




COMMENT



<https://doi.org/10.1038/s42003-020-0790-y>

OPEN

The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows


Thorsten Wagner ¹ & Stefan Raunser ¹ 

Particle selection is a crucial step when processing electron cryo microscopy data. Several automated particle picking procedures were developed in the past but most struggle with non-ideal data sets. In our recent *Communications Biology* article, we presented crYOLO, a deep learning based particle picking program. It enables fast, automated particle picking at human levels of accuracy with low effort. A general model allows the use of crYOLO for selecting particles in previously unseen data sets without further training. Here we describe how crYOLO has evolved since its initial release. We have introduced filament picking, a new denoising technique, and a new graphical user interface. Moreover, we outline its usage in automated processing pipelines, which is an important advancement on the horizon of the field.

The crYOLO particle picking procedure

A major goal of electron cryo microscopy (cryo-EM) is to obtain high-resolution three-dimensional (3D) reconstructions of proteins and protein complexes to gain novel biological insights. This process involves the selection of thousands to millions of noisy two-dimensional (2D) particle projections, a number that only keeps increasing with recent advances in hardware and software development.

In our recent work in *Communications Biology*¹ we introduced the “crYOLO” particle picking procedure. It is based on a deep neural network and the You Only Look Once (YOLO) object detection framework². This approach enables the automated picking of particles within cryo-EM micrographs with a low signal-to-noise ratio requiring minimal human supervision or intervention. CrYOLO is easy to configure and train on a specific data set. It is fast and can process up to six micrographs per second. As crYOLO sees the complete micrograph, it is able to learn the

¹Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany.
email: stefan.raunser@mpi-dortmund.mpg.de

overall context of the particles. Therefore, the approach enables highly accurate picking, e.g., it does not select particles on the carbon film or specifically picks particles attached to liposomes. In addition, a pretrained, generalized model further allows the selection of particles in previously unseen data sets with high accuracy.

Recent evolution of crYOLO

Since the release of crYOLO we have improved the software by modifying the network architecture, adding new functionalities, and increasing its usability. In particular, we have integrated a new method for denoising micrographs to increase the signal-to-noise ratio for improved particle detection. By default, crYOLO uses a standard low-pass filter for denoising. However, this method requires parameters to be manually set and has its inherent limits. To enable automated denoising, we therefore implemented the recently introduced neural-network-based approach noise2noise³ into a new tool called JANNI, that can be chosen in crYOLO as alternative denoising method. We pre-trained JANNI on movies from various cryo-EM data sets and used it to denoise previously unseen data sets (Fig. 1). JANNI might be helpful especially for data sets with low signal-to-noise ratio.

Another important new functionality of crYOLO is filament picking. Owing to their structure, the picking of filaments poses a challenge and is often not supported by automated particle

picking procedures. Optimally, only single filaments are selected and positions where filaments cross or overlap are omitted. In case of helical specimens, the boxes should be placed along the filament in a distance according to its helical rise to allow the use of helical reconstruction procedures⁴. The new filament picking procedure initially follows the general workflow of crYOLO. In a post-processing step, it uses the picked particles as support points to trace the filaments. The boxes are then placed along the filaments in a distance defined by the user (Fig. 2).

CrYOLO offers now the possibility to improve an existing model, which is of advantage when fine-tuning a general model on a specific data set. In this case, only the last few layers of the network are retrained while previous layers are fixed. This effectively reduces the amount of training data needed to improve a working model. A major advantage of this approach is a substantial speed-up along with reduced GPU memory consumption.

With the evolution of crYOLO, more options have become available to the user, which increases the complexity of the command line interface. Therefore, we most recently added a new graphical user interface, which makes crYOLO more accessible for new or less technically oriented users (Fig. 3).

Impact of crYOLO

CrYOLO has found widespread use, and since its release mid-2018 already >15 structures were solved with the support of

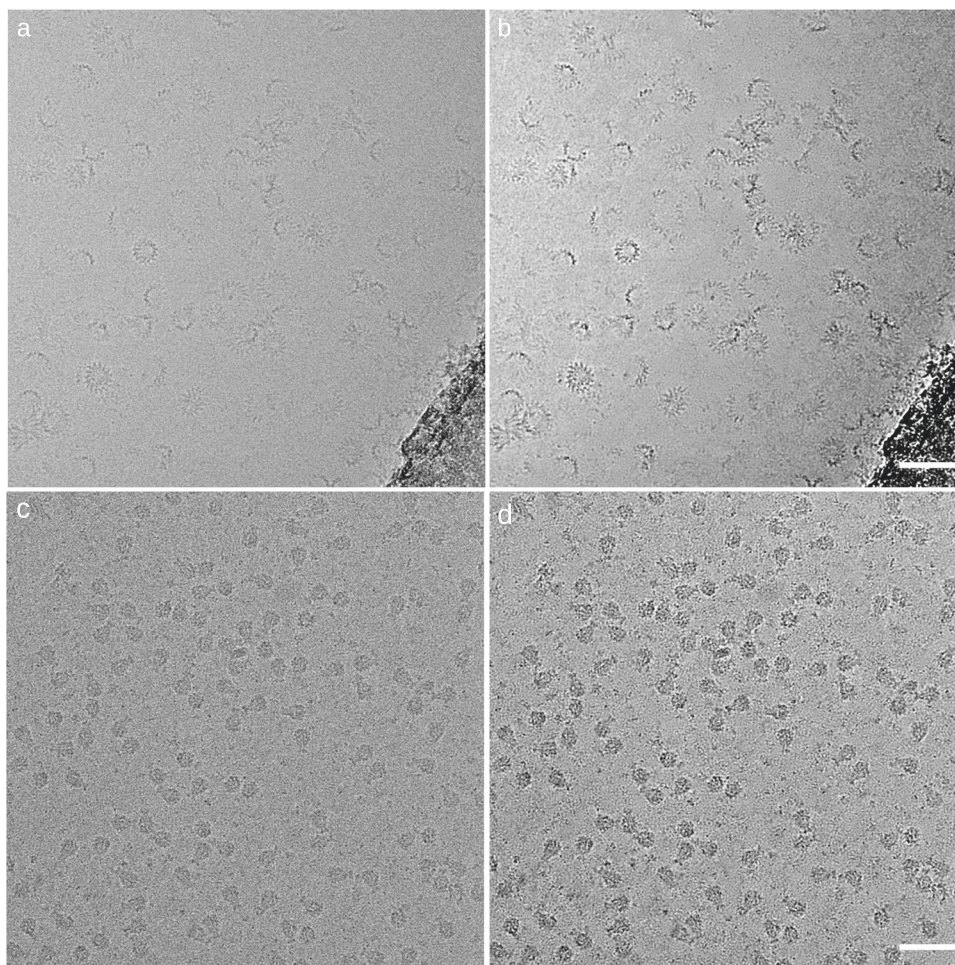


Fig. 1 Example micrographs denoised by JANNI. XaxAB toxin³² and Tc toxin³³ without denoising (**a**, **c**) and with denoising (**b**, **d**), respectively. Details for the sample and grid preparation can be found in ref. ³² for the XaxAB toxin and in ref. ³³ for the Tc toxin. Scale bars: 50 nm.

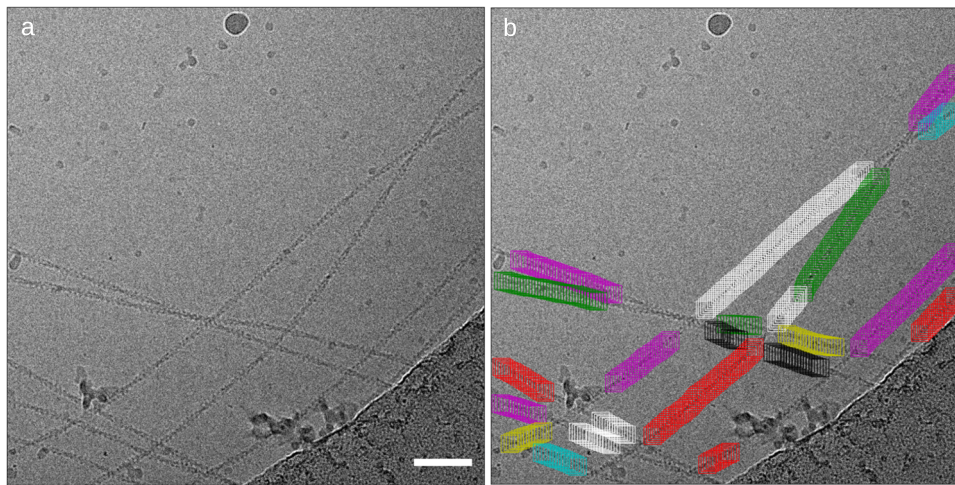


Fig. 2 Actin picked with crYOLO filament mode. **a** Sample micrograph with actin filaments. **b** CrYOLO learned to skip crossings of filaments, contaminations and the carbon edge. Details for the sample and grid preparation can be found in ref. ³⁴. Scale bar: 50 nm.

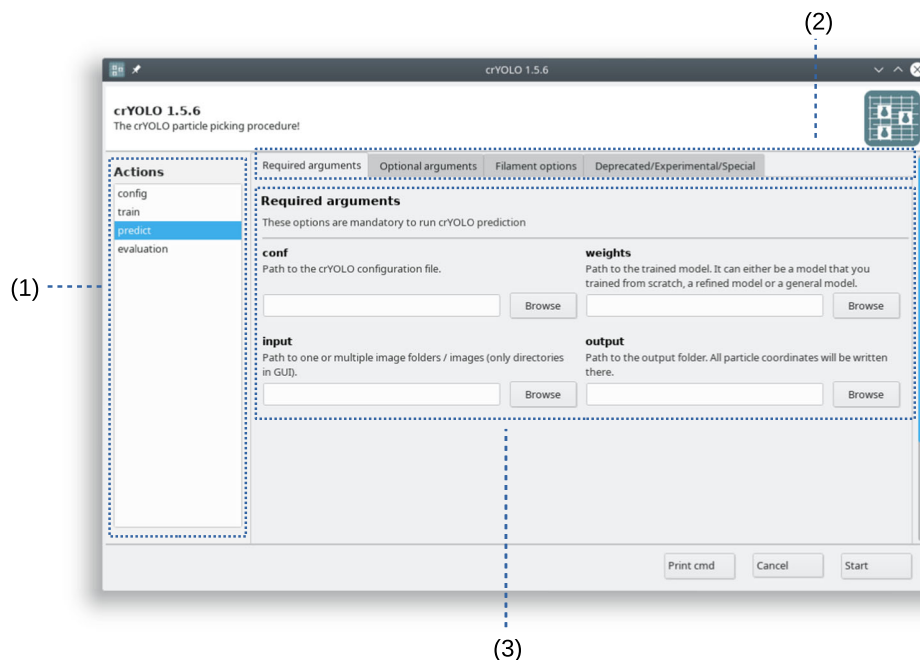


Fig. 3 Graphical user interface (GUI) of crYOLO. The GUI is divided into three parts: (1) List of all available actions like training a model or the prediction of particles on micrographs. (2) Groups of options for the respective action. (3) Available options for the selected action.

crYOLO^{5–22}. For example, Pang et al.²³ used crYOLO to selectively pick particles, which were attached to liposomes; Rogala et al.¹⁴ highlighted in their study about mTORC1 that crYOLO was especially useful to exclude particles on carbon; Joppe et al.²⁴ made use of crYOLO in a streamlined pipeline for rapid structure determination of yeast fatty acid synthase; and the new filament mode was recently used by Pospich et al. to examine the structural effects of toxins on actin filaments¹⁶.

In addition, crYOLO was made available through the SBGrid software collection²⁵, enabling easy access to crYOLO for groups without advanced computational facilities. CrYOLO was also integrated into COSMIC²⁶, a web platform for cryo-EM data processing via cloud computing. Very recently, Li et al.²⁷ used crYOLO in a user-free preprocessing pipeline. This shows that crYOLO has been broadly used by other groups and proven flexible enough for a wide variety of applications.

The general model and automated processing

Since CrYOLO provides a generalized model, it is the optimal particle selection software to be integrated in an automated cryo-EM single-particle analysis procedure. The general model of crYOLO was pretrained on >60 different data sets, including proteins of various sizes and shapes. This allows to pick previously unseen particles not included in the training data set. To this end, crYOLO is a crucial part in our software package SPHIRE that we are optimizing to be used in a completely automated fashion²⁸. In Scipion, crYOLO is supported for the construction of intelligent workflows²⁹. A recent integration of crYOLO into the automatic pipeline of Relion³⁰ is successfully used at the Electron Bio-Imaging Center (eBIC) at Diamond Light Source³¹.

Whereas the generalized model offers great opportunities for automated processing there remain limitations. The amount of

data used for the general model is still limited and might be biased towards the set of proteins used for the initial training. The general model is also not able to distinguish between intact and dissociated or fragmented particles in the same sample. This requires additional training to fine-tune the general model with particles manually picked from a few micrographs. A drawback is that this requires manual intervention and is therefore not suitable for automated processing. A better strategy is to automatically fine-tune the general model based on 2D classification, where particles representing similar views are grouped together, aligned and averaged.

During 2D classification, broken particles will be separated from intact particles. The latter ones will then be used to train a crYOLO model or fine-tune the general model.

Optimally, a fully automated pipeline would also include a deep-learning-based 2D class selection tool. Our group is currently developing such software, that we call Cinderella. While it is still under development, it is already publicly available and successfully integrated in SPHIRE²⁸. Cinderella provides a pre-trained general model and is able to separate 2D classes into good and bad classes.

In the future, a combination of Cinderella and crYOLO will allow automated feedback loops to improve the picking quality in an iterative manner. With these tools at hand, we believe that real-time automated processing even for challenging data sets is in reach.

Data availability

All data supporting the findings of this study are available from the corresponding author on reasonable request.

Received: 17 January 2020; Accepted: 24 January 2020;

Published online: 11 February 2020

References

- Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
- Redmon, J. & Farhadi, A. YOLO9000: Better, Faster, Stronger. in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* 6517–6525 (IEEE). <https://doi.org/10.1109/CVPR.2017.690> (2017).
- Lehtinen, J. et al. Noise2Noise: Learning image restoration without clean data. *ArXiv180304189 Cs Stat* at <http://arxiv.org/abs/1803.04189> (2018).
- Egelman, E. H. Reconstruction of helical filaments and tubes. *Methods Enzymol.* **482**, 167–183 (2010).
- Park, E. et al. Architecture of autoinhibited and active BRAF-MEK1-14-3-3 complexes. *Nature* **575**, 545–550 (2019).
- Zhou, B. -R. et al. Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. *Nat. Commun.* **10**, 2301 (2019).
- Gaullier, G. et al. Bridging of nucleosome-proximal DNA double-strand breaks by PARP2 enhances its interaction with HPF1. *Biochemistry* <https://doi.org/10.1101/846618> (2019).
- Leidreiter, F. et al. Common architecture of Tc toxins from human and insect pathogenic bacteria. *Sci. Adv.* **5**, eaax6497 (2019).
- Gatsogiannis, C., Balogh, D., Merino, F., Sieber, S. A. & Raunser, S. Cryo-EM structure of the ClpXP protein degradation machinery. *Nat. Struct. Mol. Biol.* **26**, 946–954 (2019).
- Shen, K. et al. Cryo-EM structure of the human FLCN-FNIP2-Rag-activator complex. *Cell* **179**, 1319–1329.e8 (2019).
- Madej, M. et al. Dynamic oligopeptide acquisition by the RagAB transporter from *Porphyromonas gingivalis*. *Microbiology* <https://doi.org/10.1101/755678> (2019).
- Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).
- Consolati, T. et al. Microtubule nucleation by single human γ TuRC in a partly open asymmetric conformation. *Biochemistry* <https://doi.org/10.1101/853218> (2019).
- Rogala, K. B. et al. Structural basis for the docking of mTORC1 on the lysosomal surface. *Science* **366**, 468–475 (2019).

- Nithianantham, S. et al. Structural basis of tubulin recruitment and assembly by microtubule polymerases with tumor overexpressed gene (TOG) domain arrays. *eLife* **7**, e38922 (2018).
- Pospich, S., Merino, F. & Raunser, S. Structural effects and functional implications of phalloidin and jasplakinolide binding to actin filaments. *Biochemistry* <https://doi.org/10.1101/794495> (2019).
- Klink, B. U., Gatsogiannis, C., Hofnagel, O., Wittinghofer, A. & Raunser, S. Structure of the human BBSome core complex in the open conformation. *Mol. Biol.* <https://doi.org/10.1101/845982> (2019).
- Singh, S., Gui, M., Koh, F., Yip, M. C. J. & Brown, A. Structure and activation mechanism of the BBSome membrane-protein trafficking complex. *Biophysics*. <https://doi.org/10.1101/849877> (2019).
- Roderer, D., Schubert, E., Sitsel, O. & Raunser, S. Towards the application of Tc toxins as a universal protein translocation system. *Nat. Commun.* **10**, 5263 (2019).
- Lill, P. et al. Towards the molecular architecture of the peroxisomal receptor docking complex. *Mol. Biol.* <https://doi.org/10.1101/854497> (2019).
- Shi, X. et al. ULK complex organization in autophagy by a C-shaped FIP200 N-terminal domain dimer. *Biochemistry*. <https://doi.org/10.1101/840009> (2019).
- Shah, P. N. M. et al. Cryo-EM structures reveal two distinct conformational states in a picornavirus cell entry intermediate. (bioRxiv). <https://doi.org/10.1101/2020.01.08.899112> (2020).
- Pang, S. S. et al. The cryo-EM structure of the acid activatable pore-forming immune effector Macrophage-expressed gene 1. *Nat. Commun.* **10**, 4288 (2019).
- Joppe, M. et al. The resolution revolution in cryoEM requires new sample preparation procedures: a rapid pipeline to high resolution maps of yeast FAS. *Biochemistry*. <https://doi.org/10.1101/829176> (2019).
- Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
- Cianfrocco, M. A., Wong-Barnum, M., Youn, C., Wagner, R. & Leschziner, A. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. in *Proc. Pract. Exp. Adv. Res. Comput. 2017 Sustain. Success Impact* 22:1–22:5 (ACM). <https://doi.org/10.1145/3093338.3093390> (2017).
- Li, Y., Cash, J. N., Tesmer, John. J. G. & Cianfrocco, M. A. High-throughput cryo-EM enabled by user-free preprocessing routines. (bioRxiv). <https://doi.org/10.1101/2019.12.20.885541> (2019).
- Moriya, T. et al. High-resolution Single Particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.* <https://doi.org/10.3791/55448> (2017).
- Maluenda, D. et al. Flexible workflows for on-the-fly electron-microscopy single-particle image processing using Scipion. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 882–894 (2019).
- Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Webb, D. relion-yolo-it. *Relion-Yolo-It 0301* at <https://pypi.org/project/relion-yolo-it/>
- Schubert, E., Vetter, I. R., Prumbaum, D., Penczek, P. A. & Raunser, S. Membrane insertion of α -xenorhabdolyisin in near-atomic detail. *eLife* **7**, e38017 (2018).
- Gatsogiannis, C. et al. Membrane insertion of a Tc toxin in near-atomic detail. *Nat. Struct. Mol. Biol.* **23**, 884–890 (2016).
- Merino, F. et al. Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM. *Nat. Struct. Mol. Biol.* **25**, 528–537 (2018).

Acknowledgements

This work was supported by the Max Planck Society (S.R.).

Author contributions

S.R. and T.W. designed the project, T.W. conceived the approach and wrote the software. T.W. and S.R. wrote the manuscript

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020