SSMRadNet: A Sample-wise State-Space Framework for Efficient and Ultra-Light Radar Segmentation and Object Detection

Anuvab Sen, Mir Sayeed Mohammad, Saibal Mukhopadhyay
Department of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA

WACV 2026
TUCSON, AZ
3/6 - 3/10

Problem Statement

High-resolution FMCW radar frames ($C \times S \times N_{rx}$) blow up **compute, memory, and latency** in point-cloud/CNN/Transformer based radar object detection pipelines, constraining **real-time operation** and **edge deployment**.

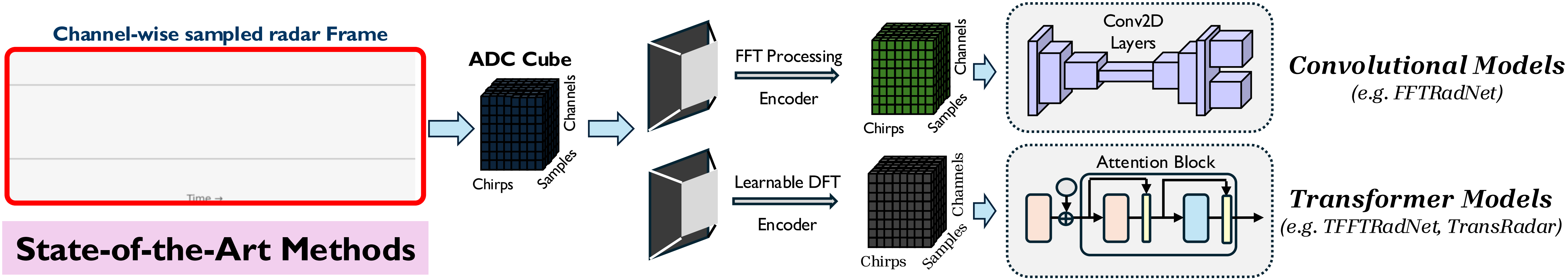
Abstract

We introduce **SSMRadNet**, the first multi-scale State Space Model (SSM) based detector for Frequency Modulated Continuous Wave (FMCW) radar that sequentially processes raw ADC samples.

Compute efficiency: Our model has significantly lower parameters, computation and processing latency compared to prior works, but maintains segmentation and detection performance.

Sample-by-sample radar detection: We propose the first sample-by-sample radar data processing architecture for detection eliminating the memory and latency of buffering ADC cubes.

AI/ML Pipeline for Radar Processing



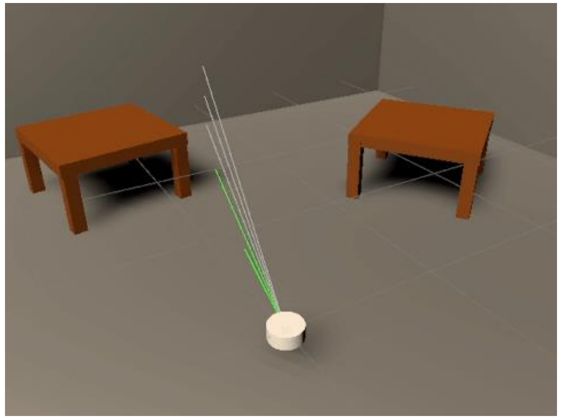
Problem

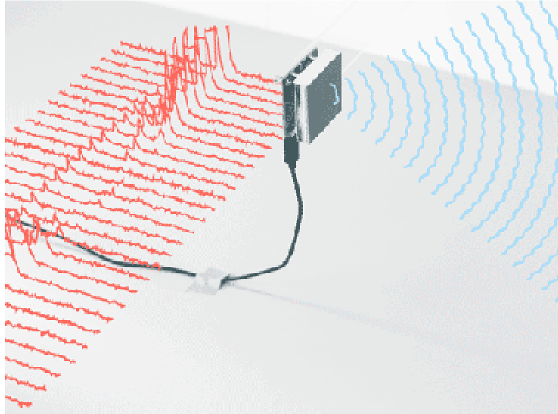
- State-of-the-Art radar detection pipelines require Fourier transform over radar frames.
  - Frame-by-frame processing leads to higher latency, buffering, and computation.
  - Fast Fourier transform (FFT) or learnable FFT front-ends ignore sequence dynamics and smear non-stationary spectra.
- The back-end ML models are based on convolution/transformer stacks
  - The computation and memory increases non-linearly as chirps/samples/antennas grow.
  - The computation latency and power dissipation increases for high-performance radar.

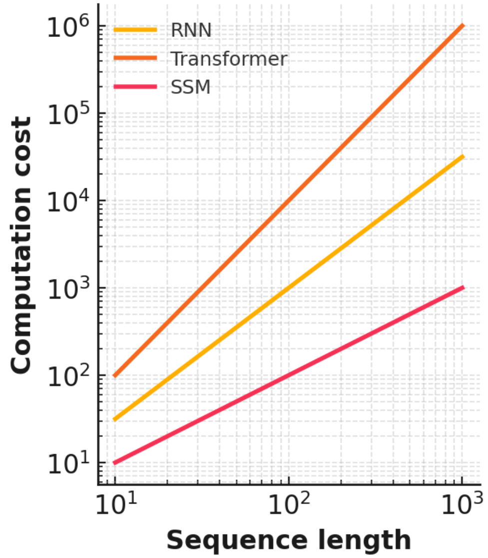
[1] Rebut, Julien, et al. "Raw high-definition radar for multi-task learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
[2] Giroux et. Al. "T-fftradnet: Object detection with swin vision transformers from raw adc radar signals." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Why SSM for LiDAR and Radar?

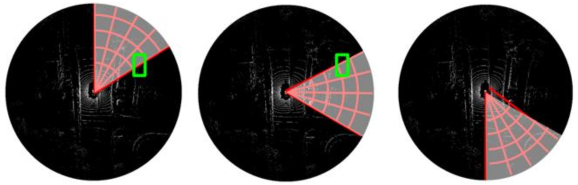
- Underexplored or unexplored for LiDAR and radar data.
- Rotating LiDAR scanners generate point clouds sequentially in the polar domain, and radar generates ADC streams in the form of chirps – making them ideal for linear sequence mode







LiDAR polar voxelization



$t_0 - t_1$ $t_1 - t_2$ $t_2 - t_3$

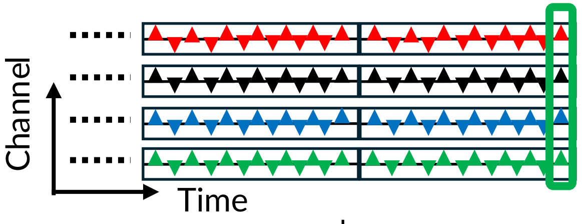
state equation

$\mathbf{h'}(t) = \mathbf{Ah}(t) + \mathbf{Bx}(t)$

output equation

$\mathbf{y}(t) = \mathbf{Ch}(t) + \mathbf{Dx}(t)$

Radar channel-wise sampled signal modeling



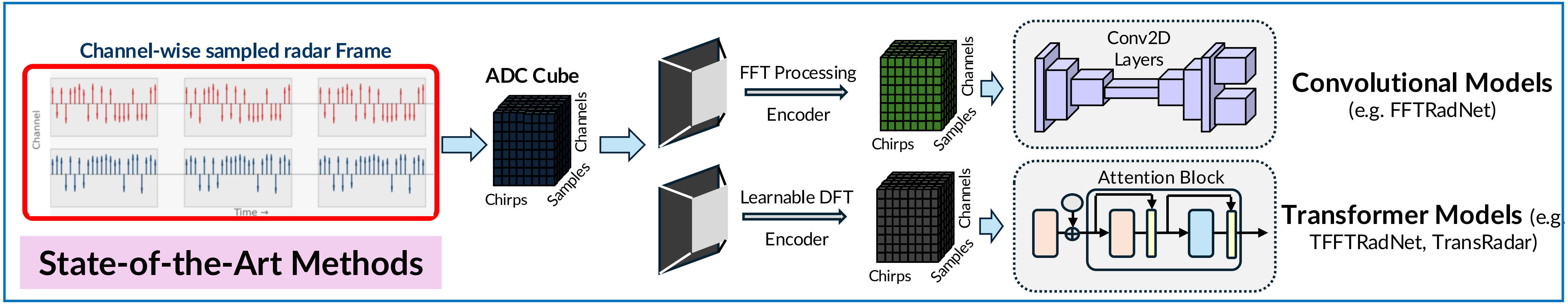
state equation

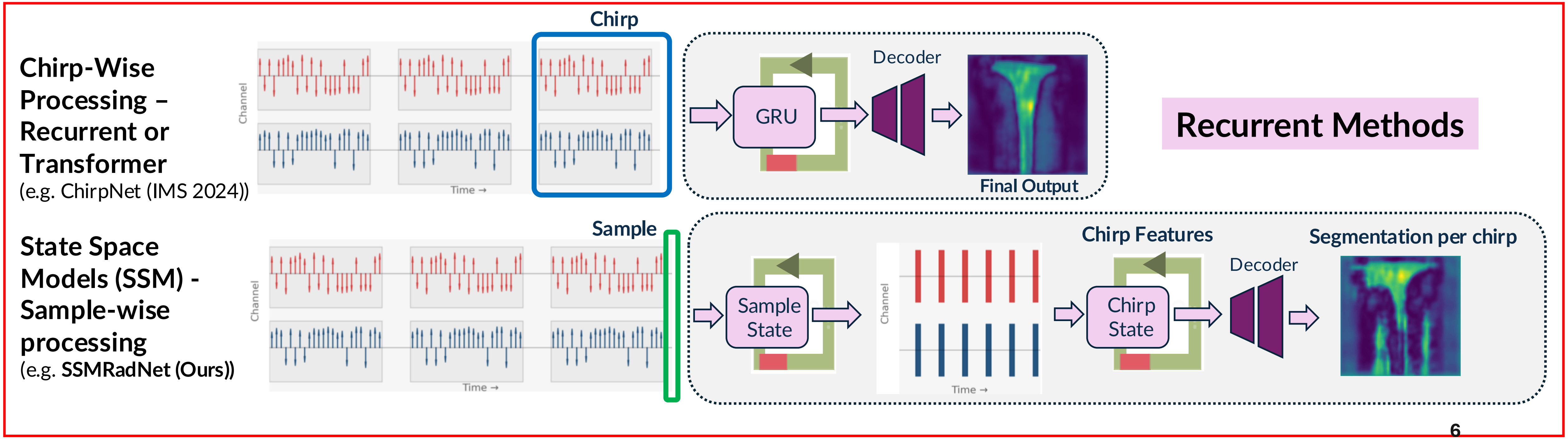
$\mathbf{h'}(t) = \mathbf{Ah}(t) + \mathbf{Bx}(t)$

output equation

$\mathbf{y}(t) = \mathbf{Ch}(t) + \mathbf{Dx}(t)$

- SSM utilizes linear state updates and its **computation scales linearly with sequence length, instead of quadratically, as in transformers**.
- Does not wait for entire LiDAR/Radar frame, achieves streaming detection capability.

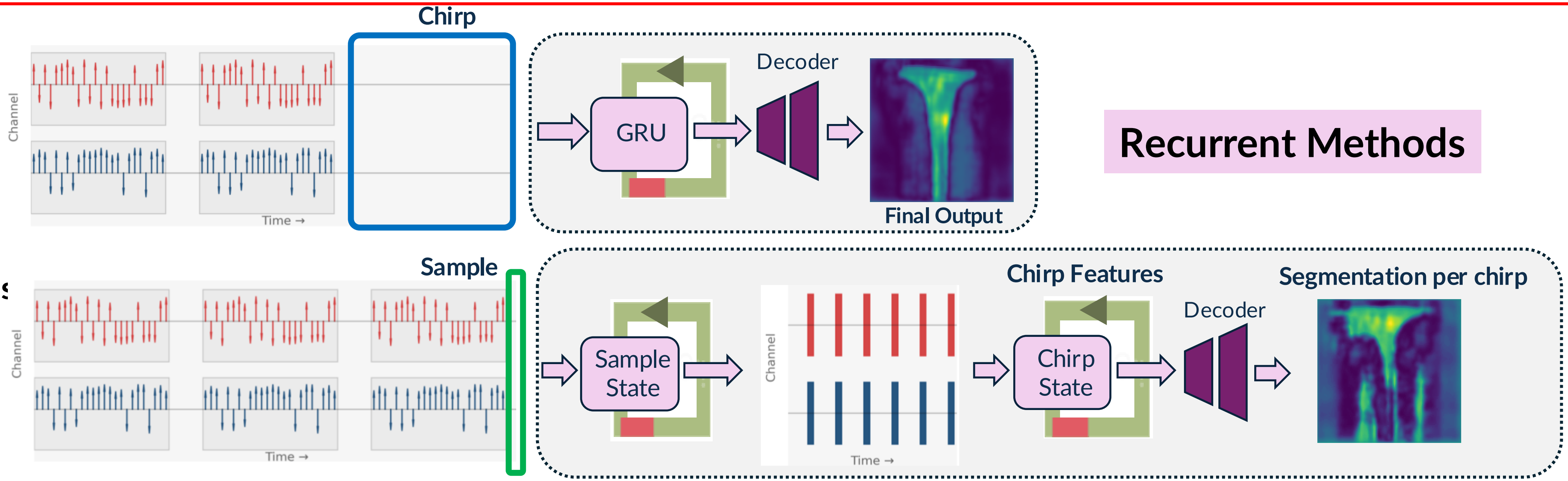

Channel-wise sampled radar Frame
Channel
Time →
ADC Cube
Channels
Chirps
Samples
FFT Processing
Encoder
Learnable DFT
Encoder
Conv2D
Layers
Attention Block
Convolutional Models
(e.g. FFTRadNet)
Transformer Models (e.g.
TFFTRadNet, TransRadar)
State-of-the-Art Methods


Chirp-Wise
Processing –
Recurrent or
Transformer
(e.g. ChirpNet (IMS 2024))
Chirp
Channel
Time →
GRU
Decoder
Final Output
Recurrent Methods
State Space
Models (SSM) -
Sample-wise
processing
(e.g. SSMRadNet (Ours))
Sample
Sample
State
Chirp Features
Chirp
State
Decoder
Segmentation per chirp

Our Approach

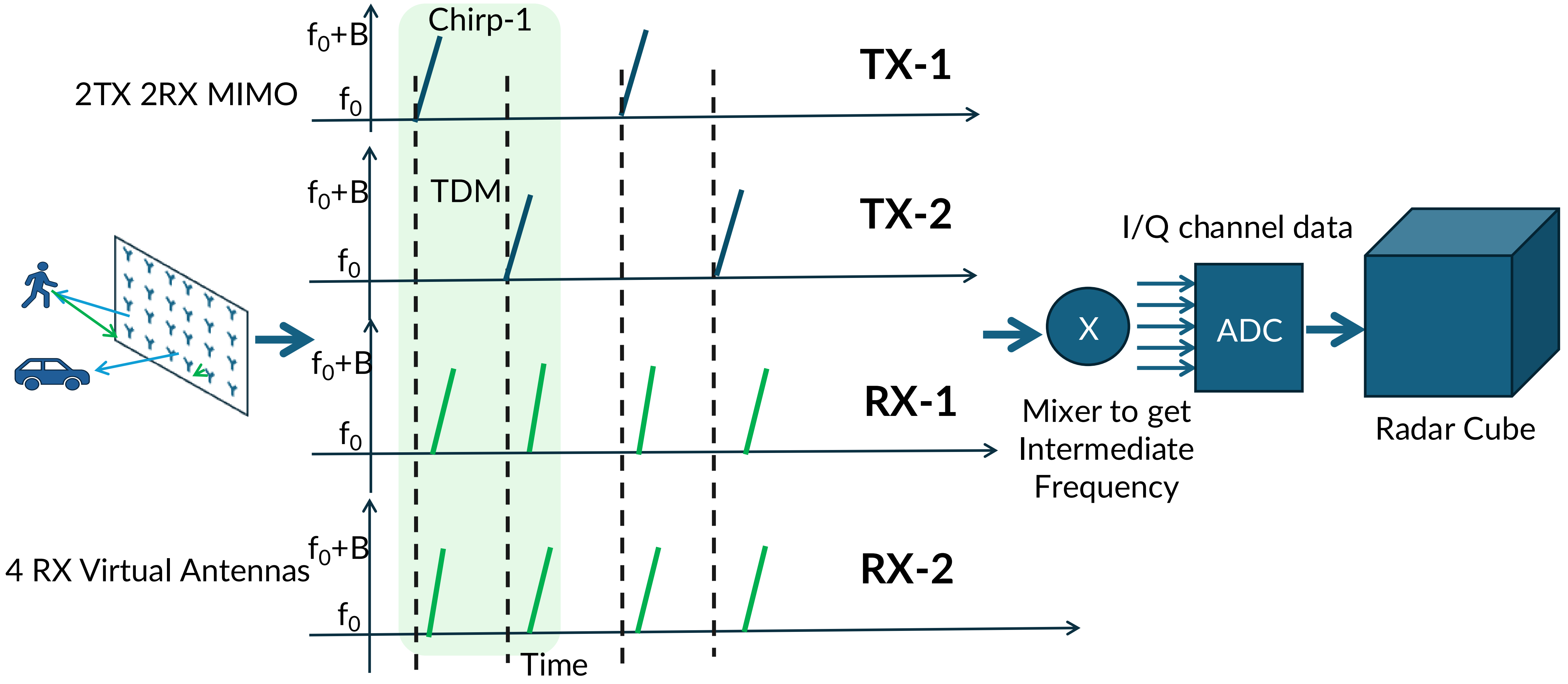
Chirp-Wise Processing – Recurrent or Transformer
(e.g. ChirpNet)

State Space Models (SSM) - Sample-wise processing
(e.g. **SSMRadNet (Ours)**)



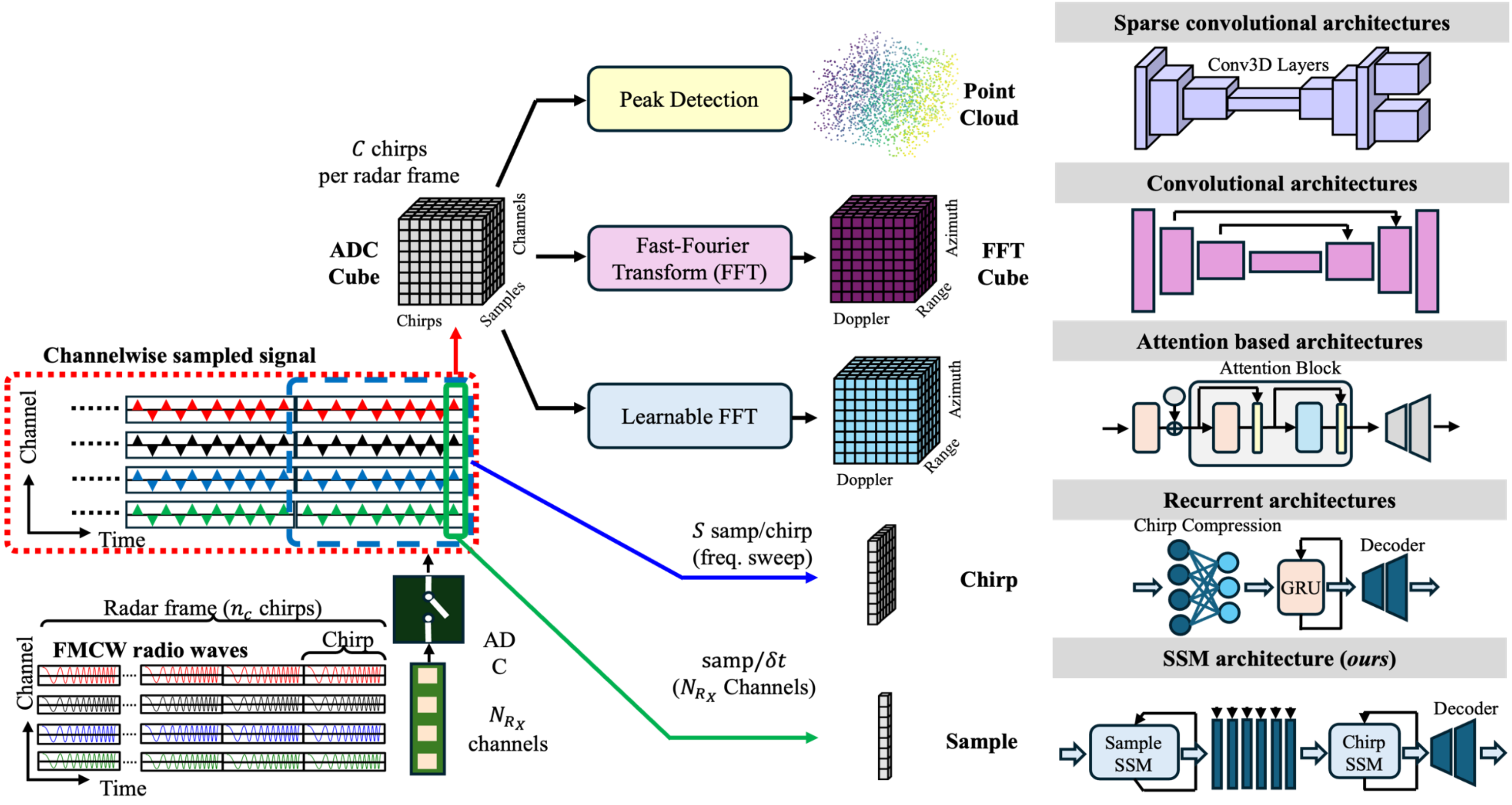
- A family of sequence processing-based AI/ML models for radar (and other types of sensors)
  - **Streaming sensor processing**: Chirp-by-chirp model and sample-by-sample model.

Background – Automotive FMCW Radar



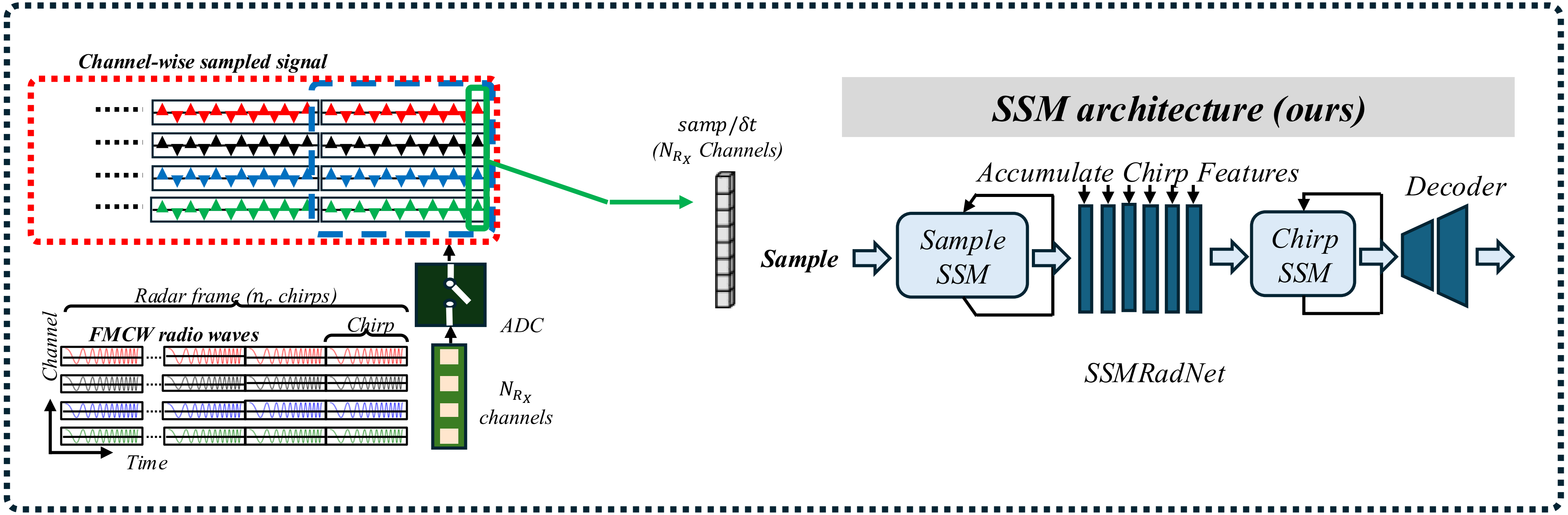
TI Automotive FMCW Radars implements MIMO through time multiplexing through the TX Antenna

Motivation from Prior Works



- ✓ Traditional convolutional [**ADC-Net, FFT-RadNet**], attention-based [**TFFT-RadNet**] and recurrent networks [**ChirpNet**] are computationally expensive because they rely on a large volume of data.
- ✓ Our state space modeling approach scales linearly in computation with increased radar resolution because of our sample-by-sample processing, achieving state of the art performance for a fraction of the computation.

Sample Wise SSM Based Radar OD



Per-sample embedding

➢ Turn every fast-time ADC sample (across all antennas) into a token.

SSM sequence Modelling

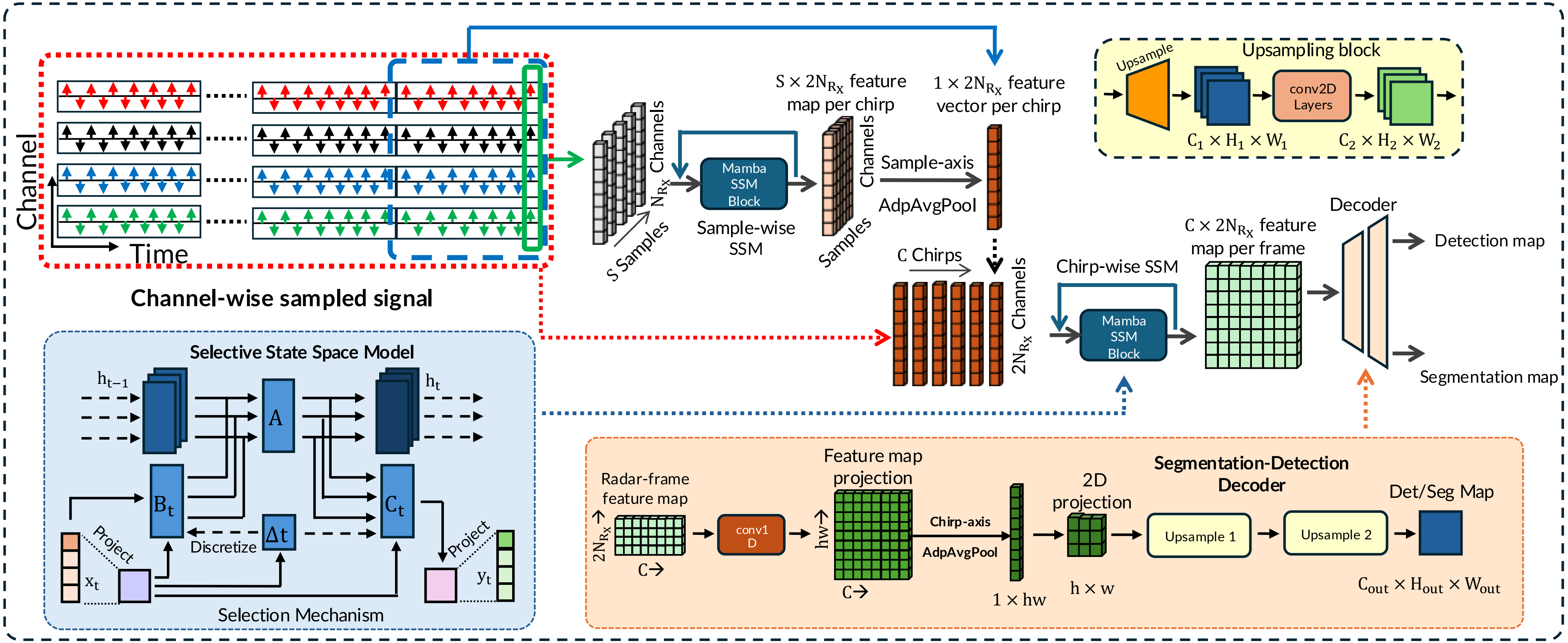
➢ Intra-chirp: model along the sample dimension of each chirp, and inter-chirp: model across chirps.

Spatial decoding

➢ Project sequence outputs to a 2-D spatial feature map. Used for free-space segmentation & vehicle detection.

Note: SSMRadNet is the first multi-scale SSM radar framework that tokenizes sequential ADC samples for sample-by-sample segmentation & detection.

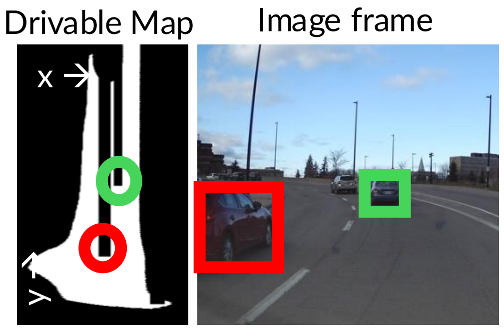
SSMRadNet Architecture



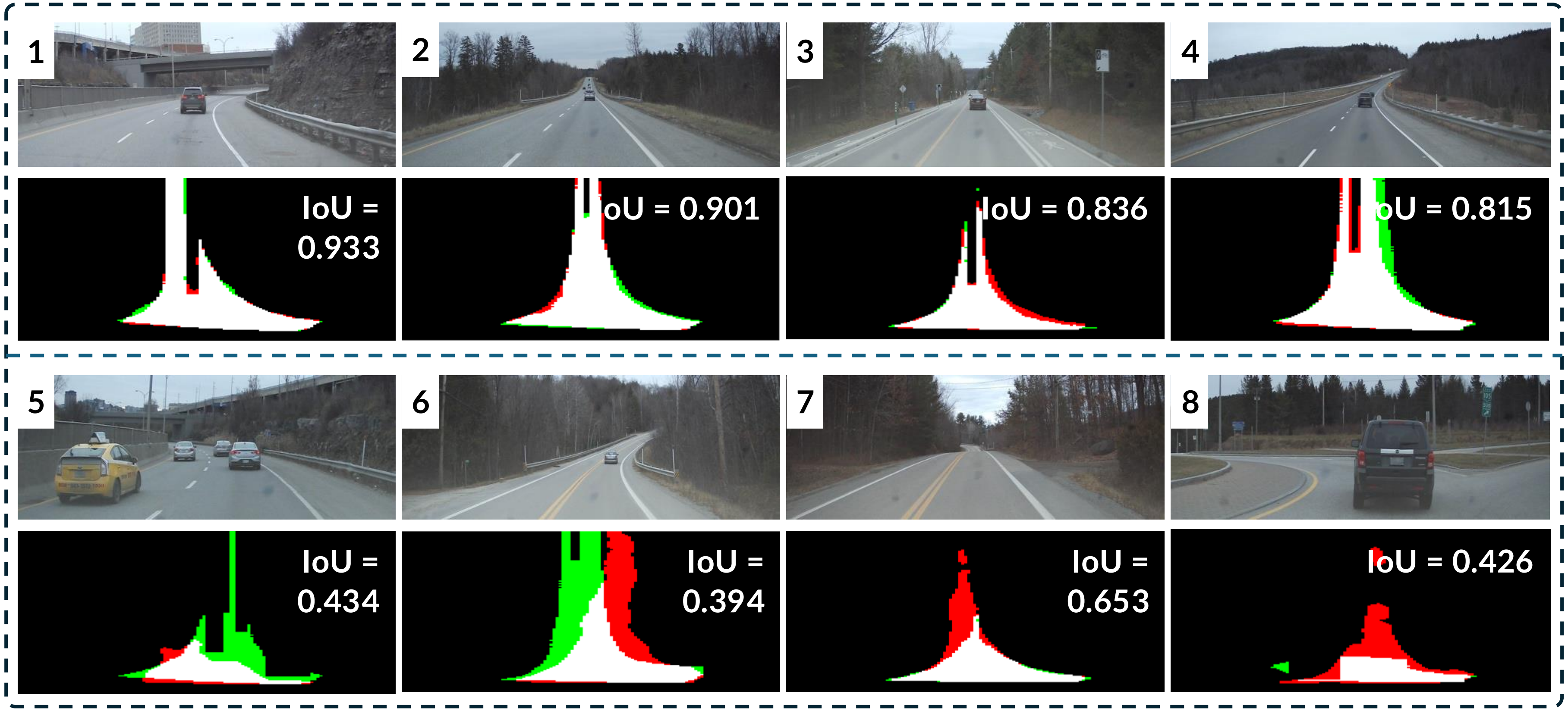
SSMRadNet is an end-to-end neural architecture that takes raw radar ADC samples as input. The Sample states/chirp are aggregated into chirp feature maps and produce a bird's-eye-view occupancy map as output, which is projected to the detection and segmentation head.

Qualitative Results — Segmentation

Dataset and Radar Specifications

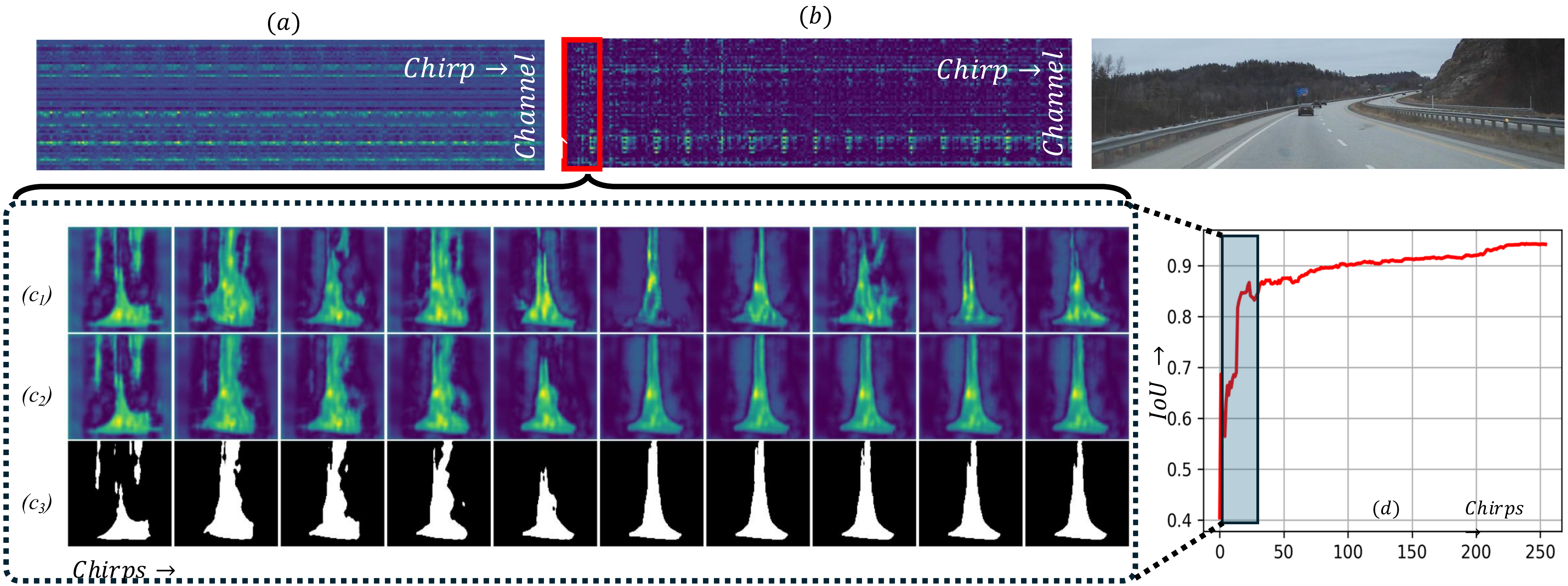


- We evaluate on **RADIal** dataset
- 91 Driving sequences spanning 2 hours (8252 labelled samples)
- HD Radar with 12Tx and 16Rx
- Radar frames: 256 chirps, 512 samples, 16 channels



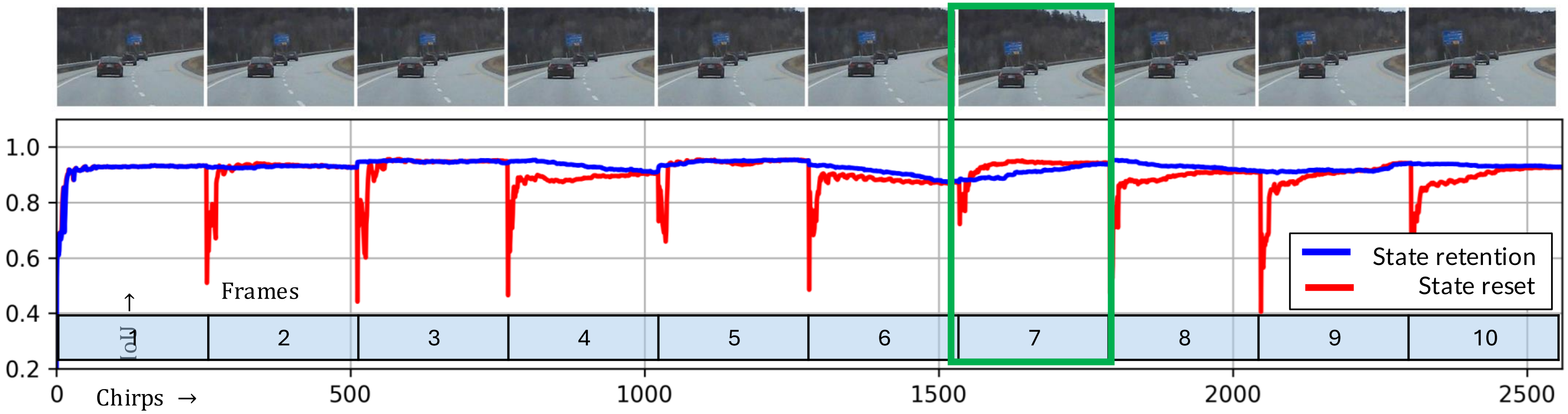
Works well on clear highways. The errors rise with traffic and grade changes. SSMRadNet prior chirp states help in better prediction. Results can be improved by adding motion/elevation priors to harden free-space predictions.

Feature Evolution Across the Model



- Early chirps rapidly accumulate spatial cues; a stable projection forms within ~25 chirps (c1→c3).
- IoU climbs sharply at first then plateaus, indicating fast convergence and state saturation.
- Early saturation ⇒ fewer chirps—cutting compute, despite a small Doppler-resolution trade-off.

State Retention vs. Reset



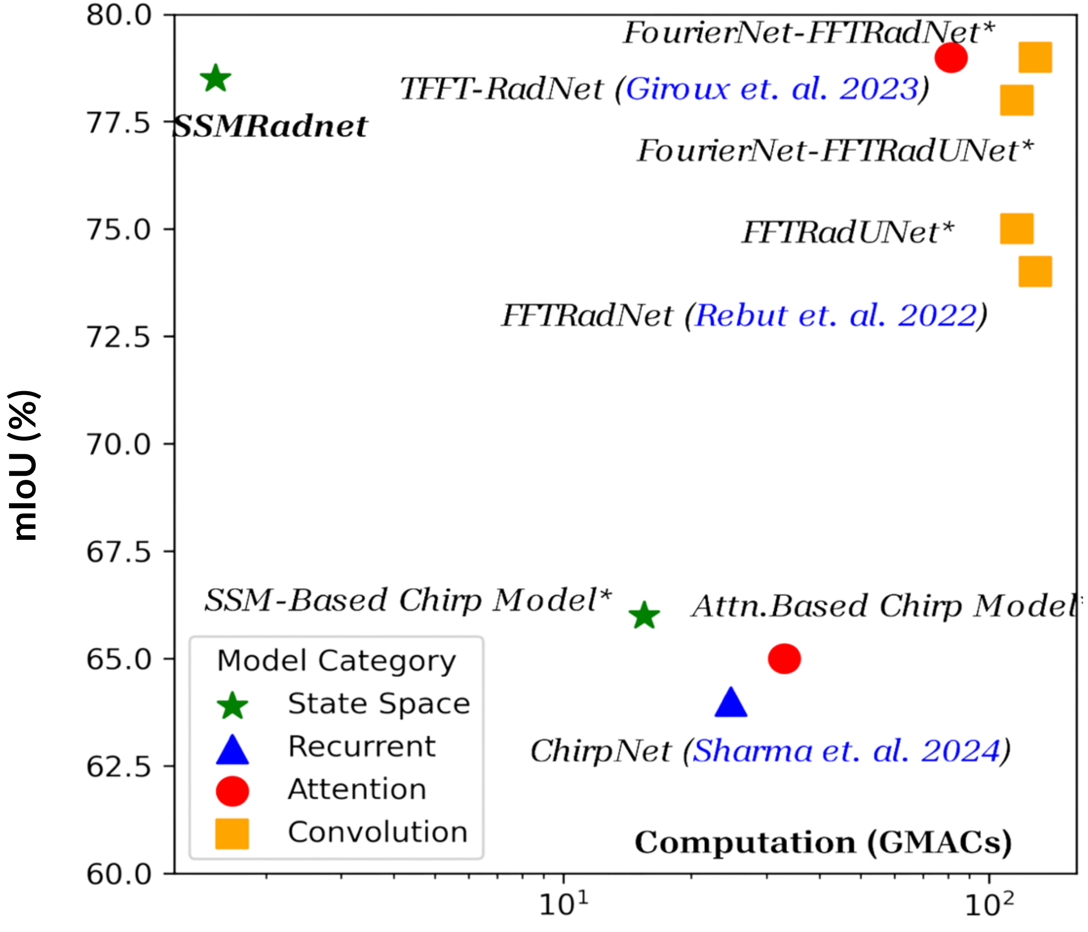
- Carrying SSM state (blue) yields smoother updates and slightly higher mean **IoU** 0.9258 vs 0.9253.
- Resets (red) show sharp dips at each new frame; retention stabilizes early in steady scenes.
- Abrupt change at **frame 7**: retention momentarily underperforms reset.

Take-away
State retention offers smoother, faster inference in steady scenes but may underperform during abrupt scene changes.

Future Take: Potential for adaptive inter-frame inference to save computation. (CVPR 2026)

Summary of key results

Computation vs Performance (RADIal Dataset)



Dataset and Training Parameters

Chirp, Sample, Receivers	*256, 512, 16*
Training Epochs	*200*
Batch Size	*8*
Optimizer	*Adam*
Learning Rate	1×10^{-4}
L2 Regularization	5×10^{-6}
Loss	*Jaccard, Focal + Smooth L1*

Quantitative Results

Performance on RADIal Dataset

Class	Model	Segmentation	Detection		
		mIoU	F1	mAP	mAR
Convolution	Pixor (PC) [31]	—	—	0.96	0.32
	Pixor (RA) [31]	—	—	0.96	0.82
	PolarNet [16]	0.61	—	—	—
	Conv3D + FFT-RadNet [29]	0.75	0.47	0.58	0.39
	FFT-RadNet [18]	0.74	—	0.97	0.82
	FFT-RadUNet[a]	0.75	0.80	0.83	0.77
	ADCNet [34]	0.79	0.89	0.93	0.86
	ADC UNet [34]	0.77	0.85	0.88	0.82
	ADC UNet (NPT) [34]	0.73	0.80	0.83	0.77
	FourierNet-FFT-RadUNet[b]	0.78	0.86	0.84	0.87
	FourierNet-FFT-RadNet[c]	0.79	0.88	0.87	0.89
Attention	Self Attention-based chirp model[d]	0.65	—	—	—
	TFFT-RADNet [6]	0.79	0.87	0.88	0.87
Recurrent	ChirpNet (GRU) [25]	0.64	—	—	—
SSM	SSM-based chirp model[e]	0.66	—	—	—
	SSMRadNet (Ours)	**0.79**	0.77	0.83	0.71

Performance on RADICal Dataset

Model	GMACs	Params (M)	Dice Coefficient (↑)	Chamfer (↓)
ChirpNet [25]	1.480	3.780	0.986	0.097
ChirpNetLite [25]	0.320	3.761	0.989	0.095
ChirpNet SSM[a]	0.340	3.761	0.990	0.088
ChirpNet-SelfAttn[b]	0.350	3.761	0.991	0.091
T-FFT-RadNet [6]	15.990	9.000	0.995	0.108
FFT-RadNet [18]	41.740	4.250	0.996	0.076
UNet [20]	15.140	17.270	0.996	0.078
SSMRadNet (Ours)	**0.108**	**0.566**	**0.996**	0.086

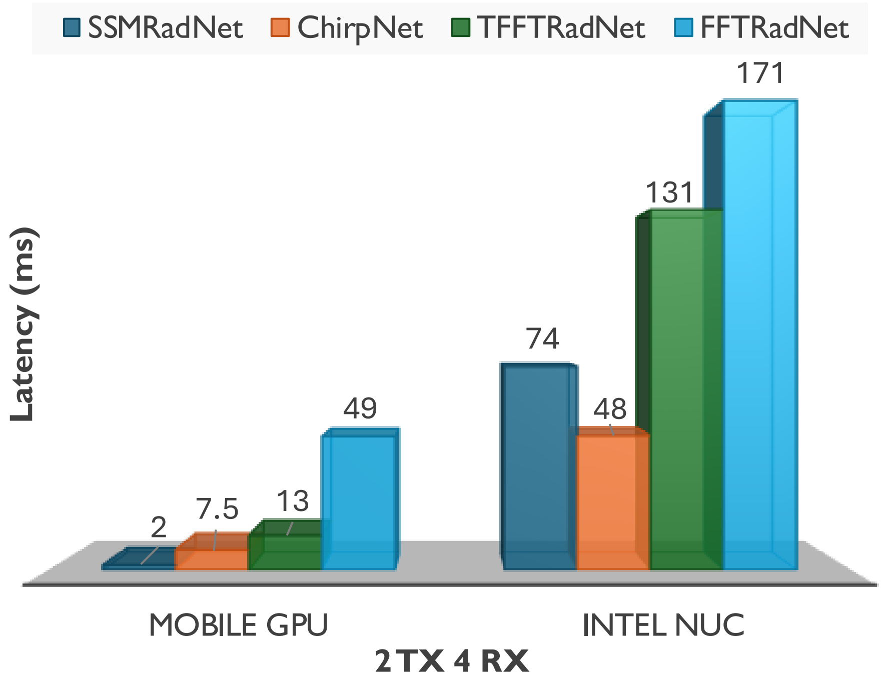
SSMRadNet achieves SoTA segmentation performance on both the RADICal and RADIal datasets at a fraction of the total parameters and computation of the existing SoTA methods.

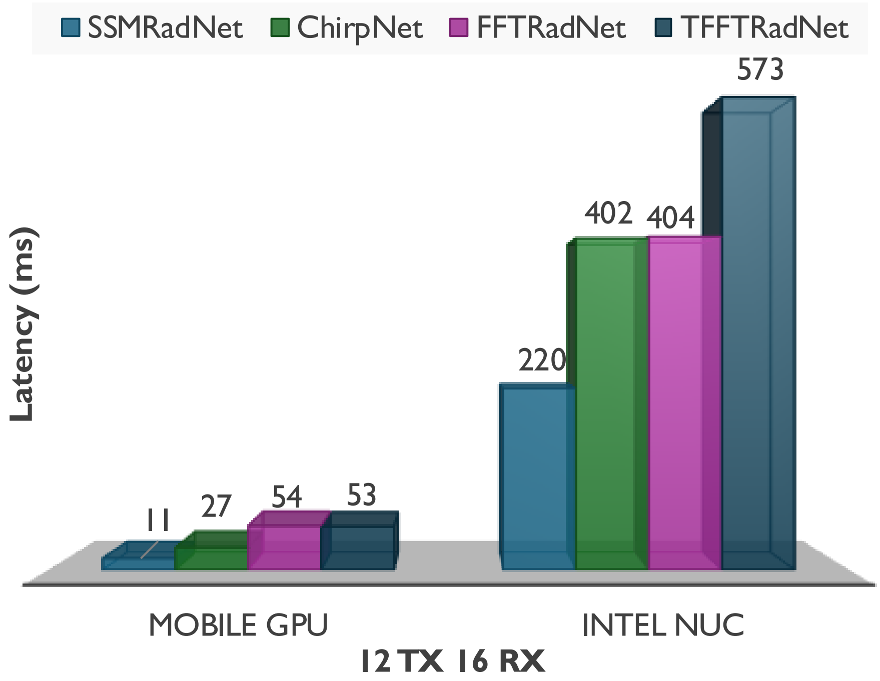
In the detection task, SSMRadNet achieves competitive scores, with a higher average recall value (more false positives) and needs further investigation for dynamic objects vs clutter identification.

SSMRadNet (S4 vs Mamba)

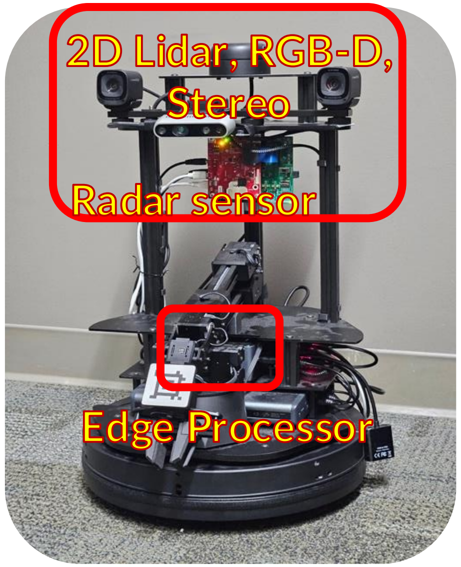
SSM Block	Detection Acc.
S4	0.76
Mamba	0.79

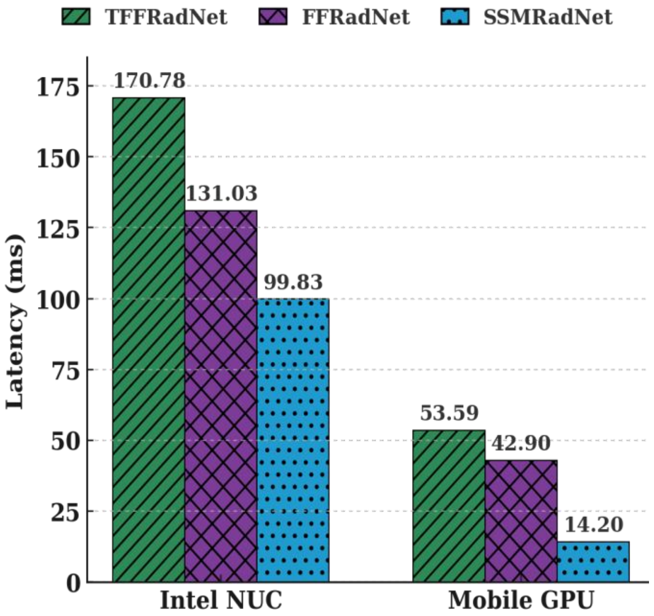
Note:
3% accuracy drop after we replace Mamba with S4. This shows further tuning is required and time varying selective SSM's model the ADC data better.





Runtime Characteristics on Edge





SSMRadNet achieves **1.7x** inference speed on our robot platform, and **3.8x** inference speed on mobile GPU compared to SoTA models

Mobile GPU:

- We characterize runtime on NVIDIA RTX 4060 mobile GPU.

Edge Hardware:

- Robot platform: Interbotix Locobot equipped with radar sensor
- Intel NUC compute unit running inference on Radar frame dims: (64, 192, 8)

Conclusion

- **SSMRadNet** achieves **33x** reduction in model complexity compared to SOTA baselines, TFFT- RadNet, and **FFTRadNet**. **88x** less computation and **3.5x** reduction in runtime.
- Our model can run on edge hardware focused on lightweight deployment while being energy efficient and achieve fast inference speeds for various applications.

Future Work:

- **SSMRadNet** achieves SoTA performance on the frees-pace segmentation task, but dynamic object detection requires further tuning and improvement.
- Retrieving Doppler information from learned representations to estimate the velocity of detected objects.
- Testing model performance under the influence of noise, clutter, and increasing efficiency through sub-sampling radar frames.

ACADEMIA • INDUSTRY • GOVERNMENT