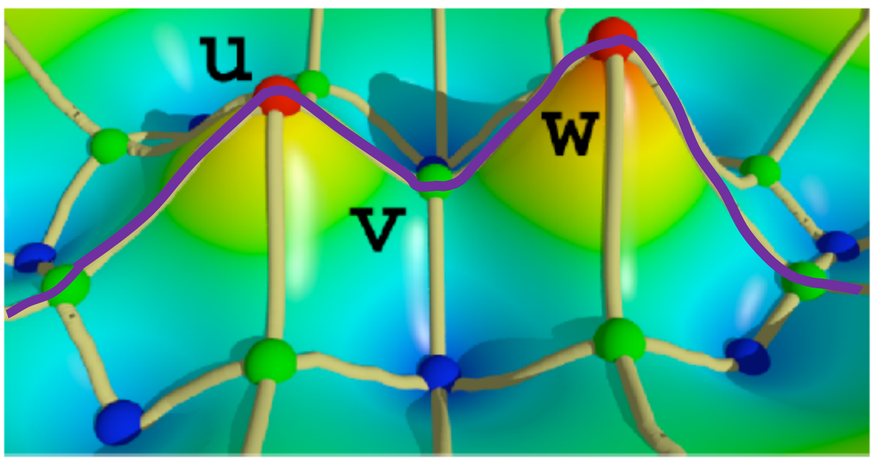
- Given a critical point x of f
  - Stable manifold $S(x) = \{ y \in R^d \mid dest(y) = x \}$
  - Unstable manifold $U(x) = \{ y \in R^d \mid ori(y) = x \}$
- Morse complex, Morse-Smale complex

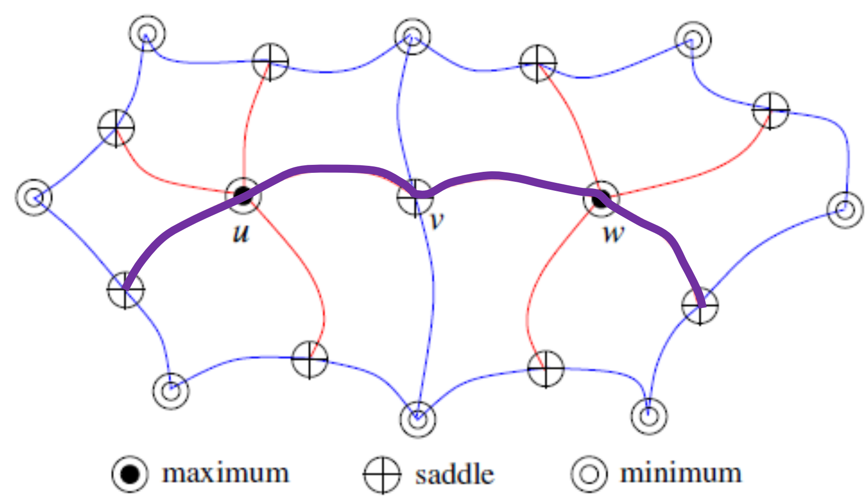




Stable / Unstable Manifolds

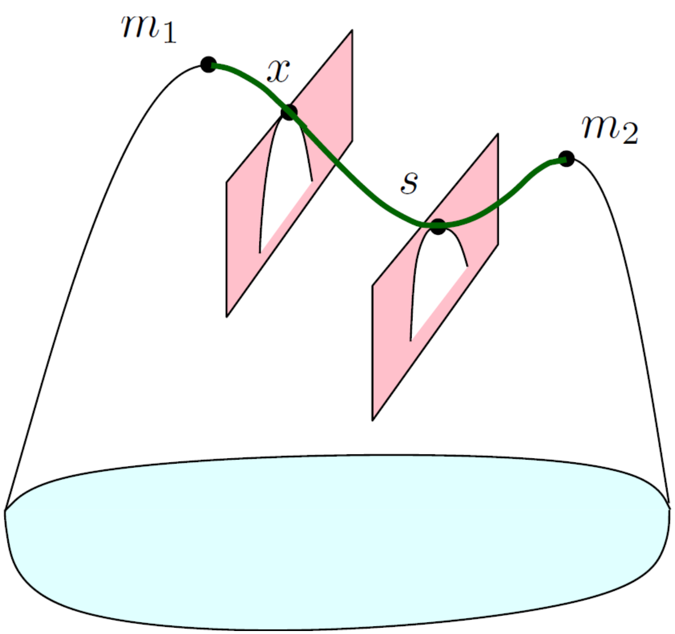
- Given a critical point x of f
  - Stable manifold $S(x) = \{ y \in R^d \mid dest(y) = x \}$
  - Unstable manifold $U(x) = \{ y \in R^d \mid ori(y) = x \}$
- Morse complex, Morse-Smale complex

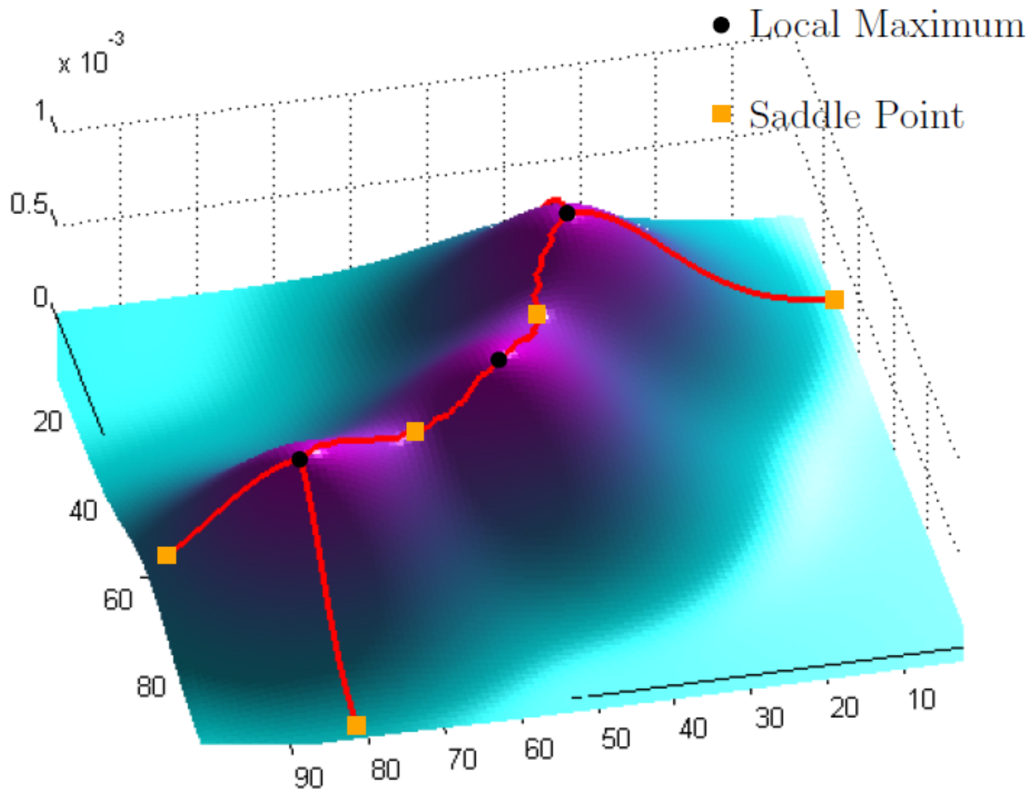




1-unstable manifold (of index $d-1$ saddle points) => mountain ridges

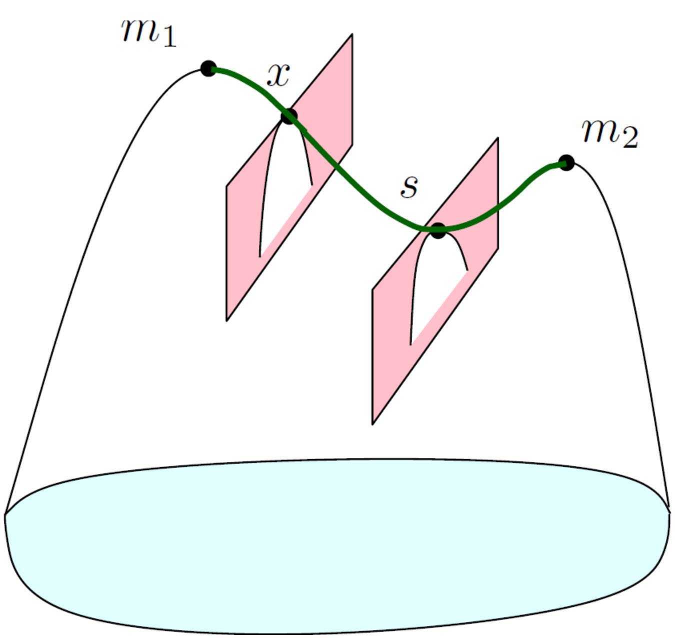
1-unstable Manifold

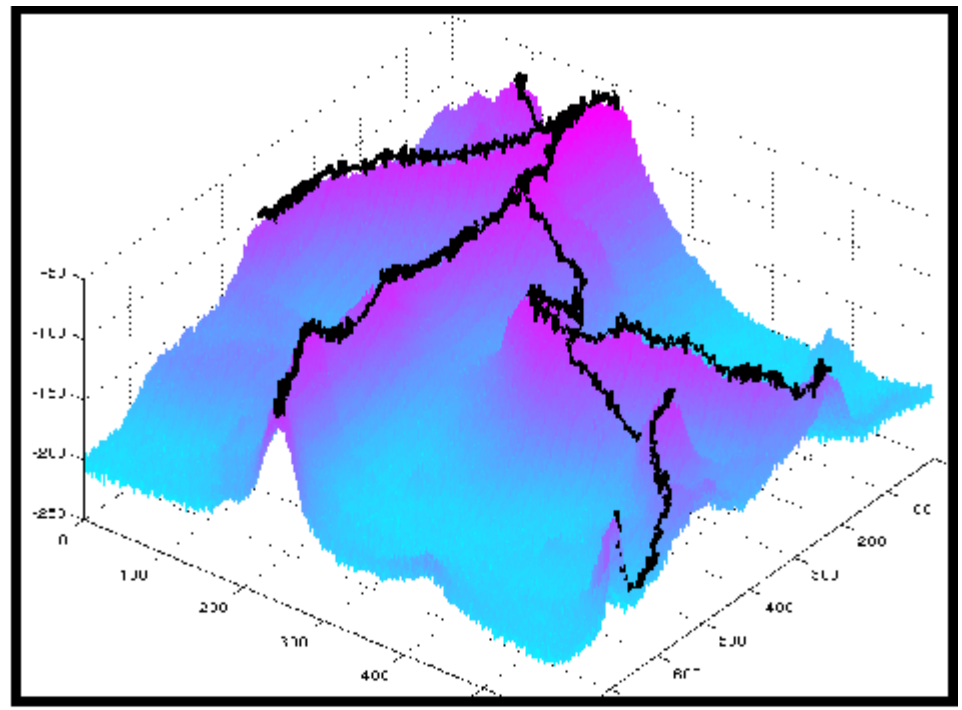




1-unstable manifold (of index $d-1$ saddle points) => mountain ridges

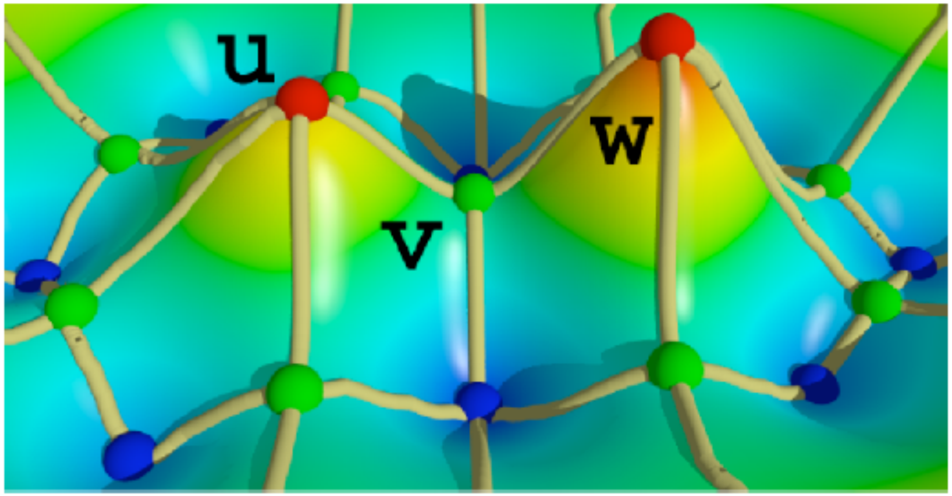
1-unstable Manifold

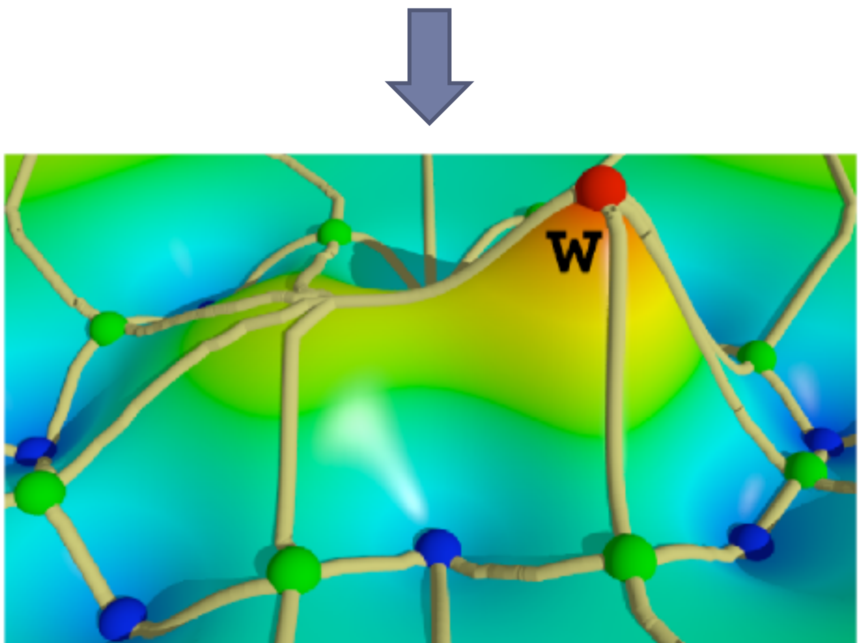




1-unstable manifold (of index $d-1$ saddle points) => mountain ridges

Simplification

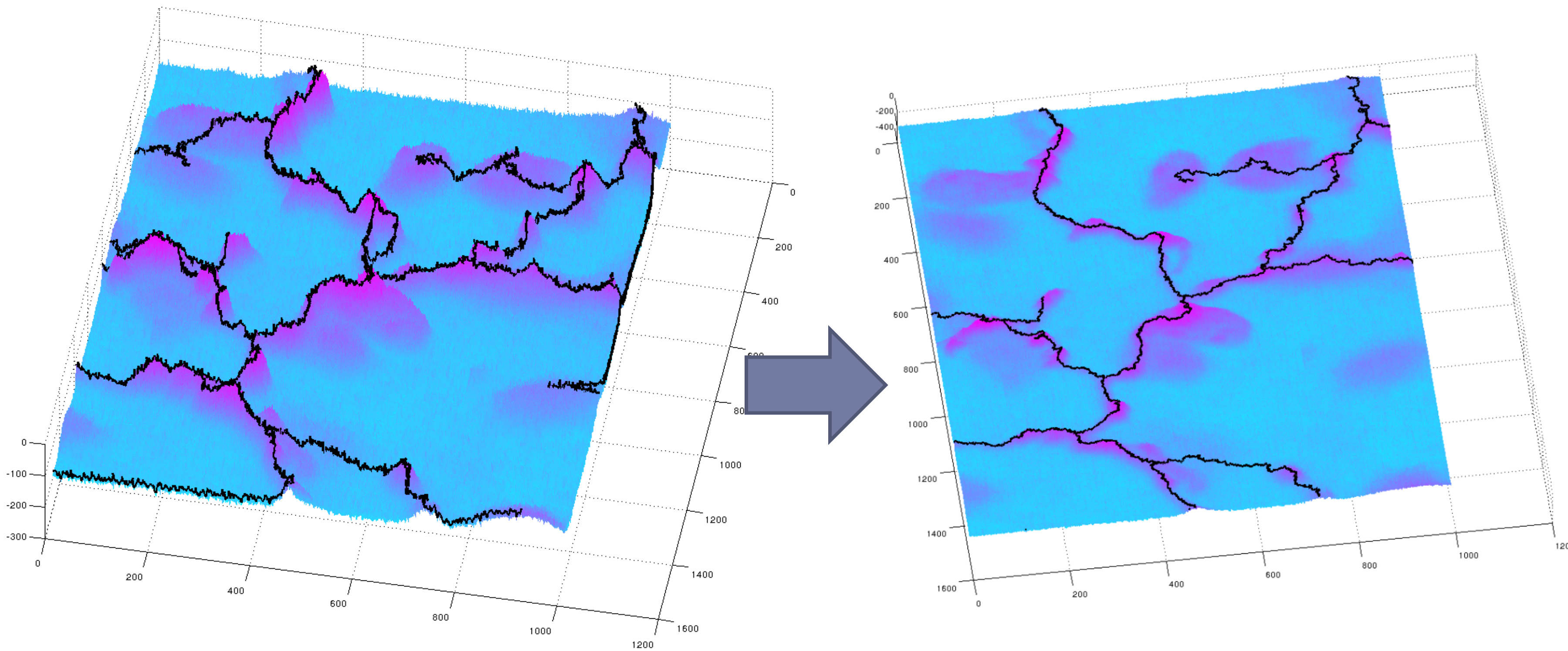




- How to decide which pair of critical points to simplify?
  - Use persistence homology

[Edelsbrunner, Letscher, Zomorodian 2002], [Zomorodian, Carlsson 2005], …

Simplification



Discrete Case

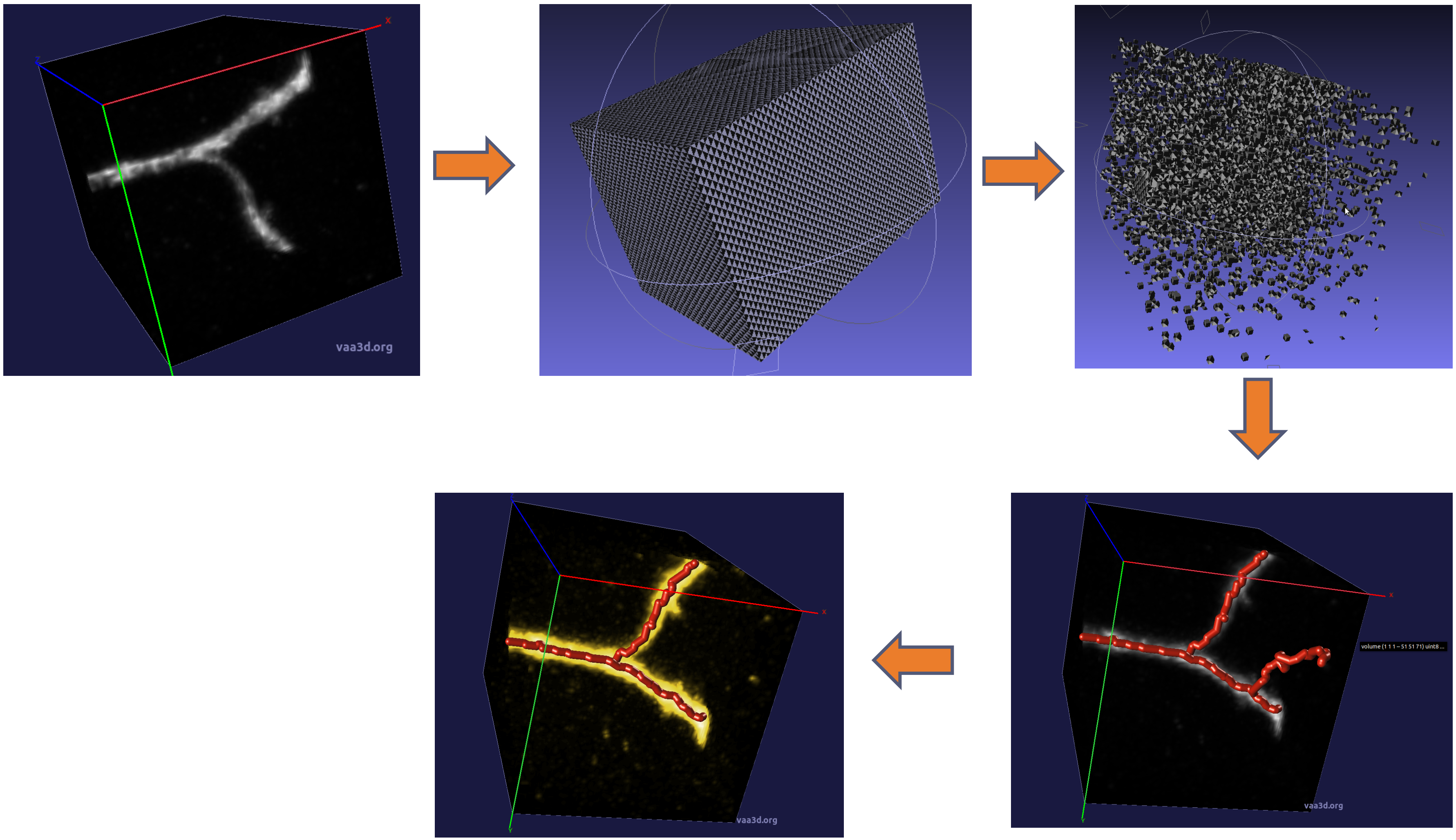
- Input: a piecewise-linear (PL) function defined on a simplicial complex domain
  - Given volumetric data (2D / 3D images), we can first triangulate it and convert it to a simplicial complex domain

- Leverage discrete Morse theory
  - *[Forman 1998, 2002]*
  - *[Gyulassy, 2008], [Sousbie 2011]* (*DisPerSE*)

Neuron Reconstruction Overview

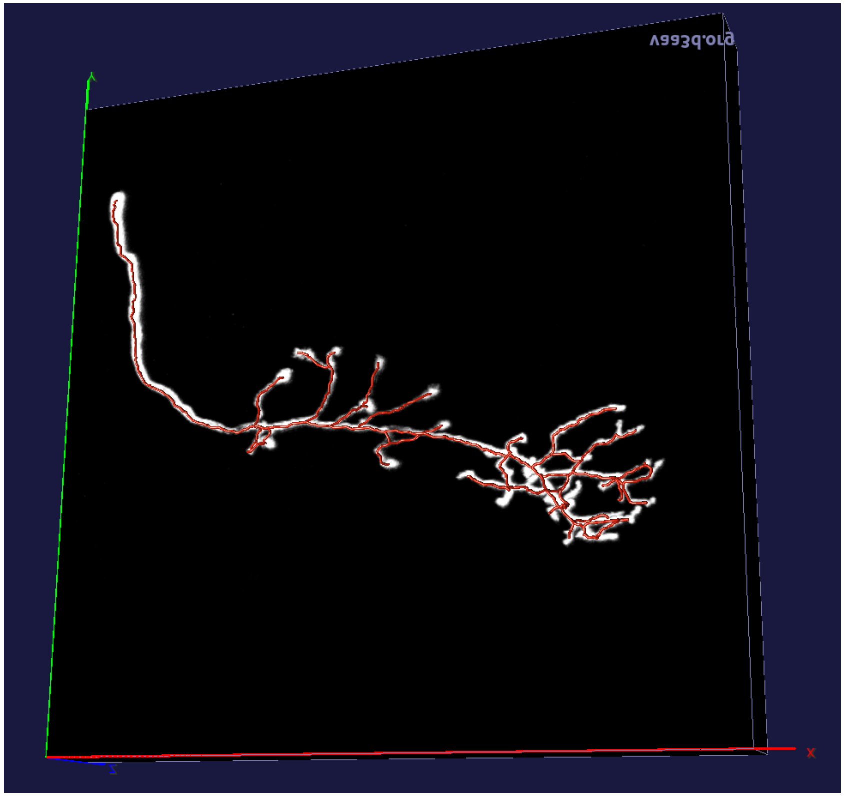
- **Input:** 2D/3D image $f: I \rightarrow R$ with f given at grid points in I
  - (1) Triangulate I to K, and potentially remove background cells to obtain PL function $f: K \rightarrow R$
  - (2) Negate f to $\hat{f} = -f$
  - (3) Compute 1-stable manifold for index-1 saddles
  - (4) Simplify to remove noise
  - (5) Output *Neuron-graph G*
  - (6) Obtain a tree structure T from G
    - Assign weights to arcs in G as integral of density f along the arc
    - Compute maximum spanning tree T

Neuron Reconstruction Overview

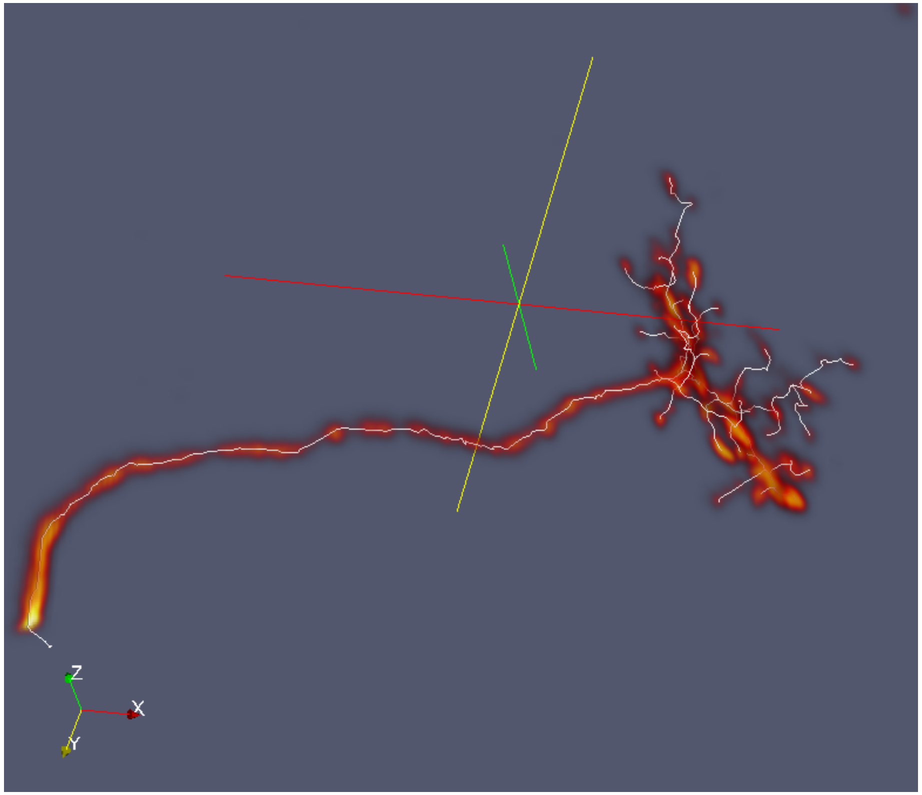


Preliminary Results

- Some DIADEM datasets



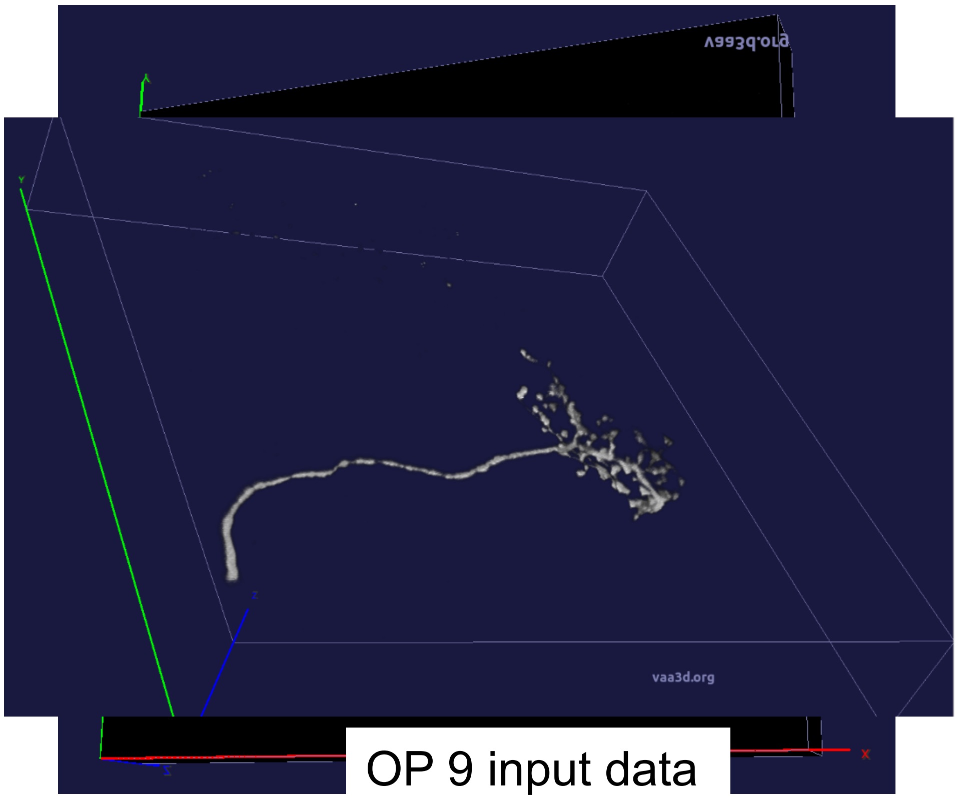
OP 1



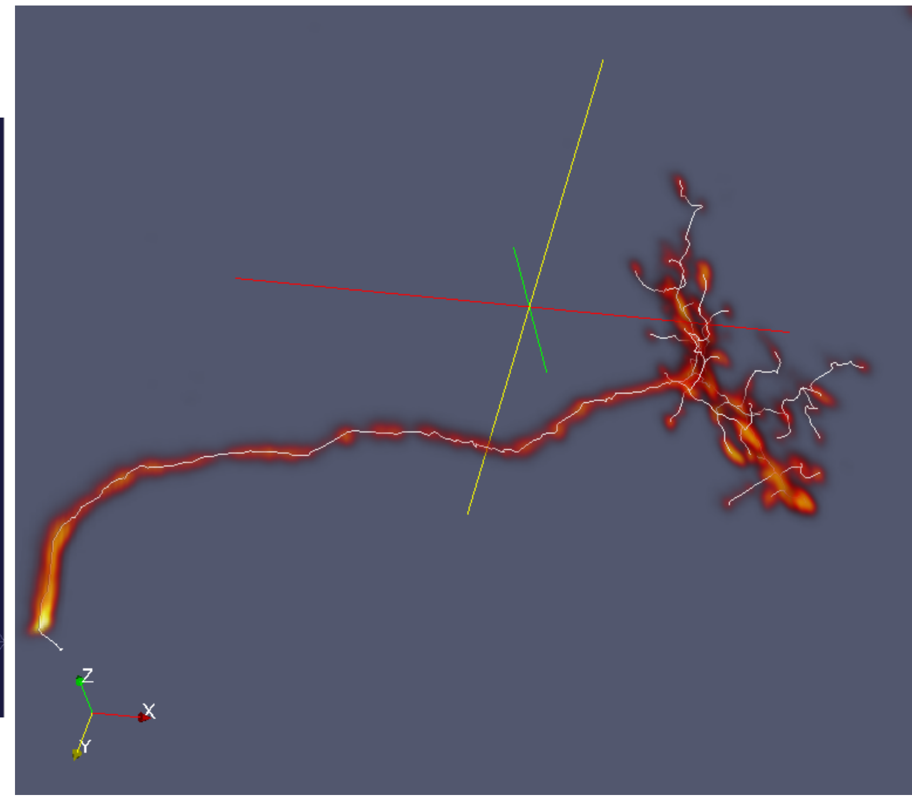
OP 9

Preliminary Results

- Some DIADEM datasets

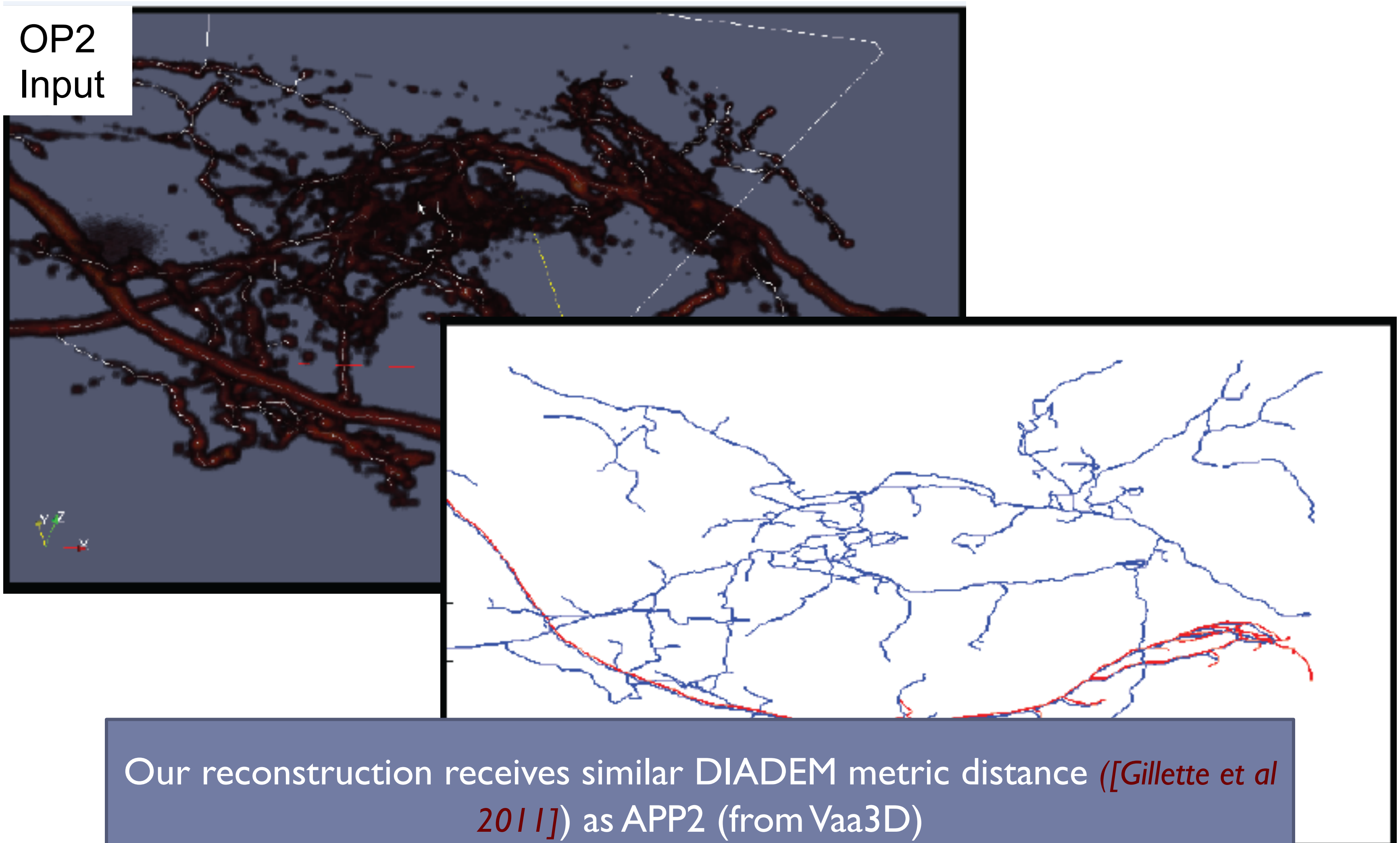


OP 9 input data



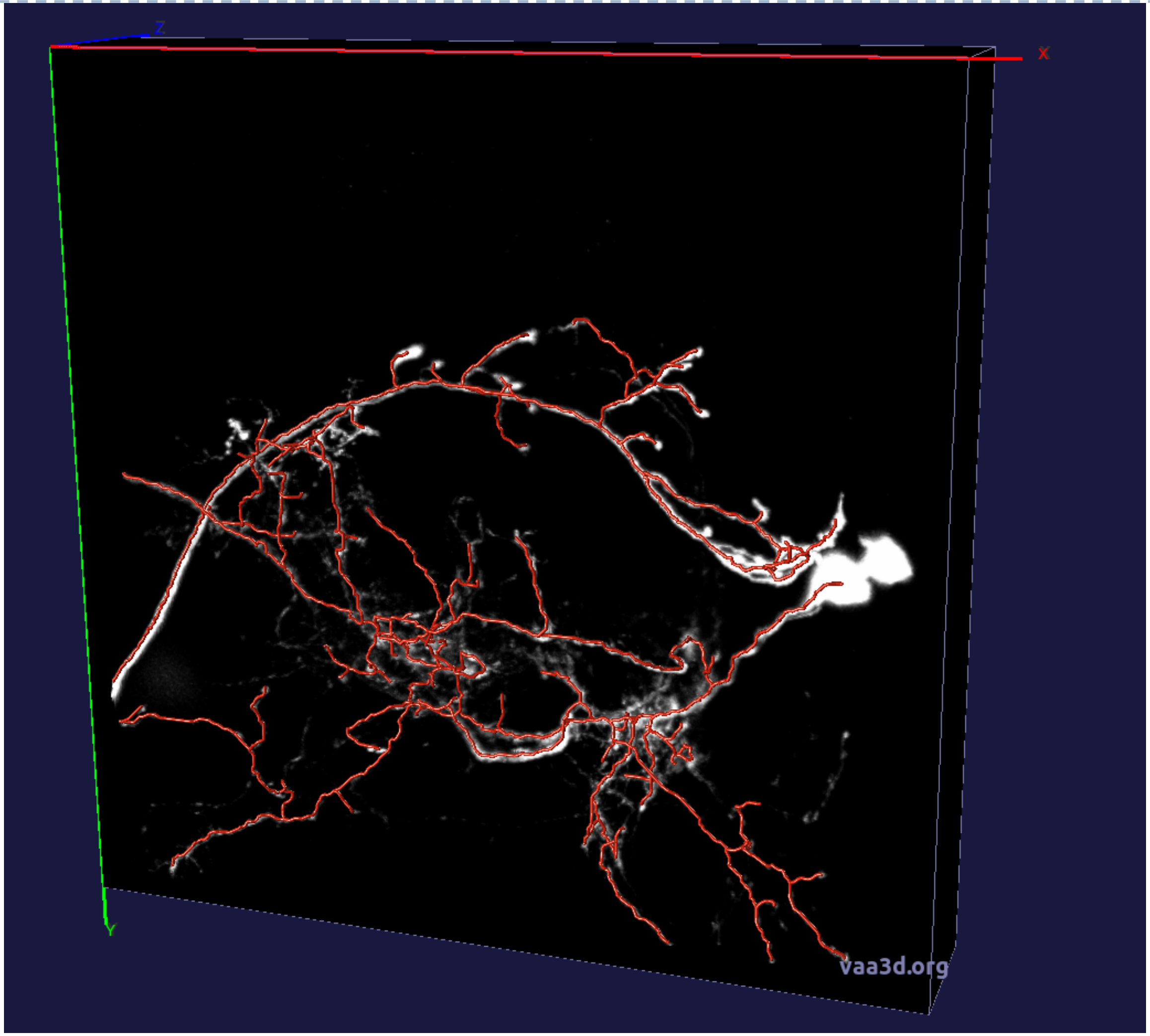
OP 9

Diadem Dataset OP2



Our reconstruction receives similar DIADEM metric distance *([Gillette et al 2011]*) as APP2 (from Vaa3D)

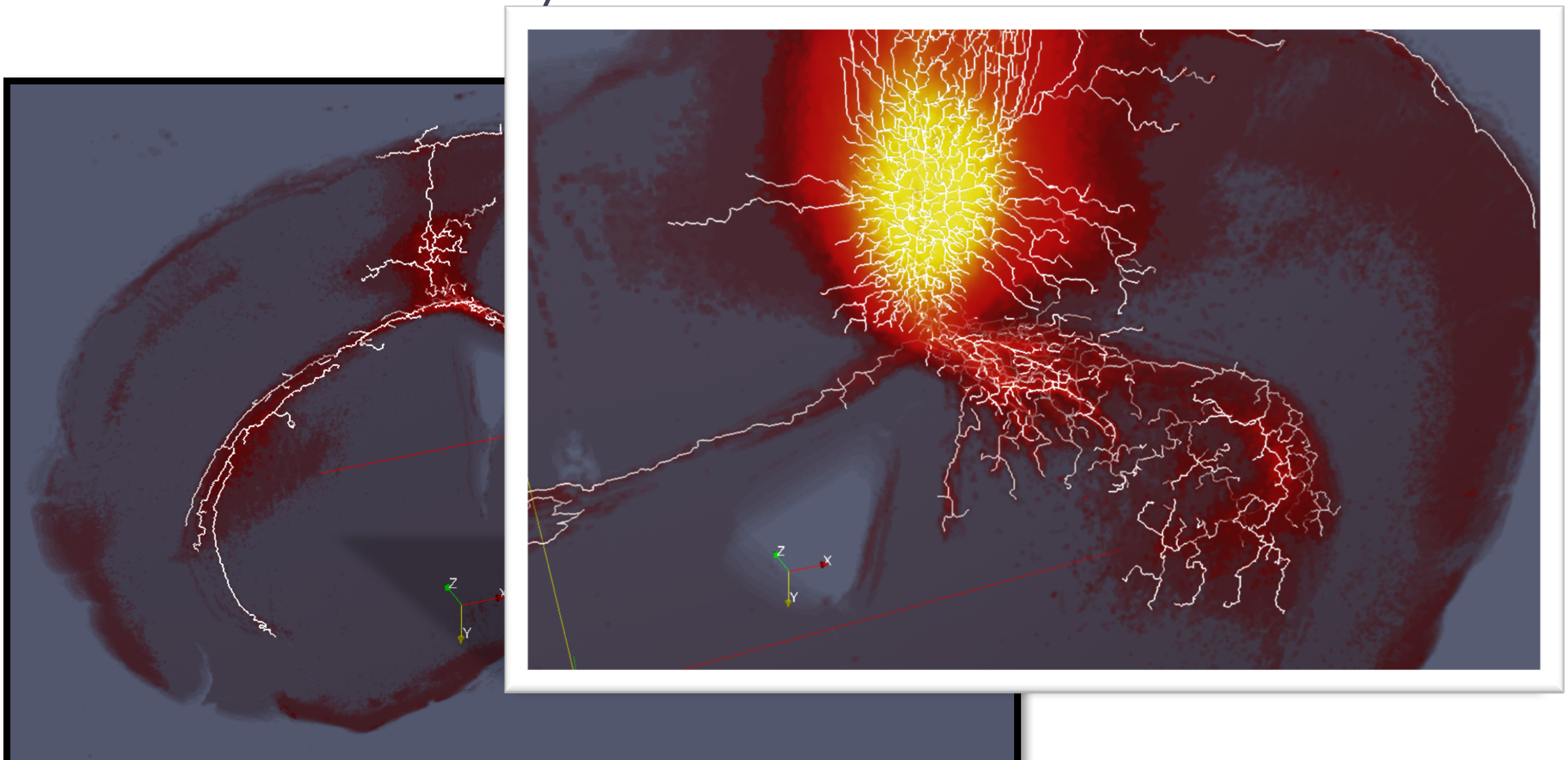
Diadem Dataset OP2



OP 2

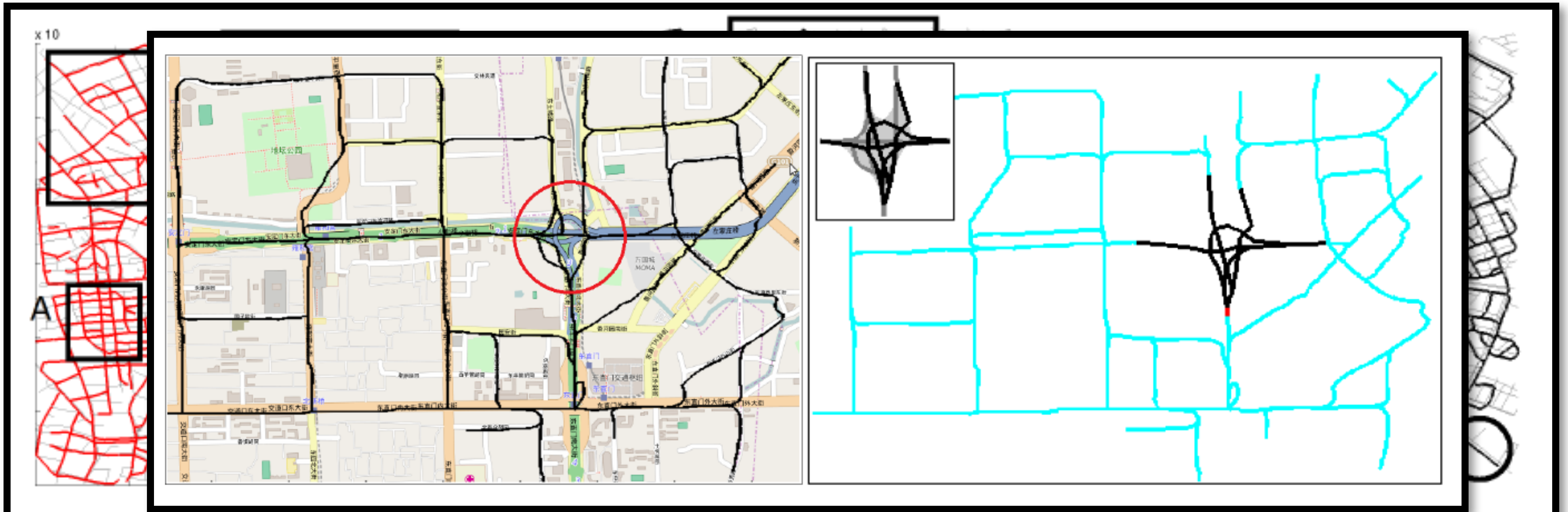
Preliminary Results

- Mouse brain LM images from an AAV viral tracer-injection
  - from Mitra laboratory at CSHL



Remarks

- Other advantages of Morse-based framework
  - Can be used to merge/integrate multiple reconstructions
  - Can be used to provide correction ability



(a) Karagiorgou 2013 (b) Our reconstruction (c) Our integration

Summary and Remarks

- Two examples of topological methods for neuronal structure analysis
  - Topological methods
    - general and robust
    - capture / leverage global structures
    - tend to be less ad-hoc
- Further develop these applications
- Provide more theoretical justification and understanding