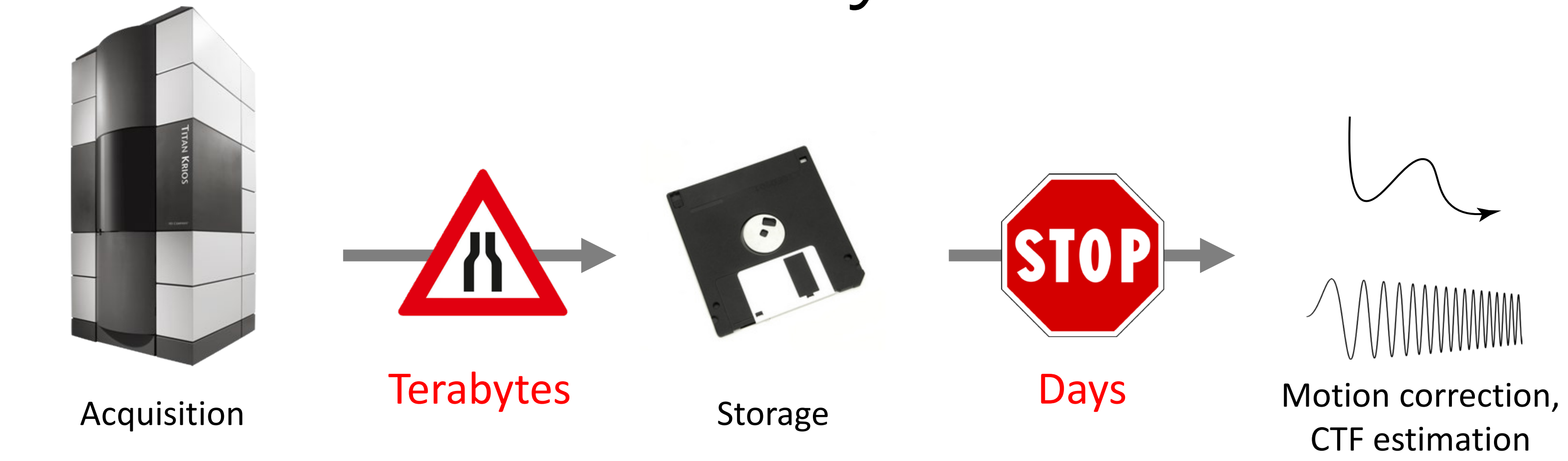
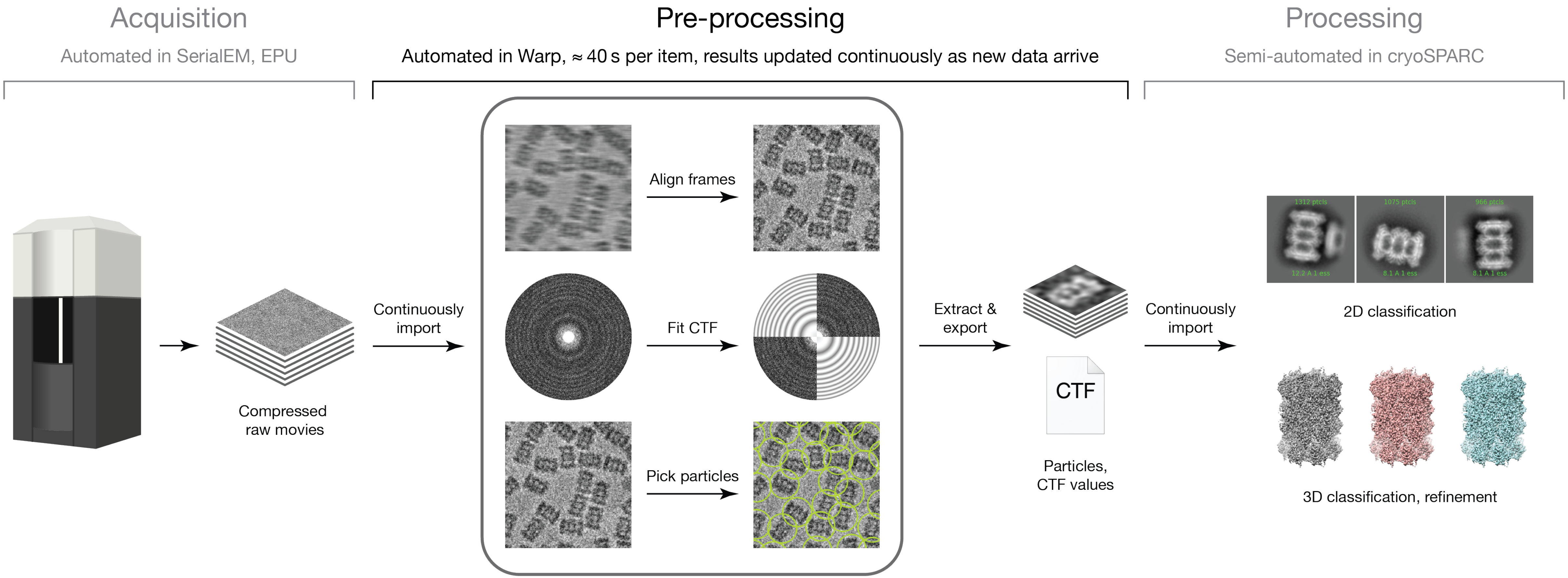
Cryo-EM Pre-Processing at Full Warp

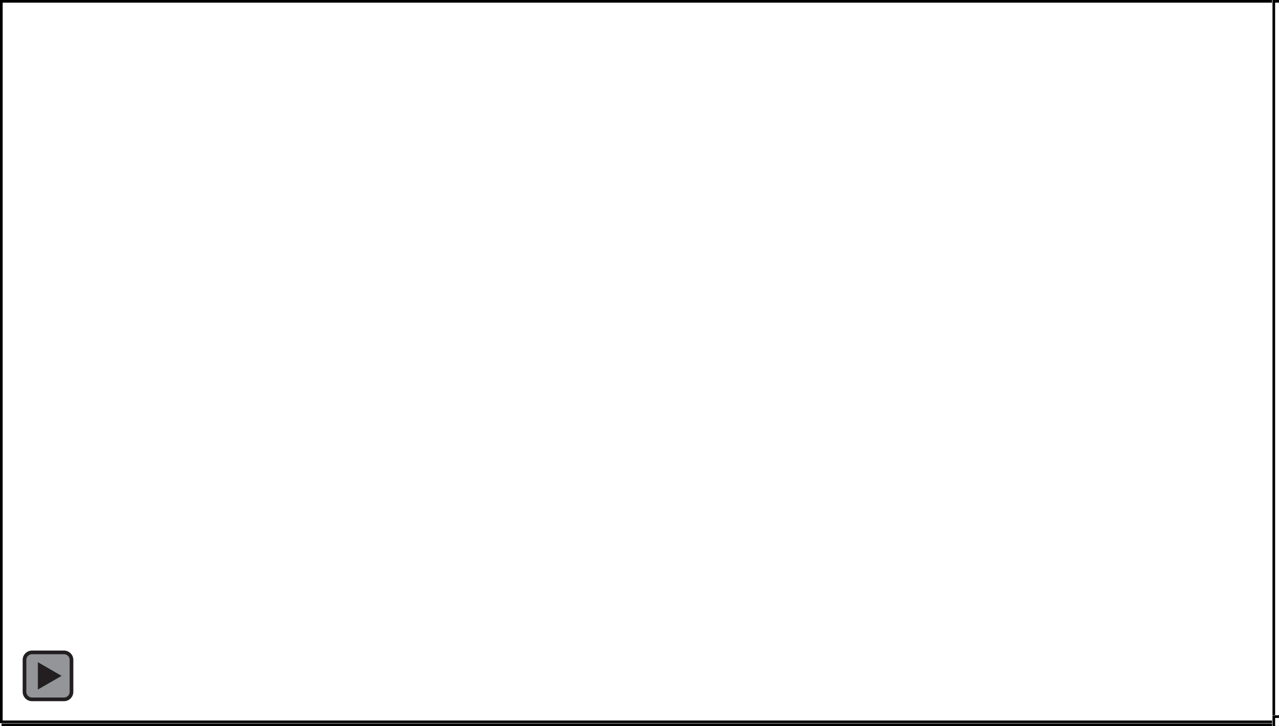
NeCEN workshop 10/2018

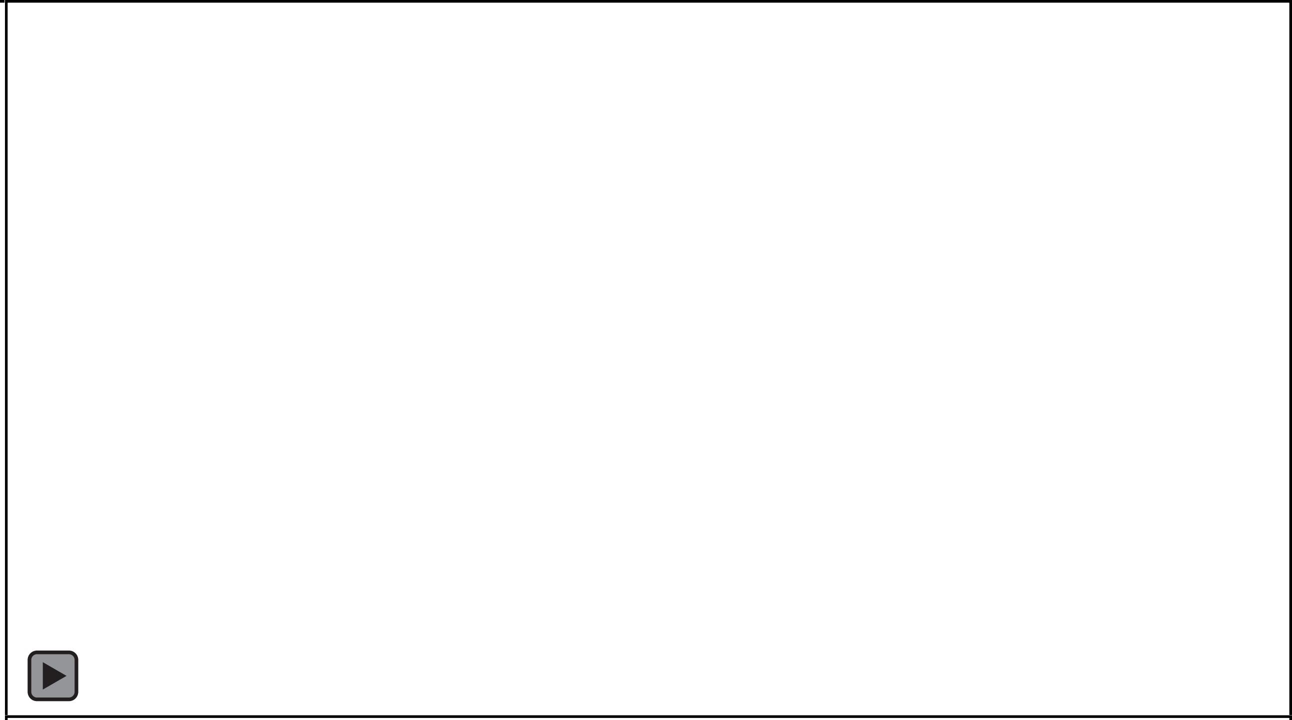
Data pre-processing, the old way

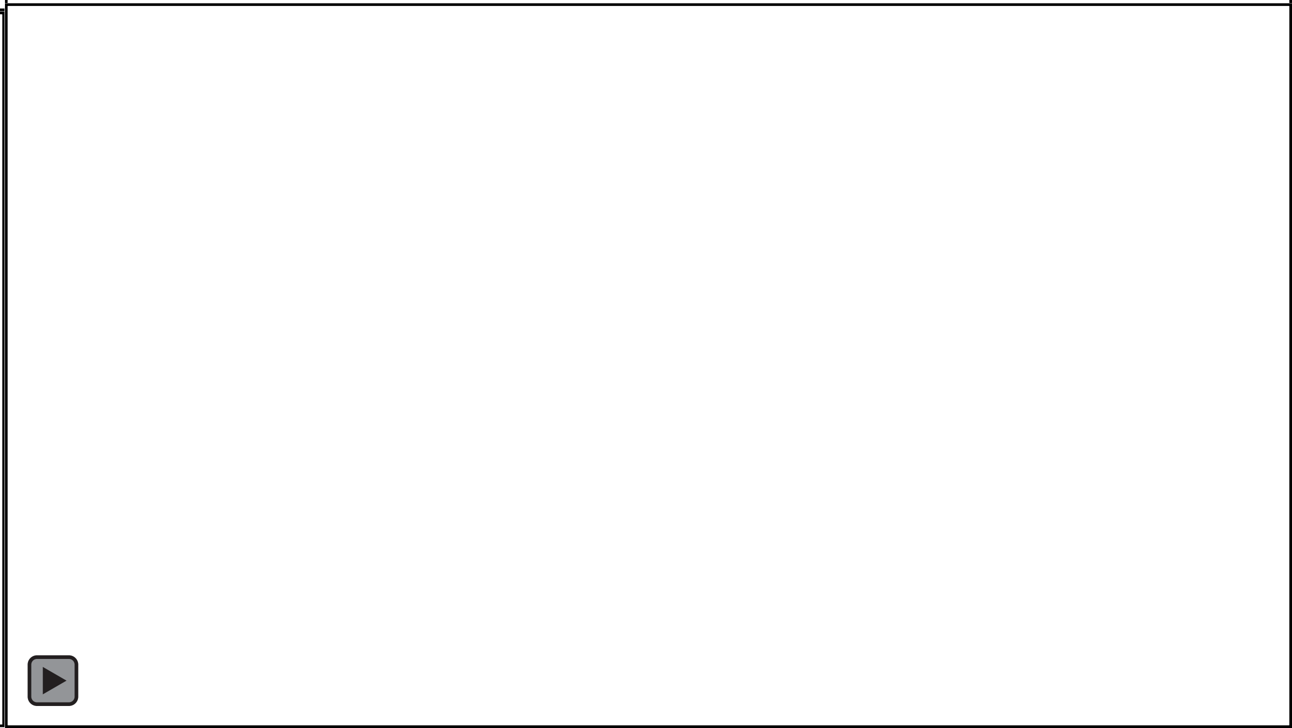


All micrographs move in one chunk, one step at a time


Acquisition
Automated in SerialEM, EPU
Pre-processing
Automated in Warp, ≈ 40 s per item, results updated continuously as new data arrive
Processing
Semi-automated in cryoSPARC
Compressed
raw movies
Continuously
import
Align frames
Fit CTF
Pick particles
Extract &
export
CTF
Particles,
CTF values
Continuously
import
2D classification
3D classification, refinement

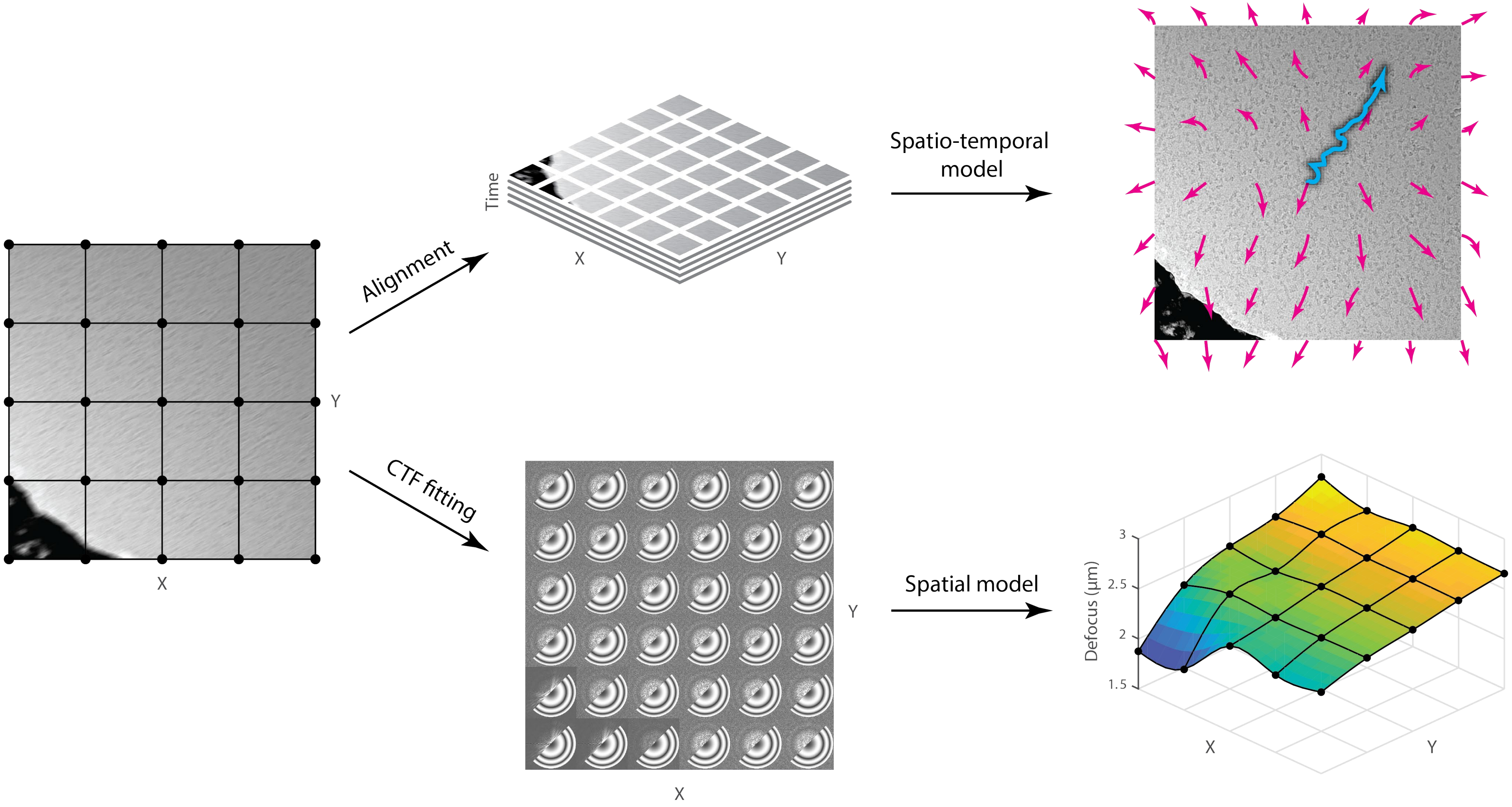






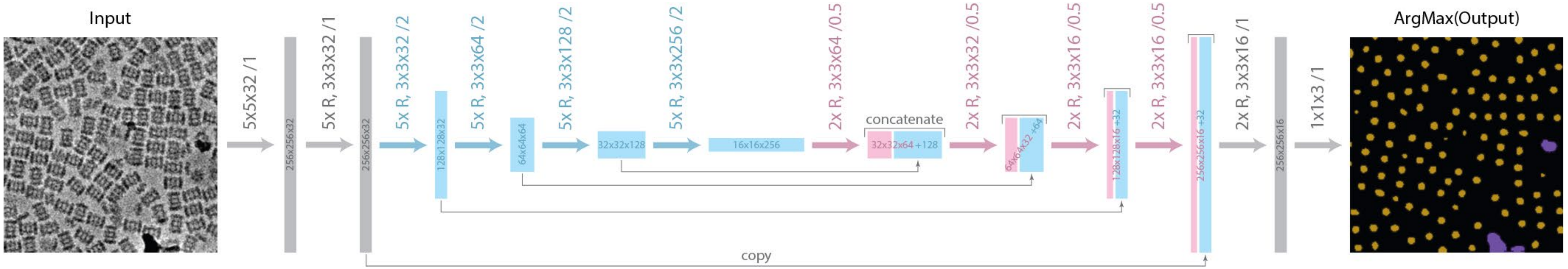

抖音
抖音号:902760646

Modeling sample deformation

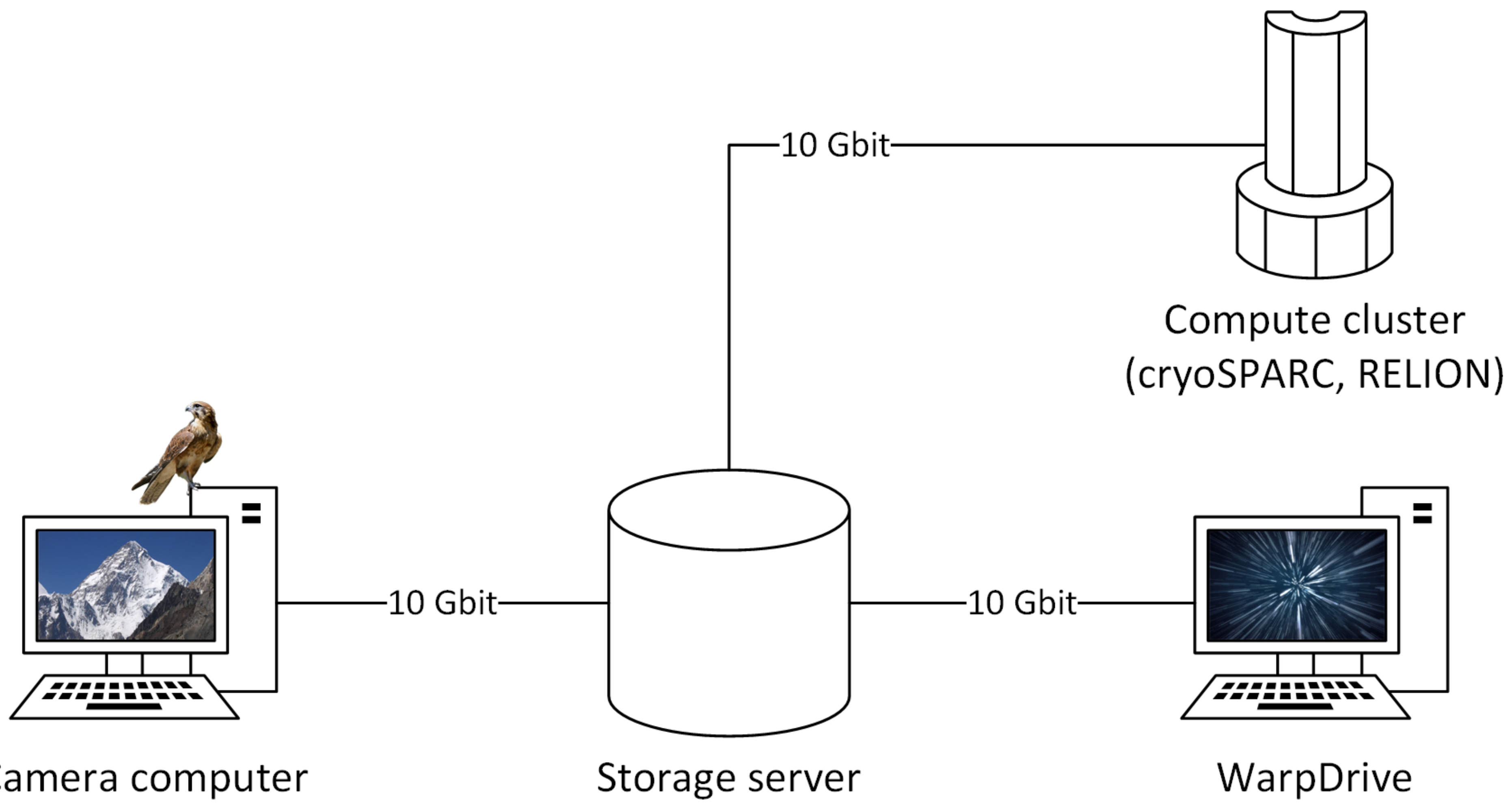


Deep learning-based picking: BoxNet

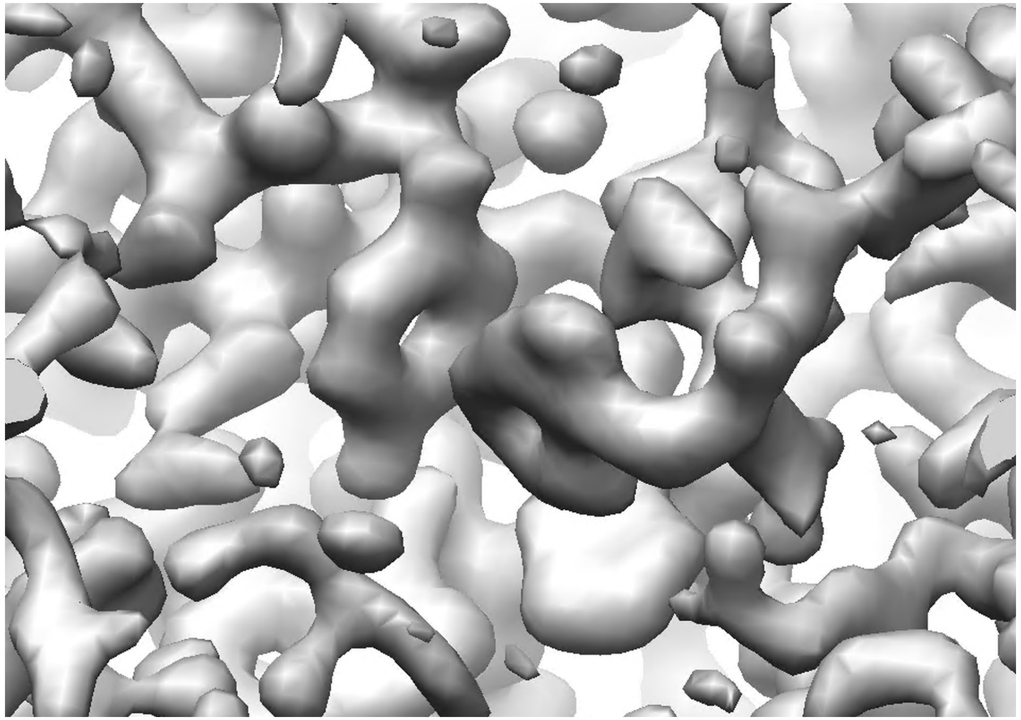
- Fully convolutional U-Net with residual blocks
- Pre-trained on 26 hand-picked data sets
- Easily retrainable within Warp
- Save and manage retrained models for projects



Infrastructure

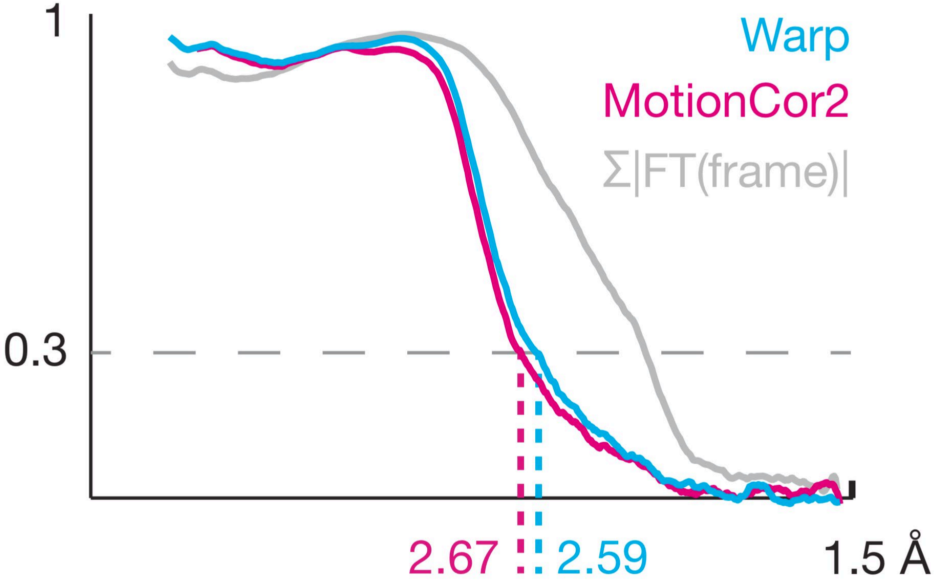


With EMPIAR-10061

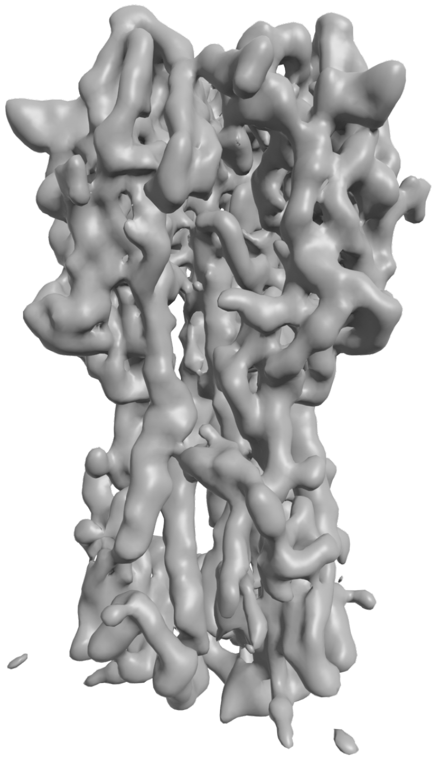


Warp + RELION: 2.09 Å
+ Beam tilt: 1.95 Å
+ Defocus: 1.95 Å
+ Polishing: 1.86 Å

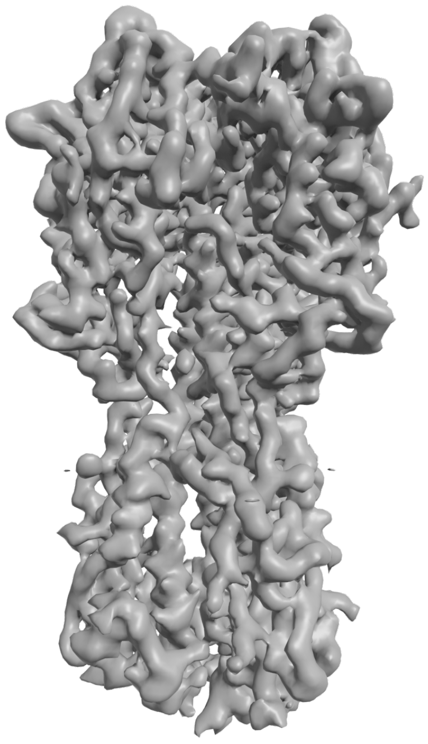
CTF fit



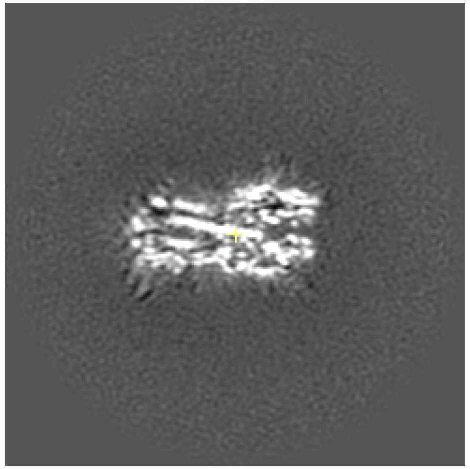
With EMPIAR-10097

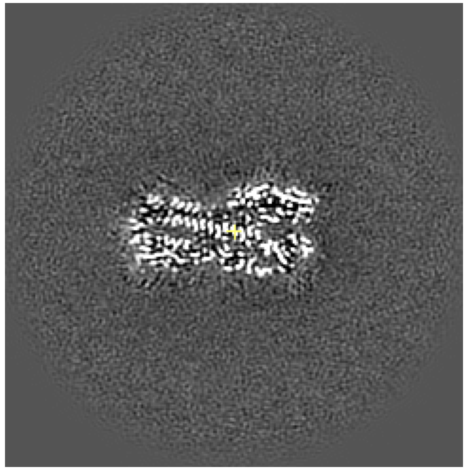


Original data, manual processing



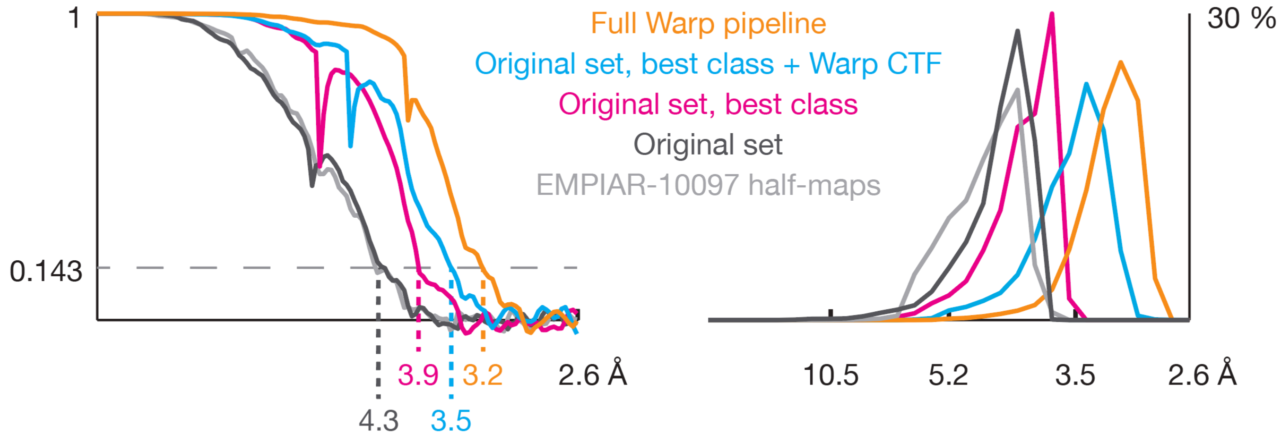
Warp pipeline, automated





Fourier shell correlation

Local resolution



Get it at warpem.com!

Denoising with deep neural nets

I have no idea why this works.

Noise2Noise: Learning Image Restoration without Clean Data

Jaakko Lehtinen [1 2] **Jacob Munkberg** [1] **Jon Hasselgren** [1] **Samuli Laine** [1] **Tero Karras** [1] **Miika Aittala** [3]
Timo Aila [1]

Abstract

We apply basic statistical reasoning to signal reconstruction by machine learning — learning to map corrupted observations to clean signals — with a simple and powerful conclusion: under certain common circumstances, it is possible to learn to restore signals without ever observing clean ones, at performance close or equal to training using clean exemplars. We show applications in photographic noise removal, denoising of synthetic Monte Carlo images, and reconstruction of MRI scans from undersampled inputs, all based on only observing corrupted data.

1. Introduction

Signal reconstruction from corrupted or incomplete measurements is an important subfield of statistical data analysis

have been reported in several applications, including Gaussian denoising, de-JPEG, text removal (Mao et al., 2016), super-resolution (Ledig et al., 2017), colorization (Zhang et al., 2016), and image inpainting (Iizuka et al., 2017). Yet, obtaining clean training targets is often difficult or tedious. A noise-free photograph requires a long exposure; full MRI sampling is slow enough to preclude dynamic subjects, etc.

In this work, we observe that under suitable, common circumstances, we can *learn to reconstruct signals from only corrupted examples, without ever observing clean signals*, and often do this just as well as if we were using clean examples. As we show below, our conclusion is almost trivial from a statistical perspective, but in practice, it significantly eases learning signal reconstruction by lifting requirements on the availability of clean data.

2. Theoretical background

Assume that we have a set of unreliable measurements

Noise2Noise: Learning Image Restoration without Clean Data

Jaakko Lehtinen [1 2] **Jacob Munkberg** [1] **Jon Hasselgren** [1] **Samuli Laine** [1] **Tero Karras** [1] **Miika Aittala** [3]
Timo Aila [1]

Abstract

We apply basic statistical reasoning t
construction by machine learning –
map corrupted observations to cl
with a simple and powerful con
certain common circumstances, i
learn to restore signals without
clean ones, at performance close o
ing using clean exemplars. We sho
in photographic noise removal, den
thetic Monte Carlo images, and reco
MRI scans from undersampled inpu
on only observing corrupted data.

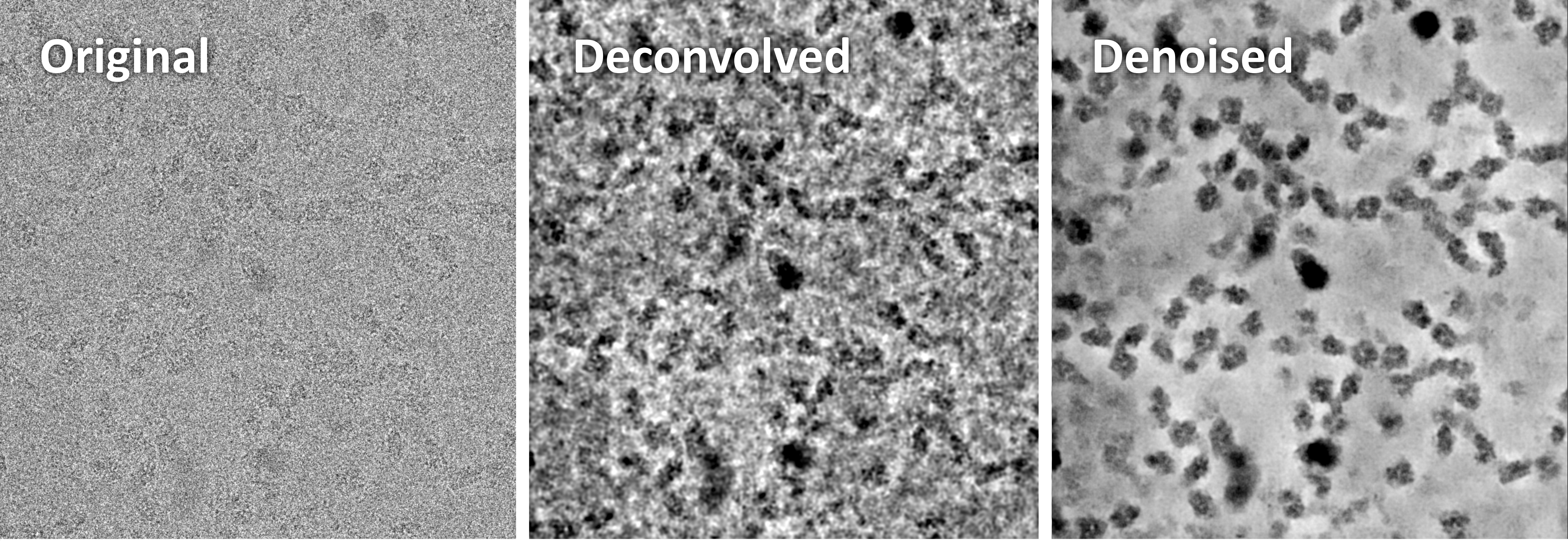
1. Introduction

Signal reconstruction from corrupted or incomplete mea-

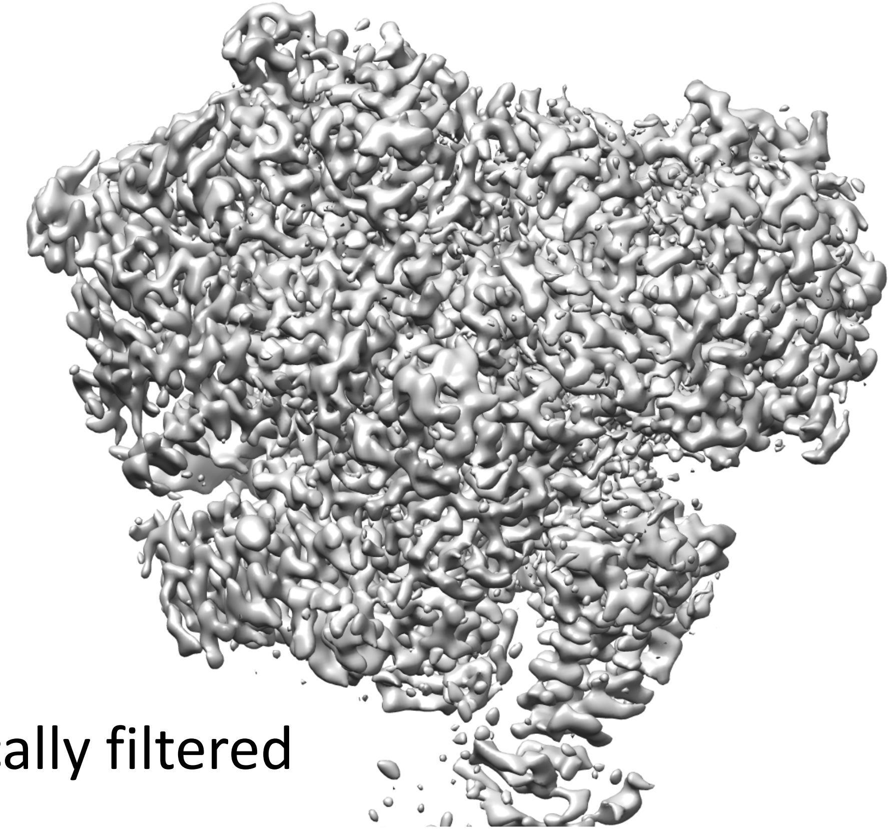
reported in several applications, including Gaus-
de-JPEG, text removal (Mao et al., 2016),
(Ledig et al., 2017), colorization (Zhang
mage inpainting (Iizuka et al., 2017). Yet,
ning targets is often difficult or tedious.
aph requires a long exposure; full MRI
ugh to preclude dynamic subjects, etc.

serve that under suitable, common cir-
learn to reconstruct signals from only
s, without ever observing clean signals,
just as well as if we were using clean ex-
ow below, our conclusion is almost trivial
perspective, but in practice, it significantly
nal reconstruction by lifting requirements
on the availability of clean data.

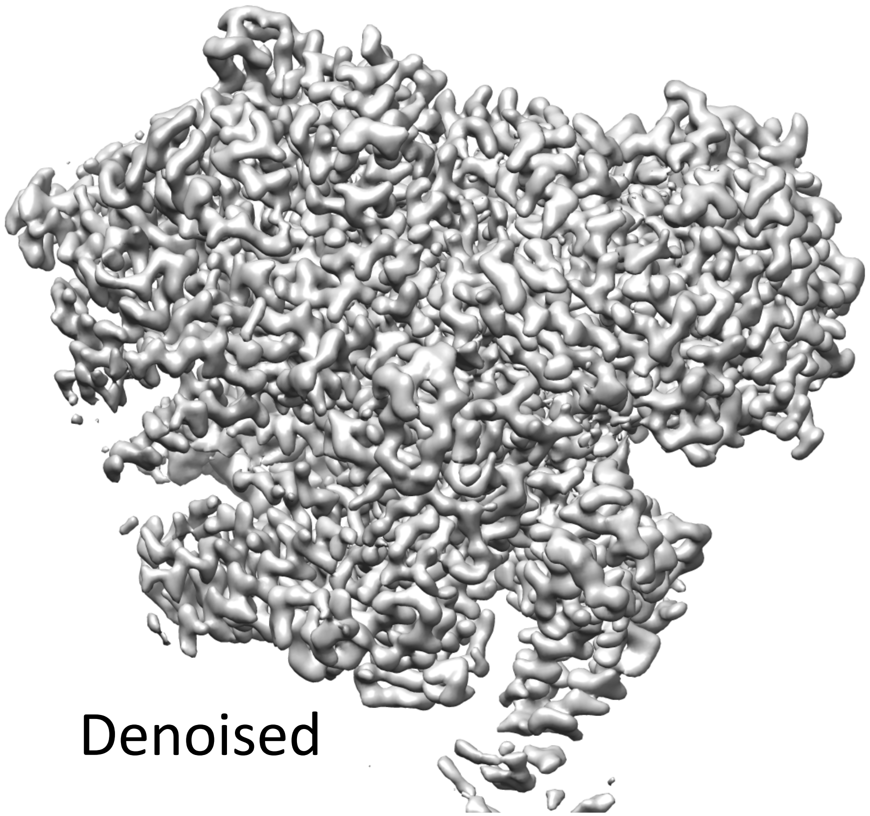
2. Theoretical background



Denoising: 2D (0.8 µm defocus)

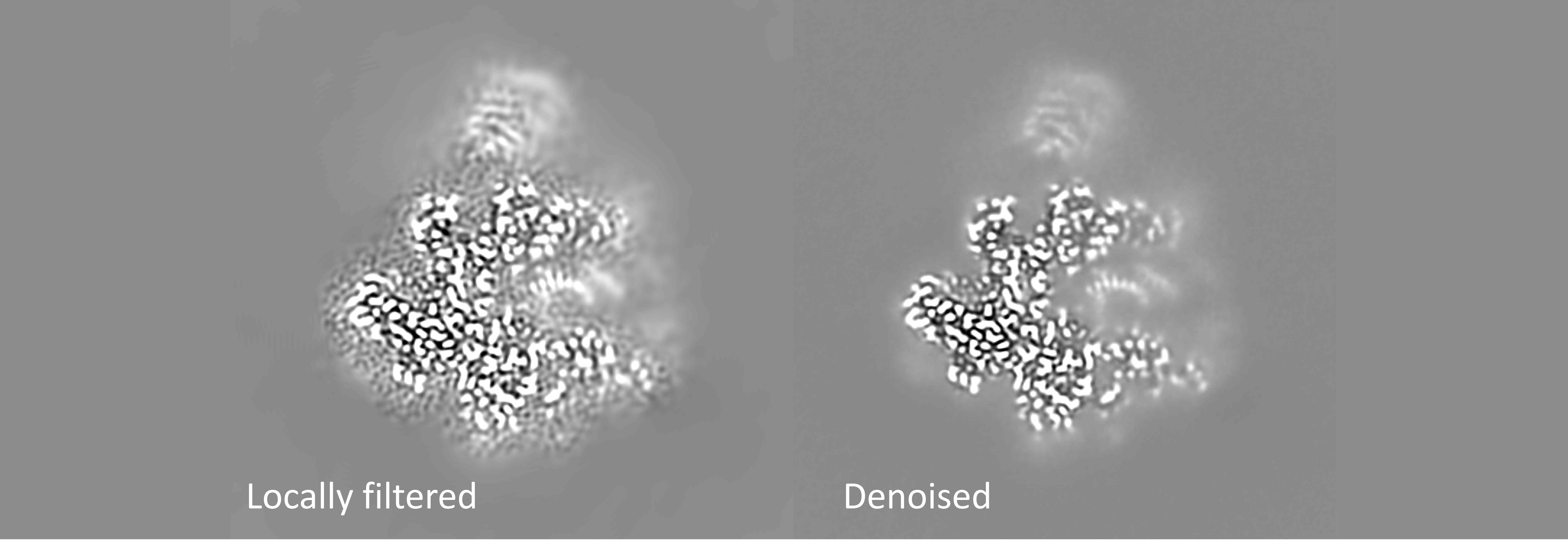


Locally filtered

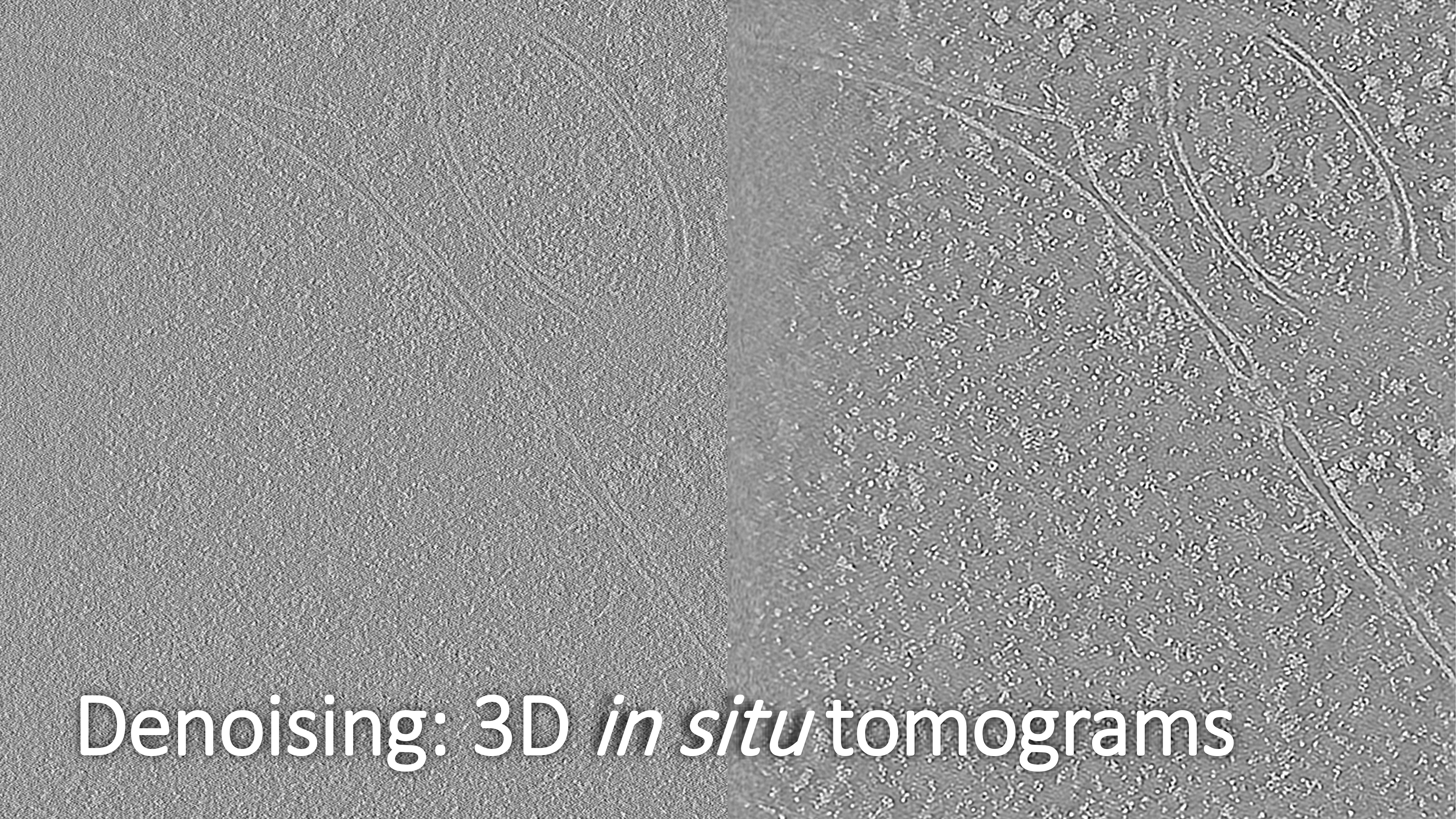


Denoised

Denoising: 3D half-maps



Denoising: 3D half-maps

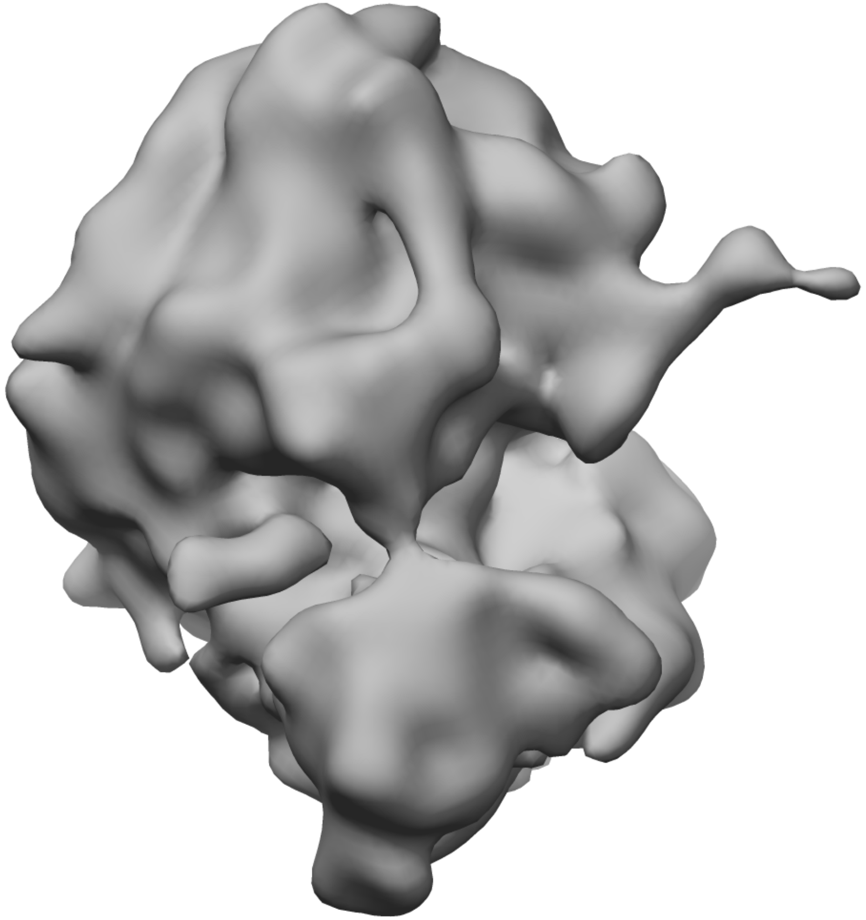


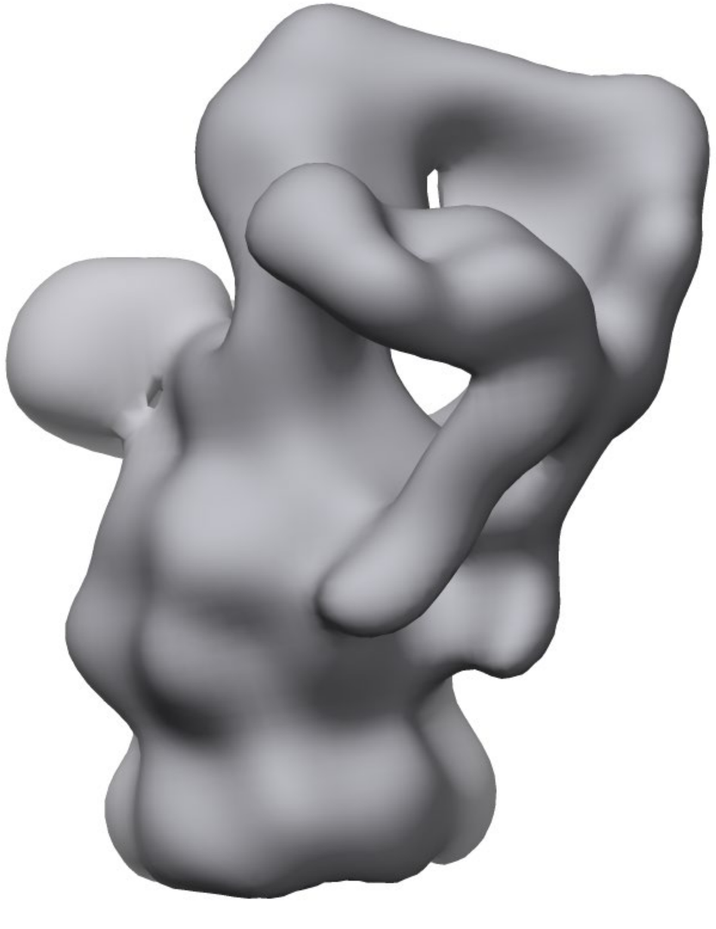
Denoising: 3D *in situ* tomograms

Where do we want to go?
Tomography.

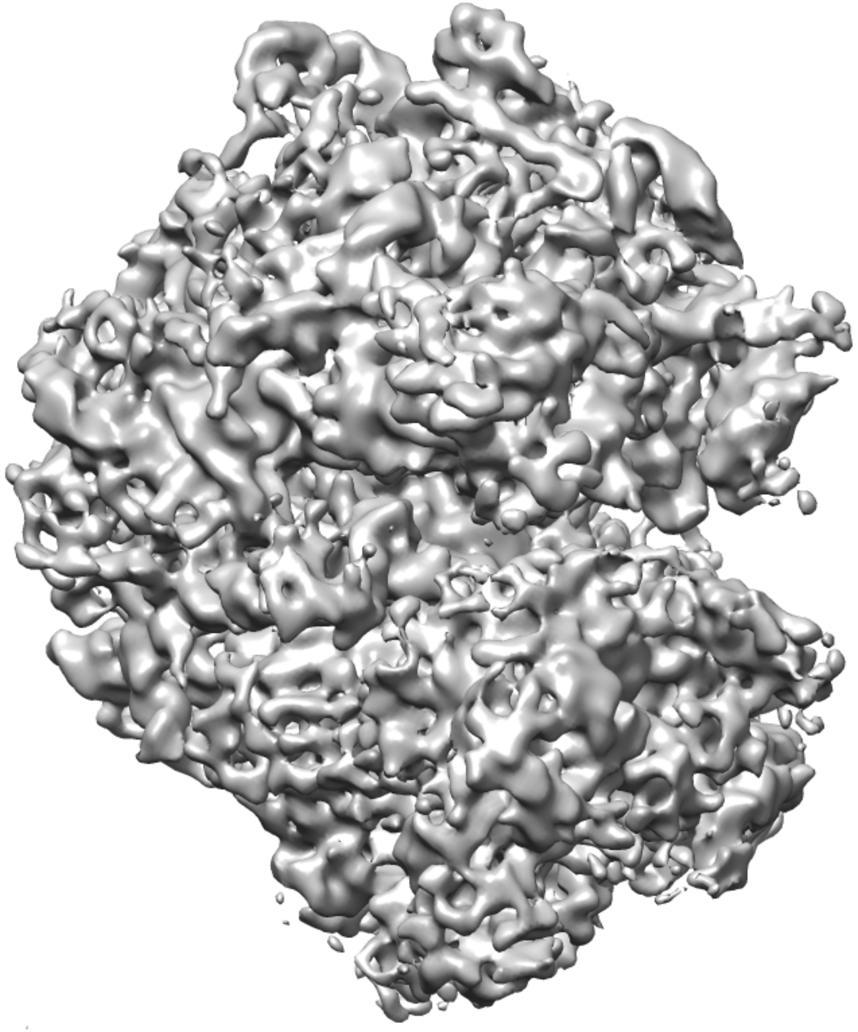
Ben Engel

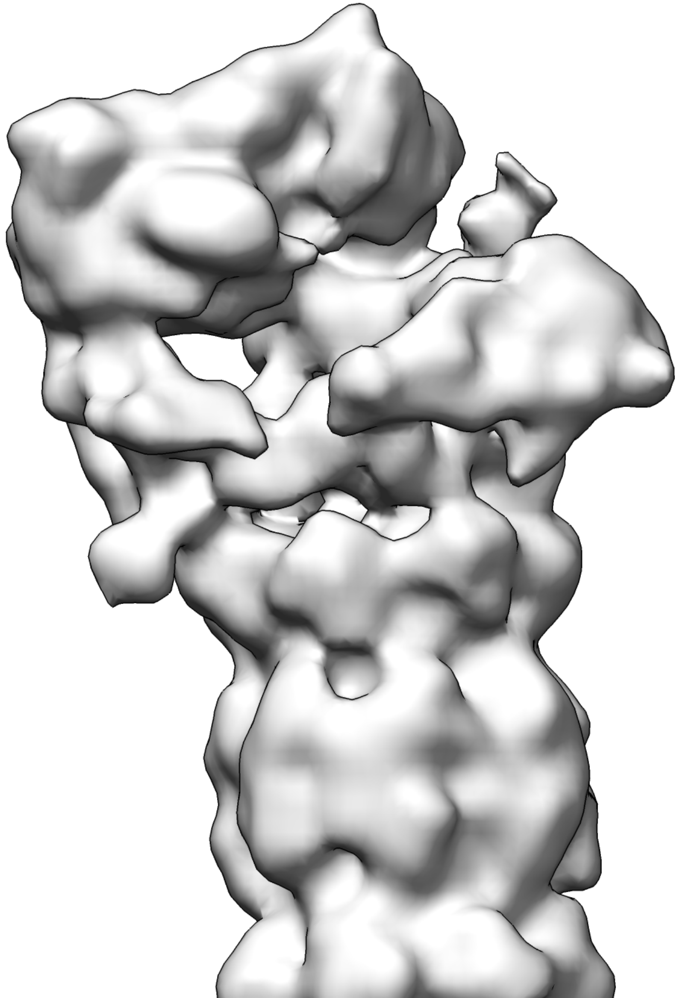
Previous results for cryo-FIBed *in situ*





Similar *in situ* data with Warp





Implications for data processing and sharing

- Every tomogram needs as many identified particles as possible
- Every particle species needs as many copies as possible
- Every *in situ* tomogram contains less than 1 particle of interest per lab
- … and 10 000+ particles of interest for other labs
- Everyone‘s resolution increases as more particles are added
- No single lab/facility will be able to produce enough data

Thanks to:
Patrick Cramer

Data from:
EMPIAR
Ben Engel
Carrie Bernecky

warpem.com
github.com/dtegunov
twitter.com/dtegunov