

การบ่งชี้วัตถุและบุคคลที่สนใจจากวิดีโอกล้องวงจรปิดในหลายแหล่งที่มา

**INDIVIDUAL AND OBJECT IDENTIFICATIONS FROM THE
COPIOUS AMOUNTS OF CCTV RECORDS**



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาก่อนอื่น เมื่อผู้จัดทำเห็นว่าไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีเหตุแห่งลิขสิทธิ์และต้องขออนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปีการศึกษา 2563

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ปริญญาโทปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การบ่งชี้วัตถุและบุคคลที่สนใจจากวิดีโอกล้องวงจรปิดในหลายแหล่งที่มา

INDIVIDUAL AND OBJECT IDENTIFICATIONS FROM THE COPIOUS
AMOUNTS OF CCTV RECORDS

ผู้จัดทำ

1. นายวโรดม ไบอุดม รหัสนักศึกษา 60010912
2. นายวัชรเกียรติ ว่องเจริญพร รหัสนักศึกษา 60010916
3. นายศิวกรณ์ บุญพามี รหัสนักศึกษา 60010986



อาจารย์ที่ปรึกษา

(ผศ.อักรเดช วัชรภูกงษ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

การบ่งชี้วัตถุและบุคคลที่สนใจจากวิดีโอกล้องวงจรปิดในหลาย

แหล่งที่มา

นายวโรดม	ใบอุดม	60010912
นายวัชรเกียรติ	ว่องเจริญพร	60010916
นายศิวกรณ์	บุญพามี	60010986
ผศ.อัครเดช	วัชรเทพพงษ์	อาจารย์ที่ปรึกษา

ปีการศึกษา 2563

บทคัดย่อ

วิดีโอ ถือเป็นเครื่องมืออย่างหนึ่งในการบันทึกเหตุการณ์ต่างๆที่เกิดขึ้น ตัวอย่างเช่น วิดีโอ จากกล้องวงจรปิด ซึ่งเป็นอุปกรณ์สำคัญที่สามารถใช้ในการตามรอยวัตถุต่างๆ หรือตามหาบุคคลที่สนใจได้ ซึ่งมีเหตุการณ์มากมายในอดีตที่ใช้ข้อมูลจากกล้องวงจรปิด มามีส่วนในการตามตัวผู้ต้องสงสัย ดังตัวอย่างเหตุการณ์ที่เคยเป็นข่าวดัง เช่น เหตุการณ์วางระเบิดที่สี่แยกราชประสงค์ หลังจากเกิดเหตุการณ์ ได้มีเจ้าหน้าที่จากหลายๆฝ่ายได้ระดมกำลังกันเพื่อที่จะค้นหาสาเหตุ หรือบุคคลที่ก่อเหตุต่อการกระทำดังกล่าว หลังจากนั้นจึงได้พยานหลักฐานโดยเป็นวิดีโอจากกล้องวงจรปิดภายในบริเวณนั้น ที่ได้บันทึกวิดีโอยืนยันถึงลักษณะของคนร้ายได้จากกล้องวงจรปิดหลายๆแหล่ง ทั้งจากของภาครัฐเอง หรือจากทางภาคเอกชนด้วย โดยคนร้ายได้แต่งกายโดยมีลักษณะเด่นคือ สวมเสื้อเชิ้ตสีเหลือง ทำให้เจ้าหน้าที่ตำรวจต้องตามหากล้องวงจรปิดในบริเวณใกล้เคียง และตามหาวิดีโอที่มีลักษณะตรงกับผู้ร้ายคนดังกล่าว ซึ่งต้องใช้กำลังของเจ้าหน้าที่จำนวนมากในการตามหาคนร้าย จากวิดีโอกล้องวงจรปิดที่มีจำนวนมาก ทำให้ใช้เวลาและทรัพยากรบุคคลมากในการตามหา จนกระทั่งพบผู้ก่อเหตุ จากปัญหาที่กล่าวมาข้างต้น จึงได้เห็นความสำคัญของปัญหาดังกล่าว และได้คิดโครงการ การบ่งชี้วัตถุที่สนใจในวิดีโอจากหลายแหล่งที่มา เพื่อใช้ในการตามหาบุคคลที่สนใจจากลักษณะที่ต้องการ เพื่อลดระยะเวลา และจำนวนของวิดีโอที่ต้องใช้ในการค้นหาให้ลดลง ซึ่งผลลัพธ์ของโปรแกรมที่จะได้ออกมา จะเป็นคลิปวิดีโอที่ถูกตัดออกมาหลายๆคลิป และภายในวิดีโอจะมีบุคคลที่สนใจอยู่ภายในนั้น ซึ่งจะทำให้วิดีโอที่ต้องค้นหาในตอนแรก มีระยะเวลาที่สั้นลง เพื่อให้การทำงานของผู้ที่เกี่ยวข้อง อย่างเช่น เจ้าหน้าที่ตำรวจ จะสามารถทำงานได้อย่างสะดวก และรวดเร็วขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดก็ตามห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

INDIVIDUAL AND OBJECT IDENTIFICATIONS FROM THE COPIOUS AMOUNTS OF CCTV RECORDS

Mr. Warodom	Baiudom	60010912
Mr. Watcharakiete	Wongcharoenbhorn	60010916
Mr. Siwakorn	Boonpamee	60010986
Asst.Prof. Akkradach	Watcharapupong	Advisor

Academic Year2020

ABSTRACT

Video is considered one of the most imperative tools for recording many incidents. For instance, CCTV camera can be deployed to track suspect in many cases in the past. One of the most tragic events is sabotaged explosive at Ratchaprasong joint whilst police had been inspecting throughout the entire recorded videos. The amount is so prolific that utilizing merely the human willpower was not sufficient and not effective enough to trace suspect characteristics afterwards. Noticeably, the perpetrator in question, for example, wore yellow shirt, and hence, a great clue to trace. Nonetheless, such tasks still are nowhere near success to inspect person holding the conditions we had from the copious amounts of data as we discussed. We, therefore, make this project in hope to help authorities investigate the crimes and bring justice back to peaceful community.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

โครงการนี้จะประสบความสำเร็จไปไม่ได้ หากขาดความช่วยเหลือจากอาจารย์อัครเดช วัชรภุพงษ์ อาจารย์ที่ปรึกษาโครงการ ที่ได้ให้คำแนะนำ แนวคิดในการแก้ปัญหา ให้คำปรึกษาและ ชี้แนะเรื่องต่างๆ และขอขอบคุณสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่เอื้อเพื่อ สถานที่สำหรับการทำงาน และการค้นคว้าหาความรู้ในการจัดทำโครงการ

ขอขอบคุณอาจารย์หลาย ๆ ท่าน ภายในภาควิชาวิศวกรรมคอมพิวเตอร์ คณะ วิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่คอยให้ความรู้ คำปรึกษา และคำแนะนำในระหว่างการทำโครงการ ตลอดจนเพื่อนๆ ในภาควิชา โดยเฉพาะอย่างยิ่ง คุณพ่อ และคุณแม่ของคณะผู้จัดทำ ที่คอยให้กำลังใจ ให้คำแนะนำตลอดการทำโครงการชิ้นนี้



ไบอูดม
ว่องเจริญพร
บุญพามี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญภาพ.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของโครงการ.....	1
1.3 เป้าหมายของโครงการ.....	2
1.4 ขอบเขตของโครงการ.....	2
1.5 แผนการดำเนินงาน.....	3
บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง.....	6
2.1 คำศัพท์และนิยามที่เกี่ยวข้อง.....	6
2.2 การตรวจนับวัตถุ.....	9
2.3 ทฤษฎีสี.....	16
2.4 เทคโนโลยีที่ใช้.....	18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่บนสื่อออนไลน์
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ (ต่อ)

	หน้า
บทที่ 3 การออกแบบและพัฒนาระบบ	21
3.1 ภาพรวมของระบบ	21
3.2 ขั้นตอนการทำโครงการ	23
บทที่ 4 ผลการทดลอง	26
4.1 การทดสอบการตรวจจับเสือ	26
4.2 การทดสอบการตรวจหาสีเสือ	31
4.3 การทดสอบการตรวจจับกระเป๋าส	34
4.4 การแสดงผล	36
บทที่ 5 สรุปการดำเนินงานของโครงการ	37
5.1 สรุปผลที่ได้จากการทำโครงการ	37
5.2 ปัญหาและอุปสรรค	37
5.3 วิธีการในการแก้ปัญหาที่เกิดขึ้น	38
5.4 แนวทางการพัฒนาต่อ	38
บรรณานุกรม	39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตาราง	หน้า
1.1 แผนการดำเนินงาน โครงการงาน 1	3
1.2 แผนการดำเนินงาน โครงการงาน 2	4
2.1 ค่าที่ใช้ในการคำนวณ Bounding Box	11
4.1 ผลลัพธ์จากโมเดลตรวจจับตำแหน่งของเสื้อ	27
4.2 Confusion Matrix ของโมเดลตรวจจับเสื้อคน	28
4.3 ผลลัพธ์จากการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ	28
4.4 Confusion Matrix ของการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ	29
4.5 เปรียบเทียบประสิทธิภาพระหว่าง One-layer system และ Two-layers system	30
4.6 เปรียบเทียบประสิทธิภาพระหว่าง YOLO และ SSD	31
4.7 ตัวอย่างเสื้อที่ได้จากการเข้า function ระบุสี	31
4.8 ผลลัพธ์จากการใช้ Pre-trained Model ในการตรวจจับกระเป๋า	35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูป	หน้า
2.1 โครงข่าย Neural Network.....	7
2.2 โครงข่าย Convolution Neural Network.....	7
2.3 การทำ Non-Maximum Suppression (NMS).....	8
2.4 Intersection over Union.....	8
2.5 ระบบการตรวจจับวัตถุของ YOLO.....	10
2.6 สถาปัตยกรรมของ YOLO ก่อนเข้าสู่การ detection และ non-maximum suppression.....	12
2.7 ระบบสี HSV.....	17
2.8 โลโก้ของ Google Colab.....	18
2.9 โลโก้ของ Labellmg.....	19
2.10 โลโก้ของ Darknet.....	19
2.11 โลโก้ของ Line Notify.....	20
3.1 ภาพรวมการทำงานของระบบ.....	21
3.2 Sequence Diagram การทำงานของระบบตรวจจับเชื้อ และกระเป๋.....	22
3.3 Sequence Diagram การทำงานภายในของระบบ.....	22
3.4 รูปแสดงข้อมูลภายในไฟล์ .txt.....	24
4.1 ผลลัพธ์ที่ส่งเข้า Line Notify.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

วิดีโอ ถือเป็นเครื่องมืออย่างหนึ่งในการบันทึกเหตุการณ์ต่าง ๆ ที่เกิดขึ้น ตัวอย่างเช่น วิดีโอจากกล้องวงจรปิด ซึ่งเป็นอุปกรณ์สำคัญที่สามารถใช้ในการตามรอยวัตถุต่าง ๆ หรือตามหาบุคคลที่สนใจได้ ซึ่งมีเหตุการณ์มากมายในอดีตที่ใช้ข้อมูลจากกล้องวงจรปิด มามีส่วนในการตามตัวผู้ต้องสงสัย ดังตัวอย่างเหตุการณ์ที่เคยเป็นข่าวดัง เช่น เหตุการณ์วางระเบิดที่สี่แยกราชประสงค์ หลังจากเกิดเหตุการณ์ ได้มีเจ้าหน้าที่จากหลายฝ่ายได้ระดมกำลังกันเพื่อที่จะค้นหาสาเหตุ หรือบุคคลที่ก่อเหตุต่อกรกระทำดังกล่าว หลังจากนั้นจึงได้พยานหลักฐานโดยเป็นวิดีโอจากกล้องวงจรปิดภายในบริเวณนั้น ที่ได้บันทึกวิดีโอยืนยันถึงลักษณะของคนร้าย ได้จากกล้องวงจรปิดหลายๆแหล่ง ทั้งจากของภาครัฐเอง หรือจากทางภาคเอกชนด้วย โดยคนร้ายได้แต่งกายโดยมีลักษณะเด่นคือ สวมเสื้อเชิ้ตสีเหลือง ทำให้เจ้าหน้าที่ตำรวจต้องตามหากกล้องวงจรปิดในบริเวณใกล้เคียง และตามหาวิดีโอที่มีลักษณะตรงกับผู้ร้ายคนดังกล่าว ซึ่งต้องใช้กำลังของเจ้าหน้าที่จำนวนมากในการตามหาคนร้าย จากวิดีโอกล้องวงจรปิดที่มีจำนวนมาก ทำให้ใช้เวลาและทรัพยากรบุคคลมากในการตามหาจนกระทั่งพบผู้ก่อเหตุ

จากปัญหาที่กล่าวมาข้างต้น จึงได้เห็นความสำคัญของปัญหาดังกล่าว และได้คิดโครงการ การบ่งชี้วัตถุที่สนใจในวิดีโอจากหลายแหล่งที่มา (CE Project63-21: Individual and object identifications from the copious amounts of CCTV records) เพื่อใช้ในการตามหาบุคคลที่สนใจจากลักษณะที่ต้องการ เพื่อลดระยะเวลา และจำนวนของวิดีโอที่ต้องใช้ในการค้นหาให้ลดลง ซึ่งผลลัพธ์ของโปรแกรมที่จะได้ออกมา จะเป็นคลิปวิดีโอที่ถูกตัดออกมาหลายๆคลิป และภายในวิดีโอจะมีบุคคลที่สนใจอยู่ภายในนั้น ซึ่งจะทำให้วิดีโอที่ต้องค้นหาในตอนแรก มีระยะเวลาที่สั้นลง เพื่อให้การทำงานของผู้ที่เกี่ยวข้อง อย่างเช่น เจ้าหน้าที่ตำรวจ จะสามารถทำงานได้อย่างสะดวกและรวดเร็วขึ้น

1.2 วัตถุประสงค์

เอกสารนี้เป็นเอกสารสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น สิ่งนี้ยังหวังจะให้คำปรึกษาและต้องอ้างอิงถึงชื่อของเอกสารหรือชื่อที่สามารถนำไปใช้

1) เพื่อทำการศึกษาหลักการการทำงานของ Neural network ที่ใช้ในการประมวลผลเกี่ยวกับภาพ

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- 2) เพื่อประยุกต์การทำงานของ Neural network ไปใช้งานกับ การตรวจจับวัตถุ
- 3) เพื่อลดปริมาณงานในการค้นหาคลิปวิดีโอในกล้องวงจรปิดและทรัพยากรที่ใช้ในการแก้ปัญหา

1.3 เป้าหมาย

- 1) นำโมเดล ที่ประยุกต์ใช้การทำงานของ Neural Network สำหรับการตรวจจับวัตถุ มาใช้งานร่วมกับ การตรวจจับสี เพื่อที่จะคัดแยกบุคคลที่สวมเสื้อในสีที่ต้องการภายในวิดีโอ
- 2) คัดเลือกวิดีโอและส่วนของวิดีโอ ที่มีบุคคลที่ตรงตามลักษณะที่ทราบหรือสนใจ เช่น สีเสื้อ ออกมาจากวิดีโอที่มีอยู่เป็นจำนวนมาก ด้วยการใช้โมเดลที่ได้จัดเตรียมขึ้น
- 3) แสดงผลช่วงของเวลาในคลิปวิดีโอที่มีบุคคลลักษณะตรงตามที่ต้องการอยู่ในคลิปหรือ ส่วนของคลิปที่ถูกคัดออกมา ซึ่งจะช่วยลดเวลาและปริมาณงานของเจ้าหน้าที่ในการตรวจสอบวิดีโอจากกล้องวงจรปิด

1.4 ขอบเขต

- 1) สนใจเฉพาะการตรวจจับบุคคลภายในวิดีโอที่มาจากกล้องวงจรปิดเป็นหลัก
- 2) ทำการศึกษาในส่วนที่เป็นเสื้อของบุคคล โดยมีขอบเขตเป็นสีต่าง ๆ โดยไม่คำนึงถึงรูปแบบและประเภทของเสื้อ
- 3) ความละเอียดของวิดีโอ ต้องมีความละเอียดขั้นต่ำ 480 (640x480) จึงจะทำให้โมเดลมีประสิทธิภาพ
- 4) ไม่คำนึงถึงเสียงที่ถูกรับในวิดีโอ
- 5) มีการตัดทอน Frame rate ของ video
- 6) วิดีโอจะต้องเป็นภาพแบบสี และมีความเข้ม ความสว่างที่ชัดเจน
- 7) ขนาดพื้นที่ของคนภายในวิดีโอ เทียบกับขนาดพื้นที่ของวิดีโอ ต้องมีขนาดมากกว่า 2.5 เปอร์เซ็นต์ของขนาดพื้นที่วิดีโอ
- 8) ความหนาแน่นของคนภายในแต่ละเฟรมของวิดีโอ ไม่เกิน 10 คน และ คนภายในวิดีโอไม่ยืนซ้อนกัน
- 9) ความสูงของกล้องวิดีโอกล้องวงจรปิดที่ติดตั้ง 2-4 เมตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

1.5 แผนการดำเนินงาน

ตาราง 1.1 แผนการดำเนินงานโครงการ 1

แผนการดำเนินงาน	พ.ศ. 2563															
	ส.ค				ก.ย				ต.ก				พ.ย			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1. รวบรวม และสืบค้นข้อมูล																
1.1 หา paper ของ single state																
1.2 หา paper ของ two state																
1.3 ศึกษาหลักการทำงานของ single state																
1.4 ศึกษาหลักการทำงานของ two state																
2. เปรียบเทียบ framework และ algorithm																
2.1 เปรียบเทียบ Accuracy ของ class ต่างๆภายในภาพที่ทำการ test จาก resource เดียวกัน																
2.2 เปรียบเทียบ ความเร็วในการประมวลผล และการ Predict																
2.3 เปรียบเทียบ Resources ที่ใช้ในการประมวลผลเช่น Ram, GPU ที่ใช้																
3. ศึกษาการทำงานของ model																
3.1 ศึกษาการทำงาน Colab																
3.2 ศึกษา code จาก tutorial																
3.3 ศึกษาการทำงานของ Yolov4																
3.4 ศึกษาการทำงานของ Darknet																
4. Implement framework and algorithm กับ รูปที่หามาได้																
5. หา dataset สำหรับ test และ train model																
6. ทำ labelling image class เพื่อหา ตำแหน่งของเสื้อ																
7. ทำ Data Augmentation																

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษานี้เท่านั้น ไม่อนุญาตให้เผยแพร่ไปใช้ประโยชน์ทางการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 1.1 แผนการดำเนินงานโครงการ 1 (ต่อ)

แผนการดำเนินงาน	พ.ศ. 2563															
	ส.ค				ก.ย				ต.ค				พ.ย			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
8. ทำการ implement framework กับ Data ที่เตรียมไว้																
8.1 ทำการเทรนโมเดลสำหรับคลาส person																
8.2 ทำการเทรนโมเดลสำหรับคลาส เสื้อ																
8.3 กำหนดจัดการค่า threshold ของสี ตาม class ต่างๆ																
8.4 นำค่า bounding box ที่ได้จาก model detect เสื้อ เข้า model classify สีเสื้อ เพื่อหาสีเสื้อ																
9. จัดทำและปรับปรุงในส่วนของ report																
9.1 Report สำหรับโปรเจก 1																

ตาราง 1.2 แผนการดำเนินงานโครงการ 2

แผนการดำเนินงาน	พ.ศ. 2564															
	ม.ค				ก.พ				มี.ค				เม.ย			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
10. Implement กับ Video																
10.1 กำหนดขอบเขตของวิดีโอสำหรับใช้งานระบบ																
10.2 ทำส่วนการอ่านไฟล์ video																
10.3 ตัด Video ออกเป็น frame ย่อยๆ เพื่อเตรียมเข้าโมเดลตรวจจับเสื้อ																
10.4 นำแต่ละ frame ที่ได้ตัดไว้มาเข้าโมเดลตรวจจับเสื้อ																
11. ปรับปรุงระบบการคำนวณสี																
11.1 เขียน function ในการคำนวณระบบสี HSV โดยใช้ค่า Nearest Distance																
11.2 รวบรวมค่าระบบสี RGB ตามสี class ต่างๆ เพื่อใช้ในฟังก์ชัน																

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้ผู้อื่นใดได้ประโยชน์อันใดจากการนำเอกสารนี้ไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 1.2 แผนการดำเนินงานโครงการ 2 (ต่อ)

แผนการดำเนินงาน	พ.ศ. 2564															
	ม.ค				ก.พ				มี.ค				เม.ย			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
11.3 ทำส่วนที่ใช้ในการเลือกสีในการประมวลผล																
12. ปรับปรุงระบบการตรวจจับและกระเป่า																
12.1 รวบรวม Dataset และจัดเตรียมข้อมูลสำหรับฝึกสอนโมเดล																
12.2 ทำการฝึกสอน โมเดลเพิ่มเติม																
12.3 ทำให้ระบบทำการประมวลผลแบบ Multiprocessor																
12.4 ทำโมเดลในการตรวจจับกระเป่าเพิ่มเติม																
12.5 เชื่อมต่อระบบตรวจจับกับกระเป่าเข้าด้วยกัน																
12.6 ทดสอบ และประเมินผลโมเดล																
12.7 Refactoring code																
13. จัดทำและปรับปรุงในส่วนของ report																
13.1 Report สำหรับ โปรเจก 2																

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 2

แนวคิดและทฤษฎีที่เกี่ยวข้อง

2.1 คำศัพท์และนิยามที่เกี่ยวข้อง

2.1.1 Machine learning and Deep learning

Machine Learning เป็นสาขาย่อยของ Artificial intelligence หรือ AI คือ การที่ระบบ หรือ โปรแกรมคอมพิวเตอร์ที่สามารถเรียนรู้ระบบรูปแบบการทำงาน หรือสิ่งต่างๆได้ด้วยตัวเอง จากการเรียนรู้จากชุดข้อมูล หรืออินพุตที่ได้รับมา ซึ่งมนุษย์ก็จะมีหน้าที่ในการเขียน Algorithm ต่างๆ เพื่อให้คอมพิวเตอร์สามารถเรียนรู้จากข้อมูลนั้นๆ ได้ ซึ่งข้อมูลก็อาจจะมียหลายแบบ เช่น ข้อมูลชุดรูปภาพ ข้อมูลเสียง ข้อมูลจากตัวหนังสือ เป็นต้น ซึ่งขึ้นอยู่กับการนำไปใช้งานตามความเหมาะสม

2.1.2 Classification, Localization, Object Detection and Bounding box

2.1.2.1 Classification

เป็นรูปแบบหนึ่งของ Supervised Learning เป็นรูปแบบการจำแนกข้อมูล ออกเป็นกลุ่มๆ ที่มีหัวข้ออย่างชัดเจน อย่างเช่น มนุษย์ รถยนต์ เป็นต้น โดยมีอัลกอริทึมเป็นตัวคัดแยกเรียกว่า Classifier

2.1.2.2 Localization

เป็นกระบวนการในการระบุตำแหน่งของวัตถุลงบนภาพ โดยจะทำการหาตำแหน่งของวัตถุและทำการระบุตำแหน่งด้วยการตีกรอบ Bounding Box

2.1.2.3 Bounding Box

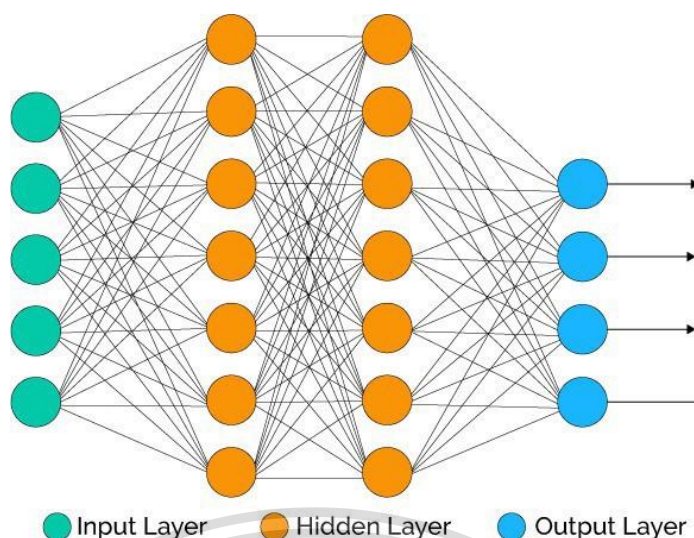
เป็นสี่เหลี่ยมจินตภาพ ถูกใช้เพื่ออ้างอิงตำแหน่งของวัตถุบนภาพให้กับอัลกอริทึม โดยจะมีคุณสมบัติอยู่ 4 อย่างนั่นคือ ตำแหน่ง X ,Y ,ความสูง และความกว้าง ซึ่งในระบบของ YOLO Object detection จะมีองค์ประกอบอยู่ 5 ส่วน โดยเพิ่มในส่วนที่ของ Objectness Confidence ซึ่งจะแสดงค่าความเชื่อมั่นของการมีอยู่ของวัตถุที่ correspond กับ Bounding box นั้น

2.1.3 Neural Network (NN), Convolutional Neural Network (CNN)

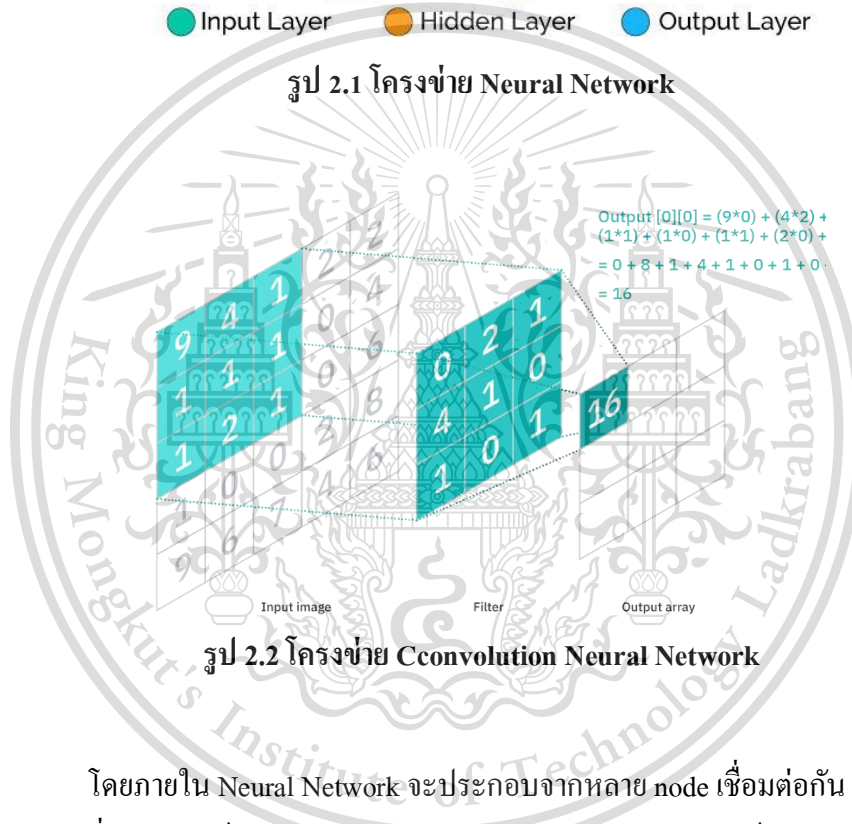
โครงข่ายประสาทเทียม เป็นศาสตร์ที่เกี่ยวข้องกับการศึกษาและสร้าง โครงข่ายประสาทที่เลียนแบบการทำงานของระบบประสาทของมนุษย์ ที่สามารถจะเรียนรู้การทำงานต่าง ๆ ที่เกี่ยวกับข้อมูล หรือแม้แต่งานทางด้านคอมพิวเตอร์วิทัศน์ (Computer vision) และ โดยเฉพาะอย่างยิ่งได้มีการศึกษาโครงข่ายแบบ Convolution (CNN) ขึ้นทำให้สะดวกสำหรับการนำมาประยุกต์ใช้ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้กับปัญหาที่เกี่ยวข้องกับการทำ image processing ได้ดังรูป 2.1 และ 2.2

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 2.1 โครงข่าย Neural Network



รูป 2.2 โครงข่าย Cconvolution Neural Network

โดยภายใน Neural Network จะประกอบจากหลาย node เชื่อมต่อกัน และประกอบกันเป็น layer เชื่อมต่อกันเป็น Network โดย Neural Network จะแบ่งออกเป็น 3 ส่วนหลักคือ input layer เป็นส่วนที่จะรับค่า input เข้ามาเพื่อทำการประมวลผลใน Neural Network, Hidden layer เป็นส่วนที่จะปรับ weight เพื่อให้เข้ากับข้อมูลที่ต้องการใช้ในการฝึกสอน และส่วนสุดท้ายคือ Output layer เป็นส่วนที่จะแสดงผล Output ตาม classes ต่างๆ

หลักการในส่วนนี้จะมีการเทรนข้อมูลให้กับโครงข่ายซึ่งจะมี Weight บันทึกลงเอาไว้ และจะสามารถเปลี่ยนแปลงได้ผ่านอัตราการเรียนรู้ (learning rate) ค่าหนึ่งให้คล้ายกับการ

เรียนรู้ของมนุษย์ เพื่อให้โครงข่ายประสาทเทียมสามารถทำนายผลจากข้อมูลชุดใหม่ที่เข้ามาได้ แต่อาจเกิดปัญหาที่เรียกว่า Overfitting ได้ ซึ่งปัญหานี้จะคล้ายกับการที่มนุษย์จำคำตอบมากเกินไป

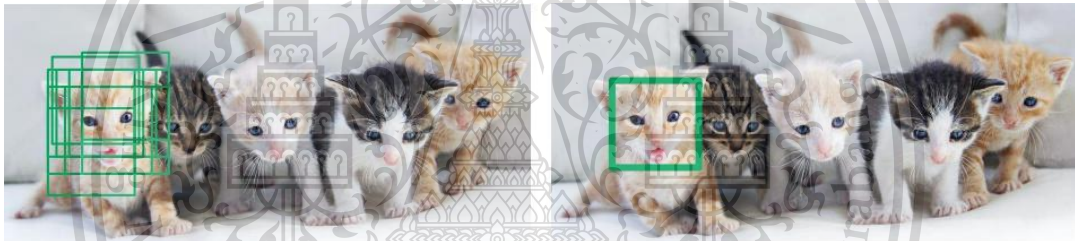
This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

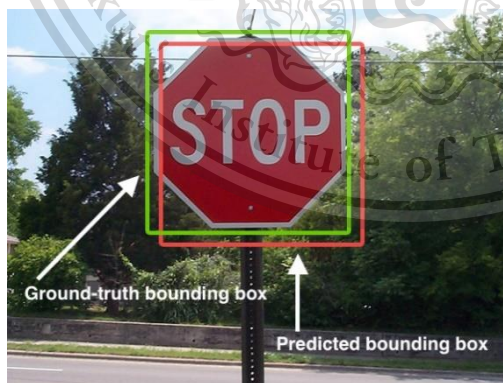
ที่มีจำนวนเซลล์มากเกินไปจนทำให้ neuron แต่ละตัวให้ความสำคัญกับข้อมูลทุกชุดที่เข้ามา ซึ่งจะ ทำให้ไม่สามารถทำนายผลลัพธ์ที่ถูกต้องสำหรับข้อมูลใหม่ ๆ ที่เข้ามาได้ แม้ว่าจะสามารถทำนายผล จากข้อมูลที่ใช้ฝึกสอนได้อย่างแม่นยำมากก็ตาม

2.1.4 Non-maximal suppression – Intersection Over Union (IOU)

ในการตรวจจับวัตถุ สำหรับ โมเดลที่แสดงผลลัพธ์ออกมาเป็น Bounding box ของ วัตถุนั้น ในการตรวจจับของแต่ละวัตถุ จะได้กรอบของวัตถุสำหรับ 1 วัตถุออกมาเป็นหลายๆกรอบ ซึ่งในความเป็นจริงแล้วควรจะต้องแสดงเป็นกรอบเดียวต่อวัตถุหนึ่งชิ้น จึงจำเป็นที่จะต้องทำ Non-maximal suppression (NMS) เพื่อจัดการปัญหานี้ โดยใน YOLO จะมีการนำ Intersection over Union เข้ามาใช้งาน ร่วมกับ Non-maximal suppression (NMS) โดยคำนวณจากอัตราส่วนของพื้นที่ ซ้อนทับกัน ระหว่าง Bounding box ที่ทำนายจาก โครงข่าย กับ Bounding box ที่เป็นคำตอบ (Ground truth) เทียบกับค่าพื้นที่ที่ Union กันระหว่างสองพื้นที่ดังกล่าว เพื่อที่จะให้ได้ผลลัพธ์ที่ดีที่สุด

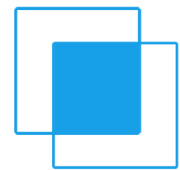


รูป 2.3 การทำ Non-Maximum Suppression (NMS)



ก)

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



ข)

รูป 2.4 Intersection over Union

ก) รูปภาพ Ground-truth และ Predicted bounding box

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับงานวิจัยเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.1.5 Dataset and Data preparation, and Data augmentation

2.1.5.1 ชุดข้อมูล (Dataset)

เป็นการรวบรวมข้อมูลเข้าไว้ด้วยกัน เพื่อการนำข้อมูลไปใช้งานสำหรับการศึกษากับการวิเคราะห์ข้อมูล การเรียนรู้ของเครื่อง การทำเหมืองข้อมูล เป็นต้น ซึ่งจำเป็นที่จะต้องมีการเตรียมข้อมูล (Data preparation) โดยสามารถทำได้ด้วยการที่ผู้ทดลองเก็บรวบรวมข้อมูลด้วยตนเองมาปริมาณหนึ่งก่อน ทำการ label สำหรับข้อมูลแบบ Supervised

2.1.5.2 Data augmentation

เป็นการ generate ข้อมูลเพิ่มเติมจากข้อมูลที่มีอยู่ ด้วยการทำการแปลงบางอย่าง เช่น ใน Image processing จะทำการแปลงรูปด้วยการ เลื่อน สะท้อน ลดขยายขนาดภาพ หรือรวมไปถึงการเข้าไปปรับแต่งในส่วนของคุณค่า สี แสง เงา และความคมชัดของภาพได้เช่นกัน

2.2 การตรวจจับวัตถุ

การตรวจจับวัตถุ หรือ Object Detection เป็นเทคโนโลยีหนึ่งของ Artificial intelligence โดยใช้หลักการของ Computer Vision และ Image Processing ซึ่งการทำงานจะเป็นกระบวนการจำแนกและตรวจจับประเภทของวัตถุ (Classification) และระบุตำแหน่งของวัตถุ (Localization) ทั้งหมดที่อยู่ในภาพ โดยการระบุขอบเขตของวัตถุสามารถทำได้หลายวิธี เช่น ใช้ Bounding Box ซึ่งเป็นการตีกรอบล้อมรอบวัตถุ หรือ การใช้ Segmentation ซึ่งเป็นการระบายสีถมที่วัตถุต่างๆเลย เป็นต้น

2.2.1 You Only Look Once model (YOLO)

YOLO เป็นอัลกอริทึมในการตรวจจับวัตถุแบบ Real-time โดยหลักการคือการมองปัญหาการตรวจจับวัตถุในรูปแบบของปัญหา Regression โดยในแต่ละโครงข่ายประสาทเทียมจะทำนายการตีกรอบของวัตถุและความน่าจะเป็นของคลาสวัตถุจากในภาพ

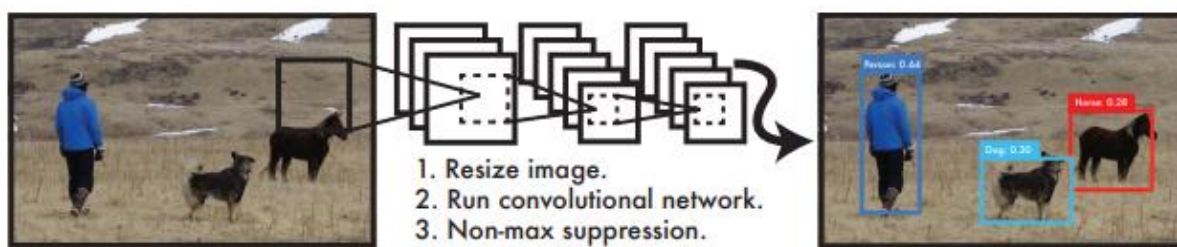
ระบบการตรวจจับวัตถุที่ใช้ในปัจจุบัน จะใช้วิธีหลักการนำ Classifier สำหรับวัตถุนั้นๆ มาคำนวณหาวัตถุนั้นๆ โดยการกำหนดพื้นที่ในภาพที่มีขนาดเฟรมที่หลากหลายและทำการ Classify object ในบริเวณนั้น ยกตัวอย่างเช่น หลักการ R-CNN จะใช้วิธีแบ่งเป็นบริเวณในภาพออกมาด้วย Regional Proposal Network เพื่อที่จะสร้าง Bounding box หลาย ๆ กรอบบนรูปและนำมา Classify ในแต่ละ box เหล่านี้เพื่อที่จะทำนาย Object ซึ่งกระบวนการเหล่านี้จะใช้เวลามากและปรับแต่งได้ยาก เนื่องจากองค์ประกอบแต่ละส่วนแยกจากกันทำให้ต้องทำการเทรนเพื่อปรับค่า Weight แยกกันเพื่อความแม่นยำ

หลักการที่ YOLO ใช้จึงเปลี่ยนมุมมองของการแก้ปัญหา Object detection เป็นปัญหา Regression โดยตรงระหว่าง Pixel ในภาพและค่า Coordinate (x, y, w, h) ของ Bounding box และค่าความน่าจะเป็นของคลาสวัตถุที่สอดคล้องกับ Pixel นั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น เมื่อผู้ผู้จัดทำเห็นใบแจ้งประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งหากมีเหตุใดเปลี่ยนแปลงเนื้อหา และต้องอ้างอิงถึงเนื้อหาของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 2.5 ระบบการตรวจจับวัตถุของ YOLO

YOLO ใช้หลักการที่ไม่ซับซ้อน ดังรูป 2.2 จะเห็นว่า ภายใน Convolutional Neural Network แต่ละตัวจะสร้าง Bounding box หลายๆกล่องขึ้นมาและคำนวณความน่าจะเป็นของคลาสวัตถุ ไปพร้อมๆกัน ซึ่ง YOLO จะใช้วิธีการ train จากทั้งภาพและทำการปรับปรุงแก้ไขประสิทธิภาพของการ detection โดยตรง ซึ่งทำให้โมเดลนี้ มีข้อได้เปรียบเหนือโมเดลแบบอื่นๆที่มีอยู่ก่อนหน้านี้ ดังนี้

ประการแรก YOLO สามารถประมวลผลได้อย่างรวดเร็ว เพราะ YOLO มองปัญหาเป็นแบบ Regression ทำให้ไม่ต้องการขั้นตอนอะไรที่ซับซ้อน ซึ่งหมายความว่า YOLO สามารถคำนวณภาพแบบ Real-time ในความล่าช้าที่น้อย แต่ YOLO ได้ความแม่นยำ (mAP) ที่มากกว่าค่าเฉลี่ยระบบ Real-time อื่นๆ

ประการที่สอง YOLO มองรูปทั้งหมดเป็นภาพรวมในการ Predict ซึ่งต่างจากการใช้วิธีการที่เป็นที่นิยมอย่าง Fast R-CNN จะมีปัญหา กับพื้นหลังมาก เพราะไม่สามารถมองภาพรวมทั้งหมดได้ เมื่อเทียบกับ YOLO ในแง่ของ False positive สำหรับพื้นหลังแล้ว YOLO มีข้อผิดพลาดน้อยกว่า

ประการที่สาม ระบบ YOLO เรียนรู้วัตถุอย่างหลากหลาย เมื่อลองนำไปทดลองกับภาพวาดแล้ว YOLO สามารถเอาชนะวิธีการอื่นๆ ได้ ด้วยเหตุนี้จึงทำให้เมื่อนำ ระบบ YOLO เข้าไปประยุกต์ใช้กับเรื่องอื่นๆ เช่น ในภาพวาด หรือในภาพถ่ายที่มีความแตกต่างของสภาวะทางสี ที่คาดไม่ถึง จึงมีโอกาที่จะผิดพลาดได้น้อยกว่าการนำระบบอื่นๆ

2.2.2 การตรวจจับของ YOLO

ระบบ YOLO รวมหลาย ๆองค์ประกอบในการตรวจจับวัตถุ เข้าเป็นโครงข่ายเดียว และใช้องค์ประกอบของภาพโดยรวมจากรูป ไปคำนวณหา Bounding box และได้คำนวณหาความน่าจะเป็นของคลาสวัตถุต่าง ๆ โดยระบบของ YOLO จะแบ่งรูปเป็นกริด ขนาด $S \times S$ ถ้าจุด

ศูนย์กลางของภาพอยู่ในช่องกริดไหน ช่องนั้น ๆ จะมีส่วนในการ detect object นั้น ๆ

ไม่ว่ากริดใด ๆ ทั้งสิ้น อีกทั้งหาจุดที่จุดใดจุดหนึ่ง และต้องอ้างอิงถึงเนื้อหาของเอกสารทุกครั้งที่มีการนำไปใช้ มี object ของ box นั้น ๆ ซึ่งค่าความเชื่อมั่นนี้บ่งบอกว่าใน box นั้นมี object อยู่หรือไม่และมีความ

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

แม่นยำเท่าใด ซึ่งแสดงโดยพจน์ของความน่าจะเป็นและค่า IOU คือ $\Pr(\text{Obj}) * \text{IOU}$ โดยค่าความเชื่อมั่นจะคำนวณได้จาก Intersection over Union ระหว่าง Predicted box กับ Ground truth box

ในแต่ละ Bounding box จะมีการคำนวณค่าทั้งหมด 5 ค่า นั่นคือ x, y, w, h และ Objectness confidence (ค่าความเชื่อมั่นในการที่จะมี object ของ box) ซึ่งค่าต่างๆ ได้แสดงไว้ในตาราง 2.1

ตาราง 2.1 ค่าที่ใช้ในการคำนวณ Bounding Box

ค่าต่างๆ	ความหมาย
(x, y)	แสดงคู่อันดับของตำแหน่งหรือกริดกึ่งกลางของ Bounding box โดยคำนวณเทียบกับขนาดความกว้างและความยาวของทั้งภาพ
w, h	แสดงค่าความกว้างและความสูงเทียบกับทั้งภาพ ตามลำดับ
ค่าความเชื่อมั่น (Objectness confidence)	คำนวณค่า IOU ระหว่าง Predicted box และ Ground truth

ยิ่งกว่านั้น ในแต่ละกริดยังมีการคำนวณในส่วนของค่าความน่าจะเป็นอย่างมีเงื่อนไขของแต่ละคลาสจำนวน C classes โดยใช้สัญลักษณ์ $\Pr(\text{Class}|\text{Object})$ แทนค่าความน่าจะเป็นนี้ ซึ่งจะมีชุดของความน่าจะเป็นหนึ่งชุดสำหรับแต่ละกริด โดยไม่ขึ้นกับจำนวนของ boxes B

$$\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{pred}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{pred}^{\text{truth}} \quad (2.1)$$

ซึ่งสมการข้างต้นจะบ่งบอกถึงค่าความเชื่อมั่นในคลาสที่สนใจในแต่ละ box ซึ่งมีทั้งค่าความน่าจะเป็นของการมีอยู่คลาสนั้นๆ และค่าที่แสดงถึงความแม่นยำในการ predict box สำหรับ object นั้น และสำหรับ YOLO ในเวอร์ชันนี้จะใช้ $S = 7, B = 2$ โดย Dataset จาก PASCAL VOC ซึ่งมีคลาสทั้งหมด 20 คลาส ดังนั้น $C = 20$ ทั้งนี้ Output ที่ได้ออกมาจะเป็น tensor ขนาด $7 \times 7 \times 30$ ซึ่ง 30 มาจาก $B \times C$ โดยมีค่าของ $B=2, C=20$ ในที่นี้นั่นเอง

YOLO ได้ประยุกต์ใช้ Convolutional Neural Network (CNN) โดยทดสอบบน PASCAL VOC dataset โดยจะมีเลเยอร์ของ CNN จำนวน 24 เลเยอร์ทำหน้าที่ดึง features ออกมา

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

มีค่าความเชื่อมั่นเป็นศูนย์ เข้ามามีบทบาทในการคำนวณ ทำให้โมเดลสับสนและอาจทำให้โมเดลหยุดไประหว่างการเทรน

การบรรเทาปัญหาในจุดนี้ลง อาจทำได้โดยการเพิ่มความสำคัญของค่า Bounding box coordinate prediction และลดความสำคัญของค่าความเชื่อมั่นที่ predict ได้จาก box ที่ไม่มี object อยู่ เข้าไปใน loss function โดยใช้ $\lambda_{\text{coord}} = 5$ และ $\lambda_{\text{noobj}} = 0.5$ เป็นค่า weight ที่ใช้

นอกจากปัญหาที่พบข้างต้นของ sum-squared error แล้วยังพบว่า การคำนวณ loss ด้วยวิธีการนี้ยังนำ error ที่เกิดบน Bounding box ที่มีขนาดใหญ่หรือเล็กมาคำนวณด้วย weight เท่าๆกัน จึงได้แก้ปัญหานี้บางส่วน ด้วยการใส่ค่ารากที่สองของความกว้างและความสูงของ Bounding box แทนที่จะใช้ขนาดเต็ม

การคาดการณ์ของ YOLO จะทำอยู่บน Bounding box หลายๆกล่องในแต่ละเซลล์กริด ซึ่งในการเทรนโมเดลจะมีการคำนวณหา IOU ที่มีค่ามากที่สุดเพื่อให้เป็นตัว predictor สำหรับแต่ละ object เพียงกล่องเดียว ในการเทรนโมเดลจะต้องทำการ optimize loss function ในพจน์ดังนี้

$$\begin{aligned} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (2.3) \end{aligned}$$

โดยที่แต่ละพจน์ใน loss function จะมีค่าเมื่อมี object ในเซลล์กริด ยกเว้นพจน์ที่มี λ_{noobj} ซึ่งจะคำนวณค่าในกรณีที่เซลล์ไม่มี object โดยที่ index j แสดงถึง Bounding box ที่ j ที่มี object แสดงอยู่บนเซลล์กริดและสอดคล้องกับการ predict นั้น

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตรวจจับวัตถุครบทั้งภาพ ทั้งนี้แม้ว่าจะมีโครงสร้าง Neural network คล้ายกัน คือจะมีส่วนของการสร้าง feature map จากภาพ input ที่เข้าไป โดย based network แล้วจึงนำไปสู่กระบวนการในการคำนวณ detection และ Non-maximum suppression ต่อไป ยังมีส่วนที่แตกต่างกันในแง่ของการที่ SSD จะมีการปรับจาก YOLO ที่จะใช้ Fully connected layer หลังจากทำ featured map ให้เป็น Convolutional layer 3-4 ชั้นซึ่งจะช่วยให้สามารถทำ Multiple scale จาก featured map ที่ได้ (ในที่นี้คือจาก based network ที่เป็น VGG16) และมีความหลากหลายมากกว่า นอกจากนี้ การคำนวณ Loss function ของทั้งคู่ยังแตกต่างกัน (รวมไปถึงการคำนวณด้วยนิยามที่ต่างกันอีกด้วย เช่นการที่โครงสร้างแบบ SSD แบ่ง bounding box เป็น positive/negative class และการนำเอาฟังก์ชัน SmoothL1 ที่ไม่ใช่ pure sum of square loss เข้ามาใช้งานร่วมกับ sigmoid loss computation ด้วยเช่นกัน) ทั้งนี้ค่า loss ของระบบจะแบ่ง weight ตาม confidence loss (L_{conf}) และ localization loss (L_{loc}) ด้วย alpha และหารด้วยจำนวนกล่อง N (ซึ่งเป็นจำนวนของ default boxes ที่ matched กับ ground truth ถ้าเป็น 0 ก็จะทำให้ force ให้ loss = 0) ดังนี้

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (2.4)$$

สำหรับในส่วนของ L_{loc} จะมีการนำ SmoothL1 loss สำหรับ prediction (l) และ ground truth (g) มาคำนวณ รวมไปถึงค่า d ที่แสดงถึงค่า bounding box ที่คำนวณมาเป็น default boxes (จากเลเยอร์ของ CNN ที่ใช้แตกต่างจากส่วนของ YOLO ดังที่กล่าวไปข้างต้น ซึ่งจะได้อีกหลายกล่อง) โดยฟังก์ชัน loss ข้างต้นจะมีการกล่าวถึงต่อไป

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \left\{ \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_j^m) \right\}$$

$$\hat{g}_j^{cz} = \frac{(g_i^{cz} - d_u^{cz})}{d_i^a}, \text{ for } z = x, y \text{ and } a = w, h \text{ respectively}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา หรืออ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

โดยค่า Confidence loss (Lconf) จะใช้การคำนวณในส่วนของ sigmoid sum ในสมการด้านขวาของรูปทางด้านล่างนี้ ก่อนจะนำไปคำนวณ contribution ในสมการทางซ้าย

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^0), \text{ where } \hat{c}_i^p = \frac{e^{c_i^p}}{\sum_p e^{c_i^p}} \quad (2.6)$$

โดยที่ SSD ได้นำฟังก์ชัน SmoothL1 loss ซึ่งจะเป็น ในลักษณะดังรูปทางด้านล่าง คือจะ คล้ายกับการมี threshold (ในที่นี้คือ beta) ที่จะแบ่งว่าฟังก์ชันจะให้ output เป็นอะไร ระหว่าง squared error term กับ absolute-wise error นั่นคือทำให้ได้ค่า ออกมาผสมผสานกันระหว่าง sum of square (ที่ใช้ใน YOLO loss) กับ linear absolute function ดังนี้

$$\text{loss}(x, y) = \frac{1}{n} \sum_i z_i; z_i = \begin{cases} \frac{1}{2\beta} (x_i - y_i)^2, & \text{if } |x_i - y_i| < \beta \\ |x_i - y_i| - \frac{1}{2}\beta, & \text{otherwise} \end{cases} \quad (2.7)$$

2.3 ทฤษฎีสี

ในปัจจุบันการจำแนกของสีต่างๆ ได้มีจัดจำแนก และแยกแยะสีโดยใช้เกณฑ์ที่แตกต่างกัน ซึ่ง โดยทั่วไปจะมีแนวคิดในการกำหนดเกณฑ์ของสีไว้คล้ายๆกัน คือ การกำหนดสีด้วยการแทนจุด ของสีในระนาบ 3 มิติ เพื่อที่จะเป็นการแทนตำแหน่งของพิกัดสีนั้นๆ เช่น RGB (Red, Green, Blue) HLS (Hue Lightness Saturation) และ HSV (Hue Saturation Value) ซึ่งในโครงงานนี้จะสนใจ ระบบสี HSV เป็นหลัก

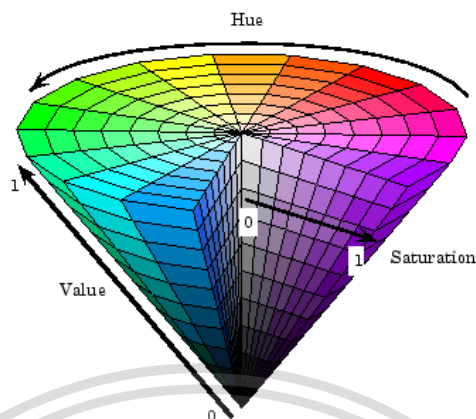
2.3.1 ระบบสี HSV (Hue Saturation Value)

ระบบสีนี้ประกอบไปด้วย 3 ค่า คือ ค่า Hue เป็นค่าสีหลัก โดยจะมีค่าอยู่ระหว่าง 0 – 360 เพื่อแทนสีต่างๆ เช่น สีแดง สีเหลือง สีเขียว เป็นต้น ค่า Saturation เป็นค่าสีที่จะอธิบายค่าความ อิ่มตัวของสี ซึ่งค่าจะอยู่ระหว่าง 0-1 โดยค่า 0 จะเป็นสีขาวซึ่งจะไม่มีค่าสี Hue อยู่เลย ส่วนค่า 1 จะ เป็นสีหลัก หรือแสดงค่าสี Hue ชัดเจนไม่มีสีขาวปน และค่าสุดท้ายคือ ค่า Value เป็นค่าที่รวมกับค่า

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Saturation ใช้ในการอธิบายความสว่าง หรือความเข้มของสี ค่าอยู่ระหว่าง 0 – 1 โดยค่า 0 คือ สีดำ ส่วนค่า 1 สว่างที่สุดและจะแสดงค่าสีนั้นชัดเจน



รูป 2.7 ระบบสี HSV

ค่าระบบสี HSV สามารถอ้างอิงมาจากค่าสี RGB ได้ โดยที่จะใช้สูตรในการแปลง คือ เริ่มจากการแปลงค่า R, G, B โดยที่จะแปลงค่า ที่อยู่ระหว่าง 0 – 255 ให้อยู่ระหว่าง 0- 1 โดยการ ทำ Normalization ด้วยหารด้วย 255 แล้วเก็บไว้ในค่าต่างๆ ดังต่อไปนี้

$$\begin{aligned}
 R' &= R/255 \\
 G' &= G/255 \\
 B' &= B/255 \\
 C_{max} &= \max(R', G', B') \\
 C_{min} &= \min(R', G', B') \\
 \Delta &= C_{max} - C_{min}
 \end{aligned}$$

(2.8)

จากนั้นจะนำค่าที่ได้ข้างต้นไปคำนวณค่า Hue(H)

$$H = \begin{cases} 0^\circ & , \Delta = 0 \\ 60^\circ \left(\frac{G' - B'}{\Delta} \pmod{6} \right) & , C_{max} = R' \\ 60^\circ \left(\frac{B' - R'}{\Delta} + 2 \right) & , C_{max} = G' \\ 60^\circ \left(\frac{R' - G'}{\Delta} + 4 \right) & , C_{max} = B' \end{cases}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

และทำการคำนวณค่า Saturation(S) ได้ดังนี้

$$S = \begin{cases} 0 & , C_{max} = 0 \\ \frac{\Delta}{C_{max}} & , C_{max} \neq 0 \end{cases} \quad (2.10)$$

โดยค่า Value(V) จะเท่ากับ Cmax และเมื่อได้ค่าทั้ง 3 ออกมาแล้ว ทำการนำเข้าสู่การคำนวณหา Distance ระหว่างค่าที่วัดได้ต่อค่าที่จัดกลุ่มไว้ (เพื่อให้สามารถคำนวณสีที่ใกล้เคียงออกมาได้) ซึ่งได้ใช้การคำนวณระยะทางแบบ L2 Norm ดังนี้

$$d = \|(dH, dS, dV)\|_2 = \left(\sum_{j \in \{dH, dS, dV\}} j^2 \right)^{\frac{1}{2}} \quad (2.11)$$

2.4 เทคโนโลยีที่ใช้

2.3.1 Google Colaboratory หรือ Colab

Google Colab เป็นเครื่องมืออย่างหนึ่งที่ใช้ในการเขียน และประมวลผลคำสั่งต่างๆ ในภาษา Python คล้ายกับ Jupyter Notebook ที่ทำงานบน Local computer ส่วน Google Colab จะทำอยู่บนบริการ cloud ของ Google ทำให้สามารถใช้งาน Hardware ต่างๆของ google ได้ไม่ว่าจะเป็น GPUs หรือ TPUs เหมาะสำหรับการทำงานเกี่ยวกับทาง Data Analysis, Machine Learning, Deep Learning



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และลิขสิทธิ์ของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.3.2 LabelImg

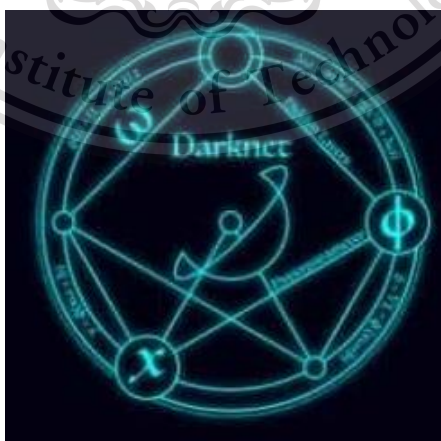
เป็นเครื่องมือสำหรับการติดกรอบวัตถุที่ต้องการ หรือ Image Annotation ซึ่งพัฒนาจากภาษา Python สามารถเก็บข้อมูลของ Annotation File ได้ 2 รูปแบบ คือ XML file และ Yolo format พัฒนาโดย MIT License



รูป 2.9 โลโก้ของ LabelImg

2.3.3 Darknet

เป็น Open source Neural Network framework ที่เขียนขึ้นมาโดยใช้ภาษา C และ CUDA ทำให้สามารถทำงานได้เร็ว และง่ายต่อการติดตั้ง ซึ่งยังสามารถทำงานได้ทั้ง CPU และ GPU โดย Darknet สร้างขึ้นมาเพื่อใช้งานกับ YOLO โดยเฉพาะ



รูป 2.10 โลโก้ของ Darknet

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.3.4 LINE Notify

เป็นบริการหนึ่งของ LINE ที่จะให้สามารถส่งข้อความ รูปภาพ หรือสติ๊กเกอร์ เข้าไปในช่องแชท หรือ กลุ่มใน LINE Application ได้ ผ่านทาง API ของ LINE ได้ ซึ่งในโครงการนี้ ได้นำ Line Notify มาใช้สำหรับในการแสดงผลลัพธ์ของระบบ



รูป 2.11 โลโก้ของ Line Notify

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 3

การออกแบบและพัฒนาระบบ

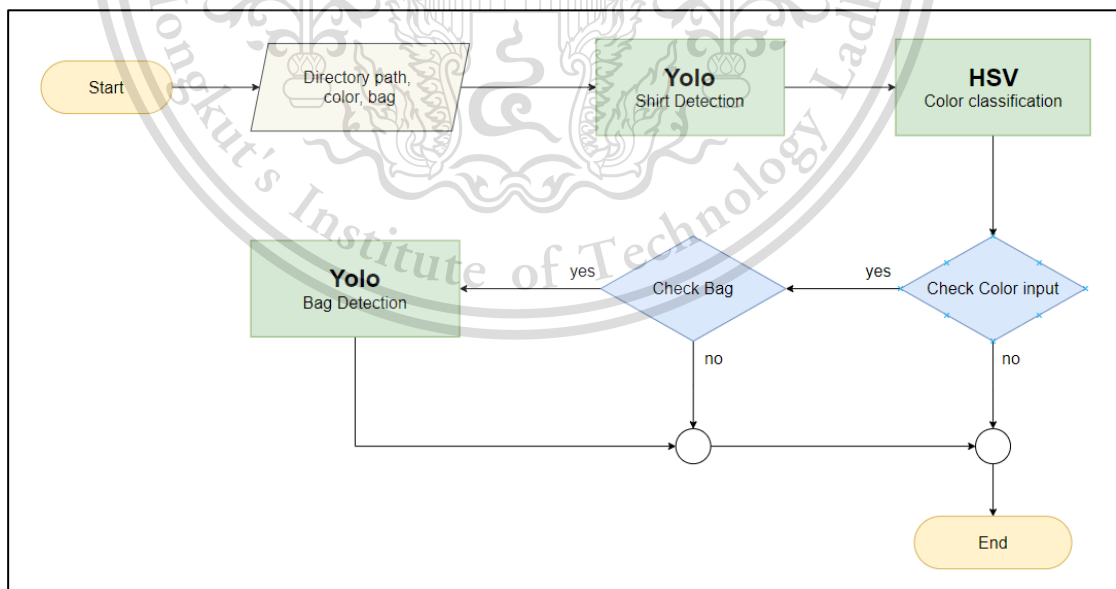
3.1 ภาพรวมของระบบ

3.1.1 Functional ของระบบ

- สามารถคัดกรองโดยใช้รูปภาพ วิดีโอ หรือ Directory ได้
- สามารถเลือกสีเสื้อที่ต้องการให้ระบบแสดงได้
- สามารถแสดงผลวิดีโอที่มีสีเสื้อที่ต้องการได้
- สามารถแสดงผลช่วงเวลาในวิดีโอที่มีสีเสื้อที่ต้องการได้
- คัดกรองกระเป๋าที่อยู่ภายในวิดีโอได้

3.1.2 การทำงานของระบบคัดกรองบุคคลในวิดีโอ

ระบบในการตรวจจับบุคคลที่สนใจจากสีเสื้อและกระเป๋า ได้ทำการออกแบบการทำงาน of ระบบออกเป็นทั้งหมด 3 ส่วน คือ ส่วนในการตรวจจับตำแหน่งของเสื้อและกระเป๋า ส่วนระบุสีเสื้อ และส่วนในการแสดงผลข้อมูล

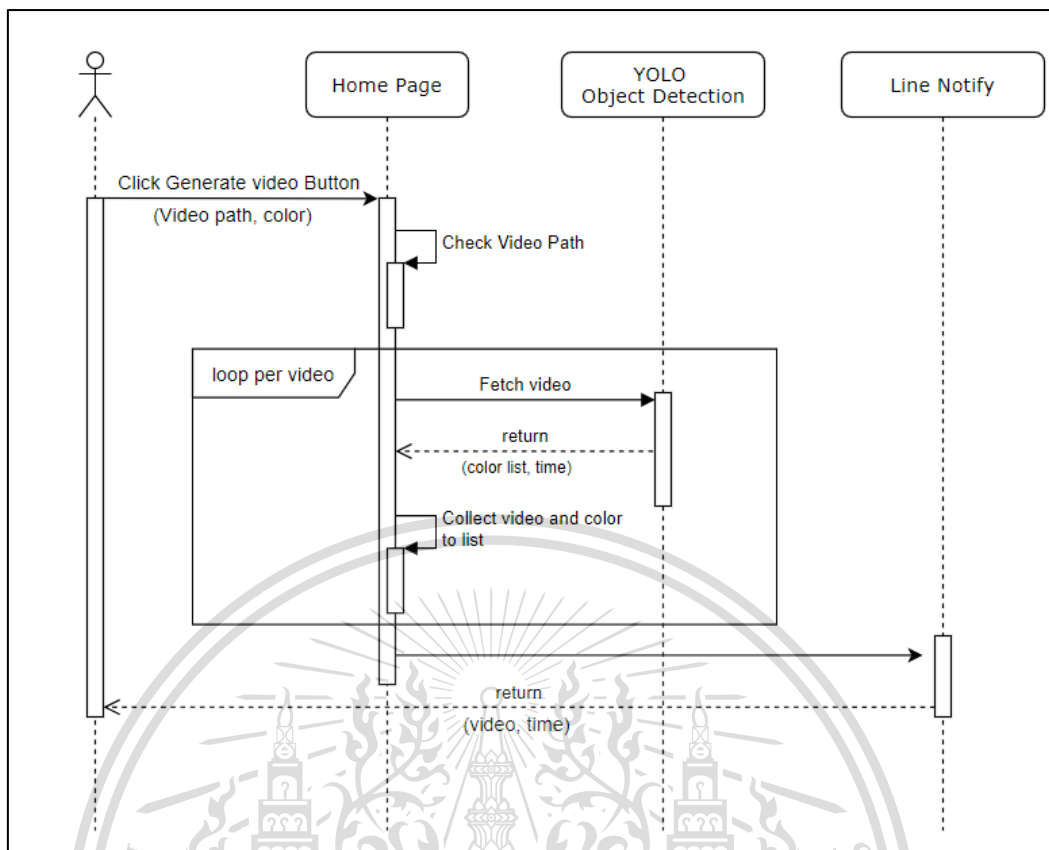


รูป 3.1 ภาพรวมการทำงานของระบบ

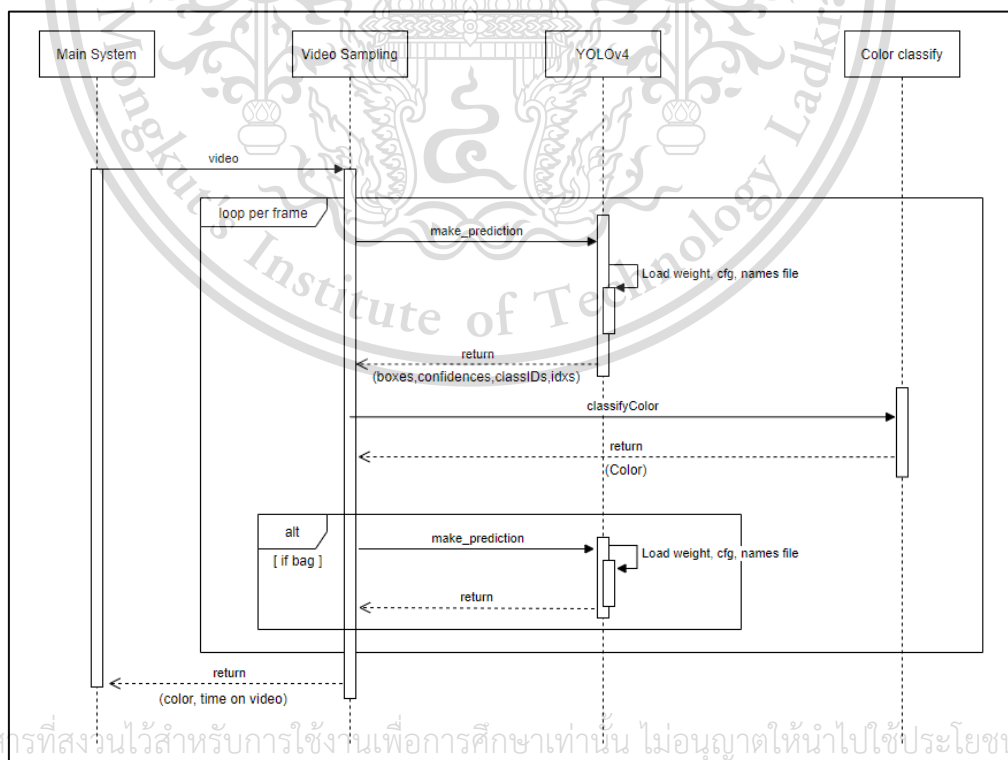
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.2 Sequence Diagram การทำงานของระบบตรวจจับและกระเป่า



รูป 3.3 Sequence Diagram การทำงานของระบบภายในของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งรูป 3.3 Sequence Diagram การทำงานของระบบภายในของระบบ ครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.1.2.1 ส่วนในการตรวจจับตำแหน่งของสี่

จากรูป 3.2 ในการใช้งานระบบ ผู้ใช้จะต้องใส่ Directory path เข้ามา หลังจากนั้นระบบจะทำการค้นหาวิดีโอใน Directory นั้น จากนั้นจะนำวิดีโอที่มีใน Directory ไปทำการประมวลผลทีละวิดีโอต่อในรูป 3.3 ซึ่งในส่วนนี้จะมีการทำ Multiprocessor เพื่อให้ระบบประมวลผลทีละหลายวิดีโอ หลังจากนั้นแต่ละ Process จะทำการประมวลผลวิดีโอโดยการตัดออกเป็นเฟรมย่อยๆ แล้วเข้าโมเดล YOLO เพื่อตรวจจับเพื่อหาตำแหน่งของสี่ภายในเฟรมนั้น

3.1.2.2 ส่วนในการระบุสี่

ซึ่งหลังจากที่ได้ตำแหน่งของสี่มาแล้ว จะทำการนำ Bounding Box ที่ได้มาทำการหาค่าของสี่เพื่อทำการระบุสี่ของสี่ที่ตรวจจับได้ ซึ่งการหาสี่ของสี่ จะใช้วิธีในการคำนวณค่าสี่ภายในกล่อง Bounding Box แล้วทำการหาค่าเฉลี่ย จากนั้นเข้าสู่ Function ที่เขียนขึ้นมาสำหรับการหาค่าสี่ที่มี Distance น้อยที่สุด ซึ่งก็จะระบุว่าเป็นสี่นั้น

3.1.2.3 ส่วนในการตรวจจับกระเป๋า

และถ้าหากผู้ใช้ระบบต้องการที่จะคัดกรองกระเป๋าด้วยจะทำการตรวจจับกระเป๋าจากเฟรมที่เจอสี่สี่นั้นๆ เพิ่มเติมเข้าไปด้วย โดยจะใช้ Pre-trained model จาก YOLOv4 มาใช้ในการตรวจจับกระเป๋า สำหรับ Feature หนึ่งของระบบ และบุคคล เพื่อการทดสอบ Architecture ของระบบ ทั้งนี้ยังได้มีการใช้งานในส่วน of Model ที่ได้ทำการฝึกสอนเองสำหรับการตรวจจับสี่ด้วยชุดข้อมูลที่มีขนาดใหญ่มากขึ้น

3.1.3 การแสดงผล

หลังจากที่ผ่านการจำแนกในส่วนแรกเสร็จเรียบร้อยแล้ว จะแสดงชื่อของไฟล์วิดีโอและแสดงช่วงเวลาที่มียุคคลตามสี่สี่และกระเป๋าที่ต้องการลงบนหน้าจอ command line ทั้งนี้จะมีการส่งข้อมูลไปยัง Line Notify เพื่อเพิ่มความสะดวกอีกด้วย

3.2 ขั้นตอนการทำโครงการ

3.2.1 การเตรียมข้อมูลสำหรับ train โมเดล

1.) การสร้างกรอบระบุตำแหน่งของสี่ภายในภาพและการระบุคลาส

ในการฝึกสอนโมเดล จะต้องนำภาพที่ได้หามา ทำการระบุตำแหน่งของวัตถุที่สนใจ ซึ่งส่วนหนึ่งในโครงการนี้ก็คือ ตำแหน่งของสี่สี่ ซึ่งได้ใช้โปรแกรม LabelImg ซึ่งเป็น

Open source ที่ใช้สำหรับการจัดทำชุดข้อมูล ในการดีกรอบวัตถุ และระบุคลาสที่ต้องการ เช่น คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปเผยแพร่บนสื่อออนไลน์

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ไฟล์ที่ได้ออกมาจากโปรแกรม LabelImg จะเป็นไฟล์ .txt ซึ่งชื่อไฟล์จะเป็นชื่อเดียวกับชื่อรูปภาพ และภายในไฟล์จะเป็นดังรูป 3.2

0	0.240625	0.543981	0.050000	0.113889
6	0.551823	0.444907	0.051562	0.100926
0	0.585938	0.515741	0.051042	0.105556
0	0.652344	0.727315	0.071354	0.171296
6	0.064844	0.943981	0.075521	0.112037
2	0.173698	0.828704	0.120313	0.222222

รูป 3.4 รูปแสดงข้อมูลภายในไฟล์ .txt

โครงสร้างภายในไฟล์ .txt ในรูป 3.2 สำหรับแต่ละ Bounding box (แต่ละแถวในไฟล์)จะประกอบไปด้วย Attribute ทั้งหมด 5 Attributes ที่แยกจากกันด้วยช่องว่างได้แก่

- i. Class number
- ii. อัตราส่วนระหว่างตำแหน่ง x ที่จุดกึ่งกลางของ Bounding box ต่อความกว้างของรูป
- iii. อัตราส่วนระหว่างตำแหน่ง y ที่จุดกึ่งกลางของ Bounding box ต่อความสูงของรูป
- iv. อัตราส่วนระหว่างความกว้างของ Bounding box ต่อความกว้างของรูป
- v. อัตราส่วนระหว่างความสูงของ Bounding box ต่อความสูงของรูป

2.) การทำ Augmentation

หลังจากการระบุตำแหน่งของเคือโดยการสร้างกล่องระบุตำแหน่งแล้ว เพื่อที่จะเตรียมข้อมูลให้เพียงพอสำหรับการเทรนได้ จะนำ Dataset ที่ได้เตรียมไว้ก่อนหน้านี้มาทำการเพิ่มจำนวนของ Dataset โดยการปรับรูปเป็นหลายๆแบบ เช่น ย่อ, ขยาย, หมุน/ซ้าย/ขวา เป็นต้น เพื่อให้ Dataset มีปริมาณและมีความหลากหลายมากขึ้น

3.2.2 การฝึกสอนโมเดล

ในการฝึกสอน YOLO model ของเสือนั้น จะใช้ Darknet สำหรับการใช้ในการสร้าง neural network ของ Yolo และใช้ Google Colab ในการฝึกสอนโมเดล ซึ่งจะทำการนำ Pre-trained model จาก yolov4.weight มาทำการตัด layer ท้ายออกและทำการฝึกสอนโมเดล โดยจำเป็นที่จะต้องทำการ configuration บางอย่าง ได้แก่

- 1) นำข้อมูล text file ที่ได้ผ่านการติกรอบตำแหน่งของเคือมารวบรวมเป็น list ของ text file สำหรับการใช้ในการเทรนโมเดล YOLO v4

2) ปรับจูนค่า config ในไฟล์ .cfg ที่ได้เตรียมไว้ ให้สอดคล้องกับจำนวนของคลาสที่ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีเหตุตเปลี่ยนแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้สนใจ ซึ่งในที่นี้ก็คือ

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- Batch บอกจำนวนภาพและ text file ที่โหลดพร้อมกัน ในการทำการ forward pass และ backpropagation ในที่นี้คือ 64
 - Subdivisions เป็นจำนวนที่ batch ถูกแบ่งเป็น blocks สำหรับ Parallel Computing ในที่นี้คือ 32
 - Max_batch เป็นจำนวนมากกว่าหรือเท่ากับ จำนวนคลาส*2000 และไม่น้อยกว่าจำนวนของ dataset
 - Steps ซึ่งเป็นส่วนสำหรับการปรับแต่ง learning rate ของโมเดลให้เหมาะสมมีค่า 80% และ 90% ตามลำดับ
 - Width and Height ของ network เป็นตัวเลขที่หารด้วย 32 ลงตัว ในที่นี้คือ 608 และ 608 ตามลำดับ
 - Classes ปรับให้เท่ากับจำนวน class ของจำนวนสีเสื้อ นั่นก็คือ 10 ดังที่กล่าวถึงในหัวข้อ การจำแนกสีเสื้อ
 - Filters ใน Convolutional layer ก่อนหน้า Yolo layer ให้เท่ากับ $(classes + coords + 1) * (number\ of\ mask) = (Classes + 5) * 3$ ซึ่งในที่นี้เท่ากับ 45
- 3) นำ pre-trained weight (yolov4.conv.137) มาใช้เป็นค่า weight เริ่มต้นสำหรับการ train ใน custom dataset ที่จัดทำขึ้น
- 4) ทำการ train จนกระทั่ง ค่า loss ไม่มีความเปลี่ยนแปลงแบบมีนัยสำคัญ และไม่แกว่งขึ้นสูงอย่างผิดปกติซึ่งเป็นสัญญาณของการเกิด Overfitting ของโมเดล

3.2.3 การใช้งานโมเดลตรวจจับเสื้อ และกระเป๋

หลังจากที่ทำการฝึกสอนเสร็จเรียบร้อยแล้ว จะนำโมเดลที่ได้ทำการฝึกสอน โมเดลนั้นมาใช้งาน โดยที่จะใช้ Open-CV ในการเรียกใช้งาน weight ของ Yolo

3.2.4 การจำแนกสีเสื้อ

การจำแนกสีจะทำการจำแนกโดยการ นำ bounding box ที่ได้จากโมเดลตรวจจับเสื้อ มาทำการระบุสี โดยการนำไปหาค่าเฉลี่ยของสี RGB ภายในกรอบ แล้วนำเข้าไป Function ที่เขียนขึ้นมา ซึ่งจะคำนวณสีที่มีค่าของระบบสี HSV โดยใช้หลักการ Nearest Distance ซึ่งจะได้อ่านสีที่ใกล้เคียงที่สุดออกมา

โดยจำนวนสีที่ได้กำหนดไว้มีทั้งหมด 9 สี คือ Pink, Red, Orange, Yellow, Green, Blue, Brown, White, Black ซึ่งในแต่ละสี ได้กำหนดจุดของสี RGB ไว้ประมาณสีละ 5-10 ขึ้นอยู่กับความกว้างของสีนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 4

ผลการทดลอง

ในส่วนนี้จะเป็นการแสดงผลการทดสอบ การวิเคราะห์ และการประเมินผลลัพธ์ที่ได้ ในส่วน of ระบบต่างๆ ซึ่งในโครงการชิ้นนี้ได้ทำการแบ่งระบบออกเป็นทั้งหมด 3 ส่วน ได้แก่ การตรวจจับเสื้อ การตรวจหาสีเสื้อ และการตรวจจับกระเป๋า

4.1 การทดสอบโมเดลตรวจจับเสื้อ

ในการเทรนตำแหน่งของเสื้อกับ โมเดล ผู้จัดทำได้ใช้จำนวนรูปภาพในการเทรนทั้งหมด 2,000 รูป เป็นจำนวน 10,000 รอบ ซึ่งได้เทรนโมเดลโดยใช้ Pre-trained weights-file ชื่อ yolov4.conv.137 โดยผลลัพธ์ที่ได้ออกมาได้ดังนี้

4.1.1 ผลลัพธ์จากโมเดลที่เทรน

เมื่อนำข้อมูลที่ได้เตรียมไว้มาทดลองกับ โมเดลที่ได้ฝึกสอนไว้จะได้ผลลัพธ์ออกมาดังตาราง 4.1 ซึ่งเป็นภาพที่ได้ จะเป็นภาพที่ทำการ Bounding box ตามที่โมเดลได้ทำนายไว้ และจะได้ตำแหน่ง Bounding Box ของเสื้อคนในแต่ละรูปเอาไว้ ซึ่งส่วนนี้จะนำไปใช้

ผลที่ได้จากการเทรนโมเดลด้วย Dataset ที่เตรียมไว้ โดยได้ทำการทดสอบโดยใช้รูปทั้งหมด 2000 รูป ซึ่งจะได้ผลลัพธ์ออกมาดังตาราง 4.1 และได้นำโมเดลมาทำการหาค่า Validation มาสรุปเป็นตาราง Confusion Matrix ลงในตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.1 ผลลัพธ์จากโมเดลตรวจจับตำแหน่งของเสื้อ

รูปภาพต้นฉบับ	รูปภาพที่ได้จากการทดสอบ โมเดล	เวลาที่ใช้ในการ ทดสอบ (วินาที)	ค่า confidence
		0.6482	0.9992
		0.8127	0.8280, 0.8832
		0.4213	0.9936

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหา และอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.2 ค่า Confusion Matrix ของการเข้าโมเดลตรวจจับตำแหน่งของเสื้อ

Predicted \ Actual	Positive	Negative
True	TP = 164.76	FP = 6.92
False	FN = 97.24	-

4.1.2 การวัดผลโมเดลที่เทรนตามโครงสร้าง Model

มีการทดสอบการทำงานของระบบ โดยการได้ทำการออกแบบโปรแกรมไว้เป็น 2 โครงสร้าง คือ

- 1.) One layer system คือ โครงสร้างที่จะมีชั้นของโมเดลตรวจจับเสื้ออยู่ชั้นเดียว โดยจะทำการตรวจจับเสื้อจากวัตถุทั้งหมดในภาพในคราวเดียว
- 2.) Two-layer system คือ โครงสร้างที่จะมีชั้นของโมเดลสองชั้น คือชั้นแรก จำนำเฟรมนั้นมาทำการตรวจจับบุคคลก่อนในชั้นแรก และในชั้นต่อมา จะนำตำแหน่ง และกรอบของบุคคลที่ได้จากการตรวจจับได้ มาเข้าโมเดลสำหรับการตรวจจับสีเสื้อในชั้นถัดมา

4.1.2.1 ผลจากการทดสอบโมเดลด้วยการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ

ตาราง 4.3 ผลลัพธ์จากการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ

รูปภาพต้นฉบับ	รูปภาพที่ได้จากการทดสอบโมเดล	เวลาที่ใช้ในการทดสอบ (วินาที)	ค่า confidence
		8.4426	0.9821

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และใช้ในงานเพื่อการวิจัยเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.3 ผลลัพธ์จากการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ(ต่อ)

รูปภาพต้นฉบับ	รูปภาพที่ได้จากการทดสอบ โมเดล	เวลาที่ใช้ในการ การทดสอบ (วินาที)	ค่า confidence
		7.8115	0.9936, 0.9844
		9.2918	0.9973

ตาราง 4.4 ค่า Confusion Matrix ของการเข้าโมเดลตรวจจับคนก่อน แล้วเข้าโมเดลตรวจจับเสื้อ

Predicted \ Actual	Positive	Negative
True	TP = 196.01	FP = 260.96
False	FN = 65.99	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

4.1.2.2 ผลจากการทดสอบโมเดลโดยนำรูปเข้าโมเดลตรวจจับเสื้อเท่านั้น

ผลลัพธ์ในส่วนนี้จะได้ผลลัพธ์เช่นเดียวกับข้อ 4.1.1 โดยผลลัพธ์การตรวจจับเสื้อคือตาราง 4.1.1.1 และผลของค่า Confusion Matrix คือตาราง 4.1.1.2 เพราะใช้ Architecture เดียวกัน

4.1.2.3 สรุปผลจากการทดสอบทั้ง 2 ระบบ

ตาราง 4.5 เปรียบเทียบประสิทธิภาพระหว่าง One-layer system และ Two-layers system

	One-layer system	Two-layers system
Precision	0.96	0.43
Recall	0.63	0.75
F1-Score	0.76	0.55
Process time	533 s	2509 s

เมื่อสังเกตจากการทดลองทั้ง 2 Systems จะสังเกตได้ว่าค่า Confidence ของแต่ละกรอบ ที่ได้ออกมาจะค่อนข้างใกล้เคียงกัน และเมื่อดูที่ค่า Precision แล้ว One-layer แสดงผลคิดค่าออกมาได้ดีกว่าถึง 0.53 ส่วน two-layer จะได้ค่าของ Recall ที่มากกว่าเพียงเล็กน้อย ประมาณ 0.12 แต่เมื่อเมื่อคิดจากค่า F1-score ซึ่งเป็นค่าเฉลี่ยระหว่าง Precision กับ recall แล้ว จะเห็นได้ว่า One-layer ให้ผลที่ดีกว่าถึง 0.21 และรวมกับเวลาที่ใช้ในการทดสอบ ในส่วนของ two-layer นั้น จะใช้เวลาในการทดสอบค่อนข้างนานกว่า One-layer ซึ่งเป็นผลมาจากถ้าในเฟรมนั้นๆ หากมีคนหลายๆ จะทำให้ต้องตรวจจับคนหลายรอบกว่าจึงทำให้ระบบยิ่งประมวลผลนานขึ้น ซึ่งจากผลการทดลองนี้ คณะผู้จัดทำจึงเลือกที่จะใช้ระบบที่จะทำการตรวจจับเพียงครั้งเดียวคือ ตรวจจับเสื้อคนจากเฟรมเต็มเลย ซึ่งได้ค่าความแม่นยำใกล้เคียงกัน และได้ผลที่เร็วกว่าการตรวจจับคนก่อน

4.2.3 การวัดผลโมเดลที่เทรนตามโครงสร้าง Network

คณะผู้จัดทำได้ทำการทดสอบ โดยใช้โครงสร้างของ Network ไว้ 2 แบบ คือ Yolo และ SSD เพื่อที่จะเปรียบเทียบประสิทธิภาพของแต่ละ โมเดล ในการตรวจจับเสื้อ ซึ่งได้ผลการทดลองดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.


ตาราง 4.6 เปรียบเทียบประสิทธิภาพระหว่าง Yolo และ SSD

	YOLO	SSD
Precision	0.96	0.52
Recall	0.63	0.69
F1-Score	0.76	0.59

4.2 การทดสอบการตรวจหาสีเสื้อ

หลังจากที่ได้ตำแหน่งของเสื้อจากโมเดลในข้อ 4.2 จะนำตำแหน่งนั้นไปเข้าโมเดลในการตรวจหาสีเสื้อต่อไป ซึ่งจะทำได้โดยวิธีการหาค่าเฉลี่ยสีในกรอบของเสื้อ จากนั้นจะนำค่า function ที่ได้เขียนไว้เพื่อคำนวณค่าสี โดยจะใช้หลักการของ Nearest Distance ผ่านค่าสีของ HSV เพื่อหาสีที่ใกล้เคียงที่สุดออกมา

ตาราง 4.7 ตัวอย่างสีเสื้อที่ได้จากการเข้า function ระบุสี

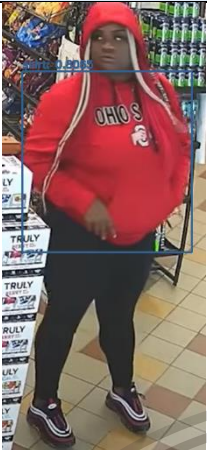
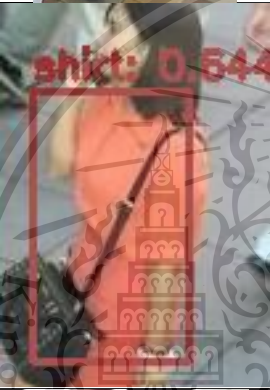

เสื้อที่ตรวจจับได้	Classes สีที่ได้จากการเข้า Function
	Pink

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.7 ตัวอย่างสีเสื้อที่ได้จากการเข้า function ระบุสี (ต่อ)


เสื้อที่ตรวจจับได้	Classes สีที่ได้จากการเข้า Function
	Red
	Orange
	Yellow

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.7 ตัวอย่างสีเสื้อที่ได้จากการเข้า function ระบุสี (ต่อ)

เสื้อที่ตรวจจับได้	Classes สีที่ได้จากการเข้า Function
	Green
	Blue
	Brown
	White

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.7 ตัวอย่างสีเสื้อที่ได้จากการเข้า function ระบุสี (ต่อ)

เสื้อที่ตรวจจับได้	Classes สีที่ได้จากการเข้า Function
	<p style="text-align: center;">Black</p>

4.3 การทดสอบโมเดลตรวจจับกระเป๋า

ส่วนสุดท้ายเป็นส่วนของการตรวจจับกระเป๋า ในการค้นหาแต่ละครั้ง ถ้าผู้ใช้ต้องการที่จะคัดกรองว่ามีกระเป๋าที่อยู่ภายในวิดีโอพร้อมกับคนที่ใส่เสื้อสีนั้นๆ อยู่ด้วย ในแต่ละ frame ของวิดีโอ ซึ่งได้ใช้ Pre-trained Model ของ YOLO เพื่อใช้ในการตรวจจับกระเป๋า โดยจะใช้เป็นตัวช่วยในการคัดกรองคนเพิ่มเติมจากการตรวจจับเสื้อของบุคคลในรูป ซึ่งในการคัดกรองจะใช้ class จำนวน 2 class ของ yolov4 คือ 'handbag' และ 'backpack' เป็นประเภทของกระเป๋าที่ค้นหา

4.3.1 ผลลัพธ์จากการทดสอบโมเดล

เนื่องจาก Pre-trained Model มีจำนวน class อยู่ทั้งหมด 80 classes แต่ในระบบของโครงการนี้จะสนใจเพียงแค่ handbag และ backpack เท่านั้น ซึ่งภายในตาราง 4.3.1 จะเป็นการแสดงรูปที่ Pre-trained Model สามารถตรวจจับได้ และเวลาที่ใช้ในการตรวจจับต่อเฟรม และค่าความมั่นใจ (confidence) ของแต่ละเฟรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

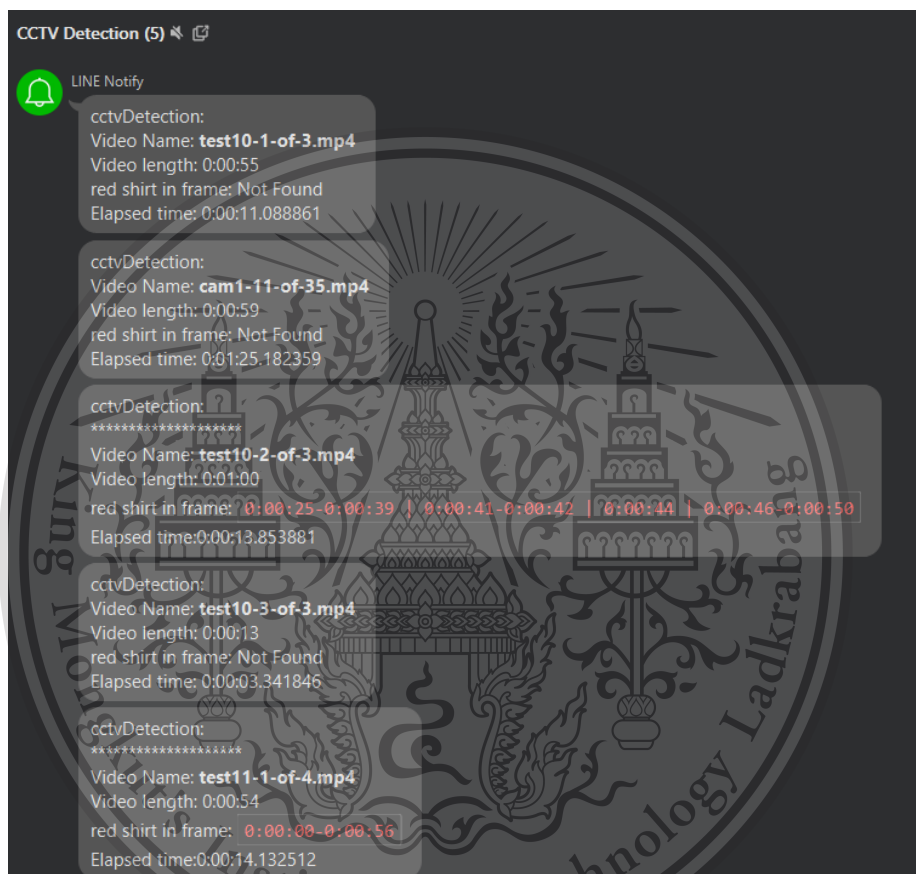
ตาราง 4.8 ผลลัพธ์จากการใช้ Pre-trained Model ในการตรวจจับกระเป๋า

รูปภาพต้นฉบับ	รูปภาพที่ได้จากการทดสอบ จาก class ต่างๆ	เวลาที่ใช้ในการ ทดสอบ (วินาที)	ค่า confidence
		Backpack & handbag 1.4130380153656006	0.713 0.7367
		handbag 1.6841223239898682	0.7034
		backpack 1.8113348484039307	0.8506

เอกสารนี้เป็นเอกสารสำหรับการใช้ภายในเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆที่ ให้ตัดแบ... อย่างยิ่งถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is for educational use only. not allowed for commercial use.

4.4 การแสดงผล

หลังจากที่แต่ละวิดีโอได้ทำการประมวลผลทั้งการตรวจจับเสื้อ ตรวจสอบหาสี หรือการตรวจหากระเป๋าสีเรียบร้อยแล้ว ได้ทำการส่งผลลัพธ์เข้า Group Line ผ่าน Line Notify โดยผลลัพธ์ที่แสดงจะแสดง ชื่อของวิดีโอ เวลาของวิดีโอทุกๆ เฟรมหรือช่วงเวลาที่เจอเสื้อสีนั้นๆ ในวิดีโอ และ เวลาที่ใช้ในการประมวลผลวิดีโอ



รูป 4.1 ผลลัพธ์ที่ส่งเข้า Line Notify

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 5

สรุปการดำเนินงานของโครงการ

5.1 สรุปผลที่ได้จากการทำโครงการ

จากการดำเนินงานของโครงการ การบ่งชี้วัตถุและบุคคลที่สนใจจากวิดีโอกล้องวงจรปิดในหลายแหล่งที่มา ทางผู้จัดทำได้ทดลองทำ และศึกษาเทคโนโลยีต่างๆ ที่จะต้องใช้ภายในระบบนี้ โดยการทำงานของระบบจะต้องใช้ Object Detection ซึ่งเป็นส่วนหนึ่งของ Neural Network เพื่อที่จะตรวจจับตำแหน่งของเสื้อของคนภายในวิดีโอ โดยที่ทางคณะผู้จัดทำได้เลือก YOLO Object Detection ในการนำมาใช้กับระบบ และได้มีการจัดเตรียม Dataset ต่างๆ ทั้งการ Label กรอบของวัตถุ การทำ Data Augmentation และ Data cleansing เพื่อเตรียมพร้อมสำหรับการฝึกสอนโมเดล และได้ทำการฝึกสอนโมเดลโดยใช้ Darknet ซึ่งเป็น tools สำหรับรัน Yolo ที่ใช้ GPU ในการรัน ซึ่งได้ฝึกสอนโมเดลโดยใช้ Google Colab จากนั้นเมื่อได้ค่าจากโมเดลในการตรวจจับเสื้อเรียบร้อยแล้วก็นำเข้าสู่ Function ที่ใช้ในการตรวจหาสี ซึ่งได้ใช้การคำนวณโดยใช้ค่าเฉลี่ย แล้วทำการหาค่าสีที่ใกล้ที่สุด โดยใช้วิธี Nearest Distance ซึ่งจะได้ออกค่าสีของตำแหน่งเสื้อนั้นๆ แล้วจึงเช็คว่าตรงกับค่าที่ Input มาหรือไม่ นอกจากนี้ ยังได้เพิ่มฟังก์ชันเสริมของระบบเข้าไป โดยการใช้ Pre-trained YOLOv4 Model เพื่อค้นหากระเป๋า ที่อยู่ในเฟรมที่มีเสื้อสีนั้นๆ ด้วย และยังได้มีการทำ Multiprocessing เพื่อให้ระบบสามารถตรวจหาได้หลายๆ วิดีโอภายในเวลาเดียวกัน

5.2 ปัญหาและอุปสรรค

ปัญหาและอุปสรรคต่างๆ ที่ได้เจอระหว่างการทำโครงการ ได้แก่

5.2.1 ปัญหาการหา Dataset

เนื่องด้วย Dataset ที่จะใช้ในการฝึกสอน โมเดล ทางคณะผู้จัดทำได้จัดทำขึ้นมาจากการรวบรวมข้อมูลเชิงรูปภาพจากแหล่งต่างๆ ในอินเทอร์เน็ต ซึ่งส่งผลให้รูปภาพที่ได้มามีความละเอียดพอสมควร แต่ไม่มีประสิทธิภาพเพียงพอ ในแง่ของการเป็น Dataset ที่ต้องการ ซึ่งจำเป็นที่จะต้องเป็นภาพจากอุปกรณ์กล้องวงจรปิดจริง

5.2.2 ปัญหาจากการฝึกสอนโมเดล

ในการฝึกสอนโมเดลตรวจจับเสื้อ สืบเนื่องจากปัญหาในข้อที่ 5.2.1 ทำให้เมื่อฝึกสอนโมเดลเสร็จเรียบร้อยแล้ว และเมื่อนำมาทดสอบกับ environment ใหม่ๆ จะทำให้โมเดลมีประสิทธิภาพที่ไม่ดีพอ

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

5.2.3 ปัญหาจาก Environment และ Resources ที่ใช้ในการทดสอบระบบ

เนื่องจากโมเดลที่ใช้ทดสอบเป็น โมเดลเกี่ยวกับ Neural Network ซึ่งต้องใช้ทรัพยากร ในการประมวลผลแต่ละภาพค่อนข้างมาก ดังนั้นในการทดสอบระบบ จึงทำให้ระบบประมวลผล ค่อนข้างช้า จึงส่งผลให้แต่ละวิดีโอต้องใช้เวลาในการประมวลผลเพิ่มมากยิ่งขึ้น

5.3 วิธีการในการแก้ปัญหาที่เกิดขึ้น

5.3.1 แนวทางการแก้ปัญหาเกี่ยวกับ Dataset

ทำการค้นหา Dataset จากอินเทอร์เน็ต และนำมาคัดกรองอีกที เพื่อคัด Dataset ที่ไม่ดี พอลออกไป จากนั้นทำการ Annotate บน Dataset ระบุกรอบตำแหน่งของวัตถุที่ต้องการ ซึ่งก็คือใน ส่วนของตำแหน่งสี่เหลี่ยม เพื่อที่จะใช้ในการฝึกสอนโมเดลต่อไป

5.3.2 แนวทางการแก้ปัญหาการฝึกสอนโมเดล

พยายามหาใช้ Dataset ที่มีความหลากหลายมาใช้ในการเทรน โมเดล เพื่อให้โมเดล สามารถทำนายได้เมื่ออยู่ใน environment ใหม่ได้มีประสิทธิภาพมากยิ่งขึ้น

5.3.3 แนวทางการแก้ปัญหา Environment

แก้ปัญหาด้วยการใช้ cloud service ต่างๆ ซึ่งจะมีทรัพยากรที่ให้เราในการประมวลผล และทำงานค่อนข้างมาก ซึ่งทางคณะผู้จัดทำได้ใช้บริการของ Colab ในการประมวลผลระบบต่างๆ แต่การใช้งาน Colab มีข้อจำกัดในการทำงาน เช่น ฟังก์ชันบางอย่างก็ไม่สามารถทำได้, syntax บางอย่างต้องเขียนเฉพาะใช้กับ Colab เท่านั้น หรือ การใช้งาน Resources บางอย่างของ Colab เช่น GPU จะถูกจำกัดต่อวัน

5.4 แนวทางการพัฒนาต่อ

นำเอาข้อมูลภาพ Dataset ของกล้องวงจรปิดในแต่ละจุดที่ต้องการติดตั้งระบบนี้ มาใช้ในการ ฝึกสอนโมเดล เพื่อให้ Dataset มีความใกล้เคียง Realistic Environment มากขึ้น ซึ่งจะส่งผลให้ ประสิทธิภาพและตรวจจับสิ่งของต่างๆได้มากยิ่งขึ้น นอกจากนี้ทางผู้จัดทำจะทำการ เพิ่มตัวคัดกรอง ของระบบ หรือจำนวน class เพิ่มเติม เช่น หมวก, แวนตา, หน้ากาก เป็นต้น เพื่อที่จะสามารถคัด กรองบุคคลภายในกล้องวงจรปิดให้อยู่ในวงที่แคบลงได้ดียิ่งขึ้น และทำการปรับแต่ง โปรแกรมให้ สามารถรันบน GPU ได้ซึ่งจะทำให้การทำงานของระบบมีประสิทธิภาพมากยิ่งขึ้น โดยเวลาที่ใช้ใน เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นับญาติหากไปใช้ประโยชน์ด้านการค้า การประมวลผลลดลง

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

Pjreddie. 2017. Darknet: Open Source Neural Networks. [online].

Available : <https://pjreddie.com/darknet/>

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. 2017. You Only Look Once: Unified, Real-Time Object Detection. [online].

Available : https://pjreddie.com/media/files/papers/yolo_1.pdf

Sam. 2020. How to train YOLOv4 for custom objects detection in Google Colab. [online].

Available : <https://medium.com/ai-world/how-to-train-yolov4-for-custom-objects-detection-in-google-colab-1e934b8ef685>

Techzizou. 2021. TRAIN A CUSTOM YOLOv4 OBJECT DETECTOR (Using Google Colab). [online].

Available : <https://medium.com/ai-world/how-to-train-yolov4-for-custom-objects-detection-in-google-colab-1e934b8ef685>

Adrian Rosebrock. 2018. YOLO object detection with OpenCV. [online].

Available : <https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/>

W3resource. 2020. Convert RGB color to HSV color. [online].

Available : <https://www.w3resource.com/python-exercises/math/python-math-exercise-77.php>

MIT license. 2018. LabelImg. [online].

Available : <https://github.com/tzatalin/labelImg>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Alexey Bochkovskiy. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. [online].

Available : <https://arxiv.org/pdf/2004.10934v1.pdf>

IBM Cloud Education. 2020. What are neural networks?. [online].

Available : <https://www.ibm.com/cloud/learn/neural-networks>

C Kuan. 2020. How to use OpenCV with GPU on Colab? [online].

Available : <https://towardsdatascience.com/how-to-use-opencv-with-gpu-on-colab-25594379945f>

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg. 2016. SSD: Single Shot MultiBox Detector [online].

Available : <https://arxiv.org/pdf/1512.02325.pdf>

Geeksforgeeks. 2019. Python | Data Augmentation [online].

Available : <https://www.geeksforgeeks.org/python-data-augmentation/>

Matus Tanonwong. 2020. Custom Object Detection Using Tensorflow in Google Colab [online]

Available : <https://medium.com/@matus.tanon/custom-object-detection-using-tensorflow-in-google-colab-e4d6e1a17f18>

Odemakinde Elisha. 2020. Real-time Object Detection using SSD MobileNet V2 on Video Streams [online]

Available : <https://heartbeat.fritz.ai/real-time-object-detection-using-ssd-mobilenet-v2-on-video-streams-3bfc1577399c>

Abhimanyu. 1990. SSD-Mobilenet-Custom-Object-Detector-Model-using-Tensorflow-2 [online]

Available : <https://github.com/abhimanyu1990/SSD-Mobilenet-Custom-Object-Detector-Model-using-Tensorflow-2>

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น เมื่อผู้รู้เห็นใช้เอกสารนี้เป็นการค้า
ไม่ว่ากรณีใดๆ หักสิทธิ์ให้มีเหตุเปลี่ยนแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Torch Contributors. 2019. SMOOTHL1LOSS [online].

Available : <https://pytorch.org/docs/stable/generated/torch.nn.SmoothL1Loss.html>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.