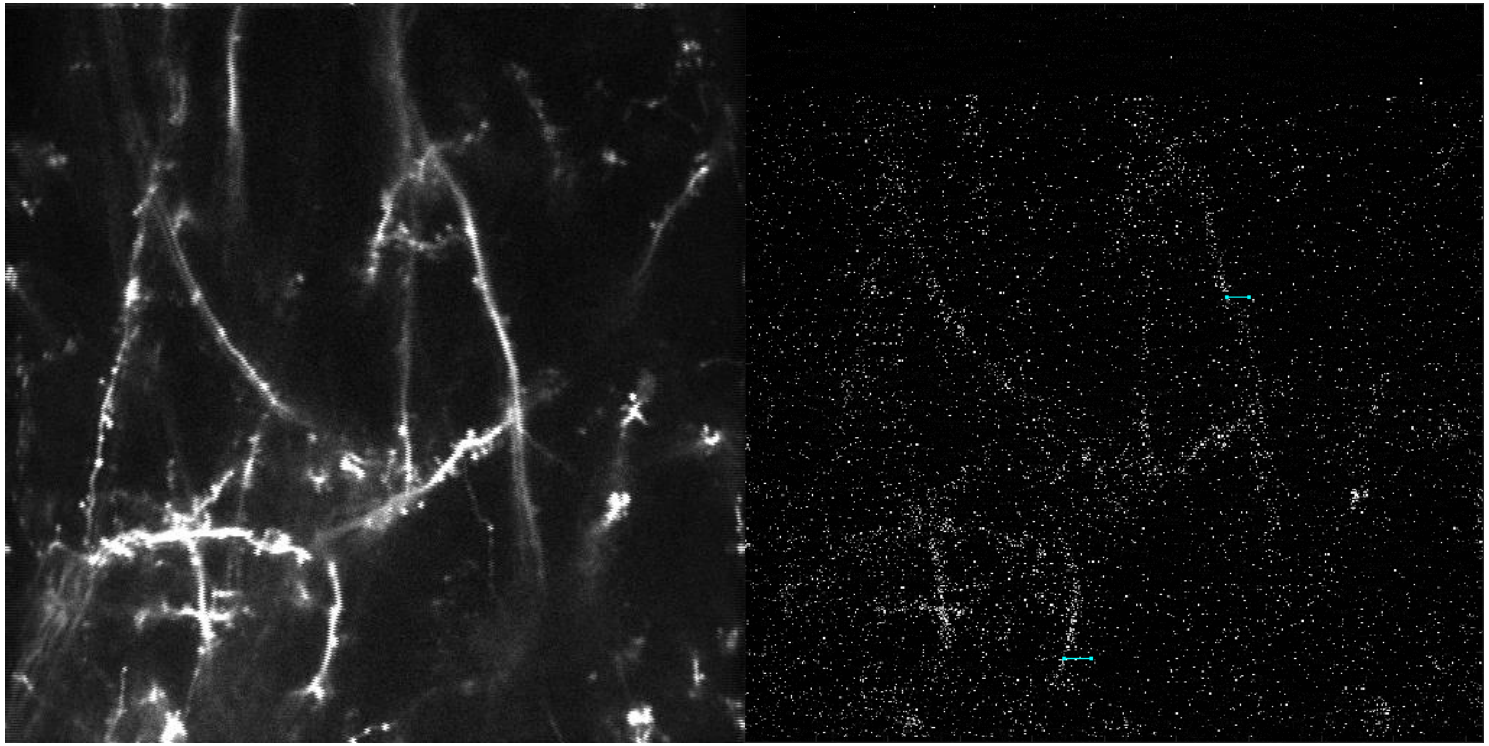
Analysing your imaging data

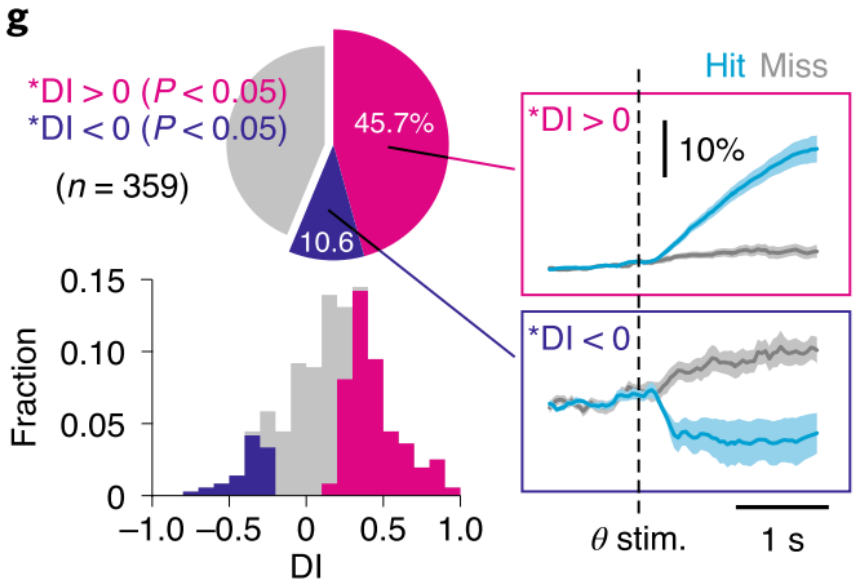
Australian Course in Advanced Neuroscience

5th July 2023

George Stuyt

From raw video to descriptive data





eLIFE elifesciences.org TOOLS AND RESOURCES

CaImAn an open source tool for scalable calcium imaging data analysis

Andrea Giovannucci[1*], Johannes Friedrich[1,2,3], Pat Gunn[1], Jérémie Kalfon[4†], Brandon L Brown[5], Sue Ann Koay[6], Jiannis Taxidis[7], Farzaneh Najafi[8], Jeffrey L Gauthier[6], Pengcheng Zhou[2,3], Baljit S Khakh[5,9], David W Tank[6], Dmitri B Chklovskii[1], Eftychios A Pnevmatikakis[1*]

ARTICLES https://doi.org/10.1038/s41592-020-01018-x nature methods

Check for updates

Cellpose: a generalist algorithm for cellular segmentation

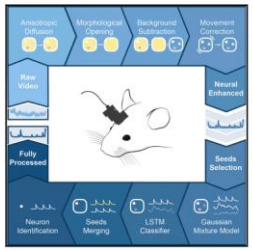
Carsen Stringer, Tim Wang, Michalis Michaelos and Marius Pachitariu

Resource

Cell Reports

MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline

Graphical Abstract



Authors
Jinghao Lu, Chunyuan Li, Jonnathan Singh-Alvarado, Zhe Charles Zhou, Flavio Fröhlich, Richard Mooney, Fan Wang

Correspondence
jinghao.lu@duke.edu (J.L.), chunyuan.li@duke.edu (C.L.), fan.wang@duke.edu (F.W.)

In Brief

Lu et al. develop an end-to-end pipeline (MIN1PIPE) for automatic processing of single-photon calcium imaging data that uses different algorithms for movement correction and signal extraction and outperforms existing methods.

ELSEVIER

Journal of Neuroscience Methods

Volume 291, 1 November 2017, Pages 83-94

NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data

Eftychios A. Pnevmatikakis, Andrea Giovannucci

Neuron

NeuroResource

CellPress

Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data

Eftychios A. Pnevmatikakis,[1,2,*] Daniel Soudry,[2] Yuanjun Gao,[2] Timothy A. Machado,[2,3,4,5] Josh Merel,[2,4] David Pfau,[2,4] Thomas Reardon,[3,4,5] Yu Mu,[6] Clay Lacefield,[4] Weijian Yang,[7] Misha Ahrens,[6] Randy Bruno,[4] Thomas M. Jessell,[3,4,5] Darcy S. Peterka,[5,7] Rafael Yuste,[4,7] and Liam Paninski[2,4,5,7,*]

OnACID: Online Analysis of Calcium Imaging Data in Real Time

Andrea Giovannucci†[1] Johannes Friedrich†*[1] Matthew Kaufman‡

Anne K. Churchland‡ Dmitri Chklovskii† Liam Paninski*

Eftychios A. Pnevmatikakis†[2]

suite2p

docs passing | build passing | coverage 67% | pypi package 0.11.1 | downloads 251k | downloads/month 3k | python 3 | license GPL-3.0 | contributors 28 | website up | repo size 98.8 MB | Stars 242 | Forks 183

Pipeline for processing two-photon calcium imaging data.
Copyright (C) 2018 Howard Hughes Medical Institute Janelia Research Campus

suite2p includes the following modules:

- Registration
- Cell detection
- Spike detection
- Visualization GUI

frontiers in NEUROINFORMATICS

METHODS ARTICLE
published: 23 September 2014
doi: 10.3389/fninf.2014.00080

SIMA: Python software for analysis of dynamic fluorescence imaging data

*Patrick Kaifosh *, Jeffrey D. Zaremba, Nathan B. Danielson and Attila Losonczy*

Department of Neuroscience, Columbia University in the City of New York, New York, NY, USA

eLife TOOLS AND RESOURCES

Minian, an open-source miniscope analysis pipeline

Zhe Dong[1], William Mau[1], Yu Feng[1], Zachary T Pennington[1], Lingxuan Chen[1], Yosif Zaki[1], Kanaka Rajan[1], Tristan Shuman[1], Daniel Aharoni[2]*, Denise J Cai[1]*

Pre-processing vs analysis

Pre-processing

- The “pre” analysis
- Data extraction
- Technical

Analysis

- Comparisons
- Statistics
- Inference

Imaging pipelines are like beer

The well known

Public release, easy, polished

The boutique

Public release

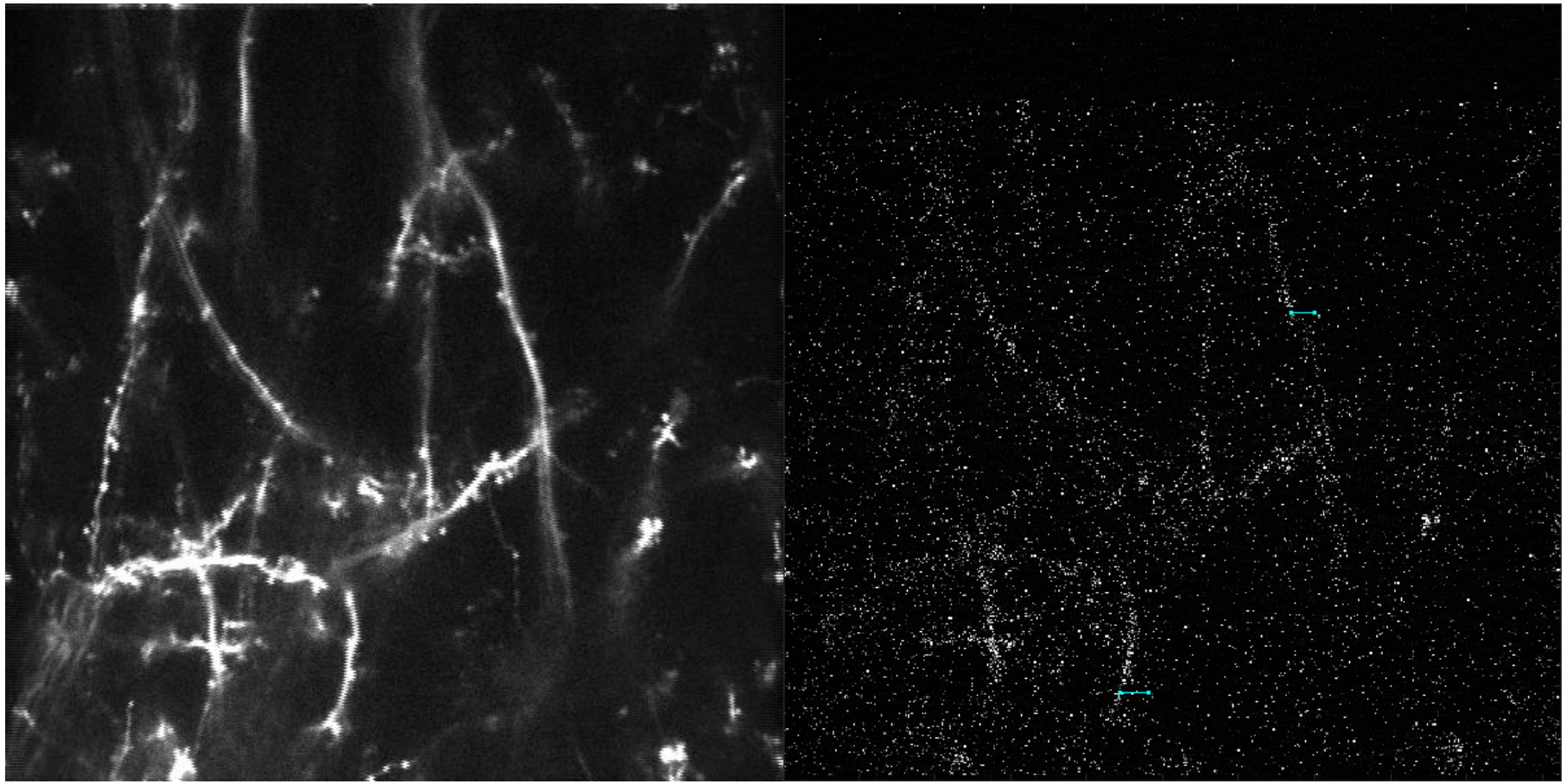
The homebrew

Precise, programming required

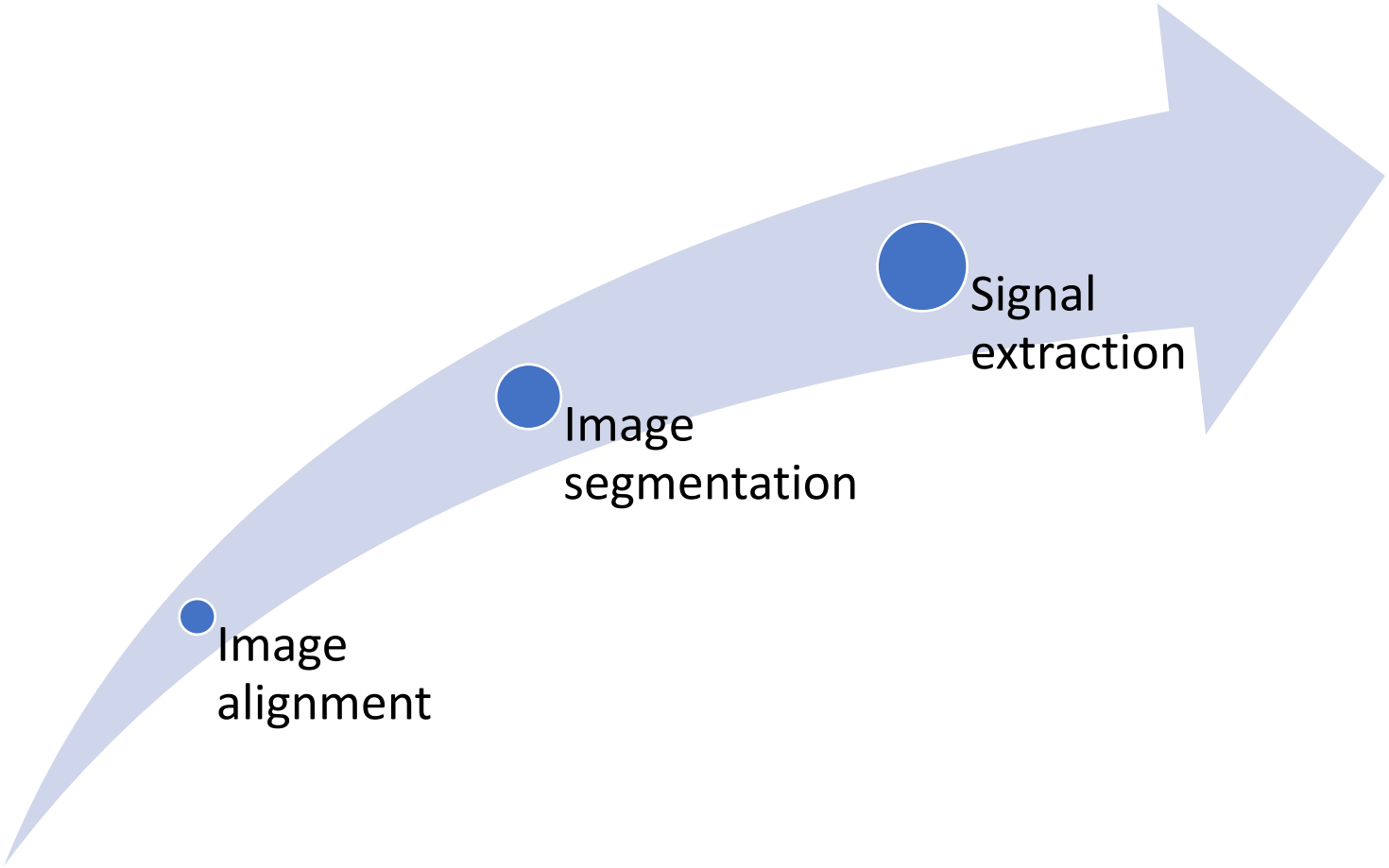
Agenda

- Pre-processing
  - Motion correcting a video
  - Classify regions of video as distinct objects
  - Extract fluorescence signals from regions
  - Normalise fluorescence signals
- Other important technical considerations
- Quantification/Analysis
  - Mean fluorescence
  - Event detection
  - Deconvolution
- Hands on example
  - DFF calculation
  - ~~Interactive Google Colab~~

The goal of pre-processing: data extraction



Three key steps to pre-processing



Alignment

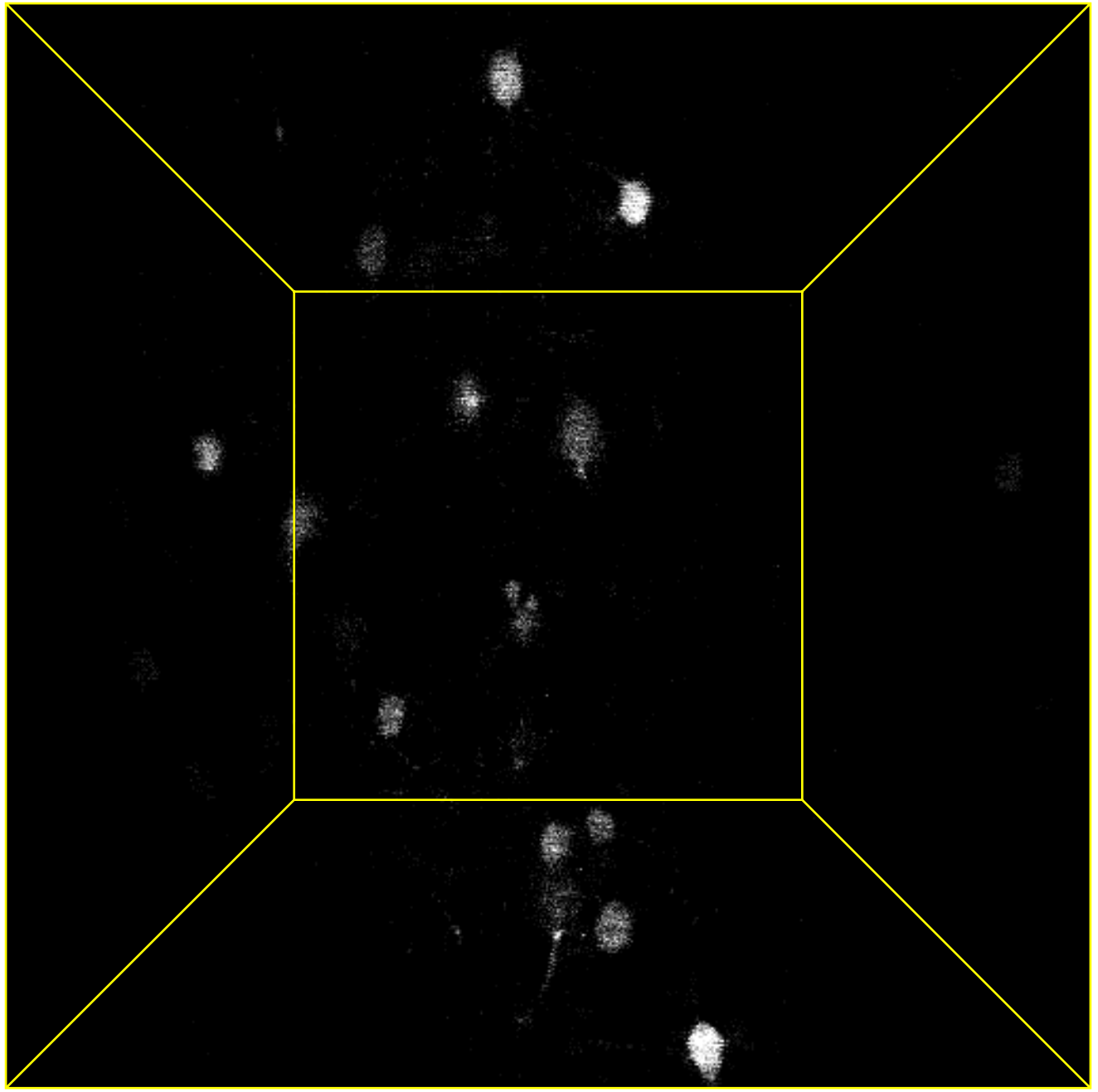
Removing movement from a video
Also called “registration”, “motion correction”

Movement is a problem

- Mouse movement
  - Arm movement while grooming
  - Licking
  - Heartbeat
  - Weight shifting
- Rig movement
  - Bumping table
  - Door closing (!)
- Water ingestion causing brain expansion

Preventing movement should happen at the prep

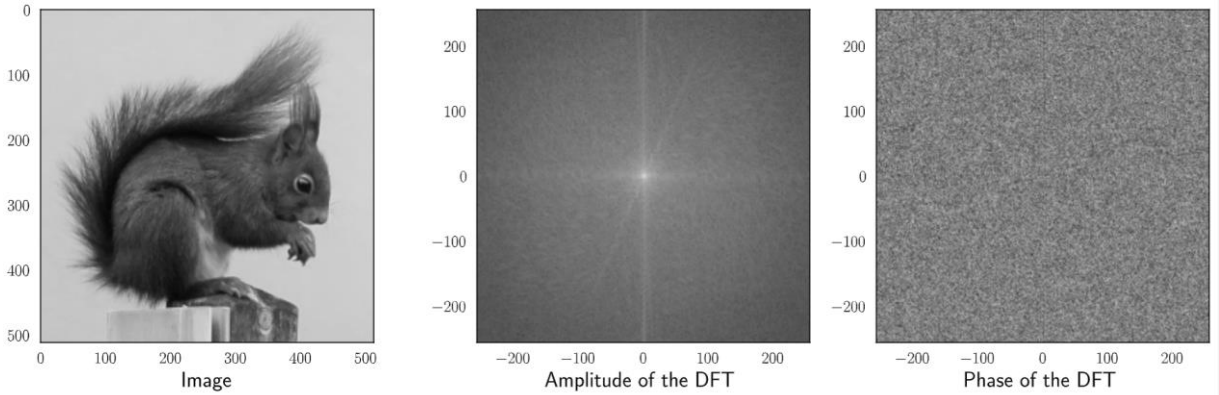
- Ex vivo
  - Sample mounting on plate
- In vivo
  - Robust headfix setup
  - Pressure on brain from coverslip



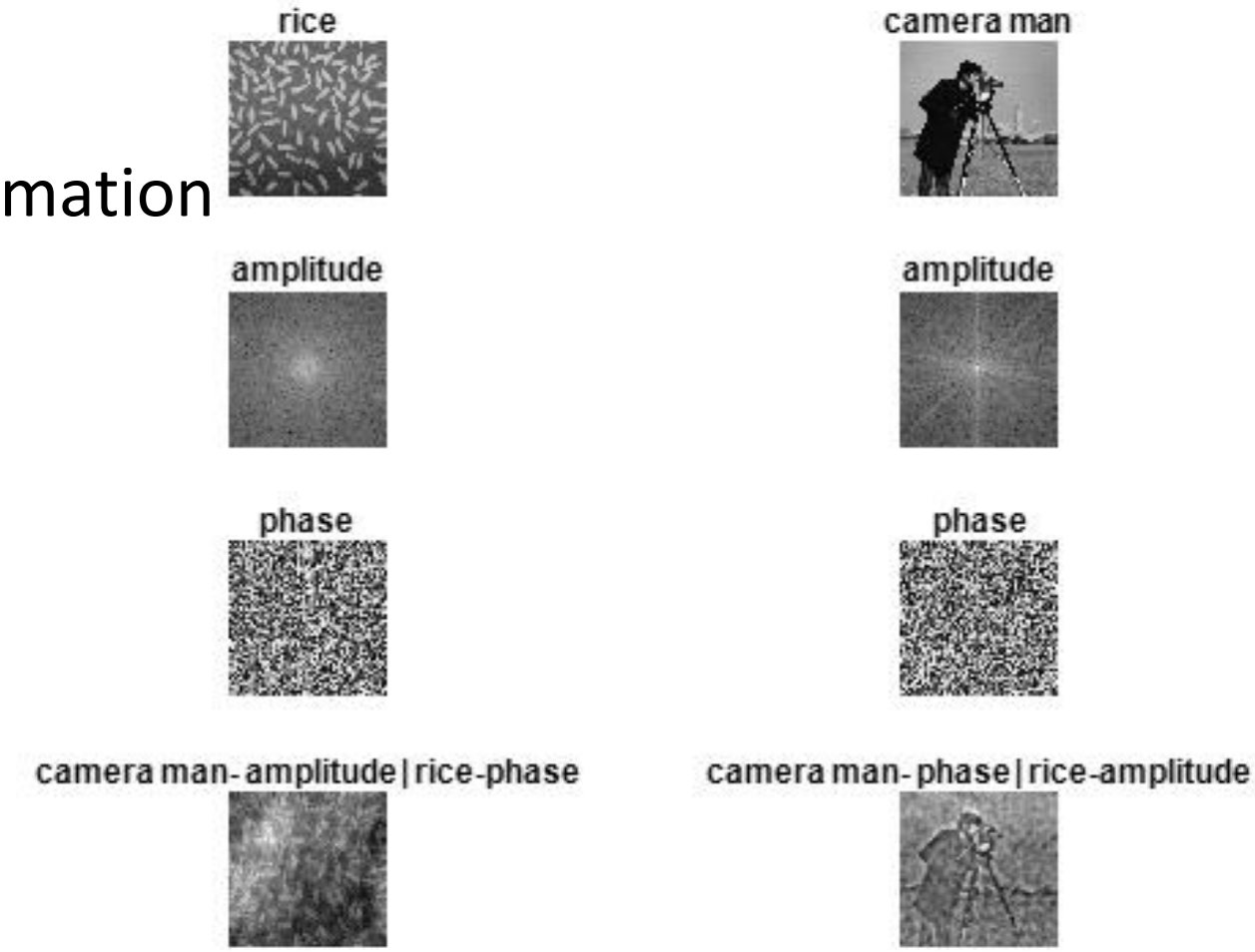
Test acquisition I did for a new rig

Phase/cross correlation

- Images can be Fourier transformed
- Amplitude = spatial/geometric information
- Phase = location information

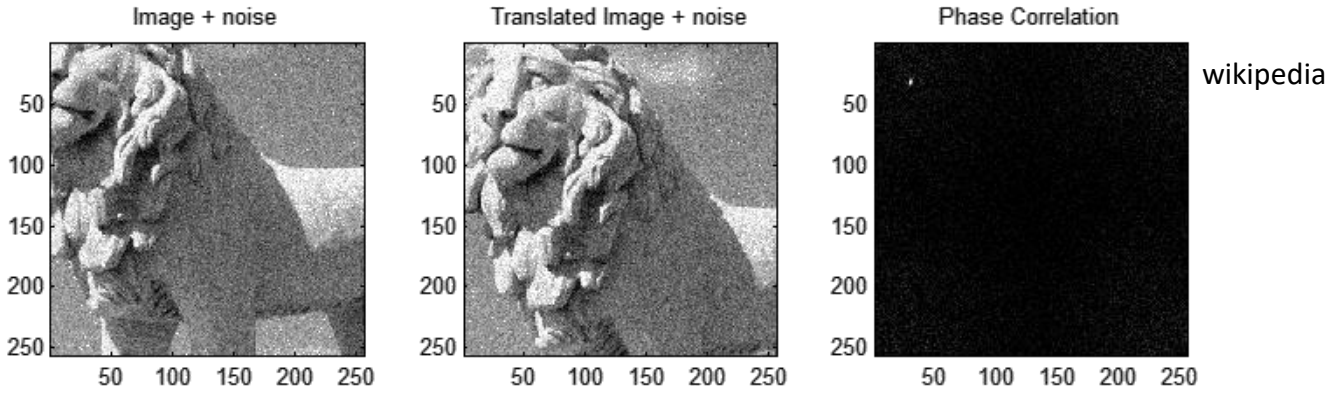


https://vincmazet.github.io/bip/filtering/fourier.html



https://dsp.stackexchange.com/a/72751

Image alignment process



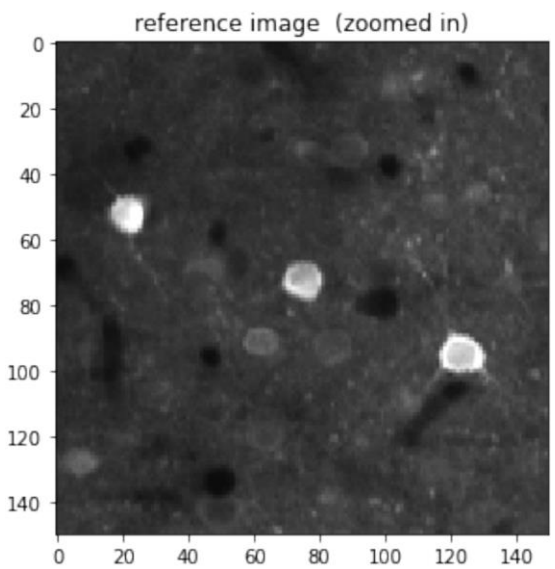
wikipedia

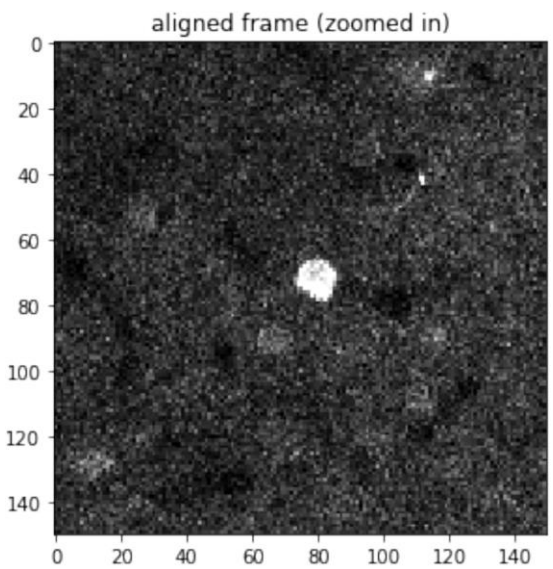
Raw image → Fourier transformed image → Correlation matrix between fourier'd raw and a fourier'd template → x-y position of maximum correlation

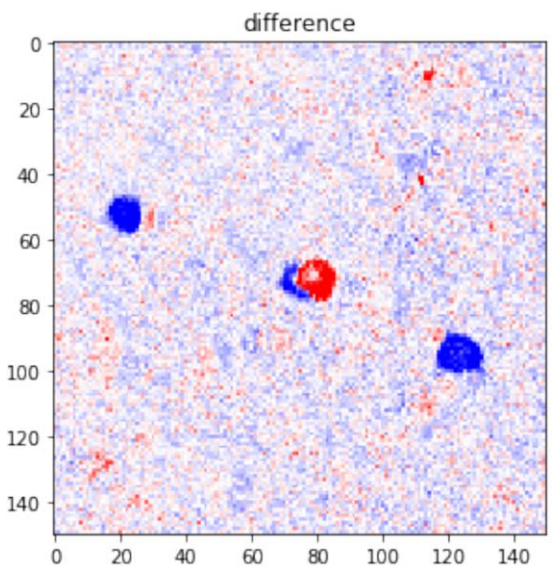
- Loop through all frames in a video
- Essentially the same thing as applying every possible shift to the image and correlating that to the template image
- Computationally expensive-ish

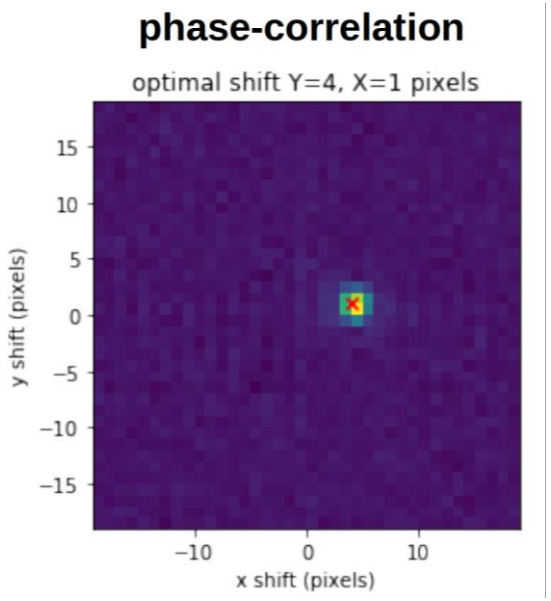
Alignment applied to calcium imaging

- Raw frame is moved by the x-y coordinates as determined in the correlation matrix





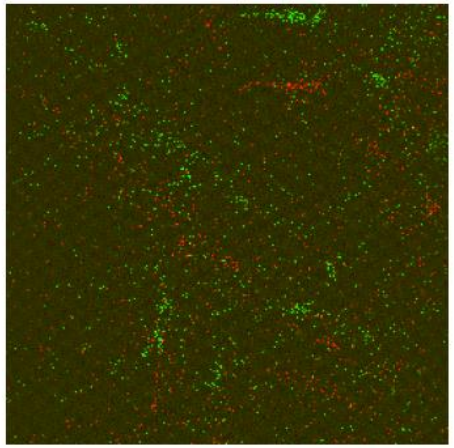




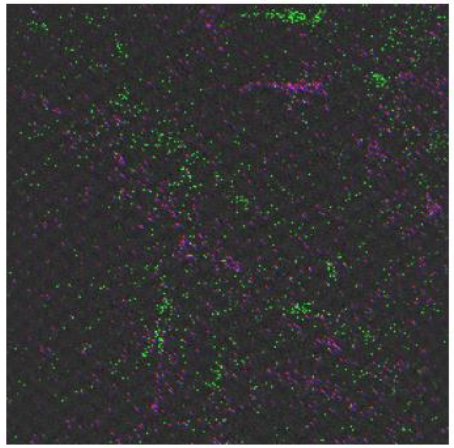
suite2p.readthedocs.io

An example of alignment failure: Restriction of maximum alignment shift

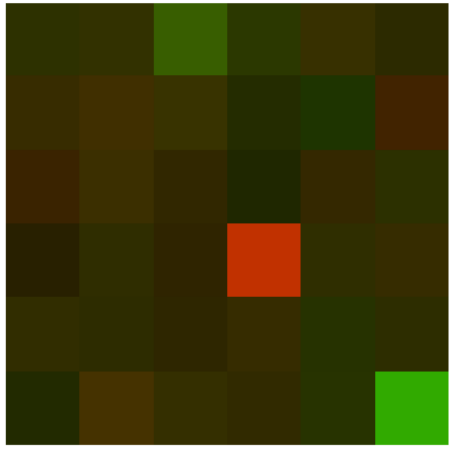
Template (green)
Frame (red)



Template (green)
Frame (red)
New Frame (blue)

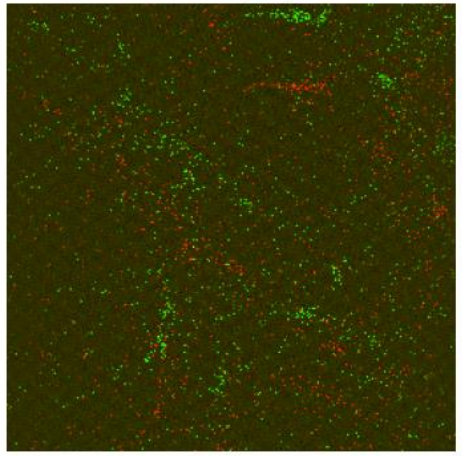


What phase correlation sees

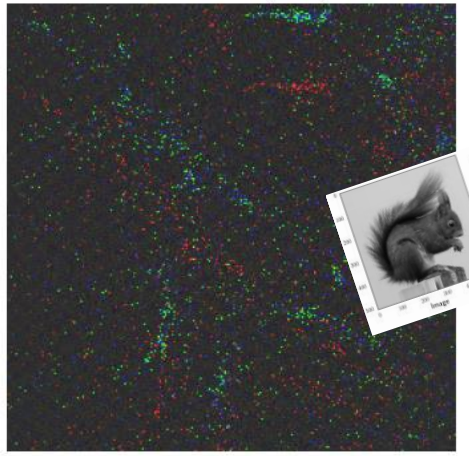


Correlation matrix output

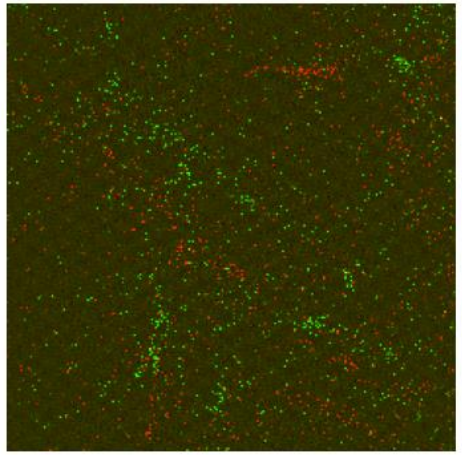
Template (green)
Frame (red)



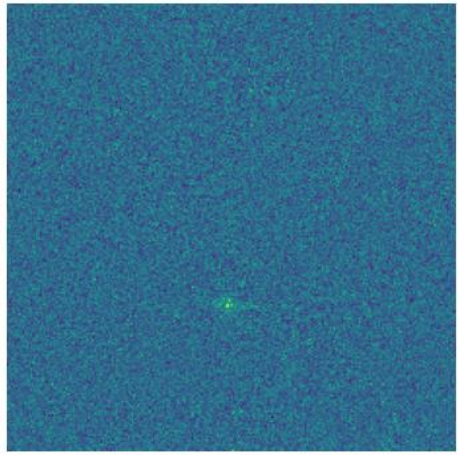
Template (green)
Frame (red)
New Frame (blue)



What phase correlation sees

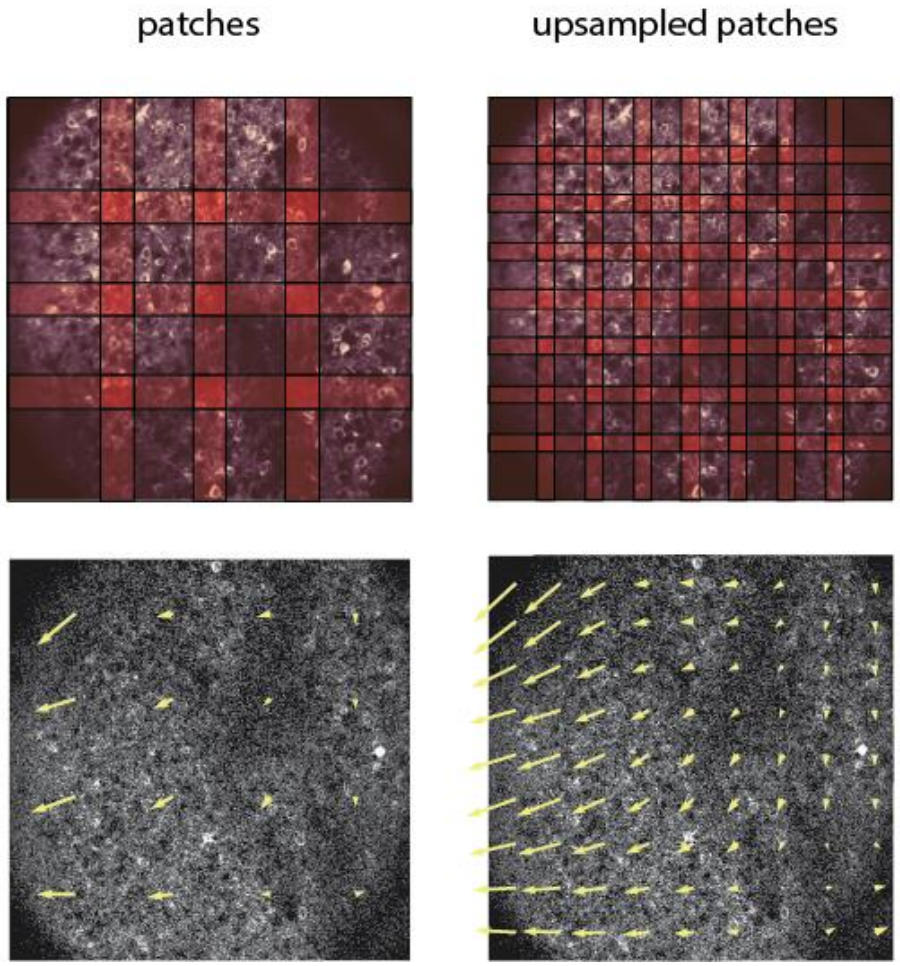


Correlation matrix output



Non-rigid registration

- Larger-scale brain deformation over time
- Computationally more expensive
- Lower-zoom increases chance you'll need to use non-rigid



https://github.com/flatironinstitute/NoRMCorre

The output

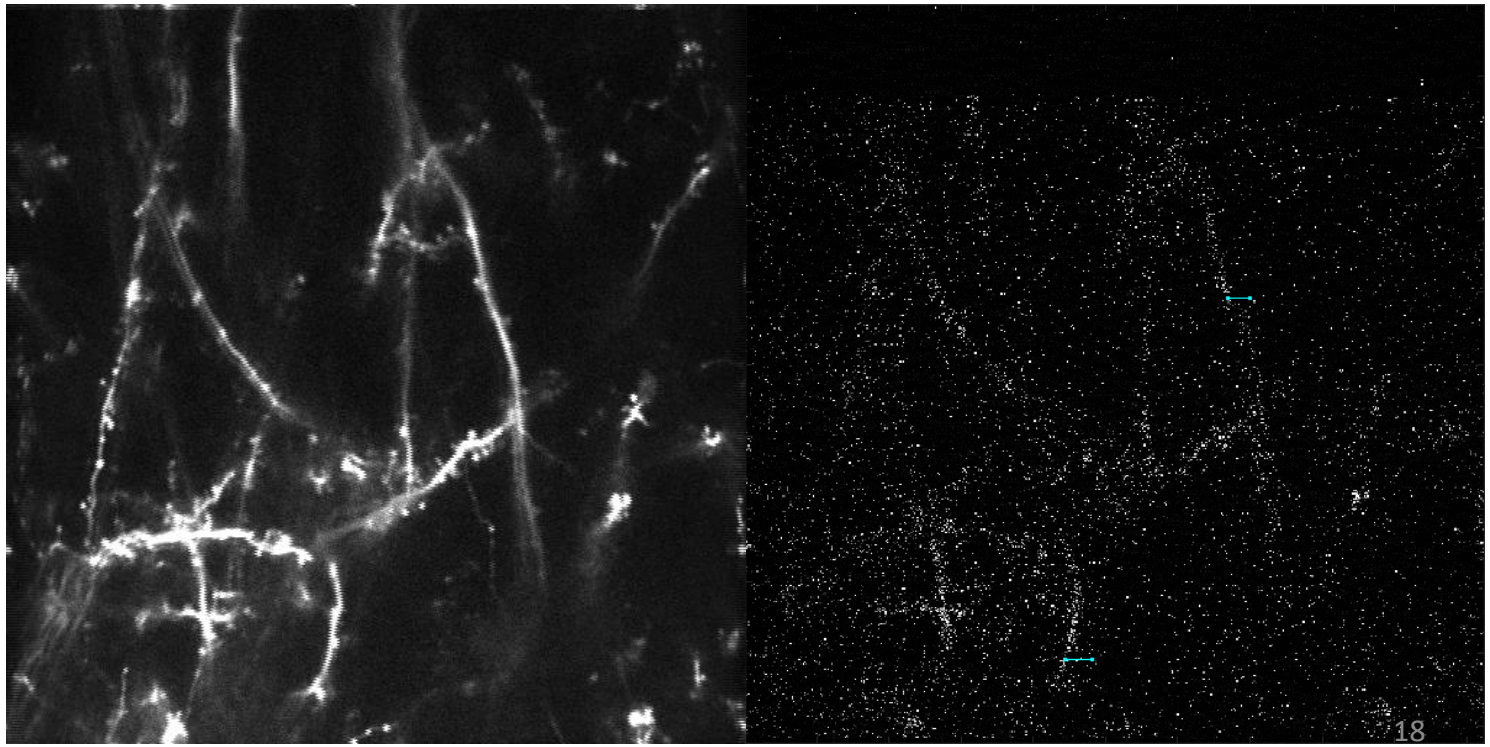

Raw videos (multiple .tiffs)
Alignment process
Aligned videos

Segmentation

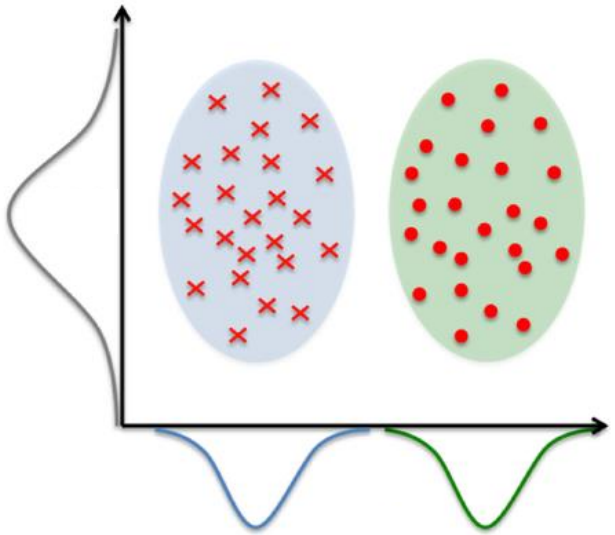
Also called “classification” or even “spatial footprint identification”

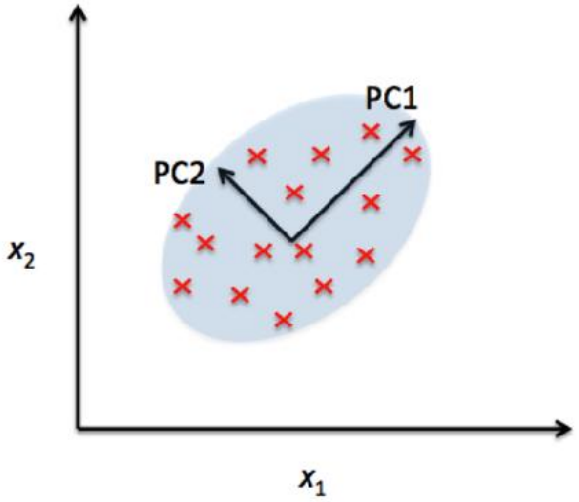
Region of interest (ROI)

- We can see objects fluorescing, so how do we *define* them?
- One method was/is to hand draw



Building an intuition for an analytical workhorse: dimensionality reduction





Adapted from Raj 2019

Non-negative matrix factorisation

- Factorise a data matrix into a feature matrix and a weight matrix

$$\underbrace{X(:,j)}_{j\text{th facial image}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\text{facial features}} \underbrace{H(k,j)}_{\substack{\text{importance of features} \\ \text{in } j\text{th image}}} = \underbrace{WH(:,j)}_{\substack{\text{approximation} \\ \text{of } j\text{th image}}} .$$

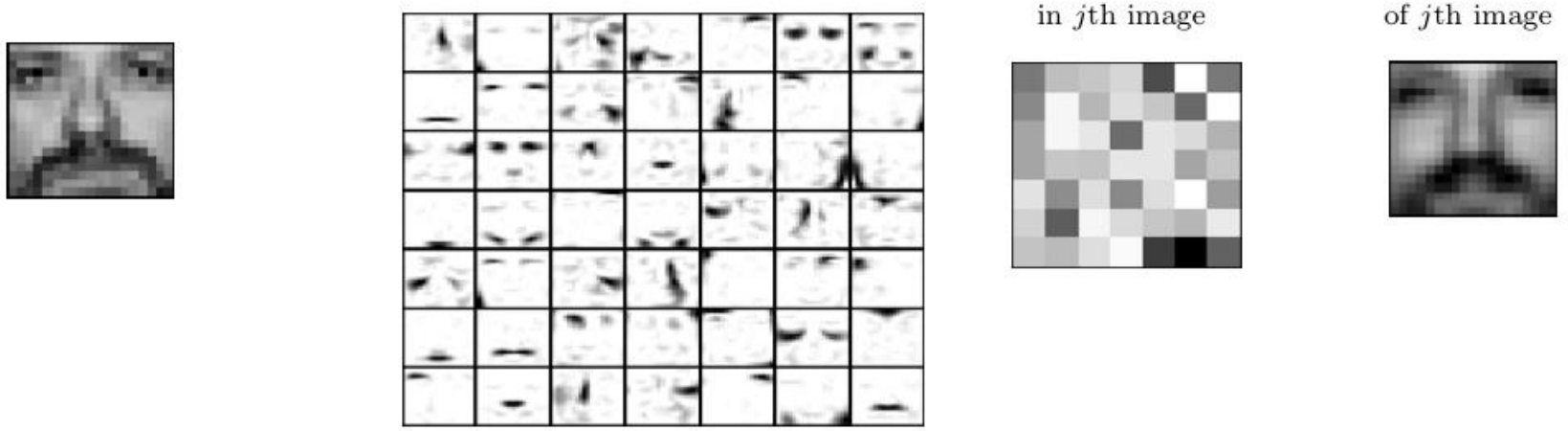
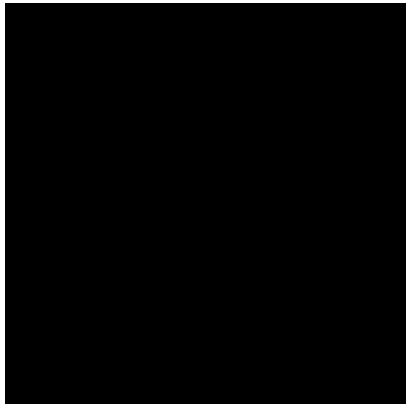


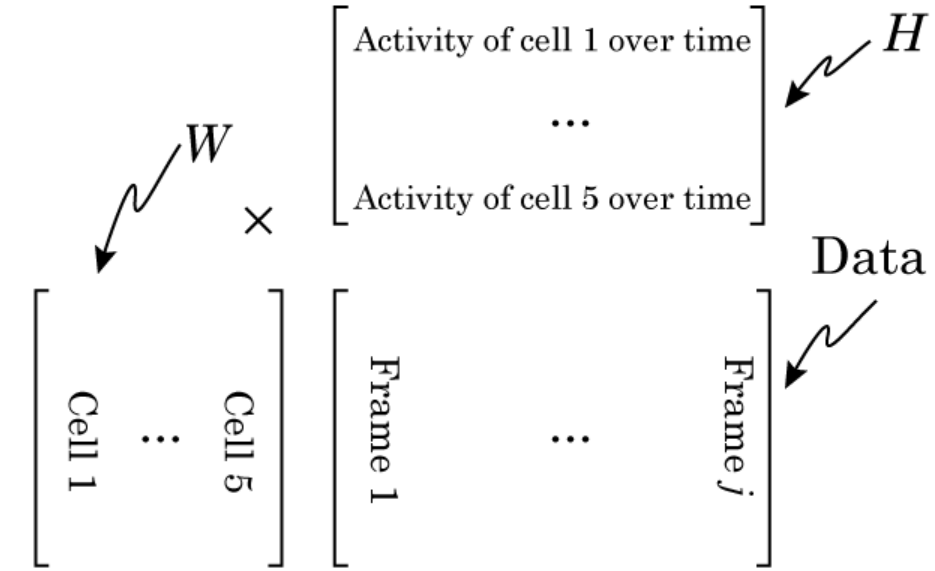
Figure 1: Decomposition of the CBCL face database, MIT Center For Biological and Computation Learning (2429 gray-level 19-by-19 pixels images) using $r = 49$ as in [79].

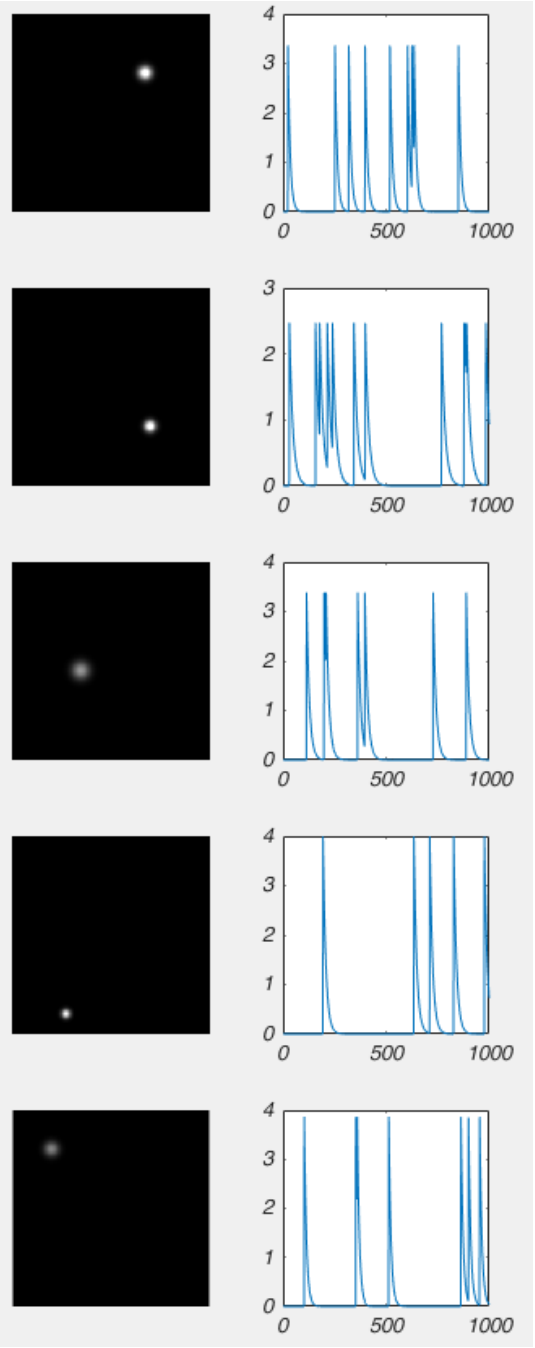
Lee & Seung (*Nature,* 1999)

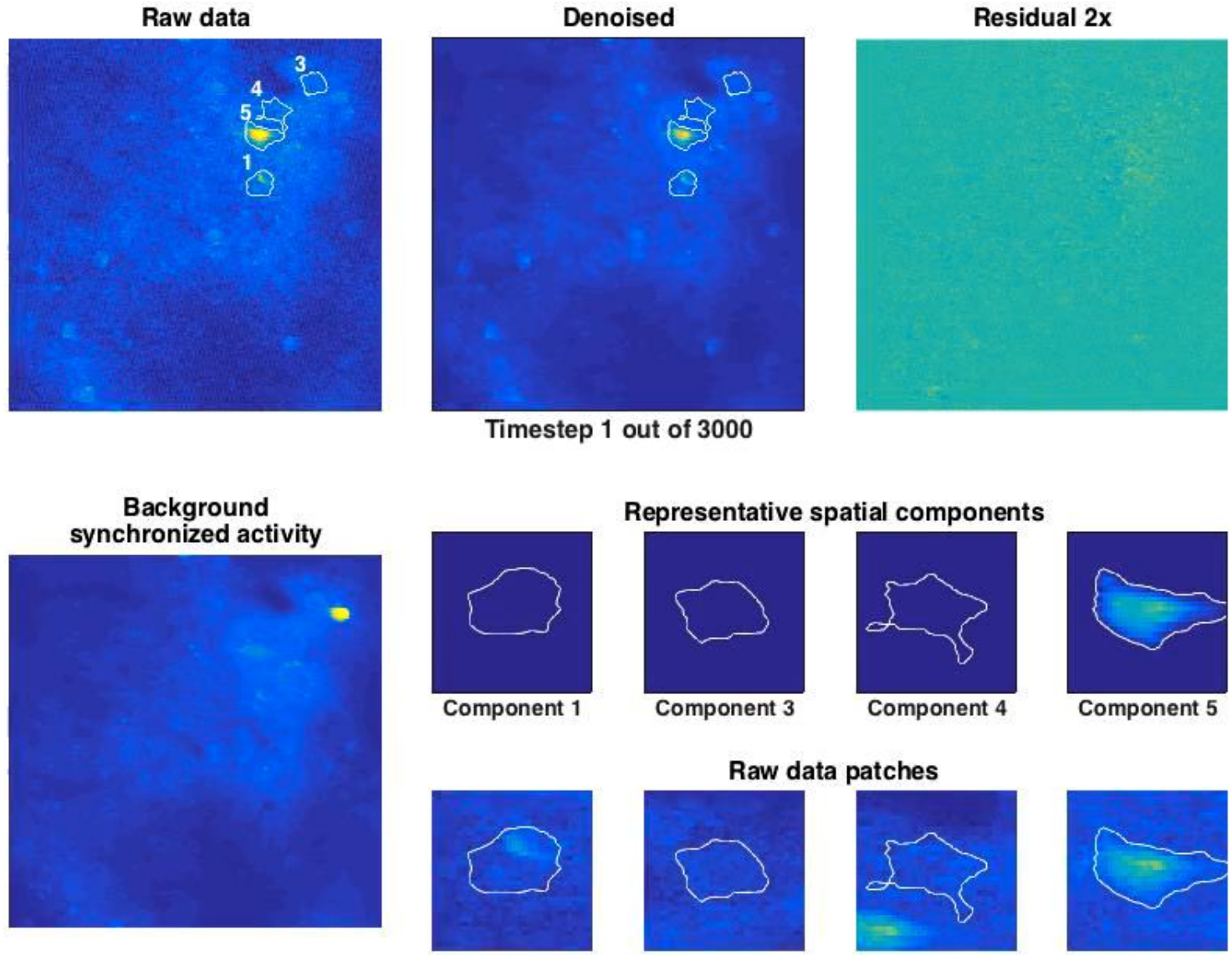
Bill Connelly's example of NNMF

- http://www.billconnelly.net/?p=534









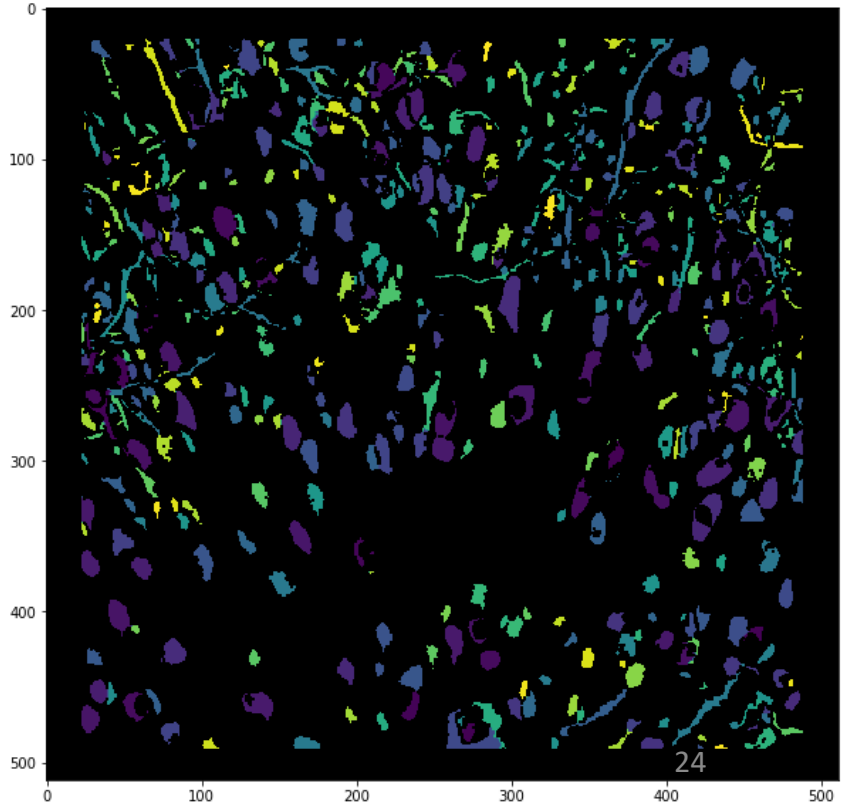
Pnevmatikakis et al. (*Neuron* 2016)

Parameter tuning

- Automated segmentation methods all involve parameters to be tuned to your experiment
- Number of iterations to run
- Assumed spatial extent of the objects
- Thresholds for correlation/activity amount

Overall output

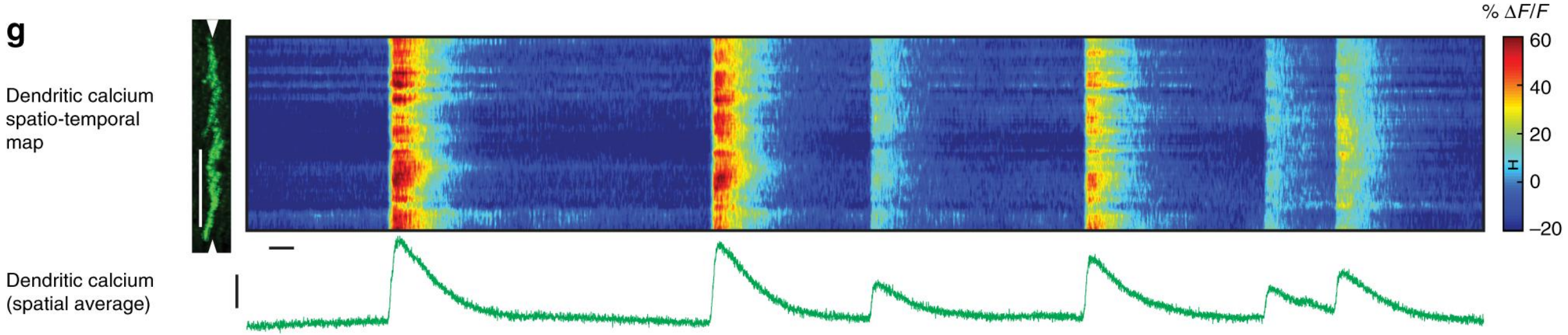
- Identity of which pixel belongs to which ROI
- Or something else for more complex methods
- Probably requires some manual sorting for real vs junk



Fluorescence extraction

A fluorescence trace

- Consider how changing the ROI changes the average



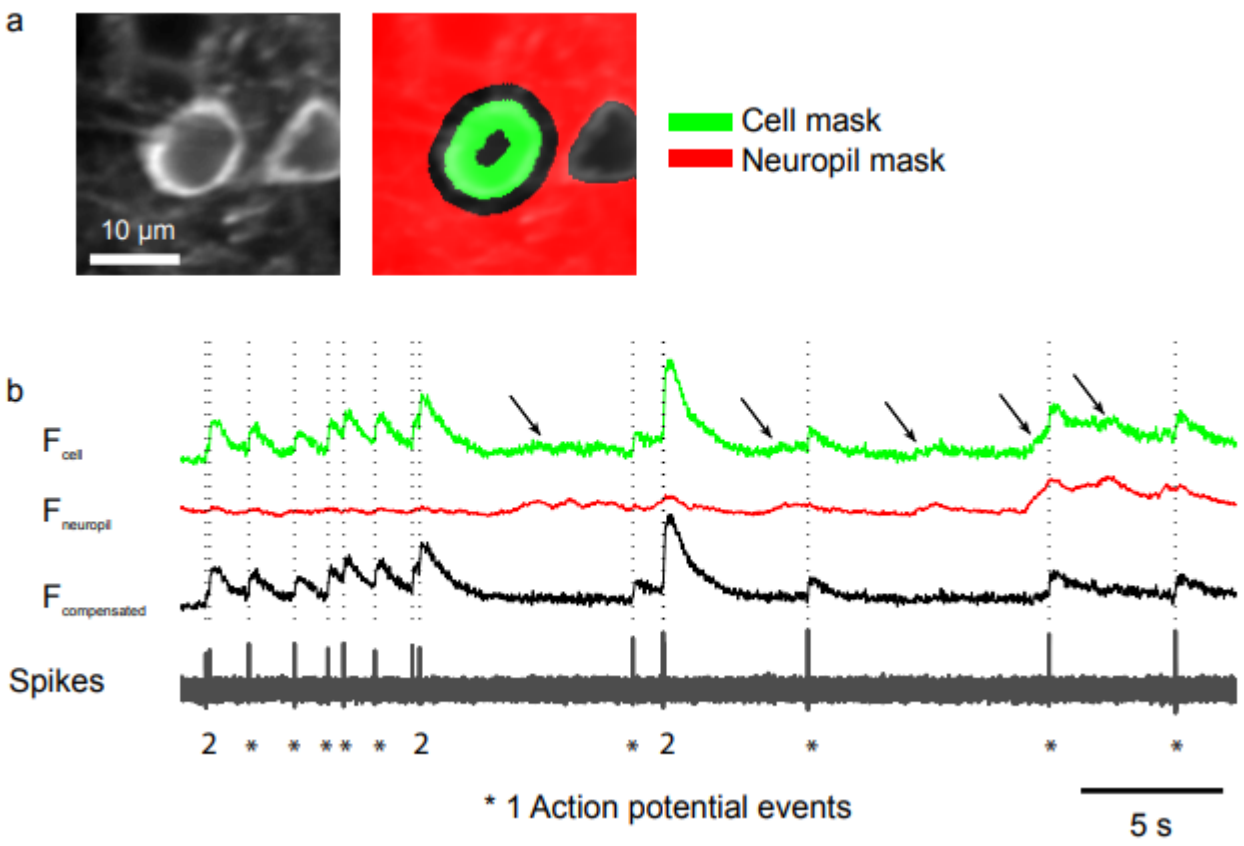
Roome & Kuhn, *Nature* Comms (2018)

Source of contamination/artefact

- Neurons overlapping each other
- *neuropil* : a dense network of interwoven nerve fibres and their branches and synapses, together with glial filaments.
- Background fluorescence
- Brain movement
- Noise

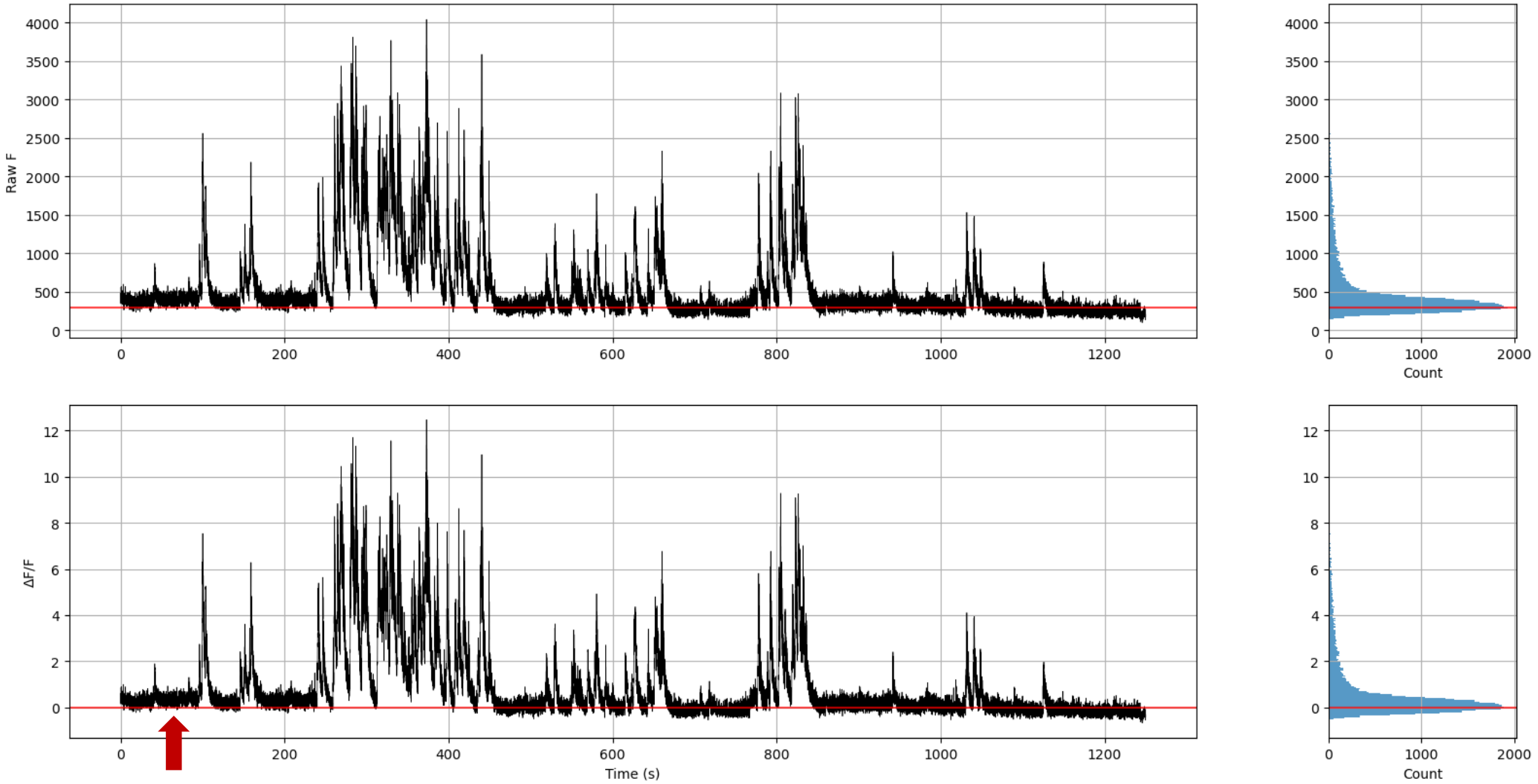
Neuropil: a segmentation addendum

- Finding neuropil fluorescence requires an ROI i.e. segmentation

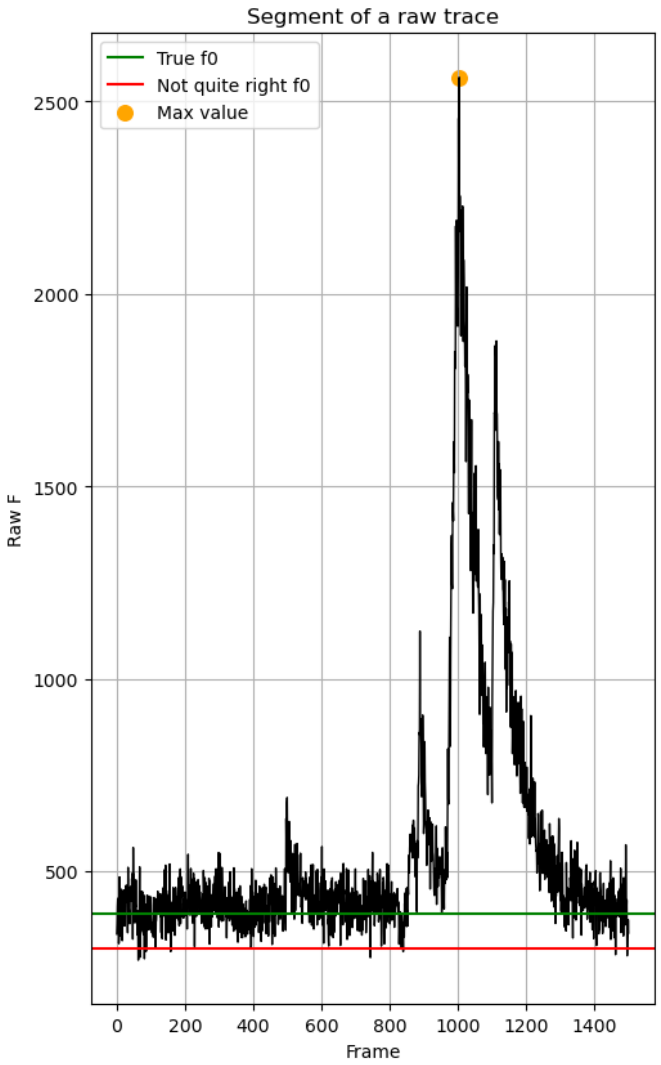


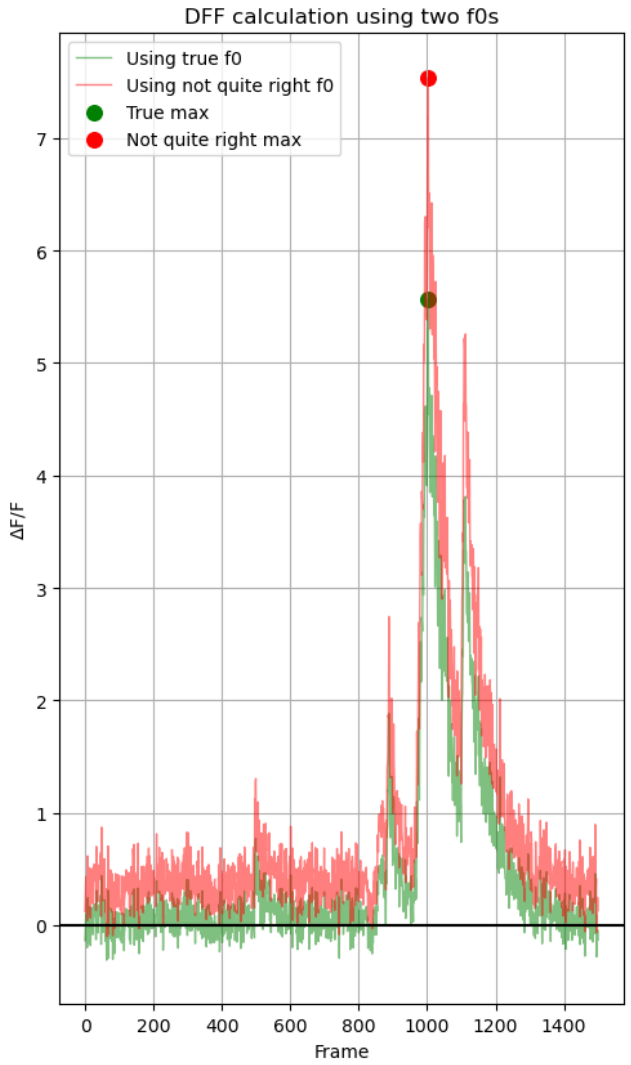
Chen … Kim, *Nature* 2016

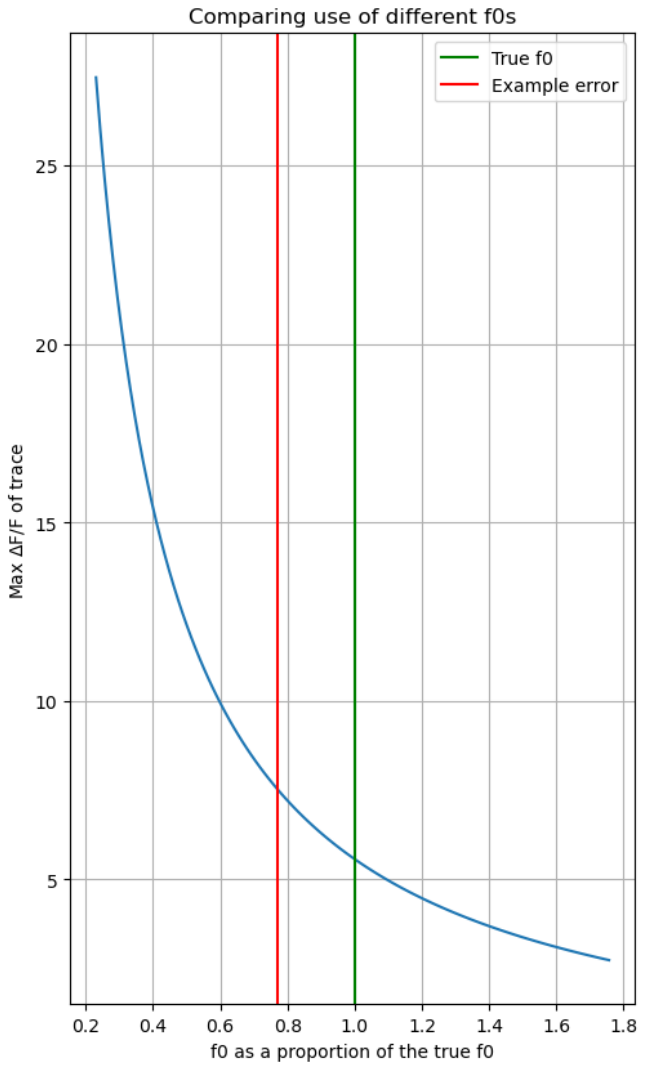
Fluorescence normalisation: $\mathbf{DFF} = \frac{\Delta F}{F} = \frac{F - F_0}{F_0}$



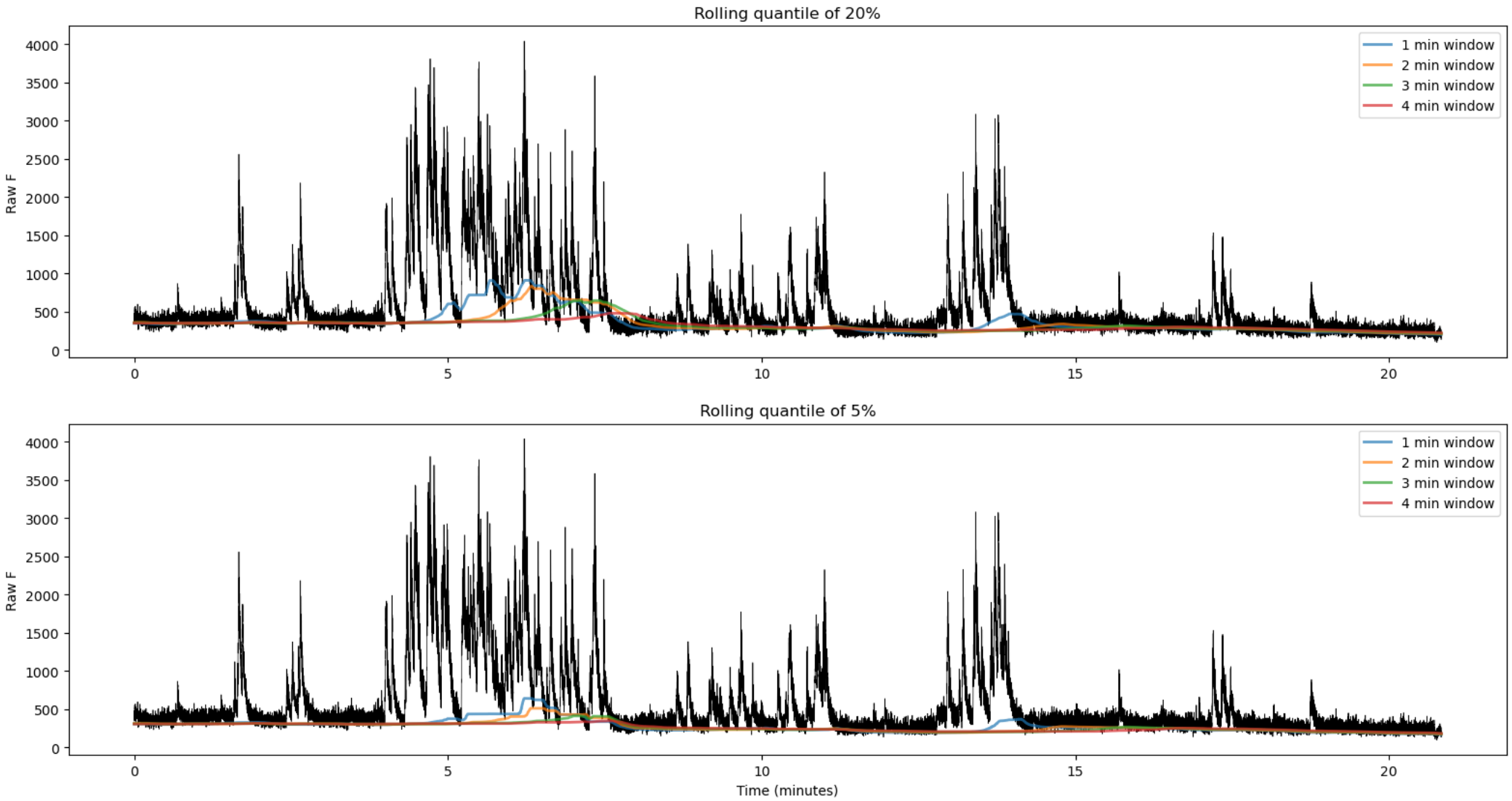
Miscalculating F0



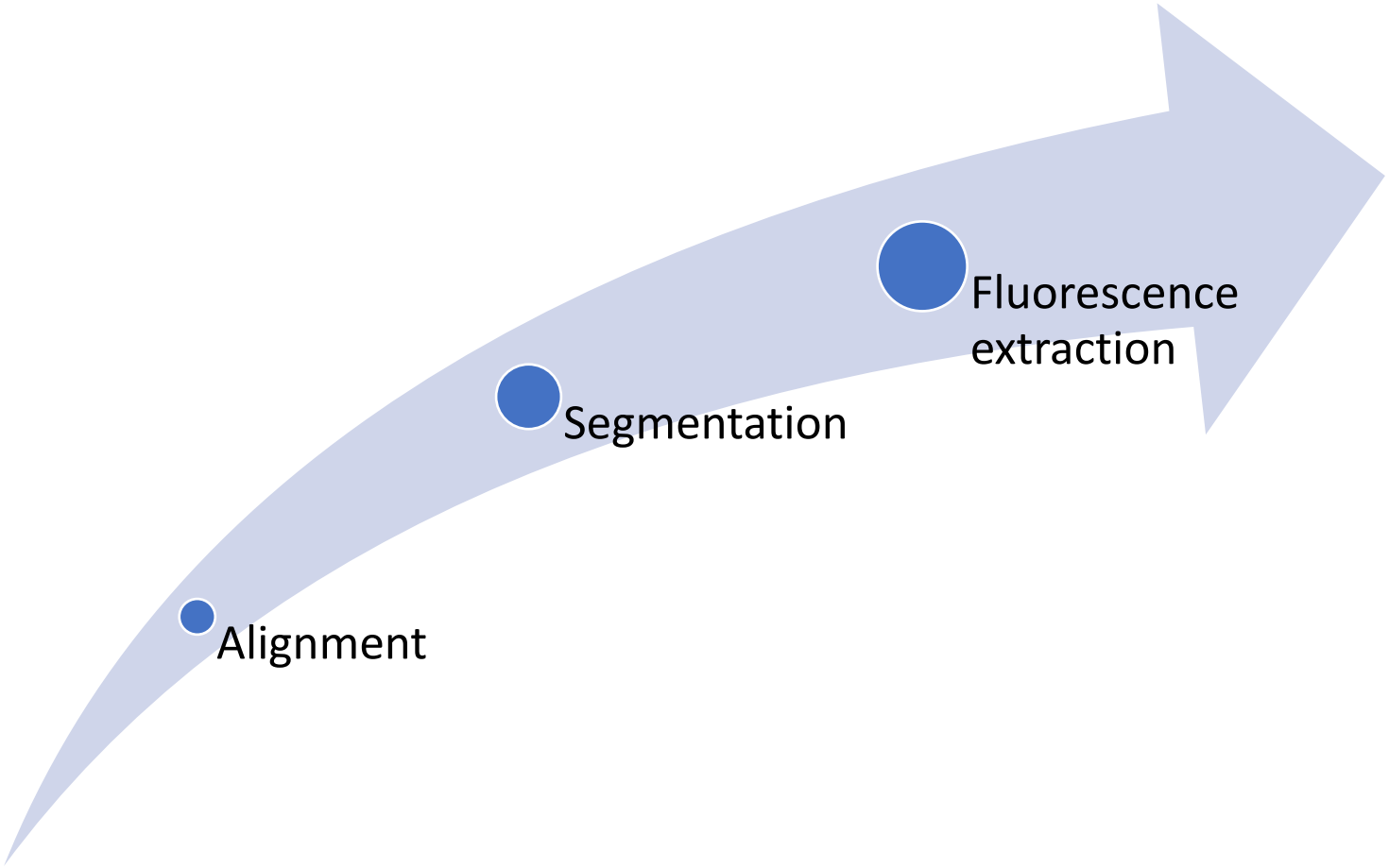




Applying a rolling method for normalisation



Three key steps to pre-processing



Things to consider

- Experimental parameters
  - Imaging system e.g., FOV size, image scale, imaging method*
  - Labelling strategy e.g., sensor kinetics, nature of objects with fluorescence
  - Activity of objects e.g., low/high, clustered/un-clustered events
- Processing/data systems
  - Storage location e.g., local/drive/server, available storage amount
  - Processing machine i.e., computational power

An opinion on employing complex tools

Having used and developed these algorithms in our own lab for our own data, we find ourselves often unwilling to use a complex method for potential small gains in performance, particularly if the method may introduce additional confounds, known or unknown.

Stringer and Pachitariu (Current Opinion in Neurobiology 2019)

*the developers of the algorithmically relatively simple suite2p

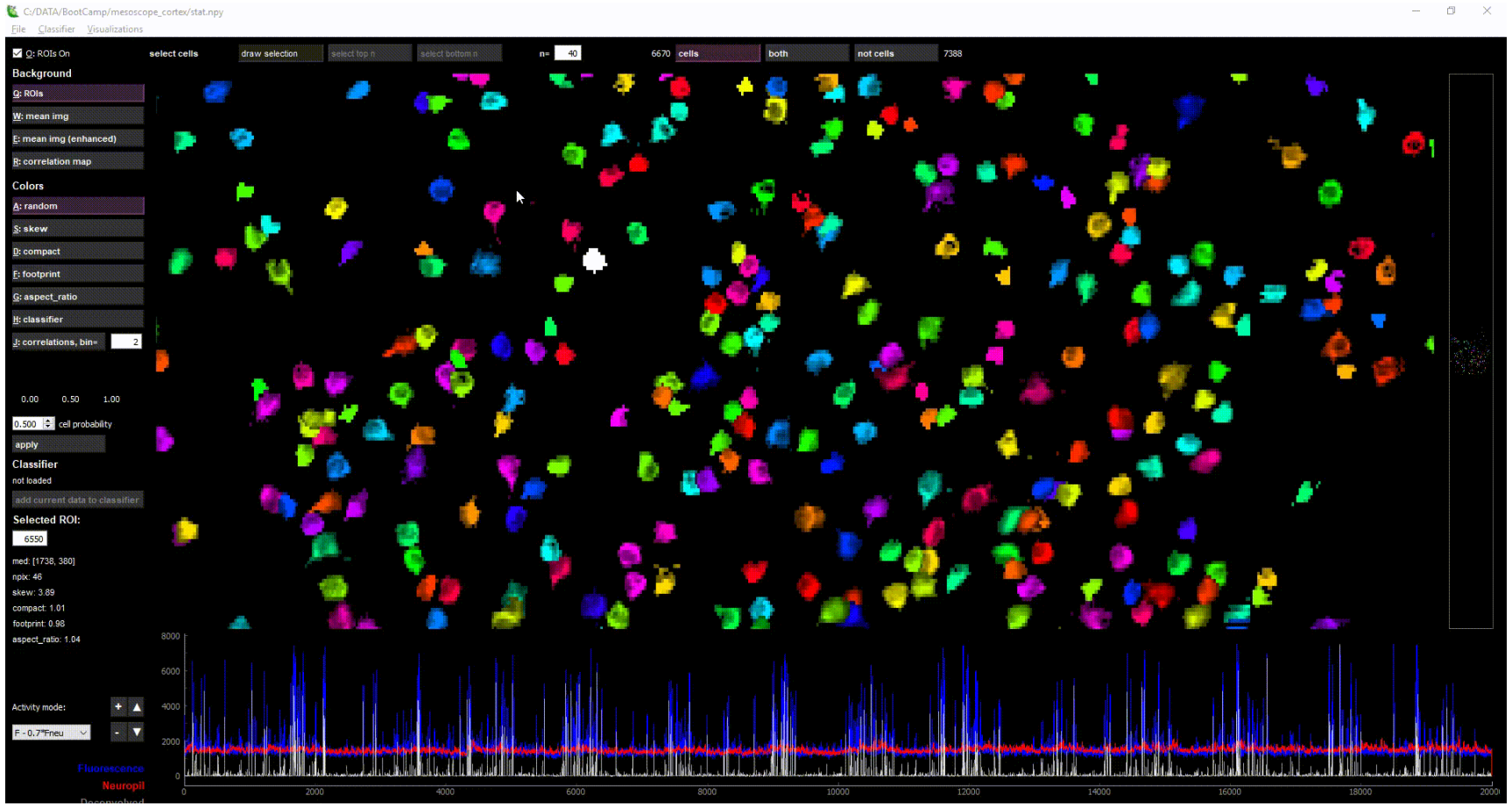
Other items

Tips and tricks

Programming

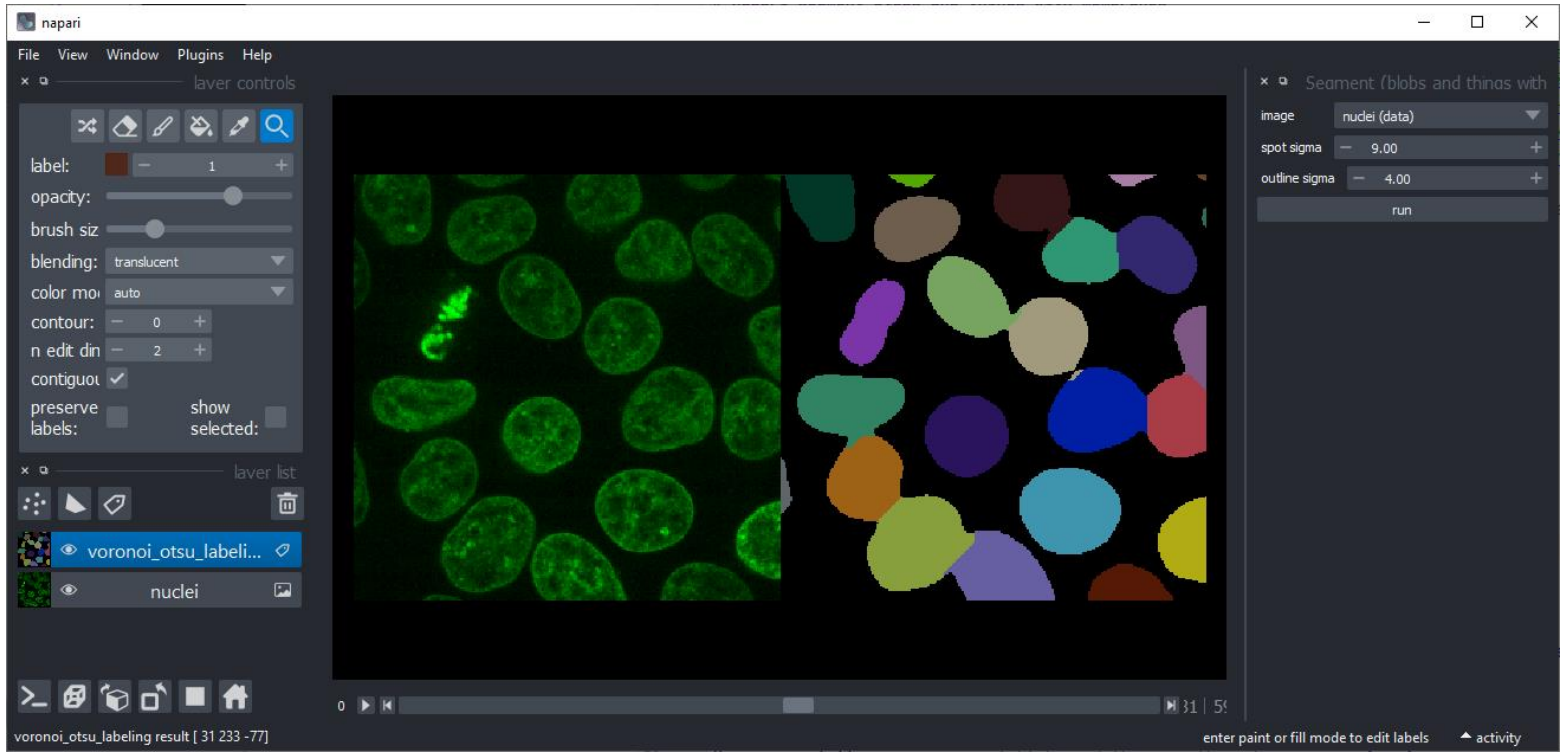
- Ya need it

Quality check:
Realise. Real eyes. Real lies. – Tupac (?)



Visualisation

- Raw video examinations
- Raw trace examinations

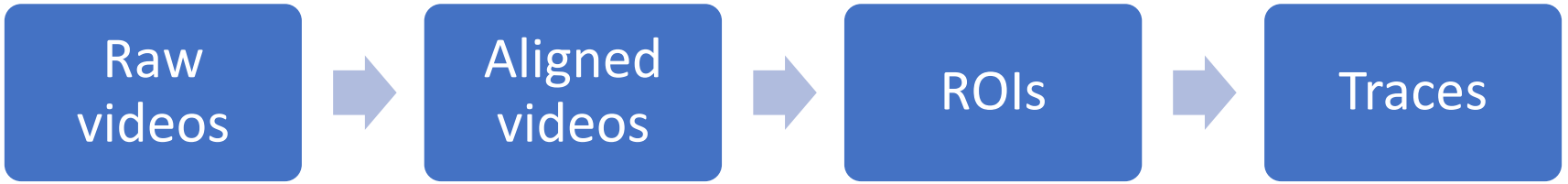


Things to be quality checked

- Alignment performance: viewing videos, plotting offsets
- Z-drift: slow alignment offset drift, raw value baseline shift in ROIs
- Fluorescence extraction
- Normalisation performance

The problem of data exclusion

1. Identifying data that should be excluded
2. Preventing/removing the data from the pipeline



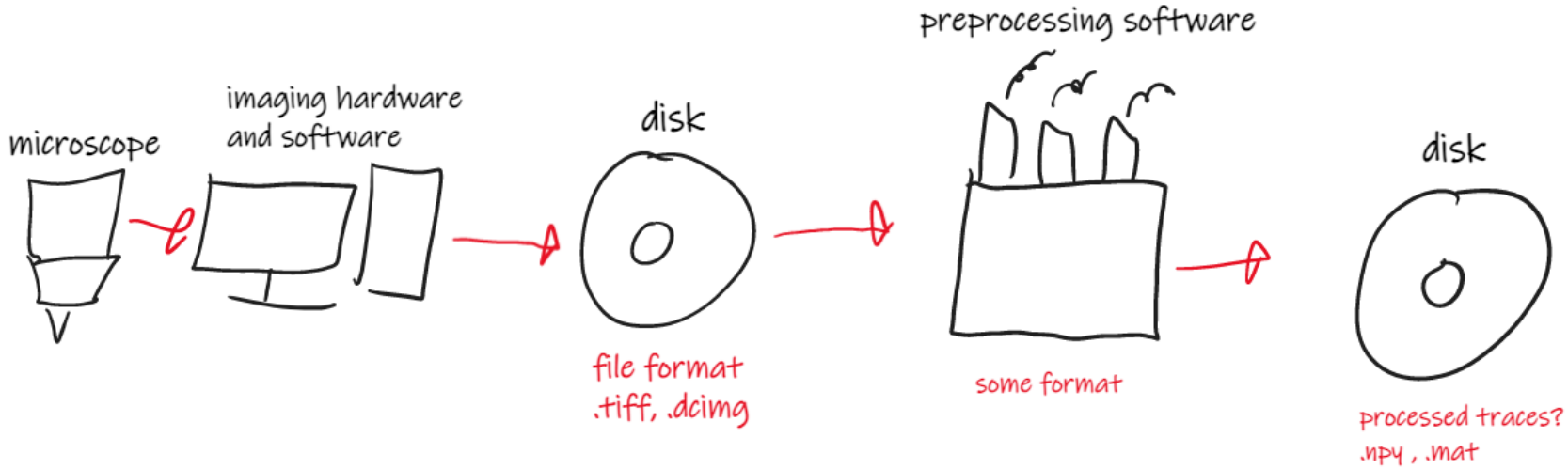
IMO: later is generally better, and explicit/tracked removal is essential

Time and space are limited

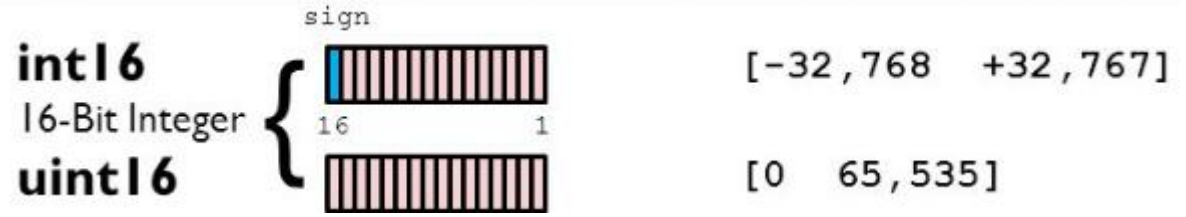
- Raw data coming out can be ~50GB/hour at (512x512 30 fps 16bit)
- Storage of the outputs of computationally expensive processes is common
- Data read/write speed can be a key bottleneck when handling imaging data depending on your setup

File and computer technicalities

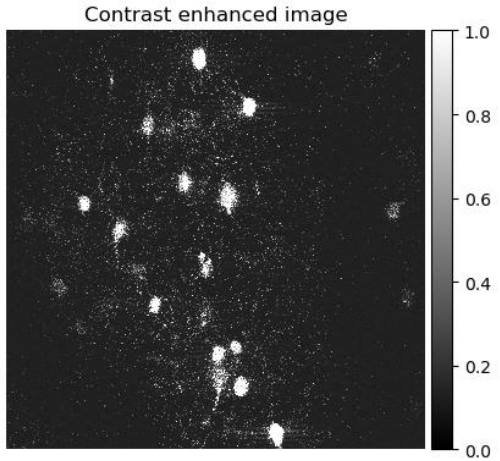
- Loading files is a time bottleneck
  - Loading method can matter
  - SSD storage with USB 3.0 plz
- Secrets of raw data

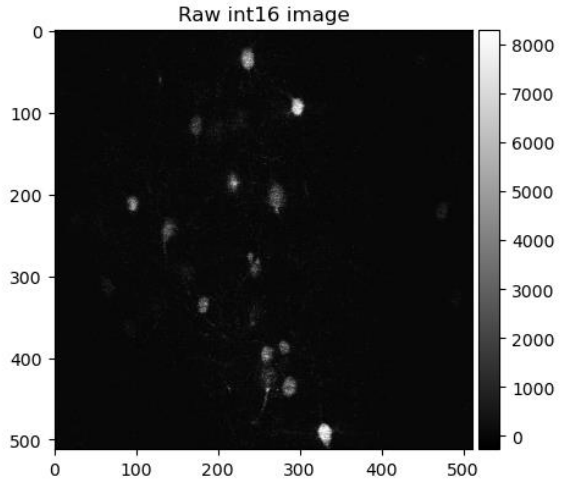


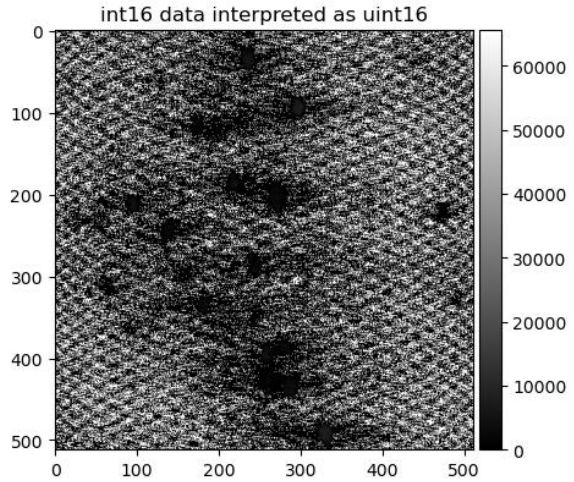
Misreading datatypes

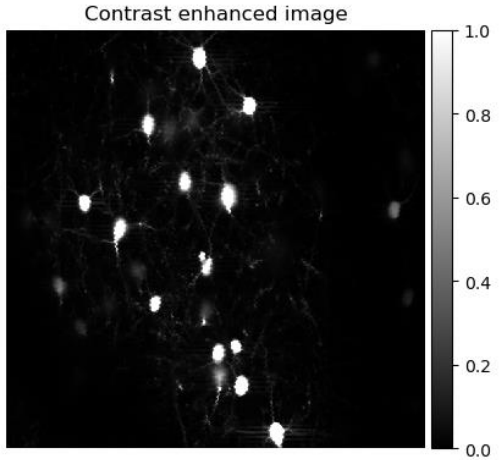


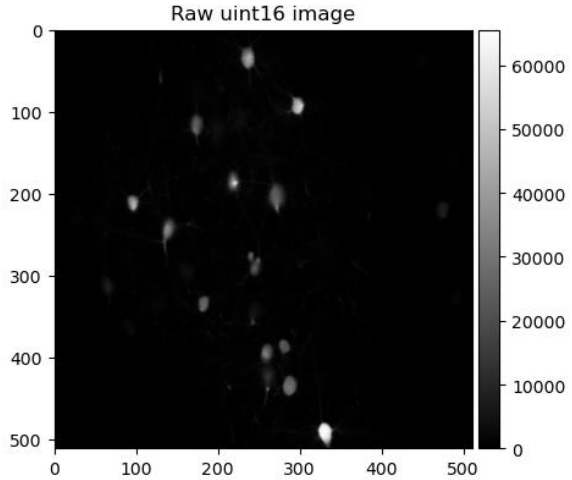
https://slideplayer.com/slide/7556788/

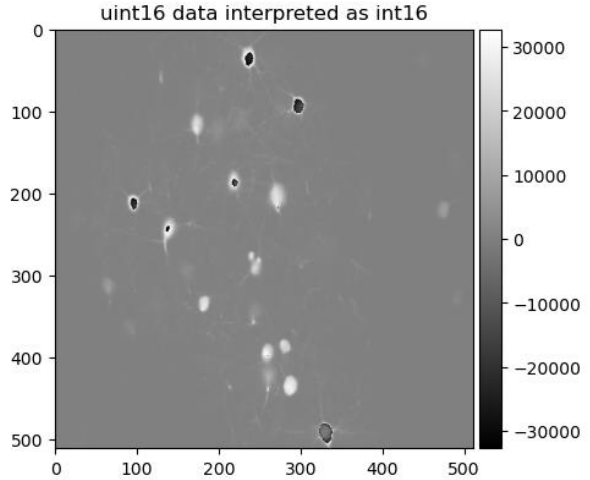












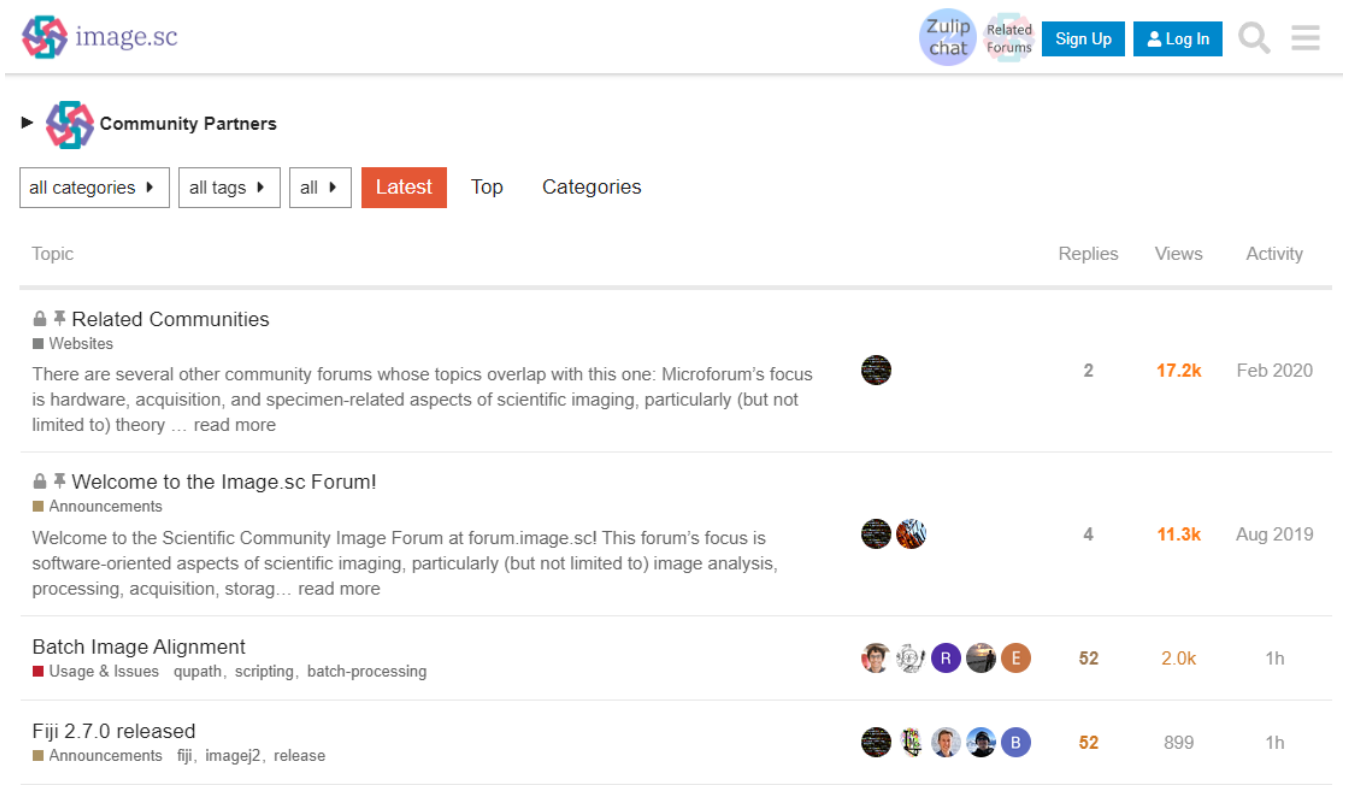
GitHub

- GitHub is basically cloud storage for text/code
- All of the calcium imaging tools that are being made are being shared by GitHub
- README.md should explain everything
- There should be a link to documentation, often stored at readthedocs.io for larger projects
- Examples of usage are typically in a examples/ or notebooks/ folder

image.sc -- internet at its finest

What is the forum's mission? (from image.sc's FAQ)

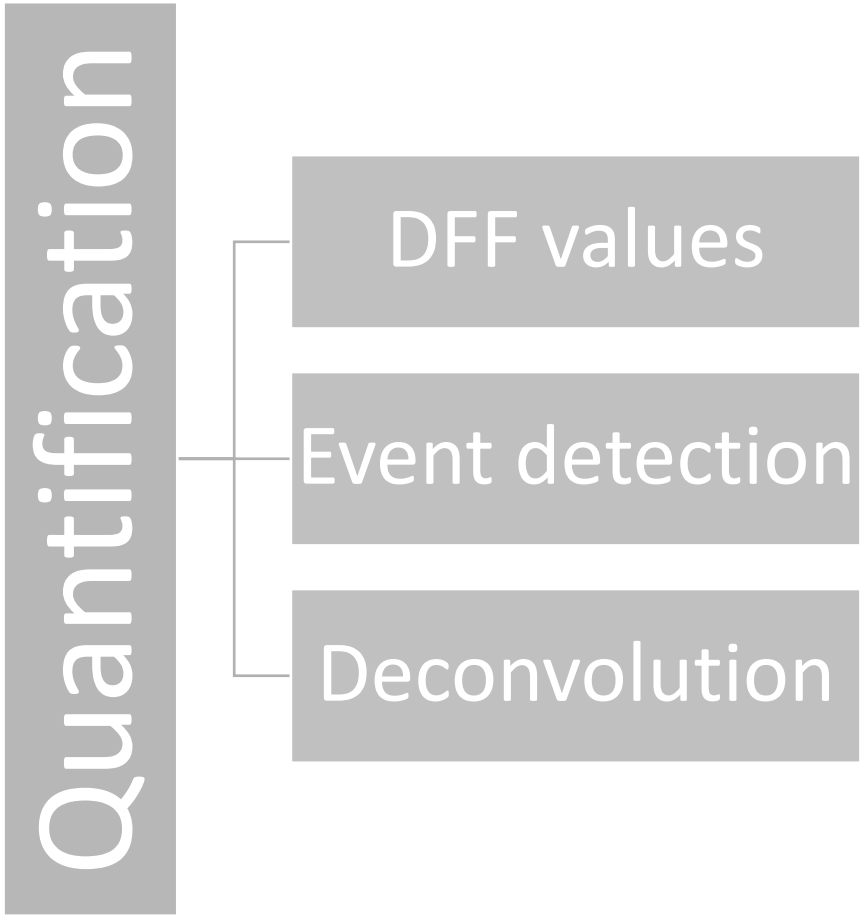
- The goal is to embrace the diversity of the scientific imaging community, while fostering independent learning:
- Enable people to ask "how do I do X?" without prior knowledge of these various software programs.
- Improve the cross-visibility of software packages.
- Make searching for previous discussions simpler.
- Give users access to a wide breadth of experts on various softwares.
- Give experts a place to have detailed discussions about elements of the software.
- Educate software developers on the capabilities of the various projects being discussed, so that they can improve the links between tools, and develop features that are more likely to be novel.
- Encourage open science and reproducible research by advocating for open tools and their interoperability.
- Foster not only scientific independent thinking, but just as importantly, independent learning 10. We want to not only teach people how to fish 17, but teach them how to learn.



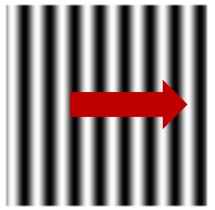
Quantification aka analysis

Core analysis methods

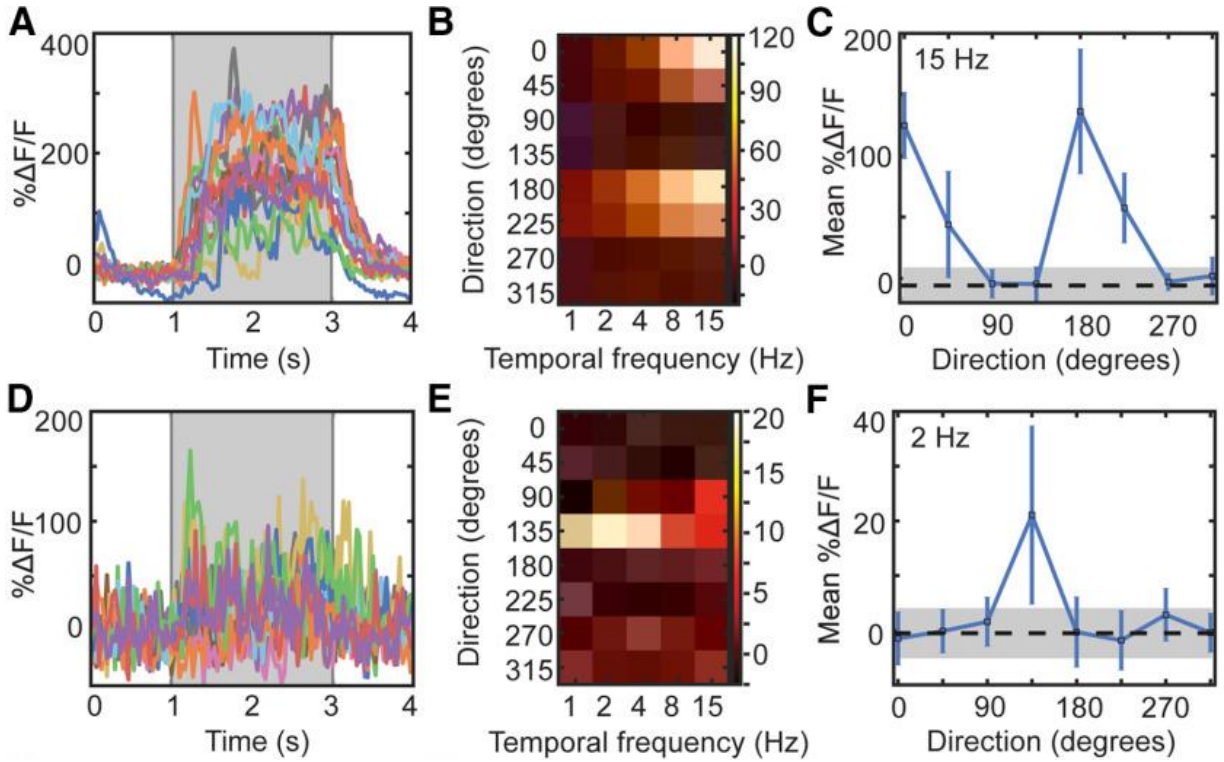
- Everything begins from DFF traces



Fluorescence itself as a metric



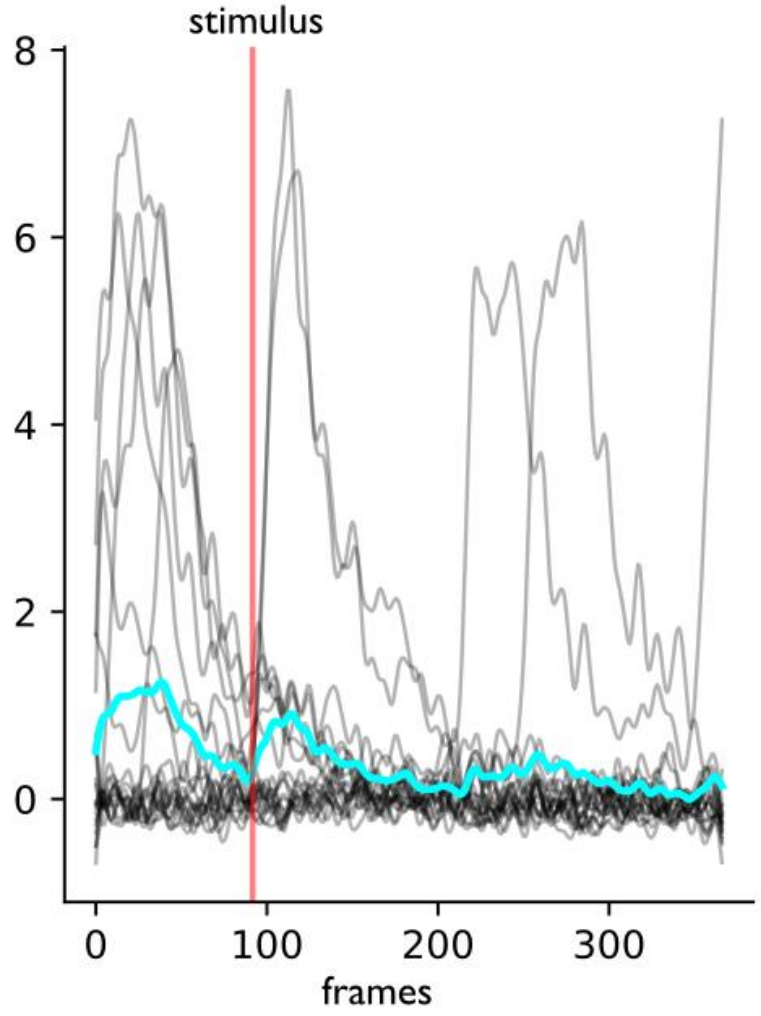
- Average or integral across epochs
- In theory, no information is lost



Mesa, Waters & de Vries (2021)

Time specificity could be janky

- Long tails result in “misattributed” activity
- Task with cue can produce start of trial events



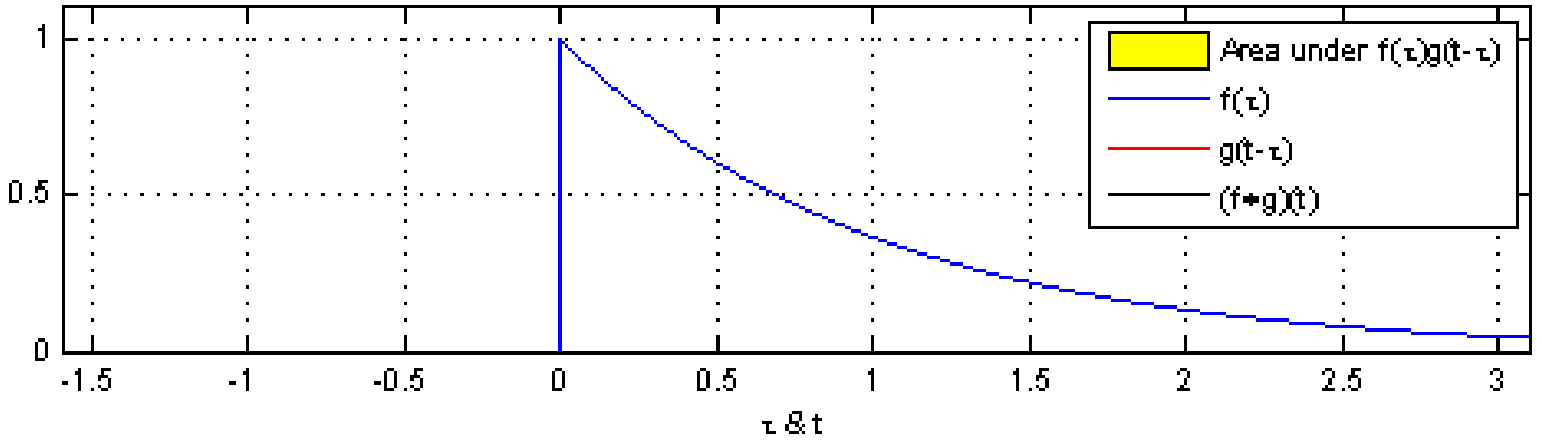
Event detection

- Specific to the dynamics of the calcium fluorescence you are observing
- Go to Methods to find those boutique specifics

Deconvolution

Convolution

- You convolve a **signal** with a **kernel** to produce its **convolution**



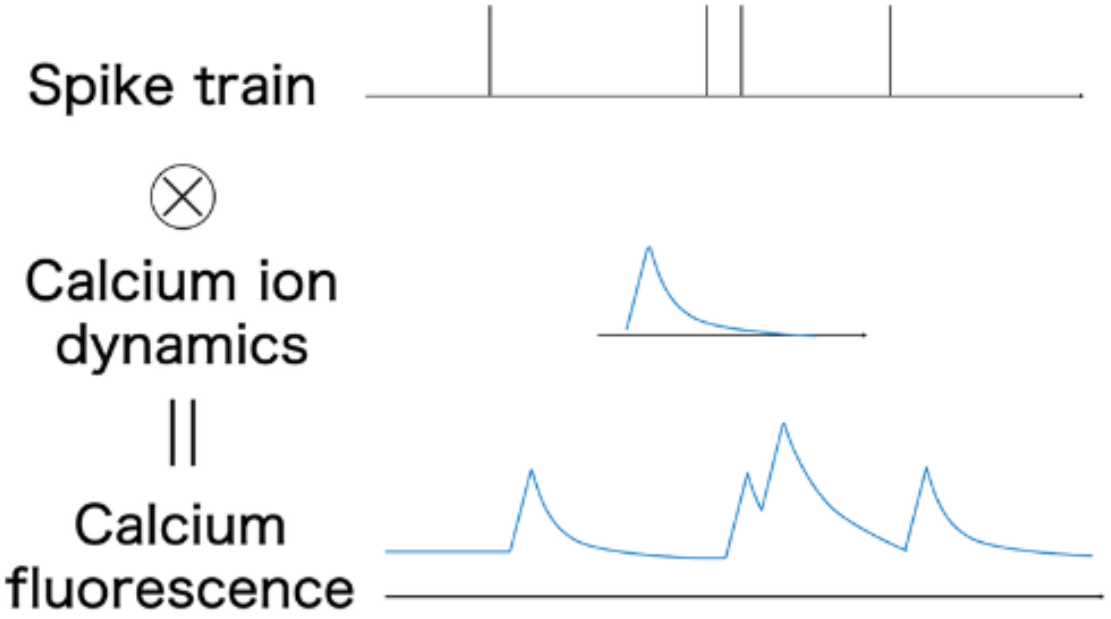
https://commons.wikimedia.org/wiki/File:Convolution_of_spiky_function_with_box2.gif

- Try out a demo: https://phiresky.github.io/convolution-demo/
  - Use sawtooth kernel and impulse train signal, see what happens

Deconvolving calcium fluorescence

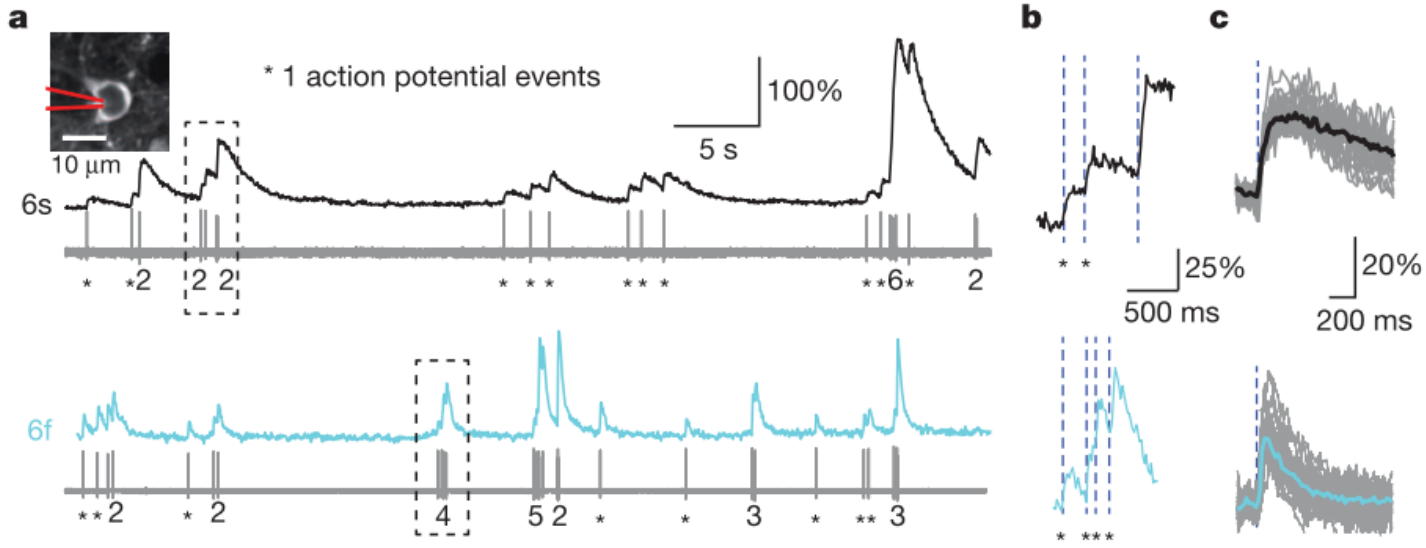
- Fluorescence signal is known
- Kernel can be determined*
- Spike train is unknown

Deconvolution of calcium fluorescence is the extraction of the underlying “spikes” that generated the fluorescence pattern.

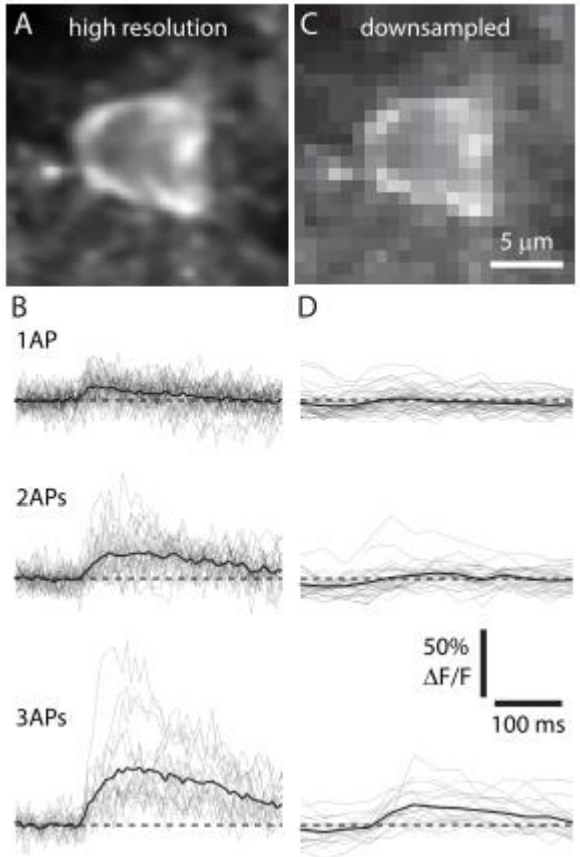


Shibue & Komaki, *PloS CompBio* 2020

Spike* deconvolution



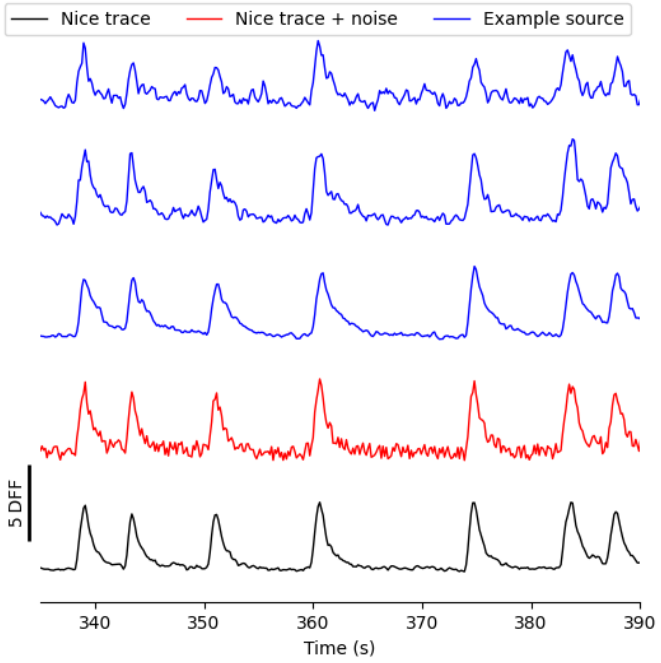
Chen … Kim, *Nature* 2013

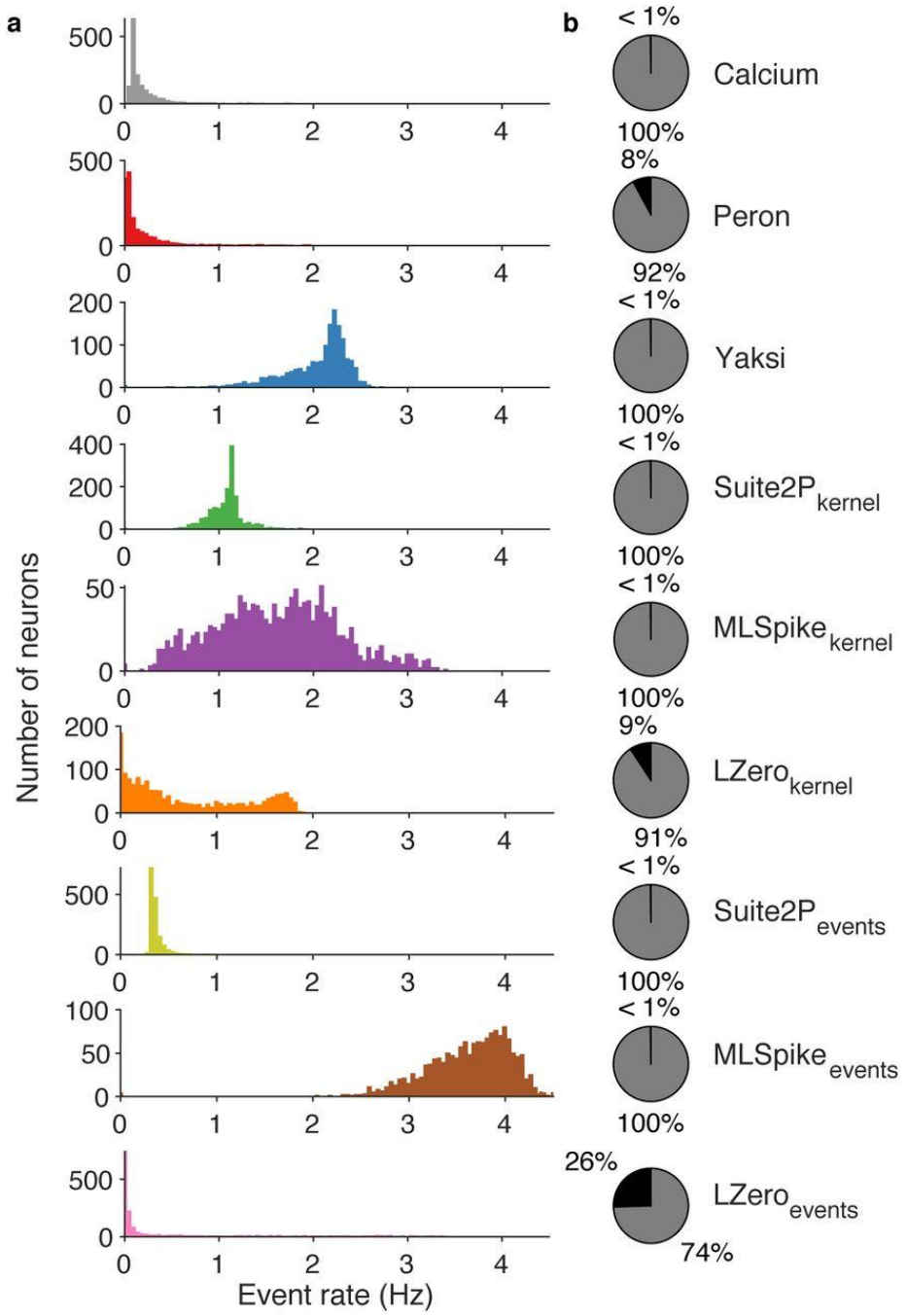


Allen Brain Observatory, *eLife*, 2021

Deconvolution may be highly method dependent

- Deconvolution algorithm try to fit a kernel to the given fluorescence timecourse
- Real traces can have non-random artefact





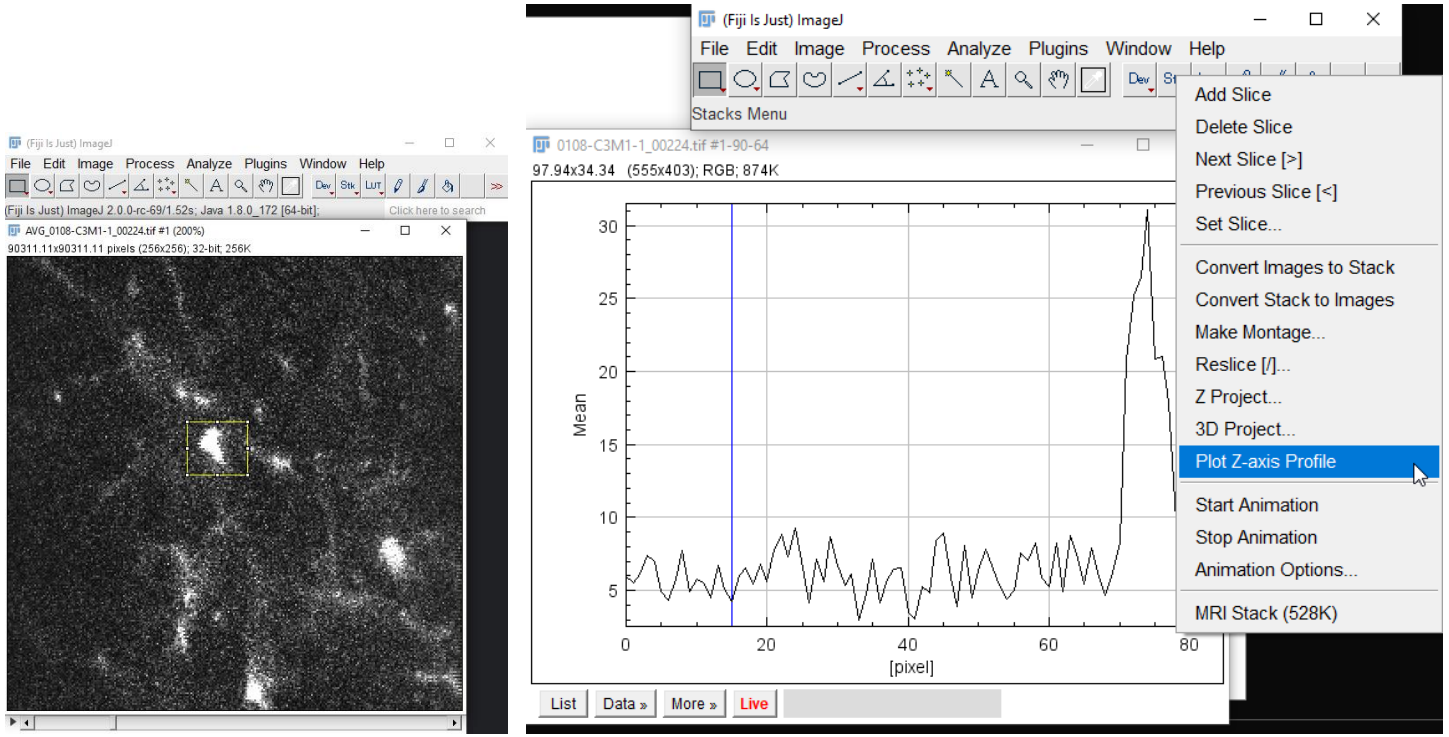
Evans, Peterson & Humphries (2020)

Remember Tupac: use your eyes

Average

Demo location

- https://github.com/ogeesan/ACAN-2023-Imaging-Workshop
- fiji.sc



Notebook (from last year)

https://github.com/ogeesan/ACAN-2022-Imaging-Workshop

1. Open Notebook
2. Runtime > Run all