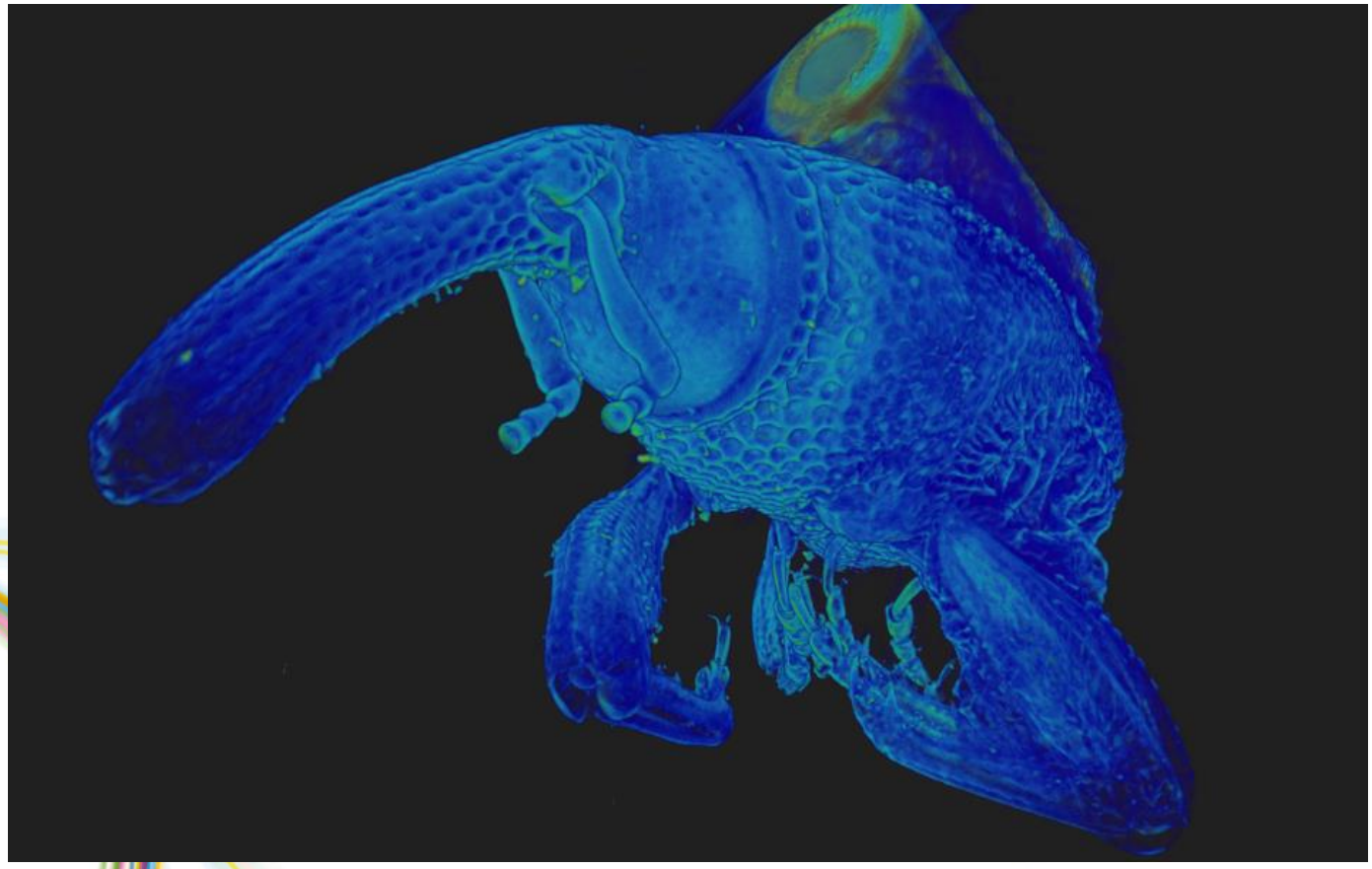
What Science can do with AI and Machine Learning

Dr Mark Basham

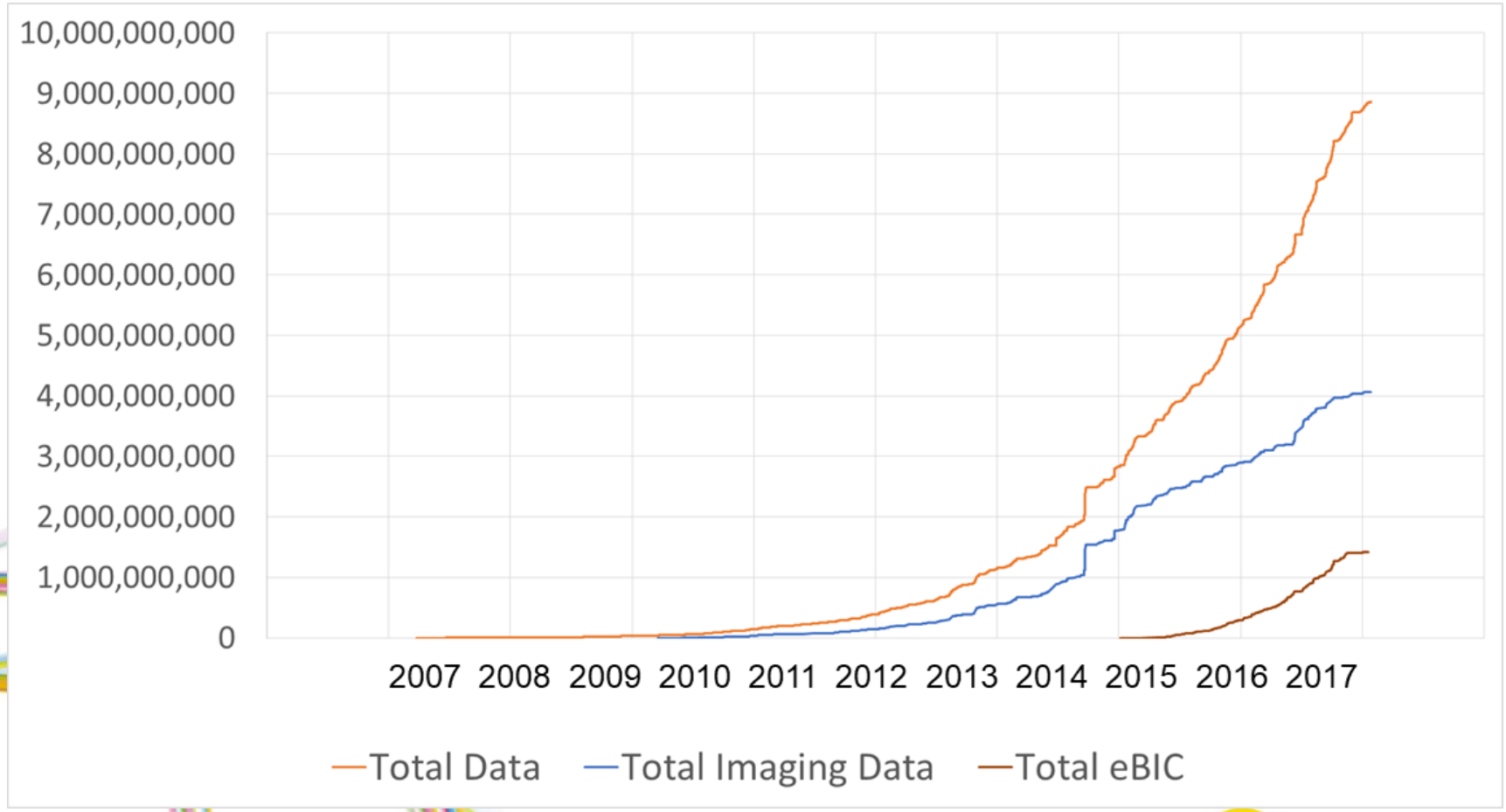
Data Analysis Group
Diamond Light Source

Focus: Imaging

- Of interest to the ALC as it is cross disciplinary (3D+ imaging)
  - DLS
  - ISIS
  - CLF
  - …



The complexity – this is now a big data problem for DLS++



The emerging challenge.

>3TB

30MB

PCCP

PAPER

View Article Online

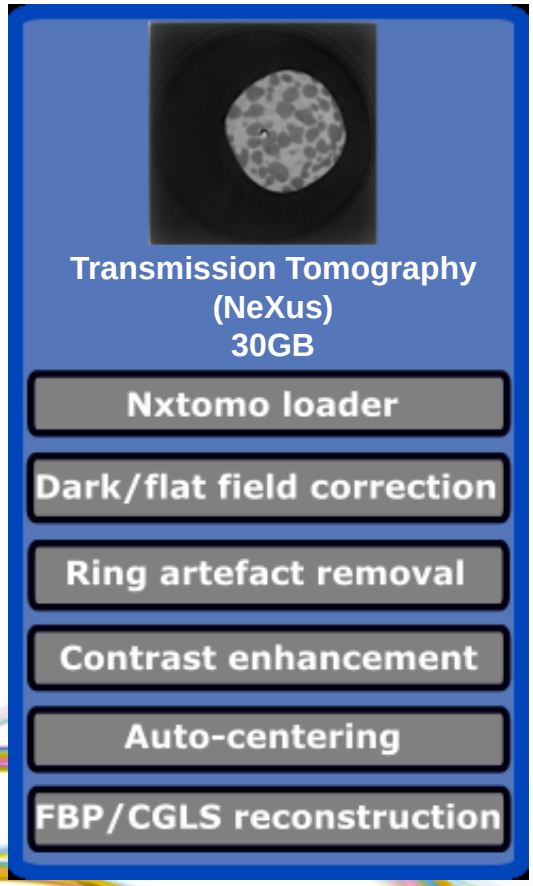
Chemical imaging of single catalyst particles with scanning μ-XANES-CT and μ-XRF-CT†

diamond

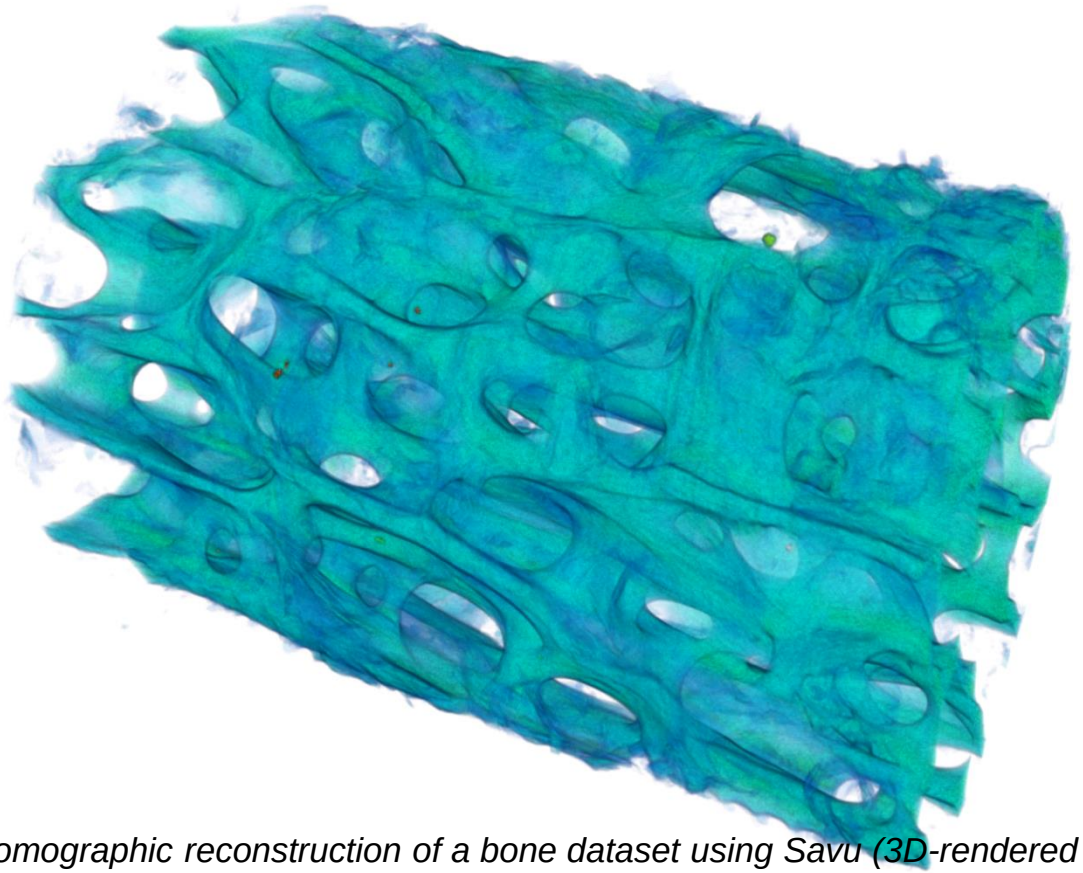
A tomography pipeline called Savu.

diamond

Full-field tomography processing with Savu at DLS



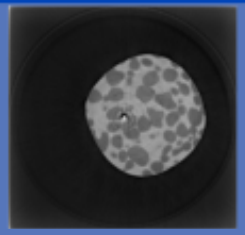
A typical process list for reconstructing full-field tomography data.



Tomographic reconstruction of a bone dataset using Savu (3D-rendered using VisIt). Courtesy of Gianluca Tozzi, Marta Pena-Fernandez, Rachna Parwani, and Asa H. Barber (2016) from Portsmouth University. Data collected on the Diamond Manchester Imaging Branchline (I13-2) with support from Andrew J. Bodey.

diamond

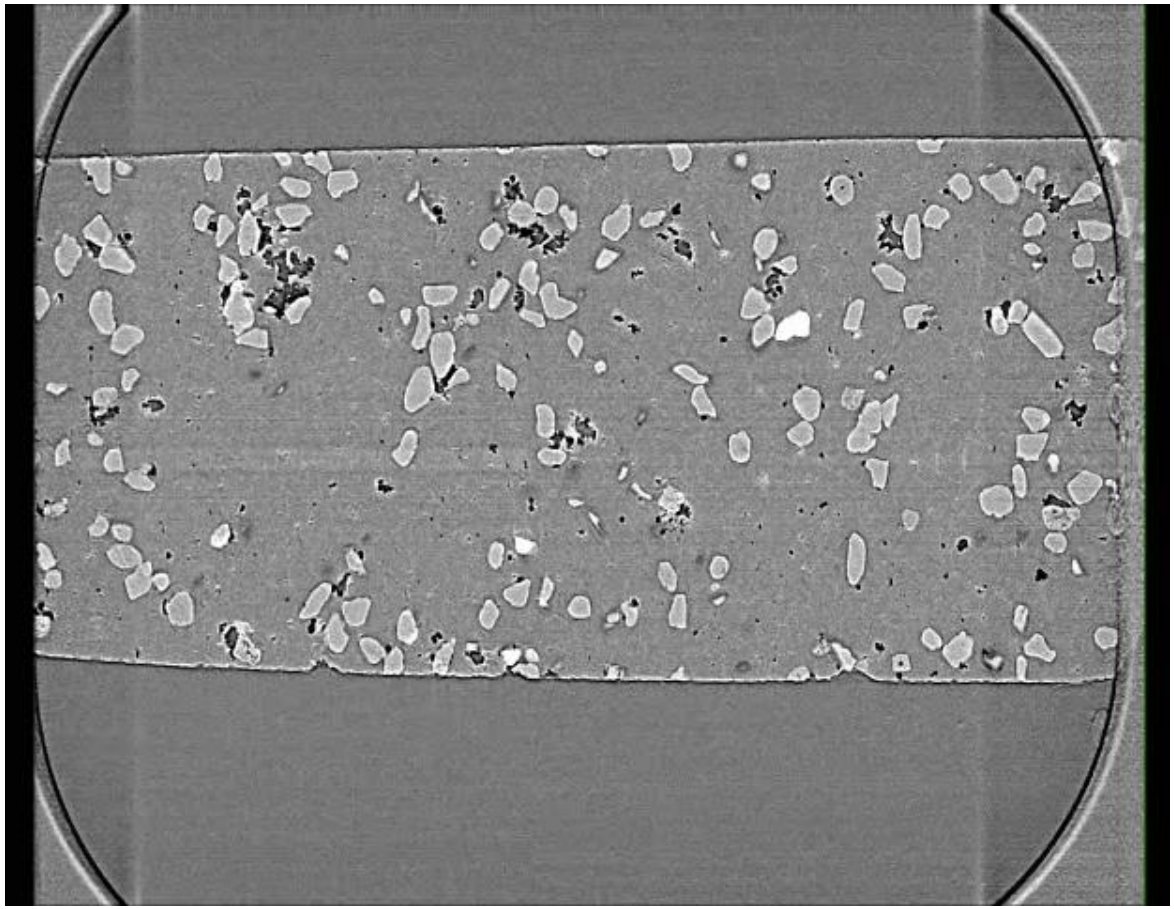
Full-field tomography processing of 4D data with Savu at DLS



Time Resolved Tomography (cine) 1500GB

- Nxtomo loader
- Dark/flat field correction
- Ring artefact removal
- Contrast enhancement
- Auto-centering
- FBP/CGLS reconstruction

A typical process list for reconstructing full-field tomography data.



diamond

Current processing and resource requirements

- During the experiment possible
- Post processing complex

The imaging challenge.

>3TB

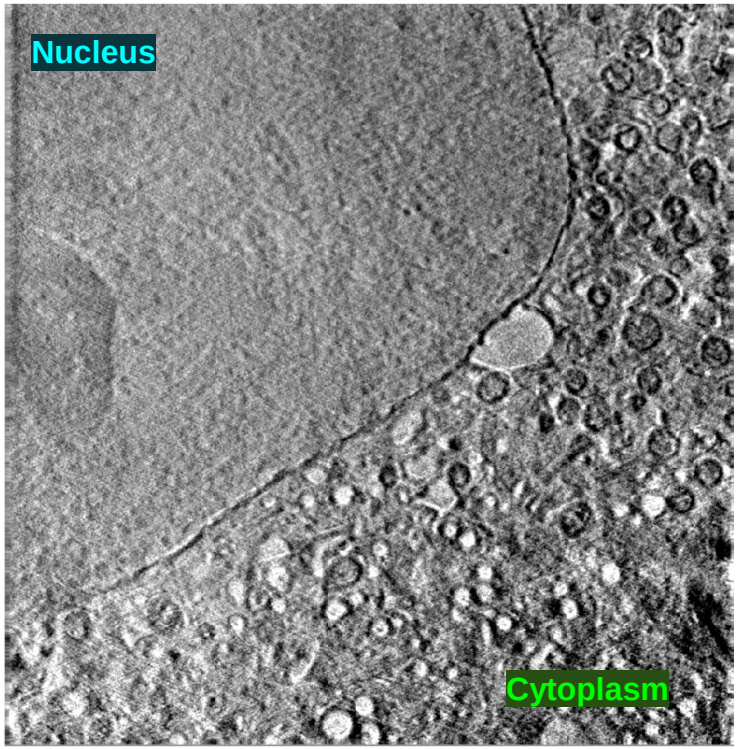
>>3TB

diamond

Now What? – Segmentation…

Challenges posed by data:

- Noisy data, missing wedge artifacts, missing boundaries
- Large amounts of data; many organelles/complexes per dataset – complex
- Tedious to manually annotate – so few training annotations available
- Each cell type/condition can look different – not generalizable
- Automated techniques usually fail



CryoSXT of neuron-like mammalian cell line; single slice

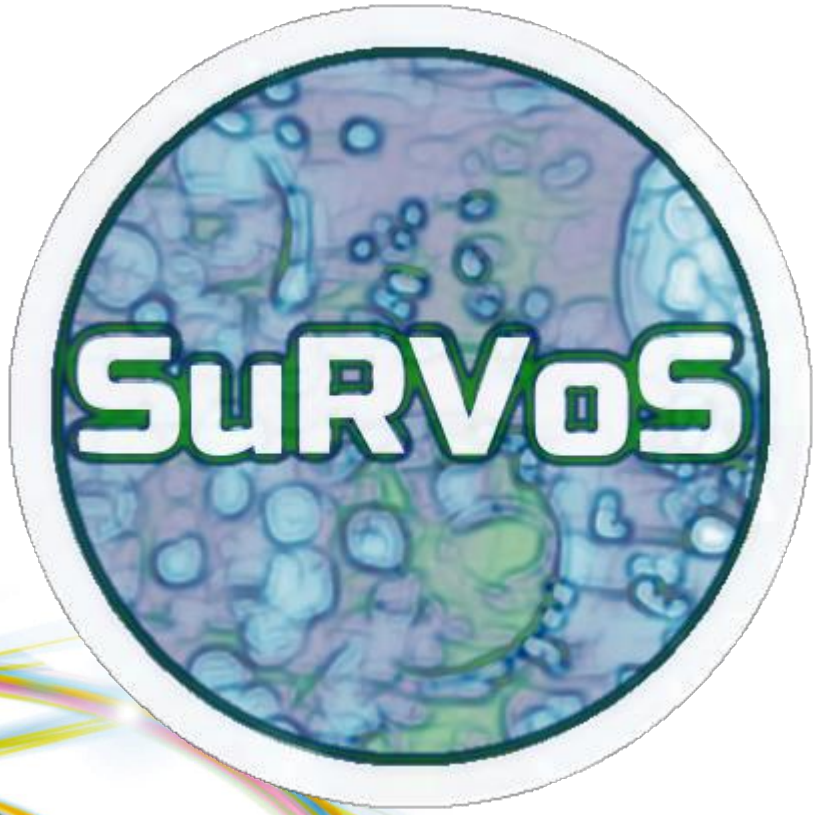
diamond

Can we apply modern machine vision methods?

- Collaborate with a Machine Vision Group
  - Nottingham university
  - 1 PhD Post to investigate the methods 2014-2017

- Conclusions
  - Yes, but not fully automatic, semi automatic
  - SuRVoS

Super-Region Volume Segmentation (SuRVoS) Workbench for Segmentation

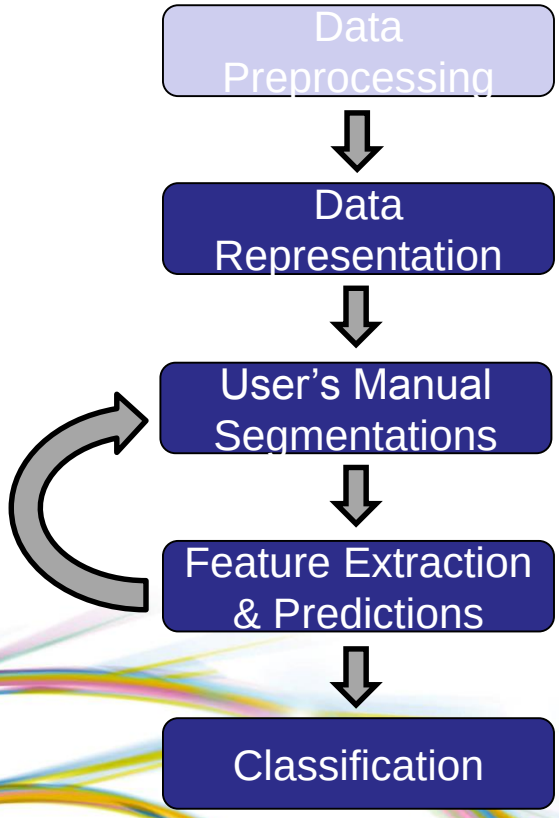


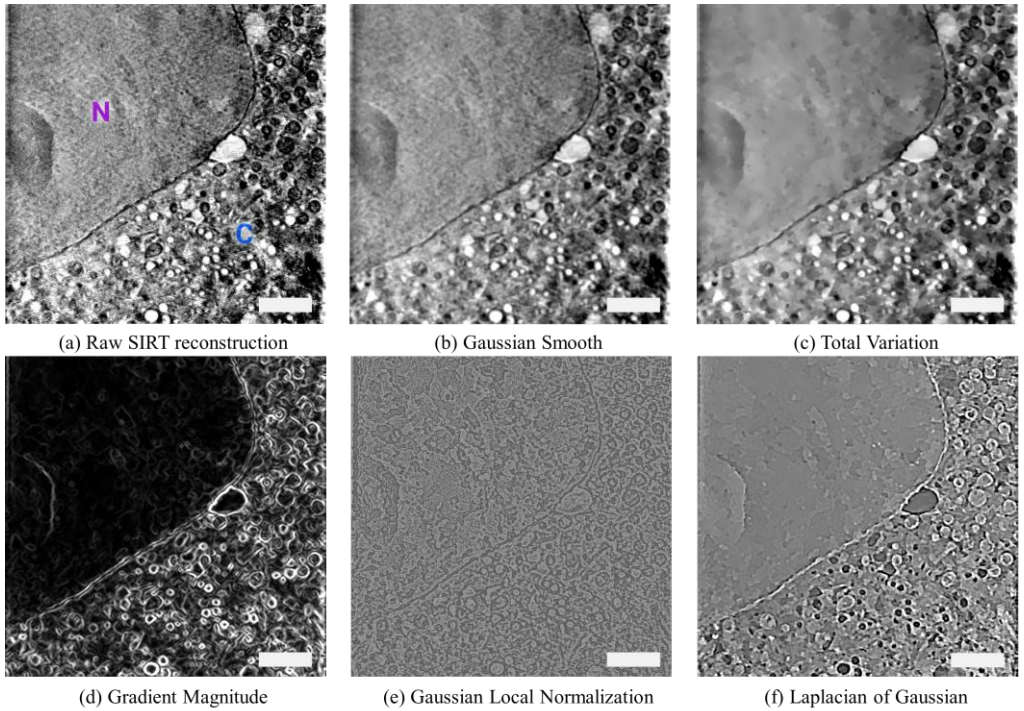
diamond
B24 beamline

The University of Nottingham
Computer Vision Laboratory

diamond

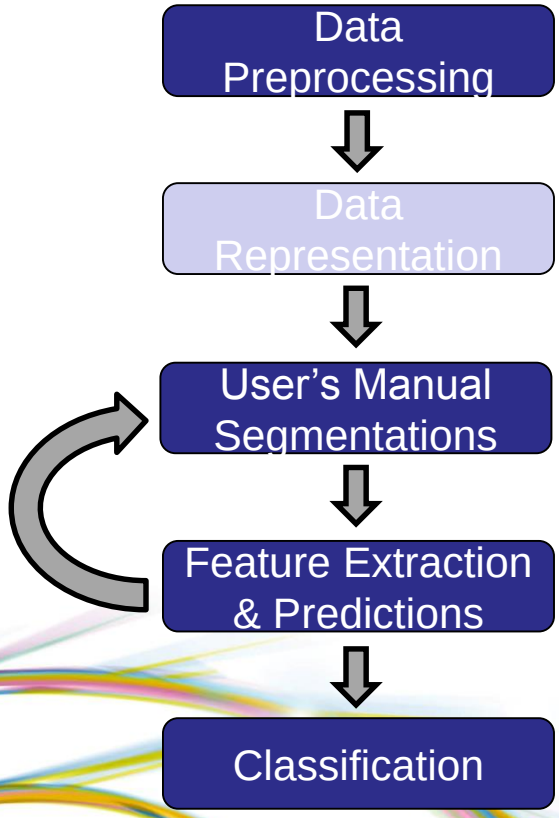
SuRVoS – Super Region Volume Segmentation Workbench



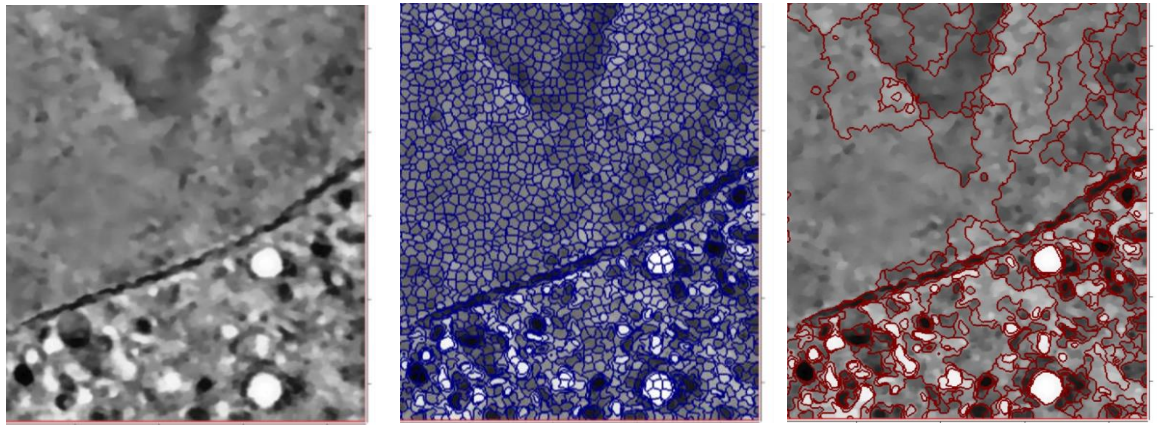


Luengo I., et al. 2017, J Struct Biol; Darrow, M., et al, 2017, JoVE

SuRVoS – Super Region Volume Segmentation Workbench



Denoised Subvolume | Super Voxels | Mega Voxels



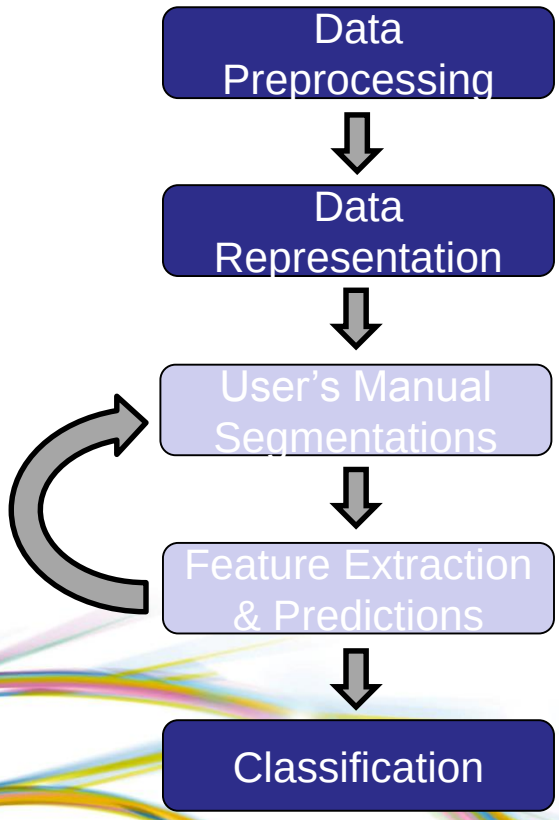
Supervoxels and megavoxels are…

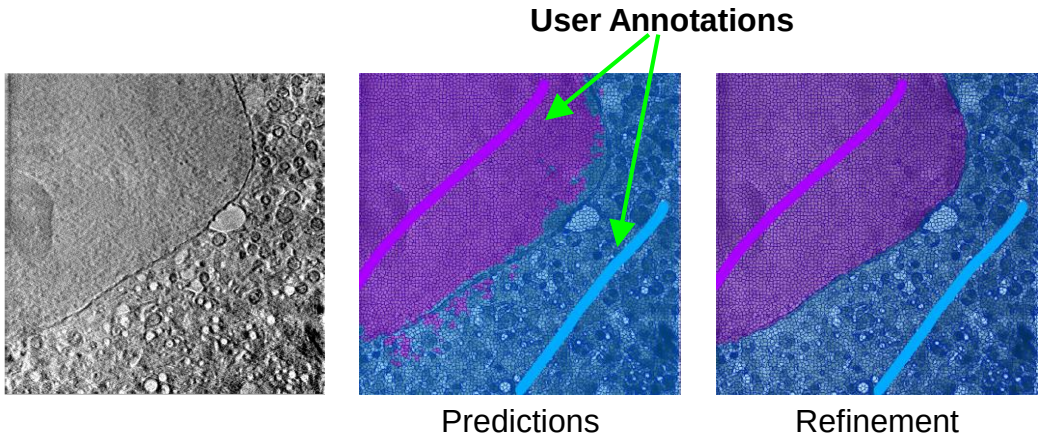
- groups of similar, adjacent voxels in 3D.
- **edge preserving**, and **three-dimensional**.
- **reduce problem complexity** by several orders of magnitude.

Luengo I., et al. 2017, J Struct Biol; Darrow, M., et al, 2017, JoVE

diamond

SuRVoS – Super Region Volume Segmentation Workbench





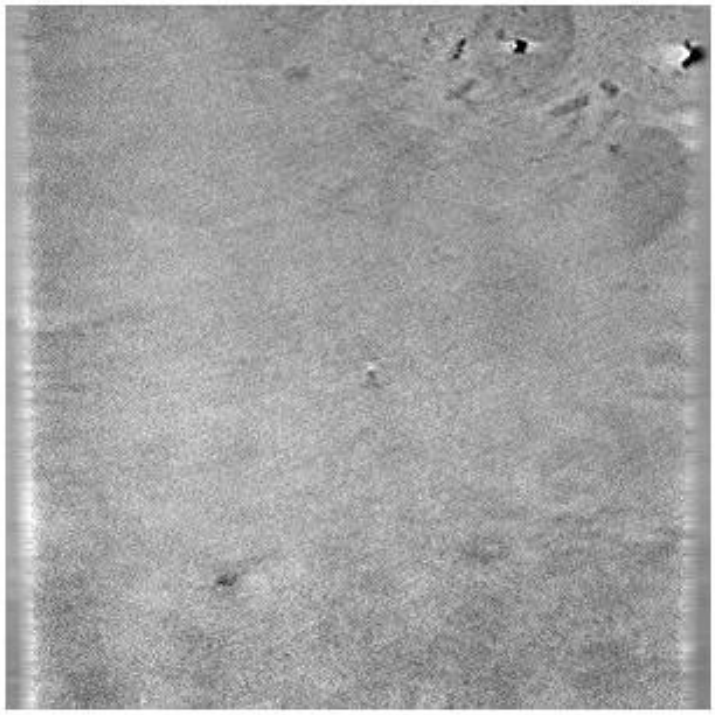
Features are extracted from voxels to represent their appearance:

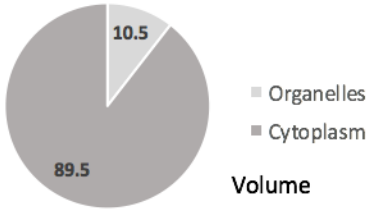
- Intensity and textural features/filters
- A machine learning classifier is trained to discriminate between different classes and predict the class of each supervoxel in the volume.
- A Markov Random Field (MRF) is then used to refine the predictions.

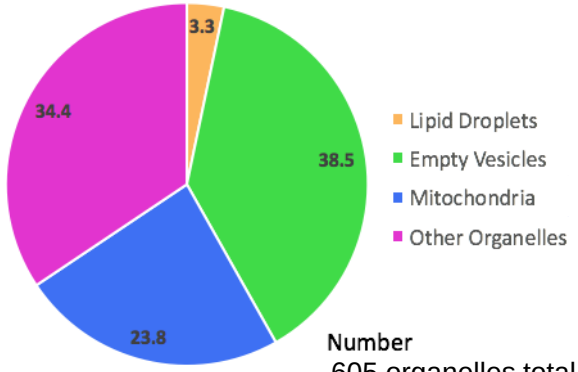
diamond

Segmented PC-12 mHTT-Ex1 (97O)

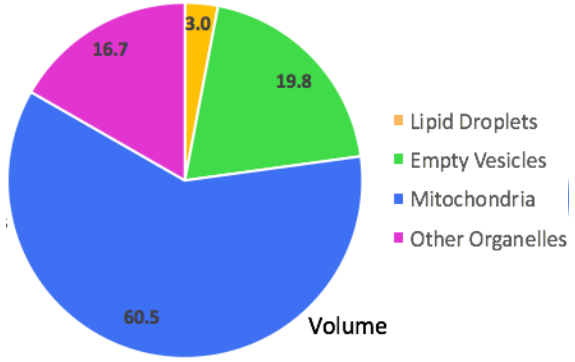
- Nucleus
- Nucleolus
- Lipid Droplets
- Mitochondria
- Empty Vesicles
- Other Organelles







605 organelles total



diamond

My Vision of ALC

- A place to bring together the facilities data analysis communities to make use of common software and expertise.
- A common pool of HPC which is capapble of dealing with this data and processing requirements. Which is available to all our facility users for post processing and analysis.

Thank You

mark.basham@diamond.ac.uk

diamond