# Supplementary Material: Contrastively Smoothed Class Alignment for Unsupervised Domain Adaptation

Shuyang Dai[1], Yu Cheng[2], Yizhe Zhang[3], Zhe Gan[2],
Jingjing Liu[2], and Lawrence Carin[1]

[1] Duke University, [2] Microsoft Dynamics 365 AI Research, [3] Microsoft Research
{shuyang.dai,lcarin}@duke.edu
{yu.cheng,yizhe.zhang,zhe.gan,jingjl}@microsoft.com

## S1 Model Architecture

The model architecture for the image datasets is listed in the following.

| Feature Classifier | Feature Generator |
|---|---|
| Input feature $G(X)$ | Input $X$ |
| MLP output $F(G(X))$ with shape 10 | $3 \times 3$ conv. 64 (96) lReLU, stride 1<br>$3 \times 3$ conv. 64 (96) lReLU, stride 1<br>$3 \times 3$ conv. 64 (96) lReLU, stride 1<br>$2 \times 2$ max pool, stride 2, dropout, $p = 0.5$, Gaussian noise, $\sigma = 1$<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>$2 \times 2$ max pool, stride 2, dropout, $p = 0.5$, Gaussian noise, $\sigma = 1$<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>$3 \times 3$ conv. 64 (192) lReLU, stride 1<br>global average pool, output feature $G(X)$ with shape 64 (192) |

**Table S1.** Model architecture for the visual domain adaptation experiments. Numbers in the $(\cdot)$ are for CIFAR→STL and STL→CIFAR.

## S2 Hyper-parameter Setup

The hyper-parameter setups for both the visual and non-visual datasets are listed in the following.

## S3 Additional Experimental Results

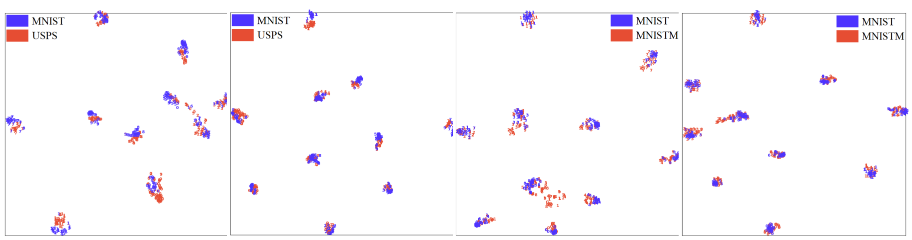
For fair comparison, the results on VisDA dataset in the main paper is reported based on ResNet101. Some of the results are reported based on ResNet152 originally, and therefore, we include them in Table S3 as follows.

| Task | $\lambda_1$ for $\mathcal{L}_{\text{MMD}}$ | $\lambda_2$ for $\mathcal{L}_{\text{adv}}$ | $\lambda_3$ for $\mathcal{L}_{\text{contras}}$ |
|---|---|---|---|
| Digits | 10.0 | 0.1 | 0.2 |
| CIFAR→STL | 5.0 | 0.1 | 0.4 |
| STL→CIFAR | 5.0 | 0.1 | 0.2 |
| Amazon Reviews | 4.0 | 0.1 | 0.8 |

**Table S2.** Hyper-parameter setup for visual and non-visual domain adaptation experiments.

| Model | plane | bcycl | bus | car | horse | knife | mcycl | person | plant | sktbrd | train | truck | mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CAN [31] | 94.5 | 76.3 | 82.2 | **71.1** | 94.3 | 86.2 | 88.3 | 81.2 | 91.4 | 89.2 | 87.3 | 50.3 | 82.7 |
| SEDA [38] | 95.9 | 87.4 | 85.2 | 58.6 | **96.2** | **95.6** | 90.6 | 80.0 | 94.8 | 90.8 | **88.4** | **47.9** | 84.3 |
| CoSCA | **96.3** | **87.9** | **86.1** | 69.8 | 95.9 | 93.7 | **91.2** | **84.1** | **95.1** | **90.9** | 86.3 | 45.8 | **85.3** |

**Table S3.** VisDA validation set results using a ResNet152 model.




(a) MCD M→U (b) CoSCA M→U (c) MCD M→MM (d) CoSCA M→MM

**Fig. S1.** t-SNE embedding of the features $G(x)$ for MNIST (M) →USPS (U), and MNIST (M) →MNISTM (MM). Color indicates domain, and the digit number is the label. The ideal situation is to mix the two colors with the same label, representing domain-invariant features.

We also include t-SNE plots for other benchmark datasets in the following. Figure S1 compares MCD and CoSCA on MNIST→USPS and MNIST→MNISTM, showing that CoSCA provides improvement over MCD.