การจำแนกและการติดตามเรือผิวน้ำด้วยระบบอัตโนมัติ

Automatic Classification and Tracking of Surface Ships

Received : April 11, 2023
Revised : November 9, 2023
Accepted : November 13, 2023

พิศณุ คูมีชัย
กองวิชาวิศวกรรมศาสตร์ ฝ่ายศึกษา โรงเรียนนายเรือ

Pisanu Kumeechai
Department of Engineering, Education Branch, Royal Thai Naval Academy
Pisanu41984198@hotmail.com

บทคัดย่อ

เทคโนโลยีการขับขี่อัตโนมัติของเรือผิวน้ำ กล้องเป็นสิ่งจำเป็นสำหรับการกำหนดเส้นทางและการตรวจจับวัตถุ เพื่อหลีกเลี่ยงสิ่งกีดขวางหรือหลีกเลี่ยงการชนกันของเรือ สิ่งที่สำคัญคือการติดตามการเคลื่อนไหวของเรือที่รู้จัก ในบทความนี้ได้ใช้ชุดข้อมูลการเดินเรือของไทย และการรวบรวมภาพข้อมูลสำหรับแบบจำลองฝึกอบรม จากนั้นจะนำเสนอการจดจำวัตถุด้วยวิธี AlexNet ด้วยการจำลองและประเมินผลการปฏิบัติงานในสภาพแวดล้อมการเดินเรือที่แตกต่างกัน จากนั้นจะเสนออัลกอริทึมการติดตามเพื่อติดตามวัตถุที่ระบุโดยเฉพาะอย่างยิ่งในการประเมินผลด้วยวิดีโอที่มีการเคลื่อนไหวสูง ในงานวิจัยนี้ใช้ภาพรวม 300 ภาพ แบ่งเป็นตัวอย่างการฝึก 200 ภาพ และตัวอย่างทดสอบ 100 ภาพ มีการวัดประสิทธิภาพของอัลกอริทึมด้วยค่า (Confusion Matrix) ผลการทดลองมีประสิทธิภาพของทั้ง precision, recall และ f1-score โดยมีค่า Precision เท่ากับ 71.3%, Recall เท่ากับ 95.4% และ F1-score เท่ากับ 81.6%

คำสำคัญ: การขับขี่อัตโนมัติของเรือผิวน้ำ, การจดจำวัตถุ, อัลกอริทึมการติดตาม, วิธี AlexNet,

Abstract

Automatic Watercraft Driving Technology: Cameras are essential for path determination and object detection to avoid obstacles and prevent ship collisions. The key focus lies in tracking the recognizable ship movements. In this article, the authors used Thai ship movement data and collected image data to develop a training model. They then presented an object recognition method using AlexNet, simulated different maritime environments, and evaluated the model's performance. Subsequently, they proposed a tracking algorithm for precise object tracking, especially in high-motion video evaluations. The study utilized a total of 300 images, divided into 200 training samples and 100 testing samples. Performance was assessed using a confusion matrix, and the experimental results revealed high efficiency, with precision, recall, and F1-score values of 71.3%, 95.4%, and 81.6%, respectively. These results demonstrate that the tracking algorithm outperforms real-time online tracking in terms of object tracking accuracy.

Keywords: automatic watercraft driving, object recognition, tracking algorithm, AlexNet method

1. บทนำ

เทคโนโลยีการขับขี่อัตโนมัตินั้นได้รับความสนใจอย่างมากในช่วงไม่กี่ปีที่ผ่านมา โดยเทคโนโลยีเหล่านี้จะใช้การประมวลผลข้อมูลจากเซ็นเซอร์เพื่อช่วยให้สามารถขับเคลื่อนได้โดยไม่ต้องมีผู้ขับขี่ โดยระบบนี้สามารถทำงานได้โดยอิงตามสภาพแวดล้อมและข้อมูลที่ได้รับจากเซ็นเซอร์ต่าง ๆ ซึ่งการออกแบบระบบสื่อสารส่งผ่านข้อมูล และการพัฒนาซอฟต์แวร์ได้พัฒนาไปอย่างรวดเร็ว ยานพาหนะต่าง ๆ เช่น รถยนต์ เรือ และโดรน ในอนาคตควรที่จะสามารถจดจำและตอบสนองถึงสิ่งกีดขวางที่อยู่เบื้องหน้าได้อย่างรวดเร็วเพื่อขับเคลื่อนตัวเองโดยปราศจากการแทรกแซงของมนุษย์ การพัฒนาความสามารถในการขับขี่อัตโนมัตินั้นอาจอาศัยเทคโนโลยีต่าง ๆ ประกอบด้วย เพื่อให้การทำงานดีขึ้น เช่น กล้องไลดาร์ LIDAR (Light Detection and Ranging) เป็นเลเซอร์การถ่ายภาพ การตรวจจับ และการวัดระยะทาง GPS (Global Positioning System) ระบบระบุตำแหน่งบนพื้นโลก และ SONAR (Sound Navigation and Ranging) ระบบการหาตำแหน่งวัตถุใต้น้ำโดยการส่งคลื่นเสียงและรับเสียงสะท้อน นอกจากนี้ในส่วนของการเดินเรือเรือผิวน้ำอัตโนมัติสามารถใช้ระบบระบุตัวตนอัตโนมัติ AIS (Automatic Identification System) ซึ่งเป็นระบบฐานข้อมูลตำแหน่งอัตโนมัติทั่วโลก โดยจะอิงจากการติดตั้งตัวรับสัญญาณขนาดเล็กเข้ากับเรือ ที่ส่งสัญญาณอย่างต่อเนื่อง แจ้งเตือนเรือลำอื่นและสถานีฝั่งที่มีเครื่องรับสัญญาณ AIS อยู่ และ GPS (Global Positioning System) ระบบการนำทางด้วยดาวเทียมซึ่งประกอบด้วยดาวเทียมอย่างน้อย 24 ดวง โดย GPS (Global Positioning System) สามารถปฏิบัติการได้ในทุกสภาพอากาศ ทุกที่ในโลก ตลอด 24 ชั่วโมงต่อวัน ทั้งนี้เพื่อเพิ่มความสามารถในการตรวจจับเรือข้างเคียงและป้องกันการชนกับเรือลำอื่น เรือขนาดเล็ก หรือทุ่นต่าง ๆ แต่ในส่วนของเรือเล็กและเรือคายัคส่วนมากจะไม่มีเครื่องส่งสัญญาณของ AIS (Automatic Identification System) เรือประมงผิดกฎหมายที่ออกหาปลาในเวลากลางคืนโดยไม่ได้รับอนุญาต อาจปิด AIS (Automatic Identification System) ส่งผลให้เกิดอุบัติเหตุทางเรือ เทคโนโลยีในอนาคตจำเป็นต้องจดจำวัตถุโดยใช้วิดีโอจากกล้องเพื่อป้องกันอุบัติเหตุทางทะเล เช่น การชนกันของเรือโดยไม่คาดคิด หลังจากการระบุเรือที่อยู่รอบ ๆ ประมวลผลข้อมูล เช่น แนวการเคลื่อนที่ ความเร็ว และทิศทางของเรือลำอื่นจำเป็นต้องวางแผนเพื่อหลีกเลี่ยงการชนกัน เพราะการรับรู้วัตถุในวิดีโอจะดำเนินการทีละเฟรม เป็นไปไม่ได้ที่จะตรวจพบการเปลี่ยนแปลงในตำแหน่งของวัตถุที่กำหนดเมื่อเวลาผ่านไป สำหรับการนำทางในเส้นทางจำเป็นต้องมีเทคโนโลยีการติดตามวัตถุด้วยเพื่อกำหนดเส้นทางการเคลื่อนที่ ความเร็ว ทิศทาง ฯลฯ มีการศึกษาหลายเรื่องเกี่ยวกับการตรวจหาและติดตามวัตถุด้วยวิดีโอ [1–9] การรู้จำวัตถุแบบ Deep-Learning เช่น R-CNN ออกแบบมาสำหรับการตรวจจับวัตถุได้อย่างยืดหยุ่น (บริเวณที่มี Convolutional Neural คุณสมบัติเครือข่าย) และ YOLO (You Only Look Once) ที่สามารถตรวจจับแม้กระทั่งวัตถุที่มันซ้อนกันได้ด้วย โดยมีโครงสร้างที่ค่อนข้างซับซ้อนของกริดในแต่ละชั้นที่เล็กลงเรื่อย ๆ ในแต่ละเลเยอร์ระบุตำแหน่งของวัตถุและทำนายประเภทของมันในสตรีมภาพจากกล้อง [1,10] การติดตามอัลกอริทึมประเมินว่าเป็นวัตถุเดียวกันหรือไม่เมื่อติดตั้งกล้องบนเรือแล้ว

เนื้อหาโดยสรุปมีเป็นการศึกษาการพัฒนาระบบการขับขี่อัตโนมัติสำหรับเรือผิวน้ำ โดยระบบนี้จะใช้เทคโนโลยีการรับรู้วัตถุและติดตามวัตถุจากกล้องวิดีโอเพื่อตรวจจับเรือที่อยู่รอบ ๆ และป้องกันการชนกัน และวัตถุประสงค์ย่อยของการศึกษา ได้แก่ ศึกษาเทคโนโลยีการรับรู้วัตถุและติดตามวัตถุจากกล้องวิดีโอ พัฒนาระบบการรับรู้วัตถุและติดตามวัตถุจากกล้องวิดีโอสำหรับเรือผิวน้ำ และการทดสอบและประเมินประสิทธิภาพของระบบการรับรู้วัตถุและติดตามวัตถุจากกล้องวิดีโอสำหรับเรือผิวน้ำ จากเนื้อหาที่ระบุไว้ การศึกษาเหล่านี้จะมุ่งเน้นไปที่การพัฒนาระบบการรับรู้วัตถุและติดตามวัตถุจากกล้องวิดีโอสำหรับเรือผิวน้ำ โดยระบบนี้จะใช้เทคโนโลยีการรู้จำวัตถุแบบ Deep-Learning เช่น R-CNN และ YOLO เพื่อตรวจจับเรือที่อยู่รอบ ๆ และติดตามการเคลื่อนไหวของเรือเหล่านั้น เพื่อป้องกันการชนกันของเรือ โดยระบบที่พัฒนาขึ้นนี้จะช่วยเพิ่มความปลอดภัยในการเดินเรือและลดอุบัติเหตุทางทะเล

สำหรับวัตถุประสงค์หลักในการวิจัยนี้คือการพัฒนาระบบการการจำแนกและติดตามวัตถุสำหรับเรือผิวน้ำ และประเมินประสิทธิภาพของระบบ โดยผลการทดลองแสดงให้เห็นว่าระบบการติดตามวัตถุมีประสิทธิภาพสูง สามารถนำไปประยุกต์ใช้เพื่อเพิ่มความปลอดภัยในการเดินเรือได้ และวัตถุประสงค์ย่อยเพื่อศึกษาเทคโนโลยีการรู้จำวัตถุแบบ Deep-Learning เพื่อนำมาประยุกต์ใช้ในการพัฒนาระบบการติดตามวัตถุ เพื่อพัฒนาชุดข้อมูลการเดินเรือของไทย และเพื่อใช้สำหรับการฝึกอบรมโมเดลการติดตามการเคลื่อนไหวของเรือที่อยู่ในสภาพแวดล้อมการเดินเรือที่แตกต่างกัน

งานวิจัยนี้นำเสนอการจำแนกและการติดตามเรือผิวน้ำด้วยระบบอัตโนมัติเพื่อศึกษาการตรวจเรือด้วยปัญญาประดิษฐ์เชิงลึก (Deep-Learning) ซึ่งที่เหลือของบทความนี้มีการจัดระเบียบดังต่อไปนี้ ส่วนที่ 2 ขอบเขตงานวิจัยที่นำเสนอ ส่วนที่ 3 กล่าวถึงทฤษฎีที่เกี่ยวข้อง ส่วนที่ 4 การดำเนินการวิจัย ส่วนที่ 5 ผลการวิจัย ส่วนที่ 6 สรุปผล และสุดท้ายส่วนที่ 7 เป็นข้อเสนอแนะ

2. ขอบเขตงานวิจัย

ชุดข้อมูลรูปภาพถูกสร้างขึ้นในลักษณะที่เป็นรูปภาพเดียวบันทึกทุกๆ ห้าเฟรม เนื่องจากเฟรมที่ต่อเนื่องกันมีความคล้ายคลึงกันในระดับสูงอาจเกิด Overfitting สำหรับวัตถุที่เปิดเผยในรูปแบบเดียวกัน และดังนั้นจึงไม่ได้เลือกทุกเฟรมของวิดีโอ ทำการดึงภาพจำนวน 300 ภาพจากวิดีโอ 5 รายการของเรือประเภทต่าง ๆ แสดงจำนวนวัตถุที่แยกจากข้อมูล เพื่อป้องกันการโอเวอร์ฟิต จึงรวบรวมภาพวัตถุจำนวน 300 ภาพ ประกอบไปด้วยเรือประเภทต่างกัน 5 ประเภท คือเรือน้ำมัน, เรือขนส่ง, เรือหาปลา, เรือรบ และเรือพายขนาดเล็ก โดยใช้กล้องเว็บแคม การทดลองการรับรู้และติดตามวัตถุดำเนินการบนเซิร์ฟเวอร์ที่มี CPU Intel Xeon E5-2620, RAM 16 GB และ GTX สองตัว GPU 1080Ti ระบบปฏิบัติการและเฟรมเวิร์กคือ window10 และ Darknet ตามลำดับ

ในเอกสารนี้ใช้ชุดข้อมูล Cifar-10 ปรับขนาดภาพ 32 × 32 นำมาแบ่ง เป็นตัวอย่างการฝึก 200 ตัวอย่าง และตัวอย่างทดสอบ 100 ตัวอย่าง

3. ทฤษฎีที่เกี่ยวข้อง

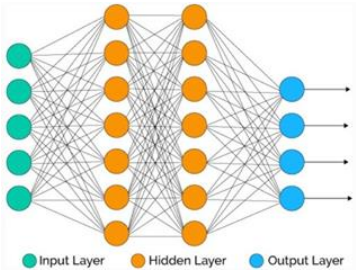
การจำแนกวัตถุจะแยกคุณลักษณะของวัตถุที่จะตรวจพบก่อนจากนั้นจะจดจำคุณสมบัติเหล่านั้นภายในภาพที่กำหนด เทคนิคการตรวจจับวัตถุต่าง ๆ เช่น ตัวตรวจจับ Canny Edge, Harris Corner, HOG (ฮิสโตแกรมของการไล่ระดับสี) และ SIFT (การแปลงคุณลักษณะที่ไม่แปรเปลี่ยนมาตราส่วน) ถูกนำเสนอในอดีต [11–14] การเรียนรู้ได้รับความนิยมมากขึ้นในช่วงไม่กี่ปีที่ผ่านมาอาทิ เช่น อัลกอริทึมการรับรู้ CNN (Convolutional Neural Network) การคำนวณนี้จะเริ่มจากการกำหนดค่าในตัวกรอง (Filter) หรือเคอร์เนล (Kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก โดยปกติตัวกรองเคอร์เนลอันหนึ่งจะดึงคุณลักษณะที่สนใจออกมาได้หนึ่งอย่างจึงจำเป็นต้องใช้ตัวกรองหลายตัวกรองด้วยเพื่อหาคุณลักษณะทางพื้นที่หลายอย่างประกอบกัน Mask R-CNN (Mask Regional Convolutional Neuron Network) ภูมิภาคที่มีคุณสมบัติโครงข่ายประสาทเทียมแบบ Convolutional และ VGG (Visual Geometry Group) ถูกแนะนำให้ใช้กลุ่ม Visual Geometry และ ResNet Residual Network สำหรับวัตถุการรับรู้ [10,15,16] การค้นหาวัตถุโดยใช้การจัดหมวดหมู่วัตถุโดยใช้การจำแนกภูมิภาค ข้อเสียของการตรวจจับแบบสองขั้นตอนคือ ตรวจจับวัตถุได้ช้าทำให้การตรวจจับตามเวลาจริงไม่สามารถทำได้ YOLO เครื่องตรวจจับแบบขั้นตอนเดียวที่จัดทำข้อเสนอระดับภูมิภาคและการจัดหมวดหมู่พร้อมกันได้รับการเสนอให้อำนวยความสะดวกในการตรวจจับตามเวลาจริง [1] YOLO เป็นการจำแนกวัตถุที่ใช้วิธีการที่เรียกว่า “Fast Single-Shot Detection” ซึ่งจะเป็นวิธีการที่สามารถตรวจจับวัตถุได้จากการส่งผ่านรูปภาพเข้าไปในระบบเพียงครั้งเดียว โดยใช้หลักการของโครงข่ายประสาทแบบคอนโวลูชัน อัลกอริทึม YOLO มีการพัฒนาอย่างต่อเนื่อง ตัวแรกที่สร้างขึ้นบน GoogLeNet มีโครงสร้างเครือข่ายแบบ 24 Convolutional เลเยอร์ และสองเลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์ด้วยอัตราการตรวจจับ 45 เฟรมต่อวินาทีใน [17] YOLOv2

หรือที่รู้จักในชื่อ YOLO9000 ได้เสนอการเพิ่มประสิทธิภาพสิบประการมีองค์ประกอบ เพื่อจัดการกับอัตราการตรวจจับที่ไม่ดีของ YOLO แรก และแสดง mAP เพิ่มขึ้น 15 เปอร์เซ็นต์ เมื่อเปรียบเทียบกับ YOLO เริ่มต้น YOLOv3 ขึ้นอยู่กับสถาปัตยกรรม YOLOv2 อย่างไรก็ตามเครือข่ายหลักคือ Darknet-53 และใช้การเชื่อมต่อแบบสั้น ResNet [18] ใน [19] YOLOv4 ใช้ CSPDarknet53 SPP (การรวมพีระมิดเชิงพื้นที่) และ PAN (การรวมเส้นทาง) และ YOLOv3 ยังใช้เพื่อปรับปรุงความแม่นยำและประสิทธิภาพ นอกจากนี้ยังแสดงให้เห็นว่าด้วยการใช้กลยุทธ์การคำนวณแบบคู่ขนาน การเรียนรู้ที่รวดเร็วและแม่นยำสามารถทำได้ทำได้แม้ในสภาพแวดล้อมการฝึกสอน GPU เดียวในการติดตามวัตถุ เทคนิคในการตรวจจับวัตถุที่เคลื่อนที่ข้ามเวลาความคล้ายคลึงกันระหว่างเฟรมต่อเนื่อง เพื่อดำเนินการติดตามวัตถุโดยเฉพาะ FairMOT (การติดตามหลายวัตถุอย่างยุติธรรม) SORT (การติดตามออนไลน์และเรียลไทม์อย่างง่าย) และ การติดตามแบบออนไลน์และเรียลไทม์อย่างง่ายเชิงลึกเป็นอัลกอริทึมการติดตามวัตถุ [2–4] ใน [2] FairMOT ดำเนินการตรวจจับวัตถุและฟังก์ชัน Re-ID อย่างสมดุล โดยมีอัตราเฟรมเท่ากับ 30 เฟรมต่อวินาที เพื่อลดต้นทุนการคำนวณ SORT ใช้ตัวกรองคาลมานและอัลกอริทึมฮังการี [3] เนื่องจากตัวกรองคาลมานคาดการณ์ตำแหน่งของวัตถุตามด้วยความเร็วก่อนหน้านี้ การติดตามในกรณีที่วัตถุสองชิ้นทับซ้อนกันหรือเปลี่ยนทิศทางอย่างรวดเร็ว ใน [4] ใช้อัลกอริทึมของฮังการีโดยสะท้อนความลึกคุณลักษณะการเรียนรู้ (Re-ID) ในตัวกรอง Kalman เพื่อติดตามวัตถุแม้ว่าจะทับซ้อนกันและเปลี่ยนทิศทางในการประเมินว่าวัตถุที่รู้จักสำหรับแต่ละเฟรมนั้นเหมือนกับเฟรมก่อนหน้าหรือไม่ การติดฉลาก ID จะดำเนินการตามความคล้ายคลึงกันของวัตถุจากนั้นจะติดตามแต่ละวัตถุจำนวนมากในเฟรมมีการศึกษาต่าง ๆ เกี่ยวกับการจดจำและติดตามเรือด้วยภาพในอุตสาหกรรมต่อเรือใน [5] แบบจำลองการจดจำวัตถุใช้ YOLOv2 และโทโพโลยีโครงข่ายประสาทเทียมได้รับการปรับแต่งเพื่อปรับปรุงความแม่นยำ ใช้ข้อมูลการฝึกสอนประกอบด้วย 159 ภาพที่รวบรวมจากชุดข้อมูลจากไทย อย่างไรก็ตามเนื่องจากฉลากไม่สม่ำเสมอกันแบบกระจายประสิทธิภาพการตรวจจับสำหรับแต่ละคลาสจะแตกต่างกันอย่างมาก นอกจากนี้ในขณะที่โมเดลที่ได้รับสามารถประมวลผลแบบเรียลไทม์ที่ 30 เฟรมต่อวินาที การเรียกคืนความแม่นยำ และ mAP แสดงผลการทำงานค่อนข้างต่ำที่ 76% 66% และ 69.79% ตามลำดับ ใน [7] Ship-YOLOv3 ถูกเสนอเพื่อเพิ่มความแม่นยำของเป้าหมายเรือการตรวจจับ เมื่อเทียบกับ YOLO3 ความแม่นยำและการเรียกคืนเพิ่มขึ้น 12.5% และ 11.5% ตามลำดับ อย่างไรก็ตาม เนื่องจากมีเพียงสามป้ายกำกับและใช้ข้อมูลจากกล้องติดตั้งบนสะพาน จึงไม่เหมาะสำหรับวิดีโอที่มีการเคลื่อนไหวสูง ใน [6] การตรวจจับวัตถุวิธีการติดตามแบบจำลองและวัตถุถูกนำมาใช้โดยใช้ YOLOv3 เมื่อเปรียบเทียบกับรุ่นก่อนหน้า YOLOv3 ที่ได้รับการปรับปรุงจะเพิ่ม mAP และ FPS 5% และ 2% ตามลำดับ ซึ่งแสดงให้เห็นว่ามีประสิทธิภาพเหนือกว่า อย่างไรก็ตามไม่เพียงพอที่จะนำไปใช้กับพื้นผิวที่เป็นอิสระในการเดินเรือ เนื่องจากได้รับการทดสอบด้วยวิดีโอสี่รายการจากกล้องคงที่เท่านั้น องค์ประกอบของภาพไม่สามารถคงที่ได้เนื่องจากการเคลื่อนที่ของเรือ เพราะวัตถุมีการเคลื่อนที่ต่อเนื่องในกรอบและออกนอกเฟรม เป็นการยากที่จะรับรู้ว่าเป็นวัตถุเดียวกันในสถานการณ์นั้น นอกจากนี้เรือหลายขนาดยังทับซ้อนกันในวิดีโอด้วยการเคลื่อนที่ไปในทิศทางต่าง ๆ เพื่อให้วัตถุที่ทับซ้อนกันหายไปหน้าจอ เมื่อวัตถุที่ซ่อนอยู่ปรากฏขึ้นอีกครั้งวัตถุนั้นจะถูกจดจำว่าเป็นวัตถุอื่น ดังนั้นจึงเสนออัลกอริทึมการจดจำและติดตามวัตถุสำหรับวิดีโอเคลื่อนไหว ข้อมูลการฝึกสอนนับพันจะถูกใช้และการถ่ายโอนใช้เทคนิคการเรียนรู้เพื่อหลีกเลี่ยงการขาดแคลนข้อมูลการฝึกสอนและความไม่สมดุลของป้ายกำกับ แบบจำลองการรับรู้ตาม YOLO ถูกเสนอสำหรับการจดจำวัตถุใช้ YOLO เพื่อการจดจำ เนื่องจากสามารถตรวจจับวัตถุได้อย่างรวดเร็วและติดตามวัตถุได้จำเป็นต้องมีอัตราการรับรู้ที่รวดเร็ว จากนั้นจึงนำเสนออัลกอริทึมการติดตามที่ใช้ IoU_Matching และ ORB_and_Size_Matching

3.1 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network) เป็นโมเดลคำนวณทางคณิตศาสตร์ที่จำลองการทำงานของระบบประสาทเทียมในสมองของมนุษย์ โดยมีวัตถุประสงค์เพื่อใช้ในการแก้ไขปัญหาทางคณิตศาสตร์และการเรียนรู้เชิงลึก (Deep Learning) โดยรับข้อมูลเข้ามาผ่านชั้นของโหนด (Node) หรือเรียกว่าเซลล์ประสาทเทียม (Neuron) แล้วนำไปประมวลผลผ่านชั้นต่าง ๆ เพื่อสร้างโมเดลการเรียนรู้และการจำแนกข้อมูล โดยโครงสร้างของโครงข่ายประสาทเทียมประกอบไปด้วยชั้นของโหนดหลาย ๆ ชั้นที่เชื่อมต่อกันเป็นเครือข่าย และมีการกำหนดน้ำหนักในการเชื่อมต่อของโหนด

เพื่อให้เกิดผลลัพธ์ที่ต้องการ โครงข่ายประสาทเทียมมีความสามารถในการเรียนรู้แบบเชิงลึก ทำให้สามารถจำแนกและรู้จำลักษณะของข้อมูลได้อย่างมีประสิทธิภาพเพิ่มขึ้น ส่วนใหญ่ถูกนำมาใช้งานในการจำแนกภาพ ประมวลผลเสียง และการแยกประเภทของข้อมูลอื่น ๆ ในช่วงไม่กี่ปีที่ผ่านมา โครงข่ายประสาทเทียมได้เป็นเครื่องมือที่สำคัญสำหรับการพัฒนาโมเดลการเรียนรู้และการจำแนกข้อมูลของโปรแกรมคอมพิวเตอร์ [20] โดยโครงสร้างของโครงข่ายประสาทเทียม แสดงดังรูปที่ 1



รูปที่ 1 โครงสร้างของโครงข่ายประสาทเทียม (Artificial Neural Networks) [21]

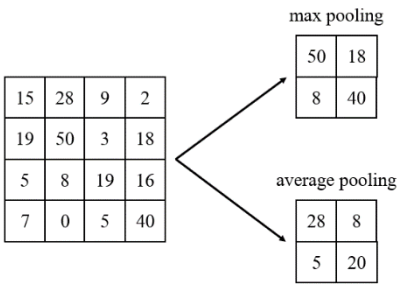
3.2 การดำเนินการของ Pooling

Pooling เป็นหนึ่งในชั้นของโมเดลโครงข่ายประสาทเทียม ที่ใช้สำหรับลดขนาดของข้อมูลภายในชั้นก่อนที่จะส่งต่อไปยังชั้นต่อไป การลดขนาดของข้อมูลจะช่วยในการลดการคำนวณและการประมวลผลที่จำเป็นในโมเดลโครงข่ายประสาทเทียม Pooling มีหลายวิธีการดำเนินการ โดยที่วิธีการที่นิยมใช้มากที่สุดคือ Max Pooling และ Average Pooling

Max Pooling: การเลือกค่าสูงสุดในช่วงของข้อมูลสำหรับส่วนที่ต้องการลดขนาด โดยเลือกค่าสูงสุดจากส่วนของข้อมูลที่กำหนดขนาด (ตัวอย่างเช่นการกำหนดส่วนของข้อมูลเป็นขนาด 2x2 จะเลือกค่าสูงสุดในช่วงของข้อมูล 4 ตัวเลข)

Average Pooling: การเลือกค่าเฉลี่ยของช่วงของข้อมูลสำหรับส่วนที่ต้องการลดขนาด โดยเลือกค่าเฉลี่ยจากส่วนของข้อมูลที่กำหนดขนาด (ตัวอย่างเช่นการกำหนดส่วนของข้อมูลเป็นขนาด 2x2 จะเลือกค่าเฉลี่ยของช่วงของข้อมูล 4 ตัวเลข)

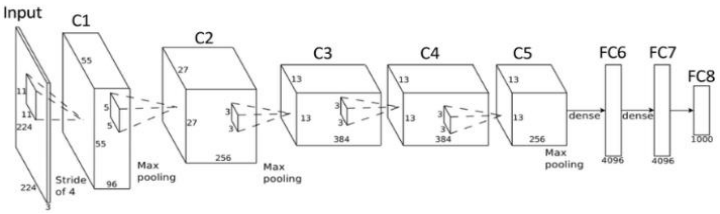
การใช้ Pooling จะช่วยลดขนาดของข้อมูล เพื่อลดการคำนวณและประหยัดความจำในการเก็บข้อมูล นอกจากนี้การใช้ Pooling ยังสามารถช่วยลดการเกิด Overfitting ในโมเดลโครงข่ายประสาทเทียมด้วยการลดความซับซ้อนของโมเดลโดยการลดขนาดของข้อมูลภายใน ซึ่งนิยมใช้การเลือกข้อมูลที่มีค่ามากที่สุด (Max Pooling) หรือ ค่าเฉลี่ย (Average Pooling) มาจากแต่ละช่วงของเมตริกซ์ เพื่อสร้างเป็นเมทริกซ์ที่มีขนาดเล็ก [22] ดังแสดงในรูปที่ 2



รูปที่ 2 ตัวอย่างชั้นการรวมโดยค่าที่มากที่สุดและค่าเฉลี่ย

3.3 อเล็กซ์เน็ต (AlexNet)

AlexNet เป็นโมเดลโครงข่ายประสาท ประเภท CNN (Convolutional Neural Network) ที่ถูกพัฒนาโดยทีมนักวิจัยของอุตสาหกรรมการค้าออนไลน์ระดับโลกของอเมริกา เพื่อใช้ในการแข่งขันการตรวจจับวัตถุในภาพ (ImageNet Large Scale Visual Recognition Challenge) ในปี 2012 โดยมีชื่อผู้วิจัยหลักคือ Alex Krizhevsky, Ilya Sutskever และ Geoffrey Hinton โมเดล AlexNet มีลักษณะการสร้างโครงข่ายที่มีความลึกมากกว่าโมเดล CNN ที่พัฒนาก่อนหน้านี้ โดยมีชั้น CNN ทั้งหมด 5 ชั้น ซึ่งจัดวางแบบลึก (Deep Architecture) และมีจำนวนนัดเชิงเส้น (Filter) ในแต่ละชั้นมากถึง 96 นัดเชิงเส้นต่อชั้น โดยใช้ฟังก์ชันเชื่อมโยง (Activation Function) แบบ ReLU (Rectified Linear Unit) สำหรับการคำนวณค่าความแตกต่างระหว่างระดับความเข้มของสัญญาณ (Signal Strength) นอกจากนี้ โมเดล AlexNet ยังใช้เทคนิคการกำหนดพารามิเตอร์ช่วยลดการเกิด Overfitting โดยใช้ Dropout ซึ่งทำการลบบางจำนวนโหนดของชั้นออกไปในระหว่างการฝึกโมเดลเพื่อลดความซับซ้อนและการเรียนรู้ที่เกินไป [20] โดยโครงสร้างของ AlexNet แสดงดังรูปที่ 3



รูปที่ 3 โครงสร้างของ AlexNet [23]

จากรูปที่ 3 แสดงโครงสร้างสถาปัตยกรรม Alexnet ภาพ Input จะถูกประมวลผลก่อนด้วยตัวกรองโดย 5 เลเยอร์ Convolutional C1-C5 ก่อนป้อนเข้าสู่เลเยอร์ FC6-FC8 ที่เชื่อมต่ออย่างสมบูรณ์3 เลเยอร์ การอ่านค่าสุดท้ายจะเกิดขึ้นที่ FC8 มีขั้นตอนการรวมสูงสุดเพิ่มเติมหลังจาก C1, C2 และ C5 ขนาดเชิงพื้นที่จะลดลงเรื่อย ๆ จาก C1 เป็น C3 ในขณะเดียวกันมีการใช้ตัวกรองมากขึ้นเรื่อย ๆ เดิมทีการประมวลผลถูกแยกออกเป็นสองสตรีมแยกกันและฝึกฝนบน GPU สองตัว

4. การดำเนินการวิจัย

ในขั้นตอนการดำเนินการวิจัยจะเริ่มการแยกวัตถุออกจากวิดีโอ และการจำแนกประเภทเรือ ดังนี้

4.1 การแยกวัตถุออกจากวิดีโอ

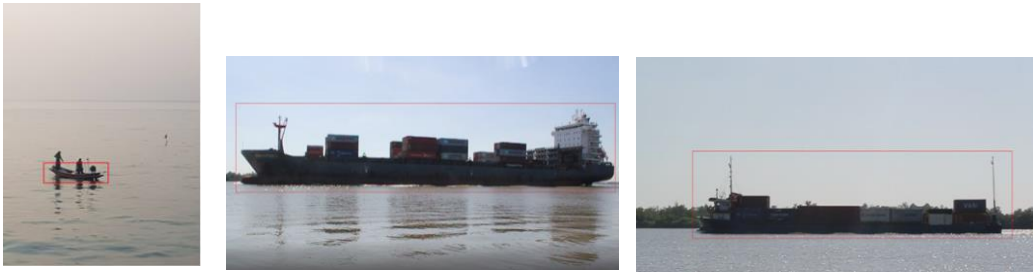
วิดีโอจะต้องถูกแปลงเป็นภาพเนื่องจากการป้อนข้อมูลในการฝึกสอนรูปแบบเป็นภาพ การแยกวัตถุออกจากวิดีโอเป็นกระบวนการที่ท้าทายและที่ยากที่สุดในการประมวลผลภาพดิจิทัล เนื่องจากวิดีโอเป็นภาพเคลื่อนไหว ซึ่งต้องใช้เทคโนโลยีที่ซับซ้อนและมีความซับซ้อนเพื่อการทำงานอย่างมีประสิทธิภาพ วิธีการแยกวัตถุออกจากวิดีโอสามารถทำได้หลายวิธี แต่ในงานวิจัยนี้ได้ใช้เทคโนโลยีการประมวลผลภาพและการเรียนรู้เชิงลึก AlexNet เป็นโมเดลโครงข่ายประสาทเทียมประเภท CNN (Convolutional Neural Network) ซึ่งมีขั้นตอนดังนี้

4.1.1 การจัดเตรียมข้อมูล ข้อมูลจะถูกแปลงเป็นภาพนิ่ง (Still image) หรือเฟรม (Frame) ของวิดีโอ และทำการปรับขนาดและตัดเฟรมตามขนาดที่ต้องการ

4.1.2 การแยกแยะวัตถุ ใช้โมเดล CNN เพื่อค้นหาวัตถุในเฟรม โดยแยกวัตถุจากพื้นหลังของเฟรม และทำการกำหนดกรอบสี่เหลี่ยม (Bounding Box) ที่อยู่บนวัตถุ

4.1.3 การติดตามวัตถุ ใช้เทคโนโลยีการเรียนรู้เชิงลึก RNN เพื่อติดตามวัตถุในวิดีโอ โดยสร้างโมเดลที่สามารถติดตามวัตถุในเฟรมก่อนหน้า และนำไปใช้ต่อเนื่องในเฟรมถัดไป

4.1.4 การคัดกรองวัตถุทำการคัดกรองวัตถุที่ไม่เกี่ยวข้อง และแสดงผลการตรวจจับภาพเรือ ดังรูปที่ 4



รูปที่ 4 ตัวอย่างผลการตรวจจับภาพเรือ

การกำหนดขั้นตอนสำหรับการแสดงภาพตัวอย่างด้วยหน้ากากเรือ และหน้ากากเข้ารหัสและถอดรหัสที่ใช้ภาพความจริงพื้นฐานในรูปแบบของมาสก์ จากนั้นโหลดโมเดล Mask-R-CNN จากที่เก็บแยก หลังจากนั้นจะสร้างคลาสที่มีสามฟังก์ชันสำหรับนำเข้าชุดข้อมูล มาสก์ และรูปภาพอ้างอิง และสร้างชั้นเรียนที่มีสามตำแหน่งสำหรับนำเข้าชุดข้อมูล รายการ และรูปภาพอ้างอิง นอกจากนี้ยังออกแบบสำหรับงานที่คล้ายกัน ซึ่งใช้ GPU การทำงาน โดย GPU เป็นสิ่งจำเป็นในสถานการณ์นี้ เนื่องจากชุดข้อมูลมีขนาดใหญ่ นอกจากนี้ เคอร์เนล Google Colab และ Kaggle เป็นทั้งผู้ให้บริการ GPU บนคลาวด์ฟรี ในขณะนี้ยังคงทำงานกับสคริปต์เพื่อโหลดชุดข้อมูลการฝึกสอน จากนั้นก็วางโมเดลผ่านขั้นตอนของมัน เพื่อให้ได้ผลลัพธ์ที่เร็วขึ้น ใช้เวลาในการฝึกฝนมากกว่านั้นมากเพื่อให้ได้การบรรจบกันที่เพียงพอ และเอฟเฟกต์ที่เหมาะสม แบบจำลองล้มเหลวในบางกรณีเมื่อเข้าใจผิดว่าเกาะเล็ก ๆ หรือก้อนกรวดชายฝั่งเป็นเรือ กล่าวอีกนัยหนึ่งโมเดลกำลังสร้างผลบวกปลอม ในบางครั้งแบบจำลองตรวจไม่พบเรือรบ กล่าวอีกนัยหนึ่ง โมเดลกำลังสร้างผลลบปลอม

4.2 การจำแนกประเภทเรือ

เมื่อได้ภาพวัตถุที่ต้องการมาแล้วจากขั้นตอนข้างต้นจะเข้าสู่ขั้นตอนของโครงข่ายประสาทเทียม โดยโครงข่ายจะทำการประมวลผลก่อนการเข้ารหัส (Pre-Processing) โดยประมวลผลภาพวัตถุที่ได้รับเข้ามาก่อนการเข้ารหัส เพื่อเตรียมข้อมูลให้เหมาะสมสำหรับการทำงานของโมเดลโครงข่ายประสาทเทียม เช่น การปรับความคมชัดของภาพ (Sharpness) การปรับความสว่างและความเข้มของภาพ (Brightness and Contrast) การลบสัญญาณรบกวน (Noise Reduction) และอื่น ๆ จากนั้นจะทำการเข้ารหัสภาพวัตถุ (Encoding) หลังจากที่โครงข่ายประสาทเทียมได้ทำการประมวลผลก่อนการเข้ารหัสแล้ว โครงข่ายจะทำการเข้ารหัสภาพวัตถุที่ได้รับเข้ามา โดยการแปลงภาพวัตถุให้อยู่ในรูปแบบของเวกเตอร์หรือพจน์ทางคณิตศาสตร์ที่เข้าใจง่ายกับโมเดลโครงข่ายประสาทเทียม เช่น การแปลงภาพวัตถุให้อยู่ในรูปแบบของรูปภาพสีเทา (Grayscale Image) หรือแปลงภาพวัตถุให้อยู่ในรูปแบบของเวกเตอร์คุณลักษณะ (Feature Vector) สุดท้ายจะเป็นการจำแนกแบบ (Learning) หลังจากที่โครงข่ายประสาทเทียมได้ทำการเข้ารหัสภาพวัตถุแล้ว โครงข่ายจะเริ่มทำการจำแนกแบบโดยการใช้ข้อมูลที่ได้รับมา

ในงานวิจัยนี้จะใช้โมเดล AlexNet ซึ่งเป็นโมเดลโครงข่ายประสาทเทียม ประเภท CNN (Convolutional Neural Network) มีโครงสร้างพื้นฐานที่ประกอบไปด้วย หน่วยประมวลผล (Processing Unit) และสถานะ (State) ซึ่งจะถูกส่งเข้าต่อกันในแต่ละรอบของการคำนวณ โดยแต่ละสถานะจะถูกเก็บไว้ เพื่อใช้ในการประมวลผลของรอบต่อไป โครงสร้างของ RNN จะประกอบไปด้วย 2 ส่วนหลัก คือ ส่วนประมวลผล และส่วนควบคุม โดยในส่วนประมวลผลจะประกอบไปด้วยหน่วยประมวลผลหลายตัวที่มีความสามารถในการรับข้อมูลจำนวนมากและเชื่อมต่อกันเป็นเครือข่าย ในขณะเดียวกันส่วนควบคุมจะทำหน้าที่ในการควบคุมการทำงานของหน่วยประมวลผลในแต่ละขั้นตอน โดยการส่งสัญญาณกลับมาจากสถานะของการคำนวณในขั้นตอนก่อนหน้านี้ นอกจากนี้ โมเดล RNN ยังสามารถมีการเพิ่มเติมส่วนประมวลผลเพิ่มเติมเข้าไป

เช่น การใช้ Long Short-Term Memory (LSTM) หรือ Gated Recurrent Unit (GRU) เพื่อเพิ่มความสามารถในการจำความหรือความสัมพันธ์ที่มีความซับซ้อนของข้อมูลได้ดีขึ้น

ในส่วนนี้จะใช้ Pooling ที่เป็นหนึ่งในชั้นของโมเดลโครงข่ายประสาทเทียม ที่ใช้สำหรับลดขนาดของข้อมูลภายในชั้นก่อนที่จะส่งต่อไปยังชั้นต่อไป

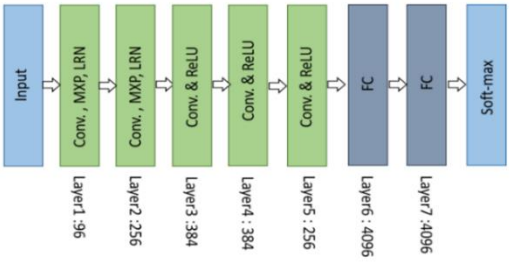
ขั้นตอนการทำงานของ AlexNet ในงานวิจัยนี้สามารถแบ่งได้เป็น 2 ขั้นตอนหลัก คือ

1) การเทรน (Training) และการทำงาน (Inference) การเทรน (Training)ในขั้นตอนแรก จะนำภาพมาเป็น Input ให้กับโมเดลแต่ละภาพจะถูกส่งผ่าน Convolutional Layer และ Pooling Layer เพื่อสกัดคุณลักษณะของภาพคุณลักษณะที่ได้จากการสกัดนี้จะถูกนำไปผ่าน Fully Connected Layer ซึ่งจะทำการหาค่าเชื่อมโยง (Weights) ระหว่าง Input กับ Output ในส่วนสุดท้าย ผลลัพธ์จะถูกนำไปใช้ในการคำนวณค่าความคลาดเคลื่อน (Loss) และใช้ Gradient Descent Algorithm ในการปรับค่าเชื่อมโยง (Weights) ให้เหมาะสมกับข้อมูลของภาพในชุดข้อมูล

2) การทำงาน (Inference) ในขั้นตอนนี้ จะนำภาพมาเป็น Input ให้กับโมเดลแต่ละภาพจะถูกส่งผ่าน Convolutional Layer และ Pooling Layer เพื่อสกัดคุณลักษณะของภาพคุณลักษณะที่ได้จากการสกัดนี้จะถูกนำไปผ่าน Fully Connected Layer ซึ่งจะทำการคำนวณเพื่อให้ได้ผลลัพธ์ที่เป็นคลาส (Class) ของภาพ โดยรวมแล้ว ขั้นตอนการทำงานของ AlexNet นั้นจะมีการใช้งาน Convolutional Layer, Pooling Layer, และ Fully Connected Layer เพื่อสกัดคุณลักษณะของภาพและจำแนกภาพให้ได้ถูกต้อง

ในการทดสอบของงานวิจัยนี้จะแสดงสถาปัตยกรรมของ AlexNet ในรูปที่ 5 โดยประการแรก Convolutional Layer ทำการ Convolution และ Max Pooling ด้วย Local Response Normalization (LRN) โดยที่ต่างกัน ใช้ตัวกรองตัวรับที่มีขนาด 11 × 11 สูงสุด ดำเนินการโดยใช้ตัวกรอง 3 x 3 การดำเนินการเดียวกันจะดำเนินการในครั้งที่สองชั้นที่มีตัวกรอง 5 × 5 ฟิลเตอร์ 3 × 3 ใช้ในอันที่สาม สี่ และเลเยอร์คอนโวลูชั่นที่ห้าที่มีคุณสมบัติ 384, 384 และ 296 ตามลำดับ ใช้เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์สองชั้นด้วยการออกกลางคันตามด้วยชั้น Softmax ในตอนท้ายสองเครือข่ายที่มีโครงสร้างใกล้เคียงกันและมีคุณลักษณะจำนวนเท่ากัน แผนที่ได้รับการฝึกฝนแบบคู่ขนานสำหรับโมเดลนี้ สองแนวคิดใหม่ Local Response Normalization (LRN) และการออกกลางคันที่แนะนำในเครือข่ายนี้ สามารถใช้ LRN ได้สองแบบวิธี: ใช้ครั้งแรกในแชนเนลเดียวหรือแผนที่คุณลักษณะโดยที่มีการเลือก N × N จากฟีเจอร์แผนที่เดียวกัน และทำให้เป็นมาตรฐานโดยอิงจากค่าย่านใกล้เคียง ประการที่สอง LRN สามารถนำไปใช้ข้ามช่องทางหรือแผนที่คุณลักษณะ (พื้นที่ใกล้เคียงตามมิติที่ 3 แต่มีพิกเซลเดียวหรือที่ตั้ง)

AlexNet มี 3 ชั้น Convolution และ 2 ชั้นที่เชื่อมต่ออย่างสมบูรณ์ เมื่อประมวลผลชุดข้อมูล ImageNet จำนวนทั้งหมดพารามิเตอร์สำหรับ AlexNet สามารถคำนวณได้ดังนี้ สำหรับค่าแรกชั้น: ตัวอย่าง Input คือ 224×224×3 Output ของชั้นการม้วนแรก คือ 55 × 55 × 96 สามารถคำนวณได้เท่านี้ก่อนชั้นนี้มีเซลล์ประสาท 290,400 (55×55×96) และ 364 (11 ×11×3 = 363) + 1) น้ำหนักพารามิเตอร์สำหรับชั้นการม้วนงอแรก คือ 290,400 × 364 = 105,705,600



รูปที่ 5 สถาปัตยกรรมของ AlexNet [24]

รูปที่ 6 การจำแนกประเภทเรือขนส่งสินค้า (Transport Ship)

5. ผลการวิจัย

แบบจำลองการจำแนกวัตถุที่นำเสนอได้รับการประเมินโดยการตรวจสอบประสิทธิภาพเมตริกต่าง ๆ เช่น ความแม่นยำ การเรียกคืน ฯลฯ สำหรับการตรวจสอบความถูกต้องของข้อมูล จากนั้น รูปแบบการติดตามวัตถุได้รับการประเมินโดยการวัดความแม่นยำในการติดตาม สำหรับการประเมินผลการปฏิบัติงาน

ความแม่นยำ การเรียกคืน คะแนน F1 เฉลี่ย ใช้ในการประเมินประสิทธิภาพของแบบจำลองการจำแนกวัตถุ มีการอ้างอิงถึงความแม่นยำของการคาดคะเนของแบบจำลองเป็นความแม่นยำ ความแม่นยำหมายถึงอัตราส่วนของผลลัพธ์ที่คาดการณ์ได้อย่างแม่นยำในบรรดาทั้งหมดผลการทำนายตามที่ระบุในสมการที่ (1)

$$precision = \frac{TP}{TP+FP} \tag{1}$$

โดยที่ TP (ผลบวกจริง) หมายถึง การทำนายผลลัพธ์จริงว่าเป็นจริง ในขณะที่ FP (ผลบวกลวง) หมายถึง การคาดคะเนผลลัพธ์ที่ผิดพลาดให้เป็นจริง สัดส่วนของวัตถุที่จดจำได้ถูกต้องจากจำนวนวัตถุทั้งหมดที่รับรู้โมเดลควรรับรู้เรียกว่าการเรียกคืน กล่าวอีกนัยหนึ่ง การเรียกคืน คือ เศษส่วนของการคาดคะเนจริงที่แบบจำลองทำนายท่ามกลางการคาดคะเนที่เป็นจริงแสดงออกมาในสมการที่ (2)

$$recall = \frac{TP}{TP+FN} \tag{2}$$

โดยที่ FN (ค่าลบเป็นเท็จ) หมายถึง การทำนายผลลัพธ์จริงเป็นเท็จ คะแนน F1 พิจารณาทั้งความแม่นยำและการเรียกคืน ซึ่งเป็นสัดส่วนผกผัน คะแนน F1 คำนวณเป็นค่าเฉลี่ยฮาร์มอนิกของความแม่นยำและการเรียกคืน ดังที่แสดงในสมการที่ (3) คะแนน F1 อยู่ในช่วงตั้งแต่ 0 ถึง 1 ซึ่งยิ่งโมเดลเข้าใกล้ 1 มากเท่าไหร่ถือว่ายิ่งดี

$$\mathrm{F1}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

หลักการเลือกการ Training และ Testing ของงานวิจัยนี้มีดังนี้ Training Set เลือกตัวแทนของข้อมูลจริงมากที่สุด เพื่อที่โมเดลจะได้เรียนรู้ลักษณะของข้อมูลจริง และสามารถทำงานได้ดีกับข้อมูลจริงมากที่สุด Testing Set เลือกชุดข้อมูลใหม่ที่ไม่ซ้ำกับ Training Set เพื่อทดสอบประสิทธิภาพของโมเดลกับข้อมูลใหม่ที่ไม่คาดคิด สัดส่วนของ Training Set และ Testing Set ได้แบ่ง Training Set ประมาณ 80% และ Testing Set ประมาณ 20% จากการศึกษานี้ได้แบ่งภาพรวมทั้งหมดออกเป็น 300 ภาพ โดยแบ่งเป็น Training Set จำนวน 200 ภาพ และ Testing Set จำนวน 100 ภาพ ซึ่งสัดส่วนนี้ถือว่าเหมาะสม เนื่องจาก Training Set มีขนาดเพียงพอที่จะเรียนรู้ลักษณะของข้อมูลจริง และ Testing Set มีขนาดเพียงพอที่จะทดสอบประสิทธิภาพของโมเดลกับข้อมูลใหม่

สำหรับการจำแนกประเภทเรือ 5 ประเภทดังกล่าว สามารถใช้ Confusion Matrix เพื่อประเมินประสิทธิภาพของการจำแนกได้ โดยมีขนาดเป็น 5 x 5 ดังตารางที่ 1

ตารางที่ 1 Confusion Matrix ประเมินประสิทธิภาพของการจำแนก

	เรือน้ำมัน	**เรือขนส่ง**	**เรือหาปลา**	**เรือรบ**	**เรือพายขนาดเล็ก**	**FN**	**TN**
ทำนายว่าเป็นเรือน้ำมัน	13 TP	7 FP	0 FP	0 FP	0 FP	0 FN	0 TN
ทำนายว่าเป็นเรือขนส่ง	7 FP	13 TP	0 FP	0 FP	0 FP	0 FN	0 TN
ทำนายว่าเป็นเรือหาปลา	0 FP	0 FP	12 TP	0 FP	3 FP	0 FN	5 TN
ทำนายว่าเป็นเรือรบ	0 FP	7 FP	0 FP	13 TP	0 FP	0 FN	0 TN
ทำนายว่าเป็นเรือพายขนาดเล็ก	0 FP	0 FP	1 FP	0 FP	11 TP	3 FN	5 TN

โดยที่ TN (True Negative) คือ จำนวนรูปภาพที่ถูกจำแนกว่าไม่ใช่เรือประเภทนั้น และ FP (False Positive) คือ จำนวนรูปภาพที่ถูกจำแนกว่าเป็นเรือประเภทนั้น แต่จริง ๆ แล้วไม่ใช่ ในงานวิจัยนี้มีจำนวนรูปภาพทั้งหมดจำนวน 300 ภาพ เป็นตัวอย่างการฝึก 200 ภาพ และตัวอย่างทดสอบ 100 ภาพ ในการฝึกสอนใช้ CONV ใช้จำนวน 5 ชิ้น ขนาด 1 1x 11 Max Pool ใช้จำนวน 3 ชิ้น ใช้ขนาด 3X3 และต้องการหาประสิทธิภาพของการจำแนก พบว่า มีจำนวนรูปภาพที่ถูกจำแนกถูกต้อง (True Positive) จำนวน 65 ภาพ และมีจำนวนรูปภาพที่ถูกจำแนกผิด (False Positive) จำนวน 35 ภาพ ซึ่งหมายความว่ามีจำนวนรูปภาพที่ถูกจำแนกว่าเป็นเรือประเภทที่ 1, 2, 3 หรือ 4 โดยไม่ใช่ประเภทที่แท้จริง ดังแสดงในตารางที่ 2

ตารางที่ 2 ผลลัพธ์การเปรียบเทียบ

	Predicted True (Positive)	**Predicted False (Negative)**
True (Positive)	True Positive (TP) = 62	False Negative (FN) = 3
False (Negative)	False Positive (FP) = 25	True Negative (TN) = 10

Recall = TP / (TP + FN) = 62 / (62 + 3) = 0.954 (หรือ 95.4%)

Precision = TP / (TP + FP) = 62 / (62 + 25) = 0.713 (หรือ 71.3%)

F1-score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.713 * 0.954) / (0.713 + 0.954) = 0.816 (หรือ 81.6%) ดังแสดงผลในตารางที่ 3

ตารางที่ 3 ค่าที่แสดงผล

	โมเดล AlexNet
Precision	71.3 %
Recall	95.4 %
F1 Score	81.6 %

ค่าที่เหมาะสมของ Precision, Recall, F1 Score ขึ้นอยู่กับบริบทของงานและเป้าหมายของโมเดล ซึ่งค่าทั้งสามยิ่งค่าสูงยิ่งแม่นยำ โดยทั่วไปแล้ว ค่า Precision ที่สูงจะบ่งชี้ว่าโมเดลสามารถระบุวัตถุได้อย่างแม่นยำ ในขณะที่ค่า Recall ที่สูงจะบ่งชี้ว่าโมเดลสามารถตรวจจับวัตถุทั้งหมดได้อย่างแม่นยำ ค่า F1 Score เป็นค่าเฉลี่ยถ่วงน้ำหนักของ Precision และ Recall ดังนั้นจึงถือเป็นตัวชี้วัดประสิทธิภาพที่ครอบคลุม สำหรับงานการติดตามวัตถุในบทความนี้มีวัตถุประสงค์ในการหลีกเลี่ยงการชนกันของเรือ ดังนั้นจึงถือว่าค่า Recall มีความสำคัญมากกว่า Precision เนื่องจากโมเดลควรตรวจจับวัตถุทั้งหมดให้ได้มากที่สุด เพื่อที่จะสามารถหลีกเลี่ยงได้ทันท่วงที จากค่า Precision, Recall, F1 Score ที่ได้จากการศึกษา พบว่า ค่า Recall สูงถึง 95.4% ซึ่งถือว่าอยู่ในระดับสูง จึงถือว่าโมเดลสามารถตรวจจับวัตถุทั้งหมดได้อย่างแม่นยำ อย่างไรก็ตาม ค่า Precision อยู่ที่ 71.3% ซึ่งถือว่ายังอยู่ในระดับปานกลาง จึงถือว่าโมเดลมีโอกาสระบุวัตถุผิดพลาดได้บ้าง ซึ่งอาจส่งผลให้เกิดการชนกันของเรือได้

6. สรุป

การเรียนรู้เชิงลึกกำลังเป็นงานวิจัยที่ได้รับความนิยมอย่างมากในปัจจุบัน ทิศทางการใช้ชั้น Convolution Neural Network Convolution ชั้น Pool และชั้นการเชื่อมต่อทั้งหมดและอื่น ๆ โครงสร้างพื้นฐานคุณสามารถปล่อยให้โครงสร้างเครือข่ายเรียนรู้และแยกคุณสมบัติที่เกี่ยวข้องและนำไปใช้ คุณลักษณะนี้ให้ความสะดวกมากมายสำหรับการศึกษาจำนวนมากกระบวนการสร้างแบบจำลองที่ซับซ้อนมาก นอกจากนี้ลึกการเรียนรู้ตอนนี้อยู่ที่การจำแนกภาพ การตรวจจับวัตถุการประมาณค่าทัศนคติและการแบ่งส่วนภาพ เป็นต้น ได้รับผลลัพธ์และความก้าวหน้าที่ยิ่งใหญ่มาก ในแง่หนึ่งความลึกแอปพลิเคชันการเรียนรู้นั้นกว้างมากและเก่งกาจสามารถทำงานต่อไปเพื่อขยายไปยังแอปพลิเคชันอื่น ๆ บนในทางกลับกัน ยังมีศักยภาพมากมายให้เรียนรู้ การศึกษาภายใต้การดูแลนั้นบรรลุผลสำเร็จอย่างมาก ความสำเร็จของการประยุกต์ใช้การเรียนรู้เชิงลึกในการเรียนรู้น่าจะเป็นแนวโน้มในอนาคต หลังจากนั้นในกรณีของมนุษย์หรือสัตว์ ส่วนใหญ่ไม่รู้ว่าสิ่งนั้นคืออะไรโดยรู้ชื่อของสิ่งนั้น ในสาขาคอมพิวเตอร์วิทัศน์ ในอนาคตคาดว่าจะเกิดขึ้นอีกในเนื้อหาของโครงข่ายประสาทเทียม (RNN) บนพื้นฐานของการเรียนรู้เชิงลึกจะกลายเป็นรูปแบบเครือข่ายที่เป็นที่นิยมมากและจะประสบความสำเร็จได้ดียิ่งขึ้นจะมีความก้าวหน้าในการวิจัยประยุกต์ที่มีความก้าวหน้ามากขึ้น

การเลือก AlexNet มาใช้เนื่องจากเป็นโมเดล Deep Learning แบบพรีเซ็ตที่มีประสิทธิภาพสูงในการจำแนกภาพ มีข้อดีหลายประการ คือ มีความลึก (Depth) ในการเรียนรู้คุณลักษณะของภาพ ทำให้สามารถสกัดคุณลักษณะของภาพได้มากขึ้น ใช้ Rectified Linear Units (ReLU) เป็น Activation Function ที่ทำให้การเทรนโมเดลเร็วขึ้นและลดการเกิดปัญหา Gradient Vanishing ใช้ Dropout ทำให้โมเดลไม่ Overfit กับข้อมูลเทรน ใช้ Local Response Normalization (LRN) เพื่อลด Overfitting และเพิ่มประสิทธิภาพในการจำแนกภาพ ใช้เทคนิค Data Augmentation เพื่อเพิ่มขนาดของชุดข้อมูลและลดความเสี่ยงของโมเดลในการ Overfit กับข้อมูลการใช้ GPU เพื่อคำนวณข้อมูลที่เร็วขึ้น ทำให้เทรนโมเดลได้เร็วขึ้น ด้วยความสามารถและประสิทธิภาพของ AlexNet ทำให้เป็นโมเดล Deep Learning ที่ใช้กับงานนี้ได้อย่างมีประสิทธิภาพ

อย่างไรก็ตามยังมีข้อเสียในด้านขนาดโมเดลที่ใหญ่และซับซ้อน โดย AlexNet มีความลึกถึง 8 ชั้น และมีจำนวนพารามิเตอร์มากกว่า 60 ล้าน ทำให้เป็นโมเดลที่ใช้เวลาในการเทรนนาน และต้องใช้ทรัพยากรการคำนวณที่มากกว่าโมเดลอื่น ๆ ความเสี่ยงต่อการ Overfitting หากไม่มีการใช้เทคนิค Regularization เช่น Dropout หรือ Weight Decay อาจทำให้โมเดล Overfit กับชุดข้อมูลเทรนได้ง่าย ต้องใช้ GPU เพื่อเทรนโมเดลโมเดลในปัจจุบันมีขนาดใหญ่และซับซ้อนมากขึ้น อันเป็นผล

ให้ต้องใช้ GPU หรือ Hardware Acceleration อื่น ๆ เพื่อคำนวณข้อมูลในการเทรนโมเดลได้เร็วขึ้น ไม่สามารถใช้กับภาพที่มีขนาดเล็ก AlexNet ถูกออกแบบมาเพื่อจำแนกภาพขนาดใหญ่ แต่ถ้ามีการใช้กับภาพขนาดเล็กจะทำให้เสียประสิทธิภาพในการจำแนกภาพ และการติดตั้งและใช้งาน AlexNet เป็นโมเดล Deep Learning ที่มีความซับซ้อน ต้องมีความรู้ความเข้าใจในการติดตั้งและใช้งานโมเดลเพื่อให้สามารถใช้งานได้อย่างถูกต้อง

ในการศึกษานี้ผลที่ได้พบว่า แบบจำลอง มีค่า Precision น้อยที่สุด และยังได้ค่า Recall และ F1 Score ที่สูง โดย Recall สูงแสดงว่าโมเดลสามารถจำแนกเรือผิวน้ำได้ดีได้ดี โดยไม่มีการพลาดของเรือผิวน้ำ (False Negative) มากนัก F1 Score สูงแสดงว่าโมเดลมีความแม่นยำสูงในการทำนายทั้งในกรณี Positive และ Negative โดยที่ไม่มีการพลาดหรือทำนายผิด (False Positive และ False Negative) มากนัก และ Precision ต่ำแสดงว่าโมเดลมีประสิทธิภาพในการจำแนกเรือผิวน้ำที่ต่ำกว่าโดยมีการทำนายเรือผิวน้ำเป็น Positive (True Positive) น้อยกว่า False Positive หรือการทำนายเป็น Positive โดยไม่มีเรือผิวน้ำ (False Positive) ดังนั้นเมื่อใช้ AlexNet ในการจำแนกและการติดตามเรือผิวน้ำ จะต้องใช้ Precision และ Recall ร่วมกันในการประเมินประสิทธิภาพของโมเดล โดยการเพิ่มประสิทธิภาพของโมเดลอาจจะต้องใช้วิธีการปรับแต่งโมเดล และการเพิ่มจำนวนข้อมูลสำหรับการฝึกโมเดลเพิ่มเติมให้มากขึ้น และใช้การทำความเข้าใจและวิเคราะห์ข้อมูลเพิ่มเติมเพื่อปรับปรุงโมเดลให้มีประสิทธิภาพสูงขึ้นในการจำแนกและการติดตามเรือผิวน้ำ

สรุปสุดท้ายคือ โมเดลสามารถจำแนกเรือผิวน้ำได้ดีเมื่อเทียบกับจำนวนเรือผิวน้ำที่มีอยู่จริงแต่ความแม่นยำในการจำแนกยังต้องพัฒนาเพิ่มเติม ดังนั้นจึงแนะนำให้ใช้โมเดลนี้ในงานที่ต้องการความแม่นยำในการจำแนกที่สูง เช่น งานวิจัยหรืองานที่มีความสำคัญในการคาดการณ์แม่นยำของผลลัพธ์ เช่นการทำนายการเกิดภัยพิบัติในทะเล และในการติดตามเรือผิวน้ำที่มีลักษณะเด่นเฉพาะอย่างเช่นเรือที่มีขนาดใหญ่

7. ข้อเสนอแนะ

AlexNet เป็นโมเดล Deep Learning ที่มีความแม่นยำสูงในการจำแนกภาพ และมีการนำไปใช้งานในหลากหลายแนวโน้ม รวมถึงการติดตามเรือผิวน้ำด้วยภาพจากดาวเทียมด้วยกัน เพื่อให้สามารถใช้ AlexNet ในการจำแนกและการติดตามเรือผิวน้ำได้อย่างเหมาะสม โดยมีข้อแนะนำเกี่ยวกับขั้นตอนในการนำไปใช้ ดังนี้

7.1 เตรียมข้อมูลในการสร้างโมเดลสำหรับการจำแนกและการติดตามเรือผิวน้ำด้วย AlexNet ต้องมีข้อมูลภาพที่เป็นรูปภาพของเรือผิวน้ำ โดยจะต้องเตรียมภาพให้มีขนาดและความละเอียดที่เหมาะสมกับโมเดล AlexNet ซึ่งสามารถพิจารณาได้จากข้อมูลของโมเดลเอง

7.2 เทรนโมเดล สามารถนำข้อมูลภาพที่เตรียมไว้มาเทรนโมเดล AlexNet โดยใช้ข้อมูลที่แบ่งแยกออกเป็นชุดเทรนและชุดทดสอบ เพื่อปรับพารามิเตอร์ของโมเดลให้มีค่าที่เหมาะสมกับข้อมูลภาพที่ใช้

7.3 ทดสอบและปรับปรุงโมเดล หลังจากนั้นก็สามารถทดสอบโมเดลในชุดข้อมูลทดสอบ เพื่อดูว่าโมเดลสามารถจำแนกและติดตามเรือผิวน้ำได้ถูกต้องหรือไม่ โดยควรมีการปรับค่าให้มีความเหมาะสมกับงาน

การใช้ Pooling ร่วมกับ AlexNet เป็นวิธีที่ช่วยเพิ่มประสิทธิภาพในการจำแนกและการติดตามเรือผิวน้ำได้ดีขึ้น โดยเฉพาะเมื่อต้องการลดขนาดของ Feature Map จากการสกัดภาพด้วย Convolutional Layer ในชั้นก่อนหน้า ด้วยเหตุนี้สามารถเลือกใช้ Pooling Layer ในชั้นถัดไป เพื่อลดขนาดของ Feature Map ลงอีก ซึ่งจะช่วยลดการคำนวณและประหยัดพื้นที่ในการเก็บข้อมูลได้ นอกจากนี้ การใช้ Pooling Layer ยังช่วยลด Overfitting ของโมเดลได้ด้วย โดยที่ Pooling Layer จะทำการลดขนาดของ Feature Map ลงในขั้นตอนการเรียนรู้ ทำให้โมเดลไม่สามารถจำจำนวนมากของรายละเอียดในรูปภาพได้ ซึ่งสามารถช่วยลด Overfitting ได้ การใช้ Pooling Layer ยังช่วยลดความซับซ้อนในโมเดลได้อีกด้วย โดยที่การใช้ Pooling Layer จะช่วยลดขนาดของ Feature Map แต่ยังรักษาข้อมูลที่สำคัญอยู่ ซึ่งจะช่วยให้โมเดลมีความเร็วในการเรียนรู้และการทำงานมากขึ้น ดังนั้น การใช้ Pooling Layer ร่วมกับ AlexNet จะช่วยเพิ่มประสิทธิภาพในการจำแนกและการติดตามเรือผิวน้ำได้อย่างมีประสิทธิภาพและประหยัดเวลาในการคำนวณ

8. กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยดี ขอขอบคุณกองวิชาวิศวกรรมศาสตร์ ฝ่ายศึกษาโรงเรียนนายเรือ กองทัพเรือที่ช่วยเอื้อเฟื้อทั้งอุปกรณ์และสถานที่ในการวิจัย จนงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

9. เอกสารอ้างอิง

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, 779–788.

[2] Zhang, Y., Wang, C., Wang, X.., Zeng,W., Liu,W., & Fairmot. (2020). On the fairness of detection and re-identification in multiple object tracking.

[3] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016, 3464–3468.

[4] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017, 3645–3649.

[5] Lee, S.J., Roh, M.I., & Oh, M.J. (2020). Image-based ship detection using deep learning. *Ocean Systems Engineering*. 10(4): 415–434.

[6] Jie, Y., Leonidas, L., Mumtaz, F., & Ali, M. (2021). Ship detection and tracking in inland waterways using improved YOLOv3 and Deep SORT. *Symmetry*. 13(2): 308.

[7] Huang, H., Sun, D., Wang, R., Zhu, C., & Liu, B. (2020). Ship target detection based on improved YOLO network. *Mathematical Problem Engineering*.2020: 1-10.

[8] Chen, X., Qi, L., Yang, Y., Luo, Q., Postolache, O., Tang, J., & Wu, H. (2020). Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis. *Journal of Advanced Transportation*. 2020: 1-12.

[9] Chen, X., Xu, X., Yang, Y., Wu, H., Tang, J., & Zhao, J. (2020). Augmented ship tracking under occlusion conditions from maritime surveillance videos. *IEEE Access*. 8: 42884–42897.

[10] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014, 580–587.

[11] Deriche, R. (1987). Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*. 91(3): 167–187.

[12] Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988, 15: 10–5244.

[13] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20-25 June 2005, 886–893.

[14] Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 60(2): 91–110.

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.

[16] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, 770–778.

[17] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017,7263–7271.

[18] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.

[19] Bochkovskiy, A., Wang, C.Y., & Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv 2020, arXiv:2004.10934.

[20] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 25(2): 1097-1105.

[21] Artificial Neural Networks. (2023). โครงข่ายประสาทเทียม. สืบค้น 9 เมษายน 2566, จาก https://shorturl.asia/PYZ2v.

[22] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324.

[23] Alexnet architecture. (2023). Alexnet architecture. สืบค้น 9 เมษายน 2566, จาก https://shorturl.asia/BD9rC.

[24] Moodley, D., & Haar, R. (2022). Automated recognition of the cricket batting backlift technique in video footage using deep learning architectures. *Nature Communications*, 13(1), 2506.