




Improved deep learning-based macromolecules structure classification from electron cryo-tomograms

Chengqian Che¹ · Ruogu Lin² · Xiangrui Zeng³ · Karim Elmaaroufi⁴ · John Galeotti¹ · Min Xu³ 

Received: 14 July 2017 / Revised: 16 January 2018 / Accepted: 18 May 2018 / Published online: 27 June 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Cellular processes are governed by macromolecular complexes inside the cell. Study of the native structures of macromolecular complexes has been extremely difficult due to lack of data. With recent breakthroughs in Cellular Electron Cryo-Tomography (CECT) 3D imaging technology, it is now possible for researchers to gain accesses to fully study and understand the macromolecular structures single cells. However, systematic recovery of macromolecular structures from CECT is very difficult due to high degree of structural complexity and practical imaging limitations. Specifically, we proposed a deep learning-based image classification approach for large-scale systematic macromolecular structure separation from CECT data. However, our previous work was only a very initial step toward exploration of the full potential of deep learning-based macromolecule separation. In this paper, we focus on improving classification performance by proposing three newly designed individual CNN models: an extended version of (Deep Small Receptive Field) DSRF3D, donated as DSRF3D-v2, a 3D residual block-based neural network, named as RB3D, and a convolutional 3D (C3D)-based model, CB3D. We compare them with our previously developed model (DSRF3D) on 12 datasets with different SNRs and tilt angle ranges. The experiments show that our new models achieved significantly higher classification accuracies. The accuracies are not only higher than 0.9 on normal datasets, but also demonstrate potentials to operate on datasets with high levels of noises and missing wedge effects presented.

Keywords Deep learning · Image classification · Medical big data learning · Cellular electron cryo-tomography

1 Introduction

As the basic unit of life, cell has always been a fundamental focus of biomedical research. Governed by macromolecules, cellular processes occur over a large length scale. To fully understand the biological processes at different levels, it is essential to gain knowledge of native structures and spatial organizations of macromolecular complexes inside single

cells. Due to the lack of data acquisition techniques, little has been known about such knowledge due to lack of suitable data acquisition techniques. Recent breakthroughs in Cellular Electron Cryo-Tomography (CECT) imaging technique enables the 3D visualization of macromolecular complex structures and their spatial organizations inside single cells at submolecular resolution and close to their native state [15,19,23,49]. CECT has made possible the discovery of numerous important structural features in prokaryotic cells, eukaryotic cells and viruses [3,11,18,21]. Therefore, CECT emerges as a very promising tool for systematically studying macromolecular complexes with unprecedented coverage, precision and fatality. In principle, a CECT image contains structural information of all macromolecular complexes inside the field of view. However, the systematic recovery of macromolecular structures from CECT is very difficult due to high degree of structural complexity and practical imaging limitations. Specifically, the densely populated cytoplasm makes a very crowded cellular environment for macromolecules. Also, macromolecules are dynamically interacting each other, forming more complex and heteroge-

✉ John Galeotti
jgaleotti@cmu.edu

✉ Min Xu
mxu1@cs.cmu.edu

¹ The Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

² Department of Automation, Tsinghua University, Beijing, China

³ Computational Biology Department, Carnegie Mellon University, Pittsburgh, USA

⁴ Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA

neous structures [28]. On the other hand, current technical limitations inherent to the process of structure determination via single-particle cryo-EM require collecting very large datasets often images of several thousands of macromolecules. It would likely require separating and averaging millions of macromolecules represented by subtomograms, potentially containing hundreds of highly heterogeneous structural classes (a *subtomogram* is a cubic subimage that contains only one macromolecule). Although advances in data acquisition automation make it no longer difficult to acquire CECT images containing such amount of macromolecules, existing computational approaches have very limited scalability and discrimination ability, making them incapable of automatic processing such large amount of data.

Given this challenging task, a number of previous works have been done for analyzing macromolecules from CECT data. In [4,26], template searching-based algorithms were proposed to localize macromolecules of known structures from CECT data. In 2013, Briggs et al. reviewed a number of subtomogram averaging methods to resolve structure of macromolecular complexes in situ [7]. In addition, unsupervised classification-based approaches were also developed (e.g., [2,6,9,32,39]). Even though these methods showed promising macromolecule structure separation and recovery results, the scalability is strictly limited by the intensive computations. Other approaches such as rotation invariant feature [40] and pose normalization [46] were proposed to address the task while reducing the computational complexity. However, these approaches are limited by anisotropic resolution from missing wedge effect, and high level of noise in CECT data.

In order to overcome the limitations mentioned above, recently, we were the *first* to propose deep learning-based approach [44] for separating particles into structurally homogeneous subgroups through supervised feature extraction using Convolutional Neural Network (CNN). Such approach achieved significantly better separation performance in terms of both accuracy and scalability compared with our previous approaches, showing that deep learning-based approach is potentially a very powerful tool for large-scale particle separation. However, our previous proof-of-principle work is only an initial step toward exploring the full power of deep learning-based large-scale particle separation. The accuracy of classification needs to be substantially improved for better structural reconstruction performance.

In this paper, we focus on improving deep learning-based separation of particles of macromolecular complexes extracted from CECT images by designing new CNN models. The three CNN models we are proposing include: an extended version of (Deep Small Receptive Field) DSRF3D [44], donated as DSRF3D-v2, a 3D residual block-based neural network [20], named as RB3D, and a convolutional 3D (C3D) [36]-based model, CB3D. Our experiment shows that

new proposed models can achieve significantly better classification performances than our previous best CNN model proposed in [44]. Among them, CB3D has the best performances and yield accuracy close to 0.9 for normal datasets. Our models also show promising classification performance over datasets of extremely low SNR (0.01).

2 Method

2.1 Convolutional neural networks

Serving as a powerful tool, convolutional deep neural networks have been widely used by researchers to resolve challenging tasks in computer vision especially for image classification. Inspired by biological processes, CNN models are composed of stacked layers including an input, an output and multiple hidden layers, which include convolutional, pooling or fully connected layers. By stacking multiple processing layers, more and more image features are learned and extracted as the training proceeds. More specifically, each convolutional layer contains numerous of filters, considered as neurons with different weights. Neurons in this layer are connected to regions of neighboring neurons in the previous layer, donated as receptive field. For instance, a 1D convolution input x with filter size of $2m + 1$ will yield an output $y_i = \sum_{j=-m}^m w_j x_{i-j}$, where w_j is the j th weight of the convolutional filter. An activation function is applied after the convolution layer. Some common activation layers include sigmoid, tanh, the rectified linear unit (ReLU) [16], Leaky ReLU [20] and Maxout [17]. These activation functions take the input and perform certain fixed mathematical operations on it. They are used to accelerate the convergence of the optimization process. For example, ReLU is defined as $\sigma^{\text{ReLU}}(x) = \max(0, x)$. Next, pooling layers are utilized to reduce the computational costs during the training process by down sampling the data. Two common ways of pooling are calculating the local maximum and average values of the pooling windows. After series of stacked convolutional and pooling layers, a fully connected layer is used to extract more global features. Each unit in fully connected layers, as name suggested, is connected to all units from the previous layer. For instance, given an i th input x_i , the j th output y_j is defined as $y_j = \sum_{i=0}^{n-1} w_{ji} x_i$, where n is the total number of inputs and w_{ji} is the weight between i th input x and j th output y . Sometimes, special techniques such as Dropout [34] and L2 regularization are used to prevent overfitting. Dropout works by simply only keeping a neuron active with some probability p or setting it to zero otherwise during the training process. Lastly, in order to perform the multi-class classification, a softmax activation function is connected to the last fully connected layer to compute a probability of a sample

being assigned to each class. The softmax function is defined in 1, where $f_j(x) = x^T w_j$. w_j are the weights with j th class and $P(j|x)$ is the probability of the subtomogram is assigned to j class.

$$o_j^{\text{softmax}}(x) = P(j|x) = \frac{e^{f_j(x)}}{\sum_{l=1}^L e^{f_l(x)}} \quad (1)$$

The input of this classification problem is a subtomogram X , (i.e., 3D gray scale image of size $n_1 \times n_2 \times n_3$), represented as a 3D array of $\mathbb{R}^{n_1 \times n_2 \times n_3}$, and the output is a label vector L , represented as a 1D array of \mathbb{R}^l , where l denotes the number of classes in the dataset. The algorithm aims at classifying subtomograms into the correct class, that is, mapping X to l correctly.

Designing CNN architectures and tuning parameters are essential for the performance of networks. In 2012, Krizhevsky et al. proposed a novel CNN architecture AlexNet [22], which was the first to show a significant improvement of image classification results on a historically difficult dataset, ImageNet [31]. From that on, CNNs have become a household name in computer vision computer community. In recent years, more advanced CNN architectures were proposed and developed such as GoogleNet [35], ResNet [20] and VGGNet [33]. These networks gradually pushed the classification error rate on ImageNet down to 3.6% [20].

In this paper, we propose three different CNN models and comparing them with our previously proposed approaches in [44].

All the models are trained using stochastic gradient descent (SGD) optimizer. We minimize the categorical cross-entropy cost function by adding Nesterov momentum of 0.9. Momentum update [30] is an update approach that often has better converge rates while using gradient descent on deep network optimization. With momentum update, the parameter vector will build up velocity in any direction that has consistent gradient. Nesterov momentum [24] is a slightly different version of the momentum update that enjoys stronger theoretical converge guarantees for convex functions and in practice it also consistently works slightly better than standard momentum. The main difference is in classical momentum it first corrects the velocity and then makes a big step according to that velocity, but in Nesterov momentum it first makes a step into velocity direction and then is corrected to a velocity vector based on new location. Previous work [24] mathematically proves that Nesterov momentum has a better convergence rate in optimization. In addition, the initial learning rate is set at 0.005 with a decay factor of $1e-7$. The training processes are performed with a batch size of 64 for 20 epochs. However, for each dataset, if the classification performance shows no improvement over 5 consecutive epochs based on the loss function, the training process will end early.

2.1.1 DSRF3D-v2 model

In this section, we propose a 3D variant VGGNet [33]-based CNN architecture called Deep Small Receptive Field (DSRF3D). This model is an extended version of our previously proposed model [44], and we donate this model as DSRF3D-v2. Just like a VGGNet, DSRF3D-v2 is featured with sequentially deep stacked layers and small 3D convolution filters with size of $3 \times 3 \times 3$. As shown in Fig. 1, the input layer is sequentially connected three sets of stacked layers, with each set consisting of 2 $3 \times 3 \times 3$ 3D convolutional layers and one $2 \times 2 \times 2$ 3D max pooling layer. Then, it is followed by two fully connected layers with 70% dropout after each layer. The final fully connected output layer has the same number of units as the structure class number. The activation layers are ReLU for all hidden layers and softmax for fully connected layers. Compared to previously designed DSRF3D model, adding more stacked layers with appropriate dropouts should improve the classification performance intuitively.

2.1.2 RB3D model

In this section, a 3-D variant residual block-based [20] CNN model is proposed, donated as RB3D. One big advantage for ResNet-based model is that it avoids negative outcomes while increasing the network overall depth. As shown in Fig. 2, this model feeds the input layer to a convolutional layer, a ReLU activation layer and a $2 \times 2 \times 2$ 3D max-polling layer. Then, four bottleneck [20] residual blocks are connected sequentially. For each block, there are two paths that merge together at the end of the block. One path contains one 1×1 layer to reduce dimension, a 3×3 layer and a 1×1 layer for restoring dimension. The other path, considered as a “shortcut”, only contains a 3×3 convolutional layer. Lastly, two fully connected layers are constructed with dropout of 50% to prevent overfitting. The usage of residual blocks might lead to higher classification accuracy.

2.1.3 CB3D model

In this section, we propose a 3D convolutional (C3D)-based model, named as CB3D. C3D [36] was originally proposed to be trained on large-scale supervised video datasets. We can think of the 3D structures of macromolecules as multiple slices of 2D images. If we look from the first slice to the last slice, we can interpret it as a continuously changing object just like a video dataset. In our model shown in Fig. 3, we concatenate eight $3D \ 3 \times 3 \times 3$ convolutional layers and each layer is activated by ReLU. Five max pooling layers are mixed among the convolutional layers. At end, two fully connected layers with 50% dropout are added and a softmax activation is appended.

Fig. 1 DSRF3D-v2 model: Each box provides configurations for each layer. '32-3 × 3 × 3-1 Conv' represents a 3D convolutional layer with 32 $5 \times 5 \times 5$ filters and stride of 1. 'ReLU' and 'Softmax' are activation layers. '2 × 2 × 2-1 MaxPool' means that max operation is implemented over $2 \times 2 \times 2$ regions with stride of 2'. 'FC-1024' and 'FC-L' represents fully connected layers with 1024 and L(total number of the classes) neurons, respectively

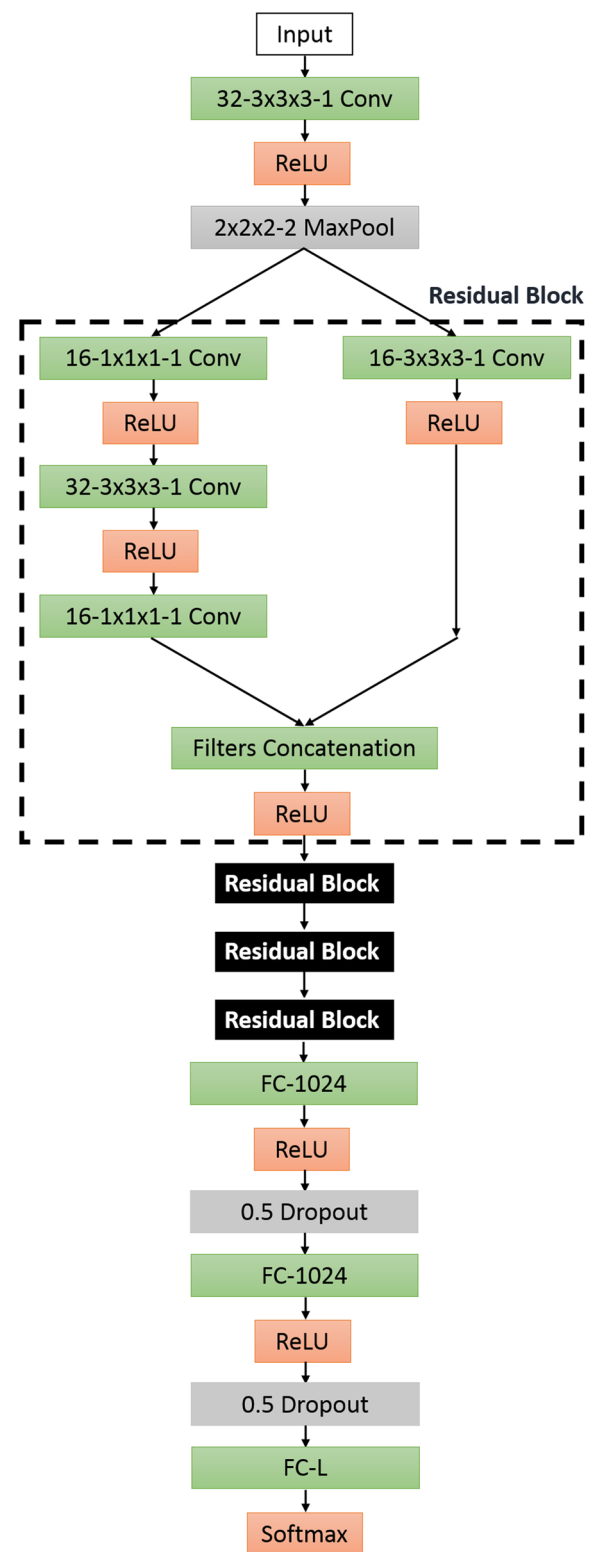
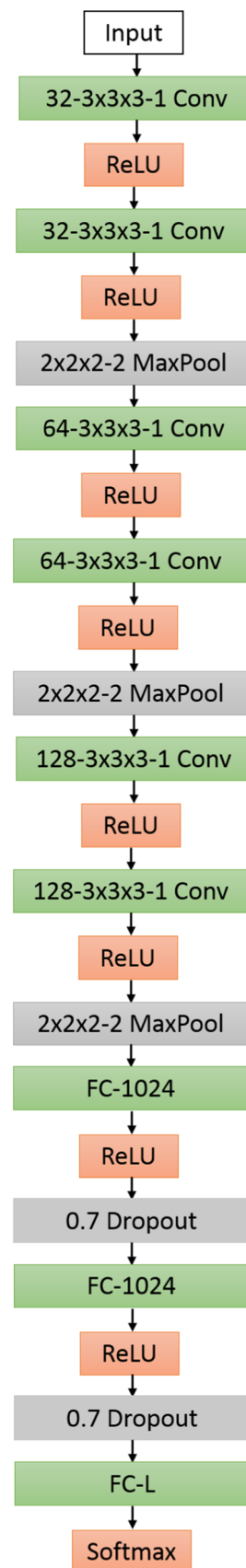
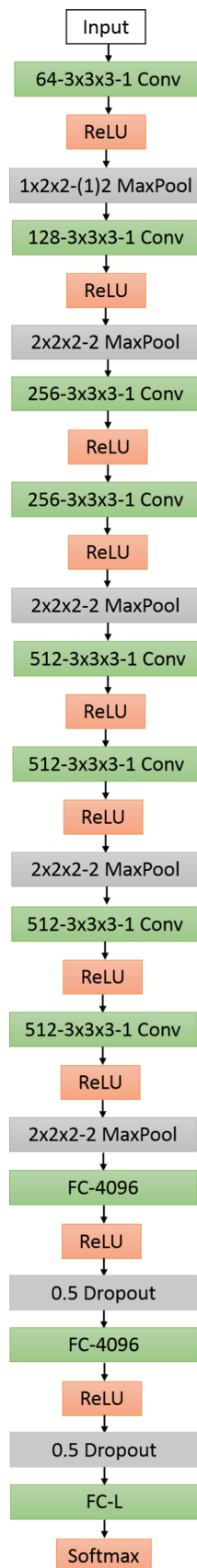


Fig. 2 RB3D model: each box provides configurations for each layer. The definition of the boxes follows Fig. 1. Four residual blocks are connected, represented by the black boxes. The specific design for a single residual block is shown in the dashlined box

Fig. 3 CB3D model: each box provides configurations for each layer. The definition of the boxes follows Fig. 1. Note that ‘ $1 \times 2 \times 2-(1)2$ MaxPool’ means that the max operation is implemented over region of $1 \times 2 \times 2$. The stride is 1 for the first dimension and 2 for the others



2.2 Generation of simulated subtomograms from experimental structures

Similar to previous works [4,12,25,29,39,39,42–45,48], we use known structures of macromolecular complexes to generate simulated subtomograms by simulating actual tomographic image reconstruction processes in order to have a reliable assessment of our proposed approaches. There are three significant aspects we focus on when simulating the subtomograms: missing wedge effects, noises and electron optical factors such as modulation transfer function (MTF) and contrast transfer function (CTF).

The contrast transfer function (CTF) mathematically describes how aberrations in a transmission electron microscope (TEM) modify the image of a sample. By considering the recorded image as a CTF-degraded true object, describing the CTF allows the true object to be reverse engineered. This is typically denoted CTF-correction, and is vital to obtain high resolution structures in three-dimensional electron microscopy, especially cryo-electron microscopy. Its equivalent in light-based optics is the optical transfer function (OTF). A typical CTF is of the following form (Eq. 2.20 of [13]):

$$\text{CTF}(\hat{k}; \Delta\hat{z}) = \sin[-\pi \Delta\hat{z} \Delta\hat{k}^2 + \frac{\pi}{2} \hat{k}^4] \quad (2)$$

where \hat{z} and \hat{k} are generalized defocus and spatial frequency, respectively.

The modulation transfer function (MTF) is a variant of the optical transfer function (OTF), neglecting phase effects. The resolution and performance of an optical microscope can be characterized by a MTF, which is a measurement of the microscope's ability to transfer contrast from the specimen to the intermediate image plane at a specific resolution. Computation of the modulation transfer function is a mechanism that is often utilized by optical manufacturers to incorporate resolution and contrast data into a single specification.

More specifically, we first generate density map volumes of 40^3 voxels with a resolution of 0.92 nm using the PDB2VOL program from the Situs [38] package. The density map volumes are generated by convoluting the atomic structures with a Gaussian kernel, whose standard deviation is assumed to be half the target resolution. We then randomly rotate and translate the volumes. Next, we generate projection images of the density maps with different tilt angles to simulate missing wedge effects. The specific tilt angle ranges are $\pm 60^\circ$, $\pm 50^\circ$ and $\pm 40^\circ$. We then convolute the projection images with CTF and MTF [13,25] to reproduce the electron optical effects to generate simulated electron micrographic images. Then, the simulated noises are added to electron micrographic images [12] with desired signal-to-noise ratio (SNR) levels so that the SNR of reconstructed subtomograms

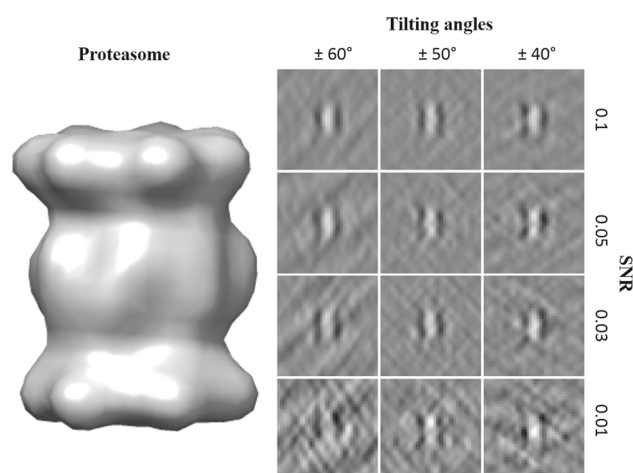


Fig. 4 Left: Isosurface of Yeast 20S proteasome (PDB ID: 3DY4); Right: Center slices of subtomograms with different levels of SNRs (0.5, 0.1, 0.05 and 0.01) and tilt angle ranges ($\pm 60^\circ$, $\pm 50^\circ$ and $\pm 40^\circ$)

are of 0.1, 0.05, 0.03 and 0.01. The acquisition parameters are set similar to [47], with spherical aberration of 2 mm, defocus of $-5\mu\text{m}$ and voltage of 300kV. Finally, with all gathered information, we construct the simulated subtomogram datasets using a direct Fourier inversion reconstruction algorithm implemented in the EMAN2 library [14]. Figure 4 shows an example of center slices of simulated subtomograms with different SNRs and tilt angle ranges.

We construct a simulated dataset for each pair of SNRs(4) and tilt angle ranges(3), which yields 12 sets of data in total. Within a single dataset, for each macromolecular complex, we generate 1000 simulated subtomograms that contain randomly rotated and translated particle of that complex. There are 22 macromolecular complexes collected from the Protein Databank (PDB) [5] shown in Fig. 5. Furthermore, we simulate 1000 subtomograms that contain no macromolecule. As an outcome, each dataset contains 23,000 simulated subtomograms of 23 structural classes.

To fully evaluate classification performances, we first split each dataset into two parts: 80% are used as training data, and 20% are used as testing data. Then, we take 20% of the training data and use it as validation data for parameter tuning during training process, and the rest 80% for training weights. Therefore, we end up with 14,720 training samples, 3680 validation samples and 4600 testing samples. We use the same partitioned datasets across all models to have a fair comparison on their classification performance.

2.3 Implementation details

This work is implemented using Keras [10] with TensorFlow [1] as back-end. Keras is a python-based, high-level neural networks API for fast deep learning experimentation. The experiments are performed on a computer with three Nvidia GTX 1080 GPUs, one Intel Core i7-6800K CPU and 128GB memory. The new proposed models are implemented with the same system as our previously proposed model [44]. For the baseline methods, the rotation invariant features is computed based on SHTools [37]. *K*-means clustering and support vector machine (SVM)-based supervised multi-class classification are implemented using the Sklearn toolbox [27].

3 Experiment results

3.1 Classification performance

In this section, we compare the classification performance of new CNN models (DSRF3D-v2, RB3D and CB3D) with our previously proposed best model DSRF3D [44], as well as a baseline method on datasets with different SNRs and tilt angles ranges. The baseline method uses spherical harmonics rotation invariant feature [41,43] with SVM using Radial Basis function kernel, denoted as RIF-SVM. The results are shown in Table 1. The best performances for each

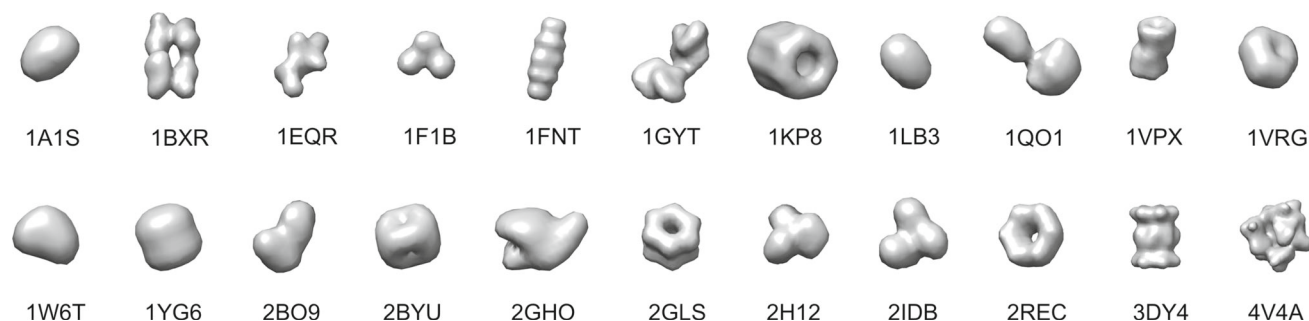


Fig. 5 Isosurfaces of all 22 types of macromolecular complexes collected from the Protein Databank (PDB), denoted with PDB ID

Table 1 The classification accuracy using four CNN models under different SNR and tilt angles

SNR/tilt angle range	$\pm 60^\circ$						$\pm 50^\circ$						$\pm 40^\circ$							
	RIF-SVM		DSRF3D		DSRF3D-v2		RB3D		CB3D		RIF-SVM		DSRF3D		DSRF3D-v2		RB3D		CB3D	
0.1	0.790	0.911	0.977	0.950	0.973	0.672	0.896	0.963	0.925	0.971	0.670	0.868	0.954	0.899	0.970					
0.05	0.620	0.844	0.925	0.852	0.933	0.493	0.753	0.910	0.750	0.899	0.495	0.735	0.876	0.671	0.877					
0.03	0.479	0.706	0.841	0.711	0.849	0.350	0.581	0.746	0.548	0.747	0.360	0.537	0.038	0.502	0.717					
0.01	0.218	0.040	0.407	0.041	0.445	0.153	0.200	0.041	0.042	0.041	0.162	0.043	0.043	0.171	0.041					

pair of imaging conditions are highlighted in bold. In general, our new proposed CNN models demonstrate significant improvements in classification performance. As shown, the best performance is achieved by newly designed models in all 11 out of 12 situations. The only exception that new models perform worse than previous models (DSRF3D and RIF-SVM) is with tilt angle ranges of $\pm 50^\circ$ and SNR of 0.001. Given such poor qualities of images, the “best” performance is only around 0.2, which is still considered as bad performance.

When comparing among the three new models, we can observe that RB3D often ends up with lower accuracy than DSRF3D-v2 and CB3D. In fact, RB3D establishes obvious improvements (greater than 0.3) only with SNR of 0.1. In situations with low SNRs, RB3D does not show apparent improvement and it performs even worse than DSRF3D. Therefore, the RB3D model is not robust to image noises. In general, residual block structures are proposed and proved to work the best with very deep neural network structures [20]. We also tested increasing the depth of RB3D model by adding one more residual block and train the new structure on three datasets with SNR of 0.01. The accuracies using 5 residual blocks structures are 0.247, 0.042 and 0.171, compared to 0.041, 0.042 and 0.171 when using 4 residual blocks. This exploration shows that there is not clear pattern of classification performance improvement when increasing the depth of RB3D. Thus, further study on improving classification performance through increasing depth or modifying residual block internal structure remains open.

In contrast, both DSRF3D-v2 and CB3D show essential accuracy increases in most cases. With SNR of 0.01 and tilt angles of $\pm 60^\circ$, both models can remarkably achieve accuracy greater than 0.4. Even though 0.4 is still not considered as high classification accuracy in Table 1, it at least shows the capability of deep learning to classify a large amount of subtomograms with extremely poor image quality.

DSRF3D-v2 and CB3D generate very similar results, usually with differences less than 0.1. It is proved that both model can accurately classify subtomograms even with much noises and missing wedge effects presented. CB3D is slightly better than DSRF3D-v2 because it achieves the best performance in 8 out of 12 datasets. With SNR of 0.03 and tilt angle ranges of $\pm 40^\circ$, CB3D can still obtain classification accuracy higher than 0.7, which is a very good performance if we consider the poor image qualities.

In addition, we also evaluate the computation time. On average, it takes 0.0056, 0.0083, 0.0141 and 0.0136s for DSRF3D, DSRF3D-v2, RB3D and CB3D, respectively, to process one subtomogram for each epoch during training process. In testing process, it takes 0.0019, 0.0027, 0.0026 and 0.0040s for DSRF3D, DSRF3D-v2, RB3D and CB3D, respectively, to process one subtomogram. Thus, compared to previous model (DSRF3D), our new proposed models cost

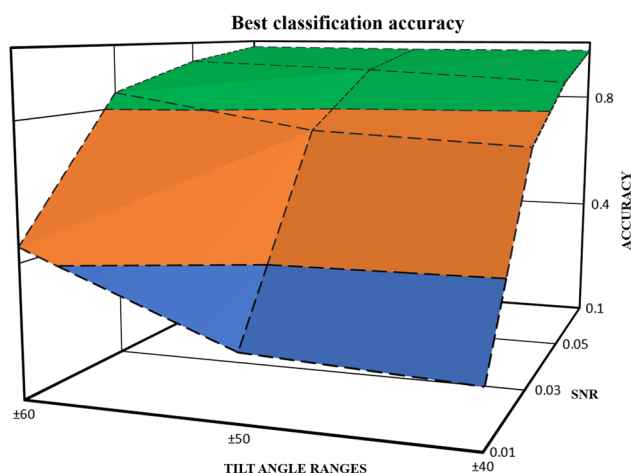


Fig. 6 The highest classification accuracy with respect to different SNRs and tilt angle ranges

48.21–151.79% more time for training and 36.84–110.53% more time for testing. Further work can be done to speed up our new deep learning models.

3.2 Classification capability

In this section, we examine the capability of deep learning to classify large-scale datasets of subtomograms. We first extract the the best performance for each of the 12 datasets and then plot the highest accuracy with respect to both SNRs and tilt angle ranges. The 3D surface is plotted in Fig. 6. As shown, the classification accuracy decreases as more noises are added to the dataset for all tilt angles. Similarly, for all SNRs, the accuracy will reduce if tilt angle ranges decrease from $\pm 60^\circ$ to $\pm 40^\circ$. Based on the plot, we can observe that for datasets whose SNRs are above or equal to 0.05, our best model can achieve classification accuracy no lower than 0.877. For datasets with poor image qualities, as long as SNR is kept above 0.03, classification can achieve higher than 0.7 for all 3 tilt angle ranges. It is proven that our proposed approach has strong abilities to accurately classify macromolecular structures from CECT images and even greater potentials to process datasets with extremely high level of noises and missing wedge effects.

4 Conclusions

In this paper, three novel CNN models are proposed to significantly improve deep learning-based separation of macromolecules extracted from CECT images. We compare them with our previously proposed model and our best model CB3D ends up with classification accuracy of approximately 0.85 for image datasets with relatively low noise level. More importantly, it demonstrates good potentials to operate

on datasets with extremely poor image qualities. After successfully and efficiently subdividing the subtomograms, the computationally intensive reference-free approaches can be applied to selected subsets separately in order to recover the structure of macromolecular complexes. The overall computational cost can be greatly reduced through such divide and conquer approach. This proof-of-principle work represents a useful step toward full systematic structural separation and recovery of millions of macromolecules extracted from CECT images.

Recently, deep learning has also been used for analyzing CECT by other groups, for example [8]. There are several major differences between our approaches and the approach in [8]. (1) We focus on the recovery of macromolecular complexes captured by CECT, instead of supervised segmentation of ultrastructures as in [8]. (2) Because the macromolecules can have arbitrary orientation in a CECT image, our models aim at learning the rotational invariant features. We used 3D filters instead of 2D filters as in [8], so that our CNN models can isotropically capture the inherent 3D spatial structure in such 3D images.

Acknowledgements This work was supported in part by U.S. National Institutes of Health (NIH) Grant P41 GM103712. John Galeotti acknowledges support from NIH R01 Grant 1R01EY021641, National Library of Medicine contract HHSN276201000580P and DoD Peer Reviewed Medical Research Program (PR130773, HRPO Log No. A-18237). Min Xu acknowledge support of Samuel and Emma Winters Foundation.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. et al.: Tensorflow: a system for large-scale machine learning (2016). [arXiv:1605.08695](https://arxiv.org/abs/1605.08695)
2. Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., Subramaniam, S.: Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* **162**(3), 436–450 (2008)
3. Beck, M., Lui, V., Förster, F., Baumeister, W., Medalia, O.: Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**(7162), 611–615 (2007)
4. Beck, M., Malmström, J.A., Lange, V., Schmidt, A., Deutsch, E.W., Aebersold, R.: Visual proteomics of the human pathogen *Leptospira interrogans*. *Nat. Methods* **6**(11), 817–823 (2009)
5. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. *Nucl. Acids Res.* **28**(1), 235 (2000)
6. Bharat, T.A.M., Russo, C.J., Löwe, J., Passmore, L.A., Scheres, S.H.W.: Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**(9), 1743–1753 (2015)
7. Briggs, J.A.G.: Structural biology in situ the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**(2), 261–267 (2013)
8. Chen, M., Dai, W., Sun, 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 (2017). [arXiv:1701.05567](https://arxiv.org/abs/1701.05567)

9. Chen, X., Chen, Y., Schuller, J.M., Navab, N., Forster, F.: Automatic particle picking and multi-class classification in cryo-electron tomograms. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 838–841. IEEE (2014)
10. Chollet, F.: keras (2015). <https://github.com/fchollet/keras>. Accessed 10 May 2017
11. Delgado, L., Martínez, G., López-Iglesias, C., Mercadé, E.: Cryo-electron tomography of plunge-frozen whole bacteria and vitreous sections to analyze the recently described bacterial cytoplasmic structure, the stack. *J. Struct. Biol.* **189**(3), 220–229 (2015)
12. Förster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S.: Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* **161**(3), 276–286 (2008)
13. Frank, J.: Three-dimensional electron microscopy of macromolecular assemblies. Oxford University Press, New York (2006)
14. Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J.: Single particle tomography in eman2. *J. Struct. Biol.* **190**(3), 279–290 (2015)
15. Gan, L., Jensen, G.J.: Electron tomography of cells. *Q. Rev. Biophys.* **45**(01), 27–56 (2012)
16. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). <http://www.deeplearningbook.org>. Accessed 15 June 2017
17. Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout networks. In: Dasgupta S., McAllester D. (eds.) Proceedings of the 30th International Conference on Machine Learning, volume 28 of Proceedings of Machine Learning Research, pp. 1319–1327, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR
18. Grünwald, K., Desai, P., Winkler, D.C., Heymann, J.B., Belnap, D.M., Baumeister, W., Steven, A.C.: Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**(5649), 1396–1398 (2003)
19. Grünwald, K., Medalia, O., Gross, A., Steven, A.C., Baumeister, W.: Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding. *Biophys. Chem.* **100**(1), 577–591 (2002)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015). [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
21. Jasnin, M., Ecke, M., Baumeister, W., Gerisch, G.: Actin organization in cells responding to a perforated surface, revealed by live imaging and cryo-electron tomography. *Structure* **24**(7), 1031–1043 (2016)
22. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)
23. Lučić, V., Rigort, A., Baumeister, W.: Cryo-electron tomography: the challenge of doing structural biology in situ. *J. Cell Biol.* **202**(3), 407–419 (2013)
24. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady* **27**, 372–376 (1983)
25. Nickell, S., Förster, F., Linaroudis, A., Net, W.D., Beck, F., Hegerl, R., Baumeister, W., Plitzko, J.M.: TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* **149**(3), 227–234 (2005)
26. Nickell, S., Kofler, C., Leis, A.P., Baumeister, W.: A visual approach to proteomics. *Nat. Rev. Mol. Cell Biol.* **7**(3), 225–230 (2006)
27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
28. Pei, L., Xu, M., Frazier, Z., Alber, F.: Simulating cryo electron tomograms of crowded cell cytoplasm for assessment of automated particle picking. *BMC Bioinform.* **17**, 405 (2016)
29. Pei, L., Xu, M., Frazier, Z., Alber, F.: Simulating cryo electron tomograms of crowded cell cytoplasm for assessment of automated particle picking. *BMC Bioinform.* **17**(1), 405 (2016)
30. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **4**(5), 1–17 (1964)
31. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
32. Scheres, S.H.W., Melero, R., Valle, M., Carazo, J.M.: Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure* **17**(12), 1563–1572 (2009)
33. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
34. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
35. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning (2016). [arXiv:1602.07261](https://arxiv.org/abs/1602.07261)
36. Tran, D., Bourdev, L.D., Fergus, R., Torresani, L., Paluri, M.: C3D: generic features for video analysis. *CoRR abs/1412.0767* **2**(7), 8 (2014)
37. Wiecek, M., Mesch, M., Sales de Andrade, E., Oshchepkov, I., Herxhbd: Shtools/shtools: Version 4.0, Dec. 2016
38. Wriggers, W., Milligan, R.A., McCammon, J.A.: Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**(2–3), 185–195 (1999)
39. Xu, M., Beck, M., Alber, F.: High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching. *J. Struct. Biol.* **178**(2), 152–164 (2012)
40. Xu, M., Li, W., James, G.M., Mehan, M.R., Zhou, X.J.: Automated multidimensional phenotypic profiling using large public microarray repositories. *Proc. Natl. Acad. Sci.* **106**(30), 12323–12328 (2009)
41. Xu, M., Zhang, S., Alber, F.: 3d rotation invariant features for the characterization of molecular density maps. In: 2009 IEEE International Conference on Bioinformatics and Biomedicine, pp. 74–78. IEEE (2009)
42. Xu, M., Alber, F.: 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)
43. Xu, M., Beck, M., Alber, F.: Template-free detection of macromolecular complexes in cryo electron tomograms. *Bioinformatics* **27**(13), i69–i76 (2011)
44. Xu, M., Chai, X., Muthakana, H., Liang, X., Yang, G., Zeev-Ben-Mordehai, T., Xing, E.: Deep learning based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. ISMB/ECCB 2017, Bioinformatics (2017, in press). Preprint. [arXiv:1701.08404](https://arxiv.org/abs/1701.08404)
45. Xu, M., Tocheva, E.I., Chang, Y.-W., Jensen, G.J., Alber, F.: De novo visual proteomics in single cells through pattern mining (2015). [arXiv:1512.09347](https://arxiv.org/abs/1512.09347)
46. Xu, X.P., Page, C., Volkmann, N.: Efficient Extraction of Macromolecular Complexes from Electron Tomograms Based on Reduced Representation Templates. Springer, Berlin (2015)
47. Zeev-Ben-Mordehai, T., Vasishtan, D., Durán, A.H., Vollmer, B., White, P., Pandurangan, A.P., Siebert, C.A., Topf, M., Grünwald, K.: Two distinct trimeric conformations of natively membrane-anchored full-length herpes simplex virus 1 glycoprotein b. *Proc. Natl. Acad. Sci.* **113**(15), 4176–4181 (2016)

48. Zeng, X., Leung, M.R., Zeev-Ben-Mordehai, T., Xu, M.: A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation. *J. Struct. Biol.* <https://doi.org/10.1016/j.jsb.2017.12.015> (2017). [arXiv:1706.04970](https://arxiv.org/abs/1706.04970)
49. Zhang, P.: Correlative cryo-electron tomography and optical microscopy of cells. *Curr. Opin. Struct. Biol.* **23**(5), 763–770 (2013)

Chengqian Che received an M.Sc. degree in Biomedical Engineering from Carnegie Mellon University (CMU) and a B.S. degree in Biomedical Engineering from University of Connecticut. He is currently pursuing his Ph.D. in Robotics at CMU. His current research focus is on differentiable rendering and computational imaging.

Ruogu Lin received B.E. degree in Automation from Tsinghua University, China, in 2018. He is pursuing his Ph.D. degree at Department of Computational Biology, Carnegie Mellon University, USA. His research focus is on computer vision and data-driven methods in computational analysis of cellular electron cryotomography.

Xiangrui Zeng received B.S. degree in Neuroscience from University of Pittsburgh, USA, in 2017. He is currently pursuing his Ph.D. degree at Department of Computational Biology, Carnegie Mellon University, USA. His research focus is on computational analysis of cellular electron cryotomography using computer vision techniques.

Karim Elmaaroufi received a B.S. and M.S. in Electrical and Computer Engineering from Carnegie Mellon University (CMU). His academic focus is on the design of embedded systems.

Dr. John Galeotti is a Systems Scientist and Adjunct Assistant Professor at Carnegie Mellon University (CMU), directing the Biomedical Image Guidance Laboratory. He received a Ph.D. in Robotics at CMU and a B.S. and M.S. in Computer Engineering at North Carolina State University. Dr. Galeotti's research focus is on applying novel, real-time computer-controlled optics, image analysis, and visualization approaches to build and control unique experimental systems for image-guided interventions, diagnosis, and biomedical research.

Dr. Min Xu is an Assistant Research Professor at the Computational Biology Department in the School of Computer Science at Carnegie Mellon University. He received a B.E. in Computer Science from the Beihang University, M.Sc. from School of Computing at the National University of Singapore, M.A. in Applied Mathematics from the University of Southern California (USC), and Ph.D. in Computational Biology and Bioinformatics from USC. He was a postdoctoral researcher at USC. Dr. Xu's career has centered on developing computational methods for the study of cellular systems using imaging and omics data.