UNIVERSAL DOMAIN ADAPTATION WITHOUT SOURCE DATA FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

Qingsong Xu[1], Yilei Shi[2], Xiaoxiang Zhu[1,3]

1 Data Science in Earth Observation, Technical University of Munich (TUM), Munich, Germany
2 Chair of Remote Sensing Technology (LMF), Technical University of Munich (TUM), Munich, Germany
3 Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9884889

ABSTRACT

Existing domain adaptation (DA) approaches are usually not well suited for practical DA scenarios of remote sensing image classification, since these methods (such as unsupervised DA) rely on rich prior knowledge about the relationship between label sets of source and target domains, and source data are usually not accessible in many cases due to the privacy or confidentiality issues. To this end, we propose a novel source data generation-based universal domain adaptation (SDG-UniDA) model, which includes two parts, i.e., the stage of source data generation and the stage of model adaptation. The first stage is to estimate the conditional distribution of source data from the pre-trained model using the knowledge of class-separability in the source domain and then to synthesize the source data. With this synthetic source data in hand, it becomes a universal DA task that requires no prior knowledge on the label sets. A novel transferable weight is proposed to distinguish the shared and private label sets to each domain, thereby promoting the adaptation in the automatically discovered shared label set and recognizing the "unknown" samples successfully. Empirical results show that SDG-UniDA is effective and practical in this challenging setting for remote sensing image scene classification.

Index Terms— Source data generation, universal domain adaptation, remote sensing image classification

1. INTRODUCTION

Domain Adaptation (DA) aims to leverage a source domain to learn a model that performs well on a different but related target domain. Most existing DA approaches for remote sensing image [1] are proposed to tackle the domain gap between different domains by learning a domain invariant feature representation. However, these DA approaches require the knowledge of the relationship between the source and target label space (category-gap). For example, the adversarial learning-based DA methods for remote sensing images [1] assume a shared label set between the source and target domains.

Recently, universal domain adaptation (UniDA) has attracted extensive attentions, which removes all constraints meanwhile includes all the above adaptation settings [2]. Two challenges are exposed in UniDA setting: 1) large domain gaps, and 2) category gaps. Most recently, the source-free domain adaptation is under continuous exploration [3]. However, existing UniDA methods have not been explored on remote sensing datasets, and usually assume that the source dataset is available when building the source classifier platform. In real application scenarios of remote sensing image classification, developing a universal domain adaptation method without source data has a high practical value and is thus highly desired. In such cases, *pre-trained models can be available*, which not only serve as baselines for the original remote sensing dataset, but also contain knowledge of the original dataset. Therefore, how to generate *synthetic source domain data* from the pre-trained model is the first problem to be solved. In this paper, we propose the **UniDA without source data** in order to firstly introduce the UniDA setting into remote sensing datasets. In this case, we merely have access to the pre-trained model from the source domain. UniDA without source data poses two major technical challenges for designing the corresponding models in the wild. 1) Distilling the knowledge of source data from the pre-trained model. The knowledge is consistent with the source in the data distribution. 2) Domain adaptation should be applied to align distributions of the *synthetic* source and target data in the presence of domain gaps and category gaps. To address these two challenges, our proposed source data generation-based universal domain adaptation (SDG-UniDA) consists of a source data generation stage and a model adaptation stage. In the source data generation stage, we reformulate the goal as to estimate the conditional distribution of source data instead of the data distribution. After obtaining the conditional distribution of source data, in order to further separate the target samples from the shared label set and those from the private label, a novel *transferable weight* is defined by considering the confidence and domain similarity. In a nutshell, our contributions are listed as follows.

- We introduce a more practical and challenging UniDA setting for remote sensing image scene classification.
- We propose a new SDG-UniDA model composed of a

source data generation stage and a model adaptation stage.

- In order to generate reliable source domain samples, a novel *conditional probability recovery method* of the source domain is designed to distill the category knowledge.
- A novel *transferable weight* is utilized to distinguish the shared label sets and the private label sets to each domain.
- Experimental results on three UniDA settings for remote sensing image scene classification demonstrate that the proposed model is effective.

2. METHODOLOGY

In this section, we elaborate the problem of UniDA without source data and address it by a novel SDG-UniDA method.

2.1. Problem setting

For the UniDA without source data, we merely have access to the pre-trained model M from the source domain, including feature extractor F and classifier C. We have no information about the source data distribution $p(x)$ that was used to train M. Considering the domain adaptation in the second stage, our goal is to synthesize the source data x_f from the pre-trained model M, which follows the source data distribution $p(x)$. The distribution is consistent with the source in the category distribution (including the shared label set and the private label set), and is as close as possible to the target in style. However, it is impracticable to estimate $p(x)$ directly since the source data space is exponential with the dimensionality of data. Thus, we generate the set by modeling a conditional probability of x given two random vectors y and z. y ($y \sim p_y(y)$) is a probability vector that represents a label, where $p_y(y)$ is an estimation of the true labeled distribution $p(y_s)$ of the source domain. z ($z \sim p_z(z)$) is a low-dimensional noise, where $p_z(z)$ is a random distribution describing the source data points. Thus, we reformulate the goal as to estimate the conditional distribution of source data $p(x \mid y, z)$ instead of the distribution $p(x)$. After obtaining the conditional distribution of source data $p(x \mid y, z)$ from the source data generation stage, it becomes a traditional UniDA task but now with synthetic source domain. A synthetic source domain and a target domain are represented by $D_f = \left\{\left(x_f^i, y_f^i\right) \sim p\right\}_{i=1}^{n_f}$ sampled from distribution $p(x \mid y, z)$ and $D_t = \left\{\left(x_t^i\right) \sim q\right\}_{i=1}^{n_t}$ sampled from distribution q, respectively. We denote by Y_f (Y_t) the label set of the fake source (target) domain. The shared label set is denoted by $Y = Y_f \cap Y_t$. The private label sets of the source and target domain are represented by $\overline{Y_f} = Y_f \backslash Y$ and $\overline{Y_t} = Y_t \backslash Y$, respectively.

2.2. Source data generation

In order to generate a reliable source domain for UniDA, the generated data x_f must meet two conditions: 1) in data content, all category distributions in the pre-trained model M can be restored, including source-share and source-private category distributions, and 2) in data style, the generated data can remain similar to the target domain style distribution. Digging further, these two conditions are to ensure the data diversity of the generated source domain.

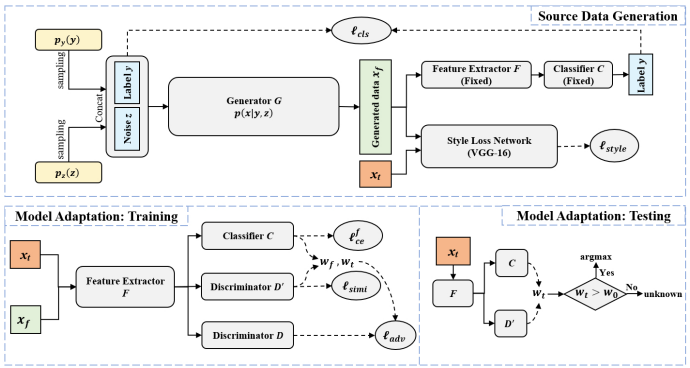

Fig. 1. Overview of the proposed UDA without source data. The model consists of a source data generation stage and a model adaptation stage.

First, in order to recover the category distributions from the pre-trained model M, As shown in the 'Source Data Generation' module in Fig. 1, a classifier loss is designed. Specifically, given a sampled class vector y and a sampled noise vector z as inputs, G is trained to produce a synthetic source domain sample that M is likely to classify as $\hat{y}$. The classifier loss can force the generated data to follow the similar class distribution from model M, by minimizing the distance between y and $\hat{y}$, which can be formulated as follows:

$$\ell_{\text{cls}}(y, \hat{y}) = -\sum_{i \in Y_f} y_i \log M(G(y, z))_i. \tag{1}$$

Notably, y and $\hat{y}$ are not scalars but probability vectors of length Y_f. Thus, the *cross-entropy* between two probability distributions is utilized to measure the distance between y and $\hat{y}$.

However, the classifier loss ℓ_{cls} easily leads to generate similar data points for each class in the synthetic source domain. Furthermore, it is necessary for domain adaptation to transfer synthetic source images to the target style. A *style loss* is presented to measure differences in style between a synthetic source image x_f and a target image x_t. Concretely, we make use of a 16-layer VGG network pretrained on the ImageNet to measure multi-scale feature style differences between images, which can be described as:

$$\ell_{\text{style}}(x_f, x_t) = \sum_{j=1}^{4} \left\| G_j^{\phi}(x_f) - G_j^{\phi}(x_t) \right\|_F^2, \tag{2}$$

$$G_j^{\phi}(x) = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{c,h,w} \phi_j^T(x)_{c,h,w}, \tag{3}$$

where $\phi_j(x)$ is the activation at the jth layer of the style loss

network, which is a feature map of shape $C_j \times H_j \times W_j$. $G_j^{\phi}(x)$ denotes Gram matrix, which is equal to the average value of the product between the feature and the transposition of the feature. The Gram matrix can grasp the general style of the entire image. The style loss $\ell_{\text{style}}(x_f, x_t)$ is the squared Frobenius norm of the difference between the Gram matrices of synthetic source image x_f and target image x_t. In addition, different layers have different feature style in the VGG network. Therefore, we sum the Gram matrices difference of four activation layers in the VGG-16.

Finally, we train generator G by minimizing,

$$\ell(\theta_g) = \min_{\theta_g}(\ell_{\text{cls}}(y, M(x_f)) + \ell_{\text{style}}(x_f, x_t)). \quad (4)$$

2.3. Model adaptation

The objective of model adaptation is to update the pre-trained model M that distinguishes samples from the target shared label set Y and those in the target privated label set $\overline{Y_t}$. One important challenge for UniDA is detecting transferable samples. In order to address this challenge, the sample transferable weight $w_f(x_f)$ or $w_t(x_t)$ is utilized to estimate the confidence that x_f or x_t is from the shared label set during the training stage. Furthermore, during testing stage, we use the transferable weight as a decision threshold w_0 to decide whether we should predict a class or mark the sample as "Unknown" that represents all labels unseen during training. This is

$$y(x_t) = \begin{cases} Class & w_t(x_t) > w_0 \\ Unknown & \text{otherwise} \end{cases} \quad (5)$$

The transferable weight is derived from uncertainty and domain similarity. Similar to [2], the domain similarity $d(x)$ is obtained by the non-adversarial domain discriminator D'. $d(x)$ can be seen as the quantification for the similarity of target domain samples to the synthetic source domain samples. Namely, a smaller $d(x_f)$ for a synthetic source sample and a larger $d(x_t)$ for a target sample mean that they are more likely to be in the shared label set.

On the other hand, we adopt the assumption that the target data in Y have a lower uncertainty than target data in $\overline{Y_t}$. Thus, in order to further separate between target samples from the shared label set and those from the private label, a well-defined criterion can be used to distinguish different degrees of uncertainty. However, the uncertainty is usually measured by entropy [2], which lacks discriminability for uncertain when the categorical distribution are relatively uniform. The confidence of predicted probabilities $\bar{y}(x)$ is a better measure when the generated categories of source samples are relatively uniform. Thus, the sample-level transferable weight for synthetic source data points and target data points can be respectively defined as,

$$w_f(x) = -d(x) - \max \bar{y}(x), \quad (6)$$

$$w_t(x) = d(x) + \max \bar{y}(x). \quad (7)$$

Note that $d(x) \in [0, 1]$ and $\max \bar{y}(x) \in [0, 1]$. The weights are also normalized into interval $[0, 1]$ during training.

Domain adaptation aims to move the target samples with higher transferable weight towards positive source categories Y. To achieve this, as shown in Fig. 1, input x from either domain is fed into the feature extractor F. The extracted features $F(x)$ is forwarded into the label classifier C and the non-adversarial domain discriminator D', to obtain the transferable weights w_f and w_t. The extracted features $F(x)$ is forwarded into the adversarial domain discriminator D to adversarially align the feature distributions of the generated source and target data falling in the shared label set. Thus, the adversarial loss function for adaptation is defined as,

$$\begin{aligned} \ell_{\text{adv}} = &- \mathbb{E}_{\mathbf{x}\sim p} w_f(x) \log D(F(x)) \\ &- \mathbb{E}_{\mathbf{x}\sim q} w_t(x) \log(1 - D(F(x))). \end{aligned} \quad (8)$$

The feature extractor F strives to confuse D. Thus, domain-invariant features in the shared label set are obtained. In order to train the classifier C on the synthetic source domain with labels, the cross-entropy loss is the following,

$$\ell_{\text{ce}} = \mathbb{E}_{(x_f, y_f)\sim p} L(y_f, C(F(x_f))), \quad (9)$$

where L is the standard cross-entropy loss. Furthermore, to better reflect domain similarity, we predict samples from synthetic source domain as 1 and samples from target domain as 0. Thus a binary cross-entropy loss is used to train non-adversarial domain discriminator D'.

$$\begin{aligned} \ell_{\text{simi}} = &- \mathbb{E}_{(x_f, y_f)\sim p} L(1, D'(F(x_f))) \\ &- \mathbb{E}_{(x_t, y_t)\sim q} L(0, D'(F(x_t))). \end{aligned} \quad (10)$$

Thus, the training of model adaptation stage can be written as a minimax game,

$$\ell(\theta_d, \theta_f, \theta_c) = \max_{\theta_d} \min_{\theta_f, \theta_c} (\ell_{\text{ce}}^f - \ell_{\text{adv}}), \quad (11)$$

$$\ell(d') = \min_{\theta_{d'}} (\ell_{\text{simi}}). \quad (12)$$

3. EXPERIMENTS

3.1. Experimental setup

Datasets. To verify our algorithm, we select the UC Merced, AID, and RSSCN7 to build the cross-domain remote sensing image scene datasets. The **UC Merced dataset** is a widely used dataset for remote sensing image scene classification. It consists of 2100 remote sensing images from 21 scene classes. Each scene class contains 100 RGB images with an image size of 256×256 pixels. The **AID dataset** is a large scale aerial image data set and acquired from Google Earth. It contains 10,000 images with a size of 600×600 pixels, which are divided into 30 classes. The **RSSCN7 dataset** contains 2800 remote sensing scene images, which are from seven typical scene categories. There are 400 images in each scene type,

Fig. 2. Sample images from five public categories.

and each image has a size of 400×400 pixels. Two UniDA tasks for remote sensing scene classification are established. The first one is from RSSCN7 to UC Merced. We use the five public categories as the shared label set (Fig. 2), namely farmland, forests, dense residential areas, rivers, and parking lot, the remaining two as the private source label set and the remaining sixteen of UC Merced as the private target label set. The other is from RSSCN7 to AID. We use the six public categories as the shared label set (adding industries), the remaining one as the private source label set and the rest of AID as the private target label set. Notably, the model is tested only on samples from the target domain and all the target-private classes are grouped into a single "Unknown" class.

Implementation details We use the standard normal vectors z of length 10 in all experiments. The generator G consists of two fully connected layers followed by seven transposed convolutional layers. The size of the generated image x_f is $3 \times 256 \times 256$. Adam with a learning rate of 0.001 is used for the generator. A ResNet-50 model is used as the backbone of the feature extractor F. The classifier network C is a fully connected network with a single layer. The discriminators D and D' consist of three fully connected layers with ReLU between the first two. During testing stage, the decision threshold $w_0 = 0.8$.

3.2. Discussion of classification results

Table 1. Average class accuracy (%) on RSSCN7 → UC Merced and RSSCN7 → AID.

RSSCN7 → UC Merced

Stage	Mehod	Farmland	Forests	Dense residential	Rivers	Parking	Unknown	Avg
UniDA	Pretrained model-only	77.00	13.00	55.00	69.00	77.00	0.44	48.57
	SDG-UniDA	92.00	64.00	63.00	76.00	93.00	17.13	67.52
	UniDA with source data	85.00	98.00	74.00	79.00	100.00	14.63	75.10
SDG	GAN-based Method	91.00	1.00	1.00	9.00	87.00	15.31	34.05
	Decoder Loss	31.00	12.00	21.00	82.00	72.00	23.13	40.19
	Our Diversity Loss	92.00	64.00	63.00	76.00	93.00	17.13	67.52

RSSCN7 → AID

Stage	Method	Farmland	Forests	Dense residential	Rivers	Parking	Industries	Unknown	Avg
UniDA	Pretrained model-only	91.89	16.00	55.37	73.41	56.67	59.49	0.36	50.46
	SDG-UniDA	97.84	69.20	78.78	37.32	96.92	62.05	17.93	65.72
	UniDA with source data	92.70	100.00	94.63	58.78	94.36	70.51	13.48	74.92

We present the UniDA setting results from the RSSCN7 dataset to the UC Merced dataset and from the RSSCN7 dataset to AID dataset in Table 1. The baseline method is a Pretrained model-only, which only uses the pretrained model from source data for training and tests on the target domain, achieving an overall accuracy of 48.57% on RSSCN7 → UC Merced and 50.46% on RSSCN7 → AID. On the contrary, UniDA with source data is the case of traditional UniDA and its performance can be considered as an upper bound for the UniDA performance. We observe that SDG-UniDA significantly outperforms the Pretrained model-only method by +18.95% and +15.26% on RSSCN7 → UC Merced and RSSCN7 → AID, respectively. It is obvious that SDG-UniDA is effective and practical in UniDA of remote sensing images, although the real source data is unavailable during the entire training process. Notably, compared with the upper bound (UniDA with source data), our SDG-UniDA maintains a more prominent performance in "Unknown" class. It has been demonstrated that data points generated by the SDG-UniDA effectively covers the distribution of the source data.

In addition, since data diversity is the key, we analyse the data diversity in the SDG stage in Table 1. It can be seen that the generating ability of our proposed diversity loss is significantly better than that of the GAN-based method [4] and the decoder loss [5], in terms of solving the problem of source domain generation in UDA without source data.

4. CONCLUSIONS

We have introduced a novel UniDA without source data framework for remote sensing image scene classification, which consists of the source data generation stage and the model adaptation stage. This work can be served as a start point in a challenging UniDA setting for remote sensing images.

5. ACKNOWLEDGEMENT

This work is supported by China Scholarship Council.

6. REFERENCES

[1] Qingsong Xu, Xin Yuan, and Chaojun Ouyang, "Class-aware domain adaptation for semantic segmentation of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[2] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan, "Universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2720–2729.

[3] Jogendra Nath Kundu, Naveen Venkat, R Venkatesh Babu, et al., "Universal source-free domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4544–4553.

[4] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu, "Model adaptation: Unsupervised domain adaptation without source data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9641–9650.

[5] Jaemin Yoo, Minyong Cho, Taebum Kim, and U Kang, "Knowledge extraction with no observable data," 2019.