Supplementary Materials

Network Architectures

For fair comparison, we follow the network in DANN [1] and simply build the autoencoder network by fully connected layers. Here we provide all the network architectures in the extensive experiments. For digit adaptation tasks with 32×32 images, we draw the specifications of CNN and fully connected layers in Table 1. For object recognition tasks on Office-Home and Image-CLEF, we draw the network architectures in Table 2. We could see that we only involve several fully connected layers compared with domain adversarial training [1], which are still computational efficient.

Table 1: The network architecture used in digit adaptation .

Layer Index	Feature Extractor	Autoencoder	Label Predictor
0	$32 \times 32 \times 3$ Image		
1	5×5 conv. 64 ReLU	2048 dense, ReLU	10 dense, softmax
2	3×3 max-pool, stride 2	512 dense, ReLU	
3	5×5 conv. 64 ReLU	2048 dense, ReLU	
4	3×3 max-pool, stride 2		
5	5×5 conv. 128 ReLU		
6	3072 dense, dropout, ReLU		
7	2048 dense, dropout ReLU		

Table 2: The network architecture used in Office-Home and VisDA-17.

Layer Index	Feature Extractor	Autoencoder	Label Predictor
0	$224 \times 224 \times 3$ Image		
1	ResNet50 CNNs	2048 dense, ReLU	10 dense, softmax
2		512 dense, ReLU	
3		128 dense, ReLU	
4		512 dense, ReLU	
5		2048 dense, ReLU	

References

1. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 1180–1189. PMLR, Lille, France (07–09 Jul 2015)