

# Multi-task Learning for Macromolecule Classification, Segmentation and Coarse Structural Recovery in Cryo-Tomography

Chang Liu<sup>1</sup>

cliu5@andrew.cmu.edu

Xiangrui Zeng<sup>1</sup>

xiangrui@andrew.cmu.edu

Kai Wen Wang<sup>1</sup>

kaiwenw1@andrew.cmu.edu

Qiang Guo<sup>2</sup>

qguo@biochem.mpg.de

Min Xu (Corresponding Author)<sup>1</sup>

mxu1@cs.cmu.edu

<sup>1</sup> School of Computer Science, Carnegie Mellon University  
Pittsburgh, PA, USA

<sup>2</sup> Max Planck Institute for Biochemistry  
Martinsried, Germany.

---

## Abstract

Cellular Electron Cryo-Tomography (CECT) is a powerful 3D imaging tool for studying the native structure and organization of macromolecules inside single cells. For systematic recognition and recovery of macromolecular structures captured by CECT, methods for several important tasks such as subtomogram classification and semantic segmentation have been developed. However, the recognition and recovery of macromolecular structures are still very difficult due to high molecular structural diversity, crowding molecular environment, and the imaging limitations of CECT. In this paper, we propose a novel multi-task 3D convolutional neural network model for simultaneous classification, segmentation, and coarse structural recovery of macromolecules of interest in subtomograms. In our model, the learned image features of one task are shared and thereby mutually reinforce the learning of other tasks. Evaluated on realistically simulated and experimental CECT data, our multi-task learning model outperformed all single-task learning methods for classification and segmentation. In addition, we demonstrate that our model can generalize to discover, segment and recover novel structures that do not exist in the training data.

## 1 Introduction

The cell is the basic unit of living organisms. Most cellular processes are governed by macromolecules. To fully understand such processes, it is necessary to precisely know the structure and spatial organization of all macromolecules inside single cells. Such information has been extremely difficult to obtain due to limitations in data acquisition. Recently, Cellular Electron Cryo-Tomography (CECT) has emerged as a dominating 3D imaging technique that captures cellular structure at sub-molecular resolution and in close-to-native state,

providing systematic 3D visualization of close-to-native state macromolecular structures in unprecedented resolution and faithfulness. However, systematic recovery of macromolecule structures is very challenging due to the imaging limitation of CECT and structural complexity of macromolecules. In particular, the images are captured at a very low signal-to-noise ratio (SNR), making it hard to identify macromolecules via simple inspection. Because of the limited imaging tilt angle range at data acquisition stage, CECT 3D images (aka tomograms) generally suffer from missing wedge effect (i.e. missing values), resulting in anisotropic resolution. Also, macromolecules are structurally highly diverse. They are often densely distributed inside the tomograms while dynamically interacting with each other, therefore introducing more complex and heterogeneous structures.

Earlier studies of macromolecule localization inside CECT image relied on template matching[1], comparing an experimental macromolecule to a known structural template. For the unbiased detection, classification, and high-resolution structure recovery of all heterogeneous complexes, reference-free methods are needed. Several unsupervised approaches have been developed on macromolecular structure recovery. The pipeline consists of three main steps: first, particle-picking methods [2] are applied for the localization of potential macromolecules and then subtomograms<sup>1</sup> are extracted from a tomogram based on those locations. Second, due to the crowded cellular environment of tomogram, extracted subtomograms may contain not only the target macromolecule but also fragments of neighboring structures. Previously, model based clustering method [3] was introduced for automatic segmentation of the target complex. However, those manually heuristic rules were not reasonable enough to deal with the inherent complexity of CECT data. For instance, the method assumes that all structural regions are spherical-like while rod or plane like structures are not being taken into account. Third, reference-free subtomogram alignment, classification and averaging methods [e.g. 4, 5, 6, 7, 8, 9] subdivide all macromolecules into several homogeneous groups, and the averaging of the subtomograms of same macromolecular structure will improve the resolution of recovered structure, achieving structure recovery from CECT data.

Despite that the promising classification and structure recovery results have been shown, existing unsupervised approaches still suffer from limited scalability and discrimination ability due to intensive computations. Recently, deep learning based methods have been analyzed to assist systematic structure recovery pipeline while maintaining great scalability and accuracy. In particular, Liu et al [10] have proposed a 3DCNN based segmentation network, named SSN3D-ED which is capable of masking out neighboring structures inside subtomograms while maintaining generalization ability to segment different types of macromolecules. Also, large-scale 3D subtomogram classification models [6, 11] such as DSRF3D-v2 have achieved solid results on subtomograms which contain no neighboring structures. Then reference-free structural recovery approaches can be applied to those classified homogeneous group of subtomograms.

Generally, deep learning based subtomogram segmentation and classification are performed separately in a cascade manner, and therefore require training and inference on different models, complicating the automatic structure recovery pipeline while taking more computational resource. Since low-level image features in the 3DCNN models[12, 13] can be shared by both tasks, it is a natural and intuitive idea to build up such a multi-task model to complete those two steps end-to-end in one shot while improving the performance via

<sup>1</sup> A *subtomogram* is a small cubic subvolume of a tomogram generally contains a single macromolecule extracted from a tomogram.

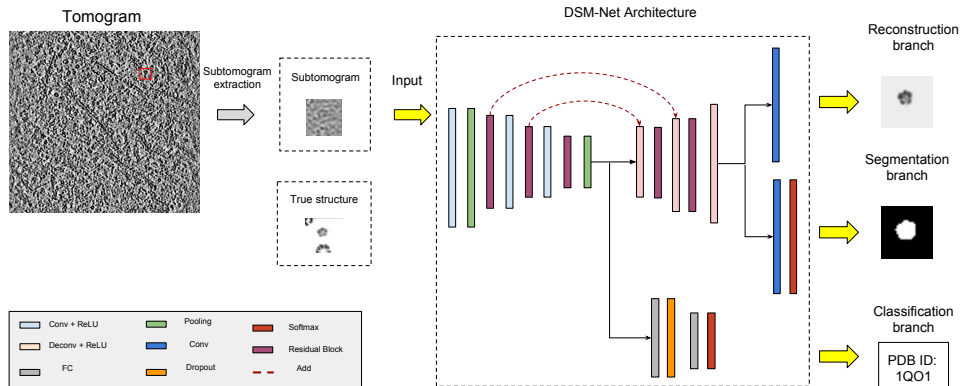


Figure 1: A conceptual diagram of DSM-Net architecture.

inductive transfer[5] compared to single-task learning.

In this paper, three CECT analysis tasks have been explored: 1) Identify the structural class of the macromolecule contained in a subtomogram if it is a known structure class in training data; 2) semantically segment the target macromolecule out of neighbor structures in subtomograms; 3) an auxiliary task of coarsely recovering the density map of macromolecular structure for the assistance of two main tasks and proof-of-principle visualization. We proposed a novel 3DCNN multi-task learning model named Deep Subtomogram Multi-task Network (DSM-Net, Fig. 1) for implementing those three tasks.

Tests on both realistically simulated and experimental dataset, our multi-task model significantly outperformed the single-task models that separately doing classification and segmentation. In addition, we show that our model has certain generalization ability to classify, segment and coarsely recover the new structures that do not exist in the training data.

## 2 Method

DSM-Net is an end-to-end, unified network which consists of a backbone network for computing convolutional feature maps, and three parallel subnetworks for subtomogram segmentation, classification and coarse structure recovery.

We use the low level image features shared by the 3DCNN models[12, 19] to build up a multi-task model that performs semantic segmentation and subtomogram classification jointly. Further, we redefine the structure recovery process as a image reconstruction problem, and extend the model by adding a small overhead on backbone to jointly learn the coarse recovered structures (represented by a density map) of target macromolecule.

### 2.1 Residual based Backbone Network and Classification Subnet

Considering the scenario that input subtomograms contain not only target complexes but also neighboring structures, previous 3D VGG based subtomogram classification model such as DSRF3D-v2 [9] failed to converge during the training process. To increase the optimization ability, 3D residual blocks plus stride-2 convolutional layer design [9] is introduced to the

backbone network. In particular, input is connected to 3D conv followed by stride-2 max-pooling layer. Then two consecutive 3D residual block plus a stride-2 3D conv layer are applied. Finally, it is followed by a 3D residual block plus 3D average pooling layer. The number of channels of three residual block is (32, 64, 128) respectively.

For classification subnet, it predicts the probability of input subtomogram categories ( $C = 22$  classes). The backbone network is followed by a fully connected layer with 1024 hidden units at the dropout rate of 70%. Equipped with a softmax activation layer, the subnet outputs C-unit vector and predicts query subtomogram to the macromolecular class that returns the highest probability.







Note that all following convolution and deconvolution layers are followed by a ReLU activation and their kernel size is  $3 \times 3 \times 3$  except the final convolution layers of segmentation and structure recovery subnet.

## 2.2 Segmentation Subnet

The backbone network and segmentation subnet follow the encoder-decoder architecture adopted from previous model SSN3D-ED which is a 3D variant of fully convolutional network. Besides 3D residual block is taken for computing feature maps at backbone network, 3D deconvolution replaces interpolation-based upsampling layers in segmentation subnet compared to SSN3D-ED. In our experiments, we found that this in-network upsampling filters outperformed simple bilinear upsampling ones for learning dense pixel prediction.

Specifically, the backbone is followed by two consecutive 3D stride-2 deconv plus a 3D residual block. Then, one more 3D stride-2 deconv is attached and followed by a  $1 \times 1 \times 1$  3D conv layer with the channel size of 2 (target region and background). After softmax activation is applied, predictions are voxel-level binary classification. In addition, skip connections from the lower layers of residual block are added to the correspondingly higher deconvolution layers, in order to integrate coarse information with fine-grained information for solving local ambiguities.

## 2.3 Structure Recovery Subnet

After subtomograms are well segmented and classified, structure recovery is usually achieved through state-of-art unsupervised alignment and the averaging of homogeneous group of complexes to obtain a high-resolution density map [e.g. , , , , , ], which are extremely computational expensive. For improving scalability while keeping recovery accuracy, we redefine the structure recovery as a supervised image reconstruction problem. Given  $x$  is a measurement of 3D CECT image,  $G$  is the previously described state-of-art unsupervised model whose response is regard as the ground truth of structure recovery and  $H$  is our supervised DSM-Net model, the objective function of image reconstruction can be written as:

$$L_{rec} = \frac{1}{N} \sum_x \|G(x) - H(x)\|_2 \quad (1)$$

where  $N$  is the number of training examples.

The overall objective loss function is a weighted sum of segmentation (seg), classification (cls), and reconstruction loss (rec). For choosing the proper weights of subtask loss, refer to

section 3.2

$$L_{obj} = \frac{1}{N}(L_{seg} + \alpha L_{cls} + \beta L_{rec}) \quad (2)$$

Structural recovery and segmentation of the macromolecule of interest are highly related tasks. Specifically, the structural recovery determines the electron density of the macromolecule of interest at each voxel location which directly correlates to subtomogram image intensity, whereas the semantic segmentation is a binary decision on whether a voxel is occupied by the macromolecule of interest by taking into account of the combination of subtomogram image intensity. With respect to distinction, segmentation task generally segments only a rough region of target complex while recovery task focuses even more on the fine-grained contour and inner cellular structure as well. One example has shown this distinction: a hole in a macromolecule will show a low electron density on the density map, but still be categorized as a part of the segment (see first column of Fig. 3 c and d for example).

Sharing parameters with segmentation task except the final convolution layer, structure recovery subnet simply add a 3D conv with channel = 1 to predict a reconstructed image while keeping model complexity. In this paper, we regard the structure recovery only as an auxiliary task which might improve the generalization ability of two main tasks and provide 3D visualization clues for discovering new structures. Further experiments which quantitatively evaluate this task and task-specified network architecture will be explored in future.

### 3 Experiments

We evaluated our proposed DSM-Net on realistically simulated and experimental dataset and compared the result to previous single-task models: DSRF3D-v2 and SSN3D-ED. Additionally, We made comprehensive ablations experiments in task-specified single models extracted from DSN-Net respectively. Further, sensitivity tests on the combination of weighted loss ratio of three tasks have been conducted.

For evaluation metrics, object level classification accuracy is adopted as classification task, while pixel level classification accuracy (Pix acc) and mean Intersection over Union (mIOU) are used for semantic segmentation.

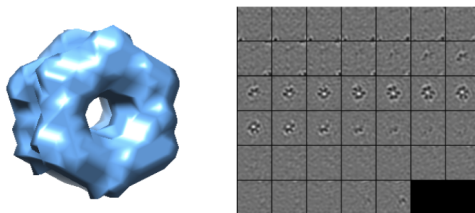


Figure 2: Left: Isosurface of Glutamine synthetase (PDB ID: 2GLS); Right: 2D slices of subtomograms with SNR = 0.5

#### 3.1 Dataset

**Simulated subtomograms from known structures** For a persuasive assessment of the approach, we generated realistically simulated tomograms with known structures of macro-

molecular complexes (class of macromolecules of the same structure) by simulating the actual tomographic image reconstruction process as previously described [14]. The limitation of CECT data such as missing wedge, image noise, electron optical factors, including the Modulation Transfer Function (MTF) and the Contrast Transfer Function (CTF), were properly included.

Specifically, 22 distinct macromolecular complexes (Tab. Supplementary S1) are chosen from the Protein Databank (PDB) [8] for experiments. Each simulated tomograms of  $600 \times 600 \times 300$  voxels contains 10000 randomly distributed macromolecules with a tilt angle range  $\pm 60^\circ$ . Given the true position of these macromolecules inside tomograms, we extracted the subvolumes of  $40^3$  voxels centering on these positions as input to our model. Removing those subtomograms outside the boundary of tomograms, we finally collected 3205 simulated subtomograms of 22 structural classes for each dataset. Datasets A, B have SNR of 0.06, and 0.01 respectively. Fig.2 shows an example of 2D slices of a subtomogram extracted from a simulated tomogram.

**Experimental tomograms** A ribosome dataset of 859 subtomograms was extracted from a tomogram of primary rat neuron culture [8]. The tomogram was captured from tilt angle  $-50^\circ$  to  $+70^\circ$ . It was then binned twice to a voxel size of 1.368 nm. Subtomograms of size  $40^3$  were extracted from the tomogram using Difference of Gaussian particle picking method [14] and coarsely filtered by a convolutional autoencoder [20]. Template search was applied to select the top 1000 subtomograms with highest structural correlation with the ribosome template. We manually inspected the 1000 subtomograms, and filtered out 141 of them which contained obvious non-ribosome structure such as fiducial.

Additionally, a dataset consisting of 386 single capped proteasome subtomograms is extracted from a tomogram of rat neuron with expression of poly-GA aggregate [8]. All subtomograms were two times binned to size  $40^3$  (voxel size: 1.368 nm). The tilt angle range was  $-50^\circ$  to  $+70^\circ$ .

To prevent class imbalance problem, we randomly select 400 ribosome subtomograms out of 859 filtered subtomograms. Overall, 400 ribosome and 386 single capped proteasome subtomograms are combined and shuffled, named Dataset C. The segmentation and density map ground truth were prepared by aligning the corresponding structural template (PDB ID: 5T2C and 5MPA).

## 3.2 Implementation Detail

Adam [10] was used to optimize the parameters. We set the learning rate to  $10^{-3}$  with the batch size of 64 and exponential decay rates to  $\beta_1 = 0.9$  and  $\beta_2 = 0.99$ . Categorical cross-entropy was used as the loss function for semantic segmentation and classification tasks, while mean squared error for structure recovery. All models were trained for no more than 100 epochs while the early stopping criterion was applied if the validation dataset did not improve for 15 consecutive epochs.

## 3.3 Results

Dataset A, B, C were randomly split into training, validation, and testing set at the ratio of 0.1 respectively. The performance result on simulated Dataset A is reported in Table 1. First, it shows that the separately trained subnetworks of DSM-Net outperformed previous methods

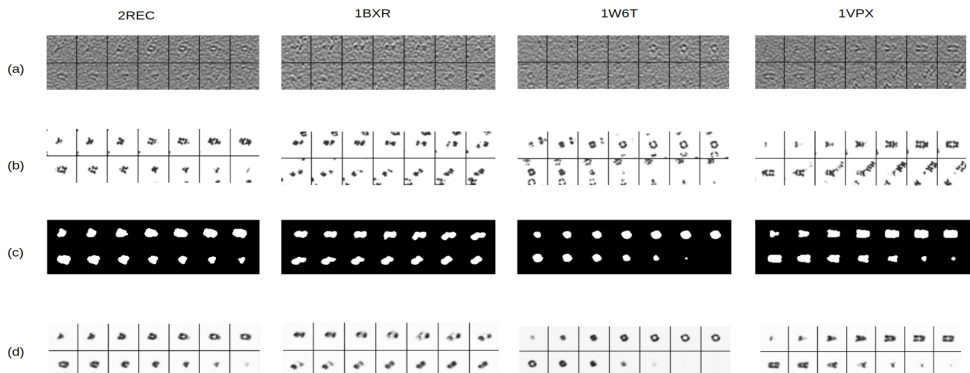


Figure 3: Examples of key 2D slices cut from input subtomograms and corresponding output segmentation predictions and regressed density maps: (a) Input subtomograms (b) True structures used to produce simulated subtomograms (c) Segmented regions of interest (d) Recovered structure of macromolecules of interest represented as density map.

such as DSRF3D-v2 and SSN3D-ED. Res-backbone for classification subnet enables the optimization of subtomograms that even contains neighboring structures while VGG-backbone failed to converge. Besides, mIOU from segmentation subnet has hugely increased by 2.39% compared to SSN3D-ED. Second, different combinations of the subnetworks of DSM-Net have been evaluated. It illustrates that all multi-task models outperformed single-task ones while DSM-Net maintained the best mIOU and object accuracy. It indicates that structure recovery branch as auxiliary task assisted the optimization of two main tasks with adding minor computational cost. Specifically, segmentation result classification accuracy reaches 93.75%, which is a substantial enhancement compared other models.

DSM-Net outputs are visualized in Figure 3. It is shown that neighboring structures in subtomograms have been masked out (Figure 3 (c)), and the structure of target macromolecules is coarsely recovered even for some inner features such as hollow (Figure 3 (d)).

In addition, challenging Dataset B and C with SNR = 0.01 have also been assessed in Table 2. DSM-Net has shown promising results on simulated Dataset B with mIoU, and accuracy increased by 1.7% and 9.37% respectively. However, DSM-Net returns similar results on experimental dataset with respect to single-segmentation model. It is probably due to the limited number of categories in dataset C (ribosome and proteasome only) and therefore has minor influence in learning further discriminative features in shared backbone network. On the other hand, multi-task scheme solved the convergence problem compared to previous classification models on experimental subtomogram, achieving zero classification error. We emphasize that the decent classification accuracy is due to the small size of dataset C (786 samples in total) with only two categories.

**Ablation Experiments** Since overall loss function is a linear combination of three tasks, different pairs of loss ratio are needed to be analyzed. 2-task learning models revised from DSM-Net provides a rough suggestion on selecting such weighting coefficient of each single-task loss. From Table 3, we find that results are insensitive to those values. Finally, 1: 10: 1 for classification, segmentation and structure recovery respectively is adopted as weighting coefficient for training DSM-Net to keep the order of magnitude for three tasks at the same level.

Table 1: Evaluation performance on simulated Dataset A with SNR = 0.06. To clarify, classification, segmentation and structure recovery subnet are individually or dually extracted from DSM-Net, abbreviated as Cls, Seg and Rec Subnet below. ‘-’ indicates incompatible tasks while ‘\*’ indicates the model fails to converge.

Method	Backbone	Segmentation		Classification
		Pix acc	mIoU	Obj acc
DSRF3D-v2[ <a href="#">6</a> ]	3D VGG-8	-	-	*
Cls Subnet	3D ResNet-9	-	-	76.24
SSN3D-ED [ <a href="#">12</a> ]	3D VGG-8	98.99	84.68	-
Seg Subnet	3D ResNet-9	99.03	87.07	-
Cls+Rec Subnet	3D ResNet-9	-	-	81.87
Cls+Seg Subnet	3D ResNet-9	<b>99.22</b>	88.70	84.37
DSM-Net	3D ResNet-9	99.21	<b>89.00</b>	<b>93.75</b>

Table 2: Evaluation performance on challenging simulated Dataset B and experimental Dataset C with SNR = 0.01.

Method	Segmentation		Classification	Dataset
	Pix acc	mIoU	Obj acc	
DSRF3D-v2[ <a href="#">6</a> ]	-	-	*	Simulated
SSN3D-ED [ <a href="#">12</a> ]	98.15	76.43	-	
Cls Subnet	-	-	43.75	
Seg Subnet	98.37	79.70	-	
DSM-Net	<b>98.59</b>	<b>81.40</b>	<b>53.12</b>	
DSRF3D-v2[ <a href="#">6</a> ]	-	-	*	Experimental
SSN3D-ED [ <a href="#">12</a> ]	95.24	58.93	-	
Cls Subnet	-	-	*	
Seg Subnet	<b>96.01</b>	60.38	-	
DSM-Net	95.80	<b>61.26</b>	<b>100.00</b>	

Table 3: Varying the loss ratio of given models on Dataset A.

Method	Segmentation		Classification
	Pix acc	mIoU	Obj acc
Cls+Rec 1:1	-	-	81.56
Cls+Rec 1:10	-	-	81.87
Cls+Seg 1:10	99.22	<b>88.70</b>	<b>84.37</b>
Cls+Seg 1:100	<b>99.23</b>	88.63	82.81



### 3.4 Discussion and Insight on Model Design

The DSM-Net has an encoder-decoder architecture. Learning through the classification module of DSM-Net encourages the encoding of subtomograms to be more discriminative between different structural classes. When the encoding is fed into the segmentation decoder module, segmentation accuracy can be improved. Conversely, learning in the segmentation decoder module provides supervision on the region of interest (ROI) of macromolecule and filters out neighbor structures. The structural outline learned by segmentation decoder module contains the structural class information. The filtration of neighbor structures helps the network to focus on the macromolecular structure of interest and reduce the bias introduced by neighbor structures. The above two factors can significantly improve classification accuracy. Since image reconstruction is highly correlated to segmentation (Explained in Sec 2.3) but involving more fine-grained structural information, successful recovery of structure will make the shared upsampling module learn even detailed contour of macromolecule, and therefore improve the segmentation result.

### 3.5 Identification and Recovery of Unseen Structure

As described in the introduction section, unsupervised reference-free structure recovery method involves clustering, alignment, and the averaging of homogeneous subtomograms, which is computationally intensive. Although 3DCNN classification model proposed by Xu et al [19] accelerates the subtomograms subdivision step, a more scalable model is needed for structure recovery to reduce computational cost.

By contrast, our proposed DSM-Net serves as an end-to-end model that can simultaneously segment, classify, and coarsely recover unseen macromolecule structures that do not exist in the training data. To test the generalization ability of DSM-Net, we trained DSM-Net on Dataset A (containing 22 classes), excluding all the GroEL subtomograms (PDB ID: 1KP8).

For discovering this new structure, we inferred all training, and testing data via DSM-Net and outputted 1024 hidden units from the fully connected layer of classification subnet. Those hidden units interpreted as structural features are invariant to missing wedge effect and rigid transformation. After that, k-means clustering with  $k = 22$  was performed on all nonlinear transformed data, and we picked out the cluster enriched with 1KP8 subtomograms. For visual assessment of this discovery of 1KP8, we embedded all clusters into an  $\mathbb{R}^2$  space using the T-SNE algorithm [23]. According to Fig.4(g), it is obvious that 1KP8 subtomograms are mostly located inside a single cluster even though this structure did not exist in the training data. Additionally, the structure recovery subnet of DSM-Net can output a coarse new structure of 1KP8 (Fig.4(b)). The performance of structural recovery is measured using Fourier Shell Correlation [24] between true and recovered structures of 1KP8 for the representation of structural discrepancy which is reported to have a decent score 5.0.

**Remark:** 1KP8 is largely distinct from other structures in training set, and therefore the successful discovery, and structural recovery of this structure can strongly support the generalization ability of our multi-task model. Regard to the cluster number for k-means, it should be chosen larger than the expected classes in dataset while not causing much computational burden. Efficient cluster number estimation methods remain to be explored.

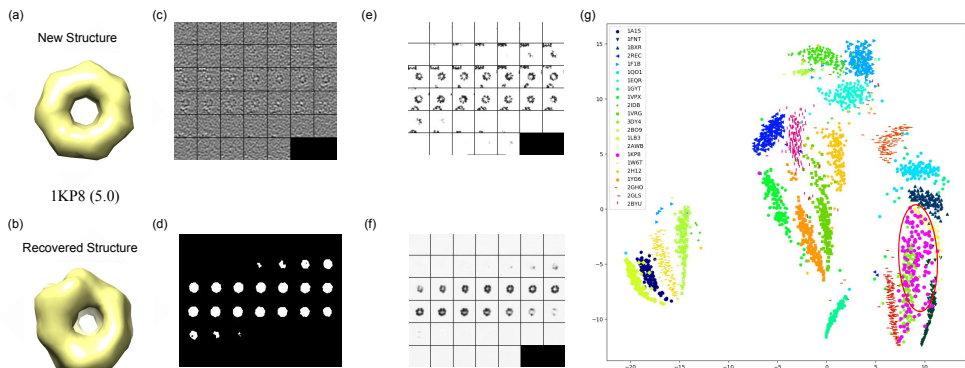


Figure 4: The outputs of DSM-Net for the new structure GroEL (PDB ID: 1KP8) with respect to training data. (a) The isosurfaces of true structure (b) The isosurfaces of recovered structure (c) Subtomogram of GroEL (d) Segmented region of GroEL (e) True structure of subtomogram (f) Recovered structure of subtomogram (g) The visualization of subtomograms in dataset  $A$  embedded to  $\mathbb{R}^2$  using T-SNE. The region enriched with GroEL subtomograms is highlighted using red circle.

## 4 Conclusion

In this paper, we present a novel multi-task 3D convolutional neural network (DSM-Net) for jointly performing classification, semantic segmentation, and coarse structural recovery of macromolecules in subtomograms. To our knowledge, this work is the first application of deep multi-task learning for CECT analysis. Evaluated on simulated and experimental dataset with different noise level, DSM-Net markedly outperforms two baseline single-task networks. In addition, we redefine structure recovery as a supervised image reconstruction problem, which serves as an auxiliary task for assisting two main tasks. The output of auxiliary task provides potential clues for discovering new structures. Further, we demonstrate that our model has certain generalization ability to classify and recover the structures that do not exist in training data. Our work serves as an important step towards systematic structural identification and recovery of macromolecules captured by CECT. For future works, structure recovery task will be quantitatively analyze, and improved. Also, improving performance in low SNR subtomograms and identifying macromolecule spatial interactions (such as interaction with cell membrane) though semantic segmentation are other important issues to be explored.

## 5 Acknowledgements

This work was supported in part by U.S. National Institutes of Health (NIH) grant P41 GM103712. M.X acknowledges support of the Samuel and Emma Winters Foundation.

## References

- [1] A Bartesaghi, P Sprechmann, J Liu, G Randall, G Sapiro, and Sriram Subramaniam. Classification and 3d averaging with missing wedge correction in biological electron tomography. *Journal of structural biology*, 162(3):436–450, 2008.
- [2] Martin Beck, Johan A Malmström, Vinzenz Lange, Alexander Schmidt, Eric W Deutsch, and Ruedi Aebersold. Visual proteomics of the human pathogen leptospira interrogans. *Nature methods*, 6(11):817, 2009.
- [3] Helen M Berman, Talapady N Bhat, Philip E Bourne, Zukang Feng, Gary Gilliland, Helge Weissig, and John Westbrook. The protein data bank and the challenge of structural genomics. *Nature Structural & Molecular Biology*, 7:957–959, 2000.
- [4] John AG Briggs. Structural biology in situ—the potential of subtomogram averaging. *Current opinion in structural biology*, 23(2):261–267, 2013.
- [5] Rich Caruana. Multitask learning. In *Learning to learn*, pages 95–133. Springer, 1998.
- [6] Chengqian Che, Ruogu Lin, Xiangrui Zeng, Karim Elmaaroufi, John Galeotti, and Min Xu. Improved deep learning based macromolecules structure classification from electron cryo tomograms. *arXiv preprint arXiv:1707.04885*, 2017.
- [7] Yuxiang Chen, Stefan Pfeffer, José Jesús Fernández, Carlos Oscar S Sorzano, and Friedrich Förster. Autofocused 3d classification of cryoelectron subtomograms. *Structure*, 22(10):1528–1537, 2014.
- [8] Q. Guo, C. Lehmer, A. Martínez-Sánchez, T. Rudack, F. Beck, H. Hartmann, M. Pérez-Berlanga, F. Frottin, M. Hipp, U. Hartl, D. Edbauer, W. Baumeister, and R Fernández-Busnadiego. In Situ Structure of Neuronal C9ORF72 Poly-GA Aggregates Reveals Proteasome Recruitment. *Cell* doi:10.1016/j.cell.2017.12.030, 2018.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [11] Hstau Y Liao and Joachim Frank. Definition and estimation of resolution in single-particle reconstructions. *Structure*, 18(7):768–775, 2010.
- [12] Chang Liu, Xiangrui Zeng, Ruogu Lin, Xiaodan Liang, Zachary Freyberg, Eric Xing, and Min Xu. Deep learning based supervised semantic segmentation of electron cryo-subtomograms. *arXiv preprint arXiv:1802.04087*, 2018.
- [13] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [14] Long Pei, Min Xu, Zachary Frazier, and Frank Alber. Simulating cryo electron tomograms of crowded cell cytoplasm for assessment of automated particle picking. *BMC bioinformatics*, 17(1):405, 2016.

- 
- [15] Sjors HW Scheres, Roberto Melero, Mikel Valle, and Jose-Maria Carazo. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure*, 17(12):1563–1572, 2009.
- [16] Min Xu and Frank Alber. Automated target segmentation and real space fast alignment methods for high-throughput classification and averaging of crowded cryo-electron subtomograms. *Bioinformatics*, 29(13):i274–i282, 2013.
- [17] Min Xu, Martin Beck, and Frank Alber. High-throughput subtomogram alignment and classification by fourier space constrained fast volumetric matching. *Journal of structural biology*, 178(2):152–164, 2012.
- [18] Min Xu, Elitza I Tocheva, Yi-Wei Chang, Grant J Jensen, and Frank Alber. De novo visual proteomics in single cells through pattern mining. *arXiv preprint arXiv:1512.09347*, 2015.
- [19] Min Xu, Xiaoqi Chai, Hariank Muthakana, Xiaodan Liang, Ge Yang, Tzviya Zeev-Ben-Mordehai, and Eric Xing. Deep learning based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *Bioinformatics*, 33(14):i13–i22, 2017. doi: 10.1093/bioinformatics/btx230.
- [20] Xiangrui Zeng, Miguel Ricardo Leung, Tzviya Zeev-Ben-Mordehai, and Min Xu. A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation. *Journal of structural biology*, 2017.