รายงานสหกิจศึกษาฉบับสมบูรณ์

การวิเคราะห์ข้อมูลลูกค้าด้วยภาพจากกล้องวงจรปิด

Customer Analysis Using CCTV Camera

นาย ฉัตรชัย แซ่ตัน

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อโครงการสหกิจศึกษา การวิเคราะห์ข้อมูลลูกค้าจากภาพกล้องวงจรปิด
ชื่อ – สกุล นักศึกษา นาย ฉัตรชัย แซ่ตัน
คณะ วิศวกรรมศาสตร์ **ภาควิชา** วิศวกรรมคอมพิวเตอร์
ชื่อ – สกุล อาจารย์นิเทศ ผศ. บัณฑิต พัสยา
ชื่อ - สกุล ผู้นิเทศงาน ดร. ธีรัช ศักดิ์เดชยนต์
ชื่อสถานประกอบการ บริษัท กสิกร บิซิเนส - เทคโนโลยี กรุ๊ป

บทคัดย่อ

ในปัจจุบันนี้เป็นยุคของ Digital Disruption เกิดธุรกิจใหม่ บริการใหม่มากมายที่นำเทคโนโลยีดิจิทัลมาใช้แก้ปัญหาและตอบสนองความต้องการของลูกค้าได้แบบที่ไม่เคยมีมาก่อน แต่ในทางกลับกันหลายธุรกิจกำลังจะถูกแทนที่ด้วยเทคโนโลยี ธุรกิจค้าปลีกอย่างเช่น ร้านเสื้อผ้า ร้านเครื่องใช้ไฟฟ้า เป็นต้น ก็เป็นธุรกิจหนึ่งที่กำลังถูกแทนที่ด้วยตลาดออนไลน์และร้านค้าออนไลน์ที่ใช้กลยุทธ์การใช้เทคโนโลยีมาช่วยให้เข้าใจพฤติกรรมลูกค้า

โครงงานนี้จึงมุ่งที่จะช่วยนำเทคโนโลยีทางด้าน Machine Learning และ Artificial Intelligence ช่วยให้ผู้ประกอบการเข้าใจถึงพฤติกรรมของลูกค้ามากขึ้นไม่ว่าจะเป็น การวิเคราะห์ตำแหน่งภายในร้านที่ลูกค้าสนใจมากที่สุด นับจำนวนปริมาณลูกค้าในแต่ละช่วงเวลา แยกแยะเพศและนับจำนวนลูกค้าที่เข้าร้านค้า เป็นต้น ช่วยให้ผู้ประกอบการสร้างการบริการที่ตรงใจกับลูกค้ามากยิ่งขึ้น

คำสำคัญ : การเรียนรู้ด้วยเครื่อง, การเรียนรู้เชิงลึก, คอมพิวเตอร์วิทัศน์, แผนภูมิความร้อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Co-operative Title: Customer Analysis using CCTV Camera
Student Inter Name: Mr. Chatchai Shaetan
Faculty: Engineering **Department**: Computer Engineering
Advisor Name: Asst.Prof. Bundit Pasaya
Mentor Name: Mr. Theerat Sakdejayon
Company: Kasikorn Business - Technology Group

ABSTRACT

Nowadays, this era is the Digital Disruption era. Rising of new business and service that use digital technology to solve problem and meet customer needs like never before. Nevertheless, many businesses will be replaced by technology. Retail business such as cloth store, electronics store is a business that will be replaced by online market and online store that use the strategy using technology to understand customer's behavior.

Therefore, this project aims to bring the technology like Machine Learning and Artificial Intelligence to help entrepreneur to improve customer's behavior understanding such as analyzing the most customer's attention positions in store, counting amount of customer in each interval time of day, counting each customer's gender. This will help the entrepreneur to create services that more satisfy to customers.

Keywords: Machine Learning, Deep Learning, Computer Vision, Heatmap

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงงานสหกิจศึกษาฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์จากหลายท่านที่ได้ให้การสนับสนุนและให้ความช่วยเหลือผู้จัดทำตั้งแต่เริ่มต้นโครงงานจนเสร็จสมบูรณ์

ขอกราบขอบพระคุณอาจารย์บัณฑิต พัสยา ที่ให้ความอนุเคราะห์ในการนิเทศโครงงานและให้คำปรึกษาในการจัดทำโครงงานเป็นอย่างดี

ขอกราบขอบพระคุณคณาจารย์คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่กรุณาสั่งสอด ถ่ายทอดความรู้ ให้ผู้จัดทำได้ประยุกต์ใช้แก้ปัญหาในโครงงานนี้

ขอกราบขอบพระคุณบริษัท กสิกร บิซิเนส เทคโนโลยี กรุ๊ป ที่ให้โอกาสแก่ผู้จัดทำเข้าปฏิบัติงานสหกิจศึกษาและขอขอบคุณนาย ธีธัช ศักดิ์เดชยนต์ พี่เลี้ยงที่ให้การแนะนำและดูแลตลอดการปฏิบัติงาน

สุดท้ายนี้ขอขอบพระคุณทุกคนในครอบครัวที่คอยให้การสนับสนุนทั้งการเรียนและการทำงานเป็นอย่างดีเสมอมา รวมทั้งเป็นกำลังใจอันสำคัญมาโดยตลอดจนสามารถประสบความสำเร็จในการจัดทำโครงงานสหกิจศึกษาชิ้นนี้

ฉัตรชัย แซ่ตัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อ I

ABSTRACT II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญภาพ VII

บทที่ 1 บทนำ 1

1.1. ความเป็นมาและความสำคัญ 1

1.2. วัตถุประสงค์ของโครงงาน 1

1.3. ขอบเขตของโครงงาน 1

1.4. วิธีการดำเนินการวิจัย 2

1.5. ประโยชน์ที่คาดว่าจะได้รับ 2

บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง 3

2.1. Neural network 3

2.2. Deep learning 4

2.3. Transfer learning 5

2.4. การตรวจจับคนในภาพ (Person Detection) 6

2.5. การติดตามตัวบุคคล (Person Tracking) 9

2.6. Python 10

2.7. NumPy 12

2.8. OpenCV 12

2.9. JSON 13

2.10. Tensorflow 14

2.11. CUDA 16

2.12. Keras 18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 วิธีการดำเนินงาน **19**

3.1. ขั้นตอนการดำเนินงาน 19

3.2. เก็บความต้องการ (Requirement) 19

3.3. รวบรวมชุดข้อมูล (Data collection) 19

3.4. การออกแบบระบบ 21

3.5. พัฒนาโมเดลทำนาย 28

3.6. ส่งออกข้อมูล 31

บทที่ 4 ผลการวิจัย **37**

4.1. รายละเอียดการทดสอบ 37

4.2. การตรวจจับตำแหน่งบุคคล (Person Detection) 38

4.3. การติดตามตัวบุคคล (Person Tracking) 39

4.4. การรู้จำเพศด้วยภาพบุคคลทั้งตัว (Full-Body Gender Recognition) 40

4.5. Heatmap 40

4.6. การนับจำนวนคนเข้า - ออก (Person Counting) 42

4.7. การส่งออกข้อมูล 43

4.8. การเปรียบเทียบความแม่นยำกับงานวิจัยที่ใกล้เคียง 46

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ **48**

5.1. ผลการดำเนินการ 48

5.2. ปัญหาและอุปสรรคที่พบ 48

5.3. ประโยชน์ที่ได้จากการปฏิบัติงานสหกิจศึกษา 48

5.4. ข้อเสนอแนะ 49

เอกสารอ้างอิง **50**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 แสดงจำนวนภาพคนของแต่ละชุดข้อมูล 20
3.2 แสดงลักษณะการแสดงผลของการวิเคราะห์ในวิดีโอ 35
3.3 แสดงลักษณะการแสดงผลของการวิเคราะห์การเข้า-ออกในวิดีโอ 35
3.4 แสดงลักษณะการแสดงผลของการวิเคราะห์การเข้า-ออกพื้นที่ในวิดีโอ 35
4.1 สรุปผลการวิเคราะห์ในแต่ละพื้นที่ 43
4.2 สรุปผลการนับจำนวนคนเข้า 43
4.3 สรุปผลการนับจำนวนคนออก 44
4.4 เปรียบเทียบความแม่นยำของโมเดลแบบ Single frame 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

ภาพที่ **หน้า**

2.1 แสดงส่วนประกอบของ Neural network 3
2.2 แสดงส่วนประกอบของ Deep learning 4
2.3 เปรียบเทียบประสิทธิภาพของเทคนิค Deep learning กับเทคนิค Machine learning ดั้งเดิมต่อจำนวนข้อมูลที่มากขึ้น 4
2.4 อธิบายการทำงานของเทคนิค Transfer learning 5
2.5 ชุดข้อมูล ImageNet 5
2.6 ตัวอย่างภาพการทำงานของระบบตรวจจับคนในภาพ (Person Detection) 6
2.7 อธิบายโครงสร้างของโมเดล YOLOv3 7
2.8 อธิบายโครงสร้างการทำงานของ Feature Pyramid Network 7
2.9 อธิบายกราฟของ Focal Loss 8
2.10 อธิบายโครงสร้างการทำงานของโมเดล RetinaNet 8
2.11 เปรียบเทียบ mAP แต่ละโมเดลต่อเวลาที่ใช้คำนวณ 9
2.12 ตัวอย่างการทำงานของระบบติดตามคน 10
2.13 ภาษา Python 10
2.14 โมดูล NumPy 12
2.15 OpenCV 12
2.16 JSON 13
2.17 ตัวอย่างโครงสร้างของ JSON 14
2.18 TensorFlow 14
2.19 ตัวอย่างโครงสร้างข้อมูลของ Tensor แต่ละมิติ 15
2.20 CUDA 16
2.21 เปรียบเทียบจำนวน core ระหว่าง CPU และ GPU 17
2.22 Keras 18
3.1 ภาพตัวอย่างชุดข้อมูล PA-100K 20
3.2 ภาพตัวอย่างชุดข้อมูล PETA 20
3.3 Class Diagram ของระบบ 21
3.4 Flowchart ของการวิเคราะห์ด้วยโมเดลแบบ Single-frame 23
3.5 Flowchart ของการวิเคราะห์ด้วยโมเดลแบบ Multi-frame 25
3.6 Flowchart ของการสร้างมโนภาพข้อมูล 27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

3.7 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ Single-frame + Mode 29
3.8 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ CNN + RNN 30
3.9 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ CNN + RNN + CTC loss 30
3.10 ภาพตัวอย่างการนำ Heatmap มาใช้ในการนำเสนอข้อมูลการตำแหน่งของนักฟุตบอลในสนาม 31
3.11 ตัวอย่าง Heatmap จากจำนวนเวลาของร้านค้า 31
3.12 ตัวอย่าง Heatmap จากจำนวนคนของร้านค้า 32
3.13 ตัวอย่างการนับจำนวนคนเข้าออกแบบเส้น 33
3.14 ตัวอย่างข้อมูลตารางของการวิเคราะห์ทั้งวิดีโอ 33
3.15 ตัวอย่างข้อมูลตารางของการวิเคราะห์เฉพาะพื้นที่ 34
3.16 ตัวอย่างของข้อมูลในรูปแบบไฟล์ JSON 34
3.17 ตัวอย่างภาพในวิดีโอหลังจากการวิเคราะห์แล้ว 36
4.1 ตัวอย่างสถานที่ทดสอบ 37
4.2 ก่อนการตรวจจับตำแหน่งบุคคล 38
4.3 แสดงผลลัพธ์หลังการตรวจจับตำแหน่งบุคคล 38
4.4 แสดงผลลัพธ์ของการติดตามตัวบุคคล 39
4.5 แสดงผลลัพธ์ของการติดตามตัวบุคคลในภาพถัดไป 39
4.6 แสดงผลลัพธ์การทำนายเพศในภาพบุคคลทั้งตัว 40
4.7 แสดงผลลัพธ์ของการวิเคราะห์ความหนาแน่นของข้อมูลจำนวนเวลาที่แต่ละคนใช้ในแต่ละพื้นที่ แสดงในรูปแบบ Heatmap 41
4.8 แสดงผลลัพธ์ของการวิเคราะห์ความหนาแน่นของข้อมูลจำนวนคนในแต่ละพื้นที่ แสดงในรูปแบบ Heatmap 41
4.9 ก่อนคนเดินออก 42
4.10 การนับจำนวนคนหลังเดินออก 42
4.11 ตัวอย่างไฟล์ JSON 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1. ความเป็นมาและความสำคัญ

ในปัจจุบันนี้เป็นยุคของ Digital Disruption เกิดธุรกิจใหม่ บริการใหม่มากมายที่นำเทคโนโลยีดิจิทัลมาใช้แก้ปัญหาและตอบสนองความต้องการของลูกค้าได้แบบที่ไม่เคยมีมาก่อน เทคโนโลยีดิจิทัลได้สร้างความสะดวกสบายและช่วยให้ชีวิตของเราดีขึ้น แต่ในทางกลับกัน หลายธุรกิจกำลังจะถูกแทนที่ด้วยเทคโนโลยีอาทิเช่น ธุรกิจการขนส่งสาธรนะ เช่น รถแท็กซี่ จักรยานยนต์รับจ้าง เป็นต้น กำลังจะถูกแทนที่ด้วย รถยนต์ไร้คนขับ (Self – Driving car) และบริการเรียกรถรับโดยสาธรณะ (Ride – Hailing) ธุรกิจสื่อสิ่งพิมพ์ เช่น หนังสือพิมพ์ นิตยสาร เป็นต้น กำลังจะถูกแทนที่ด้วยสมาร์ทโฟนและเครือข่ายสังคมออนไลน์ (Social Network) อย่าง Facebook, YouTube, Twitter เป็นต้น อีกหนึ่งธุรกิจที่มีมูลค่ามหาศาลและเกิดขึ้นมานานแล้วอย่างร้านค้า ก็เป็นธุรกิจหนึ่งที่กำลังถูกแทนที่ด้วยเทคโนโลยีอย่าง E – Commerce

การปรับตัวโดยนำเทคโนโลยีมาใช้ในธุรกิจจึงเป็นเรื่องสำคัญในยุคนี้ ธุรกิจอย่างร้านค้าหรือห้างสรรพสินค้าสามารถนำเทคโนโลยีคอมพิวเตอร์อย่างเช่น AI มาช่วยในการพัฒนาบริการ อย่างการวิเคราะห์พฤติกรรมและเข้าใจลูกค้า ก็สามารถช่วยเพิ่มประสบการณ์ที่ดีขึ้นได้

1.2. วัตถุประสงค์ของโครงงาน

1.2.1. ระบบสามารถวิเคราะห์ข้อมูลตำแหน่งและเพศของลูกค้าภายในร้านด้วยภาพวิดีโอจากกล้องวงจรปิด

1.2.2. ระบบสามารถแสดงภาพรวมของความหนาแน่นของลูกค้าในแต่ละพื้นที่ของร้านด้วยการวิเคราะห์ข้อมูลภาพวิดีโอกล้องวงจรปิด

1.2.3. ระบบสามารถนับจำนวนคนเข้า – ออกร้านค้า

1.3. ขอบเขตของโครงงาน

1.3.1. ระบบสามารถตรวจจับตำแหน่งของลูกค้าในร้านค้าได้เฉพาะลูกค้าที่เห็นภาพเต็มตัว

1.3.2. ระบบสามารถส่งออกข้อมูลการวิเคราะห์ในรูปแบบไฟล์ Excel, ไฟล์ JSON และไฟล์วิดีโอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4. วิธีการดำเนินการวิจัย

1.4.1. ศึกษางานวิจัยที่ใกล้เคียง

ศึกษางานวิจัยที่หัวข้อใกล้เคียงงานเช่น Full-Body Gender Recognition เพื่อนำองค์ความรู้ที่เคยมีผู้วิจัยมาแล้วมาปรับใช้

1.4.2. รวบรวมชุดข้อมูล

รวบรวมชุดข้อมูลที่ใช้ฝึกสอนโมเดลโดยมีเกณฑ์การรวบรวมคือต้องเป็นชุดข้อมูลภาพจากวิดีโอกล้องวงจรปิดและมีคำอธิบายประกอบ (Annotation) ด้วยเท่านั้น

1.4.3. พัฒนาระบบ

พัฒนาโมเดลทำนาย การวิเคราะห์ข้อมูล และการส่งออกข้อมูล

1.4.4. วัดผลและปรับปรุงระบบ

วัดผลโดยเปรียบเทียบกับงานวิจัยในหัวข้อเดียวกัน

1.4.5. จัดทำเอกสาร

จัดทำเอกสารอธิบายวิธีการพัฒนาและคู่มือการใช้งานเพื่อให้การพัฒนาต่อยอดสามารถทำได้อย่างง่ายดาย

1.5. ประโยชน์ที่คาดว่าจะได้รับ

1.5.1. ผู้ประกอบการสามารถเข้าใจภาพรวมของความสนใจของลูกค้าต่อสินค้าที่จัดวางอยู่ในแต่ละพื้นที่ผ่านภาพแสดงความหนาแน่นด้วยแผนที่ความร้อน (Heatmap)

1.5.2. ผู้ประกอบการสามารถนำข้อมูลจำนวนลูกค้าและสัดส่วนของลูกค้าในแต่ละเพศที่เข้า – ออกร้านค้า

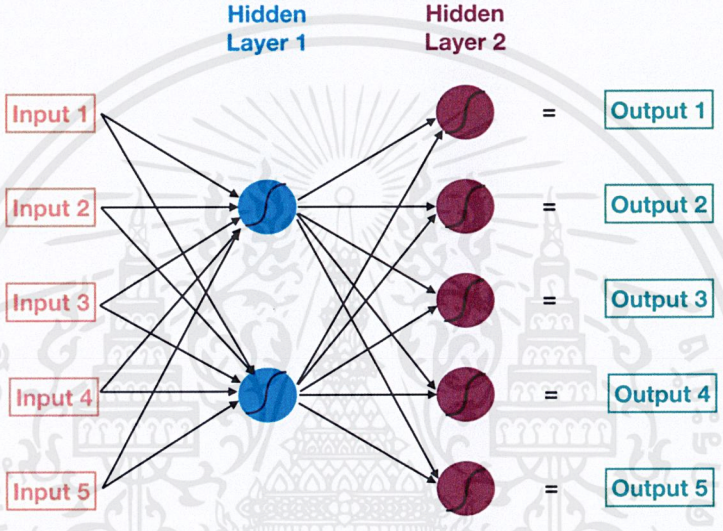
1.5.3. นักพัฒนาสามารถนำข้อมูลการวิเคราะห์ไปพัฒนาแอปพลิเคชันต่อยอดจากงานวิจัยนี้ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
แนวคิดและทฤษฎีที่เกี่ยวข้อง

2.1. Neural network

Neural network เป็นหนึ่งในเทคนิคของ Machine learning ที่มีการทำงานแบบ Connectionist คล้ายกับการทำงานของเซลล์ประสาทของสมองซึ่งประกอบไปด้วยเซลล์ประสาท (neuron) และจุดประสานประสาท (synapses) เชื่อมต่อกันเป็นเครือข่าย

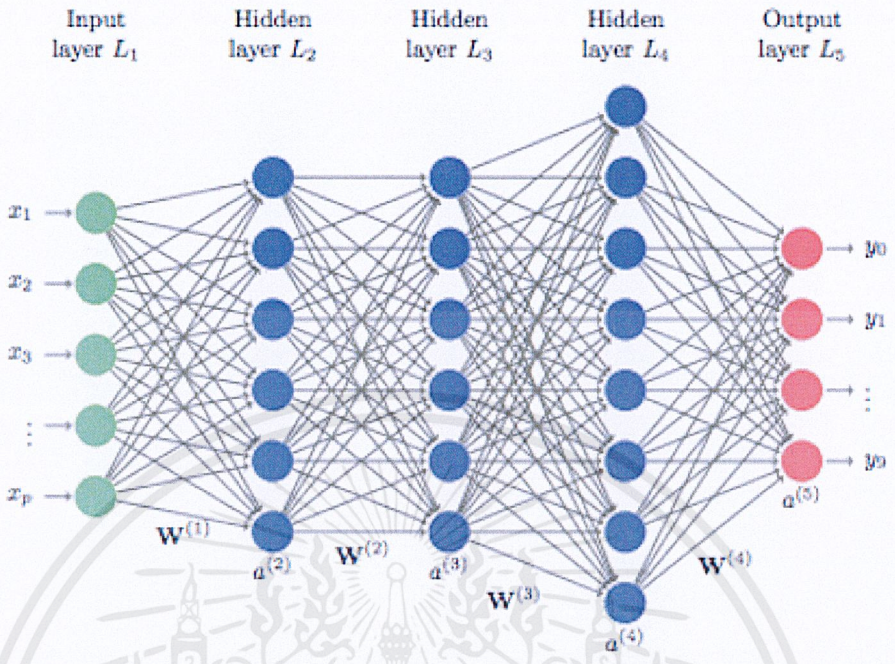


ภาพที่ 2.1 แสดงส่วนประกอบของ Neural network

การทำงานของ Neural network จะเริ่มขึ้นเมื่อรับข้อมูล input เข้ามา network นำข้อมูล input แต่ละตัวคูณกับ weight ของ node แต่ละตัว เมื่อคูณครบแล้ว แต่ละ node จะรวมค่าที่ได้จากการคูณ input แต่ละตัวและส่งผลลัพธ์ที่ได้ให้ชั้นต่อไป โดยวิธีการนี้เรียกว่า Feed – Forward เมื่อทำงานจนครบทุกชั้นจนถึงชั้น output แล้วจะนำผลลัพธ์จากการทำนายที่ได้จากชั้นสุดท้ายไปเปรียบเทียบกับค่าผลลัพธ์จริงด้วยฟังก์ชันที่ทำหน้าที่ในการคำนวณหาค่าความผิดพลาดระหว่างค่าที่ทำนายกับค่าจริง เรียกฟังก์ชันนั้นว่า loss function หรือ cost function เรียกค่าผิดพลาดว่า loss เมื่อได้ loss มาแล้ว neural network จะปรับค่า weight ของแต่ละ node ย้อนกลับจากชั้น output จนถึงชั้น input ตามขนาดของ loss เรียกวิธีนี้ว่า back propagation ด้วยการทำงานทั้งหมดนี้เอง ทำให้ neural network สามารถเรียนรู้ข้อมูลต่าง ๆ ได้มากมาย

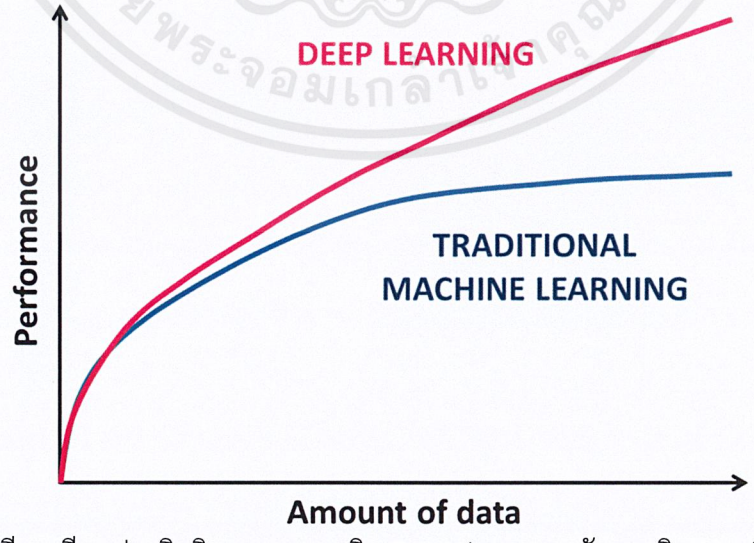
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2. Deep learning



ภาพที่ 2.2 แสดงส่วนประกอบของ Deep learning

Deep learning คือการใช้ชั้นของ neural network ต่อกันหลาย ๆ ชั้นจนเกิดความลึกของ neural network ขึ้นมา ด้วยด้วยจำนวนชั้นที่มากส่งผลให้โมเดลสามารถบรรจุตัวแปรได้มากขึ้น ทำให้ Deep Learning สามารถเรียนรู้ได้ดียิ่งขึ้นตามจำนวนข้อมูลที่สอนและแก้ไขปัญหาที่มีความซับซ้อนขึ้นได้ อีกทั้งยังสามารถทำการสกัดคุณสมบัติ (Feature Extraction) ได้ภายในตัว



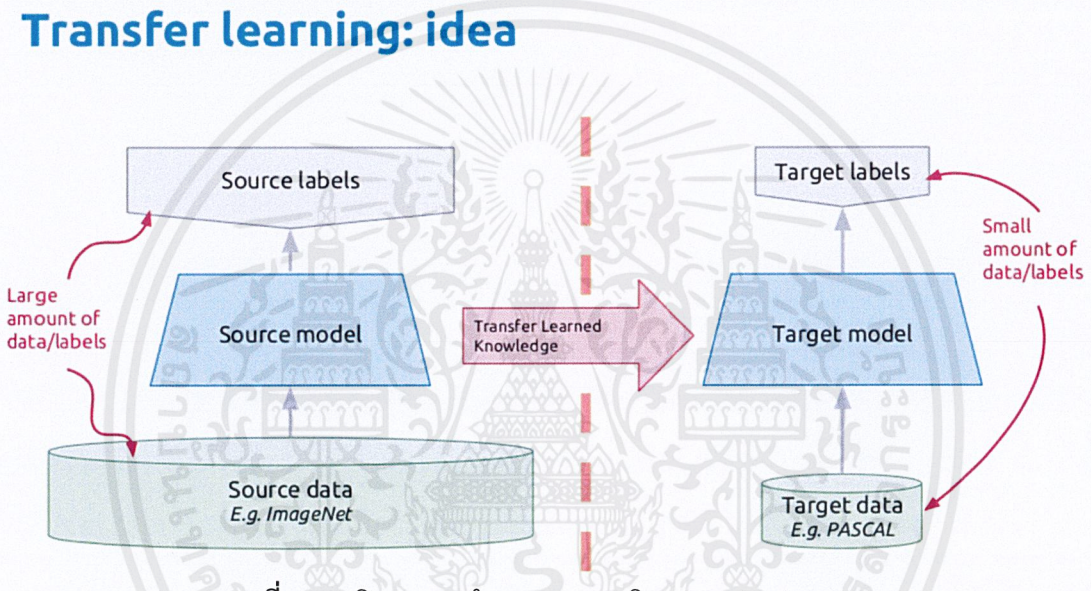
ภาพที่ 2.3 เปรียบเทียบประสิทธิภาพของเทคนิค Deep learning กับเทคนิค Machine learning ดั้งเดิมต่อจำนวนข้อมูลที่มากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัจจุบัน มีการนำ Deep learning กันอย่างกว้างขวาง นำไปใช้แก้ไขปัญหาที่เทคนิคทาง Machine Learning แบบเดิมยังทำได้ไม่ดี เช่น การแยกแยะชนิดของรูปภาพ (Image Classification) การตรวจจับวัตถุในภาพ (Object Detection) การรู้จำใบหน้า (Face Recognition) เป็นต้น

2.3. Transfer learning

Transfer Learning เป็นเทคนิคหนึ่งใน Deep Learning ที่ช่วยให้โมเดลสามารถเรียนรู้ได้เร็วขึ้นและใช้ข้อมูลฝึกสอนน้อยจากองค์ความรู้เดิมที่ใกล้เคียงที่เคยสอนให้โมเดลรู้จัก Transfer Learning เป็นที่นิยมมากในงานทางด้าน Computer Vision



ภาพที่ 2.4 อธิบายการทำงานของเทคนิค Transfer learning

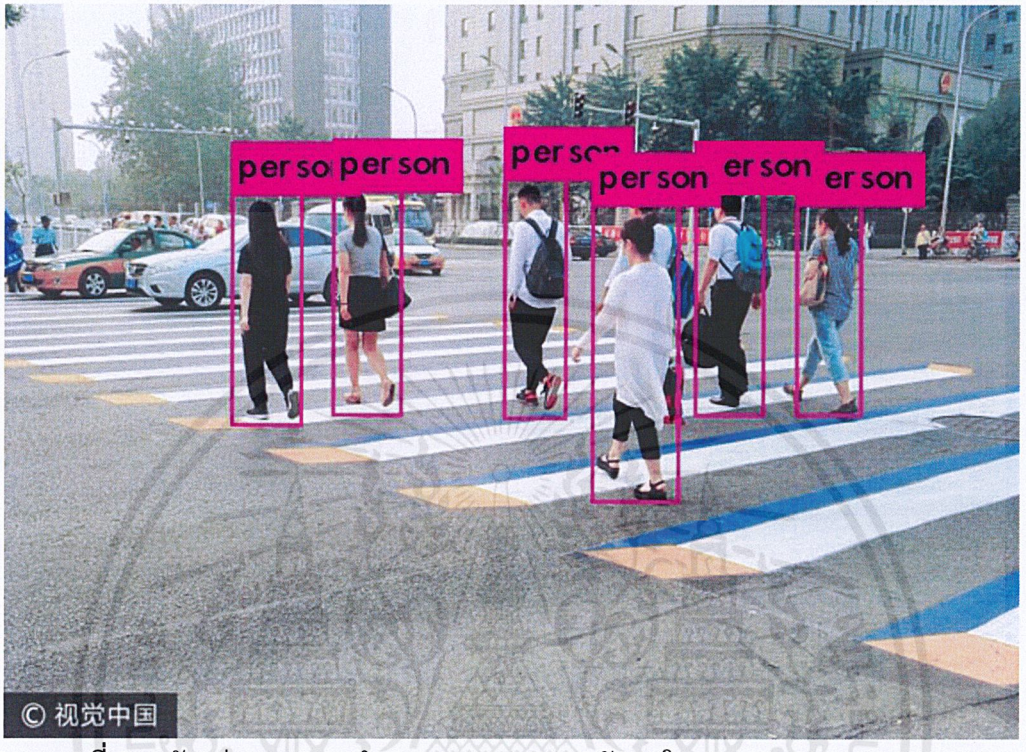
ในปัจจุบัน เนื่องจากมีการแข่งขัน ILSVRC (Image Large Scale Visual Recognition Challenge) ซึ่งเป็นการแข่งขันในโจทย์ Image Recognition ด้วยชุดข้อมูลรูปชื่อว่า ImageNet ซึ่งมีรูปหลากหลายกว่า 1000 ชนิดและจำนวนชุดข้อมูลมีจำนวนมากกว่า 10 ล้านภาพพร้อมคำอธิบาย ในการแข่งขันนี้เป็นจุดกำเนิดโมเดลที่เป็น State-of-the-art มากมายเช่น VGG, ResNet ซึ่งเป็นรากฐานของการทำ Transfer Learning ในปัจจุบัน

ภาพที่ 2.5 ชุดข้อมูล ImageNet

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4. การตรวจจับคนในภาพ (Person Detection)

งานตรวจจับคนในภาพ (Person Detection) เป็นหนึ่งในงานด้าน Computer Vision ที่นักวิจัยทั่วโลกกำลังให้ความสนใจอยู่ในปัจจุบัน



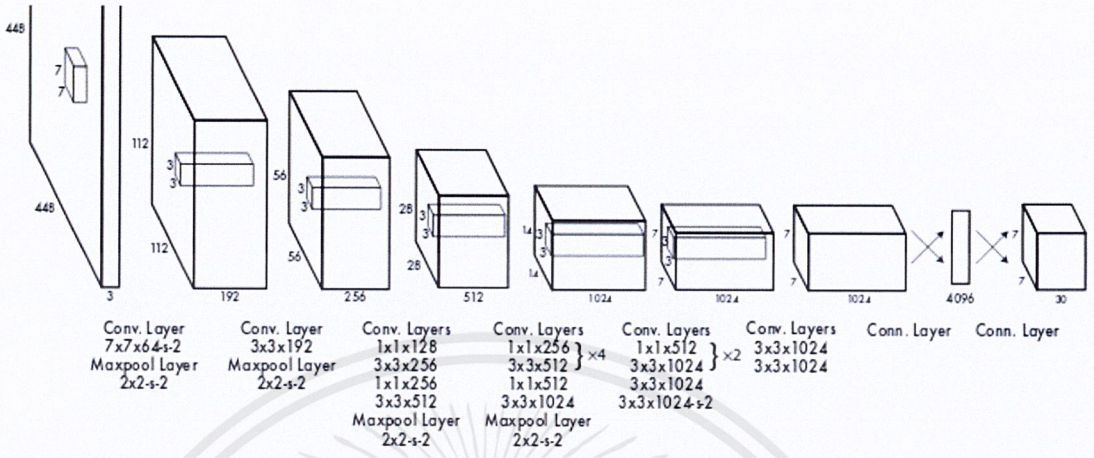
ภาพที่ 2.6 ตัวอย่างภาพการทำงานของระบบตรวจจับคนในภาพ (Person Detection)

2.4.1. YOLOv3 (You Only Look Once)

เป็นโมเดลตรวจจับวัตถุในภาพ (Object Detection) แบบ 1-stage ที่ใช้ Convolutional Neural Network (CNN) เป็นหลัก มีข้อดีคือมีความสามารถในการตรวจจับได้แม่นยำและรวดเร็ว สามารถทำงานแบบเรียลไทม์ (Real-time) ได้

โดยโมเดลที่ทดสอบเป็น YOLO รุ่น 3 ที่ถูกปรับปรุงจากรุ่นก่อนหน้าโดยมีการใช้เทคนิค Skip Connection ทำให้โมเดลมีความแม่นยำมากขึ้น

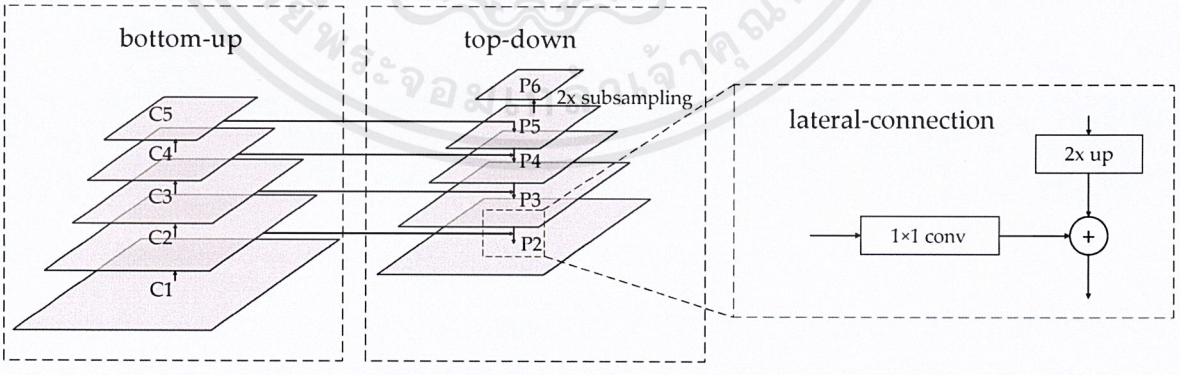
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.7 อธิบายโครงสร้างของโมเดล YOLOv3

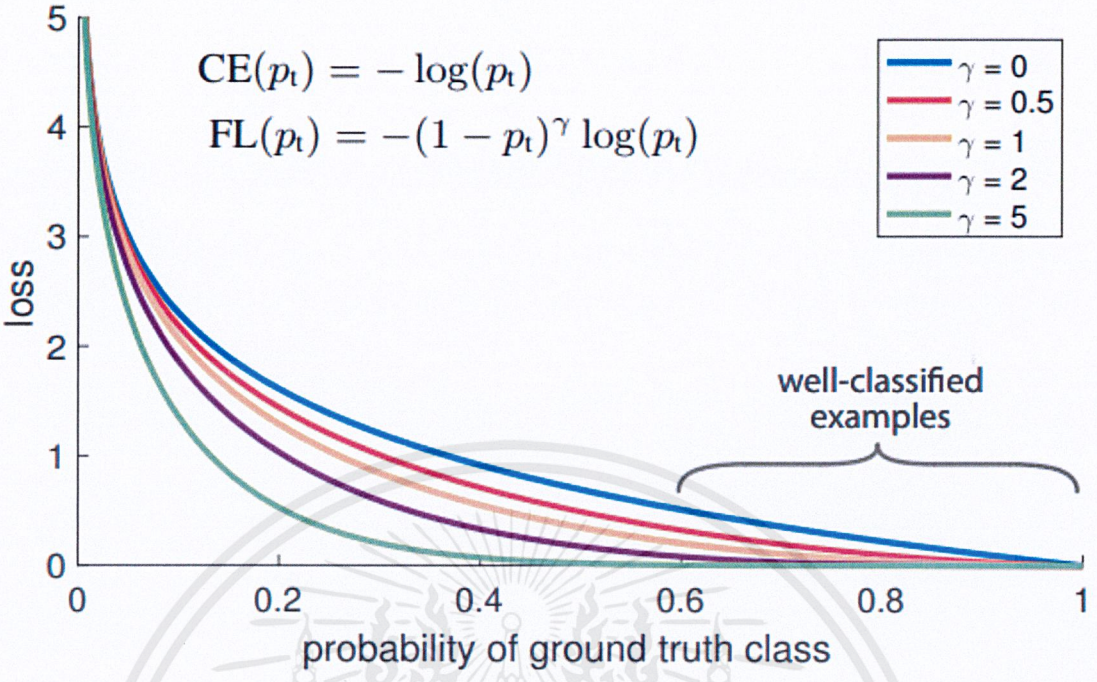
2.4.2. RetinaNet

เป็นโมเดลตรวจจับวัตถุในภาพ (Object Detection) แบบ 1-stage ที่ใช้ Convolutional Neural Network (CNN) เป็นหลัก ใช้เทคนิค Feature Pyramid Network จึงทำให้โมเดลสามารถตรวจหาวัตถุขนาดเล็กในภาพได้ดี และใช้ Focal Loss ซึ่งเป็น Loss function ที่พัฒนาเพิ่มเติมเพื่อแก้ไขจุดบกพร่องด้านการให้น้ำหนักความสำคัญในแต่ละ class ของ Cross entropy loss จึงทำให้โมเดลสามารถเรียนรู้กับชุดข้อมูลที่ไม่สมดุลได้ดี

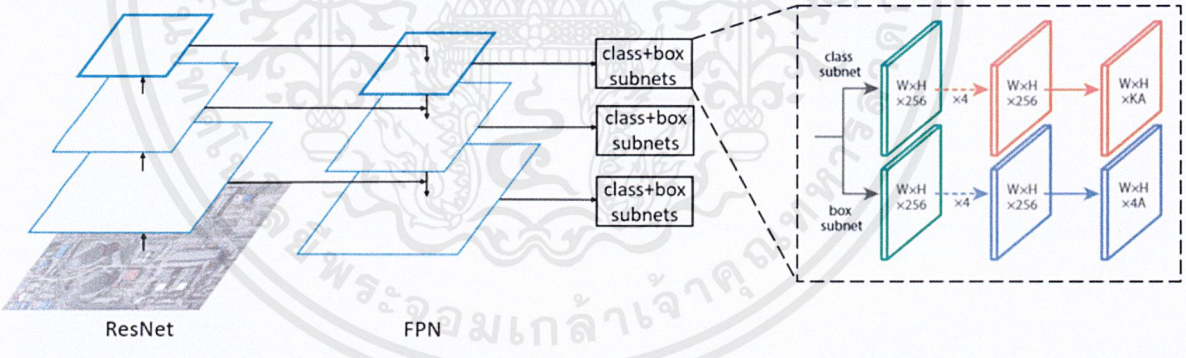


ภาพที่ 2.8 อธิบายโครงสร้างการทำงานของ Feature Pyramid Network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.9 อธิบายกราฟของ Focal Loss

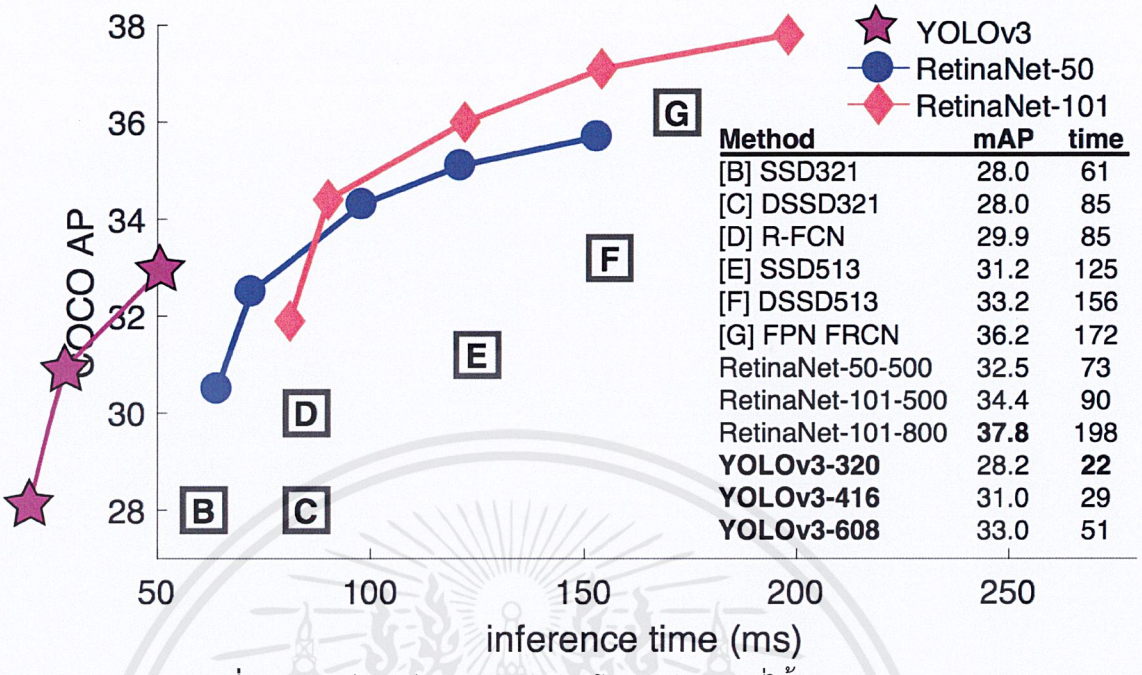


ภาพที่ 2.10 อธิบายโครงสร้างการทำงานของโมเดล RetinaNet

2.4.3. การเปรียบเทียบระหว่าง YOLOv3 กับ RetinaNet

ทั้ง 2 โมเดลที่กล่าวมาเป็นโมเดลที่ถือว่าเป็น State-of-the-art ของโจทย์การตรวจจับวัตถุในภาพ (Object Detection) ซึ่งแต่ละโมเดลก็มีความเหมาะสมต่างกันไป โดย YOLOv3 มีข้อดีคือใช้เวลาการคำนวณน้อย เหมาะกับงานที่ต้องการความเร็วการประมวลผลสูง, RetinaNet มีข้อดีคือมีความแม่นยำกับทั้งวัตถุขนาดเล็กและขนาดใหญ่แต่ว่าจะใช้เวลาการคำนวณมากกว่า YOLOv3 จึงเหมาะกับงานที่เน้นความแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



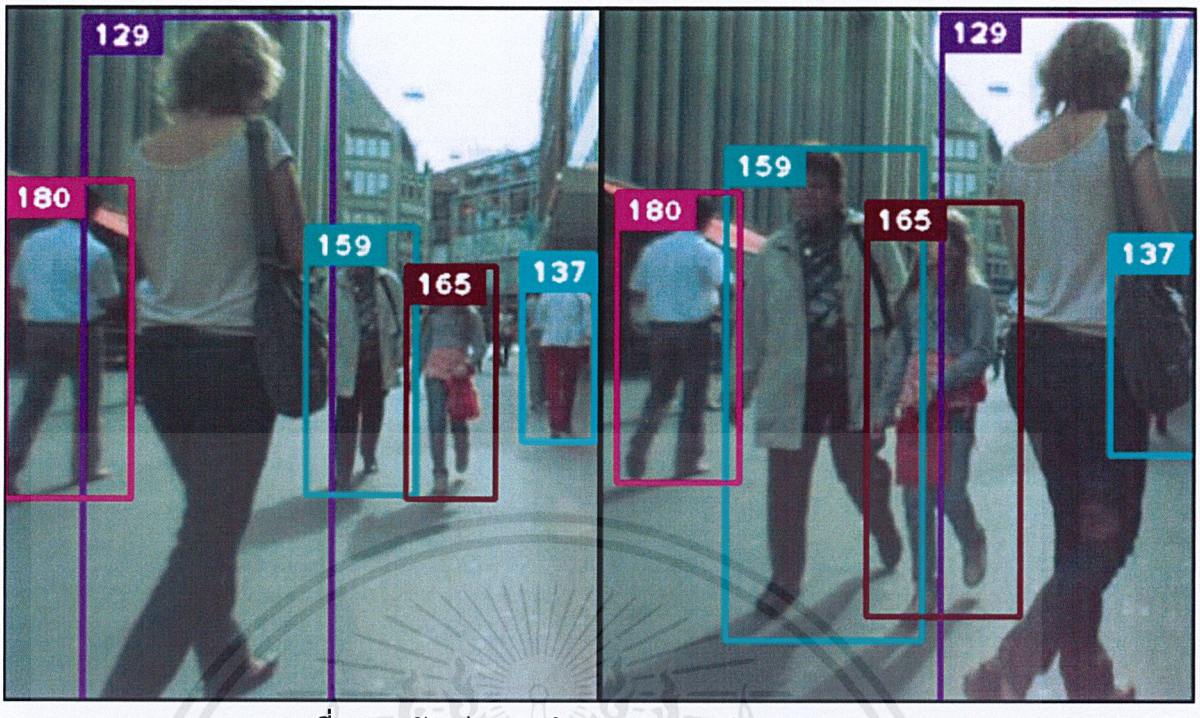
ภาพที่ 2.11 เปรียบเทียบ mAP แต่ละโมเดลต่อเวลาที่ใช้คำนวณ

2.5. การติดตามตัวบุคคล (Person Tracking)

2.5.1. DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric)

เป็น algorithm ในงานติดตามวัตถุหลายวัตถุในภาพพร้อมกัน (Real-time Multiple Object Tracking) ที่ทำงานได้รวดเร็ว ใช้ทรัพยากรน้อย และทนทานต่อปัญหาการเดินสวนกันได้ดี โดย DeepSORT พัฒนามาจาก SORT (Simple Online and Realtime Tracking) ที่ใช้ Kalman Filter ในการติดตามวัตถุโดยใช้ตำแหน่งวัตถุในภาพเพียงอย่างเดิม ใน DeepSORT มีการนำ Convolutional Neural Network (CNN) มาสกัดคุณสมบัติของคนในภาพ เพื่อนำคุณสมบัติไปช่วยในการติดตามคนในภาพควบคู่ไปกับ Kalman Filter

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.12 ตัวอย่างการทำงานของระบบติดตามคน

2.6. Python

ภาพที่ 2.13 ภาษา Python

Python เป็นภาษาชั้นสูงแบบ Interpreter ชนิดหนึ่ง Python ได้ถูกคิดค้นขึ้นในเดือนธันวาคมปี 1989 โดยนาย Guido van Rossum โปรแกรมเมอร์ชาวดัตช์ ทำงานได้ทั้งบนระบบ Unix, Linux, Windows NT, MacOS อีกทั้งภาษาตัวนี้ยังเป็น Open Source สามารถที่จะนำ Python มาพัฒนาโปรแกรมของเราได้ฟรีๆโดยไม่ต้องเสียค่าใช้จ่าย และความเป็น Open Source ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ Python มีความสามารถสูงขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

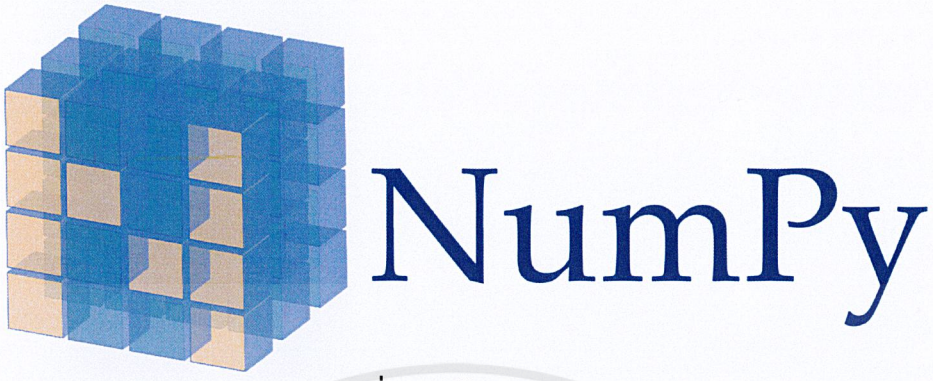
Python ถูกสร้างขึ้นมาจากภาษา C การประมวลผลจะทำในแบบ Interpreter โดยจะประมวลผลไปทีละบรรทัดและปฏิบัติตามคำสั่งที่ได้รับ Python เวอร์ชั่นแรกคือ เวอร์ชั่น 0.9.0 ออกมาเมื่อปี 2533 และเวอร์ชั่นปัจจุบันคือ 3.8

คุณลักษณะเด่นของภาษา Python

1) สนับสนุนแนวคิดการเขียนโปรแกรมเชิงวัตถุหรือ OOP (Object Oriented Programming)
2) เป็น Open Source
3) โค้ดที่เขียนด้วย Python เพียงครั้งเดียว สามารถนำไปรันบนระบบปฏิบัติการได้หลากหลาย
4) เป็นภาษาแบบ Dynamic typing
5) มี Built-in Object Types คือ โครงสร้างของข้อมูลที่สามารถใช้ได้ใน Python ประกอบด้วย List (Linked List), Tuple, Set, Dictionary (Hash-Map), String ที่ง่ายต่อการใช้งานและมีประสิทธิภาพสูง
6) ไวยากรณ์ของภาษาถูกออกแบบมาให้เน้นความเรียบง่ายในการเขียนและใกล้เคียงกับภาษามนุษย์มากที่สุด
7) มีความสามารถในการประมวลผลทางด้านคณิตศาสตร์วิทยาศาสตร์และวิศวกรรมศาสตร์ได้อย่างมีประสิทธิภาพ
8) มีชุมชนนักพัฒนา (Community) ที่แข่งแกร่ง เกิดไลบรารี่มากมายโดยเฉพาะอย่างยิ่งไลบรารี่ทางด้านการประมวลผลวิทยาศาสตร์ ปัญญาประดิษฐ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7. NumPy



ภาพที่ 2.14 โมดูล NumPy

โมดูล NumPy เป็นโมดูลส่วนเสริมของภาษา Python สำหรับใช้คำนวณทางคณิตศาสตร์และวิทยาศาสตร์ มีประโยชน์ในการประมวลผลข้อมูลที่มีโครงสร้างแบบ Array, Vector, Matrix โดยมีคำสั่งพร้อมใช้งานจำนวนมากมาย และโมดูลนี้สามารถสร้างข้อมูลชนิด Array (ที่ภาษา Python ไม่มี)

2.8. OpenCV

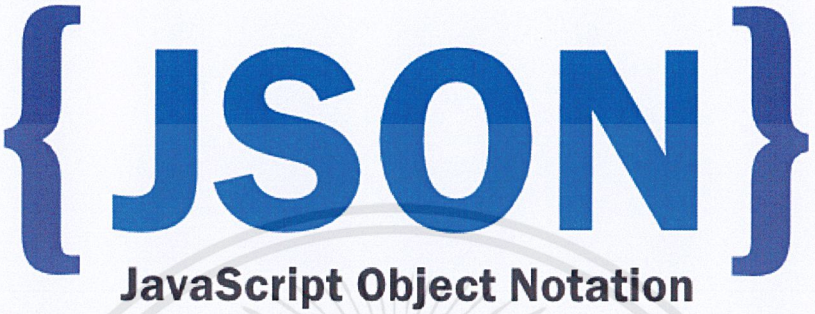
ภาพที่ 2.15 OpenCV

OpenCV หรือชื่อเต็มว่า Open Computer Vision เป็นไลบรารี่ที่ทำงานด้าน Image Processing และ Computer Vision โดย OpenCV ถูกพัฒนาขึ้นโดย Intel โดย OpenCV เป็นไลบรารีแบบข้ามแพลตฟอร์ม (Cross platform) และสามารถใช้งานได้ฟรีภายใต้ ลิขสิทธิ์ของ BSD แบบ Open Source โดย OpenCV ยังรองรับ Framework การเรียนรู้เชิงลึก Deep Learning Frameworks เช่น TensorFlow, PyTorch อีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

OpenCV ถูกเขียนด้วยภาษา C++ และรองรับ interface ของภาษาอื่นๆ เช่น Python, Java อีกด้วย ได้รับการพัฒนาเพื่อส่งเสริมการนำมาใช้งานโดยผู้ใช้ที่มีเพิ่มขึ้น

2.9. JSON



ภาพที่ 2.16 JSON

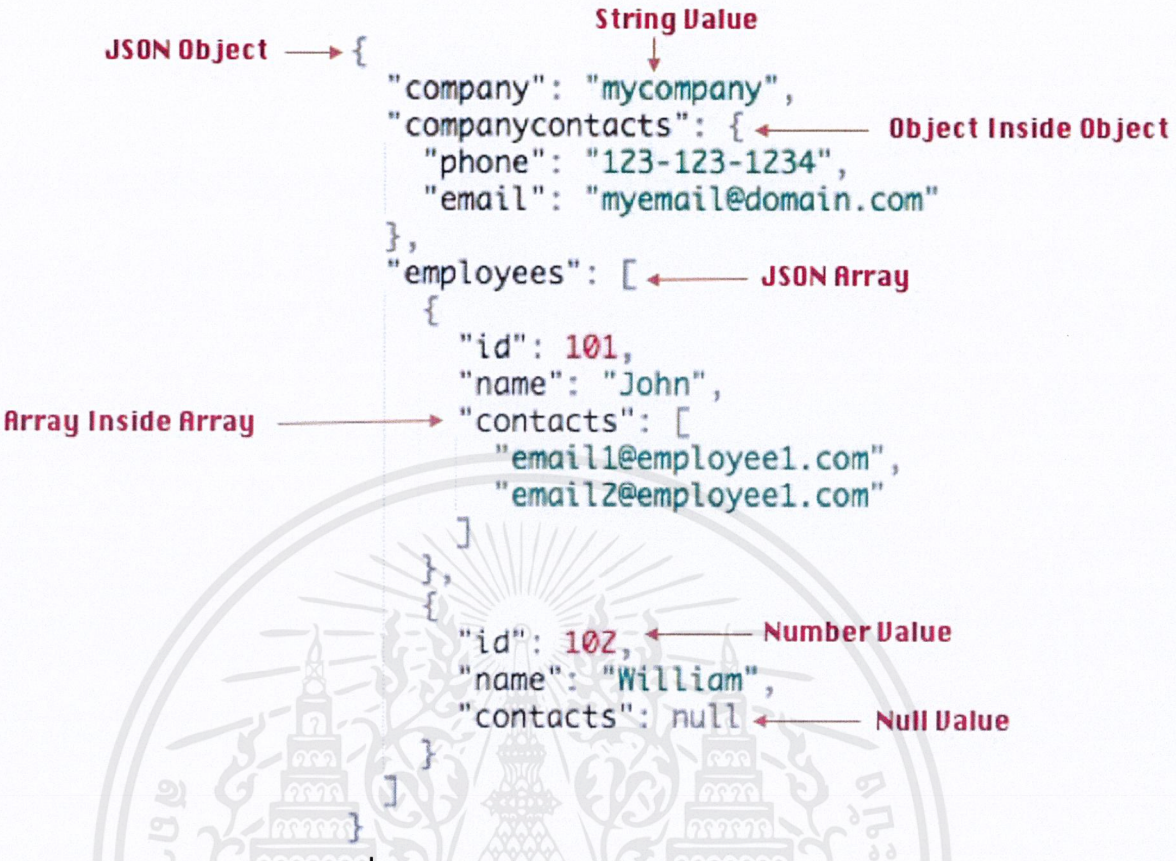
JSON ย่อมาจาก JavaScript Object Notation เป็น รูปแบบมาตรฐานกลางของการบันทึกข้อมูล Object ของภาษา JavaScript ในรูปแบบของไฟล์ text มีข้อดีคือสามารถอ่านออกได้ด้วยตาเปล่า สามารถใช้ในการสร้าง Object ขึ้นมาเพื่อส่งข้อมูลระหว่าง application หรือ Applications Program Interface (API) โดยจะมีรูปแบบเป็น คู่ Key-Value หรือเป็นแบบ Array และสามารถนำมาใช้แทน XML format ได้

JSON เป็นที่ได้นิยมในการใช้งานจากภาษา JavaScript มาก่อน แต่ปัจจุบันมีภาษาหลายชนิดที่เริ่มใช้งาน JSON โดยสามารถสร้างและแปลงไปมาได้ โดยชนิดข้อมูลที่รองรับมีดังนี้

- Number: ตัวเลขจำนวนเต็มหรือทศนิยม
- String: Unicode ใช้เครื่องหมาย double-quote (“) เป็นตัวบ่งบอก และสามารถใช้ backslash syntax ได้
- Boolean: True or False
- Array: ชุดข้อมูล ซึ่งจะเป็นชนิดใดก็ได้ ใช้สัญลักษณ์ square bracket [var1,var2] เป็นตัวแสดง และคั้นด้วย comma
- Object: ชุดข้อมูลที่เป็นคู่ Key-Value โดย Key เป็น strings ใช้สัญลักษณ์ปีกกา {key1:value1,key2:value2} ใช้ comma เป็นตัวแบ่งแต่ละคู่ และใช้ colon เป็นตัวแบ่งระหว่าง key และ value
- Null: ค่าว่าง

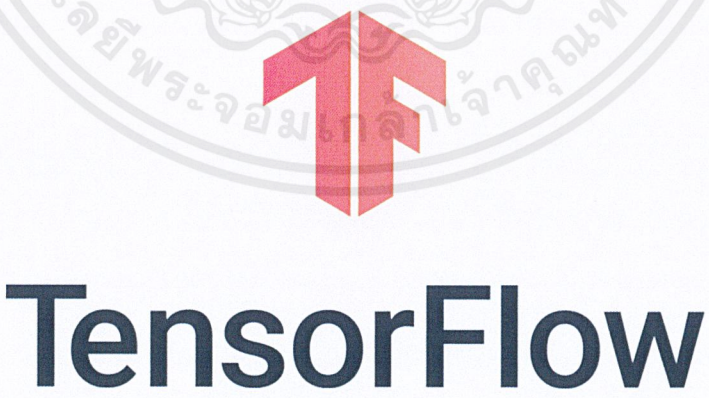
ไม่สนใจ whitespace (ช่องไฟ) มีเพียงแค่ 4 แบบที่อยู่ในกลุ่ม whitespace คือ space, tab, newline (\n) และ carriage return (\r) และไม่มีสัญลักษณ์ comment สำหรับ JSON

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.17 ตัวอย่างโครงสร้างของ JSON

2.10. Tensorflow



ภาพที่ 2.18 TensorFlow

Tensorflow เป็นเฟรมเวิร์กทางที่ทำงานทางด้าน Machine Learning ที่ถูกพัฒนาขึ้นโดย Google เพื่อให้นักวิจัยและนักพัฒนาทำงานกับโมเดล AI ได้

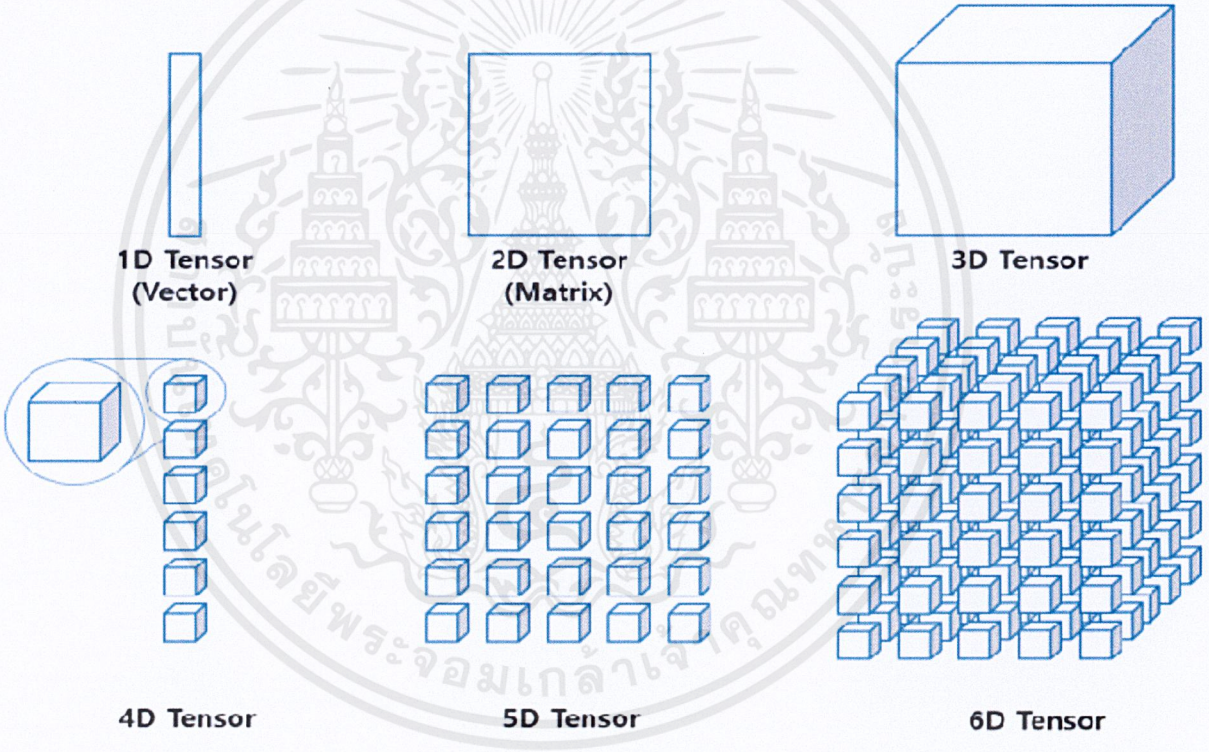
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยเปิด open source ตั้งแต่ปี 2015 และปล่อยตัวสมบูรณ์ออกมาในปี 2017 พร้อมลิขสิทธิ์แบบ Apache Open Source เปิดโอกาสให้คนทั่วไปสามารถใช้งาน, ดัดแปลง และ แจกจ่ายตัวที่ถูกดัดแปลงมาแล้ว โดยที่ไม่จำเป็นต้องจ่ายให้ Google เลย

โดยโครงสร้างของ Tensorflow จะแบ่งเป็น 3 ส่วน ได้แก่

- การเตรียมประมวลผลข้อมูล
- การสร้างแบบจำลอง
- ฝึกและประเมินแบบจำลอง

ชื่อของ Tensorflow มาจาก การที่ Tensorflow ทำงานกับข้อมูลที่เป็นอาร์เรย์หลายมิติหรือที่เรียกกันว่า tensors และเรามีหน้าที่จัดเรียงลำดับการประมวลผลเป็น flowchart (หรือที่เรียกว่ากราฟ) ข้อมูลที่ถูกป้อนไป ก็จะผ่าน (flow) แต่ละกระบวนการจนออกมาเป็นผลลัพธ์



ภาพที่ 2.19 ตัวอย่างโครงสร้างข้อมูลของ Tensor แต่ละมิติ

ชื่อของ Tensorflow มาจากชื่อเฟรมเวิร์กที่ถูกนำมาพัฒนาต่ออย่าง Tensor การคำณวนทั้งหมดจึงเกี่ยวข้องเวกเตอร์และเมทริกซ์หลายมิติที่มีข้อมูลบรรจุอยู่ ค่าทั้งหมดในหนึ่ง Tensor จะมีรูปร่างของข้อมูลแตกต่างกันไปที่เรียกว่า Shape

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tensor จะมาจากอะไรก็ได้ทั้งข้อมูลที่ป้อนเข้าไป หรือ ผลลัพธ์จากการคำนวณ ใน Tensorflow การคำนวณทั้งหมดจะเกิดขึ้นภายใน Graph โดย Graph คือลำดับของการประมวลผลอย่างต่อเนื่อง ในแต่ลำดับก็มีชื่อเรียกว่า op node และแต่ละ op node ก็เชื่อมถึงกัน

Graph เป็นโครงสร้างของตัวประมวลตัวและการเชื่อมต่อกันนะหว่าง node แต่ graph ไม่ได้เป็นตัวแสดงผล และในแต่ละ node ก็มี tensor อยู่มากมายที่รอประมวลผล

Tensorflow ใช้ Graph เป็นตัวรวบรวมและอธิบายชุดการคำนวณทั้งหมดในระหว่างการฝึก Graph จึงมีประโยชน์มากมายทั้ง

- สามารถทำงานผ่าน CPUs และ GPUs ได้หลายตัว ทั้งยังทำงานผ่านมือถือได้
- ความสามารถในเรื่องการพกพา ทำให้สามารถหยิบใช้งานได้อย่างทันที และสามารถบันทึก graph เพื่อดำเนินการต่อในอนาคต
- การคำนวณทั้งหมดใน graph เกิดจาก tensor ที่เชื่อมไว้ด้วยกัน

2.11. CUDA

ภาพที่ 2.20 CUDA

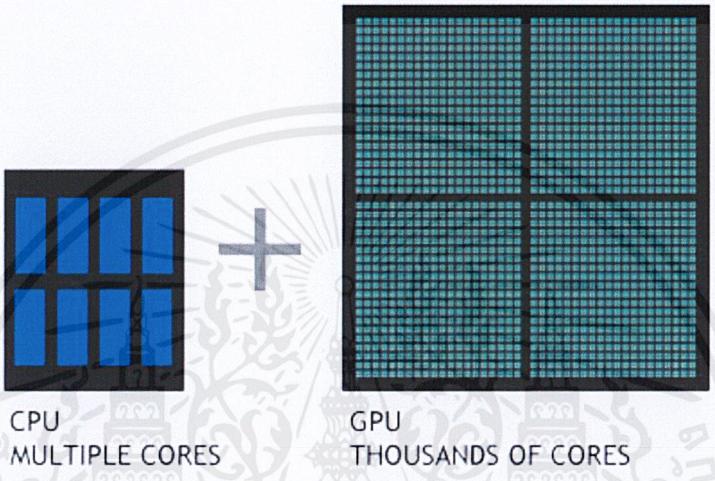
CUDA (Compute Unified Device Architecture) เป็นแพลตฟอร์มสำหรับการประมวลผลแบบคู่ขนาน (Parallel Processing) และ Application Programing Interface (API) พัฒนาโดยบริษัท Nvidia เพื่อให้นักพัฒนาและวิศวกรซอฟต์แวร์สามารถดึงศักยภาพในการประมวลผลแบบขนานของ GPU (Graphic Processing Unit) สำหรับการประมวลผลในงานต่าง ๆ หรือที่เรียกว่า GPGPU (General-Purpose computing on Graphics Processing Units)

โดยปกติแล้วในเครื่องคอมพิวเตอร์หน้าที่สำหรับการประมวลผลจะเป็นหน้าที่ของ CPU (Central Processing Unit) ซึ่งเป็นหน่วยประมวลผลกลางในการทำงานของคอมพิวเตอร์ ในสมัยก่อนการประมวลผลทางด้านกราฟฟิกจะใช้ CPU เป็นหลัก ส่งผลให้การทำงานของ CPU หนักขึ้นและไม่เพียงพอต่อการทำงานในด้านอื่น ๆ GPU จึงเข้ามามีบทบาทในการประมวลผลทางด้าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กราฟฟิกแทน เพื่อลดการทำงานของ CPU ทำให้การงานด้านกราฟฟิกมีความลื่นไหลมากยิ่งขึ้น และเมื่อ CPU ทำงานน้อยลง ความร้อนภายในเครื่องก็ลดลงด้วยเช่นกัน

CPU นั้นจะประกอบไปด้วย core เพียงไม่กี่ core จึงเหมาะกับงานในลักษณะที่เรียกว่า Sequential serial processing หรือการประมวลผลแบบลำดับ ในขณะที่ GPU จะประกอบด้วย core ขนาดเล็กจำนวนมากและถูกออกแบบให้มีการกระจายการทำงานในลักษณะ Parallel หรือการประมวลผลแบบคู่ขนาน



ภาพที่ 2.21 เปรียบเทียบจำนวน core ระหว่าง CPU และ GPU

โดยงานที่เหมาะกับ GPU จะเป็นงานในลักษณะ Single Instruction Multiple Data stream (SIMD) กล่าวคือใช้การประมวลผลด้วยชุดข้อมูลจำนวนมาก แต่ทำงานด้วยคำสั่งเดียว

ในปัจจุบันนอกจากงานทางด้านกราฟฟิกแล้ว ยังนิยมนำ GPU มาใช้ประมวลผลในงานปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) ไม่ว่าจะเป็น Deep Learning, Computer Vision หรือมาประยุกต์ใช้ในระบบรถไร้คนขับ (Self-driving car) หรืองานทางด้านวิทยาศาสตร์และคณิตศาสตร์ เนื่องจากเป็นงานที่ใช้การประมวลผลทางคณิตศาสตร์เป็นส่วนใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.12. Keras

ภาพที่ 2.22 Keras

Keras เป็นไลบรารี่ทางด้าน Neural Network และ Deep Learning หนึ่งของภาษา Python ช่วยให้การพัฒนาอัลกอริทึมเป็นไปได้โดยง่าย โดย Keras จะทำหน้าที่เป็นหน้าบ้าน (Front-End) ที่จะช่วยห่อหุ้มการทำงานระดับล่างเอาไว้ ทำให้ช่วยลดเวลาในการพัฒนาได้เป็นอย่างมาก

ปัจจุบัน Keras รองรับไลบรารี่หลังบ้านอย่าง TensorFlow, CNTK, Theano ทำให้นักพัฒนาสามารถเขียนโค๊ดครั้งแล้วทำงานได้ทั้ง TensorFlow, CNTK และ Theano ได้ทันทีโดยไม่ต้องเขียนใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงาน

3.1. ขั้นตอนการดำเนินงาน

1. เก็บความต้องการ (Requirement)
2. ศึกษางานวิจัยที่เกี่ยวข้อง
3. ออกแบบระบบ
4. รวบรวมชุดข้อมูล (Data Collection)
5. พัฒนาโมเดลทำนาย
6. วัดผลและปรับปรุงโมเดลทำนาย
7. สรุปผลและจัดทำเอกสารโครงงาน

3.2. เก็บความต้องการ (Requirement)

จากการเก็บข้อมูลความต้องการกับผู้ประกอบการ สามารถนำมาแยกเป็นหัวข้อได้ดังนี้

1. ตรวจจับตำแหน่งของลูกค้าในร้านด้วยภาพกล้องวงจรปิด
2. ตรวจจับตำแหน่งของลูกค้าในร้านเฉพาะพื้นที่ด้วยภาพกล้องวงจรปิด
3. แยกแยะเพศของลูกค้าจากภาพกล้องวงจรปิด
4. เก็บข้อมูลจำนวนลูกค้าเข้าและออกร้านค้าจากภาพกล้องวงจรปิด
5. เก็บข้อมูลจำนวนลูกค้าแยกตามเพศจากภาพกล้องวงจรปิด
6. เก็บข้อมูลจำนวนลูกค้าเข้าและออกร้านค้าเฉพาะพื้นที่จากภาพกล้องวงจรปิด
7. เก็บข้อมูลจำนวนลูกค้าแยกตามเพศเฉพาะพื้นที่จากภาพกล้องวงจรปิด
8. วิเคราะห์เวลาที่ลูกค้าใช้ภายในร้านเฉลี่ยในแต่ละคน
9. สามารถส่งออกข้อมูลลูกค้าในรูปแบบไฟล์ Excel ได้
10. สามารถส่งออกข้อมูลในรูปแบบ Video ได้
11. วิเคราะห์และสร้างภาพ Heatmap ของเวลาที่ลูกค้าใช้ในร้านค้าต่อพื้นที่
12. วิเคราะห์และสร้างภาพ Heatmap ของจำนวนลูกค้าในร้านค้าต่อพื้นที่
13. สามารถปรับแต่งระยะ (Range) การวิเคราะห์ของ Heatmap ได้

3.3. รวบรวมชุดข้อมูล (Data collection)

เราได้รวบรวมชุดข้อมูลภาพคนทั้งตัวสำหรับฝึกสอนโมเดลรู้จำเพศจากภาพคนทั้งตัว (Full – Body gender recognition)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 แสดงจำนวนภาพคนของแต่ละชุดข้อมูล

Dataset	จำนวน (ภาพ)
PA-100K	100,000
PETA	19,000

ภาพที่ 3.1 ภาพตัวอย่างชุดข้อมูล PA-100K

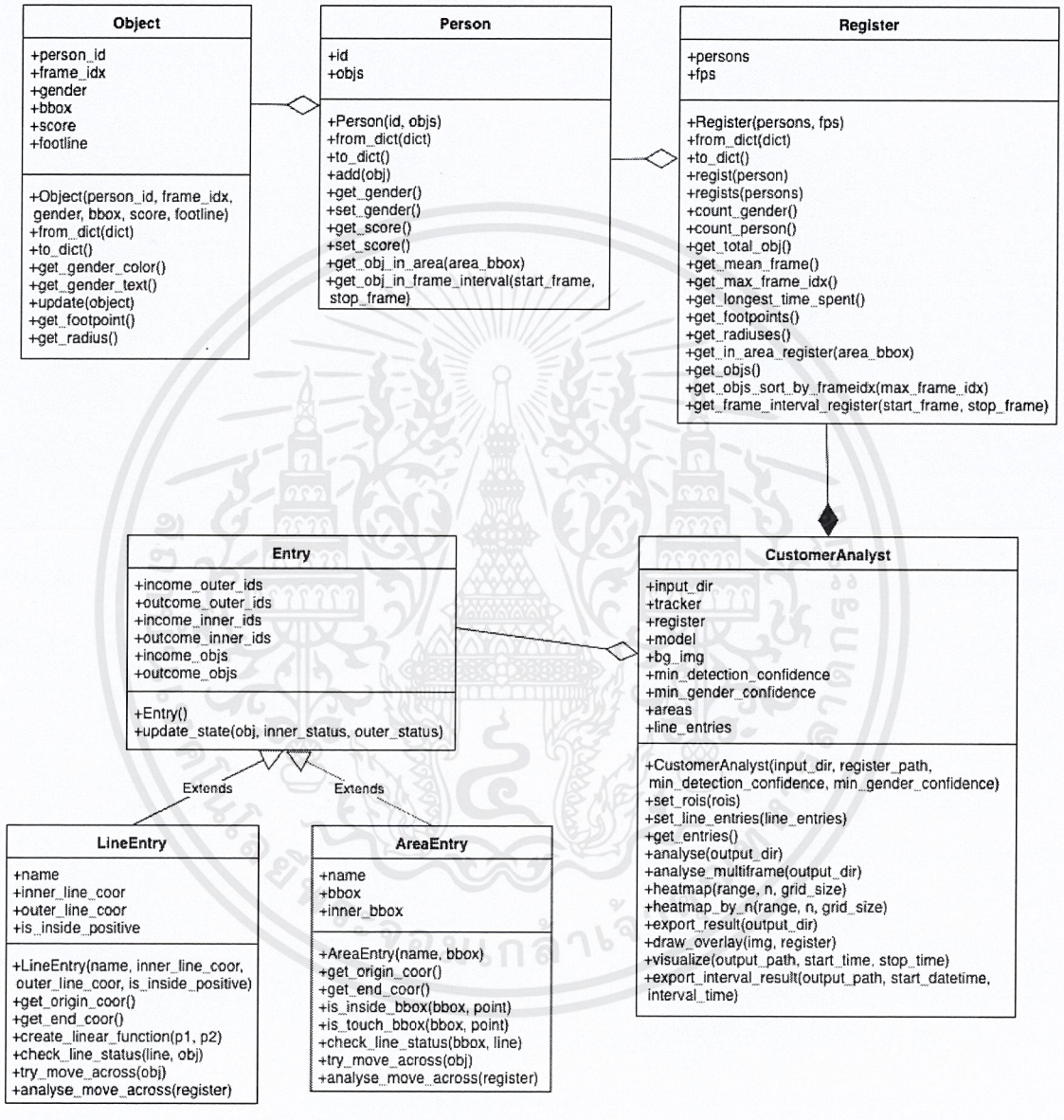
ภาพที่ 3.2 ภาพตัวอย่างชุดข้อมูล PETA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4. การออกแบบระบบ

3.4.1. Class Diagram

เนื่องจากระบบค่อนข้างมีความซับซ้อน จึงใช้ Class diagram ช่วยในการออกแบบระบบเพื่อให้ง่ายต่อการพัฒนาและการดูแลรักษา ดังนี้



ภาพที่ 3.3 Class Diagram ของระบบ

แต่ละ class ทำหน้าที่ดังนี้

- Object แทน วัตถุ 1 ชิ้นที่ตรวจจับได้ใน 1 ภาพ เก็บข้อมูลอาทิเช่น ID ของคน (person_id), พิกัดของวัตถุในภาพ (bbox), ลำดับของภาพที่เจอวัตถุนี้ (frame_idx) เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

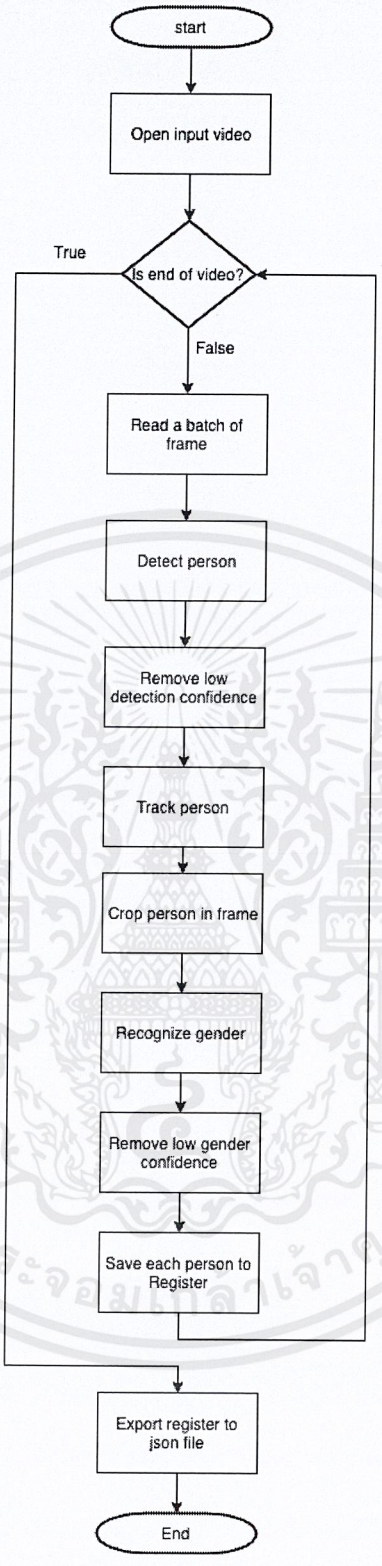
- Person แทน บุคคล เก็บข้อมูลอาทิเช่น ID ของคน (id), อาร์เรย์ของ Object (objs) เป็นต้น
- Register แทน ทะเบียนรายชื่อ เก็บข้อมูลของคนทั้งหมดที่ตรวจเจอในวิดีโอ
- Entry แทน ทางเข้า ใช้ในการเก็บพิกัดของทางเข้า – ออกและนับจำนวนคนเดินผ่าน
- LineEntry แทน ทางเข้า - ออกแบบเส้น
- AreaEntry แทน ทางเข้า – ออกแบบพื้นที่
- CustomerAnalyst แทนตัววิเคราะห์ข้อมูล ทำหน้าที่ในการวิเคราะห์วิดีโอ , Visualization, สร้าง Heatmap โดยเก็บข้อมูลเช่น ทะเบียนรายชื่อ (register) เป็นต้น

3.4.2. แผนผังแสดงการทำงานของการวิเคราะห์

3.4.2.1 การวิเคราะห์โดยโมเดลแบบ Single-frame มีการทำงานดังนี้

1. เปิดวิดีโออินพุต
2. ตรวจสอบว่าสิ้นสุดวิดีโอแล้วหรือไม่ ถ้าสิ้นสุดแล้ว ข้ามไปการทำงานที่ 12
3. อ่านชุดรูปภาพจากวิดีโอ
4. ตรวจจับตำแหน่งคนจากชุดภาพ (Person Detection)
5. ตัดทิ้งตำแหน่งที่ถูกทำนายของคนที่ได้คะแนนความมั่นใจต่ำกว่าเกณฑ์
6. ติดตามตำแหน่งคนด้วย (Person Tracking)
7. ตัดภาพเฉพาะรูปภาพคน
8. ทำนายเพศด้วยโมเดลรู้จำเพศ (Gender Recognition)
9. ตัดทิ้งเพศของคนที่ได้คะแนนความมั่นใจต่ำกว่าเกณฑ์
10. บันทึกการทำนายเพศของแต่ละคนลง Register
11. ย้อนกลับไปการทำงานที่ 2
12. ส่งออก Register โดยการบันทึกเป็นไฟล์ JSON
13. จบการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



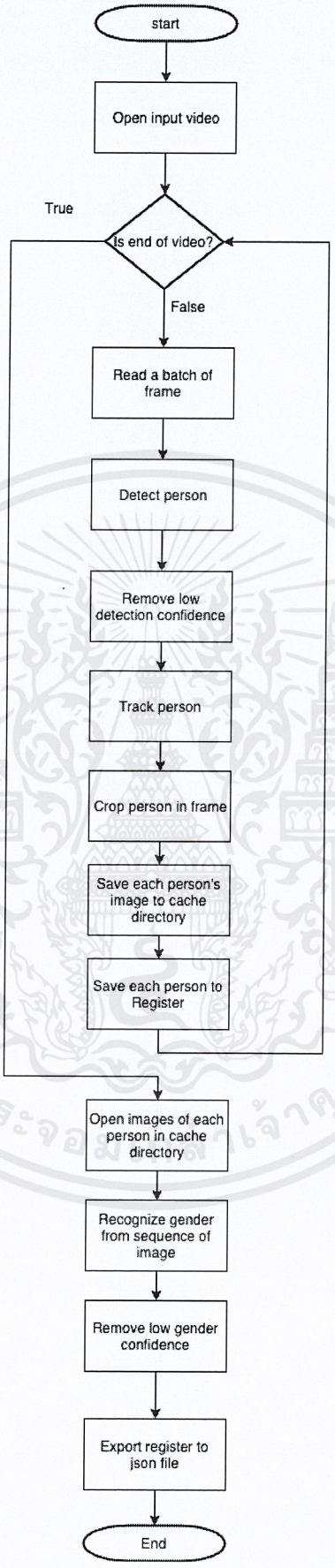
ภาพที่ 3.4 Flowchart ของการวิเคราะห์ด้วยโมเดลแบบ Single-frame

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.2.2 การวิเคราะห์โดยโมเดลแบบ Multi-frame

1. เปิดวิดีโออินพุต
2. ตรวจสอบว่าสิ้นสุดวิดีโอแล้วหรือไม่ ถ้าสิ้นสุดแล้ว ข้ามไปการทำงานที่ 11
3. อ่านชุดรูปภาพจากวิดีโอ
4. ตรวจจับตำแหน่งคนจากชุดภาพ (Person Detection)
5. ตัดทิ้งตำแหน่งที่ถูกทำนายของคนที่ได้คะแนนความมั่นใจต่ำกว่าเกณฑ์
6. ติดตามตำแหน่งคนด้วย (Person Tracking)
7. ตัดภาพเฉพาะรูปภาพคน
8. บันทึกภาพของแต่ละคนที่ได้ลงแคช (Cache)
9. บันทึกข้อมูลของแต่ละคนลง Register
10. ย้อนกลับไปการทำงานที่ 2
11. อ่านภาพของคนแต่คนละคนที่ถูกบันทึกในแคช (Cache)
12. ทำนายเพศจากลำดับของภาพคนแต่ละคนด้วยโมเดลรู้จำเพศ (Gender Recognition)
13. ตัดทิ้งเพศของคนที่ได้คะแนนความมั่นใจต่ำกว่าเกณฑ์
14. ส่งออก Register โดยการบันทึกเป็นไฟล์ JSON
15. จบการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



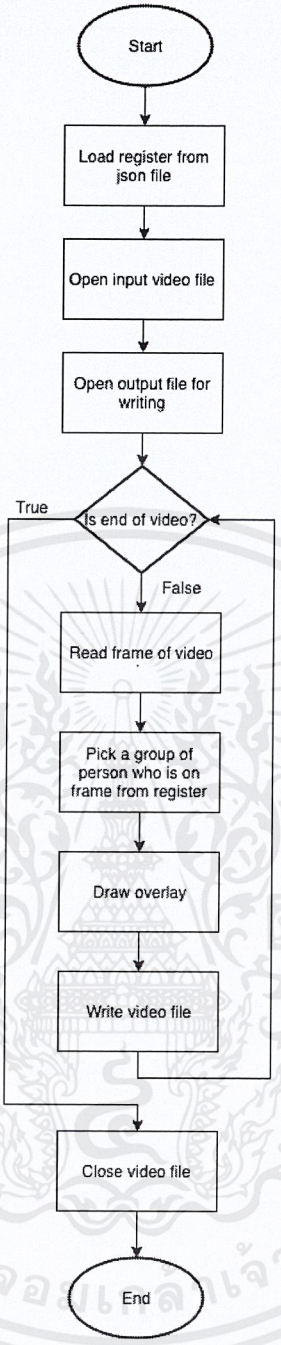
ภาพที่ 3.5 Flowchart ของการวิเคราะห์ด้วยโมเดลแบบ Multi-frame

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3. การสร้างมโนภาพข้อมูล (Visualization)

1. โหลดข้อมูล Register ที่ถูกบันทึกจากวิเคราะห์มาแล้วจากไฟล์ JSON
2. เปิดอ่านวิดีโออินพุต
3. เปิดไฟล์วิดีโอด้วยโหมดเขียนไบนารี่
4. ตรวจสอบว่าวิดีโอสิ้นสุดแล้วหรือไม่ ถ้าสิ้นสุดแล้ว ข้ามไปที่การทำงานที่ 10
5. อ่านภาพจากวิดีโอ
6. หยิบข้อมูลของคนที่อยู่ในลำดับเฟรมปัจจุบันของวิดีโอจาก Register
7. วาดโอเวอร์เลย์
8. เขียนภาพลงไฟล์วิดีโอ
9. ย้อนกลับไปการทำงานที่ 4
10. ปิดการเขียนไฟล์วิดีโอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 3.6 Flowchart ของการสร้างมโนภาพข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5. พัฒนาโมเดลทำนาย

เพื่อลดความซับซ้อนของโมเดลจึงแบ่งโมเดลออกเป็น 3 ส่วนได้แก่

3.5.1. ตรวจจับบุคคลในภาพ (Person Detection)

การตรวจจับคนในภาพ (Person Detection) คือการแยกตำแหน่งของภาพบุคคลออกจากภาพพื้นหลังที่เราไม่สนใจ โดยเราเลือกใช้โมเดลที่ทำงานใกล้เคียงกับการตรวจจับบุคคลในภาพ นั่นคือโมเดลตรวจจับวัตถุ (Object Detection) โดยโมเดลที่เลือกจะต้องมีการตรวจจับบุคคลร่วมอยู่ด้วยและเป็น State-of-the-art ของงานทางด้านการตรวจจับวัตถุ

เราได้ทดลองโมเดลตรวจจับวัตถุ 2 โมเดลด้วยกันได้แก่ YOLOv3 และ RetinaNet โดยทั้งสองโมเดลถูกฝึกสอนโดยชุดข้อมูล Common Object in Context หรือชื่อย่อว่า COCO ซึ่งเป็นชุดข้อมูลรูปภาพจำนวนกว่า 2 แสนรูปพร้อมคำอธิบายตำแหน่งและชื่อวัตถุในภาพจำนวนกว่า 1.5 ล้านชิ้น ทำให้โมเดลทั้งสองมีความฉลาดและรู้จักกับรูปภาพแบบต่าง ๆ เป็นอย่างดี

3.5.2. ติดตามตำแหน่งคนในภาพ (Person Tracking)

เนื่องจากข้อมูลที่ได้รับเป็นแบบวิดีโอ จึงจำเป็นต้องมีการติดตามตำแหน่งคนในภาพเพื่อระบุว่าคนที่ตรวจจับเจอเป็นคนเดิมกับภาพก่อนหน้าหรือไม่ โดยงานทางด้านการติดตามตำแหน่งคนในภาพเป็นงานที่มีความท้าทายเป็นอย่างมาก เพราะปัญหาเรื่องคนเดินสวนกันหรือเดินซ้อนกันเป็นปัญหาใหญ่ของการติดตามคน โดยโมเดลที่เลือกทดสอบคือ DeepSORT

3.5.3. ทำนายเพศ (Gender Recognition)

ในส่วนของการออกแบบโมเดลทำนาย ได้เลือกใช้ Convolutional Neural Network (CNN) เป็นหัวใจหลักในการทำนายเพศ และเลือกใช้สถาปัตยกรรมที่เป็นหนึ่งใน State-of-the-Art ในงานด้าน Image Classification จากการแข่งขัน ImageNet ที่ชื่อว่า Residual Network (ResNet) มาทำ Transfer learning เพื่อให้โมเดลนำองค์ความรู้จากที่เคยทำนายแยกแยะภาพมาช่วยในการเรียนรู้ในงานแยกแยะเพศของภาพตัวคนได้ง่ายยิ่งขึ้น โดยทดลอง 2 วิธีด้วยกันดังนี้

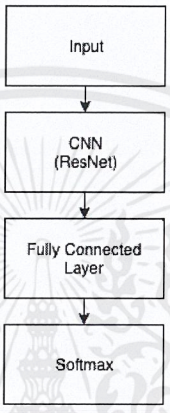
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.3.1 แบบ Single-frame + Mode

เป็นวิธีการทำนายเพศของคนในภาพครั้งละ 1 ภาพ แล้วนำผลลัพธ์ในการทำนายแต่ละครั้งมาหาฐานนิยมเพื่อเลือกทำนายเพศจากจำนวนครั้งที่โมเดลทำนายมากที่สุด

โมเดลจะแบ่งเป็น 3 ส่วน คือ

1. การสกัดคุณลักษณะจากภาพด้วย Convolutional Neural Network
2. ทำนายด้วย Fully Connected Layer
3. แปลงเป็นความน่าจะเป็นด้วย Softmax



ภาพที่ 3.7 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ Single-frame + Mode

3.5.3.2 แบบ Multi-frame

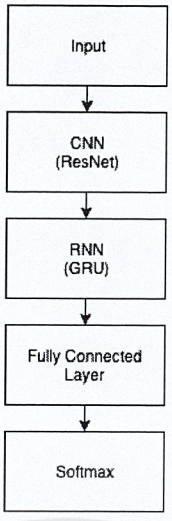
เป็นวิธีการทำนายเพศของคนในภาพโดยใช้ภาพหลาย ๆ ภาพประกอบการทำนาย โดยได้ทดลอง 2 โมเดลด้วยกันดังนี้

3.5.3.2.1 CNN + RNN

โมเดลจะแบ่งเป็น 4 ส่วน คือ

1. การสกัดคุณลักษณะจากภาพด้วย Convolutional Neural Network
2. นำคุณลักษณะของแต่ละรูปของคนในวิดีโอมาเรียนรู้ด้วย Recurrent Neural Network
3. ทำนายด้วย Fully Connected Layer
4. แปลงเป็นความน่าจะเป็นด้วย Softmax

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

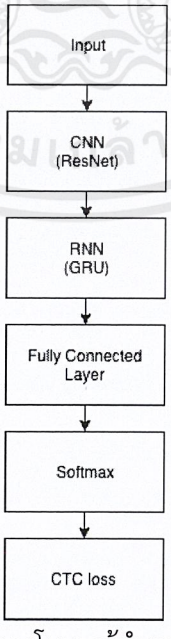


ภาพที่ 3.8 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ CNN + RNN

3.5.3.2.2 CNN + RNN + CTC loss

โมเดลจะแบ่งเป็น 4 ส่วน คือ

1. การสกัดคุณลักษณะจากภาพด้วย Convolutional Neural Network
2. นำคุณลักษณะของแต่ละรูปของคนในวิดีโอมาเรียนรู้ด้วย Recurrent Neural Network
3. ทำนายด้วย Fully Connected Layer
4. แปลงเป็นความน่าจะเป็นด้วย Softmax
5. ดูความต่อเนื่องของความน่าจะเป็นของรูปแต่ละรูปด้วย Connectionist Temporal Classification loss



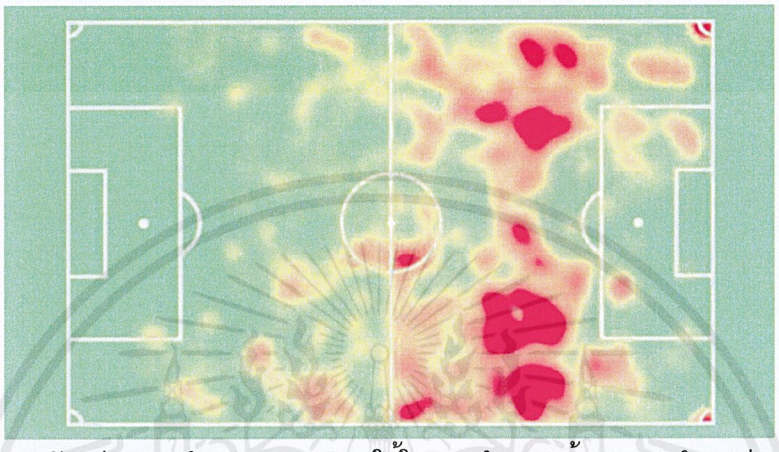
ภาพที่ 3.9 อธิบายสถาปัตยกรรมของโมเดลรู้จำเพศแบบ CNN + RNN + CTC loss

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6. ส่งออกข้อมูล

3.6.1. Heatmap

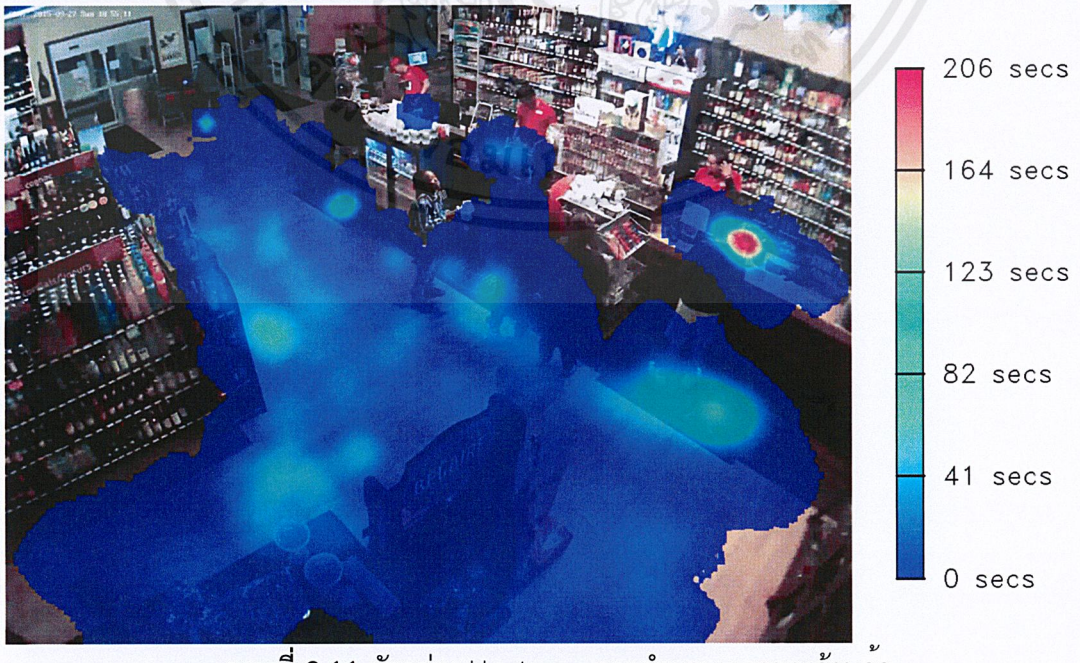
เป็นรูปแบบของการนำเสนอข้อมูลในรูปแบบของภาพชนิดหนึ่ง (Data Visualization) โดยใช้การไล่ระดับสีในการแสดงผลความหนาแน่นของข้อมูล มีประโยชน์ในการวิเคราะห์ภาพรวมของข้อมูลเพื่อนำไปใช้งานต่อไป



ภาพที่ 3.10 ภาพตัวอย่างการนำ Heatmap มาใช้ในการนำเสนอข้อมูลการตำแหน่งของนักฟุตบอลในสนาม

3.6.1.1 Heatmap จากจำนวนเวลา

เป็น Heatmap ที่แสดงความหนาแน่นของข้อมูลจากเวลาที่คนใช้ในวิดีโอ มีประโยชน์ในการมองภาพรวมเช่น ตำแหน่งที่ลูกค้าในร้านค้าใช้เวลาไปมากที่สุด



ภาพที่ 3.11 ตัวอย่าง Heatmap จากจำนวนเวลาของร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.1.2 Heatmap จากจำนวนคน

เป็น Heatmap ที่แสดงความหนาแน่นของข้อมูลจากจำนวนที่คนเดินผ่านในแต่ละพื้นที่ มีประโยชน์ในการมองภาพรวมเช่น ตำแหน่งที่ลูกค้าในร้านค้าเดินผ่านมากที่สุด

ภาพที่ 3.12 ตัวอย่าง Heatmap จากจำนวนคนของร้านค้า

3.6.2. การนับจำนวนคนเข้า – ออก (Person Counting)

เป็นการนับจำนวนคนเข้า – ออก โดยนับจากตำแหน่งของเท้าของคนที่ตรวจจับได้โดย Person Detection เป็นตำแหน่งอ้างอิงว่าคนในภาพเข้าหรือออกจากตำแหน่งที่กำหนดไว้

เนื่องจากการตรวจจับตำแหน่งคนในภาพในโครงการนี้ใช้เป็นการตรวจจับด้วยการทำนายจาก Neural Network ซึ่งมีโอกาสที่ตำแหน่งของคนจะคลาดเคลื่อนและกวัดแกว่งได้ง่าย ซึ่งจะส่งผลให้การนับคนผิดพลาดได้ง่ายดาย เราจึงแก้ปัญหาให้การนับแบบเส้นจำลองคู่ (Double Virtual line) ซึ่งเป็นการนับว่าคนผ่านเมื่อคนเดินผ่านเส้นจำลองที่ขีดไว้ทั้ง 2 เส้นแล้วเท่านั้น ด้วยวิธีนี้จึงสามารถแก้ปัญหาความคลาดเคลื่อนได้ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.2.1 แบบเส้น

ภาพที่ 3.13 ตัวอย่างการนับจำนวนคนเข้าออกแบบเส้น

3.6.3. ส่งออกข้อมูลในรูปแบบ Excel

ส่งออกข้อมูลของการวิเคราะห์ได้แก่ จำนวนคนในภาพแยกแต่ละเพศ จำนวนเวลาเฉลี่ยที่คนแต่ละคนใช้ จำนวนเวลานานที่สุดที่คนใช้ เป็นต้น และสามารถวิเคราะห์แยกตามช่วงเวลา เช่น ทุก ๆ 15 นาที เป็นต้น

	A	B	C	D	E	F	G	H	I	J	K	L
1	Entry name	Total income	Total male income	Total female income	Total unspecified income	Total outcome	Total male outcome	Total female outcome	Total unspecified outcome	Start time	End time	Interval time (sec
2	Entrance	64	29	35	0	74	40	34	0	2019-07-05 8:52:59	2019-07-05 9:07:59	
3	Entrance	72	30	42	0	101	43	58	0	2019-07-05 9:07:59	2019-07-05 9:22:59	
4	Entrance	88	33	55	0	120	46	74	0	2019-07-05 9:22:59	2019-07-05 9:37:59	
5	Entrance	115	41	74	0	152	57	95	0	2019-07-05 9:37:59	2019-07-05 9:52:28	
6	Entrance	222	79	143	0	263	98	165	0	2019-07-05 8:52:59	2019-07-05 9:52:28	3
7												

ภาพที่ 3.14 ตัวอย่างข้อมูลตารางของการวิเคราะห์ทั้งวิดีโอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A	B	C	D	E	F	G	H	I	J	K
1	Zone ID	Total	Total male	Total female	Total unspecified	Average time spent per person (seconds)	Max time spent (seconds)	Start time	End time	Interval time (seconds)	
2	0	83	38	30	15	3.36746988	54.65	2019-07-05 8:52:59	2019-07-05 9:07:59	900	
3	1	27	14	13	0	3.792592593	18.2	2019-07-05 8:52:59	2019-07-05 9:07:59	900	
4	2	17	12	2	3	1.008823529	2.15	2019-07-05 8:52:59	2019-07-05 9:07:59	900	
5	3	10	3	7	0	0.825	1.35	2019-07-05 8:52:59	2019-07-05 9:07:59	900	
6	0	39	5	24	10	3.357692308	14	2019-07-05 9:07:59	2019-07-05 9:22:59	900	
7	1	17	1	15	1	3.755882353	7.65	2019-07-05 9:07:59	2019-07-05 9:22:59	900	
8	2	7	0	3	4	0.65	1.2	2019-07-05 9:07:59	2019-07-05 9:22:59	900	
9	3	10	1	9	0	1.55	8.05	2019-07-05 9:07:59	2019-07-05 9:22:59	900	
10	0	43	14	21	8	2.620930233	13.45	2019-07-05 9:22:59	2019-07-05 9:37:59	900	
11	1	17	2	15	0	2.173529412	6.75	2019-07-05 9:22:59	2019-07-05 9:37:59	900	
12	2	8	0	3	5	0.525	1.35	2019-07-05 9:22:59	2019-07-05 9:37:59	900	
13	3	6	1	5	0	0.883333333	2.1	2019-07-05 9:22:59	2019-07-05 9:37:59	900	
14	0	40	12	24	4	4.37625	18.2	2019-07-05 9:37:59	2019-07-05 9:52:28	869.4	
15	1	22	6	16	0	3.229545455	14.45	2019-07-05 9:37:59	2019-07-05 9:52:28	869.4	
16	2	11	0	7	4	0.818181818	2.35	2019-07-05 9:37:59	2019-07-05 9:52:28	869.4	
17	3	7	1	6	0	0.828571429	2.1	2019-07-05 9:37:59	2019-07-05 9:52:28	869.4	
18	0	204	69	98	37	3.422794118	54.65	2019-07-05 8:52:59	2019-07-05 9:52:28	3569.4	
19	1	83	23	59	1	3.304216867	18.2	2019-07-05 8:52:59	2019-07-05 9:52:28	3569.4	
20	2	43	12	15	16	0.811627907	2.35	2019-07-05 8:52:59	2019-07-05 9:52:28	3569.4	
21	3	32	6	26	0	1.0890625	8.05	2019-07-05 8:52:59	2019-07-05 9:52:28	3569.4	

ภาพที่ 3.15 ตัวอย่างข้อมูลตารางของการวิเคราะห์เฉพาะพื้นที่

3.6.4. ส่งออกข้อมูลในรูปแบบ JSON

ส่งออกข้อมูลทั้งหมดของการวิเคราะห์เพื่อการนำไปใช้งานต่อได้ง่ายดายมากขึ้น มีโครงสร้างเบื้องต้นดังนี้

```
{
  "persons":{
    "1":{
      "id":1,
      "objs":[
        {
          "person_id":1,
          "frame_idx":116,
          "gender":1,
          "bbox":[
            327,
            215,
            436,
            434
          ],
          "score":0.9997922778
        },
        {
          "person_id":1,
          "frame_idx":117,
          "gender":1,
          "bbox":[
            327,
            214,
            436,
            433
          ],
          "score":0.9996839762
        },
      }
    }
}
```

ภาพที่ 3.16 ตัวอย่างของข้อมูลในรูปแบบไฟล์ JSON

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.5. ส่งออกข้อมูลในรูปแบบวิดีโอ

เพื่อให้เห็นการทำงานที่ชัดเจนของระบบจึงจำเป็นต้องมีการแสดงผลการวิเคราะห์เป็นแบบวิดีโอ โดยสิ่งที่แสดงผลมีดังนี้

ตารางที่ 3.2 แสดงลักษณะการแสดงผลของการวิเคราะห์ในวิดีโอ

ข้อมูล	ลักษณะการแสดงผล
ตำแหน่งของคน	กรอบ
เพศ	ข้อความที่มุมบนซ้ายและสีของกรอบ สีน้ำเงิน แทน เพศชาย สีแดง แทน เพศหญิง สีดำ แทน ระบุไม่ได้
คะแนนความมั่นใจของการทำนายเพศ	ตัวเลขด้านบน ถัดมาจากข้อความเพศ
ID ของคน	ตัวเลขที่มุมล่างขวา
สถิติจำนวนคนในแต่ละเพศ	ข้อความที่มุมบนขวาของวิดีโอ
ข้อมูลเวลาเฉลี่ยที่แต่ละคนในภาพใช้ในวิดีโอ	ข้อความที่มุมบนซ้าย

ตารางที่ 3.3 แสดงลักษณะการแสดงผลของการวิเคราะห์การเข้า-ออกในวิดีโอ

ข้อมูล	ลักษณะการแสดงผล
เส้นนับจำนวนคนเข้า - ออก	เส้นตรงสีแดง 2 เส้น
จำนวนคนเดินเข้า - ออก	ข้อความที่มุมขวาล่างของเส้น

ตารางที่ 3.4 แสดงลักษณะการแสดงผลของการวิเคราะห์การเข้า-ออกพื้นที่ในวิดีโอ

ข้อมูล	ลักษณะการแสดงผล
พื้นที่นับจำนวนคนเข้า - ออก	กรอบสี่เหลี่ยมสีแดง 2 กรอบ
จำนวนคนเดินเข้า - ออก	ข้อความที่มุมขวาล่างของกรอบสี่เหลี่ยม
สถิติจำนวนคนในแต่ละเพศที่เข้าพื้นที่	ข้อความที่มุมบนขวาของกรอบสี่เหลี่ยม
ข้อมูลเวลาเฉลี่ยที่แต่ละคนในภาพใช้ภายในพื้นที่	ข้อความที่มุมบนซ้ายของกรอบสี่เหลี่ยม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 3.17 ตัวอย่างภาพในวิดีโอหลังจากการวิเคราะห์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัย

4.1. รายละเอียดการทดสอบ

เนื่องจากวิดีโอกล้องวงจรปิดที่ใช้ทดสอบระบบเป็นความลับของบริษัทจึงไม่สามารถนำวิดีโอที่ทดสอบมาเผยแพร่ได้ ผู้จัดทำจึงแก้ปัญหาด้วยติดตั้งและบันทึกวิดีโอกล้องวงจรปิดด้วยตัวเองในพื้นที่ที่สามารถติดตั้งได้ ได้ทดสอบการวิเคราะห์ด้วยวิดีโอจากกล้องวงจรปิดบริเวณทางเดินชั้น 11 ของบริษัทกสิกรบิซสิเนสเทคโนโลยีกรุ๊ปจำกัดในวันที่ 7 สิงหาคม พ.ศ. 2562 เวลา 16.06 น. ถึง 16.51 น. รวมทั้งสิ้น 45 นาที

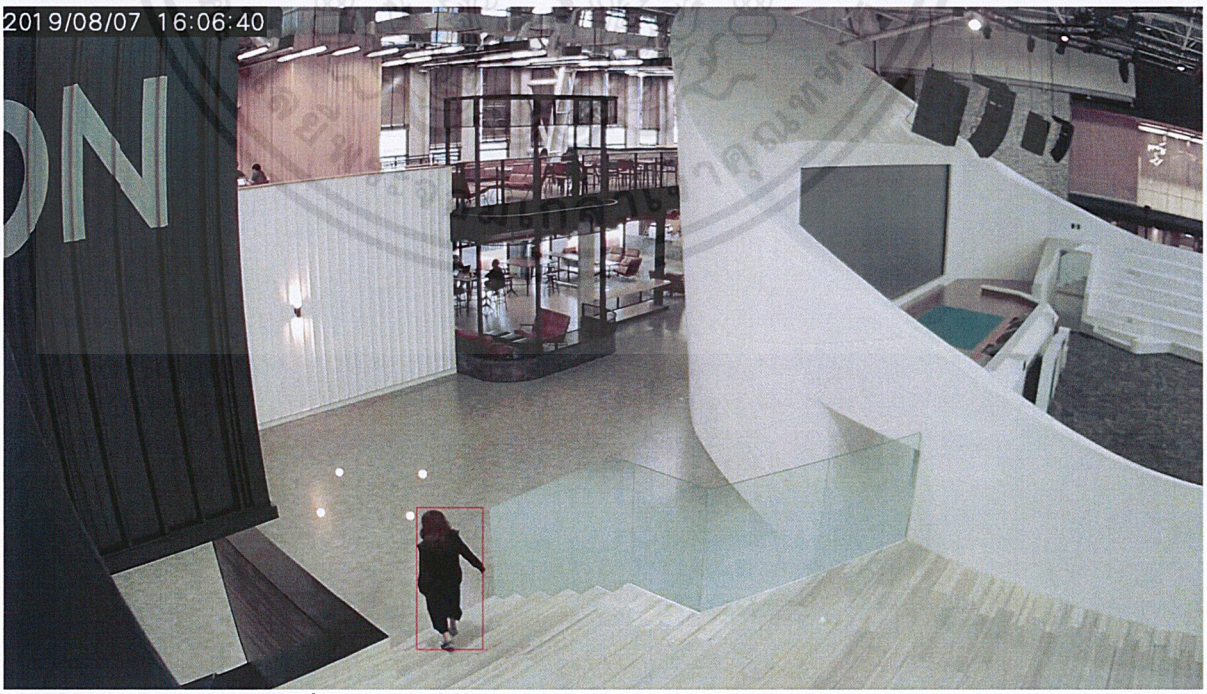
ภาพที่ 4.1 ตัวอย่างสถานที่ทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2. การตรวจจับตำแหน่งบุคคล (Person Detection)

การตรวจจับคนในภาพ (Person Detection) คือการแยกตำแหน่งของภาพบุคคลออกจากภาพพื้นหลังที่เราไม่สนใจ ตำแหน่งของบุคคลที่ตรวจจับได้จะถูกวาดกรอบสี่เหลี่ยมเพื่อแสดงผลว่าระบบตรวจจับบุคคลได้

ภาพที่ 4.2 ก่อนการตรวจจับตำแหน่งบุคคล



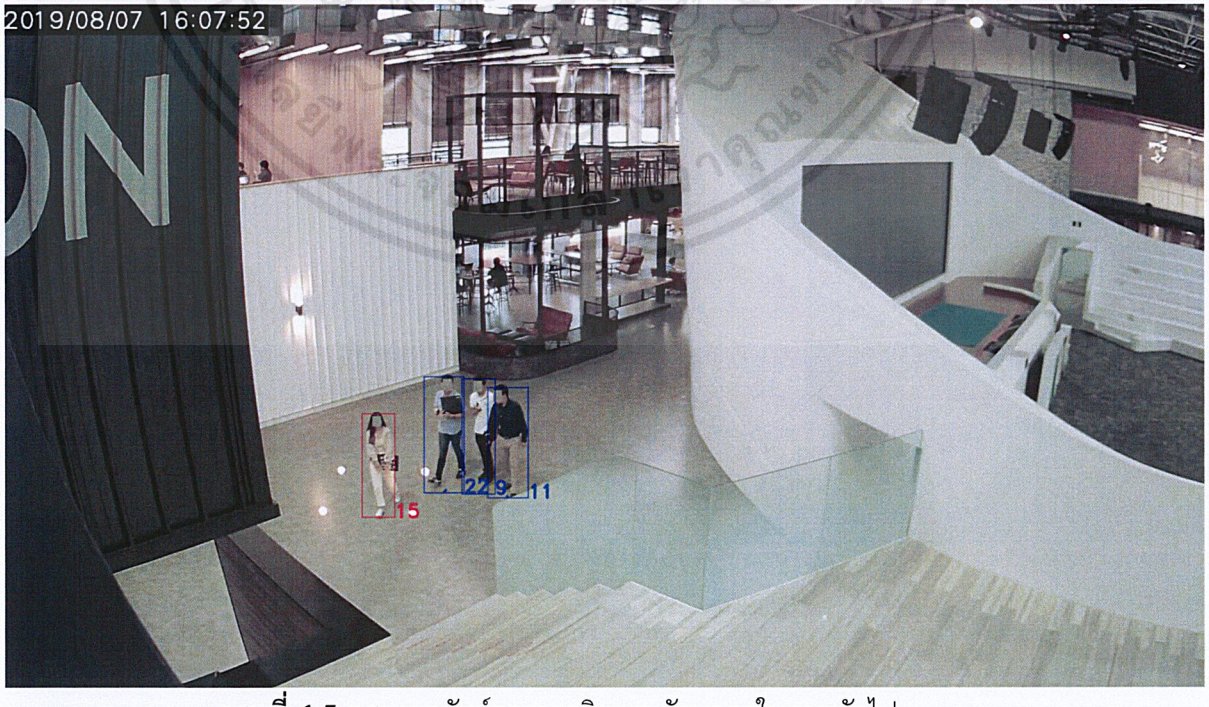
ภาพที่ 4.3 แสดงผลลัพธ์หลังการตรวจจับตำแหน่งบุคคล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3. การติดตามตัวบุคคล (Person Tracking)

การติดตามตำแหน่งคนในภาพเพื่อระบุว่าคนที่ตรวจจับเจอเป็นคนเดิมกับภาพก่อนหน้าหรือไม่ การติดตามตัวจะแสดงผลในรูปแบบของตัวเลข ID ของแต่ละคนที่มุมขวาล่างของกรอบบุคคลแต่ละคน

ภาพที่ 4.4 แสดงผลลัพธ์ของการติดตามตัวบุคคล



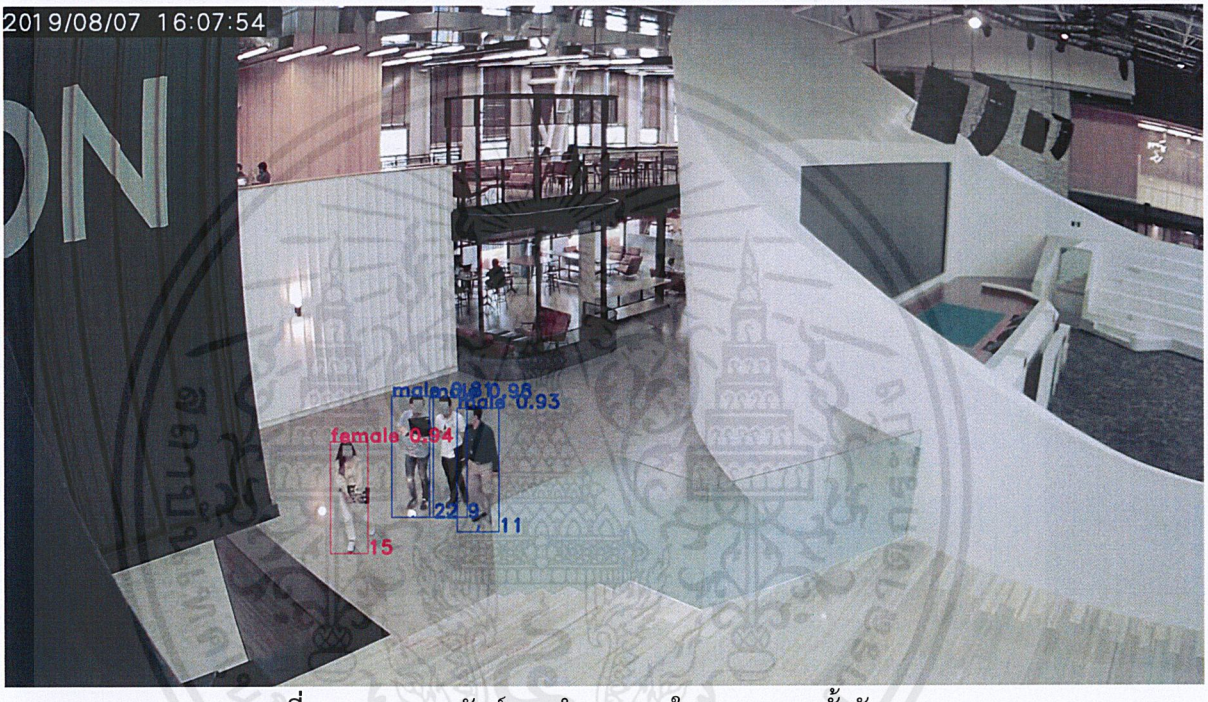
ภาพที่ 4.5 แสดงผลลัพธ์ของการติดตามตัวบุคคลในภาพถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4. การรู้จำเพศด้วยภาพบุคคลทั้งตัว (Full-Body Gender Recognition)

การรู้จำเพศบุคคลโดยใช้ภาพบุคคลทั้งตัว แสดงผลเพศที่ทำนายด้วยข้อความด้านบนของกรอบบุคคลและสีของกรอบ ดังนี้

- ข้อความ “male” และกรอบสีน้ำเงิน แทน เพศชาย
- ข้อความ “female” และกรอบสีแดง แทน เพศหญิง
- ข้อความ “unspecified” และกรอบสีดำ แทน ไม่สามารถระบุเพศได้



ภาพที่ 4.6 แสดงผลลัพธ์การทำนายเพศในภาพบุคคลทั้งตัว

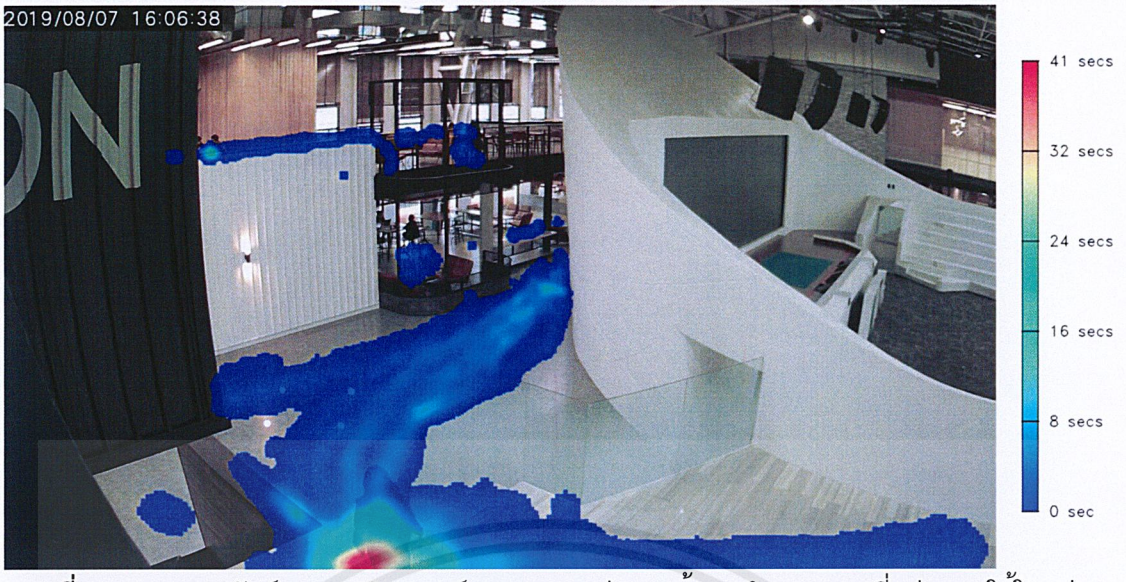
4.5. Heatmap

เป็นการแสดงความหนาแน่นของข้อมูลของแต่ละพื้นที่ด้วยการไล่ระดับสี โดยระบบจะใช้ภาพพื้นหลังเป็นภาพแรกของวิดีโอและแสดงขนาดของความหนาแน่นของแต่ละสีที่ด้านขวาของภาพ

4.5.1. Heatmap จากจำนวนเวลาที่ใช้

เป็น Heatmap ที่แสดงความหนาแน่นของข้อมูลของเวลาที่แต่ละคนใช้ในแต่ละพื้นที่ จากภาพจะสังเกตได้ว่าจะมีความหนาแน่นของเวลาที่แต่ละคนใช้สูงเป็นพิเศษบริเวณบันไดของทางเดิน ที่เป็นเช่นนี้เพราะว่าการเดินขึ้นบันไดใช้เวลามากกว่าการเดินบนทางเรียบ

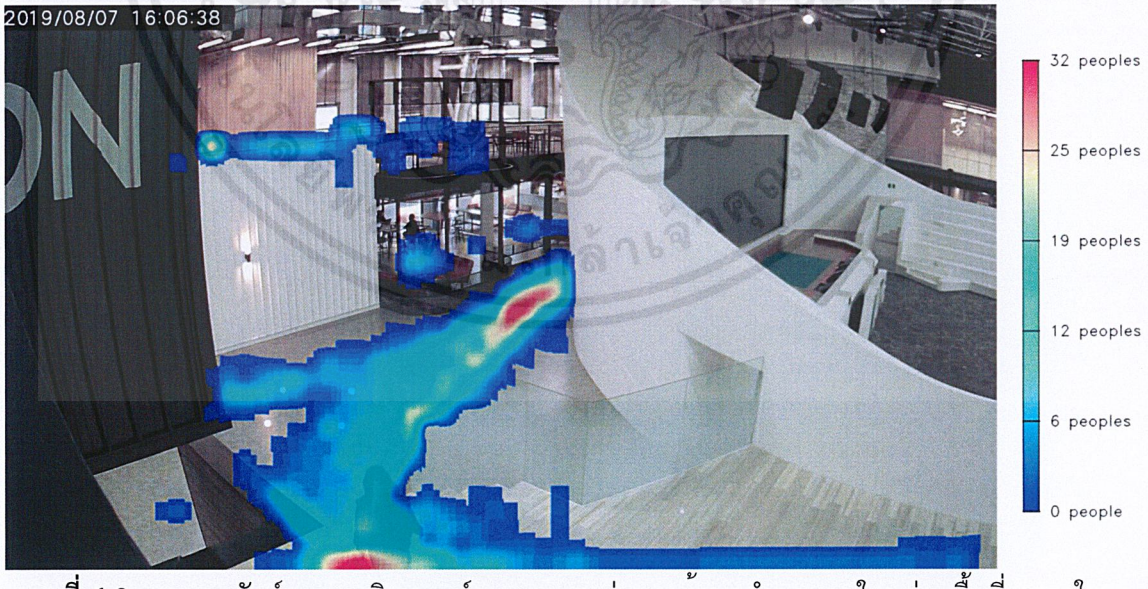
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.7 แสดงผลลัพธ์ของการวิเคราะห์ความหนาแน่นของข้อมูลจำนวนเวลาที่แต่ละคนใช้ในแต่ละพื้นที่ แสดงในรูปแบบ Heatmap

4.5.2. Heatmap จากจำนวนคน

เป็น Heatmap ที่แสดงความหนาแน่นของข้อมูลจากจำนวนที่คนเดินผ่านในแต่ละพื้นที่ จากภาพจะสังเกตได้ว่ามีความหนาแน่นของจำนวนคนสูงเป็นพิเศษบริเวณทางเดินที่กลางภาพและบันได ที่เป็นเช่นนี้เนื่องจากบริเวณทางเดินที่กลางภาพมีขนาดแคบกว่าส่วนพื้นที่โล่ง



ภาพที่ 4.8 แสดงผลลัพธ์ของการวิเคราะห์ความหนาแน่นของข้อมูลจำนวนคนในแต่ละพื้นที่ แสดงในรูปแบบ Heatmap

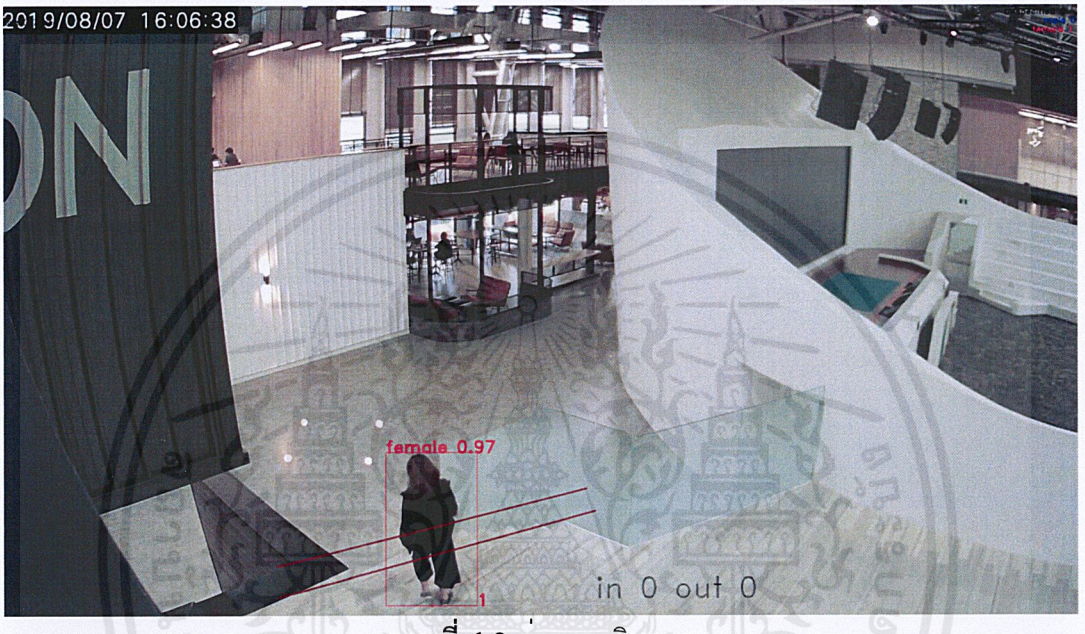
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6. การนับจำนวนคนเข้า - ออก (Person Counting)

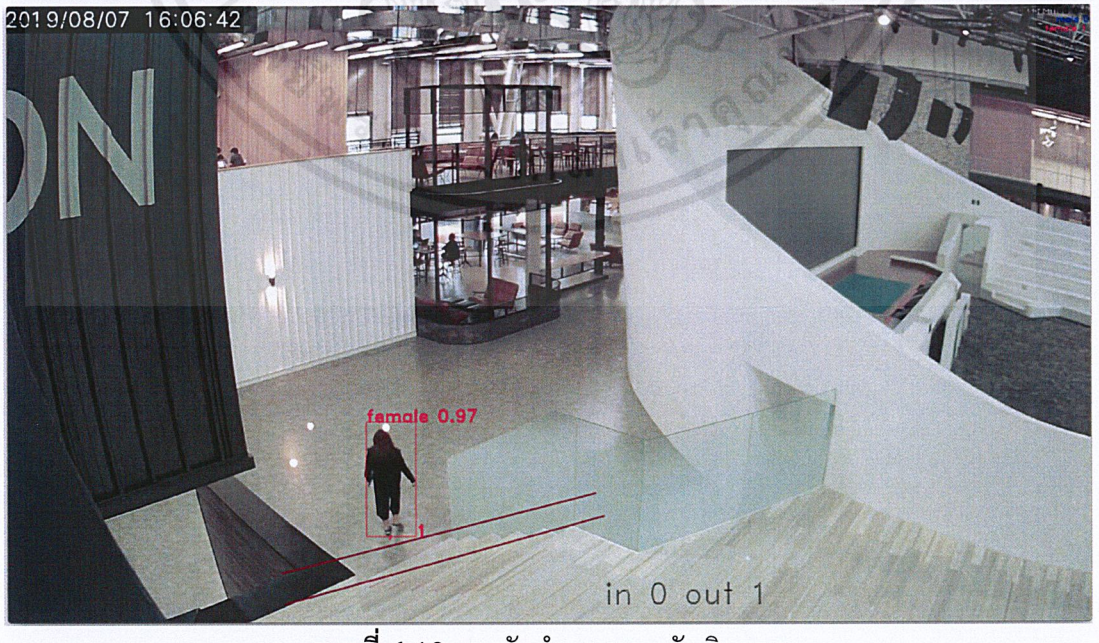
เป็นการนับจำนวนคนเข้า – ออก โดยนับจากตำแหน่งของเท้าของคนที่ตรวจจับได้โดย Person Detection เป็นตำแหน่งอ้างอิงว่าคนในภาพเข้าหรือออกจากตำแหน่งที่กำหนดไว้

4.6.1. การนับจำนวนคนเข้า – ออกแบบเส้น

นับจำนวนคนเข้า – ออกโดยการวาดเส้นจำลอง 2 เส้น โดยจะนับคนเข้าหรือออกเมื่อบุคคลนั้นเดินตัดผ่านเส้นทั้งสองเส้น แสดงผลการนับที่มุมล่างขวาของเส้นที่วาด



ภาพที่ 4.9 ก่อนคนเดินออก



ภาพที่ 4.10 การนับจำนวนคนหลังเดินออก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.7. การส่งออกข้อมูล

4.7.1. แบบไฟล์ Excel

เพื่อให้ง่ายต่อการวิเคราะห์ภาพรวม ส่งออกข้อมูลในรูปแบบตารางดังนี้

ตารางที่ 4.1 สรุปผลการวิเคราะห์ในแต่ละพื้นที่

ID พื้นที่	จำนวนคนทั้งหมด	ชาย	หญิง	ระบุไม่ได้	เวลาเฉลี่ย (วินาที)	เวลาสูงสุด (วินาที)	เวลาเริ่ม	เวลาจบ	ช่วงเวลา (วินาที)
0	61	42	17	2	7.59	31.05	2019-08-07 16:06:38	2019-08-07 16:21:38	900
0	46	30	13	3	5.58	26.55	2019-08-07 16:21:38	2019-08-07 16:36:38	900
0	28	18	9	1	6.53	22.7	2019-08-07 16:36:38	2019-08-07 16:49:13	755
0	134	89	39	6	6.74	31.05	2019-08-07 16:06:38	2019-08-07 16:49:13	2555

ตารางที่ 4.2 สรุปผลการนับจำนวนคนเข้า

ชื่อ	คนเข้าทั้งหมด	จำนวนชายเข้า	จำนวนหญิงเข้า	จำนวนไม่ระบุเข้า	เวลาเริ่ม	เวลาจบ	ช่วงเวลา (วินาที)
บันได	26	17	8	1	2019-08-07 16:06:38	2019-08-07 16:21:38	900
บันได	36	25	10	1	2019-08-07 16:21:38	2019-08-07 16:36:38	900
บันได	45	32	12	1	2019-08-07 16:36:38	2019-08-07 16:49:13	755
บันได	69	49	18	2	2019-08-07 16:06:38	2019-08-07 16:49:13	2555

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 สรุปผลการนับจำนวนคนออก

ชื่อ	คนออกทั้งหมด	จำนวนชายออก	จำนวนหญิงออก	จำนวนไม่ระบุออก	เวลาเริ่ม	เวลาจบ	ช่วงเวลา (วินาที)
บันได	26	17	8	1	2019-08-07 16:06:38	2019-08-07 16:21:38	900
บันได	37	25	11	1	2019-08-07 16:21:38	2019-08-07 16:36:38	900
บันได	46	32	13	1	2019-08-07 16:36:38	2019-08-07 16:49:13	755
บันได	69	48	20	1	2019-08-07 16:06:38	2019-08-07 16:49:13	2555

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.7.2. แบบไฟล์ JSON


```
{
  "persons":{
    "1":{
      "id":1,
      "objs":[
        {
          "person_id":1,
          "frame_idx":116,
          "gender":1,
          "bbox":[
            327,
            215,
            436,
            434
          ],
          "score":0.9997922778
        },
        {
          "person_id":1,
          "frame_idx":117,
          "gender":1,
          "bbox":[
            327,
            214,
            436,
            433
          ],
          "score":0.9996839762
        },
      }
  }
}
```


ภาพที่ 4.11 ตัวอย่างไฟล์ JSON

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.8. การเปรียบเทียบความแม่นยำกับงานวิจัยที่ใกล้เคียง

ในโจทย์การรู้จำเพศจากภาพเต็มตัว (Full-Body Gender Recognition) นั้นมีงานวิจับที่แก้ไขโจทย์นี้อยู่ไม่กี่ชิ้น การเปรียบเทียบนี้ใช้ชุดข้อมูลเดียวกันคือ PETA

ตารางที่ 4.4 เปรียบเทียบความแม่นยำของโมเดลแบบ Single frame

ชื่องานวิจัย	เทคนิคที่ใช้	ความแม่นยำชุดฝึกสอน (Train accuracy)	ความแม่นยำชุดทดสอบ (Test accuracy)
Pedestrian Attribute Detection using CNN	CNN + k-NN	-	0.5
Pedestrian Attribute Detection using CNN	CNN + SVM	-	0.55
Pedestrian Attribute Detection using CNN	CNN + FC	-	0.42
Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition	AlexNet-CNN + SVM	0.85	0.79
Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition	MiniCNN + SVM	0.80	0.75
Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition	HOG + SVM	0.72	0.56
โมเดลที่พัฒนาขึ้นเอง	ResNet50	0.9688	0.8477
โมเดลที่พัฒนาขึ้นเอง	MobileNet	0.8223	0.7871

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากชุดข้อมูล PETA มีคำอธิบายของรูปภาพบุคคล (Annotation) แบบระบุเฉพาะตัวบุคคลอยู่เพียงบางส่วน ทำให้การเปรียบเทียบความแม่นยำของโมเดลแบบ Multi frame นั้น ไม่สามารถนำไปใช้เปรียบเทียบกับงานวิจัยของโมเดลแบบ Single frame ได้

ตารางที่ 4.5 เปรียบเทียบความแม่นยำของโมเดลแบบ Multi frame

ชื่องานวิจัย	**เทคนิคที่ใช้**	**ความแม่นยำชุดฝึกสอน (Train accuracy)**	**ความแม่นยำชุดทดสอบ (Test accuracy)**
โมเดลที่พัฒนาขึ้นเอง	ResNet50 + GRU	0.8953	0.9046
โมเดลที่พัฒนาขึ้นเอง	ResNet50 + Mode	-	0.8670
โมเดลที่พัฒนาขึ้นเอง	ResNet50 + GRU + CTCloss	-	0.8096

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1. ผลการดำเนินการ

จากการทดสอบการวิเคราะห์กับวิดีโอกล้องวงจรปิด

1) ช่วยให้มองเห็นข้อมูลเชิงลึกที่มักถูกมองข้ามเช่นพื้นที่บางส่วนที่ไม่มีคนเดินผ่านเลย พื้นที่ส่วนที่มีคนเดินผ่านบ่อยเป็นพิเศษ เป็นต้น ช่วยให้ร้านค้าเข้าใจพฤติกรรมของลูกค้าได้มากขึ้น
2) ช่วยให้เข้าใจสัดส่วนของลูกค้าในแต่ละเพศที่เข้ามาในร้าน
3) ช่วยให้เข้าใจความหนาแน่นของลูกค้าแต่ละเพศในแต่ละช่วงเวลา
4) โมเดลรู้จำเพศจากภาพบุคคลทั้งตัว (Full-Body Gender Recognition) ค่อนข้างทำนายได้แม่นยำกับชุดข้อมูลวิดีโอที่คนในภาพใส่ชุดที่เห็นสัดส่วนหรือทรงผมชัดเจน เช่นชุดหน้าร้อน เป็นต้น แต่จะทำนายได้ไม่ดีกับภาพที่คนในภาพใส่ชุดที่เห็นสัดส่วนหรือลักษณะเฉพาะตัวไม่ชัดเจน เช่นชุดหน้าหนาว คนที่ใส่ชุดกันหนาวปกปิดมิดชิด เป็นต้น
5) สามารถพัฒนาแอปพลิเคชันต่อยอดการจากวิเคราะห์ได้

5.2. ปัญหาและอุปสรรคที่พบ

1) มีความไม่พร้อมในการเตรียมชุดข้อมูลวิดีโอกล้องวงจรปิด ทำให้การพัฒนาล่าช้าลง
2) ระบบเนตเวิร์กภายในบริษัทค่อนข้างเน้นความปลอดภัยซึ่งสร้างความยากลำบากในการพัฒนา
3) เนื่องจากการฝึกสอนโมเดลทางด้าน Deep Learning จำเป็นต้องใช้ฮาร์ดแวร์ที่มีคุณภาพสูง โดยเฉพาะอย่างยิ่งกับ GPU ซึ่งทรัพยากรด้านฮาร์ดแวร์ที่บริษัทจัดเตรียมไว้มีจำกัด บางครั้งไม่พอต่อการใช้งาน ทำให้เพิ่มความล่าช้าในการพัฒนา
4) ที่นั่งทำงานมีความไม่พร้อมและไม่เหมาะสม

5.3. ประโยชน์ที่ได้จากการปฏิบัติงานสหกิจศึกษา

1) ได้เรียนรู้วิธีการทำงานในองค์กรขนาดใหญ่ในด้านเทคโนโลยี เรียนรู้วิธีการทำงานในสายงานด้านการวิจัย
2) ได้เรียนรู้วิธีการมองปัญหา แตกปัญหา แก้ปัญหาในงานวิจัยให้ได้ประสิทธิภาพ
3) ได้ฝึกฝนและนำความรู้ที่มีมาใช้ในงานจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4. ข้อเสนอแนะ

1) ควรมีการพัฒนาเป็นกราฟฟิคแอปพลิเคชัน (GUI Application) เพื่อให้เพิ่มความง่ายต่อการใช้งานของผู้ใช้ปลายทาง (End user)
2) เนื่องจากระบบถูกพัฒนาเพื่อทำงานกับกล้องวงจรปิดเดียว ควรเพิ่มระบบเชื่อมต่อข้อมูลของหลายกล้องวงจรปิดเข้าด้วยกัน
3) เนื่องจากในร้านค้าจริงจะมีพนักงานของร้านรวมอยู่ในวิดีโอด้วย ควรเพิ่มการแยกแยะพนักงานออกจากการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

วรรณพงษ์ ภัททิยไพบูลย์. (2015). *เริ่มต้นกับ NumPy*. เรียกใช้เมื่อ 14 11 2019 จาก https://python3.wannaphong.com/2015/05/numpy.html

Alex Bewley, Dietrich Paulus Nicolai Wojke. (2017). *Simple Online and Realtime Tracking with a Deep Association Metric.* arXiv.

Ali Farhadi Joseph Redmon. (2017). *YOLOv3: An Incremental Improvement.* arXiv.

Athiwat. (2017). *CUDA คืออะไร*. เรียกใช้เมื่อ 14 11 2018 จาก https://medium.com/@athivvat/cuda-คืออะไร-baef4ec32963

Dipanjan (DJ) Sarkar. (15 11 2018). *A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning*. เข้าถึงได้จาก towardsdatascience.com: https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

Facebook AI Research. (2018). *Focal Loss for Dense Object Detection.* arXiv.

Haiyu Zhao, Maoqing Tian, Lu Sheng, Jing Shao, Shuai Yi, Junjie Yan, Xiaogang Wang Xihui Liu. (2017). *HydraPlus-Net: Attentive Deep Features for Pedestrian Analysis.* arXiv.

Jayanth Ramesh Agrim Gupta. (2016). *Pedestrian Attribute Detection using CNN.* arXiv.

Jingjing Wu, Jianguo Jiang, Meibin Qi, and Bo Ren Hao Liu. (2018). *Sequence-based Person Attribute Recognition with Joint CTC-Attention Model.* arXiv.

OpenCV คืออะไร ? (2018). เรียกใช้เมื่อ 14 11 2018 จาก www.mindphp.com/คู่มือ/73-คืออะไร/7061-what-is-opencv.html

Ping Luo, Chen Change Loy, Xiaoou Tang Yubin Deng. (2014). *Pedestrian Attribute Recognition At Far Distance.* ACM.

Prasit Tongpradit. (2018). *มาทำความรู้จัก Tensorflow*. เรียกใช้เมื่อ 14 11 2018 จาก www.thaiprogrammer.org/2018/12/มาทำความรู้จัก-tensorflow/

Python คืออะไร. (2018). เรียกใช้เมื่อ 14 11 2019 จาก www.mindphp.com/คู่มือ/73-คืออะไร/2417-python-คืออะไร.html

Sarayut Nonsiri. (2018). *https://www.9experttraining.com/articles/python-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%8*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4%E0%B8%A3. เรียกใช้เมื่อ 14 11 2019 จาก www.9experttraining.com/articles/python-คืออะไร

Sid-Ahmed Berrani, Natacha Ruchaud, Jean-Luc Dugelay Grigory Antipov. (2016). *Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition.* eurecom.

Suphakit Annoppornchai. (2017). *JSON คืออะไร*. เรียกใช้เมื่อ 14 11 2018 จาก https://saixiii.com/what-is-json/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้