Data Science and Artificial Intelligence

Threshold-Tunable U-Net for Small Target Detection in Maritime Radar: An Alternative to CFAR

Pepijn Lens

Supervisors (Leiden University):
Dr. P.W.H. van der Putten and Dr. D.M. Pelt

Supervisors (TNO):
MSc. Bas Jacobs and Dr. Giuseppe Papari

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

01/07/2025

Abstract

Radar-based object detection in maritime environments is a challenging task due to the presence of sea clutter and the low resolution of radar returns. Traditional detection algorithms such as Constant False Alarm Rate (CFAR) methods are widely deployed in operational radar systems because of their simplicity and real-time performance. However, their reliance on fixed clutter assumptions limits their effectiveness in dynamic sea states.

This thesis investigates an alternative approach based on deep learning, proposing a U-Net segmentation model trained on synthetic range-Doppler (RD) maps for the detection of small maritime targets. The U-Net is extended to process multiple consecutive frames as input channels, enabling temporal modeling of sea clutter dynamics. A key innovation is the interpretation of the U-Net output as a continuous confidence map, allowing real-time adjustment of detection sensitivity via threshold tuning, emulating CFAR's False Alarm Rate threshold.

The model was trained once on a diverse dataset with randomized clutter conditions and evaluated on thirteen separate datasets with varying SNRs and clutter types. Results show that the U-Net consistently outperforms CA-CFAR in both detection probability and false alarm rate, especially in the presence of sea clutter. Remarkably, the U-Net achieves a 3 ms inference time per sample, significantly faster than our CA-CFAR baseline (152 ms), though the latter may be suboptimally implemented.

This thesis demonstrates that segmentation-based detection with tunable thresholds offers a promising and high-performance alternative to traditional radar detection pipelines. An open-source Marimo application[1] accompanies this work, enabling reproducible research and further development.

[1] https://github.com/pepijn-lens/seacluttersuppression

Contents

1 Introduction

Radar systems play a critical role in navigation, surveillance, and autonomous operations. In maritime environments, radar-based object detection is vital for collision avoidance, situational awareness, and search-and-rescue operations. However, radar image interpretation is challenging due to sea clutter—non-target reflections from waves—and the typically low resolution of radar returns [1]. These effects make it especially difficult to detect small or low-signature targets, such as buoys, small vessels, drones or sea skimming missiles.

Traditional radar detection systems rely on constant false alarm rate (CFAR) techniques, which adaptively set detection thresholds based on local noise statistics [2]. CFAR is easy to implement and, thus, widely embedded into radar systems. However, it assumes relatively homogeneous clutter and does not generalize well to highly variable sea conditions. Moreover, it typically operates under fixed assumptions about clutter models, which limits its adaptability.

Recent advances in deep learning have opened up new possibilities for data-driven radar processing. Attention-based architectures such as Vision Transformers (ViT) [3], Detection Transformers (DETR) [4], and Swin Transformer [5] have achieved state-of-the-art results in many vision tasks by modeling global spatial dependencies. These models have also been applied to radar, particularly in automotive scenarios, with architectures like RadarFormer [6] and Mask-RadarNet [7] showing strong performance.

Early in this project, we experimented with Swin Transformers and CNNs for classifying RD-maps. However, Transformers performed poorly when introduced to sea clutter. Based on these limitations, we refocused our efforts on fully convolutional U-Net models [8], capable of pixel-wise prediction.

In this thesis, we propose a U-Net-based segmentation approach for detecting small maritime targets in RD-maps, even under heavy sea clutter. The model is trained once on a diverse synthetic dataset containing a broad range of sea states and SNRs. A key insight is that the U-Net outputs a continuous confidence map. By applying a tunable confidence threshold during inference, the number of detections, and thus the detection probability and false alarm rate, can be dynamically controlled.

This functionality mirrors the flexibility that made CFAR attractive, but with significantly better performance: a single trained U-Net outperforms CA-CFAR across almost all tested datasets.

To facilitate further research and reproducibility, we also release an interactive Marimo notebook[2] that enables users to simulate radar data, train U-Net models, and evaluate detection performance across custom scenarios.

Following the initial set of experiments, a follow-up study was conducted to further refine and validate the proposed U-Net-based detection method. This secondary experiment focused on improving detection performance at low signal-to-noise ratios (SNRs) by adjusting the loss function and recalibrating the threshold selection process to match specific false alarm rates. By training the model on a more challenging dataset, and systematically tuning the detection threshold, we were able to significantly enhance the trade-off between detection probability and false alarm rate. The results of this extended evaluation are presented in Section 4.6.

To summarize, our research offers the following contributions:

- We develop a configurable radar simulator capable of generating synthetic range-Doppler

[2] https://github.com/pepijn-lens/seaclutteruppression

maps with user-defined target and sea clutter characteristics.

- We propose a U-Net-based segmentation model trained on multi-frame RD inputs, and demonstrate that it generalizes across diverse sea clutter and SNR regimes.
- We show that adjusting the U-Net confidence threshold enables real-time control over false alarm rate, replicating and surpassing the functionality of CFAR.
- We compare U-Net and CA-CFAR detection performance on multiple datasets, showing that the U-Net achieves higher detection probabilities at lower false alarm rates.
- We release a fully interactive Marimo application[3] that enables users to reproduce our entire pipeline, from radar simulation to model training and evaluation.

The remainder of this thesis is organized as follows. Chapter 2 reviews relevant literature on radar detection techniques, deep learning models for radar image processing, and the limitations of existing Transformer-based methods in maritime contexts. Chapter 3 details the radar simulation pipeline, dataset generation process, U-Net model architecture, training procedure, and evaluation metrics. In Chapter 4, we present experimental results comparing the U-Net with classical CA-CFAR across varying SNR levels and clutter conditions. Chapter 5 discusses the implications of the results, the advantages and limitations of the proposed approach, and outlines directions for future work. Finally, Chapter 6 summarizes the main contributions and findings of this thesis.

2 Background and Related Work

This section provides an overview of foundational concepts and recent advancements relevant to this thesis, including the challenges of radar data interpretation in maritime environments, classical and deep learning-based detection techniques, and the evolving role of Transformers in radar perception. Special emphasis is placed on the limitations of current approaches for detecting small targets in sea clutter, motivating the use of U-Net-based segmentation.

2.1 Radar Data and Sea Clutter Challenges

Maritime radar systems produce range-Doppler (RD) maps (see Figure 1), which visualize object distance (range) and radial velocity (Doppler). A fundamental challenge in interpreting these maps is the presence of sea clutter, caused by waves reflections [9]. Sea clutter often masks small or low-SNR objects, such as small boats or buoys, complicating detection.

Traditional detection algorithms rely on constant false alarm rate (CFAR) techniques, including ordered-statistic CFAR (OS-CFAR) and cell-averaging CFAR (CA-CFAR), which dynamically set detection thresholds based on local statistics [2]. These methods are valued for their simplicity and efficiency, making them suitable for real-time deployment. However, their assumptions of homogeneity and Gaussian noise are often violated in dynamic maritime environments, leading to degraded performance. This limitation has prompted increasing interest in data-driven approaches.

[3] https://github.com/pepijn-lens/seaclutteruppression

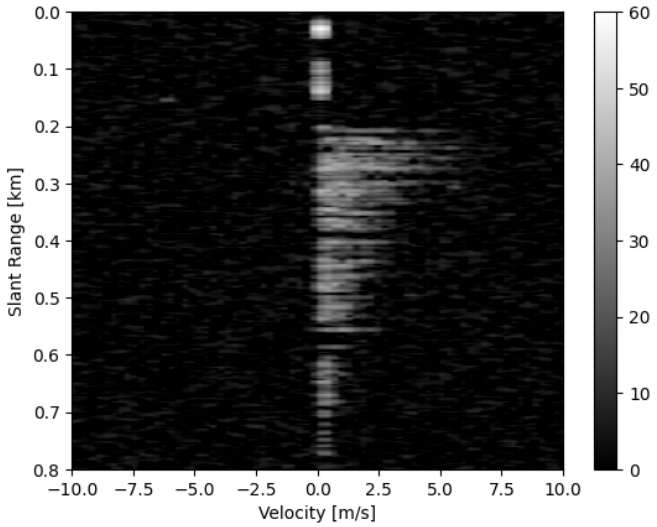


Figure 1: Example of a simulated range-Doppler map with both sea clutter and stationary clutter. The x-axis represents radial velocity, and the y-axis represents range. The signal at zero velocity and in the range of 0 to 150 meters represent dunes, bushes and other stationary clutter. The wave-like signals simulate sea clutter.

2.2 Deep Learning for Radar Image Processing

Deep learning has emerged as a powerful alternative to classical signal processing in radar perception. Convolutional neural networks (CNNs) have been applied to detect and classify targets in RD-maps, often outperforming CFAR under idealized conditions. Yavuz [10] demonstrated that CNNs can achieve better accuracy than CFAR in homogeneous L-band radar scenes. Li et al. [11] designed a dual-activation CNN to suppress sea clutter and detect small boats at low SNRs. More advanced architectures like Faster R-CNN have been adapted for airborne radar data. For example, Loran et al. [12] applied it to range-compressed X-band radar, enabling near-real-time ship detection without synthetic aperture radar (SAR) processing.

U-Net architectures are particularly well-suited for radar segmentation tasks, especially in low-resolution or cluttered settings. Originally introduced by Ronneberger et al. [8] for biomedical imaging, U-Net been used for radar signal processing in several studies. Zhai et al. [13] incorporated attention and multi-scale fusion mechanisms into U-Net to enhance sea clutter suppression and deployed the model on FPGA hardware for real-time use. Wang et al. [14] proposed a complex-valued U-Net that processes both amplitude and phase. Other studies, such as Zhang and Fan [15] and Cao et al. [16], have explored hybrid architectures combining U-Net and Vision Transformers for semantic segmentation of radar and medical imagery.

2.3 Transformers in Visual and Radar Domains

Transformers, first introduced by Vaswani et al. [17] for NLP, have been successfully adapted to vision tasks. Vision Transformers (ViTs) [3] learn spatial relationships via self-attention over image patches, and DETR [4] extended them to object detection. Swin Transformer [5] introduced a hierarchical architecture with local attention windows, improving scalability and localization.

In radar perception, Transformer-based models are a recent and growing area. Dalbah et al. [6] introduced RadarFormer, a lightweight model for automotive radar that achieves competitive performance on the CRUW dataset with reduced computational cost. Wu et al. [7] proposed Mask-RadarNet, which augments Transformer blocks with masked attention and patch-shift modules for improved spatio-temporal modeling in radar sequences. While these approaches are promising, they are largely tuned for automotive environments and do not explicitly address the unique statistical and spectral properties of maritime sea clutter.

2.4 Limitations of Transformers for Small Target Detections

In a pre-study, we implemented a Swin Transformer for a simple classification task on clutter-free range-Doppler maps to evaluate the potential of Vision Transformers (ViTs) for radar-based detection. Although the model achieved high accuracy, it was consistently outperformed by a basic CNN. The limitations became more evident when training on cluttered data, where performance degraded rapidly.

A key issue lies in the granularity of ViT patch tokens. With targets spanning just a few pixels, a small object may be encapsulated within a single token, nullifying the benefit of attention across tokens. Without multi-scale design or convolutional priors, standard ViTs struggle to detect small or noise-like targets [18].

Following feedback from domain experts at TNO, we reframed the problem as a segmentation task: identifying only target regions, with clutter treated as background. This shift leverages U-Net's pixel-level precision and allows the model to suppress sea clutter without needing to explicitly label it. This approach aligns with trends in radar and remote sensing literature, where segmentation offers improved performance for small, diffuse targets in noisy environments [13, 15].

2.5 Research Gap

Although deep learning has made considerable progress in radar-based object detection, traditional techniques such as constant false alarm rate (CFAR) detection remain the dominant approach in operational radar systems. CFAR is attractive due to its real-time performance and tunability via the false alarm rate parameter. In contrast, deep learning models, while often achieving better performance in controlled studies, are not yet widely adopted in embedded or operational maritime radar systems.

Among deep learning models, U-Net has emerged as a strong candidate for radar segmentation, especially in the presence of sea clutter [13, 14, 15]. However, most existing U-Net implementations process single-frame radar inputs and do not explicitly incorporate temporal information, despite the potential benefits of capturing wave dynamics for target discrimination. In this thesis, we explore a U-Net-based approach that uses multiple consecutive radar frames as input channels to model short-term temporal context.

Moreover, while threshold tuning in this context has been applied in CNNs [10], it has not been systematically explored for segmentation-based radar detectors. In this work, we treat the U-Net's output as a confidence map and demonstrate that varying the threshold allows real-time control over the detection probability and false alarm rate.

By combining multi-frame temporal modeling, segmentation-based detection, and post-training threshold tuning, this thesis provides a novel and highly effective alternative to CFAR.

3 Methodology

This chapter describes the simulation pipeline, dataset construction, model architecture, training procedure, and evaluation metrics used to investigate deep learning-based object detection in radar data.

3.1 Overview

The aim of this research is to evaluate the effectiveness of deep learning models—particularly the U-Net architecture—for detecting small targets in range-Doppler (RD) maps, both in ideal conditions and in the presence of sea clutter. While early explorations included Swin Transformers on classification tasks, these approaches were not pursued further due to poor generalization in cluttered scenarios. The final model is a U-Net trained to segment targets directly from RD-maps, with a classical CA-CFAR method used as a baseline for comparison. Detection performance is quantified using recall (detection probability) and false alarm rate.

3.2 Radar Simulator

The radar simulator is designed to generate synthetic range-Doppler maps with configurable sea clutter and target parameters. By setting the clutter power to a low value (e.g., -20 dB), the simulator can also produce clutter-free data. The sea clutter model is based on a compound-Gaussian (K-distribution) framework with correlated texture and speckle processes, Doppler spectral shaping, and additive thermal noise [19].

3.2.1 Texture Generation

The large-scale modulation (texture) of the clutter is modeled as a unit-mean Gamma-distributed random field:

$$T(r, p) \sim \text{Gamma}(k, 1/k), \tag{1}$$

where r indexes range bins, p indexes pulses, and k is the shape parameter. This approach follows standard K-distributed clutter modeling.

3.2.2 Speckle Generation

Speckle is generated as complex Gaussian white noise passed through an AR(1) process along the slow-time (pulse) axis:

$$x_{r,0} = \frac{w_{r,0}}{\sqrt{1-\alpha^2}}, \quad x_{r,p} = \alpha x_{r,p-1} + w_{r,p}, \tag{2}$$

where $w_{r,p} \sim \mathcal{CN}(0, 1)$. The result is normalized to have unit average power, capturing Doppler correlation consistent with real sea clutter.

3.2.3 Bragg Peak Injection

To simulate Bragg scattering from capillary waves, symmetric Doppler peaks at $\pm f_0$ are injected into the RD-map using twin Gaussian filters. However, in our final experiments, Bragg lines were

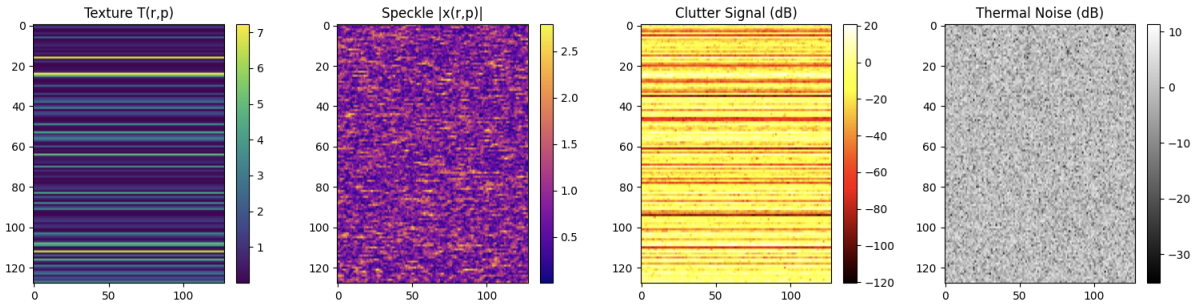


Figure 2: These plots visualize the sea clutter simulation pipeline. The main components are texture (K-distribution), speckle (AR(1) process) and thermal noise (Gaussian).

excluded to better match the characteristics of AMBER data available at TNO, which does not exhibit strong Bragg components.

3.2.4 Amplitude Scaling and Noise Addition

Clutter power P_c is computed based on the desired clutter-to-noise ratio (CNR) per pulse:

$$P_c = \left(10^{\frac{\text{CNR}_{\text{dB}}}{10}} / N_p\right) \cdot P_{\text{noise}}, \tag{3}$$

where N_p is the number of pulses. The resulting complex clutter signal is given by:

$$C(r,p) = A \cdot T(r,p) \cdot x_{r,p} + n(r,p), \tag{4}$$

with $A = \sqrt{P_c}$ and $n(r,p)$ denoting additive complex Gaussian noise (see Figure 2).

3.2.5 Target Blob Insertion

Targets are inserted as coherent blobs centered at range r_0, Doppler-shifted by frequency ν, and spanning a configurable number of range bins. The amplitude corresponds to a specified signal-to-noise ratio (SNR). Outer bins are attenuated (e.g., 70% of the center value) to simulate side lobes, a common feature in radar data.

3.2.6 Range-Doppler Map Generation

A Hamming window is applied across the slow-time axis prior to FFT-based Doppler processing. The FFT output is adjusted for power restoration and Doppler shift alignment, yielding a simulated RD-map that reflects moving clutter and target dynamics. To simulate the motion of the waves, we simply roll the texture. In between frames, new target positions are calculated and added.

3.3 Synthetic Data Generation

Thirteen datasets were generated for training and evaluation. All datasets include thermal noise with a 1 dB power level and a frame rate of 2 Hz. Each sample contains three consecutive RD-frames with spatial dimensions of 128×128 (Doppler $\times$ range bins).

Clutter-Free Datasets: Six datasets were created without sea clutter, each using a different fixed SNR value for targets: 6 dB, 8 dB, 10 dB, 12 dB, 16 dB, and 20 dB (see Figure 3a). Each dataset contains 5,500 samples.

Cluttered Datasets: Six additional datasets were created with sea clutter present (see Figure 3b). The target SNRs are the same as in the clutter-free datasets. Clutter parameters are as follows: mean clutter power = 16 dB, shape parameter $k = 0.75$, AR coefficient $\alpha = 0.9$, and wave speed = 4.0 m/s. A mean clutter power of 16 dB SNR means that the peaks of sea clutter have a higher SNR than 16 dB, making sure targets always have a lower signal power than clutter peaks. Each dataset also contains 5,500 samples.

Generalization Dataset: To evaluate the model's capability to generalize across different sea states, a thirteenth dataset was created with randomly varying parameters per sample:

- Target SNR: uniformly sampled between 10–20 dB
- Clutter power: uniformly sampled from 14–18 dB
- Shape parameter: 0.5–1.0
- AR coefficient: 0.8–0.99
- Wave speed: $[2.0, 6.0]$ m/s

This dataset contains 10,000 samples.

3.4 CA-CFAR Baseline

As a classical baseline, we implemented a 2D Cell-Averaging CFAR (CA-CFAR) detector. The algorithm uses a rectangular sliding window around each cell-under-test (CUT), with one guard cell and four training cells in both range and Doppler directions (see Figure 4). The detection threshold is determined from the average power of the surrounding training cells and scaled by a factor derived from the desired false alarm probability:

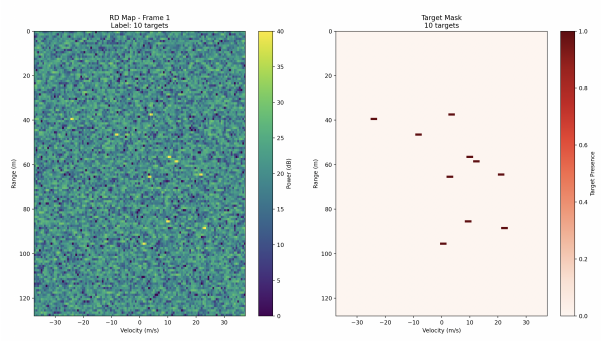
$$\alpha = N \cdot \left(\mathrm{P}_{\mathrm{FA}}^{-1/N} - 1 \right), \tag{5}$$

where N is the number of training cells. The final threshold is applied to the squared magnitude of the input signal (converted from dB to linear scale). We evaluated CA-CFAR on the same datasets as the U-Net, with false alarm rates set to $\mathrm{P}_{\mathrm{FA}} = 10^{-3}$ and 10^{-4}.

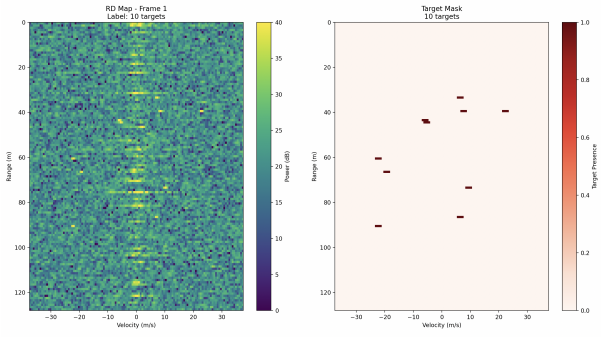
3.5 U-Net Architecture

To segment small targets in RD-maps, we implemented a U-Net architecture. The model follows an encoder–decoder structure with skip connections, enabling the preservation of spatial detail during upsampling.

The encoder consists of three convolutional blocks (`DoubleConv`) using 3×3 convolutions and ReLU activations, followed by max pooling. The decoder uses transposed convolutions for



(a) Without sea clutter



(b) With sea clutter

Figure 3: Examples of synthetic RD-maps used in this study. Targets in these samples have an SNR of 20 dB and positions are saved in the target masks on the right. The range and radical velocity of the targets can be seen on the y-axis and x-axis respectively.

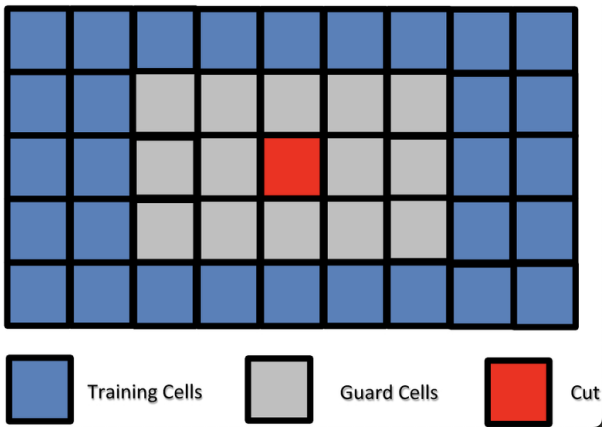


Figure 4: Visualization of the CA-CFAR algorithm, illustrating the training cells, guard cells, and the cell under test (CUT). The decision on whether the CUT represents a target is based on a comparison between its value and a threshold computed from the surrounding training cells.

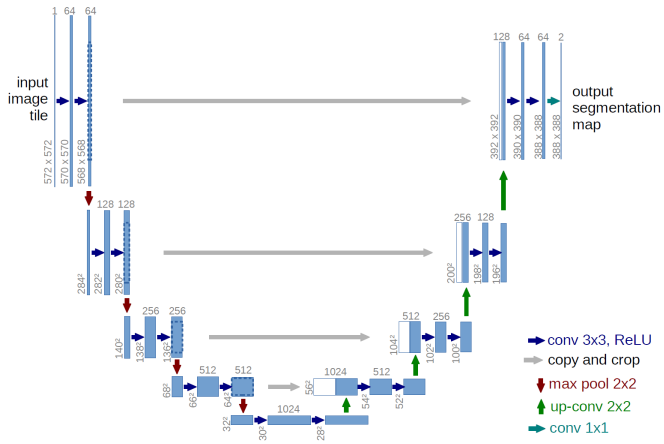


Figure 5: The U-Net architecture, which follows the original paper [8].

upsampling, concatenates the corresponding encoder features, and applies additional convolutions to refine predictions.

The architecture follows the original U-Net filter configuration with 3 input channels (temporal frames), 1 output channel and 64 base filters. The U-Net has 1.862.849 trainable parameters. The number of parameters depends on the number of base filters.

3.6 Training Details

The model was trained using a weighted combination of Binary Cross-Entropy (BCE) and Tversky loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{BCE}} \cdot \mathcal{L}_{\text{BCE}} + \lambda_{\text{Tversky}} \cdot \mathcal{L}_{\text{Tversky}} \tag{6}$$

We used weights of $\lambda_{\text{BCE}} = 0.1$ and $\lambda_{\text{Tversky}} = 0.9$, prioritizing control over class imbalance and sensitivity to false negatives.

The Tversky loss generalizes the Dice loss by introducing tunable penalties for false positives and false negatives. In this work, we set the Tversky loss parameters to $\alpha = 0.2$ and $\beta = 0.8$, emphasizing recall by assigning a higher penalty to false negatives than to false positives. This aligns with the radar detection objective, where missing a true target is typically more critical than triggering a false alarm. The equation:

$$\mathcal{L}_{\text{Tversky}} = 1 - \frac{\sum_i y_i \hat{y}_i}{\sum_i y_i \hat{y}_i + \alpha \sum_i (1 - y_i)\hat{y}_i + \beta \sum_i y_i (1 - \hat{y}_i)} \tag{7}$$

Here, y_i is the ground truth label, $\hat{y}\_i$ is the predicted probability, α controls the penalty for false positives, and β controls the penalty for false negatives.

Optimization was performed using the Adam optimizer with an initial learning rate of 1×10^{-4}, scheduled via cosine decay over the course of training. A batch size of 32 was used. Early stopping with a patience of 10 epochs was applied based on the validation loss $\mathcal{L}_{\text{val}}$. The model was trained with a maximum of 100 epochs but was stopped early after 43 epochs. Training was conducted on an Apple M4 MacBook Pro equipped with a 20-core GPU and 24 GB RAM.

3.7 Evaluation Metrics

During inference, the model produces a continuous confidence map representing the likelihood of each pixel belonging to a target. To convert this into binary detections, two different confidence thresholds are applied: 0.1 and 1×10^{-6}. These thresholds allow us to investigate the trade-off between detection sensitivity and false alarm rate.

After thresholding, connected regions are clustered using DBSCAN with parameters `eps = 1.0` and `min_samples = 1`. A minimum cluster size of one pixel is enforced to ensure that even small target responses are preserved.

Detected clusters are matched to ground truth centroids based on Euclidean distance, with a maximum matching radius of 1.5 pixels. Final detection performance is quantified using:

- **Detection Probability (P_D)**: The fraction of true targets that are successfully detected.
- **False Alarm Rate (P_{FA})**: The fraction of predicted detections that do not correspond to any ground truth target.

4 Experiments and Results

This section presents the evaluation of the U-Net model for small target detection in range-Doppler maps and compares its performance against the classical CA-CFAR baseline. The experiments focus on recall (detection probability) as a function of target SNR, under both clutter-free and cluttered conditions.

To evaluate generalization capabilities, the U-Net was trained exclusively on the final dataset containing randomized sea clutter configurations, as described in Section 3.3. The dataset was split into training, validation, and test sets with ratios of 70%, 15%, and 15%, respectively. The trained model was then evaluated on the thirteen datasets described in Section 3.3, each designed with distinct SNR levels and clutter conditions.

4.1 Detection Probability Across SNRs

Figure 6 shows detection probability (P_D) versus SNR for four methods: U-Net with thresholds of 10^{-1} and 10^{-6}, and CA-CFAR with P_{FA} of 10^{-4} and 10^{-3}.

In the clutter-free datasets (Figure 6a), both the U-Net at threshold 10^{-6} and CFAR at $P_{FA} = 10^{-3}$ show strong performance across all SNRs, with the U-Net slightly outperforming CFAR on average. Lower-threshold U-Net and high-P_{FA} CFAR outperform their stricter counterparts, highlighting the classic trade-off between sensitivity and false alarms.

In the presence of sea clutter (Figure 6b), the same trend holds: the U-Net with threshold 10^{-6} performs comparably or better than CFAR at $P_{FA} = 10^{-3}$, especially at SNRs above 8 dB. The lower performance at 6 and 8 dB SNR may be explained by the fact that the U-Net was trained only on targets with SNR $\geq$ 10 dB 3.3.

4.2 False Alarm Rate Comparison

As shown in Figure 7a, CFAR achieves its expected false alarm rates in clutter-free scenarios: approximately 1×10^{-4} and 1×10^{-3} for the two P_{FA} settings. The U-Net also remains consistent,

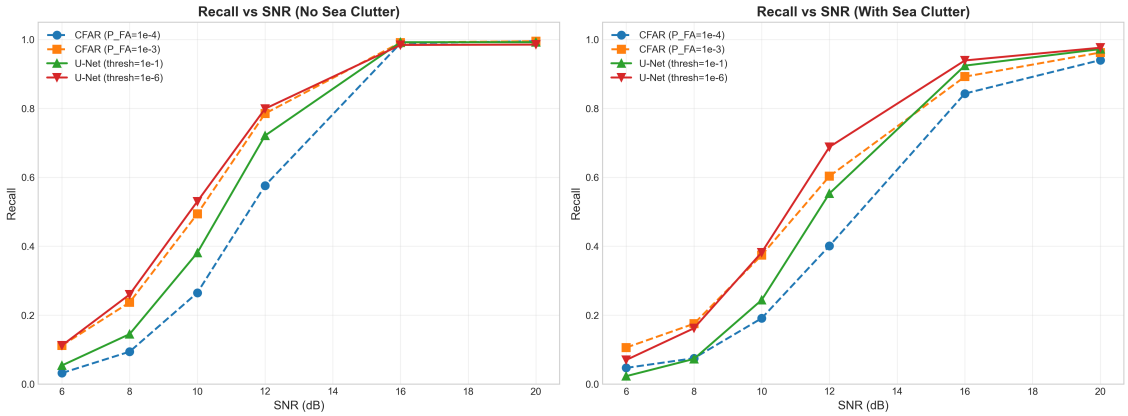


Figure 6: Detection probability of U-Net and CA-CFAR across different SNR levels. The left plot (a) shows results on datasets **without** sea clutter, and the right plot (b) shows results **with** sea clutter.

producing false alarm rates of roughly 2×10^{-4} and 7.8×10^{-4} for thresholds 10^{-1} and 10^{-6} respectively.

In cluttered scenarios, CFAR false alarm rates increase substantially, to approximately 1.6×10^{-3} and 3.4×10^{-3}, due to interference from sea clutter. In contrast, the U-Net maintains surprisingly low false alarm rates: 8.4×10^{-5} and $6\text{–}7 \times 10^{-4}$, depending on threshold. This may be attributed to the model being trained on a diverse set of clutter configurations, enabling it to generalize well.

4.3 Generalization on Mixed-Clutter Dataset

Figure 7b presents the results on the final dataset with randomized sea clutter conditions and variable SNR. Here, the U-Net clearly outperforms both CFAR configurations, achieving higher detection probabilities while simultaneously maintaining lower false alarm rates. This highlights the model's ability to generalize across unseen clutter and target settings.

4.4 Visual Comparison: CFAR vs. U-Net

To illustrate the difference between CFAR and U-Net performance, we examine Figures 8 and 9, which show predictions on a sample from the 16 dB SNR dataset with sea clutter.

The U-Net successfully detects all three targets while producing only a single false alarm. In contrast, CFAR detects only two out of the three targets and generates numerous false positives. These visual results are consistent with the findings in Figure 6, where the U-Net outperforms CFAR in terms of detection probability on cluttered datasets at 16 dB SNR.

In a more challenging scenario, the targets are heavily obscured by sea clutter, as shown in Figure 10a. In this case, the U-Net with a confidence threshold of 10^{-1} struggles to detect the occluded targets and underperforms compared to CFAR configured with $P_{FA} = 10^{-3}$. However, when the U-Net threshold is lowered to 10^{-6}, it becomes more sensitive, resulting in additional detections, but also more false alarms. This illustrates the trade-off between sensitivity and specificity in threshold tuning. Figures 10 and 11 visualize these differences in detection behavior.

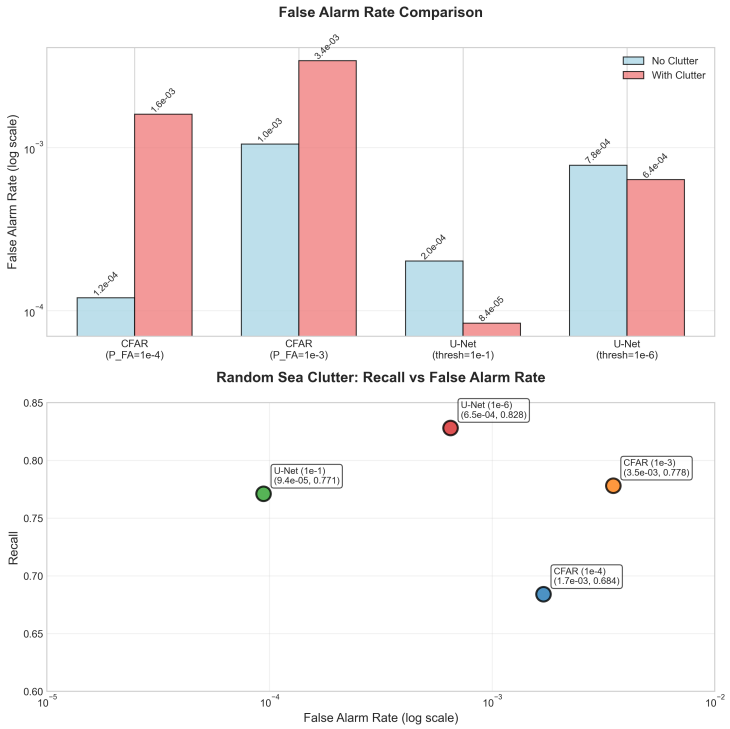
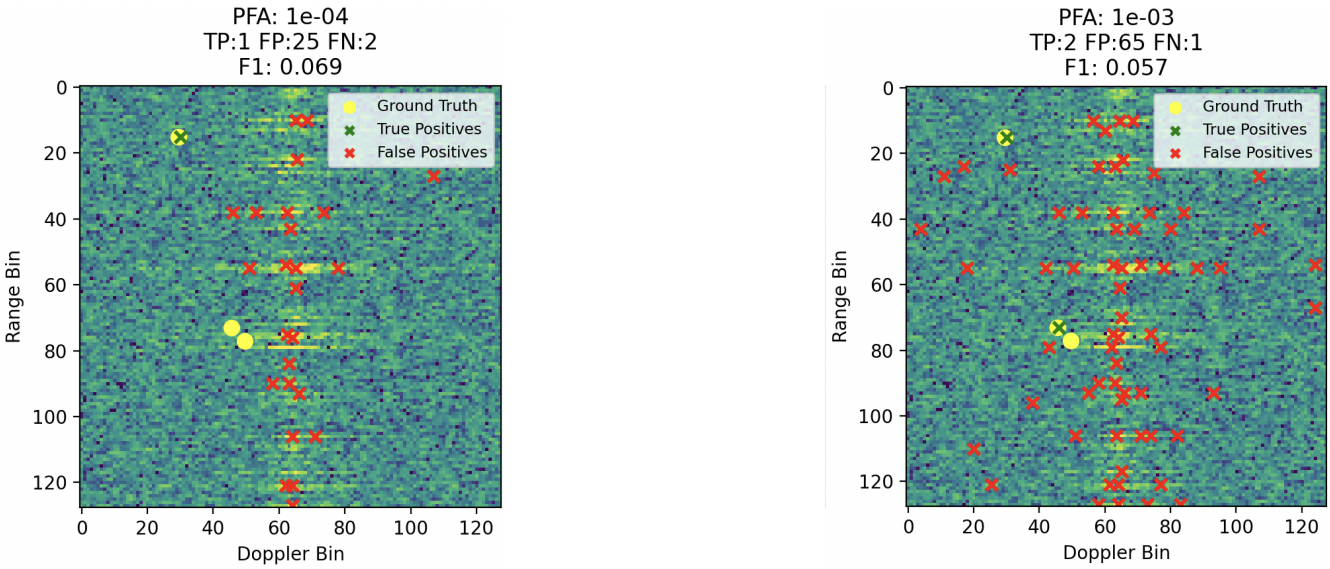
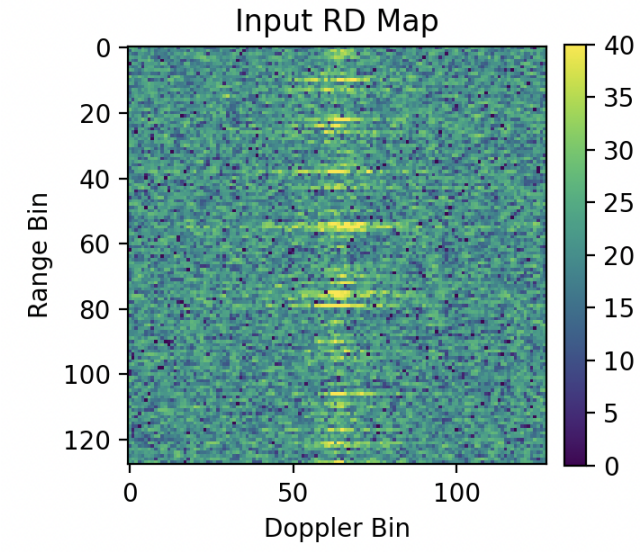


Figure 7: Upper plot (a): A histogram showing the average false alarm rate the U-Net and CFAR for different thresholds, measured from the 12 datasets 3.3. Lower plot (b): The detection probability (recall) plotted against the false alarm rate of for all method configurations, measured on the thirteenth dataset 3.3.

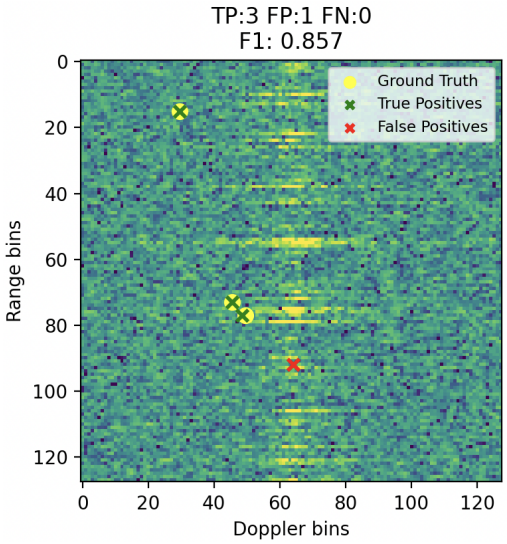


(a) $P_{FA} = 10^{-4}$ (b) $P_{FA} = 10^{-3}$

Figure 8: Example outputs of the CFAR detector at two different false alarm rate settings.

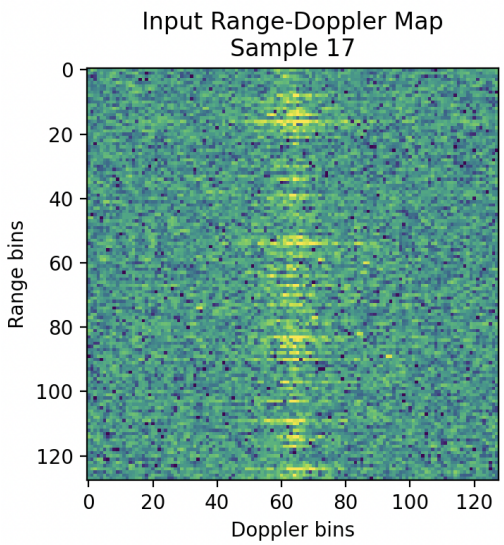


(a) U-Net input (frame 3 of RD sequence)

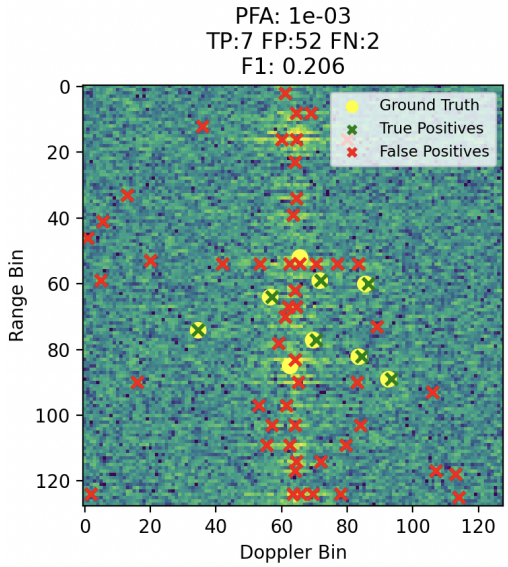


(b) U-Net output with threshold 10^{-1}

Figure 9: Example of U-Net inference on a sample from the 16SNR dataset with clutter. The left plot shows the last frame of a sample, and the right shows the predicted segmentation mask with the threshold at 10^{-1}.

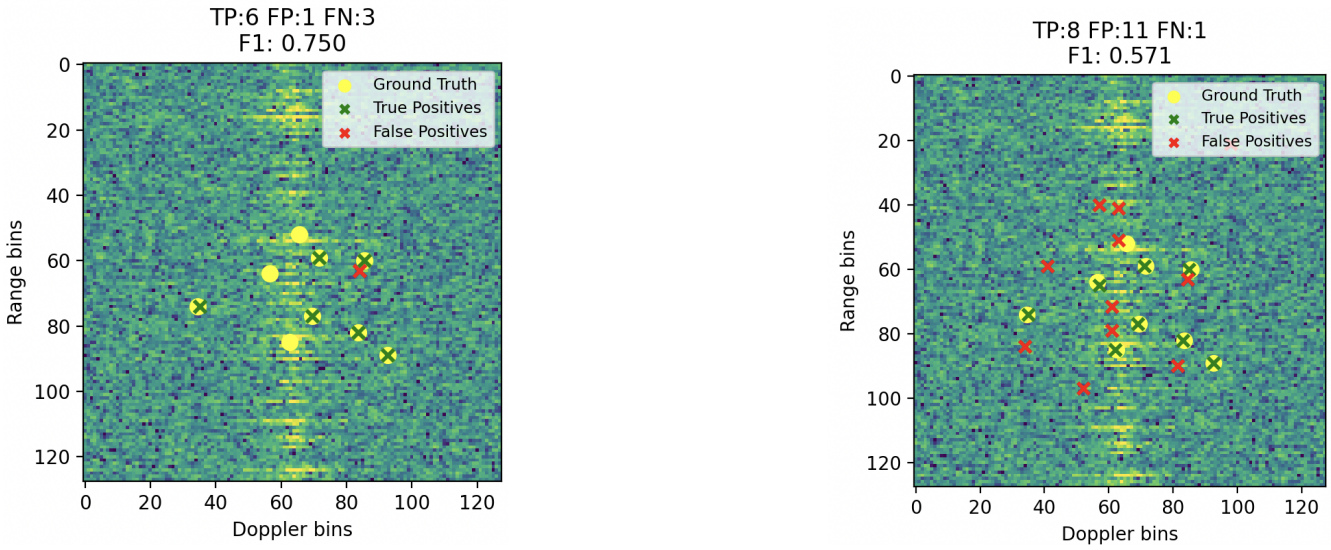


(a) Input RD-map with heavy clutter



(b) CFAR output at $P_{FA} = 10^{-3}$

Figure 10: Comparison between input image and CFAR detections under high clutter conditions.



(a) U-Net output at threshold 10^{-1}

(b) U-Net output at threshold 10^{-6}

Figure 11: U-Net detection results under heavy clutter. Lowering the threshold increases sensitivity but also introduces false alarms.

4.5 Runtime Performance

Beyond detection accuracy, computational efficiency is crucial for practical deployment in maritime systems. The trained U-Net achieves an average inference time of approximately 3 milliseconds per sample on an Apple M4 MacBook Pro equipped with a 20-core GPU. This runtime demonstrates that the model is capable of near-real-time performance, even on consumer-grade hardware.

In contrast, the CA-CFAR implementation used in this study required roughly 152 milliseconds per sample. However, it should be noted that this version was written in Python without specific optimization for speed. More efficient implementations (e.g., in C++ or on dedicated DSPs) could significantly reduce CFAR's runtime. Nevertheless, the observed gap illustrates that modern deep learning models can be not only more accurate but also faster.

4.6 Follow-Up Experiment: Improved Threshold-tuning

After the initial submission of this thesis, a follow-up experiment was conducted to further validate and improve the performance of the proposed U-Net model. While the original experiments demonstrated strong results, this second study investigates whether detection performance for low-SNR targets could be improved by adjusting the U-Net's confidence threshold more systematically.

In the initial experiment, the confidence threshold was selected independently of the false alarm rate. This approach resulted in Figure 6, where the U-Net and CFAR implementations were compared. However, such a comparison is more meaningful when both methods operate at comparable false alarm rates.

To enable this, we manually computed the U-Net threshold required to achieve a specific false alarm rate. The U-Net outputs a confidence map of size 128×128. By inputting 10,000 noise-only (newly generated without clutter or targets) images and counting the number of pixels exceeding

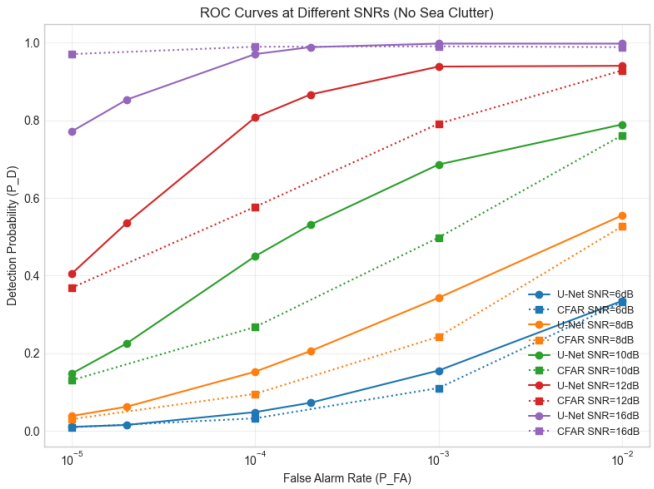


Figure 12: ROC curves comparing U-Net and CFAR for false alarm rates of 10^{-5}, 2×10^{-5}, 10^{-4}, 2×10^{-4}, 10^{-3}, and 10^{-2}, across SNR levels of 6, 8, 10, 12, and 16. The y-axis represents detection probability; the x-axis shows the false alarm rate.

a given threshold T, we estimated the threshold needed to reach a desired false alarm rate. For example, to achieve a false alarm rate of 10^{-4}, the number of false-positive pixels should be:

$$128 \times 128 \times 10000 \times 10^{-4} = 16384. \tag{8}$$

This method enabled a more accurate and fair comparison against CFAR.

To further improve recall, particularly at low SNRs, we constructed a new training dataset consisting of 10,000 samples. Each sample contained 20 targets with SNRs ranging from 8 to 12 dB. This higher target density was intended to help the model better learn target characteristics and become more sensitive to weaker signals. In addition, the mean clutter level was randomly varied between 0 and 16 dB.

Training a new model was necessary because the original one exhibited a sharp confidence gap—either predicting targets with very high confidence or not at all—making it difficult to finely adjust the threshold. Increasing the number of false alarms was thus challenging.

To address this, we modified the loss function. The original model used a combination of Binary Cross-Entropy and Tversky loss. In the follow-up experiment, we used only Tversky loss, with parameters $\alpha = 0.1$ and $\beta = 0.9$, thereby emphasizing recall and reducing the penalty for false positives. This made the model more sensitive and better suited for threshold tuning.

4.7 Results of the Additional Experiment

The new model trained for 43 epochs on the augmented dataset and performed well across all datasets used in the initial experiment. The ability to explicitly calculate thresholds corresponding to desired false alarm rates enabled us to plot Receiver Operating Characteristic (ROC) curves (see Figure 12) for the U-Net and CFAR.

In Figure 12, we observe that the U-Net achieves higher detection probabilities than CFAR across all datasets, except for the 16 dB SNR set at the lowest false alarm rates. This may be

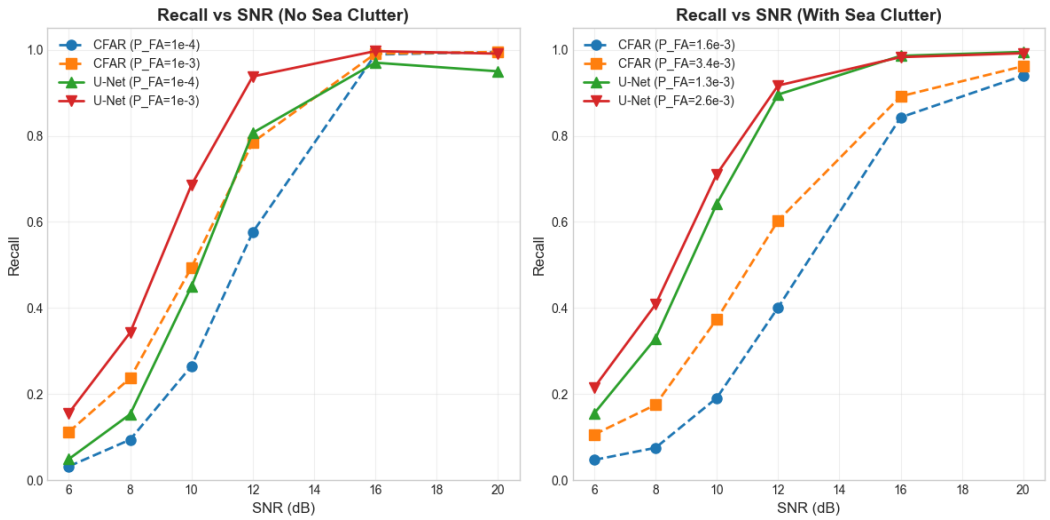


Figure 13: Detection probability versus SNR using the new model and thresholding method. Left: datasets without sea clutter. Right: datasets with sea clutter.

due to the model being trained only on targets with SNRs between 8 and 12 dB. However, for false alarm rates of 10^{-4} and higher, the model successfully detects higher-SNR targets as well, suggesting that training on low-SNR data generalizes well to higher-SNR conditions.

To improve the interpretability of the original results (Figure 6), we applied the new thresholding method to align the U-Net's false alarm rates more closely with those of CFAR.

From Figure 13, we see that in clutter-free datasets, the U-Net and CFAR achieve closely matched false alarm rates. This is expected, as thresholding on Gaussian noise is predictable and consistent. For cluttered datasets, estimating the correct threshold is more difficult due to variability in clutter characteristics. Nevertheless, the U-Net closely approaches the CFAR false alarm rates: approximately 1.3×10^{-3} and 2.6×10^{-3} compared to CFAR's 1.6×10^{-3} and 3.4×10^{-3}, respectively. Importantly, the U-Net achieves significantly higher detection probabilities—by 20 to 25 percentage points—on targets with SNRs between 8 and 12 dB.

These results demonstrate that the modifications to the training data and loss function substantially improve the recall/false alarm trade-off, validating the benefits of the proposed refinements.

5 Discussion and Future Work

This chapter discusses the key findings of the proposed U-Net approach in comparison to classical methods, highlighting its strengths and limitations.

5.1 Reflection on Results

The experiments demonstrate that a U-Net trained on synthetic radar data with randomized sea clutter conditions can generalize well across a range of unseen SNRs and clutter regimes. Most notably, the U-Net achieves higher detection probabilities and lower false alarm rates than classical CA-CFAR detectors across all tested datasets when a low confidence threshold is applied.

This is particularly evident in cluttered scenarios, where CFAR performance degrades due to its assumptions about clutter homogeneity, while the U-Net learns the clutter patterns and suppresses them.

Importantly, the ability to tune the U-Net's confidence threshold at inference time provides flexibility comparable to CFAR's configurable P_{FA}. This means that the segmentation-based detector can be calibrated post-training for operational trade-offs between sensitivity and reliability, without retraining the model. This decouples training from deployment, a practical advantage in real-world systems where operating conditions may vary significantly.

The success of the multi-frame U-Net architecture also validates the hypothesis that short-term temporal information aids in clutter suppression. By processing multiple consecutive frames as channels, the network appears to learn motion-based priors that improve its ability to distinguish moving targets from dynamic background clutter.

5.2 Known Limitations

Despite these promising findings, several limitations must be acknowledged. The most significant is the reliance on synthetic radar data. While the simulator is designed to emulate realistic clutter and target dynamics, it may not capture the full complexity of measured maritime radar data. This includes effects like hardware artifacts and complex sea-state transitions, which may impact generalization in real-world deployments.

Lastly, the current evaluation is limited to RD-maps. The absence of azimuth limits the network's ability to distinguish overlapping targets, which could be critical in multi-object tracking.

5.3 Future Work

Several avenues for further research follow naturally from this work:

- **Evaluation on real-world maritime radar data:** A crucial next step is to validate the models developed here on measured radar data. Real-world testing will reveal the extent to which the segmentation-based pipeline generalizes beyond the controlled synthetic domain.
- **Incorporating angular information (azimuth):** Extending the model input to include azimuth alongside range, Doppler, and time would enable richer spatial context and help distinguish overlapping or occluded targets. This 4D representation could improve detection accuracy.
- **Complex-valued U-Net:** Inspired by Wang et al. [14], integrating both amplitude and phase into a complex-valued U-Net could improve clutter suppression in non-homogeneous sea states. Complex-valued convolutions would allow the model to leverage additional signal structure not accessible to real-valued networks.
- **Hybrid Swin-U-Net architectures:** Given the promising results of Swin Transformers in visual tasks and recent hybrid models such as Swin-Unet [16], it would be valuable to explore a segmentation network that combines hierarchical attention with U-Net's skip-connected decoder. This could offer improved multi-scale understanding while maintaining spatial precision.

- **Longer temporal context:** While three consecutive RD-frames already improved detection performance, future work could explore the trade-offs involved in using longer temporal sequences. This may help the model learn more robust motion-based priors for differentiating between persistent clutter and true targets, potentially further reducing false negatives.

6 Conclusion

This thesis explored a deep learning-based approach for small target detection in maritime radar imagery, using range-Doppler (RD) maps. While classical methods such as Constant False Alarm Rate (CFAR) remain widely used due to their simplicity and real-time capabilities, they can struggle in environments with highly variable sea clutter. In response, we proposed a U-Net-based segmentation model trained on synthetic radar data with diverse clutter conditions. The model outputs a continuous confidence map, enabling post-training threshold tuning to control the trade-off between detection probability and false alarm rate.

Experimental results demonstrated that the U-Net achieves competitive, and in many cases improved, detection performance compared to CFAR, particularly in cluttered scenarios. The inclusion of short-term temporal context via multi-frame inputs was found to enhance clutter suppression, while the low inference time suggests the model is suitable for real-time use. However, these findings are based on simulated data, and the generalizability to real-world radar remains an area for future investigation.

A follow-up experiment conducted after the initial thesis submission introduced modifications to the training data distribution and loss function, alongside a principled method for threshold calibration based on false alarm rate targets. These adjustments led to improved detection rates—especially for low-SNR targets—while maintaining comparable false alarm levels. Together, the results suggest that segmentation-based deep learning may offer a flexible and effective alternative to traditional radar detection pipelines, though further validation on real-world data is needed to confirm practical applicability.

References

[1] M. I. Skolnik, *Radar Handbook*, 3rd ed. New York, NY: McGraw-Hill, 2008.

[2] H. Rohling, "Radar cfar thresholding in clutter and multiple target situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 4, pp. 608–621, 1983.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[4] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," 2020. [Online]. Available: https://arxiv.org/abs/2005.12872

[5] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021. [Online]. Available: https://arxiv.org/abs/2103.14030

[6] Y. Dalbah, J. Lahoud, and H. Cholakkal, "Radarformer: Lightweight and accurate real-time radar object detection model," 2023. [Online]. Available: https://arxiv.org/abs/2304.08447

[7] Y. Wu, J. Liu, G. Jiang, W. Liu, and D. Orlando, "Mask-radarnet: Enhancing transformer with spatial-temporal semantic context for radar object detection in autonomous driving," 2024. [Online]. Available: https://arxiv.org/abs/2412.15595

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[9] S. Watts, "The performance of cell-averaging cfar systems in sea clutter," in *Record of the IEEE 2000 International Radar Conference [Cat. No. 00CH37037]*, 2000, pp. 398–403.

[10] F. Yavuz, "Radar target detection with cnn," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1581–1585.

[11] G. Li, Z. Song, and Q. Fu, "A convolutional neural network based approach to sea clutter suppression for small boat detection," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 10, pp. 1504–1520, 2020. [Online]. Available: https://doi.org/10.1631/FITEE.1900523

[12] T. Loran, A. Barros Cardoso da Silva, S. K. Joshi, S. V. Baumgartner, and G. Krieger, "Ship detection based on faster r-cnn using range-compressed airborne radar data," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[13] G. Zhai, J. Zhou, H. Yang, and Y. Zhang, "A sea-surface radar target-detection method based on an improved u-net and its fpga implementation," *Electronics*, vol. 14, no. 10, p. 1944, 2025. [Online]. Available: https://doi.org/10.3390/electronics14101944

[14] Y. Wang, W. Zhao, X. Wang, J. Chen, H. Li, and G. Cui, "Nonhomogeneous sea clutter suppression using complex-valued u-net model," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[15] T. Zhang and Y. Fan, "A 3d u-net based on a vision transformer for radar semantic segmentation," *Sensors (Basel)*, vol. 23, no. 24, p. 9630, Dec 2023. [Online]. Available: https://doi.org/10.3390/s23249630

[16] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2105.05537

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017. [Online]. Available: https://arxiv.org/abs/1706.03762

[18] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," 2021. [Online]. Available: https://arxiv.org/abs/2010.04159

[19] B. Wen, Y. Wei, and Z. Lu, "Sea clutter suppression and target detection algorithm of marine radar image sequence based on spatio-temporal domain joint filtering," *Entropy*, vol. 24, no. 2, 2022. [Online]. Available: https://www.mdpi.com/1099-4300/24/2/250