****

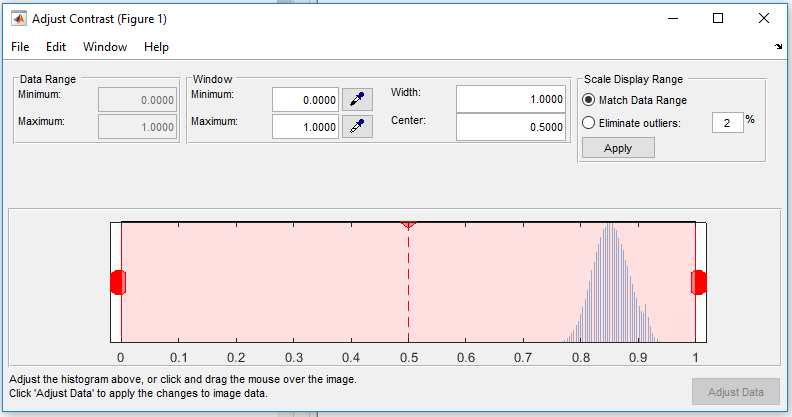
**Fig. S1. The overall single-particle cryo-EM workflow for protein structure determination**.

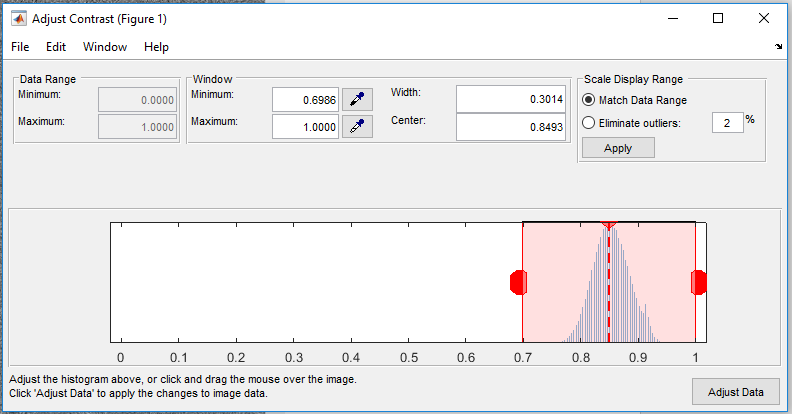


**Fig. S2. The General Framework of AutoCryoPiker: Fully Automated Single Particle Picking (comments: please replace the work “Our Algorithm” with the name of the algorithm)**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |
|  |  |
| **(e)** | **(f)** |

**Fig. S3 Cryo-EM image Samples. (a), (b), (c), and (d): cryo-EM image of frames 1, 10, 20, 50, respectively, (e) cryo-EM conversion using EMAN2, (f) cryo-EM image averaging and conversion using EMAN2**

******

****

**Fig. 4 Contrast Transfer Correction and Adjustment**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

**Fig. S5 Contrast Transfer Correction (CTC), (a) cryo-EM image before applying CTC, (b) cryo-EM image after applying CTC.**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |
|  |  |
|  |  |

**Fig. S6 The Cryo-EM histogram equalization, (a) cryo-EM image after the contrast transfer correction, (c) cryo-EM image after the histogram equalization, (b) and (d) are the histograms of the (a) and (c) respectively. (what do sub-figures (e) (f) represent?)**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |

**Fig. S8 The Cryo-EM Adaptive histogram equalization, (a) cryo-EM image before the processing, (c) cryo-EM image after the processing, (b) and (d) are the histograms of (a) and (c) respectively.**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

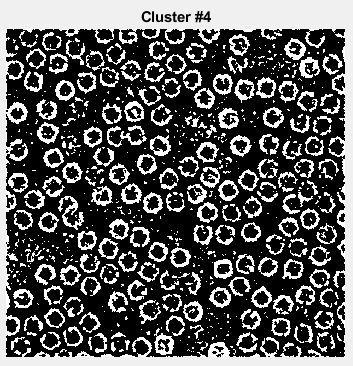
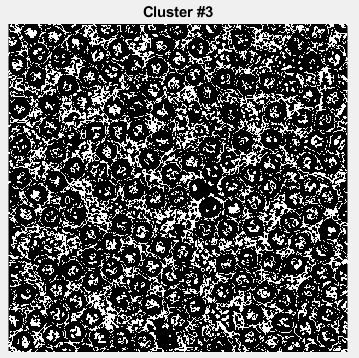
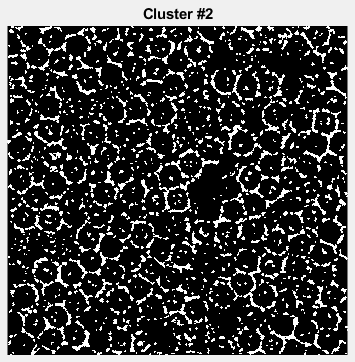
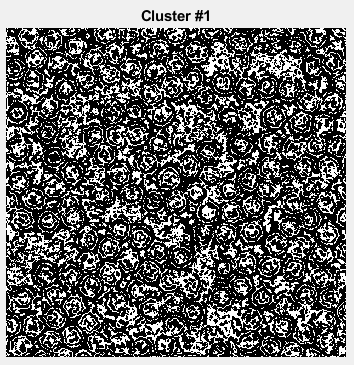
**Fig. S7 Cryo-EM Image Resonation based on Weiner Filtering, (a) cryo-EM image before applying Wiener filter, (b) cryo-EM image before applying Wiener filter.**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

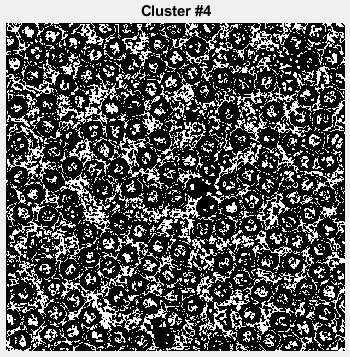
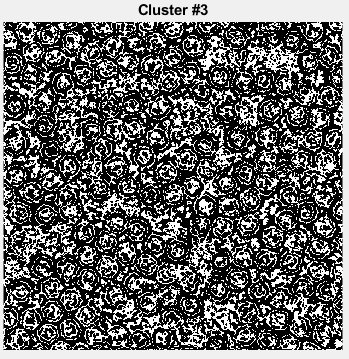
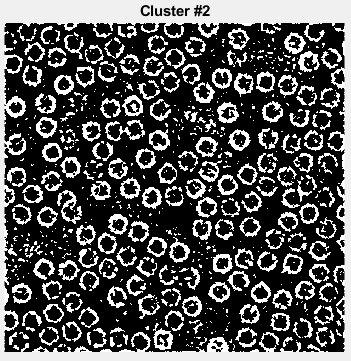
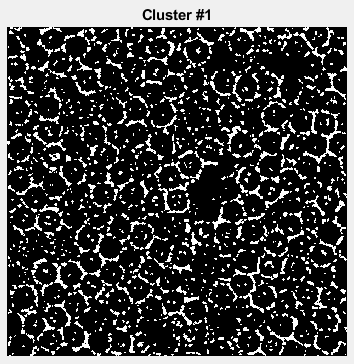
**Fig. S10. Cryo-EM morphological image operation (image closing). (a) cryo-EM before applying image closing, (b) after applying image closing. (comments: where is the image after image closing?)**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

**Fig. S9. (a) Cryo-EM image edge-preserving smoothing using guided filter, (b) histogram of (a).**



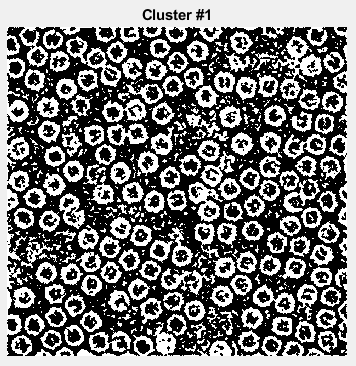
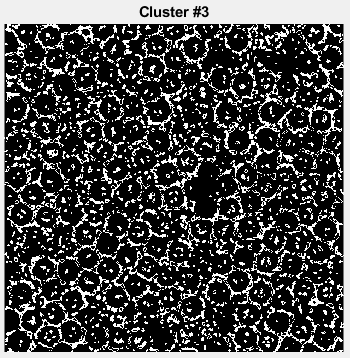
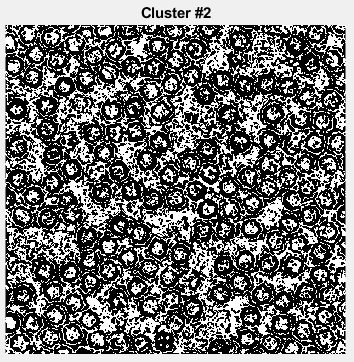
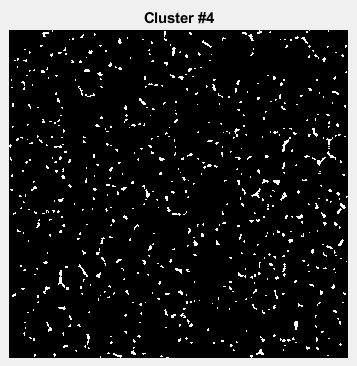
**Fig. S11 Cryo-EM image clustering based K-means clustering algorithm**



**Fig. S12 Cryo-EM image clustering based on FCM clustering algorithm**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

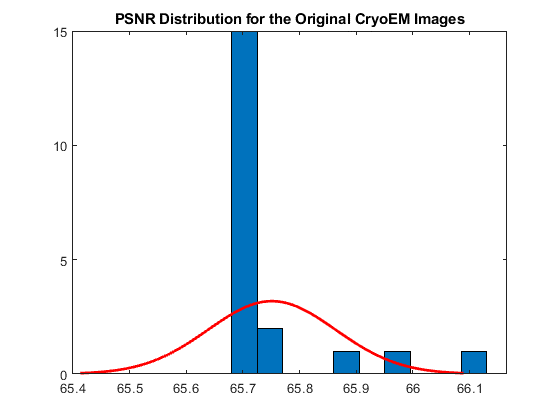
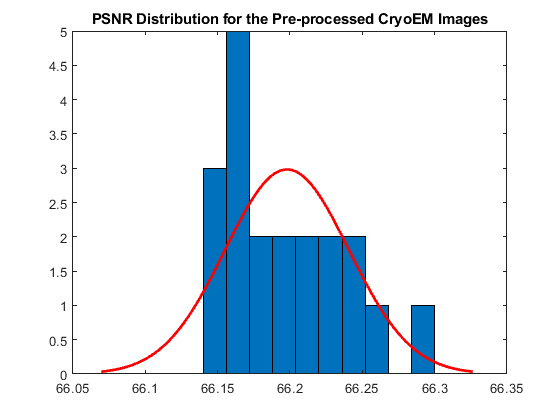
**Fig. S14 Cryo-EM cluster image cleaning. (a) cryo-EM cluster image before the image cleaning, (b) cryo-EM cluster image after image cleaning (how are circles filled here?).**

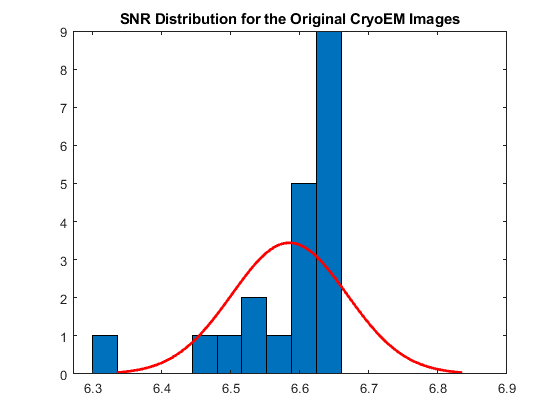
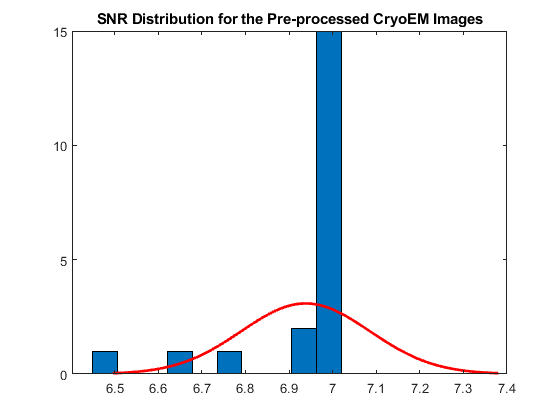
  

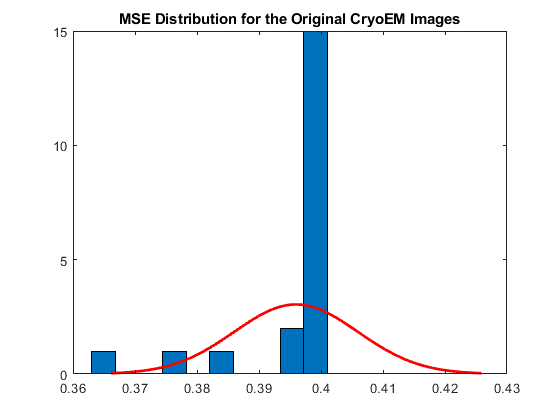
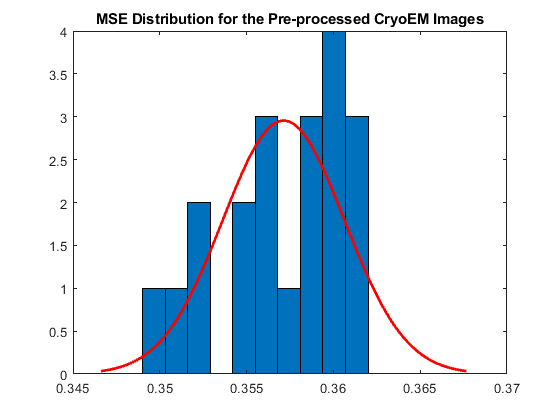
**Fig. S13. Cryo-EM image clustering based the intensity-based clustering algorithm**

|  |  |  |
| --- | --- | --- |
|  | | |
| **(a)** | | |
|  |  |  |
| **(b)** | **(c)** | **(d)** |
|  |  |  |
| **(e)** | **(f)** | **(g)** |
|  |  |  |
| **(h)** | **(i)** | **(j)** |

**Fig. S15 The General Example of AutoCryoPiker: Full Automated Single Particle Picking Based Image Clustering (need to explain what (a)-(j) are here).**

****  ****

****  ****

 ****

**Fig. S17. The quality measurement histograms of the original cryo-EM as well as the pre-processed ones**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |

**Fig. S16 The cryo-EM quality measurement before and after the preprocessing steps. (a) PSNR and SNR, (b) MSE**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |

**Fig. S18 Two cryo-EM images that were manually labeled to evaluate AutoCryoPicker. (a) original image 1, (b) the labeling of image 1 with particles highlighted in yellow, (c) original image 2, (d) the labeling of image 2 with particles highlighted in yellow.**

|  |  |
| --- | --- |
|  |  |
| **(a)** | **(b)** |
|  |  |
| **(c)** | **(d)** |

**Fig. 19 The General Experimental Results of the propose AutoCryoPicker Framework. (Need to explain (a) (b) (c) and (d))**

|  |  |
| --- | --- |
|  |  |
|  |  |
| **(a)** | |
|  |  |

|  |  |
| --- | --- |
|  |  |
| **(b)** | |
|  |  |
|  |  |
| **(c)** | |

**Fig. S20 The Fully Automated Particles Picking Based Different Clustering Approaches, (a) our Intensity-Based Clustering approach, (b) K-Means, (c) FCM.**

### K-Means Algorithm for cryo-EM particle clustering

K-means clustering is a common example of an algorithm that achieves iterative modification of cluster centroids. Suppose we have a data set involving observations of D-dimensional vectors. Each point is assigned to the closest centroid, and each group of points assigned to a centroid is a cluster based on Euclidian distance. Then, the centroid of each cluster is updated according to the points assigned to the cluster. The assignment and update steps are repeated until no point changes clusters, or likewise, until the centroids remain the same. The objective function used to derive the cluster center update is shown in equation (12) (MacQueen, 1967):

(12)

where is the position vector of the kth cluster, is the data point being evaluated, and un is the binary membership vector for data point . The k-means algorithm is shown in Algorithm 1.

|  |
| --- |
| **K-Means Algorithm** |
| **Set** number of clusters, K |
| **Initialize** cluster centers |
| **Repeat** |
| **for** n = 1 to N **do** |
| **Determine** the closest representative, , for |
| **Set** label for data point n to k |
| **end for** |
| **for** k = 1 to K **do** |
| **Update** cluster representative to the mean of the points with cluster label k |
|  |
| **end for** |
| **Until** Change in cluster centers is small |

Fig. S11 shows an example of the final clustering result using k-means algorithm (choosing 4 clusters). We can notice that the most appropriate cluster image is the cluster number 1 which includes background and the particles objects.

### FCM Algorithm for Cryo-EM Particle Picking

FCM algorithm is an unsupervised clustering algorithm in which a dataset is grouped into clusters with every point in the dataset belonging to every cluster to a certain degree. The algorithm of FCM is nearly identical to k-means minus one component. Instead of a binary u, FCM has a u ∈ [0,1]. This means that a data point can have partial membership to multiple clusters. The objective function to minimize for FCM is shown in equation (13) (J. C. Dunn, 1973):

, (13)

where is the fuzzifier and is the squared Euclidian distance between the data point and cluster. The fuzzifier determines how fuzzy the memberships will be where corresponds to crisp clustering. In order to remove the trivial solution of being equal to the zero vector the following constraint is enforced as it shown in equation (14) (J. C. Dunn, 1973):

(14)

The objective function and solving for produces the membership update equation shown in Algorithm 2.

|  |
| --- |
| **Fuzzy C-Means Algorithm** |
| **Set** number of clusters, K |
| **Initialize** cluster centers |
| **Repeat** |
| **for** n = 1 to N do |
| Update membership by taking sum of distance ratios of cluster and all clusters. |
|  |
| **end for** |
| **for** k = 1 to K do |
| **Update** cluster representative to the mean of the points with cluster label |
|  |
| **end** for |
| **Until** Change in cluster centers is small |

Fig. S12 shows an example of the final clustering result using FCM algorithm for all different clustering images based on using 4 clusters only. We can notice that the most appropriate cluster image is the cluster number 3 which includes background and the particles objects.