RDA: Robust Domain Adaptation via Fourier Adversarial Attacking

Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu*

Singtel Cognitive and Artificial Intelligence Lab for Enterprises,
School of Computer Science and Engineering, Nanyang Technological University

{Jiaxing.Huang, Dayan.Guan, Aoran.Xiao, Shijian.Lu}@ntu.edu.sg

Abstract

Unsupervised domain adaptation (UDA) involves a supervised loss in a labeled source domain and an unsupervised loss in an unlabeled target domain, which often faces more severe overfitting (than classical supervised learning) as the supervised source loss has clear domain gap and the unsupervised target loss is often noisy due to the lack of annotations. This paper presents RDA, a robust domain adaptation technique that introduces adversarial attacking to mitigate overfitting in UDA. We achieve robust domain adaptation by a novel Fourier adversarial attacking (FAA) method that allows large magnitude of perturbation noises but has minimal modification of image semantics, the former is critical to the effectiveness of its generated adversarial samples due to the existence of 'domain gaps'. Specifically, FAA decomposes images into multiple frequency components (FCs) and generates adversarial samples by just perturbating certain FCs that capture little semantic information. With FAA-generated samples, the training can continue the 'random walk' and drift into an area with a flat loss landscape, leading to more robust domain adaptation. Extensive experiments over multiple domain adaptation tasks show that RDA can work with different computer vision tasks with superior performance.

1. Introduction

Deep convolutional neural networks (CNNs) [39, 70, 21] have defined new state of the arts in various computer vision tasks [7, 44, 60, 20, 39, 70, 21], but their trained models often over-fit to the training data and experience clear performance drops for data from different sources due to the existence of *domain gaps*. Unsupervised domain adaptation (UDA) has been investigated to address the *domain gaps* by leveraging unlabeled target data. To this end, most existing UDA works [34, 35, 75, 78, 48, 80, 8, 9] involve supervised losses on source data and unsupervised losses on target data
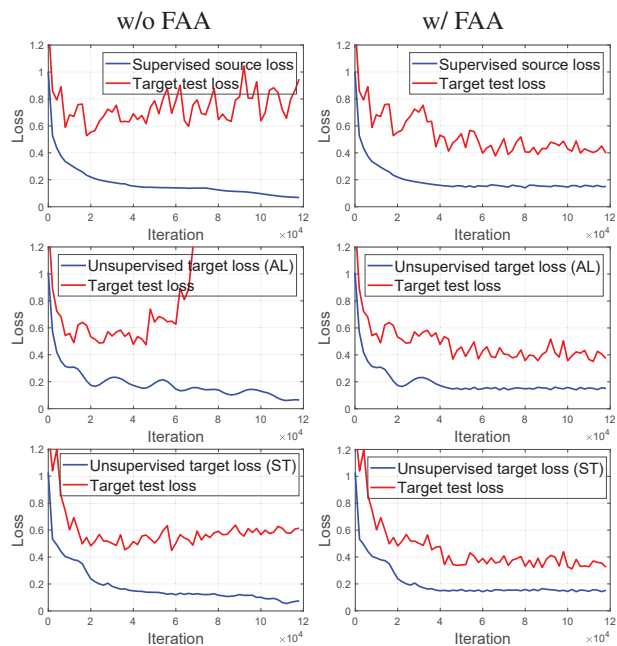
*Corresponding author.

Figure 1. Our robust domain adaptation alleviates overfitting effectively: Both supervised learning with source data (row 1) and unsupervised learning with target data (in rows 2 and 3 for adversarial learning and self-training) in unsupervised domain adaptation suffer from clear overfitting as illustrated by decreased training losses (blue curves) vs increased target test losses (red curves) in column 1. Our Fourier adversarial attacking (FAA) generates novel adversarial samples, which regulates the minimization of training losses and alleviates overfitting effectively with decreased target test loss as shown in column 2 (Best viewed in color).

for learning a model that performs well in target domains. However, as illustrated in Fig. 1, these methods often face more severe overfitting (as compared with the classical supervised learning) as supervised source losses in UDA has an extra *domain gap* (for test data in target domains) and unsupervised target losses are often noisy due to the lack of annotations.

Overfitting exists in almost all deep network training, which is undesired and often degrades the generalization of

the trained deep network models while applied to new data. One way of identifying whether overfitting is happening is to check whether the generalization gap, *i.e.*, the difference between the test loss and the training loss, is increasing or not [13] as shown in Fig. 1. Various strategies have been investigated to alleviate overfitting through weight regularization [19], dropout [71], mixup [93], label smoothing [72], batch normalization [30], etc. However, all these strategies were designed for supervised and semi-supervised learning where training data and test data usually have very similar distributions. For domain adaptive learning, they do not fit in well due to the negligence of domain gaps that widely exist between data of different domains.

We design a robust domain adaptation technique that introduces a novel Fourier adversarial attacking (FAA) technique to mitigate the overfitting in unsupervised domain adaptation. FAA mitigates overfitting by generating adversarial samples that prevent over-minimization of supervised and unsupervised UDA losses as illustrated in Fig. 1. Specifically, FAA decomposes training images into multiple frequency components (FCs) and only perturbs FCs that capture little semantic information. Unlike traditional attacking that restricts the magnitude of perturbation noises to keep image semantics intact, FAA allows large magnitude of perturbation in its generated adversarial samples but has minimal modification of image semantics. This feature is critical to unsupervised domain adaptation which usually involves clear domain gaps and so requires adversarial sample with large perturbations. By introducing the FAA-generated adversarial samples in training, networks can continue the "random walk" and avoid over-fitting and drift into an area with a flat loss landscape [6, 36, 40], leading to more robust domain adaptation.

The contributions of this work can be summarized in three aspects. *First*, we identify the overfitting issue in unsupervised domain adaptation and introduce adversarial attacking to mitigate overfitting by preventing training objectives from over-minimization. *Second*, we design an innovative Fourier Adversarial Attacking (FAA) technique to generate novel adversarial samples to mitigate overfitting in domain adaptation. FAA is generic which can work for both supervised source loss and unsupervised target losses effectively. *Third*, we conducted extensive experiments over multiple computer vision tasks in semantic segmentation, object detection and image classification. All experiments show that our method mitigates overfitting and improve domain adaption consistently.

2. Related Works

Domain Adaptation: Domain adaptation has been studied extensively for mitigating data annotation constraints. Most existing works can be broadly classified into three categories. The first category is *adversarial training* based

which employs a discriminator to align source and target domains in the feature, output or latent space [46, 78, 48, 75, 67, 24, 80, 76, 25, 88, 95, 18, 28]. The second category is *image translation* based which adapt image appearance to mitigate domain gaps [41, 92, 90, 94]. The third category is *self-training* based which predict pseudo labels or minimize entropy to guide iterative learning from target samples [99, 98, 49, 16, 26, 17].

Domain adaptation involves two typical training losses, namely, supervised loss over labeled source data and unsupervised loss over unlabeled target data. State-of-the-art methods tend to over-minimize the two types of losses which directly leads to deviated models with suboptimal adaptation as illustrated in Fig. 1. We design a robust domain adaptation technique that addresses this issue by preventing loss over-minimization.

Overfitting in Network Training: Overfitting is a common phenomenon in deep network training which has been widely studied in the deep learning and computer vision community [2, 5, 54, 62, 84, 82]. Most existing works address overfitting by certain regularization strategies such as weight decay [19], dropout [71], l_1 regularization [74], mixup [93], label smoothing [72], batch normalization [30], virtual adversarial training [51], flooding [31], etc. However, these strategies were mostly designed for supervised or semi-supervised learning which do not fit in well for domain adaptive learning that usually involves domain gaps and unsupervised losses. We design a adversarial attacking technique that mitigates the overfitting in domain adaptive learning effectively.

Adversarial Attacking: Adversarial attacking has been studied in various security problems. For example, [73] shows that adversarial samples can easily confuse CNN models. The following works improved adversarial attacking from different aspects via fast gradient signs [14], minimal adversarial perturbation [53], universal adversarial perturbations [52], gradient-free attacking [1], transferable adversarial sample generation[43], etc. Adversarial attacking has also been applied to other tasks, e.g., [96, 51] employed adversarial samples to mitigate over-fitting in supervised and semi-supervised learning, [79] generated adversarial samples for data augmentation, and [42, 89] augments transferable features for domain divergence minimization.

Most existing adversarial attacking methods commonly constrain the magnitude of perturbation noises for minimal modification of image semantics. However, such generated adversarial samples cannot tackle overfitting in domain adaptive learning well which usually involves a domain gap of fair magnitude. We design an innovative Fourier adversarial attacking technique that allows to generate adversarial samples without magnitude constraint yet with minimal modification of image semantics, more details to be described in the ensuing subsections.

3. Method

We achieve robust domain adaptation via Fourier adversarial attacking as illustrated in Fig. 2. The training consists of two phases, namely, an *Attacking Phase* and a *Defending Phase*. Given a training image, the attacking phase learns to identify the right FCs with limited semantic information that allow perturbation noises of large magnitude. It also learns to generate adversarial samples (with perturbable FCs) with minimal modification of image semantics. During the defending phase, the generated adversarial sample is applied to mitigate overfitting by preventing over-minimization of training losses, more details to be described in the ensuing subsections.

3.1. Task Definition

We focus on the problem of unsupervised domain adaptation (UDA). Given labeled source data $\{X_s, Y_s\}$ and unlabeled target data X_t, our goal is to learn a task model F that performs well on X_t. The *baseline* model is trained with the labeled source data only:

$$\mathcal{L}_{sup}(X_s, Y_s; F) = l(F(X_s), Y_s), \tag{1}$$

where $l(\cdot)$ denotes an accuracy-related loss, *e.g.*, the standard cross-entropy loss.

3.2. Fourier Adversarial Attacking

Our proposed Fourier adversarial attacking (FAA) generates adversarial samples to attack the training loss to mitigate overfitting in domain adaptation, as shown in Fig. 2. In adversarial sample generation, it first employs Fourier transformation to decompose input images into multiple frequency components (FCs) and then inject perturbation to non-semantic FCs which allows perturbation of large magnitude but with minimal modification of image semantics.

Fourier Decomposition: Inspired by JPEG [81, 58] and frequency-domain learning [27, 33], we transform an image x into frequency space and decompose it into multiple FCs which allows explicit manipulation of each FC and more controllable perturbations. We employ Fourier transformation to convert x into frequency space and further decompose it into multiple FCs of equivalent bandwidth:

$$z = \mathcal{F}(x),$$
$$z^{fc} = \mathcal{D}(z; N),$$

where $\mathcal{F}(\cdot)$ stands for Fourier transformation [4]; z denotes the frequency-space representation of x; $\mathcal{D}(z; N)$ denotes a function that decomposes z into N FCs $z^{fc} = \{z^1, z^2, ..., z^{N-1}, z^N\}$ of equivalent bandwidth.

We consider a gray-scale image $x \in \mathbb{R}^{H \times H}$ for defining $\mathcal{D}(z; N)$ formally. We thus have $z \in \mathbb{C}^{H \times H}$, where $\mathbb{C}$ denotes complex numbers. We use $z(i, j)$ to index z at (i, j),

and c_i and c_j to denote the image centroid (*i.e.*, image center at $(H/2, H/2)$). The equation $\{z^1, z^2, ..., z^{N-1}, z^N\} = \mathcal{D}(z; N)$ can be formally defined by:

$$z^n(i, j) = \begin{cases} z(i, j), & \text{if } \frac{n-1}{N} < d((i, j), (c_i, c_j)) < \frac{n}{N}, \\ 0, & \text{otherwise}, \end{cases}$$

where $d(\cdot, \cdot)$ denotes Euclidean distance, N denotes how many FCs the input is supposed to be decomposed into and n denotes the FC index.

To get $z^{fc} = \{z^1, z^2, ..., z^{N-1}, z^N\} = \mathcal{D}(z; N)$, we perform the equation N times by changing n from 1 to N. If x does not have a square size, we first up-sample the short side to be same as the long side before this process and down-sample it back to the original size after processing. If x has more than one channel (*e.g.*, RGB image $x \subset \mathbb{R}^{n \times n \times 3}$), we process each channel independently.

Adversarial Attacking: With the decomposed FCs $z^{fc} = \{z^1, z^2, ..., z^{N-1}, z^N\}$, we attack domain adaption losses by perturbing partial FCs without magnitude constraint. Specifically, we employ a learnable gate module [32, 77, 87, 50] to select certain FCs for perturbation. This gate G assigns a binary score to each FC, where '1' indicates this FC is selected for perturbation (multiplied by '1') while '0' shows this FC is discarded (reset to all zero values). G works with Gumbel-Softmax, a differentiable sampling mechanism for categorical variables that can be trained via standard back-propagation. Please refer to [32, 77, 87, 50] for details.

Given a reference image (an randomly selected target-domain image) in FC representation $z_{ref}^{fc} = \{z_{ref}^1, z_{ref}^2, ..., z_{ref}^{N-1}, z_{ref}^N\} = \mathcal{D}(\mathcal{F}(x_{ref}); N)$, we employ this gate module to block some FCs (all reset to zero), and apply the selected FCs to perturb the corresponding FCs of the input image x:

$$\hat{z}^{fc} = (1 - G(z^{fc}))z^{fc} + G(z^{fc})z_{ref}^{fc}, \tag{2}$$

where the learnable gate module G enables binary channel attention that selects which FCs to perturb.

Specifically, for the identified FCs of the input image x, we extract the corresponding FCs of a reference image X_{ref} and use them as perturbation noises. We generate adversarial samples in this manner because the perturbation noises (*i.e.*, non-semantic FCs of z_{ref}^{fc}) from target natural images are more reasonable and meaningful as compared to a random noises/signals as in many existing adversarial sample generation methods. In addition, the use of non-semantic FCs of target samples mitigates inter-domain gaps which helps to improve target-domain performance in domain adaptation.

With the perturbed FCs $\hat{z}^{fc}$, we convert them all back (via inverse Fourier transformation) to get the adversarial sample x^{FAA}:

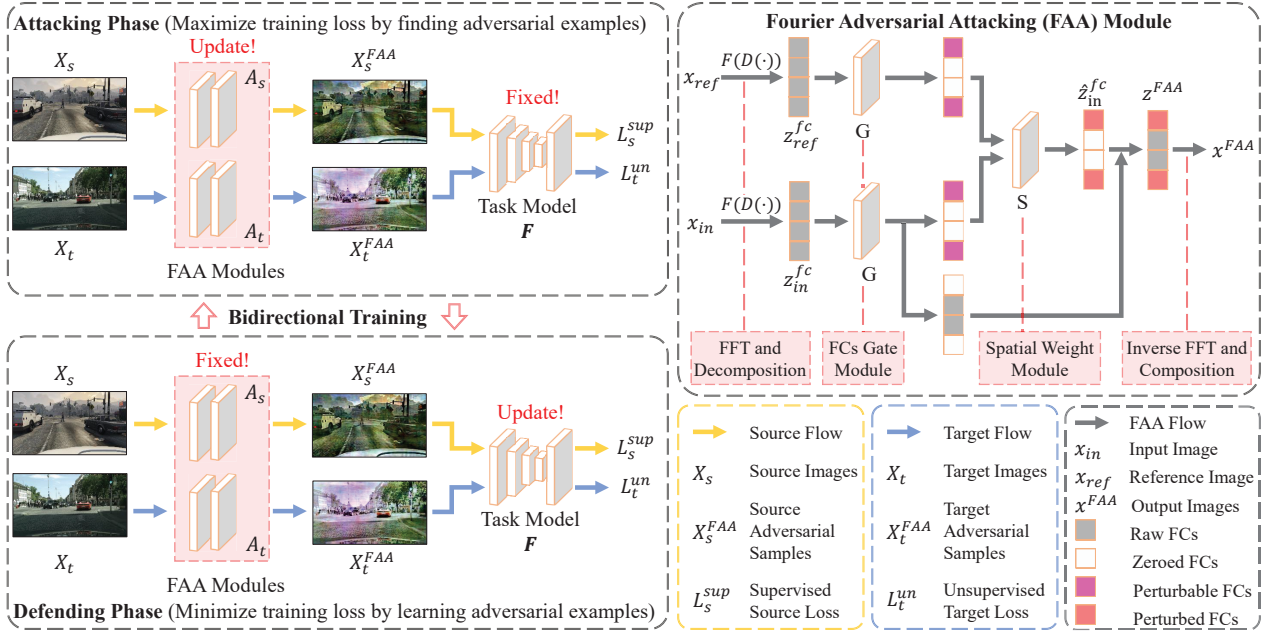$$x^{FAA} = \mathcal{F}^{-1}(\mathcal{C}(\hat{z}^{fc})), \tag{3}$$

Figure 2. Overview of our proposed robust domain adaptaion (RDA) via Fourier Adversarial Attacking (FAA). RDA has a bidirectional training process which consists of an *Attacking Phase* and a *Defending Phase*. Given a training image, the attacking phase learns to generate adversarial samples with minimal modification of image semantics. During the defending phase, the generated adversarial samples are applied to mitigate overfitting by preventing over-minimization of training losses. Given an input image, FAA decomposes it into multiple FCs representation, identifies perturbable FCs, and then perturbs them through weighted sum with the corresponding FCs from a randomly picked reference image. The objective of FFA is to maximize task losses while ensuring minimal modification of image semantics.

where $\mathcal{C}(\cdot)$ denotes an inverse process of $\mathcal{D}(\cdot; N)$ that 'compose' the decomposed z^{fc} back to the full representation by summing all the elements over frequency channels.

The *Fourier Decomposition* and *Adversarial Attacking* can be combined to form the FAA as follows:

$$x^{FAA} = \mathcal{A}(x), \tag{4}$$

where $\mathcal{A}$ takes an image x and outputs an adversarial sample x^{FAA} via the FAA as described. $\mathcal{A}$ has two sub-modules with learnable parameters, *i.e.*, the gate module G and the single-layer neural network for spatial weighting map $\mathcal{S}$ generation. The rest operations in $\mathcal{A}$ are deterministic such as Fourier transformation and its inverse ($\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$), decomposition and re-composition ($\mathcal{D}(\cdot; N)$ and $\mathcal{C}(\cdot)$)

3.3. FAA Training

The proposed FAA involves three types of losses including task loss (*i.e.*, UDA training loss in this work) that is to be attacked, the gate related loss that constrains the proportion of perturbable FCs, and the reconstruction loss that aims to minimize the attacking effects over image semantics.

Given input images $X \subset \mathbb{R}^{H \times W \times 3}$, a task model F and the Fourier attacker $\mathcal{A}$, we denote the task loss by $\mathcal{L}(X; \mathcal{A}, F)$. Note the task loss refers to UDA training loss

in this work but it can be other types of losses such as supervised or unsupervised training losses. The gate related loss and the reconstruction loss losses are defined by:

$$\mathcal{L}_{gat}(X; \mathcal{A}) = \begin{cases} \sum(G(z^{fc})), & \text{if } \sum(G(z^{fc})) > N * p, \\ 0, & \text{otherwise,} \end{cases}$$
$$\mathcal{L}_{rec}(X; \mathcal{A}) = |R(X) - R(\mathcal{A}(X))|,$$

where $G(z^{fc})$ is a gating process in $\mathcal{A}$ as described in Eq. 2. R is a band-pass filter to get the mid-frequency content [27] that captures semantic information (*e.g.*, structures and shapes) and thus the consistency loss can ensure that the selected FCs contains minimal semantic information. p is a hyper-parameter that constrains the maximum number of perturbable FCs by $N * p$,

The overall training objective of FAA is formulated by:

$$\max_{\mathcal{A}} \mathcal{L}(X; \mathcal{A}, F) - \mathcal{L}_{gat}(X; \mathcal{A}) - \mathcal{L}_{rec}(X; \mathcal{A}) \tag{5}$$

3.4. Robust Domain Adaptation

Sections 3.2 and 3.3 describe the "Attacking Phase" that generates adversarial samples via FAA. This section presents the proposed robust domain adaptation technique where FAA is used to mitigate the overfitting in UDA in a "Defending Phase". Specifically, we apply the

Algorithm 1 The proposed Robust Domain Adaptation via Fourier adversarial attacking (FAA).

Require: Input data X; task model F; Fourier attacker $\mathcal{A}$
Ensure: Learnt parameters θ of task model F

1: **for** $iter = 1$ **to** Max_Iter **do**
2:     Sample a batch of X uniformly from the dataset
3:     *Defending phase (minimize loss):*
4:     Update F via Eq. 6
5:     *Attacking phase (maximize loss):*
6:     Update $\mathcal{A}$ via Eq. 5
7: **end for**
8: **return** F

FAA-generated adversarial samples to prevent the over-minimization of UDA training losses in domain adaptation.

Given the task model F and the Fourier attacker $\mathcal{A}$, the training objective of the task model F is formulated by:

$$\min_F \mathcal{L}(X; \mathcal{A}, F) \qquad (6)$$

The optimization functions of both attacking (*i.e.*, Eq. 5) and defending (*i.e.*, Eq. 6) are generic and applicable to various *tasks* and *data*. Specifically, the model F could be for semantic segmentation, object detection or image classification task. The data could be labeled source data with supervised source losses or unlabeled target data with unsupervised target losses such as adversarial loss [75, 48, 80, 89, 55, 42, 29, 66], self-training loss [99, 98, 41, 83, 37, 90] or entropy loss [80, 15], etc. Below are a supervised loss and a self-training-based unsupervised loss here for reference:

$$\mathcal{L}_s^{sup}(X_s; \mathcal{A}, F) = l(F(\mathcal{A}(X_s)), Y_s),$$
$$\mathcal{L}_t^{un}(X_t; \mathcal{A}, F) = l(F(\mathcal{A}(X_t)), \hat{Y}_t),$$

where $l(\cdot)$ denotes an accuracy-related loss, *e.g.*, cross entropy loss; $\mathcal{A}(X_s)$ and $\mathcal{A}(X_t)$ are inputs perturbed by FAA, and $\hat{Y}_t$ is the pseudo label of unlabeled target data.

In summary, our robust domain adaptation has a bidirectional training framework including an *Attacking Phase* and a *Defending Phase* as shown in Fig. 2. In the *Attacking Phase*, the task model F is fixed and the attacker $\mathcal{A}$ generates adversarial samples to increase the training loss. In the *Defending Phase*, $\mathcal{A}$ is fixed and F is updated to reduce the training loss. These two phases are conducted in an alternative way way and form a bidirectional training framework (adversarial training). Please refer to Algorithm 1 for details.

This bidirectional training framework uses FAA to prevent over-minimization of UDA losses by forcing them oscillating around a small value. In another word, it ensures that the task model F can continue the "random walk" and drift into an area with a flat loss landscape [6, 36, 40], leading to robust domain adaptation.

4. Experiments

This section presents experiments including datasets and implementation details, domain adaptation for semantic segmentation (with ablation studies), object detection, and image classification tasks, and discussion, more details to be described in the ensuing subsections.

4.1. Datasets

Adaptation for semantic segmentation: We consider two synthesized-to-real segmentation tasks: 1) GTA5 [61] → Cityscapes [10] and 2) SYNTHIA [63] → Cityscapes. GTA5 contains $24,966$ synthetic images and shares 19 categories with Cityscapes. For SYNTHIA, we use 'SYNTHIA-RAND-CITYSCAPES' which contains $9,400$ synthetic images and shares 16 categories with Cityscapes. Cityscapes contains 2975/500 training/validation images. Following [75, 99], we adapt towards the Cityscapes training set and evaluate on the Cityscapes validation set.

Adaptation for object detection: We consider two adaptation tasks: 1) Cityscapes → Foggy Cityscapes [68] and 2) Cityscapes → BDD100k [91]. For Cityscapes, we convert instance segmentation annotations to bounding boxes in experiments. Foggy Cityscapes was created by applying synthetic fog on Cityscapes images. BDD100k contains $100k$ images including $70k$ for training and $10k$ for validation. Following [86, 66, 9], we use a BDD100k subset *daytime* in experiments, which includes $36,728$ training images and $5,258$ validation images.

Adaptation for image classification: We adopt two adaptation benchmarks VisDA17 [57] and Office-31 [64]. VisDA17 includes $152,409$ synthetic images of 12 categories as source and $55,400$ real images as target. Office-31 contains images of 31 classes from Amazon (A), Webcam (W) and DSLR (D) that have 2817, 795 and 498 images, respectively. We evaluate on six tasks A→W, D→W, W→D, A→D, D→A, and W→A as in [98, 64, 69].

4.2. Implementation Details

Semantic segmentation: We use DeepLab-V2 [7] with ResNet101 [21] as the segmentation network as in [75, 99]. We use SGD optimizer [3] with a momentum 0.9 and a weight decay $1e-4$. The initial learning rate is $2.5e-4$ and decayed by a polynomial policy of power 0.9 [7].

Object detection: As in [86, 66, 9], we use Faster R-CNN [7] with VGG-16 [70] as the detection network. We use SGD optimizer [3] with a momentum 0.9 and a weight decay $5e-4$. The initial learning rate is $1e-3$ for $50k$ iterations and reduced to $1e-4$ for another $20k$ iterations [86, 66, 9]. In all experiments, we set image shorter side to 600 and employ RoIAlign [20] for feature extraction.

Image classification: For fair comparisons, we follow [98, 64, 69] and use ResNet-101/ResNet-50 [21] (pre-

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
Baseline [21]	75.8	16.8	77.2	12.5	21.0	25.5	30.1	20.1	81.3	24.6	70.3	53.8	26.4	49.9	17.2	25.9	6.5	25.3	36.0	36.6
+FAA-S	89.8	39.1	81.7	27.6	19.9	34.2	35.9	23.3	82.1	29.5	76.6	58.3	26.0	82.1	32.5	45.2	15.3	26.9	33.5	**45.2**
AdaptSeg [75]	86.5	36.0	79.9	23.4	23.3	23.9	35.2	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.4
+FAA-S	89.1	38.7	82.0	28.5	23.6	33.2	34.9	22.2	83.7	34.0	77.5	59.0	26.6	83.5	33.8	45.8	15.8	28.3	33.9	46.0
+FAA-T	91.0	44.3	82.3	28.1	21.7	35.6	36.8	18.7	84.2	38.3	73.7	61.4	28.5	85.9	38.6	45.7	14.9	32.4	33.3	47.1
+FAA	91.2	44.7	82.5	27.7	24.7	36.7	36.5	24.7	84.9	38.1	78.4	62.2	27.4	84.9	39.3	46.7	12.9	31.9	36.7	**48.0**
ST [99]	88.5	26.3	82.0	24.5	19.9	33.6	37.6	19.8	82.7	26.2	76.8	60.0	26.6	82.9	28.9	31.0	6.3	24.5	31.2	42.6
+FAA-S	91.4	52.1	81.6	32.6	26.4	35.3	36.2	22.5	84.6	35.8	77.3	60.5	27.9	84.1	37.0	43.8	17.2	25.6	35.9	47.8
+FAA-T	91.6	49.0	83.2	32.7	26.5	38.9	46.5	23.6	83.2	35.3	77.9	60.1	28.5	85.5	38.1	41.4	16.9	25.6	36.4	48.5
+FAA	92.0	52.6	82.5	33.9	28.5	39.4	45.7	33.6	84.3	38.4	80.5	60.4	29.1	85.4	39.9	44.4	16.2	29.6	35.1	**50.1**
CLAN [48]	87.0	27.1	79.6	27.3	23.3	28.3	35.5	24.2	83.6	27.4	74.2	58.6	28.0	76.2	33.1	36.7	6.7	31.9	31.4	43.2
+FAA	92.0	53.4	82.8	32.5	29.8	36.8	36.6	28.5	83.3	35.8	77.1	61.2	30.1	83.8	36.8	46.2	11.7	28.6	35.6	**48.6**
AdvEnt [80]	89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
+FAA	92.1	49.0	82.5	31.2	28.6	38.3	37.1	27.1	84.8	39.4	79.1	61.3	29.1	85.0	38.2	46.6	14.3	32.4	36.9	**49.1**
IDA [56]	90.6	37.1	82.6	30.1	19.1	29.5	32.4	20.6	85.7	40.5	79.7	58.7	31.1	86.3	31.5	48.3	0.0	30.2	35.8	46.3
+FAA	91.4	50.3	83.4	33.0	27.2	37.7	38.4	27.0	84.4	41.8	79.9	59.7	30.6	85.1	37.7	47.5	14.2	31.2	35.3	**49.3**
CRST [98]	91.0	55.4	80.0	33.7	21.4	37.3	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	26.9	26.0	42.3	47.1
+FAA	92.7	56.4	84.4	34.3	29.2	38.2	48.8	47.1	85.5	42.4	80.6	62.3	32.9	85.3	40.8	48.9	22.7	23.0	36.8	**52.6**
BDL [41]	91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	43.6	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5
+FAA	92.8	54.1	85.3	33.8	28.1	41.0	46.1	47.6	84.8	42.7	82.8	63.5	32.5	84.1	37.0	50.3	15.5	23.2	36.5	**51.7**
CrCDA [29]	92.4	55.3	82.3	31.2	29.1	32.5	33.2	35.6	83.5	34.8	84.2	58.9	32.2	84.7	40.6	46.1	2.1	31.1	32.7	48.6
+FAA	92.9	55.2	83.4	31.8	31.1	39.4	47.8	46.4	83.4	38.4	85.7	61.2	31.0	85.7	39.1	46.3	12.0	33.4	38.9	**51.7**
SIM [83]	90.6	44.7	84.8	34.3	28.7	31.6	35.0	37.6	84.7	43.3	85.3	57.0	31.5	83.8	42.6	48.5	1.9	30.4	39.0	49.2
+FAA	92.2	53.7	84.6	34.2	29.1	38.0	47.0	45.3	85.0	43.8	84.9	60.3	29.2	85.2	39.5	47.3	12.4	32.3	44.3	**52.0**
TIR [37]	92.9	55.0	85.3	34.2	31.1	34.9	40.7	34.0	85.2	40.1	87.1	61.0	31.1	82.5	32.3	42.9	0.3	36.4	46.1	50.2
+FAA	93.0	55.5	84.0	33.0	28.2	38.1	46.6	45.8	84.9	41.7	86.1	61.2	33.7	84.2	38.3	46.5	15.3	34.9	44.6	**52.4**
FDA [90]	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	53.1	16.9	27.7	46.4	50.5
+FAA	93.1	55.0	84.7	33.1	29.5	38.7	49.3	44.9	84.8	41.6	80.2	62.3	33.2	85.6	37.3	51.3	18.5	34.6	45.3	**52.8**

Table 1. Domain adaptive semantic segmentation experiments over task GTA5 → Cityscapes (+*FAA-S*, +*FAA-T*, and +*FAA* mean to apply our FAA to attack supervised source losses, unsupervised target losses, and both types of losses, respectively).

Method	Road	SW	Build	Wall*	Fence*	Pole*	TL	TS	Veg.	Sky	PR	Rider	Car	Bus	Motor	Bike	mIoU	mIoU*
Baseline [21]	55.6	23.8	74.6	9.2	0.2	24.4	6.1	12.1	74.8	79.0	55.3	19.1	39.6	23.3	13.7	25.0	33.5	38.6
PatAlign [76]	82.4	38.0	78.6	8.7	0.6	26.0	3.9	11.1	75.5	84.6	53.5	21.6	71.4	32.6	19.3	31.7	40.0	46.5
AdaptSeg [75]	84.3	42.7	77.5	-	-	-	4.7	7.0	77.9	82.5	54.3	21.0	72.3	32.2	18.9	32.3	-	46.7
CLAN [48]	81.3	37.0	80.1	-	-	-	16.1	13.7	80.2	81.5	53.4	21.2	73.0	32.9	22.6	30.7	-	47.8
AdvEnt [80]	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	84.1	57.9	23.8	73.3	36.4	14.2	33.0	41.2	48.0
IDA [56]	84.3	37.7	79.5	5.3	0.4	24.9	9.2	8.4	80.0	84.1	57.2	23.0	78.0	38.1	20.3	36.5	41.7	48.9
CrCDA [29]	86.2	44.9	79.5	8.3	0.7	27.8	9.4	11.8	78.6	86.5	57.2	26.1	76.8	39.9	21.5	32.1	42.9	50.0
TIR [37]	92.6	53.2	79.2	-	-	-	1.6	7.5	78.6	84.4	52.6	20.0	82.1	34.8	14.6	39.4	-	49.3
+FAA	89.1	51.3	78.7	8.9	0.1	27.1	19.1	17.0	82.4	83.9	55.8	27.7	84.5	39.6	26.9	38.3	45.7	**53.4**
CRST [98]	67.7	32.2	73.9	10.7	1.6	37.4	22.2	31.2	80.8	80.5	60.8	29.1	82.8	25.0	19.4	43.8	43.8	50.1
+FAA	89.3	49.6	79.3	12.9	0.1	37.2	20.6	29.8	83.7	84.0	62.6	28.0	84.9	37.1	30.1	48.2	48.6	**55.9**
BDL [41]	86.0	46.7	80.3	-	-	-	14.1	11.6	79.2	81.3	54.1	27.9	73.7	42.2	25.7	45.3	-	51.4
+FAA	89.7	47.5	82.7	6.2	0.1	31.8	25.0	22.7	82.3	84.8	59.0	26.2	83.4	39.2	31.5	37.9	46.9	**54.8**
SIM [83]	83.0	44.0	80.3	-	-	-	17.1	15.8	80.5	81.8	59.9	33.1	70.2	37.3	28.5	45.8	-	52.1
+FAA	88.3	47.8	82.8	9.3	1.1	36.7	24.5	26.8	79.4	83.7	62.5	32.1	81.7	36.3	29.2	46.4	48.0	**55.5**
FDA [90]	79.3	35.0	73.2	-	-	-	19.9	24.0	61.7	82.6	61.4	31.1	83.9	40.8	38.4	51.1	-	52.5
+FAA	89.4	39.0	79.8	8.1	1.2	33.0	22.6	28.1	81.8	82.0	60.9	30.1	82.7	41.4	37.5	48.9	47.9	**55.7**

Table 2. Domain adaptive semantic segmentation experiments over task SYNTHIA → Cityscapes (+*FAA* means to include FAA to attack supervised source loss and unsupervised target loss in domain adaptation).

trained with ImageNet [11]) as backbones. We use SGD optimizer [3] with a momentum 0.9 and a weight decay $5e-4$. The learning rate is $1e-3$ and the batch size is 32 [98].

We set the parameter p and the number of FCs N at 0.1 and 96. The band-pass filter R follows [27] with mid-pass and low-/high-rejected designs to get the mid-frequency content that captures semantic information (e.g., structures and shapes).

4.3. Domain Adaptive Semantic Segmentation

Table 1 shows experimental results over semantic segmentation task GTA5 → Cityscapes. It can be seen that FAA is generic and can be applied to attack both *Baseline* (for preventing overfitting in supervised source loss) and state-of-the-art UDA methods (for preventing overfitting in both supervised source loss and unsupervised target loss). In addition, incorporating FAA improves both *Baseline* and UDA methods clearly and consistently.

We perform ablation studies over two representative UDA methods using adversarial alignment [75] and self-training [99], where +*FAA-S*, +*FAA-T* and +*FAA* address the overfitting of supervised source losses, unsupervised target losses and both losses, respectively. It can be seen that +*FAA-S* and +*FAA-T* both improve domain adaptation by large margins. This shows that both supervised source objective and unsupervised target objective intro-

Method	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Baseline [60]	24.4	30.5	32.6	10.8	25.4	9.1	15.2	28.3	22.0
MAF [22]	28.4	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
SCDA [97]	33.5	38.0	48.5	26.5	39.0	23.3	28.0	33.6	33.8
DA [9]	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
MLDA [85]	33.2	44.2	44.8	28.2	41.8	28.7	30.5	36.5	36.0
DMA [38]	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
CAFA [23]	41.9	38.7	56.7	22.6	41.5	26.8	24.6	35.5	36.0
SWDA [66]	36.2	35.3	43.5	30.0	29.9	42.3	32.6	24.5	34.3
+FAA	39.5	41.3	47.0	34.5	39.3	44.0	31.9	28.4	**38.3**
CRDA [86]	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
+FAA	37.4	46.4	48.3	32.6	46.5	39.3	32.3	35.3	**39.8**

Table 3. Experimental results of domain adaptive object detection over the adaptation task Cityscapes → Foggy Cityscapes (*+FAA* means to include FAA to attack supervised source loss and unsupervised target loss in domain adaptation).

Method	person	rider	car	truck	bus	mcycle	bicycle	mAP
Baseline [60]	26.9	22.1	44.7	17.4	16.7	17.1	18.8	23.4
DA [9]	29.4	26.5	44.6	14.3	16.8	15.8	20.6	24.0
SWDA [66]	30.2	29.5	45.7	15.2	18.4	17.1	21.2	25.3
+FAA	32.0	32.2	49.7	19.8	23.9	18.5	24.0	**28.6**
CRDA [86]	31.4	31.3	46.3	19.5	18.9	17.3	23.8	26.9
+FAA	32.9	33.1	51.4	21.8	24.5	19.1	26.2	**29.9**

Table 4. Experimental results of domain adaptive object detection over the adaptation task Cityscapes → BDD100k (*+FAA* means to include FAA to attack supervised source loss and unsupervised target loss in domain adaptation).

duce clear overfitting and FAA mitigates the overfitting effectively. In addition, *+FAA* performs clearly the best. This shows that preventing the two learning objectives from over-minimization is complementary as overfitting in the two learning objectives affects generalization in different manners. Specifically, supervised source loss has domain gap and over-minimizing it guides the model to over-memorize source data whereas unsupervised target loss is noisy and over-minimizing it leads to deviated solutions with accumulated errors.

Table 2 shows experimental results over semantic segmentation task SYNTHIA → Cityscapes. We can observe that FAA improves state-of-the-art UDA methods in the similar manner as in Table 1. Note we applied FAA to a few representative UDA methods only due to space limit.

4.4. Domain Adaptive Object Detection

Table 3 shows domain adaptive object detection over the task Cityscapes → Foggy Cityscapes. It can be seen that FAA boosts mAP by over +2.5% for both SWDA [66] and CRDA [86]. Note that we did not apply FAA to other listed UDA methods due to space limit.

Table 4 shows domain adaptive object detection over the task Cityscapes → BDD100k. It can be seen that including FAA outperforms state-of-the-art UDA methods consistently as in Table 3. Note that we did not apply FAA to other listed UDA methods due to space limit.

4.5. Domain Adaptive Image Classification

We presents experimental results on VisDA17 in Table 5 in per-class accuracy. It can be seen that incorporating FAA outperforms state-of-the-art UDA methods consis-

tently. This applies to UDA methods that employ stronger backbones ResNet-152 [59, 69].

Table 6 shows domain adaptive image classification experiments over Office-31 (all using the same backbone ResNet-50). We can see that incorporating FAA leads to robust domain adaptation and improve the image classification consistently by large margins.

4.6. Discussion

Qualitative illustration of domain adaptation in segmentation We compare our robust domain adaptation (RDA) with the recent domain adaptation method [75] qualitatively over semantic segmentation task. As Fig. 3 shows, the proposed RDA outperforms the baselines clearly.

Overfitting Mitigation: We compared FAA with existing overfitting mitigation methods. Most existing methods address overfitting through certain network regularization by noise injection to hidden units (*Dropout*), label-dropout (*Label smooth*), gradient ascent (*Flooding*), data and label mixing (*Mixup*), gradient based adversarial attacking (*FGSM*), virtual-label based adversarial attacking (*VAT*), etc. Table 7 shows experimental results over the task GTA→Cityscapes. It can be seen that existing regularization does not perform well in the domain adaptation task. The major reason is that existing methods were designed for supervised and semi-supervised learning where training and test data usually have little domain gap. The proposed FAA mitigates overfitting with clear performance gains as it allows large magnitude of perturbation noises which is critical to the effectiveness of its generated adversarial samples due to the existence of 'domain gaps' in UDA.

Due to the space limit, we provide more visualization of

Method	Aero	Bike	Bus	Car	Horse	Knife	Motor	Person	Plant	Skateboard	Train	Truck	Mean
Res-101 [65]	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
MMD [45]	87.1	63.0	76.5	42.0	90.3	42.9	85.9	53.1	49.7	36.3	85.8	20.7	61.1
DANN [12]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
ENT [15]	80.3	75.5	75.8	48.3	77.9	27.3	69.7	40.2	46.5	46.6	79.3	16.0	57.0
MCD [67]	87.0	60.9	83.7	64.0	88.9	79.6	84.7	76.9	88.6	40.3	83.0	25.8	71.9
ADR [65]	87.8	79.5	83.7	65.3	92.3	61.8	88.9	73.2	87.8	60.0	85.5	32.3	74.8
SimNet-Res152 [59]	94.3	82.3	73.5	47.2	87.9	49.2	75.1	79.7	85.3	68.5	81.1	50.3	72.9
GTA-Res152 [69]	-	-	-	-	-	-	-	-	-	-	-	-	77.1
CBST [99]	87.2	78.8	56.5	55.4	85.1	79.2	83.8	77.7	82.8	88.8	69.0	72.0	76.4
CBST+FAA	90.6	80.3	79.6	67.1	86.7	80.2	86.2	77.1	86.2	87.1	80.6	71.8	**81.1**
CRST [98]	88.0	79.2	61.0	60.0	87.5	81.4	86.3	78.8	85.6	86.6	73.9	68.8	78.1
CRST+FAA	91.6	80.5	81.5	70.7	89.6	81.0	87.5	79.9	87.1	86.4	81.0	75.1	**82.7**

Table 5. Experimental results of domain adaptive image classification task on VisDA17 (+*FAA* means to include FAA to attack supervised source loss and unsupervised target loss in domain adaptation).

| Target Image | Baseline | AdaptSeg [75] | **Ours(AdaptSeg+FAA)** | Ground Truth |

Figure 3. Qualitative illustration of domain adaptive semantic segmentation task GTA5 to Cityscapes task. Our robust domain adaption (RDA) employs Fourier adversarial attacking to mitigate the overfitting in domain adaptive learning effectively, which produces better semantic segmentation by mitigating over-fitting in domain adaptive learning. The differences are clearer for less-frequent categories such as traffic-sign, pole and bus, etc.

Method	A→W	D→W	W→D	A→D	D→A	W→A	Mean
ResNet-50 [21]	68.4	96.7	99.3	68.9	62.5	60.7	76.1
DAN [45]	80.5	97.1	99.6	78.6	63.6	62.8	80.4
RTN [46]	84.5	96.8	99.4	77.5	66.2	64.8	81.6
DANN [12]	82.0	96.9	99.1	79.7	68.2	67.4	82.2
ADDA [78]	86.2	96.2	98.4	77.8	69.5	68.9	82.9
JAN [47]	85.4	97.4	99.8	84.7	68.6	70.0	84.3
GTA [69]	89.5	97.9	99.8	87.7	72.8	71.4	86.5
CBST [99]	87.8	98.5	100	86.5	71.2	70.9	85.8
CBST+FAA	90.9	98.9	99.8	92.5	76.7	76.2	**89.1**
TAT [42]	92.5	99.3	100.0	93.2	73.1	72.1	88.4
TAT+FAA	92.4	99.8	99.4	94.7	78.2	77.1	**90.3**
CRST [98]	89.4	98.9	100	88.7	72.6	70.9	86.8
CRST+FAA	92.3	99.2	99.7	94.4	80.5	78.7	**90.8**

Table 6. Experimental results of domain adaptive image classification over Office-31 (+*FAA* means to include FAA to attack supervised source loss and unsupervised target loss in domain adaptation).

Method	mIoU	gain
ST [99]	42.6	N.A.
+Dropout [71]	43.1	+0.5
+Label smooth [72]	43.4	+0.8
+Mixup [93]	43.6	+1.0
+FGSM [14]	43.9	+1.3
+VAT [51]	44.3	+1.7
+Flooding [31]	44.1	+1.5
+FAA	**50.1**	**+7.5**

Table 7. Comparison with existing overfitting migitation methods: For the semantic segmentation task GTA → Cityscapes, FAA performs the best consistently by large margins.

the qualitative segmentation examples (including their comparisons with the recent domain adaptation method [75]) in the supplementary material.

5. Conclusion

In this work, we presented RDA, a robust domain adaptation technique that mitigates overfitting in UDA via a novel Fourier adversarial attacking (FAA). We achieve robust domain adaptation by a novel Fourier adversarial attacking (FAA) method that allows large magnitude of perturbation noises but has minimal modification of image semantics. With FAA-generated adversarial samples, the training can continue the 'random walk' and drift into an area with a flat loss landscape, leading to more robust domain adaptation. Extensive experiments over multiple domain adaptation tasks show that RDA can work with different computer vision tasks (*i.e.*, segmentation, detection and classification) with superior performance. We will explore disentanglement-based adversarial attacking and its applications to other computer vision tasks. We will also study how FAA could mitigate over-fitting in classical supervised learning and semi-supervised learning as well as the recent contrast-based unsupervised representation learning.

Acknowledgement

References

[1] Shumeet Baluja and Ian Fischer. Adversarial transformation networks: Learning to generate adversarial examples. *arXiv preprint arXiv:1703.09387*, 2017. 2

[2] Mikhail Belkin, Daniel Hsu, and Partha Mitra. Overfitting or perfect fitting? risk bounds for classification and regression rules that interpolate. *arXiv preprint arXiv:1806.05161*, 2018. 2

[3] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010. 5, 6

[4] Ronald Newbold Bracewell and Ronald N Bracewell. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986. 3

[5] Rich Caruana, Steve Lawrence, and Lee Giles. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in neural information processing systems*, pages 402–408, 2001. 2

[6] Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, and Riccardo Zecchina. Entropy-sgd: Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124018, 2019. 2, 5

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1, 5

[8] Qingchao Chen, Yang Liu, Zhaowen Wang, Ian Wassell, and Kevin Chetty. Re-weighted adversarial adaptation network for unsupervised domain adaptation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1

[9] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 1, 5, 7

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 5

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016. 8

[13] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016. 2

[14] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 2, 8

[15] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In *Advances in neural information processing systems*, pages 529–536, 2005. 5, 8

[16] Dayan Guan, Jiaxing Huang, Shijian Lu, and Aoran Xiao. Scale variance minimization for unsupervised domain adaptation in image segmentation. *Pattern Recognition*, 112:107764, 2021. 2

[17] Dayan Guan, Jiaxing Huang, Aoran Xiao, and Shijian Lu. Domain adaptive video segmentation via temporal consistency regularization. *arXiv preprint arXiv:2107.11004*, 2021. 2

[18] Dayan Guan, Jiaxing Huang, Aoran Xiao, Shijian Lu, and Yanpeng Cao. Uncertainty-aware unsupervised domain adaptation in object detection. *IEEE Transactions on Multimedia*, 2021. 2

[19] Stephen José Hanson and Lorien Y Pratt. Comparing biases for minimal network construction with back-propagation. In *Proceedings of the 1st International Conference on Neural Information Processing Systems*, pages 177–185, 1988. 2

[20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1, 5

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 5, 6, 8

[22] Zhenwei He and Lei Zhang. Multi-adversarial faster-rcnn for unrestricted object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6668–6677, 2019. 7

[23] Cheng-Chun Hsu, Yi-Hsuan Tsai, Yen-Yu Lin, and Ming-Hsuan Yang. Every pixel matters: Center-aware feature alignment for domain adaptive object detector. In *European Conference on Computer Vision*, pages 733–748. Springer, 2020. 7

[24] Jiaxing Huang, Dayan Guan, Shijian Lu, and Aoran Xiao. Mlan: Multi-level adversarial network for domain adaptive semantic segmentation. *arXiv preprint arXiv:2103.12991*, 2021. 2

[25] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Category contrast for unsupervised domain adaptation in visual tasks. *arXiv preprint arXiv:2106.02885*, 2021. 2

[26] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Cross-view regularization for domain adaptive panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10133–10144, 2021. 2

[27] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF Conference*

on *Computer Vision and Pattern Recognition*, pages 6891–6902, 2021. 3, 4, 6

[28] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Semi-supervised domain adaptation via adaptive and progressive feature alignment. *arXiv preprint arXiv:2106.02845*, 2021. 2

[29] Jiaxing Huang, Shijian Lu, Dayan Guan, and Xiaobing Zhang. Contextual-relation consistent domain adaptation for semantic segmentation. In *European Conference on Computer Vision*, pages 705–722. Springer, 2020. 5, 6

[30] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 2

[31] Takashi Ishida, Ikko Yamane, Tomoya Sakai, Gang Niu, and Masashi Sugiyama. Do we need zero training loss after achieving zero training error? *arXiv preprint arXiv:2002.08709*, 2020. 2, 8

[32] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016. 3

[33] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. *arXiv preprint arXiv:2012.12821*, 2020. 3

[34] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4902, 2019. 1

[35] Guoliang Kang, Liang Zheng, Yan Yan, and Yi Yang. Deep adversarial attention alignment for unsupervised domain adaptation: the benefit of target expectation maximization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 401–416, 2018. 1

[36] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016. 2, 5

[37] Myeongjin Kim and Hyeran Byun. Learning texture invariant representation for domain adaptation of semantic segmentation. *arXiv preprint arXiv:2003.00867*, 2020. 5, 6

[38] Taekyung Kim, Minki Jeong, Seunghyeon Kim, Seokeon Choi, and Changick Kim. Diversify and match: A domain adaptive representation learning paradigm for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12456–12465, 2019. 7

[39] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. 1

[40] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. *arXiv preprint arXiv:1712.09913*, 2017. 2, 5

[41] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6936–6945, 2019. 2, 5, 6

[42] Hong Liu, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable adversarial training: A general approach to adapting deep classifiers. In *International Conference on Machine Learning*, pages 4013–4022. PMLR, 2019. 2, 5, 8

[43] Yanpei Liu, Xinyun Chen, Chang Liu, and Dawn Song. Delving into transferable adversarial examples and black-box attacks. *arXiv preprint arXiv:1611.02770*, 2016. 2

[44] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[45] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, pages 97–105, 2015. 8

[46] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *Advances in Neural Information Processing Systems*, pages 136–144, 2016. 2, 8

[47] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017. 8

[48] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019. 1, 2, 5, 6

[49] Zhipeng Luo, Zhongang Cai, Changqing Zhou, Gongjie Zhang, Haiyu Zhao, Shuai Yi, Shijian Lu, Hongsheng Li, Shanghang Zhang, and Ziwei Liu. Unsupervised domain adaptive 3d detection with multi-level consistency. *arXiv preprint arXiv:2107.11355*, 2021. 2

[50] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016. 3

[51] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018. 2, 8

[52] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1765–1773, 2017. 2

[53] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2574–2582, 2016. 2

[54] Andrew Y Ng et al. Preventing" overfitting" of cross-validation data. In *ICML*, volume 97, pages 245–253. Citeseer, 1997. 2

[55] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 5

[56] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. *arXiv preprint arXiv:2004.07703*, 2020. 6

[57] Xingchao Peng, Ben Usman, Neela Kaushik, Dequan Wang, Judy Hoffman, and Kate Saenko. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2021–2026, 2018. 5

[58] William B Pennebaker and Joan L Mitchell. *JPEG: Still image data compression standard*. Springer Science & Business Media, 1992. 3

[59] Pedro O Pinheiro. Unsupervised domain adaptation with similarity learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8004–8013, 2018. 7, 8

[60] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 1, 7

[61] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *European conference on computer vision*, pages 102–118. Springer, 2016. 5

[62] Rebecca Roelofs, Vaishaal Shankar, Benjamin Recht, Sara Fridovich-Keil, Moritz Hardt, John Miller, and Ludwig Schmidt. A meta-analysis of overfitting in machine learning. *Advances in Neural Information Processing Systems*, 32:9179–9189, 2019. 2

[63] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016. 5

[64] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010. 5

[65] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Adversarial dropout regularization. *International Conference on Learning Representations*, 2017. 8

[66] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6956–6965, 2019. 5, 7

[67] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, 2018. 2, 8

[68] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018. 5

[69] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, 2018. 5, 7, 8

[70] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 5

[71] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 2, 8

[72] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 2, 8

[73] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013. 2

[74] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. 2

[75] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7472–7481, 2018. 1, 2, 5, 6, 7, 8

[76] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1456–1465, 2019. 2, 6

[77] George Tucker, Andriy Mnih, Chris J Maddison, Dieterich Lawson, and Jascha Sohl-Dickstein. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. *arXiv preprint arXiv:1703.07370*, 2017. 3

[78] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017. 1, 2, 8

[79] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In *Advances in neural information processing systems*, pages 5334–5344, 2018. 2

[80] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 1, 2, 5, 6

[81] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992. 3

[82] Yufei Wang, Haoliang Li, Lap-pui Chau, and Alex C Kot. Embracing the dark knowledge: Domain generalization using regularized knowledge distillation. *arXiv preprint arXiv:2107.02629*, 2021. 2

[83] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogerior Feris, Jinjun Xiong, Wen-mei Hwu, Thomas S Huang, and Honghui Shi. Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmentation. *arXiv preprint arXiv:2003.08040*, 2020. 5, 6

[84] Roman Werpachowski, András György, and Csaba Szepesvári. Detecting overfitting via adversarial examples. *arXiv preprint arXiv:1903.02380*, 2019. 2

[85] Rongchang Xie, Fei Yu, Jiachao Wang, Yizhou Wang, and Li Zhang. Multi-level domain adaptive learning for crossdomain detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019. 7

[86] Chang-Dong Xu, Xing-Ran Zhao, Xin Jin, and Xiu-Shen Wei. Exploring categorical regularization for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11724–11733, 2020. 5, 7

[87] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1740–1749, 2020. 3

[88] Fan Yang, Ke Yan, Shijian Lu, Huizhu Jia, Don Xie, Zongqiao Yu, Xiaowei Guo, Feiyue Huang, and Wen Gao. Part-aware progressive unsupervised domain adaptation for person re-identification. *IEEE Transactions on Multimedia*, 23:1681–1695, 2020. 2

[89] Jihan Yang, Ruijia Xu, Ruiyu Li, Xiaojuan Qi, Xiaoyong Shen, Guanbin Li, and Liang Lin. An adversarial perturbation oriented domain adaptation approach for semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12613–12620, 2020. 2, 5

[90] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020. 2, 5, 6

[91] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2(5):6, 2018. 5

[92] Fangneng Zhan, Chuhui Xue, and Shijian Lu. Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9105–9115, 2019. 2

[93] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 2, 8

[94] Jingyi Zhang, Jiaxing Huang, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. *arXiv preprint arXiv:2106.06112*, 2021. 2

[95] Jingyi Zhang, Jiaxing Huang, Zhipeng Luo, Gongjie Zhang, and Shijian Lu. Da-detr: Domain adaptive detection transformer by hybrid attention. *arXiv preprint arXiv:2103.17084*, 2021. 2

[96] Stephan Zheng, Yang Song, Thomas Leung, and Ian Goodfellow. Improving the robustness of deep neural networks via stability training. In *Proceedings of the ieee conference on computer vision and pattern recognition*, pages 4480–4488, 2016. 2

[97] Xinge Zhu, Jiangmiao Pang, Ceyuan Yang, Jianping Shi, and Dahua Lin. Adapting object detectors via selective crossdomain alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 687–696, 2019. 7

[98] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5982–5991, 2019. 2, 5, 6, 8

[99] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 289–305, 2018. 2, 5, 6, 8