

Supplementary Materials: MultiDAN: Unsupervised, Multistage, Multisource and Multitarget Domain Adaptation for Semantic Segmentation of Remote Sensing Images

Anonymous Authors

1 EXPERIMENTS

1.1 Results and Discussions

1.1.1 Visual Feature Distributions. Figure 1 exhibits the feature distributions of the baseline model without UDA, existing MSMTDA method [4] and our proposed MultiDAN on the adaptation from multisource Paris, Berlin and Tokyo datasets to multitarget Zurich, Chicago and Potsdam datasets. The 2-D space feature distributions are acquired via the t-distributed stochastic neighbor embedding (t-SNE) [3]. As shown in Figure 1, the features of different classes are mixed and confused without UDA technique. Compared with baseline model, the existing MSMTDA method [4] effectively reduce the feature confusion between various classes. By solving the multiple domain shift (simultaneous inter-domain shift and intra-domain shift) problems, the proposed MultiDAN achieve a better and clearer feature transformation against the existing MSMTDA method, showing the effectiveness and superiority of our proposed MultiDAN.

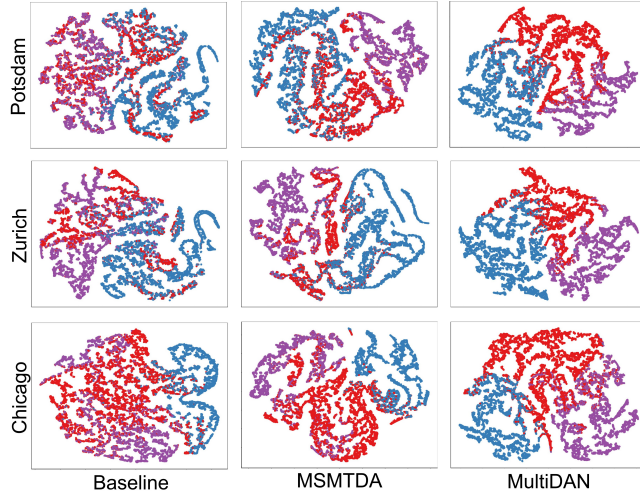


Figure 1: Visual feature distributions of baseline (source-only), MSMTDA method [4] and the proposed MultiDAN shown through t-SNE [3] on the Adaptation From multisource Paris, Berlin and Tokyo datasets to multitarget Zurich, Chicago and Potsdam datasets. Red, blue and purple denotes categories background, building and road respectively.

1.2 Ablation Study

1.2.1 Influence of Stage Number K . The stage number (subdomain number of each target domain) K is a crucial factor for our MultiDAN. Hence, to probe the effect of stage number K , we attempt MultiDAN trained with various stage number K . Figures 2 and 3 depict the visual predictions and entropy of MultiDAN trained with various stage number K . From Figures 2 and 3, we can see that the visual segmentation predictions and entropy maps firstly become better and then turn slightly worse with the continuous increasing of stage number K (from 5 to 8). This indicates our MultiDAN can achieve top performance without too many training stages.

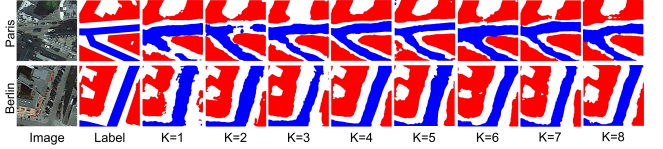


Figure 2: Visual segmentation prediction maps of the proposed MultiDAN with various stage numbers (subdomain number) K on the adaptation from Zurich and Chicago Datasets to Paris and Berlin Datasets.

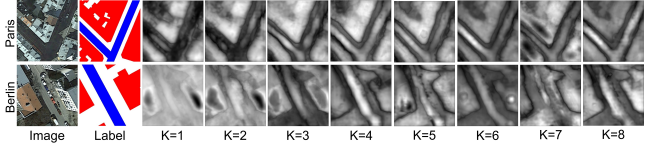


Figure 3: Visual entropy maps of the proposed MultiDAN with various stage numbers (subdomain number) K on the adaptation from Zurich and Chicago Datasets to Paris and Berlin Datasets.

1.2.2 Analysis of Pseudo Label Update Strategy. In this section, we explore the influence of the proposed PLUS by conducting numerous comparison experiments on the AIS datasets. We discuss the methods of determining the initial values of probability threshold μ and entropy threshold v , and probe the effect of initial values of μ and v .

To decide the initial values of probability threshold μ and entropy threshold v , we follow [2] and make a trade off between quality and quantity of pseudo labels. Concretely, we apply pixel ratios ρ_u and ρ_v , which corresponds to the ratio between the reserved pixels and total pixels, to determine the initial values of probability threshold μ and entropy threshold v . In other words, we only reserve the top ρ_u pixels whose probability values are higher than probability threshold μ . Similarly, we only maintain the top ρ_v pixels whose

entropy values are lower than entropy threshold v . It's notable that too small or too large pixel ratios ρ_u and ρ_v will weaken the segmentation performances. Thus, we draw the correlations between prediction entropy threshold v and pixel ratio ρ_v , prediction confidence (softmax probability) threshold μ and pixel ratio ρ_u in the left side and right side of Figure 4 respectively. From left side of Figure 4, we can see that the prediction entropy decreases approximately linearly as pixel ratio drops from 95% to 65%. When pixel ratio decreases from 65% to 40%, the prediction entropy drops rapidly. On this basis, entropy threshold v , which corresponds to inflection point $\rho_v=65\%$, is applied to reach the trade-off between quality and quantity of pseudo labels. Similarly, probability threshold μ , corresponding to inflection point $\rho_u=75\%$, is selected as depicted in right side of Figure 4.

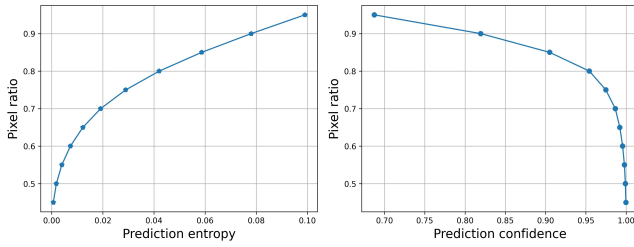


Figure 4: Correlations between the pixel ratio and prediction entropy (left), pixel ratio and prediction confidence (right).

Table 1: Influence (mIoU) of Initial Softmax Probability Threshold μ_0 of the Pseudo Label Update Strategy on the Adaptation From Zurich and Chicago datasets to Paris and Berlin datasets.

ρ_u	40%	50%	60%	70%	75%	80%	90%	100%
Paris	49.3	50.2	50.7	51.1	51.5	51.3	51.0	50.6
Berlin	51.2	51.9	52.4	52.9	53.2	53.0	52.6	52.1

To verify the effect of parameters ρ_u and ρ_v , we learn MultiDAN with various ρ_u (corresponding to initial value of probability thresholds μ) and ρ_v (corresponding to initial value of entropy thresholds v), and give the mIoU results in Tables 1 and 2. As illustrated in Tables 1 and 2, the model reaches the best segmentation performance on all the target datasets when $\rho_u=75\%$ and $\rho_v=65\%$. The too high and too low pixel ratio (ρ_u or ρ_v) will highly influence the segmentation performance of MultiDAN.

1.2.3 Analysis of Adaptive Weighting Strategy. In this section, we validate the effectiveness of adaptive weighting strategy (AWS) for our MSMTDA module, as well as comparing AWS with manual tuning methods. The comparison results (mIoU) are reported in Tables 3.

With respect to the MSMTDA module, we firstly train the model with merely source-target segmentation loss \mathcal{L}_{seg}^s . As shown in Table 3, the AWS outperforms the manual tuning by average 0.8% mIoU, which shows adaptively weighting the multiple source domains is better than applying uniform weights. Then, we add the

Table 2: Influence (mIoU) of Initial Entropy Threshold v_0 of the Pseudo Label Update Strategy on the Adaptation From Paris, Berlin, Tokyo datasets to Zurich, Chicago, Potsdam datasets.

ρ_v	40%	50%	60%	65%	70%	80%	90%	100%
Potsdam	50.6	52.1	52.4	52.8	52.6	52.2	51.7	51.3
Zurich	51.2	53.1	53.7	54.1	53.8	53.4	52.9	52.4
Chicago	48.6	50.0	50.4	50.7	50.6	50.3	49.9	49.6

Table 3: Comparison (mIoU) Between Manual Hyperparameter Tuning and Adaptive Weighting Strategy of the Proposed MSMTDA Module on the Adaptation From Zurich and Chicago datasets to Paris and Berlin datasets. "M" denotes Manual Tuning Method while "A" represents Adaptive Weighting Method.

Loss	Method	λ_{seg}^s	λ_{adv}^{st}	λ_{adv}^{tt}	Paris	Berlin
\mathcal{L}_{seg}^s	M	1	0	0	33.8	33.4
	A		-		34.6	34.1
\mathcal{L}_{adv}^{st}	M	1	0.1	0	44.8	45.1
	M	1	0.05	0	46.2	46.6
	M	1	0.02	0	45.5	46.1
	M	1	0.01	0	45.1	45.3
	A		-		46.7	47.2
\mathcal{L}_{adv}^{tt}	M	1	0.05	0.1	47.1	47.9
	M	1	0.05	0.05	47.8	48.5
	M	1	0.05	0.02	47.9	48.8
	M	1	0.01	0.01	47.5	48.1
	A		-		48.1	49.3

source-target adversarial loss \mathcal{L}_{adv}^{st} to conduct multiple adversarial training between multisource and multitarget domains. For the manual tuning methods, the model reaches the best segmentation performance when λ_{adv}^{st} is set to 0.05. In comparison, AWS improves the segmentation performances by average 0.6% mIoU without extensive tuning time and manual efforts. Finally, when adding the target-target adversarial loss \mathcal{L}_{adv}^{tt} , manual tuning methods perform the best when $\lambda_{adv}^{tt}=0.02$, while is average 0.4% mIoU worse than our AWS. These results confirm the effectiveness and superiority of adaptive weighting strategy (AWS).

1.2.4 Influence of Various Source and Target Domains. To study the effect of different source and target domains on the model performance, we test several UDA scenarios and report the mIoU results in Table 4. From Table 4, we can see using different source domains will obtain different cross-domain segmentation performance on the target domains. The reason may be different source domains have different data distributions, the source domains whose data distributions are more similar to target domains can contribute better adaptation performance. Moreover, the model performances under MSDA setting and under MTDA setting are superior to the model performance under SDA setting, showing that multisource

or multitarget training data can improve the adaptation performance over SDA, since the multisource and multitarget samples can extend the generalization and robustness of feature representation by providing complete information. Furthermore, by fully exploiting the complementary knowledge from multisource and multitarget domains, MSMTDA can reach better adaptation performance over MSDA and MTDA, with 51.5% mIoU on Paris and 53.2% mIoU on Berlin, highlighting the superiority of MSMTDA against MSDA and MTDA. Furthermore, more source and target domains can further improve the model performance. The above results indicate when seeking the optimal domains for UDA, more source and target domains are generally beneficial to improve the adaptation performance when there are sufficient source and target domains. If the labeled source domains are limited, using source domains which are more similar to target domains can achieve better adaptation performance.

Table 4: The mIoU Performance of Our Proposed MultiDAN that Uses Various Source Domains and Target Domains for Adaptation on the AIS Datasets. "/" denotes "Or". "+" denotes "And"

Scenarios	Sources	Targets	Paris Berlin	
w/o DA	Chicago	Berlin/Paris	25.9	27.3
	Zurich	Berlin/Paris	28.2	28.7
	Zurich+Chicago	Berlin/Paris	31.4	30.5
SDA	Chicago	Berlin/Paris	46.2	47.4
	Zurich	Berlin/Paris	47.4	48.8
MSDA	Zurich+Chicago	Berlin/Paris	49.8	50.6
MTDA	Chicago	Berlin+Paris	49.1	49.9
	Zurich	Berlin+Paris	49.6	50.7
MSMTDA	Zurich+Chicago	Berlin+Paris	51.5	53.2
	Zurich+Chicago+Potsdam	Berlin+Paris+Tokyo	52.7	54.9

REFERENCES

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4 (2018), 834–848.

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

[3] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).

[4] Yuxi Wang, Zhaoxiang Zhang, Wangli Hao, and Chunfeng Song. 2020. Attention guided multiple source and target domain adaptation. *IEEE Transactions on Image Processing* 30 (2020), 892–906.