Towards generalizable particle picking in Cryo-EM images by leveraging Masked AutoEncoders

Anonymous Authors[1]

Abstract

Cryo-electron microscopy (cryo-EM) is a pivotal technique for elucidating protein structures, yet particle picking remains a bottleneck due to inherent challenges such as specimen impurities, sample preparation variability, and microscope parameter fluctuations. These factors result in micrographs with diverse noise profiles, pixel characteristics, and particle dimensions, posing significant hurdles for conventional supervised methods that struggle with generalization and necessitate labor-intensive expert annotations. In this work we present a self-supervised method that leverages a Masked AutoEncoder's representation space to sequentially denoise micrographs based on clusters with different noise levels. Evaluation across 14 datasets demonstrates superior generalization capabilities compared to state-of-the-art supervised methods, showcasing consistent performance independent of pre-training data. This underscores self-supervised learning's potential for advancing cryo-EM image analysis and enabling more efficient structural biology research.

1. Introduction

Cryo-electron microscopy (cryo-EM) has *revolutionized the field of structural biology* by enabling the visualization of biological molecules at near-atomic resolution (Bai et al., 2015). In cryo-EM, protein samples are rapidly frozen in a thin layer of vitreous ice, preserving their native structures, and then imaged in an electron microscope, producing two-dimensional projection images called micrographs. Each micrograph contains numerous randomly oriented copies of the molecule of interest, the so-called particles (Milne et al., 2013; Nogales & Scheres, 2015; Doerr, 2016). However, the analysis of cryo-EM data presents *several unique challenges* that arise from the nature of the technique and the biological samples being studied (Bendory et al., 2020; Wu et al., 2021).

One of the most critical steps in cryo-EM data processing is particle picking (Chung et al., 2022), which involves accurately identifying and selecting individual particles from noisy and heterogeneous micrographs. Particle picking is complicated by several factors: **(i)** low signal-to-noise ratio (SNR) images (Agard et al., 2014), resulting in *high noise levels* (Baxter et al., 2009); **(ii)** *heterogeneity in particle appearance* arising from differences in orientation, conformational states, and complex samples imaged by cryo-EM (Skalidis et al., 2022; Elferich et al., 2022); and **(iii)** *discrepancies in appearance and contrast* due to variations in data collection parameters.

We demonstrate that state-of-the-art approaches *struggle to generalize to unseen data* and require costly annotations, rendering them impractical for real-world applications with limited data. In this work, we propose a self-supervised Masked Auto-Encoder (MAE) method that denoises micrographs using clustering on the MAE representation space, while hierarchical clustering filters out micrograph-specific noise in a sequential manner. Extensive evaluation on 14 datasets highlights the great generalization ability of our method, reporting stable performance regardless of pretrain data. In this way, our method is **annotation-free** and **independent of the pretrain data type**.

The main contributions of this paper are summarized as follows:

1. We introduce the first self-supervised method for particle picking in cryo-EM data that also **does not require any form of annotation**.

2. Our method demonstrates **stable generalization** capabilities when applied to unseen data, outperforming supervised methods. Highlighting the effectiveness of our unsupervised approach in handling diverse cryo-EM datasets.

3. We showcase the effectiveness of hierarchical denoising in learning underlying structures and patterns from noisy cryo-EM images.

2. Related work

Cryo-EM particle identification and picking are crucial steps with various proposed approaches, spanning traditional computer vision techniques and advanced machine/deep learning

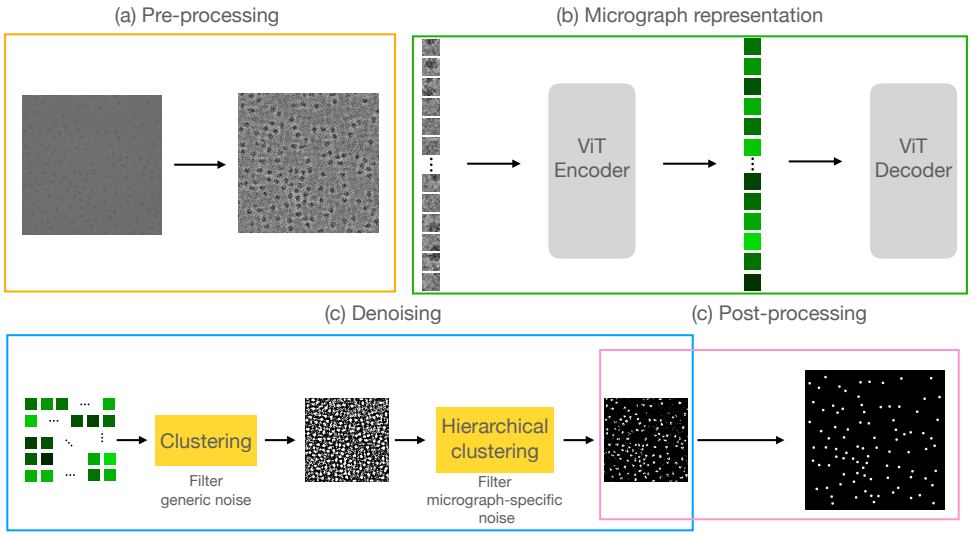

Figure 1. The pipeline starts with an input micrograph and follows these steps: (a) Pre-processing: the micrograph undergoes normalization of background noise and is filtered to enhance particle contrast, (b) Micrograph representation: patches are extracted from the pre-processed micrograph and used to map it onto the MAE representation space, (c) Denoising: the resulting embeddings form an image where a k-means trained on the train set identifies pixels with the lowest noise levels. These images undergo further denoising through micrograph-specific hierarchical clustering, (d) Post-processing: convolution-based smoothing is applied on the predictions of the particle centres for greater accuracy.

pipelines. Traditional template-based methods (Scheres, 2012; 2015) rely on projections of the target particle, introducing potential human bias through template selection. Template-free approaches like Laplacian of Gaussian (LoG) (Woolford et al., 2007b;a), Difference of Gaussians (DoG) (Voss et al., 2009), and Blob picker (Punjani et al., 2017) offer flexibility but often struggle with inherent variability in cryo-EM data, leading to high false-positive rates.

Deep learning methods to date rely on supervised learning, mostly employing convolutional neural networks. They address particle picking from three main perspectives: **(i)** Binary classification of micrograph windows as containing centered particles or not (DeepEM (Zhu et al., 2017), Topaz (Bepler et al., 2019), Warp (Tegunov & Cramer, 2019), DeepCryoPicker (Al-Azzawi et al., 2020)). **(ii)** Segmentation of micrographs into background and particle classes (DeepPicker (Wang et al., 2016), Pixer (Zhang et al., 2019), PARSED (Yao et al., 2020), DRPnet (Nguyen et al., 2021), CryoSegNet (Gyawali et al., 2023)). **(iii)** Object detection problem (crYOLO (Wagner et al., 2019), EPicker (Zhang et al., 2022), CryoTransformer (Dhakal et al., 2024)).

3. Method

3.1. Micrograph preprocessing

Analyzing Cryo-EM micrographs necessitates capturing intricate high-frequency details, that is however hindered by fluctuating levels of high-frequency noise. To enhance such information, a filtering process is vital. We employ a normalization technique (Scheres, 2015) to standardize the background noise to a zero-mean and unit standard deviation noise by adjusting for noise variations according to the particle diameter. This mitigates the influence of experimental variables, enhancing micrograph analysis consistency. Additionally, we apply a Weiner filter for denoising, Contrast Limited Adaptive Histogram Equalization (CLAHE) for contrast enhancement, and guided filtering using the CLAHE-enhanced image as reference (Dhakal et al., 2024; Gyawali et al., 2023). This selectively smooths the image while retaining structural details, balancing noise reduction and information preservation. All images are resized to 1024×1024.

3.2. Representation learning

Self-Supervised methods like Contrastive Learning (Chen et al., 2020), Self-Distillation (Caron et al., 2021) and Canonical Correlation (Zbontar et al., 2021) rely on data augmentations that preserve semantic content. However, common augmentations like Gaussian blur may not maintain the semantics of micrographs, posing a challenge in designing effective augmentations. Conversely, we speculate that learning to reconstruct random local patches of a micrograph can yield representations that capture particle-oriented local invariances, without resorting to augmentations. Given this, we propose using Masked

Autoencoders (MAEs) (He et al., 2022), which have demonstrated the ability to learn representations that encompass a broad spectrum of semantics relevant to downstream tasks, for representation learning on micrographs.

An ideal encoder should be invariant across noise levels. When clustering the representation space, distinct clusters should emerge corresponding to different noise levels and distances from particle centers. We employ k-means with four clusters on the training set, then we identify the particle cluster by reference annotations from just one micrograph. The cluster that has the higher population of particle annotations is considered to be the cluster that represents the lowest noise levels.

3.3. Inference

Given an unseen micrograph, the inference process is performed in two stages **(i)** *denoising based on the learned representation space* and **(ii)** *smoothing and filtering of predictions*.

Various noise levels in micrographs complicate accurate particle prediction. To address this, we adopt a two-stage approach during inference. First, we tackle common high-frequency noise across micrographs by denoising the input using four clusters derived from the training set. Then, to address differing noise levels in micrographs, we train micrograph-specific clusters hierarchically, enabling sequential denoising of different noise levels. Hierarchical clustering is performed on three levels based on 3, 4, and 5 clusters respectively.

The denoised image undergoes post-processing to extract the final predictions. First, convolution with a 3x3 kernel is applied to smooth the predictions, while bilinear interpolation restores the image to an appropriate resolution (512x512) for particle picking. Subsequently, a threshold is applied to the smoothed predictions to generate segmentation masks. Predicted segmentation masks are computed using various thresholds within [0, 1] range. The optimal threshold is determined by ensuring the resulting mask aligns with the statistical properties of the training data, specifically that particles occupy approximately 4% of the micrograph. Finally, further post-processing is performed to filter out **(i)** neighbouring predictions based on particle diameter **(ii)** filter predictions who are at the edges of the micrograph, and **(iii)** filter predictions whose size exceeds by a threshold the expected particle size.

4. Experimental evaluation

In this section, we compare our method with Topaz and crYOLO for particle picking, using a part of the CryoPPP dataset for training and evaluation (Dhakal et al., 2023).

4.1. Experimental set up

We compare our method to Topaz and crYOLO (Bepler et al., 2019; Wagner et al., 2019), the most widely used deep learning tools for particle picking in Single Particle Analysis (SPA) of cryo-EM micrographs, representing two distinct approaches: Topaz as an image classification model and crYOLO as an object detection model. Both utilize convolutional neural networks to learn features from cryo-EM micrographs.

All three methods were trained from scratch using four diverse EMPIAR datasets (10291, 10077, 10590, 10816) provided by CryoPPP (Dhakal et al., 2023), containing proteins of different diameters (160-250Å), types and functions. Evaluation was performed on the test sets of these four datasets and an additional set of 10 EMPIAR datasets (10028, 10081, 10096, 10240, 10406, 10289, 10737, 10059, 11183, 10017) annotated by CryoPPP, with particle diameters ranging from 100-300Å. Each training dataset consists of 300 micrographs, randomly divided into 60% training, 20% validation, 20% testing. The 14 datasets represent wide variance in protein characteristics, noise levels and micrograph concentrations.

Particle picking can be perceived as an object detection problem, therefore we use the following common evaluation metrics in the literature (Wagner et al., 2019; Xu et al., 2024): **(i)** Intersection over Union (IoU) between annotated and predicted particles, using the particle diameter as the box size, **(ii)** recall, precision, and F1 score where true positives (TP) are only counted for predicted particles that uniquely overlap with ground truth particles by more than 60% of their surface area, while all other predicted particles are classified as false positives (FP).

Micrographs used for training all three methods underwent identical preprocessing steps as discussed in Section 3.1 ensuring a fair comparison across the methodologies. For the training of Topaz, we employed the default ResNet8 model with the default training parameters, running for 400 epochs. The epoch yielding the highest area under the precision-recall curve on the validation set was selected. Micrographs were resized according to (Bepler et al., 2019). For the training of the crYOLO, micrographs were resized to the recommended resolution of 1024x1024, with anchor the particle size in each dataset relative to resolution. The number of epochs and the final model checkpoint is adjusted in the code based on validation loss during training. Our method was trained for 400 epochs.

4.2. Experimental results

Our main results include training on 4 datasets and evaluation on 14 datasets (Figure 2).

Comparison under supervised setup. When trained and

Figure 2. Bar plots depict the comparative performance of Topaz, crYOLO, and our method across four key metrics (IoU, Recall, Precision, F1), trained on distinct datasets (10291, 10077, 10590, 10816), distinguished by colour. Evaluation is conducted on 14 datasets, illustrated on the horizontal axis.

evaluated on the same datasets (10291, 10077, 10590, 10816), *our method outperforms Topaz in F1 metric and closely matches crYOLO in 3 out of 4 cases*, except dataset 10291. However, both supervised methods surpass ours in the IoU metric. Therefore, in the supervised setting, while *our method is comparable in identifying particles*, it underperforms in predicting precisely the center of the particles.

Generalization ability. The results in Figure 2 suggest that *supervised methods struggle to generalize to unseen data*, often exhibiting exceptionally low performance across metrics like F1 score, IoU, precision, or recall.

In contrast, our method demonstrates **notable cross-dataset generalization capabilities**, with nearly stable performance across various evaluation datasets, regardless of the training data. This suggests our method effectively mitigates the impact of dataset-specific noise levels and micrograph characteristics. These findings imply our method *can learn the necessary invariances irrespective of the inherent experimental nuances in cryo-EM procedures*.

5. Conclusion

We have introduced the first self-supervised approach for particle picking in cryo-EM micrographs. Our method leverages hierarchical denoising to effectively filter out different levels and types of noise present in the data. As demonstrated by our extensive experimental evaluation, our method exhibits robust generalization capabilities when applied to unseen datasets, outperforming supervised methods by a substantial margin. Our work alleviates the need for costly expert annotations, a significant bottleneck in cryo-EM data analysis, thus enabling more accessible and automated solutions in this domain.

References

Agard, D., Cheng, Y., Glaeser, R. M., and Subramaniam, S. Single-particle cryo-electron microscopy (cryo-em): Progress, challenges, and perspectives for further improvement. *Advances in imaging and electron physics*, 185:113–137, 2014.

Al-Azzawi, A., Ouadou, A., Max, H., Duan, Y., Tanner, J. J., and Cheng, J. Deepcryopicker: fully automated deep neural network for single protein particle picking in cryo-em. *BMC bioinformatics*, 21:1–38, 2020.

Bai, X.-C., McMullan, G., and Scheres, S. H. How cryo-em is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

Baxter, W. T., Grassucci, R. A., Gao, H., and Frank, J. Determination of signal-to-noise ratios and spectral snrs in cryo-em low-dose imaging of molecules. *Journal of structural biology*, 166(2):126–132, 2009.

Bendory, T., Bartesaghi, A., and Singer, A. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., and Berger, B. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods*, 16(11):1153–1160, 2019.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Chung, J. M., Durie, C. L., and Lee, J. Artificial intelligence in cryo-electron microscopy. *Life*, 12(8):1267, 2022.

Dhakal, A., Gyawali, R., Wang, L., and Cheng, J. A large expert-curated cryo-em image dataset for machine learning protein particle picking. *Scientific Data*, 10(1):392, 2023.

Dhakal, A., Gyawali, R., Wang, L., and Cheng, J. Cryotransformer: a transformer model for picking protein particles from cryo-em micrographs. *Bioinformatics*, 40 (3):btae109, 2024.

Doerr, A. Single-particle cryo-electron microscopy: a brief overview of how to solve a macromolecular structure using single-particle cryo-electron microscopy (cryo-em). *Nature methods*, 13(1):23–24, 2016.

Elferich, J., Schiroli, G., Scadden, D. T., and Grigorieff, N. Defocus corrected large area cryo-em (deco-lace) for label-free detection of molecules across entire cell sections. *Elife*, 11:e80980, 2022.

Gyawali, R., Dhakal, A., Wang, L., and Cheng, J. Accurate cryo-em protein particle picking by integrating the foundational ai image segmentation model and specialized u-net. *bioRxiv*, 2023.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Milne, J. L., Borgnia, M. J., Bartesaghi, A., Tran, E. E., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A., and Subramaniam, S. Cryo-electron microscopy–a primer for the non-microscopist. *The FEBS journal*, 280(1):28–45, 2013.

Nguyen, N. P., Ersoy, I., Gotberg, J., Bunyak, F., and White, T. A. Drpnet: automated particle picking in cryo-electron micrographs using deep regression. *BMC bioinformatics*, 22:1–28, 2021.

Nogales, E. and Scheres, S. H. Cryo-em: a unique tool for the visualization of macromolecular complexity. *Molecular cell*, 58(4):677–689, 2015.

Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. cryosparc: algorithms for rapid unsupervised cryoem structure determination. *Nature methods*, 14(3):290–296, 2017.

Scheres, S. H. Relion: implementation of a bayesian approach to cryo-em structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

Scheres, S. H. Semi-automated selection of cryo-em particles in relion-1.3. *Journal of structural biology*, 189(2): 114–122, 2015.

Skalidis, I., Kyrilis, F. L., Tüting, C., Hamdi, F., Chojnowski, G., and Kastritis, P. L. Cryo-em and artificial intelligence visualize endogenous protein community members. *Structure*, 30(4):575–589, 2022.

Tegunov, D. and Cramer, P. Real-time cryo-electron microscopy data preprocessing with warp. *Nature methods*, 16(11):1146–1152, 2019.

Voss, N., Yoshioka, C., Radermacher, M., Potter, C., and Carragher, B. Dog picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. *Journal of structural biology*, 166(2):205–213, 2009.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., et al. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications biology*, 2(1):218, 2019.

Wang, F., Gong, H., Liu, G., Li, M., Yan, C., Xia, T., Li, X., and Zeng, J. Deeppicker: A deep learning approach for fully automated particle picking in cryo-em. *Journal of structural biology*, 195(3):325–336, 2016.

Woolford, D., Ericksson, G., Rothnagel, R., Muller, D., Landsberg, M. J., Pantelic, R. S., McDowall, A., Pailthorpe, B., Young, P. R., Hankamer, B., et al. Swarmps: rapid, semi-automated single particle selection software. *Journal of structural biology*, 157(1):174–188, 2007a.

Woolford, D., Hankamer, B., and Ericksson, G. The laplacian of gaussian and arbitrary z-crossings approach applied to automated single particle reconstruction. *Journal of structural biology*, 159(1):122–134, 2007b.

Wu, J.-G., Yan, Y., Zhang, D.-X., Liu, B.-W., Zheng, Q.-B., Xie, X.-L., Liu, S.-Q., Ge, S.-X., Hou, Z.-G., and Xia, N.-S. Machine learning for structure determination in single-particle cryo-electron microscopy: A systematic review. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):452–472, 2021.

Xu, C., Zhan, X., and Xu, M. Cryomae: Few-shot cryo-em particle picking with masked autoencoders. *arXiv preprint arXiv:2404.10178*, 2024.

Yao, R., Qian, J., and Huang, Q. Deep-learning with synthetic data enables automated picking of cryo-em particle images of biological macromolecules. *Bioinformatics*, 36 (4):1252–1259, 2020.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pp. 12310–12320. PMLR, 2021.

Zhang, J., Wang, Z., Chen, Y., Han, R., Liu, Z., Sun, F., and Zhang, F. Pixer: an automated particle-selection method based on segmentation using a deep neural network. *BMC bioinformatics*, 20:1–14, 2019.

Zhang, X., Zhao, T., Chen, J., Shen, Y., and Li, X. Epicker is an exemplar-based continual learning approach for knowledge accumulation in cryoem particle picking. *Nature Communications*, 13(1):2468, 2022.

Zhu, Y., Ouyang, Q., and Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC bioinformatics*, 18:1–10, 2017.