An Empirical Study of Domain Adaptation: Are We Really Learning Transferable Representations?

I. Office-31 Dataset

We provide some additional images in 1 from the Office-31 dataset [1] to show a few more of the 31 classes present in the 3 domains: Amazon, DSLR, and Webcam.

Figure 1: Images from the Office-31 dataset. Each row is a domain, 3 sample classes (backpack, scissors, trash can) are shown.

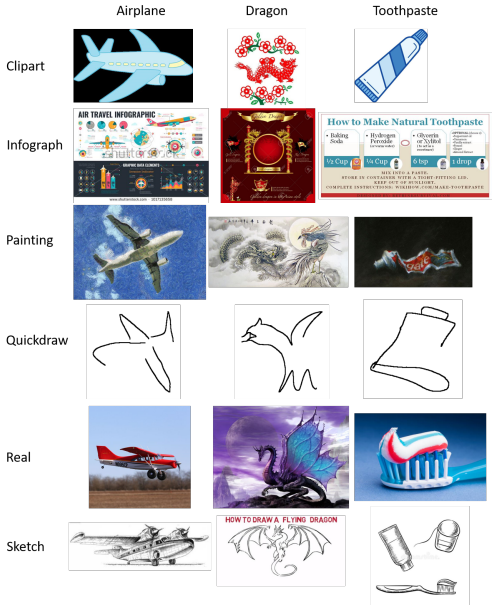


Figure 2: Images from the DomainNet dataset. Each row is a domain, 3 sample classes (airplane, dragon, toothpaste) are shown.

II. DomainNet Dataset

A. DomainNet Examples

We provide some additional images in Figure 2 from the DomainNet dataset [2] to show a few more of the 345 classes present in the 6 domains: clipart, infograph, painting, quickdraw, real, and sketch.

B. Dataset Preparation

There were a few inconsistencies identified with the original split of the DomainNet dataset as originally defined in [2] and the VisDA-2019 competition that had to be rectified for the purpose of the experiments to be conducted. We had to insure that there existed images in each class, something we found to be a problem with the original split released, and had to insure that we had enough images of each class present for each of the data subsets and train/validation/test sets we were creating for our experimental design.

The inconsistencies specifically occurred in the painting domain. In the painting domain *test* set, no images were assigned to the class "syringe" when images were available in the dataset to be assigned there. Therefore, we assigned the 4 available images to the painting domain syringe class in the test set. In the painting domain *training* set, no images were assigned to the class "t-shirt" when images were available in the dataset to be assigned there. Therefore, we assigned the 10 available images not already assigned to the test set for "t-shirt" class here.

Additionally, in order to run 5-fold cross validation, a 6 image per class minimum was enforced for all 3 subsets of the DomainNet dataset such that when creating the cross-validated folds enough images would be available for splitting. This required adding images previously not used in the defined 70-30 split (yet available in the dataset) to classes that did not have 6 images assigned per class. We added to the original 70% of training data in the painting domain 4 images to the "stereo" class, 3 images to the "underwear" class, 4 images to the "waterslide" class, 10 images to the "toe" class, and an additional 6 images were moved from the test set to the training set for the "syringe" class.

III. Experiment Protocol

We provide an additional step-by-step procedure for how experiments were conducted to fairly compare models and then present the optimal hyperparameters found for each model on the DomainNet and Offcie-31 datasets.

A. Step-by-Step Procedure

1) Domain adaptation model is chosen (e.x. JAN)
2) Data subset is chosen (full, subset-50, subset-20, or Office-31)
3) Source domain is chosen (clipart, infograph, painting, quickdraw, real, sketch, amazon, dslr, or webcam)
4) Target domain is chosen, different than the source domain (clipart, infograph, painting, quickdraw, real, sketch, amazon, dslr, or webcam)
5) 5-fold cross validation is run
6) Choose new target domain
7) Repeat steps 5 and 6 until all domains are used as target
8) Choose new source domain
9) Repeat steps 4-8 until all domains are used as source
10) Choose new data subset
11) Repeat steps 3-10 until all data subsets chosen
12) Choose new domain adaptation model
13) Repeat steps 2-12 until all domain adaptation models chosen

B. Optimal Hyperparameters

In Tables I, II we show the optimal values for the tuned hyperparameters on the DomainNet and Office-31 datsets, respectively. We report: learning rate, weight decay, loss trade-off, and temperature with a few other hyperparameter values. Pre-trained ResNet-50 models are pre-trained on ImageNet [3]. All other default hyperparameter values will be in the released code.

IV. Supplemental Results

In this section we first show additional tables for all 5 models, for all 3 data subsets, for all 30 adaptation tasks on the DomainNet dataset but with standard deviations over the 5 cross-validation folds for each adaptation task. Then we show additional tables for the Office-31 dataset, for all 6 adaptation tasks but with standard deviations over 5 cross-validation folds. We then include a set of tables with standard deviations over 5 cross-validation folds for the Office-31 dataset but without the tuned hyperparamters on the Office-31 dataset, instead showing the results on Office-31 when using the tuned hyperparameters found on DomainNet; this table is shown to emphasize the sensitivity of the MCC model to hyperparameter tuning. We then show additional classification and transfer loss curves for 4 more adaptation tasks and their corresponding test accuracy curves. Finally, we show additional T-SNE plots for all 5 models for 5 adaptation tasks (clipart as source).

A. Model Comparisons

In Tables III, IV, and V we show 5-fold cross-validated test set accuracies (means and standard deviations) for each of the 30 adaptation tasks on DomainNet for all 5 models for the full, subset-50, and subset-20 datasets, respectively.

In Table VI we show 5-fold cross-validated test set accuracies (means and standard deviations) for each of the 6 adaptation tasks on Office-31 for all 5 models with the tuned hyperparameters on the *Office-31* dataset.

In Table VII we show 5-fold cross-validated test set accuracies (means and standard deviations) for each of the 6 adaptation tasks on Office-31 for all 5 models with the tuned hyperparameters from the *DomainNet* dataset.

B. Transfer Loss Analysis

In this section we provide the classification and transfer loss curves for the remaining 4 adaptation tasks where clipart is the source domain (infograph, quickdraw, real, and sketch as target domains) as well as the test accuracy curves for each of the 4 adaptation tasks. These results are also only for the full dataset and cross-validation fold 1. Refer to Figures 4, 5, 6, 7, 8, 9, 10, and 11 for classification and transfer loss curves. Refer to Figures 12, 13, 14, and 15 for test accuracy curves.

C. Feature space alignment: T-SNE

Here we show in Figure 3 T-SNE plots for all 5 models for 5 adaptation tasks where clipart is always the source domain.

References

[1] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Eur. Conf. on Comput. Vision*. Springer, 2010, pp. 213–226.

[2] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *IEEE/CVF Int. Conf. on Comput. Vision*, 2019, pp. 1406–1415.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. on Comput. vision and pattern recognition*. IEEE, 2009, pp. 248–255.

	DANN	JAN	CDAN	AFN	MCC
Pre-trained (backbone)	ResNet-50	ResNet-50	ResNet-50	ResNet-50	ResNet-50
Batch Size	32	32	32	32	36
Momentum	0.9	0.9	0.9	0.9	0.9
Learning Rate	0.01	0.01	0.01	0.001	0.1
Weight Decay	0.001	0.001	0.003	0.001	0.0001
Loss Trade-off	1.0	1.0	2.0	0.05	5.0
Temperature	-	-	-	-	2.0

Table I: Optimal hyperparamters for each model for the DomainNet dataset.

	DANN	JAN	CDAN	AFN	MCC
Pre-trained (backbone)	ResNet-50	ResNet-50	ResNet-50	ResNet-50	ResNet-50
Batch Size	32	32	32	32	36
Momentum	0.9	0.9	0.9	0.9	0.9
Learning Rate	0.001	0.001	0.001	0.001	0.005
Weight Decay	0.01	0.01	0.001	0.01	0.001
Loss Trade-off	0.5	0.75	1.0	0.05	0.5
Temperature	-	-	-	-	2.0

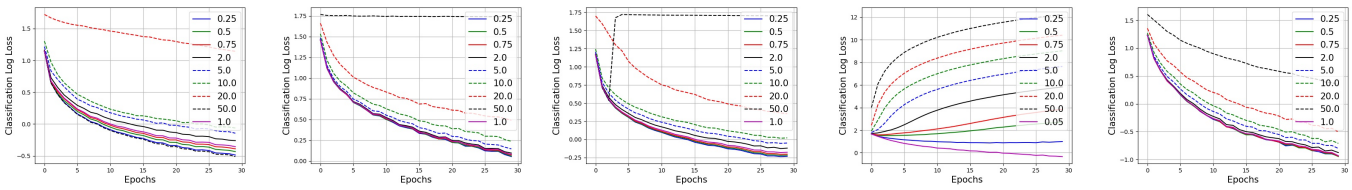
Table II: Optimal hyperparamters for each model for the Office-31 dataset.

DANN	clp	inf	pnt	qdr	rel	skt
clp	-	27.0 ± 0.2	37.3 ± 0.2	18.7 ± 0.2	47.6 ± 0.4	50.8 ± 0.3
inf	19.1 ± 0.2	-	18.6 ± 0.2	3.9 ± 0.2	22.7 ± 0.1	19.3 ± 0.1
pnt	34.3 ± 0.2	25.9 ± 0.2	-	7.9 ± 0.3	47.7 ± 0.2	41.4 ± 0.2
qdr	12.7 ± 0.5	5.5 ± 0.3	5.8 ± 0.3	-	6.5 ± 0.8	12.4 ± 0.2
rel	50.4 ± 0.2	36.9 ± 0.4	51.8 ± 0.2	12.7 ± 0.3	-	50.8 ± 0.3
skt	40.7 ± 0.1	22.6 ± 0.4	33.5 ± 0.5	13.0 ± 0.3	37.4 ± 0.3	-
JAN	clp	inf	pnt	qdr	rel	skt
clp	-	28.1 ± 0.2	38.5 ± 0.5	15.3 ± 0.6	46.6 ± 0.3	48.5 ± 0.3
inf	17.2 ± 0.3	-	17.3 ± 0.2	2.4 ± 0.1	20.6 ± 0.1	16.5 ± 0.3
pnt	33.9 ± 0.3	26.6 ± 0.9	-	6.7 ± 0.3	46.6 ± 0.1	39.2 ± 0.5
qdr	11.4 ± 0.2	4.7 ± 0.2	4.8 ± 0.3	-	4.6 ± 0.2	10.1 ± 0.4
rel	49.5 ± 0.4	39.4 ± 0.4	51.5 ± 0.3	11.7 ± 0.7	-	47.0 ± 0.3
skt	40.2 ± 0.2	24.0 ± 0.1	35.7 ± 0.2	10.9 ± 0.3	36.5 ± 0.5	-
CDAN	clp	inf	pnt	qdr	rel	skt
clp	-	26.6 ± 0.1	37.2 ± 0.2	19.9 ± 0.6	49.3 ± 0.3	51.0 ± 0.1
inf	18.7 ± 0.2	-	18.4 ± 0.3	4.0 ± 0.1	22.7 ± 0.2	18.8 ± 0.1
pnt	35.0 ± 0.2	25.2 ± 0.3	-	7.1 ± 0.2	49.0 ± 0.2	42.0 ± 0.2
qdr	10.1 ± 0.3	4.7 ± 0.4	4.7 ± 0.2	-	4.5 ± 0.5	9.3 ± 0.3
rel	52.6 ± 0.2	37.0 ± 0.3	51.6 ± 0.2	14.4 ± 0.2	-	51.1 ± 0.2
skt	40.9 ± 0.2	22.0 ± 0.4	34.8 ± 0.4	13.8 ± 0.5	38.7 ± 0.4	-
AFN	clp	inf	pnt	qdr	rel	skt
clp	-	29.1 ± 0.5	38.9 ± 0.2	17.3 ± 0.6	43.9 ± 0.5	49.6 ± 0.4
inf	15.8 ± 0.2	-	16.5 ± 0.2	2.8 ± 0.1	17.6 ± 0.2	15.6 ± 0.2
pnt	36.0 ± 0.2	32.5 ± 0.5	-	6.6 ± 0.3	46.6 ± 0.2	42.6 ± 0.4
qdr	11.3 ± 0.2	3.1 ± 0.2	4.7 ± 0.1	-	4.8 ± 0.2	11.1 ± 0.2
rel	51.9 ± 0.3	45.7 ± 0.3	53.9 ± 0.4	11.8 ± 0.4	-	51.9 ± 0.4
skt	41.2 ± 0.4	25.0 ± 0.1	36.4 ± 0.1	9.3 ± 0.3	34.4 ± 0.3	-
MCC	clp	inf	pnt	qdr	rel	skt
clp	-	31.1 ± 0.4	43.0 ± 0.4	11.6 ± 0.6	51.2 ± 0.6	53.6 ± 0.4
inf	13.5 ± 0.4	-	14.8 ± 0.4	1.1 ± 0.1	17.7 ± 0.3	13.8 ± 0.3
pnt	33.4 ± 0.6	28.7 ± 0.1	-	2.6 ± 0.4	46.1 ± 0.2	39.8 ± 0.3
qdr	12.9 ± 0.8	2.7 ± 0.4	3.7 ± 1.0	-	4.0 ± 0.7	11.7 ± 0.5
rel	49.0 ± 0.4	43.7 ± 0.6	52.5 ± 0.3	7.0 ± 0.4	-	49.1 ± 0.5
skt	39.3 ± 0.1	22.7 ± 0.7	36.1 ± 0.5	7.9 ± 0.8	35.1 ± 0.3	-

Table III: Mean and standard deviation target domain test accuracy over 5-cross-val folds for all models for all 30 adaptation tasks on the full dataset. (Columns = source domain, rows = target domain)

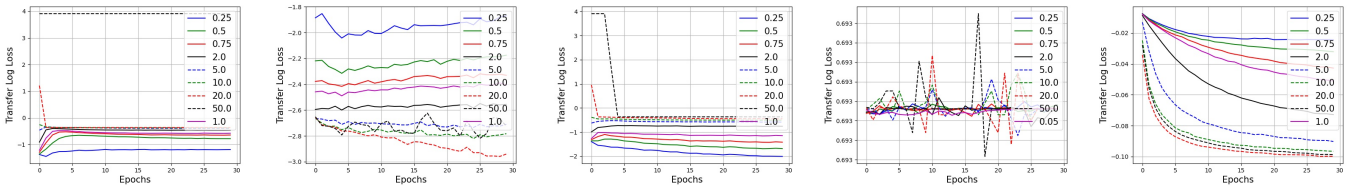
(a) C2I (b) C2P (c) C2Q (d) C2R (e) C2S

Figure 3: T-SNE plots for the airplane class only. Red represents the clipart source domain, blue represents the target domain. a-e represent 5 of the 30 adaptation tasks where clipart is always the source domain. (C=clipart, I=infograph, P=painting, Q=quickdraw, R=real, S=sketch) Each row is a model, top to bottom: DANN, JAN, CDAN, AFN, MCC.



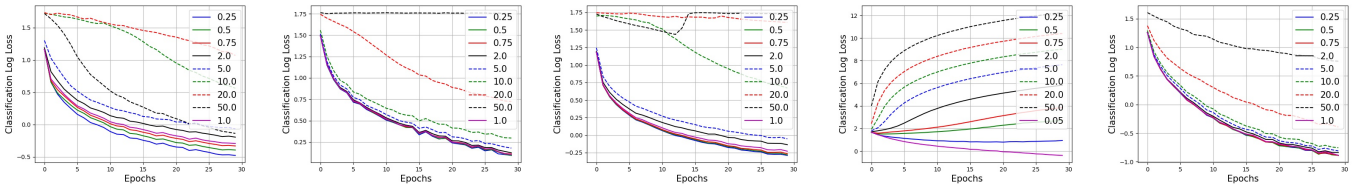
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 4: Classification loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to infograph adaptation task for full dataset.



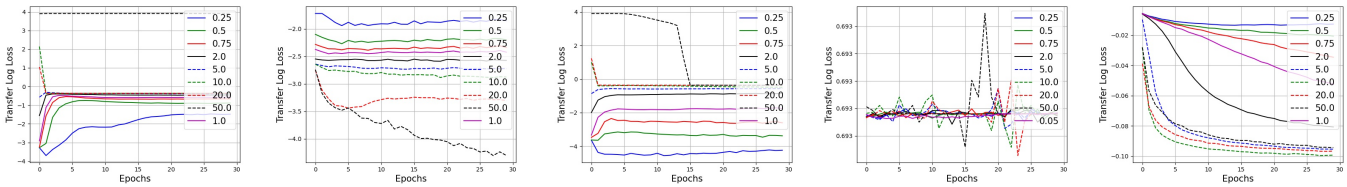
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 5: Transfer loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to infograph adaptation task for full dataset.



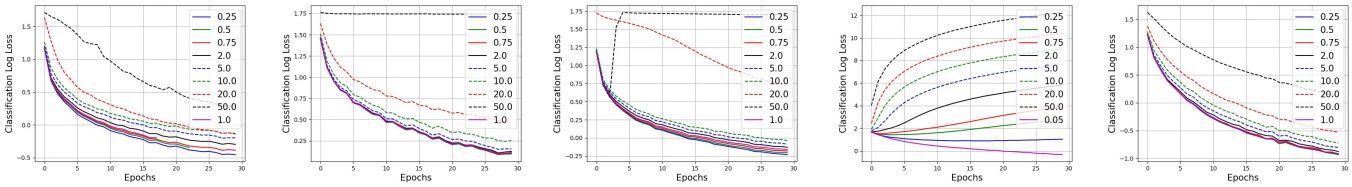
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 6: Classification loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to quickdraw adaptation task for full dataset.



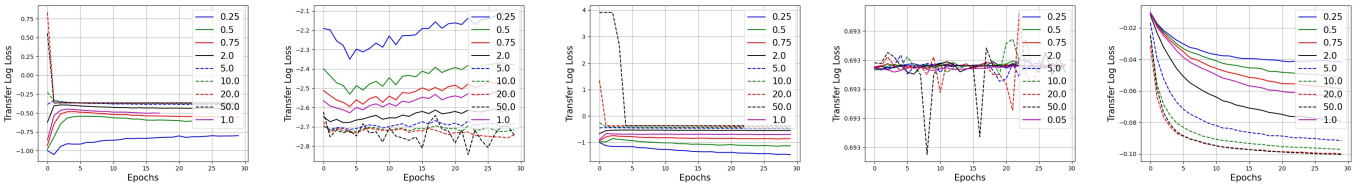
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 7: Transfer loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to quickdraw adaptation task for full dataset.



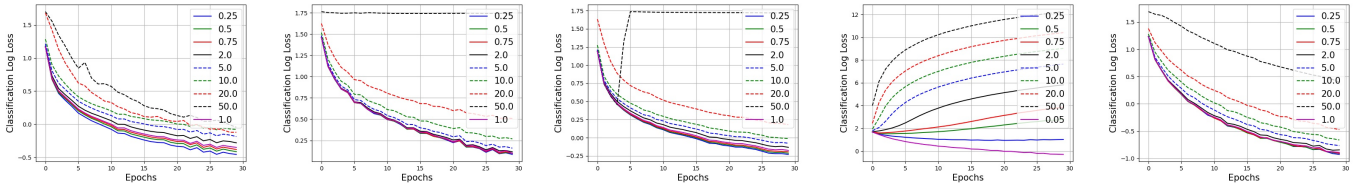
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 8: Classification loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to real adaptation task for full dataset.



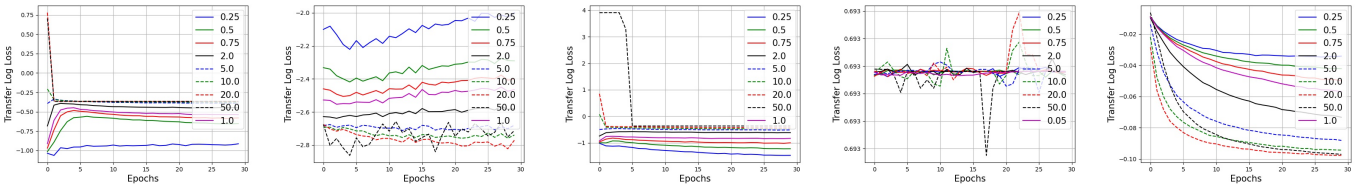
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 9: Transfer loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to real adaptation task for full dataset.



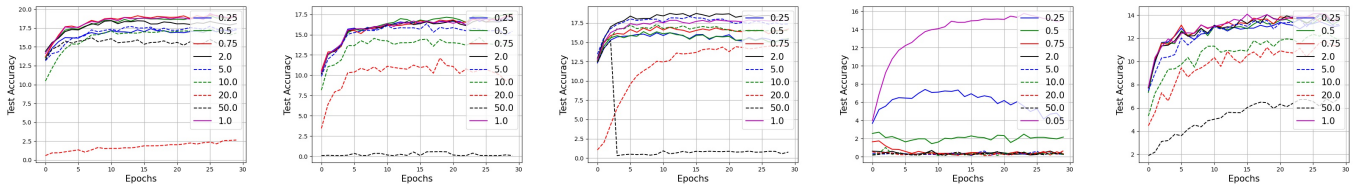
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 10: Classification loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to sketch adaptation task for full dataset.



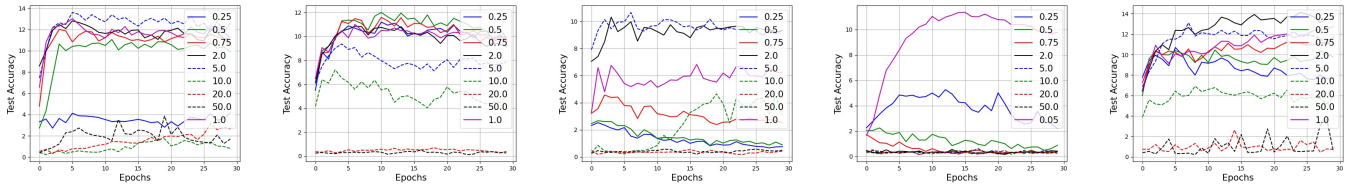
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 11: Transfer loss curves for varied loss trade-off hyperparameters (y-axis log-scale). Clipart to sketch adaptation task for full dataset.



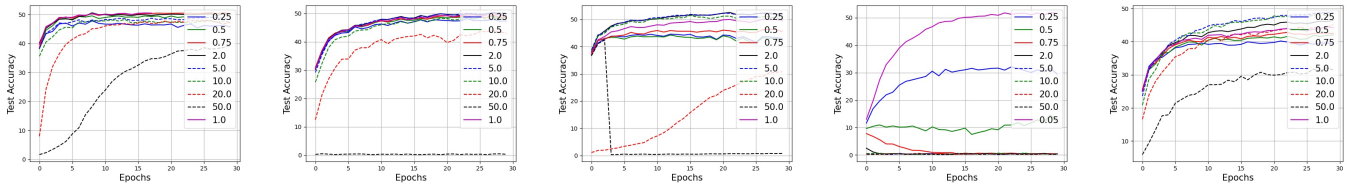
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 12: Target domain test accuracies for varied loss trade-off hyperparameters. Clipart to infograph adaptation task for full dataset.



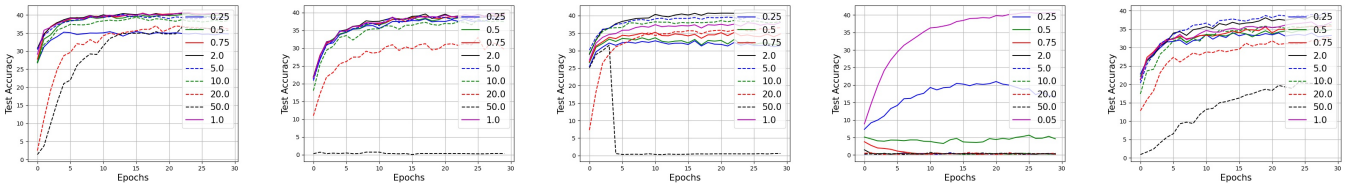
(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 13: Target domain test accuracies for varied loss trade-off hyperparameters. Clipart to quickdraw adaptation task for full dataset.



(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 14: Target domain test accuracies for varied loss trade-off hyperparameters. Clipart to real adaptation task for full dataset.



(a) DANN (b) JAN (c) CDAN (d) AFN (e) MCC

Figure 15: Target domain test accuracies for varied loss trade-off hyperparameters. Clipart to sketch adaptation task for full dataset.

DANN	clp	inf	pnt	qdr	rel	skt
clp	-	22.9 ± 0.4	34.1 ± 0.2	17.9 ± 0.3	45.3 ± 0.5	45.9 ± 0.4
inf	17.1 ± 0.2	-	17.0 ± 0.2	3.9 ± 0.3	21.6 ± 0.1	17.5 ± 0.1
pnt	32.0 ± 0.3	23.0 ± 0.3	-	7.5 ± 0.2	46.2 ± 0.2	38.7 ± 0.1
qdr	11.8 ± 0.3	5.2 ± 0.5	5.3 ± 0.5	-	6.6 ± 0.2	11.2 ± 0.4
rel	46.8 ± 0.3	33.6 ± 0.5	49.7 ± 0.1	12.8 ± 0.2	-	48.1 ± 0.2
skt	37.3 ± 0.3	19.8 ± 0.4	30.9 ± 0.4	12.5 ± 0.5	35.4 ± 0.5	-
JAN	clp	inf	pnt	qdr	rel	skt
clp	-	24.0 ± 0.4	35.4 ± 0.1	14.8 ± 0.4	44.6 ± 0.3	44.4 ± 0.5
inf	15.4 ± 0.2	-	16.0 ± 0.3	2.6 ± 0.1	19.5 ± 0.3	14.7 ± 0.2
pnt	31.5 ± 0.4	23.3 ± 0.2	-	6.9 ± 0.3	45.5 ± 0.2	36.2 ± 0.4
qdr	10.5 ± 0.4	4.3 ± 0.2	4.2 ± 0.2	-	4.6 ± 0.3	9.5 ± 0.6
rel	46.6 ± 0.3	35.2 ± 0.4	50.1 ± 0.3	12.1 ± 0.7	-	44.6 ± 0.1
skt	36.5 ± 0.4	20.0 ± 0.4	33.1 ± 0.5	11.2 ± 0.3	34.6 ± 0.4	-
CDAN	clp	inf	pnt	qdr	rel	skt
clp	-	22.4 ± 0.2	34.7 ± 0.2	19.1 ± 0.5	47.1 ± 0.2	46.2 ± 0.3
inf	16.9 ± 0.2	-	17.1 ± 0.1	4.0 ± 0.2	21.8 ± 0.1	17.1 ± 0.3
pnt	32.2 ± 0.7	21.7 ± 0.3	-	6.9 ± 0.5	47.8 ± 0.2	39.0 ± 0.2
qdr	8.4 ± 0.5	4.3 ± 0.2	3.9 ± 0.3	-	3.6 ± 0.5	8.5 ± 0.7
rel	49.1 ± 0.3	33.9 ± 0.8	50.3 ± 0.2	13.8 ± 0.4	-	48.5 ± 0.2
skt	37.1 ± 0.2	18.7 ± 0.2	31.8 ± 0.5	13.4 ± 0.5	36.4 ± 0.2	-
AFN	clp	inf	pnt	qdr	rel	skt
clp	-	27.0 ± 0.2	36.5 ± 0.4	17.3 ± 0.5	43.2 ± 0.4	47.2 ± 0.1
inf	14.8 ± 0.1	-	15.8 ± 0.2	2.8 ± 0.1	17.2 ± 0.2	15.0 ± 0.2
pnt	34.2 ± 0.3	29.6 ± 0.2	-	6.7 ± 0.2	45.9 ± 0.1	41.0 ± 0.2
qdr	10.5 ± 0.3	3.2 ± 0.2	4.2 ± 0.2	-	4.5 ± 0.1	10.6 ± 0.1
rel	49.9 ± 0.3	42.5 ± 0.3	52.8 ± 0.2	12.2 ± 0.3	-	50.3 ± 0.1
skt	38.2 ± 0.4	22.0 ± 0.2	34.3 ± 0.3	9.2 ± 0.3	33.7 ± 0.2	-
MCC	clp	inf	pnt	qdr	rel	skt
clp	-	26.4 ± 0.4	38.1 ± 0.5	11.9 ± 0.4	47.2 ± 0.2	49.4 ± 0.3
inf	11.7 ± 0.2	-	13.2 ± 0.2	1.2 ± 0.1	16.8 ± 0.3	12.4 ± 0.3
pnt	29.6 ± 0.4	24.5 ± 0.7	-	2.9 ± 0.0	43.3 ± 0.2	37.0 ± 0.4
qdr	12.0 ± 0.5	2.7 ± 0.3	4.0 ± 0.4	-	4.7 ± 0.4	11.2 ± 0.5
rel	45.1 ± 0.6	38.6 ± 0.7	49.0 ± 0.3	6.6 ± 0.5	-	46.1 ± 0.5
skt	34.9 ± 0.5	19.0 ± 0.5	32.5 ± 0.3	7.4 ± 0.9	33.2 ± 0.4	-

Table IV: Mean and standard deviation target domain test accuracy over 5-cross-val folds for all models for all 30 adaptation tasks on the subset-50 dataset. (Columns = source domain, rows = target domain)

DANN	clp	inf	pnt	qdr	rel	skt
clp	-	17.5 ± 0.5	29.0 ± 0.4	16.4 ± 0.6	41.0 ± 0.3	39.0 ± 0.4
inf	14.1 ± 0.1	-	14.9 ± 0.2	3.8 ± 0.1	19.9 ± 0.1	14.8 ± 0.3
pnt	27.1 ± 0.3	18.3 ± 0.8	-	7.2 ± 0.2	43.0 ± 0.2	34.2 ± 0.3
qdr	9.9 ± 0.1	3.9 ± 0.2	5.0 ± 0.4	-	6.1 ± 0.6	9.6 ± 0.2
rel	42.2 ± 0.5	28.3 ± 0.3	46.3 ± 0.2	12.4 ± 0.4	-	43.5 ± 0.3
skt	30.8 ± 0.3	15.0 ± 0.3	27.1 ± 0.4	11.2 ± 0.2	32.7 ± 0.2	-
JAN	clp	inf	pnt	qdr	rel	skt
clp	-	18.5 ± 0.7	30.5 ± 0.5	13.7 ± 0.5	40.3 ± 0.4	37.6 ± 0.4
inf	12.8 ± 0.2	-	14.0 ± 0.2	2.6 ± 0.2	17.8 ± 0.1	12.3 ± 0.2
pnt	26.7 ± 0.3	19.9 ± 0.3	-	6.6 ± 0.2	43.0 ± 0.3	31.6 ± 0.3
qdr	7.9 ± 0.5	3.1 ± 0.1	4.1 ± 0.3	-	3.9 ± 0.4	7.7 ± 0.5
rel	40.8 ± 0.5	31.0 ± 0.3	47.3 ± 0.3	11.3 ± 0.6	-	39.5 ± 0.6
skt	29.7 ± 0.3	15.4 ± 0.5	28.4 ± 0.3	9.9 ± 0.4	30.8 ± 0.3	-
CDAN	clp	inf	pnt	qdr	rel	skt
clp	-	16.7 ± 0.2	30.3 ± 0.7	16.1 ± 0.7	42.3 ± 0.2	38.9 ± 0.3
inf	13.7 ± 0.1	-	14.9 ± 0.1	3.6 ± 0.1	20.0 ± 0.1	14.0 ± 0.1
pnt	25.3 ± 0.2	18.9 ± 0.6	-	6.2 ± 0.3	44.6 ± 0.1	34.1 ± 0.5
qdr	4.8 ± 0.6	2.2 ± 0.4	2.5 ± 0.3	-	3.3 ± 0.2	5.4 ± 0.5
rel	42.1 ± 0.6	29.8 ± 0.7	47.8 ± 0.3	12.2 ± 0.6	-	44.1 ± 0.4
skt	28.1 ± 0.5	12.3 ± 0.3	27.4 ± 0.4	11.7 ± 0.3	33.0 ± 0.3	-
AFN	clp	inf	pnt	qdr	rel	skt
clp	-	21.4 ± 0.4	32.7 ± 0.2	17.0 ± 0.6	40.7 ± 0.3	41.2 ± 0.2
inf	13.2 ± 0.3	-	14.7 ± 0.2	2.8 ± 0.1	16.7 ± 0.1	13.0 ± 0.2
pnt	30.3 ± 0.4	24.8 ± 0.1	-	6.2 ± 0.3	44.7 ± 0.2	36.8 ± 0.3
qdr	8.9 ± 0.3	2.4 ± 0.3	3.3 ± 0.2	-	3.8 ± 0.3	9.5 ± 0.1
rel	45.3 ± 0.4	35.9 ± 0.6	50.7 ± 0.3	11.8 ± 0.3	-	45.9 ± 0.3
skt	32.5 ± 0.2	17.5 ± 0.3	30.4 ± 0.2	8.7 ± 0.3	31.8 ± 0.0	-
MCC	clp	inf	pnt	qdr	rel	skt
clp	-	19.0 ± 0.5	30.6 ± 0.4	11.6 ± 0.3	40.6 ± 0.3	40.3 ± 0.6
inf	8.7 ± 0.3	-	10.7 ± 0.2	1.2 ± 0.1	15.0 ± 0.3	9.6 ± 0.2
pnt	22.5 ± 0.4	18.4 ± 0.5	-	3.1 ± 0.3	39.6 ± 0.3	31.0 ± 0.6
qdr	9.6 ± 0.2	2.9 ± 0.6	4.3 ± 0.3	-	4.9 ± 0.6	9.8 ± 0.6
rel	37.8 ± 0.3	30.6 ± 0.4	44.3 ± 0.4	7.2 ± 0.9	-	39.6 ± 0.3
skt	26.4 ± 0.7	13.5 ± 0.5	26.1 ± 0.3	7.9 ± 0.5	27.9 ± 0.5	-

Table V: Mean and standard deviation target domain test accuracy over 5-cross-val folds for all models for all 30 adaptation tasks on the subset-20 dataset. (Columns = source domain, rows = target domain)

DANN	Amazon	DSLR	Webcam
Amazon	-	69.8 +/- 1.0	70.0 +/- 1.7
DSLR	91.7 +/- 2.6	-	99.7 +/- 0.6
Webcam	89.2 +/- 1.3	97.5 +/- 0.6	-
JAN	Amazon	DSLR	Webcam
Amazon	-	70.1 +/- 0.5	68.6 +/- 0.6
DSLR	92.5 +/- 1.2	-	100.0 +/- 0.0
Webcam	90.0 +/- 1.8	97.5 +/- 0.6	-
CDAN	Amazon	DSLR	Webcam
Amazon	-	69.9 +/- 2.9	65.9 +/- 1.0
DSLR	91.7 +/- 1.5	-	100.0 +/- 0.0
Webcam	89.7 +/- 1.5	97.3 +/- 0.4	-
AFN	Amazon	DSLR	Webcam
Amazon	-	70.3 +/- 0.8	70.2 +/- 1.0
DSLR	93.1 +/- 0.6	-	100.0 +/- 0.0
Webcam	89.2 +/- 0.8	98.2 +/- 0.4	-
MCC	Amazon	DSLR	Webcam
Amazon	-	74.7 +/- 1.8	77.2 +/- 0.6
DSLR	94.7 +/- 2.5	-	100.0 +/- 0.0
Webcam	92.2 +/- 1.9	98.3 +/- 0.6	-

Table VI: Mean and standard deviation target domain test accuracy over 5-cross-val folds for all models for all 6 adaptation tasks with tuned hyperparameters on the *Office-31* dataset. (Columns = source domain, rows = target domain)

DANN	Amazon	DSLR	Webcam
Amazon	-	72.5 +/- 1.4	74.2 +/- 1.2
DSLR	89.3 +/- 2.5	-	99.7 +/- 0.6
Webcam	88.0 +/- 2.0	97.5 +/- 1.0	-
JAN	Amazon	DSLR	Webcam
Amazon	-	69.2 +/- 0.5	69.9 +/- 0.7
DSLR	89.3 +/- 0.9	-	99.7 +/- 0.6
Webcam	94.0 +/- 2.6	96.7 +/- 0.6	-
CDAN	Amazon	DSLR	Webcam
Amazon	-	72.2 +/- 2.6	71.3 +/- 2.1
DSLR	93.3 +/- 1.9	-	100.0 +/- 0.0
Webcam	93.3 +/- 1.1	98.0 +/- 0.7	-
AFN	Amazon	DSLR	Webcam
Amazon	-	69.9 +/- 0.8	70.3 +/- 1.1
DSLR	92.0 +/- 1.6	-	100.0 +/- 0.0
Webcam	88.3 +/- 0.6	98.2 +/- 0.4	-
MCC	Amazon	DSLR	Webcam
Amazon	-	8.7 +/- 1.7	12.3 +/- 1.0
DSLR	10.7 +/- 1.6	-	56.3 +/- 4.8
Webcam	9.3 +/- 1.4	47.2 +/- 3.9	-

Table VII: Mean and standard deviation target domain test accuracy over 5-cross-val folds for all models for all 6 adaptation tasks with tuned hyperparameters from the *DomainNet* dataset. (Columns = source domain, rows = target domain)