bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Topaz-Denoise: general deep denoising models for cryoEM and cryoET

Tristan Bepler[1,2], Kotaro Kelley[3], Alex J. Noble[3,*], and Bonnie Berger[2,4,*]

[1] Computational and Systems Biology, MIT, Cambridge, MA, USA

[2] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

[3] National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, NY, NY, USA

[4] Department of Mathematics, MIT, Cambridge, MA, USA

* Corresponding authors: bab@mit.edu and anoble@nysbc.org

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Cryo-electron microscopy (cryoEM) is becoming the preferred method for resolving protein structures. Low signal-to-noise (SNR) in cryoEM images reduces the confidence and throughput of structure determination during several steps of data processing, resulting in impediments such as missing particle orientations. Denoising cryoEM images can not only improve downstream analysis but also accelerate the time-consuming data collection process by allowing lower electron dose micrographs to be used for analysis. Here, we present Topaz-Denoise, a deep learning method for reliably and rapidly increasing the SNR of cryoEM images and cryoET tomograms. By training on a dataset composed of thousands of micrographs collected across a wide range of imaging conditions, we are able to learn models capturing the complexity of the cryoEM image formation process. The general model we present is able to denoise new datasets without additional training. Denoising with this model improves micrograph interpretability, enabling us to solve the first 3D single particle closed and partially open structures of clustered protocadherin, an elongated particle with previously-elusive views. We then show for the first time that low dose collection, enabled by Topaz-Denoise, improves downstream analysis in addition to reducing data collection time. We also present the first general 3D denoising model for cryoET. Topaz-Denoise and pre-trained general models are now included in Topaz. We expect that Topaz-Denoise will be of broad utility to the cryoEM community for improving micrograph and tomogram interpretability and accelerating analysis.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Visualization of micrographs from cryo-electron microscopy (cryoEM) of biological specimens is primarily limited by the phase contrast of proteins, the low electron dose conditions required due to radiation damage accrued by the proteins, and the thickness of the ice. As researchers push towards smaller and smaller proteins, these issues hinder downstream analyses because these proteins become increasingly difficult to distinguish from noise. Certain orientations of large, non-globular proteins can also have low signal, leading to missing views. The typical signal-to-noise ratio (SNR) of a cryoEM micrograph is estimated to be only as high as 0.1[1], amongst the lowest in any imaging field, and no ground truth exists. Nonetheless, several steps during collection and processing of micrographs in single particle cryoEM rely on properly human-inspecting micrographs, identifying particles, and examining processed data. Conventional cryoEM methods for improving contrast in micrographs include downsampling, bandpass filtering, and Wiener filtering[2,3]. However, these methods do not address the specific noise properties of micrographs and often do not provide interpretable results, which increasingly hinders attempts to resolve small and non-globular proteins[4,5].

At the same time, there is a push in the field to fund large research facilities for high-throughput cryoEM. These and smaller facilities are moving towards the synchrotron model of data collection and need to increase their throughput to meet rising demand. One approach to speed up collection would be to collect shorter micrograph exposures. However, reducing total dose would exacerbate SNR-related analysis problems. Better micrograph denoising provides the opportunity to reduce total dose and increase collection throughput without compromising interpretability or downstream results.

Image denoising has long been a topic of significant interest in the computer vision and signal processing community[6], but has recently seen a surge in interest from the machine learning community. Advances in deep neural networks have enabled substantial improvements in image restoration and inpainting (i.e. filling in missing pixels) by learning complex, non-linear priors over the applied image domain. However, these methods require ground truth images to provide supervision for learning the denoising model[7,8], and are hence limited to domains where ground truth is available. To overcome this barrier, Lehtinen et al.[9] presented a general machine learning (ML) framework, called Noise2Noise, for learning denoising models from paired noisy images rather than paired noisy and ground truth images. This method has been followed by several others for learning denoising models without ground truth[10–12]. These methods offer new approaches for training deep neural network models for denoising in challenging domains. In cryoEM, neural network denoising software has only just started to emerge for dataset-by-dataset cryo-electron tomogram (cryoET) denoising[13,14] and single particle micrograph denoising[15]. However, there have not been any systematic evaluations of these methods to date nor have pre-trained general denoising models been developed.

Here, we develop Topaz-Denoise, the first large-scale, publicly available denoising models for cryoEM and cryoET. Conventional cryoEM and cryoET denoising methods are ad-hoc filters that do not model the complex image generative process. To address this, our goal is to learn the denoising process directly from data. However, deep denoising models

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

typically require ground truth signal which is not available in cryoEM. We make the key insight that the individual movie frames collected by modern direct detector devices (DDD) are many independent observations of the same underlying signal and, hence, can be used to learn denoising models directly via the Noise2Noise framework. Trained on thousands of micrographs from DDD - K2, Falcon II, and Falcon III - across a variety of imaging conditions, these general models (also called pre-trained models) provide robust denoising without the need to train on a dataset-by-dataset basis. We test and compare these denoising models on several micrographs of typical particles and of small particles, study improvements in SNR, and use denoising combined with Topaz particle picking[16] to obtain the first 3D single particle cryoEM structures of clustered protocadherin, an elongated particle with previously-elusive views and a previously unseen conformation. We also show, for the first time, that denoising enables more rapid data collection by allowing micrographs to be collected with a lower electron total dose (10%–25% typical exposure times) without sacrificing interpretability or downstream processing. Shorter exposure times allow for higher throughput microscope usage, which reduces research cost and increases research efficiency. Additionally, we develop the first general 3D denoising model for cryoET tomograms, trained on dozens of cryoET tomograms, and show that our general denoising model performs comparably to models trained on a dataset-by-dataset basis. These models are integrated into Topaz allowing easy access to the community, along with the denoising framework that allows users to train their own cryoEM and cryoET denoising models.

Topaz-Denoise source code is freely available as part of Topaz (http://topaz.csail.mit.edu) and can be installed through Anaconda, Pip, Docker, Singularity, and SBGrid[17], and is now integrated into CryoSPARC[18], Relion[19], Appion[20] and Scipion[21]. As with Topaz, Topaz-Denoise is designed to be modular and can easily be integrated into other cryoEM software suites. Topaz-Denoise includes several pre-trained models and the ability for the user to train their own models. Topaz-Denoise 2D training and inference runs efficiently on a single GPU computer, while 3D training and inference runs efficiently on multi-GPU systems. Both 2D and 3D denoising are integrated into the standalone Topaz GUI to assist with command generation.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

1. Denoising with Topaz improves micrograph interpretability and SNR

We develop a general cryoEM micrograph denoising model by training a neural network using the Noise2Noise framework on dozens of representative datasets of commonly used imaging conditions (Figure 1, Methods). By learning the denoising model directly from data, we avoid making specific assumptions about the noise-generating process leading to superior denoising performance.

Denoising with Topaz improves micrograph interpretability by eye on several datasets and improves SNR measurements in quantitative analyses. Our model correctly smoothes background areas while preserving structural features better than conventional methods (i.e. affine or low-pass filtering) (Figure 1, Supplemental Figures 1-4). Given this known smoothing behavior of micrograph areas containing primarily noise, we find that denoising allows for identification of structured background features from noise. Figure 1 shows two micrographs where the background areas between particles are flattened after denoising, while Supplemental Figure 5 shows microtubules with known small proteins in background areas properly retained after denoising. Our denoising model has the combined advantage of reducing ambiguity as to whether the background of a micrograph is generally free from contamination, allowing researchers to identify small and/or low density particle views, for example as applied to micrographs from Mao et al.[22] (Supplemental Figure 6,7). In these types of scenarios, visual assessment of denoised micrographs compared to raw micrographs increases protein density confidence, increases confidence of background content, and reduces eye strain for researchers.

We quantitatively assess denoising performance by measuring the SNR of raw micrographs, micrographs denoised with our model, and micrographs denoised with conventional methods. We chose to measure SNR using real cryoEM micrographs because the denoising models were trained on real micrographs generated under real-world conditions that no software accurately simulates. Due to the lack of ground truth in cryoEM, SNR calculations are estimates (Methods). We manually annotated paired signal and background regions on micrographs from 10 different datasets (Supplemental Figure 8). We then calculated the average SNR (in dB) for each method using these regions[23]. We present a comparison of four different denoising model architectures (affine, FCNN, U-net (small), and U-net) trained with L1 and L2 losses on either the small or large datasets (Supplemental Table 1). Note that the L2 affine filter is also the Wiener filter solution. We find only minor differences between L1 and L2 models, with L1 loss being slightly favored overall. Furthermore, we find that the training dataset is important. Intriguingly, the affine, FCNN, and U-net (small) models all perform better than the full U-net model when trained on the small dataset and perform better than the same models trained on the large dataset. The best performing model overall, however, is the full U-net model trained on the large dataset.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This model also outperforms conventional low-pass filtering denoising on all datasets except for one, where they perform equivalently (EMPIAR-10005[24]).

A summary comparison is presented in Table 1, where we report SNR results on each dataset for the best overall performing low-pass filter (16x binning) with the L2 U-net trained on the large dataset and the L1 affine model trained on the small dataset. Our pre-trained U-net model improves SNR by >2 dB on average over low-pass filtering and improves SNR by roughly 20 dB (100 fold) over the raw micrographs. The model generalizes well across different imaging conditions, improving SNR on micrographs collected on K2, Falcon II, and Falcon III cameras as well as micrographs collected in super-resolution and counting modes.

To explore the broadness of our general U-net denoising model, we applied the model to several samples across several non-DDD cameras from two screening microscopes and analyzed them visually (Supplemental Figure 9). The pixelsizes for these datasets are about twice that of the training data and camera hardware binning by two has also been applied. Despite the differing noise characteristics of these cameras relative to the DDD cameras used for training the U-net denoising model, our general denoising model performs well. We see improvements similar to those noted above. Background is reasonably smoothed while the contrast of protein densities is greatly increased in the proteasome and two apoferritin micrographs. The glutamate dehydrogenase micrograph shows slight artifacts around some proteins, but contrast is substantially improved and denoising allows for clear identification of particle aggregates. These improvements demonstrate that our pre-trained denoising model even generalizes well to micrographs collected on screening microscopes and may enable increased cryoEM screening efficiency.

2. Denoising with the general model enables more complete picking of difficult particle projections

We denoised micrographs of particles with particularly difficult-to-identify projections, clustered protocadherin (EMPIAR-10234[25]), to test whether denoising enables these views and others to be picked more completely than without denoising. Figure 2 shows a representative micrograph before and after denoising. Before denoising, many particle top-views were indistinguishable by eye from noise (Figure 2a, left inset). After denoising, top-views in particular became readily identifiable (Figure 2a, right inset and circled in green).

We manually picked 1,023 particles while attempting to balance the percentage of side, oblique, and top-views of the particle in our picks. Using these picks, we trained a Topaz[16] picking model as described in the Methods. The resulting model was used to infer a total of 23,695 particles after CryoSPARC[18] 2D classification, 3D heterogeneous refinement to identify two conformations, and 3D homogeneous refinement on each conformation (Figure 2b, right). A closed conformation consisting of 13,392 particles confirmed the previous structure obtained using sub-tomogram alignment (EMD-9197)[25]. A new, yet unseen partially open conformation consisting of 8,134 particles was obtained (Figure 2b, yellow map), which exhibits a dislocation

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on one end of the dimer and an increased twist of the whole structure relative to the closed conformation. In comparison, using only the raw micrographs for initial manual picking, the data owner picked 1,540 particles to train a Topaz model as described in Brasch et al.[25] that inferred 10,010 particles in the closed conformation after CryoSPARC 2D classification and 3D refinement (Figure 2, left). Using Topaz-Denoise to help identify particles manually enabled us to resolve a novel conformation of clustered protocadherin from single particles, resulted in 2.15x more real particles picked, and substantially increased the percentage of top- and oblique-views (Figure 2c, Supplemental Figure 10).

Interestingly, CryoSPARC ab-initio reconstruction using a minimal set of denoised particles is less reliable than using the same set of raw particles (Supplemental Figure 11). Four or five of the six ab-initio reconstructions using the raw particles resulted in the correct overall structure, while only one of the six ab-initio reconstructions using the denoised particles resulted in the correct overall structure.

3. Denoising with the general model enables shorter exposure imaging

We simulated short exposure times at the microscope by truncating frames of several datasets used during frame alignment and summed to the first 10%, 25%, 50%, and 75% of the frames. These datasets were collected with a total dose of between 40 and 69 e-/Å^2. We denoised each short exposure with our general U-net model and compared both visually and quantitatively to low-pass filtering and to the raw micrographs without denoising.

Figure 3 shows denoised and low-pass filtered example micrographs of each subset along with the raw micrographs. Visual analysis and our SNR analysis suggests that between 10% and 25% of the exposure time is comparable to the full, raw micrographs (Figure 3, Supplemental Figure 12 for FFT, Supplemental Figures 13-16). This corresponds to between 4.0 and 16.7 e-/Å^2. 3D reconstructions of frame titrations of identical apoferritin particles from 19jan04d suggests that a total dose of about 16.7 e-/Å^2 is required for accurate CTF estimation (Supplemental Figure 17). Remarkably, reconstructions from these low-dose particle stacks reach resolutions surpassing that of the full dose particle stacks. This suggests that Topaz-Denoise can enable low-dose collection, particle identification, and thus high-resolution reconstruction in practice. Furthermore, roughly double the electron dose is required for low-pass filtering to match the SNR of our neural denoised micrographs. This could allow a factor of two or more savings in exposure time. Such a significant reduction in exposure time substantially increases the efficiency of cryoEM collection sessions, allowing for microscopes to operate at higher throughput.

To account for real-world collection overhead, we tested the optimal exposure dose for 19jan04d (~17 e-/Å^2) compared to a normal exposure dose (~66 e-/Å^2) on both a Titan Krios + Gatan K2 system and a Titan Krios + Gatan K3 system for a typical stage shift collection (1 exposure per hole) and a typical image shift collection (4 exposures per hole). Table 2 shows the results. On the K2 system with stage shift collection, using the optimal exposure dose is

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

about 65% more efficient than using the normal exposure dose (178 exposures per hour compared vs. 108). With image shift collection, using the optimal exposure dose is about 57% more efficient than using the normal exposure dose (190 exposures per hour compared vs. 121). On the K3 system with stage shift collection, using the optimal exposure dose is about 25% more efficient than using the normal exposure dose (242 exposures per hour compared vs. 195). With image shift collection, using the optimal exposure dose is about 15% more efficient than using the normal exposure dose (273 exposures per hour compared vs. 237). These results show that using the Topaz-Denoise general model to optimize exposure dose can allow for on the order of 1,000 more exposures per day to be collected on K2 and K3 systems.

4. Generalized 3D cryoET tomogram denoising markedly improves contrast, SNR, and interpretability

We converted the 2D Noise2Noise framework used in the previous sections to 3D for the purpose of creating a pre-trained general denoising model for cryo-electron tomograms (Methods). To train a general denoising model, we split 32 aligned cryoET tilt-series from FEI Titan Krios + Gatan K2 BioQuantum systems of cellular and reconstituted biological environments into even/odd frame tilt-series, binned each tilt-series by 2, reconstructed each tilt-series, and trained the neural network for over one month (Methods). The average pixelsize of the trained model, called Unet-3d-10a in the Topaz-Denoise package, is 10 angstroms. To further increase the broadness of 3D denoising in Topaz-Denoise, we trained a second general 3D denoising model called Unet-3d-20a using the same data as the Unet-3d-10a model, except with all training tomograms binned by another factor of 2 in Fourier space (ie. 20 Å pixelsize tomograms). Both general 3D denoising models are included in Topaz-Denoise.

To evaluate the resulting general 3D denoising model, we applied the model to one tomogram from each of the datasets used in the training and compared the results to models trained specifically on each test tomogram ('self-trained'), in addition to low-pass filtered tomograms (Supplemental Table 2). Comparisons were made both by SNR calculations using even/odd tomograms (Supplemental Table 2, Methods), and visually. Our pre-trained 3D U-net model (Unet-3d-10a) improves SNR by >3 dB over raw tomograms and improves SNR by about 1 dB on average over the best low-passed tomograms. Self-trained models showed only a marginal improvement in SNR over Unet-3d-10a. Figure 4a and Supplemental Video 1 show a visual comparison of one of the yeast tomograms used for training and testing. The Unet-3d-10a and self-trained models show a marked improvement in contrast and detail of ribosomes, RNA, ER proteins, mitochondrial transmembrane proteins, and aggregates over the raw and low-passed tomograms, while flattening background similar to the 2D U-net model for micrographs.

We next applied the Unet-3d-10a model to a sample unlike those it was trained on in several respects: an 80S ribosome single particle unbinned tomogram with a pixelsize of 2.17 Å and defocus of 4 microns, over four times less than the average pixelsize and half the average defocus of the tomograms used for training. A visual comparison of the applied model along with binned and Gaussian low-pass filtered tomograms is shown in Figure 4b and Supplemental

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Video 2. As with the previous 2D and 3D Topaz-Denoise general model results, the Unet-3d-10a model properly flattens background while increasing contrast of proteins relative to binning and low-pass filters. The increased contrast without tomogram resampling allows for visual delineation of objects of interest while retaining their higher-resolution information, and does not require ad-hoc parameter adjustment or training required by filtering methods more complicated than low-pass filtering. Furthermore, we show that applying a Gaussian filter after denoising further increases contrast, but at the expense of higher resolution information (Figure 4b and Supplemental Video 2, last tomogram). This may be useful if researchers wish to further increase contrast and do not require all frequencies to be visualized.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

CryoEM has long been hampered by the ability for researchers to confidently identify protein particles in all represented orientations from behind sheets of noise. Several bottlenecks in the general cryoEM workflow may preclude protein structure determination due to low SNR, such as differentiating protein from noise during picking, picking homogeneous subsets of particles, picking sufficient numbers of particles in all represented orientations, and obtaining a sufficient number of particles for 3D refinement. The initial stages of *de novo* protein structure determination are particularly affected by these issues.

To ameliorate these potentially critical issues, we present Topaz-Denoise, a Noise2Noise-based convolutional neural network for learning and removing noise from cryoEM images and cryoET tomograms (section 1). By employing a network trained on dozens of datasets to account for varying sample, microscope, and collection parameters, we achieve robust general denoising for cryoEM. We show empirically that our U-net denoising models result in higher SNR relative to affine models and low-pass filters. Topaz-Denoise enables visual identification of low SNR particle views, as exemplified by the clustered protocadherin dataset where denoising allows for more representative and complete 3D reconstructions, significantly more particles picked, and a never-before-seen conformation. This newly identified, partially 'open' conformation strongly suggest the possibility that assembly of protocadherin *cis*-dimers are preformed on membranes allowing rapid assembly of lattices and triggering of an avoidance signal when two cell surfaces with identical protocadherin complements come into contact. Increased confidence in particle identification using Topaz-Denoise, as shown in a particularly difficult case in results section 2, enables novel structures to be obtained from cryoEM data due to substantially increased particle picking completeness. Moreover, due to the considerable increase in SNR of denoised single particle micrographs, exposure time may be reduced without sacrificing the ability to pick particles reliably or perform downstream processing, thus enabling an increase in collection efficiency (section 3). Finally, implementing the same Noise2Noise-based network in 3D enables denoising of cryo-electron tomograms in minutes. As shown in both cellular tomograms and single particle tomograms in section 4, the 3D general denoising model in Topaz-Denoise properly smooths areas without signal and increases contrast of areas with signal without reducing the visual resolvability of features. This results in substantially higher SNR features both visually and quantitatively. Together, the U-net model for cryoEM and the Unet-3d models for cryoET in Topaz-Denoise offer superior denoising capabilities that are broadly applicable across instruments and sample types.

Conceptually, the Noise2Noise method as applied to micrographs in Topaz-Denoise trains a neural network to transform each denoised half-micrograph into the paired raw half-micrograph, and performs this training over thousands of half-micrograph pairs. We note that this is effectively learning an SNR maximizing transformation.. This follows from the relationship

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between the SNR of a micrograph and the correlation coefficient between paired independent measurements, which has been known and used in cryoEM since at least 1975[1,26]. This relationship, SNR = CCC/(1-CCC), where CCC is the correlation coefficient, has direct connection with the Noise2Noise objective in which we seek to find a transformation of the micrograph such that the error between transformed micrographs and the paired independent micrograph is minimized. In particular, the choice of L2 loss can be motivated through direct connection with the correlation. When both the denoised micrograph and raw micrograph are normalized to have mean zero and standard deviation one, the mean squared error (MSE) and correlation coefficient (CCC) are related by MSE = 2 - 2*CCC. This suggests a direct link between the MSE objective and SNR under the framework of Frank & Al-Ali[26].

Practically in our experience, because the general models were trained on large datasets from popular DDDs (Gatan K2, FEI Falcon II, and FEI Falcon III), these models provide the best visual results on comparable DDDs. For micrographs from microscopes and detectors used in the training dataset, we find that denoising typical featureful objects, such as proteins, continuous carbon, carbon/gold edges, and crystalline ice, increases their visual contrast, while denoising amorphous objects such as vitrified water results in visual flattening (Figure 1b,c, Supplemental Figure 12 for FFTs). Micrographs from non-DDD cameras still fare well compared to DDDs in our experience (Supplemental Figure 9) despite differing physical characteristics of the microscopes and detectors. This suggests that the general U-net model in Topaz-Denoise is robust to micrographs collected on equipment outside of the training dataset. Since non-DDD cameras are often used on screening microscopes, denoising these micrographs may increase screening throughput by allowing for more rapid analysis of micrographs, thereby increasing the efficiency of grid preparation steps. These results highlight three of the main advantages of our general denoising model: 1) users do not have to spend additional time training specific denoising models using their data, 2) for cameras that do not record frames, such as most screening microscope systems, acquiring data for training is not practical, thus a general denoising model is greatly prefered, and 3) the general model enables real-time denoising during data collection because denoising takes only seconds per micrograph.

The 3D cryoET denoising model included in Topaz-Denoise, and the framework which allows users to train their own models, may allow for improved data analysis not only in the cryoET workflow, but also the cryoEM workflow. In cryoET, researchers are often exploring densely-packed, unique 3D structures that are not repetitive enough to allow for sub-volume alignment to increase the SNR. The 3D denoising model shown here and included in the software increases the SNR of tomograms, which as a consequence may make manual and automated tomogram segmentation[27] easier and more reliable. In single particle cryoEM, we anticipate that the 3D denoising model and models trained on half-maps may be used to denoise maps during iterative alignment, as has previously been shown to be useful after alignment[28]. In our experience, training models on half-maps performs a form of local b-factor correction on the full map, which may allow for more reliable and accurate iterative mask generation during single particle alignment.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Models generated using Topaz-Denoise, including the provided general models, may be susceptible to the hallucination problem in neural networks[29]. Details in denoised micrographs or tomograms may exhibit imprints of encodings from the datasets used for training. Practically, this means that denoised particles should not be relied on for reconstruction as demonstrated in Supplemental Figure 11. We suspect that the issue here is two-fold: 1) cryoEM/ET refinement and reconstruction software assume noise distributions typical of raw data, not denoised data, and 2) denoised particles may present hallucinated information from the denoising model that is not detectable by visual inspection. For these reasons, we recommend that Topaz-Denoise models be used to assist with visualization and object identification, then the objects of interest be extracted and processed from raw micrographs/tomograms. Misuse of Topaz-Denoise and other opaque augmentations of raw data may result in subtle and difficult-to-detect forms of hallucinated signal[30,31].

As cryoEM and cryoET continue to expand into adjacent fields, researchers new to micrograph and tomogram data analysis will benefit from improved methods for visualization and interpretation of these low SNR data. Topaz-Denoise provides a bridge to these researchers, in addition to assisting those experienced in cryoEM and cryoET. We expect Topaz-Denoise to become a standard component of the micrograph analysis pipeline due to its performance, modularity, and integration into CryoSPARC, Relion, Appion, and Scipion.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Training dataset preparation for 2D denoising models

To train the denoising models, we collected a large dataset of micrograph frames from public repositories[32] and internal datasets at the New York Structural Biology Center (NYSBC), as described in Supplemental Table 3. These micrograph frames were collected under a large variety of imaging conditions and contain data collected on FEI Krios, FEI Talos Arctica, and JEOL CRYOARM300 microscopes with Gatan K2, FEI Falcon II, and FEI Falcon III DDD cameras at both super-resolution (K2) and counting modes and at many defocus levels. Including several microscopes, cameras, and datasets allows for robust denoising parameters to be modelled across common microscope setups.

We form two general aggregated datasets, one we call “Large” and one called “Small”. The “Large” dataset contains micrographs from all individual datasets. To roughly balance the contribution of the individual datasets in these aggregate datasets, we randomly select up to 200 micrographs from each individual dataset for inclusion rather than all micrographs. The Small dataset contains micrographs from individual datasets selected by eye based on the denoising performance of individually-trained U-net denoising models.

The Noise2Noise framework requires paired noisy observations of the same underlying signal. We generate these pairs by splitting the micrograph frames into even/odd frames which represent independent observations. These even/odd micrograph frames are then summed directly to form the paired observations. Because micrographs are typically motion corrected before summing and this motion correction procedure can change the noise distribution of the micrographs, we also form aligned, summed micrograph pairs by aligning the even/odd micrograph frames with MotionCor2[33] using 5 by 5 patches and a b-factor of 100. This resulted in 1,929 paired micrographs for the Small dataset and 3,439 paired micrographs for the Large dataset.

Model architectures

We adopt a U-Net model architecture[34] similar to that used by Lehtinen et al.[9] except that the input and output feature maps are 1-dimensional (n=1 to match monochrome micrographs) and we replace the first two width 3 convolutional layers of Lehtinen et al. with a single width 11 convolutional layer (Supplemental Figure 18) similar to other convolutional neural networks used in cryoEM[16]. This model contains five max pooling downsampling blocks and five nearest-neighbor upsampling blocks with skip connections between down- and up-sampling blocks at each spatial resolution. We refer to this as the U-net model. For comparison, we also consider a smaller U-net model with only 3 downsampling and upsampling blocks which we refer to as the U-net (small) model. We also compare with a fully convolutional neural network consisting of three convolutional layers of width 11x11 with 64 filters each and leaky rectified linear unit activations, termed FCNN, and an affine model with a single convolutional filter of width 31x31.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Loss functions and the Noise2Noise framework

The Noise2Noise framework takes advantage of the observation that we can learn models that recover statistics of the noise distribution given paired noisy observations of the same underlying signal. Given a ground truth signal, *y*, we observe images of this signal that have been corrupted by some probabilistic noise process, $x \sim Noise(y)$. Given paired noisy observations for matched signals, $x_a \sim Noise(y)$ and $x_b \sim Noise(y)$, we can learn a function that recovers statistics of this distribution. This is accomplished by finding parameters of the denoising function, *f* with parameters θ, such that the error between the denoised sample $f(x_a)$ and raw x_b are minimized. The form of this error function determines what statistics of the noise distribution we learn to recover. Given a dataset, *X*, containing many such image pairs, minimizing the L2 error over paired samples,

$$argmin_\theta \, E_{x_z, x_b \sim X}[\, \|f(x_a) - x_b\|_2^2 \,],$$

finds *f* with mean-seeking behaviour. Minimizing the L1 error over paired samples,

$$argmin_\theta \, E_{x_z, x_b \sim X}[\, \|f(x_a) - x_b\|_1 \,],$$

finds *f* with median-seeking behaviour. Finally, minimizing the L0 error over paired samples,

$$argmin_\theta \, E_{x_z, x_b \sim X}[\, \|f(x_a) - x_b\|_0 \,],$$

finds *f* with mode-seeking behaviour. This last objective is not differentiable and requires a smoothing term to minimize with standard gradient descent. We refer the reader to Lehtinen et al.[11] for details on this training objective. In practice, the errors of denoising both x_a and x_b are used for training.

Training details

For neural networks, weights are initialized using the default initialization in PyTorch[35]. For affine models, weights are initialized to zero. All models are trained using the Adagrad variant of stochastic gradient descent[24] with a learning rate of 0.001 for 100 epochs. We train on 800 by 800 patches randomly sampled from each micrograph using a minibatch size of 4. As data augmentation, these patches are randomly rotated and mirrored. In order to avoid interpolation artifacts, images are only rotated by 90, 180, or 270 degrees. Images are first normalized at the whole micrograph level by subtracting the mean pixel intensity and dividing by the standard deviation of the pixel intensities. Models were trained on a single NVIDIA V100 GPU with 32 GB of VRAM. Training took about 15 hours per model.

Inference details

Given a trained denoising model, we denoise full size micrographs. When operating on a GPU, RAM constraints may require denoising to be performed in patches. Here, we denoise in patches of 4,000 by 4,000 pixels. In order to avoid artifacts that can occur at the patch edges

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when stitched together, we include padding of 500 pixels around each patch when denoising. We note that this padding approach perfectly resolves patch edge effects when the padding is at least as large as half the receptive field of the model. Whole micrographs are first normalized by subtracting the mean and dividing by the standard deviation of the pixel intensities. The pixel intensities of the denoised micrograph are then restored by multiplying by the standard deviation and adding back the mean. Given the trained denoising model, inference is fast. We are able to denoise 4k by 4k K2 images at a speed of about 3 seconds/micrograph on a single NVIDIA 1080 Ti.

Micrograph scaling for figure visualization

In order to properly scale the pixel intensities between raw and denoised micrographs for visualization, we scale each micrograph to be relative to the intensities of the 16x low-pass filtered micrograph. This is achieved by subtracting the mean and dividing by the standard deviation of the pixel intensities in the low-pass filtered micrograph. This ensures that the relative signal levels are identical between all processed versions of the micrograph. Furthermore, to convert these into 256-bit values for display, we bin the real number pixel intensities into uniformly spaced buckets between -4 and 4.

Signal-to-noise quantification

We quantify the SNR of raw micrographs and processed micrographs in two ways: 1) based on paired labeled signal and background regions, and 2) using two independent measurements of the same signal. For the first method, we hand-labeled 20 signal and paired background regions across up to 10 micrographs from each dataset. We sought to label a variety of signal regions and to select paired background regions as close as possible to each signal region. Labeling was performed with reference to low-pass filtered micrographs in order to prevent any possible bias towards our denoising models. Given *N* signal, background region pairs, $x_s{}^i$, $x_b{}^i$, indexed by *i*, we calculate the mean and variance of each background region, $\mu_b{}^i$ and $v_b{}^i$. From this, the signal for each region pair is defined as $s^i = (\mu_s{}^i - \mu_b{}^i)^2$ where $\mu_s{}^i$ is the mean of signal region *i*. We take the mean over the signal region to be the signal in order to reduce noise and because we sought to label small, uniform signal regions. We then calculate the average SNR in dB for the regions,

$$SNR = \frac{10}{N} \sum_{i=1}^{N} log_{10}(s^i) - log_{10}(v_b{}^i) ,$$

which is reported for each dataset given raw and denoised micrographs. This is equivalent to ten times the log of the SNR. We report SNR this way for consistency with the denoising literature.

For the second, independent method of calculating SNR, we adopt the approach of Frank and Al-Ali[26] in which the SNR is estimated from two independent measurements of the same signal using the relationship $SNR = p/(1-p)$ where *p* is the cross-correlation coefficient between the two measurements. We use this to measure the SNR of denoising by splitting

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

micrographs into even and odd frames, frame aligning each independently, denoising the odd frame micrograph and then calculating the correlation between the denoised odd frame micrograph and the raw even frame micrograph. Based on this, we report the SNR in dB as

$$SNR = 10log_{10}(p/(1-p))$$

where *p* is now the cross-correlation between the denoised odd frame micrograph and the raw even frame micrograph. An advantage of this SNR calculation over our estimate above is that it considers all pixels in the micrograph and does not require labeling signal and background regions. The disadvantage is that it requires paired measurements and, therefore, can only be calculated for datasets where we have the raw frames. We also cannot calculate the SNR of the complete micrograph.

SNR quantification for cryo-electron tomograms is performed using the second method, where a ~1,000x1,000x150 pixel sub-volume of each tomogram containing biological objects is used to calculate SNR.

Short exposure micrograph processing

To quantify our ability to improve interpretability of low electron dose micrographs, we selected between five and ten random micrographs for the four datasets presented (EMPIAR-10234, 18sep08d, 19jan04d, and 19may10e). Micrographs from each dataset were split into five frame-fractionated subsets using IMOD's newstack program[36] to simulate short exposures: 10%, 25%, 50%, 75%, and 100%. Frames were aligned with Motioncor2 using 5x5 patches and dose weighting. For each dataset, SNR quantification was performed using the second method described above. For this quantification, micrographs were split into even and odd frames. The odd frames were then dose fractionated into 10%, 17.5%, 25%, 37.5%, and 50% of the full-micrograph total doses (rounded down to the nearest frame) and denoised. These low-dose denoised micrographs were compared with their corresponding full-dose even frame micrograph to calculate the SNR. The full-micrograph total doses for EMPIAR-10234, 18sep08d, 19jan04d, and 19may10e are 67.12 e-/Å^2, 39.6 e-/Å^2, 69.46 e-/Å^2, and 64.44 e-/Å^2, respectively.

Short exposure apoferritin processing

To quantify downstream results from frame titration, 100 random independently frame-aligned fractionated micrographs of 19jan04d were prepared using Motioncor2 without dose weighting. CTF estimation of the resulting 500 frame aligned micrographs was performed using CTFFind4[37] from within Appion[20]. 9,373 particles were picked from the micrographs using the first 10% of frames (the first 1 second of the exposures), an initial model was created in CryoSPARC, and the particles were refined through homogeneous refinement. The same particle picks and initial model were then used to extract and process the 25%, 50%, 75%, and 100% subsets through de novo homogeneous refinement while retaining each independent micrograph CTF estimation. 3DFSC[38] plots were then generated from the results.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Optimized Krios K2, K3 exposure time test

To quantify impacts on collection time using exposure doses optimized with Topaz-Denoise, we set up several collection sessions on FEI Titan Krios + Gatan K2 and FEI Titan Krios + Gatan K3 systems running Leginon collection. For each system, we used an apoferritin grid with 1.2/1.3 hole spacing, 300 mesh and collected four datasets to mimic common collection parameters: one with one normal exposure (~66 e-/Å^2) and one focus per hole, one with one optimized exposure (~17 e-/Å^2) and one focus per hole, one with four normal exposures (~66 e-/Å^2) and one focus per hole, and one with four optimized exposures (~17 e-/Å^2) and one focus per hole. On the K2 system, normal exposures took 10 seconds and optimized exposures took 2.5 seconds. On the K3 system, normal exposures took 2.5 seconds and optimized exposures took 0.65 seconds. Before each exposure was collected, the stage was set to settle for 1 second. Collection for each condition lasted 20-30 minutes to account for average overhead time during collection (focusing, stage movement, camera readout, etc.). Collection times were then extrapolated to exposures per hour and exposures per day. Additional overhead such as LN2 re-filling and targeting optimal sample areas were not taken into account.

EMPIAR-10234 clustered protocadherin single particle processing

We processed the EMPIAR-10234 clustered protocadherin dataset in two seperate ways to test the whether picking in denoised micrographs was advantageous: First by using the particle picks provided by the data owner, and second by manually picking on the denoised micrographs.

The picking method used by the data owner is described in Brasch et al.[25] Briefly, 1,540 particles were manually picked by the data owner from 87 raw micrographs and used to train a Topaz[16] picking model, resulting in 14,569 particles. The following reconstruction workflow was performed in CryoSPARC v2[18] using C1 symmetry in every step and using frame-summed particles for consistency. 2D classification was performed three times to remove obvious non-particle classes, resulting in 13,739 particles. Ab-initio reconstruction with 2 classes was performed, resulting in one good class with 10,010 particles. Homogeneous refinement was performed resulting in the final reconstruction.

The picking method we used is as follows. Frame-summed micrographs were denoised with the Topaz-Denoise v0.2.1 L2 model, proprocessed with `topaz preprocess` while binning by a factor of 4, and 1,023 particles were manually picked *not* by the data owner from 215 denoised micrographs. A Topaz[16] picking model was trained using the particle coordinates on raw micrographs, resulting in 59,273 particles. The following reconstruction workflow was performed in CryoSPARC v2 using C1 symmetry in every step and using frame-summed particles for consistency. 2D classification was performed three times to remove obvious non-particle classes, resulting in 44,303 particles. Ab-initio reconstruction with 2 classes was performed, resulting in one good class with 23,695 particles. Heterogeneous refinement with 2 classes was performed, resulting in two classes with different conformations. 2D classification and heterogeneous refinement with these two classes and a junk class were performed, resulting in 13,392 particles in the closed conformation and 8,134 particles in the partially open

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conformation. Homogeneous refinement was performed on each class, resulting in the final reconstructions. UCSF Chimera[39] was used to render the final reconstructions.

Clustered protocadherin low particle number single particle processing

Denoising and picking were performed as described in the last paragraph of the previous section. Then 1,000 random particles were chosen and processed through CryoSPARC v2 ab-initio reconstruction six times using the raw particles and six times using the particles denoised by the v0.2.1 L2 model. Comparisons between the full 3D map and each set of six ab-initio models were made in UCSF Chimera[39].

Cryo-electron tomography and 3D denoising

Saccharomyces uvarum (*S.uvarum*) was grown to log phase in YPD media, applied to freshly glow discharged carbon/copper 200 mesh Quantifoil grids, back blotted with the Leica GP, and plunge-frozen into liquid ethane. Grids were clipped and loaded into a Helios NanoLab 650 dual beam. Milling protocol followed previously published methods[40] yielding lamellae ranging in thickness between ~100-150nm. Tomograms were collected with a FEI Titan Krios equipped with an energy filter and a Gatan K2 BioQuantum operated in counting mode with the following parameters: energy filter 30eV slit, ~18 µm defocus, 6.08 Å pixelsize, -48.0° to 48.0° tilt with a 2° increment, total dose of ~60 e-/Å^2 . Tomograms were collected with a bi-directional scheme with an unreleased hybrid track/predict method[41] implemented in Leginon[42]. Raw frames were aligned with MotionCor2[33] and split into even/odd pairs before summation. Whole tilt-series were aligned with Appion-Protomo[43,44], then split into even/odd frame tilt-series, and reconstructed with Tomo3D[45,46] weighted back-projection, yielding even/odd half tomograms. Thirty two tomograms of cellular and reconstituted environments collected on FEI Titan Krios + Gatan K2 BioQuantum systems with an average pixelsize of 5.2 Å (range: 1.8 - 6 Å) and defocus of 9 µm (range: 4 - 18 µm) were used for training the pre-trained general denoising models, which was performed in parallel across seven Nvidia GTX 1080 GPUs. Training the model for binned-by-two tomograms (Unet-3d-10a) took over one month and training the model for binned-by-four tomograms (Unet-3d-20a) took ten days. Denoising a 3GB binned-by-two tomogram takes about five minutes on an Nvidia RTX Titan GPU and an Nvidia RTX 2080 Ti GPU in parallel. Tomography Figures and Supplemental Videos were made with Imod[36].

The tomogram denoising model uses a U-net architecture identical to the 2D U-net presented above, except that the 2D convolutions are replaced with 3D convolutions to operate on the tomogram voxel grids and the first convolutional kernel width is 7 voxels rather than 11. Training and inference methods are otherwise identical except for the patch size, which is set to 96 for tomograms to fit in GPU RAM.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Code availability statement

Source code for Topaz-Denoise is publicly available as part of Topaz (v0.2.0 and above for 2D denoising, v0.2.4 and above for 3D denoising) on GitHub at https://github.com/tbepler/topaz. Topaz is installable through Anaconda, Pip, Docker, Singularity, SBGrid, and source. Topaz is licensed under the GNU General Public License v3.0.

Data availability statement

The general models used in this manuscript are included as options in Topaz-Denoise. Over 100 NYSBC dataset frames used for some of the models have been deposited to EMPIAR-XXXXX. The closed and partially open clustered protocadherin maps have been deposited to EMD-22060 and EMD-22059, respectively. The apoferritin maps for 10%, 25%, 50%, 75%, and 100% of the total dose have been deposited to EMD-22052, EMD-22053, EMD-22056, EMD-22057, and EMD-22058, respectively.

Acknowledgements

The authors wish to thank Simons Electron Microscopy Center (SEMC) OPs for many of the test datasets used in training and the authors of the EMPIAR entries listed in Supplemental Table 3 for additional training datasets. The authors wish to thank Dr. Julia Brasch for sharing her processing experience with EMPIAR-10234. The authors wish to thank Dr. Sargis Dallakyan for integrating Topaz-Denoise into Appion, the CryoSPARC team for integrating Topaz-Denoise into CryoSPARC, and Dr. Grigory Sharov for integrating Topaz-Denoise into Scipion. The authors wish to thank Dr. Michael Rout for the yeast specimen used for 3D cryoET denoising. The authors wish to thank Drs. Anchi Cheng, Mykhailo Kopylov, Bridget Carragher, and Clinton Potter for helpful discussions. This work is dedicated to the memory of Kotaro Kelley.

T.B. and B.B. were supported by NIH grant R01-GM081871. A.J.N. was supported by a grant from the NIH National Institute of General Medical Sciences (NIGMS) (F32GM128303). The cryoEM work was performed at SEMC and National Resource for Automated Molecular Microscopy located at NYSBC, supported by grants from the Simons Foundation (SF349247), NYSTAR, and the NIH NIGMS (GM103310) with additional support from the Agouron Institute (F00316) and NIH (OD019994).

Author contributions

T.B., A.J.N., and B.B. conceived of this project. T.B. developed and implemented the models. K.K. prepared and collected the cryo-FIB/SEM-ET data and implemented 3D denoising. T.B., K.K., and A.J.N. processed and analyzed the data and model results. T.B. and A.J.N. processed and analyzed single particle results. T.B., K.K., A.J.N., and B.B. designed the experiments and wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1: Training framework and denoising visual comparisons

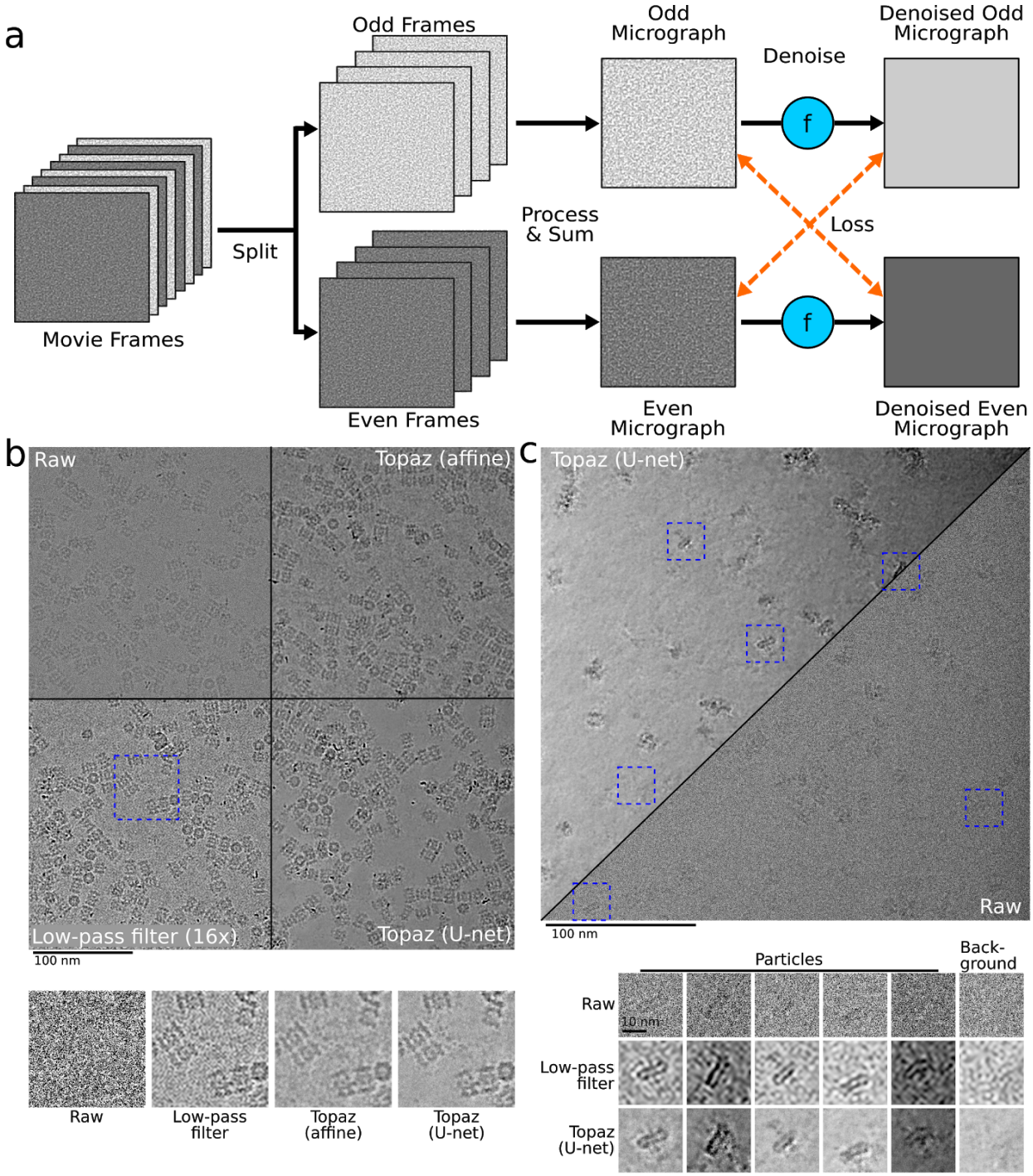


Figure 1 | Illustration of the training framework and comparison of denoising methods on two example micrographs. **(a)** The Noise2Noise method requires paired noisy observations of the same underlying signal. We generate these pairs from movie frames collected in the normal cryoEM process, because each movie frame is an independent sample of the same signal. These are first split into even/odd movie

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

frames. Then, each is processed and summed independently following standard micrograph processing protocols. The resulting even and odd micrographs are denoised with the denoising model (denoted here as f). Finally, to calculate the loss, the odd denoised micrograph is compared with the raw even micrograph and vice versa. **(b)** Micrograph from EMPIAR-10025 split into four quadrants showing the raw micrographs, low-pass filtered micrograph by a binning factor of 16, and results of denoising with our affine and U-net models. Particles become clearly visible in the low-pass filtered and denoised micrographs, but the U-net denoising shows strong additional smoothing of background noise. A detail view of the micrograph is highlighted in blue and helps to illustrate the improved background smoothing provided by our U-net denoising model. **(c)** Micrograph from EMPIAR-10261 split into the U-net denoised and raw micrographs along the diagonal. Detail views of five particles and one background patch are boxed in blue. The Topaz U-net reveals particles and reduces background noise.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: Denoising SNR comparisons

	EMPIAR datasets			NYSBC K2 datasets				NYSBC Falcon III datasets			
Method	10261[47]	10005[24]	10025[48]	Protocadherin	18sep08d	19jan04d	19may10e	18aug17l	18sep06d	18sep19l	Overall
Affine (Topaz)	5.49	1.29	0.72	4.83	4.51	**8.87**	12.02	10.65	6.90	9.15	6.44
U-net (Topaz)	**7.17**	**1.72**	**1.07**	**5.94**	**6.06**	8.43	**13.07**	**15.17**	**7.37**	**13.24**	**7.92**
Low-pass	5.19	-0.12	-0.40	4.22	3.53	6.87	9.99	9.04	6.95	8.71	5.40
Raw	-17.14	-20.13	-24.15	-14.47	-15.40	-11.73	-5.44	-6.33	-3.64	-5.63	-12.41

Table 1 | Comparison of micrograph denoising methods based on estimated SNR (in dB, larger is better). SNR was estimated from 20 paired signal and background regions selected for each dataset. In each column, the best performing model is highlighted. We report denoising results on aligned micrographs for the NYSBC K2 and Falcon III datasets. All datasets were collected in electron counting modes, except for 18sep06d which was collected using Falcon III integrating mode. Our U-net denoising model performs best overall and is best on all except for the 19jan04d dataset where our affine denoising model slightly outperforms it. We report low-pass filtering by a binning factor of 16 on all datasets, which we found to give better SNR overall compared to Gaussian low-pass filtering.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2: General denoising model makes picking elusive views possible

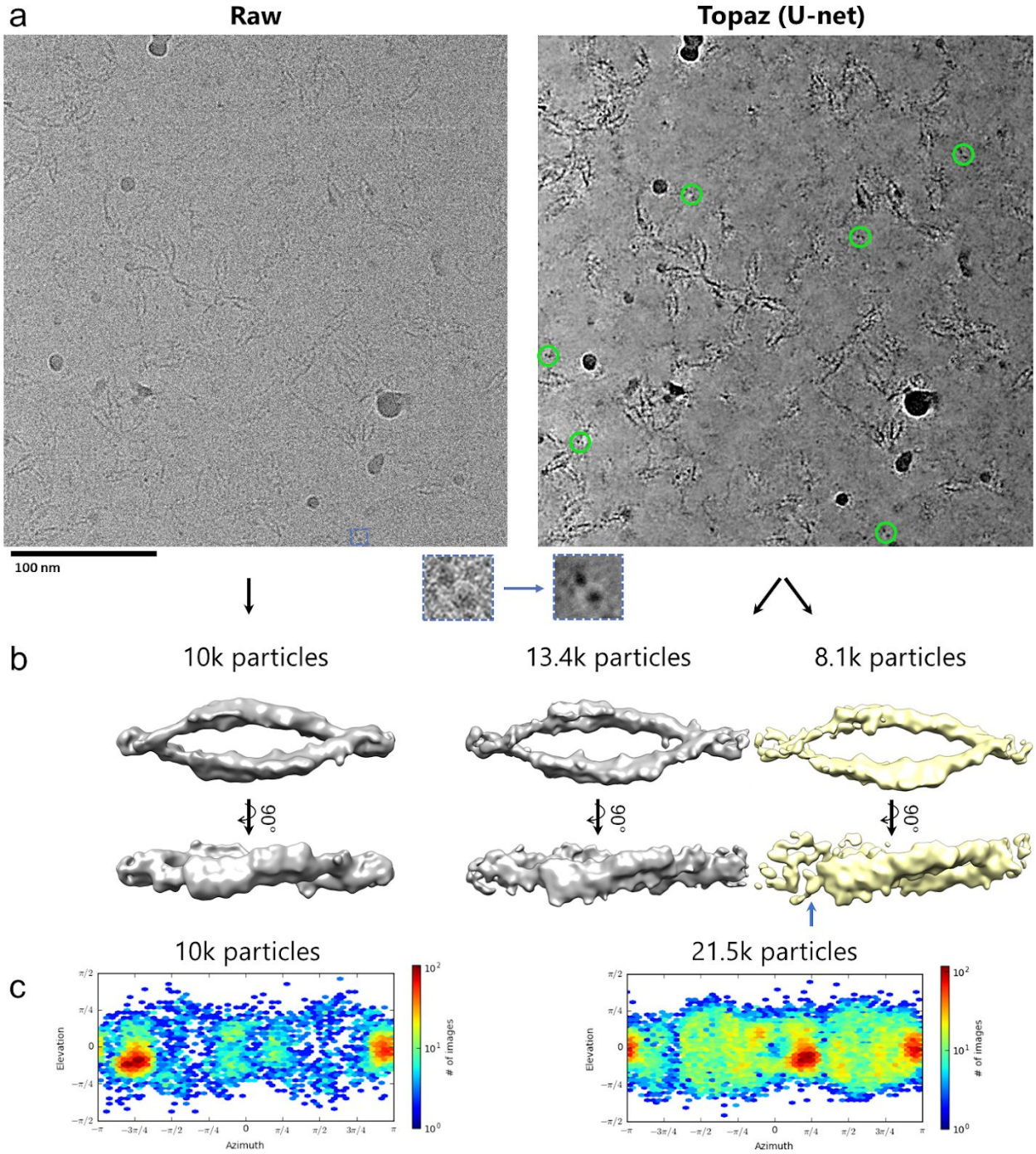


Figure 2 | Denoising with the general model in Topaz improves interpretability and picking of difficult particle projections. **(a)** A raw micrograph (left) and Topaz-Denoised micrograph (right) of the clustered protocadherin dataset (EMPIAR-10234) with a top-view boxed out (insets). Denoising allows for top-views to be clearly identified (green circles, right) and subsequently used to increase the confidence and completion of particle picking. **(b)** Topaz picking training on raw micrographs using 1,540 manually

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

picked particles from the raw micrographs resulted in the reconstruction on the left. Topaz picking training on the raw micrographs using 1,023 manually picked particles from the denoised micrographs resulted in the reconstruction on the right. Manually picking on denoised micrographs resulted in 115% more particles in the 3D reconstruction, which allowed for classification into a closed (gray) and partially open (yellow; blue arrow showing disjoint) conformation. **(c)** 3D reconstruction particle distributions for (left) Topaz picking training on raw micrographs using 1,540 manually picked particles from the raw micrographs, and (right) Topaz picking training on the raw micrographs using 1,023 manually picked particles from the denoised micrographs. All particles from the two classes in (b, right) are shown (c, right). 3DFSC plots for the three maps shown here are in Supplemental Figure 14.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3: General denoising model allows for substantially shorter exposures

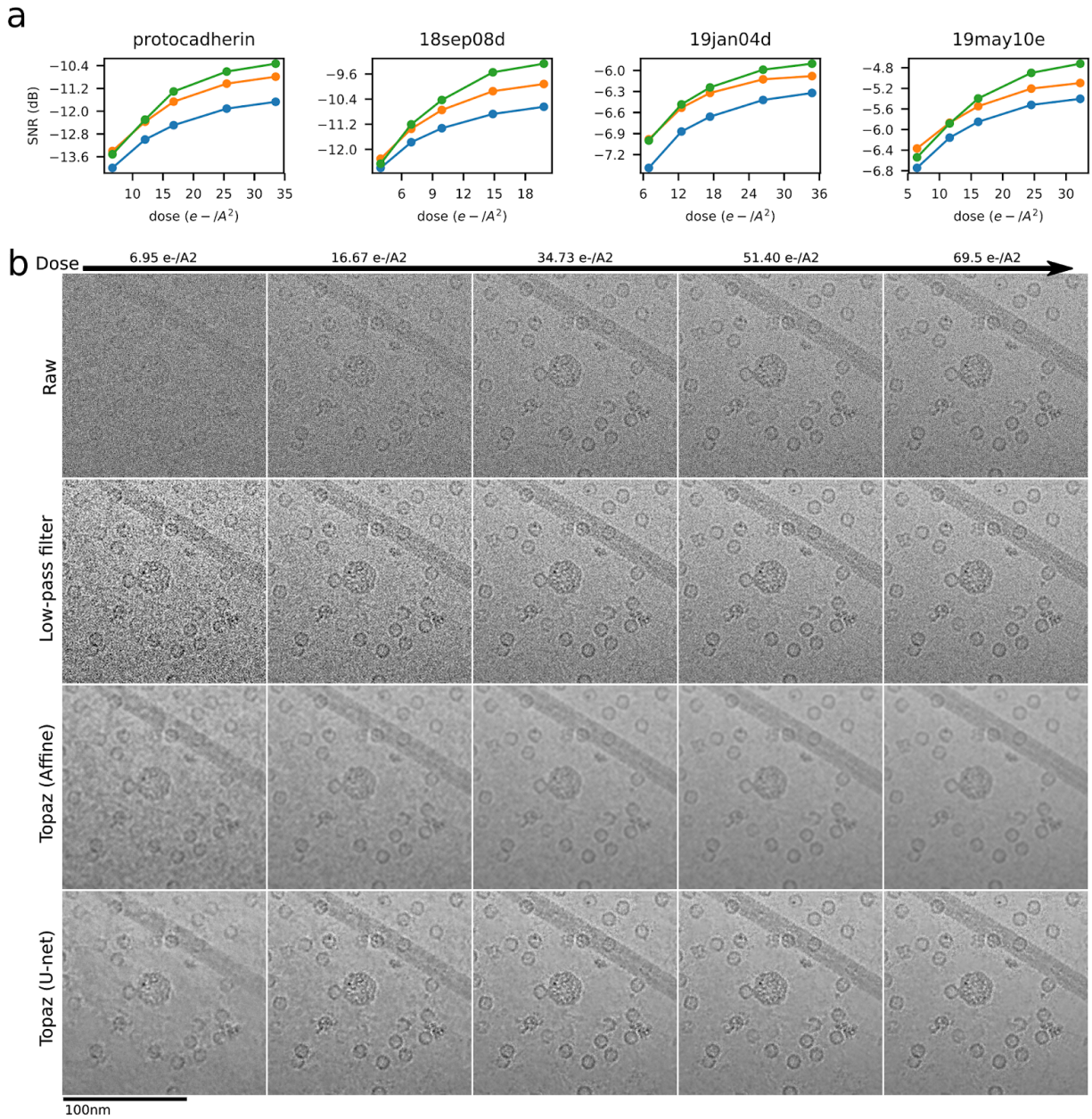


Figure 3 | Denoising with the general model in Topaz enhances SNR of short exposure micrographs. **(a)** SNR (dB) calculated using the split-frames method (see Methods) as a function of electron dose in low-pass filtered micrographs by a binning factor of 16 (blue), affine denoised micrographs (orange), and U-net denoised micrographs (green) in the four NYSBC K2 datasets. Our U-net denoising model enhances the SNR of micrographs across almost all dosages in all four datasets. U-net denoising enhances SNR by a factor of 1.5x or more over low-pass filtering at 20 e-/A^2. **(b)** Example section of a micrograph from the 19jan04d dataset of apoferritin, β-galactosidase, a VLP, and TMV (full micrograph

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in Supplemental Figure 3,4) showing the raw micrograph, low-pass filtered micrograph, affine denoised micrograph, and U-net denoised micrograph over increasing dose. Particles are clearly visible at the lowest dose in the denoised micrograph and background noise is substantially reduced by Topaz denoising.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

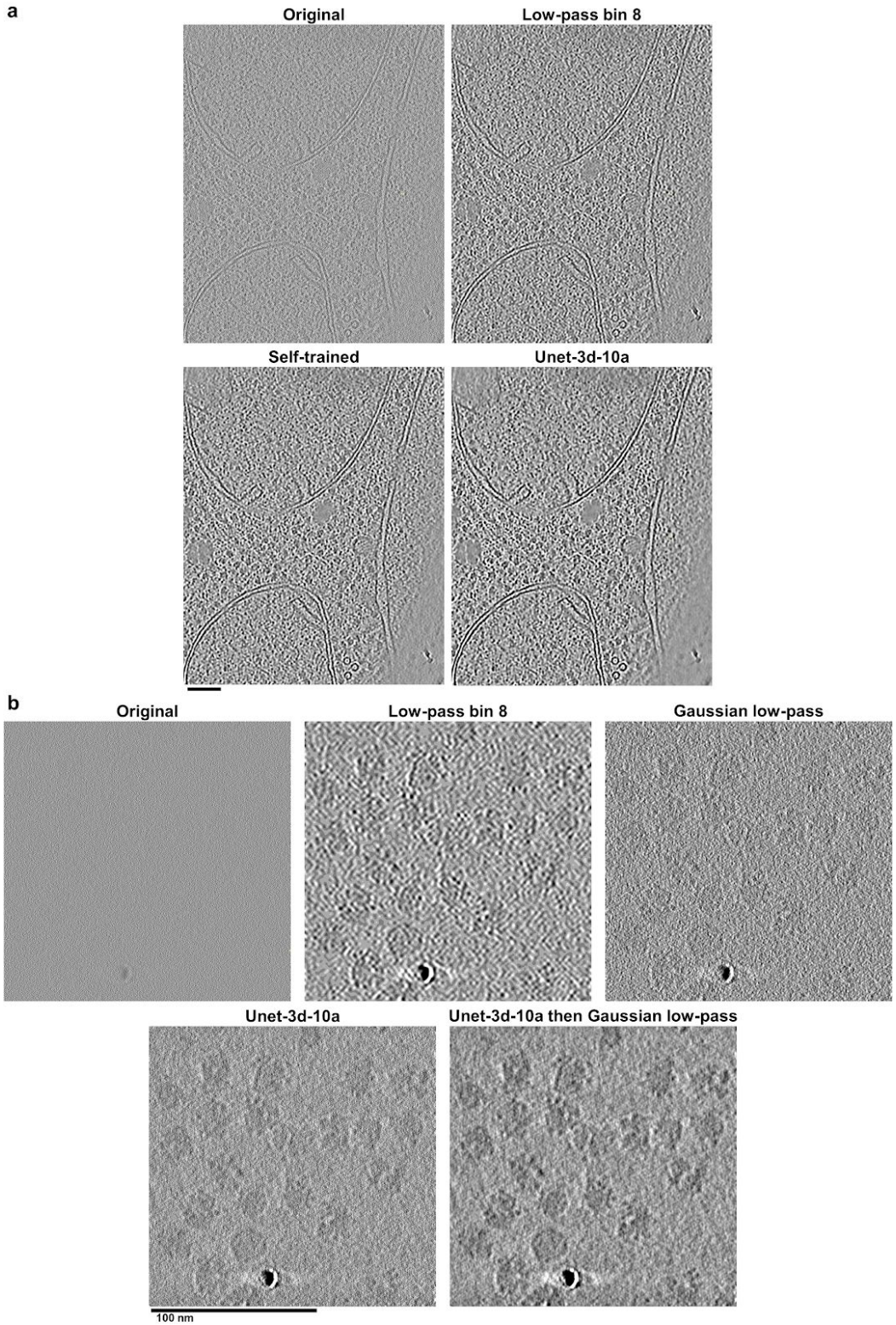
Table 2: Short exposure collection comparison

	Krios + K2 collection		Krios + K3 collection	
	Exposures per hour with 1 exposure/hole	Exposures per hour with 4 exposures/hole	Exposures per hour with 1 exposure/hole	Exposures per hour with 4 exposures/hole
Normal dose (~66 e-/Å^2)	108	121	195	237
Optimized dose (~17 e-/Å^2)	178	190	242	273
% increase in collection efficiency	65%	57%	25%	15%

Table 2 | Comparison between normal dose exposures versus optimized dose exposures using typical high-resolution microscope systems and parameters.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4: 3D general denoising model increases contrast without sacrificing resolvability



bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4 | Denoising with the 3D general model in Topaz improves cellular and single particle tomogram interpretability. **(a)** Saccharomyces uvarum lamellae cryoET slice-through (collected at 6 Å pixelsize and 18 microns defocus, then binned by 2). The general denoising model (Unet-3d-10a) is comparable visually and by SNR to the model trained on the tomogram's even/odd halves (Self-trained). Both denoising models show an improvement in protein and membrane contrast over binning by 8 while confidently retaining features of interest, such as proteins between membrane bilayers. Both denoising models also properly smooth areas with minimal protein/membrane density compared to the binning by 8. See Supplemental Video 1 for the tomogram slice-throughs. **(b)** 80S ribosomes as single particles (EMPIAR-10045; collected at 2.17 Å pixelsize and 4 microns defocus). The general denoising model (Unet-3d-10a) is markedly improved over binning by 8 and the ⅛ Nyquist Gaussian low-pass, both with smoothing background appropriately while increasing contrast and with retaining features of interest at high fidelity, such as the RNA binding pocket in all orientations. The same ⅛ Nyquist Gaussian low-pass applied to the denoised tomogram further improves contrast by suppressing high-frequencies that the user may deem unimportant. See Supplemental Video 2 for the tomogram slice-throughs.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Baxter, W. T., Grassucci, R. A., Gao, H. & Frank, J. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *J. Struct. Biol.* **166**, 126–132 (2009).
2. Sindelar, C. V. & Grigorieff, N. An adaptation of the Wiener filter suitable for analyzing images of isolated single particles. *J. Struct. Biol.* **176**, 60–74 (2011).
3. Penczek, P. A. Chapter Two - Image Restoration in Cryo-Electron Microscopy. in *Methods in Enzymology* (ed. Jensen, G. J.) vol. 482 35–72 (Academic Press, 2010).
4. Merk, A. *et al.* Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698–1707 (2016).
5. Herzik, M. A., Wu, M. & Lander, G. C. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* **10**, 1–9 (2019).
6. Milanfar, P. A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical. *IEEE Signal Process. Mag.* **30**, 106–128 (2013).
7. Xie, J., Xu, L. & Chen, E. Image Denoising and Inpainting with Deep Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 341–349 (Curran Associates, Inc., 2012).
8. Jain, V. & Seung, S. Natural Image Denoising with Convolutional Networks. in *Advances in Neural Information Processing Systems 21* (eds. Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L.) 769–776 (Curran Associates, Inc., 2009).
9. Lehtinen, J. *et al.* Noise2Noise: Learning Image Restoration without Clean Data. *ArXiv180304189 Cs Stat* (2018).

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10. Zhussip, M., Soltanayev, S. & Chun, S. Y. Extending Stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. *ArXiv190202452 Cs* (2019).

11. Batson, J. & Royer, L. Noise2Self: Blind Denoising by Self-Supervision. *ArXiv190111365 Cs Stat* (2019).

12. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - Learning Denoising from Single Noisy Images. *ArXiv181110980 Cs* (2018).

13. Buchholz, T., Jordan, M., Pigino, G. & Jug, F. Cryo-CARE: Content-Aware Image Restoration for Cryo-Transmission Electron Microscopy Data. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 502–506 (2019). doi:10.1109/ISBI.2019.8759519.

14. Buchholz, T.-O. *et al.* Chapter 13 - Content-aware image restoration for electron microscopy. in *Methods in Cell Biology* (eds. Müller-Reichert, T. & Pigino, G.) vol. 152 277–289 (Academic Press, 2019).

15. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).

16. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* 1–8 (2019) doi:10.1038/s41592-019-0575-8.

17. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

18. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

19. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

20. Lander, G. C. *et al.* Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21. de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).

22. Mao, Y. *et al.* Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proc. Natl. Acad. Sci.* **110**, 12438–12443 (2013).

23. Haider, S. A. *et al.* Fluorescence microscopy image noise reduction using a stochastically-connected random field model. *Sci. Rep.* **6**, 20640 (2016).

24. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).

25. Brasch, J. *et al.* Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **569**, 280–283 (2019).

26. Frank, J. & Al-Ali, L. Signal-to-noise ratio of electron micrographs obtained by cross correlation. *Nature* **256**, 376–379 (1975).

27. Chen, M. *et al.* Convolutional Neural Networks for Automated Annotation of Cellular Cryo-Electron Tomograms. *ArXiv170105567 Q-Bio* (2017).

28. Kokic, G. *et al.* Structural basis of TFIIH activation for nucleotide excision repair. *Nat. Commun.* **10**, 2885 (2019).

29. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).

30. Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*.

31. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18037–18041 (2013).

32. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 (2016).

33. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

34. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, 2015).

35. Paszke, A. *et al.* Automatic differentiation in PyTorch. (2017).

36. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer Visualization of Three-Dimensional Image Data Using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

37. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

38. Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

39. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

40. Schaffer, M. *et al.* Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *J. Struct. Biol.* **197**, 73–82 (2017).

41. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

42. Suloway, C. *et al.* Fully automated, sequential tilt-series acquisition with Leginon. *J. Struct. Biol.* **167**, 11–18 (2009).

43. Noble, A. J. & Stagg, S. M. Automated batch fiducial-less tilt-series alignment in Appion using Protomo. *J. Struct. Biol.* **192**, 270–278 (2015).

44. Winkler, H. & Taylor, K. A. Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. *Ultramicroscopy* **106**,

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

240–254 (2006).

45. Agulleiro, J. I. & Fernandez, J. J. Fast tomographic reconstruction on multicore computers. *Bioinformatics* **27**, 582–583 (2011).

46. Agulleiro, J.-I. & Fernandez, J.-J. Tomo3D 2.0 – Exploitation of Advanced Vector eXtensions (AVX) for 3D reconstruction. *J. Struct. Biol.* **189**, 147–152 (2015).

47. Xu, H. *et al.* Structural Basis of Nav1.7 Inhibition by a Gating-Modifier Spider Toxin. *Cell* **176**, 702-715.e14 (2019).

48. Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *eLife* **4**, e06380 (2015).

49. Yu, Q., Qu, K. & Modis, Y. Cryo-EM Structures of MDA5-dsRNA Filaments at Different Stages of ATP Hydrolysis. *Mol. Cell* **72**, 999-1012.e6 (2018).

50. Zhang, W. *et al.* Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *eLife* **8**, e43584 (2019).

51. Kujirai, T. *et al.* Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595–598 (2018).

52. Kato, T. *et al.* EMPIAR-10248 dataset: The 1.54 Å structure of Apoferritin by CRYOARM300 with cold-FEG. (2019) doi:10.6019/EMPIAR-10248.

53. Laughlin, T. G., Bayne, A. N., Trempe, J.-F., Savage, D. F. & Davies, K. M. Structure of the complex I-like molecule NDH of oxygenic photosynthesis. *Nature* **566**, 411–414 (2019).

54. Kern, D. M., Oh, S., Hite, R. K. & Brohawn, S. G. Cryo-EM structures of the DCPIB-inhibited volume-regulated anion channel LRRC8A in lipid nanodiscs. *eLife* **8**, e42636 (2019).

55. Xu, H. *et al.* Correction: Structural basis for the prion-like MAVS filaments in antiviral

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

innate immunity. *eLife* **4**, e07546 (2015).

56. Bartesaghi, A. *et al.* 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147–1151 (2015).

57. Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e03080 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplement

Supplemental Table 1: SNR comparison for model architectures, loss functions, and training datasets for the 2D denoising models

Training Dataset	Model architecture	Objective	EMPIAR datasets			NYSBC K2 datasets				NYSBC Falcon III datasets			Overall
			10261	10005	10025	Protocadherin	18sep08d	19jan04d	19may10e	18aug17l	18sep06d	18sep19l	
Small	Affine	L1	5.49	1.29	0.72	4.83	4.51	8.87	12.02	10.65	6.90	9.15	6.44
		L2	5.28	1.04	0.40	4.64	4.33	8.64	11.88	10.43	6.89	9.02	6.25
	FCNN	L1	5.45	1.25	-0.05	5.37	5.85	8.78	12.77	12.02	7.90	10.89	7.02
		L2	5.17	0.94	-0.59	4.82	5.67	8.62	12.31	11.30	7.76	10.29	6.63
	U-Net (small)	L1	5.99	0.65	0.12	5.89	**7.11**	9.71	14.17	12.05	**8.25**	10.93	7.49
		L2	5.32	0.49	0.27	5.55	6.85	**9.74**	**14.26**	11.23	7.49	10.36	7.16
	U-Net	L1	5.24	0.79	2.51	5.64	5.38	9.38	13.53	9.91	6.49	9.59	6.85
		L2	5.72	0.86	**3.01**	5.41	5.22	8.64	13.06	9.36	6.61	9.16	6.70
Large	Affine	L1	4.30	-0.24	-1.09	3.76	3.36	7.23	11.34	9.32	6.56	8.45	5.30
		L2	4.26	-0.27	-1.15	3.74	3.33	7.20	11.32	9.32	6.57	8.46	5.28
	FCNN	L1	4.69	-0.74	-1.57	3.95	5.25	7.36	11.93	10.26	8.04	9.72	5.89
		L2	3.78	-0.47	-1.57	3.76	4.79	6.81	11.21	10.15	7.34	9.62	5.54
	U-Net (small)	L1	6.47	0.51	-1.00	5.18	4.82	7.04	12.29	13.62	8.04	10.83	6.78
		L2	6.00	0.84	-0.46	5.19	5.20	7.90	12.34	12.36	7.20	10.38	6.69
	U-Net	L1	6.95	1.88	0.83	**6.33**	6.14	9.35	13.87	13.11	7.35	12.76	7.86
		L2	**7.17**	1.72	1.07	5.94	6.06	8.43	13.07	**15.17**	7.37	**13.24**	**7.92**
Low-pass		4	-5.28	-11.17	-11.92	-5.28	-6.08	-2.97	3.44	-1.04	-1.02	-0.29	-4.16
		8	0.30	-5.52	-5.89	-0.23	-0.89	2.03	7.84	4.30	2.87	4.57	0.94
		16	5.19	-0.12	-0.40	4.22	3.53	6.87	9.99	9.04	6.95	8.71	5.40
		32	3.92	**1.89**	0.08	1.65	1.63	8.25	10.13	8.58	1.90	7.92	4.59
		64	2.34	-2.22	-0.83	-3.25	5.60	5.18	11.12	10.23	6.58	-0.02	3.47
Raw			-17.14	-20.13	-24.15	-14.47	-15.40	-11.73	-5.44	-6.33	-3.64	-5.63	-12.41

Supplemental Table 1 | Comparison of denoising methods based on estimated SNR (in dB, larger is better). SNR was estimated from 20 paired signal and background regions selected for each dataset. In each column, the best performing model is highlighted. We report denoising results on aligned and dose weighted micrographs for the NYSBC K2 and Falcon III datasets.All datasets were collected in electron counting modes, except for 18sep06d which was collected using Falcon III integrating mode. The U-net denoising model trained on the “Large” dataset with L2 loss performs best on average. For the low-pass filter baselines, the amount of filtering is reported in the “Objective” column. The SNR of the raw micrographs is reported in the last row.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Table 2: SNR comparison for model architectures and training datasets for the 3D denoising models

Method	19aug10b: Yeast lamella	19jan20b: Cadherins on liposomes	19jul11a: Yeast lamella	19oct06a: Cellular lamella
Raw (bin2)	-11.12	-14.33	-9.27	-12.91
Low-pass (2x)	-9.04	-14.08	-6.71	-11.83
Low-pass (4x)	-8.44	-13.31	-5.03	-11.12
Low-pass (8x)	-12.34	-13.51	-8.63	-14.58
Topaz (self-trained)	-7.27	-11.55	-4.73	-9.78
Topaz (Unet-3d-10a)	-7.55	-11.66	-4.79	-10.00

Supplemental Table 2 | Comparison of tomogram denoising methods by SNR (in dB). For each dataset, SNR was calculated following the independent signals method. Briefly, tomograms were split into even/odd frame tomograms, denoising methods were applied to the even frame tomograms and odd frame tomograms independently. Then, denoised even frame tomograms were compared to the raw odd frame tomograms to calculate SNR and vice versa for the denoised odd and raw even frame tomograms. These values were averaged together to give the reported SNRs. The Topaz U-net models trained on each tomogram (self-trained) achieves the best SNR in each case, but is only marginally better than the general pre-trained tomogram denoising model (Unet-3d-10a). Both models improve over low-pass filtering on all four examples.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

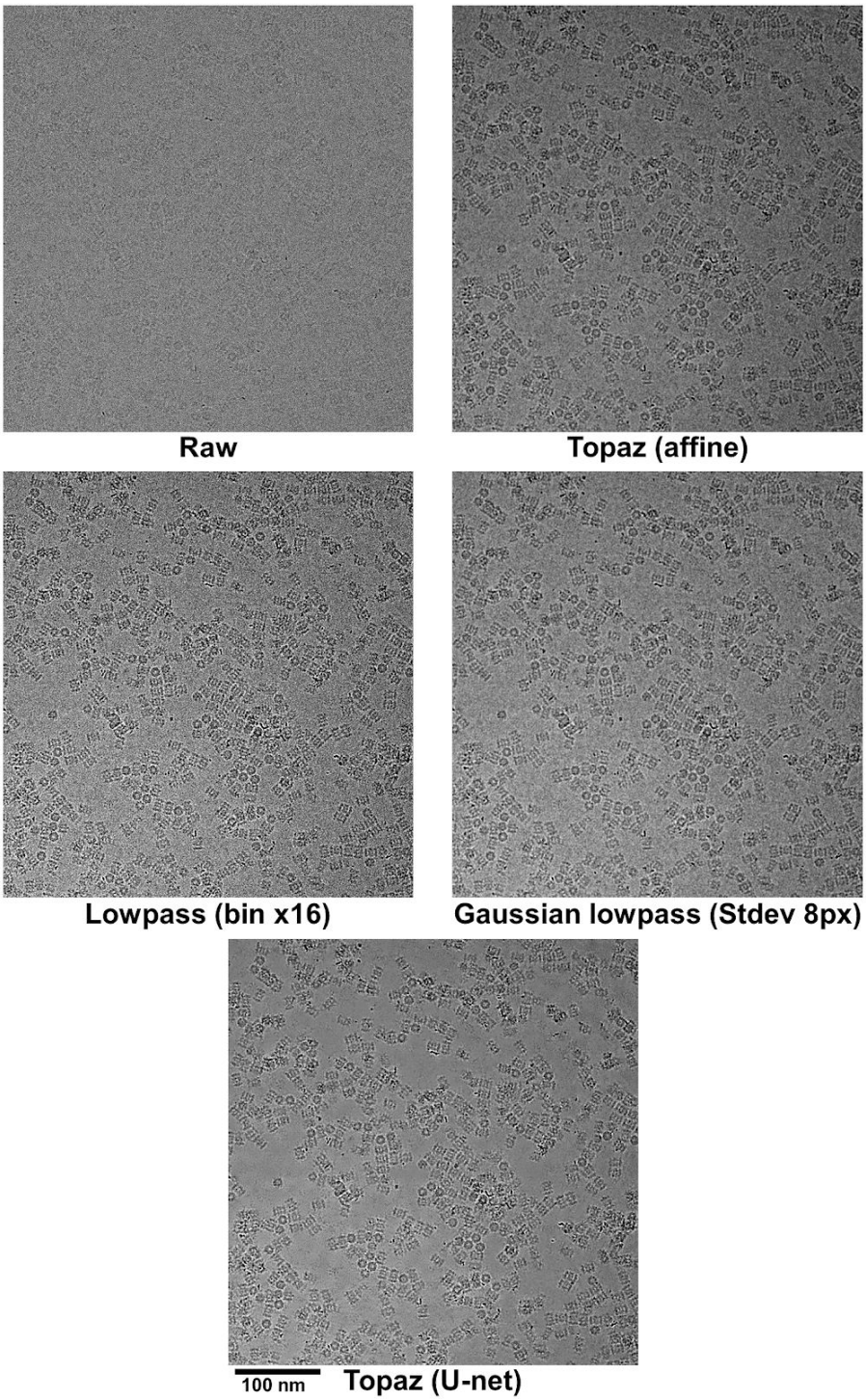
Supplemental Table 3: Training datasets for the 2D denoising models

Dataset Name	Number of Micrographs	Description
NYSBC zero defocus	671	NYSBC ladder dataset collected at zero defocus; K2 counting
NYSBC ice images	900	NYSBC micrographs taken of vitreous ice; K2 counting
EMPIAR-10025[48]	196	T20S proteasome; K2 super-resolution
EMPIAR-10005[24]	499	TRPV1; K2 counting
EMPIAR-10210[49]	170	mouse MDA5-dsRNA filaments; Falcon III counting
EMPIAR-10243[50]	142	heparin-induced 2N4R tau filaments; Falcon III counting
EMPIAR-10244[51]	642	RNA polymerase II transcribing a nucleosome; K2 super-resolution
EMPIAR-10248[52]	971	Apoferritin by CRYOARM300 with cold-FEG; K2 counting
EMPIAR-10249[5]	596	Horse liver alcohol dehydrogenase movies obtained using Talos Arctica operating at 200 kV equipped with a K2
EMPIAR-10250[5]	181	Human methemoglobin movies obtained using Talos Arctica operating at 200 kV equipped with a K2
EMPIAR-10252[5]	153	Catalytic subunit of protein kinase A bound to ATP, manganese, and IP20 movies obtained using Talos Arctica operating at 200 kV equipped with a K2
EMPIAR-10257[53]	295	NDH the complex I-like molecule of photosynthesis; K2 super-resolution
EMPIAR-10258[54]	199	LRRC8A-DCPIB in MSP1E3D1 nanodiscs; K2 super-resolution
EMPIAR-10259[54]	198	apo-LRRC8A in MSP2N2 nanodiscs; K2 super-resolution
EMPIAR-10261[47]	1461	ProTx2-bound Nav1.7 VSD2-NavAb chimeric channel; K2 counting
EMPIAR-10031[55]	512	MAVS CARD C1 filaments, Falcon II counting
EMPIAR-10061[56]	397	beta-galactosidase in complex with a cell-permeant inhibitor; K2 super-resolution
EMPIAR-10028[57]	600	Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine; Falcon II counting
Small	1929	Contains micrographs from datasets: EMPIAR-10005, -10025, -10061, -10244, -10249, -10250, -10252, -10257, -10258, and -10261
Large	3439	Contains micrographs from all individual datasets

Supplemental Table 3 | List of datasets included in model training. The individual datasets with number of micrographs from each and brief descriptions are provided in the first block. The second block describes the camera/imaging mode specific datasets. These are composed of all micrographs from subsets of the individual datasets. The last block describes the two general datasets. The “Small” dataset is composed of micrographs from a subset of the individual datasets that we found to give best performing models by eye. The “Large” dataset contains micrographs from all individual datasets. For the “Small” and “Large” datasets, individual datasets with more than 200 micrographs were subsampled to only include 200 images. This serves to approximately balance the contributions of each contained dataset.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

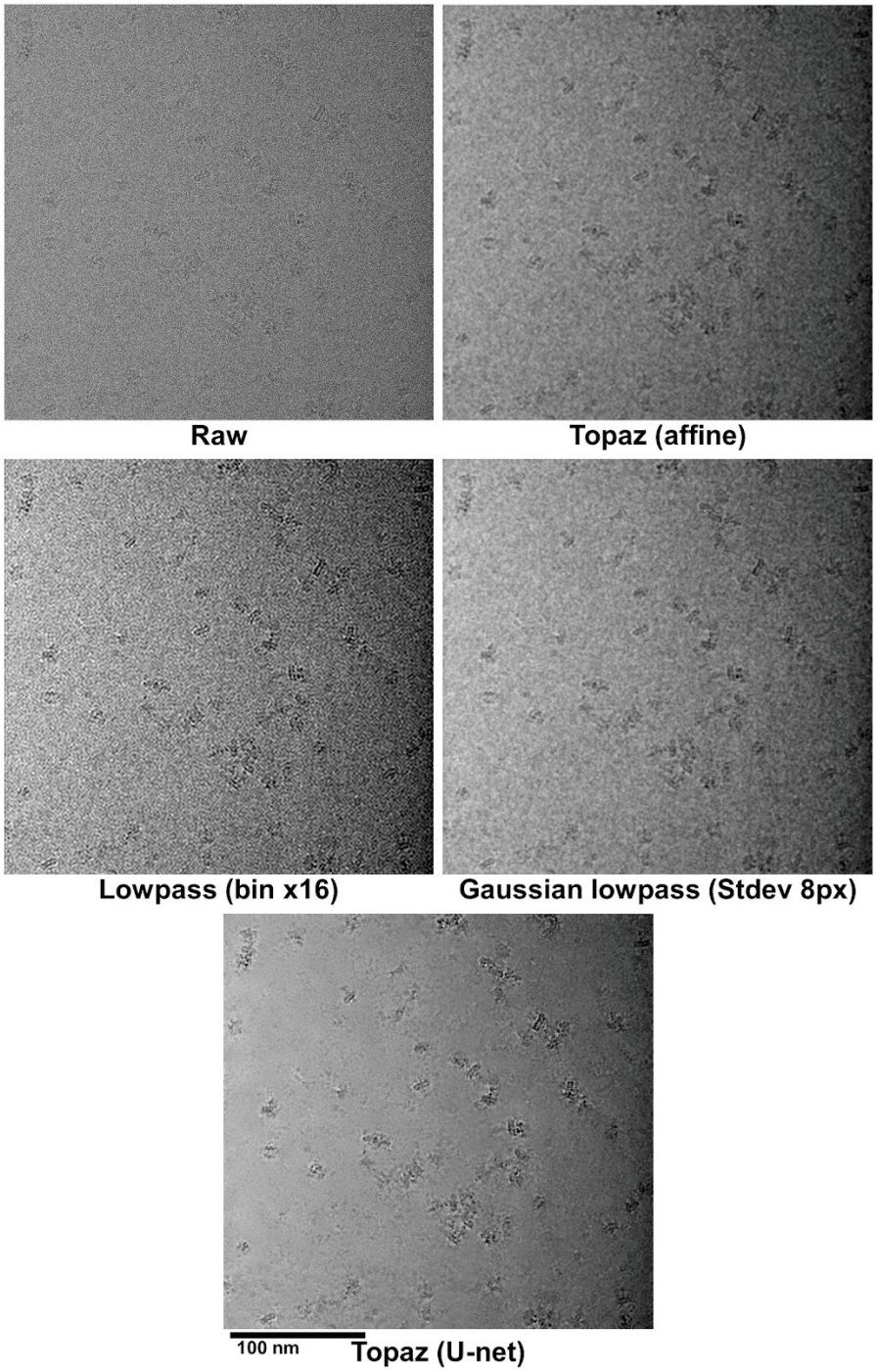
Supplemental Figure 1: Comparison of denoising methods



Supplemental Figure 1 | Figure 1b micrograph (pixel size: 0.6575 Å) processed in four different ways: Topaz affine denoising model, low-pass binning by Fourier cropping by a factor of 16 then padding, Gaussian low-pass filtering with a standard deviation of 8 pixels, and our Topaz U-net denoising model.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

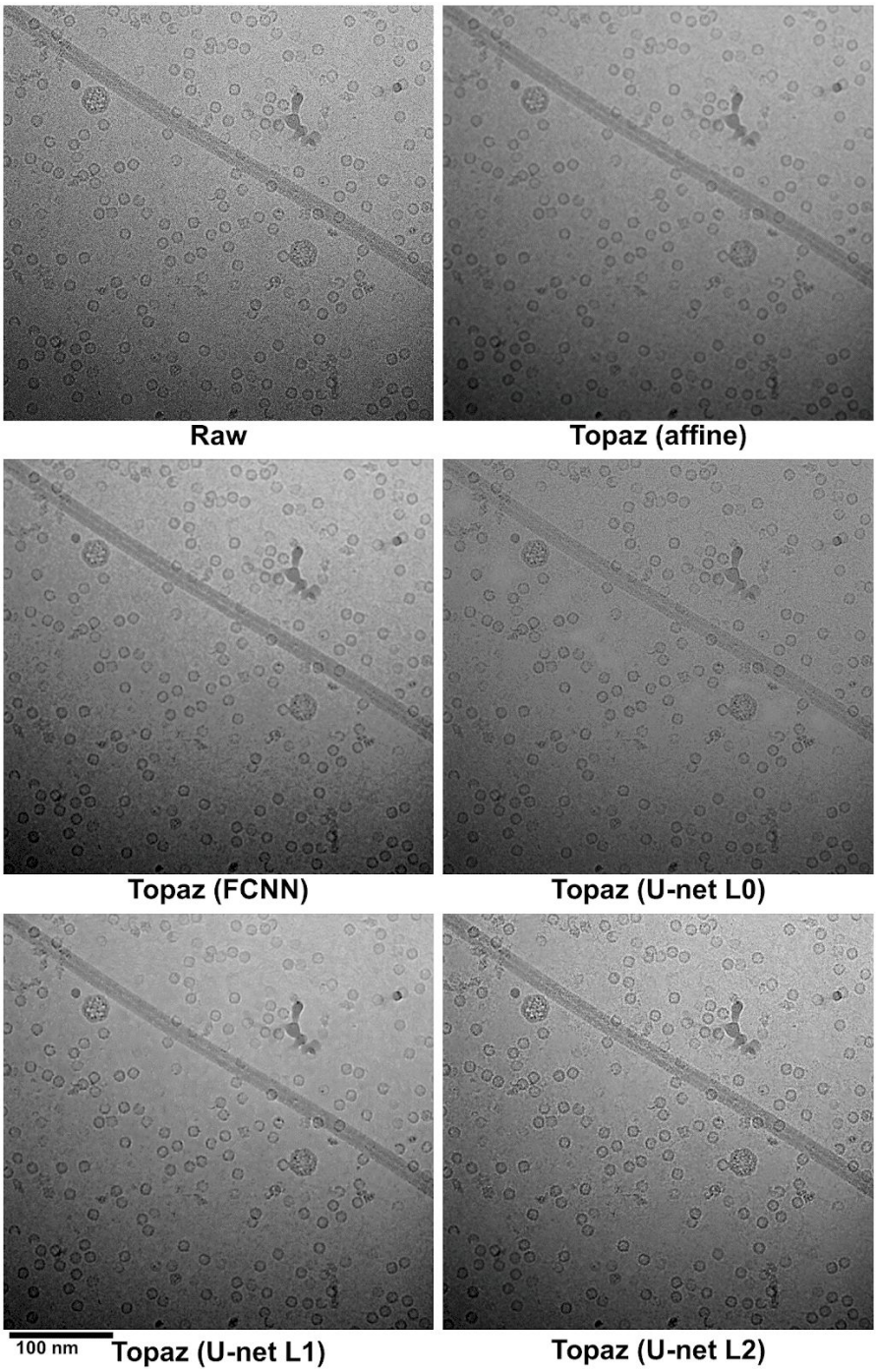
Supplemental Figure 2: Comparison of denoising methods on EMPIAR-10261



Supplemental Figure 2 | Figure 1c micrograph (pixel size: 0.849 Å) processed in four different ways: Topaz affine denoising model, low-pass binning by Fourier cropping by a factor of 16 then padding, Gaussian low-pass filtering with a standard deviation of 8 pixels, and our Topaz U-net denoising model.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

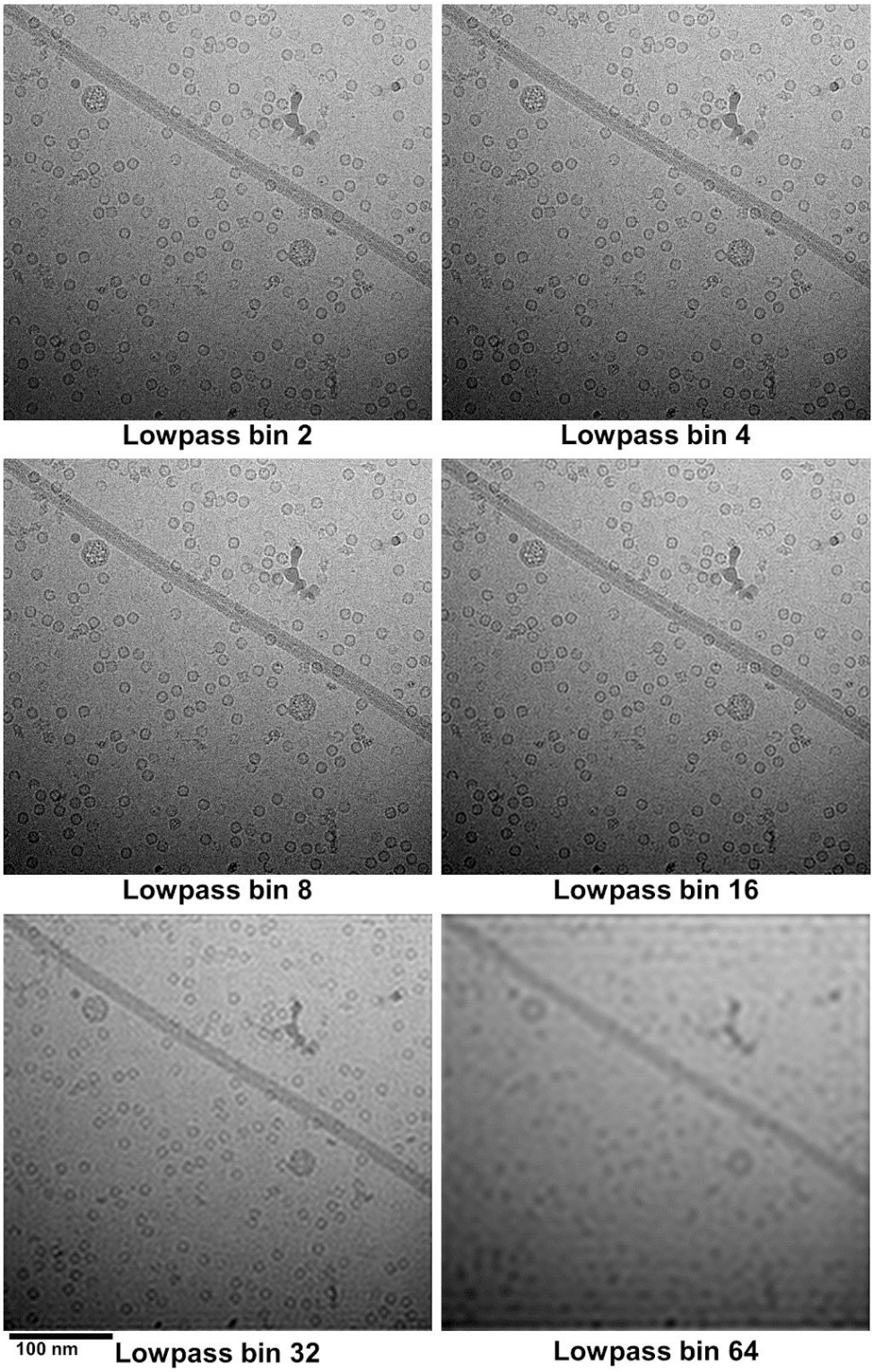
Supplemental Figure 3: Comparison of Topaz neural denoising models on 19jan04d



Supplemental Figure 3 | Comparison between denoising models on a micrograph (pixel size 1.10 Å) of apoferritin, β-galactosidase, VLPs, and TMV. The raw image denoised with the affine model, FCNN model, U-net with mode-seeking L0 loss, U-net with median-seeking L1 loss, and U-net with mean-seeking L2 loss are shown.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

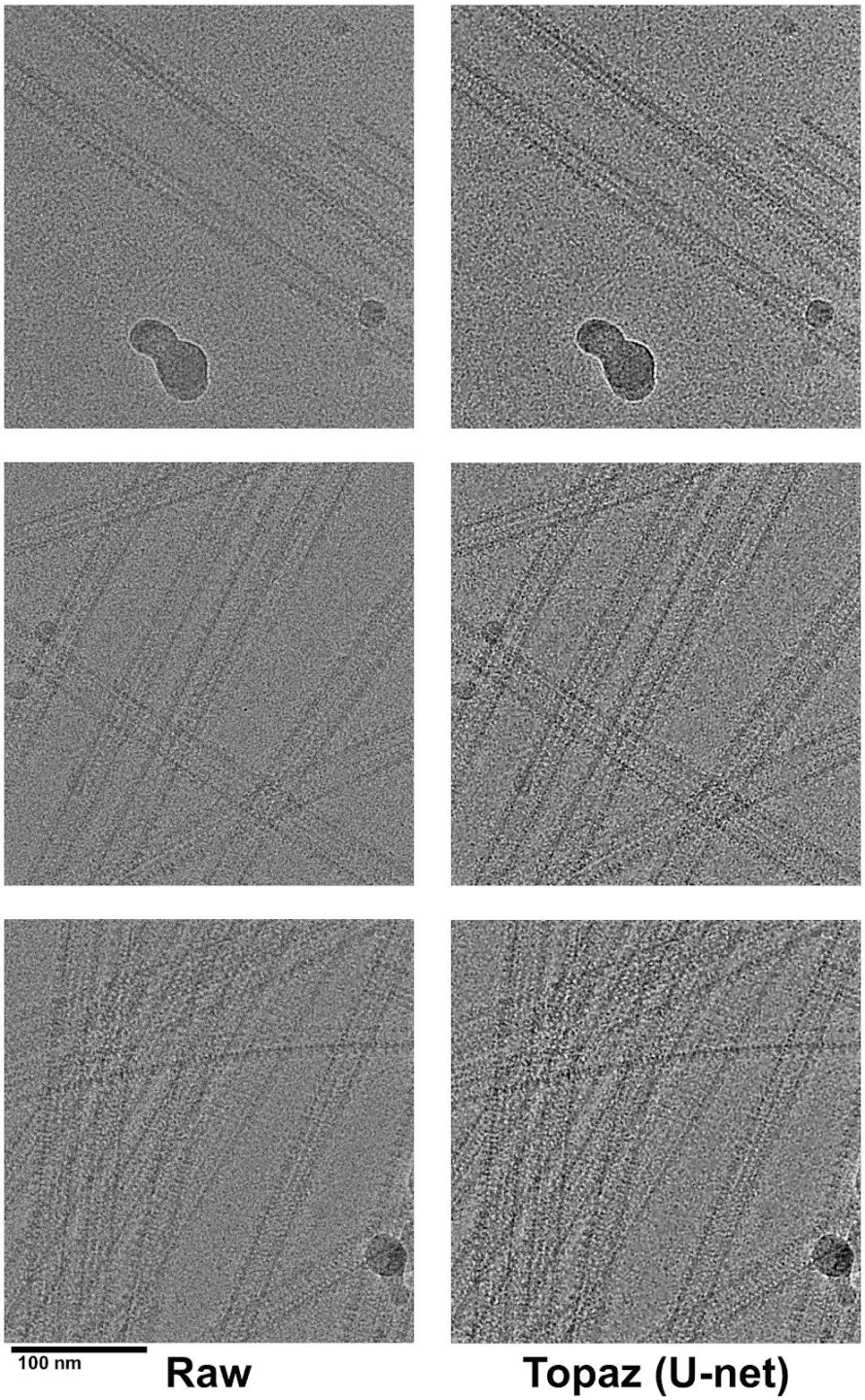
Supplemental Figure 4: Comparison of low-pass binning on 19jan04d



Supplemental Figure 4 | Comparison between low-pass binning by Fourier cropping. The raw micrograph in Supplemental Figure 3 is low-pass filtered by factors of 2, 4, 8, 16, 32, and 64.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

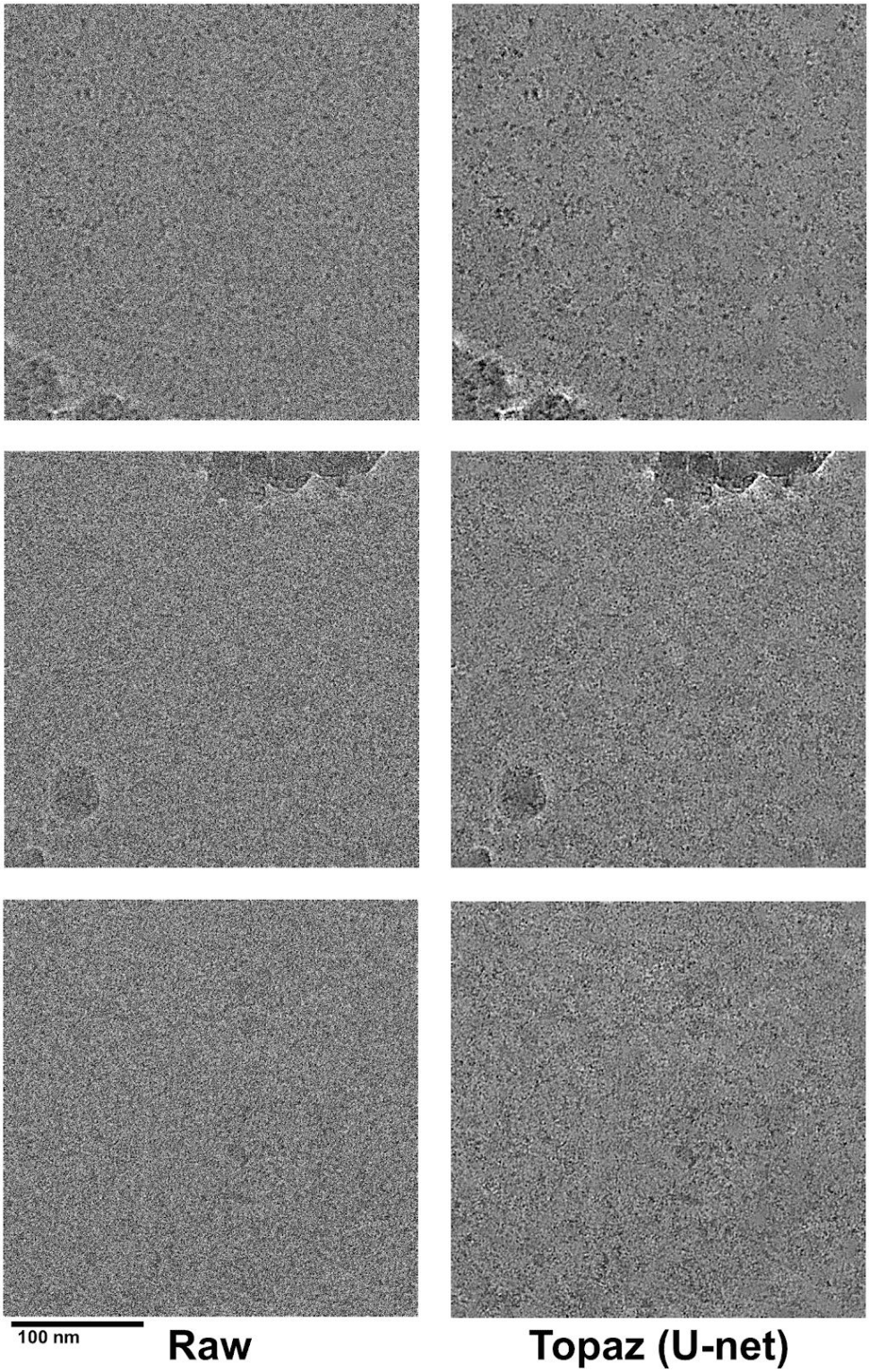
Supplemental Figure 5: Denoising sample with known background contaminants



Supplemental Figure 5 | Denoising raw micrographs (18sep15a) of microtubules with known background contaminant proteins, kinesin and tubulin. Topaz denoising appropriately accentuates features of the background proteins instead of smoothing them out.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

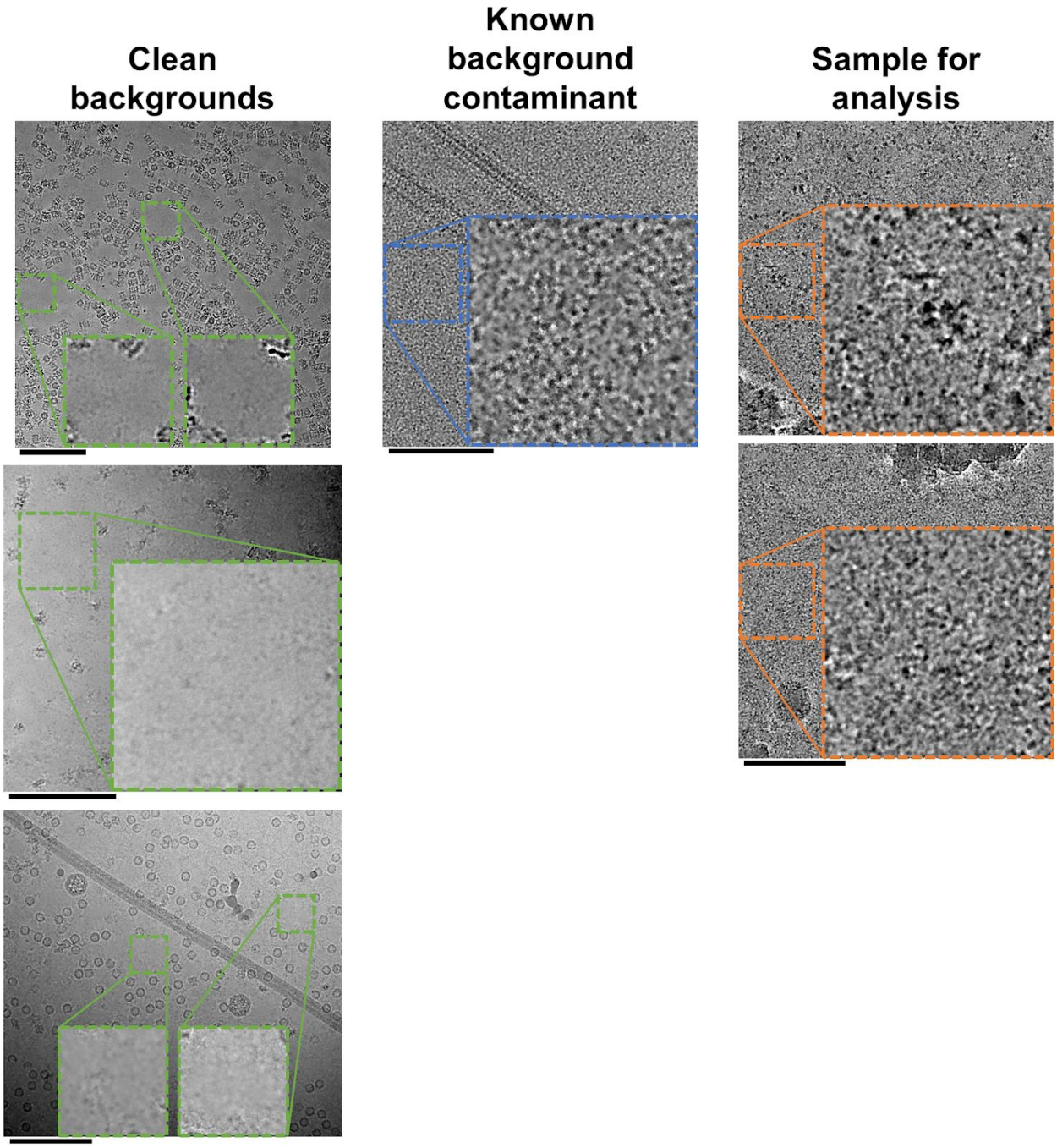
Supplemental Figure 6: Denoising EMPIAR-10003



Supplemental Figure 6 | Denoising of EMPIAR-10003 raw images (left) using the U-net model (right). Possible regions of proteins are particularly apparent in the top image.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

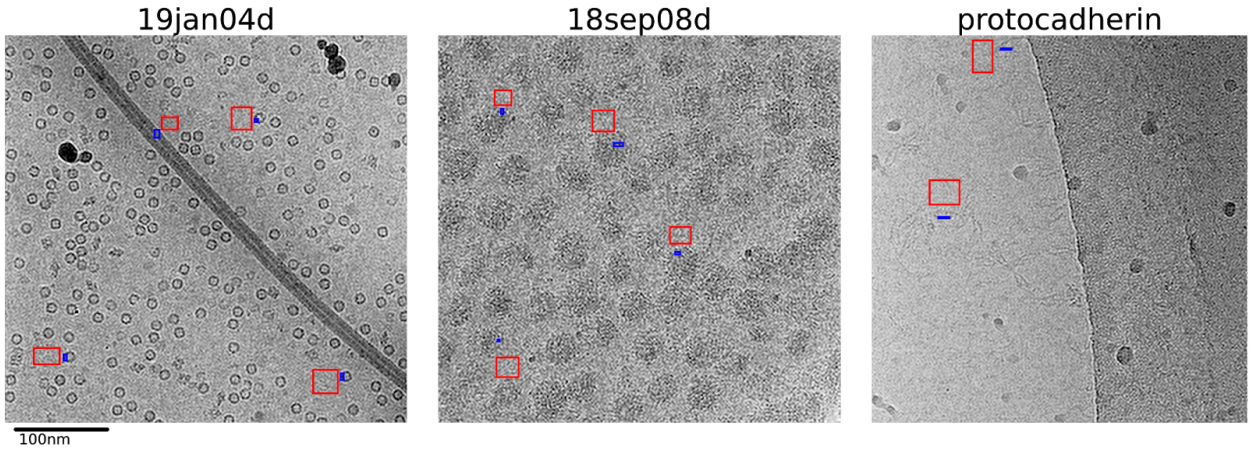
Supplemental Figure 7: Comparison of backgrounds after denoising



Supplemental Figure 7 | Denoising as a complementary method for analyzing background proteins and contamination in sample/grid preparations. Left: Three micrographs with nearly clean backgrounds (green insets). Middle: A micrograph of microtubules with known kinesin and tubulin background contaminant (blue inset). Right: Two micrographs from the EMPIAR-10003 dataset with the centers magnified (orange insets). All micrographs are denoised using our Topaz U-net model. Insets are magnified by 2x. Scalebars are 100nm.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

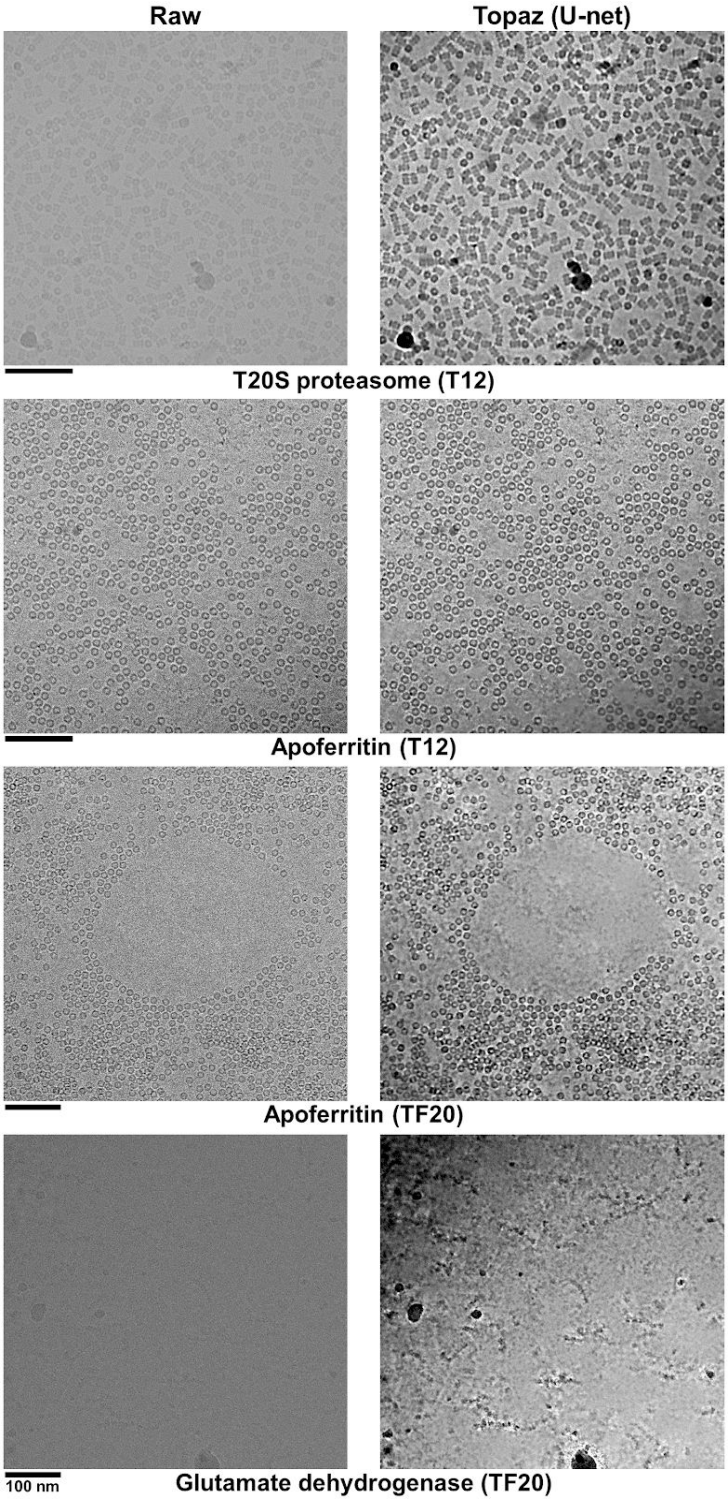
Supplemental Figure 8: Paired signal and background for SNR quantification



Supplemental Figure 8 | Example micrographs from 19jan04d, 18sep08d, and the protocadherin dataset showing labeled signal (blue) and background (red) regions overlayed over low-pass filtered images. Signal and background regions were selected close together to match local background properties as best as possible to each signal region.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

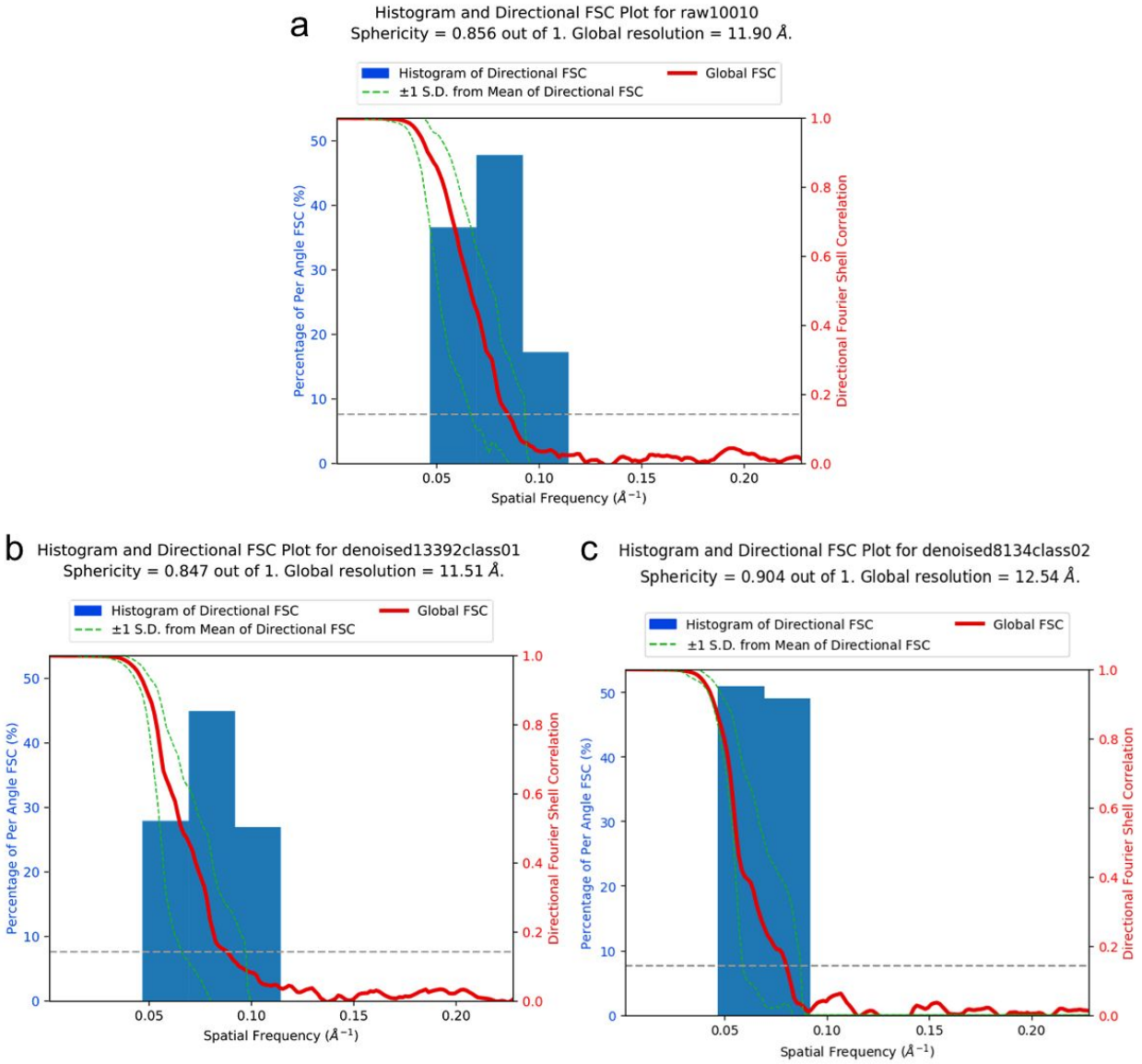
Supplemental Figure 9: Denoising micrographs from screening microscopes



Supplemental Figure 9 | Hardware-binned-by-two micrographs from 120kV (T12) and 200kV (TF20) screening microscopes with TVIPS 4kTemCam-F416 scintillator-based cameras denoised with the general U-net model.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

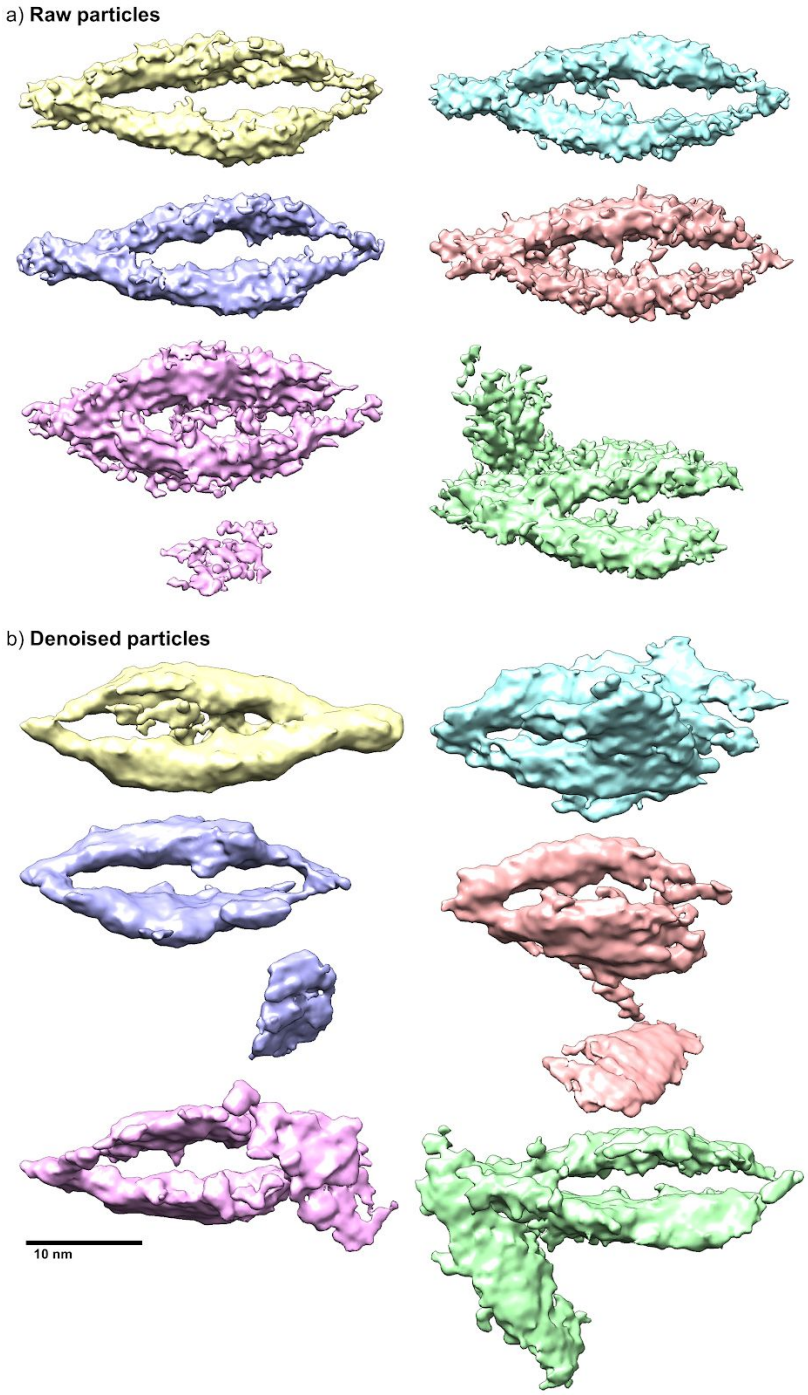
Supplemental Figure 10: Denoising improves manual picking completion for difficult particles



Supplemental Figure 10 | 3DFSC plots of clustered protocadherin (EMPIAR-10234). **(a)** Using the particle picks reported in Brasch et al., 2019, resulting in 10,010 particles in the 3D reconstruction (Figure 2b, left). **(b)** Using the particle picks reported on in the Methods (from manually picking on denoised micrographs prior to Topaz picking training on raw micrographs), resulting in 13,392 particles in the class 1 3D reconstruction (Figure 2b, middle) and **(c)** 8,134 particles in the class 2 3D reconstruction (Figure 2b, right).

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

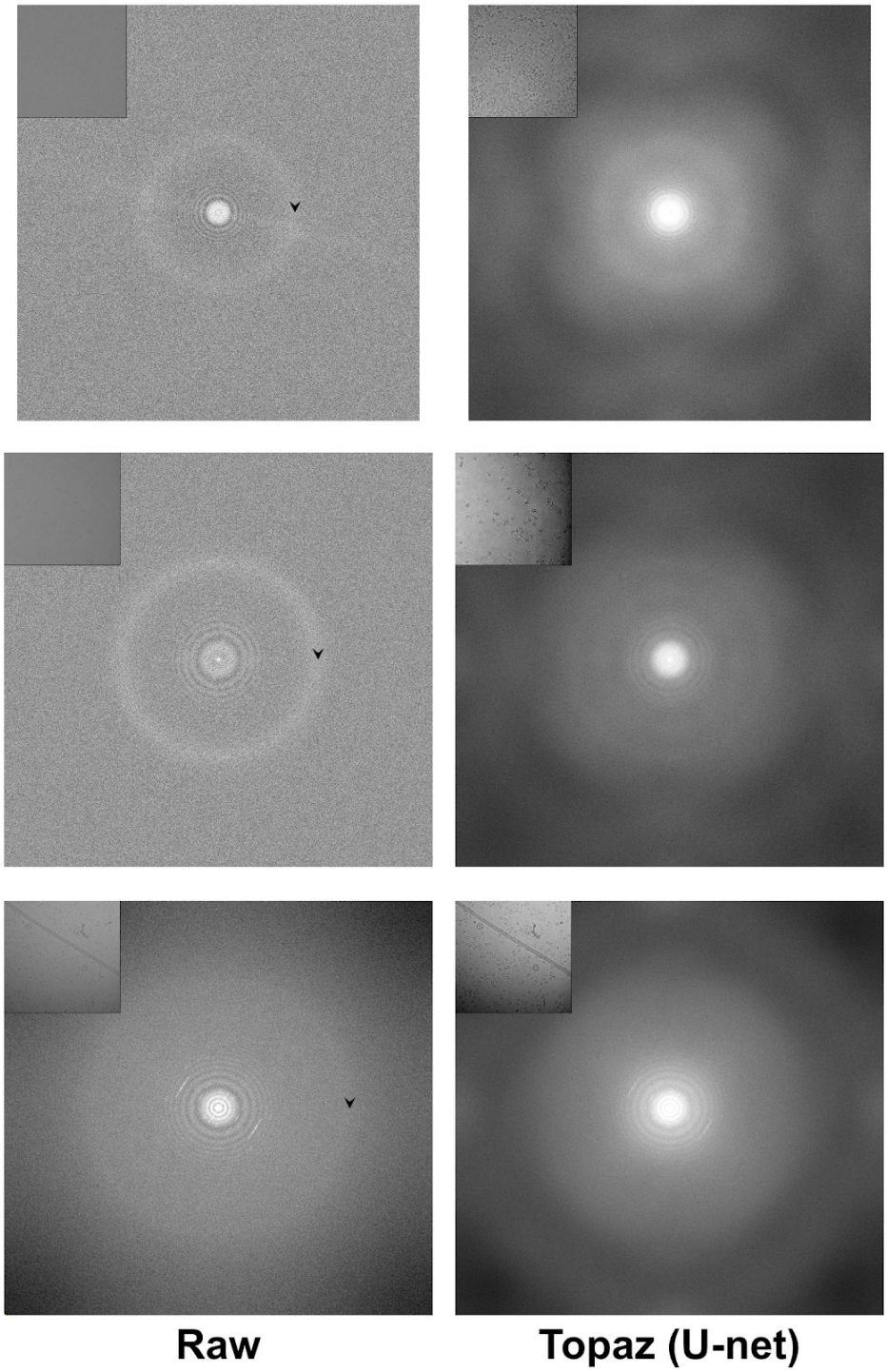
Supplemental Figure 11: Denoised particles are less reliable for ab-initio model generation



Supplemental Figure 11 | 1,000 random particles processed through CryoSPARC ab-initio reconstruction using raw particles **(a)** and denoised particles **(b)**. 4 out of 6 reconstructions using raw particles result in the correct structure (a), while at best 1 out of 6 reconstructions using denoised particles result in the correct structure (b).

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

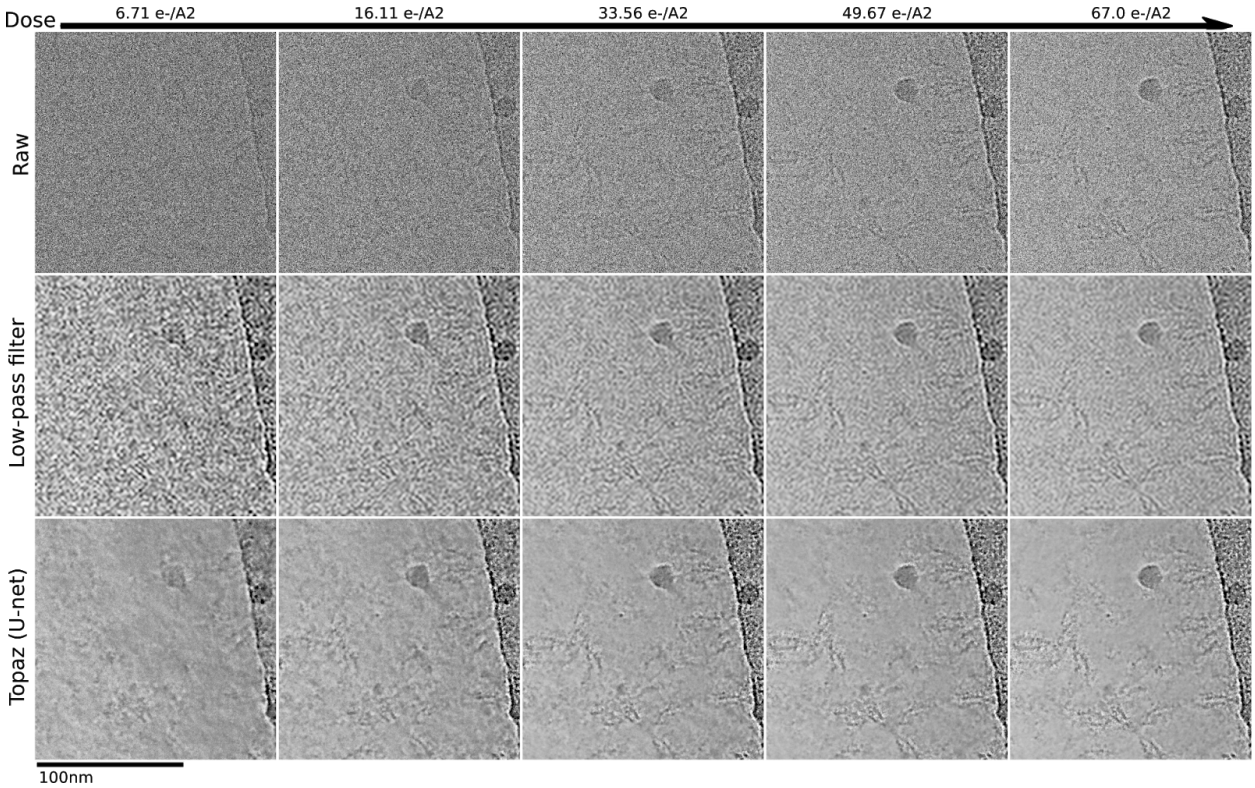
Supplemental Figure 12: Fourier transforms of raw and denoised micrographs



Supplemental Figure 12 | Fourier transforms of previously shown raw and denoised micrographs (insets). Top: Figure 1b, EMPIAR-10025. Middle: Figure 1c, EMPIAR-10261. Bottom: Figure 3, 19jan04d. Arrows show the location of the ~3.7 Å ice ring.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

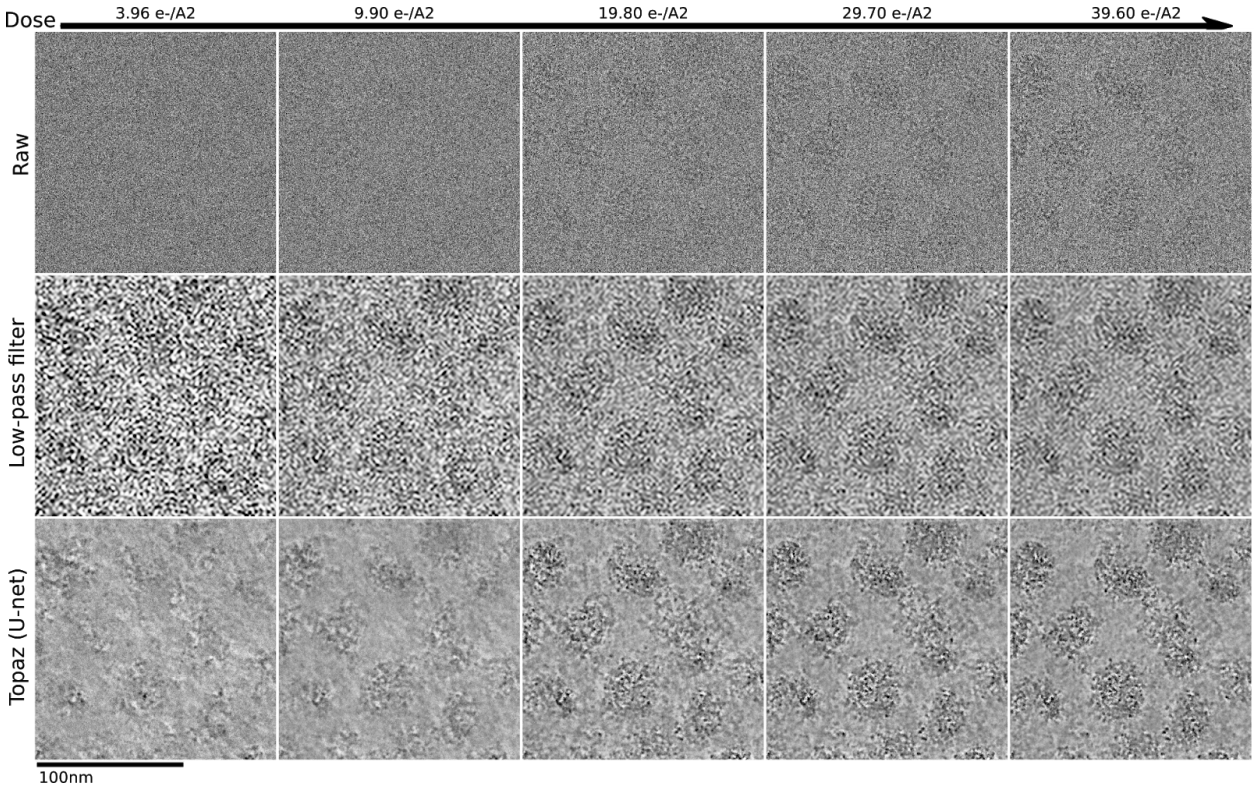
Supplemental Figure 13: Short exposure detail of protocadherin



Supplemental Figure 13 | Detail of protocadherin micrograph denoised and raw over increasing dose.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

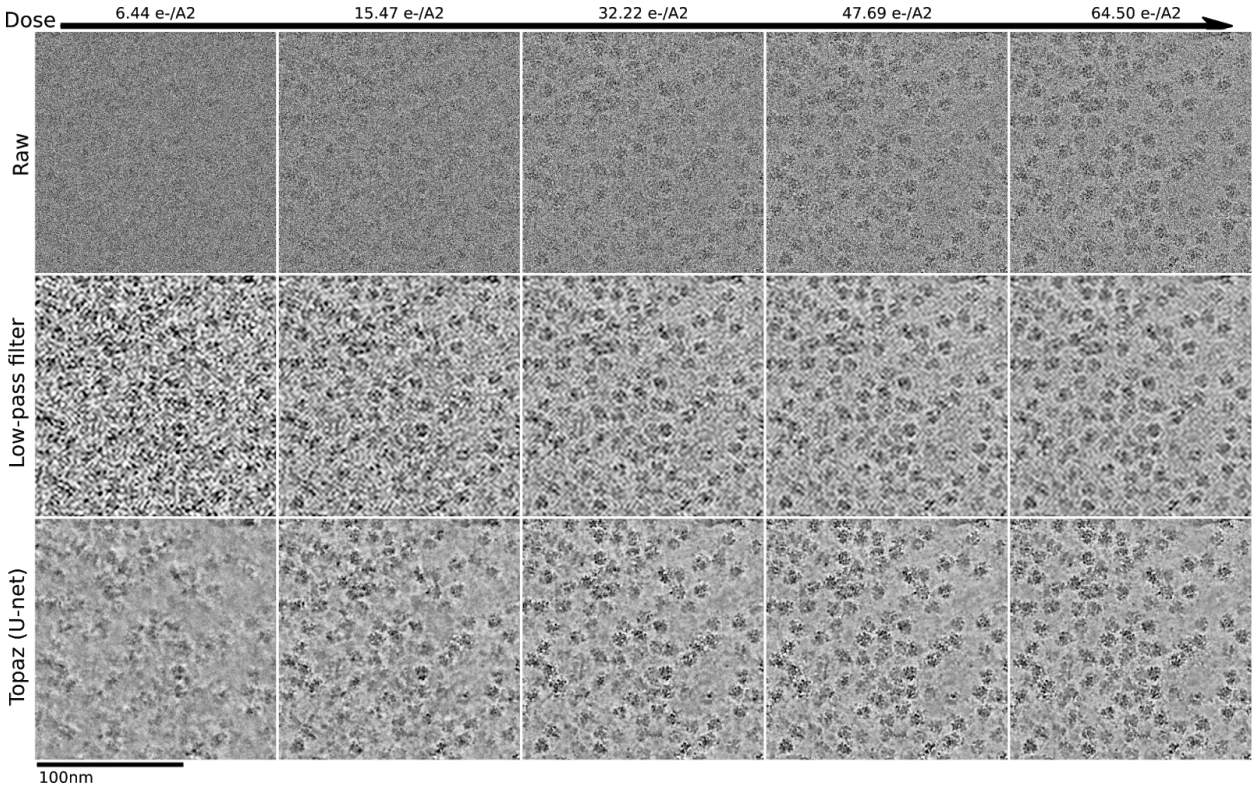
Supplemental Figure 14: Short exposure detail of 18sep08d (VLPs)



Supplemental Figure 14 | Detail of 18sep08d micrograph denoised and raw over increasing dose.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

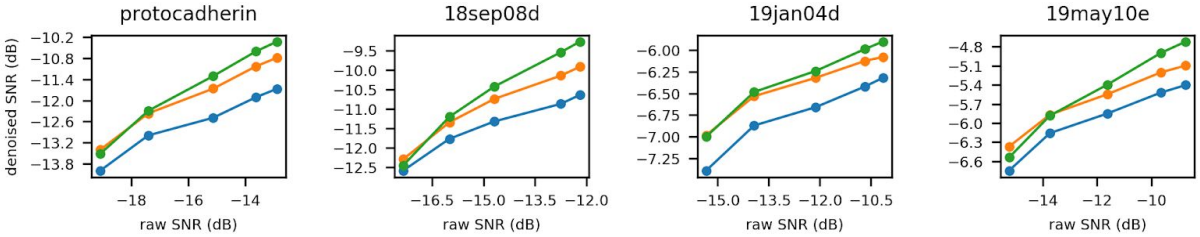
Supplemental Figure 15: Short exposure detail of 19may10e



Supplemental Figure 15 | Detail of 19may10e micrograph denoised and raw over increasing dose.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

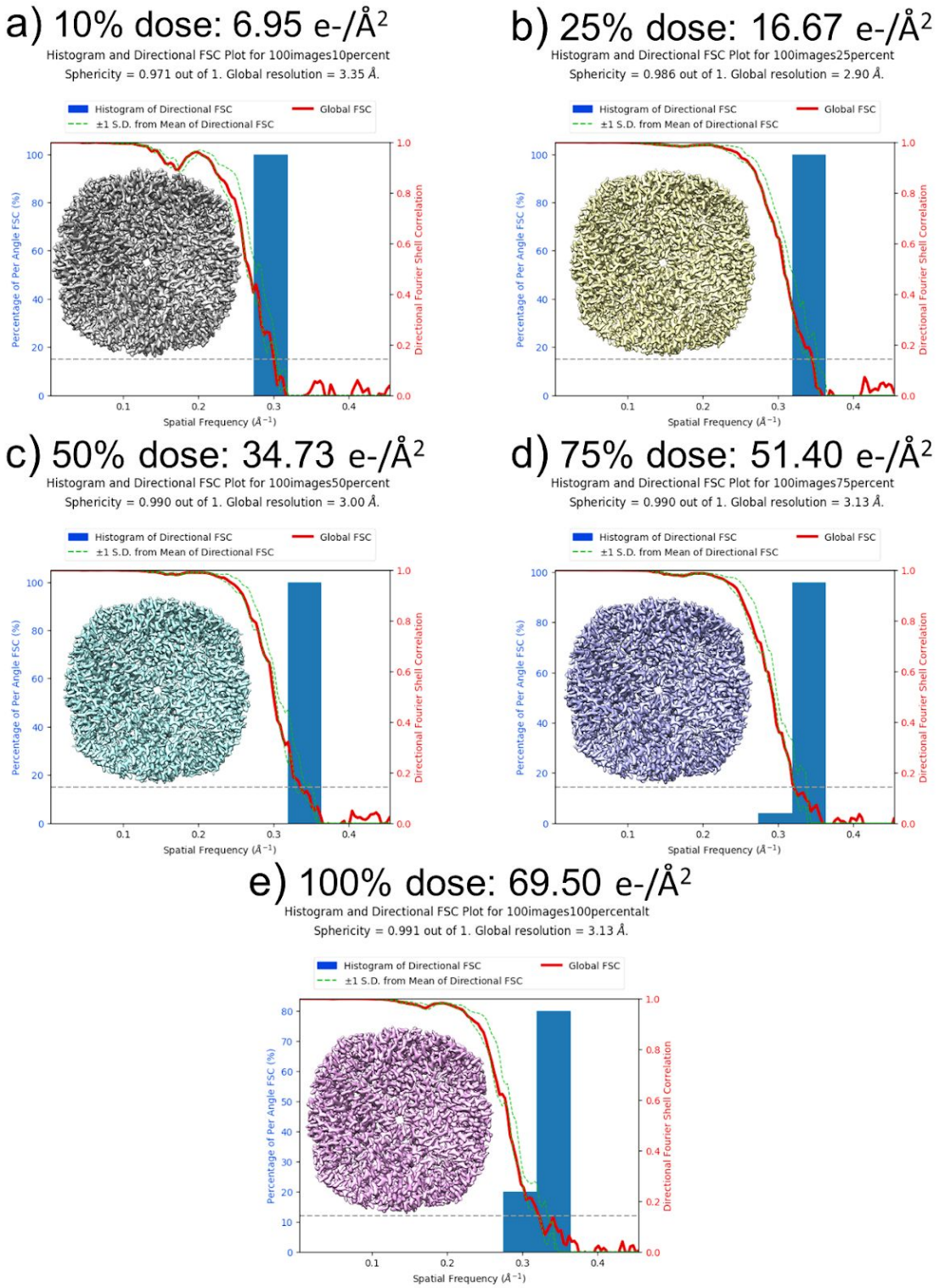
Supplemental Figure 16: Estimated raw vs estimated denoised SNRs of short exposure micrographs



Supplemental Figure 16 | Plots of estimated raw SNR vs. estimated denoised SNR for short exposure micrographs accompanying Figure 3a. Micrographs were denoised with either a 16x low-pass filter (blue), the affine denoising model (orange), or our general purpose U-net model (green). SNRs were estimated using the split-frames method described in the Methods.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

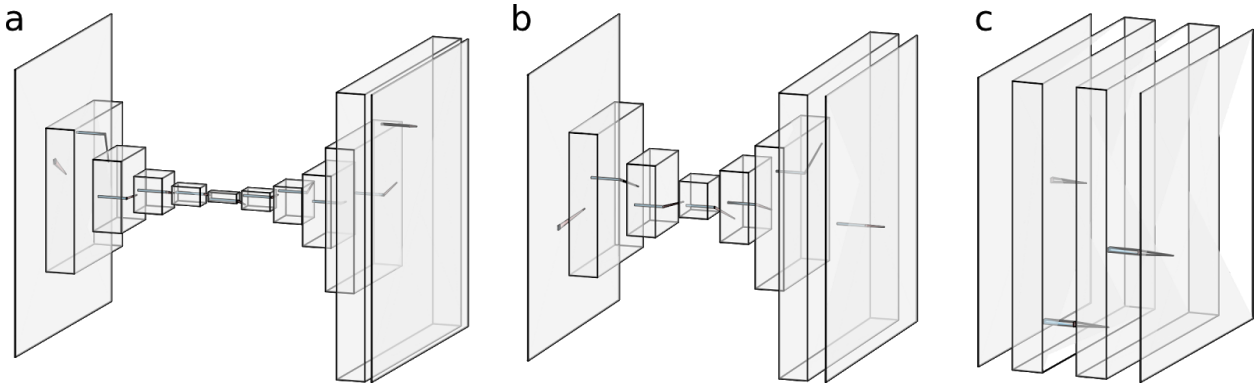
Supplemental Figure 17: Short exposure 3D reconstructions of apoferritin



Supplemental Figure 17 | 3DFSC plots of apoferritin particles fractionated by frames/exposure time: **(a)** 6.95 e-/Å^2, **(b)** 16.67 e-/Å^2, **(c)** 34.73 e-/Å^2, **(d)** 51.40 e-/Å^2, and **(e)** 69.50 e-/Å^2.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Figure 18: Model architectures



Supplemental Figure 18 | Model architecture diagrams for the U-net (a), small U-net (b), and FCNN (b) models. **(a)** The U-net model consists of 5 convolutional and downsampling blocks followed by 5 convolutional and upsampling blocks. Skip connections link each downsampling block to the mirrored upsampling block (not shown). Each convolutional layer has 48 filters. **(b)** The small U-net model has 3 convolutional and downsampling blocks followed by 3 convolutional and upsampling blocks. Skip connections link each downsampling block to the mirrored upsampling block (not shown). Each convolutional layer has 48 filters. **(c)** The FCNN model has three convolutional layers each of width 11 and 64 filters.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Video 1: 3D denoising comparison of Saccharomyces uvarum lamellae

https://drive.google.com/open?id=1-lwh0Ojr_NquzIILAHN50WyKiEHrykDu

Supplemental Video 1 | Visual comparison of cryoET weighted back-projection tomogram slice-throughs of a yeast lamellae. The tilt images were binned by 2 prior to reconstruction. Top-left: The original tomogram without filters. Top-right: The original tomogram binned by another factor of 4 in Fourier space (total binning of 8). Bottom-left: The original tomogram denoised with a 3D denoising model trained on even/odd halves of the same tomogram. Bottom-right: The original tomogram denoised with Unet-3d-10a.

bioRxiv preprint doi: https://doi.org/10.1101/838920; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Video 2: 3D denoising comparison of 80S ribosomes from EMPIAR-10045

https://drive.google.com/open?id=1iS40Gem-MN2nNbt1S6MOo7O2vrN5GPTu

Supplemental Video 2 | Visual comparison of cryoET weighted back-projection tomogram slice-throughs of single particle ribosomes. Left: The original tomogram without filters or binning. Middle: The original tomogram binned by a factor of 8 in Fourier space. Right: The original tomogram denoised with Unet-3d-10a.