

การตรวจจับโลโก้สินค้าแอลกอฮอล์ด้วยแบบจำลองการเรียนรู้เชิงลึก

พิชิตชัย พิมพ์โคตร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่  
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์  
มหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2564

# **ALCOHOL BRAND LOGO DETECTION USING DEEP LEARNING**

**PICHITCHAI PIMKOTE**

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**for the Degree of Master of Engineering**

**Department of Big Data Engineering,**

**College of Innovative Technology and Engineering,**

**Dhurakij Pundit University**

**2021**



หัวข้อวิทยานิพนธ์	การตรวจจับโลโก้สินค้าแอลกอฮอล์ด้วยแบบจำลอง
ชื่อผู้เขียน	การเรียนรู้เชิงลึก
อาจารย์ที่ปรึกษา	พิชิตชัย พิมพ์โคตร
สาขาวิชา	ดร. ธนภัทร มังคะจิตร
ปีการศึกษา	วิศวกรรมข้อมูลขนาดใหญ่
	2563

### บทคัดย่อ

การนำเทคโนโลยีด้านประมวลผลภาพถูกนำมาประยุกต์ใช้อย่างแพร่หลาย หนึ่งในงานที่เป็นที่นิยมคือการตรวจจับและการคัดแยกโลโก้สินค้าซึ่งสามารถนำไปประยุกต์ใช้ทั้งในภาครัฐและภาคธุรกิจ เช่น การตรวจจับรูปภาพเครื่องดื่มแอลกอฮอล์ที่ผิดกฎหมายโฆษณา การประเมินความพึงพอใจของลูกค้าที่มีต่อสินค้าจากภาพถ่าย เป็นต้น แต่เนื่องจากข้อมูลนั้นมีที่มาจากหลากหลายแหล่ง มีความแตกต่างกันอย่างมาก ทั้งขนาด ความละเอียดคมชัด แสงสี และสิ่งรบกวนในภาพซึ่งสิ่งเหล่านี้ล้วนแล้วแต่เป็นความท้าทายในการพัฒนาระบบให้สามารถจำแนกและตรวจจับโลโก้สินค้าได้อย่างมีประสิทธิภาพสูงสุด

งานวิจัยนี้นำเสนอขั้นตอนวิธีการเลือกใช้และเตรียมข้อมูล โดยเลือกยี่ห้อสินค้าที่ได้รับ ความนิยมในประเทศไทยและเป็นโลโก้ล่าสุดที่แต่ละยี่ห้อเลือกใช้ในปี 2562 จำนวน 6 ยี่ห้อ ได้แก่ เบียร์สิงห์, เบียร์ลีโอ, เบียร์ไฮเนเกน, เป๊ปซี่, โคคาโคล่า, สตาร์บัค โดยใช้ชุดข้อมูลที่ผู้วิจัยเก็บรวบรวมมาใช้ร่วมกับชุดข้อมูล Logos-32Plus และการเพิ่มความหลากหลายให้กับชุดข้อมูล ฟีกสอน (Data Augmentation) ให้เหมาะสม รวมไปถึงการเลือกใช้โมเดล Mask R-CNN และการ กำหนดค่าโมเดลให้เหมาะสมทำให้ได้ประสิทธิภาพมากที่สุดที่ mAP 0.982 และยังมีประสิทธิภาพ ที่ดีเมื่อทดสอบกับชุดข้อมูลที่มีคุณลักษณะที่หลากหลาย ได้แก่ ขนาดรูปภาพเล็กและมีความ ละเอียดต่ำ รูปภาพที่มีมืด แสงสว่างมากกว่าปกติ ภาพไม่คมชัด ภาพมีสิ่งรบกวน รวมไปถึงรูปภาพ ที่โลโก้ที่ไม่สมบูรณ์ ด้วย

Thesis Title	Alcohol brand logo detection using deep learning
Author	Pichitchai Pimkote
Thesis Advisor	Dr. Thanapat Kangkachit
Department	Big Data Engineering
Academic Year	2020

### ABSTRACT

The application of image processing technology is widely applied. One of the popular works is the brand logo detection which can be applied to both government and business sectors, such as the detection of illegal alcohol images, Assessing customer satisfaction from photos, but since that data comes from multiple sources. There are many differences in size, resolution, contrast, brightness, noise and blur all of which are challenges in developing a system to brand logos detection with maximum efficiency.

This research presents methodology of data selection, data preparation and modeling. By selecting the most popular brands in Thailand and the latest logos that each brand used in the 2019, consisting of 6 brands: Singha beer, Leo beer, Heineken beer, Pepsi, Coca Cola, and Starbucks using the data sets Our Logos combination with the Logos-32Plus data set and the appropriate variety of data augmentation sets. Used Mask R-CNN model and configuring the model appropriately, achieving the highest performance at mAP 0.982 and also performing well when tested with data sets that have a vary of features.

## กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จได้ด้วยการสนับสนุนและความช่วยเหลือจากหลายส่วน ขอบพระคุณที่ปรึกษาวิทยานิพนธ์ ดร. ธนภัทร ชั่งกะจิตร ที่ให้คำแนะนำ เสนอแนะ และผลักดันให้งานวิจัยนี้สำเร็จลุล่วงไปได้ ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ที่สละเวลาให้คำแนะนำและแนวทางให้งานวิจัยนี้สำเร็จ ขอขอบคุณอาจารย์และเจ้าหน้าที่หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ทุกท่านที่ให้ความรู้คำปรึกษาและประสานงานให้งานวิจัยเป็นไปอย่างราบรื่น ขอขอบคุณ ดร.เอกสิทธิ์ พัทธวงษ์ศักดิ์ ที่สนับสนุนเป็นอย่างดีมาตลอดในทุกเรื่อง ขอขอบคุณเพื่อนทุกท่านที่ให้คำปรึกษา ขอขอบคุณครอบครัวที่เป็นกำลังใจ ช่วยเหลือและให้การสนับสนุนในทุกๆเรื่อง ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นพื้นฐานในการต่อยอดองค์ความรู้ของผู้ที่สนใจศึกษาในงานด้านนี้ต่อไป

พิชิตชัย พิมพ์โคตร



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	๗
บทคัดย่อภาษาอังกฤษ.....	๗
กิตติกรรมประกาศ.....	๗
สารบัญตาราง.....	๗
สารบัญภาพ.....	๗
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตการวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์.....	3
2. ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ทฤษฎีที่เกี่ยวข้อง.....	4
2.2 งานวิจัยที่เกี่ยวข้อง.....	13
3. ระเบียบวิธีวิจัย.....	18
3.1 ขั้นตอนวิธีดำเนินการวิจัย.....	18
4. ผลงานวิจัย.....	35
4.1 ผลการทดสอบประสิทธิภาพและการอภิปรายผลแยกตามชุดข้อมูล.....	35
4.2 ผลการทดสอบประสิทธิภาพของโมเดลโดยภาพรวม.....	40
5. สรุปผลและข้อเสนอแนะ.....	41
5.1 สรุปผลการวิจัย.....	41
5.2 ข้อเสนอแนะ.....	43
บรรณานุกรม.....	44
ภาคผนวก.....	48
ประวัติผู้เขียน.....	73

## สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางคอนฟิวชันเมทริกซ์.....	11
2.2 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง.....	17
3.1 จำนวนข้อมูลแบบแยกคลาสของแต่ละชุดข้อมูล.....	20
3.2 ค่าสี RGB ของแต่ละชุดข้อมูล.....	21
3.3 จำนวนภาพแยกตามสัดส่วนขนาดโลโก้สินค้าต่อขนาดรูปภาพชุดข้อมูล.....	26
3.4 จำนวนภาพแยกตามประเภทโลโก้ในภาพสมบูรณ์ และไม่สมบูรณ์ 100 เปอร์เซ็นต์.....	27
3.5 แสดงโมเดล ข้อมูลฝึกสอนและการเพิ่มชุดข้อมูลฝึกสอน.....	35
4.1 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลปกติ (Our Logo + Logos-32).....	35
4.2 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูล Logos-32Plus.....	35
4.3 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูล FlickrLogos-47.....	36
4.4 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลโลโก้ไม่ครบ100%.....	36
4.5 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ.....	37
4.6 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลกลุ่มแสงมืดกว่าปกติ.....	37
4.7 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลสิ่งรบกวนในภาพ.....	38
4.8 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลภาพไม่คมชัด.....	38
4.9 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลชุดทดสอบสิ่งผิดปกติ 3 ชนิด.....	39



## สารบัญภาพ

ภาพที่	หน้า
2.1 โครงสร้างนิเวศอินเทอร์เน็ตเวิร์ค.....	5
2.2 โครงสร้างคอนโวลูชันนิเวศอินเทอร์เน็ตเวิร์ค.....	6
2.3 ตัวอย่างการคำนวณฟังก์ชันลักษณะ.....	7
2.4 แสดงตัวอย่างการเลือกคุณลักษณะแบบสูงสุด (Max Pooling).....	8
2.5 ความแตกต่างระหว่าง Semantic Segmentation และ Instance Segmentation.....	8
2.6 โครงสร้าง Mask R-CNN.....	9
2.7 สูตรคำนวณค่าไอโอยู.....	10
2.8 ตัวอย่างค่าไอโอยู.....	11
2.9 โครงสร้างการฝึกสอนโครงข่ายประสาทเทียมของ Deep Learning for Logo.....	13
2.10 โครงสร้างโครงข่ายประสาทเทียมของ Deep Learning for Logo Recognition..	13
2.11 ชุดข้อมูลที่ใช้ในงาน Deep Learning for Logo Recognition.....	14
2.12 เปรียบเทียบชุดข้อมูล งานวิจัย LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks.....	15
2.13 เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแต่ละแบบบนชุดข้อมูล Logos-18.....	15
2.14 เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแต่ละแบบบนชุดข้อมูล Logos-160.....	15
2.15 เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียม.....	16
2.16 แสดงจำนวนการฝึกสอนข้อมูลเทียบกับความถูกต้องแบบรวมทุกคลาส.....	16
3.1 ลำดับขั้นตอนวิธีดำเนินการวิจัย.....	18
3.2 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล Our Logos.....	22
3.3 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล Logos32-Plus.....	22
3.4 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล FlickrLogos-47.....	23
3.5 จำนวนรูปภาพชุดข้อมูล Our Logos แยกตามขนาดภาพ (หน่วยเป็นพิกเซล).....	24

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
3.6 จำนวนรูปภาพชุดข้อมูล Our Logos แยกตามขนาดภาพ (หน่วยเป็นพิกเซล).....	24
3.7 จำนวนรูปภาพชุดข้อมูล FlickrLogo-47 แยกตามขนาดภาพ (หน่วยเป็นพิกเซล)	25
3.8 ตัวอย่างการระบุประเภทวัตถุในภาพ.....	29
4.1 เปรียบเทียบประสิทธิภาพของโมเดลโดยรวม.....	40



# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญ

เทคโนโลยีที่พัฒนาไปอย่างมากในปัจจุบันทำให้ข้อมูลในรูปแบบรูปภาพเกิดขึ้นมากมายทั้งยังกระจายไปอย่างรวดเร็วจากเว็บไซต์และสังคมออนไลน์ ซึ่งเป็นความท้าทายในการพยายามจำแนกตรวจจับวัตถุในภาพเพื่ออธิบายความหมายหรือเพื่อจุดประสงค์อย่างใดอย่างหนึ่ง เทคโนโลยีการประมวลผลภาพจึงถูกนำมาใช้เพื่อลดระยะเวลาการทำงานและลดทรัพยากรบุคคล หนึ่งในงานที่เป็นที่นิยมคือการตรวจจับและการคัดแยกโลโก้สินค้าซึ่งสามารถนำไปประยุกต์ใช้ทั้งในภาครัฐและภาคธุรกิจ เช่น การตรวจจับรูปภาพเครื่องหมายการค้าที่ผิดกฎหมายโฆษณา การตรวจจับสินค้าละเมิดลิขสิทธิ์ การตรวจวัดและประเมินทัศนคติ ความพึงพอใจของลูกค้าที่มีต่อแบรนด์ เป็นต้น แต่เนื่องจากข้อมูลนั้นมีที่มาจากหลากหลายแหล่ง มีความแตกต่างกันอย่างมาก ทั้งขนาด ความละเอียดคมชัด แสงสี และสิ่งรบกวนในภาพซึ่งสิ่งเหล่านี้ล้วนแล้วแต่เป็นความท้าทายในการพัฒนาระบบให้สามารถจำแนกและตรวจจับโลโก้สินค้าได้อย่างมีประสิทธิภาพสูงสุด

ในการแข่งขัน การแยกประเภทของภาพ ImageNet Large Scale Visual Recognition Challenge (ILSVRC) ที่ใช้ฐานข้อมูล ImageNet ที่มีมากกว่า 14 ล้านภาพมีประเภทของรูปภาพมากกว่า 20,000 ประเภท วิธีการที่ได้รับความนิยมและมีความถูกต้องแม่นยำสูงก็คือการใช้โครงข่ายประสาทเทียม (Neural Networks) เป็นพื้นฐานในการสร้างแบบจำลองในการแยกประเภทของรูปภาพ ซึ่งโครงข่ายประสาทเทียม (Neural Networks) ก็ได้ถูกนำมาใช้ในงานวิจัยการแยกประเภทของโลโก้สินค้าด้วยเช่นกัน Deep Learning for Logo Recognition นำเสนอขั้นตอนวิธีการเตรียมข้อมูลเพื่อให้แบบจำลองมีประสิทธิภาพทำนายได้ถูกต้องแม่นยำที่สุดโดยใช้ชุดข้อมูล FlickrLogos-32 และ Logos-32Plus ซึ่งเป็นชุดข้อมูลโลโก้โดยเฉพาะและใช้โมเดลที่ไม่ซับซ้อนในการทำนายผล

งานวิจัย Automatic Graphic Logo Detection vis Fast Region-based Convolutional Networks[Gonçalo Oliveira, Xavier Frazão, André Pimentel, Bernardete Ribeiro.(2016)] ได้นำเสนอ

วิธีการตรวจจับรูปภาพโลโก้สินค้าด้วยการใช้โมเดล Fast Region-based Convolutional Networks (Fast R-CNN)[Ross Girshick (2015)] ใช้ชุดข้อมูล FlickrLogos-32[Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol.(2011)] งานวิจัย Logo-Net:Large-scale Deep Logo Detection and Brand Recognition with Deep Region-base Convolution Networks[Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang , Hui Xue, and Qiang Wu. (2015)]ได้นำเสนอชุดข้อมูล LOGO-Net (Logos-18, Logos-160) ที่มีจำนวนยี่ห้อสินค้าถึง 160 ยี่ห้อและงานวิจัยนี้ก็ได้นำเสนอวิธีการแยกประเภทโดยใช้โมเดล Fast Region-based Convolutional Networks (Fast R-CNN) งานวิจัย DeepLogo : Hitting Logo Recognition with the Deep Neural Network Hammer[4] นำเสนอโมเดลที่มีความซับซ้อนในการแยกภาพโดยใช้ชุดข้อมูล FlickrLogos-32

จากงานวิจัยที่อ้างถึงข้างต้นแล้วแต่ใช้โครงข่ายประสาทเทียมเป็นพื้นฐานในการสร้างโมเดลแยกตรวจจับและประเภทโลโก้สินค้าซึ่งโมเดลที่นิยมใช้และมีประสิทธิภาพและความแม่นยำในการทำนายสูงก็คือโมเดล Fast Region-based Convolutional Networks (Fast R-CNN) แต่อย่างไรก็ตามปัจจุบันโมเดล Fast Region-based Convolutional Networks (Fast R-CNN) ก็ได้ถูกพัฒนาให้มีประสิทธิภาพมากยิ่งขึ้นนั่นก็คือโมเดล Mask R-CNN ที่มีความแม่นยำสูงและสามารถคัดแยกประเภทของวัตถุในภาพได้ละเอียดมากยิ่งขึ้นอีกด้วย อีกทั้งชุดข้อมูลที่ใช้ในงานวิจัยข้างต้นเมื่อพิจารณาจากคุณสมบัติแล้วยังไม่สอดคล้องกับรูปภาพที่เกิดขึ้นจริงในอินเทอร์เน็ต หรือสื่อสังคมออนไลน์ เช่น ขนาดของภาพที่อยู่ในชุดข้อมูล FlickrLogo-32 หรือ ชุดข้อมูล Logos-32Plus[Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang , Hui Xue, and Qiang Wu. (2015)] มีขนาดใหญ่และความละเอียดสูงแต่รูปภาพที่พบในอินเทอร์เน็ตจะขนาดเล็กและความละเอียดค่อนข้างต่ำ รวมไปถึงขนาดของโลโก้ในภาพ ความสมบูรณ์ของโลโก้ในภาพสัดส่วนความสว่างความมืดของภาพ สิ่งรบกวนต่างๆในภาพที่เป็นปัจจัยในการฝึกสอนให้กับโมเดลเพื่อที่จะให้โมเดลนั้นสามารถตรวจจับรูปภาพที่มีคุณลักษณะหลากหลายขึ้นอีกด้วย

งานวิจัยนี้นำเสนอขั้นตอนวิธีการเลือกใช้และเตรียมข้อมูล โดยเลือกยี่ห้อสินค้าที่ได้รับค่านิยมในประเทศไทยและเป็นโลโก้ล่าสุดของแต่ละยี่ห้อเลือกใช้ในปี 2562 จำนวน 6 ยี่ห้อ ได้แก่ เบียร์สิงห์, เบียร์ลีโอ, เบียร์ไฮเนเกน, เป๊ปซี่, โคคาโคล่า, สตาร์บัค โดยใช้ชุดข้อมูลที่ผู้วิจัยเก็บรวบรวมมาใช้ร่วมกับชุดข้อมูล Logos-32Plus และการเพิ่มความหลากหลายให้กับชุดข้อมูลฝึกสอน (Data Augmentation)ให้เหมาะสม รวมไปถึงการเลือกใช้โมเดล Mask R-CNN และการกำหนดค่าโมเดลให้เหมาะสมทำให้ได้ประสิทธิภาพมากที่สุดที่ mAP 0.982 และยังมีประสิทธิภาพที่ดีเมื่อทดสอบกับชุด

ข้อมูลที่มีคุณลักษณะที่หลากหลาย ได้แก่ ขนาดรูปภาพเล็กและมีความละเอียดต่ำ รูปภาพที่มีดี แสงสว่างมากกว่าปกติ ภาพไม่คมชัด ภาพมีสิ่งรบกวน รวมไปถึงรูปภาพที่โลโก้ที่ไม่สมบูรณ์ ด้วย

## 1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อนำเสนอวิธีการเลือกใช้และการจัดเตรียมข้อมูลสำหรับฝึกสอน โมเดลที่ใช้สำหรับตรวจจับรูปภาพโลโก้สินค้า
2. เพื่อนำเสนอการเลือกใช้โมเดลและการตั้งค่าโมเดลให้เหมาะสมเพื่อให้มีประสิทธิภาพมากที่สุด

## 1.3 ขอบเขตงานวิจัย

1. ข้อมูลที่ใช้เลือกใช้ 6 โลโก้สินค้าจาก 6 ยี่ห้อที่ได้รับความนิยมในประเทศไทย ได้แก่ เบียร์สิงห์, เบียร์ลิโอ, เบียร์ไฮเนเกน, เป๊ปซี่, โคคาโคล่า และ สตาร์บัค และเป็นโลโก้ล่าสุดในปี 2562 ของแต่ละยี่ห้อ
2. ปัจจัยในการทดสอบเพื่อวัดประสิทธิภาพ ได้แก่ โลโก้ไม่สมบูรณ์ 100 เปอร์เซ็นต์, แสงสว่างมากกว่าปกติ, แสงมืดกว่าปกติ, สิ่งรบกวนในภาพ, ภาพไม่คมชัด

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำวิธีการที่นำเสนอไปใช้ในการพัฒนาระบบในการตรวจจับคัดกรองโลโก้หรือวัตถุอื่นในภาพที่มีคุณลักษณะหลากหลายให้มีประสิทธิภาพมากยิ่งขึ้น

## 1.5 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

Pichitchai Pimkote ; Thanapat Kangkachit,"Classification of alcohol brand logos using convolutional neural networks",2018 International Conference on Digital Arts, Media and Technology (ICDAMT)

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

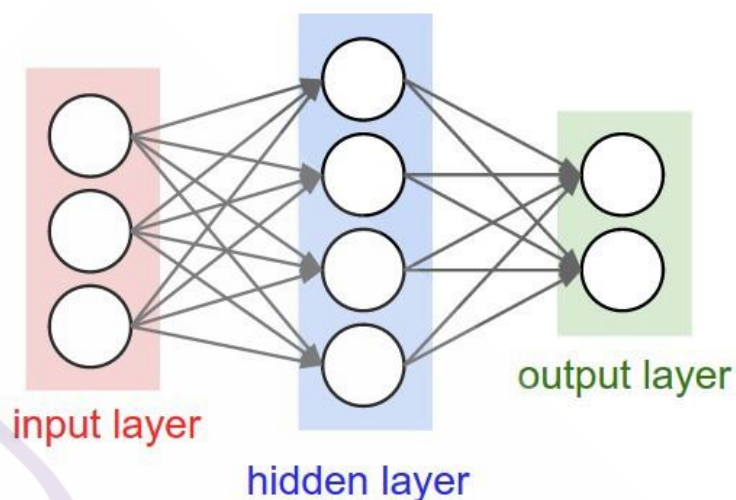
ในบทนี้จะกล่าวถึงทฤษฎีที่ใช้ในการวิจัย เทคนิคในการทดสอบประสิทธิภาพของโมเดล และวรรณกรรมที่เกี่ยวข้อง

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 นิวรอลเน็ตเวิร์ค และ คอนโวลูชันนิวรอลเน็ตเวิร์ค (Neural Network and Convolution Neural Networks)

นิวรอลเน็ตเวิร์คเป็นแบบจำลองการคำนวณทางคณิตศาสตร์ที่มีแรงบันดาลใจในการพัฒนา มาจากสมองของมนุษย์ โครงข่ายประสาทเทียมนั้นประกอบไปด้วยกลุ่มของโหนด (Node) ที่จัด ออกเป็น 3 ส่วนคือ

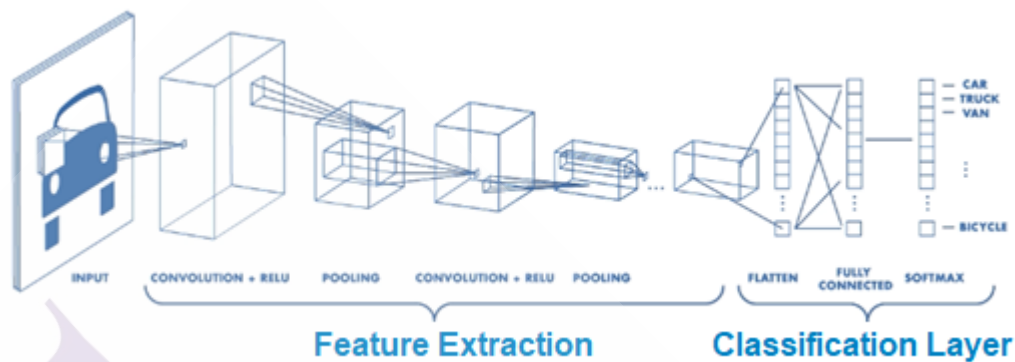
1. Input layer เป็นส่วนรับข้อมูลซึ่งจำนวนของโหนดขึ้นอยู่กับข้อมูลที่นำเข้ามาใช้ในโมเดล
2. hidden layer เป็นชั้นที่อยู่กึ่งกลางโดยสามารถกำหนดจำนวนโหนด หรือจำนวนชั้นของ hidden layer ได้ตามต้องการเพื่อความเหมาะสมและประสิทธิภาพที่ดีที่สุดกับงานที่ใช้
3. output layer เป็นชั้นคำตอบหรือชั้นที่นำข้อมูลที่ได้ไปใช้ จำนวนของโหนดก็ขึ้นอยู่กับ จำนวนคำตอบที่ต้องการ



ภาพที่ 2.1 โครงสร้างนิเวรอลเน็ตเวิร์ค

ในการใช้นิเวรอลเน็ตเวิร์คมาใช้ในการตรวจจับวัตถุข้อมูลประเภทรูปภาพ อาจใช้วิธีการการสกัดคุณลักษณะ โดยตัวสกัดคุณลักษณะ (Feature Extractor) ของวัตถุที่สนใจในรูปภาพก็มีหลากหลายวิธี เช่น Histogram of Oriented Gradients [N. Dalal and B. Triggs. (2005)], Scale-Invariant Feature Transform (SIFT) [Stefan Romberg], Speeded Up Robust Feature (SURF) [Nabeel Younus Khan, Brendan McCane, Geoff Wyvill. (2005)] เป็นต้น โดยตัวสกัดคุณลักษณะเหล่านี้จะถูกออกแบบมาให้เหมาะสมกับข้อมูลที่จะใช้ตรวจจับหากมีการนำไปใช้งานอื่นประสิทธิภาพก็อาจจะลดลง จึงมีแนวคิดในการสกัดคุณลักษณะจากชุดข้อมูลสอน เพื่อที่ไม่ต้องออกแบบตัวสกัดคุณลักษณะที่เหมาะสมกับข้อมูลทุกครั้ง วิธีการนี้ถูกเรียกว่าคอนโวลูชันนิเวรอลเน็ตเวิร์ค

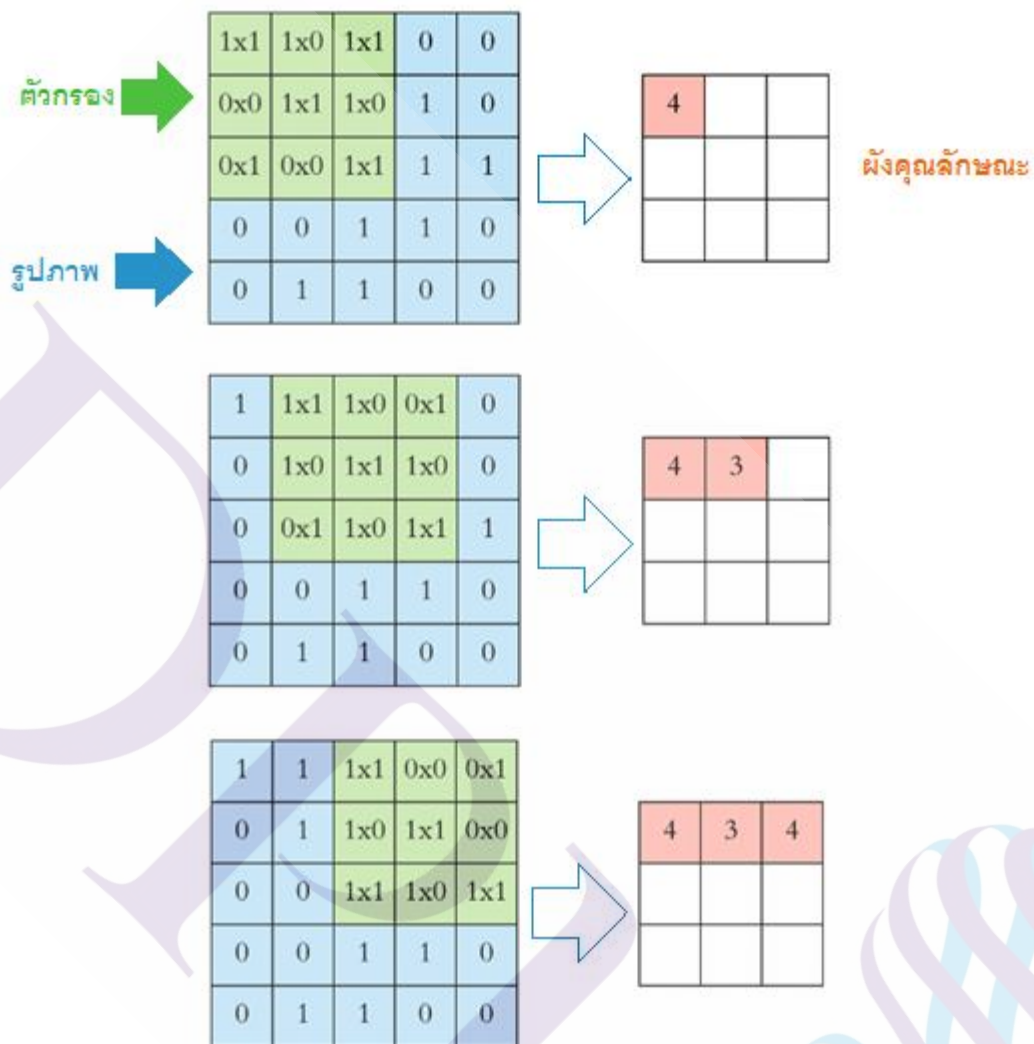
คอนโวลูชันนิเวรอลเน็ตเวิร์ค ประกอบไปด้วย ชั้นสกัดคุณลักษณะ (Feature Extraction) และ ชั้นจำแนก (Classification Layer)



ภาพที่ 2.2 โครงสร้างคอนโวลูชันนิรอลเน็ตเวิร์ค

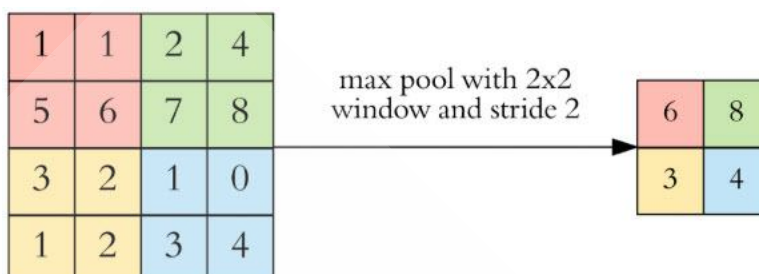
ขั้นตอนการทำสกัดคุณลักษณะรูปภาพของวิธีคอนโวลูชันเริ่มการจากใช้ ตัวกรอง(filter) คำนวณกับข้อมูลเข้า โดยแต่ละครั้งจะคำนวณตามขนาดของตัวกรอง และเลื่อน(Stride) ให้ครบทั้งภาพ ผลที่ได้จากการคำนวณจะเป็น ฟังก์ชันลักษณะ(Feature map)





ภาพที่ 2.3 ตัวอย่างการคำนวณฟังก์ชันลักษณะ

หลังจากนั้นจะนำฟังก์ชันลักษณะที่ได้ทั้งหมดมาเข้ากระบวนการเลือกคุณลักษณะ (Poling) วิธีที่นิยมใช้กันคือ การเลือกคุณลักษณะค่าสูงสุด (Max Pooling) และการเลือกคุณลักษณะค่าเฉลี่ย

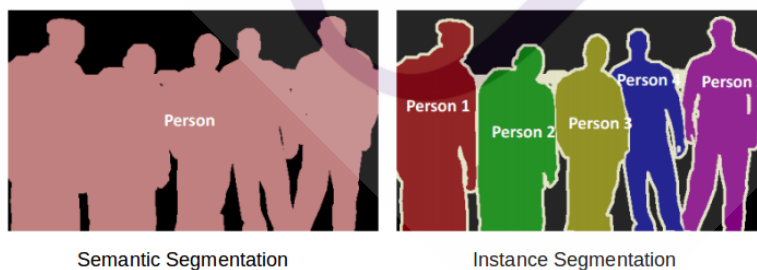


ภาพที่ 2.4 แสดงตัวอย่างการเลือกคุณลักษณะแบบสูงสุด (Max Pooling)

โดยงานวิจัยนี้ใช้โมเดล Mask-RNN ซึ่งใช้วิธีคอนโวลูชันนิรอลเน็ตเวิร์คในการสกัดคุณลักษณะรูปภาพเพื่อใช้ในกระบวนการต่อไป

### 2.1.2 การตรวจจับและแยกประเภทวัตถุแบบ Instance Segmentation ด้วยคอนโวลูชันนิรอลเน็ตเวิร์ค

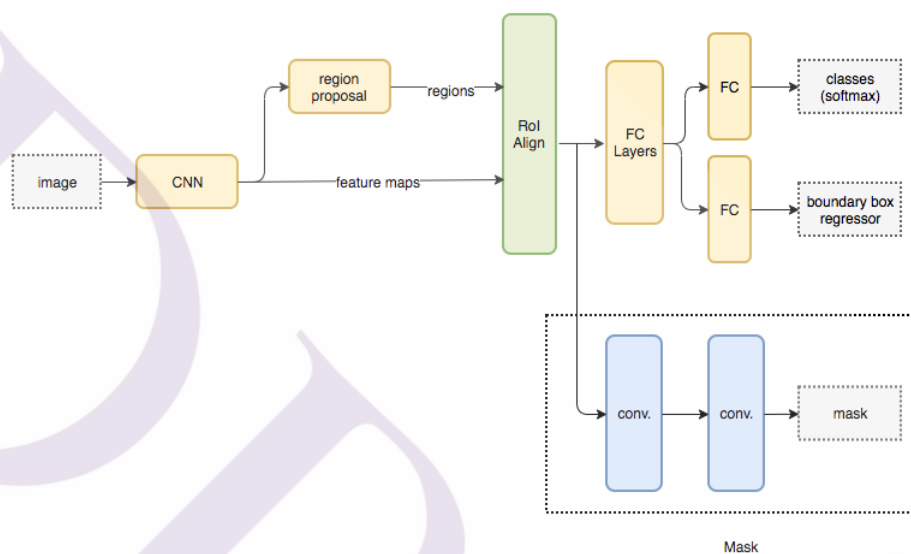
วิธีการตรวจจับวัตถุในรูปภาพด้วยถูกพัฒนาขึ้นมาเป็นอย่างมากซึ่งในงานวิจัยนี้ได้เลือกใช้วิธี Mask R- CNN ซึ่งมีแนวคิดและวิธีการพื้นฐานมาจากการใช้นิรอลเน็ตเวิร์ค ซึ่ง Mask R- CNN [Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. (2018)] นอกจากจะสามารถตรวจจับวัตถุในภาพได้แล้วยังสามารถแยกหรือระบุตำแหน่งของวัตถุแต่ละอย่างในภาพได้ในระดับพิกเซลได้อีกด้วย



ภาพที่ 2.5 ความแตกต่างระหว่าง Semantic Segmentation และ Instance Segmentation

### 2.1.2.1 Mask R-CNN [Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. (2018)]

พัฒนาต่อยอดมาจาก Faster R-CNN [S. Ren, K. He, R. Girshick, and J. Sun. (2017)] ซึ่งนอกจากจะตรวจจับวัตถุในภาพแล้วยังสามารถแยกและระบุประเภทและตำแหน่งของวัตถุในภาพได้อีกด้วย



ภาพที่ 2.6 โครงสร้าง Mask R-CNN

โครงสร้างหลักของ Mask R-CNN ประกอบไปด้วย

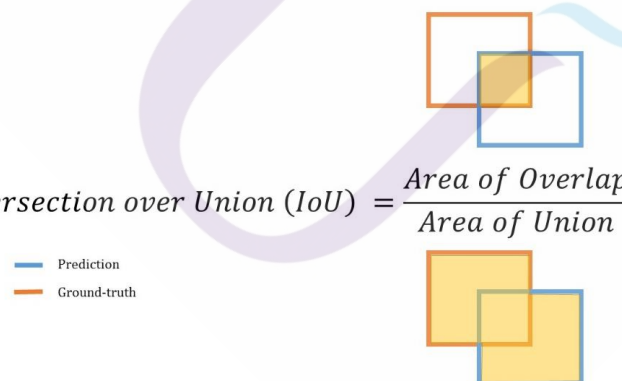
- Backbone Model เป็นส่วนที่ใช้คอนโวลูชันนิวรอลเน็ตเวิร์กสกัดคุณลักษณะ (Feature Extraction) โดยจะได้ผลลัพธ์เป็นฟังก์ชันลักษณะ (Feature map) เพื่อใช้ในขั้นตอนต่อไป สำหรับโครงสร้างหลักส่วน Backbone Model ในงานวิจัยนี้จะใช้โครงสร้าง ResNet 101 [20] และถูกฝึกสอนมาแล้วด้วยชุดข้อมูล ImageNet [Olga Russakovsky\*, Jia Deng\* (2015)]
- Region Proposal Network (RPN) เป็นส่วนที่เลือกฟังก์ชันลักษณะ (Feature map) ใดที่น่าจะมีวัตถุอยู่บ้างผลลัพธ์ที่ได้จะเป็น Region map

- ส่วน ROI Pooling เป็นส่วนที่เลือกเอา ฟังก์ชันลักษณะ Feature map และ Region map มีวัตถุประเภทใดบ้าง
- ส่วนทำนายผล (Classifier) ในส่วนนี้จะใช้นิวรอลเน็ตเวิร์ค และ คอนโวลูชันนิวรอลเน็ตเวิร์คทำนายผลโดยจะแยกทำนายผล 3 แบบคือ 1. ทำนายผลคลาส (Classification) 2. ทำนายส่วนกรอบตำแหน่งวัตถุ(boundary box regression) 3. ทำนายว่าพิกเซลไหนคือวัตถุ (Mask)

### 2.1.3 การวัดประสิทธิภาพและประเมินผล

2.3.1 ค่าไอโอยู (Intersection over Union: IoU) [Intersection over Union (IoU) for object detection]

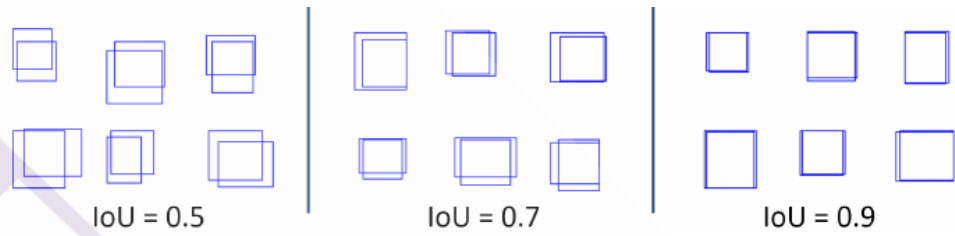
ค่าไอโอยูเป็นค่าวัดความแม่นยำของการตรวจจับวัตถุ ซึ่งเป็นการวัดอัตราส่วนพื้นที่ที่ทับซ้อนกัน(Area of Overlap) ระหว่างพื้นที่ที่ทำนาย (Prediction Area) กับ พื้นที่จริงของวัตถุ (Ground Truth Area) หารด้วยพื้นที่รวมทั้งหมด (Area of Union) ระหว่างพื้นที่ที่ทำนาย กับ พื้นที่จริงของวัตถุ

$$\text{Intersection over Union (IoU)} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


— Prediction  
— Ground-truth

ภาพที่ 2.7 สูตรคำนวณค่าไอโอยู

ในงานวิจัยนี้กำหนดค่าไอโอยูในการตรวจจับโลโก้ในภาพ หากค่าไอโอยูเกินกว่า 0.7 ถือว่าตรวจจับโลโก้ในภาพ โดยตัวอย่างของการวัดค่าไอโอยู แสดงในรูปที่ 2.7



ภาพที่ 2.8 ตัวอย่างค่าไอโอยู

### 2.3.2 คอนฟิวชันเมทริกซ์ (Confusion Matrix)

คอนฟิวชันเมทริกซ์ คือการประเมินประสิทธิภาพจากการทำนายผลเทียบกับผลลัพธ์จริง

ตารางที่ 2.1 ตารางคอนฟิวชันเมทริกซ์

		ผลลัพธ์จริง	
		Positive	Negative
ผลการทำนาย	Positive	TP	FP
	Negative	FN	TN

True Positive (TP) คือ ผลการทำนาย (Prediction) ว่ามันมีจริง (ค่าไอโออยู่มากกว่า 0.5) และ ผลลัพธ์จริง (Ground Truth) บอกว่ามีจริง

True Negative (TN) คือ ผลการทำนายว่ามันไม่มีจริง และผลลัพธ์จริงก็บอกว่ามันไม่มีจริง

False Positive (FP) คือ ผลการทำนายว่ามันมีจริง (ค่าไอโออยู่มากกว่า 0.5) แต่ผลลัพธ์จริงก็บอกว่ามันไม่มีจริง

False Negative (FN) คือ ผลการทำนายว่ามันไม่มีจริง (ค่าไอโออยู่น้อยกว่า 0.5) แต่ผลลัพธ์จริงก็บอกว่ามันมีจริง

ในงานวิจัยใช้คอนฟิวชันเมตริกซ์ในการวัดประสิทธิภาพการทำนายผลการตรวจจับโลโกในแต่ละภาพว่าโลโกที่ทำนายมานั้นถูกต้องหรือไม่

### 2.3.3 การวัดประสิทธิภาพแยกตามคลาส

ในการวัดประสิทธิภาพนั้นจะใช้วิธีวัดประสิทธิภาพแยกตามคลาส ประกอบด้วย 3 ค่า ดังนี้

การวัดค่า Precision คำนวณได้จากสูตร

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

การวัดค่า Recall คำนวณได้จากสูตร

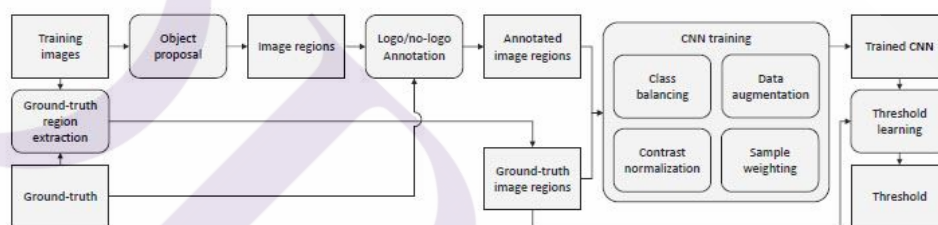
$$Recall = \frac{TP}{TP + FN}$$

การวัดความถูกต้องพิจารณาทุกคลาส (Accuracy) คำนวณได้จากสูตร

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

## 2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยของ Simone Bianco, Marco Buzzelli, Davide Mazzini, Raimondo Schettini ได้ศึกษาเรื่อง Deep Learning for Logo Recognition [Simone Bianco, Marco Buzzelli, Davide Mazzini, Raimondo Schettini.(2017)] ซึ่งนำเสนอวิธีการในการคัดกรองภาพจากโลโก้สินค้าที่มีประสิทธิภาพ โดยเน้นไปที่การจัดเตรียมข้อมูลก่อนฝึกสอนโครงข่ายประสาทเทียม ได้แก่ การตัด แยกวัตถุและกำหนดชื่อวัตถุในภาพ (Object-proposal logo annotations), การกำหนดจำนวนข้อมูลในคลาสให้มีจำนวนใกล้เคียงกัน (Class balancing), การเพิ่มความหลากหลายให้กับชุดข้อมูล (Data augmentation) และยังใช้โครงสร้างโครงข่ายประสาทเทียมที่เล็กและมีความซับซ้อนน้อย



ภาพที่ 2.9 โครงสร้างการฝึกสอนโครงข่ายประสาทเทียมของ Deep Learning for Logo Recognition

Layers	
1	Conv 32 filters of 5x5
2	Pool (max) with stride 2
3	Relu
4	Conv 32 filters of 5x5
5	Relu
6	Pool (average) with stride 2
7	Conv 64 filters of 5x5
8	Relu
9	Pool (average) with stride 2
10	Fully Connected of size 64
11	Fully Connected of size 33
12	Softmax

ภาพที่ 2.10 โครงสร้างโครงข่ายประสาทเทียมของ Deep Learning for Logo Recognition

สำหรับชุดข้อมูลในงานวิจัยนี้ใช้จะเป็น ชุดข้อมูล 2 ชุดคือ FlickrLogos-32 และ Logos-32Plus

	FlickrLogos-32	Logos-32plus
Total images	8240	7830
Images containing logo instances	2240	7830
Train + Validation annotations	1803	12302
Average annotations for class (Train + Validation)	40	400
Total annotations	3405	12302

ภาพที่ 2.11 ชุดข้อมูลที่ใช้ในงาน Deep Learning for Logo Recognition

2.2.2 งานวิจัยของ Goncalo Oliveira, Xavier Frazao, Andre Pimentel, Bernardete Ribeiro ได้ศึกษาเรื่อง Automatic Graphic Logo Detection via Fast Region-based Convolutional Networks [Gonçalo Oliveira, Xavier Frazão, André Pimentel, Bernardete Ribeiro.(2016)] นำเสนอการคัดแยกรูปภาพโลโก้สินค้าโดยใช้เครือข่ายประสาทเทียมแบบ Fast Region-based Convolutional Networks (FRCN) และใช้ชุดข้อมูล FlickrLogos-32 ให้ความถูกต้องในการคัดแยกรวมทุกคลาสอยู่ที่ 93%

2.2.3 งานวิจัยของ Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang, Hui Xue, Qiang Wu [Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang , Hui Xue, and Qiang Wu. (2015)] ได้ศึกษาเรื่อง LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks [Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang , Hui Xue, and Qiang Wu. (2015)] นำเสนอชุดข้อมูลที่ชื่อ Logos-160 และ Logos-18 ที่มีจำนวนรูปภาพมากกว่าชุดข้อมูล FlickrLogos-32 ที่เป็นที่ยอมรับซึ่ง Logos-160 มีจำนวนถึง 73,414 ภาพ 100 ยี่ห้อ



Dataset	#Image	#Logo	#Brand	#Logo Object
Logos-160	73414	160	100	130608
Logos-18	8460	18	10	16043
BelgaLogos	10000	37	37	2695
FlickrLogos-27	1080	27	27	4671
FlickrLogos-32	8240	32	32	5644

ภาพที่ 2.12 เปรียบเทียบชุดข้อมูล งานวิจัย LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks

อีกทั้งยังนำเสนอ โครงข่ายประสาทเทียม ได้แก่ RCNN(CaffeNet), FRCN(CaffeNet), FRCN(VGG1024), FRCN(VGG16), SPPnet(ZF) เปรียบเทียบประสิทธิภาพในด้านการคัดแยกรูปภาพ และ ในด้านการใช้ทรัพยากรด้วย

Algorithm(model)	mAP(%)	Accuracy (%)	AUC (%)	total train time	test time / image	GPU memory
RCNN(CaffeNet)	69.1	95.2	95.3	2444 (min)	20886(ms)	2.39 (GB)
RCNN(CaffeNet-w/o-ft)	55.1	86.5	86.4	1549 (min)	20881(ms)	2.39 (GB)
FRCN(CaffeNet)	58.8	93.2	92.0	147 (min)	448 (ms)	1.67 (GB)
FRCN(VGG1024)	59.8	94.8	93.6	253 (min)	540 (ms)	3.04 (GB)
FRCN(VGG16)	61.4	94.7	93.2	1312 (min)	859 (ms)	10.86 (GB)
SPPnet(ZF-w/o-bb)	54.5	92.5	92.3	707 (min)	968 (ms)	2.21 (GB)
SPPnet(ZF)	59.1	92.5	92.3	749 (min)	1199 (ms)	2.21 (GB)

ภาพที่ 2.13 เปรียบเทียบประสิทธิภาพของ โครงข่ายประสาทเทียมแต่ละแบบบนชุดข้อมูล Logos-18

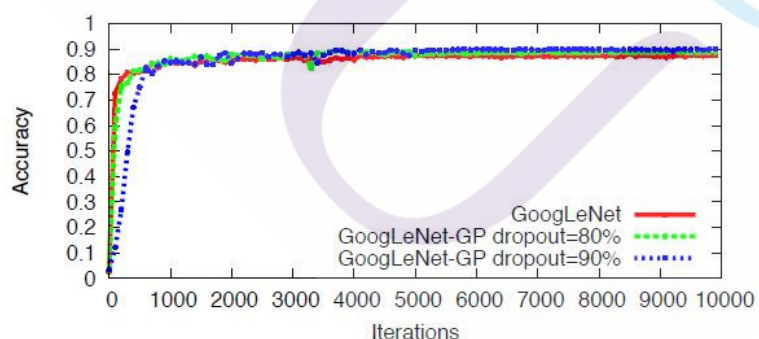
Alg(model)	mAP(%)	Accuracy (%)	AUC(%)	total train time	test time / image	GPU memory
RCNN(Caffenet)	69.9	90.0	89.6	8783 (min)	27273 (ms)	3.74(GB)
FRCN(CaffeNet)	61.0	81.6	82.6	169 (min)	685 (ms)	1.71 (GB)
FRCN(VGG1024)	60.3	81.3	82.3	283 (min)	752 (ms)	3.05 (GB)
FRCN(VGG16)	65.8	85.8	86.8	1362 (min)	1044 (ms)	10.89 (GB)
SPPnet(ZF-w/o-bb)	53.6	81.8	82.0	1360 (min)	1218 (ms)	3.63 (GB)
SPPnet(ZF)	58.1	81.8	82.0	1494 (min)	1639 (ms)	3.63 (GB)

ภาพที่ 2.14 เปรียบเทียบประสิทธิภาพของ โครงข่ายประสาทเทียมแต่ละแบบบนชุดข้อมูล Logos-160

2.2.4 งานวิจัยของ Forrest N. Iandola, Anting Shen, Peter Gao, Kurt Keutzer ได้ศึกษาเรื่อง DeepLogo : Hitting Logo Recognition with the Deep Neural Network [Forrest N. Iandola, Anting Shen, Peter Gao and Kurt Keutzer.(2015)] ซึ่งเป็นการเปรียบเทียบประสิทธิภาพการคัดแยกรูปภาพจากโครงข่ายประสาทเทียมแบบต่างๆ ได้แก่ VGG-19, AlexNet, Full-Inception, GoogLeNet ซึ่งเป็นโครงข่ายประสาทเทียมที่ได้ฝึกสอนด้วยชุดข้อมูล ImageNet-2012 และฝึกสอนเพิ่มเติมด้วยชุดข้อมูล FlickrLogos-32

Method	Dropout	Pretrain	Finetune	Accuracy
Random Chance	-	-	FlickrLogos	$\frac{1}{32} = 3.12\%$
baseline: Complete Rank Transform [4]	-	-	FlickrLogos	88.97%
VGG-19 (ours)	0.5	ImageNet-2012	FlickrLogos	3.12%
AlexNet (ours)	0.5	ImageNet-2012	FlickrLogos	70.1%
Full-Inception (ours)	0.7	ImageNet-2012	FlickrLogos	77.1%
GoogLeNet (ours)	0.7	ImageNet-2012	FlickrLogos	87.6%
GoogLeNet (ours)	0.8	ImageNet-2012	FlickrLogos	88.7%
GoogLeNet-GP (ours)	0.8	ImageNet-2012	FlickrLogos	89.1%
GoogLeNet-GP (ours)	0.9	ImageNet-2012	FlickrLogos	89.6%
FullClassify (ours)	0.9	ImageNet-2012	FlickrLogos	89.2%

ภาพที่ 2.15 เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแต่ละแบบ



ภาพที่ 2.16 แสดงจำนวนการฝึกสอนข้อมูลเทียบกับความถูกต้องแบบรวมทุกคลาส

ตารางที่ 2.2 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ชื่องานวิจัย	ผู้จัดทำ	ปีที่ตีพิมพ์	โมเดลที่ใช้	ชุดข้อมูล	ความแม่นยำ
Deep Learning for Logo Recognition[24]	Simone Bianco, Marco Buzzelli, Davide Mazzini, Raimondo Schettini	2017-05	CNN	FlicksrLogos-32 + Logos-32plus	96.0
Automatic Graphic Logo Detection via Fast Region-based Convolutional Networks[25]	Gonçalo Oliveira, Xavier Frazão†, André Pimentel†, Bernardete Ribeiro	2016-04	FRCN	FlicksrLogos-32	0.93
LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks [1]	Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang, Hui Xue, Qiang Wu	2015-11	RCNN	Logos-18,Logos160	95.2, 90.00
DeepLogo_Hitting Logo Recognition with the Deep Neural Network [4]	Forrest N. Iandola, Anting Shen , Peter Gao ,Kurt Keutzer	2015-10	GoogLeNet	FlicksrLogos-32	89.6

## บทที่ 3

### ระเบียบวิธีวิจัย

#### 3.1 ขั้นตอนวิธีดำเนินการวิจัย

ในบทนี้กล่าวถึงขั้นตอนการดำเนินการวิจัยการตรวจจับโลโก้สินค้าให้มีประสิทธิภาพโดยมีขั้นตอนดำเนินการหลัก 5 ขั้นตอนได้แก่ เก็บรวบรวมข้อมูล, การจัดเตรียมข้อมูล, การเพิ่มชุดตัวอย่างข้อมูล, การสร้างโมเดล และการประเมินผล โดยลำดับขั้นตอนตามรูปที่ 3.1



ภาพที่ 3.1 ลำดับขั้นตอนวิธีดำเนินการวิจัย

##### 3.1.1 เก็บรวบรวมข้อมูล

งานวิจัยนี้จะใช้รูปภาพที่เกิดขึ้นในชีวิตประจำวันโดยเลือกกลุ่มตัวอย่างเป็นยี่ห้อเครื่องดื่มที่ได้รับความนิยมในประเทศไทย 6 ยี่ห้อ ได้แก่ ได้แก่ เบียร์สิงห์, เบียร์ลิโอ, เบียร์ไฮเนเกน, เป๊ปซี่, โกลาโกลา, สตาร์บัค ข้อมูลที่ใช้ในการวิจัยประกอบไปด้วย 3 ชุดข้อมูล ได้แก่ ชุดที่ 1 OurLogos ชุดที่ 2. Logos-32Plus ชุดที่ 3. FlickrLogos-47

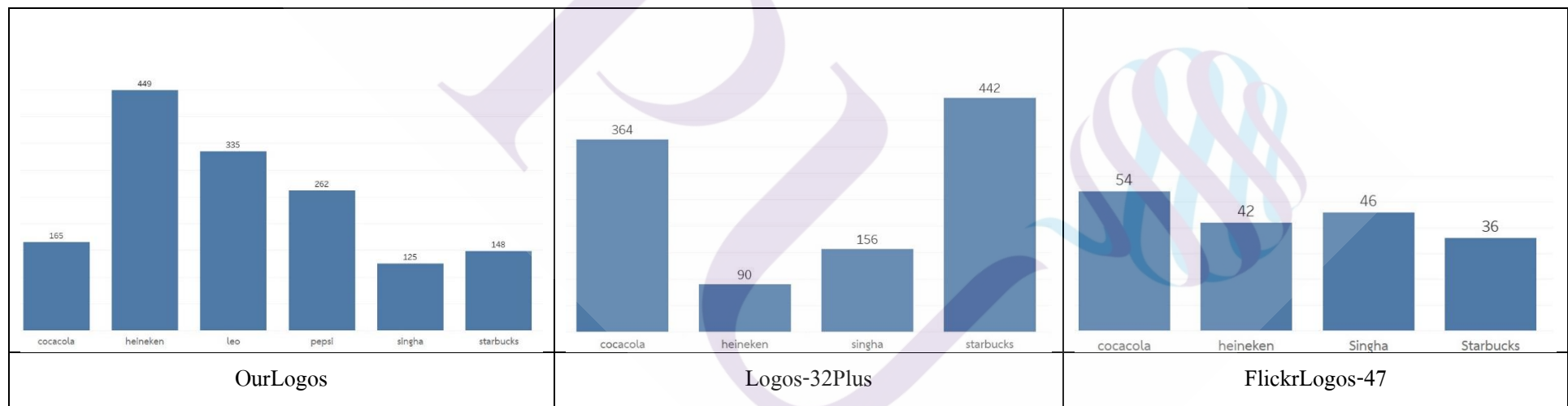
ชุดที่ 1. OurLogos เป็นข้อมูลที่เก็บรวบรวมจากจากเว็บไซต์ และสังคมออนไลน์ เป็นรูปภาพที่เกิดจากการบริโภค และสถานการณ์จริง เป็นชุดข้อมูลหลักที่ใช้ในงานวิจัยนี้ประกอบไปด้วย 6 คลาสได้แก่ เบียร์สิงห์ (singha), เบียร์ลีโอ (Leo), เบียร์ไฮเนเกน( heineken), เป๊ปซี่ (pepsi), โคคาโคลา (cocacola), สตาร์บัค (starbucks) ในการเก็บข้อมูลชุดนี้จะเลือกใช้โลโก้ล่าสุดของแต่ละยี่ห้อสินค้าและเลือกรูปภาพที่เกิดขึ้นจากชีวิตประจำวันซึ่งจะมีความหลากหลายทางคุณลักษณะ เช่น ขนาดของภาพ, ขนาดของโลโก้ในภาพ, ความคมชัด แสงสี, สิ่งรบกวน



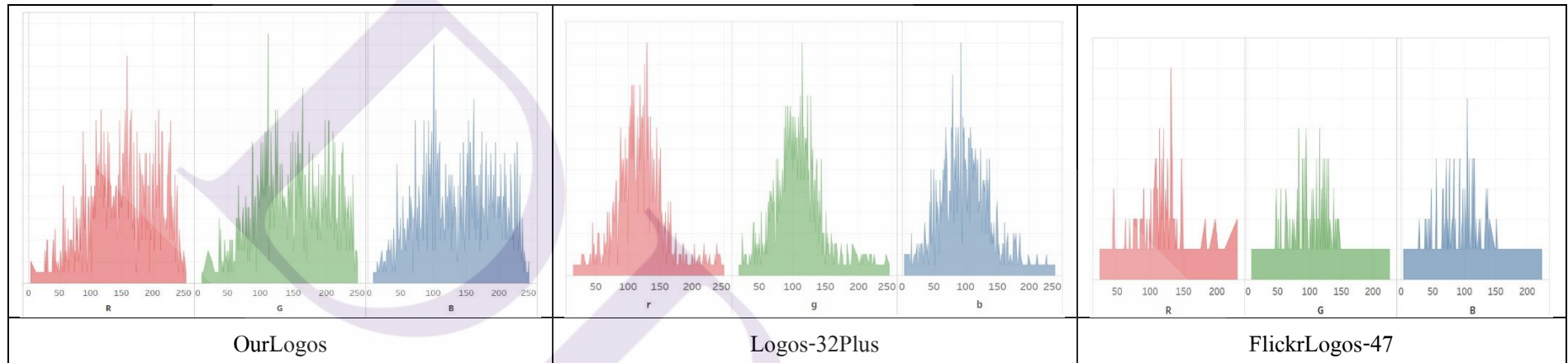
ชุดที่ 2. Logos-32Plus เป็นชุดฐานข้อมูลมาตรฐาน มีทั้งหมด 32 คลาส ได้แก่ HP, adidas, audi, apple, becks, bmw, carlsberg, chimay, cocacola, corona, dhl, erdinger, esso, fedex, ferrari, ford, fosters, google, Guinness, heineken, mika, nvidia, paulaner, pepsi, rittersport, shell, singha, starbucks, stellaarfois, texaco, tsingtao, ups แต่จะเลือกใช้เพียง cocacola, heineken, singha, starbucks เนื่องจากเป็น โลโก้ที่ตรงกันกับ คลาสที่จะใช้งาน

ชุดที่ 3. FlickrLogos-47 เป็น ชุดข้อมูลมาตรฐานที่พัฒนาเพิ่มเติมจากชุดข้อมูล FlickrLogos-32 ที่เป็นที่ยอมรับในงานวิจัยด้านการคัดแยกโลโก้

ตารางที่ 3.1 จำนวนข้อมูลแบบแยกคลาสของแต่ละชุดข้อมูล

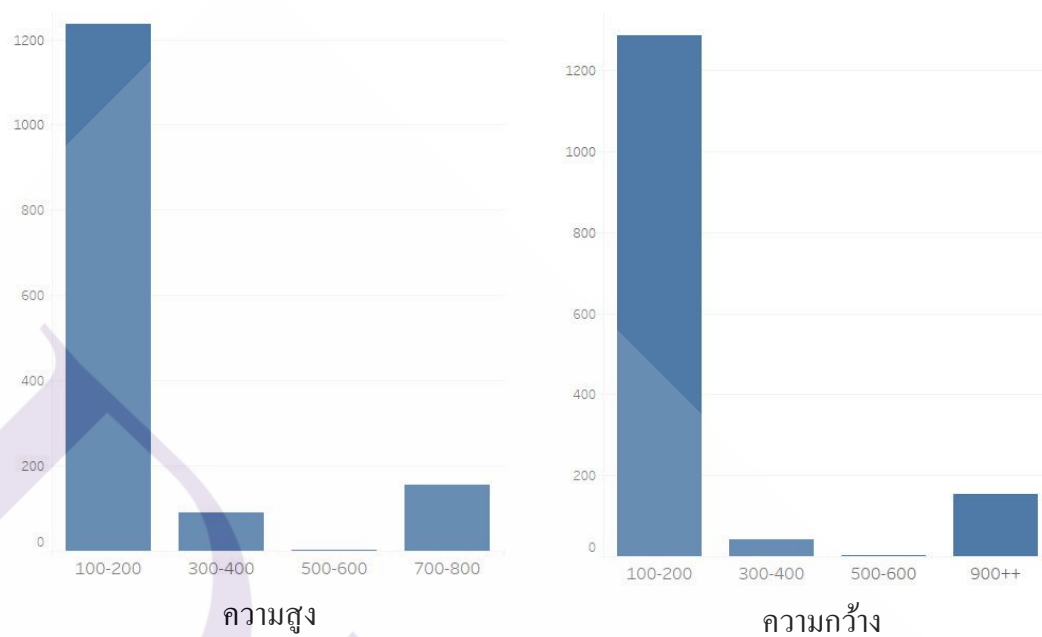


ตารางที่ 3.2 ค่าสี RGB ของแต่ละชุดข้อมูล

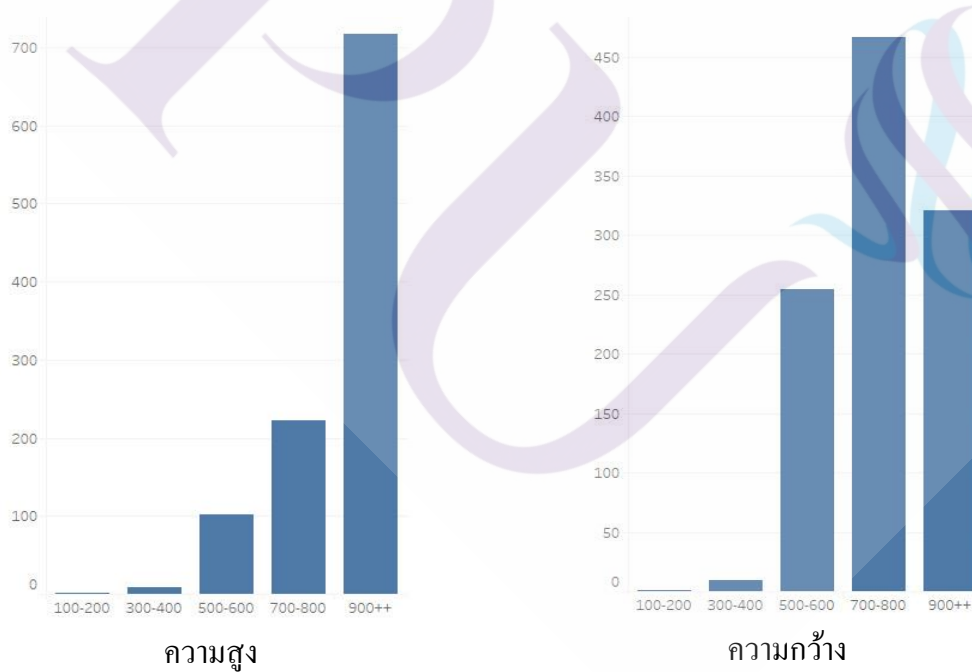


ขนาดของรูปภาพของแต่ละชุดข้อมูลแสดงแยกเป็นความสูง ความกว้างตามรูปภาพที่ 3.2-3.4 ซึ่งจะจัดกลุ่มแสดงผลเป็น 4 กลุ่มคือ

- กลุ่ม 100-200 คือกลุ่มที่มีความกว้าง หรือ สูง จาก 100 พิกเซล ขึ้นไป แต่ ต่ำกว่า 300 พิกเซล
- กลุ่ม 300-400 คือกลุ่มที่มีความกว้าง หรือ สูง จาก 300 พิกเซล ขึ้นไป แต่ ต่ำกว่า 500 พิกเซล
- กลุ่ม 700-800 คือกลุ่มที่มีความกว้าง หรือ สูง จาก 700 พิกเซล ขึ้นไป แต่ ต่ำกว่า 900 พิกเซล
- กลุ่ม 900++ คือกลุ่มที่มีความกว้าง หรือ สูง จาก 900 พิกเซล ขึ้นไป

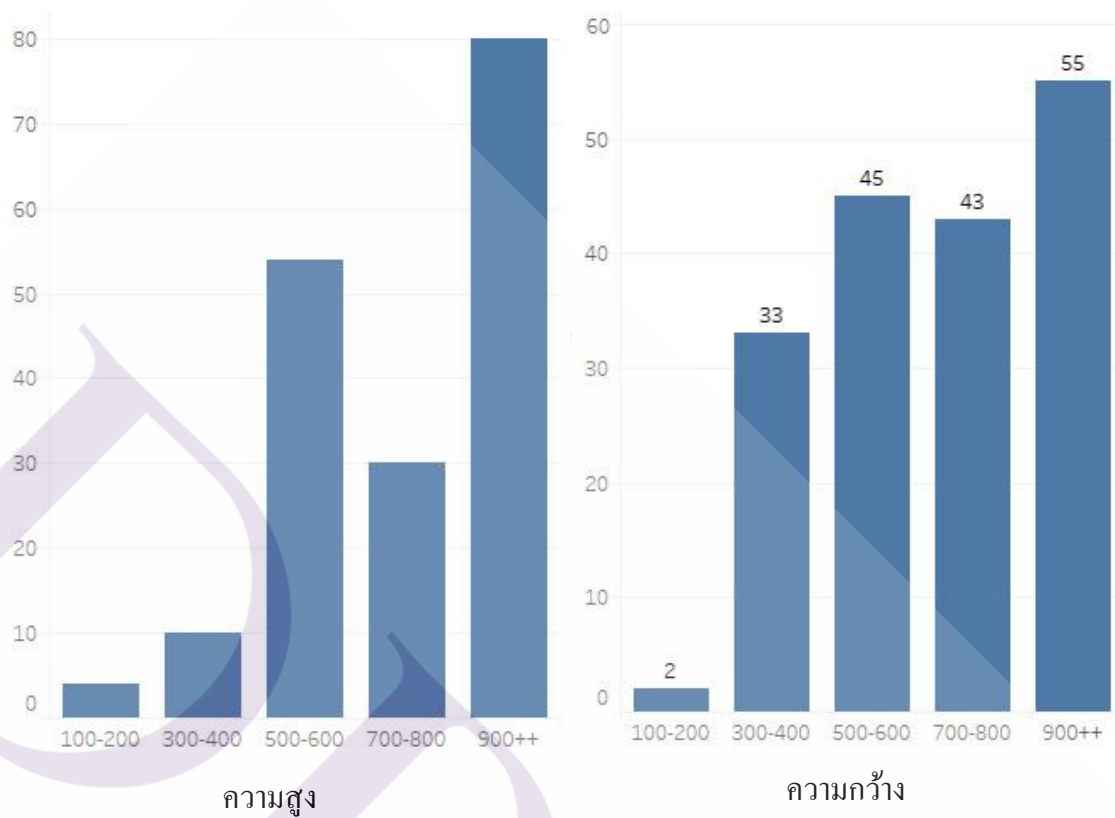


ภาพที่ 3.2 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล Our Logos



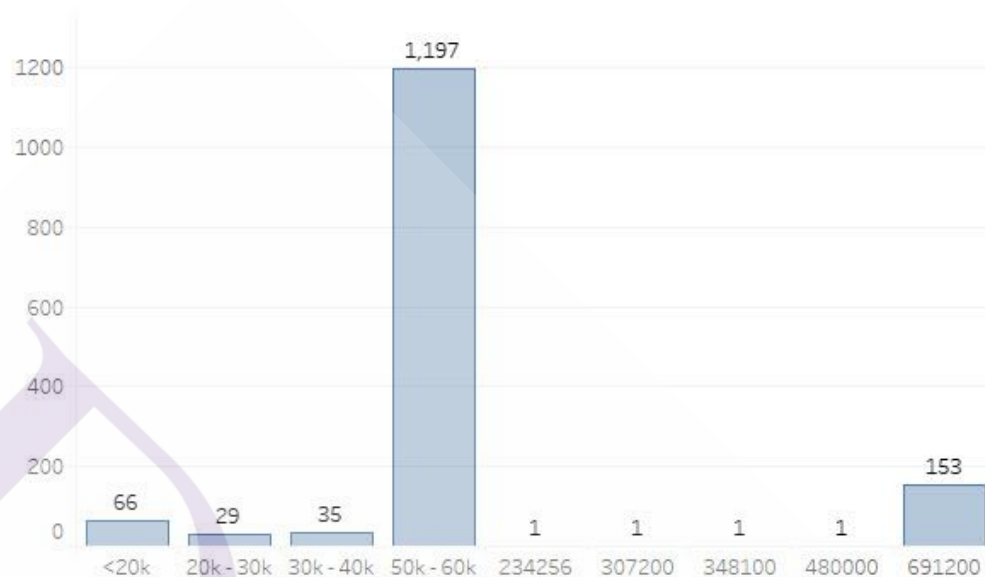
ภาพที่ 3.3 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล Logos32-Plus



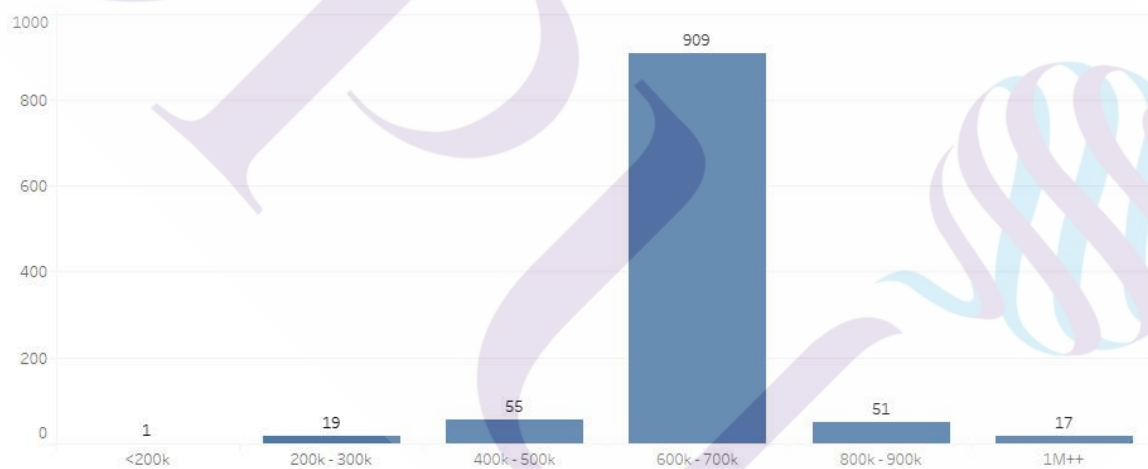


ภาพที่ 3.4 จำนวนภาพแยกตามความสูง ความกว้าง ของภาพชุดข้อมูล FlickrLogos-47

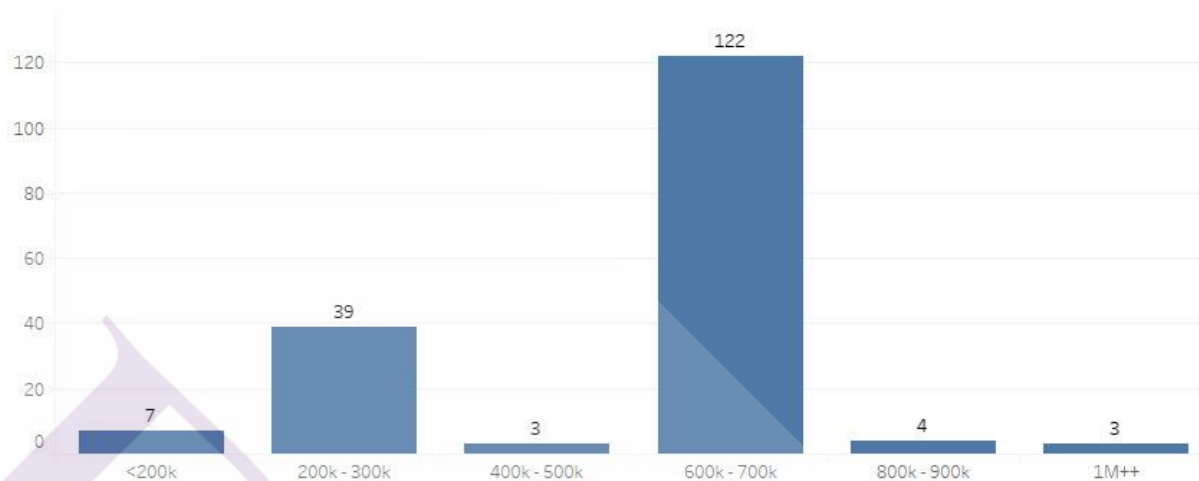
จำนวนรูปภาพแยกตามขนาดภาพของแต่ละชุดข้อมูล



ภาพที่ 3.5 จำนวนรูปภาพชุดข้อมูล Our Logos แยกตามขนาดภาพ (หน่วยเป็นพิกเซล)



ภาพที่ 3.6 จำนวนรูปภาพชุดข้อมูล Our Logos แยกตามขนาดภาพ (หน่วยเป็นพิกเซล)



ภาพที่ 3.7 จำนวนรูปภาพชุดข้อมูล FlickrLogo-47 แยกตามขนาดภาพ (หน่วยเป็นพิกเซล)

สัดส่วนขนาดโลโก้สินค้าต่อขนาดรูปภาพเมื่อจัดกลุ่ม 6 กลุ่ม ได้แก่

Ratio <1% สัดส่วนขนาดโลโก้สินค้าต่ำกว่า 1%

Ratio 1% - 2% สัดส่วนขนาดโลโก้สินค้าตั้งแต่ 1% ถึง 2%

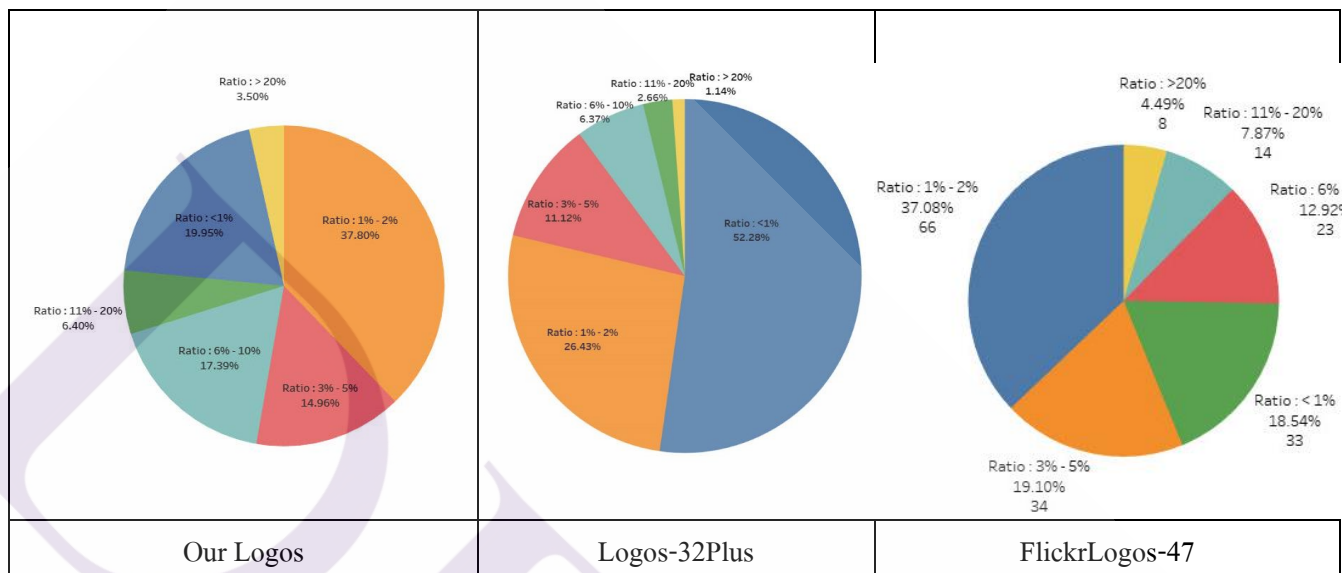
Ratio 3% - 5% สัดส่วนขนาดโลโก้สินค้าตั้งแต่ 3% ถึง 5%

Ratio 6% - 10% สัดส่วนขนาดโลโก้สินค้าตั้งแต่ 3% ถึง 5%

Ratio 11% - 20% สัดส่วนขนาดโลโก้สินค้าตั้งแต่ 3% ถึง 5%

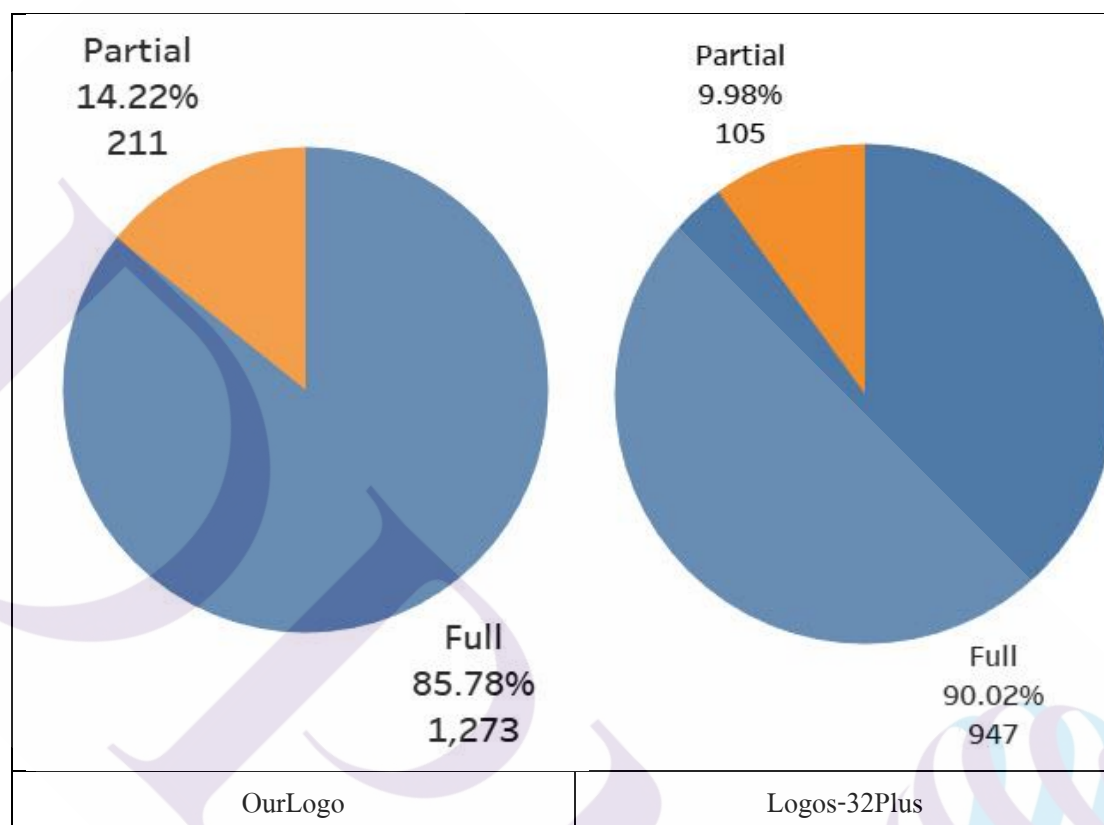
Ratio >20% สัดส่วนขนาดโลโก้สินค้ามากกว่า 20%

ตารางที่ 3.3 จำนวนภาพแยกตามสัดส่วนขนาดโลโก้สินค้าต่อขนาดรูปภาพชุดข้อมูล



ในชุดข้อมูลที่ใช้จะมีทั้งโลโก้แบบสมบูรณ์ก็คือเต็ม 100 เปอร์เซ็นต์และอีกส่วนหนึ่งที่โลโก้ขาดหายไปหรือไม่เต็ม 100 เปอร์เซ็นต์ ซึ่งสัดส่วนของภาพของแต่ละชุดข้อมูลแสดงในตารางที่ 3.4 แต่สำหรับชุดข้อมูลทดสอบของ FlickrLogos-47 นั้นไม่มีจำนวนรูปภาพไม่เต็ม 100%

ตารางที่ 3.4 จำนวนภาพแยกตามประเภทโลโก้ในภาพสมบูรณ์ และไม่สมบูรณ์ 100 เปอร์เซ็นต์



### 3.1.2 การจัดเตรียมข้อมูล

กระบวนการจัดเตรียมข้อมูลเป็นขั้นตอนการจัดเตรียมข้อมูลที่รวบรวมมาคัดแยกและทำการระบุประเภทและพื้นที่ของวัตถุในภาพ

#### 1. การคัดแยก

ข้อมูลที่ใช้งานจะถูกแยกเป็น 2 ส่วน คือ ส่วนที่ใช้ในการฝึกสอนโมเดล และส่วนที่ใช้ในการทดสอบโมเดล โดยจะใช้ข้อมูลจากชุดข้อมูล Logos-32Plus และข้อมูลที่เก็บรวบรวมจากเว็บไซต์และ

ตั้งคมออนไลน์ ซึ่งเมื่อรวมกันแล้วจะถูกแบ่งในสัดส่วนข้อมูลฝึกสอน 70 เปอร์เซ็นต์ และ 30 เปอร์เซ็นต์ สำหรับทดสอบจะแยกเป็นชุดข้อมูลเป็น 8 ชุด ได้แก่

ชุดที่ 1. ข้อมูลทดสอบปกติ เป็นชุดข้อมูลที่ถูกแบ่งออกจากชุดข้อมูล Logos-32Plus และข้อมูล Our Logos ในสัดส่วน 30 เปอร์เซ็นต์

ชุดที่ 2. ชุดข้อมูลทดสอบของ Logos-32Plus เป็นชุดข้อมูลทดสอบของข้อมูล Logos-32Plus

ชุดที่ 3. ชุดข้อมูลทดสอบของ FlickrLogos-47 เป็นชุดข้อมูลทดสอบของข้อมูล FlickrLogos-47

ชุดที่ 4. ชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ เป็นกลุ่มข้อมูลที่สำคัญมาจากชุดที่ 1. ทั้งหมด 3 ชุด ปรับคุณลักษณะภาพต่างกันโดยใช้ไลบรารี PIL ในภาษาไทยอน พารามิเตอร์ Brightness โดยชุดที่1 Brightness=1.7 ชุดที่2 Brightness =1.9 ชุดที่3 Brightness =2.1

ชุดที่ 5. ชุดข้อมูลกลุ่มแสงมืดกว่าปกติ เป็นกลุ่มข้อมูลที่สำคัญมาจากชุดที่ 1. ทั้งหมด 3 ชุด ปรับคุณลักษณะภาพต่างกันโดยใช้ไลบรารี PIL ในภาษาไทยอน พารามิเตอร์ Brightness โดยชุดที่1 Brightness=0.3 ชุดที่2 Brightness =0.5 ชุดที่3 Brightness =0.7

ชุดที่ 6. ชุดข้อมูลสิ่งรบกวนในภาพ เป็นกลุ่มข้อมูลที่สำคัญมาจากชุดที่ 1. ทั้งหมด 3 ชุด ปรับคุณลักษณะภาพต่างกันโดยใช้ไลบรารี PIL ในภาษาไทยอน พารามิเตอร์ Noise โดยชุดที่1 Noise=0.05 ชุดที่2 Noise=0.1 ชุดที่3 Noise =0.2

ชุดที่ 7. ชุดข้อมูลภาพไม่คมชัด 3 ชุด เป็นกลุ่มข้อมูลที่สำคัญมาจากชุดที่ 1. ทั้งหมด 3 ชุด ปรับคุณลักษณะภาพต่างกันโดยใช้ไลบรารี PIL ในภาษาไทยอน พารามิเตอร์ GaussianBlur ชุดที่1 GaussianBlur=2 ชุดที่2 GaussianBlur=3 ชุดที่3 GaussianBlur=5

ชุดที่ 8. ชุดข้อมูลโลโก้ไม่ครบ 100 เปอร์เซ็นต์ เป็นชุดข้อมูลที่สำคัญมาจากชุดที่ 1. แล้วทำการตัดภาพให้มีลักษณะโลโก้ไม่ครบ 100

ชุดที่ 9. เป็นกลุ่มข้อมูลที่สำคัญมาจากชุดที่ 1 และปรับสิ่งผิดปกติ 3 ชนิด 3 ชุด ได้แก่

ชุดที่ 1 Brightness 0.7 + Noise 0.05 + GaussianBlur 2 ชุดที่ 2 Brightness 0.5 + Noise 0.1 + GaussianBlur 3 ชุดที่ 3 Brightness 0.3 + Noise 0.2 + GaussianBlur 5

## 2. การระบุประเภทของวัตถุในภาพเพื่อใช้ในการสร้างโมเดลพยากรณ์

ขั้นตอนนี้เป็นการระบุว่าในรูปภาพที่ใช้ในการสร้างโมเดลนั้นมีวัตถุประเภทไหนบ้างและอยู่ตรงจุดไหนของรูปภาพ เป็นขั้นตอนที่ใช้เวลานานเนื่องจากต้องใช้แรงงานคน โดยวิธีการก็จะระบุว่าในรูปภาพนั้นตรงไหนคือโลโก้ของสินค้า ยกตัวอย่างเช่น



cocacola



starbucks



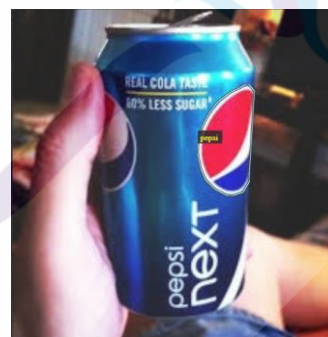
heineken



singha



leo



pepsi

ภาพที่ 3.8 ตัวอย่างการระบุประเภทวัตถุในภาพ

### 3.1.3 การเพิ่มชุดตัวอย่างข้อมูล (Augmentation)

การเพิ่มชุดตัวอย่างข้อมูลเป็นการเพิ่มรูปภาพในคุณลักษณะต่างๆเพื่อใช้ในการฝึกสอนโมเดล โดยการเพิ่มชุดตัวอย่างข้อมูลนี้จะใช้ไลบรารี imageaug ในภาษาไพธอน โดยตั้งค่าต่างๆดังนี้

#### การตั้งค่าในกลุ่มขนาดและตำแหน่ง

CropAndPad กำหนดค่า percent=-0.50 ถึง 0.50

Rotate กำหนดค่า 0 ถึง 180

Fliplr กำหนดค่า 1

Flipud กำหนดค่า 1

PiecewiseAffine กำหนดค่า scale 0.01 ถึง 0.1

#### การตั้งค่ากลุ่มภาพไม่ชัด

AverageBlur กำหนดค่า 0 ถึง 10

#### การตั้งค่ากลุ่มแสงและสี

GammaContrast กำหนดค่า 0.1 ถึง 2

Grayscale กำหนดค่า alpha 0.0 ถึง 1.0

Invert กำหนดค่า 0.25

#### การตั้งค่ากลุ่มสิ่งรบกวน

AdditiveGaussianNoise กำหนดค่า scale=0.05\*255, 0.2\*255, 0.1\*255, 0.15\*255

CoarseDropout กำหนดค่า 0.02, size\_percent=0.1, per\_channel=0.5

Dropout กำหนดค่า 0, 0.2, per\_channel=0.5



### 3.1.4 การสร้างโมเดล

รูปภาพที่มาจากสถานการณ์จริงในชีวิตประจำวันมีความหลากหลายทางคุณลักษณะมาก ในการสร้างโมเดลเพื่อที่จะทำนายผลหรือคัดแยกรูปภาพเหล่านั้นจึงจำเป็นต้องให้โมเดลมีความเสถียรภาพรองรับความหลากหลายเหล่านั้นรวมถึงความทนทานต่อสิ่งรบกวนในภาพ การปรับแต่ง หรือความไม่สมบูรณ์ต่างๆอีกด้วย

งานวิจัยนี้ใช้โมเดล Mask-RCNN ซึ่งเป็น โมเดลที่มีประสิทธิภาพในการทำนายรูปภาพ อีกทั้งยังมีคุณสมบัติในการแยกวัตถุที่สนใจในรูปภาพทำให้สามารถระบุ จำนวน ขนาดวัตถุ ซึ่งทำให้สามารถปรับปรุงแก้ไขประสิทธิภาพของโมเดลเพื่อให้รองรับคุณลักษณะที่หลากหลายได้ ในการสร้างโมเดลเองก็จะมีกรสร้าง โมเดลที่มีการกำหนดค่าที่แตกต่างกันและใช้ชุดข้อมูลแตกต่างกันเพื่อทดสอบประสิทธิภาพและหาตัวแปร ปัจจัยที่สำคัญในการกำหนดค่าเพื่อให้การสร้างโมเดลที่มีประสิทธิภาพสูงสุด

โมเดลที่ใช้ประกอบไปด้วย

1. โมเดล Mask I เป็น โมเดล Mask-RCNN ที่ฝึกสอนโดยชุดข้อมูล OpenLogos-32Plus+Our Logos ใช้วิธีการเพิ่มชุดตัวอย่างข้อมูล (Augmentation) แบบครบทุกกลุ่ม
2. โมเดล Mask II เป็น โมเดล Mask-RCNN ที่ฝึกสอนโดยชุดข้อมูล OpenLogos-32Plus+Our Logos ไม่ใช้วิธีการเพิ่มชุดตัวอย่างข้อมูล (Augmentation)
3. โมเดล Mask III เป็น โมเดล Mask-RCNN ที่ฝึกสอน โดยชุดข้อมูล OpenLogos-32Plus ใช้วิธีการเพิ่มชุดตัวอย่างข้อมูล (Augmentation) ครบทุกกลุ่ม
4. โมเดล Mask IV เป็น โมเดล Mask-RCNN ที่ฝึกสอน โดยชุดข้อมูล OpenLogos-32Plus ใช้วิธีการเพิ่มชุดตัวอย่างข้อมูล (Augmentation) ตั้งค่า Mini Mask ในการฝึกสอนเป็น  $56 \times 56$  พิกเซล

ค่าพื้นฐานของ โมเดลที่ใช้ จะใช้ Backbone resnet101 ขนาดภาพก่อนเข้าโมเดลเท่ากับ  $1024 \times 1024$  พิกเซล ซึ่งหากภาพใหญ่กว่าค่าที่กำหนดก็ทำการย่อภาพ และหากภาพมีขนาดเล็กกว่าค่ากำหนด

จะทำการเพิ่มขอบของภาพ (Padding) เพื่อให้ขนาดภาพเท่ากับค่าที่กำหนด ภาพที่ใช้เป็นภาพสี ใช้ค่าความมั่นใจขั้นต่ำ (Min Confidence) ที่ 0.7 สำหรับโมเดลที่ใช้ mini mask จะใช้ขนาด 56\*56 พิกเซล

ตารางที่ 3.5 แสดงโมเดล ข้อมูลฝึกสอนและการเพิ่มชุดข้อมูลฝึกสอน

Model	Dataset	Mini Mask	การเพิ่มชุดข้อมูล (Augmentation)			
			ตำแหน่งและขนาด (Size and Position)	แสงและสี (color and brightness)	ไม่คมชัด (Blur)	สิ่งรบกวน (Noise)
Mask I	Logos-32 + Our Logos	✗	✓	✓	✓	✓
Mask II	Logos-32 + Our Logos	✗	✗	✗	✗	✗
Mask III	Logos-32	✗	✓	✓	✓	✓
Mask IV	Logos-32	✓	✓	✓	✓	✓

### 3.1.5 การทดสอบประสิทธิภาพ

ในการทดสอบประสิทธิภาพ โมเดลประกอบไปด้วยการวัดความถูกต้องของวัตถุที่ทำนาย จะวัดจากค่า IoU (Intersection over union) ที่มากกว่า 50 เปอร์เซ็นต์ นั่นหมายความว่าพื้นที่วัตถุที่ทำนาย

ออกมาเมื่อเทียบกับพื้นที่ของวัตถุจริง ต้องมีค่ามากกว่า 50 เปอร์เซ็นต์ และผลที่ได้จะไปคำนวณค่า precision และ recall เพื่อคำนวณหาค่า mAP ซึ่งเป็นค่าที่ใช้วัดประสิทธิภาพของโมเดล

ในการทดสอบประสิทธิภาพของโมเดลทั้งหมดจะทดสอบโดยใช้ชุดข้อมูลทดสอบทั้งหมด 8 กลุ่ม ได้แก่

- ชุดที่ 1. ข้อมูลทดสอบปกติ (OurLogo + Logos-32)
- ชุดที่ 2. ชุดข้อมูลทดสอบของ Logos-32
- ชุดที่ 3. ชุดข้อมูลทดสอบของ FlickrLogos-47
- ชุดที่ 4. ชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ 3 ชุด (Brightness 1.7, 1.9, 2.1)
- ชุดที่ 5. ชุดข้อมูลกลุ่มแสงมืดกว่าปกติ 3 ชุด (Brightness 0.3, 0.5, 0.7)
- ชุดที่ 6. ชุดข้อมูลสิ่งรบกวนในภาพ 3 ชุด (Noise 0.05, 0.1, 0.2)
- ชุดที่ 7. ชุดข้อมูลภาพไม่คมชัด 3 ชุด (GaussianBlur 2, 3, 5)
- ชุดที่ 8. ชุดข้อมูลโลโก้ไม่สมบูรณ์ครบ 100%
- ชุดที่ 9. ข้อมูลสิ่งผิดปกติ 3 ชนิด 3 ชุด (Brightness 0.7 + Noise 0.05 + GaussianBlur 2 ,  
Brightness 0.5 + Noise 0.1 + GaussianBlur 3, Brightness 0.3 + Noise 0.2 + GaussianBlur 5)

## บทที่ 4

### ผลงานวิจัย

ในบทนี้จะเป็นผลการทดลองในการตรวจจับโลโก้โดยจะแยกแต่ละโมเดล และชุดข้อมูลที่  
ใช้สำหรับทดสอบเพื่อเปรียบเทียบประสิทธิภาพของแต่ละโมเดลซึ่งชุดข้อมูลที่ได้แก่

- ชุดที่ 1. ข้อมูลทดสอบปกติ (OurLogo + Logos-32Plus)
- ชุดที่ 2. ชุดข้อมูลทดสอบของ Logos-32
- ชุดที่ 3. ชุดข้อมูลทดสอบของ FlickrLogos-47
- ชุดที่ 4. ชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ 3 ชุด (Brightness 1.7, 1.9, 2.1)
- ชุดที่ 5. ชุดข้อมูลกลุ่มแสงมืดกว่าปกติ 3 ชุด (Brightness 0.3, 0.5, 0.7)
- ชุดที่ 6. ชุดข้อมูลสิ่งรบกวนในภาพ 3 ชุด (Noise 0.05, 0.1, 0.2)
- ชุดที่ 7. ชุดข้อมูลภาพไม่คมชัด 3 ชุด (GaussianBlur 2, 3, 5)
- ชุดที่ 8. ชุดข้อมูลโลโก้ไม่ครบ 100%
- ชุดที่ 9. ข้อมูลสิ่งผิดปกติมากกว่า 1 อย่าง 3 ชุด (Brightness 0.7 + Noise 0.05 +  
GaussianBlur 2 , Brightness 0.5 + Noise 0.1 + GaussianBlur 3, Brightness 0.3 + Noise 0.2  
+ GaussianBlur 5)

#### 4.1 ผลการทดสอบประสิทธิภาพและอภิปรายผลแยกตามชุดข้อมูล

แสดงผลการทดสอบโมเดลแยกตามชุดข้อมูลที่ใช้ทดสอบ ซึ่งโมเดลทั้ง 4 โมเดล ประกอบ  
ไปด้วย โมเดล Mask I ฝึกสอนด้วยชุดข้อมูล Our Logo + Logos-32Plus และ เพิ่มชุดข้อมูล  
(Augmentation) ในการฝึกสอน

โมเดล Mask II ฝึกสอนด้วยชุดข้อมูล Our Logo + Logos-32Plus และ ไม่เพิ่มชุดข้อมูล  
(Augmentation) ในการฝึกสอน

โมเดล Mask II ฝึกสอนด้วยชุดข้อมูล Logos-32Plus และ เพิ่มชุดข้อมูล (Augmentation) ในการฝึกสอน โมเดล Mask II ฝึกสอนด้วยชุดข้อมูล Logos-32Plus และ เพิ่มชุดข้อมูล (Augmentation) ในการฝึกสอน และกำหนดค่า Mini Mask ของโมเดล

#### 4.1.1 ข้อมูลทดสอบปกติ (Our Logo + Logos-32Plus)

ตารางที่ 4.1 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลปกติ (Our Logo + Logos-32)

โมเดล	mAP
Mask I	0.982
Mask II	0.958
Mask III	0.966
Mask IV	0.770

ในชุดข้อมูลทดสอบปกติโมเดลที่มีประสิทธิภาพสูงสุดคือ Mask I ส่วนในโมเดล Mask II แม้ว่าไม่ได้ทำการเพิ่มชุดจำนวนฝึกสอน (Augmentation) ก็มีประสิทธิภาพไม่ต่างกันมาก รวมถึงโมเดล Mask III ที่ฝึกสอนด้วยชุดข้อมูลต่างกันด้วย มีเพียงแต่โมเดล Mask IV ที่การตั้งค่าโมเดลใช้ Mini Mask อาจส่งผลให้ประสิทธิภาพต่ำที่สุด

#### 4.1.2 ข้อมูลทดสอบ Logos-32Plus

ตารางที่ 4.2 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูล Logos-32Plus

โมเดล	mAP
Mask I	0.980
Mask II	0.952
Mask III	0.975
Mask IV	0.809

### 4.1.3 ข้อมูลทดสอบ FlickrLogos-47

ตารางที่ 4.3 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูล FlickrLogos-47

โมเดล	mAP
Mask I	0.980
Mask II	0.952
Mask III	0.930
Mask IV	0.735

ประสิทธิภาพของโมเดลที่ทดสอบด้วยชุดข้อมูลทดสอบ Logos-32Plus (ในตารางที่ 4.2) และ FlickrLogos-47(ในตารางที่ 4.3) ผลที่ได้จะใกล้เคียงกับการทดสอบชุดข้อมูลปกติ

### 4.1.4 ข้อมูลทดสอบ ชุดข้อมูลโลโก้ไม่ครบ100%

ตารางที่ 4.4 ประสิทธิภาพโมเดลทดสอบกับชุดข้อมูลโลโก้ไม่ครบ100%

โมเดล	mAP
Mask I	0.970
Mask II	0.840
Mask III	0.920
Mask IV	0.637

ในชุดข้อมูลทดสอบโลโก้ไม่ครบ 100% โมเดล Mask I ยังมีประสิทธิภาพดีที่สุด และ โมเดล Mask IV ยังมีประสิทธิภาพต่ำที่สุด แต่ค่า mAP ของทุกๆ โมเดลนั้นต่ำลงเมื่อเทียบกับ ชุดข้อมูลทดสอบปกติ

#### 4.1.5 ข้อมูลทดสอบชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ

ตารางที่ 4.5 ประสิทธิภาพ โมเดลทดสอบกับชุดข้อมูลกลุ่มแสงสว่างมากกว่าปกติ

โมเดล	mAP		
	Brightness 1.7	Brightness 1.9	Brightness 2.1
Mask I	0.976	0.975	0.975
Mask II	0.932	0.900	0.860
Mask III	0.929	0.906	0.902
Mask IV	0.709	0.704	0.646

#### 4.1.6 ข้อมูลทดสอบชุดข้อมูลกลุ่มแสงสว่างมีดกว่าปกติ

ตารางที่ 4.6 ประสิทธิภาพ โมเดลทดสอบกับชุดข้อมูลกลุ่มแสงมีดกว่าปกติ

โมเดล	mAP		
	Brightness 0.3	Brightness 0.5	Brightness 0.7
Mask I	0.930	0.970	0.974
Mask II	0.696	0.906	0.94
Mask III	0.834	0.938	0.958
Mask IV	0.520	0.700	0.706

#### 4.1.7 ข้อมูลทดสอบชุดข้อมูลสิ่งรบกวนในภาพ

ตารางที่ 4.7 ประสิทธิภาพ โมเดลทดสอบกับชุดข้อมูลสิ่งรบกวนในภาพ

โมเดล	mAP		
	Noise 0.05	Noise 0.1	Noise 0.2
Mask I	<b>0.980</b>	<b>0.960</b>	<b>0.946</b>
Mask II	0.929	0.852	0.662
Mask III	0.942	0.92	0.84
Mask IV	0.720	0.662	0.504

#### 4.1.8 ข้อมูลทดสอบชุดข้อมูลภาพไม่คมชัด

ตารางที่ 4.8 ประสิทธิภาพ โมเดลทดสอบกับชุดข้อมูลภาพไม่คมชัด

โมเดล	mAP		
	GaussianBlur 2	GaussianBlur 3	GaussianBlur 5
Mask I	<b>0.980</b>	<b>0.970</b>	<b>0.959</b>
Mask II	0.932	0.901	0.746
Mask III	0.953	0.911	0.762
Mask IV	0.725	0.652	0.490

ชุดข้อมูลภาพไม่คมชัดเป็นชุดข้อมูลที่โมเดลมีประสิทธิภาพต่ำที่สุดเมื่อเปรียบเทียบกับผลของทุกๆชุดข้อมูลทดสอบ โดยเฉพาะ GaussianBlur 5 ซึ่งเป็นข้อมูลที่มีสิ่งรบกวนมากที่สุด



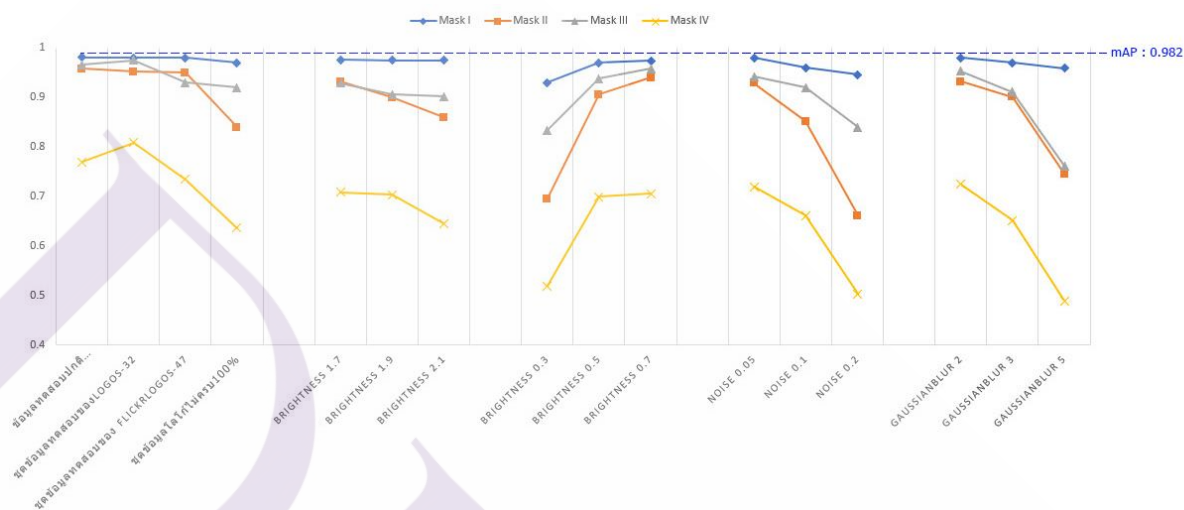
#### 4.1.9 ข้อมูลชุดทดสอบสิ่งผิดปกติมากกว่า 3 ชนิด

ตารางที่ 4.9 ประสิทธิภาพ โมเดลทดสอบกับชุดข้อมูลชุดทดสอบสิ่งผิดปกติ 3 ชนิด

โมเดล	mAP		
	Brightness 0.7 + Noise 0.05 + GaussianBlur 2	Brightness 0.5 + Noise 0.1 + GaussianBlur 3	Brightness 0.3 + Noise 0.2 + GaussianBlur 5
Mask I	<b>0.94</b>	<b>0.835</b>	<b>0.675</b>

ชุดข้อมูลสิ่งผิดปกติมากกว่า 3 ชนิด ซึ่งเป็นภาพที่มีคุณภาพต่ำมากโดยเฉพาะอย่างยิ่งเมื่อใช้ Brightness 0.3 + Noise 0.2 + GaussianBlur 5 นั่นคือเป็นภาพที่แสงน้อย มีสิ่งรบกวนในภาพสูง และความคมชัดต่ำ จึงทำให้ประสิทธิภาพของโมเดลต่ำลงอย่างมาก

## 4.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดลโดยภาพรวม



ภาพที่ 4.1 เปรียบเทียบประสิทธิภาพของ โมเดลโดยรวม

เมื่อเปรียบเทียบประสิทธิภาพของ โมเดลจะพบว่า โมเดลที่มีประสิทธิภาพดีที่สุดคือ โมเดล Mask I ซึ่งเป็น โมเดลที่ฝึกสอนด้วยชุดข้อมูล Our Logos ร่วมกับชุดข้อมูล Logos-32Plus รวมไปถึงวิธีการเพิ่มชุดข้อมูลฝึกสอน (Augmentation) ที่กำหนด โดยมีค่า mAP สูงที่สุดคือ 0.982 และประสิทธิภาพยังคงที่ที่สุดในการทดสอบในชุดข้อมูลทั้งหมด โมเดลที่ประสิทธิภาพต่ำสุดคือ โมเดล Mask IV ส่วนในโมเดล Mask II และ Mask III ถึงแม้ว่าจะมีประสิทธิภาพสูงในชุดข้อมูลทดสอบปกติ แต่เมื่อทดสอบกับชุดข้อมูลที่ผิดปกติประสิทธิภาพกลับต่ำลงอย่างมาก โดยเฉพาะเมื่อทดสอบกับชุดข้อมูลที่มีแสงสว่างที่ต่ำ มีสิ่งรบกวนในภาพสูง และเมื่อพบกับรูปภาพที่คมชัดน้อย

ถึงแม้ว่าใน โมเดล Mask I จะเป็น โมเดลที่มีประสิทธิภาพสูงที่สุด แต่เมื่อทำการทดสอบกับชุดข้อมูลที่มีการปรับสิ่งผิดปกติในภาพ 3 ปัจจัยพร้อมกันซึ่งจะพบว่า เมื่อปรับภาพให้มีแสงน้อย มีสิ่งรบกวนในภาพสูงและมีความคมชัดของภาพต่ำ ประสิทธิภาพของ โมเดลจะลดลงอย่างมาก ตามตารางที่ 4.8

## บทที่ 5

### สรุปผลและข้อเสนอแนะ

ในบทนี้ประกอบไปด้วยหัวข้อหลัก 2 หัวข้อได้แก่

#### 5.1 สรุปผลงานวิจัย

5.2 ข้อเสนอแนะ เพื่อสำหรับแนวทางในการพัฒนาต่อของงานวิจัยนี้ให้มีประสิทธิภาพมากยิ่งขึ้น

#### 5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอวิธีการเลือกข้อมูล การจัดเตรียมข้อมูล และการเลือกใช้โมเดลที่มีประสิทธิภาพนั่นคือ Mask R-CNN ที่มีการกำหนดค่า Mini Mask = false ทำให้โมเดลมีประสิทธิภาพดีกว่าโมเดลที่กำหนดค่า Mini Mask ดังจะเห็นได้จากประสิทธิภาพของโมเดล Mask IV ซึ่งเพียงแค่นำค่า Mini Mask ก็ทำให้ผลลัพธ์ที่ได้ลดลงอย่างมาก เมื่อเทียบกับ โมเดล Mask III ที่ฝึกสอนด้วยชุดข้อมูลเดียวกัน

โมเดลที่มีประสิทธิภาพมากที่สุดคือ โมเดล Mask I ที่มีค่า mAP 0.982 และยังคงมีประสิทธิภาพได้ดีถึงแม้ว่าจะทดสอบกับข้อมูลที่มีคุณลักษณะแตกต่างกันไปอย่างมาก ซึ่งวิธีการที่ใช้ในการพัฒนาโมเดล Mask I ที่มีประสิทธิภาพสามารถแก้ปัญหาต่างๆ ได้ดี สามารถสรุปได้ดังนี้

##### 5.1.1 ปัญหาภาพเล็ก และความละเอียดต่ำ

การแก้ปัญหานี้ใช้วิธีการเลือกใช้ชุดข้อมูลเพื่อฝึกสอนให้มีขนาดหลากหลาย นั่นคือการเลือกใช้ชุดข้อมูล Logos-32Plus ที่มีขนาดใหญ่และความละเอียดสูงซึ่งมีความ กว้าง สูง เริ่มต้นที่ 240 พิกเซล ถึง 1024 พิกเซล ค่าเฉลี่ยความกว้าง สูง อยู่ที่ 871 พิกเซล นำมาใช้ร่วมกับชุดข้อมูลที่ผู้วิจัยเก็บรวบรวมมา (Our Logos) ที่มีความกว้าง สูง เริ่มตั้งแต่ 70-960 พิกเซล และเฉลี่ยอยู่ที่ 298 ซึ่งจะเห็นว่า

เมื่อนำชุดข้อมูลทั้ง 2 มาใช้จะได้ทั้งข้อมูลที่มีขนาดใหญ่ความละเอียดสูง และข้อมูลที่มีขนาดเล็กความละเอียดต่ำใช้ในการฝึกสอน โมเดล

### 5.1.2 ปัญหาภาพความสว่างหรือมืด

การแก้ปัญหาที่นี้นอกจากจะใช้ข้อมูลฝึกสอนที่มีภาพกลางแจ้งที่มีความสว่างมากและภาพในอาคารและกลางคืนที่มีความมืดแต่ก็ไม่เพียงพอจึงจำเป็นต้องใช้วิธีการเพิ่มชุดข้อมูลฝึกสอน (Augmentation) โดยกำหนดตัวแปรดังนี้

- GammaContrast กำหนดค่า 0.1 ถึง 2
- Grayscale กำหนดค่า alpha 0.0 ถึง 1.0
- Invert กำหนดค่า 0.25

### 5.1.3 ปัญหาสิ่งรบกวนในภาพ และภาพไม่คมชัด

วิธีการแก้ปัญหานี้ใช้วิธีการเพิ่มชุดข้อมูลฝึกสอน (Augmentation) โดยกำหนดตัวแปรดังนี้

- AgerageBlur กำหนดค่า จาก 0 ถึง 10
- AdditiveGaussianNoise กำหนดค่า scale=0.05\*255, 0.2\*255, 0.1\*255, 0.15\*255
- CoarseDropout กำหนดค่า 0.02, size\_percent=0.1, per\_channel=0.5
- Dropout กำหนดค่า 0, 0.2, per\_channel=0.5

### 5.1.4 ปัญหาโลโก้ในภาพไม่สมบูรณ์

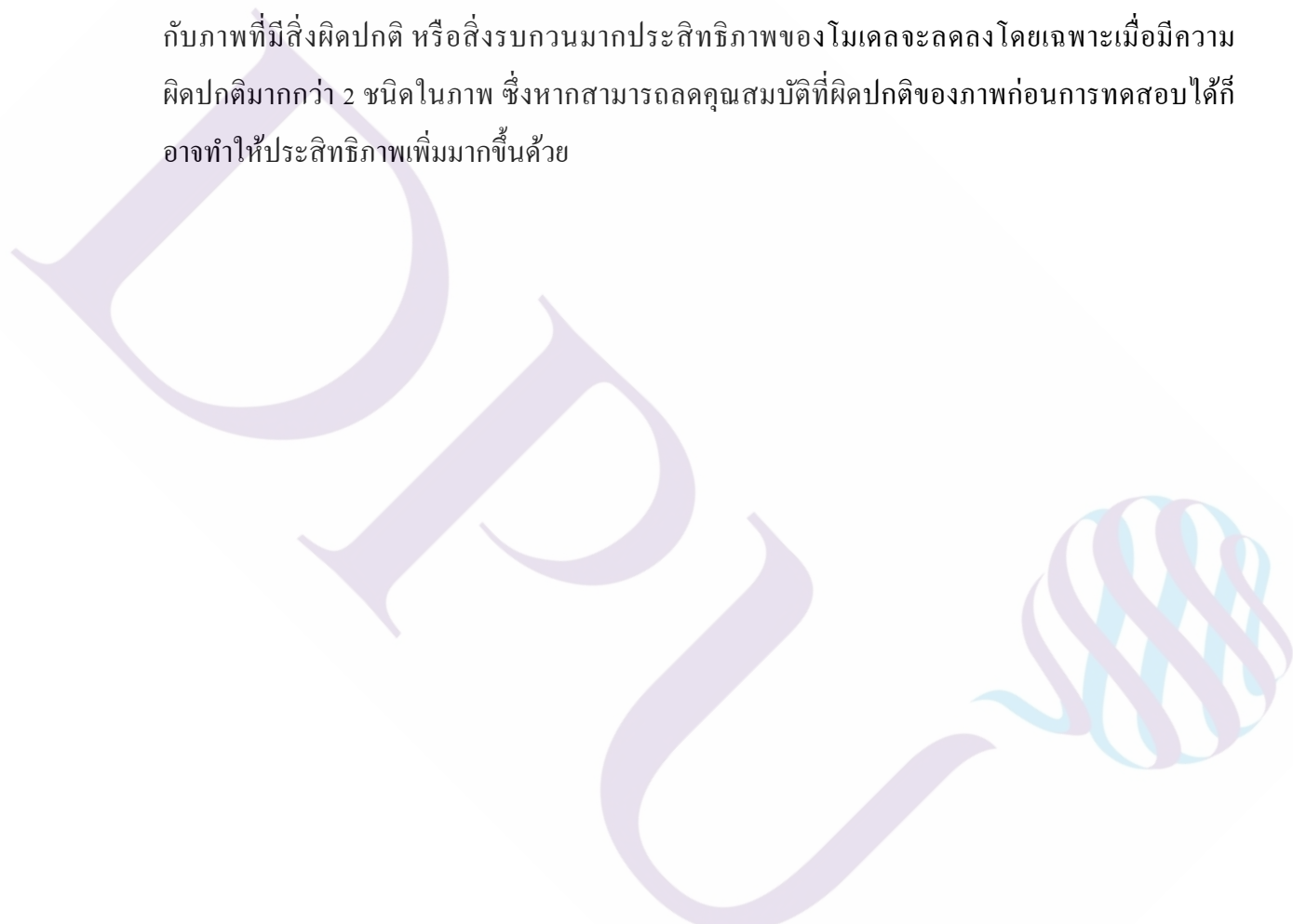
ใช้วิธีการเลือกภาพที่มีโลโก้ไม่สมบูรณ์มาฝึกสอนด้วยรวมถึงการใช้การเพิ่มชุดข้อมูลฝึกสอน (Augmentation) ร่วมด้วยโดยกำหนดค่าตัวแปรดังนี้

- CropAndPad กำหนดค่า percent=-0.50 ถึง 0.50
- PiecewiseAffine กำหนดค่า scale 0.01 ถึง 0.1
- Flipud กำหนดค่า 1
- Fliplr กำหนดค่า 1
- Rotate กำหนดค่า 0 ถึง 180

## 5.2 ข้อเสนอแนะ

การทำงานกับชุดข้อมูลรูปภาพนั้นหากไม่วางแผนในการเก็บข้อมูล และการจัดการข้อมูลที่ดี การเพิ่ม ลดหรือการแบ่งชุดข้อมูลอาจมีปัญหาในเรื่องข้อมูลซ้ำซ้อน อาจใช้วิธีแปลงข้อมูลให้อยู่ในรูปแบบอาร์เรย์ ซึ่งจะสามารถบริหารจัดการได้ดีกว่า

ถึงแม้ว่าวิธีการที่นำเสนอจะมีประสิทธิภาพในการตรวจจับโลโก้ในภาพที่ดี แต่เมื่อทดสอบกับภาพที่มีสิ่งผิดปกติ หรือสิ่งรบกวนมากประสิทธิภาพของโมเดลจะลดลงโดยเฉพาะเมื่อมีความผิดปกติมากกว่า 2 ชนิดในภาพ ซึ่งหากสามารถลดคุณสมบัติที่ผิดปกติของภาพก่อนการทดสอบได้ก็อาจทำให้ประสิทธิภาพเพิ่มมากขึ้นด้วย





**บรรณานุกรม**

## บรรณานุกรม

- Steven C.H. Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang , Hui Xue, and Qiang Wu. (2015). "LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks," arXiv.
- Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol.(2011). "Scalable Logo Recognition in Real-World Images," ICMR.
- Jia-Han Chiam, "Brand logo classification", Stanford University, unpublished.
- Forrest N. Iandola, Anting Shen, Peter Gao and Kurt Keutzer.(2015). "DeepLogo: Hitting Logo Recognition with the Deep Neural Network Hammer," arXiv .
- Raluca Boia, Alessandra Bandrabur, and Corneliu Florea.(2014). "Local description using multi-scale complete rank transform for improved logo recognition," Communications (COMM).
- Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell.(2013). "Decaf: A deep convolutional activation feature for generic visual recognition," arXiv:1310.1531.
- Stefan Duffner, and Christophe Garcia.(2006). "A neural scheme for robust detection of transparent logos in tv programs," International Conference on Artificial Neural Networks (ICANN).
- E. Francesconi, Paolo Frasconi, Marco Gori, S. Marinai, J.Q. Sheng, G. Soda, and Alessandro Sperduti.(1998). "Logo recognition by recursive neural networks," In Graphics Recognition Algorithms and Systems.
- Yannis Kalantidis, Lluis G. Pueyo, Michele Trevisiol, Roelof van Zwol, and Yannis Avrithis. (2011). "Scalable triangulation-based logo recognition," ICMR'11.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E.Hinton. (2012). "ImageNet Classification with Deep Convolutional Neural Networks," International Conference on Neural Information Processing Systems (NIPS), pp.1097-1105.

- Karen Simonyan and Andrew Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, abs/1409.1556.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich.(2014) "Going deeper with convolutions." arXiv:1409.4842.
- Guangyu Zhu and David Doermann. (2007). "Automatic document logo detection," ICDAR, September.
- A. Joly and O. Buisson.(2009). "Logo retrieval with a contrario visual query expansion," ACM International Conference on Multimedia, , pp. 581–584.
- Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. (2018)," Mask R-CNN" , arXiv:1703.06870v3 [cs.CV] .
- S. Ren, K. He, R. Girshick, and J. Sun. (2017) . "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In IEEE Trans. Pattern Anal. Mach. Intell. 39(6): 1137–1149 .
- N. Dalal and B. Triggs. (2005). "Histograms of oriented gradients for human detection," in International Conference on Computer Vision & Pattern Recognition, pp. 886–893.
- Stefan Romberg," From Local Features To Local Regions",PP.-841-845
- Nabeel Younus Khan,Brendan McCane,Geoff Wyvill. (2005). "SIFT and SURF Performance Evaluation Against Various Image Deformations on Benchmark Dataset", International Conference on Digital Image Computing: Techniques and Applications ,PP.501-506.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.(2015). "Deep Residual Learning for Image Recognition",arXiv:1512.03385 [cs.CV].
- Olga Russakovsky\*, Jia Deng\*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (2015). (\* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision.



Intersection over Union (IoU) for object detection,

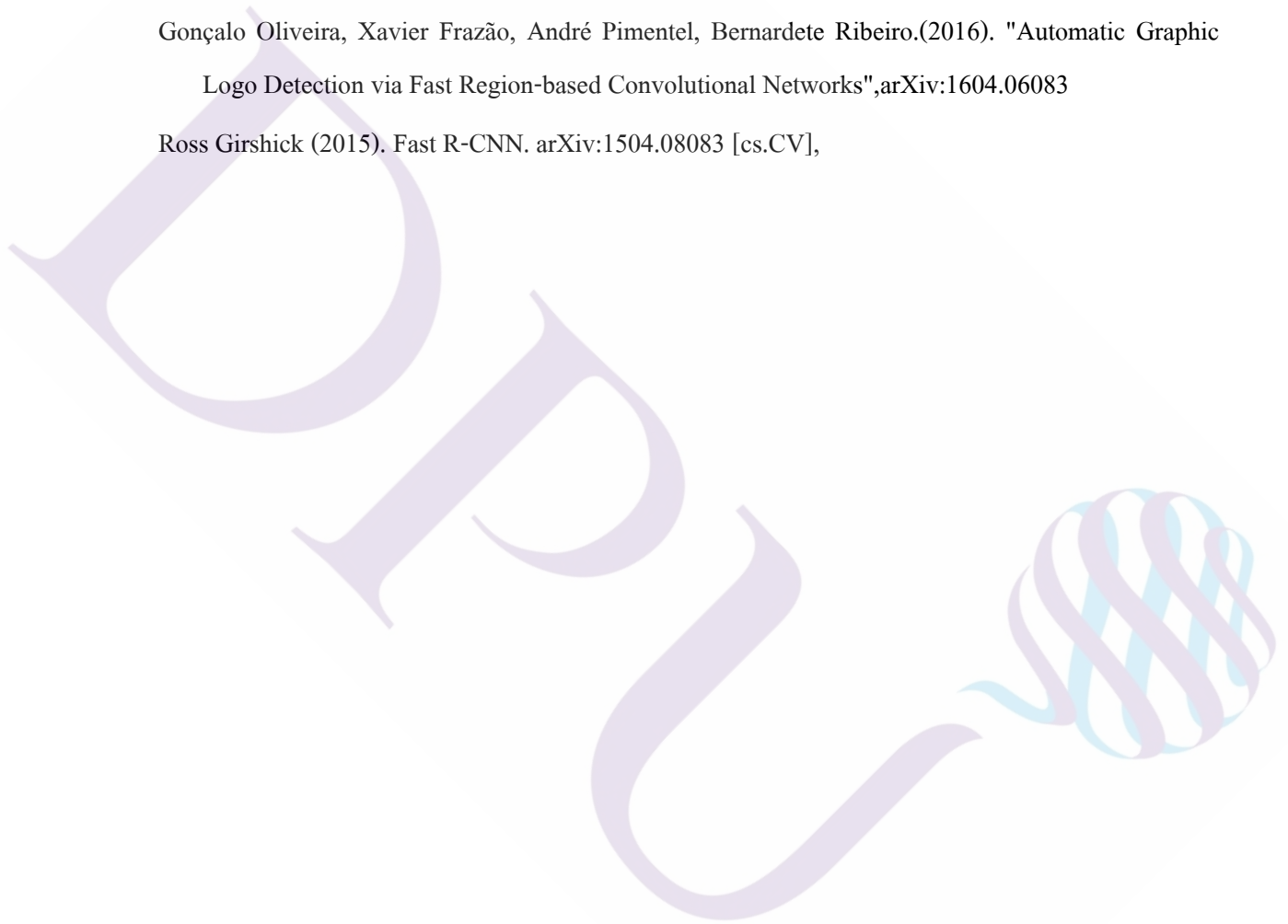
<https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>

Confusion matrix, [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix)

Simone Bianco, Marco Buzzelli, Davide Mazzini, Raimondo Schettini.(2017). "Deep Learning for Logo Recognition", arXiv:1701.02620 [cs.CV]

Gonçalo Oliveira, Xavier Frazão, André Pimentel, Bernardete Ribeiro.(2016). "Automatic Graphic Logo Detection via Fast Region-based Convolutional Networks",arXiv:1604.06083

Ross Girshick (2015). Fast R-CNN. arXiv:1504.08083 [cs.CV],





ภาคผนวก

### 6.1 ตัวอย่างชุดข้อมูลสำหรับทดสอบ

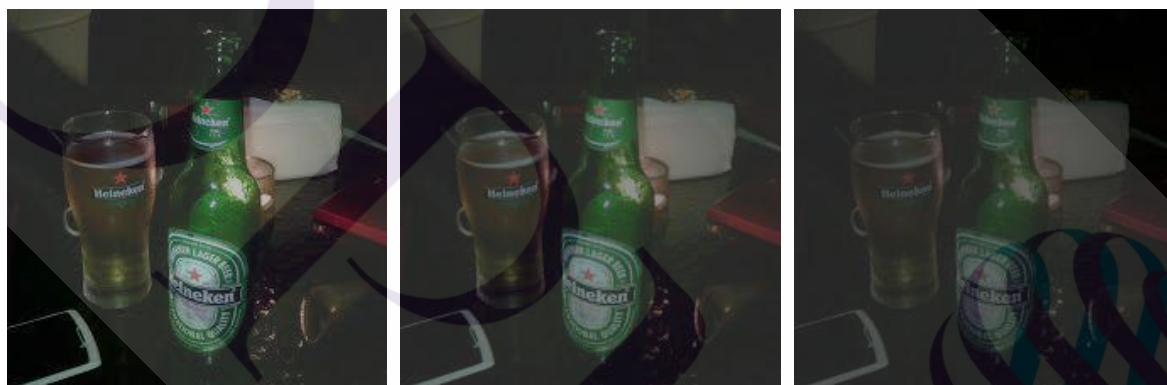


Brightness=1.7

Brightness =1.9

Brightness =2.1

รูปที่ 6.1 ตัวอย่างข้อมูลกลุ่มแสงสว่างมากกว่าปกติ



Brightness =0.7

Brightness =0.5

Brightness=0.3

รูปที่ 6.2 ตัวอย่างข้อมูลกลุ่มแสงมืดกว่าปกติ



Noise=0.05

Noise=0.1

Noise =0.2

รูปที่ 6.3 ตัวอย่างข้อมูลสิ่งรบกวนในภาพ



GaussianBlur=2

GaussianBlur=3

GaussianBlur=5

รูปที่ 6.4 ตัวอย่างข้อมูลกลุ่มภาพไม่คมชัด



รูปที่ 6.5 ตัวอย่างข้อมูลกลุ่มโลโก้ไม่ครบ 100%



Brightness 0.7 + Noise 0.05 +  
GaussianBlur 2

Brightness 0.5 + Noise 0.1 +  
GaussianBlur 3

Brightness 0.3 + Noise 0.2 +  
GaussianBlur 5

รูปภาพที่ 6.6 ตัวอย่างข้อมูลสิ่งผิดปกติ 3 ชนิด

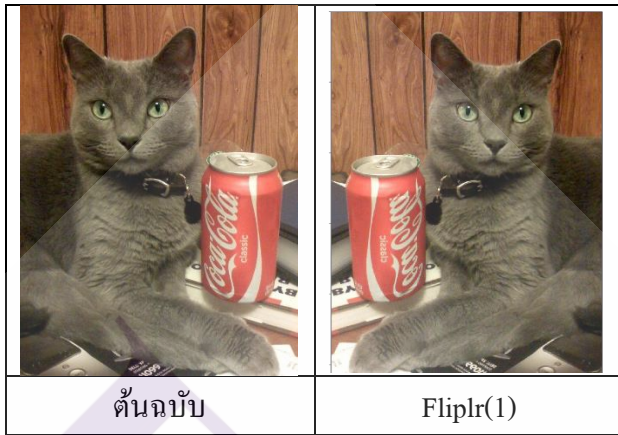
6.2 ตัวอย่างการเพิ่มชุดตัวอย่างข้อมูล (Augmentation)

ต้นฉบับ	percent=(-0.050, 0.20)	percent=(-0.30, 0.30)	percent=(-0.10, 0.50)

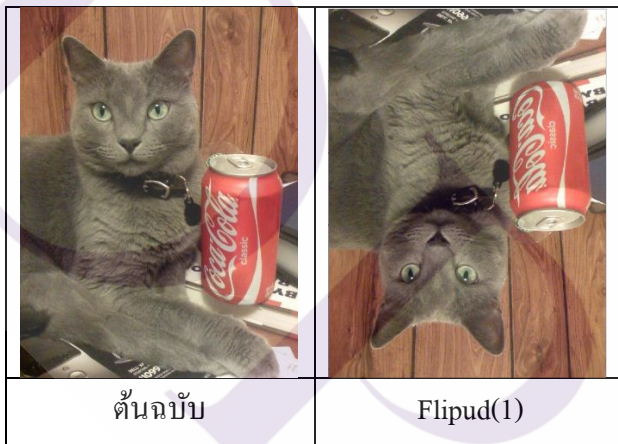
ตารางที่ 6.1 ตัวอย่างการ CropAndPad

ต้นฉบับ	rotate=(60,60)	rotate=(90,90)	rotate=(150,150)

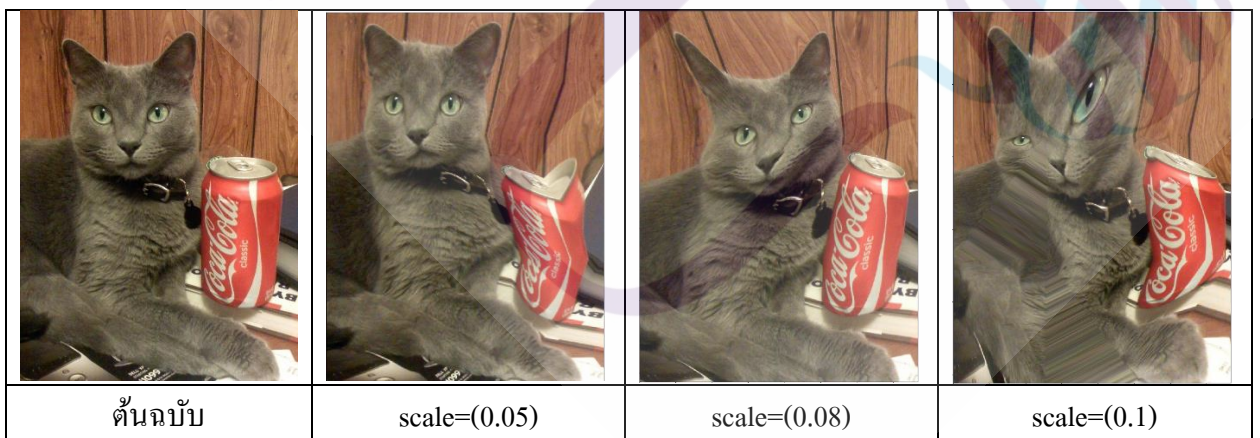
ตารางที่ 6.2 ตัวอย่างการ Affine rotate



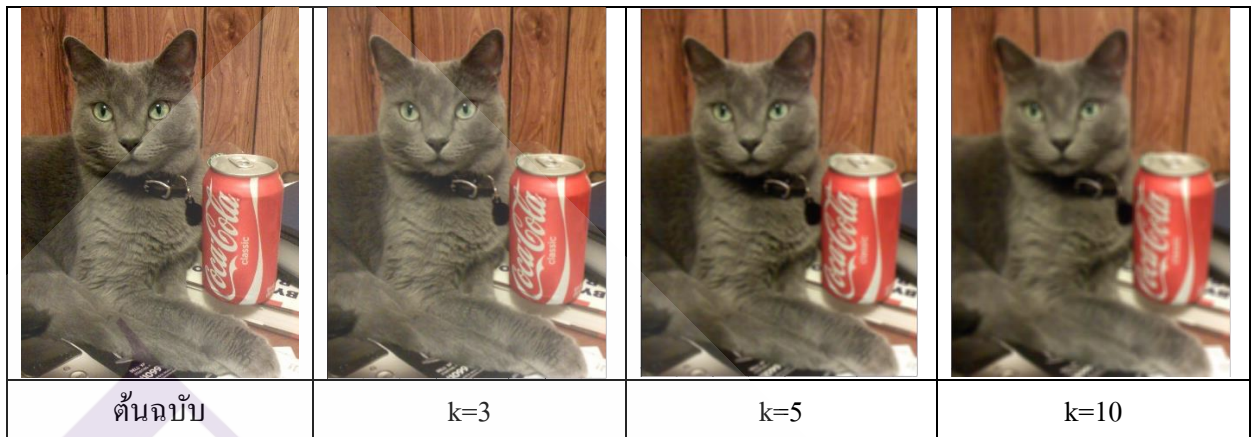
ตารางที่ 6.3 ตัวอย่างการ Fliplr



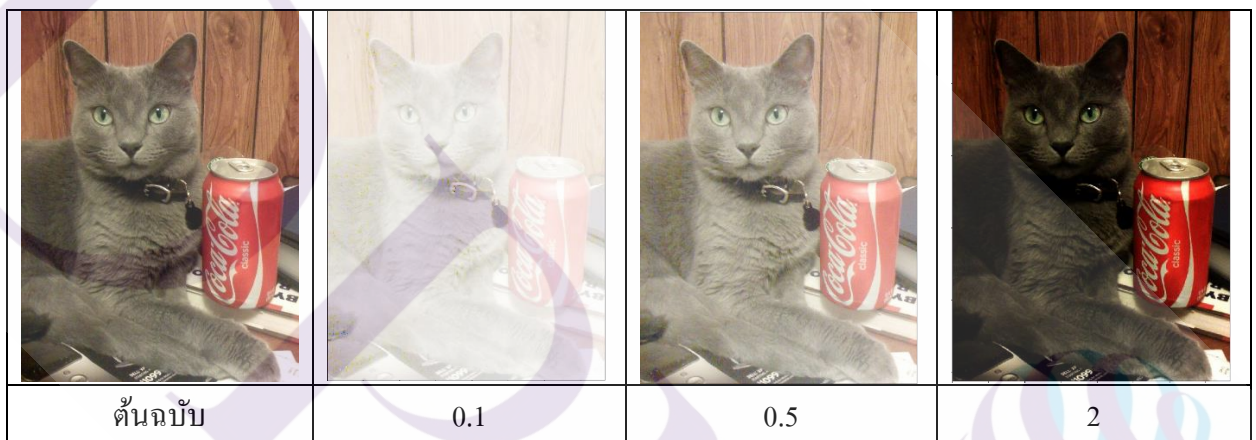
ตารางที่ 6.4 ตัวอย่างการ Flipud



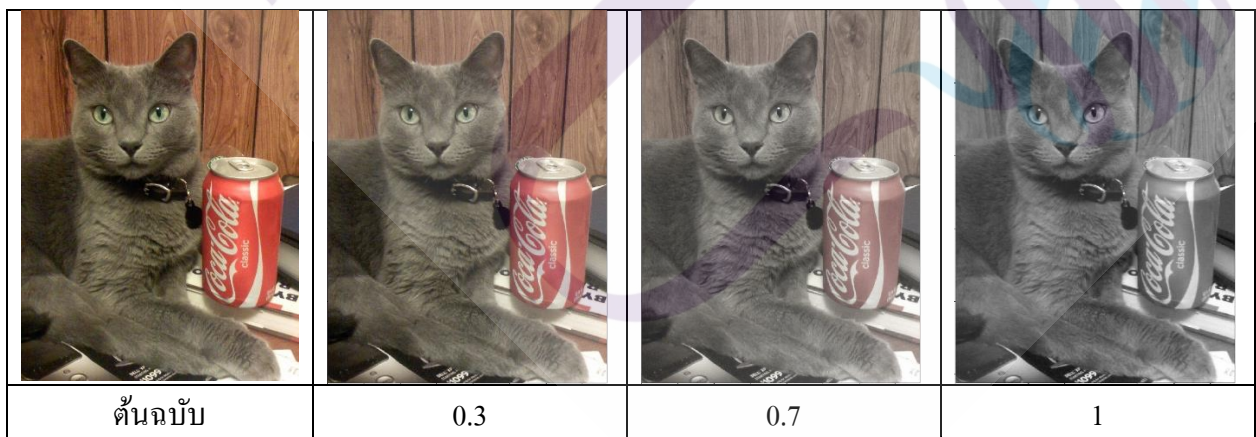
ตารางที่ 6.5 ตัวอย่างการ PiecewiseAffine



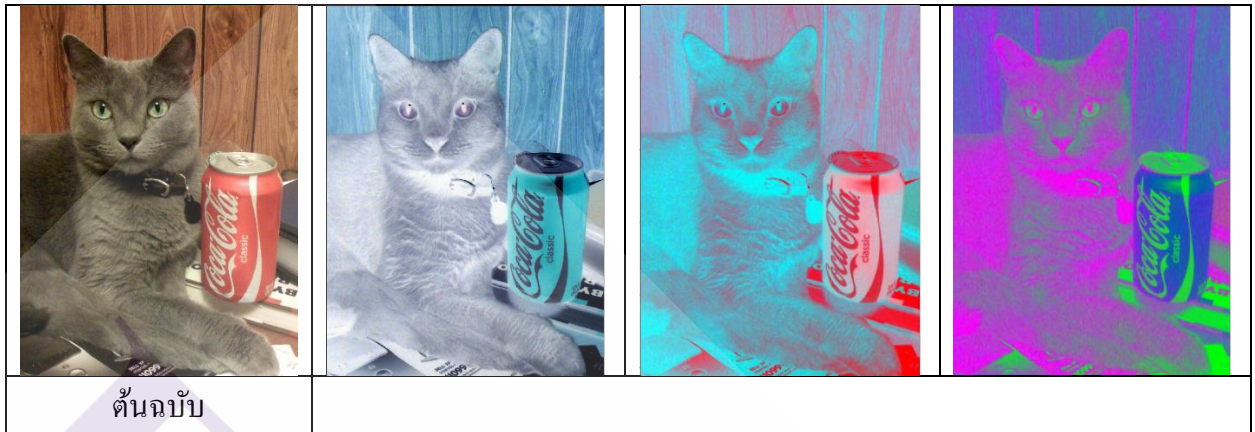
ตารางที่ 6.6 ตัวอย่างการ AgerageBlur



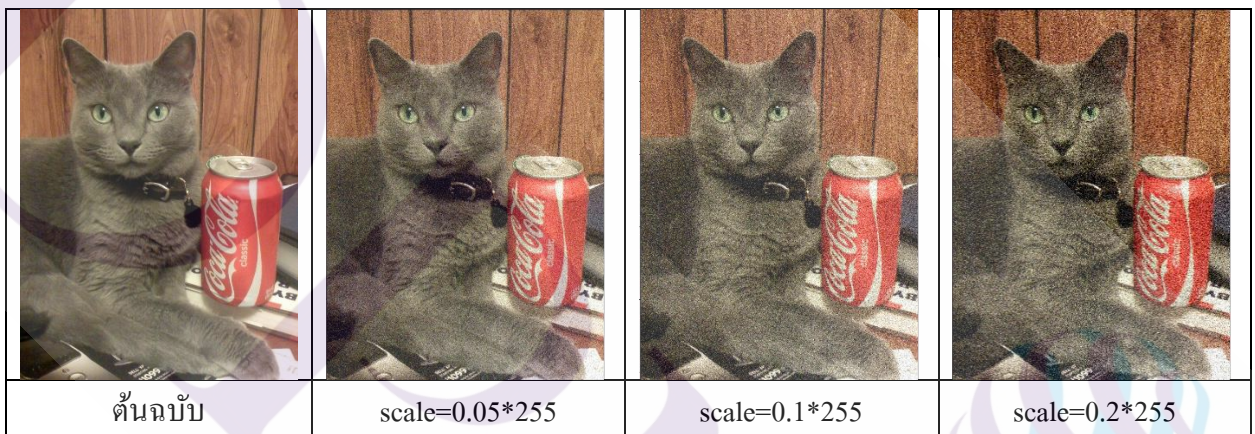
ตารางที่ 6.7 ตัวอย่างการ GammaContrast



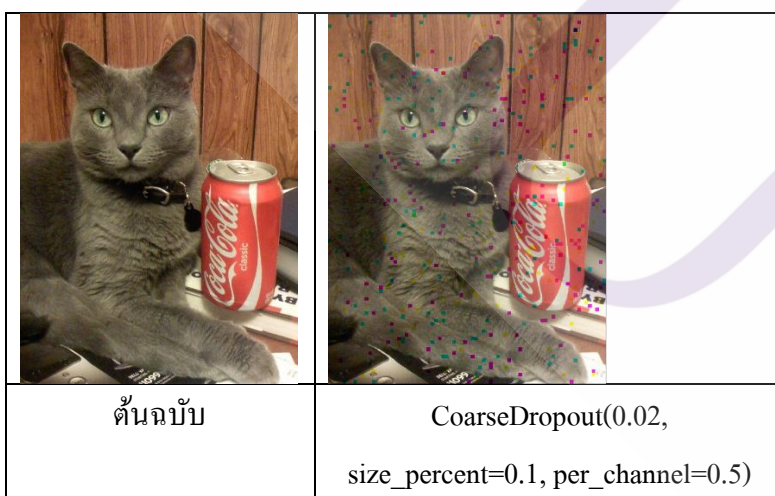
ตารางที่ 6.8 ตัวอย่างการ Grayscale



ตารางที่ 6.9 ตัวอย่างการ Invert

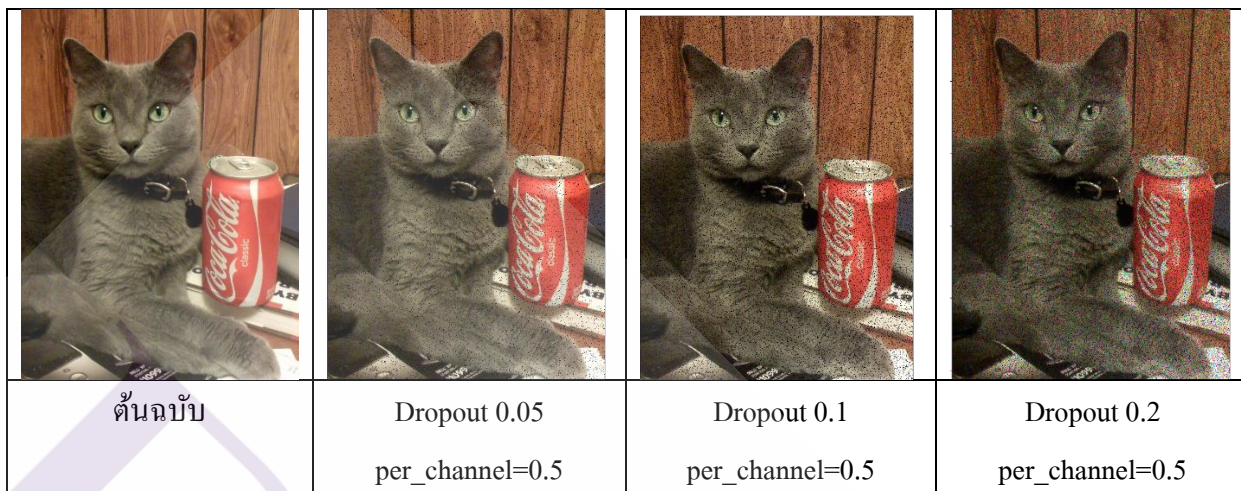


ตารางที่ 6.10 ตัวอย่างการ AdditiveGaussianNoise



ตารางที่ 6.11 ตัวอย่างการ CoarseDropout





ตารางที่ 6.12 ตัวอย่างการ Dropout

## 6.3 ตัวอย่างการทำนายผลโมเดล ทำนายถูกต้อง

ภาพต้นฉบับ	 A photograph showing a green Heineken Lager Beer bottle in the foreground and a red Coca-Cola can partially visible behind it. The Heineken label is clearly visible, featuring a red star and the brand name.
ผลการทำนาย	 The same image as above, but with red dashed bounding boxes around the Heineken label and the Coca-Cola label. The Heineken label is annotated with the text "heineken 0.999" and the Coca-Cola label is annotated with "coca-cola 0.951".

ภาพต้นฉบับ



ผลการทำนาย



<p>ภาพต้นฉบับ</p>	 A photograph of a brown glass beer bottle and a glass of beer with ice on a concrete pier. The background shows a body of water and a cloudy sky. A large, faint watermark is visible across the image.
<p>ผลการทำนาย</p>	 A predicted image of the same scene as the original. A red rectangular box is drawn around a portion of the beer bottle's label. A large, faint watermark is visible across the image.

<p>ภาพต้นฉบับ</p>	
<p>ผลการทำนาย</p>	

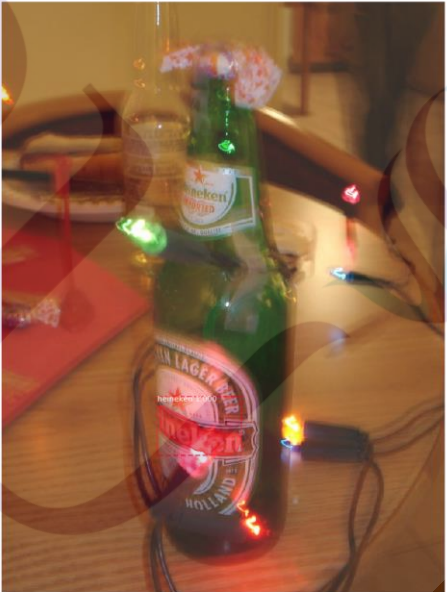
<p>ภาพต้นฉบับ</p>	 A photograph of a white plastic cup with a Heineken logo and a red star, filled with beer and a thick head of foam. The cup is sitting on a patch of green grass. A person's hand is partially visible on the right side of the frame.
<p>ผลการทำนาย</p>	 The same photograph as above, but with a red dashed rectangular box drawn around the Heineken logo and the red star. A semi-transparent watermark is overlaid on the image.

ภาพต้นฉบับ



ผลการทำนาย



ภาพต้นฉบับ	 A photograph of a green Heineken beer bottle on a wooden table. The bottle is wrapped with a string of small, multi-colored lights (red, green, yellow). A small white paper decoration is attached to the top of the bottle. The background is slightly blurred, showing a red object and some cables.
ผลการทำนาย	 A predicted or processed version of the same Heineken beer bottle image. The bottle, lights, and background are clearly visible, but there is a noticeable digital artifact or ghosting effect on the right side of the image, where a blue and white striped pattern appears to be overlaid or reflected.



ภาพต้นฉบับ



ผลการทำนาย



## 6.2 ตัวอย่างการทำนายผลโมเดล ทำนายผิด, ไม่ครบ



ภาพต้นฉบับ	 A photograph of several large plastic bottles of Pepsi lined up on a shelf. The bottles are dark with the classic Pepsi logo and branding. One bottle in the foreground has a yellow cap and a 'PEPSI STUFF' label.
ผลการทำนาย -ทำนายไม่ครบ	 The same photograph as above, but with several colored bounding boxes overlaid on the Pepsi logos. Each box is labeled with 'pepsi 0.998' or 'pepsi 0.999'. The boxes are purple, red, yellow, and green. The yellow box around the central logo is dashed, indicating a prediction error or incomplete detection.

ภาพต้นฉบับ



ผลการทำนาย  
- ทำนายไม่ครบ



<p>ภาพต้นฉบับ</p>	
<p>ผลการทำนาย - ทำนายผิด</p>	

<p>ภาพต้นฉบับ</p>	
<p>ผลการทำนาย - ทำนายไม่ครบ</p>	

ภาพต้นฉบับ



ผลการทำนาย  
- ทำนายไม่ครบ



<p>ภาพต้นฉบับ</p>	 A photograph of a white Heineken beer bucket filled with six green glass bottles of Heineken beer and ice cubes. The bucket has the Heineken logo and name on the front. The bottles are arranged in a circular pattern within the bucket.
<p>ผลการทำนาย - ทำนายไม่ครบ</p>	 A predicted image of a Heineken beer bucket, identical to the original image. It shows a white bucket with the Heineken logo, containing six green bottles of beer and ice cubes. The prediction is very accurate, capturing all the details of the original image.

ภาพต้นฉบับ



ผลการทำนาย

- ทำนายผิด





ภาพต้นฉบับ	
ผลการทำนาย - ทำนายผิด	

ภาพต้นฉบับ



ผลการทำนาย  
- ทำนายไม่ครบ



## ประวัติผู้วิจัย

ชื่อ นามสกุล

พิชิตชัย พิมพ์โคตร

ประวัติการศึกษา

พ.ศ. 2546 ปริญญาตรี

สาขาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ตำแหน่งและสถานที่ทำงานปัจจุบัน

นักวิทยาศาสตร์ข้อมูล

บริษัท อาร์เอส จำกัด (มหาชน)

