Tech Science Press

Urdnet: A Cryo-EM Particle Automatic Picking Method

Jianquan Ouyang[1], Yue Zhang[1], Kun Fang[1,2,*], Tianming Liu[3] and Xiangyu Pan[2]

[1]School of Computer Science & School of Cyberspace Science, Xiangtan University, Xiangtan, 411105, China
[2]Hunan Meteorological Information Center, Hunan Meteorological Bureau, Changsha, Hunan, 410118, China
[3]Department of Computer Science, The University of Georgia, Athens, Georgia, USA
*Corresponding Author: Kun Fang. Email: k19890823@163.com
Received: 10 November 2021; Accepted: 06 January 2022

Abstract: Cryo-Electron Microscopy (Cryo-EM) images are characterized by the low signal-to-noise ratio, low contrast, serious background noise, more impurities, less data, difficult data labeling, simpler image semantics, and relatively fixed structure, while U-Net obtains low resolution when downsampling rate information to complete object category recognition, obtains high-resolution information during upsampling to complete precise segmentation and positioning, fills in the underlying information through skip connection to improve the accuracy of image segmentation, and has advantages in biological image processing like Cryo-EM image. This article proposes A U-Net based residual intensive neural network (Urdnet), which combines point-level and pixel-level tags, used to accurately and automatically locate particles from cryo-electron microscopy images, and solve the bottleneck that cryo-EM Single-particle biological macromolecule reconstruction requires tens of thousands of automatically picked particles. The 80S ribosome, HCN1 channel and TcdA1 toxin subunits, and other public protein datasets have been trained and tested on Urdnet. The experimental results show that Urdnet could reach the same excellent particle picking performances as the mainstream methods of RELION, DeepPicker, and acquire the 3D structure of picked particles with higher resolution.

Keywords: Deep learning; convolutional neural network; particle picking; cryo-electron microscopy; single-particle reconstruction

1 Introduction

Cryo-EM has become an essential structural biology technology. It freezes the sample and keeps it in the microscope at a low temperature. Subsequently, the highly coherent electrons, used as a light source, illuminate from above and are scattered by the sample and the nearby ice layer. The scatter signal is then imaged and recorded using a detector and a lens system. Finally, signal processing is performed to obtain the structure of the sample, which is a valuable means of understanding the mechanism of biochemical reactions. Just as the catalytic sites of some proteins are known through

structures, inhibitors can be designed to deactivate these proteins, and drugs can be screened with these proteins as targets.

The other two leading technologies are X-ray crystallography and nuclear magnetic resonance (NMR), which can analyze the structure of a biomacromolecule to understand its function. For decades, X-ray crystallography has been the dominant technology for obtaining high-resolution structures of biomacromolecule. In recent years, with the improvement of cryo-EM [1,2] and the latest technological advances in sample preparation, calculation, and instrumentation, the structural analysis of protein complexes has undergone a considerable breakthrough [3] that the resolution of the large protein 3D structure increased to 3 Å [4,5].

The cryo-EM image contains 2D projections of the particles at different angles. Due to the similarity of the sample density of the protein to its surrounding solution, and the limited electron dose used in data collection, the cryo-EM image has a low SNR, low contrast, uneven background intensity, and irregular internal grain texture. Besides, bad particles that do not meet the requirements for 3D reconstruction emerged, which is caused by image impurities such as frozen liquids, carbon film, and stacked or dissociated particles. To obtain a high-resolution 3D structure, tens of thousands of projection images are typically required. Thus, particle picking is the first step toward the 3D reconstruction of macromolecules, including the particle recognition in micrograph and localization of the particle region, rather than the noise region and impurity region.

Existing methods for particle picking often use low-resolution 2D particle templates for template matching, where templates are generated from the clustering projections of manually selected particles. The basic idea of it is that the micrograph region has high cross-correlations with popular particle templates [6]. RELION [7] is a cryo-EM structure determination software that uses a template matching framework to select particles [8]. In the software, the user manually picks approximately 1000 particles from a few micrographs. These picked particle images are two-dimensionally classified to produce a small number of template images for automatically selecting particles from all micrographs. One problem with the approach is that it is susceptible to reference dependency biases resulting in high error detection rates.

In addition to RELION, many tools include automatic or semi-automated particle picking steps, such as PICKER [9], EMAN2 [10], XMIPP [11], cryoSPARC [12], most of which are based on traditional computational vision algorithms, such as edge detection, feature recognition, and template matching aforementioned. These methods are not entirely suitable for processing cryo-EM images with poor contrast and low SNR, for they do not take full advantage of the inherent and unique particle features. Moreover, their performance degrades significantly as the quality of the microscopic image decreases. Consequently, it is significant to develop an efficient, fully-automated, template-free method for particle picking.

In cryo-EM image analysis, many machine learning algorithms are widely used from support vector machines to convolutional neural networks. In the past few years, deep learning has proliferated. It can be superior to lots of traditional algorithms in computer vision by generating hierarchical features from big data analysis with deep neural networks [13,14]. Besides, some deep learning applications are robust to low SNR images [15]. As the cryo-EM image set continues to grow, and the SNR of the microscopic image is still low, deep learning seems to be very suitable for processing cryo-EM data. Inspired by some successful applications, we seek to apply deep learning methods to cryo-EM particle picking.

In this paper, an improved U-Net-based residual dense neural convolutional network, Urdnet, is proposed to select the cryo-EM particles accurately and automatically. The method introduces a

learning method combining point-level and pixel-level labels, which significantly saves the time of manual labeling and improves the efficiency and accuracy of particle picking.

2 Related Work

DeepPicker [16], a fully automatic particle selection method based on deep learning, applied the convolutional neural network to the particle picking task for the first time. Before the advent of DeepPicker, semi-automatic solutions such as RELION and EMAN2 were used to pick particles. DeepPicker converts particle picking into an image classification problem; it crops the microscopic images through a sliding window and classifies the sub-images into particles or backgrounds. Considering the lack of training data, DeepPicker uses a novel cross-molecular training strategy to train networks. However, the disadvantage is the long processing time, and the average particle picking speed is 1.5 min/mrc (mrc is a format for cryo-EM images).

DeepEM [17] uses an eight-layer convolutional neural network to recognize single particles of noisy cryo-EM micrographs and to achieve automated particle picking, selection, and verification in an integrated manner. DeepEM increases particle images through image rotation to augment the training set, but still requires manually selecting thousands of particles to train the data.

FastParticlePicker [18] is based on the object detection algorithm, Fast R-CNN, which includes a "region-of-interest proposal" network and a classification network. However, the FastParticlePicker crops the microscopic image with a sliding window instead of selecting the area of interest in the microscopic image. Therefore, its performance mainly depends on the classification network. The classification network has three types of objects, particles, ice, and background, which reduces the false-positive results caused by ice.

FCRN [19] is also an automatic particle picking method based on deep learning. It proposes a Fully Convolution Regression Network (FCRN) mapping particle images to continuous distance maps to recognize particles from different data sets. Experimental results on EMPIAR data show that FCRN achieves better particle picking performance than Faster-RCNN and RELION.

crYOLO [20] is based on the YOLO9000 algorithm to achieve particle detection. For small particles, crYOLO achieves higher precision than the original YOLO network. It achieves a processing speed of up to 6 mrcs/s and can be extended to other biomacromolecules outside the training set. Despite the modifications to the original YOLO model, the experiment did not mention how crYOLO detects particles of different sizes and aspect ratios.

Topaz [21] is an efficient and accurate particle selection method based on deep learning. Unlike other methods, it uses a positive unlabeled (PU) learning framework to reuse the remaining unlabeled particles to train neural networks of a small number of labeled particles. Even for the challenging dataset of rod-like particles with low SNR, the experimental results are superior to the general PU learning method.

PIXER [22] is an automatic particle picking method based on image segmentation using deep neural networks. To adapt to the low SNR, it uses a segmentation network to convert the microscopic image into a probability density map to detect the particles. A grid-based local maximum method is proposed to locate the particles from the probability density map. Compared to mainstream methods, PIXER can achieve as good results as the semi-automatic methods RELION and DeepEM.

Compared with natural images, cryo-EM images have the following characteristics: (1) The semantics of the images are relatively simple, and the structure is fixed. Therefore, high-level semantic information and low-level features are both important. U-Net's skip connection and U-shaped

structure are quite suitable to deal with our scenario. (2) Few sources of data and the difficulty of image labeling lead us to design a model that should not be too large. If parameters of the model are excessive, it will easily lead to overfitting. (3) Compared with natural images, cryo-EM images have a higher signal-to-noise ratio (SNR), higher complexity, a larger grayscale, and more unclear boundaries. U-Net obtains low-resolution information during down-sampling to recognize different objects and obtains high-resolution information during up-sampling to achieve segmentation and positioning of objects. In the meantime, it fills in the underlying information through skip-connections to improve segmentation accuracy. Simple CNN cannot achieve the global feature fusion of U-Net, that is, it cannot connect shallow features and deep features. Therefore, direct training of CNN is often suitable for a large number of datasets, while U-Net with a U-shaped network uses skip-connection to reduce the time of feature extraction so that the entire network can better remember the overall picture information. In addition, the goal of this article is not only to use CNN feature extraction to detect cryo-EM particles but to obtain high-quality particle images for the three-dimensional reconstruction of biological macromolecules. In addition to identifying particles, it is necessary to accurately segment artifacts to avoid picking out defective low-mass particles. However, the efficiency of U-Net in biomedical image segmentation is well known [23].

For the shortcomings of the above methods, Combine the advantages of U-Net, we developed an accurate and fully automatic particle picking approach, Urdnet, based on deep learning. Among it, the pertinent data preprocessing raised the quality of the raw Cryo-EM images. The image annotation method that combines with multiple labels significantly reduced time and effort for manually building a training dataset. Urdnet can thoroughly learn the features of particles with different particle sizes and aspect ratios via a small amount of training data. The connected component analysis is performed on the pixel map predicted by Urdnet to locate the candidate particles, which avoided selecting most of the false-positive particles. Moreover, experimental results show our method can apply to multiple types of micrographs from different cryo-EM detectors and has superior particle detection precision compared to the DeepPicker method [24–26]. Compared with the semi-automatic picking of RELION, our method resulted in a 3D single-particle structure with a higher resolution. More details of our proposed method will be introduced next.

3 Proposed Method

3.1 Data Preprocessing

3.1.1 Intensity Adjustment

Particle picking is significantly dependent on the intensity of the grayscale cryo-EM image. Due to the low-dose electron microscopy imaging mode, the randomly distributed particle and thin ice in high defocus areas are exposed to extremely the low-intensity beam, and low-contrast cryo-EM images are collected in that imaging mode. To alleviate this problem, we applied a contrast enhancement method to adjust the global image intensity and improve the SNR of cryo-EM images [27,28]. The steps are as follows: (1) calculate two segmentation thresholds x_l and x_h according to a percentage of pixel saturation of 1%, respectively indicating that the number of pixels less than x_l and more than x_h both accounts for 1%; (2) the gray value in $[x_l, x_h]$ of the original image is linearly mapped to [0, 1], as shown in Eq. (1), where x is a pixel value of the original image, and x_f is the new pixel value after the mapping.

$$x_f = \frac{1}{x_h - x_l}(x - x_l) \tag{1}$$

Fig. 1 shows the raw cryo-EM image, the intensity-adjusted image, and their corresponding histograms. (c) has higher contrast than (a), and correspondingly, (d) has a wider range of pixel distribution than (b).
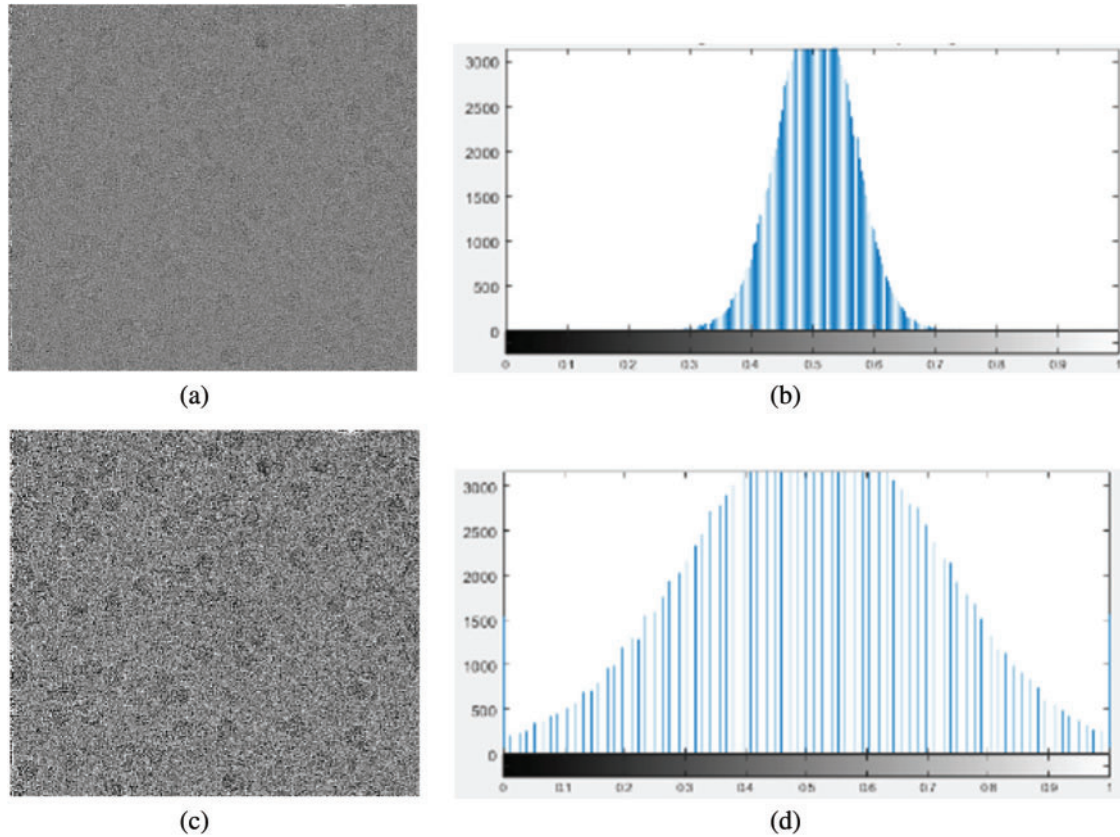

Figure 1: (a) is a cryo-EM image of the TcdA1 toxin subunit; (b) is the histogram of (a); (c) is the intensity-adjusted image of (a); (d) is the histogram of (c)

3.1.2 Noise Suppressing

Cryo-EM images tend to be so noisy and blurred that proteins and solvents are visually similar since low-electron dose imaging only produces a small number of available signals of the taken images. Hence, we apply image restoration to improve image quality. For the Gaussian noise in the cryo-EM image, we choose the Wiener filter [29] to reduce the noise of the blur area and improve the sharpness of the defocused phase-plate cryo-EM image. Wiener filtering is a method of filtering the noise-mixed signal by minimizing the overall mean square error at the inverse filtering and noise smoothing. Its mathematical expression is as in Eq. (2) Where H(f) is the Fourier transform of h in the frequency domain f, SNR(f) is the SNR, and $*$ represents the complex conjugate.

$$G(f) = \frac{H*(f)}{|H(f)|^3 + \frac{1}{SNR(f)}} \tag{2}$$

Fig. 2 shows the intensity adjusted cryo-EM image in Section 3.1.1 and its Wiener filtered image, where PSNR stands for Peak Signal-to-Noise Ratio, which is a widely used index for evaluating image

quality and signal restoration. The index is calculated by the mean square error (MSE). The MSE is calculated by Eq. (3). I and K are images of size m * n, which represent respectively the raw cryo-EM image and the processed cryo-EM image. The peak signal-to-noise ratio is defined as Eq. (4), where MAX_I represents the maximum pixel value of the image.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - K(i,j)^2\| \tag{3}$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{4}$$

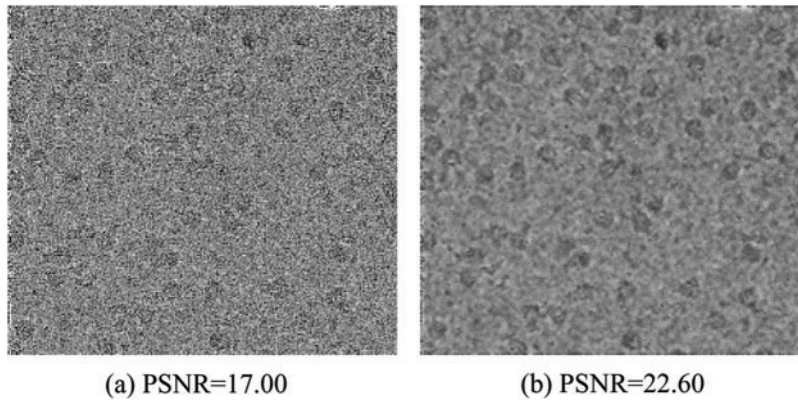(a) PSNR=17.00   (b) PSNR=22.60

Figure 2: (a) is a cryo-EM image of the TcdA1 toxin subunit (EMPIAR-10089); (b) is Wiener filtered image of (a)

As shown in Fig. 2, the PSNRs of the intensity-adjusted image aforementioned, and its noise suppressed image were calculated. Comparing the two PSNR values, 22.60 of Fig. 2a is higher than 17.00 of Fig. 2b, which indicates that the Wiener filter has a direct noise suppression effect on the cryo-EM images. Compared with Figs. 2a and 2b makes it easier to distinguish the particle boundaries.

3.2 Combination of Point Annotation and Per-Pixel Annotation

Various weak annotation forms have been explored in weakly supervised annotation methods for semantic segmentation of natural scene images, such as image-level labels (like classification labels) [30–32], extreme points [33], graffiti [34] and bounding boxes (like detection labels) [35]. The point form is the most straightforward and simplest of all weak annotation forms [36]. Since a cryo-EM image often contains a large number of particles, we proposed an annotation method of point labeling of particle center for cryo-EM particle picking to reduce the burden of manual annotation greatly. We divide objects into two classes: particles and artifacts. Artifacts are usually areas of cryo-EM impurities. We annotate the particles as points so that they are only associated with several pixels. The number of artifacts is much smaller than particles, and artifacts are particularly large in the cryo-EM images. To prevent the bad particles in the artifact from being selected, the artifacts are labeled at the pixel level like fully supervised segmentation labels, and the remaining unlabeled pixels are automatically classified as the background. Hence, we combine different forms of annotation to label particles and artifacts in a cryo-electron image. Fig. 3b shows the label image of an 80S ribosome cryo-EM image.
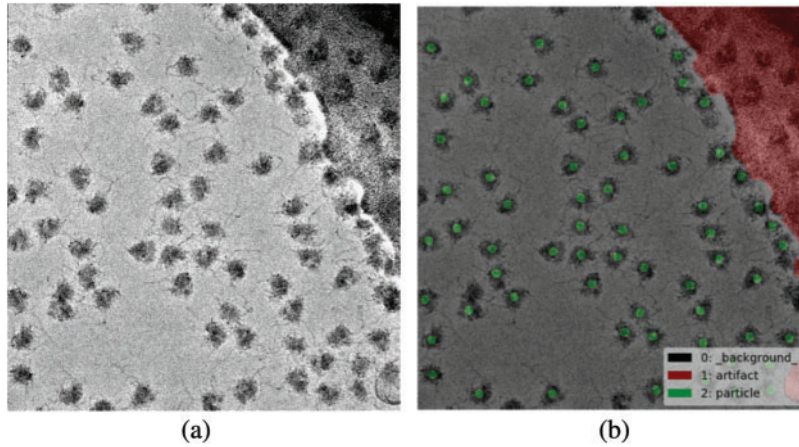
Figure 3: (a) is a cryo-EM image of the 80S ribosome; (b) is a label image of (a), the red region is the artifact's label, and the green points are the particles' labels

3.3 Convolutional Neural Network Design

U-Net [37] has achieved excellent results in medical image segmentation. Its most significant advantage is that it combines shallow and in-depth features and can reach high segmentation precision with a small training dataset. In the particle picking task, the precision of artifact segmentation affects the quality of picked particles. Imprecise artifact segmentation can easily lead to selecting false-positive particles. To improve the precision of artifact segmentation, we embedded the residual dense block (RDB) [38] in the overall model of improved U-Net, as shown in Fig. 4 when designing the Urdnet model. As shown in Fig. 5, we add two convolution layers as the input port to the basic RDB structure for local shallow feature extraction. RDB mainly integrates the residual block (Eq. (5)) [39] and the dense block (Eq. (6)) [40]. The residual block short circuits the output layer with the previous input layer by adding the feature vectors. The dense block concatenates each middle layer with all previous layers in the channel dimension to implement feature multiplexing. Combining the advantages of both, RDB performs not only local residual learning but also strengthens local feature fusion. Accordingly, it can effectively alleviate gradient vanishing, enhance feature propagation and reduce the number of parameters. RDB was proposed for processing image restoration, image super-resolution, and denoising tasks at first. Consequently, RDB is well suited for processing low-SNR and low-resolution cryo-EM images. To the best of our knowledge, our work is the first to combine U-Net with RDB to solve cryo-EM particle picking tasks.

$$x^l = f(x_{l-1}) + x^{l-1} \tag{5}$$

As shown in formula (5), the output x^l of the residual block is defined as the addition of $f(x_{l-1})$, the output of x_{l-1} after multiple convolutions and activations and x_{l-1}, the output of the $l-1$ layer.

$$x^l = H^l([x^1, x^2 \dots x^{l-1}]) \tag{6}$$

As shown in formula (6), the lth layer receives the feature matrix $x^1, x^2 \dots x^{l-1}$ from all previous layers as input, and $[x^1, x^2 \dots x^{l-1}]$ is the connection feature vector of all previous layers, $H^l(.)$ represents the amount of a single vector after convolution, activation, and other operations.
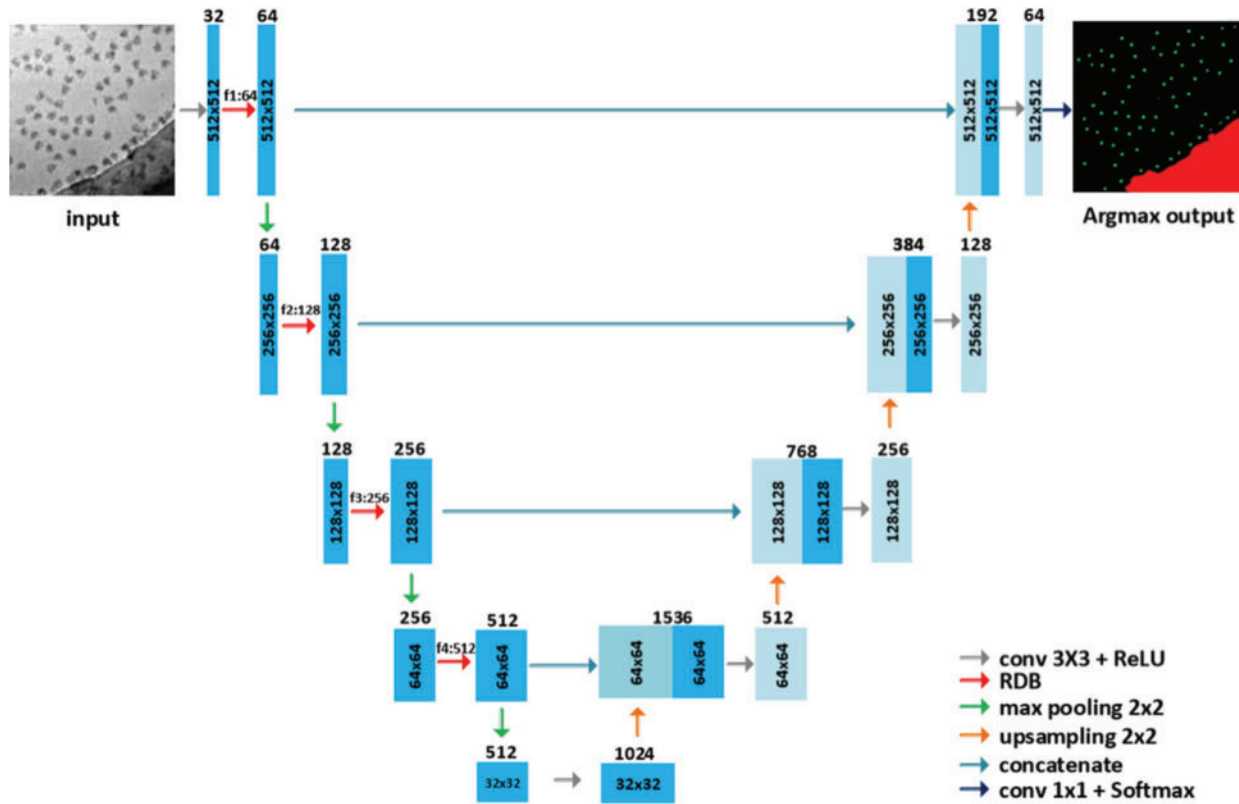
Figure 4: Structure of the Urdnet. The hyperparameter f1:64 indicates the vector channels number of the input and output of the RDB is 64, f2, f3, and f4 are the same. Different colored arrows indicate different operations on the feature vector, and the numbers above the blue vectors indicate the number of channels
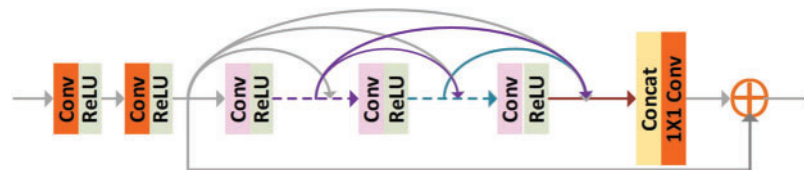

Figure 5: RDB of Urdnet

The network of Fig. 4 consists of 49 convolution layers, four max-pooling layers, four up-sampling layers, and four feature concatenations. In the encoder, i.e., feature extraction part, the first convolution layer learns the shallow global features, then followed by four RDBs for deep feature fusion. Each RDB consists of 8 standard convolution layers and one 1∗1 convolution layer. In an RDB, the number of filters in the first two convolution layers and the last 1∗1 convolution layer is determined by the external hyperparameters, such as 128 for the second RDB, and the number of filters of the remaining six convolution layers is fixed at 64, and the dashed arrows in Fig. 5 refer to the six standard convolution layers. The decoder combines the global context features with four times of upsampling and four feature concatenations, and finally restores the feature vector to the input vector size. We input a 512 × 512 2D cryo-EM image into Urdnet and apply three 1∗1 convolution filters to output

the feature map onto three channels, to obtain a softmax function (Eq. (7)) score map of $512 \times 512 \times 3$ where 3 indicates the number of object classes. While testing, the Arg max function gets the class of the highest score at each coordinate of the score map [40] and finally outputs the predicted pixel map (background: black, grain: green, artifact: red).

$$f(x^l)_i = \frac{e^{x^l_i}}{\sum_{j=1}^{C} e^{x^l_j}} \tag{7}$$

The softmax function is shown in formula (7), x^l is the output feature map of the lth network layer, e is the natural constant, $e^{x^l_i}$ represents the output value on the ith channel, C is the number of network categories, and the softmax score vector for each pixel is calculated by the formula (7).

3.4 Particle Picking Via Connected Component Analysis

We input the micrographs into the trained Urdnet model to output the prediction results of particles and artifacts. The prediction map is then binarized as shown in Figs. 6a and 6b, and the foreground data only contains pixels labeled as "particles," i.e., green pixels of Fig. 6a. Next, the foreground data is analyzed as the connected component [41], and the positions of the centroids of each connected component are taken as the particle positions. Using the connected component analysis, we can reduce interference from false-positive particles and locate the particles more accurately where the particles are close to each other. To weaken the impact caused by false positives, we remove the particles whose connected components are smaller than 30% of the mean component area. As shown in case 1 of Fig. 6, the dissociated particle in the red circle in Fig. 6c will not be picked for its small connected components. Where the bounding boxes of two particles overlap in a range of 30% area or more, the particles are considered to be stacked particles. In case 2 of Fig. 6, we reject the two particles in the red circles, since the overlapping area of them exceeds the prescribed threshold.
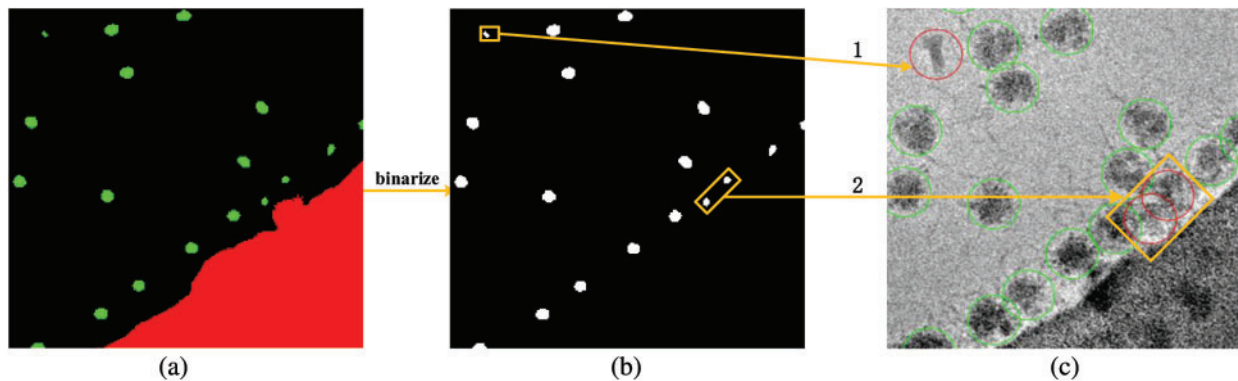

Figure 6: Particle picking workflow. (a) represents the prediction map, (b) is the corresponding binary image of (a), and (c) represents the particle picking result, where the green circle indicates the selected particles and the red circle indicates the rejected particles

4 Results and Discussion

4.1 Datasets

EMPIAR, the Electron Microscopy Public Image Archive, is a public resource for raw 2D electron microscopy images. Our experimental data includes cryo-EM images of three proteins from

EMPIAR, 80S ribosomes [42] (EMPIAR-10153), HCN1 channel [43] (EMPIAR-10081), and TcdA1 toxin subunit [44] (EMPIAR-10089). The three protein datasets from EMPIAR are new datasets released in the last three years. Besides, we test the common benchmark dataset, KLH (Keyhole limpet hemocyanin) [45] collected in 2004 to validate the performance of our automatic particle picking algorithm. Since structural analyses of viruses and large, well-ordered molecules with high point-group symmetry have been well known, and particle picking of them is relatively simple tasks, our experimental data mainly focuses on the new three datasets of small and medium protein molecules with no or low symmetry.

Human 80S ribosomes are used to synthesize proteins in cells and have complex molecular structures. The diameter of the 80S ribosomes is between 25 nm and 30 (250–300 Å), and its molecular mass is 3.9–4.5 MDa. The micrographs of EMPIAR-10153 were collected with low defocus and volta phase plate (VPP) [46]. VPP improves the phase contrast and SNR in the low-frequency range by introducing an additional phase shift in the unscattered beam. However, VPP not only enhances the contrast of the particles of interest but also the contrast of all weak phase objects, including contamination, ice dregs, carbon film, which increases the difficulty of accurately selecting the particles.

HCN1 channel is a hyperpolarization-activated cyclic nucleotide-gated ion channel that underlies the control of rhythmic activity in cardiac and neuronal pacemaker cells. Moreover, it forms a structure of the channel tetramer in a ligand-free state with a molecular weight of $\sim$74.6 kDa. Due to its small size, the number and quality of picked particles will affect the final resolution of 3D reconstruction.

KLH, a highly immunogenic protein macromolecule, is used as a carrier protein for the preparation of immunogens, which has a molecular weight of 7.9 Mda and a D5 point-group symmetry with a size of $\sim$40 nm. There are two main types of projection views of the KLH particles, the side view, and the top view. We select both views when picking particles. In addition to the particles of both views, the micrographs consist of clearly overlapping KLH particles and broken particles that make it difficult to extract unbroken single KLH particles.

Each protein dataset has different parameters such as electronic dose, defocus value, pixel size, and particle size. The main difference between the four datasets is that they have different defocus ranges and electronic dose. The SNRs of the 80S ribosome and KLH are higher for their lowest or none electron dose and low defocus value, and those of the TcdA1 and HCN1 channel are quite low for their higher electron dose and broad defocus range. The specific parameters of these data sets are listed in Tab. 1.

Table 1: Experimental datasets

Sample	Detector	Voltage (kV)	Electronic dose (e/Å∗∗2)	Defocus (µm)	Pixel size (Å)	Particle size (pix)
80S ribosome (EMPIAR-10153)	GATAN K2 SUMMIT (4k × 4k)	300	1.03	0.5	1.09	420

(Continued)

Table 1: Continued

Sample	Detector	Voltage (kV)	Electronic dose (e/Å**2)	Defocus (μm)	Pixel size (Å)	Particle size (pix)
HCN1 channel (EMPIAR-10081)	GATAN K2 SUMMIT (4k × 4k)	300	1.26	1.5~3.3	1.30	256
TcdA1 toxin subunit (EMPIAR-10089)	FEI FALCON II (4k × 4k)	300	2.5	1.3~4.5	1.14	352
KLH	CCD Tietz camera (2k × 2k)	120	—	0.6~1.5	2.22	272

4.2 Model Training

Our experimental hardware for model training is equipped with NVIDIA GeForce GTX 1080 Ti graphics, 64 GB RAM, and Intel Core i7 8700 K CPU. We trained the original U-Net and Urdnet models on the EMPIAR datasets of three proteins using the deep learning library Keras based on the Tensorflow backend. The data used for training and testing included 71 human 80S ribosomal micrographs, 30 HCN1 micrographs, and 24 TcdA1 micrographs, and each protein contains approximately 2,000–4,000 particles. We set 20% of the training set for validation. To reduce the risk of overfitting training, we augmented the training images by horizontal flip, rotation, width shift, height change, cropping, zooming, and filling, expanding one image to 32 images. The batch size was set to 2, the model optimizer is Adam [47], and the loss function is cross-entropy. The initial learning rate was set to 1E-4; as training progresses, the learning rate gradually decreases. To obtain a generic model for particles of different scales and aspect ratios, we added 16 KLH micrographs to the training dataset and separately trained Urdnet with the raw data and preprocessed data of the four proteins. The loss and accuracy of all trained models are shown in Tab. 2.

Table 2: Model training results

Data	Network	Loss	Accuracy
80S ribosome	U-Net	0.0929	0.8647
	Urdnet	0.0126	0.9570
HCN1 channel	U-Net	0.1324	0.8473
	Urdnet	0.0247	0.9336
TcdA1 toxin subunit	U-Net	0.1428	0.8070
	Urdnet	0.0521	0.9166
Four proteins (raw)	U-Net	0.1729	0.7822
Four proteins (preprocessed)	Urdnet	0.0832	0.9080

After being trained 40 epochs, all Urdnet models' losses are less than 0.1, which are lower than those of U-Net, and their accuracies are higher than those of U-Net. It sufficiently proves that the Urdnet model has better classification accuracy than the U-Net model when processing the same protein data. Furthermore, the training loss of the Urdnet generic model is shown in Fig. 7. The loss of preprocessed images is always lower than the raw images, indicating the importance of image preprocessing in Section 3.1. For the preprocessed data, the Urdnet generic model has a model loss of less than 0.1 and an accuracy of higher than 0.9, as shown in Tab. 2, which indicates that the Urdnet generic model also has excellent performance when dealing with different biomacromolecules. To a certain extent, it reveals the possibility of generalizing the Urdnet generic model to unknown biomacromolecules if more protein data of different particle shapes and sizes are trained.
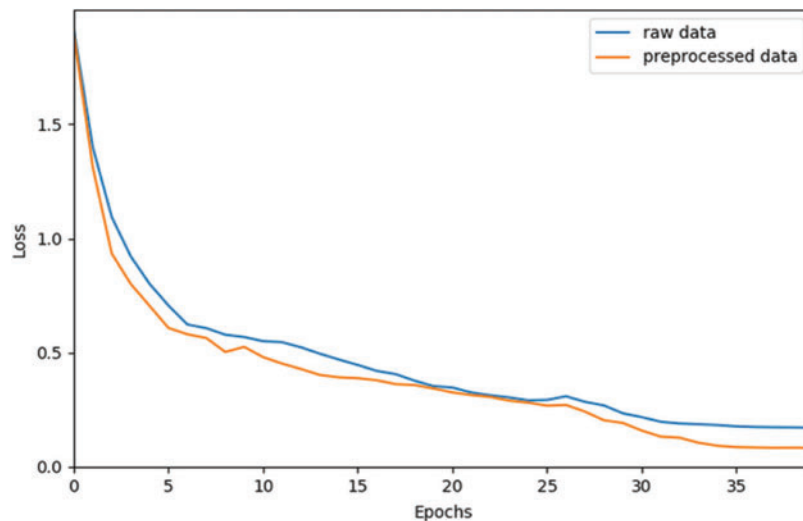
Figure 7: The training loss of raw micrographs and preprocessed micrographs

4.3 Particle Picking

To evaluate the particle picking performance of Urdnet, we calculated the precision, recall, precision-recall curve, and IoU (Intersection over Union) for the three proteins' test data, and compared them with the DeepPicker method. The results showed in Fig. 7 and Tab. 3. True Positive means a positive sample predicted by the model as positive, and True Negative means a negative sample predicted as negative. False Positive means a negative sample predicted to be positive, and False Negative means a positive sample predicted as negative. Precision is the proportion of the samples that are correctly predicted to be positive in all predicted positive samples (Eq. (8)). The recall is the proportion of the correctly predicted positive samples in all real positive samples (Eq. (9)), which indicates the ability of Urdnet to detect positive. Different precisions and recalls make up the precision-recall curve by changing the threshold of the particle prediction. As the threshold increases, the accuracy also increases, and the recall rate decreases. For the excellent performance of particle picking, we have to pick out the threshold with the highest recall before the precision drops sharply. We regard manually selected particles as Ground Truth. The mean IoU represents the accuracy of

particle position, which defined the ratio of the intersection area of ground truth and the testing result with their united area (Eq. (10)).

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \qquad (8)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \qquad (9)$$

$$\text{IoU} = \frac{\text{Testing Result} \cap \text{Ground Truth}}{\text{Testing Result} \cup \text{Ground Truth}} \qquad (10)$$

Table 3: Precision, recall, and IOU

Data	Method	Precision	Recall	IoU
80S ribosome	DeepPicker	0.8424	0.9313	0.8740
	Urdnet	0.8944	0.9587	0.9073
HCN1 channel	DeepPicker	0.7638	0.8445	0.7970
	Urdnet	0.8036	0.8878	0.8371
TcdA1 toxin subunit	DeepPicker	0.7187	0.8138	0.7713
	Urdnet	0.7524	0.8386	0.8112

As shown in Fig. 8, we find out Urdnet outperformed the DeepPicker at 80S ribosomes, HCN1 channels, and TcdA1 toxin subunits datasets. The 80S ribosome performs best on precision, recall, and IoU for its highest SNR. The precisions of the three datasets are higher than 0.75, the recalls of them are higher than 0.83, and the IoUs of them are higher than 0.81, which fully demonstrates the high performance of our method in particle picking.
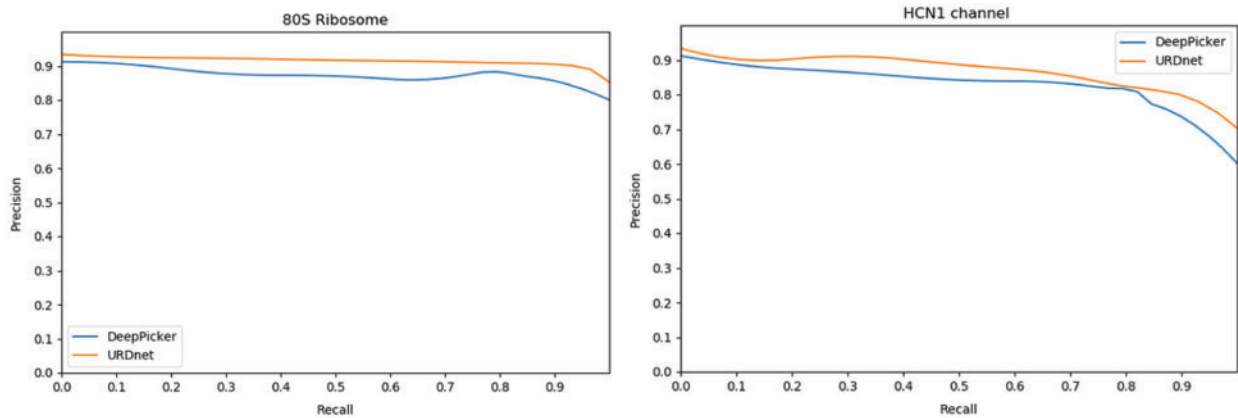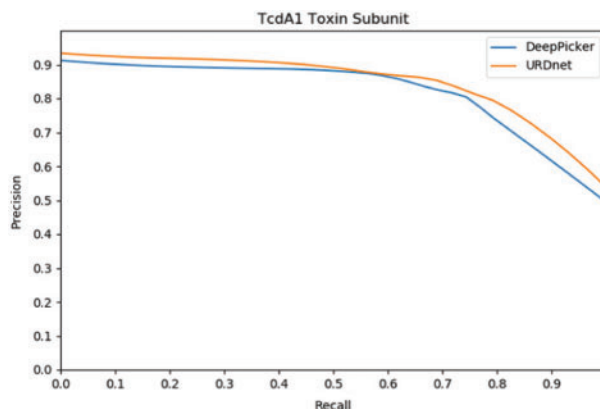

Figure 8: (Continued)

Figure 8: The precision-recall curve of three protein EMPIAR datasets

To further evaluate the quality of the particles selected by Urdnet, we compared our method to the semi-automatic selection method of RELION via the entire public datasets. 80S ribosomes (EMPIAR-10153) contain 318 micrographs, HCN1 channel (EMPIAR-10081) is the largest dataset, including 997 micrographs, TcdA1 toxin subunit (EMPIAR-10089) contains 97 micrographs, and KLH consists of 82 micrographs from the US National Resource for Automated Molecular Microscopy (NRAMM). 2D classification in the RELION software worked on the picked particles to identify suitable 2D average classes. Then, 2D average classes were further 3D classified using a low-resolution 3d map as the initial model, and good templates after 3D classification were kept and refined to construct the final 3D structure. We recorded the number of picked particles by the two methods, the "good" particle number related to the 2D class average template and the final resolution of the 3D reconstruction computed by the "gold standard" Fourier shell correlation [48]. The results are listed in Tab. 4.

Table 4: Particle auto-picking and 3D reconstruction results

Data	Method	Particles #	Particles # after 2D-class	Resolution (Å)
80S ribosome	RELION	20572	15175	3.92
	Urdnet	17612	15219	3.85
HCN1 channel	RELION	97921	55815	3.54
	Urdnet	71392	54126	3.51
TcdA1 toxin subunit	RELION	12834	10267	3.53
	Urdnet	11449	10190	3.49
KLH	RELION	1590	1160	21.10
	Urdnet	1517	1014	21.65

As shown in Tab. 4, the particles' number of 80S ribosomes, HCN1 channels, and TcdA1 extracted by RELION was approximately 16.8%, 37.1%, and 12.1% more than those of Urdnet. Due to the small number of the three protein particles used to train Urdnet, it picked fewer particles than RELION at the start. Nonetheless, the number of "good" particles after 2D classification is close to that of particles picked by RELION, i.e., the particles picked by two methods ultimately used for 3D reconstruction

are quite close. It suggests that Urdnet can identify more true-positive particles and pick fewer false-positive particles, which makes the final resolution of 3D reconstruction is slightly higher. We think that the Urdnet, different from RELION's semi auto-pick job by using the manually selected templates, picked particles with broader angular coverage, which reflects that Urdnet can choose particles with more angles and views. Especially for processing low-SNR and low-contrast datasets such as the HCN1 channel and TcDA1, Urdnet can thoroughly learn the inherent and unique particle features and avoid the dependency of users' intervention. Besides, Urdnet can significantly improve the accuracy of the picked particle. Since artifacts are labeled at the pixel level as a separate class, Urdnet avoids selecting "bad" particles in large-area artifacts, substantially eliminating the infuences of ice dregs, carbon films, and background noise. The gap between the reconstruction resolutions acquired by Urdnet and RELION is small, which is probably because the reconstructions reached resolutions close to those proteins' theoretical resolution limit. Nevertheless, the particles picked by Urdnet achieved higher resolutions of 3D reconstruction of the three protein data than RELION.

For KLH benchmark data, we find that the particles selected by RELION's semi auto-picking and the particles after 2D classification are slightly more than Urdnet, and its resolution of the final 3d reconstruction is higher than Urdnet. On the one hand, it is because a small number of KLH particles are used in the experiment. The dataset contains 82 micrographs, and each micrograph only has an average of about 25 particles, which results in little difference between the particles selected by the two methods. On the other hand, in recent years, significant changes have taken place in the imaging system of cryo-EM. Micrographs collected today (after 2012) are quite different from the older ones. Old datasets like KLH usually have higher pixel size, higher SNR, and the molecular weights of their particles are often enormous. Traditional methods, such as template matching, can easily detect these large particles. However, as the quality of microscopic images decreases, their performance dramatically decreases. Therefore, there is usually no one method to invariably get the best performance in different data sets. We think our method is quite effective when the image quality is mainly limited by low SNR, or the training data is insufficient.

4.4 Discussion

This paper proposes a residual dense convolutional network model based on multiple annotations and improved U-Net for automatic particle picking of cryo-EM biomacromolecules. It can automatically and accurately select particles from cryo-EM images and can reconstruct high-resolution 3D structures of particles from different proteins. But meanwhile, there is still room for improvement in our solution:

(1) Compared with other methods, it has little advantage in the resolution of 3D reconstruction and is not highly applicable to proteins of different shapes. It is necessary to improve the network model by increasing the types of protein training data.
(2) The image restoration performance of the residual intensive module is not further analyzed. Setting up experiments to calculate image noise indicators is requisite to evaluate whether the network has additional denoising effects after adding the RDB module.

Future work will focus on improving the performance of the Urdnet general model to achieve the goal of automatically picking a variety of challenging biological macromolecule cryo-EM image particles without retraining the model, deepening the research on image restoration issues to strive to eliminate data pre-processing steps and enabling the network to autonomously achieve image denoising to facilitate particle feature extraction.

5 Conclusion

In this paper, we proposed an automatic particle picking method, Urdnet, based on the U-Net architecture and residual dense block. We introduced a method of combining multiple forms of labels to build training data, which significantly releases the burden of manual labeling and improves annotation efficiency. Urdnet demonstrates its excellent performance on the public data of 80S ribosomal, HCN1 channel, TcdA1 toxin subunit, and KLH that it can effortlessly process cryo-EM particles from different proteins. Compared with the DeepPicker method, Urdnet achieves higher precision, recall, and IoU in particle picking. Compared with RELION's semi-automatic selection, our method can achieve higher resolutions of 3D reconstruction at most data. In future work, we will focus on improving the performance of the Urdnet generic model to pick new protein particles of multiple morphologies without the need to retrain the model.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Singer and F. J. Sigworth, "Computational methods for single-particle cryo-EM," *Computational Physics*, vol. 34, no. 6, pp. 11–20, 2020.

[2] H. Gupta, M. T. Mccann, L. Donati and M. Unser, "CryoGAN: A new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning," *IEEE Transactions on Computational Imaging*, vol. 3, no. 6, pp. 12–30, 2020.

[3] H. Jonas, S. Yashar and C. W. Muller, "The cryo-EM resolution revolution and transcription complexes," *Current Opinion in Structural Biology*, vol. 7, no. 9, pp. 8–15, 2018.

[4] M. A. Herzik, M. Wu and G. C. Lander, "Achieving better-than-3-A resolution by single-particle cryo-EM at 200 keV," *Nature Methods*, vol. 14, no. 11, pp. 1075–1078, 2017.

[5] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu *et al.,* "2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015.

[6] D. Lyumkis, "Challenges and opportunities in cryo-EM single-particle analysis," *Journal of Biological Chemistry*, vol. 7, no. 9, pp. 294, 2019.

[7] S. H. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.

[8] S. H. Scheres, "Semi-automated selection of cryo-EM particles in RELION-1.3," *Journal of Structural Biology*, vol. 189, no. 2, pp. 114–122, 2015.

[9] F. Zhang, Y. Chen, F. Ren, X. Wang, Z. Liu *et al.,* "A two-hase improved correlation method for automatic particle selection in cryo-EM," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 2, pp. 316–325, 2017.

[10] G. Tang, L. Peng, D. Mann, C. Yang, P. Penczek *et al.,* "Eman2: Software for image analysis and single particle reconstruction," *Microscopy and Microanalysis*, vol. 12, no. S02, pp. 388–389, 2016.

[11] S. Niu, Q. Chen, L. DeSisternes, Z. Ji, Z. M. Zhou *et al.,* "Robust noise region-based active contour model via local similarity factor for image segmentation," *Pattern Recognition*, vol. 61, no. 1, pp. 104–119, 2017.

[12] A. Punjani, J. L. Rubinstein, D. J. Fleet and M. A. Brubaker, "cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, no. 3, pp. 290, 2017.

[13] L. A. Gatys, A. S. Ecker and M. Bethge, "A neural algorithm of artistic style," *Journal of Vision*, vol. 16, no. 12, pp. 326–326, 2015.

[14] P. Kumar, Y. Ahmed, S. Alhumam and A. Singla, "Automatic license plate recognition system for vehicles using a CNN," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 5–50, 2022.

[15] G. J. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi *et al.,* "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 3, pp. 60–88, 2017.

[16] A. Hallou, H. G. Yevick, B. Dumitrascu and V. Uhlmann, "Deep learning for bioimage analysis in developmental biology," *Development*, vol. 148, no. 18, pp. 60–88, 2021.

[17] 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. 348–348, 2017.

[18] Y. Xiao and G. Yang, "A fast method for particle picking in cryo-electron micrographs based on fast R-CNN," *AIP Conference Proceedings*, vol. 1836, no. 1, pp. 020080, 2017.

[19] N. P. Nguyen, I. Ersoy, T. A. White and F. Bunyak, "Automated particle picking in cryo-electron micrographs using deep regression," in *IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, Spain, vol. 22, no. 1, pp. 2453–2460, 2018.

[20] T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni *et al.,* "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM," *Communications Biology*, vol. 2, no. 1, pp. 1–13, 2019.

[21] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro *et al.,* "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature Methods*, vol. 16, no. 11, pp. 1153–1160, 2019.

[22] J. Zhang, Z. Wang, Y. Chen, R. Han, Z. Liu *et al.,* "PIXER: An automated particle-selection method based on segmentation using a deep neural network," *BMC Bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019.

[23] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, vol. 70, no. 1, pp. 25–43, 2015.

[24] K. Fang, J. Ouyang and B. Hu, "Swin-HSTPS: Research on target detection algorithms for multi-source high-resolution remote sensing images," *Sensors*, vol. 21, no. 23, pp. 8113–8129. 2021.

[25] Y. Tay, M. Dehghani, S. Abnar, Y. Shen, D. Bahri *et al.,* "Long range arena: A benchmark for efficient transformers," in *Int. Conf. on Learning Representations*, Japanese, vol. 8, no. 2011, pp. 81, 2021.

[26] H. Bao, L. Dong, F. Wei, W. Wang, N. Yang *et al.,* "UniLMv2: Pseudo-masked language models for unified language model pre-training," in *Int. Conf. on Machine Learning*, Dubai, United Arab Emirates, vol. 28, no. 23, pp. 642–652, 2020.

[27] T. Bhamre, T. Zhang and A. Singer, "Denoising and covariance estimation of single particle cryo-EM images," *Journal of Structural Biology*, vol. 195, no. 1, pp. 72–81, 2016.

[28] X. Wang, W. Song, B. Zhang, B. Mausler and F. Jiang, "An early warning system for curved road based on OV7670 image acquisition and STM32," *Computers, Materials & Continua*, vol. 59, no. 1, pp. 135–147, 2019.

[29] M. Lebrun, A. Buades and J. Morel, "A nonlocal Bayesian image denoising algorithm," *Siam Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1665–1688, 2013.

[30] G. Papandreou, L. C. Chen, K. P. Murphy and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Chile, vol. 10, no. 1109, pp. 1742–1750, 2015.

[31] P. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, United States, vol. 6, no. 3, pp. 1713–1721, 2015.

[32] T. Wang, B. Han and J. Collomosse, "Touchcut: Fast image and video segmentation using single-touch interaction," *Computer Vision and Image Understanding*, vol. 120, pp. 14–30, 2014.

[33] M. Hild, M. Hashimoto and K. Yoshida, "Object recognition via recognition of finger pointing actions," in *12th Int. Conf. on Image Analysis and Processing*, Genoa, Italy, vol. 1, no. 2, pp. 88–93, 2003.

[34] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, United States, vol. 1144, no. 4308, pp. 3431–3440, 2015.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh *et al.,* "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[36] A. Bearman, O. Russakovsky, V. Ferrari and F. F. Li, "What's the point: Semantic segmentation with point supervision," *European Conference on Computer Vision*, vol. 1506, no. 2106, pp. 549–565, 2016.

[37] Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, "Residual dense network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 1109, pp. 1812–10477, 2018.

[38] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, United States, vol. 10, no. 1109, pp. 770–778, 2016.

[39] G. Huang, Z. Liu, L. V. DerMaaten and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, United States, vol. 11, no. 4, pp. 4700–4708, 2017.

[40] D. Pathak, E. Shelhamer, J. Long and T. Darrell, "Fully convolutional multi-class multiple instance learning," *Computer Ence*, vol. 12, no. 4, pp. 1412–7144, 2014.

[41] L. F. He, X. W. Ren, Q. H. Gao, X. Zhao, B. Yao *et al.,* "The connected-component labeling problem," *Pattern Recognition*, vol. 70, pp. 25–43, 2017.

[42] A. M. Anger, J. P. Armache, O. Berninghausen, M. Habeck, M. Subklewe *et al.,* "Structures of the human and drosophila 80S ribosome," *Nature*, vol. 497, no. 7447, pp. 80–85, 2013.

[43] C. H. Lee and R. MacKinnon, "Structures of the human HCN1 hyperpolarization-activated channel," *Cell*, vol. 168, no. 1–2, pp. 111–120, 2017.

[44] C. Gatsogiannis, A. E. Lang, D. Meusch, V. Pfaumann, O. Hofnagel *et al.,* "A syringe-like injection mechanism in photorhabdus luminescens toxins," *Nature*, vol. 495, no. 7442, pp. 520–523, 2013.

[45] Y. Zhu, B. Carragher, R. M. Glaeser, D. Fellmann, C. Bajaj *et al.,* "Automatic particle selection: Results of a comparative study," *Journal of Structural Biology*, vol. 145, no. 1–2, pp. 3–14, 2004.

[46] O. V. Loeffelholz, G. Papai, R. Danev, A. G. Myasnikov, S. K. Natchiar *et al.,* "Volta phase plate data collection facilitates image processing and cryo-EM structure determination," *Journal of Structural Biology*, vol. 202, no. 3, pp. 191–199, 2018.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, vol. 20, no. 3, pp. 1412–6980, 2014.

[48] P. A. Penczek, "Resolution measures in molecular electron microscopy," *Methods in Enzymology*, vol. 482, pp. 73–100, 2010.