TomoPicker: Annotation-Efficient Particle Picking in Cellular cryo-electron Tomograms

Anonymous Author(s)
Affiliation
Address
email

Abstract

Particle Picking in cellular cryo-electron tomograms (cryo-ET) is crucial for *in situ* structure detection of macromolecules and protein complexes. Given the problems associated with the traditional template-matching approaches for particle picking, learning-based solutions are necessary for particle picking. A big challenge in this regard is the lack of annotated data for training. In this work, we present TomoPicker, a Positive-Unlabeled learning-based annotation-efficient particle-picking approach that can effectively pick particles when only a minuscule portion ($\sim 0.3 - 0.5\%$) of the total particles in a cellular cryo-ET dataset are provided for training. We evaluated our method on a benchmark cryo-ET dataset of eukaryote cells, where we observed about 30% improvement by TomoPicker against the most recent state-of-the-art annotation efficient learning-based picking approaches.

1 Introduction

Cryo-electron tomography (Cryo-ET) is an emerging imaging technology that has enabled *in-situ* 3D visualization of macromolecular structures in nanometer and even subnanometer resolution inside the cells [1, 2]. Without hampering the cellular specimens, it can visualize the subcellular macromolecules inside them in their native contexts. Thanks to such unique characteristics, cryo-ET has been used extensively for *in situ* structural biology. It can recover the structures of macromolecules and protein complexes inside cells with different phenotypes and reveal their spatial organization, further facilitating the discovery of numerous novel biological insights [3].

Nevertheless, extracting the structure of macromolecules from 3D cellular cryo-ET tomograms is a complex process that involves multiple steps [4, 2]. The first and most important step is locating the macromolecules in the tomograms, called "particle picking" [5, 6]. However, particle picking is a challenging task due to several reasons. Firstly, cryo-ET tomograms are large 3D volumes, of size $\approx 1000 \times 1000 \times 500$ even after $4x$ binning [7, 8]. Secondly, these tomograms are very noisy with a low signal-to-noise ratio and contrast due to the complex cytoplasmic environment and low electron dosage [8, 4]. Finally, the concentration of macromolecules per image is very high, which can be around $500 - 1000$ per tomogram, making it even more difficult to locate them accurately.

Given the abovementioned challenges, manually picking particles in the tomograms is extremely time-consuming and burdensome. To this end, automated approaches for particle picking have been developed [5, 9, 10]. A common approach is template matching (TM), which uses templates from existing data sources as references to localize similar macromolecules in the tomograms [11]. However, TM can only be applied when a reference template is available for the macromolecules to be picked and often contains reference-dependent biases [7, 12]. In addition, TM is extremely time-consuming [7] and shows suboptimal performance [13]. To solve this issue, neural network-based deep learning approaches have been introduced [14, 10, 9]. These approaches provide high-throughput fast localization of particles without having any reference-dependent biases. However, most of these

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

approaches [10, 9] are based on supervised learning, which again requires manual annotation of many particles in the tomograms for training purposes. Given the difficulty of manual annotation in cryo-ET, annotation-efficient methods that can perform reliable annotations without requiring large annotated training data are necessary.

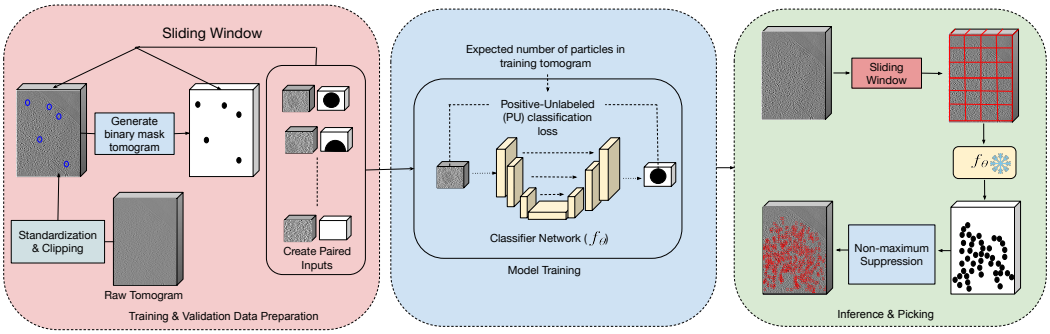

Figure 1: TomoPicker Pipeline

In recent years, a few learning-based picking approaches have addressed this annotation burden [15, 16]. Huang et al. [15] developed an algorithm to detect proteins from sparse labels by regarding particle picking as a regression problem. They consider each 3D tomogram as a single sample and predict particle coordinates for it directly at the tomogram level. This approach has two problems. First, since their method is a learning-based method and regards each tomogram as a sample, a large number of similar tomograms are required in the training set. Second, to regard each tomogram as a sample and fit them as input to convolutional networks, they significantly downsample the tomograms, increasing the crowding of particles inside the tomogram. Several tomograms, particularly from eukaryote cells, are already very crowded. In such scenarios, downsampling makes particle picking even more difficult. Another weakly-supervised algorithm has been developed very recently, named as DeepETPicker [16]. Unlike [15], DeepETPicker can be trained on a single tomogram where several particle coordinates in that tomogram are annotated. Despite achieving success for sparse single-particle and prokaryotic tomograms, its efficacy in crowded eukaryotic tomograms has not yet been explored. Moreover, this method did not adapt any mechanism tailored to deal with the annotation-efficiency issue.

In this work, we developed a novel annotation-efficient particle-picking approach called TomoPicker. Our approach only requires a small portion ($\sim 0.3 - 0.5\%$) of all particles in a tomogram dataset) of the particle's center coordinates to be annotated beforehand. Similar to [15], we regard particle picking as a voxel classification problem. For 3D cryo-ET tomograms, each voxel is classified as a binary value based on whether it contains particles or not. However, unlike [15], our method can be trained on a single tomogram since we do not treat the entire tomogram as samples and rather use subvolumes extracted from tomograms as samples for voxel classification. Given only a few portions of the voxels are labeled positive, a specific approach is necessary to deal with the large unlabeled voxels. If all unlabeled voxels are regarded as negative, it would lead to erroneous prediction and picking. To solve this problem, we introduced two positive unlabeled (PU) learning approaches.

We evaluated our methods against two well-annotated benchmark datasets of eukaryotic *S. Pombe* cell tomograms. We also evaluated the recent and popular learning-based cryo-ET picking methods (including the state-of-the-art DeepETPicker [16]) on these datasets for the first item. Our experimental results demonstrate the superior performance of the TomoPicker approach. Our proposed KL divergence-based and non-negative risk estimator-based TomoPicker method improves the particle picking performance by 30% over the state-of-the-art DeepETPicker method against the VPP and Defocus-only dataset, respectively. Thus, TomoPicker shows high efficacy even when 0.4% of the total number of annotated particles in the datasets are used for training.

2 Method

TomoPicker consists of three main components for annotation-efficient particle picking in cryo-ET tomograms (Figure 1). We briefly discuss them as follows:

2.1 Preprocessing and Data Generation for Training

We start by preprocessing the tomograms to enhance contrast, loading them as voxelized arrays. We standardize each tomogram and clip voxel values that lie beyond three standard deviations from the mean. After clipping, we re-standardize the voxels.

To generate labels for voxel classification in our particle-picking network, we create empty voxel arrays matching the shape of each training and validation tomogram. For each particle with provided coordinates (representing only a small percentage of the actual particles), we assign values of 1 to all voxels within a radius around the particle, creating spherical masks in the corresponding label arrays.

Next, we use a sliding window approach to generate subtomograms and submasks from both the tomograms and their label arrays. We save these subtomogram-submask pairs for training, while for validation, we only retain pairs with non-zero submasks. Once this data is saved, we proceed to the training phase.

2.2 Model Training with PU Leaning

We formulate the particle picking as a voxel classification problem. We assume that P is the set of labeled particle regions and U is the set of unlabeled particle and non-particle regions in the training dataset. Given P and U, we learn a classifier (f_θ) that distinguishes between particle and non-particle regions in the subtomograms. We used three different strategies (two with PU learning and one without) to train the classifier, which we discuss below.

2.2.1 Positive Negative (PN) Learning

In PN learning, we treat P as positive samples and U as negative samples, assuming most regions in U are non-particle regions. We train the classifier using the binary cross-entropy loss with the objective: $\pi E_{x\sim P}[L(f_\theta(x), 1)] + (1 - \pi)E_{x\sim U}[L(f_\theta(x), 0)]$, where π is the fraction of particle regions in the dataset and L is the binary cross-entropy loss. While this works well when all particles are labeled in P, it performs poorly (Table 1 and 2) in practice because most particles reside in U.

2.2.2 Non-negative Risk Estimator based Positive Unlabeled (PU) Leaning

To better handle the unlabeled regions, we adapt a non-negative risk estimator-based PU learning approach [17] for cryo-ET particle picking. Here, π' is the expected average voxel value in a labeled sample. It is calculated given the expected number of particles in the training tomogram, which the user can readily provide. We define the risk estimators as: $\widehat{R}_U^- = E_{x\sim U}[L(f_\theta(x), 0)]$, $\widehat{R}_P^- = E_{x\sim P}[L(f_\theta(x), 0)]$, and $\widehat{R}_P^+ = E_{x\sim P}[L(f_\theta(x), 1)]$. The PU risk estimator is then: $\widehat{R}_{PU} = \pi'(\widehat{R}_P^+ - \widehat{R}_P^-) + \widehat{R}_U^-$. We update f_θ using $\nabla\widehat{R}_{PU}$ if $\widehat{R}_U^- - \pi'\widehat{R}_P^- \geq 0$, otherwise we update using $\nabla(\pi'\widehat{R}_P^- - R_U^-)$.

2.2.3 KL based Positive Unlabeled (PU) Leaning

We propose an alternative to Non-negative Risk Estimator PU learning by minimizing the P class misclassification loss while matching the expectation over U. The classifier f_θ minimizes the term $E_{x\sim P}[L(f_\theta(x), 1)]$, subject to the constraint $E_{x\sim U}[f_\theta(x)] = \pi''$, where π'' represents the fraction of unlabeled particle regions within U. This constraint is incorporated into the objective function with a regularization term weighted by λ, resulting in the objective: $E_{x\sim P}[L(f_\theta(x), 1)] + \lambda\, KL(E_{x\sim U}[f_\theta(x)] \,||\, \pi'')$. The Kullback-Leibler divergence (KL) ensures that the expectation of the classifier over U aligns with the estimated fraction of unlabeled particles π'', and the divergence is minimized when both terms are close to each other. π'' is calculated as $\pi' - \pi$, where π and π' are defined in Section 2.2.1 and 2.2.2 respectively.

2.3 Inference and Picking

After training the classifier with the above-mentioned learning strategies, we perform particle picking on all the tomograms in the dataset, including the ones we used for training and validation. For each tomogram V, we use a sliding window strategy to obtain non-overlapping subvolumes of the same

size as the training subtomograms. Then, we perform inference for each subvolume with our learned classifier f_θ. The inference results in a score for each voxel in the subvolumes. We merge the score outputs for each subvolume in the tomogram to a volumetric array (V_{score}) with the same size as the tomogram. We then apply the picking process on this merged predicted array V_{score}. The process takes the required number of particles N or subvolume score threshold t and the particle radius r as input. It operates in 4 steps. In step 1, Find the point $(x_{\text{max}}, y_{\text{max}}, z_{\text{max}})$ with maximum score value in V_{score}. In step 2, we append $(x_{\text{max}}, y_{\text{max}}, z_{\text{max}})$ as well as the score $V_{\text{score}}(x_{\text{max}}, y_{\text{max}}, z_{\text{max}})$ to the extracted particle list. In step 3, we remove a roughly spherical region of particle radius r around $(x_{\text{max}}, y_{\text{max}}, z_{\text{max}})$ in V_{score} by setting their scores to $-\infty$. This ensures that the same particle will not be extracted more than once. Finally, we repeat steps $1-3$ until N particles are extracted or no prediction scores above the threshold t remain.

3 Experiments & Results

Baselines: We used CrYOLO [9] and DeepETPicker [16], the two most recent and publicly accessible learning-based cryo-ET picking methods as baselines. Since CrYOLO [9] is actually a bounding box predictor method for 2D cryo-EM images, it is necessary to convert the 3D tomograms into 2D slices and provide 2D annotations for each slice to train CrYOLO. In the Appendix, we describe the detailed process of 2D slice generation, crYOLO model training, prediction, and 2D-to-3D matching for evaluation. For DeepETPicker [16], we used their publicly available codebase with their default setting for picking ribosomes.

Evaluation: For evaluation, we calculated the number of True Positives (TP), False Positives (FP), False Negatives (FN), Precision, Recall, and F1-score predicted by the baseline models, and our proposed models. We use the annotations of the ribosome coordinates provided in the original dataset as ground truth. For any ground truth coordinate, if there is any predicted coordinate within 10 voxels of euclidean distance, it is regarded as a TP. The predicted coordinates that are not within 10 voxels of distance to any ground truth particle are regarded as FP. On the other hand, those ground truth coordinates, where there are no predicted coordinates within 10 voxels of euclidean distance are regarded as FNs. Precision and Recall are calculated from TP, FP, and FN as $\frac{\text{TP}}{\text{TP+FP}}$ and $\frac{\text{TP}}{\text{TP+FN}}$. Finally, F1 score is calculated as $\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$. **Experimental setup:** In our experiments for TomoPicker, we used 3D-ResUnet as the classifier network f_θ, similar to DeepETPicker [16]. We used a batch size of 8 and an initial learning rate of 2×10^{-3}, which has been reduced by a factor 0.5 if validation accuracy does not improve for 5 consecutive epochs. We trained TomoPicker and CrYOLO for 20 epochs, which we found to be sufficient. However, we trained DeepETPicker for 100 epochs. We implemented our method in pytorch and trained the models using NVIDIA RTX A5000 GPUs.

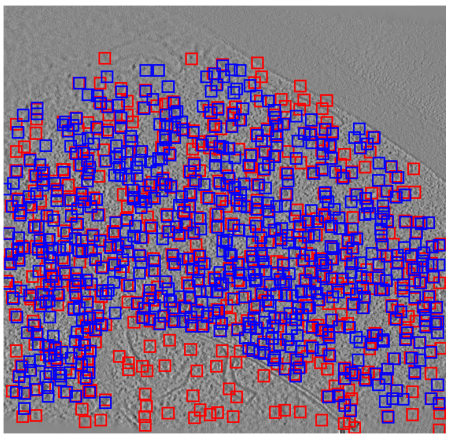

Figure 2: DeepETPicker vs Ground Truth for T0001. Blue Box = Ground Truth, Red Box = DeepETPicker Predictions.

For experiments, we use the well-annotated and very recent cellular cyo-ET datasets of eukaryotic *S. pombe* cells publicly available at EMPIAR-10988 [18]. It contains 10 volt-phase-plate (VPP) and 10 defocus-only tomograms (voxel spacing 1.348 nm) of *S. pombe* cell sections. We choose these two datasets - 1) VPP and 2) Defocus-only tomogram set for benchmarking.

Volta-Phase-Plate (VPP) *S. Pombe* cellular cryo-ET Dataset: The VPP dataset contains 10 tomograms (labeled from T0001 to T0010 consecutively) with a total of $25,311$ ribosome particles. The individual tomograms from T0001 to T0010 contains 2450, 2342, 2429, 2967, 3571, 1336, 617, 2744, 3482, and 3373 ribosome particles respectively. For training our models, we only used 100 particle coordinates from T0001 for training and 100 particle coordinates from T0002 for validation. This accounts for only $\frac{100}{25,311} = 0.4\%$ of the total particles. Since the voxel spacing is 1.348 nm and the radius of ribosome particle is $28-30$ nm, we use $\frac{28}{1.348} \approx 11$ voxels as the particle radius.

Table 1: F1 scorecomparison across different methods on VPP *S. Pombe* cellular cryo-ET datasets.

Method	**Dataset**										
	T0001	T0002	T0003	T0004	T0005	T0006	T0007	T0008	T0009	T0010	**Overall**
CrYOLO	0.36	0.27	0.28	0.28	0.24	0.13	0.07	0.18	0.25	0.38	0.25
DeepETPicker	0.71	0.23	0.27	0.05	0.52	0.40	0.45	0.13	0.37	0.43	0.35
TomoPicker (PN)	0.23	0.03	0.02	0.02	0.27	0.17	0.12	0.15	0.26	0.24	0.15
TomoPicker (PU)	0.57	0.47	0.38	0.38	0.50	0.25	0.27	0.24	0.50	0.44	*0.40*
TomoPicker (KL)	0.67	0.55	0.52	0.19	0.62	0.61	0.31	0.17	0.25	0.58	**0.45**

Defocus only *S. Pombe* cellular cryo-ET Dataset: The Defocus-only dataset contains 10 tomograms (labeled as T026, T027, T028, T029, T030, T034, T037, T041, T043, T045) with a total of $25,901$ ribosome particles. The individual tomograms on the above-mentioned sequence contains 838, 1673, 5305, 2897, 2783, 3783, 1646, 2813, 1815, and 2348 ribosome particles respectively. Among them, T026 has a very different organization compared to other tomograms, and visually looks much different. As a result, we did not use this tomogram for training, validation, or testing. Similar to VPP datasets, we only used 100 particle coordinates from T029 for training and 100 particle coordinates from T030 for validation. This accounts for only $\frac{100}{25,063} = 0.4\%$ of the total particles.

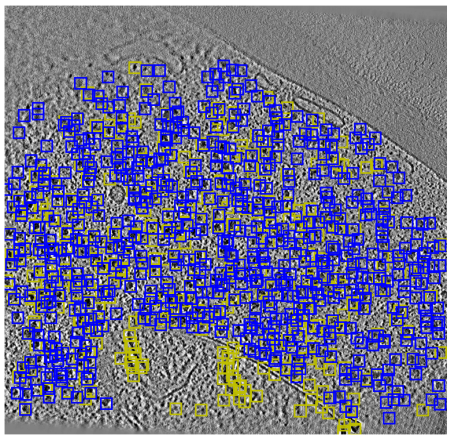

Figure 3: TomoPicker vs Ground Truth for T0001. Blue Box = Ground Truth, Yellow Box = TomoPicker Predictions.

After training, the model was tested against all the tomograms (except T026 in Defocus-only dataset). We have put the F1 score (up to 2 decimal places) obtained by each method against each tomogram and overall dataset for the VPP dataset in Table 1 and for the Defocus-only dataset in Table 2. The tables show that both of our proposed strategies for TomoPicker outperform the baseline methods. TomoPicker with KL-based PU learning outperforms state-of-the-art DeepETPicker by 29% for VPP and 17% for the Defocus-only dataset. TomoPicker, with non-negative risk estimator-based PU learning, outperforms DeepETPicker by 15% for VPP and 29% for the Defocus-only dataset.

Moreover, we provided two qualitative results of DeepETPicker prediction and TomoPicker prediction compared to ground truth in Figure 2 and Figure 3, respectively. The figure demonstrated TomoPicker's superior picking.

Table 2: F1 score comparison across different methods on Defocus-only *S. Pombe* cryo-ET datasets.

Method	**Dataset**									
	T027	T028	T029	T030	T034	T037	T041	T043	T045	**Overall**
CrYOLO	0.16	0.19	0.21	0.14	0.25	0.15	0.29	0.08	0.27	0.19
DeepETPicker	0.37	0.31	0.44	0.43	0.42	0.34	0.29	0.33	0.31	0.35
TomoPicker (PN)	0.15	0.26	0.23	0.22	0.21	0.13	0.18	0.08	0.15	0.18
TomoPicker (PU)	0.48	0.50	0.55	0.52	0.52	0.40	0.40	0.36	0.36	**0.45**
TomoPicker (KL)	0.35	0.56	0.60	0.49	0.48	0.43	0.27	0.12	0.35	*0.41*

4 Discussion & Conclusion

In this work, we have introduced a novel annotation-efficient particle-picking approach, TomoPicker, for 3D cellular cryo-ET images or tomograms. We proposed two different positive-unlabeled (PU) learning strategies to train TomoPicker. We trained and evaluated these approaches as well as recent methods on *S. Pombe* cell tomograms. We are the first to evaluate learning-based picking methods on crowded eukaryotic cell tomograms. Our exhaustive experiments demonstrate the superior ($\sim 30\%$ improvement of F1 score) performance of TomoPicker over strong baseline methods when only a minuscule portion of the particles in the tomograms are annotated for training.

References

[1] Allison Doerr. Cryo-electron tomography. *Nature Methods*, 14(1):34–34, 2017.

[2] Martin Turk and Wolfgang Baumeister. The promise and the challenges of cryo-electron tomography. *FEBS letters*, 594(20):3243–3261, 2020.

[3] Carol V Robinson, Andrej Sali, and Wolfgang Baumeister. The molecular sociology of the cell. *Nature*, 450(7172):973–982, 2007.

[4] Muyuan Chen, James M Bell, Xiaodong Shi, Stella Y Sun, Zhao Wang, and Steven J Ludtke. A complete data processing workflow for cryo-et and subtomogram averaging. *Nature methods*, 16(11):1161–1168, 2019.

[5] Guang Tang, Liwei Peng, Philip R Baldwin, Deepinder S Mann, Wen Jiang, Ian Rees, and Steven J Ludtke. Eman2: an extensible image processing suite for electron microscopy. *Journal of structural biology*, 157(1):38–46, 2007.

[6] Hannah Hyun-Sook Kim, Mostofa Rafid Uddin, Min Xu, and Yi-Wei Chang. Computational methods toward unbiased pattern mining and structure determination in cryo-electron tomography data. *Journal of Molecular Biology*, page 168068, 2023.

[7] Xiangrui Zeng, Anson Kahng, Liang Xue, Julia Mahamid, Yi-Wei Chang, and Min Xu. Disca: high-throughput cryo-et structural pattern mining by deep unsupervised clustering. *bioRxiv*, pages 2021–05, 2021.

[8] Hsuan-Fu Liu, Ye Zhou, Qinwen Huang, Jonathan Piland, Weisheng Jin, Justin Mandel, Xiaochen Du, Jeffrey Martin, and Alberto Bartesaghi. nextpyp: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods*, 20(12):1909–1919, 2023.

[9] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, et al. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications biology*, 2(1):218, 2019.

[10] Emmanuel Moebel, Antonio Martinez-Sanchez, Lorenz Lamm, Ricardo D Righetto, Wojciech Wietrzynski, Sahradha Albert, Damien Larivière, Eric Fourmentin, Stefan Pfeffer, Julio Ortiz, et al. Deep learning improves macromolecule identification in 3d cellular cryo-electron tomograms. *Nature methods*, 18(11):1386–1394, 2021.

[11] Jochen Böhm, Achilleas S Frangakis, Reiner Hegerl, Stephan Nickell, Dieter Typke, and Wolfgang Baumeister. Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences*, 97(26):14245–14250, 2000.

[12] Hsuan-Fu Liu, Ye Zhou, Qinwen Huang, Jonathan Piland, Weisheng Jin, Justin Mandel, Xiaochen Du, Jeffrey Martin, and Alberto Bartesaghi. nextpyp: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods*, pages 1–11, 2023.

[13] Guole Liu, Tongxin Niu, Mengxuan Qiu, Yun Zhu, Fei Sun, and Ge Yang. Deepetpicker: Fast and accurate 3d particle picking for cryo-electron tomography using weakly supervised deep learning. *Nature Communications*, 15(1):2090, 2024.

[14] Irene de Teresa-Trueba, Sara K Goetz, Alexander Mattausch, Frosina Stojanovska, Christian E Zimmerli, Mauricio Toro-Nahuelpan, Dorothy WC Cheng, Fergus Tollervey, Constantin Pape, Martin Beck, et al. Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods*, 20(2):284–294, 2023.

[15] Qinwen Huang, Ye Zhou, Hsuan-Fu Liu, and Alberto Bartesaghi. Accurate detection of proteins in cryo-electron tomograms from sparse labels. In *European Conference on Computer Vision*, pages 644–660. Springer, 2022.

[16] Guole Liu, Tongxin Niu, Mengxuan Qiu, Yun Zhu, Fei Sun, and Ge Yang. Deepetpicker: Fast and accurate 3d particle picking for cryo-electron tomography using weakly supervised deep learning. *Nature Communications*, 15(1):2090, 2024.

[17] Ryuichi Kiryo, Gang Niu, Marthinus C Du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. *Advances in neural information processing systems*, 30, 2017.

[18] European Bioinformatics Institute. Empiar-10988: Negative stain electron microscopy of nucleosome bound by engineered minimalist reader mbtd1 binding module, 2023. Accessed: 2024-09-20.

A Appendix

A.1 DeepETPicker Results

1. VPP Ribosome Results: Table 3
2. VPP FAS Results: Table 4
3. Defocus Ribosome Results: Table 5

Tomogram	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_0001	590	353	1,860	0.63	0.24	0.35
TS_0002	423	81	1,919	0.84	0.18	0.30
TS_0003	352	73	2,077	0.83	0.14	0.25
TS_0004	257	43	2,710	0.86	0.09	0.16
TS_0005	824	196	2,747	0.81	0.23	0.36
TS_0006	381	257	955	0.60	0.29	0.39
TS_0007	180	327	437	0.36	0.29	0.32
TS_0008	671	60	2,073	0.92	0.24	0.39
TS_0009	766	91	2,716	0.89	0.22	0.35
TS_0010	604	174	2,769	0.78	0.18	0.29
Overall				0.752	0.21	0.316

Table 3: VPP Ribosome Performance Metrics

Tomogram	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_0001	75	298	26	0.20	0.74	0.32
TS_0002	3	5	78	0.38	0.04	0.07
TS_0003	2	1	57	0.67	0.03	0.06
TS_0004	0	0	112	0	0	0
TS_0005	21	54	37	0.28	0.36	0.32
TS_0006	14	456	41	0.03	0.19	0.05
TS_0007	6	606	34	0.01	0.15	0.02
TS_0008	22	425	53	0.05	0.29	0.08
TS_0009	26	101	38	0.20	0.41	0.27
TS_0010	30	371	36	0.07	0.45	0.13
Overall				0.189	0.266	0.132

Table 4: VPP FAS Performance Metrics

Tomogram	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_026	328	84	510	0.80	0.39	0.52
TS_027	551	83	1,122	0.87	0.33	0.48
TS_028	388	28	4,917	0.93	0.07	0.14
TS_029	992	65	1,905	0.94	0.34	0.50
TS_030	856	51	1,927	0.94	0.31	0.46
TS_034	935	43	2,848	0.96	0.25	0.39
TS_037	211	59	1,435	0.78	0.13	0.22
TS_041	274	69	2,539	0.80	0.10	0.17
TS_043	403	141	1,412	0.74	0.22	0.34
TS_045	160	19	2,188	0.89	0.07	0.13
Overall				0.865	0.221	0.335

Table 5: Defocus Ribosome Performance Metrics

A.2 CrYOLO Results

1. VPP Ribosome Results: Table 6
2. VPP FAS Results: Table 7
3. Defocus Ribosome Results: Table 8

Tomogram	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_0001	1306	3322	1144	0.28	0.53	0.37
TS_0002	555	1093	1787	0.34	0.24	0.28
TS_0003	1377	5691	1052	0.19	0.57	0.29
TS_0004	1677	7016	1290	0.19	0.57	0.29
TS_0005	800	2167	2771	0.27	0.22	0.24
TS_0006	210	1566	1126	0.12	0.16	0.13
TS_0007	66	1133	551	0.06	0.11	0.07
TS_0008	471	1874	2273	0.20	0.17	0.19
TS_0009	1158	4372	2324	0.21	0.33	0.26
TS_0010	1805	4289	1568	0.30	0.54	0.38
Overall				0.22	0.34	0.25

Table 6: Cryolo VPP Ribosome Performance Metrics

Tomogram	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_0001	98	3797	3	0.03	0.97	0.05
TS_0002	43	726	38	0.06	0.53	0.10
TS_0003	17	287	42	0.06	0.29	0.09
TS_0004	25	1389	87	0.02	0.22	0.03
TS_0005	13	753	45	0.02	0.22	0.03
TS_0006	20	1991	55	0.01	0.27	0.02
TS_0007	18	5256	18	0.00	0.45	0.01
TS_0008	20	2494	55	0.01	0.27	0.02
TS_0009	15	1505	49	0.01	0.23	0.02
Overall				0.02	0.38	0.04

Table 7: Cryolo VPP FAS Performance Metrics

Ground File	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_027.coords	1296	16185	377	0.07	0.77	0.14
TS_028.coords	3495	28929	1810	0.11	0.66	0.19
TS_029.coords	1950	14039	947	0.12	0.67	0.21
TS_030.coords	1952	22392	831	0.08	0.70	0.14
TS_034.coords	2499	13838	1284	0.15	0.66	0.25
TS_037.coords	1039	10805	607	0.09	0.63	0.15
TS_041.coords	1637	6884	1176	0.19	0.58	0.29
TS_043.coords	788	16252	1027	0.05	0.43	0.08
TS_045.coords	1294	5778	1054	0.18	0.55	0.27
Overall				0.12	0.63	0.19

Table 8: Cryolo Defocus Ribosome Performance Metrics

A.3 TomoPicker Results

File	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_043.coords	222	1783	1593	0.11	0.12	0.12
TS_028.coords	2947	2213	2358	0.57	0.56	0.56
TS_045.coords	776	1253	1572	0.38	0.33	0.35
TS_030.coords	1439	1617	1344	0.47	0.52	0.49
TS_041.coords	800	2236	2013	0.26	0.28	0.27
TS_037.coords	789	1249	857	0.39	0.48	0.43
TS_034.coords	1871	2221	1912	0.46	0.49	0.48
TS_027.coords	650	1396	1033	0.32	0.39	0.35
TS_029.coords	1789	1278	1108	0.58	0.62	0.60

Table 9: KL Tomopicker Pombe Defocus Ribosome Performance Metrics

File	True Positives	False Positives	False Negatives	Precision	Recall	F1 Score
TS_0008.coords	408	1601	2336	0.20	0.15	0.17
TS_0010.coords	1879	1196	951	0.61	0.56	0.58
TS_0006.coords	873	666	463	0.57	0.65	0.61
TS_0004.coords	484	1536	2483	0.24	0.16	0.19
TS_0001.coords	1499	549	951	0.73	0.61	0.67
TS_0002.coords	1222	858	1120	0.59	0.52	0.55
TS_0009.coords	675	1437	2807	0.33	0.19	0.25
TS_0007.coords	254	758	363	0.25	0.41	0.32
TS_0005.coords	2074	1003	1497	0.67	0.58	0.62
TS_0003.coords	1162	890	1267	0.57	0.48	0.52
Overall	12530	10494	16434	0.45	0.45	0.45

Table 10: KL Tomopicker Pombe VPP Ribosome Performance Metrics