A Implementation Details

For fair comparisons, we adopt a ResNet50 as our backbone on ImageCLEF and Office-Home and a ResNet101 on DomainNet. The weights for ResNet50 and ResNet101 are from CLIP and frozen through our experiments. Prompts and auto-encoder of MPA are trained using the mini-batch SGD optimizer with a learning rate of 0.003 and 0.005 while in LST, the learned subspace is tuned with a 0.0005 learning rate. We use a batch size of 32 and adopt a cosine learning rate scheduler. For hyper-parameters, token lengths M_1 and M_2 are both set to 16. Pseudo-label threshold τ is set to 0.4 for producing reliable labels. α in Equation 7 is set to 500.

The projection function and the back projection function are implemented by:

$$\textbf{Proj}(\boldsymbol{P}_i) = \boldsymbol{W}_1(\boldsymbol{P}_i) + \boldsymbol{b}_1 \tag{9}$$

$$\textbf{Proj}_b(\boldsymbol{v}_I) = \boldsymbol{W}_3(\tanh(\boldsymbol{W}_2\boldsymbol{v}_I + \boldsymbol{b}_1)) + \boldsymbol{b}_2 \tag{10}$$

The weight matrix $\boldsymbol{W}_2$ of the back projection function in Equation 10 has a size of $\mathbb{R}^{384\times d_I}$, where d_I is 100 for ImageCLEF, 150 for OfficeHome and 250 for DomainNet. Therefore, for generalizing to new target domains using LST, only 0.02M, 0.17M, and 1.47M parameters are tuned respectively.