CryoTransformer: A Transformer Model for Picking Protein Particles from Cryo-EM Micrographs

Ashwin Dhakal[1,2], Rajan Gyawali[1,2], Liguo Wang[3], Jianlin Cheng[1,2,*]

[1] Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA
[2] NextGen Precision Health, University of Missouri, Columbia, Columbia, MO 65211, USA
[3] Laboratory for BioMolecular Structure (LBMS), Brookhaven National Laboratory, Upton, NY 11973, USA

*Corresponding author: Jianlin Cheng (chengji@missouri.edu)

Abstract

Cryo-electron microscopy (cryo-EM) is a powerful technique for determining the structures of large protein complexes. Picking single protein particles from cryo-EM micrographs (images) is a crucial step in reconstructing protein structures from them. However, the widely used template-based particle picking process requires some manual particle picking and is labor-intensive and time-consuming. Though machine learning and artificial intelligence (AI) can potentially automate particle picking, the current AI methods pick particles with low precision or low recall. The erroneously picked particles can severely reduce the quality of reconstructed protein structures, especially for the micrographs with low signal-to-noise (SNR) ratios. To address these shortcomings, we devised CryoTransformer based on transformers, residual networks, and image processing techniques to accurately pick protein particles from cryo-EM micrographs. CryoTransformer was trained and tested on the largest labelled cryo-EM protein particle dataset - CryoPPP. It outperforms the current state-of-the-art machine learning methods of particle picking in terms of the resolution of 3D density maps reconstructed from the picked particles as well as F1-score and is poised to facilitate the automation of the cryo-EM protein particle picking.

1. Introduction

Cryogenic electron microscopy (cryo-EM) is a modern biophysical technique that captures two-dimensional (2D) images of biological macromolecules, such as proteins and viruses at cryogenic temperature [1], through the use of an electron detection camera. When subjected to an electron beam within a thin vitrified sample, this technique generates 2D image projections of the specimens (e.g., protein particles). These 2D representations are stored in various image formats (like mrc, tiff, tbz, eer, png, etc.), which are called micrographs. A single micrograph can contain hundreds or thousands of particles of a protein, randomly oriented in different directions. Given the inherent challenges of ascertaining the orientations of the particles and the low SNR of micrographs, hundreds of thousands of high-quality particles are often required to be identified to determine a high-resolution three-dimensional (3D) structure of the protein.

The initial step of determining the 3D structure of the proteins from the micrographs involves the recognition and extraction of particles from 2D micrographs, which is commonly referred to as particle picking. Its primary goal is to identify and locate individual protein particles within each micrograph while excluding malformed particles, crystalline ice contamination, and background regions. Essentially, the task of particle picking involves taking a micrograph as input and generating the coordinates for all protein particles present in that micrograph as the desired output (stored in the form of *.box* or *.star* files). These

coordinates thereafter serve as the data for subsequent stages of 3D protein structure reconstruction. These 3D structures of proteins are important for understanding their biological functions [2] and their interactions with ligands [3], [4], facilitate structure-based drug discovery [3] [5] .

Because the SNR of micrographs is generally low, hundreds to thousands of micrographs need to be generated to obtain a high-resolution structure for a protein, from which as many as millions of particle images can be picked. Precise identification of true particles is important, as the presence of false positive particles complicates the down-stream 3D protein reconstruction process. The particle picking task is inherently challenging due to several factors, including high noise levels caused by ice and contamination, low contrast of particle images, heterogenous conformations of particles, and variation in the orientation of particles.

This manual picking process by human is laborious, tedious, and time-consuming, which cannot be applied to pick millions of particles from thousands of micrographs. Therefore, substantial efforts have been put to develop semi-automated or fully automated methods to pick protein particles, which can be classified into two categories: (1) template-based particle picking and (2) machine learning particle picking.

In the template-based particle picking, the identification of particles primarily hinges on measuring a potential particle's similarity to user-predefined (manually selected) reference particles called templates. Because micrographs are usually noisy due to various factors such as ice contamination, carbon areas, overlapping particles, and other impurities, the template-based particle picking is often unable to detect particles of unusual shape and suffers from high false-positive rates. As a result, subsequent steps of manual particle selection are necessary to filter the particles picked by the template-based particle picking. Typically, iterative 2D-3D classification techniques are employed to scrutinize the picked particles and remove false particles. However, this particle picking, and downstream manual curation may introduce a degree of human bias into the final particle set selection, which may mistakenly exclude rare particle views and distinct conformations that are important for building high-resolution protein structures. Thus, this approach generally necessitates a large degree of human intervention and trial and errors to obtain good results.

The machine learning particle picking consists of both unsupervised learning (clustering) methods [6] and supervised learning methods [7] [8] [9] [10]. Recently, a number of deep learning methods were developed to automate the protein particle picking, which include XMIPP [11], DeepPicker [12], DeepEM [13], Xiao et al.'s method [14], Warp [15], HydraPicker [16], McSweeney et al.'s method [17], DRPnet [18], CrYOLO [19] and Topaz [20]. Among them, CrYOLO and Topaz based on convolutional neural networks have been widely used in particle picking. However, they have been trained with limited particle data and have the difficulty to generalize to new protein types or shapes. For instance, CrYOLO usually overlooks many true protein particles, while Topaz often picks excessive numbers of duplicate particles and some false positives such as ice contaminants and false particles in carbon-rich areas.

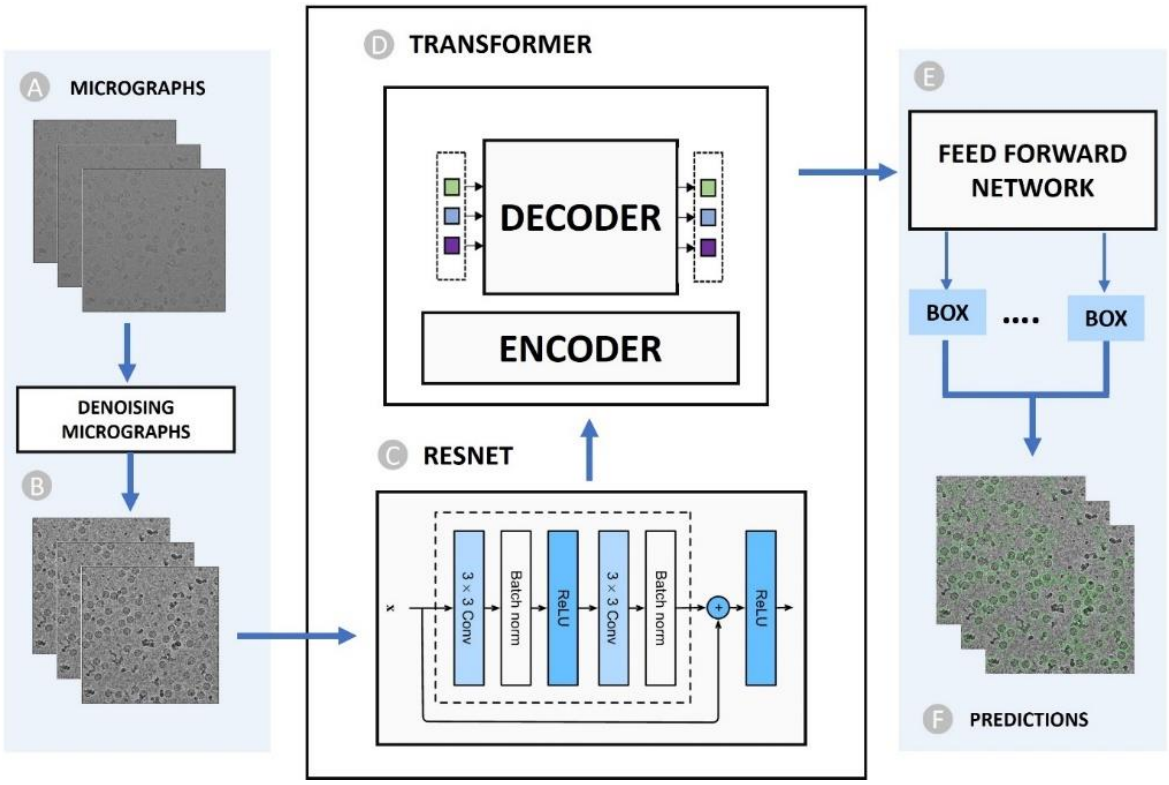

Figure 1*: Overview of the CryoTransformer Particle Picking Pipeline. (A) Input raw micrograph undergoes initial denoising. (B) Denoised micrographs serve as input for subsequent processing. (C) CNN-Based Resnet-152 architecture extracts image features. Features extracted in (C) are processed by an (D) encoder-decoder Transformer. (E) Feed-forward networks further refine the processed data. (F) Predictions of particles encircled in micrographs, eventually stored in star files as the final output.*

To overcome these obstacles, we devised a transformer-based particle picking approach and trained it on the largest, diverse, manually-labelled CryoPPP protein particle dataset [21], [22]. Inspired by Meta's Detection Transformer (DETR) [23] for detecting small objects, we designed the end-to-end detection transformer named as CryoTransformer. Briefly, it has an initial step of reducing noise in micrographs (**Figure 1A, B**), followed by the feature extraction through a ResNet-152 architecture (**Figure 1C**). Subsequently, a transformer model is used for detecting protein particles as shown in **Figure 1D**. This is succeeded by the feed-forward networks to predict particles (**Figure 1E**), which are followed by the post-processing procedures. The output (**Figure 1F**) includes particle markings on the micrographs stored in *.star* files, which can be directly used for the subsequent stages of 3D protein structure reconstruction. We conducted a rigorous evaluation of CryoTransformer. It outperforms the two popular deep learning methods: CrYOLO and Topaz. The source code and data for CryoTransformer are openly available at: https://github.com/jianlin-cheng/CryoTransformer .

2. Materials and Methods

2.1 Dataset

Dataset acquisition

We utilized the largest comprehensive CryoPPP dataset [21], [22] curated from Electron Microscopy Public Image Archive (EMPIAR) [24], to train, validate, and test CryoTransformer. The micrographs of 22 proteins (EMPIAR IDs) from the CryoPPP dataset were used, with the data of each EMPIAR ID split according to an 80%-10%-10% ratio for training, validation, and internal test. Moreover, we used the data of 6 distinct EMPIAR IDs in CryoPPP dataset different from the 22 proteins above as well as the 4 complete micrograph datasets from EMPIAR repository [24] as the independent test dataset to compare CryoTransformer with the external methods.

The selection of training and test data considered a range of protein attributes, including type, shape, size, and overall structural characteristics. The 22 proteins used for the training, validation and internal test are described in **Table 1**. **Supplementary Figure S1** illustrates the varying defocus values of the training data. The datasets encompass various protein categories, such as transport proteins, membrane proteins, viral proteins, ribosomes, signaling proteins, aldolases, and more. They are comprised of micrographs featuring diverse attributes, including those with ice patches, contaminants, varying ice thickness, and carbon areas. Different protein distribution patterns, including monodisperse, clumped clusters, and heterogeneous views, are also included. The **Supplementary Table S1 and S2** contain the information and statistics of the proteins in the independent test dataset.

Table 1: *The statistics and information of the 22 sets of micrographs for training, validation, and internal test of CryoTransformer (* Theoretical weight)*

SN	EMPIAR ID	Type of Protein	Image Size	Total Structure Weight (kDa)	# Training Micrographs	# Validation Micrographs	# Test Micrographs	# Total Micrographs
1	11183 [25]	Signaling Protein	(5760, 4092)	139.36	250	25	25	300
2	11057 [26]	Hydrolase	(5760, 4092)	149.43	250	25	20	295
3	11051 [27]	Transcription/DNA/RNA	(3838, 3710)	357.31	250	25	25	300
4	10852 [28]	Signaling Protein	(5760, 4092)	157.81	270	40	33	343
5	10816 [29]	Transport Protein	(7676, 7420)	166.62	250	25	25	300
6	10760 [30]	Membrane Protein	(3838, 3710)	321.69	250	25	25	300
7	10737 [31]	Membrane Protein	(5760, 4092)	155.83	250	25	17	292
8	10671 [32]	Signaling Protein	(5760, 4092)	77.14	250	25	23	298
9	10590 [33]	Transport Protein	(3710, 3838)	1000*	250	25	21	296
10	10526 [34]	Ribosome (50S)	(7676, 7420)	1085.81	180	20	20	220
11	10444 [35]	Membrane Protein	(5760, 4092)	295.89	250	25	21	296
12	10406 [36]	Ribosome (70S)	(3838, 3710)	632.89	200	20	19	239
13	10387 [37]	Viral Protein	(3710, 3838)	185.87	250	25	24	299
14	10291 [38]	Transport Protein	(3710, 3838)	361.39	250	25	25	300
15	10289 [38]	Transport Protein	(3710, 3838)	361.39	250	25	25	300
16	10240 [39]	Lipid Transport Protein	(3838, 3710)	171.72	250	25	24	299
17	10184 [40]	Aldolase	(3838, 3710)	150*	250	25	21	296
18	10096 [41]	Viral Protein	(3838, 3710)	150*	250	25	25	300
19	10077 [42]	Ribosome (70S)	(4096, 4096)	2198.78	250	25	25	300
20	10075 [43]	Bacteriophage MS2	(4096, 4096)	1000*	250	25	24	299

21	10059 [44]	Transport Protein	(3838, 3710)	317.88	250	25	16	291
22	10005 [45]	Transport Protein	(3710, 3710)	272.97	22	4	3	29
		Total Micrographs			**5,172**	**534**	**486**	**6,192**

Denoising and pre-processing of cryo-EM micrographs

The cryo-EM micrographs in *.mrc* format, serve as the initial input for CryoTransformer. To reduce noise and improve the signal-to-noise ratio, a Gaussian filter with a kernel size of 9 is applied to convolve with the images. Subsequently, the images undergo standard normalization to achieve consistent intensity ranges. The normalized pixel values of the images are computed using the formula [pixel = (pixel- μ)/ σ], ensuring that the data is centered and scaled appropriately for the further analysis. The normalized images are then converted to grayscale, which collapses multi-channel intensity information into a single channel, ensuring a uniform representation of pixel values ranging from 0 to 255 (**Figure 2A**).

Effective noise reduction is essential to reveal clear structural details in cryo-EM micrographs. We employ a two-step denoising process to the normalized images, involving Fast Non-Local Means (FastNLMeans) denoising followed by Weiner filtering (**Figure 2B**). FastNLMeans denoising is employed to retain image details while suppressing noise artifacts. By exploiting the redundancy present in natural images, FastNLMeans replaces the noisy pixel with a weighted average of similar pixels from a larger neighborhood. The trade-off between noise suppression and detail preservation is controlled by the choice of template window size (7 in this case) and the search window size (21 in this case).

The output of FastNLMeans denoising is subjected to Weiner filtering to further reduce the residual noise and enhance the image's structural fidelity (**Figure 2C**). It achieves this by estimating the original image's frequency spectrum and applying a correction factor to mitigate the effects of noise. Enhancing contrast in cryo-EM micrographs is crucial for improving particle visibility and overall image quality. We incorporate the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique for this purpose (**Figure 2D**). The CLAHE technique, with a clip limit of 2 and a tile grid size of 16x16, is applied to the denoised images. This technique effectively addresses non-uniform illumination and low contrast, leading to enhanced visual clarity.

To accomplish selective smoothing and fine detail preservation, guided filtering is performed using the CLAHE-enhanced image as a guide (**Figure 2E**). Guided filtering operates by estimating the local linear relationship between the guidance image and the target image (**Figure 2F**). This relationship is then used to determine the filtering weights applied to each pixel, resulting in controlled smoothing, while retaining sharp edges and fine details. The filtering fine-tunes the micrographs, achieving a balance between noise reduction and preservation of important structural information (**Figure 2G**).

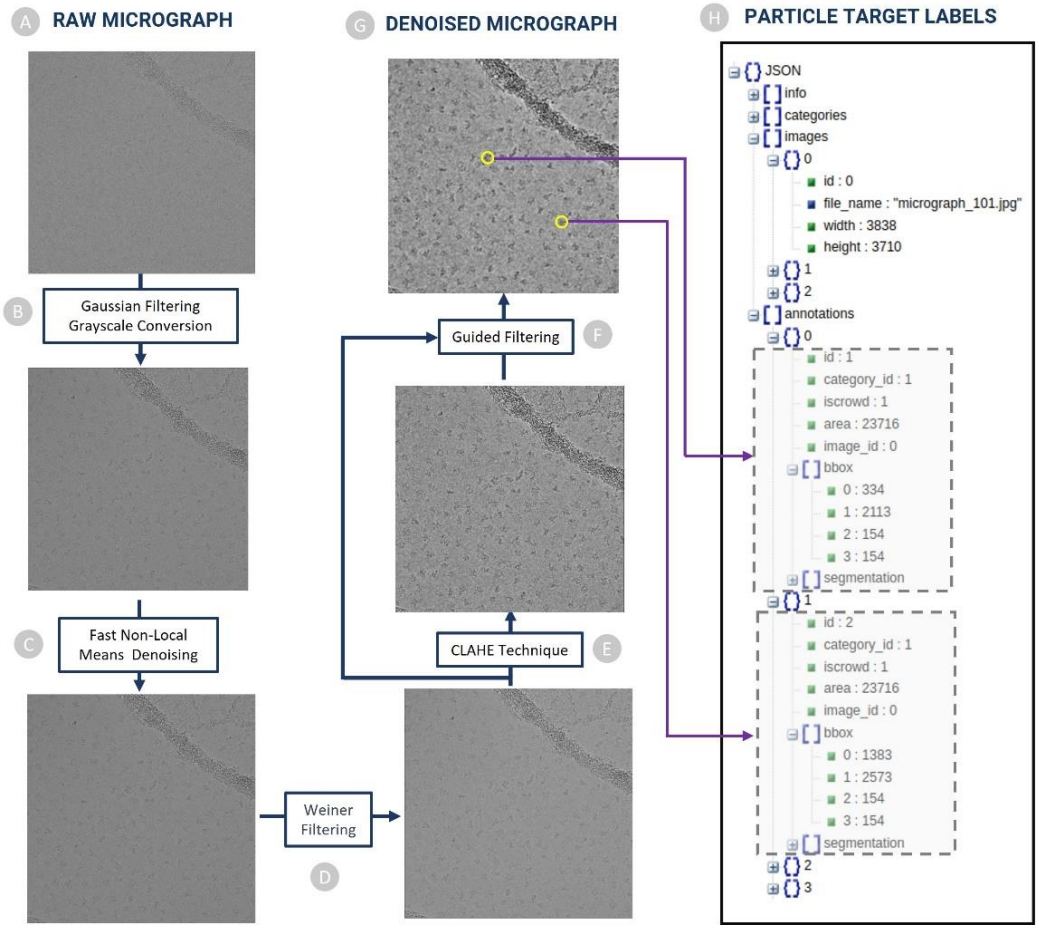

Figure 2: *The denoising used to preprocess cryo-EM micrographs in CryoTransformer. (A) Raw micrographs with low contrast and low SNR go through (B) Gaussian filtering and Grayscale conversion. Normalized micrograph undergoes (C) FastNLMeans denoising technique. (D) Weiner filtering is applied to the micrographs from the previous step, and subsequently (E) CLAHE technique is used to enhance visual clarity of the micrograph. Eventually, (F) Guided filtering is performed using the CLAHE-enhanced image as a guide to obtain (G) denoised micrographs. (H) Ground truth particle annotation data. Particle coordinates from ground truth coordinate files are extracted to create COCO-dataset that is used as target labels for training CryoTransformer.*

Generating COCO-dataset for labelled protein particles in micrographs

We used the ground truth particle coordinate data from the CryoPPP dataset [21], [22] to generate labels to train CryoTransformer. The particle labels were stored in the widely adopted Common Objects in Context (COCO) format [46]. This format is extensively used for object detection and segmentation tasks, and it adheres to a structured JSON layout that defines how labels and associated metadata are stored for an image dataset. An illustration of how these labels are stored is depicted in **Figure 2H**. In the case of all training and validation images, we have two JSON files: one for training (referred to as the "train JSON") and another for validation (referred to as the "validation JSON"). We chose to adopt this labeling data format because the COCO format imposes a standardized structure for annotations, including object category labels and bounding box coordinates. This uniformity streamlines the data preprocessing process and ensures that models can readily comprehend and learn from the annotated data. The COCO format permits the annotation of multiple objects (protein particles) within a single image (micrograph). Each object is associated with its distinct category label and bounding box. For each particle, we retain details such as its bounding box coordinates, area, category label (typically set to 1 in our case as all objects to be detected are protein particles), the corresponding image reference, and a unique particle ID.

2.2 Design and Implementation of CryoTransformer

CryoTransformer is designed to achieve the accurate prediction of bounding boxes for the protein particles within a micrograph, while minimizing the number of false positives. It undergoes an end-to-end training, using a specialized loss function that effectively combines the bipartite matching loss between predicted and ground-truth protein particles in the micrographs.

CryoTransformer Architecture

As illustrated in **Figure 3**, CryoTransformer comprises three main components: a Convolutional Neural Network (CNN) with residual connections (Resnet-152 [47]) responsible for feature extraction, an encoder-decoder transformer [23], [48] for learning the shapes of the particles in the context of an entire image, and a feed-forward network (FFN) responsible for producing the ultimate particle predictions.

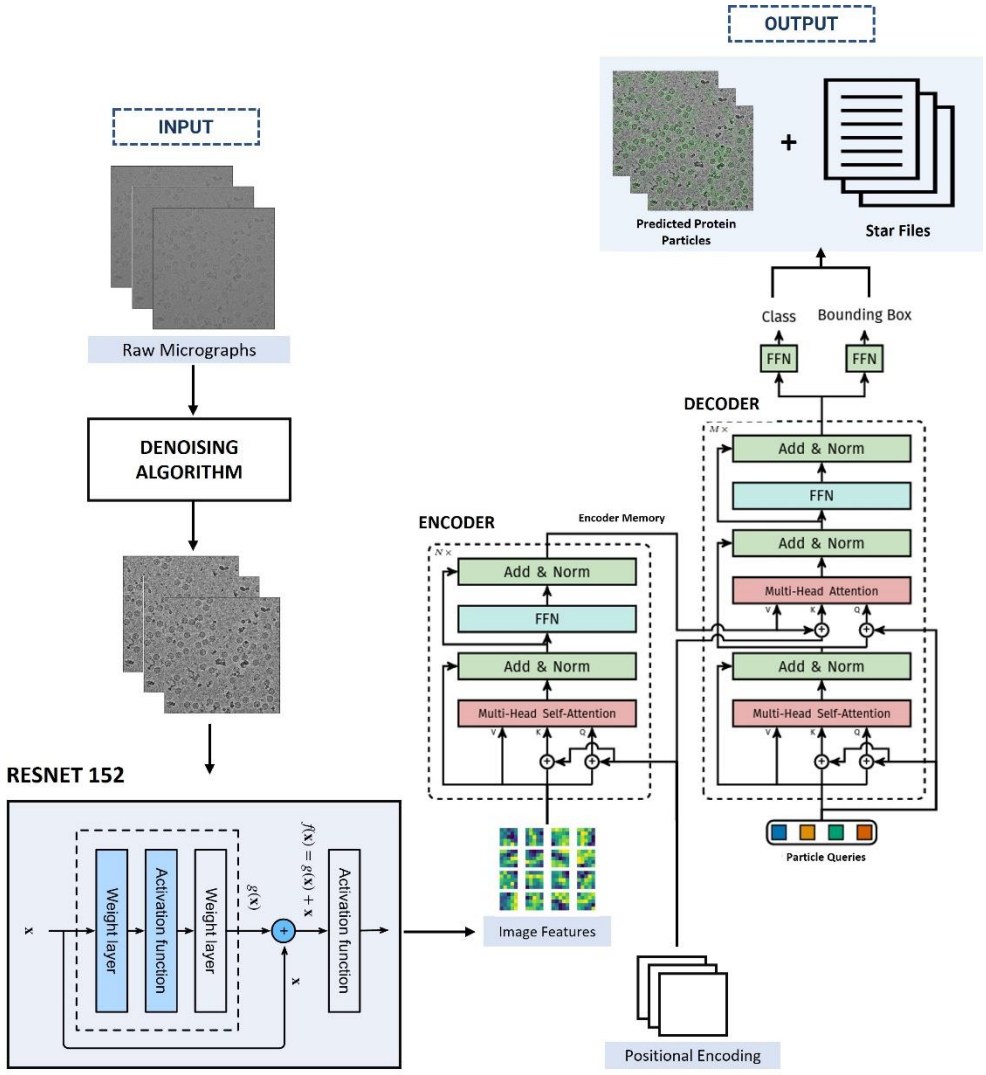

***Figure 3**: Architecture of CryoTransformer. The raw micrographs are denoised and are fed into the ResNet-152 module for feature extraction. The images features, along with positional encoding, are fed to the encoder of the transformer. The output from the encoder is subsequently passed to the decoder layer. Finally, the decoder's output is passed to the feed forward networks that*

generate the protein particle bounding box predictions. These predictions are used in generating the predicted protein particles encircled in micrographs, which are stored in the .star files.

Resnet-152 Backbone Block

The Resnet-152 receives the preprocessed micrographs $x_{\text{img}} \in \mathbb{R}^{3\times H_0\times W_0}$ (with 3 color channels)) as input and generates a lower-resolution activation map as $f \in \mathbb{R}^{C\times H\times W}$, Where $C = 2048$, and $H = \frac{H_0}{32}$, $W = \frac{W_0}{32}$. 0 padding is applied to the images in a batch to make sure that they all have same input dimensions (H_0, W_0) as the largest image size of the batch.

Transformer Module

The features extracted from the Resnet-152 are subsequently passed through the transformer. This transformer consists of two main components: encoder and decoder. The image features from the Resnet-152 backbone block are passed through the transformer along with the positional encoding and particle queries. The transformer outputs intermediate predictions, which are fed to the FFN module to predict particle labels and bounding boxes.

Transformer Encoder

The encoder plays a vital role in generating coherent and context-aware outputs. In the encoder, a 1x1 convolution operation is used to decrease the channel dimension of the high-level activation map, denoted as f, from C to a smaller dimension d, yielding a new feature map $z_0 \in \mathbb{R}^{d\times H\times W}$. Since the encoder accepts a one-dimensional sequence as input, we collapse the spatial dimensions of z_0 into a single dimension. As a result, the resultant input becomes a feature map of dimension $d \times HW$. Here, every encoder layer follows a consistent structure, comprising a multi-head self-attention component and a FFN layer. To account for the permutation-invariant nature of the transformer architecture, we enhance it by incorporating the positional encodings [49] [50], which are included in the input of every multi-head self-attention layer.

Transformer Decoder

The decoder receives the memory from encoder, positional encoding, and particle queries as input. It involves the transformation of N embeddings of size d (in our specific scenario, N = 600, meaning predicting max 600 protein particles per micrograph) through the multi-headed self- attention mechanisms. It's worth noting that since the decoder is also designed to be permutation-invariant, it requires distinct particle queries (initialized as random vectors) within the set of N inputs to generate different outcomes. These particle queries, added to the input at each attention layer, are a are updated through back propagation. Subsequently, the output of the decoder is individually used to predict box coordinates and class labels (1 in our case) through a feed-forward network, a process detailed in the following subsection, resulting in N final predictions.

Feed-Forward Networks Module

The final prediction is generated through a 3-layer perceptron with a ReLU activation function and d hidden nodes in each hidden layer, followed by a linear projection layer. This FFN is responsible for predicting the normalized center coordinates, height, and width of the bounding box relative to the input micrograph. Additionally, the linear layer predicts the class label using a softmax function. Considering that we are making predictions for a fixed-size set of N potential bounding boxes, and N is typically much larger than the actual number of protein particles in a single micrograph, we introduce a special class label denoted as

$\emptyset$. This label means that no protein particle has been detected in a particular slot. Its role is akin to the "background" class in conventional object detection.

Loss Function

CryoTransformer generates a consistent set of N predictions in a single traversal of the decoder. This number N was deliberately chosen to exceed the usual count of protein particles in a micrograph. To achieve this, the loss function is designed to establish an ideal bipartite matching between the predicted protein particles and their corresponding ground truth. Subsequently, the model optimizes the losses pertaining to individual particles in order to refine the predictions further.

We can represent the ground truth set of particles as y and the set of N predictions as $\hat{y} = \{\hat{y}_i\}_{i=1}^{N}$. When N exceeds the number of true protein particles in the micrograph, we enlarge y as a set of size N, with padding represented by $\emptyset$ (no protein particle). To find the optimal bipartite matching between these two sets, we aim to find a permutation of N elements denoted as $\sigma \in \mathfrak{S}_N$ that incurs the lowest cost. This permutation is determined by the following equation, given in equation I:

$$\hat{\sigma} = \arg\min_{\sigma \in \mathfrak{S}_N} \sum_{i}^{N} \mathcal{L}_{\text{match}}\left(y_i, \hat{y}_{\sigma(i)}\right) \quad \text{I}$$

$\mathcal{L}_{\text{match}}\left(y_i, \hat{y}_{\sigma(i)}\right)$ represents the pairwise matching cost between the ground truth particle y_i and a prediction indexed by $\sigma(i)$. This cost is calculated using the following equation II:

$$\mathcal{L}_{\text{match}}\left(y_i, \hat{y}_{\sigma(i)}\right) = -\mathbb{1}_{\{c_i \neq \emptyset\}} \hat{p}_{\sigma(i)}(c_i) + \mathbb{1}_{\{c_i \neq \emptyset\}} \mathcal{L}_{\text{box}}\left(b_i, \hat{b}_{\sigma(i)}\right) \quad \text{II}$$

We can view each element i in the ground truth set as a $y_i = (c_i, b_i)$, where c_i represents the target class label, and b_i belongs to the range $[0,1]^4$, representing a vector that specifies the center coordinates of the ground truth box, along with its height and width relative to the micrograph dimensions. This approach ensures a one-to-one matching, preventing duplicate predictions when directly predicting sets.

The next stage involves calculating the Hungarian loss using the Hungarian algorithm [51] for all pairs that were matched in the preceding step. We define this loss according to the equation III:

$$\mathcal{L}_{\text{Hungarian}}(y, \hat{y}) = \sum_{i=1}^{N} \left[-\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbb{1}_{\{c_i \neq \emptyset\}} \mathcal{L}_{\text{box}}\left(b_i, \hat{b}_{\hat{\sigma}}(i)\right)\right] \quad \text{III}$$

Here, $\hat{\sigma}$ represents the optimal assignment obtained from the initial equation I.

In practical implementation, we apply a down-weighting factor of 10 to the log-probability term when c_i is equal to $\emptyset$, denoting the absence of a particle. This adjustment is made to address the issue of class imbalance. The second part of the Hungarian loss ($\mathcal{L}_{\text{box}}(\cdot)$) scores the bounding boxes is given by the equation IV:

$$\mathcal{L}_{\text{box}}\left(b_i, \hat{b}_{\sigma(i)}\right) = \lambda_{\text{iou}} \mathcal{L}_{\text{iou}}\left(b_i, \hat{b}_{\sigma(i)}\right) + \lambda_{\text{L1}} \left\| b_i - \hat{b}_{\sigma(i)} \right\|_1 \quad \text{IV}$$

Where $\lambda_{\text{iou}}, \lambda_{\text{L1}} \in \mathbb{R}$ are hyperparameters and $\mathcal{L}_{\text{iou}}(\cdot)$ is the generalized IoU [52] given by equation V:

$$\mathcal{L}_{\text{iou}}\left(b_{\sigma(i)}, \hat{b}_i\right) = 1 - \left(\frac{\left|b_{\sigma(i)} \cap \hat{b}_i\right|}{\left|b_{\sigma(i)} \cup \hat{b}_i\right|} - \frac{\left|B\left(b_{\sigma(i)}, \hat{b}_i\right) \setminus b_{\sigma(i)} \cup \hat{b}_i\right|}{\left|B\left(b_{\sigma(i)}, \hat{b}_i\right)\right|}\right) \quad \text{V}$$

In the context provided, |.| denotes "area," and we use the terms union and intersection of box coordinates as shorthand references for the boxes themselves. To compute the areas of unions or intersections, we rely on the minimum/maximum of linear functions involving $b_{\sigma(i)}$and $\hat{b}_i$. This approach ensures that the loss behaves in a stable manner for the computation of stochastic gradients. $B\left(b_{\sigma(i)}, \hat{b}_i\right)$ refers to the largest bounding box that contains both $b_{\sigma(i)}, \hat{b}_i$.

Model Implementation and Training

We trained CryoTransformer with AdamW optimizer [52] by setting the initial transformer's learning rate to 10^{-4}, the backbone's to 10^{-5}, and weight decay to 10^{-4}. All weights are randomly initialized with Xavier initialization [53]. Additive dropout of 0.1 is applied after every multi-head attention and FFN before layer normalization. We use a training schedule of 300 epochs with a learning rate drop by a factor of 10 after 200 epochs, where a single epoch is a pass over all training images once. Training the model for 300 epochs on NVIDIA A100 80GB GPU took 2 days and 11 hours to complete.

2.3 Postprocessing Predictions and Reconstructing Protein Density Maps from Picked Particles

The FFN module of CryoTransformer predicts the coordinates of particles and their corresponding confidence scores (ranging from 0 to1). The predictions are processed in a few steps to generate final particle predictions, group the picked particles into different 2D orientation classes, and use them to build 3D density maps of proteins. The visual representation of the overall process is shown in **Figure 4**.

The predictions are first used to generate individual box files for every micrograph for a protein, containing the center coordinates (x and y) of all the predicted protein particles. We retain only the particles whose confidence score falls in the range from 25$^{\text{th}}$ percentile to 100$^{\text{th}}$ percentile. Subsequently, these box files are merged to create a *.star* file that can be accepted by CryoSPARC [54] for density map construction for the protein.

The star files generated are imported into CryoSPARC through the *'import particles'* task, accompanied by input parameters such as Acceleration Voltage (kV), Spherical Aberration (mm), and Pixel Size (Å) as well as the patch-based Contrast Transfer Function (CTF)-estimated micrographs. Subsequently, these particles are extracted using a specified extraction box size (in pixels) and fed into the 2D classification function of CryoSPARC to group them into different orientation classes.

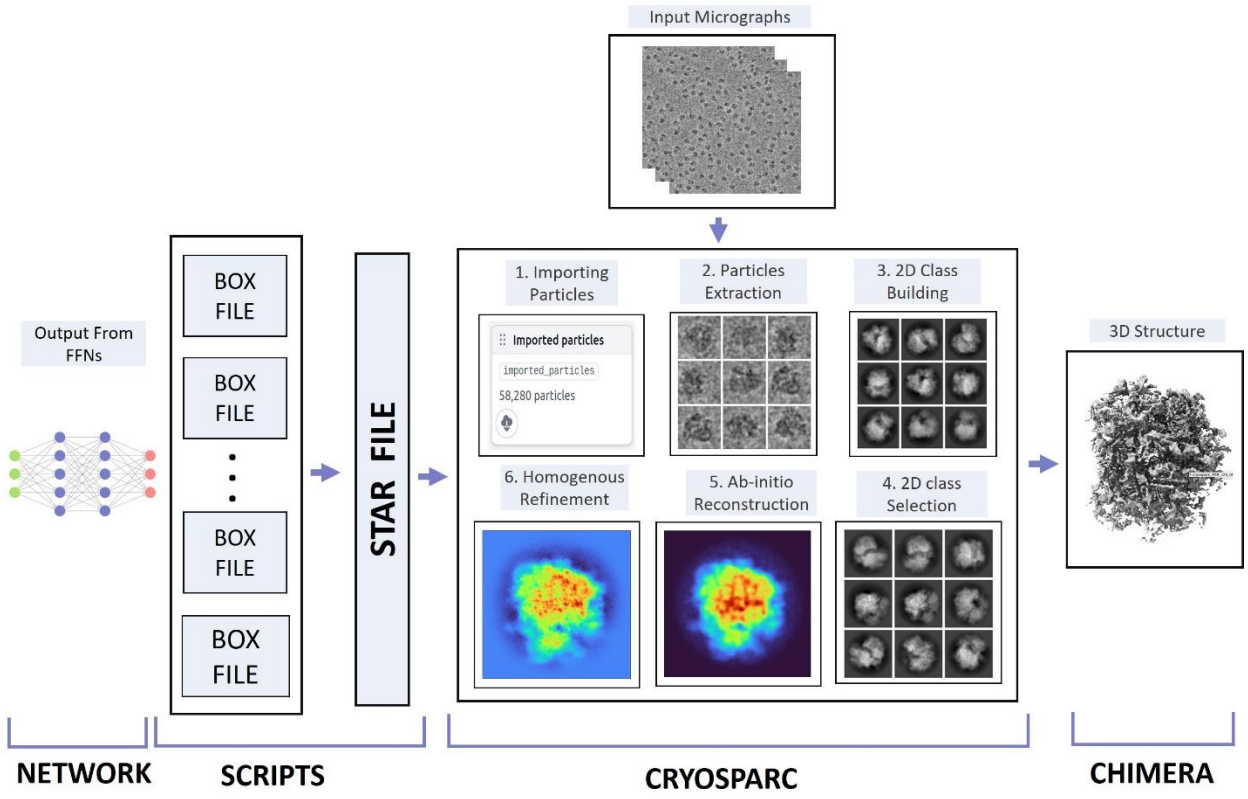

Figure 4: *Post processing steps to generate 3D protein density maps from picked particles.* CryoTransformer *outputs the .star files, which are* imported *into CryoSPARC* along *with the micrographs. Steps (1-6) are performed in CryoSPARC to generate the 3D density maps for a protein. The resolution of the density maps is employed as the main metric to evaluate the quality of the picked particles.*

This 2D classification step helps identify and exclude false particles through manual inspection, which usually can improve the resolution of the density maps reconstructed from the picked particles.

To assess the quality of the particles picked by CryoTransformer, CrYOLO and Topaz, we carried out the density map reconstruction experiments with and without the 2D selection respectively. When the 2D classification was used, we generated a total of 50 particle classes, employing a window inner radius of 0.85 and an outer radius of 0.99. Additionally, we performed 15 iterations to refine the CryoSPARC's noise model. The selected particles were used by an *ab initio* reconstruction process with the standard parameter settings, which includes 300 iterations of reconstruction with a Fourier radius step of 0.04 and a momentum of 0 and an initial learning rate of 0.4 for the stochastic gradient descent optimization. Additionally, a lowpass filter cutoff in Fourier radii of 7 was applied to the initial random structures.

After generating the initial density map for a protein, the cryoSPARC's '*homogeneous refinement'* job was employed to enhance it further. The homogeneous refinement was applied to correct the higher-order aberrations and to refine particle defocus caused by factors such as beam tilt and spherical aberration. To ensure the fairness in comparisons of the particle picking methods, the experiment was conducted three times for each method with different random seed values, and the best score (in Angstrom units) out of the three experiments was used in the comparison.

3. Results

We evaluated the particle picking performance of CryoTransformer in the following complementary ways. First, we compared it with CrYOLO and Topaz in terms of the resolution of the density maps reconstructed from the particles picked by them from the full set of micrographs in the EMPIAR repository for the four proteins in the independent test dataset. Second, we compared it with CrYOLO and Topaz in terms of the resolution of the density maps picked from a subset of labeled micrographs in the CryoPPP dataset for the proteins in the independent test dataset. Finally, we visually inspected and assessed the particles picked by the three methods.

3.1 Comparing CryoTransformer, CrYOLO, and Topaz in terms of resolution of density maps reconstructed from the particles picked from the full set of micrographs in the EMPIAR repository (~1600 micrographs per protein)

The full set of micrographs in the EMPIAR repository for the four test proteins (Human HCN1 Hyperpolarization-Activated Channel (EMPIAR 10081), Influenza Hemagglutinin (EMPIAR 10532), mechanotransduction channel NOMPC (EMPIAR 10093), and asymmetric αVβ8 (EMPIAR 10345)) in the independent test dataset were used to compare CryoTransformer, CrYOLO and Topaz. The resolution of the density map reconstructed from the particles picked by each method for each protein was calculated. The density maps were reconstructed by CryoSPARC in two modes: with 2D particle selection (*Select 2D*) or without it. The experiment for each method and each protein was conducted three times and the best results were selected for the comparison. The comparative results of the three methods are summarized in **Table 2**, while the detailed results of each trial reported in **Supplementary Table S3**.

Table 2: *Comparison of CryoTransformer with crYOLO and* Topaz's *performance in terms of the resolution of density maps reconstructed from the particles picked from the full set of micrographs of the four test proteins.*

EMPIAR ID	Number of Micrographs	Without Select 2D						With Select 2D					
		Number of Particles			3D Resolution (Å)			Number of Particles			3D Resolution (Å)		
		CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer
10081 [55]	997	59,559	383,558	293,980	7.45	6.34	**4.89**	32,472	148,378	147,662	6.39	4.19	**4.15**
10532 [56]	1,556	62,732	1,574,179	764,215	8.34	3.97	**3.86**	16,079	260,266	259,757	7.82	3.27	**3.21**
10093 [57]	1,873	53,482	791,064	596,192	6	**4.72**	6.11	40,374	359,619	204,355	5.57	**4.37**	4.65
10345 [58]	1,644	19,836	396,882	182,397	7.27	**3.5**	5.22	5,377	155,023	111,375	6.06	3.47	**3.45**

With *Select 2D*, CryoTransformer has the highest resolution of the reconstructed density maps for three out of four proteins (i.e., EMPIAR IDs: 10081, 10532, and 10345), while Topaz has the highest resolution for one protein. Without *Select 2D*, CryoTransformer and Topaz each perform best on two proteins. The detailed assessment of crYOLO, Topaz, and CryoTransformer based on the 3D resolution of Gold Standard Fourier Shell Correlation (CSFSC) curves, 3D density maps, and density projections with *Select 2D* is visualized in **Figure 5**.

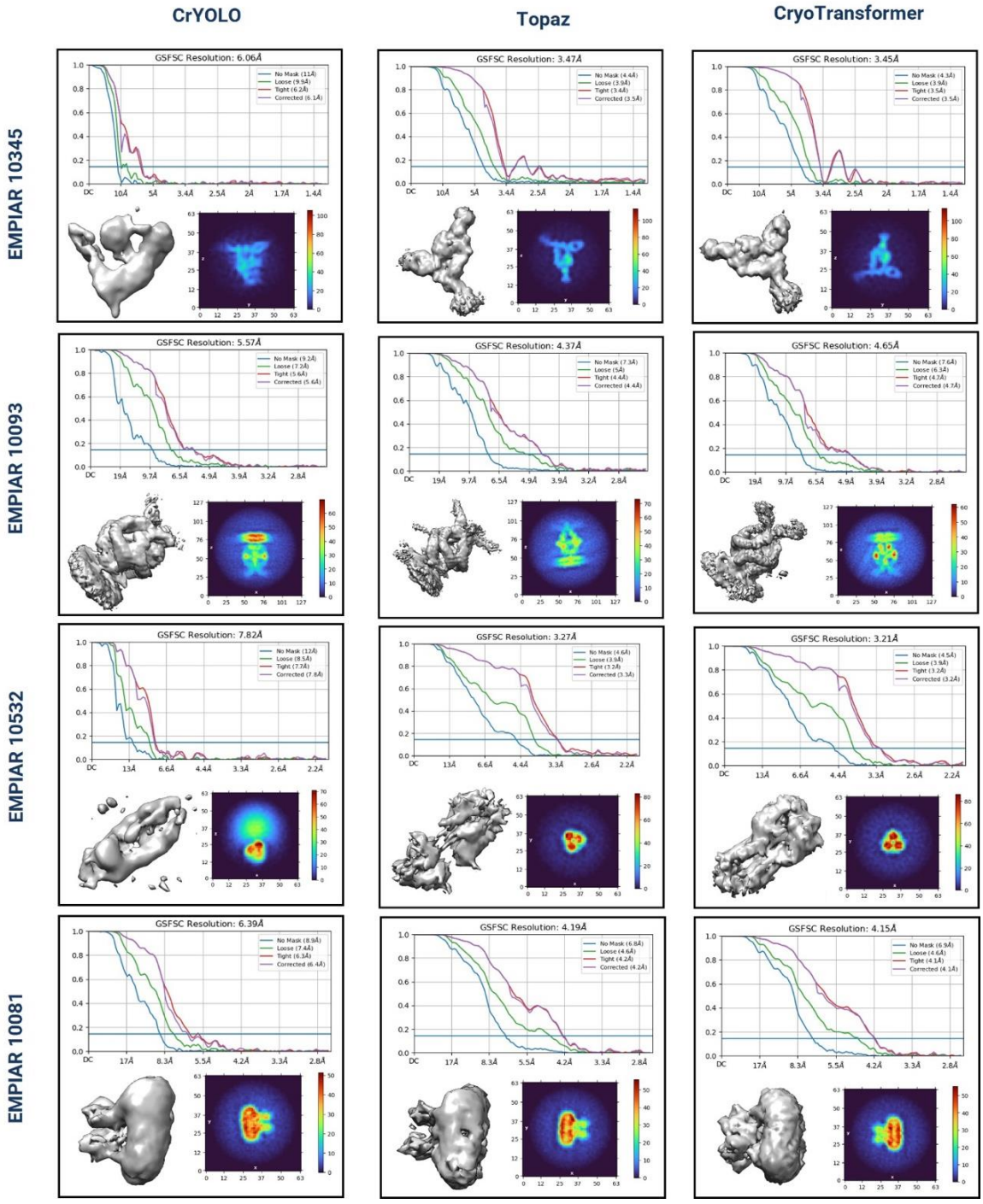

Figure 5: *Assessment of CrYOLO, Topaz, and CryoTransformer based on the 3D resolution CSFSC curves, 3D density maps, and density projections. The top diagram in each row shows CSFSC curves, which indicate the resolution of 3D density maps for proteins structures reconstructed from picked particles. Bottom-left image in each sub-figure provides a visual representation of the 3D density map. The bottom-right image in each sub-figure depicts the density projections from the intermediate output of the ab initio reconstruction phase. The integrated density values along the normal direction to that plane are displayed. The color scheme in the heatmap corresponds to the scalar density values at each voxel, with the color intensity indicating density magnitude.*

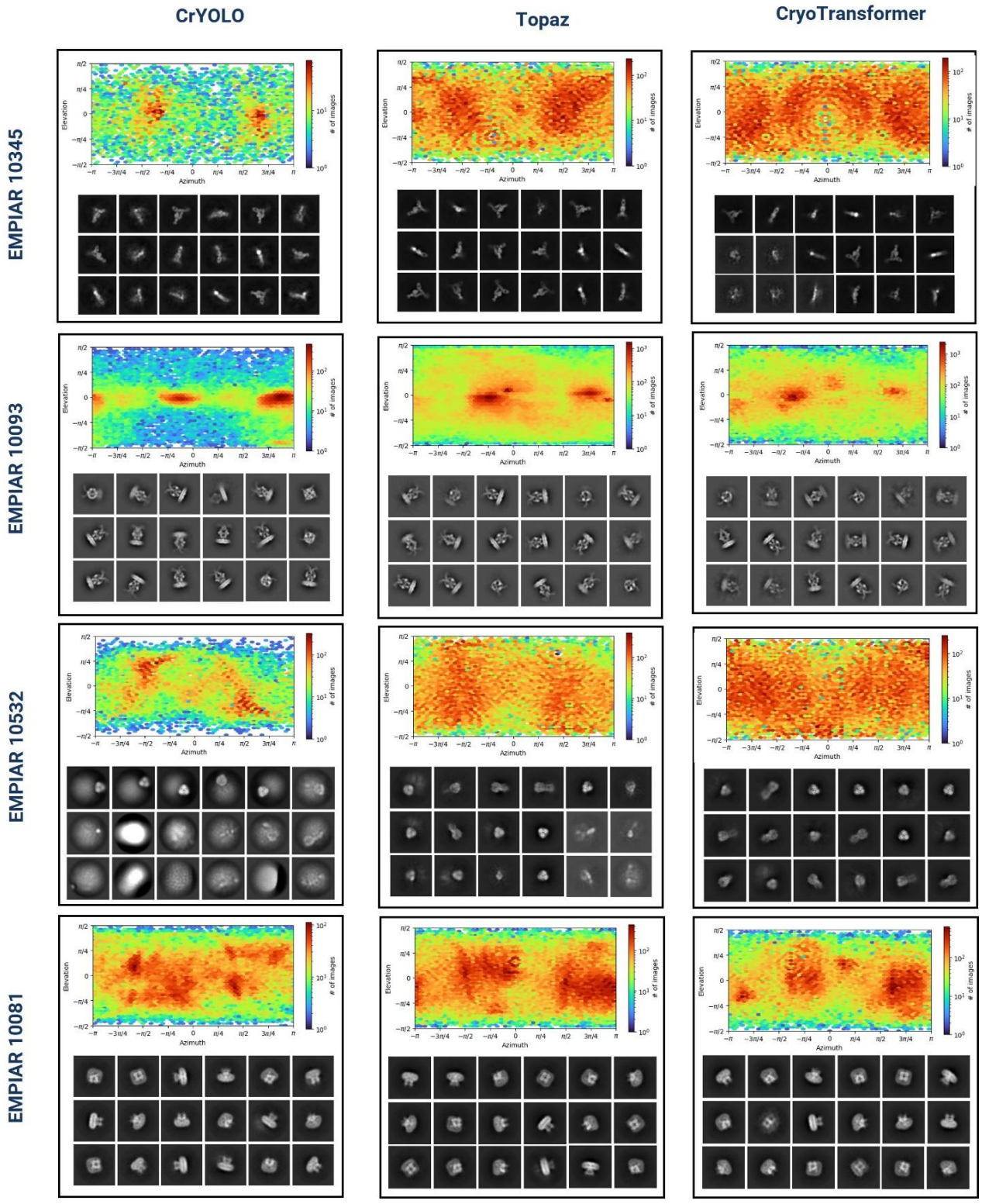

Figure 6: *Assessment of CrYOLO, Topaz, and CryoTransformer based on the 2Dl orientation classes of the picked protein particles. Each block displays two sections: the upper section presents the viewing direction plots as elevation vs azimuth plots, while the lower section showcases the averaged 2D orientation classes generated from picked particles.*

In *Figure 5***Figure 5** , CSFSC curves are plotted to assess the resolution of the obtained 3D density maps. Different variations of Fourier Shell Correlation (FSC) plots are presented: one employing an automatically generated mask with a 15 Å falloff, termed the 'loose mask' curve, and the other using an auto-generated mask with a falloff of 6 Å for all FSC plots, referred to as the 'tight mask' curve. The 3D density map reconstructed by each method for each protein is also visualized. The notable difference between the results of CrYOLO and CryoTransformer can be observed. For instance, in the case of EMPIAR 10345, the correct shape of the density map has three distinct legs, but CrYOLO failed to capture all three, yielding a lower resolution of 6.06 Å. In contrast, CryoTransformer captured all of them and achieved a high resolution of 3.45 Å. Similarly, in case of EMPIAR 10532, Topaz missed the central segment of the rod-like protein structure, whereas CryoTransformer successfully reconstructed that portion, attaining the highest resolution (3.21 Å) among all methods.

The plot located in the lower-right corner of each section in **Figure 5** represents the intermediate output of the ab-initio reconstruction phase. These plots depict density projections, but instead of slicing the density along a specific plane, the integrated density values along the normal direction to that plane are displayed. The color scheme in the heatmap corresponds to the scalar density values at each voxel, with color intensity indicating density magnitude.

In addition to this visual assessment in **Figure 5**, we conducted a comparison based on the 2D orientation classes of the picked particles (**Figure 6)**, showing that CryoTransformer picked particles in multiple orientation states that are important for obtaining high-resolution density maps. This analysis specifically involved analyzing the elevation vs azimuth plots for each test EMPIAR IDs. In the case of EMPIAR 10532 row, CrYOLO struggled to select particles representing various orientations, resulting in low-quality 2D particle classes. In contrast, Topaz performed reasonably well in picking particles with a diverse range of orientations, and CryoTransformer excelled in picking a substantial number of particles with a broad angular distribution, as indicated by the red color in the heatmap. The higher intensity of the red color in the upper section of each blocks in **Figure 6** corresponds to the higher number of particles in the elevation vs azimuth plots. Similarly, the lower section of each block depicts the averaged 2D orientation classes generated from picked particles. The diverse set of particles picked by CryoTransformer enabled the reconstruction of the density map of the highest resolution for this protein.

3.2. Comparing CryoTransformer, CrYOLO, and Topaz on a subset of micrographs in CryoPPP for the independent test proteins (~300 micrographs per protein)

Similarly, as in Section 3.1, we compared CryoTransformer, CrYOLO, and Topaz on the labeled subset of micrographs in CryoPPP for the six proteins in the independent test dataset in terms of the resolution of reconstructed density maps. The density maps were reconstructed using the *Select 2D* job from the picked particles. The 3D resolution is listed in **Table 3**.

Table 3: *Comparison of CryoTransformer with CrYOLO and Topaz's performance in terms of the resolution of 3D density maps reconstructed for six test proteins from the particles picked from a small set of micrographs in the CryoPPP*

EMPIAR ID	Number of Micrographs	Number of Particles			3D Resolution (Å)		
		CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer
10017 [59]	84	283	98,625	43,735	10.4	**5.3**	5.61
10081 [55]	300	17,550	111,752	88,632	9.78	7.81	**5.47**

10093 [57]	295	8,802	257,490	151,545	8.8	**6.06**	6.86
10345 [58]	295	4,095	93,699	105,739	10.2	8.12	**6.43**
10532 [56]	300	12,166	356,222	148,345	15.69	5.47	**3.92**
11056 [60]	305	46,582	144,600	98,193	10	8.34	**7.42**

Among the six datasets considered, CryoTransformer outperforms crYOLO and Topaz in four instances, despite picking a much smaller number of particles than Topaz in most cases. This observation underscores Topaz's tendency to pick more overlapped/duplicate particles or false positives. CrYOLO performs substantially worse than CryoTransformer and Topaz because it picks a much small number of particles, which are not sufficient to build good density maps. For the same four proteins, the best resolution of the density maps in **Table 3** is lower than that in **Table 2** because a much small number of micrographs were used for the particle picking and density map reconstruction.

Table 4: *Comparison of CryoTransformer with crYOLO and Topaz in terms of precision, recall, F1-score, and dice score of particle picking on the micrographs of six independent test proteins*

EMPIAR ID	Type of Protein	Number of Micrographs	# Particles in Ground Truth (CryoPPP dataset)	Precision			Recall			F1 Score			Dice Score		
				CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer	CrYOLO	Topaz	CryoTransformer
10017	β-galactosidase	84	49,391	0.695	0.57	**0.745**	0.024	**0.998**	0.587	0.046	**0.726**	0.657	0.041	**0.694**	0.623
10081	Transport	300	39,352	0.405	0.736	**0.860**	0.18	**0.965**	0.889	0.25	0.835	**0.874**	0.214	**0.825**	0.823
10093	Membrane	295	56,394	0.574	**0.617**	0.560	0.054	**0.537**	0.689	0.098	0.574	**0.618**	0.086	0.504	**0.600**
10345	Signaling	295	15,894	0.543	0.526	**0.744**	0.134	**0.981**	0.864	0.215	0.685	**0.799**	0.111	0.659	**0.684**
10532	Viral	300	87,933	0.715	0.672	**0.813**	0.201	**0.976**	0.665	0.313	**0.796**	0.732	0.239	**0.757**	0.614
11056	Transport	305	125,908	0.513	0.731	**0.853**	0.214	**0.832**	0.683	0.302	**0.778**	0.758	0.284	**0.692**	0.679
	Average			0.574	0.642	**0.7625**	0.1345	**0.8815**	0.7295	0.204	0.732	**0.740**	0.163	**0.689**	0.671

In addition to the evaluation based on 3D resolution and the number of picked particles, we also assessed the performance of the three methods using precision, recall, F1 score, and dice score, as detailed in **Table 4**.

Moreover, we compared the particles picked by each method with the ground truth particles labeled in CryoPPP in terms of four machine learning metrics including precision, recall, F1 score (the geometric mean of the precision and recall), and Dice score (**Table 4**).

CryoTransformer stands out with the highest average precision of 0.7625 and the highest average F1-score of 0.740, indicating that it excels in producing accurate positive predictions and achieves the best-balanced performance considering both precision and recall. Topaz has the highest average recall and dice score of 0.8815 and 0.671, highlighting its ability to correctly identify a high proportion of true positive particles and generate a strong overlap between predicted and actual positive instances.

3.3 Visual inspection of particles picked by CryoTransformer, CrYOLO, and Topaz

Figure 7 visualizes the particles picked by the three methods from the four representative micrographs of four proteins in the internal test data, which consist of 10% of micrographs from the 80%-10%-10% train-valid-test split (see detailed results in **Supplementary Table S4**). Consistent with the results in **Section 3.2**, CrYOLO tends to select few true protein particles, consequently missing many true positive across various protein types. In contrast, Topaz selects an excessive number of particles, often leading to overlaps and duplicates. Additionally, Topaz frequently picks false particles from contaminations, particle aggregates and ice patches, which can result in lower-resolution 3D density map reconstruction. Furthermore, picking a lot of redundant particles requires much more storage to store them and a lot of time and memory to reconstruct density maps from them. CryoTransformer, on the other hand, often picks most of true particles while keeping false positives to a low level, which is beneficial for 3D density map reconstruction.

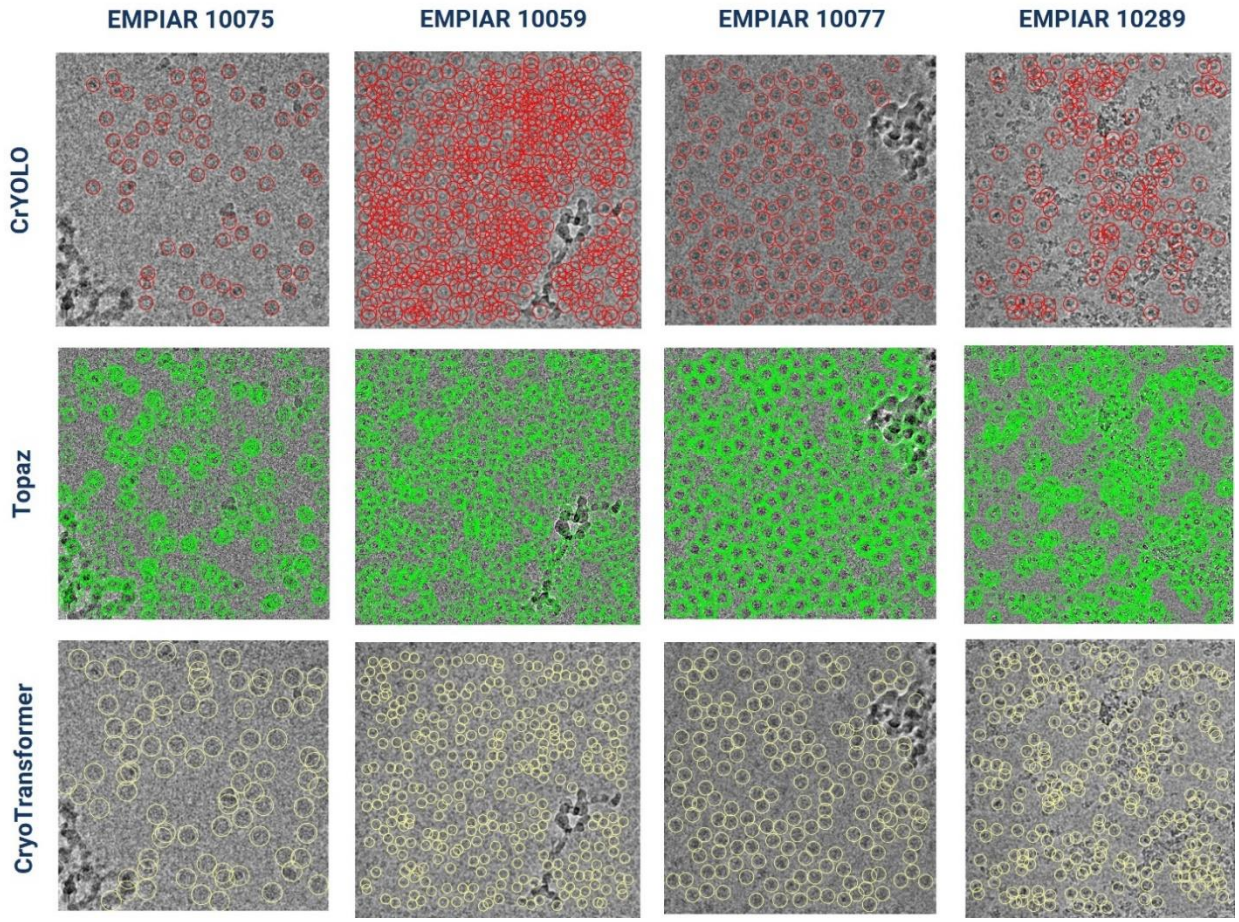

Figure 7: *Assessment of CrYOLO, Topaz, and CryoTransformer based on visual inspection of predicted particles in micrographs of four typical proteins. The first row (indicated by red circles) represents protein particles picked by CrYOLO. The second row (marked by green circles) displays protein particles picked by Topaz. The third row (with yellow circles) illustrates protein particles picked by CryoTransformer.*

Conclusion

We present CryoTransformer, a novel deep learning method for particle recognition and extraction. It leverages the power of transformers, residual networks, traditional image processing techniques, and a bipartite matching loss function. CryoTransformer was trained and tested on the largest labeled particle dataset available. According to the rigorous evaluations and comparisons, CryoTransformer achieves state-of-the-art performance, making it a robust AI-based tool to automate the process of picking protein particles from cryo-EM micrographs.

Code Availability

The source code and data are available at the GitHub repository: https://github.com/jianlin-cheng/CryoTransformer .

Acknowledgements

We thank the EMPIAR team for maintaining the Cryo-EM data archive and the scientists who contributed their cryo-EM micrographs to EMPIAR repository. We are grateful to Dr. Filiz Bunyak for her insights on handling complex microscopic images through computer vision perspectives.

Author Contributions

J.C. conceived and conceptualized this research; J.C. and L.W. provided guidance on the research and evaluation; A.D., R.G., and J.C. designed the methods; A.D. implemented the methods; A.D. performed the training and evaluations; A.D. and R.G. drafted the manuscript; J.C. and L.W. revised the manuscript; and all authors discussed the results and contributed to the final manuscript.

Competing Interests

The authors declare no competing interests.

Funding

This work was supported by a National Institutes of Health (NIH) grant (grant #: R01GM146340) to JC and LW.

References

[1] R. M. Glaeser, “Stroboscopic imaging of macromolecular complexes,” *Nat. Methods*, vol. 10, no. 6, pp. 475–476, 2013, doi: 10.1038/nmeth.2486.

[2] F. Boadu, H. Cao, and J. Cheng, “Combining protein sequences and structures with transformers and equivariant graph neural networks to predict protein function,” *Bioinformatics*, vol. 39, pp. I318–I325, 2023, doi: 10.1093/bioinformatics/btad208.

[3] A. Dhakal, C. McKay, J. J. Tanner, and J. Cheng, “Artificial intelligence in the prediction of protein-ligand interactions: recent advances and future directions,” *Briefings in Bioinformatics*, vol. 23, no. 1. 2022, doi: 10.1093/bib/bbab476.

[4] N. Giri and J. Cheng, "Improving Protein–Ligand Interaction Modeling with cryo-EM Data, Templates, and Deep Learning in 2021 Ligand Model Challenge," *Biomolecules*, vol. 13, no. 1, 2023, doi: 10.3390/biom13010132.

[5] A. Dhakal, R. Gyawali, and J. Cheng, "Predicting Protein-Ligand Binding Structure Using E(n) Equivariant Graph Neural Networks," *bioRxiv*, p. 2023.08.06.552202, 2023, [Online]. Available: http://biorxiv.org/content/early/2023/08/07/2023.08.06.552202.abstract.

[6] A. Al-Azzawi, A. Ouadou, J. J. Tanner, and J. Cheng, "AutoCryoPicker: an unsupervised learning approach for fully automated single particle picking in Cryo-EM images," *BMC Bioinformatics*, vol. 20, no. 1, p. 326, Dec. 2019, doi: 10.1186/s12859-019-2926-y.

[7] A. Al-Azzawi, A. Ouadou, H. Max, Y. Duan, J. J. Tanner, and J. Cheng, "DeepCryoPicker: fully automated deep neural network for single protein particle picking in cryo-EM," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–38, 2020, doi: 10.1186/s12859-020-03809-7.

[8] S. P. Mallick, Y. Zhu, and D. Kriegman, "Detecting particles in cryo-EM micrographs using learned features," *J. Struct. Biol.*, vol. 145, no. 1–2, pp. 52–62, 2004, doi: 10.1016/j.jsb.2003.11.005.

[9] R. Langlois, J. Pallesen, J. T. Ash, D. Nam Ho, J. L. Rubinstein, and J. Frank, "Automated particle picking for low-contrast macromolecules in cryo-electron microscopy," *J. Struct. Biol.*, vol. 186, no. 1, pp. 1–7, 2014, doi: 10.1016/j.jsb.2014.03.001.

[10] A. Heimowitz, J. Andén, and A. Singer, "APPLE picker: Automatic particle picking, a low-effort cryo-EM framework," *J. Struct. Biol.*, vol. 204, no. 2, pp. 215–227, 2018, doi: 10.1016/j.jsb.2018.08.012.

[11] R. Marabini *et al.*, "Xmipp: An image processing package for electron microscopy," *J. Struct. Biol.*, vol. 116, no. 1, pp. 237–240, 1996, doi: 10.1006/jsbi.1996.0036.

[12] F. Wang *et al.*, "DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM," *J. Struct. Biol.*, vol. 195, no. 3, pp. 325–336, 2016, doi: 10.1016/j.jsb.2016.07.006.

[13] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy," *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–10, 2017, doi: 10.1186/s12859-017-1757-y.

[14] Y. Xiao and G. Yang, "A fast method for particle picking in cryo-electron micrographs based on fast R-CNN," *AIP Conf. Proc.*, vol. 1836, no. June 2017, 2017, doi: 10.1063/1.4982020.

[15] D. Tegunov and P. Cramer, "Real-time cryo-electron microscopy data preprocessing with Warp," *Nat. Methods*, vol. 16, no. 11, pp. 1146–1152, 2019, doi: 10.1038/s41592-019-0580-y.

[16] A. Masoumzadeh and M. Brubaker, "HydraPicker: Fully automated particle picking in cryo-em by utilizing dataset bias in single shot detection," *30th Br. Mach. Vis. Conf. 2019, BMVC 2019*, 2020.

[17] D. M. McSweeney, S. M. McSweeney, and Q. Liu, "A self-supervised workflow for particle picking in cryo-EM," *IUCrJ*, vol. 7, pp. 719–727, 2020, doi: 10.1107/S2052252520007241.

[18] P. R. Baldwin and P. A. Penczek, "The Transform Class in SPARX and EMAN2," *J. Struct. Biol.*, vol. 157, no. 1, pp. 250–261, 2007, doi: 10.1016/j.jsb.2006.06.002.

[19] T. Wagner *et al.*, "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM," *Commun. Biol.*, vol. 2, no. 1, pp. 1–13, 2019, doi: 10.1038/s42003-019-0437-z.

[20] T. Bepler *et al.*, "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nat. Methods*, vol. 16, no. 11, pp. 1153–1160, 2019, doi: 10.1038/s41592-019-0575-8.

[21] A. Dhakal, R. Gyawali, L. Wang, and J. Cheng, "A large expert-curated cryo-EM image dataset for machine learning protein particle picking," *Sci. Data*, vol. 10, no. 1, pp. 1–22, 2023, doi: 10.1038/s41597-023-02280-2.

[22] A. Dhakal, R. Gyawali, L. Wang, and J. Cheng, "CryoPPP : A Large Expert-Labelled Cryo-EM Image Dataset for Machine Learning Protein Particle Picking Background & Summary I . Cryo-EM Grid Preparation and Image Acquisition II . Cryo-EM Micrographs and Single Particle Analysis," 2023.

[23] N. Carion, F. Massa, N. U. , Gabriel Synnaeve, and and S. Z. Alexander Kirillov, "End to End Object Detection Using Transformers," *Eccv*, vol. 11900, p. 26, 2020, [Online]. Available: https://ai.facebook.com/research/publications/end-to-end-object-detection-with-transformers.

[24] A. Iudin *et al.*, "EMPIAR: the Electron Microscopy Public Image Archive," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D1503–D1511, 2023, doi: 10.1093/nar/gkac1062.

[25] Y. Liu *et al.*, "Ligand recognition and allosteric modulation of the human MRGPRX1 receptor," *Nat. Chem. Biol.*, vol. 19, no. October, 2022, doi: 10.1038/s41589-022-01173-6.

[26] S. Tanaka *et al.*, "Structural Basis for Binding of Potassium-Competitive Acid Blockers to the Gastric Proton Pump," *J. Med. Chem.*, vol. 65, no. 11, pp. 7843–7853, 2022, doi: 10.1021/acs.jmedchem.2c00338.

[27] T. P. Newing *et al.*, "Molecular basis for RNA polymerase-dependent transcription complex recycling by the helicase-like motor protein HelD," *Nat. Commun.*, vol. 11, no. 1, pp. 1–11, 2020, doi: 10.1038/s41467-020-20157-5.

[28] C. Cao *et al.*, "Structure, function and pharmacology of human itch GPCRs," *Nature*, vol. 600, no. 7887, pp. 170–175, 2021, doi: 10.1038/s41586-021-04126-6.

[29] M. L. Oldham, N. Grigorieff, and J. Chen, "Structure of the transporter associated with antigen processing trapped by herpes simplex virus," *Elife*, vol. 5, no. DECEMBER2016, pp. 1–16, 2016, doi: 10.7554/eLife.21829.

[30] M. Kuzuya *et al.*, "Structures of human pannexin-1 in nanodiscs reveal gating mediated by dynamic movement of the N terminus and phospholipids," *Sci. Signal.*, vol. 15, no. 720, pp. 1–11, 2022, doi: 10.1126/scisignal.abg6941.

[31] J. Li *et al.*, "Cryo-EM structures of Escherichia coli cytochrome bo3 reveal bound phospholipids and ubiquinone-8 in a dynamic substrate binding site," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 34, 2021, doi: 10.1073/pnas.2106750118.

[32] T. M. Josephs *et al.*, "Structure and dynamics of the CGRP receptor in apo and peptide-bound forms," *Science (80-.).*, vol. 372, no. 6538, 2021, doi: 10.1126/SCIENCE.ABF7258.

[33] E. F. Pettersen *et al.*, "UCSF ChimeraX: Structure visualization for researchers, educators, and developers," *Protein Sci.*, vol. 30, no. 1, pp. 70–82, 2021, doi: 10.1002/pro.3943.

[34] Q. Li *et al.*, "Synthetic group A streptogramin antibiotics that overcome Vat resistance," *Nature*, vol. 586, no. 7827, pp. 145–150, 2020, doi: 10.1038/s41586-020-2761-3.

[35] K. Demura *et al.*, "Cryo-EM structures of calcium homeostasis modulator channels in diverse oligomeric assemblies," *Sci. Adv.*, vol. 6, no. 29, pp. 1–12, 2020, doi: 10.1126/sciadv.aba8105.

[36] D. Nicholson, T. A. Edwards, A. J. O'Neill, and N. A. Ranson, "Structure of the 70S Ribosome from the Human Pathogen Acinetobacter baumannii in Complex with Clinically Relevant Antibiotics," *Structure*, vol. 28, no. 10, pp. 1087-1100.e3, 2020, doi: 10.1016/j.str.2020.08.004.

[37] D. O. Passos *et al.*, "Structural basis for strand-transfer inhibitor binding to HIV intasomes," *Science (80-.).*, vol. 367, no. 6479, pp. 810–814, 2020, doi: 10.1126/science.aay8015.

[38] B. Burendei *et al.*, "Cryo-EM structures of undocked innexin-6 hemichannels in phospholipids," *Sci. Adv.*, vol. 6, no. 7, 2020, doi: 10.1126/sciadv.aax3157.

[39] M. E. Falzone *et al.*, "Structural basis of Ca2+-dependent activation and lipid transport by a TMEM16 scramblase," *Elife*, vol. 8, pp. 1–25, 2019, doi: 10.7554/elife.43229.

[40] L. Y. Kim *et al.*, "Benchmarking cryo-EM single particle analysis workflow," *Front. Mol. Biosci.*, vol. 5, no. JUN, 2018, doi: 10.3389/fmolb.2018.00050.

[41] Y. Zi Tan *et al.*, "Addressing preferred specimen orientation in single-particle cryo-EMthrough tilting," *Nat. Methods*, vol. 14, no. 8, pp. 793–796, 2017, doi: 10.1038/nmeth.4347.

[42] N. Fischer *et al.*, "The pathway to GTPase activation of elongation factor SelB on the ribosome," *Nature*, vol. 540, no. 7631, pp. 80–85, 2016, doi: 10.1038/nature20560.

[43] R. I. Koning *et al.*, "Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ," *Nat. Commun.*, vol. 7, pp. 1–6, 2016, doi: 10.1038/ncomms12524.

[44] Y. Gao, E. Cao, D. Julius, and Y. Cheng, "TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action," *Nature*, vol. 534, no. 7607, pp. 347–351, 2016, doi: 10.1038/nature17964.

[45] M. Liao, E. Cao, D. Julius, and Y. Cheng, "Structure of the TRPV1 ion channel determined by electron cryo-microscopy," *Nature*, vol. 504, no. 7478, pp. 107–112, 2013, doi: 10.1038/nature12822.

[46] T. Y. Lin *et al.*, "Microsoft COCO: Common objects in context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[48] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[49] N. Parmar *et al.*, "Image transformer," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 9, pp. 6453–6462, 2018.

[50] I. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens, "Attention augmented convolutional networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 3285–3294, 2019, doi: 10.1109/ICCV.2019.00338.

[51] R. Stewart, M. Andriluka, and A. Ng, "End-to-end people detection in crowded scenes," pp. 2325–2333, 2015, [Online]. Available: http://arxiv.org/abs/1506.04878.

[52] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 658–666, 2019, doi: 10.1109/CVPR.2019.00075.

[53] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.

[54] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination," *Nat. Methods*, vol. 14, no. 3, pp. 290–296, 2017, doi: 10.1038/nmeth.4169.

[55] C. H. Lee and R. MacKinnon, "Structures of the Human HCN1 Hyperpolarization-Activated Channel," *Cell*, vol. 168, no. 1–2, pp. 111-120.e11, 2017, doi: 10.1016/j.cell.2016.12.023.

[56] Y. Z. Tan and J. L. Rubinstein, "Through-grid wicking enables high-speed cryoEM specimen preparation," *Acta Crystallogr. Sect. D Struct. Biol.*, vol. 76, pp. 1092–1103, 2020, doi: 10.1107/S2059798320012474.

[57] P. Jin *et al.*, "Electron cryo-microscopy structure of the mechanotransduction channel NOMPC," *Nature*, vol. 547, no. 7661, pp. 118–122, 2017, doi: 10.1038/nature22981.

[58] M. G. Campbell *et al.*, "Cryo-EM Reveals Integrin-Mediated TGF- b Activation without Release from Latent TGF- b Article Cryo-EM Reveals Integrin-Mediated TGF- b Activation without Release from Latent TGF- b," *Cell*, vol. 180, no. 3, pp. 490-501.e16, 2020, doi: 10.1016/j.cell.2019.12.030.

[59] S. H. W. Scheres, "Semi-automated selection of cryo-EM particles in RELION-1.3," *J. Struct. Biol.*, vol. 189, no. 2, pp. 114–122, 2015, doi: 10.1016/j.jsb.2014.11.010.

[60] J. Asami *et al.*, "Structure of the bile acid transporter and HBV receptor NTCP," *Nature*, vol. 606, no. 7916, pp. 1021–1026, 2022, doi: 10.1038/s41586-022-04845-4.