# Appendix for "Rethinking the Role of Pre-Trained Networks in Source-Free Domain Adaptation"

Wenyu Zhang[1], Li Shen[1], Chuan-Sheng Foo[1,2]
[1]Institute for Infocomm Research (I$^2$R), Agency for Science, Technology and Research (A*STAR)
[2]Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR)

## 1. Detailed Results

We provide detailed results of experiments to analyze our proposed framework and strategy.

### 1.1. Main results

In Table 3a, 3b and 3c, we provide detailed results of our proposed strategy for the datasets Office-31, Office-Home and VisDA-C with co-learning networks: ResNet-50, ResNet-101, ConvNeXt-S, Swin-S, ConvNeXt-B and Swin-B, where S and B denote the small and base versions of the architectures, respectively. In general, co-learning with the more recently-released ConvNeXt and Swin networks achieves better adaptation performance than co-learning with the ResNets. In particular, co-learning with Swin-B has the best target accuracy in most cases.

In addition, we find that even networks with poor feature extraction ability, such as AlexNet, can provide useful features to improve performance when target style is similar to ImageNet. Co-learned with AlexNet, SHOT and SHOT++ overall Office-Home accuracy (71.9% and 72.7%) is little changed at 71.8% ($\downarrow$ 0.1%) and 72.7% ($=$). However, 5 out of 12 domain pairs improved by 0.1-1% and 0.1-1.2% especially when target is Product or Real World.

### 1.2. Further analysis on co-learning with pre-trained networks

**Pre-trained model branch.** We experiment with different updates to the pre-trained model branch using a subset of domain pairs from Office-Home dataset in Table 1. We choose to update no component or different component(s) of the pre-trained network, such as feature extractor, weighted nearest-centroid-classifier (NCC) and a linear 1-layer logit projection layer inserted after NCC. Finetuning just the classifier achieves the best result overall. Co-learning depends on the two branches providing different views of classification decisions. Finetuning the feature extractor or projection layer risks pre-trained model predictions converging too quickly to adaptation model predictions.

| Feat. extr. | Classifier | Projection | A → C | A → P | A → R | Avg |
|---|---|---|---|---|---|---|
| ✗ | ✗ | ✗ | 59.4 | 83.8 | 86.5 | 76.6 |
| ✓ | ✗ | ✗ | 59.6 | 82.6 | 86.4 | 76.2 |
| ✓ | ✓ | ✗ | **60.3** | 84.9 | 86.4 | 77.2 |
| ✗ | ✓ | ✗ | *59.7* | **86.3** | *87.1* | **77.7** |
| ✗ | ✓ | ✓ | 59.5 | *85.9* | **87.2** | *77.5* |
| ✗ | ✗ | ✓ | 59.5 | 84.1 | 86.5 | 76.7 |

Table 1: Co-learning experiments on the component fine-tuned in pre-trained model branch on Office-Home, with ImageNet-1k ConvNeXt-S in pre-trained model branch.

| Centroid computation | A → C | A → P | A → R | Avg |
|---|---|---|---|---|
| All samples | **59.7** | **86.3** | **87.1** | **77.7** |
| Partial samples | 59.1 | 86.1 | 87.0 | 77.4 |

(a) Samples used to estimate class centroids in classifier: all samples versus only samples pseudolabeled by MatchOrConf

| # Iterations | A → C | A → P | A → R | Avg |
|---|---|---|---|---|
| 1 | **59.7** | **86.3** | 87.1 | **77.7** |
| 2 | 58.2 | 85.5 | **87.2** | 77.0 |
| 3 | 57.7 | 85.9 | 86.6 | 76.7 |

(b) Number of iterations to estimate class centroids in classifier

Table 2: Co-learning experiments on the weighted nearest-centroid classifier in pre-trained model branch on Office-Home, with ImageNet-1k ConvNeXt-S in pre-trained model branch.

We also experiment with the computation of weighted nearest-centroid-classifier (NCC) in the pre-trained model branch in Table 2. In Table 2a, we find that class centroids in the NCC classifier should be estimated using all samples, instead of only pseudolabeled samples, to better represent the entire target data distribution. Table 2b shows that refining class centroids iteratively (as in k-means) is not beneficial, as the model predictions used for centroid estimation have inaccuracies and increased iterations can reinforce the bias that these inaccurate predictions have on the resulting

centroids.

**Pseudolabel proportion.** Figure 1 plots the proportion of pseudolabeled samples on target domains after co-learning with different ImageNet pre-trained feature extractors. Overall, the percentage of samples pseudolabeled is $96.4\%$ to $100\%$ in Office-31, $91.8\%$ to $98.7\%$ in Office-Home with percentage above $94\%$ in most cases, and $90.5\%$ to $92.5\%$ in VisDA-C. Samples are not pseudolabeled if the two model branches have the same confidence level and disagree on the class prediction. This disagreement rate is comparatively higher when the target inputs are more different from both source and ImageNet inputs, such as Amazon product images in Office-31 and Clipart images in Office-Home.

### 1.3. Further discussion on characteristics of pre-trained networks

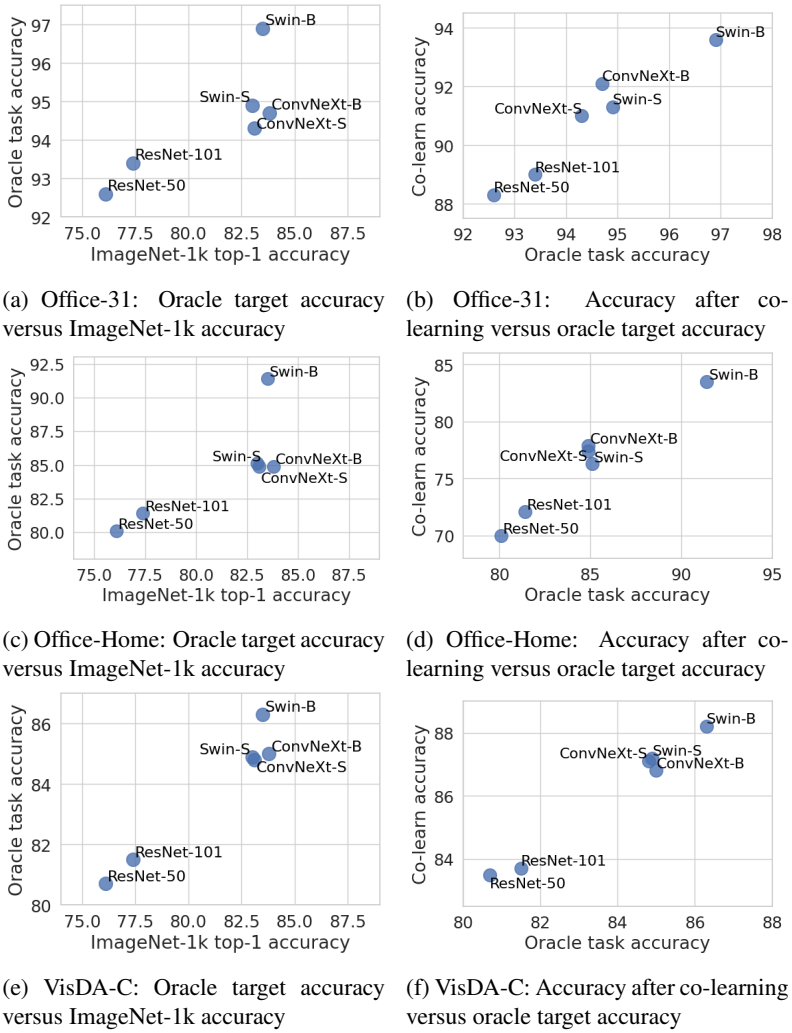
We provide detailed results on the study of preferred feature extractor characteristics for co-learning.

**Oracle target accuracy versus ImageNet-1k accuracy.** We plot the relationship of ImageNet-1k top-1 accuracy and average oracle target accuracy attained by pre-trained feature extractors for each dataset in Figure 2, and list the per-domain values in Table 4. Oracle target accuracy is computed by fitting a nearest-centroid-classifier head on each pre-trained feature extractor using fully-labeled target data. In general, higher ImageNet-1k top-1 accuracy corresponds to higher oracle target accuracy. However, there are exceptions. For instance, Swin-B has lower ImageNet-1k accuracy than ConvNext-B ($83.5\%$ vs. $83.8\%$), but higher oracle accuracy for all target domains tested. Despite being trained on the same ImageNet-1k and achieving similar ImageNet-1k performance, the feature extractors learn different features due to differences in architecture and training schemes. While ConvNeXt-B may be more robust to small amounts of input distribution shift due to train-test data splits, Swin-B may be more robust to larger cross-dataset covariate shift and more compatible with the target domains tested.

**Adapted model accuracy versus oracle target accuracy** We also plot the relationship of average oracle target accuracy of pre-trained feature extractors and the performance of adapted model after co-learning on target domains in Figure 2. In general, a pre-trained feature extractor with higher oracle target accuracy benefits co-learning as it is more capable of correctly pseudolabeling the target samples, and consequently results in higher adapted model performance. For instance in Table 5, by using ResNet architectures for co-learning in the pre-trained model branch, we see that the larger and more powerful ImageNet ResNet-101 typically results in higher adapted model performance than ImageNet ResNet-50. There are also exceptions to this correlation. With the same ResNet-50 architecture, although source ResNet-50 has higher oracle accuracy than ImageNet ResNet-50 in some source-target domain pairs (e.g. Office-31 $A \rightarrow D$, $A \rightarrow W$, $W \rightarrow D$; Office-Home $A \rightarrow R$, $P \rightarrow C$, $R \rightarrow A$), the latter results in equal or higher adapted model performance. The ImageNet feature extractors benefit co-learning as they provide an additional view of features and classification decisions different from the source feature extractor already in the adaptation model branch.


(a) Office-31 (b) Office-Home (c) VisDA-C

Figure 1: Proportion of pseudolabeled samples after co-learning with ImageNet-1k pre-trained feature extractor.




(a) Office-31: Oracle target accuracy versus ImageNet-1k accuracy

(b) Office-31: Accuracy after co-learning versus oracle target accuracy

(c) Office-Home: Oracle target accuracy versus ImageNet-1k accuracy

(d) Office-Home: Accuracy after co-learning versus oracle target accuracy

(e) VisDA-C: Oracle target accuracy versus ImageNet-1k accuracy

(f) VisDA-C: Accuracy after co-learning versus oracle target accuracy

Figure 2: Performance evaluations of ImageNet-1k feature extractors. Oracle target accuracy of ImageNet-1k feature extractor, and accuracy of adapted model after co-learning with the feature extractor, are evaluated on dataset target domains.

## References

[1] J. Liang, D. Hu, and J. Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. *ICML*, 2020. 4, 5

[2] J. Liang, D. Hu, Y. Wang, R. He, and J. Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE TPAMI*, 2021. 4, 5

| Method | Office-31 | | | | | | |
|---|---|---|---|---|---|---|---|
| | A → D | A → W | D → A | D → W | W → A | W → D | Avg |
| Source Only | 81.9 | 78.0 | 59.4 | 93.6 | 63.4 | 98.8 | 79.2 |
| Co-learn (w/ Resnet-50) | 93.6 | 90.2 | 75.7 | 98.2 | 72.5 | 99.4 | 88.3 |
| Co-learn (w/ Resnet-101) | 94.2 | 91.6 | 74.7 | 98.6 | 75.6 | 99.6 | 89.0 |
| Co-learn (w/ ConvNeXt-S) | 96.6 | 92.6 | 79.8 | 97.7 | 79.6 | 99.4 | 91.0 |
| Co-learn (w/ Swin-S) | 96.8 | 93.3 | 79.2 | 98.7 | 80.2 | 99.6 | 91.3 |
| Co-learn (w/ ConvNeXt-B) | **97.8** | 96.6 | 80.5 | 98.5 | 79.4 | 99.6 | 92.1 |
| Co-learn (w/ Swin-B) | 97.4 | **98.2** | **84.5** | **99.1** | **82.2** | **100.0** | **93.6** |
| SHOT† [1] | 95.0 | 90.4 | 75.2 | **98.9** | 72.8 | 99.8 | 88.7 |
| w/ Co-learn (w/ ResNet-50) | 94.2 | 90.2 | 75.7 | 98.2 | 74.4 | **100.0** | 88.8 |
| w/ Co-learn (w/ ResNet-101) | **96.4** | 90.7 | 77.0 | 98.2 | 75.5 | 99.8 | 89.6 |
| w/ Co-learn (w/ ConvNeXt-S) | 95.2 | 91.4 | 77.5 | **98.9** | 76.4 | **100.0** | 89.9 |
| w/ Co-learn (w/ Swin-S) | 95.2 | 91.6 | 77.7 | 98.5 | **76.8** | 99.6 | 89.9 |
| w/ Co-learn (w/ ConvNeXt-B) | 95.8 | 92.8 | 77.2 | 98.4 | 76.4 | **100.0** | 90.1 |
| w/ Co-learn (w/ Swin-B) | 95.8 | **95.6** | **78.5** | **98.9** | 76.7 | 99.8 | **90.9** |
| SHOT++† [2] | 95.6 | 90.8 | 76.0 | 98.2 | 74.6 | **100.0** | 89.2 |
| w/ Co-learn (w/ ResNet-50) | 95.0 | 90.7 | 76.4 | 97.7 | 74.9 | 99.8 | 89.1 |
| w/ Co-learn (w/ ResNet-101) | 95.2 | 90.7 | 77.3 | 98.4 | 76.1 | 99.8 | 89.6 |
| w/ Co-learn (w/ ConvNeXt-S) | 96.0 | 91.8 | 77.1 | 98.4 | 77.1 | **100.0** | 90.1 |
| w/ Co-learn (w/ Swin-S) | 96.8 | 93.0 | 78.2 | 98.7 | 77.4 | **100.0** | 90.7 |
| w/ Co-learn (w/ ConvNeXt-B) | 96.0 | 91.6 | 77.2 | **99.0** | 76.9 | **100.0** | 90.1 |
| w/ Co-learn (w/ Swin-B) | **96.6** | **93.8** | **79.8** | 98.9 | **78.0** | **100.0** | **91.2** |

(a) Office-31: 31-class classification accuracy of adapted ResNet-50

| Method | Office-Home | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A → C | A → P | A → R | C → A | C → P | C → R | P → A | P → C | P → R | R → A | R → C | R → P | Avg |
| Source Only | 43.5 | 67.1 | 74.2 | 51.5 | 62.2 | 63.3 | 51.4 | 40.7 | 73.2 | 64.6 | 45.8 | 77.6 | 59.6 |
| Co-learn (w/ Resnet-50) | 51.8 | 78.9 | 81.3 | 66.7 | 78.8 | 79.4 | 66.3 | 50.0 | 80.6 | 71.1 | 53.7 | 81.3 | 70.0 |
| Co-learn (w/ Resnet-101) | 54.6 | 81.8 | 83.5 | 68.6 | 79.3 | 80.4 | 68.7 | 52.3 | 82.0 | 72.4 | 57.1 | 84.1 | 72.1 |
| Co-learn (w/ ConvNeXt-S) | 59.7 | 86.3 | 87.1 | 75.9 | 84.5 | 86.8 | 76.1 | 58.7 | 87.1 | 78.0 | 61.9 | 87.2 | 77.4 |
| Co-learn (w/ Swin-S) | 56.4 | 85.1 | 88.0 | 73.9 | 83.7 | 86.1 | 75.4 | 55.3 | 87.8 | 77.3 | 58.9 | 87.9 | 76.3 |
| Co-learn (w/ ConvNeXt-B) | 60.5 | 85.9 | 87.2 | 76.1 | 85.3 | 86.6 | 76.5 | 58.6 | 87.5 | 78.9 | 62.4 | 88.8 | 77.9 |
| Co-learn (w/ Swin-B) | **69.6** | **89.5** | **91.2** | **82.7** | **88.4** | **91.3** | **82.6** | **68.5** | **91.5** | **82.8** | **71.3** | **92.1** | **83.5** |
| SHOT† [1] | 55.8 | 79.6 | 82.0 | 67.4 | 77.9 | 77.9 | 67.6 | 55.6 | 81.9 | 73.3 | 59.5 | 84.0 | 71.9 |
| w/ Co-learn (w/ ResNet-50) | 56.3 | 79.9 | 82.9 | 68.5 | 79.6 | 78.7 | 68.1 | 54.8 | 82.5 | 74.5 | 59.0 | 83.6 | 72.4 |
| w/ Co-learn (w/ ResNet-101) | 57.2 | 80.6 | 83.5 | 69.1 | 79.1 | 79.9 | 69.1 | 56.5 | 82.3 | 74.8 | 59.9 | 85.3 | 73.1 |
| w/ Co-learn (w/ ConvNeXt-S) | 58.9 | 81.5 | 84.6 | 71.7 | 79.7 | 82.0 | 71.3 | 57.5 | 84.2 | 75.9 | 61.9 | 86.0 | 74.6 |
| w/ Co-learn (w/ Swin-S) | 58.9 | 81.3 | 84.5 | 70.7 | 80.1 | 81.9 | 69.8 | 57.8 | 83.7 | 75.4 | 61.0 | 86.0 | 74.3 |
| w/ Co-learn (w/ ConvNeXt-B) | 60.0 | 81.1 | 84.3 | 71.4 | 80.4 | 81.5 | 70.9 | 58.8 | 83.8 | **76.2** | 62.8 | 86.3 | 74.8 |
| w/ Co-learn (w/ Swin-B) | **61.7** | **82.9** | **85.3** | **72.7** | **80.5** | **82.0** | **71.6** | **60.4** | **84.5** | 76.0 | **64.3** | **86.7** | **75.7** |
| SHOT++† [2] | 57.1 | 79.5 | 82.6 | 68.5 | 79.5 | 78.6 | 68.3 | 56.1 | 82.9 | 74.0 | 59.8 | 85.0 | 72.7 |
| w/ Co-learn (w/ ResNet-50) | 57.7 | 81.1 | 84.0 | 69.2 | 79.8 | 79.2 | 69.1 | 57.7 | 82.9 | 73.7 | 60.1 | 85.0 | 73.3 |
| w/ Co-learn (w/ ResNet-101) | 59.4 | 80.5 | 84.0 | 69.3 | 80.3 | 80.5 | 69.8 | 57.7 | 83.2 | 74.2 | 60.4 | 85.6 | 73.7 |
| w/ Co-learn (w/ ConvNeXt-S) | 60.7 | 82.4 | 84.9 | 71.1 | 80.7 | 82.3 | 70.8 | 59.4 | 84.2 | 75.8 | 63.2 | 85.9 | 75.1 |
| w/ Co-learn (w/ Swin-S) | 60.4 | 81.6 | 84.7 | 71.0 | 80.7 | 81.3 | 71.1 | 57.4 | 84.6 | 75.7 | 61.9 | 85.8 | 74.7 |
| w/ Co-learn (w/ ConvNeXt-B) | 60.8 | 80.9 | 84.5 | 70.8 | 81.2 | 83.1 | 70.9 | **60.1** | 84.3 | **76.4** | 62.5 | 85.6 | 75.1 |
| w/ Co-learn (w/ Swin-B) | **63.7** | **83.0** | **85.7** | **72.6** | **81.5** | **83.8** | **72.0** | 59.9 | **85.3** | 76.3 | **65.3** | **86.6** | **76.3** |

(b) Office-Home: 65-class classification accuracy of adapted ResNet-50

| Method | VisDA-C | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **plane** | **bike** | **bus** | **car** | **horse** | **knife** | **mcycle** | **person** | **plant** | **sktbrd** | **train** | **truck** | **Avg** |
| Source Only | 51.5 | 15.3 | 43.4 | 75.4 | 71.2 | 6.8 | 85.5 | 18.8 | 49.4 | 46.4 | 82.1 | 5.4 | 45.9 |
| Co-learn (w/ Resnet-50) | 96.2 | 76.2 | 77.5 | 77.8 | 93.8 | 96.6 | 91.5 | 76.7 | 90.4 | 90.8 | 86.0 | 48.9 | 83.5 |
| Co-learn (w/ Resnet-101) | 96.5 | 78.9 | 77.5 | 75.7 | 94.6 | 95.8 | 89.1 | 77.7 | 90.5 | 91.0 | 86.2 | 51.5 | 83.7 |
| Co-learn (w/ ConvNeXt-S) | 97.8 | 89.7 | 82.3 | **81.3** | 97.3 | 97.8 | 93.4 | 66.9 | 95.4 | 96.0 | 90.7 | **56.5** | 87.1 |
| Co-learn (w/ Swin-S) | 97.8 | 88.5 | 84.7 | 78.5 | 96.8 | 97.8 | 93.3 | 73.9 | 94.9 | 94.8 | 91.2 | 54.8 | 87.2 |
| Co-learn (w/ ConvNeXt-B) | 98.0 | 89.2 | **84.9** | 80.2 | 97.0 | 98.4 | 93.6 | 64.3 | **95.6** | **96.3** | 90.4 | 54.0 | 86.8 |
| Co-learn (w/ Swin-B) | **99.0** | **90.0** | 84.2 | 81.0 | **98.1** | **97.9** | **94.9** | **80.1** | 94.8 | 95.9 | **94.4** | 48.1 | **88.2** |
| SHOT† [1] | 95.3 | 87.1 | 79.1 | 55.1 | 93.2 | 95.5 | 79.5 | 79.6 | 91.6 | 89.5 | 87.9 | 56.0 | 82.4 |
| w/ Co-learn (w/ ResNet-50) | 95.5 | 86.1 | 76.3 | 57.6 | 93.2 | **96.9** | 83.4 | 81.1 | 92.0 | 90.5 | 84.5 | 60.7 | 83.1 |
| w/ Co-learn (w/ ResNet-101) | 94.9 | 84.8 | 77.7 | 63.0 | 94.1 | 95.6 | 85.6 | 81.0 | 93.0 | 92.2 | 86.4 | 60.4 | 84.1 |
| w/ Co-learn (w/ ConvNeXt-S) | 95.6 | **89.8** | **82.2** | **66.2** | 94.2 | 96.3 | 85.4 | 82.0 | 92.8 | 91.5 | 86.6 | 59.3 | 85.1 |
| w/ Co-learn (w/ Swin-S) | 95.2 | 88.1 | 81.5 | 63.9 | 93.9 | 95.6 | 86.7 | 82.1 | **94.0** | 92.3 | 87.3 | **63.9** | **85.3** |
| w/ Co-learn (w/ ConvNeXt-B) | 95.1 | 87.5 | 81.2 | 62.4 | **94.4** | 96.4 | 86.2 | **82.3** | 93.1 | **92.5** | 88.5 | 63.0 | 85.2 |
| w/ Co-learn (w/ Swin-B) | **96.0** | 88.1 | 81.0 | 63.0 | 94.3 | 95.9 | **87.1** | 81.8 | 92.8 | 91.9 | **90.1** | 60.5 | 85.2 |
| SHOT++† [2] | 94.5 | 88.5 | **90.4** | 84.6 | **97.9** | **98.6** | 91.9 | 81.8 | 96.7 | 91.5 | 93.8 | 31.3 | 86.8 |
| w/ Co-learn (w/ ResNet-50) | 97.5 | 89.8 | 86.4 | 82.7 | 97.8 | 98.5 | 92.7 | 83.5 | 97.3 | 90.8 | 97.5 | 36.7 | 87.4 |
| w/ Co-learn (w/ ResNet-101) | 97.9 | 88.6 | 86.8 | **86.7** | **97.9** | **98.6** | 92.4 | 83.6 | 97.4 | 92.5 | 94.4 | 32.5 | 87.4 |
| w/ Co-learn (w/ ConvNeXt-S) | 97.5 | 90.3 | 89.1 | 85.6 | 97.7 | 98.1 | **93.2** | 84.7 | 96.5 | 92.6 | **95.1** | 39.1 | 88.3 |
| w/ Co-learn (w/ Swin-S) | 97.6 | 87.3 | 86.8 | 83.3 | **97.9** | 98.2 | 92.1 | 84.4 | 97.6 | 90.2 | 94.7 | 42.0 | 87.7 |
| w/ Co-learn (w/ ConvNeXt-B) | 97.7 | 88.6 | 86.2 | **86.7** | **97.9** | 98.3 | 90.8 | 82.8 | **97.7** | 90.3 | 95.0 | 39.0 | 87.6 |
| w/ Co-learn (w/ Swin-B) | **98.0** | **91.1** | 88.6 | 83.2 | 97.8 | 97.8 | 92.0 | **85.8** | 97.6 | **93.2** | 95.0 | **43.5** | **88.6** |

(c) VisDA-C: 12-class classification accuracy of adapted ResNet-101

Table 3: Classification accuracy of adapted models. For proposed strategy, ImageNet-1k feature extractor used is given in parenthesis. † denotes reproduced results.

| Pre-training data | Feat. extr. | # Param. | IN-1k Acc@1 | Office-31 | | | |
|---|---|---|---|---|---|---|---|
| | | | | Amazon | DSLR | Webcam | Avg |
| ImageNet-1k, then Amazon | ResNet-50 | 26M | - | - | 98.6 | 98.1 | - |
| ImageNet-1k, then DSLR | ResNet-50 | 26M | - | 80.1 | - | 97.7 | - |
| ImageNet-1k, then Webcam | ResNet-50 | 26M | - | 81.7 | 99.2 | - | - |
| ImageNet-1k | ResNet-50 | 26M | 76.1 | 81.8 | 98.2 | 97.9 | 92.6 |
| ImageNet-1k | ResNet-101 | 45M | 77.4 | 83.4 | 98.8 | 97.9 | 93.4 |
| ImageNet-1k | ConvNeXt-S | 50M | 83.1 | 85.2 | 99.4 | 98.2 | 94.3 |
| ImageNet-1k | Swin-S | 50M | 83.0 | 86.2 | 99.0 | 99.4 | 94.9 |
| ImageNet-1k | ConvNeXt-B | 89M | **83.8** | 86.2 | 99.4 | 98.6 | 94.7 |
| ImageNet-1k | Swin-B | 88M | 83.5 | **90.6** | **100.0** | **100.0** | **96.9** |

(a) Oracle target accuracy of pre-trained feature extractors on Office-31

| Pre-training data | Feat. extr. | # Param. | IN-1k Acc@1 | Office-Home | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Art | Clipart | Product | Real World | Avg |
| ImageNet-1k, then Art | ResNet-50 | 26M | - | - | 69.5 | 88.1 | 86.3 | - |
| ImageNet-1k, then Clipart | ResNe-50 | 26M | - | 79.3 | - | 86.0 | 83.3 | - |
| ImageNet-1k, then Product | ResNet-50 | 26M | - | 80.0 | 67.5 | - | 85.8 | - |
| ImageNet-1k, then Real World | ResNet-50 | 26M | - | 81.8 | 69.1 | 88.1 | - | - |
| ImageNet-1k | ResNet-50 | 26M | 76.1 | 81.2 | 65.5 | 87.7 | 86.0 | 80.1 |
| ImageNet-1k | ResNet-101 | 45M | 77.4 | 82.5 | 68.0 | 88.2 | 87.0 | 81.4 |
| ImageNet-1k | ConvNeXt-S | 50M | 83.1 | 86.0 | 72.3 | 91.7 | 89.4 | 84.9 |
| ImageNet-1k | Swin-S | 50M | 83.0 | 86.6 | 72.0 | 91.5 | 90.2 | 85.1 |
| ImageNet-1k | ConvNeXt-B | 89M | **83.8** | 85.7 | 73.4 | 91.4 | 89.2 | 84.9 |
| ImageNet-1k | Swin-B | 88M | 83.5 | **92.3** | **83.1** | **95.3** | **94.7** | **91.4** |

(b) Oracle target accuracy of pre-trained feature extractors on Office-Home

| Pre-training data | Feat. extr. | # Param. | IN-1k Acc@1 | VisDA-C |
|---|---|---|---|---|
| ImageNet-1k, then Synthetic | ResNet-101 | 45M | - | 71.6 |
| ImageNet-1k | ResNet-50 | 26M | 76.1 | 80.7 |
| ImageNet-1k | ResNet-101 | 45M | 77.4 | 81.5 |
| ImageNet-1k | ConvNeXt-S | 50M | 83.1 | 84.8 |
| ImageNet-1k | Swin-S | 50M | 83.0 | 84.9 |
| ImageNet-1k | ConvNeXt-B | 89M | **83.8** | 85.0 |
| ImageNet-1k | Swin-B | 88M | 83.5 | **86.3** |

(c) Oracle target accuracy of pre-trained feature extractors on VisDA-C

Table 4: Oracle target accuracy of pre-trained feature extractors fitted with nearest-centroid-classifier heads learned on fully-labeled target domain data. IN-1k Acc@1 denotes ImageNet-1k top-1 accuracy. Feature extractors above dash line are from source models, below the dash line are from ImageNet-1k networks.

| Pre-trained feat. extr. | Office-31 | | | | | | |
|---|---|---|---|---|---|---|---|
| | **A → D** | **A → W** | **D → A** | **D → W** | **W → A** | **W → D** | **Avg** |
| Source ResNet-50 | 92.6 | 89.9 | 74.0 | 96.1 | *73.9* | *99.4* | 87.6 |
| ImageNet-1k ResNet-50 | *93.6* | *90.2* | **75.7** | *98.2* | 72.5 | *99.4* | *88.3* |
| ImageNet-1k ResNet-101 | **94.2** | **91.6** | *74.7* | **98.6** | **75.6** | **99.6** | **89.0** |

(a) Classification accuracy of adapted ResNet-50 on Office-31

| Pre-trained feat. extr. | Office-Home | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **A → C** | **A → P** | **A → R** | **C → A** | **C → P** | **C → R** | **P → A** | **P → C** | **P → R** | **R → A** | **R → C** | **R → P** | **Avg** |
| Source ResNet-50 | *53.9* | *79.0* | *81.3* | 63.5 | 75.6 | 74.1 | 63.7 | 49.8 | 80.1 | 70.5 | *54.8* | *81.4* | 69.0 |
| ImageNet-1k ResNet-50 | 51.8 | 78.9 | *81.3* | *66.7* | *78.8* | *79.4* | *66.3* | *50.0* | *80.6* | *71.1* | 53.7 | 81.3 | *70.0* |
| ImageNet-1k ResNet-101 | **54.6** | **81.8** | **83.5** | **68.6** | **79.3** | **80.4** | **68.7** | **52.3** | **82.0** | **72.4** | **57.1** | **84.1** | **72.1** |

(b) Classification accuracy of adapted ResNet-50 on Office-Home

| Pre-trained feat. extr. | VisDA-C | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **plane** | **bike** | **bus** | **car** | **horse** | **knife** | **mcycle** | **person** | **plant** | **sktbrd** | **train** | **truck** | **Avg** |
| Source ResNet-101 | 63.6 | 55.6 | *68.0* | 65.6 | 84.0 | 95.6 | 89.0 | 65.8 | 82.6 | 4.9 | 82.2 | 25.8 | 65.2 |
| ImageNet-1k ResNet-50 | *96.2* | *76.2* | **77.5** | **77.8** | *93.8* | **96.6** | **91.5** | *76.7* | *90.4* | *90.8* | *86.0* | *48.9* | *83.5* |
| ImageNet-1k ResNet-101 | **96.5** | **78.9** | **77.5** | *75.7* | **94.6** | *95.8* | *89.1* | **77.7** | **90.5** | **91.0** | **86.2** | **51.5** | **83.7** |

(c) Classification accuracy of adapted ResNet-101 on VisDA-C

Table 5: Classification accuracy of adapted model after co-learning with ResNets.