Stochastic Binary Network for Universal Domain Adaptation (Supplementary Material)

The structure of the supplementary material is as follows:

1. Notations

Table 1 offers a concise overview of the notations utilized in the paper.

2. Additional experimental details

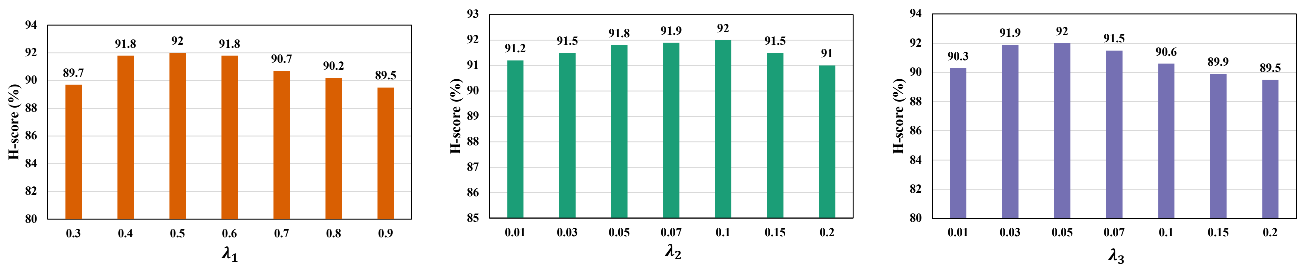
2.1. Dataset split details

The experiments were conducted in accordance with the dataset splits provided in the prior works [1, 5]. Let $\mathcal{C}$ denotes common label set between the domains, while $\overline{\mathcal{C}}_s$ and $\overline{\mathcal{C}}_t$ denotes the source private and target private label sets respectively.

In the context of the open-set domain adaptation (ODA) scenario, the dataset splits are given as follows: **(1) Office-31:** The ten common classes between Office-31 and Caltech-256 [2] are chosen as common label set ($\mathcal{C}$) and the selected 11 classes ("tape dispenser", "ring binder", "stapler", "printer", "punchers", "scissors", "ruler", "speaker", "phone", "pen" and "trash can") are chosen as target private label set ($\overline{\mathcal{C}}_t$). **(2) OfficeHome:** The initial 25 classes in alphabetical order are selected as the common label set ($\mathcal{C}$), while the remaining 40 classes constitute the target's private label set ($\overline{\mathcal{C}}_t$). **(3) VisDA:** The first six classes are selected as common label set ($\mathcal{C}$), and the remaining six classes are assigned to target private label set ($\overline{\mathcal{C}}_t$).

Table 1. Table of notations

	Symbol	Description
Network	$\mathcal{F}$	Feature extractor
	$\mathcal{D}$	Adversarial domain discriminator
	SBN	Stochastic binary network
Datasets/ Label sets	D_s	Source labeled dataset
	D_t	Target unlabeled dataset
	$\bar{D}_t$	Strongly augmented target unlabeled dataset
	$\mathcal{C}_s$	Source label set
	$\mathcal{C}_t$	Target label set
	$\mathcal{C}$	Common label set
	$\overline{\mathcal{C}}_s$	Source private label set
	$\overline{\mathcal{C}}_t$	Target private label set
Losses/factors	$\mathcal{L}_{\text{ova}}$	One-vs-all loss
	$\mathcal{L}_{\text{ent}}$	Entropy minimization loss
	$\mathcal{L}_{\text{adv}}$	Weighted adversarial learning loss
	$\mathcal{L}_{\text{cons}}$	Consistency regularization loss
	λ_1	Entropy minimization scaling factor
	λ_2	Weighted adversarial learning scaling factor
	λ_3	Consistency regularization scaling factor
Samples / Features / Miscellaneous	(x_i^s, y_i^s)	i-th labeled source sample
	x_i^t	i-th unlabeled target sample
	f_i^s	Feature of i-th source sample
	f_i^t	Feature of i-th target sample
	$\bar{f}_i^t$	Feature of i-th strongly augmented target sample
	c_i^s	Confidence score of i-th source sample
	c_i^t	Confidence score of i-th target sample
	P^t	Collective predicted probability outputs for target data
	Q^t	Collective generated auxiliary distributions for target data
	m	Number of sampled stochastic classifiers
	$\mathcal{A}$	Strong augmentation
	α	Weak augmentation
	$\mathcal{N}(\mu, \Sigma)$	Multivariate Gaussian distribution
	μ	Learnable mean vector
	Σ	Learnable diagonal covariance matrix
	τ	Confidence threshold



(a) (b) (c)

Figure 1. Hyperparameters sensitivity analysis on the Office-31 dataset under the open-set domain adaptation (ODA) scenario. (a) Effect of changing the entropy minimization scaling factor (λ_1). (b) Effect of changing the weighted adversarial learning scaling factor (λ_2). (c) Effect of changing the consistency regularization scaling factor (λ_3).

In the context of the universal domain adaptation (UniDA) scenario, the dataset splits are given as follows: **(1) Office-31:** The ten common classes between Office-31 and Caltech-256 [2] are chosen as a common label set ($\mathcal{C}$). The subsequent ten classes, listed alphabetically, compose the source private label set ($\overline{\mathcal{C}}_s$), while the remaining 11 classes constitute the target private label set ($\overline{\mathcal{C}}_t$). **(2) OfficeHome:**. The initial ten classes in alphabetical order constitute the common label set ($\mathcal{C}$). The subsequent five classes form the source private label set ($\overline{\mathcal{C}}_s$), and the remaining classes are designated as the target private label set ($\overline{\mathcal{C}}_t$). **(3) VisDA:** The first six classes are assigned to the common label set ($\mathcal{C}$), and the subsequent three classes are included in the source private label set ($\overline{\mathcal{C}}_s$). All other classes are allocated to the target private label set ($\overline{\mathcal{C}}_t$).

2.2. Evaluation metric

Following prior works [1, 5, 6], we evaluate our method using H-score. H-score is the harmonic mean of accuracy on known classes (acc$_k$) and accuracy on unknown classes (acc$_u$), and it can be written as:

$$\text{H-score} = \frac{2acc_k \cdot acc_u}{acc_k + acc_u} \quad (1)$$

The H-score metric will be high when both known and unknown accuracies are high.

2.3. Hyperparameters details

Following previous works [1,5,6], we use ResNet50 [3] pretrained on ImageNet as our feature extractor. Following [6], we set batch size as 36 and train the model for 10,000 iterations using Nesterov momentum SGD with momentum of 0.9 and weight decay of 5×10^{-4}. The initial learning rate is set as 0.01, which decays with the factor of $\left(1+\gamma\frac{i}{N}\right)^{-p}$, where i denotes the current iteration and N denotes the global iteration, and we set $\gamma = 10$ and $p = 0.75$. For the scaling factors, we empirically set the value of λ_1 and λ_2 as 0.5 and 0.1, respectively, for all datasets. The value of λ_3 is set to 0.05 for Office-31 and OfficeHome, while for the large-scale VisDA dataset, the value of λ_3 is set to 0.1 because of the presence of more fragmented distributions.

3. Additional results

3.1. Hyperparameters sensitivity analysis

To demonstrate the sensitivity of STUN to variations in the scaling factors (λ_1, λ_2 and λ_3), we conducted experiments on the Office-31 dataset using the ODA (open-set Domain Adaptation) setting.

Fig. 1a shows the sensitivity analysis on entropy minimization scaling factor (λ_1) across a broad range from 0.3 to 0.9. The performance of STUN demonstrates minimal fluctuations, underscoring the robust nature of STUN to the value of λ_1.

Likewise, in Fig. 1b and Fig. 1c, we delve into the performance analysis of STUN on changing the values of two essential scaling factors: the weighted adversarial learning scaling factor (λ_2) and the consistency regularization scaling factor (λ_3). This examination spans a comprehensive range, from 0.01 to 0.2, demonstrating that our method is less sensitive and shows the stable performance for different choices of λ_2 and λ_3.

3.2. Results on different seeds

In Table 2, we examine the stability of our framework by reporting the averaged H-score along with the standard deviation after running the experiment three times with different random seeds for the Office-31 dataset under both ODA and UniDA settings. Our results indicate that the standard deviation values remain near zero across various adaptation tasks under both ODA and UniDA settings, demonstrating our framework STUN's reliability.

Table 2. Averaged H-score (%) and standard deviation (%) based on the three runs of our framework STUN for the Office-31 dataset under both ODA and UniDA scenarios.

Setting	A2W	A2D	W2A	W2D	D2A	D2W	Avg
ODA (10/0/11)	88.3 $\pm$ 0.4	88.2 $\pm$ 0.1	89.6 $\pm$ 0.7	99.2 $\pm$ 0.3	90.5 $\pm$ 0.4	96.4 $\pm$ 0.2	92.0 $\pm$ 0.1
UniDA (10/10/11)	83.9 $\pm$0.9	89.5 $\pm$ 0.5	89.0 $\pm$ 0.3	95.8 $\pm$ 0.4	86.1 $\pm$ 0.2	94.7 $\pm$ 0.2	89.8 $\pm$ 0.2

Table 3. Studying the importance of selection criterion in consistency regularization via deep discriminative clustering (DDC) using Office-31 and VisDA dataset under UniDA scenario. STUN* represents STUN without any selection criterion in consistency regularization.

Method	Office31 (10/10/11)							VisDA
	A2W	A2D	W2A	W2D	D2A	D2W	Avg	(6/3/3)
STUN*	80.9	89.4	82.5	95.6	71.1	94.1	85.6	52.2
STUN	**83.9**	**89.5**	**89.0**	**95.8**	**86.1**	**94.7**	**89.8**	**68.3**

3.3. Importance of selection criterion in consistency regularization via DDC

We introduce the selection criterion (i.e. $\mathrm{I}\left[c_i^t \geq \tau\right]$) in consistency regularization via deep discriminative clustering (DDC) to exclude the target private samples in loss calculation because of their potential label inconsistency in neighboring data [4]. We have verified its effectiveness by removing it from the Eq. (8) of the main paper. As can be seen from Table 3, it causes a performance drop across all adaptation scenarios. This decline is especially prominent in challenging adaptation scenarios such as W2A and D2A within the Office31 and large-scale VisDA datasets. These findings demonstrate the significance of the selection criterion within consistency regularization and again validate the capability of our robust confidence scores in efficiently separating common class samples from private class samples.

References

[1] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *European Conference on Computer Vision*, 2020. 1, 2

[2] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. 2007. 1, 2

[3] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2015. 2

[4] Rundong He, Zhongyi Han, Xiankai Lu, and Yilong Yin. Ronf: reliable outlier synthesis under noisy feature space for out-of-distribution detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 4242–4251, 2022. 3

[5] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9752–9761, 2021. 1, 2

[6] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8980–8989, 2021. 2