MARSS: Radar Semantic Segmentation via Modular Attention and State Space Models

Supplementary Material

Table 1. Ablation Study on component effectiveness and RADM internal design. We assess the contribution of RADE and RFAF modules, and benchmark the proposed Mamba-based RADM against CNN and Transformer (Tr.) variants to verify the superiority of Mamba for RSS tasks. Results are in terms of mIoU and mDice. Best results are **bold**, second-best are underlined

RADE & RFAF	RADM	RD View(%)		RA View(%)	
		mIoU	mDice	mIoU	mDice
		61.16%	72.91%	41.04%	50.49%
	CNN	61.86%	72.78%	44.57%	54.46%
	Tr.	61.11%	73.34%	44.01%	55.31%
	Mamba	62.52%	74.35%	45.33%	56.69%
✓		62.89%	74.72%	45.33%	56.59%
✓	CNN	61.84%	73.68%	45.57%	55.14%
✓	Tr.	62.64%	74.49%	45.46%	56.86%
✓	Mamba	**63.26%**	**75.18%**	**46.97%**	**58.78%**

1. Extended Ablation Study

1.1. Ablation of Mamba in RADM

In order to confirm if Mamba is necessary, or if performance gains come simply from increased parameters. We performed a rigorous ablation (as shown in Table 1) replacing the Mamba backbone in RADM with CNN and Transformer equivalents.

- **Performance:** Mamba outperforms both CNN and Transformer variants in mIoU/mDice.
- **Efficiency:** The Mamba-based RADM utilizes the *fewest* parameters (9.25M) compared to Transformer (9.58M) and CNN (9.71M) variants.

This confirms that Mamba is the superior backbone choice for RSS, offering better long-range modeling with higher parameter efficiency, rather than increasing model capacity.

1.2. Ablation of two stages in RFAF

Building upon the original ablation studies, we conduct additional experiments focusing on the two distinct stages of the RFAF (Radar Feature Adaptive Fusion Module) introduced in Section 3.3 (as shown in Figure.1 and Table.3), thereby complementing the initial ablation analysis. The two feature enhancement phases of the RFAF Module operate relatively independently, with each stage capable of being employed as a standalone enhancement component.

2. Real-Time Capabilities

We provide a comprehensive runtime analysis in Table 2. On an NVIDIA RTX 3090, MARSS achieves **33 FPS**, significantly outperforming the previous SOTA, AdaPKC$^{\xi}$-Net (24 FPS), despite a moderate increase in MACs (Multiply-Accumulate operations). This demonstrates that MARSS is suitable for real-time autonomous driving deployment.

Table 2. Complexity and runtime comparison. FPS is calculated on a workstation with an NVIDIA RTX 3090 GPU.

Method	**Params(M)**	**MACs(G)**	**FPS**
TMVA-Net	5.6	102.3	63
AdaPKC$^{\xi}$-Net	6.3	109.8	24
AdaPKC$^{\theta}$-Net	6.3	110.1	23
MARSS(ours)	9.3	142.0	33

3. More Visualization Results

Figure.2 illustrates more segmentation results under challenging cases (e.g., low SNR, small fast objects like the car) on CARRADA dataset. Further qualitative visualization analysis indicates that MARSS has better segmentation performance compared to baselines.

References

Table 3. Extended Ablation Study: Assess the contribution of each module (RADE, RFAF, RADM) and their sub-stages (for RFAF) in our framework. Results are reported in terms of mIoU and mDice for RD and RA views. Best results are **bold**, second-best are <u>underlined</u>.

RADE	RFAF-Stage 1	RFAF-Stage 2	RADM	RD View		RA View	
				mIoU	mDice	mIoU	mDice
				61.16%	72.91%	41.04%	50.49%
✓				62.35%	74.24%	41.57%	51.69%
	✓			62.32%	74.29%	43.74%	54.31%
		✓		62.15%	74.03%	44.18%	55.17%
			✓	62.52%	74.35%	45.33%	56.69%
✓	✓			62.67%	74.67%	43.59%	54.29%
✓		✓		<u>62.96%</u>	74.70%	44.53%	55.51%
✓			✓	62.43%	74.09%	43.53%	54.09%
	✓	✓		62.45%	74.48%	44.70%	55.69%
		✓	✓	62.43%	74.18%	<u>45.66%</u>	<u>57.06%</u>
✓	✓	✓		62.89%	<u>74.72%</u>	45.33%	56.69%
✓	✓		✓	62.34%	74.18%	45.49%	56.80%
	✓	✓	✓	62.90%	74.64%	43.76%	54.71%
✓	✓	✓	✓	**63.26%**	**75.18%**	**46.97%**	**58.78%**

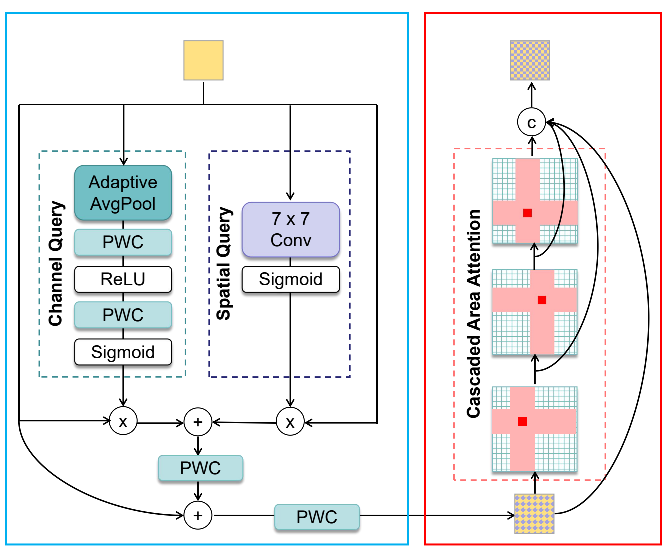
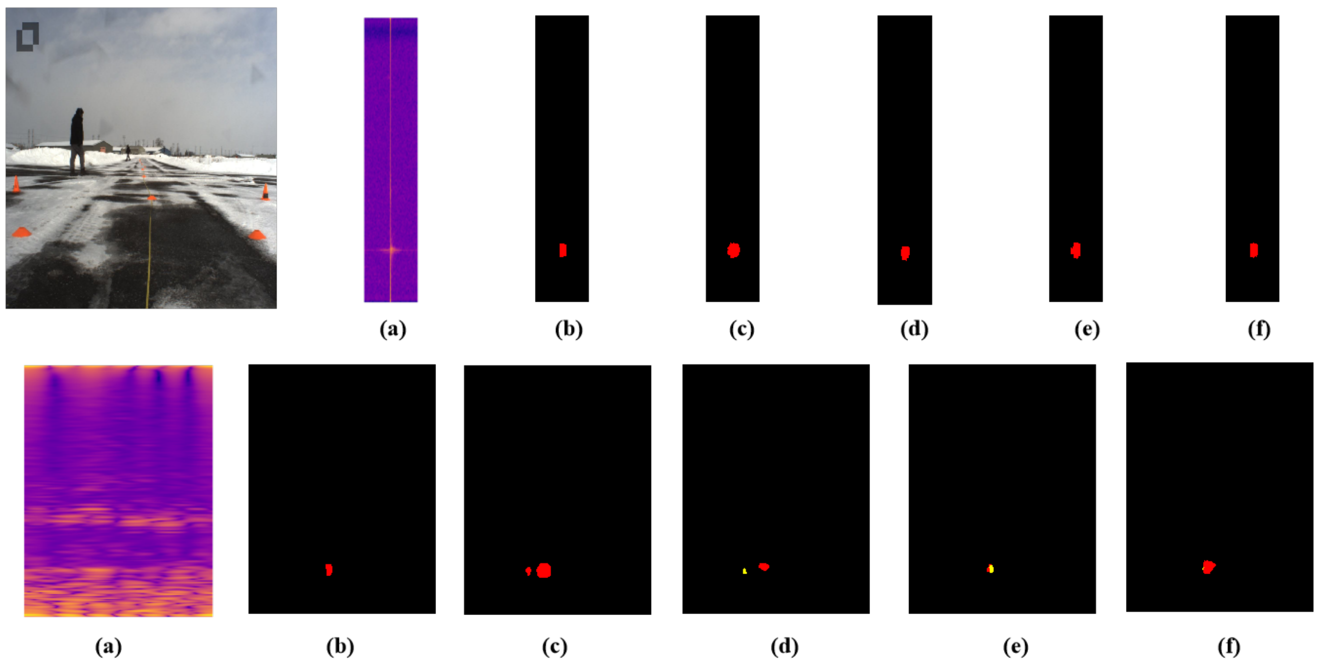
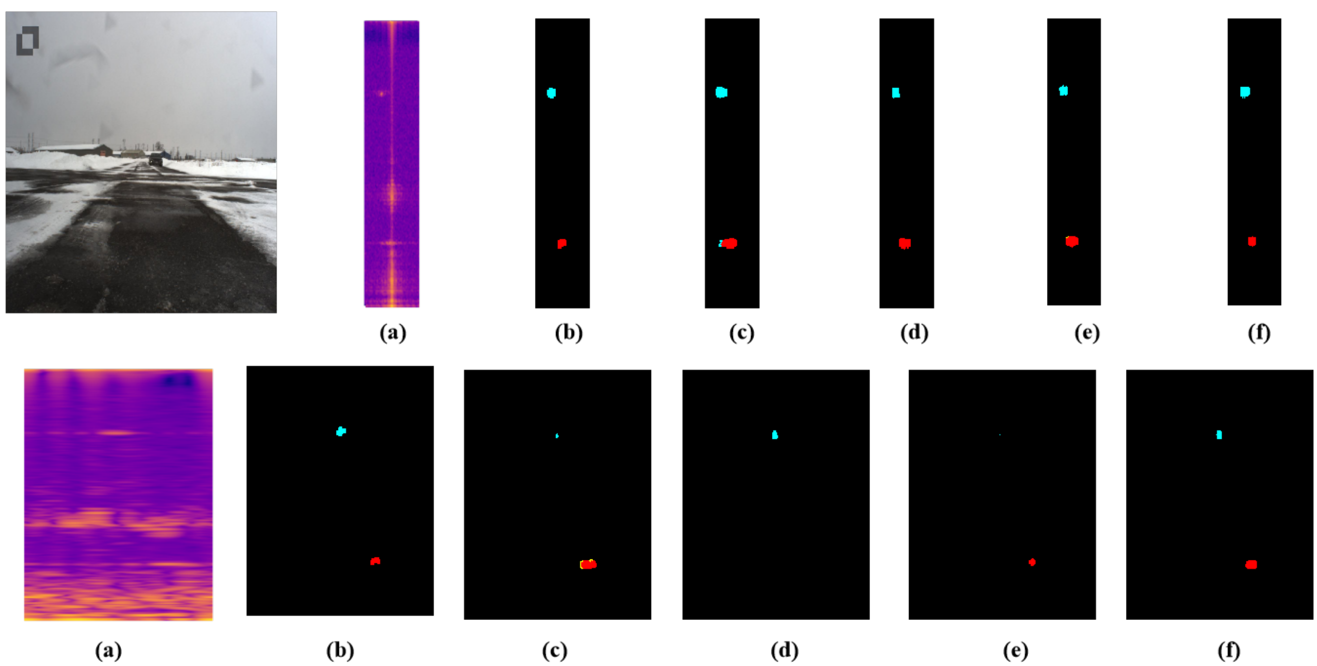


Figure 1. RFAF: Radar Feature Adaptive Fusion Module, used to enhance the intermediate feature fusion layer. PWC represents point-wise convolution (e.g., 1x1 convolution). The part in the blue box on the left is Stage 1, and the part in the red box on the right is Stage 2.



((a)) Case 1



((b)) Case 2

Figure 2. Radar segmentation results visualization: The leftmost image is the corresponding RGB camera image. The first row is RD-View, and the second row is RA-View images. From left to right, they are: (a) original RA/RD data; (b) Segmentation Mask ground truth; (c) TransRadar segmentation results; (d) TMVA-Net segmentation results; (e) AdaPKC$^{\xi}$-Net^{FiT} segmentation results; (f) MARSS(Ours) segmentation results. In the segmentation mask, red corresponds to pedestrians, yellow to cyclists, and blue to cars.