

Published in final edited form as:

Nat Methods. 2019 November ; 16(11): 1146–1152. doi:10.1038/s41592-019-0580-y.

Real-time cryo-EM data pre-processing with *Warp*

Dimitry Tegunov¹, Patrick Cramer¹

¹Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany

Abstract

The acquisition of cryo-electron microscopy (cryo-EM) data from biological specimen must be tightly coupled to data pre-processing to ensure best data quality and microscope usage. Here we provide *Warp*, a software for real-time evaluation and pre-processing of cryo-EM data during their acquisition. *Warp* corrects micrographs for global and local motion, estimates the local defocus with the use of novel algorithms, and monitors key parameters for each recorded micrograph or tomographic tilt series in real time. The software further includes deep learning-based models for accurate particle picking and image denoising. The output from *Warp* can be fed into established programs for particle classification and 3D map refinement. Our benchmarks show improvement in the nominal resolution from 3.9 Å to 3.2 Å through fully automated processing of a published cryo-EM data set for influenza virus hemagglutinin. *Warp* is easy to install, computationally inexpensive, and has an intuitive and streamlined user interface.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to D.T. (dteguno@mpibpc.mpg.de) and P.C. (patrick.cramer@mpibpc.mpg.de).

Data availability

Figures 1 and S1 use exemplary data from EMPIAR-10078 (doi:10.6019/EMPIAR-10078). Figure 2 uses a cryo-EM image of RNA Pol II complexes, available from the authors upon request. Figure 3 and the benchmark section use data from EMPIAR-10097 (doi:10.6019/EMPIAR-10097) re-analyzed in this study. The refined maps shown in Figure 3a are available in Supplementary Data 1–4. The "Full *Warp* pipeline" map shown in Figure 3a has been deposited in EMDB as EMD-0025 (<http://www.ebi.ac.uk/pdbe/entry/emdb/EMD-0025>). Figure 4 and the benchmark section use data from EMPIAR-10061 (doi:10.6019/EMPIAR-10061) re-analyzed in this study, the 1.86 Å map shown in Figure 4a is available as Supplementary Data 5. Figure 5a uses a tomogram reconstructed from data from EMPIAR-10045 (doi:10.6019/EMPIAR-10045). Figure 6 and the benchmark section use data from EMPIAR-10045 and EMPIAR-10164 (doi:10.6019/EMPIAR-10164) re-analyzed in this study, the maps shown in Figure 6a and 6b are available in Supplementary Data 6 and 7, respectively. Figure S2 uses exemplary data from EMPIAR-10061. Figure S3 uses exemplary data from EMPIAR-10097 (doi:10.6019/EMPIAR-10097). Figure S5 uses in-house data, available upon request. Figure S6 uses exemplary data from EMPIAR-10078. Figure S7 uses exemplary data from (left) EMPIAR-10078, (center) in-house data available upon request, and (right) EMPIAR-10153 (doi:10.6019/EMPIAR-10153). Training data for *BoxNet* can be accessed through <https://github.com/cramerlab/boxnet>.

Code availability

Warp binaries, source code and user guide can be downloaded from <https://github.com/cramerlab/warp>. *BoxNet* source code can be downloaded from <https://github.com/cramerlab/boxnet>.

Author contributions

D.T. designed *Warp*'s architecture and all algorithms, and carried out all implementation and application. P.C. provided scientific environment, funding and additional interpretations and implications. D.T. and P.C. wrote the manuscript.

Competing financial interests

The authors declare no competing financial or other interests.

Introduction

Modern cryo-electron microscopy (cryo-EM) allows structures of protein assemblies to be solved at an ever-increasing speed. Automation of the data acquisition process and improved instrument stability have reduced the time required to collect hundreds of thousands of particle images to mere days. As the field attracts more scientists and the research projects become more ambitious, data collection time on a high-end microscope will remain a scarce and expensive resource¹. To make the best use of microscope time, the operator needs to continuously monitor data quality². Because many artifacts only become evident at the later stages of processing, the results of as many processing steps as possible should be available quickly to enable informed decisions during data collection.

The standard cryo-EM processing pipeline begins with the alignment of dose-fractionated image frames to cancel out sample motion³. This is followed by an estimation of the contrast transfer function (CTF)⁴. Particles are then selected and extracted from the images⁵. These steps are usually referred to as ‘data pre-processing’. Subsequent 2D and 3D classification of the particles serves to separate different complex conformations. Finally, 3D refinement is performed to obtain a three-dimensional map that is used to build a model.

Separate tools exist that can handle different steps of pre-processing. Global frame alignment and, optionally, dose weighting is performed by tools such as MotionCorr⁶, Unblur⁷ and SerialEM⁸; while MotionCor2⁹, alignparts_lmbfgs¹⁰ and Zorro¹¹ can also align the images locally to account for beam-induced motion (BIM). The CTF can be estimated using CTFFIND¹² or EMAN2.1¹³, while gctf¹⁴ will also estimate per-particle defocus if coordinates are available. Due to its complexity, particle picking remains the most actively researched part of the pipeline, with correlation-based approaches available in RELION¹⁵, FindEM¹⁶ or Signature¹⁷, and more advanced machine learning techniques implemented in Xmipp¹⁸ and DeepPicker¹⁹. Software packages like Appion²⁰, FOCUS²¹ and Scipion²² are built around these tools to provide a common interface and automate the execution of all steps. Although several tools can be combined to cover the full pipeline, reliably performing pre-processing of cryo-EM raw data at the speed of data collection remains a struggle, and manual reprocessing is possibly required afterwards. Only for exceptionally favorable samples, data processing has been reduced to “one push of a button” by fully automated procedures^{23, 24}. In most cases, however, human intervention remains required, and cryo-EM data pre-processing is generally still carried out subsequent to data collection.

Here we present *Warp*, a new software that takes over all pre-processing steps for 2D and tomographic cryo-EM data. *Warp* offers improved, real-time algorithms and a fully visual, intuitive interface that assists the user with immediate feedback and an augmenting presentation during data collection and pre-processing. The new algorithms can handle both standard data collection strategies and advanced scenarios such as tilted 2D data collection²⁵ or dose-symmetric tilt series²⁶, in a transparent, unsupervised manner. Coupled with automated acquisition software^{2, 8}, *Warp* provides a continuous low-latency stream of reliably picked and corrected particle images that can be seamlessly fed into 2D classification, *ab initio* reconstruction, and 3D refinement using other packages²⁴. *Warp* offers an accurate impression of the data quality during data collection, monitors the

microscope behavior, and enables advanced data analysis during data acquisition. In favorable cases, this can result in high-resolution cryo-EM structure solution during ongoing data collection.

Results

Rationale of *Warp*

We aim to enable microscope users to evaluate, correct, and process cryo-EM raw data immediately during data acquisition. The rationale was to provide a unified, streaming interface between the data acquisition software that produces the raw data, and existing software solutions for 2D classification and 3D refinement of cryo-EM single-particle data, such as RELION²³ or cryoSPARC²⁴. We called the resulting software package ‘*Warp*’ in reference to its fast correction of object distortions that occur in cryo-EM, and its GPU-based implementation that results in quick output. *Warp* is meant to be used for the pre-processing of both 2D and tilt series cryo-EM data. *Warp* can be installed on standard platforms and operated by non-expert users via a streamlined user interface (UI) that has been developed in parallel to the underlying algorithms to augment their operation. *Warp* was designed to be widely applicable for biological data acquisition at any cryo-EM facility and substantially speeds up the process of cryo-EM structure determination with improved results.

Overall design

A schematic of the computational steps carried out by *Warp* is provided in Fig. 1. For simplicity, we first describe the workflow for 2D data, before we describe the application to tomographic tilt series at the end of the results section. At the beginning of the pipeline, *Warp* reads any new data saved by the acquisition software. *Warp* then estimates and corrects the object motion across the micrograph movie both globally and locally. Next, *Warp* fits a spatially resolved CTF model, enabling the assignment of local defocus values to any particles extracted from the micrograph later. *Warp* then uses a neural network-based approach to automatically pick particles from the data with very high accuracy. Finally, *Warp* exports the resulting dose-weighted particle images to a downstream structure determination program such as RELION²³ or cryoSPARC²⁴, which carry out 2D and 3D classification, map refinement and reconstruction. During pre-processing, *Warp* provides a comprehensive overview of important data parameters, allowing the operator to tune acquisition settings to achieve optimal results faster. In the following we describe the most important components in more detail.

User interface

Warp’s UI is designed to help the user to comprehend and interact with the thousands of micrographs generated routinely during cryo-EM data collection (Fig. S1). The ‘Overview’ tab displays various properties, such as defocus, estimated resolution, amount of motion, or particle count, for all processed items as interactive plots. The user can immediately grasp the statistics, observe intrinsic patterns, and make an informed decision to adjust the acquisition strategy, e.g. to tune the lens astigmatism, increase the stage settling time, or skip a bad grid square. A range can be specified for every plotted parameter to automatically

exclude lower-quality images from downstream processing. Any item can be quickly inspected in more detail in a tab called 'Fourier & Real Space'. Here, a display of the power spectrum and the CTF fit allows optimization of CTF fitting parameters. In the real-space view, a deconvolution filter and neural net-based denoising (Methods, Fig. S2) can be instantly applied to micrographs to improve the contrast and make the particles more visible to the human eye. The defocus variation obtained through local CTF estimation can be overlaid semi-transparently. Particle picking results can be assessed, and edited manually. Dedicated dialogs assist the user with tasks like micrograph list export, particle extraction, template matching, tomogram reconstruction, and neural network training.

Motion correction

The motion, i.e. the translational shift observed between frames, is due to two effects: movement of the mechanical sample stage, and BIM. Stage movement causes global shifts over the entire field of view, whereas BIM leads to shifts between adjacent micrograph patches^{6, 27}. While stage drift can lead to rapid shift changes between frames, BIM occurs more slowly after rapid relaxation during initial exposure³. *Warp* corrects for both global drift and local BIM at variable temporal resolution (Fig. S3). The strategy is similar to the one used by MotionCor2⁹, except that *Warp* does not apply additional *a priori* assumptions about BIM beyond those imposed by the parameter grid resolution (Methods). As a result, *Warp* corrects efficiently and thoroughly for the two types of motion that occur during cryo-EM data acquisition in any kind of sample morphology and orientation.

Estimation of local defocus and resolution

The CTF model can be estimated based on the power spectrum (PS) of a micrograph. However, the defocus varies over the micrograph area due to stage inclination, uneven sample surface, or an uneven particle distribution along the optical axis²⁸. *Warp* provides a flexible way to model local defocus in spatial and temporal dimensions without the need for *a priori* knowledge of particle positions. Instead of one global estimate, a tilted plane or more complex geometry is fitted to the PS of a movie patch to estimate local defocus. A 1D average of all local power spectra rescaled to a common defocus value allows the user to easily assess whether fitting the more complex geometry recovered more Thon rings beyond the spatial frequencies used for the fitting (Fig. S4). Thus, *Warp* goes beyond state-of-the-art CTF estimation by providing a spatially resolved model without the need for *a priori* knowledge of particle positions and hyper-parameter tuning. The spatially resolved CTF model can converge on the correct solution for tilts as high as 60°. This makes *Warp* a useful tool for tilted 2D data collection, which has been shown to improve resolution isotropy for samples with preferred orientation²⁵.

Particle picking with BoxNet

The next step in cryo-EM structure determination is the accurate selection of particles from the corrected micrographs. *Warp* includes a novel particle picking routine based on a machine learning algorithm (Methods). For several years, the computer vision community has been using convolutional neural nets (ConvNets) to vastly outperform template matching in object recognition tasks^{29, 30}. First attempts to apply ConvNets to the particle picking problem in cryo-EM have shown performance on par with traditional template matching

approaches¹⁹. Today, deep residual network (ResNet) architectures enable the training of arbitrarily deep models³¹. *Warp* employs ‘BoxNet’ – a fully convolutional ResNet architecture with 72 layers, implemented in TensorFlow³². BoxNet was trained with real data curated manually from the EMPIAR repository³³ and synthetic data simulated from PDB³⁴ structures with a molecular weight range of 0.064–18 MDa. Because of these efforts, the pre-trained neural network bundled with *Warp* performs well on many particle species and is able to accurately mask out high-contrast artifacts, such as ethane. The performance of BoxNet compares favorably with available tools when representative cryo-EM data are used as input (Fig. 2). For heterogeneous data sets, BoxNet’s generic model can provide an advantage compared to the generic model in crYOLO³⁵ or the Laplacian of Gaussian approach in RELION 3.0³⁶ (Fig. S5).

Retraining of BoxNet

Since BoxNet’s performance can vary between different samples, *Warp* offers a retraining interface for it. Such retraining leads to a very high accuracy in automated particle picking. For retraining, the user must provide positive examples of particle positions. Using ~1000 examples, retraining of BoxNet typically takes less than 10 minutes. After retraining, the user can pick the same micrographs with the re-trained network and select more examples for another round of retraining if required. To decrease the need for retraining in the future, *Warp* also provides the option of submitting training data to a central repository. *De novo* training will be carried out by us periodically with all deposited data, and the resulting updated pre-trained BoxNet offered to the community. The training set is centrally curated and a list of particle species in the current version is available from <https://github.com/cramerlab/boxnet>. The BoxNet version name will be stored in each micrograph’s metadata to ensure reproducibility of picking results obtained with older versions.

Online pre-processing during data collection

Warp is optimized for processing raw cryo-EM data immediately during data collection. Files written out by the image acquisition software are detected automatically in the specified input folder and added to the list of ‘processable items’ in *Warp*. Each item maintains its metadata in an XML file that includes the exact previous processing settings. *Warp* continuously performs the processing steps necessary to bring each item into accord with the currently specified settings. All results can be immediately inspected during processing. Items can be forcibly included (i. e. exempted from the quality filters) or excluded from downstream processing. Processing must be stopped to change the settings or to retrain the BoxNet model. If changes were made, *Warp* will first reprocess all outdated items. During online processing, *Warp* executes all processing steps for a micrograph within less than 30 seconds after the data become available. In our experience, high-quality data of complexes of RNA polymerase II enable the user to obtain detailed 2D classes of particles and 3D reconstructions at better than 5 Å resolution using *Warp* and cryoSPARC within only a few hours after the start of data collection.

Interoperability with other software

To ensure interoperability with a plethora of cryo-EM tools, *Warp* allows the user to import and export data at several steps in its workflow using widely accepted formats and standards.

Raw movie data in the MRC, TIFF and EM formats are supported, and a 'headerless' option allows the user to manually specify properties of an unknown binary format. Image data are exported in the widely used MRC format, whereas all metadata are saved in the STAR format, adhering to the conventions established by RELION and adopted by many other tools. Any pre-processing step can be turned off if required. Results obtained with other tools can be imported to skip or benefit from particular algorithms in *Warp*. For instance, particle positions can be imported to export aligned particle averages, update their CTF models with *Warp*'s local estimates, to obtain a comprehensive overview of the particle distribution in a large project, or to retrain a BoxNet model. Frame alignment data can be exported to initiate a more accurate, reference-based alignment in RELION 3.0. Micrograph and particle lists adhering to user-selected quality criteria can be quickly prepared and exported. Taken together, *Warp* is highly flexible and allows for easy interoperation with other tools used by the community.

Pre-processing tomographic data

Warp can also be used to pre-process data from cryo-electron tomographic (cryo-ET) tilt series. *Warp* can reconstruct tomograms from a tilt series and perform template matching in tomograms with available 3D structures. The (sub)-tomogram reconstruction considers the local CTF, sample distortion and magnification anisotropy. Additionally, a deconvolved version of the tomograms can be produced using the same interface to help with their visual evaluation. To ensure the CTF model is accurate, *Warp*'s CTF fitting procedure goes beyond fitting the tilt images individually. Instead, local patch 2D power spectra from all tilts are fitted simultaneously, with constraints imposed on the inter-tilt angles, and regularizing assumptions made for the progression of phase shift and astigmatism. The tilt series CTF fitting can also be performed as part of the online processing.

Template matching

In addition to picking particles of a new species of unknown shape with a system like BoxNet, finding a previously known structure in new data is central to many stages of cryo-EM data processing. The structure must be compared at many different orientations to every position in the new data while considering the CTF. *Warp* implements template matching only for 3D templates, because *de novo* 2D particle picking is better handled by a neural network such as BoxNet. A template volume can be provided by the user or automatically downloaded from the EMDB³⁷ through the same UI. For template matching, 2D micrographs are subdivided into tiles, and normalized cross-correlation is computed between the tiles and 2D projections generated from the 3D template at the specified angular intervals, convolved with the local 2D CTF (Methods). All local correlation peaks with a minimum inter-peak distance corresponding to the template particle diameter, and the corresponding best-scoring template orientations are saved so that the user can later instantly explore different peak thresholds without repeating the costly correlation step. This procedure is also implemented for tomographic volumes, where the local patches are replaced by local sub-volumes, and the local 3D CTF is considered to incorporate knowledge of defocus and the missing wedge. In both cases, the matching should be performed at a resolution significantly lower than the expected map resolution to avoid

template bias³⁸. In case the map resolution does not surpass the resolution used for template matching, the procedure should be rerun at a lower resolution, and the results reprocessed.

Software implementation

Warp is written in the programming languages C#, C++ and CUDA C. The expressiveness of C# and the availability of powerful development tools kept the high-level data management layer brief and maintainable. *Warp*'s rich UI is enabled by the Windows Presentation Foundation (WPF) framework. All performance-critical parts are implemented to run on a GPU. Central data primitives, such as 2D movies and tilt series, and all associated algorithms are wrapped in a stand-alone C# library called 'WarpLib'. The granularity of most of these methods is fine enough to make them useful for applications beyond those implemented in *Warp*. Thus, WarpLib has the potential to speed up the development of future GPU-enabled tools that provide new functionality around the same data. We intend to keep developing *Warp* to enable state-of-the-art, rapid cryo-EM data pre-processing in the future. Updates will be shared with the community via GitHub.

Benchmarking for 2D data

To test the performance of *Warp*, we reprocessed a published single-particle cryo-EM data set for the influenza hemagglutinin trimer²⁵ (Methods) (Fig. 3). We chose this case for benchmarking because the processing of a 150 kDa protein imaged at 40° tilt required a significant amount of manual screening in the original analysis²⁵, providing a challenging test case for the *Warp* pipeline. With the original set of 130,000 particles, cryoSPARC reached a similar resolution as that reported in the original analysis (Fig. 3a, b), showing that refinement in cryoSPARC and RELION yields equivalent results for this data set. However, because this particle set and the general particle population both exhibit significant heterogeneity, we draw the comparison between results obtained after subjecting all data to the same 3D classification steps in cryoSPARC (Methods). For the original set, the best class containing 57,346 particles reached a global resolution of 3.9 Å with a B-factor of -200 Å². The same particles, updated with the defocus information from *Warp*, reached a notably higher resolution of 3.5 Å with a B-factor of -170 Å². This suggests that *Warp*'s local CTF model is more accurate than the per-particle CTF fitting in gCTF¹⁴ used in the original study. *Warp* processing also estimated a narrower range of astigmatism amplitudes (Fig. 3c), in agreement with the assumption of a stable optical system. For the full, completely automated *Warp* pre-processing pipeline, the best class containing 249,495 particles reached a global resolution of 3.2 Å with a B-factor of -170 Å², accompanied by a significantly increased level of detail in the map (Fig. 3a). This improvement from 3.5 Å to 3.2 Å is due to the higher particle count, which was obtained by *Warp* in a fully automated fashion at no time cost to the user. After classification in cryoSPARC, the best classes contained 45% and 51% of all particles in the original EMPIAR-10097 set and *Warp*'s automatically picked set, respectively, suggesting a similar degree of particle 'purity' in the manual and automated approaches. Furthermore, this demonstrates that tilted data collection can lead to near-atomic resolution with minimal efforts at the data pre-processing step.

The pre-processing of 668 movies from EMPIAR-10097 in *Warp* required ca. 3 h using a system with 4 Nvidia Titan X GPUs, thus averaging to ca. 220 movies per hour. *Warp*

v1.0.7, released during the revision of this manuscript, increased the speed to ca. 1400 movies (dimensions: 3710x3838x40) per hour on the same hardware. This speed allows *Warp* to keep up with any of the current automated acquisition schemes, which collect ca. 120 movies per hour on a Titan Krios microscope (Thermo Scientific, USA), and can reach up to 300 movies per hour in the most favorable cases. Taken together, our results establish *Warp* as a very useful tool for high-performance, automated cryo-EM data pre-processing.

Complementarity of *Warp* with other tools

We further tested *Warp*'s performance on a data set of β -galactosidase particles³⁹ that is often used for benchmarking purposes (Fig. 4). Because of the sample's high structural stability, these data stress the software's ability to obtain particularly accurate CTF and motion estimates to reach the highest end of cryo-EM resolution. Furthermore, the particles are difficult to pick due to the low defocus and prevalence of high-contrast objects in the micrographs. To estimate the frame alignment accuracy independently of 3D refinement, we calculated the average CTF fit quality for aligned movie averages processed with MotionCor2⁹ or *Warp*. *Warp*'s averages could be fitted to 2.6 Å, and those of MotionCor2 to 2.7 Å (Fig. 4c), indicating slightly better frame alignment in *Warp*. After refinement of particles from the full, completely automated *Warp* pre-processing pipeline, the best class containing 127,000 particles reached a global resolution of 2.09 Å with a B-factor of -35 Å².

To test the advantage of using reference-based refinements implemented in RELION 3.0³⁶, the same particle set was first subjected to beam tilt refinement and reached a resolution of 1.95 Å with a B-factor of -29 Å². Adding per-particle defocus refinement did not improve the resolution further. Adding reference-based frame alignment, referred to as 'polishing' in RELION, improved the resolution further to 1.86 Å with a B-factor of -26 Å². This slightly surpasses the result reported³⁶ for the same set of refinements in RELION 3.0 by 0.04 Å, possibly through cleaner initial particle picking. Whereas the additional refinements improved the resolution beyond that achieved with the pure *Warp* pre-processing pipeline, we note that these procedures required the data to be fully classified and refined first. In contrast, *Warp* provided its accurate initial results before any downstream processing took place. This was evident for the hemagglutinin trimer data, where *Warp* obtained the local defocus values immediately, whereas several rounds of refinement were required by RELION 3.0³⁶. Thus, the pre-processing in *Warp* and reference-based refinements in RELION 3.0 are complementary approaches, whose combination leads to faster convergence and higher resolution.

Benchmarking for tilt series data

To assess the benefits of using the full local 3D CTF for template matching in tomograms compared to a binary missing wedge mask, we matched 7 tomograms from a publicly available 80S ribosome data set⁴⁰. Among the 3,288 top-scoring matches, the false positive rate was 1% when using the full local 3D CTF, and 15% when using the binary missing wedge mask. At 15.9 standard deviations, the CTF-aware correlation peaks rose 2.7 times higher above the respective background distribution than the 5.8 standard deviations of the CTF-unaware peaks (Fig. 5), thus allowing more accurate matching.

To benchmark the proposed CTF estimation routine and sub-tomogram export for tilt series, we pre-processed, exported and refined particles from two publicly available data sets. For EMPIAR-10045 and EMPIAR-10164 (subset of 5 tomograms used in the assessment of NovaCTF⁴¹), we obtained a resolution of 11.6 Å with 3,200 particles (Fig. 6a, b), and 3.8 Å with 22,000 particles (Fig. 6c, d), respectively. These results slightly surpass the resolution figures of 12.8 Å and 3.9 Å reported in the original studies.

Discussion

Here we present *Warp*, a novel software tool for real-time evaluation and pre-processing of cryo-EM data during data collection. *Warp* bridges between the microscope and available software for 2D classification and 3D reconstruction, taking care of all pre-processing steps in an automated fashion. *Warp* also helps the microscopist to monitor data quality during acquisition, thus allowing to optimize the acquisition strategy during ongoing data collection. This contributes to making the best use of microscope time and prevents disappointing results during traditional data processing long after the data collection is concluded. *Warp* includes the deep learning-based models to perform reliable, automated particle picking and image denoising. Benchmarking studies establish *Warp* as a very reliable tool that matches or surpasses available software in performance. *Warp* is easy to install, has an intuitive user interface, and will be maintained and improved over the next years.

To further improve cryo-EM data collection efficiency in the future, we envision an automated feedback loop between the acquisition software and *Warp*. Such feedback would help to maximize the number of high-quality images collected by quickly skipping areas with low estimated resolution or particle count, adjusting the stage settling time dynamically based on *Warp*'s motion estimate for the previous image, and performing the time-consuming focus estimation procedure less frequently, using defocus values from *Warp*'s precise CTF fits instead. Additional feedback from *Warp* users will also contribute to further development of the software, and updates will be shared with the community as they become available.

Online Methods

Spline interpolation on multi-dimensional grids

Many methods in *Warp* are based on a continuous parametrization of 1–3-dimensional spaces. This parameterization is achieved by spline interpolation between points on a coarse, uniform grid, which is computationally efficient. A grid extends over the entirety of each dimension that needs to be modeled. The grid resolution is defined by the number of control points in each dimension and is scaled according to physical constraints (e. g. number of frames or pixels) and available signal. The latter provides regularization to prevent overfitting of sparse data with too many parameters. When a parameter described by the grid is retrieved for a point in space (and time), e. g. for a particle (frame), B-spline interpolation is performed at that point on the grid. To fit a grid's parameters, in general, a cost function associated with the interpolants at specific positions on the grid is optimized. In the following, we distinguish between 2–3 spatial grid dimensions (X and Y axes in

micrographs; X, Y and Z in tomographic volumes), and a temporal dimension as a function of the accumulated electron dose. We note that B-splines are only used to interpolate parameters, not image data. For the latter higher-order schemes are used.

Motion model

Two sources contribute to the observed translational shift between frames in a dose-fractionated image sequence. First, mechanical stage instability leads to rapid shift changes that are uniform within the entire frame. Second, beam-induced motion (BIM) causes slowly changing, local motion. Warp considers the physical properties of both sources in its motion model, using two sets of grids to parametrize the frame shifts and sample deformation. Global motion is described by two grids, X_{global} and Y_{global} , with high temporal, and no spatial resolution. The temporal resolution can match the number of frames, or, in case finer dose fractionation is performed to reduce intra-frame motion, the resolution can be lower to regularize a potentially overfitted model. BIM is described by two grids, X_{local} and Y_{local} , with a temporal resolution of at most 3, and a spatial resolution of typically 4–5 in both dimensions. The overall shifts required to bring the same object in all frames into a common reference frame are then defined as $(X_{\text{global}} + X_{\text{local}}, Y_{\text{global}} + Y_{\text{local}})$.

Global and local motion correction

In the absence of known particle positions and high-resolution reference projections, individual frame patches are aligned to their averages. The movie is subdivided into groups of 512^2 px patches with a 50 % spatial overlap, masked with a raised cosine. To simplify computation, the images are transformed into Fourier space where complex multiplication replaces translation. For each group, the patches are shifted according to the interpolants at their extraction positions using the current grid values. The average of a group's shifted patches is then compared to the individual patches to calculate the patch costs as

$$C = \sum_i \sum_f |I_{i,f} - \bar{I}_f|^2,$$

where i denotes the frame index, f denotes the spatial frequency, I is the Fourier transform of a shifted patch frame, and \bar{I} is the average of all shifted patch frames. The shifts are obtained by interpolating on the current state of the parameter grids at the patch frame's position in space and time. The derivative is approximated numerically with the symmetric difference quotient. The overall cost for all grid control points is the sum of all patch costs, and the derivative for each grid control point is the weighted sum of the derivatives of all patches affected by it. The weights for each control point's derivative can be precomputed by applying a one-pixel shift to the control point and storing all resulting non-zero patch shifts. The cost and derivatives are used by the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm⁴² to optimize the values of all control points. The optimization is performed in several steps to improve global convergence. In the first step, the temporal resolution of X_{global} and Y_{global} is set to 3, and increased to the next power of 3 in subsequent steps until the desired temporal resolution is reached.

Further details of this motion fitting algorithm are available as pseudo-code in **Supplementary Note 1**.

Contrast transfer function estimation in micrographs

The CTF analytically describes the convolution applied to the images by the electron-optical system. Estimating its properties with high precision is essential for reversing the effects and obtaining high-resolution reconstructions⁴³. Whereas the methodology for measuring defocus and astigmatism from a micrograph's power spectrum (PS) has been well-established^{12, 44}, the recent increase in EM map resolution calls for a more localized approach. Local defocus variation of a seemingly flat sample can exceed 60 nm within a single micrograph, resulting in an out-of-phase CTF for some particles at resolutions beyond 3 Å. Attempts to address this issue by fitting the defocus per-particle have been made¹⁴, but they require knowledge of particle positions, and lack robustness for all but the largest particle species. Even with a local smoothing approach, per-particle defocus requires high particle density to not lose accuracy compared to a global estimate. On the other hand, strong local irregularities in the specimen surface are almost never observed in tomographic volumes *in vitro*²⁸, suggesting per-particle precision might be unnecessary.

Estimation of local defocus

To parametrize the defocus, a single grid typically consists of 5x5 spatial control points. Optionally, another single grid with exclusively temporal resolution tracks the development of the phase shift generated by a Volta Phase Plate⁴⁵. Global parameters, such as astigmatism magnitude and angle, are optimized as additional scalars in the model. In practice, the effect of using a more complex geometry or temporal resolution appears negligible. However, the increased signal of future camera hardware might make these options relevant. Like in the frame alignment procedure, groups of patches matching the desired PS size (e. g. 512²—1024² px) are extracted with a spatial overlap of 50% from the raw movie data, transformed into Fourier space, and converted to PS by taking their squared amplitudes. If no temporal resolution is desired, each group will be averaged to a single PS to save resources. Similarly, in the absence of spatial resolution, the same frame from all groups will be averaged.

For the initial, exhaustive search, a 1D rotationally averaged PS is calculated from all patches. A B-spline with 3 control points is then fitted through it and subtracted to remove most of the background. The user-defined range of defocus and, optionally, phase shift values is evaluated by matching a simulated 1D CTF. The result with the highest normalized correlation is then used to estimate the 1D PS background and envelope more accurately to consider both in subsequent 2D CTF fitting. The cost function to be minimized is calculated between a 2D PS, and the 2D CTF simulated based on the local defocus at the patch extraction position:

$$C = \sum_i \sum_f \left(\left| CTF_{i,f} \right| \cdot E_f \right)_{Norm} \cdot \left(\left| I_{i,f} \right| - B_f \right)_{Norm},$$

where i denotes the frame index (in case a temporal dimension is used), f denotes the spatial frequency within the range used for fitting, CTF is the 2D contrast transfer function, I is the FT of a patch frame, E is the envelope of the 1D PS, B is the background of the 1D PS, and $(\dots)_{Norm}$ is a normalization operator that brings the value distribution to mean = 0, standard deviation = 1. The defocus and, optionally, phase shift is obtained by interpolating on the current state of the respective parameter grid at the patch frame's position in space and time. The derivative is approximated numerically with the symmetric difference quotient. The same pre-computed weights strategy as in the frame alignment procedure is employed for the control point derivatives. An L-BFGS algorithm finally optimizes the model for all control point values.

The PS of a tilted plane will usually only show low-resolution Thon rings, regardless of what model was used for the defocus gradient. To provide the user with feedback on whether the more complex defocus model is beneficial, the 2D spectra from all patches, whose parameters are herein referred to as the “source” PS, are rescaled and rotationally averaged to a 1D PS with a single defocus value, referred to as the “target” PS, such that the CTF phases match using the following scaling function:

$$x'_\varphi = \sqrt{\frac{-\sqrt{p_\varphi'^4 \cdot p_\varphi^4 \cdot (C_S^2 \cdot \lambda^4 \cdot f_N^4 \cdot x_\varphi^4 + 2C_S \cdot \lambda^2 \cdot f_N^2 \cdot p_\varphi^2 \cdot x_\varphi^2 \cdot z_\varphi + p_\varphi^4 \cdot z_\varphi'^2)} - p_\varphi'^2 \cdot p_\varphi^4 \cdot |z_\varphi'|}{C_S \cdot \lambda^2 \cdot f_N^2 \cdot p_\varphi^4}}$$

where $'$ denotes the “target” PS coordinate system, and its absence denotes the “source” PS coordinate system; φ is the sampling angle coordinate, p is the anisotropic pixel size, C_S is the spherical aberration, λ is the electron wavelength, f_N is the spatial Nyquist frequency, z is the anisotropic defocus value, and x is the sampling radius coordinate. A similar formulation was provided before⁴⁴ for the special case of isotropic pixel size, and was used to reduce the comparison between CTF and PS to a 1D problem. However, *Warp* performs the fitting in 2D and only uses the rescaling for visualization purposes. If the complex defocus model fits the data better, the recovery of additional high-resolution Thon rings can be observed in the 1D average.

Further details of this CTF fitting algorithm are available as pseudo-code in **Supplementary Note 1**.

Resolution estimation

To estimate the useful resolution range, a normalized cross-correlation value between the averaged 1D PS and the simulated 1D CTF curve is calculated within a sliding window. The window size at any given position scales to twice the width between the zero points of the closest CTF peak, but is not allowed to fall below 16 samples. The resolution limit is then reported as the frequency where the cross-correlation falls below 0.3 for the first time. Since the higher number of optimizable parameters allows for some overfitting, it is important that the useful resolution extends beyond the range used for fitting.

Contrast transfer function estimation in tilt series

The single micrograph CTF estimation procedure with planar sample geometry described in the previous section can be used for tilted 2D data collection. However, full tilt series pose additional challenges for CTF fitting. Mechanical stage instabilities and imperfect eucentric height setting necessitate additional exposures for tracking and focusing⁸ to correct the stage position between individual tilt images. Thus, the defocus cannot be assumed to stay constant, or change smoothly over the course of a tilt series. Each tilt image requires its own defocus value, which can be challenging due to the small amount of signal available. Even at $120 \text{ e}^-/\text{\AA}^2$ for an entire series of 60 images, each tilt only has $2 \text{ e}^-/\text{\AA}^2$ to perform the same estimation as for a $40 \text{ e}^-/\text{\AA}^2$ 2D image, while striving to achieve comparable accuracy.

CTF estimation in tilt series has traditionally received less attention than its equivalent in 2D data, with the most recent publication⁴⁶ predating the introduction of direct electron detectors and phase plates. As the resolution obtainable through sub-tomogram averaging has come close to parity with 2D data⁴⁷ since then, simplifying assumptions such as the neglect of astigmatism⁴⁸ or the assumed flatness of the sample can limit the resolution. Combined with the lack of integration of dedicated tilt series CTF estimation tools into common sub-tomogram averaging pipelines⁴⁹, this has created a situation where state-of-the-art studies^{47, 50} employ tools designed for 2D data such as CTFFIND¹².

To improve the fit accuracy, the individual tilt image fits must be subjected to a common set of constraints. As the imaged sample content does not change significantly throughout the tilt series, 1D background and envelope can be derived from the average 1D spectrum of all tilt images. The relative tilt angles and the tilt axis orientation are known to a higher precision than could be derived from fitting a planar geometry *de novo*, and are kept constant throughout the optimization as suggested previously⁴⁸. However, the absolute inclination of the sample plane is unknown. This is especially critical in some of the typical applications of tomography, like the imaging of lamellae prepared through FIB milling because a lamella can be tilted by over 20° relative to the grid. This additional inclination remains constant throughout the tilt series, and is made a single optimizable parameter for all tilt images. Astigmatism and, optionally, phase shift can be kept constant throughout 2D image exposures, but can benefit from a temporally resolved model in a tilt series where the overall exposure is fractionated over a much longer time, e. g. 20—30 min. *Warp* uses 3 control points in the temporal dimension to model these parameters.

With these considerations, the full estimation process is as follows. 2D patches are extracted from all aligned tilt movie averages, as described in the micrograph CTF fitting procedure, and treated in parallel in all subsequent calculations. To provide a better initialization for the per-tilt defocus searches, an estimate for the average defocus of the entire series is obtained by preparing 1D spectra from all patches, and comparing them to simulated CTF curves for the defocus values at the respective positions and tilts, taking into account the fixed relative tilt information and the currently tested average defocus (and phase shift, optionally). This search is performed exhaustively over a range of values specified by the user. The result is used as the starting point of a more complex optimization. Defocus values for all individual tilts, 3 astigmatism magnitude/angle pairs, 3 optional phase shift values, and the two global inclination angles (i. e. the plane normal) are optimized using the L-BFGS algorithm with

the derivatives obtained numerically as described in the micrograph CTF fitting section. Upon convergence, the 1D spectra of all patches are rescaled to a common defocus value. This is especially useful for validation in tilt series since the individual images will have very noisy spectra. If the useful resolution range does not extend sufficiently beyond the fitting range, the latter is automatically decreased and the optimization repeated.

In our experience, the direction of the tilt axis is often miscalibrated. Correct handedness in structures obtained from sub-tomogram averaging does not guarantee the tilt angle sign is not flipped. In Warp, a positive rotation around the positive Y image axis is defined to result in an increased underfocus at positions to the positive X side of the tilt axis, i. e. those parts of the sample come physically closer to the electron beam source. The CTF fitting procedure in Warp can detect such mistakes by optionally repeating the optimization with the tilt angles flipped, and notifying the user if the “wrong” set of angles provides a better fit. Such a test can be used to re-calibrate the acquisition software for future data collection.

Considerations for tomogram reconstruction

Whereas the process of 3D map reconstruction from 2D images of single particles is well established today, full-tomogram reconstruction breaks some of the simplifying assumptions so they must be handled explicitly to obtain better results. In the 2D case, the CTF can be assumed to be the same for all parts of a single particle image, although corrections for a wider range of defocus values in images of large objects have been proposed⁵¹. In a tomographic tilt series, the highest tilt image can show a defocus spread of 1 μm or more. Accounting for such variations in local defocus is necessary for reaching high resolution⁴¹. Furthermore, each region in the tomographic volume is reconstructed from images with different CTFs, and the zeros and peaks of those CTFs will not overlap in Fourier space. CTF-based weighting of individual projections is commonplace for 2D data^{23, 52}, but the algorithms used in tomographic reconstruction do not go beyond CTF phase flipping, giving all spectral components equal weight^{41, 53}. This gives spectral components with pure noise (CTF = 0) the same weight as the best available signal ($|\text{CTF}| = 1$) if they overlap. Performing CTF-, dose- and tilt-based weighting later in sub-tomogram averaging has been shown to be beneficial⁴⁹, but it has an even more significant effect when applied at the level of initial tomogram reconstruction. Anisotropic magnification has been described in the past⁵⁴ and is routinely corrected in 2D data. In tomography, the real-space distortion is even more pronounced than in single particle reconstructions because the distances affected by the distortion are more than 1 μm , i. e. the extent of the entire tomogram, leading to positional errors on the order of nanometers in scenarios where the anisotropy does not coincide with the tilt axis.

Tomogram reconstruction

Warp takes the local defocus and sample distortion, as well as magnification anisotropy into account when reconstructing full or partial tomographic volumes. For a partial reconstruction at any position in the volume, the original 2D images are sampled at the following positions:

$$\mathbf{s} = \mathbf{R}_{Euler}(0, \alpha_{Tilt}, \psi_{Tilt}) \cdot \mathbf{p} + \mathbf{o}_{Tilt}$$

where \mathbf{R}_{Euler} is the rotation matrix for 3 Euler angles following the Xmipp convention⁵⁵, α is the stage tilt angle, ψ is the in-plane angle of the tilt axis, \mathbf{p} is the particle position within the tomographic volume, and \mathbf{o} is the in-plane offset of the tilt axis. The coordinates are centered within the volume and images. The CTF for each 2D image is calculated using a defocus of:

$$z = z_{Tilt} + \frac{\mathbf{s} * \mathbf{n}_{Tilt}}{\mathbf{n}_{Tilt,z}}$$

where z_{Tilt} is the average defocus estimated for the tilt image, \mathbf{n}_{Tilt} is the sample plane normal, $\mathbf{n}_{Tilt,z}$ is the z component of the normal, and * denotes the scalar product between two vectors. The reconstruction is performed in Fourier space using a gridding algorithm²³, with the data weighted by the respective CTF, and the dose- and tilt-dependent heuristic from RELION⁴⁹, but without the final deconvolution step (i. e. the weights are inserted as $|\text{CTF}|$, not as CTF^2). To obtain a full tomogram, *Warp* reconstructs a uniform grid of small, cubical volumes with an overlap of 50%, and inserts the central 50% into the overall volume to account for artifacts associated with Fourier space reconstruction at the borders of the local volumes. This ensures the corrections can be applied with local precision and remain reasonably continuous between adjacent sub-volumes.

Export of corrected data

Whereas the *Warp* model for a movie or tilt series describes the non-linear deformation of the entire particle ensemble and its environment, it is unclear whether this deformation gradient stays continuous throughout a single particle, i. e. if the protein structure is subject to the same compression and shearing as the ice around it. Many recent high-resolution maps were reconstructed using particles extracted from dose-weighted averages produced by MotionCor2⁹. The tool assumes the deformation gradient to be continuous in all parts of the image, and will thus deform images of particles and ice in the same way. This will be beneficial if the underlying physical model is indeed continuous. However, it also distorts the CTF locally without passing any knowledge of the distortion to downstream processing tools. In case of a strong local change in the motion direction, this will result in an artifact similar to lens astigmatism.

Warp assumes a continuous deformation field when exporting dose-weighted averages of whole 2D movies, i. e. each pixel will be shifted according to the grid interpolants at that exact position. This has the benefit of uniformly sharper images for visual inspection and particle picking. For particle and sub-tomogram extraction, however, the entire particle image will be shifted uniformly according to the grid interpolants at the particle's center. This keeps the CTF true to its fitted analytical description, but makes the assumption that the protein is more rigid than the surrounding ice and thus deforms less due to BIM. For whole-tomogram reconstruction, a hybrid approach is pursued: the local volumes are produced

using the same procedure as sub-tomogram extraction, but the combined volume is largely continuous depending on how small the local volumes were.

Dose weighting in *Warp* adds a B-factor of -4 \AA^2 per 1 e-/\AA^2 of dose, similar to a heuristic published previously⁷. While a different heuristic is used in MotionCor2⁹ and Unblur⁷, the accuracy of both approaches is of decreased significance as data-driven re-weighting is likely to be performed using an approach like the “particle polishing” in RELION 3.0.

Particle picking with a residual neural net

In the past years, the recipe for improving the performance of deep learning algorithms has been “deeper networks, more training data”. Outside of cryo-EM, deep ResNet architectures have been demonstrated to enable the training of very deep models by essentially solving the vanishing gradient problem³¹. At the same time, the EMPIAR raw data repository³³ has accumulated a diverse collection of 2D cryo-EM data sets that can be leveraged for training. *Warp* employs a model with 35 ResNet blocks and 2 conventional convolution layers (Fig. S6) to segment a micrograph into 3 classes: background, particle, and high-contrast artifact (e. g. ethane drops). The input window has a constant size of 256^2 px. After initial convolution with 32 5x5 kernels the data are processed by a sequence of 5 groups of each 5 ResNet blocks. At the beginning of each but the first group, the data are down-sampled by a factor of 2, while the number of channels is doubled. This enables the recognition of an increasing number of large, higher-order features. After reaching a spatial extent of 16×16 in the contractive part of the network, the data are processed by the expanding part: a sequence of 4 groups of each 2 ResNet blocks. At the beginning of each group, the data are up-sampled by a factor of 2. Each group’s output is concatenated with the output of its mirror counterpart from the contractive part in a U-Net-like fashion⁵⁶. This combines the global context and higher-order features obtained in the contractive part with the higher spatial resolution of the previous layers. After reaching the original extent of 256×256 in the expanding part, the data are processed by the final 2 ResNet blocks and projected onto 3 channels by convolution with 3 1x1 kernels. A pixelwise argmax operation finally retrieves the most likely label at each location in the original image. The model graph and variables are initialized and saved using TensorFlow’s Python API, while all subsequent training and inference are performed through its C API, wrapped in C# classes.

To segment a micrograph, it is down-scaled to the standard 8 \AA/px resolution and divided into 256^2 px tiles with 64 px overlap. Each tile is extracted, normalized, and presented to the network. All tiles’ softmax and argmax outputs are combined and subjected to a user-defined threshold for the softmax value to remove uncertain picks. Connected components of pixels labeled as “particle” are extracted, and their centroids are used as particle positions. In cases where two particles overlap according to a user-defined diameter, the particle with the bigger connected component is kept. To obtain a mask from the “artifact” label, a threshold of 0.1 is applied to the softmax values, and connected components with less than 20 pixels are removed. The remaining pixels are saved as a binary mask. The final list of particle positions only contains those with a user-defined minimum distance to the masked regions.

Initial training of BoxNet

11 EMPIAR and 14 in-house data sets (Fig. S5, **Table S1**) each contributed 20–50 micrographs to the training set. Additionally, synthetic data were prepared from 21 PDB models (**Table S1**) using a modified version of the InSilicoTEM⁵⁷ package, contributing ca. 1600 particles per species. The simulated data contained only one species per micrograph, although more heterogeneous examples might be added in the future. The training set was split 9/1 for training/validation, and trained with the momentum optimizer in TensorFlow 1.5 using a learning rate gradually decreasing from 1E-2 to 1E-5. The normalized data were augmented in each training epoch by extracting the 256² px window at random positions, and applying random rotation, flipping, shearing, and Gaussian noise with a random standard deviation between 0.0 and 0.6. This augmentation was observed to have an excellent regularizing effect, as the final training and validation scores were virtually identical. The training was performed for 800 epochs, using a batch size of 1.

Retraining of BoxNet

User-supplied positions of positive examples and, optionally, areas of increased and decreased certainty in the micrographs are automatically converted to training data. If requested, the training set is diluted with data from the latest version of the centrally curated set (in the following referred to as ‘old data’) in a 1:1 ratio to prevent possible overfitting of the new data. The retraining regime is identical to initial training, but lasts only 100 epochs. During the retraining, 4 metrics are calculated continuously for every batch: the old network’s accuracy for old and new data, and the retrained network’s accuracy for old and new data. Ideally, the retrained network’s accuracy for new data will improve to approach or even surpass the old network’s accuracy for old data by the end of the retraining process, whereas the accuracy for old data will stay constant.

Template matching in micrographs and tomograms

2D micrographs are subdivided in tiles with an overlap matching twice the template particle diameter. For each square, 2D projections of the template are prepared at user-defined angular intervals, convolved by the square’s CTF, and normalized to mean = 0, standard deviation = 1 in real space. The square’s FT is multiplied by the conjugate of the projection’s FT, and an IFT yields the cross-correlation scores for all positions within the square. These scores are normalized by the local standard deviation within the square. The scores are compared for all template orientations, and the best one is stored for each pixel within the square. Finally, the result is cropped to exclude a border matching the template particle diameter, and combined with the results from other squares to obtain the correlation scores for the entire micrograph. A local peak search is performed using the template particle diameter as the minimum distance, and all peak positions are stored for further processing. Template matching in tomographic volumes follows the same concept. Instead of square tiles, local cubes are cross-correlated with the template convolved by the local 3D CTF. Optionally, a spectrum whitening of the target micrograph/tomogram can be performed as previously described⁵⁸. This has the benefit of equalizing the spectral noise amplitudes for all spatial frequencies, effectively giving more weight to the higher frequencies and sharpening the correlation peaks.

Deconvolution

In the absence of a phase plate, the CTF will be dominated by its sine component, i. e. have very little contrast in the lowest spatial frequencies. This creates a high-pass filter effect in the raw data and, due to increasingly noisier higher frequency components, makes it hard to assess the image content visually. On the other hand, a phase plate creating the desired phase shift of $\pi/2$ will apply a low-pass filter in defocused images, rendering them blurry. Both scenarios do not affect subsequent alignment and averaging procedures significantly, and the filters will be reversed in the final reconstruction by dividing its 3D FT by the weighted average of all contributing CTFs. This becomes possible because the spectral signal-to-noise ratio (SSNR) is sufficiently high after averaging enough particles with different CTFs. However, even in single images the lowest frequency components often contain enough signal so that boosting them by inverting the CTF will increase the visible low-frequency contrast while maintaining acceptable noise levels. This provides conventional images with a better definition of object boundaries, making their manual selection easier. In defocused phase plate images, this improves sharpness.

To construct a Wiener-like filter, *Warp* makes ad hoc assumptions about the SSNR that can be adjusted by the user. The SSNR is assumed to be a combination of an exponential decay curve and a raised cosine high-pass filter:

$$SSNR_f = e^{-f \cdot 100F} \cdot 10^{3S} \cdot H_f,$$

where f denotes the spatial frequency, H is an optional high-pass filter, F is the custom fall-off parameter, and S is the custom strength parameter. The factors for F and S are empirically tuned so that the default values of 1 produce good results for typical direct electron detector data, although adjustments might be required in some cases. The SSNR is then used in a Wiener-like filter:

$$I'_f = \frac{I_f \cdot CTF_f}{CTF_f^2 + SSNR_f^{-1}},$$

where I is the FT of the image, and CTF is the 2D contrast transfer function. The shape of the SSNR curve prevents the lowest frequency components from being boosted too much, giving rise to a noisy sample background, and acts as a low-pass filter at the same time to suppress the noisy high frequency components. An example of such a filter and its effect on a 2D micrograph are shown in Fig. S2b. In practice, a higher electron dose helps to obtain good low-frequency contrast in conventional images. The commonly used dose of 30–40 $e^-/\text{\AA}^2$ works well for holey grids with thin ice, while more might be required in the presence of carbon support or thick ice. The deconvolution works especially well in tomograms, where the overall dose often surpasses 100 $e^-/\text{\AA}^2$.

Denosing improves particle visibility

Even after deconvolution, micrographs will often display a high amount of noise that makes their visual inspection difficult. Deep convolutional neural network-based denoisers have

been shown to perform better than hand-crafted statistical models^{59, 60}. [REFs]. However, their training has traditionally required pairs of noisy and noise-free observations of the same signal. This prevented denoiser training on cryo-EM data, where a noise-free ground truth is not available due to quickly increasing radiation damage in sensitive samples. Recently, the Noise2Noise principle⁶¹ has been proposed to circumvent this limitation while achieving comparable performance. Pairs of independently noisy observations of the same signal are used in training as input and output. In the absence of correlation between the noise patterns, the expectation value is the underlying noise-free signal, which the model learns to predict. Pairs of independently noisy data are readily available in cryo-EM. Since every micrograph's signal is fractionated in a multitude of frames, computing the aligned averages of all odd and even frames renders a pair of almost identical observations with independent patterns of camera shot noise. By using odd and even frames instead of the first and second halves of a movie, the effect of radiation damage is very similar in both averages.

Warp adopts the published Noise2Noise U-net-like network architecture with one change; a batch normalization layer is added before each leaky ReLU layer. The denoising is performed on overlapping 256^2 px tiles, and the central 128^2 px windows are combined to produce the final image shown in Fig. S2c. *Warp* generates training data from processed movies automatically. These images are also pre-deconvolved. While deconvolution and denoising are independent processes, their combination provides the most natural-looking result for the human observer. Once enough training examples have been generated (up to 128 are supported by default), the user can request to retrain the model, which will take a few minutes. A default pre-trained model is provided with *Warp* that covers a narrow range of conventional and VPP data. However, the best denoising performance is achieved by retraining the model for every new data set. Like deconvolution, denoising can be applied on-the-fly to any micrograph displayed in *Warp*, using the default or retrained model. While denoising provides a great visual aid for human interpretation of raw data, the process removes all signal that cannot be reliably distinguished from noise in a single observation. Thus, denoised particle images do not render themselves to averaging as a means of increasing the SNR at higher resolution.

Benchmarking for 2D data

For the influenza hemagglutinin trimer benchmark, raw movie data and pre-extracted particles from EMPIAR-10097 were downloaded. The movies were processed with the full *Warp* pipeline using the following settings: motion correction with a temporal resolution of 40 for the global motion, and 5x5 spatial resolution for the local motion, using the 0.03–0.25 Nyquist range and a B-factor of -400 \AA^2 ; CTF estimation with 6x6 spatial resolution, using the 0.1–0.35 Nyquist range; particle picking with a BoxNet model retrained on particles from 3 micrographs, using the default 0.95 threshold. Quality filters were applied in *Warp* as follows: defocus between 0.3 and $5.0 \mu\text{m}$, resolution better than 8 \AA , intra-frame motion of at most 1.5 \AA , particle count above 120. Particles were extracted from the micrographs meeting these filters and subjected to processing in cryoSPARC: no 2D classification was performed; *ab initio* refinement was performed with 6 classes and no symmetry; the 6 classes were then refined heterogeneously, with no symmetry imposed; the only class

showing the expected Hemagglutinin structure was refined with C3 symmetry. The original particle set from EMPIAR-10097 was subjected to 3 different processing strategies. First, the full set was refined in cryoSPARC with C3 symmetry using the original CTF estimates. Second, the full set was subjected to the same classification and refinement as the particles from Warp, using the original CTF estimates. Third, particles from the Hemagglutinin class obtained in the second processing branch were updated with local CTF estimates from Warp, and refined again with C3 symmetry. Resolution estimates were obtained for all maps using the respective masks automatically generated by cryoSPARC.

For the β -galactosidase benchmarking studies, raw data from EMPIAR-10061 were downloaded. The movies were processed with the full *Warp* pipeline using the following settings: motion correction with a temporal resolution of 38 for the global motion, and a 5x5 spatial resolution for the local motion, using the 0.03–0.60 Nyquist range and a B-factor of -160 \AA^2 ; CTF estimation with 5x5 spatial resolution, using the 0.08–0.60 Nyquist range; particle picking with a BoxNet model retrained on particles from 5 micrographs, using a threshold of 0.30. No quality filters were used as the data already represent a high-quality subset curated for the initial publication. Picked and extracted particles were subjected to 2D and 3D classification with C1 symmetry in RELION 2.1 to remove incomplete particles. The remaining particles were refined with D2 symmetry. The final half-maps were then used to refine beam tilt and per-particle defocus in RELION 3.0. Global motion tracks for all movies were exported from *Warp* to RELION 3.0 to perform Bayesian particle polishing.

To assess the frame alignment accuracy in *Warp* independently of downstream map refinement, β -galactosidase movies were aligned in *Warp* as described above, and using the default settings in MotionCor2. CTF fitting was performed with 5x5 spatial resolution, using the 0.08–50 Nyquist range. Frequency-dependent fit quality was calculated as described in the ‘Resolution estimation’ section, and all resulting curves averaged. The resolution was then estimated at a cut-off value of 0.3.

Benchmarking for tilt series data

For the template matching benchmark, pre-aligned tilt series from EMPIAR-10045 were downloaded. The defocus was estimated, and full tomograms were reconstructed with a pixel size of 10 \AA in Warp. The 80S ribosome map derived from these data in the original publication⁴⁰ and deposited under EMD-3228 was used as the template. Template matching was performed on the 10 \AA /px tomograms with an angular sampling of 7.5 $^\circ$, using a local 3D CTF. The same steps were performed using a binary missing wedge mask instead of the 3D CTF. The picked positions were screened manually to determine the false positive rate. Background statistics were calculated for the correlation volume trimmed to remove the vacuum region, and excluding 48 px windows around the peaks.

To benchmark CTF estimation and sub-tomogram export on EMPIAR-10045 data, the particles previously picked through template matching were exported together with their 3D CTF volumes at a pixel size of 5 \AA . The sub-tomograms were then subjected to 3D refinement in RELION 3.0 without prior classification.

To benchmark CTF estimation and sub-tomogram export on HIV-1 particles, raw data from EMPIAR-10164 were downloaded. A subset of 5 tilt series previously used by the authors of NovaCTF⁴¹ was selected. Movies were aligned in *Warp* using only global alignment with a temporal resolution of 5. Gold beads were picked manually and used to align the tilt series in IMOD⁶². Full tomograms were reconstructed with a pixel size of 5 Å in *Warp*. Template matching was performed with the EMD-4015 map with an angular sampling of 7.5 ° and C6 symmetry. A custom script was used to remove particles not fitting into a regular hexagonal grid as described previously⁴⁷. The particles were exported together with their 3D CTF volumes at a pixel size of 1.35 Å. The sub-tomograms were then subjected to 3D refinement in RELION 3.0 without prior classification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank members of the Cramer lab for beta-testing early versions of *Warp* and providing feedback on bugs in the software. We thank C. Bernecky, S. Dodonova, W. Hagen, D. Lyumkis, C. Plaschka, J. Söding and Y.Z. Tan for critical reading of the manuscript. PC was supported by ERC Advanced Grant TRANSREGULON (grant agreement No 693023) of the European Research Council, the Deutsche Forschungsgemeinschaft (SFB 860), and the Volkswagen Foundation.

References

1. Saibil HR, Grünewald K, Stuart DI. A national facility for biological cryo-electron microscopy. *Acta Crystallographica Section D: Biological Crystallography*. 2015; 71:127–135. [PubMed: 25615867]
2. Suloway C, et al. Automated molecular microscopy: the new Legimon system. *Journal of structural biology*. 2005; 151:41–60. [PubMed: 15890530]
3. Brilot AF, et al. Beam-Induced Motion of Vitriified Specimen on Holey Carbon Film. *Journal of structural biology*. 2012; 177:630–637. [PubMed: 22366277]
4. Huang Z, Baldwin PR, Mullapudi S, Penczek PA. Automated determination of parameters describing power spectra of micrograph images in electron microscopy. *Journal of structural biology*. 2003; 144:79–94. [PubMed: 14643211]
5. van Heel M. Detection of objects in quantum-noise-limited images. *Ultramicroscopy*. 1982; 7:331–341.
6. Li X, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods*. 2013; 10:584–590. [PubMed: 23644547]
7. Grant T, Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. 2015
8. Mastrorade DN. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology*. 2005; 152:36–51. [PubMed: 16182563]
9. Zheng SQ, et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. 2017; 14:331–332. [PubMed: 28250466]
10. Rubinstein JL, Brubaker MA. Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of structural biology*. 2015; 192
11. McLeod RA, Kowal J, Ringler P, Stahlberg H. Robust image alignment for cryogenic transmission electron microscopy. *Journal of structural biology*. 2017; 197:279–293. [PubMed: 28038834]
12. Rohou A, Grigorieff N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology*. 2015; 192:216–221. [PubMed: 26278980]
13. Bell JM, Chen M, Baldwin PR, Ludtke SJ. High resolution single particle refinement in EMAN2.1. *Methods (San Diego, Calif.)*. 2016; 100:25–34.

14. Zhang K. Gctf: Real-time CTF determination and correction. *J Struct Biol.* 2016; 193:1–12. [PubMed: 26592709]
15. Scheres SH. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of structural biology.* 2015; 189:114–122. [PubMed: 25486611]
16. Roseman AM. FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. *Journal of structural biology.* 2004; 145:91–99. [PubMed: 15065677]
17. Chen JZ, Grigorieff N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *Journal of structural biology.* 2007; 157:168–173. [PubMed: 16870473]
18. Sorzano C, et al. Automatic particle selection from electron micrographs using machine learning techniques. *Journal of structural biology.* 2009; 167:252–260. [PubMed: 19555764]
19. Wang F, et al. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *Journal of structural biology.* 2016; 195:325–336. [PubMed: 27424268]
20. Lander GC, et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *Journal of structural biology.* 2009; 166:95–102. [PubMed: 19263523]
21. Biyani N, et al. Focus: The interface between data collection and data processing in cryo-EM. *Journal of structural biology.* 2017; 198:124–133. [PubMed: 28344036]
22. de la Rosa-Trevin JM, et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of structural biology.* 2016; 195:93–99. [PubMed: 27108186]
23. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology.* 2012; 180:519–530. [PubMed: 23000701]
24. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods.* 2017; 14:290. [PubMed: 28165473]
25. Tan YZ, et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods.* 2017; 14:793–796. [PubMed: 28671674]
26. Hagen WJH, Wan W, Briggs JAG. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of structural biology.* 2017; 197:191–198. [PubMed: 27313000]
27. Campbell MG, et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure (London, England: 1993).* 2012; 20:1823–1828.
28. Noble AJ, et al. Routine Single Particle CryoEM Sample and Grid Characterization by Tomography. *eLife.* 2018; 7:e34257. [PubMed: 29809143]
29. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 1998; 86:2278–2324.
30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems.* 2012; 1:1097–1105.
31. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2016; 770:778.
32. Abadi M, et al. TensorFlow: a system for large-scale machine learning. *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation;* 2016. 265–283.
33. Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ, Patwardhan A. EMPIAR: a public archive for raw electron microscopy image data. *Nat Methods.* 2016; 13:387–388. [PubMed: 27067018]
34. Berman HM, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
35. Wagner T, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology.* 2019; 2:218. [PubMed: 31240256]
36. Zivanov J, et al. RELION-3: new tools for automated high-resolution cryo-EM structure determination. *eLife.* 2018; 7:e42166. [PubMed: 30412051]
37. Tagari M, Newman R, Chagoyen M, Carazo JM, Henrick K. New electron microscopy database and deposition system. *Trends in biochemical sciences.* 2002; 27:589. [PubMed: 12417136]

38. Henderson R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:18037–18041. [PubMed: 24106306]
39. Bartesaghi A, et al. 2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science (New York, N.Y.)*. 2015; 348:1147–1151.
40. Bharat TA, Scheres SH. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature protocols*. 2016; 11:2054–2065. [PubMed: 27685097]
41. Turonova B, Schur FKM, Wan W, Briggs JAG. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of structural biology*. 2017; 199:187–195. [PubMed: 28743638]
42. Nocedal J. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*. 1980; 35:773–773.
43. Sorzano CO, Otero A, Olmos EM, Carazo JM. Error analysis in the determination of the electron microscopical contrast transfer function parameters from experimental power Spectra. *BMC structural biology*. 2009; 9:18. [PubMed: 19321015]
44. Penczek PA, et al. CTER—Rapid estimation of CTF parameters with error assessment. *Ultramicroscopy*. 2014; 140:9–19. [PubMed: 24562077]
45. Danev R, Tegunov D, Baumeister W. Using the Volta phase plate with defocus for cryo-EM single particle analysis. *eLife*. 2017; 6
46. Voortman LM, Stallinga S, Schoenmakers RHM, Vliet LJv, Rieger B. A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM. *Ultramicroscopy*. 2011; 111:1029–1036. [PubMed: 21740865]
47. Schur FK, et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science (New York, N.Y.)*. 2016; 353:506–508.
48. Xiong Q, Morphew MK, Schwartz CL, Hoenger AH, Mastronarde DN. CTF determination and correction for low dose tomographic tilt series. *Journal of structural biology*. 2009; 168:378–387. [PubMed: 19732834]
49. Bharat TA, Russo CJ, Lowe J, Passmore LA, Scheres SH. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure (London, England : 1993)*. 2015; 23:1743–1753.
50. Hutchings J, Stancheva V, Miller EA, Zanetti G. Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nat Commun*. 2018; 9
51. Russo CJ, Henderson R. Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy*. 2018; 187:26–33. [PubMed: 29413409]
52. Grigorieff N. FREALIGN: high-resolution refinement of single particle structures. *Journal of structural biology*. 2007; 157:117–125. [PubMed: 16828314]
53. Kunz M, Frangakis AS. Three-dimensional CTF correction improves the resolution of electron tomograms. *Journal of structural biology*. 2017; 197:114–122. [PubMed: 27343995]
54. Grant T, Grigorieff N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *Journal of structural biology*. 2015; 192:204–208. [PubMed: 26278979]
55. Heymann JB, Chagoyen M, Belnap DM. Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *Journal of structural biology*. 2005; 151:196–207. [PubMed: 16043364]
56. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015
57. Vulovic M, et al. Image formation modeling in cryo-electron microscopy. *Journal of structural biology*. 2013; 183:19–32. [PubMed: 23711417]
58. Rickgauer JP, Grigorieff N, Denk W. Single-protein detection in crowded molecular environments in cryo-EM images. *eLife*. 2017; 6
59. Mao X-J, Shen C, Yang Y-B. Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections. *arXiv*. 2016

60. Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Transactions on Graphics (TOG)*. 2017; 36:107.
61. Lehtinen J, et al. Noise2Noise: Learning Image Restoration without Clean Data. arXiv. 2018
62. Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology*. 1996; 116:71–76. [PubMed: 8742726]

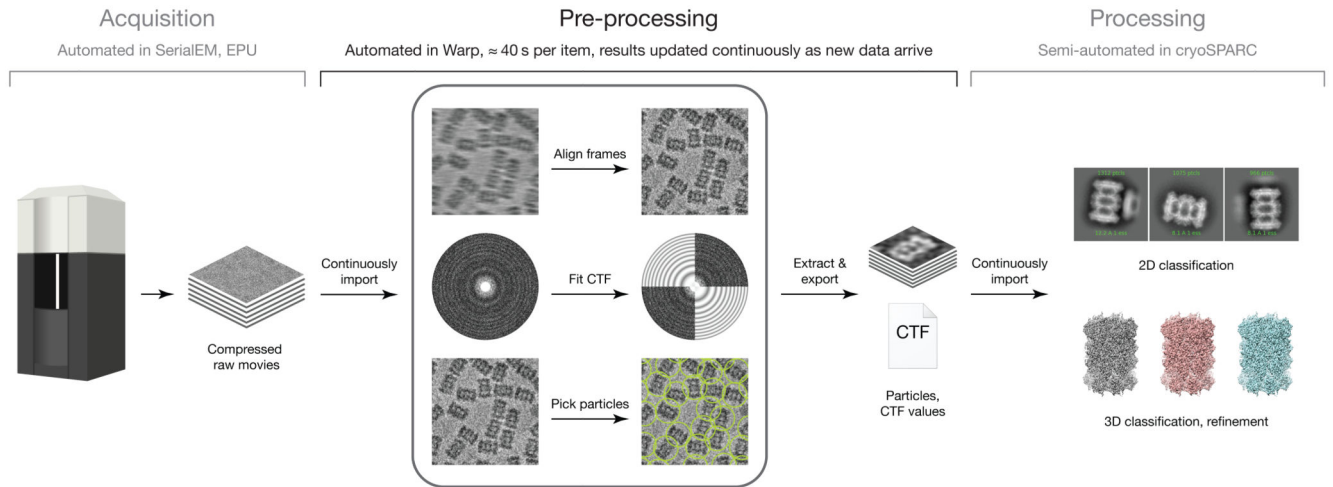


Figure 1. *Warp* handles all pre-processing steps to close a gap in the 2D cryo-EM pipeline. Data is acquired by the microscope in an automated fashion and stored as compressed stacks of movie frames. *Warp* continuously monitors its input folder for new files, and subjects them to all steps of the pre-processing pipeline: frame alignment, CTF estimation and particle picking. *Warp* writes out a stack of particles for each pre-processed micrograph and maintains a dynamically updated STAR file with references to all particles and their local CTF parameters. This file can be used as a data source in a tool such as cryoSPARC, which will periodically run subsequent processing steps like 2D classification and *ab initio* reconstruction on the latest set of particles.

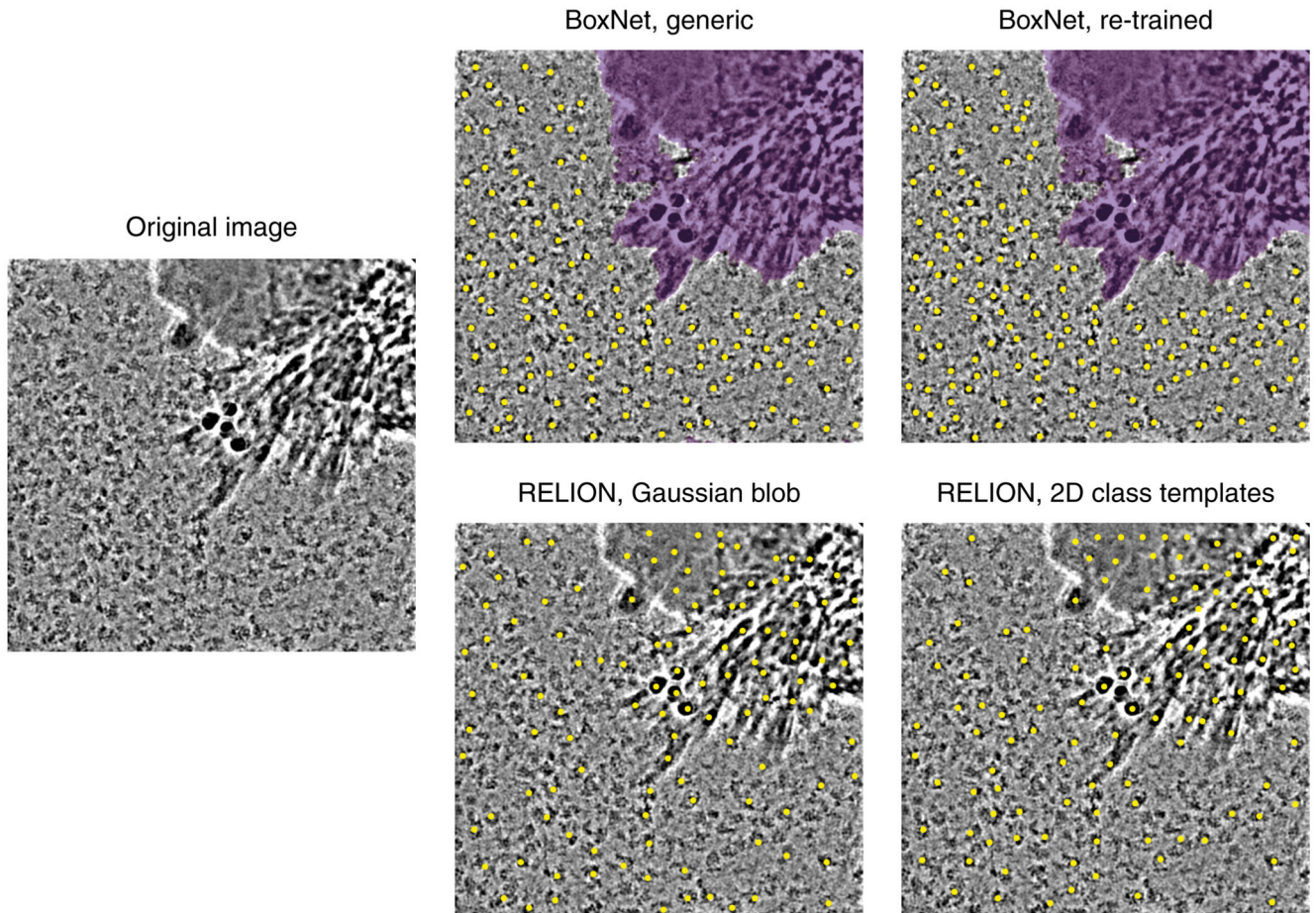


Figure 2. Automated particle picking with *Warp*'s deep learning-based BoxNet

Representative example of automated particle picking with BoxNet in *Warp* on a micrograph with high-contrast artifacts. Areas masked out automatically by BoxNet are colored purple. The generic version of BoxNet was never presented with the sample during training. The re-trained version was given 5 micrographs of the same sample, which did not include the one shown. The template-based picking in RELION used 25 class averages derived from 3000 particles, filtered to 20 Å. RELION's results show the 120 highest-scoring positions. For visualization purposes, the micrograph was deconvolved, high-pass filtered and cropped at the borders.

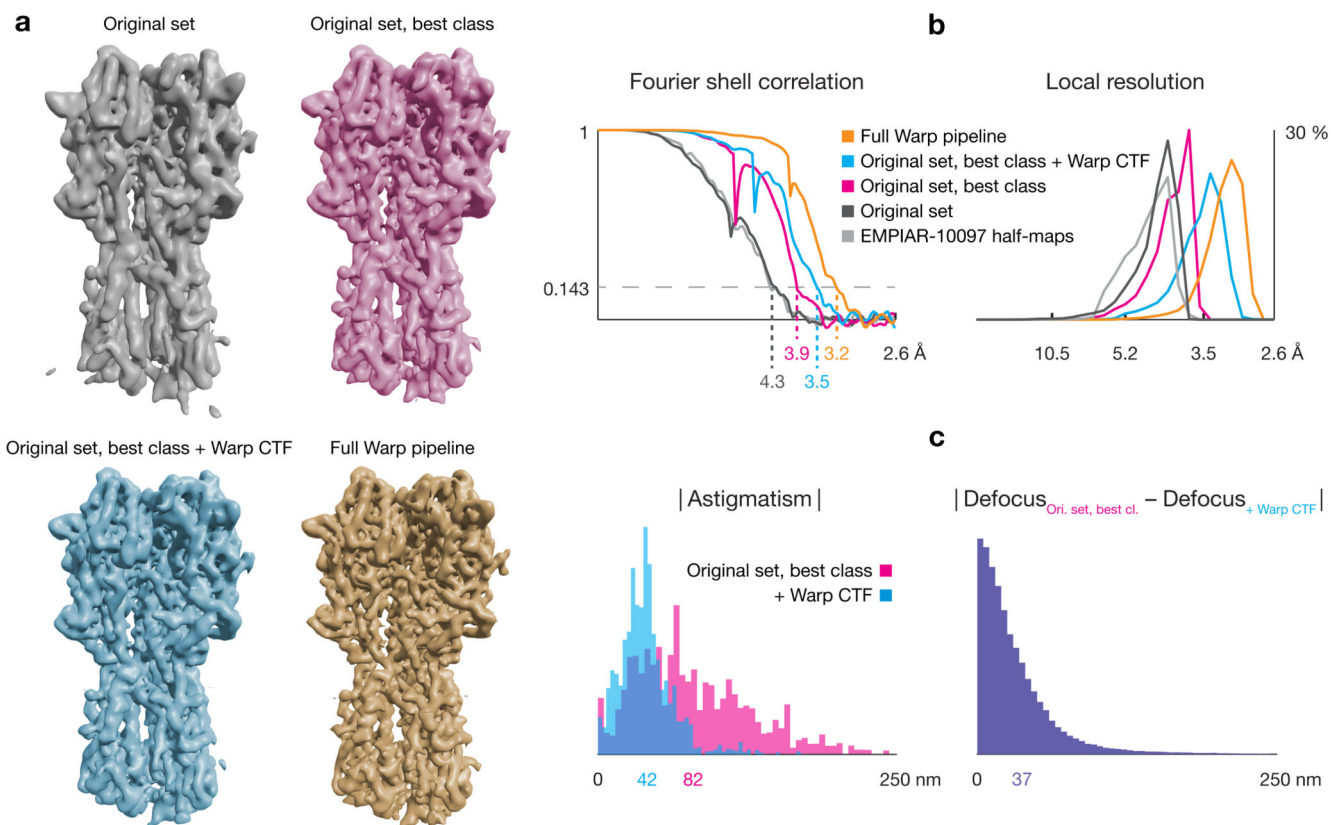


Figure 3. Warp's 2D pipeline improves cryo-EM density for influenza hemagglutinin.

As a benchmarking case we used the published EMPIAR-10097 data set containing influenza hemagglutinin trimer particles. 'Original set': 130,000 pre-extracted particles from EMPIAR-10097 with their original CTF parameters; 'Original set, best class': 57,346 particles from 'Original set' after 3D classification in cryoSPARC with their original CTF parameters; 'Original set, best class + Warp CTF': the same 57 346 particles, but with Warp's CTF estimates; 'Full Warp pipeline': 249,495 particles obtained from the raw EMPIAR-10097 data after unsupervised pre-processing in Warp and 3D classification in cryoSPARC.

(a) Isosurface renderings of the 3D maps generated with cryoSPARC using the respective sets of particles and CTF parameters, filtered to local resolution using the auto-tightened masks generated by cryoSPARC.

(b) Global masked FSC plots, and histograms of the local resolution used to filter the maps depicted in (a). 'EMPIAR-10097 half-maps' refers to the original half-map volumes deposited in EMPIAR-10097, obtained from the same 130,000 particles as 'Original set'.

(c) Histogram comparison between the original defocus parameters and those estimated by Warp for the 130 000 particles from 'Original set'. The mean value for each metric is specified underneath the horizontal axis in the same color as the corresponding histogram.

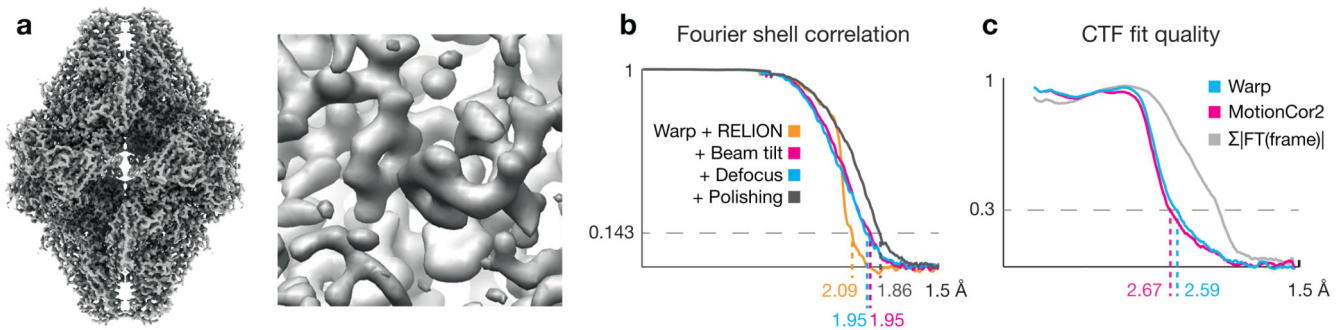


Figure 4. Warp's 2D pipeline in combination with RELION 3.0 improves cryo-EM density for β -galactosidase.

For our second benchmark, we used the published EMPIAR-10061 data set containing β -galactosidase particles. The data were processed using the full *Warp* pre-processing pipeline, and beam tilt, per-particle defocus and frame alignment were later refined against high-resolution references in RELION 3.0 to assess the additional improvement provided by these refinements.

(a) Isosurface rendering of the 1.86 Å map (left) and a detailed view of some of its sidechains, clearly displaying the aromatic rings (right).

(b) Global masked FSC plots for the map obtained with the *Warp* pipeline only, and for the additive effects of reference-based beam tilt and per-particle defocus refinement, as well as particle polishing in RELION 3.0.

(c) Average CTF fit quality curves for aligned movie averages produced with MotionCor2 and *Warp*. *Warp*'s averages can be fitted to a higher resolution, indicating more accurate frame alignment. For comparison, a fit quality curve is also included for amplitude spectra obtained from the average of individual frame spectra, which are invariant to residual inter-frame motion and radiation damage.

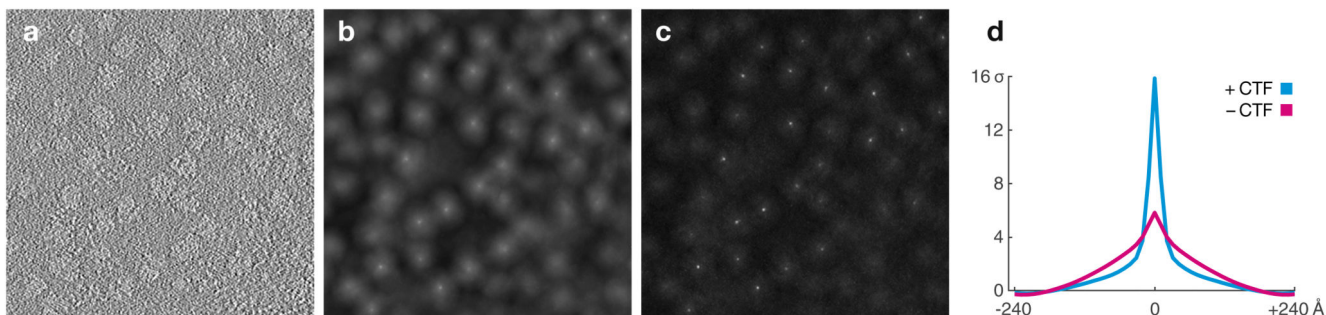


Figure 5. Effect of using the full local 3D CTF for template matching in tomograms.

During template matching, *Warp* multiplies the rotated 3D reference by the local 3D CTF before correlating it to a local portion of the tomogram volume, as opposed to multiplying it by a binary missing wedge mask. This produces sharper correlation peaks.

(a) XY slice through a tomogram reconstructed from EMPIAR-10045 data. The faint shapes of 80S ribosomes are visible.

(b) XY slice through the correlation volume at the same location as (a), using a binary missing wedge mask. White indicates higher correlation. The peaks are broad and hard to distinguish against the background.

(c) XY slice through the correlation volume at the same location as (a), using the full local 3D CTF. White indicates higher correlation. The peaks are sharper, leading to higher template matching accuracy.

(d) Rotational average of a 48 px window around all correlation peaks, mean-subtracted and normalized against the respective correlation background. 3D CTF-aware template matching (+CTF) produces peaks rising 2.7 times higher above the background compared to binary missing wedge masks (-CTF).

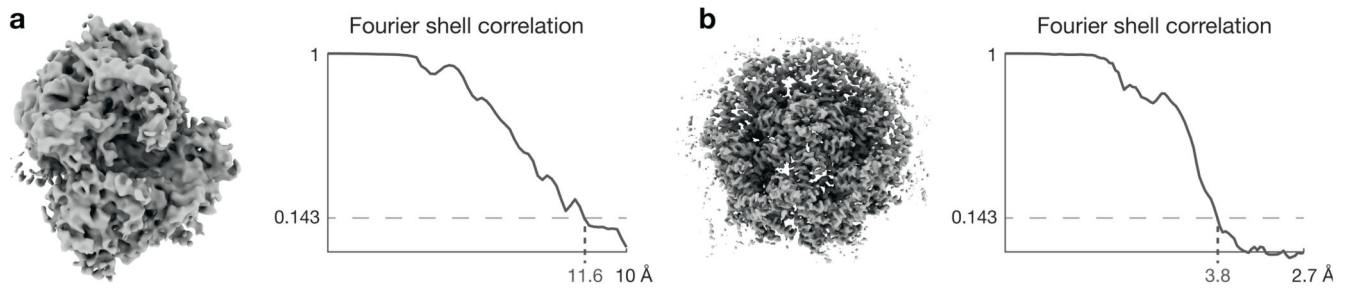


Figure 6. Sub-tomogram averaging results obtained by using *Warp*'s tilt series CTF estimation and sub-tomogram export.

To assess the benefits of the proposed CTF estimation and sub-tomogram export strategies, data from EMPIAR-10045 and EMPIAR-10164 were pre-processed and exported in *Warp*, and refined in RELION 3.0. Improved resolution was observed in both cases compared to published results.

(a) Isosurface rendering (left) and FSC plot (right) of the 80S ribosome sub-tomogram average obtained from EMPIAR-10045 data. The originally published resolution was 12.8 Å.

(b) Isosurface rendering (left) and FSC plot (right) of the HIV-1 sub-tomogram average obtained from 12% of the EMPIAR-10164 data. The originally published resolution for this subset was 3.9 Å.