UNSUPERVISED DOMAIN ALIGNMENT BASED OPEN SET STRUCTURAL RECOGNITION OF MACROMOLECULES CAPTURED BY CRYO-ELECTRON TOMOGRAPHY

Yuchen Zeng[⋆] *Gregory Howe*[⋆] *Kai Yi*[†] *Xiangrui Zeng*[⋆] *Jing Zhang*[∓] *Yi-Wei Chang*[◇] *Min Xu*[⋆]

[⋆] Computational Biology Department, Carnegie Mellon University, United States
[†] King Abdullah University of Science and Technology, Saudi Arabia
[∓] Department of Computer Science, University of California Irvine, United States
[◇] Perelman School of Medicine, University of Pennsylvania, United States

ABSTRACT

Cellular cryo-Electron Tomography (cryo-ET) provides three-dimensional views of structural and spatial information of various macromolecules in cells in a near-native state. Subtomogram classification is a key step for recognizing and differentiating these macromolecular structures. In recent years, deep learning methods have been developed for high-throughput subtomogram classification tasks; however, conventional supervised deep learning methods cannot recognize macromolecular structural classes that do not exist in the training data. This imposes a major weakness since most native macromolecular structures in cells are unknown and consequently, cannot be included in the training data. Therefore, open set learning which can recognize unknown macromolecular structures is necessary for boosting the power of automatic subtomogram classification. In this paper, we propose a method called Margin-based Loss for Unsupervised Domain Alignment (MLUDA) for open set recognition problems where only a few categories of interest are shared between cross-domain data. Through extensive experiments, we demonstrate that MLUDA performs well at cross-domain open-set classification on both public datasets and medical imaging datasets. So our method is of practical importance.

Index Terms— Cryo-Electron Tomography, Open-set learning

1. INTRODUCTION

In recent years, cryo-electron tomography (cryo-ET) has emerged as a revolutionary 3D structural biology imaging technique. Cryo-ET captures the 3D native structure and spatial distribution of macromolecules inside cells at nanometer resolutions [1]. A key step in the analysis of cryo-ET data is recognizing each macromolecule through subtomogram classification. A *subtomogram* is a 3D cubic subvolume of a tomogram that contains a single macromolecule. The subtomogram classification task is essentially a 3D grey scale image classification process. However, the high structural complexity, low signal-to-noise ratio and imaging limits have made such classification very difficult. Nowadays, the advance of automatic image acquisition has made it possible for an electron microscope to, within several days, quickly produce hundreds of tomograms that together contain millions of structurally highly heterogeneous macromolecules [2]. In such case, the traditional subtomogram alignment based subtomogram classification techniques [3] become too slow to process such large amounts of data. Recently, Convolutional Neural Network (CNN) based supervised deep learning has significantly improved the throughput of subtomogram classification [4, 5]. However, as a supervised learning approach, it cannot be directly used for recognizing the macromolecules whose structural classes do not exist in the training data. This becomes a major hurdle for the usefulness of this approach because most of the native macromolecular structures are unknown, as evidenced by genome sequencing [6], mass spectroscopy [7], and cross-species variation [8]. Therefore, it is crucial to develop a deep learning method to recognize unknown macromolecular structures. Open-set learning is an important technique to solve this problem. Another major challenge is, even for the macromolecules of known structures, the high quality annotation of corresponding subtomograms is very computation and labor intensive. Acquiring real Cryo-EM (Cryogenic electron microscopy) images is expensive, so there may not be sufficient data to train the machine learning model, but with the help of simulated data or annotated data captured from a separate data source, we can roughly make a class prediction. To solve this problem, domain adaptation is needed.

Image classification has two cases, closed set and open set recognition. A closed set recognition task indicates the source and target datsets contain identical classes. Open set recognition is different than the closed set recognition; in open set recognition, all labels in the target domain appear in the source domain. We denote the "known" classes as the labeled samples shared among the source and target data, and the "unknown" class as the unlabeled class which may or may not be

The corresponding author email: mxu1@cs.cmu.edu

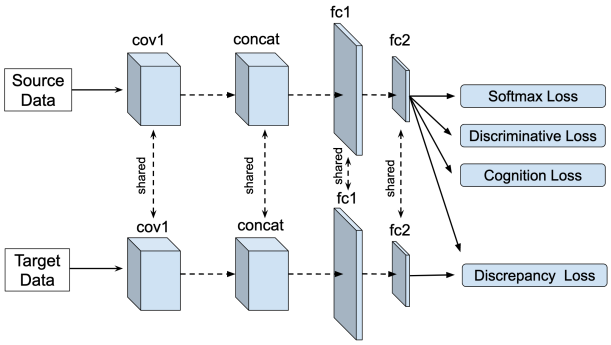


Fig. 1. Siamese network as our backbone architecture.

shared between the source and target data. An open set classifier should reject samples from unknown classes while correctly classifying samples from known classes. In traditional classification tasks, the discriminative power enhances this ability by increasing intra-class compactness and the inter-class separability. Such as, center loss [9], LM-softmax [10] and ArcFace [11]. In previous works, cross-domain open set recognition task were solved by iterative minimizing the distances of the transformation assignments, adversarial learning and etc. [12, 13, 14]. Following JDDA's work [15], in this paper, we create a margin-based loss called **Margin-based Loss for Unsupervised Domain Alignment (MLUDA)** to solve the cross-domain open set recognition on cryo-ET data. Within the MLUDA, we adapt unsupervised domain alignment techniques in order to minimize the domain discrepancy.

2. METHOD

The regular cross-entropy loss outputs a vector to represent the probability distributions of a set of potential outcomes and is widely used in deep learning classification tasks. However, the cross-entropy loss can only separate but not discriminate the deeply learned features. Learning the representative features by creating new loss functions is very effective in the close-set classification task. Open set recognition also needs strong generalization; we solve this by learning representative features from a discriminative loss function. But for cross-domain open set recognition, the feature space from the two domains is not aligned, so additionally, we introduce a domain adaptation alignment method. Together, these additions help solve the cross-domain open set recognition task.

2.1. Problem Formulation

Suppose we are given a labeled source data set $D^s = \{(x_i^s, y_i^s)_{i=1}^{n_s}\}$ from the source domain where each data point x_i^s is a cubic 3 dimensional grey scale image ie $x_i^s \in R^{H\times H\times H\times 1}$. Furthermore, we have that each x_i^s belongs to a known macromolecule class in C_{known} or an unknown class in $C_{unknown}$ where $|C_{known}| = k-1$ and $|C_{unknown}|$ is not known. We let the labels $1, ..., k-1$ represent the known classes, and the label k represent all unknown classes. Our goal is to label a new target data set $D^t = \{(x_i^t)_{i=1}^{n_t}\}$ from a potentially different domain by correctly assigning each $x_i^t \in C_{known}$ the label corresponding to its class from $\{1, ..., k-1\}$ and assigning each $x_i^t \in C_{unknown}$ the label k.

2.2. Margin-based Discrimination Loss

Center loss [9] can boost the discriminative power of the extracted features in neural networks by learning a vector-like center from deep features and combining it with the cross-entropy loss. Let x^j be the deep feature from the last layer in the Siamese network source branch and let c^j be the center of its class. Center loss is defined as the $l2$ differences between the deep features and its class center in Equation 1.

$$L_{center} = \frac{1}{2}\sum_{i=1}^{n} \|x_i^j - c^j\|_2^2 \tag{1}$$

We define a margin-based discriminative loss for the deep features from known classes. We set m_1 as the intra-class margin which specifies a maximum distance between deep features with its corresponding class center and m_2 as the inter-class margin which specifies a minimum distance between different classes. Let $H_k^s = \{(h_i^s)_{i=1}^{n_k^s}\}$ be the set of learned deep features for the data point that belong to the class k. Each learned feature representation h_i^s should be within some distance m_1 from its class's center. Furthermore, each class center should be at least some distance m_2 from all other class centers. With this in mind, we formulate the margin-based discriminative loss, Equation 2:

$$\begin{aligned} L_d(H_k^s) = &\sum_{i=1}^{n_k^s} \max(0, \|h_i^s - c_{y_i}\|_2^2 - m_1^2) \\ &+ \sum_{i=1}^{n_k^s} \max(0, m_2^2 - \|c_i - c_j\|_2^2) \end{aligned} \tag{2}$$

Intuitively, the inter-class distance should be larger than the intra-class distance, thus we require $m_1 > m_2$. In the first term of 2, $c_{y_i} \in R^d$ denotes the y_i-th class center of the deep feature y_i. In the second term, c_i and c_j represent the class centers for two arbitrary (randomly selected) macromolecule classes i and j. The second term of Equation 2 uses these class centers for the arbitrarily selected classes i and j to measure inter-class separability. We update the class centers iteratively with each batch. We use Equations 3 and 4 to update each class center, where y_i is the class of the ith data point in the batch, c_j is cluster j's center, h_i^s is the deep feature representation of the ith data point in the batch, and δ is an indicator function.

$$\Delta c_j^t = \frac{\sum_{i=1}^{b} \delta(y_i = j)(c_j - h_i^s)}{1 + \sum_{i=1}^{b} \delta(y_i = j)} \tag{3}$$

$$c_j^{t+1} = c_j^t - \gamma \cdot \Delta c_j^t \tag{4}$$

Equation 2 can classify the shared classes. However, for unknown classes the classifier also needs to separate the points from each of the known classes. Following the idea of margin-based loss, we specify a distance between the labeled known classes and the unlabeled unknown data. The margin-based cognition loss is defined by Equation 5 where $\{(h_i^s)\}_{i=1}^{n_{unk}^s}$ is the set of deep features for the unknown data from the source domain, $min(\|h_i^s - c_j\|_2^2)$ measures the closest distance from each unknown sample to all the known class centers, and m_3 assigns the unknown class margin which provides a minimum length from each unknown macromolecule to any known class center.

$$L_c(H_{unk}^s) = \sum_{i=1}^{n_{unk}^s} \max(0, m_3^2 - \min(\|h_i^s - c_j\|_2^2)) \quad (5)$$

2.3. Deep CORAL: Unsupervised Domain Alignment

For a cross-domain open set recognition task, the classifier also needs to decrease the domain discrepancy. Deep CORAL provides a simple but efficient method to match distributions of the middle features in the CNN by minimising the covariance of the source and target features [16]. The CORAL loss is expressed as the distance between the second-order statistics (covariances) of the source and target features as shown in Equation 6.

$$\begin{aligned} L_{CORAL} &= CORAL(H^s, H^t) \\ &= \frac{1}{4L^2}\|Cov(H^s) - Cov(H^t)\|_F^2 \end{aligned} \quad (6)$$

where H^s and H^t denote the deep features from the output of the bottleneck layer. $\|\cdot\|_F^2$ is the squared matrix Frobenius norm. The covariance matrices of the source and target data are given by Equation 7. h is the number of batch data and 1 is a column vector with all elements equal to 1. The calculation of the coral loss does not need the target labels as reference, so it is an unsupervised method for aligning two domains. Minimizing the correlation alignment (CORAL) can adjust shared weights and reduce the domain discrepancy.

$$Cov(H) = \frac{1}{h-1}(H^\top H - \frac{1}{h}(\mathbf{1}^\top H)^\top(\mathbf{1}^\top H)) \quad (7)$$

2.4. Training Procedure

Let H_i be the shared feature in the last fully connected layer. Combining all the losses from Equations 2 5 6 with the softmax loss in 8 to reformulate our MLUDA as Equation 9. The softmax is formulated to make the posterior probability of sample x_i.

$$L_{softmax} = \frac{1}{N}\sum_{i=1}^{N} -\log H_i \quad (8)$$

$$L = L_{softmax} + \lambda_1(L_d + L_c) + \lambda_2 L_{CORAL} \quad (9)$$

And, our MLUDA function can be trained by minimizing the weighted combination of $L_{softmax}$, L_d, L_c and L_{CORAL}. Where λ_1 and λ_2 are trade-off parameters used to balance the contribution of each loss. λ_2 should be lower than λ_1 to make sure margin-based loss dominate the total loss. The λ_2 is almost identical in different experiments. But the λ_1 need to use trial and error to find the best value. In Figure 1, the flowchart shows the training process where the loss will be calculated after the fc2 layer in the Siamese network.

3. EXPERIMENTS AND RESULTS

We conducted two experiments to verify our method for cross-domain classification on the popular 3D dataset and the real Cryo_EM dataset for practical usage.

3.1. Ablation Experiments on 3D MNIST

3D MNIST contains 3D point clouds generated from the original images of the MNIST dataset. [1] In this experiment, we use the first five classes as the source dataset and the last five classes as the target dataset. We extract the last features from our Network and used t-distributed stochastic neighbor embedding (TSNE) to display the feature space. We ablated each part of our loss and tested their performances separately. As Figure 2 shows, our method can clearly separate the class in the feature space in this cross-domain scenario. (a) means we only use cross entropy as the loss function. (b) means we use cross entropy loss and the Intra-class center as the total loss. In (c), we use cross entropy term and the margin-based discrimination loss we descibed in section 2.2. (d) is our method.

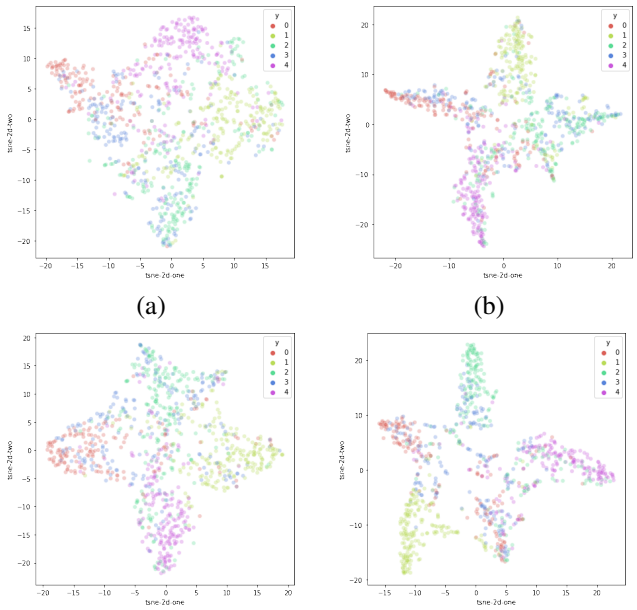
DOI	EMPIAR Name
10133	Glutamate dehydrogenase single particle
10131	Rabbit muscle aldolase single particle
10143	T20S proteasome single particle
10135	DNAB helicase-helicase loader single particle
10173	Insulin-bound insulin receptor single particle
10172	Hemagglutinin single particle
10169	Apoferritin single particle

Table 1. Experimental macromolecular complexes names

3.2. Experiments on Real Cryo-EM Tomogram

Real Dataset: We use The Noble Single Particle Dataset [17]. This dataset has 7 classes. We apply the Difference of Gaussian (DOG) particle picking process in order to extract the subtomogramss of size 28^3 [18] followed by manual selection. We denote the real dataset as R. The digital object identifier (DOI) of our real dataset in Electron Microscopy Public Image Archive (EMPIAR) is shown in Table 1.

[1] https://www.kaggle.com/daavoo/3d-mnist



(a) (b) (c) (d)

Fig. 2. Feature spaces of different ablation methods.

Simulated Dataset: we use the same structures of macromolecular complexes to generate simulated subtomograms. First we generate volume of 40^3 voxels with a resultion of 0.92 nm using the PDB2VOL program from Situs [19] package as the density map for each class. Then We apply the AITom platform to generate the simulated dataset with the same class [20]. There are two significant aspects we focus on when simulating the subtomograms: missing wedge effects and noises. We generated 500 subtomograms for each class with SNR = 10 and missing weight = 10. We denote the simulated dataset as S_3. We also generate two other simulated datasets. One, denoted as S_1, has the same real 7 complexes classes with SNR = 3 and missing weight = 10. The other not only has the 7 real experimental macromolecular complexes classes but also has another 20 new classes of subtomograms which generated by PDB2VOL program with SNR = 7 and missing weight = 10. We donoted this simulated dataset as S_2.

Method	All	OS	OS*	UNK
$L_{softmax}$	0.6070	0.4370	0.4403	0.8156
L_{center}[9]	0.5916	0.4341	0.4633	0.7811
JDDA[15]	0.6214	0.4804	0.4846	0.7922
Our	**0.6513**	**0.5112**	**0.5141**	**0.8232**

Table 2. Experiment results of $S_1 \longrightarrow S_2$

We perform the experiment between two simulated datasets, which denoted as $S_1 \longrightarrow S_2$. In S_3, we keep first 6 classes. For the target dataset, we use 4 classes as the known classes and the last two class as the unknown classes from S. We test the performance on R, which has the same 4 known classes but 3 unknown classes. This experiment denoted as $S_3 \longrightarrow R$. Our method is a new application to the cross domain classification task on Cryo-EM. Due to huge domain shifts and the complexity of 3D classification, classic classification methods cannot achieve decent results. So, we compare our method with a discriminative learning feature loss (Center Loss), and an Unsupervised Deep Domain Adaptation method (JDDA). As a standard open set recognition experiment, we define four metrics for verification. There are: 1. Overall test accuracy (ALL), 2. Accuracy averaged over all classes (OS), 3. Accuracy measured only on the known samples of the target domain (OS*), 4. Unknown class accuracy (UNK). In Table 2, our method achieves the best performance among all methods between two simulated datasets. As Table 3 shows, our results are best in the ALL, OS and OS* in real dataset.

Method	All	OS	OS*	UNK
$L_{softmax}$	0.1638	0.2582	0.2651	0.0999
L_{center}[9]	0.1404	0.2281	0.0323	0.2345
JDDA[15]	0.3506	0.3113	0.3105	**0.4027**
Our	**0.3605**	**0.4292**	**0.4304**	0.2691

Table 3. Experiment results of $S_3 \longrightarrow R$

4. ACKNOWLEDGEMENT

This work was supported in part by U.S. National Institutes of Health (NIH) grants P41GM103712, R01GM134020, and K01MH123896, U.S. National Science Foundation (NSF) grants DBI-1949629 and IIS-2007595, AMD COVID-19 HPC Fund, and Mark Foundation 19-044-ASP. XZ was supported by a fellowship from Carnegie Mellon University's Center for Machine Learning and Health.

5. CONCLUSION

Classification is a fundamental task in biomedical image analysis. The open set scenario is a more challenging problem when only a few classes share between the source data and target data. A more common situation in analyzing cryo-ET data is the cross-domain open set recognition because two datasets may be collected from different electron microscopes with various imaging conditions and resulting SNRs. In this paper, we introduce a novel loss for cross-domain open set recognition in cryo-ET. The proposed loss function has practical significance. The experiments show that our method has strong discriminative power to classify the known classes and achieve decent result on real dataset. Our method is an important step toward the systematic recognition of unknown structural classes in *situ* cryo-ET image analysis.

6. REFERENCES

[1] Vladan Lucic, Alexander Rigort, and Wolfgang Baumeister, "Cryo-electron tomography: The challenge of doing structural biology in situ," *The Journal of cell biology*, vol. 202, pp. 407–19, 08 2013.

[2] Catherine M Oikonomou and Grant J Jensen, "Cellular electron cryotomography: toward structural biology in situ," *Annual review of biochemistry*, vol. 86, 2017.

[3] Min Xu, Martin Beck, and Frank Alber, "High-throughput subtomogram alignment and classification by fourier space constrained fast volumetric matching," *Journal of structural biology*, vol. 178, no. 2, pp. 152–164, 2012.

[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*, p. 1227–1236, Jun 2018.

[5] Min Xu, Jitin Singla, Elitza Tocheva, Yi-Wei Chang, Grant Jensen, Raymond Stevens, and Frank Alber, "De novo structural pattern mining in cellular electron cryotomograms," *Structure*, vol. 27, pp. 1–13, 04 2019.

[6] Andrew D Hanson, Anne Pribat, Jeffrey C Waller, and Valérie de Crécy-Lagard, "'unknown'proteins and 'orphan'enzymes: the missing half of the engineering parts list–and how to find it," *Biochemical Journal*, vol. 425, no. 1, pp. 1–11, 2010.

[7] Mario Looso, Thilo Borchardt, Marcus Krüger, and Thomas Braun, "Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry," *Molecular & Cellular Proteomics*, vol. 9, no. 6, pp. 1157–1166, 2010.

[8] James J Elser, Claudia Acquisti, and Sudhir Kumar, "Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition," *Trends in ecology & evolution*, vol. 26, no. 1, pp. 38–44, 2011.

[9] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," 10 2016, vol. 9911, pp. 499–515.

[10] Weiyang Liu, Y. Wen, Zhiding Yu, and Meng Yang, "Large-margin softmax loss for convolutional neural networks," *ArXiv*, vol. abs/1612.02295, 2016.

[11] Jiankang Deng, J. Guo, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4685–4694, 2019.

[12] P. P. Busto and J. Gall, "Open set domain adaptation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[13] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *ECCV*, 2018.

[14] Fayin Li and H. Wechsler, "Open set face recognition using transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1686–1697, 2005.

[15] C. Chen, Zhihong Chen, Boyuan Jiang, and Xinyu Jin, "Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation," in *AAAI*, 2019.

[16] Baochen Sun and Kate Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *ECCV Workshops*, 2016.

[17] Alex J Noble, Venkata P Dandey, Hui Wei, Julia Brasch, Jillian Chase, Priyamvada Acharya, Yong Zi Tan, Zhening Zhang, Laura Y Kim, Giovanna Scapin, Micah Rapp, Edward T Eng, William J Rice, Anchi Cheng, Carl J Negro, Lawrence Shapiro, Peter D Kwong, David Jeruzalmi, Amedee des Georges, Clinton S Potter, and Bridget Carragher, "Routine single particle cryoem sample and grid characterization by tomography," *eLife*, vol. 7, may 2018.

[18] 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*, vol. 17, 10 2016.

[19] 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*, p. 125(2):185 – 195, 1999.

[20] Xiangrui Zeng and Min Xu, "Aitom: Open-source ai platform for cryo-electron tomography data analysis," *ArXiv*, vol. abs/1911.03044, 2019.