การตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าและการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความเพื่อการตรวจสอบย้อนกลับ

Batch and Logo Object Detection and Batch OCR for Traceability

ทรัพย์ตะวัน วรรณวิจิตร

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Batch and Logo Object Detection and Batch OCR for Traceability

SUBTAWON WONNAWIJIT

A COOPERATIVE EDUCATION SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2022

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	การตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ เพื่อการตรวจสอบย้อนกลับ Batch and Logo Object Detection and Batch OCR for Traceability
ชื่อนักศึกษา	นาย ทรัพย์ตะวัน วรรณวิจิตร รหัสนักศึกษา 62050773
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2565
อาจารย์ที่ปรึกษา	ดร.สกุณา ศรีอโนมัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์) ประจำปีการศึกษา 2565

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ ประธานกรรมการ	
คุณวิศิษฐ์ กิจชัยนุกูล กรรมการ	
ดร.สกุณา ศรีอโนมัย กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	การตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ เพื่อการตรวจสอบย้อนกลับ Batch and Logo Object Detection and Batch OCR for Traceability
ชื่อนักศึกษา	นายทรัพย์ตะวัน วรรณวิจิตร รหัสนักศึกษา 62050773
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2565
อาจารย์ที่ปรึกษา	ดร.สกุณา ศรีอโนมัย

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และเครื่องมือการรู้จำอักขระด้วยแสงที่ใช้ในการอ่าน และแปลงข้อความของหมายเลขสินค้า เพื่อเพิ่มประสิทธิภาพความแม่นยำ และลดระยะเวลาในการประมวลผล จึงได้นำเทคโนโลยีโยโล่เวอร์ชัน 7 มาใช้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า โดยใช้ข้อมูลรูปภาพของสินค้าจำนวน 476 ภาพ แบ่งเป็นข้อมูลฝึกสอน 80 เปอร์เซนต์ (381 ภาพ) และข้อมูลทดสอบ 20 เปอร์เซนต์ (95 ภาพ) ผลลัพธ์จากการทดสอบมีค่าความถูกต้อง 99.2 เปอร์เซนต์ และระยะเวลาในการประมวลผล 0.540 วินาทีต่อรูปภาพ ที่ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวมเท่ากับ 0.5 และนำเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ มาใช้ในการรู้จำอักขระด้วยแสงหมายเลขของสินค้า โดยใช้ข้อมูลรูปภาพของหมายเลขสินค้าจำนวน 200 ภาพ แบ่งเป็นข้อมูลฝึกสอน 80 เปอร์เซนต์ (160 ภาพ) และข้อมูลทดสอบ 20 เปอร์เซนต์ (40 ภาพ) ผลลัพธ์จากการทดสอบมีค่าความถูกต้อง 65.3 เปอร์เซนต์ และระยะเวลาในการประมวลผล 0.435 วินาทีต่อรูปภาพ จากงานวิจัยพัฒนาเครื่องมือการตรวจจับวัตถุ และการรู้จำอักขระด้วยแสง เนื่องจากปัจจุบันบริษัทกรณีศึกษาเครื่องมือที่ใช้มีค่าความถูกต้อง 90 เปอร์เซนต์ และระยะเวลาในการประมวลผลมากกว่า 5 วินาทีต่อรูปภาพ เพื่อนำเป็นต้นแบบสำหรับการพัฒนาต่อยอดให้มีประสิทธิภาพมากยิ่งขึ้น เพื่อใช้ในการตรวจสอบย้อนกลับ เพื่อตรวจสอบ ติดตาม และระบุแหล่งที่มาของสินค้าที่พบปัญหา

คำสำคัญ : การตรวจจับวัตถุ, การรู้จำอักขระด้วยแสง, การตรวจสอบย้อนกลับ

เอกสารนี้ [illegible] ใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Batch and Logo Object Detection and Batch OCR for Traceability
Students	Mr Subtawon Wonnawijit Student ID 62050773
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2022
Advisor	Dr. Sakuna Sriarnomai

Abstract

The aim of this study is to develop instruments used to detect batch, logo and optical character recognition used to read and translating batch. In order to increase efficiency in terms of accuracy and runtime by using YOLOv7, Used to detect batch and logo, using image data of 476 images, which are split data into 80% of training data (381 image) and 20% of testing data (95 image). That result at an Intersection over Union (IoU) ratio of 0.5 with accuracy 99.2% and runtime 0.540 second/image. In terms of optical character recognition using Tesseract OCR, using image data of 200 batch images, which are split data into 80% of training data (160 image) and 20% of testing data (40 image). That result are accuracy 65.3% and runtime 0.435 second/image. Through the study and development of object detection instrument and optical character recognition. Because of the company's current case studies, the instrument applied are accuracy 90% and runtime of more than 5 second/image. To serve as a model for future work to increase effectiveness to be applied to traceability in order to verify, track and identify the source of the faulty product.

Keywords : Object Detection, Optical Character Recognition (OCR), Traceability

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การทำสหกิจศึกษาฉบับนี้สำเร็จได้ เนื่องจากได้รับความเมตตาจากอาจารย์ที่ปรึกษาสหกิจศึกษา ดร. สกุณา ศรีอโนมัย ที่เมตตาให้คำปรึกษา สละเวลาอันมีค่าให้คำแนะนำ อีกทั้งยังช่วยแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่ จนกระทั่งสหกิจศึกษานี้สำเร็จลุล่วงไปด้วยดี ผู้วิจัยทราบซึ้งในความเมตตาจากท่านอาจารย์และกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ ที่ให้เกรียรติเป็นประธานกรรมการสอบสหกิจที่เมตตาให้คำปรึกษา และสละเวลามาตรวจทานแก้ไขข้อบกพร่องต่าง ๆ ตลอดการทำสหกิจฉบับนี้ อีกทั้งคณาจารย์ประจำภาคสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) ที่ให้คำแนะนำที่เป็นประโยชน์แก่ผู้วิจัย มอบวิชาความรู้ และประสบการณ์การเรียนการสอนกับผู้วิจัย

รวมถึงขอขอบคุณ คุณวิศิษฐ์ กิจชัยนุกูล ที่ให้โอกาสแก่ผู้วิจัยได้เข้าร่วมโครงการสหกิจศึกษา และให้ความอนุเคราะห์ข้อมูลรูปภาพสินค้า ในการศึกษาเพื่อให้เกิดประโยชน์กับผู้วิจัยและหน่วยงานที่เกี่ยวข้อง

สุดท้ายนี้ผู้วิจัยขอขอบคุณพระคุณบิดา มารดา ที่สนับสนุนให้กำลังใจแก่ผู้วิจัยในการศึกษาระดับปริญญาตรี ตลอดจนการทำสหกิจศึกษาครั้งนี้สำเร็จลุล่วงไปด้วยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าสหกิจศึกษานี้จะเป็นประโยชน์และจะเป็นแนวทางในการพัฒนาต่อไป

ทรัพย์ตะวัน วรรณวิจิตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ก
บทคัดย่อภาษาอังกฤษ..... ข
กิตติกรรมประกาศ..... ค
สารบัญ..... ง
สารบัญตาราง..... ช
สารบัญรูป..... ซ

บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1
1.2 วัตถุประสงค์ของงานวิจัย..... 2
1.3 ขอบเขตของงานวิจัย..... 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 4
1.5 นิยามศัพท์เฉพาะ..... 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การเรียนรู้ของเครื่อง (Machine Learning) 6
2.2 การเรียนรู้เชิงลึก (Deep Learning) 7
2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN)..... 8
2.4 การตรวจจับวัตถุ (Object Detection)..... 10
2.4.1 R-CNN (Region-based Convolutional Neural Networks) 10
2.4.2 Fast R-CNN กับ Faster R-CNN 11
2.5 อัลกอริทึมโยโล่ (You Only Look Once: YOLO)..... 11
2.5.1 อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7)..... 14
2.6 การรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR)..... 15
2.7 เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)..... 16
2.8 ภาษาโปรแกรมไพทอน (Python)..... 16
2.9 ไลบรารี่ (Library) ที่เกี่ยวข้องในงานวิจัย..... 17
2.9.1 Tensorflow..... 17
2.9.2 Matplotlib..... 17
2.9.3 Numpy (Numerical Python) 17
2.9.4 OpenCv (Open Source Computer Vision Library)..... 18
2.9.5 Pillow..... 18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

2.9.6 PyYAML..... 18
2.9.7 requests..... 18
2.9.8 SciPy..... 18
2.9.9 Torch..... 19
2.9.10 torchvision..... 19
2.9.11 tqdm..... 19
2.10 การตรวจสอบย้อนกลับ (Traceability) 19
2.11 ค่าประเมอนประสิทธิภาพแบบจำลอง..... 21
2.11.1 ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) 21
2.11.2 ค่าความถูกต้อง ค่าความแม่นยำ และค่าความระลึก (Accuracy, Precision and Recall)..... 22
2.11.3 ค่าวัดประสิทธิภาพ (F1-Score) 23
2.11.4 ค่าความแม่นยำเฉลี่ย (Average Precision) 23
2.11.5 ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER)..... 24
2.11.6 ระยะเวลาในการประมวลผล (Runtime) 24
2.12 งานวิจัยที่เกี่ยวข้อง..... 24
บทที่ 3 วิธีการดำเนินงานวิจัย
3.1 เครื่องมือที่ใช้ในงานวิจัย..... 27
3.2 ขั้นตอนการดำเนินงานวิจัย..... 32
3.2.1 ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุ..... 32
3.2.2 ขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสง..... 34
3.3 การเก็บรวบรวมข้อมูล..... 36
3.3.1 ชุดข้อมูลสำหรับการตรวจจับวัตถุ (Object Detection)..... 36
3.3.2 ชุดข้อมูลสำหรับการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) 38
3.4 การเตรียมข้อมูล..... 38
3.4.1 การกำกับประเภทของวัตถุบนรูปภาพสำหรับการตรวจจับวัตถุ..... 38
3.4.2 การกำกับตัวอักษรของข้อความบนรูปภาพสำหรับการรู้จำอักขระด้วยแสง..... 42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

3.5 การแบ่งชุดข้อมูล..... 44
3.5.1 การแบ่งชุดข้อมูลสำหรับการตรวจจับวัตถุ..... 44
3.5.2 การแบ่งชุดข้อมูลสำหรับการรู้จำอักขระด้วยแสง..... 46
3.6 ขั้นตอนในการสร้างแบบจำลองการตรวจจับวัตถุ..... 47
3.6.1 การฝึกสอนแบบจำลอง..... 47
3.6.2 ทดสอบประสิทธิภาพของแบบจำลองในการตรวจจับหมายเลขและโลโก้ของสินค้า..... 52
3.7 ขั้นตอนในการสร้างแบบจำลองการรู้จำอักขระด้วยแสง..... 53
3.7.1 การติดตั้งเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)..... 53
3.7.2 การฝึกสอนแบบจำลองเทสเซอร์แรคต์ (Tesseract)..... 54
3.7.3 ทดสอบประสิทธิภาพของแบบจำลองในการรู้จำข้อความของหมายเลขสินค้า..... 57
3.8 การทำงานของแบบจำลองการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง..... 60

บทที่ 4 ผลการวิจัย

4.1 การประเมินแบบจำลองการตรวจจับวัตถุ (Object Detection)..... 63
4.2 การประเมินแบบจำลองการรู้จำอักขระด้วยแสง (OCR)..... 72
4.3 การอภิปรายผล..... 76

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย..... 78
5.2 ข้อจำกัด..... 80
5.3 ข้อเสนอแนะ..... 80

เอกสารอ้างอิง..... 82
ภาคผนวก ก. 86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
3.1	จำนวนของรูปภาพสำหรับการตรวจจับวัตถุ	38
3.2	จำนวนของวัตถุแต่ละประเภทสำหรับการตรวจจับวัตถุ	39
3.3	แสดงจำนวนของรูปภาพสำหรับการรู้จำอักขระด้วยแสง	42
3.4	จำนวนของรูปภาพในอัตราส่วน 80:20 สำหรับการตรวจจับวัตถุ	44
3.5	จำนวนของวัตถุแต่ละประเภทในแต่ละชุดข้อมูลสำหรับการตรวจจับวัตถุ	45
3.6	แสดงจำนวนของรูปภาพในอัตราส่วน 80:20 สำหรับการรู้จำอักขระด้วยแสง	47
4.1	รายละเอียดจำนวนวัตถุของแต่ละประเภทวัตถุ	63
4.2	รายละเอียดการตรวจจับของหมายเลขสินค้า	65
4.3	รายละเอียดการตรวจจับของโลโก้สินค้ารูปแบบที่ 1	67
4.4	รายละเอียดการตรวจจับของโลโก้สินค้ารูปแบบที่ 2	69
4.5	ตารางแสดงผลลัพธ์ของการทดสอบแบบจำลองการตรวจจับวัตถุทั้ง 3 ประเภทโดยใช้ชุดข้อมูลสำหรับการทดสอบ	71
4.6	รายละเอียดระยะเวลาที่ใช้ในการประมวลผลการตรวจจับ	72
4.7	ตารางแสดงค่าความถูกต้อง (Accuracy) และค่าความผิดพลาดระดับตัวอักษร (CER) ของการรู้จำตัวอักษรกับตัวอักษรที่ถูกต้อง (Ground Truth)	73
4.8	ตัวอย่างผลลัพธ์การทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 1	75
4.9	ตัวอย่างผลลัพธ์การทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 2	75
4.10	รายละเอียดระยะเวลาที่ใช้ในการประมวลผลการรู้จำหมายเลขสินค้า	76

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
1.1	ตัวอย่างประเภทของรูปภาพ 3 รูปแบบ	2
1.2	ตัวอย่างรูปแบบของตัวอักษร 2 รูปแบบ	3
2.1	โครงสร้างของโครงข่ายประสาทเทียม (Artificial Neural Networks)	8
2.2	ตัวอย่างการทำงานชั้นคอนโวลูชัน (Convolutional Layer)	8
2.3	ตัวอย่างการทำงานของชั้นพูลลิ่ง (Pooling Layer)	9
2.4	ตัวอย่างการทำงานชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer)	9
2.5	ตัวอย่างของกล่องแบบสี่เหลี่ยม (Boundary Box) ในการตรวจจับวัตถุ	10
2.6	ตัวอย่างขั้นตอนการตรวจจับวัตถุในภาพของอัลกอริทึมโยโล่ (YOLO)	12
2.7	ตัวอย่างสถาปัตยกรรมของอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7 Architecture)	14
2.8	ขั้นตอนการทำงานของ เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)	16
2.9	กระบวนการตรวจสอบย้อนกลับ	20
2.10	วิธีการคำนวณค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (IoU)	22
3.1	โลโก้ของภาษาไพทอน (Python)	28
3.2	หน้าต่างโปรแกรม Anaconda	28
3.3	หน้าต่างการทำงานของโปรแกรม Jupyter notebook	28
3.4	หน้าต่างการทำงานของโปรแกรม Visual Studio Code	29
3.5	หน้าต่างการทำงานของเว็บไซต์ Make Sense	29
3.6	หน้าต่างการทำงานของโปรแกรม Qt Box Editor	30
3.7	โลโก้ของโยโล่เวอร์ชันที่ 7 (YOLOv7)	30
3.8	โลโก้ของเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)	31
3.9	ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุ	32
3.10	ขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสง	34
3.11	ตัวอย่างรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1	36
3.12	ตัวอย่างรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1	37
3.13	รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2	37
3.14	ตัวอย่างรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1	38
3.15	ตัวอย่างรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2	38
3.16	ตัวอย่างขั้นตอนการใช้งานหน้าเว็บไซต์ Make Sense (1)	40
3.17	ตัวอย่างขั้นตอนการใช้งานหน้าเว็บไซต์ Make Sense (2)	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่		หน้า
3.18	ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (3)	41
3.19	ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (4)	41
3.20	ตัวอย่างไฟล์ .xml ที่ได้จากเว็บไซต์ Make Sense	42
3.21	ตัวอย่างไฟล์ .box	43
3.22	ตัวอย่างภาพโปรแกรม Qt Box Editor	44
3.23	ตัวอย่างหน้าเว็บไซต์กิตฮับ YOLOv7	48
3.24	Jepyter notebook จากโปรแกรม Anaconda	48
3.25	คำสั่ง git clone ทำการดาวน์โหลดโยโล่เวอร์ชันที่ 7 (YOLOv7)	48
3.26	ติดตั้งไลบรารี่ (Library) ที่จำเป็นทั้งหมดโดยภายไฟล์ที่ชื่อว่า “requirements.txt	49
3.27	รูปตัวอย่างโฟลเดอร์สำหรับการฝึกสอบ และสำหรับการทดสอบ	49
3.28	รูปตัวอย่างโฟลเดอร์รูปภาพ ป้ายกำกับ	50
3.29	ไฟล์ coco.yaml	50
3.30	ตัวอย่างของการกำหนดค่าต่าง ๆ จากไฟล์ coco.yaml	50
3.31	ชุดน้ำหนักที่ได้รับการฝึกสอบล่วงหน้าสำหรับแบบจำลองโยโล่เวอร์ชันที่ 7	51
3.32	ตัวอย่างคำสั่งเริ่มกระบวนการฝึกสอบแบบจำลอง	52
3.33	กำหนดเป็น TP เมื่อค่า IoU > 0.5 และมีการทำนายประเภทวัตถุที่ถูกต้อง	52
3.34	กำหนดเป็น FP เมื่อค่า IoU < 0.5 และมีการทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพ	53
3.35	กำหนดเป็น FN เมื่อค่า IoU > 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้อง	53
3.36	ตัวอย่างหน้าต่างแก้ไข Environment	54
3.37	ตัวอย่างไฟล์นามสกุล .tr	54
3.38	Unicharset	55
3.39	Font_properties	55
3.40	Shapetable, inttemp, pffmtable, normproto	56
3.41	แก้ไขชื่อไฟล์ทั้ง 4 ไฟล์ข้อมูล	56
3.42	ไฟล์นามสกุล .traineddata	57
3.43	ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 1	57
3.44	ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 2	58
3.45	ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 3	59
3.46	แผนผังการทำงานของการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง	61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่		หน้า
3.47	ตัวอย่างผลลัพธ์ที่ได้จากการทำงานของการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง.....	62
4.1	ผลลัพธ์จากโปรแกรมทดสอบการตรวจจับวัตถุ.....	63
4.2	กำหนดเป็น TP เมื่อค่า IoU > 0.5 และมีการทำนายประเภทวัตถุที่ถูกต้อง.....	64
4.3	กำหนดเป็น FP เมื่อค่า IoU < 0.5 และมีการทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพ	65
4.4	กำหนดเป็น FN เมื่อค่า IoU > 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้อง.....	65
4.5	รูปภาพตัวอย่างที่ 1 ของการตรวจจับหมายเลขสินค้า.....	67
4.6	รูปภาพตัวอย่างที่ 2 ของการตรวจจับหมายเลขสินค้า.....	67
4.7	รูปภาพตัวอย่างของการตรวจจับโลโก้สินค้ารูปแบบที่ 1.....	68
4.8	รูปภาพตัวอย่างของการตรวจจับโลโก้สินค้ารูปแบบที่ 2.....	70
4.9	ผลลัพธ์จากโปรแกรมทดสอบการรู้จำอักขระด้วยแสง.....	72

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากในปัจจุบันแนวโน้มธุรกิจ และอุตสาหกรรมไทยกำลังเติบโตอย่างต่อเนื่อง (ธนาคารกรุงศรีอยุธยา, 2565) การตรวจสอบย้อนกลับ (Traceability) เป็นกระบวนการที่สำคัญ และจำเป็นในภาคอุตสาหกรรมที่เกี่ยวข้องกับการผลิต และจัดจำหน่ายสินค้า เพราะผู้บริโภคมีความคาดหวังกับคุณภาพของสินค้า โดยเฉพาะอย่างยิ่งกับสินค้าประเภทอาหาร และเครื่องดื่มต้องมีความปลอดภัย และมีคุณภาพสำหรับการบริโภค การระบุ และตรวจสอบสินค้าที่เคลื่อนย้ายในระบบจัดการสินค้าเป็นปัจจัยสำคัญในการประสบความสำเร็จในการจัดการสินค้า และควบคุมคุณภาพ นอกจากนี้ยังเป็นองค์ประกอบสำคัญในการเพิ่มประสิทธิภาพ และเสริมสร้างความไว้วางใจในกระบวนการขนส่ง และจัดเก็บสินค้า การตรวจสอบย้อนกลับสามารถทำได้โดยใช้หมายเลขสินค้า และโลโก้สินค้าเป็นตัวระบุ หมายเลขสินค้าเป็นรหัสที่กำหนดให้แก่สินค้าเพื่อระบุ และแยกแยะสินค้า โลโก้สินค้าเป็นสัญลักษณ์ที่ใช้แทนตราสินค้าหรือผู้ผลิตสินค้า (มาดีนา และกนกวรรณ, 2556)

ในปัจจุบันบริษัทกรณีศึกษาที่ผู้วิจัยได้รับอนุเคราะห์ในการเข้าร่วมโครงการสหกิจศึกษา ทางบริษัทได้มีการใช้เครื่องมือการตรวจจับวัตถุ (Object Detection) ที่ใช้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ที่ใช้ในการอ่าน และแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ โดยได้ทำการพัฒนาด้วยเทคโนโลยีเทนเซอร์โฟล (Tensorflow) และมีประสิทธิภาพของเครื่องมือที่ค่าความถูกต้อง (Accuracy) 90 เปอร์เซนต์ พบข้อจำกัดด้านระยะเวลาในการประมวลผลที่ใช้ระยะเวลามากกว่า 5 วินาทีต่อภาพ งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือการตรวจจับวัตถุ (Object Detection) จึงได้นำเทคโนโลยีโยโล่เวอร์ชันที่ 7 (YOLOv7) (WongKinYiu, 2022) มาใช้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ได้นำเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) (Amitdo, 2016) มาใช้ในการรู้จำอักขระด้วยแสงของหมายเลขของสินค้าในการอ่าน และแปลงข้อความของหมายเลขสินค้า จากการนำเทคโนโลยีโยโล่เวอร์ชันที่ 7 (YOLOv7) และเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) มาใช้ในการพัฒนาเครื่องมือ เพื่อเพิ่มประสิทธิภาพในเรื่องของค่าความถูกต้อง (Accuracy) และลดระยะเวลาที่ใช้ในการประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อพัฒนาเครื่องมือการตรวจจับวัตถุ (Object Detection) เพื่อตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าในรูปภาพ โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองที่มากกว่า 90 เปอร์เซนต์

2. เพื่อพัฒนาเครื่องมือการตรวจจับวัตถุ (Object Detection) เพื่อตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าในรูปภาพ โดยมีระยะเวลาในการประมวลผลที่น้อยกว่า 5 วินาทีต่อภาพ

3. เพื่อพัฒนาเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เพื่อแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบของข้อความ โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองที่มากกว่า 90 เปอร์เซนต์

4. เพื่อพัฒนาเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เพื่อแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบของข้อความ โดยมีระยะเวลาในการประมวลผลที่น้อยกว่า 5 วินาทีต่อภาพ

1.3 ขอบเขตของงานวิจัย

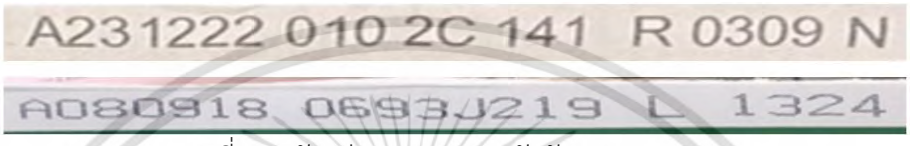
1.3.1 ขอบเขตด้านข้อมูลที่ใช้ในงานวิจัย

1. ชุดข้อมูลที่ใช้สำหรับการตรวจจับวัตถุ (Object Detection) ชุดข้อมูลที่ใช้สำหรับการฝึกสอนแบบจำลอง และการทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection) เป็นรูปภาพของสินค้าจำนวน 476 ภาพ ซึ่งรูปภาพทั้งหมดเป็นภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x640 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG มีประเภทของวัตถุที่ต้องการตรวจจับ 3 ประเภท คือ หมายเลขของสินค้า โลโก้ของสินค้ารูปแบบที่ 1 และโลโก้ของสินค้ารูปแบบที่ 2 ซึ่งมีประเภทของรูปภาพ 3 รูปแบบ โดยมีรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 มี 10 ภาพ รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1 มี 442 ภาพ และรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2 มี 24 ภาพ

รูปที่ 1.1 ตัวอย่างประเภทของรูปภาพ 3 รูปแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ชุดข้อมูลที่ใช้สำหรับการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ชุดข้อมูลที่ใช้สำหรับการฝึกสอนแบบจำลอง และการทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เป็นรูปภาพที่มีเฉพาะหมายเลขของสินค้าจำนวน 200 ภาพ ซึ่งรูปภาพทั้งหมดเป็นภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x40 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG โดยมีรูปแบบของตัวอักษร 2 รูปแบบ โดยรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 จำนวน 100 ภาพ และรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 จำนวน 100 ภาพ



รูปที่ 1.2 ตัวอย่างรูปแบบของตัวอักษร 2 รูปแบบ

1.3.2 ขอบเขตด้านอุปกรณ์และเครื่องมือที่ใช้ในการพัฒนา

1. ด้านซอฟแวร์
    1.1 ภาษาโปรแกรมไพทอน (Python)
    1.2 โปรแกรม Anaconda
        1.2.1 Jupyter notebook
        1.2.2 Visual Studio Code (VS Code)
    1.3 เว็บไซต์ Make Sense
    1.4 โปรแกรม Qt Box Editor
    1.5 อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7)
    1.6 เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)
    1.7 ไลบรารี่ (Library) ที่ใช้
        1.7.1 matplotlib
        1.7.2 numpy
        1.7.3 opencv-python
        1.7.4 Pillow
        1.7.5 PyYAML
        1.7.6 requests
        1.7.7 scipy
        1.7.8 torch

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7.9 torchvision

1.7.10 tqdm

2. เครื่องคอมพิวเตอร์ที่มีคุณสมบัติดังนี้

CPU: Intel Core i5-9400F CPU @ 2.90GHz

GPU: NVIDIA GeForce RTX 2060

System type: 64-bit Operating System, 64 based processor

OS: Windows version 10

Memory (Ram): 16 GB

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นการทดสอบว่าเทคโนโลยีโยโล่เวอร์ชันที่ 7 (YOLOv7) สามารถที่จะตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า โดยมีประสิทธิภาพของค่าความถูกต้อง (Accuracy) ที่มากกว่าเครื่องมือที่ใช้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ที่ใช้อยู่ในปัจจุบันได้

2. เพื่อเป็นการทดสอบว่าเทคโนโลยีโยโล่เวอร์ชันที่ 7 (YOLOv7) สามารถที่จะตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า โดยใช้ระยะเวลาในการประมวลผลที่น้อยกว่าเครื่องมือที่ใช้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ที่ใช้อยู่ในปัจจุบันได้

3. เพื่อเป็นการทดสอบว่าเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) สามารถที่จะอ่าน และแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ โดยมีประสิทธิภาพของค่าความถูกต้อง (Accuracy) ที่มากกว่าเครื่องมือที่ใช้ในการอ่าน และแปลงข้อความของหมายเลขสินค้า ที่ใช้อยู่ในปัจจุบันได้

4. เพื่อเป็นการทดสอบว่าเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) สามารถที่จะอ่าน และแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ โดยใช้ระยะเวลาในการประมวลผลที่น้อยกว่าเครื่องมือที่ใช้ในการอ่าน และแปลงข้อความของหมายเลขสินค้า ที่ใช้อยู่ในปัจจุบันได้

1.5 นิยามศัพท์เฉพาะ

การตรวจสอบย้อนกลับ (Traceability) คือความสามารถในการตรวจสอบ และติดตามแหล่งที่มาของวัตถุดิบ บรรจุภัฒฑ์ หรือส่วนที่เกี่ยวข้อง ซึ่งรวมถึงทั้งขั้นตอนการผลิต กระบวนการขนส่งสินค้าจากผู้ผลิตจนถึงผู้บริโภค (มาดีนา และกนกวรรณ, 2556)

การตรวจจับวัตถุ (Object Detection) คือเทคโนโลยีในการตรวจจับ และระบุตำแหน่งของวัตถุที่สนใจว่าเป็นวัตถุอะไรพร้อมทั้งระบุตำแหน่งว่าวัตถุนั้นอยู่ที่ตำแหน่งไหนภายในภาพ (ปฏิญญา และคณะ, 2564)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) คือกระบวนการแปลงข้อความจากรูปภาพหรือภาพสแกนเป็นตัวอักษรที่สามารถแก้ไขได้ในคอมพิวเตอร์ (จักรพันธ์, 2561)

โลโก้ (Logo) คือเครื่องหมาย ตราหรือสัญลักษณ์ทางการค้า ที่ผ่านการออกแบบเพื่อสร้างเอกลักษณ์แก่สินค้า และบริการให้กับบริษัท ห้างร้าน องค์กร หรือผู้ผลิต (พจนานุกรม ฉบับราชบัณฑิตยสถาน, 2555)

อุตสาหกรรม (Industry) คือกิจกรรมที่ใช้ทุน และแรงงานเพื่อผลิตสิ่งของหรือจัดให้มีบริการ เช่น อุตสาหกรรมการผลิต อุตสาหกรรมท่องเที่ยว (พจนานุกรม ฉบับราชบัณฑิตยสถาน, 2554)

ภาษาโปรแกรมไพทอน (Python) คือภาษาโปรแกรมที่ใช้ในการเขียนโปรแกรมคอมพิวเตอร์ โดยมีลักษณะที่สะดวก และเข้าใจง่ายสำหรับบุคคลที่ไม่มีความรู้หรือประสบการณ์ในการเขียนโปรแกรมมาก่อน (เรือนขวัญ, 2563)

แบบจำลอง (Model) คือแบบจำลองความคิดสิ่งที่ถูกสร้างขึ้นเพื่อใช้ในการทำนายหรือการประมวลผลข้อมูล (วิกิพีเดีย, 2565)

ชุดแบบ หรือรูปแบบของตัวอักษร (Font) คือตัวอักษรที่มีความแตกต่างกัน ทั้งด้านของขนาด ความหนา ความกว้าง และความเอียง (ธวัชชัย, 2549)

ไลบรารี่ (Library) คือการรวบรวมชุดคำสั่งโค้ดที่เก็บฟังก์ชันการทำงานที่เฉพาะทาง (วิกิพีเดีย สารานุกรมเสรี, 2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ได้อธิบายถึงแนวคิด ทฤษฎี และหลักการที่เกี่ยวข้องของการตรวจจับวัตถุ (Object Detection) และการรู้จำอักขระด้วยแสง (Optical Character Recognition) เพื่อนำประยุกต์ใช้ในขั้นตอนต่าง ๆ เพื่อพัฒนากระบวนการในส่วนของการตรวจจับหมายเลขสินค้าและโลโก้ของสินค้า และการรู้จำอักขระด้วยแสงข้อความของหมายเลขสินค้า

2.1 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ที่จะช่วยให้โปรแกรมเกิดการเรียนรู้จากชุดข้อมูลที่จะสามารถทำนายผลได้อย่างแม่นยำ โดยหลักการคือจะมีโมเดลทางคณิตศาสตร์ และมีอัลกอริทึมในหลากหลายรูปแบบ มาเรียนรู้จากชุดข้อมูลจนทำให้คอมพิวเตอร์เกิดการเรียนรู้ และฉลาดในการตัดสินใจบนข้อมูลที่ไม่เคยเห็นมาก่อนได้ โดยการเรียนรู้ของเครื่อง (Machine Learning) แบ่งออกเป็น 3 ประเภท ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised Learning) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) (ศิวกร และวราพร, 2565)

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือการเรียนรู้แบบมีการกำกับ (Labeled Data) ซึ่งประกอบไปด้วยข้อมูลรับเข้า (Input Data) และเป้าหมาย (Target) หรือผลลัพธ์ที่ต้องการให้แบบจำลองทำนาย (Output) โดยแบบจำลอง (Model) จะเรียนรู้จากข้อมูลที่รับเข้า และผลลัพธ์เป้าหมาย เพื่อหาความสัมพันธ์ระหว่างข้อมูลที่รับเข้า และผลลัพธ์เป้าหมาย โดยการเรียนรู้แบบมีการกำกับจะมีการใช้ข้อมูลเป้าหมายในการทดสอบผลของแบบจำลอง (Model) ว่าทำนายได้ถูกต้องหรือไม่ และใช้ผลการทดสอบเหล่านี้ในการปรับปรุงแบบจำลอง (Model) เพื่อให้มีประสิทธิภาพในการทำนายที่ดีขึ้น

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือการเรียนรู้แบบไม่มีการกำกับ (Labeled Data) เข้ามาเทียบกับผลลัพธ์ที่ได้ เมื่อมีการให้ข้อมูลเข้ามาในระบบ (Input Data) ระบบจะทำการเรียนรู้เองว่าข้อมูลเหล่านั้นมีลักษณะแบบไหน โดยจะพยายามหารูปแบบ (Pattern) หรือโครงสร้างภายในของข้อมูลเหล่านั้น เพื่อแบ่งกลุ่มข้อมูลที่มีลักษณะคล้ายกันเข้าด้วยกัน ซึ่งจะช่วยในการทำการจัดกลุ่ม (Clustering) ข้อมูลหรือสกัดคุณลักษณะ (Feature Extraction) ของข้อมูลได้ โดยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ไม่มีคำตอบที่ถูกต้องหรือผิดในขั้นตอนการเรียนรู้ เนื่องจากไม่มีการกำกับข้อมูลที่ถูกต้องมาเทียบ

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) คือการเรียนรู้แบบการลองผิดลองถูกซึ่งทำให้ระบบสามารถเรียนรู้จากประสบการณ์ ซึ่งระบบจะได้รับการสั่งซ้ำ ๆ โดยระบบจะหาวิธีใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเพิ่มประสิทธิภาพของการทำงานโดยดำเนินการทดลอง และตัดสินใจโดยอ้างอิงจากผลการทำงานที่เกิดขึ้นก่อนหน้า กล่าวคือการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นกระบวนการเรียนรู้ที่เน้นการตัดสินใจในสถานการณ์ และการเรียนรู้จากประสบการณ์ก่อนหน้าเพื่อทำให้เกิดผลลัพธ์ที่ดีขึ้นโดยไม่มีการกำหนดขั้นตอนการเรียนรู้ไว้ล่วงหน้า (มาณวิกา, 2562)

2.2 การเรียนรู้เชิงลึก (Deep Learning)

การเรียนรู้เชิงลึก (Deep Learning) เป็นส่วนหนึ่งในการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) ในการเรียนรู้ และทำนายข้อมูล โดยการเรียนรู้เชิงลึก (Deep Learning) มุ่งเน้นที่การสร้างแบบจำลอง (Model) ที่มีความซับซ้อน เพื่อที่จะสามารถจัดการกับปัญหาที่ซับซ้อนได้ การเรียนรู้เชิงลึก (Deep Learning) มีประสิทธิภาพมากในการประมวลผลข้อมูลที่มีขนาดใหญ่ และซับซ้อน เช่น การจำแนกภาพ การตรวจจับวัตถุ และการแปลภาษา (พิชยุทธ และยุทธนา, 2564)

โครงข่ายประสาทเทียม (Artificial Neural Networks) เป็นแบบจำลองทางคณิตศาสตร์ที่มีการจำลองการทำงานของระบบประสาทในสมองของมนุษย์ โครงข่ายประสาทเทียม (Artificial Neural Networks) มีโครงสร้างที่คล้ายกับโครงข่ายประสาทในสมองของมนุษย์ ประกอบด้วยหน่วยประมวลผลที่เรียกว่าโหนด (Nodes) ซึ่งเชื่อมต่อกันเป็นชั้น (Layer) ต่าง ๆ สามารถเรียนรู้ และเก็บความรู้ให้อยู่ในรูปแบบค่าน้ำหนัก (Weight) ซึ่งสามารถปรับเปลี่ยนค่าได้เมื่อมีการเรียนรู้สิ่งใหม่ ๆ เข้ามา โครงสร้างพื้นฐานของโครงข่ายประสาทเทียมประกอบด้วย 3 ชั้น (Layer) หลักดังนี้

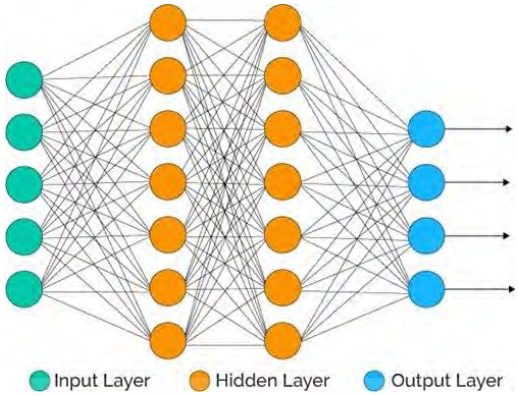
1. ชั้นนำเข้า (Input Layer) รับข้อมูลเข้าสู่โครงข่าย แต่ละโหนด (Nodes) ในชั้นนี้แทนคุณลักษณะหรือคุณสมบัติของข้อมูลนำเข้า

2. ชั้นซ่อน (Hidden Layer) เป็นชั้นที่ต่อจากชั้นนำเข้า และสามารถมีหลายชั้นได้ แต่ละโหนดในชั้นซ่อน จะประมวลผลของข้อมูลจากชั้นก่อนหน้า และส่งผลลัพธ์ไปยังชั้นถัดไป

3. ชั้นส่งออก (Output Layer) เป็นชั้นสุดท้ายของโครงข่าย ที่รับข้อมูลมาจากชั้นก่อนหน้า และส่งผลลัพธ์เป็นค่าส่งออกที่สอดคล้องกับปัญหาหรืองานที่กำหนดไว้

การทำงานของโครงข่ายประสาทเทียม (Artificial Neural Networks) เกิดขึ้นจากการให้โครงข่ายเรียนรู้จากชุดข้อมูลที่ให้มา โดยมีกระบวนการที่เรียกว่าการฝึกสอน (training) เพื่อปรับค่าน้ำหนัก (Weight) และการเชื่อมต่อของโหนด (Nodes) เพื่อให้โครงข่ายสามารถทำนายผลลัพธ์ของข้อมูลที่ไม่เคยเห็นมาก่อนได้ โครงข่ายประสาทเทียม (Artificial Neural Networks) มีความสามารถในการแก้ไขปัญหาที่ซับซ้อน และมีความยืดหยุ่นสูง เช่น การจดจำรูปภาพ การแปลภาษา การจำแนกประเภทข้อมูล และการคาดเดาผลลัพธ์ในอนาคต การเรียนรู้เชิงลึก (Deep Learning) เป็นเทคนิคที่ใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) เพื่อแก้ไขปัญหาที่ซับซ้อนอย่างมีประสิทธิภาพ (พิชยุทธ และยุทธนา, 2564)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

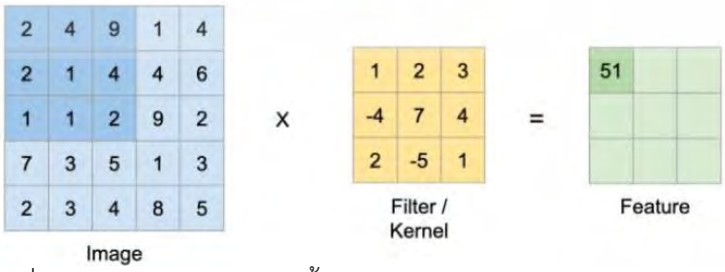


รูปที่ 2.1 โครงสร้างของโครงข่ายประสาทเทียม (Artificial Neural Networks)

2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN) คือแบบจำลองโครงข่ายประสาทเทียมที่ถูกออกแบบมาเพื่อใช้ในการจดจำภาพ และวิดีโอ โดยมีความเหมือนกับโครงข่ายประสาทเทียมปกติ แต่มีการเพิ่มเติมชั้นแบบพิเศษขึ้นมาเพื่อรองรับการแยกแยะภาพ และวิดีโอที่ซับซ้อนและมีขนาดใหญ่ โดยการออกแบบ โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ได้ถูกสร้างขึ้นโดยอิงจากโครงสร้างของสมอง และระบบประสาทที่มนุษย์มี โดยเฉพาะการทำงานของเซลล์ประสาทในส่วนของการจดจำภาพ โดยทั่วไปโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) จะประกอบไปด้วย 3 ชั้นที่สำคัญ ได้แก่ชั้นคอนโวลูชัน (Convolutional Layer) , ชั้นพูลลิ่ง (Pooling Layer) และชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) (ธนภัทร์, 2559)

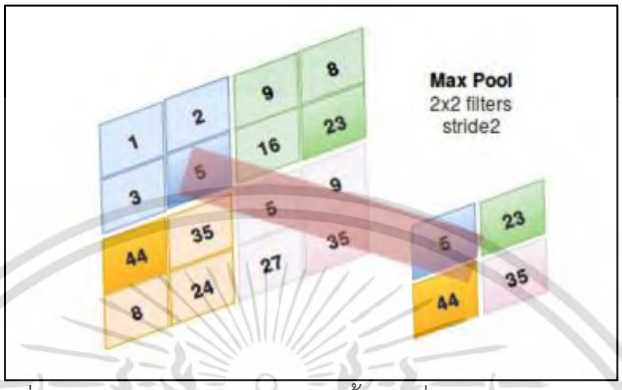
1. ชั้นคอนโวลูชัน (Convolutional Layer) เป็นการคำนวณเพื่อหาผลลัพธ์ของนิวรอล (Neuron) ที่เชื่อมต่อมาจากพื้นที่ส่วนย่อย (Local Region) ของรูปภาพด้วยวิธีการคอนโวลูชัน ซึ่งเป็นการคำนวณการคูณแบบจุด (Dot Product) กับเคอร์เนล (Kernel) แบบจัตุรัสขนาดเล็ก เช่น 3x3, 5x5 หรือ 7x7 เป็นต้น ผลลัพธ์ที่ได้จากการคำนวณเรียกว่าฟีเจอร์แมพ (Feature Map)



รูปที่ 2.2 ตัวอย่างการทำงานชั้นคอนโวลูชัน (Convolutional Layer)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

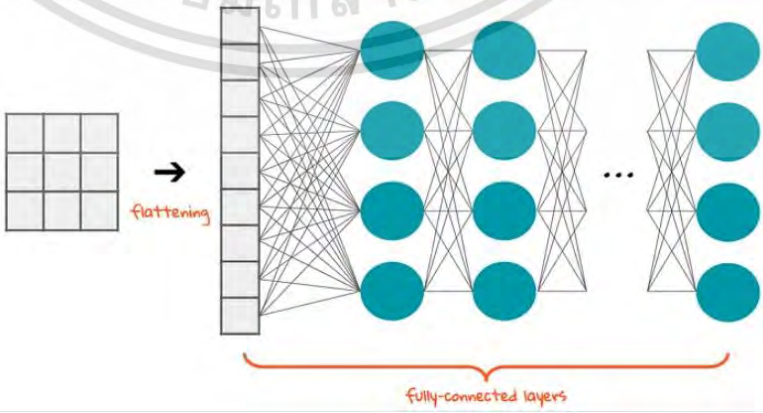
2. ชั้นการรวม (Pooling Layer) เป็นชั้นที่อยู่คั่นกลางระหว่างชั้นคอนโวลูชัน โดยมีวัตถุประสงค์เพื่อลดขนาด (Downsampling) ของฟีเจอร์แมพ (Feature Map) ให้มีขนาดเล็กลงซึ่งสามารถใช้ฟังก์ชันค่าเฉลี่ย ฟังก์ชันค่าต่ำสุด และฟังก์ชันค่าสูงสุดในการคำนวณ ดังนั้นหากเลือกใช้ฟังก์ชันค่าสูงสุดในการคำนวณ จะเรียกว่าการรวมโดยใช้ค่ามากสุด (Max Pooling)



รูปที่ 2.3 ตัวอย่างการทำงานของชั้นพูลลิ่ง (Pooling Layer)

ดังนั้นการทำการรวมโดยใช้ค่ามากสุด (Max Pooling) สามารถทำได้โดยนำฟีเจอร์แมพ (Feature Map) มาแบ่งออกเป็นพื้นที่ย่อย (Local Region) โดยกำหนดให้มีขนาดเป็น *PxP* ส่วน (จากรูปที่ 2.4 กำหนดให้มีขนาด 2x2 ส่วน) จากนั้นหาค่าที่มากที่สุดในแต่ละพื้นที่ย่อย จากพื้นที่แรกที่ประกอบด้วยตัวเลข 1, 2, 3, 5 สังเกตได้ว่าตัวเลขที่มากที่สุดคือ 5 ดังนั้นในการหาค่าที่มากที่สุดในแต่ละพื้นที่ย่อยนี้จึงเลือกค่า 5 มาพิจารณา

3. ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) เป็นชั้นที่เชื่อมโยงระหว่าง ฟีเจอร์แมพ (Feature Map) และผลลัพธ์ (Output) แบบสมบูรณ์ นั่นหมายถึงทุก ๆ นิวรอน (Neuron) ที่อยู่ในชั้นสุดท้ายของฟีเจอร์แมพ (Feature Map) จะถูกนำไปเปลี่ยนรูป (Reshape) หรือเรียกว่าแฟลทเท็น (Flatten) (ธนภัทร์, 2559)



รูปที่ 2.4 ตัวอย่างการทำงานชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การตรวจจับวัตถุ (Object Detection)

การตรวจจับวัตถุ (Object Detection) เป็นเทคโนโลยีในทางคอมพิวเตอร์ที่ใช้การเรียนรู้เชิงลึก (Deep Learning) และการประมวลผลภาพ (Image Processing) เป็นการตรวจจับ และระบุตำแหน่งของวัตถุที่สนใจภายในภาพด้วยคอมพิวเตอร์ ว่าเป็นวัตถุอะไรพร้อมทั้งระบุตำแหน่งว่าวัตถุนั้นอยู่ที่ตำแหน่งไหนภายในภาพ เนื่องจากภาพแต่ละภาพสามารถมีวัตถุได้หลายชนิด และมีขนาดที่แตกต่างกัน ซึ่งการตรวจจับวัตถุ (Object Detection) เป็นเทคโนโลยีที่พัฒนาขึ้นมาเพื่อให้คอมพิวเตอร์สามารถค้นหา และระบุวัตถุที่สนใจภายในภาพแทนการทำงานของมนุษย์ จุดเด่นที่สำคัญของการตรวจจับวัตถุ (Object Detection) คือการตรวจคำนวณหาขอบเขตของกล่องแบบสี่เหลี่ยม (Boundary Box) ซึ่งมีวิธีการที่เป็นที่นิยมได้แก่ R-CNN, Fast-RCNN, Faster-RCNN และ YOLO เป็นต้น (ปฏิญญา และคณะ, 2564)

รูปที่ 2.5 ตัวอย่างของกล่องแบบสี่เหลี่ยม (Boundary Box) ในการตรวจจับวัตถุ

2.4.1 R-CNN (Region-based Convolutional Neural Networks)

R-CNN เป็นหนึ่งในเครื่องมือแรกของการเรียนรู้เชิงลึก (Deep Learning) ที่ใช้ในการตรวจจับวัตถุ และเป็นตัวอย่างของแนวคิดที่ตรวจจับด้วย 2 ขั้นตอน ดังนี้

1. การสร้าง Region Proposal R-CNN ใช้วิธีการเรียกว่า Selective Search หรือเทียบเท่าเพื่อสร้างขอบเขตที่อาจมีวัตถุอยู่ในจุด ๆ นั้น โดยการแบ่งภาพออกเป็นส่วนย่อย ๆ (sub-regions) และทำการค้นหาพื้นที่ที่เป็นตัวเลือกในการสร้างขอบเขตวัตถุ ในกระบวนการนี้จะสร้างจำนวน Region Proposal ประมาณ 2,000 กรอบขอบเขต (Bounding Box)

2. จากนั้นจะถูกนำเข้าสู่โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) เพื่อนำไปใช้ในการจำแนกประเภทของวัตถุ ในขั้นตอนนี้ โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) จะเรียนรู้และแยกแยะคุณลักษณะของวัตถุจากภาพหรือวิดีโอ ซึ่งผลลัพธ์ที่ได้จะประกอบด้วยคุณลักษณะที่ดึงออกมาได้จากภาพ และฟีเจอร์ (Feature) ที่แยกออกมานั้นจะถูกนำไปเข้าสู่เครื่องมือการจำแนกประเภทอื่น ๆ เช่น Support Vector Machine (SVM) เพื่อ

เอกสารนี้ ผลลัพธ์ที่ได้ ไม่ว่ากรณี

ทำการจำแนกประเภทของวัตถุในพื้นที่นั้น ๆ อย่างไรก็ตาม ปัญหาที่เกิดขึ้นในวิธีการมาตรฐานของ R-CNN คือความช้าในการประมวลผล ทำให้ไม่เหมาะสมเป็นเครื่องมือในการตรวจจับวัตถุแบบครบวงจรที่สมบูรณ์ (พุฒิพงศ์, 2563)

2.4.2 Fast R-CNN กับ Faster R-CNN

Fast R-CNN เป็นการพัฒนาต่อมาจาก R-CNN ที่เป็นเครื่องมือในการตรวจจับวัตถุแบบครบวงจรที่สมบูรณ์ โดยกำจัดข้อจำกัดในการค้นหาของเวอร์ชันก่อนหน้า และใช้ Region Proposal Network (RPN) เพื่อแทนที่การค้นหาของ Selective Search ใน R-CNN Fast R-CNN ใช้ RPN เพื่อสกัดคุณลักษณะที่มีความน่าจะเป็นว่าเป็นวัตถุจาก Feature Map และทำการเปลี่ยนจากการนำคุณลักษณะที่แยกออกมาได้จากรูปเข้า SVM เป็นการต่อเติมโครงข่ายประสาทเทียมแทนที่ อย่างไรก็ตาม Fast R-CNN ยังมีปัญหาเรื่องความเร็วซึ่งสามารถรองรับได้เพียง 5 เฟรมต่อวินาที บนการ์ดจอประมวลผลภาพ (GPU) (พุฒิพงศ์, 2563)

ซึ่งต่อมาจึงมีการพัฒนา Faster R-CNN ขึ้นมาเพื่อแก้ไขปัญหาดังกล่าว โดยแนวคิดหลักคือการรวมโครงข่ายประสาทเทียมกับ Selective Search เพื่อเพิ่มประสิทธิภาพในการตรวจจับวัตถุ ใน Faster R-CNN จะมีการสร้าง Anchor Box ที่กำหนดจำนวนและขนาดไว้ล่วงหน้า เพื่อใช้ในการทำนายตำแหน่งของวัตถุ รวมทั้งความน่าจะเป็นที่จะมีกรอบขอบเขตวัตถุอยู่ในแต่ละ Anchor Box โดยกระบวนการทั้งหมดจะแยกแยะ และทำงานตามขั้นตอนที่แตกต่างกัน การพัฒนาในรูปแบบนี้ทำให้ Faster R-CNN มีความเร็วเพิ่มขึ้นถึง 10 เท่าเมื่อเปรียบเทียบกับ R-CNN ในการตรวจจับวัตถุ (พุฒิพงศ์, 2563)

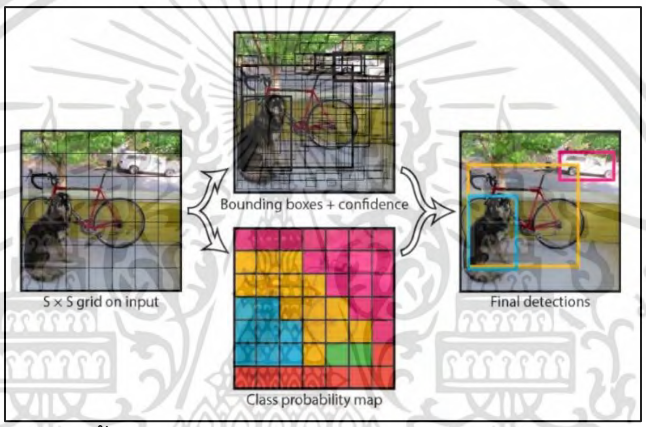
ในขณะที่ R-CNNs มีแนวโน้มที่จะมีความแม่นยำมากขึ้น แต่ปัญหาที่ใหญ่ที่สุดในตระกูล R-CNN คือความเร็วในการประมวลผล ซึ่งจึงมีการพัฒนางานที่เน้นความเร็วในการตรวจจับวัตถุและแก้ไขปัญหาที่เกิดขึ้น โดยมีการพัฒนาการตรวจจับวัตถุแบบขั้นตอนเดียว ในการตรวจจับวัตถุแบบขั้นตอนเดียว ไม่มีขั้นตอนการสกัด Region Proposal แยกไปจากการจำแนกวัตถุ ซึ่งทำให้เครื่องมือสามารถทำงานได้เร็วขึ้น โดยตรงจากภาพนำเข้าไปยังการจำแนกวัตถุ ตัวอย่างของเครื่องมือที่พัฒนาขึ้นจากแนวคิดนี้ได้แก่ อัลกอริทึมโยโล่ (You Only Look Once: YOLO) ซึ่งเป็นโมเดลที่มีความเร็วสูงและสามารถทำนายวัตถุแบบเรียลไทม์ได้ โดยไม่ต้องพึ่งพาขั้นตอนการสกัด Region Proposal เพิ่มเติม การพัฒนาการตรวจจับวัตถุแบบขั้นตอนเดียว ช่วยให้เครื่องมือตรวจจับวัตถุมีความเร็วสูงขึ้นและสามารถประมวลผลได้ในเวลาเรียลไทม์ ทำให้เหมาะสำหรับงานที่ต้องการความเร็วและประสิทธิภาพในการตรวจจับวัตถุได้อย่างมีประสิทธิภาพ (พุฒิพงศ์, 2563)

2.5 อัลกอริทึมโยโล่ (You Only Look Once: YOLO)

อัลกอริทึมโยโล่ (You Only Look Once: YOLO) เป็นอัลกอริทึมการตรวจจับวัตถุ (Object Detection) ที่มีลักษณะการทำงานที่เร็ว และแม่นยำ เนื่องจากสามารถคำนวณผลลัพธ์ทั้งหมดในภาพได้ในครั้งเดียว โดยไม่ต้องทำการสแกน (sliding window) หรือการตัดภาพ (image pyramid)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เหมือนกับการทำงานของอัลกอริทึมการตรวจจับวัตถุอื่น ๆ เช่น R-CNN และ Faster R-CNN ขั้นตอนการตรวจจับวัตถุในภาพของอัลกอริทึมโยโล่ (You Only Look Once: YOLO) คือการแบ่งภาพเป็นหลาย ๆ ช่อง (grid) จากนั้นใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network) เพื่อคำนวณผลลัพธ์ในแต่ละช่องว่ามีวัตถุอยู่หรือไม่ และให้แบบจำลองคำนวณค่าความน่าจะเป็น (probability) และพิกัดของวัตถุในแต่ละช่อง จากนั้นอัลกอริทึมโยโล่ (You Only Look Once: YOLO) จะใช้การลดการซ้อนทับของค่าสูงสุด (Non-Maximum Suppression) เพื่อกรองผลลัพธ์ที่ซ้ำซ้อนกัน และเลือกผลลัพธ์ที่ดีที่สุดของแต่ละวัตถุ ซึ่งทำให้อัลกอริทึมโยโล่ (You Only Look Once: YOLO) มีความแม่นยำสูง และสามารถทำงานได้อย่างรวดเร็ว (Wang, Bochkovskiy andMark Liao, 2022)



รูปที่ 2.6 ตัวอย่างขั้นตอนการตรวจจับวัตถุในภาพของอัลกอริทึมโยโล่ (YOLO)

ปัจจุบันอัลกอริทึมโยโล่ (You Only Look Once: YOLO) มีหลายเวอร์ชัน (Version) ให้เลือกใช้ และในแต่ละเวอร์ชัน (Version) มีโครงสร้างที่ไม่เหมือนกัน โดยในปัจจุบันที่มีการเผยแพร่ให้ใช้งานเวอร์ชันล่าสุดคือโยโล่เวอร์ชันที่ 8 (YOLOv8) โดยอัลกอริทึมโยโล่มีการวัฒนาการมาจนถึงปัจจุบันดังนี้

1. โยโล่ (YOLO) เป็นโมเดลการตรวจจับวัตถุ (Object Detection) และแบ่งแยกภาพออกเป็นส่วน ๆ (Image Segmentation) ที่ได้รับความนิยมมาก ถูกพัฒนาโดย โจเซฟ เรดมอน (Joseph Redmon) และ อาลี ฟาร์ฮาดี (Ali Farhadi) ที่มหาวิทยาลัยวอชิงตัน โดยเปิดตัวในปี 2015 โยโล่ได้รับความนิยมอย่างรวดเร็วเนื่องจากความเร็ว และความแม่นยำที่สูง

2. โยโล่เวอร์ชันที่ 2 (YOLOv2) เปิดตัวในปี 2016 ได้ปรับปรุงจากเวอร์ชันก่อนหน้าให้ดียิ่งขึ้น โดยการนำเทคนิคในการปรับปรุงการเรียนรู้ และการสร้างแบบจำลองของระบบประมวลผลเชิงลึก (deep learning) ในการฝึกสอนแบบจำลองประมวลผลเชิงลึก (Batch Normalization) เทคนิคการหาตำแหน่ง และขอบเขตของวัตถุในภาพ (Anchor Boxes) และเทคนิคการจัดกลุ่มคุณลักษณะของข้อมูล (Dimension Clusters) เข้ามาในโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. โยโล่เวอร์ชันที่ 3 (YOLOv3) เปิดตัวในปี 2018 และปรับปรุงประสิทธิภาพของโมเดลโดยใช้เครือข่ายหลักที่มีประสิทธิภาพสูงขึ้น โดยใช้เทคนิคการหาตำแหน่ง และขอบเขตของวัตถุในภาพ (Anchor Boxes) หลายตัวเพื่อระบุตำแหน่ง และขอบเขตของวัตถุที่ต้องการตรวจจับในภาพ (Multiple Anchors) และเทคนิคที่ใช้ในการปรับขนาดของฟีเจอร์แมพ (feature map) เพื่อให้สามารถรับรู้และสร้างคำสั่งจากฟีเจอร์แมพที่มีขนาดต่างกันได้อย่างมีประสิทธิภาพในการประมวลผลภาพ (Spatial Pyramid Pooling)

4. โยโล่เวอร์ชันที่ 4 (YOLOv4) ได้ถูกเปิดตัวในปี 2020 และนำเสนอนวัตกรรมใหม่เช่นเทคนิคในการเพิ่มข้อมูล (data augmentation) ในการฝึกสอนแบบจำลองการตรวจจับวัตถุ (object detection) โดยทำการรวมภาพจากหลาย ๆ ภาพมาเป็นภาพเดียว เพื่อสร้างชุดข้อมูลใหม่ที่มีความหลากหลายและความซับซ้อนมากขึ้น (Mosaic data augmentation) เทคนิคการตรวจจับวัตถุที่ไม่ใช้การหาตำแหน่ง และขอบเขตของวัตถุในภาพ (Anchor Boxes) เพื่อทำการคาดเดาตำแหน่งและความเป็นไปได้ของวัตถุในภาพโดยตรง และฟังก์ชันที่ใช้ในกระบวนการคำนวณค่าความสูญเสีย (loss) ในการฝึกสอนแบบจำลองเพื่อปรับปรุงความแม่นยำของแบบจำลองในการทำนายผลลัพธ์

5. โยโล่เวอร์ชันที่ 5 (YOLOv5) ได้ปรับปรุงประสิทธิภาพของแบบจำลอง และเพิ่มคุณสมบัติใหม่ เช่น การปรับแต่งพารามิเตอร์อัตโนมัติ การติดตามการทดลองแบบรวม และการส่งออก (Export) อัตโนมัติไปยังรูปแบบการส่งออกที่ได้รับความนิยม

6. โยโล่เวอร์ชันที่ 6 (YOLOv6) ได้เปิดตัวในรูปแบบของโอเพนซอร์ส (Open Source) โดยเหม่ยถวน (Meituan) ในปี 2022 และใช้งานในหลาย ๆ รุ่นของหุ่นยนต์ส่งอาหารอัตโนมัติ

7. โยโล่เวอร์ชันที่ 7 (YOLOv7) เพิ่มส่วนเสริมเพิ่มเติมเช่น การประเมินท่าทาง (pose estimation) ซึ่งเป็นกระบวนการหาหรือคาดเดาตำแหน่ง และทิศทางของส่วนต่าง ๆ ของร่างกายหรือวัตถุในภาพหรือวิดีโอ เช่น การติดตามท่าของมนุษย์ในระบบควบคุมการเคลื่อนไหว และในงานวิจัยทางการแพทย์เพื่อวิเคราะห์การเคลื่อนไหวของร่างกายในการวินิจฉัย และรักษาโรค วิธีการในการประเมินท่าสามารถใช้วิธีการตรวจหาจุดสำคัญบนร่างกาย (keypoint detection) และเชื่อมต่อจุดสำคัญเหล่านั้นเข้าด้วยกันเพื่อสร้างรูปร่างท่า (pose) ที่เป็นไปได้

8. โยโล่เวอร์ชันที่ 8 (YOLOv8) เป็นเวอร์ชันล่าสุดของโยโล่ ที่ถูกพัฒนาโดย อัลตราลิสติกส์ (Ultralytics) โดยเป็นโมเดลที่ทันสมัย และเป็นที่ยอมรับในระดับสูง โดย โยโล่เวอร์ชันที่ 8 (YOLOv8) ได้พัฒนาต่อยอดจากความสำเร็จของเวอร์ชันก่อนหน้า โดยนำเสนอคุณสมบัติ และการปรับปรุงใหม่เพื่อเพิ่มประสิทธิภาพ เสถียรภาพ และประสิทธิภาพการทำงาน โยโล่เวอร์ชันที่ 8 (YOLOv8) สนับสนุนงานทั้งหมดในการประมวลผลภาพ AI รวมถึงการตรวจจับ (detection) เซ็กเมนต์ (segmentation) การประเมินท่าทาง (pose estimation) การติดตาม (tracking) และการจำแนก (classification) ความหลากหลายนี้ช่วยให้ผู้ใช้สามารถใช้ประโยชน์จากความสามารถของ โยโล่เวอร์ชันที่ 8 (YOLOv8) ได้ในแอปพลิเคชันและสาขางานที่หลากหลาย (Glenn, et al., 2023)

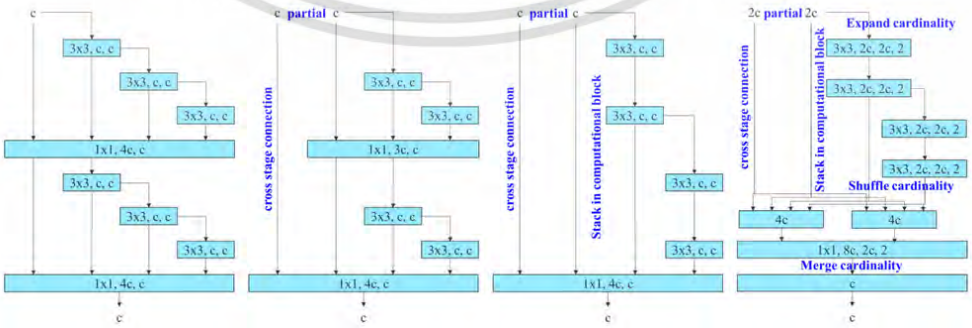
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยในงานวิจัยนี้ได้เลือกใช้อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) ในการตรวจจับหมายเลขสินค้า และโลโก้ของสินค้าจากรูปภาพ เหตุผลที่เลือกใช้อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) เนื่องจากอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) มีวิธีการใช้งานที่เผยแพร่ให้ศึกษาเรียนรู้มากกว่าอัลกอริทึมโยโล่เวอร์ชันที่ 8 (YOLOv8) ที่เป็นเวอร์ชันล่าสุด

2.5.1 อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7)

โยโล่เวอร์ชันที่ 7 (YOLOv7) เป็นอัลกอริทึมตรวจจับวัตถุที่พัฒนาโดย อเล็กเซย์ บอชโคฟสกี (Alexey Bochkovskiy) และเชียน เยา วัง (Chien-Yao Wang) โดยอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) เป็นเวอร์ชันที่พัฒนามาจากเวอร์ชันก่อนหน้า เช่น โยโล่เวอร์ชันที่ 5 (YOLOv5) และโยโล่เวอร์ชันที่ 6 (YOLOv6) ในโยโล่เวอร์ชันที่ 7 (YOLOv7) ใช้วิธีการตรวจจับวัตถุที่มีการปรับขนาดภาพให้เหมาะสมกับวัตถุที่ต้องการตรวจจับ (Multi-Scale Approach) เพื่อตรวจจับวัตถุในภาพ คือการแบ่งภาพนำเข้าเป็นกริด (Grid) ขนาด SxS และทำการทำนายพื้นที่ และความเป็นไปได้ของประเภทของวัตถุ (Class) สำหรับแต่ละเซลล์ (Cell) ของกริด (Grid) โดยการใช้วิธีการตรวจจับวัตถุที่มีการปรับขนาดภาพให้เหมาะสมกับวัตถุที่ต้องการตรวจจับ (Multi-Scale Approach) นี้ช่วยให้โยโล่เวอร์ชันที่ 7 (YOLOv7) สามารถตรวจจับวัตถุในภาพที่มีขนาด และมุมมองที่หลากหลายได้อย่างมีประสิทธิภาพ (Wang, Bochkovskiy andMark, 2022)

โครงสร้างหรือสถาปัตยกรรมของอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7 Architecture) ประกอบด้วยสามส่วนหลักคือ ซีเอสพีดาร์กเน็ต (CSPDarknet) เป็นส่วนหลักที่ใช้ในสกัดลักษณะที่มีความสำคัญจากข้อมูล (Feature Extraction) พีเอเน็ท (PANet) เป็นส่วนที่ใช้ในการรวมลักษณะที่ถูกสกัดมาจากข้อมูล เพื่อให้ได้ลักษณะที่รวม ๆ และรวมกันได้อย่างมีประสิทธิภาพ (Feature Fusion Module) และชั้นหรือเลเยอร์ของโยโล่เวอร์ชันที่ 7 (YOLOv7 Layer) เป็นส่วนที่ใช้ในการสร้างผลลัพธ์ และส่งผลลัพธ์จากการตรวจจับ ซึ่งประกอบไปด้วย ประเภทของวัตถุ (Class) ค่าความเชื่อมั่น (Confidence Score) ตำแหน่ง (Bounding Box Coordinates) และขนาดของวัตถุที่ตรวจจับ (Size) (Wang, Bochkovskiy andMark, 2022)



รูปที่ 2.7 ตัวอย่างสถาปัตยกรรมของอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7 Architecture)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 การรู้จำอักขระด้วยแสง (Optical Character Recognition – OCR)

การรู้จำอักขระด้วยแสง (Optical Character Recognition) คือกระบวนการแปลงข้อความจากรูปภาพหรือภาพสแกนเป็นตัวอักษรที่สามารถแก้ไขได้ในคอมพิวเตอร์ โดยไม่ต้องทำการพิมพ์ใหม่ ซึ่งสามารถช่วยลดระยะเวลาในการพิมพ์เอกสาร และยังสามารถเก็บข้อมูลในคอมพิวเตอร์ได้ทันที โดยกระบวนการทำการรู้จำอักขระด้วยแสง (Optical Character Recognition) มีขั้นตอนดังนี้

1. การตรวจจับตัวอักษร (Character Detection) ในขั้นตอนนี้ซอฟต์แวร์ (Software) การรู้จำอักขระด้วยแสง (Optical Character Recognition) จะทำการตรวจจับว่าในรูปภาพมีตัวอักษรอยู่หรือไม่

2. การตรวจจับกรอบของตัวอักษร (Character Bounding Box Detection) หลังจากขั้นตอนการตรวจจับอักษรแล้วซอฟต์แวร์ (Software) การรู้จำอักขระด้วยแสง (Optical Character Recognition) จะทำการวาดกรอบ (Bounding Box) รอบตัวอักษรเพื่อระบุตำแหน่ง และขอบเขตของแต่ละตัวอักษรในรูปภาพ ซึ่งจะใช้ในขั้นตอนต่อไปในการแยกแยะตัวอักษร

3. การแยกแยะตัวอักษร (Character Recognition) เมื่อทำการวาดกรอบของตัวอักษรแล้วซอฟต์แวร์ (Software) การรู้จำอักขระด้วยแสง (Optical Character Recognition) จะทำการแยกแยะตัวอักษรแต่ละตัวในกรอบ โดยใช้แบบจำลอง (Model) การเรียนรู้เชิงลึก (Deep Learning) เพื่อรู้จำ และระบุตัวอักษรที่อยู่ในกรอบ โดยส่วนใหญ่แล้วจะใช้แบบจำลองที่ถูกฝึกสอนด้วยข้อมูลตัวอักษรจำนวนมากเพื่อให้สามารถรู้จำ และจำแนกตัวอักษรได้อย่างแม่นยำ

เทคโนโลยีการรู้จำอักขระด้วยแสง (Optical Character Recognition) ได้รับการพัฒนาให้มีความสามารถในการแปลงข้อมูลจากเครื่องพิมพ์ได้แม่นยำยิ่งขึ้น อย่างไรก็ตาม การแปลงข้อมูลที่เกิดจากการเขียนด้วยลายมือยังมีข้อจำกัด เนื่องจากลายมือ และวิธีการเขียนของแต่ละบุคคลนั้นไม่เหมือนกัน จึงจำเป็นต้องมีการฝึกสอนตัวอย่างตัวอักษรแต่ละตัวให้กับซอฟต์แวร์ เพื่อให้สามารถรู้จำและสามารถอ่านข้อมูลลายมือที่คุ้นเคยแล้วแปลงเป็นตัวอักษรที่ถูกต้อง อย่างไรก็ตาม วิธีการฝึกสอนแบบนี้มีข้อจำกัดในเรื่องของกลุ่มลายมือ เนื่องจากถ้าเป็นลายมือที่แตกต่างจากต้นแบบแล้วการแปลงข้อมูลอาจคลาดเคลื่อนหรือผิดพลาดได้ (จักรพันธ์, 2561)

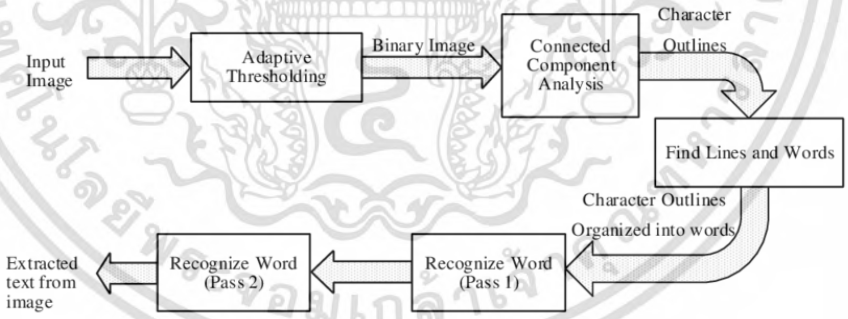
โดยในงานวิจัยนี้ได้เลือกใช้เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ในการรู้จำอักขระด้วยแสง (Optical Character Recognition) ของหมายเลขสินค้า เหตุผลที่เลือกใช้เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เนื่องจากเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) สามารถใช้งานได้ฟรี เนื่องจากเป็นโอเพนซอร์ส (Open Source) เมื่อเทียบกับ ไอสคริปต์ (aiScript) หรือกูเกิลวิชัน (Google Vision) ที่มีค่าใช้จ่ายในการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ถูกพัฒนาขึ้นครั้งแรกระหว่างปี ค.ศ.1985-1994 โดยในปี ค.ศ.2005 เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ได้เปิดเป็นโครงการโอเพนซอร์ส (Open Source) โดยบริษัท เอชพี (HP) ในปัจจุบันโครงการนี้ได้รับการพัฒนาต่อภายใต้การดูแลของบริษัท กูเกิล (Google) (Amitdo, 2016)

รูปแบบการทำงานของ เทสเซอร์แรคต์ (Tesseract) เริ่มต้นด้วยการรับภาพที่เป็นรูปภาพเฉดสีเทา (Gray Scale) เป็นข้อมูลรับเข้า (Input) เข้ามาแปลงภาพเป็นภาพไบนารี (Binary Image) ด้วยการใช้เทคนิคอะแดพทีฟเทรชโฮลดิ่ง (Adaptive Thresholding) ขั้นตอนถัดไปคือการวิเคราะห์พิกเซลที่เชื่อมต่อกันของตัวอักษร (Connected Component Analysis) ซึ่งใช้สำหรับแยกเส้นเค้าโครงของตัวอักษรออกมา จากนั้นจะทำการวิเคราะห์แถว และพื้นที่เพื่อหาแถวของตัวอักษร และหาตัวอักษร จากนั้นการรู้จำตัวอักษรจะเริ่มต้นด้วยขั้นตอน 2 รอบ (Two Pass) การทำงานในรอบที่ 1 เทสเซอร์แรคต์ (Tesseract) ทำการรู้จำตัวอักษรแต่ละตัวอักษร โดยส่งตัวอักษรที่ได้ไปยัง ตัวจัดลำดับอัลกอริทึมที่เรียนรู้ได้ (Adaptive Classifier) เพื่อรู้จำตัวอักษรที่มีความแม่นยำมากยิ่งขึ้น โดยเรียนรู้จากข้อมูลการฝึกสอน (Training Data) แต่จะมีตัวอักษรจำนวนหนึ่งที่ไม่สามารถทำการจดจำได้ในรอบที่ 1 เทสเซอร์แรคต์ (Tesseract) จะทำการส่งตัวอักษรเหล่านั้นไปประมวลผลในรอบที่ 2 เพื่อที่จะถูกแก้ไขปัญหาต่าง ๆ และสกัดตัวอักษร (Extracted Text) ออกจากภาพ (Chirag and Dharmendra, 2012)



รูปที่ 2.8 ขั้นตอนการทำงานของ เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

2.8 ภาษาโปรแกรมไพทอน (Python)

ภาษาโปรแกรมไพทอน (Python) คือภาษาโปรแกรมคอมพิวเตอร์ระดับสูงที่ใช้กันอย่างกว้างขวาง ถูกพัฒนาขึ้นโดย Guido van Rossum โปรแกรมเมอร์ชาวเยอรมัน ตั้งแต่ปี ค.ศ. 1990 โดยถูกออกแบบมาให้เป็นภาษาสคริปต์ที่อ่านง่าย โครงสร้างและไวยกรณ์ไม่ซับซ้อน ดังนั้นภาษาไพทอน (Python) จึงเขียนและอ่านง่ายกว่าภาษา C, C++ และ Java

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาษาไพทอน (Python) ถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษาไพทอน (Python) ได้ทั้งบน Windows, Linux และ macOS เป็นต้น อักทั้งยังเป็นโอเพนซอร์ส (Open Source) สามารถเข้าถึงได้ง่าย ทำให้ทุกคนสามารถนำภาษาไพทอน (Python) มาพัฒนาโปรแกรมได้โดยไม่ต้องเสียค่าใช้จ่าย และด้วยความเป็นโอเพนซอร์ส (Open Source) ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ภาษาไพทอน (Python) มีความสามารถสูงขึ้น และใช้งานได้กับทุกลักษณะงาน และที่สำคัญคือภาษาไพทอน (Python) มีไลบรารี่ (Library) ที่สนับสนุนการพัฒนางานทางด้านปัญญาประดิษฐ์ (เรือนขวัญ, 2563)

2.9 ไลบรารี่ (Library) ที่เกี่ยวข้องในงานวิจัย

2.9.1 Tensorflow

Tensorflow เป็นเครื่องมือ และไลบรารี่ที่ใช้ในการพัฒนา และการประมวลผลข้อมูลด้านเทคโนโลยีการเรียนรู้ของเครื่อง (Machine learning) ได้รับการพัฒนาโดยบริษัท กูเกิล (Google) ได้ทำการเปิดตัวเมื่อวันที่ 11 กุมภาพันธ์ 2017 ซึ่ง Tensorflow นั้นเป็นโอเพนซอร์ส (Open Source) ที่จะใช้ภาษาไพทอน (Python) ในการเขียนโปรแกรม รองรับปฏิบัติการ Linux, macOS, Windows และ Android (ณัฏฐณิชา, 2562)

2.9.2 Matplotlib

Matplotlib เป็นไลบรารีภาษาไพทอน (Python) ที่ใช้ในการสร้างกราฟและการพล็อตข้อมูลต่าง ๆ ในรูปแบบต่าง ๆ โดย Matplotlib นั้นสามารถสร้างกราฟเชิงสถิติ และกราฟเชิงเส้นได้ รวมถึงการพล็อตแผนภูมิและภาพถ่ายอื่น ๆ ที่เกี่ยวข้องกับข้อมูลที่สนใจ Matplotlib ให้ความยืดหยุ่นและความสามารถในการปรับแต่งกราฟให้เหมาะสม สามารถกำหนดสีและลักษณะต่าง ๆ ของเส้น ขนาดแกน ตัวแปรต่าง ๆ ที่ต้องการแสดงผล รวมถึงรูปแบบ และขนาดของข้อความที่แสดงบนกราฟได้อีกด้วย Matplotlib เป็นไลบรารีที่นิยมใช้ในการทำกราฟและการพล็อตข้อมูลในชุมชนของนักวิทยาศาสตร์ข้อมูล นักวิเคราะห์ข้อมูล นักฟิสิกส์ วิศวกร และนักพัฒนาซอฟต์แวร์ ซึ่งมีการสนับสนุนครอบคลุมสำหรับการสร้างกราฟและแผนภูมิที่หลากหลายในภาษาไพทอน (Python) อย่างมาก (The Matplotlib development team, 2023)

2.9.3 Numpy (Numerical Python)

NumPy (Numerical Python) เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการทำงานกับข้อมูลที่มีลักษณะเป็นอาร์เรย์ (Array) หรือเมทริกซ์ (Matrix) โดย NumPy มีเครื่องมือ และฟังก์ชันที่ช่วยให้ง่ายต่อการทำงานกับข้อมูลทางตัวเลข รวมถึงการประมวลผล และการคำนวณทางคณิตศาสตร์ NumPy มีโครงสร้างข้อมูลหลักที่เรียกว่า ndarray (N-dimensional array) ซึ่งเป็นอาร์เรย์หมายเลขที่มีขนาด และรูปร่างที่ถูกกำหนดล่วงหน้า โดย NumPy สามารถจัดการกับข้อมูลที่มีมิติต่าง ๆ ได้ตั้งแต่มิติเดียว (Vector) ไปจนถึงมิติสอง (Matrix) หรือมากกว่านั้น NumPy มีฟังก์ชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทางคณิตศาสตร์ที่มาพร้อมกับการทำงานกับอาร์เรย์ เช่น การบวก การลบ การคูณ และการหาร เป็นต้น (Jake, 2016)

2.9.4 OpenCv (Open Source Computer Vision Library)

OpenCV (Open Source Computer Vision Library) เป็นไลบรารีโอเพนซอร์ส (Open Source) ที่ใช้ในการประมวลผลภาพและวิดีโอ โดยมีเป้าหมายในการสนับสนุนและสร้างเครื่องมือที่ใช้ในการพัฒนาแอปพลิเคชันทางการมองเห็นคอมพิวเตอร์ (Computer Vision) และระบบที่เกี่ยวข้องกับการมองเห็นคอมพิวเตอร์ OpenCV เป็นไลบรารีที่มีความยืดหยุ่นและสามารถใช้งานได้ในหลายแพลตฟอร์ม รวมถึงรองรับการทำงานกับภาษาโปรแกรมหลายภาษา เช่น C++, Python, Java, และ MATLAB ซึ่งทำให้สามารถใช้ OpenCV ในโปรแกรมที่พัฒนาด้วยภาษาที่ต้องการได้อย่างสะดวกและยืดหยุ่น (สุทธิรักต์, 2557)

2.9.5 Pillow

Pillow เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการประมวลผลภาพ และการจัดการกับรูปภาพ โดยมีการพัฒนาขึ้นจากไลบรารีภาษา Python ชื่อว่า PIL (Python Imaging Library) ที่เป็นไลบรารีที่นิยมใช้ในการประมวลผลภาพในอดีต Pillow มีความสามารถในการอ่านและเขียนรูปภาพในรูปแบบต่าง ๆ เช่น JPEG, PNG, BMP, GIF, TIFF และอื่น ๆ (Jeffrey, 2023)

2.9.6 PyYAML

PyYAML เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการอ่านและเขียนข้อมูลในรูปแบบ YAML (YAML Ain't Markup Language) ซึ่งเป็นรูปแบบการซีรีเอชข้อมูล (Serialization) ที่ออกแบบมาเพื่อความอ่านง่ายและสะดวกในการใช้งานกับภาษาสคริปต์ รูปแบบ YAML สามารถแสดงข้อมูลในรูปแบบที่มนุษย์สามารถอ่านและเข้าใจได้ง่าย และยังให้ความยืดหยุ่นในการสร้างโครงสร้างข้อมูลที่ซับซ้อน (Kirill, 2021)

2.9.7 requests

requests เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการทำการส่งคำขอ HTTP ไปยังเซิร์ฟเวอร์เพื่อรับหรือส่งข้อมูล โดยส่วนใหญ่ใช้สำหรับการสื่อสารกับ API (Application Programming Interface) เพื่อเข้าถึงแหล่งข้อมูลหรือบริการต่าง ๆ ผ่านโปรโตคอล HTTP หรือ HTTPS (Kenneth, 2023)

2.9.8 SciPy

SciPy เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการประมวลผลทางวิทยาศาสตร์และคณิตศาสตร์ โดยมีเป้าหมายในการให้เครื่องมือและฟังก์ชันที่สำคัญสำหรับการทำงานทางคณิตศาสตร์ SciPy มีความสามารถในการทำงานกับข้อมูลและการประมวลผลทางวิทยาศาสตร์ที่หลากหลาย เช่น การแก้สมการเชิงเส้นและเชิงอนุพันธ์ การทำความสัมพันธ์ข้อมูลและการวิเคราะห์สถิติ การประมวลผลภาพและภาพเคลื่อนไหว การแก้ปัญหาทางวิทยาศาสตร์ด้วยการคำนวณเชิงตัวเลข การจำลองและการคำนวณทางกายภาพ และอื่น ๆ (Scipy developers, 2023)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9.9 Torch

Torch เป็นไลบรารีในภาษาไพทอน (Python) ที่มีเป้าหมายในการสนับสนุนการคำนวณทางคณิตศาสตร์และการเรียนรู้ของโมเดลปัญญาประดิษฐ์ (Artificial Intelligence) โดยเฉพาะในเชิงลึก (Deep Learning) ซึ่งมีพื้นฐานมาจากไลบรารี Lua Torch (ณัฏฐณิชา, 2562)

2.9.10 torchvision

torchvision เป็นไลบรารีชุดหนึ่งที่เป็นส่วนหนึ่งของไลบรารี Torch ที่ใช้ในการประมวลผลภาพและการทำงานกับชุดข้อมูลภาพในการเรียนรู้เชิงลึก (Deep Learning) torchvision มีฟังก์ชันที่มาพร้อมกับ Torch เพื่อการจัดการและประมวลผลข้อมูลภาพ เช่น การโหลดและจัดการกับชุดข้อมูลภาพต่าง ๆ (datasets) การทำการเตรียมและกำหนดการปรับปรุงภาพ (transforms) การสร้างและฝึกโมเดลปัญญาประดิษฐ์ที่ใช้กับภาพ (models) การวาดและแสดงผลภาพ (visualization) การทำคำนวณทางคณิตศาสตร์ที่เกี่ยวข้องกับภาพ (metrics) การจัดการกับภาพสีและกระบวนการทางวิทยาศาสตร์ทางการมองเห็น (computer vision) (ณัฏฐณิชา, 2562)

2.9.11 tqdm

tqdm เป็นไลบรารีในภาษาไพทอน (Python) ที่ใช้ในการสร้างแถบความคืบหน้า (progress bar) ที่แสดงความคืบหน้าของการทำงานหรือการวนซ้ำ เพื่อให้ผู้ใช้งานสามารถรับรู้ได้ว่ากระบวนการกำลังดำเนินอย่างไรและเป็นเครื่องมือสำหรับการตรวจสอบความคืบหน้าของการทำงาน (tqdm developers, 2015)

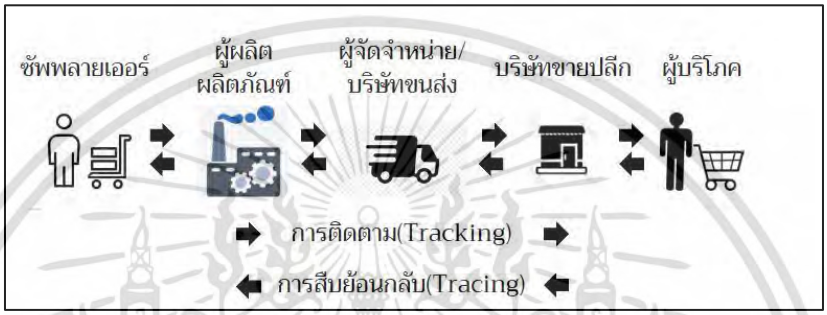
2.10 การตรวจสอบย้อนกลับ (Traceability)

การตรวจสอบย้อนกลับ (Traceability) คือการติดตาม และตรวจสอบการเคลื่อนไหวของสินค้าหรือวัตถุดิบตั้งแต่จุดเริ่มต้นที่แหล่งการผลิต ไปจนถึงการจัดจำหน่าย โดยมีเป้าหมายเพื่อให้สามารถตรวจสอบได้ว่าสินค้าหรือวัตถุดิบเหล่านั้นผลิตขึ้นมาจากแหล่งผลิตที่ได้รับการรับรองว่ามีคุณภาพและมีความปลอดภัย รวมถึงตรวจสอบว่ามีการใช้วัตถุดิบที่ถูกต้อง และตรงกับข้อกำหนดของผู้ซื้อ โดยมีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพการบริหารจัดการกระบวนการผลิต ลดความเสี่ยงของการเกิดภัยคุกคามต่อผู้บริโภค และเพิ่มความไว้วางใจในสินค้าจากผู้ซื้อ และผู้บริโภค กระบวนการตรวจสอบย้อนกลับประกอบด้วย 2 กระบวนการได้แก่

1. การติดตาม (Tracking) คือกระบวนการที่สามารถติดตาม และบันทึกข้อมูลเกี่ยวกับการเคลื่อนไหวหรือการดำเนินงานในระบบ การติดตามสามารถระบุได้ว่าสินค้านั้นไปอยู่ ณ ที่ใด เช่น ผู้ผลิต พบว่าวัตถุดิบในการผลิตสินค้าชุดหนึ่งมีปัญหา แต่สินค้าได้ถูกส่งออกไปจำหน่ายเรียบร้อยแล้ว ผู้ผลิตจึงต้องมีการเรียกคืนสินค้าที่ผลิตจากวัตถุดิบชุดที่มีปัญหาคืนทั้งหมด ผู้ผลิตต้องติดตามเส้นทางการผลิตและการจัดจำหน่าย เพื่อจะได้ทราบว่าสินค้าที่มีปัญหามีการจัดจำหน่ายอยู่ ณ ที่ใดบ้าง และสามารถเรียกคืนสินค้าได้อย่างถูกต้อง การดำเนินการดังกล่าวคือ การค้นหาปลายทางของสินค้า

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การสืบย้อนกลับ (Tracing) คือกระบวนการที่ใช้ในการตรวจสอบ และติดตามทางย้อนกลับของเหตุการณ์ของสินค้า ซึ่งมักเกี่ยวข้องกับการตรวจสอบความถูกต้องหรือการหาสาเหตุของปัญหาหรือข้อผิดพลาด การสืบย้อนกลับสามารถระบุได้ว่าสินค้าที่มีปัญหา มีการผลิตขึ้นเมื่อใด จากสายการผลิตไหน และรับวัตถุดิบมาจากแหล่งไหนเป็นต้น เพื่อค้นหาว่าจุดใดที่ก่อให้เกิดปัญหา และจุดที่ก่อให้เกิดปัญหาได้ผลิตสินค้าไปมากน้อยเพียงใด และมีข้อมูลรายละเอียดของการผลิตอย่างไร เพื่อทำการติดตามสินค้าคืนได้อย่างถูกต้อง การดำเนินการดังกล่าว คือการค้นหาต้นทางของสินค้าเพื่อใช้เป็นข้อมูลในการหาปลายทางของสินค้าต่อไป (มาดีนา และกนกวรรณ, 2556)



รูปที่ 2.9 กระบวนการตรวจสอบย้อนกลับ

ระบบการตรวจสอบย้อนกลับเป็นมาตรการที่ประเทศคู่ค้าสำคัญ ไม่ว่าจะเป็นสหรัฐอเมริกา สหภาพยุโรป และญี่ปุ่น ได้กำหนดเป็นกฎระเบียบในการนำเข้าสินค้าให้ประเทศผู้ส่งออกต้องนำไปปฏิบัติตาม ทั้งนี้ประเทศไทยในฐานะที่เป็นผู้ส่งออกสินค้า จึงจำเป็นอย่างยิ่งที่จะต้องนำระบบการตรวจสอบย้อนกลับมาประยุกต์ใช้ในกระบวนการผลิตสินค้า เพื่อรองรับสถานการณ์ทางการค้าที่เกิดขึ้นในปัจจุบัน และอนาคต นอกจากนั้นการนำระบบการตรวจสอบย้อนกลับมาประยุกต์ใช้ ยังก่อให้เกิดประโยชน์ทั้งกับกับผู้ผลิตสินค้า และผู้บริโภค ดังนี้

1. สามารถลดปริมาณการเรียกคืนสินค้าที่มีปัญหา โดยสามารถเรียกคืนเฉพาะสินค้าล็อตที่มีปัญหา

2. ช่วยในการป้องกันชื่อเสียงของบริษัทไม่ให้เสียหายเช่น ในกรณีที่สามารถพิสูจน์ได้ว่าจุดที่ก่อให้เกิดปัญหาไม่ได้เป็นเพราะขั้นตอนการผลิต แต่เป็นเพราะการจัดเก็บของผู้จัดจำหน่ายสินค้า

3. สามารถสืบย้อนกลับ สืบค้นแหล่งที่มาของสินค้าได้อย่างรวดเร็ว แม่นยำ และช่วยลดต้นทุนการเรียกคืนสินค้า

4. เป็นการรับประกันคุณภาพ และสามารถสืบค้นแหล่งที่มาของสินค้าได้ทั้งระบบของวงจรการผลิต

5. สร้างความมั่นใจให้กับผู้บริโภคได้ว่าจะได้รับสินค้าที่มีความปลอดภัย

6. เพื่อเป็นการปฏิบัติให้ตรงตามกฎระเบียบการค้าของประเทศคู่ค้าที่สำคัญเช่น สหภาพยุโรป ญี่ปุ่น และสหรัฐอเมริกา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. สามารถลดปริมาณการใช้กระดาษสำหรับการบันทึกข้อมูลได้

ในกระบวนการตรวจสอบย้อนกลับสินค้าอาหาร มีผู้ที่เกี่ยวข้องต่อการเก็บรวบรวมข้อมูลในขั้นตอนต่าง ๆ เพื่อให้กระบวนการมีความสมบูรณ์ ตั้งแต่การผลิตวัตถุดิบ กระบวนการแปรรูป จนถึงการกระจายผลิตภัณฑ์อาหารสู่ผู้บริโภค ซึ่งผู้ที่มีส่วนเกี่ยวข้องหลักกับระบบการตรวจสอบย้อนกลับแบ่งออกได้ดังนี้

1. ผู้ผลิตวัตถุดิบ คือบุคคลหรือองค์กรที่จัดหา และจัดส่งวัตถุดิบที่ใช้ในกระบวนการผลิตสินค้า เป็นผู้รับผิดชอบในการเก็บข้อมูลเกี่ยวกับแหล่งที่มาของวัตถุดิบ เช่น การสั่งซื้อ การนำเข้า การจัดเก็บ และรายละเอียดทางเทคนิคอื่น ๆ เกี่ยวกับวัตถุดิบ

2. ผู้ผลิตสินค้า เป็นผู้รับผิดชอบในการติดตามข้อมูล และบันทึกข้อมูลเกี่ยวกับกระบวนการผลิตสินค้า ซึ่งรวมถึงข้อมูลเกี่ยวกับการผลิตสินค้าตามมาตรฐาน คุณภาพ วันที่ผลิต และรายละเอียดของการบรรจุสินค้า

3. ผู้กระจายสินค้า เป็นผู้ที่มีบทบาทในการจัดส่งสินค้าจากผู้ผลิตไปยังผู้บริโภค ผู้กระจายสินค้ามีการติดตาม และบันทึกข้อมูลเพิ่มเติมเกี่ยวกับการจัดส่ง การจัดเก็บสินค้า หรือข้อมูลเกี่ยวกับการตลาด

4. ผู้บริโภค เป็นผู้ที่มีบทบาทในการใช้สินค้าหรือบริการ ผู้บริโภคจะสามารถทราบถึงข้อมูลที่มาของสินค้าที่บริโภคได้ โดยหากเกิดอันตรายอันเนื่องมาจากการใช้งานหรือการบริโภคสินค้าอาหารนั้น ผู้บริโภคสามารถสืบย้อนกลับถึงแหล่งที่มาของสินค้าหรือตรวจสอบจากแหน่งกำเนิดไปจนถึงผลิตภัณฑ์ขั้นตอนสุดท้ายได้ (มนัส และมาเรียม, 2558)

2.11 ค่าประเมินประสิทธิภาพแบบจำลอง

การประเมินประสิทธิภาพการทำนายของโปรแกรม เปรียบเทียบกับผลลัพธ์จริง ๆ ที่ถูกต้องโดยมีความหมายแต่ละตัวดังนี้ (พิชยุทธ และยุทธนา, 2564)

True Positive – TP คือ สิ่งที่โปรแกรมทำนายว่าจริง และเป็นวัตถุซึ่งกำลังสนใจอยู่

True Negative – TN คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง และเป็นวัตถุซึ่งไม่ได้สนใจอยู่

False Positive – FP คือ สิ่งที่โปรแกรมทำนายว่าจริง แต่เป็นวัตถุซึ่งไม่ได้สนใจอยู่

False Negative – FN คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง แต่เป็นวัตถุซึ่งกำลังสนใจอยู่

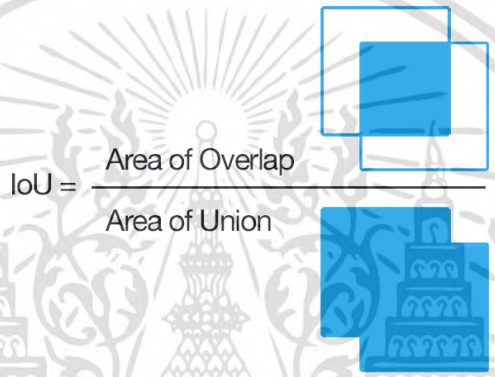
2.11.1 ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU)

ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) ใช้ในการประเมินประสิทธิภาพของอัลกอริทึมการตรวจจับวัตถุ (Object Detection) ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) แสดงถึงการซ้อนทับระหว่างกรอบบริเวณ (Bounding Box) ของผลเฉลย (Ground Truth) และค่าที่โปรแกรมทำนายได้ (Predicted) แบบจำลองการตรวจจับวัตถุมีประสิทธิภาพมากขึ้นเมื่อค่าอัตราส่วนระหว่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) เพิ่มขึ้น เนื่องจากค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) ที่มากขึ้นแสดงถึงความแตกต่างที่น้อยลงของกรอบบริเวณ (Bounding Box) ที่ทำนายได้กับกรอบบริเวณจริง (Luxolo, Dane andAlden, 2023) โดยนิยมใช้ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union – IoU) ที่ 0.5 หรือวัตถุที่ตรวจจับตรงกับตำแหน่งที่กำหนดของผลเฉลย (Ground Truth) เท่ากับหรือมากกว่า 50 เปอร์เซนต์ (ธนดล, 2564)

$$IntersectionoverUnion(IoU) = \frac{AreaofOverlap}{AreaofUnion} \quad (1)$$



รูปที่ 2.10 วิธีการคำนวณค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (IoU)

2.11.2 ค่าความถูกต้อง ค่าความแม่นยำ และค่าความระลึก (Accuracy, Precision and Recall)

ค่าความถูกต้อง (Accuracy) ใช้ในการประเมินความถูกต้องของแบบจำลองในภาพรวม ค่าความถูกต้อง (Accuracy) คืออัตราส่วนของวัตถุที่แบบจำลองตรวจจับได้ถูกต้องเทียบกับจำนวนวัตถุทั้งหมดในชุดข้อมูลทดสอบ โดยในงานวิจัยนี้เลือกใช้ค่าความถูกต้อง (Accuracy) ในการประเมินผลของแบบจำลอง เนื่องจากต้องการจะเปรียบเทียบค่าความถูกต้อง (Accuracy) กับเครื่องมือการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน (ศรัญญา, 2563)

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

ค่าความแม่นยำ (Precision) เป็นค่าที่ใช้ในการประเมินความถูกต้องของแบบจำลองในการตรวจจับวัตถุ (Object Detection) ในสถานการณ์ที่แบบจำลองทำนายว่าเป็นวัตถุที่สนใจ และจะสนใจเฉพาะเหตุการณ์ที่แบบจำลองทำนายว่าเป็นวัตถุที่สนใจได้ถูกต้อง (ศรัญญา, 2563)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Precision = \frac{TP}{(TP+FP)} = \frac{TP}{All\ Predict} \quad (3)$$

ค่าความระลึก (Recall) เป็นการประเมินค่าความถูกต้องของแบบจำลอง ในงานตรวจจับวัตถุ (Object Detection) ที่ต้องการตรวจจับวัตถุที่สนใจที่มีอยู่จริงในภาพ (ศรัญญา, 2563)

$$Recall = \frac{TP}{(TP+FN)} = \frac{TP}{All\ Ground\ Truth} \quad (4)$$

2.11.3 ค่าวัดประสิทธิภาพ (F1-Score)

ค่าวัดประสิทธิภาพ (F1-Score) เป็นค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) (ศรัญญา, 2563)

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

2.11.4 ค่าความแม่นยำเฉลี่ย (Average Precision)

ค่าความแม่นยำเฉลี่ย (Average Precision) เป็นการหาค่าเฉลี่ยของค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของแบบจำลองในการตรวจจับวัตถุ ใช้ประเมินประสิทธิภาพความแม่นยำของแบบจำลอง (Model) ในการตรวจจับวัตถุโดยคำนวณจากพื้นที่ใต้กราฟของค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) และเรียงความมั่นใจที่การทำนายสูงสุดตามลำดับ เมื่อกำหนดให้ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) เท่ากับ 0.5 จากนั้นปรับค่าความแม่นยำ (Precision) ด้วยฟังก์ชันความเที่ยงที่ได้รับการประมาณค่าระหว่างข้อมูล (Interpolated Precision) เพื่อหาพื้นที่ใต้กราฟ แล้วจะทำการแบ่งกราฟออกเป็นส่วน ๆ เพื่อคำนวณค่าความแม่นยำเฉลี่ย (Average Precision) ในขั้นตอนสุดท้ายดังนี้ (พิชยุทธ และยุทธนา, 2564)

$$P_{interp}(r) = \max_{\tilde{r}:\tilde{r} \geq r} p(\tilde{r}) \quad (6)$$

โดยที่ p คือ ค่าความแม่นยำ (Precision)

$\tilde{r}$ คือ ค่าความระลึก (Recall)

การแบ่งกราฟออกเป็นส่วน ๆ เพื่อหาพื้นที่ใต้กราฟ ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$AP = \sum(r_{n+1} - r_n)p_{interp}(r_{n+1}) \quad (7)$$

2.11.5 ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER)

ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) ใช้ในการประเมินความแม่นยำของการรู้จำอักขระด้วยแสง (Optical Character Recognition) โดยการประเมินความผิดพลาดระหว่างข้อความที่แบบจำลองสร้างขึ้น และข้อความที่ถูกต้อง ผลลัพธ์ที่ต่ำแสดงให้เห็นถึงระดับความแม่นยำที่สูง หากค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) เป็น 0 เปอร์เซ็นต์ แสดงว่าแบบจำลองได้ผลลัพธ์ที่ถูกต้องทุกตัวอักษร ส่วนถ้าเป็น 100 เปอร์เซ็นต์ แสดงว่าแบบจำลองไม่สามารถแปลงหรือระบุตัวอักษรใด ๆ ได้ถูกต้องเลย

$$CER = \frac{Insertions+Substitutions+Deletions}{Length(Ground\ truth\ label)} \quad (8)$$

จากสูตรการหาค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) ด้านบนโดยที่ค่า Insertions คือจำนวนตัวอักษรที่เพิ่มขึ้น Substitutions คือจำนวนตัวอักษรที่เปลี่ยนไป Deletions คือจำนวนตัวอักษรที่หายไป และ Length คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล (ธนดล, 2564)

2.11.6 ระยะเวลาในการประมวลผล (Runtime)

ระยะเวลาในการประมวลผล (Runtime) ใช้เพื่อเปรียบเทียบระยะเวลาในการประมวลผลกับเครื่องมือที่ทางบริษัทกรณีศึกษาใช้อยู่ในปัจจุบัน เพื่อเป็นตัวเปรียบเทียบว่าเทคโนโลยีที่ใช้ในงานวิจัยนี้มีประสิทธิภาพในเรื่องของระยะเวลาในการประมวลผลที่เร็วกว่า เครื่องมือที่ใช้ในปัจจุบันหรือไม่ เพื่อปรับปรุง และพัฒนาโดยการพิจารณา นำเทคโนโลยีที่ทำการวิจัยมาพัฒนาต่อยอดเพื่อทดแทนเครื่องมือเดิมที่ใช้ในปัจจุบัน

2.12 งานวิจัยที่เกี่ยวข้อง

จักรภัทร และไตรปิฎก (2560) ได้ทำการศึกษาระบบที่สามารถนำรูปภาพป้ายโฆษณามาค้นหาเว็บไซต์ที่เกี่ยวข้องกับตัวโฆษณาแบบอัตโนมัติ และแสดงข้อมูลแก่ผู้ใช้งานระบบ โดยทำงานบนระบบปฏิบัติการแอนดรอยด์ (Android) โดยใช้ข้อมูลรูปภาพ และข้อความจากป้ายโฆษณามาประมวลผลและแสดงเว็บไซต์ที่เกี่ยวข้องให้กับผู้ใช้งานโดยกดเพียงปุ่มเดียวแบบเรียลไทม์ ผลจากการประเมินตัวตรวจจับวัตถุ ใช้รูปภาพจำนวน 1000 ภาพ (250 ภาพต่อ 1 ประเภทวัตถุ) ในสี่ประเภทวัตถุที่ต่างกัน ได้แก่ ภาพแมว ส้ม ทีวี และตุ๊กตาหมี โดยกำหนดให้คะแนน IoU ต่อตำแหน่งของวัตถุนั้นเมื่อมีค่า

เอกสารนี้ ... ไม่ว่ากรณีใดๆ ...

เท่ากับ 0.5 จะกล่าวเป็นตัวอย่างในเชิงบวก (Positive Sample) ในขณะที่ต่ำกว่าให้เป็นตัวอย่างในเชิงลบ (Negative Sample) ผลจากการทดสอบได้ค่าเฉลี่ยความแม่นยำโดยรวม (mAP) จากประเภทของวัตถุทั้ง 4 ประเภทเท่ากับ 0.73 และในส่วนของการประเมินการตรวจจับข้อความ โดยใช้เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) จากภาพข้อความ 5 ประเภทได้แก่ ภาพข้อความที่เขียนด้วยมือ ภาพข้อความที่สังเคราะห์ขึ้น ภาพข้อความที่มีความชัดเจนสูง ภาพข้อความที่มีความชัดเจนต่ำ ตัวอักษรขาดเสียหาย ภาพข้อความบนป้ายโฆษณา ได้ค่าความแม่นยำเฉลี่ยโดยรวม (mAP) อยู่ที่ 0.65 และค่าความเหมือนเฉลี่ยอยู่ที่ 0.62

ณรงค์เดช (2563) ได้ทำการศึกษาการพัฒนาแอปพลิเคชันแอนดรอยด์ (Android) สำหรับช่วยนักท่องเที่ยว และคนต่างชาติใช้แปลตัวอักษรภาษาไทยบนป้ายสาธารณะ โดยเลือกใช้เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ในการรู้จำตัวอักษรภาษาไทย ต่อจากนั้นใช้บริการแปลภาษาของไป่ตู้ (Baidu) ทำการแปลภาษาไทยเป็นภาษาอังกฤษ และจีน ผลการทดสอบแสดงให้เห็นผลลัพธ์เป็นที่น่าพึงพอใจ จากผลการทดลองป้ายที่อยู่บนถนนจำนวน 14 ป้าย (207 ตัวอักษร) ซึ่งมีความแม่นยำอยู่ที่ 0.985

พิชยุทธ และยุทธนา (2564) ได้ทำการศึกษาเทคโนโลยีการตรวจจับวัตถุตรวจจับความเสียหายของถนนเช่น หลุม ถนนแตก ถนนซ่อมปะ โดยใช้การเรียนรู้เชิงลึกด้วยวิธี Faster R-CNN ซึ่งสามารถใช้งานการตรวจจับได้ผ่าน Mobile Application และ Desktop Application โดยเรียนรู้และทดสอบจากภาพถนนจริง ๆ จากการทดสอบประเมินประสิทธิภาพการตรวจจับความเสียหายบนถนนรวมทั้งสิ้น 2,471 ภาพ ผลลัพธ์ความถูกต้องของแต่ละประเภทของถนนดังนี้ หลุม ถนนแตก ถนนซ่อมปะ ได้ผลลัพธ์ตามลำดับต่อไปนี้ 0.2370, 0.7266, 0.3764 และความถูกต้องเฉลี่ยอยู่ที่ 44 เปอร์เซ็นต์

Fang, Wang, and Ren (2020) ได้ทำการศึกษาวิธีการตรวจจับวัตถุแบบเรียลไทม์สำหรับสภาพแวดล้อมที่จำกัดด้วยอัลกอริทึม Tinier-YOLO ซึ่ง Deep Neural Networks (DNNs) มักจะ 22 ทำงานบนอุปกรณ์ที่มีประสิทธิภาพซึ่งมีความสามารถในการวิเคราะห์และมีหน่วยความจำที่สูง ซึ่งด้วยข้อจำกัดในส่วนของอุปกรณ์ทำให้ไม่สามารถใช้งานได้ และ Tiny-YOLO-V3 เป็นรุ่นที่มีขนาดเล็กสามารถทำงานบนอุปกรณ์ที่ฝังมาพร้อมกับคอมพิวเตอร์ทุกเครื่องได้ วัตถุประสงค์ของการวิจัยต้องการนำแบบจำลองดังกล่าวมาย่อขนาดให้เล็กลงมากยิ่งขึ้นโดยยังคงความแม่นยำในการตรวจจับที่ดี ยิ่งขึ้นและมีประสิทธิภาพเมื่อใช้กับข้อมูลแบบเรียลไทม์ ประสิทธิภาพการตรวจจับวัตถุได้รับการปรับปรุงใน Tinier-YOLO โดยใช้ Passthrough Layer ที่มีการผสานใน Map Feature จาก Layer ด้านหน้าเพื่อให้ได้คุณสมบัติที่ละเอียดซึ่งสามารถให้แบบจำลอง Tinier-YOLO มีขนาด 8.9 MB ซึ่งเล็ก กว่า Tiny-YOLO-V3 เกือบ 4 เท่า เมื่อนำไปทดสอบประสิทธิภาพกับข้อมูลเรียลไทม์ Jetson TX1 ผลลัพธ์มีค่าเท่ากับ 25 FPS และค่าความแม่นยำเฉลี่ยโดยรวม (mAP) เท่ากับ 65.7 % ในข้อมูล PASCAL และเท่ากับ 34.0 % บนชุดข้อมูล COCO

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Luxolo, Dane and Alden (2023) ได้ทำการศึกษาเทคโนโลยีการตรวจจับวัตถุเพื่อช่วยในการตรวจจับปลาฉลามบูลล์ที่โมซัมบิก โดยใช้ YOLOv7 และ ESRGAN เพื่อช่วยในการตรวจจับและระบุปลาฉลามหัวบาตร เพื่อช่วยนักวิทยาศาสตร์และผู้วิจัยเกี่ยวกับปลาฉลามหัวบาตร ในการติดตามและระบุปลาฉลามหัวบาตรผ่านภาพถ่ายและวิดีโอที่ได้รับจากทางนักวิทยาศาสตร์และผู้วิจัย เพื่อศึกษาให้สาธารณชนเข้าใจความสำคัญของปลาฉลามในทรัพยากรทะเล และลดโอกาสที่จะเกิดการสูญพันธุ์ของปลาฉลามหัวบาตร โดยผลลัพธ์ที่ได้จากการทดสอบกับข้อมูลทดสอบ แบบจำลอง YOLOv7 ได้คะแนน ค่าความแม่นยำ ค่าความระลึก และค่าความแม่นยำเฉลี่ยอยู่ที่ 100 เปอร์เซ็นต์ของทุกค่า

ธนดล (2564) ได้ทำการศึกษาการนำวิธีการเรียนรู้เชิงลึกมาใช้โดยใช้วิธีโครงข่ายประสาทแบบคอนโวลูชัน (CNN) ร่วมกับวิธีหน่วยความจำระยะสั้นระยะยาว (LSTM) ซึ่งเรียกว่า CNN-LSTM เพื่อที่จะนำมารู้จำคำบรรยายจากวีดิทัศน์ ได้สร้างตัวอย่างต้นแบบ CNN ที่มีจำนวน 16 ชั้น โดยชั้นสุดท้ายเป็น การย่อ ขนาดโดยใช้ค่าสูงสุด (Max-pooling) ก่อนที่จะส่งเข้า LSTM โดยในการเรียนรู้ได้ใช้รูปภาพคำ บรรยายที่มีรูปแบบ ขนาด และพื้นหลังที่หลากหลาย แล้วใช้การจำแนกการเชื่อมต่อชั่วคราว (CTC loss) ในการคำนวนหาค่า loss และถอดรหัสเป็นผลลัพธ์ สำหรับข้อมูลที่นำมาใช้ในการเรียนรู้นั้น ได้มาจากการรวมรวม 24 วีดิทัศน์จาก Youtube และ Facebook ที่มีคำบรรยายภาษาไทย อังกฤษ ตัวเลขไทยและตัวเลขอารบิก ซึ่งมีทั้งหมด 157 ตัวเพื่อนำมาถอดรหัสข้อมูลในชุดรูปภาพนั้นมีทั้งหมด 4,224 รูป ซึ่งได้ค่าเฉลี่ยความผิดพลาดที่น้อยที่สุดคือ 11.06%

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง ทำให้ผู้วิจัยสามารถทราบถึงกระบวนการที่เหมาะสมในการฝึกสอนระบบการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง จึงได้ทำการศึกษางานวิจัยที่ผู้วิจัยท่านอื่น ๆ ได้ทำการศึกษาเพื่อนำมาพิจารณาเพิ่มในส่วนของเทคนิคในการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง ให้มีประสิทธิภาพมากยิ่งขึ้น และจากงานวิจัยของ Luxolo, Dane and Alden (2023) ที่ได้นำอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) มาใช้ในการวิจัย และเป็นการตรวจจับปลาฉลาม ซึ่งโลโก้ของสินค้าทั้ง 2 รูปแบบในงานวิจัยนี้ก็มีรูปร่างเป็นสัตว์เช่นเดียวกัน ได้ผลลัพธ์ที่ค่าความแม่นยำเฉลี่ย (mAP) จากการทดสอบเท่ากับ 1.00 และจากงานวิจัยของ ธนดล (2564) ได้ใช้สถาปัตยกรรมที่คล้ายคลึงกับเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ข้อมูลที่นำมาใช้ในการเรียนรู้นั้น ได้มาจากการรวมรวม 24 วีดิทัศน์จาก Youtube และ Facebook ที่มีคำบรรยายภาษาไทย อังกฤษ ตัวเลขไทย และตัวเลขอารบิก ซึ่งมีทั้งหมด 157 ตัวเพื่อนำมาถอดรหัสข้อมูลในชุดรูปภาพนั้นมีทั้งหมด 4,224 รูป ซึ่งได้ค่าเฉลี่ยความผิดพลาดที่น้อยที่สุดคือ 11.06 เปอร์เซนต์ หรือคิดเป็นมีค่าความถูกต้อง เท่ากับ 88.94 เปอร์เซนต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

ในปัจจุบันบริษัทกรณีศึกษาได้มีการใช้เครื่องมือการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า (Object Detection) และเครื่องมือที่ใช้ในการแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ (Optical Character Recognition: OCR) โดยได้ทำการพัฒนาด้วยเทคโนโลยีเทนเซอร์โฟล (Tensorflow) และมีประสิทธิภาพของเครื่องมือที่ค่าความถูกต้อง (Accuracy) 90 เปอร์เซนต์ โดยใช้ในการตรวจสอบย้อนกลับ เพื่อตรวจสอบ และติดตามข้อมูลของสินค้า เพื่อระบุแหล่งที่มาของสินค้าที่พบปัญหามีการผลิตขึ้นเมื่อใด จากสายการผลิตไหน และรับวัตถุดิบมาจากแหล่งไหน การตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และการแปลงข้อความของหมายเลขสินค้า ช่วยในการเก็บรวบรวมข้อมูลของโลโก้สินค้า และหมายเลขของสินค้า ที่ต้องการทำการตรวจสอบย้อนกลับจากรูปภาพของสินค้าได้โดยอัตโนมัติ พบข้อจำกัดด้านระยะเวลาในการประมวลผล งานวิจัยนี้ใช้เครื่องมือโยโล่เวอร์ชันที่ 7 (YOLOv7) และเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เพื่อศึกษาการทำงาน และทดสอบประสิทธิภาพในเรื่องของค่าความถูกต้อง (Accuracy) และลดระยะเวลาที่ใช้ในการประมวลผล เพื่อนำเป็นต้นแบบสำหรับการพัฒนาต่อยอดให้มีประสิทธิภาพมากยิ่งขึ้น ในบทนี้จึงกล่าวถึงขั้นตอน และเครื่องมือที่ใช้ในการดำเนินงานวิจัย

3.1 เครื่องมือที่ใช้ในงานวิจัย

1. ภาษาโปรแกรมไพทอน (Python) คือภาษาโปรแกรมคอมพิวเตอร์ระดับสูงที่ใช้กันอย่างกว้างขวาง ถูกพัฒนาขึ้นโดย Guido van Rossum โปรแกรมเมอร์ชาวเยอรมัน ตั้งแต่ปี ค.ศ. 1990 โดยถูกออกแบบมาให้เป็นภาษาสคริปต์ที่อ่านง่าย โครงสร้างและไวยกรณ์ไม่ซับซ้อน ดังนั้นภาษาไพทอน (Python) จึงเขียนและอ่านง่ายกว่าภาษา C, C++ และ Java

ภาษาไพทอน (Python) ถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษาไพทอน (Python) ได้ทั้งบน Windows, Linux และ macOS เป็นต้น อักทั้งยังเป็นโอเพนซอร์ส (Open Source) สามารถเข้าถึงได้ง่าย ทำให้ทุกคนสามารถนำภาษาไพทอน (Python) มาพัฒนาโปรแกรมได้โดยไม่ต้องเสียค่าใช้จ่าย และด้วยความเป็นโอเพนซอร์ส (Open Source) ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ภาษาไพทอน (Python) มีความสามารถสูงขึ้น และใช้งานได้กับทุกลักษณะงาน และที่สำคัญคือภาษาไพทอน (Python) มีไลบรารี่ (Library) ที่สนับสนุนการพัฒนาด้านปัญญาประดิษฐ์ด้วย

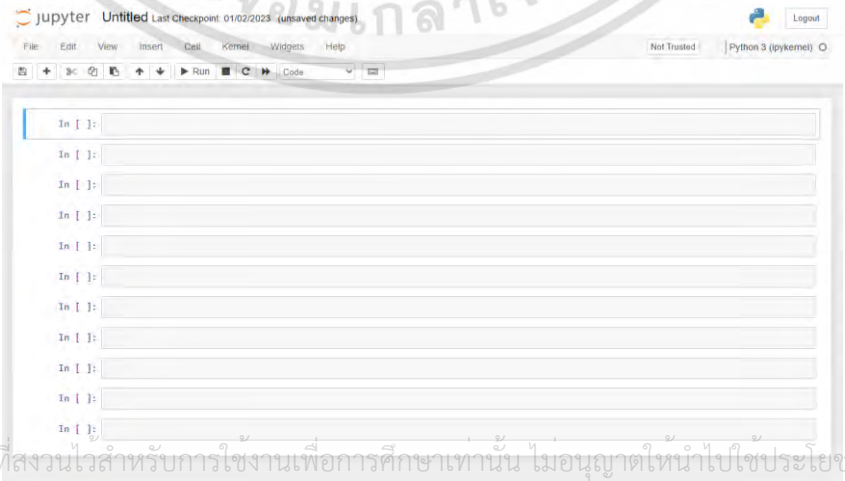
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.1 โลโก้ของภาษาไพทอน (Python)

2. โปรแกรม Anaconda คือโปรแกรมที่รวบรวมเครื่องมือในการคำนวณทางคณิตศาสตร์ และการพัฒนาปัญญาประดิษฐ์เอาไว้มากมาย แต่ละเครื่องมือมีลักษณะและการใช้งานที่แตกต่างกัน ในงานวิจัยนี้เลือกใช้เครื่องมือที่เหมาะสม นั้นคือ Jupyter notebook และ Visual Studio Code (VS Code)

รูปที่ 3.2 หน้าต่างโปรแกรม Anaconda

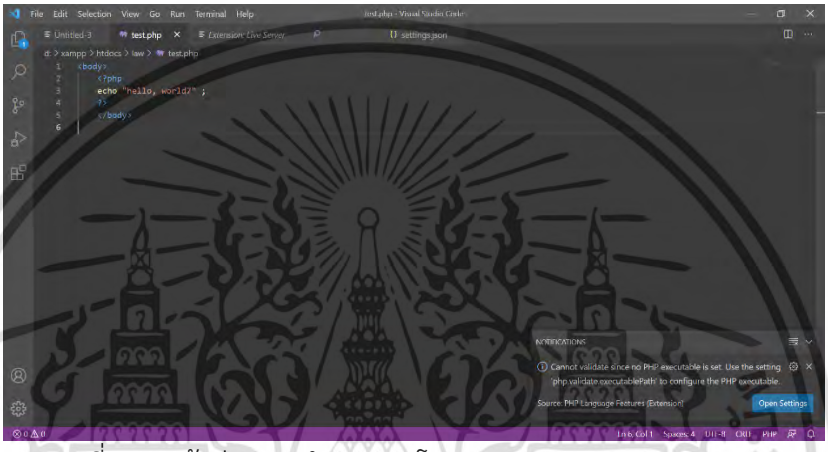
3. Jupyter notebook เป็นเครื่องมือที่นิยมใช้ในการพัฒนางานด้าน data science และงานด้านปัญญาประดิษฐ์ เป็นเครื่องมือที่ใช้งานง่าย สามารถเรียกใช้ไลบรารี่ (Library) พร้อมกับเขียนโค้ดและสามารถรันทีละบรรทัดเพื่อดูผลลัพธ์ได้ทันที



รูปที่ 3.3 หน้าต่างการทำงานของโปรแกรม Jupyter notebook

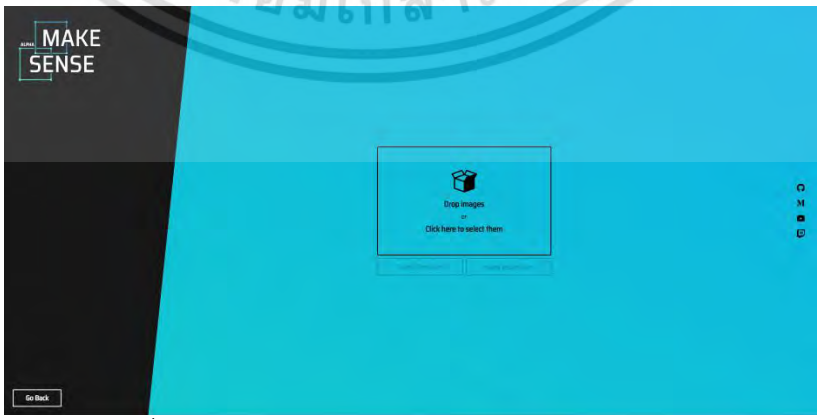
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

4. Visual Studio Code (VS Code) เป็นโปรแกรม Code Editor ที่ใช้ในการแก้ไข และปรับแต่งโค้ด จากบริษัทไมโครซอฟท์ มีการพัฒนาออกมาในรูปแบบของโอเพนซอร์ส (Open Source) จึงสามารถนำมาใช้งานได้แบบไม่เสียค่าใช้จ่าย รองรับการใช้งานทั้งบน Windows, Linux และ macOS สนับสนุนทั้งภาษา JavaScript, TypeScript และ Node.js สามารถเชื่อมต่อกับ Git ได้นำมาใช้งานได้ง่ายไม่ซับซ้อน มีเครื่องมือส่วนขยายต่าง ๆ ให้เลือกใช้อย่างมาก เช่น การเปิดใช้งานภาษาอื่น ๆ ทั้งภาษา C++, C#, Java, Python, PHP เป็นต้น



รูปที่ 3.4 หน้าต่างการทำงานของโปรแกรม Visual Studio Code

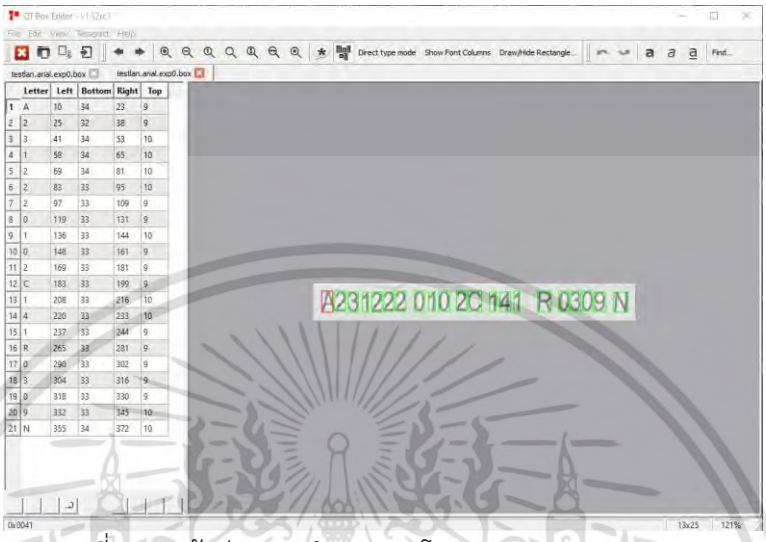
5. เว็บไซต์ Make Sense เป็นเครื่องมือออนไลน์ที่เปิดให้ใช้บริการแบบไม่มีค่าใช้จ่ายใช้สำหรับระบุตำแหน่ง และขอบเขตของวัตถุในรูปภาพ ด้วยการใช้เบราว์เซอร์ไม่ต้องการการติดตั้งที่ซับซ้อน เพียงเข้าเว็บไซต์ก็สามารถใช้งานได้ทันที เป็นเครื่องมือที่ช่วยในกระบวนการระบุชื่อของวัตถุที่สนใจ และการระบุตำแหน่งบนรูปภาพ เพื่อระบุหรือบ่งชี้คุณลักษณะหรือข้อมูลอื่น ๆ ที่สำคัญบนภาพ เพื่อการใช้งานในงานด้านการประมวลผลภาพ



รูปที่ 3.5 หน้าต่างการทำงานของเว็บไซต์ Make Sense

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. โปรแกรม Qt Box Editor เป็นเครื่องมือที่ใช้สำหรับการปรับเปลี่ยน และแก้ไขกล่องขอบเขต (Bounding Box) ของตัวอักษรบนภาพได้โดยตรง สามารถเพิ่มหรือลบกล่องตรวจจับ ปรับขนาด ย้าย หรือปรับเปลี่ยนตำแหน่งของกล่องตรวจจับ รวมถึงสามารถบันทึกและโหลดข้อมูลของกล่องตรวจจับ เพื่อนำไปใช้ในการฝึกสอนและทดสอบแบบจำลองการรู้จำอักขระด้วยแสงได้



รูปที่ 3.6 หน้าต่างการทำงานของโปรแกรม Qt Box Editor

7. อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) ใช้ในการฝึกสอนแบบจำลองเพื่อให้สามารถตรวจจับหมายเลขสินค้า และโลโก้ของสินค้าในภาพ อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) ถูกออกแบบให้มีประสิทธิภาพสูงทั้งในด้านการตรวจจับวัตถุ และความเร็วในการประมวลผล อีกทั้งอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) มีความยืดหยุ่นในการปรับแต่งและการปรับใช้งานตามความต้องการของผู้ใช้ สามารถตรวจจับวัตถุหลายวัตถุได้ในเวลาเดียวกัน

รูปที่ 3.7 โลโก้ของโยโล่เวอร์ชันที่ 7 (YOLOv7)

8. เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เป็นเครื่องมือที่ใช้ในการรู้จำและแปลงข้อความจากภาพหรือเอกสารที่สแกนหรือถ่ายมาให้อยู่ในรูปแบบข้อความ เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เป็นโอเพนซอร์ส (Open Source) ซึ่งทำให้สามารถเข้าถึงและใช้งานได้โดยไม่มีค่าใช้จ่าย และสามารถแก้ไข และปรับปรุงตามความต้องการของผู้ใช้ได้ตามต้องการ รองรับหลาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาษา สามารถรองรับการรู้จำและแปลงอักขระในหลายภาษาได้ รวมถึงภาษาที่ซับซ้อนและมีรูปแบบการเขียนที่ซับซ้อน เช่น ภาษาอังกฤษ ภาษาไทย ภาษาญี่ปุ่น ภาษาจีน เป็นต้น

Tesseract OCR

รูปที่ 3.8 โลโก้ของเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

9. ไลบรารี่ (Library) ที่ใช้
    1. Matplotlib ใช้สำหรับการสร้างกราฟและแผนภาพ
    2. Numpy ใช้สำหรับการทำงานกับอาเรย์หลายมิติและการคำนวณทางคณิตศาสตร์
    3. OpenCv ใช้ในการประมวลผลภาพและวิดีโอ โดยมีฟังก์ชันหลากหลายสำหรับการแปลงภาพ ตรวจจับวัตถุ สกัดคุณลักษณะ
    4. Pillow ใช้ในการจัดการภาพ รวมถึงการโหลดและบันทึกภาพในรูปแบบต่าง ๆ
    5. PyYAML ใช้ในการอ่านและเขียนไฟล์ในรูปแบบ YAML (YAML Ain't Markup Language) ที่ใช้ในการกำหนดค่าและการกำหนดแบบแผน
    6. requests ใช้ในการทำการส่งคำขอ HTTP เพื่อเชื่อมต่อกับเซิร์ฟเวอร์หรือรับส่งข้อมูลผ่านเครือข่าย ใช้ในการดาวน์โหลดหรือส่งไฟล์ที่เกี่ยวข้อง
    7. SciPy ใช้ในการทำงานทางวิทยาศาสตร์ รวมถึงการจัดการข้อมูลทางคณิตศาสตร์ ใช้ในกระบวนการประมวลผลและการวิเคราะห์ข้อมูล
    8. Torch เป็นไลบรารีที่เชื่อมโยงกับการเรียนรู้เชิงลึก (Deep Learning) โดยมีโมเดลปัญญาประดิษฐ์ที่สนับสนุนต่าง ๆ ใช้ในการสร้างและฝึกโมเดลปัญญาประดิษฐ์
    9. torchvision ช่วยในการประมวลผลภาพ รวมถึงการโหลดและจัดการกับชุดข้อมูลภาพ ใช้ในการประมวลผลภาพและการทำงานกับโมเดลปัญญาประดิษฐ์
    10. tqdm ใช้สร้างแถบความคืบหน้า (progress bar) ที่แสดงว่างานวิจัยหรือกระบวนการทำงานกำลังดำเนินอยู่อย่างไร ช่วยให้รับรู้และติดตามความคืบหน้าของการทำงานในเวลาที่กำหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัย แบ่งออกเป็น 2 ส่วนหลักได้แก่ ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุ และขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสง

3.2.1 ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุ

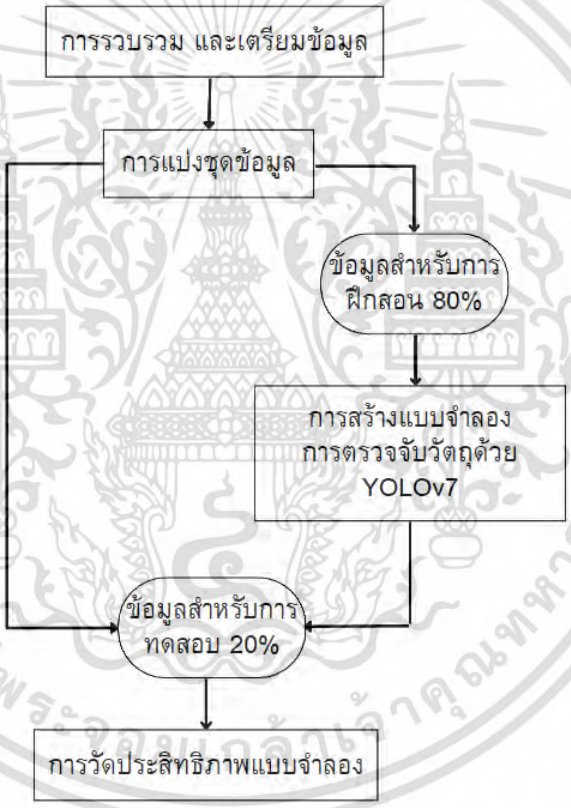
ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุแบ่งเป็น 4 ขั้นตอนได้แก่

1 การเก็บรวบรวมข้อมูล และการเตรียมข้อมูล

2 การแบ่งชุดข้อมูล

3 การสร้างแบบจำลองการตรวจจับวัตถุด้วยโยโล่เวอร์ชันที่ 7 (YOLOv7)

4 การทดสอบประสิทธิภาพของแบบจำลองการตรวจจับวัตถุ



รูปที่ 3.9 ขั้นตอนการดำเนินงานวิจัยของการตรวจจับวัตถุ

1. การเก็บรวบรวมข้อมูล และการเตรียมข้อมูล

1.1 การเก็บรวบรวมข้อมูลสำหรับการฝึกสอนแบบจำลอง และการทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection) ชุดข้อมูลเป็นรูปภาพของสินค้าที่ทางบริษัทจัดหามาให้ และได้ไปถ่ายรูปภาพสินค้าจากพื้นที่จัดจำหน่ายจริงจำนวน 476 ภาพ ซึ่งรูปภาพทั้งหมดเป็นภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x640 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG มีประเภทของวัตถุที่ต้องการตรวจจับ 3 ประเภท คือ หมายเลขของสินค้า โลโก้ของสินค้ารูปแบบที่ 1 และโลโก้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้ง

ของสินค้ารูปแบบที่ 2 ซึ่งมีประเภทของรูปภาพ 3 รูปแบบ โดยมี รูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 มีจำนวน 10 ภาพ รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1 มีจำนวน 442 ภาพ และรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2 มีจำนวน 24 ภาพ

1.2 การเตรียมข้อมูลสำหรับการฝึกสอนแบบจำลอง และการทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection) ทำการระบุตำแหน่งของวัตถุในรูปภาพเพื่อที่จะนำมาใช้สำหรับการฝึกสอน และการทดสอบแบบจำลองในการตรวจจับวัตถุ ในการระบุชื่อของวัตถุจะแบ่งออกเป็น 3 ประเภทวัตถุได้แก่ หมายเลขสินค้า โลโก้รูปแบบที่ 1 และโลโก้รูปแบบที่ 2 โดยเครื่องมือที่ใช้ในการระบุตำแหน่ง และขอบเขตของวัตถุในรูปภาพ คือเว็บไซต์ เมค เซ็นส์ (Make Sense) www.makesense.ai

2. การแบ่งชุดข้อมูล

จากชุดข้อมูลรูปภาพจำนวน 476 ภาพ รูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 มี 10 ภาพ รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1 มี 442 ภาพ และรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2 มี 24 ภาพ ทำการแยกชุดข้อมูลออกเป็นแต่ละประเภทของรูปภาพ แล้วทำการแบ่งชุดข้อมูลด้วยวิธี Split Test แบ่งข้อมูลออกเป็น 2 ส่วนคือ ชุดข้อมูลฝึกสอน (Train set) และชุดข้อมูลทดสอบ (Test set) ด้วยอัตราส่วน 80:20 ตามลำดับ จากแต่ละประเภทของรูปภาพ

3. การสร้างแบบจำลองการตรวจจับวัตถุด้วยโยโล่เวอร์ชันที่ 7 (YOLOv7)

ในการสร้างแบบจำลองการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าด้วยโยโล่เวอร์ชันที่ 7 (YOLOv7) ใช้ชุดข้อมูลสำหรับการฝึกสอนจำนวน 381 ภาพ

4. การทดสอบประสิทธิภาพของแบบจำลองการตรวจจับวัตถุ

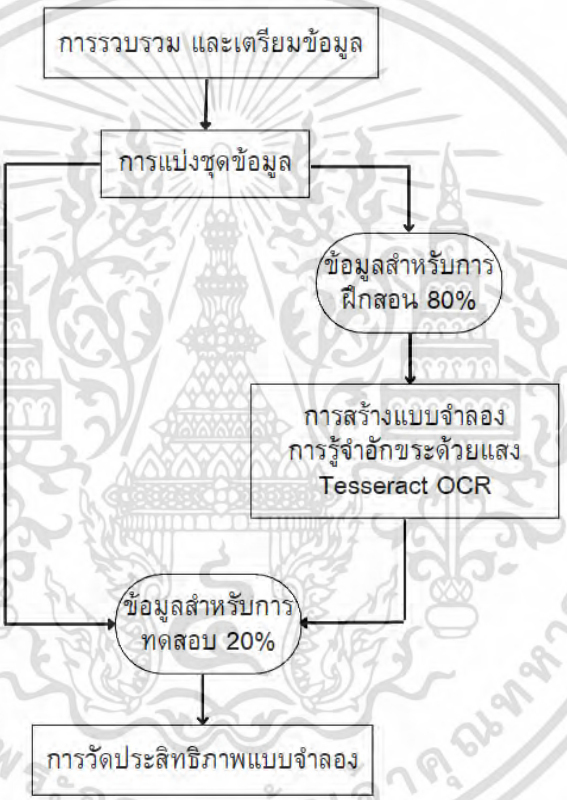
ทดสอบประสิทธิภาพของแบบจำลองการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า โดยใช้ชุดข้อมูลสำหรับการทดสอบจำนวน 95 ภาพ ทดสอบประสิทธิภาพของแบบจำลองโดยพิจารณาจากค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผล โดยกำหนดให้มีค่า IoU เท่ากับ 0.5 เนื่องจากต้องการจะเปรียบเทียบกับเครื่องมือการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน ที่มีค่าความถูกต้อง (Accuracy) อยู่ที่ 90 เปอร์เซนต์ และระยะเวลาในการประมวลที่มากกว่า 5 วินาทีต่อภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 ขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสง

ขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสงแบ่งเป็น 4 ขั้นตอน ได้แก่

1. การเก็บรวบรวมข้อมูล และการเตรียมข้อมูล
2. การแบ่งชุดข้อมูล
3. การสร้างแบบจำลองการรู้จำอักขระด้วยแสงด้วยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)
4. การทดสอบประสิทธิภาพของแบบจำลองการรู้จำอักขระด้วยแสง



รูปที่ 3.10 ขั้นตอนการดำเนินงานวิจัยของการรู้จำอักขระด้วยแสง

1. การเก็บรวบรวมข้อมูล และการเตรียมข้อมูล

1.1 การเก็บรวบรวมข้อมูลสำหรับการสร้างแบบจำลอง และการทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ชุดข้อมูลเป็นรูปภาพของสินค้าที่ทางบริษัทจัดหามาให้ และได้ไปถ่ายรูปภาพสินค้าจากพื้นที่จัดจำหน่ายจริงจำนวน 200 ภาพ ซึ่งรูปภาพทั้งหมดเป็นภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x40 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG โดยมีรูปแบบของตัวอักษร 2 รูปแบบ ซึ่งเป็นตัวอักษรภาษาอังกฤษและตัวเลขอารบิกที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มีจำนวน 20-21 หลัก โดยรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 มีจำนวน 100 ภาพ และรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 มีจำนวน 100 ภาพ

1.2 การเตรียมข้อมูลสำหรับการสร้างแบบจำลอง และการทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ทำการกำกับตัวอักษรของหมายเลขสินค้าในรูปภาพ เพื่อที่จะนำมาใช้สำหรับการฝึกสอน และการทดสอบแบบจำลองในการรู้จำตัวอักษร โดยขั้นตอนการกำกับตัวอักษรของแต่ละภาพตัวอักษร โดยข้อมูลที่ต้องทำการระบุคือ ข้อมูลที่บอกตำแหน่งของแต่ละตัวอักษรในภาพ และข้อมูลที่ระบุว่าแต่ตัวอักษรในภาพเป็นตัวอักษรอะไร โดยข้อมูลเหล่านี้จะถูกเก็บในไฟล์ .box ซึ่งการสร้างไฟล์ .box เพื่อทำการระบุข้อมูลตำแหน่งของแต่ละตัวอักษร และข้อมูลที่ระบุว่าแต่ตัวอักษรในภาพเป็นตัวอักษรอะไร ทำได้โดยคำสั่งโปรแกรมเทสเซอร์แรคต์ (Tesseract) ซึ่งผลลัพธ์ที่ได้ความถูกต้องของข้อมูลมีความไม่แม่นยำ จึงจำเป็นต้องเปลี่ยนหรือแก้ไขตัวอักษรที่ระบุผิดในไฟล์ .box โดยโปรแกรมที่เลือกใช้ในการปรับเปลี่ยนหรือแก้ไขตัวอักษรที่ระบุผิดคือโปรแกรม Qt Box Editor

2. การแบ่งชุดข้อมูล

จากชุดข้อมูลรูปภาพของหมายเลขสินค้าจำนวน 200 รูป มีรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 จำนวน 100 ภาพ และรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 จำนวน 100 ภาพ ทำการแบ่งชุดข้อมูลด้วยวิธี Split Test ออกเป็น 2 ส่วนคือ ชุดข้อมูลฝึกสอน (Train set) และชุดข้อมูลทดสอบ (Test set) ด้วยอัตราส่วน 80:20 ตามลำดับ

3. การสร้างแบบจำลองการรู้จำอักขระด้วยแสงด้วยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

ในการสร้างแบบจำลองการรู้จำอักขระด้วยแสง โดยใช้ชุดข้อมูลสำหรับการฝึกสอนจำนวน 160 ภาพ

4. การทดสอบประสิทธิภาพของแบบจำลองการรู้จำอักขระด้วยแสง

ทดสอบประสิทธิภาพของแบบจำลองการรู้จำหมายเลขของ โดยใช้ชุดข้อมูลสำหรับการทดสอบจำนวน 40 ภาพ ทดสอบประสิทธิภาพโดยพิจารณาจากค่าความถูกต้อง ในการแปลงตัวอักษรที่ได้จากค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผล เนื่องจากต้องการจะเปรียบเทียบกับเครื่องมือการรู้จำหมายเลขของสินค้า ของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน ที่มีความแม่นยำในการแปลงตัวอักษรอยู่ที่ 90 เปอร์เซนต์ และ ระยะเวลาในการประมวลผลที่มากกว่า 5 วินาทีต่อภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การเก็บรวบรวมข้อมูล

ชุดข้อมูลรูปภาพที่ได้นำมาใช้ในงานวิจัยเป็นรูปภาพของหมายเลขสินค้า และรูปภาพของสินค้าที่ทางบริษัทจัดหามาให้ และได้ไปถ่ายรูปภาพสินค้าจากพื้นที่จัดจำหน่ายจริง เป็นจำนวน 676 ภาพ โดยจะแบ่งเป็นชุดข้อมูลสำหรับการตรวจจับวัตถุ (Object Detection) 476 ภาพ และชุดข้อมูลสำหรับการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) 200 ภาพ

3.3.1 ชุดข้อมูลสำหรับการตรวจจับวัตถุ (Object Detection)

ชุดข้อมูลที่ใช้สำหรับการสร้างแบบจำลอง และการทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection) เป็นรูปภาพของสินค้าจำนวน 476 ภาพ ซึ่งมีประเภทของรูปภาพสินค้า 3 รูปแบบ ดังนี้

รูปแบบที่ 1 เป็นรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 จำนวน 10 ภาพ สินค้าเป็นผลิตภัณฑ์ประเภทกล่อง เป็นรูปภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x640 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG ดังรูปที่ 3.11

รูปที่ 3.11 ตัวอย่างรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1

รูปแบบที่ 2 เป็นรูปภาพที่มีหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1 มีจำนวน 442 ภาพ สินค้าเป็นผลิตภัณฑ์ประเภทกล่อง เป็นรูปภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x640 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG ดังรูปที่ 3.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.12 ตัวอย่างรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1

รูปแบบที่ 3 เป็นรูปภาพที่มีหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2 มีจำนวน 24 ภาพสินค้าเป็นผลิตภัณฑ์ประเภทกล่อง เป็นรูปภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x640 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG ดังรูปที่ 3.13

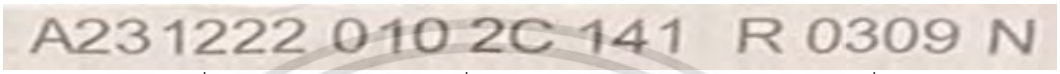
รูปที่ 3.13 รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 ชุดข้อมูลสำหรับการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR)

ชุดข้อมูลที่ใช้สำหรับการสร้างแบบจำลอง และการทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เป็นรูปภาพที่มีเฉพาะหมายเลขของสินค้าจำนวน 200 ภาพ ซึ่งมีประเภทของรูปภาพของหมายเลขสินค้า 2 รูปแบบ ดังนี้

รูปแบบที่ 1 เป็นรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 มีจำนวน 100 ภาพ เป็นรูปภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x40 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG ซึ่งเป็นตัวอักษรภาษาอังกฤษและตัวเลขอารบิกที่มีจำนวน 20-21 หลัก ดังรูปที่ 3.14



รูปที่ 3.14 ตัวอย่างรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1

รูปแบบที่ 2 เป็นรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 มีจำนวน 100 ภาพ ภาพ เป็นรูปภาพถ่ายมุมตรง ที่มีขนาดรูปภาพเท่ากับ 640x40 พิกเซล (Pixel) เป็นไฟล์นามสกุล JPG ซึ่งเป็นตัวอักษรภาษาอังกฤษและตัวเลขอารบิกที่มีจำนวน 20-21 หลัก ดังรูปที่ 3.15

รูปที่ 3.15 ตัวอย่างรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2

3.4 การเตรียมข้อมูล

3.4.1 การกำกับประเภทของวัตถุบนรูปภาพสำหรับการตรวจจับวัตถุ

สำหรับการฝึกสอนแบบจำลองการตรวจจับวัตถุ และการทดสอบแบบจำลองการตรวจจับวัตถุ จำเป็นที่จะต้องทำการระบุตำแหน่งของวัตถุในรูปภาพเพื่อที่จะนำมาใช้สำหรับการฝึกสอน และการทดสอบแบบจำลองในการตรวจจับวัตถุ รูปภาพที่ใช้สำหรับการฝึกสอน และทดสอบแบบจำลองการตรวจจับวัตถุจากข้อมูลที่เก็บรวบรวมมาได้รายละเอียด ดังตารางที่ 3.1

ตารางที่ 3.1 จำนวนของรูปภาพสำหรับการตรวจจับวัตถุ

ประเภทของรูปภาพ	จำนวนรูปภาพ
เฉพาะโลโก้ของสินค้ารูปแบบที่ 1	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทของรูปภาพ	จำนวนรูปภาพ
มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1	442
มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2	24
รวม	476

จากข้อมูลรูปภาพที่ได้ทำการเก็บรวบรวมมา ในการระบุชื่อของวัตถุจะแบ่งออกเป็น 3 ประเภทวัตถุได้แก่ หมายเลขสินค้า โลโก้สินค้ารูปแบบที่ 1 และโลโก้สินค้ารูปแบบที่ 2 รายละเอียดจำนวนของแต่ประเภทของวัตถุแสดงดังตารางที่ 3.2

ตารางที่ 3.2 จำนวนของวัตถุแต่ละประเภทสำหรับการตรวจจับวัตถุ

ประเภทของวัตถุ	จำนวนวัตถุ
หมายเลขสินค้า A231222 010 2C 141 R 0309 N	466
โลโก้สินค้ารูปแบบที่ 1	452

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทของวัตถุ	จำนวนวัตถุ
โลโก้สินค้ารูปแบบที่ 2	24
รวม	942

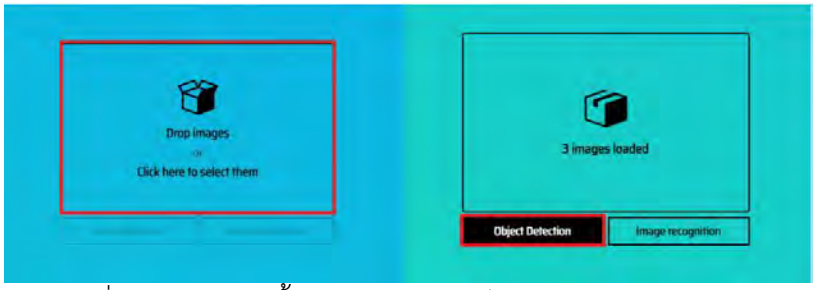
โดยได้ใช้เว็บไซต์ เมค เซ็นส์ (Make Sense) www.makesense.ai ซึ่งเป็นเว็บไซต์สำหรับใช้ในการระบุตำแหน่ง และขอบเขตของวัตถุในรูปภาพที่เปิดให้ใช้บริการแบบไม่มีค่าใช้จ่าย สามารถใช้งานได้ทันทีโดยไม่จำเป็นต้องดาวน์โหลด (Download) โดยจะทำการระบุชื่อของวัตถุที่สนใจ พิกัดแกน x และ y ของตำแหน่งของวัตถุที่ทำการกำกับไว้ (Bounding Box) โดยมีขั้นตอนการทำงานดังนี้

1. เข้าสู่หน้าเว็บไซต์ เมค เซ็นส์ (Make Sense) www.makesense.ai จากนั้นทำการกดที่ปุ่มคำสั่ง Get Started ที่มุมขวาล่างเพื่อเริ่มการใช้งาน ดังรูปที่ 3.16

รูปที่ 3.16 ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (1)

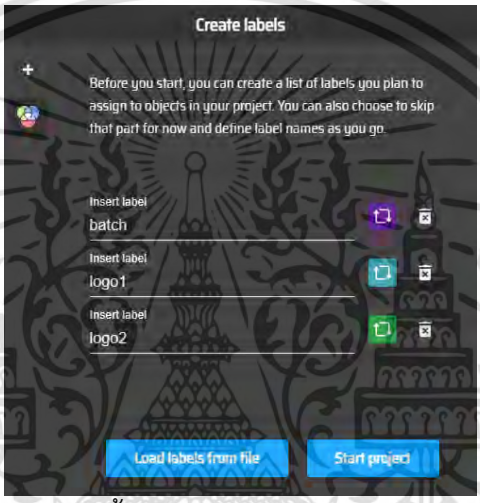
2. ทำการกดที่ปุ่มคำสั่ง Drop images or Click here to select them เพื่อทำการนำรูปภาพทั้งหมด 476 ภาพ ที่ต้องทำการกำกับประเภทของวัตถุภายในภาพ เมื่อทำการนำรูปภาพทั้งหมด 476 ภาพสำหรับการกำกับประเภทของวัตถุแล้ว ให้ทำการกดที่ปุ่มคำสั่ง Object Detection เพื่อเริ่มกระบวนการทำการกำกับประเภทของวัตถุ ดังรูปที่ 3.17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



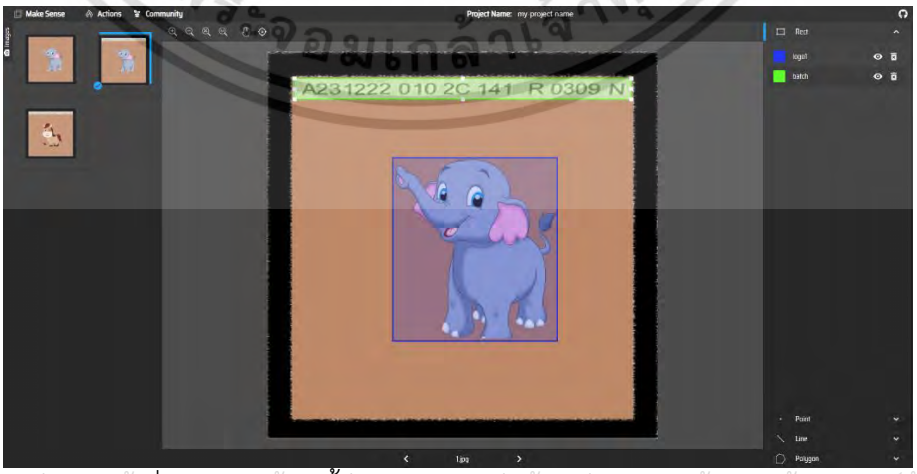
รูปที่ 3.17 ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (2)

3. ทำการระบุชื่อประเภทของวัตถุที่ต้องการจะกำกับบนรูปภาพ ดังรูปที่ 3.18



รูปที่ 3.18 ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (3)

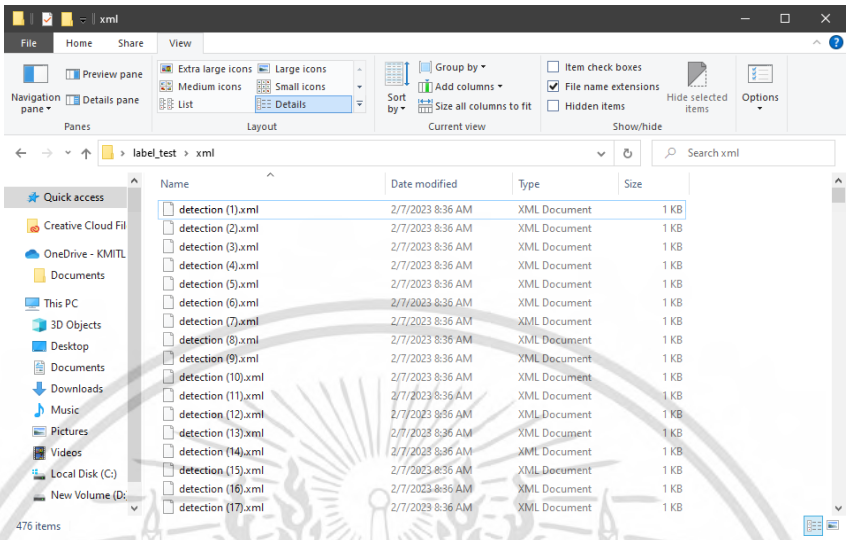
4. ทำการกำกับประเภทของวัตถุบนรูปภาพ โดยการระบุตำแหน่ง และขอบเขตของวัตถุนั้น ๆ บนรูปภาพในขั้นตอนนี้จำเป็นต้องลากกรอบขอบเขตด้วยมือทีละภาพจนครบ ดังรูปที่ 3.19



รูปที่ 3.19 ตัวอย่างขั้นตอนการใช้งานเว็บไซต์ Make Sense (4)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. หลังจากทำการกำกับประเภทของวัตถุบนรูปภาพจนเสร็จสิ้นทำการบันทึกข้อมูลที่ได้ทำการกำกับประเภทของวัตถุ ผลลัพธ์ที่ได้เป็นไฟล์ .xml ดังรูปที่ 3.20

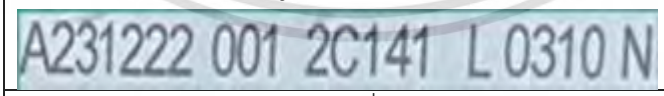


รูปที่ 3.20 ตัวอย่างไฟล์ .xml ที่ได้จากเว็บไซต์ Make Sense

3.4.2 การกำกับตัวอักษรของข้อความบนรูปภาพสำหรับการรู้จำอักขระด้วยแสง

ขั้นตอนการกำกับตัวอักษรสำหรับภาพของหมายเลขสินค้า รูปภาพที่ใช้สำหรับการฝึกสอนแบบจำลอง และทดสอบการรู้จำอักขระด้วยแสง ได้แก่รูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 และรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 จำนวนทั้งหมด 200 ภาพ รายละเอียดแสดงดังตารางที่ 3.3

ตารางที่ 3.3 จำนวนของรูปภาพสำหรับการรู้จำอักขระด้วยแสง

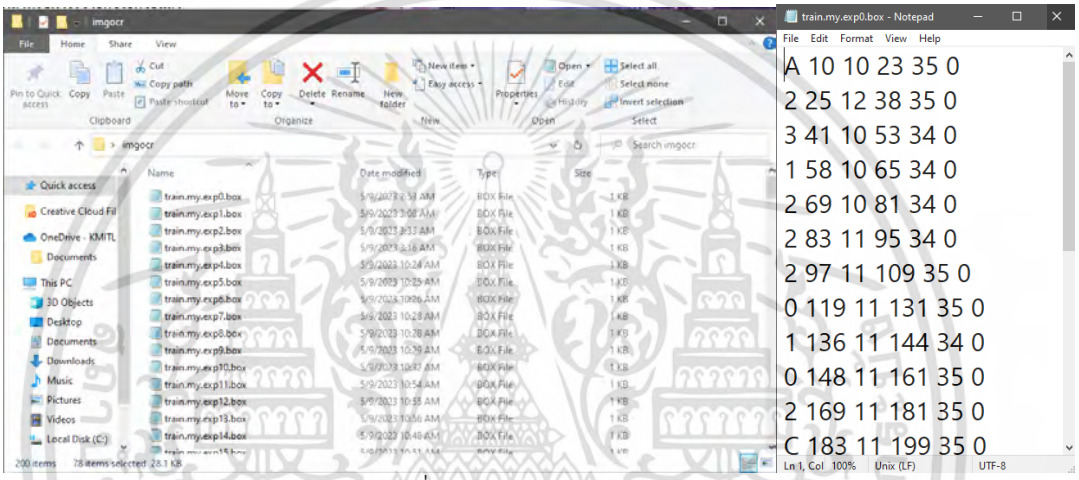
ประเภทของรูปแบบตัวอักษร	จำนวนรูปภาพ
รูปแบบที่ 1 	100
รูปแบบที่ 2	100
รวม	200

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยขั้นตอนการกำกับตัวอักษรของแต่ละภาพตัวอักษร โดยข้อมูลที่ต้องทำการระบุคือข้อมูลที่บอกตำแหน่งของแต่ละตัวอักษรในภาพ และข้อมูลที่ระบุว่าแต่ตัวอักษรในภาพเป็นตัวอักษรอะไร โดยข้อมูลเหล่านี้จะถูกเก็บในไฟล์ .box ซึ่งการสร้างไฟล์ .box เพื่อทำการระบุข้อมูลตำแหน่งของแต่ละตัวอักษร และข้อมูลที่ระบุว่าแต่ตัวอักษรในภาพเป็นตัวอักษรอะไร ทำได้โดยคำสั่งโปรแกรมเทสเซอร์แรคต์ (Tesseract) ดังนี้

คำสั่ง : tesseract [ชื่อภาษา].[ชื่อแบบอักษร].[expN].[นามสกุลไฟล์ของรูปภาพ] [ชื่อภาษา].[ชื่อแบบอักษร].[expN] batch.nochop makebox

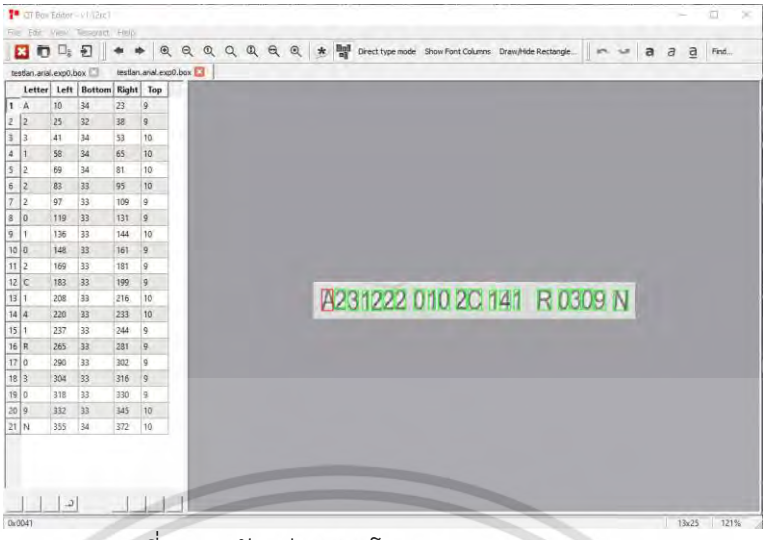
ตัวอย่าง : tesseract train.my.exp0.jpg train.my.exp0 batch.nochop makebox



รูปที่ 3.21 ตัวอย่างไฟล์ .box

ซึ่งผลลัพธ์ที่ได้จากการใช้คำสั่งโปรแกรมเทสเซอร์แรคต์ (Tesseract) ข้างต้นเพื่อสร้างไฟล์ .box ผลลัพธ์ที่ได้ความถูกต้องของข้อมูลมีความไม่แม่นยำ จึงจำเป็นต้องเปลี่ยนหรือแก้ไขตัวอักษรที่ระบุผิดในไฟล์ .box โดยโปรแกรมที่เลือกใช้ในการปรับเปลี่ยนหรือแก้ไขตัวอักษรที่ระบุผิดคือโปรแกรม Qt Box Editor เหตุผลที่เลือกใช้โปรแกรมนี้ เนื่องจากโปรแกรม Qt Box Editor เป็นเครื่องมือสำหรับปรับเปลี่ยน และแก้ไขไฟล์ .box สำหรับเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) จุดมุ่งหมายของโครงการ Qt Box Editor คือการจัดหาวิธีที่ง่าย และมีประสิทธิภาพในการปรับเปลี่ยนและแก้ไขไฟล์ .box โดยไม่คำนึงถึงขนาดไฟล์ และยังเปิดให้ใช้บริการโดยไม่มีค่าใช้จ่าย (Zdenop, 2023) ดังรูปที่ 3.22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.22 ตัวอย่างภาพโปรแกรม Qt Box Editor

3.5 การแบ่งชุดข้อมูล

3.5.1 การแบ่งชุดข้อมูลสำหรับการตรวจจับวัตถุ

จากชุดข้อมูลรูปภาพจำนวน 476 ภาพ รูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 มี 10 ภาพ รูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1 มี 442 ภาพ และรูปภาพที่มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2 มี 24 ภาพ และมีประเภทของวัตถุ 3 ประเภทได้แก่ หมายเลขสินค้า โลโก้สินค้ารูปแบบที่ 1 และโลโก้สินค้ารูปแบบที่ 2 เพื่อการสร้างแบบจำลองการตรวจจับวัตถุจะทำการแยกชุดข้อมูลออกเป็นแต่ละประเภทของรูปภาพ แล้วทำการแบ่งชุดข้อมูลด้วยวิธี Split Test แบ่งข้อมูลออกเป็น 2 ส่วนคือ ชุดข้อมูลฝึกสอน (Train set) และชุดข้อมูลทดสอบ (Test set) แล้วทำการสุ่มด้วยอัตราส่วน 80:20 ตามลำดับ (อุบลรัตน์ และจรัญ, 2562) จากแต่ละประเภทของรูปภาพ เพื่อให้ได้ข้อมูลที่ครอบคลุมทุกประเภทของรูปภาพโดยจะแบ่งได้ดังตารางที่ 3.4

ตารางที่ 3.4 จำนวนของรูปภาพในอัตราส่วน 80:20 สำหรับการตรวจจับวัตถุ

ประเภทของรูปภาพ	ชุดข้อมูลฝึกสอน (Train set) 80%	ชุดข้อมูลทดสอบ (Test set) 20%	รวม
เฉพาะโลโก้รูปแบบที่ 1	8	2	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

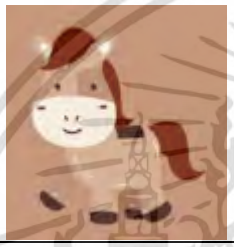
ประเภทของรูปภาพ	ชุดข้อมูลฝึกสอน (Train set) 80%	ชุดข้อมูลทดสอบ (Test set) 20%	รวม
มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 1	354	88	442
มีทั้งหมายเลขของสินค้าและโลโก้ของสินค้ารูปแบบที่ 2	19	5	24
รวม	381	95	476

จากตารางที่ 3.4 แสดงจำนวนของรูปภาพในอัตราส่วน 80:20 ของแต่ละประเภทของรูปภาพเป็นสำหรับชุดข้อมูลฝึกสอน 80 เปอร์เซนต์ และสำหรับชุดข้อมูลทดสอบ 20 เปอร์เซนต์ ดังนั้นจากรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 จำนวน 10 ภาพ จะได้ชุดข้อมูลฝึกสอน 8 ภาพ และชุดข้อมูลทดสอบ 2 ภาพ จากรูปภาพที่มีทั้งหมายเลขและโลโก้ของสินค้ารูปแบบที่ 1 จำนวน 442 ภาพ จะได้ชุดข้อมูลฝึกสอน 354 ภาพ และชุดข้อมูลทดสอบ 88 ภาพ และจากรูปภาพที่มีทั้งหมายเลขและโลโก้ของสินค้ารูปแบบที่ 2 จำนวน 24 ภาพ จะได้ชุดข้อมูลฝึกสอน 19 ภาพ และชุดข้อมูลทดสอบ 5 ภาพ ดังนั้นจากรูปภาพทั้งหมด 476 ภาพจะได้ชุดข้อมูลฝึกสอน 381 ภาพ และชุดข้อมูลทดสอบ 95 ภาพ

ตารางที่ 3.5 จำนวนของวัตถุแต่ละประเภทในแต่ละชุดข้อมูลสำหรับการตรวจจับวัตถุ

ประเภทของวัตถุ	ชุดข้อมูลฝึกสอน (Train set) 80%	ชุดข้อมูลทดสอบ (Test set) 20%	รวม
หมายเลขสินค้า A231222 010 2C 141 R 0309 N	373	93	466

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทของวัตถุ	ชุดข้อมูลฝึกสอน (Train set) 80%	ชุดข้อมูลทดสอบ (Test set) 20%	รวม
โลโก้สินค้ารูปแบบที่ 1	362	90	452
โลโก้สินค้ารูปแบบที่ 2 	19	5	24
รวม	754	188	942

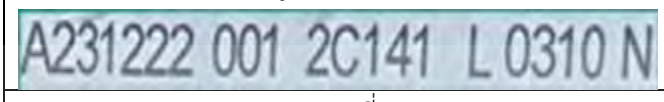
จากตารางที่ 3.5 แสดงจำนวนของวัตถุแต่ละประเภทที่ได้จากรูปภาพแต่ละประเภทของรูปภาพที่ได้ทำการแยกองค์ประกอบจากรูปภาพที่มีเฉพาะโลโก้ของสินค้ารูปแบบที่ 1 จำนวน 10 ภาพ มีวัตถุ 1 ประเภทในรูปภาพคือ โลโก้สินค้ารูปแบบที่ 1 จำนวน 10 รายการ จากรูปภาพที่มีทั้งหมายเลขและโลโก้ของสินค้ารูปแบบที่ 1 จำนวน 442 ภาพ มีวัตถุ 2 ประเภทในรูปภาพคือ หมายเลขสินค้าจำนวน 442 รายการ และโลโก้สินค้ารูปแบบที่ 1 จำนวน 442 รายการ และจากรูปภาพที่มีทั้งหมายเลขและโลโก้ของสินค้ารูปแบบที่ 2 จำนวน 24 ภาพ มีวัตถุ 2 ประเภทในรูปภาพคือ หมายเลขสินค้าจำนวน 24 รายการ และโลโก้สินค้ารูปแบบที่ 1 จำนวน 24 รายการ ดังนั้นจากรูปภาพทั้งหมด 476 รูปภาพ จะได้จำนวนวัตถุประเภทของ หมายเลขสินค้าจำนวน 466 รายการ โลโก้สินค้ารูปแบบที่ 1 จำนวน 452 รายการ และโลโก้สินค้ารูปแบบที่ 2 จำนวน 24 รายการ รวมทั้งสิ้นจากรูปภาพจำนวน 476 ภาพมีจำนวนวัตถุทั้งหมด 942 รายการ

3.5.2 การแบ่งชุดข้อมูลสำหรับการรู้จำอักขระด้วยแสง

จากชุดข้อมูลรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 1 จำนวน 100 ภาพ และรูปภาพที่มีเฉพาะหมายเลขของสินค้ารูปแบบที่ 2 จำนวน 100 ภาพ รวมทั้งสิ้นเป็นรูปภาพจำนวน 200 รูป เพื่อการสร้างแบบจำลองการตรวจจับวัตถุจะทำการแบ่งชุดข้อมูลด้วยวิธี Split Test ออกเป็น 2 ส่วนคือ ชุดข้อมูลฝึกสอน (Train set) และชุดข้อมูลทดสอบ (Test set) ด้วยอัตราส่วน 80:20 ตามลำดับ (อุบลรัตน์ และจรัญ, 2562) โดยจะแบ่งได้ดังตารางที่ 3.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 แสดงจำนวนของรูปภาพในอัตราส่วน 80:20 สำหรับการรู้จำอักขระด้วยแสง

รูปแบบตัวอักษร	ชุดข้อมูลฝึกสอน (Train set) 80%	ชุดข้อมูลทดสอบ (Test set) 20%	รวม
รูปแบบที่ 1 	80	20	100
รูปแบบที่ 2	80	20	100
รวม	160	40	200

จากตารางที่ 3.6 แสดงจำนวนของรูปภาพหมายเลขสินค้าในอัตราส่วน 80:20 ของแต่ละรูปแบบของตัวอักษรเป็นสำหรับชุดข้อมูลฝึกสอน 80 เปอร์เซนต์ และสำหรับชุดข้อมูลทดสอบ 20 เปอร์เซนต์ ดังนั้นจากรูปภาพของหมายเลขสินค้าจำนวน 200 ภาพ สามารถแบ่งเป็นรูปภาพที่เป็นรูปแบบของตัวอักษรรูปแบบที่ 1 จำนวน 100 ภาพ เป็นชุดข้อมูลฝึกสอนจำนวน 80 ภาพ และชุดข้อมูลทดสอบจำนวน 20 ภาพ และรูปภาพที่เป็นรูปแบบของตัวอักษรรูปแบบที่ 2 จำนวน 100 ภาพ จะได้ชุดข้อมูลฝึกสอนจำนวน 80 ภาพ และชุดข้อมูลทดสอบจำนวน 20 ภาพ ดังนั้นจากรูปภาพหมายเลขสินค้าทั้งหมด 200 ภาพจะได้ชุดข้อมูลฝึกสอนจำนวน 160 ภาพ และชุดข้อมูลทดสอบจำนวน 40 ภาพ

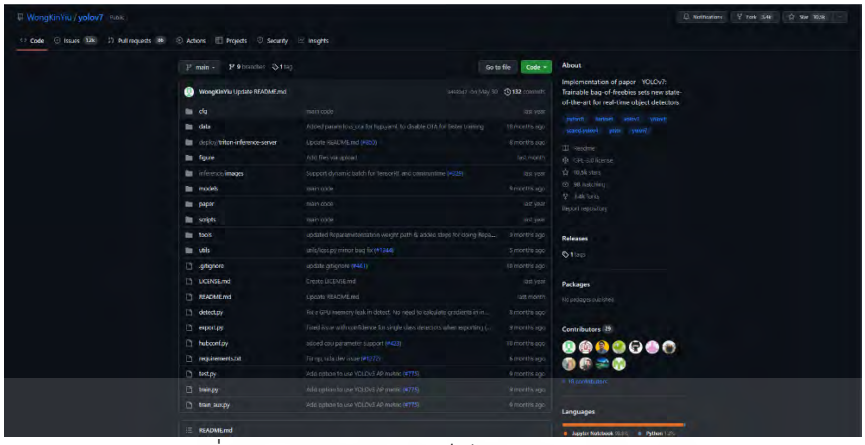
3.6 ขั้นตอนในการสร้างแบบจำลองการตรวจจับวัตถุ

3.6.1 การฝึกสอนแบบจำลอง

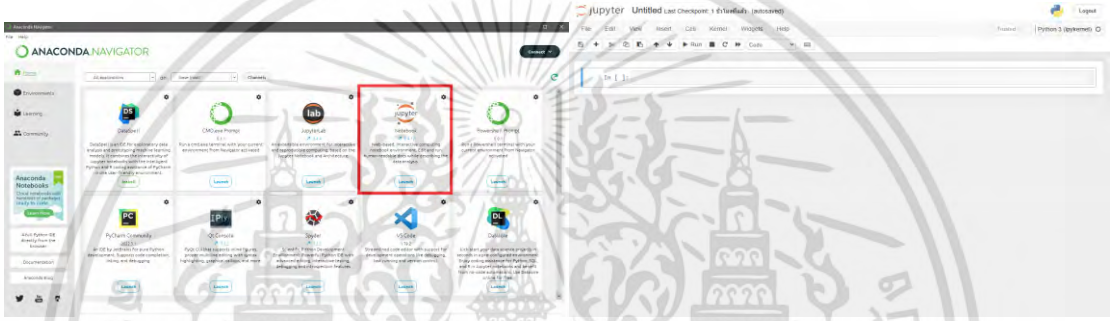
ในงานวิจัยนี้ได้เลือกใช้การตรวจจับวัตถุด้วยอัลกอริทึม โยโล่เวอร์ชันที่ 7 (YOLOv7) ในการสร้างแบบจำลองการตรวจจับหมายเลขสินค้าและโลโก้ของสินค้า โดยใช้ชุดข้อมูลสำหรับการฝึกสอนจำนวน 381 ภาพ มีขั้นตอนการทำงานดังนี้

1. ทำการดาวน์โหลด (Download) โยโล่เวอร์ชันที่ 7 (YOLOv7) จากทางเว็บไซต์ กิตฮับ (Github) ของทางโยโล่เวอร์ชันที่ 7 (YOLOv7) (WongKinYiu, 2022) https://github.com/WongKinYiu/yolov7 ด้วยคำสั่ง git clone โดยใช้ jupyter notebook (รูปที่ 3.25) ผ่านโปรแกรม Anaconda ดังรูปที่ 3.24

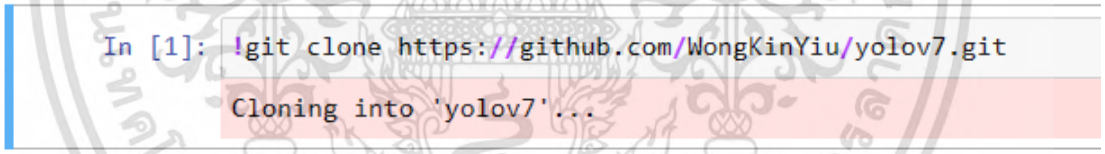
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.23 ตัวอย่างหน้าเว็บไซต์กิตฮับ YOLOv7



รูปที่ 3.24 jupyter notebook จากโปรแกรม Anaconda



รูปที่ 3.25 คำสั่ง git clone ทำการดาวน์โหลดโยโล่เวอร์ชันที่ 7 (YOLOv7)

2. ติดตั้งไลบรารี่ (Library) ที่จำเป็นทั้งหมดสำหรับการใช้งาน โดยภายในไฟล์ของ โยโล่เวอร์ชันที่ 7 (YOLOv7) จะมีไฟล์ที่ชื่อว่า requirements.txt ที่ภายในไฟล์จะทำการรวมทุกไลบรารี่ (Library) ที่จำเป็นสำหรับการใช้งานโดยติดตั้งผ่านทาง jupyter notebook ดังรูปที่ 3.26 โดยมีรายละเอียดของไลบรารี่ ดังนี้

2.1 Matplotlib ใช้สำหรับการสร้างกราฟและแผนภาพ

2.2 Numpy ใช้สำหรับการทำงานกับอาเรย์หลายมิติและการคำนวณทางคณิตศาสตร์

2.3 OpenCv ใช้ในการประมวลผลภาพและวิดีโอ โดยมีฟังก์ชันหลากหลายสำหรับการแปลงภาพ ตรวจจับวัตถุ สกัดคุณลักษณะ

2.4 Pillow ใช้ในการจัดการภาพ รวมถึงการโหลดและบันทึกภาพในรูปแบบต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 PyYAML ใช้ในการอ่านและเขียนไฟล์ในรูปแบบ YAML (YAML Ain't Markup Language) ที่ใช้ในการกำหนดค่าและการกำหนดแบบแผน

2.6 requests ใช้ในการทำการส่งคำขอ HTTP เพื่อเชื่อมต่อกับเซิร์ฟเวอร์หรือรับส่งข้อมูลผ่านเครือข่าย ใช้ในการดาวน์โหลดหรือส่งไฟล์ที่เกี่ยวข้อง

2.7 SciPy ใช้ในการทำงานทางวิทยาศาสตร์ รวมถึงการจัดการข้อมูลทางคณิตศาสตร์ ใช้ในกระบวนการประมวลผลและการวิเคราะห์ข้อมูล

2.8 Torch เป็นไลบรารีที่เชื่อมโยงกับการเรียนรู้เชิงลึก (Deep Learning) โดยมีโมเดลปัญญาประดิษฐ์ที่สนับสนุนต่าง ๆ ใช้ในการสร้างและฝึกโมเดลปัญญาประดิษฐ์

2.9 torchvision ช่วยในการประมวลผลภาพ รวมถึงการโหลดและจัดการกับชุดข้อมูลภาพ ใช้ในการประมวลผลภาพและการทำงานกับโมเดลปัญญาประดิษฐ์

2.10 tqdm ใช้สร้างแถบความคืบหน้า (progress bar) ที่แสดงว่างานวิจัยหรือกระบวนการทำงานกำลังดำเนินอยู่อย่างไร ช่วยให้รับรู้และติดตามความคืบหน้าของการทำงานในเวลาที่กำหนด

รูปที่ 3.26 ติดตั้งไลบรารี่ (Library) ที่จำเป็นทั้งหมดโดยภายไฟล์ที่ชื่อว่า “requirements.txt”

3. การเตรียมข้อมูลสำหรับการฝึกสอนแบบจำลอง สร้าง 2 โฟลเดอร์ สำหรับการฝึกสอน (Train) และสำหรับการทดสอบ (Test) โดยสำหรับการฝึกสอนเป็น 381 ภาพ และสำหรับการทดสอบเป็น 95 ภาพ จากรูปภาพทั้งหมด 476 ภาพ ที่ได้ทำการแบ่งไว้จากขั้นตอนการแบ่งชุดข้อมูลสำหรับการตรวจจับวัตถุ

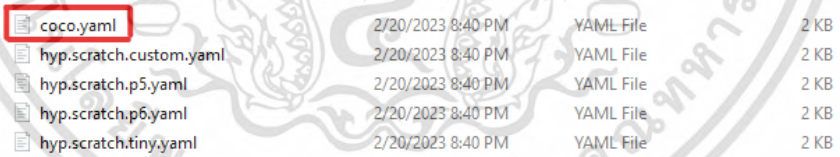
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.27 รูปตัวอย่างโฟลเดอร์สำหรับการฝึกสอน และสำหรับการทดสอบ

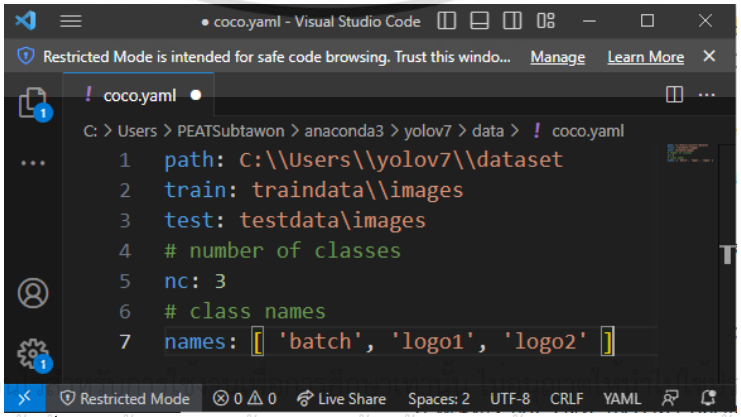
โดยภายในแต่ละโฟลเดอร์จะมีโฟลเดอร์ 2 โฟลเดอร์โดยมี โฟลเดอร์รูปภาพ (Images) และป้ายกำกับ (Labels) ที่ได้จากการทำการกำกับประเภทของวัตถุบนรูปภาพ ด้วยเว็บไซต์ เมคเซ็นส์ (Make Sense)

รูปที่ 3.28 รูปตัวอย่างโฟลเดอร์รูปภาพ และป้ายกำกับ

4. ทำการแก้ไขไฟล์กำหนดค่าจากไฟล์ coco.yaml ที่ได้รับจากภายในโฟลเดอร์ data ของโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยกำหนดเส้นทางโฟล์เดอร์รูปภาพที่ใช้สำหรับการฝึกสอน และสำหรับการทดสอบ ต่อจากนั้นให้กำหนดจำนวนของประเภทของวัตถุที่ใช้ในการฝึกสอน และชื่อของประเภทของวัตถุที่ใช้ฝึกสอน โดยมีวัตถุทั้งหมด 3 ประเภท คือ หมายเลขสินค้า (batch) โลโก้รูปแบบที่ 1 (logo1) และโลโก้รูปแบบที่ 2 (logo2) โดยแก้ไขไฟล์ผ่านโปรแกรม Visual Studio Code (VS Code) ดังรูปที่ 3.30



รูปที่ 3.29 ไฟล์ coco.yaml



รูปที่ 3.30 ตัวอย่างของการกำหนดค่าต่าง ๆ จากไฟล์ coco.yaml

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. ดาวน์โหลด (Download) ชุดน้ำหนักที่ได้รับการฝึกสอนล่วงหน้าสำหรับแบบจำลองโยโล่เวอร์ชันที่7 (Pre-trained YOLOv7 Weights) จากทางเว็บไซต์ กิตฮับ (Github) ของทางโยโล่เวอร์ชันที่ 7 (YOLOv7) (WongKinYiu, 2022) จากนั้นให้ย้ายไฟล์ yolov7.pt ที่ได้ดาวน์โหลด (Download) มาไปยังโฟลเดอร์ของโยโล่เวอร์ชันที่7 (YOLOv7)

MS COCO

Model	Test Size	AP^{test}	AP_{50}^{test}	AP_{75}^{test}	batch 1 fps	batch 32 average time
YOLOv7	640	51.4%	69.7%	55.9%	161 *fps*	2.8 *ms*
YOLOv7-X	640	53.1%	71.2%	57.8%	114 *fps*	4.3 *ms*
YOLOv7-W6	1280	54.9%	72.6%	60.1%	84 *fps*	7.6 *ms*
YOLOv7-E6	1280	56.0%	73.5%	61.2%	56 *fps*	12.3 *ms*
YOLOv7-D6	1280	56.6%	74.0%	61.8%	44 *fps*	15.0 *ms*
YOLOv7-E6E	1280	56.8%	74.4%	62.1%	36 *fps*	18.7 *ms*

รูปที่ 3.31 ชุดน้ำหนักที่ได้รับการฝึกสอนล่วงหน้าสำหรับแบบจำลองโยโล่เวอร์ชันที่7

6. เริ่มการฝึกสอนแบบจำลอง ด้วยคำสั่ง train,py เพื่อเริ่มกระบวนการฝึกสอนแบบจำลอง โดยตั้งค่าพารามิเตอร์ที่เกี่ยวข้องดังต่อไปนี้ batch = 5 , epochs = 300 , data = coco.yaml , weights = yolov7.pt และ device = 0 ซึ่งแต่ละพารามิเตอร์มีความหมายดังนี้

1. batch = 5 คือการกำหนดขนาดของชุดข้อมูล (batch size) กำหนดเป็น 5 ซึ่งหมายถึงการส่งข้อมูลเข้าสู่แบบจำลองเป็นชุด 5 รูปภาพพร้อมกันในแต่ละรอบการฝึกสอน ซึ่งการกำหนดขนาดของ batch เปลี่ยนตามการ์ดจอประมวลผลภาพ (GPU) ของเครื่องที่ใช้ หากการฝึกสอนเกิดข้อผิดพลาดของหน่วยความจำให้ทำการลดขนาดของ batch ลง
2. epochs = 300 คือการกำหนดจำนวนรอบการฝึกสอนเป็น 300 รอบ ซึ่งหมายถึงจำนวนครั้งที่แบบจำลองจะรับข้อมูลฝึกสอนทั้งหมด 300 รอบ
3. data = coco.yaml คือการกำหนดไฟล์ข้อมูล (data file) สำหรับการฝึกสอน ไฟล์นี้มีคำอธิบายของวัตถุต่าง ๆ ที่แบบจำลองควรจะรู้จัก และต้องฝึกสอน
4. weights = yolov7.pt คือการกำหนดไฟล์น้ำหนัก (weights file) เริ่มต้นของแบบจำลองเป็น yolov7.pt ซึ่งอาจเป็นไฟล์ที่บันทึกการเรียนรู้ของแบบจำลองก่อนหน้านี้ แบบจำลองจะใช้น้ำหนักเริ่มต้นเหล่านี้เพื่อเริ่มกระบวนการฝึกสอน
5. device = 0 คือการกำหนดอุปกรณ์ (device) ที่ใช้ในการฝึกสอนเป็นอุปกรณ์หมายเลข 0 ซึ่งหมายถึงการใช้งานการ์ดจอ (GPU) ในการประมวลผล

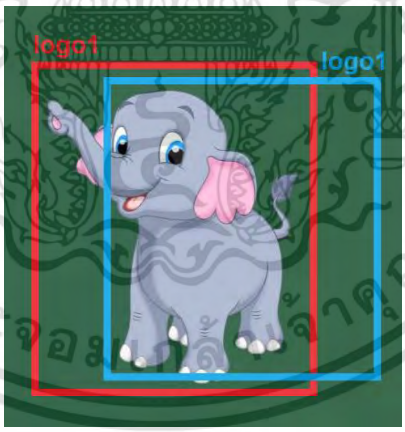
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# train
!python train.py --batch 5 --epochs 20 --data cococustom.yaml --weights 'yolov7.pt' --device 0

  0%|          | 0/71 [00:00<?, ?it/s]
     0/299     1.05G   0.06059   0.02663    0.0232    0.1104        22       640:   0%|          | 0/71 [00:12<?, ?it/s]
     0/299     1.05G   0.06059   0.02663    0.0232    0.1104        22       640:   1%|1         | 1/71 [00:12<14:06, 12.09s/i
t]
     0/299     4.16G     0.067   0.02439   0.02292    0.1143        14       640:   1%|1         | 1/71 [00:12<14:06, 12.09s/i
t]
     0/299     4.16G     0.067   0.02439   0.02292    0.1143        14       640:   3%|2         | 2/71 [00:12<05:56,  5.16s/i
t]
     0/299     4.16G   0.06954   0.02403   0.02281    0.1164        23       640:   3%|2         | 2/71 [00:12<05:56,  5.16s/i
t]
     0/299     4.16G   0.06954   0.02403   0.02281    0.1164        23       640:   4%|4         | 3/71 [00:12<03:19,  2.94s/i
```

รูปที่ 3.32 ตัวอย่างคำสั่งเริ่มกระบวนการฝึกสอนแบบจำลอง

3.6.2 ทดสอบประสิทธิภาพของแบบจำลองในการตรวจจับหมายเลขและโลโก้ของสินค้า

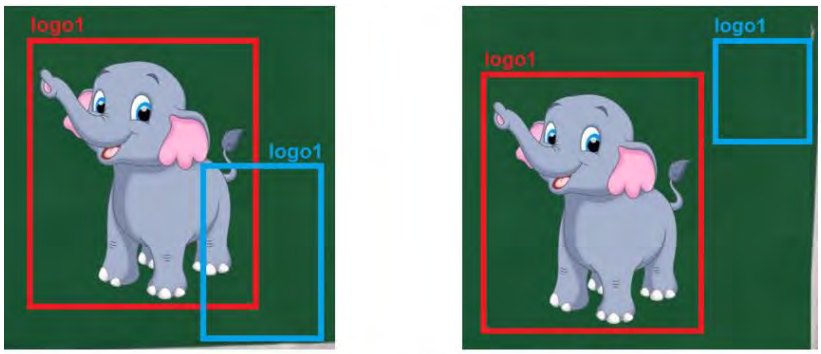
ทดสอบประสิทธิภาพของแบบจำลองการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ของอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยใช้ชุดข้อมูลสำหรับการทดสอบจำนวน 95 ภาพ ทดสอบประสิทธิภาพของแบบจำลองโดยพิจารณาจากค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผล เนื่องจากต้องการจะเปรียบเทียบกับเครื่องมือการตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน โดยเปรียบค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผล โดยกำหนดให้มีค่า IoU เท่ากับ 0.5 เป็นค่าที่นิยมใช้ในการทดสอบประสิทธิภาพของแบบจำลอง (ธนดล, 2564) ถ้าค่า IoU เท่ากับ 0.5 และแบบจำลองทำนายประเภทของวัตถุถูกต้องจะถือว่าการตรวจจับวัตถุนั้นตรวจจับถูกต้อง (True Positive: TP) ดังรูปที่ 3.33



รูปที่ 3.33 กำหนดเป็น TP เมื่อค่า IoU > 0.5 และมีการทำนายประเภทวัตถุที่ถูกต้อง

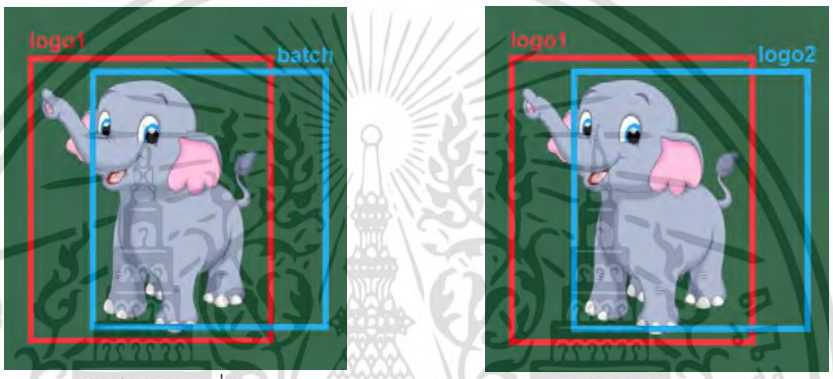
ถ้าค่า IoU น้อยกว่า 0.5 และแบบจำลองทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพจะถือว่าการตรวจจับวัตถุนั้นตรวจจับผิด (False Positive: FP) ดังรูปที่ 3.34 และถ้าค่า IoU มากกว่า 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้องจะถือว่าการตรวจจับวัตถุนั้นตรวจจับไม่เจอวัตถุ (False Negative: FN) ดังรูปที่ 3.35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



IoU < 0.5 No IoU

รูปที่ 3.34 กำหนดเป็น FP เมื่อค่า IoU < 0.5 และมีการทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพ



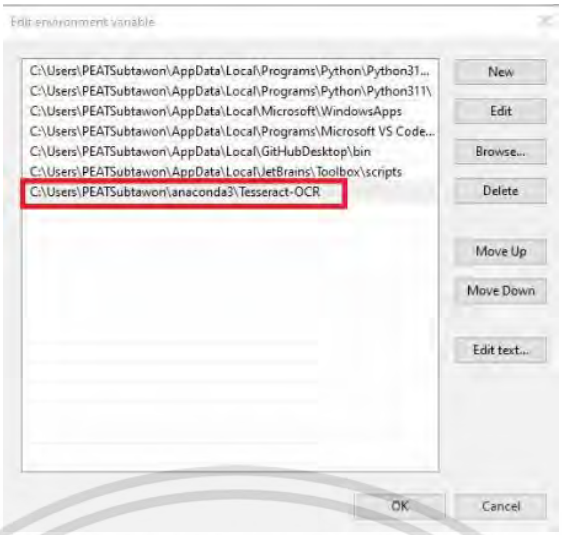
รูปที่ 3.35 กำหนดเป็น FN เมื่อค่า IoU > 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้อง

3.7 ขั้นตอนในการสร้างแบบจำลองการรู้จำอักขระด้วยแสง

3.7.1 การติดตั้งเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

ทำการดาวน์โหลด (Download) เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) จากทางเว็บไซต์ โดยในงานวิจัยนี้ได้เลือกใช้ เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เวอร์ชันที่ 5.3.0 หลังจากทำการติดตั้ง เทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เสร็จสิ้น ทำการนำเส้นทาง (Path) ของโฟลเดอร์ที่ได้ทำการติดตั้ง นำไปแก้ไขในเอ็นไวรอนเม้นท์ (Environment) ของเครื่องคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.36 ตัวอย่างหน้าต่างแก้ไข Environment

3.7.2 การฝึกสอนแบบจำลองเทสเซอร์แรคต์ (Tesseract)

1. สร้างไฟล์นามสกุล .tr ซึ่งมีหน้าที่รวมไฟล์ภาพ และไฟล์ .box ที่ได้จากขั้นตอนการกำกับตัวอักษรเข้าด้วยกัน ด้วยคำสั่งโปรแกรมเทสเซอร์แรคต์ (Tesseract) ดังนี้

คำสั่ง : tesseract [ชื่อภาษา].[ชื่อแบบอักษร].[expN].[นามสกุลไฟล์] [ชื่อภาษา].[ชื่อแบบอักษร].[expN] box.train

ตัวอย่าง : tesseract train.my.exp0.jpg train.my.exp0 box.train

รูปที่ 3.37 ตัวอย่างไฟล์นามสกุล .tr

2. สร้างไฟล์ข้อมูลชื่อ unicharset ซึ่งทำหน้าที่เก็บคุณลักษณะของแต่ละภาพตัวอักษร

คำสั่ง : unicharset_extractor [ชื่อภาษา].[ชื่อแบบอักษร].[expN].box

ตัวอย่าง : unicharset_extractor train.my.exp0.box

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.38 unicharset

3. สร้างไฟล์ข้อมูล font_properties ไฟล์ font_properties ทำหน้าที่กำหนดคุณสมบัติของตัวอักษร เพื่อให้ระบบสามารถจัดรูปแบบ และอ่านตัวอักษรได้อย่างถูกต้อง

คำสั่ง : echo "[ชื่อแบบอักษร] [ตัวเอียง (0 หรือ 1)] [ตัวหนา (0 หรือ 1)] [ความกว้างเท่ากัน (0 หรือ 1)] [แบบมีเชิง (0 หรือ 1)] [ลวดลายซับซ้อน (0 หรือ 1)]" > font_properties

ตัวอย่าง : echo "my 0 0 0 0 0" > font_properties

โดย 1 หมายถึงมีแอตทริบิวต์ และ 0 หมายถึงไม่มีแอตทริบิวต์

รูปที่ 3.39 font_properties

4. สร้างไฟล์ฝึกสอน ประกอบด้วย 3 ไฟล์หลัก shapetable เป็นไฟล์ที่เก็บข้อมูลเกี่ยวกับรูปแบบ และรูปทรงของตัวอักษร inttemp เป็นไฟล์ที่เก็บข้อมูลเกี่ยวกับคุณสมบัติต่าง ๆ ของตัวอักษร และ pffmtable เป็นไฟล์ที่เก็บข้อมูลเกี่ยวกับความน่าจะเป็นในการพบตัวอักษรในตำแหน่งและรูปทรงที่ต่าง ๆ ในภาพ

คำสั่ง : mftraining -F font_properties -U unicharset -O [ชื่อภาษา].unicharset [ชื่อภาษา].[ชื่อแบบอักษร].[expN].tr

ตัวอย่าง : mftraining -F font_properties -U unicharset -O train.unicharset train.my.exp0.tr

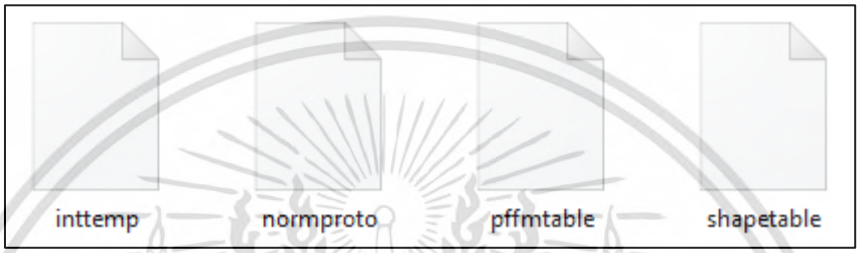
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. สร้างไฟล์ข้อมูล normproto เป็นไฟล์ที่ใช้ในกระบวนการปรับสภาพหรือปรับรูปแบบตัวอักษรในระบบการแปลงตัวอักษรหรือระบบการปรับปรุง และปรับให้ตัวอักษรเป็นรูปแบบที่เหมือนหรือใกล้เคียงกันทั้งในด้านรูปแบบ และการจัดวางตัวอักษร

คำสั่ง : cntraining [ชื่อภาษา].[ชื่อแบบอักษร].[expN].tr

ตัวอย่าง : cntraining train.my.exp0.tr

หลังจากทำตามขั้นตอนที่ 5 และขั้นตอนที่ 6 เสร็จสิ้นจะได้รับไฟล์สี่ไฟล์ ได้แก่ shapetable, inttemp, pffmtable, normproto



รูปที่ 3.40 shapetable, inttemp, pffmtable, normproto

6. ทำแก้ไขชื่อไฟล์ทั้ง 4 ไฟล์ข้อมูล shapetable , inttemp , pffmtable , normproto ให้เป็น [ชื่อภาษา].shapetable , [ชื่อภาษา].inttemp , [ชื่อภาษา].pffmtable , [ชื่อภาษา]. normproto

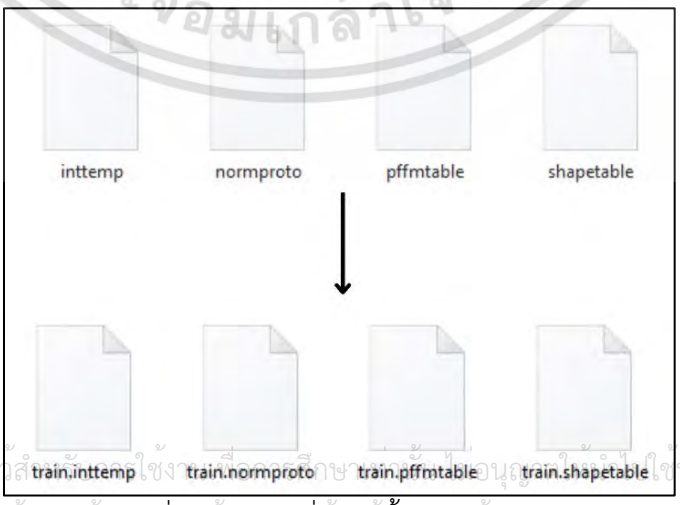
คำสั่ง : rename ชื่อไฟล์เดิม ชื่อไฟล์ใหม่

ตัวอย่าง : rename shapetable train.shapetable

rename inttemp train.inttemp

rename pffmtable train.pffmtable

rename normproto train.normproto



รูปที่ 3.41 แก้ไขชื่อไฟล์ทั้ง 4 ไฟล์ข้อมูล

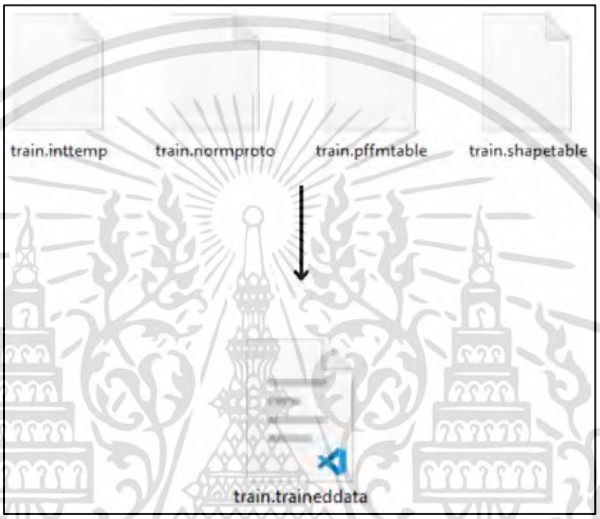
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้

7. สร้างไฟล์นามสกุล .traineddata โดยทำการรวมไฟล์ทุกไฟล์ที่ได้จากการฝึกแบบจำลองเข้าด้วยกันในรูปแบบของไฟล์เดียว ด้วยคำสั่งดังนี้

คำสั่ง : combine_tessdata [ชื่อภาษา].

ตัวอย่าง : combine_tessdata train.

หลังจากเสร็จสิ้นขั้นตอนจะได้ไฟล์นามสกุล .traineddata ให้ทำการย้ายไฟล์ดังกล่าวไปยังโฟลเดอร์ย่อยชื่อ tessdata ของโปรแกรม tesseract ก็เป็นอันเสร็จสิ้นกระบวนการ

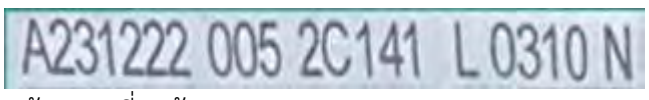


รูปที่ 3.42 ไฟล์นามสกุล .traineddata

3.7.3 ทดสอบประสิทธิภาพของแบบจำลองในการรู้จำข้อความของหมายเลขสินค้า

ทดสอบด้วยประสิทธิภาพของแบบจำลองการรู้จำอักขระด้วยแสงด้วยค่าความถูกต้อง (Accuracy) เพื่อเปรียบเทียบกับเครื่องมือการรู้จำอักขระด้วยแสงหมายเลขของสินค้า ของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน และหาค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) และระยะเวลาในการประมวลผล ตัวอย่างผลลัพธ์ที่ได้จากการทดสอบ

1. การทดสอบตัวอย่างที่ 1 จากข้อความหมายเลขสินค้าตัวอักษรรูปแบบที่ 1 จำนวน 21 ตัวอักษร



ข้อความที่ถูกต้อง = A231222 005 2C141 L 0310 N

ข้อความที่ทำนาย = 1A231222 008 201311 L 0810 0

รูปที่ 3.43 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.43 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 1 มีจำนวนตัวอักษรที่เพิ่มขึ้นจากข้อความที่ถูกต้อง 2 ตัวคือ “1”, “1” และมีจำนวนตัวอักษรที่เปลี่ยนไปจากข้อความที่ถูกต้อง 5 ตัวคือ “5” เป็น “8” , “C” เป็น “0” , “4” เป็น “3” , “3” เป็น “8” และ “N” เป็น “0” และไม่มีจำนวนตัวอักษรที่หายไปจากข้อความที่ถูกต้อง

จากสูตรการหาค่า $CER = \frac{I+S+D}{N}$

โดยที่ ค่า I คือจำนวนตัวอักษรที่เพิ่มขึ้น

ค่า S คือจำนวนตัวอักษรที่เปลี่ยนไป

ค่า D คือจำนวนตัวอักษรที่หายไป

และ ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $CER = \frac{2+5+0}{21} = 0.33$

จากสูตรการหาค่า $Accuracy = \frac{T}{N}$

โดยที่ ค่า T คือจำนวนตัวอักษรที่ทายถูกต้อง

ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $Accuracy = \frac{16}{21} = 0.76$

ผลลัพธ์ได้ค่าความผิดพลาดระดับตัวอักษร (CER) เท่ากับ 0.33 หรือคิดเป็น 33 เปอร์เซนต์ และค่าความถูกต้อง (Accuracy) เท่ากับ 0.76 หรือคิดเป็น 76 เปอร์เซนต์

2. การทดสอบตัวอย่างที่ 2 จากข้อความหมายเลขสินค้าตัวอักษรรูปแบบที่ 1 จำนวน 21 ตัวอักษร

ข้อความที่ถูกต้อง = A231222 001 2C141 L 0310 N

ข้อความที่ทำนาย = 223222 001 20711 L 0310 A3

รูปที่ 3.44 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 2

จากรูปที่ 3.44 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 2 มีจำนวนตัวอักษรที่เพิ่มขึ้นจากข้อความที่ถูกต้อง 1 ตัวคือ “3” มีจำนวนตัวอักษรที่เปลี่ยนไปจากข้อความที่ถูกต้อง 5 ตัวคือ “A” เป็น “2” , “C” เป็น “0” , “1” เป็น “7” , “4” เป็น “1” และ “N” เป็น “A” และมีจำนวนตัวอักษรที่หายไปจากข้อความที่ถูกต้อง 1 ตัวคือ “1”

จากสูตรการหาค่า $CER = \frac{I+S+D}{N}$

โดยที่ ค่า I คือจำนวนตัวอักษรที่เพิ่มขึ้น

ค่า S คือจำนวนตัวอักษรที่เปลี่ยนไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า D คือจำนวนตัวอักษรที่หายไป

และ ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $$CER = \frac{1+5+1}{21} = 0.33$$

จากสูตรการหาค่า $Accuracy = \frac{T}{N}$

โดยที่ ค่า T คือจำนวนตัวอักษรที่ทายถูกต้อง

ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $$Accuracy = \frac{15}{21} = 0.71$$

ผลลัพธ์ได้ค่าความผิดพลาดระดับตัวอักษร (CER) เท่ากับ 0.33 หรือคิดเป็น 33 เปอร์เซนต์ และค่าความถูกต้อง (Accuracy) เท่ากับ 0.71 หรือคิดเป็น 71 เปอร์เซนต์

3. การทดสอบตัวอย่างที่ 3 จากข้อความหมายเลขสินค้าตัวอักษรรูปแบบที่ 2 จำนวน 20 ตัวอักษร

ข้อความที่ถูกต้อง = A180818 0623K353 R 2215

ข้อความที่ทำนาย = A180818 082307383 8 2215

รูปที่ 3.45 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 3

จากรูปที่ 3.45 ตัวอย่างผลลัพธ์การทดสอบตัวอย่างที่ 3 มีจำนวนตัวอักษรที่เพิ่มขึ้นจากข้อความที่ถูกต้อง 1 ตัวคือ “7” มีจำนวนตัวอักษรที่เปลี่ยนไปจากข้อความที่ถูกต้อง 4 ตัวคือ “6” เป็น “8” , “K” เป็น “0” , “5” เป็น “8” และ “R” เป็น “8” และไม่มีจำนวนตัวอักษรที่หายไปจากข้อความที่ถูกต้อง

จากสูตรการหาค่า $CER = \frac{I+S+D}{N}$

โดยที่ ค่า I คือจำนวนตัวอักษรที่เพิ่มขึ้น

ค่า S คือจำนวนตัวอักษรที่เปลี่ยนไป

ค่า D คือจำนวนตัวอักษรที่หายไป

และ ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $$CER = \frac{1+4+0}{20} = 0.25$$

จากสูตรการหาค่า $Accuracy = \frac{T}{N}$

โดยที่ ค่า T คือจำนวนตัวอักษรที่ทายถูกต้อง

ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น $$Accuracy = \frac{16}{20} = 0.8$$

ผลลัพธ์ได้ค่าความผิดพลาดระดับตัวอักษร (CER) เท่ากับ 0.25 หรือคิดเป็น 25 เปอร์เซนต์ และค่าความถูกต้อง (Accuracy) เท่ากับ 0.8 หรือคิดเป็น 80 เปอร์เซนต์

3.8 การทำงานของแบบจำลองการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง

เมื่อได้รับแบบจำลองการตรวจจับวัตถุ (Object Detection) และแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition) ที่ถูกฝึกสอนแล้ว จึงทำการรวมทั้งสองแบบจำลองเข้าด้วยกัน เพื่อใช้ในการทำงานในขั้นตอนเดียว เมื่อได้รับผลลัพธ์จากการตรวจจับวัตถุ แบบจำลองตรวจจับวัตถุจะให้ข้อมูลเกี่ยวกับตำแหน่ง และขอบเขตของวัตถุที่ตรวจจับได้ ซึ่งจะถูกนำมาใช้เป็นข้อมูลนำเข้าให้กับแบบจำลองการรู้จำอักขระด้วยแสง เพื่อให้แบบจำลองรู้ว่าต้องรู้จำอักขระในบริเวณใดของภาพ โดยแบบจำลองการรู้จำอักขระด้วยแสงจะทำการวิเคราะห์ และรู้จำข้อความที่ปรากฏในบริเวณนั้น ๆ โดยการเพิ่มส่วนขยายของโค้ดเพื่อทำการรู้จำอักขระด้วยเทสเซอร์แรคต์ (Tesseract) ในไฟล์โค้ดที่ชื่อ detect.py ของทางโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยตั้งชื่อไฟล์ที่ได้ทำการรวมแบบจำลองการตรวจจับและการรู้จำอักขระว่า detectandocr.py

ขั้นตอนการทำงานของการตรวจจับหมายเลขสินค้าและโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ มีดังนี้

1. เริ่มต้นด้วยการรับรูปภาพของสินค้าที่ต้องการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และแปลงหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ เข้าแบบจำลองโยโล่เวอร์ชันที่ 7 (YOLOv7) ที่ได้ทำการฝึกสอนแล้ว

2. หลังที่นำรูปภาพสินค้าเข้าสู่อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) จะทำการตรวจจับวัตถุที่เป็นหมายเลขของสินค้า และโลโก้ของสินค้า แล้วทำการระบุตำแหน่งของวัตถุที่ตรวจพบ

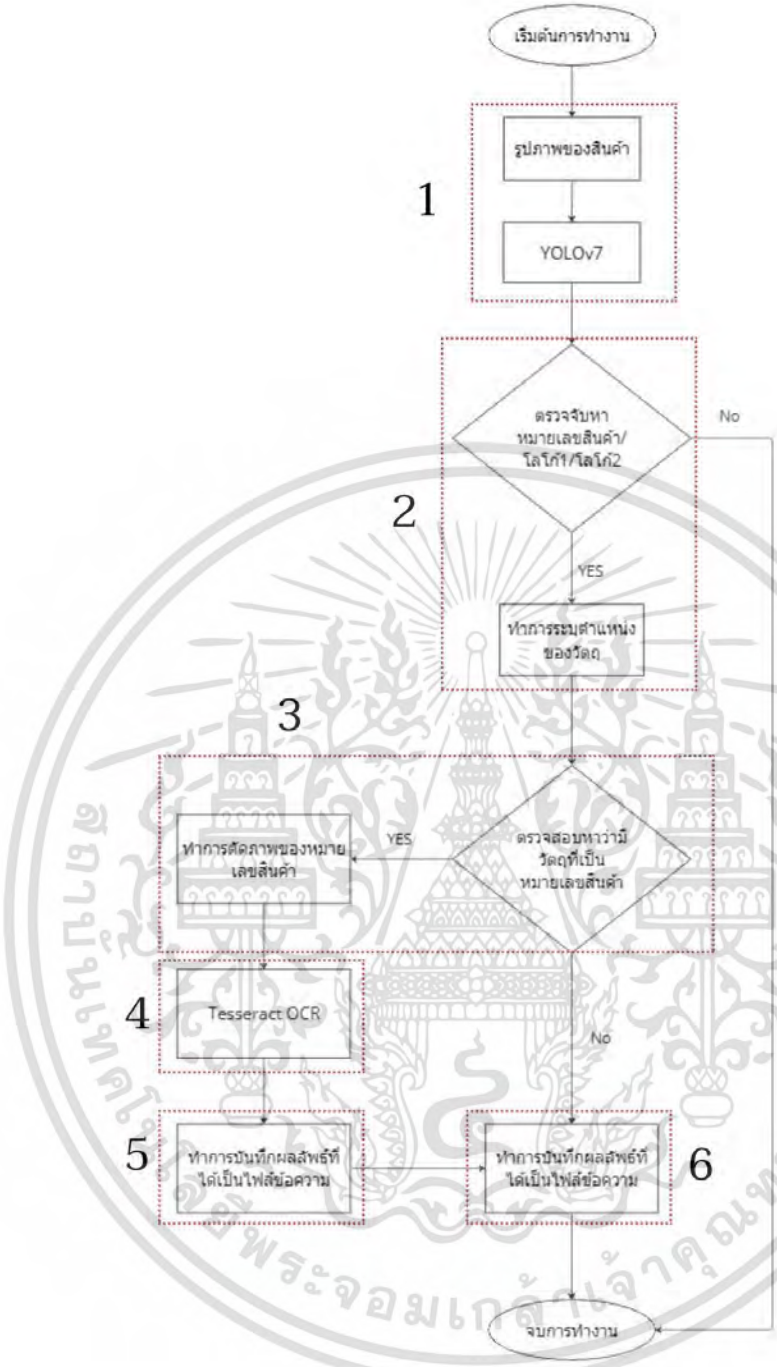
3. หลังจากทำการตรวจจับวัตถุภายในภาพ จะทำการตรวจสอบว่าวัตถุที่ตรวจพบนั้นเป็นหมายเลขของสินค้า หากพบวัตถุที่เป็นหมายเลขของสินค้า จะทำการตัดภาพของหมายเลขสินค้า โดยใช้ข้อมูลตำแหน่งจากขอบเขตที่ระบุตำแหน่งของหมายเลขสินค้า

4. หลังจากตัดภาพของหมายเลขสินค้า จะนำภาพที่ตัดออกมาได้นำไปทำการรู้จำอักขระด้วยแสงด้วยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) เพื่อแปลงภาพของหมายเลขสินค้าในอยู่ในรูปแบบของข้อความ

5. ทำการบันทึกผลลัพธ์ที่ได้จากการตรวจจับวัตถุ และรู้จำอักขระด้วยแสงในภาพ โดยบันทึกผลลัพธ์ที่ได้เป็นไฟล์ข้อความ

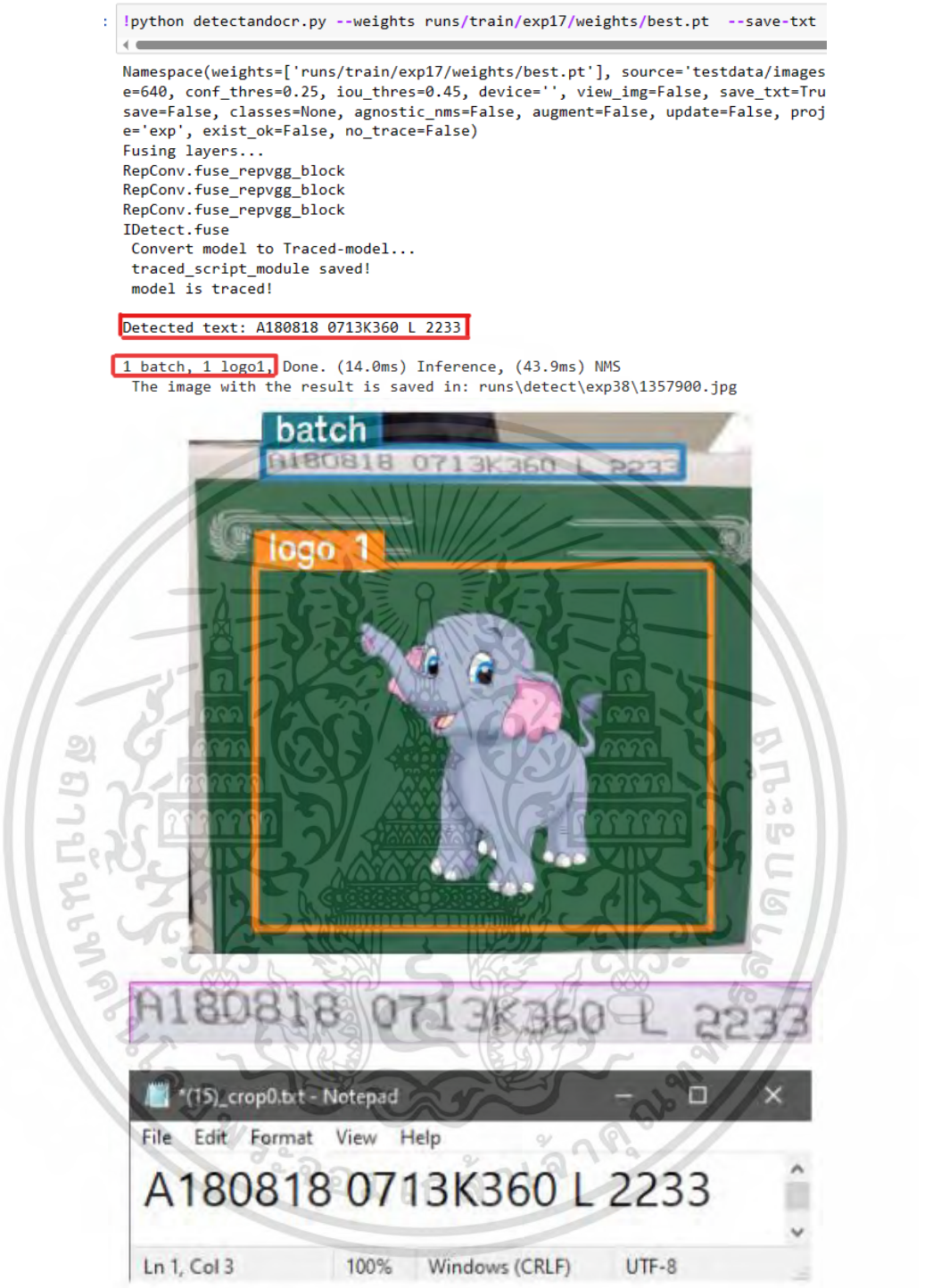
6. แสดงผลลัพธ์ที่ได้บนหน้าต่างของโปรแกรม jupyter notebook

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.46 แผนผังการทำงานของการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.47 ตัวอย่างผลลัพธ์ที่ได้จากการทำงานของการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง

จากรูปที่ 3.47 ผลลัพธ์ที่ได้จากการทำงานของการตรวจจับวัตถุ และการรู้จำอักขระด้วยแสง เป็นรูปภาพที่มีการระบุตำแหน่ง และชื่อของประเภทวัตถุ และข้อความจากการทำการรู้จำอักขระด้วยแสงของหมายเลขสินค้าที่ได้ทำการบันทึกเป็นรูปแบบของไฟล์ข้อความ โดยใช้โปรแกรม jupyter notebook เพื่อรันไฟล์ detectandocr.py

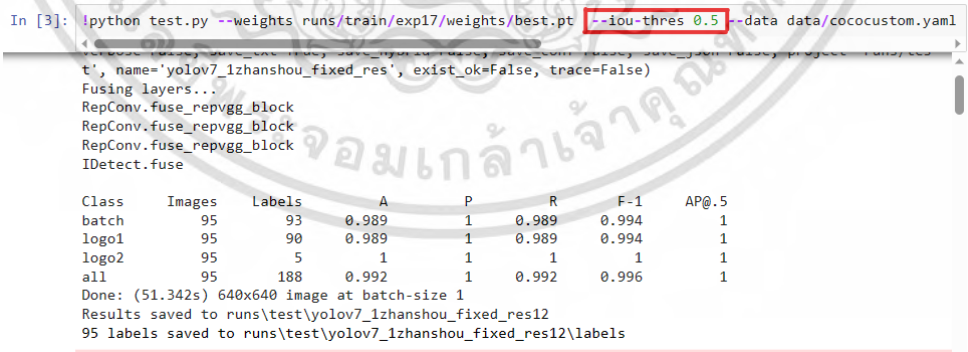
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

ในบทนี้แสดงถึงผลลัพธ์ที่ได้จากการทดสอบการตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าจากอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยทดสอบประสิทธิภาพการตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าด้วยค่าความถูกต้อง (Accuracy) และระยะเวลาที่ใช้ในการประมวลผล ที่ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) เท่ากับ 0.5 (ธนดล, 2564) และผลลัพธ์ที่ได้จากการทดสอบการรู้จำอักขระด้วยแสงของหมายเลขสินค้าด้วยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) โดยใช้ค่าความถูกต้อง (Accuracy) และระยะเวลาที่ใช้ในการประมวลผล เป็นค่าสำหรับทดสอบประสิทธิภาพของการรู้จำหมายเลขของสินค้า

4.1 การทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection)

ในการทดสอบแบบจำลองการตรวจจับวัตถุ (Object Detection) จากอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยใช้ชุดข้อมูลสำหรับการทดสอบเป็นจำนวน 95 ภาพ หรือคิดเป็น 20 เปอร์เซ็นต์จากข้อมูลทั้งหมด 476 ภาพ มีประเภทของวัตถุ 3 ประเภทคือ หมายเลขของสินค้า โลโก้สินค้ารูปแบบที่ 1 และโลโก้สินค้ารูปแบบที่ 2 โดยผลลัพธ์ที่ได้จากการทดสอบดังรูปที่ 4.1 และตารางที่ 4.1 แสดงรายละเอียดของจำนวนวัตถุของแต่ละประเภทวัตถุ ทดสอบโดยใช้โปรแกรม jupyter notebook รันไฟล์ชื่อ test.py ที่ได้รับจากภายในโฟลเดอร์ ของโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยกำหนดค่า IoU เท่ากับ 0.5



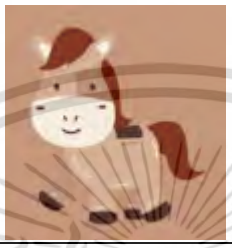
```
In [3]: !python test.py --weights runs/train/exp17/weights/best.pt --iou-thres 0.5 --data data/cococustom.yaml
t', name='yolov7_1zhanshou_fixed_res', exist_ok=False, trace=False)
Fusing layers...
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
IDetect.fuse

Class     Images     Labels          A          P          R      F-1      AP@.5
batch         95         93      0.989          1      0.989    0.994          1
logo1         95         90      0.989          1      0.989    0.994          1
logo2         95          5          1          1          1        1          1
all           95        188      0.992          1      0.992    0.996          1
Done: (51.342s) 640x640 image at batch-size 1
Results saved to runs\test\yolov7_1zhanshou_fixed_res12
95 labels saved to runs\test\yolov7_1zhanshou_fixed_res12\labels
```

รูปที่ 4.1 ผลลัพธ์จากโปรแกรมทดสอบการตรวจจับวัตถุ

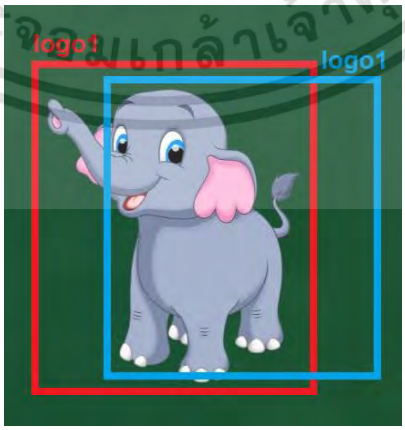
ตารางที่ 4.1 รายละเอียดจำนวนวัตถุของแต่ละประเภทวัตถุ

ประเภทของวัตถุ	จำนวนวัตถุ
หมายเลขสินค้า	93

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
A231222 010 2C 141 R 0309 N

ประเภทของวัตถุ	จำนวนวัตถุ
โลโก้สินค้ารูปแบบที่ 1	90
โลโก้สินค้ารูปแบบที่ 2 	5
รวม	188

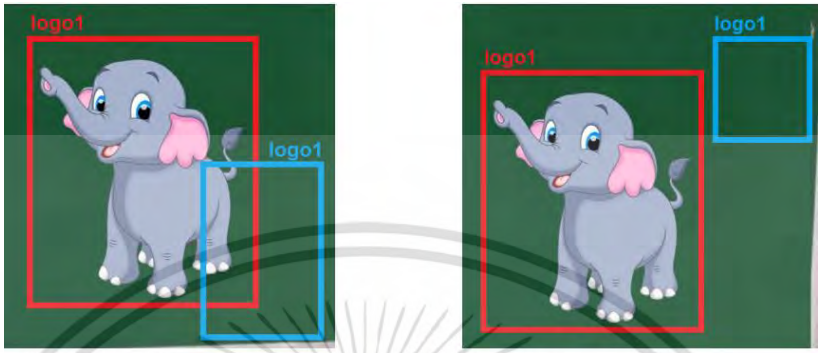
โดยการทดสอบประสิทธิภาพของแบบจำลองการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ของอัลกอริทึมโยโล่เวอร์ชันที่ 7 พิจารณาจากค่าความถูกต้อง (Accuracy) และระยะเวลาที่ใช้ในการประมวลผล เนื่องจากต้องการจะเปรียบเทียบกับเครื่องมือการตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้าของทางบริษัทกรณีศึกษาที่ใช้อยู่ในปัจจุบัน โดยเปรียบค่าความถูกต้อง (Accuracy) และ ระยะเวลาที่ใช้ในการประมวลผล โดยกำหนดให้มีค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (Intersection over Union: IoU) เท่ากับ 0.5 ดังรูปที่ 4.1 ถ้าค่า IoU เท่ากับ 0.5 และแบบจำลองทำนายประเภทของวัตถุถูกต้องจะถือว่าการตรวจจับวัตถุนั้นตรวจจับถูกต้อง (True Positive: TP) ดังรูปที่ 4.2



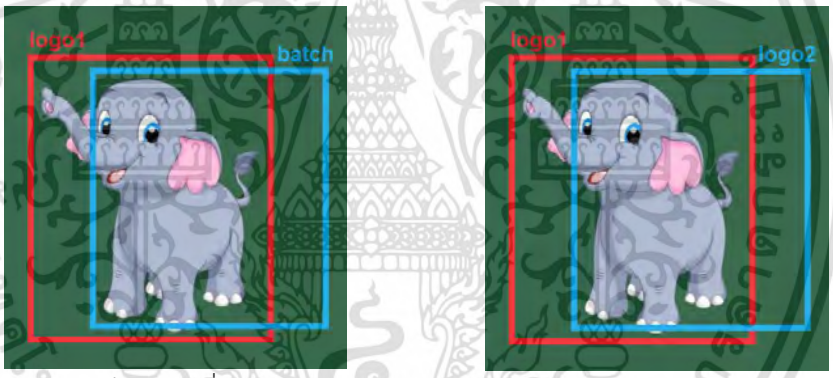
รูปที่ 4.2 กำหนดเป็น TP เมื่อค่า IoU > 0.5 และมีการทำนายประเภทวัตถุที่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้าค่า IoU น้อยกว่า 0.5 และแบบจำลองทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพจะถือว่าการตรวจจับวัตถุนั้นตรวจจับผิด (False Positive: FP) ดังรูปที่ 4.3 และถ้าค่า IoU มากกว่า 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้องจะถือว่าการตรวจจับวัตถุนั้นตรวจจับไม่เจอวัตถุ (False Negative: FN) ดังรูปที่ 4.4



รูปที่ 4.3 กำหนดเป็น FP เมื่อค่า IoU < 0.5 และมีการทำนายว่ามีวัตถุที่สนใจอยู่ในรูปภาพ



รูปที่ 4.4 กำหนดเป็น FN เมื่อค่า IoU > 0.5 แต่แบบจำลองทำนายประเภทของวัตถุไม่ถูกต้อง

1. ผลลัพธ์ของการตรวจจับประเภทของวัตถุหมายเลขสินค้า รูปภาพสินค้าสำหรับการทดสอบจำนวน 95 ภาพ มีจำนวนวัตถุของหมายเลขสินค้า 93 รายการ เนื่องจากมีรูปภาพของสินค้าที่ไม่มีหมายเลขสินค้าภายในภาพของสินค้าจำนวน 2 ภาพ ได้ผลลัพธ์ดังนี้

ตารางที่ 4.2 รายละเอียดการตรวจจับของหมายเลขสินค้า

ผลการตรวจจับของหมายเลขสินค้า	
จำนวนที่พบวัตถุ	93
จำนวนผลเฉลย	93
ทำนายถูก	92
ทำนายผิด	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.2 รายละเอียดของการตรวจจับของหมายเลขสินค้าจากรูปภาพ 95 ภาพ มีจำนวนวัตถุที่เป็นหมายเลขสินค้า 93 รายการ ได้ทำการทดสอบเทียบกับจำนวนผลเฉลย (Ground Truth) ผลลัพธ์ที่ได้ ตรวจพบวัตถุที่เป็นหมายเลขสินค้าจำนวน 93 รายการ จำนวนผลที่ทำนายถูกว่าคือหมายเลขสินค้าจำนวน 92 รายการ และทำนายผิด 1 รายการ จากผลลัพธ์ที่ได้จึงสามารถคำนวณหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

จากสูตร $$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

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

และ $$Recall = \frac{TP}{(TP+FN)}$$

โดยที่ ค่า TP คือจำนวนเมื่อค่า IoU > 0.5 และทำนายถูก

ค่า TN คือจำนวนทำนายถูกตรงกับผลเฉลยที่ไม่พบวัตถุ

ค่า FP คือจำนวนเมื่อค่า IoU < 0.5

ค่า FN คือจำนวนเมื่อค่า IoU > 0.5 แต่มีการทำนายผิด

ดังนั้น $$Accuracy = \frac{(92+0)}{(92+0+1+0)} = 0.989$$

$$Precision = \frac{92}{(92+0)} = 1$$

และ $$Recall = \frac{92}{(92+1)} = 0.989$$

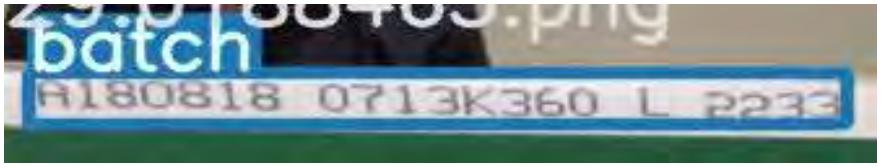
จึงได้ผลลัพธ์ค่าความถูกต้อง (Accuracy) เท่ากับ 0.989 ค่าความแม่นยำ (Precision) เท่ากับ 1 และค่าความระลึก (Recall) เท่ากับ 0.989 หลังจากทราบค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) สามารถที่จะหาค่าวัดประสิทธิภาพ (F1-Score)

จากสูตร $$F1 - Score = \frac{2 \times precision \times recall}{precision + recall}$$

ดังนั้น $$F1 - Score = \frac{2 \times 1 \times 0.989}{1 + 0.989} = 0.994$$

จึงได้ผลลัพธ์ค่าวัดประสิทธิภาพ (F1-Score) เท่ากับ 0.994

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.5 รูปภาพตัวอย่างที่ 1 ของการตรวจจับหมายเลขสินค้า

รูปที่ 4.6 รูปภาพตัวอย่างที่ 2 ของการตรวจจับหมายเลขสินค้า

2. ผลลัพธ์ของการตรวจจับประเภทของวัตถุโลโก้สินค้ารูปแบบที่ 1 รูปภาพสินค้าสำหรับการทดสอบจำนวน 95 ภาพ มีจำนวนวัตถุของโลโก้สินค้ารูปแบบที่ 1 จำนวน 90 รายการจาก 95 ภาพ ได้ผลลัพธ์ดังนี้

ตารางที่ 4.3 รายละเอียดการตรวจจับของโลโก้สินค้ารูปแบบที่ 1

ผลการตรวจจับของโลโก้สินค้ารูปแบบที่ 1	
จำนวนที่พบวัตถุ	90
จำนวนผลเฉลย	90
ทำนายถูก	89
ทำนายผิด	1

จากตารางที่ 4.3 รายละเอียดของการตรวจจับของโลโก้สินค้ารูปแบบที่ 1 จากรูปภาพ 95 ภาพ มีจำนวนวัตถุที่เป็นโลโก้สินค้ารูปแบบที่ 1 จำนวน 90 รายการ ได้ทำการทดสอบเทียบกับจำนวนผลเฉลย (Ground Truth) ผลลัพธ์ที่ได้ ตรวจพบวัตถุที่เป็นโลโก้สินค้ารูปแบบที่ 1 จำนวน 90 รายการ จำนวนผลที่ทำนายถูกว่าคือโลโก้สินค้ารูปแบบที่ 1 จำนวน 89 รายการ และทำนายผิด 1 รายการ จากผลลัพธ์ที่ได้จึงสามารถคำนวณหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

จากสูตร $$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

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

และ $$Recall = \frac{TP}{(TP+FN)}$$

โดยที่ ค่า TP คือจำนวนเมื่อค่า IoU > 0.5 และทำนายถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า TN คือจำนวนทำนายถูกตรงกับผลเฉลยที่ไม่พบวัตถุ

ค่า FP คือจำนวนเมื่อค่า IoU < 0.5

ค่า FN คือจำนวนเมื่อค่า IoU > 0.5 แต่มีการทำนายผิด

ดังนั้น $$Accuracy = \frac{(89+0)}{(89+0+1+0)} = 0.989$$

$$Precision = \frac{89}{(89+0)} = 1$$

และ $$Recall = \frac{89}{(89+1)} = 0.989$$

จึงได้ผลลัพธ์ค่าความถูกต้อง (Accuracy) เท่ากับ 0.989 ค่าความแม่นยำ (Precision) เท่ากับ 1 และค่าความระลึก (Recall) เท่ากับ 0.989 หลังจากทราบค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) สามารถที่จะหาค่าวัดประสิทธิภาพ (F1-Score)

จากสูตร $$F1 - Score = \frac{2 \times precision \times recall}{precision + recall}$$

ดังนั้น $$F1 - Score = \frac{2 \times 1 \times 0.988}{1 + 0.989} = 0.994$$

จึงได้ผลลัพธ์ค่าวัดประสิทธิภาพ (F1-Score) เท่ากับ 0.994

รูปที่ 4.7 รูปภาพตัวอย่างของการตรวจจับโลโก้สินค้ารูปแบบที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ผลลัพธ์ของการตรวจจับประเภทของวัตถุโลโก้สินค้ารูปแบบที่ 2 รูปภาพสินค้าสำหรับการทดสอบจำนวน 95 ภาพ มีจำนวนวัตถุของโลโก้สินค้ารูปแบบที่ 2 จำนวน 5 รายการจาก 95 ภาพ ได้ผลลัพธ์ดังนี้

ตารางที่ 4.4 รายละเอียดการตรวจจับของโลโก้สินค้ารูปแบบที่ 2

ผลการตรวจจับของโลโก้สินค้ารูปแบบที่ 2	
จำนวนที่พบวัตถุ	5
จำนวนผลเฉลย	5
ทำนายถูก	5
ทำนายผิด	0

จากตารางที่ 4.4 รายละเอียดของการตรวจจับของโลโก้สินค้ารูปแบบที่ 2 จากรูปภาพ 95 ภาพ มีจำนวนวัตถุที่เป็นโลโก้สินค้ารูปแบบที่ 2 จำนวน 5 รายการ ได้ทำการทดสอบเทียบกับจำนวนผลเฉลย (Ground Truth) ผลลัพธ์ที่ได้ ตรวจพบวัตถุที่เป็นโลโก้สินค้ารูปแบบที่ 2 จำนวน 5 รายการ จำนวนผลที่ทำนายถูกว่าเป็นโลโก้สินค้ารูปแบบที่ 2 จำนวน 5 รายการ และไม่มีการทำนายผิด จากผลลัพธ์ที่ได้จึงสามารถคำนวณหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

จากสูตร $$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

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

และ $$Recall = \frac{TP}{(TP+FN)}$$

โดยที่ ค่า TP คือจำนวนเมื่อค่า IoU > 0.5 และทำนายถูก

ค่า TN คือจำนวนทำนายถูกตรงกับผลเฉลยที่ไม่พบวัตถุ

ค่า FP คือจำนวนเมื่อค่า IoU < 0.5

ค่า FN คือจำนวนเมื่อค่า IoU > 0.5 แต่มีการทำนายผิด

ดังนั้น $$Accuracy = \frac{(5+0)}{(5+0+0+0)} = 1$$

$$Precision = \frac{5}{(5+0)} = 1$$

และ $$Recall = \frac{5}{(5+0)} = 1$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จึงได้ผลลัพธ์ค่าความถูกต้อง (Accuracy) เท่ากับ 1 ค่าความแม่นยำ (Precision) เท่ากับ 1 และค่าความระลึก (Recall) เท่ากับ 1 หลังจากทราบค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) สามารถที่จะหาค่าวัดประสิทธิภาพ (F1-Score)

จากสูตร $F1 - Score = \frac{2 \times precision \times recall}{precision + recall}$

ดังนั้น $F1 - Score = \frac{2 \times 1 \times 1}{1+1} = 1$

จึงได้ผลลัพธ์ค่าวัดประสิทธิภาพ (F1-Score) เท่ากับ 1

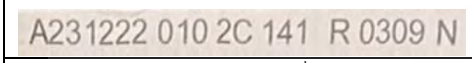
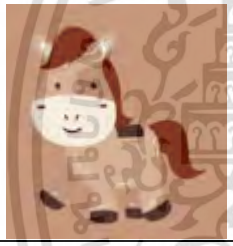
รูปที่ 4.8 รูปภาพตัวอย่างของการตรวจจับโลโก้สินค้ารูปแบบที่ 2

4. จากผลลัพธ์ของตารางที่ 4.2 ตารางที่ 4.3 และตารางที่ 4.4 แสดงรายละเอียดของการตรวจจับวัตถุทั้ง 3 ประเภทและสามารถที่จะคำนวณหาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าวัดประสิทธิภาพ (F1-Score) ได้ จากค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ที่ได้สามารถคำนวณหาค่าความแม่นยำเฉลี่ย (Average Precision) ของแต่ละวัตถุ

จากสูตร $AP = \sum(r_{n+1} - r_n)p_{interp}(r_{n+1})$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ตารางแสดงผลลัพธ์ของการทดสอบแบบจำลองการตรวจจับวัตถุทั้ง 3 ประเภทโดยใช้ชุดข้อมูลสำหรับการทดสอบ

ประเภทของวัตถุ (Class)	ค่าความถูกต้อง (Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่าวัดประสิทธิภาพ (F1-Score)	ค่าความแม่นยำเฉลี่ย (Average Precision)
หมายเลขสินค้า 	0.989	1	0.989	0.994	1
โลโก้รูปแบบที่ 1	0.989	1	0.989	0.994	1
โลโก้รูปแบบที่ 2 	1	1	1	1	1
รวม	0.992	1	0.992	0.996	1

จากตารางที่ 4.5 จากผลทดสอบการตรวจจับวัตถุทั้ง 3 ประเภท ผลลัพธ์ของการตรวจจับที่ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (IoU) เท่ากับ 0.5 ได้ค่าความถูกต้อง (Accuracy) เท่ากับ 0.992 หรือคิดเป็น 99.2 เปอร์เซนต์ ค่าความแม่นยำ (Precision) เท่ากับ 1 หรือคิดเป็น 100 เปอร์เซนต์ ค่าความระลึก (Recall) เท่ากับ 0.992 หรือคิดเป็น 99.2 เปอร์เซนต์ ค่าวัดประสิทธิภาพ (F1-Score) เท่ากับ 0.996 หรือคิดเป็น 99.6 เปอร์เซนต์ และค่าความแม่นยำเฉลี่ย (Average Precision) เท่ากับ 1 หรือคิดเป็น 100 เปอร์เซนต์

5. ระยะเวลาที่ใช้ในการประมวลผลของแบบจำลองการตรวจจับวัตถุ ผลจากของการทดสอบการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า จำนวน 95 ภาพ ได้ผลลัพธ์ระยะเวลาที่ใช้ในการประมวลผลของการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ดังตารางที่ 4.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

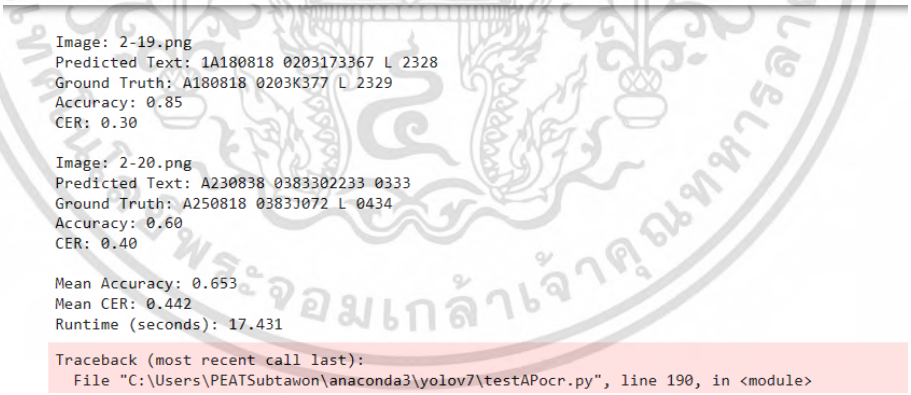
ตารางที่ 4.6 รายละเอียดระยะเวลาที่ใช้ในการประมวลผลการตรวจจับ

ระยะเวลาที่ใช้ในการประมวลผล (วินาที)	
ระยะเวลาต่อข้อมูลทั้งหมด	51.342
ระยะเวลาต่อรูปภาพ	0.540

จากตารางที่ 4.6 ระยะเวลาที่ใช้ในการประมวลผลการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าจำนวน 95 ภาพคือ 51.342 วินาที คิดเป็นใช้ระยะเวลาในการประมวลผลการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าเท่ากับ 0.540 วินาทีต่อรูปภาพ

4.2 การทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (OCR)

ในการทดสอบแบบจำลองการรู้จำอักขระด้วยแสง (OCR) จากเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ได้นำรูปภาพจำนวน 40 ภาพสำหรับการทดสอบหรือเป็น 20 เปอร์เซ็นต์จากทั้งหมด 200 ภาพ จากชุดแบบตัวอักษร (Font) 2 รูปแบบตัวอักษรแบ่งเป็นรูปแบบละ 20 ภาพ โดยพิจารณาผลลัพธ์ที่ได้จากแบบจำลองการรู้จำอักขระด้วยแสง (OCR) เปรียบเทียบกับข้อความที่ถูกต้อง (Ground Truth) เพื่อทดสอบค่าความถูกต้อง (Accuracy) ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) และระยะเวลาที่ใช้ในการประมวลผล ทดสอบโดยใช้โปรแกรม jupyter notebook ผลลัพธ์จากโปรแกรมดังรูปที่ 4.9



รูปที่ 4.9 ผลลัพธ์จากโปรแกรมทดสอบการรู้จำอักขระด้วยแสง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ตารางแสดงค่าความถูกต้อง (Accuracy) และค่าความผิดพลาดระดับตัวอักษร (CER) ของการรู้จำตัวอักษรกับตัวอักษรที่ถูกต้อง (Ground Truth)

รูปแบบของตัวอักษร (Font)	จำนวนรูปภาพทั้งหมด (รูป)	จำนวนตัวอักษรทั้งหมด (ตัวอักษร)	จำนวนตัวอักษรที่ถูกต้อง (ตัวอักษร)	จำนวนตัวอักษรที่เพิ่มขึ้น (ตัวอักษร)	จำนวนตัวอักษรที่เปลี่ยนไป (ตัวอักษร)	จำนวนตัวอักษรที่หายไป (ตัวอักษร)	ค่าความผิดพลาดระดับตัวอักษร (CER)	ค่าความถูกต้อง (Accuracy)
ตัวอักษรรูปแบบที่ 1	20	420	248	38	143	29	0.500	0.590
ตัวอักษรรูปแบบที่ 2	20	402	289	41	96	17	0.383	0.718
รวม	40	822	537	79	239	46	0.442	0.653

จากตารางที่ 4.7 จากผลทดสอบการรู้จำอักขระด้วยแสงตัวอักษร 2 รูปแบบ โดยใช้ค่าความถูกต้อง (Accuracy) และค่าความผิดพลาดระดับตัวอักษร (CER) ในการทดสอบความแม่นยำของการรู้จำอักขระด้วยแสง (Optical Character Recognition) โดยการทดสอบความผิดพลาดระหว่างข้อความที่แบบจำลองสร้างขึ้น และข้อความที่ถูกต้อง

จากการทดสอบตัวอักษรรูปแบบที่ 1 มีจำนวนตัวอักษรทั้งหมด 420 ตัว ได้ผลลัพธ์จากการทดสอบการรู้จำอักขระด้วยแสง มีจำนวนตัวอักษรที่ถูกต้อง 248 ตัว มีจำนวนตัวอักษรที่เพิ่มขึ้นจากข้อความผลเฉลย 38 ตัว มีจำนวนตัวอักษรที่เปลี่ยนไปจากข้อความผลเฉลย 143 ตัว และมีจำนวนตัวอักษรที่หายไปจากข้อความผลเฉลย 29 ตัว

จากสูตรการหาค่า $CER = \frac{I+S+D}{N}$

โดยที่ ค่า I คือจำนวนตัวอักษรที่เพิ่มขึ้น

ค่า S คือจำนวนตัวอักษรที่เปลี่ยนไป

ค่า D คือจำนวนตัวอักษรที่หายไป

และ ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $CER = \frac{38+143+29}{420} = 0.5$

จากสูตรการหาค่า $Accuracy = \frac{T}{N}$

โดยที่ ค่า T คือจำนวนตัวอักษรที่ทายถูกต้อง

ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $Accuracy = \frac{248}{420} = 0.59$

ผลลัพธ์ได้ค่าความผิดพลาดระดับตัวอักษร (CER) เท่ากับ 0.5 หรือคิดเป็น 50 เปอร์เซนต์ และค่าความถูกต้อง (Accuracy) เท่ากับ 0.59 หรือคิดเป็น 59 เปอร์เซนต์

จากการทดสอบตัวอักษรรูปแบบที่ 2 มีจำนวนตัวอักษรทั้งหมด 402 ตัว ได้ผลลัพธ์จากการทดสอบการรู้จำอักขระด้วยแสง มีจำนวนตัวอักษรที่ถูกต้อง 289 ตัว มีจำนวนตัวอักษรที่เพิ่มขึ้นจากข้อความผลเฉลย 41 ตัว มีจำนวนตัวอักษรที่เปลี่ยนไปจากข้อความผลเฉลย 96 ตัว และมีจำนวนตัวอักษรที่หายไปจากข้อความผลเฉลย 46 ตัว

จากสูตรการหาค่า $CER = \frac{I+S+D}{N}$

โดยที่ ค่า I คือจำนวนตัวอักษรที่เพิ่มขึ้น

ค่า S คือจำนวนตัวอักษรที่เปลี่ยนไป

ค่า D คือจำนวนตัวอักษรที่หายไป

และ ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $CER = \frac{41+96+17}{402} = 0.383$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสูตรการหาค่า $Accuracy = \frac{T}{N}$

โดยที่ ค่า T คือจำนวนตัวอักษรที่ทายถูกต้อง

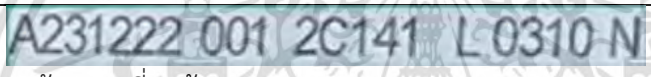
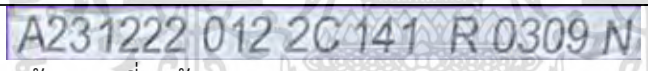
ค่า N คือจำนวนตัวอักษรทั้งหมดที่ได้มาจากเฉลยของชุดข้อมูล

ดังนั้น $$Accuracy = \frac{289}{402} = 0.718$$

ผลลัพธ์ได้ค่าความผิดพลาดระดับตัวอักษร (CER) เท่ากับ 0.383 หรือคิดเป็น 38.3 เปอร์เซนต์ และค่าความถูกต้อง (Accuracy) เท่ากับ 0.718 หรือคิดเป็น 71.8 เปอร์เซนต์

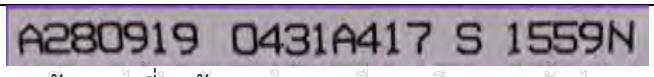
จากการทดสอบแบบจำลองการรู้จำอักขระด้วยแสงโดยใช้ชุดข้อมูลสำหรับการทดสอบจำนวน 40 ภาพ แบบจำลองการรู้จำอักขระด้วยแสงโดยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ได้ค่าความผิดพลาดระดับตัวอักษร (CER) ของแบบจำลองเท่ากับ 0.442 หรือคิดเป็น 44.2 เปอร์เซนต์ และได้ค่าความถูกต้อง (Accuracy) ในการแปลงตัวอักษรเท่ากับ 0.653 หรือคิดเป็น 65.3 เปอร์เซนต์

ตารางที่ 4.8 ตัวอย่างผลลัพธ์การทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 1

ตัวอย่างที่	รูปภาพ และข้อความที่ถูกต้อง	ผลลัพธ์การรู้จำอักขระด้วยแสง
1	 ข้อความที่ถูกต้อง: A231222 001 2C141 L 0310 N	1A231222 008 20811 L 0810 0
2	 ข้อความที่ถูกต้อง: A190319 0022C 118 R 0227 N	A190319 0022C 118 R 0227 N
3	 ข้อความที่ถูกต้อง: A231222 012 2C 141 R 0309 N	18231222 032 20 330 8 0838 0

จากตารางที่ 4.8 แสดงตัวอย่างของการทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 1 ที่มีรูปของหมายเลขสินค้ารูปแบบที่ 1 กับข้อความที่ถูกต้อง และผลลัพธ์ของการรู้จำอักขระด้วยแสงของเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

ตารางที่ 4.9 ตัวอย่างผลลัพธ์การทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 2

ตัวอย่างที่	รูปภาพ และข้อความที่ถูกต้อง	ผลลัพธ์การรู้จำอักขระด้วยแสง
1	 ข้อความที่ถูกต้อง: A180818 0623K353 R 2215	A180818 082307333 8 2218
2	 ข้อความที่ถูกต้อง: A180818 0713K360 L 2233	1A180818 0671317360 L 2233
3	 ข้อความที่ถูกต้อง: A280919 0431A417 S 1559N	A280919 0431A417 S 1559N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.9 แสดงตัวอย่างของการทดสอบการรู้จำอักขระด้วยแสงของตัวอักษรรูปแบบที่ 2 ที่มีรูปของหมายเลขสินค้ารูปแบบที่ 2 กับข้อความที่ถูกต้อง และผลลัพธ์ของการรู้จำอักขระด้วยแสงของเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR)

ผลลัพธ์ของระยะเวลาที่ใช้ในการประมวลผลของแบบจำลองการรู้จำอักขระด้วยแสง จากการทดสอบการรู้จำหมายเลขของสินค้า จำนวน 40 ภาพ ได้ผลลัพธ์ระยะเวลาที่ใช้ในการประมวลผลของการรู้จำหมายเลขของสินค้า ดังตารางที่ 4.10

ตารางที่ 4.10 รายละเอียดระยะเวลาที่ใช้ในการประมวลผลการรู้จำหมายเลขสินค้า

ระยะเวลาที่ใช้ในการประมวลผล (วินาที)	
ระยะเวลาต่อข้อมูลทั้งหมด	17.431
ระยะเวลาต่อรูปภาพ	0.435

จากตารางที่ 4.10 ระยะเวลาที่ใช้ในการประมวลผลการรู้จำอักขระด้วยแสงของรูปภาพของหมายเลขสินค้าจำนวน 40 ภาพคือ 17.431 วินาที คิดเป็นใช้ระยะเวลาในการประมวลผลการรู้จำอักขระด้วยแสงของรูปภาพของหมายเลขสินค้าเท่ากับ 0.435 วินาทีต่อรูปภาพ

4.3 การอภิปรายผล

จากการทดสอบแบบจำลองการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าด้วยอัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) โดยใช้ข้อมูลสำหรับการทดสอบเป็นจำนวน 95 ภาพ ที่ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (IoU) ต่อตำแหน่งของวัตถุนั้นที่ค่าเท่ากับ 0.5 ได้ค่าความแม่นยำเฉลี่ย (mAP) เท่ากับ 100 เปอร์เซนต์ เมื่อเปรียบเทียบกับงานวิจัยของ Luxolo, Dane and Alden (2023) ที่ทำการศึกษาการตรวจจับปลาฉลามหัวบาตร ซึ่งมีการใช้อัลกอริทึมโยโล่เวอร์ชันที่ 7 (YOLOv7) เหมือนกัน และประเภทของวัตถุที่ตรวจจับมีลักษณะที่เป็นสัตว์ คล้ายกับโลโก้สินค้าในวิจัยนี้ที่มีรูปร่างเป็นสัตว์เหมือนกัน ผลลัพธ์ของงานวิจัยของ Luxolo, Dane and Alden (2023) ได้ค่าความแม่นยำเฉลี่ย (mAP) เท่ากับ 100 เปอร์เซนต์ที่ค่าอัตราส่วนระหว่างพื้นที่ซ้อนทับกันและพื้นที่รวม (IoU) เท่ากับ 0.5 เมื่อทำการเปรียบเทียบกับผลลัพธ์ของงานวิจัยนี้ผลลัพธ์ของงานวิจัยของ Luxolo, Dane and Alden (2023) มีค่าความแม่นยำเฉลี่ย (mAP) เท่ากันที่ 100 เปอร์เซนต์

จากการทดสอบแบบจำลองการรู้จำหมายเลขสินค้า ด้วยเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) โดยใช้ข้อมูลสำหรับการทดสอบเป็นจำนวน 40 ภาพ ที่มีทั้งตัวอักษรภาษาอังกฤษ และตัวเลขอารบิก ได้ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) ที่ 44.2 เปอร์เซนต์ เมื่อเทียบกับงานวิจัยของ ธนดล (2564) ที่ได้ศึกษาเกี่ยวกับการรู้จำคำบรรยายจากวีดิทัศน์ ซึ่งได้ใช้สถาปัตยกรรมที่คล้ายคลึงกับของเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) โดยในการฝึกสอนแบบจำลองข้อมูลที่นำมาใช้จากการรวมรวม 24 วีดิทัศน์จาก Youtube และ Facebook ที่มีคำ

เอกสารนี้ ... ไม่ว่ากรณี ...

บรรยายภาษาไทย อังกฤษ ตัวเลขไทย และตัวเลขอารบิก ผลลัพธ์จากการทดสอบได้ค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) ที่ 11.06 เปอร์เซนต์ เมื่อทำการเปรียบเทียบ งานวิจัยของ ธนดล (2564) มีค่าความผิดพลาดระดับตัวอักษร (Character Error Rate: CER) ที่น้อยกว่าหรือมีความแม่นยำที่ดีกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยเรื่องการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความเพื่อการตรวจสอบย้อนกลับ มีวัตถุประสงค์เพื่อนำเครื่องมือเทคโนโลยีในปัจจุบันเข้ามาทดลองเพื่อที่จะทดสอบว่าสามารถที่จะนำมาปรับปรุง และนำมาแทนที่เทคโนยีเดิมที่ใช้ เพื่อเพิ่มประสิทธิภาพในเรื่องของค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผลได้ดีขึ้นหรือไม่ ในการตรวจจับ (Object Detection) หมายเลขของสินค้า และโลโก้ของสินค้า และการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ในการอ่าน และแปลงข้อความของหมายเลขสินค้า ซึ่งสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

ปัจจุบันบริษัทกรณีศึกษาที่ผู้วิจัยได้รับอนุเคราะห์ในการเข้าร่วมโครงการสหกิจศึกษา ได้มีการใช้เครื่องมือการตรวจจับวัตถุ (Object Detection) ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ในการอ่านและแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ เพื่อการตรวจสอบย้อนกลับ เพื่อตรวจสอบ และติดตามข้อมูลของสินค้า เพื่อระบุแหล่งที่มาของสินค้าที่พบปัญหามีการผลิตขึ้นเมื่อใด จากสายการผลิตไหน และรับวัตถุดิบมาจากแหล่งไหน ข้อมูลที่ใช้ในการตรวจสอบย้อนกลับได้แก่ หมายเลขของสินค้า และโลโก้ของสินค้า จึงได้ใช้เครื่องมือการตรวจจับวัตถุเพื่อแยกแยะโลโก้ของสินค้า และตรวจจับหมายเลขของสินค้าเพื่อนำหมายเลขสินค้าที่ตรวจจับได้ นำไปทำการแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความด้วยเครื่องมือการรู้จำอักขระด้วยแสง เพื่อใช้ในการเก็บข้อมูลของสินค้า โดยเครื่องมือที่ใช้ได้ทำการพัฒนาด้วยเทคโนโลยีเทนเซอร์โฟล (Tensorflow) และมีประสิทธิภาพของเครื่องมือที่ค่าความถูกต้อง (Accuracy) 90 เปอร์เซนต์ และระยะเวลาในการประมวลผลที่ใช้ระยะเวลามากกว่า 5 วินาทีต่อภาพ ในงานวิจัยจึงมีวัตถุเพื่อทำการพัฒนาเครื่องมือการตรวจจับวัตถุ (Object Detection) ด้วยเทคโนโลยีโยโล่เวอร์ชันที่ 7 (YOLOv7) ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และเครื่องมือการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ด้วยเทคโนโลยีเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ในการรู้จำอักขระด้วยแสงของหมายเลขของสินค้าในการอ่าน และแปลงข้อความของหมายเลขสินค้าให้มีประสิทธิภาพค่าความถูกต้อง (Accuracy) และระยะเวลาในการประมวลผลให้ดียิ่งขึ้น โดยขอสรุปผลแยกตามวัตถุประสงค์ดังต่อไปนี้

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วัตถุประสงค์ข้อที่ 1. เพื่อสร้างแบบจำลองการตรวจจับวัตถุ (Object Detection) เพื่อตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าในรูปภาพ โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองที่มากกว่า 90 เปอร์เซนต์

ปัจจุบันบริษัทกรณีศึกษาได้พัฒนาเครื่องมือการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าโดยใช้เทคโนโลยีเทนเซอร์โฟล (Tensorflow) โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองอยู่ที่ 90 เปอร์เซนต์ งานวิจัยนี้ได้ผลลัพธ์จากการทดสอบประสิทธิภาพของแบบจำลองโยโล่เวอร์ชันที่ 7 (YOLOv7) ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า ได้ค่าความถูกต้อง (Accuracy) เท่ากับ 99.2 เปอร์เซนต์ ซึ่งมีค่าความถูกต้อง (Accuracy) ที่ดีขึ้น งานวิจัยนี้เป็นการทดสอบเบื้องต้นว่าในปัจจุบันมีเทคโนโลยีที่มีประสิทธิภาพค่าความถูกต้อง (Accuracy) ที่มากกว่าเพื่อนำไปพัฒนาต่อยอดเครื่องมือที่ใช้ในปัจจุบันในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าได้

วัตถุประสงค์ข้อที่ 2. เพื่อสร้างแบบจำลองการตรวจจับวัตถุ (Object Detection) เพื่อตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าในรูปภาพ โดยมีระยะเวลาในการประมวลผลที่น้อยกว่า 5 วินาที

ปัจจุบันบริษัทกรณีศึกษาได้พัฒนาเครื่องมือการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าโดยใช้เทคโนโลยีเทนเซอร์โฟล (Tensorflow) โดยมีระยะเวลาในการประมวลผลของแบบจำลองที่มากกว่าหรือเท่ากับ 5 วินาที งานวิจัยนี้ได้ผลลัพธ์จากการทดสอบประสิทธิภาพของแบบจำลองโยโล่เวอร์ชันที่ 7 (YOLOv7) ในการประมวลผลใช้ระยะเวลาเท่ากับ 0.540 วินาทีต่อรูปภาพ ซึ่งใช้ระยะเวลาในการประมวลผลที่ลดลง งานวิจัยนี้เป็นการทดสอบเบื้องต้นว่าในปัจจุบันมีเทคโนโลยีที่มีประสิทธิภาพด้านระยะเวลาในการประมวลผลที่น้อยกว่า เพื่อนำไปพัฒนาต่อยอดเครื่องมือที่ใช้ในปัจจุบันในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าได้

วัตถุประสงค์ข้อที่ 3. เพื่อสร้างแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เพื่อแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบของข้อความ โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองที่มากกว่า 90 เปอร์เซนต์

ปัจจุบันบริษัทกรณีศึกษาได้พัฒนาเครื่องมือการรู้จำอักขระด้วยแสงโดยใช้เทคโนโลยีเทนเซอร์โฟล (Tensorflow) โดยมีค่าความถูกต้อง (Accuracy) ของแบบจำลองอยู่ที่ 90 เปอร์เซนต์ งานวิจัยนี้ได้ผลลัพธ์จากการทดสอบประสิทธิภาพแบบจำลองเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ได้ค่าความถูกต้อง (Accuracy) เท่ากับ 65.3 เปอร์เซนต์ ซึ่งมีค่าความถูกต้อง (Accuracy) ที่น้อยกว่าเครื่องมือที่พัฒนาโดยใช้เทคโนโลยีเทนเซอร์โฟล (Tensorflow) จึงไม่เป็นไปตามวัตถุประสงค์ที่ตั้งไว้

วัตถุประสงค์ข้อที่ 4. เพื่อสร้างแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) เพื่อแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบของข้อความ โดยมีระยะเวลาในการประมวลผลที่น้อยกว่า 5 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัจจุบันบริษัทกรณีศึกษาได้พัฒนาเครื่องมือการรู้จำอักขระด้วยแสงโดยใช้เทคโนโลยีเทนเซอร์โฟล (Tensorflow) โดยมีระยะเวลาในการประมวลผลของแบบจำลองที่มากกว่าหรือเท่ากับ 5 วินาที งานวิจัยนี้ได้ผลลัพธ์จากการทดสอบประสิทธิภาพแบบจำลองเทสเซอร์แรคต์ โอซีอาร์ (Tesseract OCR) ในการประมวลผลใช้ระยะเวลาเท่ากับ 0.435 วินาทีต่อรูปภาพ ซึ่งใช้ระยะเวลาในการประมวลผลที่ลดลง งานวิจัยนี้เป็นการทดสอบเบื้องต้นว่าในปัจจุบันมีเทคโนโลยีที่มีประสิทธิภาพด้านระยะเวลาในการประมวลผลที่น้อยกว่า เพื่อนำไปพัฒนาต่อยอดเครื่องมือที่ใช้ในปัจจุบันในการแปลงหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบของข้อความได้

หลังจากทำการพัฒนาแบบจำลองการตรวจจับวัตถุ (Object Detection) ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และแบบจำลองการรู้จำอักขระด้วยแสง (Optical Character Recognition: OCR) ในการอ่าน และแปลงข้อความของหมายเลขสินค้าจากรูปภาพให้อยู่ในรูปแบบข้อความ ได้ทำการรวมแบบจำลอง และได้ทำการทดลองการใช้งาน เพื่อเป็นประโยชน์ และแนวทางในการนำไปใช้งานจริงในอนาคต

5.2 ข้อจำกัด

1. ข้อจำกัดในเรื่องของชุดข้อมูลที่ใช้ในงานวิจัยของการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าเป็นรูปภาพมุมตรงเพียงอย่างเดียว ทำให้ในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าในมุมอื่น ๆ อาจพบปัญหาในเรื่องของความแม่นยำ

2. ข้อจำกัดในเรื่องของชุดข้อมูลในการฝึกสอนแบบจำลองการรู้จำอักขระด้วยแสง ในเรื่องคุณภาพของรูปภาพหมายเลขสินค้าที่มีความเบลอ แสงที่ไม่เพียงพอ ทำให้รูปภาพหมายเลขสินค้ามีความไม่ชัดเจน

3. ข้อจำกัดในเรื่องของผลลัพธ์ค่าความถูกต้อง (Accuracy) ในการรู้จำตัวอักษร เนื่องจากการทดสอบเป็นผลลัพธ์ค่าความถูกต้อง (Accuracy) ที่ได้จากทุกตัวอักษร จึงไม่ทราบค่าความถูกต้อง (Accuracy) ที่แท้จริงของแต่ละตัวอักษร

4. ข้อจำกัดในเรื่องการนำแบบจำลองไปใช้งานจริง ในงานวิจัยนี้แบบจำลองการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง ไม่สามารถนำไปใช้งานจริงได้ เนื่องจากงานวิจัยนี้เป็นการทดสอบเพื่อพิสูจน์แนวคิดว่าเทคโนโลยีในปัจจุบันสามารถที่จะนำมาใช้ในการปรับปรุง หรือแทนที่เครื่องมือที่ทางบริษัทกรณีศึกษาใช้อยู่ในปัจจุบันได้หรือไม่ งานวิจัยนี้เป็นส่วนนึงที่นำไปพัฒนาต่อยอด

5.3 ข้อเสนอแนะ

1. เพิ่มชุดข้อมูลของภาพสินค้าในมุมที่หลากหลายเพื่อเพิ่มความแม่นยำในการตรวจจับในมุมอื่น ๆ ที่ไม่ใช่มุมตรง โดยการเพิ่มชุดข้อมูลของภาพสินค้าในมุมที่หลากหลายในการฝึกสอนแบบจำลอง เพื่อให้แบบจำลองสามารถที่จะตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้าได้อย่างแม่นยำในสถานการณ์ และมุมมองที่แตกต่างมากยิ่งขึ้น

เอกสารนี้ [illegible] ไม่ว่ากรณีใดๆ [illegible] และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เพิ่มความชัดเจนของรูปภาพหมายเลขสินค้าโดยการปรับปรุงคุณภาพรูปภาพหมายเลขสินค้าโดยใช้เทคโนโลยีอื่น ๆ เข้ามาช่วยในการปรับปรุงคุณภาพรูปภาพ ทำให้รูปภาพมีความคมชัดมากขึ้นและลดการเกิดความเบลอหรือการสะท้อนที่ส่งผลต่อความแม่นยำในการตรวจจับ

3. ทำการปรับปรุงการทดสอบค่าความถูกต้อง (Accuracy) ในการรู้จำตัวอักษรของหมายเลขสินค้า โดยทำการทดสอบเฉพาะเจาะจงของแต่ละตัวอักษร เพื่อที่จะทราบค่าความถูกต้อง (Accuracy) ที่แท้จริงของแต่ละตัวอักษร ทำให้สามารถเพิ่มชุดข้อมูลของตัวอักษรที่มีค่าความถูกต้อง (Accuracy) ต่ำ เพื่อทำการฝึกสอนเพิ่มเติม เพื่อให้แบบจำลองมีประสิทธิภาพในการรู้จำได้แม่นยำมากยิ่งขึ้น

4. พิจารณาการนำเทคโนโลยีที่ได้ทำการวิจัยไปพัฒนาต่อยอด เพื่อปรับปรุงหรือแทนที่เครื่องมือที่ทางบริษัทกรณีศึกษาใช้อยู่ในปัจจุบัน ทำให้เครื่องมือในการตรวจจับหมายเลขของสินค้า และโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ มีประสิทธิภาพการทำงานในเรื่องของความแม่นยำ (Accuracy) และระยะเวลาในการประมวลผลที่ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

ธนาคารกรุงศรีอยุธยา. **แนวโน้มธุรกิจและอุตสาหกรรมไทยปี 2565-2567**. [Online]. Available: https://www.krungsri.com/th/research/industry/summary-outlook/outlook-2022-2024. เข้าถึงเมื่อวันที่ 1 มิ.ย. 66

มาดีนา น้อยทับทิม และกนกวรรณ สุขขจรวงษ์. 2556. **การประยุกต์ใช้ระบบการตรวจสอบย้อนกลับ (Traceability) ในอุตสาหกรรมอาหาร**. วารสารมหาวิทยาลัยนราธิวาสราชนครินทร์, ปีที่ 5, ฉบับที่ 4, หน้า 206-215.

ปฏิญญา ตันฑวิวัฒน์ และคณะ. 2564. **การวิเคราะห์ภาพคนและสัมภาระสำหรับการตรวจจับวัตถุที่ปราศจากเจ้าของ**. วารสารวิศวกรรมศาสตร์และเทคโนโลยีดิจิทัล, ปีที่ 9, ฉบับที่ 2, หน้า 49-60.

พิชยุทธ บุญตน และยุทธนา สีขวา. 2564. **ระบบจำแนกถนนชำรุด**. ปริญญานิพนธ์ปริญญาบัณฑิต. มหาวิทยาลัยมหาสารคาม.

จักรพันธ์ วาศพุฒิสิทธิ์. 2561. **การเพิ่มประสิทธิภาพการจัดทำดัชนีบนกล่องเอกสารด้วยเทคนิคการรู้จำอักขระภาพลายมือ**. สารนิพนธ์ปริญญามหาบัณฑิต. สถาบันเทคโนโลยีไทย-ญี่ปุ่น.

วิกิพีเดีย. 2565. **แบบจำลอง หรือ โมเดล**. [online]. Available: https://th.wikipedia.org/wiki/. เข้าถึงเมื่อวันที่ 1 มิ.ย. 66.

จักรภัทร แก้วทอง และไตรปิฎก อินทสุวรรณ. 2560. **โปรแกรมตรวจจับวัตถุและข้อความบนป้ายโฆษณา.** วิทยานิพนธ์ปริญญาบัฒฑิต. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

ณรงค์เดช กีรติพรานนท์. 2563. **การรับรู้ป้ายสาธารณะและการแปลภาษาไทย-จีน บนระบบแอนดรอยด์**. วารสารบัณฑิตศึกษา. มหาวิทยาลัยธุรกิจบัณฑิตย์, ปีที่ 8, ฉบับที่ 3, หน้า 754-766.

พจนานุกรม ฉบับราชบัณฑิตยสถาน. 2554. **อุตสาหกรรม**. [online]. Available: https://dictionary.orst.go.th/. เข้าถึงเมื่อวันที่ 5 มิ.ย. 66.

ศิวกร บรรลือทรัพย์ และวราพร จิระพันธุ์ทอง. 2565. **อัลกอริธึมการเรียนรู้ของเครื่องสำหรับการทำนายการคัดแยกผู้ป่วย COVID-19**. วารสารมหาวิทยาลัยธุรกิจบัณฑิตย์, ปีที่ 13, ฉบับที่ 2, หน้า 47-60.

มาณวิกา กิตติพร. 2562. **ระบบแนะนำทางการศึกษา และเทคนิคการเรียนรู้ของเครื่องจักร**. วารสารหมาวิทยาลัยมหาสารคาม, ปีที่ 4, หน้า 92-109.

ธนภัทร์ คุ้มสุภา. 2559. **การจำแนกประเภทข้อความในภาษาไทยโดยใช้นิวรอลเน็ตเวิร์กคอนโวลูชันระดับตัวอักษร**. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยจุฬาลงกรณ์.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มนัส บุญวงศ์ และมาเรียม นะมิ. 2558. **กระบวนการตรวจสอบย้อนกลับในโซ่อุปทานของอุตสาหกรรมเนื้อสัตว์แปรรูปฮาลาลในประเทศไทยเพื่อรองรับการเปิดประชาคมเศรษฐกิจอาเซียน**. มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร.

ธนดล สิงขรอาสน์. 2564. **การเรียนรู้เชิงลึกสำหรับการตรวจจับและรู้จำคำบรรยายในวีดิทัศน์**. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยมหาสารคาม.

อุบลรัตน์ ศิริสุขโภคา และจรัญ แสนราช. 2562. **การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลปัจจัยสนับสนุนการเรียนรู้ของผู้เรียนระดับอุดมศึกษาในสภาพแวดล้อมการเรียนรู้เสมือนจริง**. วารสารพระจอมเกล้าพระนครเหนือ, ปีที่ 10, ฉบับที่ 3, หน้า 131-139.

พจนานุกรม ฉบับราชบัณฑิตยสถาน. 2555. **โลโก้**. [online]. Available: สำนักงานราชบัณฑิตยสภาโลโก้ (๑ ธันวาคม ๒๕๕๕) - สำนักงานราชบัณฑิตยสภา (orst.go.th). เข้าถึงเมื่อวันที่ 10 มิ.ย. 66.

เรือนขวัญ พลฤทธิ์. 2563. **การพัฒนาแอปพลิเคชันเพื่อการศึกษาร่วมกับกิจกรรมการเรียนรู้แบบใช้ปัญหาเป็นฐาน วิชาวิทยาการคำนวณ เรื่องการเขียนโปรแกรมภาษา Python ที่มีต่อความสามารถในการเขียน โปรแกรมและผลงานการเขียนโปรแกรมของนักเรียนชั้นมัธยมศึกษาปีที่ 1**. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยศิลปากร.

พุฒิพงศ์ จันทร์แจ่ม. 2563. **การปรับปรุงประสิทธิภาพของกระบวนการบริการลูกค้าโดยใช้การวิเคราะห์จากกล้องวงจรปิด**. วิทยานิพนธ์ปริญญาบัณฑิต. สถาบันบัณฑิตพัฒนบริหารศาสตร์.

สุทธิรักต์ ตาละลักษณ์. 2557. **ระบบการรู้จำใบหน้าบน Smart Phone กรณีศึกษา การตรวจสอบการออกทำงานนอกพื้นที่ของพนักงานขาย บริษัท วอลล์ เทคโนโลยี จำกัด.** วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยศิลปากร

ธวัชชัย ศรีสุเทพ. 2549. **ฟอนต์ไหนดี?**. กรุงเทพฯ : มาร์คมายเว็บ, 2549.

ศรัญญา นาทองห่อ. 2563. **การสรุปใจความสำคัญของข้อความแบบสกัดสำหรับข่าวท่องเที่ยวภาษาไทย**. วิทยานิพนธ์ปริญญามหาบัณฑิต. มหาวิทยาลัยจุฬาลงกรณ์

Jake Vanderplas. 2016. **Python Data Science Handbook**. Sebastopol: O'Reilly Media, Inc, 2016.

Jeffrey A Clark. 2023. **Pillow**. [online]. Available: https://pillow.readthedocs.io.

Kirill Simonov. 2021. **PyYAML**. [online]. Available: https://pypi.org/project/PyYAML.

Kenneth Reitz. 2023. **Requests**. [online]. Available: https://pypi.org/project/requests.

Scipy developer. 2023. **Numpy and Scipy Documentation**. [online]. Available: https://docs.scipy.org/doc.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tqdm developers. 2015. **tqdm**. [online]. Available: https://pypi.org/project/tqdm/2.2.3.

Wang, C. Y., Bochkovskiy, A and Mark Liao, H. Y. 2022. YOLOv7: **Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors**. Taiwan: Sinica.

Chirag , Dharmendra. 2012. **Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study. Journal of Computer Applications**. 55 (10): 50-56.

Fang, W., Wang, L., Ren, P. 2020. Tinier-YOLO:**A Real-Time Object Detection Method for Constrained Environments**. IEEE Access. 8: 1935-1944.

Luxolo, K., Dane, B., and Alden, B. 2023. **Exploring The Incremental Improvements of YOLOv7 on Bull Sharks in Mozambique**. South Africa: Rhodes.

Amitdo. 2016. **Tesseract Open Source OCR Engine (main repository)**. [online] Available: https://github.com/tesseract-ocr/tesseract.

WongKinYiu. 2022. **Yolov7**. [online]. Available: https://github.com/WongKinYiu/yolov7.

MAKESENSE. [online]. Available: https://www.makesense.ai/.

Tesseract. [online]. Available: https://github.com/UB-Mannheim/tesseract/wiki.

Zdenop. 2023. **Qt-box-editor**. [online]. Available: GitHub - zdenop/qt-box-editor: QT4 editor of tesseract-ocr box files.

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. 2018. **You Only Look Once: Unified, Real-Time Object Detection**. University of Washington.

Glenn Jocher, Ayush Chaurasia, Sergiu Waxmann, Laughing. 2023. **Ultralytics YOLOv8 Docs**. [online]. Available: Home - Ultralytics YOLOv8 Docs.

The Matplotlib development team. 2023. **Maplotlib**. [online]. Available: https://matplotlib.org.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ภาคผนวก ก ชุดคำสั่งภาษาโปรแกรมไพทอน (Python) ที่ใช้ในการตรวจจับวัตถุและการรู้จำอักขระด้วยแสง

ภาคผนวก ก.1 ชุดคำสั่งที่ใช้ในการตรวจจับและการรู้จำอักขระด้วยแสงของสินค้า

```
import argparse

import time

from pathlib import Path

import pytesseract

from PIL import Image

import cv2

import torch

import torch.backends.cudnn as cudnn

from numpy import random

from models.experimental import attempt_load

from utils.datasets import LoadStreams, LoadImages

from utils.general import check_img_size, check_requirements, check_imshow,
non_max_suppression, apply_classifier, \ scale_coords, xyxy2xywh,
strip_optimizer, set_logging, increment_path

from utils.plots import plot_one_box

from utils.torch_utils import select_device, load_classifier, time_synchronized,
TracedModel

def detect(save_img=False):source, weights, view_img, save_txt, imgsz, trace =
opt.source, opt.weights, opt.view_img, opt.save_txt, opt.img_size, not
opt.no_trace

save_img = not opt.nosave and not source.endswith('.txt')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
webcam = source.isnumeric() or source.endswith('.txt') or
source.lower().startswith(('rtsp://', 'rtmp://', 'http://', 'https://'))

# Directories

save_dir = Path(increment_path(Path(opt.project) / opt.name,
exist_ok=opt.exist_ok))  # increment run

(save_dir / 'labels' if save_txt else save_dir).mkdir(parents=True, exist_ok=True)
# make dir

# Initialize

set_logging()

device = select_device(opt.device)

half = device.type != 'cpu'  # half precision only supported on CUDA

# Load model

model = attempt_load(weights, map_location=device)  # load FP32 model

stride = int(model.stride.max())  # model stride

imgsz = check_img_size(imgsz, s=stride)  # check img_size

if trace:

model = TracedModel(model, device, opt.img_size)

if half:

model.half()  # to FP16

# Second-stage classifier

classify = False

if classify:

modelc = load_classifier(name='resnet101', n=2)  # initialize

modelc.load_state_dict(torch.load('weights/resnet101.pt',
map_location=device)['model']).to(device).eval()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# Set Dataloader

vid_path, vid_writer = None, None

if webcam:

view_img = check_imshow()

cudnn.benchmark = True  # set True to speed up constant image size inference

dataset = LoadStreams(source, img_size=imgsz, stride=stride)

else:

dataset = LoadImages(source, img_size=imgsz, stride=stride)

# Get names and colors

names = model.module.names if hasattr(model, 'module') else model.names

colors = [[random.randint(0, 255) for _ in range(3)] for _ in names]

# Initialize OCR engine

pytesseract.pytesseract.tesseract_cmd = 'C:\\Users\\Tesseract-
OCR\\tesseract.exe'# update this with your Tesseract installation path

ocr_config = '-l train --oem 3 --psm 11'  # OCR engine configuration

# Run inference

if device.type != 'cpu':

model(torch.zeros(1, 3, imgsz,
imgsz).to(device).type_as(next(model.parameters())))  # run once

old_img_w = old_img_h = imgsz

old_img_b = 1

t0 = time.time()

for path, img, im0s, vid_cap in dataset:

img = torch.from_numpy(img).to(device)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
img = img.half() if half else img.float()  # uint8 to fp16/32

img /= 255.0  # 0 - 255 to 0.0 - 1.0

if img.ndimension() == 3:

img = img.unsqueeze(0)

# Warmup

if device.type != 'cpu' and (old_img_b != img.shape[0] or old_img_h !=
img.shape[2] or old_img_w != img.shape[3]):

old_img_b = img.shape[0]

old_img_h = img.shape[2]

old_img_w = img.shape[3]

for i in range(3):

model(img, augment=opt.augment)[0]

# Inference

t1 = time_synchronized()

with torch.no_grad():   # Calculating gradients would cause a GPU memory leak

pred = model(img, augment=opt.augment)[0]

t2 = time_synchronized()

# Apply NMS

pred = non_max_suppression(pred, opt.conf_thres, opt.iou_thres,
classes=opt.classes, agnostic=opt.agnostic_nms)

t3 = time_synchronized()

# Apply Classifier

if classify:

pred = apply_classifier(pred, modelc, img, im0s)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# Process detections

for i, det in enumerate(pred):  # detections per image

if webcam:  # batch_size >= 1

p, s, im0, frame = path[i], '%g: ' % i, im0s[i].copy(), dataset.count

else:

p, s, im0, frame = path, '', im0s, getattr(dataset, 'frame', 0)

p = Path(p)  # to Path

save_path = str(save_dir / p.name)  # img.jpg

txt_path = str(save_dir / 'labels' / p.stem) + ('' if dataset.mode == 'image' else
f'_{frame}')  # img.txt

gn = torch.tensor(im0.shape)[[1, 0, 1, 0]]  # normalization gain whwh

if len(det):

# Rescale boxes from img_size to im0 size

det[:, :4] = scale_coords(img.shape[2:], det[:, :4], im0.shape).round()

# Print results

c in det[:, -1].unique():

n = (det[:, -1] == c).sum()  # detections per class

s += f"{n} {names[int(c)]}{'s' * (n > 1)}, "  # add to string

# Write results

for *xyxy, conf, cls in reversed(det):

if save_txt:  # Write to file

xywh = (xyxy2xywh(torch.tensor(xyxy).view(1, 4)) / gn).view(-1).tolist()  #
normalized xywh

line = (cls, *xywh, conf) if opt.save_conf else (cls, *xywh)  # label format
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
with open(txt_path + '.txt', 'a') as f:
f.write(('%g ' * len(line)).rstrip() % line + '\n')
if save_img or view_img:  # Add bbox to image
label = f'{names[int(cls)]} {conf:.2f}'
plot_one_box(xyxy, im0, label=label, color=colors[int(cls)], line_thickness=1)
# OCR on detected regions
for *xyxy, conf, cls in reversed(det):
if cls == 0:  # assuming class 0 is text
# Crop and save image region
box = [int(x) for x in xyxy]
cropped_img = im0[box[1]:box[3], box[0]:box[2]]
if save_img:save_crop_path = str(save_dir / f"{p.stem}_crop{i}.jpg")
Image.fromarray(cropped_img).save(save_crop_path)
# Run OCR
text = pytesseract.image_to_string(Image.fromarray(cropped_img),
config=ocr_config)
print(f"Detected text: {text}")
if save_txt:
save_txt_path = str(save_dir / 'labels' / f"{p.stem}_crop{i}.txt")
with open(save_txt_path, 'w') as f:
f.write(text)
# Print time (inference + NMS)
print(f'{s}Done. ({(1E3 * (t2 - t1)):.1f}ms) Inference, ({(1E3 * (t3 - t2)):.1f}ms)
NMS')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# Stream results
if view_img:
cv2.imshow(str(p), im0)
cv2.waitKey(1)  # 1 millisecond
# Save results (image with detections)
if save_img:
if dataset.mode == 'image':
cv2.imwrite(save_path, im0)
print(f" The image with the result is saved in: {save_path}")
else:  # 'video' or 'stream'
if vid_path != save_path:  # new video
vid_path = save_path
if isinstance(vid_writer, cv2.VideoWriter):
vid_writer.release()  # release previous video writer
if vid_cap:  # video
fps = vid_cap.get(cv2.CAP_PROP_FPS)
w = int(vid_cap.get(cv2.CAP_PROP_FRAME_WIDTH))
h = int(vid_cap.get(cv2.CAP_PROP_FRAME_HEIGHT))
else:  # stream
fps, w, h = 30, im0.shape[1], im0.shape[0]
save_path += '.mp4'
vid_writer = cv2.VideoWriter(save_path, cv2.VideoWriter_fourcc(*'mp4v'), fps, (w,
h))
vid_writer.write(im0)
if save_txt or save_img:
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
s = f"\n{len(list(save_dir.glob('labels/*.txt')))} labels saved to {save_dir / 'labels'}" if save_txt else ''

#print(f"Results saved to {save_dir}{s}")

print(f'Done. ({time.time() - t0:.3f}s)')

if __name__ == '__main__':

parser = argparse.ArgumentParser()

parser.add_argument('--weights', nargs='+', type=str, default='yolov7.pt', help='model.pt path(s)')

parser.add_argument('--source', type=str, default='inference/images', help='source')  # file/folder, 0 for webcam

parser.add_argument('--img-size', type=int, default=640, help='inference size (pixels)')

parser.add_argument('--conf-thres', type=float, default=0.25, help='object confidence threshold')

parser.add_argument('--iou-thres', type=float, default=0.45, help='IOU threshold for NMS')

parser.add_argument('--device', default='', help='cuda device, i.e. 0 or 0,1,2,3 or cpu')

parser.add_argument('--view-img', action='store_true', help='display results')

parser.add_argument('--save-txt', action='store_true', help='save results to *.txt')

parser.add_argument('--save-conf', action='store_true', help='save confidences in --save-txt labels')

parser.add_argument('--nosave', action='store_true', help='do not save images/videos')

parser.add_argument('--classes', nargs='+', type=int, help='filter by class: --class 0, or --class 0 2 3')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
parser.add_argument('--agnostic-nms', action='store_true', help='class-agnostic
NMS')

parser.add_argument('--augment', action='store_true', help='augmented
inference')

parser.add_argument('--update', action='store_true', help='update all models')

parser.add_argument('--project', default='runs/detect', help='save results to
project/name')

parser.add_argument('--name', default='exp', help='save results to
project/name')

parser.add_argument('--exist-ok', action='store_true', help='existing
project/name ok, do not increment')

parser.add_argument('--no-trace', action='store_true', help='don`t trace model')

opt = parser.parse_args()

print(opt)

#check_requirements(exclude=('pycocotools', 'thop'))

with torch.no_grad():

if opt.update:  # update all models (to fix SourceChangeWarning)

for opt.weights in ['yolov7.pt']:

detect()

strip_optimizer(opt.weights)

else:

detect()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานทะเบียนคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
คำรับรองเล่มสหกิจศึกษา

วันที่ 9 เดือน กรกฎาคม พ.ศ 2566

ข้าพเจ้า นายทรัพย์ตะวัน วรรณวิจิตร รหัสประจำตัว 62050773
นักศึกษาหลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ ขอรับรองว่าโครงงานสหกิจศึกษา เรื่อง

ชื่อภาษาไทย การตรวจจับหมายเลขของสินค้าและโลโก้ของสินค้า และการเปลี่ยนหมายเลขของสินค้าให้อยู่ในรูปแบบของข้อความ เพื่อการตรวจสอบย้อนกลับ

ชื่อภาษาอังกฤษ Batch and Logo Object Detection and Batch OCR for Traceability

ปีการศึกษา 2565

เป็นผลงานวิจัยที่มิได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อนเรียบร้อยแล้ว และได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่มโครงงานสหกิจศึกษาฉบับสมบูรณ์แล้ว
โปรแกรมอักขราวิสุทธิ์ 1.55%

ลงชื่อ.....ทรัพย์ตะวัน.....
(นายทรัพย์ตะวัน วรรณวิจิตร)
นักศึกษา

ข้าพเจ้า ดร.สกุณา ศรีอโนมัย อาจารย์ที่ปรึกษาโครงงานสหกิจศึกษา ได้ตรวจสอบโครงงานสหกิจศึกษาของนักศึกษาข้างต้นแล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ.....สกุณา.....
อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้