**MINISTRY OF EDUCATION AND TRAINING**

**HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY AND ENGINEERING**

**NGO DUC DAT**

# SHIP IDENTIFICATION AND CLASSIFICATION IN THE COASTAL SURVEILLANCE USING ARTIFICIAL INTELLIGENCE.

**PhD THESIS SUMMARY**

**FIELD: ELECTRONICS ENGINEERING**

**HO CHI MINNH CITY, APRIL 2026**

## LIST OF PUBLISHED WORKS

1. Image-Based Ship Detection Using Deep Variational Information Bottleneck.

Sensors, 23(19), 8093.

https://doi.org/10.3390/s23198093

2. Transformer based ship detector: An improvement on feature map and tiny training set. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 12(1).

https://doi.org/10.4108/eetinis.v12i1.6794

3. Clustering based ship classification using radar signal and neuron network. In Proceedings of the 2021 International

Conference on System Science and Engineering (ICSSE) (pp. 122-127). IEEE.

https://doi.org/10.1109/ICSSE52999.2021.9538475.

4. A Vision-based Container-Code Checking System: Case Study at International Terminal *(*IWIS*)*

https://doi.org/10.1109/IWIS58789.2023.10284525

# INTRODUCTION

## 1. Rationale for the Research Topic

Ship detection, identification, and classification play a crucial role in maritime surveillance and sovereignty protection. Currently, coastal monitoring systems mainly rely on pulse radar and surveillance cameras.

Pulse radar is widely used for long-range target detection and tracking under various weather and lighting conditions. It provides important information such as target position, velocity, and direction, enabling early detection of intrusions or abnormal activities. However, radar signals only reflect the electromagnetic characteristics of targets and cannot provide visual information, which limits the ability to accurately identify vessel types.

In contrast, surveillance cameras provide visual images that allow detailed observation of vessel shape, size, flags, and behavior. Cameras are increasingly deployed at radar stations and coastal observation centers. Nevertheless, their detection range is limited and their performance is strongly affected by weather and lighting conditions.

To overcome these limitations, modern coastal surveillance systems often integrate radar and cameras. Radar performs long-range detection while cameras provide detailed observation when targets approach coastal areas. However, ship identification and classification in these systems still depend largely on operator experience, leading to subjectivity and reduced efficiency in complex and high-density maritime environments.

Therefore, research on applying artificial intelligence to automate ship detection, identification, and classification based on radar signals and camera imagery is necessary.

## 2. Research Objectives

The thesis aims to develop artificial intelligence–based methods for ship detection, identification, and classification using radar reflection signals and surveillance camera imagery, thereby improving accuracy, objectivity, and automation in coastal surveillance systems.

## 3. Research Tasks

To achieve these objectives, the thesis focuses on the following tasks:

- Study and process surveillance camera image data for ship detection.
- Analyze the characteristics of radar signals.
- Develop and evaluate a deep learning model for ship detection based on radar signals.
- Develop and evaluate a deep learning model for ship detection based on images.

## 4. Research Scope

The research focuses on radar reflection signals and camera image data collected from public datasets and in-house datasets. The study mainly develops, trains, and evaluates artificial intelligence models in a prototype environment without large-scale hardware deployment.

The work concentrates on technical aspects such as radar signal processing, image processing, and algorithm efficiency, without addressing legal, policy, or international cooperation issues in maritime surveillance.

## 5. Research Approach and Methodology

### 5.1 Research Approach

The research integrates radar signals and surveillance camera data to exploit their complementary strengths in coastal monitoring systems. Artificial intelligence, machine learning, and deep learning techniques are applied to automatically extract features and classify ships, reducing dependence on human operators.

### 5.2 Research Methodology

For image data, deep learning models are used to detect ships and classify their categories based on available ship image datasets.

For radar signals, signal preprocessing, noise filtering, and feature extraction techniques are applied. Clustering methods are used to group signal patterns, and the results are integrated into classification models for ship identification.

System performance is evaluated based on recognition accuracy, classification capability, processing speed, and stability in complex maritime environments.

# CHAPTER 1: LITERATURE REVIEW

Chapter 1 presents an overview of coastal surveillance and the importance of ship identification and classification in increasingly complex maritime environments. Early detection, tracking, and classification of maritime targets are essential for maritime safety, resource management, and sovereignty protection.

Current coastal surveillance systems mainly use radar and camera sensors with complementary capabilities. Radar enables long-range detection and tracking under various environmental conditions but lacks visual information for accurate ship identification. In contrast, camera systems provide detailed visual information for target recognition, although their performance is limited by distance, lighting, and weather conditions.

Previous international studies highlight the important role of radar systems such as X-band radar and Pulse-Doppler radar in maritime target detection, often exploiting radar signal characteristics like Doppler features and radar cross section. Meanwhile, deep learning methods using camera images have achieved promising results in ship detection and classification tasks.

However, domestic research mainly focuses on HF radar for oceanographic monitoring, while studies combining pulse radar data with image data for ship recognition remain limited.

Therefore, integrating radar signals, camera imagery, and artificial intelligence is a promising approach to improving ship recognition and classification performance in coastal surveillance systems.

# CHAPTER 2: THEORETICAL FOUNDATIONS

Chapter 2 presents the theoretical foundations for ship detection and classification in coastal surveillance systems using image data. It first introduces surveillance camera systems and the role of image sensors in collecting visual data for object analysis and recognition. Image data from these systems provide an important source for applying image processing and artificial intelligence techniques.

The chapter reviews object detection approaches, including traditional methods based on manually designed features such as color, edges, and shape, and modern deep learning methods. Deep learning models, particularly Convolutional Neural Networks (CNNs), can automatically learn representative features from images with higher effectiveness. Representative detection architectures such as YOLOX and Transformer-based models like DETR are also introduced.

Common evaluation metrics for object detection are presented, including Precision, Recall, F1-score, Intersection over Union (IoU), Average Precision (AP), and mean Average Precision (mAP). These metrics are used to evaluate detection accuracy and model performance. In addition, clustering and classification methods used in data analysis and signal processing are reviewed to support dataset organization and model training. Finally, model optimization techniques and training strategies are discussed to improve deep learning performance, providing a theoretical basis for the methods proposed in subsequent chapters.

# CHAPTER 3: SHIP DETECTION FROM CAMERA SIGNALS

## 3.1 Dataset and Experimental Setting

To evaluate the proposed method, this research uses the SeaShips [15] dataset. Previous works shows that many studies use 80% of published data for training/validation and 20% for testing. We select $D^{Train}{}_{1,}$ consisting of 5,600 images, for training and $D^{Test}{}_{1,}$ consisting of 1,400 images, for testing. Recent works also use more challenging contexts with 50% training data and 50% test data. We prepare $D^{Train}{}_{2}$ and $D^{Test}{}_{2}$ following this setting for comparison. To evaluate performance on very small datasets, we randomly select subsets $S_1$, $S_2$, $S_3$ comprising 30%, 70%, and 100% of samples from $D^{Train}{}_{2}$ for subsequent experiments.

## 3.2 Experimental Results with YOLoX Model

### 3.2.1 Selection of Hyperparameter $\alpha_{KL}$

This section discusses how to select hyperparameter $\alpha_{KL}$ for training. The results are described in Figure 3.1. The results show that the optimal value of $\alpha_{KL}$ is 0.125.

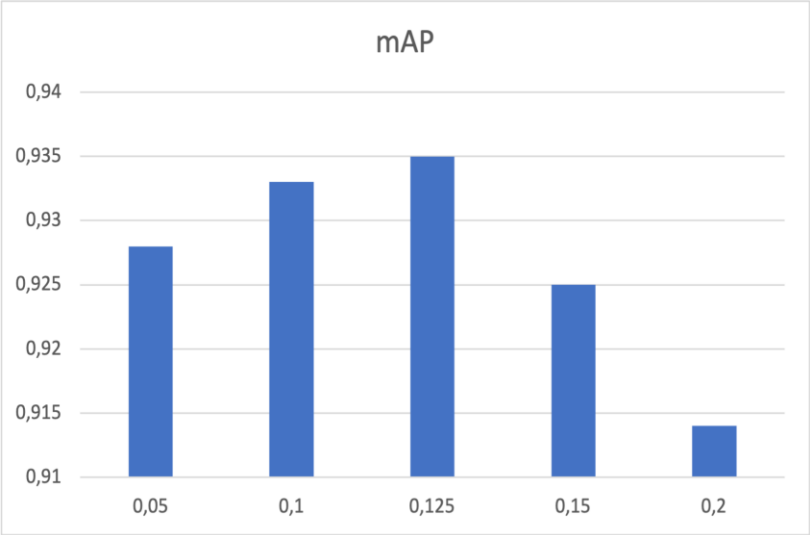



Figure 3.1: The mAP values corresponding to different $\alpha_{VIB}$ parameters. The x-axis represents the $\alpha_{KL}$ parameter, and the y-axis represents the average accuracy.

Without using the feature selection technique, mAP reaches only 0.923. When $\alpha_{KL}$ is 0.05, the mAP increases to 0.928. When $\alpha_{KL}$ is increased, the mAP increases progressively. However, when $\alpha_{KL}$ is 0.15, the mAP begins to decrease, and when $\alpha_{KL}$ = 0.2, the mAP is reduced to 0.914. When $\alpha_{KL}$ is 0.2, the mAP is lower the baseline without feature selection. This phenomenon occurs because the feature selection reduces the number of features selected for the primary task. When too many features are reduced, the classifier may lack information for classification. Therefore, $\alpha_{KL}$ = 0.125 is optimal for the dataset.

### 3.2.2 Comparison with State-of-the-Art Methods for Ship Detection

This section compares our proposed method with state-of-the-art methods. Based on 70,000 images published from the SeaShip [12], Zhang_2022[15] and Zhang_2021[16] used 90% data for training and validation; the remaining 10% is the test dataset. Liu_2020[17], Liu_2022[18], Han_2021[19], and Light_SDNet[20] used 80% data for training and validation, with 20% as the test dataset. To compare with these methods, we use $D^{Train}{}_1$ for training and $D^{Test}{}_1$ for testing. The results in Table 3.1 show that our method achieves higher accuracy compared to these SoTA.

Table 3.1: Comparison of performance of different methods. The best results are shown **in bold.**

| **Method** | **Train+Val / Test (in %)** | **fishing boat** | **container ship** | **ore carrier** | **bulk cargo carrier** | **passenger ship** | **general cargo ship** | **mAP** |
|---|---|---|---|---|---|---|---|---|
| Zhang_2022 [15] | 90/10 | 0.824 | 0.940 | 0.859 | 0.915 | 0.787 | 0.914 | 0.873 |
| Zhang_2021 [16] | 90/10 | - | - | - | - | - | - | 0.946 |
| Liu_2020 [17] | 80/20 | - | - | - | - | - | - | 0.908 |
| Liu_2022 [18] | 80/20 | - | - | - | - | - | - | 0.964 |
| Han_2021[19] | 80/20 | - | - | - | - | - | - | 0.906 |
| Light_SDNet [20] | 80/20 | **0.986** | 0.995 | **0.989** | 0.990 | 0.982 | 0.989 | 0.988 |
| Proposed method | 80/20 | 0.979 | **1** | 0.987 | **0.994** | **0.994** | **0.993** | **0.991** |
| Yani_2022 (ESDT) [13] | 50/50 | - | - | - | - | - | - | 0.593 |
| Yani_2022 (DETR)[13] | 50/50 | - | - | - | - | - | - | 0.965 |
| Biaohua_2022 [14] | 50/50 | 0.940 | 0.987 | 0.966 | 0.978 | 0.937 | 0.972 | 0.963 |
| Proposed method | 50/50 | 0.970 | 0.986 | 0.984 | 0.991 | 0.964 | 0.989 | **0.98** |

### 3.3 Experimental Results with DETR

#### 3.3.1 Hyperparameter Selection

The number of queries can significantly affect the output quantity of a DETR-based detector. This section focuses on selecting appropriate parameters to control the model. We compare performance when $n_{queries}$ takes values [300, 200, 100, 50]. The sets $D^{Train}{}_2$ and $D^{Test}{}_2$ are selected as training and test sets to ensure a high-challenge problem setup, meaning 50% of data is used for testing.

Experimental results are presented in detail in Table 3.2. In the mmdetection library, the default value for $n_{queries}$ is 300. Observation shows that using this default setting leads to a significant increase in detection quantity. For example, the detection quantity for the fishing boat class is 204,449. This trend results in reduced average precision (AP) to 0.798, while recall increases to 0.913. Additionally, the high detection quantity for fishing boats may be due to more labels of the category in the dataset.

By reducing $n_{queries}$, bias in detections is significantly minimized. Specifically, when n_queries is reduced to 200, 100, and 50, the fishing boat detection quantities decrease correspondingly to 208,817, 119,774, and 96,577. Furthermore, reducing this parameter helps balance detection quantities across different vessel types. With $n_{queries}$ = 300, the lowest detection quantity belongs to container ships (6,300). However, when $n_{queries}$ = 50, the lowest detection quantity increases to approximately 18,949, and detection distribution among container ships, ore carriers, and passenger ships becomes relatively similar.

Table 3.2: Comparison of Deformable DETR performance with different numbers of queries. The best results are shown **in bold**.

| Class | 300 query | | | 200 query | | | 100 query | | | 50 query | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **dets** | **recall** | **AP** | **dets** | **recall** | **AP** | **dets** | **recall** | **AP** | **dets** | **recall** | **AP** |
| fishing boat | 204449 | 0,913 | 0,798 | 208817 | 0,985 | 0,970 | 119774 | **0,986** | **0,969** | 96577 | 0,972 | 0,949 |
| container ship | 6300 | 0,989 | 0,894 | 11460 | 0,995 | 0,995 | 31083 | 0,995 | 0,989 | 18949 | **0,998** | **0,997** |
| ore carrier | 53520 | 0,987 | 0,923 | 49362 | **0,997** | **0,992** | 62122 | 0,996 | 0,989 | 18949 | 0,993 | 0,987 |
| bulk cargo carrier | 45541 | 0,985 | 0,912 | 47754 | 0,996 | **0,992** | 46490 | **0,997** | 0,986 | 76230 | 0,996 | **0,989** |
| passenger ship | 22498 | 0,968 | 0,610 | 16199 | 0,970 | **0,957** | 14866 | 0,968 | 0,936 | 20043 | **0,972** | 0,926 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| general cargo ship | 17692 | 0,987 | 0,930 | 16408 | 0,995 | **0,992** | 75665 | 0,995 | 0,991 | 98537 | **0,996** | **0,990** |
| mAP | | | 0.844 | | | **0.982** | | | 0.977 | | | 0.973 |

### 3.3.2 Comparison with State-of-the-Art Methods (SoTA)

This section compares our proposed method with state-of-the-art methods (SoTA) based on the mAP index. For each method, corresponding training and test datasets are used. Specifically, we use $D^{Train}{}_1$ and $D^{Test}{}_1$ to train our model and compare with Zhang_2022, Zhang_2021[16], Liu_2020, Liu_2022, Han_2021, SDNet_2022, and DETR-based methods. Additionally, we use $D^{Train}{}_2$ and $D^{Test}{}_2$ to train another model and compare it to Biaohua_2022, Yani_2022, and DETR-based methods.

Because the mAP is maximum when $n_{queries}$ = 200 for our method, we select this setting for this experiment. Given the results in Table 3.3, several conclusions can be drawn:

- The base framework is a key factor for achieving better results: Cui_2019 and Liu_2020 were developed based on YoloV3. Therefore, their performance is not as good as Liu_2022, which is based on the SSD framework. Han_2021 is based on YoloV4, but performance shows no improvement over Liu_2020 (based on YoloV3). Leveraging YoloV5 advantages, SDNet_2022 achieved significant improvement over Liu_2020. The YoloV5 framework increases mAP index by up to 8% compared to YoloV3. The DETR-based method is based on YOLoX, while our method is based on the DETR backbone. Recently, these frameworks have become predominant methods for object detection tasks. Therefore, the achieved results are better than other methods. Notably, ship detection research typically leverages an object detection framework as a base, accompanied by task-specific modifications. Therefore, inheriting the capabilities of such a new and powerful framework inevitably leads to improved results.
- When the number of training samples decreases, DETR-based methods tend to perform more effectively than CNN-based methods. Specifically, in the table, Biaohua_2022 is a CNN-based detection model, while Yani_2022 is a DETR-based model. The mAP indices of Yani_2022 and Biaohua_2022 are 0.965 and 0.9963, respectively. However, this performance can be improved by selecting appropriate hyperparameters. In our research, by setting $n_{queries}$ = 200, the mAP index was improved to 0.981.

- The proposed method is equivalent to the current best CNN-based method for ship detection. If the training dataset has more samples than the test set, our method is equivalent to YOLO-based methods and DETR-based methods. Specifically, when using $D^{Train}{}_1$ and $D^{Test}{}_1$ for training and testing, the mAP index of both our method and DETR-based methods are similar. However, when the number of samples in the training set equals the number in the test set, our method slightly outperforms the DETR-based method.

  Table 3.3: Comparison of performance of different methods. The best results are shown **in bold.**

| **Method** | **Train+Val / Test (in %)** | **fishing boat** | **container ship** | **ore carrier** | **bulk cargo carrier** | **passenger ship** | **general cargo ship** | **mAP** |
|---|---|---|---|---|---|---|---|---|
| Zhang_2022 [20] | 90/10 | 0.824 | 0.940 | 0.859 | 0.915 | 0.787 | 0.914 | 0.873 |
| Zhang_2021 [12] | 90/10 | - | - | - | - | - | - | 0.946 |
| Cui_2019 [21] | 80/20 | 0.900 | 0.940 | 0.90 | 0.910 | 0.910 | 0.900 | 0.910 |
| Liu_2020 [22] | 80/20 | - | - | - | - | - | - | 0.908 |
| Han_2021 [19] | 80/20 | - | - | - | - | - | - | 0.906 |
| Liu_2022 [16] | 80/20 | - | - | - | - | - | - | 0.964 |
| SDNet_2022 [20] | 80/20 | **0.986** | 0.995 | **0.989** | 0.990 | 0.982 | 0.989 | 0.988 |
| Method based on DETR [23] | 80/20 | 0.979 | **1** | 0.987 | **0.994** | 0.994 | **0.993** | **0.991** |
| Ours ($n_{query}$ =200) | 80/20 | 0.982 | **1** | **0.989** | 0.991 | **0.995** | 0.990 | **0.991** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Yani_2022 (ESDT) [13] | 50/50 | - | - | - | - | - | - | 0.593 |
| Yani_2022 (DETR) [13] | 50/50 | - | - | - | - | - | - | 0.965 |
| Biaohua_2022 [14] | 50/50 | 0.940 | 0.987 | 0.966 | 0.978 | 0.937 | 0.972 | 0.963 |
| Method base on DETR [23] | 50/50 | 0.970 | 0.986 | 0.984 | **0.991** | **0.964** | 0.989 | 0.98 |
| Ours ($n_{query}$ =200) | 50/50 | 0.970 | **0.995** | **0.992** | 0.992 | 0.957 | **0.992** | **0.982** |

### 3.3.3 Ablation Study

This section discusses an ablation study on loss functions during training. Deformable DETR uses multiple loss functions for training, including focal loss, GIoU loss, and L1 loss, each serving a specific purpose: focal loss for classification, GIoU for bounding box regression, and L1 for object detection. The combination of these loss functions ensures training success. Although all loss functions play crucial roles, their contribution levels can be adjusted. By default, weights are set to 2.0 for focal loss, 2.0 for GIoU loss, and 5.0 for L1 loss. To evaluate the impact of each component, we reduced the weights of these loss functions by 10 times (performed sequentially for each) and compared results with default settings. Results in Table 3.4 demonstrate that $L_{cls}$ (classification loss function) is the most critical factor; reducing its weight leads to significant performance degradation. Conversely, reducing the object loss function has less impact, as performance remains at levels nearly equivalent to the initial setting. Localization capability is slightly affected by the GIoU loss function, as shown by mAP decreasing to 0.931 from 0.941 in the initial setting.

Table 3.4: Comparison of mAP of Deformable DETR when reducing training loss functions.

| | *LGIoU* | $L_1$ | *Lcls* |
|---|---|---|---|
| fishing boat | 0.899 | 0.899 | 0.227 |
| container ship | 0.976 | 0.985 | 0.246 |
| ore carrier | 0.953 | 0.947 | 0.212 |

| | | | |
|---|---|---|---|
| bulk cargo carrier | 0.951 | 0.966 | 0.151 |
| passenger ship | 0.845 | 0.849 | 0.0173 |
| general cargo ship | 0.964 | 0.969 | 0.160 |
| **mAP** | 0.931 | 0.936 | 0.178 |

### 3.3.4 Feature Analysis

Experimental results in Section 3.3.2 indicate that DETR methods perform better than CNN methods when the training sample is limited. However, it is worth explaining why DETR based method is better than the Yolo base method

To address the question, we visualize features generated by both methods after the backbone, neck, and head sections of the model. For each input image, feature maps are extracted after each module. The total value of a feature map represents its importance level. Thus, we selected the 20 most important feature maps for each backbone, neck, and head section to generate a heatmap. This map is computed as the average of all important feature maps and represents focal regions in the image. Figure 3.2 illustrates examples of heatmaps generated from an input image. The first row shows feature maps from DETR, the second row shows heatmaps from the DETR-based method (using a feature selection loss function), and the third row presents heatmaps from YOLOX based entirely on CNN networks. Results show that DETR, with its attention mechanism, can better focus on non-background objects. For example, after the head section, feature maps highlight text displaying scores in the survival system, and the ship is also highlighted, though less clearly than the text. As features are processed through the neck section, higher-level semantic features are learned, making the ship more prominent while the text focus decreases. Main focal points concentrate on the ship in the head section, with text attention diminishing.

Conversely, the DETR-based method generates sparse heatmaps where many pixels at image edges show no response. However, these maps do not concentrate precisely on the object. This occurs because the DETR-based method uses a feature selection loss function to determine important features, resulting in sparse, highly discriminative feature maps that do not precisely center on the object. The overall heatmap distribution is relatively similar, with minor improvements from backbone to head. Without the feature selection loss function, the heatmap distribution would likely be more uniform, as shown in the third row.

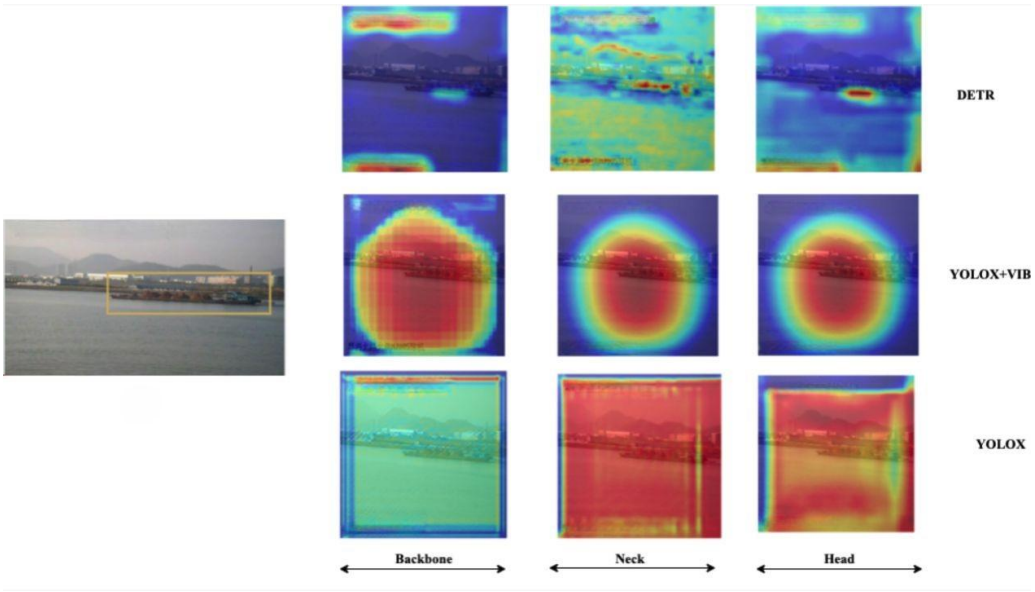



Figure 3.2: Feature map at the head, neck, and backbone. (Text in the original image has been removed.)

# CHAPTER 4: RADAR DATA ANALYSIS USING DEEP LEARNING TECHNIQUES

## 4.1 Expert Knowledge-Based Clustering Method for Ship Classification

### 4.1.1 System Overview

The system overview is presented in Figure 4.1. The radar signal dataset is annotated with attributes based on expert knowledge. The deep learning model learns features that both reconstruct radar reflection signal information and predict labeled attributes for each sample. Finally, these features are used for data clustering.

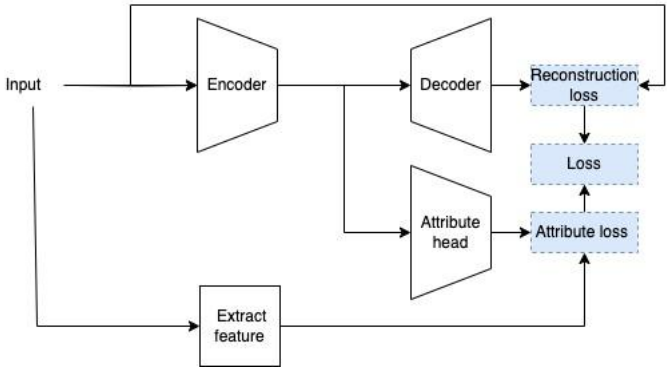



Figure 4.1: Overview of the proposed methods

Section 4.1.2 presents how attributes describe vessel types and the rationale for selecting them. Section 4.1.3 describes attribute annotation for each sample in the dataset. Section 4.1.4 presents the model architecture and objective function.

### 4.1.2 Expert - based attributes

To analyze attributes that support vessel clustering, fishing boat, military, and cargo vessel samples were separated and displayed in Figures 4.2, 4.3, and 4.4. Fishing boats have small sizes and hull characteristics, resulting in low-amplitude radar reflection. Additionally, fishing boats often travel together, potentially creating multiple wave peaks in a single reflection, as shown in Figure 4.2. In Figure 4.3. When a military vessel is far from the radar station, the reflected signal has a small peak amplitude, as in Figure 4.3a. Conversely, when near the radar station, the reflection has large peak and trough amplitudes as in Figure 4.3b. For cargo vessels: when a cargo vessel is distant with a tendency to run parallel to the coast, the reflection has low amplitude and higher peak as in Figure 4.4a. When a ship is near the coast but not heading into port, the reflection has a larger amplitude. Also amplitude is large; additionally, peak fluctuations remain large as in Figure 4.4b. When near the coast and heading into port, signal amplitude is large, but signal width is only moderate because only the vessel bow contacts the radar wave as in Figure 4.4c.

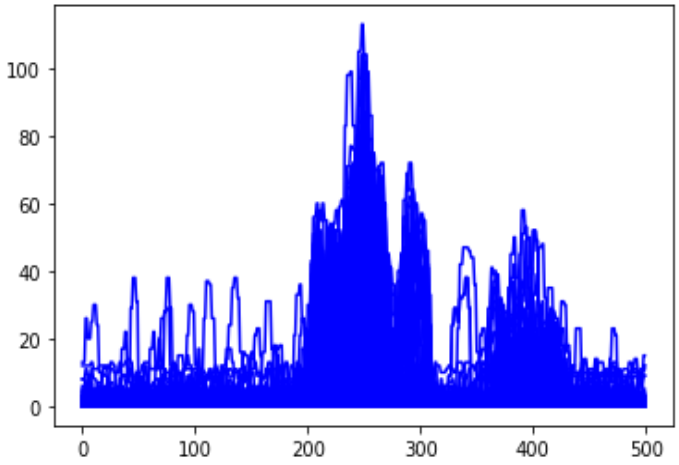
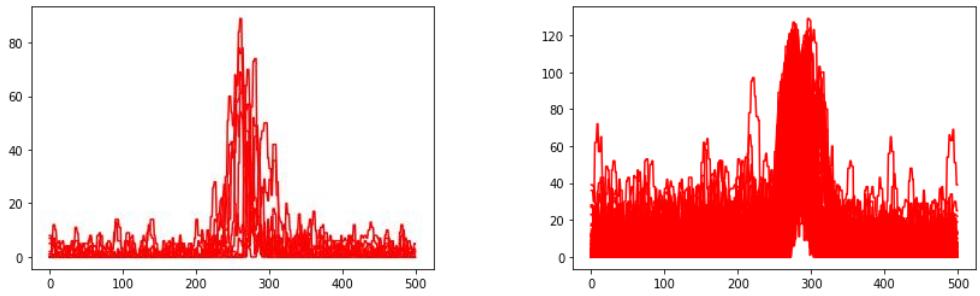


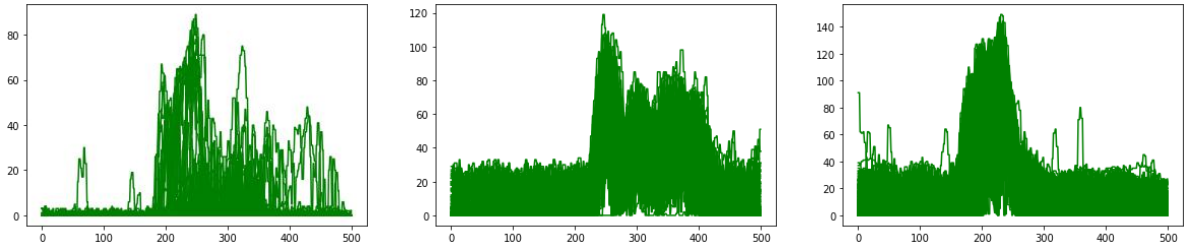

Figure 4.2: Waveform of fishing vessels, the vessels in only group 1

Based on these analyses, the following characteristics are used to assess vessel type based on radar reflection:



(a) From group 1 (b) From group 2

Figure 4.3: Waveforms of military ships



(a) From group 1 (b) From group 2 (c) From group 3

Figure 4.4: Waveforms of cargo ships.

- Number of peaks in the reflected signal: typically, radar reflection waveforms have only one peak. For fishing boats, due to the small object size and occasional interference from nearby fishing boats, multiple peaks may appear.
- Peak amplitude of the reflected signal: this is the maximum value of the peak section, significantly affecting clustering results if using raw sensor signals.
- Peak width of the reflected signal: this characteristic is very small for fishing boat reflections. For military vessels, this value can be small or moderate depending on the distance. The cargo vessel's peak is very large due to stacked containers. However, this characteristic only takes a moderate value when heading toward shore.
- Peak fluctuation of the reflected signal: this characteristic describes fluctuation at the peak section. For fishing boats or military vessels, this value tends to be smaller. Cargo vessels typically have large values according to this characteristic.
- Trough amplitude of the reflected signal: this characteristic describes the bottom section of a waveform. For fishing boats, this value is very small due to the small vessel size. Cargo vessels or military vessels have larger values due to hull characteristics.

- Through fluctuation of the reflected signal: with this characteristic, fishing boats typically have small values, while military and cargo vessels typically have large values.

### 4.1.3 Attribute Extraction

Let the radar reflection signal be x. The process for extracting attributes in Section 4.1.2 is described as follows:

1. Apply a first-order Butterworth filter to extract low-pass component $x^H$ and high-pass component $x^L$ .
2. Apply peak detection algorithm [24] to detect the number of peaks $n_{peak}$ in waveform x.
3. Using the low-frequency signal $x^L$, separate peak and trough sections based on start time B and end time E of the highest peak. The description of times B and E is shown in Figure 4.5. Let $min_{30}(x)$ be the set of 30% smallest values of x; *Base* value in Figure 4.5 is determined by the formula $Base = mean(min_{30}(x))$

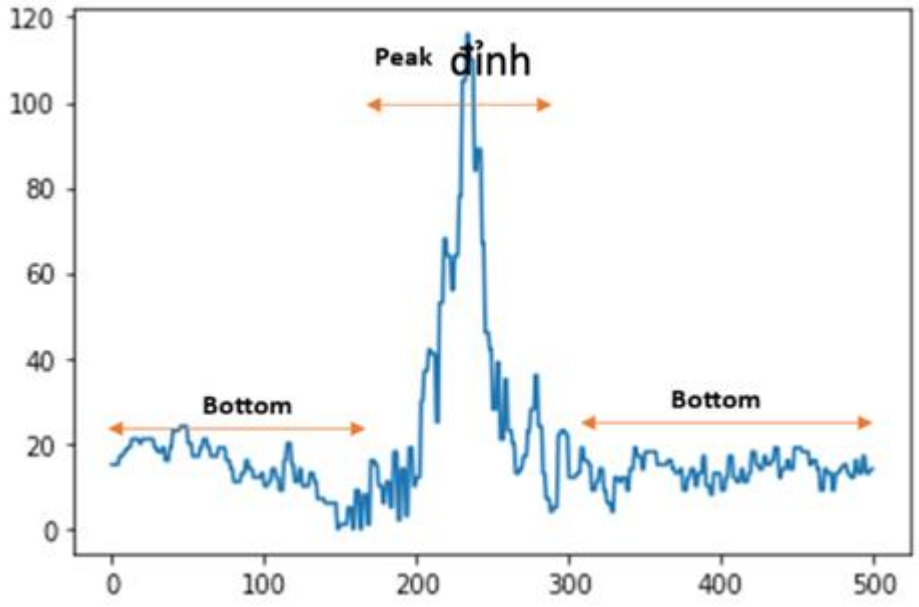



Figure 4.5 Peaks and troughs of the radar reflection waveform

4. Based on start time B and end time E of the highest peak, extract information on peak and trough sections from low-frequency component $x^L$ and high-frequency component

$x^H$ Specifically, $x_{BE}^L$ is the peak section of the low-frequency signal; while $x_{\sim}^L{}_{BE}$ is the trough section of the low-frequency signal.

5. Let N be the number of elements in a radar reflection signal sample. Characteristics on peak width, peak amplitude, peak fluctuation, trough amplitude, and trough fluctuation are calculated by Equations 4.1, 4.2, 4.3, 4.4, and 4.5 respectively.

$$L_{peak} = \frac{E-B}{N} \quad (4.1)$$

$$A_{peak} = mean\left(max_{30}(x_{BE}^L)\right) \quad (4.2)$$

$$S_{peak} = mean\left[x_{BE}^H{}^2\right] \quad (4.3)$$

$$A_{base} = mean(x_{\sim BE}) \quad (4.4)$$

$$S_{base} = mean\left[x_{\sim BE}^H{}^2\right] \quad (4.5)$$

### 4.1.4 Feature Extraction Model and Objective Function

The model architecture used in the paper is described as follows:

Encoder model: $\boldsymbol{:f_{\theta^E}}(.)$ is used for feature extraction. This block is a sequence of blocks from the list [*nn.linear,nn.relu,*].

Decoder model: $\boldsymbol{f_{\theta^D}}(.)$ is used to recover the original information. This block is a sequence of blocks from the list [*nn.linear,nn.relu,*].

Attribute prediction model: $\boldsymbol{f_{\theta^D}}(.)$ is used to predict attributes from features extracted by the encoder block. Layers in the attribute prediction block are described as [*nn.linear,nn.relu,*]. Let x be a 500-element vector describing the sensor input signal, and $\hat{x}$ be the reconstructed waveform. The objective function for data reconstruction is described in Equation 4.6.

$$L_R(x, \hat{x}) = \left||x - \hat{x}|\right|^2 \quad (4.6)$$

Let a be the attribute labeled in Section 4.1.3 and â be the attribute predicted from the attribute prediction model. The attribute objective function is described in Equation 4.7.

$$L_A(a, \hat{a}) = \left||a - \hat{a}|\right|^2 \quad (4.7)$$

The parameter α controls the balance between two constraints $L_R(x, \hat{x})$ and $L_A(a, \hat{a})$ .The objective function for training the feature extraction model is described as follows:

$$Loss(x, a) = L_R\left(x, f_{\theta^D}\left(f_{\theta^E}(x)\right)\right) + \alpha L_A\left(a, f_{\theta^A}\left(f_{\theta^E}(x)\right)\right) \quad (4.8)$$

## 4.2 Clustering Results for Ship Signals

### 4.2.1 Experimental Results with Traditional Methods

This section presents experimental results based on traditional clustering algorithms. The K-means algorithm is used with different features such as FFT[25], DCT[26], and DWT[27]. Additionally, data normalization schemes are applied to features to compensate for nonlinearity when converting signals from the time domain to the frequency domain. Experimental results are described in Table 4.1. Results show that features in the time domain provide better results than FFT or DCT features.

Table 4.1: Clustering results based on Kmeans and features.

| Đặc trưng | Chuẩn hoá | MI | A_MI | NorMI | A_RS | Complex | Flow |
|---|---|---|---|---|---|---|---|
| Raw | TRUE | 14.41 | 17.48 | 17.61 | 03.05 | 24.78 | 51.59 |
| Raw | FALSE | 76.26 | 71.27 | 71.30 | 72.11 | 70.32 | 81.93 |
| FFT | TRUE | 49.11 | 46.22 | 46.28 | 46.96 | 46.00 | 65.86 |
| FFT | FALSE | 60.71 | 57.35 | 57.40 | 54.57 | 57.22 | 70.83 |
| DCT | TRUE | 60.81 | 59.58 | 59.63 | 47.23 | 61.75 | 67.72 |
| DCT | FALSE | 57.81 | 55.71 | 55.76 | 47.55 | 56.75 | 67.23 |
| DWT | TRUE | 44.63 | 45.79 | 45.86 | 34.65 | 50.05 | 61.95 |
| DWT | FALSE | 76.26 | 71.27 | 71.30 | 72.11 | 70.32 | 81.93 |

### 4.2.2 Comparison with Advanced Deep Learning Methods

Table 4.2: Comparison with deep learning clustering methods.

| Methods | MI | A_MI | NorMI | A_RS | Complex | Flow |
|---|---|---|---|---|---|---|
| AE [28] | 77.28 | 72.19 | 72.22 | 73.03 | 71.19 | 82.51 |
| VAE [29] | 79.18 | 74.16 | 74.19 | 76.56 | 73.33 | 84.84 |
| DEC [30] | 77.59 | 72.47 | 72.50 | 73.44 | 71.46 | 82.27 |

| | | | | | | |
|---|---|---|---|---|---|---|
| DeepCluster [31] | 75.33 | 70.39 | 70.43 | 87.12 | 71.75 | 69.45 |
| Proposed method | 84.37 | 80.01 | 80.03 | 84.98 | 80.07 | 90.39 |
| Only attribute | 28.48 | 26.72 | 26.81 | 19.92 | 26.60 | 26.61 |

This section compares the proposed method with deep learning clustering methods including AE[28], VAE[19], DEC[30], and DeepCluster[31]. AE[28] and VAE[29] are classical methods where feature extraction blocks and data clustering blocks are separated. Unlike classical deep learning methods, the proposed method benefits from attribute information to improve feature extraction. In conditions with limited data, leveraging expert knowledge to train the feature extraction model is an appropriate approach. Accordingly, the feature extraction block not only helps reconstruct original information but sometimes assists in predicting sample attributes. We see the MI value significantly increases to 84.13% corresponding to $\alpha = 0.1$. It should be noted that clustering based solely on attribute information results in MI, A_MI, N_MI, RS, A_RS, CS, and FM_S values of 0.2848, 0.2672, 0.2680, 0.5083, 0.1992, 0.2660, and 0.4838, respectively. This shows that attribute annotation does not directly assist clustering; however, when combined with the signal reconstruction objective function, learned features will support finding better features for clustering. For this reason, when designing the clustering block, we need to use multiple neural network layers to reduce the impact of attribute prediction on clustering features.

### 4.3 Ship Classification Through Clustering Techniques

#### 4.3.1 System Overview

Due to the limited nature of collected datasets, this dissertation uses only traditional methods for ship classification. Clustering algorithm is a technique dividing a set of input samples into different clusters. A cluster is a group of data points similar based on their relationships with surrounding points. One of the simplest and most easily implemented clustering algorithms is the K-means technique.

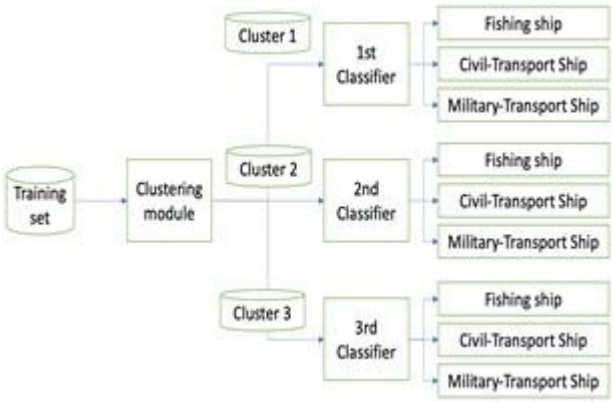



Figure 4.6: System Overview

The K-means algorithm works best on smaller datasets because it iterates over all data points. This means it takes longer to classify if there is a large amount of data. With dataset $X = [x_1,x_2,...,x_N] \in R^{dxN}$ consisting of N samples represented by d features, the K-means algorithm divides this dataset into K clusters. Within each cluster, members must be as similar as possible. Each cluster is represented by a cluster center $m_k \in R^{dx1}(k = 1 \sim K)$; and for each point in the dataset, a label must be assigned to determine which cluster the sample belongs to.

## 4.4 Experimental Results

### 4.4.1 Comparison of Features in Frequency Domain and Time Domain.

Traditionally, 1D[33], [34] signal classification problems must segment signals before classification. Because signals in the time domain are easily affected by segmentation, frequency features of 1D signals are extracted to eliminate this effect. However, in this application, radar emitted signals are modulated by square pulses, so the segmentation problem is well-solved. Therefore, signals in the time domain can also be reliable signals for machine learning algorithms. In this experiment, we investigate the impact of features in the frequency domain and the spatial domain on machine learning algorithms. Two basic machine learning tasks, clustering and classification, are applied with FFT features and features in the time domain. For each task, we use 33% of data for testing and 66% for training. Algorithms are implemented based on the sklearn library.

For the clustering task, we use the K-means algorithm to divide data into 3 clusters. Clustering results are presented in Table 4.3.

During evaluation, a test sample is input to the cluster generator created during training. The sample is assigned to one of the three clusters created during training.

Table 4.3: Classification results with different characteristics of ship types.

| Training process | | | | |
|---|---|---|---|---|
| | | Fishing boat | Military transport ship | regular cargo ship |
| FFT | C0 | 418 | 14 | 21 |
| | C1 | 0 | 13 | 205 |
| | C2 | 1 | 186 | 232 |
| FFT + Scaling | C0 | 0 | 18 | 299 |
| | C1 | 399 | 33 | 28 |
| | C2 | 20 | 162 | 201 |
| FFT + Log | C0 | 367 | 9 | 9 |
| | C1 | 2 | 45 | 308 |
| | C2 | 50 | 159 | 141 |
| Time domain | C0 | 0 | 0 | 341 |
| | C1 | 419 | 7 | 28 |
| | C2 | 0 | 206 | 89 |
| The evaluation process | | | | |
| FFT | C0 | 192 | 6 | 10 |
| | C1 | 0 | 9 | 114 |
| | C2 | 0 | 100 | 106 |
| | C0 | 0 | 16 | 116 |

| | | | | |
|---|---|---|---|---|
| FFT + Scaling | C1 | 186 | 15 | 12 |
| | C2 | 6 | 84 | 102 |
| FFT + Log | C0 | 175 | 2 | 6 |
| | C1 | 1 | 32 | 160 |
| | C2 | 16 | 81 | 84 |
| Time domain | C0 | 0 | 0 | 180 |
| | C1 | 192 | 4 | 14 |
| | C2 | 0 | 111 | 36 |

Table 4.4: Summary of neural network configurations and training parameters

| Parameter | Solver | Coefficient Regulization | Hidden Layer Size | Random State |
|---|---|---|---|---|
| S1 | lbfgs | 1e-5 | (300,200,100,50) | 100 |
| S2 | lbfgs | 1e-5 | (300,100,50) | 100 |
| S3 | lbfgs | 1e-5 | (200,75) | 100 |
| S4 | lbfgs | 1e-5 | (100,75) | 100 |
| S5 | lbfgs | 1e-5 | (75,50) | 100 |

Not only did the students perform clustering using different features, but they also performed ship classification directly based on the features used in the clustering problem. The students used three neural network configurations (S1, S3, S5) to evaluate the role of the features. Among these classifiers, configuration S1 had the highest complexity because it consisted of many layers, and the number of nodes in each layer was high. Model S5 had the lowest complexity because it only had 2 hidden layers, and the number of nodes for each layer was low. Details of these classifiers are described in Table 4.4. Experimental results in Table 4.5 show that the FFT feature is not suitable for ship classification. The very low classification results demonstrate that the model cannot converge well to classify ships. If nonlinear factors

are compensated for by data normalization [35] and log functions, the model can learn but the accuracy is not high. The best possible value achieved on the test set was 96%, while the time-domain classification accuracy was 98%. This result is consistent with the results obtained when implementing the clustering algorithm, indicating that the time-domain signal is good enough for processing maritime radar signals.

Table 4.5: Classification results using different characteristics with training/test ratios.

| Pha test | | | |
|---|---|---|---|
| | | ACC | F1-Score |
| FFT | S1 | 0.1195 | 0.2364 |
| | S3 | 0.5680 | 0.6312 |
| | S5 | 0.3910 | 0.4357 |
| FFT + Scaling | S1 | 0.9606 | 0.9608 |
| | S3 | 0.9606 | 0.9608 |
| | S5 | 0.9492 | 0.9497 |
| FFT + Log | S1 | 0.9414 | 0.9422 |
| | S3 | 0.9243 | 0.9255 |
| | S5 | 0.0945 | 0.9255 |
| Time domain | S1 | 0.9813 | 0.9814 |
| | S3 | 0.9795 | 0.9795 |
| | S5 | 0.9795 | 0.9795 |

### 4.4.2 Impact of Clustering Algorithm on Classification Task

This section demonstrates the effectiveness of the proposed method using clustering algorithms for data preprocessing before classification. Like traditional classification problems, we use accuracy and F1-score to evaluate system effectiveness. However, to

emphasize the contribution of the clustering process, we use the enhancement ratio of ACC and F1-score as in Equation 4.13.

$$E = \frac{P_{clus} - P}{P} 100 \quad (4.13)$$

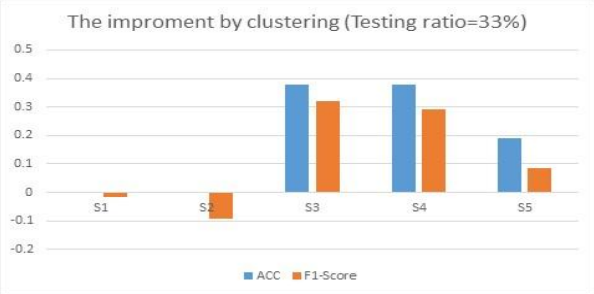



Figure 4.7: Enhancement of ACC and F1-Score when the training/testing ratio is 66%/33%

Here, P is the system performance without preprocessing clustering; P_clus is system performance with clustering preprocessing. Performance here includes accuracy (ACC) or F1-score. If this value is greater than 0, we can conclude that preprocessing through clustering helps improve classification quality. Conversely, if this value is less than 0, clustering preprocessing does not help enhance performance. We conduct experiments with multiple different neural network configurations from very dense (S1) to very sparse (S5). Details of these configurations are listed in Table 4.4. Figure 4.7 and 4.8 show the enhancement level of accuracy and F1-Score when selecting training/testing ratios of 66%/33% and 66%/50%.

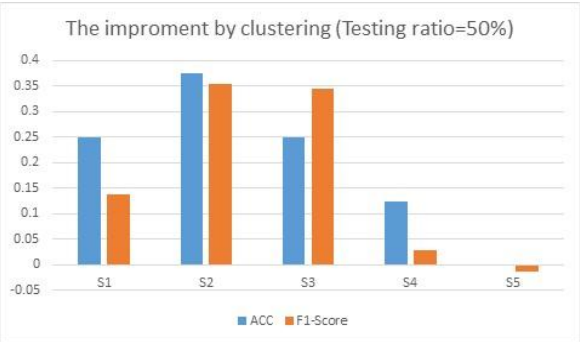



Figure 4.8: Enhancement of ACC and F1-Score when the training/testing ratio is 66%/50%

## CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

Maritime surveillance and security assurance are of critical importance for marine economic development, the protection of national sovereignty, and the maintenance of maritime order. In the context of increasing maritime activities in both scale and complexity, single-sensor surveillance systems exhibit significant limitations, particularly in ship detection, identification, and classification tasks. Addressing these practical requirements, this

dissertation focuses on methods for coastal surveillance that combine pulsed radar signals and camera image data, supported by modern artificial intelligence techniques.

Through a systematic review of domestic and international publications, the dissertation shows that most existing studies treat radar and image data separately. While radar systems excel at target detection and tracking under all weather conditions, cameras provide rich semantic information useful for identification and classification. However, integrating these two sensor types in realistic coastal environments characterized by noise and variability remains a significant research challenge.

For image data, the dissertation proposes two deep learning–based models for ship detection. The first model is based on convolutional neural networks combined with feature selection blocks designed to remove redundant information and highlight highly discriminative features. Experimental results demonstrate that this model achieves higher accuracy compared with many advanced methods, particularly when training data are limited. The second model is based on a transformer architecture, which exploits attention mechanisms to focus on important regions in images. This model shows strong generalization ability with small datasets and provides meaningful attention maps that support result interpretation and explanation.

For pulsed radar signals, the dissertation constructs a realistic dataset collected from coastal radar stations in Vietnam, addressing the limitations of previous studies that primarily rely on simulated data. Based on this dataset, expert knowledge is combined with machine learning techniques to extract features reflecting the physical and kinematic characteristics of targets. The proposed ship clustering and classification method based on radar signals demonstrates the capability to distinguish major vessel groups while also revealing limitations of traditional approaches that are strongly affected by observation distance.

In summary, this dissertation contributes both theoretically and practically to the field of coastal surveillance. From an academic perspective, it clarifies the role of deep learning and selective feature representation in ship recognition under limited data conditions. From a practical perspective, the developed models and datasets provide a foundation for the development of intelligent coastal surveillance systems aimed at greater automation, improved reliability, and reduced operational dependence on human factors.

## References


[1] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You only look one-level feature," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 13034–13043.

[2] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," 2021. [Online]. Available: https://arxiv.org/abs/2107.08430

[3] S. Cope et al., "Coastal radar as a tool for continuous and fine-scale monitoring of vessel activities," PLOS ONE, 2022, open-access journal article.

[4] "On a new atr tool for coastal surveillance: Pulse-doppler radar applications," conference or technical paper; biblio- graphic details incomplete.

[5] D. Yang et al., "A review of intelligent ship marine object detection based on rgb camera," IET Image Processing, 2024.

[6] "Leveraging deep learning and computer vision for coastal fisheries monitoring," Scientific Reports, 2024.

[7] J. Molina et al., "Robust sensor fusion in real maritime surveillance scenarios," 2010, classic paper on maritime sensor fusion.

[8] Y. Zhou et al., "Review on millimeter-wave radar and camera fusion: Techniques and applications," Sustainability, 2022.

[9] L. Trinh, S. Mercelis, and A. Anwar, "A comprehensive review of datasets and deep learning techniques for vision in unmanned surface vehicles," Ocean Engineering, vol. 334, p. 121501, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801825011850

[10] T. H. Tran, A. Sentchev, T. To Duy, M. Herrmann, S. Ouillon, and K. C. Nguyen, "Surface circulation characterization along the middle southern coastal region of vietnam from high-frequency radar and numerical modeling," Ocean Science, vol. 21, no. 1, pp. 1–18, 2025. [Online]. Available: https://os.copernicus.org/articles/21/1/2025/

[11] B. Tran, "Vietnam's quest for enhanced maritime domain awareness," ISEAS – Yusof Ishak Institute, Singapore, ISEAS Perspective 2023/96, December 2023, iSSN 2335-6677. [Online]. Available: https:

//www.iseas.edu.sg/wp-content/uploads/2024/01/ISEAS_Perspective_2023_96.pdf

[12] N. K. Cuong, T. N. Anh, N. X. Loc, P. D. H. Binh, and V. H. Dang, "Advanced high-resolution measurements of surface waves and currents using two land-based hf radars for offshore operations," in Proceedings of the 3rd Vietnam Symposium on Advances in Offshore Engineering, D. V. K. Huynh, H. Doan, T. M. Cao, and P. Watson, Eds. Singapore: Springer Nature Singapore, 2025, pp. 61–69.

[13] N. Mai, A. Sentchev, and T. Cuong, “Applying the method of eof interpolation and 2dvar to complete the ocean surface current data obtained from hf radar system,” VNU Journal of Science: Earth and Environmental Sciences, vol. 34, 12 2018.

[14] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” 2015. [Online]. Available: https://arxiv.org/abs/1506.01497

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot MultiBox detector,” in Computer Vision – ECCV 2016. Springer International Publishing, 2016, pp. 21–37. [Online]. Available: https://doi.org/10.1007%2F978-3-319-46448-0_2

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[17] 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

[18] ——, End-to-End Object Detection with Transformers, 11 2020, pp. 213–229.

[19] H. W. Kuhn, “The Hungarian Method for the Assignment Problem,” Naval Research Logistics Quarterly, vol. 2, no. 1–2, pp. 83–97, March 1955.

[20] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658–666.

[21] P. Hinz, “The layer-wise l1 loss landscape of neural nets is more complex around local minima,” 2021. [Online].Available: https://arxiv.org/abs/2105.02831

[22] H. S. Emadi and S. M. Mazinani, “A novel anomaly detection algorithm using dbscan and svm in wireless sensor networks,” Wireless Personal Communications, vol. 98, pp. 2025–2035, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:39736486

[23] Y. Meng, Y. Zhang, J. Huang, Y. Zhang, C. Zhang, and J. Han, “Hierarchical topic mining via joint spherical tree and text embedding,” in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining, ser. KDD ’20. ACM, Aug. 2020. [Online]. Available: http://dx.doi.org/10.1145/3394486.3403242

[24] S. Wibisono, M. Anwar, A. Supriyanto, and I. Amin, “Multivariate weather anomaly detection using dbscan clustering algorithm,” Journal of Physics: Conference Series, vol. 1869, p. 012077, 04 2021.

[25] J.S.Walker, Fast Fourier transforms. Boca Raton, Fla.: CRC Press, 1996. [Online]. Available: http://www.amazon.co.uk/gp/search?index=books&linkCode=qs&keywords=9780849371639

[26] N. Ahmed, T. Natarajan, and K. Rao, “Discrete cosine transform,” IEEE Transactions on Computers, vol. C-23, no. 1, pp. 90–93, 1974.

[27] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, “Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing,” 2021.

[28] M. West, “On scale mixtures of normal distributions,” Biometrika, vol. 74, no. 3, pp. 646–648, 1987. [Online]. Available: http://biomet.oxfordjournals.org/content/74/3/646.abstract

[29] I. Jolliffe, Principal Component Analysis.Springer Verlag, 1986.

[30] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, “Euclidean distance geometry and applications,” 2012.

[31] S. R. Blackburn, C. Homberger, and P. Winkler, “The minimum manhattan distance and minimum jump of permu-tations,” 2018.

[32] J. A. Hartigan and M. A. Wong, “A k-means clustering algorithm,” JSTOR: Applied Statistics, vol. 28, no. 1, pp. 100–108, 1979.

[33] P. Macgregor, “Fast and simple spectral clustering in theory and practice,” 2023.

[34] D.-y. Xia, F. Wu, X.-q. Zhang, and Y.-t. Zhuang, “Local and global approaches of affinity propagation clustering for large scale data,” Journal of Zhejiang University-SCIENCE A, vol. 9, no. 10, p. 1373–1381, Oct. 2008. [Online].Available: http://dx.doi.org/10.1631/jzus.A0720058

[35] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” 2021.

[36] D. P. Kingma and M. Welling, “An introduction to variational autoencoders,” Foundations and Trends® in Machine Learning, vol. 12, no. 4, p. 307–392, 2019. [Online]. Available: http://dx.doi.org/10.1561/2200000056

[37] X. Liu, F. Zhang, Z. Hou, Z. Wang, L. Mian, J. Zhang, and J. Tang, “Self-supervised Learning: Generative or Contrastive,” arXiv:2006.08218 [cs, stat], Jul. 2020, arXiv: 2006.08218. [Online]. Available: http://arxiv.org/abs/2006.08218

[38] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features.” CoRR, vol. abs/1807.05520, 2018. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1807.html#abs-1807-05520

[39] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” 2015.

[40] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” 2016.

[41] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” 2014.

[42] J. Zhao, M. Mathieu, R. Goroshin, and Y. LeCun, “Stacked what-where auto-encoders,” 2016.

[43] S. J. Wetzel, “Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders,” Physical Review E, vol. 96, no. 2, Aug. 2017. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.96.022140

[44] J. Zheng and Y. Liu, “A study on small-scale ship detection based on attention mechanism,” IEEE Access, vol. 10, Pp 77940–77949, 2022.

[45] B. Ye, T. Qin, H. Zhou, J. Lai, and X. Xie, “Cross-level attention and ratio consistency network for ship detection,” in 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 4644–4650.

[46] H. Cui, Y. Yang, M. Liu, T. Shi, and Q. Qi, “Ship detection: An improved yolov3 method,” in OCEANS 2019 -Marseille, 2019, pp. 1–4.

[47] T. Liu, B. Pang, S. Ai, and X. Sun, “Study on visual detection algorithm of sea surface targets based on improved yolov3,” Sensors, vol. 20, no. 24, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/24/7263

[48] H. Li, L. Deng, C. Yang, J. Liu, and Z. Gu, “Enhanced yolo v3 tiny network for real-time ship detection from visual image,” IEEE Access, vol. 9, pp. 16692–16706, 2021.

[49] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” 2016. [Online]. Available: https://arxiv.org/abs/1612.03144

[50] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 2018.

[51] S. Woo, J. Park, J.-Y. Lee, and I. Kweon, “Cbam: Convolutional block attention module: 15th european conference, munich, germany, september 8–14, 2018, proceedings, part vii,” 09 2018, pp. 3–19.

[52] T. Liu, B. Pang, L. Zhang, W. Yang, and X. Sun, “Sea surface object detection algorithm based on yolo v4 fused with reverse depthwise separable convolution (rdsc) for usv,” Journal of Marine Science and Engineering, vol. 9, no. 7, 2021. [Online]. Available: https://www.mdpi.com/2077-1312/9/7/753

[53] J. Guo, Y. Li, W. Lin, Y. Chen, and J. Li, “Network decoupling: From regular to depthwise separable convolutions,” 2018.

[54] X. Han, L. Zhao, Y. Ning, and J. Hu, “Shipyolo: An enhanced model for ship detection,” Journal of Advanced Transportation, vol. 2021, pp. 1–11, 06 2021.

[55] M. Zhang, X. Rong, and X. Yu, “Light-sdnet: A lightweight cnn architecture for ship detection,” IEEE Access, vol. 10, pp. 86647–86662, 2022.

[56] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, “Ghostnet: More features from cheap operations,” 2020.

[57] R. Ye, F. Liu, and L. Zhang, “3d depthwise convolution: Reducing model parameters in 3d vision tasks,” 2018.

[58] Q. Zhang, Y. Huang, and R. Song, “A ship detection model based on yolox with lightweight adaptive channel feature fusion and sparse data augmentation,” in 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2022, pp. 1–8.

[59] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8759–8768.

[60] Y. Zhang, M. J. Er, W. Gao, and J. Wu, “High performance ship detection via transformer and feature distillation,” in 2022 5th International Conference on Intelligent Autonomous Systems (ICoIAS), 2022, pp. 31–36.

[61] Z. Zhang, L. Zhang, Y. Wang, P. Feng, and R. He, “Shiprsimagenet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 8458–8472, 2021.

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

[63] N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” in Proc. of the 37-th Annual Allerton Conference on Communication, Control and Computing, 1999, pp. 368–377. [Online]. Available: https://arxiv.org/abs/physics/0004057

[64] C. E. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, pp. 379–423, 1948. [Online]. Available: http://plan9.bell labs.com/cm/ms/what/shannonday/shannon1948.pdf

[65] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, “Deep variational information bottleneck,” 2016. [Online]. Available: http://arxiv.org/abs/1612.00410

[66] “Seaship dataset,” 10, 20, 2022. [Online]. Available: https://www.kaggle.com/datasets/tangwenyang/seaship

[67] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[68] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” 2021.

[69] M. Lueken, W. ten Kate, J. P. Batista, C. Ngo, C. Bollheimer, and S. Leonhardt, “Peak detection algorithm for gait segmentation in long-term monitoring for stride time estimation using inertial measurement sensors,” in 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), 2019, pp. 1–4.

[70] L. Du, L. Li, B. Wang, and J. Xiao, “Micro-doppler feature extraction based on time-frequency spectrogram for ground moving targets classification with low-resolution radar,” IEEE Sensors Journal, vol. 16, pp. 1–1, 05 2016.

[71] M. E. Demirhan and O. Salor, “Classification of targets in sar images using svm and k-nn techniques,” in 2016 24th Signal Processing and Communication Application Conference (SIU), 2016, pp. 1581–1584.

[72] M. Ren, J. Cai, Y. Zhu, and M. He, “Radar emitter signal classification based on mutual information and fuzzy support vector machines,” in 2008 9th International Conference on Signal Processing, 2008, pp. 1641–1646.

[73] C. Wang, J. Pei, R. Wang, Y. Huang, and J. Yang, “A new ship detection and classification method of spaceborne sar images under complex scene,” in 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), 2019, pp1–4.

[74] U. Kaydok, “Chaff discrimination using convolutional neural networks and range profile data,” in 2020 IEEE International Radar Conference (RADAR), 2020, pp. 373–377.

[75] S. Rajkamal, “Selecting reviewers for research by clustering proposals using expectation maximization clustering algorithm,” in 2017 International Conference on Technical Advancements in Computers and Communications (IC-TACC), 2017, pp. 56–60.

[76] A. E. Vincent and K. Sreekumar, “A survey on approaches for ecg signal analysis with focus to feature extraction and classification,” in 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 140–144.

[77] S.-S. Kim and T. Kasparis, “A modified domain deformation theory on 1-d signal classification,” IEEE Signal Processing Letters, vol. 5, no. 5, pp. 118–120, 1998.

[78] M. S. Azmi, N. A. Arbain, A. K. Muda, Z. A. Abas, and Z. Muslim, “Data normalization for triangle features by adapting triangle nature for better classification,” in 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2015, pp. 1–6.