

Table 1: **MNIST-USPS-SVHN datasets.** Classification accuracy for domain adaptation over the MNIST, USPS, and SVHN datasets. \mathcal{M} , \mathcal{U} , and \mathcal{S} stand for MNIST, USPS, and SVHN domain. LB is our base model without adaptation. FT and FADA stand for fine-tuning and our method, respectively.

	LB	Traditional UDA			Adversarial UDA			SDA	1	2	3	4	5	6	7
		[60]	[45]	[15]	[33]	[59]	[49]								
$\mathcal{M} \rightarrow \mathcal{U}$	65.4	47.8	60.7	91.8	91.2	89.4	92.5	FT	82.3	84.9	85.7	86.5	87.2	88.4	88.6
								[38]	85.0	89.0	90.1	91.4	92.4	93.0	92.9
								FADA	89.1	91.3	91.9	93.3	93.4	94.0	94.4
$\mathcal{U} \rightarrow \mathcal{M}$	58.6	63.1	67.3	73.7	89.1	90.1	90.8	FT	72.6	78.2	81.9	83.1	83.4	83.6	84.0
								[38]	78.4	82.2	85.8	86.1	88.8	89.6	89.4
								FADA	81.1	84.2	87.5	89.9	91.1	91.2	91.5
$\mathcal{S} \rightarrow \mathcal{M}$	60.1	-	-	82.0	76.0	-	84.7	FT	65.5	68.6	70.7	73.3	74.5	74.6	75.4
								FADA	72.8	81.8	82.6	85.1	86.1	86.8	87.2
$\mathcal{M} \rightarrow \mathcal{S}$	20.3	-	-	40.1	-	-	36.4	FT	29.7	31.2	36.1	36.7	38.1	38.3	39.1
								FADA	37.7	40.5	42.9	46.3	46.1	46.8	47.0
$\mathcal{S} \rightarrow \mathcal{U}$	66.0	-	-	-	-	-	-	FT	69.4	71.8	74.3	76.2	78.1	77.9	78.9
								FADA	78.3	83.2	85.2	85.7	86.2	87.1	87.5
$\mathcal{U} \rightarrow \mathcal{S}$	15.3	-	-	-	-	-	-	FT	19.9	22.2	22.8	24.6	25.4	25.4	25.6
								FADA	27.5	29.8	34.5	36.0	37.9	41.3	42.9

feature space. UDA only looks for domain confusion and does not address class separability, because of the lack of labeled target samples.

Connection with conditional GANs: Concatenation of outputs of different inferences has been done before in conditional GANs. For example, [43, 44, 64] concatenate the input text to the penultimate layers of the discriminators. [25] concatenates positive and negative pairs before passing them to the discriminator. However, all of them use the vanilla binary discriminator.

Relationship between g_s and g_t : There is no restriction for g_s and g_t and they can be constrained or unconstrained. An obvious choice of constraint is equality (weight-sharing) which makes the inference functions symmetric. This can be seen as a regularizer and will reduce overfitting [38]. Another approach would be learning an asymmetric inference function [45]. Since we have access to very few target samples, we use weight-sharing ($g_s = g_t = g$).

Choice of g_s , g_t , and h : Since we are interested in visual recognition, the inference functions g_s and g_t are modeled by a convolutional neural network (CNN) with some initial convolutional layers, followed by some fully connected layers which are described specifically in the experiments section. In addition, the prediction function h is modeled by fully connected layers with a softmax activation function for the last layer.

Training Process: Here we discuss the training process for the weight-sharing regularizer ($g_s = g_t = g$). Once the inference functions g and the prediction function h are chosen, FADA takes the following steps: First, g and h are initialized using the source dataset \mathcal{D}_s . Then, the mentioned four groups of pairs should be created using \mathcal{D}_s and \mathcal{D}_t . The next step is training DCD using the four groups of pairs. This should be done by freezing g . In the next step, the inference function g and prediction function h should be updated in order to confuse DCD and maintain high classification accuracy. This should be done by freezing DCD. See Algorithm 1 and Figure 2. The training process for the non weight-sharing case can be derived similarly.

4 Experiments

We present results using the Office dataset [47], the MNIST dataset [32], the USPS dataset [24], and the SVHN dataset [40].

4.1 MNIST-USPS-SVHN Datasets

The MNIST (\mathcal{M}), USPS (\mathcal{U}), and SVHN (\mathcal{S}) datasets have recently been used for domain adaptation [12, 45, 59]. They contain images of digits from 0 to 9 in various different environments including in the wild in the case of SVHN [40]. We considered six cross-domain tasks. The first two tasks include $\mathcal{M} \rightarrow \mathcal{U}$, $\mathcal{U} \rightarrow \mathcal{M}$, and followed the experimental setting in [12, 45, 33, 59, 49], which involves randomly selecting 2000 images from MNIST and 1800 images from USPS. For the rest of