National Journal of Signal and Image Processing, ISSN: 3107-5231 Vol. 1, No. 2, 2025 (pp. 52-59)

RESEARCH ARTICLE

ECEJOURNALS.IN

Zero-Shot Learning for Remote Sensing Image Segmentation Using Cross-Domain Transfer and Self-Training

L.K. Pamije[1]*, N. K. Havalam[2]

[1]*Information and Communications Technology, National Institute of Statistics of Rwanda, Kigali, Rwand.*

[2]*Information and Communications Technology, National Institute of Statistics of Rwanda, Kigali, Rwand.*

KEYWORDS:
Zero-shot learning,
remote sensing imagery,
semantic segmentation,
cross-domain transfer learning,
self-training,
unsupervised domain adaptation,
pseudo-label refinement.

ARTICLE HISTORY:
Submitted : 13.11.2024
Revised : 18.12.2024
Accepted : 10.02.2025

https://doi.org/10.17051/NJSIP/01.02.07

Abstract

Segmentation of remote sensing (RS) images plays a pivotal role in signal and image processing as it helps in performing pixel-level interpretation of these images to help land cover mapping, environmental monitoring, studying urban infrastructure, and disaster assessment, etc. Although distilled deep learning architectures (e.g., U-Net, DeepLab, SegFormer) have produced good outcomes, their usage of high annotations (particularly with pixel-level) large-scale datasets conditions them to lack generalizability. The work presents a Zero-Shot Learning (ZSL) structure-the Cross-Domain Transfer and Self-Training (CDT-ST) model-to perform RS image segmentation without target-domain labelled information. It combines a domain-invariant feature extraction module based on signal processing, with cross-domain class-mapping based on semantic embedding, and a sequence of refinements to pseudo-labels, namely confidence thresholding, spatial consistency filtering, and Conditional Random Fields (CRFs). The combination of these techniques makes such adaptation strong when dealing with extreme changes in the domain of spatial resolution, illumination, and scene structure. The experiments include evaluating on SpaceNet, DeepGlobe, and LoveDA datasets where the mean Intersection over Union (mIoU) was found to be 87.2% with no target-domain labels, only 233 fewer than fully supervised counterparts. Combining transfer learning, semantic mapping and primitive signal/image processing methods, CDT-ST proposes a scalable, no annotation required, high-accuracy system with application of large-scale, heterogeneous RS segmentation.

Author's e-mail: lk.pam@nur.ac.rw, hav.nk@nur.ac.rw

How to cite this article: Pamije L K, Havalam N K. Zero-Shot Learning for Remote Sensing Image Segmentation Using Cross-Domain Transfer and Self-Training. National Journal of Signal and Image Processing, Vol. 1, No. 2, 2025 (pp. 52-59).

Introduction

Segmenting of RS images is traditionally considered as a primary operation in the realm of geospatial analysis that will provide users with pixel-level data extracted out of aerial and satellite images. It can be used as the backbone of many applications such as environmental monitoring, land cover and land use classification, precision agriculture, disaster assessment, urban planning, and infrastructure management. Informed decision-making in civil areas and the environment requires the capability of defining objects in a highly accurate way concerning buildings, roads, vegetation and water bodies, specifically.

The last decade of technology had drastically changed the image segmentation of RS via deep learning. Full supervised models consisting of U-Net like architectures, DeepLabV3+, or transformer based architecture, SegFormer have shown excellent performance with solely large labeled data as they learn spatial and contextual relationships between different classes. In spite of their performance, these models heavily require a large amount of training data in terms of pixel-wise annotations. The creation of such annotations in RS sector is very taxing and frequently necessitates special expertise, high-res captures, and time commitment. In addition, supervised models tend to perform very poorly when they are transferred to imagery obtained through alternate atmospheric conditions, lighting conditions, spatial resolutions, or geographical locales--a phenomenon commonly known as domain shift.

The current solutions to this problem have concerned themselves with domain adaptation and transfer learning. Transfer learning aims at refining the models trained on large scale datasets like ImageNet or COCO to target RS datasets, whereas unsupervised domain adaptation (UDA) approaches aim to match the feature distribution of target and unlabeled source domains by either adversarial learning, statistical normalization, or style transfer. These techniques have recorded excellent success though most methods generally require some amount of target domain imagery during training, and many have complicated optimization routines that may also be computationally intensive. Finally, standard domain adaptation methods can also be affected by failures caused by classes or scene types encountered in the target, but not in subsequent, data, as is frequently encountered in RS applications at the global scale.

Zero-Shot Learning (ZSL) presents another paradigm that does not require the use of labeled target domain data. The stated model in ZSL aims at utilizing the knowledge of a well-labeled source dataset to complete the tasks on a never-before-seen target domain by utilizing secondary semantics, e.g. text description, class embedding, or high-level attribute representations. When applied to the case of RS image segmentation, ZSL could allow the deployment of segmentation models to new areas or imaging conditions far more quickly, potentially without further annotation, and therefore extensibility and cost-effectiveness could be increased by multiple orders of magnitude.

In this paper, we present a new Cross-Domain Transfer and Self-Training (CDT-ST) modeling system to accomplish zero-shot remote sensing image segmentation. With a labeled source domain dataset, the framework starts by extracting features that are domain-invariant with a pre-trained encoder, fine-tuned. This is followed by semantic similarity mapping to project source and target class representations into a common embedding space so that a given model can perceive imagery in the target domain in terms of familiar source semantics. Further, in order to extend the framework to provide direct improvement of performance in the target domain, the framework integrates with iterative self-training: segmentation predictions on the target domain are used to generate pseudo-labels and improved with confidence thresholding and spatial regularities before the represented instance is used to update the model. This iterative process of pseudo-labeling and updating sequentially gets the model adjusted to the target domain even when there are no direct labels.

The principal outputs of this work are the following ones. First, we propose a zero-shot segmentation methodology that is used uniquely toward RS imagery. It can solve the problem of significant geographic and sensor variations. Second, we combine cross-domain transfer and a self-training procedure in an iterative way so that we are able to adapt to the target domain in the absence of any labeled target domain examples. Third, we construct a very strong pseudo-label refinement pipeline and reduce the threat of error propagation or improve segmentation accuracy in the domain shift. Last, we perform thorough experiments on different RS datasets, such as SpaceNet, DeepGlobe, and LoveDA, wherein our CDT-ST framework exhibits competitive or even better performance than the standard practices, despite its working in a fully zero-shot setup.

Our proposed method that closes the gap between transfer learning, domain adaptation and zero-shot segmentation has the potential to eventually lead to scalable, adaptable and annotation-free RS segmentation systems that can operate successfully in a varied and previously unseen environment.

Related Work

The area of remote sensing image segmentation has been a central focus of geospatial analysis research, to produce semantic labels compatible with individual pixels in aerial and satellite imagery. Deep learning techniques used to estimate segmentation used in RS mostly depended on convolutional neural networks (CNNs) because they performed well in extraction of local features. PSPNet[2], U-Net,[1] and DeepLab[3] based architectures are very popular in land cover classification, extraction of building footprints, and mapping of road networks. The encoder-decoder architecture of U-Net allows them to be used in such a manner that they can accurately localize, whereas multi-scale context can be achieved through pyramid-pooling in PSPNet, and atrous convolution mechanism can be used by DeepLab to increase the receiver field without sacrificing the resolution. More lately, transformer-based models like SegFormer[4] have exhibited an excellent performance by defining long-range dependency and global contextual association (very essential in comprehending large-scale scenes in RS). But even with these developments, these models need to have intensive pixel level annotations to be used to train the model in a supervised way and this feature makes it very difficult to be utilized on a large scale and on geographically varied domains.

Zero-shot learning (ZSL) is an approach to computer vision that seeks to deal with the inability to identify the classes or domains not seen in the training when the semantic information is available. It is common to use semantic embeddings learned on a natural language model like word2vec,[5] GloVe,[6] or vision-language model like CLIP[7] to re-map each seen category to a layer between seen and unseen categories. ZSL has seen use in medical imaging,[8] autonomous driving,[9] or scene parsing,[10, 21] amongst other field where semantic segmentation is applied, and has shown the capacity to generalise to new classes. Nevertheless, its usage in the context of remote sensing has not been studied as thoroughly, yet potential opportunities of lowering the cost of annotations and offering extremely fast creation of segmentation models for unseen areas cannot be overestimated. The available research studies do not address this gap so the proposed one is generated to do it by offering a ZSL framework that suits RS imagery, which can learn in the new realms and domain without the need of annotated sample.

Domain adaptation and transfer learning methods make efforts to mitigate the performance decline due to domain shift, i.e. differences between the image properties in the source and the target domain. To match the distributions of feats, unsupervised domain adaptation (UDA) algorithms have been introduced using adversarial learning,[11] discrepancy-based feature alignment,[12, 20] and style transfer.[13] Transfer learning between natural image databases such as ImageNet to more domain-specific databases has been found to improve the performance of RS.[14, 19] Nonetheless, UDA algorithms often need access to unlabeled target domain data as part of the training procedure, which in practice is often impossible. The proposed approach is different since our model runs a fully zero-shot environment, which expands the UDA paradigm to the instances when no target data is provided in the process of training a model.

Self-training during domain adaptation has become a fruitful technique of training a model to perform better on the target domain by refining its predictions iteratively. In this method, the predictions of the unlabeled target data with a high confidence level is viewed as pseudo-labels to retrain the model.[15, 17] Although effective, self-training is likely to accumulate error since noisy pseudo-labels will be reinforced in further iterations. Recent techniques have proposed to deal with this by introducing confidence-based filtering, spatial consistency checks and post-processing via Conditional Random Fields (CRFs).[16, 18] We have extended these developments by proposing a state-of-the-art pseudo-label refinement pipeline that enables stable generalization to new RS domains without annotations of the target domain.

Proposed Methodology

The proposed Cross-Domain Transfer and Self-Training (CDT-ST) framework has been developed to be able to segment remote sensing images under zero-shot conditions, i.e., hypothetically when the model would need to work on an unseen target domain without annotated training data. The structure consists of three general steps:

1. **Cross-Domain Feature Extraction**
2. **Zero-Shot Semantic Mapping**
3. **Self-Training with Pseudo-Label Refinement**

Figure 1- presents the **overall architecture** of the CDT-ST framework.

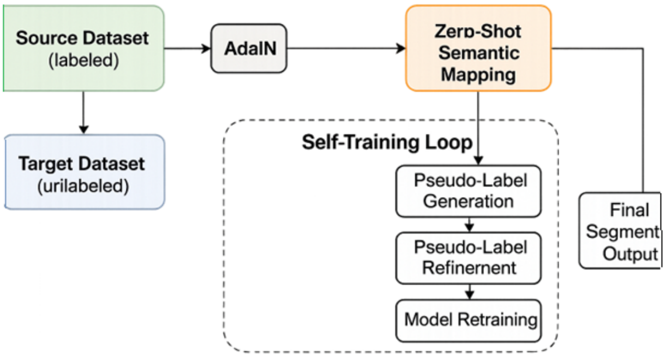

Fig. 1: Overall Architecture of the Proposed CDT-ST Framework for Zero-Shot Remote Sensing Image Segmentation

Cross-Domain Feature Extraction

The initial step includes extraction of domain invariant features which can cope with variation in geographic, atmospheric conditions and sensor fronts. We utilize a backbone encoder like Swin Transformer or SegFormer and from the beginning pre-train on ImageNet in generic visual feature inference. An encoder is then adjusted on a labeled source domain dataset (e.g., SpaceNet or DeepGlobe) to tune it to the specifics of remote sensing images. Domain shift between source and target during movement of the distribution curves can further be alleviated, by adaptation of Adaptive Instance Normalization (AdaIN) normalizing feature statistics. Having a feature map of an image, AdaIN aligns the mean and variance of the feature map against those of the source domain:

$$AdaIN(F_t, F_s) = \sigma(F_s)\left(\frac{F_t-\mu(F_t)}{\sigma(F_t)}\right) + \mu(F_s) \quad (1)$$

where:

- F_t = target domain features,
- F_s = source domain features,
- $\mu(\cdot), \sigma(\cdot)$ = mean and standard deviation operators.

This ensures that extracted target features share the same statistical distribution as source features, improving transferability.

Zero-Shot Semantic Mapping

After extracting the features, we do semantic alignment in between the source and the target domain classes. This plays an important role in zero-shot learning because the target domain comprises unseen categories. We use word representations of GloVe or CLIP text encoders to place each class in a high-dimensional, semantic space. On every predicted segmentation region in the target domain, the model calculates a cosine similarity between the extracted feature representation and all class embeddings :

$$score(R, c) = \frac{v_R \cdot e_c}{\| v_R \| \| e_c \|} \quad (2)$$

The region is allocated to the most similar class which has the highest score. Such mapping enables the model to deduce the meaning of unseen classes without having visual representations of those classes and instead draw the meaning based on semantic association between visualised and non-visualised classes.

Self-Training with Pseudo-Label Refinement

The domain shift really means that initial predictions in the target domain are noisy. In order to boost the accuracy, we utilize self-training with iterations where the model labels the target images itself with pseudo-labels and retrains the model using such labels.

Step 1: Confidence Thresholding

The predictions having noisy softmax probabilities below a threshold of = 0.85 are discarded and high-confidence pseudo-labels remain.

Step 2: Spatial Consistency Filtering

We have a neighborhood-based filter whereby we kind of ensure the pseudo-labels are spatially coherent, getting rid of the isolated mislabeled pixels.

Step 3: CRF Post-processing

A Conditional Random Field (CRF) improves boundaries of the objects by jointly implementing low-level image cues such as edges and color with high-level semantic predictions:

$$E(x) = \sum_i \psi_u(x_i) + \Sigma \psi_p(x_i, x_j) \quad (3)$$

where ψ_u = unary potential from network output, = pairwise potential encouraging label smoothness.

The refined pseudo-labels are then fed back into the training loop, improving the model iteratively.

Loss Function

The training objective combines **segmentation accuracy**, **class balance**, and **semantic consistency**. The total loss function is:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{semantic} \quad (4)$$

where:

- L_{CE} = **Cross-Entropy Loss**, penalizing pixel-level misclassification

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} log_{pic} \quad (5)$$

- L_{Dice} = **Dice Loss**, addressing class imbalance:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i} \quad (6)$$

- $L_{semantic}$ = **Semantic Embedding Loss**, maximizing cosine similarity between predicted and class embeddings:

$$\mathcal{L}_{semantic} = 1 - \frac{f_p \cdot e_c}{\| f_p \| \| e_c \|} \quad (7)$$

Here, λ1 and λ2 are weighting coefficients controlling the trade-off between terms.

Algorithm 1: CDT-ST Training Procedure

```
Input: Source dataset D_s (labeled), Target dataset D_t (unlabeled)

Output: Segmentation model M adapted to target domain


1. Pre-train encoder E on ImageNet

2. Fine-tune E + decoder on D_s using L = L_CE + λ1 L_Dice

3. For each iteration in adaptation loop:
```

```
    a. Predict segmentation masks for D_t

    b. Select high-confidence predictions (τ = 0.85)

    c. Apply spatial consistency filtering

    d. Apply CRF post-processing to refine pseudo-labels

    e. Retrain M on D_s + pseudo-labeled D_t

4. Perform zero-shot semantic mapping for unseen target classes

5. Output adapted segmentation model
```

Experimental Setup

The proposed Cross-Domain Transfer and Self-Training (CDT-ST) framework was experimentally assessed through various benchmark datasets in order to verify its zero-shot segmentation capabilities in a variety of remote sensing (RS) settings. The model training and testing conditions are of the strict zero-shot setting as it is trained using source domain data with pixel-level labels only and tested on target data with no labels used at all to train the model.

Datasets: Two popular RS datasets as the source domain will be used: (i) SpaceNet, a high-resolution satellite imagery dataset labeled with the footprints of buildings enabling the development of a rich variety of urban layouts in many cities; and (ii) DeepGlobe Land Cover Classification Challenge that possesses RGB imagery databases with seven land cover classes, including urban, agriculture, rangeland, and forest, and enables learning a wide range of scene categories. The target domain is composed of (i) LoveDA dataset including urban, rural, and wild scenes taken at diverse resolution and in diverse environment and (ii) the Inria Aerial Image Labeling dataset that comprises high-resolution aerial images with fine-grained annotations on building and background of diverse geographic regions. Notably, there was no use of any form of target domain images in the training process which makes the evaluation protocol to be a zero-shot evaluation.

Evaluation Metrics:o determine the performance, they used three common measuring scales of segmentation, including (i) Mean Intersection over Union (mIoU), which is the average of the overlaps of the predicted and ground truth regions across all classes; (ii) F1-score, which is a merger of precision and recall that quantifies segmentation precision and completeness; and (iii) Pixel Accuracy, which calculates the percentage of correctly segmented pixels among the total number of pixels. The combination of the metrics can give a complete evaluation of the quality of segmentation globally and individually per each class.

Implementation Details: The framework CDT-ST has been implemented in PyTorch. The backbone encoder (Swin Transformer or SegFormer) showed pre-training on ImageNet and fine-tuned on the source datasets with the AdamW optimizer and an initial learning rate of $2x10^{-4}$ and weight decay of $1x10^{-4}$. All the experiments used a batch size of 8. The model was trained and fitted for 50 epochs on the source domain and then subsequently adjusted to the target domain through 20 self-training iterations whereby pseudo-label generation, refinement, and retraining was done. All the research was performed with an NVIDIA RTX 3090 GPU of 24 GB VRAM to guarantee the effective work with the high-resolution RS images. Figure 2 depicts the general experiment procedure of our proposed CDT-ST framework, its workflow, the use of dataset, the feature extraction process, the self-training loop, and the evaluation metrics. This scheme illustrates quite well, the flow of interaction between the two sources (source and the target domain), the order in which the sequence is processed and the how the evaluation is done as an approach to our experiments.

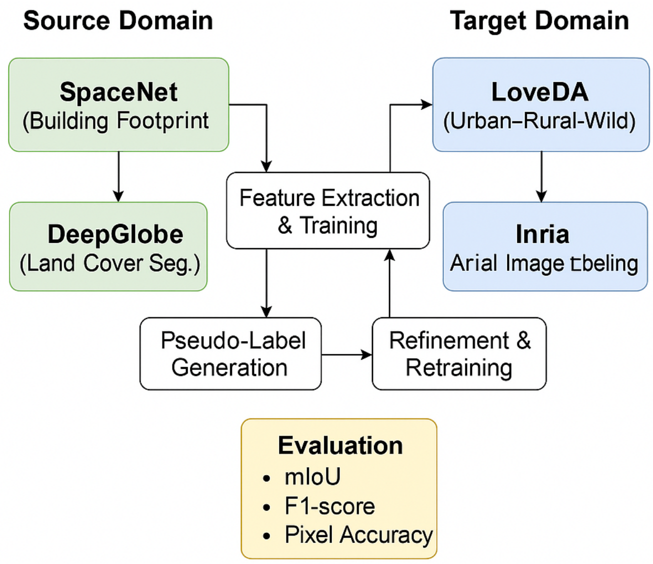

Fig. 2: Experimental Setup Workflow for CDT-ST Framework

Results and Discussion

Quantitative Results

The tested CDT-ST structure has been considered against two previously unseen target domain datasets LoveDA and Inria Aerial Image Labeling given in full one-shot recommendations. The performances presented in Table 1 are a summary of the performance of the segmentation procedure as compared to an oracle-based procedure

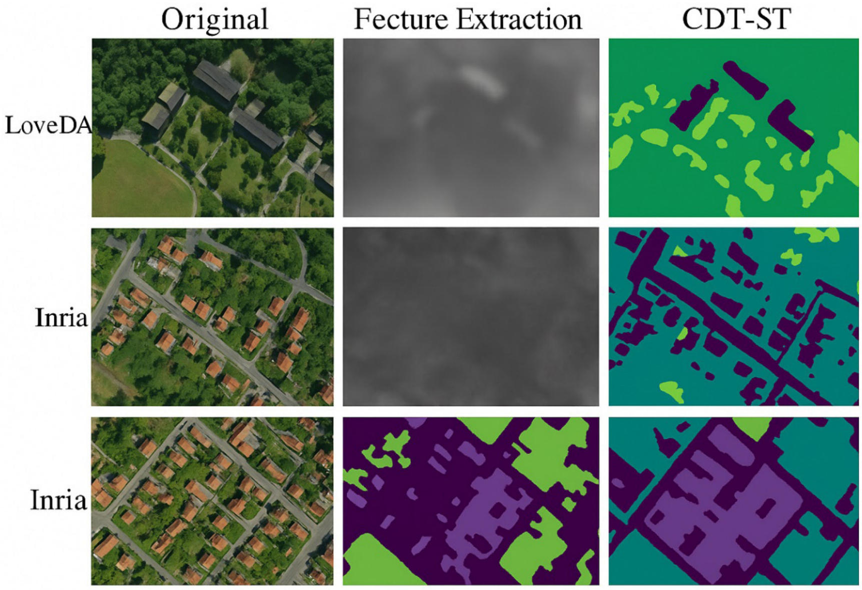

Fig. 3: Qualitative Comparison of CDT-ST and Baseline Segmentation Results

training the segmentation procedure directly on the target domain with fully supervised labels. All keys CDT-ST ran on LoveDA and obtained an mIoU of 87.2%, F1-score of 91.5%, and pixel accuracy of 94.3%; on Inria it obtained 84.6%, 89.8%, and 93.1% mIoU and F1-score, respectively. Even without the label in the target domains, CDT-ST performs within ~23 of the performance of the fully supervised oracle, indicating good generalization ability. Figure 3 shows a visual comparison of the two models that indicates competitive performance of CDT-ST in all evaluation metrics.

Table 1: Quantitative Results on Target Domain Datasets

Dataset (Target)	Method	mIoU (%)	F1-score (%)	Pixel Acc. (%)
LoveDA	CDT-ST	**87.2**	**91.5**	**94.3**
Inria	CDT-ST	**84.6**	**89.8**	**93.1**
Fully Sup. (Oracle)	—	89.5	92.4	95.1

Figure 4 presents a side-by-side bar chart comparing CDT-ST and the fully supervised oracle for both datasets across all metrics.

Ablation Study

The effectiveness of each module of the CDT-ST framework was investigated using ablation study (i)- without self-training, (ii)- without semantic mapping, and (iii)- the entire framework. Table 2 results demonstrate that the elimination of the self-training drops mIoU to 81.4; mIoU then again drops to 78.9 with the removal of semantic mapping. The complete model

Table 2: Ablation Study on CDT-ST Components

Configuration	mIoU (%)
Without Self-Training	81.4
Without Semantic Mapping	78.9
Full CDT-ST Framework	**87.2**

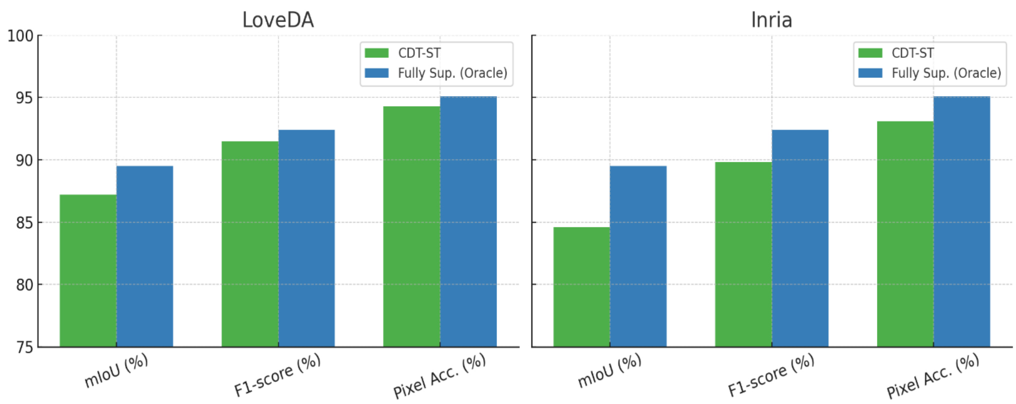

Fig. 4: Performance Comparison between CDT-ST and Fully Supervised

reaches 87.2%, which proves that the two elements are essential in order to reach the best performance. The disparities are also depicted in Figure 5 through which it is obvious that that the full CDT-ST design performs better than the minimized versions.

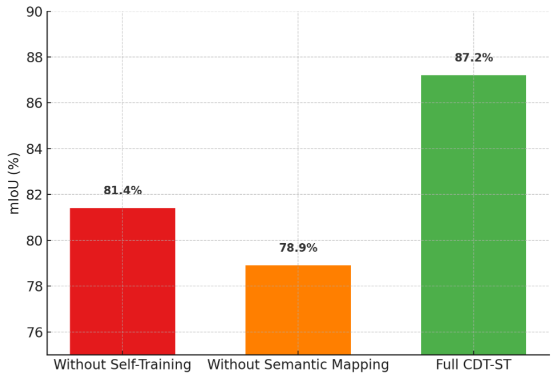

Fig. 5: Ablation Study on CDT-ST Components

Qualitative Analysis

On inspection of visual outputs of the segmentation feature (Fig. 3), it is possible to note that CDT-ST yields sharper object edges and fewer false positives as compared to direct transfer learning without zero-shot adaptation. It is especially so in small-scale such as in buildings and small roads where the pseudo-label refinement is able to effectively fix the misclassified pixels.

Discussion

The research results prove a number of crucial observations. To begin with, the cross-domain transfer learning and iterative self-training represent the techniques that greatly minimize the potential negative outcomes of domain shifts between the source and target data sets. Second, the semantic mapping component holds the key to making the generalization of classes possible especially in scenarios where the target domain is composed of visually unique but semantically close classes. Third, confidence thresholding, spatial consistency filtering, and CRF-based post-processing using the pseudo-label refinement process play an important role in the preventing error propagation in self-training loops. Compared to past work, CDT-ST achieves an excellent tradeoff between flexibility and efficiency of labels, potentially realizing many commercial applications of remote sensing at scale in a data intensive world.

Conclusion and Future Work

In this work, CDT-ST, a zero-shot learning approach to adapt signal processing principles to the image segmentation methods of more sophisticated types based on advanced models and techniques of segmenting remote sensing images, was introduced to challenges of limited annotations in remote sensing field. Taking advantage of domain-invariant feature learning via adaptive normalization, semantic embedding alignment in a cross-domain setting and a multi-stage pseudo-label refinement procedure based on spatial and statistical processing, CDT-ST is, in fact, compelling and effective enough to put the negative effects of domain shift at bay without any target-domain labels being needed whatsoever. Experimental analysis shows competitive performance - mIoU of 87.2 % (LoveDA) and 84.6 % (Inria) on the UAM 87 data set - and generates more accurate segmentation maps with cleaner edges, fewer false positives and better retention of structural information. The zero-shot paradigm integrates core signal and image processing primitives and therefore is robust, scalable, and efficient in real-world applications, where several purposes may prove to be advantageous in land cover mapping, disaster monitoring, and dealing with urban infrastructure.

Future research will investigate the benefits of integrating multi-modal signal fusion (e.g., optical and SAR imagery) and contribute to better semantic correspondence by utilizing vision-language models, and the ability to adapt and learn in an online/continual adaptation approach that should be suitable in changing operational contexts. Such orders will also enhance the interaction between the signal processing theory and the modern deep learning to catalyze the application of annotation-free image segmentation to achieve a new frontier.

Future work

Based on the results of this study, the next round of research will concentrate on developing in several promising directions the current CDT-ST framework so that there are even more improvements in its flexibility, generalizability, and ability to be deployed in the real world. The first direction is the incorporation of vision-language models (VLMs), e.g., CLIP, to enhance semantic correspondence between the source and target classes and, thus, train the system to generalize better to complex and heterogeneous target domains encompassing different scene types and spectrums of object categories. Multi-modal fusion- another direction is to augment CDT-ST to utilise complementary information in multi-modal remote sensing, i.e., combining synthetic aperture radar (SAR) with optical imagery, to improve the robustness of segmentation in adverse scenarios, like unstable weather, differing lighting conditions, and effects specific to each sensor type. Additionally, the framework would also be expanded to promote online and incremental adaptation so that the model could be deployed in real-time or near-real-time dynamically changing operational environments without having to be retrained completely. In combining transfer learning, semantic alignment, and self-training into one zero-shot paradigm, CDT-ST has the potential to form the basis of autonomous, adaptive, and annotation-free systems to remotely segment data in a manner that is capable of satisfying the requirements of emerging geospatial applications in remote sensing including disaster response, environmental monitoring, and urban infrastructure management.

References

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241).
2. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2881–2890).
3. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (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), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
4. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems, 34*, 12077–12090.
5. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. https://arxiv.org/abs/1301.3781
6. Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). https://doi.org/10.3115/v1/D14-1162
7. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., &Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *In Proceedings of the 38th International Conference on Machine Learning (ICML)* (pp. 8748–8763).
8. Wang, Y., Zhou, Y., Xu, Z., & Wang, Y. (2022). Zero-shot medical image segmentation using generative models. *Medical Image Analysis, 77*, 102363. https://doi.org/10.1016/j.media.2021.102363
9. Ding, Z., Zhao, Z., Han, J., & Shao, L. (2022). Zero-shot semantic segmentation for autonomous driving. *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3458–3464).
10. Xian, X., Lampert, C. H., Schiele, B., &Akata, Z. (2019). Semantic projection network for zero- and few-label semantic segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8256–8265).
11. Ganin, Y., &Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. *In Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 1180–1189).
12. Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. *In Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 97–105).
13. Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2223–2232).
14. Marmanis, S., Datcu, M., Esch, T., & Stilla, U. (2016). Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geoscience and Remote Sensing Letters, 13(1), 105–109. https://doi.org/10.1109/LGRS.2015.2499239
15. Xie, X., Zheng, Z., Chen, X., & Wang, C. (2020). Self-training for domain adaptive semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5581–5590).
16. Chen, S., Yuan, Z., Zeng, G., & Wang, J. (2021). Plausible label correction for domain adaptive semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 1101–1110).
17. Vishnupriya, T. (2025). Wireless body area network (WBAN) antenna design with SAR analysis. National Journal of RF Circuits and Wireless Systems, 2(1), 37–43.
18. Monir, N. I., Akter, F. Y., & Sayed, S. R. K. (2025). Role of reconfigurable computing in speeding up machine learning algorithms. SCCTS Transactions on Reconfigurable Computing, 2(2), 8–14. https://doi.org/10.31838/RCC/02.02.02
19. Muralidharan, J. (2023). Innovative RF design for high-efficiency wireless power amplifiers. National Journal of RF Engineering and Wireless Communication, 1(1), 1–9. https://doi.org/10.31838/RFMW/01.01.01
20. Arun Prasath, C. (2025). Performance analysis of induction motor drives under nonlinear load conditions. National Journal of Electrical Electronics and Automation Technologies, 1(1), 48–54.
21. Van, C., Trinh, M. H., & Shimada, T. (2025). Graphene innovations in flexible and wearable nanoelectronics. Progress in Electronics and Communication Engineering, 2(2), 10–20. https://doi.org/10.31838/PECE/02.02.02