Jyothi Hills, Panjal Road,
Vettikattiri PO, Cheruthuruthy , Thrissur,
Kerala 679531

NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL

Jyothi Engineering College
NAAC Accredited College with NBA Accredited Programmes*
Approved by AICTE & affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKATTIRI P.O, CHERUTHURUTHY, THRISSUR. PIN-679531 PH : +91- 4884-259000, 274423 FAX : 04884-274777
NBA accredited B.Tech Programmes in Computer Science & Engineering, Electronics & Communication Engineering, Electrical & Electronics Engineering and Mechanical Engineering valid for the academic years 2016-2022. NBA accredited B.Tech Programme in Civil Engineering valid for the academic years 2019-2022.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SEMINAR REPORT

Ship Detection Using YOLOv3

Submitted by

NITHIN PETER
JEC17CS075

Supervised by

Mr. ANIL ANTONY
Assistant Prof., Dept. of CSE

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY (B.Tech)

in

COMPUTER SCIENCE & ENGINEERING

of

A P J ABDUL KALAM TECHNOLOGICAL UNIVERSITY

CREATING TECHNOLOGY LEADERS OF TOMORROW

DECEMBER 2020

NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL

Jyothi Hills, Panjal Road,
Vettikattiri PO, Cheruthuruthy , Thrissur,
Kerala 679531

Jyothi Engineering College
NAAC Accredited College with NBA Accredited Programmes*
Approved by AICTE & affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKATTIRI P.O, CHERUTHURUTHY, THRISSUR. PIN-679531 PH : +91- 4884-259000, 274423 FAX : 04884-274777
NBA accredited B.Tech Programmes in Computer Science & Engineering, Electronics & Communication Engineering, Electrical & Electronics Engineering and Mechanical Engineering valid for the academic years 2016-2022. NBA accredited B.Tech Programme in Civil Engineering valid for the academic years 2019-2022.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SEMINAR REPORT

Ship Detection Using YOLOv3

Submitted by

NITHIN PETER
JEC17CS075

Supervised by

Mr. ANIL ANTONY
Assistant Prof., Dept. of CSE

---

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY (B.Tech)

in

COMPUTER SCIENCE & ENGINEERING
of

A P J ABDUL KALAM TECHNOLOGICAL UNIVERSITY

CREATING TECHNOLOGY LEADERS OF TOMORROW

DECEMBER 2020

Department of Computer Science and Engineering

JYOTHI ENGINEERING COLLEGE, CHERUTHURUTHY

THRISSUR 679 531

DECEMBER 2020

BONAFIDE CERTIFICATE

This is to certify that the seminar report entitled **Ship Detection Using YOLOv3** submitted by **Nithin Peter (JEC17CS075)** in partial fulfillment of the requirements for the award of **Bachelor of Technology** degree in **Computer Science and Engineering** of **A P J Abdul Kalam Technological University** is the bonafide work carried out by him under our supervision and guidance.

Mr. Anil Antony	**Mr. Shaiju Paul**	**Dr. Vinith R**
Seminar Guide	Seminar Coordinator	Head of The Dept
Assistant Professor	Assistant Professor	Professor
Dept. of CSE	Dept. of CSE	Dept. of CSE

Jyothi Engineering College
NAAC Accredited college with NBA Accredited programmes*
Approved by AICTE & Affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKKATTIRI P.O., CHERUTHURUTHY, THRISSUR PIN-679531 PH: +91-4884-259000, 274423 FAX : 04884-274777
www.jecc.ac.in | E-mail : info@jecc.ac.in

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

COLLEGE VISION

Creating eminent and ethical leaders through quality professional education with emphasis on holistic excellence.

COLLEGE MISSION

- To emerge as an institution par excellence of global standards by imparting quality engineering and other professional programmes with state-of- the-art facilities.
- To equip the students with appropriate skills for a meaningful career in the global scenario.
- To inculcate ethical values among students and ignite their passion for holistic excellence through social initiatives.
- To participate in the development of society through technology incubation, entrepreneurship and industry interaction.

Jyothi Engineering College
NAAC Accredited college with NBA Accredited programmes*
Approved by AICTE & Affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKKATTIRI P.O., CHERUTHURUTHY, THRISSUR PIN-679531 PH: +91-4884-259000, 274423 FAX : 04884-274777
www.jecc.ac.in | E-mail : info@jecc.ac.in

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

DEPARTMENT VISION

Creating eminent and ethical leaders in the domain of computational sciences through quality professional education with a focus on holistic learning and excellence.

DEPARTMENT MISSION

- To create technically competent and ethically conscious graduates in the field of Computer Science & Engineering by encouraging holistic learning and excellence.
- To prepare students for careers in Industry, Academia and the Government.
- To instill Entrepreneurial Orientation and research motivation among the students of the department.
- To emerge as a leader in education in the region by encouraging teaching, learning, industry and societal connect

PROGRAMME OUTCOMES (POs)

1. **Engineering Knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem Analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/Development of Solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations of Complex Problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern Tool Usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The Engineer and Society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and Team Work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project Management and Finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-Long Learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

1. The graduates shall have sound knowledge of Mathematics, Science, Engineering and Management to be able to offer practical software and hardware solutions for the problems of industry and society at large.
2. The grad uates shall be able to establish themselves as practising professionals, researchers or Entrepreneurs in computer science or allied areas and shall also be able to pursue higher education in reputed institutes.
3. The graduates shall be able to communicate effectively and work in multidisciplinary teams with team spirit demonstrating value driven and ethical leadership.

Programme Specific Outcomes (PSOs)

1. An ability to apply knowledge of data structures and algorithms appropriate to computational problems.
2. An ability to apply knowledge of operating systems, programming languages, data management, or networking principles to computational assignments.
3. An ability to apply design, development, maintenance or evaluation of software engineering principles in the construction of computer and software systems of varying complexity and quality.
4. An ability to understand concepts involved in modeling and design of computer science applications in a way that demonstrates comprehension of the fundamentals and tradeoffs involved in design choices.

Course Outcomes (COs)

C418.1 **Presentation Skills in terms of Content** : Students will be able to show competence in identifying relevant information, defining and explaining topics under discussion. They will demonstrate depth of understanding, use primary and secondary sources; they will demonstrate the working, complexity, insight, cogency, independent thought, relevance, and persuasiveness. They will be able to evaluate information and use and apply relevant theories.

C418.2 **Presentation Skills in terms of Organization** : Students will be able to show competence in working with a methodology, structuring their oral work, and synthesizing information. They will make a detailed study on the previous works related to their topic and will present the observations.

C418.3 **Presentation Skills in terms of Delivery** : Students will use appropriate registers and vocabulary, and will demonstrate command of voice modulation, voice projection, and pacing. They will be able to make use of visual, audio and audio-visual material to support their presentation, and will be able to speak cogently with or without notes.

C418.4 **Discussion Skills** : Students will be able to judge when to speak and how much to say, speak clearly and audibly in a manner appropriate to the subject, ask appropriate questions, use evidence to support claims, respond to a range of questions, take part in meaningful discussion to reach a shared understanding, speak with or without notes, show depth of understanding.

C418.5 **Listening Skills** : Students will demonstrate that they have paid close attention to what others say and can respond constructively. Through listening attentively, they will be able to build on discussion fruitfully, supporting and connecting with other discussants.

C418.6 **Argumentative Skills and Critical Thinking** : Students will develop persuasive speech, present information in a compelling, well-structured, and logical sequence, respond respectfully to opposing ideas, show depth of knowledge of complex subjects, and develop their ability to synthesize, evaluate and reflect on information.

Course Outcome							
Programme Outcomes		C418.1	C418.2	C418.3	C418.4	C418.5	C418.6
	1	3	3	3	3	3	3
	2	3	3	3	3	3	3
	3	3	3	3	3	3	3
	4	3	3	3	3	3	3
	5	3	3	3	3	3	3
	6	3	3	3	3	3	3
	7	3	3	3	3	3	3
	8	3	3	3	3	3	3
	9	3	3	3	3	3	3
	10	3	3	3	3	3	3
	11	3	3	3	3	3	3
	12	3	3	3	3	3	3

PO - CO Mapping

PEO - CO Mapping

Course Outcome							
Programme Educational Objective		C418.1	C418.2	C418.3	C418.4	C418.5	C418.6
	1	3	3	1	1	-	2
	2	3	3	3	3	1	3
	3	1	2	3	3	1	3

PSO - CO Mapping

Course Outcome							
Programme Specific Outcomes		C418.1	C418.2	C418.3	C418.4	C418.5	C418.6
	1	3	3	3	3	3	3
	2	3	3	3	3	3	3
	3	3	3	3	3	3	3
	4	3	3	3	3	3	3

Seminar Outcome

1. Studied about the concept of Deep Learning.
2. Studied about different neural networks.
3. Analyzed and compared the general architecture of CNN,RNN and CRNN.
4. Studied about different ship detection methods.
5. Analyzed the methodology of ship detection using YOLOv3.

Seminar Outcome - CO Mapping

	Course Outcome						
		C418.1	C418.2	C418.3	C418.4	C418.5	C418.6
Seminar Outcome	1	3	3	3	1	3	3
	2	3	3	1	1	3	3
	3	3	3	3	1	3	1
	4	3	3	3	3	1	1
	5	3	1	3	3	1	1

ACKNOWLEDGEMENT

I take this opportunity to express my heartfelt gratitude to all respected personalities who had guided, inspired and helped me in the successful completion of this seminar. First and foremost, I express my thanks to **The Lord Almighty** for guiding me in this endeavour and making it a success.

I take immense pleasure in thanking the **Management** of Jyothi Engineering College and **Dr. Sunny Joseph Kalayathankal**, Principal, Jyothi Engineering College for having permitted me to carry out this seminar. My sincere thanks to **Dr. Vinith R**, Head of the Department of Computer Science and Engineering for permitting me to make use of the facilities available in the department to carry out the seminar successfully.

I express my sincere gratitude to **Mr. Shaiju Paul & Dr. Swapna B Sasi**, Seminar Coordinators for their invaluable supervision and timely suggestions. I am very happy to express my deepest gratitude to my mentor **Mr. Anil Antony**, Assistant Professor,Department of Computer Science and Engineering, Jyothi Engineering College for his able guidance and continuous encouragement.

Last but not least, I extend my gratefulness to all teaching and non-teaching staff who directly or indirectly involved in the successful completion of this seminar work and to all friends who have patiently extended all sorts of help for accomplishing this undertaking.

ABSTRACT

Ship detection plays an important role in marine transportation, fishery management, and maritime disaster rescue. Nowadays, the current researches almost are focusing on improving detection accuracy while detection speed is neglected. However, it is also extraordinarily important to increase the ship detection speed, because it can provide real-time ocean observation and timely ship rescue. Boat identification is also a very important and challenging task in maritime traffic monitoring, the difficulty of this task lies in the accurate positioning and identification of relatively small boats in complex scenes. Synthetic aperture radar (SAR) imagery has been used as a promising data source for monitoring maritime activities, and its application for oil and ship detection has been the focus of many previous research studies. Many object detection methods ranging from traditional to deep learning approaches have been proposed. However, majority of them are computationally intensive and have accuracy problems. The huge volume of the remote sensing data also brings a challenge for real time object detection. YOLO is a network for object detection. The detection task consists in determining the location on the image where certain objects are present, and classifying those objects. Previous methods for this, like R-CNN and its variations,used a pipeline to perform this task in multiple steps. This may be slow to run and also hard to optimize because each individual component must be trained separately. YOLO does it all with one neural network. You Only Look Once v3 (YOLOv3) can be optimised to be used for ship detection. This algorithm makes it possible to do real time ship detection in an efficient manner.

Keywords - Detection, Identification, Synthetic aperture radar, You Only Look Once, Convolutional neural networks.

CONTENTS

List of Figures

List of Abbreviations

CNN	: *Convolutional Neural Network*
RNN	: *Recurrent Neural Network*
CRNN	: *Convolutional Recurrent Neural Network*
SAR	: *Synthetic Aperture Radar*
YOLO	: *You Only Look Once*
CRNN	: *Convolutional Recurrent Neural Network*
IOU	: *Intersection Over Union*
PSI	: *Persistent Scatterer Interferometry*

CHAPTER 1
INTRODUCTION

1.1 Overview

With the development of economic globalization, maritime transportation is becoming more and more frequent. Due to complex shipping scenes, boat crossing, fog shielding and other problems, traffic accidents are constantly occurring. Ship detection has great demands in civil and military fields.For example, in the civil field, ship detection can supervise transportation, marine traffics and illegal smuggling. In the military field, one can monitor for cross-border smuggling or other illegal behaviors. However, traditional ship detection is based on naked eyes monitoring, which causes huge labor costs.

Computer-aided detection method greatly saves the labor cost and improves detection efficiency at the same time. Nowadays, with the development of artificial intelligence technology, more and more experts begin to study the methods of ship detection based on data-driven and artificial intelligence. Probably a simple explanation to this fact comes from that artificial intelligence methods can automatically extract ship's features, avoiding the manual feature engineering of traditional methods, which greatly improves the detection efficiency.

Synthetic aperture radar (SAR), an all-weather and all-time microwave sensor, is one of the most important tools in remote sensing filed. Up to now, many scholars have proposed many SAR ship detection methods, which have greatly promoted the development of SAR image interpretation. According to my survey, in SAR ship detection field, many scholars are focusing on improving the accuracy of ship detection while the detection speed is neglected. In fact, it is also extraordinarily important to increase the ship detection speed, because it can provide real-time ocean observation and timely ship rescue.

In this paper we will look at various modifications that can be made to the YOLOv3 algorithm to optimize it for ship detection specifically. We will look at methods that not only improve the accuracy but also increase the speed of detection.

1.1.1 Artificial Intelligence

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal.

Artificial intelligence is based on the principle that human intelligence can be defined in a way that a machine can easily mimic it and execute tasks, from the most simple to those that are even more complex. The goals of artificial intelligence include learning, reasoning, and perception.

As technology advances, previous benchmarks that defined artificial intelligence become outdated. For example, machines that calculate basic functions or recognize text through optimal character recognition are no longer considered to embody artificial intelligence, since this function is now taken for granted as an inherent computer function.

AI is continuously evolving to benefit many different industries. Machines are wired using a cross-disciplinary approach based in mathematics, computer science, linguistics, psychology, and more.

Types of AI:

- Reactive Machines AI: Based on present actions, it cannot use previous experiences to form current decisions and simultaneously update their memory. Example: Deep Blue

- Limited Memory AI: Used in self-driving cars. They detect the movement of vehicles around them constantly and add it to their memory.

- Theory of Mind AI: Advanced AI that has the ability to understand emotions, people and other things in the real world.

- Self Aware AI: AIs that posses human-like consciousness and reactions. Such machines have the ability to form self-driven actions.

- Artificial Narrow Intelligence (ANI): General purpose AI, used in building virtual assistants like Siri

- Artificial General Intelligence (AGI): Also known as strong AI. An example is the Pillo robot that answers questions related to health.

- AI that possesses the ability to do everything that a human can do and more. An example is the Alpha 2 which is the first humanoid ASI robot.

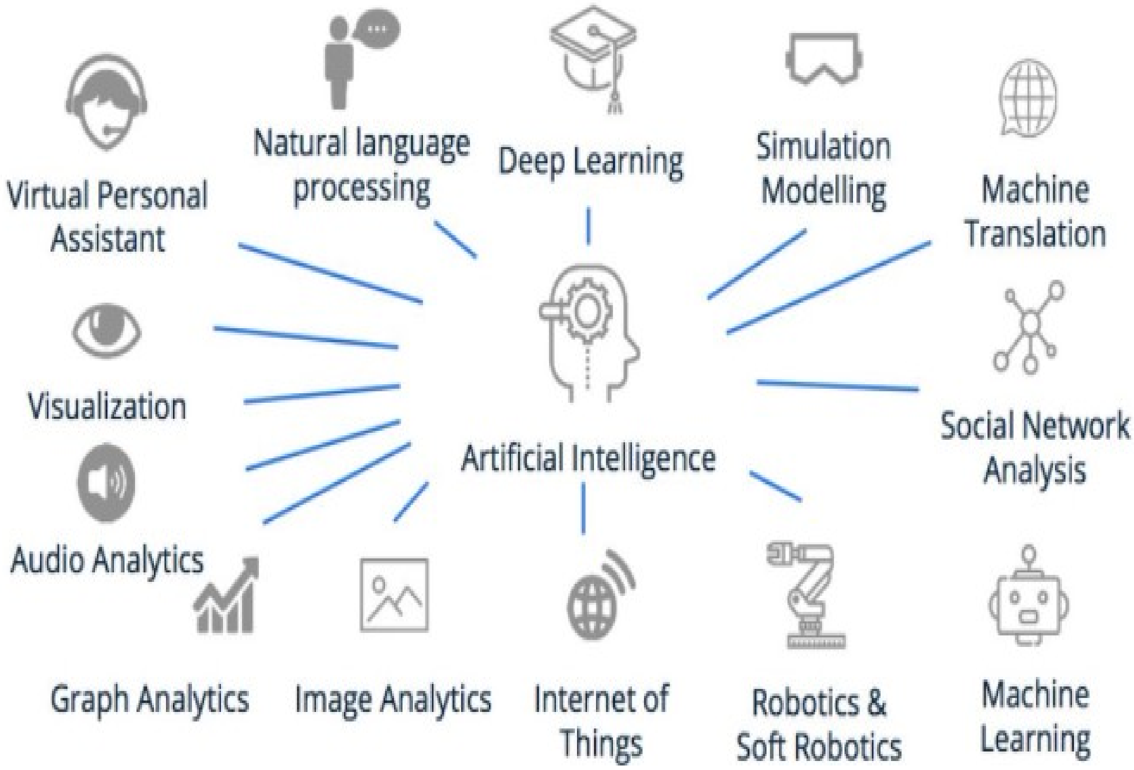


Figure 1.1: Artificial intelligence

1.1.2 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Machine learning algorithms are often categorized as supervised or unsupervised. Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training data set, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from data sets to describe hidden structures from unlabeled data.

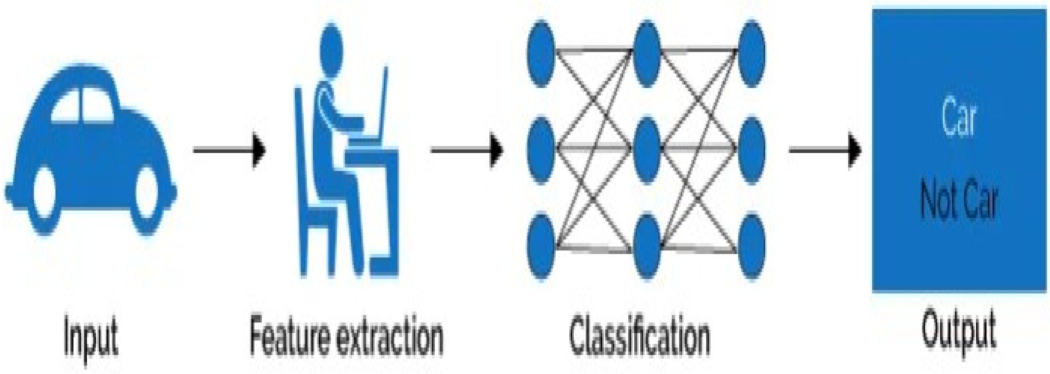


Figure 1.2: Machine Learning

1.1.3 Deep Learning

Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

Deep learning has evolved hand-in-hand with the digital era, which has brought about an explosion of data in all forms and from every region of the world. This data, known simply as big data, is drawn from sources like social media, internet search engines, e-commerce platforms, and online cinemas, among others. This enormous amount of data is readily accessible and can be shared through fintech applications like cloud computing.

However, the data, which normally is unstructured, is so vast that it could take decades for humans to comprehend it and extract relevant information. Companies realize the incredible potential that can result from unraveling this wealth of information and are increasingly adapting to AI systems for automated support

In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. In an image recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face. Importantly, a deep learning process can learn which features to optimally place in which level on its own.

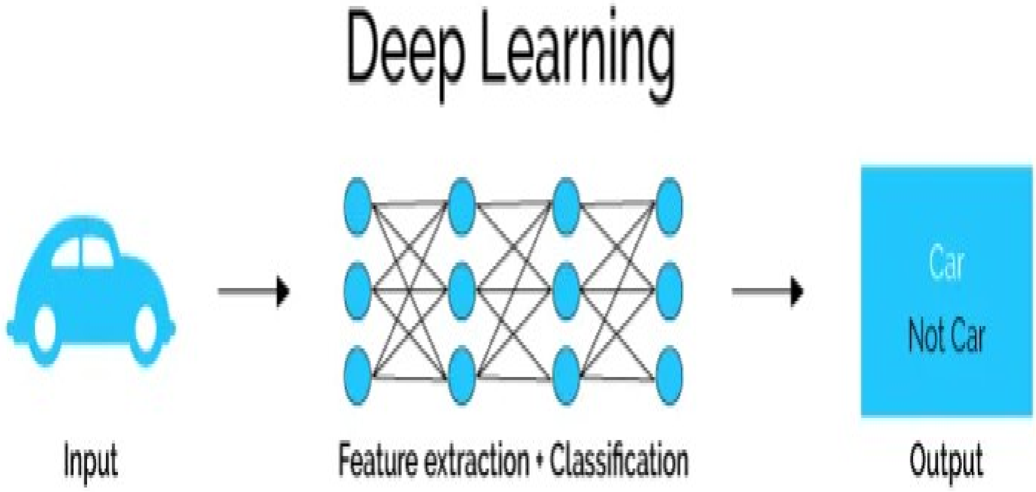


Figure 1.3: Deep Learning

1.1.4 Artificial Neural Networks

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data becomes available.

An ANN has hundreds or thousands of artificial neurons called processing units, which are interconnected by nodes. These processing units are made up of input and output units. The input units receive various forms and structures of information based on an internal weighting system, and the neural network attempts to learn about the information presented to produce one output report

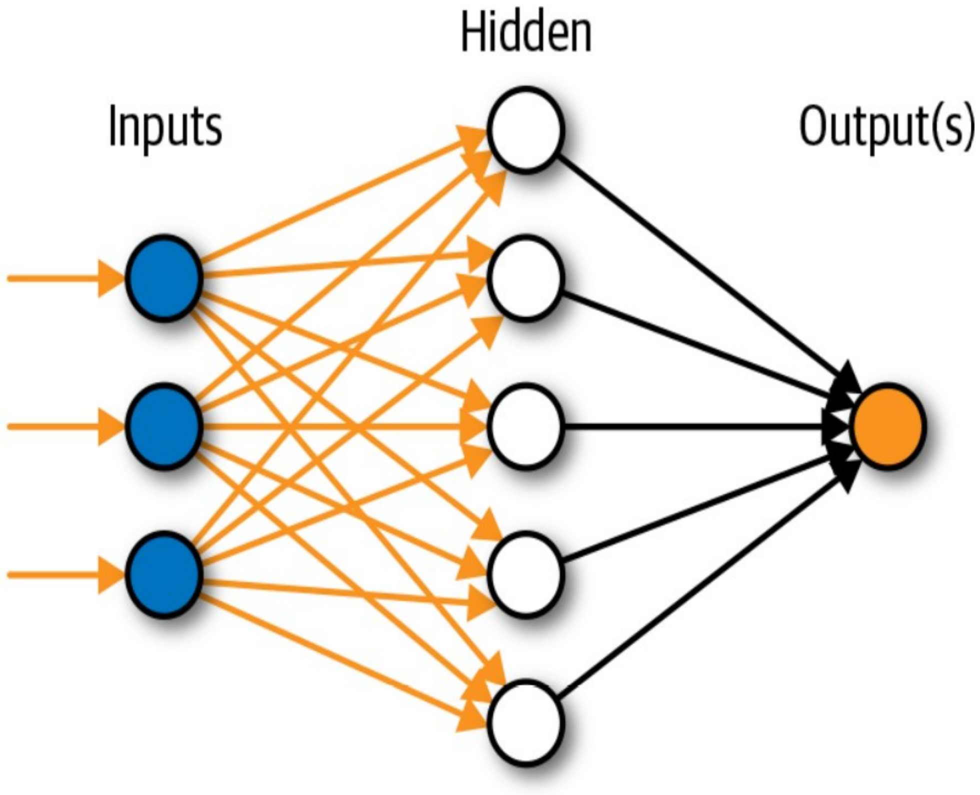


Figure 1.4: Artificial Neural Networks

1.1.5 Recurrent Neural Networks

Recurrent Neural Network(RNN) are a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

RNN have a "memory" which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks. [8]

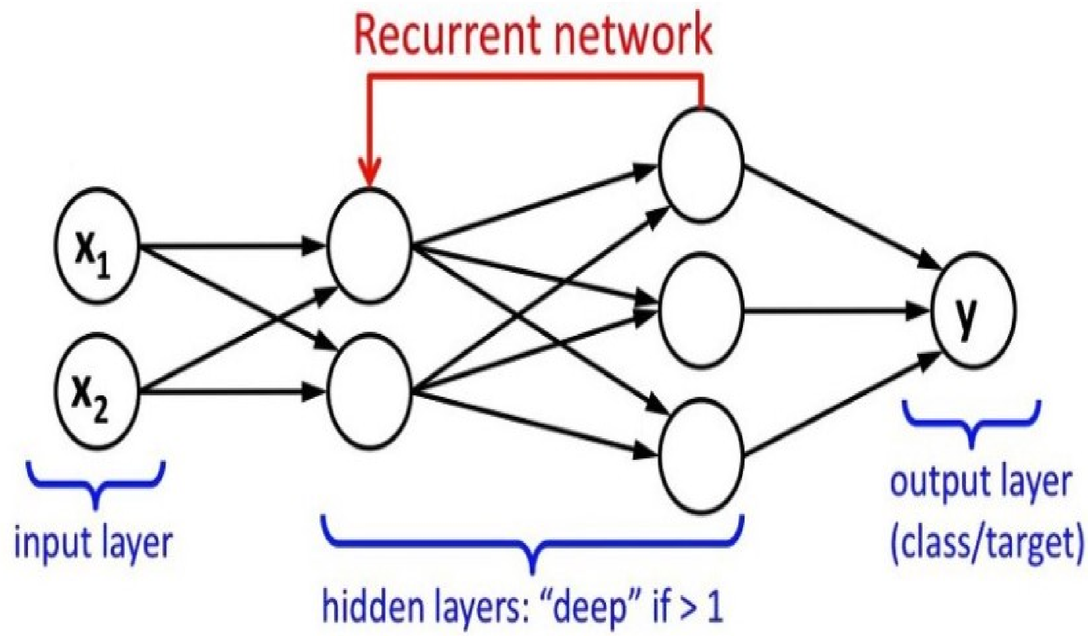


Figure 1.5: Recurrent Neural Networks

1.1.6 Convolutional Neural Networks

A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks.

CNNs are a fundamental example of deep learning, where a more sophisticated model pushes the evolution of artificial intelligence by offering systems that simulate different types of biological human brain activity.

CNNs are regularized versions of multilayer perceptrons. Multiplier perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to over-fitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme. [3]

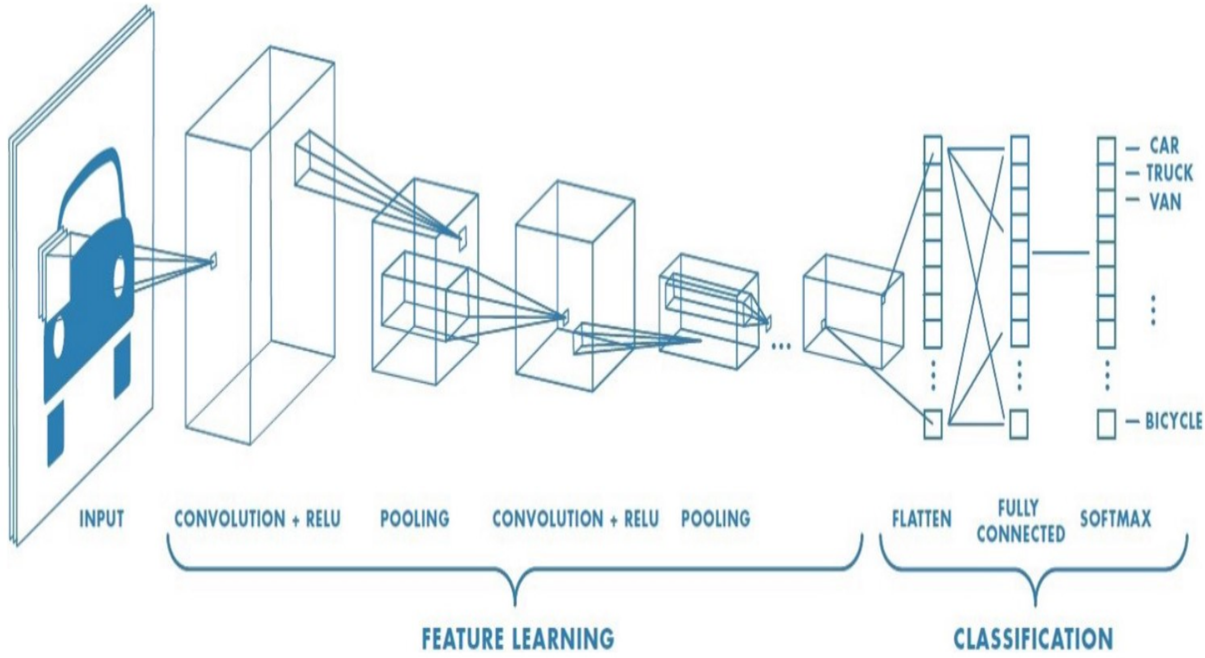


Figure 1.6: Convolutional Neural Networks

1.1.7 Object Detection

Object detection is a computer vision technique that works to identify and locate objects within an image or video. Specifically, object detection draws bounding boxes around these detected objects, which allow us to locate where said objects are in (or how they move through) a given scene.

Object detection is commonly confused with image recognition, so before we proceed, it's important that we clarify the distinctions between them. Image recognition assigns a label to an image. A picture of a dog receives the label "dog". A picture of two dogs, still receives the label "dog". Object detection, on the other hand, draws a box around each dog and labels the box "dog" as shown in fig.1.9. The model predicts where each object is and what label should be applied. In that way, object detection provides more information about an image than recognition.

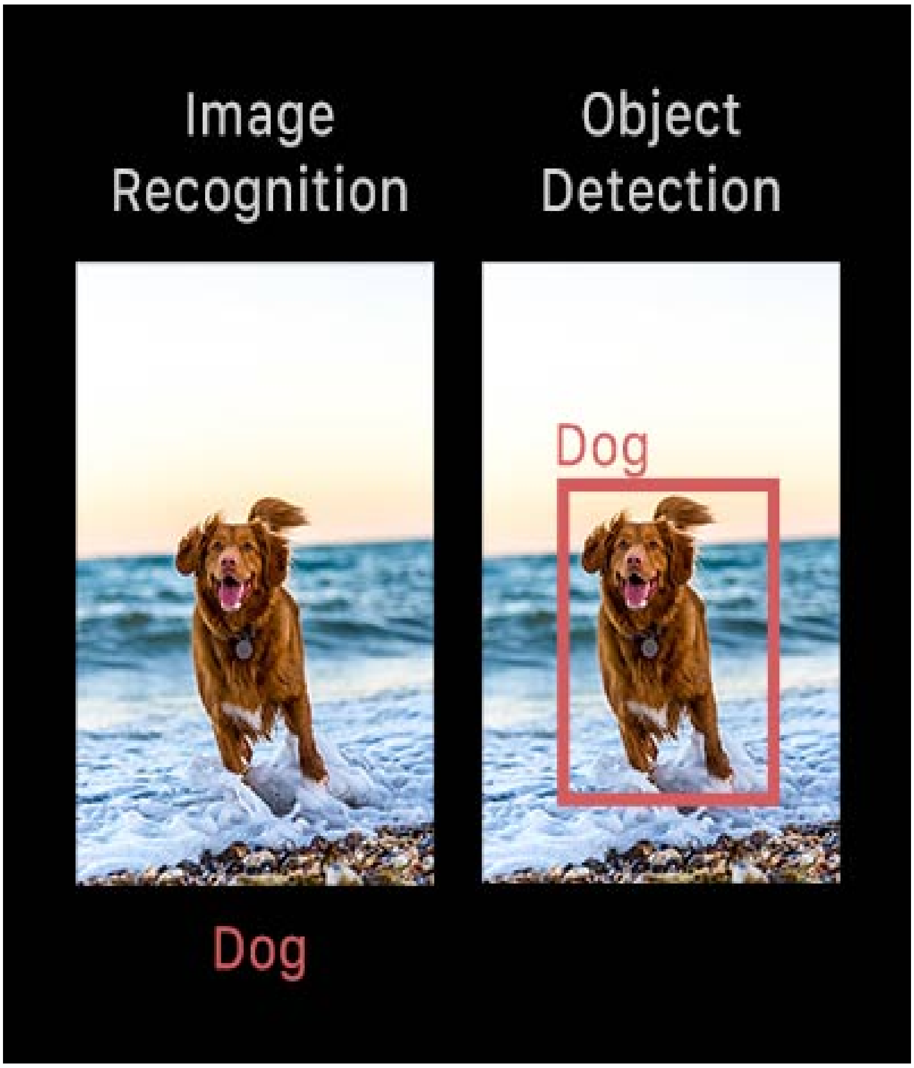


Figure 1.7: Object Detection VS Image Recognition

Broadly speaking, object detection can be broken down into machine learning-based approaches and deep learning-based approaches. In more traditional ML-based approaches, computer vision techniques are used to look at various features of an image, such as the color histogram or edges, to identify groups of pixels that may belong to an object. These features are then fed into a regression model that predicts the location of the object along with its label.

On the other hand, deep learning-based approaches employ convolutional neural networks (CNNs) to perform end-to-end, unsupervised object detection, in which features don't need to be defined and extracted separately. [6]

Object detection is inextricably linked to other similar computer vision techniques like image recognition and image segmentation, in that it helps us understand and analyze scenes in images or video. But there are important differences. Image recognition only outputs a class label for an identified object, and image segmentation creates a pixel-level understanding of a scene's elements. What separates object detection from these other tasks is its unique ability to locate objects within an image or video. This then allows us to count and then track those objects.

Given these key distinctions and object detection's unique capabilities, we can see how it can be applied in a number of ways. A few of them are :

- Crowd counting
- Self-driving cars
- Video surveillance
- Face detection
- Anomaly detection
- Ship Detection

1.1.8 Synthetic-aperture radar

Synthetic aperture radar (SAR) refers to a technique for producing fine-resolution images from a resolution-limited radar system. It requires that the radar be moving in a straight line, either on an airplane or, as in the case of NISAR, orbiting in space.

The basic principle of any imaging radar is to emit an electromagnetic signal (which travels at the speed of light) toward a surface and record the amount of signal that bounces/echoes back, or "backscatters," and its time delay. The resulting radar imagery is built up from the strength and time delay of the returned signal, which depends primarily on the roughness and electrical conducting properties of the observed surface and its distance from the orbiting radar. [4]

The wavelengths that remote sensing radars use to observe Earth's surface are microwaves, typically in the range of a few to tens of centimeters. Because the radar signal loses energy as it travels – at a rate equivalent to the beam width (wavelength / antenna size) – by the time it hits the surface, the beam has spread dramatically. For example, with a signal wavelength of 10 centimeters and an antenna of 10 meters in diameter, the beam width is 1/100 radians (0.6 degrees). From an altitude of 1,000 kilometers, the resulting beam width on the ground becomes a very large 10 km, producing an image resolution which is insufficient for most applications. SAR is the solution to this dilemma as it can vastly improve the resolution.

SAR techniques take advantage of the fact that the radar is moving in orbit to synthesize a virtual 10-km-long antenna from the physical 10-m antenna in the direction of flight. As the radar moves along its path, it sweeps the antenna's footprint across the ground while continuously transmitting pulses – short signal bursts separated by time – and receiving the echoes of the returned pulses.

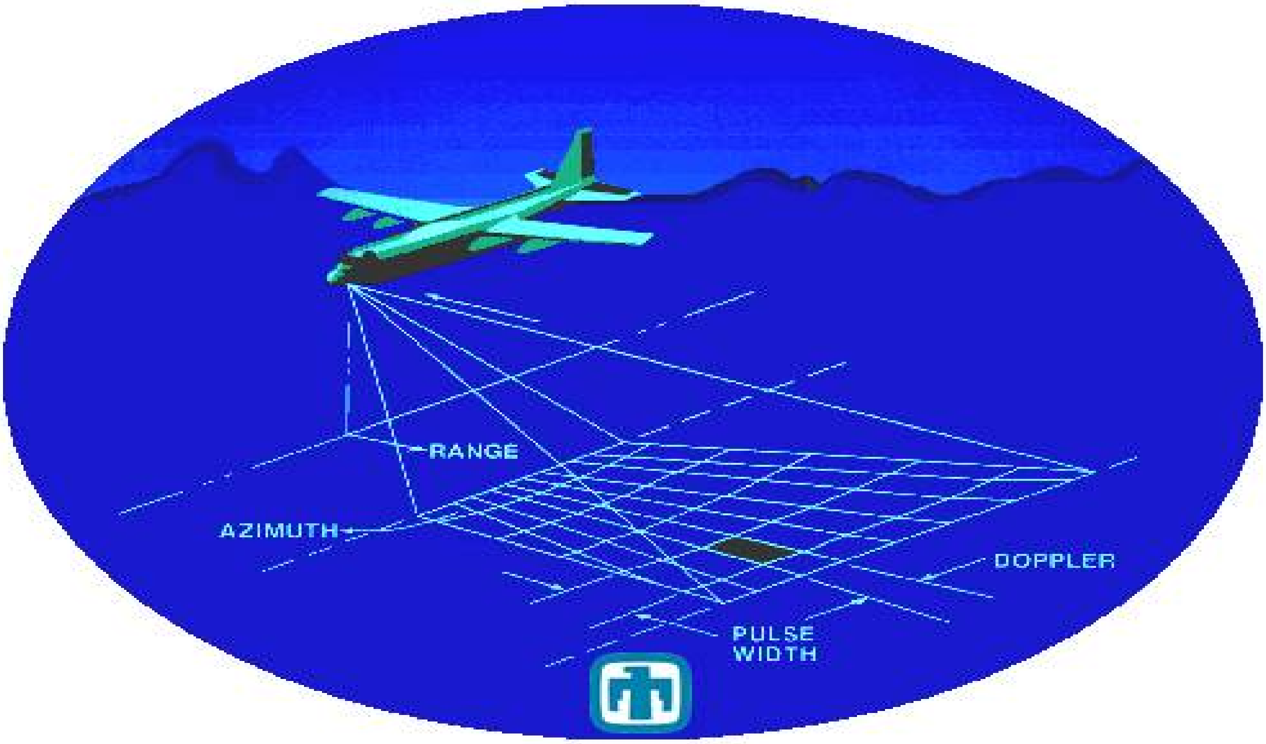


Figure 1.8: Synthetic-aperture radar

To create a SAR image, successive pulses of radio waves are transmitted to "illuminate" a target scene, and the echo of each pulse is received and recorded. The pulses are transmitted and the echoes received using a single beam-forming antenna, with wavelengths of a meter down to several millimeters. As the SAR device on board the aircraft or spacecraft moves, the antenna location relative to the target changes with time. Signal processing of the successive recorded radar echoes allows the combining of the recordings from these multiple antenna positions. This process forms the synthetic antenna aperture and allows the creation of higher-resolution images than would otherwise be possible with a given physical antenna. [1]

Figure 1.9: Synthetic-aperture radar image

1.2 Objective

- Learn about YOLOv3 algorithm
- Use YOLOv3 for ship detection
- Study various modifications that can be made to YOLOv3
- Optimize YOLOv3 to have Better accuracy and speed for ship detection

1.3 Organization Of The Report

The report is organised as follow:

- **Chapter 1: Introduction**
  Gives an Introduction about ship detection and the various technologies used for it.

- **Chapter 2: Literature Survey**
  Summarises the different ship detection techniques and technologies.

- **Chapter 3: YOLOv3-SHIP**
  Specifies details of modifications done to YOLOv3 so that it can be optimized for ship detection.

- **Chapter 4: High-Speed Ship Detection By Improved YOLOv3**
  Details modification to YOLOv3 to improve its speed in ship detection.

- **Chapter 5: YOLO-H Based on Improved Loss Function**
  Details changes to the Loss Function of YOLOv3 to specifically for ship detection.

- **Chapter 6: Conclusion**
  Summarizes the different changes to YOLOv3 to optimize it for ship detection.

- **References**
  Reference papers are included for use in the future.

CHAPTER 2

LITERATURE SURVEY

2.1 You Only Look Once: Unified, Real-Time Object Detection

We unify the separate components of object detection into a single neural network. Our network uses features from the entire image to predict each bounding box. It also predicts all bounding boxes across all classes for an image simultaneously. This means our network reasons glob-ally about the full image and all the objects in the image.The YOLO design enables end-to-end training and real-time speeds while maintaining high average precision.Our system divides the input image into an S×S grid.If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object.

Each grid cell predicts B bounding boxes and confidence scores for those boxes. These confidence scores reflect how confident the model is that the box contains an object and also how accurate it thinks the box that it predicts is.

If no object exists in the cell, the confidence scores should be zero. Otherwise we want the confidence score to equal the intersection over union (IOU) between the predicted box and the ground truth. Each bounding box consists of 5 predictions: x,y,w,h and confidence. The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. Finally the confidence prediction represents the IOU between the predicted box and any ground truth box. [7]

2.1.1 Network Design

We implement this model as a convolutional neural net-work and evaluate it on the PASCAL VOC detection data-set. The initial convolutional layers of the network extract features from the image while the fully connected layers predict the output probabilities and coordinates.Our network architecture is inspired by the GoogLeNet model for image classification . Our network has 24 convolutional layers followed by 2 fully connected layers.Instead of the

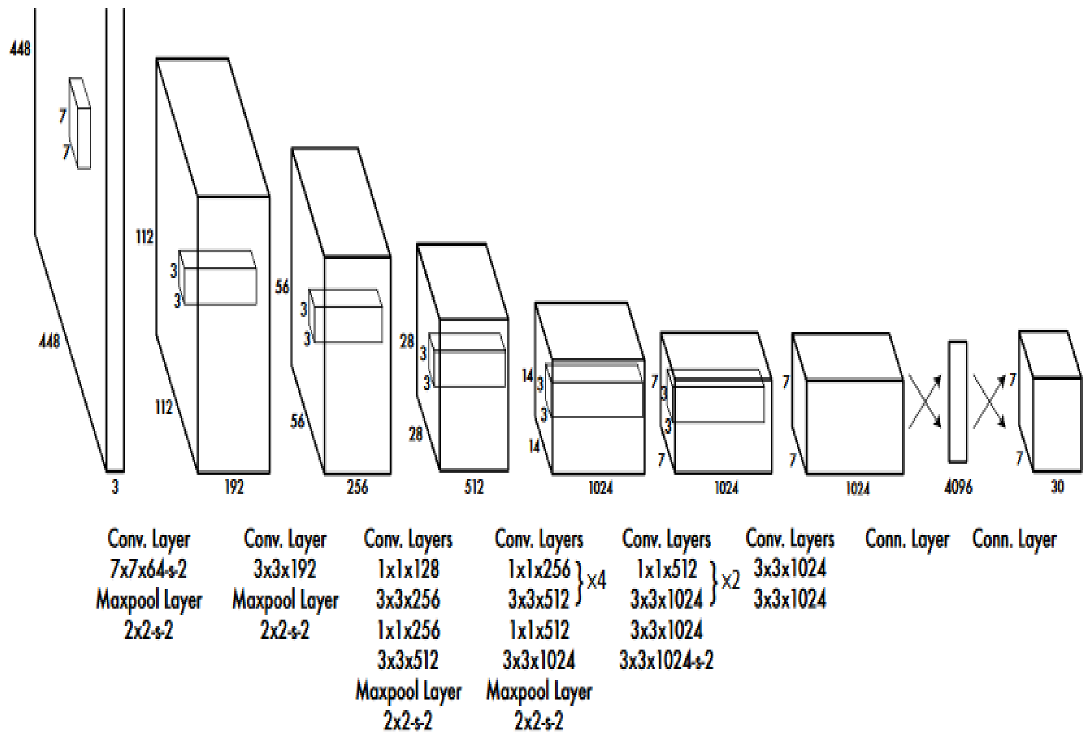


Figure 2.1: The Architecture.

inception modules used by GoogLeNet, we simply use 1×1 reduction layers followed by 3×3 convolutional layers, similar to Lin et al .We also train a fast version of YOLO designed to push the boundaries of fast object detection. Fast YOLO uses a neural network with fewer convolutional layers (9 instead of 24) and fewer filters in those layers. Other than the size of the network, all training and testing parameters are the same between YOLO and Fast YOLO. The final output of our network is the 7×7×30 tensor of predictions.

2.2 CNN Based Re-normalization Method For Detection In VHR Remote Sensing Images

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area. A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better. [9]

2.2.1 Model

The proposed method named CNN based re-normalization method. Localization and classification are separately performed. Main framework of the proposed method is given.Input image is partitioned into patches by sliding window. At the first stage, the posture parameters (the rotate angle , scale and shifting Xoff and Yoff) are predicted by CNN.In R-CNN,Bounding-box regression method is used to predict the scale and shifting. We also use Bounding-box regression method and an Alexnet to predict the location parameters. Rotate angle is predicted by another Alexnet. We can get the posture parameters of ships in patches of input image if the uniformed patches are ships. At last, according to the posture parameters and the origin windows, we can calculate the bounding boxes by a feedback algorithm which re-scale, relocate and rotate the origin window's bounding box to get target's bounding box. The proposed method finally gets the bounding boxes of ships. Unlike normal CNN, the bounding boxes may have different rotation and more fit the ship's edge and shapes.The whole framework of ship detection which includes the proposed re-normalization and final classification is implement with Caffe. Recently,some deep neural networks have been developed for object detection, such as Overfeat and STN. Compared with Overfeat and STN, main merit of the proposed method consist in rectangle object such as ships can be detected using more proper bounding boxes

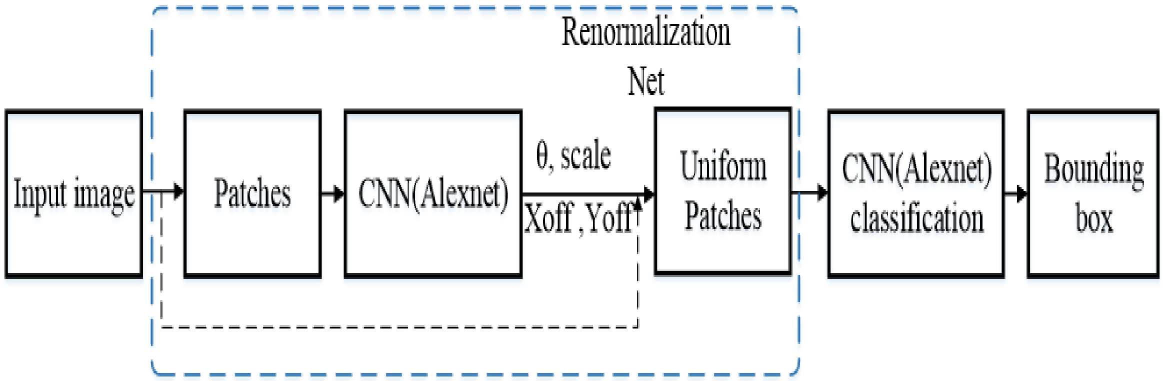


Figure 2.2: CNN-based re-normalization method for ship detection

2.3 Enhanced Resolution in SAR/ISAR Imaging Using Iterative Sidelobe Apodization

Synthetic-aperture radar (SAR) is a form of radar that is used to create two-dimensional images or three-dimensional reconstructions of objects, such as landscapes. SAR uses the motion of the radar antenna over a target region to provide finer spatial resolution than conventional beam-scanning radars. SAR is typically mounted on a moving platform, such as an aircraft or spacecraft, and has its origins in an advanced form of side looking airborne radar (SLAR). The distance the SAR device travels over a target in the time taken for the radar pulses to return to the antenna creates the large synthetic antenna aperture (the size of the antenna). Typically, the larger the aperture, the higher the image resolution will be, regardless of whether the aperture is physical (a large antenna) or synthetic (a moving antenna) – this allows SAR to create high-resolution images with comparatively small physical antennas. Additionally, SAR has the property of having larger apertures for more distant objects, allowing consistent spatial resolution over a range of viewing distances. [11]

2.3.1 Working Of SAR

Electromagnetic waves are transmitted sequentially, the echoes are collected and the system electronics digitizes and stores the data for subsequent processing. As transmission and reception occur at different times, they map to different positions. The well ordered combination of the received signals builds a virtual aperture that is much longer than the physical antenna width. That is the source of the term "synthetic aperture," giving it the property of an imaging

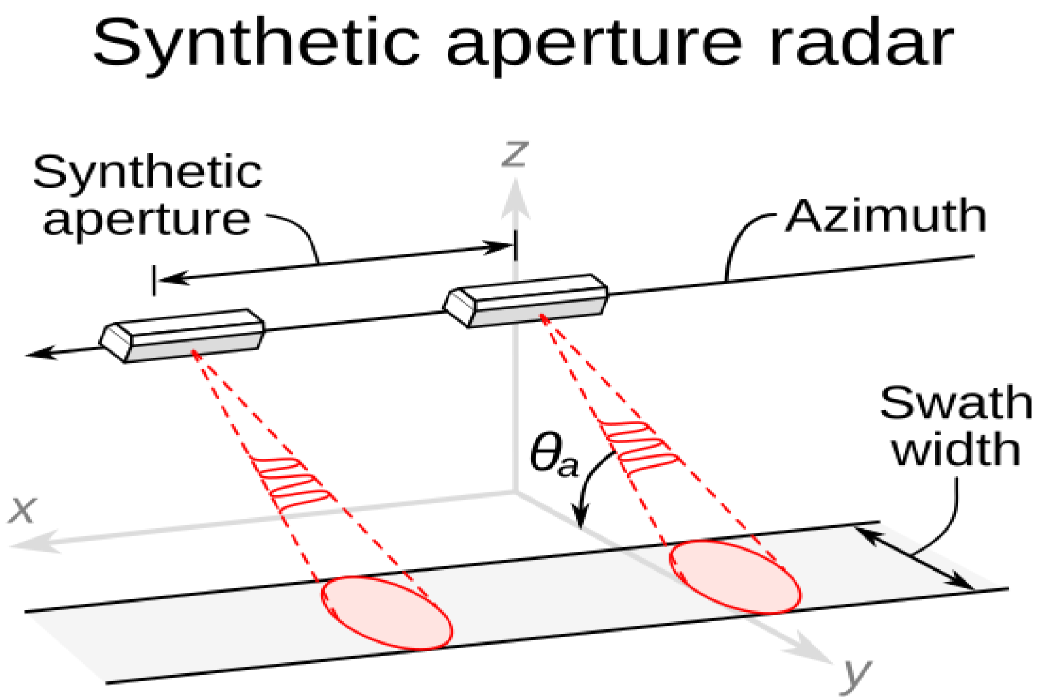


Figure 2.3: Basic Principle Of SAR

radar. The range direction is parallel to the flight track and perpendicular to the azimuth direction, which is also known as the along-track direction because it is in line with the position of the object within the antenna's field of view. [10]

The 3D processing is done in two stages. The azimuth and range direction are focused for the generation of 2D (azimuth-range) high-resolution images, after which a digital elevation model (DEM) is used to measure the phase differences between complex images, which is determined from different look angles to recover the height information. This height information, along with the azimuth-range coordinates provided by 2-D SAR focusing, gives the third dimension, which is the elevation. The first step requires only standard processing algorithms, for the second step, additional pre-processing such as image co-registration and phase calibration is used. In addition, multiple baselines can be used to extend 3D imaging to the time dimension. 4D and multi-D SAR imaging allows imaging of complex scenarios, such as urban areas, and has improved performance with respect to classical interferometric techniques such as persistent scatterer interferometry (PSI).

2.3.2 Algorithm For Creating SAR Images

A three-dimensional array (a volume) of scene elements is defined, which will represent the volume of space within which targets exist. Each element of the array is a cubical voxel representing the probability (a "density") of a reflective surface being at that location in space. (Note that two-dimensional SARs are also possible, showing only a top-down view of the target area.)

Initially, the SAR algorithm gives each voxel a density of zero. Then for each captured waveform, the entire volume is iterated. For a given waveform and voxel, the distance from the position represented by that voxel to the antenna(s) used to capture that waveform is calculated. That distance represents a time delay into the waveform. The sample value at that position in the waveform is then added to the voxel's density value. This represents a possible echo from a target at that position. Note there are several optional approaches here, depending on the precision of the waveform timing, among other things. For example, if phase cannot be accurately determined, only the envelope magnitude (with the help of a Hilbert transform) of the waveform sample might be added to the voxel. If waveform polarization and phase are known and are accurate enough, then these values might be added to a more complex voxel that holds such measurements separately.

After all waveforms have been iterated over all voxels, the basic SAR processing is complete. What remains, in the simplest approach, is to decide what voxel density value represents a solid object. Voxels whose density is below that threshold are ignored. Note the threshold level chosen must be higher than the peak energy of any single wave, otherwise that wave peak would appear as a sphere (or ellipse, in the case of multistatic operation) of false "density" across the entire volume. Thus to detect a point on a target, there must be at least two different antenna echoes from that point. Consequently, there is a need for large numbers of antenna positions to properly characterize a target. The voxels that passed the threshold criteria are visualized in 2D or 3D. Optionally, added visual quality can sometimes be had by use of a surface detection algorithm like marching cubes. [10]

2.4 Ship Detection: An Improved YOLOv3 Method

YOLOv3 is the state of art detector, which performs an excellent balance in detection speed and accuracy. In this paper, an improved YOLOv3 model named YOLOv3-ship is proposed for the ship detection. The main contributions to the YOLOv3-ship consists of dimension Clusters, network Improve-ment and embedding of the Squeeze-and-Excitation (SE) module.The experiments results show that the detection accuracy has been significantly improved by the YOLOv3-ship. [2]

The main contributions of this method can be listed as follows:

- Determine the anchor settings for the ship dataset by kmeans++ algorithm.
- Design a convolutional neural network named Darknet-ship to solve the problem of excessive YOLOv3 parameters.
- Embed the Squeeze-and-Excitation module in YOLOv3 to increase the network's ability to extract global features.

2.5 High-Speed Ship Detection In SAR Images By Improved YOLOv3

Ship detection in synthetic aperture radar (SAR) images plays an important role in marine transportation, fishery management, and maritime disaster rescue. Nowadays, the current researches almost are focusing on improving detection accuracy while detection speed is neglected. However, it is also extraordinarily important to increase the ship detection speed, because it can provide real-time ocean observation and timely ship rescue. Therefore, in order to solve this problem, this paper proposes a high-speed SAR ship detection approach by improved you only look once version 3 (YOLOv3). We experimented on a public SAR ship detection dataset (SSDD) which has been used by many other scholars. Finally, the experimental results indicated that the detection speed of our proposed improved YOLOv3 is faster than current other methods, such as faster-regions convolutional neural network (Faster R-CNN), single shot multi-box detector (SSD), and original YOLOv3 under a same hardware environment. Meanwhile, the detection accuracy remains basically unchanged. [12]

2.6 Research on Boat Identification Based on Improved Loss Function of Deep Convolutional Neural Networks

Abstract-Boat identification is a very important and challenging task in maritime traffic monitoring, the difficulty of this task lies in the accurate positioning and identification of relatively small boats in complex scenes, In view of the above situation, this paper proposes a new PL loss function to replace the original cross entropy function and constitute the single-stage detector YOLO-H model. this model effectively solves the problem that the model precision is low because of the sample imbalance. To prove the effectiveness of the loss function, we used the ms-coco data set to evaluate the YOLO-H network. The experimental results show that YOLO-H on MS-COCO realizes COCO test-dev AP 34.3, and the obtained AP is 1.3 higher than YOLOv3. In the actual task boat identification, the problem of low resolution and difficult detection of small targets is tested. The experimental results prove that the YOLO-H model can effectively identify small targets and solve the difficulty of the YOLOv3 algorithm in the detection of small targets.[5]

CHAPTER 3

YOLOv3-SHIP

The main contributions of this method is to determine the anchor settings for the ship dataset by kmeans++ algorithm. Here we Design a convolutional neural network named Darknet-ship to solve the problem of excessive YOLOv3 parameters.Embedind the Squeeze-and-Excitation module in YOLOv3 to increases the network's ability to extract global features. [2]

3.0.1 Dimension Clusters

YOLOv3 introduces anchors, a set of initial candidate boxes with fixed width and height. The settings of anchors boxes affect the detection accuracy and speed. Kmeans algorithm is selected to conduct dimension clusters in YOLOv3. How-ever, the kmeans algorithm is sensitive to the initial points.Therefore, an improved clusters algorithm named kmeans++ is introduced to solve this problem. The distance function of K-means++ algorithm is defined as:

$$d(a,b) = 1 - IOU(a,b)$$

where a is the size of rectangular box, b is centroid of the rectangular box. The IOU function represents the overlapping ratio of two rectangular boxes, as shown in Fig. 3.1.

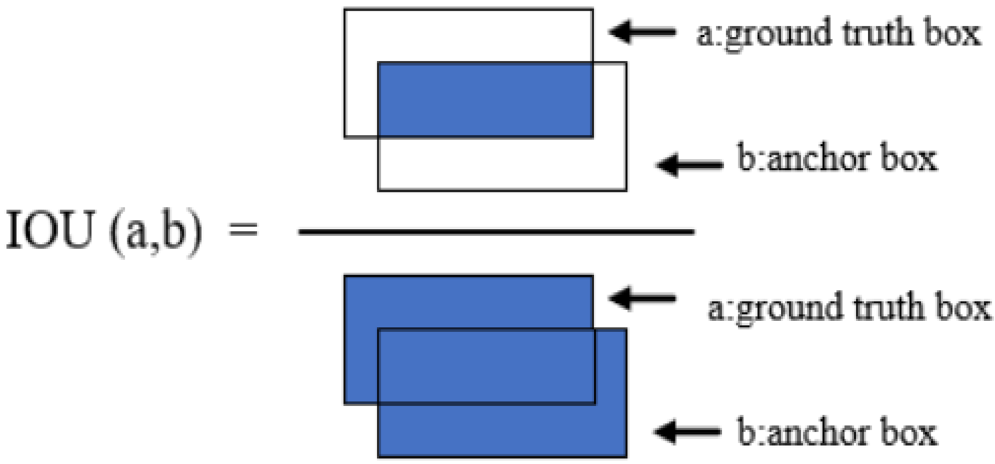


Figure 3.1: The description of IOU function

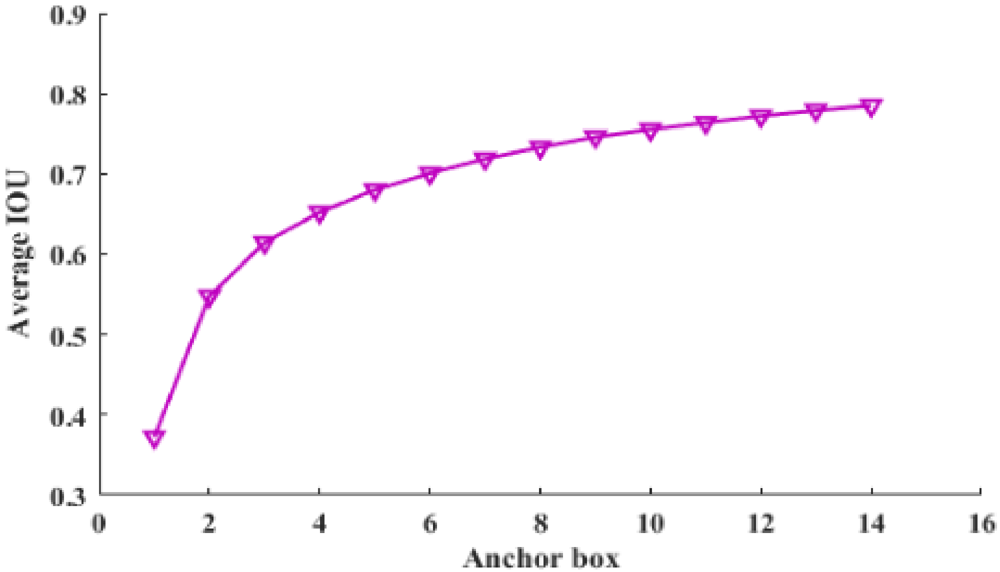


Figure 3.2: The relationship between the number of anchor boxes and average IOU

To determine the settings of anchor box, the relationship between average IOU and anchor box is depicted in Fig.3.2. According to the inflection point method, we selects six clusters and divide up the six clusters on three scales. the corresponding sizes of six clusters are: (31, 15), (65, 26), (115,42), (156, 28), (221, 55), (304, 104)

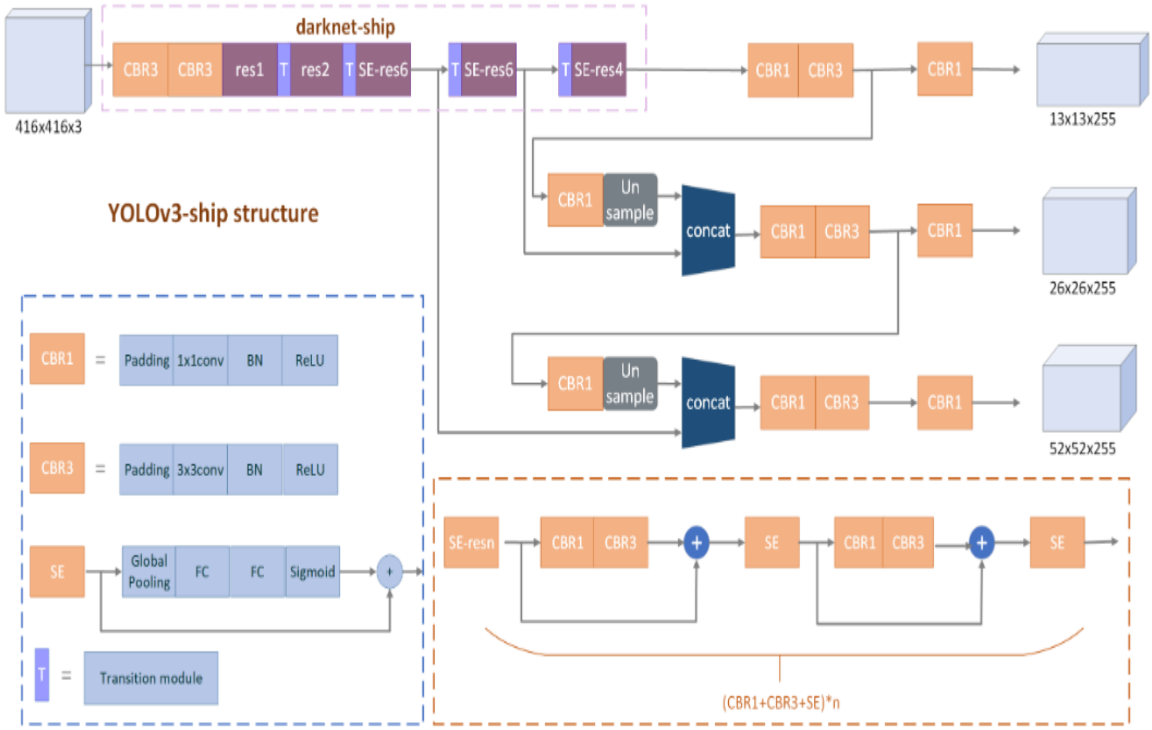


Figure 3.3: The description of YOLOv3-ship structure

3.1 Network Structure Improvement

Yolov3 establishes a Darknet53 ConvNet as a feature ex-tractor. However, it seems too complex and redundant forship detection, which may lead to more complex trainingand slower detection speed. Based on Darknet-53, a ConvNetnamed Darknet-ship is designed to reduce the parameters andimprove the performance of network to extract features. InDarknet-ship, the 1*1 convolution kernels are used to reducethe parameters in the transition module, as defined in Fig.3.4. [2]

Type	Filters	Size	Output
Conv	32	3x3 conv stride=2	208x208
Residual block(1)	32	3x3 conv stride=1	208x208
x1	64	3x3 conv stride=1	
Transition module	32	1x1 conv stride=1	104x104
	64	3x3 conv stride=2	
Residual block(2)	64	1x1 conv stride=1	104x104
x2	128	3x3 conv stride=1	
Transition module	64	1x1 conv stride=1	52x52
	128	3x3 conv stride=2	
Residual block(3)	128	1x1 conv stride=1	52x52
x6	256	3x3 conv stride=1	
Transition module	128	1x1 conv stride=1	26x26
	256	3x3 conv stride=2	
Residual block(4)	256	1x1 conv stride=1	26x26
x6	512	3x3 conv stride=1	
Transition module	256	1x1 conv stride=1	13x13
	512	3x3 conv stride=2	
Residual block(5)	256	1x1 conv stride=1	13x13
x4	512	3x3 conv stride=1	

Figure 3.4: The network structure of Darknet-ship.

3.2 Embedding of the Squeeze-and-Excitation Module

Squeeze-and-Excitation(SE) module is a ConvNet structure proposed by J. Hu in 2017, which won the championship of the imagenet classification competition. The SE module improves the expressive ability of network by accurately modeling the interaction between channels of convolution features.In the last three resnet blocks in Darknet-ship, the SE module is introduced to increase the receptive field and enhance the ability of network to extract global information. The SE module and the YOLOv3-ship structure are shown in Fig.3.5

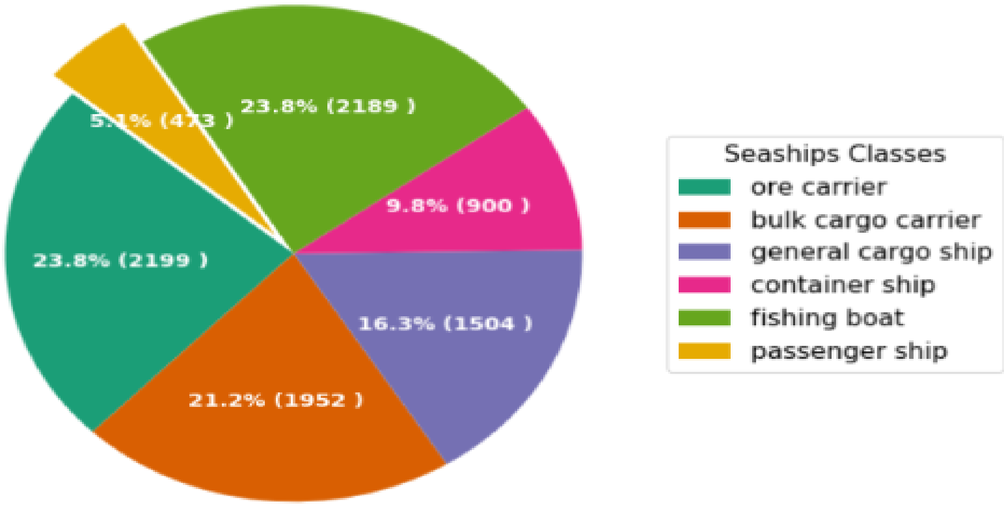
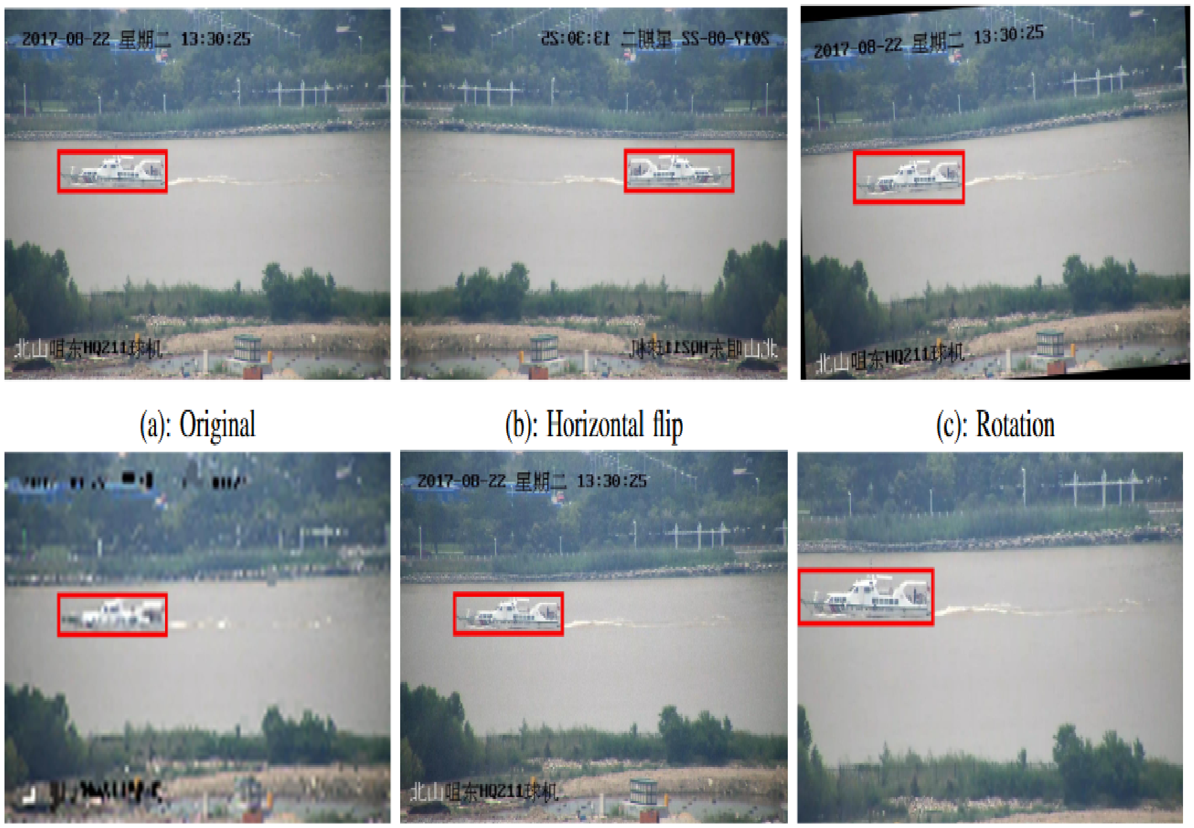


Figure 3.5: The numbers of bounding boxes for each class



(a): Original (b): Horizontal flip (c): Rotation

(d): Super pixel (e): Gaussian noise (f): Scale

Figure 3.6: Some samples of the augmentation methods

CHAPTER 4

HIGH-SPEED SHIP DETECTION BY IMPROVED YOLOV3

Ship detection in synthetic aperture radar (SAR) images plays an important role in marine transportation, fishery management, and maritime disaster rescue. Synthetic aperture radar (SAR), an all-weather and all-time microwave sensor, is one of the most important tools in remote sensing filed. Up to now, many scholars have proposed many SAR ship detection methods, which have greatly promoted the development of SAR image interpretation. In this method we modify YOLOv3 to increase speed of detection without compromising the accuracy in ship detection from SAR images.

4.1 Improved Method

In order to improve the speed of ship detection, we have improved the original YOLOv3. Different from the original YOLOv3 where 20 types of targets need to be detected, SAR target detection this method contains only one class that is ship, so the reduction of network size does not significantly reduce accuracy by our research findings.The network structure of improved YOLOv3 is shown in fig.4.2. The detailed improvements are shown in the table1. From fig.4.2 and fig.4.3, to reduce the size of the network, we will use Darknet-19 as the backbone of the improved YOLOv3, which can reduce detection time. We also deleted repeated layers in YOLOv3-Scale1, YOLOv3-Scale2, and YOLOv3-Scale3. Finally, in order to make full use of the features extracted from the network, we have added two feature concatenation paths, which can improve the detection accuracy.[12]

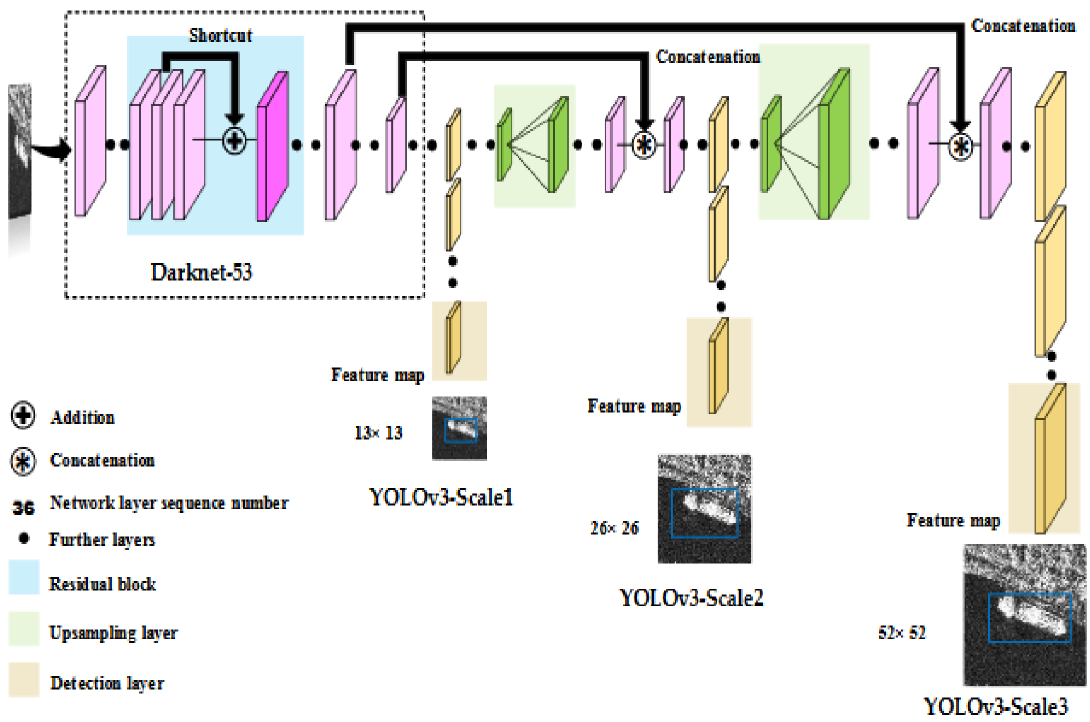


Figure 4.1: The network structure of original YOLOv3

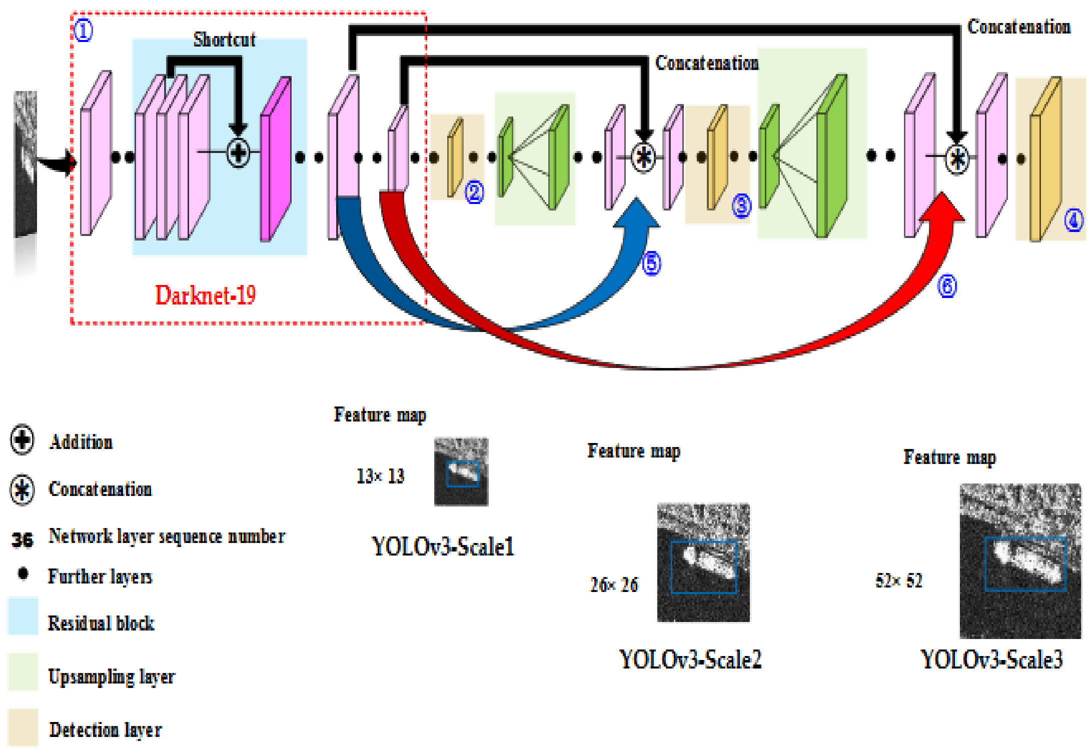


Figure 4.2: The network structure of improved YOLOv3

Number	Explanation
1	Change the backbone from Darknet-53 to Darknet-19.
2	Delete repeated layers of YOLOv3-Scale1.
3	Delete repeated layers of YOLOv3-Scale2.
4	Delete repeated layers of YOLOv3-Scale3.
5	Increase a concatenation.
6	Increase a concatenation.

Figure 4.3: The detailed improvements.

4.2 SAR Ship Detection Results

After some reasonable trainings, when the loss reaches the minimum, we get the final detection model. Then, we perform the actual SAR ship detection on the test set. The detection results of some samples are shown in the fig.4.4. From fig.4.4, almost all real ships can be detected correctly, and the original YOLOv3 and the improved YOLOv3 have similar detection accuracy.

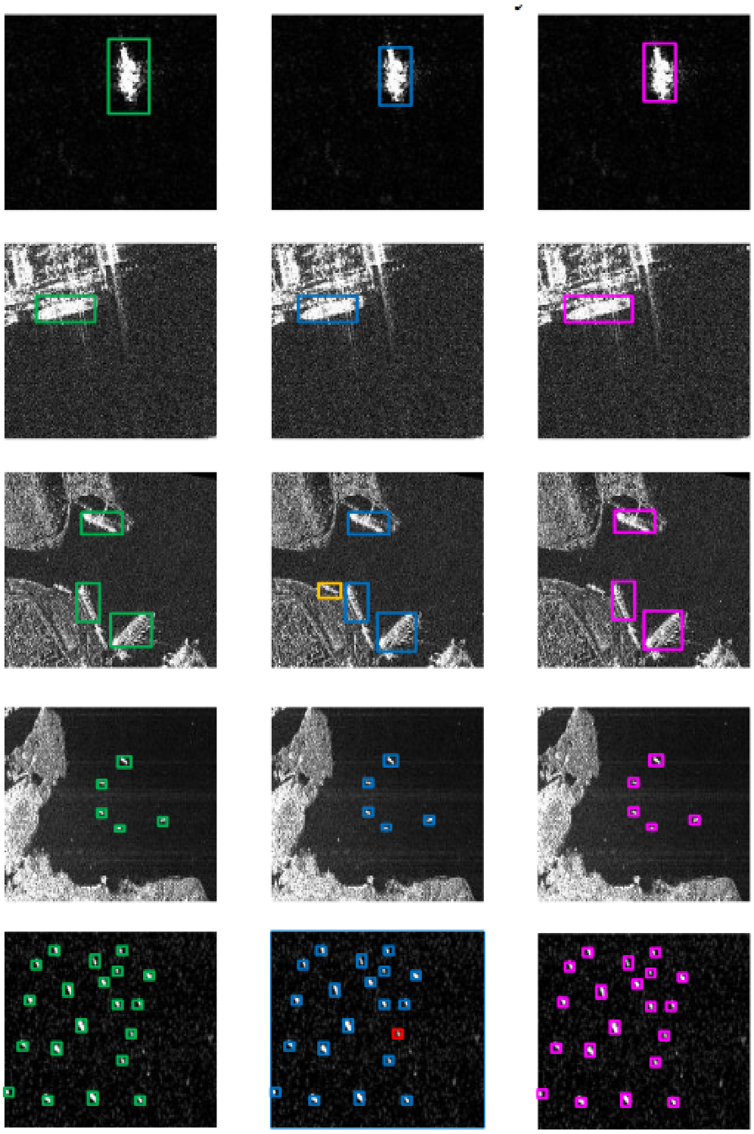


Figure 4.4: SAR ship detection results. (a) Ground truth. (b) Original YOLOv3. (c) Improved YOLOv3. Red is miss-detection and yellow is false alarm.

Fig.4.5 is the comparison of their quantitative evaluation indicators. From fig.3 and fig.4.5, the performance of the improved YOLOv3 is slightly better than the original YOLOv3, but the gap is narrow. More importantly, on the premise of keeping the accuracy basically unchanged, the improved network is smaller and lighter, which can reduce the detection time. In addition, due to the added two concatenation, the detection accuracy can remain unchanged. [12]

Method	Recall	Precision	mAP	F1-score
Original YOLOv3	89.12%	88.57%	90.14%	0.89
Improved YOLOv3	91.10%	89.36%	90.08%	0.90

Figure 4.5: Improvement Comparison.

CHAPTER 5

YOLO-H Based on Improved Loss Function

Most methods for ship detection ignore the case of small boats. But small boat detection is also very important and harder to do that regular ship detection as the size of the object is very small. This method aims to modify the YOLOv3 algorithm so that it can be used for small boat detection also. The method tries to detect boats even against complex backgrounds and in adverse weather conditions.

5.1 Basic Algorithm

The main idea of single-stage detector is to correlate the anchor box of each target feature graph with the result of convolution operation. The single-stage detector is composed of a backbone network and two task-specific sub-networks The backbone is responsible for calculating the convolution feature graph on the entire input image, which is a specific convolution network. The first subnet classifies the output of the trunk network by convolution. The second subnet performs convolution reference bounding box regression. In the regression subnet work, the relative offset between anchor box and ground truth box was predicted through the four position information output. The offset was minimized through the appropriate loss function. The boundary box was optimized through NMS and the overlapped parts were deleted. In this paper, we use darknet-53 network model and YOLOv3 algorithm for target recognition research, because it has a good performance in terms of efficiency and accuracy. [5]

5.2 Network Structure

To improve the cross entropy loss function of YOLOV3, FL function is used to replace the traditional cross entropy function. In order to facilitate the comparison of the improved network in subsequent experiments, the improved network is named YOLO-H. The YOLO-H network is a single stage regression detector. The backbone network is made of Darknet-53 feature extractor and YOLO interaction layer, Darknet-53 is a full convolution network for basic feature processing, and YOLO layer realizes local feature interaction between feature graphs by means of convolution kernel (3*3 and 11). The structure of YOLO-H network is shown in fig.5.1.

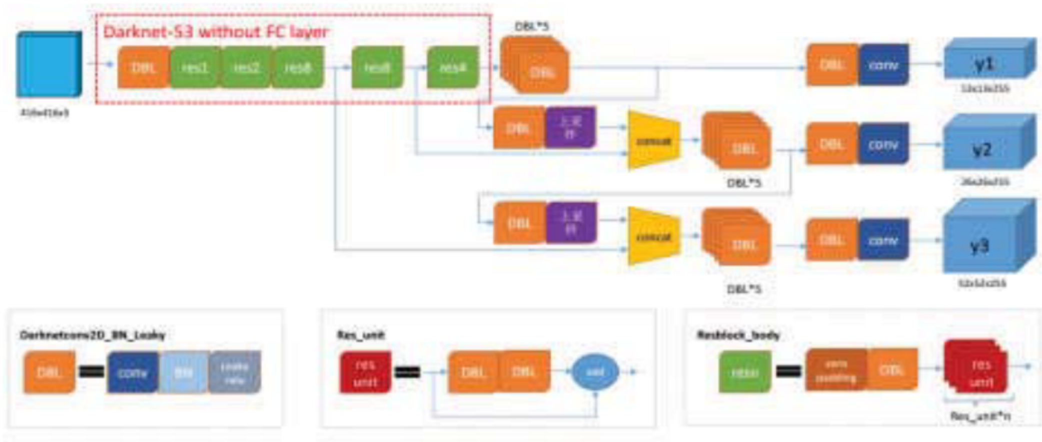


Figure 5.1: YOLO-H Network.

The darknet-53 network is a blend of darknet-19 and a new neural network technology called residual network, the network USES the convolution layer of 53 3x3 and 1x1 convolution kernels, and takes some jump connection modes between the layers to form the res layer, DBL includes two dimensional convolution operation (conv2D), BN layer and local response normalization (Leaky relu). Input data will be processed by two-dimensional convolution function, and the depth of each node matrix will be increased correspondingly. The shallow features of the image will be further analyzed to obtain the deeper features with higher abstraction, as shown in fig. 5.2.

$$S_i^j = g\left(\sum_k S_i^{j-1} * P_{ik}^j + W_i^j\right)$$

Here S term represents the i characteristic output of the j convolution layer, g is the activation function of the delta function, P is the connection weight of the i output feature and the j input feature, W term represents the offset of the i characteristic output of the j convolution layer.[5]

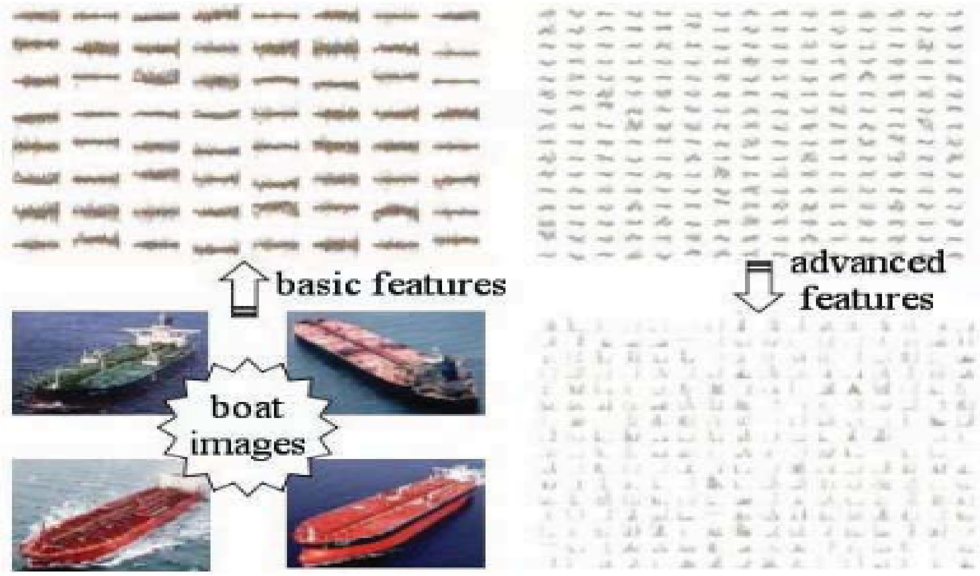


Figure 5.2: Basic features of boat data are extracted to advanced features.

5.3 Feature interaction layer

The Yolo layer is composed of three scales, in which local feature interactions are carried out through convolution kernel. The size of these three scales is 13x13, 26x26 and 52x52, respectively, forming the small, medium and large scale yolo layer. The characteristics of these sizes are extracted by the similar FPN (feature pyramid network) network to form the pyramid network. In this way, the up-sampling feature and fine-grained feature in the early feature mapping can be found and more meaningful semantic information can be obtained. The same network design is used to predict the final size of the bounding box, a process that actually helps with classification prediction by filtering out more detailed features from earlier images.YOLOv3 splices and fuses the upper sampling layer with the previous layer, which preserves the more detailed features in the image, which is very helpful for detecting objects of small size.

5.4 Result

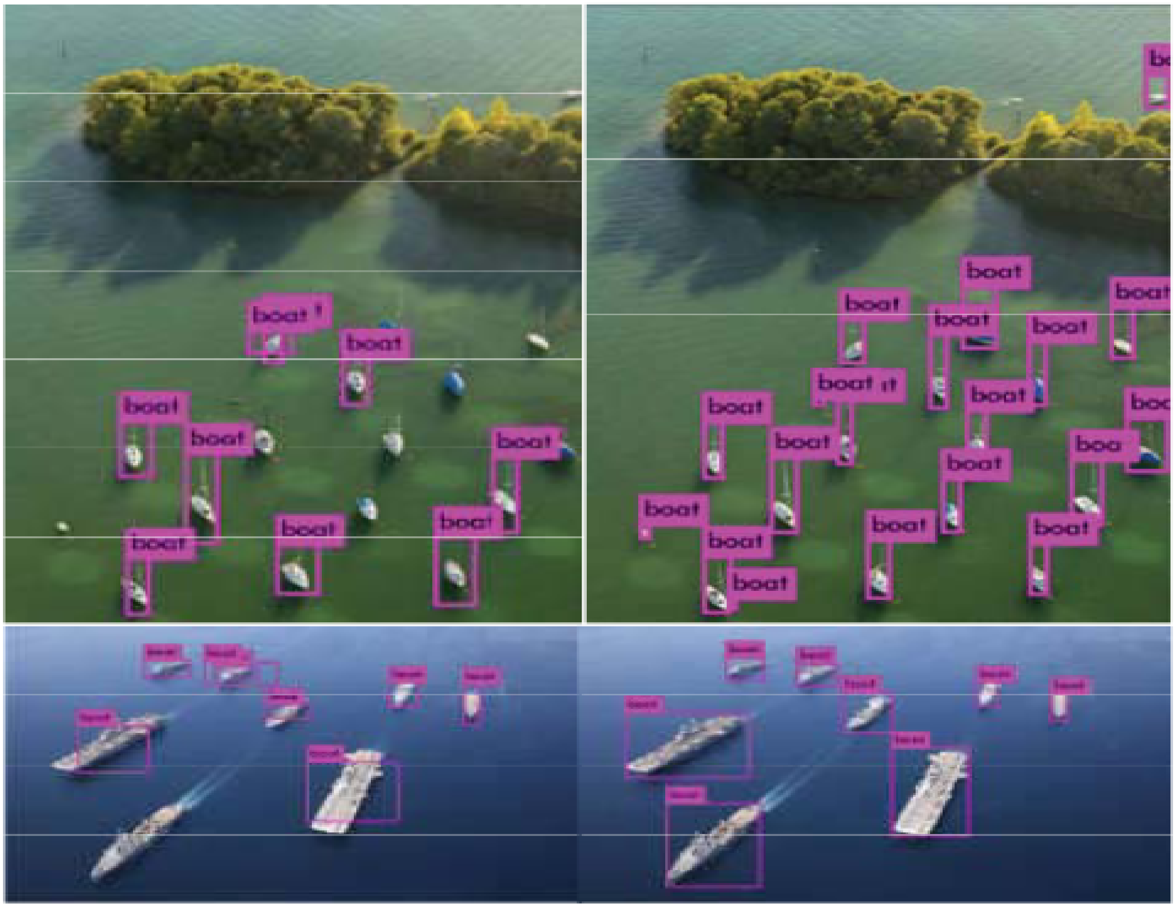


Figure 5.3: Identification and comparison of minimal targets.

The test image contains a large number of extremely small targets. The left image shows repeated recognition and omission recognition, while the right image effectively identifies cach extremely small target in the image.

Experimental results are shown in fig. 5.3, target size is too small, shade, before - background is difficult to distinguish between environment, YOLOv3 relatively large boat detection which can identify the target size, but for the target image resolution is low, small and will not be able to effectively identify, improved YOLO - H model all the target in the image recognition, thus proving FL function for the target detection has a great effect. [5]

CHAPTER 6

CONCLUSION

YOLOv3 is a Good Object Detection Algorithm but the improvements we made make it optimized for ship detection. In YOLO-ship we Used Darknet53 and SE module to improve the performance and to increase the accuracy for large and medium size objects. YOLO-H is more geared towards small objects where the Fl loss function and advanced single layer model makes it perfect for detecting small boats. Improved YOLO method removes basically increases the speed of detection while maintaining the same accuracy by removing parts of the original YOLOv3 that are not needed in ship detection.

REFERENCES

[1] Bindhya Bhadran and Jyothisha J Nair. Classification of patterns on high resolution sar images. In *2015 International Conference on Computing and Network Communications (CoCoNet)*, pages 784–792. IEEE, 2015.

[2] H. Cui, Y. Yang, M. Liu, T. Shi, and Q. Qi. Ship detection: An improved yolov3 method. In *OCEANS 2019 - Marseille*, pages 1–4, 2019.

[3] Reagan L Galvez, Argel A Bandala, Elmer P Dadios, Ryan Rhay P Vicerra, and Jose Martin Z Maningo. Object detection using convolutional neural networks. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 2023–2027. IEEE, 2018.

[4] Ievgen M Gorovyi and Dmytro S Sharapov. Pattern detection and recognition in sar images. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pages 123–126. IEEE, 2017.

[5] Y. Guo, S. Wang, H. He, L. Sun, and S. Ma. Research on boat identification based on improved loss function of deep convolutional neural networks. In *2019 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*, pages 278–283, 2019.

[6] Constantine P Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pages 555–562. IEEE, 1998.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[8] Francesco Visin, Kyle Kastner, Kyunghyun Cho, Matteo Matteucci, Aaron Courville, and Yoshua Bengio. Renet: A recurrent neural network based alternative to convolutional networks. *arXiv preprint arXiv:1505.00393*, 2015.

[9] Tengfei Wang and Yanfeng Gu. Cnn based renormalization method for ship detection in vhr remote sensing images. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1252–1255. IEEE, 2018.

[10] Wikipedia contributors. Synthetic-aperture radar — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Synthetic-aperture_radar&oldid=980796719`, 2020. [Online; accessed 28-November-2020].

[11] Xiaojian Xu and Ram M Narayanan. Enhanced resolution in sar/isar imaging using iterative sidelobe apodization. *IEEE Transactions on image processing*, 14(4):537–547, 2005.

[12] T. ZHANG, X. ZHANG, J. SHI, and S. WEI. High-speed ship detection in sar images by improved yolov3. In *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, pages 149–152, 2019.