A Reproducible and Realistic Evaluation of Partial Domain Adaptation Methods

Supplementary material

Outline. The supplementary material of this paper is organized as follows:

- In Section A, we give more details on our experimental protocol.
- In Section B, we provide additional results from our experiments.

A Additional details on Experimental Protocol

A.1 Implementations in BenchmarkPDA

In order to reimplement the different PDA methods, we adapted the code from the official repository associated with each of the paper. We list them in Table 1.

Method	Code Repository
PADA	`https://github.com/thuml/PADA/blob/master/pytorch/src/`
SAFN	`https://github.com/jihanyang/AFN/blob/master/partial/OfficeHome/SAFN/code/`
BA3US	`https://github.com/tim-learn/BA3US/`
AR	`https://github.com/XJTU-XGU/Adversarial-Reweighting-for-Partial-Domain-Adaptation`
JUMBOT	`https://github.com/kilianFatras/JUMBOT`
M-POT	`https://github.com/UT-Austin-Data-Science-Group/Mini-batch-OT/tree/master/PartialDA`

Table 1: Office Github code repositories for the PDA methods considered in this work.

One of our main claims regarding previous work is the use of target labels to choose the best model along training. This can be easily verified by inspecting the code. For PADA it can be seen on line 240 of the script "train_pada.py", for BA3US in line 116 for the script "run_partial.py", for M-POT it can be seen line 164 of the file "run_mOT.py", for SAFN it can be seen in the "eval.py" file and finally for AR in line 149 of the script "train.py".

For JUMBOT and M-POT which are based on optimal transport, we used the optimal transport solvers from (Flamary et al., 2021).

A.2 Model Selection

DEV requires learning a discriminative model to distinguish source samples from target samples. Its neural network architecture must be specified as well the training details. You et al. (2019) (DEV) use a multilayer perceptron, while Saito et al. (2021) (SND) use a Support Vector Machine in their reimplementation of DEV. We empirically observed the latter to yield more stable weights and so that was the one we used. In order to train the SVM discriminator, following (Saito et al., 2021), we take 3000 feature embeddings from source samples used in training and 3000 random feature embeddings from target samples, both chosen randomly. We do a 80/20 split into training and test data. The SVM is trained with a linear kernel for a maximum of 4000 iterations. Of 5 different SVM models trained with decay values spaced evenly on log space between 10^{-2} and 10^{4} the one that leads to the highest accuracy (in distinguishing source from target features) on the test data split is the chosen one.

As for SND, it also requires specifying a temperature for temperature scaling component of the strategy. We used the default value of 0.05 that is suggested in (Saito et al., 2021).

Finally, we mention that the samples used for 100-RND were randomly selected and their list is made available together with the code. As for the samples used for 1-SHOT, they are the same as the ones used in semi-supervised domain adaptation.

A.3 Optimizer

In general, all methods claim to adopt Nesterov's acceleration method as the optimization method with a momentum of 0.9 and setting the weight decay set to 5×10^{-4}. The learning rate follows the annealing strategy as in Ganin et al. (2016):

$$\mu_p = \mu_0(1 + \alpha \mathring{u} p)^{-\beta},$$

where p is the training progress linearly changing from 0 to 1, $\mu_0 = 0.01$ and $\alpha = 10$ and $\beta = 0.75$.

However, inspecting the Official code repo for each PDA method, the actual learning schedule is given by

$$\mu_i = \mu_0(1 + \alpha \mathring{u} i)^{-\beta},$$

where i is the iteration number in the training procedure, $\mu_0 = 0.01$ and $\alpha = 0.001$ and $\beta = 0.75$. Only when the total number of iterations is 10000 do the learning rate schedules match. In this work, we followed the latter since it is the one indeed used. For OFFICE-HOME, all methods are trained for 5000 iterations, while for VISDA they are trained for 10000 iterations, with the exception of the S. ONLY which is trained for 1000 iterations on OFFICE-HOME and 5000 iterations on VISDA.

A.4 Hyper-Parameters

In Table 2, we report the values used for each hyper-parameter in our grid search. We report in Table 3 the hyper-parameters chosen by each model selection strategy for each method on both datasets. In addition, for the reproducibility of AR with the proposed architecture in Gu et al. (2021),a feature normalization layer is added in the bottleneck which requires specifying r, the value to which the 2-norm is set. This hyper-parameter is therefore included in the hyper-parameter grid search with the possible values of $\{5, 10, 20\}$ which are the different values used in the experiments in (Gu et al., 2021).

B Additional Discussion of Results

In this section, we provide additional results that we could not add to the main paper due to the space constraints.

In Table 4, we show the accuracy per task on OFFICE-HOME averaged over three different seeds (2020, 2021, 2022) for all pairs of methods and model selection strategies.

In Table 5, we compare previously reported results with ours on VISDA. While proposed methods reported results on OFFICE-HOME, only PADA and AR results are reported in the original papers for VISDA. Gu et al. (2021) AR) also report results for BA3US. Analysing the results, we see a 9 percentage point decrease in average task accuracy for PADA, but our experiments show that there is a significant seed dependence which we discuss in detail below. This is particularly important since Cao et al. (2018) (PADA) report results from a single run. Comparing our best seeds for PADA on the SR and RS tasks, we achieve 58.01% and 67.9% accuracy versus a reported 53.53% and 76.5%. Moreover, we point out that the official code repository for PADA does not include the details to reproduce the VISDA experiments, so it is possible that minor tweaks (e.g learning rate) are necessary. As for BA3US, our results are within the standard deviation being better on the SR task and worse on the RS task. Finally as for AR we see a decrease in performance which, as the results on OFFICE-HOME show, can be explained by the differences in the neural network architecture.

Finally in Table 6, we show all the average task accuracies from all pairs of methods and model selection strategies on the OFFICE-HOME and VISDA datasets including the 50-RND model selection strategy.

Method	HP	Values
PADA	λ	$[0.1, 0.5, 1.0, 5.0, 10.0]$
BA3US	λ_{wce}	$[0.1, 0.5, 1, 5, 10]$
	λ_{ent}	$[0.01, 0.05, 0.1, 0.5, 1]$
SAFN	λ	$[0.005, 0.01, 0.05, 0.1, 0.5]$
	Δ_r	$[0.01, 0.1, 1.0]$
AR	ρ_0	$[2.5, 5.0, 7.5, 10.0]$
	A_{up}	$[5.0, 10.0]$
	A_{low}	$-A_{up}$
	λ_{ent}	$[0.01, 0.1, 1.0]$
JUMBOT	τ	$[0.001, 0.01, 0.1]$
	η_1	$[0.00001, 0.0001, 0.001, 0.01, 0.1]$
	η_2	$[0.1, 0.5, 1.]$
	η_3	$[5, 10, 20]$
MPOT	ϵ	$[0.5, 1.0, 1.5]$
	η_1	$[0.0001, 0.001, 0.01, 0.1, 1.0]$
	η_2	$[0.1, 1.0, 5.0, 10.0]$
	m	$[0.1, 0.2, 0.3, 0.4]$

Table 2: Hyper-Parameter values for each PDA method considered in the grid search.

References

Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 135–150, 2018.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL `http://jmlr.org/papers/v22/20-451.html`.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

Xiang Gu, Xi Yu, Yan Yang, Jian Sun, and Zongben Xu. Adversarial reweighting for partial domain adaptation. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. URL `https://openreview.net/forum?id=f5liPryFRoA`.

Kuniaki Saito, Donghyun Kim, Piotr Teterwak, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Tune it the right way: Unsupervised validation of domain adaptation via soft neighborhood density. *arXiv preprint arXiv:2108.10860*, 2021.

Kaichao You, Ximei Wang, Mingsheng Long, and Michael Jordan. Towards accurate model selection in deep unsupervised domain adaptation. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7124–7133. PMLR, 09–15 Jun 2019. URL `https://proceedings.mlr.press/v97/you19a.html`.

Method	Dataset	HP	ORACLE	1-SHOT	50-RND	100-RND	S-ACC	ENT	DEV	SND
PADA	OFFICE-HOME	λ	0.5	0.1	0.1	0.5	0.1	1.0	5.0	0.5
	VISDA	λ	0.5	1.0	10.0	0.5	1.0	0.5	5.0	0.1
SAFN	OFFICE-HOME	λ	0.005	0.1	0.005	0.01	0.005	0.01	0.005	0.005
		Δr	0.1	0.01	0.01	0.01	0.01	0.1	0.1	0.1
	VISDA	λ	0.005	0.005	0.05	0.05	0.005	0.05	0.005	0.05
		Δr	0.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01
BA3US	OFFICE-HOME	λ_{wce}	5.0	10.0	5.0	5.0	5.0	0.1	10.0	1.0
		λ_{ent}	0.05	0.05	0.01	0.05	0.01	0.1	0.05	0.01
	VISDA	λ_{wce}	1.0	1.0	0.1	1.0	5.0	1.0	5.0	5.0
		λ_{ent}	0.5	0.5	0.5	0.5	0.05	0.5	0.05	1.0
AR	OFFICE-HOME	ρ_0	2.5	2.5	5.0	5.0	2.5	5.0	7.5	10.0
		A_{up}	5.0	5.0	10.0	5.0	5.0	10.0	10.0	10.0
		A_{low}	-5.0	-5.0	-10.0	-5.0	-5.0	-10.0	-10.0	-10.0
		λ_{ent}	0.1	0.1	1.0	1.0	0.01	1.0	0.01	1.0
	VISDA	ρ_0	2.5	2.5	2.5	2.5	2.5	7.5	2.5	10.0
		A_{up}	10.0	10.0	10.0	10.0	5.0	10.0	10.0	10.0
		A_{low}	-10.0	-10.0	-10.0	-10.0	-5.0	-10.0	-10.0	-10.0
		λ_{ent}	0.1	0.1	0.1	0.1	0.01	0.1	0.01	0.01
JUMBOT	OFFICE-HOME	τ	0.01	0.01	0.01	0.001	0.1	0.01	0.01	0.001
		η_1	0.0001	0.0001	0.001	0.0001	0.01	1e-05	0.01	1e-05
		η_2	0.5	1.0	0.5	0.1	0.1	0.5	1.0	1.0
		η_3	10.0	5.0	5.0	5.0	5.0	20.0	10.0	5.0
	VISDA	τ	0.01	0.01	0.01	0.01	0.001	0.01	0.001	0.01
		η_1	0.001	0.001	0.001	0.001	0.01	1e-05	0.01	0.0001
		η_2	1.0	1.0	0.5	1.0	0.1	0.5	1.0	1.0
		η_3	5.0	5.0	5.0	5.0	10.0	5.0	20.0	5.0
MPOT	OFFICE-HOME	ϵ	0.5	0.5	1.0	0.5	1.0	1.5	1.0	1.5
		η_1	0.01	0.01	0.01	0.01	0.001	0.0001	1.0	0.01
		η_2	10.0	1.0	1.0	1.0	1.0	10.0	0.1	1.0
		m	0.3	0.1	0.1	0.2	0.3	0.4	0.2	0.4
	VISDA	ϵ	0.5	0.5	0.5	0.5	1.0	1.0	1.0	0.5
		η_1	0.01	0.001	0.01	0.01	0.001	0.0001	0.0001	0.01
		η_2	1.0	1.0	1.0	1.0	1.0	10.0	1.0	10.0
		m	0.3	0.1	0.3	0.3	0.2	0.4	0.2	0.3

Table 3: Hyper-parameters selected for the different methods for each model selection strategy on both OFFICE-HOME and VISDA.

Task	Method	s-acc	ent	dev	snd	1-shot	100-rnd	oracle
AC	s. only	44.50 ± 1.7	45.27 ± 1.1	43.74 ± 1.8	42.23 ± 1.3	43.84 ± 1.7	43.28 ± 1.6	45.43 ± 0.9
	PADA	50.15 ± 2.8	46.03 ± 2.9	44.70 ± 1.3	50.43 ± 0.8	52.98 ± 0.2	50.41 ± 0.8	50.53 ± 0.7
	SAFN	47.36 ± 0.1	47.08 ± 2.0	48.12 ± 0.4	49.57 ± 0.3	31.40 ± 3.7	47.58 ± 0.8	49.57 ± 0.3
	BA3US	54.89 ± 4.7	59.26 ± 0.9	41.67 ± 18.9	62.21 ± 0.9	44.60 ± 21.0	62.53 ± 2.0	63.26 ± 1.0
	AR	51.12 ± 1.2	54.91 ± 1.8	49.25 ± 2.8	54.37 ± 1.6	56.00 ± 2.3	54.89 ± 2.0	57.33 ± 1.7
	JUMBOT	49.07 ± 0.2	57.69 ± 5.6	46.11 ± 0.1	56.60 ± 2.8	61.59 ± 1.7	61.07 ± 0.9	61.87 ± 1.4
	MPOT	53.07 ± 0.3	52.94 ± 2.0	46.07 ± 0.7	32.96 ± 0.4	53.97 ± 1.3	61.59 ± 1.2	64.48 ± 1.2
AP	s. only	67.71 ± 2.4	68.91 ± 1.4	67.81 ± 1.2	68.91 ± 1.4	66.52 ± 3.1	68.76 ± 1.6	68.91 ± 1.4
	PADA	66.93 ± 1.2	62.09 ± 2.8	61.61 ± 5.4	66.72 ± 1.5	63.03 ± 1.6	67.21 ± 1.8	67.45 ± 1.6
	SAFN	66.82 ± 1.9	66.83 ± 0.5	67.30 ± 0.5	68.18 ± 1.3	49.73 ± 4.3	67.53 ± 0.8	68.55 ± 1.0
	BA3US	71.34 ± 0.8	76.38 ± 1.5	50.05 ± 28.7	83.29 ± 0.4	51.39 ± 29.8	82.09 ± 0.8	82.75 ± 0.9
	AR	72.79 ± 0.7	78.45 ± 1.8	70.20 ± 1.7	79.01 ± 2.2	78.58 ± 1.9	78.54 ± 1.4	79.61 ± 1.6
	JUMBOT	65.45 ± 0.4	75.44 ± 1.4	66.33 ± 0.6	68.48 ± 1.5	76.86 ± 3.4	77.87 ± 1.4	78.19 ± 2.4
	MPOT	72.61 ± 1.2	68.94 ± 1.2	65.43 ± 0.8	49.73 ± 1.1	68.78 ± 1.7	75.56 ± 1.7	80.88 ± 3.3
AR	s. only	78.37 ± 0.3	79.26 ± 0.7	78.28 ± 0.7	79.35 ± 0.6	77.38 ± 0.9	77.97 ± 1.2	79.53 ± 0.3
	PADA	76.73 ± 1.7	76.05 ± 1.4	68.99 ± 11.3	79.72 ± 1.8	78.06 ± 2.6	79.97 ± 1.5	80.14 ± 1.4
	SAFN	77.62 ± 0.2	77.73 ± 0.2	77.43 ± 0.5	77.86 ± 0.5	62.82 ± 2.0	77.91 ± 0.4	78.26 ± 0.2
	BA3US	81.91 ± 3.9	86.03 ± 0.6	63.74 ± 26.1	88.50 ± 0.6	65.47 ± 27.2	88.28 ± 0.4	89.16 ± 0.2
	AR	77.91 ± 0.2	84.23 ± 0.9	79.73 ± 2.5	84.54 ± 0.8	82.77 ± 2.0	84.34 ± 0.6	86.31 ± 0.4
	JUMBOT	77.14 ± 0.3	85.24 ± 2.7	76.42 ± 0.3	84.70 ± 2.1	86.45 ± 2.1	86.01 ± 1.3	88.11 ± 1.5
	MPOT	78.50 ± 0.7	75.98 ± 0.6	76.46 ± 0.4	57.39 ± 1.4	78.04 ± 2.1	82.59 ± 0.6	86.78 ± 0.5
CA	s. only	52.56 ± 0.9	54.21 ± 2.1	51.42 ± 2.7	51.76 ± 3.7	50.47 ± 2.4	53.75 ± 1.1	55.59 ± 0.7
	PADA	58.00 ± 1.4	55.07 ± 2.7	35.08 ± 13.1	57.30 ± 1.9	51.67 ± 5.0	56.69 ± 1.5	57.30 ± 1.9
	SAFN	57.85 ± 0.6	56.54 ± 2.2	56.75 ± 0.3	57.91 ± 0.3	48.88 ± 2.4	56.47 ± 1.0	57.91 ± 0.3
	BA3US	61.68 ± 5.2	68.96 ± 1.8	60.70 ± 2.2	68.50 ± 0.9	65.63 ± 1.4	69.15 ± 1.2	69.91 ± 0.2
	AR	63.21 ± 1.5	64.86 ± 2.3	62.72 ± 1.0	64.52 ± 1.6	68.99 ± 0.2	64.95 ± 2.4	69.45 ± 0.5
	JUMBOT	60.09 ± 0.1	75.97 ± 1.4	56.81 ± 0.1	71.81 ± 1.8	74.20 ± 0.9	74.56 ± 0.4	77.69 ± 0.1
	MPOT	61.92 ± 0.5	60.58 ± 0.8	56.44 ± 1.0	44.11 ± 2.4	69.24 ± 0.4	72.48 ± 1.0	76.22 ± 0.1
CP	s. only	54.81 ± 0.1	55.52 ± 0.6	54.55 ± 1.2	53.48 ± 2.1	53.24 ± 2.0	55.57 ± 2.2	57.42 ± 1.2
	PADA	56.13 ± 1.4	47.28 ± 0.1	24.24 ± 20.5	52.10 ± 1.7	56.28 ± 0.4	53.86 ± 1.6	54.47 ± 1.7
	SAFN	57.89 ± 0.7	59.07 ± 0.7	58.17 ± 1.2	58.17 ± 1.2	45.27 ± 0.7	58.19 ± 0.4	59.29 ± 0.5
	BA3US	67.13 ± 3.9	71.07 ± 0.8	59.08 ± 10.9	71.45 ± 3.6	59.78 ± 15.3	71.65 ± 1.5	71.93 ± 1.6
	AR	60.54 ± 4.0	68.16 ± 3.5	61.85 ± 4.6	68.05 ± 3.2	68.35 ± 1.9	69.00 ± 3.7	71.88 ± 0.9
	JUMBOT	59.59 ± 1.3	74.85 ± 3.3	56.36 ± 0.5	71.84 ± 1.6	73.43 ± 3.3	76.40 ± 1.4	76.75 ± 0.8
	MPOT	64.16 ± 1.8	65.99 ± 2.2	57.95 ± 1.0	38.66 ± 1.2	65.88 ± 0.5	69.77 ± 0.9	77.95 ± 1.3
CR	s. only	62.88 ± 0.9	63.19 ± 0.3	63.94 ± 1.7	63.94 ± 1.7	61.77 ± 1.1	63.94 ± 0.4	65.23 ± 0.8
	PADA	66.45 ± 0.8	60.92 ± 2.4	61.66 ± 2.4	63.11 ± 1.9	64.00 ± 1.4	63.94 ± 1.3	64.55 ± 1.1
	SAFN	66.92 ± 0.9	66.22 ± 0.5	65.64 ± 1.3	66.13 ± 1.0	57.26 ± 2.2	65.88 ± 0.2	66.81 ± 0.5
	BA3US	72.96 ± 1.0	76.22 ± 1.2	67.88 ± 0.9	76.96 ± 0.6	68.49 ± 1.3	77.21 ± 0.6	77.58 ± 0.9
	AR	72.76 ± 0.9	80.45 ± 0.8	70.86 ± 5.6	79.16 ± 2.8	77.25 ± 1.4	79.57 ± 0.2	79.94 ± 0.8
	JUMBOT	66.67 ± 1.3	79.75 ± 1.2	66.70 ± 0.8	80.91 ± 0.9	79.85 ± 0.3	81.54 ± 1.7	84.15 ± 1.3
	MPOT	70.22 ± 0.2	71.51 ± 0.8	66.35 ± 1.0	50.06 ± 1.0	71.42 ± 0.7	75.41 ± 0.5	82.59 ± 0.7
PA	s. only	58.77 ± 0.5	56.96 ± 1.5	57.94 ± 0.9	55.37 ± 0.6	56.11 ± 1.7	58.37 ± 0.4	59.32 ± 0.7
	PADA	60.33 ± 2.1	56.69 ± 2.8	57.91 ± 1.7	60.82 ± 3.0	58.92 ± 3.3	60.27 ± 2.7	61.07 ± 3.0
	SAFN	58.80 ± 0.7	56.75 ± 2.1	59.08 ± 0.5	59.14 ± 0.8	42.33 ± 1.6	59.69 ± 0.1	59.87 ± 0.7
	BA3US	68.90 ± 5.0	73.16 ± 0.6	64.62 ± 1.6	76.19 ± 1.2	68.38 ± 1.7	75.15 ± 1.3	75.73 ± 1.3
	AR	63.39 ± 3.1	67.58 ± 0.4	61.65 ± 1.0	65.60 ± 1.7	69.67 ± 1.5	66.73 ± 0.3	70.28 ± 1.0
	JUMBOT	60.24 ± 1.0	72.85 ± 2.4	58.03 ± 1.1	70.28 ± 0.8	74.96 ± 3.4	72.60 ± 1.2	76.83 ± 1.9
	MPOT	64.13 ± 0.9	58.28 ± 0.9	57.64 ± 0.8	43.74 ± 4.3	70.31 ± 1.0	72.64 ± 0.9	75.18 ± 1.3
PC	s. only	39.28 ± 0.8	38.75 ± 0.6	39.40 ± 0.9	37.35 ± 1.0	37.35 ± 1.0	39.12 ± 0.4	40.80 ± 0.9
	PADA	43.50 ± 1.2	38.43 ± 3.0	38.03 ± 0.6	39.26 ± 2.0	43.62 ± 1.0	40.56 ± 1.8	40.94 ± 1.6
	SAFN	42.49 ± 0.6	39.58 ± 2.0	43.00 ± 1.1	43.90 ± 0.5	29.77 ± 2.6	43.14 ± 1.7	45.29 ± 0.7
	BA3US	55.92 ± 1.3	57.91 ± 2.5	56.74 ± 1.3	59.94 ± 1.7	57.83 ± 1.3	58.17 ± 1.0	59.94 ± 0.7
	AR	48.36 ± 1.7	52.34 ± 1.0	43.72 ± 0.7	51.28 ± 1.6	51.98 ± 1.8	50.85 ± 1.4	53.57 ± 0.2
	JUMBOT	43.60 ± 0.0	60.18 ± 0.9	41.99 ± 0.8	50.69 ± 4.9	62.87 ± 0.6	59.92 ± 0.4	63.72 ± 0.5
	MPOT	50.87 ± 1.1	49.87 ± 2.6	43.60 ± 0.6	28.66 ± 2.6	53.03 ± 0.7	57.67 ± 1.6	64.60 ± 0.0
PR	s. only	75.08 ± 0.5	75.65 ± 1.3	74.91 ± 0.6	74.10 ± 2.8	71.97 ± 1.8	75.56 ± 1.3	75.80 ± 1.2
	PADA	76.70 ± 0.4	77.08 ± 0.2	73.11 ± 3.4	79.33 ± 1.3	74.27 ± 4.1	78.91 ± 1.8	79.55 ± 1.4
	SAFN	75.46 ± 0.4	73.90 ± 0.9	74.64 ± 0.4	75.81 ± 0.7	63.52 ± 3.2	75.00 ± 0.7	75.98 ± 0.6
	BA3US	79.13 ± 4.7	85.59 ± 1.2	75.21 ± 0.6	86.31 ± 1.4	82.05 ± 1.0	85.92 ± 1.3	86.89 ± 0.5
	AR	78.02 ± 1.7	82.48 ± 1.9	76.29 ± 0.7	83.05 ± 1.1	78.72 ± 1.0	82.39 ± 1.9	83.78 ± 1.0
	JUMBOT	74.43 ± 0.9	83.21 ± 1.1	74.97 ± 0.5	83.89 ± 1.5	81.83 ± 0.9	84.63 ± 2.3	84.80 ± 1.3
	MPOT	77.40 ± 0.1	73.77 ± 1.3	74.86 ± 1.3	58.40 ± 1.9	76.88 ± 1.3	82.02 ± 0.6	84.87 ± 1.4
RA	s. only	68.90 ± 0.6	69.24 ± 1.0	69.27 ± 1.0	68.53 ± 1.4	68.96 ± 0.5	69.02 ± 0.5	69.88 ± 0.9
	PADA	69.27 ± 3.5	69.48 ± 1.3	66.33 ± 0.7	73.09 ± 1.5	68.26 ± 3.1	72.70 ± 1.4	73.09 ± 1.5
	SAFN	67.92 ± 0.0	67.80 ± 0.2	68.11 ± 0.9	68.17 ± 1.6	56.11 ± 3.2	69.64 ± 1.0	69.08 ± 0.6
	BA3US	72.27 ± 3.5	78.11 ± 1.4	70.92 ± 2.0	79.46 ± 1.4	80.78 ± 1.1	79.86 ± 2.1	80.93 ± 0.8
	AR	70.00 ± 1.1	74.75 ± 2.1	70.31 ± 1.7	75.02 ± 1.6	76.19 ± 0.7	74.66 ± 2.3	77.26 ± 0.6
	JUMBOT	70.19 ± 0.5	81.97 ± 1.0	67.43 ± 0.3	81.21 ± 0.6	78.48 ± 2.0	81.85 ± 1.7	81.79 ± 0.8
	MPOT	70.40 ± 0.6	64.98 ± 0.4	67.68 ± 0.5	56.90 ± 1.9	76.52 ± 0.4	79.80 ± 0.5	80.59 ± 0.6
RC	s. only	45.33 ± 1.0	45.31 ± 1.0	45.33 ± 1.0	43.78 ± 0.6	46.13 ± 2.0	43.46 ± 0.2	47.20 ± 0.9
	PADA	53.93 ± 1.3	49.73 ± 3.5	29.97 ± 21.0	45.77 ± 1.6	54.25 ± 1.6	53.39 ± 2.2	54.63 ± 0.9
	SAFN	49.73 ± 0.1	48.76 ± 0.1	50.53 ± 0.6	49.59 ± 1.6	37.55 ± 0.8	50.85 ± 0.3	51.68 ± 0.8
	BA3US	51.84 ± 0.5	62.85 ± 2.7	58.39 ± 2.3	65.35 ± 1.9	63.10 ± 0.8	66.57 ± 1.5	66.77 ± 1.5
	AR	52.52 ± 1.0	55.64 ± 1.2	49.61 ± 0.8	55.02 ± 1.8	55.48 ± 2.1	55.42 ± 1.6	59.68 ± 1.1
	JUMBOT	51.12 ± 1.1	61.81 ± 4.6	48.12 ± 0.5	58.85 ± 1.7	61.59 ± 2.2	64.84 ± 1.0	64.70 ± 1.1
	MPOT	53.99 ± 1.5	57.53 ± 0.6	48.12 ± 0.8	39.34 ± 1.2	57.39 ± 1.7	64.64 ± 0.1	67.04 ± 0.6
RP	s. only	76.34 ± 0.7	76.47 ± 0.8	75.99 ± 1.3	75.84 ± 1.6	73.33 ± 2.2	75.28 ± 2.3	77.31 ± 0.1
	PADA	78.88 ± 0.8	78.00 ± 1.7	71.07 ± 11.3	80.62 ± 0.4	78.62 ± 0.4	80.73 ± 0.9	80.93 ± 0.6
	SAFN	76.23 ± 0.8	76.23 ± 0.7	75.65 ± 0.5	76.64 ± 0.5	67.00 ± 1.4	76.41 ± 0.8	77.29 ± 0.5
	BA3US	81.85 ± 4.1	84.84 ± 0.6	78.06 ± 1.3	86.35 ± 0.9	79.20 ± 1.1	85.66 ± 1.0	86.93 ± 0.2
	AR	77.55 ± 2.6	83.06 ± 1.2	75.61 ± 0.4	83.40 ± 0.9	82.73 ± 1.0	82.80 ± 0.4	83.72 ± 0.6
	JUMBOT	77.12 ± 1.3	86.33 ± 1.6	76.04 ± 0.1	88.18 ± 0.4	87.34 ± 0.2	87.64 ± 0.7	87.17 ± 1.7
	MPOT	77.61 ± 0.3	73.17 ± 2.7	75.89 ± 0.4	63.14 ± 0.6	77.95 ± 1.4	82.60 ± 0.5	86.52 ± 1.2
Avg	s. only	60.38 ± 0.5	60.73 ± 0.2	60.22 ± 0.3	59.55 ± 0.3	58.92 ± 0.4	60.34 ± 0.4	61.87 ± 0.3
	PADA	63.08 ± 0.3	59.74 ± 0.5	52.72 ± 2.8	62.36 ± 0.4	62.00 ± 0.5	63.22 ± 0.1	63.72 ± 0.3
	SAFN	62.09 ± 0.2	61.37 ± 0.3	62.03 ± 0.4	62.59 ± 0.1	49.30 ± 0.7	62.36 ± 0.2	63.30 ± 0.2
	BA3US	68.32 ± 1.1	73.36 ± 0.6	62.25 ± 7.1	75.37 ± 0.8	65.56 ± 7.6	75.19 ± 0.4	75.98 ± 0.3
	AR	65.68 ± 0.3	70.58 ± 0.4	64.32 ± 0.9	70.25 ± 0.2	70.56 ± 0.7	70.34 ± 0.2	72.73 ± 0.3
	JUMBOT	62.89 ± 0.2	74.61 ± 0.8	61.28 ± 0.1	72.29 ± 0.2	74.95 ± 0.1	75.74 ± 0.3	77.15 ± 0.4
	MPOT	66.24 ± 0.1	64.46 ± 0.1	61.37 ± 0.2	46.92 ± 0.4	68.28 ± 0.2	73.06 ± 0.3	77.31 ± 0.5

Table 4: Average accuracy of different PDA methods based on different model selection strategies on the 12 tasks of Partial OFFICE-HOME. Average is done over three seeds (2020, 2021, 2022). For each method, we highlight the best and worst label-free model selection strategies in green and red, respectively.

Algorithm	S2R	R2S	Avg
S. ONLY†	45.26	64.28	54.77
S. ONLY (Ours)	51.86	67.11	59.48
PADA†	53.53	76.50	65.02
PADA (Ours)	49.34	59.81	54.57
SAFN†	67.65	-	-
SAFN (Ours)	56.88	68.40	62.64
BA3US†	69.86	67.56	68.71
BA3US (Ours)	71.77	63.56	67.67
AR†*	85.30	74.82	80.06
AR (Ours)	76.33	71.36	73.85
JUMBOT†	-	-	-
JUMBOT (Ours)	90.55	77.46	84.01
MPOT†	-	-	-
MPOT (Ours)	87.23	86.67	86.95

Table 5: Comparison between reported (†) accuracies on partial VISDA from published methods with our implementation using the ORACLE model selection strategy. * denotes different bottleneck architectures.

Dataset	Method	S-ACC	ENT	DEV	SND	1-SHOT	50-RND	100-RND	ORACLE
OFFICE-HOME	S. ONLY	60.38±0.5	60.73±0.2	60.22±0.3	59.55±0.3	58.92±0.4	60.28±0.4	60.34±0.4	61.87±0.3
	PADA	63.08±0.3	59.74±0.5	52.72±2.8	62.36±0.4	62.00±0.5	63.82±0.4	63.22±0.1	63.72±0.3
	SAFN	62.09±0.2	61.37±0.3	62.03±0.4	62.59±0.1	49.30±0.7	62.00±0.2	62.36±0.2	63.30±0.2
	BA3US	68.32±1.1	73.36±0.6	62.25±7.1	75.37±0.8	65.56±7.6	73.22±0.3	75.19±0.4	75.98±0.3
	AR	65.68±0.3	70.58±0.4	64.32±0.9	70.25±0.2	70.56±0.7	70.26±0.2	70.34±0.2	72.73±0.3
	JUMBOT	62.89±0.2	74.61±0.8	61.28±0.1	72.29±0.2	74.95±0.1	64.95±0.3	75.74±0.3	77.15±0.4
	MPOT	66.24±0.1	64.46±0.1	61.37±0.2	46.92±0.4	68.28±0.2	69.90±0.5	73.06±0.3	77.31±0.5
VISDA	S. ONLY	55.15±2.4	55.24±3.2	55.07±1.2	55.02±2.9	55.72±2.2	57.90±1.1	58.16±0.6	59.48±0.4
	PADA	47.48±4.8	32.32±4.9	43.43±5.3	56.83±1.0	53.15±2.9	55.67±2.5	54.38±2.7	54.57±2.6
	SAFN	58.20±1.7	42.83±6.3	58.62±1.3	44.82±8.8	56.89±2.1	57.90±3.3	59.09±2.8	62.64±1.5
	BA3US	55.10±3.7	65.58±1.4	58.40±1.4	51.07±4.3	64.77±1.4	66.66±2.4	67.44±1.2	67.67±1.3
	AR	66.68±1.0	64.27±3.6	67.20±1.5	55.69±0.9	70.29±1.7	71.91±0.3	72.60±0.8	73.85±0.9
	JUMBOT	60.63±0.7	62.42±2.4	59.86±0.6	77.69±4.2	78.34±1.9	82.85±2.9	83.49±1.9	84.01±1.9
	MPOT	70.02±2.0	74.64±4.4	61.62±1.3	78.40±3.9	70.96±3.7	86.65±5.1	86.69±5.1	86.95±5.0

Table 6: Task accuracy average over seeds 2020, 2021, 2022 on Partial OFFICE-HOME and Partial VISDA for the PDA methods and model selection strategies. For each method, we highlight the best and worst label-free model selection strategies in green and red, respectively.