Supplementary materials for paper "Multi-Source open-set deep adversarial domain adaptation"

Sayan Rakshit[1][0000−0002−0189−7257], Dipesh Tamboli[1], Pragati Shuddhodhan Meshram[1], Biplab Banerjee[1][0000−0001−8371−8138], Gemma Roig[2][0000−0002−6439−8076], and Subhasis Chaudhuri[1]

[1] Indian Institute of Technology Bombay, Mumbai, India
{sayan1by2, dipeshtamboli1, pragatimeshram30, getbiplab}@gmail.com, sc@iitb.ac.in
[2] Goethe University Frankfurt, Frankfurt, Germany
roig@cs.uni-frankfurt.de

1 Introduction

In the supplementary materials, we mention the followings:

- We showcase results on a subset of domains of the recently introduced Domain-net dataset [4].
- We show results on the ImageCLEF dataset.
- We include the results of Office-Home dataset where the openness is high $\mathcal{O} \to 1$.
- Analysis of Proxy-distance ($\mathcal{A}$) for the setting **A,D** $\mapsto$ **W** of Office-31.
- Stability of MOSDANET.
- Incremental analysis of the different components of MOSDANET.
- We depict an algorithm (Algorithm 1) of the training process of MOSDANET. The notations and loss functions mentioned in the main paper are used in Algorithm 1.

2 Results on the subset of the Domain-net dataset

The extremely large-scale Domain-net [4] dataset was introduced as a multi-source dataset for the VisDA challenge and the dataset has only been used for multi-source domain adaptation so far. We propose the following experimental protocol for carrying out MS-OSDA for Domain-net. The original dataset contains six visual domains with images from 345 categories per domain. Due to resource constraints, we consider four domains in our experiments: Painting (**P**), Clipart (**C**), Infograph (**I**), and Sketch (**S**), respectively. We further consider 100 known classes and 245 open-set classes as per the alphabetic order for two experimental settings: **I,P** $\mapsto$ **S** and **I,P** $\mapsto$ **C** and we randomly sample 50 images per class for all the domains. If the number of images for a class is less than 50, we use all the images. This leads to approximately $16,000$ images per domain. For constructing the feature encoder $\mathcal{E}$, we consider a modified version of the Imagenet pre-trained VGG-16 [7] model where the classification layer is replaced by three new fully-connected layers. Batch-normalization and Leaky-ReLU non-linearity are used in conjunction to all the three new layers (similar to the Office dataset, Section 4 main paper). While the original VGG-16 layers are freezed during model training, the parameters relating to the three new layers are updated. We compare the performance of MOSDANET with the single-best and source-combine versions of OSVM [6], OSDA-BP [5], and IOSDA-BP [2], respectively. Due to the high complexity, all the techniques perform poorly for this dataset. MOSDANET, in the other hand achieves at least a 9% improvement in the performance with respect to OSDA-BP[5], which is the best performing among the state-of-the-art methods.

Method	I,P - S		I,P - C		Average	
	OS*	OS	OS*	OS	OS*	OS
OSVM[6]	9.6	9.1	10.9	8.9	10.2	9.0
OSDA-BP [5] (single best)	21.6	25.3	23.0	24.5	22.3	24.9
OSDA-BP[5] (source combine)	1.0	1.1	1.6	1.6	1.3	1.3
IOSDA-BP[2] (single best)	14.5	12.6	17.2	17.0	15.8	14.8
IOSDA-BP[2] (source combine)	12.0	10.2	20.6	20.4	16.3	15.3
MOSDANET	**27.2**	**28.3**	**29.2**	**29.5**	**28.2**	**28.9**

Table 1. Performance comparison for MS-OSDANET for 100 shared and 245 unknown classes for VisDA dataset (in %) with the margin parameter $\tau = 0.6$.

3 Results on ImageClef dataset

The ImageClef dataset contain images from 12 shared classes of Imagenet-2012 (**I**), Caltech-256 (**C**), and Pascal-VOC (**P**), respectively, with 50 images per class per domain. We consider all the three experimental settings with two source domains and one target domain: $\mathbf{I}, \mathbf{P} \mapsto \mathbf{C}$, $\mathbf{C}, \mathbf{P} \mapsto \mathbf{I}$ [3], and $\mathbf{I}, \mathbf{C} \mapsto \mathbf{P}$, respectively, with 8 shared and 4 open-set classes. We follow similar structure of the feature encoder $\mathcal{E}$ as of the Office datasets which is based on Resnet-50 backbone (Section 4 main paper). Table 2 depicts the comparative analysis for ImageClef. As we observe from the table, MOSDANET results outperform other methods in most cases, and in $\mathbf{C},\mathbf{I} \mapsto \mathbf{P}$, $OS*$ achieves competitive results to the best performing one(OSDA-BP[5] source combine)

[3] Since part of our feature extractor $\mathcal{E}$ contains Imagenet pre-trained Resnet-50 model, it may be argued whether this particular setup violates the assumptions of unsupervised DA. However, note that we keep the Resnet-50 part fixed and train the newly added layers only without utilizing any target-domain supervision, this particular experiment may be justified.

Method	C,I - P		C,P - I		I,P - C		Average	
	OS*	OS	OS*	OS	OS*	OS	OS*	OS
OSVM[6]	74.3	57.8	89.7	67.2	90	72.8	84.7	65.9
OSVM+DANN[3]	64.7	62.9	89.0	72.5	89.7	76.8	81.1	70.7
OSVM+[4]	82.5	55.0	92.5	61.7	96.5	64.4	90.5	60.4
OSDA-BP [5] (single best)	80.1	75.3	90.4	80.0	99.1	88.3	89.7	81.2
OSDA-BP[5] (source combine)	**83.2**	62.1	94.3	70.5	94.0	69.3	90.6	67.3
IOSDA-BP[2] (single best)	80.4	56.2	94.5	64.4	94.7	68.4	89.8	63.0
IOSDA-BP[2] (source combine)	73.4	72.7	89.3	83.0	93.9	90.0	85.5	81.9
MOSDANET	77.6	**75.8**	**94.5**	**86.3**	**99.7**	**92.1**	**90.6**	**84.7**

Table 2. Performance comparison for MS-OSDANET for 8 shared and 4 unknown classes for ImageClef dataset (in %) with the margin parameter $\tau = 0.6$.

Method	A,C,P - R	A,P,R - C	P,C,R - A	A,C,R - P	Average
	OS	OS	OS	OS	OS
OSVM[6]	26.1	11.1	18.5	19.2	18.7
OSVM+DANN[3]	27.9	10.3	27.5	26.8	23.1
OSVM+[4]	25.9	20.5	25.4	22.1	23.5
OSDA-BP[5] (single best))	50.6	63.6	64.2	61.5	60.0
OSDA-BP[5](source combine))	44.1	42.5	47.2	49.3	45.8
IOSDA-BP[2](single best)	43.4	30.3	39.5	29.2	35.6
IOSDA-BP[2](source combine)	54.9	42.4	49.4	58.2	51.2
MOSDANET	**88.8**	**73.5**	**75.3**	**88.8**	**81.6**

Table 3. Performance comparison for MOSDANET for 25 shared and 40 unknown classes for Office-Home dataset (in %) with the margin parameter $\tau = 0.6$.

4 Experiments on Office-Home on a new known/unknown split

In order to further assess the robustness of our method to different openness scores, we conduct another experiment on Office-Home [8] dataset when 25 shared and 40 open-set classes are considered as per the alphabetic order (Table 3). This is a challenging scenario considering the large domain-gap among the four domains of Office-Home and the fine-grained nature of the different categories. This accounts for an openness value $\mathcal{O} \to 1$. We hypothesize in the main paper (Section 4.2) that MOSDANET is robust to high openness scenarios. In the same line, we find in Table 3 that MOSDANET is able to produce extremely high OS values (more than 30% than the literature) for all the cases in comparison to the other techniques considered.

5 Analysis of the proxy distance $\mathcal{A}$ of MOSDANET

$\mathcal{A}$-distance is a measure of cross-domain discrepancy [1]. Along with the source risk, $\mathcal{A}$ is used to bound the true risk in the target domain. Generally speaking,

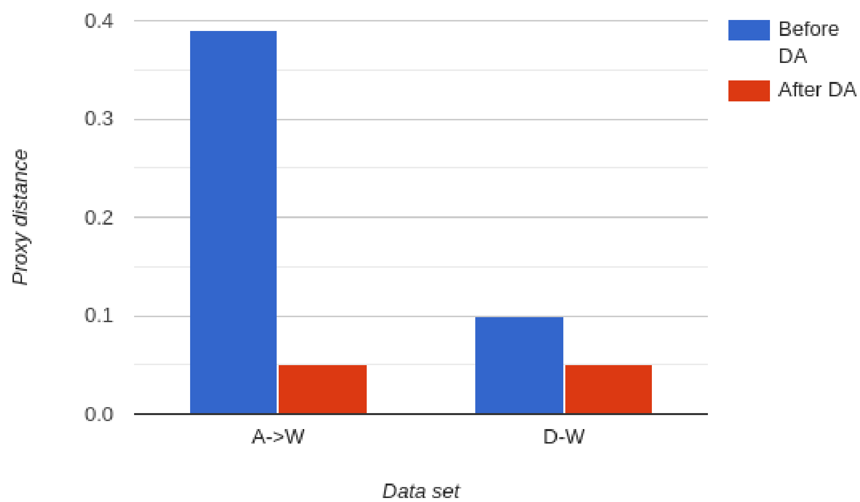

Fig. 1. proxy distance analysis for **A,D** $\mapsto$ **W** (Office-31) with 20 known and 11 open-set classes.

$\mathcal{A}$-distance is defined as $d_{\mathcal{A}} = 2(1 - 1\epsilon)$ where ϵ is the generalization error of a binary classification task devoted to the discrimination of the domains. A small $d_{\mathcal{A}}$ corresponds to high domain similarity. Figure 1 depicts the analysis on $\mathcal{A}$ for **A,D** $\mapsto$ **W** (Office-31). It can be observed that the $d_{\mathcal{A}}$ is large before adaptation both for **A** $\mapsto$ **W** and **D** $\mapsto$ **W** but reduces drastically after adaptation. It is also to be noted that $d_{\mathcal{A}}$ for both the source-domains become equal after adaptation, which signifies proper alignment between the source-domains in the shared feature space.

6 Stability of MOSDANET

In order to assess the stability of MOSDANET, we perform multiple runs of the algorithm for every dataset and consider the variance in the reported accuracies. We find that MOSDANET produces variance ≤ 1 in all the cases for ten runs.

7 Incremental analysis of MOSDANET

In this case, we consider three models: i) Base model-1 (without margin loss and pseudo-labeling), ii) Base model-2 (with margin loss and without pseudo-labeling), and iii) the full model. Below, we show the analysis with respect to these three models.

	Office-31	Office-Home	Office-CalTech	Digits
Base model-1	85.4 ± 0.6	66.1 ± 1.1	87.6 ± 0.9	80.2 ± 0.5
Base model-2	90.3 ±0.7	70.3 ± 0.9	92.5 ± 0.9	84.5 ± 0.5
Full model	91.9 ± 0.6	72.1 ± 0.5	95.8 ± 0.4	87.1 ± 0.2

Table 4.

8 Training algorithm

Algorithm 1 Training Algorithm for MOSDANET

Input: $\mathbb{X}_s = \{\mathbb{X}_l\}_{l=1}^{L}$, $\mathbb{X}_t$, Initial $\theta_{\mathcal{E}}$ and $\theta_{\mathcal{F}}$
Output: $\hat{\theta}_{\mathcal{E}}$, $\hat{\theta}_{\mathcal{F}}$ {Trained model parameters}
1: $ep = 0$ {training iteration}
2: $\mathbb{X} = \mathbb{X}_s$ {$\mathbb{X}$ is the temporary variable which is used to evaluate $\mathcal{L}_{SA}$}
3: **while** Not converged **do**
4: Optimize Equation 5 (main paper)
5: Optimize Equation 6 (main paper)
6: Obtain $\hat{\mathbb{X}}_t^{ep}$ using Rule mentioned in Equation 4 (main paper) {$\hat{\mathbb{X}}_t^{ep}$ denotes target-domain samples with pseudo known class-labels}
7: $\mathbb{X} = \mathbb{X} \cup \hat{\mathbb{X}}_t^{ep}$ {Consideration of selected target samples with pseudo-labels for next iteration}
8: $ep = ep + 1$
9: **end while**
10: Obtain $\hat{\theta}_{\mathcal{E}}$, $\hat{\theta}_{\mathcal{F}}$

References

1. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1-2), 151–175 (2010)
2. Fu, J., Wu, X., Zhang, S., Yan, J.: Improved open set domain adaptation with backpropagation. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 2506–2510. IEEE (2019)
3. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. arXiv preprint arXiv:1409.7495 (2014)
4. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. arXiv preprint arXiv:1812.01754 (2018)
5. Saito, K., Yamamoto, S., Ushiku, Y., Harada, T.: Open set domain adaptation by backpropagation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 153–168 (2018)
6. Scheirer, W.J., Jain, L.P., Boult, T.E.: Probability models for open set recognition. IEEE transactions on pattern analysis and machine intelligence **36**(11), 2317–2324 (2014)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
8. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5018–5027 (2017)