Open-set Recognition of Unseen Macromolecules in Cellular Electron Cryo-Tomograms by Soft Large Margin Centralized Cosine Loss

Xuefeng Du
xuefengdu1@gmail.com
Xiangrui Zeng
xiangruz@andrew.cmu.edu
Bo Zhou
bzhou2@andrew.cmu.edu
Alex Singh
alexs1@andrew.cmu.edu
Min Xu (Corresponding Author)
mxu1@cs.cmu.edu

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

Abstract

Cellular Electron Cryo-Tomography (CECT) is a 3D imaging tool that visualizes the structure and spatial organization of macromolecules at sub-molecular resolution in a near native state, allowing systematic analysis of seen and unseen macromolecules. Methods for high-throughput subtomogram classification on known macromolecules based on deep learning have been developed. However, the learned features guided by either the regular Softmax loss or traditional feature descriptors are not well applicable in the open-set recognition scenarios where the testing data and the training data have a different label space. In other words, the testing data contain novel structural classes unseen in the training data. In this paper, we propose a novel loss function for deep neural networks to extract discriminative features for unseen macromolecular structure recognition in CECT, called Soft Large Margin Centralized Cosine Loss (Soft LMCCL). Our Soft LMCCL projects 3D images into a normalized hypersphere that generates features with a large inter-class variance and a low intra-class variance, which can better generalize across data with different classes and in different datasets. Our experiments on CECT subtomogram recognition tasks using both simulation data and real data demonstrate that we are able to achieve significantly better verification accuracy and reliability compared to classic loss functions. In summary, our Soft LMCCL is a useful design in our detection task of unseen structures and is potentially useful in other similar open-set scenarios.

1 Introduction

The cell is one of the most important structural units of living organisms. Most cellular processes are governed by macromolecules. To further understand their functions and in-

teractions, Cellular Electron Cryo-Tomography (CECT) has been proposed to visualize the structure and organization of macromolecules at sub-molecular resolution in a near native state. However, the systematic recovery of macromolecules captured by CECT are challenging due to both a high level of structural complexity and imaging limits. Particularly, the images are captured at a low Signal-to-Noise Ratio (SNR) which are hard to identify via simple inspection. To aid in automatically identifying single particles in subtomograms (A *subtomogram* is a subvolume of a tomogram that is likely to contain a single macromolecule) for structural recovery, several classification methods are proposed [22, 26]. Recently, Convolutional Neural Network (CNN) based methods were developed for supervised segmentation and classification in CECT tasks [1, 2, 10, 13, 28]. However, a majority of them are only evaluated in close-set protocols where the label space in the testing set is identical to the training set, allowing low generalization and discrimination ability for identifying unseen macromolecular structures. Since the native structures of most macromolecular complexes are unknown [6], only a limited amount of paired known image and labels can be used for training, which cannot cover every possible structure appearing in real-world applications. Sometimes, even if we have sufficient labeled training data in one domain (*i.e* dataset in our case), the macromolecules in the testing set may come from another domain. The domain shift between them prevents our model from an increased performance. Therefore, a discriminative model that can detect novel structures in both inter-domain or cross-domain scenarios is particularly desirable. In previous works, unsupervised subtomogram classification or pattern mining approaches [26] were used to recover the representative structures in these novel data. The particles of the new structures can be further included into training data for more comprehensive disentangling of structural features with enhanced discrimination ability [25]. However, they require complex processing pipeline and are computationally intensive. In this paper, we aim to build a discriminative macromolecule recognition framework under the open-set protocol to facilitate unseen structure identification and recovery without retraining. Compared to the close-set setting, open-set recognition requires that the testing set has a different label space from the training set with unseen classes.

Since the regular Softmax loss cannot handle both the label prediction and feature learning because of its low discrimination capacity, several methods have been proposed to mitigate this issue. The majority of them focus on the design of loss functions in order to learn efficient discriminative feature representations, such as contrastive loss [4] and triplet loss [16]. Most of them are *verification* (given an image pair, identify whether they belong to the same class or not) losses rather than *identification* (given a probe image, find its identity by comparing with other images in the testing set) losses. These loss functions ensure that the extracted features have low intra-class variance and high inter-class variance in Euclidean space. Besides, the combination of identification and verification losses have been recently proposed for mutual-complementary effects. For instance, Wen *et al.* [19] proposed Center loss and combined it with the Softmax loss in order to aggregate the features from the same class to the class center. [21] combined the Softmax loss with exclusive loss to jointly train one-shot Person re-ID tasks. Furthermore, the newly-proposed Large Margin Cosine Loss (LMCL) [17] achieved a state-of-the-art facial recognition result by learning an optimal feature representation space. In the hyperspace, the features from different classes are uniformly and sparsely distributed by setting a cosine margin m in the angular space. However, little attention is paid to intra-class variations. By contrast, the Center loss in [19] can pull features from the same class as close as possible which is helpful empirically to rectify the potential bias in the LMCL-driven open-set recognition tasks.

To address the open-set macromolecule recognition problem, we propose a novel loss

function called Soft Large Margin Centralized Cosine Loss (Soft LMCCL) by combining Center loss, Softmax loss and LMCL together. Specifically, we jointly enlarge the separability of features from different classes with LMCL and Softmax loss while Center loss forces the features learned by LCML and Softmax loss to be more centralized for each class. Given Soft LMCCL, we formulate the general workflow of the open-set subtomogram recognition in Figure 1. For evaluation, we see if our model can correctly compare the relationship between the given image pairs in the testing set, which is referred to the verification task in Figure 1. Our experiments on both intra-domain (*i.e* dataset from same acquisition protocols) and cross-domain (*i.e* dataset from different acquisition protocols) open-set recognition demonstrate that our Soft LMCCL can better train the deep model for unseen macromolecular structure recognition task.

Contributions: The proposed method enables our model to detect the unknown macromolecular structures in the wild, which significantly improves the generalization ability of supervised classification models. Our contributions are three-fold:

- We present a novel loss function, Soft LMCCL, for 3D open-set macromolecule recognition by combining three effective loss units, such that the model learned on known structures can be utilized to detect large amounts of unseen macromolecule structures in CECT.

- In our Soft LMCCL, we add two constraints on the distances of subtomogram representations globally and locally. The learned representations are optimized within every class in the training set which gain better discriminative ability for subsequent recognition tasks.

- We conduct both intra-domain subtomogram recognition and cross-domain subtomogram recognition experiments to evaluate the performance. In the challenging cross-domain setting where the label space in the testing and training set is not only non-overlapping but also from different datasets, our method also demonstrates superior performance.

2 Method

2.1 Problem Formulation

Our proposed method aims to solve an important problem of open-set subtomogram recognition which is not covered by deep learning based CECT researches. Suppose we have a training dataset D_A which contains n training subtomograms $D_i^A, i = 1, ..., n$ from label space A. However, the testing set consists of novel data $D_i^A, i = n+1, ..., n+m$ from a different label space in the same domain or data $D_i^B, i = n+1, ..., n+m$ in a different domain. Instead of direct classification, we desire to learn a discriminative embedding space for subtomogram verification task. Given millions of query macromolecule structures X, we compare each of them with the known structures and the remaining structures in X in order to determine if the macromolecule is novel or not and what relationship it has with other query structures (In this case, the class distribution in X becomes apparent for biologists). We propose a new loss function in the CNN embedding module which ensures a promising verification accuracy and a safe and reliable recognition result.

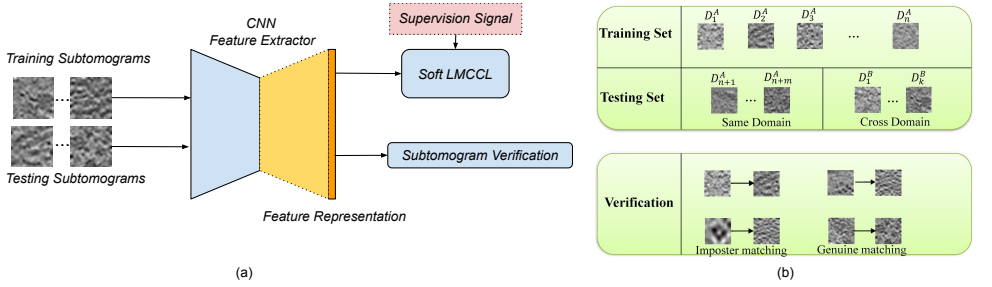

Figure 1: (a) The flowchart of open-set macromolecule recognition. (b) The configuration of our training and evaluation protocols. The upper part denotes the configuration of the training and testing set where D_n^A stands for the n^{th} class in dataset A and D_k^B denotes the k^{th} class in dataset B. Inter-domain and cross-domain open-set scenarios are tested. The performance is evaluated by subtomogram verification which is shown in the lower part.

2.2 Soft Large Margin Centralized Cosine Loss

In this section, we discuss the components of our proposed Soft LMCCL in detail.

Center Loss for Intra-Class Variance Minimization: Traditional softmax loss emphasize the separability of features from different classes but they do not consider the compactness of learned features, meaning that features from the same class cluster together [19]. Wen *et al.* [19] proposed Center loss to shrink the intra-class variance of the learned representations. Denote x_i^j as the feature representation of a 3D subtomogram from class j, c^j as the center vector of class j. The intra-class variance is formulated as the average L2 distance as follows:

$$L_{Center} = \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^j - c^j\right\|_2, \tag{1}$$

where N is the number of samples in class j. In each iteration, the centers of each class and the representations of training samples are updated simultaneously. However, updating every class center in each training iteration causes a huge computational burden. Therefore, the mini-match updating rule is proposed by keeping a fraction of the class centers updated, and is implemented in Equation 2 and 3.

$$\Delta c_t^j = \frac{\sum_{i=1}^{N}\delta\left(y_i \in j\right)\cdot\left(c^j - x_i\right)}{1+\sum_{i=1}^{N}\delta\left(y_i \in j\right)}, \tag{2}$$

$$c_{t+1}^j = c_t^j - \alpha\cdot\Delta c_t^j, \tag{3}$$

where Δc_t^j denotes the deviation between the class center and the data points from the j^{th} class. y_i is the ground truth label of the i^{th} 3D image. $\delta(x)$ is a function such that if argument x is true, then $\delta(x)=1$, otherwise $\delta(x)=0$. α is the learning rate for updating the center vectors so that the learned representations have a low intra-class variance.

Large Margin Cosine Loss for Inter-Class Variance Maximization: Initially, the softmax loss is formulated to make the posterior probability of sample x_i being correctly classified:

$$L_{softmax} = \frac{1}{N}\sum_{i=1}^{N}-\log p_i = \frac{1}{N}\sum_{i=1}^{N}-\log\frac{e^{W_{y_i}^T\cdot x_i+b_{y_i}}}{\sum_{j=1}^{C}e^{W_j^T\cdot x_i+b_j}} = \sum_{i=1}^{N}-\log\frac{e^{\|W_{y_i}\|\|x_i\|\cos\theta_{y_i}+b_{y_i}}}{\sum_{j=1}^{C}e^{\|W_j\|\|x_i\|\cos\theta_j+b_j}}, \tag{4}$$

where C is the number classes in training set. p_i is the posterior probability of a sample being correctly classified. $W_j, W_{y_i}, b_j, b_{y_i}$ are the weight matrix and bias term in the last fully connected layer. θ_j is the angle between the weight vector and the feature vector in the high dimensional Euclidean space. Obviously, both the norm and angle of vectors contribute to the posterior probability, so the difference of the feature norm between the training and testing set will influence the recognition result as well. Therefore, Large Margin Cosine Loss is proposed to improve the separability of features from different classes [17], which transforms the feature vectors to the angular space to develop the discriminative power. Specifically, we set the norm of feature and weight vectors in Equation 4 to a constant where $\|W_j\| = 1, \|x_i\| = s$. Therefore, Equation 4 can be reformulated as:

$$L_{softmax} = \frac{1}{N}\sum_{i=1}^{N} -\log \frac{e^{s\cdot\cos\theta_{y_i}+b_{y_i}}}{\sum_{j=1}^{C} e^{s\cdot\cos\theta_j+b_j}}. \tag{5}$$

Based on the normalization on the feature and weight vectors, the performance of our model relies on separability in the angular space, which removes the influence of feature and weight norm when facing novel testing data. However, the modified Softmax loss is still lacking discrimination power when applied in the open-set scenario, especially when the distribution of testing data is far from what the training data presents. To enhance the inter-class variance, a cosine margin term m is introduced to modify the decision boundary of neural networks.

For simple illustration, we take the binary classification as an example. For models trained by the normalized Softmax loss, the decision boundary is simply derived: $\cos\theta_1 - \cos\theta_2 = 0$ where θ_1 and θ_2 denote the angle between the feature vector and its corresponding weight vector separately for class C_1 and C_2. If $\cos\theta_1 > \cos\theta_2$, the image will be predicted to come from class C_1 and otherwise it will be class C_2. In order to more rigorously guide the model at open-set scenarios, an additional fixed cosine margin m is proposed to enlarge the distance between two classes in the angular space. Therefore, the new decision boundary are obtained in Equation 6 and 7.

$$C_1 : \cos\theta_1 > \cos\theta_2 + m. \tag{6}$$

$$C_2 : \cos\theta_2 > \cos\theta_1 + m. \tag{7}$$

By introducing such a cosine distance, the trained model can provide an error margin which is sufficient to cover the outliers in the representation space so that the discrimination ability is improved. Here the formal LMCL is defined as follows:

$$L_{LMCL} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\left(\cos\theta_{y_i}-m\right)+b_{y_i}}}{e^{s\left(\cos\theta_{y_i}-m\right)+b_j} + \sum_{j=1, j\neq y_i}^{C} e^{s\cos\theta_j+b_j}} \tag{8}$$

Joint Supervision of Soft Large Margin Centralized Cosine Loss: Though the discrimination power of the trained CNN feature extractor is increased with Center loss and LMCL, there several problems still manifest. First, the center updating rule in center loss cares only about the minimization of the intra-class variance so the learned representations from different classes may overlap with each other in the high-dimensional space. The cosine margin in LMCL essentially takes the inter-class variance into account but determining an appropriate margin m is a non-trivial work. The optimal value of m is hard to find when facing heterogeneous testing data in CECT applications. With an improper margin m, the intra-class

variance will increase which even makes the model hard to converge [17]. Therefore, we propose a new loss function for open-set subtomogram recognition where we roughly find a sub-optimal cosine margin m and then jointly train a CNN feature extractor with the supervision of Center loss, LMCL and Softmax loss, which maximizes the inter-class variance and minimizes the intra-class variance simultaneously. The overall loss functions are formulated as follows:

$$L_{all} = L_{LMCL} + \lambda \cdot L_{Center} + L_{Softmax} + L_{reg}, \tag{9}$$

where λ is the weighting parameter. L_{reg} is the regularization loss in order to prevent overfitting. Specifically, $L_{Softmax}$ and L_{Center} are jointly applied to ensure the separability of features from different classes. In that Softmax loss is an auxiliary training objective which has a comparable value to LMCL, we do not add a weighting parameter on it. The empirical comparison of our proposed loss function and other different loss functions are demonstrated in Figure 2.

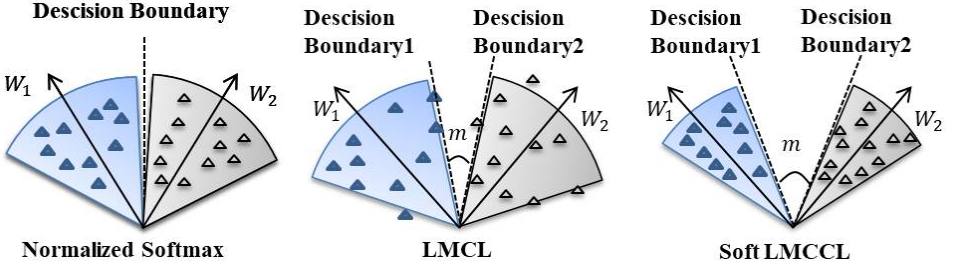

Figure 2: Comparison of three different loss functions in binary classification.

From Figure 2, the data points residing near the decision boundary prevent us from setting a large margin m, otherwise they will become outliers. In contrast, Soft LMCCL shrinks the intra-class gap in the angular space which relaxes the requirement of a predefined optimal cosine margin.

3 Experiments and Results

3.1 Datasets

We evaluated our method on both simulated and experimental CECT datasets. Specifically, we use two simulation datasets and one real dataset. For simulation datasets, we utilize the PDB2VOL program [20] to generate 23 classes of subtomograms at two different SNR levels, including $SNR = 0.03$ (S_1) and $SNR = 0.05$ (S_2). Each class contains approximately $1,000$ subtomograms with size of 40^3 voxels. These simulation datasets were used in our 23 classes open-set recognition task experiments. For real datasets, we used a set of rat neuron tomograms from [3]. We apply Difference of Gaussian (DoG) particle picking method [14] to extract $2,394$ subtomograms with size of 40^3 from several tomograms. The $2,394$ subtomograms contain 6 classes detected and classified by template matching (R_1) [3]. We then evaluated the 6 classes open-set recognition task on it.

3.2 Experimental Setting

To evaluate the performance of our proposed loss function, we conduct open-set verification in two different settings, which is intra-domain verification and cross-domain verification. Concretely, given pairs of subtomograms in the testing set, we use our model to obtain their normalized feature embeddings. Then we calculate the L2 distance between them. Denote the different image pairs from class m and n as $\{x_i^m, x_j^n\}$, we generate an appropriate threshold k by *Sklearn K-FOLD* cross evaluation module such that if $\|f(x_i^m) - f(x_j^n)\| > k$, they are not from the same class where f means the CNN feature extractor. For intra-domain verification, we split the three datasets into training and testing datasets. Then after optimization by Soft LMCCL on the training datasets, we apply the model to conduct verification tasks on the testing set. The training and testing ratio is set to 17:6 in S_1 and S_2 dataset and 1:1 in R_1 dataset. Training is conducted on 17,000 simulated subtomograms from 17 classes and 1,468 experimental subtomograms from 3 classes. The evaluation is conducted on 6,000 simulated subtomograms from 6 unseen classes and 926 experimental subtomograms from 3 classes. For cross-domain verification, a CNN feature extractor is trained separately on dataset S_1, S_2 and R_1 and then we generate the embeddings on subtomograms from other datasets. Specifically, we conduct 4 sets of cross-domain open-set verification, which are simplified as $S_1 \rightarrow S_2, S_2 \rightarrow S_1, S_1 \rightarrow R_1$ and $S_2 \rightarrow R_1$. We don't consider transferring from R_1 to S_1, S_2 because the 6 classes in the experimental dataset are not sufficient for cross-domain adaptation compared to 23 classes in the simulated dataset. For both intra-domain evaluation and cross-domain evaluation, we manually generate 6,000 subtomogram pairs from the same class and 12,000 subtomogram pairs from different classes. Then, we report four metrics based on the distances between these subtomogram pairs. They are: 1) **Verification accuracy**: it was calculated through dividing the number of correctly predicted samples by the overall number of matching pairs. 2) **True Accepted Rate (TAR) under a specific False Accepted Rate (FAR)**: a high TAR at a low FAR is important because it avoids that an imposter matching is accepted, which improves the reliability and sensitivity of the open-set CECT recognition model. 3) **Area Under Curve (AUC)**: the area under the Receiver Operating Characteristic (ROC) curve. 4) **Equal Error Rate (EER)**: the error rate at which the False Acceptance Rate (FAR) equals False Recognition Rate (FRR).

The first three metrics are in the positive relationship with the model performance such that a higher value of the metrics indicate the a better recognition performance except the last one.

3.3 Implementation Details

The CNN feature extractor used in our experiments is similar to ResNet-20 [8] but with 3D convolutional kernels. We change the dimension of the first fully connected layer to 1,024 and regard its output as the feature embeddings. We set the norm of the features to 64. The learning rate for updating the neural network is set to 0.01 with Adam optimizer [9] while the learning rate for class center update is set to 0.05. On each dataset, we train the neural network by 30,000 epochs with the batch size of 90. The regularization loss is implemented by weight decay of 0.0005. Upon testing, we obtain the embeddings of the testing 3D volumes for distance calculation and evaluation.

Table 1: Intra-domain Subtomogram Verification on Dataset S_1

Loss Functions	Verification Acc	TAR @ (FAR=0.1%)	TAR @ (FAR=0.01%)	AUC	EER (%)
$L_{Softmax}$	0.726+-0.010	0.02514+-0.00687	0.00117+-0.00130	0.698	0.347
L_{Center}	0.667+-0.012	0.01817+-0.00548	0.00017+-0.00050	0.501	0.500
L_{LMCL}-0.35	0.742+-0.009	0.00847+-0.00240	0.00034+-0.00067	0.760	0.284
L_{LMCL}-0.5	0.772+-0.010	0.01604+-0.00481	0.00317+-0.00330	0.749	0.304
L_{LMCL}-0.65	0.742+-0.010	0.01545+-0.00599	0.00230+-0.00208	0.684	0.357
L_{LMCL}-0.7	0.753+-0.008	0.01735+-0.00476	0.00134+-0.00102	0.664	0.379
$L_{LMCL}+L_{Center}$	0.667+-0.007	0.53698+-0.00514	**0.13784+-0.01125**	0.599	0.418
$L_{LMCL}+0.01\cdot L_{Center}$	0.734+-0.010	0.51546+-0.00402	0.1050+-0.00107	0.726	0.335
$L_{LMCL}+0.1\cdot L_{Center}$	0.743+-0.005	0.51116+-0.00595	0.09031+-0.00093	0.736	0.321
$L_{LMCL}+0.2\cdot L_{Center}$	0.720+-0.008	0.55828+-0.00798	0.12495+-0.00478	0.670	0.376
$L_{LMCL}+0.5\cdot L_{Center}$	0.667+-0.010	0.50046+-0.02283	0.09918+-0.00226	0.543	0.471
Soft LMCCL	**0.794+-0.008**	**0.54761+-0.01468**	0.13613+-0.01532	**0.862**	**0.222**

Table 2: Intra-domain Subtomogram Verification on Dataset S_2

Loss Functions	Verification Acc	TAR @ (FAR=0.1%)	TAR @ (FAR=0.01%)	AUC	EER(%)
$L_{Softmax}$	0.696+-0.008	0.00802+-0.00257	0.00475+-0.00125	0.667	0.380
L_{Center}	0.667+-0.012	0.03001+-0.00168	0.07157+-0.00015	0.500	0.500
L_{LMCL}-0.35	**0.746+-0.007**	0.00882+-0.00405	0.00066+-0.00110	0.734	0.323
L_{LMCL}-0.5	0.733+-0.010	0.00703+-0.00288	0.00184+-0.00162	0.681	0.362
L_{LMCL}-0.65	0.730+-0.011	0.01020+-0.00358	0.00117+-0.00168	0.660	0.384
L_{LMCL}-0.7	0.719+-0.012	0.01418+-0.00390	0.00288+-0.00191	0.645	0.385
$L_{LMCL}+L_{Center}$	0.667+-0.008	0.55115+-0.01564	0.13256+-0.01751	0.651	0.415
$L_{LMCL}+0.01\cdot L_{Center}$	0.667+-0.011	0.49875+-0.00256	0.11564+-0.02397	0.617	0.407
$L_{LMCL}+0.1\cdot L_{Center}$	0.688+-0.006	0.51403+-0.00340	0.10083+-0.00083	0.633	0.402
$L_{LMCL}+0.2\cdot L_{Center}$	0.667+-0.005	0.55997+-0.00553	0.10000+-0.00672	0.545	0.475
$L_{LMCL}+0.5\cdot L_{Center}$	0.682+-0.007	0.50218+-0.00151	0.08756+-0.00124	0.638	0.389
Soft LMCCL	0.705+-0.009	**0.60121+-0.02561**	**0.15481+-0.01569**	**0.802**	**0.277**

3.4 Intra-domain Open-set Subtomogram Recognition

In this section, we present the experiment results of intra-domain open-set recognition in dataset S_1, S_2 and S_3. For comparison, we report the results with Softmax loss, Center loss and LMCL separately. Besides, we also conduct ablation studies on two key hyperparameters, namely cosine margin m and weighting parameter λ. We first coarsely determine an appropriate cosine margin by testing on the model trained with LMCL only. Afterwards, we jointly train a CNN model supervised by Center loss and LMCL and we set different values of λ in order to find an optimal weighting parameter. Then, we integrate all three parts together with the empirically determined m and λ and report the four metrics. The performance is summarized in Table 1, 2 and 3. Note that the statistics is reported in the form of its mean value followed by the standard deviation in the *K-Fold* evaluation protocol.

From the above tables, it is obvious that our proposed loss function can guide the deep

Table 3: Intra-domain Subtomogram Verification on Dataset R_1

Loss Functions	Verification Acc	TAR @ (FAR=0.1%)	TAR @ (FAR=0.01%)	AUC	EER(%)
$L_{Softmax}$	**0.649+-0.021**	0.00876+-0.00419	0.00602+-0.00356	**0.701**	0.346
L_{Center}	0.554+-0.015	0.02766+-0.00435	0.00000+-0.00000	0.500	0.500
L_{LMCL}-0.35	0.560+-0.015	0.02232+-0.00596	0.02232+-0.00596	0.525	0.489
L_{LMCL}-0.5	0.545+-0.013	0.01637+-0.00168	0.00165+-0.00058	0.516	0.489
L_{LMCL}-0.65	0.556+-0.017	0.01867+-0.0713	0.00549+-0.00025	0.513	0.496
L_{LMCL}-0.7	0.555+-0.021	0.00732+-0.00223	0.00012+-0.00000	0.527	0.472
$L_{LMCL}+L_{Center}$	0.554+-0.014	0.66512+-0.01145	0.08965+-0.00457	0.501	0.499
$L_{LMCL}+0.01\cdot L_{Center}$	0.579+-0.016	0.67969+-0.00256	0.10580+-0.01145	0.583	0.441
$L_{LMCL}+0.1\cdot L_{Center}$	0.554+-0.013	0.66386+-0.01042	0.09856+-0.00128	0.509	0.495
$L_{LMCL}+0.2\cdot L_{Center}$	0.554+-0.014	0.66439+-0.00791	0.10323+-0.01154	0.509	0.495
$L_{LMCL}+0.5\cdot L_{Center}$	0.554+-0.018	0.66397+-0.01374	0.10081+-0.01087	0.508	0.495
Soft LMCCL	0.602+-0.010	**0.69546+-0.02116**	**0.13575+-0.02293**	0.671	**0.339**

Table 4: Cross-domain Subtomogram Verification Results

Loss Functions	Verification Acc	TAR @ (FAR=0.1%)	TAR @ (FAR=0.01%)	AUC	EER(%)
		$S_1 \rightarrow S_2$			
$L_{Softmax}$	0.730+-0.011	0.01867+-0.00254	0.00564+-0.00025	0.728	0.329
L_{Center}	0.667+-0.011	0.03265+-0.01254	0.01002+-0.00000	0.500	0.500
L_{LMCL}-0.5	**0.789+-0.009**	0.02521+-0.00542	0.00533+-0.00336	0.799	0.263
Soft LMCCL	0.788+-0.005	**0.50930+-0.01185**	**0.11786+-0.00790**	**0.862**	**0.225**
		$S_2 \rightarrow S_1$			
$L_{Softmax}$	0.694+-0.009	0.01710+-0.00595	0.00256+-0.00012	0.659	0.381
L_{Center}	0.667+-0.011	0.05135+-0.00145	0.00425+-0.00151	0.500	0.500
L_{LMCL}-0.35	0.731+-0.009	0.00833+-0.00404	0.00083+-0.00083	0.697	0.343
Soft LMCCL	**0.734+-0.007**	**0.53915+-0.00732**	**0.10301+-0.00238**	**0.784**	**0.293**
		$S_1 \rightarrow R_1$			
$L_{Softmax}$	0.554+-0.012	0.03428+-0.00409	0.00672+-0.00138	0.503	0.489
L_{Center}	0.554+-0.019	0.01439+-0.00147	0.01125+-0.00243	0.552	0.462
L_{LMCL}-0.5	0.619+-0.017	0.01708+-0.00607	0.00838+-0.00399	0.525	0.489
Soft LMCCL	**0.637+-0.011**	**0.50668+-0.00263**	**0.09066+-0.00133**	**0.663**	**0.371**
		$S_2 \rightarrow R_1$			
$L_{Softmax}$	0.565+-0.021	0.11230+-0.00514	0.07519+-0.02561	0.573	0.436
L_{Center}	0.554+-0.015	0.01025+-0.00002	0.00108+-0.00727	0.500	0.500
L_{LMCL}-0.35	**0.647+-0.020**	0.09845+-0.00771	0.01564+-0.00357	0.658	0.378
Soft LMCCL	0.637+-0.017	**0.52199+-0.01184**	**0.11635+-0.00838**	**0.669**	**0.373**

neural network to learn more discriminative features. Compared to Center loss, Softmax loss and LMCL generally perform slightly better. However, all of the models trained independently under the three loss functions are much weaker than the model trained with Center loss and LCML and the model with Soft LMCCL. Specifically, after combining Center loss and LMCL, the CNN feature extractor improved by more than 50% in all 3 datasets in terms of TAR at a low FAR, indicating it will be less likely to accept imposter samples. Therefore, the reliability of the open-set recognition framework is boosted. However, the verification accuracy in this scenario is less than the model trained with LMCL alone. After jointly training with Center loss, LMCL and Softmax loss, the verification accuracy is recovered and even boosted because of the inter-class separability of Softmax loss. With respect to the influence of cosine margin m and the weighting parameter λ, we find that $m = 0.35$ is optimal for dataset S_2, R_1 and $m = 0.5$ is optimal for dataset S_1. When m is larger, the performance decreases because the noisy unseen testing data will reside beyond the decision boundary obtained in the training set. Similarly, the weighting parameter is set to 0.1 for S_1, S_2 and 0.01 for R_1. For the Soft LMCCL, the cosine margin and the weighting parameter is determined according to the optimal values obtained in the ablation studies.

3.5 Cross-domain Open-set Subtomogram Recognition

Similar to the experiment setting in the previous section, the cross-domain open-set recognition is conducted by training on one entire dataset and then be tested on given subtomogram pairs in another dataset. This setting is more difficult because of the larger distribution gap between the training and unseen data. The results are reported in Table 4.

From Table 4, the model trained using Soft LMCCL performs better than three independent loss functions, especially in terms of TAR at a low FAR. The performance is even close to the intra-domain open-set recognition, which demonstrates the robustness of the learned discriminative feature embeddings in cross-domain applications. This type of application is closer to the practical setting where biologists face and want to research the macromolecule structures in an entirely unseen CECT dataset.

4 Conclusions

Computational analysis, deep learning approaches in particular, has been playing increasingly important role for obtaining molecular machinery insights from CECT data [7, 11, 12, 18, 27, 29, 30]. However, supervised deep learning approaches cannot be directly used to detect macromolecular structures that do not exist in training data. In this paper, we propose a novel loss function, Soft LMCCL, for unseen macromolecule recognition in CECT tasks, which consists of Softmax loss, Large Margin Cosine Loss and Center loss. These components jointly supervise the feature extraction in deep neural networks in order to learn a discriminative feature representation. Our experiments on both simulated and experimental datasets demonstrated that our proposed loss function can enlarge the inter-class variance while shrinking the intra-class variance, which significantly improves CNN's generalization ability compared to previous methods. In summary, our Soft LMCCL can detect and recognize unseen macromolecule structures captured by CECT. Our work serves as an important step towards systematic recognition and recovery of native structures and spatial organizations of macromolecules, which would lead to a broad range of applications in life science. 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 [15, 22], averaging [5, 29], and segmentation [12, 23, 24].

5 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] Chengqian Che, Ruogu Lin, Xiangrui Zeng, Karim Elmaaroufi, John M. Galeotti, and Min Xu. Improved deep learning-based macromolecules structure classification from electron cryo-tomograms. *Mach. Vis. Appl.*, 29(8):1227–1236, 2018.

[2] 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.

[3] Qiang Guo, Carina Lehmer, Antonio Martinez-Sanchez, Till Rudack, Florian Beck, Hannelore Hartmann, Manuela Perez-Berlanga, Frederic Frottin, Mark S. Hipp, and F. Ulrich Hartl. In situ structure of neuronal c9orf72 poly-ga aggregates reveals proteasome recruitment. *Cell*, 172(4):696, 2018.

[4] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pages 1735–1742, 2006.

[5] 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.

[6] Bong-Gyoon Han, Ming Dong, Haichuan Liu, Lauren Camp, Jil Geller, Mary Singer, Terry C Hazen, Megan Choi, H Ewa Witkowska, David A Ball, Dieter Typke, Kenneth H Downing, Maxim Shatsky, Steven E Brenner, John-Marc Chandonia, Mark D Biggin, and Robert M Glaeser. Survey of large protein complexes in d. vulgaris reveals great structural diversity. In *Proceedings of the National Academy of Sciences*, volume 106, pages 16580–16585, 2009.

[7] 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, 07 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz323.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016.

[9] 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.

[10] 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.

[11] Ruogu Lin, Xiangrui Zeng, 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, 07 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz364.

[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. In *2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018*, pages 1578–1582, 2018. doi: 10.1109/ICIP.2018.8451386.

[13] 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.

[14] Pei Long, 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] 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*.

[16] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 815–823, 2015.

[17] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 5265–5274, 2018.

[18] Kaiwen Wang, Xiangrui Zeng, Xiaodan Liang, Zhiguang Huo, Eric Xing, and Min Xu. Image-derived generative modeling of pseudo-macromolecular structures - towards the statistical assessment of electron cryotomography template matching. In *British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3-6, 2018*, page 130, 2018.

[19] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VII*, pages 499–515, 2016.

[20] Willy Wriggers, Ronald A. Milligan, and J.Andrew McCammon. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *Journal of Structural Biology*, 125(2):185 – 195, 1999. ISSN 1047-8477.

[21] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, and Y. Yang. Progressive learning for person re-identification with one example. *IEEE Transactions on Image Processing*, pages 1–1, 2019. ISSN 1057-7149.

[22] M. Xu, M Beck, and F Alber. High-throughput subtomogram alignment and classification by fourier space constrained fast volumetric matching. *Journal of Structural Biology*, 178(2):152–164, 2012.

[23] 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.

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

[25] 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.

[26] 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(4):679–691, 2019.

[27] 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*, 202(2):150–160, 2018.

[28] 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.

[29] 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. doi: 10.1093/bioinformatics/bty267.

[30] Bo Zhou, Qiang Guo, Kaiwen Wang, Xiangrui Zeng, Xin Gao, and Min Xu. Feature decomposition based saliency detection in electron cryo-tomograms. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2467–2473. IEEE, 2018.