Generative Alignment of Posterior Probabilities for Source-free Domain Adaptation - Supplementary

Sachin Chhabra
Arizona State University,
Tempe, AZ, USA
schhabr6@asu.edu

Hemanth Venkateswara
Georgia State University,
Atlanta, GA, USA
hvenkateswara@gsu.edu

Baoxin Li
Arizona State University,
Tempe, AZ, USA
baoxin.li@asu.edu

1. High-Resolution Figure

We show the high resolution tsne plot for Office31 - Amazon $\rightarrow$Webcam (Fig. 5) in source-present setting in Fig. 1.

2. Class prior distribution experiments

We perform experiments to check the impact of different class prior distributions (using a source-prior to generate labels for the source replicator instead of a uniform distribution). The results are shown in Table 1 and Table 2 for Office-31 and VisDa datasets respectively. As shown in the result tables, using a uniform prior yields slightly higher performance than using the source prior distribution.

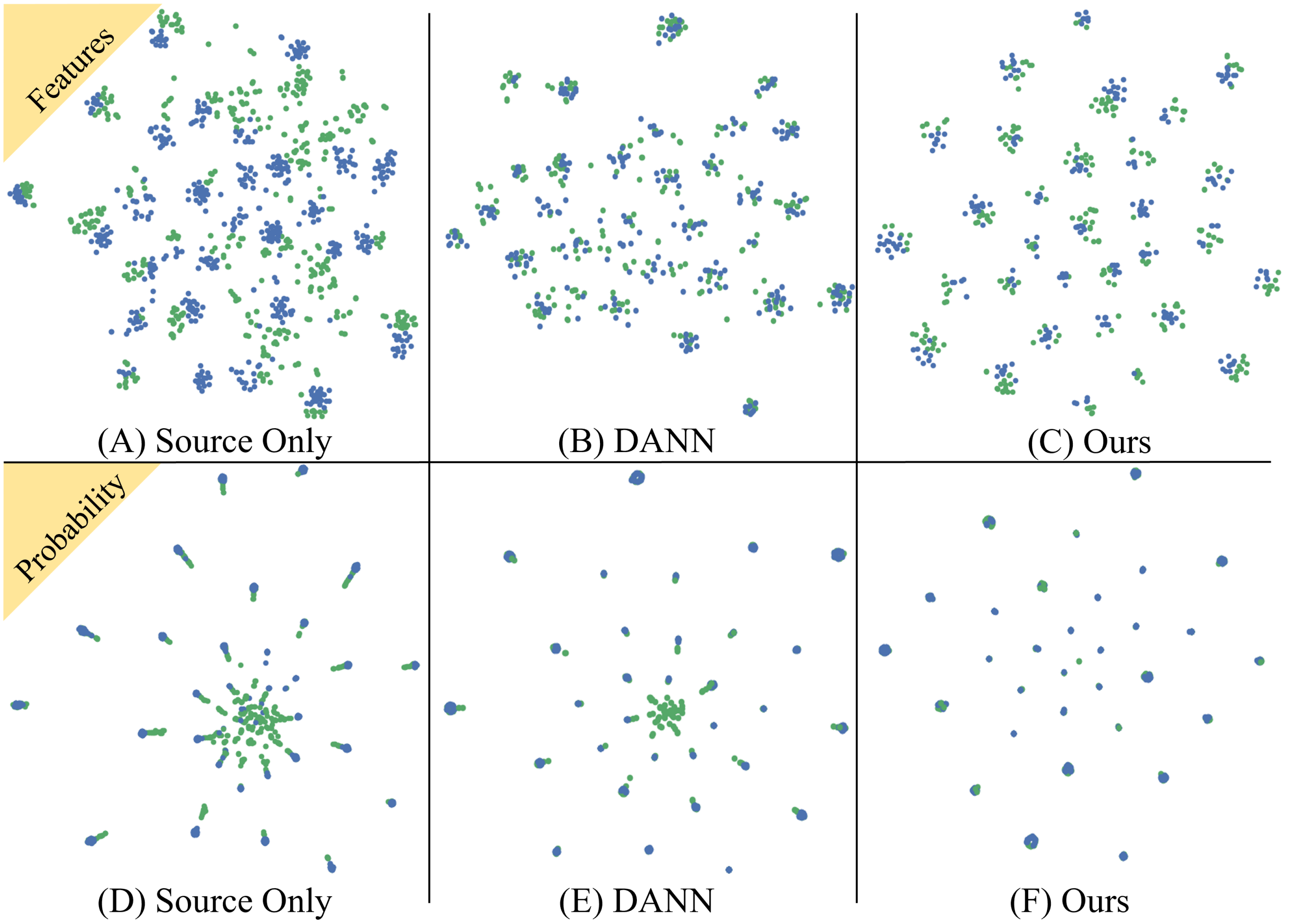

Figure 1. t-SNE visualization of the penultimate layer features and the the output probability space for Office31 - Amazon → Webcam with source present setting. Blue denotes Source domain (Amazon) and Green denotes Target domain (Webcam). Best viewed in color.

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg
Source + LS	80.8	76.9	60.3	95.3	63.6	98.7	79.3
GAP with uniform prior	90.6	90.9	74.5	98.7	73.9	99.8	88.1
GAP with source prior	88.3	89.6	74.9	96.4	73.8	99.8	87.1

Table 1. Comparison of class prior sampling for source-free classification accuracies on the Office-31 dataset (ResNet-50). LS stands for label smoothing.

Method	Plane	Bcycl	Bus	Car	Horse	Knife	Mcycl	Person	Plant	Sktbrd	Train	Truck	Avg
Source + LS	60.9	21.6	50.9	7.6	65.8	6.3	82.2	23.2	57.3	30.6	84.6	8.0	46.6
GAP with uniform prior	94.2	84.8	82.5	57.2	93.8	95.1	86.5	78.2	83.1	87.8	86.3	53.5	81.9
GAP with source prior	94.4	76.4	85.8	58.1	92.9	95.2	89.9	78.2	82.3	87.4	88.3	41.3	80.8

Table 2. Comparison of class prior sampling source-free classification accuracies on on the VisDA dataset (ResNet-101). LS stands for label smoothing.

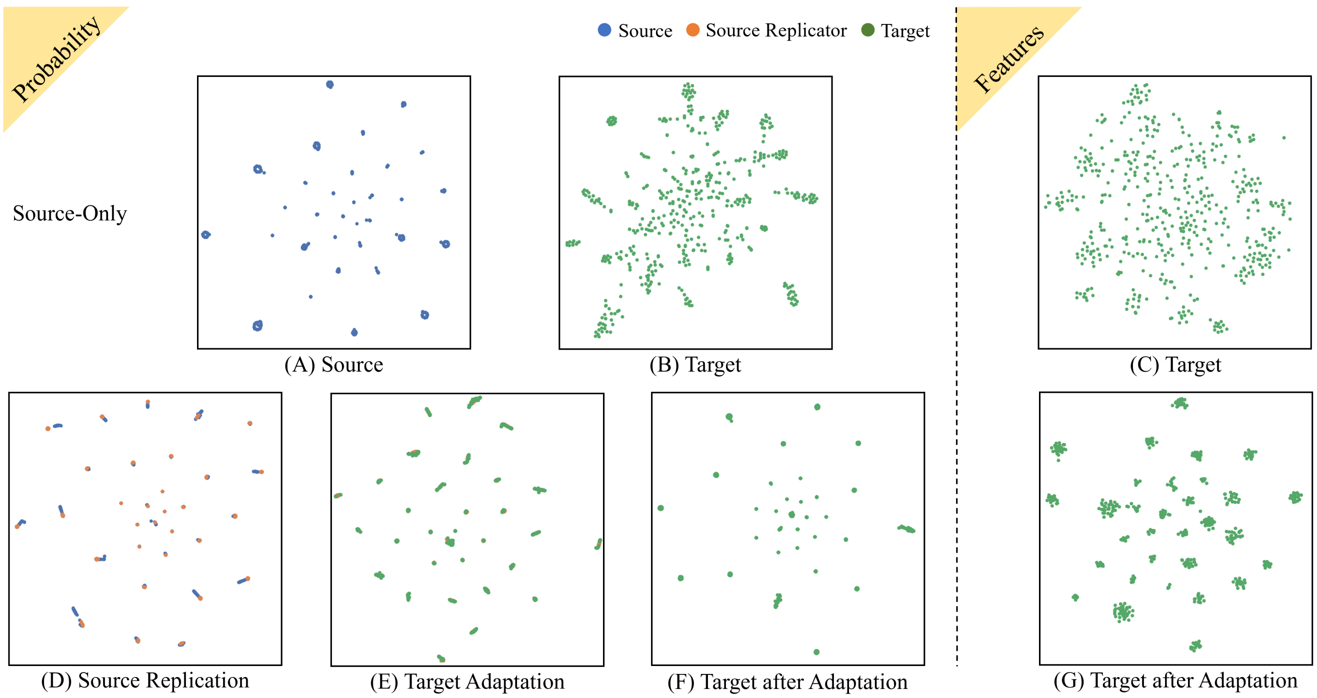

Figure 2. t-SNE visualization for Office31 - DSLR → Amazon in source-free setting. Different colors represents different domains: Blue: Source; Orange: Source Replicator; Green: Target. The probability distributions are showcased on the left section and the penultimate layer features on the right. The first row is before adaptation and the second row is our approach. (A) Source probability distribution on source trained model. (B) Target probability distribution on source trained model. (C) Target features using source trained model. (D) Source and Source Replicator probability distribution. (E) Source Replicator and Target probability distribution after adaptation. (F) Target probability distribution after adaptation. (G) Target features after adaptation. Best viewed in color.

3. t-SNE Visualization

Similar to section 5.3, we show t-SNE plots for visualizing the output probability space and the penultimate layer features for Office-31 - DSLR → Amazon and SVHN → MNIST tasks in Fig. 2 and Fig. 3 respectively. We observe that GAP clusters the penultimate layer features as well as the probabilities, similar to that observed for the Visda dataset.

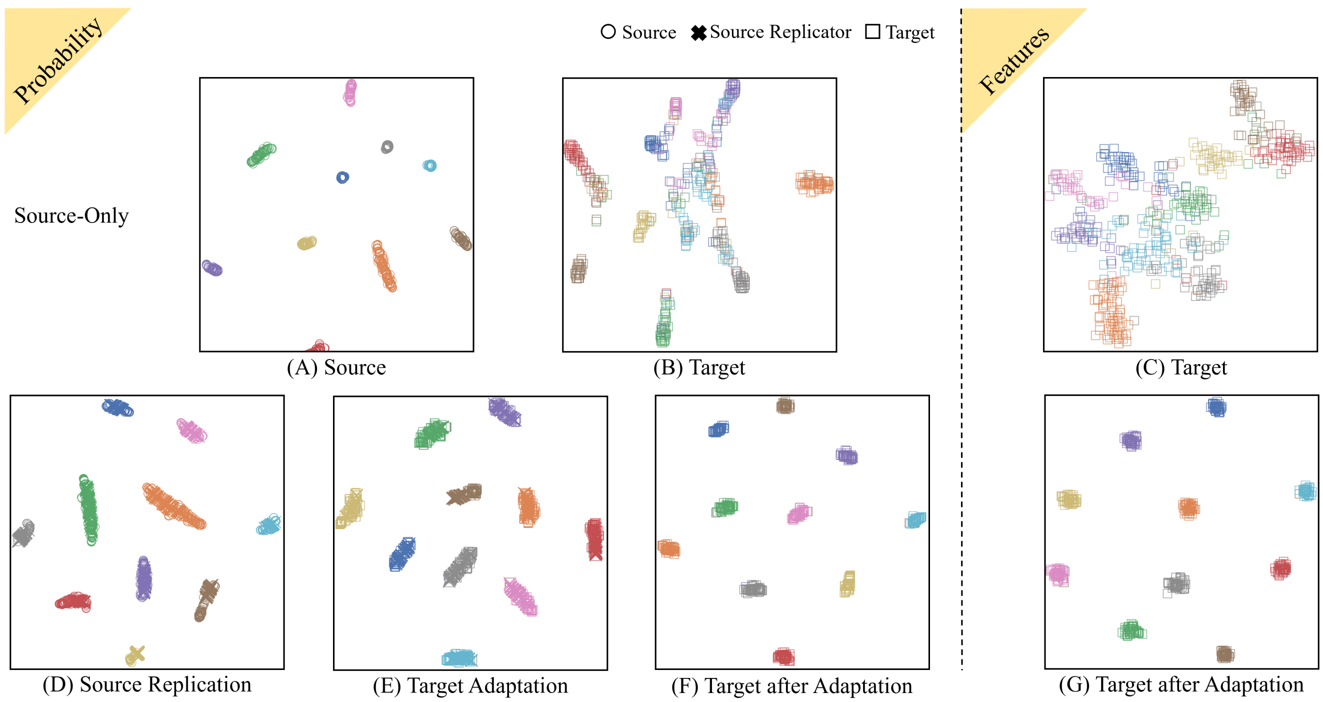

Figure 3. t-SNE visualization for digits in a source-free setting. Different colors represent different classes and different shapes represent different domains: Hollow Circle: Source; Cross: Source Replicator; Hollow Square: Target. The probability distributions are showcased on the left section and the penultimate layer features on the right. The first row is before adaptation and the second row is our approach. (A) Source probability distribution on source trained model. (B) Target probability distribution on source trained model. (C) Target features using source trained model. (D) Source and Source Replicator probability distribution. (E) Source Replicator and Target probability distribution after adaptation. (F) Target probability distribution after adaptation. (G) Target features after adaptation. Best viewed in color.

Method	A→D	A→W	D→A	D→W	W→A	W→D	Mean
Source	68.9	68.4	62.5	96.7	60.7	99.4	76.1
Source with LS	80.8	76.9	60.3	95.3	63.6	98.7	79.3
GAP without source replicator	89.2	91.4	73.9	98.8	73.9	100.0	87.9
GAP with source replicator	90.6	90.9	74.5	98.7	73.9	99.8	88.1

Table 3. Comparison of source-free classification accuracies on the Office-31 dataset (ResNet-50). Bold numbers represent the highest accuracy and the underline denotes the second highest. LS stands for label smoothing.

Source	Ar			Cl			Pr			Rw			Mean
Target	Cl	Pr	Rw	Ar	Pr	Rw	Ar	Cl	Rw	Ar	Cl	Pr	
Source	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
Source with LS	44.6	67.3	74.8	52.7	62.7	64.8	53.0	40.6	73.2	65.3	45.4	78.0	60.2
GAP without source replicator	55.4	75.8	81.3	67.4	73.7	75.9	67.0	54.6	82.3	74.8	57.9	82.6	70.7
GAP with source replicator	55.4	73.4	80.8	67.2	75.5	78.3	65.5	54.0	82.4	74.3	59.4	84.0	70.8

Table 4. Comparison of source-free classification accuracies on the Office-Home dataset (ResNet-50). Bold numbers represent the highest accuracy and the underline denotes the second highest. LS stands for label smoothing.

Method	Plane	Bcycl	Bus	Car	Horse	Knife	Mcycl	Person	Plant	Sktbrd	Train	Truck	Mean
Source	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
Source with LS	60.9	21.6	50.9	67.6	65.8	6.3	82.2	23.2	57.3	30.6	84.6	8.0	46.6
GAP without Source Replicator	94.5	80.3	80.3	70.5	93.5	97.8	87.4	79.4	88.7	83.3	89.5	29.1	81.2
GAP with Source Replicator	94.2	84.8	82.5	57.2	93.8	95.1	86.5	78.2	83.1	87.8	86.3	53.5	81.9

Table 5. Comparison of source-free classification accuracies on on the VisDA dataset (ResNet-101). Bold numbers represent the highest accuracy and the underline denotes the second highest. LS stands for label smoothing.

4. The case of no Source Replicator

We explored and evaluated the case when it is not possible to train a GAN to model the source variations. In this situation, we model $\tilde{\mathcal{Y}}$ synthetic probabilities that would replicate source variations in an ideal scenario. First, we add label smoothening $\epsilon = 0.1$ to the fake label's one-hot vectors $\boldsymbol{y}_f$ as $(1-\epsilon).\boldsymbol{y}_f + \frac{\epsilon}{K}$ where K is the number of classes. Next, we introduce a jitter to these probabilities for randomness. The jitter magnitude is sampled δ from a uniform distribution between $-\gamma$ and γ. The δ is multiplied with a random probability vector $\boldsymbol{u} \in \mathbb{R}^K$ with $\sum_{k=1}^{K} u_k = 1$ and $0 \leq u_k \leq 1$ and combined with the $\tilde{\mathcal{Y}}$ as $(1-\epsilon).\boldsymbol{y}_f - \delta.\boldsymbol{y}_f + \frac{\epsilon}{K} + \frac{\delta}{K}.\boldsymbol{u}$. Jitter randomly increases/decreases the magnitude of the fake class by δ and adjusts it with the other classes. The sythetic source probabilites are sampled as,

$$\tilde{\mathcal{Y}} := \{\tilde{\boldsymbol{y}} | \tilde{\boldsymbol{y}} = (1-(\epsilon+\delta)).\boldsymbol{y}_f + \frac{\epsilon+\delta}{K}.\boldsymbol{u}\} \qquad (1)$$

We use these probability vectors to align with the target labels using eq. 3 (main text). This equation introduces γ as a hyper-parameter and is set to 0.03 for the experiments. We test this approach on Office31, OfficeHome, and Visda datasets. The results are in Table 3, 4 and 5 respectively. Approach skips the source replicator training and yet achieves comparable performance at the cost of a hyper-parameter.