A Unified Framework for Heterogeneous Semi-supervised Learning Supplementary Materials

Marzi Heidari*, Abdullah Alchihabi*, Hao Yan*, Yuhong Guo*†
*School of Computer Science, Carleton University, Ottawa, Canada
†Canada CIFAR AI Chair, Amii, Canada
{marziheidari@cmail., abdullahalchihabi@cmail., haoyan6@cmail., yuhong.guo@}carleton.ca

1. Training Algorithm

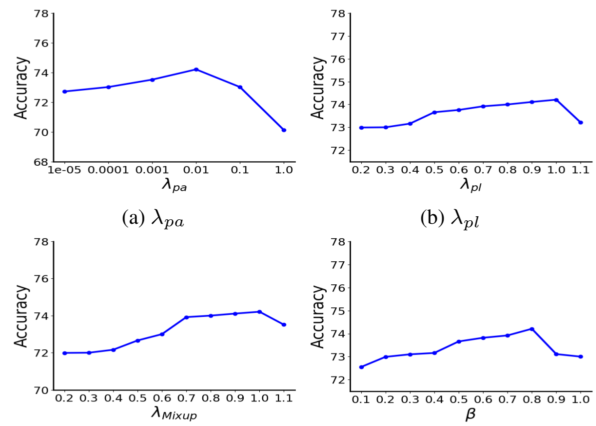
The training algorithm for the proposed method is provided in Algorithm 1.

Algorithm 1 Training Algorithm for Uni-HSSL

Input: $\mathcal{D}_L$, $\mathcal{D}_U$; initialized prediction model (f, h) and pseudo-labels $\{\hat{\mathbf{y}}_i^0\}_{i=1}^{N_u}$
Output: Trained feature encoder f, $2C$-class classifier h
for t = 1 **to** T **do**
  Compute the supervised loss $\mathcal{L}_{cl}^L$ on $\mathcal{D}_L$ using Eq.(4)
  Update the pseudo-labels $\{\hat{\mathbf{y}}_i^t\}$ on $\mathcal{D}_U$ using Eq.(5)
  Compute the classification loss $\mathcal{L}_{pl}^U$ on $\mathcal{D}_U$ using Eq.(6)
  for each semantic class $k \in \{1, \cdots, C\}$ **do**
    Compute labeled class prototype $\mathbf{p}_k$ using Eq.(7)
    Compute unlabeled prototype $\mathbf{p}_{C+k}$ using Eq.(8)
  end for
  Calculate prototype alignment loss $\mathcal{L}_{pa}$ using Eq.(9)
  Generate $\mathcal{D}_{\text{Mixup}}$ via Eq.(10)—λ sampled via Eq.(11)
  Calculate $\mathcal{L}_{\text{Mixup}}$ using Eq.(12)
  $\mathcal{L}_{total} = \mathcal{L}_{cl}^L + \lambda_{pl}\mathcal{L}_{pl}^U + \lambda_{pa}\mathcal{L}_{pa} + \lambda_{\text{Mixup}}\mathcal{L}_{\text{Mixup}}$
  Update parameters of f and h using gradient descent.
end for

2. Hyperparameter Sensitivity

We conduct sensitivity analysis for the proposed Uni-HSSL framework over four hyperparameters: the trade-off hyperparameters controlling the contribution of each loss term λ_{pa}, λ_{pl} and λ_{Mixup} and the β hyperparameter controlling the rate of update for the pseudo-labels. We conducted the experiments on Office-31 using Webcam (W) as the labeled domain and Amazon (A) as the unlabeled domain by testing a range of different values for each of the four hyperparameters independently. The obtained results are reported in Figure 1. From the figure, we can see that values either too small or too large cause performance degradation for the proposed framework for all the four hyper-parameters, while values in between yield the best results. In the case of the trade-off hyper-parameters controlling the contribution of each loss term, this highlights the importance of the balanced interplay between the loss terms to obtain the best results. As for β, very large values prevent the framework from applying larger updates to the pseudo-labels in the early iteration of the training process, while very small values lead to oscillating updates across the training iterations. A value between 0.6 and 0.8 is required to obtain the best results.



(a) λ_{pa} (b) λ_{pl} (c) λ_{Mixup} (d) β

Figure 1. Sensitivity analysis for four hyper-parameters λ_{pa}, λ_{pl}, λ_{Mixup}, and β on Office-31 using Webcam (W) as the labeled domain and Amazon (A) as the unlabeled domain.

	MEL	NV	BCC	AK	BKL	DF	VASC	SCC	Total
BCN (B)	2,857	4,206	2,809	737	1,138	124	111	431	12,413
HAM (H)	1,113	6,705	514	130	1,099	115	142	197	10,015

Table 1. The number of samples of each class in the BCN and HAM domains of the ISIC-2019 dataset.

3. ISIC dataset

The details of BCN (B) and HAM (H) domains of ISIC-2019 dataset are presented in Table 1.