SPHIRE-crYOLO:
A fast and well-centering automated particle picker for cryo-EM

Thorsten Wagner[1], Felipe Merino[1], Markus Stabrin[1], Toshio Moriya[1], Christos Gatsogiannis[1], and Stefan Raunser[1]

[1]Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Str. 11, 44227 Dortmund, Germany

Correspondence: stefan.raunser@mpi-dortmund.mpg.de

Abstract

Selecting particles from digital micrographs is an essential step in single particle electron cryomicroscopy (cryo-EM). Since manual selection is a tedious and time-consuming process, many automatic particle pickers have been developed. However, they have problems especially with non-ideal datasets. Here, we present a novel automated particle picking software called crYOLO, which is based on the deep learning object detection system "You Only Look Once" (YOLO). After training the network with 500 – 2,500 particles per dataset, it automatically selects the particles with high accuracy and precision reaching a speed of up to six micrographs per second. Importantly, we show that crYOLO can be trained to select previously unseen datasets paving the way for completely automated "on-the-fly" cryo-EM data pre-processing during data acquisition. CrYOLO is available as standalone program under http://sphire.mpg.de/ and will be part of the image processing workflow in SPHIRE.

Introduction

In recent years, single particle electron cryomicroscopy (cryo-EM) became one of the most important and versatile methods for investigating the structure and function of biological macromolecules. In single particle cryo-EM many images of identical but randomly oriented macromolecular particles are selected from raw cryo-EM micrographs, then digitally aligned, classified, averaged, and back-projected to obtain a three-dimensional structure of the protein.

Since usually more than 100,000 particles have to be selected for a near-atomic cryo-EM structure, many automatic particle picking procedures, often based on heuristic approaches, have been recently developed[1-6]. A popular selection approach, called template matching, cross-correlates the micrographs with pre-calculated templates of the particle of interest. However, this procedure is error-prone and the parameter optimization is often complicated. Although it works with optimal data, it often fails when dealing with non-ideal datasets where part of the particles overlap or are degraded and the background of the micrographs is contaminated with ice. In these cases, either the number of selected particles decreases or the number of false positives increases tremendously. The last resort is then often manual selection of particles, which is laborious and time-intensive.

To solve this problem two particle selection programs[7,8] have been recently developed, employing deep convolutional neural networks (CNN). CNNs are extremely successful in processing data with a grid-like topology[9]. For images this is undoubtedly the state-of-the-art method for pattern recognition and object detection. Similar to learning the weights and biases of specific neurons in a multi-layer perceptron[10], a CNN learns the elements of convolutional kernels. A convolution is an operation which calculates the weighted sum of the local neighborhood at a specific position in an image. The weights are the elements of the kernel which extracts specific local features of an image, such as corners or edges. In convolutional neural networks, several layers of convolutions are stacked, where the output of one layer is the input of the next layer. This enables CNNs to learn hierarchies of features.

Common modern object detection systems, like the particle selection tools by Wang et al. and Zhu et al.[7,8], employ a specific classifier for an object and evaluate it at every position. Here, they slide a window over the micrograph to crop out single regions, pass it through a CNN network and finally classify the extracted region. The confidence of classification is transferred into a map and the object positions are

estimated by finding the local maxima in this map. Using this approach, it is possible to handle more challenging datasets. However, as it classifies many cut out regions independently, it comes with a high computational cost. Moreover, since the classifier only sees the windowed region, it is not able to learn the larger context of a particle (e.g. to not pick regions near ice contamination).

In 2016, Joseph Redmon introduced the 'You Only Look Once' (YOLO) framework[11] as an alternative to the sliding window approach and reformulated the classification problem into a regression task, where the exact position of the particle is predicted by a network. In contrast to the sliding window approach, the YOLO framework needs only a single pass of the full image instead of multiple passes of cropped-out regions. Thus, the YOLO approach basically splits the image into a grid and predicts for each grid cell if it contains the center of a particle in the bounding box. If that is the case, it applies regression to predict the relative position of the particle center inside the cell as well as the width and height of the bounding box (Figure 1a). This procedure makes the detection pipeline rather simple and reduces the number of convolutions tremendously, which renders it fast but still accurate. During training, the YOLO approach only requires labeling positive examples, whereas the sliding window approach also needs the background and contamination to be labeled as negative examples. Moreover, since the network sees the complete image, it is able to learn the larger context and provides therefore an excellent framework to detect single particles in electron micrographs reliably and fast (Figure 1b).

Here, we present the novel single particle selection procedure crYOLO which utilizes the YOLO algorithm to select single particles in cryo-EM micrographs. We evaluated our procedure on three recently published high resolution cryo-EM datasets. The results demonstrate that crYOLO is able to accurately and precisely select single particles in micrographs of varying quality with a speed of up to six micrographs per second on a single GPU. This leads to a tremendous acceleration of particle selection, and significantly improves not only the quality of the datasets but also the final structure. Furthermore we show that crYOLO, trained on multiple datasets, is able to select previously unseen particles.

Results

Convolutional Neural Network

The program crYOLO builds upon a Python-based open source implementation[12] of YOLO and uses the deep learning library Keras[13]. Beyond the basic implementation we added support for MRC micrographs, single channel data and EMAN1[14] box files. We added a graphical tool to read and create box files for training data generation or visualization of the results in a user-friendly manner (Figure 2) and made changes in the original neural network architecture of YOLO[15].

The YOLO network consists of 22 convolutional and five max-pooling layers. In addition, it contains a passthrough layer to exploit fine grain features and the network is followed by a (1x1) convolutional layer that is used as a detection layer. YOLO uses a relatively coarse grid, which could lead to problems when detecting smaller particles as each grid cell can only detect a single particle. To tackle this, we took out the max-pooling layer after the 13th convolutional layer. Thus, the adapted network downsamples the input image only by a factor of 16 instead of 32 in original YOLO resulting in a finer effective grid size of 64x64 cells on an input image size of 1024x1024. By default, crYOLO bins the micrographs automatically to 1024x1024 pixels to reduce memory consumption and increase the detection speed. The finer grid is now suitable for detecting smaller particles. Finally, we removed the last six convolutional layers, the passthrough layer and placed a dropout layer in front of the detection layer to reduce overfitting (Table 1).

To increase the amount of training data and to reduce overfitting, we augmented our datasets for each training step of the crYOLO network by a random selection of the following methods: flipping, blurring, adding noise and random contrast changes.

Test datasets

We validated the performance of our program on three different cryo-EM datasets which are publicly available in the electron microscopy public image archive (EMPIAR) and produced high-resolution structures. These include the TcdA1 toxin subunit from *Photorhabdus luminescens*[16] (EMPIAR-10089, EMDB-3645), the *Drosophila* transient receptor channel NOMPC[17] (EMPIAR-10093, EMDB-8702) and the human peroxiredoxin-3 (Prx3)[18] (EMPIAR-10050, EMDB-3233).

Layer	Type	Filters	Size
Feature Extraction			
1.	Convolutional	32	3x3
	Max-Pooling		2x2
2.	Convolutional	64	3x3
	Max-Pooling		2x2
3.	Convolutional	128	3x3
4.	Convolutional	64	1x1
5.	Convolutional	128	3x3
	Max-Pooling		2x2
6.	Convolutional	256	3x3
7.	Convolutional	128	1x1
8.	Convolutional	256	3x3
	Max-Pooling		2x2
9.	Convolutional	512	3x3
10.	Convolutional	256	1x1
11.	Convolutional	512	3x3
12.	Convolutional	256	1x1
13.	Convolutional	512	3x3
14.	Convolutional	1024	3x3
15.	Convolutional	512	1x1
16.	Convolutional	1024	3x3
Detection			
	Dropout (0.2)		
1.	Convolutional	6	1x1

Table 1. CrYOLO network architecture. The CNN consists of 16 convolutional and four max-pooling layers for feature extraction. For detection, a single convolutional layer is used with a dropout layer in front to reduce overfitting. The latter is only used during training, not during prediction.

TcdA1 assembles in the soluble prepore state in a large bell-like shaped pentamer with a molecular weight of 1.4 MDa[19]. It has been our test specimen to develop the software package SPHIRE[16]. We previously used a dataset of only 7,475 particles of TcdA1 to obtain a reconstruction at a resolution of 3.5 Å[16]. Due to the large molecular weight of the specimen and its characteristic shape, particles are clearly discernible in the micrographs although a carbon support film was used. Therefore, the selection of the particles is straightforward in this case. We chose this test dataset

because of its small size, since the quality and number of selected particles will have likely an influence on the quality of the final reconstruction.

In the case of NOMPC, which was reconstituted in nanodiscs[17], the overall shape of the particle, sample concentration, and limited contrast make it difficult to accurately select the particles, even though it has a molecular weight of 754 kDa. Furthermore, the density of the nanodisc is significantly higher than the density of the extramembrane domain of the protein, thus the center of mass is not located at the center of particle but shifted towards the nanodisc density. A challenge for the selection program is therefore to accurately detect the center of the particles and avoid selection of empty nanodiscs. We therefore chose NOMPC as second test dataset.

Prx3 has a molecular weight of 257 kDa and shows a characteristic ring-like shape. The dataset is one of the first near-atomic resolution datasets recorded using a Volta phase plate[18] (VPP). The VPP introduces an additional phase shift in the unscattered beam. This increases the phase contrast significantly, providing a significant boost of the signal-to-noise ratio in the low-resolution range. This makes the structural analysis of low molecular weight complexes at high resolution possible and has revolutionized the cryo-EM field[20]. The VPP, however, enhances not only the contrast of the particles of interest, but also the contrast of all weak phase objects, including smaller particles (impurities, contamination, dissociated particle components) that would otherwise not be visible in conventional cryo-EM. This poses new complications with regard to particle selection, especially for automated particle picking procedures that cross-correlate raw EM images with templates. We therefore chose Prx3 as third test dataset.

Training of crYOLO

To train crYOLO, we manually selected initial training datasets. Dependent on the density of particles, the heterogeneity of the background, and the variation of defocus, more or fewer micrographs are needed. We found that 500 to 2,500 particles from at least five micrographs were sufficient to properly train the networks for the three datasets used in this study (Table 2). Ideally, each micrograph should be picked to completion. It should be emphasized that picking of negative examples (including background, carbon edges, ice contamination and “bad particles”) is not required at all

in crYOLO. Twenty percent of the micrographs that were used for manual selection were chosen for validation purposes.

	Manually picked particles	Micrographs
TcdA1	1,100	10
NOMPC	585	9
Prx3	2,500	5

Table 2: The number of manually picked particles and the corresponding number of micrographs.

To determine the influence of the size of the training dataset on the selection quality, we trained the crYOLO network on subsets of the manually picked particle dataset. We decreased the size of the subsets systematically and evaluated the results by calculating the precision and recall scores[21]. The recall score measures the ability of the classifier to detect positive examples and the precision score specifies the strength of the classifier not to label a true negative as true positive. Both scores are summarized with the integral of the precision-recall curves (AUC), which is an aggregate measure for the quality of detection. The larger the AUC value, the better the performance.

The selection performance was excellent, when using as few as 400 TcdA1 particles from four micrographs (Figure 3a). However, when fewer particles were used, such as 200, the training dataset was not large and representative enough to enable a precise selection (Figure 3a). Interestingly, training with 400 particles from four micrographs worked better than with 600 particles from six micrographs (Figure 3a). The reason for this unexpected behavior is the fact that four different defocus groups were used for the recording of the TcdA1 dataset. Therefore, selection of six micrographs instead of four results in an overrepresentation of certain defoci. However, this effect only appears when working with very small training datasets. In general, the user should make sure that the selection of micrographs covers the full defocus range when training crYOLO on manually picked particles.

We furthermore compared crYOLO with the original YOLO network using the full training dataset comprising 1,100 TcdA1 particles from 10 micrographs (Figure 3b). At an input image size of 1024x1024 the performance of the networks is very similar. However, for input images with the size of 512x512, crYOLO with an AUC of 0.71 is superior to YOLO which reaches only a very low AUC value of 0.31 in this case

(Figure 3b). Therefore, it is possible to train crYOLO on GPUs with less memory. However, the 1024x1024 input image size is preferable due to the overall better performance (Figure 3b). Importantly, for Prx3 crYOLO significantly outperformed YOLO using a training dataset of 2,500 particles from 5 micrographs (Figure 3c), clearly highlighting the power of crYOLO for small particles.

To quantify how well the particles are centered, we calculated the mean intersection over union (IOU) value of manually selected particles versus the automatically picked boxes using crYOLO. The IOU is defined as ratio of the intersecting area of two boxes and the area of their union and is a common measure for the localization accuracy. Picked particles with an IOU higher than 0.6 were classified as true positives. The mean IOU for TcdA1 and Prx3 was 0.88 ± 0.009 and 0.80 ± 0.004, respectively. This indicates a high localization accuracy of crYOLO. Since it was difficult to determine the proper center of each NOMPC particle even when selected manually, we did not calculate an IOU for this dataset.

Besides this classical evaluation, we additionally calculated 2-D classes using the iterative stable alignment and clustering approach ISAC[22] implemented in the SPHIRE software package[16]. 2-D clustering using ISAC relies on the concepts of stability and reproducibility and is capable of extracting validated and homogeneous subsets of images that can be reliably processed in further steps of the workflow. Therefore, the number of particles in these subsets indirectly reflects the performance of the particle picker. Moreover, we calculated 3-D reconstructions and compared it to the published reconstructions.

For TcdA1, crYOLO selected 11,436 particles from 98 micrographs. This is ~36% more particles than previously identified by the Gauss-Boxer in EMAN2[14] (Figure 4a-c). Also the number of 'good' particles after 2-D classification is significantly higher (Figure 4d,e), indicating that crYOLO is able to identify more true positive particles. This results in a slightly improved resolution of the 3-D reconstruction (Figure 4f,g).

For NOMPC, the authors picked originally 337,716 particles from 1,873 micrographs using RELION[23] (Figure 5b,c) and obtained a 3-D reconstruction at a resolution of 3.55 Å from 175,314 particles after extensive cleaning of the dataset by 2-D and 3-D classification in RELION[17]. CrYOLO was trained on 9 micrographs with 585 particles and selected only 228,499 particles (> 30% less) (Figure 5a,c). After 2-D classification and the removal of bad classes (Figure 5c,d), 178,458 particles were used

for the 3-D refinement yielding a slightly improved cryo-EM structure of NOMPC of 3.4 Å (Figure 5e-f). The reconstruction shows high-resolution features even in the region of the ankyrin-repeats (Figure 5g). Although the number of initially selected particles differs tremendously from crYOLO, the number of particles used for the final reconstruction is very similar, indicating that crYOLO selects less false positive particles than the selection tool in RELION. In this case, this also drastically reduced the steps and overall time of image processing, since 2-D classification was performed on a much lower starting number of particles and further cleaning of the dataset by a laborious 3-D classification was obsolete.

For Prx3, crYOLO identified 345,362 particles from 802 digital micrographs, which is comparable to the number of particles picked in the original article using EMAN2[18] (Figure 6a-c). We used two consecutive ISAC rounds to classify the particles in 2-D (Figure 6d). In the first one, we classified the whole dataset into large classes to identify the overall orientation of the particles. In the second, we split them into top, side and tilted views, and ran independent classifications on each of them. This procedure is necessary to avoid centering errors in ISAC. After discarding bad classes and removing most of the preferential top view orientations, the dataset was reduced to 43,852 particles. 3-D refinement of this particle stack in SPHIRE yielded a map at a nominal resolution of 4.7 Å (Figure 6e, f). In comparison, the final stack in the original article was composed of only 8,562 particles, which were refined to a 4.4 Å resolution map in RELION. Compared to our processing pipeline, the authors used three rounds of 2-D and four rounds of 3-D classification in RELION[23] to clean this dataset.

The small differences in the final resolution of the reconstructions can be attributed to the fact that all reconstructions reach resolutions near the theoretical resolution limit; at a pixel size of 1.74 Å, 4.4 Å represents approximately 0.8 times Nyquist. Consistent with this, the FSC curves of the reconstructions from both sets of particles are practically superimposable (Figure 6e). Interestingly, a large fraction of the particles was discarded during refinement, both by the original authors – 97.6% – and by us – 87.7% kept. In our case, this is not due to the quality of the particle selection. Indeed, 310,123 of our picks were included into the stably aligned images after ISAC. Instead, we deleted most of the particles found in the preferred top orientation, which represents roughly half the dataset. The remaining discarded images correspond to correct picks in regions where the protein forms clusters. In those cases, while the

particles were correctly identified by crYOLO, the particles were not used for single particle reconstruction.

Computational efficiency

We used a desktop computer equipped with a NVIDIA Geforce GTX 1080 graphics card with 8 GB memory and an Intel Core i7 6900k to train the networks and select the particles. The time needed for training the networks was 13 to 25 minutes (Figure 7a). Whereas the same was needed for training crYOLO on the TcdA1 and NOMPC datasets, the training time was lower for Prx3.

The selection process was very fast for all three datasets (Figure 7b). The selection of the Prx3, TcdA1 and NOMPC datasets reached a speed of ~1.1, ~4.6 and ~6.0 micrographs per second, respectively. This difference in speed can be mainly explained by the density of particles within each micrograph. The fewer particles per micrograph, the faster crYOLO performs. While the micrographs of the TcdA1 and NOMPC datasets have on average ~ 140 particles, the Prx3 micrographs contain ~ 456 particles. We expect even faster performance for micrographs containing fewer particles.

Generalization to unseen datasets

Ideally, crYOLO would specifically recognize and select particles that it has not seen before, paving the way for fully automated particle selection. In order to reach this level of generalization, we trained the network with a combination of 34,813 manually picked particles from 331 micrographs derived from two VPP and nine conventional cryo-EM datasets, including the Prx3 dataset described above, the T20S proteasome[24] (EMPIAR 10025) and nine other datasets from different projects that are currently being processed in our research group. The molecular weight of the complexes from the respective datasets ranges from 64 kDa to 1.1 MDa.

Using this generalized crYOLO network, we automatically selected particles of TcdA1 (EMPIAR 10089), an ATP-synthase[25] (EMPIAR 10023), an in-house ribosome and an influenza hemagglutinin dataset[26] (EMPIAR 10097) (Figure 8a-d). Although crYOLO had never seen these particles before, it performed very well in selecting them specifically while avoiding to pick ice contamination or particles surrounded by contamination (Figure 8c). This clearly shows that crYOLO is able to learn general features for identifying particles in context of the micrograph.

To assess the quality of the generalized crYOLO network, we compared its performance selecting TcdA1 with that of the network that was directly trained on this protein (see above). As expected, the AUC and localization accuracy are better for the directly trained network (Figure 8e). However, the generalized crYOLO network selected a similar number of particles (Figure 4d); the AUC of 0.748 (Figure 8e) and the IOU of 0.77 show that its performance is of sufficient quality to select a good set of particles. Indeed, the TcdA1 particles, selected by the generalized crYOLO network resulted in a reconstruction of similar resolution and quality as the particles from the TcdA1-trained crYOLO (Figure 4f). We expect that increasing the number of particles on different micrographs from different complexes to train the network will further expand the capabilities of crYOLO.

Discussion

Here, we present crYOLO, a novel automated particle-picking procedure for cryo-EM. crYOLO employs a state-of-the-art deep learning object detection system which contains a 16-layer convolutional neural network. The excellent performance of crYOLO on several state-of-the-art direct detector datasets, reflects the effectiveness of the program and its efficiency to detect good particles at an accuracy comparable to manual particle selection. This underlines its potential to become a crucial component in a streamlined automated workflow for single particle analysis and thus eliminates one of the few remaining bottlenecks.

To close the other remaining gaps in the workflow between the electron microscope and data processing, such as evaluation, drift and CTF correction, file conversion and transfer, our lab has developed an on-the-fly processing pipeline called TranSPHIRE, which is documented and freely available on www.sphire.mpg.de (manuscript in preparation). The TranSPHIRE pipeline includes crYOLO.

The crYOLO package is available as a standalone program and will also soon be integrated into the SPHIRE image processing workflow. In addition to the command line interface, crYOLO provides an easy-to-use graphical tool (Figure 2), that makes it easier for users to generate training data and evaluate the results. CrYOLO outputs the coordinate files in EMAN .box-file format which can be easily imported to all available software packages for single particle analysis. CrYOLO picks micrographs rapidly on a standard GPU. If a trained network is available, particle picking can also be performed

on multi-core CPUs. However, this decreases the speed of selection to few seconds per micrograph.

Our evaluation demonstrated that the accuracy of crYOLO yields a set of particles with a tremendously low number of false positives. The IOU values of the selected particles show the excellent centering performance of crYOLO. Both strengths improve the quality of the input data set for image processing and ultimately improve the final reconstruction in many cases. At the same time it dramatically reduces the number of subsequent processing steps, such as computationally intensive 3-D classifications, and thereby shortens the overall processing time.

In comparison to template-based particle picking approaches, crYOLO is not prone to template bias and therefore the danger of picking "Einstein from noise"[22] is significantly reduced. The presented datasets have obtained optimal results after training the network on 5 to 10 micrographs. However, the particles in the training data set must be representative to those of the full dataset.

Importantly, crYOLO can be trained to select previously unseen datasets. With only 34,813 manually picked particles from 331 micrographs derived from 11 cryo-EM datasets, we have already obtained a powerful generalized network. Adding more training datasets from different projects manually picked by experienced users will improve the performance of crYOLO even further. CrYOLO is especially suited for this procedure, because in contrast to the sliding-window approaches, training YOLO only requires particles to be labeled and not the background or contamination. Therefore, previously processed datasets can be directly used to train crYOLO without manually picking background and contamination afterwards. In principle, it is also feasible to train generalized networks, that are specialized on certain types of particles, such as viruses, elongated or very small particles to improve the performance of crYOLO. At the moment, the performance of convolutional neural networks is moving fast and new CNN networks which outperform earlier networks are released frequently. The flexible software architecture of crYOLO allows an easy incorporation of new CNN networks and therefore a straightforward adaption to new developments if necessary.

Taken together, crYOLO enables rapid automated on-the-fly particle selection at the precision of manual picking without bias from reference templates. Furthermore, increasing the size of the training database will result in a generalized particle picker

that can be used without a template or human intervention to select particles on most single particle cryo-EM projects within few minutes.

Acknowledgments

We would like to thank Amir Apelbaum, Evelyn Schubert, and Philine Hagel for providing datasets for optimization purposes and Tanvir R. Tapu Shaikh for proofreading and fruitful discussions. We thank Yifan Cheng and Radostin Danev for kindly providing us with the original coordinates of the NOMPC and Prx3 particles, respectively. This work was supported by the Max Planck Society (S.R.) and the European Council under the European Union's Seventh Framework Programme (FP7/2007–2013) (Grant 615984 to S.R.).

Author contributions

S.R. designed the project, T.W. conceived the approach and wrote the software. T.W., C.G., M.S., T.M, optimized the approach, T.W., C.G., F.M. evaluated the performance, T.W. and S.R. wrote the manuscript with input from all authors. All authors discussed the results and the manuscript.

Data availability

The training datasets for this study are available from the corresponding author upon reasonable request. CrYOLO is available for download under http://sphire.mpg.de

Online Methods

CrYOLO architecture and training

CrYOLO trains a deep convolutional neural network (CNN) for automated particle selection. A typical CNN consists of multiple convolutional and pooling layers, and is characterized by the depth of the network, which is the number of convolutional layers[27]. The input of a CNN is data with a grid-like topology, most often images which may have multiple channels. A convolutional layer consists of a fixed number of filters, and each filter consists of multiple convolutional kernels with a fixed size (typical 1x1, 3x3 or 5x5). The number of convolutional kernels per filter is equal to the number of

channels in the input data of the layer. During forward propagation, each filter creates a feature map which is a channel in the final output of the layer. Each kernel is convolved with the corresponding input channel. The results are summed up along the channels, resulting in the first channel of the output. The same procedure is applied for the second filter which results in the second channel and so on. After all channels are calculated, the so-called batch normalization[28] is applied. Batch normalization normalizes the feature map and leads to faster training. Furthermore it has some regularization power which makes further regularization often unnecessary. Finally, a non-linear activation function is applied element-wise to the normalized feature map which results in the final feature map.

In typical CNNs max-pooling layers are inserted between some of the convolutional layers. Max-pooling layers divide the input image into equal sized tiles (e.g. 2x2) which are then used to calculate a condensed feature map. Therefore for each tile a cell is created, the maximum for the tile is computed and inserted into the cell. This leads to a reduced dimensionality of the feature map, and makes the network more memory-efficient and robust against small perturbations in the pixel values.

The general crYOLO network is summarized in Table 1. The feature extraction part consists of 16 convolutional layers and each convolutional layer consists either of multiple 3x3 filters or 1x1 filters. All convolutional layers use padding so that they do not reduce the size of their input. We used the leaky rectified linear unit[29] (LRELU) as activation function. LRELU simply returns the element itself if the element is positive and the element multiplied by a fixed constant α (here α=0.1) if it is negative.

Four max-pooling layers downsample the image by a factor of 16. During training, a dropout unit in the detection part sets 20 % of the entries in the feature map after layer 16 to zero. This regularizes the network and reduces overfitting. Finally, a convolution layer with six 1x1 filters does the actual detection. With the YOLO[15] approach, each cell in the final feature map is used (i) to classify if the center of a particle box is inside this grid cell and (ii) if that is the case, it predicts the exact position of the box center relative to the cell as well as the width and height of the box.

The network was trained using backpropagation with the ADAM optimizer[30]. Backpropagation applies the chain-rule to compute the gradient values in every layer. The gradient determines how the kernel elements in each convolutional layer should be updated to get a lower loss. The optimizer determines how the gradient of the loss is

used to update the network parameters. For YOLO, the loss function that is minimized in that way is given by:

$$L = \lambda_{coord} \sum_{i=0}^{S^2} \mathbb{1}_i^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2$$

$$+\lambda_{coord} \sum_{i=0}^{S^2} \mathbb{1}_i^{obj} \left[\left(\sqrt{w_i} - \sqrt{\widehat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right]$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{obj} (C_i - \hat{C}_i)^2$$

$$+\lambda_{noobj} \sum_{i=0}^{S^2} \mathbb{1}_i^{noobj} (C_i - \hat{C}_i)^2$$

where $\mathbb{1}_i^{obj}$ is 1 if the center of a particle is in cell *i* and 0 otherwise, λ_{coord} and λ_{noobj} constant weights, (x_i, y_i) the centrum coordinates of the training boxes and $(\hat{x}_i, \hat{y}_i)$ the predicted coordinates. The width and height of the box with their estimates are given by (w_i, h_i) and $(\widehat{w}_i, \hat{h}_i)$ respectively. The confidence that a cell contains a particle is C_i. The first term of the loss function penalizes bad localization of particle boxes. The second term penalizes inaccurate estimates for the width and height of the boxes. The third term tries to increase the confidence for cells with a particle inside. The last term decreases the confidence of those cells containing no particle center. The loss function slightly differs from the one used in Redmon et al.[11] since we have only a single class to predict and also only one reference box (anchor box in Redmon et al.[11]).

Data Augmentation

During training each image is augmented before passing it through the network. This means that it is slightly altered by random selection methods instead of passing the original image through the network. As each image is passed multiple times through the network, it is randomly modified in different ways. This helps the network to reduce overfitting and also the amount of training data needed. The applied methods are:

- Gaussian blurring: A random standard deviation between 0 and 3 is selected and then a corresponding filter mask is created. This mask is then convolved with the input image.

- Average blurring: A random mask size between 2 and 7 is chosen. This mask is shifted over the image. At each position the central element is replaced with the mean values of its neighbors.
- Flip: The image is mirrored along the horizontal axis.
- Noise: Gaussian noise with randomly selected standard deviation is added to the image.
- Dropout: Randomly replaces 1 to 10% of the image pixels with the image mean value.
- Contrast normalization: The contrast is changed by subtracting the median pixel value from each pixel, multiply them by a random constant and finally add the median value again.

References

1. Ogura, T. & Sato, C. An Automatic Particle Pickup Method Using a Neural Network Applicable to Low-Contrast Electron Micrographs. *Journal of Structural Biology* **136,** 227–238 (2001).
2. Nicholson, W. V. & Glaeser, R. M. Review: Automatic Particle Detection in Electron Microscopy. *Journal of Structural Biology* **133,** 90–101 (2001).
3. Huang, Z. & Penczek, P. A. Application of template matching technique to particle detection in electron micrographs. *Journal of Structural Biology* **145,** 29–40 (2004).
4. Volkmann, N. An approach to automated particle picking from electron micrographs based on reduced representation templates. 145, 152–156 (2004).
5. Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology* **189,** 114–122 (2015).
6. Roseman, A. M. *FindEM—a fast, efficient program* **for** *automatic selection* **of** *particles* **from** *electron micrographs*. *Journal of Structural Biology* **145,** 91–99 (2004).
7. Wang, F. *et al.* DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *Journal of Structural Biology* **195,** 325–336 (2016).
8. Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics* **18,** 29 (2017).
9. Rawat, W. & Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation* **29,** 2352–2449 (2017).
10. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks* **61,** 85–117 (2015).

11. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. in 779–788 (IEEE). doi:10.1109/CVPR.2016.91
12. Anh, H. N., Suppa, M., Sirotenko, M. & Schoen, A. J. experiencor/basic-yolo-keras. (2018). doi:10.5281/zenodo.1168755
13. Chollet, F. C. O.others. *Keras*. (GitHub, 2015).
14. EMAN2: An extensible image processing suite for electron microscopy. **157,** 38–46 (2007).
15. Redmon, J. & Farhadi, A. YOLO9000: Better, Faster, Stronger. in 6517–6525 (IEEE). doi:10.1109/CVPR.2017.690
16. Moriya, T. *et al.* High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *JoVE* e55448–e55448 (2017). doi:10.3791/55448
17. Jin, P. *et al.* Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature* **547,** 118–122 (2017).
18. Khoshouei, M. *et al.* Volta phase plate cryo-EM of the small protein complex Prx3. *Nat Comms* **7,** 10534 (2016).
19. A syringe-like injection mechanism in Photorhabdus luminescens toxins. **495,** 520–523 (2013).
20. Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat Comms* **8,** 16099 (2017).
21. Goutte, C. & Gaussier, E. in *Advances in Information Retrieval* **3408,** 345–359 (Springer Berlin Heidelberg, 2005).
22. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative Stable Alignment and Clustering of 2D Transmission Electron Microscope Images. *Structure* **20,** 237–247 (2012).
23. RELION: implementation of a Bayesian approach to cryo-EM structure determination. **180,** 519–530 (2012).
24. 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,** e01963 (2015).
25. Allegretti, M. *et al.* Horizontal membrane-intrinsic α-helices in the stator a-subunit of an F-type ATP synthase. *Nature* **521,** 237–240 (2015).
26. Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Meth* **14,** 793–796 (2017).
27. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86,** 2278–2324 (1998).
28. Ioffe, S. & Szegedy, C. Batch Normalization - Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ICML* (2015).
29. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. in **30,** (2013).
30. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **abs/1412.6980,** (2014).
31. Vinayagam, D. *et al.* Electron cryo-microscopy structure of the canonical TRPC4 ion channel. *eLife* **7,** 213 (2018).

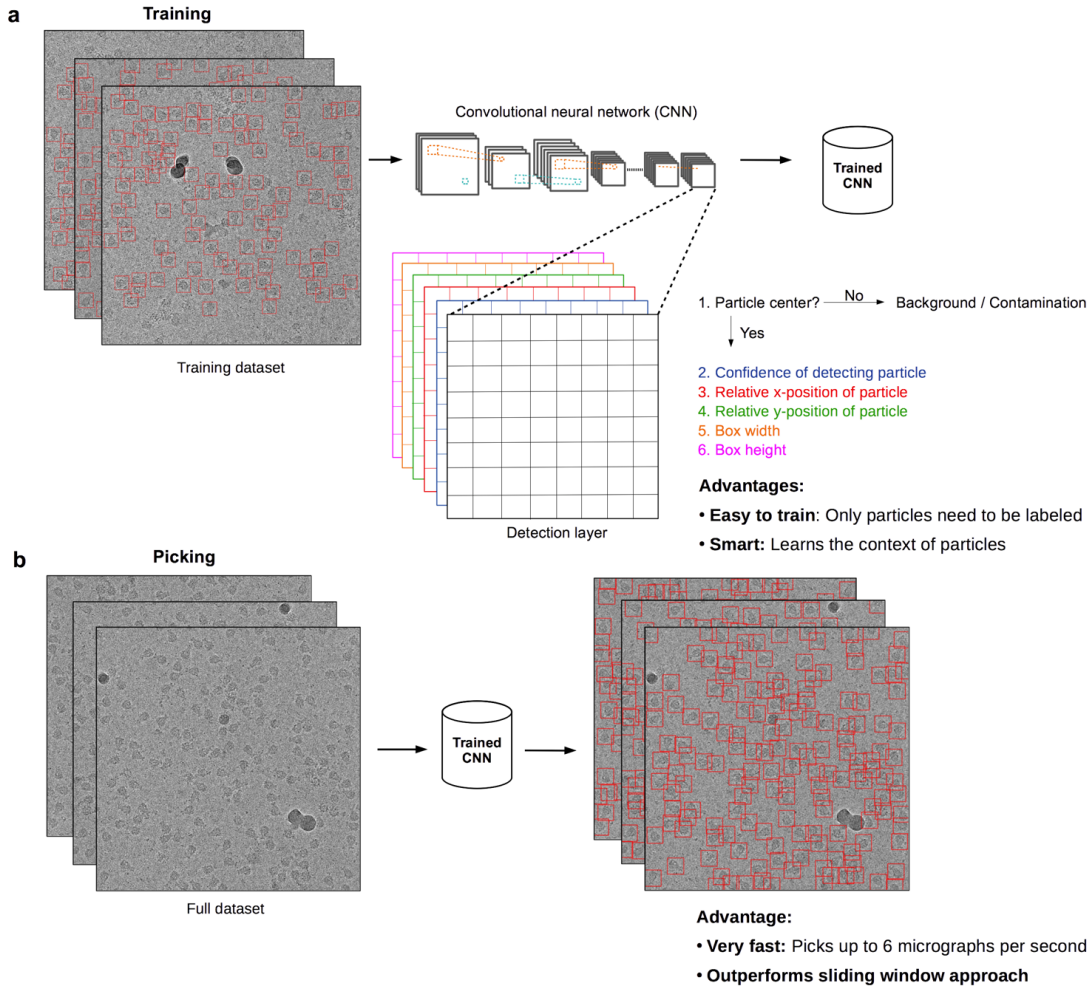


Figure 1. Training and picking in crYOLO. a) The convolutional neural network of crYOLO is trained with a small dataset. For this purpose, only particles have to be labelled and not background or contamination. During training and later picking, the image is spatially downsampled by the CNN to a small grid. Then crYOLO predicts for each grid cell if it contains the center of a particle. If this is the case, it estimates the relative position of the particle center inside the grid cell as well as the width and height of the particle box. If not the grid cell is classified as background. Since the network "sees" the complete micrograph it learns the context of the particle. **b)** CrYOLO uses the trained CNN to select particles from the full dataset. Since every micrograph is only processed once by the network, this procedure is very fast and outperforms the sliding window approach; crYOLO selects up to six micrographs per second.

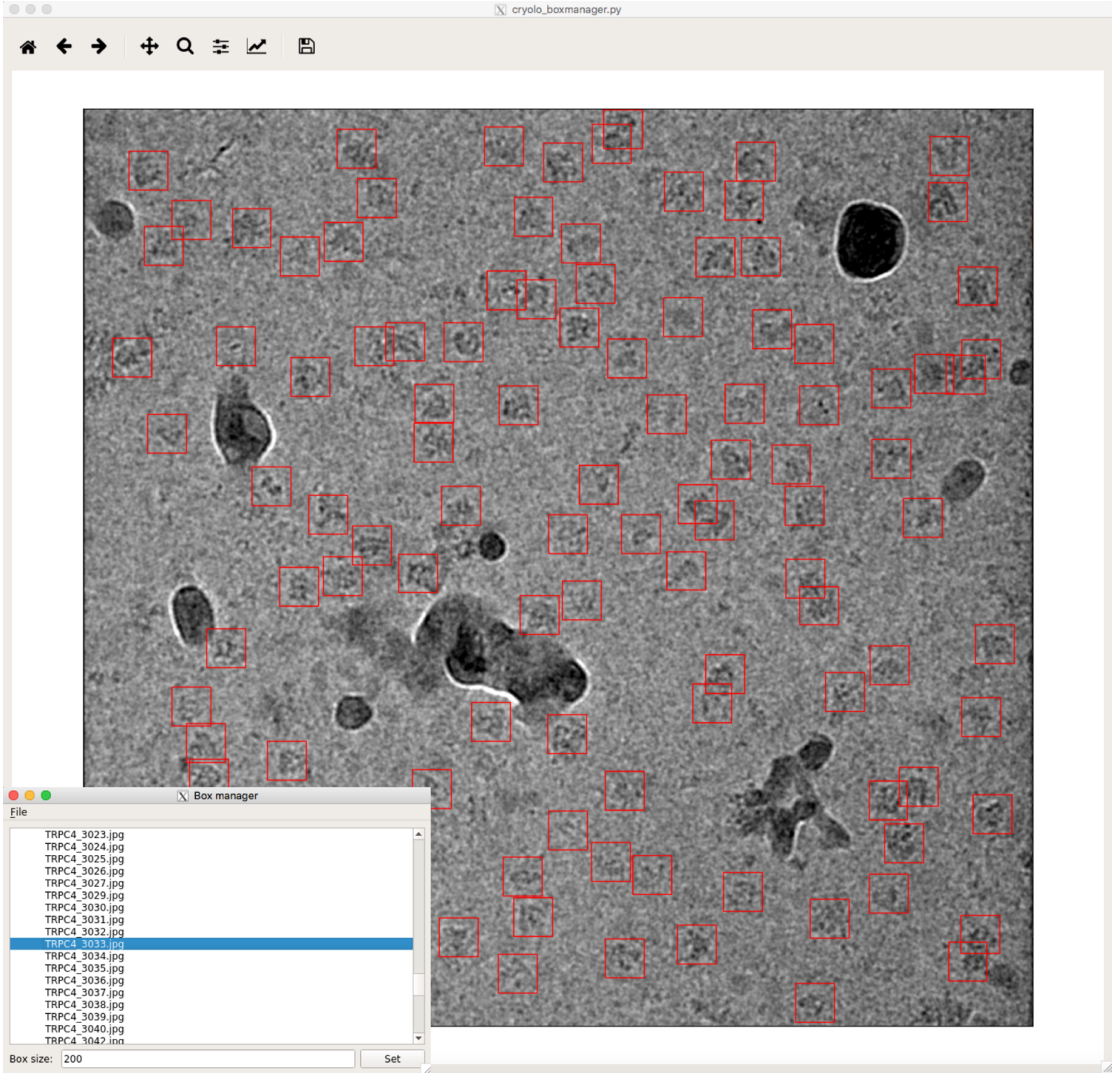


Figure 2. Graphical tool for creating training data and visualizing results. The tool can read images in MRC, TIFF and JPG format and box files in EMAN1 format. The example shown is a micrograph of TRPC4[31] with many contaminants.

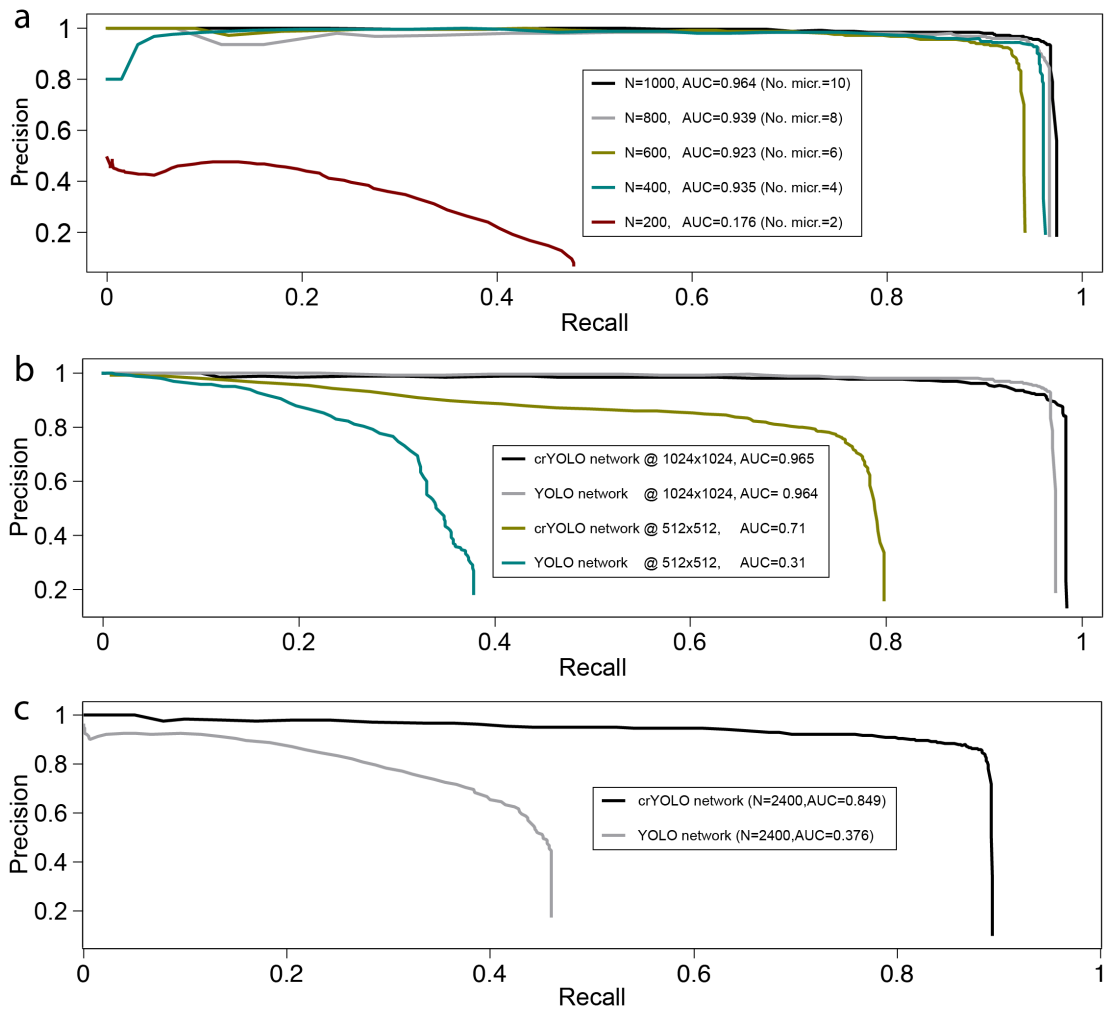
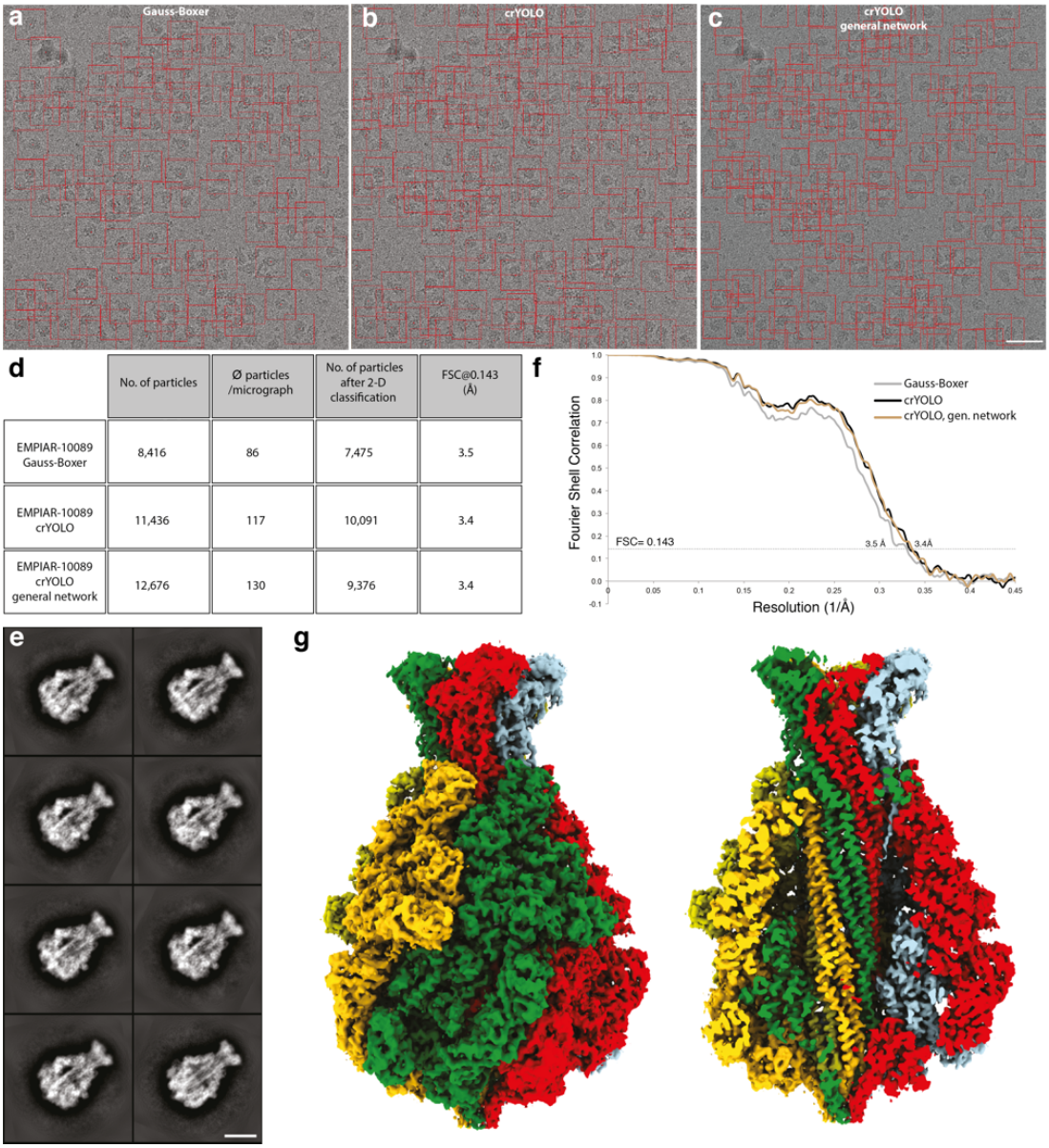


Figure 3. Training of the crYOLO network. a) Precision-recall curves for the TcdA1 dataset using several training set sizes. The curves were estimated based on four randomly selected test micrographs out of the full dataset (size 1024x1024, one micrograph for each defocus class). The AUC values are 0.964 (black), 0.939 (grey), 0.923 (yellow), 0.935 (cyan), 0.176 (red). **b)** The crYOLO network (black) trained on 10 micrographs shows a similar performance on 1024x1024 micrographs in terms of AUC but with slightly lower precision and higher recall as the YOLO network (grey). On low resolution images of size 512x512 the crYOLO network (yellow) achieves an AUC = 0.71 and the YOLO network (cyan) an AUC of 0.31. **c)** Precision-recall curves for the Prx3 dataset using crYOLO network (black curve) and YOLO network (grey curve). Both networks were trained on the same dataset. The AUC was 0.846 and 0.376 respectively.



	No. of particles	Ø particles /micrograph	No. of particles after 2-D classification	FSC@0.143 (Å)
EMPIAR-10089 Gauss-Boxer	8,416	86	7,475	3.5
EMPIAR-10089 crYOLO	11,436	117	10,091	3.4
EMPIAR-10089 crYOLO general network	12,676	130	9,376	3.4

Figure 4. Selection of TcdA1 particles and structural analysis. a-c) Representative digital micrograph (micrograph number 0169) taken from the EMPIAR-10089 dataset. Red boxes indicate the particles selected by Gauss-Boxer (**a**), crYOLO (**b**) or the generalized crYOLO network (**c**), respectively. Scale bar, 50 nm. **d**) Summary of particle selection and structural analysis of the three datasets. All datasets were processed using the same workflow in SPHIRE. **e)** Representative reference-free 2-D class averages of TcdA1 obtained using the ISAC and Beautifier tools (SPHIRE) from particles picked using crYOLO. Scale bar, 10 nm. **f**) Fourier shell correlation (FSC) curves of the 3-D reconstructions calculated from the particles selected in crYOLO and Gauss-Boxer. The FSC 0.143 between the independently refined and masked half-maps indicates resolutions of ~3.4 and ~3.5 Å, respectively. **g)** The final density map of TcdA1 obtained from particles picked by crYOLO is shown from the top and side, and is colored by subunit. The reconstruction using particles from the generalized crYOLO network is indistinguishable.

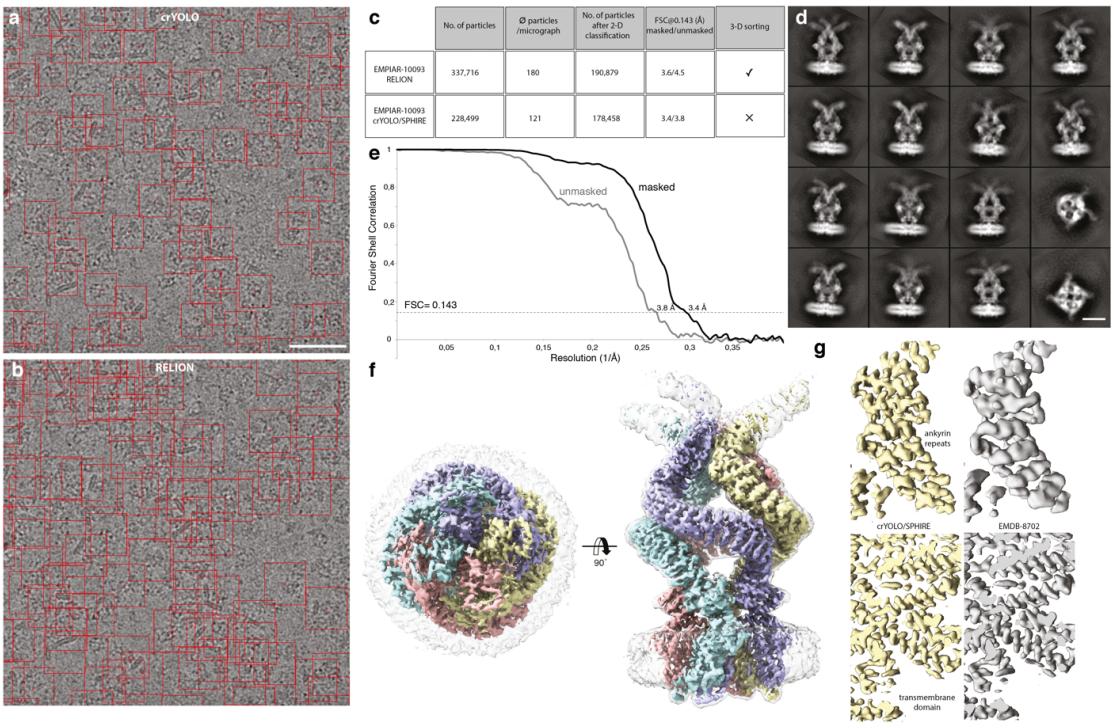
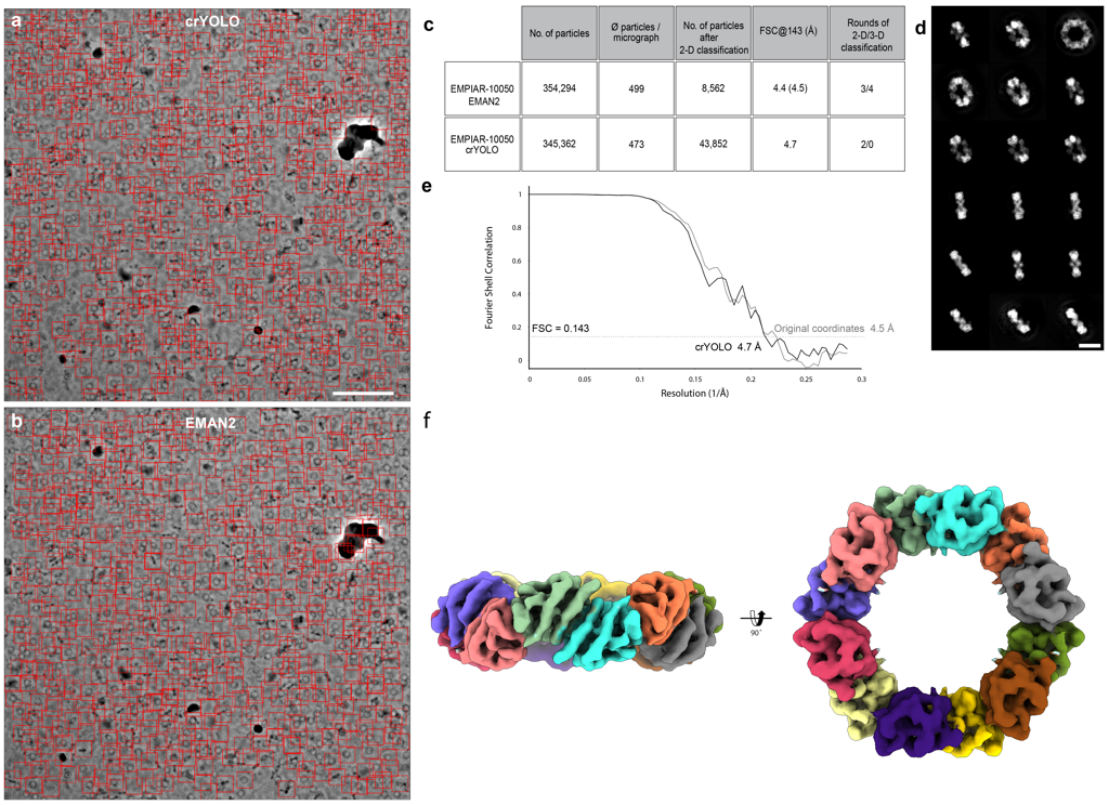


Figure 5. Selection of NOMPC particles and structural analysis. a, b) Representative micrograph (micrograph number 1,854) of the EMPIAR-10093 dataset. Particles picked by crYOLO **(a)** or RELION **(b)** are highlighted by red boxes, respectively. Scale bar, 50 nm. **c)** Summary of particle selection and structural analysis using RELION and crYOLO/SPHIRE. **d)** Representative reference-free 2-D class averages obtained using the ISAC and Beautifier tools (SPHIRE) from particles selected by crYOLO. Scale bar, 10 nm. **e)** FSC curves **(f)** and final 3-D reconstruction of the NOMPC dataset obtained from particles picked using crYOLO and processed with SPHIRE. The 0.143 FSC between the masked and unmasked half-maps indicates resolutions of 3.4 and 3.8 Å, respectively. The 3-D reconstruction is shown from the top and side. To allow better visualization of the nanodisc density the unsharpened (gray, transparent) and sharpened map (colored by subunits) are overlaid. **g)** Comparison of the density map obtained by crYOLO/SPHIRE with the deposited NOMPC 3-D reconstruction EMDB-8702.



	No. of particles	Ø particles / micrograph	No. of particles after 2-D classification	FSC@143 (Å)	Rounds of 2-D/3-D classification
EMPIAR-10050 EMAN2	354,294	499	8,562	4.4 (4.5)	3/4
EMPIAR-10050 crYOLO	345,362	473	43,852	4.7	2/0

Figure 6. Selection of Prx3 particles and structural analysis. a-b) Particles selected on a representative micrograph (micrograph number 19.22.14) of the EMPIAR-10050 dataset using either crYOLO **(a)** or EMAN2 **(b)**. Scale bar, 100 nm. **c)** Summary of particle selection and structural analysis. The resolution in parentheses is the result obtained after a 3-D refinement performed in SPHIRE, using the final 8,562 particles of the original dataset. **d)** Representative 2-D class averages obtained from 2 rounds of classification using the crYOLO-selected particles and ISAC. Scale bar, 10 nm. Well-centered examples for all views showing high-resolution details can be readily obtained from the data. **e)** Fourier shell correlation plots for the final 3-D reconstruction (black) using the crYOLO-selected particles or the 8,562 particles from the original dataset (gray). The average resolution of our 3-D reconstruction is ~ 4.7 Å while that one from the originally used particles is ~ 4.5 Å. **f)** Top and side views of the 3-D reconstruction obtained with crYOLO/SPHIRE. For guidance, all subunits are colored differently in the volume.

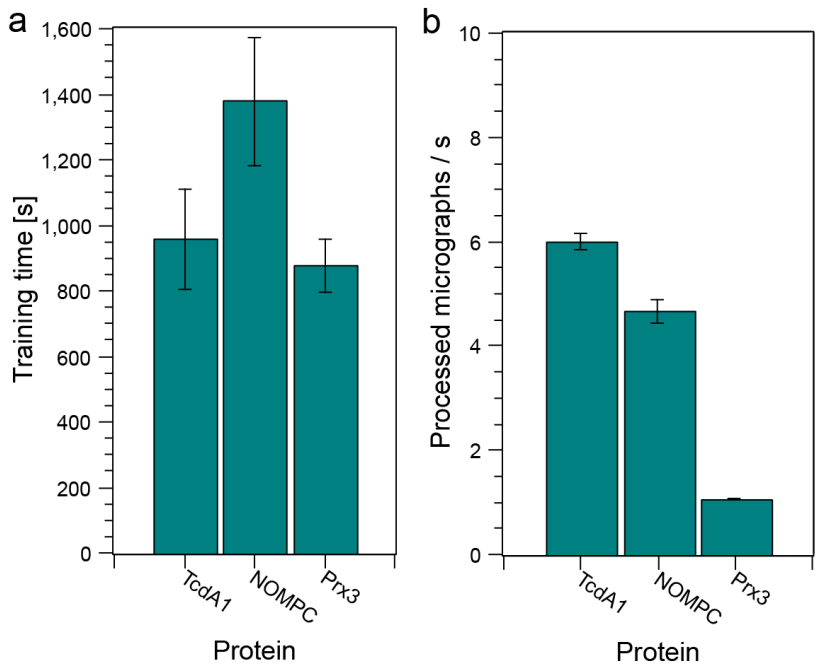


Figure 7. Computational efficiency statistics of training and particle selection. a) The crYOLO training time for TcdA1, NOMPC and Prx3 was 956 s, 1,378 s and 876 s, respectively. **b)** Number of processed micrographs per second during particle selection. All datasets were picked in less than a second per image. For TcdA1, crYOLO needed 0.16 s per micrograph, for NOMPC 0.21 s and 0.9 s for Prx3.

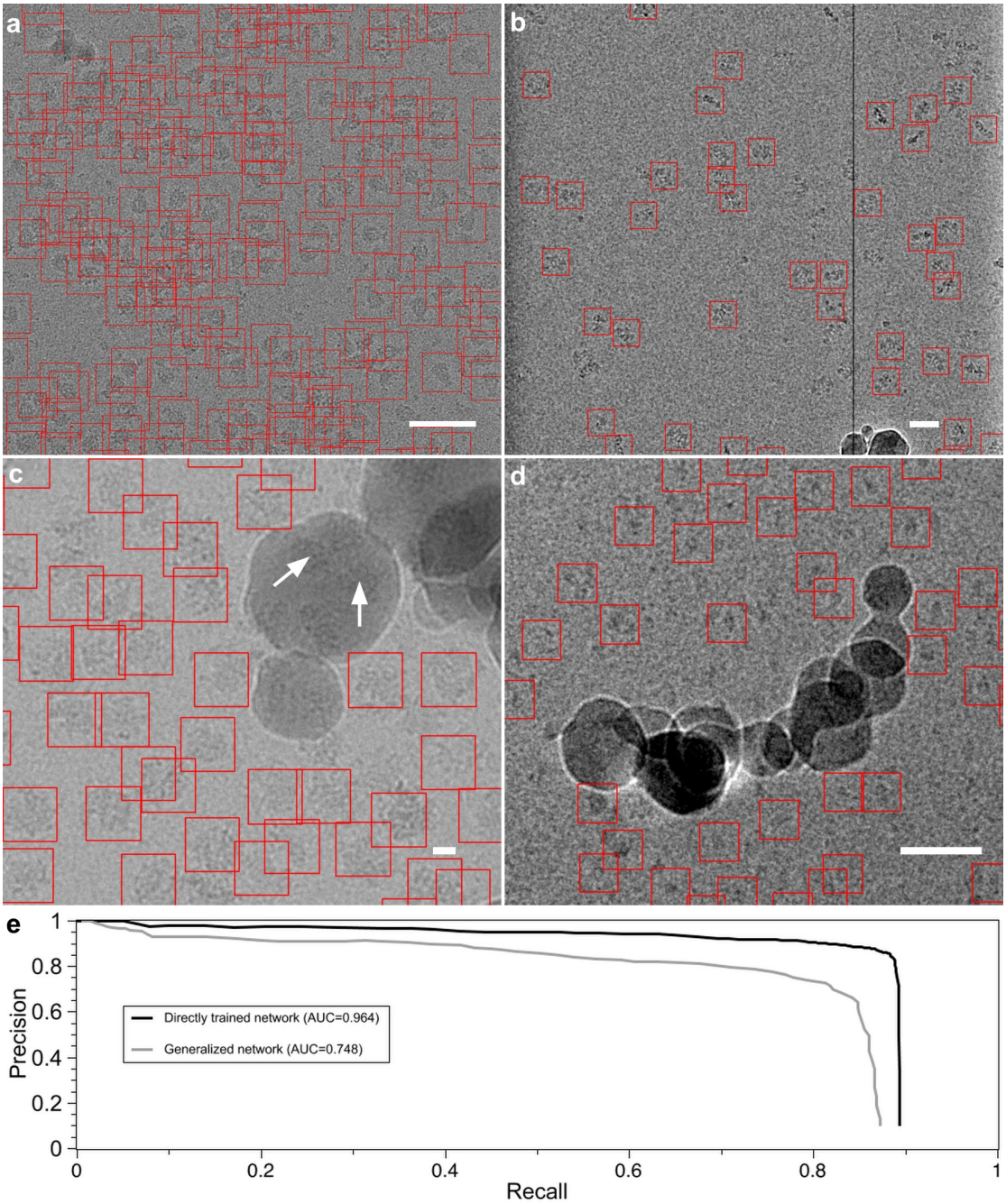


Figure 8. Generalized crYOLO network. **a-d)** Particles selected on a representative micrograph of TcdA1 (EMPIAR 10089) **(a)**, ATP synthase (EMPIAR 10023) **(b)**, ribosome (in-house dataset) **(c)** and influenza hemagglutinin (EMPIAR 10097) **(d)** by crYOLO using the generalized network. None of the datasets was included in the set used for training the generalized crYOLO network. As the YOLO approach enables to learn the context of an object, crYOLO did not select particles enclosed in or overlaid by ice blobs (**c**, white arrows). Scale bars, 40 nm **e**) Precision-recall curves for TcdA1 picked with either a network directly trained on the TcdA1 dataset (black) and a generalized model trained on several datasets but not on TcdA1 (grey).