Zero-shot Domain Adaptation without Source Irrelevant Task-of-interest

Anonymous ECCV submission

Paper ID 7002

Abstract. Zero-shot domain adaptation (ZSDA) enables models adapt to new domains without further task-of-interest (ToI) data collection in the target domains. For using conventional ZSDA schemes, following requirements should be met: i) irrelevant task-of-interest (IrT) datasets in both of the source and target domains, and ii) a paired condition (or labels [13]) between the source IrT and target IrT datasets. The above requirements make impractical to real world applications although they are successful in synthetic benchmarks.

We propose a new zero-shot domain adaptation scheme without IrT samples of source domain, i.e. source IrT-free zero-shot domain adaptation. Our method uses geometry-consistent generative adversarial network (GcGAN) with some modifications in the generator architecture and training objective constraints. The method guarantees low-effort domain adaptation without source irrelevant task-of-interest samples but with only target ones and thus, no paired condition is required.

We evaluate the method on X-MNIST benchmarks with synthetic domain pairs and WIFI channel state information (CSI) human activity recognition (HAR) task datasets by our manual collection. We empirically prove that our method successfully adapts to novel domains in above benchmarks and show the feasibility of the source-IrT free zero-shot domain adaptation.

Keywords: Transfer Learning; Domain Adaptation; Zero-shot Domain Adaptation; Geometry-consistent Generative Adversarial Networks

1 Introduction

Unsupervised DA approach [17, 26, 20, 10, 22, 18, 28, 12] guarantees a good adaptation performance without labor-intensive labeling process, under an assumption that the enough amount of data already exists. However, unsupervised DA requires data acquisition process in the target domains. For instance, in human activity recognition task, to adapt models to new environments, expensive human labor is required to collect data on target environments.

Zero-shot domain adaptation methods [23, 29, 30, 13] are good options to deploy models in novel domains without **task-of-interest (ToI)** target domain data but with **irrelevant task-of-interest (IrT)**, if exists. It means it is possible to adapt a model to new domain if there exists enough samples on both of the source and target domains and the samples are not necessarily the same

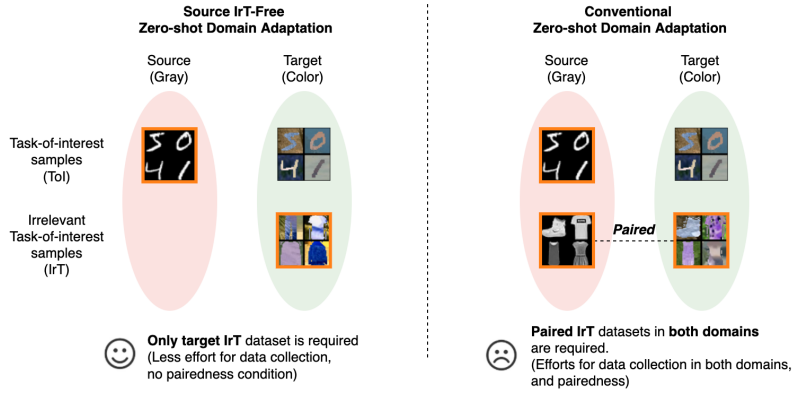


Fig. 1: The motivation of the source IrT-free zero-shot domain adaptation. Orange border means available during training. Note that only source ToI samples have labels. (Best viewed in color.)

task. At the same time, it means the domain adaptation can be done without collecting ToI classes, which might be expensive or laborious, but of low-effort or zero-effort classes instead; for instance, in human activity recognition task, a "no activity" class (a sample collected during the period without any human activities/movements) in two different rooms is the IrT class that can be leveraged to adapt a model to classify the ToI classes such as walk, sitdown, and etc.

In this paper, we introduce a new zero-shot domain adaptation scheme, source IrT-free zero-shot domain adaptation, which remove the source IrT existence requirement and thus, the paired condition. We propose the unpaired zero-shot pixel-level domain adaptation using geometry-consistent generative adversarial network (GcGAN) [8]. We modify the GcGAN generator into the siamese generator to factorize features into domain-specific features and task-specific features. Then, we add cross-combination invariance and semantic identity mapping to the training objective. It requires the IrT samples only in target domain while conventional ZSDA approaches need them in both domains.

Not only conventional X-MNIST classification benchmarks (MNIST, FashionMNIST, EMNIST) with synthetic domains, our experiments include a real world example, WIFI CSI-based human activity recognition (HAR) task benchmarks. We argue the WIFI CSI-based HAR task is a good test-bed for evaluating domain adaptation for following reasons; Firstly, the task exploits environment-dependent multi-path information of radio frequency signal, which is highly environment-dependent. So, adaptation to new domains is a one of the key issues to develop WIFI CSI human activity recognition system [14, 36, 34]. Secondly, human activity recognition tasks require lots of physical efforts by human for data collection. This nature of human activity recognition task leads substantial cost for adapting pre-trained model to new environments. Lastly, there are various cases of domain shifts, for instance, domain shifts occur when the position

of transmitters or receivers is changed or the listening frequency is changed or the room is changed. The three domain pairs of our benchmark cover the each of the cases.

Note that the conventional zero-shot domain adaptation is not applicable to our WIFI CSI-based HAR benchmark, since it is impossible to replicate the exact motion behavior of human in one domain to another while it is possible for the MNIST synthetic domains; the samples are synthesized from the identical samples. The data generation process in real world scenarios is not independent and identically distributed and thus, no paired condition can be established in general.

Here, our contributions can be summarized as follows:

- To the best of our knowledge, we firstly introduce the zero-shot domain adaptation scheme without source irrelevant task-of-interest samples, so called, the source IrT-free zero-shot domain adaptation. The method reduces data collection cost significantly and guarantees successful adaptation on not only synthetic benchmarks but non i.i.d real world task benchmark, since no paired condition is required.
- For the source IrT-free ZSDA task, our proposed method tackles an image-to-image translation between domains having disjoint classes, which is an ill-posed problem. We propose a siamese generator architecture and semantic identity mapping to modify the geometry-consistent generative adversarial networks. We empirically observe our proposed method actually mitigate the corruption of discriminative information during translation.

2 Related Works

2.1 Unsupervised Domain Adaptation

Domain adaptation models are trained with the labeled source domain samples and labeled/semi-labeled/unlabeled target domain samples, in general. We call supervised DA if the target domain samples has the (few) labels [22, 21], semi-supervised DA if only part of samples have the labels, and unsupervised DA if no labels are available [17, 10, 26, 20, 18, 19, 28, 12, 24].

2.2 Generative Adversarial Approach for Pixel-level Domain Transfer

Pixel-level domain transfer approach is closely related to image-to-image translation. One can naturally think that transferring target domain image to source domain image can be done by converting target image to source-like image sample and then utilizing a pretrained classifier on source domain dataset. The early apporach is shown in [16]. It proposes a coupled generative adversarial network training with weight sharing constraint to learn a joint distribution of two domains. [4] also employs generative adversarial network whose generator translates

source image conditioned with a random noise vector z to target image and discriminator distinguishes generated image and real target image. [37] introduces a cycle-consistency to the generative adversarial objective along with the identity mapping constraint. It shows visually impressive results on various applications without paired condition and can work with larger images. [12] extends the above work by adding semantic consistency to the objective for domain adaptation task and show significant improvements on classification task and segmentation task as well. Then, [8] tackles the geometric deformation problem in image-to-image translation models by proposing equivariant geometric consistency constraint with one-sided mapping architecture [3].

2.3 Zero-shot Domain Adaptation

Zero-shot domain adaptation (ZSDA) enables domain adaptation using IrT data from both of source and target domains, proposed in [23]. The method has two steps and one optional step. It trains a source feature extractor to simulates target representations from IrT classes data of source and target domain with each feature extractor. Then, it trains the source feature extractor and classifier with given labeled ToI data from source domain, while the weight of source feature extractor is shared with the target feature extractor's weight. Here, the target feature extractor is fixed to simulate target feature representations as the previous step. [29, 30] extends coupled generative adversarial network (CoGAN) [16] for ZSDA. [29] propose conditional coupled generative adversarial network (CoCoGAN) using conditioning variable c. The architecture consists of a pair of GAN models for source and target domains. For obtaining joint distribution of dual-domain samples for both IrT and ToI classes, the task label loss conditioned by c is introduced to make the ToI and IrT task classes in-distinguishable. Then, we get image pairs of source and target and train target domain classifier with source domain labels. [30] adopts CoGAN to learn domain shifts between IrT source and target domain pairs. Here, the underlying assumption is that the domain shifts learned from one task can be transferred to another. A pair of co-training classifiers predict labels and its logits are used for cross-domain consistency. Recently, [13] proposes non-generative feature-level adversarial ZSDA framework. The overhead of the method is significantly lower than generative ones and it does not require the paired condition by using labeled IrT samples on both of the domains.

2.4 Environment-independent WIFI CSI-based Human Activity Recognition

WiFi CSI-based activity recognition has several advantages over the camera-based human activity recognition: 1) robust to illumination and occlusion problems, 2)no deployment costs of image-capturing devices such as RGB/depth cameras since WiFi is universally deployed around the world and its RF signal is everywhere in these days, 3) free from privacy problem since CSI spectrogram image is not informative to human eyes' perception. However, since channel state

information is derived from the radio frequency signal which is reflected, attenuated and diffracted by nearby objects, it inevitably has strong dependency on nearby environments.

To achieve environment-independent recognition performance, one of the early attempts is CARM [32], which provides an environment-independent feature extraction technique for human activity recognition. It denoises the raw CSI amplitude values using principle component analysis. Then, it extracts the latent motion speed information which is independent to nearby environments using discrete wavelet transform. For neural network approaches, domain adaptation models show good environment-independent performance [14], but it requires expensive data collection in target environment (domain). [36] propose Widar3.0, a zero-efforts cross-domain gesture recognition method to reduce the data collection efforts further in gesture recognition. It extracts domain-invariant features using Doppler frequency shift from CSI values and generates body-coordinate velocity profile and trains gated recurrent unit (GRU) networks [5].

3 Preliminary

3.1 Domain Adaptation

A domain is a specific distribution $\mathcal{D}$ over the instance set $\mathcal{X}$ [2]. It is often represented as a pair of instance set $\mathcal{X}$ and marginal distribution $P(X)$, where $X = \{x_1, x_2, ...\} \in \mathcal{X}$ [31], i.e. $\mathcal{D} = \langle \mathcal{X}, P(X) \rangle$. Then, one can define domain adaptation is transferring a model f trained on one domain to another domain sharing a common task $\mathcal{T}$. Here, note that the source domain $\mathcal{D}_S = \langle \mathcal{X}_S, P(X_S) \rangle$ is a domain where the model f is originally trained and the target domain $\mathcal{D}_T = \langle \mathcal{X}_T, P(X_T) \rangle$ is a domain to be transferred. In unsupervised domain adaptation, the labels Y_S for X_S from $\mathcal{D}_S$ instances are accessible while the labels of X_T from $\mathcal{D}_T$ are not.

The underlying assumption above is that all data samples are of the common **task-of-interest** (ToI). That is, what we denote as D_S and D_T above can be denoted as D_S^{ToI} and D_T^{ToI}, respectively, in zero-shot domain adaptation. Zero-shot domain adaptation is a domain adaptation approach using **irrelevant task-of-interest** (IrT) samples. It utilizes IrT samples in both of the source and target domains, i.e. $x_s^{irt} \sim D_S^{IrT}$ and $x_t^{irt} \sim D_T^{IrT}$, to transfer ToI task models from the source to new target domains.

3.2 Pixel-level Domain Adaptation

One of the intuitive approach for domain adaptation is translating a unlabeled target domain samples to label-rich source domain so that one can leverage the pretrained model f_S of the source domain, so called pixel-level domain adaptation. The pixel-level domain adaptation approach is closely related to image-to-image translation task. The image-to-image translation is also called style transfer task that translates an image in one domain to another. The neural

style transfer [4, 37, 8] has been a major computer vision task and the generative adversarial networks-based models is one of the dominant models.

The generative adversarial networks for image-to-image translation have the generator-discriminator architecture with minimax optimization objective, as the one for image generation with an uniform random noise z but with source domain image $x_s \sim X_S$. Then, a basic form of generative adversarial objective function for the image-to-image translation is

$$\mathcal{L}_{\text{adv}}\left(G_{ST}, D_T, X_T, X_S\right) = \mathbb{E}_{x_t \sim X_T}\left[\log D_T\left(x_t\right)\right] + \mathbb{E}_{x_s \sim X_S}\left[\log\left(1 - D_T\left(G_{ST}\left(x_s\right)\right)\right)\right] \tag{1}$$

where G_{ST} is a generator (or mapper) translating source domain image in $\mathcal{D}_S$ to target domain $\mathcal{D}_T$ and D is a discriminator to determine whether the generated images are real or fake.

3.3 Geometry-consistent Generative Adversarial Network

A mapping $\Phi : S \to T$ is equivariant to a geometric transformation T satisfying

$$\Phi\left(Tx\right) = T'\Phi(x) \tag{2}$$

for any x in S. [7] argues that the equivariance helps generalization capability. [8] introduces the notion of equivariance to the image-to-image translation task as a geometric consistency constraints. The generator is trained to be equivariant to the geometric transformations using geometric consistency constraints along with an identity mapping [27]. The geometric consistency can be written as follows:

$$\begin{aligned}\mathcal{L}_{\text{geo}}\ \left(G_{ST}, G_{\tilde{S}\tilde{T}}, S, T\right) =& \mathbb{E}_{x_s \sim X_S}\left[\left\|G_{ST}(x_s) - f^{-1}\left(G_{\tilde{S}\tilde{T}}(f(x_s))\right)\right\|_1\right] \\ &+ \mathbb{E}_{x_s \sim X_S}\left[\left\|G_{\tilde{S}\tilde{T}}(f(x_s)) - f\left(G_{ST}(x_s)\right)\right\|_1\right]\end{aligned} \tag{3}$$

where f is a geometric transformation such as 90° clockwise rotation or vertical flip, and $\tilde{S}$ and $\tilde{T}$ are the domains transformed by f, respectively. Note that the parameters of G_{ST} and $G_{\tilde{S}\tilde{T}}$ are shared. One can see the geometry constraint as a reconstruction loss under the predefined geometric transformation f. Note that we choose the geometric transformation $f =$ (90° clockwise rotation) in this paper.

3.4 WIFI Channel State Information

The IEEE 802.11 standards used in recent commercial off-the-shelf devices is using multiple inputs and multiple outputs (MIMO) system with orthogonal frequency-division (OFDM) modulation scheme. MIMO guarantees advantageous gains in spatial diversity and multiplexing, and reduces interference. OFDM uses 52 orthogonal subcarriers (48 data subcarriers + 4 pilot subcarriers) with

small bandwidth, and therefore, each subcarrier suffers a narrow, small-scale flat fading.

To guarantee reliable communications between transmitters (Tx) and recievers (Rx), the channel state of single subcarrier for each antenna should be estimated. Note that S be the number of measured subcarriers, A^{Tx} be the number of antenna in Tx array, and A^{Rx} be the number of antenna in Rx array. Then, a channel state information matrix H_i of single subcarrier $i \in S$, can be represented as

$$y_i = H_i x_i + n_i \tag{4}$$

where $x_i \in \mathbb{C}^{A^{Tx}}, y_i \in \mathbb{C}^{A^{Rx}}$ is a transmitted symbol vector and a received symbol of subcarrier i, respectively, and n_i is a white Gaussian noise vector. A channel state information matrix H at a time step t has $S \times A^{Tx} \times A^{Rx}$ entries, where $H = \{H_1, H_2, ...H_S\}$. The entry of H_i matrix is called channel frequency response (CFR) which is a complex form. In general, its amplitude values is often preferred as reliable source for feature extraction in human activity/gesture recognition task rather than its phase values [35].

4 Method

4.1 Source IrT-free Zero-shot Domain Adaptation

Source IrT-free zero-shot domain adaptation (ZSDA) task is a chellengingly ill-posed problem which it only uses labeled source ToI samples and unlabeled target IrT samples, while conventional ZSDA methods use labeled source ToI samples and unlabeled IrT samples from both of the source and target domains. It means, in the source IrT-free ZSDA setting, we have no access to both of the target ToI samples and the source IrT samples, and thus, no paired condition during the training phase.

source	MNIST (0.99168)		
target	MNIST_C	MNIST_E	MNIST_N
Src Only	0.6327	0.5194	0.2746
DANN	0.7584	0.9545	0.6311
ADDA	0.8175	0.7237	-
CycleGAN	0.9658	0.1057	-
GcGAN	**0.9786**	**0.9809**	**0.9886**

Table 1: The accuracy comparison of unsupervised DA methods. Empty entry means fail to converge.

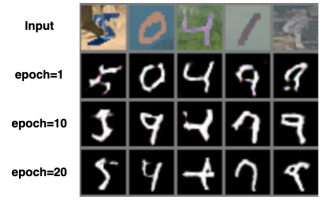


Fig. 2: The translated results of GcGAN on different epochs trained under the source IrT-free ZSDA setting.

To circumvent the task discrepancy between given source and target domain datasets, our strategy is learning a task-independent target→source image-to-

image translation mapping G_{TS} using IrT target dataset X_T^{IrT} and a ToI source dataset X_S^{ToI}. Then, in inference phase, we can leverage source domain classifier f_S to make predictions on the translated images $G_{TS}(X_T)$. Our task-independent mapping is a modified GcGAN, called siamese GcGAN. We choose GcGAN as base architecture since it is empirically proved that the geometry-consistency preserves the disriminability well after the translation in several experiments [8] and our comparison with other unsupervised domain adaptation methods Table 1 on MNIST benchmark with four synthetic domains.

Applying GcGAN directly in the source IrT-free ZSDA setting fails to learn the task-independent target→source mapping according to Table 2 and Figure 2. Here, there are two challenges we observe: the target ToI test accuracy keeps decreasing sharply during training (See Figure 7 in Supplementary Material) as the translated images keep losing its (i) its geometry and (ii) semantic information. Figure 2 shows that, as the epoch goes, results of 5, 4, 9 keeps its original semantics but suffers a geometric deformation problem, and 0, 1 totally lost its original semantics. To tackle the issues, we propose a siamese generator architecture with cross-combination invariance and semantic identity mapping.

4.2 Siamese Generator

Motivation Given source and target datasets are of different tasks, for successful translation of target → source domain, we argue the generator should be able to discriminate the task-specific information and domain-specific information during training. Hence, we propose a modified generator architecture, a siamese generator for factorizing features into domain-specific features and task-specific features.

Domain-invariant/task-invariant feature encoder The siamese generator has two feature extraction encoders, a task-specific feature encoder E_c and a domain-specific feature encoder E_d, and an upsampling decoder U_{TS}, shown in Figure 5. Our assumption is that it is possible to obtain task-specific features and domain-specific features, by making the encoder E_c to be domain-invariant, and the encoder E_d to be task-invariant, respectively. The two encoders are adversarially-trained with domain-specific discriminator D_c and task-specific discriminator D_d, to obtain task-invariant features and domain-invariant features, respectively. Then, the objectives can be written as follows.

$$\begin{aligned}\mathcal{L}_{\text{adv}}^{\text{u}}\left(U_{TS}, D_S, X_T^{IrT}, X_S^{ToI}\right) &= \mathbb{E}_{x_s^{toi}\sim X_S^{ToI}}\left[\log D_S\left(x_s^{toi}\right)\right] \\ &+ \mathbb{E}_{c_{irt}\sim E_c\left(X_T^{IrT}\right), d_t\sim E_d\left(X_T^{IrT}\right)}\left[\log\left(1 - D_S\left(U_{TS}\left(c_{irt} + d_t\right)\right)\right)\right]\end{aligned} \tag{5}$$

$$\begin{aligned}&\mathcal{L}_{\text{adv}}^{\text{c}}\left(E_c, D_c, X_T^{IrT}, X_S^{ToI}\right) = \\ &\mathbb{E}_{x_t^{irt}\sim X_T^{IrT}}\left[\log D_c\left(x_t^{irt}\right)\right] + \mathbb{E}_{x_s^{toi}\sim X_S^{ToI}}\left[\log\left(1 - D_c\left(E_c\left(x_s^{toi}\right)\right)\right)\right]\end{aligned} \tag{6}$$

$$\begin{aligned}&\mathcal{L}_{\text{adv}}^{\text{d}}\left(E_d, D_d, X_T^{IrT}, X_S^{ToI}\right) = \\ &\mathbb{E}_{x_t^{irt}\sim X_T^{IrT}}\left[\log D_d\left(x_t^{irt}\right)\right] + \mathbb{E}_{x_s^{toi}\sim X_S^{ToI}}\left[\log\left(1 - D_d\left(E_d\left(x_s^{toi}\right)\right)\right)\right]\end{aligned} \tag{7}$$

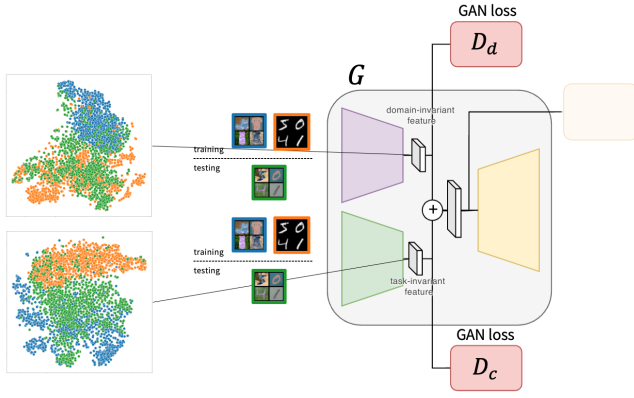


Fig. 3: The TSNE feature visualizations. Blue dots are FashionMNIST_C samples, orange dots are MNIST samples, and green dots are MNIST_C samples. To teach a task-independent target→source image translation mapping, we use domain-invariant encoder and task-invariant encoder to mitigate the coupling of task-related information and domain-related information. One can see MNIST_C and FasionMNIST_C are aligned in task-invariant space sharing the same domain, and MNIST_C and MNIST in domain-invariant space sharing the same task. (Best viewed in color.)

where $G_{TS}(x) = U(E_c(x) + E_d(x))$, D_S is source domain discriminator.

Cross-combination invariance During training, the generator module is trained with the IrT target domain samples but have no access to ToI target domain samples which will be the inputs during testing phase. So, we prevent the upsampling module from knowing whether the given combined features are from seen distribution (X_S^{ToI}, X_T^{IrT}) or unseen distribution (X_T^{ToI}, X_S^{IrT}) during translation Figure 4. This is done by an additional cross-combination invariance adversarial constraint below.

$$
\begin{aligned}
&\mathcal{L}_{\text{cross}}\left(E_c, E_d, D_{cross}, X_T^{IrT}, X_S^{ToI}\right) = \\
&\quad \mathbb{E}_{x_t^{irt}\sim X_T^{IrT}}\left[\log D_{cross}\left(E_c\left(x_t^{irt}\right) + E_d\left(x_t^{irt}\right)\right)\right] \\
&\qquad + \mathbb{E}_{x_t^{irt}\sim X_T^{IrT}, x_s^{toi}\sim X_S^{ToI}}\left[\log\left(1 - D_{cross}\left(E_c\left(x_t^{irt}\right) + E_d\left(x_s^{toi}\right)\right)\right)\right] \\
&\qquad\quad + \mathbb{E}_{x_s^{toi}\sim X_S^{ToI}}\left[\log D_{cross}\left(E_c\left(x_s^{toi}\right) + E_d\left(x_s^{toi}\right)\right)\right] \\
&\qquad\qquad + \mathbb{E}_{x_t^{irt}\sim X_T^{IrT}, x_s^{toi}\sim X_S^{ToI}}\left[\log\left(1 - D_{cross}\left(E_c\left(x_s^{toi}\right) + E_d\left(x_t^{irt}\right)\right)\right)\right]
\end{aligned}
\tag{8}
$$

The combined features $E_c\left(x_t^{irt}\right)+E_d\left(x_t^{irt}\right)$, $E_c\left(x_s^{toi}\right)+E_d\left(x_s^{toi}\right)$ and $E_c\left(x_t^{irt}\right)+E_d\left(x_s^{toi}\right)$, $E_c\left(x_s^{toi}\right)+E_d\left(x_t^{irt}\right)$ will not be discriminable, and thus it helps the generator generalize its translation performance when the ToI target domain samples is given, where the samples are of the unseen combination (ToI + target domain) during training.

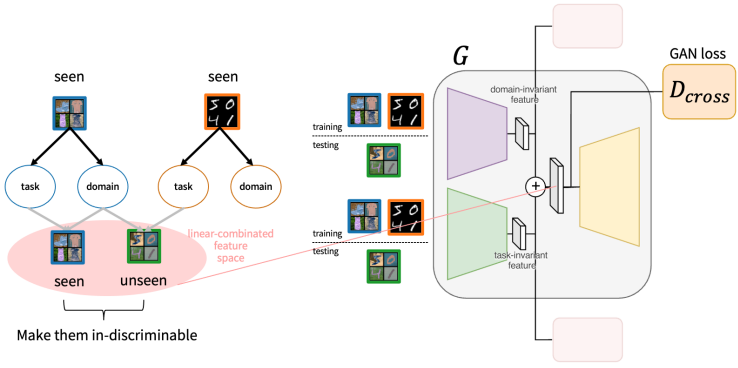


Fig. 4: The Motivation of cross-combination invariance. To enhance the generalization capability, we use cross-combination discriminator to make indistinguishable between combinations seen during training phase (source-toi, target-irt) and combinations unseen during training, but seen during testing phase (target-toi, source-irt). (Best viewed in color.)

4.3 Semantic Identity Mapping

Motivation The training graph of siamese GcGAN Figure 7 shows that the siamese generator alone is not enough to keep discriminability of ToI images after translation. [12] solves a similar problem with semantic consistency constraints enforcing source classifier output logits before and after translations are same. Since our source and target datasets during training have different tasks, semantic consistency [12] is not applicable. Therefore, we propose a semantic identity mapping, replacing the conventional identity mapping [27].

Semantic identity mapping The semantic identity mapping is based on the assumption that the semantics of source domain samples are preserved after the translation of generator G_{TS}. We define semantic identity mapping as

$$\mathcal{L}_{\text{sem\_idt}}\left(G_{TS}, X_S^{ToI}, f_S\right) = \mathbb{E}_{x_s^{toi} \sim X_S^{ToI}}\left[\|f_S\left(G_{TS}\left(x_s^{toi}\right)\right) - f_S\left(x_s^{toi}\right)\|_2\right] \quad (9)$$

where f_S is a pretrained classifier on source domain dataset. Replacing identity mapping to semantic identity mapping shows improvements in Table 2 and Table 3.

4.4 Full Objective

Combining the objectives for the siamese generator, geometric consistency constraint, and cross invariant constraint, the complete objective is shown below.

$$\begin{aligned}\mathcal{L}_{\text{zsgcgan}}\left(G_{TS}, D_S, G_{\tilde{T}\tilde{S}}, D_{\tilde{S}}, E_c, E_d, D_c, D_d, X_S^{ToI}, X_T^{IrT}, f_S\right) = \\ \mathcal{L}_{\text{adv}}^{u}\left(U_{TS}, D_S, X_S^{ToI}, X_T^{IrT}\right) + \mathcal{L}_{\text{adv}}^{u}\left(U_{\tilde{T}\tilde{S}}, D_{\tilde{S}}, X_{\tilde{S}}^{ToI}, X_{\tilde{T}}^{IrT}\right) \\ + \mathcal{L}_{\text{adv}}^{\text{d}}\left(E_d, D_d, X_T^{IrT}, X_S^{ToI}\right) + \mathcal{L}_{\text{adv}}^{\text{c}}\left(E_c, D_c, X_T^{IrT}, X_S^{ToI}\right) \\ + \mathcal{L}_{\text{geo}}\left(G_{TS}, G_{\tilde{T}\tilde{S}}, S, T\right) + \mathcal{L}_{\text{cross}}\left(E_c, E_d, D_{cross}, X_T^{IrT}, X_S^{ToI}\right) \\ + \lambda_{sem\_idt}\mathcal{L}_{\text{sem\_idt}}\left(G_{TS}, X_S^{ToI}, f_S\right)\end{aligned} \tag{10}$$

where λ_{sem_idt} is a hyperparameter.

4.5 Postprocessing

Weighted logit ensemble We denote the translated image $x_{s'}^{toi} = G_{TS}\left(x_t^{toi}\right)$ During the inference stage, rather than using the translated image's output $z_{(i),s'}^{toi} = f_S\left(x_{(i),s'}^{toi}\right)$, we predict the final classification results with weighted sum of the sigmoids of both $z_{(i),s'}^{toi}$ and $z_{(i),t}^{toi} = f_S\left(x_{(i),t}^{toi}\right)$. That is,

$$c^* = argmax\left(\lambda_{pp}\frac{\exp\left(z_{(i),t}^{toi}\right)}{\sum_j \exp\left(z_{(j),t}^{toi}\right)} + (1-\lambda_{pp})\frac{\exp\left(z_{(i),s'}^{toi}\right)}{\sum_j \exp\left(z_{(j),s'}^{toi}\right)}\right) \tag{11}$$

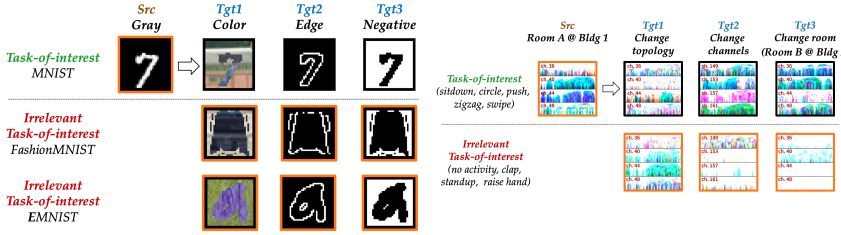
where λ_{pp} is an hyperparameter to control the contributions of the two logits. $\lambda_{pp} = 1$ equals the prediction on the original images $f_S\left(x_t^{toi}\right)$, and $\lambda_{pp} = 0$ equals the prediction on translated images $f_S\left(x_{s'}^{toi}\right)$.

5 Experiment

5.1 Datasets

MNIST Synthetic Domain Benchmark MNIST Digit classification [15] has been a widely used synthetic datasets for domain adaptation. [9] introduces MNIST-M, a synthetic domain dataset for unsupervised domain adaptation. Then, [23] extends the concept to FashionMNIST [33], EMNIST [6], NIST [11]. [29] adds more synthetic domains, such as negative, edge.

In our evaluation, to show the feasibility of the proposed method, we choose MNIST as task-of-interest (ToI) task and the FasionMNIST, EMNIST as irrelevant task-of-interest (IrT). MNIST and FashionMNIST tasks have 70,000 images with 10 classes, EMNIST task have 145,000 images with 26 classes. The class distribution for all tasks are balanced. As done in recent ZSDA works [29, 30], four domains for each task are used: grayscale domain, colored domain (C), edged domain (E) and negative domain (N) and each of the domain can be obtained by blending the patches of BSDS500 dataset [1], applying canny edge detector and subtracting each pixel value from 255, i.e. $255 - i$ where each entry i in image matrix I, respectively.



(a) X-MNIST Synthetic Domain Benchmark

(b) WIFI CSI-based Human Activity Recognition Benchmark

Fig. 5: The overview of the two source IrT-free zero-shot DA benchmarks. Two IrT datasets for X-MNIST benchmarks and one IrT splitted dataset is used and three src-tgt domain pairs are evaluated for both of the benchmarks. Data available during training is indicated with orange borders.

WIFI CSI-based Human Activity Recognition Benchmark The data collection has been done in two rooms from 14 people. There are five classes for ToI (sitdown, circle, push, zigzag, swipe), and four classes (no activity, clap, standup, raise hand) for IrT, eight balanced classes in total and the number of samples are 2496. Our evaluation has four domains: Bldg1 Room A, Bldg1 Room A with the router topology changes, Bldg1 Room A with listening channel changes (to ch.149, ch.153, ch.157, ch.161) and Bldg2 Room B (room changes). Note that the non-negligible domain shifts are caused by not only changing the room, but changing other factors such as the WIFI topology changes or listening different RF channels Table 3. We describe more details on data collection and preprocessing in Supplementary Material section.

5.2 Result & Analysis

Training Details All models in our evaluation are implemented with Pytorch and Pytorch Lightning. MNIST results are obtained by averaging runs from four random seeds on a single RTX3090 and WIFI CSI-based HAR results from three random seeds on two RTX3090s. We use FP16 half-precision training, adam optimizer with learning rate 0.0001 for all experiments. The number of epochs for each task is fixed to 1 for EMNIST_X→MNIST, 10 for FashionMNIST_X→MNIST, and 50 for all WIFI CSI-based har adaptation pairs, respectively.

Quantitative Analysis For quantitative analysis, our metric is the test accuracy after 10 epochs.

X-MNIST synthetic domain benchmark We simply replace the generator to the siamese one and surpasses the source only performance in some pairs, such as G→C (+2.66%) in EMNIST task and G→N (+19.09%) in FashionMNIST task. Another observation is the semantic identity mapping improves

		MNIST (99.16)			EMNIST (93.59)			FMNIST (90.99)		
		C	E	N	C	E	N	C	E	N
GcGAN	IrT1	58.15	29.18	4.63	28.71	11.14	5.93	39.03	**28.58**	3.45
w/ idt	IrT2	53.12	36.53	15.39	31.65	8.12	3.86	33.18	10.80	5.91
Siamese GcGAN	IrT1	48.85	27.32	25.41	34.44	11.69	2.75	33.91	27.61	**30.61**
w/ idt	IrT2	53.81	30.59	21.10	32.49	7.45	15.52	24.76	8.35	6.08
Siamese GcGAN	IrT1	61.54	49.97	**41.33**	31.71	**22.04**	13.49	**43.84**	23.85	18.78
w/ sem idt	IrT2	62.72	49.22	32.07	**38.95**	20.52	**18.10**	35.73	17.12	7.72
Siamese GcGAN	IrT1	64.10	**52.33**	41.13	33.36	21.43	8.40	41.74	20.52	14.16
w/ sem idt, w/ p.p.	IrT2	**64.98**	52.01	36.66	36.62	21.36	10.55	38.99	24.19	8.70
LeNet (Src Only)	-	63.27	51.94	27.46	31.78	21.04	4.75	39.73	24.72	11.58

Table 2: The accuracy table of X-MNIST benchmark circularly training with one ToI dataset and two IrT datasets, FashionMNIST_C/E/N and EMNIST_C/E/N. Each entry is the averaged value of runs with four different seeds. Note that IrT1=EMNIST, IrT2=FashionMNIST when ToI=MNIST, IrT1=MNIST, IrT2=FashionMNIST when ToI=EMNIST, IrT1=EMNIST, IrT2=MNIST when ToI=FashionMNIST. The proposed model successfully adapt to all three target domains by only using corresponding target domain IrT samples. Note that the λ_{pp}=0.5 in all IrT-ToI pairs except for the case EMNIST→MNIST, λ_{pp}=0.9

the quality of translations between the domains with distinct tasks and helps to achieve better accuracy. Especially for EMNIST task, the siamese GcGAN with the semantic identity mapping ourperforms the source only performance by +7.17%, +1.00%, +8.74%, to G→C/E/N, respectively. Also, the post-processing step is helpful in MNIST and EMNIST task, as it shows better accuracy compared to the source only performance in all cases of MNIST and EMNIST tasks, with the margins ranging from +0.39% to +13.87%.

WIFI CSI human activity recognition benchmark From Table 3, the proposed model with postprocessing surpasses source only model and adapt to novel domain successfully except different topology domain. The accuracy gain is +10.27% on different channel and +20.41% on different rooms, respectively.

The source only model shows robust performance to topology change and the siamese GcGAN fails to adapt. However, in listening channel changed domain and room changed domain, even though the source only model performs much worse than the topology changed domain, our method outperforms the source only model with large margin.

Ablation Study Siamese generator The proposed siamese GcGAN shows the stable accuracy while the GcGAN decreases sharply, during the epochs from Figure 7 in Supplementary Material section. It is natural that it fails to learn a good target→source translation mapping preserving discriminability since given source and target datasets have no overlapping classes. GcGAN trained for 10 epochs produces lower test accuracy images by -21.68% compared to the initial epoch, while the siamese GcGAN by -10.01%. Furthermore, the siamese GcGAN

	Bldg1 Room A (0.6761)		
	different topology	different channel	different room
Siamese GcGAN w/ sem idt	0.4037	0.3858	0.3667
Siamese GcGAN w/ sem idt w/ pp	0.4719	**0.4605**	**0.4222**
MobileNetv2 (Src Only)	**0.5367**	0.3578	0.2181

Table 3: The accuracy table on WIFI CSI-based human activity recognition in Bldg1 Room A to domains created by changing topology, listening channel and room. All runs are averaged with three random seeds. Except the domain with different topology, our model successfully adapt to target domains in source IrT-free ZSDA setting. λ_{pp} is fixed to 0.5 in all runs.

with semantic identity mapping for 10 epochs is only decreased by -6.30%. From this observation, we can say the siamese generator helps the generator learns the mapping from target IrT dataset and source ToI dataset by minimizing discriminability lost.

Semantic identity mapping Table 2 shows there are large gaps before and after applying semantic identity mapping. It means the semantic identity mapping has the key role to make the translated images keeping its original task-related information.

6 Conclusion

We propose a novel source IrT-free zero-shot domain adaptation scheme. Different from conventional zero-shot domain adaptation, it guarantees a successful adaptation by collecting only target domain irrelevant task-of-interest samples samples and thus no paired condition is required.

Then, we propose the siamese generator architecture and semantic identity mapping. The siamese GcGAN helps the model factorize features into task-specific and domain-specific features. We integrate additional constraints to decouple domain-related and task-related information so that the learned image translation mapping be the task-independent. The semantic identity mapping prevents the siamese GcGAN lose the semantically meaningful information for the target ToI task. The ablation study supports both of siamese GcGAN and semantic identity mapping are the essential components for successful adaptation.

Our evaluation is done in conventional MNIST synthetic benchmark and the manually collected WIFI CSI-based human activity recognition benchmark. The quantitative results show successful adaptation in both of the benchmarks and the margin from source only accuracy is large particularly for the pairs showing low source only model accuracy.

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 898–916 (May 2011). https://doi.org/10.1109/TPAMI.2010.161, http://dx.doi.org/10.1109/TPAMI.2010.161
2. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1), 151–175 (2010)
3. Benaim, S., Wolf, L.: One-sided unsupervised domain mapping. In: NIPS (2017)
4. Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3722–3731 (2017)
5. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
6. Cohen, G., Afshar, S., Tapson, J., Van Schaik, A.: Emnist: Extending mnist to handwritten letters. In: 2017 International Joint Conference on Neural Networks (IJCNN). pp. 2921–2926. IEEE (2017)
7. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: International conference on machine learning. pp. 2990–2999. PMLR (2016)
8. Fu, H., Gong, M., Wang, C., Batmanghelich, K., Zhang, K., Tao, D.: Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2427–2436 (2019)
9. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015)
10. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. The journal of machine learning research **17**(1), 2096–2030 (2016)
11. Grother, P.J.: Nist special database 19-hand-printed forms and characters database. Technical Report, National Institute of Standards and Technology (1995)
12. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning. pp. 1989–1998. PMLR (2018)
13. Jhoo, W.Y., Heo, J.P.: Collaborative learning with disentangled features for zero-shot domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8896–8905 (2021)
14. Jiang, W., Miao, C., Ma, F., Yao, S., Wang, Y., Yuan, Y., Xue, H., Song, C., Ma, X., Koutsonikolas, D., et al.: Towards environment independent device free human activity recognition. In: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. pp. 289–304 (2018)
15. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
16. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 29. Curran Associates, Inc. (2016), https://proceedings.neurips.cc/paper/2016/file/502e4a16930e414107ee22b6198c578f-Paper.pdf
17. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International conference on machine learning. pp. 97–105. PMLR (2015)

18. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. arXiv preprint arXiv:1705.10667 (2017)
19. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Unsupervised domain adaptation with residual transfer networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. pp. 136–144 (2016)
20. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Deep transfer learning with joint adaptation networks. In: International conference on machine learning. pp. 2208–2217. PMLR (2017)
21. Luo, Y., Liu, P., Guan, T., Yu, J., Yang, Y.: Adversarial style mining for one-shot unsupervised domain adaptation. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 20612–20623. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/ed265bc903a5a097f61d3ec064d96d2e-Paper.pdf
22. Motiian, S., Jones, Q., Iranmanesh, S., Doretto, G.: Few-shot adversarial domain adaptation. Advances in Neural Information Processing Systems **30**, 6670–6680 (2017)
23. Peng, K.C., Wu, Z., Ernst, J.: Zero-shot deep domain adaptation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 764–781 (2018)
24. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3723–3732 (2018)
25. Schulz, M., Wegemer, D., Hollick, M.: Nexmon: The c-based firmware patching framework (2017), https://nexmon.org
26. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: European conference on computer vision. pp. 443–450. Springer (2016)
27. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation (2016)
28. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7167–7176 (2017)
29. Wang, J., Jiang, J.: Conditional coupled generative adversarial networks for zero-shot domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3375–3384 (2019)
30. Wang, J., Jiang, J.: Adversarial learning for zero-shot domain adaptation. In: European Conference on Computer Vision. pp. 329–344. Springer (2020)
31. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. Neurocomputing **312**, 135–153 (2018)
32. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of wifi signal based human activity recognition. In: Proceedings of the 21st annual international conference on mobile computing and networking. pp. 65–76 (2015)
33. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
34. Xue, H., Jiang, W., Miao, C., Ma, F., Wang, S., Yuan, Y., Yao, S., Zhang, A., Su, L.: Deepmv: Multi-view deep learning for device-free human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **4**(1), 1–26 (2020)
35. Yousefi, S., Narui, H., Dayal, S., Ermon, S., Valaee, S.: A survey on behavior recognition using wifi channel state information. IEEE Communications Magazine **55**(10), 98–104 (2017)

36. Zheng, Y., Zhang, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., Yang, Z.: Zero-effort cross-domain gesture recognition with wi-fi. In: Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services. pp. 313–325 (2019)
37. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)