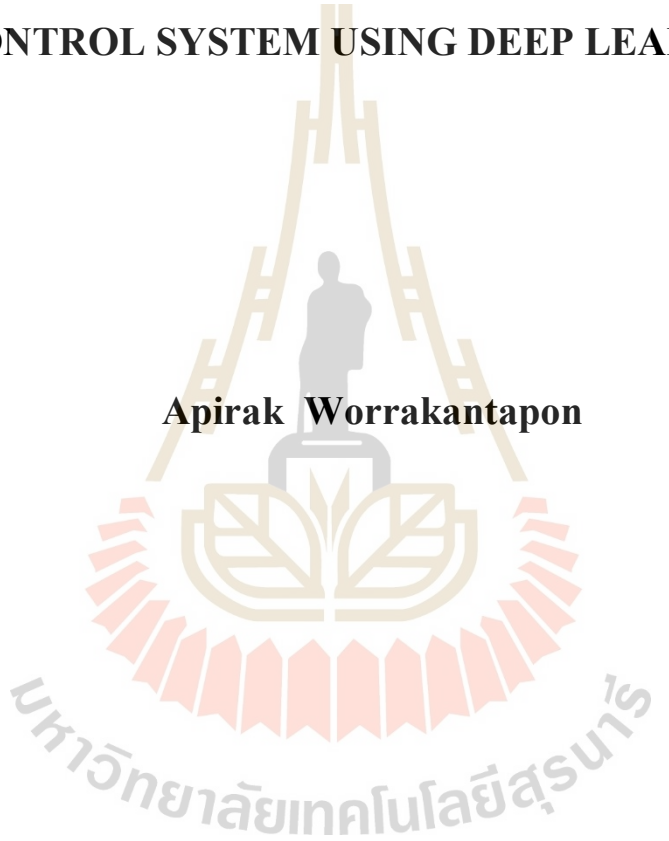
การพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์โดยใช้การเรียนรู้เชิงลึก

นายอภิรักษ์ วรกานตพล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโทรคมนาคมและคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2563

DEVELOPMENT OF A SYSTEM FOR HUMAN DETECTION AND RISK-LEVEL IDENTIFICATION TO PREVENT ACCIDENTS IN TRUCK DUMPER CONTROL SYSTEM USING DEEP LEARNING

Apirak Worrakantapon



A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Telecommunication and Computer Engineering

Suranaree University of Technology

Academic Year 2020

การพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์โดยใช้การเรียนรู้เชิงลึก

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

(รศ. ดร.กิตติศักดิ์ เกิดประสพ)
ประธานกรรมการ

(รศ. ดร.นิตยา เกิดประสพ)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

(ผศ. ดร.นันทวุฒิ คะอังกุ)
กรรมการ

(อ. ดร.รติพร จันทร์กลั่น)
กรรมการ

(รศ. ร.อ. ดร.กนต์ธร ชำนิประศาสน์)
รองอธิการบดีฝ่ายวิชาการและพัฒนาความเป็นสากล

(รศ. ดร. พรศิริ จงกล)
คณบดีสำนักวิชาวิศวกรรมศาสตร์

อภิรักษ์ วรกานตพล : การพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์โดยใช้การเรียนรู้เชิงลึก (DEVELOPMENT OF A SYSTEM FOR HUMAN DETECTION AND RISK-LEVEL IDENTIFICATION TO PREVENT ACCIDENTS IN TRUCK DUMPER CONTROL SYSTEM USING DEEP LEARNING) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา เกิดประสพ, 119 หน้า.

ในปัจจุบัน โรงงานอุตสาหกรรมได้ก้าวเข้าสู่ยุคของการผลิตด้วยระบบที่ชาญฉลาดและมีการควบคุมด้วยระบบอัตโนมัติ โดยเฉพาะในโรงงานผลิตอาหารสัตว์ ซึ่งประกอบด้วยเครื่องจักรที่ควบคุมด้วยระบบอัตโนมัติ เช่น เครื่องผสมอาหาร เครื่องอัดเม็ดอาหาร ทรัดดัมป์ เป็นต้น ทรัดดัมป์ถูกใช้ในการยกรถบรรทุกให้เอียงขึ้น เพื่อให้วัตถุดิบบนรถบรรทุกไหลลงมารวมกันยังบ่อรับวัตถุดิบ อย่างไรก็ตามในขณะที่ทรัดดัมป์ถูกยกตัวขึ้นนั้น พื้นที่ใกล้เคียงถือเป็นพื้นที่อันตราย เนื่องจากหากมีคนยืนอยู่ด้านบนทรัดดัมป์ขณะที่เครื่องจักรกำลังทำงาน อาจทำให้คนตกลงไปในบ่อรับวัตถุดิบได้และได้รับบาดเจ็บหรือเกิดอันตรายถึงแก่ชีวิตได้ ถึงแม้ว่าในปัจจุบันจะมีระบบอัตโนมัติมาช่วยสนับสนุนการทำงานของเครื่องจักรแต่ในการควบคุมทรัดดัมป์นั้นยังจำเป็นต้องใช้พนักงานในการสั่งงานจากในห้องควบคุม พนักงานต้องคอยสังเกตการณ์ผู้คนในพื้นที่อันตรายอย่างเข้มงวดก่อนเริ่มเดินเครื่องเพื่อความปลอดภัย แต่อย่างไรก็ตามด้วยตำแหน่งของห้องควบคุมนั้น พนักงานไม่สามารถมองเห็นบุคคลที่ยืนอยู่บนพื้นที่อันตรายได้ทั่วทุกจุด เนื่องจากมุมในการมองเห็นของพนักงานจากในห้องควบคุมนั้นถูกบดบังด้วยสิ่งกีดขวางต่าง ๆ เช่น กำแพง บันได รถบรรทุก เป็นต้น

เทคโนโลยีการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกได้ถูกนำมาใช้อย่างกว้างขวางในงานด้านการตรวจจับวัตถุ ในงานวิจัยนี้ ผู้วิจัยได้พัฒนาโมเดลสำหรับการตรวจจับบุคคลอัตโนมัติ โดยใช้โครงข่ายประสาทเทียมคอนโวลูชันในการเรียนรู้และจดจำรูปแบบของคุณลักษณะสำคัญของวัตถุเป้าหมาย วัตถุประสงค์งานวิจัยคือ พัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยง เพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์ ผู้วิจัยได้ทดลองและคัดเลือกสถาปัตยกรรมการตรวจจับวัตถุที่มีประสิทธิภาพและเหมาะสมต่อการประยุกต์ใช้ร่วมกับระบบควบคุมทรัดดัมป์ จากผลการวิจัยพบว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 มีประสิทธิภาพสูงกว่า Faster R-CNN ทั้งทางด้านความแม่นยำและความเร็วในการประมวลผล โดยมีค่าความเที่ยงตรงเฉลี่ยเมื่อทดสอบกับภาพเวลากลางวันสูงถึง 99.93% และ 94.25% ในเวลากลางคืน มีความถูกต้องในการระบุความเสี่ยงสูงถึง 94.18% จากชุดข้อมูลทดสอบทั้งหมด โดยมีความเร็วในการประมวลผลภาพต่อวินาทีเฉลี่ยสูงถึง 31.96 ภาพต่อวินาที

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา อภิรักษ์ วรกานตพล

ปีการศึกษา 2563 ลายมือชื่ออาจารย์ที่ปรึกษา

APIRAK WORRAKANTAPON : DEVELOPMENT OF A SYSTEM FOR HUMAN DETECTION AND RISK-LEVEL IDENTIFICATION TO PREVENT ACCIDENTS IN TRUCK DUMPER CONTROL SYSTEM USING DEEP LEARNING. THESIS ADVISOR : ASSOC. PROF. NITAYA KERDPRASOP, Ph.D. 119 PP.

DEEP LEARNING/CONVOLUTION NEURAL NETWORK/OBJECT DETECTION/HUMAN DETECTION/RISK-LEVEL IDENTIFICATION

Presently, industrial factories have moved toward the era of intelligent automation systems, especially, animal feed industries. They contained several automation machines such as a mixer, pellet mill, and truck dumper. The truck dumper has been used to lift the whole truck and dump the raw material into an intake hopper. However, during its process, neighbor areas are identified as a risk area. Thus, if there is a person standing on the truck dumper platform at the time, this may cause fatal injury or death. Even though, the automation system has an important role to control the machines, but not for the truck dumper. It still needs staff to control the system process in a control room. The staff has to strictly observe people nearby before operating the machine. However, it is difficult to observe every person in the risk area because the staff's vision has been blocked by obstructions, e.g., walls, stairs, trucks, etc.

Machine learning and deep learning techniques are widely used in object detection tasks. In this study, a new proposal to develop a model of automatic human detection using a convolutional neural network that learns and recognizes the important characteristics of the target objects has been introduced. The objective of the study was to develop a system of human detection and risk-level identification in the truck dumper control system to prevent accidents. Herein, several experiments have been conducted in order to select the effective and optimal object detection architecture to apply to the truck dumper control system. In the results, the YOLOv4 model outperformed the Faster R-CNN model in both precision and processing speed. The average precision was 99.93% on a day-time dataset and 94.25% on a night-time dataset. The overall accuracy of risk identification was 94.18%. An average processing speed was 31.96 frames per second.

School of Computer Engineering | Student's Signature ______________

Academic Year 2020 | Advisor's Signature ______________

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ บุคคล และกลุ่มบุคคลต่าง ๆ ที่ได้กรุณาให้คำปรึกษา แนะนำ ช่วยเหลืออย่างดียิ่ง ทั้งในด้านวิชาการ และด้านการดำเนินงานวิจัย ดังต่อไปนี้

รองศาสตราจารย์ ดร.กิตติศักดิ์ เกิดประสพ และรองศาสตราจารย์ ดร.นิตยา เกิดประสพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้โอกาสทางการศึกษา อบรม ให้คำปรึกษา คำแนะนำในการทำงานวิจัย ตลอดจนถึงการจัดรูปแบบวิทยานิพนธ์ และช่วยตรวจทานความถูกต้องของวิทยานิพนธ์

อาจารย์ ดร.ศรัญญา กาญจนวัฒนา และอาจารย์ ดร.วรรธนะ พงษ์เสนา ที่คอยให้กำลังใจ และให้คำปรึกษาในการดำเนินงานวิจัยมาโดยตลอด

คุณอนุพงษ์ บรรจงการ คุณอนุสรา หิรัญวนากุล และคุณปิยวัฒน์ ศรีประเสริฐ ที่ให้ความช่วยเหลือในการช่วยตรวจทาน ติชม เพื่อปรับปรุงแก้ไขวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น

คุณปราณี กฐินใหม่ เลขานุการสาขาวิชาวิศวกรรมคอมพิวเตอร์ ที่ให้ความช่วยเหลือในการประสานงานระหว่างศึกษา

ขอขอบคุณนักศึกษาร่วมชั้นเรียนทั้งปริญญาโท และปริญญาเอก ที่ให้คำแนะนำคำปรึกษาด้านวิชาการ และช่วยสนับสนุนด้วยดีมาตลอด

ขอบคุณมหาวิทยาลัยเทคโนโลยีสุรนารี ที่ให้การสนับสนุนทุนการศึกษา ทุนวิจัย ทั้งยังค่าใช้จ่ายต่าง ๆ

นอกจากนี้ขอขอบคุณ ครู อาจารย์ ทั้งในอดีตและปัจจุบันที่ให้ความรู้แก่ผู้วิจัยอย่างมากมาย จนประสบความสำเร็จ

ท้ายที่สุดขอกราบขอบพระคุณ บิดา มารดา ที่ให้กำเนิด อบรม เลี้ยงดูด้วยความรัก ส่งเสริมการศึกษาเป็นอย่างดีมาโดยตลอด ทำให้ผู้วิจัยมีความรู้ ความสามารถ มีจิตใจที่เข้มแข็งและไม่ย่อท้อต่ออุปสรรคต่าง ๆ อีกทั้งยังเป็นกำลังใจแก่ผู้วิจัยจนประสบความสำเร็จในชีวิต

อภิรักษ์ วรกานตพล

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย) ..... ก
บทคัดย่อ (ภาษาอังกฤษ) ..... ข
กิตติกรรมประกาศ ..... ง
สารบัญ ..... จ
สารบัญตาราง ..... ซ
สารบัญรูป ..... ฎ
บทที่
1 บทนำ 1
1.1 ความสำคัญและที่มาของปัญหาการวิจัย 1
1.2 วัตถุประสงค์การวิจัย 4
1.3 ขอบเขตของการวิจัย 4
1.4 ประโยชน์ที่คาดว่าจะได้รับ 5
2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 6
2.1 ระบบควบคุมการทำงานของทรัดดัมป์ 6
2.1.1 หน้าจอระบบควบคุมเครื่องจักร (Human Machine Interface) 6
2.1.2 ระบบรับส่งข้อมูลระหว่างอุปกรณ์ 8
2.1.3 การทำงานของโปรแกรมพีแอลซี (Programmable Logic Controller) 9
2.1.4 ระบบควบคุมทรัดดัมป์ (Truck Dumper Control System) 10
2.2 เทคนิคการเรียนรู้เชิงลึก 12
2.2.1 การเรียนรู้เชิงลึก (Deep Learning) 13
2.2.2 โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network) 9
2.2.3 การตรวจจับวัตถุ (Object Detection) 25
2.2.4 สถาปัตยกรรม Faster R-CNN 27

สารบัญ (ต่อ)

หน้า

2.2.5 สถาปัตยกรรม YOLO (You Only Look Once) 31

2.3 มาตรวัดประสิทธิภาพในงานด้านการตรวจจับวัตถุ 39

2.3.1 ค่าความเที่ยงตรง (Precision) 40

2.3.2 ค่าการระลึก (Recall) 40

2.3.3 ค่าประสิทธิภาพโดยรวม (F-measure) 40

2.3.4 ค่าความเที่ยงตรงเฉลี่ย (Average Precision) 40

2.3.5 ความเร็วในการประมวลผลภาพต่อวินาที (Frame per Second) 44

2.4 งานวิจัยที่เกี่ยวข้อง 45

3 วิธีดำเนินงานวิจัย 50

3.1 กรอบแนวคิดงานวิจัย 50

3.2 กระบวนการวิจัย 51

3.2.1 การเตรียมข้อมูล 52

3.2.2 การแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ 55

3.2.3 การวัดประสิทธิภาพโมเดลเพื่อคัดเลือกสถาปัตยกรรมการตรวจจับระหว่าง Faster R-CNN และ YOLOv4 56

3.2.4 การพัฒนาโมเดลและวัดประสิทธิภาพเพื่อคัดเลือกโมเดลการบุคคลจากชุดฝึกสอนในแต่ละแบบ 56

3.2.5 การนำโมเดลมาพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยงของบุคคล 57

3.3 การรวมระบบที่นำเสนอเข้ากับระบบควบคุมเครื่องจักร 59

3.4 เครื่องมือที่ใช้ในการวิจัย 60

4 การทดสอบและอภิปรายผล 62

4.1 ข้อมูลที่ใช้ในการทดสอบ 62

4.1.1 ชุดข้อมูลสำหรับทดสอบสถาปัตยกรรมการตรวจจับวัตถุ 63

4.1.2 ชุดข้อมูลสำหรับทดสอบโมเดลการตรวจจับบุคคลและระบุระดับความเสี่ยง 63

4.2 ผลการทดสอบประสิทธิภาพสถาปัตยกรรมการตรวจจับวัตถุ 64

สารบัญ (ต่อ)

หน้า

4.2.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN 64
4.2.2 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4....66
4.2.3 สรุปผลการทดสอบโมเดลจากสถาปัตยกรรม Faster R-CNN และ YOLOv4....69
4.3 ผลการทดสอบประสิทธิภาพโมเดลการตรวจจับบุคคลและระบุระดับความเสี่ยง....71
4.3.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท 71
4.3.2 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล 76
4.3.3 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง....81
4.3.4 สรุปผลการทดสอบโมเดลที่พัฒนาขึ้นทั้งหมด....86
4.4 การพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยง....91
4.5 อภิปรายผล....93
5 บทสรุปและข้อเสนอแนะ....95
5.1 สรุปผลการวิจัย....95
5.2 การประยุกต์ผลการวิจัย....97
5.3 ข้อเสนอแนะ....98
รายการอ้างอิง....100
ภาคผนวก บทความวิจัยที่ตีพิมพ์ระหว่างการศึกษา....103
ประวัติผู้เขียน....119

สารบัญตาราง

ตารางที่ **หน้า**

2.1 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยง 48

3.1 สรุปสัดส่วนการแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ 56

4.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบ กับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด 65

4.2 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด 66

4.3 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด 67

4.4 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด 68

4.5 สรุปผลการทดสอบโมเดลจากสถาปัตยกรรม Faster R-CNN และ YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด 69

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.6 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 72

4.7 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 73

4.8 ผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 74

4.9 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 76

4.10 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 78

4.11 ผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 79

4.12 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมดทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 82

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.13 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 83

4.14 ผลการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด 84

4.15 สรุปผลการทดสอบโมเดลพี่พัฒนาขึ้นทั้งหมด 87

5.1 สรุปผลการทดสอบประสิทธิภาพโมเดลในงานวิจัยนี้ทั้งหมด 97

สารบัญรูป

รูปที่ **หน้า**

2.1 ตัวอย่างหน้าจอระบบควบคุมเครื่องจักร 7
2.2 โครงสร้างการทำงานของระบบ SCADA 7
2.3 โครงสร้างการรับส่งข้อมูลผ่าน OPC Server 8
2.4 สถาปัตยกรรมการทำงานภายใน PLC 9
2.5 การทำงานของโปรแกรม Ladder Logic ภายใน PLC 10
2.6 ตัวอย่างการทำงานของแท่นทรัคดัมป์ (Truck Dumper) 11
2.7 แนวคิดเบื้องต้นสำหรับการใช้เทคนิคการเรียนรู้เชิงลึกกับข้อมูลภาพลายมือเขียนตัวเลข 12
2.8 กระบวนการเรียนรู้ภายในโมเดลการเรียนรู้เชิงลึก 13
2.9 สถาปัตยกรรมพื้นฐานของโครงข่ายประสาทเทียม (Neural Network) 14
2.10 โครงสร้างเพอร์เซปตรอน (Perceptron) 15
2.11 โครงสร้างเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron) 15
2.12 รูปแบบการปรับค่าน้ำหนักจากเทคนิคการแพร่ย้อนกลับ (Backpropagation) 16
2.13 ตัวอย่างข้อมูลรูปภาพดิจิทัล แบบ Grayscale และ RGB 17
2.14 ตัวอย่างการแทนค่าในรูปภาพโหมด RGB 18
2.15 ตัวอย่างการจัดเรียงตำแหน่งของพิกเซลในรูปภาพ 18
2.16 ตัวอย่างงานด้านการจำแนกประเภทรูปภาพ 19
2.17 โครงสร้างโครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network) 19
2.18 ตัวอย่างการดำเนินการคอนโวลูชัน 21
2.19 ตัวอย่างตัวกรอง 96 แบบ ของ Krizhevsky 22
2.20 ตัวอย่างการดำเนินการพูลลิ่ง 23
2.21 ตัวอย่างโครงสร้างชั้นเชื่อมโยงสมบูรณ์ (Fully Connected Layer) 24
2.22 ตัวอย่างเทคนิคการตรวจจับวัตถุภายในรูปภาพ 25
2.23 ตัวอย่างการทำกรอบล้อมรอบวัตถุจากข้อมูลจริง (Ground Truth) 26

สารบัญรูป (ต่อ)

รูปที่ หน้า

2.24 โครงสร้างการทำงานของการตรวจจับวัตถุ 27
2.25 โครงสร้างการทำงานของ Region Proposal Network 28
2.26 ตัวอย่าง Anchors Boxes 28
2.27 โครงสร้างการทำงานของ Region Proposal Network ในขั้นตอนการเรียนรู้ 29
2.28 ตัวอย่างการทำงานของเทคนิค Non-maximum Suppression 30
2.29 ตัวอย่างการทำงานของ ROI Pooling 30
2.30 ตัวอย่างการทำงานของสถาปัตยกรรม YOLO (You Only Look Once)....31
2.31 โครงสร้างของโครงข่าย YOLOv1....32
2.32 โครงสร้างของโครงข่าย Darknet-19 ใน YOLOv2....33
2.33 โครงสร้างของโครงข่าย Darknet-53 ใน YOLOv3....34
2.34 โครงสร้างทั่วไปของสถาปัตยกรรมการตรวจจับวัตถุ....35
2.35 โครงสร้างโครงข่ายของเทคนิค Cross Stage Partial DenseNet 36
2.36 ตัวอย่างเทคนิค Mosaic Data Augmentation 37
2.37 ตัวอย่างลักษณะของกราฟ Mish เปรียบเทียบกับ Activation Function อื่น ๆ 38
2.38 ตัวอย่างการคำนวณอัตราส่วนระหว่างพื้นที่ทับซ้อนของกรอบล้อมรอบวัตถุ ที่ทำนายได้เทียบกับกรอบล้อมรอบวัตถุขนาดจริง Intersection over Union (IoU)....39
2.39 ตัวอย่างกราฟ Precision-Recall....41
2.40 ตัวอย่างการประมาณค่าความเที่ยงตรงจากค่าการระลึกที่สนใจทั้งหมด 11 จุด....42
2.41 ตัวอย่างการประมาณค่าความเที่ยงตรงจากค่าการระลึกที่สนใจทุกจุดเมื่อค่าความเที่ยงตรงมีการเปลี่ยนแปลง....43
3.1 กรอบแนวคิดงานวิจัย....51
3.2 แผนภาพแสดงขั้นตอนการวิจัยทั้ง 5 ขั้นตอน 52
3.3 ตัวอย่างข้อมูลรูปภาพช่วงเวลากลางวันโดยแบ่งตามระดับความเสี่ยงปานกลางและระดับความเสี่ยงสูง....53
3.4 ตัวอย่างข้อมูลรูปภาพช่วงเวลากลางคืนแบ่งตามระดับความเสี่ยงปานกลางและระดับความเสี่ยงสูง....54

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

3.5 ตัวอย่างการกำหนดขอบเขตของบุคคลภายในรูปภาพ 54
3.6 ภาพรวมการทำงานของระบบตรวจจับบุคคลและระบุระดับความเสี่ยง 58
3.7 กระบวนการระบุระดับความเสี่ยงจากตำแหน่งของบุคคล 59
3.8 แผนผังการรับส่งข้อมูลระหว่างระบบที่นำเสนอและระบบควบคุมเครื่องจักร 60
3.9 การกำหนดเงื่อนไขการส่งสัญญาณออกจากโปรแกรม PLC ไปยังเครื่องจักร 60
4.1 ตัวอย่างรูปภาพชุดทดสอบจำนวน 2 กลุ่ม แบ่งภาพออกเป็น 2 ขนาดภาพ 63
4.2 ตัวอย่างรูปภาพชุดทดสอบจำนวน 4 กลุ่ม แบ่งภาพออกเป็น 2 ขนาดภาพ 64
4.3 กราฟแสดงความเที่ยงตรงเฉลี่ยและความเร็วในการประมวลผลภาพต่อวินาที ระหว่างโมเดล Faster R-CNN และ YOLOv4 แบ่งตามเวลากลางวัน กลางคืน และรวมทั้งหมด 70
4.4 กราฟแสดงความเที่ยงตรงเฉลี่ยและความเร็วในการประมวลผลภาพต่อวินาที ระหว่างโมเดล YOLOv4 ในแต่ละชุดข้อมูล แบ่งตามเวลากลางวัน กลางคืน และรวมทั้งหมด 88
4.5 กราฟแสดงความถูกต้องในการระบุความเสี่ยงบุคคลเมื่อเทียบกับชุดข้อมูลจริง ทั้งหมด ระหว่างโมเดล YOLOv4 ที่ถูกพัฒนาขึ้นในแต่ละชุดข้อมูล 88
4.6 ตัวอย่างภาพผลลัพธ์การตรวจจับบุคคลและระบุระดับความเสี่ยงจาก การทดลอง 90
4.7 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุ ในระบบควบคุมทรัดดัมป์ในขอบเขตพื้นที่ความเสี่ยงสูง 91
4.8 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุ ในระบบควบคุมทรัดดัมป์ในขอบเขตพื้นที่ความเสี่ยงปานกลาง 92
4.9 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุ ในระบบควบคุมทรัดดัมป์ในมุมมองอื่น 92

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

ปัจจุบันเทคโนโลยีอัตโนมัติและหุ่นยนต์ได้เข้ามามีบทบาทสำคัญในกระบวนการผลิตของโรงงานอุตสาหกรรมทั้งในประเทศไทยและต่างประเทศ ซึ่งถือเป็นหัวใจหลักในการสร้างผลผลิตตลอดทั้งกระบวนการเพื่อให้ได้ประสิทธิภาพมากที่สุด การทำงานของเครื่องจักรนั้นย่อมมีส่วนที่ใช้คนควบคุมและตัดสินใจในการสั่งการ โดยเครื่องจักรจะมีการทำงานที่แตกต่างกันไปขึ้นอยู่กับกระบวนการของแต่ละธุรกิจ

โรงงานผลิตอาหารสัตว์โดยทั่วไปจะมีกระบวนการเริ่มต้นจาก การรับวัตถุดิบ จัดเก็บวัตถุดิบ กระบวนการผสมอาหาร อัดเม็ดอาหาร บรรจุอาหาร และส่งมอบอาหารไปยังรถลูกค้า เพื่อเป็นการเพิ่มประสิทธิภาพรวมถึงคุณภาพของการผลิตอาหาร จึงต้องมีเครื่องจักรเข้ามาเกี่ยวข้องตลอดทั้งกระบวนการ ด้วยเหตุนี้ในขณะที่เครื่องจักรกำลังทำงานนั้นพื้นที่โดยรอบจะถือเป็นเขตอันตรายและห้ามมีบุคคลใด ๆ เข้ามายังบริเวณใกล้เคียง เนื่องจากการทำงานของเครื่องจักรอาจก่อให้เกิดอุบัติเหตุแก่พนักงานหรือบุคคลที่เกี่ยวข้องกับเครื่องจักรในบริเวณดังกล่าวได้

ในกระบวนการรับวัตถุดิบนั้น โรงงานทั่วไปจะรับวัตถุดิบจากรถบรรทุกของผู้ขายวัตถุดิบ โดยผู้ขายจะนำรถบรรทุกมายังจุดเทวัตถุดิบและทำการยกท้ายรถบรรทุกเพื่อเทวัตถุดิบลงไปยังบ่อรับวัตถุดิบ หลังจากนั้นพนักงานรับวัตถุดิบจะสั่งงานเครื่องจักรเพื่อลำเลียงวัตถุดิบจากบ่อรับวัตถุดิบไปยังสถานที่จัดเก็บวัตถุดิบต่อไป โดยรถบรรทุกของลูกค้าจะแบ่งออกเป็น 2 ประเภท คือ รถที่สามารถยกท้ายรถบรรทุกได้ด้วยตัวเอง และรถที่ไม่สามารถยกท้ายรถบรรทุกได้ด้วยตัวเอง ในส่วนของรถบรรทุกที่ไม่สามารถยกท้ายได้ด้วยตัวเอง เมื่อผู้ขายนำรถมาจอดยังช่องเทวัตถุดิบเรียบร้อยแล้ว คนขับรถจะต้องลงจากรถและเดินออกไปจากพื้นที่ที่เรียกว่าทรัคดัมป์ ในพื้นที่ส่วนนี้ทางโรงงานจะมีเครื่องจักรสำหรับยกรถบรรทุกทั้งคันให้เอียงขึ้นเพื่อให้วัตถุดิบภายในรถสามารถไหลลงมายังบ่อรับวัตถุดิบได้ เครื่องจักรนี้เรียกว่า ทรัคดัมป์เปอร์ (Truck Dumper) ซึ่งถูกควบคุมด้วยพนักงานภายในห้องควบคุมผ่านระบบควบคุมทรัคดัมป์ (Truck Dumper Control System) โดยระหว่างที่เครื่องจักรกำลังทำงานจะห้ามมิให้บุคคลใด ๆ เข้ามายังพื้นที่ใกล้เคียงแท่นทรัคดัมป์ โดยเฉพาะอย่างยิ่งบนพื้นที่ที่ทรัคดัมป์ถูกยกตัวขึ้น เพราะในขณะที่พื้นถูกยกตัวให้เอียงขึ้นถ้ามีคนยืนอยู่จะทำให้ลื่นไหลตกไปยังบ่อรับวัตถุดิบได้ ซึ่งส่งผลให้บาดเจ็บหรืออาจเกิดอันตรายถึงแก่ชีวิตได้ การป้องกันการเกิดอุบัติเหตุดังกล่าวในโรงงานอุตสาหกรรมทั่วไป ยังใช้คนเพื่อคอยสังเกต

ซึ่งมีโอกาสผิดพลาดสูง อีกทั้งยังไม่สามารถมองเห็นได้ในหลายมุมมองจากภายในห้องควบคุม

การแก้ปัญหาดังกล่าวสามารถใช้กล้องวงจรปิดหรือกล้อง IP Camera (Internet Protocol Camera) ติดตั้งในมุมสูงของพื้นที่อันตราย และใช้พนักงานควบคุมในการสังเกตการณ์ภายในห้องควบคุมได้ แต่เนื่องด้วยระหว่างปฏิบัติงานพนักงานควบคุมอาจมีภาระงานที่ต้องทำจำนวนมากรวมถึงการควบคุมเครื่องจักรในหลาย ๆ ส่วน ซึ่งอาจทำให้พนักงานละเลยหรือพลาดการสังเกตการณ์ในบางช่วงเวลาได้ และนั่นอาจหมายถึงชีวิตคนที่ต้องสูญเสียถ้าเกิดอุบัติเหตุร้ายแรงขึ้น ดังนั้นเพื่อทดแทนข้อจำกัดของมนุษย์ การใช้เทคโนโลยีคอมพิวเตอร์วิทัศน์ (Computer Vision) เพื่อใช้ตรวจจับบุคคลในเขตพื้นที่อันตรายจึงมีความสำคัญมากในการแก้ปัญหานี้ โดยในอดีตได้มีการใช้เทคนิคทางด้านการประมวลผลภาพ (Image Processing) ช่วยในการตรวจจับบุคคลและวัตถุที่เคลื่อนไหว โดยใช้หลักการคำนวณทางคณิตศาสตร์มาวิเคราะห์ค่าสีภายในรูปภาพ รวมถึงมีการหาค่าสหสัมพันธ์กับรูปแบบของวัตถุที่สนใจ แต่ยังพบข้อจำกัดเมื่อภาพที่นำมาวิเคราะห์อยู่ในสภาพแสงและมุมมองที่เปลี่ยนไป ทำให้มีการพัฒนาโดยใช้เทคนิคการสกัดพื้นหลังออกโดยยึดจากวัตถุที่เคลื่อนไหวได้ (Background Subtraction) แต่ก็ยังพบปัญหาจากกรณีที่ภายในภาพมีวัตถุเคลื่อนไหวจำนวนมาก ต่อมามีการประยุกต์ใช้เทคโนโลยีกล้องสามมิติ (3D Depth Camera) เพื่อตรวจจับบุคคลภายในรูปซึ่งมีความเร็วและความแม่นยำสูงแต่ยังมีข้อจำกัดที่ไม่สามารถตรวจจับบุคคลในระยะไกลได้และกล้องประเภทนี้ยังมีต้นทุนสูง

การพัฒนาด้านประสิทธิภาพของไมโครโปรเซสเซอร์หรือ CPU (Central Processing Unit) รวมถึงหน่วยประมวลผลด้านกราฟฟิก 3 มิติหรือ GPU (Graphics Processing Unit) ทำให้คอมพิวเตอร์สามารถประมวลผลได้รวดเร็วยิ่งขึ้น ซึ่งส่งผลโดยตรงกับเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ซึ่งมีการคำนวณค่าน้ำหนักภายในโครงข่ายประสาทเทียมที่ซับซ้อนได้รวดเร็วมากยิ่งขึ้น เทคนิคการเรียนรู้เชิงลึกนั้นนิยมใช้กับงานด้านการประมวลผลภาพ เช่น เทคนิคการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network) โดยเป็นเทคนิคที่ใช้ในการสกัดคุณลักษณะเด่นภายในรูปภาพจากตัวกรอง (Filters) ที่แตกต่างกัน ตัวกรองเหล่านี้ช่วยรองรับต่อการเปลี่ยนตำแหน่งของวัตถุ การเปลี่ยนขนาด รวมถึงการเปลี่ยนแปลงของค่าสีจากแสงที่แตกต่างกัน หลังจากพยายามสกัดคุณลักษณะเด่นจากตัวกรองต่าง ๆ กันแล้ว จะตามมาด้วยขั้นตอนการทำพูลลิ่ง (Pooling) เพื่อเป็นการทำซ้ำเพื่อรวบรวมลักษณะเด่นของวัตถุภายในรูปภาพที่ขนาดต่าง ๆ กัน จนได้ข้อมูลที่สำคัญและนำมาจำแนกได้อย่างชัดเจนในขั้นตอนสุดท้าย

นอกจากการจำแนกวัตถุภายในรูปภาพแล้ว ยังมีเทคนิคการระบุตำแหน่งของวัตถุและตรวจจับวัตถุภายในรูปภาพ เรียกว่า Object Detection ซึ่งมีความจำเป็นอย่างมากกับงานที่ต้องใช้คอมพิวเตอร์ในการมองเห็น เช่น ระบบการมองเห็นของหุ่นยนต์ การตรวจจับสิ่งผิดปกติในสายการผลิต การวิเคราะห์วัตถุของรถยนต์ไร้คนขับ ซึ่งผลจากความก้าวหน้าทางด้านเทคโนโลยี

การเรียนรู้เชิงลึก ทำให้มีการนำโครงข่ายประสาทแบบคอนโวลูชันมาใช้กับงานด้าน Object Detection โดยในยุคแรกจะมีโครงสร้างการทำงานแบ่งออกเป็น 2 ส่วน (Two-stage Detection) โดยส่วนแรกเป็นการเสนอพื้นที่ที่น่าจะมีวัตถุอยู่ (Region Proposal Network) และส่วนที่สองคือส่วนที่ทำหน้าที่รู้จำวัตถุ (Recognition) ซึ่งข้อดีคือสามารถรู้จำวัตถุได้ดีและมีความแม่นยำสูง แต่ใช้เวลาในการประมวลผลค่อนข้างนาน ยกตัวอย่างเช่น โครงข่ายแบบ R-CNN, Fast R-CNN, Faster R-CNN ซึ่งอาจจะไม่เหมาะกับงานที่ต้องทำงานแบบทันที (Real-time)

ต่อมาได้มีการเสนอโครงข่ายแบบ One-stage detection คือสามารถทำนายตำแหน่งและขนาดกรอบล้อมรอบวัตถุรวมถึงความน่าจะเป็นของชนิดวัตถุภายในรูปภาพออกมาพร้อมกันได้ ซึ่งมีความเร็วในการประมวลผลสูงมากแต่ต้องแลกกับความแม่นยำที่ลดลง โดยเฉพาะโครงสร้างแบบ YOLO (You Only Look Once) ในปัจจุบันถือเป็นความก้าวหน้าที่ทันสมัยในงานด้านการตรวจจับวัตถุที่มีความสามารถสูงโดยเฉพาะงานที่ต้องการการทำงานแบบทันที (Real-time)โดยการทำงานพื้นฐานจะแบ่งรูปภาพออกเป็นตาราง S × S ช่อง หลังจากนั้นจะทำนายตำแหน่งและขนาดของวัตถุที่สนใจภายในแต่ละช่อง พร้อมกับทำนายค่าความเชื่อมั่นของประเภทวัตถุที่เราสนใจ (Confidence) โดยถ้าค่าความเชื่อมั่นมากหมายถึงความน่าจะเป็นของวัตถุที่ตรวจจับได้จะเป็นประเภทวัตถุนั้นมาก การทำงานรูปแบบนี้ทำให้โมเดลมีการทำงานที่รวดเร็วมากยิ่งขึ้น จึงเหมาะกับการนำมาใช้กับงานที่มีความต้องการความเร็วในประมวลผลสูง เช่น งานในด้านเทคโนโลยีอัตโนมัติ การมองเห็นของหุ่นยนต์ รถยนต์ไร้คนขับ โดยเฉพาะงานทางด้านความปลอดภัยที่จำเป็นต้องใช้คอมพิวเตอร์ช่วยในการมองเห็นและตรวจจับเพื่อป้องกันอันตรายที่อาจจะเกิดขึ้นได้ทันเวลา

ในงานวิจัยนี้ ผู้วิจัยจึงได้พัฒนาระบบเพื่อป้องกันการเกิดอุบัติเหตุที่อาจเกิดแก่บุคคลในพื้นที่การทำงานของทรัคดัมป์ โดยจะแบ่งการทำงานออกเป็น 2 ส่วน คือ ระบบตรวจจับบุคคล และระบบระบุความเสี่ยง โดยระบบตรวจจับบุคคลจะมีการคัดเลือกโมเดลที่ใช้โครงสร้างการทำงานแบบ Faster R-CNN และ YOLO เพื่อให้ได้ตำแหน่งของบุคคลภายในรูปภาพส่งให้ระบบประเมินความเสี่ยง ซึ่งจะทำงานโดยการหาตำแหน่งของบุคคลเปรียบเทียบกับพื้นที่อันตรายที่ได้ระบุไว้ล่วงหน้า โดยจะมีการระบุพื้นที่ความเสี่ยงปานกลางและความเสี่ยงสูง ซึ่งระบบจะทำงานร่วมกับระบบควบคุมเครื่องจักรโดยตรง เพื่อที่จะสามารถป้องกันอุบัติเหตุได้ทันเวลา โดยมีเงื่อนไขดังนี้ เมื่อระบบสามารถตรวจจับบุคคลภายในพื้นที่ความเสี่ยงปานกลางได้ ระบบจะทำการส่งสัญญาณเตือนภัยไปยังหน้าจอควบคุมในห้องควบคุม เพื่อเฝ้าระวังการเคลื่อนไหวของบุคคลดังกล่าว นอกจากนี้เมื่อระบบสามารถตรวจจับบุคคลภายในพื้นที่ความเสี่ยงสูง ระบบจะทำการส่งสัญญาณไปหยุดการทำงานของเครื่องจักรทันที ในการทดลองเพื่อทดสอบความแม่นยำของระบบนั้น ผู้วิจัยได้ใช้ข้อมูลภาพจากวิดีโอที่ติดตั้งในสถานที่จริงจากหลาย ๆ เหตุการณ์ ทั้งในช่วงเวลากลางวันและ

กลางคืน โดยมีการทดสอบความแม่นยำจากตัวชี้วัดต่าง ๆ รวมถึงความเร็วในการประมวลผล เพื่อเป็นการยืนยันว่าระบบนี้สามารถนำไปใช้ในสถานการณ์จริงได้

1.2 วัตถุประสงค์การวิจัย

1) เพื่อพัฒนาระบบสำหรับป้องกันอุบัติเหตุในเขตพื้นที่อันตรายภายในโรงงานแบบอัตโนมัติเพื่อทดแทนการสังเกตการณ์ของมนุษย์ที่อาจเกิดความผิดพลาดได้

2) เพื่อทดสอบประสิทธิภาพและคัดเลือกสถาปัตยกรรมการตรวจจับวัตถุแบบต่าง ๆ ที่เหมาะสมสำหรับใช้ในงานตรวจจับบุคคลเพื่อป้องกันอันตราย โดยคำนึงถึงความแม่นยำและความเร็วในการประมวลผล

3) เพื่อพัฒนาและทดสอบประสิทธิภาพของโมเดลในงานตรวจจับบุคคลจากสถาปัตยกรรมการตรวจจับวัตถุที่เลือกจากข้อที่ 2

4) เพื่อพัฒนาระบบการประเมินความเสี่ยงจากพื้นที่อันตรายในระดับต่าง ๆ จากการระบุตำแหน่งของวัตถุที่ได้จากโมเดลการตรวจจับบุคคลที่พัฒนาขึ้นจากข้อที่ 3 เปรียบเทียบกับพื้นที่อันตรายในระดับต่าง ๆ ที่ได้กำหนดไว้แล้ว

1.3 ขอบเขตของการวิจัย

งานวิจัยได้นำความรู้จากเทคนิคการเรียนรู้เชิงลึกมาประยุกต์เพื่อพัฒนาระบบสำหรับตรวจจับบุคคลและประเมินความเสี่ยงในพื้นที่อันตราย เพื่อใช้ในการแจ้งเตือนและป้องกันอุบัติเหตุโดยทำงานร่วมกับระบบควบคุมเครื่องจักรโดยตรง ในส่วนของการดำเนินงานได้มีการกำหนดขอบเขตของการวิจัยไว้ดังต่อไปนี้

1) ชุดข้อมูลที่ใช้ในงานวิจัย ใช้ข้อมูลรูปภาพที่ถูกตัดออกมาจากหลาย ๆ เหตุการณ์ในวิดีโอคลิปจากกล้องวงจรปิดที่ถูกติดตั้งไว้ด้านหน้าของช่องทางเทวัตถุดิบบริเวณทรัคดัมป์ในสถานการณ์จริงภายในโรงงาน โดยแบ่งออกเป็นช่วงเวลากลางวันและกลางคืน

2) การศึกษาเปรียบเทียบประสิทธิภาพระบบการตรวจจับบุคคลด้วยโมเดลใช้โครงสร้างสองแบบคือ Faster R-CNN และ YOLOv4 โดยการเปรียบเทียบประสิทธิภาพด้วยมาตรวัด ค่าความเที่ยงตรง ค่าการระลึก ค่าประสิทธิภาพโดยรวม ค่าความเที่ยงตรงเฉลี่ย และความเร็วในการประมวลผลภาพต่อวินาที

3) การศึกษาเปรียบเทียบค่าความแม่นยำจากการระบุระดับความเสี่ยงจากตำแหน่งของบุคคลที่ได้จากระบบตรวจจับบุคคลและขอบเขตพื้นที่อันตรายที่กำหนดไว้ใช้วิธีเปรียบเทียบกับข้อมูลจริง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่ได้รับจากการศึกษาและพัฒนางานวิจัยนี้ ได้แก่

1) ได้ระบบสำหรับป้องกันอุบัติเหตุภายในโรงงานที่ประยุกต์ใช้เทคนิคต่าง ๆ ทางด้านปัญญาประดิษฐ์มาใช้ทดแทนข้อจำกัดจากการสังเกตการณ์ด้วยมนุษย์ที่อาจเกิดข้อผิดพลาดได้

2) ได้ประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกกับงานด้านคอมพิวเตอร์วิทัศน์ในการตรวจจับบุคคลโดยทำการทดสอบในสถานการณ์จริง

3) ได้พัฒนาระบบต่อยอดจากผลลัพธ์ที่ได้จากเทคนิคการตรวจจับวัตถุโดยใช้ตำแหน่งมาวิเคราะห์ความเสี่ยงเพื่อใช้กับงานทางด้านความปลอดภัยได้อย่างมีประสิทธิภาพและใช้งานได้จริง

4) ได้ระบบการเตือนภัยและป้องกันอันตรายจากความสูญเสียต่าง ๆ จากการใช้งานเครื่องจักรที่มีส่วนปฏิสัมพันธ์กับมนุษย์โดยทำงานร่วมกับระบบควบคุมด้วยเทคโนโลยีอัตโนมัติ

5) สามารถนำระบบเตือนภัยไปใช้งานกับระบบที่ใช้เครื่องจักรหรือไม่ได้ใช้เครื่องจักรภายในเขตพื้นที่อันตรายได้ เช่น พื้นที่กำลังก่อสร้าง พื้นที่อับอากาศ เป็นต้น

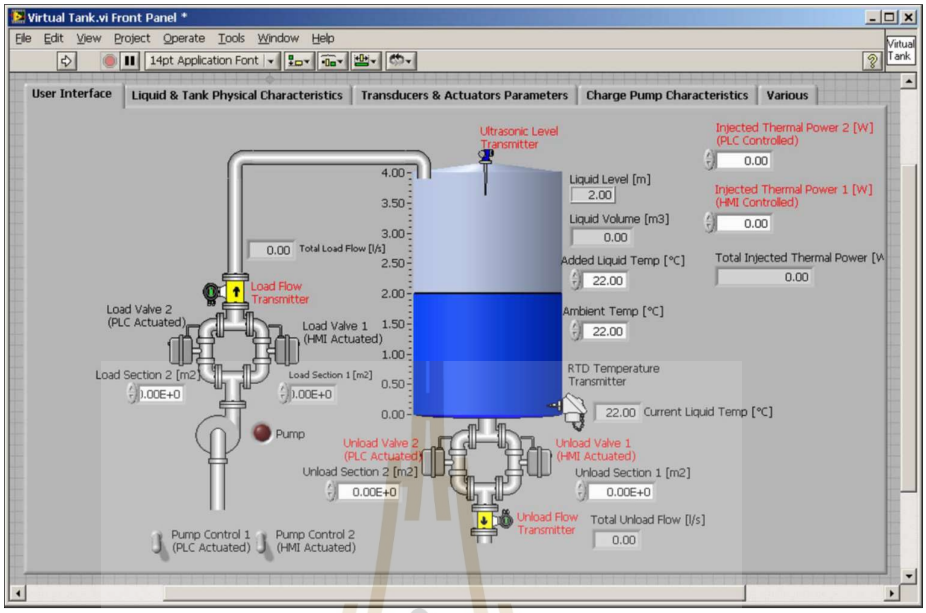
บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้เป็นการกล่าวถึง การศึกษาในหลักการ ทฤษฎีพื้นฐาน และงานวิจัยที่เกี่ยวข้อง สำหรับนำมาประยุกต์ใช้เพื่อพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์โดยใช้การเรียนรู้เชิงลึก ซึ่งประกอบด้วยรายละเอียดเกี่ยวกับระบบควบคุมการทำงานทรัคดัมป์ เทคนิคการเรียนรู้เชิงลึก เทคนิคการจำแนกรูปภาพด้วยโครงข่ายประสาทเทียมคอนโวลูชัน เทคนิคการตรวจจับวัตถุ งานวิจัยที่เกี่ยวข้อง ตามลำดับ

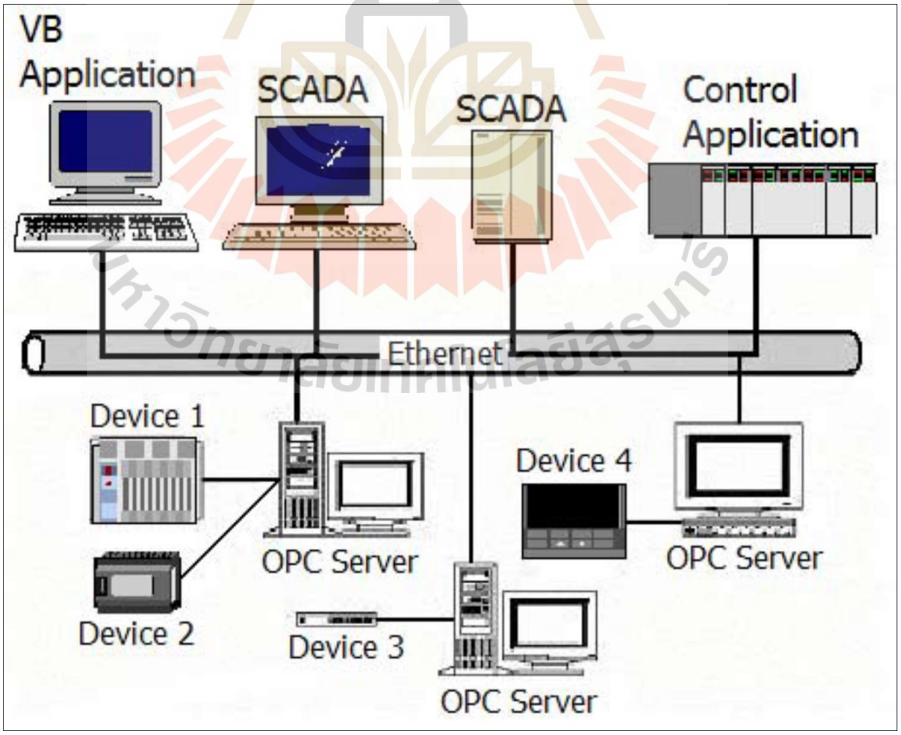
2.1 ระบบควบคุมการทำงานของทรัคดัมป์

2.1.1 หน้าจอระบบควบคุมเครื่องจักร (Human Machine Interface)

หน้าจอระบบควบคุมเครื่องจักร หรือ Human Machine Interface คือส่วนรับคำสั่งจากมนุษย์หรือผู้ควบคุมเพื่อนำสัญญาณไปสั่งงานเครื่องจักรให้ทำงาน รวมถึงยังเป็นส่วนแสดงผลข้อมูลที่ได้รับมาจากเซนเซอร์ต่าง ๆ ที่จำเป็นต่อการควบคุมเครื่องจักรให้ทำงานได้อย่างมีประสิทธิภาพหรือใช้เพื่อสังเกตการณ์ความผิดปกติที่อาจเกิดขึ้นได้จากการทำงานของเครื่องจักร โดยที่หน้าจอควบคุมระบบจะเป็นการแสดงผลภาพกราฟิกบนคอมพิวเตอร์ที่นิยมสร้างจากแผนผังภาพรวมของเครื่องจักรที่ผู้ควบคุมต้องควบคุมทั้งหมด เพื่อให้ง่ายต่อความเข้าใจและใช้เวลาในการเรียนรู้ได้อย่างรวดเร็ว ดังตัวอย่างในรูปที่ 2.1 หน้าจอควบคุมระบบโดยทั่วไปจะมีปุ่มเพื่อให้ผู้ควบคุมสามารถสั่งงานได้โดยง่ายและสะดวกต่อการใช้งาน และมีช่องสำหรับกรอกข้อมูลเพื่อใช้ปรับแต่งการทำงานของเครื่องจักร การสั่งงานเพื่อให้เครื่องจักรทำงานนั้นจึงจำเป็นต้องมีระบบสำหรับรับส่งข้อมูลระหว่างอุปกรณ์ภายในเครือข่าย โดยระบบควบคุมหลักที่โรงงานอุตสาหกรรมทั่วไปส่วนใหญ่นิยมใช้นั้นเรียกว่าระบบ SCADA (Supervisory Control and Data Acquisition) (Gaushell & Darlington, 1987) ซึ่งจะเป็นส่วนควบคุมการทำงานภาพรวมทั้งหมดของระบบ ดังรูปที่ 2.2 นอกจากนี้ยังมีระบบซอฟต์แวร์ที่สามารถออกแบบและสร้างหน้าจอควบคุมการทำงานเป็นภาพกราฟิกได้



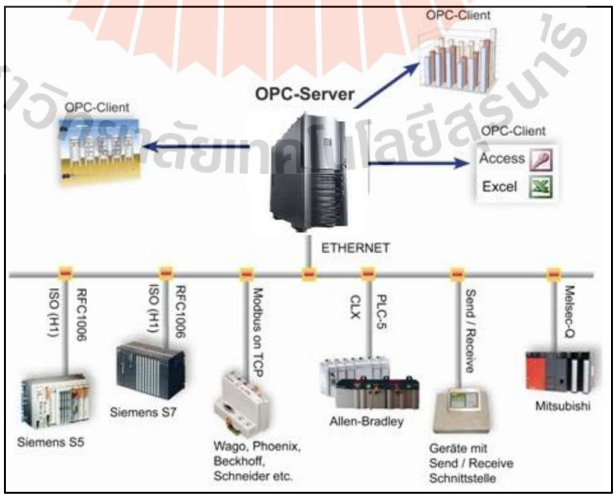
รูปที่ 2.1 ตัวอย่างหน้าจอระบบควบคุมเครื่องจักร (Adamo et al., 2007)



รูปที่ 2.2 โครงสร้างการทำงานของระบบ SCADA (Hernandez et al., 2007)

2.1.2 ระบบรับส่งข้อมูลระหว่างอุปกรณ์

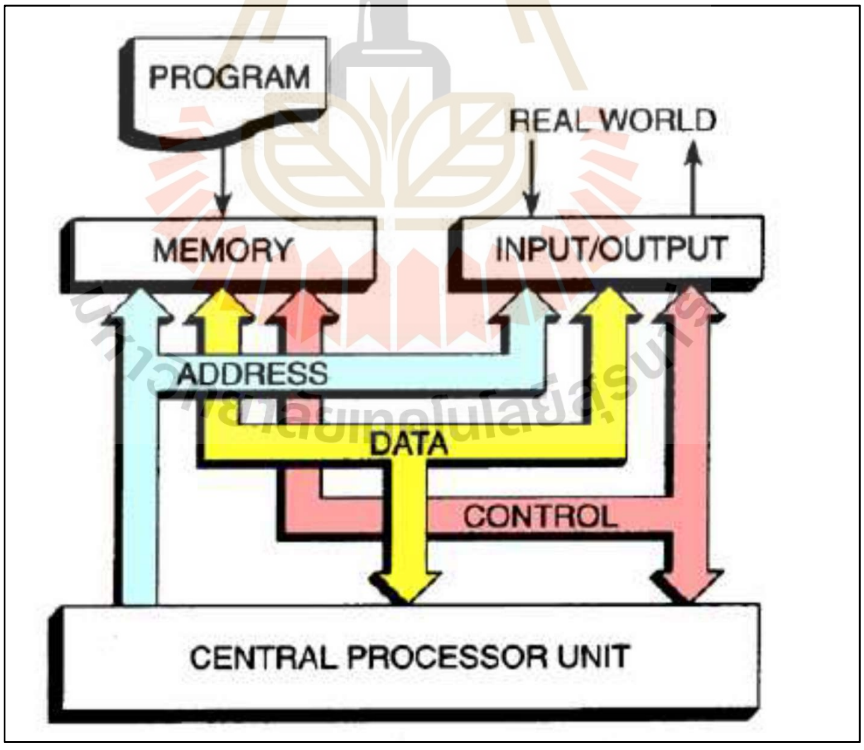
ในการรับส่งข้อมูลระหว่างหน้าจอระบบควบคุมเพื่อส่งสัญญาณนำออกไปสั่งงานเครื่องจักรนั้น จำเป็นต้องมีการส่งข้อมูลไปยังหน่วยความจำภายในอุปกรณ์พีแอลซี (Programmable Logic Controller : PLC) โดยที่อุปกรณ์ในแต่ละรุ่นจะมีโปรแกรมสำหรับขับอุปกรณ์ให้สามารถทำงานได้ (Driver) และเพื่อใช้เชื่อมต่อข้อมูลกับหน่วยความจำ ซึ่งมีความยุ่งยากในการเปลี่ยนแปลงในภายหลังและไม่สะดวกต่อการใช้งานอุปกรณ์ที่แตกต่างกัน โรงงานอุตสาหกรรมโดยทั่วไปจึงนิยมใช้ซอฟต์แวร์ที่ทำหน้าที่เป็นตัวกลางในการรับส่งข้อมูลหรือเรียกว่า OPC (Open Platform Communications) ซึ่งมีการพัฒนามาจากมาตรฐานดั้งเดิมคือ OLE for Process Control (Object Linking and Embedding for Process Control) ที่ถูกจำกัดไว้ภายใต้ระบบปฏิบัติการวินโดวส์ (Windows) เท่านั้น โดยการทำงานของมาตรฐานนี้ประกอบด้วยซอฟต์แวร์ฝั่งลูกข่าย (OPC Client) และซอฟต์แวร์ฝั่งแม่ข่าย (OPC Server) ซึ่งเป็นซอฟต์แวร์สำหรับรับส่งข้อมูลระหว่างคอมพิวเตอร์และอุปกรณ์ต่าง ๆ ภายในเครือข่าย โดยภายในซอฟต์แวร์ได้มีการรวบรวมโปรแกรมสำหรับขับอุปกรณ์ที่ใช้งานอย่างแพร่หลายไว้แล้ว ในการรับส่งข้อมูลจะมีการกำหนดตัวแปรข้อมูล หรือเรียกว่า Tag ซึ่งจะถูกกำหนดชื่อและที่อยู่ของหน่วยความจำภายใน PLC ไว้ การทำงานใช้ลักษณะวนซ้ำเพื่อรับส่งข้อมูลระหว่าง Tag ภายใน OPC Server กับหน่วยความจำภายใน PLC โดยตรง โดยโครงสร้างการทำงานของระบบควบคุมนั้น ส่วนที่รับคำสั่งจากผู้ควบคุมหรือหน้าจอควบคุมจะทำหน้าที่เป็น OPC Client และรับคำสั่งจากหน้าจอเพื่อส่งข้อมูลไปยัง Tag ที่ได้กำหนดไว้ เพื่อให้ OPC Server ส่งข้อมูลไปยังหน่วยความจำภายใน PLC ต่อไป ดังรูปที่ 2.3



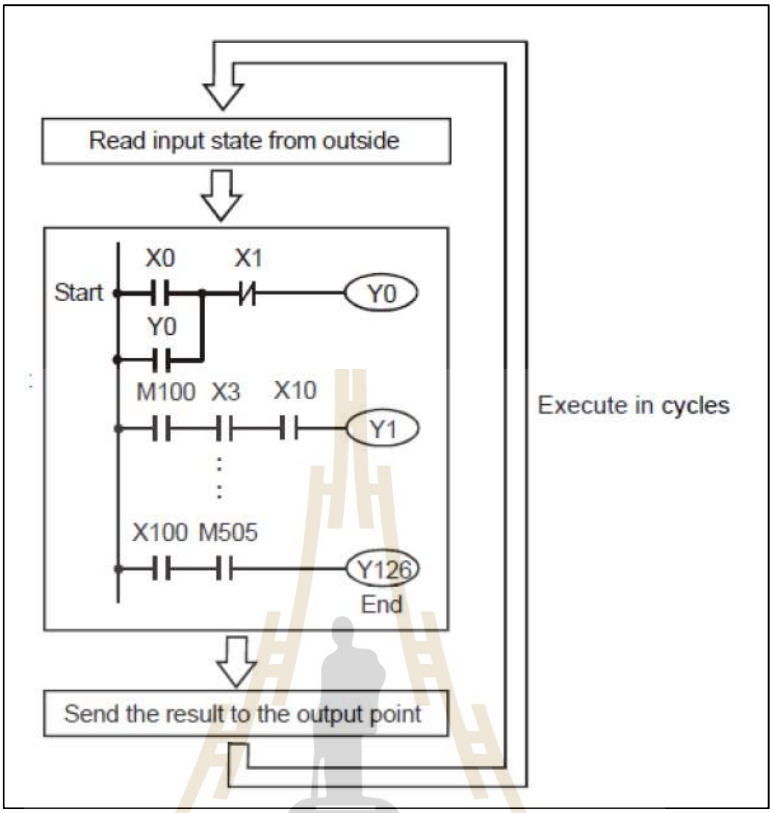
รูปที่ 2.3 โครงสร้างการรับส่งข้อมูลผ่าน OPC Server (Nicola et al., 2018)

2.1.3 การทำงานของโปรแกรมพีแอลซี (Programmable Logic Controller)

พีแอลซี (PLC) หรือ โปรแกรมเมเบิลลอจิกคอลโทรลเลอร์ (Programmable Logic Controller) เป็นอุปกรณ์ไฟฟ้าสำหรับควบคุมการทำงานของเครื่องจักรที่นิยมใช้ภายในโรงงานอุตสาหกรรมทั่วไปในปัจจุบัน จากเดิมงานควบคุมเครื่องจักรนั้นจะควบคุมผ่านการเดินสายวงจรไฟฟ้า ซึ่งยากต่อการเปลี่ยนแปลงการทำงานในภายหลังและยังมีค่าใช้จ่ายสูง จึงได้มีการเปลี่ยนมาใช้ PLC เป็นตัวกลางในการควบคุมการทำงานของเครื่องจักรทั้งหมด เพราะสามารถเขียนโปรแกรมควบคุมลงไปยังหน่วยความจำของ PLC ได้และง่ายต่อการเปลี่ยนแปลงโปรแกรมการทำงานในภายหลัง โดยภายในอุปกรณ์จะมีหน่วยประมวลผลภายในตัวเอง (Central Processing Unit) และมีหน่วยรับสัญญาณนำเข้าจากโปรแกรมภายนอกหรือสวิตช์สั่งงาน รวมทั้งรับค่าจากอุปกรณ์เซนเซอร์ต่าง ๆ เข้ามาประมวลผลภายในโปรแกรมที่ได้เขียนไว้ (Ladder Logic) แสดงสถาปัตยกรรมการทำงานได้ดังรูปที่ 2.4 โดยโปรแกรมจะมีการทำงานในลักษณะวนซ้ำเพื่อนำค่าสัญญาณนำเข้า (Input) จากเซนเซอร์ต่าง ๆ รวมถึงค่าที่มีอยู่ในหน่วยความจำมาประมวลผล หลังจากนั้นจะส่งสัญญาณส่งออก (Output) ออกไปให้เครื่องจักรทำงาน ดังรูปที่ 2.4 และ 2.5



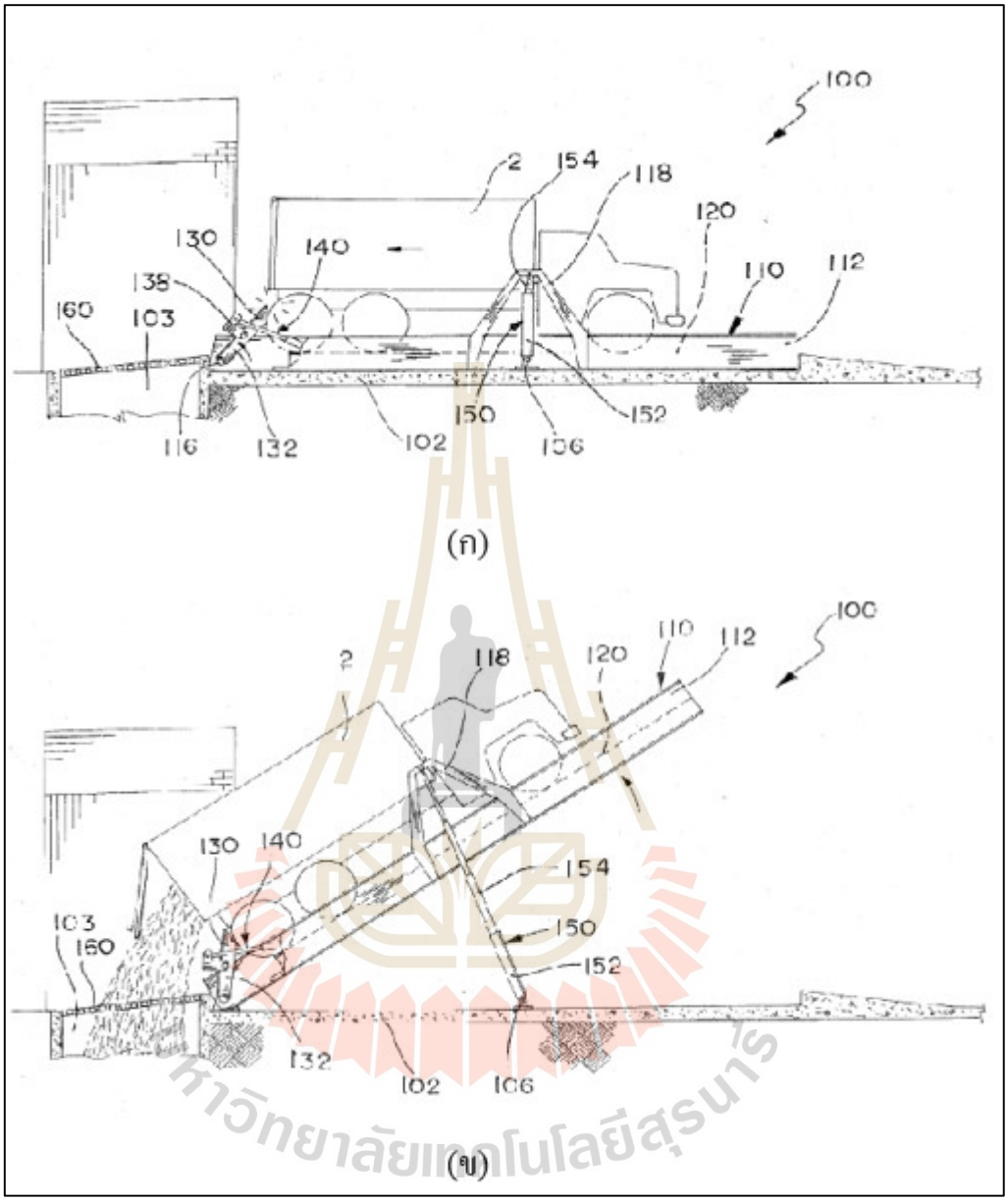
รูปที่ 2.4 สถาปัตยกรรมการทำงานภายใน PLC (Htay & Mon, 2014)



รูปที่ 2.5 การทำงานของโปรแกรม Ladder Logic ภายใน PLC (Htay & Mon, 2014)

2.1.4 ระบบควบคุมทรัดดัมป์ (Truck Dumper Control System)

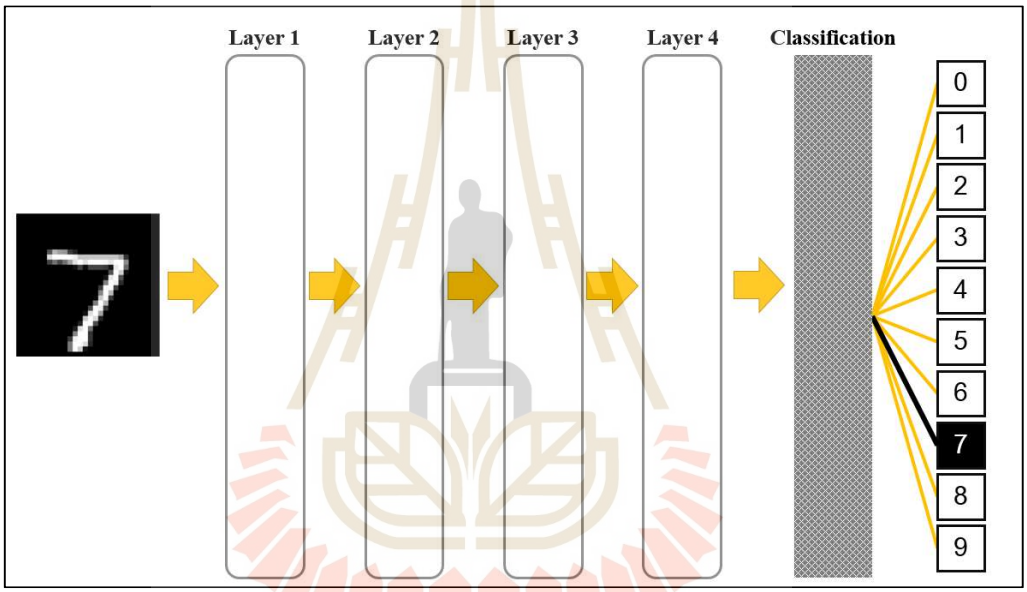
ระบบควบคุมทรัดดัมป์ภายในโรงงานทั่วไปจำเป็นต้องใช้พนักงานในการควบคุมการยกตัวของแท่นทรัดดัมป์เพื่อทำให้รถบรรทุกยกตัวเอียงขึ้นเพื่อให้วัตถุดิบไหลลงสู่บ่อรับวัตถุดิบได้ รวมถึงการลดระดับลงเมื่อเทวัตถุดิบเสร็จเรียบร้อยแล้ว โดยการทำงานจะรับคำสั่งจากหน้าจอระบบควบคุม เมื่อผู้ควบคุมกดปุ่มสั่งงานบนหน้าจอ ระบบจะส่งสัญญาณมายังหน่วยความจำภายใน PLC เพื่อส่งสัญญาณออกไปสั่งงานให้ไฮดรอลิคดัมป์ยกขึ้นหรือลงตามเงื่อนไขที่ได้เขียนโปรแกรม Ladder Logic ในหน่วยความจำของ PLC โดยการทำงานเบื้องต้นของระบบทรัดดัมป์จะแสดงดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างการทำงานของแท่นทรัคดัมป์ (Truck Dumper) (Hobbs, 2015)
(ก) ทรัคดัมป์ยังไม่ทำงาน (ข) ทรัคดัมป์กำลังทำงาน

2.2 เทคนิคการเรียนรู้เชิงลึก

เทคนิคการเรียนรู้เชิงลึกเป็นการพัฒาต่อยอดจากอัลกอริทึมโครงข่ายประสาทเทียมที่เป็นการจำลองการทำงานของสมองในสิ่งมีชีวิตและเป็นส่วนหนึ่งของเทคนิคการเรียนรู้ของเครื่อง โดยพื้นฐานแนวคิดของการเรียนรู้เชิงลึกนั้นเป็นการสร้างตัวแทนการเรียนรู้ระหว่างข้อมูลนำเข้าและข้อมูลส่งออกโดยใช้อัลกอริทึมโครงข่ายประสาทเทียมเช่นเดียวกันกับการเรียนรู้ของเครื่อง แต่จะมีการประมวลผลหลาย ๆ ชั้น และมีกระบวนการคำนวณที่ซับซ้อนกว่า ซึ่งประกอบด้วยการแปลงข้อมูลทั้งแบบเชิงเส้นและไม่เป็นเชิงเส้น

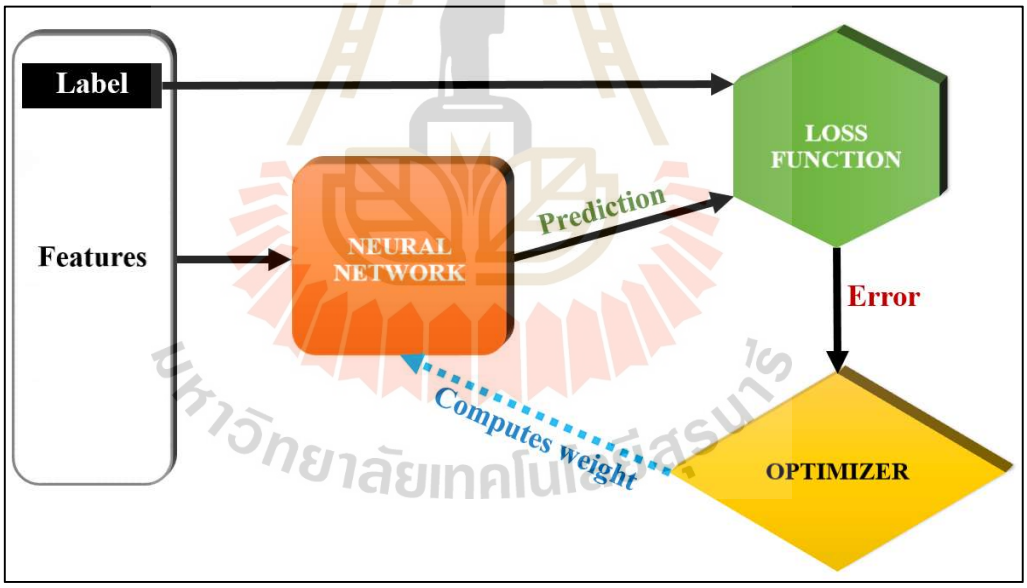


รูปที่ 2.7 แนวคิดเบื้องต้นสำหรับการใช้เทคนิคการเรียนรู้เชิงลึกกับข้อมูลภาพลายมือเขียนตัวเลข

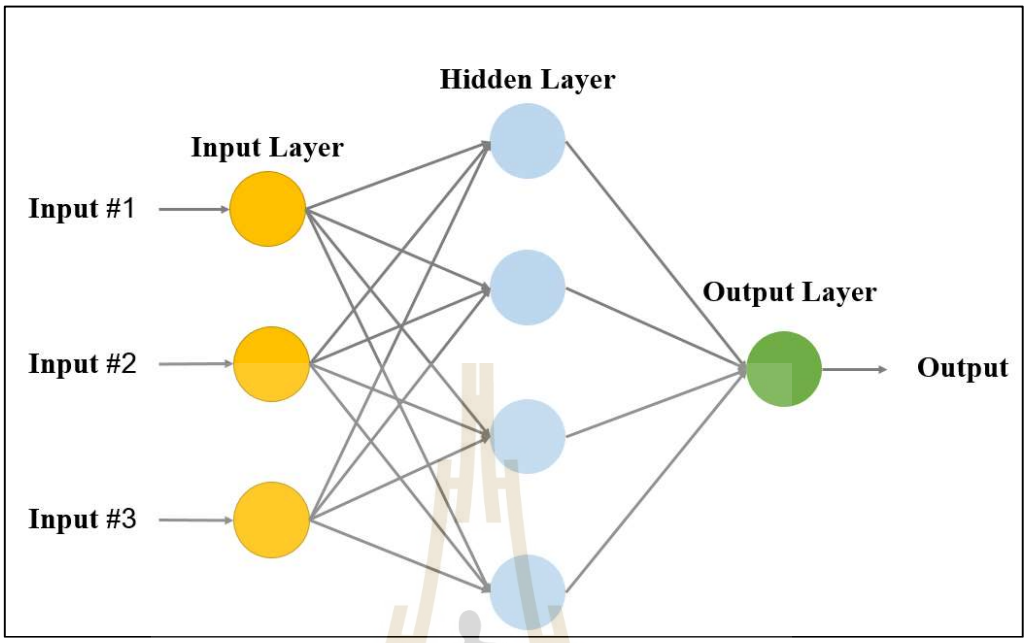
จากรูปที่ 2.7 แสดงถึงแนวคิดเบื้องต้นในการใช้เทคนิคการเรียนรู้เชิงลึกในการจำแนกรูปภาพตัวเลขที่เขียนด้วยลายมือ เริ่มต้นจากโครงข่ายประสาทเทียมรับข้อมูลรูปภาพนำเข้าจากทางด้านซ้ายมือเรียกว่าชั้นอินพุต (Input Layer) ผ่านการประมวลผลภายในแต่ละชั้นที่ต่อเนื่องกันตามลำดับ เรียกชั้นภายในโครงข่ายที่อยู่ตรงกลางระหว่างชั้นข้อมูลนำเข้าและข้อมูลนำออกว่าชั้นซ่อนเร้น (Hidden Layer) โดยการทำงานเบื้องต้นของชั้นที่ต่อเนื่องกันนี้จะเป็นการดึงคุณลักษณะที่มีความหมายสำคัญออกมาเพื่อรวมคุณลักษณะสำคัญที่สกัดได้มารวมกันและส่งไปยังชั้นสุดท้ายหรือที่เรียกว่าชั้นเอาต์พุต (Output Layer) เพื่อทำการจำแนกประเภทที่ใกล้เคียงกับคำตอบมากที่สุด

2.2.1 การเรียนรู้เชิงลึก (Deep Learning)

การทำงานของการเรียนรู้เชิงลึกเริ่มต้นจากการนำข้อมูลนำเข้าเข้าสู่โครงข่ายประสาทเทียม (Neural Network) โดยภายในข้อมูลที่นำเข้าจะมีคุณลักษณะหลายชนิด (Features) โดยมีแนวคิดให้โมเดลพยายามเรียนรู้คุณลักษณะสำคัญจากข้อมูลนำเข้าและทำการทำนายคำตอบโดยเปรียบเทียบกับคำตอบจริงที่มีการระบุคำตอบไว้ (Label) ผ่านฟังก์ชันที่ใช้สำหรับการคำนวณค่าความคลาดเคลื่อน (Error) จากการเปรียบเทียบกันระหว่างคำตอบจริงและคำตอบที่ทำนายได้ (Loss Function หรือ Cost Function) ซึ่งฟังก์ชันในการคำนวณค่าความคลาดเคลื่อนมีหลายประเภทขึ้นอยู่กับวัตถุประสงค์ของงานที่ต้องการคำตอบที่ต่างกันรวมถึงลักษณะของข้อมูลที่ใช้ในการเรียนรู้ หลังจากที่ได้ค่าความคลาดเคลื่อน จะมีการเพิ่มประสิทธิภาพของโมเดลระหว่างการเรียนรู้ในแต่ละรอบโดยพยายามทำให้ค่าความคลาดเคลื่อนที่ได้นั้นลดน้อยลงเรื่อย ๆ จนได้ค่าที่ต่ำที่สุดเหมาะสมที่สุด หรือครบตามจำนวนรอบของการเรียนรู้ที่ได้กำหนดไว้ เรียกกระบวนการนี้ว่า Optimizer ดังแสดงตัวอย่างกระบวนการเรียนรู้ในรูปที่ 2.8

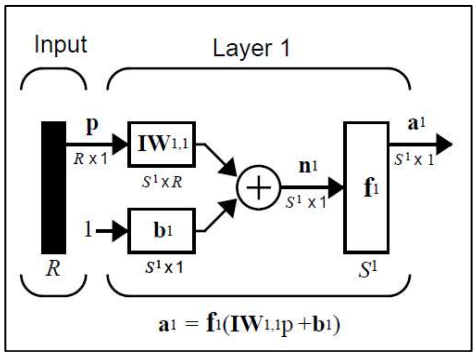


รูปที่ 2.8 กระบวนการเรียนรู้ภายในโมเดลการเรียนรู้เชิงลึก



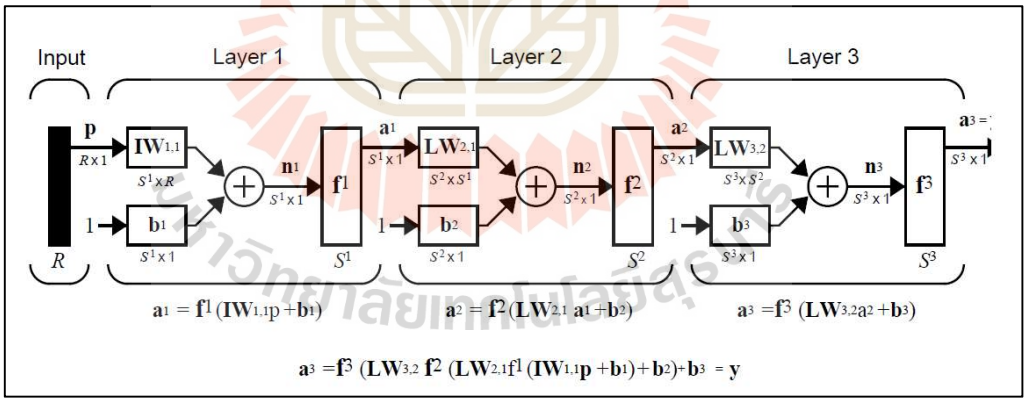
รูปที่ 2.9 สถาปัตยกรรมพื้นฐานของโครงข่ายประสาทเทียม (Neural Network)

จากรูปที่ 2.9 แสดงถึงสถาปัตยกรรมพื้นฐานของโครงข่ายประสาทเทียมซึ่งประกอบไปด้วยชั้น Input Layer, Hidden Layer และ Output Layer ในแต่ละชั้นจะประกอบไปด้วยเซลล์ประสาทหรือที่เรียกว่านิวรอน การเรียนรู้เชิงลึกนั้นจะประกอบไปด้วย Hidden Layer จำนวนหลายชั้น การกำหนดจำนวนนิวรอนในแต่ละชั้นและจำนวนชั้น Hidden Layer ที่เชื่อมต่อกันอยู่หลาย ๆ ชั้นนั้นมีความสำคัญต่อผลความแม่นยำในการทำนาย ยกตัวอย่างการใช้นิวรอนและจำนวนชั้นที่น้อยเกินไปอาจจะทำให้โมเดลไม่สามารถจดจำรูปแบบของข้อมูลได้เพียงพอซึ่งอาจส่งผลให้โมเดลมีความแม่นยำที่ต่ำหรือเรียกว่าโมเดลเกิดการ Underfitting ในทางตรงกันข้ามถ้าใช้นิวรอนและจำนวน Hidden Layer มากจนเกินไปอาจจะทำให้โมเดลเรียนรู้และจดจำคุณลักษณะของข้อมูลที่จำเพาะกับข้อมูลฝึกมากจนเกินไปหรือเรียกว่าโมเดลเกิดการ Overfitting ซึ่งจะทำให้โมเดลที่ได้ไม่มีประสิทธิภาพกับข้อมูลใหม่ ๆ นอกจากนี้ยังส่งผลให้การประมวลผลใช้เวลานานขึ้น ดังนั้นการพยายามเลือกจำนวนนิวรอนและจำนวนชั้น Hidden Layer ให้สมดุลจึงมีความสำคัญต่อการสร้างโมเดลการเรียนรู้ด้วยโครงข่ายประสาทเทียม



รูปที่ 2.10 โครงสร้างเพอร์เซปตรอน (Perceptron) (Demuth et al., 1992)

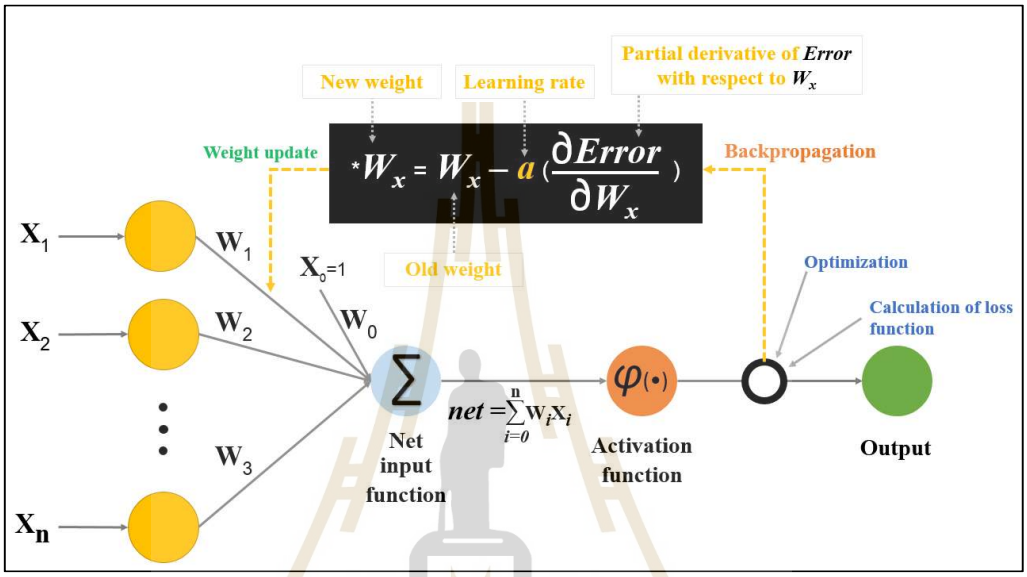
เพอร์เซปตรอน (Perceptron) คือหน่วยย่อยที่สุดของโครงข่ายประสาทเทียม โดยจะมีชั้น Hidden Layer เพียง 1 ชั้น โดยแต่ละ Input ข้อมูลจะถูกนำเข้ามาที่นิวรอน ผ่านการคูณด้วยค่าน้ำหนัก (Weight) และบวกด้วยค่าความลำเอียง (Bias) ซึ่งในขั้นตอนสุดท้ายจะถูกแปลงค่าด้วยฟังก์ชันกระตุ้น (Activation Function) เพื่อให้ได้ค่าผลลัพธ์ส่งออกไปในรูปแบบของข้อมูลที่ต้องการไปยัง Output Layer ดังแสดงในรูปที่ 2.10



รูปที่ 2.11 โครงสร้างเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron) (Demuth et al., 1992)

เพอร์เซ็ปตรอนหลายชั้น (Multi-layer Perceptron) คือ โครงสร้างประสาทเทียมแบบมี Hidden Layer หลายชั้น มีการทำงานลักษณะเดียวกันกับเพอร์เซ็ปตรอน ซึ่งข้อมูลนำออกของชั้นแรก จะเป็นข้อมูลนำเข้าของชั้นถัดไป ทำซ้ำแบบเดียวกันจนกว่าจะส่งผลลัพธ์ไปยัง Output Layer เพื่อทำนายคำตอบ ดังแสดงในรูปที่ 2.11 หลังจากได้ค่าที่ถูกทำนายด้วยโมเดลแล้ว สามารถ

นำค่าที่ได้มาเปรียบเทียบกับค่าจริง (Label) เพื่อหาค่าความคลาดเคลื่อน (Error) กระบวนการทั้งหมดนี้เรียกว่าโครงข่ายประสาทแบบป้อนไปข้างหน้า (Feed Forward) ซึ่งในการเรียนรู้เชิงลึกในแต่ละรอบจะมีการปรับค่าน้ำหนักและค่าความลำเอียงโดยพยายามลดค่าความคลาดเคลื่อนจากโครงข่ายให้มีค่าออกมาน้อยที่สุดโดยใช้เทคนิคการแพร่ย้อนกลับ (Backpropagation)



รูปที่ 2.12 รูปแบบการปรับค่าน้ำหนักจากเทคนิคการแพร่ย้อนกลับ (Backpropagation)

เทคนิคการแพร่ย้อนกลับ หรือ Backpropagation นั้นเป็นเทคนิคการคำนวณหาค่าความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนที่ได้จาก Loss Function จากโครงข่ายประสาทแบบป้อนไปข้างหน้าเปรียบเทียบกับค่าน้ำหนักที่นำเข้ามาในแต่ละชั้น โดยใช้เทคนิคการหาค่าอนุพันธ์ย่อย (Partial Differential) และกฎลูกโซ่ (Chain Rule) ในการคำนวณย้อนกลับไปยังค่าน้ำหนักของทุก ๆ นิวรอนที่มีผลต่อค่าที่ส่งออกไป (Partial derivative of Error with respect to W_X) เมื่อได้ค่าความสัมพันธ์ของแต่ละ W_X ในแต่ละนิวรอนแล้ว กระบวนการเรียนรู้จะได้ค่าความชันของกราฟออกมาด้วย จึงทำให้สามารถทราบว่าจะเลื่อนไปยังทิศใดเพื่อให้ได้ค่า Error ต่ำที่สุด กล่าวคือจุดที่ต่ำที่สุดของกราฟ กระบวนการนี้เรียกว่าการทำ Optimization ดังแสดงในรูปที่ 2.12 โดยอัลกอริทึมที่นิยมใช้คือ Gradient Descent โดยที่หลังจากคำนวณย้อนกลับด้วย Backpropagation ครบแล้วจะถูกนำไปคูณด้วยค่าคงที่สำหรับการเรียนรู้หรือ Learning Rate ที่ถูกกำหนดไว้ เพื่อปรับค่าน้ำหนักใหม่ ซึ่งหลังจากทำงานครบรอบแล้วจะกลับไปทำงานด้วยโครงข่ายประสาทแบบป้อนไปข้างหน้า

ใหม่อีกครั้งเพื่อคำนวณหาค่าความคลาดเคลื่อนใหม่ และจะทำซ้ำกระบวนการเดิมไปเรื่อย ๆ จนครบจำนวนรอบการเรียนรู้ที่ได้กำหนดไว้ หรือ ได้ค่าความคลาดเคลื่อนที่ต่ำและเหมาะสมที่สุด

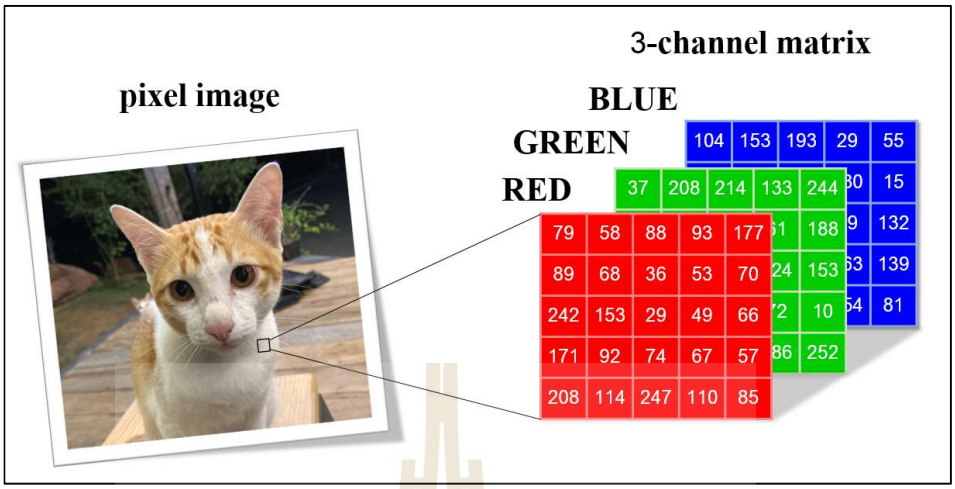
2.2.2 โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network)

โครงข่ายประสาทคอนโวลูชันเป็นเทคนิคการเรียนรู้เชิงลึกประเภทหนึ่งที่นิยมใช้กับงานทางด้านการประมวลผลภาพในงานด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) ซึ่งมีความคาดหวังให้คอมพิวเตอร์สามารถมองเห็นหรือจำแนกวัตถุจากภาพได้ โดยมีรายละเอียดพื้นฐานดังหัวข้อต่อไปนี้

1) พื้นฐานเกี่ยวกับรูปภาพดิจิทัล (Digital Image Fundamentals)

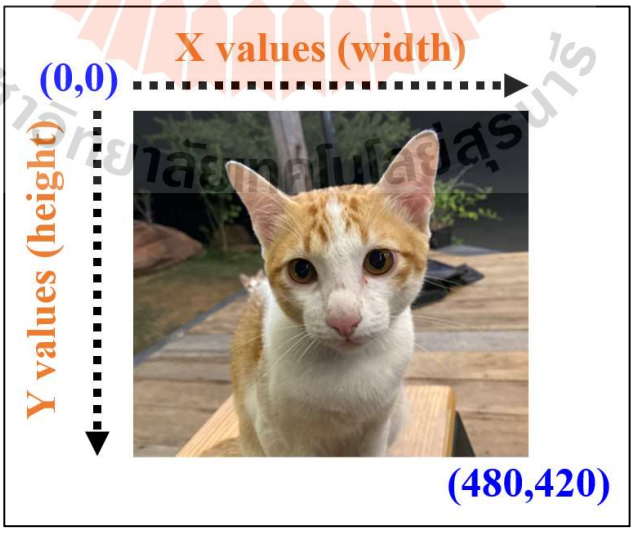
ข้อมูลรูปภาพดิจิทัลถูกเก็บค่าของรูปภาพเป็นจุดสีของพิกเซลในรูปแบบของเมทริกซ์ จากตัวอย่างภาพที่มีขนาดความกว้าง 1,920 พิกเซล และความสูง 1,080 พิกเซล ในกรณีเป็นภาพแบบโหมดขาวดำ (Grayscale) จะมีจำนวนพิกเซลเท่ากับ 2,073,600 พิกเซล โดยจะเก็บค่าเป็นเลขจำนวนเต็มตั้งแต่ 0 ถึง 255 โดยที่ 0 คือสีดำ และ 255 คือสีขาว ซึ่งหมายความว่าในระหว่างค่า 0 ถึง 255 จะเป็นการไล่ระดับสีจากดำไปขาว นั่นก็คือเป็นช่วงสีของสีเทานั่นเอง ในส่วนของภาพโหมดสีแบบ RGB ใน 1 พิกเซลจะถูกแทนด้วยค่า 3 ค่า คือ ค่าสีแดง ค่าสีเขียว และค่าสีน้ำเงิน และจะแบ่งแต่ละสีเป็นแชนแนลของสีนั้น ๆ (Channel) โดยในแต่ละสีจะมีค่าตั้งแต่ 0 ถึง 255 ซึ่งหมายความว่าใน 1 พิกเซลของแต่ละแชนแนลนั้น ๆ มีค่าสีอยู่มากน้อยเพียงใด ดังแสดงตัวอย่างรูปภาพแบบ Grayscale และ RGB ในรูปที่ 2.13

รูปที่ 2.13 ตัวอย่างข้อมูลรูปภาพดิจิทัล แบบ Grayscale (ซ้าย) และ RGB (ขวา)



รูปที่ 2.14 ตัวอย่างการแทนค่าในรูปภาพโหมด RGB

จากรูปที่ 2.14 แสดงถึงการแทนค่าสีในรูปภาพในโหมด RGB จากตัวอย่างพื้นที่ขนาด 5 × 5 พิกเซล จะถูกแบ่งเป็น 3 แชนแนล คือ สีแดง สีเขียว และสีน้ำเงิน โดยมีค่าตั้งแต่ 0-255 เมื่อถูกแสดงออกมาเป็นรูปภาพดิจิทัลจะใช้หลักการการผสมสีแบบบวก (Additive Color) คือการนำสีที่มีความเข้มมากมาผสมกันแล้วจะเกิดเป็นสีขาว ในการกำหนดข้อมูลรูปภาพสีแบบ RGB จึงประกอบไปด้วย ความกว้าง × ความยาว × ความลึก (Weight × Height × Depth)

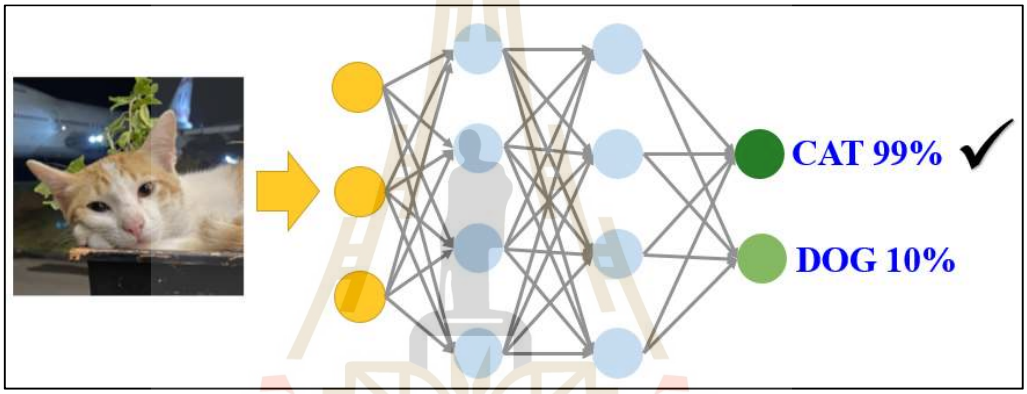


รูปที่ 2.15 ตัวอย่างการจัดเรียงตำแหน่งของพิกเซลในรูปภาพ

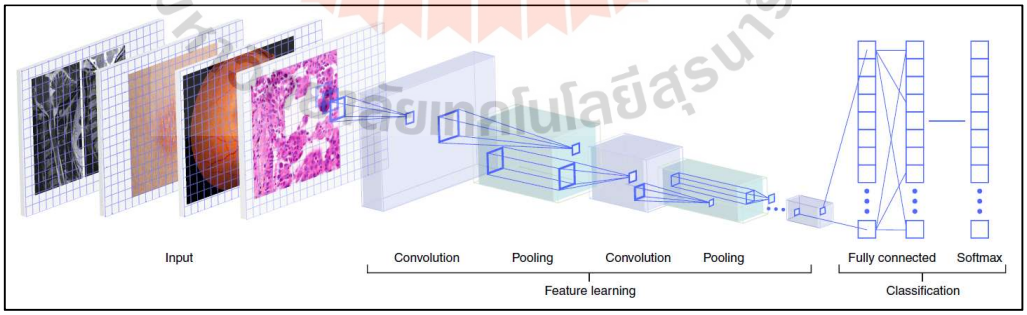
จากรูปที่ 2.15 ข้อมูลภาพดิจิทัลนั้นจะถูกแทนด้วยข้อมูลค่าพิกเซลในรูปแบบตาราง (Grid) โดยตำแหน่งซ้ายบนจะเป็นจุดเริ่มต้น หมายถึงจุด (0,0) และตำแหน่งขวาล่างจะเป็นจุดสิ้นสุด หมายถึงตำแหน่ง (ความกว้าง, ความสูง) ของรูปภาพ โดยในแต่ละจุดจะมีความลึกเท่ากัน

2) การจำแนกประเภทรูปภาพ (Image Classification)

งานทางด้านการจำแนกประเภทรูปภาพคือการระบุชนิดหรือคลาส (Label) ให้กับรูปภาพจากคลาสที่มีการกำหนดไว้แล้ว โดยการวิเคราะห์คุณลักษณะสำคัญของรูปภาพที่ถูกป้อนเข้าไปในโมเดล ผลลัพธ์ที่ได้จะเป็นข้อมูลที่โมเดลทำนายออกมาในลักษณะของชนิดรูปภาพและความน่าจะเป็นของแต่ละชนิด เช่น แมว 99% สุนัข 1% เป็นต้น ดังรูปที่ 2.16



รูปที่ 2.16 ตัวอย่างงานด้านการจำแนกประเภทรูปภาพ



รูปที่ 2.17 โครงสร้างโครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network) (Esteva et al., 2019)

โครงข่ายประสาทเทียมคอนโวลูชัน คือ เพอร์เซ็ปตรอนหลายชั้นอีกรูปแบบหนึ่ง ซึ่งมีความใกล้เคียงกับโครงข่ายประสาทเทียมแบบทั่วไปที่มีการเรียนรู้และปรับค่าน้ำหนักด้วยเทคนิคการแพร่ย้อนกลับ แต่ถูกออกแบบมาเพื่อใช้ในงานทางด้านการประมวลผลภาพจากลักษณะข้อมูลพิกเซลภายในรูปภาพโดยตรงซึ่งใช้เวลาน้อย ดังแสดงตัวอย่างโครงสร้างในรูปที่ 2.17 โดยโครงข่ายประสาทเทียมคอนโวลูชันนั้นสามารถจดจำรูปแบบภาพหรือวัตถุที่มีความแปรปรวนสูงได้อย่างดี เช่น ข้อมูลลายมือ ข้อมูลภาพทางการแพทย์ เป็นต้น และมีความทนทานต่อการเปลี่ยนแปลงของข้อมูลรูปภาพเช่น การเปลี่ยนค่าแสง หรือมุมมองของวัตถุที่ต่างกัน โดยผู้คิดค้นเริ่มแรกคือ Yann Lecun ในปี 1980 โดยสถาปัตยกรรม Convolutional Neural Network หรือ ConvNets ในยุคแรก ๆ มีโครงสร้างชื่อว่า LeNet (LeCun et al., 1988) การทำงานเบื้องต้นเริ่มจากมีข้อมูลนำเข้าเป็นภาพเข้ามาในชั้นคอนโวลูชัน ซึ่งจะมีการใช้งานตัวกรอง (Filter หรือ Kernel) เพื่อดักจับคุณลักษณะบางอย่างที่สำคัญภายในรูปภาพ ตามด้วยชั้นพูลลิ่ง (Pooling) เพื่อลดขนาดข้อมูลเพื่อให้เหลือข้อมูลที่มีคุณลักษณะที่สำคัญหรือเรียกว่า Feature Extraction และในขั้นสุดท้ายคือการนำข้อมูลสำคัญที่ได้ส่งให้กับกระบวนการจำแนกประเภท (Classification) ในชั้นเชื่อมโยงสมบูรณ์ (Fully Connected Layer) เพื่อทำการจำแนกประเภทต่อไป

องค์ประกอบเบื้องต้นของโครงข่ายประสาทเทียมคอนโวลูชัน

Filter, Kernel or Feature Detector

คือ เมทริกซ์ ขนาดเล็ก ใช้สำหรับตรวจจับคุณลักษณะสำคัญในชั้นคอนโวลูชัน

Convolved Feature, Activation Map or Feature Map

คือ ผลลัพธ์ที่ได้จากการคูณเวกเตอร์แบบดอท (Dot Product) ของข้อมูลจุดพิกเซลกับตัวกรอง โดยเลื่อนตัวกรองภายในรูปภาพทั้งรูป

Receptive Field

คือ พื้นที่บางส่วนของรูปภาพที่มีขอบเขตจำกัดสำหรับให้นิวรอนแต่ละนิวรอนรับผิดชอบ ซึ่งมีขนาดเท่ากับขนาดของตัวกรอง

Stride

คือ ตัวเลขที่ใช้กำหนดจำนวนการเลื่อนของตัวกรองภายในรูปภาพ

Zero-padding

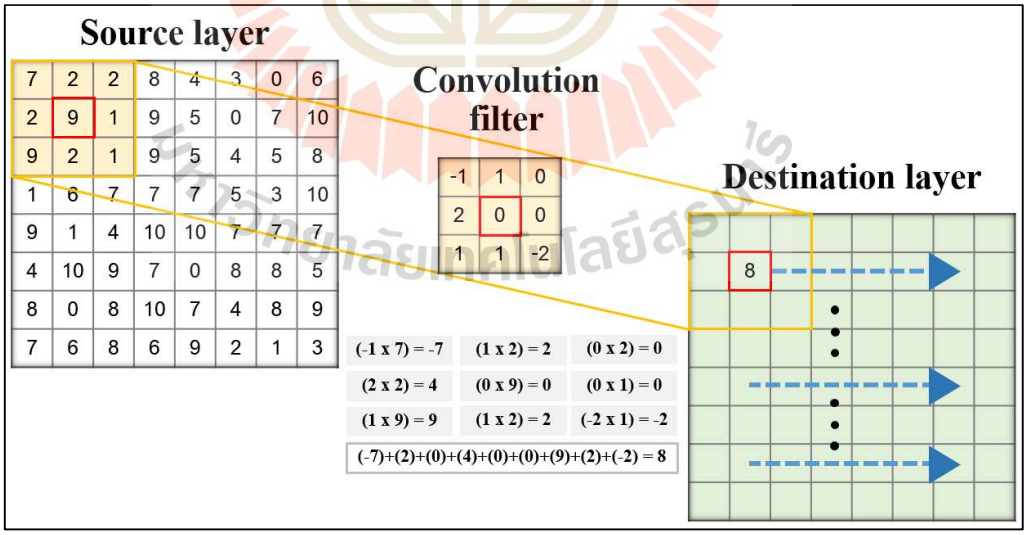
คือ การกำหนดค่า 0 ล้อมรอบไปยังรูปภาพเพื่อให้ขนาดผลลัพธ์หลังจากผ่านการคอนโวลูชันมีขนาดเท่าเดิมและเพื่อไม่ให้สูญเสียข้อมูลสำคัญที่อยู่บริเวณขอบภาพ

ReLU layer

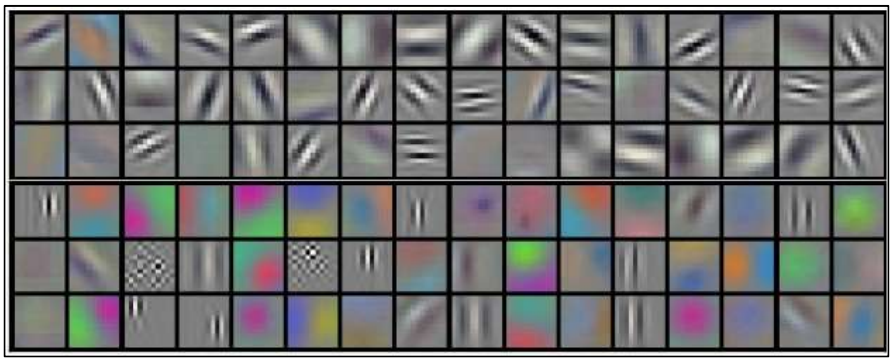
คือ การใช้ ReLU (Rectified Linear Unit) ในส่วนของฟังก์ชันกระตุ้น โดย y จะมีค่า 0 เมื่อ $x<0$ และ $y=x$ เมื่อ $x>=0$

3) การดำเนินการคอนโวลูชัน (Convolution Operation)

การคอนโวลูชันคือการนำข้อมูลพิกเซลในรูปภาพนำเข้ามาดำเนินการคูณเวกเตอร์แบบดอทกับตัวกรองที่ได้กำหนดไว้ เพื่อเป็นการดึงคุณลักษณะภายในขอบเขตบริเวณที่สนใจ โดยจะพิจารณาพื้นที่จากขนาดของตัวกรองเริ่มจากตำแหน่งซ้ายบนสุดของรูปภาพเลื่อนจากซ้ายไปขวา เมื่อดำเนินการจนสุดขอบภาพทางขวาแล้วจะเริ่มต้นใหม่ที่ทางด้านซ้าย วนซ้ำเรื่อย ๆ จนครบขนาดของภาพ โดยที่การเลื่อนแต่ละครั้งจะเลื่อนตามจำนวน Stride ที่ได้กำหนดไว้ ดังตัวอย่างในรูปที่ 2.18 ทำการคอนโวลูชันโดยใช้ตัวกรองขนาด 3 × 3 และกำหนด Stride มีค่าเท่ากับ 1 กับรูปภาพขนาด 8 × 8 พิกเซล และไม่มีการใช้ Zero-padding เมื่อนำตัวกรองไปดำเนินการคูณเวกเตอร์แบบดอทกับรูปภาพนำเข้า จากตัวอย่างในช่องแรก จะได้ผลลัพธ์ = (-1×7) + (1×2) + (0×2) + (2×2) + (0×9) +(0×1) +(1×9) + (1×2) + (-2×1) = 8 หลังจากนั้นจะทำการเลื่อนตัวกรองไปด้านขวาทีละ 1 ช่อง ไปจนสุดขอบภาพและขึ้นบรรทัดใหม่เริ่มจากด้านซ้ายสุดเหมือนเดิม ในกรณีของข้อมูลรูปภาพสี RGB นั้นข้อมูลของรูปภาพจะเก็บอยู่ในลักษณะ 3 มิติ ดังนั้นจึงต้องใช้ตัวกรองที่มีขนาด 3 มิติเช่นเดียวกัน ดังรูปที่ 2.19 แสดงตัวอย่างตัวกรอง 96 แบบของ Krizhevsky ที่ใช้ในการแข่งขันการจำแนกภาพจากฐานข้อมูล ImageNet โดยใช้ตัวกรองมีขนาด 11 × 11 × 3 ในชั้นแรกของการคอนโวลูชัน โดย 48 ตัวกรองด้านบนจะคัดกรองทิศทางการจัดเรียงของค่าสี และ 48 ตัวกรองด้านล่างจะเน้นการคัดกรองคุณลักษณะที่มีค่าสีที่แตกต่างกันออกไป



รูปที่ 2.18 ตัวอย่างการดำเนินการคอนโวลูชัน



รูปที่ 2.19 ตัวอย่างตัวกรอง 96 แบบ ของ Krizhevsky (Krizhevsky et al., 2012)

ขนาดของผลลัพธ์สุดท้าย $(Output)$ จากการคอนโวลูชันจะได้ข้อมูลอยู่ในรูปแบบเมทริกซ์ โดยมีสมการคำนวณขนาดของผลลัพธ์ ดังสมการที่ 2.1, 2.2, 2.3 และ 2.4

$$Output = W_2 \times H_2 \times D_2 \tag{2.1}$$

$$W_2 = \left[\frac{(W_1 - F + 2P)}{S}\right] + 1 \tag{2.2}$$

$$H_2 = \left[\frac{(H_1 - F + 2P)}{S}\right] + 1 \tag{2.3}$$

$$D_2 = D_1 \tag{2.4}$$

เมื่อ F แทนขนาด Filter

S แทนจำนวน Stride

P แทนจำนวน Zero-padding

W_1 แทนขนาดความกว้างของภาพนำเข้า

H_1 แทนขนาดความสูงของภาพนำเข้า

D_1 แทนขนาดความลึกของภาพนำเข้า

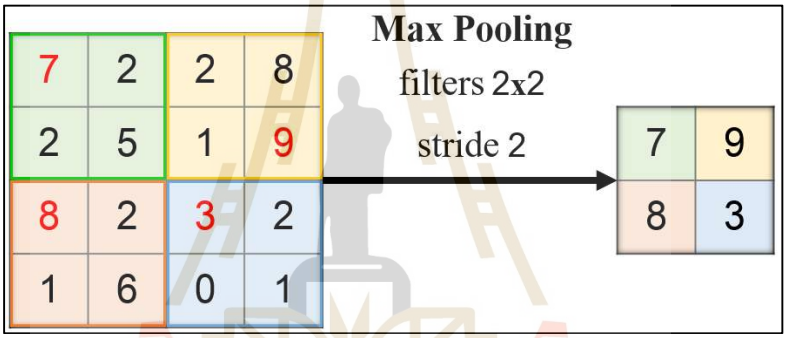
W_2 แทนขนาดความกว้างของผลลัพธ์

H_2 แทนขนาดความสูงของผลลัพธ์

D_2 แทนขนาดความลึกของผลลัพธ์

4) การดำเนินการพูลลิ่ง (Pooling Operation)

การดำเนินการพูลลิ่ง หรือ Pooling Layer เป็นเทคนิคการลดขนาดเชิงพื้นที่ของข้อมูลนำเข้าและลดการคำนวณที่ซับซ้อนของโมเดล นอกจากนี้ยังเป็นการลดปัญหา Overfitting ของโมเดล โดยมีเทคนิคหลายอย่าง เช่น Max Pooling, Average Pooling, L2-norm Pooling ซึ่งโดยทั่วไปจะนิยมใช้ Max Pooling คือการหาค่าสูงสุดในกรอบข้อมูลที่สนใจ เพราะสามารถดักจับค่าที่สำคัญจากข้อมูลสีในพิกเซลที่มีค่ามากได้ โดยจะมี 2 พารามิเตอร์ คือ การกำหนดขนาดของ Filter และ Stride ดังตัวอย่างในรูปที่ 2.20 ทำการพูลลิ่งด้วยการหาค่าสูงสุดโดยกำหนด Filter ขนาด 2 × 2 และ Stride มีค่าเท่ากับ 2 จากขนาดต้นทาง 4 × 4 ผลลัพธ์ที่ได้จึงมีขนาดเท่ากับ 2 × 2 และค่ามากที่สุดในแต่ละกรอบที่คำนวณได้จะมีค่าเท่ากับ 7, 9, 8, 3 ตามลำดับ



รูปที่ 2.20 ตัวอย่างการดำเนินการพูลลิ่ง

ขนาดของผลลัพธ์สุดท้าย $(Output)$ จากการพูลลิ่งจะได้ข้อมูลอยู่ในรูปแบบเมทริกซ์ โดยมีสมการคำนวณขนาดของผลลัพธ์ ดังสมการที่ 2.5, 2.6, 2.7 และ 2.8

$$Output = W_2 \times H_2 \times D_2 \tag{2.5}$$

$$W_2 = \left[\frac{(W_1 - F)}{S}\right] + 1 \tag{2.6}$$

$$H_2 = \left[\frac{(H_1 - F)}{S}\right] + 1 \tag{2.7}$$

$$D_2 = D_1 \tag{2.8}$$

เมื่อ F แทนขนาด Filter

S แทนจำนวน Stride

W_1 แทนขนาดความกว้างของภาพนำเข้า

H_1 แทนขนาดความสูงของภาพนำเข้า

D_1 แทนขนาดความลึกของภาพนำเข้า

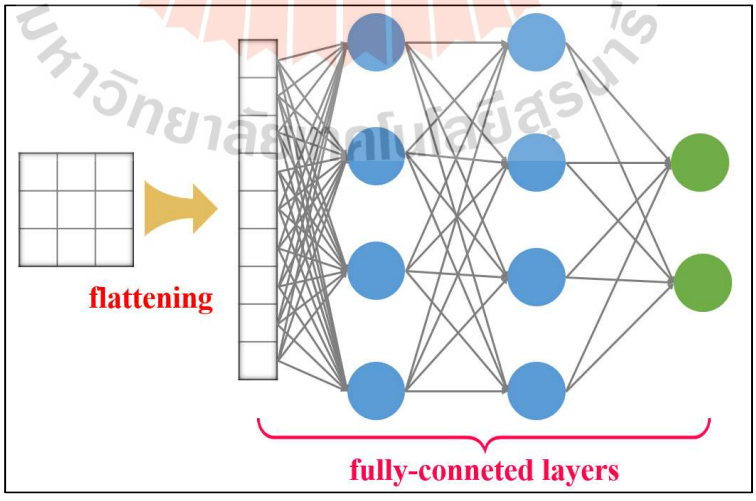
W_2 แทนขนาดความกว้างของผลลัพธ์

H_2 แทนขนาดความสูงของผลลัพธ์

D_2 แทนขนาดความลึกของผลลัพธ์

5) ชั้นเชื่อมโยงสมบูรณ์ (Fully Connected Layer)

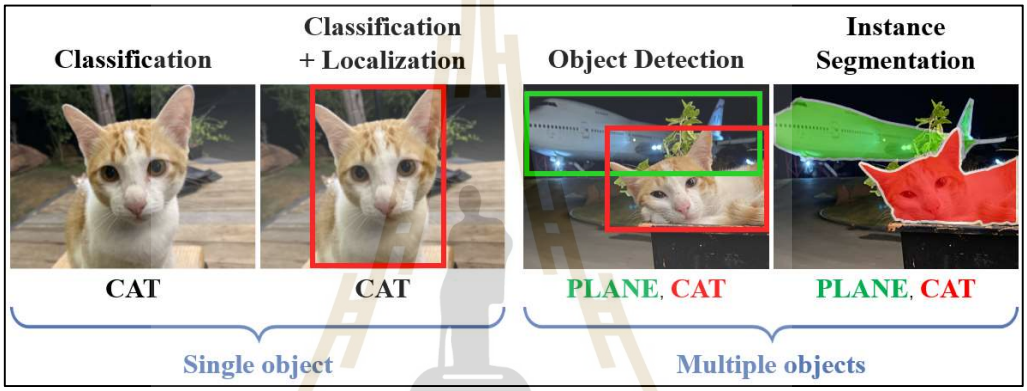
ในชั้นเชื่อมโยงสมบูรณ์ หรือ Fully Connected Layer จะเป็นการตัดสินใจขั้นสุดท้ายของโครงข่ายประสาทเทียมคอนโวลูชัน โดยผลลัพธ์จะเป็นการระบุประเภทหรือคลาส (Class Label) และค่าความน่าจะเป็นของประเภทที่ทำนายได้ โดยก่อนจะเข้าสู่ชั้นนี้จะมีการนำข้อมูลคุณลักษณะที่สกัดได้ในรูปแบบ 2 มิติแปลงให้อยู่ในรูปแบบ 1 มิติก่อน เรียกกระบวนการนี้ว่า Flattening เมื่อได้ข้อมูลในรูปแบบ 1 มิติแล้วจึงทำการเชื่อมต่อกับนิวรอนทุกนิวรอนในชั้นถัด ๆ ไป ดังรูปที่ 2.21 โดยมีการทำงานลักษณะเดียวกันกับโครงข่ายประสาทเทียม แต่จะมีการใช้ฟังก์ชันกระตุ้นเป็นซอฟต์แมก (Softmax Activation Function) เพื่อทำการปรับค่าความน่าจะเป็นในแต่ละคลาสที่คำนวณได้ออกมาให้มีผลรวมกันเท่ากับหนึ่ง



รูปที่ 2.21 ตัวอย่างโครงสร้างชั้นเชื่อมโยงสมบูรณ์ (Fully Connected Layer)

2.2.3 การตรวจจับวัตถุ (Object Detection)

งานด้านการตรวจจับวัตถุ หรือ Object Detection เป็นหนึ่งในงานหลักที่ใช้กับงานด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) ซึ่งเป็นศาสตร์หนึ่งที่พยายามทำให้คอมพิวเตอร์สามารถเข้าใจภาพหรือวิดีโอได้อย่างฉลาดเช่นเดียวกับการมองเห็นของมนุษย์และยังเป็นแขนงวิชาหนึ่งในสาขาวิชาปัญญาประดิษฐ์ โดยเทคนิคการตรวจจับวัตถุนั้นสามารถระบุตำแหน่งของวัตถุที่ปรากฏอยู่ภายในภาพพร้อมระบุชนิดของวัตถุได้ จึงเป็นจุดเริ่มต้นสำคัญสำหรับการประมวลผลและวิเคราะห์ภาพได้อย่างชาญฉลาด และสามารถนำข้อมูลที่ได้ไปสร้างประโยชน์ได้ต่อไป



รูปที่ 2.22 ตัวอย่างเทคนิคการตรวจจับวัตถุภายในภาพ

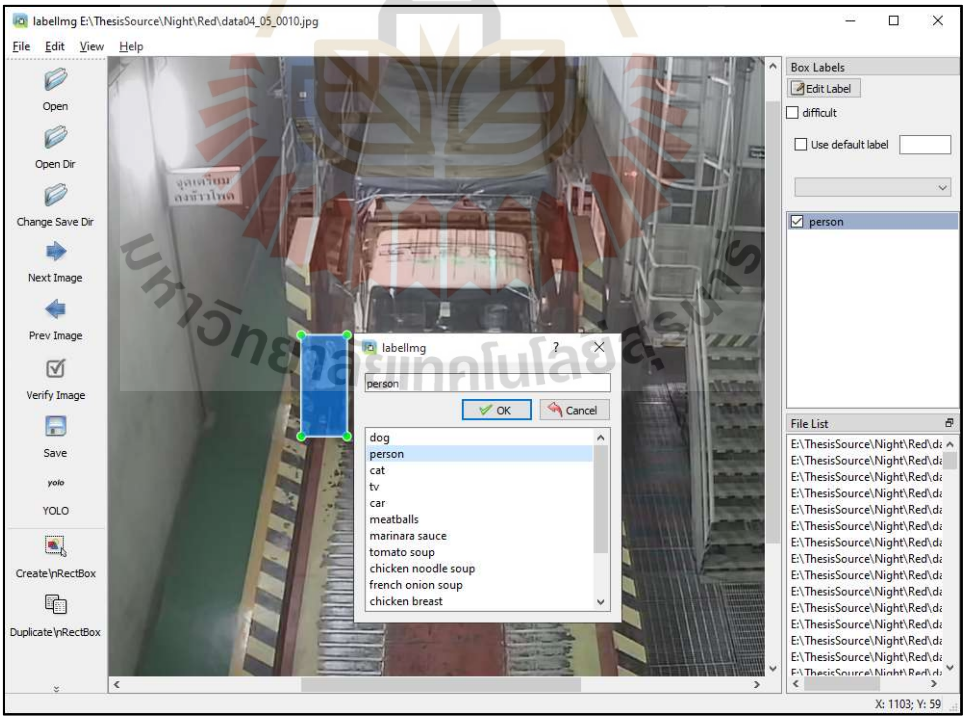
งานจำแนกรูปภาพทั่วไปหรือ Image Classification เป็นการระบุประเภทของรูปภาพนั้น ๆ ว่าเป็นประเภทไหน ต่อมาจึงได้มีการพยายามจำแนกรูปภาพพร้อมกับหาตำแหน่งวัตถุภายในรูปภาพว่าอยู่ตำแหน่งใด (Localization) เพื่อกำหนดชื่อประเภทวัตถุพร้อมกับตีกรอบล้อมรอบตำแหน่งของวัตถุภายในภาพ (Bounding Boxes) รวมถึงการพยายามค้นหาวัตถุและตำแหน่งภายในรูปภาพที่มีหลายวัตถุภายในรูปภาพเดียวกัน งานประเภทนี้เรียกว่างานด้านการตรวจจับวัตถุ (Object Detection) นอกจากการตีกรอบล้อมรอบวัตถุภายในรูปภาพแล้ว ยังมีงานประเภทคัดแยกวัตถุออกมาจากพื้นหลังอย่างชัดเจนจากเส้นขอบของวัตถุพร้อมระบุชนิดวัตถุ เรียกงานประเภทนี้ว่า Instance Segmentation ดังแสดงตัวอย่างในรูปที่ 2.22

งานด้านการตรวจจับวัตถุโดยทั่วไปจะต้องมีการกำหนดหรือพยายามหาตำแหน่งของวัตถุเบื้องต้นที่คิดว่าวัตถุจะอยู่บริเวณนั้น (Region Proposal หรือ Region of Interrest) โดยผลลัพธ์ที่ได้จากขั้นตอนนี้คือชุดข้อมูลที่เก็บข้อมูลตำแหน่งกรอบล้อมรอบที่เป็นไปได้ของวัตถุไว้จำนวนมาก (Set of Bounding Boxes) เพื่อใช้ในขั้นตอนต่อไป โดยจะเป็นการพยายามหาว่าใน

กรอบล้อมรอบที่ได้มานั้นมีวัตถุอยู่ภายในกรอบนั้นหรือไม่ หรือมีค่าความเป็นไปได้มากน้อยเพียงใด

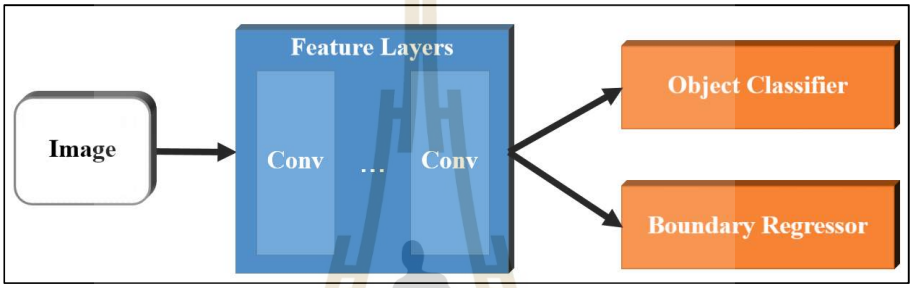
หลังจากผ่านขั้นตอน Region Proposal มาแล้ว จะเป็นขั้นตอนที่นำชุดข้อมูลตำแหน่งกรอบล้อมรอบที่ได้มาวิเคราะห์ว่าภายในกรอบนั้นเป็นวัตถุหรือพื้นหลัง (Final Classification) ซึ่งในขั้นตอนนี้สามารถส่งข้อมูลไปให้โครงข่ายคอนโวลูชันช่วยประมวลผลและวิเคราะห์ โดยที่ถ้าจำนวนของชุดข้อมูลกรอบล้อมรอบมีจำนวนเยอะมาก ๆ จะส่งผลให้การประมวลผลมีความช้าตามไปด้วย เนื่องด้วยข้อด้อยนี้จึงทำให้การใช้โครงสร้างแบบคอนโวลูชันในยุคแรก ๆ ยากที่จะใช้งานได้ในสถานการณ์จริงกับงานที่ต้องการความเร็วในการประมวลผลสูง หรืองานจำพวก Real-time เช่น รถยนต์ไร้คนขับ การมองเห็นของเครื่องจักร เป็นต้น

การฝึกสอนโครงข่ายประสาทเทียมคอนโวลูชันให้สามารถตรวจจับและจำแนกวัตถุได้โดยอัตโนมัติจำเป็นต้องมีข้อมูลจริง หรือ Ground Truth ซึ่งหมายถึงข้อมูลตำแหน่งจริงของวัตถุภายในรูปภาพ โดยส่วนใหญ่ข้อมูลจริงที่ใช้ระบุภายในรูปภาพจะจัดเก็บเป็นชนิดของวัตถุพร้อมกับตำแหน่งกรอบล้อมรอบในรูปแบบของ Coordinates(X, Y, Width, Height) ดังแสดงตัวอย่างการกำหนดข้อมูลจริงดังรูปที่ 2.23



รูปที่ 2.23 ตัวอย่างการทำกรอบล้อมรอบวัตถุจากข้อมูลจริง (Ground Truth)

สถาปัตยกรรมการตรวจจับวัตถุมีหลากหลายชนิด แต่โครงสร้างส่วนใหญ่นั้นจะมีรูปแบบที่คล้ายกันซึ่งมีความใกล้เคียงกับโครงสร้างแบบคอนโวลูชัน คือ เมื่อนำข้อมูลรูปภาพเข้าสู่โมเดลแล้วจะมีชั้นที่สามารถตรวจจับคุณลักษณะที่สำคัญออกมา และต่อเข้ากับชั้นเชื่อมโยงสมบูรณ์เพื่อเป็นการระบุประเภทรูปภาพ ในขณะที่งานด้านการตรวจจับวัตถุจะมีความซับซ้อนเพิ่มยิ่งขึ้น คือเมื่อผ่านชั้นคอนโวลูชันแล้วจะถูกส่งผ่านโครงสร้างโครงข่ายแบบ 2 ประเภทคือ ตัวจำแนกประเภทของวัตถุ และ ตัวคำนวณขอบเขตกรอบล้อมรอบ ดังรูปที่ 2.24

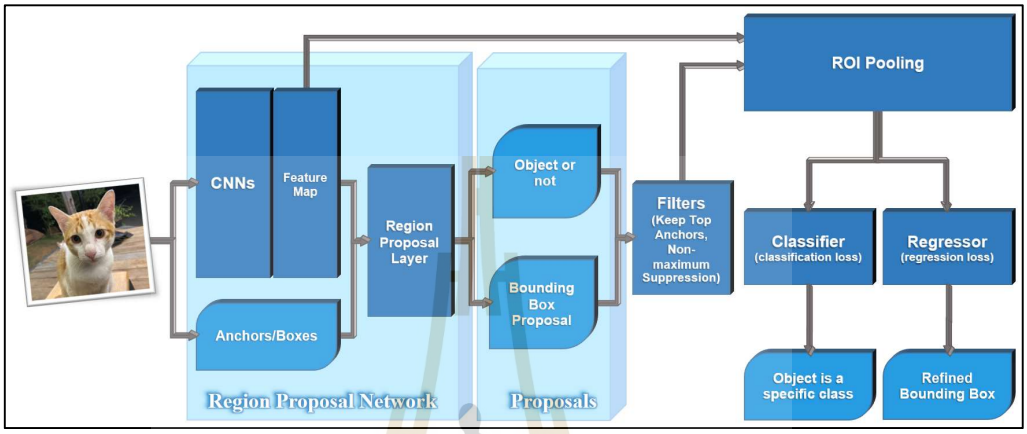


รูปที่ 2.24 โครงสร้างการทำงานของการตรวจจับวัตถุ

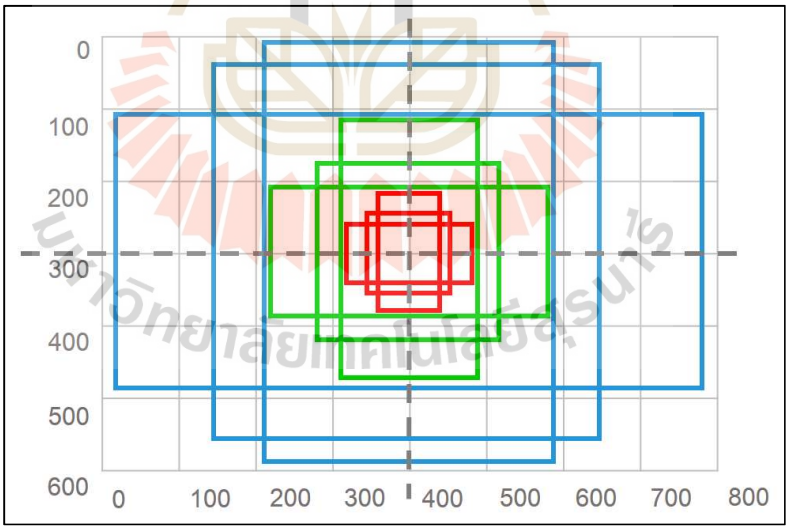
2.2.4 สถาปัตยกรรม Faster R-CNN

Faster R-CNN (Ren et al., 2017) เป็นสถาปัตยกรรมด้านการตรวจจับวัตถุที่ถูกพัฒนาต่อยอดขึ้นมาจาก R-CNN (Girshick et al., 2014) และ Fast R-CNN (Girshick, 2015) เพื่อเพิ่มความเร็วในการประมวลผล โดยมีการปรับปรุงโครงสร้างในส่วนของ Region Proposal Network เพื่อใช้ทดแทนเทคนิค Selective Search (Uijlings et al., 2013) โดยมีการทำงานเบื้องต้นเริ่มจากการนำข้อมูลภาพเข้าสู่กระบวนการคอนโวลูชันเพื่อสกัดคุณลักษณะและได้ผลลัพธ์ออกมาเป็น Feature Map ในขณะเดียวกันจะมีการกำหนดชุดข้อมูลของกรอบล้อมรอบ (Anchors/Boxes) ไว้ล่วงหน้าซึ่งจะถูกนำไปใช้ร่วมกับ Feature Map ในชั้นของ Region Proposal Layer ซึ่งจุดประสงค์ของชั้นนี้คือการนำเสนอขอบเขตที่น่าสนใจ (Region of Interest) ที่มีความเป็นไปได้ว่าจะมีวัตถุปรากฏอยู่ตำแหน่งนั้น ๆ โดยส่วนแรกจะมีการพิจารณาว่าในขอบเขตมีวัตถุอยู่หรือไม่ และอีกส่วนคือการละทิ้งกรอบรอบล้อมที่ไม่มีวัตถุอยู่โดยจะคำนึงถึงกรอบล้อมรอบที่มีวัตถุอยู่เท่านั้นเพื่อส่งให้ส่วนถัดไป คือส่วน Filters ซึ่งมีหน้าที่ละทิ้งจำนวนกรอบล้อมรอบที่มีมากเกินไปโดยจะคัดเลือกเฉพาะกรอบล้อมรอบที่มีความเป็นไปได้สูงที่จะเจอวัตถุ เทคนิคนี้เรียกว่า Non-maximum Suppression ซึ่งจะส่งข้อมูลกรอบล้อมรอบที่มีความเป็นไปได้สูงที่จะมีวัตถุอยู่ให้กับส่วนถัดไป พร้อมกับ Feature Map ที่ได้จากชั้นคอนโวลูชันให้กับส่วน ROI Pooling (Region of

Interest Pooling) โดยมีหลักการทั่วไปคล้ายกับในชั้น Pooling ของโครงข่ายประสาทเทียมคอนโวลูชัน ซึ่งหลังจากผ่านชั้นนี้แล้วจะแบ่งการทำงานออกเป็น 2 ส่วนคือ ส่วนระบุประเภทวัตถุ (Classifier) และ ส่วนปรับแต่งขนาดกรอบล้อมรอบวัตถุ (Refined Bounding Box) ดังรูปที่ 2.25



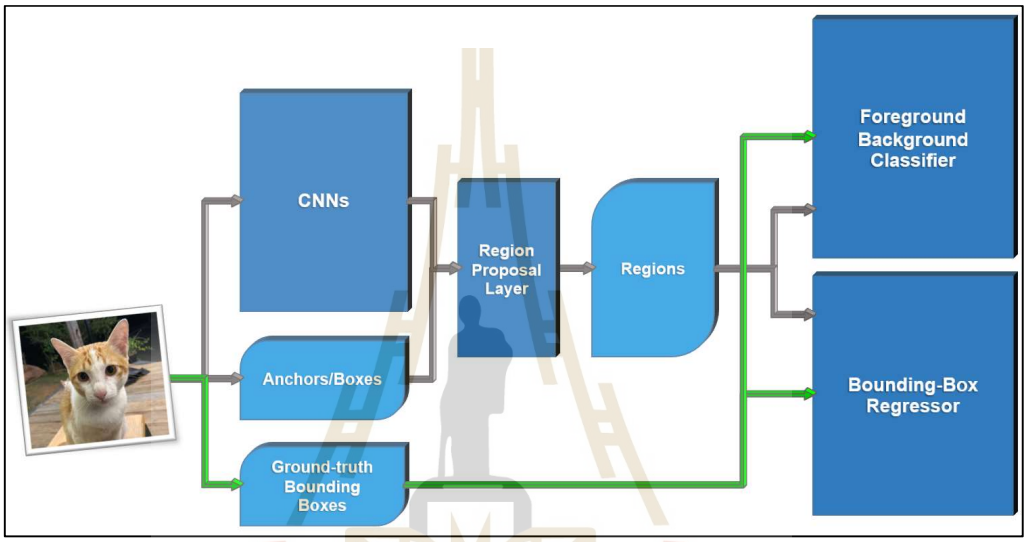
รูปที่ 2.25 โครงสร้างการทำงานของ Region Proposal Network



รูปที่ 2.26 ตัวอย่าง Anchors Boxes

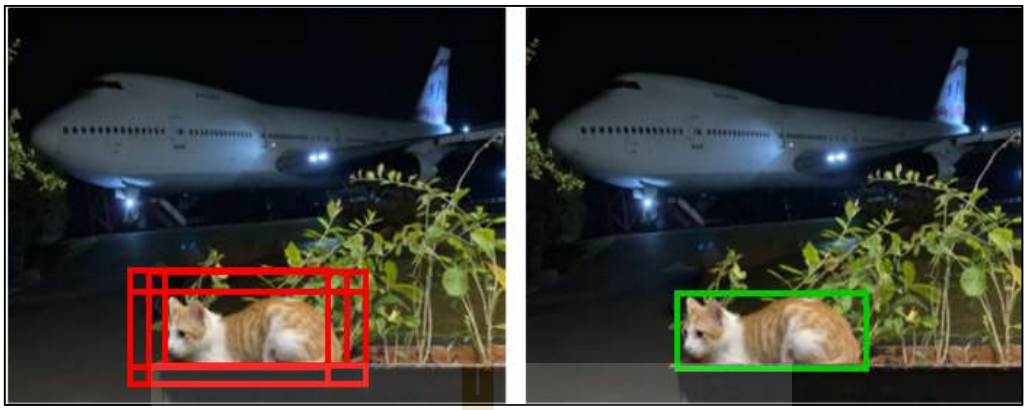
Anchors คือ ชุดข้อมูลของกรอบที่มีขนาดแตกต่างกันใช้สำหรับตรวจจับวัตถุในขนาดต่าง ๆ กัน เช่น รถ คน สัตว์ชนิดต่าง ๆ และวัตถุชนิดอื่น ๆ ดังตัวอย่างในรูปที่ 2.26 มีทั้งหมด

9 Anchors แบ่งเป็น 3 สี มีทั้งหมด 3 ขนาด ดังนี้ 128×128, 256×256, 512×512 โดยในแต่ละขนาดจะถูกแบ่งเป็น 3 อัตราส่วนดังนี้ 1:1, 1:2, 2:1 ดังนั้นถ้า Feature Map ที่ได้มีขนาด 38 × 56 × 9 Anchors และมี Stride เป็น 16 จะได้ผลลัพธ์กรอบที่นำเสนอทั้งหมดเท่ากับ (38 × 56 × 9) / 16 = 1,197 ซึ่งมีจำนวนมากและใช้เวลามากในการประมวลผล ทั้งนี้จึงต้องมีการนำเทคนิค Region Proposal Network เข้ามาช่วยพิจารณา



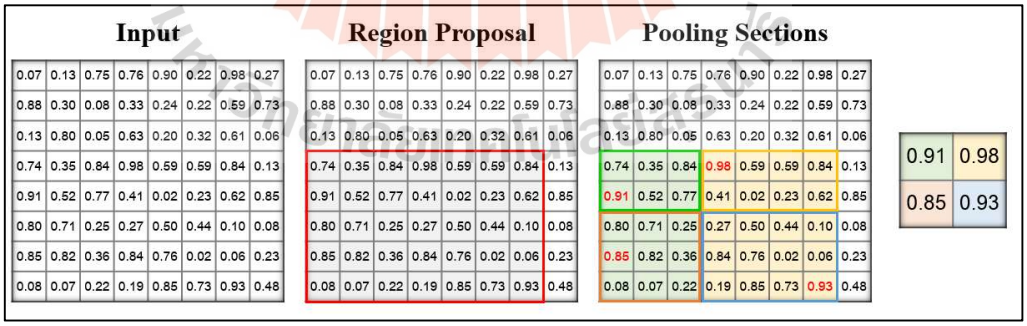
รูปที่ 2.27 โครงสร้างการทำงานของ Region Proposal Network ในขั้นตอนการเรียนรู้

จากรูปที่ 2.27 แสดงถึงโครงข่าย Region Proposal Network ในขั้นตอนการเรียนรู้เริ่มต้นจากการนำ Anchors และผลลัพธ์ที่ได้จากการคอนโวลูชันส่งให้กับ Region Proposal Layer เพื่อนำผลลัพธ์ที่ได้คือกรอบต่าง ๆ (Region) เข้าสู่กระบวนการปรับความแม่นยำให้กับกรอบล้อมรอบ (Bounding Box Regressor) และส่งเข้าสู่ตัวจำแนกชนิดระหว่างภาพวัตถุและภาพพื้นหลัง (Foreground Background Classifier) โดยมีการเปรียบเทียบกับข้อมูลจริงที่นำเข้ามาฝึกสอน (Ground-truth Bounding Boxes) เพื่อคำนวณหาค่าความคลาดเคลื่อน (Error) ซึ่งจะมีทั้งส่วนที่ได้จากกระบวนการปรับความแม่นยำให้กับกรอบล้อมรอบและส่วนจำแนกชนิดระหว่างภาพวัตถุและภาพพื้นหลัง ทำงานลักษณะนี้ในขั้นตอนการเรียนรู้วนซ้ำเรื่อย ๆ เพื่อลดค่าความคลาดเคลื่อนลงจนกว่าจะได้ค่าที่ต่ำสุดหรือค่าที่เหมาะสม หลังจากผ่านกระบวนการ Region Proposal Network แล้วจะยังเหลือกรอบล้อมรอบวัตถุจำนวนหนึ่ง ซึ่งจะส่งข้อมูลส่งให้กระบวนการถัดไปใช้เทคนิค Non-maximum Suppression เพื่อลดจำนวนกรอบล้อมรอบวัตถุ



รูปที่ 2.28 ตัวอย่างการทำงานของเทคนิค Non-maximum Suppression

จากตัวอย่างทางด้านซ้ายในรูปที่ 2.28 เมื่อได้รับข้อมูลกรอบล้อมรอบวัตถุจาก Region Proposal Network จะเห็นว่ามีกรอบล้อมรอบวัตถุจำนวนมากที่ถูกจำแนกประเภทออกมาเหมือนกัน ซึ่งในความเป็นจริงแล้วคือวัตถุเดียวกันจึงควรมีกรอบล้อมรอบแค่เพียงกรอบเดียวดังรูปด้านขวา ดังนั้นในขั้นตอนนี้จึงมีการนำเทคนิค Non-maximum Suppression เพื่อเป็นการผสานกรอบล้อมรอบที่ซ้อนทับกันอยู่ภายในวัตถุเดียวกันให้เหลือเพียงกรอบเดียว โดยพยายามละทิ้งกรอบล้อมรอบวัตถุที่มีความน่าจะเป็นน้อยกว่า และนำกรอบที่เหลือมาวนซ้ำเพื่อพิจารณาจากค่าความน่าจะเป็นที่สูง โดยใช้ค่า IoU (Intersection over Union) ประกอบด้วย



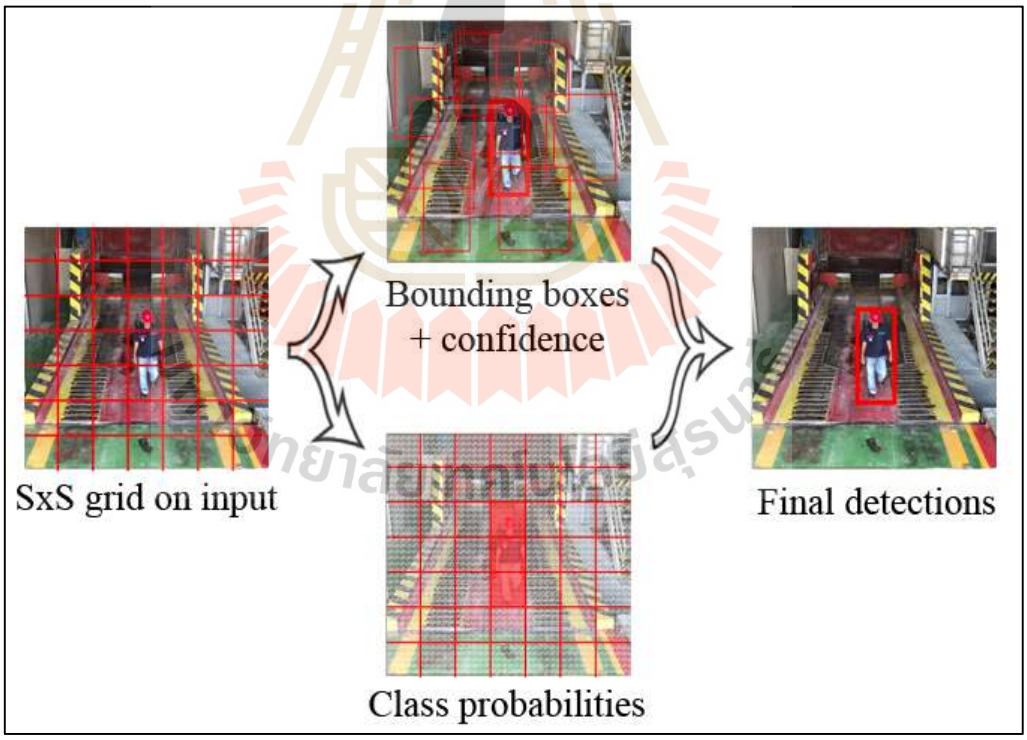
รูปที่ 2.29 ตัวอย่างการทำงานของ ROI Pooling

การทำพูลลิ่งกับขอบเขตที่สนใจหรือ ROI Pooling (Region of Interest Pooling) คือการลดขนาด Feature Map ลงให้มีขนาดเท่ากัน ไม่เหมือนกันการทำ Max Pooling ในโครงข่าย

ประสาทเทียมคอนโวลูชัน แต่จะเป็นการแบ่ง Feature Map ออกเป็น N ส่วน (Region) โดยมีขอบเขตที่ใกล้เคียงกันโดยประมาณ หลังจากนั้นจะทำ Max Pooling ในขอบเขตที่ถูกแบ่งออกทุกส่วน ดังตัวอย่างในรูปที่ 2.29 ผลลัพธ์ที่ได้จะมีขนาดเท่ากับ N เสมอ ซึ่งข้อมูลที่ได้คือคุณลักษณะที่จะถูกส่งต่อไปให้ส่วนระบุประเภทวัตถุและส่วนคำนวณกรอบล้อมรอบวัตถุต่อไป

2.2.5 สถาปัตยกรรม YOLO (You Only Look Once)

การทำงานของสถาปัตยกรรมการตรวจจับวัตถุชนิด YOLO (You Only Look Once) (Redmon et al., 2016) จะแตกต่างออกไปจาก Faster R-CNN ที่มีการทำงานแบบ Two-stage Detection โดยที่ YOLO จะมีการทำงานเพียงแค่ครั้งเดียว (One-stage Detection) โดยมีแนวคิดที่จะทำการจำแนกประเภท (Classification) พร้อมกับกำหนดขอบเขตของกรอบล้อมรอบวัตถุ (Localization) ในกระบวนการเดียว ทั้งนี้เพื่อเพิ่มความเร็วในการประมวลภาพสำหรับงานที่ต้องใช้ความเร็วสูงและต้องทำงานแบบทันที (Real-time) โดยการทำงานเริ่มต้นจากแบ่งรูปภาพนำเข้าเป็นตารางจำนวน S × S (Grid of Cells) ดังรูปที่ 2.30

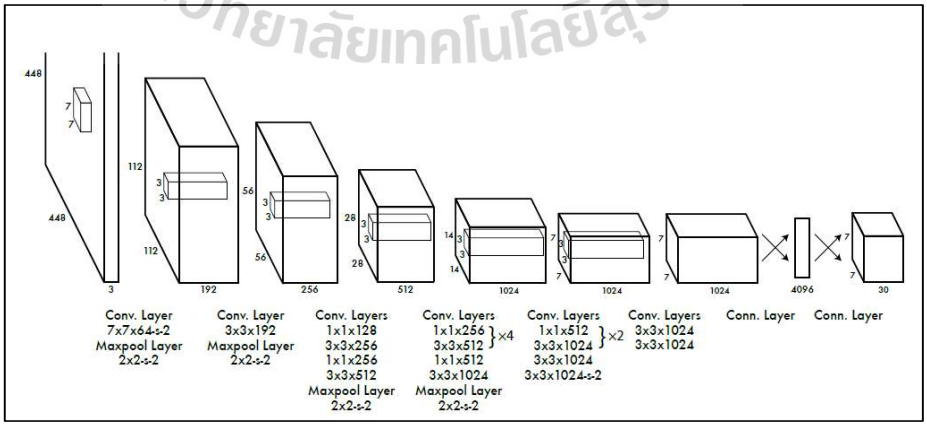


รูปที่ 2.30 ตัวอย่างการทำงานของสถาปัตยกรรม YOLO (You Only Look Once)

โดยในแต่ละช่องจะมีหน้าที่พยายามตรวจจับวัตถุภายในช่องโดยมีแนวคิดที่ว่าวัตถุจะปรากฏอยู่กึ่งกลางของช่องนั้น ๆ ในขั้นตอนการฝึกสอนนั้น YOLO จะพยายามทำนายกรอบล้อมรอบวัตถุหลาย ๆ กรอบในทุก ๆ ช่อง และสุดท้ายจะต้องการเพียงกรอบเดียวเพื่อล้อมรอบวัตถุ โดยพิจารณาจากค่า IoU เปรียบเทียบกับค่าจริง (Ground Truth) โดยในแต่ละกรอบล้อมรอบจะมีองค์ประกอบ 5 อย่าง ดังนี้ (X, Y, Width, Height, Confidence Score) โดยที่ X และ Y คือพิกัดของวัตถุที่อยู่ภายในรูป Width และ Height คือขนาดความกว้างและความสูงของกรอบล้อมรอบวัตถุ และ Confidence Score คือ ค่าความมั่นใจของวัตถุที่อยู่ภายในกรอบล้อมรอบนั้นซึ่งหมายถึงความน่าจะเป็นของวัตถุภายในกรอบที่โมเดลทำนายได้ การพัฒนาสถาปัตยกรรม YOLO ในแต่ละเวอร์ชันมีรายละเอียดดังต่อไปนี้

1) สถาปัตยกรรม YOLOv1

YOLO ในเวอร์ชันแรก (Redmon et al., 2016) นั้นถูกฝึกสอนด้วยข้อมูลรูปภาพจาก ImageNet-1000 ซึ่งสามารถประมวลผลได้รวดเร็วและมีโครงสร้างที่ไม่ซับซ้อนมาก โดยมีการใช้ชั้นคอนโวลูชัน 24 ชั้นตามด้วยชั้นเชื่อมต่อแบบสมบูรณ์ 2 ชั้น ดังรูปที่ 2.31 ข้อจำกัดของ YOLO เวอร์ชันนี้คือการค้นหาวัตถุขนาดเล็กทำได้ไม่ดีนัก ยกตัวอย่างในกรณีที่รูปภาพถูกแบ่งออกเป็น 7 × 7 ช่อง หมายความว่าทั้งรูปภาพนี้จะสามารถตรวจจับวัตถุได้เพียงแค่ 49 วัตถุ โดยที่ในแต่ละช่องจะมีการทำนายกรอบล้อมรอบไว้ 2 กรอบ และจะพิจารณาค่าความแม่นยำในการทำนายประเภทวัตถุจากค่าความมั่นใจที่ได้ (Confidence Score) ซึ่งได้จากการคำนวณ IoU เปรียบเทียบกับค่าจริง (Ground Truth) และในกรณีที่มีวัตถุจำนวนหลาย ๆ วัตถุอยู่ภายในช่องเดียวกันหรือมีตำแหน่งใกล้กันโมเดลจะไม่สามารถระบุวัตถุออกมาได้ทั้งหมด นอกจากนี้ยังพบว่าโครงสร้างดังกล่าวมีความยากลำบากในการเรียนรู้วัตถุที่มีขนาดแตกต่างกัน



รูปที่ 2.31 โครงสร้างของโครงข่าย YOLOv1 (Redmon et al., 2016)

2) สถาปัตยกรรม YOLOv2

YOLO ในเวอร์ชันที่ 2 (Redmon & Farhadi, 2016) ได้ถูกพัฒนาขึ้นโดย Joseph Redmon และ Ali Farhadi ในปี 2016 ผู้พัฒนาได้พยายามเพิ่มประสิทธิภาพให้ดีขึ้นโดยเน้นไปที่การเพิ่มค่าการระลึก (Recall) และการคำนวณตำแหน่งของกรอบล้อมรอบวัตถุ (Localization) โดยที่ยังรักษาค่าความแม่นยำในการทำนายของโมเดล และพยายามเพิ่มความแม่นยำให้เทียบเท่ากับ Faster R-CNN ที่มีการใช้เทคนิค Region Proposal Network นอกจากนี้ยังมีการใช้เทคนิคเพิ่มเติม เช่น มีการใช้เทคนิค Batch Normalization ในทุก ๆ ชั้นของคอนโวลูชัน มีการใช้เทคนิค Higer Resolution Classifier คือเพิ่มขนาดข้อมูลนำเข้าจาก 224 × 224 เป็น 448 × 448 และการเปลี่ยนแปลงที่สำคัญอีกอย่างคือมีการใช้ Anchor Boxes ในการทำนายกรอบล้อมรอบโดยมีการใช้เทคนิค K-means Clustering เข้ามาช่วยในการพิจารณา มีการใช้เทคนิค Fine-grained Features เพื่อแก้ปัญหาในการตรวจจับวัตถุขนาดเล็ก ๆ ของ YOLO เวอร์ชันแรก นั่นคือมีการแบ่งรูปภาพออกเป็น 13 × 13 ช่อง จึงทำให้ YOLO เวอร์ชัน 2 สามารถตรวจจับวัตถุขนาดเล็กได้ดีขึ้นและยังคงมีประสิทธิภาพในการตรวจจับวัตถุขนาดใหญ่ และจากปัญหาด้านการตรวจจับวัตถุที่มีขนาดแตกต่างกันจึงได้มีการนำเสนอเทคนิค Multi-scale Training คือมีการสุ่มรูปภาพของวัตถุในขนาดที่แตกต่างกันในเวอร์ชันที่ 2 ตั้งแต่จาก 320 × 320 ถึง 608 × 608 ในขั้นตอนการฝึกสอน ส่งผลให้โครงสร้างแบบใหม่นี้สามารถเรียนรู้ จดจำ และทำนายวัตถุในขนาดที่แตกต่างกันได้ดียิ่งขึ้น

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

รูปที่ 2.32 โครงสร้างของโครงข่าย Darknet-19 ใน YOLOv2 (Redmon & Farhadi, 2016)

จากรูปที่ 2.32 ใน YOLO เวอร์ชัน 2 ได้มีการปรับเปลี่ยนโครงสร้างโครงข่ายเป็นแบบใหม่เรียกว่าสถาปัตยกรรม Darknet-19 โดยมีการใช้ชั้นคอนโวลูชัน 19 ชั้น ชั้นพูลลิ่ง 5 ชั้น และมีการใช้ Softmax Activation ในขั้นตอนการจำแนกประเภทวัตถุ ซึ่งสถาปัตยกรรม Darknet-19 เป็น Neural Network Framework ที่ถูกพัฒนาขึ้นด้วยภาษา C และ CUDA โดยมีความเร็วในการประมวลผลสูงมากและสามารถใช้กับงาน Real-time ได้

3) สถาปัตยกรรม YOLOv3

เป็นการพัฒนา YOLO เวอร์ชัน 2 เพิ่มเติมเพื่อเพิ่มความแม่นยำมากยิ่งขึ้น โดยมีการใช้เทคนิค Logistic Regression ในการทำนายกรอบล้อมรอบวัตถุโดยในแต่ละกรอบจะมีการคำนวณค่าวัตถุประสงค์หรือค่าความมั่นใจลงไปด้วยนอกจากนี้ใน YOLO เวอร์ชัน 3 (Redmon & Farhadi, 2018) ยังมีการใช้ตัวจำแนกแบบโลจิสติก (Logistic Classifiers) แทน Softmax จึงทำให้โมเดลสามารถทำนายวัตถุหลาย ๆ ประเภทพร้อมกันได้ ยกตัวอย่างเช่น ในรูปที่มีคน สามารถระบุว่าเป็นวัตถุประเภทคน 95% พร้อมกับระบุประเภทวัตถุว่าเป็นผู้ชาย 85% ได้ด้วยเช่นกัน ต่างจากการใช้ Softmax ที่ผลสุดท้ายแล้วค่าความมั่นใจจะถูกลดทอนลงตามประเภทวัตถุที่พบแล้วมีผลรวมกันเป็น 1 เพราะจากตัวอย่างคือวัตถุประเภทคนสามารถเป็นได้ทั้งคนและผู้ชายในเวลาเดียวกันได้ นอกจากนี้ยังมีการปรับเปลี่ยนในส่วนของการทำนายโดยมีความคล้ายคลึงกับเทคนิค Feature Pyramid Networks (FPN) คือมีการทำนาย 3 ครั้งในแต่ละขนาดของรูปภาพโดยทำการดึงคุณลักษณะออกมาทำนายในแต่ละครั้งด้วยขนาดที่แตกต่างกัน ซึ่งส่งผลให้ YOLO เวอร์ชัน 3 มีความสามารถในการทำนายรูปภาพที่หลากหลายขนาดได้แม่นยำยิ่งขึ้น

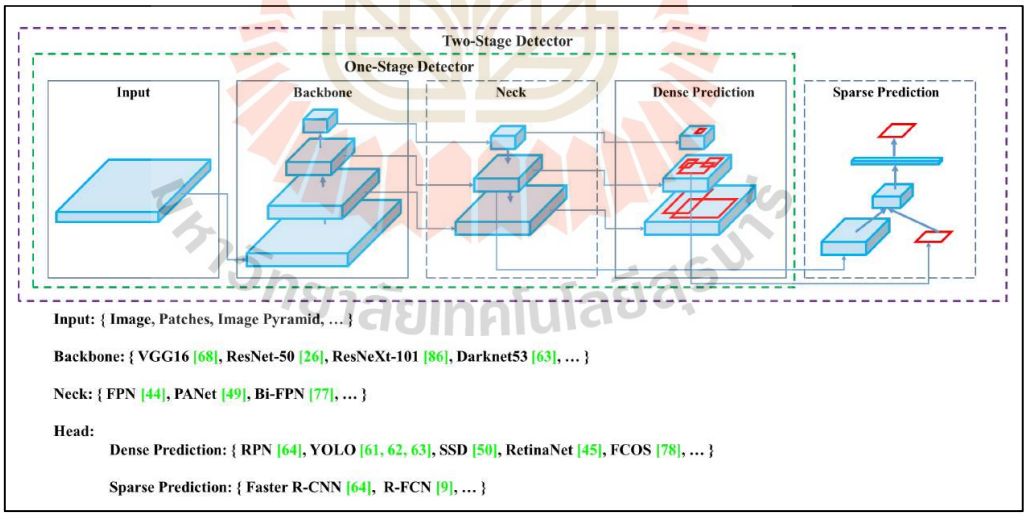
	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3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

รูปที่ 2.33 โครงสร้างของโครงข่าย Darknet-53 ใน YOLOv3 (Redmon & Farhadi, 2018)

จากรูปที่ 2.33 ในสถาปัตยกรรม YOLO เวอร์ชัน 3 ได้มีการปรับปรุงโครงข่าย Darknet-19 ร่วมกับเทคนิค Residual Network คือมีการนำเทคนิคข้ามการเชื่อมต่อระหว่างชั้นคอนโวลูชันในแต่ละรอบเพื่อลดความสูญเสียไม่ให้คุณลักษณะบางอย่างเลือนหายไป โครงข่ายแบบใหม่นี้ใช้ชื่อว่า Darknet-53 ซึ่งมีการใช้คอนโวลูชันทั้งหมด 53 ชั้น ด้วยตัวสกัดคุณลักษณะแบบใหม่นี้ทำให้ YOLO เวอร์ชัน 3 มีความแม่นยำในการตรวจจับวัตถุมากยิ่งขึ้นและถือเป็นสถาปัตยกรรมการตรวจจับวัตถุที่เน้นความเร็วในการประมวลผลที่ก้าวหน้าที่สุด ณ เวลานั้นเมื่อเปรียบเทียบกับสถาปัตยกรรมอื่น

4) สถาปัตยกรรม YOLOv4

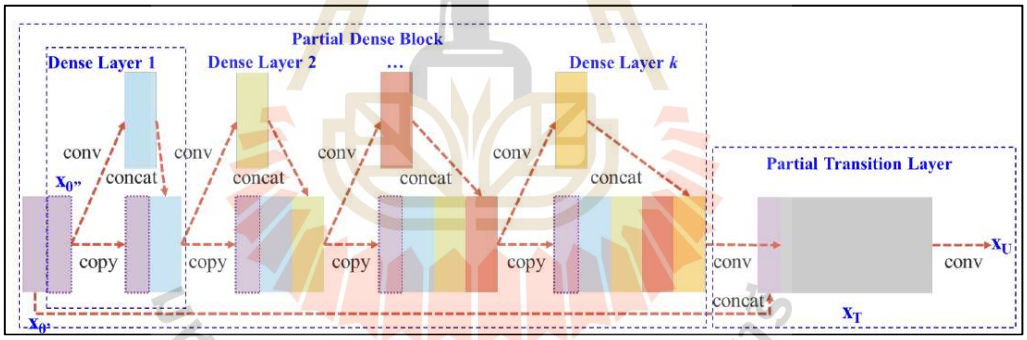
เนื่องจาก Joseph Redmon ผู้สร้าง YOLO ได้ประกาศยุติการวิจัยงานที่เกี่ยวข้องกับคอมพิวเตอร์วิทัศน์เพื่อหลีกเลี่ยงการใช้เทคโนโลยีไปในทางที่ผิดโดยคำนึงถึงจริยธรรมเป็นหลัก ในปี 2020 Alexey Bochkovskiy, Chien-Yao Wang และ Hong-Yuan Mark Liao จึงได้ปรับปรุงสถาปัตยกรรม YOLOv3 ต่อในชื่อ YOLOv4 (Bochkovskiy et al., 2020) โดยเพิ่มประสิทธิภาพในด้านความแม่นยำและความเร็วในการประมวลผล โดยนำเทคนิคใหม่ ๆ ทางด้านการเรียนรู้เชิงลึกมาปรับใช้กับโครงสร้างการตรวจจับวัตถุแบบเดิม ทั้งในส่วนของ Backbone, Neck และ Head โดยมีลักษณะโครงสร้างดังรูปที่ 2.34



รูปที่ 2.34 โครงสร้างทั่วไปของสถาปัตยกรรมการตรวจจับวัตถุ (Bochkovskiy et al., 2020)

YOLOv4 Backbone Network

โดยทั่วไปในงานตรวจจับวัตถุจะมีการใช้โครงข่ายเบื้องหลัง (Backbone) จากโครงข่ายที่ถูกฝึกสอนไว้ล่วงหน้าแล้ว (Pre-trained) จากข้อมูล ImageNet ซึ่งหมายความว่าโครงข่ายดังกล่าวได้รับการเรียนรู้และปรับค่าน้ำหนักให้เข้ากับคุณลักษณะของชุดข้อมูลฝึกสอนไว้แล้วเพื่อนำไปใช้กับงานใหม่ได้สะดวกและรวดเร็วขึ้น โดยใน YOLO เวอร์ชัน 4 ผู้พัฒนาได้พัฒนาโครงสร้างแบบ CSPDarknet53 ซึ่งมีโครงข่ายแบบ DenseNet คือ โครงสร้างโครงข่ายที่ออกแบบมาเพื่อเชื่อมต่อชั้นคอนโวลูชันในแต่ละชั้นเข้าด้วยกัน ดังรูปที่ 2.35 เพื่อลดปัญหาเกรเดียนต์ที่มีขนาดเล็กลงเรื่อย ๆ จนไม่ส่งผลต่อการปรับปรุงค่าน้ำหนักอีกต่อไปสำหรับโครงข่ายที่มีความลึกมาก ๆ (Vanishing Gradient) และยังช่วยเพิ่มการเผยแพร่คุณลักษณะที่สกัดออกมาได้ให้มีการนำกลับมาใช้ใหม่ นอกจากนี้ยังเป็นการลดจำนวนพารามิเตอร์ในโครงข่ายอีกด้วย โดยในโครงข่ายแบบ CSPDarknet53 ได้มีการปรับปรุงเพิ่มเติมโดยแยกคุณลักษณะของชั้นฐานโดยคัดลอกและส่งต่อไปให้กับชั้นถัด ๆ ไป ซึ่งเป็นการลดปัญหาคอขวดทางด้านการคำนวณที่ซับซ้อนและยังเป็นการปรับปรุงกระบวนการเรียนรู้โดยส่งคุณลักษณะที่ยังไม่ถูกปรับปรุงไปยังชั้นถัด ๆ ไปด้วย



รูปที่ 2.35 โครงสร้างโครงข่ายของเทคนิค Cross Stage Partial DenseNet (Wang et al., 2020)

YOLOv4 Neck - Feature Aggregation

ในขั้นตอนนี้เป็นการนำคุณลักษณะที่สกัดได้จากชั้นคอนโวลูชันมารวมกัน (Neck) เพื่อเตรียมพร้อมสำหรับการตรวจจับวัตถุในขั้นตอนถัดไป โดย YOLO เวอร์ชัน 4 ได้เลือกใช้เทคนิค PANet (Path Aggregation Network) (Liu et al., 2018) ในการรวมคุณลักษณะของโครงข่าย

YOLOv4 Head - The Detection Step

ในส่วนการตรวจจับ (Head) ของ YOLO เวอร์ชัน 4 ได้ใช้เทคนิคเดียวกันกับ YOLO เวอร์ชัน 3 คือมีการใช้ Anchors ในการคำนวณและมีการตรวจจับแบบ 3 ระดับ นอกจากขั้นตอนที่กล่าวมายังมีการนำเทคนิคอื่น ๆ มาช่วยเพิ่มประสิทธิภาพ ดังนี้

เทคนิค Bag of Freebies (BoF)

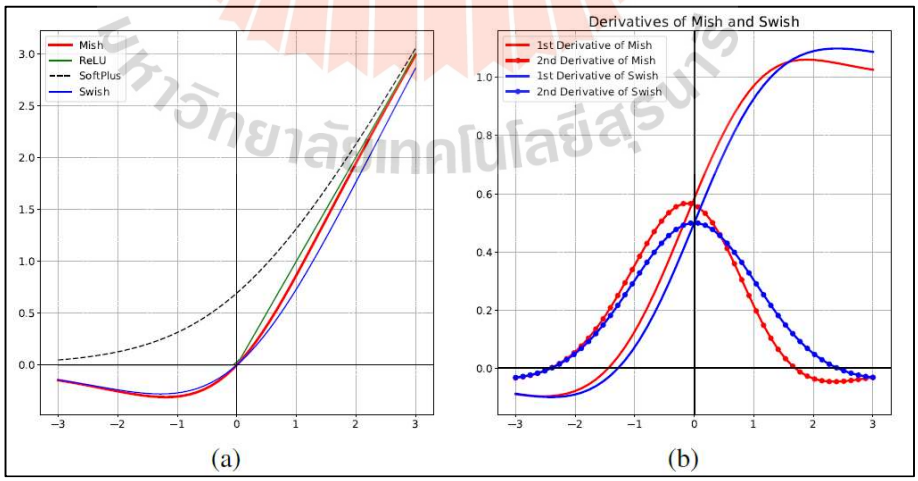
เป็นชุดของเทคนิคที่ถูกปรับใช้ในส่วน Neck และ Backbone เพื่อเพิ่มความแม่นยำโดยไม่ทำให้เวลาในการประมวลผลในการใช้งานจริงเพิ่มขึ้น แต่จะส่งผลเฉพาะขั้นตอนการเรียนรู้ โดยชุดของเทคนิคดังกล่าว มีดังนี้ Augmentation, Random Cropping, Drop Block เป็นต้น ซึ่งเทคนิคเหล่านี้คือเทคนิคทั่วไปของงานด้านคอมพิวเตอร์วิทัศน์ โดย YOLO เวอร์ชัน 4 นำมาใช้เพื่อเพิ่มขนาดของชุดข้อมูลในขั้นตอนการฝึกสอนรวมถึงใช้ในการตรวจสอบเพื่อเพิ่มความแม่นยำมากยิ่งขึ้น นอกจากนี้ยังมีการนำเสนอเทคนิคใหม่คือ Mosaic Data Augmentation มาใช้งาน ซึ่งเป็นการรวมภาพ 4 ภาพเข้าด้วยกัน โดยสอนให้โมเดลค้นหาวัตถุที่มีขนาดเล็กและลดความสนใจกับฉากรอบ ๆ ที่ไม่ใช่วัตถุลง ดังรูปที่ 2.36

รูปที่ 2.36 ตัวอย่างเทคนิค Mosaic Data Augmentation (Bochkovskiy et al., 2020)

อีกหนึ่งเทคนิคใหม่ที่ได้ถูกนำมาใช้งานคือ เทคนิค Self Adversarial Training (SAT) โดยมีหลักการทำงานที่จะพยายามค้นหาส่วนของรูปภาพที่มีผลต่อกระบวนการเรียนรู้ของโครงข่ายมากที่สุด ต่อจากนั้นจะพยายามบดบังส่วนดังกล่าวเพื่อให้โครงข่ายพยายามที่จะหาคุณลักษณะสำคัญใหม่ ๆ ที่จะมาช่วยตรวจจับเพิ่มเติมได้ นอกจากนี้ยังมีการปรับเปลี่ยนในส่วนของ Loss Function โดยมีการใช้ Complete IoU (CIoU) เข้ามาใช้แทนในส่วนของการทำนายกรอบล้อมรอบวัตถุเปรียบเทียบกับขนาดกรอบจริง โดยมีแนวคิดพื้นฐานจากข้อจำกัดในการพิจารณาเฉพาะการทับซ้อนของกรอบล้อมรอบที่ทำนายได้กับขนาดกรอบจริงนั้นอาจไม่เพียงพอต่อการตัดสินใจ จำเป็นต้องใช้ปัจจัยอื่น ๆ เพิ่มเติมในการตรวจสอบ เช่น ปัจจัยทางด้านรูปทรงเรขาคณิต ลักษณะการทับซ้อน ระยะห่างของกรอบ รวมถึงอัตราส่วนของกรอบ เป็นต้น ยกตัวอย่างในกรณีที่ไม่มีการทับซ้อนกัน อาจจะต้องมีการพิจารณาระยะทางระหว่างกรอบล้อมรอบที่ถูกทำนายกับกรอบล้อมรอบจากข้อมูลจริงว่ามีระยะทางห่างกันมากเพียงใดและจะทำอย่างไรให้โครงข่ายพยายามเลื่อนกรอบที่ถูกทำนายให้เข้าใกล้กับข้อมูลจริงได้มากที่สุด

เทคนิค Bag of Specials (BoS)

เป็นอีกหนึ่งชุดของเทคนิคที่ได้นำมาใช้กับ YOLO เวอร์ชัน 4 ในส่วนของ Neck และ Backbone แต่มีผลกับเวลาในการประมวลเพิ่มขึ้นเล็กน้อยเพื่อเพิ่มประสิทธิภาพและความแม่นยำของโครงข่ายยิ่งขึ้น ซึ่งชุดของเทคนิคดังกล่าว มีดังนี้ Cross Stage Partial Connection (CSP), Spatial Attention Module (SAM), Path Aggregation Network (PAN), Spatial Pyramid Pooling (SPP), Mish Activation ดังรูป 2.37 เป็นต้น ดังที่ได้กล่าวไว้แล้วก่อนหน้านี้



รูปที่ 2.37 ตัวอย่างลักษณะของกราฟ Mish เปรียบเทียบกับ Activation Function อื่น ๆ (Misra, 2019)

2.3 มาตรวัดประสิทธิภาพในงานด้านการตรวจจับวัตถุ

การวัดประสิทธิภาพของโมเดลในงานด้านการตรวจจับวัตถุมีหน่วยวัดที่นิยมใช้อยู่หลากหลายชนิดโดยส่วนใหญ่มีการคำนวณเบื้องต้นจากการหาค่าอัตราส่วนระหว่างพื้นที่ทับซ้อนของกรอบล้อมรอบวัตถุที่ทำนายได้เปรียบเทียบกับกรอบล้อมรอบวัตถุขนาดจริง (Intersection over Union หรือ IoU) ดังรูปที่ 2.38 โดยจะมีการตั้งค่าเกณฑ์ (Threshold) ไว้เช่น 0.5 0.75 0.95 เป็นต้น ยิ่งค่า IoU มีค่ามากหมายความว่าโมเดลสามารถทำนายกรอบล้อมรอบวัตถุนั้นได้แม่นยำมาก โดยจะมีการกำหนดมาตรวัดประสิทธิภาพต่าง ๆ ดังนี้

TP (True Positive)

คือ จำนวนวัตถุที่โมเดลสร้างกรอบล้อมรอบวัตถุได้ถูกต้อง (IoU ≥ Threshold)

FP (False Positive)

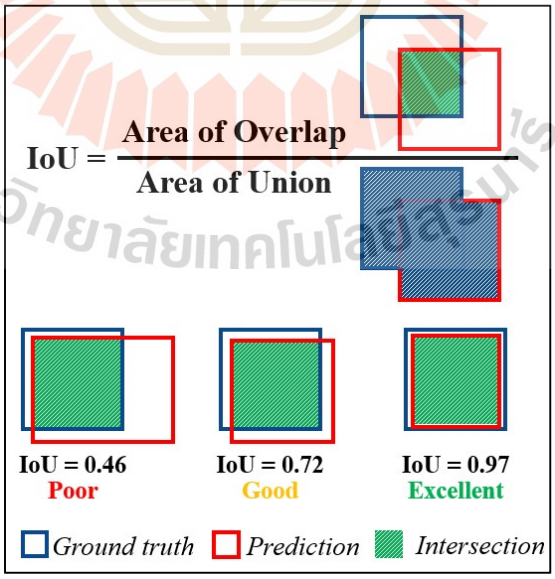
คือ จำนวนวัตถุที่โมเดลสร้างกรอบล้อมรอบวัตถุได้ไม่ถูกต้อง (IoU < Threshold)

FN (False Negative)

คือ จำนวนวัตถุที่โมเดลไม่สร้างกรอบล้อมรอบวัตถุ

TN (True Negative)

คือ จำนวนวัตถุที่โมเดลไม่สร้างกรอบล้อมรอบวัตถุเนื่องจากเป็นวัตถุที่ไม่อยู่ในความสนใจ ดังนั้น TN จึงไม่ถูกนำมาใช้ในงานด้านการตรวจจับวัตถุ



รูปที่ 2.38 ตัวอย่างการคำนวณอัตราส่วนระหว่างพื้นที่ทับซ้อนของกรอบล้อมรอบวัตถุที่ทำนายได้เปรียบเทียบกับกรอบล้อมรอบวัตถุขนาดจริง Intersection over Union (IoU)

2.3.1 ค่าความเที่ยงตรง (Precision)

คือ หน่วยวัดค่าความเที่ยงตรงของโมเดลโดยเปรียบเทียบจากจำนวนกรอบล้อมรอบวัตถุที่ทำนายได้ถูกต้องกับจำนวนกรอบที่ถูกโมเดลสร้างขึ้นมาทั้งหมด ดังสมการที่ 2.9

$$Precision = \frac{TP}{TP + FP} \tag{2.9}$$

2.3.2 ค่าการระลึก (Recall)

คือ หน่วยวัดค่าการระลึกหรือการจดจำได้ของโมเดลโดยเปรียบเทียบจากจำนวนกรอบล้อมรอบวัตถุที่ทำนายได้ถูกต้องกับจำนวนกรอบจากข้อมูลจริงทั้งหมด ดังสมการที่ 2.10

$$Recall = \frac{TP}{TP + FN} \tag{2.11}$$

2.3.3 ค่าประสิทธิภาพโดยรวม (F-measure)

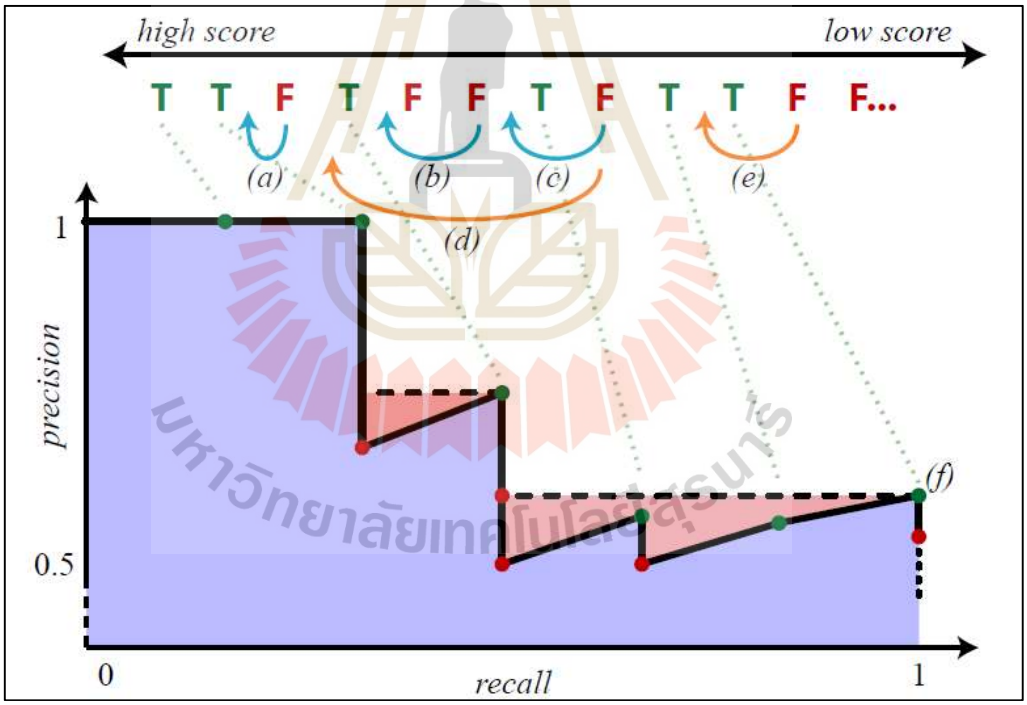
F-measure หรือ F1 Score คือ หน่วยวัดประสิทธิภาพของโมเดลที่ได้จากการคำนวณค่าความเที่ยงตรงและค่าการระลึกเพื่อให้ได้ค่าผลลัพธ์ออกมาเพียงหน่วยเดียวที่แสดงถึงค่าประสิทธิภาพโดยรวมของโมเดล ดังสมการที่ 2.11

$$F - measure = 2 \times (\frac{Precision \times Recall}{Precision + Recall}) \tag{2.10}$$

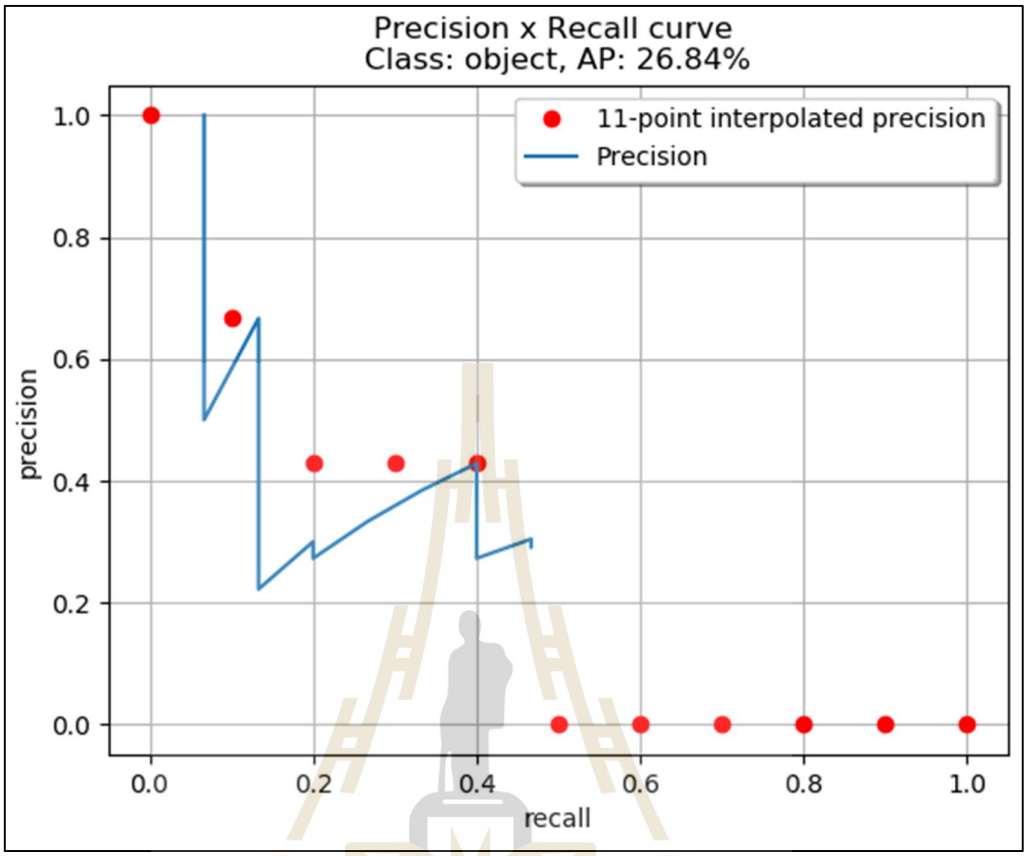
2.3.4 ค่าความเที่ยงตรงเฉลี่ย (Average Precision)

คือ หน่วยวัดความแม่นยำของโมเดลที่นิยมใช้วัดประสิทธิภาพในงานด้านการตรวจจับวัตถุโดยคำนวณหาพื้นที่ใต้กราฟระหว่างค่าความเที่ยงตรง (Precision) และ ค่าการระลึก (Recall) ในแต่ละวัตถุที่โมเดลตรวจจับได้ โดยนำข้อมูลผลลัพธ์ที่ได้จากการตรวจจับวัตถุประเภทเดียวกันมาเรียงลำดับตามค่าความมั่นใจจากมากไปหาน้อย (Confidence Score) โดยในแต่ละวัตถุที่ตรวจจับได้จะนำผลลัพธ์ TP FP และ FN ที่ได้จากการพิจารณาจากค่าเกณฑ์ IoU ที่กำหนดไว้ มาคำนวณหาค่าความเที่ยงตรงและค่าการระลึกไล่ตามลำดับที่เรียงไว้เพื่อนำมาสร้างเป็นกราฟ ดังรูปที่ 2.39 โดยที่ T คือ True Positive และ F คือ False Positive โดยพิจารณาลำดับจากซ้ายไปขวา เมื่อผลลัพธ์เป็น T ค่าความเที่ยงตรงและค่าการระลึกจะสูงขึ้นและเมื่อผลลัพธ์ที่ได้เป็น F จะส่งผลต่อค่าความเที่ยงตรงที่ลดลง ดังสมการ Precision และ Recall กราฟที่ได้ส่วนใหญ่จึงมีลักษณะคดไปคดมา (Zigzag Pattern) ในการหาพื้นที่ใต้กราฟหรือค่า AP นั้นมีการนำเสนอเทคนิคในหลายรูปแบบ

โดยในยุคแรก ๆ จะใช้เทคนิคการหาพื้นที่ใต้กราฟแบบ 11-Point Interpolated Precision โดยเทคนิคนี้จะเป็นการใช้ค่าความเที่ยงตรงจากการประมาณ ณ ตำแหน่งของค่าการระลึกที่สนใจ 11 จุด โดยหาจากค่าความเที่ยงตรงที่มากที่สุดจากตำแหน่งค่าการระลึกที่สนใจไปยังตำแหน่งด้านขวาสุดของกราฟ ค่า AP ที่ได้จึงเป็นพื้นที่ใต้กราฟสีม่วงรวมกับสีชมพู โดยที่สีชมพูแสดงถึงพื้นที่ถูกเพิ่มเติมจากการประมาณ ดังนั้นในบางกรณีที่มีการสลับลำดับของผลลัพธ์จากตัวอย่างลูกศรสีฟ้า (a) (b) และ (c) จะไม่ส่งผลต่อพื้นที่ใต้กราฟ แต่กรณีลูกศรสีส้ม (e) และ (d) จะส่งผลให้พื้นที่ใต้กราฟที่ได้นั้นเปลี่ยนไป การหาพื้นที่ใต้กราฟแบบ 11-Point Interpolated Precision นั้นจะใช้ตำแหน่งการระลึกทั้งหมด 11 จุด ดังนี้ {0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0} ดังรูปที่ 2.40 เช่นเดียวกับมาตรวัดที่ใช้ในการแข่งขัน The Pascal Visual Object Classes (VOC) Challenge 2007 (Everingham et al., 2010) โดยที่ Average Precision (AP) หรือค่าความเที่ยงตรงเฉลี่ยสามารถคำนวณได้ดังสมการที่ 2.12 และ 2.13



รูปที่ 2.39 ตัวอย่างกราฟ Precision-Recall (Henderson & Ferrari, 2016)



รูปที่ 2.40 ตัวอย่างการประมาณค่าความเที่ยงตรงจากค่าการระลึกที่สนใจทั้งหมด 11 จุด (Padilla et al., 2016)

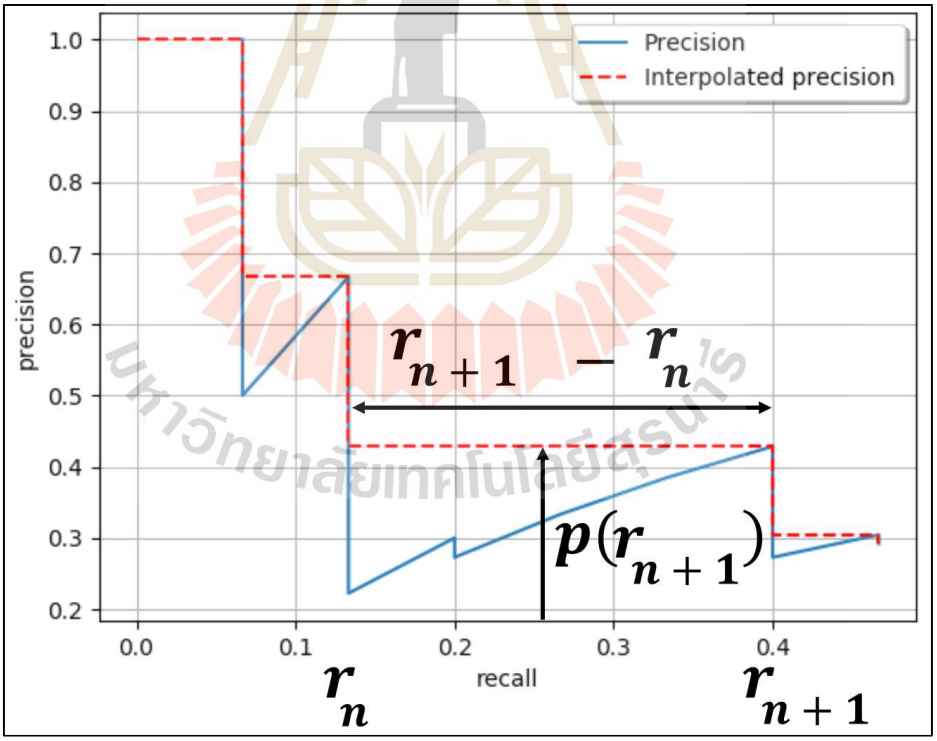
$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} Pinterp(r) \tag{2.12}$$

เมื่อ $Pinterp(r)$ คือค่าความเที่ยงตรงจากการประมาณ ณ ตำแหน่งที่ r โดยที่ r คือตำแหน่งของค่าการระลึกที่สนใจ มีทั้งหมด 11 จุด จาก 0.0 ถึง 1.0 เพิ่มขึ้นทีละ 0.1

$$Pinterp(r) = \max_{r' : r' \geq r} P(r') \tag{2.13}$$

เมื่อ $P(r')$ คือค่าความเที่ยงตรงใด ๆ ในช่วง r ถึง r' โดยที่ r' คือตำแหน่งของค่าการระลึกใด ๆ ที่มีค่ามากกว่าหรือเท่ากับ r ดังนั้น ค่าความเที่ยงตรงที่สูงที่สุดภายในช่วงจึงถูกใช้เป็นค่าความเที่ยงตรงจากการประมาณเพื่อนำไปใช้ในการหาค่า AP

แต่เนื่องจากการหาพื้นที่ใต้กราฟแบบ 11-Point Interpolated Precision นั้นยังไม่ละเอียดพอสำหรับการทดสอบกับข้อมูลจำนวนมากและผลลัพธ์ที่มีค่าความเที่ยงตรงที่แตกต่างกันไม่มากนักในช่วงค่าการระลึกที่สนใจในช่วงเดียวกัน เนื่องจากการเติมเต็มให้กับกราฟที่มีลักษณะคดไปคดมาเป็นค่าสูงสุดในแต่ละตำแหน่งของค่าการระลึกที่สนใจนั้น ทำให้ความแตกต่างที่ได้นั้นหายไปและส่งผลให้ค่าความเที่ยงตรงเฉลี่ยมีค่ามากกว่าที่ควรจะเป็น ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงได้ใช้วิธีการคำนวณหาค่าความเที่ยงตรงเฉลี่ยแบบใหม่หรือที่เรียกว่า Area under curve (AUC) ในการทดสอบประสิทธิภาพของโมเดล เช่นเดียวกับการปรับเปลี่ยนการคำนวณหาค่าความเที่ยงตรงเฉลี่ยในการแข่งขัน The Pascal Visual Object Classes (VOC) Challenge หลังจากปี 2010 เป็นต้นมา ซึ่งเป็นการปรับเปลี่ยนวิธีการคำนวณการหาพื้นที่ใต้กราฟโดยการใช้ตำแหน่งจากค่าการระลึกที่สนใจทั้งหมด 11 จุด เปลี่ยนเป็นการใช้ตำแหน่งค่าการระลึกทั้งหมดโดยพิจารณาจากค่าความเที่ยงตรงที่มากที่สุดนับจากตำแหน่งค่าการระลึกปัจจุบันจนถึง 1 ดังแสดงตัวอย่างดังรูปที่ 2.41 และสามารถคำนวณได้ดังสมการที่ 2.14 และ 2.15



รูปที่ 2.41 ตัวอย่างการประมาณค่าความเที่ยงตรงจากค่าการระลึกที่สนใจทุกจุดเมื่อค่าความเที่ยงตรงมีการเปลี่ยนแปลง (Padilla et al., 2016)

$$AP = \sum (r_{n+1} - r_n) Pinterp(r_{n+1}) \tag{2.12}$$

เมื่อ $Pinterp(r_{n+1})$ คือค่าความเที่ยงตรงจากการประมาณ ณ ตำแหน่งที่ r_{n+1} โดยที่ r_n คือตำแหน่งของค่าการระลึกปัจจุบัน และ r_{n+1} คือตำแหน่งของค่าการระลึกถัดไปโดยพิจารณาจากค่าความเที่ยงตรงมากที่สุดนับจากตำแหน่งค่าการระลึกปัจจุบันจนถึง 1

$$Pinterp(r_{n+1}) = \max_{r' : r' \geq r_n} P(r') \tag{2.13}$$

เมื่อ $P(r')$ คือค่าความเที่ยงตรงใด ๆ ในช่วง r_n ถึง r' โดยที่ r' คือตำแหน่งของค่าการระลึกใด ๆ ที่มีค่ามากกว่าหรือเท่ากับ r_n ดังนั้น ค่าความเที่ยงตรงที่สูงที่สุดภายในช่วงจึงถูกใช้เป็นค่าความเที่ยงตรงจากการประมาณเพื่อนำไปใช้ในการหาค่า AP

2.3.5 ความเร็วในการประมวลผลภาพต่อวินาที (Frame per Second)

ความเร็วในการประมวลผลภาพต่อวินาทีเป็นหน่วยวัดประสิทธิภาพในด้านความเร็วในการประมวลผลของโมเดล โดยคำนวณจากเวลาที่โมเดลใช้ในการทำนายใน 1 วินาที สามารถทำนายภาพที่มีวัตถุอยู่ภายในภาพได้จำนวนกี่ภาพ มีหน่วยเป็น รูปภาพ ต่อ วินาที (Frame per Second หรือ FPS) ดังสมการที่ 2.14 ในงานวิจัยนี้ผู้วิจัยใช้ความเร็วในการประมวลผลภาพต่อวินาทีเฉลี่ย (AVERAGE FPS) ของแต่ละกลุ่มข้อมูลที่ใช้ในทดสอบ โดยสามารถคำนวณได้ดังสมการที่ 2.15

$$FPS = \frac{1000}{Processing\ Time\ (ms)} \tag{2.14}$$

เมื่อ $Processing\ Time$ คือเวลาที่โมเดลใช้ในการประมวลผลภาพ (หน่วยมิลลิวินาที)

$$AVERAGE\ FPS = \frac{1}{n} \sum_{i=0}^{n} \left(\frac{1000}{Processing\ Time\ (ms)} \right) \tag{2.15}$$

เมื่อ $Processing\ Time$ คือเวลาที่โมเดลใช้ในการประมวลผลภาพ (หน่วยมิลลิวินาที)

n คือจำนวนภาพที่ใช้ในการทดสอบ

2.4 งานวิจัยที่เกี่ยวข้อง

ในปัจจุบันมีงานวิจัยมากมายที่เกี่ยวข้องกับการตรวจจับบุคคลภายในรูปภาพรวมถึงการระบุตำแหน่งของบุคคล โดยมีการประยุกต์ใช้เทคนิคต่าง ๆ เช่น การใช้เทคนิคการประมวลผลภาพ การใช้ความสามารถของอุปกรณ์ในการตรวจจับ รวมถึงการใช้เทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก โดยผู้วิจัยได้ทบทวนวรรณกรรมที่เกี่ยวข้องกับเทคนิคการตรวจจับข้อมูลบุคคลภายในภาพและงานวิจัยที่ใช้เทคนิคคอมพิวเตอร์วิทัศน์เข้ามาใช้ในงานด้านความปลอดภัย ทั้งภายในโรงงานและภายนอกโรงงาน ดังนี้

การตรวจจับมนุษย์หรือบุคคลภายในรูปภาพนั้นได้รับความนิยมอย่างมากจากงานวิจัยในปี ค.ศ. 2005 ของ Dalal และ Triggs (2005) โดยประยุกต์ใช้การคัดแยกคุณลักษณะด้วยเทคนิค Histograms of Oriented Gradients (HOG) ในการตรวจจับมนุษย์โดยเฉพาะ ซึ่งเป็นวิธีค้นหาเส้นขอบโดยพิจารณาทิศทางการกระจายตัวของเส้นขอบ จากการแบ่งรูปภาพเป็นช่องหรือเรียกว่าเซลล์ย่อย (Cells) เพื่อใช้ในการเก็บค่า Histogram ของในแต่ละเซลล์ และใช้หน้าต่างที่ทับซ้อนกัน (Block) ในการคำนวณหาค่าเกรเดียนต์ ในที่สุดจะได้คุณลักษณะที่สำคัญออกมาทั้งขนาดและทิศทางของเส้นขอบในแต่ละเซลล์ และจะถูกส่งข้อมูลต่อให้กับอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) สำหรับใช้ในการจำแนกวัตถุออกมาโดยจะพิจารณาว่าเป็นรูปแบบภาพบุคคลหรือไม่ใช่บุคคล โดยใช้ข้อมูลจาก MIT และ INRIA ในการฝึกสอนและทดสอบ ซึ่งผลลัพธ์ที่ได้ของงานวิจัยนี้สามารถตรวจจับบุคคลได้อย่างแม่นยำและเป็นต้นแบบของงานวิจัยในด้านการตรวจจับบุคคลในเวลาต่อมา

ปี ค.ศ. 2006 Zhu และคณะ (2006) ได้นำเสนอเทคนิค Cascade of Histograms of Oriented Gradients (HOG) โดยมีการประยุกต์ใช้เทคนิค Cascade of Rejector ร่วมกับการใช้ AdaBoost มาใช้ในกระบวนการคัดเลือก Block ของงานวิจัย Dalal และ Triggs ในปี ค.ศ. 2005 โดยมีความพยายามที่จะเพิ่มความเร็วในการประมวลของของแต่ละหน้าต่างที่ใช้ในขั้นตอนการตรวจจับ รวมถึงเพิ่มความแม่นยำให้มากยิ่งขึ้น โดยเพิ่มจำนวนหน้าต่างตรวจจับจากเดิม 800 เป็น 12,800 ผลลัพธ์ที่ได้ในงานวิจัยนี้คือโมเดลสามารถทำงานได้รวดเร็วขึ้นกว่าเดิมถึง 70 เท่า โดยมีความเร็วภาพต่อวินาทีอยู่ระหว่าง 5 – 30 ภาพต่อวินาที ซึ่งต่างจากงานของ Dalal และ Triggs ที่สามารถทำความเร็วได้เพียงแค่ 1 ภาพต่อวินาที สำหรับภาพขนาด 320 × 240 พิกเซล

นอกจากเทคนิคทางด้านการประมวลผลภาพและการเรียนรู้ของเครื่องแล้ว ในงานด้านการตรวจจับบุคคล ยังมีการประยุกต์ใช้ความสามารถของอุปกรณ์ต่าง ๆ มาช่วยในการตรวจจับ โดยในปี ค.ศ. 2011 Xia และคณะ (2011) ได้ทดลองนำกล้อง Kinect ที่ถูกพัฒนาโดยบริษัท Microsoft มาประยุกต์ใช้กับงานด้านการตรวจจับบุคคล โดยใช้ข้อมูลความลึกที่ได้จากเซนเซอร์ตรวจวัดความลึกภายในกล้อง (3D Depth Sensor) มาใช้ในการตรวจจับรูปแบบลักษณะของศีรษะโดยใช้เทคนิค

2D Chamfer Matching จากการคำนวณเส้นขอบในลักษณะภาพ 2 มิติและพิจารณาเพิ่มเติมอีกครั้งโดยสร้างรูปทรงของศีรษะในรูปแบบ 3 มิติโดยใช้ข้อมูลความลึกที่ได้ นอกจากนี้ยังมีการนำเสนออัลกอริทึมในการติดตามการเคลื่อนไหวของคน (Human Tracking) โดยใช้ข้อมูลในแต่ละเฟรมที่ได้จากกล้องวัดความลึก ผลการวิจัยนี้ได้แสดงให้เห็นถึงประสิทธิภาพในการประยุกต์ใช้กล้องวัดความลึกกับงานทางด้านการตรวจจับบุคคลซึ่งมีค่าความแม่นยำสูงถึง 98.4% แต่ก็มีข้อเสียในกรณีที่ศีรษะถูกบดบังด้วยสิ่งกีดขวางจะไม่สามารถตรวจจับบุคคลได้

หลังจากนั้นในเวลาต่อมาเทคนิคการเรียนรู้เชิงลึกได้รับความนิยมเป็นอย่างมากจึงได้มีการประยุกต์ใช้ในการตรวจจับวัตถุอย่างแพร่หลายรวมถึงการตรวจจับบุคคล ในปี ค.ศ. 2016 Zhang และคณะ (2016) ได้ค้นพบว่าการตรวจจับข้อมูลคนเดินเท้าด้วยโครงสร้างแบบ Faster R-CNN นั้นยังทำได้ไม่ดีนัก และพยายามค้นหาสาเหตุโดยทดลองโครงข่ายในแบบฉบับของตัวเองโดยใช้ข้อมูลคนเดินเท้าจากฐานข้อมูล Caltech, INRIA, ETH และ KITTI ในการทดลอง การปรับโครงข่ายการตรวจจับวัตถุโดยประยุกต์ใช้เทคนิคต่าง ๆ นั้น ผู้วิจัยสรุปได้ว่าสาเหตุที่ Faster R-CNN ไม่มีความแม่นยำพอกับข้อมูลคนเดินเท้าเนื่องจากความละเอียดของคุณลักษณะไม่เพียงพอกับข้อมูลคนที่มีขนาดเล็ก ผลลัพธ์ที่ดีที่สุดของโครงข่ายจากงานวิจัยนี้คือโครงข่ายที่ใช้เทคนิค Region Proposal Network (RPN) ร่วมกับเทคนิค Boosted Forest (BF) ซึ่งสามารถลด False Positive และ Miss Rate ลงมากกว่าโครงข่ายอื่น ๆ เมื่อเปรียบเทียบกับทุกชุดข้อมูล รวมถึงมีค่า Mean Average Precision มากที่สุดกับชุดข้อมูล KITTI เมื่อเปรียบเทียบกับโครงข่ายอื่น ๆ

ในงานด้านคอมพิวเตอร์วิทัศน์นั้นค่อนข้างมีข้อจำกัดด้านแสงสว่าง ซึ่งมีผลโดยตรงต่อความแม่นยำในการตรวจจับวัตถุภายในภาพ การตรวจจับวัตถุในเวลากลางคืนจึงเป็นงานที่ท้าทายและยากกว่างานทั่ว ๆ ไป ในปี ค.ศ. 2018 Kim และคณะ (2018) จึงได้นำเสนอเทคนิคการตรวจจับคนเดินเท้าในเวลากลางคืนโดยใช้โครงข่ายแบบ Faster R-CNN ซึ่งมีการเพิ่มเทคนิคการผสมผสานคุณลักษณะที่ได้จากชั้นคอนโวลูชันของเฟรมที่ต่อเนื่องกันมาพิจารณาร่วมด้วย (Fusion of Deep Convolutional Features) โดยใช้ข้อมูลภาพคนเดินเท้าที่มีทั้งเวลากลางวันและในเวลากลางคืนในสภาพแสงที่ดวงตามนุษย์สามารถมองเห็นได้ (Visible-light) จากฐานข้อมูล KAIST, Caltech และ NICTA ในการทดลอง ผลลัพธ์จากการทดลองแสดงให้เห็นว่าเทคนิคที่นำเสนอมีความแม่นยำมากขึ้นกว่าโครงข่ายแบบเดิมทั้งเวลากลางวันและกลางคืน ซึ่งจะส่งผลที่แม่นยำในการนำไปใช้จริงกับกล้องวงจรปิดที่ติดตั้งตามท้องถนน

ในปี ค.ศ. 2019 Zengeler และคณะ (2019) ได้ทำการวิจัยเพื่อทดสอบประสิทธิภาพการตรวจจับบุคคลภายในโรงงานด้วยเทคนิคแตกต่างกัน เพื่อเปรียบเทียบความแม่นยำของการตรวจจับบุคคลในสภาพแวดล้อมภายในโรงงานที่ยากต่อการตรวจจับ เช่น มีฝุ่น สภาพแสงที่แตกต่างกันในแต่ละบริเวณ และบุคคลที่ไม่สามารถมองเห็นได้เต็มตัว โดยมีการใช้เทคนิค

Histogram of Oriented Gradients (HOG), You Only Look Once (YOLO), OpenPose (OP) ในการทดลองกับข้อมูลภาพที่ได้จากวิดีโอที่ถูกบันทึกภายในห้องปฏิบัติการอย่างต่อเนื่องในเหตุการณ์ที่แตกต่างกัน และมีการใช้เทคนิคการเพิ่มค่าแสง ค่ารบกวน และเพิ่มการบิดรูป เพื่อทำการทดสอบ โดยผลลัพธ์ที่ได้นั้นเทคนิค HOG สามารถทำความเร็วในการประมวลผลได้เร็วที่สุดคือ 33.5 ภาพต่อวินาที และ YOLO ทำเวลาได้ดีเป็นอันดับถัดมาที่ 18.4 ภาพต่อวินาที และ OpenPose ประมวลผลได้ช้าที่สุดคือ 11.3 ภาพต่อวินาที แต่เมื่อพิจารณาถึงความแม่นยำแล้ว OpenPose มีความแม่นยำมากที่สุดและแสดงให้เห็นถึงความทนทานต่อภาพที่มีสัญญาณรบกวนรวมถึงภาพในพื้นที่ที่มีสภาพแสงน้อย ตามมาด้วย YOLO และ HOG ตามลำดับ

ในปี ค.ศ. 2019 Shim และ Choi (2019) ได้มีการประยุกต์ใช้การตรวจจับวัตถุด้วยเทคนิคการเรียนรู้เชิงลึกกับงานทางด้านความปลอดภัยภายในพื้นที่ก่อสร้างเพื่อเป็นการลดความสูญเสียจากการเกิดอุบัติเหตุภายในสถานที่ก่อสร้าง โดยใช้การตรวจจับวัตถุที่อยู่บริเวณด้านหน้าของรถที่ใช้ในการก่อสร้างโดยเฉพาะคนงานและสัตว์ โดยการประมวลผลภาพจากกล้องบริเวณด้านหน้าของรถด้วยสถาปัตยกรรม YOLOv3 หลังจากนั้นจะคำนวณจุดกึ่งกลางของวัตถุจากขนาดของกรอบล้อมรอบวัตถุที่ถูกตรวจจับได้และนำไปคำนวณหาความเสี่ยงจากขอบเขตพื้นที่ที่มีความเสี่ยงในการเกิดอุบัติเหตุสูงบริเวณด้านหน้าของรถ ถ้าพบว่ามีวัตถุอยู่ในบริเวณที่อาจเกิดอุบัติเหตุได้ ระบบจะแจ้งเตือนไปยังคนขับรถเพื่อเป็นการป้องกันอุบัติเหตุในเบื้องต้น ผลลัพธ์การทำงานของงานวิจัยนี้มีความเร็วในการประมวลผลวิดีโออยู่ที่ 15.06 เฟรมต่อวินาที และยังมีความแม่นยำในการประเมินสถานการณ์อันตรายได้ถึง 90.48%

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่ามีงานวิจัยจำนวนมากที่ศึกษาเกี่ยวกับงานด้านการตรวจจับบุคคลภายในภาพ และได้มีการนำเสนอเทคนิคต่าง ๆ เพื่อช่วยเพิ่มความแม่นยำในการตรวจจับบุคคลทั้งในสภาพแวดล้อมที่มีสภาพแสงปกติและสภาพแสงน้อย ทั้งนี้ยังมีการพัฒนาและปรับปรุงโครงสร้างการทำงานภายในอัลกอริทึมที่แตกต่างกันเพื่อเพิ่มความเร็วในการประมวลผลภาพ ซึ่งมีความจำเป็นมากในงานที่ต้องการความเร็วในการประมวลผลสูง ในบางงานวิจัยจึงมีการประยุกต์ใช้การตรวจจับบุคคลกับงานทางด้านความปลอดภัยเพื่อป้องกันอุบัติเหตุที่อาจเกิดขึ้นได้ ดังนั้นวิทยานิพนธ์นี้จึงประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกที่มีความแม่นยำสูงและใช้เวลาในการประมวลผลน้อย โดยคัดเลือกโมเดลการตรวจจับวัตถุที่ถูกพัฒนาขึ้นด้วยสถาปัตยกรรมที่แตกต่างกัน คือ Faster R-CNN และ YOLOv4 มาพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์

งานวิจัยที่เกี่ยวข้องประกอบด้วย

ก แทนงานวิจัยของ Dalal และ Triggs (2005)

ข แทนงานวิจัยของ Zhu และคณะ (2006)

ค แทนงานวิจัยของ Xia และคณะ (2011)

ง แทนงานวิจัยของ Zhang และคณะ (2016)

จ แทนงานวิจัยของ Kim และคณะ (2018)

ฉ แทนงานวิจัยของ Zengeler และคณะ (2019)

ช แทนงานวิจัยของ Shim และ Choi (2019)

ซ* แทนงานวิจัยของวิทยานิพนธ์ฉบับนี้

ตารางที่ 2.1 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยง

กระบวนการทำงาน	**งานวิจัยที่เกี่ยวข้อง**							
	ก	**ข**	**ค**	**ง**	**จ**	**ฉ**	**ช**	**ซ***
จุดประสงค์การวิจัย								
การตรวจจับบุคคล	✓	✓	✓	✓	✓	✓	✓	✓
การระบุระดับความเสี่ยงของบุคคล							✓	✓
เทคนิคที่ใช้ในการตรวจจับบุคคล								
3D Depth Sensors			✓					
Chamfer Matching			✓					
Histograms of Oriented Gradients	✓	✓				✓		
Support Vector Machine	✓	✓				✓		
R-CNN				✓				
Fast R-CNN				✓				
Faster R-CNN				✓	✓			✓
YOLOv1						✓		
YOLOv3							✓	
YOLOv4								✓
OpenPose						✓		

ตารางที่ 2.1 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยง (ต่อ)

กระบวนการทำงาน	งานวิจัยที่เกี่ยวข้อง							
	ก	ข	ค	ง	จ	ฉ	ช	ซ*
เทคนิคที่ใช้ในการระบุความเสี่ยงของบุคคล								
พิจารณาจากตำแหน่งจุดกึ่งกลางของของกรอบล้อมรอบวัตถุ							✓	
พิจารณาจากตำแหน่งจุด 3 จุดด้านล่างของกรอบล้อมรอบวัตถุ (มุมซ้าย, กึ่งกลาง, มุมขวา)								✓
ข้อมูลที่ใช้ในการทดสอบ								
นำมาจากแหล่งอ้างอิงที่มีอยู่แล้ว	✓	✓		✓	✓			✓
เก็บรวบรวมข้อมูลด้วยผู้วิจัย			✓			✓	✓	✓
ลักษณะข้อมูลที่ใช้ในการทดสอบ								
ข้อมูลรูปภาพในสภาพแสงปกติ	✓	✓	✓	✓	✓	✓	✓	✓
ข้อมูลรูปภาพในสภาพแสงน้อย					✓	✓		✓

บทที่ 3
วิธีดำเนินงานวิจัย

วัตถุประสงค์ของงานวิจัยนี้คือการพัฒนาระบบตรวจจับบุคคลในพื้นที่เสี่ยงอันตรายเพื่อป้องกันอันตรายที่อาจเกิดขึ้นได้จากการทำงานของทรัคดัมป์ โดยประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกมาพัฒนาระบบตรวจจับบุคคลพร้อมกับการประเมินความเสี่ยงจากตำแหน่งของบุคคล ซึ่งมีการวิจัยเปรียบเทียบประสิทธิภาพของโมเดลในโครงสร้างแบบต่าง ๆ เพื่อคัดเลือกโมเดลการตรวจจับวัตถุ (Object Detection) ที่เหมาะสม งานวิจัยนี้ได้ใช้โครงสร้างแบบ Faster R-CNN และ YOLO (You Only Look Once) เพื่อตรวจจับบุคคลและส่งตำแหน่งของบุคคลที่ตรวจจับได้มาประเมินความเสี่ยงโดยเปรียบเทียบกับพื้นที่อันตรายที่ถูกกำหนดไว้ เพื่อส่งการแจ้งเตือนไปยังระบบควบคุมเครื่องจักรรวมถึงสามารถหยุดการทำงานของเครื่องจักรได้ทันที เพื่อให้การดำเนินงานวิจัยสอดคล้องตามวัตถุประสงค์ดังกล่าว ในบทนี้จึงเป็นการนำเสนอในส่วนของ กรอบแนวคิดงานวิจัย วิธีดำเนินงานวิจัย และเครื่องมือที่ใช้สำหรับการวิจัย ตามลำดับ

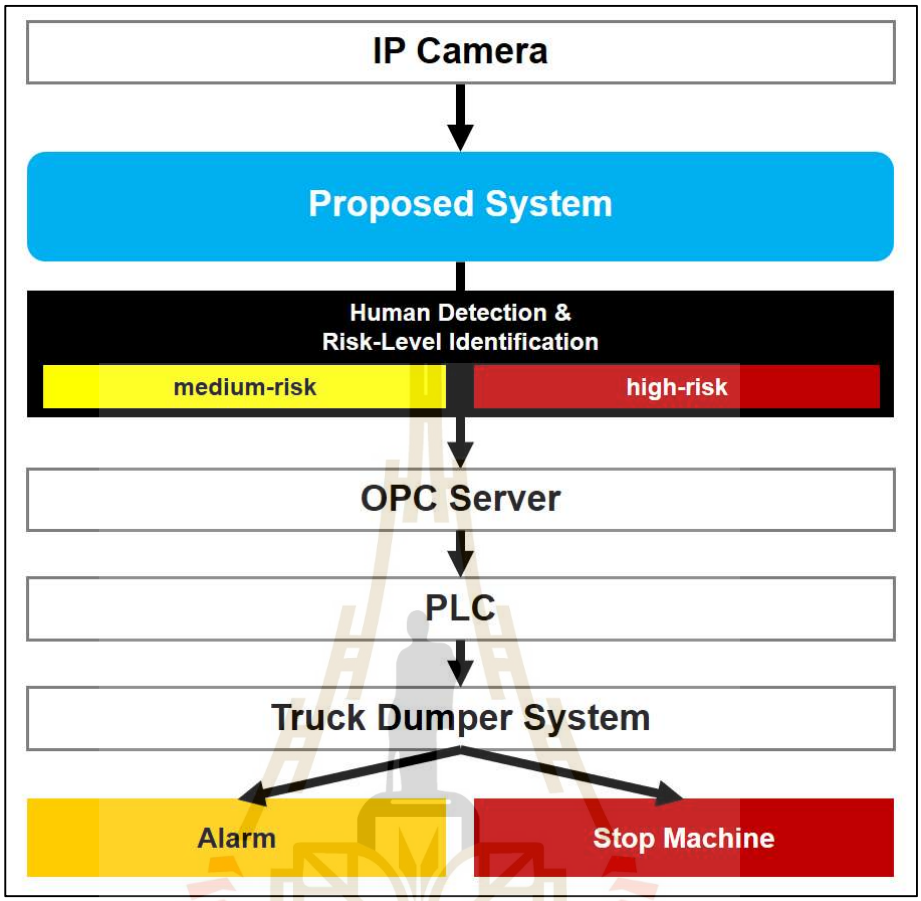
3.1 กรอบแนวคิดงานวิจัย

กรอบแนวคิดของงานวิจัยเพื่อพัฒนาระบบตรวจจับบุคคลและระบุความเสี่ยงสำหรับทำงานร่วมกับระบบควบคุมเครื่องจักรภายในโรงงาน ได้ถูกออกแบบดังรูปที่ 3.1 โดยมีขั้นตอนการทำงาน ดังนี้

1) ระบบรับข้อมูลภาพแบบ Real-time จากกล้อง IP Camera เพื่อส่งให้กับระบบตรวจจับบุคคลและระบุความเสี่ยง

2) ระบบตรวจจับบุคคลและระบุความเสี่ยง แบ่งการทำงานออกเป็น 2 ส่วน คือ ระบบตรวจจับบุคคลใช้ตรวจจับบุคลลภายในกรอบพื้นที่ที่กำหนดไว้ เพื่อหาตำแหน่งล้อมรอบวัตถุ แล้วส่งข้อมูลต่อไปยังส่วนที่ 2 คือ ระบบระบุความเสี่ยง โดยทำหน้าที่ประเมินความเสี่ยงอันตรายจากตำแหน่งของบุคคลที่ตรวจจับได้ และทำการระบุระดับความเสี่ยงเพื่อส่งจำนวนคนในเขตพื้นที่อันตรายให้ระบบควบคุมเครื่องจักรต่อไป

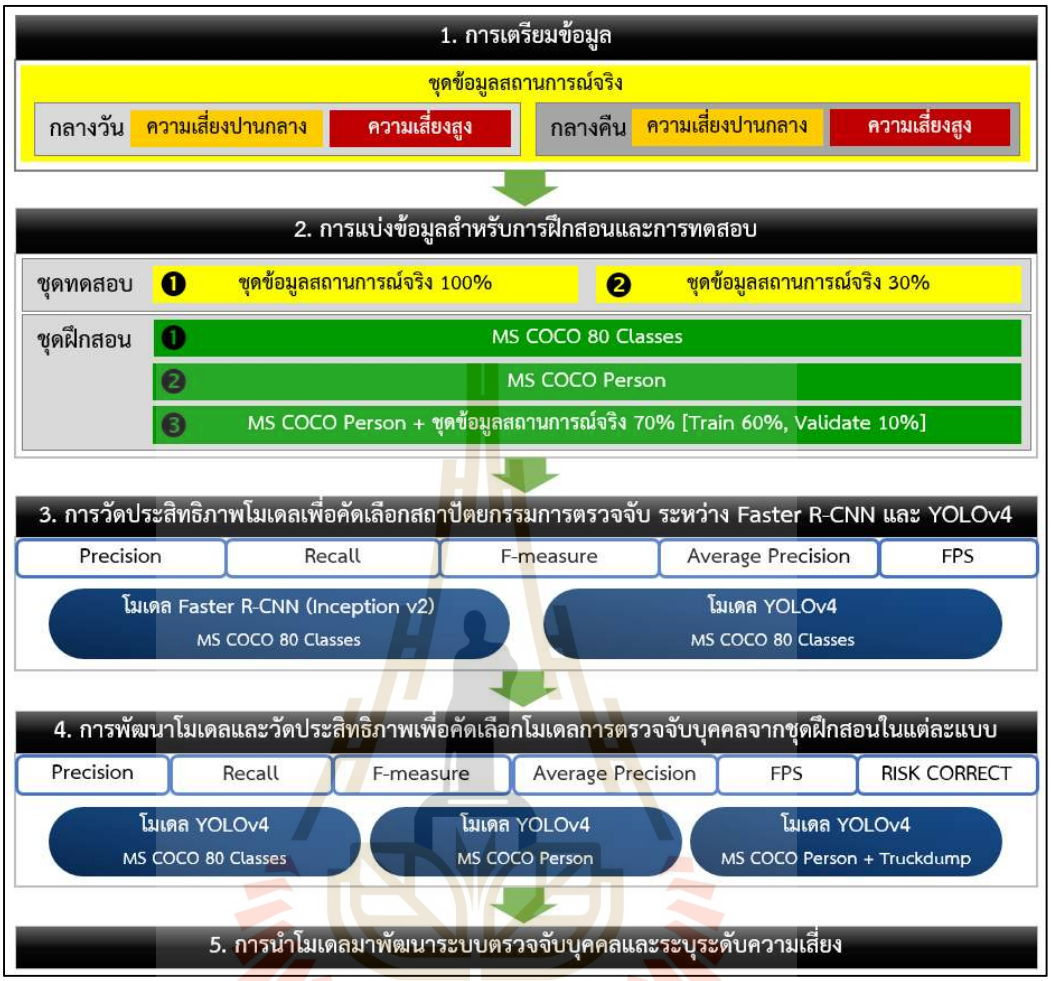
3) ระบบควบคุมเครื่องจักรรับข้อมูลจากระบบตรวจจับบุคคลและระบุความเสี่ยง เพื่อใช้ในการทำงานของเงื่อนไขสำหรับการแจ้งเตือนไปยังหน้าจอควบคุม และเงื่อนไขสำหรับการหยุดการทำงานของเครื่องจักรเมื่อมีคนเข้ามายังพื้นที่อันตราย



รูปที่ 3.1 กรอบแนวคิดงานวิจัย

3.2 กระบวนการวิจัย

ในงานวิจัยนี้ผู้วิจัยได้แบ่งกระบวนการวิจัยออกเป็น 5 ขั้นตอนดังนี้ 1) การเตรียมข้อมูล 2) การแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ 3) การวัดประสิทธิภาพโมเดลเพื่อคัดเลือกสถาปัตยกรรมการตรวจจับ ระหว่าง Faster R-CNN และ YOLOv4 4) การพัฒนาโมเดลและวัดประสิทธิภาพเพื่อคัดเลือกโมเดลการตรวจจับบุคคลจากชุดฝึกสอนในแต่ละแบบ 5) การนำโมเดลมาพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยง ดังแสดงในรูปที่ 3.2โดยในแต่ละขั้นตอนมีรายละเอียด ดังนี้



รูปที่ 3.2 แผนภาพแสดงขั้นตอนการวิจัยทั้ง 5 ขั้นตอน

3.2.1 การเตรียมข้อมูล (ขั้นตอนที่ 1)

ในขั้นตอนการเตรียมข้อมูลสำหรับการวิจัยผู้วิจัยได้เตรียมข้อมูลภาพบุคคลจากสถานการณ์จริง โดยมีขั้นตอนทั้งหมดดังนี้

1) การคัดเลือกภาพจากคลิปวิดีโอ

ในขั้นตอนนี้ผู้วิจัยได้เตรียมข้อมูลโดยนำข้อมูลมาจากคลิปวิดีโอที่บันทึกไว้ ณ สถานที่ปฏิบัติงานจริงที่ถูกติดตั้งไว้ด้านหน้าของช่องทางเทวัตถุดิบ โดยเป็นกล้องมุมสูงแบบไม่สามารถปรับมุมกล้องได้แต่สามารถมองเห็นขอบเขตพื้นที่บริเวณทรัคดัมป์ได้อย่างทั่วถึง คลิปวิดีโอได้ถูกบันทึกอย่างต่อเนื่องตลอดทั้งวัน เป็นระยะเวลา 10 วัน เพื่อสะดวกต่อการวัดประสิทธิภาพของระบบการตรวจจับบุคคลและระบบระบุความเสี่ยง ผู้วิจัยจึงได้ทำการคัดเลือก

เหตุการณ์ภายในวิดีโอเฉพาะช่วงที่มีบุคคลปรากฏอยู่ภายในบริเวณพื้นที่อันตรายทั้ง 2 ระดับ คือ ความเสี่ยงปานกลางและความเสี่ยงสูง โดยเหตุการณ์ที่มีความเสี่ยงปานกลางคัดเลือกจากตำแหน่งการยืนของบุคคลบริเวณใกล้เคียงแท่นทรัคดัมป์และความเสี่ยงสูงคัดเลือกจากตำแหน่งการยืนของบุคคลบนแท่นทรัคดัมป์ ทั้งนี้ผู้วิจัยจะทำการดึงข้อมูลภาพออกจากวิดีโอทุก ๆ เฟรมในเหตุการณ์ที่ถูกเลือก และเพื่อเป็นการวัดความทนทานต่อสภาพแสงที่เปลี่ยนไป ผู้วิจัยจึงได้เลือกเหตุการณ์ทั้งช่วงเวลากลางวันและกลางคืนมาใช้ในงานวิจัยนี้

2) การแยกกลุ่มข้อมูล

เมื่อได้ข้อมูลรูปภาพที่มีบุคคลยืนอยู่อย่างน้อย 1 คน ภายในพื้นที่ระดับความเสี่ยงปานกลางและความเสี่ยงสูงแล้ว ผู้วิจัยได้นำข้อมูลรูปภาพมาคัดแยกเป็นกลุ่มข้อมูลเพื่อใช้สำหรับทดสอบระบบโดยมีการแบ่งข้อมูล ดังนี้

2.1) ข้อมูลรูปภาพในสถานที่ปฏิบัติงานจริงในช่วงเวลากลางวัน ดังรูปที่ 3.3

ก) มีบุคคลอยู่ในพื้นที่ระดับความเสี่ยงปานกลาง

ข) มีบุคคลอยู่ในพื้นที่ระดับความเสี่ยงสูง

2.2) ข้อมูลรูปภาพในสถานที่ปฏิบัติงานจริงในช่วงเวลากลางคืน ดังรูปที่ 3.4

ก) มีบุคคลอยู่ในพื้นที่ระดับความเสี่ยงปานกลาง

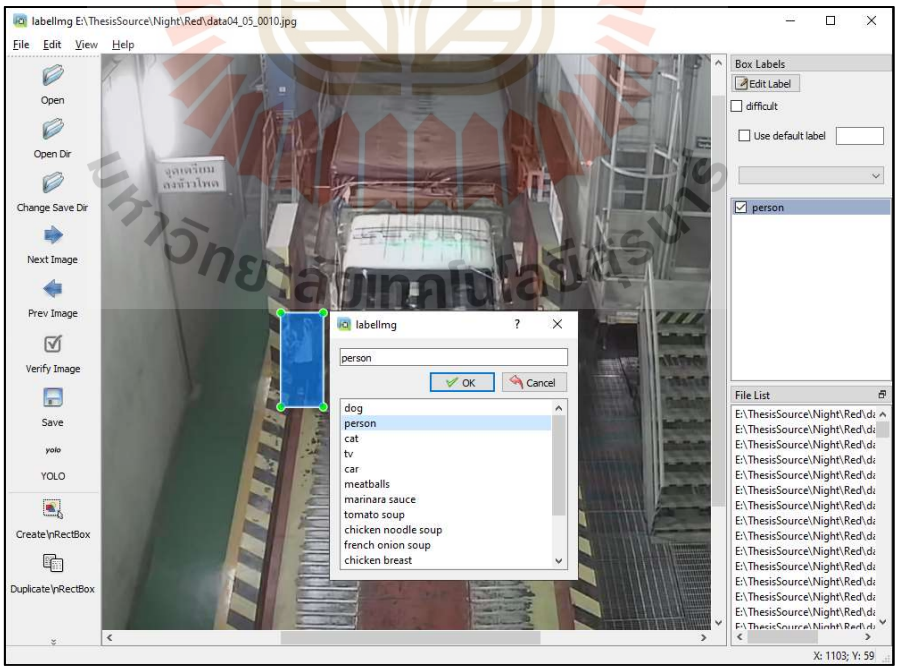
ข) มีบุคคลอยู่ในพื้นที่ระดับความเสี่ยงสูง

รูปที่ 3.3 ตัวอย่างข้อมูลรูปภาพช่วงเวลากลางวันโดยแบ่งตาม
ระดับความเสี่ยงปานกลาง (ซ้าย) และระดับความเสี่ยงสูง (ขวา)

รูปที่ 3.4 ตัวอย่างข้อมูลรูปภาพช่วงเวลากลางคืนโดยแบ่งตาม
ระดับความเสี่ยงปานกลาง (ซ้าย) และระดับความเสี่ยงสูง (ขวา)

3) การกำหนดขอบเขตบุคคลให้กับข้อมูล

เนื่องจากข้อมูลรูปภาพที่เตรียมสำหรับงานวิจัยในขั้นตอนต่อไปยังไม่มีการระบุตำแหน่งของบุคคลภายในรูปภาพ ผู้วิจัยจึงต้องทำการกำหนดขอบเขตของบุคคลภายในภาพเพื่อใช้เป็นคำตอบ (Ground Truth) สำหรับการวัดประสิทธิภาพของระบบตรวจจับบุคคล โดยผู้วิจัยได้ใช้ซอฟต์แวร์ LabelImg สำหรับการกำหนดขอบเขตวัตถุภายในรูปภาพ ดังรูปที่ 3.5



รูปที่ 3.5 ตัวอย่างการกำหนดขอบเขตของบุคคลภายในรูปภาพ

3.2.2 การแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ (ขั้นตอนที่ 2)

หลังจากเตรียมข้อมูลเรียบร้อยแล้วผู้วิจัยได้มีการแบ่งสัดส่วนข้อมูลสำหรับฝึกสอนและทดสอบ โดยมีขั้นตอนทั้งหมดดังนี้

1) การแบ่งข้อมูลสำหรับการทดสอบ

ข้อมูลที่ใช้ในการทดสอบแบ่งออกเป็น 2 ส่วน ดังตารางที่ 3.1 โดยส่วนแรกใช้สำหรับทดสอบประสิทธิภาพการตรวจจับบุคคลระหว่างโมเดลที่สร้างด้วยสถาปัตยกรรม Faster R-CNN และ YOLOv4 ในส่วนที่ 2 ใช้สำหรับทดสอบประสิทธิภาพการตรวจจับบุคคลและความถูกต้องในการระบุความเสี่ยงจากโมเดลที่ผู้วิจัยพัฒนาขึ้นด้วยสถาปัตยกรรมที่ถูกเลือกใช้จากส่วนแรก โดยมีรายละเอียดดังนี้

ชุดข้อมูลทดสอบส่วนที่ 1 ใช้ชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด โดยแบ่งเป็นภาพในเวลากลางวัน 5,043 ภาพ และภาพในเวลากลางคืน 4,068 ภาพ มีการกำหนดขอบเขตบุคคลไว้ภาพละ 1 กรอบ

ชุดข้อมูลทดสอบส่วนที่ 2 ใช้ชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด โดยคัดเลือกเหตุการณ์ที่แตกต่างกันด้วยสัดส่วน ชุดฝึกสอน 60% ชุดตรวจสอบ 10% และชุดทดสอบ 30% จากชุดข้อมูลทั้งหมด แบ่งเป็นภาพในเวลากลางวัน ความเสี่ยงสูง 685 ภาพ ความเสี่ยงปานกลาง 828 ภาพ และภาพในเวลากลางคืน ความเสี่ยงสูง 644 ภาพ ความเสี่ยงปานกลาง 576 ภาพ มีการกำหนดขอบเขตบุคคลไว้ภาพละ 1 กรอบ พร้อมกับกำหนดความเสี่ยงเพื่อเป็นคำตอบไว้ในแต่ละกรอบ

2) การแบ่งข้อมูลสำหรับการฝึกสอน

ข้อมูลที่ใช้ในการฝึกสอนแบ่งออกเป็น 3 ชุด ดังนี้

ชุดฝึกสอนที่ 1 ใช้ชุดข้อมูล MS COCO ทั้งหมด โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท

ชุดฝึกสอนที่ 2 ใช้ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล โดยแบ่งเป็นชุดฝึกสอน 64,115 ภาพ มีการกำหนดขอบเขตบุคคลไว้ทั้งหมด 257,252 กรอบ และชุดตรวจสอบ 2,693 ภาพ มีการกำหนดขอบเขตบุคคลไว้ทั้งหมด 10,777 กรอบ

ชุดฝึกสอนที่ 3 ใช้ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล 100% ของชุดข้อมูลทั้งหมด รวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด โดยแบ่งเป็นชุดฝึกสอนทั้งหมด 60% ของชุดข้อมูลทั้งหมด มีจำนวน 5,466 ภาพ จากการแบ่งออกมาจากกลุ่มภาพในเวลากลางวัน ความเสี่ยงสูง 1,370 ภาพ ความเสี่ยงปานกลาง 1,655 ภาพ และภาพในเวลากลางคืน ความเสี่ยงสูง 1,289 ภาพ ความเสี่ยงปานกลาง 1,152 ภาพ และชุดตรวจสอบ 10% ของชุดข้อมูลทั้งหมด มีจำนวน 912 ภาพ จากการแบ่งออกมาจากกลุ่มภาพในเวลา

กลางวัน ความเสี่ยงสูง 229 ภาพ ความเสี่ยงปานกลาง 276 ภาพ และภาพในเวลากลางคืน ความเสี่ยงสูง 215 ภาพ ความเสี่ยงปานกลาง 192 ภาพ เมื่อรวมข้อมูลทั้งหมดแล้วจะมีข้อมูลชุดฝึกสอนทั้งหมด 262,718 กรอบ และชุดตรวจสอบ 11,689 กรอบ

ตารางที่ 3.1 สรุปสัดส่วนการแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ

MODEL	TRAIN SET			TEST SET
	MS COCO 80 Classes	MS COCO Person	Truckdump	Truckdump
MODEL 1.1	100%	-	-	100%
MODEL 1.2	100%	-	-	100%
MODEL 2.1	100%	-	-	30%
MODEL 2.2	-	100%	-	30%
MODEL 2.3	-	100%	70%	30%

3.2.3 การวัดประสิทธิภาพโมเดลเพื่อคัดเลือกสถาปัตยกรรมการตรวจจับ ระหว่าง Faster R-CNN และ YOLOv4 (ขั้นตอนที่ 3)

ในขั้นตอนนี้มีวัตถุประสงค์เพื่อวัดประสิทธิภาพโดยรวมของการตรวจจับบุคคลระหว่างโมเดลที่พัฒนาจากสถาปัตยกรรม Faster R-CNN และ YOLOv4 โดยผู้วิจัยจะคัดเลือกสถาปัตยกรรมการตรวจจับวัตถุที่เหมาะสมที่สุดกับงานวิจัยนี้ โดยทำการทดสอบประสิทธิภาพของโมเดลด้วยชุดข้อมูลทดสอบส่วนที่ 1 คือ ชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 100% โดยใช้มาตรวัดประสิทธิภาพดังนี้ 1) ค่าความเที่ยงตรง 2) ค่าการระลึก 3) ค่าประสิทธิภาพโดยรวม 4) ค่าความเที่ยงตรงเฉลี่ย และ 5) ความเร็วในการประมวลผลภาพต่อวินาที โดยมีการกำหนดค่าเกณฑ์ IoU ที่ใช้สำหรับการตรวจจับวัตถุไว้ที่ 0.5 กับทุกโมเดล

3.2.4 การพัฒนาโมเดลและวัดประสิทธิภาพเพื่อคัดเลือกโมเดลการตรวจจับบุคคลจากชุดฝึกสอนในแต่ละแบบ (ขั้นตอนที่ 4)

ในขั้นตอนนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลและวัดประสิทธิภาพโดยรวมของการตรวจจับบุคคลและระบุระดับความเสี่ยงของบุคคล ระหว่างโมเดลที่พัฒนาจากสถาปัตยกรรมการตรวจจับวัตถุที่ถูกเลือกใช้จากการทดลองในขั้นตอนที่แล้ว โดยผู้วิจัยจะพัฒนาโมเดลด้วยชุด

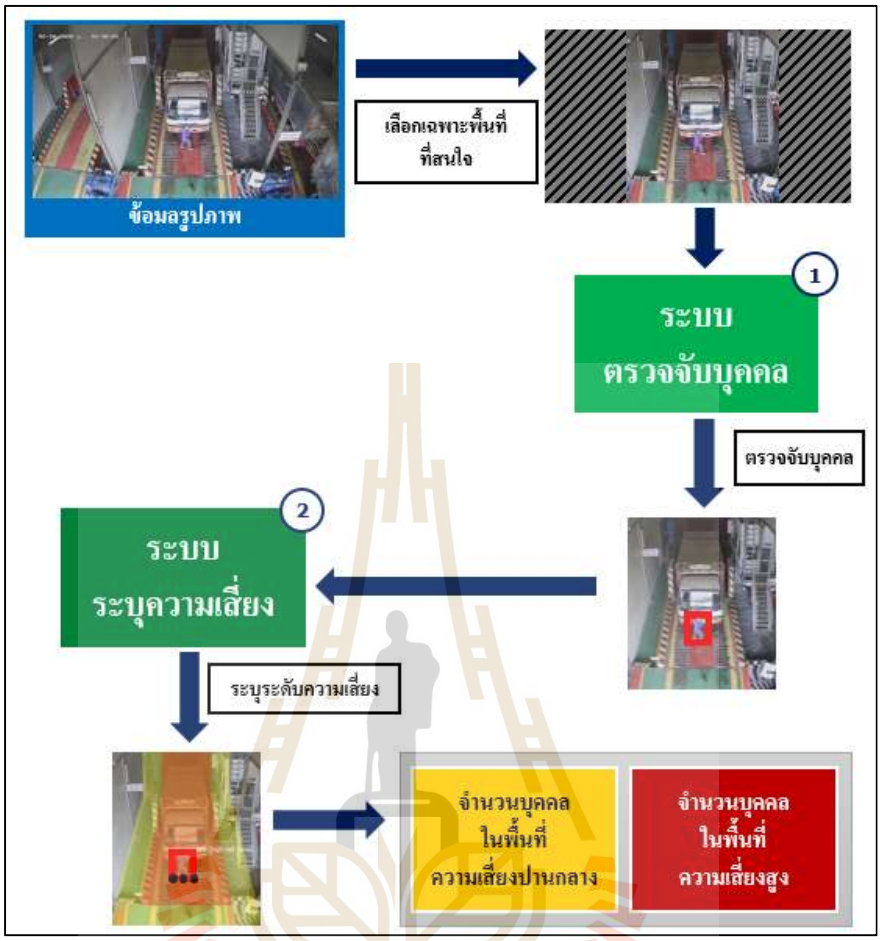
ข้อมูลที่แตกต่างกันจำนวน 3 ชุด ดังนี้ ชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล และ ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด โดยทำการทดสอบประสิทธิภาพของโมเดลด้วยชุดข้อมูลทดสอบส่วนที่ 2 คือ ชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด โดยใช้มาตรวัดประสิทธิภาพดังนี้ 1) ค่าความเที่ยงตรง 2) ค่าการระลึก 3) ค่าประสิทธิภาพโดยรวม 4) ค่าความเที่ยงตรงเฉลี่ย 5) ความเร็วในการประมวลผลภาพต่อวินาที และ 6) ค่าความถูกต้องในการระบุความเสี่ยงเมื่อเปรียบเทียบคำตอบกับชุดข้อมูลจริงทั้งหมด โดยมีการกำหนดค่าเกณฑ์ IoU ที่ใช้สำหรับการตรวจจับวัตถุไว้ที่ 0.5 กับทุกโมเดล

3.2.5 การนำโมเดลมาพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยงของบุคคล (ขั้นตอนที่ 5)

ในการพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยงนั้นมีจุดประสงค์หลักเพื่อรับข้อมูลภาพที่ถูกเตรียมไว้มาผ่านระบบเพื่อตรวจหาตำแหน่งของบุคคลภายในรูปภาพ และทำการระบุความเสี่ยงจากตำแหน่งที่บุคคลปรากฏอยู่ สุดท้ายระบบจะส่งผลลัพธ์ออกมาเป็นจำนวนบุคคลที่อยู่ภายในพื้นที่บริเวณความเสี่ยงปานกลางและความเสี่ยงสูง ดังรูปที่ 3.6 โดยกระบวนการสำหรับการพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยง จะถูกแบ่งออกเป็นระบบย่อย 2 ขั้นตอน คือ การพัฒนาระบบตรวจจับบุคคล และการพัฒนาระบบระบุความเสี่ยง ซึ่งมีรายละเอียดดังนี้

1) การพัฒนาระบบตรวจจับบุคคล

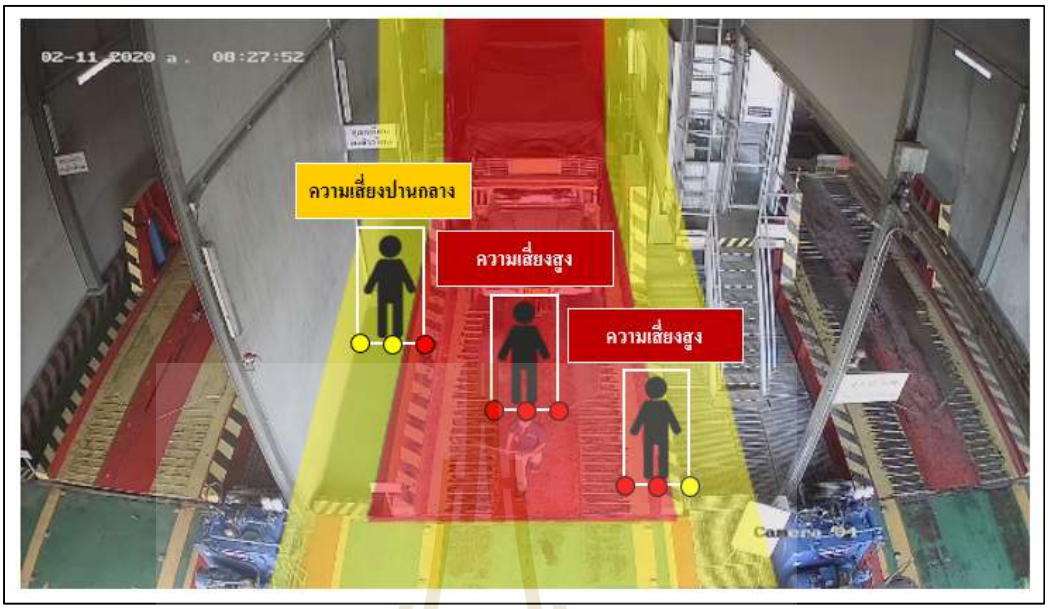
จุดประสงค์หลักของระบบนี้คือการพัฒนาระบบที่สามารถหาตำแหน่งบุคคลภายในรูปภาพได้ ในขั้นตอนนี้ผู้วิจัยจะนำข้อมูลที่เตรียมไว้มาทำการทดสอบเพื่อคัดเลือกโมเดลในโครงสร้างแบบ Faster R-CNN (Inception v2) และ YOLOv4 ซึ่งทั้งสองโมเดลนี้ได้ถูกฝึกสอนด้วยชุดข้อมูล MS COCO (Microsoft Common Objects in Context) (Lin et al., 2014) ซึ่งเป็นชุดข้อมูลขนาดใหญ่ มีจำนวนภาพมากกว่า 330,000 รูปภาพ มีจำนวนประเภทของวัตถุทั้งหมด 80 ประเภท โดยในงานวิจัยนี้จะพิจารณาเฉพาะรูปภาพประเภทบุคคล (Person) โดยผู้วิจัยจะทำการทดสอบประสิทธิภาพของโมเดลทั้ง 2 แบบ และคัดเลือกโมเดลที่เหมาะสมมาใช้ในการพัฒนาระบบตรวจจับบุคคล โดยใช้ผลลัพธ์ที่ได้จากการทำนายรูปภาพประเภทบุคคลและหาประสิทธิภาพที่เหมาะสมที่สุดกับงานวิจัยนี้ โดยคำนึงถึงค่าความถูกต้องและความเร็วในการประมวลผล



รูปที่ 3.6 ภาพรวมการทำงานของระบบตรวจจับบุคคลและระบุระดับความเสี่ยง

2) การพัฒนาระบบระบุความเสี่ยง

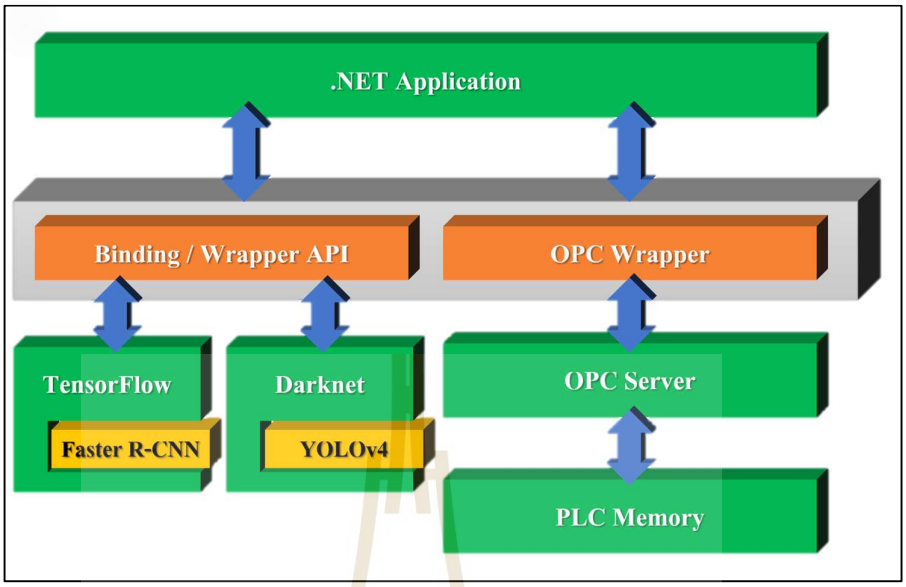
ในขั้นตอนแรกของระบบระบุความเสี่ยง จะต้องมีการกำหนดพื้นที่ของอาณาเขตในระดับความเสี่ยงปานกลางและความเสี่ยงสูงไว้ในการใช้งานครั้งแรก เมื่อระบบได้รับข้อมูลตำแหน่งกรอบล้อมรอบวัตถุ (Bounding Boxes) จากระบบการตรวจจับบุคคล ระบบจะประเมินความเสี่ยงจากตำแหน่งของบุคคลที่ตรวจจับได้ โดยมีการกำหนดจุด 3 จุดไว้ด้านล่างกรอบของวัตถุ ดังนี้ 1) มุมซ้ายล่างของกรอบ 2) จุดกลางของเส้นใต้กรอบ 3) มุมขวาล่างของกรอบ จากนั้นระบบจะคำนวณว่าจุดส่วนใหญ่อยู่ในเขตพื้นที่อันตรายระดับใดมากที่สุด เพื่อเป็นระบุระดับความเสี่ยงอันตราย ดังแสดงในรูปที่ 3.7 และในขั้นตอนสุดท้ายระบบจะรวบรวมจำนวนคนในเขตพื้นที่อันตรายปานกลางและพื้นที่อันตรายสูงส่งให้กับระบบควบคุมเครื่องจักรต่อไป



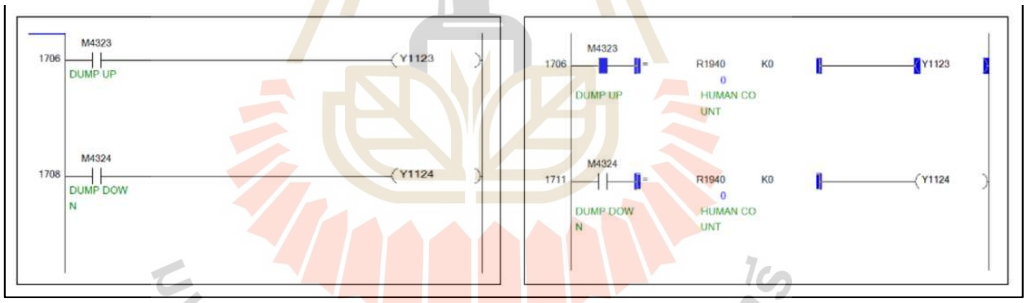
รูปที่ 3.7 กระบวนการระบุระดับความเสี่ยงจากตำแหน่งของบุคคล

3.3 การรวมระบบที่นำเสนอเข้ากับระบบควบคุมเครื่องจักร

ในการทำงานของระบบตรวจจับบุคคลและระบุระดับความเสี่ยงในสถานการณ์จริง สามารถนำระบบในงานวิจัยนี้ไปติดตั้งให้ทำงานร่วมกับระบบควบคุมเครื่องจักรได้โดยตรง โดยระบบสามารถส่งข้อมูลจำนวนบุคคลที่อยู่ในพื้นที่ระดับความเสี่ยงปานกลางและจำนวนบุคคลที่อยู่ในพื้นที่ระดับความเสี่ยงสูงไปยังหน่วยความจำของโปรแกรมพีแอลซี (PLC) ผ่านซอฟต์แวร์ OPC Server ดังแสดงในรูปที่ 3.8 โดยภายในพีแอลซีต้องกำหนดหน่วยความจำสำหรับรับตัวแปรทั้ง 2 ตัว คือจำนวนบุคคลระดับความเสี่ยงปานกลางและความเสี่ยงสูงไว้ล่วงหน้า โดยในส่วนการแจ้งเตือนที่หน้าจอระบบควบคุม ต้องมีการกำหนดเงื่อนไขในการแสดงหน้าจอแจ้งเตือนเพิ่มเติมโดยคิดเงื่อนไขจากตัวแปรที่เก็บจำนวนความเสี่ยงปานกลาง ในส่วนของการหยุดเครื่องจักรสามารถเพิ่มเงื่อนไขในการส่งสัญญาณขาออกโดยมีการกำหนดด้วยตัวแปรจำนวนระดับความเสี่ยงสูง ดังแสดงในรูปที่ 3.9



รูปที่ 3.8 แผนผังการรับส่งข้อมูลระหว่างระบบที่นำเสนอและระบบควบคุมเครื่องจักร



รูปที่ 3.9 การกำหนดเงื่อนไขการส่งสัญญาณออกจากโปรแกรม PLC ไปยังเครื่องจักร
โปรแกรมก่อนกำหนดเงื่อนไข (ซ้าย) และโปรแกรมหลังกำหนดเงื่อนไข (ขวา)

3.4 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในงานวิจัยและพัฒนาระบบ ประกอบด้วย

1) เครื่องคอมพิวเตอร์สำหรับพัฒนาระบบและทดสอบประสิทธิภาพ

- หน่วยประมวลผลกลาง : Inter® Core™ i7-9750H 2.60GHz
- หน่วยประมวลผลกราฟิก : NVIDIA Geforce RTX 2070 with Max-Q Design
- หน่วยความจำหลัก : 16 GB

2) ระบบปฏิบัติการและโปรแกรมประยุกต์สำหรับพัฒนาระบบ ประกอบด้วย

- ระบบปฏิบัติการ : Windows 10 Pro
- ซอฟต์แวร์ช่วยประมวลผล : NVIDIA CUDA 10.1 and cuDNN SDK 7.6
- ซอฟต์แวร์เฟรมเวิร์ค : TensorFlow-GPU 2.3.1, Darknet
- ซอฟต์แวร์ไลบรารี : OpenCV 4.3.0
- ซอฟต์แวร์สำหรับพัฒนาโปรแกรม : Visual Studio 2019 Community

บทที่ 4
การทดสอบและอภิปรายผล

ในการทดสอบประสิทธิภาพของโมเดลจะแบ่งออกเป็น 2 ส่วน คือ การวัดประสิทธิภาพของโมเดลจากสถาปัตยกรรมการตรวจจับวัตถุที่แตกต่างกันระหว่าง Faster R-CNN และ YOLOv4 เพื่อคัดเลือกสถาปัตยกรรมการตรวจจับวัตถุที่เหมาะสมทั้งทางด้านความแม่นยำและความเร็วในการประมวลผลมาพัฒนาโมเดลสำหรับการตรวจจับบุคคล โดยในส่วนที่ 2 จะทำการพัฒนาโมเดลและวัดประสิทธิภาพของโมเดลเพื่อคัดเลือกโมเดลที่มีความแม่นยำที่สุดทั้งด้านการตรวจจับบุคคลและความถูกต้องในการระบุระดับความเสี่ยงมาพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์ โดยใช้เกณฑ์การพิจารณาจากจำนวนกรอบล้อมรอบวัตถุที่สร้างถูกต้อง (True Positive : TP) จำนวนกรอบล้อมรอบวัตถุที่สร้างไม่ถูกต้อง (False Positive : FP) และจำนวนกรอบล้อมรอบวัตถุที่ไม่ถูกสร้าง (False Negative : FN) มาคำนวณหาค่าความเที่ยงตรง (Precision) ค่าการระลึก (Recall) ค่าประสิทธิภาพโดยรวม (F-measure : F1) และค่าความเที่ยงตรงเฉลี่ย (Average Precision : AP) รวมถึงความเร็วในการประมวลผลภาพต่อวินาที (Frame per Second : FPS) และในการวัดความถูกต้องในการระบุความเสี่ยงบุคคลจะพิจารณาจากจำนวนบุคคลที่สามารถระบุได้ถูกต้องเมื่อเปรียบเทียบกับข้อมูลจริง (Ground Truth : GT) ที่นำมาทดสอบทั้งหมด

4.1 ข้อมูลที่ใช้ในการทดสอบ

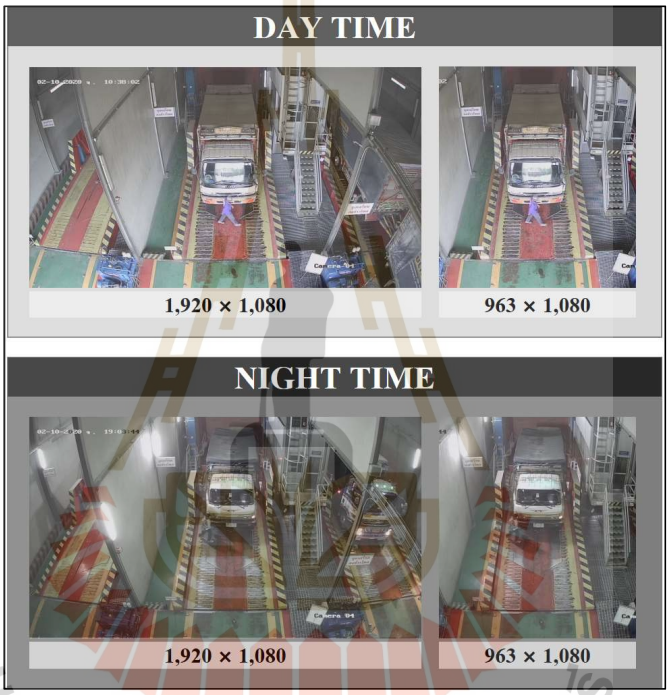
ข้อมูลที่ใช้ในการทดสอบโมเดลทั้ง 2 ส่วน จะใช้ข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริงในพื้นที่ความเสี่ยงสูงและความเสี่ยงปานกลาง ทั้งในเวลากลางวันและเวลากลางคืน โดยทุกภาพจะมีการเตรียมภาพไว้ 2 ขนาด คือ ภาพต้นฉบับจากกล้องมีขนาดความกว้าง 1,920 × ความสูง 1,080 พิกเซล และ ภาพที่มีการตัดบริเวณที่ไม่จำเป็นออกโดยเลือกเฉพาะพื้นที่ของทรัคดัมป์ที่กล้องติดตั้งอยู่โดยเลือกจากตำแหน่งพิกัด X=502 จนถึง X=1,465 ในขณะที่ความสูงคงเดิม ดังนั้นภาพที่ถูกลดขนาดลงจะเหลือขนาดความกว้าง 963 × ความสูง 1,080 พิกเซล ทั้งนี้จะมีการแบ่งกลุ่มข้อมูลที่แตกต่างกันและมี 2 ชุดข้อมูลสำหรับการทดสอบ ดังนี้

4.1.1 ชุดข้อมูลสำหรับทดสอบสถาปัตยกรรมการตรวจจับวัตถุ

ในการทดสอบสถาปัตยกรรมการตรวจจับวัตถุจะใช้ภาพบุคคลจากสถานที่ปฏิบัติงานจริงทั้งหมด 100% ของชุดข้อมูลทั้งหมด โดยแบ่งข้อมูลออกเป็น 2 กลุ่ม ดังรูปที่ 4.1 โดยมีรายละเอียดดังต่อไปนี้

1) ภาพบุคคลในเวลากลางวัน จำนวน 5,043 ภาพ

2) ภาพบุคคลในเวลากลางคืน จำนวน 4,068 ภาพ



รูปที่ 4.1 ตัวอย่างรูปภาพชุดทดสอบจำนวน 2 กลุ่ม แบ่งภาพออกเป็น 2 ขนาด

4.1.2 ชุดข้อมูลสำหรับทดสอบโมเดลการตรวจจับบุคคลและระบุระดับความเสี่ยง

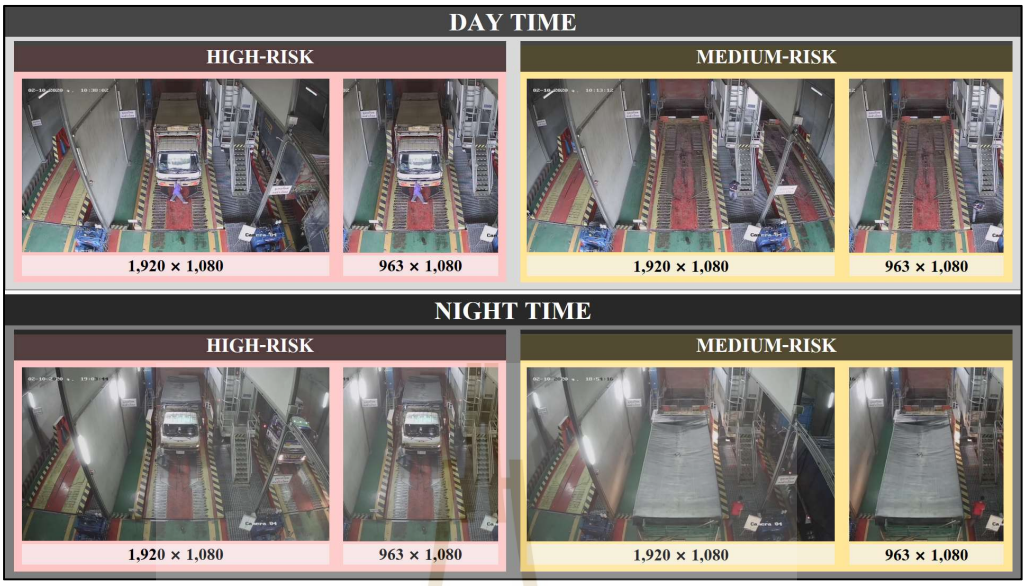
ในการทดสอบโมเดลที่พัฒนาขึ้นจากสถาปัตยกรรมที่ถูกคัดเลือกแล้วในข้อ 4.1.1 จะใช้ภาพบุคคลจากสถานที่ปฏิบัติงานจริงทั้งหมด 30% ของชุดข้อมูลทั้งหมด โดยแบ่งข้อมูลออกเป็น 4 กลุ่ม ดังรูปที่ 4.2 โดยมีรายละเอียดดังต่อไปนี้

1) ภาพบุคคลในเวลากลางวัน ความเสี่ยงสูง จำนวน 685 ภาพ

2) ภาพบุคคลในเวลากลางวัน ความเสี่ยงปานกลาง จำนวน 828 ภาพ

3) ภาพบุคคลในเวลากลางคืน ความเสี่ยงสูง จำนวน 644 ภาพ

4) ภาพบุคคลในเวลากลางคืน ความเสี่ยงปานกลาง จำนวน 576 ภาพ



รูปที่ 4.2 ตัวอย่างรูปภาพชุดทดสอบจำนวน 4 กลุ่ม แบ่งภาพออกเป็น 2 ขนาด

4.2 ผลการทดสอบประสิทธิภาพสถาปัตยกรรมการตรวจจับวัตถุ

การทดสอบความแม่นยำในการตรวจจับบุคคลรวมถึงความเร็วในการประมวลผลของทั้ง 2 สถาปัตยกรรมการตรวจจับวัตถุจะทดสอบกับโมเดลที่พัฒนาไว้แล้วด้วยฐานข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ในการทดลองจะทดสอบเฉพาะประเภทบุคคลกับข้อมูลสถานที่ปฏิบัติงานจริงทั้งหมด 2 กลุ่ม คือ ภาพบุคคลจากสถานที่ปฏิบัติงานจริงในเวลากลางวันและกลางคืน แบ่งภาพออกเป็น 2 ขนาด คือ 1,920 × 1,080 พิกเซล และ 963 × 1,080 พิกเซล โดยมีการกำหนดค่าเกณฑ์ IoU ไว้ที่ 0.5

4.2.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN

การทดสอบประสิทธิภาพโมเดลที่ถูกพัฒนาจากสถาปัตยกรรม Faster R-CNN โดยใช้โครงข่ายประสาทเทียมเบื้องหลังแบบ Inception v2 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบบน Tensorflow Framework กับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงทั้งหมด 100% ของชุดข้อมมูลทั้งหมด ให้ผลการทดลองดังตารางที่ 4.1 และตารางที่ 4.2

ตารางที่ 4.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด

Faster R-CNN MS COCO 80 Classes	GT	HUMAN DETECTION				
		BOXES	TP	FP	FN	AVG TIME (ms)
DAY [1920 × 1080]	5,043	5,078	2,564	217	2,297	69.8308
DAY [963 × 1080]	5,043	5,103	3,112	230	1,761	58.0252
NIGHT [1920 × 1080]	4,068	4,109	2,055	86	1,968	71.1380
NIGHT [963 × 1080]	4,068	4,159	2,562	128	1,469	59.3144

จากตารางที่ 4.1 จะเห็นได้ว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม Faster R-CNN ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีผลลัพธ์การตรวจจับบุคคลในเวลากลางวันค่อนข้างต่ำเนื่องจากมีกรอบล้อมรอบวัตถุที่สร้างถูกต้องเพียง 2,564 กรอบ คิดเป็น 50.84% และมีกรอบที่ไม่ถูกสร้างจำนวน 2,297 กรอบ คิดเป็น 45.55% นอกจากนี้ยังมีกรอบที่สร้างผิดจำนวน 217 กรอบ คิดเป็น 4.30% จากข้อมูลทดสอบทั้งหมด ในขณะที่ผลลัพธ์เมื่อทดสอบกับภาพบุคคลในเวลากลางคืน โมเดลสามารถสร้างกรอบได้ถูกต้องจำนวน 2,055 กรอบ คิดเป็น 50.52% ไม่สร้างกรอบจำนวน 1,968 คิดเป็น 48.38% และสร้างกรอบผิดจำนวน 86 คิดเป็น 2.11% จากข้อมูลทดสอบทั้งหมด เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดรูปภาพทั้งในเวลากลางวันและกลางคืน สามารถให้ผลลัพธ์ที่ดีขึ้น โดยภาพบุคคลในเวลากลางวันสามารถเพิ่มการสร้างกรอบที่ถูกต้องได้ถึง 10.87% และภาพในเวลากลางคืนสามารถเพิ่มการสร้างกรอบที่ถูกต้องได้ถึง 12.46% แต่ทั้งภาพในเวลากลางวันและกลางคืนหลังจากลดขนาดภาพลงจะส่งผลให้มีการสร้างกรอบที่ผิดพลาดมากขึ้นตามด้วยเล็กน้อย ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 70.41 มิลลิวินาที และ ภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 58.63 มิลลิวินาที

ตารางที่ 4.2 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม Faster R-CNN ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด

Faster R-CNN MS COCO 80 Classes	HUMAN DETECTION				
	AVG FPS	PRECISION	RECALL	F1	AP
DAY [1920 × 1080]	14.3555	0.9220	0.5275	0.6710	51.5875
DAY [963 × 1080]	17.2620	0.9312	0.6386	0.7576	63.0854
NIGHT [1920 × 1080]	14.0888	0.9598	0.5108	0.6668	50.6053
NIGHT [963 × 1080]	16.9047	0.9524	0.6356	0.7624	63.3128

จากตารางที่ 4.2 จะเห็นได้ว่าผลสรุปการตรวจจับบุคคลของสถาปัตยกรรม Faster R-CNN มีค่าความเที่ยงตรงค่อนข้างสูง คือ 0.9220 จากภาพในเวลากลางวันและ 0.9598 จากภาพในเวลากลางคืน แต่เมื่อพิจารณาค่าการระลึกยังพบว่ามีค่าเพียง 0.5275 จากภาพในเวลากลางวันและ 0.6356 จากภาพในเวลากลางคืน ซึ่งส่งผลให้ค่าประสิทธิภาพโดยรวมมีค่าเพียง 0.6710 จากภาพในเวลากลางวันและ 0.6668 จากภาพในเวลากลางคืน และเมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยพบว่าภาพในเวลากลางวันมีค่าเพียง 51.59% และ 50.61% ในเวลากลางคืน ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมทั้งหมดดีขึ้น โดยเฉพาะค่าความเที่ยงตรงเฉลี่ยสูงขึ้นเป็น 63.09% จากภาพในเวลากลางวันและ 63.31% จากภาพในเวลากลางคืน ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพกับภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 14.24 ภาพต่อวินาที และภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 17.10 ภาพต่อวินาที

4.2.2 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4

การทดสอบประสิทธิภาพโมเดลที่ถูกพัฒนาจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงทั้งหมด 100% ของชุดข้อมูลทั้งหมด ให้ผลการทดลองดังตารางที่ 4.3 และ ตารางที่ 4.4

ตารางที่ 4.3 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO 80 Classes	GT	HUMAN DETECTION				
		BOXES	TP	FP	FN	AVG TIME (ms)
DAY [1920 × 1080]	5,043	5,043	2,738	4	2,301	38.2100
DAY [963 × 1080]	5,043	5,043	3,435	3	1,605	31.3117
NIGHT [1920 × 1080]	4,068	4,068	2,108	4	1,956	35.9861
NIGHT [963 × 1080]	4,068	4,072	2,389	11	1,672	31.3050

จากตารางที่ 4.3 จะเห็นได้ว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีผลลัพธ์การตรวจจับบุคคลในเวลากลางวันดีกว่าสถาปัตยกรรม Faster R-CNN เพียงเล็กน้อย โดยโมเดลสร้างกรอบล้อมรอบวัตถุที่สร้างถูกต้องเพียง 2,738 กรอบ คิดเป็น 54.29% และมีกรอบที่ไม่ถูกสร้างจำนวน 2,301 กรอบ คิดเป็น 45.63% แต่พบว่ามีกรอบที่สร้างผิดจำนวนน้อยมากคือ 4 กรอบ คิดเป็น 0.08% จากข้อมูลทดสอบทั้งหมด ในขณะที่ผลลัพธ์เมื่อทดสอบกับภาพบุคคลในเวลากลางคืน โมเดลสามารถสร้างกรอบได้ถูกต้องจำนวน 2,108 กรอบ คิดเป็น 51.82% ไม่สร้างกรอบจำนวน 1,956 คิดเป็น 48.08% และสร้างกรอบผิดจำนวน 4 กรอบ คิดเป็น 0.10% จากข้อมูลทดสอบทั้งหมด เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดรูปภาพทั้งในเวลากลางวันและกลางคืน สามารถให้ผลลัพธ์ที่ดีขึ้น โดยภาพบุคคลในเวลากลางวันสามารถเพิ่มการสร้างกรอบที่ถูกต้องได้ถึง 13.82% และภาพในเวลากลางคืนสามารถเพิ่มการสร้างกรอบที่ถูกต้องได้ 6.91% นอกจากนี้ยังสามารถลดจำนวนกรอบที่ไม่ถูกสร้างและจำนวนกรอบที่สร้างผิดลงได้ ยกเว้นภาพในเวลากลางคืนมีจำนวนกรอบที่สร้างผิดเพิ่มขึ้นเพียงเล็กน้อย คือ 7 กรอบ ทั้งนี้โมเดลใช้เวลาในการประมวลผลเฉลี่ยเพียง 37.22 มิลลิวินาทีกับภาพขนาด 1,920 × 1,080 พิกเซล และ 31.31 มิลลิวินาที กับภาพขนาด 963 × 1,080 พิกเซล

ตารางที่ 4.4 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO 80 Classes	HUMAN DETECTION				
	AVG FPS	PRECISION	RECALL	F1	AP
DAY [1920 × 1080]	26.4409	0.9985	0.5434	0.7038	54.3163
DAY [963 × 1080]	31.9617	0.9991	0.6815	0.8103	68.1349
NIGHT [1920 × 1080]	28.0942	0.9981	0.5187	0.6826	51.8455
NIGHT [963 × 1080]	31.9679	0.9954	0.5883	0.7395	58.8033

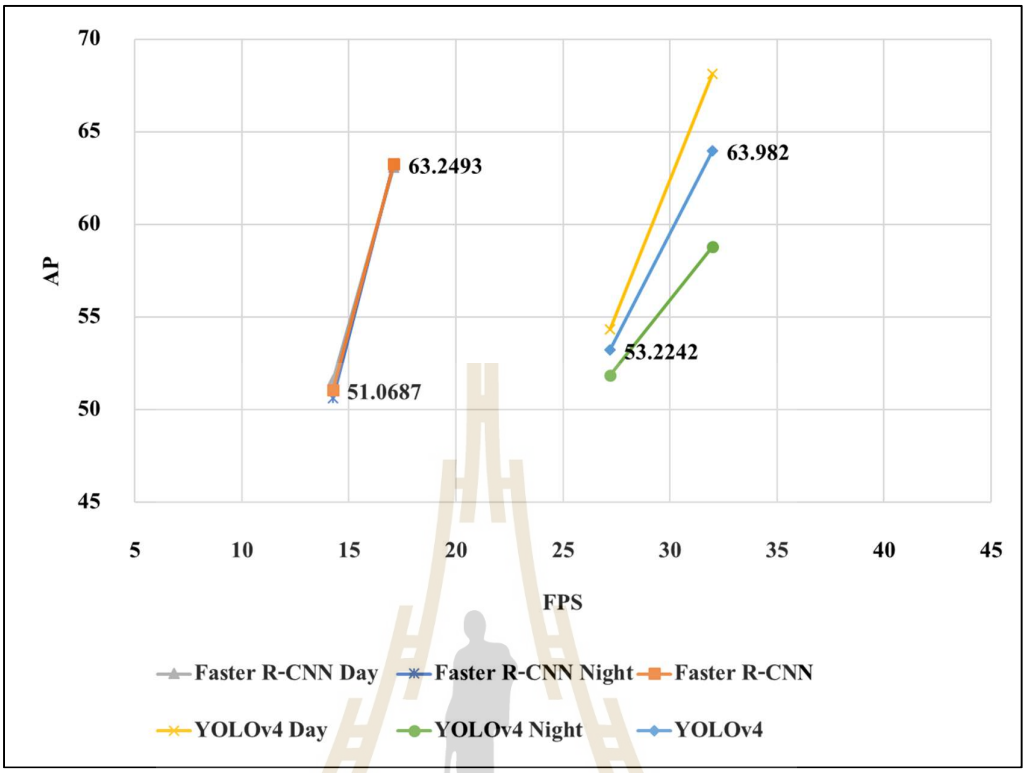
จากตารางที่ 4.4 จะเห็นได้ว่าผลสรุปการตรวจจับบุคคลของสถาปัตยกรรม YOLOv4 มีค่าความเที่ยงตรงค่อนข้างสูงมาก คือ 0.9985 จากภาพในเวลากลางวันและ 0.9981 จากภาพในเวลากลางคืน แต่เมื่อพิจารณาค่าการระลึกยังพบว่ามีค่าต่ำที่ 0.5434 จากภาพในเวลากลางวันและ 0.5187 จากภาพในเวลากลางคืน ซึ่งส่งผลให้ค่าประสิทธิภาพโดยรวมมีค่าเพียง 0.7038 จากภาพในเวลากลางวันและ 0.6826 จากภาพในเวลากลางคืน และเมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยพบว่าภาพในเวลากลางวันมีค่าเพียง 54.32% และในเวลากลางคืน 51.85% ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมดีขึ้น ยกเว้นค่าความเที่ยงตรงจากภาพในเวลากลางคืนที่มีค่าต่ำลงเพียงเล็กน้อย แต่เมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยทั้งหมดแล้วยังเพิ่มสูงขึ้นเป็น 68.14% จากภาพในเวลากลางวันและ 58.80% จากภาพในเวลากลางคืน ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพกับภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 27.18 ภาพต่อวินาที และภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 31.96 ภาพต่อวินาที

4.2.3 สรุปผลการทดสอบโมเดลจากสถาปัตยกรรม Faster R-CNN และ YOLOv4

สรุปข้อมูลผลการทดสอบประสิทธิภาพโมเดลจากสถาปัตยกรรม Faster R-CNN และ YOLOv4 ที่ถูกพัฒนาโมเดลขึ้นจากชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริงทั้งหมด 100% ของชุดข้อมูลทั้งหมด สรุปผลการทดลองได้ดังตารางที่ 4.5 และรูปที่ 4.3

ตารางที่ 4.5 สรุปผลการทดสอบโมเดลจากสถาปัตยกรรม Faster R-CNN และ YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด

MODEL	AVG FPS	AP DAY	AP NIGHT	AP
Faster R-CNN MS COCO 80 Classes [1920 × 1080]	14.2366	51.5875	50.6053	51.0687
Faster R-CNN MS COCO 80 Classes [963 × 1080]	17.0958	63.0854	63.3128	63.2493
YOLOv4 MS COCO 80 Classes [1920 × 1080]	27.1791	54.3163	51.8455	53.2242
YOLOv4 MS COCO 80 Classes [963 × 1080]	31.9644	68.1349	58.8033	63.9820



รูปที่ 4.3 กราฟแสดงความเที่ยงตรงเฉลี่ยและความเร็วในการประมวลผลภาพต่อวินาที ระหว่างโมเดล Faster R-CNN และ YOLOv4 แบ่งตามเวลากลางวัน กลางคืน และรวมทั้งหมด

จากตารางที่ 4.5 และรูปที่ 4.3 พบว่าเมื่อพิจารณาเปรียบเทียบผลลัพธ์ทั้งหมดระหว่างโมเดลที่พัฒนาจากสถาปัตยกรรม Faster R-CNN กับ YOLOv4 ทั้งด้านความแม่นยำในการตรวจจับด้วยค่าความเที่ยงตรงเฉลี่ยและความเร็วในการประมวลภาพต่อวินาทีพบว่าภาพที่ถูกลดขนาดพิกเซลลงได้ผลลัพธ์ที่ดีกว่าในทุก ๆ ด้าน โดยค่าความเที่ยงตรงเฉลี่ยจากภาพในเวลากลางคืนจากสถาปัตยกรรม YOLOv4 มีค่าน้อยกว่า Faster R-CNN อยู่ 4.51% แต่เมื่อพิจารณาประสิทธิภาพโดยรวมแล้ว YOLOv4 มีค่าความเที่ยงตรงเฉลี่ยจากภาพในเวลากลางวันสูงกว่าถึง 5.05% และทำให้ค่าความเที่ยงตรงเฉลี่ยจากผลการทดสอบทั้งหมดมีค่าสูงกว่า Faster R-CNN อยู่ 0.73% โดยมีเวลาประมวลผลภาพสูงถึง 31.96 ภาพต่อวินาที ซึ่งสูงกว่า Faster R-CNN ถึง 1.87 เท่า ดังนั้น โมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 จึงเหมาะสมในการนำมาพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์ทั้งในด้านความแม่นยำและความเร็วในการประมวลผล ผู้วิจัยจึงพัฒนาโมเดลด้วยสถาปัตยกรรม YOLOv4 ในชุดข้อมูลที่แตกต่างกันเพื่อเปรียบเทียบประสิทธิภาพของโมเดลซึ่งแสดงในหัวข้อถัดไป

4.3 ผลการทดสอบประสิทธิภาพโมเดลการตรวจจับบุคคลและระบุระดับความเสี่ยง

การทดสอบความแม่นยำในการตรวจจับบุคคลและระบุระดับความเสี่ยงจะทดสอบกับโมเดลที่ได้พัฒนาขึ้นจากชุดข้อมูลฝึกสอนที่แตกต่างกันจำนวน 3 ชุด คือ ชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท (MS COCO 80 Classes) ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล (MS COCO Person) และชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด (MS COCO Person + Truckdump) โดยในการทดลองจะทดสอบโมเดลกับข้อมูลสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด แบ่งออกเป็น 4 กลุ่ม คือ ภาพบุคคลจากสถานที่ปฏิบัติงานจริงอยู่ภายในพื้นที่ความเสี่ยงสูง ความเสี่ยงปานกลางในเวลากลางวันและภาพบุคคลอยู่ภายในพื้นที่ความเสี่ยงสูง ความเสี่ยงปานกลางในเวลากลางคืน แบ่งภาพออกเป็น 2 ขนาด คือ 1,920 × 1,080 พิกเซล และ 963 × 1,080 พิกเซล โดยมีการกำหนดค่าเกณฑ์ IoUไว้ที่ 0.5

การวัดประสิทธิภาพการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้นั้นจะกำหนดขอบเขตพื้นที่เป็นพิกัดจุด (X,Y) และสร้างเส้นตรงเชื่อมแต่ละจุดเข้าหากันจนเกิดเป็นขอบเขตพื้นที่ โดยขอบเขตความเสี่ยงสูงมีพิกัดทั้งหมด 6 พิกัด ดังนี้ (834,0) (1136,0) (1136,259) (1275,950) (690,950) (834,259) และขอบเขตความเสี่ยงปานกลางมีทั้งหมด 6 พิกัด ดังนี้ (714,0) (1256,0) (1256,259) (1465,1080) (502,1080) (714,259) โดยจะนำผลลัพธ์จากกรอบล้อมรอบวัตถุที่ทำนายถูกว่าเป็นประเภทบุคคลมาทำการหาจำนวนจุด 3 จุด ดังนี้ จุดซ้ายล่างของกรอบล้อมรอบวัตถุ จุดกึ่งกลางจากเส้นขอบล่างกรอบล้อมรอบวัตถุ และจุดขวาล่างของกรอบล้อมรอบวัตถุ เทียบกับตำแหน่งภายในพื้นที่ของขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลางตามลำดับ โดยถ้าในขอบเขตความเสี่ยงสูงมีจุดอยู่ 2 จุดขึ้นไปจะระบุว่าบุคคลที่ตรวจจับได้อยู่ในพื้นที่ความเสี่ยงสูง และถ้าในขอบเขตความเสี่ยงปานกลางมีจุดอยู่ 2 จุดขึ้นไปจะระบุว่าบุคคลที่ตรวจจับได้อยู่ในพื้นที่ความเสี่ยงปานกลาง โดยภาพขนาด 963 × 1,080 พิกเซลนั้นตำแหน่งกรอบที่ตรวจจับได้ในพิกัดแกน Y จะเท่าเดิม แต่ในพิกัดแกน X จะถูกเพิ่มค่าเพื่อให้เทียบเท่ากับภาพขนาด 1,920 × 1,080 พิกเซล ด้วยการบวกค่าขอบภาพทางด้านซ้ายที่ถูกตัดออกไปจากภาพต้นฉบับ คือ 502 พิกเซล

4.3.1 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท

การทดสอบประสิทธิภาพโมเดลที่ถูกพัฒนาจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงทั้งหมด 30% ของชุดข้อมูลทั้งหมด ให้ผลการทดลองการตรวจจับบุคคลดังตารางที่ 4.6 และตารางที่ 4.7 ส่วนผลการทดลองการระบุระดับความเสี่ยงแสดงดังตารางที่ 4.8

ตารางที่ 4.6 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO 80 Classes	GT	HUMAN DETECTION				
		BOXES	TP	FP	FN	AVG TIME (ms)
DAY HIGH-RISK [1920 × 1080]	685	685	433	0	252	31.4901
DAY HIGH-RISK [963 × 1080]	685	685	564	0	121	30.4756
DAY MEDIUM-RISK [1920 × 1080]	828	828	497	0	331	31.6411
DAY MEDIUM-RISK [963 × 1080]	828	828	524	0	304	30.7318
NIGHT HIGH-RISK [1920 × 1080]	644	644	406	0	238	31.6327
NIGHT HIGH-RISK [963 × 1080]	644	644	446	3	195	30.6844
NIGHT MEDIUM-RISK [1920 × 1080]	576	576	304	0	272	31.8312
NIGHT MEDIUM-RISK [963 × 1080]	576	580	355	4	221	31.3076

ตารางที่ 4.7 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO 80 Classes	HUMAN DETECTION				
	AVG FPS	PRECISION	RECALL	F1	AP
DAY HIGH-RISK [1920 × 1080]	31.4901	1.0000	0.6321	0.7746	0.6307
DAY HIGH-RISK [963 × 1080]	30.4756	1.0000	0.8234	0.9031	0.8219
DAY MEDIUM-RISK [1920 × 1080]	31.6411	1.0000	0.6002	0.7502	0.5990
DAY MEDIUM-RISK [963 × 1080]	30.7318	1.0000	0.6329	0.7751	0.6316
NIGHT HIGH-RISK [1920 × 1080]	31.6327	1.0000	0.6304	0.7733	0.6289
NIGHT HIGH-RISK [963 × 1080]	30.6844	0.9933	0.6958	0.8183	0.6940
NIGHT MEDIUM-RISK [1920 × 1080]	31.8312	1.0000	0.5278	0.6909	0.5260
NIGHT MEDIUM-RISK [963 × 1080]	31.3076	0.9889	0.6163	0.7594	0.6144

ตารางที่ 4.8 ผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO 80 Classes	GT	TP	RISK-LEVEL IDENTIFICATION		
			TP CORRECT (BOXES)	TP CORRECT (%)	GT CORRECT (%)
DAY HIGH-RISK [1920 × 1080]	685	433	398	91.9200	58.1000
DAY HIGH-RISK [963 × 1080]	685	564	423	75.0000	61.7500
DAY MEDIUM-RISK [1920 × 1080]	828	497	489	98.3900	59.0600
DAY MEDIUM-RISK [963 × 1080]	828	524	521	99.4300	62.9200
NIGHT HIGH-RISK [1920 × 1080]	644	406	406	100.0000	63.0400
NIGHT HIGH-RISK [963 × 1080]	644	446	446	100.0000	69.2500
NIGHT MEDIUM-RISK [1920 × 1080]	576	304	304	100.0000	52.7800
NIGHT MEDIUM-RISK [963 × 1080]	576	355	355	100.0000	61.6300

จากตารางที่ 4.6 จะเห็นได้ว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีผลลัพธ์การตรวจจับบุคคลในภาพเวลากลางวันดีกว่าภาพในเวลากลางคืน โดยภาพในเวลากลางวันโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงปานกลางได้ดีกว่าความเสี่ยงสูง โดยมีจำนวนกรอบล้อมรอบวัตถุที่สร้างถูกต้อง 63.21% ในขอบเขตความเสี่ยงสูงและ 60.02% ในขอบเขตความเสี่ยงปานกลาง โดยที่โมเดลไม่มีการตรวจจับกรอบล้อมรอบวัตถุผิดพลาด แต่ยังพบว่ามีกรอบล้อมรอบวัตถุที่ไม่ถูกสร้างสูงถึง 36.79% ในขอบเขตความเสี่ยงสูงและ 39.98% ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลง สามารถเพิ่มความถูกต้องในการสร้างกรอบได้ถึง 19.13% ในขอบเขตความเสี่ยงสูง และ 3.27% ในขอบเขตความเสี่ยงปานกลาง ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดีกว่าความเสี่ยงปานกลาง โดยมีจำนวนกรอบล้อมรอบวัตถุที่สร้างถูกต้อง 63.04% ในขอบเขตความเสี่ยงสูงและ 52.78% ในขอบเขตความเสี่ยงปานกลาง โดยที่โมเดลไม่มีการตรวจจับกรอบล้อมรอบวัตถุผิดพลาด แต่ยังพบว่ามีกรอบล้อมรอบวัตถุที่ไม่ถูกสร้างสูงถึง 36.96% ในขอบเขตความเสี่ยงสูงและ 47.22% ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลง สามารถเพิ่มความถูกต้องในการสร้างกรอบได้ถึง 6.21% ในขอบเขตความเสี่ยงสูงและ 8.88% ในขอบเขตความเสี่ยงปานกลาง และมีการสร้างกรอบล้อมรอบวัตถุผิดพลาดเพิ่มขึ้นเพียงเล็กน้อย ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพเฉลี่ยเพียง 31.64 มิลลิวินาทีต่อภาพขนาด 1,920 × 1,080 พิกเซล และ 30.78 มิลลิวินาทีต่อภาพขนาด 963 × 1,080 พิกเซล

จากตารางที่ 4.7 จะเห็นได้ว่าผลสรุปการตรวจจับบุคคลของสถาปัตยกรรม YOLOv4 มีค่าความเที่ยงตรงค่อนข้างสูงมาก โดยเฉพาะภาพในเวลากลางวัน มีค่าสูงสุด คือ 1 ทั้งภาพในขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลาง โดยมีค่าการระลึกอยู่ที่ 0.6321 ในขอบเขตความเสี่ยงสูงและ 0.6002 ในขอบเขตความเสี่ยงปานกลาง ซึ่งส่งผลให้ค่าประสิทธิภาพโดยรวมมีค่าเพียง 0.7746 จากภาพในขอบเขตความเสี่ยงสูงและ 0.7502 ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยพบว่าภาพในขอบเขตความเสี่ยงสูงมีค่าเพียง 63.07% และในความเสี่ยงปานกลาง 59.90% ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมทั้งหมดดีขึ้น โดยเฉพาะค่าประสิทธิภาพโดยรวมสามารถเพิ่มขึ้นเป็น 0.9031 และค่าความเที่ยงตรงเฉลี่ยเพิ่มสูงขึ้นถึง 82.19% จากภาพในขอบเขตความเสี่ยงสูง ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดีโดยมีค่าความเที่ยงตรงสูงถึง 1.0000 ทั้ง 2 ขอบเขตความเสี่ยง แต่ค่าการระลึก ค่าประสิทธิภาพโดยรวม และค่าความเที่ยงตรงเฉลี่ยมีค่าเพียง 0.6958 0.7733 และ 63.16% ตามลำดับ ในขอบเขตความเสี่ยงสูง และ 0.5278 0.6909 และ 69.40% ในขอบเขตความเสี่ยงปานกลาง ในขณะที่ผลลัพธ์

จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมทั้งหมดดีขึ้นโดยมีค่าความเที่ยงตรงลดลงเพียงเล็กน้อย แต่ค่าประสิทธิภาพโดยรวมและค่าความเที่ยงตรงเฉลี่ยเพิ่มขึ้นเป็น 0.8183 และ 69.40% ตามลำดับในขอบเขตความเสี่ยงสูงและ 0.7594 และ 61.44% ตามลำดับ ในขอบเขตความเสี่ยงปานกลาง ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 31.70 ภาพต่อวินาที และภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 32.56 ภาพต่อวินาที

จากตารางที่ 4.8 จะเห็นได้ว่าผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีความแม่นยำสูงมาก โดยผลลัพธ์จากภาพในเวลากลางวัน เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้นั้นมีความถูกต้องสูงถึง 91.92% ในขอบเขตความเสี่ยงสูงและ 98.39% ในขอบเขตความเสี่ยงปานกลาง แต่เมื่อพิจารณาเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดที่อยู่ในชุดข้อมูลพบว่าโมเดลสามารถระบุความเสี่ยงได้ถูกต้องเพียง 58.10% ในขอบเขตความเสี่ยงสูงและ59.06% ในขอบเขตความเสี่ยงปานกลาง เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่ามีจำนวนกรอบที่ทำนายได้ถูกต้องเพิ่มขึ้นจากเดิม แต่กลับพบว่าความถูกต้องในการระบุความเสี่ยงในขอบเขตความเสี่ยงสูงลดลงถึง 16.92% เมื่อเปรียบเทียบกับจำนวนที่ตรวจจับได้ แต่ความถูกต้องโดยรวมเมื่อเปรียบเทียบกับชุดข้อมูลจริงนั้นสูงขึ้นทั้งในขอบเขตความเสี่ยงสูงและความปานกลาง ในขณะที่ผลลัพธ์จากภาพในเวลากลางคืน เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้นั้นมีความถูกต้องสูงถึง 100% ทั้ง 2 ขอบเขตและเมื่อพิจารณาเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดที่อยู่ในชุดข้อมูลพบว่าโมเดลสามารถระบุความเสี่ยงได้ถูกต้องเพียง 63.04% ในขอบเขตความเสี่ยงสูง และ 52.78% ในขอบเขตความเสี่ยงปานกลาง เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าค่าความถูกต้องของการระบุความเสี่ยงเพิ่มขึ้นทั้งในกรณีที่เปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้และเมื่อเปรียบเทียบกับชุดข้อมูลจริงทั้งหมด โดยมีความถูกต้องเพิ่มขึ้น 6.21% ในขอบเขตความเสี่ยงสูงและ 8.85% ในขอบเขตความเสี่ยงปานกลาง

4.3.2 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล

การทดสอบประสิทธิภาพโมเดลที่ถูกพัฒนาจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงทั้งหมด 30% ของชุดข้อมูลทั้งหมด ให้ผลการทดลองการตรวจจับบุคคลดังตารางที่ 4.9 และตารางที่ 4.10 ส่วนผลการทดลองการระบุระดับความเสี่ยงแสดงดังตารางที่ 4.11

ตารางที่ 4.9 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person	GT	HUMAN DETECTION				
		BOXES	TP	FP	FN	AVG TIME (ms)
DAY HIGH-RISK [1920 × 1080]	685	685	685	0	0	42.0550
DAY HIGH-RISK [963 × 1080]	685	685	685	0	0	30.4069
DAY MEDIUM-RISK [1920 × 1080]	828	830	828	2	0	42.2156
DAY MEDIUM-RISK [963 × 1080]	828	828	828	0	0	30.5299
NIGHT HIGH-RISK [1920 × 1080]	644	644	621	0	23	42.1461
NIGHT HIGH-RISK [963 × 1080]	644	644	627	7	10	30.9208
NIGHT MEDIUM-RISK [1920 × 1080]	576	576	515	0	61	43.5158
NIGHT MEDIUM-RISK [963 × 1080]	576	722	499	176	47	31.0835

ตารางที่ 4.10 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person	HUMAN DETECTION				
	AVG FPS	PRECISION	RECALL	F1	AP
DAY HIGH-RISK [1920 × 1080]	23.8539	1.0000	1.0000	1.0000	0.9985
DAY HIGH-RISK [963 × 1080]	32.9619	1.0000	1.0000	1.0000	0.9985
DAY MEDIUM-RISK [1920 × 1080]	23.8287	0.9976	1.0000	0.9988	0.9988
DAY MEDIUM-RISK [963 × 1080]	32.8111	1.0000	1.0000	1.0000	0.9988
NIGHT HIGH-RISK [1920 × 1080]	23.7560	1.0000	0.9643	0.9818	0.9627
NIGHT HIGH-RISK [963 × 1080]	32.4319	0.9890	0.9843	0.9866	0.9825
NIGHT MEDIUM-RISK [1920 × 1080]	23.0478	1.0000	0.8941	0.9441	0.8924
NIGHT MEDIUM-RISK [963 × 1080]	32.3016	0.7393	0.9139	0.8174	0.8876

ตารางที่ 4.11 ผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person	GT	TP	RISK-LEVEL IDENTIFICATION		
			TP CORRECT (BOXES)	TP CORRECT (%)	GT CORRECT (%)
DAY HIGH-RISK [1920 × 1080]	685	685	579	84.5300	84.5300
DAY HIGH-RISK [963 × 1080]	685	685	600	87.5900	87.5900
DAY MEDIUM-RISK [1920 × 1080]	828	828	826	99.7600	99.7600
DAY MEDIUM-RISK [963 × 1080]	828	828	825	99.6400	99.6400
NIGHT HIGH-RISK [1920 × 1080]	644	621	621	100.0000	96.4300
NIGHT HIGH-RISK [963 × 1080]	644	627	627	100.0000	97.3600
NIGHT MEDIUM-RISK [1920 × 1080]	576	515	515	100.0000	89.4100
NIGHT MEDIUM-RISK [963 × 1080]	576	499	499	100.0000	86.6300

จากตารางที่ 4.9 จะเห็นได้ว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 ด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล มีผลลัพธ์การตรวจจับบุคคลในภาพเวลากลางวันดีกว่าภาพในเวลากลางคืน โดยภาพในเวลากลางวันโมเดลสร้างกรอบล้อมรอบวัตถุได้ถูกต้องถึง 100% ทั้งในขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลาง และมีการสร้างกรอบล้อมรอบวัตถุผิดพลาดเพียง 2 กรอบเท่านั้น เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าโมเดลมีความแม่นยำมากโดยสามารถสร้างกรอบล้อมรอบวัตถุได้ถูกต้องครบทุกกรอบและไม่มีการสร้างกรอบที่ผิดพลาด ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดีกว่าความเสี่ยงปานกลาง โดยมีจำนวนกรอบล้อมรอบวัตถุที่สร้างถูกต้อง 96.43% ในขอบเขตความเสี่ยงสูงและ 89.41% ในขอบเขตความเสี่ยงปานกลาง โดยที่โมเดลไม่มีการสร้างกรอบล้อมรอบวัตถุผิดพลาด และมีกรอบล้อมรอบวัตถุที่ไม่ถูกสร้างเพียงเล็กน้อย คือ 3.57% ในขอบเขตความเสี่ยงสูงและ 10.59% ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลง สามารถเพิ่มความถูกต้องในการสร้างกรอบได้อีก 0.93% ในขอบเขตความเสี่ยงสูง แต่กลับพบว่าความถูกต้องลดลง 2.78% ในขอบเขตความเสี่ยงปานกลาง และมีการสร้างกรอบล้อมรอบวัตถุผิดพลาดเพิ่มขึ้นเป็น 1.09% ในขอบเขตความเสี่ยงสูงและเพิ่มขึ้นเป็น 30.56% ในขอบเขตความเสี่ยงปานกลาง ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ยเพียง 42.43 มิลลิวินาทีต่อภาพ และภาพขนาด 963 × 1,080 พิกเซล ใช้เวลาประมวลผลภาพเฉลี่ยเพียงแค่ 30.72 มิลลิวินาทีต่อภาพ

จากตารางที่ 4.10 จะเห็นได้ว่าผลสรุปการตรวจจับบุคคลของสถาปัตยกรรม YOLOv4 มีค่าความเที่ยงตรงสูงมาก โดยเฉพาะภาพในเวลากลางวันมีค่า 1 ในขอบเขตความเสี่ยงสูงและ 0.9976 ในขอบเขตความเสี่ยงปานกลาง โดยมีค่าการระลึกอยู่ที่ 1 ในทั้ง 2 ขอบเขตความเสี่ยง ซึ่งส่งผลให้ค่าประสิทธิภาพโดยรวมมีค่าสูงเช่นกันคือ 1 จากภาพในขอบเขตความเสี่ยงสูงและ 0.9988 ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยพบว่าภาพในขอบเขตความเสี่ยงสูงมีค่าสูงถึง 99.85% และ 99.88% ในความเสี่ยงปานกลาง ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมทั้งหมดดีขึ้น โดยมีค่าความเที่ยงตรง ค่าการระลึก ค่าประสิทธิภาพโดยรวม เท่ากับ 1 ทั้งในขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลาง โดยค่าความเที่ยงตรงเฉลี่ยยังคงมีค่าเท่าเดิม ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดี โดยมีค่าความเที่ยงตรงสูงถึง 1 ทั้ง 2 ขอบเขตความเสี่ยง และค่าการระลึก ค่าประสิทธิภาพโดยรวม และค่าความเที่ยงตรงเฉลี่ย มีค่าค่อนข้างสูงดังนี้ 0.9643 0.9818 และ 96.27% ตามลำดับ ในขอบเขตความเสี่ยงสูงและ 0.8941 0.9441 และ 89.24% ในขอบเขตความเสี่ยงปานกลาง ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าค่าความเที่ยงตรงลดลงเพียงเล็กน้อยทั้ง 2 ขอบเขตความเสี่ยง แต่ค่าการระลึกเพิ่มขึ้น

เป็น 0.9843 ในขอบเขตความเสี่ยงสูงและ 0.9139 ในขอบเขตความเสี่ยงปานกลาง โดยที่ค่าประสิทธิภาพโดยรวมและค่าความเที่ยงตรงเฉลี่ยในขอบเขตความเสี่ยงสูงนั้นเพิ่มมากขึ้นเป็น 0.9866 และ 98.25% ตามลำดับ ในขณะที่ขอบเขตความเสี่ยงต่ำลดลงเหลือ 0.8174 และ 88.76% ตามลำดับ ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 23.65 ภาพต่อวินาที และ ภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 32.64 ภาพต่อวินาที

จากตารางที่ 4.11 จะเห็นได้ว่าผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลมีความแม่นยำสูงมาก โดยผลลัพธ์จากภาพในเวลากลางวันมีความถูกต้องสูงถึง 84.53% ในขอบเขตความเสี่ยงสูงและ 99.76% ในขอบเขตความเสี่ยงปานกลาง เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้และเปรียบเทียบกับชุดข้อมูลจริงทั้งหมด เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าจำนวนกรอบที่ทำนายได้ถูกต้องมีจำนวนคงเดิม และพบว่าความถูกต้องในการระบุความเสี่ยงในขอบเขตความเสี่ยงสูงมีค่าเพิ่มขึ้น 3.06% แต่ความถูกต้องในขอบเขตความเสี่ยงปานกลางมีค่าลดลงเล็กน้อยเพียง 0.12% ในขณะที่ผลลัพธ์จากภาพในเวลากลางคืน เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้นั้นมีความถูกต้องสูงถึง 100% ทั้ง 2 ขอบเขตและเมื่อพิจารณาเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดที่อยู่ในชุดข้อมูลพบว่าโมเดลสามารถระบุความเสี่ยงได้ถูกต้องสูงถึง 96.43% ในขอบเขตความเสี่ยงสูง และ 89.41% ในขอบเขตความเสี่ยงปานกลาง เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าค่าความถูกต้องของการระบุความเสี่ยงเมื่อเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดเพิ่มขึ้น 0.93% ในขอบเขตความเสี่ยงสูง แต่กลับลดลง 2.78% ในขอบเขตความเสี่ยงปานกลาง

4.3.3 ผลทดสอบการตรวจจับบุคคลด้วยโมเดล YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง

การทดสอบประสิทธิภาพโมเดลที่ถูกพัฒนาจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงทั้งหมด 30% ของชุดข้อมูลทั้งหมด ให้ผลการทดลองการตรวจจับบุคคลดังตารางที่ 4.12 และตารางที่ 4.13 ส่วนผลการทดลองการระบุระดับความเสี่ยงแสดงดังตารางที่ 4.14

ตารางที่ 4.12 ผลทดสอบการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person + Truckdump	GT	HUMAN DETECTION				
		BOXES	TP	FP	FN	AVG TIME (ms)
DAY HIGH-RISK [1920 × 1080]	685	686	684	1	1	42.6613
DAY HIGH-RISK [963 × 1080]	685	685	685	0	0	30.5433
DAY MEDIUM-RISK [1920 × 1080]	828	828	828	0	0	41.9212
DAY MEDIUM-RISK [963 × 1080]	828	828	828	0	0	30.7289
NIGHT HIGH-RISK [1920 × 1080]	644	644	612	0	32	42.4991
NIGHT HIGH-RISK [963 × 1080]	644	644	622	5	17	30.9790
NIGHT MEDIUM-RISK [1920 × 1080]	576	576	517	0	59	42.7554
NIGHT MEDIUM-RISK [963 × 1080]	576	656	494	113	49	30.7430

ตารางที่ 4.13 สรุปผลการตรวจจับบุคคลด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person + Truckdump	HUMAN DETECTION				
	AVG FPS	PRECISION	RECALL	F1	AP
DAY HIGH-RISK [1920 × 1080]	23.4691	0.9985	0.9985	0.9985	0.9971
DAY HIGH-RISK [963 × 1080]	32.8065	1.0000	1.0000	1.0000	0.9985
DAY MEDIUM-RISK [1920 × 1080]	24.0136	1.0000	1.0000	1.0000	0.9988
DAY MEDIUM-RISK [963 × 1080]	32.5785	1.0000	1.0000	1.0000	0.9988
NIGHT HIGH-RISK [1920 × 1080]	23.5413	1.0000	0.9503	0.9745	0.9488
NIGHT HIGH-RISK [963 × 1080]	32.3470	0.9920	0.9734	0.9826	0.9718
NIGHT MEDIUM-RISK [1920 × 1080]	23.4020	1.0000	0.8976	0.9460	0.8958
NIGHT MEDIUM-RISK [963 × 1080]	32.5942	0.8138	0.9098	0.8591	0.8993

ตารางที่ 4.14 ผลการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด

YOLOv4 MS COCO Person + Truckdump	GT	TP	RISK-LEVEL IDENTIFICATION		
			TP CORRECT (BOXES)	TP CORRECT (%)	GT CORRECT (%)
DAY HIGH-RISK [1920 × 1080]	685	684	592	86.5500	86.4200
DAY HIGH-RISK [963 × 1080]	685	685	627	91.5300	91.5300
DAY MEDIUM-RISK [1920 × 1080]	828	828	826	99.7600	99.7600
DAY MEDIUM-RISK [963 × 1080]	828	828	825	99.6400	99.6400
NIGHT HIGH-RISK [1920 × 1080]	644	612	612	100.0000	95.0300
NIGHT HIGH-RISK [963 × 1080]	644	622	622	100.0000	96.5800
NIGHT MEDIUM-RISK [1920 × 1080]	576	517	517	100.0000	89.7600
NIGHT MEDIUM-RISK [963 × 1080]	576	494	494	100.0000	85.7600

จากตารางที่ 4.12 จะเห็นได้ว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 ด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มีผลลัพธ์การตรวจจับบุคคลในภาพเวลากลางวันดีกว่าภาพในเวลากลางคืน โดยภาพในเวลากลางวันโมเดลสร้างกรอบล้อมรอบวัตถุได้ถูกต้องถึง 99.85% ในขอบเขตความเสี่ยงสูงและ 100% ในขอบเขตความเสี่ยงปานกลาง และมีการสร้างกรอบล้อมรอบวัตถุผิดพลาดเพียง 1 กรอบเท่านั้น เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าโมเดลมีความแม่นยำมากโดยสามารถสร้างกรอบล้อมรอบวัตถุได้ถูกต้องครบทุกกรอบและไม่มีการสร้างกรอบที่ผิดพลาด ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดีกว่าความเสี่ยงปานกลาง โดยมีจำนวนกรอบล้อมรอบวัตถุที่สร้างถูกต้อง 95.03% ในขอบเขตความเสี่ยงสูงและ 89.76% ในขอบเขตความเสี่ยงปานกลาง โดยที่โมเดลไม่มีการสร้างกรอบล้อมรอบวัตถุผิดพลาด และมีกรอบล้อมรอบวัตถุที่ไม่ถูกสร้างเพียงเล็กน้อย คือ 2.64% ในขอบเขตความเสี่ยงสูงและ 8.51% ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลง สามารถเพิ่มความถูกต้องในการสร้างกรอบได้อีก 1.55% ในขอบเขตความเสี่ยงสูง แต่กลับพบว่าความถูกต้องลดลง 4% ในขอบเขตความเสี่ยงปานกลาง และมีการสร้างกรอบล้อมรอบวัตถุผิดพลาดเพิ่มขึ้นเล็กน้อยคือ 0.78% ในขอบเขตความเสี่ยงสูงและ 19.62% ในขอบเขตความเสี่ยงปานกลาง ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ยเพียง 42.42 มิลลิวินาทีต่อภาพ และภาพขนาด 963 × 1,080 พิกเซล ใช้เวลาเฉลี่ยเพียงแค่ 30.75 มิลลิวินาทีต่อภาพ

จากตารางที่ 4.13 จะเห็นได้ว่าผลสรุปการตรวจจับบุคคลของสถาปัตยกรรม YOLOv4 มีค่าความเที่ยงตรง ค่าการระลึก และค่าประสิทธิภาพโดยรวมสูงมาก โดยภาพในเวลากลางวันมีค่าเท่ากันคือ 0.9985 ในขอบเขตความเสี่ยงสูงและ 1 ในขอบเขตความเสี่ยงปานกลาง และเมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยพบว่าภาพในขอบเขตความเสี่ยงสูงมีค่าสูงถึง 99.71% และในความเสี่ยงปานกลาง 99.88% ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าประสิทธิภาพการตรวจจับโดยรวมทั้งหมดดีขึ้น โดยมีค่าความเที่ยงตรง ค่าการระลึก และค่าประสิทธิภาพโดยรวมเท่ากับ 1 ทั้งในขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลาง โดยค่าความเที่ยงตรงเฉลี่ยในขอบเขตความเสี่ยงสูงมีค่าเพิ่มขึ้นเป็น 99.85% ส่วนในขอบเขตความเสี่ยงปานกลางยังคงมีค่าเท่าเดิม ในขณะที่ผลลัพธ์จากการทดสอบภาพในเวลากลางคืนโมเดลสามารถตรวจจับบุคคลในขอบเขตความเสี่ยงสูงได้ดี โดยมีค่าความเที่ยงตรงสูงถึง 1 ทั้ง 2 ขอบเขตความเสี่ยง และค่าการระลึก ค่าประสิทธิภาพโดยรวม และค่าความเที่ยงตรงเฉลี่ยมีค่าค่อนข้างสูงดังนี้ 0.9503 0.9745 และ 94.88% ตามลำดับ ในขอบเขตความเสี่ยงสูงและ 0.8976 0.9460 และ 89.58% ตามลำดับ ในขอบเขตความเสี่ยงปานกลาง ในขณะที่ผลลัพธ์จากภาพที่ถูกลดขนาดลงพบว่าค่าความเที่ยงตรง

ลดลงเพียงเล็กน้อยทั้ง 2 ขอบเขตความเสี่ยง แต่ค่าการระลึกยังคงเพิ่มขึ้นเป็น 0.9734 ในขอบเขตความเสี่ยงสูงและ 0.9098 ในขอบเขตความเสี่ยงปานกลาง ส่งผลให้ค่าประสิทธิภาพโดยรวมในขอบเขตความเสี่ยงสูงมีค่าเพิ่มขึ้นเป็น 0.9826 แต่ในขอบเขตความเสี่ยงปานกลางกลับมีค่าลดลงเหลือ 0.8591 แต่เมื่อพิจารณาค่าความเที่ยงตรงเฉลี่ยทั้ง 2 ขอบเขตความเสี่ยงยังคงเพิ่มขึ้นเป็น 97.18% ในขอบเขตความเสี่ยงสูงและ 89.93% ในขอบเขตความเสี่ยงปานกลาง ทั้งนี้โมเดลใช้เวลาในการประมวลผลภาพขนาด 1,920 × 1,080 พิกเซล เฉลี่ย 23.64 ภาพต่อวินาที และ ภาพขนาด 963 × 1,080 พิกเซล เฉลี่ย 32.58 ภาพต่อวินาที

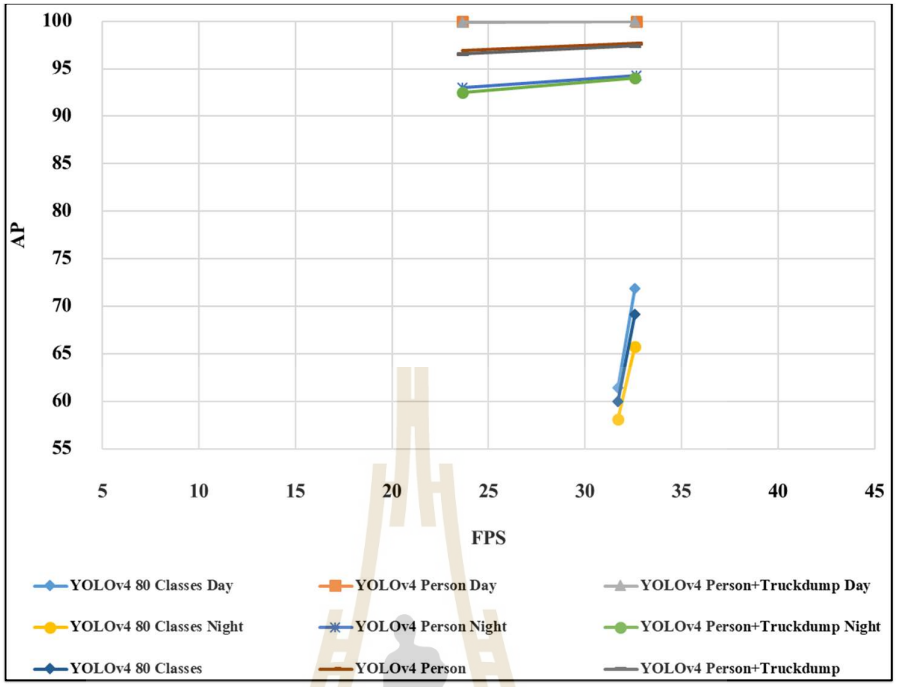
จากตารางที่ 4.14 จะเห็นได้ว่าผลทดสอบการระบุความเสี่ยงจากตำแหน่งบุคคลที่ตรวจจับได้ด้วยโมเดลจากสถาปัตยกรรม YOLOv4 ฝึกสอนด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มีความแม่นยำสูงมาก โดยผลลัพธ์จากภาพในเวลากลางวันมีความถูกต้องสูงถึง 86.55% ในขอบเขตความเสี่ยงสูงและ 99.76% ในขอบเขตความเสี่ยงปานกลาง เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้ และเมื่อเปรียบเทียบกับชุดข้อมูลจริงทั้งหมด พบว่ามีค่าความถูกต้องอยู่ที่ 86.42% ในขอบเขตความเสี่ยงสูงและ 99.76% ในขอบเขตความเสี่ยงปานกลาง เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าจำนวนกรอบที่ทำนายได้ถูกต้องมีจำนวนเพิ่มขึ้น และพบว่าความถูกต้องในการระบุความเสี่ยงในขอบเขตความเสี่ยงสูงมีค่าเพิ่มขึ้น 5.11% แต่ความถูกต้องในขอบเขตความเสี่ยงปานกลางมีค่าลดลงเล็กน้อยเพียง 0.12% เมื่อเปรียบเทียบกับชุดข้อมูลจริงทั้งหมด ในขณะที่ผลลัพธ์จากภาพในเวลากลางคืน เมื่อเปรียบเทียบกับจำนวนกรอบที่ตรวจจับได้นั้นมีความถูกต้องสูงถึง 100% ทั้ง 2 ขอบเขตและเมื่อพิจารณาเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดที่อยู่ในชุดข้อมูลพบว่าโมเดลสามารถระบุความเสี่ยงได้ถูกต้องสูงถึง 96.03% ในขอบเขตความเสี่ยงสูง และ 89.76% ในขอบเขตความเสี่ยงปานกลาง เมื่อพิจารณาผลลัพธ์ที่ได้จากการลดขนาดภาพลงพบว่าค่าความถูกต้องของการระบุความเสี่ยงเมื่อเปรียบเทียบกับชุดข้อมูลจริงทั้งหมดเพิ่มขึ้น 1.55% ในขอบเขตความเสี่ยงสูงแต่กลับลดลง 4% ในขอบเขตความเสี่ยงปานกลาง

4.3.4 สรุปผลการทดสอบโมเดลที่พัฒนาขึ้นทั้งหมด

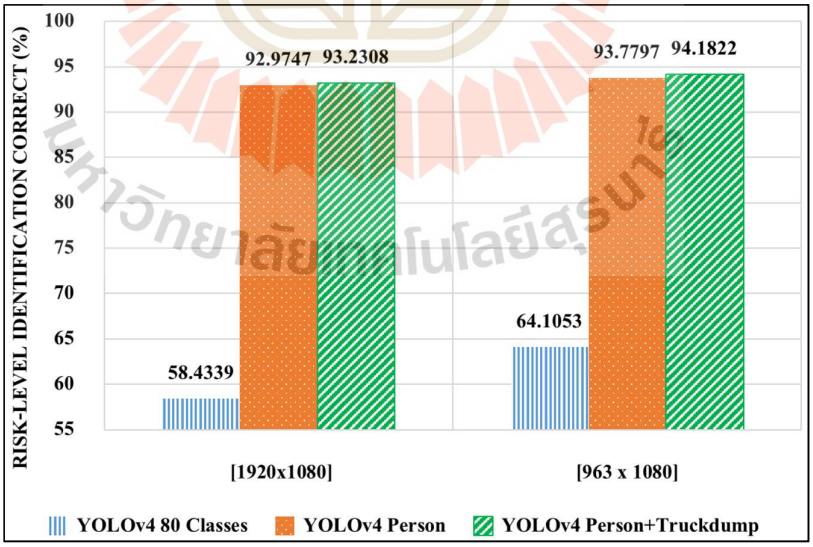
สรุปข้อมูลผลการทดสอบประสิทธิภาพโมเดลที่พัฒนาขึ้นจากสถาปัตยกรรม YOLOv4 ด้วยชุดข้อมูลที่แตกต่างกันจำนวน 3 ชุด คือ ชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล และชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด ทดสอบกับชุดข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริงทั้งหมด 30% ของชุดข้อมูลทั้งหมด สรุปผลการทดลองได้ดังตารางที่ 4.15 รูปที่ 4.4 และรูปที่ 4.5

ตารางที่ 4.15 สรุปผลการทดสอบโมเดลพี่พัฒนาขึ้นทั้งหมด

MODEL	AVG FPS	AP DAY	AP NIGHT	AP	RISK CORRECT (%)
YOLOv4 MS COCO 80 Classes [1920 × 1080]	31.7037	61.4012	58.1148	59.9707	58.4339
YOLOv4 MS COCO 80 Classes [963 × 1080]	32.5561	71.8440	65.7173	69.1575	64.1053
YOLOv4 MS COCO Person [1920 × 1080]	23.6541	99.9339	93.0328	96.8899	92.9747
YOLOv4 MS COCO Person [963 × 1080]	32.6383	99.9339	94.2544	97.6780	93.7797
YOLOv4 MS COCO Person + Truckdump [1920 × 1080]	23.6371	99.8678	92.4590	96.5971	93.2308
YOLOv4 MS COCO Person + Truckdump [963 × 1080]	32.5804	99.9339	94.0626	97.4597	94.1822



รูปที่ 4.4 กราฟแสดงความเที่ยงตรงเฉลี่ยและความเร็วในการประมวลผลภาพต่อวินาที ระหว่างโมเดล YOLOv4 ในแต่ละชุดข้อมูล แบ่งตามเวลากลางวัน กลางคืน และรวมทั้งหมด



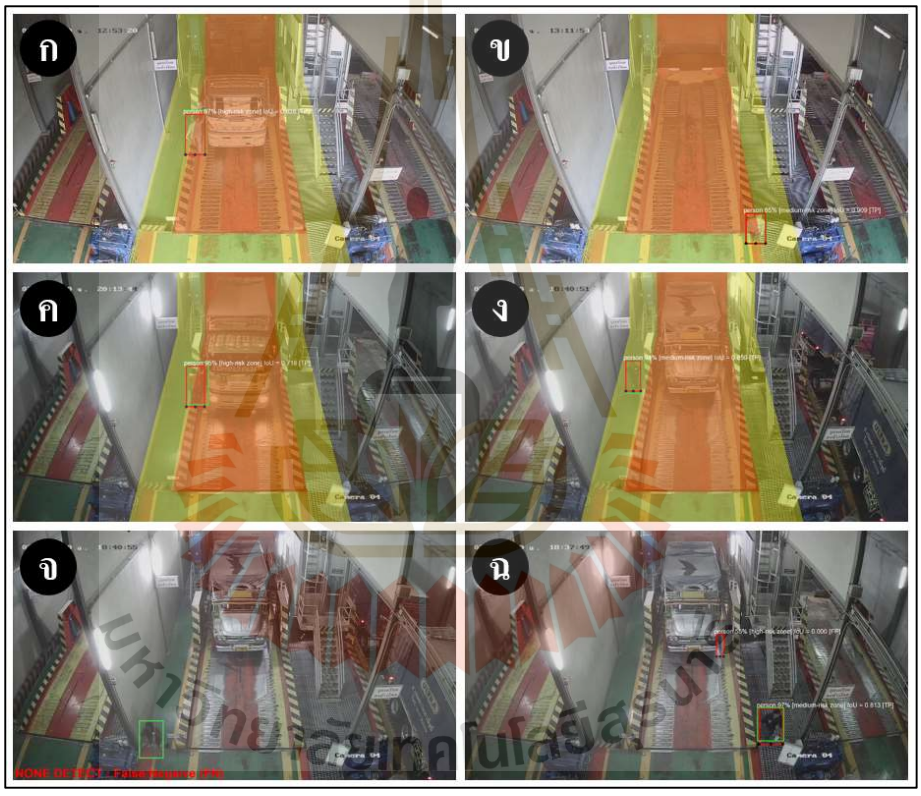
รูปที่ 4.5 กราฟแสดงความถูกต้องในการระบุความเสี่ยงบุคคลเมื่อเทียบกับชุดข้อมูลจริงทั้งหมดระหว่างโมเดล YOLOv4 ที่ถูกพัฒนาขึ้นในแต่ละชุดข้อมูล

จากตารางที่ 4.5 และรูปที่ 4.4 เมื่อพิจารณาเปรียบเทียบผลลัพธ์ทั้งหมดระหว่างโมเดลที่พัฒนาจากสถาปัตยกรรม YOLOv4 ทั้ง 3 ชุดข้อมูล พบว่าโมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีประสิทธิภาพในการตรวจจับบุคคลน้อยที่สุด โดยมีค่าความเที่ยงตรงเฉลี่ยของภาพในเวลากลางวันอยู่ที่ 61.40% ภาพในเวลากลางคืน 58.12% และภาพรวมทั้งชุดข้อมูลได้เพียง 59.97% ด้วยความเร็วประมวลผลภาพ 31.70 ภาพต่อวินาที และสามารถเพิ่มความแม่นยำได้จากการลดขนาดภาพ โดยมีค่าความเที่ยงตรงเฉลี่ยของภาพในเวลากลางวันเพิ่มขึ้นเป็น 71.84% ภาพในเวลากลางคืนเพิ่มขึ้นเป็น 65.72% และภาพรวมทั้งชุดข้อมูลเพิ่มขึ้นเป็น 69.16% ด้วยความเร็วประมวลผลภาพ 32.57 ภาพต่อวินาที ในขณะที่โมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล มีความแม่นยำในการตรวจจับบุคคลมากที่สุด โดยที่ผลลัพธ์จากขนาดภาพต้นฉบับมีความเที่ยงตรงเฉลี่ยของภาพในเวลากลางวันสูงถึง 99.93% ภาพในเวลากลางคืน 93.03% และภาพรวมทั้งชุดข้อมูลสูงถึง 96.89% ด้วยความเร็วประมวลผลภาพ 23.65 ภาพต่อวินาที และสามารถเพิ่มความแม่นยำได้จากการลดขนาดภาพลงโดยมีค่าความเที่ยงตรงเฉลี่ยของภาพในเวลากลางวันยังคงเท่าเดิม แต่ภาพในเวลากลางคืนเพิ่มขึ้นเป็น 94.25% และภาพรวมทั้งชุดข้อมูลเพิ่มขึ้นเป็น 97.68% ด้วยความเร็วประมวลผลภาพสูงที่สุดคือ 32.64 ภาพต่อวินาที และสุดท้ายโมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มีความแม่นยำในการตรวจจับบุคคลรองลงมาเพียงเล็กน้อย โดยที่ผลลัพธ์จากขนาดภาพต้นฉบับมีความเที่ยงตรงเฉลี่ยกับภาพในเวลากลางวันสูงถึง 99.87% ภาพในเวลากลางคืน 92.46% และภาพรวมทั้งชุดข้อมูลสูงถึง 96.60% ด้วยความเร็วประมวลผลภาพ 23.64 ภาพต่อวินาที และสามารถเพิ่มความแม่นยำได้จากการลดขนาดภาพลง โดยมีค่าความเที่ยงตรงเฉลี่ยของภาพในเวลากลางวันเพิ่มขึ้นเป็น 99.93% ภาพในเวลากลางคืนเพิ่มขึ้นเป็น 94.06% และภาพรวมทั้งชุดข้อมูลเพิ่มขึ้นเป็น 97.46% ด้วยความเร็วประมวลผลภาพ 32.58 ภาพต่อวินาที

จากรูปที่ 4.5 เมื่อพิจารณาเปรียบเทียบผลลัพธ์จากกราฟแสดงความถูกต้องในการระบุความเสี่ยงบุคคลเมื่อเทียบกับชุดข้อมูลจริงทั้งหมดระหว่างโมเดล YOLOv4 ที่ถูกพัฒนาขึ้นในแต่ละชุดข้อมูล พบว่าโมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท มีความถูกต้องในการระบุความเสี่ยงน้อยที่สุด คือ 58.43% ในภาพขนาดต้นฉบับและ 64.11% เมื่อลดขนาดภาพลง ในขณะที่โมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลมีความถูกต้องในการระบุความเสี่ยงเพิ่มขึ้นมาเป็น 92.98% ในภาพขนาดต้นฉบับและ 93.78% เมื่อลดขนาดภาพลง และโมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มี

ความถูกต้องในการระบุความเสี่ยงสูงที่สุด คือ 93.23% ในภาพขนาดต้นฉบับและ 94.18% เมื่อลดขนาดภาพลง โดยตัวอย่างภาพผลลัพธ์จากการทดลองแสดงดังรูปที่ 4.6

ดังนั้นเมื่อพิจารณาด้วยความแม่นยำในการระบุตำแหน่งของบุคคลโดยรวมแล้ว โมเดลที่พัฒนาด้วยข้อมูลชุด MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด จึงมีความเหมาะสมที่จะนำไปพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์ที่สุด



รูปที่ 4.6 ตัวอย่างภาพผลลัพธ์การตรวจจับบุคคลและระบุระดับความเสี่ยงจากการทดลอง

ก) ตัวอย่างภาพการตรวจจับบุคคลและระบุความเสี่ยงสูงในเวลากลางวัน

ข) ตัวอย่างภาพการตรวจจับบุคคลและระบุความเสี่ยงปานกลางในเวลากลางวัน

ค) ตัวอย่างภาพการตรวจจับบุคคลและระบุความเสี่ยงสูงในเวลากลางคืน

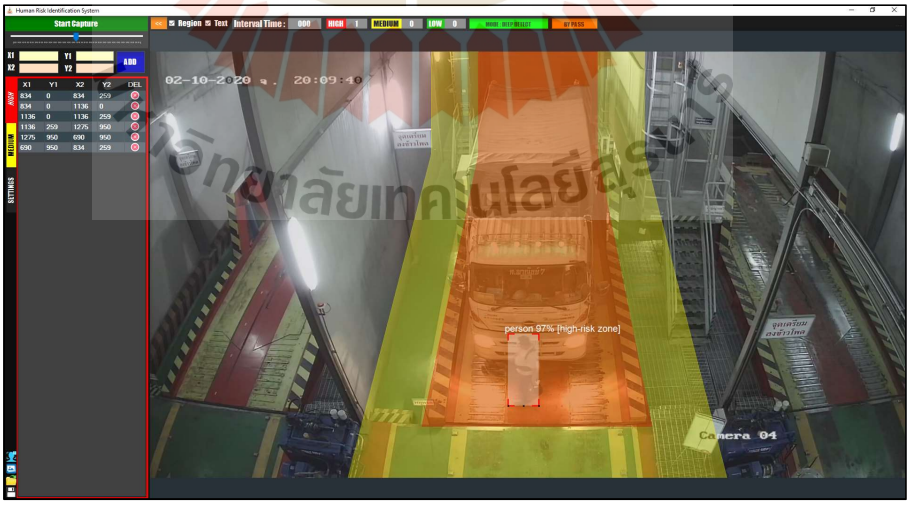
ง) ตัวอย่างภาพการตรวจจับบุคคลและระบุความเสี่ยงปานกลางในเวลากลางคืน

จ) ตัวอย่างภาพที่โมเดลไม่สามารถตรวจจับบุคคลในภาพได้

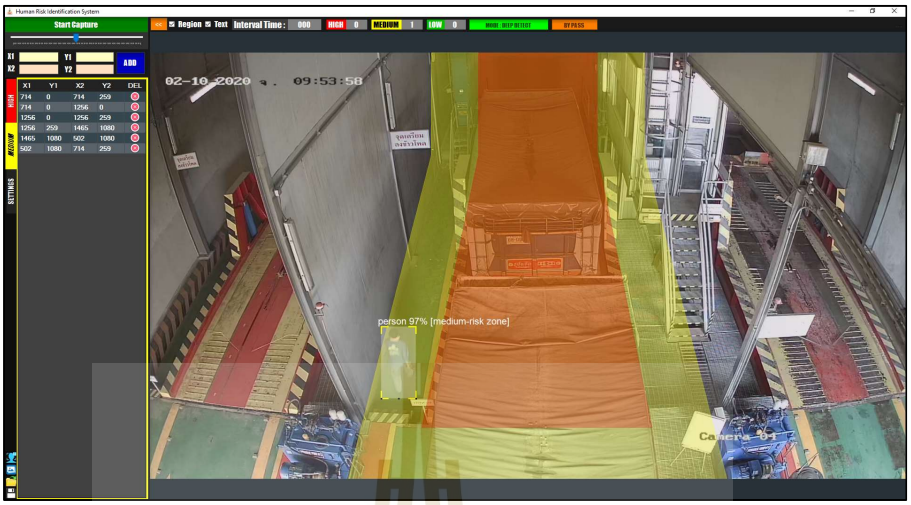
ฉ) ตัวอย่างภาพที่โมเดลตรวจจับบุคคลผิดพลาดจากวัตถุขนาดเล็ก

4.4 การพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยง

ผู้วิจัยได้พัฒนาระบบที่นำเสนอโดยนำโมเดลการตรวจจับบุคคลและอัลกอริทึมสำหรับระบุความเสี่ยงที่ได้จากการทดลองในหัวข้อ 4.3 มาพัฒนาโปรแกรมด้วยภาษา C# และใช้ซอฟต์แวร์ไลบรารี OpenCV ในการทำงาน โดยเชื่อมต่อกล้องวงจรปิดภายในโรงงานผ่าน Real Time Streaming Protocol (RTSP) ในการติดตั้งระบบครั้งแรกพนักงานจำเป็นต้องกำหนดขอบเขตของพื้นที่ความเสี่ยงสูงและความเสี่ยงปานกลางได้ผ่านทางหน้าจอโปรแกรม ดังรูปที่ 4.7 และรูปที่ 4.8 ในการทำงานของโปรแกรมจะเริ่มต้นจากการตรวจสอบภาพที่ได้รับจากกล้องวงจรปิดและประมวลผลทีละเฟรม และดำเนินการส่งผลลัพธ์ให้กับซอฟต์แวร์ไลบรารีส่วนกลางสำหรับส่งจำนวนบุคคลที่อยู่ในขอบเขตพื้นที่ความเสี่ยงสูงและพื้นที่ความเสี่ยงปานกลางให้กับ OPC Server เพื่อส่งข้อมูลต่อไปยังหน่วยความจำของ PLC และเมื่อโปรแกรมภายใน PLC ทำงาน จะเป็นการส่งสัญญาณการแจ้งเตือนไปยังหน้าจอ HMI หรือหยุดเครื่องจักรตามคำสั่งที่ได้กำหนดเงื่อนไขไว้ โดยที่ผลลัพธ์หลังจากการติดตั้งระบบนั้นมีความเร็วในการประมวลผลภาพเฉลี่ยอยู่ที่ 80.24 มิลลิวินาทีต่อภาพ และใช้เวลาในการส่งข้อมูลไปยังหน่วยความจำภายใน PLC เฉลี่ย 5.43 มิลลิวินาที ทั้งนี้ปัจจัยในการทำงานจริงขึ้นอยู่กับประสิทธิภาพโดยรวมของอุปกรณ์และปริมาณการรับส่งข้อมูลภายในเครือข่าย และเมื่อทดสอบกับกล้องวงจรปิดที่ติดตั้งในมุมมองที่ต่างออกไปโมเดลสามารถตรวจจับบุคคลได้อย่างแม่นยำได้ทุกมุมมอง ดังรูปที่ 4.9 แสดงให้เห็นว่าระบบที่ผู้วิจัยได้นำเสนอสามารถนำไปใช้งานในสถานที่ปฏิบัติงานจริงได้อย่างเหมาะสมและมีประสิทธิภาพ



รูปที่ 4.7 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์ในขอบเขตพื้นที่ความเสี่ยงสูง



รูปที่ 4.8 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์ในขอบเขตพื้นที่ความเสี่ยงปานกลาง

รูปที่ 4.9 ตัวอย่างการตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัดดัมป์ในมุมมองอื่น

4.5 อภิปรายผล

จากการทดสอบประสิทธิภาพของสถาปัตยกรรมการตรวจจับวัตถุด้วยข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด โดยแบ่งข้อมูลทดสอบออกเป็น 2 กลุ่ม คือ ภาพในเวลากลางวันและกลางคืน ใน 2 ขนาดภาพ และการทดสอบประสิทธิภาพการตรวจจับบุคคลและการระบุความเสี่ยงของโมเดลที่พัฒนาขึ้นเอง ทดสอบกับข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด โดยแบ่งข้อมูลออกเป็น 4 กลุ่ม คือ ภาพในเวลากลางวันและกลางคืน ในขอบเขตความเสี่ยงสูงและความเสี่ยงปานกลาง ใน 2 ขนาดภาพ สามารถสรุปการเปรียบเทียบประสิทธิภาพ และนำมาวิเคราะห์ได้ ดังต่อไปนี้

1) การทดลองเปรียบเทียบประสิทธิภาพสถาปัตยกรรมการตรวจจับวัตถุระหว่างสถาปัตยกรรม Faster R-CNN และ YOLOv4 นั้น พบว่าภาพที่ถูกลดขนาดลงนั้นมีประสิทธิภาพที่ดีกว่าทั้งในด้านความแม่นยำและความเร็วในการประมวลผลกับทั้ง 2 สถาปัตยกรรม โดยโมเดลที่ถูกพัฒนาด้วยสถาปัตยกรรม YOLOv4 มีความแม่นยำในการตรวจจับบุคคลเมื่อวัดจากค่าความเที่ยงตรงเฉลี่ยเท่ากับ 63.98% ซึ่งมากกว่า Faster R-CNN 0.73% จากชุดข้อมูลทดสอบภาพบุคคลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด โดยที่ความเร็วการประมวลผลภาพสูงถึง 31.97 ภาพต่อวินาที ซึ่งสูงกว่า Faster R-CNN ถึง 0.54 เท่า หรือคิดเป็น 14.88 ภาพต่อวินาที

2) การทดลองเปรียบเทียบประสิทธิภาพการตรวจจับบุคคลระหว่างโมเดลที่ถูกพัฒนาขึ้นจากชุดข้อมูลที่แตกต่างกันจำนวน 3 ชุด พบว่าภาพที่ถูกลดขนาดลงนั้นมีประสิทธิภาพที่ดีกว่าทั้งในด้านความแม่นยำและความเร็วในการประมวลผลกับทุกโมเดล เมื่อทดสอบกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด โดยที่โมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล มีความแม่นยำในการตรวจจับบุคคลมากที่สุดเมื่อวัดจากค่าความเที่ยงตรงเฉลี่ย โดยมีค่าเท่ากับ 97.68% ซึ่งมากกว่าโมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด เพียง 0.22% ซึ่งมีค่าเท่ากับ 97.46% และโมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท นั้นมีความแม่นยำน้อยที่สุด โดยมีค่าความเที่ยงตรงเฉลี่ยเพียง 64.11% ทั้งนี้ทุกโมเดลมีความเร็วในการประมวลผลภาพต่อวินาทีไม่แตกต่างกันมาก โดยมีค่าเท่ากับ 32.64 32.58 และ 32.56 ภาพต่อวินาที ตามลำดับ

3) การทดลองเปรียบเทียบความถูกต้องในการระบุความเสี่ยงจากตำแหน่งบุคคลระหว่างโมเดลที่ถูกพัฒนาขึ้นจากชุดข้อมูลที่แตกต่างกันจำนวน 3 ชุด ทดสอบกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด พบว่าภาพที่ถูกลดขนาดลงนั้นมีความถูกต้องในการระบุความเสี่ยงเพิ่มมากขึ้นเมื่อเปรียบเทียบจากจำนวนที่ระบุความเสี่ยงได้ถูกต้องกับจำนวนข้อมูลจริงทั้งหมด เนื่องจากการลดขนาดภาพลงทำให้โมเดลสามารถตรวจจับบุคคลได้มากขึ้นและ

ส่งผลให้กรอบล้อมรอบวัตถุที่สร้างขึ้นใกล้เคียงกับขนาดจริงมากยิ่งขึ้น โดยที่โมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มีความถูกต้องในการระบุความเสี่ยงบุคคลมากที่สุดซึ่งมีค่าเท่ากับ 94.18% เมื่อเทียบกับจำนวนข้อมูลจริงทั้งหมด ซึ่งมากกว่าโมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล 0.40% ซึ่งมีค่าเท่ากับ 93.78% และโมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท นั้นมีความถูกต้องในการระบุความเสี่ยงบุคคลน้อยที่สุด เนื่องจากมีจำนวนบุคคลที่ตรวจจับได้น้อยที่สุด โดยมีความถูกต้องในการระบุความเสี่ยงบุคคลเพียง 64.11% เมื่อเทียบกับจำนวนข้อมูลจริงทั้งหมด

บทที่ 5
บทสรุปและข้อเสนอแนะ

งานทางด้านการตรวจจับวัตถุนั้นมีความสำคัญมากในสาขาวิชาคอมพิวเตอร์วิทัศน์ซึ่งเป็นส่วนหนึ่งของการพัฒนาระบบคอมพิวเตอร์ให้สามารถมองเห็นและวิเคราะห์ภาพด้วยเทคโนโลยีทางด้านปัญญาประดิษฐ์ ในปัจจุบันจึงมีงานวิจัยจำนวนมากที่พยายามค้นคว้าและนำเสนอเทคนิคสำหรับเพิ่มความแม่นยำและความเร็วในการตรวจจับเพื่อให้สามารถนำมาประยุกต์ใช้กับงานในสถานการณ์จริงได้ทัดเทียมกับความสามารถของมนุษย์

ดังนั้นงานวิจัยนี้จึงมุ่งเน้นในการศึกษาและพัฒนาระบบการตรวจจับวัตถุโดยเฉพาะอย่างยิ่งการตรวจจับบุคคลภายในบริเวณพื้นที่อันตรายในสถานการณ์เสี่ยงที่มีโอกาสจะได้รับบาดเจ็บจากการทำงานของเครื่องจักร รวมถึงสถานที่หวงห้ามต่าง ๆ ภายในโรงงานที่อาจเกิดเหตุการณ์ที่ไม่ปลอดภัยต่อชีวิตมนุษย์ได้ โดยผู้วิจัยได้นำเสนอการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกที่ถูกนำมาใช้กับเทคโนโลยีการตรวจจับวัตถุในสถาปัตยกรรมต่าง ๆ ในปัจจุบัน โดยทำการวิจัยเปรียบเทียบประสิทธิภาพจากการทดสอบโมเดลต่าง ๆ กับชุดข้อมูลบุคคลในสถานที่ปฏิบัติงานจริงเพื่อศึกษาความเป็นไปได้ในการพัฒนาระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุโดยเฉพาะในระบบควบคุมทรัคดัมป์ เพื่อแจ้งเตือนและหยุดการทำงานของเครื่องจักรได้ทันทีก่อนที่จะเกิดอันตราย จากผลการศึกษาทั้งหมดสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

วัตถุประสงค์ของงานวิจัยนี้มุ่งเน้นไปที่การพัฒนาระบบเพื่อให้สามารถตรวจจับบุคคลในสถานการณ์จริงได้อย่างมีประสิทธิภาพ ผู้วิจัยจึงเริ่มต้นจากการคัดเลือกสถาปัตยกรรมการตรวจจับวัตถุที่มีความแม่นยำและความเร็วสูง โดยทดสอบจากข้อมูลบุคคลจากสถานที่ปฏิบัติงานจริงทั้งในเวลากลางวันและกลางคืนเพื่อทดสอบผลลัพธ์ให้ครอบคลุมและใกล้เคียงกับสภาพแวดล้อมการใช้งานจริงมากที่สุด โดยคัดเลือกโมเดลจากการทดสอบประสิทธิภาพของสถาปัตยกรรม Faster R-CNN และ YOLOv4 และนำไปพัฒนาเป็นโมเดลการตรวจจับบุคคลโดยฝึกสอนด้วยชุดข้อมูลที่แตกต่างกันเพื่อเปรียบเทียบประสิทธิภาพในด้านการตรวจจับ ความเร็วในการประมวลผล และความถูกต้องในการระบุความเสี่ยง โดยมีกระบวนการวิจัยดังต่อไปนี้

1) การเตรียมข้อมูลโดยนำข้อมูลวิดีโอคลิปที่ถูกบันทึกจากกล้องวงจรปิดมาแปลงให้เป็นไฟล์ภาพ โดยคัดเลือกเฉพาะภาพที่มีบุคคลอยู่ภายในรูปและดำเนินการระบุตำแหน่งบุคคลภายในรูปภาพเพื่อใช้เป็นคำตอบสำหรับฝึกสอนและทดสอบโมเดล

2) การแบ่งข้อมูลสำหรับการฝึกสอนและการทดสอบ โดยในส่วนแรกจะนำข้อมูลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด ทดสอบประสิทธิภาพของสถาปัตยกรรมการตรวจจับวัตถุ และในส่วนต่อไปจะนำข้อมูลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมดไว้ฝึกสอน และอีก 30% ของชุดข้อมูลทั้งหมด ไว้ทดสอบ

3) การวัดประสิทธิภาพโมเดลเพื่อคัดเลือกสถาปัตยกรรมการตรวจจับ ระหว่าง Faster R-CNN และ YOLOv4 โดยฝึกสอนด้วยชุดข้อมูล MS COCO ทั้งหมด 80 ประเภท และนำมาทดสอบกับชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 100% ของชุดข้อมูลทั้งหมด ด้วยตัวชี้วัด ค่าความเที่ยงตรง ค่าการระลึก ค่าประสิทธิภาพโดยรวม ค่าความเที่ยงตรงเฉลี่ย และความเร็วในการประมวลผลภาพต่อวินาที

4) การพัฒนาโมเดลและวัดประสิทธิภาพเพื่อคัดเลือกโมเดลการตรวจจับบุคคลจากชุดฝึกสอนในแต่ละแบบ คือ ชุดข้อมูล MS COCO โดยมีข้อมูลประเภทวัตถุทั้งหมด 80 ประเภท (MS COCO 80 Classes) ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคล (MS COCO Person) และ ชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด (MS COCO Person + Truckdump) และนำมาทดสอบกับชุดข้อมูลจากสถานที่ปฏิบัติงานจริง 30% ของชุดข้อมูลทั้งหมด ด้วยตัวชี้วัด ค่าความเที่ยงตรง ค่าการระลึก ค่าประสิทธิภาพโดยรวม ค่าความเที่ยงตรงเฉลี่ย ความเร็วในการประมวลผลภาพต่อวินาที และความถูกต้องในการระบุความเสี่ยง

5) การนำโมเดลที่มีประสิทธิภาพและเหมาะสมกับการใช้งานจริงมากที่สุดมาพัฒนาระบบตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อใช้กับระบบควบคุมทรัดดัมป์

หลังจากดำเนินการวิจัยดังกระบวนการที่กล่าวมาข้างต้น สามารถสรุปข้อมูลการวิจัยได้ดังตารางที่ 5.1 โดยสรุปผลได้ดังต่อไปนี้

โมเดลที่มีประสิทธิภาพการตรวจจับได้แม่นยำที่สุดคือโมเดลที่พัฒนาด้วยสถาปัตยกรรม YOLOv4 และยังมีความเร็วในการประมวลผลภาพที่สูงกว่า Faster R-CNN ถึง 1.87 เท่า กล่าวคือในเวลา 1 วินาที โมเดล YOLOv4 สามารถประมวลผลภาพได้เฉลี่ยเป็นจำนวน 31.96 ภาพ ซึ่งมีจำนวนมากกว่า Faster R-CNN ถึง 14.87 ภาพ และเมื่อพัฒนาโมเดลการตรวจจับบุคคลโดยใช้สถาปัตยกรรม YOLOv4 ด้วยชุดข้อมูลที่แตกต่างกันพบว่าโมเดลที่พัฒนาด้วยชุดข้อมูล MS COCO เลือกเฉพาะข้อมูลประเภทบุคคลรวมกับชุดข้อมูลภาพบุคคลจากสถานที่ปฏิบัติงานจริง 70% ของชุดข้อมูลทั้งหมด มีความถูกต้องในการระบุความเสี่ยงของบุคคลมากที่สุด โดยมีค่าความถูกต้อง

เท่ากับ 94.18% โดยมีค่าความเที่ยงตรงเฉลี่ยในการตรวจจับบุคคลสูงถึง 99.93% กับภาพในเวลากลางวันและ 94.06% กับภาพในเวลากลางคืน ซึ่งมีค่าความเที่ยงตรงเฉลี่ยเท่ากับ 97.46% เมื่อคิดจากชุดข้อมูลทั้งหมด

ตารางที่ 5.1 สรุปผลการทดสอบประสิทธิภาพโมเดลในงานวิจัยนี้ทั้งหมด

MODEL	TEST SET	AVG FPS	AP DAY	AP NIGHT	AP	RISK CORRECT (%)
Faster R-CNN MS COCO 80 Classes	100%	17.0958	63.0854	63.3128	63.2493	
YOLOv4 MS COCO 80 Classes	100%	31.9644	68.1349	58.8033	63.9820	
YOLOv4 MS COCO 80 Classes	30%	32.5561	71.8440	65.7173	69.1575	**64.1053**
YOLOv4 MS COCO Person	30%	32.6383	99.9339	94.2544	97.6780	93.7797
YOLOv4 MS COCO Person + Truckdump	30%	32.5804	99.9339	94.0626	97.4597	**94.1822**

5.2 การประยุกต์ผลการวิจัย

จากผลการวิจัยแสดงให้เห็นว่าระบบสำหรับตรวจจับบุคคลและระบุระดับความเสี่ยงเพื่อป้องกันอุบัติเหตุในระบบควบคุมทรัคดัมป์โดยใช้การเรียนรู้เชิงลึกที่ถูกพัฒนาขึ้นสามารถทำงานได้ดีในสถานที่ปฏิบัติงานจริงในสภาพแวดล้อมภายในโรงงานที่มีปัจจัยต่อความแม่นยำในการตรวจจับที่ลดลง เช่น ฝุ่นละอองสูง พื้นที่สภาพแสงน้อย ความคมชัดของสัญญาณภาพต่ำ หรือการติดตั้งกล้องในมุมสูง เป็นต้น ซึ่งนอกจากพื้นที่ภายในระบบควบคุมทรัคดัมป์แล้ว ยังสามารถนำไปประยุกต์ใช้กับพื้นที่เสี่ยงอื่น ๆ ภายในโรงงาน เช่น พื้นที่ที่กำลังดำเนินการก่อสร้าง พื้นที่หวงห้าม หรือพื้นที่อับอากาศ เป็นต้น โดยสามารถพัฒนาระบบแจ้งเตือนโดยเชื่อมต่อกับโปรแกรมภายนอก

(Application Programming Interface) ทดแทนการแจ้งเตือนผ่านหน้าจอควบคุมได้ ทั้งนี้ยังสามารถพัฒนาระบบเพื่อติดตั้งไปยังอุปกรณ์คอมพิวเตอร์ขนาดเล็กหรือเทคโนโลยีการเชื่อมต่อของสรรพสิ่ง (Internet of Things) โดยการพอร์ตซอฟต์แวร์ไปยังระบบปฏิบัติการอื่น ๆ โดยใช้ภาษา C++ หรือ Python ในการพัฒนาแทน

5.3 ข้อเสนอแนะ

จากผลการวิจัยสามารถสรุปประเด็นสำคัญต่อการนำไปปรับปรุงและพัฒนาต่อยอดได้ดังต่อไปนี้

1) เนื่องจากรูปภาพบุคคลในชุดข้อมูล MS COCO ที่ใช้ฝึกสอนนั้นมีความหลากหลายของรูปแบบข้อมูลอยู่มาก รวมถึงขนาดของบุคคลภายในรูปค่อนข้างหลากหลายจึงส่งผลให้โมเดลเกิดความสับสนกับวัตถุขนาดเล็กภายในภาพได้ การใช้ข้อมูลบุคคลจากสถานการณ์จริงในการฝึกฝนปริมาณที่มากขึ้นจึงส่งผลโดยตรงกับความซับซ้อนของลักษณะข้อมูลที่โมเดลต้องจดจำ การลดข้อมูลฝึกสอนที่ไม่จำเป็นและเพิ่มข้อมูลสถานการณ์จริงจึงทำให้โมเดลมีความแม่นยำมากยิ่งขึ้น

2) การตรวจจับที่ผิดพลาด (False Positive) นั้นจะส่งผลโดยต่อการการแจ้งเตือนที่ผิดพลาด (False Alarm) ในเบื้องต้นนั้นสามารถแก้ปัญหาได้ด้วยการปรับค่าเกณฑ์ IoU เพิ่มขึ้น แต่ผลที่ตามมาคือจำนวนการตรวจจับบุคคลไม่ได้ (False Negative) จะเพิ่มขึ้นแต่ในสถานการณ์ความเป็นจริงอาจจะไม่ส่งผลต่อระบบโดยรวมมากนักเพราะโมเดลสามารถตรวจจับความเคลื่อนไหวได้แบบต่อเนื่องโดยใช้เวลาเพียงเล็กน้อย และอาจนำเสนอการใช้เทคนิคแบบโหวตจากจำนวนเฟรมต่อเนื่องล่าสุดก่อนจะตัดสินใจ ในขณะเดียวกันจากผลการทดลองพบว่าการตรวจจับที่ผิดพลาดนั้นเกิดจากวัตถุขนาดเล็กเกินความจริงที่จะเป็นลักษณะของมนุษย์ได้ จึงสามารถนำเสนอการคัดกรองการตรวจจับที่ผิดพลาดเบื้องต้นด้วยการตั้งเกณฑ์คัดกรองวัตถุที่มีขนาดเล็กออก โดยคำนวณพื้นที่จากกรอบล้อมรอบวัตถุที่ตรวจจับได้ ซึ่งสามารถป้องกันการแจ้งเตือนที่ผิดพลาดได้

3) การป้องกันการแจ้งเตือนที่ผิดพลาดนั้นสามารถเพิ่มโมเดลในการจำแนกประเภทในการตรวจสอบขั้นที่ 2 ได้โดยฝึกสอนโมเดลด้วยข้อมูลเพียง 2 ประเภท ระหว่างวัตถุที่ใกล้เคียงมนุษย์และวัตถุที่เป็นมนุษย์ โดยใช้ภาพที่ตรวจจับได้มาพิจารณาก่อนส่งสัญญาณเตือน

4) การนำระบบไปใช้กับกล้องประเภทที่สามารถปรับองศาการมองเห็นได้นั้นอาจจะใช้เทคนิค Semantic Segmentation หรือ Instance Segmentation ร่วมกับระบบการตรวจจับบุคคล เพื่อเป็นการระบุพื้นที่บริเวรณพื้นหลังโดยเฉพาะอย่างยิ่งพื้นที่บริเวณทรัดดัมป์ ทั้งนี้จะเป็นการยกเลิกการตั้งค่าขอบเขตพื้นที่ความเสี่ยงในระดับต่าง ๆ ที่ต้องตั้งค่าในการใช้งานครั้งแรกได้ แต่อาจจะส่งผลต่อความเร็วในการประมวลภาพที่ต้องใช้เวลามากขึ้น และความแม่นยำในการระบุความเสี่ยงที่ลดลง

5) การเพิ่มความแม่นยำในการระบุความเสี่ยงบุคคลอาจทดลองใช้เทคนิค Perspective Transformation กับตำแหน่งวัตถุที่ตรวจจับได้ซึ่งทำมุมกับจุดติดตั้งกล้องวงจรปิด เพื่อทำการแปลงตำแหน่งจากมุมมองปัจจุบันเป็นตำแหน่งจากมุมมองด้านบน (Bird's-eye View) โดยจะสามารถใช้จุดกึ่งกลางเพียงจุดเดียวในการคำนวณตำแหน่งเพื่อระบุความเสี่ยงได้

รายการอ้างอิง

Adamo, F., Attivissimo, F., Cavone, G., & Giaquinto, N. (2007). SCADA/HMI systems in advanced educational courses. **IEEE Transactions on Instrumentation and Measurement**, 56(1), 4–10.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. **ArXiv Preprint ArXiv:2004.10934**.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. **Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 886–893.

Demuth, H., Beale, M., & Hagan, M. (1992). Neural network toolbox. **For Use with MATLAB. The MathWorks Inc**, 2000.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. **Nature Medicine**, 25(1), 24–29.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. **International Journal of Computer Vision**, 88(2), 303–338.

Gaushell, D. J., & Darlington, H. T. (1987). Supervisory control and data acquisition. **Proceedings of the IEEE**, 75(12), 1645–1658.

Girshick, R. (2015). Fast r-cnn. **Proceedings of the IEEE International Conference on Computer Vision**, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 580–587.

Henderson, P., & Ferrari, V. (2016). End-to-end training of object class detectors for mean average precision. **Proceedings of Asian Conference on Computer Vision**, 198–213.

Hernandez R., S. I., Ochoa M., P., & Ramirez V., J. C. (2007). Bluetooth and OPC (OLE for

Process Control) for the Distributed Data Integration. **Proceedings of Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)**, 27–32.

Hobbs, J. R. (2015). Pit-less truck dumper. Google Patents.

Htay, S., & Mon, S. S. Y. (2014). Implementation of plc based elevator control system. **International Journal of Electronics and Computer Science Engineering, IJECSE**, 3(2), 91–100.

Kim, J. H., Batchuluun, G., & Park, K. R. (2018). Pedestrian detection based on faster R-CNN in nighttime by fusing deep convolutional features of successive images. **Expert Systems with Applications**, 114, 15–33.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. **Proceedings of Advances in Neural Information Processing Systems**, 1097–1105.

LeCun, Y., Haffner, P., Bottou, L., Bengio, Y., Bottou, L., Haffner, P., Howard, P., Simard, P., Bengio, Y., LeCun, Y., & others. (1988). Object recognition with gradient-based learning. **Feature Grouping**, 66, 233–240.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. **Proceedings of European Conference on Computer Vision**, 740–755.

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 8759–8768.

Misra, D. (2019). Mish: A self regularized non-monotonic neural activation function. **ArXiv Preprint ArXiv:1908.08681**.

Nicola, M., Nicola, C.-I., Duta, M., & others. (2018). SCADA systems architecture based on OPC and Web servers and integration of applications for industrial process control. **International Journal of Control Science and Engineering**, 8(1), 13–21.

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. **Proceedings of 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)**, 237–242.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. **Proceedings of the IEEE Computer Society Conference on**

Computer Vision and Pattern Recognition, 779–788.

Redmon, J., & Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. **ArXiv Preprint ArXiv:1612.08242**.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. **ArXiv Preprint ArXiv:1804.02767**.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 39(6), 1137–1149.

Shim, S., & Choi, S.-I. (2019). Development on identification algorithm of risk situation around construction vehicle using YOLO-v3. **Journal of the Korea Academia-Industrial Cooperation Society**, 20(7), 622–629.

Uijlings, J. R. R., Van De Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. **International Journal of Computer Vision**, 104(2), 154–171.

Wang, C.-Y., Mark Liao, H.-Y., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., & Yeh, I.-H. (2020). CSPNet: A new backbone that can enhance learning capability of cnn. **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**, 390–391.

Xia, L., Chen, C.-C., & Aggarwal, J. K. (2011). Human detection using depth information by kinect. **Proceedings of CVPR 2011 Workshops**, 15–22.

Zengeler, N., Grimm, M., Borgmann, C., Jansen, M., Eimler, S., & Handmann, U. (2019). An evaluation of human detection methods on camera images in heavy industry environments. **Proceedings of 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)**, 205–210.

Zhang, L., Lin, L., Liang, X., & He, K. (2016). Is faster R-CNN doing well for pedestrian detection? **Proceedings of European Conference on Computer Vision**, 443–457.

Zhu, Q., Yeh, M.-C., Cheng, K.-T., & Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. **Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)**, 2, 1491–1498.

ภาคผนวก

บทความวิจัยที่ตีพิมพ์ระหว่างการศึกษา

รายชื่อบทความที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างการศึกษา

Apirak Worrakantapon, Kittisak Kerdprasop and Nittaya Kerdprasop. (2020). A Framework for Human Detection and Risk-Level Identification in Truck Dumper Control System. In **Proceedings of SUT International Virtual Conference on Science and Technology**, (pp. 156-162).

Apirak Worrakantapon, Wattana Pongsena, Kittisak Kerdprasop and Nittaya Kerdprasop. (2021). Real-Time Human Detection in a Restricted Area for Safety in Truck Dumper Control System Using Deep Learning. **International Journal of Electrical and Electronic Engineering & Telecommunications**, vol. 10, no. 1, (pp. 29-35).

SUT International Virtual Conference on Science and Technology
Nakhon-Ratchasima, Thailand
28th August 2020

EAT0027

A Framework for Human Detection and Risk-Level Identification in Truck Dumper Control System

Apirak Worrakantapon*, Kittisak Kerdprasop, and Nittaya Kerdprasop

Data and Knowledge Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand

* Corresponding Author: apirak.wor@gmail.com

Abstract. Any accident may cause a loss in human resources; thus, safety in industries should be primarily concerned. Truck dumpers are an industrial area for receiving raw materials from suppliers' trucks. The suppliers have to walk around the workplace while machines are working; this may cause serious accidents. Our conceptual idea to prevent accidents is that if there is a person close to the dangerous area, the machine should stop immediately and also alert workers. Therefore, a system to automatically detect and predict the existence of humans was necessary. In this paper, we introduce a novel framework for human detection and risk-level identification. This framework provides an intelligent monitoring system with the ability to detect persons through the camera images and be able to identify their locations that are divided into high-risk and medium-risk zones. The high-risk zone is the region at the truck dumper. The medium-risk zone is the region around the truck dumper. In our experiments, we measured precision and recall comparing between two real-time deep learning methods: YOLOv3 and YOLOv4. According to the experiment results, YOLOv4 showed better performance in both precision and recall. The processing speed of YOLOv4 was not much different from YOLOv3. Thus, we apply YOLOv4 in real situations due to efficient performance and suitable speed.

Keywords: Truck Dumper, Human Detection, Risk-Level Identification, Deep Learning, Accident Prevention.

1. Introduction

Safety in animal feed industries should be prioritized over other things to reduce risks occurring at workplaces, especially to prevent loss in human resources. In the industries, truck dumpers are risky areas for employees and suppliers. The suppliers convey raw material, e.g., corn, wheat grain, and soybean, to the truck dumpers. After the suppliers' trucks move to the truck dumpers area, the truck dumpers lift the whole truck and dump the material into the intake hoppers. Due to safety concerns, persons are prohibited to approach the area. However, accidents may occur because there are some blind corners in the area that the employees who control the dumping machine may not be able to notice human beings around. This probably

leads to an unpleasant accidental situation. Therefore, automatic person detection and risk identification system is necessary to be implemented in the risky area to prevent any fatal accident.

During past decades, there were several studies relating to industrial safety. Rafael Mosberger and teammates [1] proposed a system for detecting multiple humans to prevent collisions between vehicle and pedestrian by using NIR camera captured with IR flash to detect reflective clothing. Zdenek Kolar et al. [2] used transfer learning technique from VGG-16 [3] model based on deep convolutional neural networks to detect safety guardrail and inspect safety area in the construction industry. Hao Wu et al. [4] presented a new fire detection system applied in chemical factories and high- fire- risk industries using deep learning, i. e. , Convolutional Neural Networks (CNN) and You Only Look Once (YOLO) [5]. Their experimental results showed that the system could classify fire and fire-like images precisely; thus, it could avoid false alarm problems. Seungbo Shim et al. [6] developed a system to identify the risk situation of workers location based on YOLOv3 [7] for preventing injuries at the construction site. Apirak et al. [8] provided a real-time human detection system in a truck dumper control system. They used YOLO to detect humans and evaluated the system by comparing YOLOv2 [9] and YOLOv3. Experimental results from YOLOv3 outperformed YOLOv2 in terms of precision.

In this study, we propose a novel human detection and risk-level identification framework for the truck dumper control system using deep learning based on YOLOv3 and YOLOv4 [10]. It includes two stages. The first stage is human detection. The second stage is the risk-level identification based on the specific regions around the truck dumper area. We collect videos from surveillance cameras installed near the truck dumpers. The cameras recorded all events happening in the areas, and we found that there were a few people walking around the workplaces including the risky area. This system can detect people and identify human location. If someone is close to the risky area, the system should alert a warning to the employees in a control room. The objectives of this study are to prevent accidents that may occur in the truck dumper control system and to develop a new framework based on a deep learning technique to be implemented on the automation system.

2. A Novel Framework for Human Detection and Risk-Level Identification

In this study, the novel human detection and risk-level identification framework has been developed. The main purpose is to reduce fatal accidents that can affect humans by detecting human positions in a risky region and commanding machines to stop for safety in case that humans exist in the region.

In general, the surveillance camera is always installed in the frontal view of every truck dumper. This framework requires to set up the region of interest in each truck dumper for the human detection system and defines the medium-risk area and high-risk area for the risk-level identification system. Figure 1 illustrates an overview of this proposed framework which contains four main steps to be explained in the following subsections.

2.1. Data Collection and Preparation

The data used in this study had been collected by industrial surveillance cameras installed at the truck dumpers. The cameras record all events during daytime including the event with suppliers getting off the trucks (Figure 1(a)) and the event with the employees cleaning around the area (Figure 1(b)).

The prior assumption of our framework is that the obtained data should be ready and in a proper format for processing by the detection and identification systems. We thus prepare the

data by selecting the scenes containing at least one person in a frame and then extracting the specific video clips to be images in a frame-by-frame basis (Figure 1(a)).

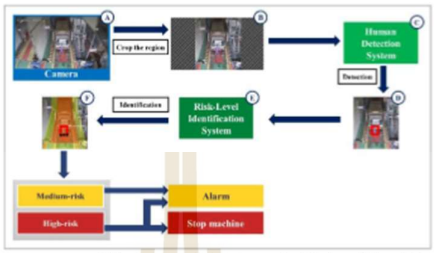


Figure 1. The processes of the human detection and risk-level identification systems

2.2. Human Detection and Risk-Level Identification

The proposed framework consists of two systems: human detection and risk- level identification. The human detection system is used to detect humans existing in the images with bounding boxes. The input data are the images extracted from the video clips containing humans. This system crops only the interested image region, which is the region around the truck dumper, and discards the rest of the images (Figure 1(b)). The human detection system applies YOLO to capture humans in the images (Figure 1(c)). This YOLO had been trained by the MS COCO dataset [11] containing objects from 80 classes or categories including the Person class that we used in this study. Other classes, except for the Person class, were ignored. Finally, the system provides output data, which are object class with a percentage of confidence score and a bounding box locating the human position (Figure 1(d)).

Another system was a risk-level identification (Figure 1(e). The input data are the output from the previous human detection system. This identification system has been configured to assign high-risk and medium-risk zones. The high-risk zone is the region right at the truck dumper because the heavy machine can lift itself up and down and may cause danger to living beings. Thus, this zone is strictly prohibited for persons, particularly during the machine is working. If there are persons in the high-risk zone, the system should alarm the employees in the control room and the lifting machine should stop immediately. In Figure 1F, we mark this high-risk zone with red color. Moreover, the medium-risk zone is also monitored as it is an area next to the high-risk zone. This area is not as strict as the high-risk zone. The persons are allowed to stay near the truck dumper but not too close. If there are persons in the area, the system should also alarm the employees in the control room, but the machines are still working. We set the zone color as yellow (Figure 1(f)).

To differentiate between high and medium risky level of the person in the dumper area, we define three points at the bounding box (Figure 2). The points are located at the lower-left corner of the bounding box, the lower-middle at the bottom edge, and the lower-right corner of the bounding box. If there are at least two points found to be in the high-risk zone, the system should identify the object in the bounding box as being at a high-risk level. In the same way, if there are at least two points found to be in the medium-risk zone, the system should identify the object in the bounding box as being at a medium-risk level.

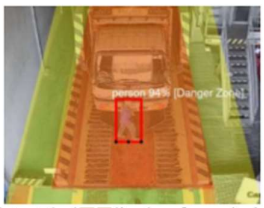


Figure 2. The three points at the bounding box for analyzing risk-level of a person

3. Experiment and Evaluation

3.1. Experimental Data

In the experiment, we used the collected data as mentioned in Section 2. We had recorded the video clips for 10 days, only during the daytime. After extracting the images from the video clips, we obtained totally 6,637 images that could be categorized as high-risk zone for 3,011 images and as medium-risk zone for 3,626 images. An image resolution was 1920x1080 pixels. We labeled the images by using LabelImg software [12] which is an efficient software to annotate graphical images for testing performance of the identification system.

3.2. Tools

The hardware and software used in this study described as follow:

Hardware
- CPU: Inter® Core™ i7-9750H 2.60GHz
- RAM: 16.0 GB
- GPU: NVIDIA GeForce RTX 2070

Software
- Windows 10 Pro
- NVIDIA CUDA 10.2 and cuDNN 8.0
- OpenCV 4.3.0

3.3. Evaluation

In this study, the performance of the proposed framework to detect humans was validated using precision and recall because the main target of this study was to detect the human in the specifics area. The precision represents how accurate the system predicts as compared to the correct result (or ground truth) of all detectable persons. The computation is shown in equation (1). The recall, as shown in equation (2), represents how well the system recalls or recognizes the true objects in the class by comparing correct results to all actual persons appearing in the data. Therefore, precision and recall are the two main metrics used for assessing the performance of the proposed framework of this study.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2)$$

where:

True positive (TP) is the number of persons being predicted correctly. False positive (FP) is the number of non-human objects being incorrectly predicted as "Person". False negative (FN) is the number of persons being undetectable by the system; i.e., a person is classified as "Not-a-Person" but the person actually exists in the images. True Negative (TN) is not used in

object detection problems. There are many possible bounding boxes that are not detected correctly.

4. Experimental Results

The experiments had been conducted on both YOLOv3 and YOLOv4 during the study. The obtained results were presented here by separating into two systems, i.e., human detection system and risk-level identification system. Performance of the human detection system is presented in Table 1, whereas the performance of the risk-level identification is illustrated in Table 2. Example results are also displayed in Figures 3 and 4.

Table 1. Experimental results of the human detection system.

	All Samples	TP	FP	FN	Precision	Recall
YOLOv3 medium-risk	3,626	2,290	9	1,336	0.996	0.631
YOLOv4 medium-risk	3,626	3,204	14	422	0.995	0.883
YOLOv3 high-risk	3,011	1,474	3	1,537	0.998	0.489
YOLOv4 high-risk	3,011	2,574	2	437	0.999	0.854

It can be seen from the experimental results that the human detection system provides high precisions on both YOLOv3 and YOLOv4 with precision value as high as 0.99. Our system can detect and predict persons correctly with only minor incorrect predictions. From the total of 6,637 images, wrong predictions are only 12 images for YOLOv3 and 16 images for YOLOv4. The incorrect predictions occurred in some specific images. For example, if there is a reflection on the truck's windscreen, then detection system misunderstood the reflection as the person. However, we observe that the bounding box of the incorrect result was presented with a low confidence value. Therefore, to mitigate this problem, we can increase threshold confidence to a higher value.

For the recall measurement, YOLOv4 outperforms YOLOv3 explicitly in that the average recall is 0.567 for YOLOv3 and 0.870 for YOLOv4. Thus, YOLOv4 can detect people more effective than YOLOv3. After considering the results, we found that there are some image characteristics that YOLO cannot detect correctly, for example, an image with a person holding a broom or equipment, an image with a person bending down, and an image with incomplete human structure due to a part of building obstructs the camera sight.

In this study, the image data had been fixed as a front-view received from a fixed angle camera. However, the camera angle may change if it is installed in different factories or locations. Due to the flexibility of this framework, it is possibly applicable to various image data even if they are from a different perspective because the camera settings are not the main factor for this framework process. In every first use of this framework, a user has to define the region of interest as low-, medium- and high-risk regions manually; thus, the framework process depends on how the user defines the regions.

Table 2. Experimental results of the risk-level identification.

	Ground Truths	Medium-Risk	High-Risk	Accuracy
YOLOv3 medium-risk	2,290	2,290	0	1.000
YOLOv4 medium-risk	3,204	3,203	1	0.999
YOLOv3 high-risk	1,474	64	1,410	0.956
YOLOv4 high-risk	2,574	194	2,380	0.924

On assessing performance of the risk-level identification system, we measure its accuracy, which is the ratio of correctly classified images (as high or medium risk) to all tested images. For both YOLOv3 and YOLOv4, we can observe high accuracy of 0.978 and 0.961, respectively. However, there are some cases that lead to mistaken identification. For example, a person extends arms that causes the bounding box to expand, and this the three points are shifted to another zone. This case should be identified as the high-risk level, but the system instead classifies the case as the medium-risk level.

For safety concern on timely warning in the real application, we also consider the system speed. The speed of YOLOv3 was 70.048 milliseconds per image and 78.643 milliseconds per image for YOLOv4. Even though YOLOv3 was faster than YOLOv4, the difference was insignificant. Therefore, YOLOv4 is appropriate to use for this study because of higher precision and recall as well as providing good speed.

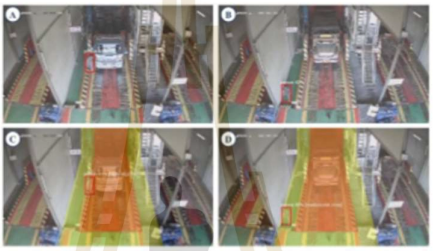


Figure 3. Example results from human detection and risk-level identification: (A) human detectable in high-risk zone, (B) human detectable in medium-risk zone, (C) high-risk zone identification and (D) medium-risk zone identification.

Figure 4. Example of false positive and false negative results from human detection system

Conclusion

In this study, we proposed a new framework to prevent accidents in susceptible industrial areas. The framework had been designed based on an intelligent paradigm to include two systems: human detection and risk-level identification. The main objectives were to prevent accidents at the truck dumper control system as well as to develop a new framework based on a deep learning technique implemented on the automation system. We conducted the experiments and compared the results between two deep learning systems: YOLOv3 and YOLOv4. We found that the human detection system implemented by YOLOv4 can detect persons precisely with the high precision rate at 0.87. For the risk-level identification system, it provided high accuracy on identifying high/medium risk level of human detected around the dumper area. Therefore, our framework has been proven efficient enough to be applicable in a real situation because it is reliable, accurate, and also provides a suitable processing speed.

References

[1] Mosberger R, Andreasson H and Lilienthal A J 2013 Multi-human tracking using high-visibility clothing for industrial safety *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* pp 638–44

[2] Kolar Z, Chen H and Luo X 2018 Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images *Autom. Constr.* **89** 58–70

[3] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *arXiv Prepr. arXiv1409.1556*

[4] Wu H, Wu D and Zhao J 2019 An intelligent fire detection approach through cameras based on computer vision methods *Process Saf. Environ. Prot.* **127** 245–56

[5] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: Unified, real-time object detection *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol 2016-December pp 779–88

[6] Shim S and Choi S-I 2019 Development on identification algorithm of risk situation around construction vehicle using YOLO-v3 *J. Korea Acad. Coop. Soc.* **20** 622–9

[7] Redmon J and Farhadi A 2018 YOLOv3: An incremental improvement *arXiv Prepr. arXiv1804.02767*

[8] Worrakatapon A, Pongsena W, Kerdprasop K and Kerdprasop N (in press) Real-time human detection in a restricted area for safety in truck dumper control system using deep learning *Int. J. Electr. Electron. Eng. Telecommun.*

[9] Redmon J and Farhadi A 2017 YOLO9000: Better, faster, stronger *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp 7263–71

[10] Bochkovskiy A, Wang C-Y and Liao H-Y M 2020 YOLOv4: Optimal speed and accuracy of object detection *arXiv Prepr. arXiv2004.10934*

[11] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L 2014 Microsoft coco: Common objects in context *European conference on computer vision* pp 740–55

[12] Tzutalin L Git code (2015) https://github.com/tzutalin/labelImg

International Journal of Electrical and Electronic Engineering & Telecommunications Vol. 10, No. 1, January 2021

Real-Time Human Detection in a Restricted Area for Safety in Truck Dumper Control System Using Deep Learning

Apirak Worrakantapon, Wattana Pongsena, Kittisak Kerdprasop, and Nittaya Kerdprasop
Data and Knowledge Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand
Email: apirak.wor@gmail.com; pongsena@hotmail.com; {kerdpras, nittaya}@sut.ac.th

***Abstract*—A process to receive raw materials from suppliers in an animal feed industry utilizes both automatic and semi-automatic machine control systems. The process called "truck dumper system" is the procedure that the suppliers provide raw materials carried by trucks; then, their tailgates open, and the raw materials are discharged by raising front end part of a truck to gather raw materials in a collection area. In general, the truck dumper system has been controlled manually by staff in a control room, not by a truck driver. However, serious accidents may occur during the process because when the dumper lifts up, the staff's vision has been blocked by the raised part of a truck. Therefore, if the staff controls the dumper to lift down by lacking safety awareness, people in the restricted area can be endangered. In this study, we proposed a framework of automatic human detection to prevent any accident that may occur from the truck dumper in the restricted area. The human detection model was developed to detect humans possibly in different blind corners that are difficult for staff in a control room to monitor these unseen areas for safety-awareness. The main technology of the proposed framework was the real-time human detection with fully convolutional neural network architecture called You Only Look Once, or YOLO. The framework has been designed to send a signal to terminate the truck dumper system immediately after the model detects people in the restricted area. In experiments, we discovered that the model could detect a human in all blind corners, including the corners that the staff's sight was completely bloacked by some barriers. The overall efficiency of this framework in an aspect of speed was high. The average time to process per image was 397 milliseconds by using CPUs and only 52 milliseconds by using GPUs. The results also showed that the model was effectively applicable to detect human in real-time due to its high-speed process.**

***Index Terms*—Truck dumper, human detection, safe-dumping system, deep learning, YOLO, convolutional neural network**

I. Introduction

In the present era of industry 4.0, more and more industrial factories have moved toward the smart and automation systems that in the previous decade have usually been operated by humans. However, which industrial procedures being able to be automated can be varied from industry to industry due to the differences in the systems. The focus of this work is automation process in the animal feed industry. Traditionally, a feedmill, which is a factory to produce animal feed, includes many processes: receiving raw material, storing and managing the material, mixing and pelleting, packing, and loading out to silo trucks. Indeed, the mentioned processes involve many heavy machines. Therefore, safety in the work place for employees and other people nearby is very important because accidents from heavy machinery can be easily occurred.

In the process of raw-material receiving, general factories obtain the raw material from suppliers conveyed by big dump trucks. The trucks dump the material to raw-material intake; then, the machine transfers the raw material to storage such as silos or bulk storages. Supplier's dump trucks can be distinguished into two main types: a truck with a self-dumping dump body and a truck fixed the dump body part. For the truck fixed the dump body part, the factories use truck dumper systems to lift the whole truck and dump the material into a designated area. During the process, this are is strictly prohibited by not allowing people to get close because it may cause fatal danger. To prevent any unpleasant accidental situation to occur, a staff in a control room is assigned a particular job to observe the workspace area for making sure that there is no human in the area. However, blind-side corners in the area are difficult to notice by the staff who monitors in a remote room.

We therefore propose an automation system to help securing the controlled access zone in the truck dumping system. As far as we know, this framework is the first proposal to turn a human work toward the automation process. The main benefit of our proposed framework is for human safety in a truck dumping system, especially in the animal feed factory. In the next section, preliminaries regarding human detection with convolutional neural network and You Only Look Once (YOLO) are briefly presented. Our proposed automated safety framework for human detection in the truck dumping system is explained in Section III. Section IV describes experimental details and evaluation metric. Experimental

Manuscript received January 2, 2020 revised March 5, 2020; accepted April 10, 2020.
Corresponding author: Apirak Worrakantapon (email: apirak.wor@gmail.com).

doi: 10.18178/ijeetc.10.1.29-35

results are illustrated in Section V. We finally conclude our work in Section VI.

II. PRELIMINARIES

In the past, computer vision based on various statistical and mathematical methods has been a popular technique for detecting humans in still images and video stream and for recognizing action [1]-[5]. However, recognition accuracy is sometimes unsatisfiable because light and sight aspects affect the performance of the computer vision techniques. Common solution for such problems are applying the background subtraction technique, but it is inapplicable to various budging objects. Another solution is object capturing with a 3D camera that has been proven satisfiable detected results; however, the 3D camera is costly [6], [7].

With the high performances of the central processing unit (CPU) and graphics processing unit (GPU), the complex deep learning algorithms such as convolutional neural networks (CNN) have been extensively applied to solve the human detection problem [8]-[11], [12]. The success of CNN application for object detection is because of the multi-layer network-based architecture that makes it excellent in extracting representative features from images [13], [14]. Its architecture contains two parts: a convolution part and a classification part. A convolution part is for feature extraction, whilst the classification part is for object recognition. However, CNN is still hard to use in real-time applications because of its huge time-consuming during the learning phase.

To improve time complexity during learning phase of a CNN deep learning method, a faster image recognition scheme called You Only Look Once, or YOLO, has been proposed by Redmon and colleagues in 2015 [15]. As the name suggested, YOLO speeds up typical CNN learning style by performing a one-stage scan over image and transforming an image classification problem to be a logistic regressional learning. Presently, YOLO has been adopted to solve a wide rage of machine vision and engineering problems that need object detection in real-time [16]-[22].

YOLO works faster than ordinary CNNs by locating objects in the image and classifying the objects at the same time. It firstly divides an image into SxS grids. Then, it predicts positions and size of interest objects and estimates their confidence scores for possibly object types on each grid simultaneously [15], [20]. This concept is shown in Fig. 1. The high confidence score represents the high probability of the target. This concept helps reducing time consumption potentially.

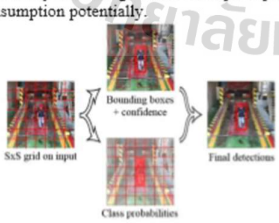


Fig. 1. The overall object detection concept of YOLO

In this study, we proposed a framework to adopt YOLO for real-time human detection in the restricted area for the safety of truck dumper control systems. The main focus of our framework is for preventing accidents that may harm people involved in truck dumper systems.

III. A SAFE-DUMPING SYSTEM FRAMEWORK

In this study, we proposed a framework (as shown in Fig 2) for preventing accidents during the raw-material receiving process managed by truck dumper control systems. This framework operates in a real-time scenario. Firstly, the system captures images from the IP Camera (step 1) and processes them using the human detection system implemented with YOLO (step 2). It counts the number of people appearing in the risky areas of the image. Then, the counting results are forwarded to the Object Linking and Embedding for Process Control Server, or OPC Server, system (step 3).

After that, the number of detected people were recorded in the Programmable Logic Controller (PLC) memory (step 4). We add additional conditions to a ladder logic in the PLC program to operate the truck dumper control system. The ladder logic is a program written in a graphical diagram that specifies particular conditions for signaling to the machine. Details of steps 1 throuh 4 are explained in the sub-sections: A, B, and C.

In the proposed framework, if a human detection system in step 2 can detect people in the restricted areas, that is, counting result is greater than zero. A signal should be submitted to the machine, and all machine operations must stop. If the specified memory in the PLC is equal to 0, the machine operation should resume to normal situations.

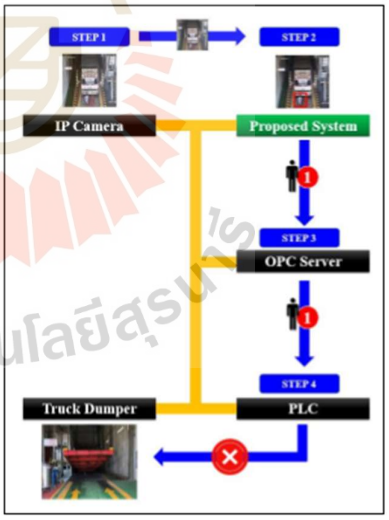


Fig. 2. A proposed framework for human safety in the truck dumping system in a feedmill factory

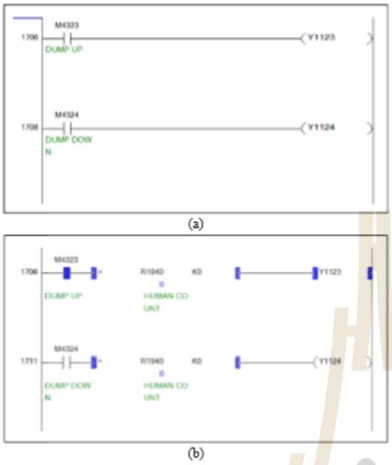


Fig. 3. PLC program: (a) Normal PLC program: interface of a program before adding conditions to a ladder logic and (b) a PLC program with safety conditions: a modified version according to the proposed framework that safety conditions have been added to a logic

Fig. 4. Illustration of the truck dumper: (a) machine in the idle stage, and (b) machine in the working stage (Photos courtesy: Kasetphand Industry Co., Ltd., Thailand)

Fig. 3 (a) shows the original PLC program that was implemented in the truck dumper system to control raw-material dumping in the animal feedmill factory. The program control two operations: dump up and dump down. The modification of this original PLC program to provide a safer environment for employees is the one shown in Fig. 3 (b). The number of human count as a result of human detection system has been added to the program. The new program controls truck dumping like the original version if the number of human count in the restricted area is zero. For some unusal situations that a human appears in the prohibited zone, the truck dumping operations will be blocked. Fig. 4 illustrates truck dumper system at work. In Fig. 4 (a), the dumper system is in the idle stage, whereas it is in the working stage (dump-up) in Fig. 4 (b).

A. Internet Protocol Camera (IP Camera) (Fig. 2 Step 1)

IP Camera is a type of digital video camera that can directly connect to a network system. It is usually used for surveillance camera around houses, organizations, or factories for security monitoring. This camera can be set as a web server; thus, all devices located in the same network can access the camera in real-time.

B. Proposed Safety System: Human Detection with Deep Learning (Fig. 2 Step 2)

The human detection system had been implemented by Python programming. Our system has been designed to access the camera streaming in real-time. Images are then imported to our YOLO model in order to detect human in the image. Comparing to other deep learning models such as CNN and Faster Region-based Convolutional Neural Networks (Faster R-CNN) [23], the YOLO model works faster with acceptable recognition accuracy.

C. OPC Server (Fig. 2 Step 3) and PLC (Fig. 2 Step 4)

OPC Server is a software interface standard used for communicating among devices in the network of the control system. It communicates with other devices through the Human Machine Interface (HMI), which is a graphical screen allowing persons to control the system comfortably.

PLC is a small modular device with multiple inputs and outputs (I/O) used for controling the machine. A traditional controler controls the machine via electrical circuit wiring which is difficult to modify and costly to maintain the circuit afterward. A better solution is to use PLC for machine control. Programmers or technicians can write a control program into a PLC memory and can easily modify the program subsequently. The PLC has its own CPU and input signal receivers including sensors for program processing. Also, it can generate output signals for activating the machine. In this study, we add a program into PLC to identify the conditions for filtering output signals to the machine.

HMI or OPC Client can send and receive both data and commands to OPC Server. Then, the server passess data and command to PLC memory. After that, the data and command in the memory are processed into the designed ladder logic. Finally, the machine address mapped by ladder logic should be activated. In this study, we implement a program as an OPC Client. It sends the number of detected people from the human detection system to OPC Server and records the number into the PLC memory for the program processing.

IV. Experiment and Evaluation

A. Data Preparation

The input data of this framework are a collection of images in the real situation. These images are extracted

International Journal of Electrical and Electronic Engineering & Telecommunications Vol. 10, No. 1, January 2021

from video clips captured by the IP Camera installed at the animal feedmill factory in Thailand. The data set of this work contains 300 images with only one perspective camera but the image data contain three different human positions in the images, i.e., a human at left-side of the image (Fig. 5 (a)), a human at the center of the image (Fig. 5 (b)), and a human at right-side of the image (Fig. 5 (c)).

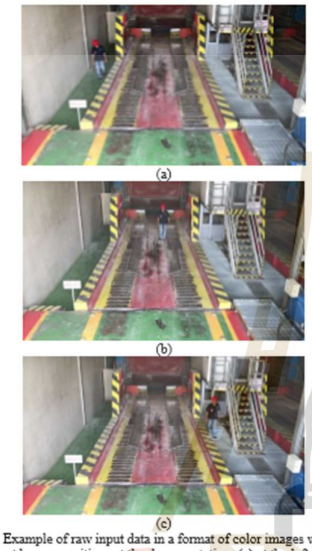


Fig. 5. Example of raw input data in a format of color images with three different human positions at the dumper station: (a) at the left-side, (b) at the center, and (c) at the right-side.

The image with a human at the left side of the dumper station represents a situation that there is a human in the prohibited area which is the opposite side of the control room. The human at the center of the image represents scenario that there is a human standing at the top of the truck dumper. The human at the right side of the image represents the event that a human appears at the side of the control room. We collect 100 images of each of the three scenarios. The image resolution was 1928×1080 pixels.

B. Tools used in the Experiment

The hardware and software tools used for training the deep learning model and for testing the recognition accuracy of the human detection system are listed as follws.

- Hardware
  - OS: Windows 10 Pro
  - CPU: Inter® Core™ i7-9750H 2.60GHz
  - RAM: 16.0 GB
  - GPU: NVIDIA Geforce RTX 2070
- Software
  - NVIDIA CUDA 10.1 and cuDNN 7.6

C. Evaluation

In this research, the performance of the deep model for detecting human in the restricted area was evaluated using precision as a measurement metric for assessing performance. The precision can be calculated as the proportion between the number of results that the human detection system predicts accurately and the total number of the test data. The calculation is shown in equation (1).

$$\text{Precision} = \frac{\text{True positive}}{\text{Actual results}} \quad (1)$$

where True positive is the number of positive results that the model predicts accurately and Actual results are the total number of the test data.

V. Experimental Results

Our input data was collected from the fixed-perspective IP Camera with three different human positions (as shown some examples in Fig. 5). After inquiring about possible unsafe situations from the factory staff, we found that only two human positions had been blocked from the staff's vision: a human at the left-side of the image and a human at the center of the image. These may lead to serious accidents if any people exist in the restricted area while the truck dumper system is activated. Fortunately, the model based on deep learning can detect people who are in the restricted area from all positions.

Results of human detection at all three positions are summarized and presented in Table I. We implement both YOLO versions 2 and 3, and then compare human detection performance against a human staff working in a control room. Results from YOLOv3 human detection for both detectable and undetectable cases are demonstrated in Fig. 6.

TABLE I. Results Showing Effectiveness of the Model for Human Detection in the Restricted Areas

Experimentation		Human Staff	Model-based YOLOv2	Model-based YOLOv3
left-side	Precision	0.00	0.85	1.00
	Processing Time (milliseconds)	-	203	399
center	Precision	0.00	0.72	1.00
	Processing Time (milliseconds)	-	209	396
right-side	Precision	1.00	0.60	0.91
	Processing Time (milliseconds)	-	195	398

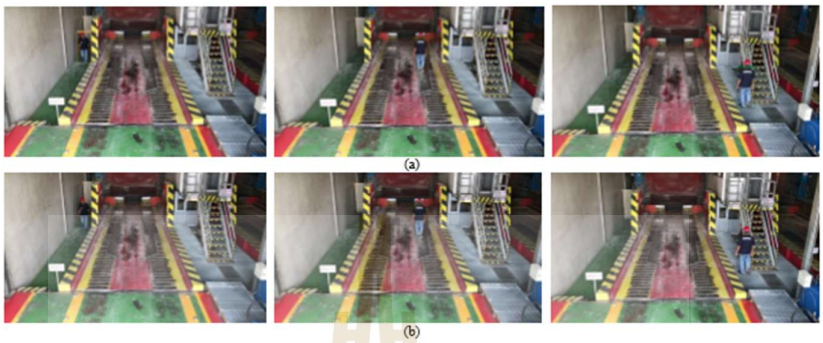


Fig. 6. Performance of a human detection model: (a) detectable a human image at the left-side, center and right-side, and (b) undetectable a human image at the left-side, center and right-side.

In terms of processing time, the experimental results in Table I show that the model based on YOLOv2 can recognize human faster than the model based on YOLOv3, approximately two times faster. In the aspect of human safety, the speed of the model is highly important because it obviously helps reducing any accident that may occur when people enter the restricted area. Once the human is detected, the control system can automatically pause the dumper machine immediately.

In terms of accuracy, the results demonstrate that the model based on YOLOv3 provides significantly higher accuracy than the model based on YOLOv2. The model based on YOLOv2 provides the average precision for automatic detection in three different human positions as 72%, whereas the average accuracy of YOLOv3 is as high as 97%. The precision of the YOLOv2 model to detect the human at the left-side, center, and right side of the image was 85%, 72%, and 60%, respectively. YOLOv3 can detect human at the same positions with accuracy rate 100%, 100%, and 91%, respectively. It can be observed from the result that the model based on YOLOv3 provides the average precision significantly higher than the model based on YOLOv2 up to 25%.

The main objective of this study was to develop the framework to prevent accidents in the industrial factory that might occur in the restricted area. If the accidents in the restricted area really happen in the truck dumper area, it will risk a human's life. Thus, we select the model based on YOLOv3 due to its high precision and acceptable speed. To compare the results obtained by staff monitoring, our model can detect the human at the blind-side of the image in which the staff cannot notice human in that area. Therefore, we are confident to apply this framework in a real workplace of truck dumper.

VI. Conclusion

This work presents the design of a framework for detecting human unexpected to appear in the prohibited zone of truck dumping in the feedmill factory. The main purpose of this design is for improving safety in the factory. The automatic process of the proposed framework is anticipated to provide a better safety system than the current human staff controling system. The experimental results of this study show that the proposed framework has a high potential to prevent serious accidents for people working in the restricted area.

In the proposed framework, we adopt a deep learning model to detect the number of people in all three possible positions. The model can detect human in restricted area with precision as high as 97%. The precision of the proposed model can possibly be further improved by increasing the number of images in the training dataset. However, the current model is effectively applicable even when the human appears in blind spots that invisible by the staff in a control room. The processing time of the model is quite low with an average time per image of 397 milliseconds. Moreover, it can process images in real-time with the GPU that a rate of processing speed was up to 19.2 images per second. This indicated that the proposed framework with the deep learning model can be used in real-world situations.

Based on the success of this preliminary design and implementation, the proposed idea can be deployed not only in the truck dumper control system of the feedmill factory, but the idea can also be applied to many production and manufacturing sectors. Since the main concern of the proposed framework is for improving safety for workers in the factory, any kind of application that seeks for safety improvement is thus able to adopt and adjust the proposed fraework to suit a specific application area. For instance, the safety control system in loading/unloading goods in large warehouses is one area of application.

In future work, we plan to improve the performance of the human detection model that can work under different environments such as different illumination conditions. Such conditions can normally occur in a workplace of our truck dumper control system, for example, the blur scene during the cloudy days. Therefore, some preprocessing steps to handle sub-optimal illumination should be helpful for the increase in accuracy of our system. We

also plan to include a classifier to distinguish between outsiders and employees. Even though this aspect has no direct effect toward the performance of our system, such extension is more or less improving the intelligent level of our system. Then, we plan to combine this framework of human-detection and a notification system to notify security guards when people invade the prohibited areas, for instance, a confined space and under-construction area, to reduce the probability of accidents. This kind of extension is necessary for a practical application aspect of the proposed framework to aid other kinds of workplaces such as hardware manufacturing, construction sites, and many dangerous industries.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Apirak Worrakantapon conducted the research and experiments, analyzed the data as well as proposed the method of the study; Apirak Worrakantapon and Wattana Pongsena wrote the paper; Kittisak Kerdprasop and Nittaya Kerdprasop proved the correctness of the results and had approved the final version of this manuscript

REFERENCES

[1] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, vol. 2, pp. 1491–1498.

[2] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proc. European Conference on Computer Vision*, 2006, pp. 428–441.

[3] C. Raghavachari, V. Aparna, S. Chithira, and V. Balasubramanian, "A comparative study of vision based human detection techniques in people counting applications," *Procedia Computer Science*, vol. 58, pp. 461-469, 2015.

[4] H. Zhang, Y. Zhang, B. Zhong, Q. Lei, L. Yang, J. Du, and D. Chen, "A comprehensive survey of vision-based human action recognition methods," *Sensors*, vol. 19, 2019.

[5] J. K. Agrawal and M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Survey*, vol. 43, nn. 3, 2011.

[6] X. Yang and Y. Tian, "Effective 3D action recognition using EigenJoints," *Journal of Visual Coomunication and Image Reoresentation*, vol. 25, no. 1, pp. 2-11, 2014.

[7] D. I. B. Hagebeuker and P. Marketing, "A 3D time of flight camera for object detection," *PMD Technol. GmbH, Siegen*, 2007.

[8] J. Zhao, G. Zhang, L. Tian, and Y. Q. Chen, "Real-time human detection with depth camera via a physical radius-depth detector and a CNN descriptor," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 1536–1541.

[9] J. H. Kim, H. G. Hong, and K. R. Park, "Convolutional neural network-based human detection in nighttime images using visible light camera sensors," *Sensors*, vol. 17, no. 5, 2017.

[10] T. Liu and T. Stathaki, "Faster R-CNN for robust pedestrian detection using semantic segmentation network," *Frontiers in Neurorobotics*, vol. 12, 2018.

[11] N. AlDahoul, A. Q. M. Sabri, and A. M. Mansoor, "Real-time human detection for aerial captured video sequence via deep models," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[12] X. Wang and S. Hosseinyalamdary, "Human detection based on a sequence of thermal images using deep learning," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W13, pp. 127-132, 2019.

[13] D. Strigl, K. Kofler, and S. Podlipnig, "Performance and scalability of GPU-based convolutional neural networks," in *Proc. 18th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 2010, pp. 317–324.

[14] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 650–657.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[16] R. Laroca, *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–10.

[17] Q. Peng, *et al.*, "Pedestrian detection for transformer substation based on gaussian mixture model and YOLO," in *Proc. 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2016, vol. 2, pp. 562–565.

[18] D. Shen, X. Chen, M. Nguyen, and W. Q. Yan, "Flame detection using deep learning," in *Proc. 4th International Conference on Control, Automation and Robotics (ICCAR)*, 2018, pp. 416–420.

[19] C. Tang, Y. Ling, X. Yang, W. Jin, and C. Zheng, "Multi-view object detection based on deep learning," *Applied Sciences*, vol. 8, no. 9, 2018.

[20] J. Li, J. Gu, Z. Huang, and J. Wen, "Application research of improved YOLO v3 algorithm in PCB electronic component detection," *Applied Sciences*, vol. 9, 2019.

[21] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417-426, 2019.

[22] X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, and L. Kurach, "A deep learning-based framework for an automated defect detection system for sewer pipes," *Automation in Construction*, vol. 109, 2020.

[23] C. Han, G. Gao, and Y. Zhang, "Real-time small traffic sign detection with revised faster-RCNN," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13263-13278, 2019.

Apirak Worrakantapon is a master student, School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. He received his B.E. in computer engineering from SUT, in 2008. He has long experiences with automation system in the feedmill factory. His current research of interest includes software engineering, automation system, artificial intelligence, machine learning and deep learning.

Watthana Pongsena is a Ph.D.student at School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. He received his B.E. and M.E. in computer engineering from SUT in 2008 and 2012, respectively. His research of interest includes software engineering, data mining, artificial intelligence, and human-computer interaction.

Kittisak Kerdprasop is an Associate Professor and Chair, School of Computer Engineering, SUT. He received his bachelor's degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes machine learning and artificial intelligence.

Nittaya Kerdprasop is an associate professor and the head of Data and Knowledge Engineering Research Unit, School of Computer Engineering, SUT. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her research of interest includes data mining, artificial intelligence, logic, and constraint programming.

ประวัติผู้เขียน

นายอภิรักษ์ วรกานตพล เกิดเมื่อวันที่ 19 ตุลาคม พ.ศ. 2529 ที่อำเภอเมือง จังหวัดราชบุรี เริ่มเข้ารับการศึกษาระดับชั้นอนุบาล 1 ที่โรงเรียนวัดไผ่ล้อม(เจริญราษฎร์วิทยาคม) อำเภอเมือง จังหวัดราชบุรี จนจบการศึกษาระดับประถมศึกษาชั้นปีที่ 6 จากนั้นได้เข้าศึกษาต่อในระดับมัธยมศึกษาตอนต้นที่โรงเรียนดรุณานุเคราะห์ อำเภอบางคนที จังหวัดสมุทรสงคราม และศึกษาต่อในระดับมัธยมศึกษาตอนปลายที่โรงเรียนเบญจมราชูทิศ ราชบุรี อำเภอเมือง จังหวัดราชบุรี ในปีการศึกษา 2548 ได้เข้าศึกษาระดับปริญญาตรีในสาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี อำเภอเมือง จังหวัดนครราชสีมา และได้สำเร็จการศึกษาเมื่อปี พ.ศ. 2552

ภายหลังจากสำเร็จการศึกษาในระดับปริญญาตรี ได้เข้าทำงานที่บริษัท KPMG Phoomchai Holdings Co., Ltd. ในตำแหน่ง Web Developer เมื่อปี พ.ศ. 2552-2553 จากนั้นได้เข้าทำงานที่บริษัท CPF (Thailand) Public Co., Ltd. ในตำแหน่งผู้ชำนาญการพัฒนาสารสนเทศด้านเทคโนโลยีอัตโนมัติ เมื่อปี พ.ศ. 2552-2562

ในปี พ.ศ. 2562 ได้เข้ารับการศึกษาในระดับปริญญาโท สาขาวิชาวิศวกรรมโทรคมนาคมและคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี โดยทำการวิจัยในด้านปัญญาประดิษฐ์เพื่อพัฒนาความสามารถในการมองเห็นของคอมพิวเตอร์ด้วยการใช้เทคนิคการเรียนรู้เชิงลึกกับงานทางด้านการตรวจจับวัตถุและนำไปประยุกต์ใช้กับระบบควบคุมเครื่องจักรแบบอัตโนมัติ โดยในระหว่างการศึกษาได้รับการอนุเคราะห์เป็นอย่างดีจากอาจารย์ที่ปรึกษาและอาจารย์ประจำรายวิชาต่าง ๆ และได้ทำการตีพิมพ์บทความวิชาการซึ่งมีรายละเอียดสามารถดูได้ที่ภาคผนวก