Supplementary Materials:
Source-Free Domain Adaptation via Distribution Estimation

Ning Ding[1], Yixing Xu[2], Yehui Tang[1,2], Chao Xu[1], Yunhe Wang[2*], Dacheng Tao[3]

[1] Key Lab of Machine Perception (MOE), School of Artificial Intelligence, Peking University

[2] Huawei Noah's Ark Lab [3] JD Explore Academy, China

dingning@stu.pku.edu.cn, {yixing.xu, yunhe.wang}@huawei.com, yhtang@pku.edu.cn

xuchao@cis.pku.edu.cn, dacheng.tao@gmail.com

1. L2 norm of anchors and features

In Sec. 3.3, we mention that the L2 norm of anchors and feature representations are usually different. Therefore, we combine the norm of target features and the direction of corresponding anchors to estimate the mean of surrogate source distribution, which is $\hat{\mu}_k^s = \|\bar{f}_k^t\|_2 \cdot \frac{\mathbf{w}_k^G}{\|\mathbf{w}_k^G\|_2}$, $k \in \mathcal{C}$. As shown in Fig. 1, we visualize the L2 norm of the mean of feature representations classwisely for both source and target domains, where $\bar{f}_k^s = \frac{\sum_i f_{i,k}^s}{\sum_{x_i^s \in \mathcal{D}_s} \mathbb{1}(y_i^s = k)}$ and $\bar{f}_k^t = \frac{\sum_i f_{i,k}^t}{\sum_{x_i^t \in \mathcal{D}_t'} \mathbb{1}(\hat{y}_i^t = k)}$. Meanwhile, the L2 norm of source anchors $\mathbf{w}_k^G$ is also shown in Fig. 1. It can be clearly seen that the gap between the norm of anchors and features is huge.

Besides, the norm of the bias term b_k^G of the classifier $\mathbf{G}$ is even smaller, specifically $\|\mathrm{b}_k^G\|_2 < 2\text{e-}3$, so it's reasonable to omit this term in Eq. 1. The general form of classifier function is $\hat{y}_i = \mathbf{G}(f_i) = \arg\max_k f_i^\top \mathbf{w}_k^G + \mathrm{b}_k^G$, $k \in \mathcal{C}$.

2. Avoiding overlaps of surrogate distributions

We use multiple $\hat{\mu}_k^s$ with different L2 norms to construct class-conditioned surrogate source distributions $\mathcal{N}_k^{sur}(\hat{\mu}_k^s, \hat{\Sigma}_k^s)$, and then sample surrogate features from different distributions. As shown in Fig. 2, surrogate features of different classes start to overlap as the L2 norm of $\hat{\mu}_k^s$ becomes smaller. This is the reason for the severe performance drop shown in Tab. 5 of Sec. 4.3.

3. Reversing surrogate features to image space

To clearly demonstrate the diverse semantic information contained by surrogate features derived from SDE, we reversely map the class-conditioned surrogate features back to the image space and show the results in Fig. 3. Since there is no closed-form inverse function for the convolutional feature extractor $f = \mathbf{F}(x)$, we adopt the reverse mapping algorithm proposed in [2] by utilizing a BigGAN [1] pretrained on ImageNet. Specifically, given a fixed pretrained generator $\mathcal{G}$, the class-conditioned image derived from a

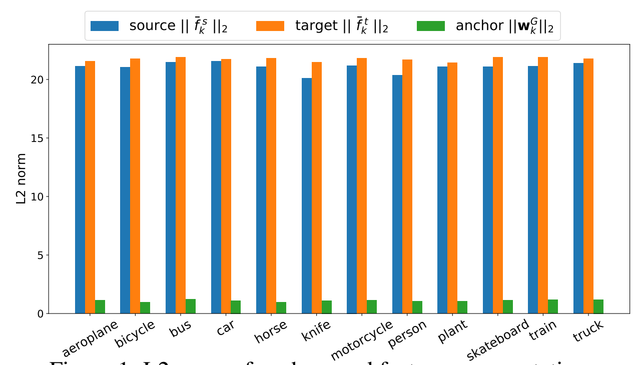


Figure 1. L2 norm of anchors and feature representations.

random noise vector z is denoted as $x = \mathcal{G}(z|k)$, where k is the class label. If the feature representation of a generated image is identical to a given surrogate feature, then this generated image can represent the reversely mapped image of the given surrogate feature. Thus, we can find the desired image x_i corresponding to a given surrogate feature $f_{i,k}^{sur} \sim \mathcal{N}_k^{sur}$ by solving for a specific noise vector

$$z^* = \arg\min_z \| \mathbf{F}(\mathcal{G}(z|k)) - f_{i,k}^{sur} \|_2 ,$$

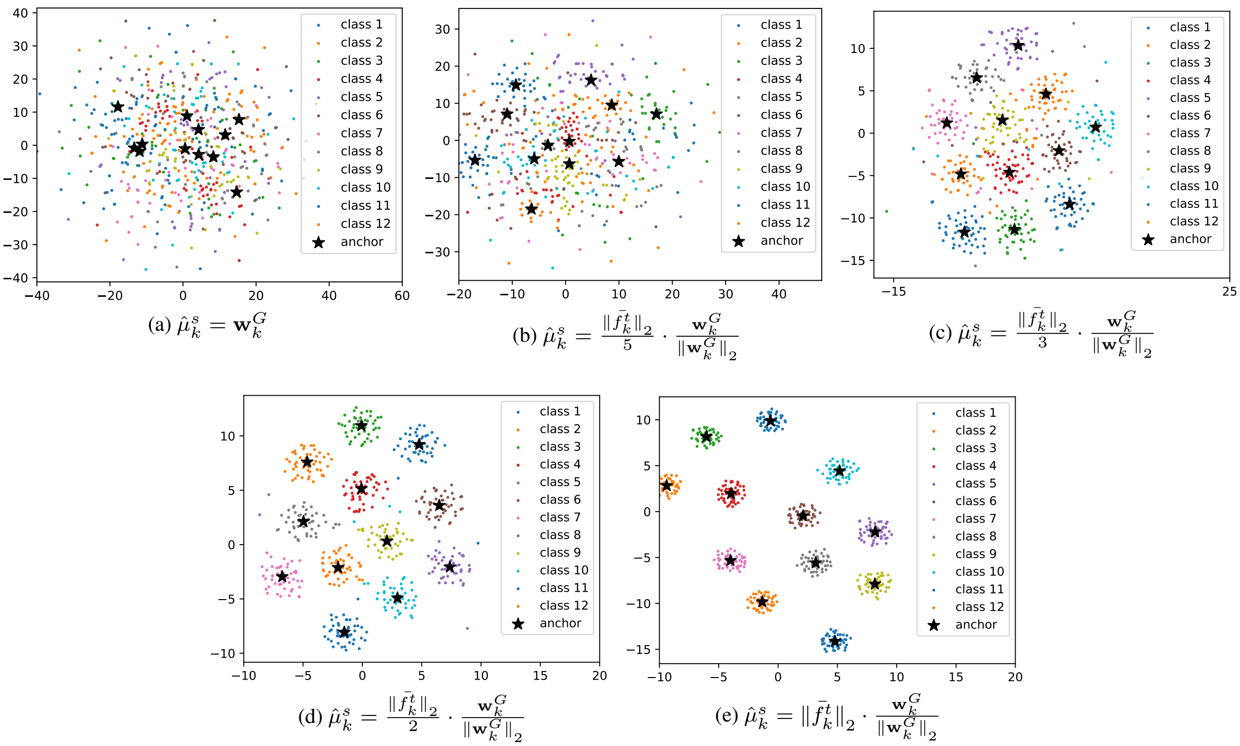
and $x_i = \mathcal{G}(z^*|k)$ is the mapped image. By mapping features back to the image space, we can directly observe the intra-class semantic richness of sampled surrogate features.

We use the standard gradient descent to solve the above equation and choose the **R**eal-**w**orld domain of Office-Home dataset to conduct reverse mapping and visualization, since the images in **Rw** domain are similar to those of the ImageNet dataset.

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *ICLR 2019*. 1

[2] Yulin Wang, Gao Huang, Shiji Song, Xuran Pan, Yitong Xia, and Cheng Wu. Regularizing deep networks with semantic data augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1



(a) $\hat{\mu}_k^s = \mathbf{w}_k^G$

(b) $\hat{\mu}_k^s = \frac{\|\bar{f}_k^t\|_2}{5} \cdot \frac{\mathbf{w}_k^G}{\|\mathbf{w}_k^G\|_2}$

(c) $\hat{\mu}_k^s = \frac{\|\bar{f}_k^t\|_2}{3} \cdot \frac{\mathbf{w}_k^G}{\|\mathbf{w}_k^G\|_2}$

(d) $\hat{\mu}_k^s = \frac{\|\bar{f}_k^t\|_2}{2} \cdot \frac{\mathbf{w}_k^G}{\|\mathbf{w}_k^G\|_2}$

(e) $\hat{\mu}_k^s = \|\bar{f}_k^t\|_2 \cdot \frac{\mathbf{w}_k^G}{\|\mathbf{w}_k^G\|_2}$

Figure 2. Features sampled from surrogate source distributions with different $\hat{\mu}_k^s$.

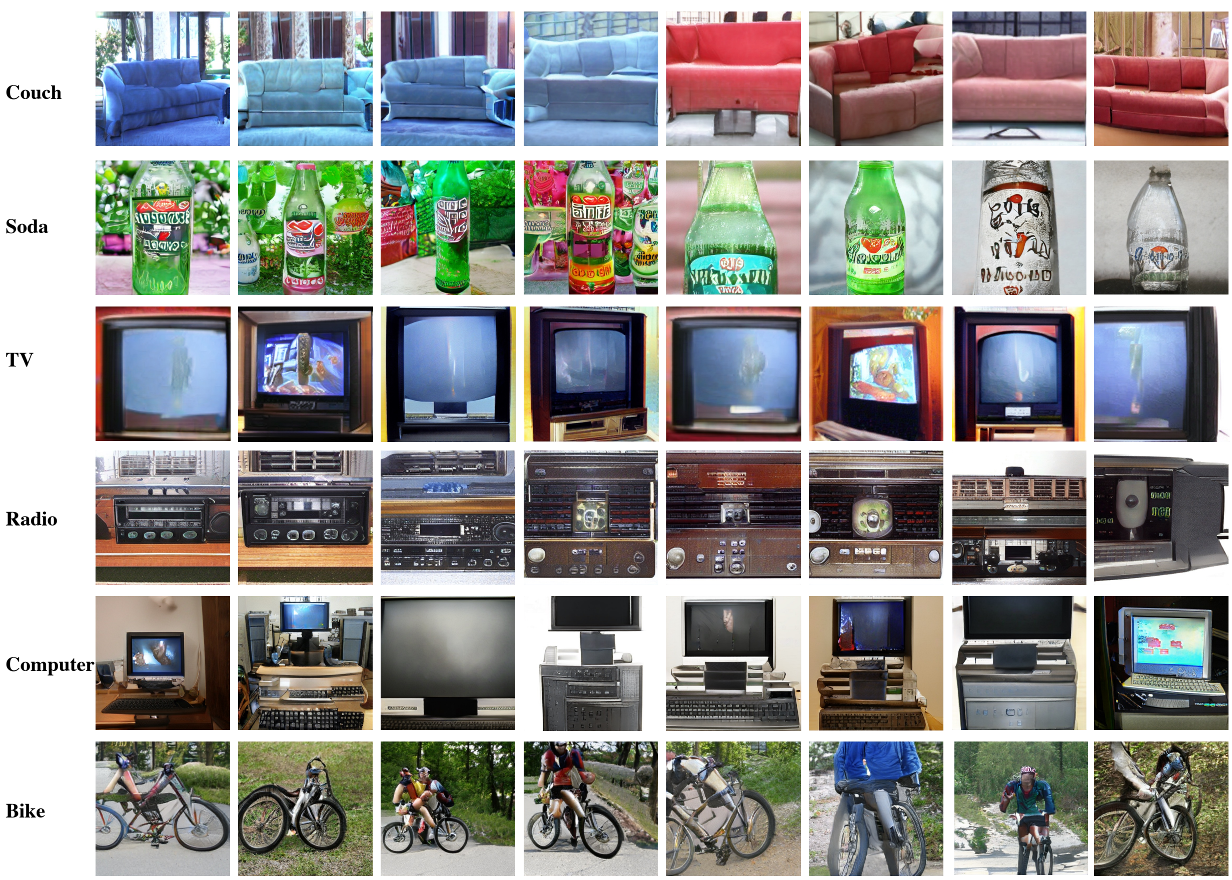


Figure 3. Visualization result of class-conditioned surrogate features reversed to image space on **Rw** domain of Office-Home dataset.