Supplementary materials
"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]
{sayan1by2@gmail.com, sbalasubramanian@sssihl.edu.in, hmrishavbandyopadhyay@gmail.com
pbharambe196@gmail.com, nandand1998@gmail.com, getbiplab@gmail.com, sc@ee.iitb.ac.in}
Indian Institute of Technology Bombay[1], SSSIHL[2], Jadavpur University [3], India

We discuss the following aspects regarding our paper in this supplementary text:

- Details about the newly curated remote sensing dataset (NPU-RSDA).
- Qualitative results for cross-domain retrieval task corresponding to the shared classes under the FosDA scenario.
- Description of the model architecture in detail
- Detail Discussion on Eq-4 and Eq-5 (main paper) and the effect with and without repetitive pseudo labeling on total loss convergence.
- More quantitative results for Adaptiope and mini-domainNet (3% case) and NPU-RSDA (3-shot and 1-shot), respectively.

1. More details about the NPU-RSDA dataset

As already stated in the main paper (Section 6), the proposed NPU-RSDA dataset has been curated from three publicly available benchmark remote sensing optical image datasets. In addition to what is already stated in the main paper, we provide further details regarding these datasets in the following.

UC-Merced [7]: This is a widely used dataset for the classification of land use and land cover classes. This dataset consists of 21 different classes and each class contains 100 aerial images of size 256×256 with the spatial resolution of 0.3m. The images are extracted from large aerial images downloaded from the U.S. Geological Survey (USGS) National Map over the regions of Birmingham, Boston, Buffalo, Columbus, Dallas, Harrisburg, Houston, Jacksonville, Las Vegas, Los Angeles, Miami, Napa, New York, Reno, San Diego, Santa Barbara, Seattle, Tampa, Tucson, and Ventura. The 21 classes are: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, free- way, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts.

NWPU-RESISC45 [2] : Currently, NWPU-RESISC45 is considered as one of the largest benchmark datasets for remote sensing scene classification. This dataset was extracted from the Google Earth (Google Inc.) images, which are obtained from satellite imagery, aerial photography, and geographic information system (GIS) onto a 3-D globe, respectively. This dataset contains 45 scene classes and each class consists of 700 images. The classes are airplane, airport, baseball diamond, basketball court, beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense-residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium-residential, mobile home park, mountain, overpass, palace, parking-lot, railway, railway station, rectangular farmland, river, round-about, runway, sea ice, ship, snow berg, sparse residential, stadium, storage tank, tennis court, terrace, thermal power station, and wetland. The images are acquired from over 100 countries and regions all over the world and the spatial resolution of most of the classes varies in the range of 0.2m to 30m except some of the classes like lake, mountain and island, which tend to have a lower resolution.

PatternNet [8]: This dataset was created for the purpose of remote sensing image retrieval but the dataset can also be used for remote sensing image classification. The images in this dataset were collected from Google earth imagery or via the GoogleMap API for different US cities and representing 38 visual classes. Each class consists of 800 images of size 256×256. The classes are airplane, baseball field, basketball court, beach, bridge, cemetery, chaparral, Christmas tree farm, closed road, coastal man-sion, crosswalk, dense residential, ferry terminal, football field, for-est, freeway, golf course, harbor, intersection, mobile

home park,nursing home, oil gas field, oil well, overpass, parking lot, parking-space, railway, river, runway, runway marking, shipping yard, solar panel, sparse residential, storage tank, swimming pool, tennis court,transformer station and wastewater treatment plant. The spatial resolution is quite high and ranges from 0.062 to 4.693 meters.

Our Contribution(NPU-RSDA): We identify 17 classes which are common or closely related among all the three domains(UC-Merced, NPU-RESISC45, PatternNet). The classes are represented as airplane, baseball diamond, beach, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, mobile home park, overpass, parking lot, river, sparse residential, storage tank, and tennis court, respectively. For UC-Merced(U) and PatternNet(P) the total number of images are 1700 and 13600 respectively. For NWPU-RESISC45(N), we remove 200 noisy images over these 17 classes and a total of 11700 images present in this domain. Overall, we have 27000 images of 17 classes spanning across the three datasets (domains) that constitute the NPU-RSDA data. The variations in the spatial, spectral properties, different climatic conditions during image acquisition, geographical aspects, etc. induce the domain differences in this dataset. Sample images for this dataset can be seen in Fig. 1. We hope that NPU-RSDA will be helpful to the remote sensing community for bench-marking different domain adaptation tasks.

2. Cross-domain image retrieval results

In order to showcase the quality of domain alignment for the closed-set classes, we conduct an image retrieval experiment where the query sample comes from the source domain and the task is to retrieve semantically identical images from the target domain. Fig. 4 depicts results for cross-domain retrieval for images from A-D (Office-31), A-R (Office-Home), and P-N (NPU-RSDA). As we observe, FosDANet is able to produce highly precise retrieval results for the different cases.

3. Model Architecture

Fig. 2 highlights the model architecture in detail corresponding to section 4 of the main paper. We specify the number of layers, size of the latent space, etc. in this figure part by part of the full model diagram in the main paper (Figure 2).

4. Discussion on Eq. 4-5 and total Loss convergence with respect to with and without repetitive pseudo labelling

Let the sample prob. w.r.t the open class be p_o. In Eq. 4(main paper), $\mathcal{F}$ and $\mathcal{G}_1$ are pitted against each other. While $\mathcal{G}_1$ attempts to keep $p_o = 0.5$, $\mathcal{F}$ tries to push it to 0 or 1. It is easier for $\mathcal{F}$ to push more open samples' prob. closer to 1, and the less open samples' prob. closer to 0, leading to over-fitting with the closed classes. To tackle this, we pit $\mathcal{F}$ against $\mathcal{G}_2$ in Eq. 5. $\mathcal{G}_2$ attempts to keep $p_o = 1$ while $\mathcal{F}$ tries to push it below 1. Since there's equal competition amongst the adversaries, the more open samples would have p_o much greater than 0.5. Importantly, note that the less open samples' prob. are now being pushed to 0.5 by $\mathcal{G}_1$ and also to 1 by $\mathcal{G}_2$. Thus, $\mathcal{F}$ has *dual* resistance in trying to ensure p_o gets a low value. Thus, average p_o is likely to be higher than when Eq. 4 is used alone. This has been empirically validated in Fig. 5(b)(main paper), and is illustrated in the figure-3(b). Thus, unknown samples that are similar to known ones are likely to be more confidently predicted as unknown rather than being over-fitted to a known class. This is the regularization effect of Eq. 5(main paper).

From figure-3(a), it can be seen that the without repetitive pseudo labeling gives us more stable convergence than the repetitive pseudo labeling.

5. Evaluation on Adaptiope and mini-domainNet

We mention the detailed results for mini-domainNet and Adaptiope for the 3% labeled data, and NPU-RSDA for the 1-shot and 3-shot cases in Table 1-3. The average accuracy on mini-domainNet and Adaptiope are preset in the Table:4 of the main paper where we also mentioned the average accuracy without the pseudo-labeling.

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. 4, 6

[2] 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. 1

[3] 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. 6

[4] 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. 4

[5] 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. 4, 6

[6] 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. 4, 6

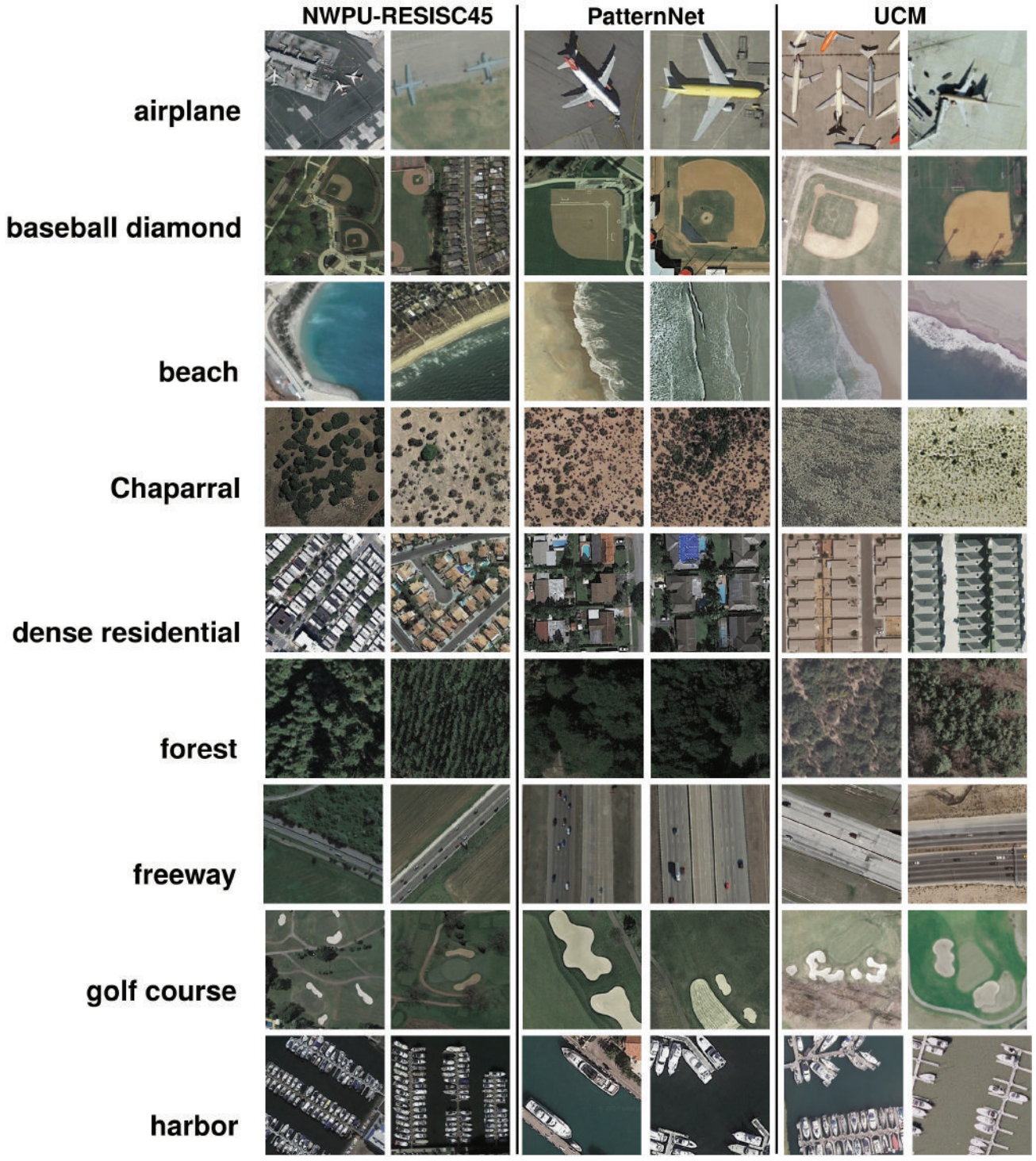

Figure 1. Examples of Domain-wise images of alphabetically the first 10 classes of NPU-RSDA

[7] 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. 1

[8] 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. 1

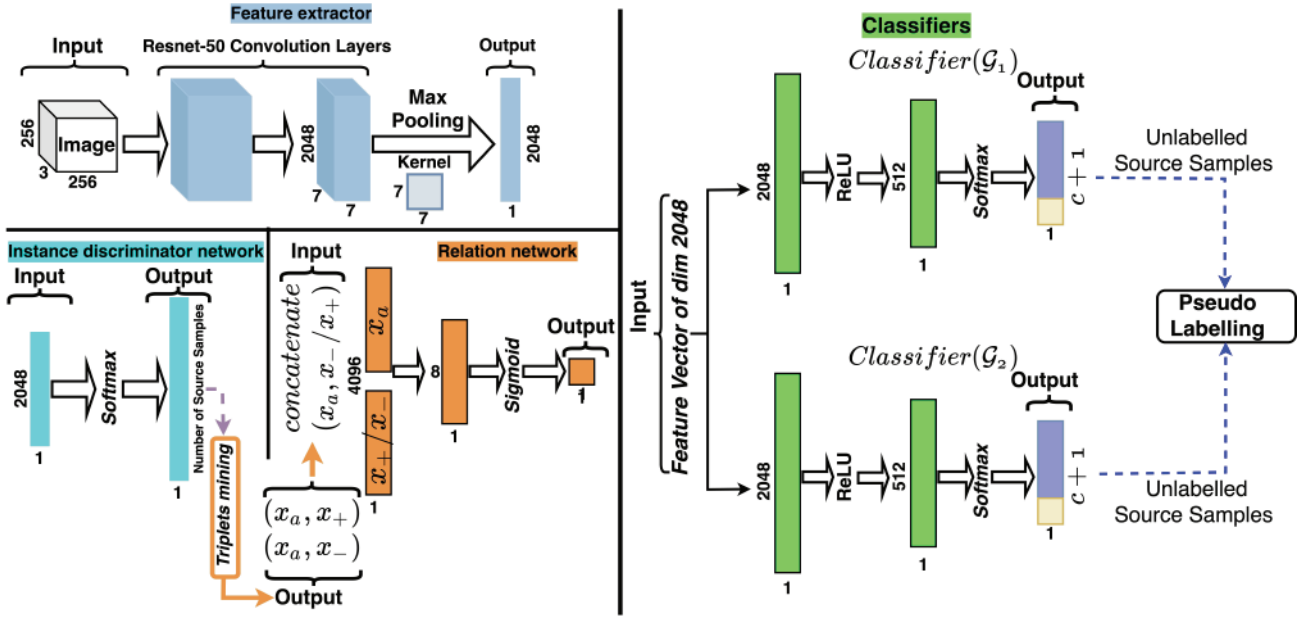

Figure 2. Detained architecture design for FosDANet. We show the architectures for the feature extractor, instance discriminator, relation network, and the classifiers, respectively, in separate diagrams. The number of dense layers in a sub-module is specified numerically.

Method	N - U		N - P		U - P		U - N		P - U		P - N		AVG	
	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS
1-Shot														
OSDA-BP [6]	27.1	26.8	26.9	29.5	25.5	24.0	25.5	29.9	35.5	33.3	25.3	29.4	27.6 (±0.3)/ 26.1	28.8(±0.2)/28.2
STA [5]	14.7	13.4	23.4	21.3	24.3	27.2	29.1	27.4	18.7	17.0	14.4	13.1	20.8 (±0.4)/ 24.9	19.9 (±0.5)/23.6
FosDANet (1-shot)	**58.7**	**57.3**	**55.3**	**55.3**	**77.3**	**72.4**	**84.1**	**80.5**	**78.4**	**74.4**	**58.2**	**56.7**	**68.7**(±0.7)/**53.6**	**66.1**(±1.3)/**55.2**
3-Shot														
OSVM [4](Source only)	14.8	22.0	20.8	26.9	29.1	27.2	10.9	10.3	43.3	40.3	27.5	26.1	-/ 24.4	-/ 25.5
OSDA-BP [6]	21.5	20.3	24.8	23.7	24.5	24.1	22.9	21.9	32.1	34.2	21.3	20.5	24.5(±1.1)/ 24.3	24.1(±0.6)/23.6
STA [5]	33.3	36.6	64.5	66.0	52.1	54.0	28.8	29.4	47.7	49.2	36.8	40.5	43.8(±0.7)/45.0	45.9(±0.4)/46.9
ROT [1]	52.6	51.1	13.3	20.7	19.7	26.9	18.6	20.3	45.7	46.4	23.8	27.6	28.9 (±0.3)/36.5	32.2(±0.4)/35.0
FosDANet (3-shot)	**66.1**	**64.8**	**88.5**	**84.0**	**96.6**	**92.1**	**88.8**	**85.5**	**90.1**	**83.6**	**77.1**	**74.0**	**84.5**(±0.7)/**73.9**	**80.7**(±1.3)/**71.5**

Table 1. Comparison to the literature for the NPU-RSDA dataset for the 1-shot and 3-shot cases. For each method, for the average values (last column), we show the performance of the model with (in black) and without (in blue) the pseudo-labeling. (%)

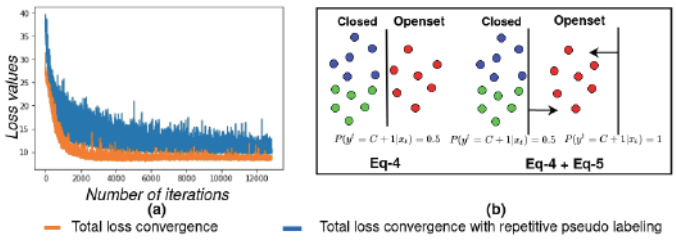

Figure 3. (a)Total loss convergence with and without repetitive pseudo labeling (b) Semantically compared observation of the effectiveness of single Eq-4 and combination of Eq-4 + Eq-5 in main paper

Figure 4. Qualitative results for the cross-domain retrieval task. The query image is taken from the source domain while the retrieved images are from the target domain. Retrieval is performed on the domain aligned feature space. We showcase the top-5 retrieval results.

Method	P - R		P - S		S - P		S - R		R - P		R - S		AVG	
	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS
OSDA-BP [6]	7.8	8.8	5.1	5.6	4.8	5.3	4.5	4.8	10.4	10.8	3.4	5.2	6.0	6.7
STA [5]	24.2	24.4	8.9	9.6	6.5	6.9	7.7	8.0	19.3	20.3	4.2	5.3	11.8	12.4
ATD [3]	5.6	6.3	4.6	5.3	2.8	3.4	3.6	4.5	4.1	5.2	2.1	2.9	3.8	4.6
ROT [1]	19.7	20.8	0.1	2.0	4.3	5.7	0.7	1.8	1.9	2.4	1.0	1.7	4.6	5.7
FosDa-Net	**61.6**	**61.5**	**49.5**	**50.6**	**41.6**	**42.2**	**20.4**	**21.5**	**73.7**	**73.3**	**32.2**	**32.9**	**46.5**	**47.0**

Table 2. Average accuracy comparisons in % for Adaptiope for 3% labeled data.

Method	P - R		P - S		S - P		S - R		R - P		R - S		AVG	
	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS	OS*	OS
OSDA-BP [6]	4.3	4.6	3.7	3.6	3.2	3.5	3.9	4.5	4.0	4.3	4.1	4.6	3.9	4.2
STA [5]	6.0	6.1	5.0	5.1	3.6	3.6	4.0	4.1	7.9	8.0	7.6	7.7	5.7	5.8
ATD [3]	4.4	3.9	3.5	3.7	2.7	2.0	2.5	3.1	3.4	4.4	2.7	3.9	3.2	3.5
ROT [1]	1.1	1.1	2.0	2.2	3.0	2.9	1.5	1.4	10.3	10.3	4.7	4.9	3.7	3.8
FosDa-Net	**43.4**	**44.1**	**31.1**	**32.2**	**32.3**	**33.2**	**43.9**	**44.7**	**41.0**	**41.4**	**37.1**	**37.9**	**38.1**	**38.9**

Table 3. Average accuracy comparisons in % for mini-domainNet for 3% labeled data.