2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) | 978-1-6654-8739-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/CVPRW56347.2022.00448

# Open-Set Domain Adaptation Under Few Source-Domain Labeled Samples

Sayan Rakshit[1], Balasubramanian S[2], Hmrishav Bandyopadhyay[3], Piyush Bharambe[1],
Sai Nandan Desetti[2], Biplab Banerjee[1], Subhasis Chaudhuri[1]
Indian Institute of Technology Bombay[1], Sri SSIHL[2], Jadavpur University [3], India

## Abstract

*Recently, the notion of closed-set few-shot domain adaptation (FSDA) has been introduced where limited supervision is present in the source domain. However, FSDA overlooks the fact that the unlabeled target domain may contain new classes unseen in the source domain. To this end, we introduce the novel problem definition of few-shot open-set DA (FosDA) where the source domain contains few labeled samples together with a large pool of unlabeled data, and the target domain consists of test samples from the known as well as new categories. We propose an end-to-end model called FosDANet to tackle such a scenario which operates on two principles: to generate confident pseudo-labels for the unlabeled source samples and to classwise align the source and target domains for the known classes while rejecting the unknown-class data. A combination of a self-supervised loss and a novel triplet-based relation learning module is devised to aid in confident pseudo-labeling, and a dual adversarial learning scheme is proposed for domain alignment. Extensive experiments were performed on five datasets: Office-31, Office-Home, Adoptiape, and two new datasets we designed: Mini-domainNet and a remote sensing benchmark called NPU-RSDA. FosDANet is found to consistently outperform the relevant literature.*

## 1. Introduction

The unsupervised domain adaptation (UDA) paradigm [29, 34, 25, 11] presumes the availability of a label-rich source domain and a label-scarce target domain. The goal is to learn a domain-independent representation space where a classifier trained on the source domain can generalize well on the target data. The closed-set UDA has majorly been researched where both the domains share the same categories [23, 25]. However, this is a very restricted setup and is often easy to handle in practice. In open-set UDA [21, 16], the target domain may contain outliers arising from previously unknown classes. On the other hand, substantial annotations may not be available in the source domain since labeling is expensive and laborious whereas it is easy to obtain sufficient unlabeled data. The recently studied few-shot DA (FSDA) problem [33, 13] concerns a type of closed-set UDA setting where the source domain contains very few labeled and many unlabeled samples per class. A more realistic situation would be to extend FSDA to assume the presence of outlier samples in the target domain. In this regard, we propose to tackle the novel problem of few-shot open-set DA (FosDA) in this paper (Fig. 1). We note that FosDA is very different from the semi-supervised DA [20] (target domain supervision available) and the generic few-shot open-set learning [15] (base classes with ample supervision are present) and is more challenging to deal with.

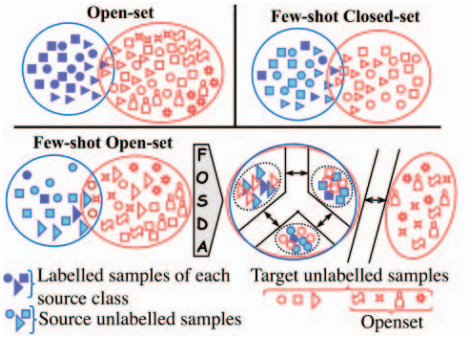



Figure 1: FosDA is a novel problem definition combining the challenges of few-shot DA and open-set DA, respectively. Similar to FSDA, FosDA has a small number of labeled samples in the source domain. Further, FosDA does not have any prior knowledge regarding the novel classes in the target domain. Our goal is to learn a discriminative feature space aligning the known-class samples of both the domains while identifying the unknown-class data.

The existing OSDA methods [21, 16] cannot handle FosDA effectively given that the distributions of the source domain classes cannot be approximated meaningfully from a small set of labeled data. The existing FSDA techniques [33, 13] employ the notion of self-supervised learning (SSL) in terms of cross-domain instance-to-instance matching. Although such matching may be useful for a closed-set setup, it will fail in an open-set setting where an

978-1-6654-8739-9/22/$31.00 ©2022 IEEE
DOI 10.1109/CVPRW56347.2022.00448

outlier sample from the target domain may erroneously be mapped to the instances of the source domain. Hence, a naive extension of such techniques is not advocated to solve FosDA. In this regard, we outline three requisites to deal with FosDA: i) to learn a discriminative latent space given all the source domain samples, ii) to classwise align the known-class target domain samples with the source-domain data, and iii) to ensure separation between the known and unknown-class target domain samples. The main challenge in solving (i) lies in generating confident pseudo-labels for the unlabeled source data. Similarly, (ii) and (iii) require learning a soft boundary to demarcate the unknown-class samples from known-class data with high precision.

**Contributions:** Considering the aforesaid concerns, we propose a novel architecture called FosDANet to tackle FosDA. We expand the proportion of labeled data in the source domain by pseudo-labeling the unlabeled data, which aids in better domain alignment. For pseudo-labeling, we ensure good feature learning from both the labeled and unlabeled source domain samples, and further constrain the feature space to be class-discriminative. While we depend on the instance discrimination based SSL strategy [30] for instance-level feature learning (Sec. 5.1), a novel relation network is proposed to relate different instances through a metric loss (Sec. 5.3), thereby defining a way to distinguish between classes. As a result, confident pseudo-labeling of the unlabeled source samples can be performed through an ensemble of two classifiers (Sec. 5.2). Besides pseudo-labeling, the classifier system is utilized to deploy the domain alignment loss functions. Specifically, a dual adversarial learning strategy to align the domains is implemented by forcing adversarial games between both the classifiers and the shared feature extractor (Sec. 5.4). The first classifier relies on the adversarial model of [21] to directly learn the open-set boundary and also align the domains. However, this is not sufficient because [21] does not assume any prior regarding the open-set. This leads to a low-confidence prediction for the outliers. To combat this, we force the second classifier also to play an adversarial game with the feature extractor through a novel adversarial loss, where the classifier believes that all the target samples are outliers. This will encourage the feature extractor to learn distinctive features for the open-set samples, thereby enhancing the confidence on outlier prediction. We highlight our major contributions as follows:

i) We introduce the problem of FosDA to the community and propose an end-to-end trainable network called FosDANet as a solution. ii) FosDaNet improves the feature learning capability of instance discrimination based SSL through the relation network. A combination of both the modules is able to generate higher quality pseudo-labels than the direct probability thresholding method. iii) Our dual adversarial learning performs better source-target alignment while ensuring separability from the unknown-class data. iv) We devise the experimental protocol for FosDA and conduct extensive experiments on five benchmark and large-scale datasets, among these we newly curate NPU-RSDA and Mini-domainNet.

## 2. Related works

**Open-Set domain adaptation (OSDA):** The OSDA problem comes in a variety of flavours where classes in both the domains can be divided into two groups, viz. shared classes and domain-specific classes. For example, the unlabelled target domain may contain samples from an additional previously unknown set of classes, thus making the target task an open-set recognition problem. One of the early attempts [12] used the class-wise probability thresholds to recognise the open-set samples. Subsequently, [21] tackled the OSDA problem through adversarial learning to distinguish the unseen and seen-class samples and perform adaptation between domains. [16, 6] are two recent methods that used the concept of adversarial learning for domain alignment by virtue of a coarse to fine level sample weighting scheme. [6] exploited the semantic structure of the latent space to identify the outliers. The notion of rotation angle prediction based SSL has been used in [1] for OSDA.
**Few-shot domain adaptation (FSDA):** As already pointed out, FSDA is a type of UDA setup where the source domain contains very few labeled samples (eg. 1-shot) together with a large pool of unlabelled data. FSDA is a comparatively newly studied problem and only a handful of works exist in the literature. [33] leveraged in-domain and cross-domain prototypical learning through SSL followed by the notion of instance to prototype matching. Another endeavor in this regard [13] is based on the paradigm of instance discrimination for cross-domain instance-level matching. Both the methods follow the closed-set setup and to best of our knowledge, no prior attempt has been made to solve FosDA.
**Self-supervised learning (SSL):** SSL is a type of unsupervised learning which is used to highlight meaningful abstract concepts from the data without any semantic information. A common tendency to achieve self-supervision is through designing a pretext task, such as image colorization, image imprinting, jigsaw puzzle [2, 3], to be solved jointly with the downstream task. Some recent contrastive learning based SSL approaches [9, 4] have gained impressive performance in the area of representation learning. The notion of instance discrimination seeks to classify the samples into different instance identities with the hope to learn improved class-wise feature representations [30, 28].

Self-supervision has also been incorporated within some of the UDA models. Reconstruction based SSL has been used in [7, 8] to learn a domain invariant feature space. Besides, jigsaw puzzle [2] and instance discriminator [33] have also been utilized to obtain domain invariant features

from images. [24] incorporated multiple SSL tasks within a DA framework in a multi-task fashion. The use of SSL for FSDA and OSDA has already been mentioned above.

## 3. Problem definition

Let us consider a source domain $\mathcal{S}$ equipped with very few labeled samples per-class: $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ together with a large volume of unlabeled samples $\mathcal{D}^{su} = \{(x_j^{su})\}_{j=1}^{N_{su}}$, where (i) $N_s \ll N_{su}$, (ii) a given $x_i$ represents the image data and (iii) the samples from both $\mathcal{D}^s$ and $\mathcal{D}^{su}$ use the same label-set, $C_s = \{1, 2, \cdots, \mathcal{C}\}$. The target domain $\mathcal{T}$ consists of unlabeled test samples $\mathcal{D}^t = \{(x_k^t)\}_{k=1}^{N_t}$ spreading over the label set $C_t$. As per the proposed setting, $C_s \subset C_t$ and $C_{t/s} = C_t - C_s$ denotes the set of class labels private to $\mathcal{T}$. Further, $\mathcal{D}^s$ and $\mathcal{D}^{su}$ arise from the same underlying distributions: $P(\mathcal{D}^s) = P(\mathcal{D}^{su})$ whereas the data distributions of $\mathcal{D}^t$ are different from $\mathcal{D}^s \cup \mathcal{D}^{su}$: $P(\mathcal{D}^t) \neq P(\mathcal{D}^s \cup \mathcal{D}^{su})$.

We seek to perform empirical risk minimization for $\mathcal{S}$ leveraging $\mathcal{D}^s \cup \mathcal{D}^{su}$ in a domain-invariant latent space where $P(\mathcal{D}^t) \approx P(\mathcal{D}^s \cup \mathcal{D}^{su})$ such that the trained classifier is able to discriminate the samples from $\mathcal{D}^t$ into $\mathcal{C}+1$ class labels where the $\mathcal{C}+1^{th}$ index denotes the common 'reject' class referring to the outliers.

## 4. Model overview

Broadly, the model architecture of FosDaNet (Fig. 2) consists of three sub-modules, namely, a shared feature extractor $\mathcal{F}$, two separate $\mathcal{C}+1$-class classifiers $(\mathcal{G}_1, \mathcal{G}_2)$, and a module for discriminative feature learning for $\mathcal{S}$ consisting of an instance discriminator based SSL network $\mathcal{H}$ and the triplet-based relational network $\mathcal{R}$, respectively.

**Feature extractor ($\mathcal{F}$):** To extract features from the images, we consider the Resnet-50 [10] pre-trained on ImageNet as our backbone architecture where we generate the features from the final convolution layer followed by average pooling. $\mathcal{F}$ is shared by both $\mathcal{S}$ and $\mathcal{T}$, respectively.

**Instance discriminator network ($\mathcal{H}$):** The main goal of this part is to learn a meaningful $\mathcal{F}$ given $\mathcal{D}^s \cup \mathcal{D}^{su}$ by capturing the inherent similarities of the samples. For this purpose, the SSL method based on the notion of instance discrimination is considered by imposing a distinctive instance identifier to every image. Features are learned by training $\mathcal{F}$ such that an image is classified to its own identity while treating all the other images as negatives. The instance discriminator network $\mathcal{H}$ consists of a linear layer and it considers the input $\mathcal{F}(\mathcal{D}^s \cup \mathcal{D}^{su})$ to produce an output of size $(N_s + N_{su})$ through a softmax-type activation function.

**Relation network ($\mathcal{R}$):** The prime focus of the relation network $\mathcal{R}$ is to maximize the pairwise similarity of the source domain samples sharing the class labels in the latent space produced by $\mathcal{F}$ while making different-class data more separable. $\mathcal{R}$ is driven by the notion of pairwise similarity optimization and works with triplets of data. Given a triplet $(x_a, x_+, x_-)$, $\mathcal{R}$ is composed of a multi-layer perceptron with a sigmoid activation at the final layer and considers $|\mathcal{F}(x_a); \mathcal{F}(x_+/x_-)|$ as input where $|;|$ defines the vector concatenation operation. $\mathcal{R}$ outputs a single value in the range $[0, 1]$ depicting the similarity of $x_a$ with $x_+/x_-$.

**Classifiers $(\mathcal{G}_1, \mathcal{G}_2)$:** We use two classifiers on top of $\mathcal{F}$, each outputting a $\mathcal{C}+1$-dimensional class membership vector. The classifiers are made of two dense layers each and are utilized for multiple purposes: i) to classify the samples from $\mathcal{S}$ into one of the $\mathcal{C}$ categories, ii) to estimate the pseudo-labels for the samples in $\mathcal{D}^{su}$, and iii) to align the known-class samples from $\mathcal{D}^t$ with $\mathcal{S}$ while enforcing the unknown-class samples from $\mathcal{T}$ to take the label $\mathcal{C}+1$.

## 5. Training & inference

In this section, we discuss the loss functions applied on $\mathcal{S}$ and $\mathcal{T}$ separately. We define three loss functions based on $\mathcal{D}^s \cup \mathcal{D}^{su}$: classification loss for $(\mathcal{G}_1, \mathcal{G}_2)$, instance discrimination loss for $\mathcal{H}$, and the similarity learning loss for $\mathcal{R}$, respectively. The adversarial domain alignment losses are used for $\mathcal{D}^t$. Details are presented below.

### 5.1. Instance discrimination based SSL on $\mathcal{S}$

Since $\mathcal{D}^s$ has limited supervisory signal, the classifiers trained on $\mathcal{D}^s$ are bound to overfit. Alternatively, $\mathcal{D}^{su}$ can be utilized along with $\mathcal{D}^s$ to learn generic features for the classes in $\mathcal{S}$. With this aim, non-parametric instance discrimination [30] is used to learn abstract visual representations in $\mathcal{F}$ for $\mathcal{S}$ by discriminating the input images of $\mathcal{D}^s \cup \mathcal{D}^{su}$ into different identities and overlooking the presence of the semantic class-labels. A memory bank $\mathcal{V}^s$ is initialized on source domain features as: $\mathcal{V}^s = [v_1, v_2, \cdots, v_{N_s+N_{su}}]$ where $v_i = \mathcal{H}(\mathcal{F}(x_i))$ for a given image $x_i$. After the initialization, $\mathcal{V}^s$ is subsequently updated with a momentum in every batch. To perform instance discrimination on $\mathcal{S}$, we compute the similarity distribution $P^s$ given the features $f_i^s = \mathcal{H}(\mathcal{F}(x_i))$ as per Eq. 1. Following [30], we use two different notations $v$ and $f$ to distinguish between the memory bank elements and the features, respectively. The feature normalization is performed for both $v$ and $f$ by calculating Eq. 1.

$$P_{ij}^s = \frac{\exp((v_j^s)^T f_i^s/\tau)}{\sum_{l=1}^{N_s+N_{su}} \exp((v_l^s)^T f_i^s/\tau)} \tag{1}$$

$\tau$ is the temperature parameter and controls the concentration level of $P^s$ and is set to $0.005$ following [30]. The instance discrimination is optimized via the cross-entropy loss, $\mathcal{L}_{ID}$, where $i$ is considered to be the instance label for a given $x_i$.

$$\mathcal{L}_{ID} = \sum_{i=1}^{N_s+N_{su}} -i \log P_{ii}^s \tag{2}$$

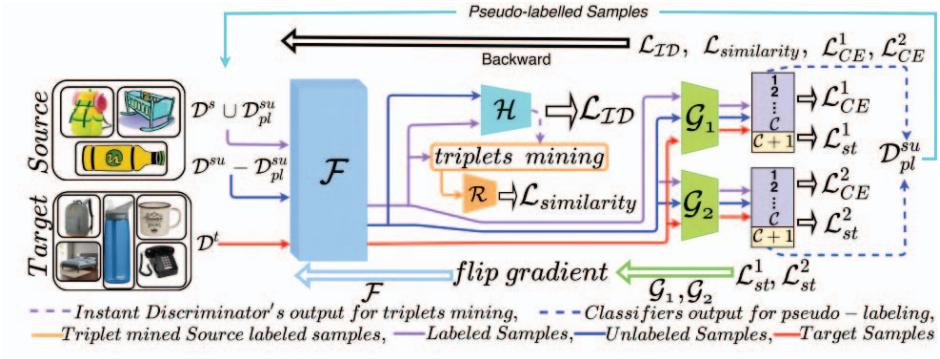



Figure 2: A detailed illustration of the FosDANet model. The major model components are the feature extractor $\mathcal{F}$, the instance discriminator $\mathcal{H}$, the relation network $\mathcal{R}$, and the classifiers $\mathcal{G}_1$ and $\mathcal{G}_2$.

### 5.2. Classifiers training and pseudo-labeling on $\mathcal{S}$

The classifiers $\mathcal{G}_1$ and $\mathcal{G}_2$ are trained with respect to the multi-class cross-entropy loss functions $\mathcal{L}^1_{CE}$ and $\mathcal{L}^2_{CE}$, respectively. Although the training commences with the available labeled data $\mathcal{D}^s$, subsequently, samples from $\mathcal{D}^{su}$ are pseudo-labeled, and a set of confident pseudo-labeled samples $\mathcal{D}^{su}_{pl}$ ($\mathcal{D}^s_{pl} \subset \mathcal{D}^{su}$) are used together with $\mathcal{D}^s$ ($\mathcal{D}^s \cup \mathcal{D}^{su}_{pl}$) to train the classifiers. Here, only the first $\mathcal{C}$ indices are used in $\mathcal{L}^1_{CE}$ and $\mathcal{L}^2_{CE}$ since $\mathcal{S}$ contains samples from $\mathcal{C}$ classes. In order to inject diversity in training $\mathcal{G}_1$ and $\mathcal{G}_2$, the training batches are made partially non-overlapping so that the classifiers get to observe disjoint samples.

**Generation of the pseudo-labeled samples $\mathcal{D}^{su}_{pl}$:** The main goal of pseudo-labeling is to expand the label information of $\mathcal{S}$. The initial training iterations for $(\mathcal{G}_1, \mathcal{G}_2)$ are based on $\mathcal{D}^s$ only, letting the classifiers gain enough knowledge to predict informatively regarding the class-distributions for the samples in $\mathcal{D}^{su}$. Subsequently, we evaluate the class probability scores for these samples with respect to both $\mathcal{G}_1$ and $\mathcal{G}_2$. For a given sample, if we find that the predicted softmax probabilities for a specific class are more than a pre-defined threshold of $\alpha$ for both $\mathcal{G}_1$ and $\mathcal{G}_2$ ($\alpha$ is set to $0.95$ for ensuring high confidence), then the sample is included in $\mathcal{D}^{su}_{pl}$ together with the concerned class label. Employing an ensemble of two classifiers instead of a single classifier enforces a degree of confidence in the predictions. We note that a given sample, once included in $\mathcal{D}^{su}_{pl}$, is not considered for pseudo-labeling subsequently.

### 5.3. Similarity optimization on $\mathcal{S}$ using $\mathcal{R}$

Although the instance discriminator is expected to highlight some meaningful features from the samples, it does not ensure a dense feature space as $\mathcal{L}_{ID}$ does not utilize the semantic class information. Furthermore, the inherent image ambiguities may hamper the feature learning abilities of $\mathcal{F}$ if it is optimized solely based on $\mathcal{L}_{ID}$. If the feature space is not class discriminative, then it remains difficult for $(\mathcal{G}_1, \mathcal{G}_2)$ to generate highly confident pseudo-labels. Hence, we aim to ensure feature-level consistency for the samples class-wise using the relation network $\mathcal{R}$, which, in turn, makes the latent space highly class-concentrated. In some sense, $\mathcal{R}$ and $\mathcal{H}$ demonstrate mutual co-regularization effects. As already mentioned, $\mathcal{R}$ works with triplets of data. However, triplet selection is a non-trivial problem in our context. Our goal is to select a moderate number of highly representative triplets in such a way that $\mathcal{R}$ can subsequently ensure a globally discriminative feature space for $\mathcal{S}$. Our proposal in this regard is as follows.

**Triplets mining:** The triplets are curated from the updated $\mathcal{D}^s \cup \mathcal{D}^{su}_{pl}$ at a given iteration. We leverage the fact that $\mathcal{H}$ promotes similarity between instances akin to how supervised class learning promotes similarity between classes. Hence, for an anchor image $x_a \in \mathcal{D}^s \cup \mathcal{D}^{su}_{pl}$, selected as the one having a high $P^s$ score, we define (i) a hard positive as the least similar sample $x_+$ to $x_a$ sharing the same semantic class label, i.e. $x_+$ is selected from the same class as $x_a$ with lowest $P^s$ score, (ii) a hard negative as the most similar instance $x_-$ to $x_a$ from another semantic category, i.e. $x_-$ is selected from a class different from that of $x_a$ with a high $P^s$ value. The premise behind such a selection strategy signifies that while $\mathcal{H}$ is able to bring out superior semantic features from $x_a$, the features of $x_+$ are relatively poor. The goal of $\mathcal{R}$ is to bring such extreme samples of a given class closer, which subsequently reduces the intra-class variations. Similarly, pushing $x_a$ and $x_-$ far will direct $\mathcal{F}$ to learn class-discernible features.

**Proposed similarity loss**: Given $|(\mathcal{F}(x_a); \mathcal{F}(x_+))|$ and $|(\mathcal{F}(x_a); \mathcal{F}(x_-))|$ as the inputs, we seek $\mathcal{R}$ to produce a high similarity score ($\approx 1$) for $(x_a, x_+)$ and a low similarity score for $(x_a, x_-)$ ($\approx 0$) simultaneously, as follows,

$$\mathcal{L}_{similarity} = \mathbb{E}_{(x_a, x_+, x_-) \in \mathcal{S}}[(\mathcal{R}(|(\mathcal{F}(x_a); \mathcal{F}(x_+))|) - 1)^2 + (\mathcal{R}(|(\mathcal{F}(x_a); \mathcal{F}(x_-))|)^2] \quad (3)$$

We note that the pseudo-labeling is not considered for $\mathcal{T}$ given the unknown openness factor which may severely affect the outcomes of the instance discriminator by wrongly matching open and closed-set samples.

### 5.4. Dual adversarial alignment of $\mathcal{S}$ and $\mathcal{T}$

We propose an adversarial learning based framework to align the known-class samples of $\mathcal{D}^t$ with $\mathcal{D}^s \cup \mathcal{D}^{su}$ and identify the unknown class samples in $\mathcal{T}$. We define two adversarial loss objectives $\mathcal{L}^1_{st}$ and $\mathcal{L}^2_{st}$, implemented by $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, to carry out these goals. The adversarial games are played between $\mathcal{F}$ and the classifiers $\mathcal{G}_1$ and $\mathcal{G}_2$: the classifiers aim to minimize $\mathcal{L}^1_{st}$ and $\mathcal{L}^2_{st}$ respectively, while $\mathcal{F}$ tries to deceive the classifiers. While the adversarial loss on $\mathcal{G}_1$ performs a coarse-level alignment by purposefully discriminating the samples from $\mathcal{D}^t$ between the open-set and closed-set super-classes, the adversarial loss on $\mathcal{G}_2$ performs a more confident class assignment for the target data, in particular, the unknown-class samples.

Our coarse-level alignment loss is inspired by the OSDA-BP [21] idea which seeks to learn a pseudo open-class boundary separating the open-set samples from the inliers. In this way, the inlier samples of $\mathcal{T}$ are aligned with $\mathcal{S}$ by sharing the first $\mathcal{C}$ class labels as produced by $\mathcal{G}_1$ whereas the potential outlier samples are assigned the label $\mathcal{C}+1$. The loss is implemented as a binary cross-entropy loss with the ground-truth probabilities for the open and closed classes set to $0.5$ each.

$$\mathcal{L}^1_{st} = \max_{\mathcal{F}} \min_{\mathcal{G}_1} \frac{1}{N_t} \sum_{k=1}^{N_t} -0.5 \log(\mathcal{G}_1(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) \\ -0.5 \log(1 - \mathcal{G}_1(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) \tag{4}$$

We highlight the potential issues of $\mathcal{L}^1_{st}$ in confidently classifying the open-set samples from $\mathcal{T}$. If an open-set sample is highly similar to some of the closed-set classes (e.g, a pair of fine-grained classes are present in the open-set and closed-set), Eq. 4 shares the posterior probability between those classes, thus reducing its open-set posterior probability $\mathcal{G}_1(y_t = \mathcal{C}+1|\mathcal{F}(x_t))$ incorrectly. Also, since the ground-truth is set to $0.5$, a little deviation in the posterior probability provides $\mathcal{F}$ with the signal to fool $\mathcal{G}_1$. Hence, the average open-set probability scores for the unknown-class samples are not very high in this case, leading to their less confident predictions and occasional misclassifications.

We propose another adversarial loss between $\mathcal{F}$ and $\mathcal{G}_2$ to combat the aforesaid issues judiciously through the multi-class loss as follows, where the ground-truth labels for all the target samples are fixed as $\mathcal{C}+1$.

$$\mathcal{L}^2_{st} = \max_{\mathcal{F}} \min_{\mathcal{G}_2} -\frac{1}{N_t} \sum_{k=1}^{N_t} \log(\mathcal{G}_2(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) \tag{5}$$

Now for the potential known-class samples from $\mathcal{T}$, $\mathcal{F}$ tries to reduce their open-class probability to deceive both $\mathcal{G}_1$ and $\mathcal{G}_2$, making those samples to overfit with the samples from $\mathcal{S}$. Let us consider two types of open-set samples in $\mathcal{T}$: *less-open* samples which have some similarity with the known-class data, and *more-open* samples, respectively. For these samples, $\mathcal{F}$ tries to make $\mathcal{G}_1(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) > 0.5$ (as per Eq. 4) and $\mathcal{G}_2(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) < 1.0$ (as per Eq. 5) simultaneously while $\mathcal{G}_1$ and $\mathcal{G}_2$ will restrict $\mathcal{G}_1(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) = 0.5$ and $\mathcal{G}_2(y^t_k = \mathcal{C}+1|\mathcal{F}(x^t_k)) = 1.0$. This provides a dual restriction to the potential open-set samples and restricts their open-class probability from going $< 0.5$. As a result, the *more-open* samples will have an open-class probability much higher while the open-class probability for the *less-open* samples will be higher than Eq. 4 alone.

### 5.5. Total loss

We follow an alternate optimization strategy to train FosDaNet. The model is trained for multiple repeats with a number of training epochs taking place within each repeat. The pseudo-labeling is performed at the end of the final epoch of each repeat. The sequence of the training losses are mentioned below with the weighting parameter $\gamma$,

$$\min_{\mathcal{F},\mathcal{H},\mathcal{R}} (\mathcal{L}_{ID} + \mathcal{L}_{similarity}) + \min_{\mathcal{G}_1,\mathcal{G}_2} (\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE} + \mathcal{L}^1_{st} + \gamma \mathcal{L}^2_{st}) \tag{6}$$

$$\min_{\mathcal{F}} (\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE} - \mathcal{L}^1_{st} - \gamma \mathcal{L}^2_{st}) \tag{7}$$

The first part of Eq. 6 is devoted to the SSL task and optimizes the ID loss together with the similarity loss. The remaining part of Eq. 6 and Eq. 7 implement the dual adversarial loss together with the classification losses on $\mathcal{S}$ in the min-max optimization fashion given $\mathcal{F}$ and $(\mathcal{G}_1, \mathcal{G}_2)$, respectively. During inference, we combine the class predictions for both $\mathcal{G}_1$ and $\mathcal{G}_2$ through average pooling, and the class-label with the maximum probability is considered.

## 6. Experiments

**Datasets:** We evaluate the performance of FosDANet on five datasets, out of which two are introduced in this paper. For Office-31 [19] (three domains: Amazon (A), Webcam (W), DSLR (D), 31 classes), Office-Home [27] (four domains: Art (A), Clipart (C), Product (P), Realworld (W), 65 classes), and Adaptiope [18] (three domains: Product (P), Real (R), Synthetic (S), 123 classes), we consider $[20, 40, 63]$ closed-set classes selected in the alphabetical order whereas the remaining classes constitute the open-set (label $\mathcal{C}+1$). **Mini-domainNet**: is created from the publicly available very large-scale DomainNet dataset [17] and tailored to the few-shot task as the FosDA performance on domainNet was found to be extremely poor. Out of the original six domains, we consider three (Real (R), painting (P) and sketch (S)) in Mini-domainNet with 200 classes per domain selected in the alphabatical order and 50 images per class. If a given class contains less than 50 images

in domainNet, then all the images are taken. Alphabetically, the first 70 of them constitute the closed-set, whereas the remaining classes denote the open-set. The new **NPU-RSDA**[1] dataset is created from three benchmark optical remote sensing datasets, UC-Merced (U), PatternNet (P), and NWPU-RESISC45 (N), respectively, captured across the globe. The task is to perform scene recognition from the images. The images in UC-Merced [32] are extracted from the USGS National Map Urban Area Imagery collection. A total of 21 categories are present in UC-Merced and the spatial resolution of the images is 0.3 m. The NWPU-RESISC45 [5] dataset contains 45 scene categories which are extracted from the Google Earth Engine platform. The spectral resolution of the images varies in the range of 0.2m to 30m. Finally, *PatternNet* [35] is a high resolution dataset created from the Google Earth Engine containing 38 land-cover classes. The spectral resolution varies from 0.062m to 4.693m. We identify the 17 common classes to constitute the three domains containing a total of 27000 images. Alphabetically, the first ten classes are considered as the closed-set and the remaining classes comprise the open-set.

**Experimental and evaluation protocols:** The training samples were considered randomly and the same set of samples was used for all the comparative methods. We train the model using the SGD optimizer with a batch size of 32 and an initial learning rate of $0.1$ for training $(\mathcal{R}, \mathcal{H}, \mathcal{G}_1, \mathcal{G}_2)$ and $0.001$ for training $\mathcal{F}$, respectively. We fix $\gamma = 0.6$ so that Eq. 5 can provide a soft regularization effect to Eq. 4 in confidently classifying the outliers. The entire model including the feature extractor is trained. We consider five repeats and $50$ epochs inside each repeat. Model convergence could be observed for all the cases. We report the $OS$ and $OS^*$ scores (mean$\pm$std over three runs) to denote the average class-wise accuracy measures for the known+unknown and the known classes, respectively. The classification performance for the outlier samples (UNK) can be calculated as $OS \times (\mathcal{C}+1) - OS^* \times \mathcal{C}$. For unbiased comparisons, we report the harmonic mean (HOS)[1] of $OS^*$ and $UNK$.

### 6.1. Comparison to the literature

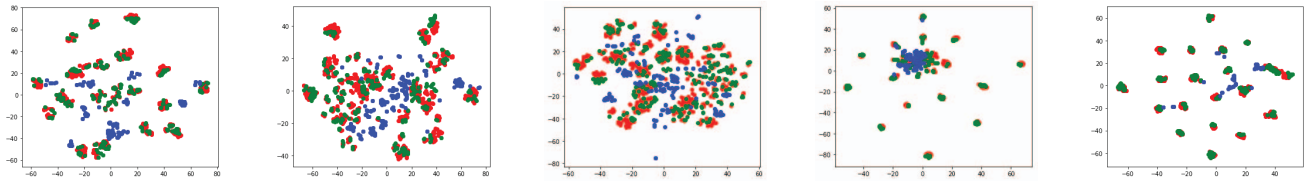
We compare FosDANet with three representative OSDA techniques from the literature: OSDA-BP [21], STA [16], and an SSL based approach: ROT [1], respectively, whereas open-set SVM (OSVM) [22] is selected as the baseline. We consider two variants of these algorithms for fair comparison: one without any pseudo-labeling and the other with pseudo-labeling on $\mathcal{D}^{su}$ with a probability threshold of $0.95$. Note that pseudo-labeling is not possible for OSVM as it does not include unlabeled data during training. For Office-31 and NPU-RSDA, we report the results of 1-shot and 3-shot source labels per class (Detailed results for NPU-RSDA is mentioned in the supplementary text). For office-home, we conducted experiments with $3\%$ and $6\%$ labeled images per class following [14]. For Adoptiape and Mini-domainNet, we consider $3\%$ per-class labeled images, which means not more than three labeled samples exist at the class-level. As we observe from Tables 1-4, FosDANet is able to sharply outperform the other approaches for all the datasets. For example, FosDANet produces an $OS$ score of $73.1\%$ (1-shot) and $79.1\%$ (3-shot) for office-31, $47.7\%$ ($3\%$) and $48.1\%$ ($6\%$) for Office-Home, and $80.7\%$ (3-shot) for NPU-RSDA, respectively. For Adaptiope and Mini-domainNet, our $OS$ values are $47\%$ and $38.8\%$ for the $3\%$ case. We further observe that FosDA with pseudo-labeling outperform the naive FosDANet without pseudo-labeling substantially (by more than $25\%$ and around $15\%$ for Adaptiope and Mini-domainNet, respectively). Barring a few cases in Tables 1-4, most of the methods from literature report comparable or lower performance with the naive probability thresholding based pseudo-labeling against the case without pseudo-labeling. One reason could be that naive probability thresholding can introduce many noisy labels, which can confuse the classifier. This is in complete contrast to our FosDANet which significantly benefits from pseudo-labeling, underlining the pseudo-labeling strategy adopted here. For example, Fig. 5 (a) shows a comparison on the number of correctly retrieved pseudo-labels after the first repeat by all the methods on W-D (Office-31). As can be observed, FosDANet produces the most number of correct pseudo-labels (270) while the next best performing ROT [1] produces only 170 pseudo-labels. The other reason could be the inability of the models to deal with fine-grained classes unlike our FosDANet. In fact, the t-SNE [26] plots for W-A (Office-31) in Fig. 3 lucidly highlight the significantly better class-discriminative feature space learnt by our FosDANet, in comparison to the other methods.

### 6.2. Ablation analysis [2]

We analyze the effects of each component of the proposed cost function in Table 5 for two cases: $R-A$ (Office-Home), $U-N$ (NPU-RSDA), respectively. We first consider the base model consisting of $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$, $\mathcal{L}^1_{St}$, $\mathcal{L}^2_{St}$ and without any pseudo-labeling on $\mathcal{D}^{su}$. The use of pseudo-labeling in the base model shows an improvement of around $8$ to $9\%$ for both the cases. We consider two pseudo-labeling strategies here, i) randomly picking up one classifier with a probability of $0.5$ and use it for obtaining the pseudo-label (PL-random), ii) the proposed confidence based pseudo-labeling combining $\mathcal{G}_1$ and $\mathcal{G}_2$ (PL). Ours is found to be better. At convergence, both $\mathcal{G}_1$ and $\mathcal{G}_2$ produce identical class distributions, hence, the best component classifier and the ensemble do not show much differences in performance. We subsequently include $\mathcal{L}_{ID}$ and $\mathcal{L}_{similarity}$ to the base model sequentially which shows fur-

[1] More details about the dataset in the supplementary text.

[2] The cross domain retrieval results are shown in the supplementary text.



(a) Resnet50 features. (b) OSDA-BP [21]. (c) STA [16]. (d) ROT [1]. (e) FosDANet (Full).

Figure 3: t-SNE plots for W-D (Office-31) (3-shot) by different techniques. Red and Green represent the known-class samples from $\mathcal{S}$ and $\mathcal{T}$ whereas Blue denotes the open-set samples private to $\mathcal{T}$.

| Method | A - W | | A - D | | W - D | | W - A | | D - A | | D - W | | AVG | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OS* | OS | OS* | OS | OS* | OS | OS* | OS | OS* | OS | OS* | OS | OS* | OS | HOS |
| **OSVM**[12] | 4.7 | 4.4 | 5.2 | 5.0 | 4.6 | 4.4 | 4.9 | 4.6 | 5.1 | 4.8 | 5.2 | 5.1 | -/ 4.9 | -/ 4.7 | -/6.9 |
| | 0.7 | 5.4 | 3.3 | 7.9 | 17.7 | 21.6 | 9.8 | 14.1 | 5.1 | 4.9 | 5.0 | 4.8 | -/ 6.9 | -/ 12.1 | -/9.8 |
| **OSDA-BP**[21] | 16.7 | 17.3 | 19.7 | 19.8 | 29.0 | 31.8 | 15.0 | 18.3 | 13.5 | 16.4 | 34.7 | 32.8 | 21.4(±0.8)/14.4 | 22.7(±0.7)/15.8 | 29.7/21.7 |
| | 19.1 | 21.2 | 21.4 | 22.0 | 35.6 | 32.8 | 18.8 | 21.2 | 17.2 | 19.1 | 23.9 | 27 | 22.7(±0.8)/17.0 | 23.9(±0.5)/18.6 | 30.8/25.3 |
| **STA**[16] | 8.7 | 7.7 | 7.0 | 6.4 | 34.5 | 37.2 | 17.3 | 20.6 | 30.2 | 32.4 | 39.3 | 40.6 | 22.8(±1.3)/17.9 | 24.1 (±0.9)/18.3 | 31.3/21.3 |
| | 40.6 | 41.3 | 37.5 | 37.2 | 64.4 | 63.7 | 35.4 | 36.8 | 33.9 | 34.9 | 63.5 | 64.7 | 45.9(±0.4)/40.9 | 46.4(±0.5)/40.5 | 50.6/36.2 |
| **ROT**[1] | 24.7 | 25.2 | 18.9 | 19.1 | 56.2 | 57.8 | 24.9 | 27.9 | 24.5 | 27.4 | 26.8 | 29.0 | 29.3(±0.6)/27.9 | 31.0(±0.8)/27.4 | 40.4/21.4 |
| | 62.1 | 62.2 | 62.2 | 61.0 | 73.6 | 74.7 | 34.3 | 36.1 | 52.3 | 52.7 | 73.8 | 74.7 | 61.0(±0.5)/48.3 | 60.2(±0.4)/47.6 | 51.3/39.6 |
| **FosDANet** | **57.6** | **57.1** | **69.1** | **68.6** | **91.3** | **89.2** | **67.6** | **66.5** | **66.2** | **65.5** | **93.3** | **92.5** | **74.1**(±1.1)/**64.2** | **73.3**(±1.3)/**63.3** | **64.6/53.1** |
| | **80.2** | **79.7** | **73.1** | **72.7** | **96.4** | **95.9** | **64.5** | **64.1** | **66.5** | **66.2** | **99.2** | **97.2** | **79.9**(±0.8)/**68.7** | **79.1**(±0.9)/**68.4** | **70.5/65.4** |

Table 1: Comparison to the literature for the Office-31 dataset for 1-shot and 3-shot cases. For each method, the first row denotes the results of 1-shot and the second row (in gray) shows the results of 3-shot settings, respectively. For the average values (last column), we show the performance of the model with (in black) and without (in blue) the pseudo-labeling. (%)

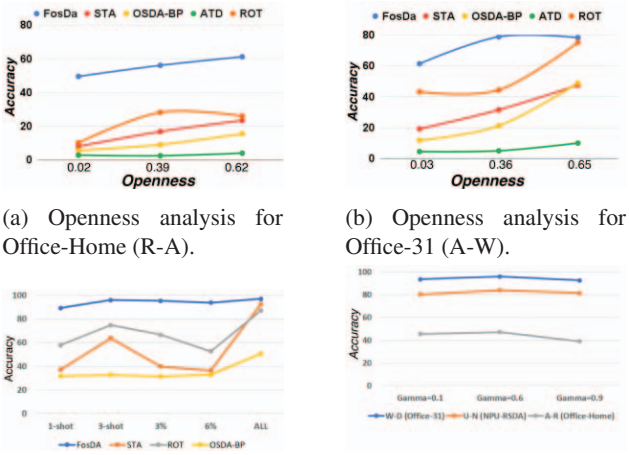
ther improvements of at least 4%. Finally, given all the source domain loss functions, we consider the two adversarial losses individually to analyze their effects. We find that the combination of Eq. 4-5 boosts the performance of the individual adversarial loss measures by $2-4\%$ majorly.

**Openness analysis:** Openness is defined as $\mathcal{O} = 1 - \frac{|C_s|}{|C_t|}$ where $|C_s|/|C_t|$ denotes the number of classes in $\mathcal{S}$ and $\mathcal{T}$, respectively. A large $\mathcal{O}$ signifies that the number of unknown classes is higher than the number of known classes. We compare different values of $\mathcal{O}$ for FosDANet to the other approaches for two experimental cases of Office-31 and Office-Home datasets. As per Fig. 4a-4b, FosDANet maintains high $OS$ scores for different $\mathcal{O}$.

### 6.3. Critical analysis

**Effects of the weight $\gamma$:** The selection of $\gamma$ affects the target performance substantially. If a small $\gamma$ ($\approx 0$) is chosen, it does not help in enhancing the confidence of Eq. 4 whereas a large $\gamma$ ($\approx 1$) may force all the target domain samples to be misclassified as closed-set. We find an intermediate value of 0.6 provides a balanced classification (Fig. 4d).

**Analysis of relation network and the adversary $(\mathcal{G}_2, \mathcal{F})$:** In order to assess the effects of the relation network $\mathcal{R}$, we consider the average class-probability for the known-classes with and without the usage of $\mathcal{L}_{similarity}$ for W-D (Office-31). Basically, the feature learning is governed by $\mathcal{L}_{ID}$ when $\mathcal{R}$ is inactive. As can be observed from Fig. 5c, $\mathcal{R}$ enforces the source samples to be class-concentrated as the class probability is very high with $\mathcal{R}$. Further, the use of $\mathcal{R}$ helps in generating more pseudo-labels (at least 10%) than the models only with classifiers and $\mathcal{H}$. Fig. 5b shows the importance of the dual adversarial alignment over the individual loss of Eq. 4 for W-D (Office-31). Here, we plot the probability for the top two classes for the potential outliers. Ideally, a large gap between the probabilities signi-




(a) Openness analysis for Office-Home (R-A). (b) Openness analysis for Office-31 (A-W).

(c) Sensitivity to more labels. (d) Sensitivity analysis of $\gamma$.

Figure 4: Critical analysis of FosDANet.

| Method | A - C OS | A - P OS | A - R OS | C - A OS | C - P OS | C - R OS | P - A OS | P - C OS | P - R OS | R - C OS | R - A OS | R - P OS | AVG OS | AVG HOS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **OSVM**[12](Source only) | 7.1 | 8.3 | 3.5 | 6.1 | 2.5 | 3.0 | 6.1 | 6.2 | 5.2 | 2.5 | 2.5 | 2.5 | -/ 4.6 | -/7.0 |
| | 2.5 | 2.5 | 4.5 | 15.3 | 27.1 | 7.5 | 14.4 | 16.5 | 13.3 | 27.1 | 25.2 | 43.0 | -/16.6 | -/20.5 |
| **OSDA-BP**[21] | 7.2 | 6.2 | 10.9 | 7.1 | 9.3 | 6.1 | 10.7 | 8.9 | 13.0 | 9.5 | 9.3 | 12.4 | 9.2(±0.9)/10.1 | 10.3/14.5 |
| | 8.8 | 10.8 | 12.5 | 10.1 | 9.6 | 7.9 | 8.4 | 9.0 | 15.4 | 13.1 | 10.5 | 16.0 | 11.0(±0.5)/11.2 | 15.2/15.5 |
| **STA**[16] | 5.1 | 11.2 | 12.0 | 5.9 | 11.1 | 12.8 | 15.1 | 11.4 | 15.9 | 12.2 | 16.8 | 20.6 | 12.5(±0.6)/11.4 | 17.5/14.2 |
| | 12.9 | 15.0 | 21.0 | 12.4 | 18.4 | 19.4 | 17.4 | 19.4 | 26.9 | 20.3 | 21.4 | 25.7 | 19.2(±0.4)/27.9 | 25.3/35.8 |
| **ROT**[1] | 8.0 | 13.5 | 11.5 | 13.9 | 10.5 | 17.6 | 17.2 | 15.3 | 23.2 | 16.0 | 28.2 | 28.8 | 16.9(±1.3)/16.2 | 24.6/21.2 |
| | 12.9 | 18.3 | 25.3 | 17.6 | 14.9 | 20.9 | 20.1 | 24.4 | 38.9 | 29.4 | 51.8 | 46.8 | 26.8(±0.9)/24.7 | 36.1/31.4 |
| **FosDANet** | **28.29** | **37.2** | **47.1** | **43.5** | **49.3** | **51.9** | **46.7** | **44.2** | **61.3** | **44.7** | **56.0** | **62.5** | **47.7**(±1.9)/**40.0** | **49.6/48.6** |
| | **32.3** | **36.1** | **47.3** | **45.1** | **49.4** | **52.2** | **47.9** | **44.9** | **63.4** | **44.9** | **56.2** | **63.1** | **48.1**(±1.2)/**40.9** | **53.2/51.6** |

Table 2: Comparison to the literature for Office-Home when 3% and 6% labeled training data are used in $\mathcal{S}$. For each method, the first row denotes the results of 3% whereas the second row (in gray) shows the results of 6% settings, respectively. In the last column, the average $OS$ values are reported for each case with (in black) and without (in blue) pseudo-labeling. (%)

| Method | AVG OS* | AVG OS | AVG HOS |
|---|---|---|---|
| **OSVM**[12](Source only) | -/ 24.4 | -/ 25.5 | -/29.2 |
| **OSDA-BP**[21] | 24.5(±1.1)/ 24.3 | 24.1(±0.6)/23.6 | 22.1/19.7 |
| **STA**[16] | 43.8(±0.7)/45.0 | 45.9(±0.4)/46.9 | 52.9/53.5 |
| **ROT**[1] | 28.9 (±0.3)/36.5 | 32.2(±0.4)/35.0 | 40.0/25.8 |
| **FosDANet** | **84.5**(±0.7)/**73.9** | **80.7**(±1.3)/**71.5** | 56.7/57.8 |

Table 3: Average accuracy comparisons in % for NPU-RSDA dataset with (in black) and without (in blue) the pseudo-labeling for 3-shot case.

| Method | Adaptiope OS* | Adaptiope OS | Adaptiope HOS | Mini-domainNet OS* | Mini-domainNet OS | Mini-domainNet HOS |
|---|---|---|---|---|---|---|
| **OSDA-BP**[21] | 6.0/6.3 | 6.7/6.6 | 10.7/10.1 | 3.9/3.9 | 4.2/4.0 | 6.7/5.8 |
| **STA**[16] | 11.8/11.5 | 12.4/12.0 | 19.1/18.2 | 5.7/3.3 | 5.8/3.4 | 7.9/5.0 |
| **ATD**[31] | 3.8/3.7 | 4.6/4.2 | 7.1/6.7 | 3.2/3.3 | 3.5/3.4 | 5.7/5.0 |
| **ROT**[1] | 4.6/2.7 | 5.7/2.8 | 8.7/4.3 | 3.7/3.8 | 3.8/3.9 | 5.5/5.6 |
| **FosDANet** | **46.5/20.2** | **47.0/20.1** | **58.4/16.4** | **38.1/23.5** | **38.9/24.2** | **54.4/35.6** |

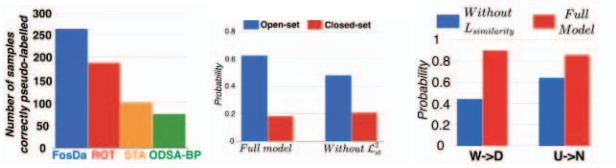
Table 4: Average accuracy comparisons in % for Adaptiope and Mini-domainNet with (in black) and without (in blue) the application of pseudo-labeling for 3% labeled data. Detailed report is provided in the supplementary text.

| Loss function | NPU-RSDA U-N OS* | NPU-RSDA U-N OS | Office-Home R-A OS* | Office-Home R-A OS |
|---|---|---|---|---|
| $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$ + **Eq. 4-5** | 76.2 | 72.4 | 43.2 | 43.2 |
| $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$ + **pseudo-labeling (PL) + Eq. 4-5** | 84.1 | 80.2 | 52.4 | 51.2 |
| $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$ + **PL** + $\mathcal{L}_{ID}$ + **Eq. 4-5** | 86.9 | 83.6 | 54.8 | 54.1 |
| $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$ + **PL** + $\mathcal{L}_{ID}$ + $\mathcal{L}_{similarity}$ + **Eq. 4** | 86.4 | 82.3 | 53.7 | 54.2 |
| $\mathcal{L}^1_{CE} + \mathcal{L}^2_{CE}$ + **PL** + $\mathcal{L}_{ID}$ + $\mathcal{L}_{similarity}$ + **Eq. 5** | 85.9 | 80.6 | 54.1 | 51.8 |
| **FosDANet with PL-random** | 86.7 | 84.1 | 54.3 | 54.9 |
| **FULL FosDANet** | **88.8** | **85.5** | **57.0** | **56.2** |

Table 5: Ablation analysis on the loss components for two cases: U-N (NPU-RSDA) and R-A (Office-Home). (%)

fies a more confident open-set classification which is clearly achieved by FosDANet.

**Effect of the amount of labeled samples in $\mathcal{S}$:** In order to analyze the scalability of FosDANet with more labeled data in $\mathcal{S}$, we plot the performance comparison of all the methods with increasing label information in Fig. 4c for W-D (Office-31) for a fixed $\mathcal{O}$ ($\mathcal{O} = 0.36$ here). It is found that FosDANet provides superior and more consistent performance for all the cases than the comparative techniques.




(a) W-D (Office-31). (b) W-D (Office-31). (c) W-D / U-N.

Figure 5: (a) Correct pseudo-labeled samples after first repeat by different methods. (b) Effect of Eq. 5 on classifying the outliers. (c) Impact of $\mathcal{R}$ on enhancing the confidence on source domain classification.

## 7. Conclusions

We introduce the novel and realistic problem definition of few-shot open-set DA in this paper and propose an end-to-end trainable model called FosDANet as a solution. FosDANet aims to generate a discriminative feature space for the source domain by combining SSL and similarity metric learning based on a novel relation network. This subsequently helps in producing confident pseudo-labels for the unlabeled source data. A dual adversarial learning strategy is then introduced to align the known-class samples from the target domain with the source data in the latent space while rejecting the unknown-class samples with high confidence. Our experimental results on five datasets confirm that FosDANet outperforms the relevant literature by a large margin consistently. We introduce a new benchmark DA dataset consisting of optical satellite images from three visual domains and propose a Mini-domainNet version of the large-scale domainNet dataset. We hope that the considered problem and the proposed solution will open up new avenues in DA research in the low-data regime.

## References


[1] S. Bucci, M. R. Loghmani, and T. Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *European Conference on Computer Vision*, pages 422–438. Springer, 2020. 2, 6, 7, 8

[2] F. M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo, and T. Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019. 2

[3] T. Chen, X. Zhai, M. Ritter, M. Lucic, and N. Houlsby. Self-supervised gans via auxiliary rotation loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12154–12163, 2019. 2

[4] X. Chen, H. Fan, R. Girshick, and K. He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 2

[5] G. Cheng, J. Han, and X. Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017. 6

[6] Q. Feng, G. Kang, H. Fan, and Y. Yang. Attract or distract: Exploit the margin of open set. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2

[7] M. Ghifary, W. B. Kleijn, M. Zhang, and D. Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *Proceedings of the IEEE international conference on computer vision*, pages 2551–2559, 2015. 2

[8] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *European conference on computer vision*, pages 597–613. Springer, 2016. 2

[9] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020. 2

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[11] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5001–5009, 2018. 1

[12] L. P. Jain, W. J. Scheirer, and T. E. Boult. Multi-class open set recognition using probability of inclusion. In *European Conference on Computer Vision*, pages 393–409. Springer, 2014. 2, 7, 8

[13] D. Kim, K. Saito, T.-H. Oh, B. A. Plummer, S. Sclaroff, and K. Saenko. Cross-domain self-supervised learning for domain adaptation with few source labels. *arXiv preprint arXiv:2003.08264*, 2020. 1, 2

[14] D. Kim, K. Saito, T.-H. Oh, B. A. Plummer, S. Sclaroff, and K. Saenko. Cross-domain self-supervised learning for domain adaptation with few source labels. *arXiv preprint arXiv:2003.08264*, 2020. 6

[15] B. Liu, H. Kang, H. Li, G. Hua, and N. Vasconcelos. Few-shot open-set recognition using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2020. 1

[16] H. Liu, Z. Cao, M. Long, J. Wang, and Q. Yang. Separate to adapt: Open set domain adaptation via progressive separation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2, 6, 7, 8

[17] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1406–1415, 2019. 5

[18] T. Ringwald and R. Stiefelhagen. Adaptiope: A modern benchmark for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 101–110, January 2021. 5

[19] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010. 5

[20] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058, 2019. 1

[21] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 153–168, 2018. 1, 2, 5, 6, 7, 8

[22] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult. Towards open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 35, July 2013. 6

[23] B. Sun, J. Feng, and K. Saenko. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016. 1

[24] Y. Sun, E. Tzeng, T. Darrell, and A. A. Efros. Unsupervised domain adaptation through self-supervision. *arXiv preprint arXiv:1909.11825*, 2019. 3

[25] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 1

[26] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 6

[27] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5018–5027, 2017. 5

[28] B. Wallace and B. Hariharan. Extending and analyzing self-supervised learning across domains. In *European Conference on Computer Vision*, pages 717–734. Springer, 2020. 2

[29] M. Wang and W. Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. 1

[30] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin. Unsupervised feature learning via non-parametric instance discrimination. In

*Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018. 2, 3

[31] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018. 8

[32] Y. Yang and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010. 6

[33] X. Yue, Z. Zheng, S. Zhang, Y. Gao, T. Darrell, K. Keutzer, and A. S. Vincentelli. Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13834–13844, June 2021. 1, 2

[34] Y. Zhang, P. David, and B. Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 2020–2030, 2017. 1

[35] W. Zhou, S. Newsam, C. Li, and Z. Shao. Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS journal of photogrammetry and remote sensing*, 145:197–209, 2018. 6