



การระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ
สั่งอัตโนมัติ

Vehicle Physical Appearance Identification Using
Convolutional Neural Network

คหบดี ประกอบชาติ
Kahabodee Prakobchat

วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Data Science
Prince of Songkla University

2566

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์



การระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ
สั่งอัตโนมัติ

Vehicle Physical Appearance Identification Using
Convolutional Neural Network

คหบดี ประกอบชาติ
Kahabodee Prakobchat

วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Data Science
Prince of Songkla University

2566

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์


ชื่อวิทยานิพนธ์ การระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ
 สัณฐานการ

ผู้เขียน นายคหบดี ประกอบชาติ

สาขาวิชา วิทยาการข้อมูล

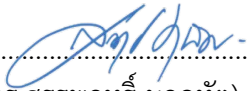
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

คณะกรรมการสอบ



 (ผู้ช่วยศาสตราจารย์ ดร.ขวัญกมล ดิฐักัญจน์)


ประธานกรรมการ
 (ผู้ช่วยศาสตราจารย์ ดร.สร้าง มุสิกสุวรรณ)


กรรมการ
 (ดร.สรรพทธี มฤคหัต)


กรรมการ
 (ผู้ช่วยศาสตราจารย์ ดร.ขวัญกมล ดิฐักัญจน์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็น
 ส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

.....
 (ผู้ช่วยศาสตราจารย์ ดร.กวินพัฒน์ สิริกานติโสภณ)
 รักษาการแทนคณบดีบัณฑิตวิทยาลัย

(3)

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มี
ส่วนช่วยเหลือแล้ว

ลงชื่อ..... 

(ผู้ช่วยศาสตราจารย์ ดร. ชวัญกมล ดิฐักัญจน์)
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ลงชื่อ..... 

(นาย คหบดี ประกอบชาติ)

(4)

นักศึกษา ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....

(นาย คหบดี ประกอบชาติ)

นักศึกษา

ชื่อวิทยานิพนธ์ การระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ
สังวัตนาการ
ผู้เขียน นายคหบดี ประกอบชาติ
สาขาวิชา วิทยาการข้อมูล
ปีการศึกษา 2566

บทคัดย่อ

การละเมิดกฎจราจรโดยผู้ขับขี่ยานพาหนะเป็นปัญหาสำคัญระดับโลก โดยเฉพาะในเขตเมืองเพราะสิ่งเหล่านี้ทำให้เกิดอุบัติเหตุจราจรเพิ่มมากขึ้น การศึกษานี้เสนอแนวทางใหม่ในการระบุยานพาหนะที่เกี่ยวข้องกับการละเมิดกฎจราจรโดยการสร้างตัวโมเดลการตรวจจับยานพาหนะโดยใช้เทคนิคการประมวลผลภาพ และการเรียนรู้ของเครื่อง การวิจัยมุ่งเน้นไปที่คุณลักษณะสำคัญของยานพาหนะ 3 ประเภท ได้แก่ ประเภทยานพาหนะ สี และแบรนด์ งานวิจัยนี้ศึกษาการใช้เทคนิคการถ่ายโอนการเรียนรู้ซึ่งเป็นวิธีการเรียนรู้ของเครื่องสร้างตัวโมเดลการจำแนกผลการวิจัยพบว่า YOLO V8 มีความแม่นยำสูงสุดในการจำแนกการตรวจจับประเภท และสีของยานพาหนะ โดยมีความแม่นยำร้อยละ 83, และ 79 ตามลำดับ เมื่อเปรียบเทียบกับ YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 มีความแม่นยำต่ำตามลำดับ ส่วนในแง่ของการจำแนกการตรวจจับแบรนด์ของยานพาหนะ YOLO V8 มีความแม่นยำถึงร้อยละ 78 ซึ่งมีความแม่นยำมากกว่า YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 งานศึกษาวิจัยนี้แสดงให้เห็นถึงประสิทธิภาพของเทคนิคการประมวลผลภาพ และการถ่ายโอนการเรียนรู้ของโมเดลในการระบุยานพาหนะที่เกี่ยวข้องกับการละเมิดกฎจราจรได้อย่างแม่นยำ และสามารถพัฒนาตัวโมเดลที่มีประสิทธิภาพในการลดจำนวนอุบัติเหตุจราจรที่เกิดจากการละเมิดกฎจราจร ผลที่ได้จากงานวิจัยนี้สามารถนำไปประยุกต์ใช้และต่อยอดในการเพิ่มความปลอดภัยทางถนนและส่งเสริมเทคโนโลยีเพื่อแก้ไขปัญหาในสถานการณ์จริง

คำสำคัญ: การเรียนรู้เชิงลึก, การตรวจจับวัตถุ, YOLO, ยานพาหนะการตรวจจับลักษณะกายภาพ, การประมวลผลภาพ, โครงข่ายประสาทเทียมแบบสังวัตนาการ

Thesis Title Vehicle Physical Appearance Identification using Convolutional Neural Network
Author Mr. Kahabodee Prakobchat
Major Program Master of Science (Data Science)
Academic Year 2023

ABSTRACT

Traffic rule violations by drivers are a significant global concern, particularly in urban areas, as they contribute to increasing traffic accidents. This study proposed a novel approach to identifying vehicles involved in such violations by building vehicle detection model using image processing and machine learning techniques. The research focused on three key vehicle characteristics: type, colour, and brand. The study employed a transfer learning mechanism as the machine learning method to generate the prediction models. The results revealed that the YOLO V8 achieved the highest accuracy in predicting vehicle type and colour, with an accuracy of 83 % and 79 %, respectively. Comparatively, YOLO V7, YOLO V6, YOLO V5, YOLO V4, and YOLO V3 achieved lower accuracies. In terms of predicting vehicle brand, the YOLO V8 achieved an accuracy of 78 %, surpassing the accuracies of the YOLO V7, YOLO V6, YOLO V5, YOLO V4 and YOLO V3. These findings demonstrated the potential of image processing and machine learning techniques in accurately identifying vehicles involved in traffic violations and highlighted the opportunity to develop effective strategies to reduce the number of traffic accidents caused by rule violations. This research has significant implications for enhancing road safety and promoting advanced technologies to address real-world problems.

Keywords: Deep learning, Object detection, YOLO, Vehicle appearance detection, Image processing, Convolutional neural network

กิตติกรรมประกาศ

ขอขอบพระคุณ ดร.ชวัญกมล ดิฐกัญจน์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำปรึกษา และช่วยเหลือในทุก ๆ ด้านไม่ว่าจะเป็นการผลักดันและกระตุ้นในการทำงานวิจัย ข้าพเจ้ารู้สึกขอบคุณสำหรับทุกอย่างที่คอยช่วยเหลือมาตลอดอย่างสุดซึ้ง

ขอขอบพระคุณคณะกรรมการสอบ และคณาจารย์วิทยาลัยการคอมพิวเตอร์ทุกท่านที่คอยแนะนำ ช่วยเหลือ สำหรับงานวิจัยนี้และอื่น ๆ

ขอขอบคุณข้อมูลจากหน่วยงาน ในจังหวัดภูเก็ตที่มอบข้อมูลสำหรับการทำวิจัย

ขอขอบคุณทุก ๆ กำลังใจจากครอบครัว ทั้งนี้รวมทั้งคนที่มีส่วนเกี่ยวข้องและไม่เกี่ยวข้องในงานวิจัยนี้ที่คอยสนับสนุน ผลักดัน และให้กำลังใจมาตลอด ขอขอบคุณ ณ โอกาสนี้

คหบดี ประกอบชาติ

สารบัญ

	หน้า
บทคัดย่อ.....	(5)
ABSTRACT.....	(6)
กิตติกรรมประกาศ.....	(7)
สารบัญ.....	(8)
รายการตาราง.....	(10)
รายการภาพประกอบ.....	(11)
บทที่	
1 บทนำ.....	1
1.1 ที่มาและความสำคัญของงานวิจัย.....	1
1.2 คำถามงานวิจัย.....	5
1.3 วัตถุประสงค์.....	6
1.4 ระเบียบวิธีวิจัย.....	6
1.5 ข้อยกเว้น.....	8
1.6 ขอบเขตงานวิจัย.....	8
1.7 สรุป.....	9
2 การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง.....	10
2.1 บทนำ.....	10
2.2 โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network) และการเรียนรู้เชิงลึก.....	10
2.3 การประมวลผลภาพ (Image Processing).....	17
2.4 การจำแนกประเภทข้อมูลภาพ (Image Classification).....	20
2.5 การตรวจจับวัตถุ (Object detection).....	21
2.6 การถ่ายโอนการเรียนรู้ (Transfer learning).....	25
2.7 You Only Look Once (YOLO).....	27
2.8 งานวิจัยที่เกี่ยวข้อง.....	47
2.9 การตรวจวัดประสิทธิภาพ.....	53
2.10 สรุป.....	57
3 ชุดข้อมูลภาพและกระบวนการเตรียมความพร้อมของภาพ.....	58
3.1 บทนำ.....	58
3.2 ที่มาของชุดข้อมูล.....	58
3.3 แยกเฟรมภาพจากวิดีโอ.....	59
3.4 การเพิ่มปริมาณข้อมูล.....	60
3.5 ตัดป้ายกำกับ.....	61

สารบัญ(ต่อ)

	หน้า
3.6 สรุป.....	63
4 การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทของยานพาหนะ.....	64
4.1 บทนำ.....	64
4.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set).....	64
4.3 การวัด และ ประเมินผลจำแนกประเภทของยานพาหนะ.....	67
4.4 สรุปผล.....	71
5 การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทของสียานพาหนะ.....	73
5.1 บทนำ.....	73
5.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set).....	73
5.3 การวัด และ ประเมินผลการจำแนกประเภทของสียานพาหนะ.....	78
5.4 สรุปผล.....	83
6 การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทแบรนด์ของยานพาหนะ.....	84
6.1 บทนำ.....	84
6.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set).....	84
6.3 การวัด และ ประเมินผลการจำแนกประเภทของแบรนด์ของยานพาหนะ.....	88
6.4 สรุปผล.....	92
7 สรุปและข้อเสนอแนะ.....	94
7.1 สรุปผลการทดลอง.....	94
7.2 ข้อค้นพบหลักและสิ่งที่ค้นพบใหม่.....	95
7.3 ปัญหาและอุปสรรค.....	96
7.4 ข้อจำกัดของงานวิจัย.....	97
7.5 ลักษณะของการนำไปใช้ประโยชน์.....	97
7.6 งานในอนาคต.....	97
บรรณานุกรม.....	98
ภาคผนวก.....	108
ภาคผนวก ก ขั้นตอนการทำงานของ Algorithm Fram Extraction.....	108
ภาคผนวก ข ตัวอย่างรูปภาพที่ใช้โมเดลในการจำแนกทั้ง 3 โมเดล.....	110
ประวัติผู้วิจัย.....	116

รายการตาราง

ตารางที่	หน้า
2.1 งานวิจัยที่ผ่านมา.....	49
2.2 เทคนิคการใช้งานสำหรับการจำแนกยานพาหนะ.....	53
3.1 ตารางการแยกเฟรมวิดีโอ.....	60
3.2 จำนวนภาพหลังจากใช้กระบวนการเสริมภาพ.....	62
4.1 จำนวนชุดข้อมูลภาพประเภทของยานพาหนะที่ใช้ในงานวิจัย.....	65
4.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ.....	68
4.3 ผลลัพธ์ที่ได้ตรวจจับถูกรหัสหรือผิดสำหรับการจำแนกประเภทของยานพาหนะ.....	69
4.4 การประเมินประสิทธิภาพสำหรับการพัฒนาการจำแนกประเภทยานพาหนะ.....	70
5.1 จำนวนชุดข้อมูลภาพประเภทสีของยานพาหนะที่ใช้ในงานวิจัย.....	74
5.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ.....	79
5.3 ผลลัพธ์ที่ได้ตรวจจับถูกรหัสหรือผิดสำหรับการจำแนกประเภทสีของยานพาหนะ.....	80
5.4 การประเมินประสิทธิภาพการพัฒนาโมเดลการจำแนกสียานพาหนะ.....	81
6.1 จำนวนชุดข้อมูลภาพแบรนด์ของยานพาหนะที่ใช้ในงานวิจัย.....	85
6.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ.....	89
6.3 ผลลัพธ์ที่ได้ตรวจจับถูกรหัสหรือผิดสำหรับการจำแนกแบรนด์ของยานพาหนะ.....	91
6.4 การประเมินประสิทธิภาพการพัฒนาโมเดลการจำแนกแบรนด์ของยานพาหนะ.....	92

รายการภาพประกอบ

ภาพประกอบที่	หน้า
1.1 แสดงการตรวจจับประเภทยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ สังวัตนาการ.....	3
1.2 กรอบงานวิจัย.....	7
2.1 โครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการ.....	11
2.2 คอนโวลูชัน (Convolution layer).....	12
2.3 การก้าว และ การ Padding.....	12
2.4 Pooling layer.....	14
2.5 เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์.....	14
2.6 การใช้การปรับปรุงภาพ.....	18
2.7 การปรับคอนทราสต์.....	18
2.8 การกรอกภาพ.....	19
2.9 การปิดอัดภาพ.....	19
2.10 ตัวอย่างรูปภาพการจำแนกประเภทของรูปภาพ.....	20
2.11 ตัวอย่างรูปภาพการตรวจจับวัตถุ.....	21
2.12 รายละเอียดการทำงานของ RCNN.....	22
2.13 สถาปัตยกรรมของ Fast R-CNN.....	23
2.14 ROI Pooling.....	23
2.15 สถาปัตยกรรม Faster R-CNN.....	24
2.16 โครงข่าย Region Proposal Network.....	25
2.17 การถ่ายโอนการเรียนรู้.....	26
2.18 S x S Residual blocks.....	28
2.19 Bounding boxes.....	28
2.20 ไอโอยู = พื้นที่ Intersection หารด้วยพื้นที่ทั้งหมด.....	29
2.21 สถาปัตยกรรม YOLO.....	30
2.22 สถาปัตยกรรม YOLO V2.....	31
2.23 โครงสร้างของ DarkNet-19.....	32
2.24 สถาปัตยกรรม YOLO V3.....	33
2.25 โครงสร้างของ Feature-Pyramid.....	33
2.26 โครงสร้างของ DarkNet-53.....	34
2.27 สถาปัตยกรรม YOLO V4.....	35
2.28 Object Detection Workflow.....	36
2.29 Bag of freebies.....	37
2.30 กระบวนการเพิ่มข้อมูล.....	38
2.31 ตัวอย่างภาพประกอบที่ ทำกระบวนการ data augmentation.....	39

รายการภาพประกอบ (ต่อ)

ภาพประกอบที่	หน้า
2.32 Bag of Specials.....	39
2.33 DenseNet.....	40
2.34 สถาปัตยกรรม DenseNet แบบ end-to-end.....	41
2.35 โครงสร้างของ Spatial Pyramid Pooling.....	41
2.36 backbone ของ FPN.....	42
2.37 สถาปัตยกรรม YOLO V5.....	42
2.38 EfficientRep Backbone และ Rep-PAN.....	43
2.39 แผนภาพโครงสร้างหัวแยกที่มีประสิทธิภาพ segmentation.....	44
2.40 E-ELAN (Extended Efficient Layer Aggregation Network).....	45
2.41 เมทริกซ์ความสับสน (Confusion Matrix).....	54
2.42 ค่าความมั่นใจไอโอยู.....	56
2.43 เกณฑ์ไอโอยู.....	56
3.1 แผนที่แสดงที่ตั้งกล้อง.....	59
3.2 ตัวอย่างภาพจากกล้องวงจรปิด.....	59
3.3 ตัวอย่างภาพการเพิ่มปริมาณข้อมูล.....	61
3.4 ตัวอย่างภาพการติดป้ายกำกับคลาส.....	63
3.5 ตัวประมวลผลผลลัพธ์ข้อมูลที่ติดป้ายกำกับโดยใช้ Roboflow.....	62
3.6 ตัวอย่างภาพไฟล์เก็บชุดข้อมูลที่ใช้ในการพัฒนาโมเดลจาก Roboflow.....	63
4.1 ตัวอย่างภาพประกอบประเภทยานพาหนะ.....	66
4.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทยานพาหนะระหว่างข้อมูลชุด Train และ Test.....	68
4.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทของยานพาหนะ.....	70
4.4 กราฟ loss ชุดตรวจสอบความถูกต้องของโมเดลการจำแนกประเภทยานพาหนะ.....	71
5.1 ภาพตัวอย่างประเภทสีของยานพาหนะ.....	77
5.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทสีของยานพาหนะระหว่างข้อมูลชุด Train และ Test.....	79
5.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทสีของยานพาหนะ.....	81
5.4 กราฟ loss ชุดตรวจสอบความถูกต้องของโมเดลการจำแนกประเภทสีของยานพาหนะ.....	82

รายการภาพประกอบ (ต่อ)

ภาพประกอบที่	หน้า
6.1 ภาพตัวอย่างแบรนต์ของยานพาหนะ.....	87
6.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทแบรนต์ของยานพาหนะระหว่าง ข้อมูลชุด Train และTest.....	89
6.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทแบรนต์ ของยานพาหนะ.....	90
6.4 กราฟ loss ชุดตรวจสอบความถูกต้องของโมเดลการจำแนกของยานพาหนะ.....	92

บทที่ 1

บทนำ

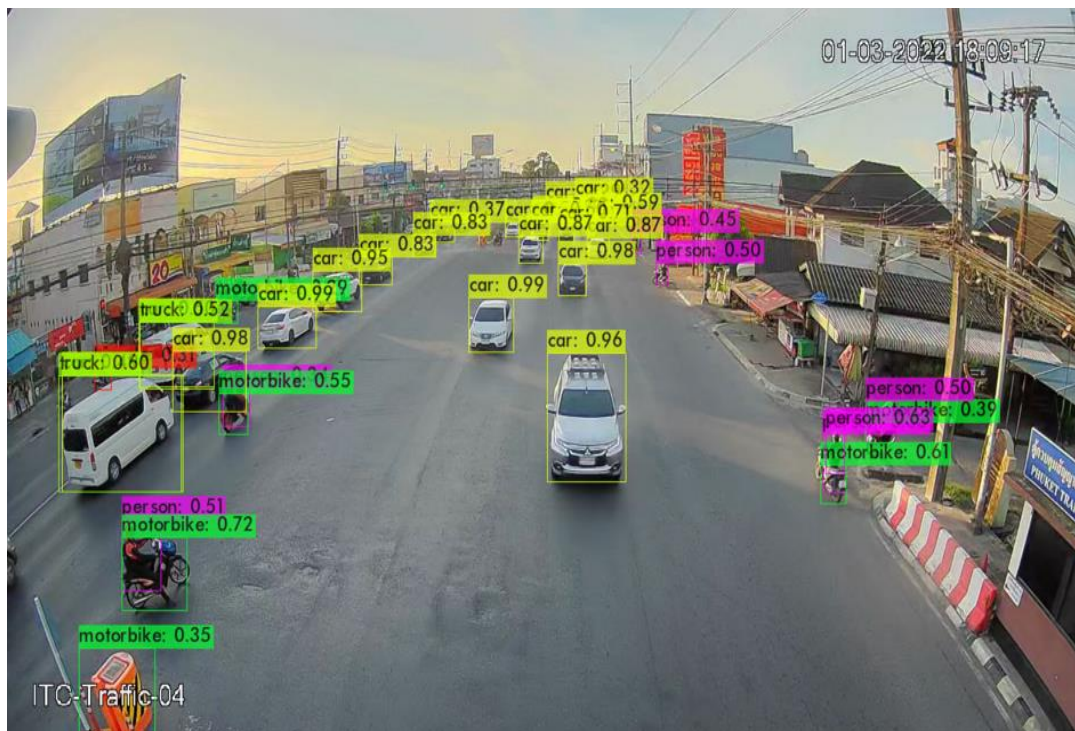
1.1 ที่มาและความสำคัญของงานวิจัย

การละเมิดกฎจราจร หรือการเกิดโจรกรรมเป็นปัญหาที่ไม่สามารถแก้ไขได้อย่างถาวรแต่สามารถเตรียมการรับมือได้ด้วยการเข้ามาของหน่วยงานราชการที่ช่วยในการเฝ้าระวังภัยการเกิดโจรกรรมหรือการละเมิดกฎจราจรแต่ในปัจจุบันมีจำนวนพาหนะ และจำนวนประชากรในแหล่งชุมชนที่เพิ่มขึ้นทุกปี ทำให้หน่วยราชการไม่สามารถเฝ้าระวังภัยได้อย่างครอบคลุมทุกพื้นที่ ดังนั้นจึงมีการนำเทคโนโลยีการตรวจจับวัตถุมาช่วยงานซึ่งการตรวจจับวัตถุมีการพัฒนาอย่างรวดเร็วในปัจจุบันทุกหน่วยงานได้นำการตรวจจับวัตถุนำไปประยุกต์เข้ากับหน่วยงานต่าง ๆ เพื่อประหยัดเวลา และ ลดภาระการทำงานของคนโดยเฉพาะในหน่วยงานทางจราจร ที่นำการตรวจจับยานพาหนะมาใช้ซึ่งเป็นหนึ่งในกระบวนการที่มีปัจจัยสำคัญสำหรับการเฝ้าระวังการจราจรด้วยจำนวนประชากรจำนวนมาก เนื่องจากประเทศต่าง ๆ และ รัฐบาลต้องการให้ความน่าเชื่อถือแก่ประชาชน และเป็นธรรมทุกวันนี้อย่างไรก็ตามประชาชนในประเทศไทยส่วนใหญ่มียานพาหนะเป็นของตนเอง เช่น รถยนต์หรือรถจักรยานยนต์โดยที่จำนวนประชากร และ ยานพาหนะบนท้องถนนที่เพิ่มขึ้นก่อให้เกิดปัญหามากมาย เช่น การละเมิดกฎจราจร ความปลอดภัยสาธารณะ อุบัติเหตุบนท้องถนน และ การโจรกรรมยานพาหนะที่เพิ่มขึ้น สำหรับสถิติการแจ้งยานพาหนะหายทางวิทยุข่าวสาร และ การจราจรได้รายงานไว้ระหว่างเดือนกรกฎาคมถึงธันวาคม 2563 มียานพาหนะหายรวม 200 คัน และ ติดตามได้ 24 คัน จากรายงานสรุปได้ว่าการสูญเสียส่วนใหญ่เป็นยานพาหนะจักรยานยนต์ และ ตามมาด้วยยานพาหนะ และ ยานพาหนะประเภทอื่น ๆ (จส.100, 2021) อย่างไรก็ตามในการหาข้อมูลยานพาหนะหายทางตำรวจจราจร หรือเจ้าหน้าที่ของรัฐที่รับผิดชอบมักจะค้นหา และ สอบถามข้อมูลจากพยานที่เห็นเหตุการณ์ แต่อาจทำให้การตรวจสอบข้อมูลในการจับกุมมีความล่าช้า และ รายละเอียดของยานพาหนะที่ถูกโจรกรรมไม่ถูกต้อง และ เพิ่มภาระงานของตำรวจจราจรและภาครัฐในการหาข้อมูล

จากรายงานขององค์การอนามัยโลก (World Health Organization [WHO], 2018) ในปี 2561 ประเทศไทยเป็นประเทศที่มีอัตราการเกิดอุบัติเหตุสูงเป็นอันดับ 9 ของโลก เสียชีวิตประมาณ 22,491 คนต่อปี หรือ 32.7 ต่อประชากร 100,000 คน โดยเฉพาะในช่วงเทศกาลต่างๆ เช่น ปีใหม่หรือ เทศกาลสงกรานต์ ประเทศไทยยังคงน่าเป็นห่วงเนื่องจากสถิติอุบัติเหตุยังคงสูงเกินค่าเฉลี่ยทั่วโลกถึง 2 เท่า สำนักงานนโยบาย และ แผนการขนส่ง และ จราจร กระทรวงคมนาคมพบสาเหตุหลักของการเกิดอุบัติเหตุ 3 ประการ ได้แก่ (i) การชนกันโดยทั่วไป (ii) เกิดจากบุคคล เช่น ขับยานพาหนะด้วยความเร็วหรือขับยานพาหนะโดยผิดกฎหมายจราจร (iii) เกิดจากอุปกรณ์ เช่น เบรกขัดข้อง ปัจจุบันเกือบทุกประเทศมีการติดตั้งกล้องวงจรปิด (CCTV) เพื่อติดตามการจราจรในรูปแบบวิดีโอ วิดีโอที่บันทึกไว้ในข้อมูลตามเวลาจริงของเหตุการณ์

ปัจจุบันทุกประเทศมีการติดตั้งกล้องวงจรปิด เพื่อติดตามการจราจรในรูปแบบวิดีโอ ข้อมูลวิดีโอเป็นข้อมูลเรียลไทม์ของเหตุการณ์ กล้องวงจรปิดถูกใช้งานอย่างแพร่หลายในองค์กรต่างๆ ในบริบทของถนนก็มีการนำกล้องวงจรปิด มาใช้ โดยเฉพาะของกรมทางหลวง และ สำนักงานตำรวจ กล้องวงจรปิดที่ติดตั้งส่วนใหญ่จะถูกตรวจสอบโดยระบบกลางที่คอยตรวจสอบ และ บันทึกข้อมูล วิดีโอหรือภาพตลอดเวลา วัตถุประสงค์หลักของระบบนี้คือการเฝ้าระวังความปลอดภัยของบุคคล และ สถานที่ซึ่งมีข้อดีสำหรับผู้ใช้งานพาหนะใช้ถนนดังนี้ (i) ยานพาหนะวิ่งด้วยความเร็วสูงเกินกำหนด เมื่อกล้องจับภาพ ระบบจะทำงานควบคุมที่ศูนย์กลางเพื่อตรวจสอบยานพาหนะคันนั้น เพื่อไม่ให้เกิดอันตรายต่อผู้ใช้งานพาหนะใช้ถนน (ii) ข้อดีของกล้องวงจรปิด แม้ว่ายานพาหนะส่วนใหญ่จะติดตั้งกล้องติดยานพาหนะไว้ แต่ยานพาหนะที่เกิดอุบัติเหตุบางคันอาจไม่ได้ติดตั้งกล้องชนิดนี้ไว้ ตำรวจจึงต้องอาศัยตัวช่วยอย่างกล้องวงจรปิดเพื่อหาหลักฐานของยานพาหนะเป้าหมาย อีกทั้งยังมีการเพิ่มหรือลดความสว่างของไฟหน้ายานพาหนะที่ส่องมาที่กล้องเพื่อให้จับภาพป้ายทะเบียนได้ง่ายขึ้นและยังปรับปรุงภาพจากกล้องให้คงความคมชัดมากขึ้นเพื่อประโยชน์ของผู้ใช้ โดยใช้เทคโนโลยีใหม่ในการจับภาพป้ายทะเบียนให้ชัดเจน แมื่อยานพาหนะจะวิ่งด้วยความเร็วสูงก็ยังสามารถใช้งานได้ในทุกสภาพแสงและทุกสภาพอากาศ

งานวิจัยนี้ได้นำเสนอการแก้ไขปัญหาเพื่อตรวจจับยานพาหนะที่ต้องการค้นหาหรือระบุบุรุษลักษณะของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ โดยคุณลักษณะที่จะใช้ตรวจจับมีสามประเภทได้แก่ (i) ประเภทของยานพาหนะ (ii) สีของยานพาหนะ และ (iii) แบรินด์ของยานพาหนะโดยจุดมุ่งหมายของงานมุ่งไปที่การวิเคราะห์ ใช้ชุดข้อมูลบนทางจราจรเพื่อใกล้เคียงยานพาหนะในชุมชนมากที่สุด เพื่อพัฒนา โครงข่ายประสาทเทียมแบบสังวัตนาการในการระบุลักษณะทางกายภาพของยานพาหนะทั้ง 3 ประเภท งานวิจัยนี้เบื้องต้นหวังว่าจะช่วยต่อยอดในการพัฒนาตัวตรวจจับยานพาหนะบนท้องถนนได้ในอนาคตซึ่งผลสำเร็จจะมีประโยชน์ได้ดังนี้ (i) การตรวจจับยานพาหนะมีวัตถุประสงค์เพื่อให้ข้อมูลช่วยในการตรวจจับยานพาหนะ (ii) เพื่อให้เจ้าหน้าที่จราจรหรือผู้ขับขี่ทราบเมื่อพบยานพาหนะที่ไม่ปลอดภัย และ (ii) อัตราการเกิดอุบัติเหตุยานพาหนะ และการโจรกรรมยานยนต์ อัตราอาชญากรรมลดลงได้ จากภาพประกอบที่ 1.1 แสดงการจำแนกประเภทยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ



ภาพประกอบที่ 1.1 แสดงการตรวจจับประเภทยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

การตรวจจับวัตถุเป็นเทคนิคการมองเห็นด้วยคอมพิวเตอร์สำหรับระบุตำแหน่งของวัตถุในภาพหรือวิดีโอ โดยทั่วไปแล้วอัลกอริทึมการตรวจจับวัตถุจะใช้ประโยชน์จากการเรียนรู้ของเครื่องหรือการเรียนรู้เชิงลึกเพื่อสร้างผลลัพธ์ที่มีความหมาย

โครงข่ายประสาทเทียม (Artificial Neural Networks หรือ ANN) (Grossi and Buscema, 2007) คือกระบวนการการสร้างคอมพิวเตอร์ที่จำลองการทำงานของสมองหรือโครงข่ายประสาทของคน เป็นการที่คนสร้างอัลกอริทึมให้คอมพิวเตอร์ให้เกิดกระบวนการ คิด วิเคราะห์ ในการทำงานในรูปแบบต่าง ๆ ตามที่คนต้องการที่จะสอนให้คอมพิวเตอร์ได้เรียนรู้ในการทำงาน โดยโครงข่ายประสาทเทียมมีจุดส่งกระแสประสาท และ จุดรับประสาทซึ่งมีหน้าที่รับข้อมูลที่ได้นำเข้ามาเรียนรู้ในโครงข่ายประสาทในสมองเกิดการเรียนรู้ในโครงข่ายเพื่อนำมาตอบซึ่งก็คือผลลัพธ์แต่ถ้าผลลัพธ์ไม่ใช่ก็นำคำตอบนั้นกลับมาเรียนรู้ในโครงข่ายประสาทใหม่เพื่อให้ได้คำตอบที่ถูกต้องซึ่งจะอธิบายในภายหลัง การมีเซลล์ประสาทนิวรอน (Neurons) และ จุดประสานประสาท (Synapses) ซึ่งเป็นเซลล์ประสาททำหน้าที่รับกระแสประสาท เดนไดรต์ (Dendrite) ที่เป็นตัวรับข้อมูล (Input) ตัวรับข้อมูลเข้ามาเพื่อนำมาวิเคราะห์ แอกซอน (Axon) ทำงานเป็นการตัวประมวลผลผลลัพธ์ของเซลล์เหมือนส่งไปยังสมองว่าวัตถุนั้นคืออะไร หลักการทำงานของโครงข่ายประสาทเทียมสำหรับคอมพิวเตอร์จะประกอบไปด้วย นิวรอนที่มีการนำเข้า ชั้นซ่อน (Hidden Layer) คือชั้นที่อยู่ระหว่างกลาง มีผลต่อการเรียนรู้ของรูปโมเดลชั้นนี้การทำงานเหมือนการเรียนรู้ของตัวประสาท และ ตัวประมวลผลผลลัพธ์ (Output) ที่จะเป็นการนำผลลัพธ์ออกมาว่าสิ่งที่เห็นคืออะไร โดยมีการฝึกที่

จะต้องมีคือค่าน้ำหนัก (Weight) เป็นตัวกำหนดค่าน้ำหนักในการฝึกเพื่อฝึกให้คอมพิวเตอร์เกิดเรียนรู้ว่าวัตถุที่เห็นคืออะไร ซึ่งจะทำให้การเรียนรู้ช้า ๆ ซึ่งเรียนรู้หลายรอบยิ่งฉลาดนั่นเอง โดยจะมีค่า เกณฑ์ (Threshold) เป็นตัวกำหนดค่าน้ำหนักของนำเข้า และ ตัวประมวลผลผลลัพธ์ให้ทดสอบหาผลลัพธ์

การทำงานของโครงข่ายประสาทเทียมมีตัวรับข้อมูลมาในโครงข่ายประสาทเทียม เพื่อนำตัวรับข้อมูลมาเรียนรู้ด้วยเทคนิคทางคณิตศาสตร์แล้วเอาผลลัพธ์ที่ได้จากตัวรับข้อมูลที่เกิดการเรียนรู้ นำมาเทียบกับเกณฑ์ ที่กำหนดไว้ผลรวมที่มีค่าเกินกว่าค่าเกณฑ์จะส่งค่าตัวประมวลผลผลลัพธ์ไปยังตัวรับข้อมูลตัวอื่นที่เชื่อมกันในโครงข่ายประสาทเทียมถ้าค่าต่ำกว่าเกณฑ์ก็จะไม่เกิดการส่งออกประมวลผลผลลัพธ์ออกมาแต่นำค่านั้นมาคำนวณใหม่เพื่อให้ได้ผลลัพธ์ที่ผ่านค่าเกณฑ์ออกไป ในโครงข่ายประสาทเทียมนั้นจะมีกระบวนการที่เรียกว่า โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back propagation) เป็นกระบวนการนำค่าที่ไม่ผ่านเกณฑ์มาปรับปรุงเพื่อที่จะได้ค่าใหม่ในแต่ละรอบ เพื่อให้ได้ค่าตัวรับข้อมูล ที่คาดหวังไว้ การเรียนรู้ของเครื่อง (Machine learning) เริ่มต้นมาตั้งแต่ปี พ.ศ. 2493 ได้มีนักวิทยาศาสตร์ท่านหนึ่งชื่อว่า Arthur Lee Samuel ได้คิดค้นอัลกอริทึมที่ทำให้คอมพิวเตอร์คิดหาวิธีสอนคอมพิวเตอร์ให้เล่นหมากรุก และ การคิดกระบวนการคำนวณเลข ทำให้คอมพิวเตอร์สามารถเข้าใจ คาดการณ์ จดจำรูปแบบแผนในการเล่นได้ และ ยังสามารถเรียนรู้ความผิดพลาดเพื่อนำมาแก้ไขปัญหาคือได้ด้วยปัจจุบันจึงนำการเรียนรู้ของเครื่องมาประยุกต์ในการทำงานต่าง ๆ เช่น การคาดการณ์สถานการณ์ทางการเกษตร, การกีฬา, การเรียน รวมไปถึงการแก้ปัญหาด้วยตัวเองของคอมพิวเตอร์อัลกอริทึมของการเรียนรู้ของเครื่องที่ต้องมีถ้าขาดตัวนี้ไปก็ไม่สามารถฝึกมันได้ ก็คือ การให้ชุดการฝึก (Training set) ของข้อมูลเพื่อให้คอมพิวเตอร์สามารถหาคำตอบได้จากเฉลย ซึ่งเฉลยที่ใช้ชุดการฝึกในการฝึกการเรียนรู้ของเครื่อง แล้วให้ใช้ข้อมูลดังกล่าวเพื่อตอบคำถามที่ตั้งขึ้น โดยใช้ชุดทดสอบ (Test set) เช่น การตอบคำถามว่า หมาคือหมา แมวคือแมว คนคือคน ยานพาหนะคือยานพาหนะ ยังมีชุดการฝึกมากก็ยังมีประสิทธิภาพมากในการเรียนรู้ของเครื่องต่อมามีการพัฒนาเป็นการเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep Learning) คือ มีการทำงานที่เหมือนกับโครงข่ายประสาท ในสมองของคนซึ่งเป็นการเรียนรู้เชิงลึกเป็นส่วนหนึ่งเล็ก ๆ ของการเรียนรู้ของเครื่องซึ่งการเรียนรู้เชิงลึกเป็นโครงข่ายประสาทเทียมที่สร้างขึ้นใหม่ซึ่งจะใช้ชั้นเลเยอร์ (Layers) ที่ซับซ้อนมากขึ้นเพื่อเรียนรู้ข้อมูลความซับซ้อนของตัวโมเดล (Model) เป็นการพัฒนาที่ทำให้คอมพิวเตอร์ได้มีความซับซ้อนคิดหลายกระบวนการเพื่อหาผลลัพธ์ที่มีความถูกต้องมากขึ้น ช่วงการเรียนรู้จะทำผ่านโครงข่ายประสาทเทียมที่เป็นสถาปัตยกรรมมีชั้นเลเยอร์เรียงทับซ้อน ๆ ตัวรับข้อมูลเป็นตัวแรกที่ตัวรับข้อมูลเข้าไปในนิรอรอน และ ถูกคำนวณด้วยคณิตศาสตร์ด้วยการคูณด้วยค่าน้ำหนักผลลัพธ์จากผลคูณหลังจากนั้นก็ยังมีผลลัพธ์ไปยังชั้นเลเยอร์ต่อไปซึ่งก็คือ กระบวนการนี้ทำได้ซ้ำในแต่ละชั้นเลเยอร์ในโครงข่ายนี้ชั้นเลเยอร์สุดท้ายถูกเรียกว่า ชั้นเลเยอร์ตัวประมวลผลผลลัพธ์ที่เป็นผลลัพธ์ว่าจะมีความถูกต้องแม่นยำมากกว่าแค่ไหน มันจะคอยช่วยสนับสนุนค่าจริงที่กำหนดไว้ และ ความน่าจะเป็นของแต่ละประเภทของงานแบ่งประเภท (Classification) โครงข่ายประสาทเทียมใช้อัลกอริทึมทางคณิตศาสตร์เพื่อปรับปรุงค่าน้ำหนักของข้อมูลของทุกโครงข่ายประสาทเทียมที่ถูกฝึกอบรมมาอย่างเต็มที่แล้วเมื่อค่าของค่าน้ำหนักในการฝึกอบรมจะให้ค่าผลลัพธ์เข้าใกล้กับค่าจริงที่กำหนดไว้ในชุด

ข้อมูลในการฝึกอบรม กระบวนการเรียนรู้ของเครื่องที่เป็นส่วนสำคัญจะมี การสกัดคุณลักษณะ (Feature extraction) การสกัดคุณลักษณะประกอบไปด้วยคุณลักษณะ (Feature) ทำหน้าที่เพื่อ สกัดคุณสมบัติคุณลักษณะวัตถุในรูปภาพหาจุดเด่นนั้นในรูปภาพที่ให้คอมพิวเตอร์ตรวจจับมัน เช่น การสกัดคุณลักษณะ จมูก, ตา, และ ปาก ของรูปภาพคน ที่ถูกเลือกมาแล้วจะส่งต่อโมเดลการแบ่ง ประเภท (Classification model) ซึ่งการเรียนรู้เชิงลึก และ โครงข่ายประสาทเทียมแบบสังวัตนาการ จะอธิบายต่อไปในบทที่ 2

การเรียนรู้เชิงลึกนั้นโดดเด่นในหลากหลายด้านในปัจจุบันซึ่งมีการพัฒนาอย่างมาก และมีโครงข่ายหนึ่งเรียกว่า โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolution neural network หรือ CNN) ซึ่งโครงข่ายประสาทเทียมแบบสังวัตนาการจะขยายความในหัวข้อการทบทวนวรรณกรรม และ ทฤษฎีที่เกี่ยวข้อง

ดังนั้นงานวิจัยนี้จึงเป็นการพัฒนาสถาปัตยกรรมโครงข่ายประสาทเทียมแบบสังวัตนาการรวมที่จะสามารถระบุลักษณะทางกายของยานพาหนะได้ดังนั้นผู้วิจัยจึงจะเสนอใช้ร่วมกับการ ถ่ายโอนความรู้เชิงลึกกับชุดข้อมูล และ ในการ พัฒนาตัวโมเดลให้สามารถตรวจจับยานพาหนะที่สามารถระบุลักษณะทางกายภาพได้โดยแยกเป็น 3 โมเดล คือ (i) ประเภทของยานพาหนะ (ii) สีของยานพาหนะ และ (iii) แบรินด์ของยานพาหนะ โดยผู้วิจัยจะติดป้ายกำกับ กับ ชุดข้อมูลให้เหมาะสมกับชุดข้อมูล 3 โมเดล จากนั้นใช้การถ่ายโอนการเรียนรู้กับตัวโมเดลจำแนกเพื่อพัฒนาในจำแนกการตรวจจับโดยระบุลักษณะทางกายภาพของยานพาหนะ และพัฒนาชุดข้อมูลเพื่อเพิ่มความแม่นยำในจำแนกการตรวจจับ ซึ่งประโยชน์ที่ได้รับจากงานวิจัยนี้ คือ ตัวโมเดลจำแนกจะสามารถระบุลักษณะทางกายภาพของตัวยานพาหนะ และเพื่อพัฒนาต่อไปในอนาคตและเพื่อการศึกษาในตัวโมเดลในแต่ละสถาปัตยกรรมว่าสถาปัตยกรรมไหนที่มีประสิทธิภาพที่ดีที่สุดในแต่ละสถาปัตยกรรม ว่าสถาปัตยกรรมไหนจะสามารถจำแนกการตรวจจับยานพาหนะที่ระบุลักษณะทางกายภาพได้ดีที่สุด

1.2 คำถามงานวิจัย

จากที่กล่าวมาข้างต้น เพื่อศึกษา วิเคราะห์ และกระบวนการที่สามารถระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียม ซึ่งแสดงเป็นคำถามงานวิจัยดังนี้

“การพัฒนาโมเดลสำหรับการจำแนกลักษณะทางกายภาพของยานพาหนะจากข้อมูลวิดีโอโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถทำได้อย่างไร”

เป้าหมายของงานวิจัยนี้แนะนำเสนอการกำหนดวิธีการระบุลักษณะทางกายภาพของยานพาหนะโดยใช้กลไกโครงข่ายประสาทเทียมแบบสังวัตนาการ คำถามการวิจัยนี้ครอบคลุมประเด็นต่างๆ ดังนี้

- 1) ขั้นตอนการเตรียมความพร้อมที่เสนอเพื่อพัฒนาตัวโมเดลการจำแนกที่เหมาะสมสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ
- 2) ขั้นตอนวิธีใดที่ใช้ในการเตรียมข้อมูลวิดีโอก่อนที่จะเริ่มกระบวนการสร้างโมเดลสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ

3) กลไกการเรียนรู้โครงข่ายประสาทเทียมแบบสังวัตนาการใด ที่เหมาะสมที่สุดสำหรับ การสร้างโมเดลการจำแนกวัตถุเพื่อระบุลักษณะทางกายภาพของยานพาหนะจากข้อมูลที่กำหนด

1.3 วัตถุประสงค์

จากหัวข้อก่อนหน้านี้ที่กล่าวถึง การวิเคราะห์ พัฒนา การกำหนดวิธีการระบุลักษณะทางกายภาพของยานพาหนะโดยใช้กลไกโครงข่ายประสาทเทียมเพื่อให้การดำเนินงานวิจัยได้อย่างถูกต้อง แสดงวัตถุประสงค์ดังนี้

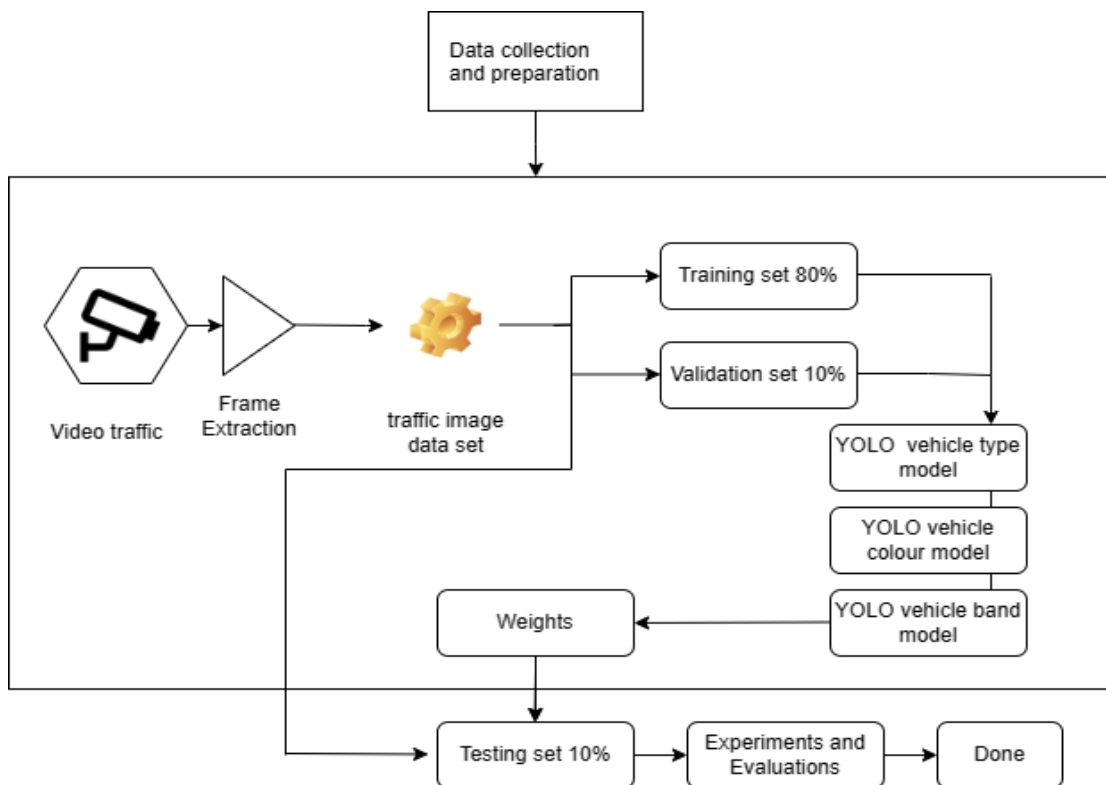
1) เพื่อตรวจสอบ และประเมินเทคนิคการประมวลผลล่วงหน้าในการประมวลผลข้อมูลวิดีโอเตรียมความพร้อมสำหรับการพัฒนาตัวติดป้ายกำกับให้กับชุดข้อมูลที่จะนำไปพัฒนาตัวโมเดล

2) เพื่อค้นหา และ ตรวจสอบกลไกโครงข่ายประสาทเทียมที่เหมาะสมต่อการพัฒนาตัวโมเดลในการจำแนกลักษณะทางกายภาพของยานพาหนะ

3) เพื่อพัฒนาโมเดลสำหรับตรวจจับวัตถุ และเปรียบเทียบประสิทธิภาพโดยใช้สถาปัตยกรรมการตรวจจับวัตถุที่มีอยู่

1.4 ระเบียบวิธีวิจัย

ระเบียบวิธีวิจัยแสดงไว้ในส่วนนี้ ในข้อเสนอการวิจัยนี้ได้เสนอกรอบการสร้างโมเดลการเรียนรู้เชิงลึกสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ เพื่อให้มีมุมมองที่ชัดเจนเกี่ยวกับข้อเสนอโครงการวิจัย จึงได้จัดทำกรอบโครงร่าง และ แสดงไว้ในภาพประกอบที่ 1.2



ภาพประกอบที่ 1.2 กรอบงานวิจัย

จากภาพประกอบที่ 1.2 แสดงกรอบงานวิจัยสำหรับการระบุลักษณะทางกายภาพของยานพาหนะจากข้อมูลวิดีโอ กรอบการทำงานนี้ประกอบด้วย (i) การรวบรวมข้อมูล (ii) การแยกเฟรมออกจากวิดีโอ (iii) การฝึกอบรมด้วยข้อมูลเพื่อตรวจจับประเภทยานพาหนะ (iv) การฝึกอบรมด้วยข้อมูลเพื่อจำแนกสียานพาหนะ และ (v) การฝึกอบรมด้วยข้อมูลเพื่อตรวจจับแบรนด์ยานพาหนะ

กระบวนการเริ่มต้นด้วยการรวบรวมข้อมูล ข้อมูลที่ใช้เกี่ยวกับงานที่นำเสนอในข้อเสนอนี้เป็นข้อมูลวิดีโอที่ได้มาจากกล้องวงจรปิดที่ติดตั้งที่ถนนสี่แยกในจังหวัดภูเก็ต เมื่อรวบรวมข้อมูลวิดีโอแล้วนำมาแยกเฟรมออกเป็นเฟรมรูปภาพตามเฟรมของชุดข้อมูลวิดีโอเพื่อนำมาสร้างชุดข้อมูลที่ใช้สำหรับการฝึกอบรมกับตัวโมเดล ตัวโมเดลมีการใช้โมเดลการคาดการณ์ที่สร้างขึ้นมาสืบค้นแบบเพื่อระบุลักษณะทางกายภาพของยานพาหนะ ได้แก่ (i) การจำแนกประเภทยานพาหนะ (ii) การจำแนกสียานพาหนะ และ (iii) การจำแนกแบรนด์ยานพาหนะ

ประเภทของยานพาหนะที่เกี่ยวข้องกับงานวิจัยนี้ ได้แก่ (i) รถยนต์ (ii) รถจักรยานยนต์ (iii) รถกระบะ และ (iv) รถโดยสาร

ประเภทของสียานพาหนะสืบสีได้แก่ (i) สีขาว, (ii) สีดำ, (iii) สีเทา/เงิน/สีบรอนซ์, (iv) สีน้ำเงิน, (v) สีแดง, (vi) สีเหลือง, (vii) สีเขียว, (viii) สีน้ำตาล, (ix) สีส้ม, และ (x) สีอื่น ๆ

ประเภทของแบรนด์ยานพาหนะที่น่าสนใจ ได้แก่ (i) โตโยต้า, (ii) ฮีซูซู, (iii) ฟอร์ด, (iv) มาสด้า, (v) บีเอ็มดับเบิลยู, (vi) มิตซูบิชิ, (vii) ฮอนด้า, (viii) ซูซูกิ, (ix) นิสสัน, และ (x) แบรนด์อื่น ๆ

1.5 ข้อค้นพบ

จากที่มาและความสำคัญ วัตถุประสงค์ และกรอบงานวิจัยนี้ สามารถค้นพบกระบวนการพัฒนาโมเดลสำหรับการจำแนกลักษณะทางกายภาพของยานพาหนะจากข้อมูลวิดีโอโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการดังนี้

- 1) กระบวนการเก็บรวบรวมข้อมูลวิดีโอการจราจรสำหรับการใช้ในการทำวิจัยจากกล้องวงจรปิดจังหวัดภูเก็ต
- 2) กระบวนการจัดการวิดีโอการจราจรออกเป็นรูปภาพเพื่อง่ายสำหรับสร้างชุดข้อมูลรูปภาพ
- 3) กระบวนการเตรียมความพร้อมของภาพการจราจรโดยการปรับแสงในวิดีโอและความคมชัดเพื่อความเหมาะสมต่อการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ
- 4) กระบวนการเพิ่มปริมาณข้อมูลภาพทางจราจรเพื่อลดระยะเวลาในการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ ดีกว่าการสร้างโมเดลขึ้นมาใหม่
- 5) การคิดพื้นที่สนใจของยานพาหนะที่สำคัญต่อการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ
- 6) กระบวนการตีขอบเขตที่ทำนายในพื้นที่สนใจ และ การตัดป้ายกำกับเพื่อแบ่งแย่งประเภทสำหรับการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบ สังวัตนาการ
- 7) โมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการชื่อว่า YOLO (You Only Look Once)
- 8) กระบวนการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้ YOLO ทั้ง 6 สถาปัตยกรรมเพื่อเปรียบเทียบคุณภาพการพัฒนาการจำแนกโดยใช้ชุดข้อมูลที่กำหนด

1.6 ขอบเขตงานวิจัย

ขอบเขตงานวิจัยนี้กล่าวถึงชุดข้อมูลที่นำมาวิจัย และ เครื่องมือในการทำงานวิจัย เพื่อตรวจจับลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

- 1) ชุดข้อมูลที่ใช้ คือ ชุดวิดีโอจากกล้องวงจรปิด ของบริษัทเอกชนจากจังหวัดภูเก็ต ตำบลเกาะแก้ว ถนนเทพกระษัตรี อำเภอเมืองภูเก็ต จังหวัดภูเก็ต
- 2) การระบุลักษณะทางกายภาพของยานพาหนะแบ่งออกเป็น 3 โมเดลได้แก่ (i) ประเภทของยานพาหนะ (ii) สีของยานพาหนะ (iii) แบรินด์ของยานพาหนะ

- 3) ภาษาที่ใช้ในการพัฒนาเป็น Python
- 4) เครื่องมือที่ใช้ในการพัฒนาได้แก่ Google Colab และ Roboflow
- 5) เทคโนโลยีที่ใช้สำหรับการเรียนรู้เชิงลึก (Deep learning) คือ Pytorch เวอร์ชัน

1.10.1 และ Tensorflow

1.7 สรุป

ในบทนี้ได้อธิบายถึงที่มา และ ความสำคัญของการดำเนินงานวิจัยนี้ โดยปัจจุบันการตรวจจับยานพาหนะเป็นสิ่งสำคัญมากมายตามสถานการณ์ต่าง ๆ เช่น ตรวจจับเพื่อจับกุม หรือค้นหาในพื้นที่ จราจร สถานที่สาธารณะ เพื่อความปลอดภัยของประชาชนโดยงานวิจัยนี้จึงเป็นการพัฒนาสถาปัตยกรรมโครงข่ายประสาทเทียมแบบสังวัตนาการรวมที่จะสามารถระบุลักษณะทางกายของยานพาหนะโดยใช้ชุดข้อมูลที่กำหนด และในบทที่ 2 จะเป็นการอธิบายถึงการทบทวนวรรณกรรม และทฤษฎีที่เกี่ยวข้องที่ใช้ในงานวิจัยนี้

บทที่ 2

การทบทวนวรรณกรรม และทฤษฎีที่เกี่ยวข้อง

2.1 บทนำ

การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้องในงานวิจัยนี้ จะอธิบายถึงสิ่งต่าง ๆ ที่เกี่ยวข้องกับงานวิจัยซึ่งประกอบไปด้วยอธิบายสถาปัตยกรรมโครงข่ายประสาทเทียมแบบสังวัตนาการ และการเรียนรู้เชิงลึกที่ใช้ในงานตรวจจับวัตถุ อธิบายการประมวลผลภาพ การตรวจจับวัตถุ และแสดงตารางเปรียบเทียบงานวิจัยก่อนหน้านี้ที่เกี่ยวข้องกับการตรวจจับวัตถุโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการที่ใช้ในการวิจัย รวมถึงวิธีการการประมวลผลภาพ (Image processing) ซึ่งเป็นวิธีการพื้นฐานเป็นการใช้เทคโนโลยีดิจิทัลการประมวลผลภาพด้วยคอมพิวเตอร์เพื่อให้ได้ภาพประกอบที่ละเอียดคมชัดเพื่อให้ชุดข้อมูลที่มีคุณภาพมากขึ้น สามารถนำมาใช้งานในเรื่องอื่น ๆ ได้ และ ช่วยเพิ่มคุณภาพของภาพเพื่อช่วยในการฝึกอบรมชุดข้อมูล และ สุดท้ายอธิบายถึงวิธีการวัดและ ประเมินผลเพื่อหาวิธีที่เหมาะสมต่อในการใช้ตัวโมเดล

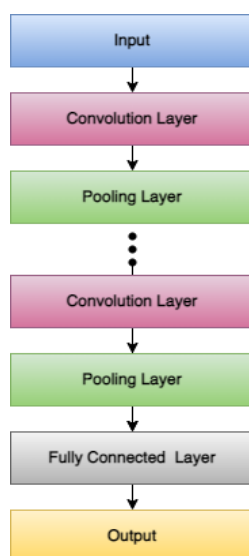
หัวข้อที่จะกล่าวถึงในบทนี้ มีดังนี้ 2.2 โครงข่ายประสาทเทียมแบบสังวัตนาการ เพื่ออธิบายรายละเอียดกระบวนการที่อยู่ในโครงข่าย และ การเรียนรู้เชิงลึก 2.3 การประมวลผลภาพ อธิบายกระบวนการทำงานของการประมวลผลภาพ 2.4 การประมวลผลภาพเพื่ออธิบายกระบวนการประมวลผลภาพ 2.5 การตรวจจับวัตถุ 2.6 การถ่ายโอนการเรียนรู้ (Transfer learning) 2.7 สถาปัตยกรรม You Only Look Once (YOLO) YOLO, YOLO V2, YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7, และ YOLO V8 ที่ใช้ในการวิจัยการจำแนกการตรวจจับยานพาหนะ 2.8 งานวิจัยที่เกี่ยวข้องแสดงตารางการเปรียบเทียบของงานวิจัยที่เกี่ยวข้อง 2.9 การตรวจวัดประสิทธิภาพ และ 2.10 สรุป ในหัวข้อสุดท้ายของบทที่ 2

2.2 โครงข่ายประสาทเทียมแบบสังวัตนาการ และ การเรียนรู้เชิงลึก

วัตถุประสงค์ของการวิจัยนี้มีเป้าหมายเพื่อสร้างโมเดลการจำแนกการตรวจจับโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อระบุลักษณะทางกายภาพของยานพาหนะ เช่น ประเภท สี และ แบรินด์ของยานพาหนะ โครงข่ายประสาทเทียมแบบสังวัตนาการเป็นส่วนเล็กของการเรียนรู้ของเครื่อง การเรียนรู้ของเครื่องเป็นแขนงหนึ่งของปัญญาประดิษฐ์ที่คนเป็นคนออกแบบ และ ใช้ อัลกอริทึมในการสอนคอมพิวเตอร์ในการเรียนรู้ คิด วิเคราะห์ แยกแยะ สิ่งต่าง ๆ ที่ต้องการในช่วยทำงานในหน่วยงานต่าง ๆ และ ยังพัฒนาต่อไปเพื่อให้มีประสิทธิภาพมากขึ้น ในการทำงานของเครื่องขึ้นอยู่กับข้อมูลที่นำเข้าไปฝึกอบรมในตัวเครื่องยังมีข้อมูลเยอะยังมีประสิทธิภาพมากขึ้นเท่านั้น

การเรียนรู้เชิงลึก (Deep learning) คือ อย่างที่กล่าวไปในบทที่ 1 การเรียนรู้เชิงลึก (Fourie, 2003) เป็นการจำลองรูปแบบการประมวลผลของสมองคน โดยใช้โครงข่ายคล้ายเซลล์ประสาทคน เมื่อได้รับข้อมูลมาทำการแบ่งแยกเรียนรู้ส่วนของข้อมูล และ บันทึกรายละเอียดต่าง ๆ ที่

ได้รับมาทั้งหมด แล้วนำมาประมวลผลหาคุณลักษณะเด่นในชั้นเซลล์ประสาท และ จุดเด่นของการเรียนรู้เชิงลึกคือ ข้อมูลในเชิงลึกทำงานเป็นส่วนเป็นชั้น ๆ แล้วสรุปผลข้อมูลออกมาเป็นตัวประมวลผลผลลัพธ์ และ ตรวจสอบว่าข้อมูลนั้นส่งผลตรงกับสิ่งที่เห็นหรือต้องการหรือไม่ โครงข่ายประสาทเทียมแบบสังวัตนาการหรือ โครงข่ายประสาทแบบคอนโวลูชัน (Convolution) (Alfarraj, 2020) เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม bio-inspired โดยทั่วไปแล้วโครงข่ายประสาทเทียมแบบสังวัตนาการจะใช้ในการจดจำภาพ และ ประมวลผลออกมานำข้อมูลที่รับเข้ามาแปลงรูปภาพนั้นเป็นตารางตัวเลขพิกเซลเพื่อให้คอมพิวเตอร์ง่ายต่อการคำนวณตามหลักคณิตศาสตร์ โครงข่ายประสาทเทียมแบบสังวัตนาการเป็นปัญญาประดิษฐ์ในการประมวลผลภาพที่มีประสิทธิภาพซึ่งใช้การเรียนรู้เชิงลึกในการทำงาน การใช้ลูกตาเสมือนของคอมพิวเตอร์ (Computer Vision) (Paragios and Zabih, 2008) ในการจดจำภาพ และ วิดีโอเพื่อตรวจจับวัตถุที่สนใจ และ มีการทำงานร่วมกับการประมวลผลภาพเพื่อเพิ่มประสิทธิภาพในกระบวนการทำงานของโครงข่ายประสาทเทียมแบบสังวัตนาการ จากภาพประกอบที่ 2.1

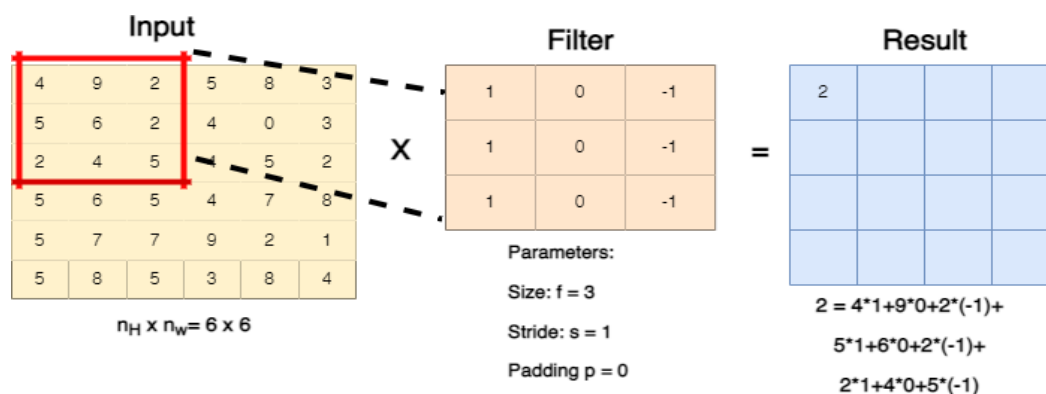


ภาพประกอบที่ 2.1 โครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการ

จากภาพประกอบที่ 2.1 คือ โครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบสังวัตนาการ จะประกอบไปด้วยชั้นทั้งหมด 3 ชั้น (Ghosh et al., 2019) ได้แก่ (i) ชั้นคอนโวลูชันเลเยอร์ (ii) ชั้นพูลลิง (Pooling layer) และ (iii) ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully connected layer)

ชั้นที่ 1 ชั้นคอนโวลูชัน ทำหน้าที่แยกคุณลักษณะเด่นจากรูปภาพหรือดึงคุณลักษณะสำคัญของรูปภาพออกมาเพื่อหาจุดเด่นของภาพ ซึ่งจะชั้นคอนโวลูชัน มีหน้าที่เก็บตัวเลขพิกเซลของรูปภาพเอาไว้ให้คงเดิมมากที่สุดเต็มที่ 0 - 255 ชั้นคอนโวลูชันทำการหาคุณลักษณะเด่นของรูปภาพออกมาเพื่อนำค่าพิกเซลที่ดึงออกมาจากรูปภาพส่งไปยังตัวกรอง (Filter) หรือ เคอร์เนล (Kernel) ทำหน้าที่ดึงคุณลักษณะที่ต้องการตรวจจับเพื่อจดจำลักษณะเด่นของวัตถุเอาไว้ไว้มากเป็นลักษณะ 2

มิติ ใช้เลื่อนหน้าต่างให้ตารางที่อยู่ศูนย์กลางวิ่งผ่านทีละจุดภาพแล้วแทนค่าจุดภาพนั้นๆ เช่น รูปภาพคนร้องไห้ เคอร์เนลจะทำการเลื่อนหน้าต่างตารางไปเรื่อย ๆ เพื่อหาหยดน้ำตาซึ่งเป็นจุดเด่นของรูปภาพด้วยตัวเลขค่าเฉลี่ยด้วยวิธีทางคณิตศาสตร์ เช่น เมทริกซ์ 3×3 ดำเนินการคูณองค์ประกอบตามสูตรคณิตศาสตร์ โดยที่จะทำการคำนวณตามตาราง และ เคลื่อนที่ไปเรื่อยตามสเกลรูปภาพตั้งภาพประกอบที่ 2.2



ภาพประกอบที่ 2.2 คอนโวลูชัน (Convolution layer)

ตัวกรองเมทริกซ์ 3×3 ที่ใช้กับรูปภาพจะถูกคำนวณสำหรับพิกเซลตัวรับข้อมูล และ คำนวณหน้าต่างตัวกรอง จากนั้นตัวกรองจะเลื่อนตามจำนวนพิกเซล n หรือที่เรียกว่า ก้าว (Stride) ที่ขยับกรอบไปยังกรอบต่อไปเพื่อคำนวณ และการคำนวณนี้ทำให้เมทริกซ์ผลลัพธ์มีขนาดเล็กลง ตัวอย่าง หากทำชั้นคอนโวลูชันหลายชั้น และยังมี Padding เป็นการกำหนดกรอบของตารางคุณลักษณะให้มีขนาดเท่าตัวเต็มส่วนใหญ่จะเติมกรอบรอบนอกที่เพิ่มมาเป็นสีดำหรือสีเทาเพื่อไม่ให้เวลาขยับกรอบขนาดตัวเลขที่สำคัญไป ช่องว่างภายในระหว่างการขยายขอบของตัวรับข้อมูลจากทุกด้านจะต้องใส่ขนาด 6×6 เพื่อทำการขยาย $p = 1$ จะทำให้ขนาดตัวรับข้อมูลออกเป็น 8×8 ดังสมการที่ 1 - 3 ใช้ในการคำนวณมิติตัวรับข้อมูล หากการกรองที่ถูกต้องคำนวณโดยสมการ 1 สมการ 1 จะแสดงการคำนวณการก้าวเมื่อการกรองที่ถูกต้อง (Uchida et al., 2018) ขนาดตัวประมวลผลผลลัพธ์คือ:

$$(n - f + 1), (n - f + 1) \quad (1)$$

โดยที่ n คือ input ตัวรับข้อมูล f คือมิติของตัวกรอง เมื่อการคำนวณครบรอบเดียวกันจะถูกเติมให้ตารางใหม่เพื่อให้มิติตัวประมวลผลผลลัพธ์เหมือนกับมิติตัวรับข้อมูล ขนาดตัวประมวลผลผลลัพธ์คือ

$$(n + 2p - f + 1), (n + 2p - f + 1) \quad (2)$$

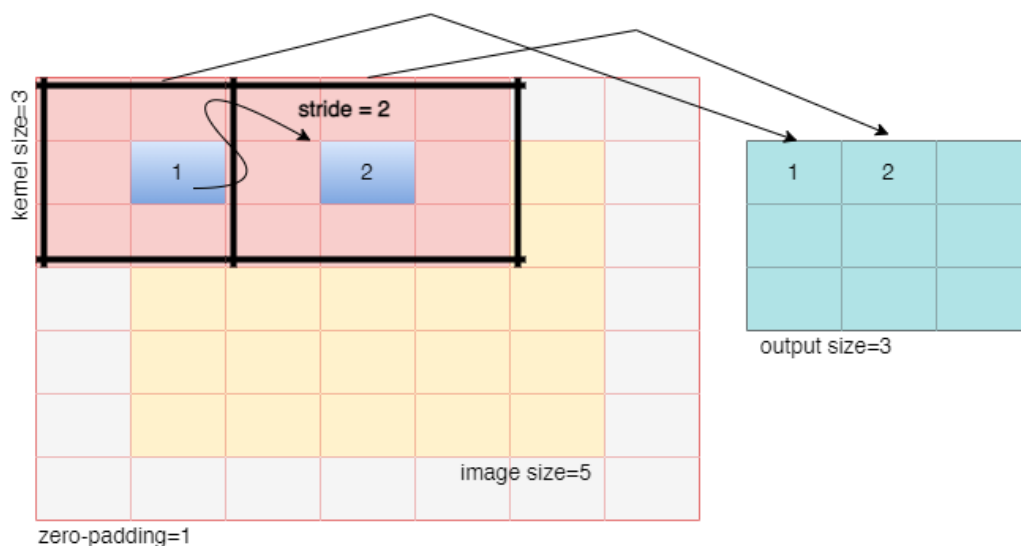
$$p = \frac{f-1}{2} \quad (3)$$

เมื่อ p คือขนาดพิกเซลของช่องว่างภายในแต่ละด้าน ตัวอย่างเช่น หากตัวกรองมีขนาด 5×5 ต้องแทนลงสมการที่ 4

$$p = \frac{5-1}{2} = 2 \text{ pixel} \quad (4)$$

ดังนั้น f ควรเป็นเลขคี่เพื่อให้ได้สมมาตร และมีจุดศูนย์กลาง การก้าว คือจำนวนของช่องที่จะย้ายกรอบการคำนวณกระบวนการในแต่ละครั้ง เหมือนกับว่าแต่ละช่องเลื่อน ก้าว $s = 1$ แต่เลื่อนออกไปสองช่องคือ $s = 2$ สูตรสำหรับคำนวณขนาดตัวประมวลผลผลลัพธ์เมื่อใช้ ก้าว ดังสมการที่ 5 ตัวอย่างการคำนวณระยะห่างจากขอบ และ ช่วงก้าวดึงภาพประกอบที่ 2.3

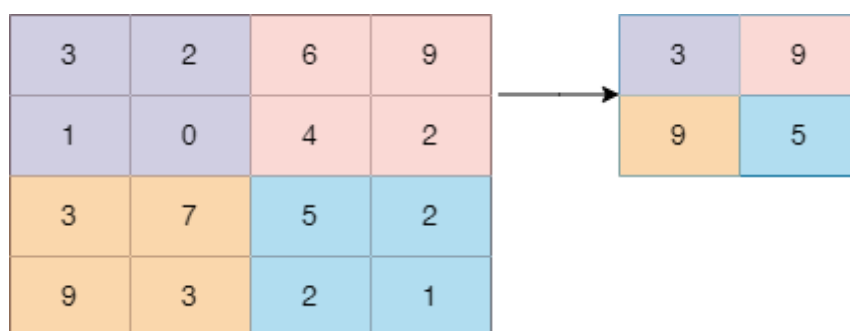
$$\left\lceil \frac{n+2p-f}{s} \right\rceil + 1, \left\lceil \frac{n+2p-f}{s} \right\rceil + 1 \quad (5)$$



ภาพประกอบที่ 2.3 การก้าว และการ Padding

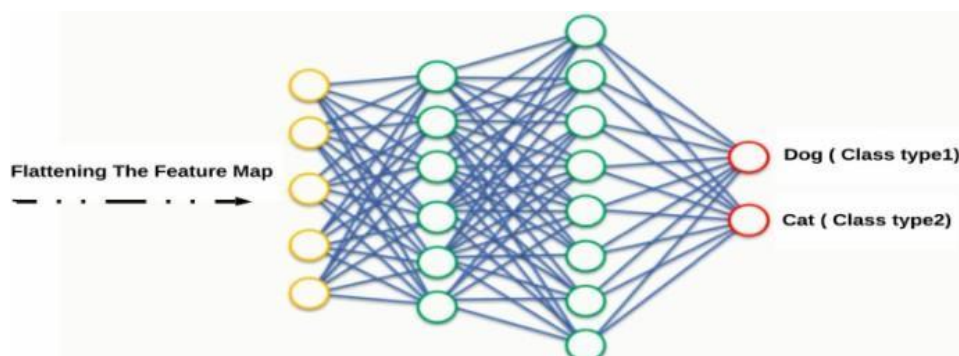
การแปลงฟังก์ชันเรกติไฟด์ไลน์เนี่ยยูนิต (Rectified Linear Unit or ReLU) (Bodianskiy et al., 2022) จะถูกนำไปใช้หลังจากการดำเนินการบิดทุกครั้งเพื่อให้แน่ใจว่าไม่เป็นเชิงเส้น ReLU เป็นฟังก์ชันการเปิดใช้งานที่นิยมใช้มากที่สุดใน การกำหนดตัวประมวลผลผลลัพธ์ของโครงข่ายประสาทเทียม เช่น ใช่ หรือ ไม่ใช่ มันจับคู่ค่าผลลัพธ์ระหว่าง 0 ถึง 1 หรือ -1 ถึง 1 เป็นต้น

ชั้นที่ 2 ชั้นพูลลิ่งเลเยอร์ (Pooling layers) (Turchenko et al., 2019) ใช้ในการสุ่มตัวอย่างแบบใช้เพื่อลดขนาดของแผนผังตารางคุณลักษณะ มีการพูลลิ่งวิธีต่าง ๆ โดยทั่วไปคือการรวมสูงสุด ขนาดของตัวกรองการรวมมักจะเป็นเมทริกซ์ 2×2 ในการรวมกลุ่มสูงสุด ตัวกรอง 2×2 จะเลื่อนไปเหนือแผนที่คุณลักษณะ และ เลือกค่าที่ใหญ่ที่สุดหรือมากที่สุดในกลุ่มที่กำหนด การดำเนินการนี้ส่งผลให้เกิดแผนผังคุณลักษณะแบบรวมค่าตัวเลขไว้ การรวมที่นิยมกันโดยทั่วไปมี 2 ประเภท: (i) การพูลลิ่งรวมค่าสูงสุด (Max pooling) และ (ii) การพูลลิ่งรวมค่าเฉลี่ย (Average pooling) การใช้ค่าสูงสุดของแต่ละกลุ่มของเซลล์ประสาทในแทนพื้นที่ในแผนผังคุณลักษณะ ตัวอย่างของการรวมค่าสูงสุดดังภาพประกอบที่ 2.4 จากภาพประกอบที่ 2.4 ชั้นการรวมที่แยกเฉพาะค่าสูงสุดของกริดเซลล์ที่จะถูกเก็บไว้ในตารางใหม่



ภาพประกอบที่ 2.4 Pooling layer

ชั้นเชื่อมต่ออย่างสมบูรณ์ Fully Connected (FC) ประกอบด้วยเลเยอร์ที่เชื่อมทุกอย่างก่อนหน้าไว้เพื่อนำผลลัพธ์ออกมาแสดงผลตัวประมวลผลผลลัพธ์ของโครงข่ายประสาทเทียมแบบสังวัตนาการดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.5 เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์ (SuperDataScience, 2018)

โครงข่ายประสาทเทียมแบบสังวัตนาการ มีอิทธิพลอย่างมากในปัจจุบันในด้านการตรวจจับวัตถุต่าง ๆ โครงข่ายประสาทเทียมแบบสังวัตนาการมีหลายหน่วยงานที่ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการเป็นพัฒนาพื้นฐานในงานตรวจจับเช่น การตรวจจับแยกสินค้าในโรงงาน (Wang et al., 2017) มีการใช้โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อตรวจจับแยกสินค้าใน

โรงงานเพื่อตรวจสอบความปลอดภัยของสินค้าเพื่อลดภาระของคนในหน่วยงาน โครงข่ายรับรูปภาพเป็นตัวรับข้อมูล และรับชุดค่าเลขที่แสดงถึงลักษณะเด่นของสินค้าหรือลักษณะในส่วนต่าง ๆ ของภาพ โครงข่ายประสาทเทียมแบบสังวัตนาการได้แสดงความแม่นยำที่ดียิ่งขึ้นเหนืออัลกอริธึมก่อนหน้านี้ โดยระบุสินค้าได้อย่างแม่นยำ

การติดตาม และวิเคราะห์พฤติกรรมของสัตว์ป่าที่วิจัยของ เวอร์มา และคุปตะ (Verma and Gupta, 2018) เสนอวิธีการติดตามสัตว์ป่าผ่านการตรวจจับสัตว์ในรูปภาพที่ติดตั้งกล้องเพื่อส่องสัตว์ป่า เพื่อจัดการกับการพบเจอสัตว์ที่พบได้ยากในการเจอกว่าที่สัตว์ป่าจะออกมา และมีการใช้ฐานข้อมูลกล้องที่มีการระบุของสัตว์ เพื่อให้สามารถดำเนินการตรวจสอบได้โดยมีการใช้โครงข่ายประสาทเทียมเชิงลึก Deep Convolutional Neural Network (DCNN) ในการตรวจจับสัตว์ป่า โดยมีความแม่นยำ 91.4% จากข้อมูลกล้อง ทั้งนี้มีการใช้ในการจดจำอารมณ์ในหน่วยงานแพทย์ เช่น วิศวกรรมชีวการแพทย์ จิตวิทยา ประสาทวิทยาศาสตร์ และการดูแลสุขภาพ โดยช่วยในการวินิจฉัยความผิดปกติทางสมองและจิตโดยทีมผู้วิจัย ออซเดเมียร์ และคณะ (Ozdemir et al., 2020) ใช้ประโยชน์จากการเรียนรู้เชิงลึก โดยเฉพาะสถาปัตยกรรม LeNet ที่ใช้ โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อเรียนรู้จดจำการแสดงออกทางสีหน้า งานวิจัยนี้ได้รวมชุดข้อมูลสามชุด (JAFFE, KDEF และชุดข้อมูลที่กำหนดเอง) และฝึกอบรมโมเดล LeNet เพื่อจำแนกสถานะทางอารมณ์จากการแสดงออกทางสีหน้า การศึกษารายงานผลลัพธ์ที่ด้วยความแม่นยำ 96.43% และความแม่นยำในการตรวจสอบที่ 91.81% สำหรับการแยกแยะอารมณ์ที่แตกต่างกันเจ็ดอารมณ์ผ่านการแสดงออกทางสีหน้า

โครงข่ายประสาทเทียมเชิงลึกสำหรับการตรวจจับการรั่วไหลของน้ำมันจากอิมเมจ SAR โดยผู้วิจัย เซง และหวัง (Zeng and Wang, 2020) ใช้ OSCNet เป็นโครงข่ายประสาทเทียมแบบโครงข่ายประสาทเทียมเชิงลึกสำหรับการตรวจจับการรั่วไหลของน้ำมันจากเรดาร์จะรับรู้แสงสังเคราะห์ (SAR) OSCNet ได้รับการออกแบบโดยใช้ VGG-16 โดยใช้ประโยชน์จากชุดข้อมูลมากกว่า 20,000 รูปและการเพิ่มข้อมูลมีประสิทธิภาพเพื่อให้ชุดข้อมูลมีความหลากหลายมากขึ้นในการตรวจจับ โดยปรับปรุงความแม่นยำ การเรียกคืน และความแม่นยำ การปรับปรุงนี้ทำให้ความสามารถของ OSCNet ในการเรียนรู้คุณลักษณะที่โดดเด่นมากขึ้นจากข้อมูล เมื่อเปรียบเทียบกับคุณลักษณะที่สร้างขึ้นด้วยแบบดั้งเดิม

การตรวจจับและการจดจำอาหารโดยทีมผู้วิจัย คากายะ และคณะ (Kagaya et al., 2014) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการกับงานตรวจจับ และจดจำรูปภาพอาหาร เนื่องจากอาหารมีหลากหลายประเภท โดยทั่วไปแล้วการจดจำรายการอาหารด้วยรูปภาพเพื่อแยกประเภทอาหาร

ยานพาหนะทางอากาศไร้คนขับโดยทีมผู้วิจัย เบนญาดิรา และคณะ (Benjdira et al., 2019) นำมาใช้มากขึ้นในการเฝ้าระวัง และติดตามการจราจรความท้าทายที่สำคัญประการหนึ่งคือการตรวจจับ และนับรถยนต์อย่างแม่นยำแบบเรียลไทม์โดยใช้ภาพถ่ายทางอากาศเพื่อติดตามการจราจร

การแปลอัตโนมัติ มีการใช้โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้ในบริบทของการเรียนรู้เชิงลึกสำหรับการแปลอัตโนมัติระหว่างคู่ภาษา เช่น ภาษาอังกฤษ และ ภาษาฝรั่งเศส โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้ในการแปลระหว่างคู่ภาษา เช่น ภาษาจีน และ ภาษาอังกฤษด้วยความแม่นยำระดับสูง ทำให้ไม่จำเป็นต้องใช้การแปลแบบคำต่อคำหรือผู้ช่วยที่พูดได้สองภาษา

การทำนายคำถัดไปในประโยคโดยทีมผู้วิจัย กรานรอต-ไวลด์ดิง และคลาร์ก (Granroth-Wilding and Clark, 2016) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้เพื่อทำนายคำถัดไปของประโยคโดยคำนึงถึงบริบทบางอย่าง โมเดลของโครงข่ายประสาทเทียมแบบสังวัตนาการสามารถประมวลผลหลายประโยค และ เรียนรู้ว่าคำใดมักจะตามหลังคำอื่นๆ เช่น “ฉันมาจากอินเดีย” ตามด้วย “ฉันพูดภาษาฮินดี”

การรู้จำอักขระที่เขียนด้วยลายมือโดยทีมผู้วิจัย อะห์ลาวัต และคณะ (Ahlawat et al., 2020) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ เพื่อจดจำอักขระที่เขียนด้วยลายมือได้ โครงข่ายประสาทเทียมแบบสังวัตนาการใช้ภาพของตัวละครเป็นข้อมูลเข้า และ แบ่งออกเป็น ส่วนย่อย ๆ ระบุจุดที่สามารถเชื่อมต่อหรือทับซ้อนกับจุดอื่น ๆ เพื่อกำหนดรูปร่างของตัวอักษรที่ใหญ่ขึ้น โมเดลของซีเอ็นเอ็นได้ถูกสร้างขึ้นที่สามารถระบุภาษาต่าง ๆ รวมทั้งภาษาจีน ภาษาอาหรับ และ ภาษารัสเซีย แม้ว่าภาษาเหล่านั้นจะเขียนต่างกัน

การวิเคราะห์ภาพเอ็กซ์เรย์โดยทีมผู้วิจัย โอ้ และคณะ (Oh et al., 2020) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้สำหรับการถ่ายภาพทางการแพทย์เพื่อระบุ COVID-19 หรือความผิดปกติอื่น ๆ ในภาพเอ็กซ์เรย์ โมเดลของ โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถถ่ายภาพส่วนหนึ่งของร่างกายคน เช่น ข้อเข่า และ ระบุตำแหน่งที่ภายในภาพนั้นอาจมีเนื้องอก โดยอ้างอิงจากภาพประกอบที่ คล้ายคลึงกันก่อนหน้าซึ่งประมวลผลโดยโครงข่าย โครงข่ายประสาทเทียมแบบสังวัตนาการนอกจากนี้ยังสามารถใช้โมเดล โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อระบุความผิดปกติจากภาพเอ็กซ์เรย์ มีการใช้ โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อระบุว่าภาพเอกซเรย์บริเวณใดมีเนื้องอกหรือความผิดปกติอื่น ๆ เช่น กระดูกหัก

การตรวจหามะเร็งโดยทีมผู้วิจัย เอ็ม และคณะ (M et al., 2017) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้เพื่อตรวจหามะเร็งในภาพทางการแพทย์ เช่น แมมโมแกรม และ ซีทีสแกน โมเดลของ โครงข่ายประสาทเทียมแบบสังวัตนาการถ่ายภาพของผู้ป่วย และ เปรียบเทียบกับภาพฐานข้อมูลที่มีลักษณะคล้ายกัน โดยระบุเมื่อมีสัญญาณอยู่ในภาพซึ่งบ่งบอกถึงความร้ายกาจหรือความเสียหายต่อเซลล์เนื่องจากทั้งปัจจัยทางพันธุกรรม และ สิ่งแวดล้อม เช่น พฤติกรรมการสูบบุหรี่ โมเดลของ โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถให้ผลลัพธ์ที่แม่นยำสูง และ โครงข่ายประสาทเทียมแบบสังวัตนาการยังสามารถระบุเซลล์มะเร็งได้ด้วย ความแม่นยำ

การตอบคำถามด้วยภาพโดยทีมผู้วิจัย มาลินอฟสกี และคณะ (Malinowski et al., 2015) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการยังนำไปใช้กับการตอบคำถามด้วยภาพ ซึ่งเป็นงาน

ของอากาศยานพาณิชย์คำถามภาษาธรรมชาติเกี่ยวกับภาพ จากนั้นจึงให้คำตอบที่ถูกต้องตามภาพนั้น
โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถถ่ายภาพเป็นตัวรับข้อมูล และ ส่งกลับคำตอบใน
ภาษาธรรมชาติสำหรับคำถามใดๆ ที่ถามเกี่ยวกับภาพนั้น

คำบรรยายภาพโดยทีมผู้วิจัย อัมริตการ์ และจาบาเด (Amritkar and Jabade, 2018) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการยังใช้ในการสร้างคำบรรยายสั้นๆ ที่อธิบายสิ่งที่มีอยู่
ในภาพใหม่ที่ป้อนเข้าสู่โครงข่าย โครงข่ายประสาทเทียมแบบสังวัตนาการ หรือประมวลผลภาพหลาย
ภาพพร้อมกันกับภาพประกอบที่มีอยู่ เช่น ภาพประกอบที่ พบในเว็บไซต์โซเซียลมีเดีย โมเดลของ
โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถใช้ลำดับของภาพเป็นตัวรับข้อมูล และ สร้างประโยค
หนึ่งประโยคขึ้นไปเพื่ออธิบายสิ่งที่มีอยู่ในภาพเหล่านั้น

การพิสูจน์ตัวตนด้วยไบโอเมตริกโดยทีมผู้วิจัย สูดาก้า และกาเพโตรวา (Sudhakar and Gavrilova, 2020) ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการถูกนำมาใช้เพื่อพิสูจน์ตัวตนของ
ผู้ใช้ด้วยไบโอเมตริกโดยการระบุลักษณะทางกายภาพบางอย่างที่เกี่ยวข้องกับใบหน้าของบุคคล
โมเดลของซีเอ็นเอ็น

จากการใช้งานจริงของ โครงข่ายประสาทเทียมแบบสังวัตนาการที่ระบุไว้ข้างต้น
สามารถสังเกตได้ว่า โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถทำงานร่วมกับ (i) การจัด
ประเภทภาพ (ii) การตรวจจับวัตถุ และ (iii) การแบ่งส่วนภาพ โดยมีรายละเอียดของแต่ละงานใน
หัวข้อที่ 2.3 การประมวลผลภาพ

2.3 การประมวลผลภาพ

ประมวลผลภาพเป็นการเตรียมความพร้อมภาพสำหรับงานตรวจจับวัตถุในรูปภาพ
แบบพื้นฐานของกระบวนการประมวลผลข้อมูลภาพได้หลากหลายวิธีการ การประมวลผลรูปภาพเป็น
กระบวนการปรับปรุงคุณภาพของรูปภาพ หรือ การเลือกพื้นที่ของข้อมูลรูปภาพวิธีการประมวลผล
ภาพประกอบด้วยสามขั้นตอนหลัก: (i) นำเข้าข้อมูลรูปภาพจากที่เก็บรูปภาพ (ii) การวิเคราะห์รูปภาพ
และ การจัดการรูปภาพ และ (iii) ผลการวิเคราะห์รูปภาพ มีกลไกการประมวลผลภาพหลายอย่าง
(Kour et al., 2013) การประมวลผลภาพได้มีการปรับปรุงรูปภาพหลากหลายวิธีดังนี้

การปรับปรุงภาพ (Image Enhancement) คือกระบวนการแปลงข้อมูลภาพเป็น
ตัวเลขเพื่อสร้างภาพประกอบที่ เน้นรายละเอียดที่ต้องการหรือปรับช่วงของโทนแสงที่ต้องการของ
ภาพเทียบกับข้อมูลหรือรายละเอียดอื่น ๆ ของภาพ การปรับปรุงรูปภาพใช้เพื่อโฟกัสหรือเลือก
คุณลักษณะที่สำคัญของรูปภาพ เช่น การปรับภาพให้คมชัดชัดเจนขึ้นเพื่อแยกรายละเอียด เช่น เลข
ทะเบียนยานพาหนะ หรือฟิล์มเอ็กซเรย์ในบางพื้นที่ ในภาพถ่ายทางอากาศ อาจจำเป็นต้องปรับปรุง
ขอบหรือเส้นเพื่อเลือกอาคารหรือวัตถุอื่นๆ ส่วนประกอบสเปกตรัมบางอย่างของภาพอาจจำเป็นต้อง
ปรับปรุงในภาพประกอบที่ ได้จากกล้องโทรทรรศน์หรือยานสำรวจอวกาศ ในบางกรณีอาจต้องใช้
ความคมชัด (Stanimirović et al., 2015) ดังภาพประกอบที่ 2.6



ภาพประกอบที่ 2.6 การใช้การปรับปรุงภาพ

การปรับความเข้ม (Contrast Adjustment) จะแมปค่าความเข้มของภาพใหม่กับช่วงการแสดงผลทั้งหมดของประเภทข้อมูล ภาพประกอบที่ มีการปรับความเข้มของภาพจะมีความแตกต่างอย่างชัดเจนระหว่างภาพขาวดำ ดังภาพประกอบที่ 2.7



ภาพประกอบที่ 2.7 การปรับคอนทราสต์

การกรองรูปภาพ (Image Filter) ส่งภาพผ่านตัวกรองเพื่อให้ได้ภาพประกอบที่ ได้รูปภาพประกอบที่ ได้จะมีคุณสมบัติที่แตกต่างจากรูปภาพเริ่มต้น จุดประสงค์หลักของการกรองรูปภาพคือเพื่อเพิ่มหรือลดทอนคุณสมบัติบางอย่างของรูปภาพ เพื่อให้ได้ภาพประกอบที่ มีคุณสมบัติตามต้องการ ประโยชน์บางประการของการใช้ฟิลเตอร์สามารถพบเห็นได้บ่อย ๆ เช่น การเบลอ การเน้นขอบของวัตถุในภาพ การลบจุดขาวดำในภาพ ดังภาพประกอบที่ 2.8



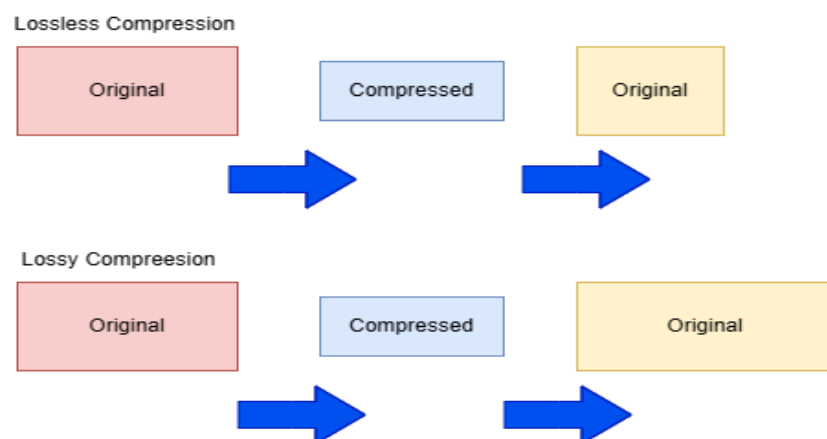
ภาพประกอบที่ 2.8 การกรอกภาพ

Thresholding เป็นหลักการที่ใช้ค่าคงที่เพื่อเปรียบเทียบกับพิกเซลในกราวด์ แต่ถ้าค่าของพื้นที่พิกเซลน้อยกว่าค่าคงที่ มันจะเปลี่ยนค่าพิกเซลของพื้นที่เป็น 0 แต่ถ้าค่าพิกเซลของพื้นที่นั้นมากกว่า จะเปลี่ยนค่าพิกเซลของพื้นที่เป็น 1 รูปภาพตัวรับข้อมูลต้องเป็นโทนสีเทา และ ตัวประมวลผลผลลัพธ์จะเป็นรูปภาพไบนารี สมการของฟังก์ชันเกณฑ์แสดงไว้ดังนี้:

$$g(x,y) = 0 \quad \text{if } f(x,y) \geq \text{threshold value} \quad (6)$$

$$g(x,y) = 1 \quad \text{if } f(x,y) < \text{threshold value} \quad (7)$$

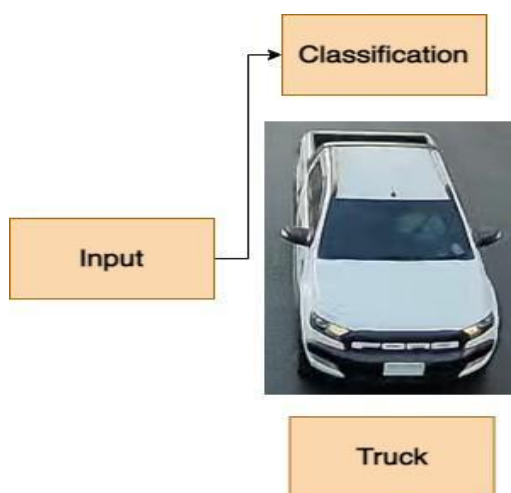
การบีบอัดภาพ (Image Compression) การบีบอัดรูปภาพเป็นกระบวนการลดขนาดไฟล์รูปภาพและยังคงรักษาคุณภาพของรูปภาพเอาไว้ วิธีการบีบอัดรูปภาพนี้จะช่วยประหยัดการเก็บพื้นที่ของคอมพิวเตอร์โดยที่ยังรักษาคุณภาพของรูปภาพอยู่



ภาพประกอบที่ 2.9 การบีบอัดภาพ

2.4 การจำแนกประเภทข้อมูลภาพ (Image Classification)

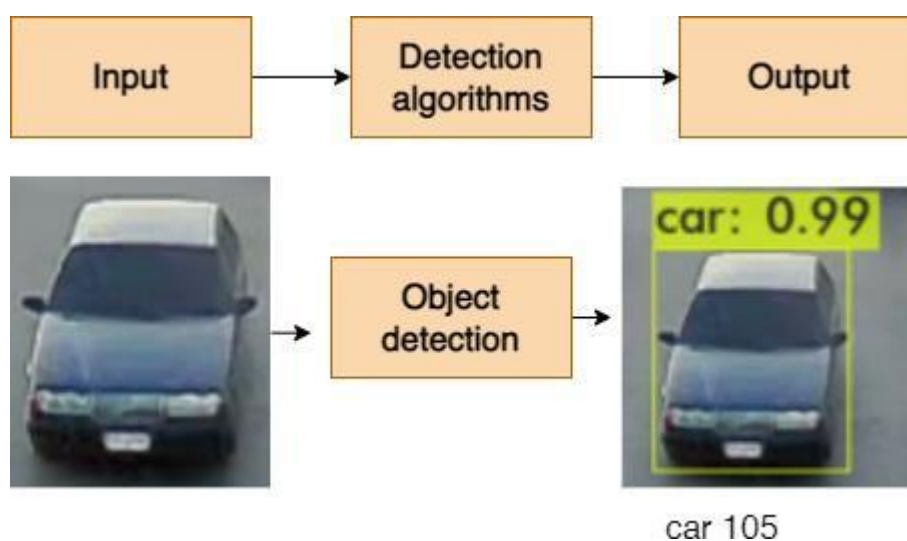
การจำแนกประเภทข้อมูลภาพเป็นกระบวนการที่สำคัญในขอบเขตของคอมพิวเตอร์วิทัศน์และปัญญาประดิษฐ์ ซึ่งเกี่ยวข้องกับกระบวนการจัดหมวดหมู่ภาพเป็นคลาสหรือหมวดหมู่ที่กำหนดไว้ล่วงหน้าตามเนื้อหาภาพ ด้วยการพัฒนาเทคนิคการเรียนรู้เชิงลึกอย่างรวดเร็วโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ ความสำคัญของการจำแนกภาพ คือ การจำแนกภาพมีจุดมุ่งหมายเพื่อสอนโมเดลถึงวิธีจัดจำรูปแบบคุณสมบัติและวัตถุภายในภาพและกำหนดให้กับหมวดหมู่ที่เกี่ยวข้องกัน การจดจำนี้ช่วยให้โมเดลสามารถระบุข้อมูลภาพ และสามารถพัฒนานำไปใช้ร่วมกับแอปพลิเคชันที่หลากหลายในหน่วยงานต่าง ๆ ขั้นตอนการฝึกอบรมเริ่มต้นต้องให้โมเดลแมชชีนเลิร์นนิงเรียนรู้จากชุดข้อมูลที่มีป้ายกำกับ ชุดข้อมูลที่ใช้จะประกอบด้วยรูปภาพที่จับคู่กับป้ายกำกับคลาสที่สนใจให้ตัวโมเดลจำแนกโดยสอนให้โมเดล ในระหว่างการฝึกอบรมโมเดลจะแยกคุณสมบัติจากรูปภาพ เช่น ขอบพื้นผิว และ รูปร่าง และใช้อัลกอริทึมต่าง ๆ เพื่อระบุตำแหน่งคุณสมบัติเหล่านี้กับคลาสเฉพาะ (Lu and Weng, 2007) การจำแนกภาพใช้ร่วมกับแอปพลิเคชันจำนวนมาก เช่น การถ่ายภาพทางการแพทย์ (Mall et al., 2019) การจำแนกภาพช่วยในการวินิจฉัยเงื่อนไขทางการแพทย์โดยการระบุความผิดปกติในภาพทางการแพทย์เช่นรังสีเอกซ์ MRIs และการสแกน CT หมวดหมู่ผลิตภัณฑ์ ระบบการแนะนำผลิตภัณฑ์ที่ใช้การจำแนกรูปภาพเพื่อทำความเข้าใจและจัดหมวดหมู่ผลิตภัณฑ์ปรับปรุงประสบการณ์ผู้ใช้ และ เพิ่มยอดขาย ความปลอดภัยและการเฝ้าระวัง การจำแนกภาพช่วยในการระบุภัยคุกคามความปลอดภัยจดจำใบหน้าและตรวจสอบพื้นที่ที่น่าสนใจแบบเรียลไทม์ การตรวจสอบสิ่งแวดล้อมช่วยตรวจสอบที่อยู่อาศัยตามธรรมชาติและสัตว์ป่า โดยการระบุและนับสายพันธุ์ในรูปภาพและวิดีโอโดยอัตโนมัติเป็นต้น ตัวอย่างการจำแนกรูปภาพดังภาพประกอบที่ 2.10



ภาพประกอบที่ 2.10 ตัวอย่างรูปภาพการจำแนกประเภทของรูปภาพ

2.5 การตรวจจับวัตถุ (Object detection)

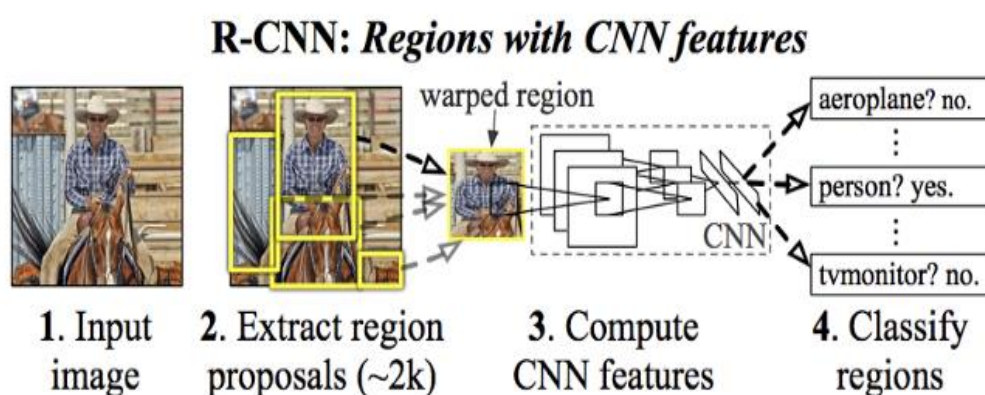
การตรวจจับวัตถุเป็นเทคนิคการมองเห็นด้วยคอมพิวเตอร์สำหรับค้นหาวัตถุในภาพหรือวิดีโอ ใช้การเรียนรู้ของเครื่องหรือการใช้อัลกอริทึมกับการตรวจจับวัตถุการเรียนรู้เชิงลึกเพื่อระบุวัตถุที่ต้องการ หลักการคล้ายกับสมองของคนซึ่งสามารถจดจำวัตถุที่มองเห็นได้ด้วยตา ด้วยวิธีการเรียนรู้เชิงลึกสามารถนำไปใช้ร่วมกันได้โดยการใช้ชุดข้อมูลในการอบรมเพื่อระบุวัตถุในภาพ หรือใช้ในแอปพลิเคชันต่าง ๆ เช่น การรักษาความปลอดภัย และการเฝ้าระวัง การค้นหา การค้นหาใบหน้า และอื่น ๆ มากมายในงานตรวจจับวัตถุ แต่แนวทางที่ใช้กันมากที่สุดคือการใช้การเรียนรู้เชิงลึก โดยเฉพาะ โครงข่ายประสาทเทียมแบบสังวัตนาการ (Zhiqiang and Jun, 2017) อย่างที่กล่าวไปในหัวข้อ โครงข่ายประสาทเทียมแบบสังวัตนาการ เป็นโครงข่ายประสาทเทียมประเภทหนึ่งซึ่งได้รับการออกแบบมาโดยเฉพาะสำหรับงานประมวลผลภาพ เนื่องจากมีกระบวนการทำงานที่ทางคณิตศาสตร์ที่ทำให้คอมพิวเตอร์มองเห็นเหมือนดวงตาคน และการจับคุณลักษณะของวัตถุในรูปภาพจึงเหมาะสมสำหรับงานตรวจจับวัตถุ โครงข่ายประสาทเทียมแบบสังวัตนาการ ที่ได้รับการฝึกอบรมสำหรับการตรวจจับวัตถุจะแสดงกล่องขอบเขต ที่แสดงตำแหน่งที่ตรวจพบวัตถุในภาพพร้อมกับป้ายกำกับตัวอย่างของการตรวจจับวัตถุ ดังภาพประกอบที่ 2.11



ภาพประกอบที่ 2.11 ตัวอย่างรูปภาพการตรวจจับวัตถุ

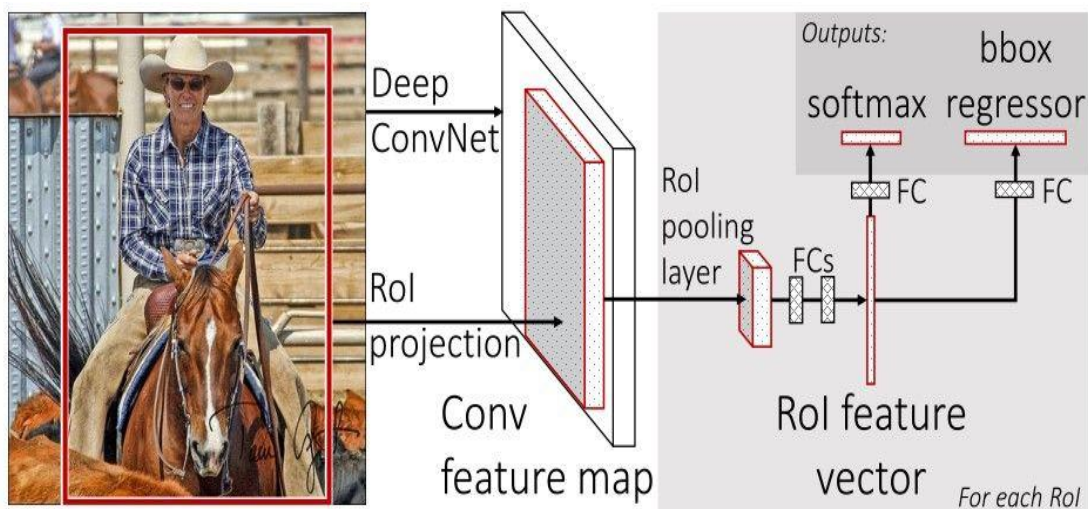
ตัวอย่างของโครงข่ายดังกล่าวคือโมเดล Faster R-CNN (Region-based Convolutional Neural Network) ที่เป็นที่ยอมรับ ซึ่งเป็นโครงข่ายแบบ end-to-end ที่ฝึกอบรมได้อย่างสมบูรณ์ โมเดล R-CNN ที่เร็วขึ้นจะสลับระหว่างการปรับแต่งอย่างละเอียดสำหรับงานข้อเสนอภูมิภาค คือ การคาดคะเนภูมิภาคในภาพที่อาจมีวัตถุอยู่ จากนั้นจึงปรับแต่งพัฒนาโมเดลอย่างละเอียดสำหรับการตรวจจับวัตถุ ซึ่งการตรวจจับวัตถุมีหลายสถาปัตยกรรมมากมายที่เป็นนิยมจะยกตัวอย่าง ดังนี้

สถาปัตยกรรม R-CNN ย่อมาจาก Region-based Convolutional Neural Network (Girshick et al., 2014) หลักแนวคิดการทำงานของ R-CNN คือ การตรวจจับวัตถุที่อยู่ในรูปภาพประกอบที่ 2.12 โดยใช้หลักการตีกรอบกล่องรอบวัตถุเอาไว้เพื่อในการตรวจจับ จะแสดงให้เห็นกระบวนการทำงานของ R-CNN ดังนี้ (i) การเสนอพื้นที่สนใจในรูปภาพวัตถุในรูปภาพด้วย Selective Search (ii) การฝึก และ ปรับแต่งอย่างละเอียดบนตัวของโครงข่ายประสาทเทียมแบบสังวัตนาการ (iii) การฝึกตัวโมเดลประเภทด้วยกระบวนการ ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) ที่จะช่วยแยกประเภทของวัตถุด้วยการแยกคลาสแต่ละคลาสของวัตถุ และ กระบวนการสุดท้ายคือ (iv) การฝึกการตีขอบเขตที่ทำนายให้แม่นยำสำหรับการตรวจจับวัตถุรูปภาพ



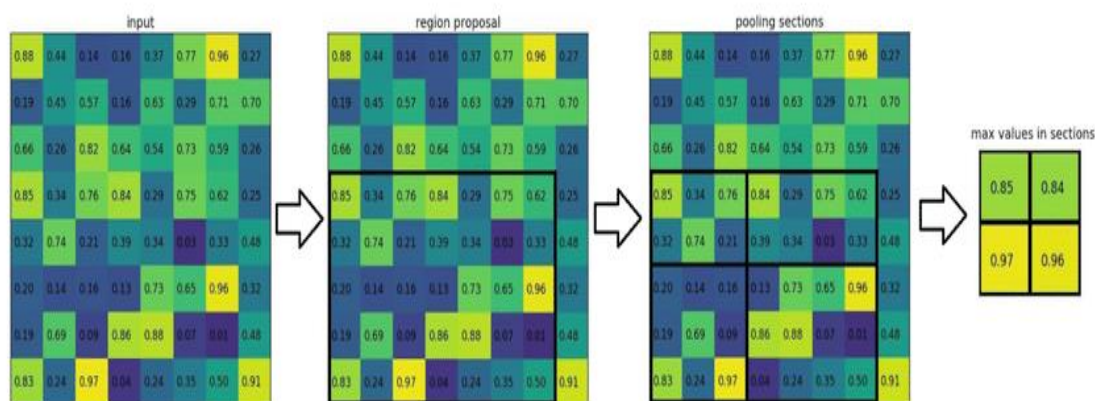
ภาพประกอบที่ 2.12 รายละเอียดการทำงาน RCNN โดย (Girshick et al., 2014)

สถาปัตยกรรม Fast R-CNN เป็นตัวโมเดลตรวจจับวัตถุที่พัฒนาขึ้น (Girshick, 2015) แก้ไขปัญหาต่าง ๆ ที่เกิดขึ้น ใน R-CNN ปกติส่งพื้นที่ที่ตีกรอบไว้ส่งเข้า คอนโวลูชันเลเยอร์ทีละชั้น แต่ Fast R-CNN จะทำการส่งผ่านโดยตรงเลย และข้อดีของ Fast R-CNN คือ ความเร็วในตัวของ Fast R-CNN ได้มีการเสนอ เลเยอร์ใหม่ขึ้นมาเรียกว่า ROI pooling ซึ่งจะทำหน้าที่แยกเวกเตอร์ที่มีคุณลักษณะความยาวเท่ากันคงที่เพื่อส่งไปยังเลเยอร์สมบูรณ์แบบ กับ softmax เพื่อให้ทำการแบ่งประเภทของวัตถุที่ถูกตีกรอบนั้นเป็นวัตถุอะไร และ ข้อได้เปรียบสำหรับตัวโมเดล Fast R-CNN กับตัว R-CNN คือกระบวนการต่าง ๆ ของตัวโมเดล R-CNN (การเสนอพื้นที่, การดึงคุณลักษณะสำคัญของวัตถุในรูปภาพ, การจัดหมวดหมู่แบ่งแยกกลุ่มข้อมูลโดยการใช้ ซัพพอร์ตเวกเตอร์แมชชีน) Fast R-CNN จะเร็วกว่าด้วยการสร้างโครงข่ายรวมกันให้เหลือเพื่อขั้นตอนเดียวและ การมีเลเยอร์ใหม่ ROI pooling ช่วยให้การคำนวณเร็วขึ้นตัวโมเดล Fast R-CNN แม่นยำกว่า R-CNN สถาปัตยกรรมทั่วไปของ Fast R-CNN (Jiang and Learned-Miller, 2017) ดังดังภาพประกอบที่ 2.13



ภาพประกอบที่ 2.13 สถาปัตยกรรมของ Fast R-CNN โดย (Girshick, 2015)

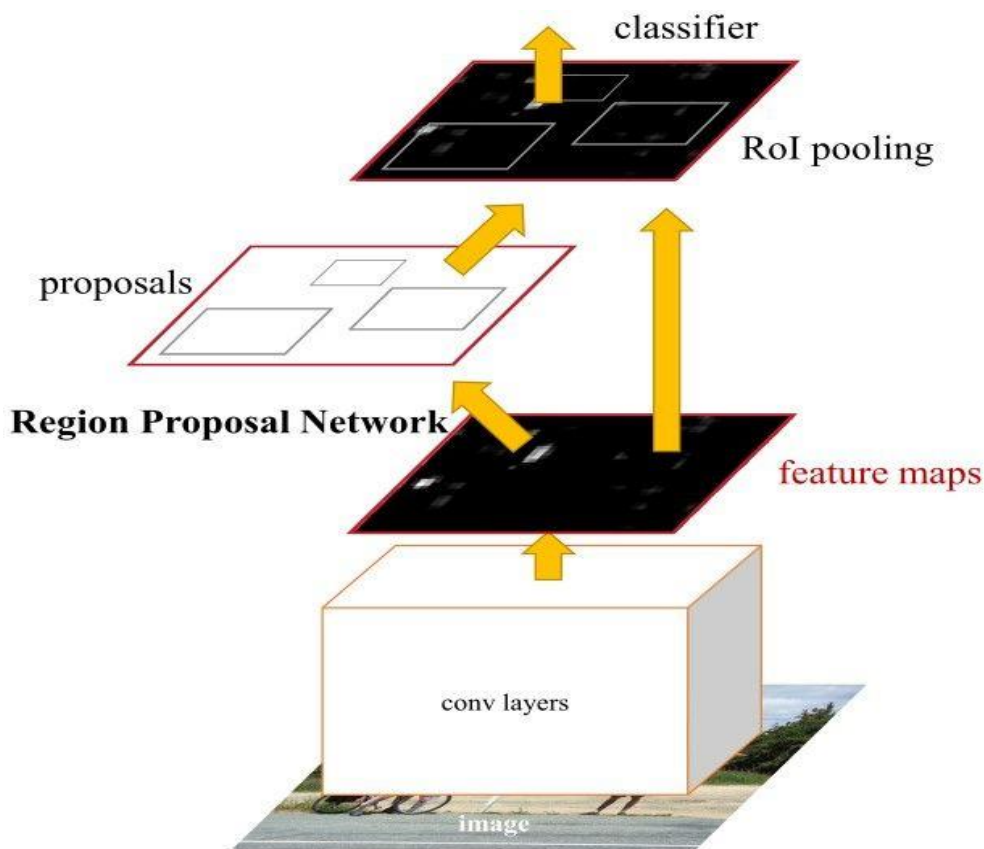
ROI Pooling คือการทำงานโดยทำการแบ่งเขตเสนอพื้นที่ของรูปภาพออกเป็นกริดของเซลล์ กระบวนการจะทำการรวบรวมค่าสูงสุด และ นำไปใช้กับแต่ละกริดเพื่อส่งกลับค่าเดียว ค่าทั้งหมดจากทุกกริดเซลล์จะแสดงถึงเวกเตอร์คุณลักษณะ ถ้าขนาดกริดคือ 3×3 การคำนวณ จะแบ่งช่องเท่าจำนวนที่ตั้งไว้แล้วเลือกค่ามากที่สุดของแต่ละช่องออกมาวิธีการนี้คล้ายกับของ max pooling แต่ ROI ยึดขนาดให้คลุมกับกรอบที่ตีไว้ความยาวเวกเตอร์จะเป็น 4 ดังดังภาพประกอบที่ 2.14



ภาพประกอบที่ 2.14 ROI Pooling โดย (Grel, 2017)

Faster R-CNN (Ren et al., 2017) ได้มีการปรับปรุงกระบวนการในตัวโมเดลให้มีความเร็วกว่าตัวของโมเดล Fast R-CNN ด้วยโครงข่ายที่เพิ่มเข้ามา คือ Region Proposal Network หรือ โครงข่าย RPN ที่เป็นการสกัดคุณสมบัติคุณลักษณะเด่นของวัตถุในรูปออกมา และ ทำการทำนายว่าจะเป็วัตถุในรูปภาพ แล้วนำส่วนที่คาดการณ์เข้าในชั้นของคอนโวลูชันเลเยอร์เพื่อคำนวณหาคุณลักษณะต่อไป การเสนอพื้นที่ RPN เป็นรูปแบบสี่เหลี่ยมผืนผ้า กว้าง คุณ ยาวใช้คำสั่งด้วย

โครงข่ายประสาทโดยให้ความสนใจจุดที่สนใจเพื่อบอกการตรวจจับวัตถุ เทคนิคการคำนวณอัตราพื้นที่ทับซ้อนกันช่วยในการช่วยลดเวลาในการคำนวณ RPN มีหน้าที่สร้างเสนอพื้นที่ (Region) โดยใช้แนวคิดของในโครงข่ายประสาทเทียมแบบสังวัตนาการทำให้โมเดลตรวจจับได้แม่นยำ และรวดเร็วกว่าตัวโมเดล Fast R-CNN ไปยังตำแหน่งวัตถุในภาพ ดังภาพประกอบที่ 2.15

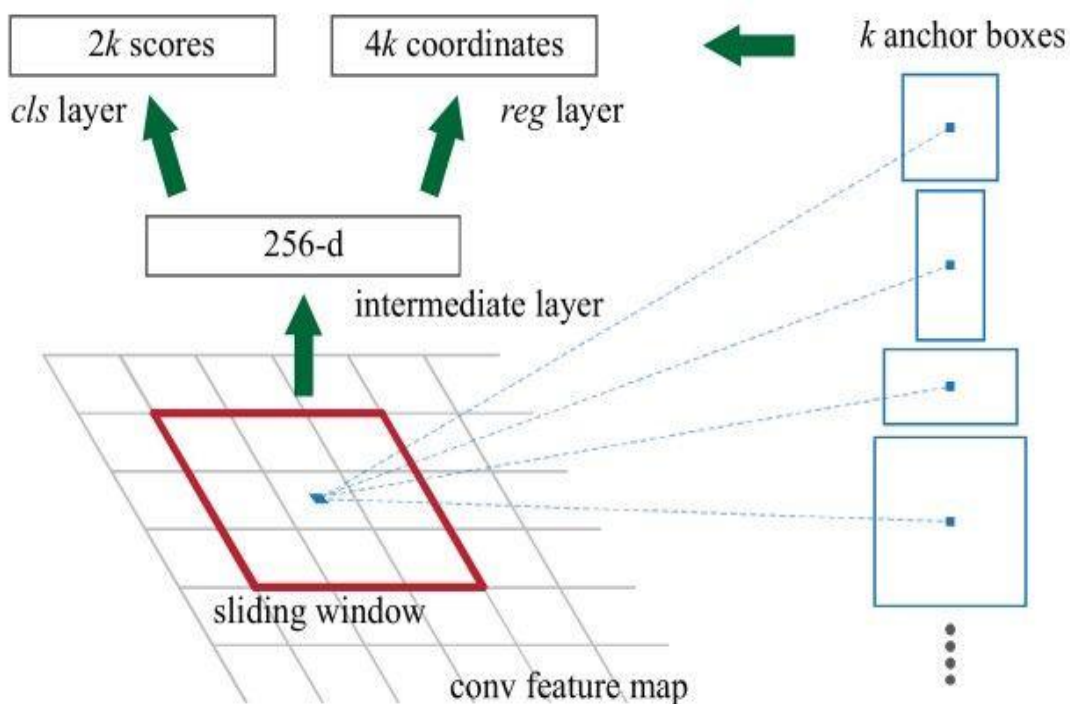


ภาพประกอบที่ 2.15 สถาปัตยกรรม Faster R-CNN โดย (Ren et al., 2017)

คอนโวลูชันทำงานร่วมกับกระบวนการ RPN และ Fast R-CNN Faster R-CNN มี RPN ไว้สร้างข้อเสนอพื้นที่ในแต่ละพื้นที่ และ สำหรับข้อเสนอพื้นที่ทั้งหมดในรูปภาพเวกเตอร์คุณลักษณะที่มีความยาวคงที่จะถูกแยกออกจากโดยใช้เลเยอร์ ROI Pooling เวกเตอร์คุณลักษณะที่แยกออกมาจะถูกจัดประเภทโดยใช้ Fast R-CNN

โครงข่าย Region Proposal Network (RPN) ใน Faster R-CNN ประมวลผลภาพโดยผ่านคอนโวลูชันเลเยอร์เดียวกันกับที่ใช้ในโครงข่ายการตรวจจับ Fast R-CNN ดังนั้นเลเยอร์เครื่องข่ายเดียวกันได้ ส่วนการฝึกอบรมจะทำเพียงครั้งเดียว RPN ทำงานบนแผนการทำงานของคุณลักษณะตัวประมวลผลผลลัพธ์ที่ย้อนกลับมาจากคอนโวลูชันเลเยอร์ (König et al., 2017) ภาพประกอบที่ 2.16 จะเห็นหน้าต่างสี่เหลี่ยมขนาด $n \times n$ หน้าต่างนี้จะทำหน้าที่เลื่อนผ่านคุณลักษณะที่นำเข้ามา และ สกัดคุณสมบัติของคุณลักษณะของภาพในแต่ละหน้าต่าง ดังนั้นจึงมีการสร้างกรอบเสนอพื้นที่หลายกรอบขึ้นมา ส่วนของกล่องอ้างอิง (Anchor box) คือ กล่องสี่เหลี่ยมที่ทำ

การครอบคลุมวัตถุหลาย ๆ กล้อง และ คอนโวลูชันที่ใช้ร่วมกันจะนำส่งผ่านหน้าต่าง $n \times n$ ซึ่ง $n = 3$ ตามรูปภาพด้านล่าง และ หน้าต่างแต่ละหน้าต่างจะเป็นตัวแปร K ซึ่งเกิดจากข้อเสนอพื้นที่ที่ถูกสร้างขึ้น และ กำหนดเป็นพารามิเตอร์ตามกล่องอ้างอิง ซึ่งมีพารามิเตอร์ 2 ตัวคือ มาตรฐาน และ อัตราส่วนของภาพ โดยทั่วไปนั้นจะมี 3 สเกล และ 3 อัตราส่วน ดังนั้นจึงมี $K = 9$ กล่องอ้างอิงทั้งหมด แต่ K มีมากกว่าคือ 9 นั่นก็คือ K เป็นภูมิภาคต่าง ๆ ของรูปภาพจะเกิดขึ้นจากข้อเสนอแต่ละภูมิภาค ซึ่งแต่ละ K ภูมิภาคจะแตกต่างกันไปทั้งในระดับมาตรฐานหรืออัตราส่วนกว้างยาว



ภาพประกอบที่ 2.16 โครงข่าย Region Proposal Network โดย (Ren et al., 2017)

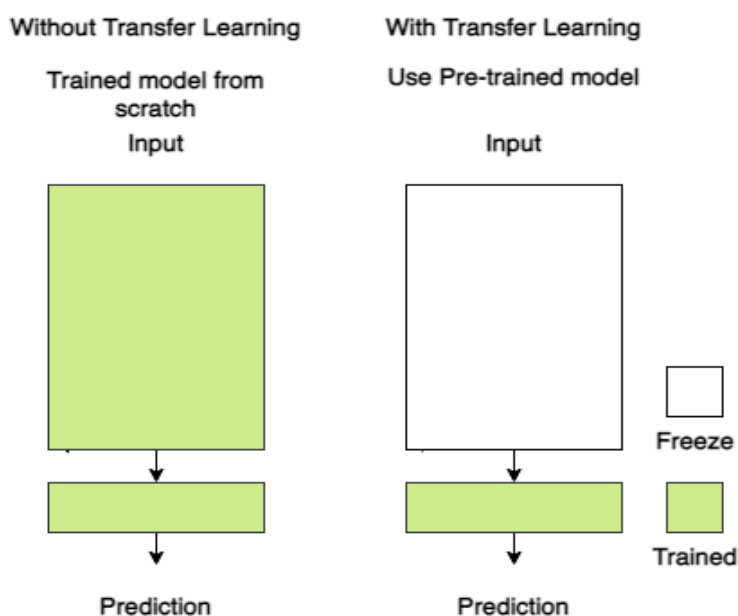
Objectness Score คือ คะแนนที่ถูกกำหนดขึ้นมาเพื่อวัดค่าว่าโมเดลการตรวจจับวัตถุของแต่ละคลาสของวัตถุได้มากน้อยแค่ไหนในระหว่างตรวจจับ เลเยอร์จะแสดงเวกเตอร์ขององค์ประกอบของ 2 รายการสำหรับข้อเสนอแต่ละพื้นที่ หากองค์ประกอบของวัตถุแรกคือ เลข 1 และองค์ประกอบของวัตถุที่ 2 คือ เลข 0 (Dai et al., 2016)

2.6 การถ่ายโอนการเรียนรู้

การถ่ายโอนการเรียนรู้ (Transfer learning) เป็นการฝึก และ การปรับแต่งโมเดลบนโครงข่ายประสาทเทียมแบบสังวัตนาการ โดยปกติจะไม่ฝึกใหม่ตั้งแต่ศูนย์หรือเริ่มใหม่ แต่จะใช้โมเดลที่ผ่านการฝึกอบรมมาแล้วเพื่อลดเวลาในการทำงานโดยมีโมเดลที่มีการใช้ฐานข้อมูลการถ่ายโอนการเรียนรู้เป็นเทคนิคที่ช่วยลดปัญหาการฝึกอบรมในการเรียนรู้ของเครื่องได้เยอะที่ไม่ต้องเริ่มใหม่จากศูนย์ ในการเรียนรู้เชิงลึกก็ใช้การถ่ายโอนการเรียนรู้เป็นเทคนิคที่ใช้โมเดลโครงข่ายประสาทเทียมได้รับการฝึกอบรมมาแล้วมาเป็นอย่างดี นำมาฝึกอบรมด้วยชุดข้อมูลที่ต้องการ เลเยอร์หนึ่งหรือ

หลายเลเยอร์จากโมเดลที่ผ่านการฝึกอบรมจะถูกนำไปใช้ในโมเดลใหม่ที่ได้รับการฝึกอบรมกับชุดข้อมูลใหม่ (Shin et al., 2016) ซึ่งเป้าหมายที่ได้รับการฝึกอบรมก็เปลี่ยนไป ตัวอย่างเช่น ใส่ชุดข้อมูลฝึกอบรมเรียนรู้เกี่ยวกับหมวดหมู่ภาพชุดหนึ่งโมเดลแรกได้รับการฝึกอบรมข้อมูลสำหรับตรวจจับสุนัขและ แมวหลังจากนั้นเจอนกตัวโมเดลไม่สามารถตรวจจับได้ในโมเดลที่ยังไม่ใช้การถ่ายโอนการเรียนรู้พอโมเดลที่สองใช้การถ่ายโอนการเรียนรู้ให้ตรวจจับนก โมเดลตรวจจับประเภทปลา (Raza and Hong, 2020) ก็สามารถตรวจจับได้ (Chiba and Sasaoka, 2021)

ประโยชน์หลักของการถ่ายโอนการเรียนรู้สำหรับการเรียนรู้ของเครื่อง ได้แก่ (i) ความสะดวกจำเป็นสำหรับชุดข้อมูลการฝึกอบรมที่มีป้ายกำกับจำนวนมากสำหรับโมเดลใหม่ทุกสถาปัตยกรรม (ii) ปรับปรุงประสิทธิภาพของการพัฒนาแมชชีนเลิร์นนิงและการปรับใช้สำหรับโมเดลหลาย ๆ สถาปัตยกรรม (iii) แนวทางทั่วไปในการแก้ปัญหาเครื่องจักร การใช้ประโยชน์จากอัลกอริธึมที่แตกต่างกันเพื่อแก้ปัญหาความท้าทายใหม่ๆ และ (iv) สามารถฝึกโมเดลภายในการจำลองแทนสภาพแวดล้อมจริง คำนี้นักในชั้นที่ใช้ซ้ำอาจใช้เป็นจุดเริ่มต้นสำหรับกระบวนการฝึกอบรมและปรับเปลี่ยนเพื่อตอบสนองต่อปัญหาใหม่ การใช้งานนี้ถือว่าการเรียนรู้การถ่ายโอนเป็นรูปแบบการเริ่มต้นค่าน้ำหนักประเภทหนึ่ง สิ่งนี้อาจมีประโยชน์เมื่อปัญหาที่เกี่ยวข้องกันแรกมีข้อมูลที่ติดป้ายกำกับมากกว่าปัญหาที่สนใจ ตัวอย่างของการถ่ายโอนการเรียนรู้ดังภาพประกอบที่ 2.17 ดังนั้นแนวคิดคือการกำหนดแก้ไขค่าน้ำหนัก ของบางเลเยอร์ระหว่างการฝึกและปรับแต่งส่วนที่เหลือให้ตรงกับวัตถุที่ต้องการตรวจจับ



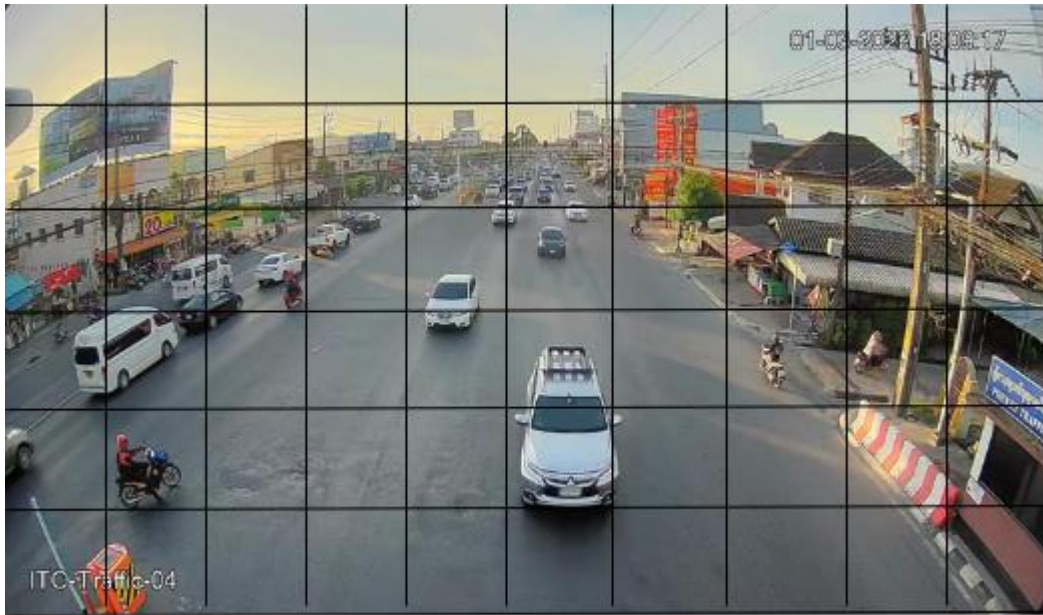
ภาพประกอบที่ 2.17 การถ่ายโอนการเรียนรู้

2.7 You Only Look Once (YOLO)

YOLO เป็นตัวย่อของคำว่า “คุณมองเพียงครั้งเดียว” คืออัลกอริทึมที่ตรวจจับและจดจำวัตถุต่าง ๆ ในภาพแบบเรียลไทม์ การตรวจจับวัตถุใน YOLO มีทั้งความเร็วในการตรวจจับ และการตรวจจับหลายวัตถุที่ทับซ้อนกันได้ในเวลาเดียวกันได้อย่างแม่นยำ การตรวจจับของ YOLO ใช้ อัลกอริทึม โครงข่ายประสาทเทียมแบบสังวัตนาการ เพื่อตรวจจับวัตถุแบบเรียลไทม์ตามชื่อที่แนะนำ อัลกอริทึมต้องการการตรวจจับแบบรวดเร็วโดยการเรียนรู้ที่เร็วที่ YOLO ได้ตรวจจับวัตถุต่าง ๆ ในภาพหรือวิดีโอ วัตถุบางอย่างที่ตรวจพบมีทั้งคน, ยานพาหนะ, แก้ว, ก้อนหิน, สิ่งก่อสร้าง, และ อีกมากมาย การตรวจจับวัตถุประกอบไปด้วยหลากหลายโมเดล เช่น R-CNN, Retina-Net, และ Single-Shot MultiBox Detector (SSD) ในการการสร้างโมเดลในการตรวจจับวัตถุที่มีมาก่อน YOLO อัลกอริทึม YOLO ได้รับความนิยมเนื่องจากประสิทธิภาพที่เหนือกว่าเทคนิคการตรวจจับวัตถุดังกล่าว คือ ความเร็ว และ แม่นยำในการตรวจจับวัตถุ YOLO สามารถใช้การถ่ายโอนการเรียนรู้ได้ดีหัวข้อนี้ จึงอธิบาย YOLO สถาปัตยกรรมต่าง ๆ ดังนี้ i) YOLO, ii) YOLO V2, iii) YOLO V3, iv) YOLO V4, v) YOLO V5, vi) YOLO V6, vii) YOLO V7 และ viii) YOLO V8

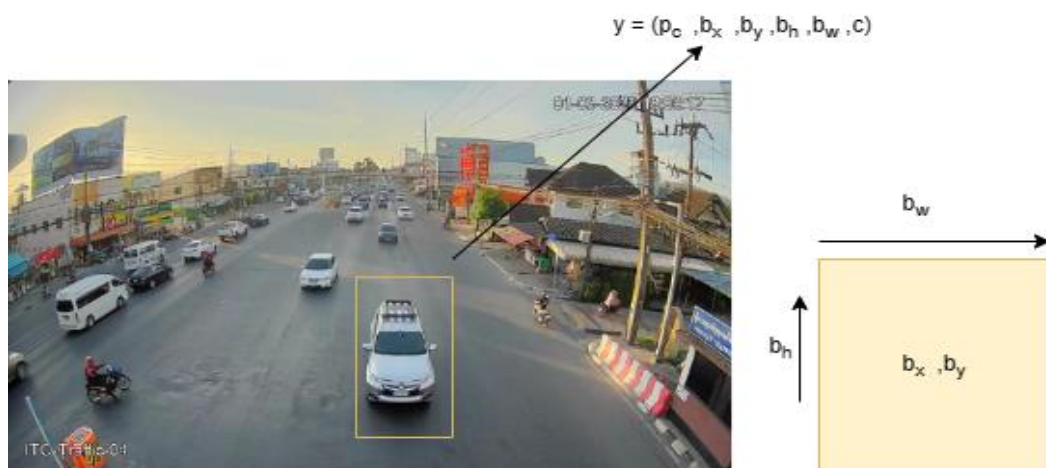
YOLO โมเดล YOLO ได้รับการเผยแพร่ครั้งแรกในปี 2016 (Redmon et al., 2016) หลักการทำงานที่สำคัญของ YOLO คือความสามารถในการตรวจจับวัตถุในรอบเดียว ซึ่งต่างจากกระบวนการสองขั้นตอน YOLO แบ่งรูปภาพออกเป็นตารางและคาดการณ์กรอบขอบเขตและความน่าจะเป็นของคลาสใช้โครงข่าย backbone (Backbone network) เป็นแกนหลัก YOLO ใช้ GoogLeNet ที่ดัดแปลงมาเป็นโครงข่ายแกนหลัก หลังจากนั้น Redmond ได้สร้างโมเดลใหม่ซึ่งได้ชื่อว่า DarkNet-19 ซึ่งเป็นการโครงข่ายประสาทเทียมแบบสังวัตนาการ ได้ทำการใช้ตัวกรอง 3×3 การเพิ่มจำนวนชั้นเป็นสองเท่าในแต่ละชั้นตอนชั้นพูลลิง 1×1 ยังใช้ในการบีบอัดคุณสมบัติของรูปภาพเป็นระยะเรื่อย ๆ ทั่วไปของโครงข่าย (Redmon and Farhadi, 2017) หลังจากพัฒนา DarkNet-19 แล้วได้มีการพัฒนาตัว DarkNet-53 (Redmon and Ali, 2018) ซึ่งทำให้ตัวโมเดลมีการพัฒนาที่ดีขึ้นซึ่งตัวของโมเดล YOLO ได้มีการฝึกอบรมล่วงหน้ามาเรียบร้อยแล้วมากกว่า 1,000,000 ภาพก่อนจะปรับให้เหมาะสมสำหรับการตรวจจับ จากนั้นพบว่าการฝึกอบรมหลายครั้งได้มีการเพิ่มประสิทธิภาพของตัวโมเดล YOLO สูงขึ้นเมื่อสิ้นสุดการฝึกอบรมในการตรวจจับวัตถุสรุปได้ว่า ยังมี การเพิ่มข้อมูล และ การฝึกอบรมที่มากขึ้นทำให้ตัวโมเดลมีการตรวจจับที่ดีทั้งยังให้ค่าความแม่นยำที่มากขึ้น การปรับโครงข่ายการจำแนกสำหรับการตรวจจับวัตถุจะประกอบไปด้วยการลบเลเยอร์สองสามเลเยอร์สุดท้ายของเครื่องข่าย และ เพิ่มเลเยอร์คอนโวลูชันชั้นเลเยอร์ $B(5 + C)$ และ กรองด้วย $N \times N \times B$ สำหรับคาดการณ์กล่องขอบเขต หลักการทำงานของ YOLO หลัก ๆ มีดังนี้ (i) Residual blocks, (ii) Bounding boxes, และ (iii) ไอโอยู

Residual blocks เป็นวิธีการแบ่งภาพแบ่งออกเป็นกริดต่าง ๆ แต่ละตารางมีมิติของ $S \times S$ ในรูปภาพประกอบจะแสดงตารางกริดภาพด้านล่างมีเซลล์กริดจำนวนมากที่มีขนาดเท่ากัน เซลล์กริดทุกเซลล์จะตรวจจับวัตถุที่อยู่ในเซลล์กริด ซึ่งหมายความว่าหากศูนย์กลางวัตถุปรากฏขึ้นภายในเซลล์กริดบางเซลล์เซลล์นี้จะทำหน้าที่ในการตรวจจับ (Lu et al., 2020) ดังภาพประกอบที่ 2.18



ภาพประกอบที่ 2.18 S x S Residual blocks

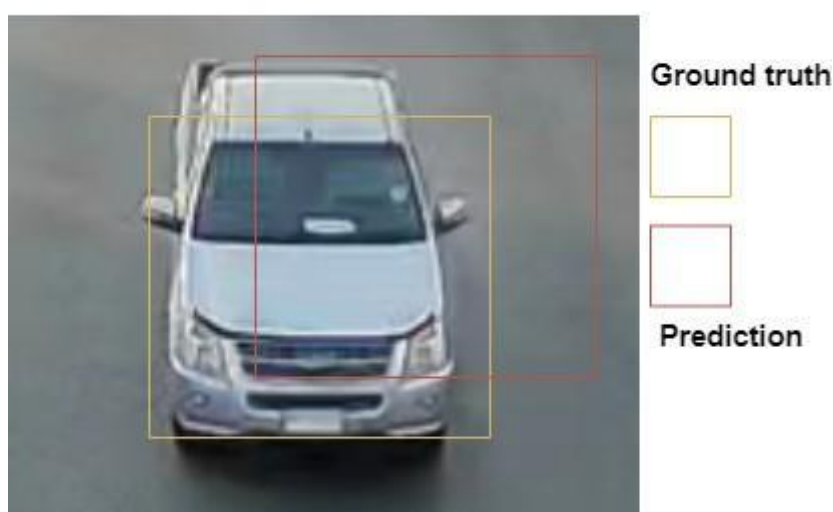
Bounding boxes หรือ กล่องขอบเขตในตัวโมเดลของ YOLO นั้นใช้กล่องขอบเขตในการทำนายวัตถุที่ตรวจจับเหมือนกันซึ่ง YOLO จะทำนายโดยใช้การอธิบายของกล่องขอบเขต คือ x และ y เป็นพิกัดโดยกล่องขอบเขตทำหน้าที่ล้อมขอบเขตที่ทำนายในรูปภาพที่สนใจไว้ (Jana et al., 2018) โดยจะคำนวณตามเซลล์กริดไว้โดยมี ค่าพิกัดนั้นจะมีค่ากำหนดเป็น 0 ถึง 1 ซึ่งจะเป็นตัวกำหนดความกว้างความสูงของกล่องขอบเขตประกอบไปด้วย ความกว้าง (b_w) ความสูง (b_h) คลาส (เช่น คน, สุนัข, ยานพาหนะ, แมว, และ อื่น ๆ) แทนด้วย c สุดท้ายคือศูนย์กล่องขอบเขต (b_x, b_y) ดังภาพประกอบที่ 2.19



ภาพประกอบที่ 2.19 Bounding boxes

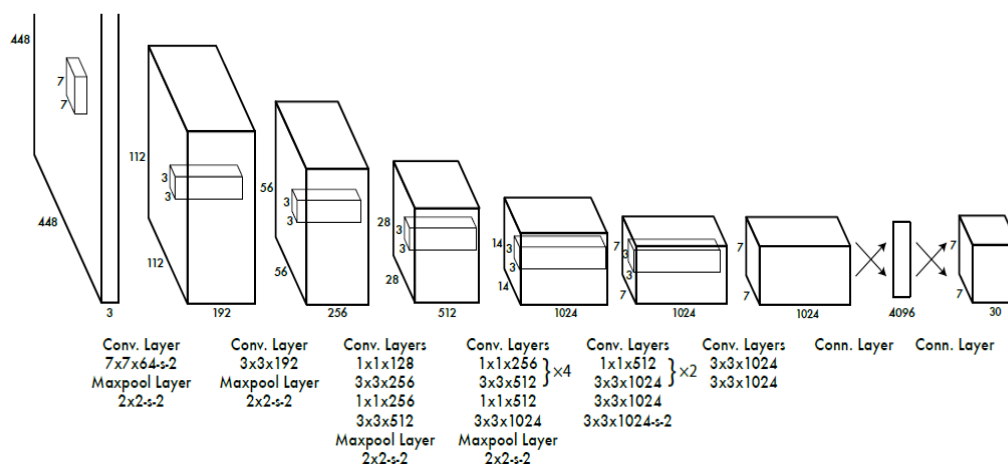
ไอโอยู คือ อัตราส่วนระหว่างพื้นที่ที่เป็น intersection ของ 2 ขอบเขตที่ทำนายหาร ด้วยพื้นที่รวมของกรอบทั้งสองเรียกว่าจุดตัดเหนือสหภาพเป็นการวัดขนาดของการทับซ้อนกัน ระหว่างสองสิ่งกล่องล้อมรอบ (หรือในกรณีทั่วไปคือวัตถุสองชิ้น) ตัวอย่างของขอบเขตการทำนาย และ เฉลยจริงเซลล์กริดแต่ละเซลล์มีหน้าที่ในการทำนายขอบเขตที่ทำนายและคะแนนความเชื่อมั่น ไอโอยูจะเท่ากับ 1 หากขอบเขตที่ทำนายที่คาดการณ์ไว้เหมือนกับขอบเขตที่ทำนายจริง กลไกนี้กำจัด ขอบเขตที่ทำนายที่ไม่เท่ากับกรอบจริง ในภาพด้านล่างมีขอบเขตที่ทำนายสองกรอบ หนึ่งเป็นสีส้ม และ อีกกรอบหนึ่งเป็นสีแดงกล่องสีแดงเป็นขอบเขตที่ทำนายไว้ในขณะที่กรอบสีส้มเป็นกรอบจริงดัง ภาพประกอบที่ 2.20 จะกล่าวอีกครั้งในหัวข้อการตรวจวัดประสิทธิภาพ

$$\text{IoU} = \text{area of overlap} / \text{area of union}$$



ภาพประกอบที่ 2.20 ไอโอยู = พื้นที่ Intersection หารด้วยพื้นที่ทั้งหมด

ขอบเขตที่ทำนายและ ให้คะแนนความเชื่อมั่นเป็นการตรวจประสิทธิภาพการ ตรวจจับของตัวโมเดล YOLO นั้นจะทำนายความน่าจะเป็นของคลาสทำให้มั่นใจได้ว่าขอบเขตที่ ทำนายไว้จะเท่ากับกล่องจริงของวัตถุหรือไม่ ซึ่งเป็นการคำนวณของกรอบที่ดีที่สุดขอบเขตที่ทำนายเอาไว้ (เช่น ความสูง และ ความกว้าง) การตรวจจับขั้นสุดท้ายจะประกอบด้วยขอบเขตที่ทำนายที่ไม่เหมือน โมเดลอื่นซึ่งพอดีกับวัตถุอย่างสมบูรณ์แบบ โครงข่ายของ YOLO มี 24 ชั้นคอนโวลูชัน ตามด้วย 2 ชั้น ที่เชื่อมต่อย่างสมบูรณ์ แทนที่จะใช้โมเดลการเริ่มต้นที่ใช้โดย GoogLeNet เพียงแค่ใช้เลเยอร์การลด 1×1 ชั้นตามด้วยเลเยอร์ที่จับซ้อน 3×3 ชั้น ดังภาพประกอบที่ 2.21



ภาพประกอบที่ 2.21 สถาปัตยกรรม YOLO โดย (Redmon et al., 2016)

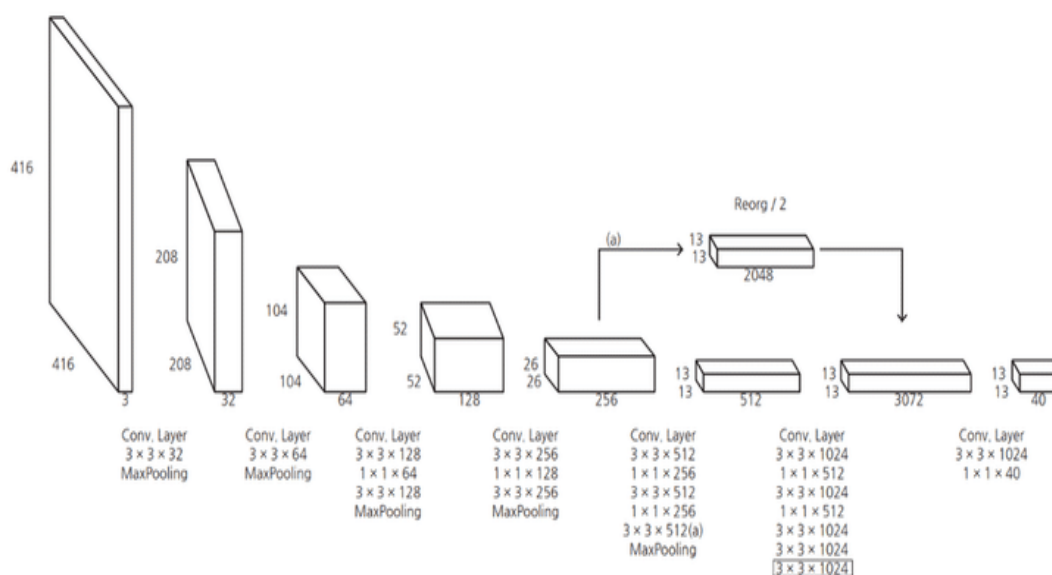
สถาปัตยกรรม YOLO ได้ใช้ DarkNet-53 ซึ่งก็คือ เป็นโครงข่ายประสาทเทียมแบบสังวัตนาการ ที่มีความลึก 53 ชั้นเป็นโครงข่ายเวอร์ชันที่ได้รับการฝึกอบรมมาล่วงหน้าซึ่งได้รับการฝึกอบรมเกี่ยวกับภาพมากกว่าหนึ่งล้านภาพจาก ImageNet ฐานข้อมูลโครงข่ายที่ได้รับการฝึกอบรมล่วงหน้าสามารถจำแนกภาพออกเป็นหมวดหมู่วัตถุ เช่น ดินสอ, ยานาหนะ และ สัตว์มากมาย ด้วยเหตุนี้โครงข่ายจึงได้เรียนรู้คุณสมบัติที่หลากหลาย การเป็นตัวแทนสำหรับภาพประกอบที่หลากหลาย โครงข่ายมีขนาดตัวรับข้อมูลภาพ 256 สำหรับโครงข่ายที่ได้รับการฝึกอบรมล่วงหน้า

จากรูปภาพสถาปัตยกรรมของ YOLO (Redmon et al., 2016) ได้รับแรงบันดาลใจจาก GoogleLeNet (Szegedy et al., 2015) โครงข่ายประกอบด้วยคอนโวลูชันเลเยอร์ 24 ชั้นสำหรับการสกัดคุณลักษณะตามด้วยเลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์ 2 ชั้นสำหรับการทำนายพิกัดกล่องขอบเขต และ ความน่าจะเป็นของวัตถุที่เกี่ยวข้องแทนที่โมดูลการเริ่มต้นของ GoogleLeNet ด้วยเลเยอร์คอนโวลูชันเลเยอร์ 1×1 เพื่อลดมิติความซับซ้อนของแผนที่คุณลักษณะง่ายต่อการทำงาน การฝึกอบรมของ YOLO ประกอบด้วย 2 ขั้นตอน (i) ฝึกโครงข่ายล่วงหน้าเพื่อทำการจำแนกด้วยความละเอียดตัวรับข้อมูล 224×224 บน ImageNet (Deng et al., 2009) โดยใช้เลเยอร์คอนโวลูชันเลเยอร์ 20 ชั้นแรกหลังจากนั้นจะตามด้วยการ pooling เฉลี่ย และ เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์ (ii) ฝึกโครงข่ายสำหรับการตรวจจับโดยการเพิ่มเลเยอร์คอนโวลูชันเลเยอร์ 4 ชั้น และ เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์สองชั้น สำหรับการตรวจจับฝึกตัวรับข้อมูลความละเอียด 448×448 เนื่องจากภาพจะมีขนาดรายละเอียดมากขึ้นเพื่อเพิ่มความแม่นยำในการตรวจจับของตัวโมเดล

ข้อจำกัดของ YOLO จะทำงานได้ไม่ดีกับวัตถุขนาดเล็ก โดยเฉพาะวัตถุที่รวมกลุ่มกัน ตัวของโมเดลจะตรวจจับได้ยากเนื่องจากเซลล์กริดแต่ละเซลล์คาดการณ์จะทับซ้อนคลาสเดียวกัน ทำให้เกิดข้อจำกัดเชิงพื้นที่ที่จำกัดจำนวนวัตถุที่คาดการณ์ต่อเซลล์กริด

สถาปัตยกรรม YOLO V2 (Redmon and Farhadi, 2017) พัฒนาต่อจาก YOLO เนื่องจาก YOLO ได้รับข้อผิดพลาดในการแปลงการคาดการณ์การเรียกคืนที่ต่ำดังนั้นจึงมีการพัฒนา

มาเป็น YOLO V2 ซึ่งสถาปัตยกรรมนี้จะมีการปรับปรุงตามโมเดลก่อนหน้านี้ที่พัฒนามาอย่าง SSD (Liu et al., 2016) เพื่อแก้ไขจุดบกพร่องของข้อจำกัดของตัวโมเดล ซึ่งผู้พัฒนาต้องการให้ YOLO เร็วขึ้น และ แม่นยำ YOLO V2 มีการปรับเปลี่ยนแปลงมาตรฐานของแบทช์ (Batch normalization) มีหน้าที่ทำให้แต่ละชั้นเลเยอร์ ในโครงข่ายประสาทเทียม สามารถเรียนรู้ได้ด้วยตัวเอง ลดภาระการทำงานเลเยอร์ที่ไม่จำเป็นออก YOLO V2 มีการแก้ไขปัญหาโดยเริ่มการฝึกลักษณะนามบทความละเอียด 224×224 ก่อนหลังจากนั้นจะปรับเปลี่ยนคุณลักษณะของรูปภาพบทความละเอียด 448×448 แทนเพื่อให้รูปภาพมีคุณภาพมากขึ้นซึ่งหลังการฝึกทำให้โครงข่ายปรับให้เข้ากับตัวรับข้อมูลความละเอียดสูงขึ้นมา จากนั้นปรับแต่งโครงข่ายผลลัพธ์ในการตรวจจับการปรับแต่งลักษณะนามอย่างละเอียดบนตัวรับข้อมูลความละเอียดสูงจะเพิ่มความแม่นยำขึ้น 4% แสดงสถาปัตยกรรม YOLO V2 ดังภาพประกอบที่ 2.22



ภาพประกอบที่ 2.22 สถาปัตยกรรม YOLO V2 โดย (Redmon and Farhadi, 2017)

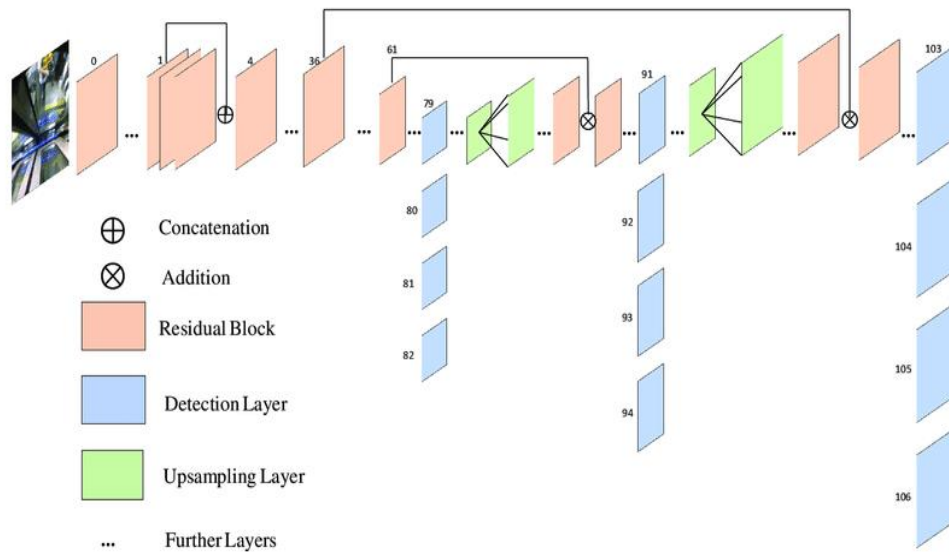
YOLO V1 ใช้จากสถาปัตยกรรม GoogleLeNet แต่ก็ขาดความแม่นยำ 88.0% mAP บน ImageNet เมื่อเทียบกับ 90.0% mAP ด้วย VGG-16 (Simonyan and Zisserman, 2014) YOLO V2 ผู้พัฒนาพยายามเพิ่มประสิทธิภาพสูงสุดโดยการพัฒนาโครงข่ายที่กำหนดเองซึ่งมีทั้งความแม่นยำและรวดเร็วเพื่อรองรับแอปพลิเคชันที่ต้องการความเร็วใกล้เคียงเรียลไทม์เช่นในยานพาหนะที่ขับเคลื่อนด้วยตนเอง นี่คือนำเสนอ Darknet-19 มีความแม่นยำ 91.2%

Darknet-19 ประกอบด้วยคอนโวลูชันเลเยอร์ 19 เลเยอร์ส่วนใหญ่ใช้ฟิลเตอร์ 3×3 ที่มีเลเยอร์การรวมสูงสุด 5 ชั้น ใช้ตัวกรอก 1×1 เพื่อลดช่องว่างระหว่างคอนโวลูชัน 3×3 ตามภาพประกอบที่ 2.23

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

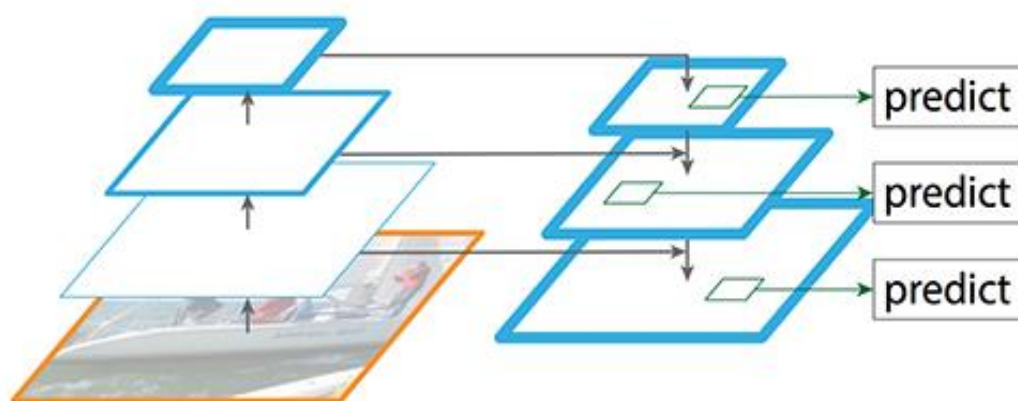
ภาพประกอบที่ 2.23 โครงสร้างของ DarkNet-19 โดย (Redmon and Farhadi, 2017)

สถาปัตยกรรม YOLO V3 ได้รับแรงบันดาลใจจากสถาปัตยกรรมของคุณสมบัติ ResNet (He et al., 2016) และ FPN (Feature-Pyramid Network) Separator YOLO V3 เรียกว่า Darknet-53 (พร้อม 52 ชั้นคอนโวลูชัน) รวมถึงการเชื่อมต่อกับพีรามิดที่คาดการณ์ไว้ FPN แต่ละภาพประมวลผลการบีบอัดเชิงพื้นที่ที่แตกต่างกัน โครงสร้าง YOLO V3 จากภาพประกอบที่ 2.24



ภาพประกอบที่ 2.24 สถาปัตยกรรม YOLO V3 โดย (Dai et al., 2020)

Feature-Pyramid ทำหน้าที่ลดขนาดแผนที่คุณลักษณะจะค่อย ๆ ลดลงในมิติเชิงพื้นที่ลงก่อนจะนำคุณลักษณะกลับไปลดภาพขนาดใหญ่ไปเล็ก และ เล็กไปใหญ่จะขึ้นนอกจากนี้ยังมีกระบวนการส่งข้ามกันระหว่างสเกลด้วยกันด้วยซึ่งก็คือ FPN (Feature-Pyramid Network) นั้นอิงแนวความคิดว่าการนำเอาข้อมูลจากภาพใหญ่ แต่ได้จากการสร้างขาขึ้น นั้นทำให้ได้ข้อมูลที่มีความละเอียดสูงในภาพขนาดใหญ่ FPN สามารถดึงคุณลักษณะได้ทุกขนาด YOLO V3 ใช้ตัวกรองแบบคอนโวลูชันบนแผนที่ที่ฝสานเพื่อสร้างการคาดการณ์ ทำซ้ำ 2 อีกครั้งเพื่อให้เลเยอร์แผนที่คุณลักษณะที่ได้มีข้อมูลโครงสร้างระดับที่ดี และข้อมูลรูปภาพที่มีความละเอียดดีเกี่ยวกับตำแหน่งของวัตถุ (Redmon and Ali, 2018) ตัวอย่างภาพประกอบที่ 2.25



ภาพประกอบที่ 2.25 โครงสร้างของ Feature-Pyramid โดย (Lin et al., 2016)

จากภาพประกอบที่ 2.25 กระบวนการ FPN ช่วยให้ YOLO V3 เรียนรู้วัตถุในขนาดต่าง ๆ กลุ่มการตรวจจับ มีขนาดที่กว้างขึ้น และ มีความละเอียดต่ำกว่าเมื่อเทียบกับกลุ่มการตรวจจับอื่น ๆ ดังนั้นจึงมีคุณภาพในการตรวจจับวัตถุขนาดใหญ่ในขณะที่กลุ่มมีคุณภาพในการตรวจจับวัตถุขนาดเล็ก ในส่วนของหัวจำแนกการตรวจจับแต่ละหัวมีกระบวนการตรวจจับแยกต่างหาก

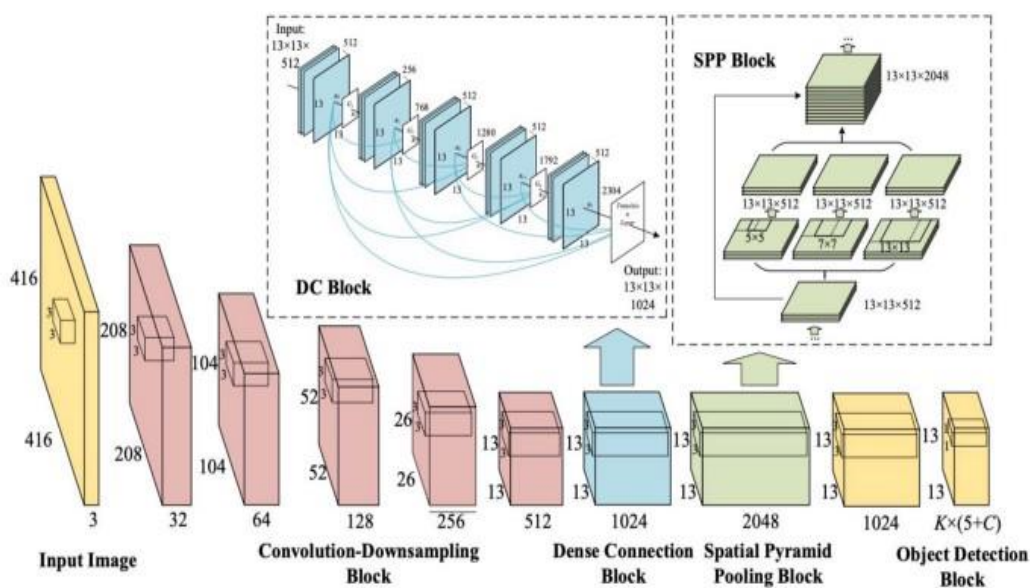
โดยที่กระบวนการตรวจจับวัตถุของตัว YOLO ก็เหมือนกับการสถาปัตยกรรมก่อน ๆ จะเริ่มขั้นตอนแรกจากการแบ่งส่วนรูปภาพประกอบที่ จะวิเคราะห์ออกเป็นเซลล์กริด $S \times S$ ที่จะกำหนดฐานไว้ $S = 7$ ในแต่ละ Grid Cell ต่อมาก็เข้าสู่กระบวนการใช้กล่องขอบเขตในการทำนายเป็นจำนวนเท่ากับ 2 ($B = 2$) จากนั้นจะเข้าสู่กระบวนการคำนวณประเภทของวัตถุที่ติดรอบกล่องทั้งหมด ขอบเขตของกล่องที่จะทำนายนั้น จะมีค่าพารามิเตอร์คือ (p, x, y, a, b) โดย x, y แทนพิกัดจุดกลางของความกว้างของเขตกรอบกล่องในแกน x พิกัดกลางความสูงของขอบเขตกล่องในแนวแกน y (a, b) แทน (ความกว้างของขอบเขตกล่อง, ความสูงของขอบเขตกล่อง) และ p แทน ค่าความเชื่อมั่น จะเป็นการวัดขอบกล่องที่จะทำนายวัตถุของ C โดยการวัดค่าระหว่าง $0 \leq p \leq 1$ และใช้ค่า $p > 0.5$ เท่านั้นจึงจะสามารถอธิบายได้ว่าสิ่งที่สนใจในกล่องได้

	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	$3 \times 3 / 2$	128×128
1x	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	$3 \times 3 / 2$	64×64
2x	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	$3 \times 3 / 2$	32×32
8x	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3 / 2$	16×16
8x	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024	$3 \times 3 / 2$	8×8
4x	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

ภาพประกอบที่ 2.26 โครงสร้างของ DarkNet-53 โดย (Redmon and Farhadi, 2018)

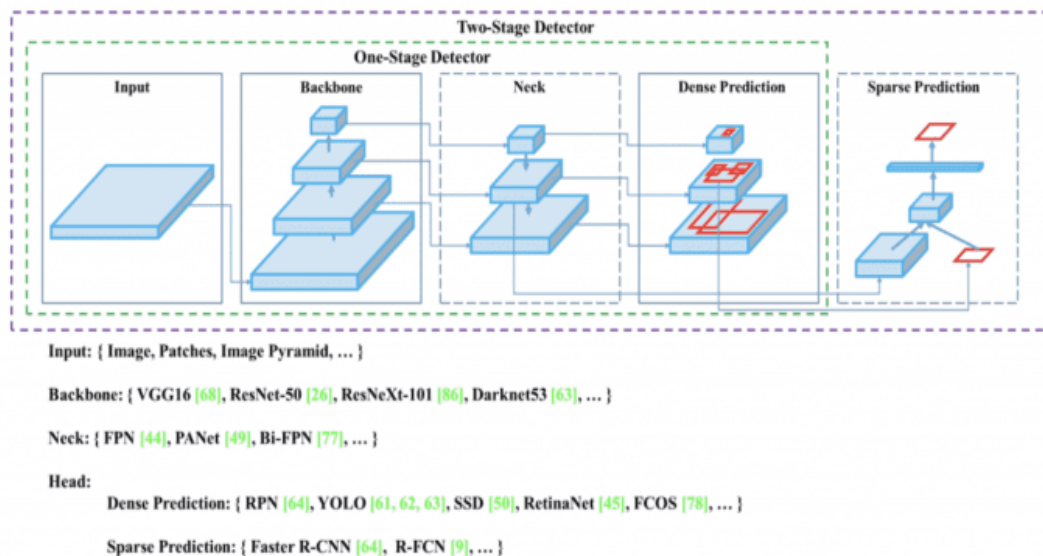
ประสิทธิภาพ YOLO V3 เมตริก COCO AP ของ YOLO V3 เทียบเท่ากับ SSD แต่เร็วกว่า 3 เท่า แต่ AP ของ YOLO V3 ยังคงอยู่เบื้องหลัง RetinaNet (Lin et al., 2017) โดยเฉพาะอย่างยิ่ง AP@ ไอโอยู = 0.75 ลดลงอย่างมีนัยสำคัญเมื่อเทียบกับ RetinaNet ซึ่งแสดงให้เห็นว่า YOLO V3 มีข้อผิดพลาดในการแปลที่สูงกว่า YOLO V3 ยังแสดงให้เห็นถึงการปรับปรุงที่สำคัญในการตรวจจับวัตถุขนาดเล็ก

สถาปัตยกรรม YOLO V4 พัฒนาโดยทีมวิจัย Bochkovskiy และ คณะ (Bochkovskiy et al., 2020) เริ่มจาก YOLO V1 - YOLO V3 เป็นโครงข่ายการตรวจจับวัตถุแรกที่มีรวมปัญหาในการวาดกล่องขอบเขต และการระบุป้ายกำกับคลาสในโครงข่ายที่แตกต่างแบบ end-to-end ต่อมา YOLO V2 ได้ทำการปรับปรุงซ้ำ ๆ หลายครั้งนอกเหนือจาก YOLO รวมถึงความละเอียดสูงขึ้น และ กล่องขอบเขตที่มีความเชื่อมั่นที่สูงขึ้น YOLO V3 สร้างขึ้นจากสถาปัตยกรรมก่อนหน้าโดยการเพิ่มคะแนนความเที่ยงธรรมในการทำนายกล่องขอบเขตเพิ่มการเชื่อมต่อกับเลเยอร์โครงข่าย backbone และ ทำการคาดการณ์ในระดับความละเอียดแยกกันสามระดับเพื่อปรับปรุงประสิทธิภาพบนวัตถุขนาดเล็ก 3 สถาปัตยกรรมได้ทำการอธิบายไปแล้วก่อนหน้านี้ ต่อมาเป็นสถาปัตยกรรมที่พัฒนาต่อจาก YOLO V3 คือ YOLO V4 เป็นเครื่องตรวจจับแบบขั้นตอนเดียวที่มีความรวดเร็วละเอียดและแม่นยำ (Cai, et al., 2021) ดังภาพประกอบที่ 2.27



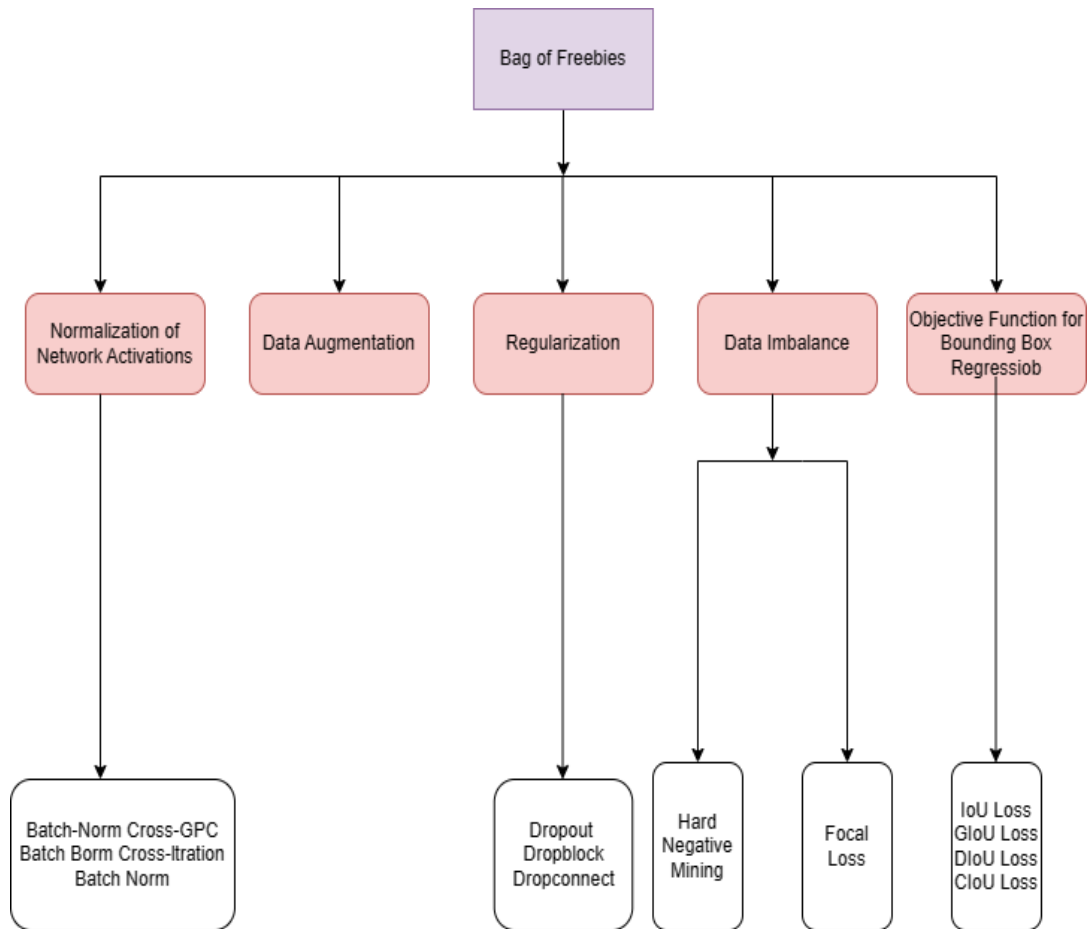
ภาพประกอบที่ 2.27 สถาปัตยกรรม YOLO V4 โดย (Huang, et al., 2020)

ส่วนประกอบหลักของ YOLO V4 ที่ใช้ก็เหมือน YOLO V3 สถาปัตยกรรมก่อนหน้าซึ่งใช้งาน CSPDarknet-53 ที่แตกต่างกันดังภาพประกอบที่ 2.28



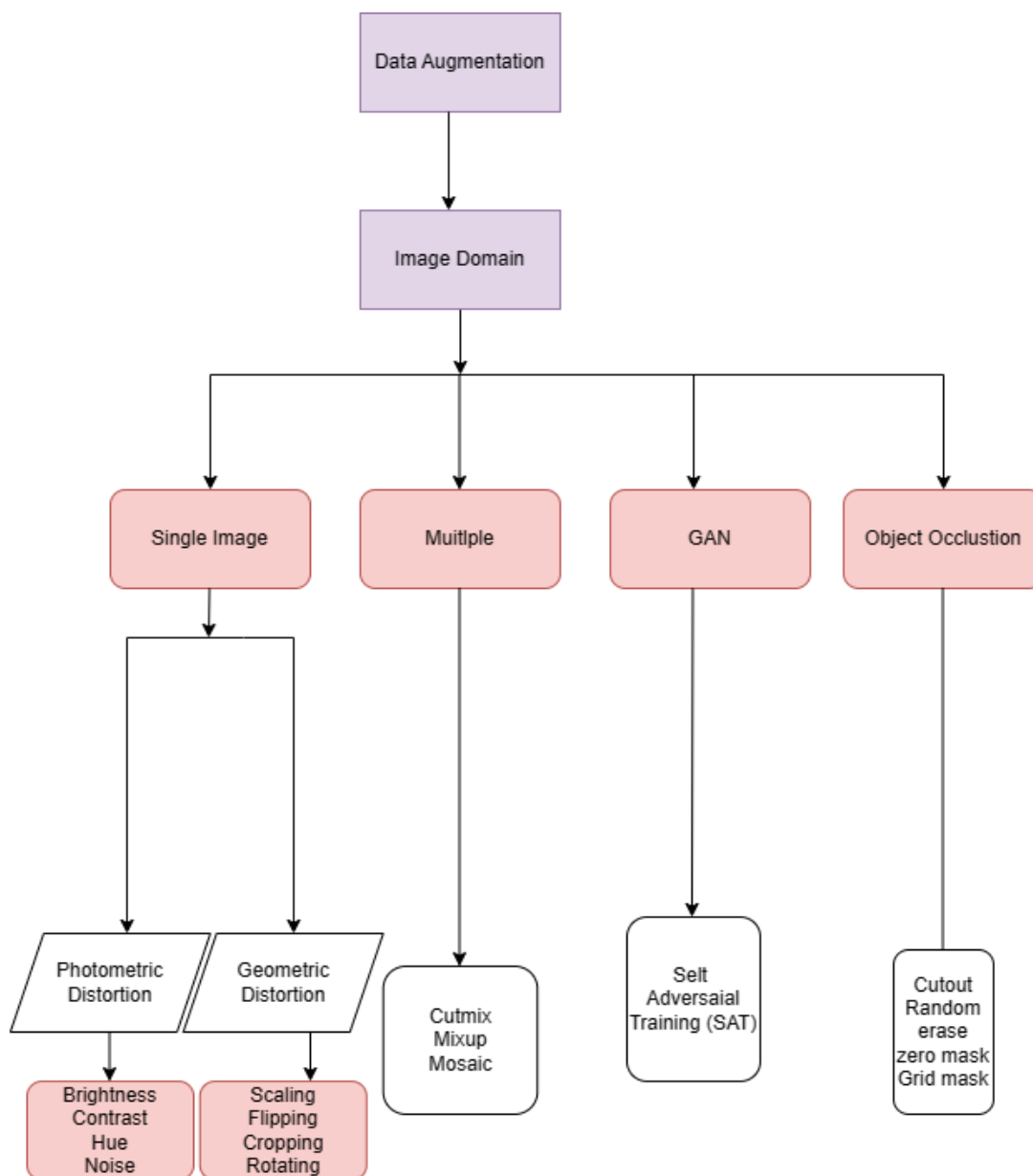
ภาพประกอบที่ 2.28 Object Detection Workflow โดย (Bochkovskiy et al., 2020)

ทางด้านส่วนของ Backbone YOLO V4 สถาปัตยกรรมการเรียนรู้เชิงลึกที่โดยทั่วไปทำหน้าที่เป็นตัวแยกคุณลักษณะ YOLO V4 มีกระบวนการที่เรียกว่า Bag of freebies เป็นกระบวนการที่เปลี่ยนกระบวนการฝึกอบรมจึงทำให้ลดประหยัดเวลาที่ใช้ในการฝึกอบรมเพื่อปรับปรุงค่าความแม่นยำของโมเดล ตามดังภาพประกอบที่ 2.29



ภาพประกอบที่ 2.29 Bag of freebies

กระบวนการฝึกอบรมที่ใช้ในการเสริมข้อมูลตั้งดังภาพประกอบที่ 2.30 วัตถุประสงค์ของการเพิ่มข้อมูลคือการเพิ่มปริมาณข้อมูลเพื่อให้เกิดความหลากหลายของข้อมูลที่มีอยู่แล้ว และ ยังลดเวลาในการหาข้อมูลเพิ่มจากภายนอก (Muhammad et al., 2022)



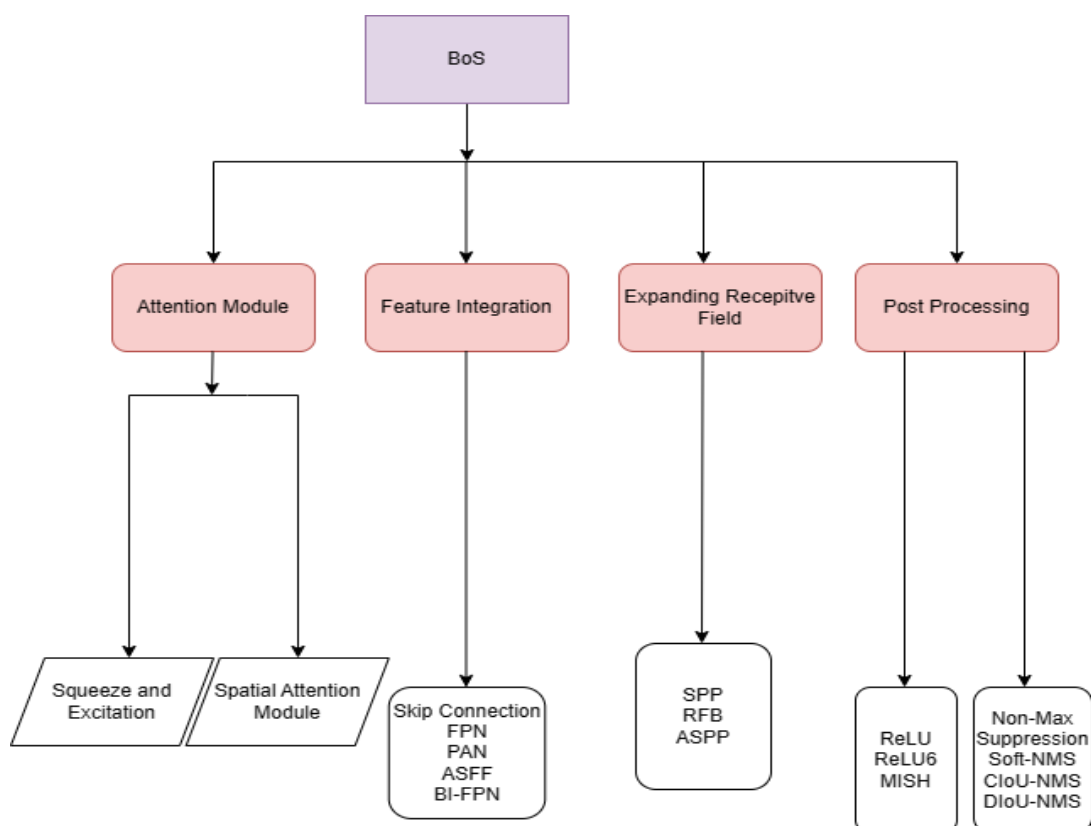
ภาพประกอบที่ 2.30 กระบวนการเพิ่มข้อมูล

การเพิ่มปริมาณข้อมูลที่มีหลากหลายแบบในดังภาพประกอบที่ 2.31 ได้ใช้กับโมเดล YOLO V4 เพิ่มชุดการฝึกอบรมโดยการเพิ่มของชุดข้อมูลโดยใช้ข้อมูลที่มีอยู่ ซึ่งรวมถึงการเปลี่ยนแปลงลักษณะในชุดข้อมูลหรือใช้การเรียนรู้เชิงลึกเพื่อสร้างจุดข้อมูลใหม่ ตัวอย่าง เช่น การกลับด้านของรูปภาพ, การเพิ่มแสงสว่าง, การกลับหัวรูปภาพ เพื่อเพิ่มความแตกต่างในชุดของข้อมูล (Sowmya and Radha, 2021) ตัวอย่างดังภาพประกอบที่ 2.31



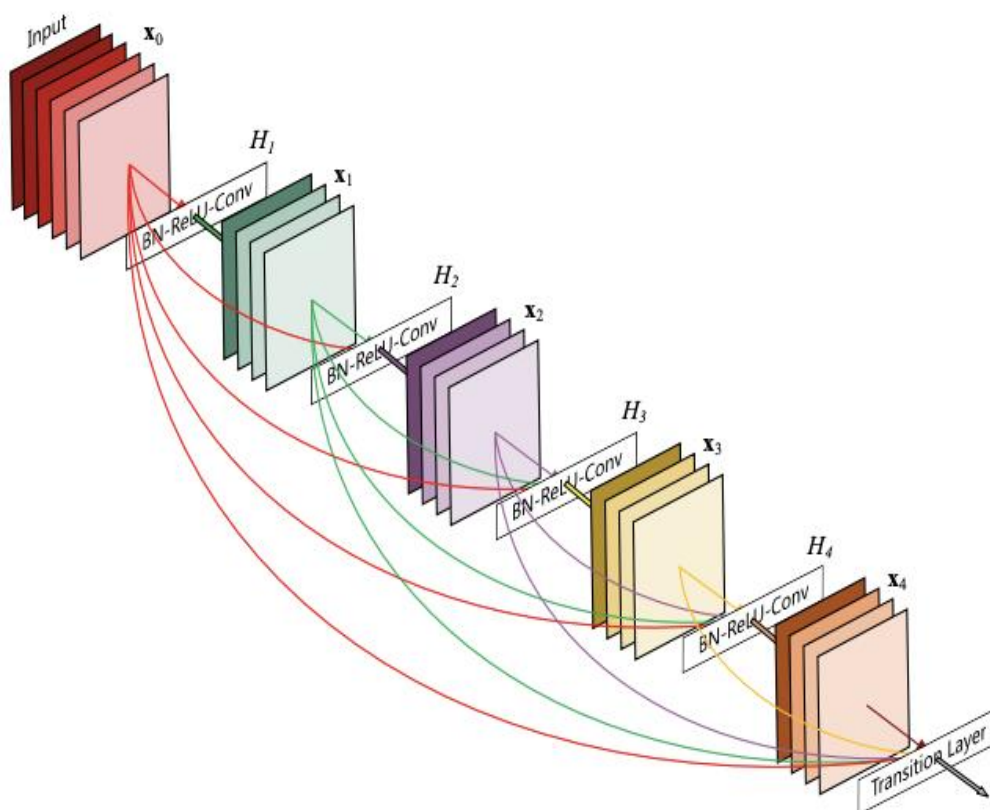
ภาพประกอบที่ 2.31 ตัวอย่างภาพประกอบที่ ทำกระบวนการ data augmentation

หลังจาก BoF ต่อมาคือ Bag of Specials หรือ BoS เป็นวิธีหลังการประมวลผล และการสร้างความแตกต่างเพียงหลังเล็กน้อยหรือการใช้อนุมานเพียงเล็กน้อย แต่เพิ่มคุณภาพได้มาก ซึ่งกระบวนการเหล่านี้มีไว้สำหรับการเพิ่มคุณลักษณะเฉพาะในโมเดลออกมา เช่น การใช้ Spatial Pyramid Pooling (SPP) ที่ใช้สำหรับการขยายรูปภาพเพื่อหาจุดสำคัญในรูปภาพ ภาพประกอบที่ 2.32 เป็นแผงผังที่แสดงตัวเลือกในการใช้โมดูล หลังการประมวลผลในกระบวนการทำงานของ YOLO V4 (Karlita et al., 2022)



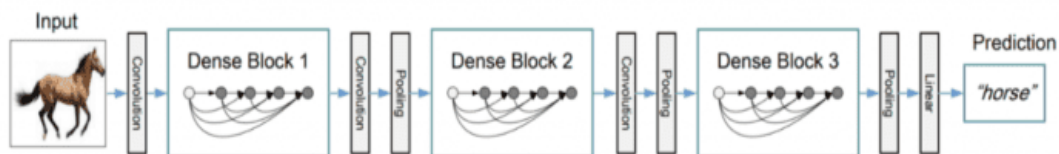
ภาพประกอบที่ 2.32 Bag of Specials

สถาปัตยกรรม YOLO V4 ใช้ backbone ใหม่เป็น CSPDarkNet-53 (Chen et al., 2021 ; Wang and Wang, 2021) ใช้ DenseNet ในการตรวจจับวัตถุ CSP (Cross-Stage Partial Connection) หลังจากการพัฒนาตัวของ DarkNet-53 ได้ใช้สถาปัตยกรรม DenseNet แต่ตัวของ DarkNet-53 มีการดัดแปลงกระบวนการของ DenseNet จะประกอบไปกับการผ่านกล่องคอนโวลูชันหลายกล่องในกระบวนการเรียกว่า Dense Block และประกอบไปด้วยเลเยอร์ของคอนโวลูชันตามดังภาพประกอบที่ 2.33



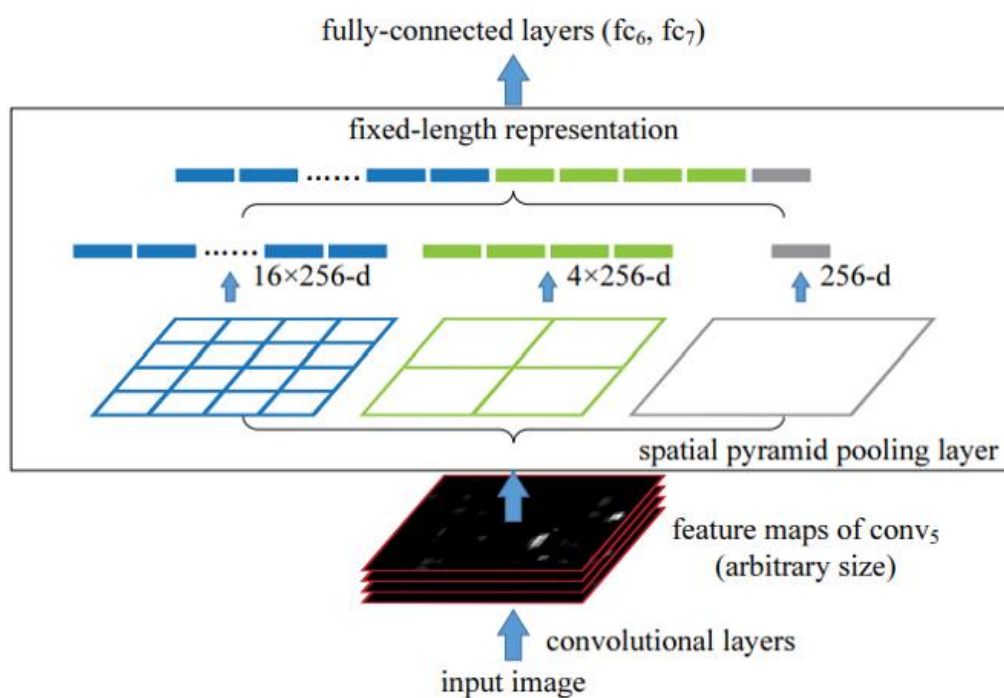
ภาพประกอบที่ 2.33 DenseNet โดย (Haque et al., 2021)

ดังภาพประกอบที่ 2.33 จะมีคอนโวลูชันเลเยอร์ที่มีตัวแปร H_i ซึ่งหลักการทำงานเหมือนโครงข่ายประสาทเทียมแบบสังวัตนาการ คือ มีแบทช์ และ ReLU H_i รูปภาพด้านบนจะแสดงแผนที่คุณลักษณะสี่แบบ ดังนั้นในแต่ละเลเยอร์จำนวนแผนที่คุณลักษณะจะเพิ่มขึ้นสี่ตัวในรูปภาพประกอบที่แสดง การเชื่อมต่อกันทั้ง 4 เลเยอร์จะช่วยเพิ่มประสิทธิภาพให้กับตัวโมเดล ดังภาพประกอบที่ 2.34 จะแสดง สถาปัตยกรรม DenseNet แบบ end-to-end



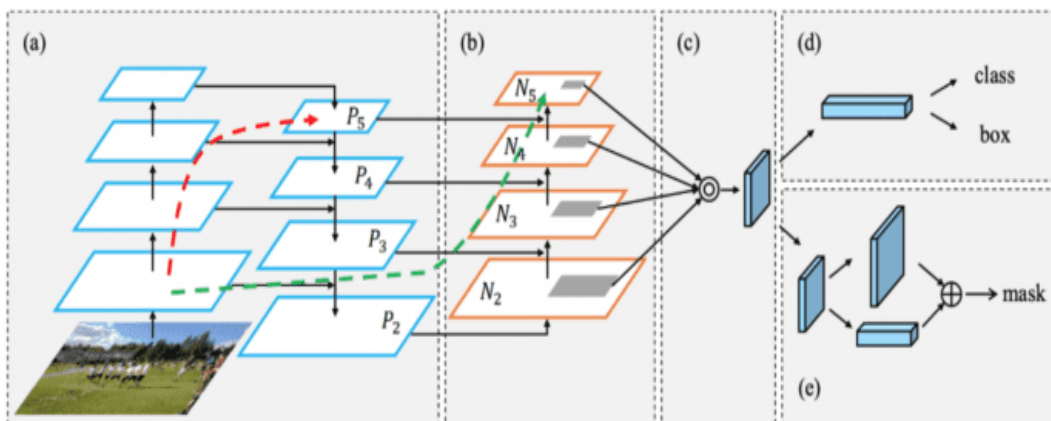
ภาพประกอบที่ 2.34 สถาปัตยกรรม DenseNet แบบ end-to-end โดย (Haque et al., 2021)

หลักการทำงานของ Spatial Pyramid Pooling จะทำการรับตัวรับข้อมูล ที่เป็นคุณลักษณะสำคัญที่สนใจของภาพ ตัวกรอกเป็นรูปแบบรายละเอียดที่ 256 ซึ่งเป็นแผนที่คุณลักษณะที่ทำหน้าที่ลดขนาดรูปภาพให้เล็กลงเพื่อสกัดคุณสมบัติคุณลักษณะเด่นวัตถุในรูปภาพที่นำเข้ามา Spatial ที่กำหนด เช่น ปรับขนาดข้อมูล 1×1 ต่อมา ปรับขนาดข้อมูล 2×2 และ จะทำการแบ่งภาพออกเป็น 4 ส่วน หรือ ปรับขนาดข้อมูล 3×3 เป็นการแบ่งภาพเป็น 16 ส่วนสำหรับ 3×3 จากนั้นตัวประมวลผลผลลัพธ์ไปยังด้านผลลัพธ์ SPP (Tai et al., 2020) ดังภาพประกอบที่ 2.35



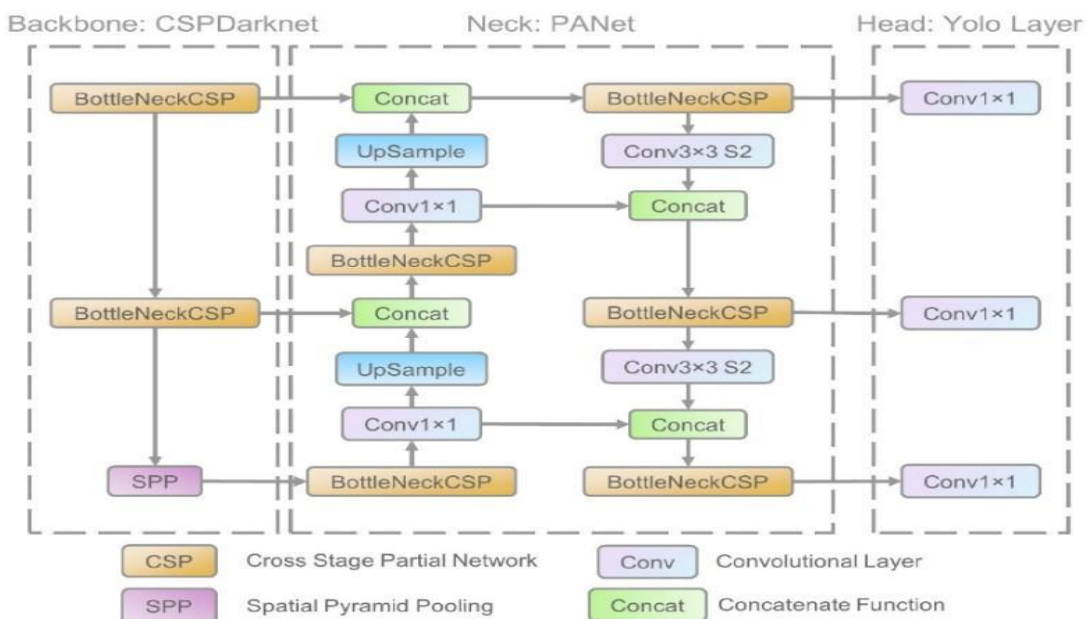
ภาพประกอบที่ 2.35 โครงสร้างของ Spatial Pyramid Pooling โดย (He et al., 2015)

Path Aggregation Network หรือ PANet มีหน้าที่เพิ่มอัตราการนำเข้าข้อมูลให้มีคุณภาพมากขึ้นตามอินสแตนซ์ตามที่กำหนด ซึ่งมีการพัฒนาการจับคุณลักษณะรูปภาพที่ดีขึ้นเพิ่มเส้นทางจากล่างขึ้นบนที่ด้านบนของ FPN (Liu et al., 2018) ดังภาพประกอบที่ 2.36



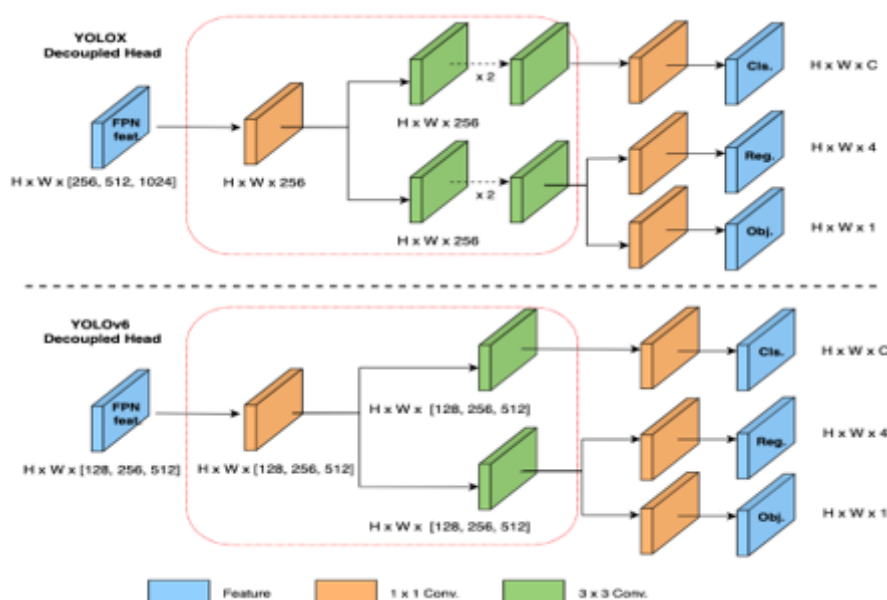
ภาพประกอบที่ 2.36 (a) backboneของ FPN (b) การเพิ่มเส้นทางจากล่างขึ้นบนสำหรับ PAN (c) การรวมคุณลักษณะที่ปรับเปลี่ยนได้ (d) กล่องแยกคลาส (e) เลเยอร์เชื่อมต่ออย่างสมบูรณ์ โดย (Liu et al., 2018)

สถาปัตยกรรม YOLO V5 ได้รับการพัฒนาโดย Glenn Jocher ใช้งาน YOLO V3 โดย PyTorch สถาปัตยกรรม YOLO V5 ทั้งหมดประกอบด้วยส่วนประกอบ 3 ส่วนเดียวกัน: CSP-Darknet-53 เป็น backbone, SPP และ PANet ที่คอ และ หัวที่ใช้ใน YOLO V4 และ YOLO V5 ใช้ CSP-Darknet-53 เป็นแกนหลัก CSP-Darknet-53 เป็นโครงข่ายที่ซับซ้อน Darknet-53 ที่ใช้เป็นแกนหลักสำหรับ YOLO V3 ซึ่งโครงข่ายความคล้ายคลึงกันระหว่าง 2 สถาปัตยกรรม คือ CSP backbone, PANet และ ส่วนเสริมข้อมูล YOLO V5 ประสิทธิภาพได้รับการปรับปรุง และ พิสูจน์แล้วว่าความแม่นยำจะต่ำ แต่ความเร็วของ YOLO V5 นั้นสูงอย่างมาก สิ่งนี้ทำให้สามารถใช้ CPU ได้อย่างเต็มประสิทธิภาพ (Jiang et al., 2022; Xu et al., 2021) ภาพสถาปัตยกรรมที่ 2.37



ภาพประกอบที่ 2.37 สถาปัตยกรรม YOLO V5 โดย (Xu et al., 2021)

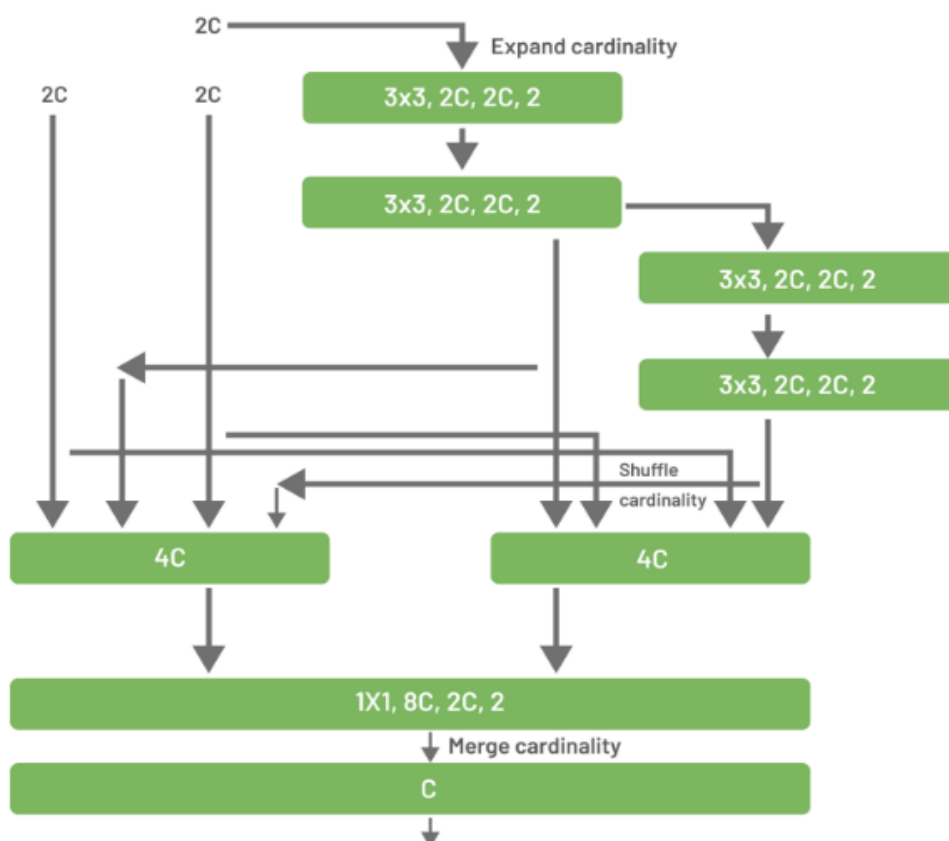
RepBlock โดยชั้น RepConv ใช้แปลงตำแหน่งช่องภาพมิติ และ ปรับแต่งออกแบบ SimSPPF (Sun and Chen, 2022) ได้ดีขึ้นสำหรับนำภาพมารวมพูลสูงสุดและส่งออกด้านผลลัพธ์โดย SimSPPF สามารถลดการสูญเสียคุณสมบัติคุณลักษณะเด่นของภาพจนได้การลดข้อมูลพารามิเตอร์ให้มีอย่างมีประสิทธิภาพในส่วนของ Rep-PAN ได้มีการพัฒนาเพื่อให้การอนุมานฮาร์แวร์มีประสิทธิภาพมากขึ้น เพื่อให้ได้ความแม่นยำและ ความเร็วในการเทรนโมเดลตามภาพประกอบด้านบน หลังจากนั้น YOLO V6 ได้ใช้ Decoupled Head for Efficiency ที่นำโครงสร้างหัวมาแยกเพื่อนำมาปรับปรุงการออกแบบ โดยปกติหัวของ YOLO V5 ทำได้โดยการแบ่งแยกการจำแนกประเภทภาพประกอบที่ 2.39 แสดงให้เห็นว่าหัวของ YOLOX (Zhang et al., 2022) ได้สร้างหัวในการแยกประเภทออกมากด้วยการใช้คอนโวลูชันเลเยอร์ขนาด 3×3 และ เพิ่มเป็นสองชั้นช่วยในการเพิ่มความแม่นยำในการตรวจจับข้อเสียคือความล่าช้าในการฝึกอบรมโมเดลขณะนั้นตัวของ YOLO V6 ออกแบบให้แบ่งหัวแยกส่วนให้มีประสิทธิภาพมากขึ้นให้ได้ค่าสมดุระหว่างการทำหนดลักษณะของงานที่ต้องการตรวจจับและการทำงานของฮาร์แวร์



ภาพประกอบที่ 2.39 แผนภาพโครงสร้างหัวแยกที่มีประสิทธิภาพ โดย (Gupta et al., 2023)

สถาปัตยกรรม YOLO V7 (Wang et al., 2022) ได้รับการพัฒนาให้เพิ่มความเร็วและความแม่นยำในการตรวจจับมากขึ้นโดยใช้สถาปัตยกรรมที่คล้ายกับสถาปัตยกรรม Scaled YOLO V4 (Wang et al., 2021) โดย backbone ของ YOLO V7 ไม่ได้ใช้การฝึกอบรมล่วงหน้ากับชุดข้อมูลของ ImageNet แต่ใช้กับชุดข้อมูล COCO (Lin et al., 2014) แทนสถาปัตยกรรม YOLO V7 ใช้ E-ELAN (Extended Efficient Layer Aggregation Network) และ ปรับขนาดโมเดลเพิ่มเติมการปรับคอนโวลูชันพารามิเตอร์ใหม่สถาปัตยกรรม E-ELAN เป็นบล็อกที่คำนวณในกระบวนการทำงานของ backbone ของ E-ELAN ได้รับการออกแบบมาเพื่อเพิ่มประสิทธิภาพและความแม่นยำของ YOLO V7 โดยการนำเทคนิคการรวมเลเยอร์รูปแบบใหม่ที่ช่วยลดความซับซ้อนในการคำนวณ

ของแบบจำลองในขณะที่ยังคงมีความแม่นยำสูง E-LAN ใช้การบิดเบือนเชิงลึกแบบไฮบริด และบล็อกความสนใจของช่องสัญญาณที่เรียกว่าบล็อก E-LAN บล็อก E-LAN ประกอบด้วยชั้นคอนโวลูชันเชิงลึกตามด้วยเลเยอร์ความสนใจของชั้นคอนโวลูชัน เชิงลึกใช้เพื่อลดจำนวนพารามิเตอร์ในแบบจำลองในขณะที่ยังคงความสนใจของช่องถูกใช้เพื่อเพิ่มพลังการเลือกปฏิบัติของโมเดลโดยการถ่วงน้ำหนักแผนที่คุณลักษณะตามความสำคัญ ตามภาพประกอบที่ 2.40



ภาพประกอบที่ 2.40 E-LAN (Extended Efficient Layer Aggregation Network)

โดย (Muhammad Rizwan Munawar, 2023)

Lead Head Guided Label Assigner และ Coarse-to-Fine Lead Head Guided Label Assigner เป็นสองวิธีที่แตกต่างกันในการเรียนรู้แบบมัลติทาสก์ ซึ่งเป็นการเรียนรู้ของเครื่องประเภทหนึ่งที่โมเดลเดียวได้รับการฝึกอบรมให้ทำงานหลายอย่างพร้อมกันในการเรียนรู้แบบหลายงานงานที่ต่างกันอาจมีระดับความซับซ้อนที่ต่างกันและต้องการข้อมูลในปริมาณที่ต่างกันเพื่อให้ได้ประสิทธิภาพที่ดี Lead Head Guided Label Assigner และ Coarse-to-Fine Lead Head Guided Label Assigner เป็นสองวิธีในการแก้ไขปัญหาใน Lead Head Guided Label Assigner แต่ละงานจะมีหัวส่งออกของตัวเองซึ่งมีหน้าที่ในการทำนายป้ายกำกับงานที่เกี่ยวข้อง ในระหว่างการฝึกอบรมการสูญเสียของแต่ละหัวจะถูกคำนวณอย่างอิสระและการไล่ระดับสีจะถูกย้อนกลับผ่านหัวที่เกี่ยวข้องเท่านั้น วิธีการนี้ถือว่าการทั้งหมดมีความสำคัญเท่าเทียมกันและ

แบบจำลองจะให้ความสนใจกับแต่ละงานอย่างเท่าเทียมกันในทางกลับกัน Coarse-to-Fine Lead Head Guided Label Assigner เป็นวิธีการที่ซับซ้อนมากขึ้น ในวิธีนี้งานจะถูกจัดระเบียบเป็นลำดับชั้นตามความซับซ้อนและงานที่ด้านบนสุดของลำดับชั้นถือเป็นงานต้นในขณะที่ยานที่ด้านล่างถือเป็นงานที่ดี แบบจำลองได้รับการฝึกฝนให้ทำนายงานต้นก่อนจากนั้นจึงทำผลลัพธ์ที่ดี ในระหว่างการฝึกอบรมการไล่ระดับสีจะถูกย้อนกลับจากงานละเอียดไปยังงานต้นทำให้แบบจำลองสามารถเรียนรู้คุณสมบัติที่มีประโยชน์สำหรับงานต้น และ ละเอียด Coarse-to-Fine Lead Head Guided Label Assigner ได้รับการแสดงเพื่อปรับปรุงประสิทธิภาพของโมเดลการเรียนรู้แบบหลายงานโดยเฉพาะอย่างยิ่งเมื่องานมีความซับซ้อนในระดับที่แตกต่างกัน อย่างไรก็ตามต้องใช้ความรู้ล่วงหน้าเกี่ยวกับงานและความสัมพันธ์ซึ่งอาจไม่สามารถใช้ได้เสมอไป Lead Head Guided Label Assigner เป็นวิธีการที่ง่ายกว่าซึ่งไม่จำเป็นต้องมีความรู้ล่วงหน้าเกี่ยวกับงาน แต่อาจไม่มีประสิทธิภาพเท่ากับ Coarse-to-Fine Lead Head Guided Label Assigner (Wang et al., 2023)

สถาปัตยกรรม YOLO V8 ซึ่งเป็นการพัฒนาล่าสุดของอัลกอริธึม YOLO เหนือกว่ารุ่นก่อนด้วยชุดการปรับปรุงต่าง ๆ โดยนำเสนอความสนใจเชิงพื้นที่การผสมผสานคุณลักษณะ และ โมดูลการรวมบริบทซึ่งร่วมกันขับเคลื่อนประสิทธิภาพที่ได้รับการปรับปรุงการปรับปรุงเหล่านี้ไม่เพียงเพิ่มความแม่นยำ แต่ยังเร่งการตรวจจับวัตถุอีกด้วย คุณสมบัติที่สำคัญของ YOLO V8 ความแม่นยำที่เพิ่มขึ้น YOLO V8 ยกระดับความแม่นยำในการตรวจจับวัตถุให้เหนือกว่ารุ่นก่อนโดยใช้เทคนิคและการเพิ่มประสิทธิภาพแบบใหม่ ความเร็วที่เร่งความเร็ว YOLO V8 มีความเร็วอนุมาณที่เร็วกว่าเมื่อเทียบกับโมเดลการตรวจจับวัตถุของคู่แข่ง ในขณะที่ยังคงความแม่นยำในระดับสูง backbone ที่หลากหลาย YOLO V8 รองรับ backbone ที่หลากหลายรวมถึง EfficientNet, ResNet และ CSPDarknet ทำให้ผู้ใช้มีอิสระในการเลือกรุ่นที่เหมาะสมที่สุดสำหรับการใช้งานเฉพาะของตน การฝึกอบรมแบบปรับเปลี่ยนได้ YOLO V8 ใช้กลยุทธ์การฝึกอบรมแบบปรับเปลี่ยนได้ซึ่งปรับแต่งอัตราการเรียนรู้และปรับสมดุลฟังก์ชันที่สูญเสียไปตลอดกระบวนการฝึกอบรม ส่งผลให้โมเดลมีประสิทธิภาพเหนือกว่า การเพิ่มข้อมูลที่ทันสมัย YOLO V8 ใช้เทคนิคการเพิ่มข้อมูลขั้นสูง เช่น MixUp และ CutMix เพื่อเสริมความแข็งแกร่งของโมเดลและเพิ่มขีดความสามารถทั่วไป สถาปัตยกรรมที่ปรับแต่งได้ สถาปัตยกรรมของ YOLO V8 สามารถปรับเปลี่ยนได้เป็นพิเศษ ทำให้ผู้ใช้สามารถปรับแต่งโครงสร้างและพารามิเตอร์ของโมเดลได้อย่างง่ายดายเพื่อให้สอดคล้องกับความต้องการของผู้ใช้ โมเดลที่ได้รับการฝึกอบรมล่วงหน้า YOLO V8 นำเสนอโมเดลที่ได้รับการฝึกอบรมล่วงหน้าพร้อมใช้งานเพื่อการปรับใช้และถ่ายโอนการเรียนรู้ที่ราบรื่นไปยังชุดข้อมูลต่าง ๆ คุณสมบัติหลักเหล่านี้ทำให้ YOLO V8 เป็นตัวเลือกที่น่าเกรงขามในด้านการตรวจจับวัตถุ การผสมผสานความแม่นยำ ความเร็ว และความสามารถในการปรับตัว เพื่อตอบสนองความต้องการที่หลากหลาย (Reis et al., 2023; Li et al., 2023)

2.8 งานวิจัยที่เกี่ยวข้อง

ในการศึกษาการตรวจจับวัตถุที่เหมาะสม เบื้องต้น เริ่มต้นด้วยการหาปัญหาในปัจจุบันบนทางจราจรของผู้ขับขี่หรือจากปัจจัยต่าง ๆ ที่เกี่ยวข้องกัน เพื่อตรวจสอบการแก้ไขปัญหาเบื้องต้นการป้องกันการเกิดปัญหาทางจราจรที่ทำงานได้อย่างรวดเร็ว และ ลดภาระในการหาข้อมูลยานพาหนะที่ต้องการค้นหา เพื่อจับกุม หรือ แจ้งเตือนโดยผู้วิจัยได้ทำการศึกษางานวิจัยที่ผ่านโดยใช้ตัวโมเดล YOLO ในการตรวจจับที่เหมาะสมกับข้อมูลในพื้นที่รวมทั้งข้อมูลในด้านโครงข่ายประสาทเทียมแบบสังวัตนาการ และ การถ่ายโอนการเรียนรู้

ตรวจจับยานพาหนะทางอากาศโดยทีมผู้วิจัย สารibas และคณะ (Saribas et al., 2018) โมเดล YOLO V2 สำหรับการตรวจจับยานพาหนะ ชุดข้อมูลภาพโดยใช้ควอดคอปเตอร์ผลลัพธ์ถูกเปรียบเทียบโดยใช้ทั้ง Extended Polyhedral Conic Classifier และ You Only Look Once (YOLO) ค่าความแม่นยำที่ได้รับมากกว่า 80%

ตรวจจับยานพาหนะทางอากาศพัฒนาโดยทีมผู้วิจัย ซู และคณะ (Xu et al., 2019) พัฒนาโมเดล YOLO V3 สำหรับการตรวจจับยานพาหนะโดยใช้ชุดข้อมูลใหม่ VEDAI ที่มาจากภาพถ่ายดาวเทียมของ Utah AGRC เปลี่ยนหมายเลขเลเยอร์การบิดของโครงข่ายที่ปรับปรุงแล้วเป็น 75 แผนที่คุณลักษณะระดับบนสุดมีข้อมูลเพิ่มเติมเกี่ยวกับเพื่อตรวจจับวัตถุขนาดเล็กเพื่อตรวจจับยานพาหนะทางอากาศได้ความแม่นยำมากขึ้นถึง 90%

การตรวจจับและติดตามวัตถุสำหรับชุมชน การเฝ้าระวังโดยใช้การถ่ายโอนการเรียนรู้โดยทีมผู้วิจัย มาชิราจู และคณะ (Machiraju et al., 2021) ได้ทำการปรับแต่ง YOLO V3 โดยใช้การถ่ายโอนการเรียนรู้สำหรับพัฒนาโมเดล และสร้างชุดข้อมูลเฉพาะสำหรับการตรวจจับวัตถุบนทางเท้าที่โดยโมเดลตรวจจับวัตถุมีความสามารถในการแยกวัตถุเล็กได้ และ พัฒนาต่อยอดไปใช้ประโยชน์สำหรับแอปพลิเคชันที่ต้องการการตรวจจับวัตถุ

เฉิน และคณะ (Chen et al., 2019) นำเสนอการตรวจจับวัตถุที่พัฒนาขึ้นโดยทั้ง SSD และ YOLO V3 ซึ่งมีส่วนร่วมตามแนวทางการเรียนรู้เชิงลึกสามแนวทางสำหรับ (i) การตรวจจับวัตถุ (ii) การติดตาม และ (iii) การประมาณระยะทางสำหรับแอปพลิเคชันการเคลื่อนที่อัจฉริยะ (ถนนจราจร และ ทางยานพาหนะไฟ) ในงานนี้

การตรวจจับผู้ใช้ทางจราจรแบบเรียลไทม์โดยใช้อัลกอริทึม YOLO โดยทีมผู้วิจัย เอโรวิช และคณะ (Ćorović et al., 2018) ได้นำเสนอการประยุกต์ใช้อัลกอริทึม YOLO V3 สำหรับการตรวจจับผู้เข้าร่วมทางการจราจรแบบเรียลไทม์ คำนวณน้ำหนักของโครงข่ายประสาทเทียมได้รับการเริ่มต้นโดยใช้โมเดลที่ได้รับการฝึกอบรมล่วงหน้าซึ่งได้รับการฝึกอบรมในชุดข้อมูล COCO โครงข่ายประสาทเทียมได้รับการฝึกอบรมเพิ่มเติมใน Berkley Deep ด้วยชุดข้อมูลเพื่อในการตรวจจับผู้เข้าร่วมการจราจรห้าประเภท: (i) ยานพาหนะ (ii) ยานพาหนะบรรทุก (iii) คนเดินเท้า (iv) ป้ายจราจร และ (v) ไฟที่ครอบคลุมสภาพอากาศ และ สภาพแสงที่แตกต่างกัน เพื่อความปลอดภัยของยานพาหนะ และ ผู้โดยสารจะไม่ตกอยู่ในอันตรายหากใช้อัลกอริทึมนี้ในการตรวจจับผู้เข้าร่วมการจราจร การทำงาน

จาง และคณะ (Zhang et al., 2018) ใช้โครงข่าย YOLO V2 ที่ให้ผลลัพธ์ที่ยอดเยี่ยมในฟิลด์การตรวจจับเป้าหมายซึ่งผ่านการฝึกอบรมสัญญาณไฟจราจร และการวางตำแหน่งสัญญาณไฟจราจร ร่วมกับโมเดลสี HSV ออกแบบอัตราส่วนของสีแดง และ สีเขียว สีเหลือง

การตรวจจับและติดตามวัตถุที่เคลื่อนไหวโดยใช้โครงข่ายประสาทเทียมการเรียนรู้เชิงลึกและตัวกรองสหสัมพันธ์โดยทีมผู้วิจัย สุปรีธ และปาติล (Supreeth and Patil, 2018) เสนอเครื่องตรวจจับพื้นหน้าแบบ Gaussian mix model (GMM) ถูกนำมาใช้สำหรับการตรวจจับวัตถุ พื้นหลังในเฟรมวิดีโอ วัตถุเคลื่อนที่ที่ตรวจพบจะถูกครอบตัด และ กำหนดให้เป็นตัวรับข้อมูลสำหรับสถาปัตยกรรมของ ImageNet-Caffe-Alexnet ซึ่งใช้ในงานนี้ การฝึกอบรมล่วงหน้าของ Alexnet สามารถจัดหมวดหมู่ได้ 1,000 คลาส ดังนั้นทั้งสามเลเยอร์นี้จึงเหมาะสำหรับงานจัดหมวดหมู่ งานการจัดหมวดหมู่ใหม่โดยใช้การถ่ายโอนการเรียนรู้เชิงลึกไปยังการตรวจจับ และ ติดตามวัตถุที่เคลื่อนไหวในงานนี้

การสร้างระบบการจำแนกยานพาหนะตามการเรียนรู้เชิงลึกโดยผู้วิจัย ยาง (Yang, 2020) ได้เสนอให้ระบบใช้อัลกอริทึม SSD ใน TensorFlow เพื่อให้บรรลุการจัดประเภท และ ตำแหน่งยานพาหนะ ตั้งแต่การรวบรวมรูปภาพ การปรับเทียบรูปภาพ การฝึกอบรมโมเดล การตรวจจับโมเดล หลายแง่มุมของการแนะนำโดยละเอียดของกระบวนการจำแนกยานพาหนะ สามารถใช้ป้ายกำกับล่วงหน้าในคำอธิบายประกอบรูปภาพเพื่อปรับปรุงประสิทธิภาพคำอธิบายประกอบระบบนี้ตระหนักถึงการสอบเทียบยานพาหนะ การฝึกอบรมโมเดล การตรวจจับ

ตารางที่ 2.1 ตารางงานวิจัยที่เกี่ยวข้อง

ชื่องานวิจัย	ผู้เขียน	ปี	คลาส	รายละเอียด	หลักการ	ผล	ภาพ	วิดีโอ
Vehicle Detection in Aerial Images Using Modified YOLO.	Xu et al	2019	Car, Human	บทความนี้เสนอปรับปรุงอัลกอริทึม YOLO V3 สำหรับการตรวจจับยานพาหนะ โดยการฝึกอบรมด้วยชุดข้อมูลใหม่ในพื้นที่	YOLO V3	Precision 89.6% AP 91.72%	ใช่	-
Car Detection in Images Taken from Unmanned Aerial Vehicles	Saribas et al.	2018	Car	ชุดข้อมูลภาพโดยใช้ Quadcopters และ ประเภทต่างๆของกล้อง เพื่อแก้ปัญหาการตรวจจับผลลัพธ์นำมาเปรียบเทียบโดยใช้ทั้งสองอย่าง รูปกรวยหลายเหลี่ยมแบบขยายลักษณะนาม โดยใช้ YOLO	YOLO V2, YOLO Tiny, EPCC8, EPCC4	ค่าความแม่นยำ 83.35%, 80.19%, 84.49%, 81.19%	ใช่	-

ตารางที่ 2.1 ต่อ

ชื่องานวิจัย	ผู้เขียน	ปี	คลาส	รายละเอียด	หลักการ	ผล	ภาพ	วิดีโอ
Object Detection and Tracking for Community Surveillance using Transfer Learning	Machiraju et al	2019	Car, Human, etc	ในการศึกษานี้มีประสิทธิภาพและการตรวจจับคนเดินเท้าที่แม่นยำอัลกอริทึมนำเสนอโดยใช้ปรับปรุงโครงข่าย Yolov3 ในขั้นต้น กระบวนการเริ่มต้นด้วยการตรวจจับภาพและการขยายการทำงานสู่วิดีโอแบบเรียลไทม์การตรวจจับ	YOLO V3	-	ใช้	ใช้
Vehicle classification in video using deep learning.	Faruque, Hadi et al	2019	Car, human, etc	งานวิจัยนี้ผู้เขียนได้ใช้ Faster R-CNN และ YOLO คือใช้สำหรับการนับและจัดประเภทยานพาหนะในวิดีโอ โดยใช้ชุดข้อมูล 3 ชุดในการเปรียบเทียบและฝึกอบรม	YOLO , Faster R-CNN	Classification สำเร็จ 99.76% and 99.38% ของชุดที่ดีที่สุดของทั้ง 2 โมเดล	ใช้	ใช้

ตารางที่ 2.1 ต่อ

ชื่องานวิจัย	ผู้เขียน	ปี	คลาส	รายละเอียด	หลักการ	ผล	ภาพ	วิดีโอ
Real-time Object detection and Classification for Autonomous Driving	Naghavi and Hamidreza	2018	Car	งานวิจัยนี้ผู้เขียนได้ใช้ Single Deep Convolutional neural ในการตรวจจับยานพาหนะบนท้องถนน	Faster R-CNN, YOLO, SSD300	mAP 72.3%, 64.8%, 73.2%	ใช่	ใช่
The Real-Time Detection of Traffic Participants Using YOLO Algorithm	Aleksa Ćorović et al.	2018	Object Traffic Participants	งานวิจัยนี้ผู้เขียนได้ใช้อบรมโครงข่ายสำหรับวัตถุ 5 คลาส (ยานพาหนะ ยานพาหนะบรรทุกทางเท้า ป้ายจราจร และ ไฟ) และได้แสดงให้เห็นประสิทธิภาพของวิธีการในความหลากหลายของสภาพการขับขี่ (สว่างและท้องฟ้ามีดครึ้ม หิมะ หมอก และ กลางคืน)	YOLO V3	mAP 46.60%	ใช่	-

ตารางที่ 2.1 ต่อ

ชื่องานวิจัย	ผู้เขียน	ปี	คลาส	รายละเอียด	หลักการ	ผล	ภาพ	วิดีโอ
Moving object detection and tracking using deep learning neural network and correlation filter	Supreeth and Patil	2018	Car, Human	งานวิจัยนี้ผู้เขียนได้ใช้อบรมโครงข่ายสำหรับวัตถุ 5 คลาส (ยานพาหนะ ยานพาหนะบรรทุกทางเท้า ป้ายจราจร และ ไฟ) และได้แสดงให้เห็นประสิทธิภาพของวิธีการในความหลากหลายของสภาพการขับขี่ (สว่างและท้องฟ้ามีดครึ้ม หิมะ หมอก และ กลางคืน)	YOLO V3	mAP 46.60%	ใช่	ใช่

ตารางที่ 2.2 เทคนิคการใช้งานสำหรับการตรวจจับยานพาหนะ

ลำดับ	วิธีการ	หลักการ
1	การเตรียมความพร้อมข้อมูล	ตัดป้ายคำกำกับ, แยกเฟรมจากวิดีโอ, ปรับปรุงชุดข้อมูล
2	การเรียนรู้เชิงลึก	YOLO V2 , YOLO V3, YOLO V4, CNN, and SSD
3	ชุดข้อมูล	ชุดข้อมูลจากกล้องวงจรปิด, ชุดข้อมูลพื้นฐาน

จากตารางที่ 2 สามารถสังเกตเห็นได้ว่าเทคนิคการใช้งานประกอบด้วยสามกระบวนการ: (i) การประมวลผลล่วงหน้า (ii) วิธีการเรียนรู้เชิงลึก และ (iii) การเก็บภาพ เมื่อรวบรวมชุดข้อมูลรูปภาพ แนะนำกระบวนการก่อนการประมวลผล การตัดป้ายกำกับเป็นวิธีการตัดป้ายกำกับวัตถุที่จะตรวจพบซึ่งสามารถนำไปใช้กับโมเดลได้ในภายหลัง การกรองรูปภาพเพื่อแยกลักษณะเป็นวิธีการแยกคุณสมบัติของรูปภาพ วิธีการเรียนรู้เชิงลึกตามวรรณกรรม ได้แก่ YOLO V3, YOLO V4, และ YOLO V5 อุปกรณ์จับภาพคือดาวเทียม และ กล้องวงจรปิด การเรียนรู้แบบถ่ายโอน คือ การนำความรู้ที่ได้รับการฝึกอบรมมาล่วงหน้าเกี่ยวกับปัญหาเดียวกันมาประยุกต์ใช้กับปัญหาที่คล้ายคลึงกัน

เกี่ยวกับงานที่นำเสนอในข้อเสนอ ข้อมูลวิดีโอมีวัตถุประสงค์เพื่อรวบรวมโดยใช้กล้องวงจรปิด เนื่องจากกล้องวงจรปิดสามารถเก็บข้อมูลประเภทยานพาหนะต่างๆ ได้ง่าย ดังนั้น กล้องวงจรปิดจึงเป็นอุปกรณ์ที่เหมาะสมในการบันทึกข้อมูล ปัจจุบันมีการตรวจตราจราจรโดยใช้กล้องวงจรปิด การระบุบุคคลเพื่อความปลอดภัย การตรวจจับ จำแนกโรค และอื่น ๆ อีกมากมาย การวิจัยมุ่งพัฒนาสถาปัตยกรรมการเรียนรู้เชิงลึกใหม่เพื่อตรวจหาเงื่อนงำที่แท้จริงของการระบุลักษณะทางกายภาพของยานพาหนะ สรุปผลงานที่ผ่านมาที่เกี่ยวข้องกับงานที่นำเสนอในรายงานนี้

2.9 การตรวจวัดประสิทธิภาพ

อธิบายรายละเอียดเกี่ยวกับเมตริกการประเมินการจำแนกประเภทที่จะใช้ในการวิจัยนี้ ในการประเมินประสิทธิภาพการจำแนกประเภทจะใช้มาตรการการประเมินต่างๆ การประเมินวิธีการที่กล่าวถึงในงานวิจัยนี้คือ (i) ความสับสนเมตริกซ์ และ (ii) หมายถึงความแม่นยำเฉลี่ย (mAP) รายละเอียดของแต่ละวิธีจะกล่าวถึงด้านล่าง

เมตริกซ์ความสับสน (Confusion Matrix) (Shind et al., 2018) เมตริกซ์ความสับสนเป็นอัลกอริทึมสำหรับการเปรียบเทียบประสิทธิภาพระหว่างค่าที่คาดการณ์ และ ค่าจริงในรูปแบบของกริด $a \times n$ เมตริกซ์ที่ใช้สำหรับการประเมินประสิทธิภาพของโมเดลการจำแนกประเภทสำหรับการจำแนกประเภทไบนารี เมตริกซ์ความสับสนคือเมตริกซ์ 2×2 หากคลาสเป้าหมายคือ 3 เมตริกซ์ความสับสนจะเป็นเมตริกซ์ 3×3 ไปเรื่อย ๆ ตัวอย่างของเมตริกซ์ 2×2 ดังภาพประกอบที่ 2.41

	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

ภาพประกอบที่ 2.41 เมทริกซ์ความสับสน (Confusion Matrix)

จากภาพประกอบที่ 2.40

True Positive (TP) คือค่าที่ทำนายว่าการทำนายเป็นจริงซึ่งสอดคล้องกับค่าจริง

True Negative (TN) คือค่าที่คาดคะเนว่าการคาดคะเนไม่เป็นจริงซึ่งตรงกับค่าความจริง

False Positive (FP) คือค่าการทำนายที่คาดการณ์ว่าเป็นผิดซึ่งไม่ตรงกับค่าจริง

False Negative (FN) คือค่าที่คาดการณ์ไว้ซึ่งไม่ตรงกับค่าจริง

มีการใช้ค่าสี่ค่าเพื่อวัดโดยการคำนวณความแม่นยำ การเรียกคืน ความแม่นยำ นอกจากนี้โมเดลการตรวจจับวัตถุแล้ว การประเมินยังมีฟังก์ชัน mAP และการสูญเสีย

ความแม่นยำ (Accuracy) เป็นการวัดความแม่นยำโดยรวมของระบบระหว่างค่าจริง และ ค่าทำนาย ถ้าค่าความแม่นยำสูงแสดงว่าค่าทำนายนั้นถูกต้อง ใกล้เคียงกับค่าจริง (Visa et al., 2011) ตามสมการที่ 8

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Precision คือ การวัดความแม่นยำที่วัดได้จากความสามารถในการทำซ้ำของค่าทำนาย 8 ค่าที่ทำนายไว้ซึ่งมีความแม่นยำถึงค่าจริง ถ้าค่าความแม่นยำสูงแสดงว่าค่าทำนายใกล้เคียงกับค่าจริง (Bittrich et al., 2019) ดังสมการที่ 9

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

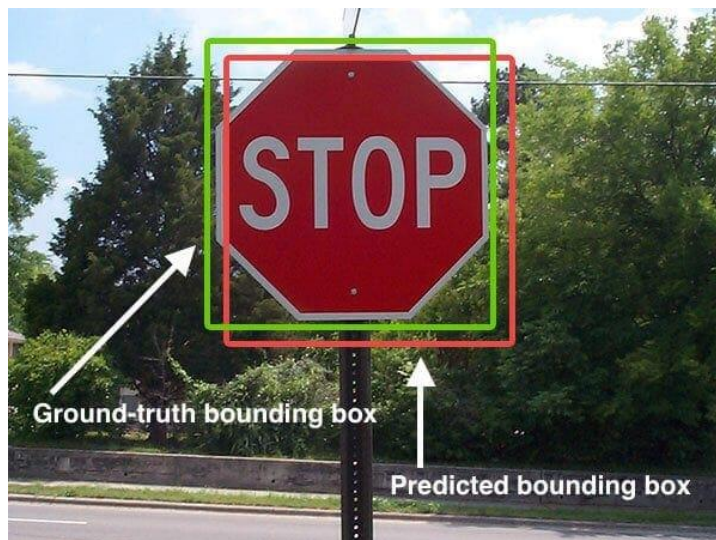
การเรียกคืน Recall เป็นการวัดความสมบูรณ์ ซึ่งหมายความว่าอัตราส่วนของการทำนายที่คาดการณ์ไว้วัดค่าที่ถูกต้องตรงกับค่าจริงจากจำนวนค่าจริงทั้งหมด ถ้าเรียกคืนค่ามีค่ามาก ซึ่งหมายความว่าค่าการทำนายสามารถคาดการณ์ได้อย่างสมบูรณ์ใกล้เคียงกับค่าจริง (Bittrich, et al., 2019) ดังสมการที่ 10

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP+FN}} \quad (10)$$

ค่าเฉลี่ยความแม่นยำ (Mean Average Precision: mAP) เป็นเมตริกที่ใช้ในการประเมินโมเดลการตรวจจับวัตถุ เช่น Fast R-CNN, YOLO, Mask R-CNN เป็นต้น Average Precision (AP) เป็นความแม่นยำเฉลี่ยจะวัดความแม่นยำ และความแม่นยำของระบบตรวจจับวัตถุ หรือเรียกค้นข้อมูลสำหรับคลาสหรือหมวดหมู่เฉพาะ โดยจะวัดว่าระบบจัดลำดับและดึงข้อมูล อินสแตนซ์ที่เกี่ยวข้องของคลาสนั้นได้ดีเพียงใด ค่าเฉลี่ยความแม่นยำ ใช้เพื่อประเมินประสิทธิภาพโดยรวมของระบบการตรวจจับวัตถุหรือเรียกค้นข้อมูลในหลายคลาสหรือหมวดหมู่ โดยจะคำนวณค่าเฉลี่ยของค่า AP สำหรับทุกคลาส โดยให้ค่าสเกลาร์เดียวที่สรุปประสิทธิภาพของระบบในการระบุ คลาสอีอบเจ็กต์หลายคลาส ในการคำนวณ mAP จะคำนวณค่า AP สำหรับแต่ละคลาสแยกกัน จากนั้นจึงหาค่าเฉลี่ยของค่า AP และคำนึงถึงความผันแปรของประสิทธิภาพในคลาสต่าง ๆ (Prasetyo et al., 2020) ดังแสดงในสมการที่ 11

$$\mathbf{mAP} = \frac{1}{2} \sum_{i=1}^n \mathbf{AP}_i \quad (11)$$

ไอโอยู (IoU) อย่างที่อธิบายไปในหัวข้อการตรวจจับวัตถุเป็นการวัดขนาดของการทับซ้อนกันระหว่างสองสิ่งก่อดล้อมรอบ (หรือในกรณีทั่วไปคือวัตถุสองชิ้น) โดยคำนวณขนาดของการทับซ้อนกันระหว่างวัตถุสองชิ้นหารด้วยพื้นที่รวมของวัตถุทั้งสองที่รวมกัน ตัวอย่างของกรอบขอบเขตการทำนาย และความจริงพื้นฐานดังภาพประกอบที่ 2.42



ภาพประกอบที่ 2.42 ค่าความมั่นใจไอโอยู โดย (Rosebrock , 2016)

สำหรับแนวคิดนี้พิจารณาที่มุมมองของฉาก มีสองเซตคือ P ซึ่งแทนเซตของพิกเซลในกรอบที่เครื่องทำนายเลือกมา และ G คือเซตของพิกเซลในกรอบที่เป็นเฉลย (ใช้ P เพราะแสดงถึงกรอบเครื่องที่ทำนาย และ G คือกรอบที่เฉลย) การทำโมเดลการตรวจจับวัตถุจะหาค่าความเชื่อมั่นหรือวัดประสิทธิภาพระบุตำแหน่งของวัตถุได้จาก อัตราส่วนระหว่างพื้นที่ อินเตอร์เซกชันของทั้ง 2 bounding box หารด้วยพื้นที่รวมของกรอบทั้ง 2 กรอบ (Yun et al., 2021) ดังสมการที่ 12

$$IoU(P, G) = \frac{|P \cap G|}{|P \cup G|} \quad (12)$$



ภาพประกอบที่ 2.43 เกณฑ์ไอโอยู

จากภาพประกอบที่ 2.42 ตัวเลขตั้งแต่ 0 ถึง 1 ที่ระบุจำนวนของการทับซ้อนกันระหว่างกล่องขอบเขตความจริงที่ทำนายและกรอบเฉลี่ย ไอโอยู ของ 0 หมายความว่าไม่มีการทับซ้อนกันระหว่างกล่องที่ไม่สามารถทำนายได้ โดยทั่วไป ไอโอยู ที่ 0.5 ถือว่าเป็นการคาดการณ์ที่ดี ไอโอยู ของ 1 หมายความว่าสภาพของกล่องจะเหมือนกันกับการทับซ้อนกันซึ่งบ่งชี้ว่าพวกมันซ้อนทับกันอย่างสมบูรณ์ถือเป็นการทำนายอย่างสมบูรณ์

2.10 สรุป

ในบทนี้ได้ทบทวนวรรณกรรม และ งานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ ประกอบด้วย โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network) เพื่ออธิบายรายละเอียดกระบวนการที่อยู่ในโครงข่าย, การเรียนรู้เชิงลึก, การประมวลผลภาพอธิบายกระบวนการทำงานของการประมวลผลภาพ, การประมวลผลภาพเพื่ออธิบายกระบวนการประมวลผลภาพ และการตรวจจับวัตถุ เพื่ออธิบายกระบวนการตรวจจับวัตถุ การถ่ายโอนการเรียนรู้เทคนิคในการลดเวลาการฝึกอบรม สถาปัตยกรรม YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 ที่ใช้ในการศึกษาเพื่อวิจัยตรวจจับลักษณะทางกายภาพของยานพาหนะ ตารางการเปรียบเทียบของงานวิจัยที่เกี่ยวข้อง และ สุดท้าย การวัด และ ประเมินผลสำหรับงานวิจัย โดยการทบทวนวรรณกรรม และ งานวิจัยที่เกี่ยวข้องกับงานวิจัยในบทนี้จะนำไปประยุกต์ในบทที่ 3 ที่เกี่ยวกับการเตรียมความพร้อมของชุดข้อมูลที่ใช้ในการพัฒนาโมเดลการจำแนกยานพาหนะที่ระบุบุลักษณะทางกายภาพโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

บทที่ 3

ชุดข้อมูลภาพ และ กระบวนการเตรียมความพร้อมของภาพ

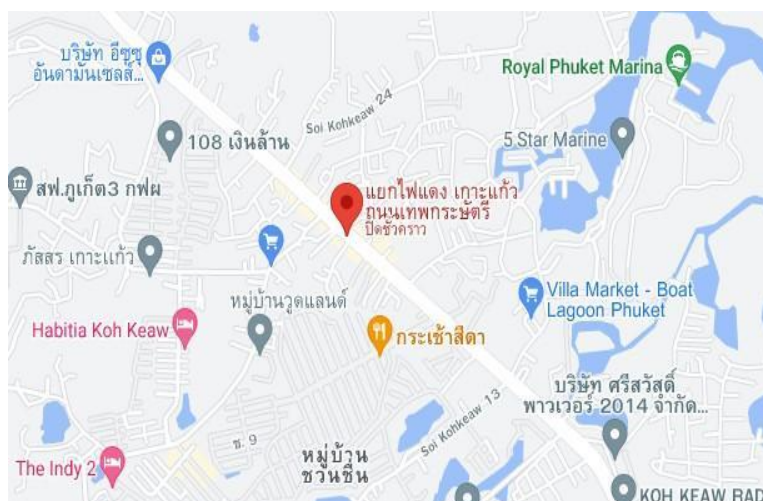
3.1 บทนำ

ในบทนี้จะเป็นการอธิบายถึงการได้มาของชุดข้อมูลวิดีโอการจราจรที่ใช้ในงานวิจัยนี้ ซึ่งเป็นงานการจำแนกลักษณะทางกายภาพของยานพาหนะ โดยจะกล่าวถึงแหล่งที่มาของชุดข้อมูลวิดีโอว่ามาจากที่ใด และ กระบวนการแยกเฟรมภาพจากวิดีโอเพื่อจะติดค่ากำกับที่จะใช้ในการฝึกอบรมในตัวโมเดลเพื่อตรวจจับลักษณะทางกายภาพของยานพาหนะ โดยกลุ่มที่ผู้วิจัยสนใจในงานวิจัยนี้มีชุดข้อมูล 3 กลุ่มได้แก่ (i) ประเภทของยานพาหนะ 4 ประเภท ยานพาหนะ, ยานพาหนะกระบะ, ยานพาหนะจักรยานยนต์, และ ยานพาหนะโดยสาร หรือยานพาหนะบรรทุก (ii) สีของยานพาหนะ 10 สี สีขาว, สีดำ, สีเทา/เงิน/สีบรอนซ์, สีน้ำเงิน, สีแดง, สีเหลือง, สีเขียว, สีน้ำตาล, สีส้ม, และอื่นๆ (iii) แบรินต์ของยานพาหนะ 10 แบรินต์ โตโยต้า, อีซูซุ, ฟอร์ด, มาสด้า, บีเอ็มดับเบิลยู, มิตซูบิชิ, ฮอนด้า, ซูซูกิ, นิสสัน, และ อื่นๆ โดยจะมีเพิ่มเติมเทคนิคการเพิ่มปริมาณข้อมูลของชุดข้อมูลทั้งหมด 3 แบบ (i) ความน่าจะเป็น 50% ของการพลิกแนวนอน (ii) การหมุนแบบสุ่มระหว่าง -15° ถึง $+15^{\circ}$ องศา และ (iii) การเอียงแบบสุ่มระหว่าง -15° ถึง $+15^{\circ}$ ในแนวนอน และ -15° ถึง $+15^{\circ}$ ในแนวตั้ง เพื่อให้มีคุณภาพมากขึ้นเพื่อเพิ่มความแม่นยำตรวจจับของตัวโมเดลก่อนแยกชุดข้อมูลจำแนกเป็น ชุดข้อมูลเรียนรู้สำหรับการอบรมตัวโมเดล (Training set) , ชุดตรวจสอบความถูกต้อง (Validation Set) และ ข้อมูลชุดการทดสอบใช้สำหรับการทดสอบโมเดลที่สร้าง (Test set)

หัวข้อที่จะกล่าวต่อไปนี้จะประกอบด้วย หัวข้อ 3.2 ที่มาของชุดข้อมูล โดยบอกที่มาของชุดข้อมูลวิดีโอที่ใช้ในการสร้างชุดข้อมูลสำหรับพัฒนาโมเดลการจำแนกยานพาหนะที่ระบุลักษณะทางกายภาพ 3.3 แยกเฟรมภาพจากวิดีโอขั้นตอนนี้สามารถวิเคราะห์และตรวจสอบเนื้อหาวิดีโอโดยละเอียด โดยตรวจสอบแต่ละเฟรมเพื่อรวบรวมข้อมูลเชิงลึก 3.4 เพิ่มปริมาณข้อมูลเพื่อเพิ่มความหลากหลายและขนาดของชุดข้อมูลการฝึกอบรมโดยการใส่การแปลงและการแก้ไขต่าง ๆ กับข้อมูลต้นฉบับ วัตถุประสงค์หลักของการเพิ่มข้อมูลคือการปรับปรุงลักษณะทั่วไป ความหลากหลายของภาพ และเพิ่มประสิทธิภาพของโมเดลการเรียนรู้ของเครื่อง 3.5 ติดป้ายกำกับเพื่อระบุคำอธิบายหรือหมวดหมู่ของข้อมูล ช่วยให้โมเดลการเรียนรู้ของเครื่องเข้าใจและเรียนรู้รูปแบบภายในชุดข้อมูล 3.6 กล่าวสรุปทั้งหมดของบทที่ 3

3.2 ที่มาของชุดข้อมูล

รายละเอียดข้อมูลที่ใช้ในการนำเสนอนี้ วิดีโอความยาว 30 นาทีมีอัตราเฟรม 10 fps ที่ได้จากกล้องวงจรปิดรุ่น Dahua IPC-HFW5541E-ZE พร้อมกับ 5 MP, IP, Vari-Focal Bullet, กล้องเครือข่าย WizMind และ IP67 ที่ตำบลเกาะแก้ว ถนนเทพกระษัตรี อำเภอเมืองภูเก็ต จังหวัดภูเก็ต เมื่อวันที่ 1 มีนาคม พ.ศ. 2565 ในแผนที่ภาพประกอบที่ 3.1



ภาพประกอบที่ 3.1 แผนที่แสดงที่ตั้งกล้อง



ภาพประกอบที่ 3.2 ตัวอย่างภาพจากกล้องจราจรปิด

3.3 แยกเฟรมภาพจากวิดีโอ

เมื่อได้ชุดข้อมูลวิดีโอแล้ว เมื่อได้รับข้อมูลวิดีโอแล้ว เริ่มต้นด้วยการแยกข้อมูลวิดีโอออกเป็นชุดของเฟรมภาพวิดีโอทุก 1 วินาทีกำหนดให้สามารถแยกออกเป็น 10 เฟรม ได้ดังตารางที่ 3.1 แสดงจำนวนการแยกเฟรมได้ เหตุผลที่เลือกเฟรมภาพที่ไม่เท่ากันเนื่องจากความยากง่ายของการจำแนกยานพาหนะที่แตกต่างกันที่ใช้ในการฝึกอบรม และ ทดสอบโมเดลเนื่องจากบางรูปภาพมียานพาหนะคลาสที่เหมือนกันมากกว่า 1 คันสำหรับการตรวจจับประเภทของยานพาหนะที่วัดภูมิความใหญ่ตรวจจับได้ง่ายดังนั้นจึงใช้ชุดข้อมูลภาพที่ไม่ต้องเยอะและยังมีเทคนิคเพิ่มปริมาณข้อมูลที่อยู่ในขั้นตอนต่อไป ส่วนอีก 2 โมเดลมีเหตุผลที่คล้ายคลึงกันคือเป็นการจำแนกที่ต้องใช้ปริมาณข้อมูลที่ยะอะแต่ยังต้องรักษาความสมดุลของคลาสไม่มากไม่น้อยจนเกินไป รายละเอียดของข้อมูลที่ได้แสดงในตารางที่ 3.1

ตารางที่ 3.1 ตารางการแยกเฟรมวิดีโอ

ประเภท	วิดีโอ	เวลาต่อวินาที	เฟรมที่ใช้ในการทดลอง
ยานพาหนะ	1	15,989	757
สี่	1	15,989	3,550
แบนด์	1	15,989	5,283

3.4 การเพิ่มปริมาณข้อมูล

เพิ่มข้อมูลมีความหลากหลายมากขึ้น หลังจากเพิ่มข้อมูลกระบวนการติดป้ายกำกับแล้ว เพื่อโมเดลสำหรับการจำแนกที่ระบุลักษณะทางกายภาพของยานพาหนะที่มีประสิทธิภาพมากขึ้น การปรับปรุงข้อมูลสำหรับการฝึกอบรมที่มีอยู่ให้ครอบคลุมในสถานการณ์อื่น ๆ ช่วยให้โมเดลสามารถเรียนรู้จากสถานการณ์ที่หลากหลายขึ้น หนึ่งในกระบวนการที่สำคัญที่สุดในการพัฒนาโมเดลการทำนาย คือ การเพิ่มข้อมูลเพื่อสร้างภาพประกอบที่หลากหลายกระบวนการเสริมข้อมูลใน Roboflow ถูกนำมาใช้เพื่อสร้างข้อมูลเสริม จากการศึกษาเบื้องต้นนี้ มีการใช้การเพิ่มข้อมูลทฤษฎีการ (i) ความน่าจะเป็น 50% ของการพลิกแนวนอน (ii) การหมุนแบบสุ่มระหว่าง -15 ถึง $+15$ องศา (iii) การเฉือนแบบสุ่มระหว่าง -15° ถึง $+15^\circ$ (Jubayer et al., 2021) ภาพประกอบที่ 3.3 แสดงตัวอย่างการเพิ่มข้อมูล เมื่อทำการเสริมข้อมูลแล้ว จำนวนชุดภาพก็เปลี่ยนไปดังตารางที่ 3.2



(a) ภาพดั้งเดิม



(b) 50% probability of horizontal flip



(c) Random rotation of between -15 and +15 degrees



(d) Random shear of between -15° to +15° horizontally and -15° to +15° vertically

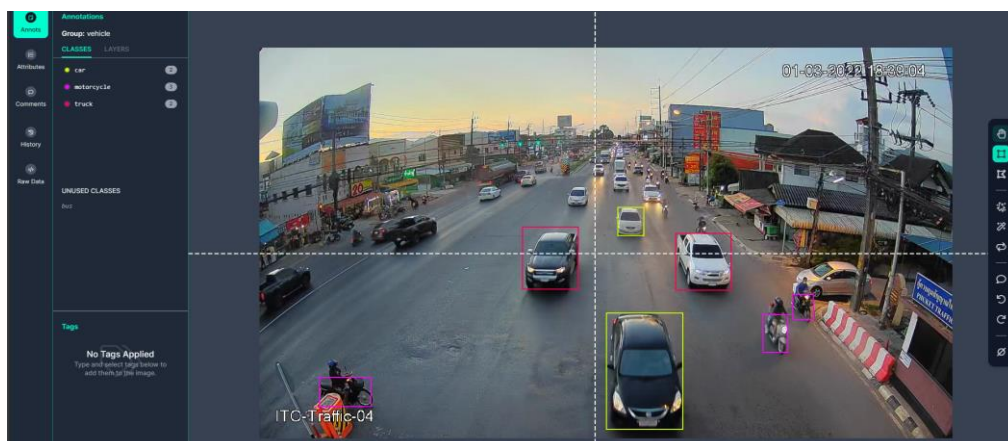
ภาพประกอบที่ 3.3 ตัวอย่างภาพการเพิ่มปริมาณข้อมูล

ตารางที่ 3.2 จำนวนภาพหลังจากใช้กระบวนการเสริมภาพ

ประเภท	รูปภาพ	หลังใช้เทคนิคเพิ่มปริมาณข้อมูล
จำแนกการตรวจจับประเภทยานพาหนะ	757	2,271
จำแนกการตรวจจับประเภทสียานพาหนะ	3,550	10,650
จำแนกการตรวจจับแบรนด์ยานพาหนะ	5,283	15,849

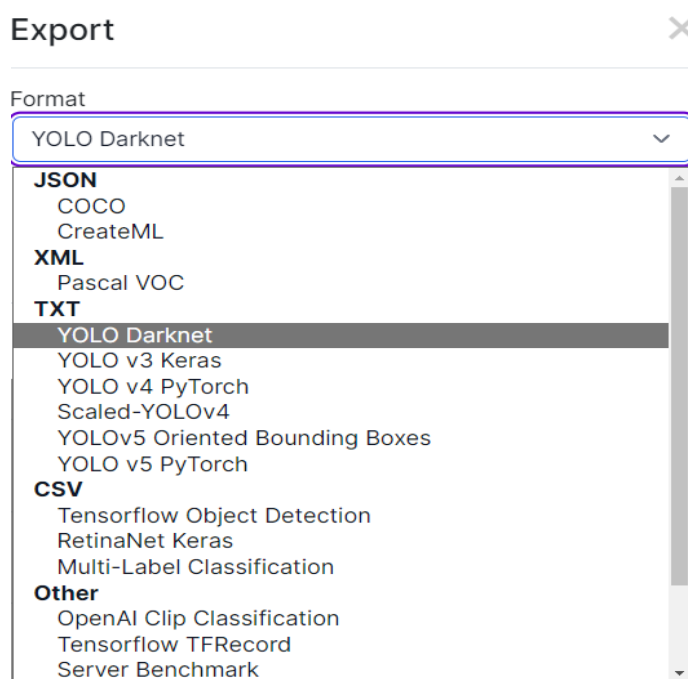
3.5 ติดป้ายกำกับ

การติดป้ายกำกับคลาส (Label) วัตถุประสงค์ของการติดป้ายกำกับวัตถุ คือ เพื่อระบุวัตถุคลาสที่จะทำการจำแนกในภาพโดยแบ่งชุดข้อมูลออกเป็น 3 ชุดได้แก่ i) ประเภทของยานพาหนะ, ii) สีของยานพาหนะ และ iii) แบรินด์ของยานพาหนะโดยจะใช้ในกระบวนการต่อไปจากการศึกษาเบื้องต้นนี้ แอปพลิเคชัน Roboflow ถูกนำมาใช้เป็นเครื่องมือในการติดป้ายกำกับทั้ง 3 ชุด



ภาพประกอบที่ 3.4 ตัวอย่างภาพการติดป้ายกำกับคลาส

จากภาพประกอบที่ 3.4 เป็นตัวอย่างภาพการติดป้ายกำกับคลาสเมื่อทำการติดป้ายกำกับคลาสแล้ว ขั้นตอนต่อไปเป็นการเอ็กซ์พอร์ตไฟล์ชุดข้อมูลในแอปพลิเคชัน Roboflow ตัวอย่างการเอ็กซ์พอร์ตไฟล์ดังภาพประกอบที่ 3.5



ภาพประกอบที่ 3.5 ตัวอย่างประมวลผลผลลัพธ์ข้อมูลที่ติดป้ายกำกับโดยใช้ Roboflow

เมื่อเอ็กซ์พอร์ตไฟล์แล้วจะได้ข้อมูลที่แยกออกเป็น 3 ชุดได้แก่ i) ชุดทดสอบ, ii) ชุดฝึกอบรม และ iii) ชุดตรวจสอบความถูกต้อง ดังภาพประกอบที่ 3.6



ภาพประกอบที่ 3.6 ตัวอย่างภาพไฟล์เก็บชุดข้อมูลที่ใช้ในการพัฒนาโมเดลจาก Roboflow

3.6 สรุป

บทนี้เป็นการอธิบายถึงการเตรียมชุดข้อมูลภาพ และ กระบวนการเตรียมความพร้อมของภาพด้วยเทคนิคต่าง ๆ เพื่อที่สร้างชุดข้อมูลให้มีคุณภาพที่ได้จากกล้องวงจรปิดที่นำมาจากทางจราจรสี่แยกในจังหวัดภูเก็ตที่กล่าวไปจากนั้นนำมาแยกเฟรมให้เป็นรูปภาพเพื่อที่จะนำไปพัฒนาตัวโมเดลการระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ โดยใช้ชุดข้อมูลทั้งหมด 3 ชุดได้แก่ (i) ชุดข้อมูลประเภทของยานพาหนะ (ii) ชุดข้อมูลสี่ของยานพาหนะ และ (iii) ชุดข้อมูลของแบรนด์ยานพาหนะ ชุดข้อมูลทั้งหมดจะนำไปใช้ในการพัฒนาโมเดลสำหรับ i) ประเภทของยานพาหนะ, ii) สีของยานพาหนะ และ iii) แบนด์ของยานพาหนะ

บทที่ 4

การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทของยานพาหนะ

4.1 บทนำ

งานวิจัยนี้ได้สนใจการจำแนกการตรวจจับประเภทยานพาหนะเป็นเทคโนโลยีอเนกประสงค์ที่มีการใช้งานครอบคลุมทั้งการขนส่ง การวางผังเมือง ความปลอดภัย การอนุรักษ์สิ่งแวดล้อม และขอบเขตอื่น ๆ ซึ่งมีส่วนช่วยในการปรับปรุงความปลอดภัย ประสิทธิภาพ และคุณภาพชีวิตโดยรวมในเมืองและชุมชนสมัยใหม่ จากสถิติศูนย์ป้องกันและปราบปรามการโจรกรรมรถยนต์ พบว่า ประเทศไทยมีรถหาย มากกว่า 20,000 คันต่อปี แบ่งออกเป็นรถจักรยานยนต์จำนวน 10,000 กว่าคัน รถกระบะ ประมาณ 3,000 คัน โดยจังหวัดที่มีสถิติรถหายมากที่สุดคือ กรุงเทพมหานคร ที่มีจำนวนรถหายมากกว่า 7,500 คันต่อปี โดยได้คืนเพียง 10% (Editorj, 2022) งานวิจัยนี้ใช้ชุดข้อมูลวิดีโอที่ได้ทำการแตกเฟรมออกจากวิดีโอเป็นรูปภาพแล้วโดยจากกล้องวงจรปิดสี่แยกจรัญจงห์เกิดจำนวนทั้งหมด 1 ชุดได้แก่ 2,171 ภาพสำหรับชุดข้อมูลประเภทยานพาหนะประเภทยานพาหนะที่สนใจทั้งหมด 4 ประเภท i) รถยนต์, ii) รถกระบะ iii) รถจักรยานยนต์ และ iv) รถโดยสาร แบ่งออกเป็น 3 ชุดข้อมูล คือ ชุดข้อมูลที่ใช้สำหรับการฝึกอบรม, ชุดตรวจสอบความถูกต้อง และ ชุดข้อมูลที่ใช้สำหรับทดสอบ ในบทนี้จะกล่าวถึงการใช้ออนไลน์เรียนรู้เชิงลึกให้กับตัวสถาปัตยกรรมโครงข่ายประสาทเทียมแบบสังวัตนาการของตัว YOLO ทั้ง 6 สถาปัตยกรรมได้แก่ YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 เพื่อเปรียบเทียบว่าสถาปัตยกรรมใดจะมีประสิทธิภาพดีที่สุดสำหรับการจำแนกประเภทของยานพาหนะ

หัวข้อย่อยที่จะกล่าวถึงภายในบทนี้จะประกอบด้วย หัวข้อย่อย 4.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set) รวมทั้งอธิบายถึงชุดข้อมูลที่ใช้ การเตรียมความพร้อมของข้อมูลภาพ อีกทั้งยังกล่าวถึงวิธีการเลือกชุดข้อมูลสำหรับการสร้างตัวจำแนก และ แสดงผลลัพธ์ของงานวิจัยใช้ในบทนี้ หัวข้อย่อย 4.3 การวัด และ ประเมินผล จะกล่าวถึงวิธีการวัด และ ประเมินผลรวม แสดงผลลัพธ์ของงานวิจัย หัวข้อย่อย 4.4 สรุปผลการทดลองภายในบทที่ 4

4.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set)

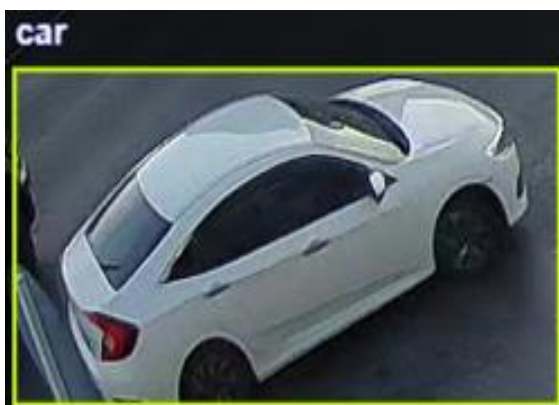
ในการพัฒนาโมเดลสำหรับการจำแนกประเภทของยานพาหนะมีชุดข้อมูลมีทั้งหมด 2,271 ภาพได้แก่ i) รถยนต์ 687 ภาพ ii) รถกระบะ 492 ภาพ iii) รถจักรยานยนต์ 791 ภาพ และ iv) รถโดยสาร 301 ภาพ และ ทำการแบ่งชุดข้อมูล 3 ชุดได้แก่ i) ชุดข้อมูลฝึกอบรม ร้อยละ 80, ii) ชุดตรวจสอบความถูกต้องร้อยละ 10 และ iii) ชุดทดสอบร้อยละ 10 ที่ใช้สำหรับการพัฒนาตัวโมเดลสำหรับตรวจจับลักษณะทางกายภาพของยานพาหนะ

ตารางที่ 4.1 จำนวนชุดข้อมูลภาพประเภทของยานพาหนะที่ใช้ในงานวิจัย

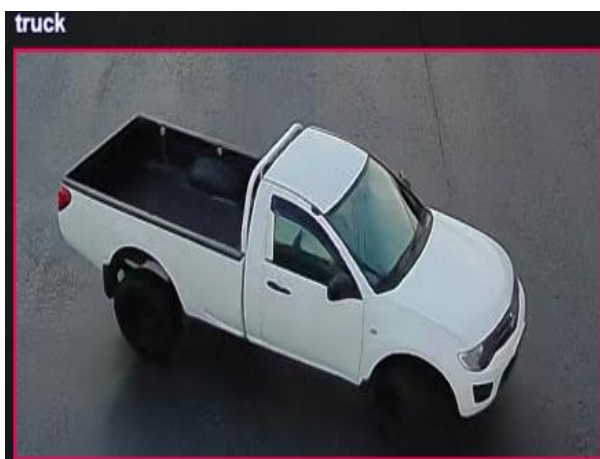
ชุดข้อมูล	ชุดข้อมูลฝึกอบรม	ชุดตรวจสอบความถูกต้อง	ชุดทดสอบ	จำนวนทั้งหมด
รถยนต์	549	69	69	687
รถกระบะ	394	49	49	492
รถจักรยานยนต์	633	79	79	791
รถโดยสาร	241	30	30	301
จำนวนทั้งหมด	1,817	227	227	2,271

การเตรียมความพร้อมของภาพ (Image Pre-Processing)

จากที่กล่าวมาในหัวข้อก่อนหน้า ที่ได้อธิบายถึงชุดข้อมูลที่ใช้ในงานวิจัย เริ่มต้นด้วยการได้ชุดข้อมูลภาพประเภทของยานพาหนะแบ่งออกเป็น 4 ประเภทจากตารางที่ 4.1 ในหัวข้อที่ 4.2 กระบวนการเตรียมความพร้อมของภาพที่จะนำมาพัฒนาตัวโมเดลการจำแนกประเภทของยานพาหนะโดยจะเพิ่มปริมาณข้อมูลหลังจากการตัดป้ายกำกับของชุดข้อมูลโดยใช้เทคนิคการเพิ่มปริมาณข้อมูลทั้งหมด 3 แบบ ดังกล่าวในบทที่ 3 หลังเพิ่มปริมาณข้อมูลมีจำนวนรูปภาพเพิ่มขึ้น 2,271 ภาพ ชุดข้อมูลภาพทั้งหมดหลังจากผ่านกระบวนการเตรียมความพร้อมของภาพจะนำไปพัฒนาตัวโมเดลสำหรับการจำแนกประเภทของยานพาหนะ โดยภาพประกอบ ที่ 4.1 จะแสดงภาพตัวอย่างของการตัดป้ายกำกับประเภทของยานพาหนะทั้ง 4 ประเภท



(a) รถยนต์



(b) รถกระบะ



(c) รถจักรยานยนต์



(d) รถโดยสาร

ภาพประกอบที่ 4.1 ภาพตัวอย่างประเภทของยานพาหนะ

จากภาพประกอบที่ 4.1 เป็นภาพประกอบประเภทยานพาหนะหลังจากการติดป้ายกำกับสำหรับยานพาหนะ 4 ประเภท เพื่อที่จะนำไปพัฒนาตัวโมเดลสำหรับการจำแนกประเภทของยานพาหนะโดยยานพาหนะแต่ละประเภทที่คุณลักษณะเด่นที่แตกต่างกัน เช่น รถยนต์โดยทั่วไปมีรูปแบบสี่ประตู ตัวรถมีความเพรียวบางที่กะทัดรัด รถจักรยานยนต์มีคุณลักษณะเด่นเป็น ยานพาหนะสองล้อดีไซน์เล็กกะทัดรัดซึ่งมีจุดเด่นมากกว่ายานพาหนะประเภทอื่น รถกระบะเป็นรถที่มีท้ายกระบะเป็นจุดเด่นซึ่งบางครั้งได้นำไปตกแต่งปรับเปลี่ยนรถ เช่น ปิดท้ายกระบะเป็นที่กันแดด ทางผู้วิจัยก็ได้ตรวจสอบเตรียมความพร้อมข้อมูลติดป้ายกำกับของแต่ละคันอย่างละเอียด ประเภทสุดท้าย รถโดยสารเป็นรถที่มีขนาดใหญ่และแตกต่างที่สุดในบรรดา 4 ประเภททั้งรูปร่างของตัวรถโดยทั้งหมดจะนำไปพัฒนาตัวโมเดลสำหรับการจำแนกประเภทของยานพาหนะต่อไป

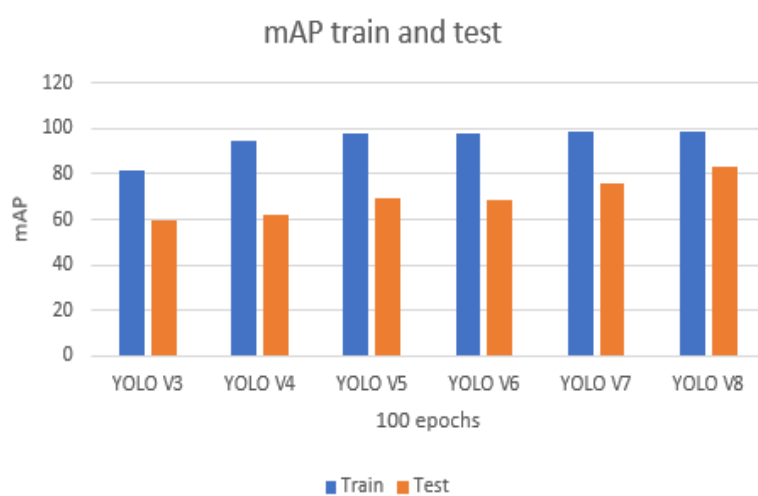
4.3 การวัด และ ประเมินผลจำแนกประเภทของยานพาหนะ

การวัด และ การประเมินผลในบทนี้ จะกล่าวถึงกระบวนการสร้างตัวตรวจจับลักษณะทางกายภาพของยานพาหนะจากโครงข่ายประสาทเทียมแบบสังวัตนาการ หรือ YOLO ที่ได้ นำเสนอที่จะใช้ในงานวิจัยนี้ จากข้อมูลภาพประกอบที่ ได้จากกล้องวงจรปิด จำนวนทั้งหมด 1 ชุด ได้แก่ 2,271 ภาพสำหรับชุดข้อมูลประเภทยานพาหนะ แบ่งออกเป็น 3 ชุดข้อมูล คือ i) ชุดฝึกอบรม, ii) ชุดตรวจสอบความถูกต้อง และ iii) ชุดข้อมูลทดสอบ ซึ่งแต่ละภาพผ่านกระบวนการเตรียมความพร้อมข้อมูลเรียบร้อยแล้ว และ การเพิ่มปริมาณข้อมูลเพื่อเพิ่มคุณภาพของชุดข้อมูล

วัตถุประสงค์ของการทดลองและการประเมินผล เพื่อจะหาโครงข่ายประสาทเทียมแบบสังวัตนาการแบบ YOLO สถาปัตยกรรมไหนเหมาะสมกับชุดข้อมูลที่เตรียมไว้สำหรับการจำแนกประเภทของยานพาหนะ ซึ่งในหัวข้อนี้จะอธิบายถึงการวัด และ การประเมินผลของการจำแนกลักษณะทางกายภาพของยานพาหนะ โดยมีการประเมินประสิทธิภาพของโมเดลจากค่าความแม่นยำ ค่าความถูกต้อง ค่าเรียกคืน ของตัวสถาปัตยกรรมทั้ง 6 สถาปัตยกรรม

การพัฒนาโมเดลสำหรับตรวจจับประเภทของยานพาหนะ

ข้อมูลที่ใช้ทั้งหมด 2,271 ภาพรวมเทคนิคการเพิ่มปริมาณรูปภาพแล้วซึ่งช่วยในการสร้างหลากหลายของชุดข้อมูลสำหรับการพัฒนาแบบจำแนกการจำแนกประเภทยานพาหนะที่แบ่งชุดข้อมูลออกเป็น 3 ชุด จากหัวข้อที่ 4.2 หลักจากติดป้ายกำกับประเภทยานพาหนะโดยใช้แอปพลิเคชัน Roboflow เสร็จสิ้น i) รถยนต์ 687 ภาพ, (ii) รถกระบะ 492 ภาพ, (iii) รถจักรยานยนต์ 791 ภาพ และ (iv) รถโดยสารติดป้ายกำกับทั้งหมด 301 ภาพ หลังจากนั้นกำหนดการฝึกอบรมตัวโมเดล โดยมีรอบการเรียนรู้ (Epochs) จำนวน 100 รอบ กำหนดค่าพารามิเตอร์ดังนี้ ขนาด batch size = 32 และ ขนาดภาพ 416 x 416 โดยตารางที่ 4.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ การจำแนกประเภทของยานพาหนะ โดยสถาปัตยกรรม YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 และ ภาพประกอบที่ 4.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทยานพาหนะระหว่างข้อมูลชุด Train และ Test



ภาพประกอบที่ 4.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทยาพาหนะระหว่างข้อมูลชุด Train และ Test

จากภาพประกอบที่ 4.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทยาพาหนะระหว่างแท่งสีฟ้าคือข้อมูลชุด Train และ สีส้มคือข้อมูลชุด Test สถาปัตยกรรม YOLO V8, YOLO V7, YOLO V6, YOLO V5 และ YOLO V4 มีค่า mAP ในข้อมูลชุด Train มีค่าใกล้เคียงกัน โดยตัวสถาปัตยกรรม YOLO V3 มีค่า mAP น้อยที่สุดในบรรดา 6 สถาปัตยกรรม ในส่วนของค่า mAP ข้อมูลชุด Test ในการจำแนกประเภทของยานพาหนะ สถาปัตยกรรม YOLO V8 ได้ค่า mAP Test มากที่สุดและ YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ตามลำดับ ซึ่งจะรายละเอียดจะอธิบายในหัวข้อถัดไป

ตารางที่ 4.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ

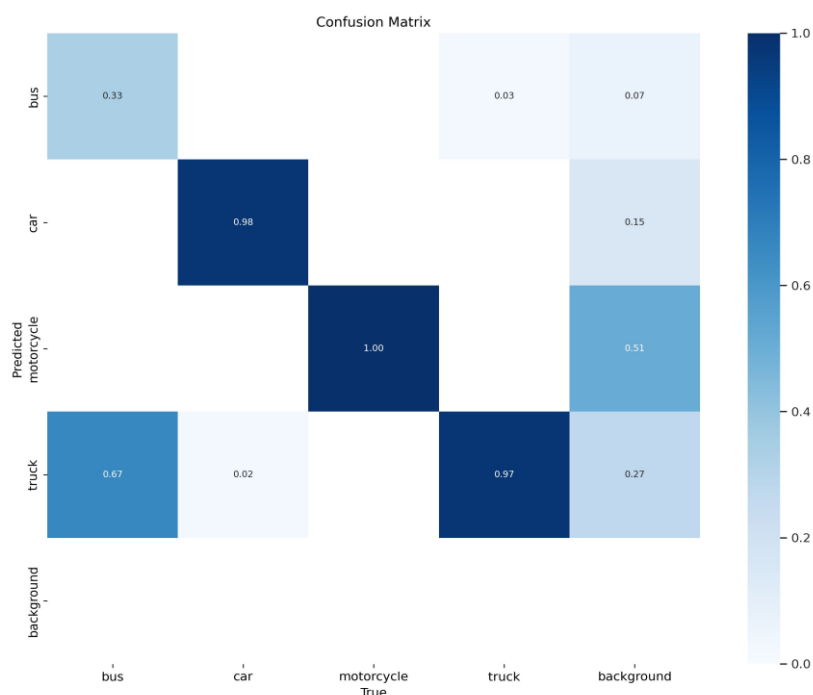
โมเดล	Train	Test
	mAP	mAP
YOLO V3	0.81	0.59
YOLO V4	0.94	0.62
YOLO V5	0.98	0.69
YOLO V6		0.68
YOLO V7		0.75
YOLO V8		0.83

จากตารางที่ 4.2 แสดงให้เห็นว่าสถาปัตยกรรม YOLO V8, YOLO V7, YOLO V6, และ YOLO V5 มีค่า mAP ร้อยละ 98 ซึ่งใกล้เคียงกันในชุด Train และ YOLO V4, YOLO V3 ให้ค่า mAP ร้อยละ 94 , 81 ตามลำดับ ส่วนของชุด Test สถาปัตยกรรม YOLO V8 ได้ค่า mAP ร้อยละ 83 มากที่สุดในบรรดา 6 สถาปัตยกรรมซึ่งจากการทดลองทั้ง 2 ชุดสรุปได้ว่า สถาปัตยกรรม YOLO V8 มีประสิทธิภาพมากที่สุดด้วยความแตกต่างด้านพัฒนาสถาปัตยกรรมในปัจจุบันซึ่ง สถาปัตยกรรม YOLO V7 – YOLO V3 ก็ยังมีประสิทธิภาพที่ดีในการจำแนกประเภทของยานพาหนะได้ดี ตารางที่ 4.3 แสดงผลลัพธ์ที่ได้เป็นจริงหรือผิดสำหรับการจำแนกประเภทยานพาหนะที่นำรูปภาพที่ไม่เคยเห็นมาก่อนมาตรวจแล้วนับว่ามีกี่รูปภาพที่จับผิดหรือถูก และภาพประกอบที่ 4.3 แสดงภาพ confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทของยานพาหนะ

ตารางที่ 4.3 ผลลัพธ์ที่ได้ตรวจจับถูกหรือผิดสำหรับการจำแนกประเภทของยานพาหนะ

โมเดล คลาส	รูปภาพ ทั้งหมด	YOLOV3		YOLO V4		YOLO V5		YOLO V6		YOLO V7		YOLO V8	
		ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด
รถยนต์	100	80	20	94	6	94	6	94	6	95	5	95	5
รถกระบะ	100	90	10	90	10	90	10	90	10	90	10	90	10
รถจักรยานยนต์	100	100	0	100	0	100	0	100	0	100	0	100	0
รถโดยสาร	60	50	10	50	10	50	10	50	10	50	10	50	10

ตารางที่ 4.3 แสดงให้เห็นได้ชัดว่าจำแนกประเภทรถกระบะ และรถโดยสารจำแนกตรวจจับผิดมากที่สุดในประเภททั้งหมด ส่วนประเภทที่ตรวจจับถูกที่สุดคือ รถจักรยานยนต์ที่ทำการจำแนกถูกมากที่สุดในประเภทของยานพาหนะจากกล้องวงจรปิด และ ตัวโมเดล YOLO V8, YOLO V7, YOLO V6 และ YOLO V5 ผลตรวจจับผิดถูกได้ใกล้เคียงกัน ซึ่งรูปภาพที่ใช้ในการจำแนกอาจเป็นปัจจัยในการจำแนกผิดถูกเช่นกัน สรุปได้ว่าการจำแนกประเภทยานพาหนะทั้ง 6 สถาปัตยกรรมมีประสิทธิภาพโดยรวมที่ดี การประเมินประสิทธิภาพสำหรับการพัฒนาการจำแนกประเภทยานพาหนะแสดงดังตารางที่ 4.4

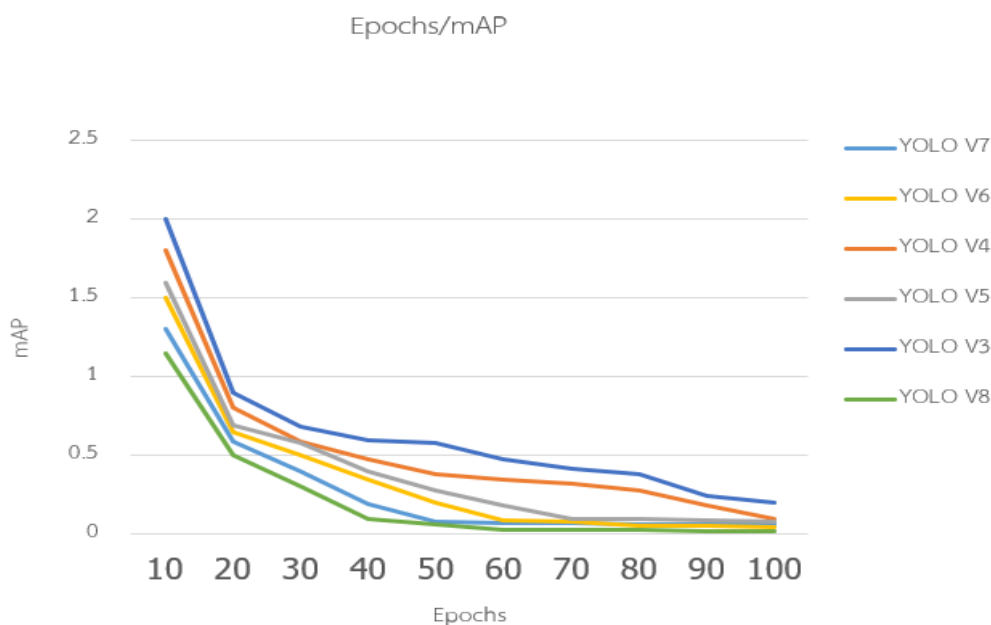


ภาพประกอบที่ 4.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทของยานพาหนะ

ตารางที่ 4.4 การประเมินประสิทธิภาพสำหรับการพัฒนาการจำแนกประเภทยานพาหนะ

ผลการวัดประสิทธิภาพ	mAP	Precision	Recall
YOLO V3	0.59	0.72	0.48
YOLO V4	0.62	0.75	0.56
YOLO V5	0.69	0.78	0.58
YOLO V6	0.68	0.79	0.62
YOLO V7	0.75	0.81	0.65
YOLO V8	0.83	0.88	0.69

ตารางที่ 4.4 แสดงให้เห็นประสิทธิภาพของแต่ละสถาปัตยกรรม เมื่อพิจารณาค่าประสิทธิภาพมากที่สุด พบว่าสถาปัตยกรรม YOLO V8 มีค่า mAP ร้อยละ 83 ค่า Precision ร้อยละ 88 และ ค่า Recall ร้อยละ 69 มากที่สุดในบรรดาทั้ง 6 สถาปัตยกรรม ผลจากชุดทดสอบแสดงให้เห็นว่าโมเดลการจำแนกประเภทของยานพาหนะที่ใช้ทั้ง 6 สถาปัตยกรรมมีประสิทธิภาพที่ดีเหมาะสมกับชุดข้อมูลที่ใช้พัฒนาการจำแนกประเภทยานพาหนะ ที่ทุกสถาปัตยกรรมมีค่า mAP มากกว่าร้อยละ 50



ภาพประกอบที่ 4.4 กราฟ loss ชุดตรวจสอบความถูกต้องของโมเดลการจำแนกประเภท ยานพาหนะ

ภาพประกอบที่ 4.4 ค่าการสูญเสียโดยรวมนี้เป็นการสูญเสียของชุดข้อมูลตรวจสอบความถูกต้องของชุดข้อมูลจำแนกประเภทของยานพาหนะ สำหรับสถาปัตยกรรม YOLO ทั้ง 6 สถาปัตยกรรมในชุดการตรวจสอบความถูกต้องน้อยกว่า 0 ค่านี้นับชี้ว่าโมเดลทำงานได้ดีเพียงใดในแง่ของความแม่นยำในการจำแนกออกเบ้จกต์และความแม่นยำในการแปลข้อมูลชุดตรวจสอบความถูกต้องซึ่งบอกได้ว่าชุดข้อมูลที่ใช้ในการฝึกอบรมมีคุณภาพไปทิศทางบวก

4.4 สรุปผล

การจำแนกประเภทของยานพาหนะโดยใช้ชุดข้อมูลที่ถูกต้อง หลังการจากชุดฝึกอบรมได้วัดประสิทธิภาพของตัวโมเดล YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 แล้วพบว่า YOLO V8 มีค่า mAP.5 ร้อยละ 98 ในชุด Train และ ร้อยละ 83 ในชุด Test ตามด้วยสถาปัตยกรรม YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ให้ค่า mAP ตามลำดับ การที่ได้ค่า mAP เยอะอาจเป็นเพราะข้อมูลคลาสจรรย์านยนต์ที่มีข้อมูลภาพที่เยอะ และ ตรวจจับได้ถูกเยอะโดยมีจุดเด่นในการจำแนกรถจรรย์านยนต์ ไม่เหมือนข้อมูลยานพาหนะคลาสอื่น ๆ YOLO V5, YOLO V6 และ YOLO V7 มีความเร็วในการฝึกซ้อมใกล้เคียงกัน แต่ YOLO V8 มีประสิทธิภาพเหนือกว่า YOLO V7 รวมทั้งรุ่นก่อนในแง่ของความเร็ว ความเร็วเป็นปัจจัยสำคัญในการจำแนกวัตถุแบบเรียลไทม์ และ YOLO V8 ให้การจำแนกที่รวดเร็วยิ่งขึ้นโดยไม่กระทบต่อความแม่นยำ นอกจากนี้ YOLO V8 ยังแสดงความแม่นยำที่เพิ่มขึ้น โดยเฉพาะอย่างยิ่งในการระบุวัตถุขนาดเล็กเนื่องการพัฒนาสถาปัตยกรรมและเทคโนโลยีของ YOLO V8 ในปัจจุบัน ผลการทดสอบระบุว่า YOLO V8 บรรลุ mAP ที่ใกล้เคียงกันสำหรับการจำแนกประเภทยานพาหนะ เช่น YOLO V5, YOLO V6

และ YOLO V7 ซึ่งในอนาคตสามารถเพิ่มชุดข้อมูลให้กับงานวิจัยเพื่อให้มีประสิทธิภาพ และหลากหลายมากขึ้น ในบทถัดไปจะเป็นการพัฒนาโมเดลสำหรับการจำแนกประเภทของสัญญาณพาหนะ เพื่อทำตามวัตถุประสงค์งานวิจัย

บทที่ 5

การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทของสัญญาณพาด

5.1 บทนำ

ในหัวข้อนี้ใช้ชุดข้อมูลสัญญาณพาด 10 สี ได้แก่ i) สีเทา/เงิน/บรอนซ์, ii) สีดำ, iii) สีขาว, iv) สีแดง, v) สีน้ำเงิน, vi) สีน้ำตาล/เบจ, vii) สีเหลือง, viii) สีเขียว, ix) สีส้ม, และ x) สีอื่น ๆ รวม 10,650 ภาพ โดยชุดข้อมูลการจำแนกสีของสัญญาณพาดได้ใช้เทคนิคเดียวกันกับการจำแนกประเภทของสัญญาณพาด เนื่องการทดลองก่อนหน้าเป็นการจำแนกประเภทของสัญญาณพาด ซึ่งไม่เพียงพอสำหรับระบุลักษณะทางกายภาพของสัญญาณพาด ดังนั้นจึงต้องเพิ่มโมเดลการจำแนกสีของสัญญาณพาดเพื่อให้ครอบคลุมสำหรับการการจำแนกระบุลักษณะทางกายภาพของสัญญาณพาด ในบทนี้จะกล่าวถึงการใช้การถ่ายโอนเรียนรู้เชิงลึกให้กับตัวสถาปัตยกรรมโครงข่ายประสาทเทียมแบบสังวัตนาการของตัว YOLO ทั้ง 6 สถาปัตยกรรม เพื่อเปรียบเทียบว่าสถาปัตยกรรมใดที่มีประสิทธิภาพในการจำแนกสีของสัญญาณพาดดีที่สุดโดยใช้ชุดข้อมูลที่กำหนด

หัวข้อย่อยที่จะกล่าวถึงภายในบทนี้จะประกอบด้วย หัวข้อย่อย 5.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set) รวมทั้งอธิบายถึงชุดข้อมูลที่ใช้ การเตรียมความพร้อมของข้อมูลภาพ อีกทั้งยังกล่าวถึงวิธีการเลือกชุดข้อมูลสำหรับการสร้างตัวจำแนก และ แสดงผลลัพธ์ของงานวิจัยใช้ในบทนี้ หัวข้อย่อย 5.3 การวัด และ ประเมินผล จะกล่าวถึงวิธีการวัด และ ประเมินผลรวม แสดงผลลัพธ์ของงานวิจัย หัวข้อย่อย 5.4 สรุปผลการทดลองภายในบทที่ 5

5.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set)

ชุดข้อมูลภาพประกอบที่ใช้ในงานวิจัยนี้มีชุดข้อมูลสัญญาณพาด 10 สี ได้แก่ i) สีเทา/เงิน/บรอนซ์ 2,160 ภาพ, ii) สีดำ 1,960 ภาพ, iii) สีขาว 1,865 ภาพ, iv) สีแดง 1,563 ภาพ, สีน้ำเงิน 1,530 ภาพ, vi) สีน้ำตาล/เบจ 755 ภาพ, vii) สีเหลือง 344 ภาพ, viii) สีเขียว 263 ภาพ, ix) สีส้ม 110 ภาพ และ x) สีอื่น ๆ 100 ภาพสำหรับชุดข้อมูลสัญญาณพาด และ ทำการแบ่งชุดข้อมูล 3 ชุด i) ชุดข้อมูลฝึกอบรม ร้อยละ 80, ii) ชุดตรวจสอบความถูกต้องร้อยละ 10 และ iii) ชุดทดสอบร้อยละ 10 สำหรับการพัฒนาการจำแนกสีของสัญญาณพาด 5.1 จำนวนชุดข้อมูลภาพประเภทสีของสัญญาณพาดที่ใช้ในการพัฒนาโมเดลสำหรับการจำแนกประเภทของสัญญาณพาด

ตารางที่ 5.1 จำนวนชุดข้อมูลภาพประเภทสีของยานพาหนะที่ใช้ในงานวิจัย

ชุดข้อมูล	ชุดข้อมูล ฝึกอบรม	ชุดตรวจสอบ ความถูกต้อง	ชุดทดสอบ	จำนวนทั้งหมด
สีเทา/เงิน/บรอนซ์	1,728	216	216	2,140
สีดำ	1,568	196	196	1,960
สีขาว	1,492	187	186	1,865
สีแดง	1,251	156	156	1,563
สีน้ำเงิน	1,224	153	153	1,530
สีน้ำตาล/เบจ	604	76	75	755
สีเหลือง	275	34	35	344
สีเขียว	210	26	27	263
สีส้ม	88	11	11	110
สีอื่นๆ	80	10	10	100
จำนวนทั้งหมด	8,520	1,065	1,065	10,652

การเตรียมความพร้อมของภาพ (Image Pre-Processing)

การเตรียมความพร้อมของภาพสำหรับการสร้างชุดข้อมูลสีของยานพาหนะแบ่งสีเป็น 10 สีโดยติดป้ายกำกับข้อมูลเป็นชนิดของสีที่จะทำการพัฒนาโมเดลการจำแนกสีของยานพาหนะ โดยภาพประกอบที่ 5.1 จะแสดงตัวอย่างการติดป้ายกำกับประเภทสีของยานพาหนะ หลังจากนั้นใช้เทคนิคการเพิ่มปริมาณข้อมูลเพื่อเพิ่มความหลากหลายของชุดข้อมูลสีโดยใช้เทคนิค 3 แบบ เช่นเดียวกันกับการพัฒนาการจำแนกประเภทของยานพาหนะ



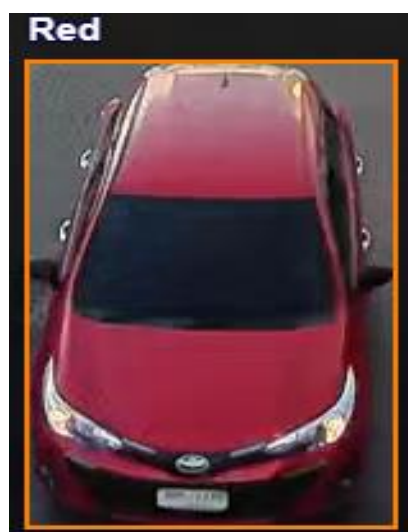
(a) สีเทา/เงิน/บรอนซ์



(b) สีดำ



(c) สีขาว



(d) สีแดง



(e) สีน้ำเงิน



(f) สีน้ำตาล/เบจ



(g) สีเหลือง



(h) สีเขียว



(i) สีส้ม



(j) สีอื่น ๆ

ภาพประกอบที่ 5.1 ภาพตัวอย่างประเภทสีของยานพาหนะ

จากภาพประกอบที่ 5.1 เป็นภาพประกอบประเภทสีของยานพาหนะหลังจากการตัดป้ายกำกับสำหรับสีของยานพาหนะ 10 ประเภท เพื่อที่จะนำไปพัฒนาตัวโมเดลสำหรับการจำแนกประเภทสีของยานพาหนะโดยสีแต่ละสีมีความเด่นชัดของตัวสี ดังนั้นการตัดป้ายกำกับสำหรับสีของยานพาหนะจึงดีกรอบรอบตัวยานพาหนะเพื่อให้ตัวโมเดลสามารถเรียนรู้และจดจำสีของแต่ละสีได้ซึ่งต้องใช้ชุดข้อมูลสีที่มีปริมาณเยอะในการพัฒนาตัวโมเดลการจำแนกสีของยานพาหนะ

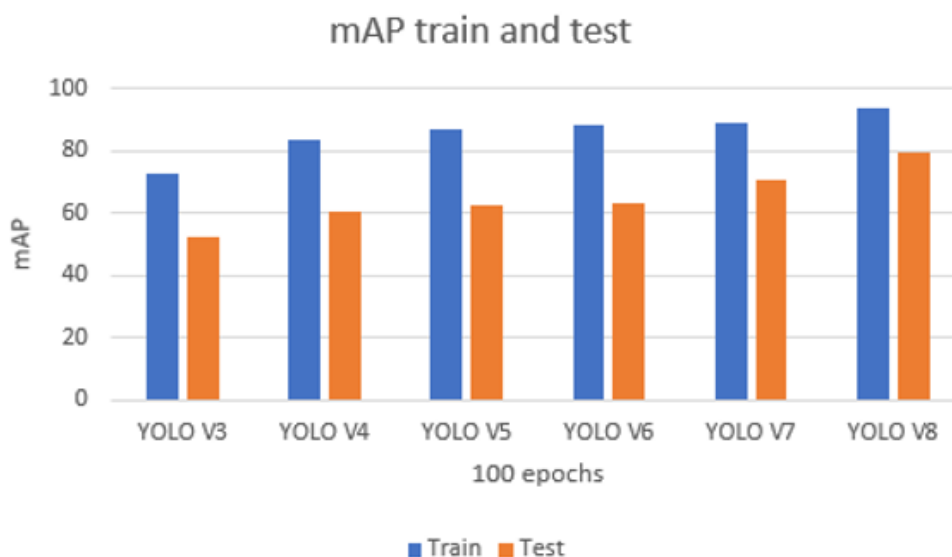
5.3 การวัด และ ประเมินผลตรวจจับประเภทของสียานพาหนะ

การวัด และ การประเมินผลในบทนี้ จะกล่าวถึงกระบวนการสร้างตัวตรวจจับลักษณะทางกายภาพของยานพาหนะจากโครงข่ายประสาทเทียมแบบสังวัตนาการ หรือ YOLO ที่ได้นำเสนอที่จะใช้ในงานวิจัยนี้ จากข้อมูลภาพประกอบที่ได้จากกล้องวงจรปิดจำนวนทั้งหมด 1 ชุดได้แก่ 10,650 ภาพสำหรับชุดข้อมูลประเภทสีของยานพาหนะ แบ่งออกเป็น 3 ชุดข้อมูล คือ i) ชุดข้อมูลฝึกอบรม, ii) ชุดตรวจสอบความถูกต้อง และ iii) ชุดข้อมูลที่ใช้สำหรับทดสอบ ซึ่งแต่ละภาพผ่านกระบวนการเตรียมความพร้อมข้อมูลเรียบร้อยแล้ว และ เพิ่มปริมาณข้อมูลเพื่อเพิ่มคุณภาพของชุดข้อมูล

วัตถุประสงค์ของการทดลองและการประเมินผล เพื่อจะหาโครงข่ายประสาทเทียมแบบสังวัตนาการแบบ YOLO สถาปัตยกรรมไหนเหมาะสมกับชุดข้อมูลที่เตรียมไว้สำหรับการจำแนกประเภทสีของยานพาหนะ ซึ่งในหัวข้อนี้จะอธิบายถึงการวัด และ การประเมินผลของการจำแนกประเภทสีของยานพาหนะ โดยมีการประเมินประสิทธิภาพของโมเดลจากค่าความแม่นยำ ค่าความถูกต้อง ค่าเรียกคืน ของตัวสถาปัตยกรรมทั้ง 6 สถาปัตยกรรม

การจำแนกประเภทสีของยานพาหนะ

ข้อมูลที่ใช้มีทั้งหมด 10,650 ภาพรวมขั้นตอนการเพิ่มปริมาณข้อมูลซึ่งเป็นการเพิ่มความหลากหลายของชุดข้อมูล และ ลดระยะเวลาการฝึกอบรมของตัวโมเดลสำหรับการพัฒนาแบบจำแนกการจำแนกประเภทสีของยานพาหนะที่แบ่งชุดข้อมูลออกเป็น 3 ชุด จากหัวข้อที่ 4.2 และ ตัดป้ายกำกับ 10 สี ได้แก่ i) สีเทา/เงิน/บรอนซ์ 2,160 ภาพ, ii) สีดำ 1,960 ภาพ, iii) สีขาว 1,865 ภาพ, iv) สีแดง 1,563 ภาพ, สีน้ำเงิน 1,530 ภาพ, vi) สีน้ำตาล/เบจ 755 ภาพ, vii) สีเหลือง 344 ภาพ, viii) สีเขียว 263 ภาพ, ix) สีส้ม 110 ภาพ และ x) สีอื่น ๆ 100 ภาพสำหรับชุดข้อมูลสีของยานพาหนะโดยกำหนดการทดลองให้มีรอบการเรียนรู้ (Epochs) จำนวน 100 รอบ กำหนดค่าพารามิเตอร์ ต่าง ๆ ดังนี้ ค่า ขนาด batch size = 32 ขนาดภาพ 416 x 416 โดยมีตารางแสดงประสิทธิภาพการจำแนกประเภทสีของยานพาหนะหลังการฝึกอบรมด้วยชุดข้อมูลการฝึกอบรม และ ชุดทดสอบ โดยสถาปัตยกรรม YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 แสดงไว้ในตารางที่ 5.2 และ โดยแสดงกราฟแท่งในภาพประกอบที่ 5.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทสีของยานพาหนะระหว่างข้อมูลชุด Train และ Test และ



ภาพประกอบที่ 5.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทสี่ของยานพาหนะระหว่างข้อมูลชุด Train และ Test

จากภาพประกอบที่ 5.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทสี่ของยานพาหนะระหว่างแท่งสีฟ้าคือข้อมูลชุด Train และ สีส้มคือข้อมูลชุด Test สถาปัตยกรรม YOLO V8 มีค่า mAP มากที่สุด รองลงมา YOLO V7, YOLO V6, YOLO V5 และ YOLO V4 มีค่า mAP รองลงมาในข้อมูลชุด Train มีค่าใกล้เคียงกัน โดยตัวสถาปัตยกรรม YOLO V3 มีค่า mAP น้อยที่สุดในบรรดา 6 สถาปัตยกรรม ในส่วนค่า mAP ข้อมูลชุด Test ในการจำแนกประเภทของยานพาหนะ สถาปัตยกรรม YOLO V8 ได้ค่า mAP Test มากที่สุดและ YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ตามลำดับซึ่งการจำแนกประเภทสี่ของยานพาหนะ จากกราฟแสดงให้เห็นว่าการพัฒนาโมเดลโดยใช้ชุด Train ที่สร้างขึ้นเป็นไปในเชิงบวก แม้ว่าค่า mAP ของชุด Test จะลดลงเนื่องจากการจำแนกสี่ของยานพาหนะ ซึ่งจะรายละเอียดจะอธิบายในหัวข้อถัดไป

ตารางที่ 5.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และชุดทดสอบ

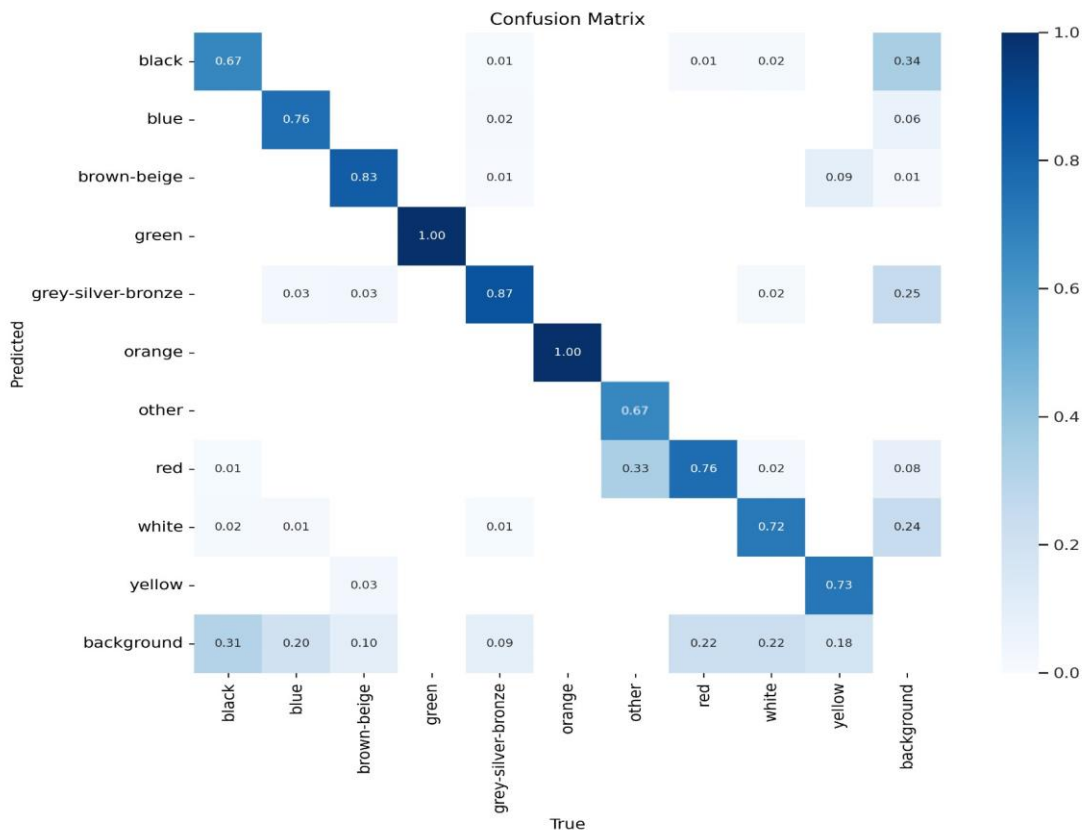
โมเดล	Train	Test
	mAP	mAP
YOLO V3	0.72	0.52
YOLO V4	0.83	0.60
YOLO V5	0.87	0.62
YOLO V6	0.88	0.63
YOLO V7	0.89	0.70
YOLO V8	0.93	0.79

จากตารางที่ 5.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และชุดทดสอบ สถาปัตยกรรม YOLO V8 มีค่า mAP มากที่สุดร้อยละ 93 และ YOLO V7, YOLO V6, YOLO V5 และ YOLO V4 มีค่า mAP รองลงมาตามลำดับในข้อมูลชุด Train นอกจากนี้ YOLO V8 ที่มีค่า mAP ร้อยละ 93 โดยเฉลี่ย สถาปัตยกรรมอื่นมีค่า mAP ประมาณร้อยละ 83 – 89 โดยตัวสถาปัตยกรรม YOLO V3 มีค่า mAP น้อยที่สุดในบรรดา 6 สถาปัตยกรรม ในใส่ส่วนของค่า mAP ข้อมูลชุด Test ในการจำแนกประเภทของยานพาหนะ สถาปัตยกรรม YOLO V8 ได้ค่า mAP Test มากที่สุดโดยมีค่า mAP ร้อยละ 79 และ YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ตามลำดับ ถัดไปแสดงผลลัพธ์ที่ได้เป็นจริงหรือผิดสำหรับการตรวจสอบยานพาหนะที่ไม่เคยเห็นมาก่อน แสดงดังตารางที่ 5.3 และ ภาพประกอบที่ 5.3 แสดงภาพ confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทสีของยานพาหนะ

ตารางที่ 5.3 ผลลัพธ์ที่ได้จำแนกถูกหรือผิดสำหรับการจำแนกประเภทสีของยานพาหนะ

โมเดล คลาส	รูปภาพ ทั้งหมด	YOLO V3		YOLO V4		YOLO V5		YOLO V6		YOLO V7		YOLO V8	
		ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด
สีเทา/ เงิน/ บรอนซ์	100	80	20	85	15	90	10	90	10	90	10	90	10
สีดำ	100	85	15	90	10	90	10	92	8	92	8	93	7
สีขาว	100	80	20	90	10	90	10	90	10	90	10	90	10
สีแดง	100	100	0	100	0	100	0	100	0	100	0	100	0
สีน้ำเงิน	100	75	25	80	20	80	20	80	20	80	20	80	20
สี น้ำตาล/ เบจ	30	20	10	20	10	20	10	20	10	20	10	20	10
สีเหลือง	60	53	7	53	7	53	7	53	7	53	7	53	7
สีเขียว	50	50	0	50	0	50	0	50	0	50	0	50	0
สีส้ม	20	20	0	20	0	20	0	20	0	20	0	20	0
สีอื่น ๆ	10	28	2	28	2	28	2	28	2	28	2	28	2

จากตารางที่ 5.3 แสดงให้เห็นได้ชัดว่า สีเงิน, สีขาว, สีน้ำเงิน, และ สีน้ำตาลจำแนกผิดมากที่สุดในประเภททั้งหมด ส่วนประเภทที่จำแนกถูกที่สุดคือ สีแดง, สีเขียว, และ สีส้มที่ทำการจำแนกถูกมากที่สุดในประเภทของยานพาหนะจากกล้องวงจรปิด ตัว YOLO V8 ถึง YOLO V5 จำแนกได้ใกล้เคียงซึ่งการจำแนกถูกผิดสีของยานพาหนะมีผลลัพธ์ที่ดี ถัดไปการวัดประสิทธิภาพการตรวจของตัวโมเดล แสดงดังตารางที่ 5.4



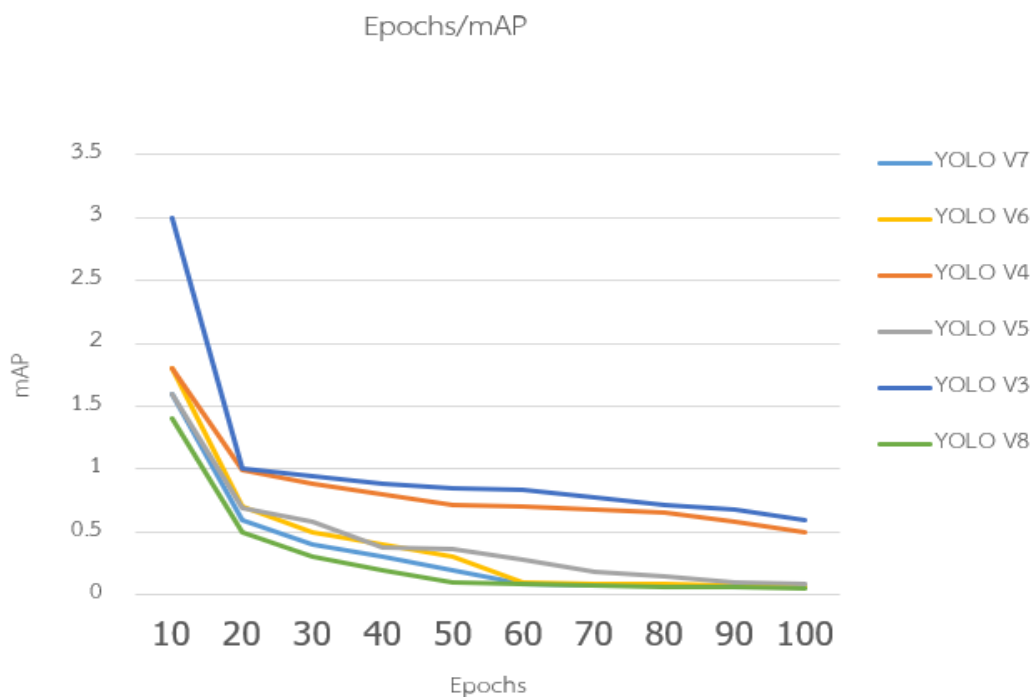
ภาพประกอบที่ 5.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทสีของยานพาหนะ

ตารางที่ 5.4 การประเมินประสิทธิภาพการพัฒนาโมเดลการจำแนกสียานพาหนะ

ผลการวัดประสิทธิภาพ	mAP	Precision	Recall
YOLO V3	0.52	0.68	0.35
YOLO V4	0.60	0.72	0.41
YOLO V5	0.62	0.75	0.48
YOLO V6	0.63	0.75	0.52
YOLO V7	0.70	0.78	0.58
YOLO V8	0.79	0.83	0.62

จากตารางที่ 5.4 แสดงให้เห็นประสิทธิภาพของการจำแนกของแต่ละสถาปัตยกรรม เมื่อพิจารณาค่าประสิทธิภาพมากที่สุด พบว่า YOLO V8 มีค่า mAP.50 ร้อยละ 79 ค่า Precision ค่าร้อยละ 83 และ ค่า Recall ค่าร้อยละ 62 มากที่สุดในบรรดาทั้ง 6 สถาปัตยกรรมผลหลังการจากชุดฝึก

ทดสอบ และ ภาพประกอบที่ 5.4 แสดงกราฟค่า loss ชุดตรวจการสอบความถูกต้องการจำแนกประเภทสี่ของยานพาหนะ



ภาพประกอบที่ 5.4 กราฟ loss ชุดตรวจสอบความถูกต้องของโมเดลการจำแนกประเภทสี่ของยานพาหนะ

จากภาพประกอบที่ 5.4 ค่าการสูญเสียโดยรวมนี้เป็นการสูญเสียของชุดข้อมูลตรวจสอบความถูกต้องของชุดข้อมูลตรวจจับสี่ของยานพาหนะ สำหรับสถาปัตยกรรม YOLO ทั้ง 6 สถาปัตยกรรมในชุดการตรวจสอบความถูกต้องน้อยกว่า 0 คำนี้นับชี้ว่าโมเดลทำงานได้ดีเพียงใดในแง่ของความแม่นยำในการจำแนกออกเบ้จ้กต์และความแม่นยำในการแปลข้อมูลชุดตรวจสอบความถูกต้องทั้งนี้ YOLO V3 และ YOLO V4 อาจมีความแตกต่างในสถาปัตยกรรม ไฮเปอร์พารามิเตอร์ กลยุทธ์การฝึกอบรม และปัจจัยอื่น ๆ ระหว่าง YOLO V3 และ YOLO V4 สามารถนำไปสู่การเปลี่ยนแปลงของค่าการสูญเสียในชุดการตรวจสอบ

5.4 สรุปผล

การจำแนกประเภทของสียานพาหนะโดยใช้ชุดข้อมูลที่กำหนด หลังการจากชุดฝึกอบรม และ วัดประสิทธิภาพของสถาปัตยกรรม YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 แล้วพบว่า YOLO V8 มีค่า mAP ร้อยละ 93 ต่อไป YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLOV3 ตามลำดับในชุดฝึกอบรม และ YOLO V8 มีค่า mAP ร้อยละ 79 ในชุดทดสอบการทดสอบสียานพาหนะ ในระหว่างการทดสอบนี้มีอุปสรรคในการจำแนกสีเงิน สีดำ และสีน้ำตาลของยานพาหนะ ซึ่งอาจเนื่องมาจากความคล้ายคลึงกัน การแก้ไขปัญหานี้อาจจำเป็นต้องเพิ่มชุดข้อมูลด้วยข้อมูลที่เกี่ยวข้องมากขึ้น ในด้านความแม่นยำในการจำแนกสี สีแดงมีความโดดเด่นและตรวจจับด้วยความแม่นยำสูงสุดในเรื่องความเร็วในการฝึกซ้อม YOLO V8 นั้นเร็วที่สุด ซึ่งผลลัพธ์ของ YOLO V3 ทำงานได้ดีกว่าที่คาดไว้ในแง่ของ mAP ที่ร้อยละ 50 ข้อเสียคือความเร็วการฝึก YOLO V3 ยังช้ากว่ารุ่นอื่น ๆ ซึ่งชุดข้อมูลที่ใช้ในการพัฒนามียานพาหนะที่สี และลักษณะคล้ายกันเยอะจึงเป็นปัจจัยทำให้ผลลัพธ์ออกมาได้ดิบถัดไปจะเป็นการพัฒนาโมเดลสำหรับการจำแนกแบรนด์ของยานพาหนะซึ่งเป็นโมเดลสุดท้ายของวัตถุประสงค์ของงานวิจัยนี้

บทที่ 6

การพัฒนาโมเดลสำหรับจำแนกการตรวจจับประเภทแบรนด์ของยานพาหนะ

6.1 บทนำ

ในการทดลองสุดท้ายเป็นการทดลองเพื่อพัฒนาสถาปัตยกรรมในการจำแนกการตรวจจับแบรนด์ของยานพาหนะ เนื่องจากปัจจุบันยังไม่พบเจองานวิจัยที่ใช้สถาปัตยกรรม YOLO ในการจำแนกแบรนด์ของยานพาหนะ และ วัตถุประสงค์ของงานวิจัยนี้คือการพัฒนาสถาปัตยกรรมเพื่อการระบุลักษณะทางกายภาพของยานพาหนะ แบรนด์ที่สนใจในการทดลองมีทั้งหมด 10 แบรนด์ เป็นจำนวนทั้งหมด 1 ชุดได้ทั้งสิ้น 15,849 ภาพ ได้แก่ i) โตโยต้า, ii) อีซูซุ, iii) ฟอर्ड, iv) มาสด้า, v) บีเอ็มดับเบิลยู, vi) มิตซูบิชิ, vii) ฮอนด้า, viii) ซูซูกิ, ix) นิสสัน และ x) อื่น ๆ โดยแบรนด์ที่เลือกมาเป็นแบรนด์ที่นิยมในประเทศไทย จากนั้นนำข้อมูลแบ่งออกเป็น 3 ชุดข้อมูล คือ ชุดข้อมูลที่ใช้สำหรับการฝึกอบรม ชุดตรวจสอบความถูกต้อง และ ชุดที่ใช้สำหรับทดสอบ โดยนำข้อมูลมาพัฒนาสถาปัตยกรรมทั้ง 6 สถาปัตยกรรมของ YOLO และ เปรียบเทียบประสิทธิภาพของทั้ง 6 สถาปัตยกรรมที่ใช้ข้อมูลที่กำหนดว่ามีประสิทธิภาพเท่าใด

หัวข้อย่อยที่จะกล่าวถึงภายในบทนี้จะประกอบด้วย หัวข้อย่อย 6.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set) รวมทั้งอธิบายถึงชุดข้อมูลที่ใช้ การเตรียมความพร้อมของข้อมูลภาพ อีกทั้งยังกล่าวถึงวิธีการเลือกชุดข้อมูลสำหรับการสร้างตัวจำแนก และ แสดงผลลัพธ์ของงานวิจัยใช้ในบทนี้ หัวข้อย่อย 6.3 การวัด และ ประเมินผล จะกล่าวถึงวิธีการวัด และ ประเมินผลรวม แสดงผลลัพธ์ของงานวิจัย หัวข้อย่อย 6.4 สรุปผลการทดลองภายในบทที่ 6

6.2 ข้อมูลที่ใช้ในงานวิจัย (Data Set)

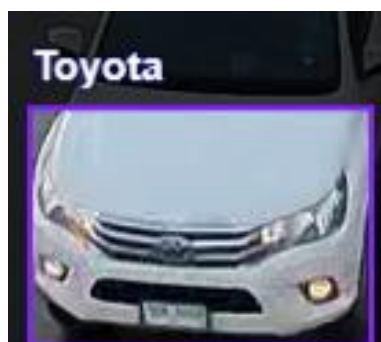
ชุดข้อมูลภาพประกอบที่ การพัฒนาโมเดลสำหรับตรวจจับแบรนด์ของยานพาหนะ 10 แบรนด์ได้แก่ i) โตโยต้า 2,101 ภาพ, ii) อีซูซุ 1,971 ภาพ, iii) ฟอर्ड 1,693 ภาพ, iv) มาสด้า 1,653, v) บีเอ็มดับเบิลยู 1,469 ภาพ, vi) มิตซูบิชิ 1,469 ภาพ, vii) ฮอนด้า 1,854 ภาพ, viii) ซูซูกิ 1,430 ภาพ, ix) นิสสัน 1,763 ภาพและ x) อื่น ๆ 187 ภาพ ซึ่งตารางที่ 6.1 จำนวนชุดข้อมูลภาพแบรนด์ของยานพาหนะที่ใช้ในงานวิจัย รวมทั้งทั้งหมด 15,849 ภาพสำหรับชุดข้อมูลแบรนด์ของยานพาหนะหลังเพิ่มปริมาณข้อมูลทั้ง 3 เทคนิค และ ทำการแบ่งชุดข้อมูลร้อยละ 80 ของชุดข้อมูลสำหรับการฝึกอบรม ชุดตรวจสอบความถูกต้องร้อยละ 10 และ ชุดทดสอบร้อยละ 10 สำหรับการทดสอบตัวโมเดลสำหรับตรวจจับลักษณะทางกายภาพของยานพาหนะ

ตารางที่ 6.1 จำนวนชุดข้อมูลภาพแบรนด์ของยานพาหนะที่ใช้ในงานวิจัย

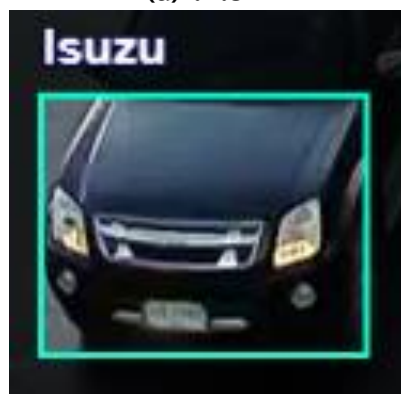
ชุดข้อมูล	ชุดฝึกอบรม	ชุดตรวจสอบความถูกต้อง	ชุดทดสอบ	จำนวนทั้งหมด
โตโยต้า	1,681	210	210	2,101
อีซูซุ	1,577	197	197	1,971
ฟอร์ด	1,355	169	169	1,693
มาสด้า	1,322	166	166	1,653
บีเอ็มดับเบิลยู	1,175	147	147	1,469
มิตซูบิชิ	1,382	173	173	1,728
ฮอนด้า	1,483	186	185	1,854
ซูซูกิ	1,144	143	143	1,430
นิสสัน	1,410	177	176	1,763
อื่น ๆ	150	19	18	187
จำนวนทั้งหมด	12,679	1,587	1,583	15,849

การเตรียมความพร้อมของภาพ (Image Pre-Processing)

การเตรียมความพร้อมของภาพสำหรับการสร้างชุดข้อมูลแบรนด์ของยานพาหนะโดยติดป้ายกำกับแบรนด์ยานพาหนะทั้ง 10 แบรนด์ เพื่อเตรียมนำข้อมูลภาพไปพัฒนาโมเดลการจำแนกแบรนด์ของยานพาหนะโดยภาพประกอบที่ 6.1 จะแสดงตัวอย่างการติดป้ายกำกับประเภทแบรนด์ของยานพาหนะ



(a) โตโยต้า



(b) อีซูซุ



(c) ฮอนด้า



(d) มิตรชูบิชิ



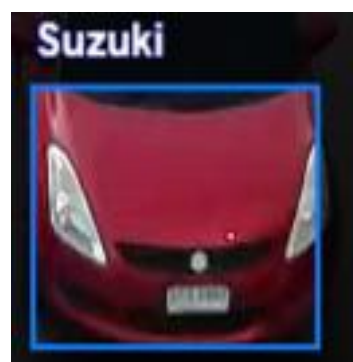
(e) นิสสัน



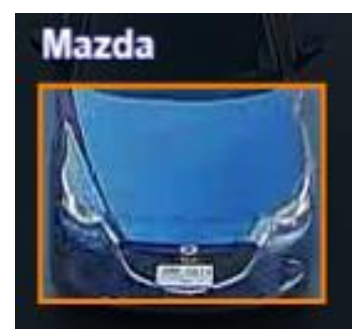
(f) ฟอर्ड



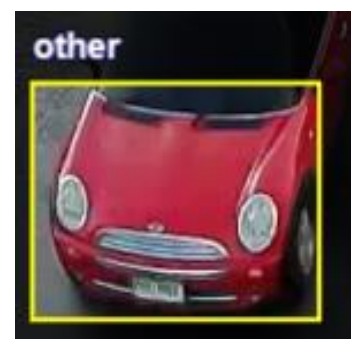
(g) บีเอ็มดับเบิลยู



(h) ซูซูกิ



(i) มาสด้า



(j) แบรินต์อื่น ๆ

ภาพประกอบที่ 6.1 ภาพตัวอย่างแบรนด์ของยานพาหนะ

จากภาพประกอบที่ 6.1 เป็นภาพประกอบประเภทแบรนต์ของยานพาหนะหลังจากการตัดป้ายกำกับทั้ง 10 แรนต์เนื่องจากข้อมูลที่นำมาใช้ในการพัฒนาโมเดลการจำแนกแบรนต์มีความชัดที่ไม่ค่อยดีในการตัดป้ายกำกับโลโก้ของแบรนต์ซึ่งอาจทำให้ตัวของสถาปัตยกรรม YOLO ตรวจจับหรือเรียนรู้ได้ไม่ดีจึงแก้ไขโดยการตัดป้ายกำกับทั้งกระโปรงของยานพาหนะเลย เพื่อที่จะให้ตัวของ YOLO ได้จดจำทั้งโลโก้และช่องลมของยานพาหนะเพราะชุดข้อมูลที่นำมาใช้มียานพาหนะที่มีแบรนต์เดียวกันและรุ่นเดียวกันแต่ต่างกันแต่สีเท่านั้น

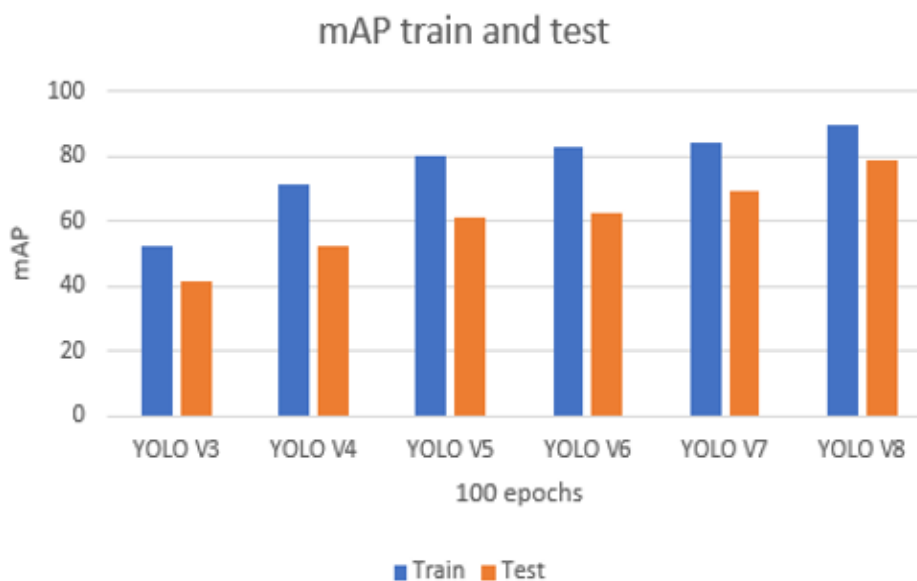
6.3 การวัด และ ประเมินผลการจำแนกประเภทของแบรนต์ยานพาหนะ

การวัด และ การประเมินผลในบทนี้ จะกล่าวถึงกระบวนการสร้างตัวตรวจจับลักษณะทางกายภาพของยานพาหนะจากโครงข่ายประสาทเทียมแบบสังวัตนาการ หรือ YOLO ที่ได้นำเสนอที่จะใช้ในงานวิจัยนี้ จากข้อมูลภาพประกอบที่ ได้จากกล้องวงจรปิด จำนวนทั้งหมด 1 ชุด ได้แก่ 15,849 ภาพสำหรับชุดข้อมูลประเภทยานพาหนะ แบ่งออกเป็น 3 ชุดข้อมูล คือ i) ชุดข้อมูลฝึกอบรม, ii) ชุดตรวจสอบความถูกต้อง และ iii) ชุดข้อมูลที่ใช้สำหรับทดสอบ ซึ่งแต่ละภาพผ่านกระบวนการเตรียมความพร้อมข้อมูลเรียบร้อยแล้ว และ การเพิ่มปริมาณข้อมูลเพื่อเพิ่มคุณภาพของชุดข้อมูล

วัตถุประสงค์ของการทดลองและการประเมินผล เพื่อจะหาโครงข่ายประสาทเทียมแบบสังวัตนาการแบบ YOLO สถาปัตยกรรมไหนเหมาะสมกับชุดข้อมูลที่เตรียมไว้สำหรับการจำแนกประเภทแบรนต์ของยานพาหนะทั้ง 10 แรนต์

การจำแนกแบรนต์ของยานพาหนะ

การพัฒนาโมเดลสำหรับจำแนกตรวจจับแบรนต์ของยานพาหนะใช้ 15,849 ภาพแบรนต์ยานพาหนะ 10 แรนต์ ได้แก่ ได้แก่ i) โตโยต้า 2,101 ภาพ, ii) อีซูซุ 1,971 ภาพ, iii) ฟอर्ड 1,693 ภาพ, iv) มาสด้า 1,653, v) บีเอ็มดับเบิลยู 1,469 ภาพ, vi) มิตซูบิชิ 1,469 ภาพ, vii) ฮอนด้า 1,854 ภาพ, viii) ซูซูกิ 1,430 ภาพ, ix) นิสสัน 1,763 ภาพและ x) แรนต์อื่น ๆ 187 ภาพ สำหรับการพัฒนาโมเดลสำหรับตรวจจับแบรนต์ของยานพาหนะ โดยกำหนดการทดลองให้มีรอบการเรียนรู้ (Epochs) จำนวน 100 รอบ กำหนดค่าพารามิเตอร์ ดังนี้ ค่า ขนาด batch size = 32 ขนาดภาพ 416 x 416 ในตารางที่ 6.2 แสดงตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และ ชุดทดสอบ YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 แสดง และ ภาพประกอบที่ 6.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทแบรนต์ของยานพาหนะระหว่างข้อมูลชุด Train และ Test



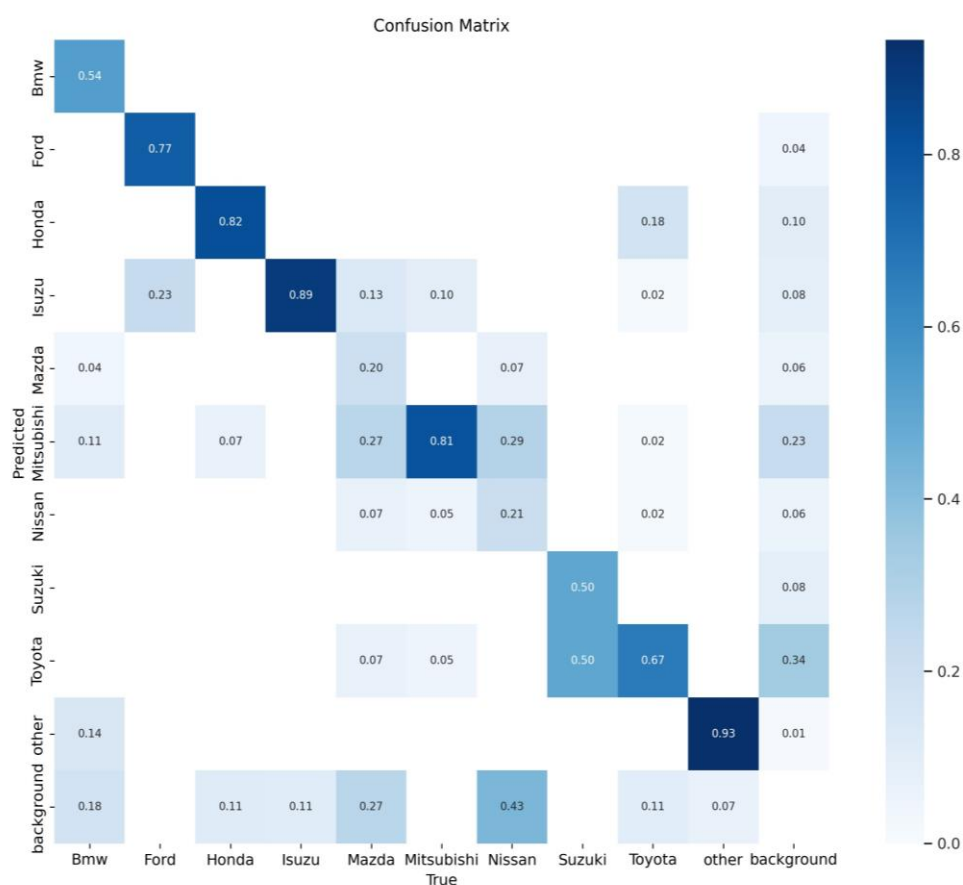
ภาพประกอบที่ 6.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทแบรนด์ของยานพาหนะ ระหว่างข้อมูลชุด Train และ Test

จากภาพประกอบที่ 6.2 กราฟแท่งแสดงค่า mAP ของโมเดลจำแนกประเภทแบรนด์ของยานพาหนะระหว่างแท่งสีฟ้าคือข้อมูลชุด Train และ สีส้มคือข้อมูลชุด Test สถาปัตยกรรม YOLO V8 มีค่า mAP มากที่สุด รองลงมา YOLO V7, YOLO V6, YOLO V5 และ YOLO V4 มีค่า mAP รองลงมาในข้อมูลชุด Train มีค่าใกล้เคียงกัน โดยตัวสถาปัตยกรรม YOLO V3 มีค่า mAP ที่ดีในการจำแนกแบรนด์ที่เป็นทั้งการจำแนกช่องลมและโลโก้ ในส่วนของค่า mAP ข้อมูลชุด Test ในการจำแนกประเภทของยานพาหนะ สถาปัตยกรรม YOLO V8 ได้ค่า mAP Test มากที่สุดและ YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ตามลำดับ ซึ่งได้ผลลัพธ์ดีกว่าที่คาดไว้

ตารางที่ 6.2 ตารางเปรียบเทียบผลทดลองค่าความแม่นยำชุดฝึกอบรม และชุดทดสอบ

โมเดล	Train	Test
	mAP	mAP
YOLO V3	0.52	0.41
YOLO V4	0.71	0.52
YOLO V5	0.80	0.61
YOLO V6	0.82	0.62
YOLO V7	0.84	0.69
YOLO V8	0.89	0.78

จากตารางที่ 6.2 แสดงให้เห็นได้ชัดว่าค่า mAP ในการจำแนกแบรนด์ยานพาหนะ YOLO V8 อยู่ที่ย่อยละ 89 ถัดไป YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ในชุดฝึกอบรมตามลำดับ ในส่วนของค่า mAP ทดสอบ YOLO V8 อยู่ที่ย่อยละ 78 เยอะที่สุดในสถาปัตยกรรมทั้ง 6 สถาปัตยกรรมเป็นผลมาจากโครงสร้างสถาปัตยกรรมที่พัฒนาขึ้นในปัจจุบันซึ่งค่าของ mAP มากขึ้นตามการพัฒนาสถาปัตยกรรมในปัจจุบันเช่นเดียวกับชุดข้อมูลที่ใช้ในการพัฒนาโมเดลทั้งนี้สถาปัตยกรรมอาจจะเหมาะสมกับงานอื่นมากกว่าขึ้นอยู่กับงานที่นำไปใช้ ถัดไปแสดงผลลัพธ์ที่ได้ถูกหรือผิดสำหรับการจำแนกแบรนด์ยานพาหนะซึ่งไม่เคยเห็นมาก่อน แสดงดังตารางที่ 6.3 และภาพประกอบที่ 6.3 แสดงภาพ confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทแบรนด์ของยานพาหนะ



ภาพประกอบที่ 6.3 confusion matrix ของตัวโมเดล YOLO V8 จำแนกการตรวจจับประเภทแบรนด์ของยานพาหนะ

ตารางที่ 6.3 ผลลัพธ์ที่ได้ตรวจจับถูกหรือผิดสำหรับการจำแนกแบรนด์ของยานพาหนะ

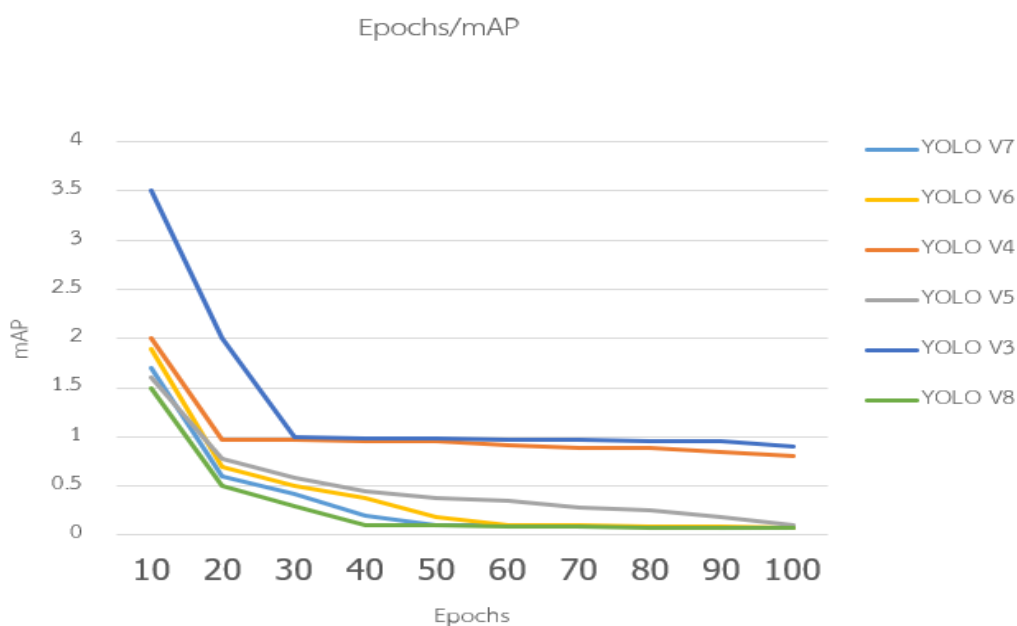
โมเดล คลาส	รูปภาพ ทั้งหมด	YOLO V3		YOLO V4		YOLO V5		YOLO V6		YOLO V7		YOLO V8	
		ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด	ถูก	ผิด
โตโยต้า	100	75	25	79	21	82	12	82	12	82	12	82	12
อีซูซุ	100	80	10	85	15	90	10	90	10	90	10	90	10
ฟอร์ด	100	84	16	84	16	84	16	84	16	84	16	84	16
มาสด้า	100	70	30	70	30	72	28	72	28	72	28	75	25
บีเอ็ม ดับเบิลยู	100	70	30	70	30	70	30	70	30	70	30	74	26
มิตซูบิชิ	100	75	25	82	18	95	5	95	5	95	5	96	4
ฮอนด้า	100	80	20	80	20	85	15	85	15	85	15	86	14
ซูซูกิ	100	48	52	48	52	68	32	68	32	68	32	68	32
นิสสัน	100	70	30	72	28	83	17	83	17	83	17	84	16
อื่น ๆ	30	20	10	20	10	20	10	20	10	20	10	20	10

จากตารางที่ 6.3 แสดงให้เห็นได้ชัดว่าแบรนด์ ซูซูกิ, นิสสัน, มาสด้า, จำแนกผิดมากที่สุด
 ในบรรดาแบรนด์ทั้งหมด ส่วนประเภทที่จำแนกถูกที่สุดคือ มิตซูบิชิ, อีซูซุ, โตโยต้า, ฟอร์ด
 เนื่องจากแบรนด์ที่กล่าวมานี้มีคุณลักษณะที่เด่นชัดในการจำแนกถูกมากที่สุด
 ในประเภทแบรนด์ของยานพาหนะ

ตารางที่ 6.4 การประเมินประสิทธิภาพการพัฒนาโมเดลการจำแนกแบรนด์ของยานพาหนะ

ผลการวัด ประสิทธิภาพ	mAP	Precision	Recall
YOLO V3	0.41	0.64	0.32
YOLO V4	0.52	0.68	0.45
YOLO V5	0.61	0.72	0.52
YOLO V6	0.62	0.79	0.58
YOLO V7	0.69	0.78	0.62
YOLO V8	0.78	0.85	0.72

จากตารางที่ 6.4 แสดงให้เห็นประสิทธิภาพของการจำแนกของแต่ละสถาปัตยกรรม เมื่อพิจารณาค่าประสิทธิภาพมากที่สุด พบว่า YOLO V8 มีค่า mAP.50 ร้อยละ 78 ค่า Precision ร้อยละ 85 และ ค่า Recall ร้อยละ 72 ซึ่ง YOLO V 7 ถึง YOLO V3 มีค่าน้อยลงตามลำดับมากไปน้อย ภาพประกอบที่ 6.4 ค่า loss ชุดตรวจสอบความถูกต้องการจำแนกแบรนด์ของยานพาหนะ



ภาพประกอบที่ 6.4 ค่า loss ชุดตรวจสอบความถูกต้องการจำแนกแบรนด์ของยานพาหนะ

จากภาพประกอบที่ 6.4 ค่าการสูญเสียโดยรวมนี้เป็นการสูญเสียของชุดข้อมูล ตรวจสอบความถูกต้องของชุดข้อมูลตรวจจับแบรนด์ของยานพาหนะ สำหรับสถาปัตยกรรม YOLO ทั้ง 6 สถาปัตยกรรมในชุดการตรวจสอบความถูกต้องน้อยกว่า 0 ค่านี้นับชี้ว่าโมเดลทำงานได้ดีเพียงใด ในแง่ของความแม่นยำในการจำแนกออกเจ็ทและความแม่นยำในการแปลข้อมูลชุดตรวจสอบความถูกต้องทั้งนี้ YOLO V3 อาจมีความแตกต่างในสถาปัตยกรรม ไฮเปอร์พารามิเตอร์ กลยุทธ์การฝึกอบรม และปัจจัยอื่น เช่น ปัญหาการจำแนกโลโก้แบรนด์ที่มีขนาดเล็ก หรือ เกิดขึ้นจากชุดข้อมูลที่ใช้ในการพัฒนาตัวโมเดล

6.4 สรุปผล

การจำแนกประเภทของแบรนด์ยานพาหนะโดยใช้ชุดข้อมูลที่กำหนด หลังผลการฝึกอบรม และ วัดประสิทธิภาพของตัวโมเดล YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8 แล้วพบว่า YOLO V7 มีค่า mAP.50 ที่ร้อยละ 89 ต่อด้วย YOLO V7, YOLO V6, YOLO V5, YOLO V4 และ YOLO V3 ให้ค่า mAP.50 ตามลำดับ ในบรรดาโมเดลต่างๆ YOLO V8 แสดงให้เห็นถึงประสิทธิภาพสูงสุด ในขณะที่ YOLO V3 ทำงานได้ค่อนข้างดี แต่ข้อเสียในการ

โดยเฉพาะอย่างยิ่งในการจำแนกวัตถุขนาดเล็ก ข้อจำกัดนี้ยังคงอยู่แม้ว่าจะใช้ชุดข้อมูลที่มีความละเอียดปานกลางหรือติดป้ายกำกับทั้งด้านหน้าของรถ ในด้านระยะเวลาการฝึก YOLO V8 มีความเร็วการฝึกที่เร็วที่สุดในบรรดารุ่นทั้ง 6 รุ่น ในระหว่างกระบวนการจำแนกแบรนด์ ได้พบอุปสรรคบางประการ การจำแนกแบรนด์ซูซูกิ ถือเป็นอุปสรรคในการทดลองการจำแนกแบรนด์ ในขณะที่ มิตซูบิชิตรวจพบได้ดีกว่า ปัญหาหลักอาจเกิดจากรูปลักษณะที่คล้ายคลึงกันของแบรนด์ โดยโลโก้ทำให้ตัวโมเดลสับสน และ ต้องกำหนดระยะทางการจำแนกประเภทของแบรนด์เพื่อปรับปรุงความแม่นยำ และ ประสิทธิภาพการจำแนกประเภทของแบรนด์ต้องรวบรวมข้อมูลเพิ่มเติมไว้ในกระบวนการฝึกอบรมบดบังไปเป็นการสรุปผลการทดลองทั้งหมด

บทที่ 7

สรุป และ ข้อเสนอแนะ

บทสุดท้ายเป็นสรุปงานวิจัยการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ จากชุดข้อมูลวิดีโอกล้องวงจรปิด ซึ่งจะอธิบายในบทนี้ทั้งหมดโดยหัวข้อ 7.1 เป็นการสรุปผลการทดลองของงานวิจัยนี้ หัวข้อ 7.2 ข้อค้นพบหลักและสิ่งที่ค้นพบใหม่ หัวข้อ 7.3 ปัญหาและอุปสรรค หัวข้อ 7.4 กล่าวการนำไปประยุกต์ใช้ให้เกิดประโยชน์ และ หัวข้อ 7.5 เป็นหัวข้อสุดท้าย จะกล่าวถึงงานในอนาคต

7.1 สรุปผลการทดลอง

จากที่กล่าวมาทั้งหมดในงานวิจัยนี้ มี 2 กระบวนการหลัก ๆ กระบวนการที่ 1 เป็นการศึกษามอเดลจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ และการสร้างชุดข้อมูล หรือ การเตรียมความพร้อมของชุดข้อมูล กระบวนการที่ 2 การพัฒนาตรวจจับลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ โดยใช้การถ่ายโอนการเรียนรู้ และ ชุดข้อมูลที่กำหนดมาเพื่อพัฒนาโมเดลโมเดลตรวจจับลักษณะทางกายภาพของยานพาหนะและ มีการประเมินประสิทธิภาพสถาปัตยกรรมของโครงข่ายประสาทเทียมของโครงข่ายประสาทเทียมแบบสังวัตนาการทั้งหมด 6 สถาปัตยกรรมได้แก่ YOLO V3, YOLO V4, YOLO V5, YOLO V6, YOLO V7 และ YOLO V8

กระบวนการที่ 1 การศึกษาพัฒนาโมเดลตรวจจับลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการของสถาปัตยกรรมที่ใช้ชื่อว่า YOLO ซึ่ง YOLO ได้มีหลักการทำงานการจำแนกประเภทโดยการแยกกล่องขอบเขตเชิงพื้นที่ (bounding box) ที่ของวัตถุที่ถูกรอบไว้โดยใช้โครงข่ายประสาทเทียม สถาปัตยกรรม YOLO มีความคล้ายกับ GoogleNet ที่มีคอนโวลูชัน (Convolution Layers) 24 ชั้นรวมไปถึงชั้นเชื่อมโยงสมบูรณ์ (Fully connected Layers) และ ยังมีชั้น ReLU อยู่ในกระบวนการ จากนั้นการเตรียมชุดข้อมูลสำหรับการพัฒนาโมเดลตรวจจับลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ โดยชุดข้อมูลที่จับแบ่งออกเป็น 3 ชุด ชุดได้แก่ (i) ประเภทของยานพาหนะ (ii) ประเภทสีของยานพาหนะ และ (iii) แบรินด์ของยานพาหนะ โดยการสร้างชุดข้อมูลนี้ได้มาจากการแยกเฟรมของชุดข้อมูลวิดีโอออกเป็นรูปภาพเพื่อสร้างชุดข้อมูลรูปภาพ โดยชุดข้อมูลประเภทของยานพาหนะมี 757 ภาพ ประเภทสีของยานพาหนะมี 3,550 ภาพ และ แบรินด์ของยานพาหนะมี 5,283 ภาพ .ซึ่งมีการใช้เทคนิคการเพิ่มปริมาณข้อมูลเพื่อสร้างความแตกต่างในชุดข้อมูลทำให้มีคุณภาพมากขึ้นเป็น 2,271 ภาพ, 10,650 ภาพ, และ 15,849 ภาพ เรียงตามชุดข้อมูลโดยแต่ละชุดจะแบ่งออกเป็น 3 ชุด คือ (i) ชุดข้อมูลสำหรับฝึกอบรม, (ii) ชุดตรวจสอบความถูกต้อง และ (iii) ชุดข้อมูลสำหรับทดสอบ ชุดข้อมูลสำหรับประเภทของยานพาหนะแบ่งภาพชุดข้อมูลสำหรับการฝึกอบรม 1,817 ภาพ ชุดตรวจสอบความถูกต้อง 227 ภาพ และ ภาพชุดข้อมูลสำหรับการทดสอบ 227 ภาพ ชุดข้อมูลสำหรับ

ประเภทสี่ของยานพาหนะภาพชุดข้อมูลสำหรับการฝึกอบรมแบ่งเป็น 8,520 ภาพ ชุดตรวจสอบความถูกต้อง 1,065 ภาพ และ ภาพชุดข้อมูลสำหรับการทดสอบ 1,065 ภาพ สุดท้ายชุดข้อมูลสำหรับตรวจจับแบรนด์ยานพาหนะเป็น ภาพชุดข้อมูลสำหรับการฝึกอบรม 12,679 ภาพ ชุดตรวจสอบความถูกต้อง 1,587 ภาพ และ ภาพชุดข้อมูลสำหรับการทดสอบ 1,583 ภาพ

กระบวนการที่ 2 การพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะ โดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการโดยใช้การถ่ายโอนการเรียนรู้กับ 6 สถาปัตยกรรมด้วยชุดข้อมูลทั้งหมด 3 ชุดข้อมูล (i) ประเภทของยานพาหนะ (ii) ประเภทสี่ของยานพาหนะ และ (iii) แบรินด์ของยานพาหนะ เพื่อหาสถาปัตยกรรมที่ให้ค่า mAP ดีที่สุดซึ่ง YOLO V8 โดยให้ค่า mAP การจำแนกประเภทของยานพาหนะ ร้อยละ 83 และ ร้อยละ 79 สำหรับการจำแนกสี่ของยานพาหนะ และ ร้อยละ 78 สำหรับการจำแนกแบรนด์ของยานพาหนะโดยใช้ชุดทดสอบ

7.2 ข้อค้นพบหลักและสิ่งที่ค้นพบใหม่

ข้อค้นพบหลักและสิ่งที่ค้นพบใหม่ จะเป็นการอธิบายสิ่งที่ได้จากงานวิจัยนี้ เพื่อตอบคำถามงานวิจัยที่ได้กล่าวไว้ในบทที่ 1

“การพัฒนาโมเดลสำหรับการจำแนกลักษณะทางกายภาพของยานพาหนะจากข้อมูลวิดีโอโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการสามารถทำได้อย่างไร”

1) ขั้นตอนการเตรียมความพร้อมของภาพที่เสนอเพื่อพัฒนาตัวโมเดลการจำแนกที่เหมาะสมสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ

ขั้นตอนวิธีการในการเตรียมความพร้อมของภาพเพื่อหาภาพที่เหมาะสมสำหรับการพัฒนาโมเดลสำหรับการจำแนกลักษณะทางกายภาพของยานพาหนะ โดยใช้การถ่ายโอนการเรียนรู้กับ 3 สถาปัตยกรรมด้วยชุดข้อมูลทั้งหมด 3 ชุดข้อมูล (i) ประเภทของยานพาหนะ (ii) ประเภทสี่ของยานพาหนะ และ (iii) แบรินด์ของยานพาหนะ

2) ขั้นตอนวิธีที่ในการเตรียมข้อมูลวิดีโอก่อนที่จะเริ่มการพัฒนาตัวโมเดลการจำแนกที่เหมาะสมสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ

ขั้นตอนวิธีการในการเตรียมข้อมูลวิดีโอ และ แดกเฟรมภาพเพื่อสร้างชุดข้อมูลที่เหมาะสมสำหรับการพัฒนาตัวโมเดลการจำแนกที่เหมาะสมสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ

3) กลไกการเรียนรู้โครงข่ายประสาทเทียมแบบสังวัตนาการ ที่เหมาะสมที่สุดสำหรับการสร้างโมเดลการจำแนกวัตถุเพื่อระบุลักษณะทางกายภาพของยานพาหนะจากข้อมูลที่กำหนด

การเรียนรู้โครงข่ายประสาทเทียมแบบสังวัตนาการเพื่อให้ทราบถึงองค์ประกอบของโครงข่ายประสาทเทียมแบบสังวัตนาการที่เหมาะสมสำหรับการพัฒนาตัวโมเดลการจำแนกที่เหมาะสมสำหรับการระบุลักษณะทางกายภาพของยานพาหนะ

ผลงานหลักของงานวิจัยที่นำเสนอในวิทยานิพนธ์นี้ที่กล่าวในหัวข้อที่ 1.5 ของบทที่ 1 เพื่อสรุปจึงมีการนำเสนออีกครั้งในบทนี้

1) กระบวนการเก็บรวบรวมข้อมูลวิดีโอการจราจรสำหรับการใช้ในการทำวิจัยจากกล้องวงจรปิดจังหวัดภูเก็ต

2) กระบวนการจัดการวิดีโอการจราจรออกเป็นรูปภาพเพื่อง่ายสำหรับสร้างชุดข้อมูลรูปภาพ

3) กระบวนการเตรียมความพร้อมของภาพการจราจรโดยการปรับแสงในวิดีโอ และความคมชัดเพื่อความเหมาะสมต่อการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

4) กระบวนการเพิ่มปริมาณข้อมูลภาพทางจราจรเพื่อลดระยะเวลาในการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ ดีกว่าการสร้างโมเดลขึ้นมาใหม่

5) การคิดพื้นที่สนใจของยานพาหนะที่สำคัญต่อการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

6) กระบวนการตีขอบเขตที่ทำนายในพื้นที่สนใจ และ การตัดป้ายกำกับเพื่อแบ่งแย่งประเภทสำหรับการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

7) โมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการชื่อว่า “YOLO (You Only Look Once)”

8) กระบวนการพัฒนาโมเดลการจำแนกลักษณะทางกายภาพของยานพาหนะโดยใช้ YOLO ทั้ง 6 สถาปัตยกรรมเพื่อเปรียบเทียบคุณภาพการพัฒนาการจำแนกโดยใช้ชุดข้อมูลที่กำหนด

7.3 ปัญหาและอุปสรรค

ปัญหา และ อุปสรรคที่เกิดขึ้นในงานวิจัยนี้ พบว่าได้มีหลายอย่างที่กระทบต่องานวิจัยดังนี้

1) ข้อมูลที่ใช้ในการทำชุดข้อมูลวิจัยนี้ วิดีโอความยาว 30 นาทีที่ได้จากกล้องวงจรปิด ตำบลเกาะแก้ว ถนนเทพกระษัตรี อำเภอเมืองภูเก็ต คุณภาพความละเอียดของวิดีโอมีส่วนในการสร้างชุดข้อมูลความละเอียดวิดีโอที่ได้รวบรวมมามีความละเอียดค่อนข้างที่จะไม่ดีนัก ทำให้ผู้วิจัยต้องวิเคราะห์หุ้รูปยานพาหนะในรูปภาพที่แยกเฟรมออกมาจากวิดีโอใช้เวลานานมากขึ้น

2) หลังจากการทดลองตัวโมเดลเพื่อตรวจจับลักษณะทางกายภาพของยานพาหนะแล้วพบว่ามีปัญหาส่วนของตัวสียานพาหนะ เนื่องจากสีของตัวยานพาหนะค่อนข้างที่จะคล้ายคลึงกัน ดังนั้นจึงจำเป็นต้องตัดป้ายกำกับให้เยอะมากกว่าประเภทของยานพาหนะ และ แบรินด์ จึงใช้เวลามากในการเตรียมชุดข้อมูลสี

7.4 ข้อจำกัดของงานวิจัย

หัวข้อนี้จะกล่าวถึงข้อจำกัดของงานวิจัยการการระบุลักษณะทางกายภาพของยานพาหนะโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

- 1) โมเดลที่ได้จากการศึกษาครั้งนี้มีประสิทธิภาพกับชุดข้อมูลที่ใช้ในการศึกษาเท่านั้น ซึ่งอาจจะให้ผลลัพธ์แตกต่างไปจากนี้หากนำไปใช้กับชุดข้อมูลจากสถานที่อื่น ๆ
- 2) โมเดลที่ได้จากการศึกษานี้ยังมีข้อจำกัดในการตรวจจับในบางมุมของยานพาหนะในรูปภาพอาจจะตรวจจับไม่ได้
- 3) ข้อมูลภาพที่ใช้ในการศึกษาในงานวิจัยนี้เป็นภาพที่ได้จากกล้องวงจรปิดที่จับในช่วงเวลากลางวันเท่านั้น ดังนั้นโมเดลที่ได้อาจจะไม่เหมาะสมสำหรับการใช้งานตรวจจับภาพในช่วงเวลากลางคืน

7.5 ลักษณะของการนำไปใช้ประโยชน์

การนำไปใช้ประโยชน์ของงานวิจัยนี้ สามารถนำไปพัฒนาต่อยอดหรือเปรียบเทียบสำหรับกับ YOLO สถาปัตยกรรมใหม่ในอนาคตเพื่อเป็นตัวเลือกในการใช้งานในการพัฒนาการจำแนกลักษณะทางกายภาพของยานพาหนะเพื่อที่จะนำไปประยุกต์ใช้จริงในอนาคตสำหรับหน่วยงานราชการในการจำแนกลักษณะทางกายภาพของยานพาหนะ

7.6 งานในอนาคต

หัวข้องานในอนาคตของงานวิจัยนี้สามารถนำไปต่อยอดได้ในอนาคต รวมไปถึงการพัฒนาอุปกรณ์กล้องวงจรปิด เพื่อพัฒนาให้มีประสิทธิภาพ และ แม่นยำมากขึ้นได้ ดังนี้

- 1) งานวิจัยนี้สามารถนำไปพัฒนาตัวโมเดลหรือประยุกต์ได้ โดยจะต้องมีการเพิ่มชุดข้อมูลหรือพัฒนาตัวโมเดลให้สามารถรวมทั้ง 3 ประเภทเข้าด้วยกัน เพื่อให้ตัวกล้องวงจรปิดสามารถตรวจจับลักษณะทางกายภาพของยานพาหนะได้
- 2) กระบวนการพัฒนาตัวโมเดลตรวจจับลักษณะทางกายภาพของยานพาหนะชนิดอื่นได้ด้วยการเพิ่มชุดข้อมูลลงไป และศึกษาการปรับแต่งโครงสร้างของโมเดลเพื่อให้มีประสิทธิภาพมากขึ้น
- 3) สามารถเพิ่มชุดข้อมูลช่วงเวลากลางคืนหรือ ตามสภาพแวดล้อมเพื่อให้โมเดลมีการจำแนกการตรวจจับที่ครอบคลุมมากขึ้น

บรรณานุกรม

- Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B. (2020, June 12). Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN). *Sensors*, 20(12), 3344. doi:10.3390/s20123344
- Alfarraj, O. (2020, October 8). Internet of things with bio-inspired co-evolutionary deep-convolution neural-network approach for detecting road cracks in smart transportation. *Neural Computing and Applications*, 1-16. doi:10.1007/s00521-020-05401-9
- Amritkar, C., & Jabade, V. (2018, August 1). *Image Caption Generation Using Deep Learning Technique*. Retrieved August 4, 2022, from <https://ieeexplore.ieee.org/abstract/document/8697360>
- Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3. In *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*. Muscat: IEEE. doi:10.1109/UVS.2019.8658300
- Bittrich, S., Kaden, M., Leberecht, C., Kaiser, F., Villmann, T., & Labudde, D. (2019). Application of an interpretable classification model on Early Folding Residues during protein folding. *BioData mining*, 12, 1-16. doi:10.1186/s13040-018-0188-2
- Bochkovskiy, A., Wang, C.-Y., & Mark Liao, H.-Y. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. doi:<https://doi.org/10.48550/arXiv.2004.10934>
- Bodyanskiy, Y., Deineko, A., Skorik, V., & Brodetskyi, F. (2022). Deep neural network with adaptive parametric rectified linear units and its fast learning. *International Journal of Computing*, 21(1), 11-18. doi:10.47839/ijc.21.1.2512
- Cai, Y., Luan, T., Gao, H., Wang, H., Chen, L., Li, Y., . . . Li, Z. (2021). Yolov4-5D: An effective and efficient object detector for autonomous driving. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-13. doi:10.1109/tim.2021.3065438
- Chen, S.-H., Wang, C., Tai, I.-H., Weng, K.-P., Chen, Y., & Hsieh, K.-S. (2021, January 1). Modified YOLOv4-DenseNet Algorithm for Detection of Ventricular Septal Defects in Ultrasound Images. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(7), 101-108. doi:10.9781/ijimai.2021.06.001

บรรณานุกรม (ต่อ)

- Chen, Z., Khemmar, R., Decoux, B., Atahouet, A., & Ertaud, J.-Y. (2019). Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility. *In 2019 Eighth International Conference on Emerging Security Technologies (EST)* doi:10.1109/est.2019.8806222
- Chiba, S., & Sasaoka, H. (2021, May 20). Basic Study for Transfer Learning for Autonomous Driving in Car Race of Model Car. *In 2021 6th International Conference on Business and Industrial Research (ICBIR)* doi:10.1109/icbir52339.2021.9465856
- Ćorović, A., Velibor, I., Đurić, S., Marijan, M., & Pavković, B. (2018). The real-time detection of traffic participants using YOLO algorithm. *In 2018 26th Telecommunications Forum (TELFOR)* doi:10.1109/telfor.2018.8611986
- Dai, J., Li, Y., He, K., & Sun, J. (2016, May 20). R-FCN: Object Detection via Region-based Fully Convolutional Networks. *arXiv (Cornell University)*, 29, 379-387.
- Dai, Y., Liu, W., Li, H., & Liu, L. (2020). Efficient foreign object detection between PSDS and metro doors via Deep Neural Networks. *IEEE Access*, 8, 46723-46734. doi:10.1109/access.2020.2978912
- Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255).
- Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). RepVGG: Making VGG-style convnets great again. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. doi:10.1109/cvpr46437.2021.01352
- Faruque, M. O., Hadi, G., & Chengjun, L. (2019). Vehicle classification in video using deep learning. *Machine Learning and Data Mining in Pattern Recognition*, 117-131.
- Fourie, C. (2003, 1 1). Deep learning? What deep learning? *South African Journal of Higher Education*, 17(1). doi:10.4314/sajhe.v17i1.25201
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2019). Fundamental concepts of Convolutional Neural Network. *Intelligent Systems Reference Library*, 519 - 567. doi:10.1007/978-3-030-32644-9_36
- Girshick, R. (2015). Fast r-cnn. *In Proceedings of the IEEE international conference on computer vision*, 1440-1448.

บรรณานุกรม (ต่อ)

- 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*, 580-587.
- gps tracking. (2022, December 13). Retrieved December 2, 2022, from https://www.forthtrack.co.th/2022/12/13/gps_tracking_01/
- Granroth-Wilding, M., & Clark, S. (2016, March 5). What Happens Next? Event Prediction Using a Compositional Neural Network Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1). doi:10.1609/aaai.v30i1.10344
- Grel, T. (2017, February 28). *Region of interest pooling explained*. Retrieved November 27, 2022, from <https://deepsense.ai/region-of-interest-pooling-explained/>
- Grossi, E., & Buscema, M. (2007, December). Introduction to artificial neural networks. *European Journal of Gastroenterology & Hepatology*, 19(12), 1046-1054. doi:10.1097/meg.0b013e3282f198a0
- Gupta, C., Gill, N. S., Gulia, P., & Chatterjee, J. M. (2023). A novel finetuned YOLOv6 transfer learning model for real-time object detection. *Journal of Real-Time Image Processing*, 20(3). doi:10.1007/s11554-023-01299-3
- Haque, A., Sutradhar, I., Rahman, M., Hasan, M., & Sarker, M. (2021, July 31). Convolutional Nets for Diabetic Retinopathy Screening in Bangladeshi Patients. *arXiv (Cornell University)*. doi:10.48550/arXiv.2108.04358
- He, K., Zhang, X., Ren, S., & Sun, J. (2015, September 1). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904-1916. doi:10.1109/tpami.2015.2389824
- 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*, 770-778.
- Huang, Z., Wang, J., Fu, X., Yu, T., Guo, Y., & Wang, R. (2020, June). DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection. *Information Sciences*, 522, 241-258. doi:10.1016/j.ins.2020.02.067
- Jana, A. P., Biswas, A., & Mohana. (2018). Yolo based detection and classification of objects in video records. *In 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. 2448-2452.

บรรณานุกรม (ต่อ)

- Jiang, H., & Learned-Miller, E. (2017). Face Detection with the Faster R-CNN. *In 2017 12th IEEE international conference on automatic face & gesture recognition* 650-657.
- Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022, January 1). A Review of Yolo Algorithm Developments. *Procedia Computer Science*, 1066-1073. doi:10.1016/j.procs.2022.01.135
- Jubayer, F., Soeb, J., Mojumder, A., Paul, M., Barua, P., Kayshar, S., Islam, A. (2021). Detection of mold on the food surface using YOLOv5. *Current Research in Food Science*, 4, 724-728. doi:10.1016/j.crfs.2021.10.003
- Kagaya, H., Aizawa, K., & Ogawa, M. (2014). Food Detection and Recognition Using Convolutional Neural Network. *In Proceedings of the 22nd ACM international conference on Multimedia*, 1085-1088. doi:10.1145/2647868.2654970
- Karlita, T., Adiputra, A. A., & Kusumaningtyas, E. M. (2022). Identification of the effect of different feature samples on handwritten Japanese character recognition using Yolov4. *In 2022 International Electronics Symposium (IES)*, 472-477.
- König, D., Adam, M. J., Jarvers, C., Layher, G., Neumann, H., & Teutsch, M. (2017). Fully Convolutional Region Proposal Networks for Multispectral Person Detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 49-56.
- Kour , A., Yadav, V. K., Maheshwari, V., & Prashar, D. (2013). A Review on Image Processing. *International Journal of Electronics Communication and Computer Engineering*, 4(1), 270-275.
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., . . . Wei, X. (2022, September 7). YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. doi:10.48550/arxiv.2209.02976
- Li, Y., Fan, Q., Huang, H., Han, Z., & Gu, Q. (2023). A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. *Drones*, 7(3), 304. doi:10.3390/drones7050304
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. *In Proceedings of the IEEE international conference on computer vision*, 2980-2988. doi:10.1109/iccv.2017.324

บรรณานุกรม (ต่อ)

- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L. (2014). Microsoft Coco: Common Objects in Context. *In Computer Vision–ECCV 2014: 13th European Conference Zurich, Switzerland, 740-755.* Switzerland, September 6-12, 2014, Proceedings, Part V 13: Springer International Publishing.
- Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for instance segmentation., 8759-8768.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands. Part I 14* , pp. 21-37. Amsterdam, The Netherlands: Springer International Publishing. Retrieved October 11-14, 2016
- Lu, D., & Weng, Q. (2007, March). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823-870. doi:10.1080/01431160600746456
- Lu, Y., Zhang, L., & Xie, W. (2020). Yolo-Compact: An efficient yolo network for single category real-time object detection. *In 2020 Chinese control and decision conference (CCDC)*, 1931-1936.
- M, J. J., B.R, J., K, J., J, J., & S, S. (2017, December 27). Lung Cancer Detection using Artificial Neural Netw. *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, 05(06), 01-05. doi:10.9756/sijcsea/v5i6/05010050101
- Machiraju, G. S., Kumari, K. A., & Sharif, S. K. (2021). Object detection and tracking for community surveillance using transfer learning. *In 2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 1035-1042. doi:10.1109/icict50816.2021.9358698
- Malinowski, M., Rohrbach, M., & Fritz, M. (2015). Ask Your Neurons: A Neural-based Approach to Answering Questions about Images. *In Proceedings of the IEEE international conference on computer vision*, 1-9.
- Mall, P. K., Singh, P. K., & Yadav, D. (2019, December 1). *GLCM Based Feature Extraction and Medical X-RAY Image Classification using Machine Learning Techniques*. Retrieved April 22, 2023, from https://ieeexplore.ieee.org/abstract/document/9066263?casa_token=8FQYHDKuwZMAAAA:KqREvxL2Q8_3qicWpYHy5v2fToc8hjo4KDagT3Y5xk8n-NHg6xZjYjg31K63-TC0XltGNOY

บรรณานุกรม (ต่อ)

- Muhammad Rizwan Munawar. (2023). *YOLOv7 architecture explanation*. Retrieved March 2, 2023, from <https://www.cameralyze.co/blog/yolov7-architecture-explanation>
- Muhammad, A. R., Utomo, H. P., Hidayatullah, P., & Syakrani, N. (2022, April 26). Early Stopping Effectiveness for YOLOv4. *Journal of Information Systems Engineering and Business Intelligence*, 8(1), 11-20. doi:10.20473/jisebi.8.1.11-20
- Naghavi, S. H., & Pourreza, H. (2018, October). Real-time object detection and classification for autonomous driving. In *2018 8th International Conference on Computer and Knowledge Engineering (ICCCKE)*, 274-279. doi:10.1109/iccke.2018.8566491
- Oh, Y., Park, S., & Ye, J. (2020). Deep Learning COVID-19 Features on CXR using Limited Training Data Sets. *IEEE Transactions on Medical Imaging*, 39(8), 2688-2700. doi:10.1109/TMI.2020.2993291
- Ozdemir, M., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R., & Akan, A. (2020, October). Real Time Emotion Recognition from Facial Expressions Using CNN Architecture. In *2019 Medical Technologies Congress (TIPTEKNO)*, 1-4. doi:10.1109/tiptekno.2019.88952152
- Paragios, N., & Zabih, R. (2008). Discrete optimization in computer vision. *Computer Vision and Image Understanding*, 112(1), 1-2. doi:10.1016/j.cviu.2008.09.001
- Prasetyo, E., Suciati, N., & Fatichah, C. (2020). A comparison of Yolo and mask R-CNN for segmenting head and tail of fish. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 1-6. doi:10.1109/icicos.51170.2020.9299024
- Raza, K., & Hong, S. (2020). Fast and Accurate Fish Detection Design with Improved YOLO-v3 Model and Transfer Learning. *International Journal of Advanced Computer Science and Applications*, 11(2). doi:10.14569/ijacsa.2020.0110202
- Redmon, J., & Ali, F. (2018). YOLOv3: An incremental improvement. doi:10.48550/arXiv.1804.02767
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263-7271.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. doi:10.1109/cvpr.2016.91

บรรณานุกรม (ต่อ)

- Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023, May 17). Real-Time Flying Object Detection with YOLOv8. doi:10.48550/arxiv.2305.09972
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149. doi:10.1109/tpami.2016.2577031
- Rosebrock, A. (2016, November 7). *Intersection over Union (IoU) for object detection*. Retrieved April 30, 2022, from PyImageSearch: <https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>
- Sagar, I., Prakash, K. B., & G. R., K. (2021). PyTorch. *Programming with TensorFlow: Solution for Edge Computing Applications*, 87-104.
- Saribas, H., Cevikalp, H., & Kahvecioglu, S. (2018, May, May). Car detection in images taken from unmanned aerial vehicles. *In 2018 26th Signal Processing and Communications Applications Conference (SIU)*, (pp. 1-4). doi:10.1109/siu.2018.8404201
- Shin, H.-C., Roth, H., Gao, M., Lu, L., Xu, Z., Nogues, I., Summers, R. M. (2016, May). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285-1298. doi:10.1109/tmi.2016.2528162
- Shinde, S., Kothari, A., & Gupta, V. (2018). YOLO based Human Action Recognition and Localization. *Procedia computer science*, 133, 831-838. doi:10.1016/j.procs.2018.07.112
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. doi:10.48550/arXiv.1409.1556
- Sowmya, V., & Radha, R. (2021). Heavy-Vehicle Detection Based on YOLOv4 featuring Data Augmentation and Transfer-Learning Techniques. *In Journal of Physics: Conference Series*, 1911(1), 012029. doi:10.1088/1742-6596/1911/1/012029
- Stanimirović, P., Stojanović, I., Katsikis, V., Pappas, D., & Zdravev, Z. (2015). Application of the Least Squares Solutions in Image Deblurring. *Mathematical Problems in Engineering*, 2015, 1-18. doi:10.1155/2015/298689
- Sudhakar, T., & Gavrilova, M. (2020). Cancelable Biometrics Using Deep Learning as a Cloud Service. *IEEE Access*, 8, 112932-112943. doi:10.1109/access.2020.3003869

บรรณานุกรม (ต่อ)

- Sun, Z., & Chen, B. (2022). Research on Pedestrian Detection and Recognition Based on Improved YOLOv6 Algorithm. *In International Conference on Artificial Intelligence in China*. Singapore: Springer Nature Singapore.
- SuperDataScience. (2018, August 18). Retrieved August 17, 2022, <https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection>
- Supreeth, H., & Patil, C. M. (2018). Moving object detection and tracking using deep learning neural network and correlation filter. *In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 1775-1780. doi:10.1109/icicct.2018.8473354
- Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. (2015). Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.
- Tai, S.-K., Dewi, C., Chen, R.-C., Liu, Y.-T., Jiang, X., & Yu, H. (2020, October 7). Deep Learning for Traffic Sign Recognition Based on Spatial Pyramid Pooling with Scale Analysis. *Applied Sciences*, 10(19), 6997. doi:10.3390/app10196997
- Turchenko, V., Chalmers, E., & Luczak, A. (2019, March 31). A Deep convolutional auto-ecocoder with polling – unpooling layers in caffe. *International Journal of Computing*, 8-31. doi:10.47839/ijc.18.1.1270
- Uchida, K., Tanaka, M., & Okutomi, M. (2018, September). Coupled convolution layer for convolutional neural network. *Neural Networks*, 105, 197-205. doi:10.1016/j.neunet.2018.05.002
- Verma, G., & Gupta, P. (2018, December). Wild Animal Detection from Highly Cluttered Images Using Deep Convolutional Neural Network. *International Journal of Computational Intelligence and Applications*, 17(04), 1850021. doi:10.1142/s1469026818500219
- Visa, S., Ramsay, B., Ralescu, A., & Knaap, E. v. (2011). Confusion matrix-based feature selection. *Maics*, 710(1), 120-127.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. (2021). Scaled-YOLOv4: Scaling cross stage partial network. *In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 13029-13038. doi:10.1109/cvpr46437.2021.01283

บรรณานุกรม (ต่อ)

- Wang, C.-Y., Bochkovski, A., & Liao, H.-Y. M. (2023). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7464-7475. doi:10.1109/cvpr52729.2023.00721
- Wang, T., Chen, Y., Qiao, M., & Snoussi, H. (2017, August 15). A fast and robust convolutional neural network-based defect detection model in product quality control. *The International Journal of Advanced Manufacturing Technology*, 94(9-12), 3465-3471. doi:10.1007/s00170-017-0882-0
- Wang, W., & Wang, Y. (2021, May). Underwater Target Detection System based on Yolo V4. *In 2021 2nd International Conference on Artificial Intelligence and Information Systems*, 1-5. doi:10.1145/3469213.3470310
- Wang, Y., Wang, H., & Xin, Z. (2022). Efficient Detection Model of Steel Strip Surface Defects Based on YOLO-V7. *IEEE Access*, 10, 133936-133944. doi:10.1109/ACCESS.2022.3230894
- World Health Organization. (2018). *Global status report on road safety 2018*. World Health Organization. Retrieved March 12, 2021, from <https://www.who.int/publications/i/item/9789241565684>
- Xu, B., Wang, B., & Gu, Y. (2019). Vehicle Detection in Aerial Images Using Modified YOLO. *In 2019 IEEE 19th International Conference on Communication Technology (ICCT)*, 1669-1672. doi:10.1109/icct46805.2019.8947049
- Xu, R., Lin, H., Lu, K., Cao, L., & Liu, Y. (2021, February 13). A Forest Fire Detection System Based on Ensemble Learning. *Forests*, 12(2), 217. doi:10.3390/f12020217
- Yan, B., Fan, P., Lei, X., Liu, Z., & Yang, F. (2021, April 21). A real-time apple targets detection method for picking robot based on improved yolov5. *Remote Sensing*, 13(9), 1619. doi:10.3390/rs13091619
- Yang, Y. (2020). Realization of Vehicle Classification System Based on Deep Learning. *In 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*. doi:10.1109/icpics50287.2020.9202376
- Yun, G.-h., Oh, S.-j., & Shin, S.-c. (2021, December 23). Image preprocessing method in radiographic inspection for automatic detection of ship welding defects. *Applied Sciences*, 12(1). doi:10.3390/app12010123

บรรณานุกรม (ต่อ)

- Zeng, K., & Wang, Y. (2020, March 22). A Deep Convolutional Neural Network for Oil Spill Detection from Spaceborne SAR Images. *Remote Sensing*, 12(6), 1015. doi:10.3390/rs12061015
- Zhang, X., Qiu, Z., Huang, P., Hu, J., & Luo, J. (2018). Application Research of YOLO v2 Combined with Color Identification. In *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 138-1383. doi:10.1109/cyberc.2018.00036
- Zhang, Y., Zhang, W., Yu, J., He, L., Chen, J., & He, Y. (2022, July). Complete and accurate holly fruits counting using YOLOX object detection. *Computers and Electronics in Agriculture*, 198, 107062. doi:10.1016/j.compag.2022.107062
- Zhiqiang, W., & Jun, L. (2017). A review of object detection based on convolutional neural network. In *2017 36th Chinese control conference (CCC)*, 11104-11109.
- จส.100. (2021, February 2). *จส. 100*. Retrieved November 20, 2564 from www.js100.com: https://www.js100.com/en/site/lost_found/view/98050

ภาคผนวก ก

ขั้นตอนการทำงานของ Algorithm Frame Extraction

Algorithm 1: Frame Extraction

```

1: initialize video reader
2: open video file
3: get video width and height
4: set frame rate to 10 fps
5: set video duration to 30 minutes
6: frame_count = frame_rate * (video_duration * 60) // Convert video duration to
seconds
7: frame_interval = 1 / frame_rate
8: output_directory = "path/to/save/frames"
9: for i = 1 to frame_count:
10: current_time = i * frame_interval
11: frame = extract_frame_from_video(current_time)
12: save_frame_as_image(frame, i, output_directory)
13: close video reader

```

รายละเอียดขั้นตอนการทำงานของ Algorithm Frame Extraction

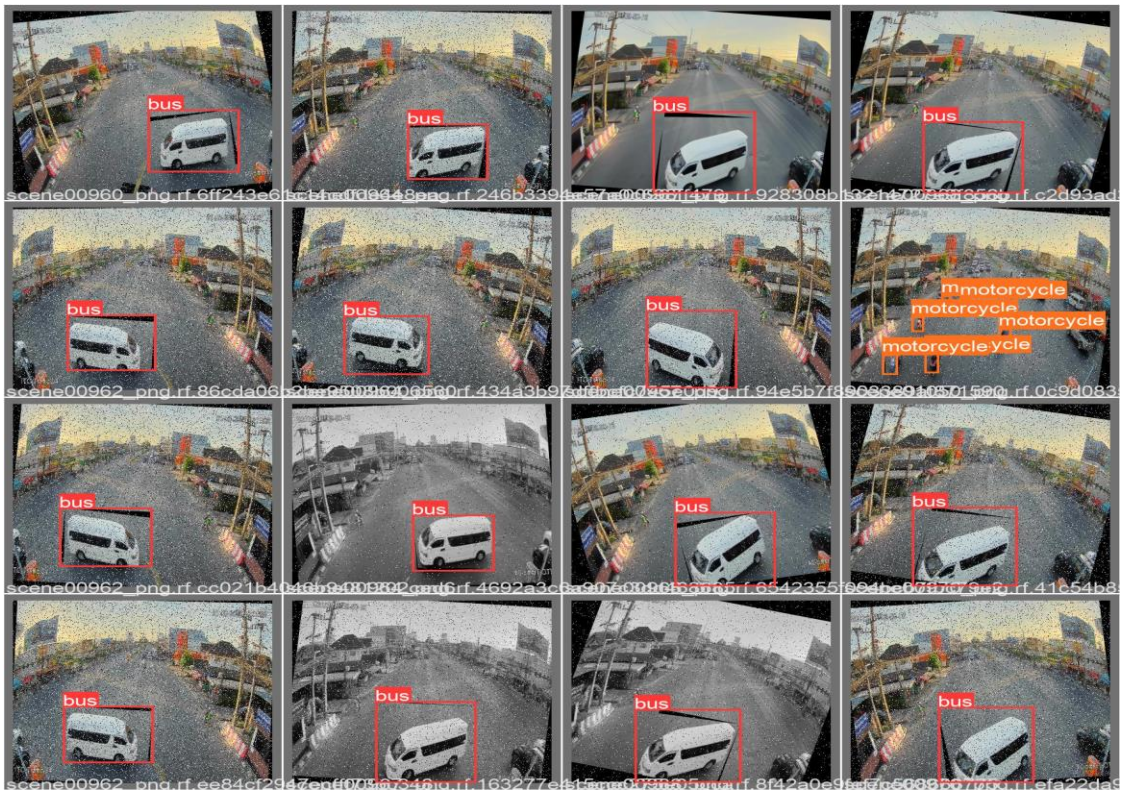
1. เริ่มต้นโปรแกรมอ่านวิดีโอ (Initialize video reader) เริ่มต้นตั้งค่าส่วนประกอบที่จำเป็นเพื่ออ่านและจัดการไฟล์วิดีโอ
2. เปิดไฟล์วิดีโอ (Open video file) โค้ดจะเปิดไฟล์วิดีโอซึ่งหมายความว่า จะสร้างการเชื่อมต่อกับวิดีโอและเตรียมอ่านเนื้อหา
3. รับความกว้างและความสูงของวิดีโอ (Get video width and height) ขั้นตอนนี้ดึงความกว้างและความสูงของเฟรมวิดีโอ
4. ตั้งค่าอัตราเฟรมเป็น 10 fps (Set frame rate to 10 fps) อัตราเฟรมคือจำนวนเฟรมที่แสดงต่อวินาทีในวิดีโอ ในขั้นตอนนี้ตั้งค่าเป็น 10 เฟรมต่อวินาที (fps) ซึ่งหมายความว่าวิดีโอจะถูกประมวลผลและบันทึกในอัตรา 10 เฟรมต่อวินาที
5. ตั้งระยะเวลาวิดีโอเป็น 30 นาที (Set video duration to 30 minutes) ระยะเวลาของวิดีโอถูกตั้งค่าเป็น 30 นาที ข้อมูลนี้จะใช้ในภายหลังเพื่อคำนวณจำนวนเฟรมทั้งหมดในวิดีโอ
6. $frame_count = frame_rate * (video_duration * 60)$ บรรทัดนี้คำนวณจำนวนเฟรมทั้งหมดในวิดีโอ มันคูณอัตราเฟรม (10 fps) ด้วยระยะเวลาของวิดีโอเป็นวินาที (30 นาที * 60 วินาที / นาที) เพื่อกำหนดจำนวนเฟรมที่ต้องแยก

7. `frame_interval = 1 / frame_rate`: ช่วงเวลาเฟรมคำนวณเป็นส่วนกลับของอัตราเฟรม ในกรณีนี้คือ $1/10$ ซึ่งเท่ากับ 0.1 วินาทีต่อเฟรม
8. `output_directory = "path/to/save/frames"`: สิ่งนี้จะตั้งค่าไดเรกทอรีที่เฟรมที่แยกออกมาจะถูกบันทึกเป็นไฟล์รูปภาพที่ด้วยเส้นทางที่ต้องการบันทึกเฟรม "path/to/save/frames"
9. จากนั้นโค้ดจะส่งเข้าสู่ลูปที่วิ่งจากไปยัง `i = 1` ถึง `frame_count` ลูปนี้มีหน้าที่แยกแต่ละเฟรมออกจากวิดีโอ
10. `current_time = i * frame_interval` ขั้นตอนนี้จะคำนวณเวลาเป็นวินาทีสำหรับเฟรมปัจจุบันตามดัชนีลูปและช่วงเวลาเฟรม `i`
11. `frame = extract_frame_from_video(current_time)` ฟังก์ชันนี้มีหน้าที่แยกเฟรมออกจากวิดีโอในเวลาที่กำหนด
12. `save_frame_as_image(frame, i, output_directory)` ฟังก์ชันนี้จะบันทึกเฟรมที่แยกออกมาเป็นไฟล์รูปภาพที่มีชื่อไฟล์ที่สอดคล้องกับหมายเลขเฟรม (`i`) ในไดเรกทอรีเอาต์พุตที่ระบุ `i`
13. หลังจากวนซ้ำผ่านเฟรมทั้งหมดโค้ดจะทำงานถึงขั้นตอนสุดท้าย `close video reader` คำสั่งนี้จะปิดโปรแกรมอ่านวิดีโอที่เล่นอยู่

ภาคผนวก ข
ตัวอย่างรูปภาพที่ใช้โมเดลในการจำแนกทั้ง 3 โมเดล

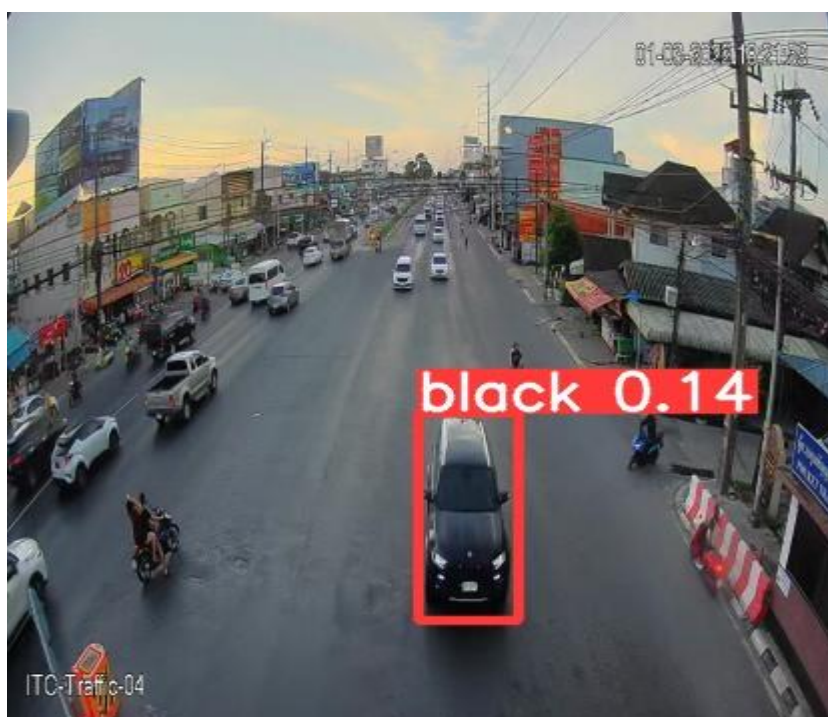
ข1. ตัวอย่างรูปภาพการจำแนกประเภทของยานพาหนะ





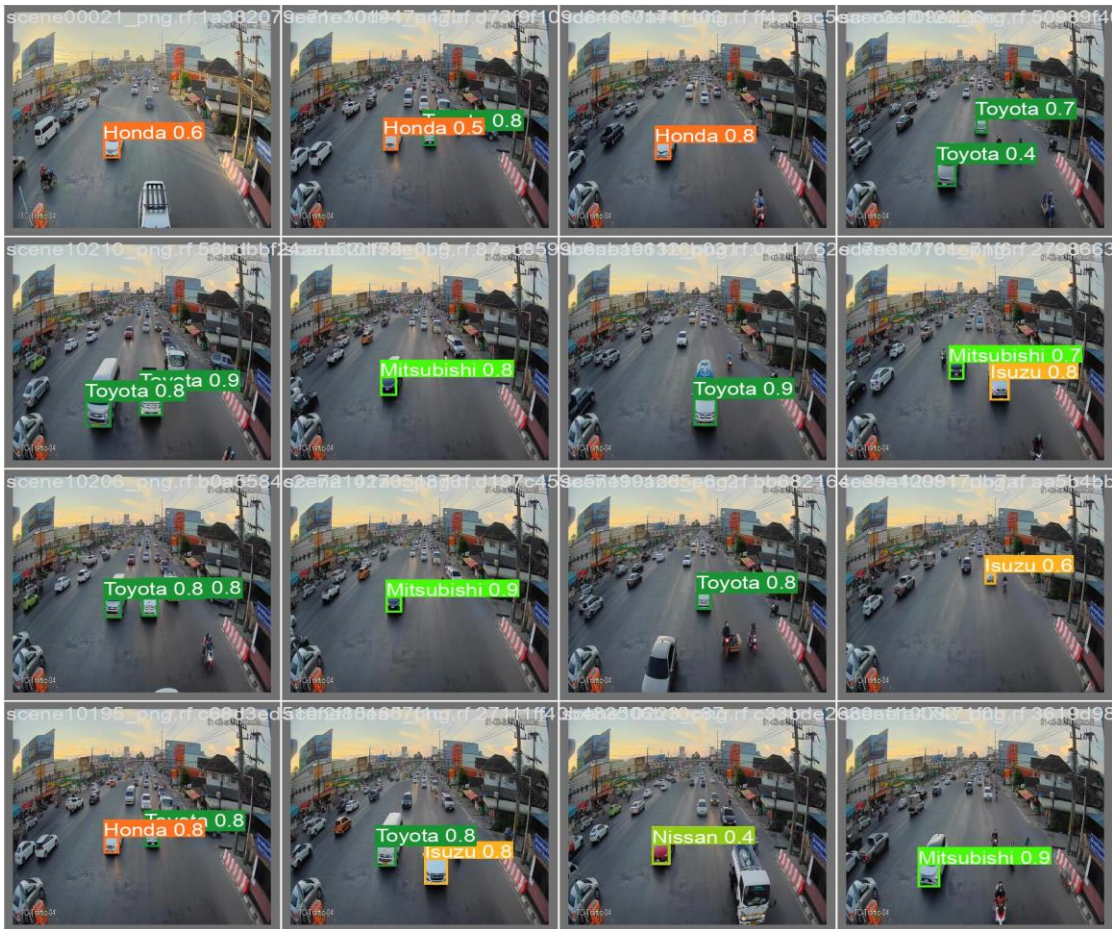
ข2. ตัวอย่างรูปภาพการจำแนกประเภทสีของยานพาหนะ





ข3. ตัวอย่างรูปภาพการจำแนกประเภทแบรนด์ของยานพาหนะ





ประวัติผู้เขียน

ชื่อ สกุล นายคหบดี ประกอบชาติ

รหัสประจำตัวนักศึกษา 6410025005

วุฒิการศึกษา

วุฒิ
เทคโนโลยีบัณฑิต
(เทคโนโลยีปิโตรเลียม)

ชื่อสถาบัน
มหาวิทยาลัยเทคโนโลยี
ราชมงคลศรีวิชัย

ปีที่สำเร็จการศึกษา
พ.ศ. 2562

การตีพิมพ์เผยแพร่ผลงาน

Prakobchat, K., Dittakan, K., & Musikasuwan, S. (2023) A Comparative Study on Vehicle Physical Appearance Identification using Transfer Learning Methods. *Engineering Access*, 9(2), 154-167.