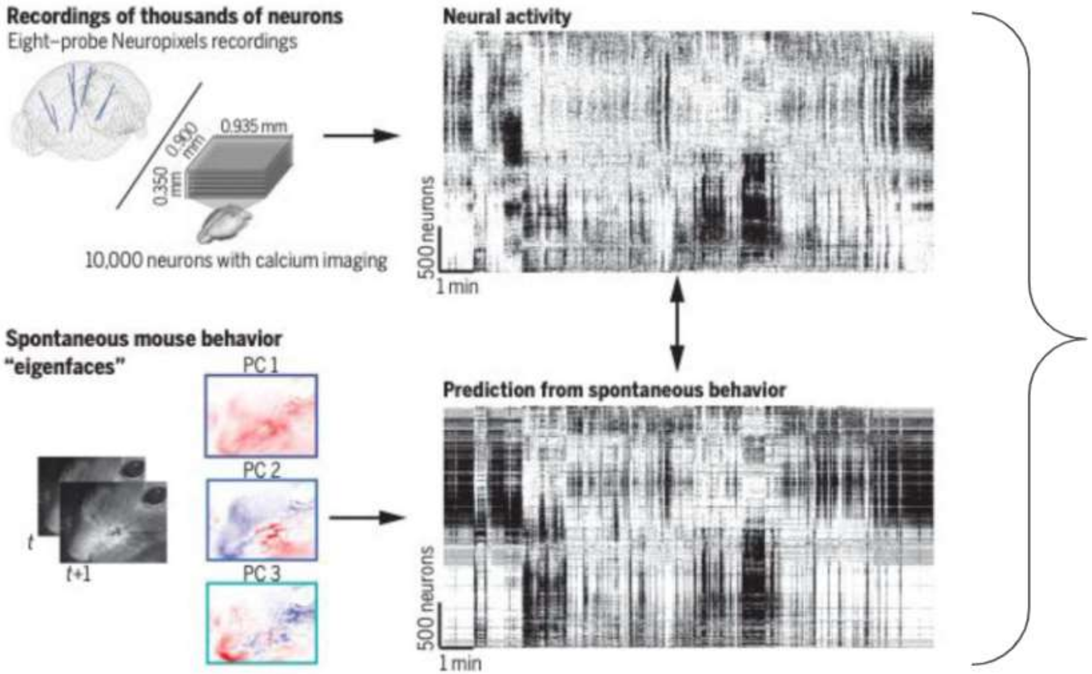
Exploring Cell Assemblies in Large Datasets

Pod: 112-charcoal-atika

Group: (overachieving-)theta-hats

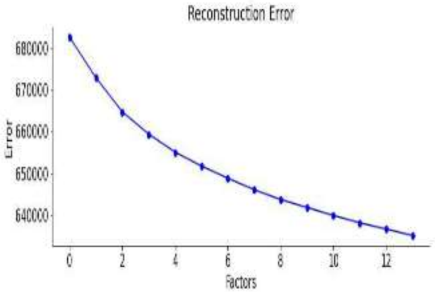
Dataset: Stringer 10,000 neurons, calcium imaging

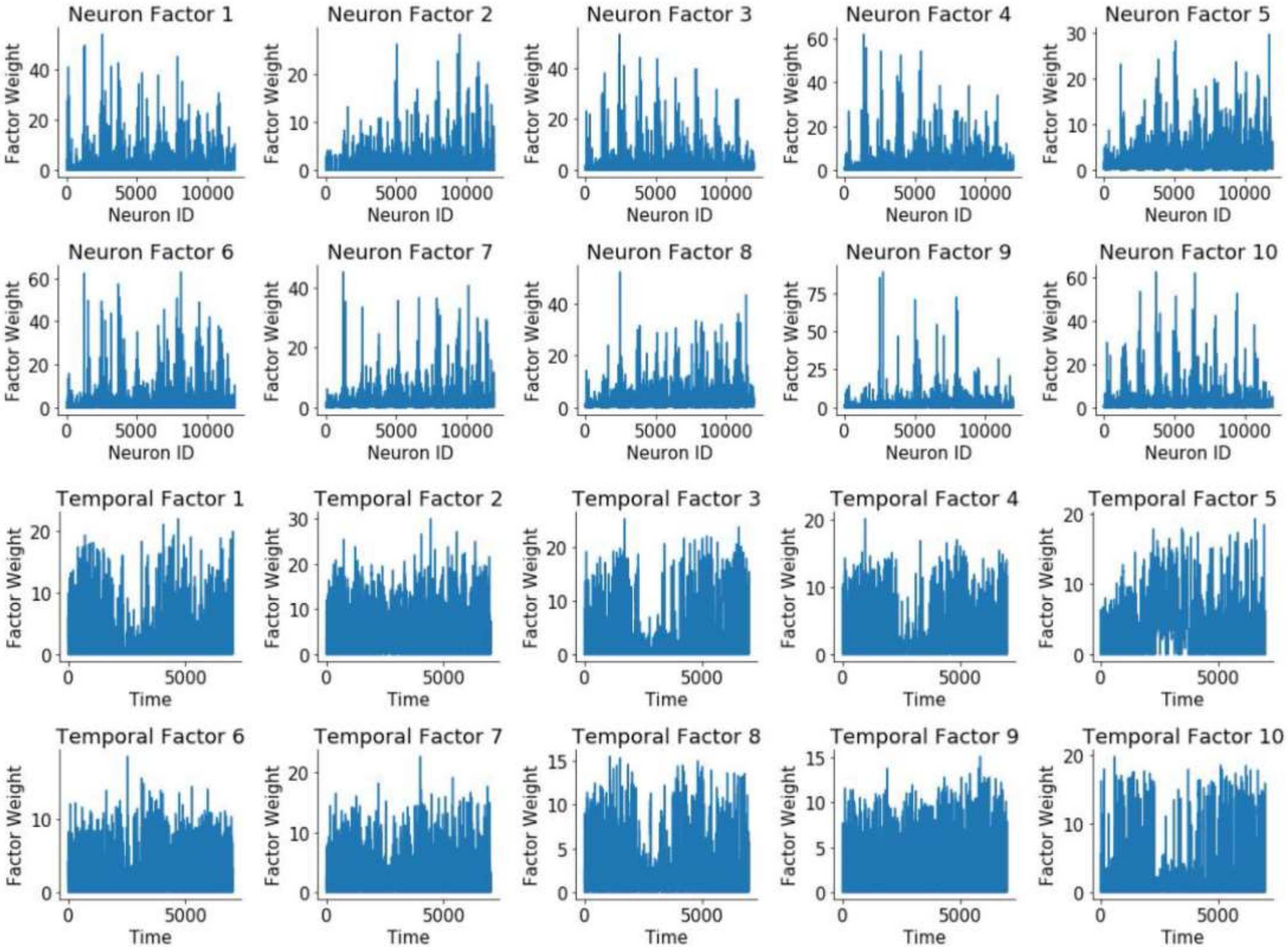
Project Summary



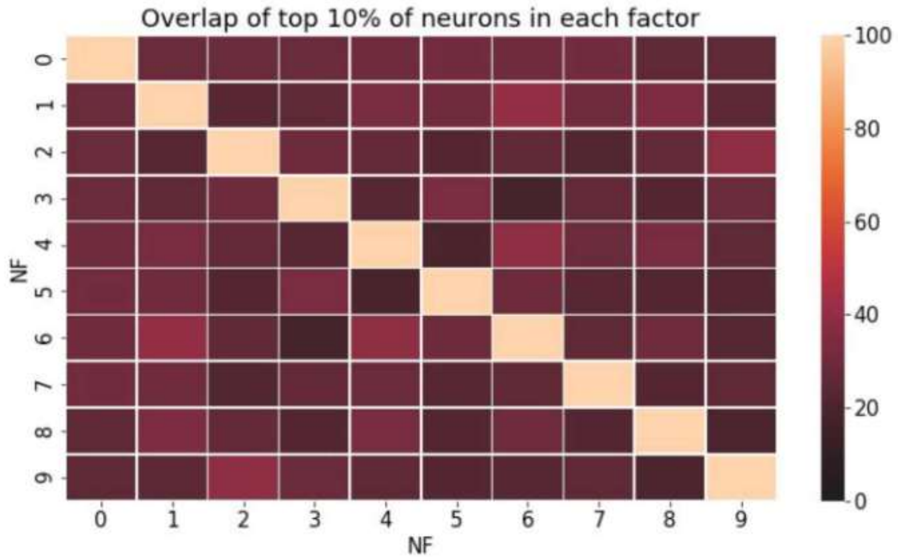
identify potential cell assemblies during specific behavioral or stimulus-driven events

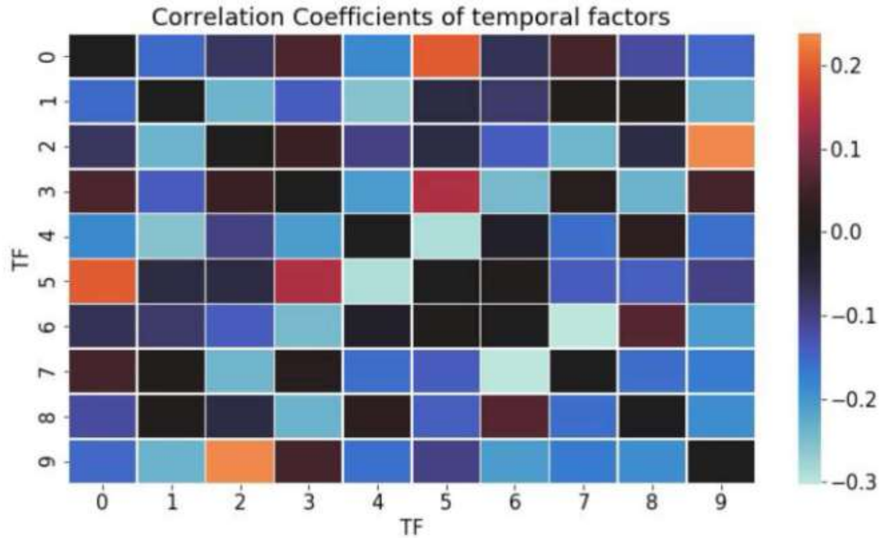
Non Negative Matrix Factorization (NMF)





NMF Overlap & Correlations

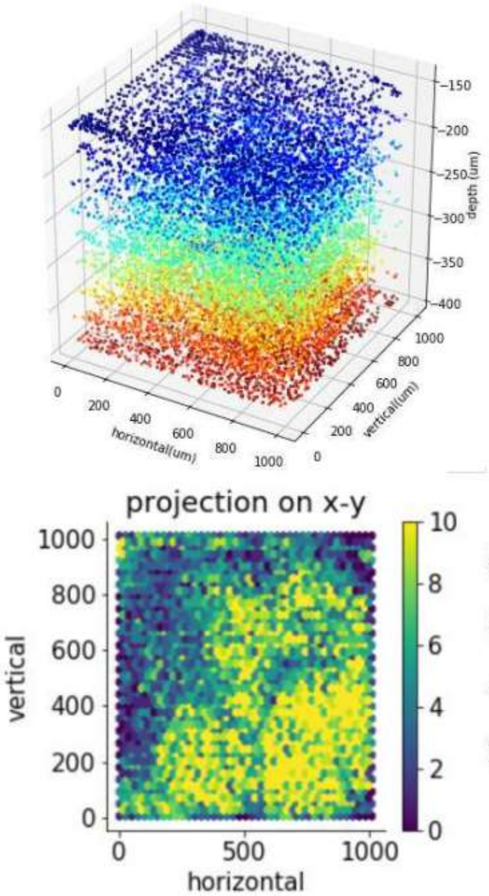


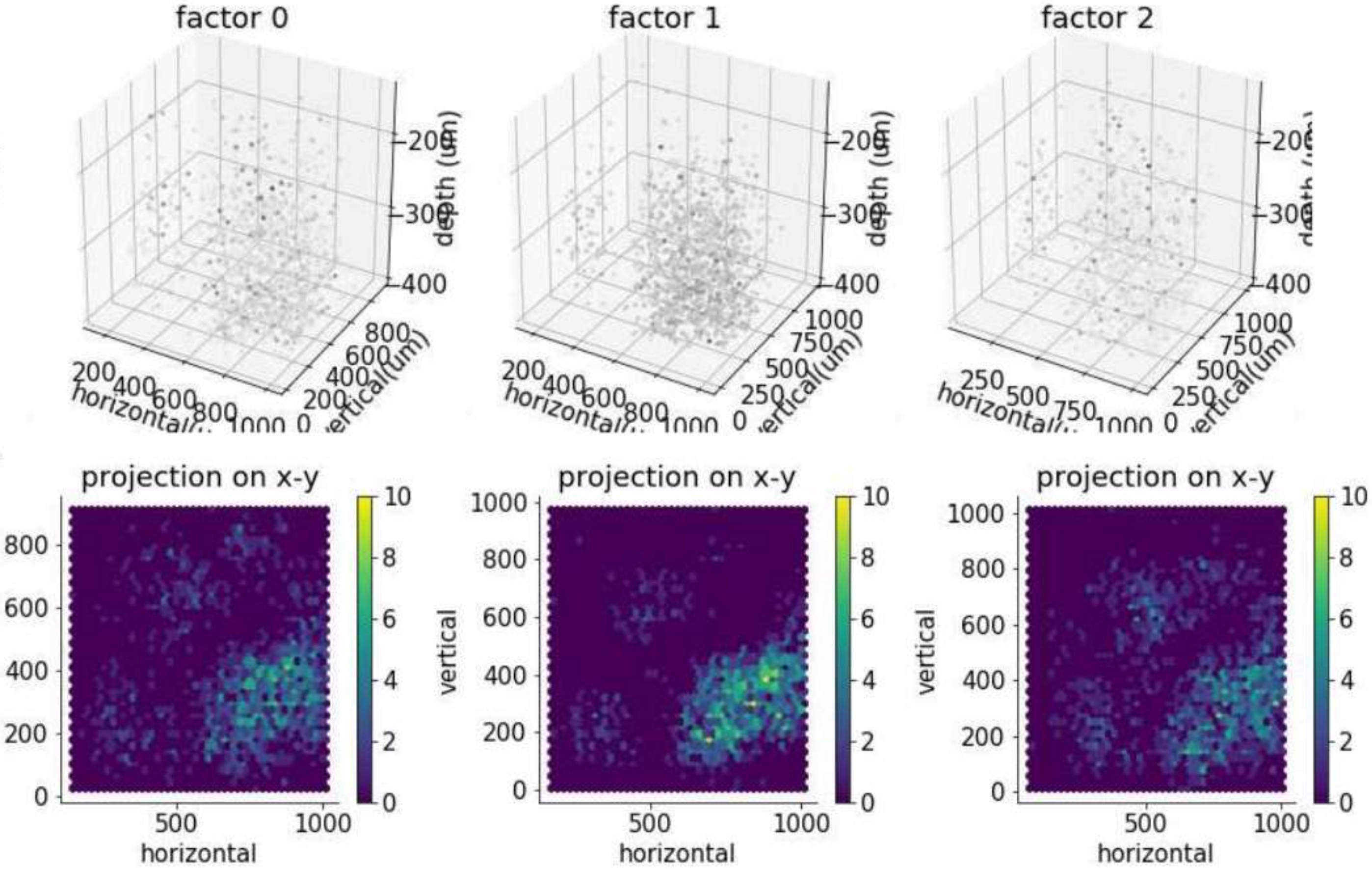


Spatial representation of factors

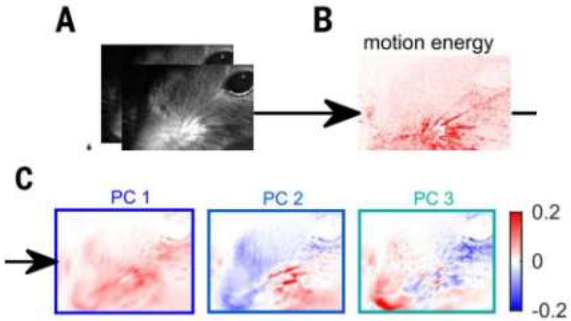
Spatial distribution - all neurons
(note the spatial sampling bias)

Plotting neurons with weights > 90 percentile
Darker color = higher weights

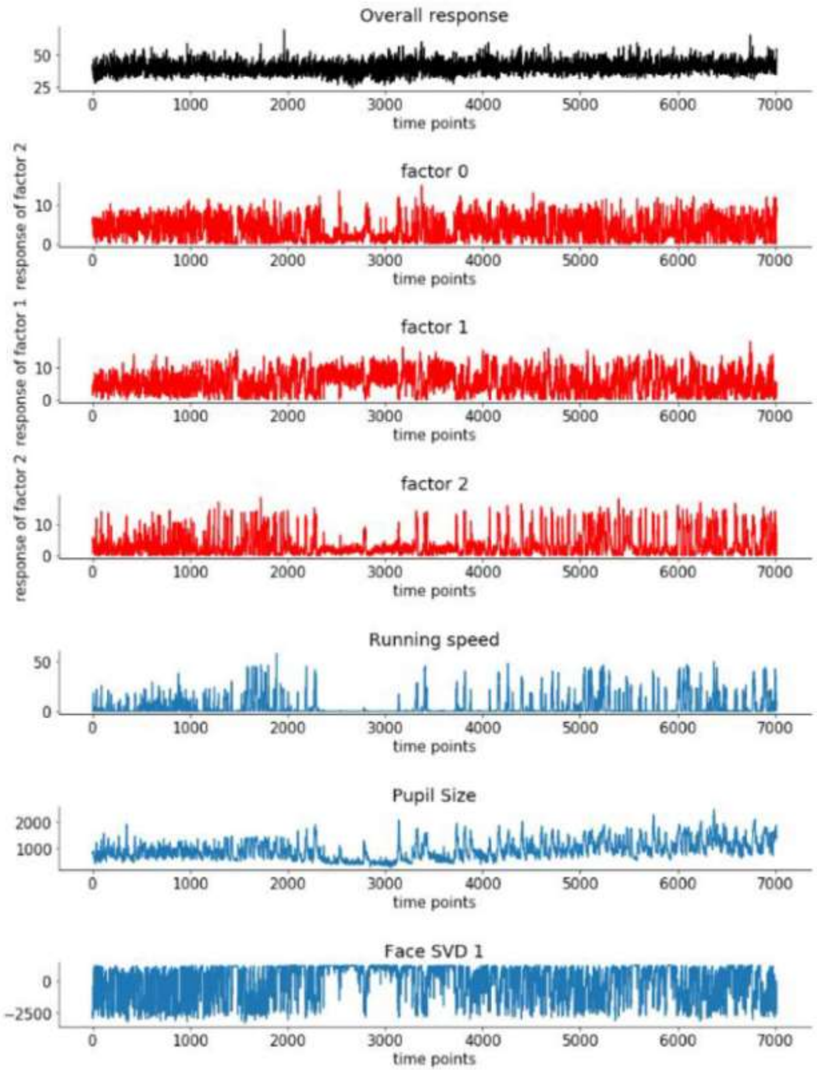




Correlation between factor response and behaviors

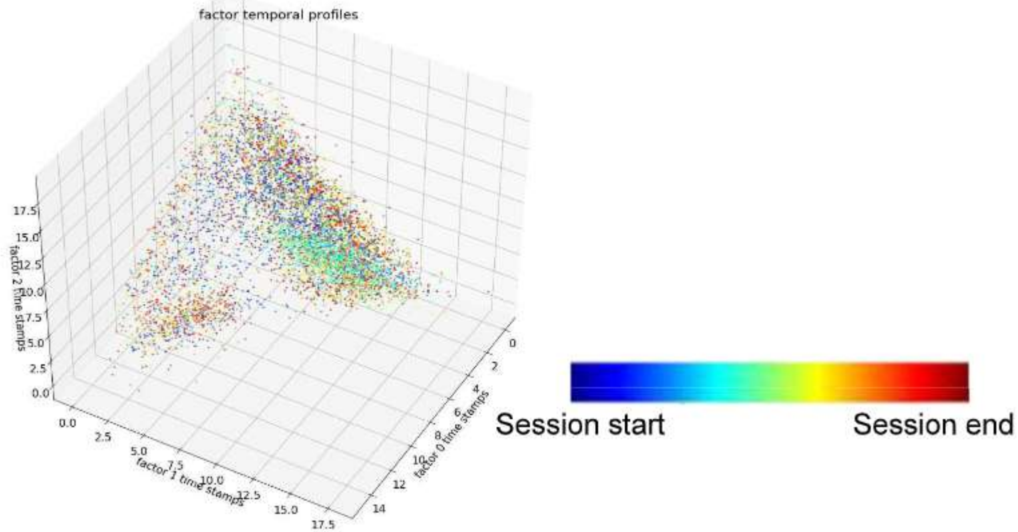


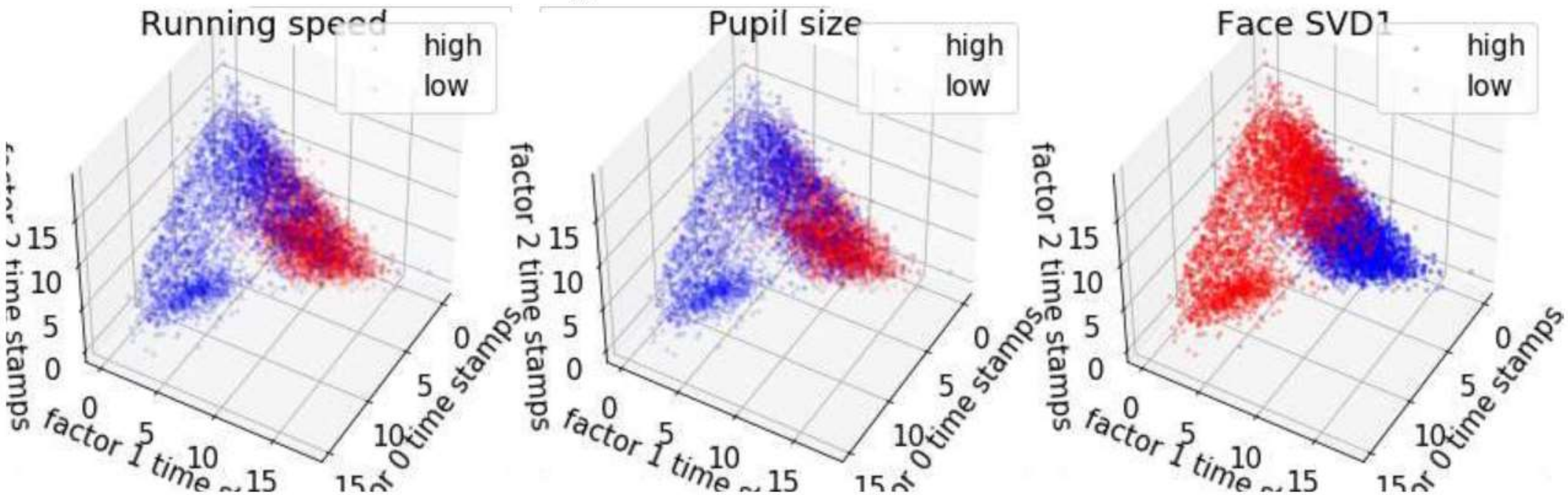
Variable 1	Variable 2	Correlation coefficient
Factor 0	Running speed	-0.204
	Pupil size	0.210
	Face SVD 1	-0.404
Factor 1	Running speed	-0.403
	Pupil size	-0.649
	Face SVD 1	0.759
Factor 2	Running speed	0.752
	Pupil size	0.653
	Face SVD 1	-0.652



(correlation with SVD2 and SVD3 was small, max ~0.37)

3 Factors' responses represented in 3D space





(High/low threshold defined by 50 percentile)

Comparing Supervised Models

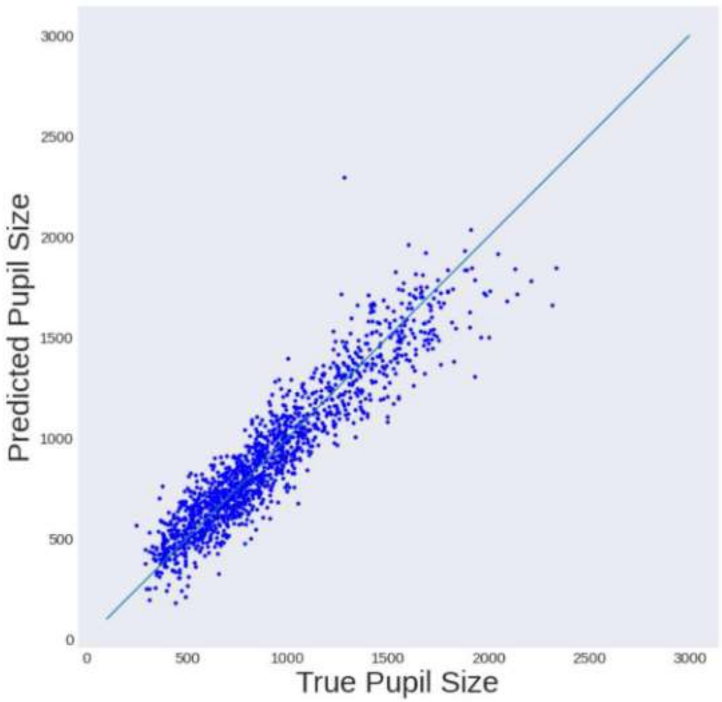
Which is the 'better' model?

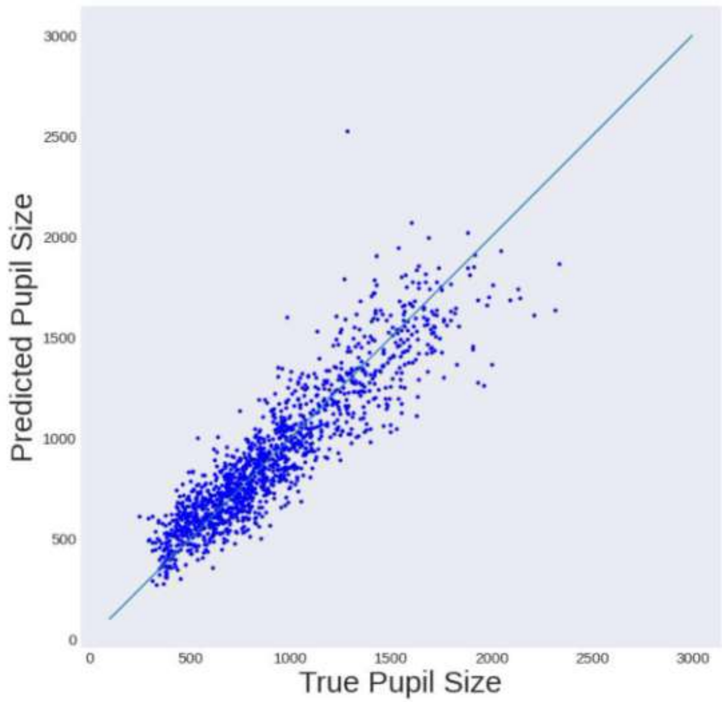
<u>Lasso Regression</u>
Cross-validation: 88.1%
Test accuracy: 87.4%

Assembly size:
1855 Neurons

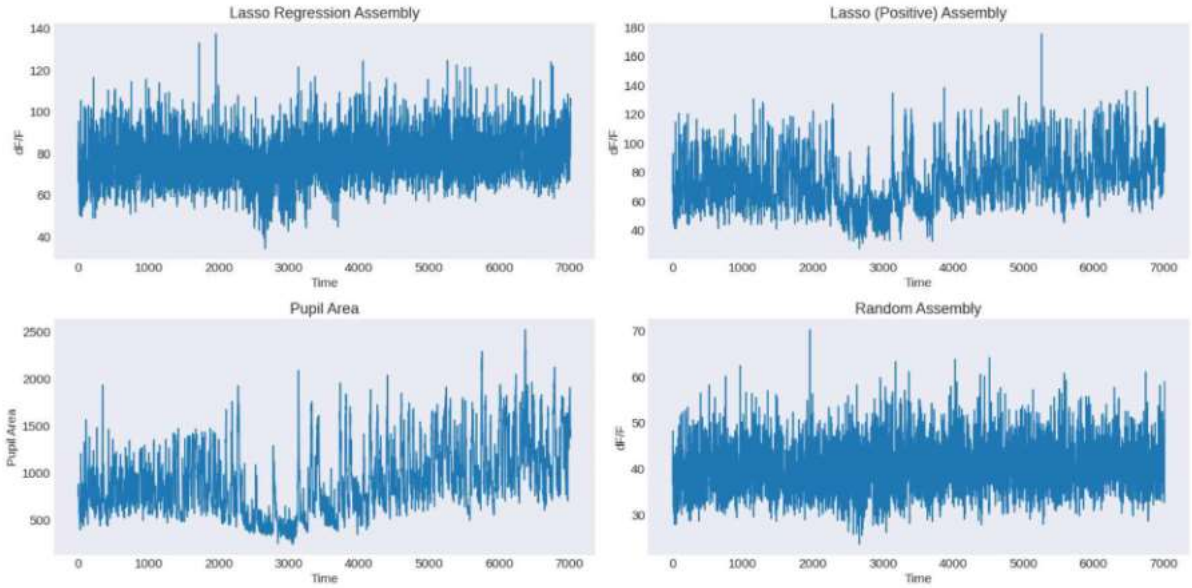
<u>Lasso Regression (Positive)</u>
Cross-validation: 84.4%
Test accuracy: 82.9%

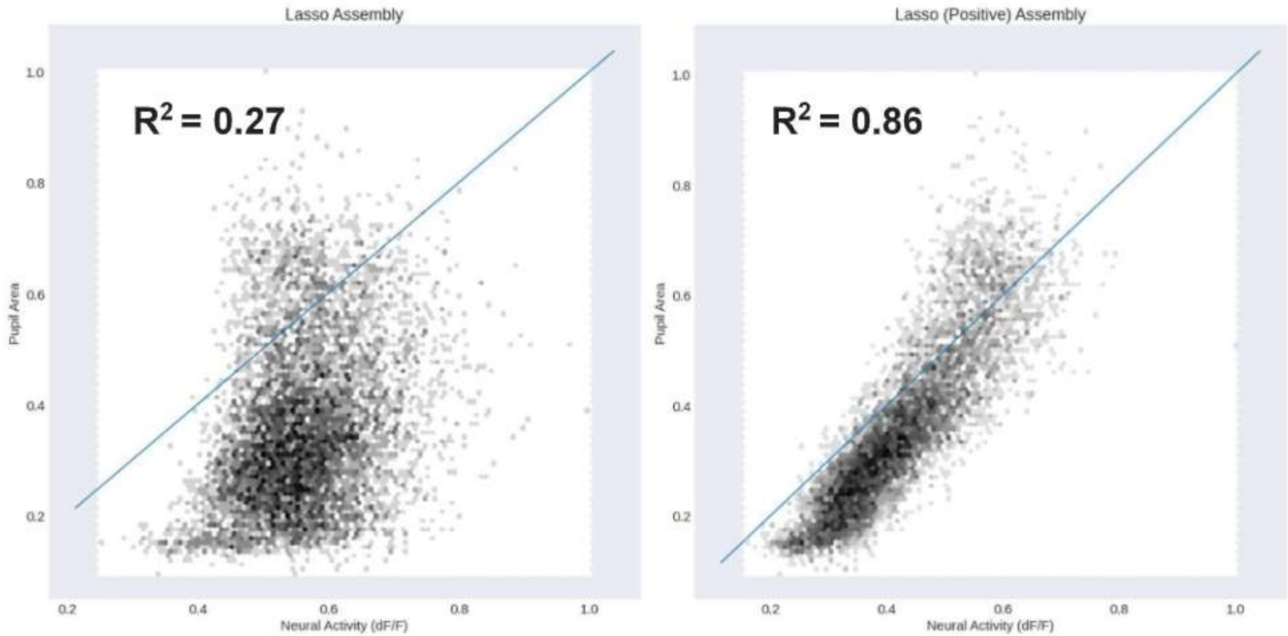
Assembly size:
559 Neurons





Comparing Supervised Models





'Better' models are not necessarily more useful!

Future project

Use marked point process to with a Bayesian model predict pupil size based on the calcium imaging data

- Marks are the amplitudes of the calcium spikes
- States are the pupil sizes within a certain range
    - We will have three states
- Using the marks we will create a probability distribution of the marks seen at each state
- We will create a probability distribution of the all the states observed during the trial
- We then use the spiking activity to predict the pupil sizes during the experiment
    - Use the two probabilities to create a posterior distribution of how likely it is that the pupil is at each state for each time point.
    - The maximum posterior is our prediction

Conclusions

- NMF gave us an unsupervised methodology for pulling out assemblies.
- However, it became difficult to choose the proper number of factors to understand the network.
- Assemblies revealed interesting spatial clustering patterns. Some ensemble activities showed correlation with behaviors.
- Supervised learning efficiently identified assemblies with correlated activity to behavioral measures (i.e., pupil area) of interest. We learned that the best fitting model is not necessarily the most useful model for interpretation, depending on our research question.

Acknowledgements

Yueqi Guo

Zeeshan Haqqee

Justin Pastore

Francisco Aparicio

Thiago Arzua

Anna Nadtochiy
(TA)

Surya Ganguli
(Mentor)

STRINGER THINGS

By Alessandro M. Vargas, Benedetta Zattera, Bianca Trovò, Chiara Baston and Marco Monforte (@pod-075-solid-jackals)

neuromatch academy

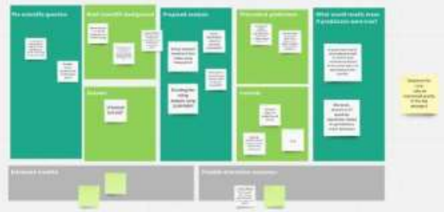
Brainstorm to Project

NEUROSCIENCE

Spontaneous behaviors drive multidimensional, brainwide activity

Carsen Stringer*†, Marius Pachitariu*†, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini‡, Kenneth D. Harris†‡

From this…



Project codename: Stringer Things

…to this

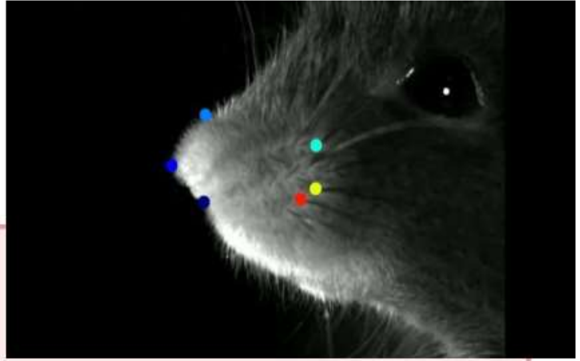
Are spontaneous motor behaviors encoded in V1?

➔ **Mouse visual cortex** is strongly modulated by running and facial movements (*Stringer et al. 2019, Science; Musall et al. 2019, Nature Neurosc.; Niell and Stryker, 2010, Neuron*).

➔ **Facial movements** may directly affect the visual input, which might explain why they affect V1 responses.

➔ Our aim is to investigate **if** facial movements (such as **whiskers**, **pupils**...) are **encoded in V1 neurons**.

Project Data and Methods



DATA: Stringer dataset.
- Calcium imaging from V1 area.
- Behavioral data: pupils movements, area and motion energy.

AIM:
Explore which predictors can better explain the neural activity.

METHODS (working):
- dimensionality reduction of neural data
  - **UMAP**
- Regression model
  - **GLM**

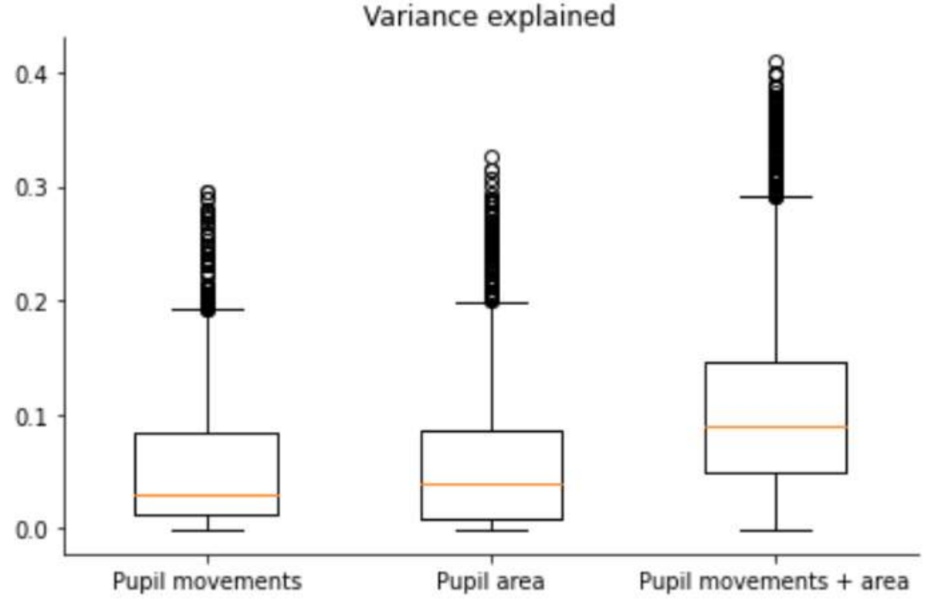
AIM:

Extract spatio-temporal information about the neural states in the latent space.

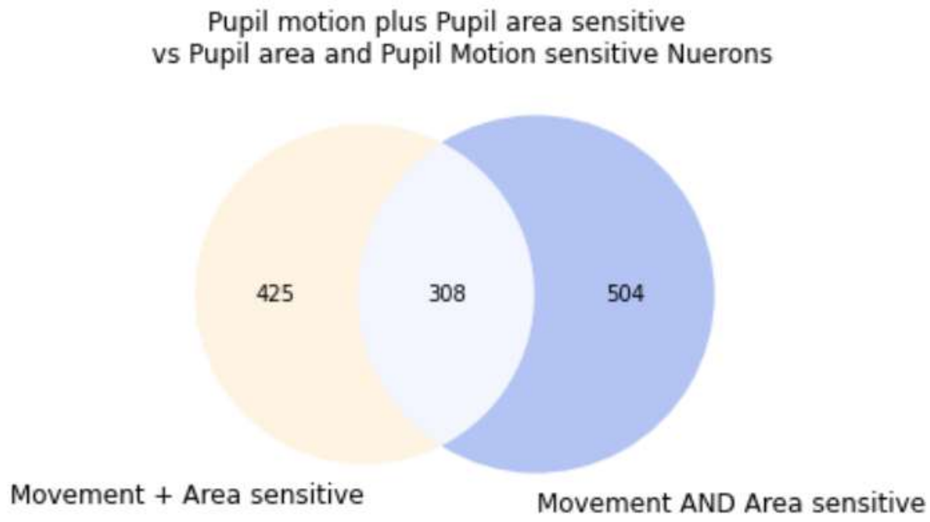
METHODS (tried):
- dimensionality reduction of neural data
  - **Gaussian-Process Factor Analysis** (rank deficient matrix on the whole dataset)
  - **Functional Principal Component Analysis** (infinite computational time)
- state space model of neural data
  - **Hidden Markov Models** (ran Out of Memory)

Results and Experience

Pupils' movements and area can explain about **10% of the neural activity** in V1.



Neurons whose activity is predicted by pupil's motion and area together are only **partially overlapping** with neurons whose activity is predicted by the sum of their components taken separately.



Intrinsic Structure of Orientation Response Manifold

By: Elizaveta Kozlova, Anna Vasilevskaya, Viktoryia Kuryla, Egor Zverev
pod-162-great-mule

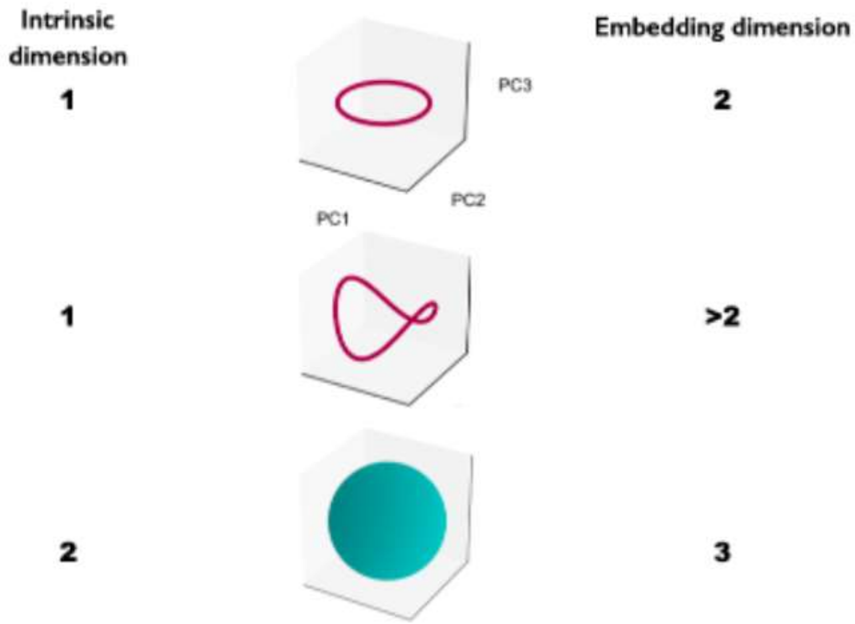
neuromatch academy

Proposal

- **question:** What is intrinsic and embedding dimensionality of mice orientation response manifold? What is its topological type?
- **dataset:** Stringer orientations
- **methods:**
  - dimensionality reduction (PCA, FA, Isomap, LLE, ICA)
  - dimensionality estimation (Grassberger-Procaccia, PCA/FA modifications)
  - topology analysis (persistent homology)

Dimensionality

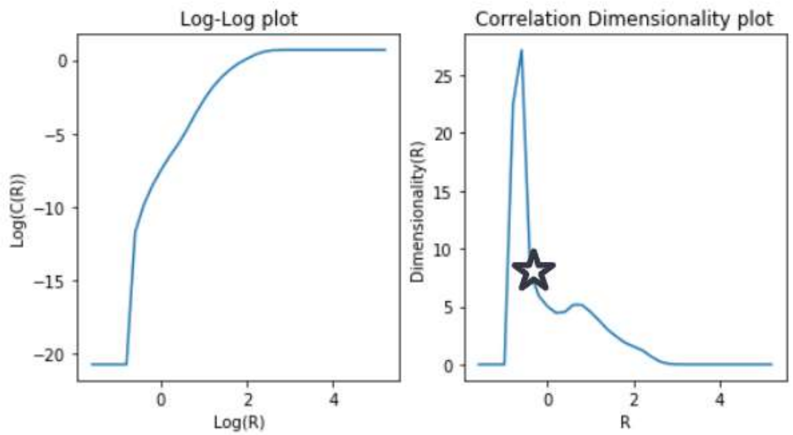
- **intrinsic dimensionality** —
  the number of variables needed in a minimal representation of the manifold
- **embedding dimensionality** —
  the dimensionality of the minimal linear space that contains the manifold



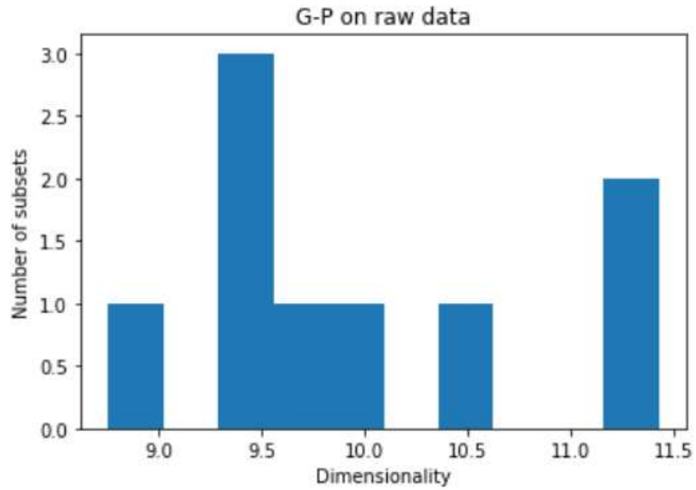
Dimensionality

- estimated embedding dimensionality: ~23-24

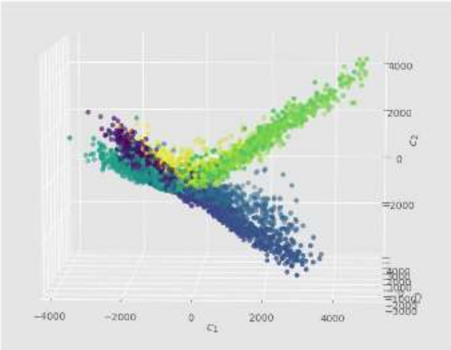
Dimensionality

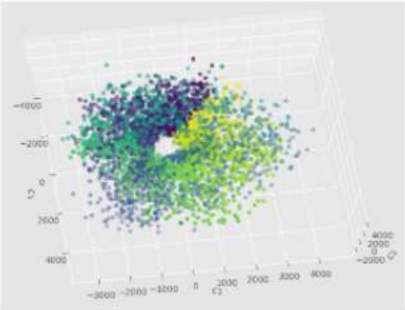


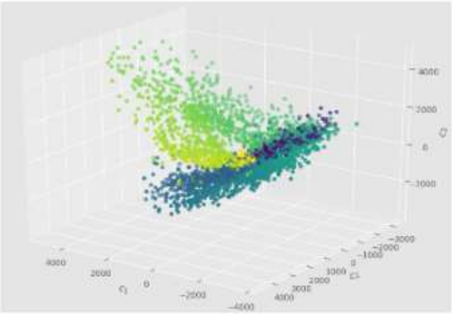
- estimated intrinsic dimensionality: ~8-10



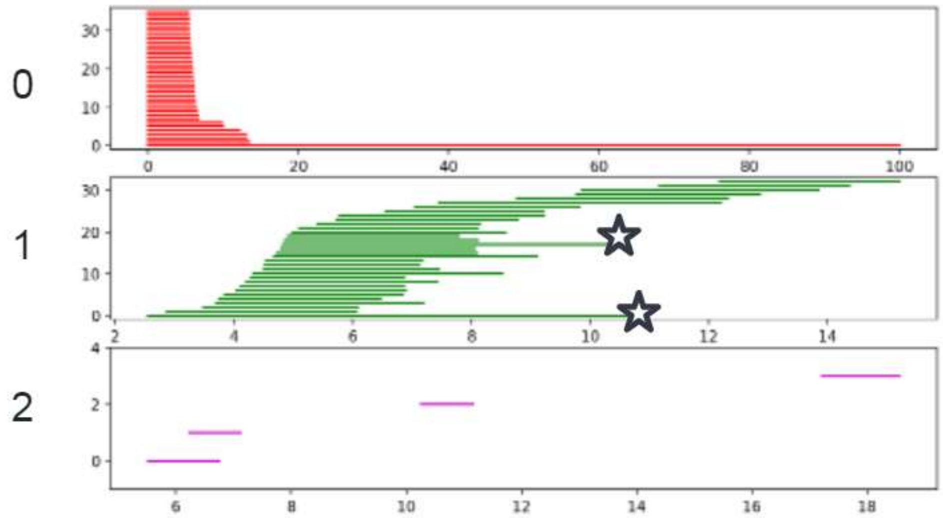
PCA topology analysis



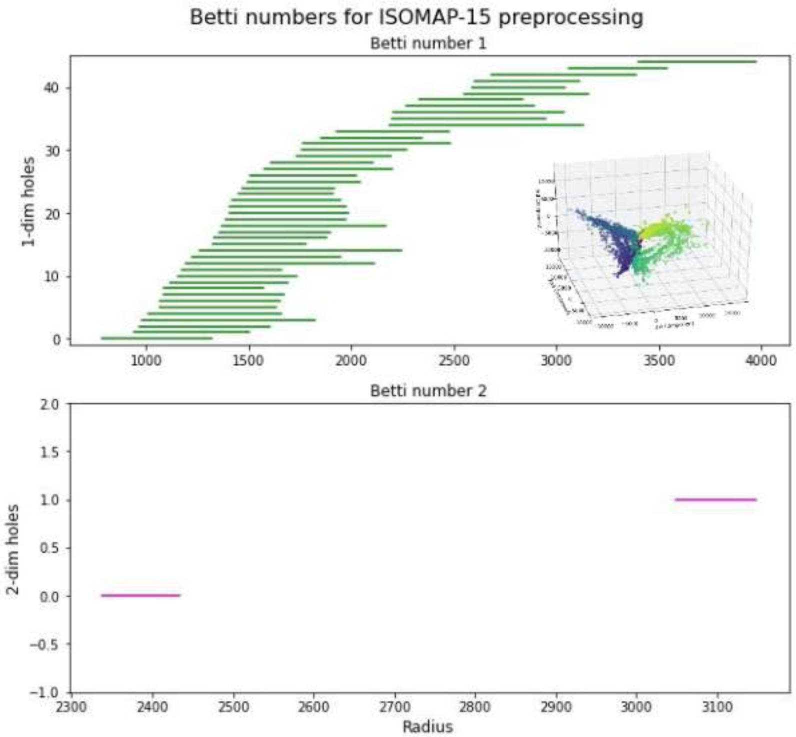


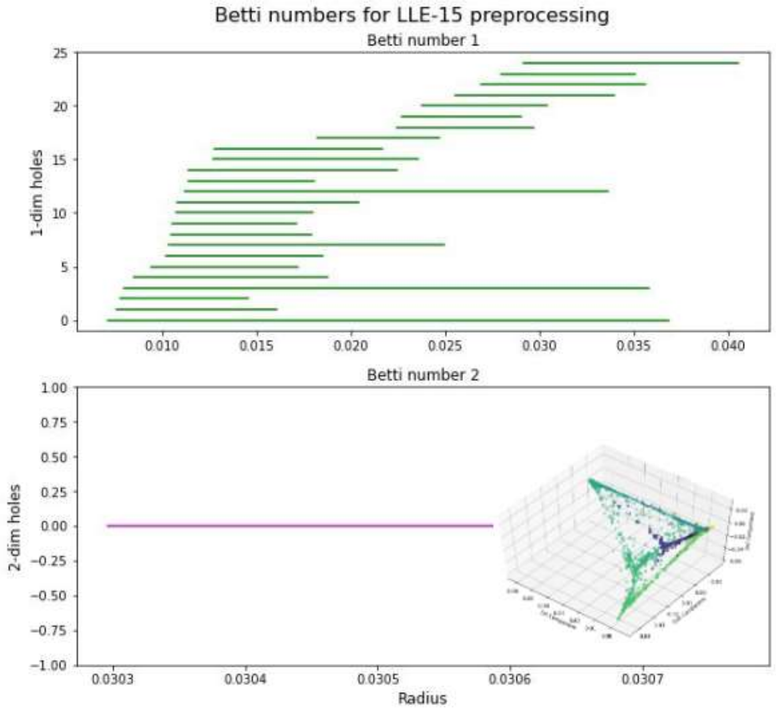


Betti numbers
(holes of different dimensionality)



Preprocessing influence





Discussion

Are our results reliable?
No, but we explored a lot!
(it would be interesting to make a more detailed analysis)

Many thanks

to our mentor Srdjan Ostojic!

Decoding behavioral states from V1 neuronal population activity

By: Paul, Dean, Bashir, Will, Jacques
Matrix Mice - Faithful Millipedes

neuromatch academy

[Matrix Mice]

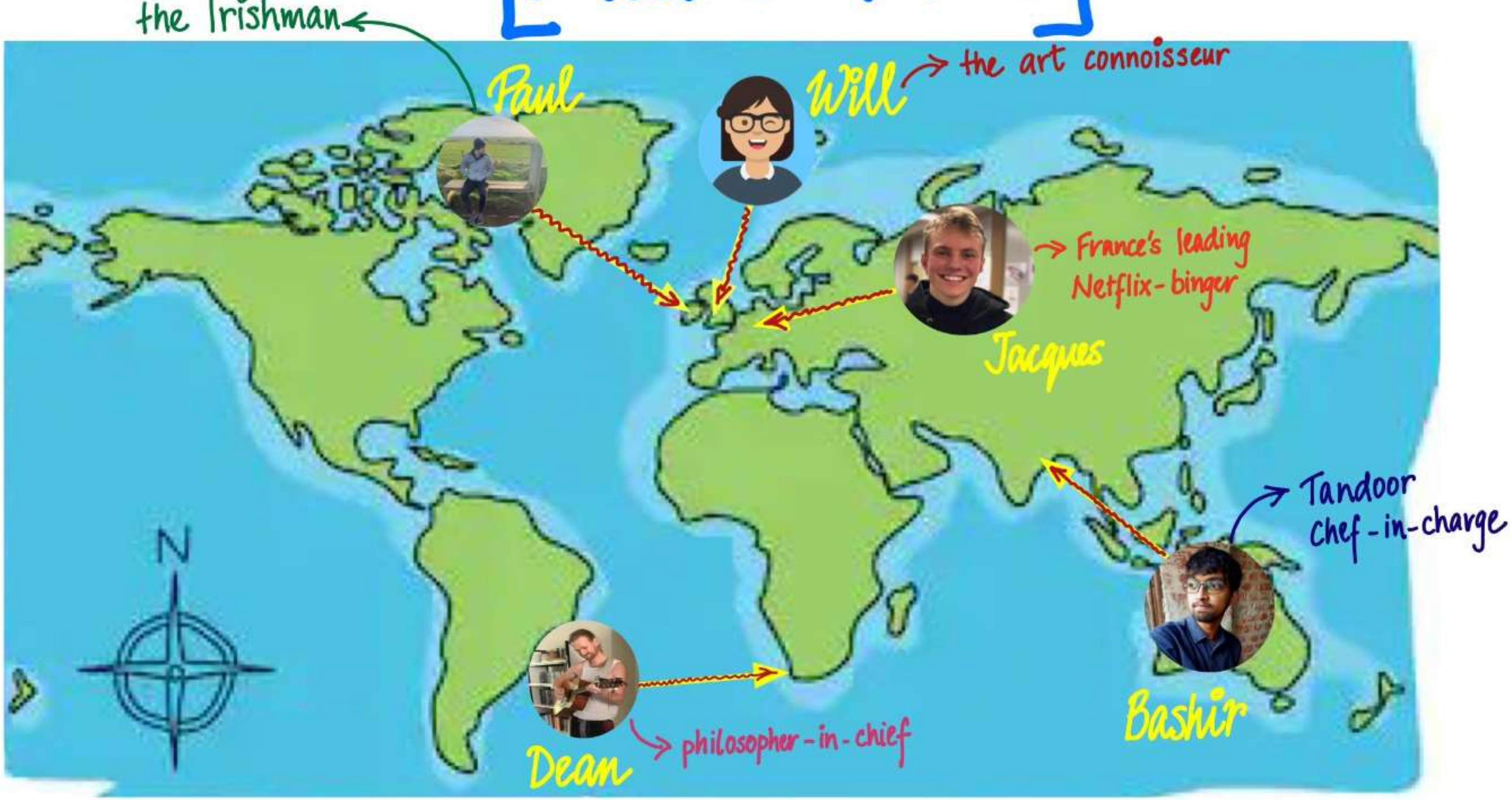


Image by Astarina https://www.shutterstock.com/image-vector/world-map-doodle-598322657

What did we do?

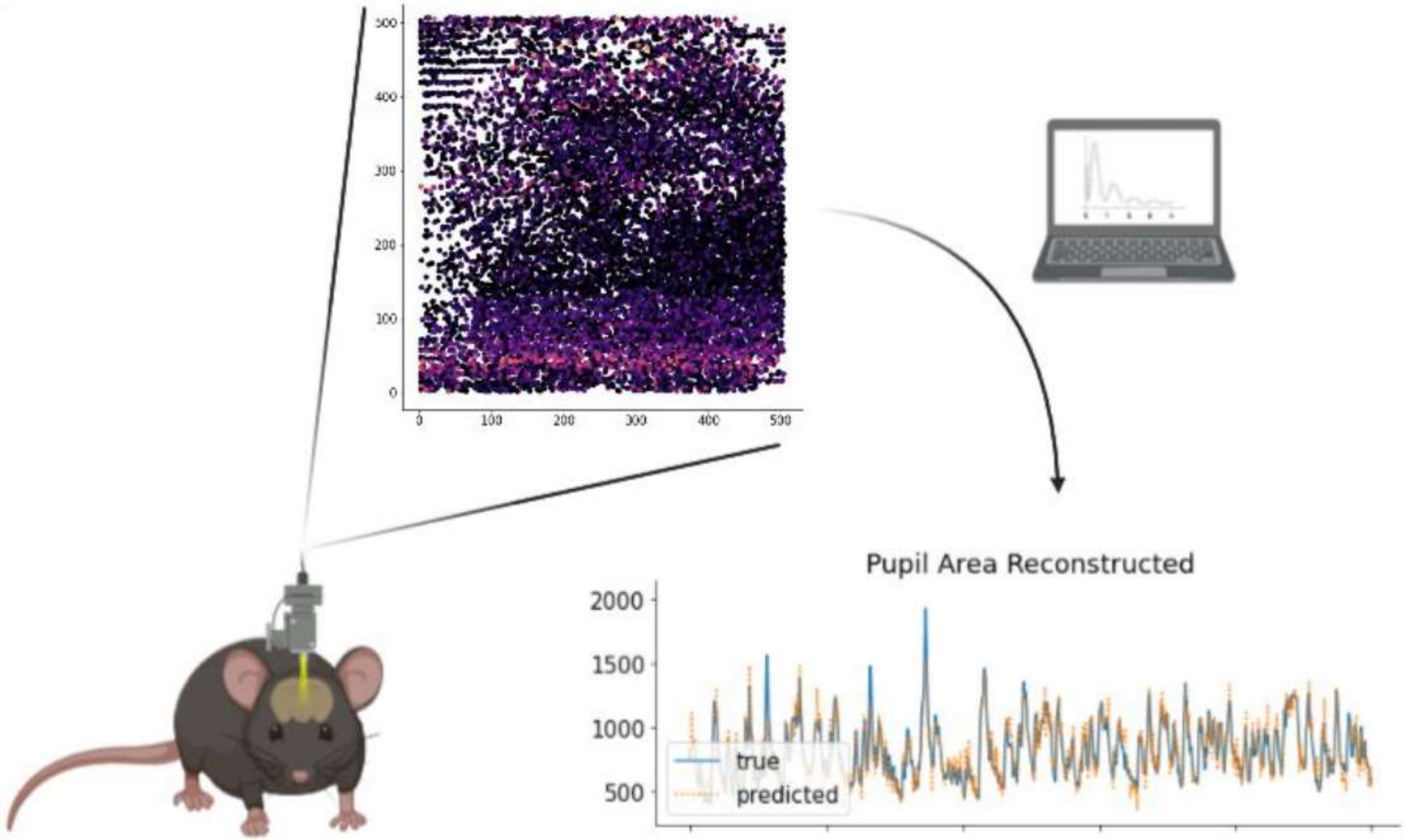
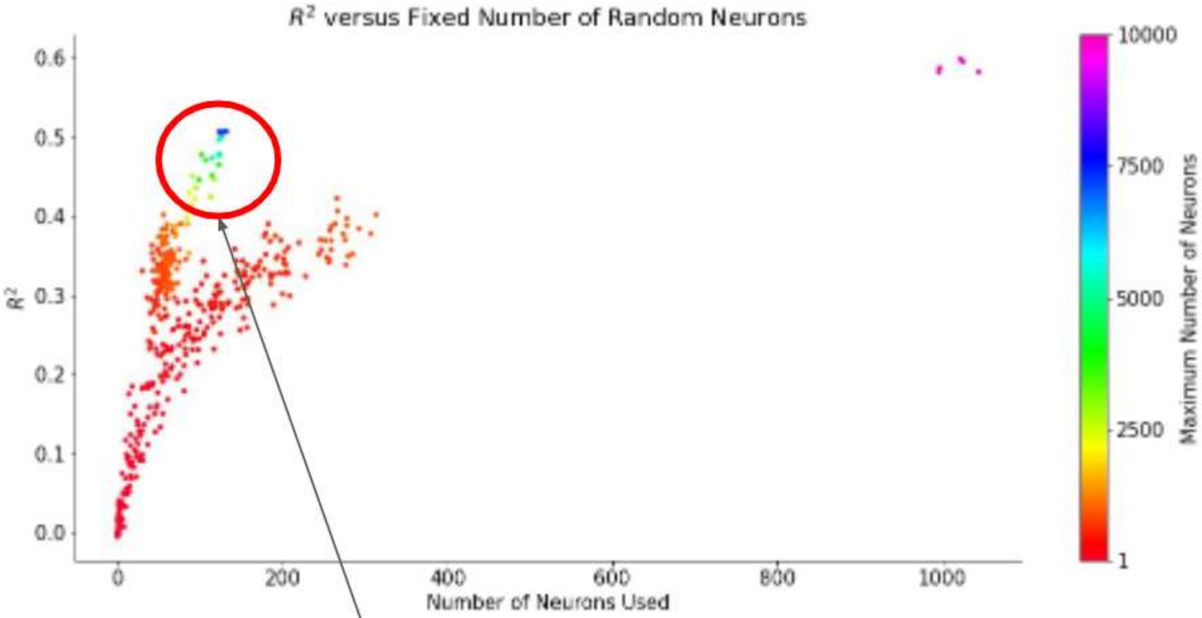


Image created on biorender.com

Stringer et al.:
Behaviour predicts neural activity

Us:
Behaviour can predict behaviour
Neural activity can predict behaviour

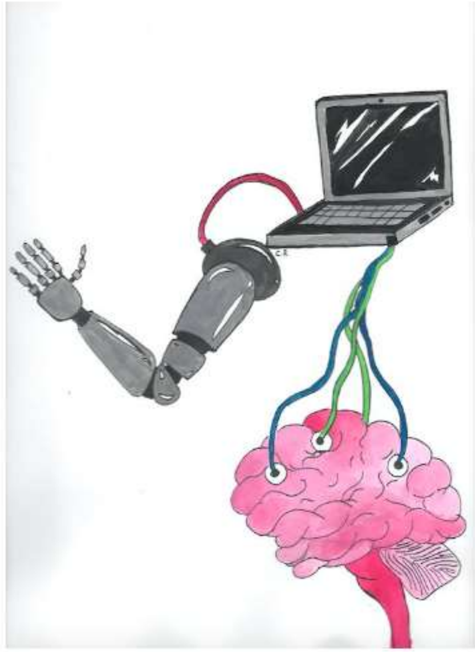
What would we do?



What's happening here?

(they are not the same neurons…)

- "150" the magic number?
- The existence of sufficient neurons?



The McGill Daily

Effects of Multiple DR Techniques on Decoding of Orientations from V1 Responses

Nicholas Cimaszewski, Yang Xiang, Alireza Ghadimi, Javan Tahir
Advised by Dr. Ruben Coen-Cagli

Stringer Orientation Dataset

Meet our team

Mentor: Ruben

Alireza Yang

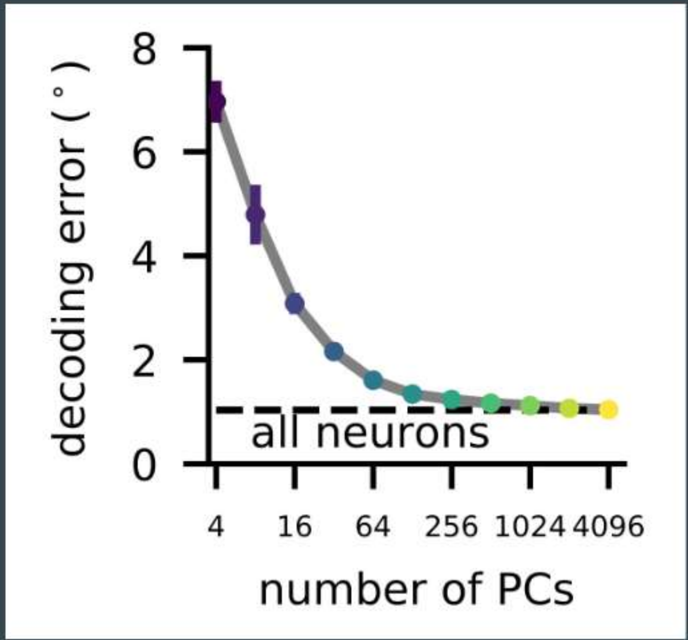
Javan Nick

Motivation

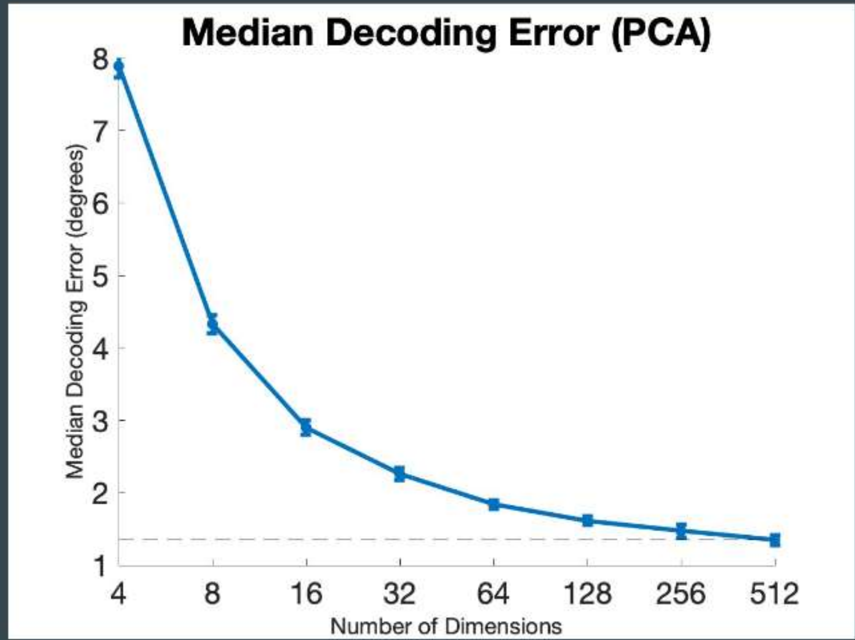
- We were interested in finding a lower dimensional representation of the Stringer orientation dataset.
- After reading Cunningham & Yu (2014), we wondered which dimensionality reduction techniques would best capture the structure of the data?
- Initially considered investigating the inherent dimensionality of the response manifold, but for such a low dimensional stimulus set it seemed trivial
- Instead investigated performance of a linear decoder as a function of dimensionality for different techniques

The path we “stumbled down”

Sanity check

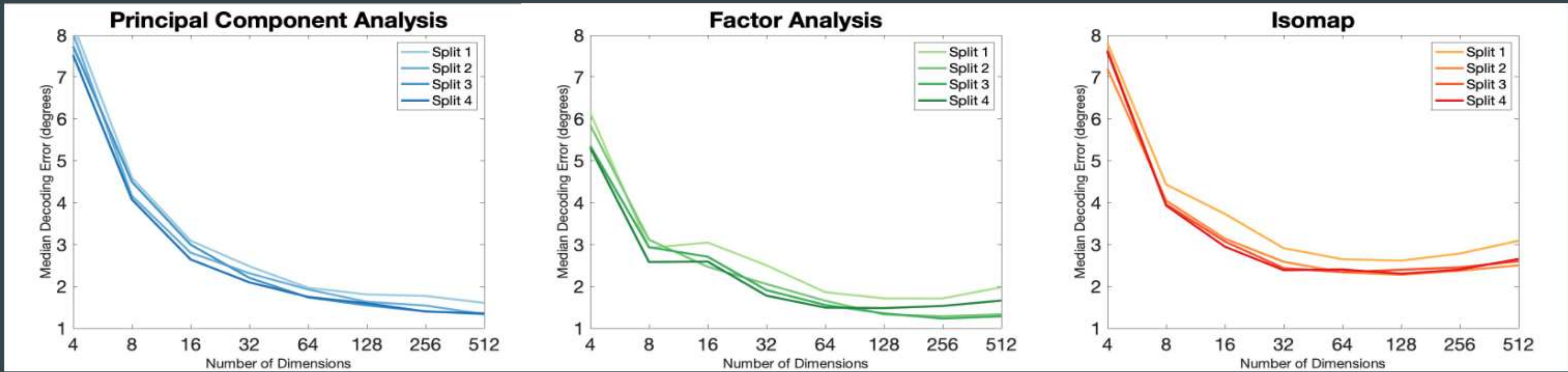


Stringer et al., 2019

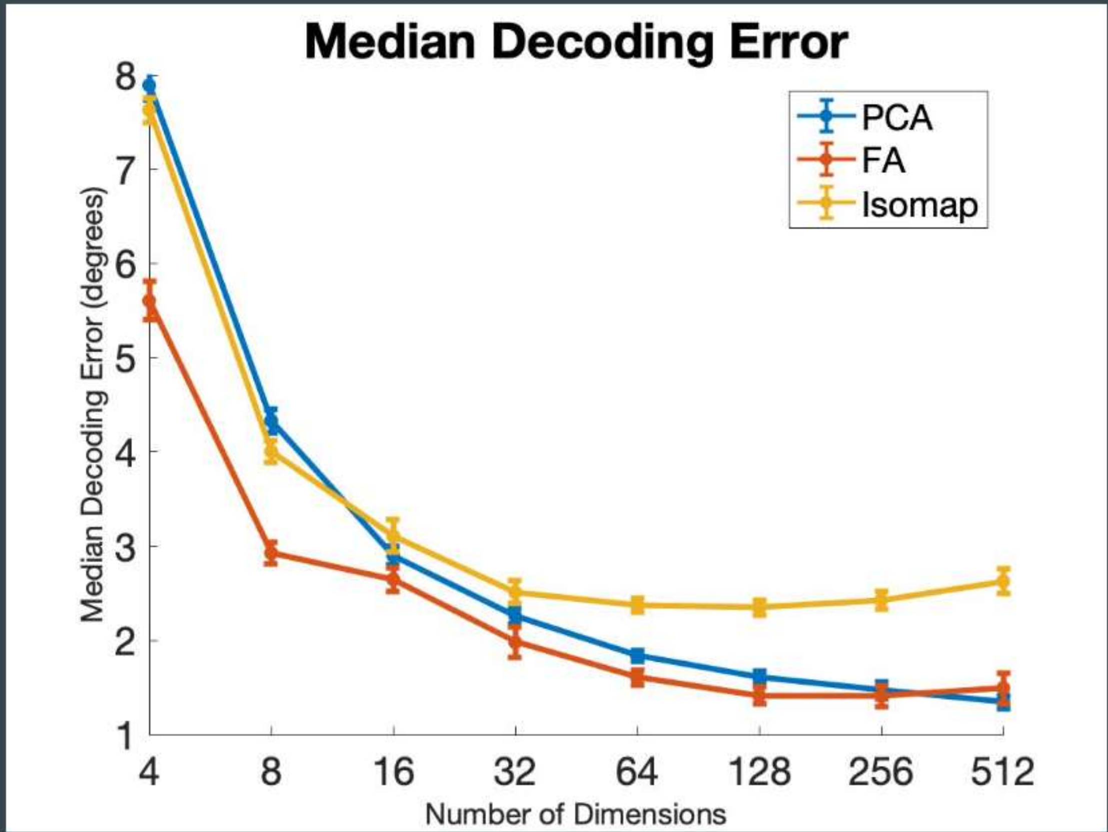


Our analysis

Comparing PCA, FA, and Isomap

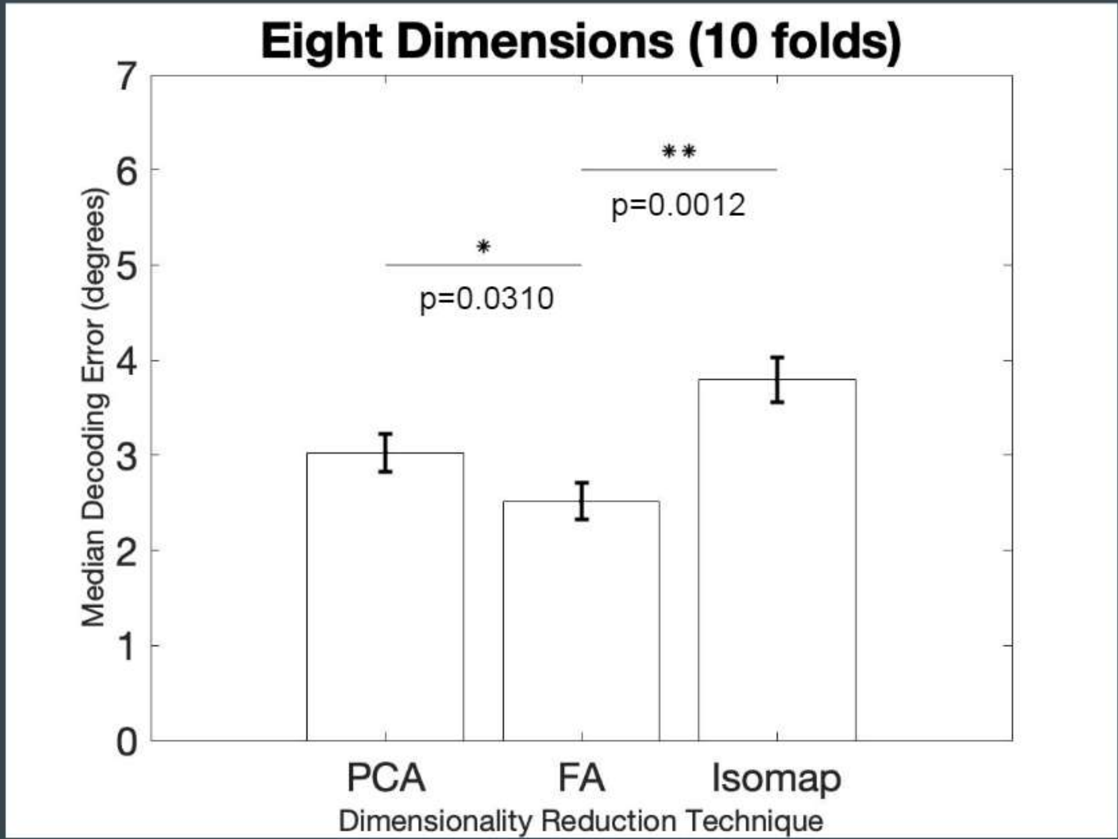


Comparing PCA, FA, and Isomap



Error bars: SE

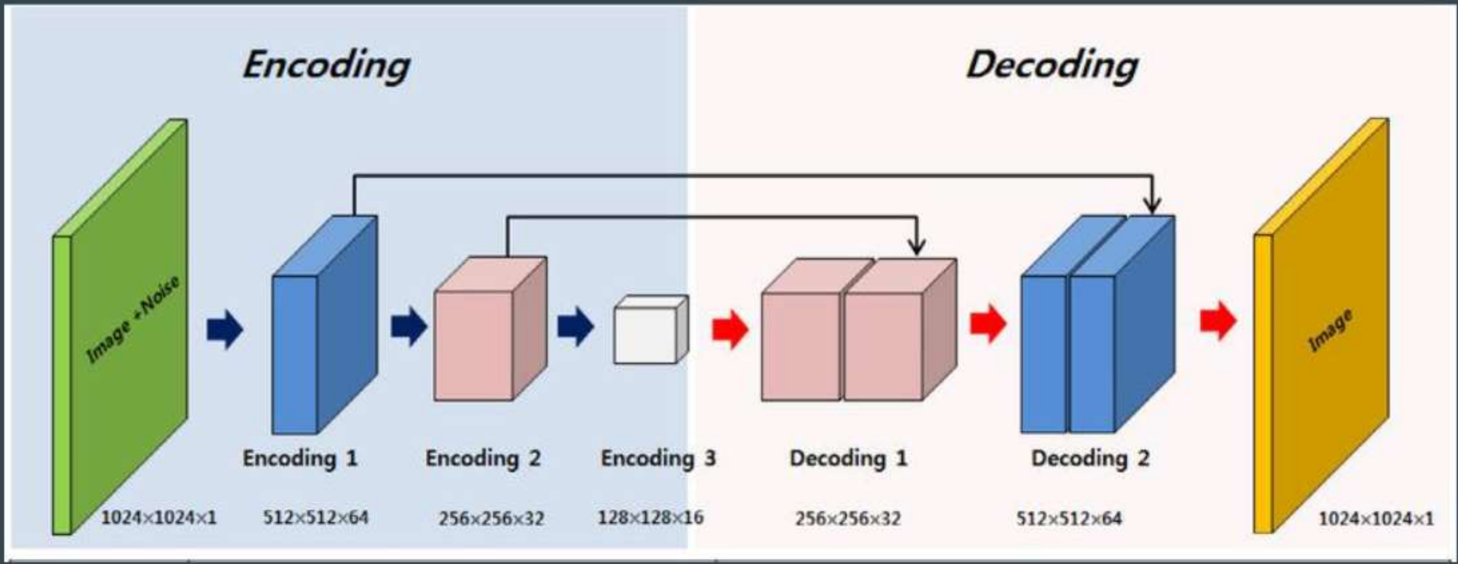
Comparing PCA, FA, and Isomap



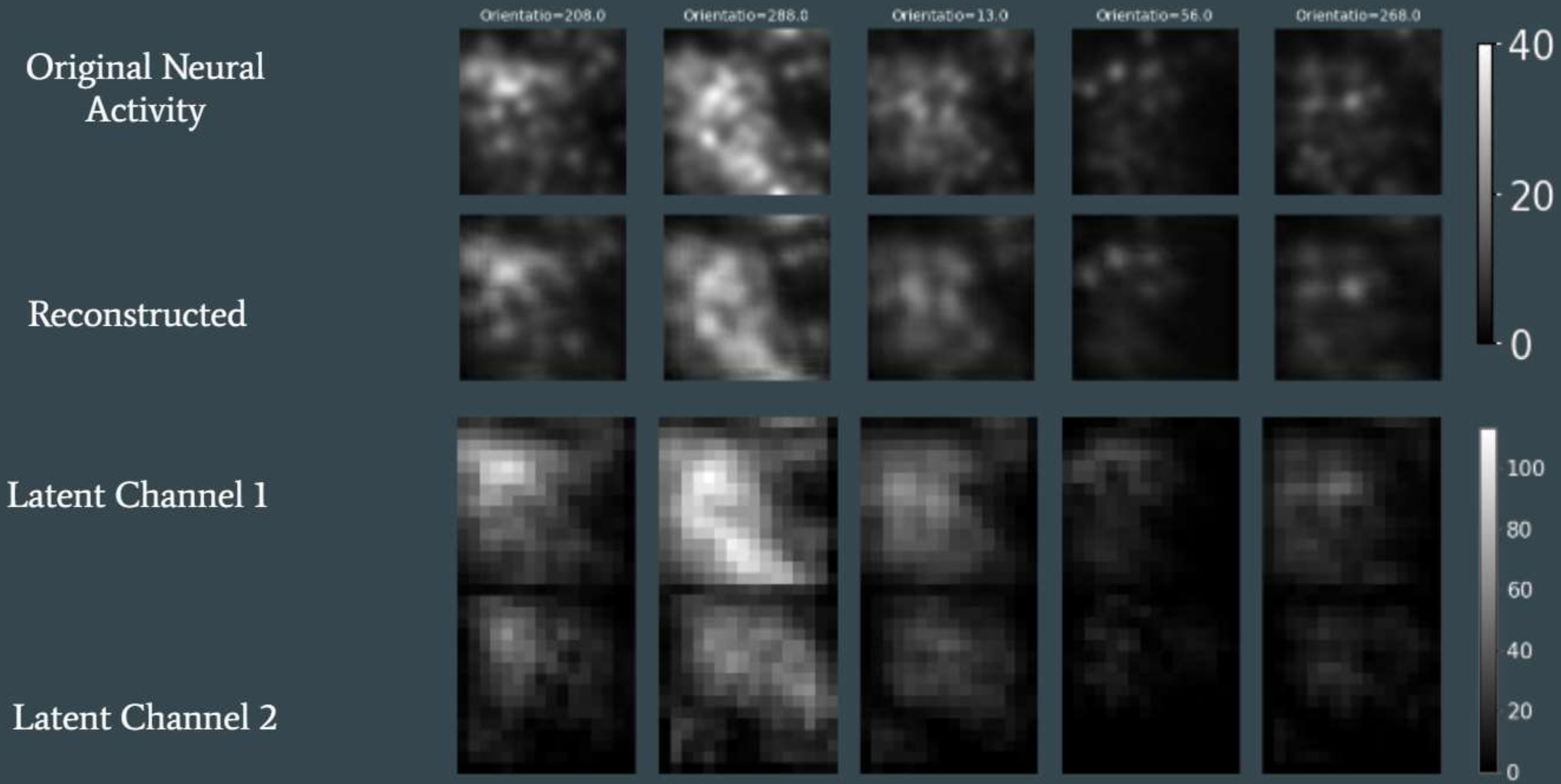
Error bars: SE

Autoencoder

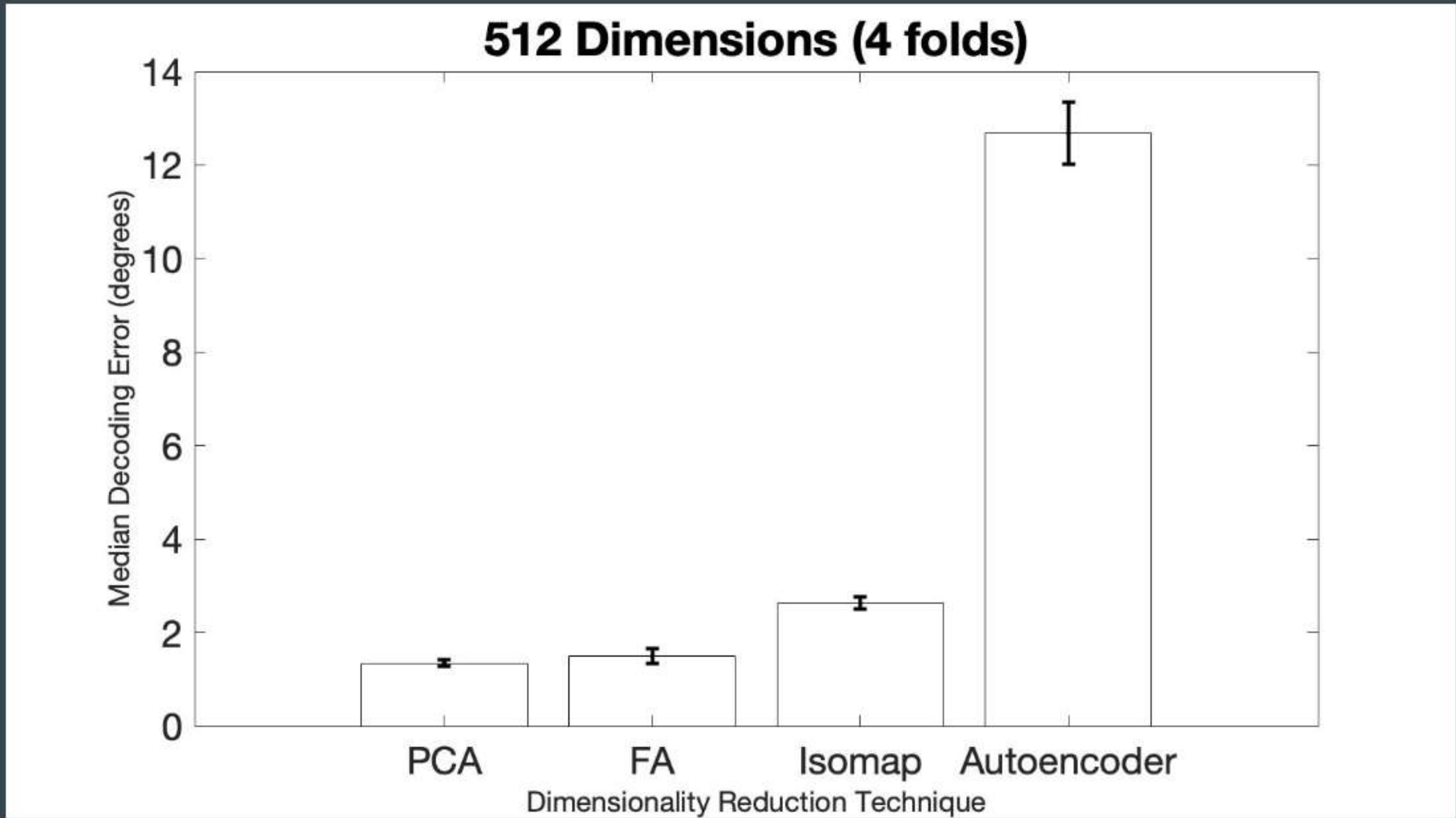
- Convolutional autoencoder may be able to exploit the spatial structure of the data
- May learn a better latent representation than traditional dimensionality reduction methods



Autoencoder



Comparing PCA, FA, Isomap, and autoencoder



Error bars: SE

Conclusion

- **Linear vs non-linear**: Isomap never achieved as high a performance as the linear methods beyond ~32 dimensions
- **PCA vs FA**: FA did slightly but significantly better than PCA for low number of dimensions
  - Perhaps the assumption of different variances for the latent variables has some benefit for low dimensional projections of firing rates
  - For dimensions as low as 8, the runtime for FA is not significantly greater than for PCA
- **Autoencoder vs everything**: Autoencoder failed to capture the relevant structure of the data even in a relatively high latent representation

Acknowledgement

We would like to say thank you to the NMA organizers -- This was a wonderful experience!

We also want to thank Carsen Stringer et al. Without them we would not have had this curated dataset to work on :)

Special thanks to our mentor Dr. Ruben Coen-Cagli, who is super well-versed, super nice, super helpful, and super supportive!

Reference

Cunningham, J. P. and Byron, M. Y. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509.

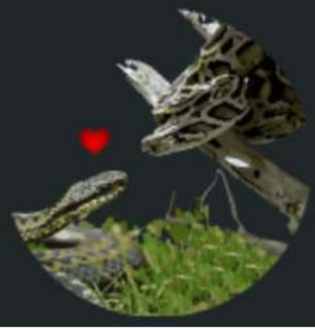
Stringer, C., Michaelos, M., and Pachitariu, M. (2019a). High precision coding in visual cortex. *BioRxiv*, page 679324.

Decoding stimulus orientation in visual cortex in the context of running

By: Antonio Ortega, Agustina Frechou, Habiba Noamany, Daniela Buchwald

Pod: 001 - Tunneling Adder - Group: Mind Readers

neuromatch academy

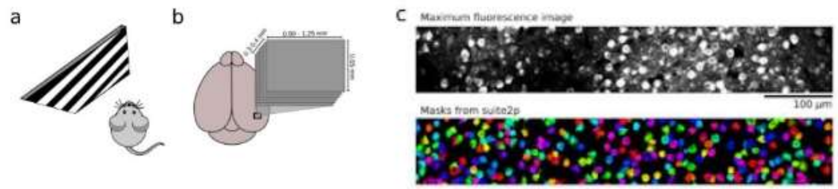


Objective Can we predict stimulus orientation from neural activity in visual cortex? Does running speed affect neural activity in the visual cortex?

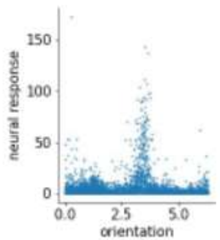
Dataset: Stringer (orientations)

Background: One important characteristic of living beings is the ability to produce behaviours. While many different brain areas are involved in the execution of behaviours, preceding neuronal activity can be found in many areas that are seemingly unrelated to this behaviour.

Task:



https://www.biorxiv.org/content/10.1101/679324v2.full.pdf



Approach:

- Linear Regression
- LDA
- PCA
- Bayesian Modeling

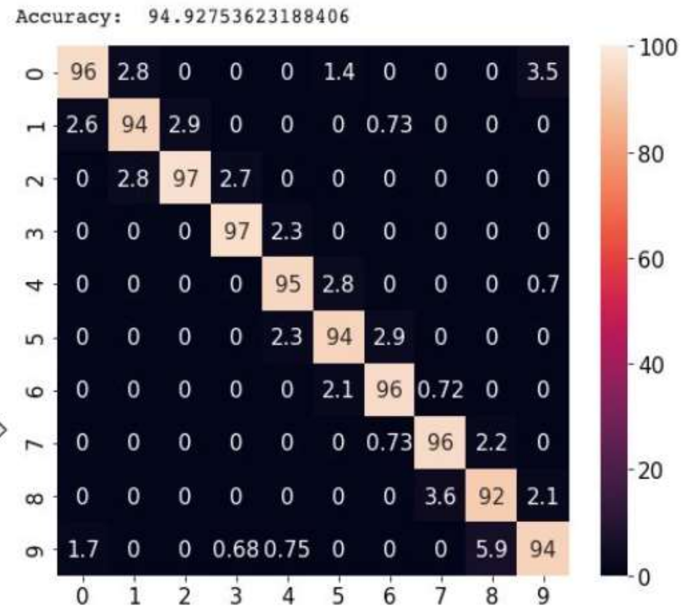
Can we use simple linear regression and classification methods to decode angle and running speed?

	R^2		√MSE	
	Train	Test	Train	Test
Neurons → angle	1	.86	0 °	40 °
200 PCs → angle	.83	.77	42 °	50 °
200 PCs →speed	.88	.80	6.5	8.3

overfitting!

How about classification?

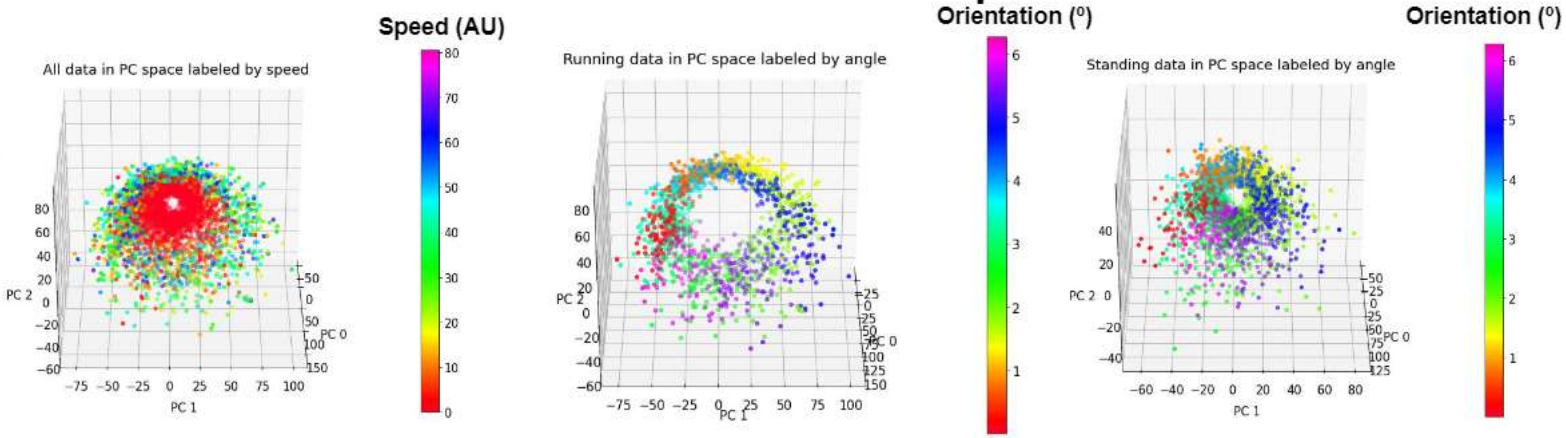
Decoding orientations using Linear discriminant analysis (LDA)



How might the PCs allow for decoding of both running and stimulus orientation?

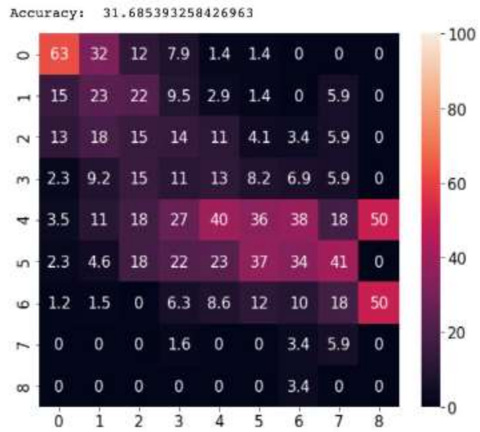
Investigating the segregation of speed and stimulus in neural responses transformed to PC space

1 Splitting data by running and standing, we see a difference in the spread of the data

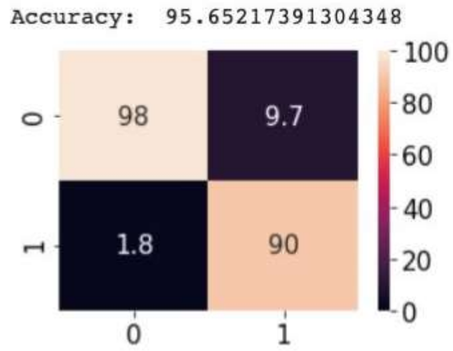


2 Can a decoder classify speed from V1 activity?

Binning speed into 10 categories and trying to decode using LDA results in low accuracy (~32%)



3 Splitting the data into running and non running (based on the segregation of the PCs), however, allows for decoding above chance level

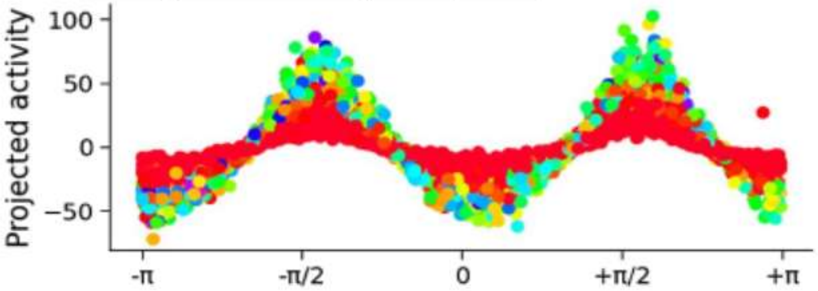


Estimating effect of running speed on neural activity

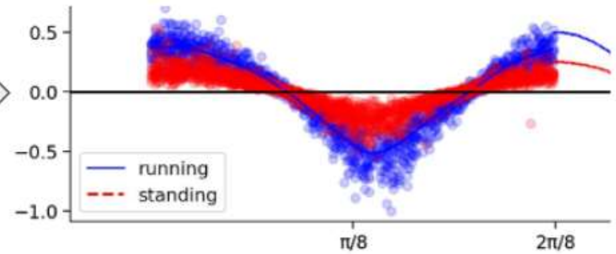
How much does neural activity increase with running speed?

1 Project neural activity on PC#2

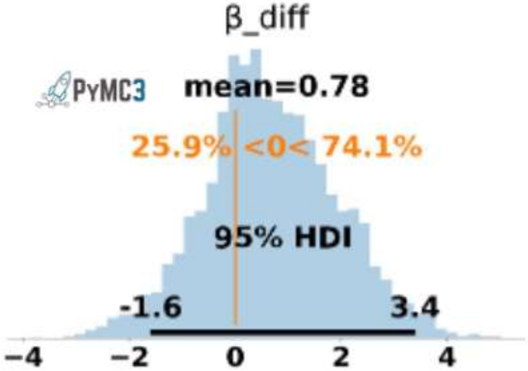
Running trials exhibit higher amplitude!!



2 Fit to a cosine and bin running speed



3 Bayesian modeling



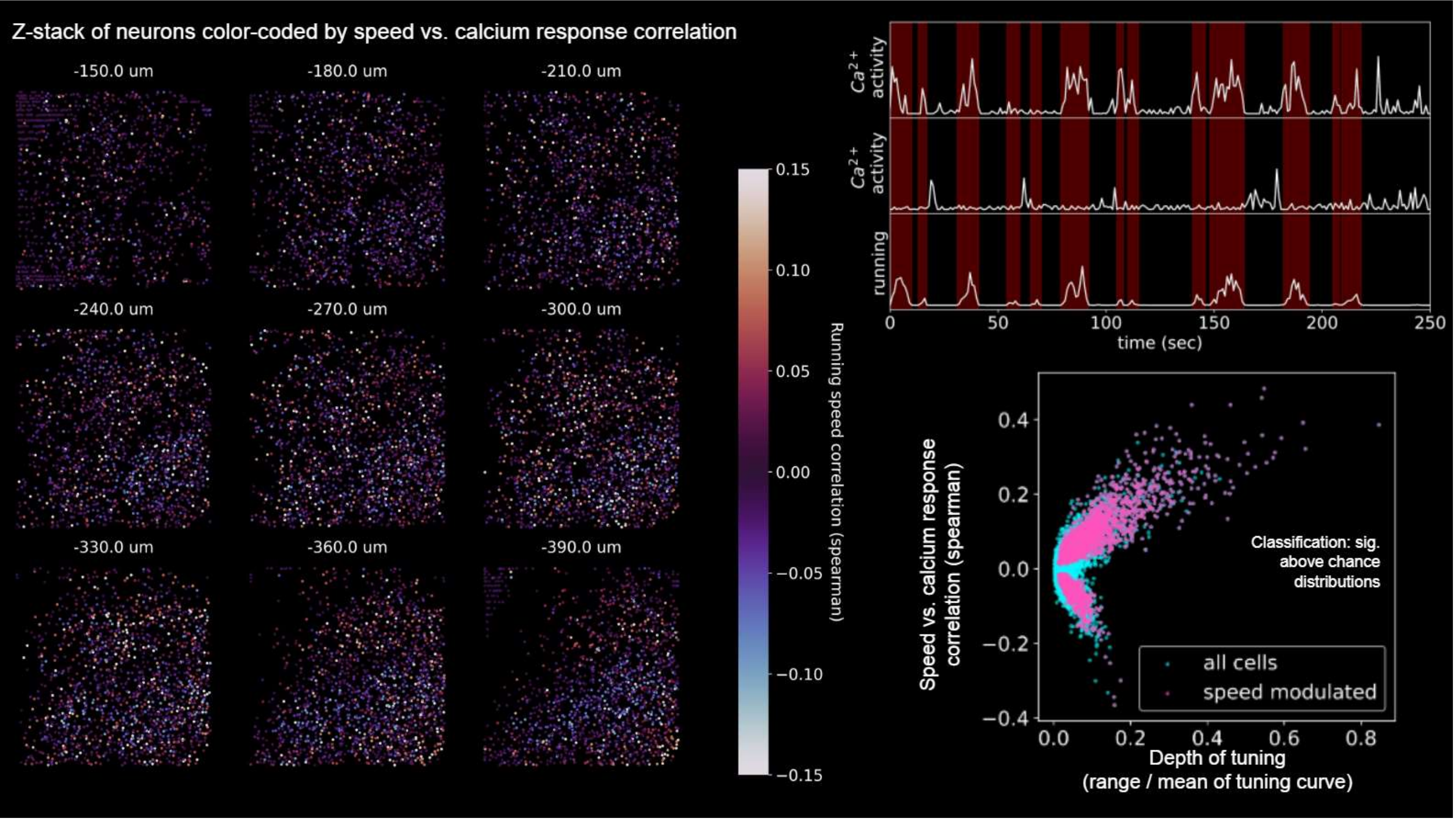
4 Linear Regression neural activity → angle

	R^2		√MSE	
	Train	Test R^2	Train	Test
Standing	.86	.70	38º	55º
Running	.89	.84	33º	42º

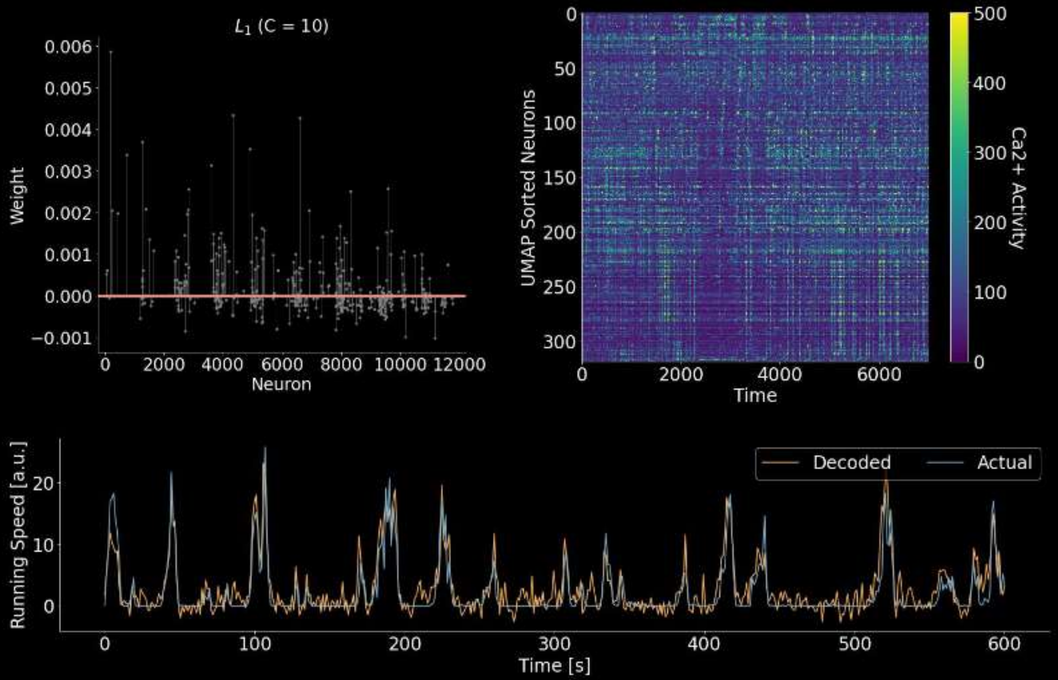
https://colab.research.google.com/drive/1ZFTCpUiYIGpDrpJcVWm2_pe9K0W_hwtv?usp=sharing

Conclusion and prospects

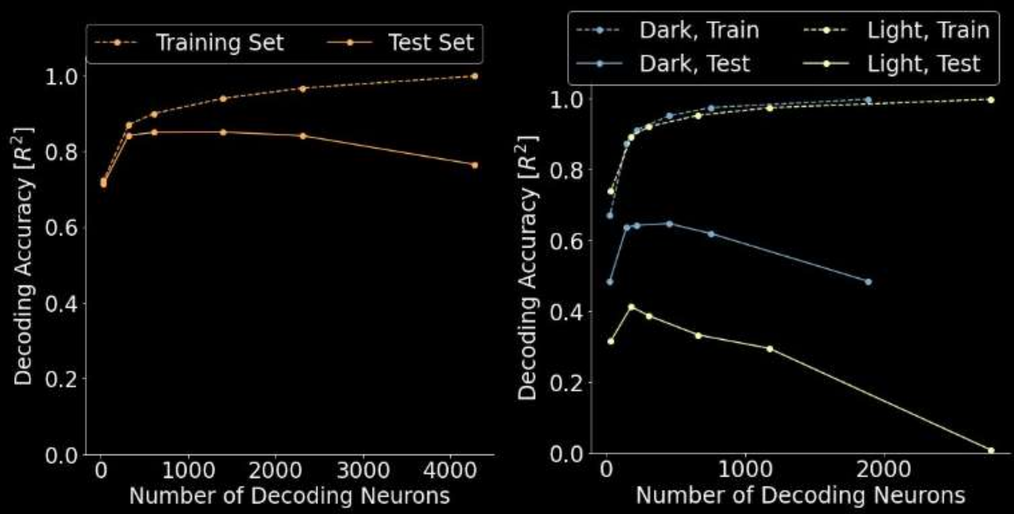
- We can modestly decode angles from neural activity using simple models
- Running speed has a qualitative effect on neural activity spread in PC space
- A Bayesian model was deployed to try to quantify this effect
- A decoder can differentiate between trials where the mouse was moving vs. standing still, but not the speed
- Decoding angles based on running trials compared to standing trials seems to have slightly better accuracy, but this would need to be validated using cross validation


Z-stack of neurons color-coded by speed vs. calcium response correlation
-150.0 um
-180.0 um
-210.0 um
-240.0 um
-270.0 um
-300.0 um
-330.0 um
-360.0 um
-390.0 um
Running speed correlation (spearman)
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
Ca^{2+} activity
Ca^{2+} activity
running
0
50
100
150
200
250
time (sec)
Speed vs. calcium response correlation (spearman)
0.4
0.2
0.0
−0.2
−0.4
0.0
0.2
0.4
0.6
0.8
Depth of tuning
(range / mean of tuning curve)
Classification: sig. above chance distributions
all cells
speed modulated

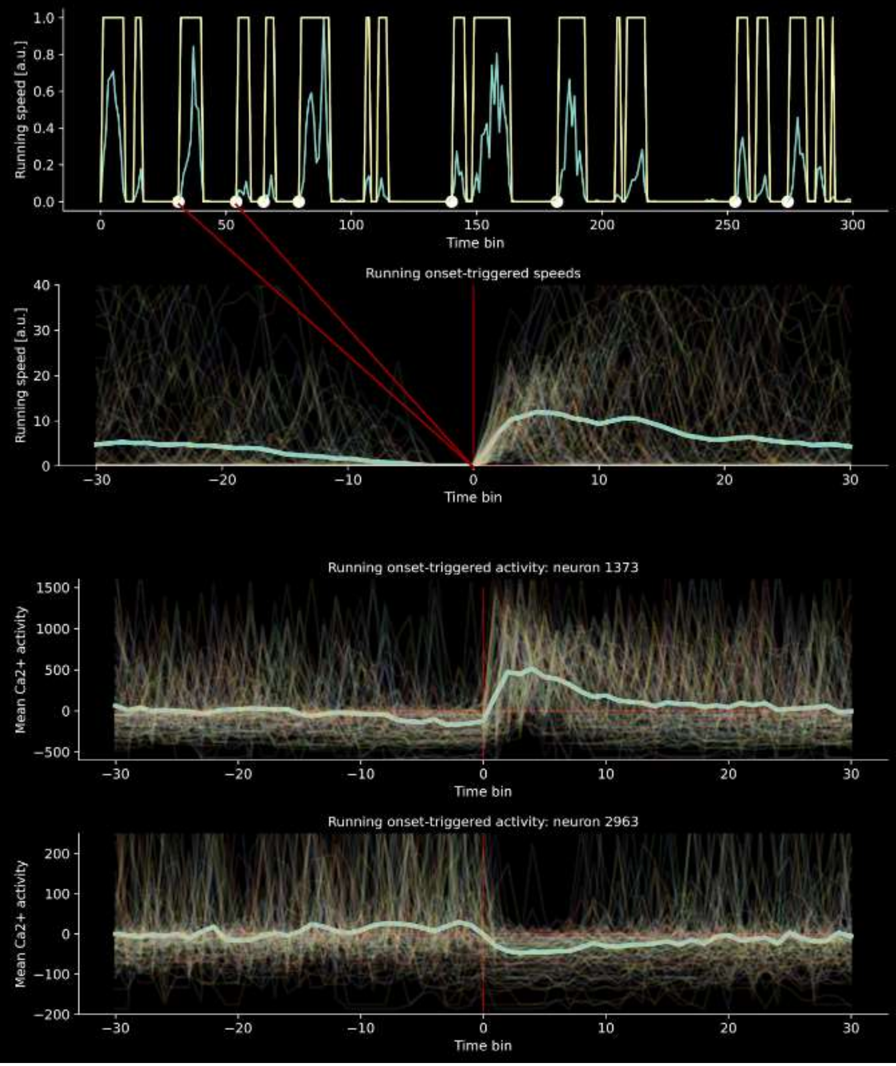
Decoding running speed using sparse neuron population

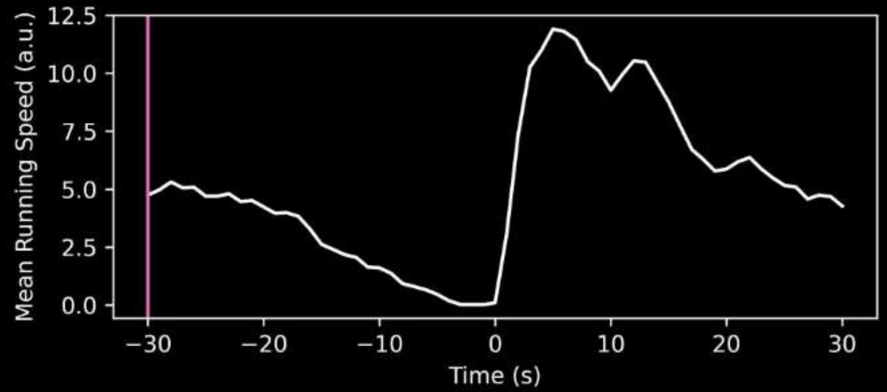


Visual stimulus reduces decoding accuracy of V1

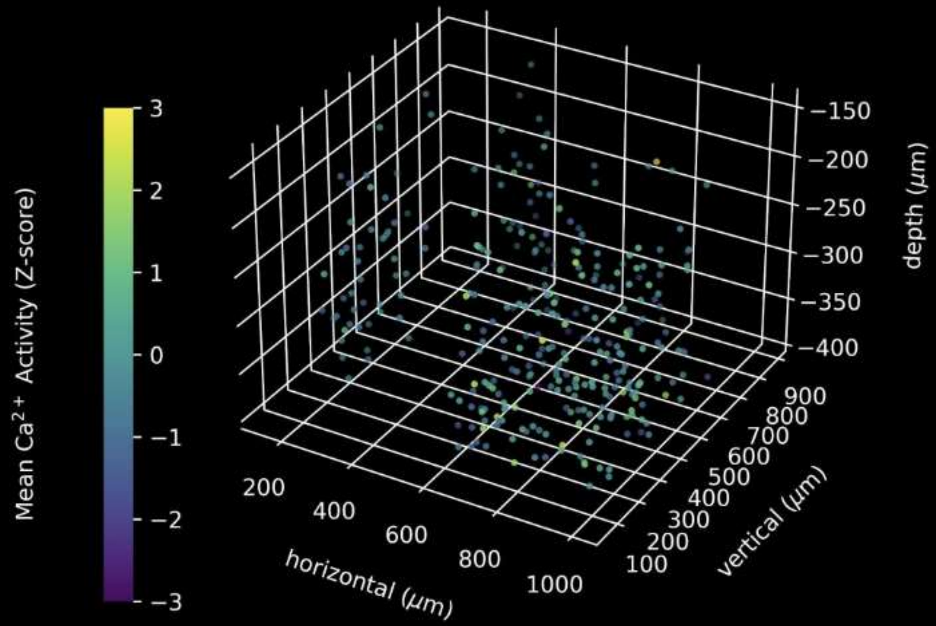


Running onset-triggered calcium activity in positively and negatively tuned neurons





Spatiotemporal dynamics
https://youtu.be/4aWvZXFcbg8

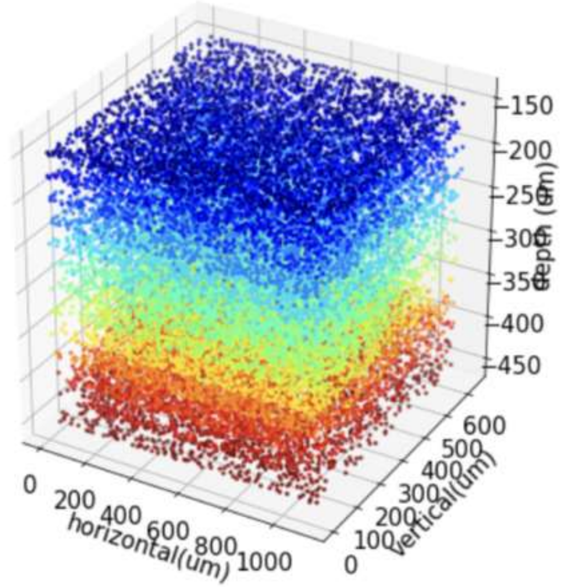


What Question did we study?

How are different classes of neurons spatially distributed across V1? And could spatial distance between different classes neurons explain trial-to-trial variability of sensory information encoding neurons?

Which dataset did we work with?

Stringer dataset: two-photon calcium imaging data of ~20,000 neurons from primary visual cortex of mice. Consisting of two parts: during presentation of visual orientation stimulus, during sitting in complete darkness.



How we tried to answer… aka the struggle

We did some literature search on the subject matter.

Tried running analysis software that came with the dataset (rastermap and suite2p).

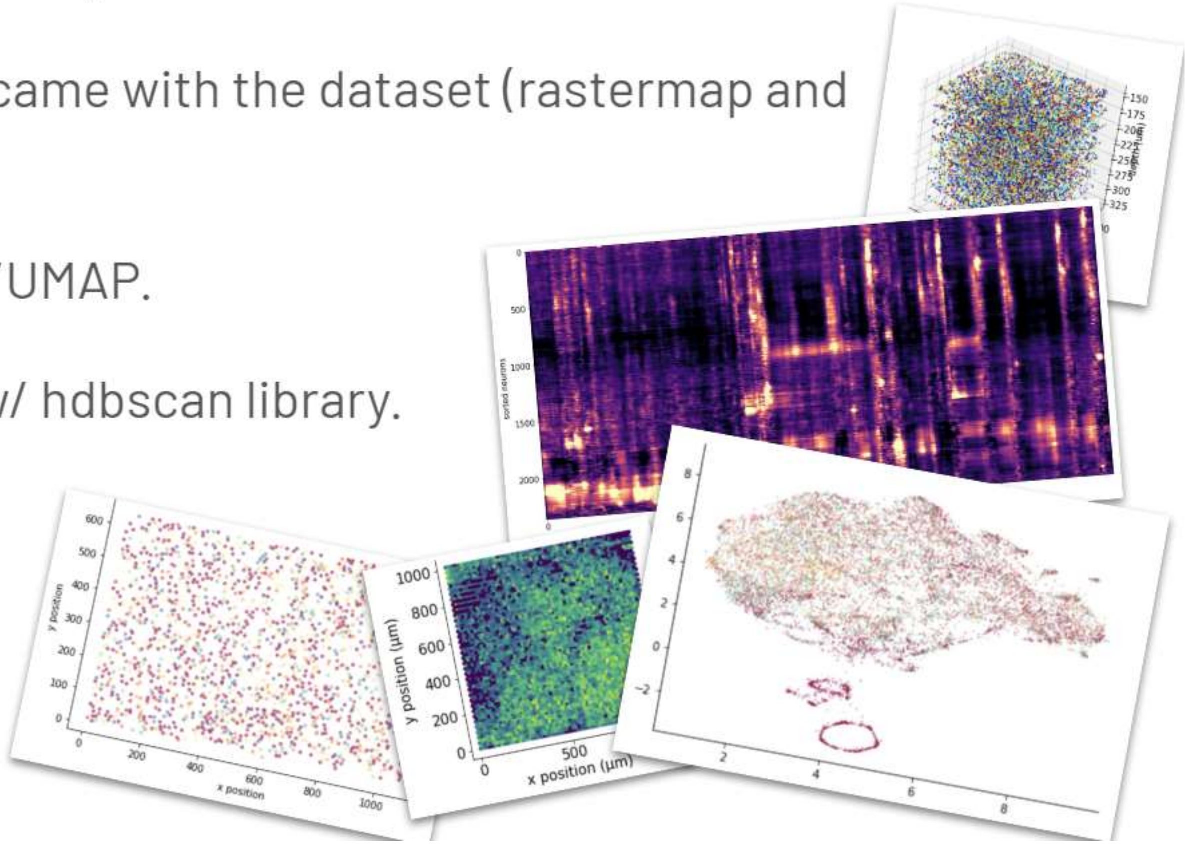
1~3D manifold embedding using PCA/UMAP.

Clustering through K-means, UMAP w/ hdbscan library.

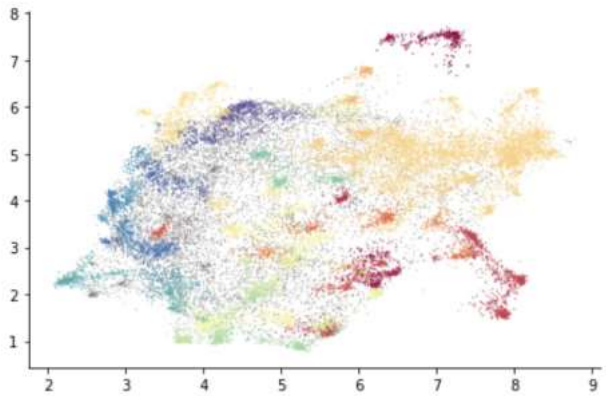
Plotting heatmaps of neurons.

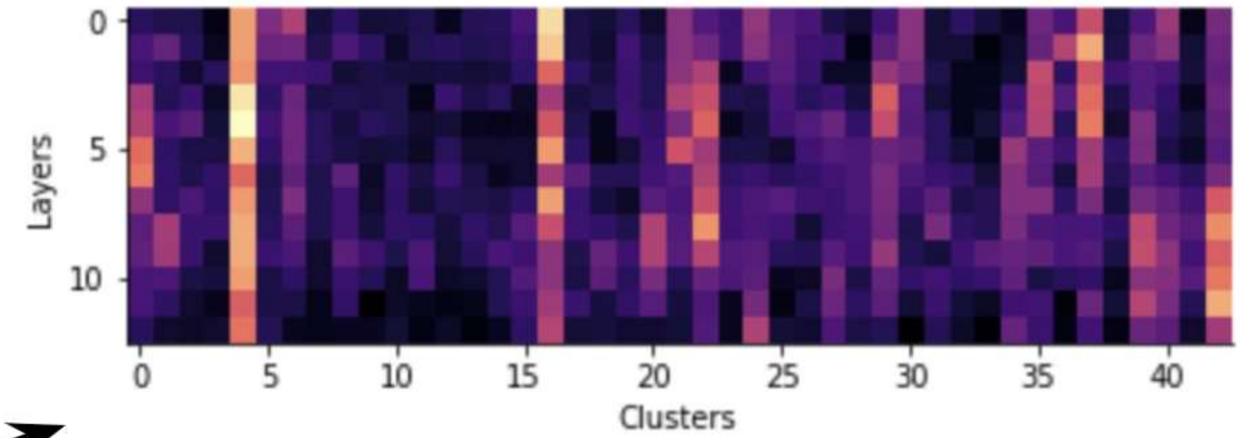
Labeling clustering in 2D.

⋮



Results? and more...





Take-away: Some implications that z-position of neurons matter...

We will continue to work on it after NMA under our mentor's guidance!