

Development of a Lightweight SAR Semantic Segmentation Framework with ASFEM for Monitoring Pastoral Ecosystems

Narimane Benouakta¹, Zohra Slimane², Belkacem Benadda³

¹University of Ain Temouchent, Faculty of Technology, Department Electronics and Telecommunications, SSL Laboratory, Ain Temouchent, Algeria.

²Abou BekrBelkaid University, Faculty of Technology, Department of Telecommunication, STIC Laboratory, Tlemcen, Algeria.

³Abou BekrBelkaid University, Faculty of Technology, Departement of Electronic, Tlemcen, Algeria.

Corresponding Author : Narimane BENOOUAKTA

Abstract

Synthetic Aperture Radar (SAR) provides reliable land-cover observations under varying illumination and weather conditions, but semantic segmentation remains challenging because of speckle noise and heterogeneous backscatter. This study adapts the lightweight PP-LiteSeg architecture to Sentinel-1 SAR semantic segmentation and introduces an Adaptive Spatial Feature Enhancement Module (ASFEM) to improve spatial feature representation. ASFEM combines convolutional processing, Coordinate Attention, depthwise convolution, and residual learning with limited additional computational cost. Experiments on the TurkeySAR dataset show that the adapted PP-LiteSeg baseline achieves 69.45% mIoU, while PP-LiteSeg-ASFEM improves mIoU to 70.70%, with an overall accuracy of 89.67%. The proposed module increases the parameter count by only 0.17% and the computational cost by approximately 0.06%, while maintaining an inference speed of 187.81 FPS. These results demonstrate that PP-LiteSeg can be effectively adapted to SAR imagery and that ASFEM improves its segmentation capability while largely preserving the efficiency of the original lightweight architecture.

Keywords

Synthetic Aperture Radar (SAR); Sentinel-1; semantic segmentation; land-cover mapping; PP-LiteSeg; ASFEM; Coordinate Attention; lightweight deep learning; remote sensing

1. Introduction

Accurate land-cover information is essential for monitoring environmental and human-driven changes. Remote sensing enables observations over large areas, while semantic segmentation provides land-cover information at the pixel level. Synthetic Aperture Radar (SAR) is particularly valuable because it operates independently of sunlight and can acquire imagery under cloud-covered conditions [1], [2]. However, SAR segmentation remains challenging because speckle can obscure fine spatial structures, while heterogeneous or similar backscatter responses can make neighboring land-cover classes difficult to distinguish [3], [4].

Deep-learning approaches have increasingly been applied to these challenges. Šćepanović et al. [10] evaluated several semantic segmentation architectures for wide-area land-cover mapping using Sentinel-1 data, while Erten et al. [1] investigated U-Net, PSPNet, and HRNet on high-resolution SAR imagery. Although established segmentation networks can provide effective spatial representations [5]–[7], their computational requirements may restrict their use in resource-constrained applications. This has encouraged the development of more efficient architectures and feature-processing strategies [7], [12], [13].

PP-LiteSeg is a lightweight segmentation architecture designed around efficient feature extraction and decoding [8], with STDC providing a compact feature-extraction structure [11]. At the same time, attention-based feature refinement offers a way to improve spatial representations without relying on substantially larger networks. Coordinate Attention incorporates directional positional information with limited computational requirements [9], while recent remote-sensing and SAR studies have explored attention, boundary enhancement, saliency-aware fusion, and feature-refinement mechanisms to improve spatial representation [12], [14], [15]. However, the application of PP-LiteSeg specifically to SAR semantic segmentation remains largely unexplored in the literature considered in this study.

To address this gap, this study adapts PP-LiteSeg to multi-class Sentinel-1 SAR land-cover segmentation and introduces an Adaptive Spatial Feature Enhancement Module (ASFEM) to refine its intermediate spatial representations. ASFEM combines convolutional processing, Coordinate Attention, depthwise convolution, and residual learning while introducing limited additional complexity. The framework is evaluated on the TurkeySAR dataset against the adapted PP-LiteSeg baseline, with previously reported segmentation results included for broader context. The main contributions are:

- adaptation and evaluation of PP-LiteSeg for multi-class Sentinel-1 SAR land-cover segmentation;
- development of ASFEM for lightweight spatial feature refinement;
- an improvement in mIoU from 69.45% to 70.70%, with approximately 0.17% additional parameters and 0.06% additional computational cost; and

- evaluation through overall, class-wise, qualitative, and computational-complexity analyses.

2. Materials and Methods

2.1. Proposed PP-LiteSeg-ASFEM Framework

The proposed framework adapts PP-LiteSeg to Sentinel-1 SAR semantic segmentation and extends it with the Adaptive Spatial Feature Enhancement Module (ASFEM), as illustrated in Fig. 1. PP-LiteSeg was selected for its efficient real-time segmentation architecture, which combines hierarchical feature extraction with lightweight multi-scale processing and decoding [8].

Given a preprocessed Sentinel-1 SAR image, the PP-LiteSeg encoder extracts hierarchical features at multiple spatial resolutions. Because speckle and complex backscatter can reduce the spatial separability of land-cover classes [3], [4], [13], [14], ASFEM is introduced between the encoder and decoder to refine the intermediate representation. The module combines convolutional processing, Coordinate Attention, depthwise convolution, and residual feature fusion to strengthen spatial information with limited additional computational complexity.

The refined features are subsequently processed by the PP-LiteSeg decoder through multi-scale feature fusion to produce the final pixel-wise prediction across the five land-cover categories. Rather than increasing the depth or width of the backbone, the proposed approach focuses on targeted feature refinement while retaining the lightweight structure of PP-LiteSeg.

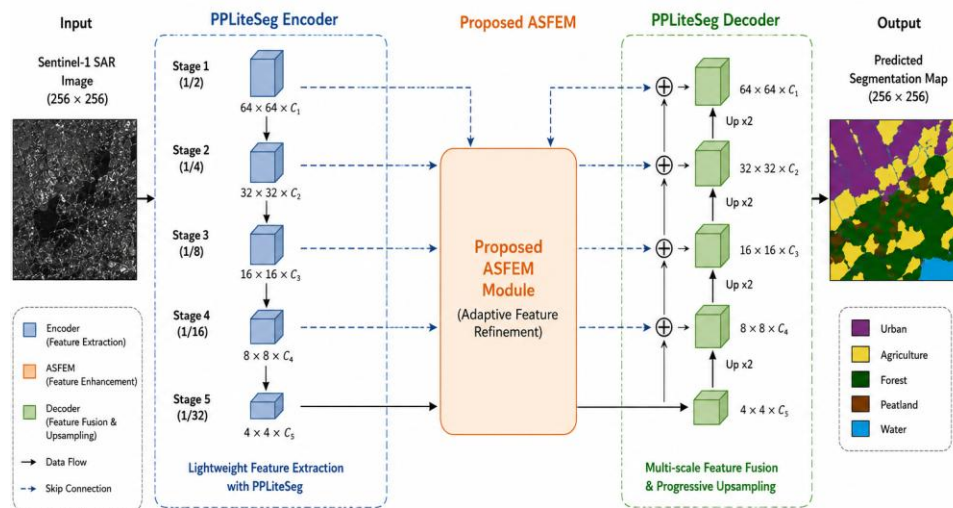


Figure 1: Overall architecture of the proposed PP-LiteSeg-ASFEM framework. ASFEM is inserted between the PP-LiteSeg encoder and decoder to refine intermediate feature representations before multi-scale decoding.

2.2. Adaptive Spatial Feature Enhancement Module (ASFEM)

The proposed ASFEM architecture is illustrated in Fig. 2. Given an intermediate encoder feature map, ASFEM first applies a 3×3 convolution followed by batch normalization and ReLU activation to extract local spatial features. The resulting representation is processed by Coordinate Attention (CA) [9], which encodes information along the horizontal and vertical directions to retain positional information during feature weighting. A 3×3 depthwise convolution, followed by batch normalization and ReLU activation, then refines the spatial representation with fewer parameters than a standard convolution with equivalent kernel and channel dimensions. Finally, a residual connection combines the refined features with the module input before the representation is passed to the decoder.

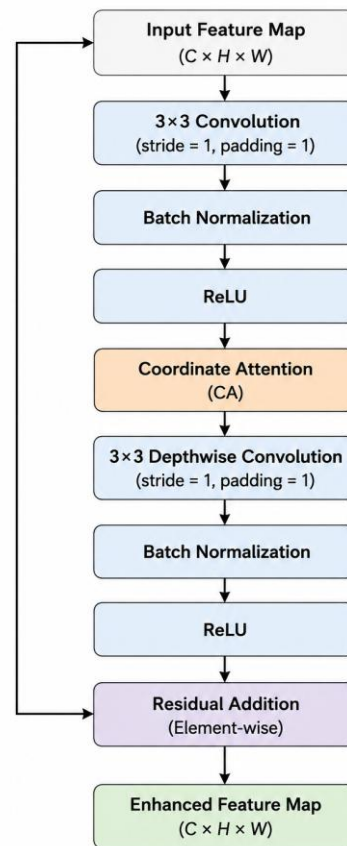


Figure 2: Detailed architecture of the proposed Adaptive Spatial Feature Enhancement Module (ASFEM).

2.2.1. Mathematical Formulation

Let the encoder output feature map to ASFEM be denoted by

$$F \in R^{C \times H \times W} \quad (1)$$

where C, H and W represent the number of channels, height, and width, respectively. The initial feature transformation is expressed as:

$$F_1 = \text{ReLU} \left(\text{BN}(\text{Conv}_{3 \times 3}(F)) \right) \quad (2)$$

Coordinate Attention [9] then aggregates directional information as

$$z_c^h(h) = \frac{1}{W} \sum_{i=1}^W F_{1,c}(h, i), \quad (3)$$

$$z_c^w(w) = \frac{1}{H} \sum_{j=1}^H F_{1,c}(j, w) \quad (4)$$

The resulting descriptors are transformed into directional attention maps, producing

$$F_2 = F_1 \odot A_h \odot A_w \quad (5)$$

where A^h and A^w denote the directional attention maps and $\odot$ represents element-wise multiplication. Local spatial refinement is then performed as

$$F_3 = \text{ReLU} \left(\text{BN}(\text{DWConv}_{3 \times 3}(F_2)) \right) \quad (6)$$

Finally, residual fusion produces the ASFEM output:

$$F_{out} = F_3 + F \quad (7)$$

which is subsequently passed to the PP-LiteSeg decoder.

2.3. Dataset and Experimental Configuration

Experiments were conducted using the TurkeySAR dataset introduced by Erten et al. [1], comprising 6,722 Sentinel-1 GRD SAR image-mask pairs with pixel-level annotations across five land-cover classes: Urban, Agricultural Areas, Forest, Peatland/Bogs/Marshes, and Water. Images were resized to 256×256 pixels, while masks were resized using nearest-neighbor interpolation to preserve discrete class labels. The original labels were converted to zero-based indexing.

The dataset was divided into separate training, validation, and test subsets with a fixed random seed of 42. The same partition was used for PP-LiteSeg and PP-LiteSeg-ASFEM to ensure a controlled comparison.

The models were implemented in PyTorch and trained on the Kaggle platform using an NVIDIA Tesla T4 GPU. Training was performed for 50 epochs with an initial learning rate of 1×10^{-3} , weight decay of 1×10^{-4} , cross-entropy loss, and a batch size of 16.

Horizontal and vertical flips and random 90° rotations were applied jointly to training images and masks. No augmentation was applied during validation, and the model with the best validation performance was retained for evaluation.

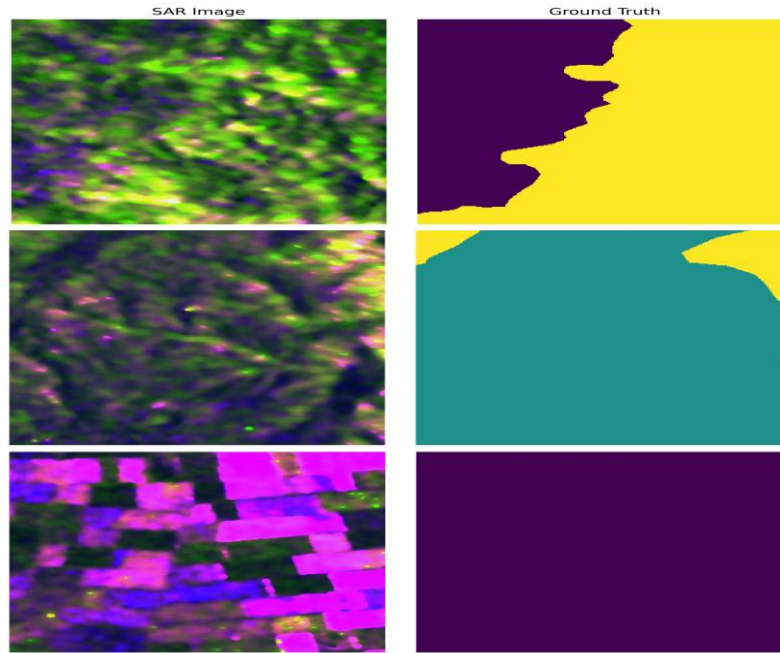


Figure 3: Representative samples from the TurkeySAR dataset showing Sentinel-1 SAR images and corresponding ground-truth masks.

2.4. Evaluation Metrics

Performance was evaluated using Overall Accuracy (OA), Precision, Recall, F1-score, Intersection over Union (IoU), and Mean Intersection over Union (mIoU). mIoU was considered the primary segmentation metric because it gives equal weight to each semantic class. IoU for class i is defined as

$$IoU = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (8)$$

where TP_i , FP_i , and FN_i denote true-positive, false-positive, and false-negative pixels, respectively. Mean IoU is computed as

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU_i \quad (9)$$

where C is the number of semantic classes.

3. Results

3.1. Overall Performance and Computational Complexity

The original PP-LiteSeg and proposed PP-LiteSeg-ASFEM models were evaluated under the same experimental configuration. For broader context, previously reported U-Net, PSPNet, and HRNet results on TurkeySAR are also included in Table 1.

Compared with the PP-LiteSeg baseline, ASFEM increased mIoU from 69.45% to 70.70%, corresponding to a gain of 1.25 percentage points, while OA changed only slightly from 89.81% to 89.67%. The resulting mIoU is also competitive with the 70.60% reported for HRNet on the same dataset [1].

Table 2 shows that this improvement is obtained with limited additional complexity. ASFEM adds 10,488 parameters, increasing the parameter count by approximately 0.17%, while computational cost increases by approximately 0.06%, from 2.2190 to 2.2203 GFLOPs. Model size changes only from 24.18 MB to 24.23 MB.

Table 1: Quantitative comparison of the proposed PPLiteSeg-ASFEM with representative semantic segmentation models on the TurkeySAR dataset.

Method	OA (%)	MIoU (%)
UNet	88.59	62.87
PSPNet	90.74	65.99
HRNet	92.23	70.60
PPLiteSeg (Baseline)	89.81	69.45
PPLiteSeg + ASFEM	89.67	70.70

* U-Net, PSPNet, and HRNet results are reported by Erten et al. [1]. PP-LiteSeg and PP-LiteSeg-ASFEM were evaluated in the present study.

Table 2: Computational comparison of PP-LiteSeg and PP-LiteSeg-ASFEM

Model	Parameters (M)	Size (MB)	GFLOPs	mIoU (%)
PPLiteSeg	6.3105	24.18	2.2190	69.45
PPLiteSeg-ASFEM	6.3210	24.23	2.2203	70.70

3.2. Class-wise Performance Analysis

Table 3 and Figures 4–5 provide the class-wise performance analysis of PP-LiteSeg-ASFEM. Water achieved the strongest performance, with 97.46% IoU and 98.72% F1-score, followed by Forest and Agricultural Areas. Urban achieved 67.64% IoU, while Peatland, Bogs and Marshes remained the most challenging class, with 31.55% IoU and 47.96% F1-score. Its lower performance is likely influenced by its limited representation in the training data. The confusion matrix shows the same general pattern, with greater

prediction errors for Urban and particularly for the underrepresented Peatland/Bogs/Marshes category.

Table 3: Per-class Intersection over Union (IoU) achieved by the proposed PPLiteSeg-ASFEM model on the TurkeySAR dataset.

Land-cover class	IoU (%)
Urban	67.64
Agricultural Areas	77.53
Forest	79.30
Peatland, Bogs and Marshes	31.55
Water	97.46
Mean IoU	70.70

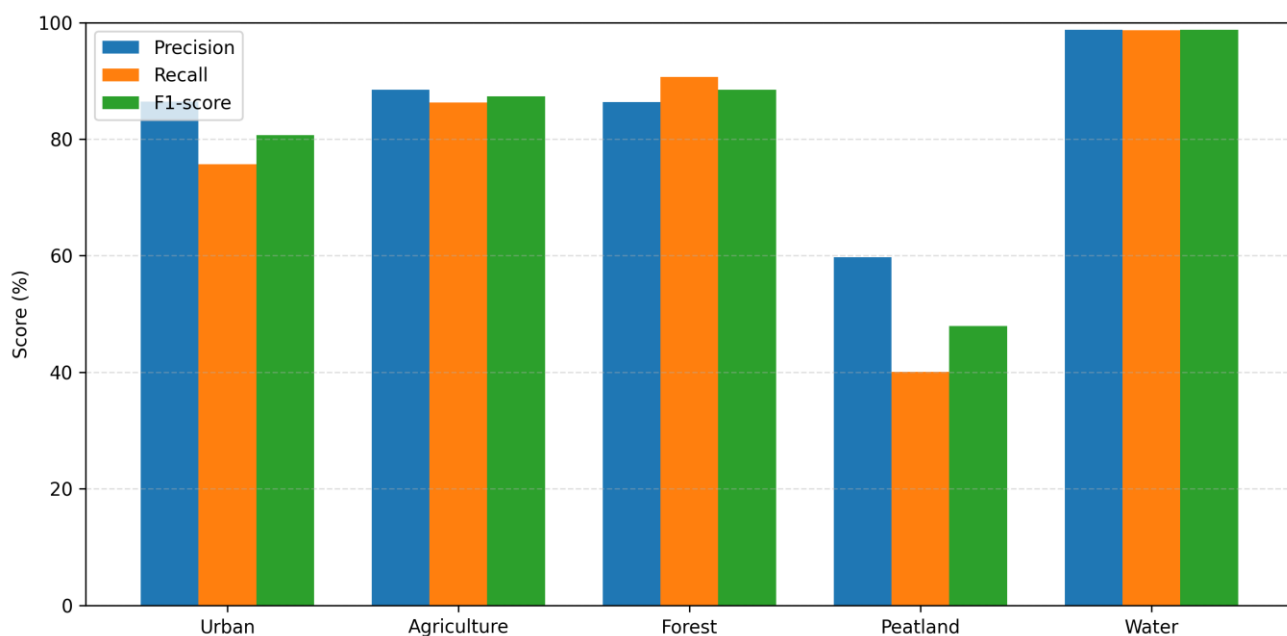


Figure 4: Class-wise Precision, Recall, and F1-score obtained by the proposed PPLiteSeg-ASFEM model.



Figure 5: Confusion matrix obtained by the proposed PPLiteSeg-ASFEM model on the TurkeySAR dataset.

3.3. Qualitative Results

Figure 6 presents representative PP-LiteSeg-ASFEM predictions alongside the corresponding SAR inputs and ground-truth masks. The predictions generally reproduce the dominant land-cover regions, while some discrepancies remain in smaller areas and around complex class boundaries. These observations are consistent with the class-wise quantitative results.

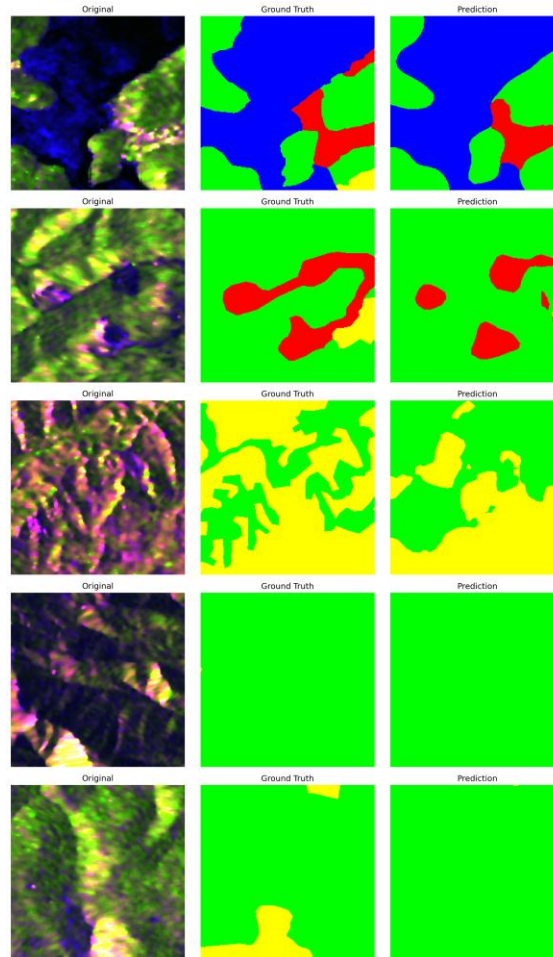


Figure 6: Representative qualitative segmentation results produced by the proposed PPLiteSeg-ASFEM model on the TurkeySAR dataset.

4. Discussion

The results demonstrate that PP-LiteSeg can be effectively adapted to Sentinel-1 SAR semantic segmentation and further improved through targeted feature refinement. Integrating ASFEM increased mIoU from 69.45% to 70.70%, a gain of 1.25 percentage points, while OA remained comparable (89.81% versus 89.67%). Importantly, this improvement required only approximately 0.17% additional parameters and 0.06% additional computational cost, indicating that ASFEM improves segmentation performance without substantially increasing the complexity of the adapted backbone.

For broader context, PP-LiteSeg-ASFEM achieved 70.70% mIoU, compared with the 70.60%, 65.99%, and 62.87% reported for HRNet, PSPNet, and U-Net, respectively, on TurkeySAR [1]. Although the models were not reproduced under identical experimental

conditions, they indicate that the proposed framework achieves competitive segmentation performance. More importantly, the controlled comparison with the adapted PP-LiteSeg baseline directly demonstrates the contribution of ASFEM.

Class-wise results show strong performance for Water, Forest, and Agricultural Areas, whereas the substantially underrepresented Peatland/Bogs/Marshes class remained the main challenge. Overall, the findings demonstrate the applicability of PP-LiteSeg to Sentinel-1 SAR segmentation, an application that remains largely unexplored in the literature considered in this study, and show that ASFEM can improve this lightweight adaptation with minimal additional complexity.

5. Conclusion

This study adapted PP-LiteSeg to multi-class Sentinel-1 SAR semantic segmentation and introduced ASFEM to improve its spatial feature representation. The adapted PP-LiteSeg baseline achieved 69.45% mIoU, demonstrating its applicability to SAR land-cover segmentation, while the proposed PP-LiteSeg-ASFEM framework increased mIoU to 70.70%. This 1.25-percentage-point improvement required only approximately 0.17% additional parameters and 0.06% additional computational cost, showing that the segmentation gain can be achieved while largely preserving the lightweight characteristics of the original architecture.

The findings therefore support two main conclusions: PP-LiteSeg provides an effective lightweight foundation for SAR semantic segmentation, an application that has received limited attention in the existing literature, and the proposed ASFEM further improves its segmentation performance without substantially increasing architectural complexity. The remaining performance gap for the underrepresented Peatland/Bogs/Marshes class also indicates that class imbalance remains an important challenge.

Future work will investigate strategies for improving underrepresented classes, including more balanced training data and class-aware learning approaches. The framework will also be evaluated on additional SAR datasets and geographic regions to assess its generalization beyond TurkeySAR. Further work may examine robustness under SAR-specific degradations, including stronger speckle and reduced spatial resolution, as well as optimization for deployment on resource-constrained hardware.

References

- [1] H. Erten et al., "Semantic Segmentation with High-Resolution Sentinel-1 SAR Data," *Applied Sciences*, vol. 13, no. 10, p. 6025, May 2023, doi: 10.3390/app13106025.
- [2] J. M. Álvarez-Martínez et al., "Four decades of remote sensing for monitoring terrestrial ecosystems: a global review and future challenges," *Science of Remote Sensing*, vol. 13, p. 100341, June 2026, doi: 10.1016/j.srs.2025.100341
- [3] Q. Li and Y. Kong, "An Improved SAR Image Semantic Segmentation Deeplabv3+ Network Based on the Feature Post-Processing Module," *Remote Sensing*, vol. 15, no. 8, p. 2153, Apr. 2023, doi: 10.3390/rs15082153.
- [4] J. Pluto-Kossakowska and S. Wangiyana, "Supervised Semantic Segmentation of Urban Area Using SAR," *Remote Sensing*, vol. 17, no. 9, p. 1606, May 2025, doi: 10.3390/rs17091606.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234-241, 2015.
- [6] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2881-2890.
- [7] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021.
- [8] J. Peng et al., "PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model," *arXiv preprint arXiv:2204.02681*, 2022.
- [9] Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13713-13722.
- [10] S. Šćepanović, O. Antropov, P. Laurila, Y. Rauste, V. Ignatenko, and J. Praks, "Wide-area land cover mapping with Sentinel-1 imagery using deep learning semantic segmentation models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10357–10374, 2021, doi: 10.1109/JSTARS.2021.3116094.
- [11] M. Fan et al., "Rethinking BiSeNet for Real-Time Semantic Segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 9716-9725.
- [12] X. Li, L. Xie, C. Wang, J. Miao, H. Shen, and L. Zhang, "Boundary-enhanced dual-stream network for semantic segmentation of high-resolution remote sensing images," *GIScience & Remote Sensing*, 2024.
- [13] Qiu, Y., Zou, B., and L. Zhang, "A lightweight and fast method for water and airport segmentation in SAR images," in *Proc. 15th Eur. Conf. Synthetic Aperture Radar (EUSAR 2024)*, 2024, p. 6.

[14] R. Xu, C. Wang, J. Zhang, S. Xu, W. Meng, and X. Zhang, "RSSFormer: Foreground saliency enhancement for remote sensing land-cover segmentation," *IEEE Transactions on Image Processing*, 2023.

[15] W. Wei, Y. Ye, G. Chen, Y. Zhao, X. Yang, L. Zhang, and Y. Zhang, "SAR remote sensing image segmentation based on feature enhancement," *Neural Networks*, vol. 185, p. 107190, 2025, doi: 10.1016/j.neunet.2025.107190.