Eurographics Workshop on 3D Object Retrieval (2019)
SHREC - 3D Shape Retrieval Contest (2019)
S. Biasotti, G. Lavoué, B. Falcidieno, and I. Pratikakis (Editors)
R. C. Veltkamp (SHREC organizer)

SHREC'19 Track: Classification in Cryo-Electron Tomograms

Ilja Gubins[†1], Gijs van der Schot[†2], Remco C. Veltkamp[†1], Friedrich Förster[†2], Xuefeng Du[‡3], Xiangrui Zeng[‡3], Zhenxi Zhu[‡3], Lufan Chang[‡3], Min Xu[‡3], Emmanuel Moebel[‡4], Antonio Martinez-Sanchez[‡5], Charles Kervrann[‡4], Tuan M. Lai[‡6], Xusi Han[‡7], Genki Terashi[‡7], Daisuke Kihara[‡7,6], Benjamin A. Himes[‡8], Xiaohua Wan[‡9], Jingrong Zhang[‡9], Shan Gao[‡9], Yu Hao[‡9], Zhilong Lv[‡9], Xiaohua Wan[‡9], Zhidong Yang[‡9], Zijun Ding[‡10], Xuefeng Cui[‡10], Fa Zhang[‡9]

[1] Department of Information and Computing Sciences, Utrecht University, Netherlands
[2] Department of Chemistry, Utrecht University, Netherlands
[3] Computational Biology Department, Carnegie Mellon University, USA
[4] Inria Rennes Bretagne Atlantique, France
[5] Max Planck Institute for Biochemistry, Germany
[6] Department of Computer Science, Purdue University, USA
[7] Department of Biological Sciences, Purdue University, USA
[8] HHMI Janelia Research Campus, USA
[9] Institute of Computing Technology, Chinese Academy of Sciences, China
[10] Institute for Interdisciplinary Information Sciences, Tsinghua Uninversity, China

Abstract

Different imaging techniques allow us to study the organization of life at different scales. Cryo-electron tomography (cryo-ET) has the ability to three-dimensionally visualize the cellular architecture as well as the structural details of macro-molecular assemblies under near-native conditions. Due to beam sensitivity of biological samples, an inidividual tomogram has a maximal resolution of 5 nanometers. By averaging volumes, each depicting copies of the same type of a molecule, resolutions beyond 4 Å have been achieved. Key in this process is the ability to localize and classify the components of interest, which is challenging due to the low signal-to-noise ratio. Innovation in computational methods remains key to mine biological information from the tomograms.

To promote such innovation, we organize this SHREC track and provide a simulated dataset with the goal of establishing a benchmark in localization and classification of biological particles in cryo-electron tomograms. The publicly available dataset contains ten reconstructed tomograms obtained from a simulated cell-like volume. Each volume contains twelve different types of proteins, varying in size and structure. Participants had access to 9 out of 10 of the cell-like ground-truth volumes for learning-based methods, and had to predict protein class and location in the test tomogram.

Five groups submitted eight sets of results, using seven different methods. While our sample size gives only an anecdotal overview of current approaches in cryo-ET classification, we believe it shows trends and highlights interesting future work areas. The results show that learning-based approaches is the current trend in cryo-ET classification research and specifically end-to-end 3D learning-based approaches achieve the best performance.

CCS Concepts

•***Information systems*** → ***Evaluation of retrieval results; Specialized information retrieval; Multimedia and multimodal retrieval;*** *Retrieval models and ranking;*

1. Introduction

There is a resolution gap in knowledge of cellular life between the molecular level (obtained by techniques such as X-ray crystallography and cryo-electron microscopy single particle analysis) and the cellular level (typically obtained by light microscopy techniques). With the advent of the direct detectors and the associated resolution revolution, cryo-electron tomography (cryo-ET) has the potential to bridge this gap by simultaneously visualizing the cellular architecture, as well as the structural details of macromolecular assemblies thee-dimensionally. The technique may offer insights into key cellular process and improve our understanding of essential life processes.

The biological samples imaged by cryo-ET are sensitive to beam-induced radiation, which limits the maximal resolution of individual tomograms to 5 nm. One common approach to increase

[†] Track organizers. For any questions related to the track, please contact Ilja Gubins. E-mail: i.gubins@uu.nl
[‡] Track participants.

DOI: 10.2312/3dor.20191061

resolution is to average volumes of particle, bringing a challenge of correctly localizing and identifying specific particles in the first place.

Due to the low signal-to-noise ratio of the tomograms and the large amount of data, manual localization of the particles by experts is rarely feasible. Instead, automated approaches utilize the structural signatures within a tomogram. One such common approach is based on applying Difference of Gaussian (DoG) [VYR*09]: a band-pass filter that removes noisy high frequency components and homogeneous low frequency areas, obtaining edges of particle. Based on the edges, a subtomogram can be extracted and classification can be done on a volume that, theoretically, has only one bio-particle. Other methods are based on reference information. For example, known particles can be found in the tomogram by template matching [FBF*02]: using particle as a template for cross-correlation over tomogram to find peak locations, voxels where the template matches best.

Machine learning approaches have been successfully applied to cryo-ET. Support vector machines have been used for both detection and classification [CHP*12]. With ever increasing amount of data captured by cryo-EM and -ET methods [BTE*18], deep learning methods are gaining popularity. Models were proposed for localization [WGL*16], classification [CLZ*18], end-to-end segmentation [CDS*17] and structural pattern mining [XST*19], providing potentially faster, reference-free, and often more accurate results than template matching.

2. Benchmark

We propose a task of localization (e.g. particle picking) and classification (e.g. template matching) of particles in the cryo-electron tomogram volume. A benchmark is conducted on a simulated cryo-electron tomogram populated with 2540 proteins of 12 different classes. To facilitate application of learning-based methods, we also provided nine tomograms simulated in the same fashion as the benchmark, but with ground truth data that was used for simulation.

2.1. Dataset

Our dataset generation starts with creating the original density maps (grandmodels). First, to evaluate localization and classification for various size and shape proteins we chose 12 different biomolecular complexes of known structure with the following PDB entry identifiers:

- 1bxn
- 1qvr
- 1s3x
- 1u6g
- 2cg9
- 3cf3
- 3d2f
- 3gl1
- 3h84
- 3qm1
- 4b4t
- 4d8q

The protein volumes were placed in the grandmodel at random locations, in random orientations, without overlapping each other. For each protein volume we saved its class, its center coordinates and the Euler angles of its orientation (in ZXZ notation). The space in-between proteins is filled with vitreous water (molecular density $0.94g/cm^3$), which was subjected to structural noise (stdev = 0.05). Consecutively we created a series of projection images of the grandmodel with a signal-to-noise ratio of 0.02, applied a contrast transfer function correction to each projection image, added shot-noise, and did a weighted back-projection reconstruction. The resulting reconstructions have a resolution of 1nm/voxel and have a size of 512x512x512 voxels. Each reconstructed tomogram is filled with on average 2500 proteins.

2.2. Evaluation

The main goal of the track is to localize and classify biological particles in the tomogram. The performance of the methods will be evaluated solely on the test tomogram, the only tomogram for which ground truth is not provided.

First, based on ground truth information an automatic script builds a "hitbox" volume. This volume consists of bounding boxes that can be traced back to corresponding ground truth particle. Next, we parse the submitted results and for each predicted particle we try to see if it lies within any bounding box, and record statistical information, such as whether the predicted class is correct, how far from the real center predicted particle center is, and many others.

To have a comprehensive evaluation of the methods, we employ some commonly adopted performance metrics and compute them separately for classification and localization. This separation allows us to compare separate steps of different methods, even if they are not done in end-to-end fashion. The metrics that we are going to evaluate are precision (percentage of results which are relevant), recall (percentage of total relevant results correctly classified), F1 score (harmonic average of the precision and recall) and false negative rate (percentage of results which yield negative test outcomes).

3. Participants

There were eleven groups registered for the track: seven from USA, one from Switzerland, one from Germany, one from France and one from China. The participants had two and a half weeks to send in their results and a one-page description of the method used to obtain the results. Out of eleven, five groups submitted eight result sets. We have assigned short names to each of the result set for easier referencing in the text, and if no title for the method was provided, we also took liberty of giving them a relevant full name.

1. *DoG-CB3D* submitted by Xuefeng Du, Xiangrui Zeng, Zhenxi Zhu, Lufan Chang, Min Xu (section 4.1),
2. *DeepFinder* submitted by Emmanuel Moebel, Antonio Martinez-Sanchez, Charles Kervrann (Section 4.2),
3. *2.5D-Resnet* submitted by Tuan M. Lai, Xusi Han, Genki Terashi, Daisuke Kihara (Section 4.3),
4. *3D-TM* submitted by Benjamin A. Himes (Section 4.4),
5. *3D-HN-localization* submitted by Shan Gao, Zhidong Yang, Jingrong Zhang, Xuefeng Cui, Fa Zhang (Section 4.5),

6., 7. *3D-Unet-CNN-8/12* submitted by Zijun Ding, Shan Gao, Zhidong Yang, Fa Zhang, Xuefeng Cui (Section 4.6),

8. *2.5D-SSD-3D-CNN* submitted by Yu Hao, Zhilong Lv, Xiaohua Wan, Zhidong Yang, Xuefeng Cui, Fa Zhang (Section 4.7).

4. Methods

4.1. DoG-CB3D: Convolution-Based 3D neural network model

By Xuefeng Du, Xiangrui Zeng, Zhenxi Zhu, Lufan Chang, Min Xu

This method separates localization and classification [CLZ*18]. For particle picking, reference-free Difference of Gaussian image transform is used. By subtracting two versions of Gaussian ($s_1 = 3.0, k = 1.2$) filtered image they obtain DoG map which peaks correspond to potential particles. Peaks closer than 10 voxels from each other are filtered to contain only one maximum. The volume of 28^3 voxels around the candidates is then extracted, representing a subtomogram with just one particle.

Next, using provided annotated tomograms, a convolutional neural network (CNN) is trained to classify subtomograms into 13 classes (12 structural classes and 1 none-of-the-above class). The model consists of eight 3D convolutional layers each activated by ReLU, and five max-pooling layers mixed between (Figure 1). The last max-pooling layer is then connected to two fully connected layers with 50% dropout and a softmax final activation.

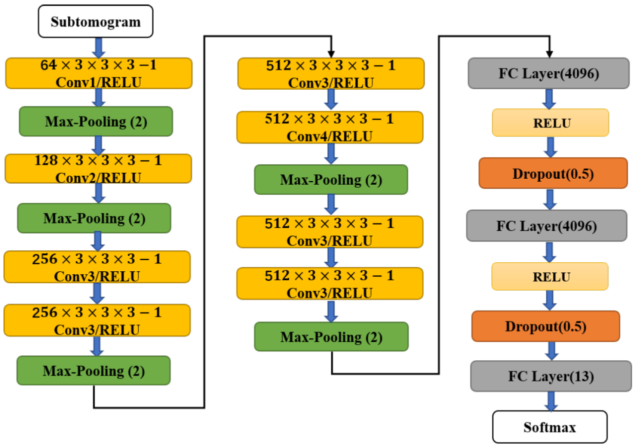

Figure 1: *DoG-CB3D's subtomogram classification model. "64x3x3x3-1 Conv" represents a 3D convolutional layer with 64x3x3x3 filters and stride of 1. "ReLU" and "Softmax" are activation layers. "Max-Pooling (2)" means that max operation is implemented over 2x2x2 regions with stride of 2. "FC Layer (4096)" represents fully connected layers with 1024 neurons.*

The model is trained using stochastic gradient descent with a Nesterov momentum of 0.9 on the categorical cross-entropy loss. To prevent overfitting, an l_2 weight decay regularizer of 0.0005 is added. In addition, the initial learning rate is set at 0.001 with a decay factor of $1e^{-7}$. The training was done in batches of 128 for 100 epochs.

4.2. DeepFinder: Semantic segmentation using 3D U-Net CNN

By Emmanuel Moebel, Antonio Martinez-Sanchez, Charles Kervrann

This method is based on semantic segmentation of tomogram by deep learning and applying clustering on segmentation map to retrieve center coordinates of individual macromolecules [MML*18]. Based on provided ground truth, training segmentation targets are first obtained. A 3D U-Net [RFB15] CNN (Fig. 2) is trained from segmentation maps and used to classify each tomogram's voxel into 13 classes (12 structural and 1 none). DeepFinder estimates each particle center by clustering neighboring voxels into 3D connected components and computing object's centroid.

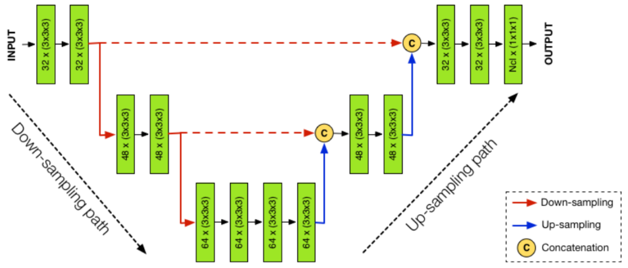

Figure 2: *DeepFinder semantic segmentation model. Green rectangles represent convolutional layers labeled with (#filter x (filter size)). In the last layer, "Ncl" stands for the number of classes.*

The 3D CNN architecture is trained with Adam optimizer, using 0.0001 as learning rate, 0.9 as exponential decay rate for the first moment estimate and 0.999 for the second moment estimate. A Dice loss is used to estimate the network parameters. The training took ~12 hours on an Nvidia K80 GPU, and segmentation and clustering of a 512^3 tomogram takes ~25 minutes.

4.3. 2.5D-Resnet: 2.5D semantic segmentation using 2.5D ResNet

By Tuan M. Lai, Xusi Han, Genki Terashi, Daisuke Kihara

The method is based on 2.5D semantic segmentation of the tomogram. Given a voxel of a tomogram, the proposed deep learning model takes three 2D slices along XY, XZ and YZ-planes around the voxel and outputs 13 probability scores for 12 proteins in dataset and one for whether the voxel is not the center of any particle. The size of each 2D input slice was selected to be 32x32. Each input slice is encoded into a vector consisting of 128 numbers using a CNN with resembling ResNet [HZRS16] architecture. After this step, all three encoded vectors are concatenated into one vector of 384. This new vector is then fed into a feed forward neural network consisting of two hidden layers, producing 13 probability scores.

In order to train the proposed deep learning model, the dataset tomograms were split into training examples. Each example consists of three 2D slices whose centers are located at the same point and the correct protein label for the point. In order to generate negative examples, they randomly sampled points in the tomogram that are not too close to any protein. In the end, they generated about 20000 positive examples and about 20000 negative examples. Since neighbouring voxels can produce multiple center estimates, the neighbourhoods are locally filtered by majority voting to choose one class and one center.

4.4. 3D-TM: 3D template matching

By Benjamin A. Himes

This method is based on using template matching program "templateSearch“ that is part of *emClarity* [HZ18] software package with an estimated particle number of 300 per reference and 2.5% false positive rate based on the estimated number of particles. The threshold is applied to the detection of SNR, similar to [RGD17] with the exception that the normalization of the cross-correlation map is done per-rotation, prior to maximum intensity projection.

The SNR-based threshold is only accurate when the template matching uses a well-matched filter. Due to the bug in projection images of the first version of the dataset, the full process of correctly whitening the input images was not possible, leading to a sub-optimally matched filter. To compensate for this, the estimated threshold was compared to the SNR with the best precision/recall from the training set, and the estimated for each reference in the test data set.

The references themselves were converted to MRC density from the PDB coordinates using a developmental program to create scattering potentials in the software *cisTEM* (unpublished results). In case references had biological assemblies associated with them, the first one of any multiples was picked. The author admits that it was for convenience and predicts that it may not have been the best choice. The tomograms were padded by 10 pixels (100 Å) in each dimension to at least not strictly rule out the edge cases. Normally in *emClarity*, the border of the tomogram, one full particle radius is excluded from the search. Since the tomograms are normalized with a sliding window based on the particle radius, even zero-padded "half“ particles should be weighted appropriately such that the energy they contribute to the cross-correlation is equal to a full particle.

The angular sampling rate was chosen based on the resolution limit, which was 20 Å. Using an average particle radius of 100 Å, this gives an angular step of ~6 degrees ((20/2) / 100 radians). Searches were run on Nvidia v100 GPUs on the Janelia shared resource cluster.

4.5. 3D-HN-localization: 3D-HybridNet localization

By Shan Gao, Jingrong Zhang, Xiaohua Wan, Fa Zhang

The proposed method delegates localization and classification to two neural networks. First, using a 3D segmentation network the tomogram is segmented. By shift-based clustering, segmentations are reduced to particle centers, and subtomograms of 32^3 voxels are extracted. Next, the extracted subtomograms are fed to a hybrid classification network that would determine the final class of particle with an ensemble of two networks.

Only localization part of the method was evaluated in this benchmark.

4.6. 3D-Unet-CNN-8/12: U-Net segmentation and CNN subtomogram classification

By Shan Gao, Jingrong Zhang, Xiaohua Wan, Fa Zhang

The method uses two neural networks for segmentation and classification. Using provided ground truth data, the additional ground truth segmentation labels are generated. Two neural networks were pre-trained separately, a U-Net model for segmentation, and a CNN for classification. Then the two models are combined and trained together using the alternating direction method of multipliers algorithm (ADMM), allowing classification model to supervise the segmentation model. After, the approximate locations of particles are estimated by center shifting clustering of 8 pixels or 12 pixels (which corresponds to two different submissions).

4.7. 2.5D-SSD-3D-CNN: 2.5D particle picking and 3D classifying scheme using deep learning

By Yu Hao, Zhilong Lv, Xiaohua Wan, Zhidong Yang, Xuefeng Cui, Fa Zhang

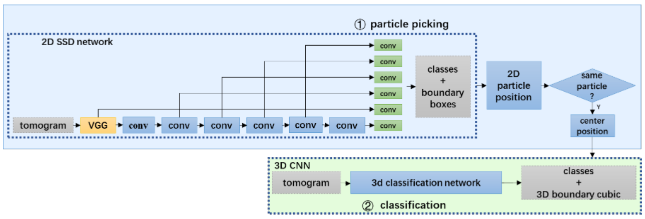

Figure 3: *The particle picking stage (blue) consists of SSD network and a center determination algorithm. The classification stage (green) consists of extracting particles and classifying with 3D hybrid classification network.*

This network considers the tomogram as serial XY sections along z direction. Single shot multi box detector (SSD) [LAE*16] uses VGG [SZ14] model as a feature extractor, followed by six convolutional layers to get multi-scale convolutional feature maps (Fig. 3). The input layer has a fixed size of 512x512, and all 512 slices are processed by the detector.

First, the SSD network is trained on their original simulated data to reach 90 mAP and then fine tune the network with augmented provided data, using tomogram 4 for validation and everything else for training. Then, using graph clustering, they compute particle centers. Such approach takes

Later, a hybrid classification network (HCN) is used to classify the extracted subtomograms around computer center coordinates. The HCN is first trained on 13 (12 structural and 1 none) classes for coarse classification, then an additional small scale fine classification network is trained for fine classification of false negative samples. The final result is then determined by thresholding the most credible score between coarse and fine classifications.

5. Results

In this section, we perform a comparative evaluation of eight result sets submitted by five groups. We measure localization and classification performance based on the metric mentioned in Section 2.2.

Size	Proteins
Tiny	1s3x, 3qm1, 3gl1
Small	3d2f, 1u6g, 2cg9, 3h84
Medium	1qvr, 1bxn, 3cf3
Large	4b4t, 4d8q

Table 3: *Grouping proteins included in the dataset by their size.*

Submission	Tiny	Small	Medium	Large
DoG-CB3D	0.144	0.300	0.714	0.566
DeepFinder	**0.222**	**0.400**	**0.830**	**0.929**
2.5D-Resnet	0.065	0.200	0.352	0.424
3D-TM	0.014	0.000	0.299	0.550
3D-Unet-CNN-8	0.009	0.200	0.587	0.697
3D-Unet-CNN-12	0.012	0.200	0.572	0.699
2.5D-SSD-3D-CNN	0.003	0.000	0.274	0.306

Table 4: *F1 scores of each submission for size classes defined in Table 3. The best results in each column are highlighted.*

5.1. Localization

Table 1 shows detailed localization performance evaluation on the test tomogram. DeepFinder produced the best localization results, achieving F1 score of 0.7971 with precision of 0.7492 at 0.8515 recall. The second best method, DoG-CB3D follows closely with F1 score of 0.7764 with precision of 0.9321 at 0.6653 recall.

5.2. Classification

Table 2 shows classification performance evaluation. DeepFinder method produced the best classification results, achieving highest F1 scores for almost all classes.

For a more in-depth analysis, we group proteins by their size (Table 3), and average the resulting F1 score, which can be extrapolated to estimate classification for other particles of similar size (Table 4).

6. Discussion

This track allows us to identify and compare state-of-the-art methods, as well as highlight current challenges and recognize future research directions.

First of all, we want to note that learning-based methods are increasingly more popular with cryo-ET researchers. Not without a reason: the learning-based methods show better performance.

If we compare just learning-based methods, it is noticeable that 3D methods achieve better performance, suggesting that a neural network can benefit from using all input information, agreeing with conclusions of [DXH*18], that for 3D data one architecture with 3D convolutions performs better than same architecture with 2D convolutions. Results also seem to agree with the popular idea of making networks end-to-end, as in feeding them all existing information instead of separating the work into multiple steps (e.g. end-to-end segmentation vs. localization and classification).

Predictably, methods performance directly correlates with protein sizes, with only one exception of DoG-CB3D approach, for which F1 of "medium" particles outperform F1 of "large" particles. This shows that some additional methods must be developed to overcome limited resolution and increase signal-to-noise ratio, for example with a non-linear denoising or microscope hardware improvements.

We also noted multiple improvement points for our simulation process, which we hope to improve in the future:

1. In the future, the dataset generation method needs to be tested even more thoroughly, preferably automatically. Due to a mistake in our projection process that was found and fixed after dataset publishing, participant 5, *3D-TM*, unknowingly used a faulty early version of dataset, and the results of their methods could have been better.
2. Current dataset generation process uses a constant CTF, which is not realistic.
3. De-focusing gradient must be added.
4. The tomograms could have been more packed, and should be, as in real biological samples.

The timeframe for the SHREC contest adds additional pressure on the participants, and it would be better if such a benchmark could have been done continuously. The training dataset could be published first, and the evaluation on a test data could be done online. Then, once a year, for each SHREC contest, newly submitted results could be highlighted and described in a paper like this one.

References

[BTE*18] BALDWIN P. R., TAN Y. Z., ENG E. T., RICE W. J., NOBLE A. J., NEGRO C. J., CIANFROCCO M. A., POTTER C. S., CARRAGHER B.: Big data in cryoem: automated collection, processing and accessibility of em data. *Current Opinion in Microbiology 43* (2018), 1 – 8. Environmental Microbiology * The New Microscopy. URL: http://www.sciencedirect.com/science/article/pii/S1369527417301315, doi:https://doi.org/10.1016/j.mib.2017.10.005. 2

[CDS*17] CHEN M., DAI W., SUN S. Y., JONASCH D., HE C. Y., SCHMID M. F., CHIU W., LUDTKE S. J.: Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *nature methods 14*, 10 (2017), 983. 2

[CHP*12] CHEN Y., HRABE T., PFEFFER S., PAULY O., MATEUS D., NAVAB N., FÖRSTER F.: Detection and identification of macromolecular complexes in cryo-electron tomograms using support vector machines. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)* (2012), IEEE, pp. 1373–1376. 2

[CLZ*18] CHE C., LIN R., ZENG X., ELMAAROUFI K., GALEOTTI J., XU M.: Improved deep learning-based macromolecules structure classification from electron cryo-tomograms. *Machine Vision and Applications 29*, 8 (Nov 2018), 1227–1236. URL: https://doi.org/10.1007/s00138-018-0949-4, doi:10.1007/s00138-018-0949-4. 2, 3

[DXH*18] DENIZ C. M., XIANG S., HALLYBURTON R. S., WELBECK A., BABB J. S., HONIG S., CHO K., CHANG G.: Segmentation of the proximal femur from mr images using deep convolutional neural networks. *Scientific reports 8*, 1 (2018), 16485. 5

[FBF*02] FRANGAKIS A. S., BÖHM J., FÖRSTER F., NICKELL S., NICASTRO D., TYPKE D., HEGERL R., BAUMEISTER W.: Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proceedings of the National Academy of Sciences 99*, 22 (2002), 14153–14158. 2

Submission	**RR**	**TP**	**FP**	**FN**	**MH**	**RO**	**AD**	**Recall**	**Precision**	**Miss rate**	**F1 Score**
DoG-CB3D	1813	1690	**110**	850	**13**	**1**	**2.4519**	0.6653	**0.9321**	0.3346	0.7764
DeepFinder	2887	**2163**	709	**377**	15	24	3.5063	**0.8515**	0.7492	**0.1484**	**0.7971**
2.5D-Resnet	4524	1507	1185	1033	876	**1**	3.9866	0.5933	0.3331	0.4066	0.4266
3D-TM	2429	814	356	1726	425	313	2.5608	0.3204	0.3351	0.6795	0.3276
3D-HN-localization	2127	455	867	2085	311	48	5.9316	0.1791	0.2139	0.8208	0.1949
3D-Unet-CNN-8	2500	1367	372	1173	480	13	4.1660	0.5381	0.5468	0.4618	0.5424
3D-Unet-CNN-12	2500	1438	555	1102	352	12	4.4083	0.5661	0.5752	0.4338	0.5706
2.5D-SSD-3D-CNN	1977	710	196	1830	485	7	4.6453	0.2795	0.3591	0.7204	0.3143

Table 1: *Results of localization evaluation. RR: results reported; TP: true positive, unique particles found; FP: false positive, reported non-existant particles; FN: false negative, unique particles not found; MH: multiple hits: unique particles that had more than one result; RO: results otside of volume; AD: average euclidean distance from predicted particle center; Recall: uniquely selected true locations divided by 2540, number of particles in the test tomogram; Precision: uniquely selected true locations divided by RR; Miss rate: percentage of results which yield negative results* $(1 - recall)$*; F1 Score: harmonic average of the precision and recall. The best results in each column are highlighted.*

Submission	**1bxn**	**1qvr**	**1s3x**	**1u6g**	**2cg9**	**3cf3**	**3d2f**	**3gl1**	**3h84**	**3qm1**	**4b4t**	**4d8q**
DoG-CB3D	0.866	0.619	0.047	0.446	**0.343**	0.657	0.358	0.225	0.25	0.160	0.222	0.911
DeepFinder	**0.904**	**0.800**	**0.154**	**0.522**	0.330	**0.784**	**0.584**	**0.318**	**0.332**	**0.193**	**0.907**	**0.951**
2.5D-Resnet	0.087	0.405	0.119	0.263	0.018	0.566	0.366	0.039	0.293	0.037	0.489	0.359
3D-TM	0.684	0.020	0.005	0.024	0.008	0.194	0.008	0.019	0.032	0.018	0.211	0.890
3D-Unet-CNN-8	0.702	0.559	0	0.234	0.268	0.501	0.209	0.029	0.008	0	0.684	0.711
3D-Unet-CNN-12	0.663	0.577	0	0.243	0.273	0.477	0.209	0.038	0.008	0	0.671	0.728
2.5D-SSD-3D-CNN	0.312	0.343	0	0.054	0	0.166	0.040	0.010	0	0	0.379	0.234

Table 2: *Results of classification evaluation for all classes. The values correspond to F1 score achieved by participants on specific classes. The best results in each column are highlighted.*

[HZ18] Himes B. A., Zhang P.: emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature methods 15*, 11 (2018), 955. 4

[HZRS16] He K., Zhang X., Ren S., Sun J.: Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778. 3

[LAE*16] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. C.: Ssd: Single shot multibox detector. In *European conference on computer vision* (2016), Springer, pp. 21–37. 4

[MML*18] Moebel E., Martinez A., Larivière D., Ortiz J., Baumeister W., Kervrann C.: 3d convnet improves macromolecule localization in 3d cellular cryo-electron tomograms. 3

[RFB15] Ronneberger O., Fischer P., Brox T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241. 3

[RGD17] Rickgauer J. P., Grigorieff N., Denk W.: Single-protein detection in crowded molecular environments in cryo-em images. *Elife 6* (2017), e25648. 4

[SZ14] Simonyan K., Zisserman A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014). 4

[VYR*09] Voss N., Yoshioka C., Radermacher M., Potter C., Carragher B.: Dog picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. *Journal of structural biology 166*, 2 (2009), 205–213. 2

[WGL*16] Wang F., Gong H., Liu G., Li M., Yan C., Xia T., Li X., Zeng J.: Deeppicker: A deep learning approach for fully automated particle picking in cryo-em. *Journal of Structural Biology 195*, 3 (2016), 325 – 336. URL: http://www.sciencedirect.com/science/article/pii/S1047847716301472, doi:https://doi.org/10.1016/j.jsb.2016.07.006. 2

[XST*19] Xu M., Singla J., Tocheva E. I., Chang Y.-W., Stevens R. C., Jensen G. J., Alber F.: De novo structural pattern mining in cellular electron cryotomograms. *Structure* (2019). URL: http://www.sciencedirect.com/science/article/pii/S096921261930005X, doi:https://doi.org/10.1016/j.str.2019.01.005. 2