Semi-supervised Macromolecule Structural Classification in Cellular Electron Cryo-Tomograms using 3D Autoencoding Classifier

Siyuan Liu[1]
siyuanl2@andrew.cmu.edu
Xuefeng Du[2]
xuefengdu1@gmail.com
Rong Xi[1]
rxi@andrew.cmu.edu
Fuya Xu[1]
fuyax@andrew.cmu.edu
Xiangrui Zeng[2]
xiangruz@andrew.cmu.edu
Bo Zhou[2]
bzhou2@andrew.cmu.edu
Min Xu (Corresponding Author)[2]
mxu1@cs.cmu.edu

[1] Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, USA

[2] Computational Biology Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, USA

Abstract

Recent advances in the Cellular Electron Cryo-Tomography (CECT) imaging technique have enabled the 3D visualization of macromolecules and other sub-cellular components in single cells in their near-native state. Automatic structural classification of macromolecules is increasingly desirable for researchers to better study and understand the features of different macromolecular complexes. However, accurate classification of macromolecular complexes is still impeded by the lack of annotated training data due to the limited expert resource for labeling full datasets. In this paper, we introduce a semi-supervised classification framework to reduce annotation burden in the macromolecule structural classification tasks. Specifically, we propose a 3D autoencoding classifier framework for simultaneous macromolecule structural reconstruction and classification. Our framework jointly optimizes two branches of network using both labeled and unlabeled data during training phase. Extensive experiments demonstrate the effectiveness of our approach against other semi-supervised classification approaches on both real and simulated datasets. Our approach also achieves competitive results in terms of macromolecule reconstruction. To our best knowledge, this is the first work to address the task of semi-supervised macromolecule structural classification in CECT.

1 Introduction

Macromolecules play an important role in driving molecular processes in cells, which are the basic structural and functional unit of living organisms. The structures and spatial organizations of macromolecules are critical for the functioning of many biological pathways. However, due to data acquisition limitations, it is hard to obtain the native structural information of macromolecules in single cells, which significantly hinders our understanding of the machinery of the macromolecules. Recent advances in Cellular Electron CryoTomography (CECT) imaging technique have enabled 3D visualization of sub-cellular structures at sub-molecular resolution in a near-native state, which makes it a promising tool for 3D visualization of macromolecules in single cells [20]. However, automatic classification of macromolecular complexes is still restricted by the highly heterogeneous structural complexity of macromolecules and the lack of annotated data for training a well-performed classifier. Researchers have proposed to use deep convolutional neural networks (CNNs) [4, 7, 14, 21, 38] for discriminative feature extraction and supervised classification of macromolecule complexes. Though these supervised approaches have achieved a promising classification accuracy, they require proper human annotation and expertise in specific domain which are very costly and not easily accessible, especially for large-scale macromolecule datasets. In clinical and biomedical research, the amount of data/annotation pairs is often limited due to the insufficient expert resource for labeling full datasets. An unsupervised convolutional autoencoder [37] has been recently introduced to first learn the hidden representations for the input subtomograms (A subtomogram is a cubic subvolume of a tomogram that likely to contain a single macromolecule) using 3D autoencoder network. After that, K-means clustering is applied to coarsely generate a structural grouping of input subtomograms. However, the unsupervised settings in this method prevent us from obtaining the exact labels of raw input data. The performance of this method is not even comparable to fully supervised models due to the lack of supervision signals.

In order to reduce the annotation burden in the subtomogram classification tasks while achieving high classification accuracy, we introduce a semi-supervised learning framework which is able to take advantage of both labeled and unlabeled data for learning feature representations. Specifically, we propose a 3D autoencoding classifier network for simultaneous subtomogram classification and structural reconstruction. Our network consists of an autoencoder network for unsupervised feature mining and a classifier network for supervised classification. An encoder network is utilized as a feature extractor shared by both the classifer and the decoder network. We use the output of the encoder network to train a classifier using only the labeled data. Meanwhile, the learned representations by the encoder network are projected back to their initial shape by the decoder network. For the unlabeled data, we also input them into the autoencoder and the classifier but force them to not influence the parameter updating in the classifier. Instead of training the autoencoder first and then fine-tuning the classifier in a cascaded way, we jointly train the subtomogram classification and reconstruction branch using both labeled and unlabeled data so that the whole network is optimized in an end-to-end scheme. The flowchart of our joint subtomogram classification and reconstruction process is illustrated in Figure 1. Our contributions are summarized as follows:

- We propose a 3D autoencoding classifier network for more effective semi-supervised subtomogram classification, which is able to significantly reduce the annotation cost for deep model's training. To our best knowledge, this is the first work to address semi-supervised macromolecule structural classification in CECT.

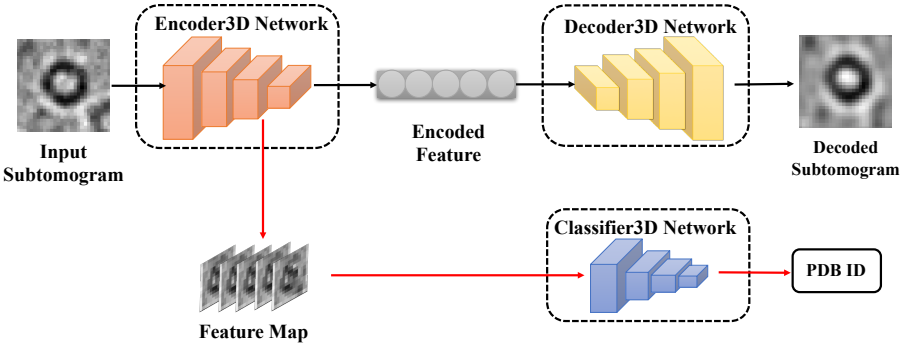

Figure 1: The overview of our proposed 3D autoencoding classifier network architecture. The black arrows show the original reconstruction branch based on 3D autoencoder [37] for subtomogram reconstruction. The red arrows indicate the extra classification branch designed for subtomogram classification.

- In our proposed framework, we adopt a joint learning scheme for the subtomogram classification and the structural reconstruction branch so that the two branches can complement each other. We experimentally demonstrate that jointly training the two branches outperforms training in a cascaded way.
- We evaluate our model on both real and simulated subtomogram datasets under settings where training data are rarely labeled. Experimental results show an apparent advantage of our approach against other semi-supervised classification approaches. Our model also achieves a promising reconstruction performance compared to baseline 3D autoencoder approach. Visualization results demonstrate that the reconstructed subtomogram by our approach can significantly reduce image noise and suppress the irrelevant image features of neighboring structures.

2 Related Work

Subtomogram classification is becoming vital to the study of macromolecular complexes. There have been several works that address supervised classification algorithms on subtomograms. Xu *et al.* [34] propose two 3D CNN models based on VGGNet and GoogleNet for subtomogram feature extraction and classification. Chang *et al.* [18] present a three-branch CNN framework for simultaneous subtomogram classification, structural segmentation and recovery. This framework adopts a multi-task learning technique to jointly optimize each individual task. Che *et al.* [4] propose a customized convolutional C3D [29] based CNN structure for macromolecule classifications. This network is robust to subtomograms with various SNRs and tilt angles. One limitation of these supervised approaches is the requirement of abundant labeled data, which often needs professional human time and effort. These approaches will suffer from overfitting when only a small portion of data is labeled. On the other hand, Lin *et al.* [16] propose a domain adaptation framework to improve the subtomogram classification model trained on simulated data to apply to experimental data. However, it does not make use of unlabeled data and the accuracy still has a large room to improve. In that case, semi-supervised learning becomes more popular for being able to learn from both labeled and unlabeled data.

Semi-supervised learning is an important tool for automatic data annotation when only a small portion of data is artificially labeled. Traditional semi-supervised learning approaches are divided into self-training [27, 28], graph-based [2, 3, 40] and generative model based [6, 10, 13]. Due to their capability to extract deep features from high-dimensional data samples, CNN-based models are increasingly popular in the task of semi-supervised classification. Chen *et al*. [5] proposes a Feed-Forward CNN architecture along with an innovative unlabeled data selection method specialized for image classification. They construct the convolutional layers and fully-connected layers in an unsupervised manner where no back propagation is used. Li *et al*. [15] proposes a disentangled Variantional Autoencoder structure along with reinforcement learning to deal with insufficient training data. Many approaches [11, 22, 30] divide the training phase of CNN into two disjoint steps. They pre-train the model using both labeled and unlabeled data under unsupervised settings, which is followed by supervised fine-tuning with only labeled data. This approach is able to well leverage both labeled and unlabeled data in training a CNN model. One limitation of this approach is that the learned parameters using unsupervised learning might not be an optimal initialization for supervised classification. Others [26, 36] utilize a hybrid autoencoder structure which jointly optimizes the two steps with both labeled and unlabeled data. However, these approaches are still limited to 2D settings. When it comes to 3D images which contain higher dimensions and more complicated structural information, these approaches might not work out as expected.

3 Method

3.1 Overview

In general semi-supervised classification task, a large training set is given but only a small portion is manually labeled. Our goal is to utilize such a training set to train a well-performed classifier, which is able to significantly reduce the annotation cost.

The key issue of semi-supervised classification task is to leverage unlabeled datasets in learning generalized feature representation. Motivated by [37], we assume a 3D autoencoder network is an effective approach for unsupervised feature mining of subtomograms and can be well employed to extract features from unlabeled data. In addition, we notice that the encoder network can be utilized as a feature extractor for classification task. Based on these assumptions, we propose a 3D autoencoding classifier network specialized for semi-supervised classification on subtomogram datasets.

Figure 1 illustrates the general framework of our proposed 3D autoencoding classifier network. This framework contains two branches specialized for two different tasks: subtomogram reconstruction and classification. The Encoder3D network is shared by two branches for feature extraction of the input subtomogram. The Classifier3D network takes the feature map output by the last max-pooling as input and is trained to predict the label (PDB ID) of the subtomogram. The Decoder3D network aims to reconstruct the 3D structure of the original subtomogram from the encoded feature vector.

3.2 Network Architecture

The network architecture of our 3D Autoencder network is shown in Figure 2. Our encoder network takes a $28 \times 28 \times 28$ 3D subtomogram as input and outputs a 128 1-D encoded feature vector. It consists of two 3D convolutional layers, two 3D max pooling layers and one fully connected layer. Each convolutional layer is followed by a ReLU activation layer

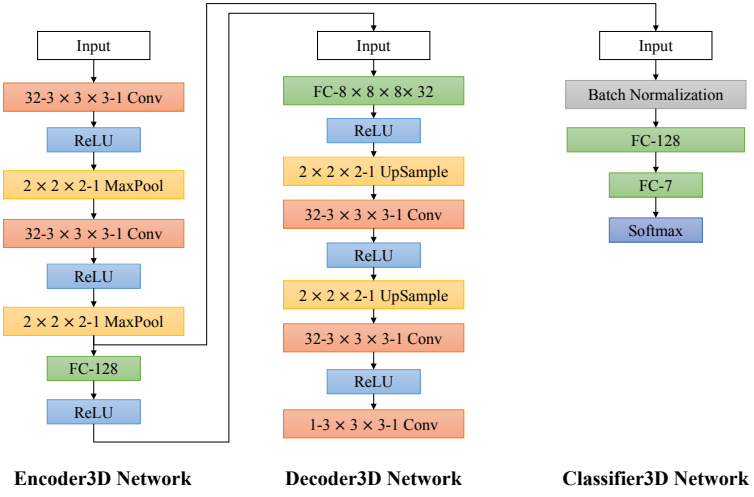

Figure 2: Network Architecture of Encoder3D, Decoder3D and Classifier3D networks. The function and parameter settings of each layer is shown inside the box. "32-3 × 3 × 3-1 Conv" indicates a Convolutional layer with 32 filters. The kernel size is 3 and the stride is 1. "2 × 2 × 2-1 MaxPool" means a MaxPooling layer with kernel size 2 and stride 1.

to increase output sparsity and prevent overfitting. We further apply L1 normalization in the fully connected layer to encourage sparsity of the encoded features. As is shown in previous works [23, 37], sparsity regularization can greatly improve the performance of autoencoder.

The Decoder3D network structure is symmetric to the Encoder3D network. It takes the encoded feature vectors by the Encoder3D network as input and outputs a 28 × 28 × 28 3D structure, which is of the same dimension as the input to the Encoder3D network. It consists of two 3D UpSampling layers corresponding to the two 3D MaxPooling layers in the Encoder3D network. The upsampling ratio is set to 2 in all three dimensions.

The Classifier3D network takes the feature map produced by the last MaxPooling layer of the Encoder3D network as its input. We choose not to use the encoded feature vector from the last layer because the encoding pattern for the two tasks are different and sparse encoding might not be well applicable for the classification task. We then introduce a Batch normalization layer immediately after the input layer to normalize the input and improve the generalization of the Classifier3D network. The feature maps are encoded by two fully connected layers and a softmax layer is applied to predict the possibilities of different classes for classification.

3.3 Semi-supervised Classification

We discuss how this architecture can be utilized for our semi-supervised classification task in the training phase. Let $D = \{X_1, X_2, ...X_n\}$ be the training set of subtomograms and $D_{labeled} = \{(X_1, y_1), (X_2, y_2), ...(X_m, y_m)\}$ be the labeled subset, where $m << n$. Our goal is to use the limited labeled data and large portions of unlabeled data to train a high-performance classifier.

Our framework contains two branches which are regularized by different loss functions during training phase. The reconstruction branch employs the Mean Square Error (MSE)

Algorithm 1 Training procedure of 3D autoencoding classifier

Input: Training Set $D = \{X_1, X_2, ...X_n\}$
Output: Trained $Encoder3D$, $Decoder3D$ and $Classifier3D$

Initialize network parameters for $Encoder3D$, $Decoder3D$, and $Classifier3D$
for i = 1 to n **do**
    Fetch data X_i from training set D
    Get decoded data $\hat{X}_i = Decoder3D(Encoder3D(X_i))$
    if X_i is labeled **then**
        Fetch label y_i for X_i
        Predict label $P = Classifier3D(Encoder3D.MaxPool2(X_i))$
        Calculate loss $L = \alpha L_{MSE}(X_i, \hat{X}_i) + \beta L_{CE}(y_i, P)$
    else
        Calculate loss $L = L_{MSE}(X_i, \hat{X}_i)$
    for each parameter w in $Encoder3D$, $Decoder3D$ and $Classifier3D$ **do**
        Update w using $\nabla L(w)$

loss function as follows:

$$L_{MSE}(X, \hat{X}) = \frac{1}{N} \sum ||X - \hat{X}||^2 \tag{1}$$

where $X \in R^{w \times w \times w}$ is the original subtomogram and $\hat{X} \in R^{w \times w \times w}$ is the decoded subtomogram. $N = w \times w \times w$ is the size of X.
The classification branch utilizes Cross-Entropy loss function:

$$L_{CE}(y, P) = -ylog(P^T) \tag{2}$$

where $y \in R^{1 \times C}$ is the one-hot encoding of the ground-truth label. $P \in R^{1 \times C}$ is the feature vector output by softmax layer. C represents the number of classes to predict.

We adopt an end-to-end scheme in training the proposed 3D autoencoding classifier network where the two branches are jointly trained and optimized with both labeled and unlabeled data. Consequently, joint loss is applied to optimize the whole architecture of our 3D Autoencoder:

$$L_{joint} = \alpha L_{MSE}(X_i, \hat{X}_i) + \beta L_{CE}(y, P) \tag{3}$$

where α and β are the weights of the two loss values.

The training procedure for one epoch is illustrated in Algorithm 1. As we can see, the network is optimized differently for labeled and unlabeled data. The labeled data can be used to optimize both the reconstruction branch and the classification branch, where the final loss is the weighted sum of MSE loss and Cross-Entropy loss. The unlabeled data can only be used to update the parameters in the reconstruction branch with MSE loss. Therefore, α is set to 1 and β is set to 0 in Eq 3. In testing phase for the classification task, only the classification branch is employed to predict the actual class of testing data.

4 Experiments

We conduct experiments on both real and simulated datasets to evaluate the performance of our proposed framework. The proposed 3D autoencoding classifier is evaluated in two aspects. To begin, we compare the classification performance of our approach with one

Table 1: The experimental macromolecular complexes used for tomogram simulation

PDB ID	Macromolecular Complex
1FNT	Yeast 20S proteasome with activator PA26
2GLS	Glutamine Synthetase
1F1B	E. coli asparate transcarbamoylase P268A
2IDB	3-octaprenyl-4-hydroxybenzoate decarboxylase
1KP8	GroEL-KMgATP
3DY4	Yeast 20S proteasome
4V4A	E. Coli 70S Ribosome
5T2C	Human Ribosome

supervised approach and three more semi-supervised classification approaches. We then measure the subtomogram reconstruction performance of our approach compared with the baseline 3D autoencoder network. We utilize visualized subtomograms to have a better understanding of the reconstruction performance.

4.1 Dataset

Real Dataset This dataset contains 2800 subtomograms of size 28^3 from 7 classes of macro-molecules, which are extracted from Noble Single Particle Dataset collected by Noble *et al*. [24]. For each tomogram in the original set, subtomograms of size 28^3 were extracted using a Difference of Gaussian(DOG) particle picking process [25]. We then apply a template search approach as described in [37] to select the top 1000 subtomograms according to the cross-correlation scores. Four hundred subtomograms are manually selected for each class which contain clear macro-molecule structures. In our experiments, we select 1400 subtomograms for training and the remaining 1400 for testing.

Simulated Dataset This dataset consists of 8000 simulated subtomograms of size 64^3 from eight classes of macro-molecules. These subtomograms are extracted from realistically simulated tomograms, which are generated by simulating the actual tomographic image reconstruction process [25] based on well-recognized structures of macromolecular complexes. We select eight classes of macromolecular complexes from Protein Databank (PDB) [1] for our experiments. The eight classes of macromolecular complexes are shown in Table 1. In addition, 1000 simulated subtomograms of size 64^3 and signal-to-noise (SNR) 0.03 are extracted for each class so we get 8000 subtomograms in total. In our experiment, we select 3200 subtomograms for training and the remaining 4800 for testing.

4.2 Experiments on Semi-supervised Classification

In this experiment, we compare the classification performance of our proposed 3D autoencoding classifier with supervised 3D CNN, self-trained 3D CNN [27, 28], pre-trained 3D CNN [11, 22, 30] and deep generative model [13].

Supervised 3D CNN We utilize the classification branch of our proposed 3D autoencoding classifier network as the bottleneck architecture of the supervised 3D CNN. This network is trained in an end-to-end scheme using only the labeled data and regularized with Eq 2.

Self-trained 3D CNN This method is based on the supervised 3D CNN as a classifier. During each training epoch, the unlabeled data is classified by the trained classifier and the

Table 2: Classification accuracy (%) on real and simulated testing sets. The classification performance is evaluated under three settings of labeled proportions: 5%, 10% and 20%

	Real Dataset			Simulated Dataset		
	5%	10%	20%	5%	10%	20%
Supervised 3D-CNN	14.29	30.78	87.71	17.78	28.92	33.56
Self-trained 3D-CNN [27, 28]	28.57	42.86	57.57	17.35	25.56	32.19
Pre-trained 3D-CNN [11, 22, 30]	14.29	71.57	94.03	38.04	59.19	75.15
Deep Generative Model [13]	72.14	76.86	76.64	16.67	18.00	17.19
Our approach	**78.21**	**84.64**	**95.36**	**50.27**	**71.44**	**77.85**

data with the top confidence score is added to the training set. We set the threshold as 0.6 in the real dataset's experiment and 0.95 in the simulated dataset's experiment. These thresholds are derived using grid search on validation set.

Pre-trained 3D CNN The network architecture is the same as our proposed 3D autoencoding classifier model except that the two branches of the network are trained in a cascaded way. The network is pre-trained with the unlabeled data, which is followed by supervised fine-tuning with only labeled data.

Deep Generative Model We implement this approach using the source code published by [13] and evaluate it on our subtomogram datasets.

Experiment settings The network architecture we use for real dataset is shown in Figure 2. As for the simulated dataset, we change the dimensions of the last fully connected layer of Classifier3D network to be eight, thus corresponding to the total number of classes. In response to the increased dimensions and complexities of the simulated data, we add two additional (Convolution3D + MaxPooling) layers to Encoder3D network and two additional (Upsampling3D + Convolution3D) layers to Decoder3D network. The hyperparameters of the newly added layers are the same as those of the original architecture.

In terms of the hyperparameters for training, we set learning rate to 0.0001 and branch size to 64. Adam [12] is applied as an optimizer with decay rate $\beta 1 = 0.9$ and $\beta 2 = 0.99$. We randomly sample 10% of the training set as a validation set and the validation loss is used as a metric for early stopping. All the models are trained for 100 epochs on the training set. Both α and β in Eq 3 are set to 1.0 in our approach for labeled data. These settings are applied across all the experiments using CNN. As for the deep generative model approach, we strictly follow all the parameter settings in [13].

Testing results We evaluate our model using three labeled settings:5%, 10% and 20%, which are the proportions of the labeled data in training set. The testing results of the five approaches on the real dataset and the simulated dataset are shown in Table 2. Our approach clearly outperforms all the other methods under all settings of the labeled proportions on both real and simulated datasets. Our model significantly outperforms supervised 3D-CNN model, which means that 3D autoencoder network is an effective way to leverage unlabeled data in learning generalized feature representation. Besides, the apparent advantage over the Pre-trained 3D-CNN indicates that the joint optimization of the two branches is a more suitable scheme for training proposed model. We notice that the classification accuracy on the simulated dataset is lower than on the real dataset of all three settings. It is primarily because simulated dataset contains subtomogram with higher resolutions and smaller voxel spacing, which makes this dataset more challenging than the real dataset. Compared with deep generative model, our approach is more robust to input subtomogram with higher resolution and

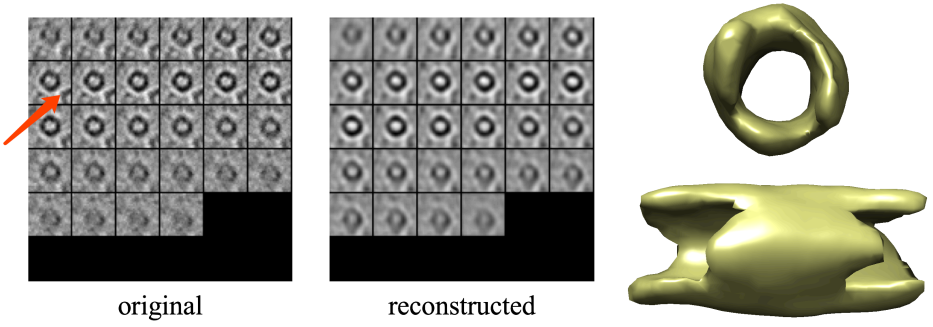

Figure 3: Reconstructed apoferritin macromolecular structure by our proposed 3D autoencoding classifier network. (a) Comparison of original and reconstructed apoferriti subtomogram. Red arrow indicates the neighbor structure in original subtomogram, which is alleviated by reconstruction (b) Reconstructed apoferritin 3D structure from two side views

reaches significantly better performance.

4.3 Experiments on Subtomogram Reconstruction

In this experiment, we compare the reconstruction performance of our approach with the baseline 3D autoencoder model on the two datasets.

Experiment settings All the network architecture and the experiment settings are the same as in Section 4.2 for semi-supervised classification. Our proposed model is trained in the same way as in Section 4.2. The 3D autoencoder model utilizes the structure of the reconstruction branch of our method as its bottleneck architecture, which is regularized by Eq 1 in training phase. We use three settings of the labeled proportion (5%, 10%, 20%) to evaluate the reconstruction performance of our method. The reconstruction performance is measured by Mean Square Error as in Eq 1 in testing set.

Testing results Table 3 shows the reconstruction performance of our approach and the 3D autoencoder model. All settings of our approach outperform the original 3D autoencoder in terms of reconstruction performance. It is also observed that our approach reaches the best reconstruction performance when only 10% of the subtomograms are labeled. In the real dataset, there exists little difference in the reconstruction performance of all three settings. However, in the simulated dataset, the reconstruction performance greatly improves when the labeled portion is increased to 10%. And this performance decreases slightly when increased to 20%. This demonstrates that proper portions of labeled data will benefit the reconstruction performance. However, excessive labeled data will in some ways compete with the reconstruction branch in feature learning, which will adversely affect the reconstruction performance. Figure 4.3 shows the reconstructed apoferritin structure generated by our best performed model as an example. We can see that the reconstructed subtomogram contains less noise compared to the original input subtomogram. Besides, the signal of neighboring structure features is suppressed in our reconstructed subtomogram, which is primarily attributed to the discriminative feature representations learned through supervised classification of labeled data. The reconstructed subtomograms are promising to be utilized for further segmentation and course recovery.

Table 3: Reconstruction performance on real and simulated testing sets. The performance of our model is evaluated under three settings of labeled data proportions in training set: 5%, 10% and 20%. The performance is measured using mean square error between original subtomogram and the reconstructed subtomogram

	Real set	Simulated set
3D Autoencoder	0.0147	1.2464
Our approach (5% labeled)	0.0050	0.8357
Our approach (10% labeled)	**0.0049**	**0.1491**
Our approach (20% labeled)	0.0054	0.1793

5 Conclusion

In this paper, we introduce semi-supervised classification to reduce annotation burden in the subtomogram classification tasks. Specifically, we propose a 3D autoencoding classifier based deep learning framework for simultaneous subtomogram reconstruction and classification. In training phase, we jointly optimize the two branches of network using both labeled and unlabeled data. Extensive experiments demonstrate that our approach is not only able to accurately label the subtomogram but can also improve the reconstruction performance of 3D autoencoder network. The reconstructed subtomogram can greatly reduce image noise and the overlapping with the neighboring structure. Our works act as an important step toward automated structure classification and recovery in CECT. As a future work, we will evaluate our approach on more challenging subtomogram datasets with higher resolution and smaller voxel spacing. We will also integrate our approach to complement other tasks such as subtomogram alignment [19, 33], averaging [8, 39], pattern mining [35], segmentation [17, 31, 32], and tomogram reconstruction [9].

6 Acknowledgements

This work was supported in part by U.S. National Institutes of Health (NIH) grant P41 GM103712. XZ was supported by a fellowship from Carnegie Mellon University's Center for Machine Learning and Health.

References

[1] 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(11s):957, 2000.

[2] Gustavo Camps-Valls, Tatyana V Bandos Marsheva, and Dengyong Zhou. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3044–3054, 2007.

[3] Olivier Chapelle and Alexander Zien. Semi-supervised classification by low density separation. In *AISTATS*, volume 2005, pages 57–64. Citeseer, 2005.

[4] Chengqian Che, Ruogu Lin, Xiangrui Zeng, Karim Elmaaroufi, John Galeotti, and

Min Xu. Improved deep learning-based macromolecules structure classification from electron cryo-tomograms. *Machine Vision and Applications*, 29(8):1227–1236, 2018.

[5] Yueru Chen, Yijing Yang, Min Zhang, and C.-C. Jay Kuo. Semi-supervised learning via feedforward-designed convolutional neural networks. *CoRR*, abs/1902.01980, 2019.

[6] Jonathan Gordon and José Miguel Hernández-Lobato. Bayesian semisupervised learning with deep generative models. *arXiv preprint arXiv:1706.09751*, 2017.

[7] Jialiang Guo, Bo Zhou, Xiangrui Zeng, Zachary Freyberg, and Min Xu. Model compression for faster structural separation of macromolecules captured by cellular electron cryo-tomography. In *International Conference Image Analysis and Recognition*, pages 144–152. Springer, 2018.

[8] Wim JH Hagen, William Wan, and John AG Briggs. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of structural biology*, 197(2):191–198, 2017.

[9] Renmin Han, Zhipeng Bao, Xiangrui Zeng, Tongxin Niu, Fa Zhang, Min Xu, and Xin Gao. A joint method for marker-free alignment of tilt series in electron tomography. *Bioinformatics*, 35(14):i249–i259, 2019. doi: 10.1093/bioinformatics/btz323.

[10] Jingrui He, Jaime G Carbonell, and Yan Liu. Graph-based semi-supervised learning as a generative model. 2007.

[11] Le Hou, Kunal Singh, Dimitris Samaras, Tahsin M Kurc, Yi Gao, Roberta J Seidman, and Joel H Saltz. Automatic histopathology image analysis with cnns. In *2016 New York Scientific Data Summit (NYSDS)*, pages 1–6. IEEE, 2016.

[12] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[13] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in neural information processing systems*, pages 3581–3589, 2014.

[14] Ran Li, Xiangrui Zeng, Stephanie E Sigmund, Ruogu Lin, Bo Zhou, Chang Liu, Kaiwen Wang, Rui Jiang, Zachary Freyberg, Hairong Lv, et al. Automatic localization and identification of mitochondria in cellular electron cryo-tomography using faster-rcnn. *BMC bioinformatics*, 20(3):132, 2019.

[15] Yang Li, Quan Pan, Suhang Wang, Haiyun Peng, Tao Yang, and Erik Cambria. Disentangled variational auto-encoder for semi-supervised learning. *Information Sciences*, 482:73–85, 2019.

[16] Ruogu Lin, Xiangrui Zheng, Kris Kitani, and Min Xu. Adversarial domain adaptation for cross data source macromolecule in situ structural classification in cellular electron cryo-tomograms. *Bioinformatics*, 35(14):i260–i268, 2019. doi: 10.1093/bioinformatics/btz364.

[17] 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. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1578–1582. IEEE, 2018.

[18] Chang Liu, Xiangrui Zeng, Kaiwen Wang, Qiang Guo, and Min Xu. Multi-task learning for macromolecule classification, segmentation and coarse structural recovery in cryo-tomography. In *British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3-6, 2018*, page 271, 2018.

[19] Yongchun Lu, Xiangrui Zeng, Xiaofang Zhao, Shirui Li, Hua Li, Xin Gao, and Min Xu. Fine-grained alignment of cryo-electron subtomograms based on mpi parallel optimization. *BMC bioinformatics*.

[20] Vladan Lučić, Alexander Rigort, and Wolfgang Baumeister. Cryo-electron tomography: the challenge of doing structural biology in situ. *J Cell Biol*, 202(3):407–419, 2013.

[21] Ziqian Luo, Xiangrui Zeng, Zhipeng Bao, and Min Xu. Deep learning-based strategy for macromolecules classification with imbalanced data from cellular electron cryotomography. In *International Joint Conference on Neural Networks (IJCNN)*, 2019.

[22] Xiaorui Ma, Hongyu Wang, and Jie Geng. Spectral–spatial classification of hyperspectral image based on deep auto-encoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4073–4085, 2016.

[23] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.

[24] Alex J Noble, Venkata P Dandey, Hui Wei, Julia Brasch, Jillian Chase, Priyamvada Acharya, Yong Zi Tan, Zhening Zhang, Laura Y Kim, Giovanna Scapin, et al. Routine single particle cryoem sample and grid characterization by tomography. *Elife*, 7: e34257, 2018.

[25] 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.

[26] Thomas Robert, Nicolas Thome, and Matthieu Cord. Hybridnet: Classification and reconstruction cooperation for semi-supervised learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 153–169, 2018.

[27] Chuck Rosenberg, Martial Hebert, and Henry Schneiderman. Semi-supervised self-training of object detection models. 2005.

[28] Maja Stikic, Kristof Van Laerhoven, and Bernt Schiele. Exploring semi-supervised and active learning for activity recognition. In *2008 12th IEEE International Symposium on Wearable Computers*, pages 81–88. IEEE, 2008.

[29] Du Tran, Lubomir D Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. C3d: generic features for video analysis. *CoRR, abs/1412.0767*, 2(7):8, 2014.

[30] Qian Wang, Hang Li, Zhi Chen, Dou Zhao, Shuang Ye, and Jiansheng Cai. Supervised and semi-supervised deep neural networks for csi-based authentication. *arXiv preprint arXiv:1807.09469*, 2018.

[31] 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.

[32] Min Xu, Martin Beck, and Frank Alber. Template-free detection of macromolecular complexes in cryo electron tomograms. *Bioinformatics*, 27(13):i69–i76, 2011.

[33] 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.

[34] Min Xu, Xiaoqi Chai, Hariank Muthakana, Xiaodan Liang, Ge Yang, Tzviya Zeev-Ben-Mordehai, and Eric P Xing. Deep learning-based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *Bioinformatics*, 33(14):i13–i22, 2017.

[35] Min Xu, Jitin Singla, Elitza I Tocheva, Yi-Wei Chang, Raymond C Stevens, Grant J Jensen, and Frank Alber. De novo structural pattern mining in cellular electron cryotomograms. *Structure*, 27(64):679–691, 2019.

[36] Weidi Xu, Haoze Sun, Chao Deng, and Ying Tan. Variational autoencoder for semi-supervised text classification. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[37] Xiangrui Zeng, Miguel Ricardo Leung, Tzviya Zeev-Ben-Mordehai, and Min Xu. A convolutional autoencoder approach for mining features in cellular electron cryotomograms and weakly supervised coarse segmentation. *Journal of structural biology*, 202(2):150–160, 2018.

[38] Guannan Zhao, Bo Zhou, Kaiwen Wang, Rui Jiang, and Min Xu. Respond-cam: Analyzing deep models for 3d imaging data by visualizations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 485–492. Springer, 2018.

[39] Yixiu Zhao, Xiangrui Zeng, Qiang Guo, and Min Xu. An integration of fast alignment and maximum-likelihood methods for electron subtomogram averaging and classification. *Bioinformatics*, 34(13):i227–i236, 2018.

[40] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.