Tracking Neural Activity

Automated Image Analysis

B. Jones[1] A. Stanley[2]

[1]Department of Mathematics
State University of New York at Geneseo

[2]Department of Mathematics
Grand Valley State University

January 26, 2019

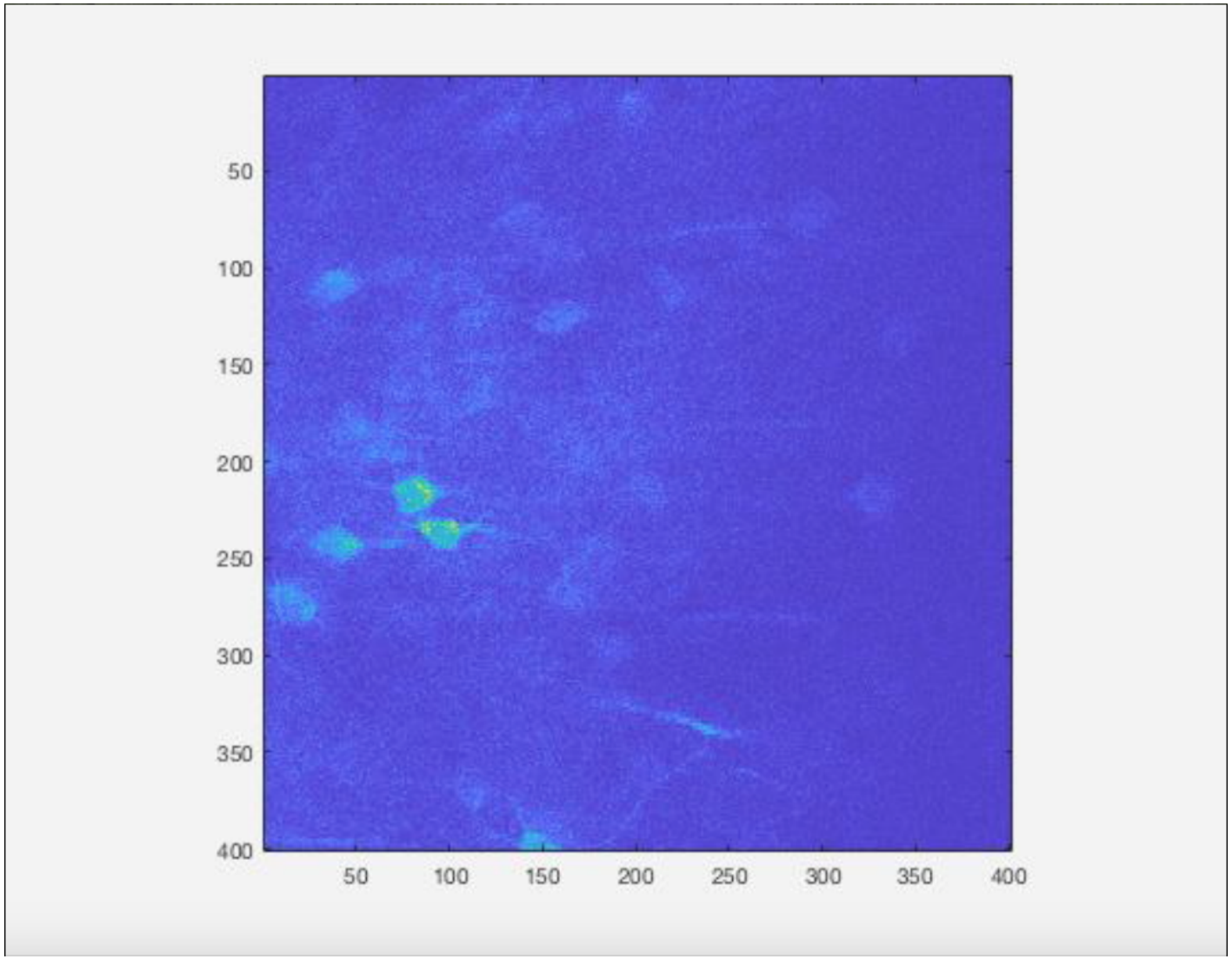
Outline

Outline

Understanding the Brain

- We wish to understand how brain activity correlates with specific actions
- Current leading algorithms to detect cells need improvement
  - Based off of generic criteria (size and roundness) - not ideal
  - Not all cells are perfectly round
  - Not all cells are perfectly visible in the brain images
  - There is also background noise, or **neuropil**
  - Only detect $\sim$50% of cells, leading scientists to manually pick out cells

The Zaneta Mouse Data Set: An Example



Outline

This is Your Brain on Math

We represent a data set (a video of the brain) as a 3-dimensional matrix $\mathbf{Y} \in \mathbb{R}^{N \times D_y \times D_x}$ where

$$N = \text{number of frames}$$
$$D_y = \text{number of pixels along y}$$
$$D_x = \text{number of pixels along x.}$$

This is Your Brain on Math

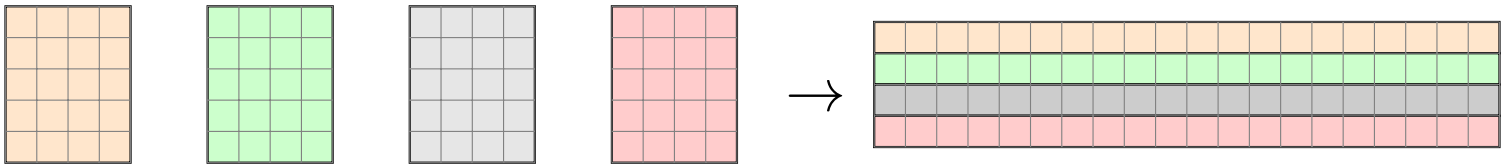
We represent a data set (a video of the brain) as a 3-dimensional matrix $\mathbf{Y} \in \mathbb{R}^{N \times D_y \times D_x}$ where

$$N = \text{number of frames}$$
$$D_y = \text{number of pixels along y}$$
$$D_x = \text{number of pixels along x.}$$

Note that $\mathbf{Y}$ is very large.

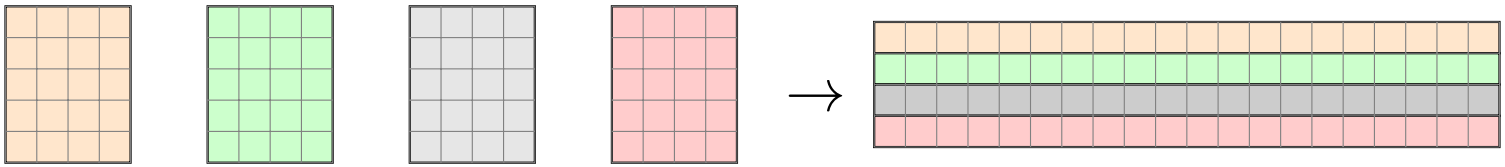
In order to make these data sets more manageable, we convert to a rectangular array with dimensions $N \times D$ where $D = D_x \times D_y$

From 3-Dimensional to 2-Dimensional



We reshape a $4 \times 5 \times 4$ matrix into single a 4×20 matrix

From 3-Dimensional to 2-Dimensional



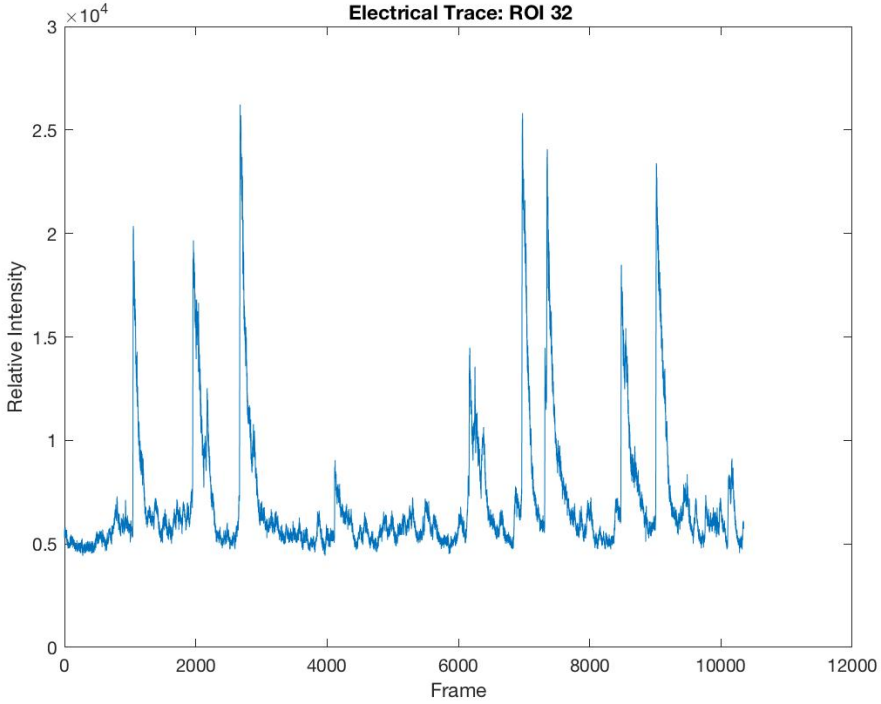
We reshape a $4 \times 5 \times 4$ matrix into single a 4×20 matrix

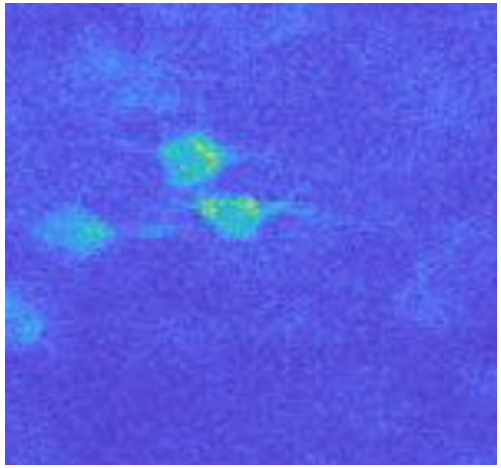
Question:

How do we work with our high-dimensional $N \times D$ array and find our sparse set of **sources** (neurons)?

Dimension Reduction

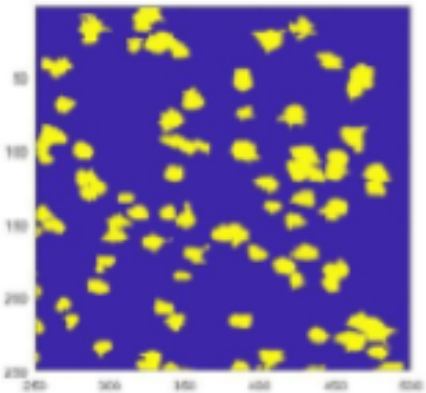
Each source has a signal and a footprint. A neuron's **signal** is a measurement over time (activity). A neuron's **footprint** is its spacial measurement (location, size, and shape).



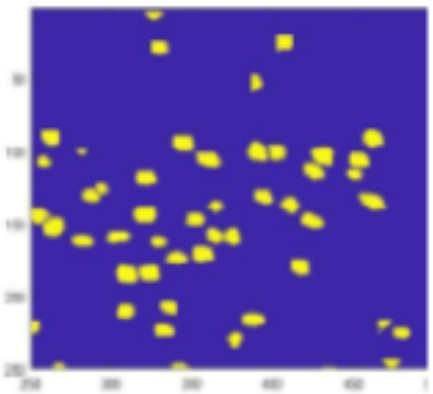


We aim to find these true sources (neurons) in our data and eliminate background noise caused by neuropil.

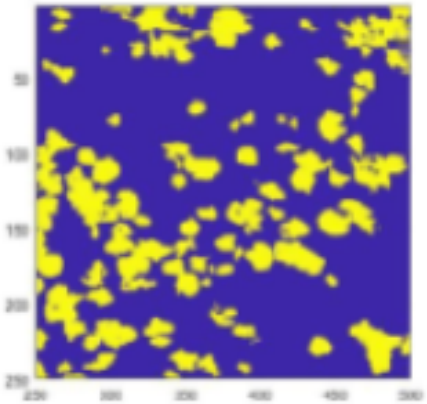
Leading Algorithms



Suite2P



"Ground truth"

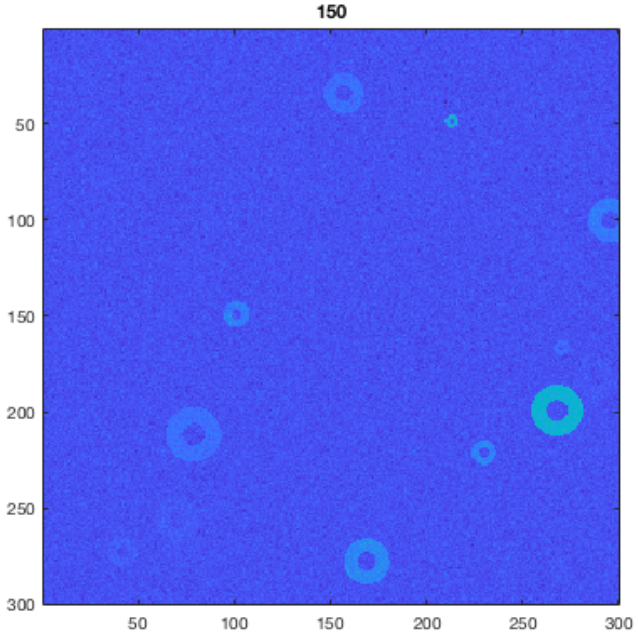


CaImAn

To identify potential sources, or **"regions of interest" (ROIs)**, Suite2P uses singular value decomposition (SVD) and CaImAn uses non-negative matrix factorization (NNMF).

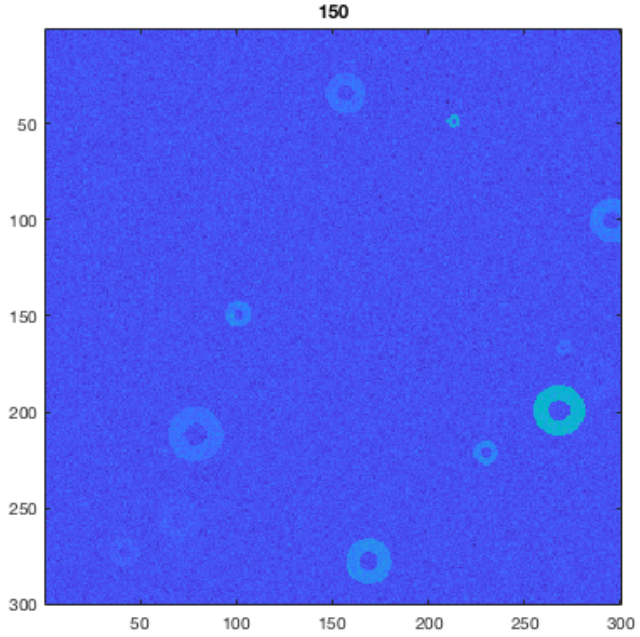
Synthetic Neurons

To understand how these programs use SVD, we randomly-generated a mini video $\mathbf{A} \in \mathbb{R}^{151 \times 300 \times 300}$ of synthetic neurons including



Synthetic Neurons

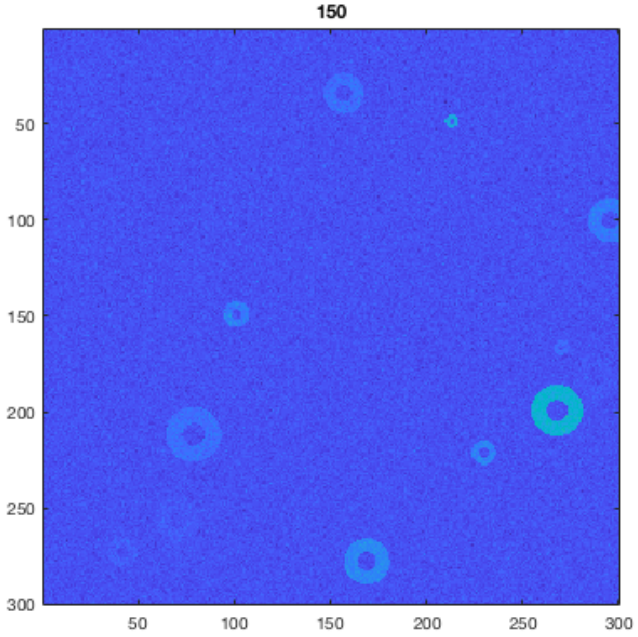
To understand how these programs use SVD, we randomly-generated a mini video $\mathbf{A} \in \mathbb{R}^{151 \times 300 \times 300}$ of synthetic neurons including



- Individual signals
- Individual footprints
- Background noise

Synthetic Neurons

To understand how these programs use SVD, we randomly-generated a mini video $\mathbf{A} \in \mathbb{R}^{151 \times 300 \times 300}$ of synthetic neurons including



- Individual signals
- Individual footprints
- Background noise

Note that we generated 15 neurons in $\mathbf{A}$.

Singular Value Decomposition

Let $\mathbf{A}$ be an $m \times n$ matrix.

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}$$

where

- $\mathbf{U}$ is an $m \times m$ orthogonal matrix
- $\boldsymbol{\Sigma}$ is an $m \times n$ diagonal matrix of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of $\mathbf{A}$ where $r = \min(m, n)$
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix

Singular Value Decomposition

Let $\mathbf{A}$ be an $m \times n$ matrix.

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}$$

where

- $\mathbf{U}$ is an $m \times m$ orthogonal matrix
- $\boldsymbol{\Sigma}$ is an $m \times n$ diagonal matrix of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of $\mathbf{A}$ where $r = \min(m, n)$
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}_{4\times 3} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}$$

Singular Value Decomposition

Let $\mathbf{A}$ be an $m \times n$ matrix.

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}$$

where

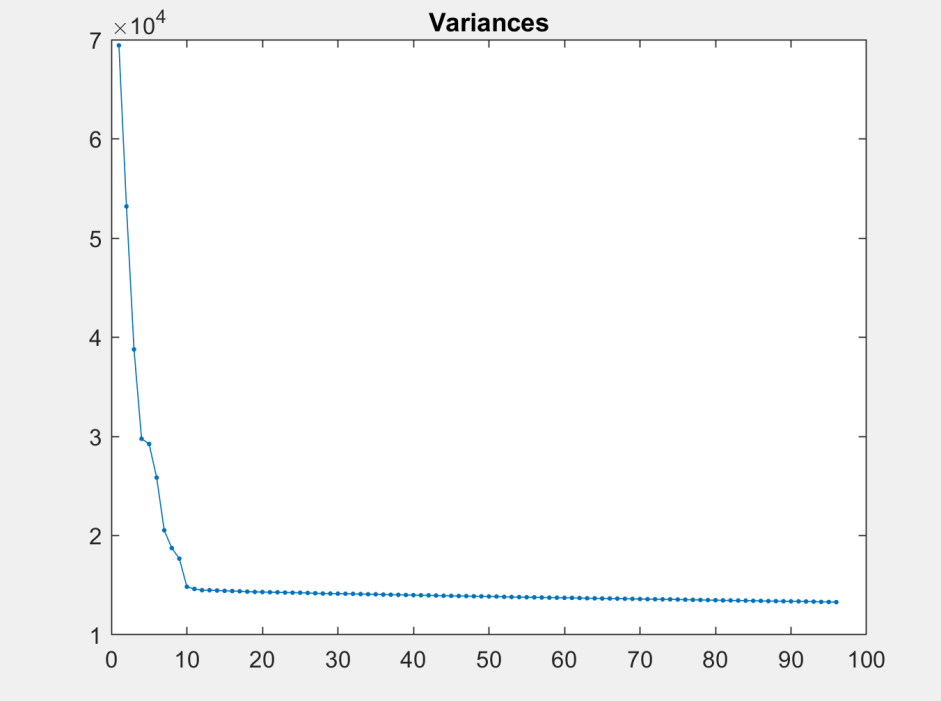
- $\mathbf{U}$ is an $m \times m$ orthogonal matrix
- $\boldsymbol{\Sigma}$ is an $m \times n$ diagonal matrix of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of $\mathbf{A}$ where $r = \min(m, n)$
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix

Let $\mathbf{u_i}$ and $\mathbf{v_i}$ be the column vectors of $\mathbf{U}$ and $\mathbf{V}$ respectively. Then

$$\mathbf{A} = \sum_{i}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

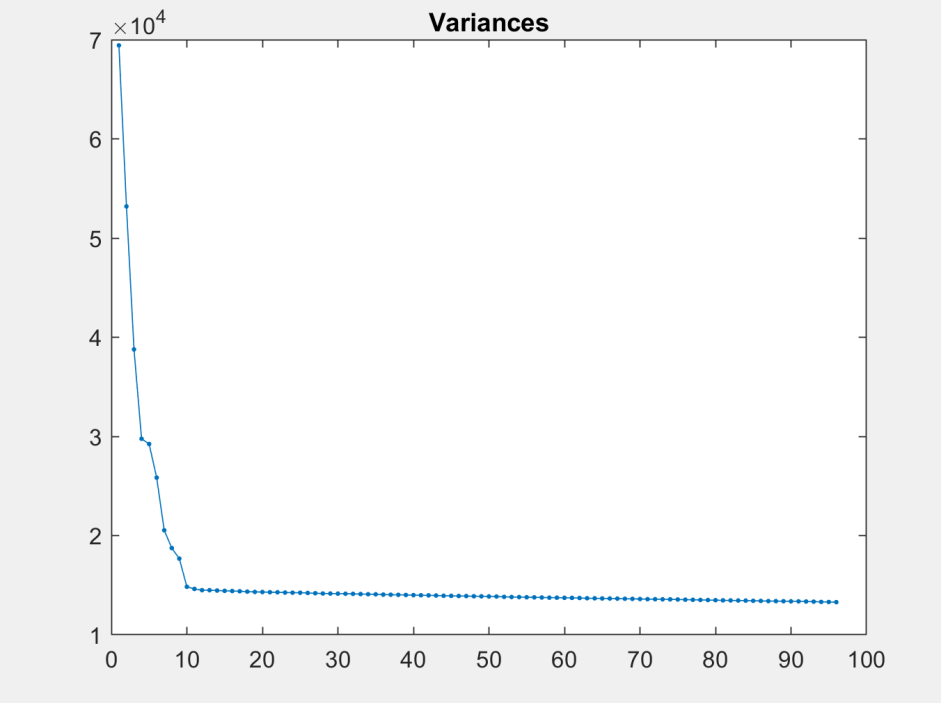
Synthetic Neurons

The largest singular values in Σ will contribute most to $\mathbf{A}$. Since our goal is to find neurons, the factors that contribute most to our data set, these values become very important.



Synthetic Neurons

The largest singular values in Σ will contribute most to $\mathbf{A}$. Since our goal is to find neurons, the factors that contribute most to our data set, these values become very important.



After about the fifth singular value, the values become insignificant.

Synthetic Neurons

Let $k = 5 \leq r$.

Synthetic Neurons

Let $k = 5 \leq r$. Recall

$$\begin{aligned}\mathbf{A} &= \sum_{i}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \\ &= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \ldots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T \\ &= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \ldots + \sigma_k \mathbf{u_k} \mathbf{v_k}^T + \ldots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T\end{aligned}$$

Synthetic Neurons

Let $k = 5 \leq r$. Recall

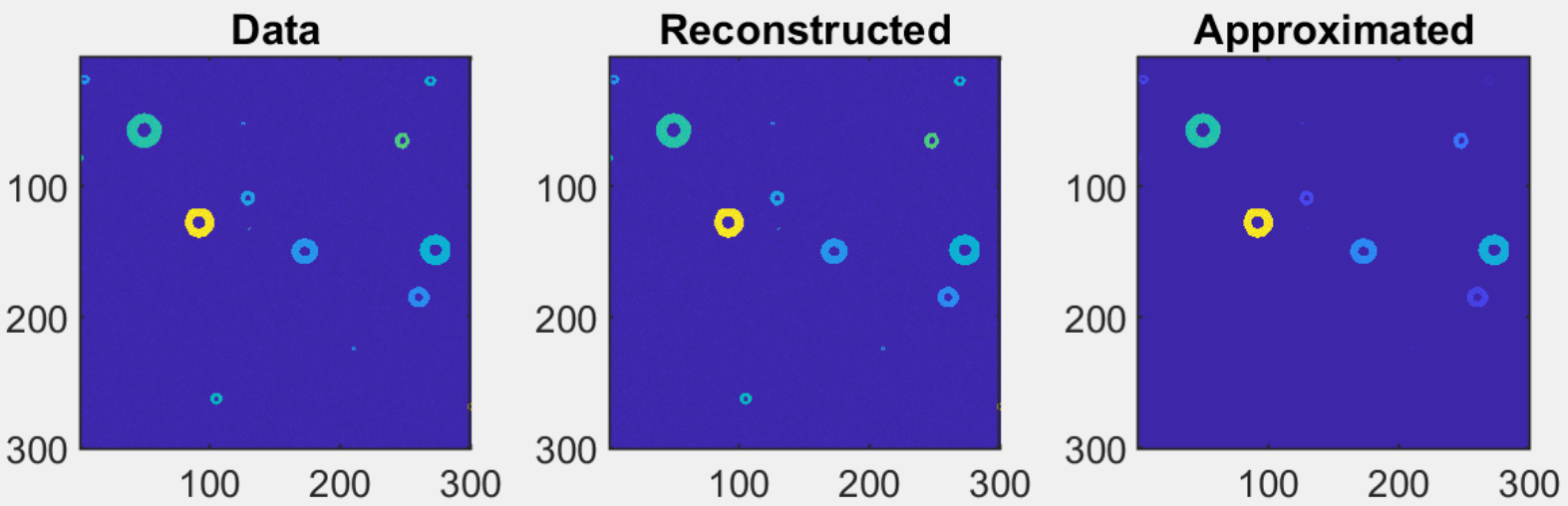
$$\begin{aligned}\mathbf{A} &= \sum_{i}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \\ &= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \ldots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T \\ &= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \ldots + \sigma_k \mathbf{u_k} \mathbf{v_k}^T + \ldots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T\end{aligned}$$

We can approximate $\mathbf{A}$ using k:

$$\begin{aligned}\tilde{\mathbf{A}} &= \sum_{i}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \\ &= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \sigma_2 \mathbf{u_2} \mathbf{v_2}^T + \ldots + \sigma_k \mathbf{u_k} \mathbf{v_k}^T\end{aligned}$$

Synthetic Neurons

Using our approximation **A**, we potentially have a clearer view of the neurons in our video, eliminating the background noise:



Outline

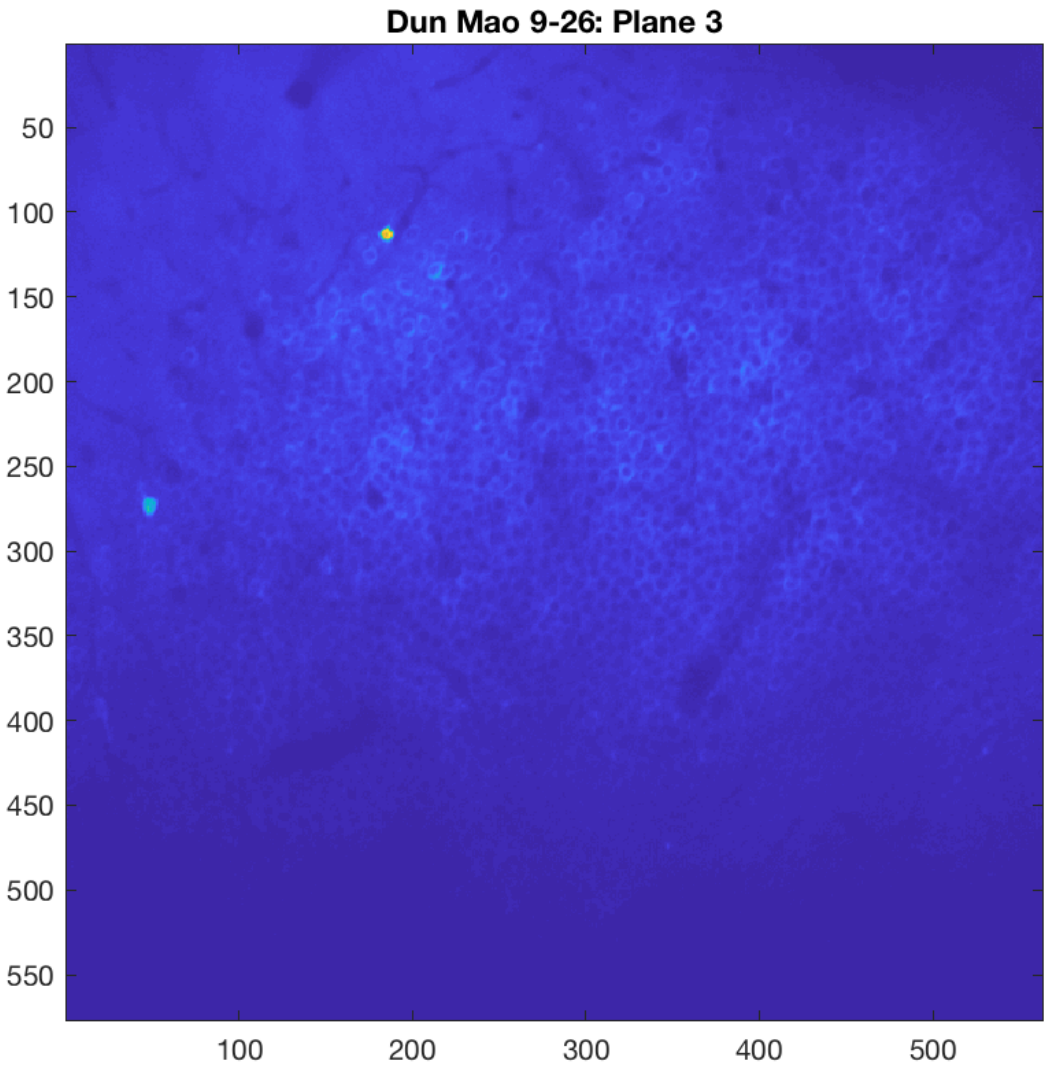
Dun Mao Data

The goal of the lab at Lethbridge is to study how memory and learning affects the brain.

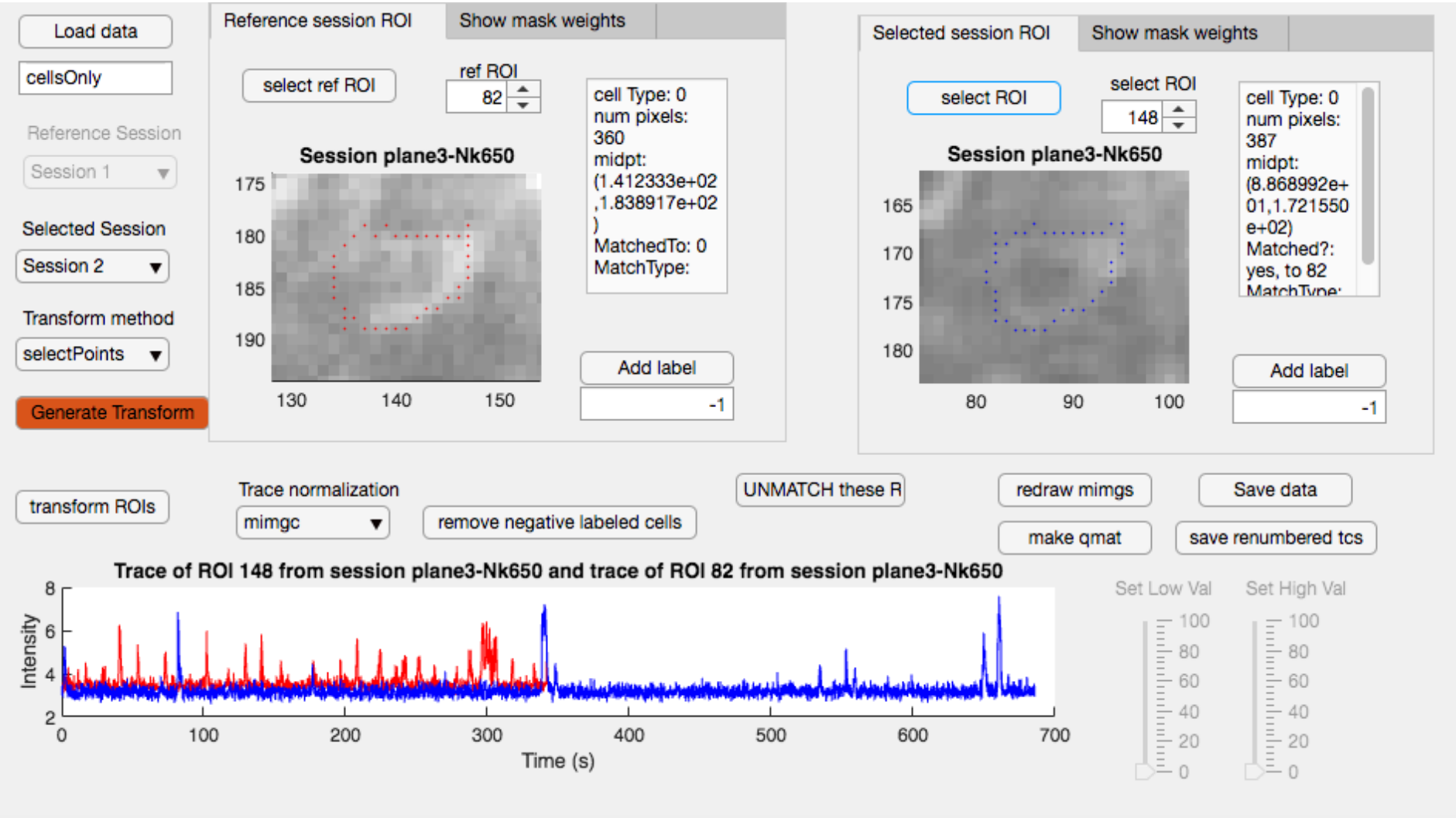
Matching cells between the sessions will allow us to see, as the mouse repeats actions, how neural firing activity changes:

- Location
  - Which area of the brain is active
  - Which/how many cells in that area are active
- Frequency
- Coordinated firing

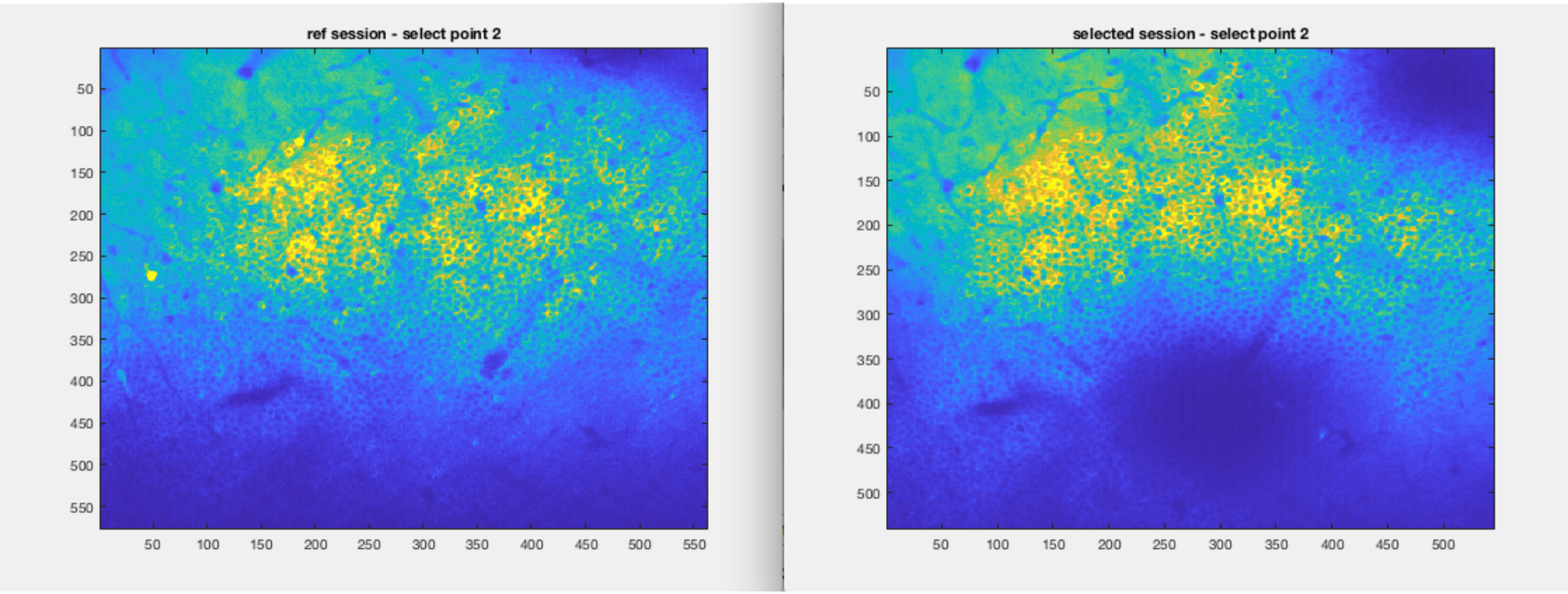
Mean Image


Dun Mao 9-26: Plane 3
50
100
150
200
250
300
350
400
450
500
550
100
200
300
400
500

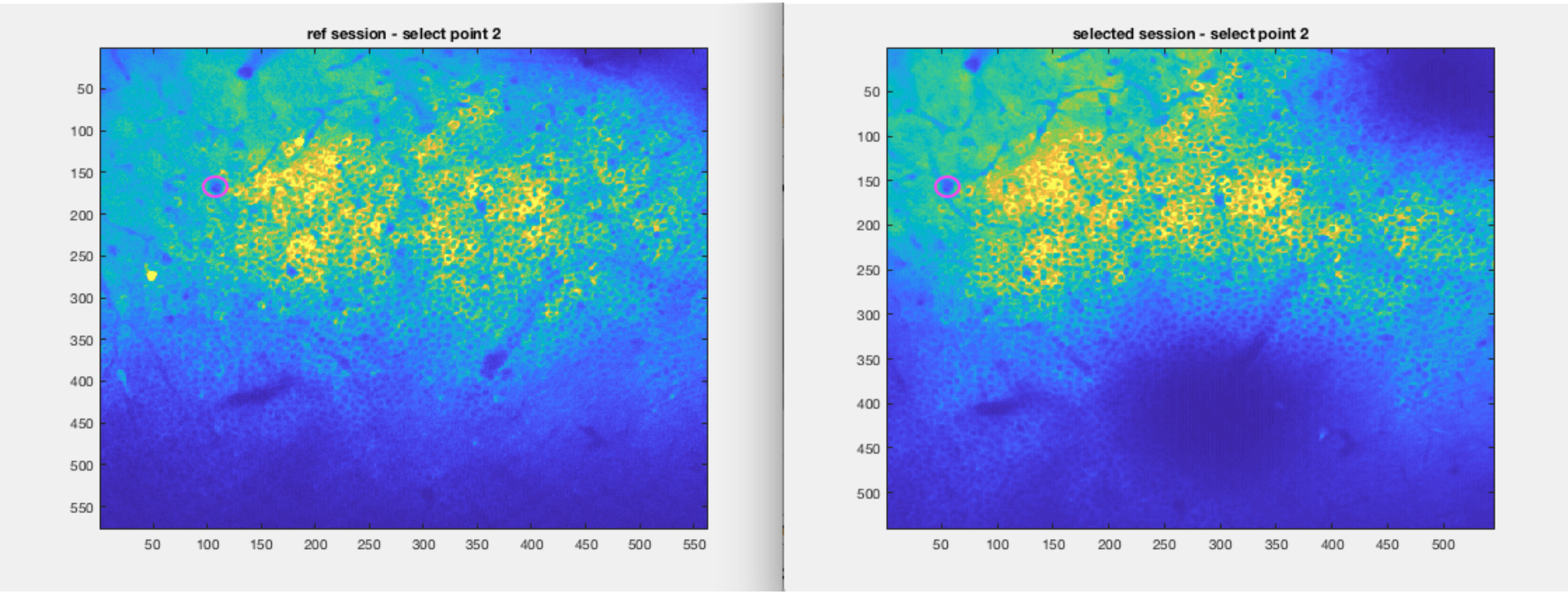
ROI Transformation



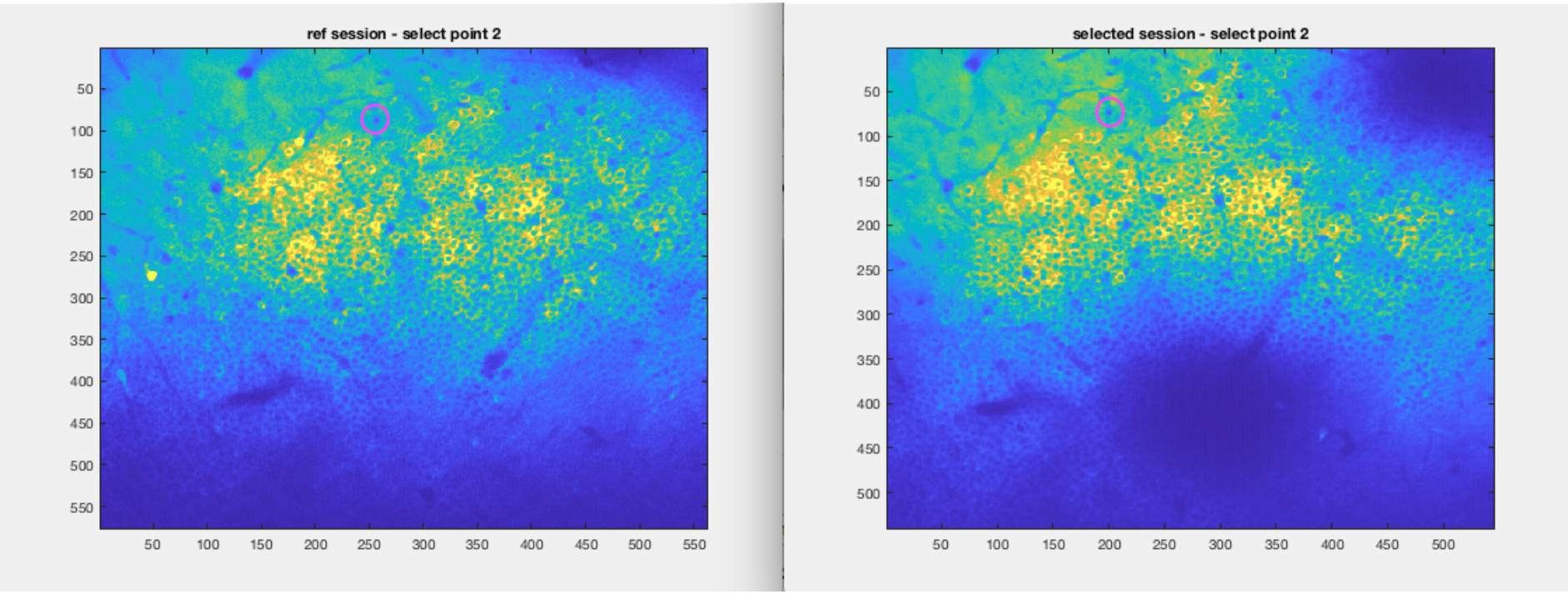
ROI Transformation



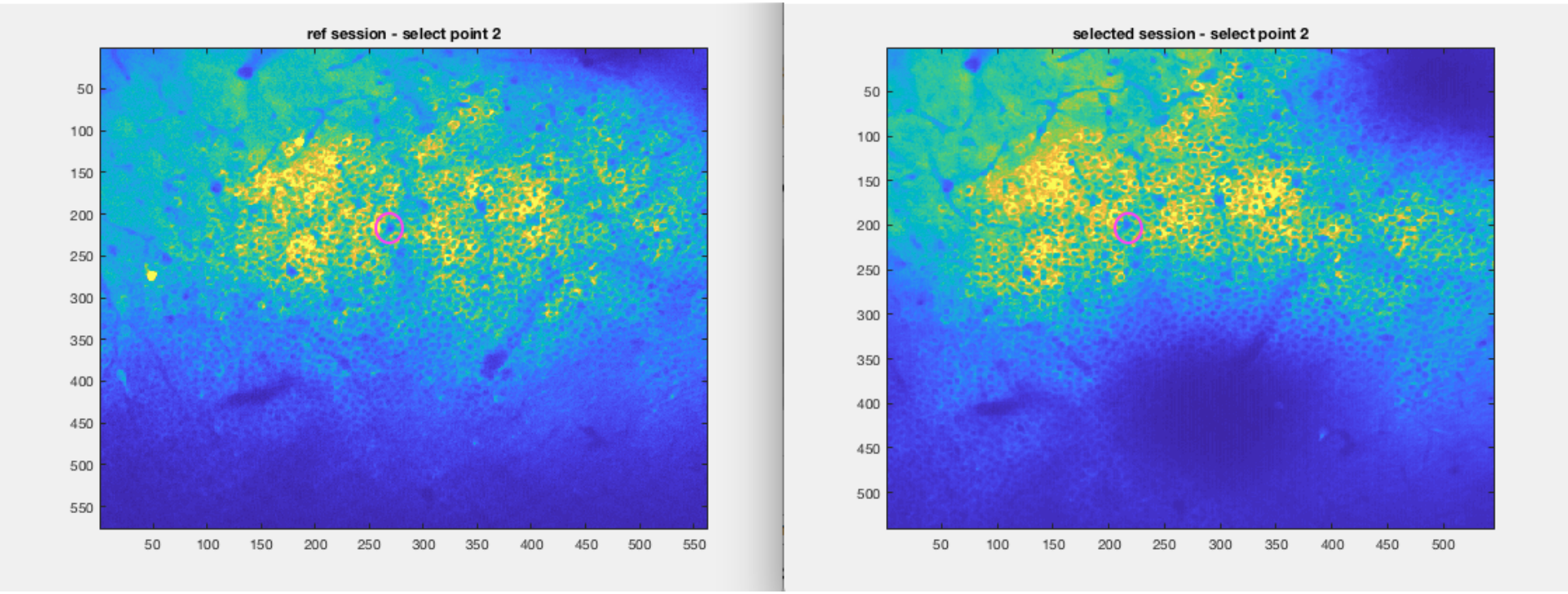
ROI Transformation



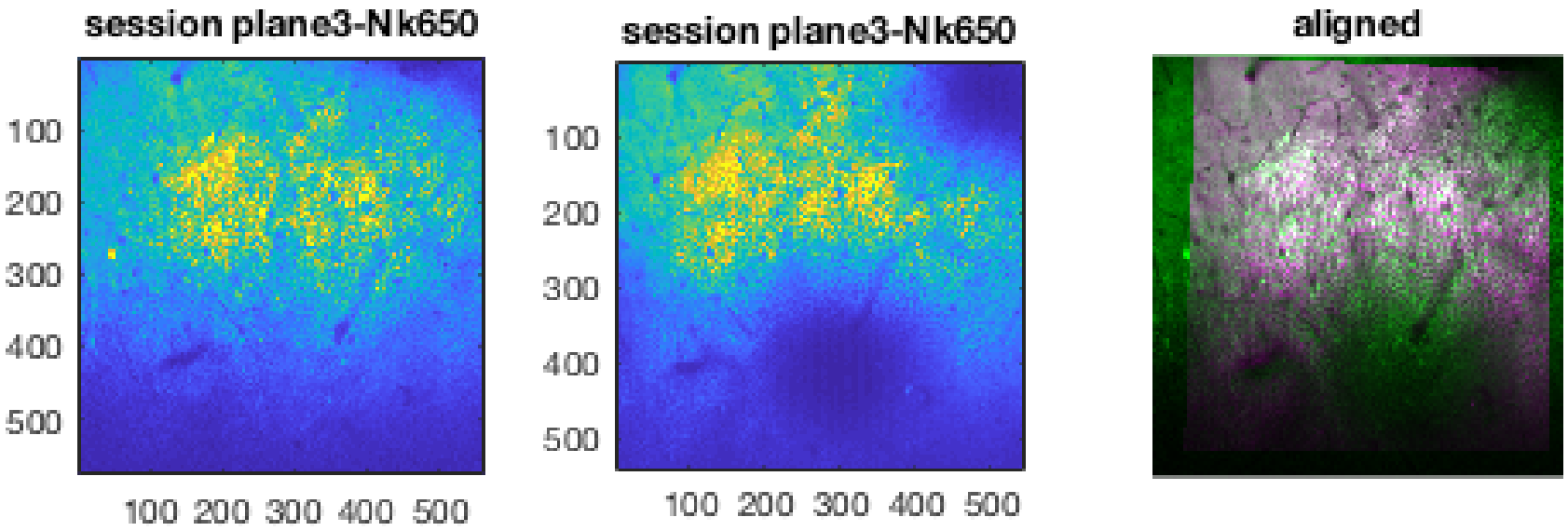
ROI Transformation



ROI Transformation

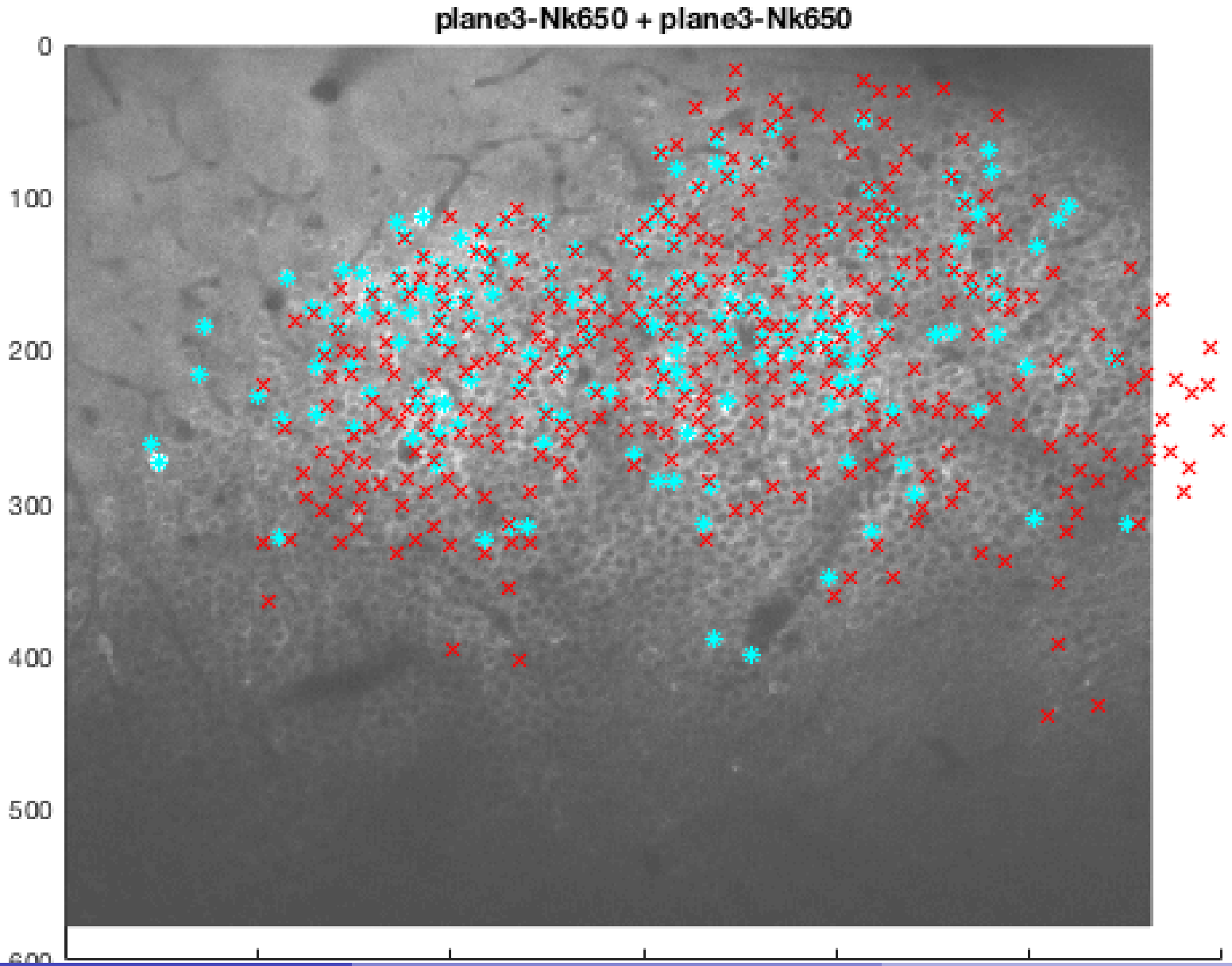


ROI Transformation

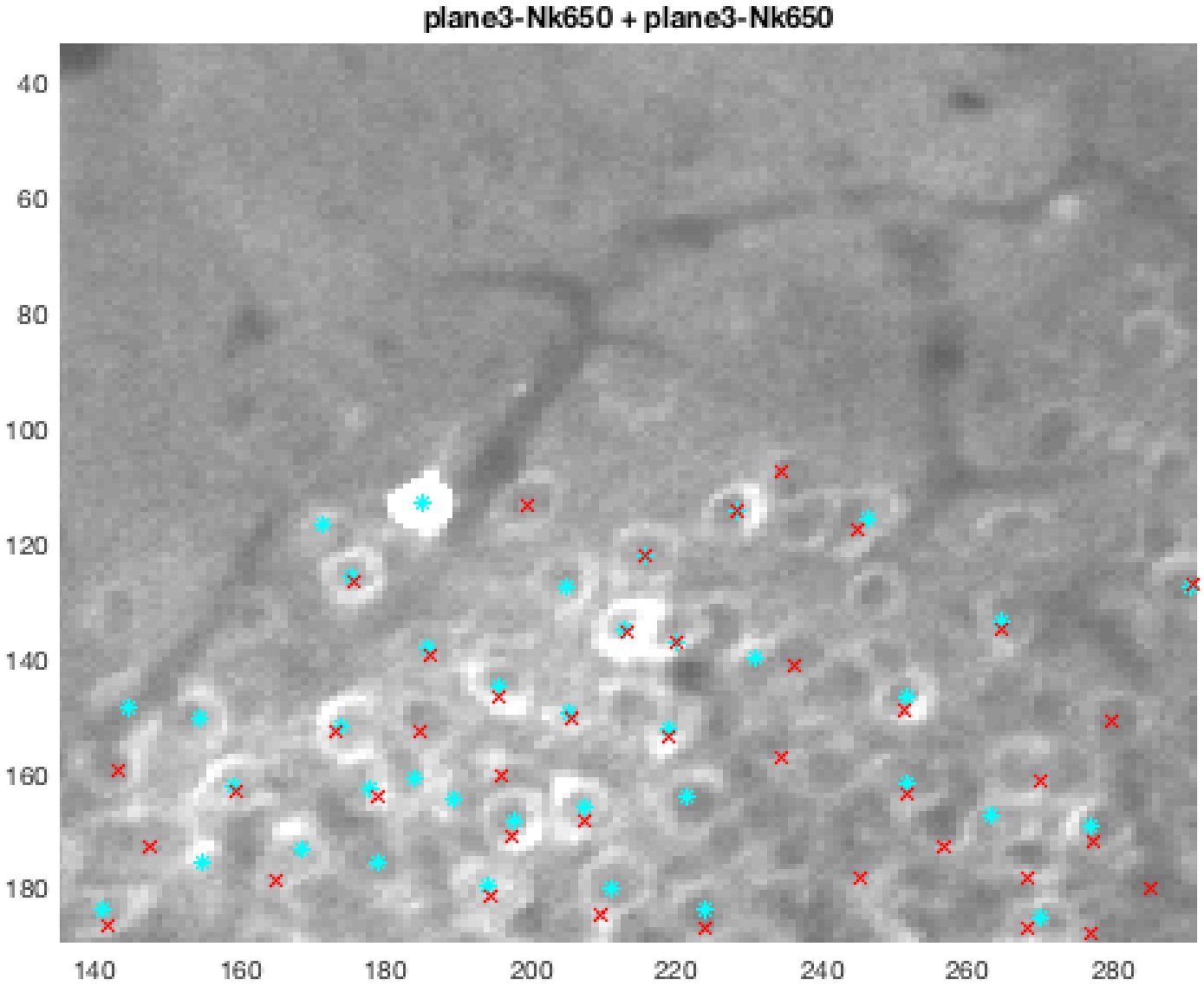


The two sessions are aligned (left: 9/26 at 170μm will become green/blue stars, middle: 10/1 at 180μm will become purple/red crosses)

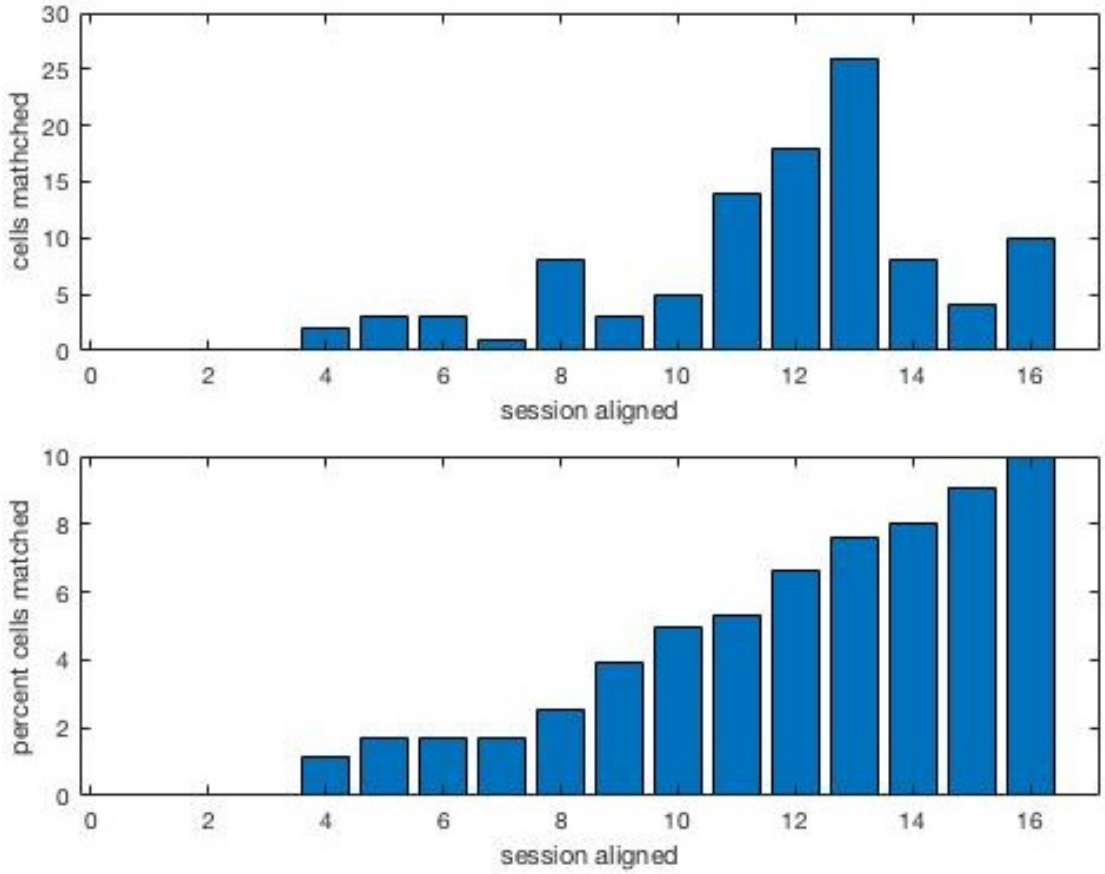
Matching Cells



Matching Cells



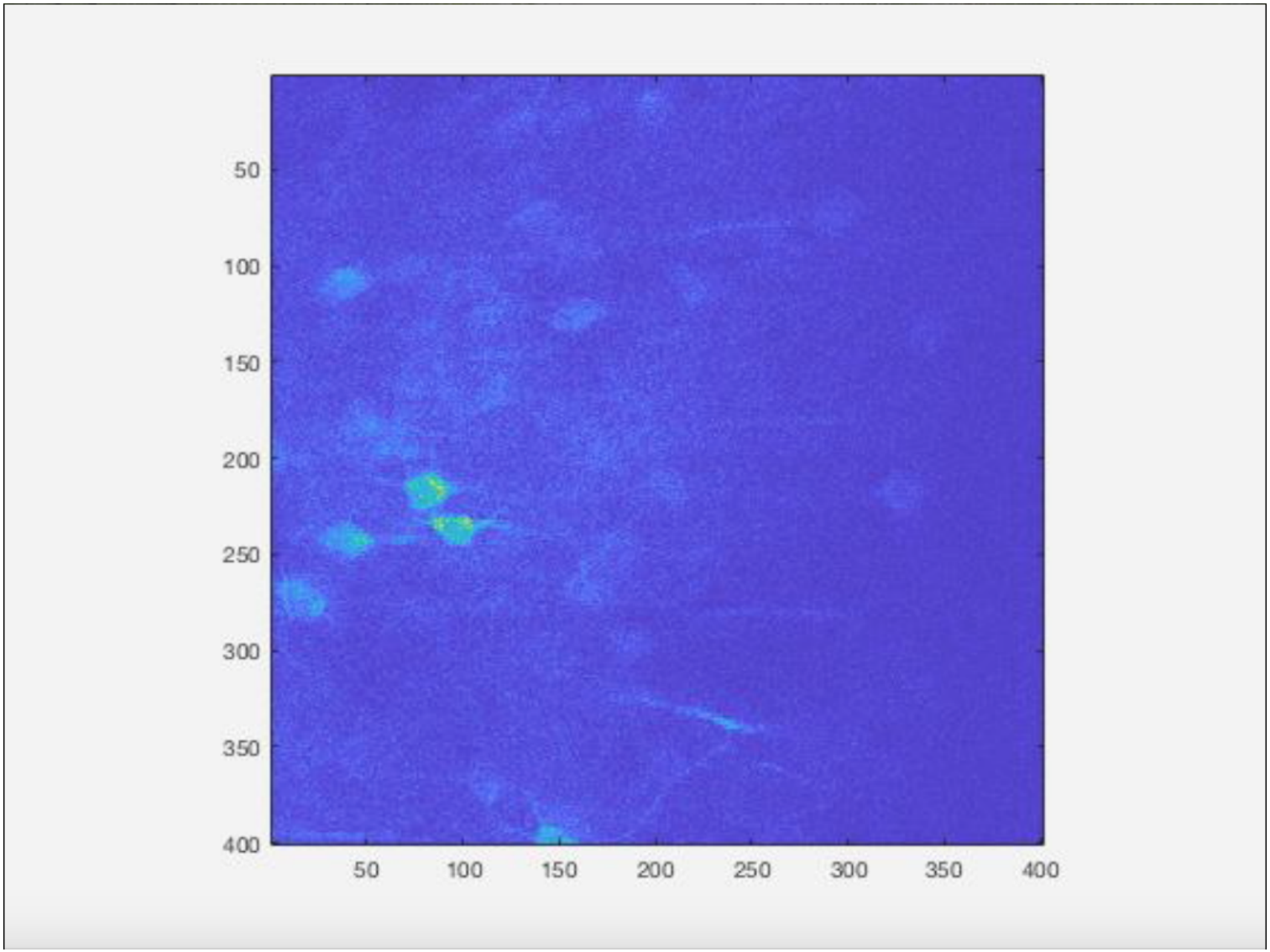
Matching Cells



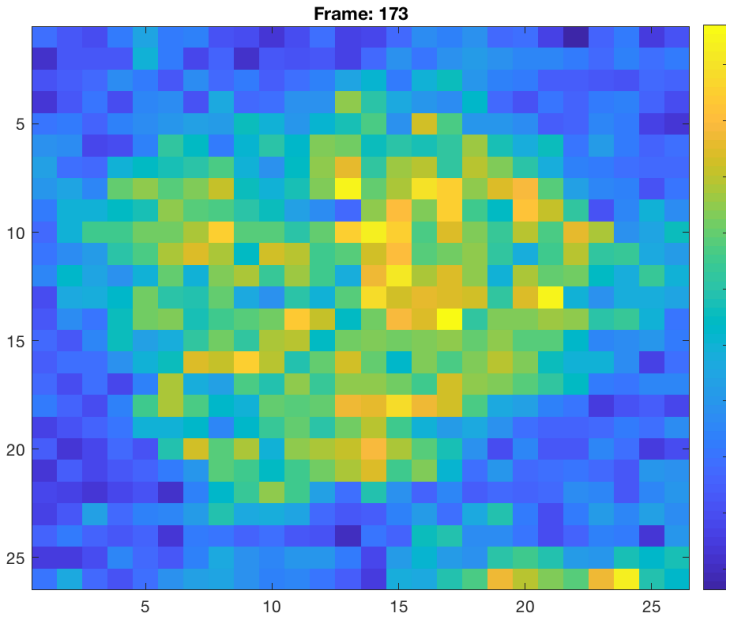
Here, we see that three of the sixteen aligned sessions did not match any cells. This was most likely due to the lack of visual geometric features.

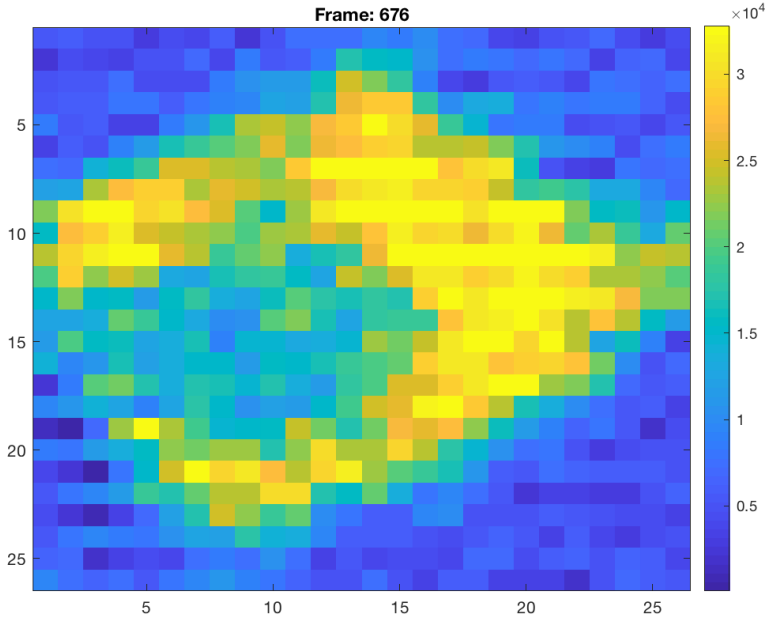
Outline

Cell Substructure



Cell Substructure





Principal Component Analysis

Definition

The **principal components** of a data set are a set of linearly uncorrelated variables that account for the variability in the data. The first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and the remaining principal components continue to decrease or stay the same in the amount of variability they account for.

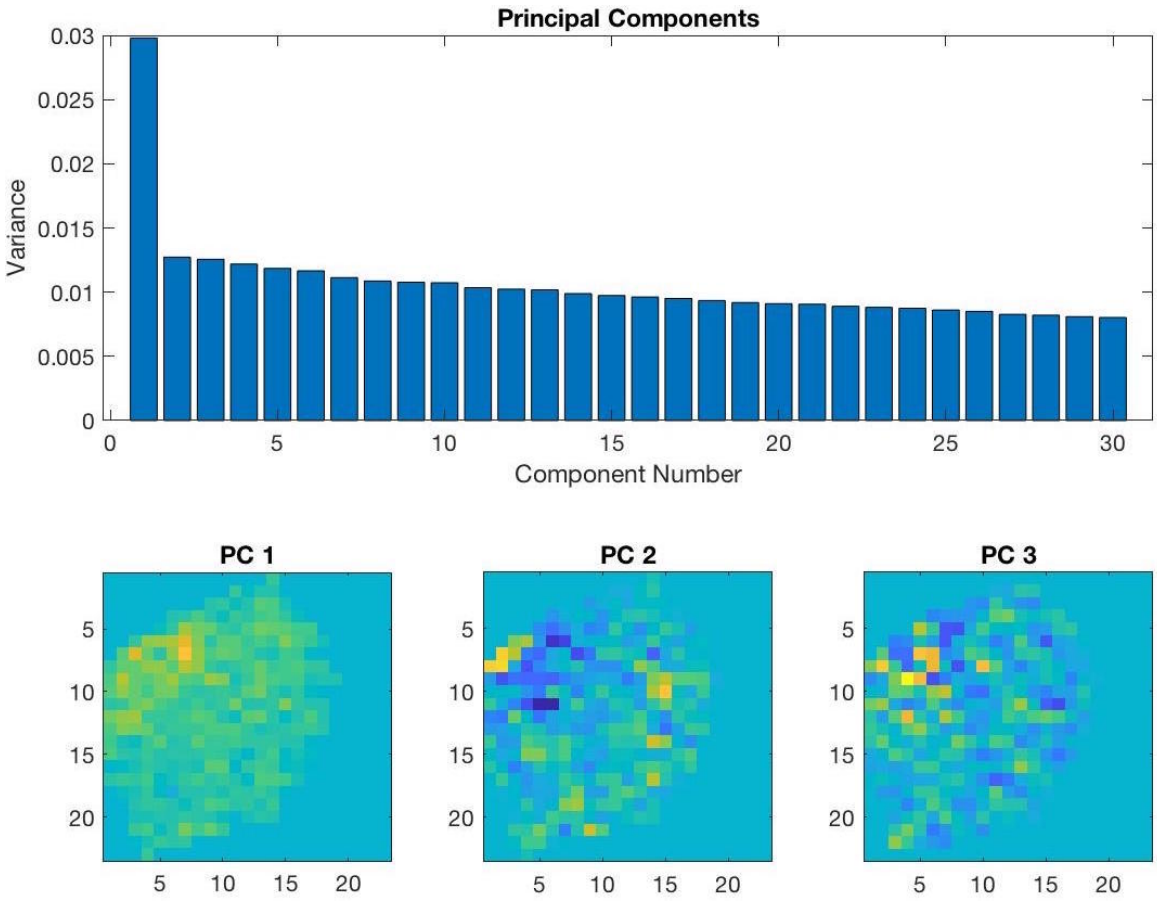
Principal Component Analysis

Definition

The **principal components** of a data set are a set of linearly uncorrelated variables that account for the variability in the data. The first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and the remaining principal components continue to decrease or stay the same in the amount of variability they account for.

Leading algorithms assume there is one principal component contributing to the majority of cell activity (the other components are background noise).

Principal Component Analysis



Summary

To analyze videos of mice brains, we used

- Singular Value Decomposition

in order to

- Match neurons from different recording sessions to help study memory and learning
- Establish that future algorithms should allow for multiple principal components in each cell.

Acknowledgements

Many thanks to the other students in our lab and our sponsors and mentors at Michigan State University, Drs. Mark Reimers, Michael Moore, and Samanthule Nola.

Thanks also to all the volunteers and sponsers at NCUWM for giving us the opportunity to talk!

And thank *you* for listening!

NSA Award No. H98230-18-1-0042

NSF Award No. 1559776

MSU SURIEM program

NCUWM 2019