Reply to Editor

This paper was reviewed by two experts. The major concern is the motivation is not very clear. Besides, they also identify many issues from equations, experiments to writing. The authors should address the comments in revision.

Response:

Thanks so much for your precious comments on our manuscript. The suggestions from the expert reviewers are of great help to improve our work.

This paper is about *Open Set Domain Adaptation (OSDA)*. Existing methods developed for OSDA attempt to assign smaller weights to target samples of unknown class to alleviate negative transfer. Despite promising performance achieved by existing methods, the samples of the unknown class are still used for distribution alignment, which make the model suffer from the risk of the negative transfer. Instead of reweighting, this paper presents a novel Thresholded Domain Adversarial Network (*ThDAN*) for OSDA to progressively select transferable target samples for domain adversarial training.

In the revised manuscript, the following contents have been modified,

Abstract and Introduction

According to the reviewers' advice, we have reorganized the abstract and introduction to give a tighter logical connection and clearer motivation. In the revised manuscript, the abstract has been rewritten as follows,

> *In recent years, many unsupervised domain adaptation (UDA) methods have been proposed to tackle the domain shift problem. Most existing UDA methods are derived for Close Set Domain Adaptation (CSDA) in which source and target domains are assumed to share the same label space. However, target domain may contain unknown class different from the known ones in the source domain in practice, i.e., Open Set Domain Adaptation (OSDA). Due to the presence of unknown class, aligning the whole distribution of the source and target domain for OSDA as in the previous methods will lead to negative transfer. Existing methods developed for OSDA attempt to assign smaller weights to target samples of unknown class. Despite promising performance achieved by existing methods, the samples of the unknown class are still used for distribution alignment, which make the model suffer from the risk of negative transfer. Instead of reweighting, this paper presents a novel method namely Thresholded Domain Adversarial Network (ThDAN), which progressively selects transferable target samples for distribution alignment. Based on the fact that samples from the known classes must be more transferable than target samples of the unknown one, we derive a criterion to quantify the transferability by constructing classifiers to categorize known classes and to discriminate unknown class. In ThDAN, an adaptive threshold is calculated by averaging transferability scores of source domain samples to select target samples for training. The threshold is tweaked progressively during the training process so that more and more target samples from the known classes can be correctly selected for adversarial training. Extensive experiments show that the proposed method outperforms state-of-the-art domain adaptation and open set recognition approaches on benchmarks.*

According to the redesigned abstract, we have modified the **introduction** to make it connect to the abstract tightly. The following Table 1 shows the logical connection between each paragraph of the introduction and each sentence of the abstract.

Table 1: The logical connection between paragraphs of the introduction and sentences of the abstract.

Paragraph of Introduction	Sentence of Abstract
# 1	*In recent years, many unsupervised domain adaptation (UDA) methods have been proposed to tackle the domain shift problem.*
# 2	*Most existing UDA methods are derived for Close Set Domain Adaptation (CSDA) in which source and target domains are assumed to share the same label space. However, target domain may contain unknown class different from the known ones in the source domain in practice, i.e., Open Set Domain Adaptation (OSDA).*
# 3	*Existing methods developed for OSDA attempt to assign smaller weights to target samples of unknown class. Despite promising performance achieved by existing methods, the samples of the unknown class are still used for training, which make the model suffer from the risk of negative transfer.*
# 4	*Instead of reweighting, this paper presents a novel method namely Thresholded Domain Adversarial Network (ThDAN), which progressively selects transferable target samples for distribution alignment.* .....

Section 3. Preliminaries

According to the reviewers' comments, we have added a new section for better understanding of preliminaries of the proposed method. This new section contains the following two subsections,

- *3.1. Open Set Domain Adaptation*
- *3.2. Domain Adversarial Training*

In Section 3.1, we clarify the settings of the *Open Set Domain Adaptation* [1], while *domain adversarial training* [2] is introduced in Section 3.2.

Section 4. Method

We have reorganized section *4. Method* and added the following subsection to clarify the proposed method and equations,

- *4.1. Thresholded Domain Adversarial Network*

In section 4.1, we introduce the idea and training procedure of the proposed ThDAN. Then, details about the transferability based sample selection algorithm used in ThDAN is given in the following subsections. We believe that organizing section 4 in this way can help readers better understand our work.

Section 5. Experiments

To perform more comprehensive evaluation, the following modifications have been made in section *5.2. Classification Results*,

- Experiment results of recently proposed *Factorized Representations for Open Set Domain Adaptation* ***(FRFOSDA)*** model [3] are added in sections *5.2.1. Result on Office-31* and *5.2.2. Result on Office-Home* for comparison.
- More detailed analysis of the model performance on VisDA dataset are added in section *5.2.3. Result on VisDA*.

In section *5.3. Analysis*, we revise the paper with the following modifications,

- Section *5.3.1 Ablation Study* is added to better understand how the proposed method performs under ablation settings.
- More analysis is added in section *5.3.4 Varing the Number of Unknown Samples and Number of Unknown Classes* to explain how the performance changes with different sampling proportions between the data from known and unknown classes.

- In section *5.3.6 Change in Upper Bound of Transferability Offset*, we analyze why the proposed method is insensitive to the setting of γ_0 with more details.

Figure Caption

According to the reviewers' suggestions, the captions of Figures 1 and 7 have been modified for better understanding.

Current work aligns the entire target domain with the source domain without excluding unknown samples, which may give rise to negative transfer due to the mismatch between unknown and known classes.

References

[1] K. Saito, S. Yamamoto, Y. Ushiku, T. Harada, Open set domain adaptation by backpropagation, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 153–168.

[2] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The Journal of Machine Learning Research 17 (1) (2016) 2096–2030.

[3] M. Baktashmotlagh, M. Faraki, T. Drummond, M. Salzmann, Learning factorized representations for open-set domain adaptation, in: International Conference on Learning Representations, 2019.