GRAPH NEURAL NETWORK BASED OPEN-SET DOMAIN ADAPTATION

Shan Zhao[1], Sudipan Saha[1], Xiao Xiang Zhu[1,2]*

[1]Department of Aerospace and Geodesy, Data Science in Earth Observation, Technical University of Munich, Ottobrunn, Germany
[2]Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany

KEY WORDS: Graph Neural Networks, Open-set, Domain Adaptation, Progressive Learning.

ABSTRACT:

Owing to the presence of many sensors and geographic/seasonal variations, domain adaptation is an important topic in remote sensing. However, most domain adaptation methods focus on close-set adaptation, i.e., they assume that the source and target domains share the same label space. This assumption often does not hold in practice, as there can be previously unseen classes in the target domain. To circumnavigate this issue, we propose a method for open set domain adaptation, where the target domain contains additional unknown classes that are not present in the source domain. To improve the model's generalization ability, we propose a Progressive Weighted Graph Learning (PWGL) method. The proposed method exploits graph neural networks in aggregating similar samples across source and target domains. The progressive strategy gradually separates the unknown samples apart from known samples and upgrades the source domain by incorporating the pseudolabeled known target samples. The weighted adversarial learning promotes the alignment of known classes across different domains and rejects the unknown class. The experiments performed on a multi-city dataset show the effectiveness of the proposed approach.

1. INTRODUCTION

Satellite Remote Sensing (RS) enables us to obtain contact-free large-scale information about physical properties of the Earth system from space. Thanks to the large number of such satellites launched by different space agencies, timely Earth monitoring is currently possible with a variety of sensors that can capture different properties. While the multi-temporal, multi-sensory, and large coverage information can benefit Earth observation classification tasks, such temporal, spectral or geography shifts also serve as a challenge for training models. Different geographic locations exhibit different characteristics on remote sensing imagery. The impact of geographic locations on the earth observation can be observed by the difference in climate, which impacts the type and growth of vegetation and crops, as well as the cultural and anthropogenic divides that result in various building styles and densities. Such shifts will degrade the performance of the model when it is applied on the dataset that has a different data distribution to the training set.

Capabilities of deep learning models are generally dependent on the annotated training data used to train the model. However, data labeling is a tedious and expensive process, especially for the RS data (Zhu et al., 2017). Models trained for one setting may not generalize well for the other settings. Narrowing the distribution shift is crucial to enhance the robustness of the models (Gawlikowski et al., 2021). To deal with the shift between training and test data, the research on Domain Adaption (DA) has sprung up (Saha et al., 2016). Given the labeled source domain(s) and unlabeled target domain(s), the DA aims at training a classifier for the target domains.

Most DA methods assume the source and target domains share the same classes of objects, which is not always true in many practical scenarios. A more practical setting is the open-set adaptation, where the target domain has more classes than the source domain. These classes that have not been seen by the source domain are collectively referred as 'unknown' class. Thus

the open-set DA is aiming at identifying the unknown class while classifying the known classes.

A number of works have been proposed in the computer vision literature to solve the open-set domain shift problem. (Luo et al., 2020) proposed an end-to-end Progressive Graph Learning (PGL) framework. The target data is pseudo labeled as unknown or one of the known classes based on the confidence score in a progressive paradigm whose rate is controlled by an enlarge parameter. (Zhang et al., 2021) explored the transferability and discriminability for RS image scene classification. Transferability is minimized to suppress the global distribution difference among domains as well as the local distribution discrepancy of the same classes in different domains. The discriminability is encouraged to enlarge the distribution divergence of different classes in different datasets.

Inspired by (Luo et al., 2020) and (Roy et al., 2021), we exploit a graph neural network (GNN)-based architecture. We follow the curriculum learning scheme to progressively separate the unknown away from known samples. Besides, adversarial training can effectively close the gap between known source samples and known target samples.

The contributions of our work include:

1. Devise a graph neural network based architecture to encourage the within-class compactness and domain closeness while forcing the unknown class farther away from known classes. The proposed architecture can benefit from all source and target samples in a batch.
2. Propose a curriculum learning based strategy for the target domain adaptation.
3. Conduct the experiment on a dataset with geographical shift showing the effectiveness of the proposed method.

The rest of the paper is organized as follows. Related works are discussed in Section 2. We present the proposed method in Section 3. Experimental results are discussed in Section 4. Finally, we conclude the paper in Section 5.

---

* Corresponding author

2. RELATED WORK

2.1 Domain Adaptation

A large number of domain adaptation methods focus on the statistics alignment. Towards this, the most popular indices are maximum mean discrepancy (MMD), HδH-distance, KL divergence, moments, etc. (Blitzer et al., 2007), (Pan et al., 2019), (Peng et al., 2019), (Rakshit et al., 2019). (Venkateswara et al., 2017) proposed an unsupervised Domain Adaptive Hashing (DAH) network. The domain shifts are addressed by minimizing multi-kernel Maximum Mean Discrepancy (MK-MMD). The Hash technique helps to encode the samples from the same categories similar hash code, and the hash values are used to develop a unique loss function for the target data. Though MMD helps to reduce the distribution shift, it fails to prompt within-domain and within-class compactness (Chen et al., 2019). To solve this problem, a complementary term derived from graph embedding is appended to the empirical MMD to revise the similarity matrix of the intrinsic graph. A popular paradigm in domain adaptation is adversarial training, which is generally achieved using a Gradient Reversal Layer (Ganin and Lempitsky, 2015). Such models train the network for domain prediction in addition to the usual classification in the label-space. Using the GRL layer, the model can effectively extract both discriminative and domain-invariant features. Most of the adversarial networks solely match the feature representation across domains but ignore multimodal distributions. Conditional domain adversarial networks (CDAN) (Long et al., 2017) compensate for such shortcomings by conditioning the adversarial adaptation models on the discriminative information conveyed in the classifier predictions. Generative adversarial networks (GANs) provide a way to learn deep representations without using extensively annotated training data (Goodfellow et al., 2014). (Tzeng et al., 2017) proposed an unsupervised Adversarial Discriminative Domain Adaptation (ADDA) approach. When the source and the target feature representations become sufficiently inseparable that a domain discriminator fails to separate them, the source classifier can be tasked to classify target samples on the common feature space.

The open-set DA assumes the target domain has more number of classes than the source domain while for the universal DA it allows each domain to hold private classes that are not seen by other domains. Here we mainly focus on the open-set domain adaptation (Panareda Busto and Gall, 2017). Formally speaking, the open-set DA is formulated as the following. Assume that we have a labeled source dataset $\mathcal{S} = \{(\mathbf{x}_{s,i}, y_{s,j})\}_{i=1}^{n_s} \sim P^s$, where n_s is the number of labeled samples; and the unlabeled target dataset $\mathcal{T}$, where $\mathcal{T} = \{\mathbf{x}_{t_j,k}\}_{k=1}^{n_j} \sim Q_X^t$ with each having n_j unlabeled samples. P^s is the joint probability distribution of the source domain, and Q_X^t is the marginal distribution of the target domain. The data distribution of source and target domain is different. The goal is to learn an optimal target classifier $h : Q_X^t \rightarrow \mathcal{Y}_t$. Here the target label space $\mathcal{Y}_t = \{\mathcal{Y}_s, \text{unk}\} = \{1, ..., C+1\}$ includes the additional unknown class $C+1$, which is not present in the source label space $\mathcal{Y}_s$. The research towards DA in remote sensing images has sprung up recently (Adayel et al., 2020), (Zhang et al., 2020), (Damodaran et al., 2018), (Saha et al., 2011), (Tuia et al., 2016). (Saha et al., 2022) explores the graph neural networks to adapt the model on several target domains. (Nirmal et al., 2020) applies the domain adaptation technique on the hyperspectral images to enhance the model efficiency (Saha et al., 2022).

2.2 Curriculum learning

Curriculum learning describes a type of learning in which the model starts out with only easy examples of a task and then gradually increases the task difficulty. It is popular in many DA methods. (Roy et al., 2021) determine the 'easy-to-hard' learning sequence for Multi-target DA by computing the entropy of target domains returned by the current model. (Luo et al., 2020) gradually select the target samples with extreme confidence score in multiple steps to separate unknowns and upgrade the source domain. (Liu et al., 2019) follow a 'coarse-to-fine' learning strategy to incrementally force the unknown samples far away from any known sets. More importantly, it has been proven that the curriculum strategy can achieve a tighter upper bound of the target error (Luo et al., 2020). However, how to estimate the difficulty level of samples and the right pacing function for introducing more difficult tasks are the key challenges for the curriculum learning (Narvekar et al., 2020).

Inspired by the above works, in this paper, we explore the level of difficulty in the sample's level and rank their difficulty using the confidence score, which is the maximum of the softmax of the prediction logits. For in-domain examples which are confidently predicted, the cross-entropy loss maximizes the logit value of the correct class. On the contrary, the network tends to produce uniformly negative logits for the unknown class. Our model uses a single GCN head for prediction. With the simplified architecture, our model is still capable to achieve comparable accuracy to the dual heads in (Roy et al., 2021). Moreover, only a small set of the target samples are pseudolabeled to avoid the possible negative transfer in (Luo et al., 2020).

2.3 Graph neural networks

Recently GNNs have shown excellent performance in many remote sensing tasks owing to their capability to handle both local and global context and complex interrelationship between data. GNN encodes the local information by generating a node's representation as an aggregation of its own features and those of its neighbors. Similarly, the global information is encoded by stacking multiple graph convolution layers. Semi-supervised frameworks are popular in GNN, especially for node-level classification (Kipf and Welling, 2016). Given a graph for which some nodes have known label, the GNN is capable to assign labels to the unknown samples. Our work borrows from the same idea where labels for the source domain samples are known and labels for target domain samples are assigned by the GNN in the adaptation process. For the creation of a graph, some works decompose an image into many superpixels and treat each superpixel as a node in the graph (Saha et al., 2020). Other works treat each image as a node in the graph (Roy et al., 2021). In this work, we treat each image as a node in the graph and form minibatches drawing images from both source and the target domain. Following this, the adjacency matrix is formed by defining the relationship among images (i.e., nodes) in the minibatch.

3. PROPOSED METHOD

The proposed architecture consists of a backbone feature extractor F, a graph network G_{gcn}, and a domain discriminator D, as shown in Figure 1. Batches of images are processed by the feature extractor F. Extracted features are fed to the G_{gcn} that outputs the class label of samples. G_{gcn} is first pretrained on the batches sampled solely from source domain for

K epochs. Following this, the target samples are also used to form batches and GCN is used to assign pseudolabels to the target samples in a progressive scheme. The confidence score of the target samples gives us an indication to separate known classes and the unknown class in the target domain. Besides, the confidence scores are also propagated to the domain discriminator for the weighted adversarial learning to promote the alignment of the known class while forcing the unknown class to be far from the shared feature space. We call the proposed framework as Progressive Weighted Graph Learning (PWGL).

3.1 Pretrain on source samples

Labels of source domain samples are known and hence to learn class-specific discriminative features we firstly train the G_{gcn} only on the source domain samples using the cross entropy loss of f_{node} and binary cross entropy loss of f_{edge}. f_{edge} is a neural network that receives the node representation and outputs the activation between 0 and 1 that encodes the similarity between every two samples. The output is an affinity matrix indicating the graph structure in image batches. The ground truth affinity matrix connects the samples from the same class as neighbors. f_{node} is another neural network that allows the communication from edge embedding to node embedding and produces the updated node representation. The final output of f_{node} has C logits, corresponding to C classes.

3.2 Target adaptation

When the pre-training finishes, we expect that unknown samples will produce a uniformly low probability belonging to any one of the shared classes. Here we use the confidence score in Equation 1 as an indication to measure the similarity between each target sample to the source data.

$$w_i = \max\left\{\sigma(\mathbf{g}_i)\right\} = \max\left\{\frac{\exp(\mathbf{g}_{i,j})}{\sum_{j=1}^C \exp(\mathbf{g}_{i,j})_{i=1}^C}\right\} \quad (1)$$

Target samples of the shared known class will have relatively higher w_i compared with target unknown samples, since the latter has low probability to all classes in the source sample.

The target adaptation is performed using progressive pseudolabeling strategy. In each step, firstly, we filter the target samples with extreme high and low confidence score. We assume that the subset of the target samples whose similarity is extreme provides us a relatively reliable indicator of whether the samples are from unknown set. To select the samples with extreme similarities, we follow the incremental learning strategy. There are two parameters controlling the pseudolabeling procedure. First is the enlarge factor ϵ similar to (Luo et al., 2020), which is a decimal number that lies between 0 and 1. In each progressive step, ϵ of unlabeled target samples will be pseudolabeled as one of the known classes or unknown according to their confidence score, and the pseudolabeled known target samples will join the source set for the next adaptation. However, we notice that there is computation redundancy and potential negative transfer if all target samples are pseudolabeled. To overcome this, we introduce another parameter which is the extreme percentage $\gamma \in (0, 1]$. Only γ of the total target samples will be pseudolabeled because only the extreme part gives us reliable information. Also, other pseudolabel selection methods could be used. For example, we can use the fixed value as the threshold, or to filter the pseudolabeled samples based on the statistics of the confidence score distribution.

With the above progressive separation procedure, unknown samples from the target domain will be gradually identified by G_{gcn}. The target sample being pseudolabeled is noted as $\mathcal{S}_p$. If $\text{len}\{\mathcal{S}_p\} \leq \gamma \times \text{len}\{\mathcal{T}\}$:

$$\hat{y}^{(m)} = \begin{cases} C + 1, & \text{if } \operatorname{rank} \max_{i \in Y_s} p(i|x)) \leq \tau_{m_u}^m \\ \hat{y}, & \text{if } \operatorname{rank} \max_{i \in Y_s} p(i|x)) \geq \tau_{m_k}^m \end{cases} \quad (2)$$

m is the labeling step ranging from 0 to γ/ϵ. $\tau_{m_u}^m = r \cdot \epsilon \cdot n_t$ and $\tau_{m_u}^m = r \cdot (1-\epsilon) \cdot n_t$. The hyperparameter r controls the openness of the target domain which is defined by the ratio of the number of unknown samples to the number of known samples. $\operatorname{rank}(\cdot)$ is a global ranking function that ranks the predicted probabilities in ascending order and returns the index list as an output (Luo et al., 2020). After the pseudolabeling, the target sample with the extreme high rank ($\geq \tau_{m_k}^m$) will be assigned to one of C classes and the samples with the extreme low rank ($\leq \tau_{m_u}^m$) will be assigned to $C + 1$ class (unknown). The pseudolabeling procedure considers the openness of the target domain and the ratio of unknown samples to known samples remains as r during the training.

The upgrade of the source domain is done by progressively concatenating source samples with the pseudolabeled known target samples $\mathcal{S}_p$. The new source batch $\mathbf{b_s}$ is sampled from this upgraded source domain, and the target batch $\mathbf{b_t}$ is sampled from the remaining target samples that have not been upgraded to the source domain.

$$\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{S}_p \quad (3)$$
$$\mathcal{T} \leftarrow \mathcal{T} \backslash \mathcal{S}_p \quad (4)$$
$$\mathbf{b_s} \sim \mathcal{S}; \ \mathbf{b_t} \sim \mathcal{T} \quad (5)$$

G_{gcn} is further trained using the cross entropy of f_{node} and the binary cross entropy of f_{edge} on the updated $\mathbf{b_s}$. The ground truth affinity matrix $\mathcal{A}^g$ encodes the samples from the same class as neighbors, i.e., $a_{i,j}^g = 1$ if sample i and j are from the same class and 0 otherwise.

3.3 Known-set only alignment

Except for the hard pseudolabels, G_{gcn} is capable to give the confidence of its prediction. A smaller w_j means a higher probability that $\mathbf{x}_{t,i}$ comes from the unknown class, i.e we interpret $1 - w_j$ as the probability that a target sample is unknown. Instead of aligning the whole source domain to the target domain, here we only concentrate on the known classes to promote the samples in the same class from the different domain getting closer and to force the unknown class far away. The adversarial training is performed on aligning known classes only, and w_j is used to exploit the known-set only alignment forming the adversarial loss by

$$\ell_{adv} = \frac{1}{|\mathcal{B}_s|} \sum_{i=1}^{\mathcal{B}_s} y_{s,i} \log D(F(\mathbf{x}_{s,i}))$$
$$+ \frac{1}{\sum_{x_j \in |\mathcal{B}_t|} w_j} \sum_{i=1}^{\mathcal{B}_t} w_j (1 - y_{t,i}) \log[1 - D(F(\mathbf{x}_{t,i}))] \quad (6)$$

A simple yet effective solution is to make use of the hard label $\hat{y}$ computed from Equation 2 and reject all the unknown samples directly by setting their weight as zero.
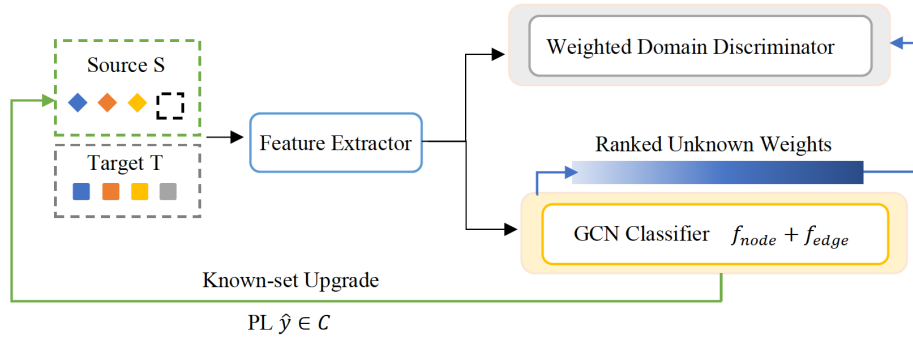
Figure 1. Openset DA Model

The overall objective is

$$\ell = \lambda_{node}\ell_{node} + \ell_{edge} - \lambda_{adv}\ell_{adv} \qquad (7)$$

The G_{gcn} is trained to minimize the loss while the D is trained to maximize the loss.

4. EXPERIMENTAL VALIDATION

4.1 Dataset

We conduct experiments on a multi-city subset from So2Sat LCZ42 Dataset (Zhu et al., 2020). More specifically, we select two cities Moscow and Mumbai from Europe and Asia, respectively. Due to their geographic and anthropogenic variation, they exhibit significant difference. We chose 6 classes (Compact mid-rise/low-rise, Open high-rise/mid-rise/low-rise, Large low-rise, Sparsely built, Trees, and Bush and Low plants) and approximately 800 images per class per city. We show results taking Mumbai as source city and Moscow as the target city and vice-versa.

4.2 Settings

We use the pretrained ResNet-50 as the feature extractor backbone. Among 6 classes, the last one is set as unknown (present only in the target domain) in our experiments and the rest as known classes (present in both source and target domains). While there are several hyperparamters, the most important parameters are listed as bellow. The number of known classes (C) is 5. The openness (r) is set as 1/C, i.e., 1/5. K is 1000 epochs. The extreme percentage (γ) is 0.2, and ϵ is 0.05. In total, 20% of the target samples is pseudo labeled in 5 steps. The network is optimized by ADAM with the weight decay of 5e-5 and an initial learning rate 1e-4. The λ_{adv} is 1 and λ_{node} is 0.3. In each adaptation iteration, the size of both $\mathbf{b_s}$ and $\mathbf{b_t}$ is 32.

We use the normalized accuracy for all classes (OS) and normalized accuracy for the known classes (OS*) to evaluate the performance of the model.

$$OS = \frac{1}{C+1}\sum_{i=1}^{C+1}\frac{x : x \in \mathcal{D}_t^i \cap \hat{y}(x) = i}{|x : x \in \mathcal{D}_t^i|} \qquad (8)$$

$$OS^* = \frac{1}{C}\sum_{i=1}^{C}\frac{x : x \in \mathcal{D}_t^i \cap \hat{y}(x) = i}{|x : x \in \mathcal{D}_t^i|} \qquad (9)$$

$\mathcal{D}_t^i$ is the target samples in the i-th class, and $\hat{y}$ is the prediction of the node classifier.

4.3 Compared methods

Our work is one of the first works in remote sensing for openset domain adaptation. We compare the proposed method to the following methods:

1. Without adaptation (**w.o**). The model is trained on the labeled source domain and tested directly on the target domain.

2. Curriculum Graph Co-Teaching for Multi-Target Domain Adaptation (**CGCT**) (Roy et al., 2021). Two classifier heads GCN and MLP that co-teach each other to identify known and unknown samples using the fixed thresholds (we adapt the original CGCT by adding another mask for selecting low-confidence samples). We report the accuracy returned by the GCN head.

4.4 Result

Table 1 reports the accuracy of each class, OS, and OS*. We notice that some of the class accuracies are high, such as class 1, while for some classes the performance is bad, such as class 2. The adaptation Mumbai → Moscow achieves the best performance: 67.17% samples are correctly classified with 74.5% of the total unknown samples is recognized. The accuracy of Moscow → Mumbai is less prominent. Such deviation could be explained by the difference of the data distribution. The model learnt on an intrinsically abundant domain obtains higher generalization ability and is able to achieve better performance on other domains. Figure 2 is the t-SNE visualization of feature distribution on Mumbai → Moscow task. The left one is the feature distribution at the beginning of the adaptation and the right one is the feature distribution after the adaptation by PWGL. When the model is not well trained, the unknown class is mixed with other known classes and the features cluster together make them indistinguishable. However, after the adaptation, the unknown class is forced farther away from any known set and the rest known classes are grouped into C clusters.

Table 2 is the performance comparison between PWGL and other methods. We adapt the CGCT to open-set adaptation problem by adding another mask to select the low confidence samples as unknown set. Without adaptation, very few samples are identified as unknown, due to the reason that the cross entropy is computed based on the known class only and the model
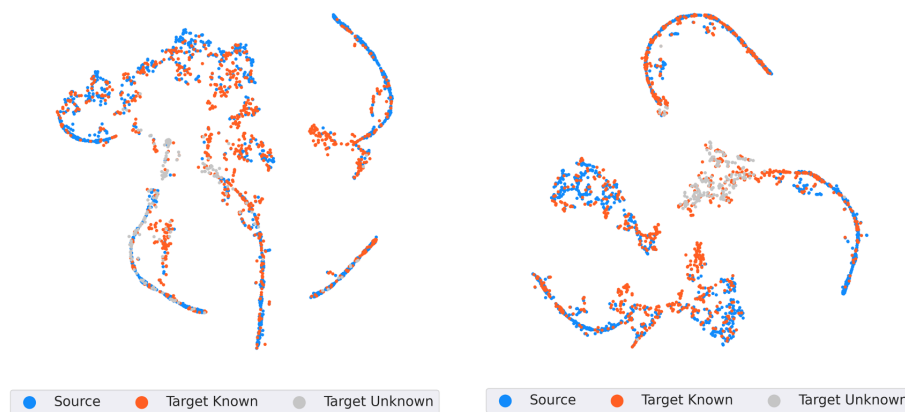
Figure 2. T-SNE before and after adaptation, taking Mumbai as the source and Moscow as the target

Result	Mu-Mo	Mo-Mu
Class 1	71.9	69.6
Class 2	35.4	21.1
Class 3	62.6	60.4
Class 4	74.1	86.4
Class 5	84.5	65.9
Unknown	74.5	31.2
OS	67.17	55.77
OS*	65.70	60.68

Table 1. Performance of the proposed model for Mumbai (Mu) as source and Moscow (Mo) as target and vice-versa. Individual accuracies on the 5 known classes and the unknown class are shown. Furthermore, OS and OS* are shown.

Domain	Mu-Mo		Mo-Mu	
Index	unk	OS	unk	OS
w.o	1.40	37.69	0.25	40.00
CGCT*	24.10	62.93	0.63	31.21
Proposed	74.51	67.17	31.23	55.77

Table 2. Performance comparison of the proposed method to the existing methods.

hasn't seen unknown data. The improvement of the unknown accuracy brought by CGCT is very small, but PWGL improves the accuracy of unknown to a large extent. These data demonstrate that the known-set only alignment in PWGL is necessary for unknown separation. It tries to force the unknown set away from the common feature space and thus it's easier for PWGL to recognize the unknown samples. On the contrary, CGCT attempts to close the gap between the source domain and the target domain without differentiating known or unknown samples. As a result, the unknown samples become closer to the source domain and make the model hard to identify them. More importantly, the rank-based filtering strategy guarantees that in each round, there is a sufficient number of samples are labeled as unknowns despite of the shift of the confidence distribution during the training procedure. Generally speaking, our experiment verifies the efficacy of PWGL in identifying both unknown and known classes. This also emphasizes the role of extreme percentage γ in suppressing negative transfer.
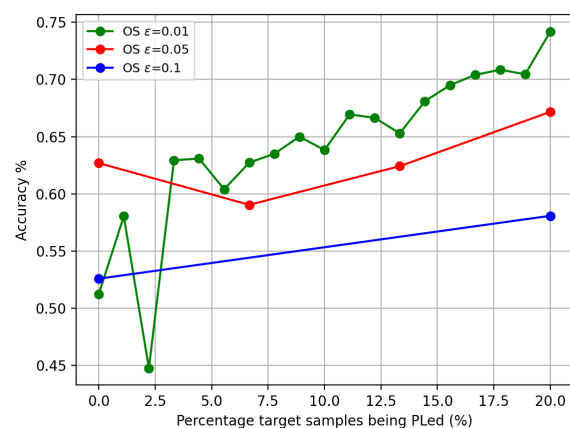

Figure 3. Impact of the progressive step to the OS accuracy

4.5 Effect of progressive learning step

One important parameter in the proposed method is the progressive learning step size. To further understand its impact on the model performance, we use different values of ϵ, i.e., we use $\epsilon = 0.05$, $\epsilon = 0.01$, and $\epsilon = 0.1$. The OS is plotted as Figure 3. The figure reveals that the smaller step size can return a smoother increase of the normalized accuracy over all classes (OS), and it produces a higher final OS that the model could be able to deliver. But it increases the computation cost. The OS improvement from $\epsilon = 0.01$ to $\epsilon = 0.05$ is 7.02%, but the training effort is increased by 5 times. With the aggressive learning strategy, the computation cost is reduced with the compromise of accuracy.

5. CONCLUSION

This paper proposes a method for open-set domain adaptation, which is a practical however challenging extension of the more popularly addressed close-set domain adaptation. Towards this, we propose a GNN-based architecture PWGL to recognize the additional unknown set in a progressive way while encouraging within-class compactness. We start from the observation of the logits distribution of unknown set and known set. The unknown samples are separated by identifying low confidence samples.

PWGL utilizes the GNNs to aggregate similar samples across domains by promoting the communication between edge and node embedding. The unknown class is rejected in the process of closing domain gaps to force the unknown samples away from the shared feature space. The result on multi-city dataset validates the effectiveness of the proposed method. In the future, we plan to experiment on other settings, e.g., domains described by different sensors. Furthermore, we plan to extend the task by considering the scenario where some classes from the source domain are not present in the target domain. In addition, the work can be extended to semantic segmentation to address the shifts between various semantic scenarios.

ACKNOWLEDGEMENTS

REFERENCES

Adayel, R., Bazi, Y., Alhichri, H., Alajlan, N., 2020. Deep open-set domain adaptation for cross-scene classification based on adversarial learning and pareto ranking. *Remote Sensing*, 12(11), 1716.

Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Wortman, J., 2007. Learning bounds for domain adaptation. *Advances in Neural Information Processing Systems*, 20.

Chen, Y., Song, S., Li, S., Wu, C., 2019. A graph embedding framework for maximum mean discrepancy-based domain adaptation algorithms. *IEEE Transactions on Image Processing*, 29, 199–213.

Damodaran, B. B., Kellenberger, B., Flamary, R., Tuia, D., Courty, N., 2018. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 447–463.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. *International Conference on Machine Learning*, PMLR, 1180–1189.

Gawlikowski, J., Saha, S., Kruspe, A., Zhu, X. X., 2021. Towards out-of-distribution detection for remote sensing. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 8676–8679.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

Kipf, T. N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Liu, H., Cao, Z., Long, M., Wang, J., Yang, Q., 2019. Separate to adapt: Open set domain adaptation via progressive separation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2927–2936.

Long, M., Cao, Z., Wang, J., Jordan, M. I., 2017. Conditional adversarial domain adaptation. *arXiv preprint arXiv:1705.10667*.

Luo, Y., Wang, Z., Huang, Z., Baktashmotlagh, M., 2020. Progressive graph learning for open-set domain adaptation. *International Conference on Machine Learning*, PMLR, 6468–6478.

Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M. E., Stone, P., 2020. Curriculum learning for reinforcement learning domains: A framework and survey. *arXiv preprint arXiv:2003.04960*.

Nirmal, S., Sowmya, V., Soman, K., 2020. Open set domain adaptation for hyperspectral image classification using generative adversarial network. *Inventive Communication and Computational Technologies*, Springer, 819–827.

Pan, Y., Yao, T., Li, Y., Wang, Y., Ngo, C.-W., Mei, T., 2019. Transferrable prototypical networks for unsupervised domain adaptation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2239–2247.

Panareda Busto, P., Gall, J., 2017. Open set domain adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, 754–763.

Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B., 2019. Moment matching for multi-source domain adaptation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1406–1415.

Rakshit, S., Banerjee, B., Roig, G., Chaudhuri, S., 2019. Unsupervised multi-source domain adaptation driven by deep adversarial ensemble learning. *German Conference on Pattern Recognition*, Springer, 485–498.

Roy, S., Krivosheev, E., Zhong, Z., Sebe, N., Ricci, E., 2021. Curriculum graph co-teaching for multi-target domain adaptation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5351–5360.

Saha, A., Rai, P., Daumé, H., Venkatasubramanian, S., DuVall, S. L., 2011. Active supervised domain adaptation. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, 97–112.

Saha, S., Banerjee, B., Merchant, S. N., 2016. Unsupervised domain adaptation without source domain training samples: a maximum margin clustering based approach. *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, 1–8.

Saha, S., Mou, L., Zhu, X. X., Bovolo, F., Bruzzone, L., 2020. Semisupervised Change Detection Using Graph Convolutional Network. *IEEE Geoscience and Remote Sensing Letters*.

Saha, S., Zhao, S., Zhu, X. X., 2022. Multi-target domain adaptation for remote sensing classification using graph neural network. *IEEE Geoscience and Remote Sensing Letters*.

XXIV ISPRS Congress (2022 edition), 6–11 June 2022, Nice, France

Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 41–57.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 7167–7176.

Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S., 2017. Deep hashing network for unsupervised domain adaptation. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 5018–5027.

Zhang, J., Liu, J., Pan, B., Chen, Z., Xu, X., Shi, Z., 2021. An Open Set Domain Adaptation Algorithm via Exploring Transferability and Discriminability for Remote Sensing Image Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*.

Zhang, J., Liu, J., Shi, L., Pan, B., Xu, X., 2020. An open set domain adaptation network based on adversarial learning for remote sensing image scene classification. *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 1365–1368.

Zhu, X. X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., Bagheri, H., Haberle, M., Hua, Y., Huang, R., Hughes, L., Li, H., Sun, Y., Zhang, G., Han, S., Schmitt, M., Wang, Y., 2020. So2Sat LCZ42: A Benchmark Data Set for the Classification of Global Local Climate Zones. *IEEE Geoscience and Remote Sensing Magazine*, 8(3), 76-89.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.