

ระบบการตรวจจับสัตว์ป่าอัตโนมัติโดยใช้ YOLOv12 ร่วมกับไมโครคอนโทรลเลอร์ ESP32-S3 Automated Wildlife Detection System Using YOLOv12 and ESP32-S3 Microcontroller

ณัชพล เกิดชนะ และ ชัยพร เขมะภตะพันธ์*

Natchapol Kerdchana and Chaiyaporn Khemapatapan*

สาขาวิชาวิศวกรรมคอมพิวเตอร์ วิทยาลัยวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาวิทยาลัทธิธรรกิจบัณฑิตย

กรุงเทพมหานคร 10210 ประเทศไทย

Department of Computer Engineering, College of Engineering and Technology, Dhurakij Pundit University,

Bangkok 10210, Thailand

*E-mail: chaiyaporn@dpu.ac.th

Received: Jun 05, 2025

Revised: Jul 23, 2025

Accepted: Jul 31, 2025

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลตรวจจับและจำแนกสัตว์ป่าด้วยเทคโนโลยีการเรียนรู้เชิงลึก โดยใช้โมเดล YOLOv12 ซึ่งได้รับการฝึกด้วยชุดภาพสัตว์ป่าจำนวน 12,500 ภาพ ที่รวบรวมจากฐานข้อมูล Kaggle, Google, Pixabay และภาพถ่ายจากพื้นที่จริง การทดลองประกอบด้วยเปรียบเทียบประสิทธิภาพของกล้องเลนส์มุมกว้าง 66 และ 160 องศา ที่ใช้ร่วมกับเซ็นเซอร์ตรวจจับความเคลื่อนไหวแบบอินฟราเรด (HC-SR501) และไมโครเวฟ (RCWL-0516) นอกจากนี้ ยังมีการพัฒนาระบบแจ้งเตือนแบบเรียลไทม์ โดยใช้ไมโครคอนโทรลเลอร์ ESP32-S3 เพื่อส่งภาพที่บันทึกไว้ผ่าน Wi-Fi ไปยังเซิร์ฟเวอร์ FastAPI โดยใช้โปรโตคอล HTTPS และส่งการแจ้งเตือนผ่าน Telegram Bot กลุ่มเป้าหมายที่สนใจตรวจจับของงานวิจัยนี้คือสัตว์ป่าขนาดใหญ่ที่มีความเสี่ยงต่อความขัดแย้งระหว่างมนุษย์กับสัตว์ป่า ได้แก่ ช้าง เสือ หมี และกระทิง การประเมินประสิทธิภาพของโมเดลใช้ค่า mean Average Precision (mAP), Precision และ Recall ผลการวิจัยพบว่า YOLOv12 ที่ปรับแต่งเพื่อการตรวจจับสัตว์ป่ามีค่า mAP@0.5 เท่ากับ 0.908 โดยจำแนกเสือได้แม่นยำที่สุด (mAP = 0.947) รองลงมาคือ ช้าง (mAP = 0.925), กระทิง (mAP = 0.901) และหมี (mAP = 0.859) กล้องเลนส์ 66 องศาให้ความแม่นยำในการตรวจจับเฉลี่ย 82.28% สูงกว่าแบบ 160 องศา (72.72%) เนื่องจากภาพจากเลนส์มุมกว้างลดขนาดสัตว์ในภาพเมื่ออยู่ระยะไกล ระบบสามารถแจ้งเตือนแบบเรียลไทม์ได้อย่างมีประสิทธิภาพ โดยเซ็นเซอร์ HC-SR501 เหมาะกับการตรวจจับระยะใกล้ โดยให้ค่าความแม่นยำในการตรวจจับเท่ากับ 84.43% ที่ระยะ 4 เมตร สำหรับช้าง และให้ค่าความแม่นยำในการตรวจจับเท่ากับ 80.46% ที่ 6 เมตรสำหรับกระทิง ในขณะที่เซ็นเซอร์ RCWL-0516 มีความแม่นยำมากกว่าในระยะไกลและในสภาพแวดล้อมที่มีสิ่งกีดขวาง โดยให้ค่าความแม่นยำในการตรวจจับเท่ากับ 69.43% ที่ระยะ 8 เมตร สำหรับช้าง และให้ค่าความแม่นยำในการตรวจจับเท่ากับ 65.17% ที่ 10 เมตรสำหรับหมี ระบบที่พัฒนานี้แสดงให้เห็นศักยภาพในการติดตามและแจ้งเตือนสัตว์ป่าอย่างคุ้มค่าและแม่นยำในการใช้งานจริง

คำสำคัญ: การตรวจจับวัตถุ ปัญญาประดิษฐ์ ไมโครคอนโทรลเลอร์ การเรียนรู้เชิงลึก อินเทอร์เน็ตประสานสรรพสิ่ง

Abstract

This research aimed to develop a wildlife detection and classification system using deep learning techniques. The model employed is YOLOv12, trained on a dataset of 12,500 wildlife images collected from Kaggle, Google, Pixabay and actual field photographs. The study involved performance comparison of 66 degree and 160 degree wide-angle lenses used in combination with infrared (HC-SR501) and microwave (RCWL-0516) motion sensors. In addition, a real-time alert system was also developed using the ESP32-S3 microcontroller to transmit captured images via Wi-Fi to a FastAPI server using the HTTPS protocol and to send alerts through a

Telegram Bot. The target group of interest in detection of this research was large wildlife species considered high-risk to human-wildlife conflict including elephants, tigers, bears, and gaurs. Model performance was evaluated using standard metrics including mean Average Precision (mAP), Precision, and Recall. The results revealed that the fine-tuned YOLOv12 model for wildlife detection had a mAP@0.5 of 0.908, with the highest accuracy in classifying tigers (mAP = 0.947), followed by elephants (mAP = 0.925), gaurs (mAP = 0.901), and bears (mAP = 0.859). The camera with 66-degree lens achieved an average detection accuracy of 82.28%, higher than that with 160-degree lens (72.72%) because the image from the wide-angle lens reduced the size of the animals at a distance in the image. The developed system successfully delivered real-time notifications with high reliability. The HC-SR501 sensor was suitable for close-range detection with 84.43% detection accuracy at 4 meters for elephants and 80.46% detection accuracy at 6 meters for gaurs. On the other hand, the RCWL-0516 sensor was more accurate at long range and in environments with obstacles with 69.43% detection accuracy at 8 meters for elephants and 65.17% detection accuracy at 10 meters for bears. This developed system demonstrates the potential for cost-effective and accurate wildlife monitoring and alerting in real-world applications.

Keywords: Object detection, Artificial intelligence, Microcontroller, Deep learning, Internet of things

1. บทนำ

การเฝ้าระวังและตรวจจับสัตว์ป่าในพื้นที่ธรรมชาติ มีบทบาทสำคัญต่อการลดความเสี่ยงจากการเผชิญหน้าระหว่างสัตว์ป่ากับมนุษย์ อย่างไรก็ตาม พื้นที่ห่างไกลหรือเข้าถึงได้ยากมักเผชิญความท้าทายในการติดตั้งและบำรุงรักษา ระบบเฝ้าระวัง เนื่องจากข้อจำกัดด้านพลังงานและการสื่อสาร จึงจำเป็นต้องมีระบบอัตโนมัติที่สามารถตรวจจับได้อย่างแม่นยำ รวดเร็ว ใช้พลังงานต่ำ และทำงานได้อย่างต่อเนื่อง

โมเดลตรวจจับวัตถุแบบ Single-stage detection อย่าง YOLO (You Only Look Once) มีจุดเด่นด้านความเร็วและความแม่นยำ เหมาะกับงานหลากหลายประเภท [1]-[3] โดยได้รับการพัฒนาอย่างต่อเนื่องตั้งแต่เวอร์ชันแรกจนถึง YOLOv12 เพื่อเพิ่มประสิทธิภาพในการตรวจจับวัตถุขนาดเล็ก และวัตถุในสภาพแวดล้อมซับซ้อน เช่น สัตว์ป่าที่กลมกลืนกับธรรมชาติ [1], [3], [4] อีกทั้งยังมีการพัฒนาเพื่อตอบโจทย์เฉพาะทาง เช่น การตรวจจับและติดตามหลายวัตถุพร้อมกัน [2], [5] รวมถึงการประยุกต์ใช้ในภาพถ่ายทางอากาศ [1] เทคโนโลยีการตรวจจับยังถูกนำไปประยุกต์ใช้อย่างหลากหลาย เช่น อุปกรณ์พกพาอย่าง Smartwatches ที่ใช้ Motion sensors เพื่อตรวจจับพฤติกรรมมนุษย์ [6] ระบบรักษาความปลอดภัยในบ้านอัจฉริยะ [7] และการนับจำนวนคนแบบเรียลไทม์ [8] ซึ่งแสดงให้เห็นถึงความยืดหยุ่นและความหลากหลายของระบบตรวจจับในบริบทต่าง ๆ

ในด้านฮาร์ดแวร์ ไมโครคอนโทรลเลอร์ ESP32 S3 เหมาะสำหรับการประมวลผลภาพ ด้วยขนาดกะทัดรัด

ประหยัดพลังงาน และมี Wi-Fi ในตัว [9] การผสมผสานความสามารถของ ESP32 S3 กับโมเดล YOLO ที่ปรับแต่งเฉพาะ ช่วยสร้างระบบตรวจจับสัตว์ป่าอัตโนมัติที่มีประสิทธิภาพ ต้นทุนต่ำ และพร้อมใช้งานในพื้นที่ห่างไกลที่มีข้อจำกัดด้านพลังงาน นอกจากนี้ งานวิจัยนี้ยังทดลองใช้เลนส์กล้องมุมกว้างสองชนิด ได้แก่ กล้องมุม 66 องศา และ 160 องศา พร้อมทั้งเพิ่มเซ็นเซอร์ตรวจจับความเคลื่อนไหว HC-SR501 และ RCWL-0516 เพื่อลดการใช้พลังงานโดยการกระตุ้นการทำงานของกล้องเฉพาะเมื่อมีการเคลื่อนไหวในพื้นที่ตรวจจับ ซึ่งช่วยเพิ่มความแม่นยำและประสิทธิภาพของระบบในสภาพแวดล้อมธรรมชาติที่มีข้อจำกัดด้านพลังงาน

จากพื้นฐานดังกล่าว งานวิจัยนี้มุ่งเน้นการพัฒนา ระบบตรวจจับสัตว์ป่าโดยใช้โมเดล YOLOv12 ที่ได้รับการปรับแต่งเฉพาะเพื่อการตรวจจับสัตว์ป่า โดยระบบจะถ่ายภาพและส่งข้อมูลไปประมวลผลบนเซิร์ฟเวอร์ FastAPI ผ่าน HTTPS พร้อมระบบแจ้งเตือนผ่าน Telegram ออกแบบให้ทำงานแบบเรียลไทม์ มีความแม่นยำสูง ใช้พลังงานต่ำ และสามารถใช้งานได้จริงในพื้นที่ธรรมชาติอย่างมีประสิทธิภาพ

2. วัสดุ อุปกรณ์และวิธีการวิจัย

การวิจัยนี้เป็นการพัฒนาระบบตรวจจับและจำแนกประเภทสัตว์ป่าด้วยไมโครคอนโทรลเลอร์ ESP32-S3 ร่วมกับเซ็นเซอร์ตรวจจับความเคลื่อนไหวและโมดูลกล้อง โดยภาพที่ได้จะส่งผ่านเซิร์ฟเวอร์เพื่อประมวลผลด้วยโมเดล YOLOv12 ที่ปรับแต่งเพื่อการตรวจจับสัตว์ป่า เช่น หมี เสือ กระต๊อง และ

ข้าง การวิจัยดำเนินการในรูปแบบเชิงทดลองและการพัฒนาระบบ โดยครอบคลุมตั้งแต่การรวบรวมข้อมูล การฝึกโมเดล การออกแบบและติดตั้งระบบฮาร์ดแวร์ซอฟต์แวร์ จนถึง การทดสอบและวิเคราะห์ผลลัพธ์ในสภาพแวดล้อม ดังแสดงใน Figure 1

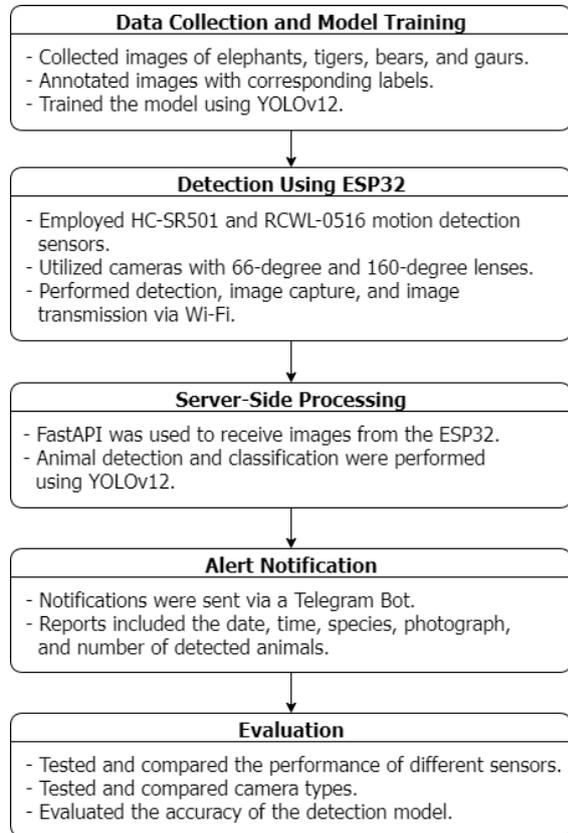


Figure 1 Workflow of the automated wildlife detection system

2.1. การรวบรวมข้อมูล

การรวบรวมข้อมูลสำหรับฝึกโมเดลสามารถทำได้จากแหล่งต่าง ๆ เช่น Kaggle Datasets, Google, Pixabay และถ่ายภาพจากพื้นที่จริง (สวนสัตว์เปิดในประเทศไทย) รวมภาพประมาณ 12,500 ภาพ หลังจากได้ภาพแล้วจะทำการเตรียมข้อมูล (Data preprocessing) โดยการทำความสะอาดข้อมูล เช่น ลบภาพซ้ำหรือมีคุณภาพต่ำ เช่น ภาพเบลอและแสงน้อย และปรับสมดุลของชุดข้อมูลให้แต่ละประเภทสัตว์มีจำนวนภาพใกล้เคียงกัน จากนั้นจะใช้เครื่องมือ Labelimg ในการ Annotate ข้อมูล โดยกำหนด Bounding Box และ Label สำหรับสัตว์แต่ละชนิด เช่น ช้าง, เสือ, กระต๊อง, หมี แบ่งชุดข้อมูลฝึกโมเดล 80% ตรวจสอบผลระหว่างการฝึก 10% และทดสอบหลังการฝึก 10%

2.2. การพัฒนาโมเดล YOLOv12 ที่ปรับแต่งเพื่อการตรวจจับสัตว์ป่า

การพัฒนาโมเดลตรวจจับวัตถุด้วย YOLOv12 เพื่อการตรวจจับสัตว์ป่าในงานวิจัยนี้ ได้มีการปรับแต่งพารามิเตอร์ในไฟล์คอนฟิกอย่างละเอียดเพื่อเพิ่มประสิทธิภาพและความแม่นยำของโมเดล โดยกำหนดจำนวนรอบการฝึกหรือ epochs ไว้ที่ 100 รอบ เพื่อให้โมเดลมีโอกาสเรียนรู้ฟีเจอร์ต่าง ๆ ของข้อมูลภาพสัตว์ป่าอย่างครบถ้วนและลึกซึ้ง โดยเฉพาะเมื่อมีการใช้เทคนิคการเพิ่มข้อมูลแบบ Augmentation ที่ช่วยเพิ่มความหลากหลายของข้อมูล ฝึกที่ขนาดภาพอินพุตที่ 640x640 พิกเซล เพื่อให้เกิดสมดุลระหว่างความแม่นยำและความเร็วในการประมวลผล ขนาดของ batch ถูกตั้งไว้ที่ 12 เพื่อให้เหมาะสมกับทรัพยากร GPU ที่มีอยู่และช่วยให้การฝึกมีความเสถียร นอกจากนี้ยังใช้การฝึกแบบ multi-GPU กับอุปกรณ์ 6 ตัว (Device "0,1,2,3,4,5") เพื่อเพิ่มความเร็วและประสิทธิภาพในการเรียนรู้ สำหรับการปรับแต่ง Optimizer เลือกใช้ AdamW ซึ่งเป็นเวอร์ชันที่พัฒนาจาก Adam โดยมีการแยก Weight decay ออกจากกลไกการปรับ Learning rate ช่วยลดปัญหา Overfitting ได้ดีกว่า ค่า Learning rate เริ่มต้นถูกตั้งไว้ที่ 0.0001 และลดลงไปจนถึง 0.01 ของค่าเริ่มต้นในช่วงท้ายของการฝึก โดยใช้กลยุทธ์ Cosine annealing ที่ช่วยให้ learning rate ลดลงอย่างนุ่มนวลตลอดการฝึก เพื่อให้โมเดลสามารถปรับน้ำหนักได้ละเอียดและแม่นยำมากขึ้น นอกจากนี้ยังตั้งค่า Patience ไว้ที่ 40 เพื่อใช้ Early stopping หาก validation loss ไม่ดีขึ้นภายใน 40 Epochs ช่วยลดเวลาการฝึกและป้องกันการ Overfitting พร้อมกับเปิดใช้ Automatic Mixed Precision (AMP) เพื่อเพิ่มความเร็วในการฝึกและลดการใช้หน่วยความจำ โดยใช้ข้อมูลทั้งแบบ float16 และ float32 ร่วมกัน มีการตั้งค่าการ Augmentation อย่างละเอียด โดยมีการพลิกภาพในแนวตั้งและแนวนอนแบบสุ่ม 50% เพื่อให้โมเดลสามารถเรียนรู้วัตถุจากมุมมองที่หลากหลาย ปรับค่าสี Hue แบบสุ่ม $\pm 1.5\%$ และปรับค่าสี Saturation สูงสุด 70% พร้อมปรับความสว่าง (Value) $\pm 40\%$ เพื่อจำลองสภาพแสงที่หลากหลายในธรรมชาติ นอกจากนี้ยังใช้การขยาย-ย่อภาพสูงสุด 30% การเลื่อนภาพไม่เกิน 20% ของขนาดภาพ หมุนภาพ ± 20 องศา และเอียงภาพ (Shear) ± 10 องศา เพื่อช่วยให้โมเดลเรียนรู้วัตถุในมุมที่หลากหลาย รวมถึงเพิ่ม Perspective distortion เล็กน้อยเพื่อเพิ่มความสมจริงในภาพ และใช้เทคนิค Mosaic augmentation ด้วยความน่าจะเป็น 10% ผสมภาพ 4 ภาพ

เข้าด้วยกัน รวมถึง Mixup augmentation ที่ผสมภาพและ label ของสองภาพด้วยความน่าจะเป็น 10% และ Copy-paste augmentation ที่คัดลอกวัตถุจากภาพหนึ่งไปวางในอีกภาพ เพื่อจำลองสภาพแวดล้อมที่ซับซ้อนและมีวัตถุซ้อนกัน

การปรับแต่งพารามิเตอร์และการใช้เทคนิค Augmentation ต่าง ๆ ทำให้การฝึกโมเดล YOLOv12 มีประสิทธิภาพ โดยในระหว่างกระบวนการฝึกจะมีการตรวจสอบค่าความผิดพลาด (Loss) บนชุดฝึกและชุดทดสอบอย่างต่อเนื่อง รวมถึงติดตามค่าประสิทธิภาพของโมเดล เช่น Precision, Recall และ mean Average Precision (mAP) หากพบว่าค่า confidence ของการตรวจจับต่ำกว่าเกณฑ์ที่ตั้งไว้ 80% จะดำเนินการปรับค่าพารามิเตอร์ใหม่ เช่น Learning rate, Batch size หรือ Anchor boxes และทำการฝึกซ้ำ เพื่อให้ค่าประสิทธิภาพเหล่านี้มีแนวโน้มเพิ่มขึ้นจนถึงเกณฑ์ที่กำหนดไว้ กระบวนการนี้ช่วยให้สามารถพัฒนาโมเดลที่มีความแม่นยำเพียงพอสำหรับการนำไปประยุกต์ใช้งานจริงในภายหลัง โดยเฉพาะในการตรวจจับสัตว์ป่าในสภาพแวดล้อมธรรมชาติ ดังแสดงใน Figure 2

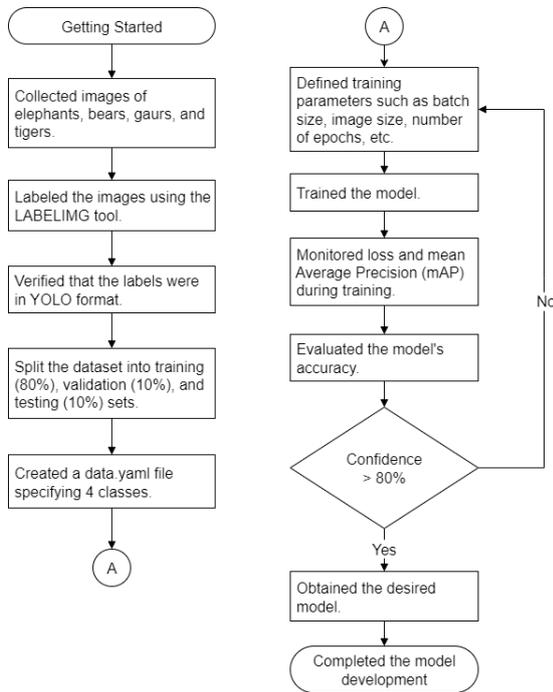


Figure 2 Workflow of the YOLOv12 model training process

2.3. การพัฒนาไมโครคอนโทรลเลอร์ด้วย ESP32

ระบบออกแบบการทำงานโดยอัตโนมัติสำหรับตรวจจับและส่งภาพเมื่อมีการเคลื่อนไหวของสัตว์ป่า โดยเริ่มจากการตั้งค่าเริ่มต้น เช่น การติดตั้ง SD card, เชื่อมต่อ Wi-Fi และตั้งเวลาโดยใช้ NTP จากนั้นเข้าสู่สถานะพร้อมทำงาน และเริ่มตรวจจับการเคลื่อนไหวผ่านเซนเซอร์ หากตรวจพบการเคลื่อนไหว ระบบจะทำการยืนยันอีกครั้งก่อนถ่ายภาพ และส่งภาพไปยังเซิร์ฟเวอร์ FastAPI ผ่านโปรโตคอล HTTPS เมื่อส่งภาพเสร็จ ระบบจะกลับเข้าสู่สถานะรอตรวจจับใหม่

การพัฒนาไมโครคอนโทรลเลอร์ด้วย ESP32-S3 สำหรับระบบตรวจจับสัตว์ป่า ใช้ฮาร์ดแวร์หลักคือ กล้องที่มีเลนส์มุมกว้างแตกต่างกัน 2 รุ่น ได้แก่ มุมกว้าง 66 และ 160 องศาเพื่อรองรับการจับภาพที่แตกต่างกัน โดยกล้องทั้งสองรุ่นถูกตั้งค่าความละเอียดระดับ SVGA และบันทึกภาพในรูปแบบ JPEG ซึ่งทำให้ได้ภาพที่มีความคมชัดและเหมาะสมต่อการนำไปวิเคราะห์

เลนส์มุม 66 องศา ถูกออกแบบมาเพื่อจับภาพมุมแคบทำให้ได้ภาพที่วัตถุมีรายละเอียดสูงและชัดเจน ขณะที่เลนส์มุมกว้าง 160 องศา เหมาะกับการตรวจจับภาพในพื้นที่กว้าง ช่วยให้ครอบคลุมบริเวณที่กว้างขึ้นแต่รายละเอียดของภาพต่ำลง ระบบกล้องติดตั้งฟังก์ชันควบคุมแสงอัตโนมัติ เช่น Auto exposure, Auto white balance (AWB), Gamma correction และ Lens correction เพื่อให้มั่นใจได้ว่าภาพที่ได้มีคุณภาพสูงในทุกสภาพแสงในเวลากลางวัน และกล้องได้ใช้หน่วยความจำ PSRAM ในการจัดเก็บเฟรมภาพระหว่างการจับภาพ ซึ่งช่วยลดภาระการใช้หน่วยความจำหลักของไมโครคอนโทรลเลอร์และเพิ่มความเร็วในการจับภาพ ทำให้สามารถบันทึกและส่งข้อมูลภาพได้รวดเร็วและต่อเนื่องมากขึ้น ระบบมีการใช้งาน SD card ทั้งสำหรับอ่านไฟล์ Configuration ที่ใช้ในการตั้งค่าการทำงาน และสำหรับเก็บภาพถ่ายที่บันทึกไว้ในกรณีที่มีการส่งข้อมูลผ่าน Wi-Fi ล้มเหลวหรือไม่เสถียร เพื่อเพิ่มความน่าเชื่อถือของระบบในการใช้งานในพื้นที่ที่มีข้อจำกัดด้านอินเทอร์เน็ต

ระบบตรวจจับการเคลื่อนไหวที่พัฒนาขึ้นนี้ ใช้เซนเซอร์หลักสองประเภท ได้แก่ Passive Infrared Sensor (PIR) ซึ่งตรวจจับการเปลี่ยนแปลงของความร้อนจากสิ่งมีชีวิต และ Radar sensor ซึ่งใช้คลื่นไมโครเวฟในการตรวจจับการ

เคลื่อนไหว โดย Radar sensor สามารถทำงานได้อย่างมีประสิทธิภาพแม้ในสภาพแวดล้อมที่มีสิ่งกีดขวางหรือในที่มืด

เซ็นเซอร์ทั้งสองชนิดเชื่อมต่อกับไมโครคอนโทรลเลอร์ ESP32-S3 ผ่านขา GPIO14 และส่งสัญญาณดิจิทัลระดับ High เมื่อมีการตรวจพบการเคลื่อนไหว เพื่อใช้เป็นตัวกระตุ้นให้กล้องทำการถ่ายภาพทันที กลไกนี้ช่วยให้ระบบประหยัดพลังงานโดยไม่จำเป็นต้องเปิดใช้งานกล้องตลอดเวลา

ระบบทั้งหมดใช้พลังงานจากแบตเตอรี่ลิเธียมไอออนขนาด 18650 จำนวน 2 ก้อน ที่เชื่อมต่อกันแบบอนุกรมให้แรงดันรวมประมาณ 7.4 โวลต์ จากนั้นผ่านโมดูล DC-DC Converter เพื่อลดแรงดันลงเหลือ 5 โวลต์ ซึ่งเหมาะสมกับการจ่ายไฟให้กับ ESP32-S3 และอุปกรณ์อื่น ๆ ในระบบ ดังแสดงใน Figure 3

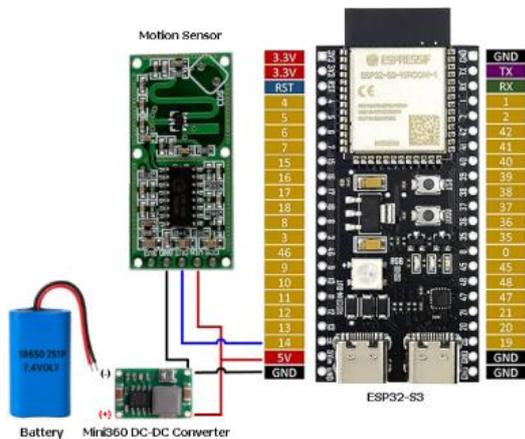


Figure 3 Interfacing devices with the ESP32-S3

ระบบยังใช้ฟังก์ชัน Deep sleep เพื่อเข้าสู่โหมดพักเมื่อไม่มีการเคลื่อนไหว ทำให้ใช้พลังงานต่ำมาก และสามารถปลุกตัวเองได้เมื่อเซ็นเซอร์ตรวจพบการเคลื่อนไหว เมื่อตัวกล้องถ่ายภาพเสร็จ ESP32-S3 จะส่งข้อมูลภาพผ่าน Wi-Fi ไปยังเซิร์ฟเวอร์ FastAPI ในรูปแบบ Multipart/form-data โดยมีการตรวจสอบความแรงสัญญาณ Wi-Fi ก่อนส่ง เพื่อให้การส่งข้อมูลมีความเสถียรและน่าเชื่อถือ พร้อมทั้งตั้งเวลาซิงค์ผ่าน NTP เพื่อให้เวลาบันทึกภาพตรงกับเวลาจริง ทำให้ข้อมูลภาพที่ได้มีความถูกต้องและสามารถนำไปใช้งานในระบบตรวจสอบได้อย่างแม่นยำ ดังแสดงใน Figure 4

2.4. การพัฒนาเซิร์ฟเวอร์ FastAPI

ผู้วิจัยเลือกใช้ FastAPI เป็นเซิร์ฟเวอร์หลักในการรับและประมวลผลภาพ เนื่องจากเป็นเฟรมเวิร์กที่มีความเร็วสูง

ใช้งานง่าย และรองรับการสร้าง RESTful API ได้อย่างยืดหยุ่น ระบบสามารถรับภาพจากกล้องหลายตัวพร้อมกันผ่านคำสั่ง HTTPS POST โดยใช้โครงสร้าง List[UploadFile] = File(...) ซึ่งเอื้อต่อการส่งภาพจากหลายอุปกรณ์เข้าสู่ระบบพร้อมกัน ช่วยลดเวลาและเพิ่มความสามารถของระบบ

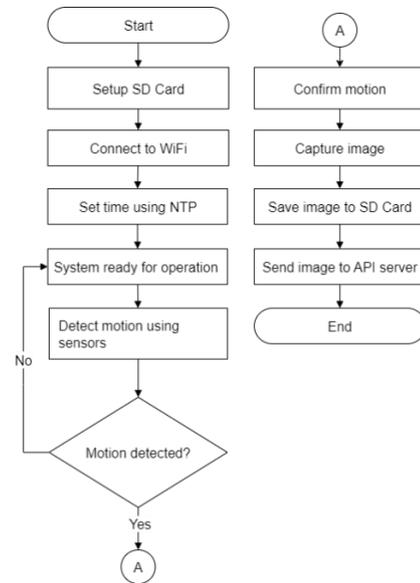


Figure 4 Operational workflow of the ESP32-S3 microcontroller

เพื่อจัดการกับปริมาณภาพจำนวนมากที่ส่งเข้ามา จึงได้นำเทคนิค Multithreading มาใช้ในการประมวลผล โดยแต่ละภาพจะถูกแยกออกมามีการจัดการใน Thread ของตัวเอง ทำให้สามารถประมวลผลภาพหลายภาพพร้อมกันได้ และช่วยลดระยะเวลาหน่วงในการตอบสนองของระบบโดยรวม เพิ่มความเร็วในการวิเคราะห์ข้อมูล

หลังจากรับภาพเข้ามาแล้ว ภาพทั้งหมดจะถูกจัดเก็บชั่วคราวไว้ในโฟลเดอร์ Upload/ เพื่อรอการประมวลผลต่อไป ระบบจะใช้โมเดล YOLOv12 ที่ปรับแต่งเพื่อการตรวจจับสัตว์ป่า โดยประมวลผลร่วมกับ OpenCV ซึ่งทำหน้าที่แปลงภาพเป็น Blob เพื่อให้เหมาะกับอินพุตของโมเดล เมื่อโมเดลตรวจพบสัตว์เป้าหมาย และมีค่าความเชื่อมั่น (Confidence score) มากกว่า 0.8 ภาพนั้นจะถูกจัดเก็บในโฟลเดอร์ Detected/ ส่วนภาพที่ตรวจไม่พบสัตว์เป้าหมายจะถูกลบทิ้งโดยอัตโนมัติ เพื่อลดภาระด้านการจัดเก็บ

ระบบยังออกแบบให้มีการตรวจสอบโฟลเดอร์ Upload/ เป็นระยะ เพื่อค้นหาไฟล์ที่อาจค้างจากปัญหาการประมวลผลก่อนหน้า หากพบไฟล์ที่ยังไม่ได้รับการวิเคราะห์ ระบบจะนำ

กลับเข้าสู่ขั้นตอนการตรวจจับอีกครั้งโดยอัตโนมัติ เพื่อให้มั่นใจว่าทุกภาพได้รับการประมวลผลอย่างครบถ้วน ดังแสดงใน Figure 5

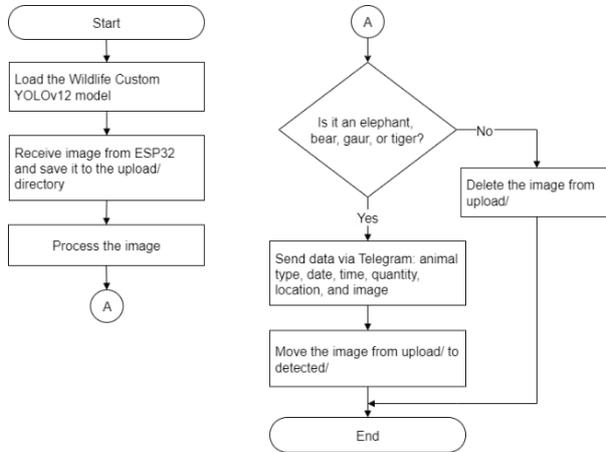


Figure 5 Workflow of the FastAPI server

เมื่อตรวจจับสำเร็จ ระบบจะดำเนินการแจ้งเตือนผ่าน Telegram Bot API โดยส่งข้อมูลครบถ้วนทั้งชนิดของสัตว์ จำนวนที่พบ สถานที่ เวลา และภาพที่ถ่ายได้ ไปยังผู้ใช้งานทันที เพิ่มขีดความสามารถของระบบในการเฝ้าระวังอย่างต่อเนื่องและแม่นยำตัวอย่างดังแสดงใน Figure 6

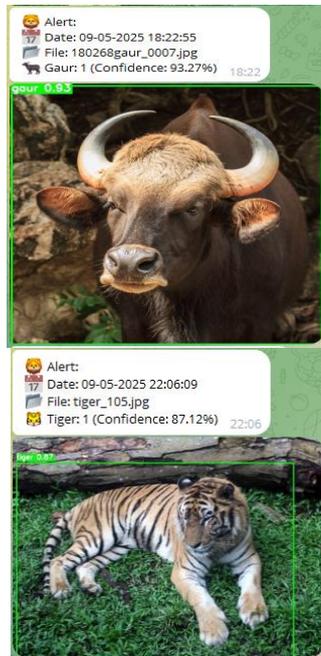


Figure 6 Notification results via Telegram application

3. ผลการวิจัยและอภิปรายผลการวิจัย

การวิเคราะห์ข้อมูล Annotation สำหรับการฝึกโมเดลตรวจจับวัตถุแสดงให้เห็นถึงลักษณะเชิงสถิติที่สำคัญของชุดข้อมูล โดยเฉพาะความไม่สมดุลของจำนวนตัวอย่างระหว่างคลาส โดยคลาส bear, elephant และ tiger มีจำนวน Bounding boxes ประมาณ 4,000-6,000 ภาพ ในขณะที่คลาส gaur มีเพียง 2,000 ภาพ ซึ่งส่งผลให้โมเดลมีความแม่นยำลดลง นอกจากนี้การกระจายของ Bounding boxes แสดงให้เห็นว่าวัตถุส่วนใหญ่อยู่ใกล้ศูนย์กลางของภาพ (พิกัดศูนย์กลางเฉลี่ยประมาณ $x \approx 0.5, y \approx 0.6$) และมีขนาดเล็กถึงปานกลาง โดยช่วงความกว้างและความสูงอยู่ในช่วง 0.1-0.3 ของขนาดภาพ ข้อมูลจาก Heatmap ยังยืนยันถึงความหนาแน่นของวัตถุที่บริเวณกึ่งกลางภาพ ซึ่งสอดคล้องกับพฤติกรรมการถ่ายภาพที่เน้นวัตถุเป็นจุดสนใจหลัก ข้อมูลเชิงลึกเหล่านี้มีประโยชน์อย่างยิ่งในการออกแบบกระบวนการ เช่น การสุ่มครอบภาพ การเลือก Anchor boxes ให้เหมาะสม และการปรับสมดุลของภาพก่อนฝึกโมเดล เพื่อเพิ่มประสิทธิภาพการตรวจจับวัตถุในทุกคลาสอย่างรอบด้าน ดังแสดงใน Figure 7

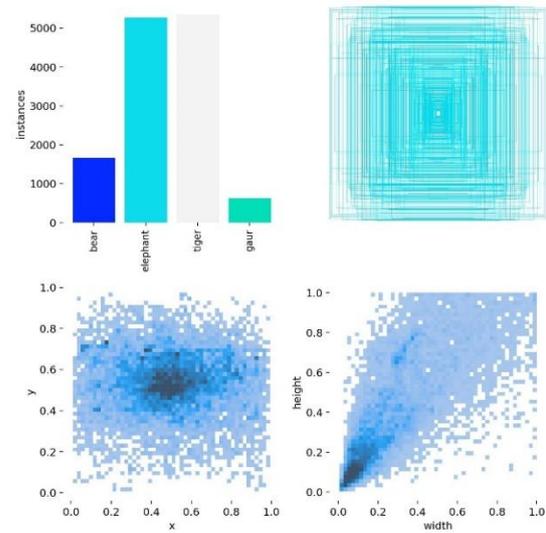


Figure 7 Annotation data for training an object detection model

สำหรับการจำแนกวัตถุ 5 ประเภท ได้แก่ bear, tiger, gaur, elephant และ background ช่วยให้เห็นภาพรวมของความแม่นยำและข้อผิดพลาดในการทำนายของโมเดลอย่างชัดเจน โดยค่าทางทแยงมุมแสดงจำนวนตัวอย่างที่โมเดลทำนาย

ถูกต้อง ส่วนค่านอกที่ยังมุงคือการทำนายผิด จากผลลัพธ์พบว่าโมเดลสามารถทำนายคลาส tiger ได้แม่นยำที่สุด (ถูกต้อง 511 ตัวอย่าง) รองลงมาคือ elephant (432 ตัวอย่าง) และ bear (364 ตัวอย่าง) อย่างไรก็ตาม พบว่ามีการสับสนกับคลาส background ค่อนข้างมาก โดยเฉพาะ bear ที่มีตัวอย่าง background ถูกทำนายผิดเป็น bear ถึง 202 ตัวอย่าง และ elephant ที่ถูกทำนายผิดเป็น background 111 ตัวอย่าง คลาส gaur มีจำนวนตัวอย่างน้อยและอัตราการทำนายผิดสูง โดยถูกทำนายผิดเป็น background ถึง 40 ตัวอย่าง จากจำนวนตัวอย่างจริง 171 ตัวอย่าง (ทำนายถูก 131) ส่งผลให้ความแม่นยำในคลาสนี้ต่ำกว่าคลาสนอื่น จึงควรเพิ่มข้อมูลฝึกฝนในคลาส gaur เพื่อช่วยให้โมเดลเรียนรู้ได้ดีขึ้นและลดความสับสนกับ background ดังแสดงใน Figure 8

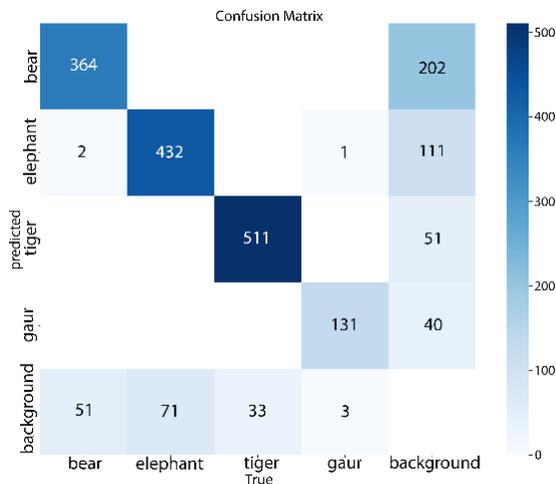


Figure 8 Confusion matrix results

เพื่อประเมินประสิทธิภาพของโมเดลสำหรับการตรวจจับวัตถุโดยพิจารณาจากความแม่นยำเทียบกับระดับความเชื่อมั่น การวิเคราะห์ Precision-confidence curve ในการทดลองนี้พบว่า

- ค่าความแม่นยำมีแนวโน้มเพิ่มขึ้นเมื่อค่า Confidence เพิ่มขึ้น เมื่อปรับเพิ่ม Threshold ของ Confidence โมเดลจะเลือกยอมรับเฉพาะการตรวจจับที่มีความมั่นใจสูง ซึ่งช่วยลดจำนวนการตรวจจับผิดพลาด (False positive)

- ประสิทธิภาพของโมเดลแตกต่างกันในแต่ละคลาส โมเดลสามารถตรวจจับคลาส tiger และ gaur ได้แม่นยำ แม้ในช่วง Confidence ต่ำ ในขณะที่คลาส bear แสดงค่าความแม่นยำต่ำในทุกระดับความเชื่อมั่น ดังแสดงใน Figure 9

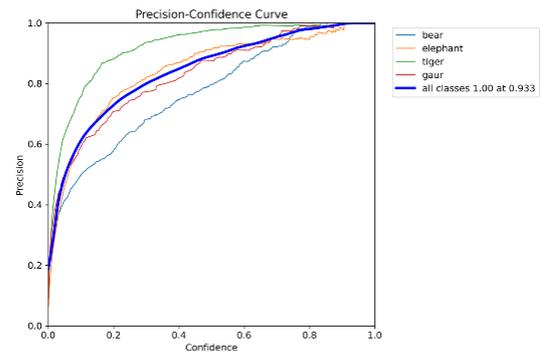


Figure 9 Precision-confidence curve of the YOLOv12 model

การประเมินประสิทธิภาพของโมเดลสำหรับการตรวจจับวัตถุ โดยเฉพาะในกรณีที่มีข้อมูลมีความไม่สมดุลระหว่างคลาส (Class imbalance) การวิเคราะห์ PR curve ในการทดลองนี้พบว่า

- คลาส tiger และ gaur มีค่า Average precision (AP) เท่ากับ 0.956 และ 0.943 ตามลำดับ แสดงว่าโมเดลสามารถตรวจจับวัตถุในคลาสนี้ได้อย่างแม่นยำและครอบคลุม ซึ่งสะท้อนถึงค่า Precision และ Recall ที่สูง

- คลาส elephant และ bear มีค่า AP ในระดับกลางถึงสูง แสดงถึงความสามารถของโมเดลในการตรวจจับวัตถุในคลาสดังกล่าว

- ค่า mAP@0.5 รวมอยู่ที่ 0.908 ซึ่งถือว่าอยู่ในระดับที่ดีมาก สะท้อนให้เห็นว่าโมเดลมีความสามารถในการตรวจจับวัตถุในหลายคลาสได้อย่างมีประสิทธิภาพ ดังแสดงใน Figure 10

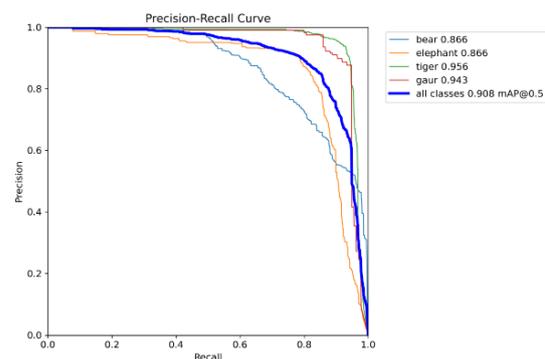


Figure 10 Precision-Recall curve of the YOLOv12 model

โมเดล YOLOv12 ที่ปรับแต่งเพื่อการตรวจจับสัตว์ป่า ได้รับการฝึกด้วยชุดข้อมูลจำนวน 12,500 ภาพ โดยดำเนินการฝึกทั้งหมด 100 รอบ (Epoch) เพื่อให้โมเดลสามารถเรียนรู้

ลักษณะของวัตถุเป้าหมายได้อย่างมีประสิทธิภาพ จากการประเมินผลหลังการฝึก พบว่าโมเดลมีความสามารถในการตรวจจับอยู่ในระดับที่น่าพอใจ โดยมีค่า Precision เฉลี่ย 86.7% และ Recall เฉลี่ย 84.0% ขณะที่ค่า mAP@0.5 ที่ 90.8% และค่า mAP@0.5:0.95 ที่ 66.1

จากการวิเคราะห์ตามชนิดของสัตว์ พบว่าโมเดลสามารถตรวจจับเสือ และ กระต๊องได้อย่างแม่นยำที่สุด โดยมีค่า mAP สูงกว่า 94% ขณะที่ ช้าง มีค่า Precision 88.9% และ Recall ที่ 78.9% ซึ่งอยู่ในเกณฑ์ดี ส่วน หมู มีค่า Precision 76.3% และ Recall ที่ 75.6% ซึ่งถือว่าอยู่ในระดับพอใช้ ดังแสดงใน Table 1

Table 1 Results of model training by animal category

Class	Precision	Recall	mAP	
			@0.5	@0.5:0.95
Bear	0.763	0.756	0.866	0.711
Tiger	0.967	0.868	0.956	0.616
Gaur	0.848	0.948	0.943	0.705
Elephant	0.889	0.789	0.866	0.612

ผลลัพธ์ดังกล่าวสะท้อนว่าโมเดล YOLOv12 ที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในงานตรวจจับสัตว์ป่าได้อย่างมีประสิทธิภาพ โดยเฉพาะในระบบตรวจจับแบบเรียลไทม์ที่ต้องการความแม่นยำและการตอบสนองที่รวดเร็ว

จากการทดลองตรวจจับสัตว์เป้าหมายในพื้นที่จริง (ในสวนสัตว์เปิดในประเทศไทย) โดยใช้เซ็นเซอร์ตรวจจับความเคลื่อนไหว HC-SR501 และ RCWL-0516 ร่วมกับกล้องเลนส์มุมมอง 66 องศา และ 160 องศา ดังแสดงใน Figure 11 และ Figure 12

ประสิทธิภาพการทำงานของระบบแตกต่างกันตามประเภทสัตว์ ระยะห่าง และอุปกรณ์ที่ใช้ โดยสัตว์ขนาดใหญ่ที่อยู่ใกล้ เช่น ช้าง (ระยะ 4 เมตร) สามารถตรวจจับได้แม่นยำที่สุด โดยเซ็นเซอร์ HC-SR501 ร่วมกับเลนส์ 66 องศา ได้ผลลัพธ์แม่นยำถึง 84.43% ส่วน RCWL-0516 ได้ 83.18% ในขณะที่เลนส์ 160 องศา มีประสิทธิภาพต่ำกว่าเล็กน้อยสำหรับกระต๊อง (ระยะ 6 เมตร) ประสิทธิภาพการตรวจจับยังอยู่ในระดับดี โดยอยู่ที่ประมาณ 80% ในกล้องมุมมอง 66 องศา และลดลงเมื่อใช้กล้อง 160 องศา เสือและหมูที่อยู่ในระยะไกลกว่า (8-10 เมตร) สามารถตรวจจับได้เพียงเมื่อใช้เลนส์ 66 องศา ร่วมกับ RCWL-0516 โดยเสียมีค่าแม่นยำ

69.43% และหมู 65.17% แสดงให้เห็นว่าเลนส์มุมกว้าง 160 องศาไม่เหมาะกับการตรวจจับสัตว์ที่อยู่ไกลและมีขนาดเล็กในภาพ ดังแสดงใน Table 2



Figure 11 66° and 160° lenses with RCWL-0516 motion sensor



Figure 12 66° and 160° lenses with HC-SR501 motion sensor

Table 2 Confidence results from sensor and camera module relative to object ranges

Type	Range (m)	HC-SR501		RCWL-0516	
		Field of view		Field of view	
		66°	160°	66°	160°
Elephant	4	84.43	73.31	83.18	74.31
Gaur	6	80.46	72.16	81.06	71.27
Tiger	8	-	-	69.43	-
Bear	10	-	-	65.17	-

ผลจากการถ่ายภาพด้วยเลนส์มุมแคบ 66 องศา (Figure 13 และ Figure 14) พบว่าวัตถุในภาพมีขนาดใหญ่เพียงพอสำหรับการตรวจจับได้อย่างชัดเจน เหมาะสำหรับการกรณีที่ต้องการความแม่นยำในการจำแนกวัตถุ โดยเฉพาะในระยะใกล้ซึ่งรายละเอียดของสัตว์สามารถรับรู้ได้อย่างชัดเจน เลนส์มุมกว้าง 160 องศา (Figure 15 และ Figure 16) แม้จะครอบคลุมพื้นที่ได้กว้างกว่า จึงมีข้อดีในการตรวจสอบบริเวณที่กว้างขึ้น แต่ส่งผลให้ขนาดของวัตถุในภาพเล็กลง โดยเฉพาะในกรณีที่วัตถุมีขนาดเล็กหรืออยู่ในระยะไกล เช่น เสือที่ปรากฏใน Figure 16 ไม่สามารถตรวจจับได้ ขณะที่ช้างใน Figure 15 ซึ่งมีขนาดใหญ่ สามารถตรวจจับได้ชัดเจน

ระยะห่างระหว่างวัตถุกับเซนเซอร์ตรวจจับความเคลื่อนไหวมีผลต่อการส่งงานกล้องถ่ายภาพ โดยกรณีเสือที่อยู่ห่างประมาณ 10 เมตร พบว่าเซนเซอร์ RCWL-0516 ซึ่งมีระยะตรวจจับได้ไกล สามารถสั่งให้กล้องทำงานได้ ขณะที่เซนเซอร์ HC-SR501 ซึ่งมีระยะตรวจจับสั้นกว่าไม่สามารถสั่งให้กล้องถ่ายภาพได้ อย่างไรก็ตาม แม้เซนเซอร์ RCWL-0516 จะสามารถสั่งถ่ายภาพได้ในระยะไกล แต่หากวัตถุมีขนาดเล็กเกินไปหรือปรากฏอยู่บริเวณขอบภาพของเลนส์มุมกว้าง ความแม่นยำในการตรวจจับของโมเดลจะลดลง แสดงให้เห็นว่าการเลือกใช้เลนส์และเซนเซอร์ ต้องทำงานร่วมกันได้อย่างเหมาะสมโดยคำนึงถึงระยะการตรวจจับและขนาดของวัตถุเป้าหมายที่ต้องการ

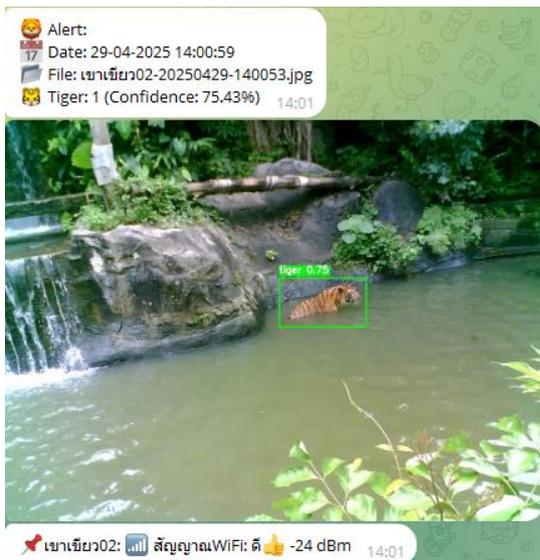


Figure 13 66° lens with RCWL-0516 sensor (tiger)

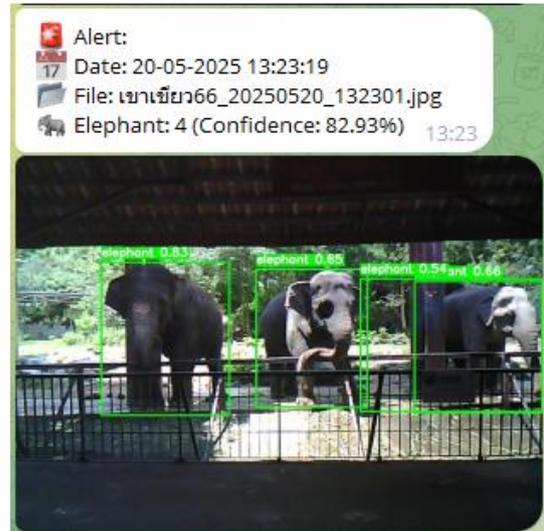


Figure 14 66° lens with HC-SR501 sensor (elephant)

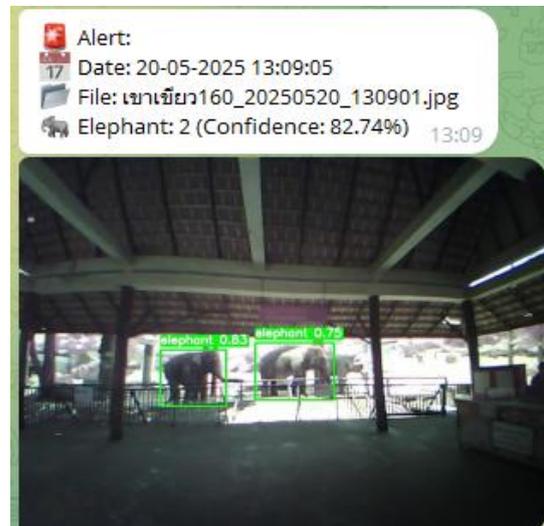


Figure 15 160° lens with RCWL-0516 sensor (elephant)

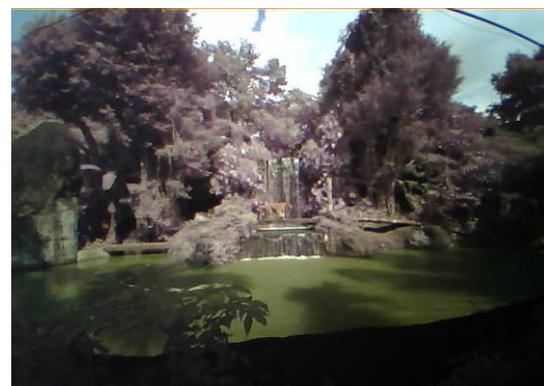


Figure 16 160° lens with RCWL-0516 sensor (tiger)

4. บทสรุป

งานวิจัยนี้พัฒนาโมเดลตรวจจับสัตว์ป่าโดยใช้ YOLOv12 ร่วมกับไมโครคอนโทรลเลอร์ ESP32-S3 เพื่อรองรับการทำงานแบบเรียลไทม์ในพื้นที่จริง โดยรวบรวมและเตรียมข้อมูลภาพสัตว์จำนวน 12,500 ภาพ พร้อมทำ Annotation และปรับแต่งพารามิเตอร์เพื่อฝึกโมเดลให้สามารถตรวจจับสัตว์ ได้แก่ ช้าง เสือ กระทิง และหมิ

โมเดลที่ผ่านการฝึกมีประสิทธิภาพสูง โดยได้ค่า mAP@0.5 เท่ากับ 90.8% พร้อม Precision เฉลี่ย 86.7% และ Recall เฉลี่ย 84.0% โดยเฉพาะเสือและกระทิงที่มีความแม่นยำสูงสุด แสดงถึงความสามารถของโมเดลในการเรียนรู้ลักษณะ เฉพาะของสัตว์เป้าหมายอย่างแม่นยำ

ระบบฮาร์ดแวร์ประกอบด้วยกล้องเลนส์สองมุมมอง (66 และ 160 องศา) ร่วมกับเซ็นเซอร์ PIR และ Radar เชื่อมต่อกับ ESP32-S3 ซึ่งสามารถเข้าสู่โหมด Deep sleep เพื่อประหยัดพลังงาน และส่งภาพผ่าน Wi-Fi ไปยังเซิร์ฟเวอร์ FastAPI พร้อมแจ้งเตือนผ่าน Telegram Bot แบบเรียลไทม์

ผลการทดสอบถ่ายภาพสัตว์เป้าหมายพบว่ากล้องมุมแคบ 66 องศา ร่วมกับเซ็นเซอร์ Radar ให้ผลตรวจจับแม่นยำที่สุด โดยเฉพาะกับสัตว์ขนาดใหญ่ในระยะไกล เช่น ช้างที่ระยะ 4 เมตร ได้ความแม่นยำสูงถึง 84.43% ด้วยเซ็นเซอร์ PIR และ 83.18% ด้วย Radar ส่วนเสือและหมิที่อยู่ระยะไกลกว่า (8-10 เมตร) ตรวจจับได้แม่นยำ 69.43% และ 65.17% ตามลำดับ แสดงถึงประสิทธิภาพของระบบในการตรวจจับและเฝ้าระวังได้อย่างมีประสิทธิภาพ

หากเปรียบเทียบระบบที่พัฒนาขึ้นกับการใช้กล้องดักถ่าย (Camera trap) กล้องดักถ่ายไม่สามารถประมวลผลภาพและแจ้งเตือนได้ทันที ในขณะที่ระบบที่พัฒนาขึ้นสามารถประมวลผลและส่งแจ้งเตือนได้ในทันที เพิ่มประสิทธิภาพในการเฝ้าระวังและตอบสนองต่อเหตุการณ์ได้รวดเร็วกว่า นอกจากนี้เมื่อเปรียบเทียบกับแพลตฟอร์มอย่าง Raspberry Pi ที่แม้จะมีประสิทธิภาพด้านการประมวลผลสูงกว่า แต่มีข้อจำกัดด้านการใช้พลังงาน ระบบที่ใช้ ESP32-S3 จึงเหมาะสมกว่า เนื่องจากสามารถเข้าสู่โหมดประหยัดพลังงาน (Deep sleep) และทำงานร่วมกับเซ็นเซอร์ต่าง ๆ ได้อย่างมีประสิทธิภาพ

สำหรับการพัฒนาต่อยอดในอนาคต ควรมุ่งเน้นการเพิ่มความละเอียดของกล้องถ่ายภาพและเซ็นเซอร์ตรวจจับ การเคลื่อนไหว เพื่อยกระดับความแม่นยำในการตรวจจับสัตว์ที่มีขนาดเล็กและอยู่ในระยะไกล รวมถึงการปรับปรุงและ

พัฒนาโมเดลปัญญาประดิษฐ์ให้สามารถแยกแยะพฤติกรรมของสัตว์ได้อย่างละเอียดมากขึ้น พร้อมทั้งรองรับการตรวจจับชนิดสัตว์ที่หลากหลาย เพื่อให้ระบบเฝ้าระวังสัตว์ป่ามีความครอบคลุมและแม่นยำยิ่งขึ้น นอกจากนี้การนำเทคโนโลยีเซ็นเซอร์ตรวจจับที่มีประสิทธิภาพสูงและตอบสนองได้รวดเร็ว มาใช้ร่วมกับระบบ ยังจะช่วยเสริมประสิทธิภาพการตรวจจับที่มีความซับซ้อนและเปลี่ยนแปลงได้อย่างมีประสิทธิภาพมากขึ้น

5. References

- [1] Yan, X., Shen, B. and Li, H. 2023. Small object detection method for UAVs aerial image based on YOLOv5s. In: **Proceedings of the 2023 IEEE 6th International Conference on Electronic Information and Communication Technology**, July 21-24, 2023. Qingdao, China.
- [2] Sahay, A., Singh, K.V. and Ponsam, G. 2023. Multi-object detection and tracking using machine learning. In: **Proceedings of the 2023 International Conference on Computer Communication and Informatics**, 23-25 January 2023. Coimbatore, India.
- [3] Leonid, T.T. and et al. 2023. Human wildlife conflict mitigation using YOLO algorithm. In: **Proceedings of the 2023 Eighth International Conference on Science Technology Engineering and Mathematics**, 6-7 April 2023. Chennai, India.
- [4] Cardelicchio, A. and et al. 2023. Tomato detection in challenging scenarios using YOLO-based single stage detectors. In: **Proceedings of the 2023 IEEE International Workshop on Metrology for Agriculture and Forestry**, 6-8 November 2023. Pisa, Italy.
- [5] Chen, X. and Zhai, Y. 2023. A multi-objective traffic flow detection system based on an improved YOLOv4 algorithm. In: **Proceedings of the 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms**, 24-26 February 2023. Changchun, China.
- [6] Ono, S., Nishiyama, Y. and Sezaki, K. 2022.

- Detecting face-mask wearing status using motion sensors in commercially available smartwatches. In: **Proceedings of the 2022 IEEE International Conference on E-health Networking, Application & Services**, 17-19 October 2022. Genoa, Italy.
- [7] Kumar, V.M., Ajina, A. and Deepak, D.J. 2023. Improving smart home safety with face recognition using machine learning. In: **Proceedings of the 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems**, 19-21 April 2023. Bangalore, India.
- [8] Elaoua, A., Nadour, M. and Cherroun, L. 2023. Real-time people counting system using YOLOv8 object detection. In: **Proceedings of the 2023 2nd International Conference on Electronics, Energy and Measurement**, 28-29 November 2023. Medea, Algeria.
- [9] Nale, P. and Gite, S. 2023. Real-time weapons detection system using computer vision. In: **Proceedings of the 2023 Third International Conference on Smart Technologies, Communication and Robotics**, 9-10 December 2023. Sathyamangalam, India.