

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 of reviewers are of great help to our work.

The subject of this paper is *Open Set Domain Adaptation (OSDA)*. Existing methods developed for OSDA try to assign smaller weights to target samples of unknown class to alliterate negative transfer. Despite promising performance, the samples of the unknown class are still used for training, which exposes the model to 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

Following the advice of reviewers, we redesigned the abstract and introduction to have a tighter logical connection and clearer motivation. In the revised manuscript, the abstract has being redesigned as follows,

Unsupervised domain adaptation (UDA) is a paradigm to tackle domain shift problem. Most existing UDA methods are proposed for Close Set Domain Adaptation (CSDA) which assumes source and target domains share the same label space. More practically, target domain contains unknown class different from the known ones in the source domain, i.e., Open Set Domain Adaptation (OSDA). Due to the presence of unknown classes, aligning the whole distribution of the source and target domain for OSDA as in the previous method will lead to negative transfer. Existing methods developed for OSDA try to assign smaller weights to target samples of unknown class. Despite promising performance, the samples of unknown class are still used for training, which exposes the model to 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.

In the revised abstract, we first introduce the unsupervised domain adaptation (UDA). Then we make an introduction to OSDA and the challenge of it. To clarify the motivation of our work, we then state the working mechanism of existing models proposed for OSDA and the drawbacks of it. Finally, we present the general idea of the proposed Thresholded Domain Adversarial Network (ThDAN).

Following the logic of the redesigned abstract, we modified the **introduction** to make it more tightly connected to the abstract. For example, in the first paragraph of the introduction, we first introduce the domain shift problem, then lead to the topic of unsupervised domain adaptation, which is logically related to the first sentence of the abstract.

Method Section

We reorganized section 3. *Method* and added the following subsection to facilitate the understanding of the proposed methods and equations,

- 3.1. *Preliminary: Domain Adversarial Training*
- 3.2. *Thresholded Domain Adversarial Network*

In Section 3.1, we introduce *domain adversarial training* [1] as the preliminary of the proposed model. Then in Section 3.2, we introduce the idea and training procedure of the proposed ThDAN. In the following sections, we detailed introduce the transferability based sample selection algorithm used in ThDAN. We consider such an organization can help readers better understand our work.

Experiments Section

In order to better demonstrate the performance of the proposed model, the following modifications have been made in section 4.2. *Classification Results*,

- Experiment results of newly proposed *Factorized Representations For Open Set Domain Adaptation (FRFOSDA)* model [2] have been added in sections 4.2.1. *Result on Office-31* and 4.2.2. *Result on Office-Home*.
- More detailed analysis of model performance on VisDA dataset has been added in section 4.2.3. *Result on VisDA*.

In section 4.3. *Analysis*, we made following modifications to deliver more convincing analysis for the proposed model,

- The writing of the ablation study in the original is confusing since it separated in different subsections. In the revised manuscript, we unify the ablation study and reorganized in section 4.3.1 *Ablation Study*.
- More analysis has been added in section 4.3.4 *Varying the Number of Unknown Samples and Number of Unknown Classes* to indicate the effect of sampling proportion between the data from known and unknown classes on model performance.
- More detailed reasons on why the model is insensitive to the setting of γ_0 have been added in section 4.3.6 *Change in Upper Bound of Transferability Offset*.

Figure Caption

After reading the suggestions of reviewers, we notice the claim of some figures is confusing. Therefore we modified the caption of Figure 1 and 7 to make clearer claims

References

- [1] 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.
- [2] M. Baktashmotlagh, M. Faraki, T. Drummond, M. Salzmann, Learning factorized representations for open-set domain adaptation, in: *International Conference on Learning Representations*, 2019.