# LIDAR-TO-SAR POINT CLOUD SEGMENTATION VIA UNSUPERVISED DOMAIN ADAPTATION NETWORK

*Muhan Wang*[1,3], *Xiaolan Qiu*[1,2], *Silin Gao*[3], *Zhe Zhang*[2],

[1]China National Key Laboratory of Microwave Imaging Technology,
Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China.
[2]Suzhou Key Laboratory of Microwave Imaging, Processing and Application Technology,
Suzhou Aerospace Information Research Institute, Suzhou 215123, China.
[3]School of Electronic, Electrical and Communication Engineering,
University of Chinese Academy of Sciences, Beijing 100094, China

IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641953

## ABSTRACT

Synthetic Aperture Radar (SAR) 3D point cloud reconstruction improves target identification by mitigating issues like overlay masking and shadows found in 2D SAR image projections. Nevertheless, challenges arise from limited SAR data availability and complexities in deciphering and labeling, hindering research in SAR 3D reconstruction point cloud semantic segmentation. To address these hurdles, we propose an alternative training approach—shifting from LiDAR to SAR point clouds. Leveraging benchmark datasets in the LiDAR domain, we advocate using LiDAR point clouds for training to counter the scarcity of SAR training sets. However, applying a segmentation model across different domains leads to a performance decline, particularly in cross-modal SAR-reconstructed point clouds, attributed to distinct roughness introduced by outliers. This paper introduces a 3D semantic segmentation framework based on unsupervised domain adaptation (UDA) for cross-modal learning from LiDAR to SAR. Additionally, we present a simple yet effective geometric transformation data augmentation technique to handle highly imbalanced data distribution. Experimental results confirm the feasibility and effectiveness of our proposed method for SAR 3D reconstructed point cloud semantic segmentation.

***Index Terms*—** Synthetic Aperture Radar (SAR) 3D point clouds, semantic segmentation, unsupervised domain adaptation, cross-modal learning, data augmentation

## 1. INTRODUCTION

Synthetic Aperture Radar (SAR) [1] imaging is pivotal in remote sensing, offering unique advantages over optical imagery. However, traditional 2D SAR images lack elevation direction resolving ability, leading to issues like layover and shadowing. Recently, three-dimensional(3D) SAR [2, 3] has demonstrated significant importance in directly reconstructing the electromagnetic scattering of intrinsic target structures, overcoming 2D limitations. The 3D SAR technique holds great potential for topographic mapping, urban inspection, and forestry detection, making the interpretation of reconstructed SAR 3D point clouds essential for a comprehensive view.

Recent advancements in deep neural networks (DNNs) have propelled progress in point cloud understanding, particularly in tasks like semantic segmentation [4] and object detection [5]. Semantic segmentation in 3D point clouds currently plays a crucial role in applications such as autonomous driving, urban planning, and disaster response. However, the application of deep learning to SAR three-dimensional point cloud segmentation faces challenges. The high cost of SAR systems and limited data availability hinder the creation of extensive training datasets, resulting in an insufficient number of training samples. Reflective mechanisms in SAR imaging introduce secondary scattering, significantly increasing the complexity of interpretation. Additionally, 3D point clouds obtained from SAR data reconstruction suffer from substantial noise, further complicating interpretation and annotation processes.

This work focuses on effective and efficient semantic segmentation for better learning from limited SAR 3D point cloud data. And thanks to the advancements in unsupervised domain adaptation (UDA) [6] technology, it has become feasible to train models on a source domain with abundant training samples and subsequently apply them to a target domain with limited or missing labels. To reduce the domain gap, various UDA approaches have been proposed [7, 8]. Therefore, we propose that training can be initially conducted on a LiDAR source domain with a certain amount of annotated samples. Subsequently, leveraging UDA technology, the trained model can be transferred to the target domain of SAR

Thanks to NSFC grant 61991421, 61991420 and AIRCAS grant "Structural sparsity signal high performance adaptive sensing theory and its applications in microwave imaging" for funding.

three-dimensional point clouds, thereby achieving semantic segmentation of reconstructed SAR 3D point clouds.

It is important to note that there are significant differences between LiDAR point clouds and SAR point clouds in practical data scenarios. Firstly, the acquisition methods for these two types of data differ, leading to variations in point cloud density and scale. Secondly, LiDAR point clouds exhibit relatively fewer noise elements, whereas SAR point clouds have a lower signal-to-noise ratio, resulting in an overall coarser appearance compared to LiDAR point clouds. Lastly, the composition of objects in scenes varies across different countries and regions (urban versus rural), exacerbating the gap between different datasets.

To address these challenges, we have employed methods such as data alignment, global data augmentation, and local data augmentation to mitigate the gap between the datasets. The contribution of this work can be summarized in three major aspects: Firstly, we have achieved, for the first time, semantic segmentation of reconstructed SAR 3D point clouds using LiDAR point cloud as training dataset. We propose a semantic segmentation framework for urban scene SAR 3D point clouds based on UDA technique. Additionally, We propose a data augmentation method specifically designed for the target domain consisting of SAR point cloud data, which is used to reduce the discrepancy between LiDAR rural scenes and SAR urban scenes. Moreover, We constructed a SAR 3D point cloud based on the MV3DSAR dataset and verified the effectiveness of the proposed method for semantic segmentation of SAR point cloud. In addition, the reference labels of the corresponding SAR 3D point cloud scenes are provided, which lays a solid foundation for future research on the highly significant problem of SAR 3D point cloud semantic segmentation.

## 2. METHODOLOGY

Fig. 1 illustrates the architecture of our proposed UDA-based 3D SAR point cloud segmentation framework for bridging the domain gap. We first acquire SAR three-dimensional point clouds using sparse reconstruction technique. For a given LiDAR training set and the reconstructed SAR three-dimensional point clouds, we align and normalize the point cloud density and coordinates, ensuring uniformity across different domain. The proposed framework focuses on eliminating the domain gap between different scenarios through the proposed class balancing strategy of local geometric enhancement. By leveraging this strategy, we ensure that point cloud features from the two domains have a similar distribution, thereby enabling the segmentation of SAR three-dimensional reconstructed point clouds.

### 2.1. Reconstruction to Obtain SAR 3D Point Clouds

3D SAR technique uses a stack of 2D SAR images collected from different cross-track angles of the same observation area to reconstruct the scattering information along the elevation direction and achieve the three-dimensional resolving ability.

For a single look complex (SLC) SAR image, $y(x_0, r_0)$ represents the value of an azimuth-range pixel $(x_0, r_0)$. Consider the same target is observed $N$ times from slightly different viewing angles to obtain $N$ SLCs (via an antenna array or multiple flight passes). If the SLCs are perfectly aligned, denote the $n$-th acquisition as $y_n$. Each azimuth-range pixel $(x_0, r_0)$ can be expressed as the integration of the scattering coefficient along the elevation. With the additive noise, it can be written in a matrix-vector form as (1).

$$\mathbf{y} = \mathbf{R}\gamma + \varepsilon \tag{1}$$

where $y$ is the measurement vector stacked by $y_n$, $\mathbf{R}$ plays as an $N \times L$ mapping matrix with

$$R_{nl} = \exp(j4\pi b_n s_l / \lambda r_0) \tag{2}$$

Now, 3D SAR inversion boils down to a signal recovery problem, where the goal is to obtain the corresponding scattering parameters such as elevation and reflectivity profile $\gamma$ of each range-azimuth cell by solving (1), where $\mathbf{R}$ is usually known from model and $y$ is the observation. Hence, the echo signal along the elevation direction is sufficiently sparse and can be solved via a sparse reconstruction problem (1) using the compressed sensing (CS) technique. The SAR 3D reconstructed point cloud used in this paper is also reconstructed from SAR 2D images, based on the orthogonal matching pursuit algorithm [9].

### 2.2. Modality Gap and Proposed UDA Framework

#### *2.2.1. Data preprocessing of different modes*

From the comparison of different scenes in Fig. 2, it is evident that the reconstructed density of SAR three-dimensional point clouds is significantly greater than that of the LiDAR point cloud training set. Additionally, the sizes and scales of the two scenes vary, posing challenges to the learning of point cloud features and domain adaptation, potentially causing confusion in recognition results. Therefore, it is imperative to perform data alignment and coordinate normalization on point cloud data from different modalities.

#### *2.2.2. Data augmentation for cross-scene class balancing*

The LiDAR point cloud training set includes rural scenes from Vahingen, while SAR three-dimensional reconstruction point clouds originate from urban residential areas. Notable distinctions in scenes include the prevalence of low, pointed-roof houses in rural settings and a mix of high-rise apartments

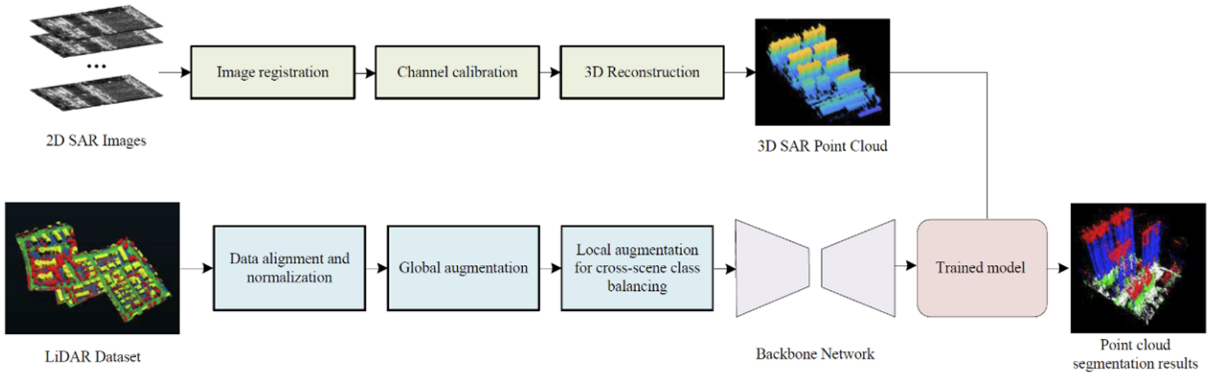
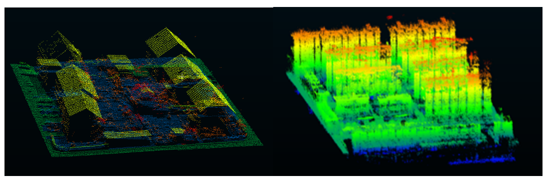



**Fig. 1**: Overview of proposed 3D SAR semantic segmentation network.



(a) (b)

**Fig. 2**: Comparison between LiDAR point clouds and SAR three-dimensional reconstructed point clouds reveals notable differences.

and single-story houses in urban areas. Rural scenes also feature more trees, surpassing the height of houses, whereas urban neighborhoods have fewer trees and lower vegetation.

Identifying these distinctions, we propose a UDA framework to bridge the domain gap for precise segmentation of SAR three-dimensional reconstruction point clouds.

Considering limited training samples and to prevent overfitting, essential global data augmentation techniques, such as rotation and scaling, are employed. Due to substantial noise in SAR point clouds, noise introduction becomes necessary for training. Additionally, Mix3D [10] augmentation further enhances generalization.

Beyond global augmentations, this study incorporates local enhancements based on distinct target scene characteristics. Geometric scale reduction transformations are applied to the "Tree" category in the LiDAR training set, addressing scene disparities. This involves randomly discarding 70% of points and merging the scaled-down tree category with shrubs and low vegetation, collectively categorized as "Vegetation." This aims to mitigate differences in vegetation quantity and height disparity between vegetation and buildings in urban and rural scenes.

### *2.2.3. Backbone network*

Our selection of MinkowskiNet [11] as the backbone of our network is driven by its exceptional proficiency in efficiently handling sparse and high-dimensional data. MinkowskiNet capitalizes on the Minkowski Engine, an open-source library specifically crafted for streamlined operations on sparse tensors, making it an optimal choice for tasks involving point clouds, 3D scenes, and other forms of unstructured data. Additionally, MinkowskiNet stands out as a current state-of-the-art (SOTA) point cloud segmentation model, further underscoring its prominence in the field. Fig .3 illustrates the fundamental architecture of MinkowskiNet.

### *2.2.4. Optimization*

Due to the uneven distribution of ground objects in wild, LiDAR point cloud datasets [12, 13] often face a serious data-imbalanced problem, which negatively leads the model to focus on classes with more samples. Take Vaihingen3D dataset [14] as example, classes including tree, road, roof and so on take majority of data while classes like facad have very limited sample points, which can be seen in Fig. 2. Hence, we replace the cross-entropy loss function commonly used in point cloud segmentation with a weighted cross-entropy objective to focus more on classes with less samples, in which the weight of loss for each class is inverse to its frequency $f$ on training set:

$$\mathcal{L}_{wcent} = -\sum_{j=1}^{C} w_j p(y_j) \log\left(p(\hat{y}_j)\right), w_j = {f_j}^{-1} \quad (3)$$

where $y_j$ and $\hat{y}_j$ are ground truth label and predicted label of class $j$. In this case, classes with small number of samples have larger losses and imbalanced data distribution problem could be alleviated.

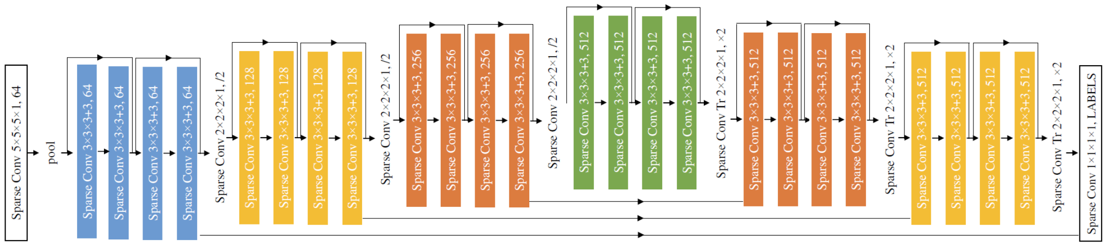


**Fig. 3**: Overview architecture of MinkowskiNet [11].

**Table 1**: Segmentation results on the SARMV3D dataset

| Methods | Vegetation | Road | Roof | Facade | mIoU |
|---|---|---|---|---|---|
| MinkNet | **6.46** | 13.40 | 15.86 | 15.66 | 12.85 |
| + Global Augmentation | 0.83 | **60.61** | 22.60 | 16.85 | 25.22 |
| + Weighted Cross Entropy Loss | 2.76 | 56.27 | 51.10 | 64.22 | 43.59 |
| + Proposed Local Augmentation | 3.79 | 59.88 | **53.11** | **68.57** | **46.34** |

## 3. EXPERIMENTS

### 3.1. Dataset

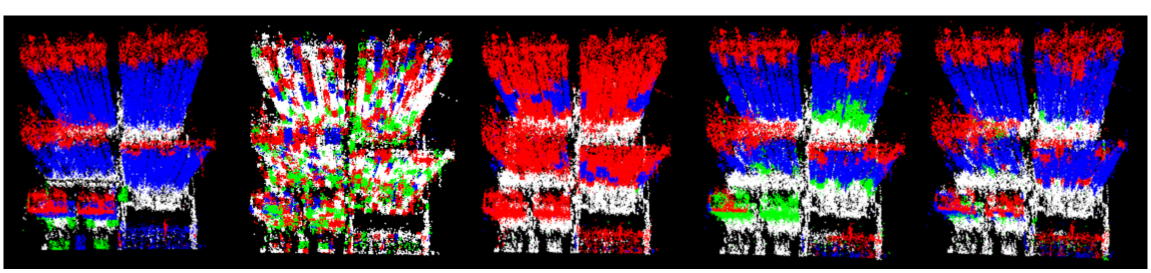
The training set on the source domain is adopted from ISPRS 3D Building Reconstruction Test Data on Vaihingen and the SAR 3D reconstructed point cloud on the target domain is from the MV3DSAR dataset [15]. We compared the ablation experiments with the addition of different modules and performed a visualization, as shown in Fig. 4. In addition to this, for further quantitative analysis, we labeled the SAR 3D reconstructed point cloud with reference semantics based on the optical map. Performances are evaluated via per-class IoU and mean IoU (mIoU).

### 3.2. Experiment Results

From the results shown in Fig. 4(b), it can be seen that training MinkNet with only LiDAR point clouds without any processing makes the results overfitted and unable to recognize SAR point clouds. Due to the severe class imbalance in the training set, simply performing global enhancement without class balancing processing leads to the result of indistinguishable roofs and buildings in Fig. 4(c). Fig. 4(d) shows the difference in target classes between scenes, with the low building being recognized as vegetation. And after the proposed geometric transformation, local augmentation of the class of vegetation can improve this and achieve more accurate semantic segmentation, as shown in Fig. 4(e). Experimental results verify the effectiveness and feasibility of the proposed UDA framework for SAR point cloud semantic segmentation.

## 4. CONCLUSION

We introduces a 3D semantic segmentation framework based on UDA for cross-modal learning from LiDAR to SAR. Our scheme aims to reduce the domain gap between cross-modal point clouds through data augmentation. Experimental results affirm the feasibility and effectiveness of our proposed method for the segmentation of SAR 3D reconstructed point clouds. Our framework demonstrates not only theoretical innovation but also significant potential in practical applications, laying a foundation for future research and applications on SAR 3D point cloud semantic segmentation.



(a) (b) (c) (d) (e)

**Fig. 4**: Semantic segmentation results of SAR 3D reconstructed point cloud with different methods. (a) Ground truth. (b) Training MinkNet directly on LiDAR point cloud data without any data augmentation. (c) Training MinkNet after global data augmentation only (d) Weighted cross-entropy loss function as class balancing strategy after global data augmentation. (e) The proposed UDA framework

## 5. REFERENCES


[1] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[2] Andreas Reigber and Alberto Moreira, "First demonstration of airborne sar tomography using multibaseline l-band data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2142–2152, 2000.

[3] Xiao Xiang Zhu and Richard Bamler, "Very high resolution spaceborne sar tomography in urban environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, 2010.

[4] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, 2022.

[5] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[6] Yaroslav Ganin and Victor Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[7] Cristiano Saltori, Fabio Galasso, Giuseppe Fiameni, Nicu Sebe, Elisa Ricci, and Fabio Poiesi, "Cosmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 586–602.

[8] Longkun Zou, Hui Tang, Ke Chen, and Kui Jia, "Geometry-aware self-training for unsupervised domain adaptation on object point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6403–6412.

[9] Joel A. Tropp and Anna C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[10] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann, "Mix3d: Out-of-context data augmentation for 3d scenes," in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 116–125.

[11] Christopher Choy, JunYoung Gwak, and Silvio Savarese, "4d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3075–3084.

[12] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9297–9307.

[13] Nina Varney, Vijayan K Asari, and Quinn Graehling, "Dales: A large-scale aerial lidar data set for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 186–187.

[14] Joachim Niemeyer, Franz Rottensteiner, and Uwe Soergel, "Contextual classification of lidar data and building object detection in urban areas," *ISPRS journal of photogrammetry and remote sensing*, vol. 87, pp. 152–165, 2014.

[15] Xiaolan Qiu, Zekun Jiao, Zhenli Yang, Yao Cheng, Bei Lin, Yitong Luo, Wei Wang, Yongwei Dong, Liangjiang Zhou, and Chibiao Ding, "Key technology and preliminary progress of microwave vision 3d sar experimental system," *Journal of Radars*, vol. 11, no. 1, pp. 1–19, 2022.