Consensus Adversarial Domain Adaptation

Han Zou,[1*] **Yuxun Zhou,**[1] **Jianfei Yang,**[2] **Huihan Liu,**[1] **Hari Prasanna Das,**[1] **Costas J. Spanos**[1]
[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA
[2]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore
{hanzou, yxzhou, liuhh, hpdas, spanos}@berkeley.edu, yang0478@ntu.edu.sg

Abstract

We propose a novel domain adaptation framework, namely Consensus Adversarial Domain Adaptation (CADA), that gives freedom to both target encoder and source encoder to embed data from both domains into a common domain-invariant feature space until they achieve consensus during adversarial learning. In this manner, the domain discrepancy can be further minimized in the embedded space, yielding more generalizable representations. The framework is also extended to establish a new few-shot domain adaptation scheme (F-CADA), that remarkably enhances the ADA performance by efficiently propagating a few labeled data once available in the target domain. Extensive experiments are conducted on the task of digit recognition across multiple benchmark datasets and a real-world problem involving WiFi-enabled device-free gesture recognition under spatial dynamics. The results show the compelling performance of CADA versus the state-of-the-art unsupervised domain adaptation (UDA) and supervised domain adaptation (SDA) methods. Numerical experiments also demonstrate that F-CADA can significantly improve the adaptation performance even with sparsely labeled data in the target domain.

Introduction

In recent years, a booming development of deep learning methods has been witnessed, partially as a consequence of the availability of a large amount of labeled data to train and validate more advanced models. More often than not, these recognition models trained with the large datasets perform extremely well in one domain, i.e., the source domain. However, they often fail to generalize well to new datasets or new environments, i.e., the target domain, due to domain shift or dataset bias (Tzeng et al. 2017).

To alleviate the issue of domain shift, a large body of research has been carried out on domain adaptation, which aims to distill the shared knowledge across domains and therefore improving the generalization of the learned model. Domain adaptation methods can be categorized into 2 classes, unsupervised domain adaptation (UDA) and supervised domain adaptation (SDA), depending on whether labeled data is available in the target domain. In many real-world cases, it is time-consuming, labor-intensive and expensive to collect and annotate a huge number of samples in the target domain. Thus, in practice, research on UDA is arguably more popular than SDA since collecting unlabeled target data is usually a trivial task. Conventional UDA methods (e.g., DDC (Tzeng et al. 2014), RevGrad (Ganin and Lempitsky 2015) and DRCN (Ghifary et al. 2016)) map data from both domains into a common feature space to reduce the domain shift. This is achieved by minimizing some measure of distance between the target and source feature distributions, e.g., correlation distances or maximum mean discrepancy. The goal is to identify a feature space in which samples from both target and source domain are indistinguishable. Once the task is accomplished, the model constructed in the source domain can be applied to the tasks in the target domain by embedding the dataset with the learned transformation.

Meanwhile, with the unprecedented success of Generative Adversarial Network (GAN) (Goodfellow et al. 2014), some researchers have proposed to construct an adversarial loss to accommodate the domain shift, which is commonly referred as adversarial domain adaptation (ADA) or adversarial UDA (Tzeng et al. 2017). GAN trains a generator and a discriminator in a min-max fashion, where the generator learns to generate high-quality data to fool the discriminator, and the discriminator aims to distinguish the real and synthetic data. Similar to the setup of GAN, ADA aims to minimize an approximate domain discrepancy distance through an adversarial objective with respect to a domain discriminator. Through adversarial learning, it trains a source encoder and a target encoder such that a well-formed domain discriminator cannot determine the domain label of the encoded samples. Adversarial UDA methods, e.g., CoGAN (Liu and Tuzel 2016) and ADDA (Tzeng et al. 2017), achieve appealing performance compared to traditional UDA methods. However, the feature mapping is usually defined by the source encoder in these methods. More specifically, previous methods align embedded feature representations of the target domain to the source domain by fixing parameters of the source encoder during adversarial learning. Additionally, their network settings follow that of

*Han Zou is the corresponding author.

GAN's exactly, where the real image distribution remains fixed and synthesized one is learned to match it. Similar to GAN, the source feature representation is considered as an absolute good reference for the target, leading to the stagnated parameters of the source encoder. However, in practice, the domain discrepancy between source and target is considerably significant. This assumption, therefore, might not always hold and the target data cannot be completely embedded into the imposed representation space. Both of the aforementioned concerns would result in sub-optimal adaptation, particularly when the target representation is far from source features in the latent space or the source encoder already exhibits over-fitting.

In this paper, we propose **C**onsensus **A**dversarial **D**omain **A**daptation (CADA), a novel unsupervised ADA scheme that gives freedom to both target encoder and source encoder. As such, they can achieve consensus and transform data from both domains into a general domain-invariant feature space to further accommodate the domain discrepancy and avoid over-fitting models in neither domain. After obtaining a source encoder and a source classifier as a good reference in the source domain, CADA trains a target encoder and also gives freedom to the source encoder by fine-tuning it through adversarial learning. In this manner, both unlabeled target data and labeled source data are embedded to a domain-invariant feature space defined by both domains, in a way that a domain discriminator cannot distinguish the domain labels of them. The original source classifier is further fine-tuned as a shared classifier with the source dataset and the source encoder is refined via ADA. In the target domain, we employ the trained target encoder to embed the target samples into the domain-invariant feature space and infer its class using the shared classifier. In certain applications, a few labeled data samples can be collected opportunistically in the target domain. This availability is scarce in practice but is precious for model improvement. To leverage the extra information, we also propose a few-shot version of CADA (F-CADA) which exploits the prior labeled data by greedy label propagation for further performance enhancement. In a nutshell, given a well-defined metric in the latent feature space, F-CADA assigns presumptive labels to unlabeled data points in the target domain, by greedily minimizing an information entropy loss function. The target encoder is then fine-tuned and a target classifier is constructed using both the prior and presumptively labeled data. The whole process can be repeated until convergence. The class of each target test sample is inferred using the final target encoder and classifier.

The performance of CADA and F-CADA are validated to the task of digit recognition across domains on standard digit adaptation dataset (MNIST, USPS, and SVHN digits datasets) and the task of spatial adaptation for WiFi-enabled device-free gesture recognition (GR). Experimental results demonstrate that CADA achieves outstanding domain adaptation results and outperforms state-of-the-art methods on both digit adaptation and spatial adaptation for GR. For the challenging SVHN $\Rightarrow$ MNIST scenario, it improves the digit recognition accuracy from 60% to 91%. It also enhances the GR accuracy by 25% over non-adapted classifier under environmental dynamics. Moreover, F-CADA achieves further performance gain over the best few-shot ADA methods when only one labeled target sample per class is available. It validates that the proposed label learning method indeed contributes to the overall performance improvement.

Related Work

Unsupervised Domain Adaptation The performance of conventional classifiers degrade severely when the data distribution in source domain and target domain are different. Unsupervised domain adaptation (UDA) aims to reduce the difference in the feature distribution between the source and target domain to improve generalization performance without requiring any labeled data in target domain (Tzeng et al. 2017). Some metrics have been proposed to measure the domain shift between source and target domains for their difference minimization. For instance, maximum mean discrepancy is leveraged by DDC (Tzeng et al. 2014), that estimates the norm of mean difference and matches higher order statistics of the two distributions in a reproducing kernel Hilbert space. RevGrad (Ganin and Lempitsky 2015) and DRCN (Ghifary et al. 2016) treat domain invariance as a binary classification problem and maximize the classifier loss by reversing its gradients.

Adversarial Domain Adaptation Recently, with the booming development of Generative Adversarial Network (GAN) (Goodfellow et al. 2014), researchers have proposed to construct an adversarial loss to accommodate the domain shift, which is commonly referred to as adversarial domain adaptation (ADA) (Shen et al. 2018). Similar to the learning configuration of GAN, the generator of ADA aims to fool the discriminator to make the target domain samples look like the source domain ones, and the discriminator tries to identify the domain labels (source or target) instead of fake or real image in GAN. CoGAN (Liu and Tuzel 2016) trains 2 GANs to synthesize both source and target images and achieves a domain invariant feature space by tying the high-level layer parameters of the 2 GAN to solve the domain transfer problem. ADDA (Tzeng et al. 2017) learns a discriminative representation using the labels in the source domain and then a separate encoding that maps the target data to the same space using an asymmetric mapping learned through a standard GAN loss without weights sharing. A cycle-consistency loss is designed in CyCADA (Hoffman et al. 2017) to enforce both structural and semantic consistency during ADA. One major limitation for these methods is that the adversarial discriminative models focus on aligning feature embeddings of target domain to source domain defined by the source encoder. Since the parameters of source encoder are fixed during ADA, there is no freedom for the source encoder. Thus, the ADA performance is not guaranteed when the target representation is far from source features.

Supervised Domain Adaptation Though UDA achieves acceptable performance using large amounts of unlabeled target data, it still cannot deal with large covariate shift in the distributions of the samples between two datasets. In reality, it is reasonable to label only a few samples for each class in the target dataset and then supervised domain adaptation

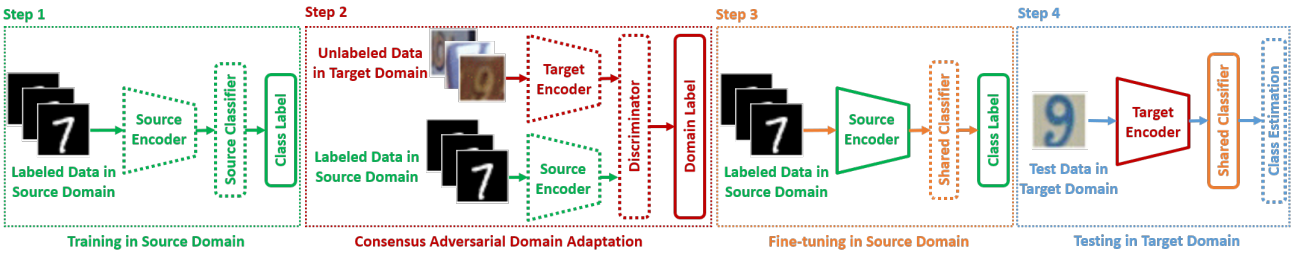


Figure 1: An overview of CADA. Step 1: A source encoder and a source classifier are trained with the labeled source data. Step 2: A target encoder is trained and the source encoder is fine-tuned through unsupervised adversarial domain adaptation to map both target data and source data to a domain-invariant feature space such that a domain discriminator cannot distinguish the domain labels of the data. Step 3: A shared classifier is constructed with the labeled source data. Step 4: During testing, the class of each target test sample is inferred using the target encoder trained in Step 2 and shared classifier obtained in Step 3. The network parameters in solid line boxes are fixed and those in dashed line boxes are trained in each step.

(SDA) can efficiently transfer the knowledge from source domain to this sparsely labeled target domain. Also, SDA does not demand large annotation overhead commonly required by standard supervised learning approaches. In (Luo et al. 2017), the authors propose a framework that learns a representation transferable across different domains and tasks in a label efficient manner. It tackles the problem of high sensitivity and overfitting during fine-tuning stage using a novel end-to-end SDA approach. Apart from feature-based SDA method, Ren et al. (2018) designed a prototypical network from the perspective of metric learning. It maps the samples into a space where the samples from the same class are close and those from different classes are far apart. Recently, few-shot learning has become attractive because only a few labeled data is required for adaptation. In domain adaptation, few-shot adversarial domain adaptation (FADA) (Motiian et al. 2017a) is proposed to transfer knowledge with only several annotated samples in each class. It exploits adversarial learning to learn an embedded subspace that simultaneously maximizes the confusion between two domains while semantically aligning their embeddings. However, they only consider a few labeled target samples but never use unlabeled target samples that are more easily obtained. In our approach, we build the few-shot domain adaptation based on semi-supervised learning, which minimizes the domain confusion via adversarial training and guides the adaptation process in a few-shot manner.

Consensus Adversarial Domain Adaptation

The objective of CADA is to improve the generalization capability of a classifier across domains without collecting labeled data in the target domain via ADA. The rationale behind CADA is to embed data from both domains into a common feature space until they achieve consensus during ADA. It is different from existing methods, which force representation alignment of the target to the source. The systematic training procedure of CADA is demonstrated in Fig. 1, which is consisted of 4 steps. The detailed methodology of each step is elaborated as follows.

Step 1: Suppose N_s samples $\mathbf{X}_s$ with labels Y_s are collected in the source domain with L possible classes. As the first step of CADA, we train a source encoder M_s and a source classifier C_s so that the source samples can be recognized with high classification accuracy. Mathematically, Step 1 solves the following minimization via back-propagation:

$$\min_{M_s, C_s} \mathcal{L}_{C_s}(\mathbf{X}_s, Y_s) = \\ - \mathbb{E}_{(\mathbf{x}_s, y_s)\sim(\mathbf{X}_s, Y_s)} \sum_{l=1}^{L} [\mathbb{I}_{[l=y_s]} \log C_s(M_s(\mathbf{x}_s))] \quad (1)$$

This is indispensable since a good baseline of the feature space and the classifier for the sub-sequent steps is needed.

Step 2: More often than not, data labeling in the target domain is a time-consuming and expensive process. On the other hand, accumulating unlabeled data in the target domain, denoted by X_t, is usually a trivial task. As the most essential step of CADA, in Step 2 we train a target encoder M_t and fine-tune the source encoder M_s such that a discriminator D cannot tell whether a sample is from the source domain or from the target domain after the associated feature mapping. In other words, after the feature embedding in the target and source domain via $M_t(X_t)$ and $M_s(X_s)$, respectively, the domain label cannot be effectively recognized by a well-formed discriminator D. This task is similar to the original GAN, that aims to generate a fake image that is indistinguishable from the real image. In our case, the labels for the discriminator D are domain labels (source and target) instead of fake and real. We formulate this step as an optimization of the following adversarial loss,

$$\min_{M_s, M_t} \max_{D} \mathcal{L}_D(\mathbf{X}_s, \mathbf{X}_t, M_s, M_t) = \\ \mathbb{E}_{\mathbf{x}_s\sim\mathbf{X}_s}[\log D(M_s(\mathbf{x}_s))] + \mathbb{E}_{\mathbf{x}_t\sim\mathbf{X}_t}[\log(1 - D(M_t(\mathbf{x}_t)))] \quad (2)$$

The GAN loss for the source encoder M_s is

$$\min_{M_s} \mathcal{L}_{M_s}(\mathbf{X}_s, \mathbf{X}_t, D) = -\mathbb{E}_{\mathbf{x}_s\sim\mathbf{X}_s}[\log D(M_s(\mathbf{x}_s))] \quad (3)$$

and the inverted label GAN loss (Goodfellow et al. 2014) is employed to train the target encoder M_t as follows,

$$\min_{M_t} \mathcal{L}_{M_t}(\mathbf{X}_s, \mathbf{X}_t, D) = -\mathbb{E}_{\mathbf{x}_t\sim\mathbf{X}_t}[\log D(M_t(\mathbf{x}_t))]. \quad (4)$$

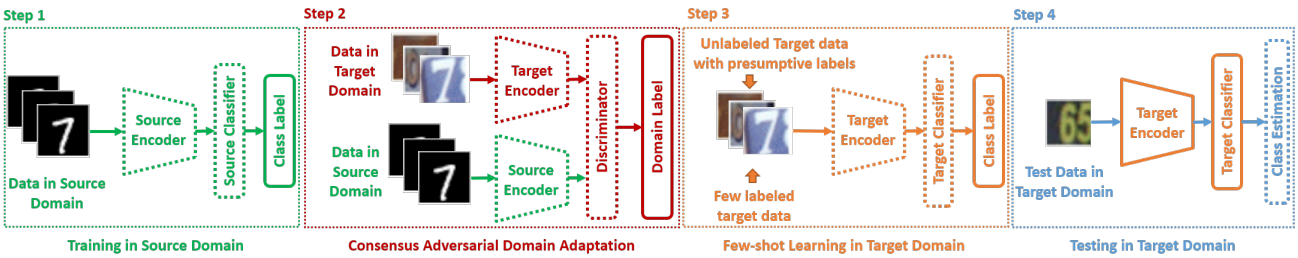


Figure 2: An overview of F-CADA. Step 1 and Step 2 are the same as CADA. Step 3: in the target domain, presumptive labels are generated for target unlabeled data with target few labeled data. Then, the target encoder is fine-tuned and a target classifier is constructed with unlabeled and few labeled target data. Step 4: During testing, the class of each target test sample is inferred using the target encoder and target classifier obtained in Step 3. The network parameters in solid line boxes are fixed and those in dashed line boxes are trained in each step.

The parameters in M_t and M_s are initialized with those of the source encoder M_s learned in the Step 1 for burn-in training. It is worth pointing out the novelty of CADA and its difference from the state-of-the-art ADA methods (e.g. ADDA (Tzeng et al. 2017) and DIFA (Volpi et al. 2017)). Under previous methods, the parameters of the source encoder are fixed during the training process of the target encoder via ADA. Consequently, the feature mapping is defined by the source encoder and ADA essentially tries to align the feature embeddings of the target domain with the source domain. In this way, the obtained source encoder is used as an absolute reference, which may deteriorate the domain adaptation performance because the alignment could be sub-optimal when the target samples cannot be completely embedded into the imposed representation space. The issue can be substantial particularly when the source and the target domain exhibit material discrepancy or the source encoder already bears some overfitting. This bottleneck is well-addressed in the proposed CADA framework, where the parameters of the source encoder are not fixed but are instead given the freedom to be fine-tuned together with the target encoder. Therefore, the feature space is defined by the consensus between M_t and M_s, yielding better generalization in both the domains.

Step 3: When the discriminator D in Step 2 is not able to identify the domain label of target samples and source samples, it is an indication that the target encoder M_t and the source encoder M_s achieved consensus by mapping the corresponding input data to a shared domain-invariant feature space. Given that, we fix the parameters of the source encoder M_s and train a shared classifier C_{sh} using the labeled source domain data $\{\mathbf{X}_s, Y_s\}$. The learning process is equivalent to minimizing the cross-entropy loss:

$$\min_{C_{sh}} \mathcal{L}_{C_{sh}}(\mathbf{X}_s, Y_s) = \\ -\mathbb{E}_{(\mathbf{x}_s,y_s)\sim(\mathbf{X}_s,Y_s)} \sum_{l=1}^{L} [\mathbb{I}_{[l=y_s]} \log C_{sh}(M_s(\mathbf{x}_s))] \quad (5)$$

The shared classifier C_{sh} can be directly used in the target domain since the target encoder M_t has embedded the target samples to the domain-invariant feature space.

Step 4: During testing in the target domain, we map the target test samples to the domain-invariant feature space through the target encoder M_t trained in Step 2, and then use the shared classifier C_{sh} obtained in Step 3 to identify category of the samples in the target domain without collecting any labeled target data.

In summary, the complete learning objective of CADA can be formulated as follows:

$$\mathcal{L}_{CADA}(\mathbf{X}_s, \mathbf{X}_t, Y_s, D, M_s, M_t) = \mathcal{L}_{C_s}(\mathbf{X}_s, Y_s) \\ + \mathcal{L}_D(\mathbf{X}_s, \mathbf{X}_t, M_s, M_t) + \mathcal{L}_{M_s}(\mathbf{X}_s, \mathbf{X}_t, D) \\ + \mathcal{L}_{M_t}(\mathbf{X}_s, \mathbf{X}_t, D) + \mathcal{L}_{C_{sh}}(\mathbf{X}_s, Y_s) \quad (6)$$

The training of CADA is thusly equivalent to solving:

$$\min_{C_{sh}} \min_{M_t, M_s} \max_{D} \min_{M_s, C_s} \mathcal{L}_{CADA}(\mathbf{X}_s, \mathbf{X}_t, Y_s, D, M_s, M_t)$$

As is illustrated in Fig. 1, we firstly train a source encoder M_s and a source classifier C_s with the labeled source data by optimizing $\mathcal{L}_{C_s}$ as described in equation (1). After that, in Step 2, we train a target encoder M_t and fine-tune the source encoder M_s via adversarial learning by optimizing $\mathcal{L}_D$, $\mathcal{L}_{M_s}$ and $\mathcal{L}_{M_t}$, i.e., equation (2)-(4). Then, a shared classifier C_{sh} is constructed in Step 3 with the labeled source domain data by optimizing $\mathcal{L}_{C_{sh}}$ as described in equation (5). During testing in Step 4, we employ the trained target encoder M_t to map the test sample from the target domain into the domain-invariant feature space and use the shared classifier C_{sh} directly to identify the category of each testing sample in the target domain.

Few-shot Consensus Adversarial Domain Adaptation (F-CADA)

A powerful extension of the CADA learning framework is to enrich it with the ability to integrate a few labeled data that may be available in the target domain for information fusion and model improvement. This task, although seems challenging in many other learning paradigms, can be achieved efficiently with F-CADA. Notation-wise, we assume that N_s samples $\mathbf{X}_s$ with labels Y_s are available in the source domain and the target domain contains N_t^u unlabeled samples $\mathbf{X}_t^u$. Additionally, a few samples, numbered N_t^l and denoted

by $\mathbf{X}_t^l$, are assumed available with associated labels Y_t^l in the target domain. To conform with the scenario of few-shot learning, the number of labeled samples in the target domain is much smaller than the that of unlabeled ones, i.e., $N_t^l \ll N_t^u$. Moreover, it is also much smaller than the number of source domain samples, i.e., $N_t^l \ll N_s$.

The overall training procedure of F-CADA is presented in Fig. 2. All the other steps are pretty much similar to CADA except for step 3, detailed below:

Step 3: Suppose few labeled samples $\{\mathbf{X}_t^l, Y_t^l\}$ are available in the target domain. As the most vital step in F-CADA, we design a label learning algorithm to assign presumptive labels $\tilde{Y}_t^l$ to target unlabeled samples $\mathbf{X}_t^u$. Then, we fine-tune the target encoder obtained in Step 2 and build up a target classifier C_t using both unlabeled target samples with presumptive labels $\{\mathbf{X}_t^u, \tilde{Y}_t^l\}$ and labeled target samples $\{\mathbf{X}_t^l, Y_t^l\}$.

Assume k_i labeled samples are available for class i in the target domain, we can compute in the embedded space (1) the centroid vector c_i for each class and (2) a similarity metric between each unlabeled target sample $x_{t,j}^u \in \mathbf{X}_t^u$ and the specific centroid, denoted by $\psi\left(f(x_{t,j}^u), c_i\right)$. Depending on the dimension of the transformed feature space, this similarity metric can simply be a Gaussian kernel to capture local similarity (Maaten and Hinton 2008), or the inverse of Wasserstein distance (Shen et al. 2018) for better generalization with complex networks.

Ideally, the semi-supervised scheme should be able to (1) identify the correct labels of unlabeled target samples, and (2) update the encoder with the additional information. Using information entropy as the measure of "goodness of separation", we can formulate the joint objective into the following minimization:

$$\min_{y_{t,j}^u \in \tilde{Y}_t^u, f \in \mathcal{H}} \mathcal{L}_U(\mathbf{X}_t^u, \mathbf{X}_t^l, \tilde{Y}_t^l) = \sum_{x_{t,j}^u \in \mathbf{X}_t^u} H\left(\sigma(\psi(f(x_{t,j}^u), c_{y_{t,j}^u})/\tau)\right)$$

where the $H(\cdot)$ is the entropy function, $\sigma(\cdot)$ is the softmax function, and τ is a decay factor that controls the neighborhood proximity. The above problem is combinatorial in nature due to the discrete presumptive labels $y_{t,j}^u$. We establish an alternating approach that recursively performs (1) fixing the feature mapping f and propagating presumptive labels using a greedy assignment, i.e., the j^{th} unlabeled sample is presumed to have the same label to its closest centroid, and (2) updating the feature mapping (the encoder) as supervised learning by treating the presumptive labels as true labels.

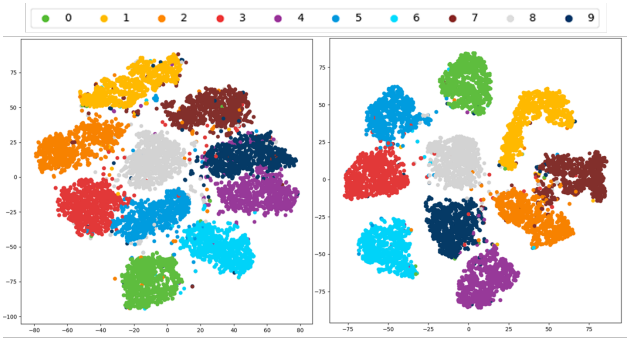
The proposed greedy propagation, intuitively simple and practically easy to implement, in fact has theoretical guarantees since the entropy objective is approximately submodular when the feature mapping is fixed. Interested readers are referred to (Zhou and Spanos 2016) for a detailed theoretical analysis. The above is conducted alternately until the convergence of the feature mapping and presumptive label assignment. In practice, it is observed that the convergence is usually achieved in few iterations. Adding the above objective function to that of CADA in equation (6), we obtain the overall learning formulation of F-CADA. In the testing step (step 4 in Fig. 2), we map the target testing samples to the latent feature space through the updated target encoder M_t, and then apply the updated target classifier C_t to identify their classes.

Experiments

We evaluate CADA and F-CADA for 2 real-world domain adaptation problems: 1) digit classification adaptation across 3 benchmark splits of public digit datasets; 2) spatial adaptation for WiFi-enabled device-free gesture recognition.

Digit Adaptation

3 public digit datasets, MNIST (LeCun et al. 1998), USPS (Hull 1994), and SVHN (Netzer et al. 2011), which consist 10 classes of digits are used in our digit adaptation experiments. We evaluate our methods across 3 adaptation shifts: MNIST ⇒ USPS, USPS ⇒ MNIST, and SVHN ⇒ MNIST, that are commonly adopted for digit adaptation assessment. The models are trained using the full training sets and evaluated on the full testing sets. We leverage a variant of the LeNet architecture as the encoder and the classifier for CADA and F-CADA for all digit shifts. We repeat the experiment 50 times for each digit adaptation case and performed model selection based on the recent Bayesian optimization technique (Malkomes, Schaff, and Garnett 2016) to identify optimal choices of all hyper-perimeters, e.g., the structure and dropout rate of the encoder, the decay fact of F-CADA, etc.



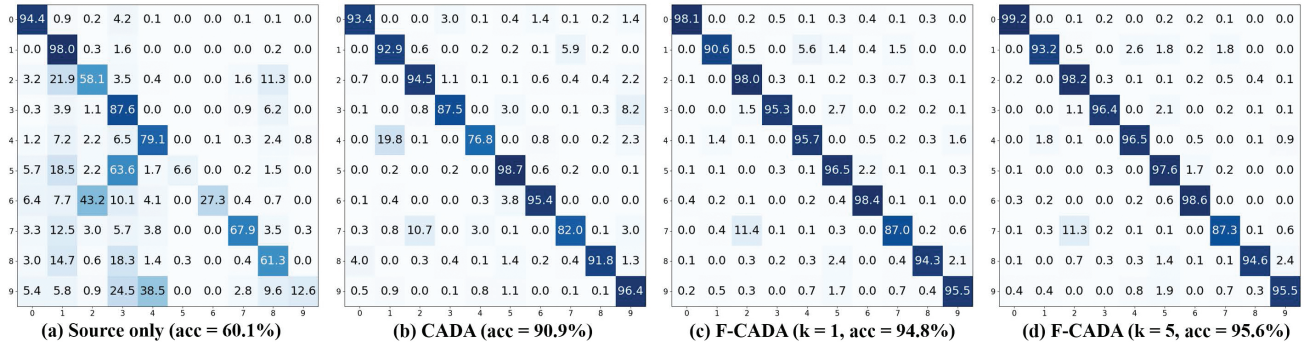
(a) Non-adapted source encoder (b) CADA source encoder

Figure 3: The t-SNE visualization of features embedded using distinct encoders in target domain. (SVHN ⇒ MNIST).

Performance of CADA We compare the performance of CADA with 3 traditional UDA methods (DDC, RevGrad, DRCN), and 3 state-of-the-art adversarial UDA methods (CoGAN, ADDA, CyCADA). Table 1 reports the classification accuracies of these methods for each shift. The 2^{nd} column shows the accuracies when the non-adapted source classifiers are applied as the lower-baseline, and the last column reports the accuracies when the target classifiers are trained with full target training samples as the upper-baseline. It can be observed that CADA outperforms others

Table 1: Digit adaptation across MNIST-USPS-SVHN datasets.

Scenario	Source only	*Traditional UDA*			*Adversarial UDA*				*SDA*								Target fully supervised
		DDC	RevGrad	DRCN	CoGAN	ADDA	CyCADA	**CADA**	SDA	1	2	3	4	5	6	7	
MNIST ⇒ USPS	75.2 ± 1.6	79.1 ± 0.5	77.1 ± 1.8	91.8 ± 0.1	91.2 ± 0.1	89.4 ± 0.2	95.6 ± 0.2	**96.4 ± 0.1**	CCSA	85.0	89.0	90.1	91.4	92.4	93.0	92.9	98.9 ± 0.1
									FADA	89.1	91.3	91.9	93.3	93.4	94.0	94.4	
									F-CADA	**97.2**	**97.5**	**97.9**	**98.1**	**98.3**	**98.4**	**98.6**	
USPS ⇒ MNIST	57.1 ± 1.7	66.5 ± 3.3	73.0 ± 2.0	73.7 ± 0.04	89.1 ± 0.8	90.1 ± 0.8	96.5 ± 0.1	**97.0 ± 0.1**	CCSA	78.4	82.2	85.8	96.1	88.8	89.6	89.4	99.2 ± 0.1
									FADA	81.1	84.2	87.5	89.9	91.1	91.2	91.5	
									F-CADA	**97.5**	**97.8**	**98.1**	**98.4**	**98.6**	**98.8**	**98.9**	
SVHN ⇒ MNIST	60.1 ± 1.1	68.1 ± 0.3	73.9	82.0 ± 1.6	-	76.0 ± 1.8	90.4 ± 0.4	**90.9 ± 0.2**	FT	65.5	68.6	70.7	7.3	74.5	74.6	75.4	99.2 ± 0.1
									FADA	72.8	81.8	82.6	85.1	86.1	86.8	87.2	
									F-CADA	**94.8**	**95.1**	**95.4**	**95.5**	**95.6**	**95.9**	**96.1**	



(a) Source only (acc = 60.1%) (b) CADA (acc = 90.9%) (c) F-CADA (k = 1, acc = 94.8%) (d) F-CADA (k = 5, acc = 95.6%)

Figure 4: Confusion matrices for the digit adaptation (SVHN⇒ MNIST).

for all the aforementioned digit adaptation scenarios. For relatively easy shift between MNIST and USPS (both greyscale hand-written digit datasets), CADA enhances the accuracy in both adaptation directions by **at least 21%** compared to the lower-baseline. It elevates the performance closer to supervised learning methods as well as the upper-baseline as demonstrated in Table 1.

The adaptation for SVHN ⇒ MNIST is much more challenging since SVHN is a color digit dataset of house number plates while MNIST contains unified greyscale digits. Even in this case, CADA improves the accuracy by **31%** over the lower-baseline and outperforms the prior works. We use t-SNE (Maaten and Hinton 2008) to map the embedded feature representations through different encoders to a 2-D space for better visualization of the domain shift. Fig. 3(a) and Fig. 3(b) depict the embedded features using the non-adapted source encoder and the CADA source encoder, respectively (different color represents different digits). Confusion matrices before and after using CADA for this adaptation are presented in Fig. 4(a) and Fig. 4(b). If we directly apply the non-adapted source encoder in the target domain, as shown in Fig. 3(a), the clusters of 3s and 5s, 4s and 9s overlap with each other, which leads to corresponding large misclassification among these digits as shown in Fig. 4(a). After employing CADA, the digit clusters of these common confusions are separated in the latent feature space (Fig. 3(b)), that indeed contributes to the corresponding performance gain as presented in confusion matrix (Fig. 4(b)).

Performance of F-CADA We randomly chose ($k = 1, \ldots, 7$) labeled samples per class as the labeled target samples and utilized them for Step 3 label learning of F-CADA. The performance of F-CADA is compared with one SDA method: CCSA (Motiian et al. 2017b), and one advanced few-shot adversarial SDA method: FADA (Motiian et al. 2017a). For the scenario of SVHN ⇒ MNIST, we compared F-CADA to the source only model on available labeled target data with fine-tuning (denote as FT in Table 1). It can be observed from Table 1 that F-CADA achieves significant performance gain over the current best SDA benchmarks in all scenarios. Another noteworthy point is that it achieves comparable accuracy to the upper-baseline with only 7 labeled target samples per category. This impressive performance comes from 2 main reasons. Firstly, F-CADA inherits the advantages of CADA. As shown in Table 1, the accuracy of CADA is already higher than SDA methods for several cases. The embedded target dataset via CADA is the ideal input dataset for the following label learning of F-CADA. Secondly, the proposed label learning method can fully make use of the few labeled target samples for accuracy enhancements. For instance, Fig. 5 depicts the label learning accuracy and classification accuracy of SVHN ⇒

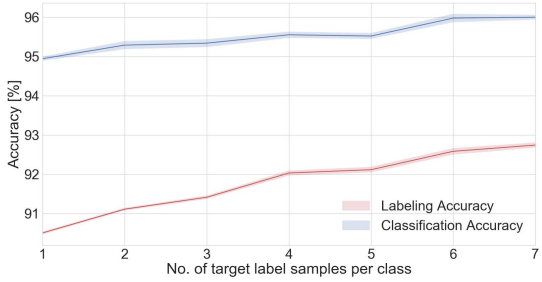


Figure 5: Impact of number of labeled target samples on labeling accuracy and classification accuracy (SVHN ⇒ MNIST). The shaded area is the 5% and 95% percentile.

Table 2: Spatial adaptation for gesture recognition across different environments.

Scenario	source only	*Traditional UDA*		*Adversarial UDA*			**F-CADA**			Target fully supervised
		RevGrad	DRCN	GoGAN	ADDA	**CADA**	**1**	**3**	**5**	
Large ⇒ Small	58.4 ± 0.7	68.1 ± 0.2	69.3 ± 0.3	69.4 ± 0.2	71.5 ± 0.3	**88.8 ± 0.1**	**92.3**	**96.3**	**98.7**	99.2 ± 0.1
Small ⇒ Large	62.2 ± 0.6	66.6 ± 1.1	65.8 ± 0.7	70.2 ± 0.5	67.7 ± 0.6	**87.4 ± 0.1**	**91.7**	**96.0**	**98.3**	99.1 ± 0.1

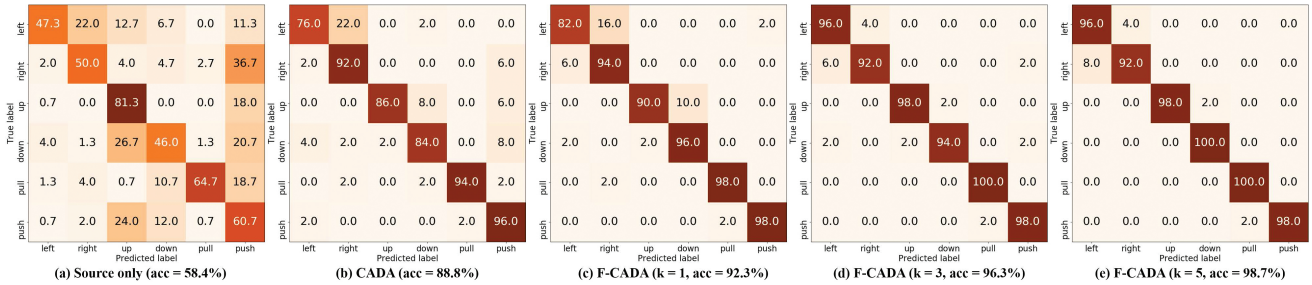


Figure 6: Confusion matrices for gesture recognition (large conference room ⇒ small conference room).

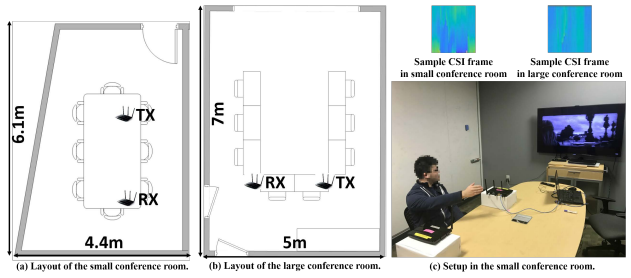


Figure 7: Floor plan of the testbeds for the spatial adaptation experiments and sample CSI frames from different spatial sources.

MNIST when distinct numbers of labeled target samples are available. We can easily observe the positive correlation between label learning accuracy and classification accuracy as the number of labeled target samples is increased. Comparing the confusion matrices between CADA (Fig. 4(b)) and F-CADA ($k = 1$) (Fig. 4(c)), the digit misclassification between 1s and 4s is reduced significantly by F-CADA with only one labeled target sample per class. It leads to **4%** overall accuracy improvement. These analyses prove the excellent few-shot domain adaptation performance of F-CADA even when the number of labeled target samples is tiny.

Spatial Adaptation for Gesture Recognition

We also implement our methods to enhance the spatial adaptation capability of WiFi-enabled device-free gesture recognition (GR). By leveraging fine-grained channel measurement (Channel State Information (CSI)) from WiFi physical layer and advanced machine learning methods, numerous occupancy sensing tasks, e.g. crowd counting (Zou et al. 2018b), human activity recognition (Zou et al. 2018c), and even human identification (Zou et al. 2018a), have been realized in a device-free, privacy-preserving and non-intrusive manner. Since human gestures also alter the WiFi signal propagation among WiFi-enabled IoT devices, we can identify the gestures in a device-free manner via the Channel State Information (CSI) enabled sensing platform proposed in (Zou et al. 2018a). One major bottleneck being, tedious data collection and labeling process required to train a new gesture classifier when the system is to be employed in a new environment. The classifier is also vulnerable to spatial variations. Thus, we aim to use our methods to improve the accuracy and resilience of the classifier over spatial dynamics without collecting 1) any labeled target samples, 2) sparsely labeled ones. As shown in Fig. 7, the experiments were conducted in 2 conference rooms with different sizes (i.e. a large conference room ($7m \times 5m$) and a small conference room ($6.1m \times 4.4m$). Volunteers performed 6 common gestures, moving a hand right and left, up and down, push and pull between the two IoT devices. 200 samples per gesture were collected in each room during different days. After transforming CSI time series data into CSI frames (each CSI frames size: 400×228 as depicted in Fig. 7), we modified the LeNet architecture and designed a dedicated encoder and classifier for our methods.

Performance of CADA We compare the performance of CADA with 2 state-of-the-art UDA methods (RevGrad and DRCN) and 2 adversarial UDA methods (CoGAN and ADDA). Table 2 summarizes the gesture classification accuracies of these methods in both adaptation directions between the 2 conference rooms. CADA enhances the accuracy by **at least 25%** over the lower-baseline (non-adapted source encoder is adopted) in both adaptation scenarios, without tedious labeled target data collection and training process. Comparing the confusion matrices before and after using CADA (Fig. 6(a) and Fig. 6(b) (large ⇒ small)), the recognition accuracy of every gesture is improved. It can be easily observed that CADA outperforms all the traditional and adversarial UDA approaches. It realizes resilient WiFi-enabled device-free gesture recognition against spatial variations without time-consuming and labor-intensive data collection and labeling process in a new environment.

Performance of F-CADA We randomly chose ($k = 1, 3, 5$) labeled samples per gesture as the labeled target samples and used them for the Step 3 of F-CADA training process. Similar to the digit adaptation results, F-CADA achieves significant performance gain over UDA methods by **at least 3.5%** when only one labeled sample per gesture is available in the target domain. Moreover, as demonstrated in Table 2 and Fig. 6, its accuracy is further increased when a little more labeled samples are available and employed. Its accuracy reaches **98.5%** with 5-shot learning, which is only 0.6% lower than the upper-baseline (train a new classifier with full labeled target samples).

Conclusion

In this paper, we proposed Consensus Adversarial Domain Adaptation (CADA) that gives freedom to both target encoder and source encoder in adversarial learning, by embedding data from both domains into a consensus domain-invariant feature space. In this manner, the domain discrepancy can be further minimized. CADA's feature representations are more robust to large domain shift and have the capacity to avoid over-fitting models in both domains. A novel few-shot domain adaptation scheme (F-CADA) is also proposed to enhance the ADA performance by exploring few labeled target data in an efficient way. By inheriting CADA's feature representation, F-CADA assigns presumptive labels to unlabeled data points in the target domain, by greedily minimizing an information entropy loss function. The greedy label learning method has theoretical guarantees since the entropy objective is approximately submodular. Extensive real-world experiments on digit recognition across multiple benchmark digit datasets and WiFi-enabled device-free gesture recognition under spatial dynamics are conducted. The results validate that CADA achieves compelling results and outperforms state-of-the-art UDA and SDA methods. F-CADA can further enhance the adaptation performance even with sparsely labeled target data.

Acknowledgments

This work is supported by a 2018 Seed Fund Award from CITRIS and the Banatao Institute at the University of California. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

Ganin, Y., and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, 1180–1189.

Ghifary, M.; Kleijn, W. B.; Zhang, M.; Balduzzi, D.; and Li, W. 2016. Deep reconstruction-classification networks for unsupervised domain adaptation. In *European Conference on Computer Vision*, 597–613. Springer.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A. A.; and Darrell, T. 2017. Cycada: Cycle-consistent adversarial domain adaptation. *arXiv preprint arXiv:1711.03213*.

Hull, J. J. 1994. A database for handwritten text recognition research. *IEEE Transactions on pattern analysis and machine intelligence* 16(5):550–554.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.

Liu, M.-Y., and Tuzel, O. 2016. Coupled generative adversarial networks. In *Advances in neural information processing systems*, 469–477.

Luo, Z.; Zou, Y.; Hoffman, J.; and Fei-Fei, L. F. 2017. Label efficient learning of transferable representations acrosss domains and tasks. In *Advances in Neural Information Processing Systems*, 164–176.

Maaten, L. v. d., and Hinton, G. 2008. Visualizing data using t-sne. *Journal of machine learning research* 9(Nov):2579–2605.

Malkomes, G.; Schaff, C.; and Garnett, R. 2016. Bayesian optimization for automated model selection. In *Advances in Neural Information Processing Systems*, 2900–2908.

Motiian, S.; Jones, Q.; Iranmanesh, S.; and Doretto, G. 2017a. Few-shot adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, 6673–6683.

Motiian, S.; Piccirilli, M.; Adjeroh, D. A.; and Doretto, G. 2017b. Unified deep supervised domain adaptation and generalization. In *The IEEE International Conference on Computer Vision (ICCV)*, volume 2.

Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; and Ng, A. Y. 2011. Reading digits in natural images with unsupervised feature learning. In *NIPS workshop on deep learning and unsupervised feature learning*, volume 2011, 5.

Ren, M.; Triantafillou, E.; Ravi, S.; Snell, J.; Swersky, K.; Tenenbaum, J. B.; Larochelle, H.; and Zemel, R. S. 2018. Meta-learning for semi-supervised few-shot classification. *arXiv preprint arXiv:1803.00676*.

Shen, J.; Qu, Y.; Zhang, W.; and Yu, Y. 2018. Wasserstein distance guided representation learning for domain adaptation. In *AAAI*.

Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; and Darrell, T. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. *arXiv preprint arXiv:1702.05464*.

Volpi, R.; Morerio, P.; Savarese, S.; and Murino, V. 2017. Adversarial feature augmentation for unsupervised domain adaptation. *arXiv preprint arXiv:1711.08561*.

Zhou, Y., and Spanos, C. J. 2016. Causal meets submodular: Subset selection with directed information. In *Advances In Neural Information Processing Systems*, 2649–2657.

Zou, H.; Zhou, Y.; Yang, J.; Gu, W.; Xie, L.; and Spanos, C. J. 2018a. Wifi-based human identification via convex tensor shapelet learning. In *AAAI*.

Zou, H.; Zhou, Y.; Yang, J.; and Spanos, C. J. 2018b. Device-free occupancy detection and crowd counting in smart buildings with wifi-enabled iot. *Energy and Buildings* 174:309–322.

Zou, H.; Zhou, Y.; Yang, J.; and Spanos, C. J. 2018c. Towards occupant activity driven smart buildings via wifi-enabled iot devices and deep learning. *Energy and Buildings* 177:12–22.