การวิเคราะห์ภาพคนและสัมภาระสำหรับการตรวจจับวัตถุที่ปราศจากเจ้าของ

ปฏิญญา ตันฑวิวัฒน์[1] ไตรรัตน์ สบายใจ[2] ดัชกรณ์ ตันเจริญ[3] ณัฐชัย วัชราภินชัย[4] ศีตภา รุจิเกียรติกำจร[5*]

[1,2,3]คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์, นนทบุรี, ประเทศไทย
[4,5]ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, ปทุมธานี, ประเทศไทย*

*ผู้ประพันธ์บรรณกิจ อีเมล: sitapa.ruj@nectec.or.th

รับต้นฉบับ: 17 พฤษภาคม 2564; รับบทความฉบับแก้ไข: 17 กรกฎาคม 2564; ตอบรับบทความ: 31 สิงหาคม 2564
เผยแพร่ออนไลน์: 13 ธันวาคม 2564

บทคัดย่อ

บทความนี้นำเสนอวิธีการวิเคราะห์ภาพวิดีโอคนและสัมภาระสำหรับการตรวจจับวัตถุที่ปราศจากเจ้าของ และตรวจจับบุคคลที่มีการถือกระเป๋าประเภทต่าง ๆ เนื่องจากเหตุการณ์วางทิ้งกระเป๋าเป็นเหตุการณ์ที่มีความเสี่ยงและเกิดขึ้นได้จริงโดยเฉพาะพื้นที่สาธารณะที่ต้องการเฝ้าระวังเป็นพิเศษ เช่น สถานีรถไฟ สนามบิน หรือสถานที่สำคัญภายในอาคาร เป็นต้น การตรวจจับเพื่อระบุตำแหน่งคนและกระเป๋าใช้วิธีการเรียนรู้เชิงลึกที่ถูกฝึกสอนด้วยภาพจำนวน 12,000 ภาพ ที่ประกอบด้วยภาพคนและกระเป๋า เช่น กระเป๋าเป้ กระเป๋าถือ และกระเป๋าเดินทาง งานวิจัยนี้ใช้แบบจำลอง YOLOv3 สำหรับการตรวจจับวัตถุซึ่งประมวลผลได้แบบเรียลไทม์และมีประสิทธิภาพความถูกต้อง (Accuracy) ถึง 98% ส่วนการวิเคราะห์ความเป็นเจ้าของ และการละทิ้งกระเป๋าจะใช้วิธีการพิจารณาความสัมพันธ์ในเชิงตำแหน่งและการเคลื่อนที่ของคนและกระเป๋า ซึ่งประสิทธิภาพการระบุความเป็นเจ้าของได้ค่าความถูกต้อง (Accuracy) 65.1% และประสิทธิภาพการระบุการละทิ้งวัตถุที่ 66.6%

คำสำคัญ : การตรวจจับวัตถุที่ปราศจากเจ้าของ การวิเคราะห์การเคลื่อนที่ การประมวลผลภาพ การเรียนรู้ของเครื่องจักร

Human and Luggage Analysis for Abandoned Object Detection

Patinya Tantawiwat[1] Trairat Sabaichai[2] Datchakorn Tancharoen[3]
Nattachai Watcharapinchai[4] Sitapa Rujikietgumjorn[5*]

[1,2,3]*Faculty of Engineering and Technology, Panyapiwat Institute of Management (PIM), Nonthaburi, Thailand*
[4,5*]*National Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand*

*Corresponding Author. E-mail address: sitapa.ruj@nectec.or.th

Received: 17 May 2021; Revised: 17 July 2021; Accepted: 31 August 2021
Published online: 13 December 2021

Abstract

This paper presents a method to analyze videos for Abandoned Object Detection (AOD) and ownership identification. Abandoned luggage is a risky event that can actually occur in many public areas such as train stations, airports, or other important places in the building. Deep learning is used for person and luggage detection. The dataset is trained for both people and luggage images including backpacks, handbags and suitcases for more than 12,000 images. In this work, we use the YOLOv3 model that can be processed in real time with 98% accuracy. For ownership identification and abandoned detection, we propose a method that evaluates the spatial relationships and trajectories of each person and luggage. Ownership identification yields 65.1% accuracy while abandoned detection provides 66.6% accuracy.

***Keywords*:** Abandoned object detection, Motion analysis, Image processing, Machine learning

1) บทนำ

ความก้าวหน้าของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) และการเรียนรู้ของเครื่อง (Machine Learning) มีประสิทธิภาพความถูกต้องและความแม่นยำสูงขึ้น ก่อให้เกิดการต่อยอดเทคโนโลยีบนระบบกล้องวงจรปิดอัจฉริยะ (Intelligent CCTV System) ที่สามารถทำการวิเคราะห์ภาพได้อย่างอัตโนมัติ เนื่องจากปัจจุบันจำนวนกล้องมีปริมาณมากขึ้น ทำให้ผู้ดูแลไม่สามารถสังเกตเหตุการณ์ผ่านระบบกล้องวงจรปิดแบบเดิม และแจ้งเตือนเหตุการณ์ต่าง ๆ ได้ทั้งหมด จึงเป็นหน้าที่ของเทคโนโลยีปัญญาประดิษฐ์ที่จะมีส่วนช่วยผู้ดูแลสถานที่ให้มีความปลอดภัยยิ่งขึ้น

การตรวจจับเหตุการณ์วางกระเป๋าหรือสิ่งของโดยปราศจากผู้ครอบครอง เป็นตัวอย่างเหตุการณ์ที่เกิดขึ้นจริงในระบบกล้องวงจรปิด และเป็นที่ต้องการเพื่อนำมาใช้ในพื้นที่สาธารณะในการเฝ้าระวังสถานที่บางแห่งเป็นพิเศษ เช่น สถานีรถไฟ สนามบิน หรือพื้นที่ภายในอาคาร โดยจะเป็นการวิเคราะห์การวางวัตถุโดยปราศจากผู้ครอบครอง (Abandoned Objects Detection หรือ AOD) หมายถึง การวิเคราะห์การเคลื่อนที่ของวัตถุในสถานการณ์ที่ว่าบุคคลในภาพนั้นมีความประสงค์จะทอดทิ้งวัตถุนั้นหรือไม่ [1]–[3] ภายใต้การใช้งานในสถานการณ์จริง AOD ยังคงมีความท้าทายในแง่ของเทคโนโลยีหลายประการ ได้แก่ 1) วัตถุที่ต้องการวิเคราะห์มีความหลากหลาย เช่น กระเป๋า กล่อง หรือถุง เป็นต้น 2) สถานการณ์จริงมักจะมีคนอยู่พลุกพล่านทำให้เกิดการบังวัตถุและยากต่อการวิเคราะห์ความเป็นเจ้าของ 3) ปัจจัยที่เกี่ยวข้องกับคุณภาพของภาพ เช่น ภาพที่มีสัญญาณรบกวน ภาพที่มีคุณภาพต่ำ เป็นต้น 4) ปริมาณกล้องที่ต้องการวิเคราะห์ที่มีอยู่อย่างมหาศาล แต่หน่วยประมวลผลมีจำกัด

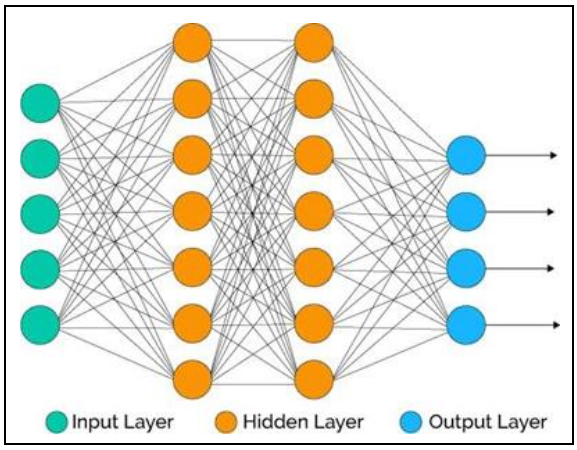
สำหรับงานวิจัยนี้ มีวัตถุประสงค์เพื่อให้สามารถตรวจจับเหตุการณ์วางกระเป๋าโดยปราศจากเจ้าของ และตรวจจับบุคคลที่มีการถือกระเป๋าประเภทต่าง ๆ ได้แก่ กระเป๋าเป้สะพายหลัง (Backpack) กระเป๋าถือ (Handbag) และกระเป๋าเดินทาง (Suitcase) งานวิจัยนี้นำเสนอวิธีการระบุความเป็นเจ้าของสัมภาระ (Owner Analysis) และการระบุวัตถุที่ปราศจากผู้ครอบครอง (Abandoned Object Detection) โดยใช้การวิเคราะห์ความสัมพันธ์ในเชิงตำแหน่งและการเคลื่อนที่ของคนและกระเป๋า และมีการนำเสนอการติดตามวัตถุ (Object Tracking) เพื่อใช้ในการติดตามการเคลื่อนที่ของคนและกระเป๋า นอกจากนั้นมีการเปรียบเทียบแบบจำลองที่เหมาะสมสำหรับการตรวจจับวัตถุเพื่อให้สามารถประมวลผลได้เร็วและมีความถูกต้อง โดยตัวอย่างการตรวจจับภาพคนกับกระเป๋า และตัวอย่างความเป็นเจ้าของ และการละทิ้งกระเป๋าแสดงในรูปที่ 1

รูปที่ 1 : ตัวอย่างการตรวจจับภาพคนกับกระเป๋า ความเป็นเจ้าของ และการละทิ้ง

2) ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1) โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network) มีหลักการทำงานโดยจำลองมาจากสมองของสิ่งมีชีวิตที่มีความซับซ้อน โดยใช้โมเดลทางคณิตศาสตร์ ซึ่งแนวคิดเริ่มต้นได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพในสมอง โครงข่ายประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทจนเป็นเครือข่ายที่ทำงานร่วมกัน ดังรูปที่ 2

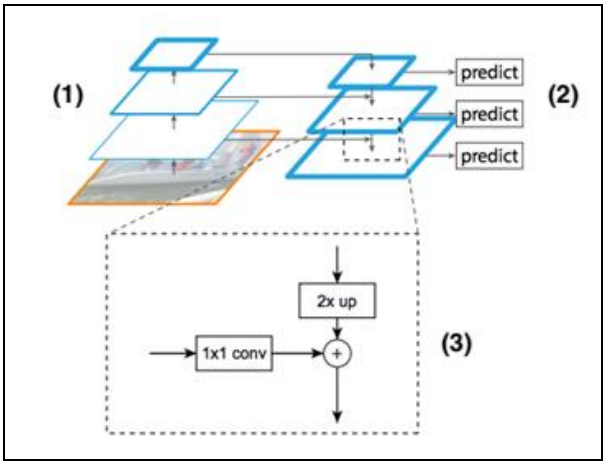


รูปที่ 2 : โครงข่ายประสาทเทียม

2.2) การตรวจจับวัตถุ

การตรวจจับวัตถุ (Object Detection) มีเป้าหมายเพื่อต้องการตรวจหาตำแหน่งของวัตถุที่สนใจ ว่าเป็นวัตถุอะไรพร้อมทั้งระบุตำแหน่งว่าวัตถุนั้นอยู่ในตำแหน่งไหนของภาพหรือวิดีโอ โดยที่วัตถุนั้นอาจจะเป็นได้ทั้ง คน สัตว์ หรือสิ่งของ ซึ่งเป็นระเบียบกรรมวิธีที่พัฒนาขึ้นเพื่อให้เครื่องจักรค้นหาสิ่งของในภาพหรือวีดีโอแทนคน จุดเด่นที่สำคัญของการตรวจจับวัตถุ คือ การตรวจคำนวณหาขอบเขตของกล่องแบบสี่เหลี่ยม (Boundary Box) ซึ่งมีวิธีการที่เป็นที่นิยมได้แก่ R-CNN และ YOLO [4], [5] เป็นต้น

สำหรับ YOLO เป็นหนึ่งในการตรวจจับวัตถุที่ใช้กันแพร่หลาย ซึ่งถูกพัฒนาโดย Joseph Redmon และ Ali Ferhadi จากมหาวิทยาลัย Washington งานวิจัยนี้เลือกแบบจำลอง YOLOv3 สำหรับการตรวจจับวัตถุเนื่องจากมีความเร็วในการประมวลผลและความแม่นยำสูง โดยจุดเด่นหลักของ YOLOv3 คือ การใช้หลักการของ Feature Pyramid Network [6]



รูปที่ 3 : หลักการของ Feature Pyramid Network

รูปที่ 3 แสดงถึงหลักการทำงานของ Feature Pyramid Network โดยมี 3 ขั้นตอน ดังนี้

2.2.1) Bottom-up Pathway ที่จะนำภาพตั้งต้นมาทำการลดขนาดลงทีละขั้นดังรูปทรงพีระมิด

2.2.2) Top-down Pathway เป็นขั้นตอนที่จะขยายขนาดภาพให้กลับมาในขนาดดังเดิม และเพิ่มขึ้นตอน Predict เข้าไปในแต่ละขั้นของพีระมิด

2.2.3) Lateral Connection จะเกิดขึ้นในทุกขั้นระหว่าง Bottom-up Pathway และ Top-down Pathway โดยจะเชื่อมพีระมิดทั้งสองด้วย 1x1 convolution เพื่อปรับให้ทั้งสองมีขนาดเท่ากัน

2.3) การวิเคราะห์วัตถุที่ปราศจากเจ้าของ

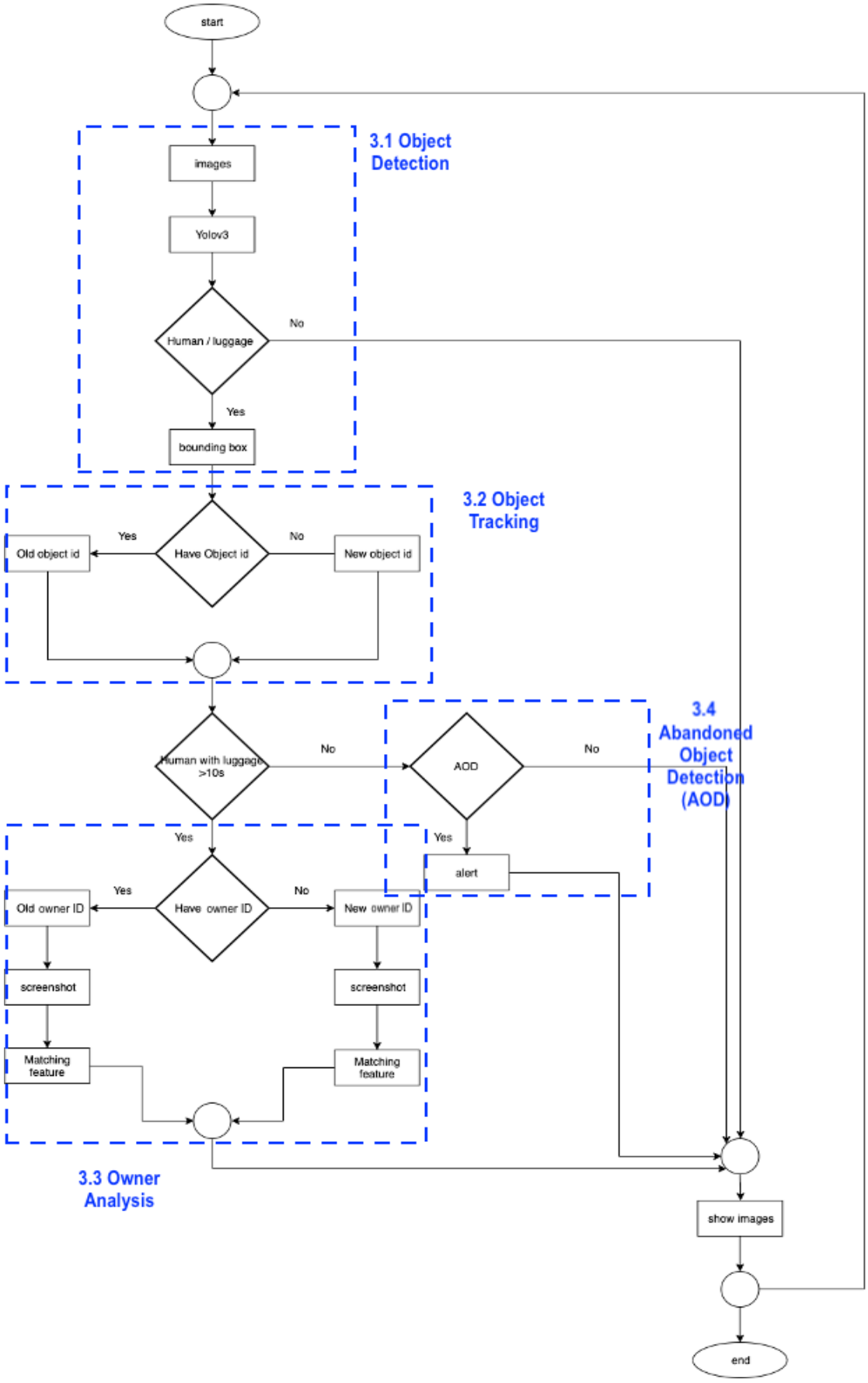
เทคนิคการวิเคราะห์การวางวัตถุโดยปราศจากผู้ครอบครองมีแนวทางการวิจัย 2 แนวทาง สำหรับแนวทางที่ 1 คือ การตรวจจับวัตถุว่าไม่มีการเคลื่อนที่เกินกว่าระยะเวลาหนึ่ง ซึ่งมักนิยมใช้เทคนิคแยกฉากหน้า (Foreground Segmentation) [7]–[9] หรือการหาตำแหน่งและแยกวัตถุด้วยเทคนิคการเรียนรู้เชิงลึก [10]–[12] ซึ่งเป็นการประยุกต์ใช้การเรียนรู้เชิงลึกเพื่อหาตำแหน่งวัตถุภายในภาพ ส่วนแนวทางที่ 2 คือ การวิเคราะห์พฤติกรรมของคนที่ไม่เอาใจใส่ (Unattended) [13]–[15] เป็นการวิเคราะห์พฤติกรรมของคนที่เป็นเจ้าของวัตถุได้ละทิ้งวัตถุนั้นหรือไม่ โดยพิจารณาด้วยระยะห่างระหว่างวัตถุถึงคนเป็นเกณฑ์ในการพิจารณา

3) วิธีดำเนินการวิจัย

การวิเคราะห์ภาพคนและสัมภาระสำหรับการตรวจจับวัตถุที่ปราศจากเจ้าของที่นำเสนอประกอบด้วย 4 ขั้นตอน ได้แก่ การตรวจจับวัตถุ (Object Detection) การติดตามวัตถุเพื่อระบุเลขประจำตัวของวัตถุ (Object Tracking) การวิเคราะห์ความเป็นเจ้าของ (Owner Analysis) และการวิเคราะห์วัตถุที่ปราศจากเจ้าของ (Abandoned Object Analysis) โดยรูปที่ 4 แสดงภาพรวมของการทำงานของทั้ง 4 ขั้นตอน

3.1) การตรวจจับวัตถุ (Object Detection)

ขั้นตอนการตรวจจับวัตถุเป็นการระบุตำแหน่งของวัตถุแบบกรอบสี่เหลี่ยมรอบวัตถุโดยเป็นพิกัดตำแหน่งมุมซ้ายบน (x1, y1) และมุมขวาล่าง (x2, y2) ซึ่งชนิดของวัตถุที่สนใจในงานวิจัยนี้ คือ คน และ กระเป๋า



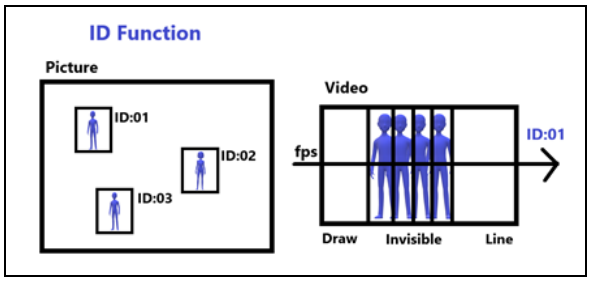
รูปที่ 4 : ภาพรวมของการระบุเจ้าของสัมภาระและการตรวจจับสัมภาระที่ปราศจากเจ้าของที่ประกอบด้วย 4 ขั้นตอนหลัก

เนื่องจากงานวิจัยนี้มีเป้าหมายที่จะประมวลผลแบบเรียลไทม์ จึงต้องเลือกใช้แบบจำลองที่ประมวลผลได้เร็วและพิจารณาความเร็วในการประมวลผลควบคู่กับความถูกต้อง แบบจำลองที่ใช้ในงานวิจัยนี้สำหรับส่วนฝึกสอน (Training) เพื่อการตรวจจับวัตถุ คือ YOLOv3 และ YOLOv3-tiny ซึ่งแบบจำลองหลังเป็นแบบจำลองที่ถูก Optimize ให้มีขนาดเล็กลงเพื่อให้ประมวลผลได้เร็วขึ้น แต่มีส่วนทำให้ประสิทธิภาพการตรวจจับวัตถุลดลง

3.2) การติดตามวัตถุ (Object Tracking)

ในการระบุเลขประจำตัววัตถุ (Object ID) การตรวจจับวัตถุเป็นการหาตำแหน่งวัตถุในแต่ละเฟรม สำหรับวิดีโอประกอบด้วยภาพหลายเฟรมมาต่อกัน จึงมีขั้นตอนเพื่อหาว่าตำแหน่งวัตถุในเฟรมปัจจุบัน คือ ตำแหน่งวัตถุใดในเฟรมถัดไป โดยใช้การระบุเลขประจำตัววัตถุ เพื่อแสดงว่าเป็นวัตถุเดียวกันในแต่ละเฟรม การระบุเลขประจำตัววัตถุมีความสำคัญอย่างมากในหลายด้าน ไม่ว่าจะเป็นการระบุตัวตน การนับ หรือการตรวจสอบความเป็นเจ้าของ

การระบุเลขประจำตัววัตถุในงานวิจัยนี้ใช้วิธีการสร้างจุดที่มองไม่เห็น (Invisible Point) เพื่อที่จะนำไว้ลากเส้นที่มองไม่เห็น (Invisible Line) จึงทำให้ ID ที่ออกมานั้นไม่เปลี่ยนแปลงแม้จะทำการเปลี่ยนเฟรมไป ดังรูปที่ 5



รูปที่ 5 : การสร้างเลขประจำตัววัตถุ Object ID ของวีดีโอ

3.3) การวิเคราะห์ความเป็นเจ้าของสัมภาระ (Owner Analysis)

ขั้นตอนการวิเคราะห์ความเป็นเจ้าของเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างคนและกระเป๋า โดยจะตรวจสอบว่าบุคคลใดเป็นเจ้าของสัมภาระชิ้นไหน เพื่อทำการบันทึกและสร้างเลขประจำตัวระหว่างเจ้าของกับสัมภาระ วิธีการที่ใช้ในงานวิจัยนี้จะใช้การพิจารณาตำแหน่งและการเคลื่อนที่ของคนและกระเป๋าโดยคนและกระเป๋าจะต้องอยู่ไม่เกินระยะห่างที่กำหนดและอยู่ใกล้กันเกินช่วงระยะเวลาที่กำหนด ซึ่งในการทดสอบได้กำหนดระยะเวลาไว้ที่ 10 วินาที ดังนั้นคนที่แค่เดินผ่านกระเป๋าจะไม่ถูกพิจารณาว่าเป็นเจ้าของกระเป๋า

3.4) การวิเคราะห์วัตถุที่ปราศจากผู้ครอบครอง (Abandoned Object Detection)

ขั้นตอนการวิเคราะห์วัตถุที่ปราศจากผู้ครอบครองเป็นการวิเคราะห์หาสัมภาระที่ถูกวางหรือละทิ้งโดยไม่มีผู้ครอบครอง ซึ่งมีหลักการโดยใช้การคาดการณ์ว่าสัมภาระมีแนวโน้มว่าจะถูกทิ้งหรือไม่ โดยพิจารณาระยะเวลาและระยะห่างของคู่ความเป็นเจ้าของระหว่างคนและกระเป๋า เมื่อบุคคลที่เป็นเจ้าของกระเป๋าเดินออกห่างจากกระเป๋าเกินระยะเวลาที่กำหนด ระบบจะถือว่าเป็นเหตุการณ์ที่มีแนวโน้มความเสี่ยงแล้วระบบจะทำการแจ้งเตือนรวมถึงแสดงภาพก่อนหน้าเพื่อแสดงบุคคลที่มีแนวโน้มว่าจะเป็นเจ้าของสัมภาระนี้

4) ผลการวิจัย

การวิเคราะห์ภาพคนและสัมภาระสำหรับการตรวจจับวัตถุที่ปราศจากเจ้าของในงานวิจัยนี้ ใช้ชุดข้อมูลแบบเปิดสาธารณะสำหรับการฝึกสอนและทดสอบ โดยมีการประเมินผลใน 4 ส่วน ดังนี้ ประสิทธิภาพแบบจำลอง YOLOv3 เปรียบเทียบกับ YOLOv3-tiny, ประสิทธิภาพการตรวจจับวัตถุแต่ละชนิด, ประสิทธิภาพการระบุความเป็นเจ้าของ, ประสิทธิภาพการระบุการละทิ้งกระเป๋า

4.1) ชุดข้อมูล (Dataset)

ข้อมูลที่ใช้ในงานวิจัยนี้มี 2 ส่วน คือ ข้อมูลสำหรับฝึกสอนการรู้จำคนและกระเป๋า และข้อมูล i-Lids AVSS-AB [13] สำหรับการวิเคราะห์ความเป็นเจ้าของและการละทิ้งกระเป๋า

4.1.1) ข้อมูลสำหรับฝึกสอนการรู้จำคนและกระเป๋า

การฝึกสอนการรู้จำวัตถุจำเป็นต้องใช้ตัวอย่างฝึกสอนจำนวนมากซึ่งชุดข้อมูล i-Lids AVSS-AB มีจำนวนภาพคนและกระเป๋าน้อย ไม่หลากหลายเพียงพอที่จะใช้ฝึกสอนเพื่อการรู้จำคนและกระเป๋า เพื่อให้ได้แบบจำลองที่มีความทนทานมากขึ้น งานวิจัยนี้จึงรวบรวมภาพจากชุดข้อมูลเปิดสำหรับฝึกสอนเพื่อการรู้จำคนและกระเป๋า โดยผู้วิจัยรวมรวมภาพของคนและกระเป๋าจำนวน 12,000 ภาพ สำหรับการฝึกสอน (Training) และ 100 ภาพสำหรับการทดสอบ (Testing) โดยมีแหล่งที่มา ได้แก่ ชุดข้อมูล COCO [16] ชุดข้อมูล Open Image [17] และ ภาพจาก Google Image Search

เนื่องจากกระเป๋ามีความหลากหลายมาก ผู้วิจัยจึงเลือกกระเป๋า 3 ประเภท ได้แก่ กระเป๋าสัมภาระ (Suitcase) กระเป๋าสะพาย (Backpack) และ กระเป๋าถือ (Handbag) ดังรูปที่ 6 โดยในการฝึกสอนนั้นกระเป๋าทั้งสามประเภทจะถูกนำไปใช้เพื่อฝึกสอน ส่วนในการทดสอบจะไม่มีการแบ่งประเภทของกระเป๋า โดยจะทำนายกระเป๋าประเภทต่าง ๆ เป็นกระเป๋าเหมือนกัน

รูปที่ 6 : ภาพคนและกระเป๋าทั้ง 3 ประเภทจากชุดข้อมูล Open Image

4.1.2) ข้อมูล i-Lids AVSS-AB

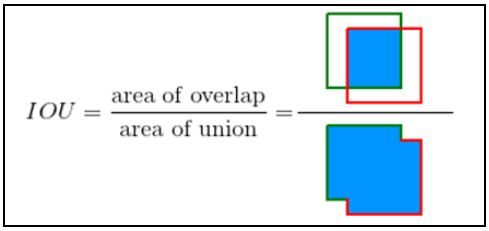
ชุดข้อมูล i-Lids AVSS-AB [18] สำหรับการวิเคราะห์ความเป็นเจ้าของและการละทิ้งกระเป๋า เป็นชุดข้อมูลวิดีโอจากกล้อง CCTV ของสถานีรถไฟที่มีคนถือและไม่ถือกระเป๋าเดินไปมา โดยรวมถึงกรณีที่มีคนทิ้งกระเป๋าด้วย ชุดข้อมูลเปิด (Public Dataset) ที่มีคนทิ้งกระเป๋าค่อนข้างมีอยู่อย่างจำกัด สำหรับข้อมูลวิดีโอนี้มีความยาวประมาณ 3 นาที จำนวน 3 วิดีโอ โดยแบ่งตามความยากง่ายเป็นระดับง่าย (AVSS AB Easy) ระดับกลาง (AVSS AB Medium) และระดับยาก (AVSS AB Hard) ในแต่ละกรณีจะมีเหตุการณ์ละทิ้งกระเป๋าที่มีความยากง่ายต่างกัน และมีความหนาแน่นและการซ้อนทับของคนต่างกันเช่นกัน

4.2) วิธีการประเมินประสิทธิภาพ

4.2.1) การวัดประสิทธิภาพการตรวจจับวัตถุ การวัดประสิทธิภาพการตรวจจับวัตถุส่วนมากจะใช้วิธี Intersection Over Union (IOU) เพื่อประเมินความถูกต้องของกรอบสี่เหลี่ยมที่ทำนายเทียบกับกรอบสี่เหลี่ยมของ Ground Truth ดังรูปที่ 7

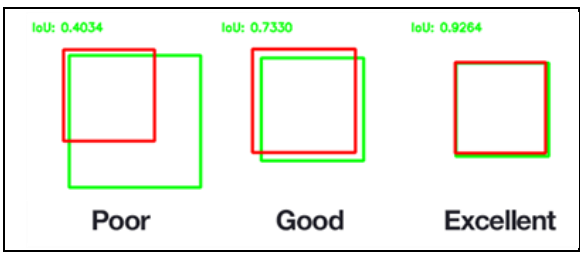
รูปที่ 7 : ผลการทำนาย (กรอบสีแดง) เทียบกับ Ground Truth (กรอบสีเขียว)

IOU เป็นการทดสอบโดยวัดตามดัชนีบน Jaccard Similarity ที่ประเมินการทับซ้อนระหว่างกรอบสี่เหลี่ยม Ground Truth และ กรอบสี่เหลี่ยมที่ทำนาย โดยมีวิธีการคำนวณ ดังรูปที่ 8



รูปที่ 8 : วิธีการคำนวณค่า Intersection Over Union (IOU)

IOU จะวิเคราะห์ได้ว่าการตรวจจับนั้นถูกต้อง (True Positive) หรือไม่ถูกต้อง (False Positive) ตามตัวอย่างดังรูปที่ 9 ทั้งกล่องสีแดงและกล่องสีเขียวเป็นวัตถุเดียวกัน โดยกล่องสีเขียวเป็นข้อมูลจริงและกล่องสีแดงเป็นข้อมูลจากการทำนาย ซึ่งคลาดเคลื่อนจากข้อมูลจริงเล็กน้อย และแสดงการซ้อนทับ 3 ลักษณะ โดยภาพทางซ้ายเป็นการทำนายที่ไม่ค่อยดีเนื่องจากตรวจจับได้ขนาดใหญ่เกินจริง ส่วนภาพกลางและภาพขวาแสดงการทำนายที่ยอมรับได้ว่าถูกต้อง



รูปที่ 9 : ผลการทำนายเทียบกับข้อมูลจริงในกรณีต่าง ๆ

4.2.2) การวัดประสิทธิภาพความถูกต้อง

การวัดประสิทธิภาพความถูกต้อง พิจารณาจากค่าความถูกต้อง (Accuracy) ดังสมการที่ 1 ค่าความแม่นยำ (Precision) ดังสมการที่ 2 และค่าความครบถ้วน (Recall) ดังสมการที่ 3

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$\Pr ecision = \frac{tp}{tp + fp} \quad (2)$$

$$\mathrm{Re}call = \frac{tp}{tp + fn} \quad (3)$$

โดยที่ tp คือ True Positive, fp คือ False Positive,
tn คือ True Negative, fn คือ False Negative

4.2.3) การวัดประสิทธิภาพความเร็ว

การเปรียบเทียบความเร็วพิจารณาได้จากค่า FLOPS (Floating Point Operations per Second) ที่แสดงจำนวน Operation ของแบบจำลองทั้งสองในตารางที่ 1 และค่า Frame Rate เป็นเฟรมต่อวินาที (Frame Per Seconds หรือ fps) ที่ประมวลผลบนหน่วยประมวลผลภาพ GPU RTX2070 บนระบบปฏิบัติการ Ubuntu 19.04

4.3) การเปรียบเทียบประสิทธิภาพของ YOLOv3 กับ YOLOv3-tiny

เนื่องจากแบบจำลอง YOLOv3-tiny ถูก Optimize ให้มีขนาดเล็กลงเมื่อเทียบกับ YOLOv3 เพื่อให้ประมวลผลได้เร็วขึ้น จึงมีผลทำให้ประสิทธิภาพความถูกต้องลดลง การทดสอบส่วนนี้จึงศึกษาเปรียบเทียบประสิทธิภาพความเร็วและความถูกต้องของแบบจำลองทั้งสองที่ถูกฝึกสอนด้วยภาพจำนวน 12,000 ภาพ จากชุดข้อมูล i-Lids AVSS-AB และทดสอบด้วยภาพจำนวน 100 ภาพ โดยแสดงผลดังตารางที่ 1 ซึ่งจะเห็นว่า YOLOv3-tiny ประมวลผลได้ประมาณ 30 fps ซึ่งเร็วกว่า YOLOv3 ประมาณ 3 เท่า แต่ค่า Accuracy ลดลงจาก 98% เหลือเพียง 37% ซึ่งการ Optimize นี้ มีผลกระทบกับวัตถุขนาดเล็ก เช่น กระเป๋าและคน ทำให้ไม่สามารถตรวจจับได้ สำหรับการนำไปใช้งาน YOLOv3 จึงมีความเหมาะสมมากกว่าในแง่ของความถูกต้องและความเร็วในการประมวลผล ดังนั้นการทดสอบประสิทธิภาพในหัวข้อที่ 4.4 ถึง 4.6 จึงเลือกใช้แบบจำลอง YOLOv3

ตารางที่ 1 : ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง YOLOv3 และ YOLOv3-tiny

Model	YOLOv3	YOLOv3-tiny
Accuracy (%)	98	37
FLOPS (Bn)	140.69	5.56
Frame rate (fps)	10	30

4.4) ผลการทดสอบประสิทธิภาพการตรวจจับวัตถุแต่ละชนิด

การทดสอบส่วนนี้เป็นการประเมินผลการถ่ายทอดการเรียนรู้ (Transfer Learning) จากแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูลหนึ่ง แล้วนำมาใช้กับชุดข้อมูลอื่น เช่น ชุดข้อมูล i-Lids AVSS-AB โดยตารางที่ 2 แสดงผลค่าความถูกต้อง True Positive (TP), False Positive (FP), False Negative (FN) แบ่งตามชนิดของวัตถุ คือ คนและกระเป๋า โดยมีการแบ่งแยกตามระดับความยาก 3 ระดับ คือ ระดับง่าย ระดับปานกลาง และระดับยาก โดยในการทดสอบนี้ได้สุ่มเลือกภาพมากรณีละ 20 ภาพ รวมทั้งสิ้น 60 ภาพ เพื่อสร้าง Ground Truth สำหรับการทดสอบ ดังรูปที่ 10 เป็นตัวอย่างผลของการตรวจจับวัตถุแต่ละชนิด โดยที่กรอบสีเหลือง (คน) และกรอบสีน้ำเงิน (กระเป๋า) คือ ความสามารถในการตรวจจับวัตถุได้ และวงกลมสีแดง คือ ส่วนที่ไม่สามารถตรวจจับวัตถุได้

โดยผลการทดลองในตารางที่ 2 และตารางที่ 3 แสดงให้เห็นว่าการตรวจจับคนมีความถูกต้องสูงมาก ส่วนการตรวจจับกระเป๋ามีความยากกว่าการตรวจจับคนมาก การตรวจจับกระเป๋าให้ค่าความแม่นยำที่สูงมาก แต่ความครบถ้วนยังค่อนข้างต่ำ ซึ่งเกิดได้จากการถูกบัง กระเป๋ามีขนาดเล็ก และลักษณะของกระเป๋ามีความหลากหลายดังรูปที่ 10 โดยประสิทธิภาพของการตรวจจับวัตถุแต่ละชนิดแสดงในตารางที่ 2 และตารางที่ 3

ตารางที่ 2 : ผลการตรวจจับวัตถุแต่ละชนิด

LEVEL	CATEGORY	TP	FP	FN
AVSS AB Easy	People	43	0	0
	Luggage	22	1	17

ตารางที่ 2 : ผลการตรวจจับวัตถุแต่ละชนิด (ต่อ)

LEVEL	CATEGORY	TP	FP	FN
AVSS AB Medium	People	98	0	7
	Luggage	17	0	45
AVSS AB Hard	People	122	0	8
	Luggage	16	0	36
Overall	People	263	0	15
	Luggage	55	1	98

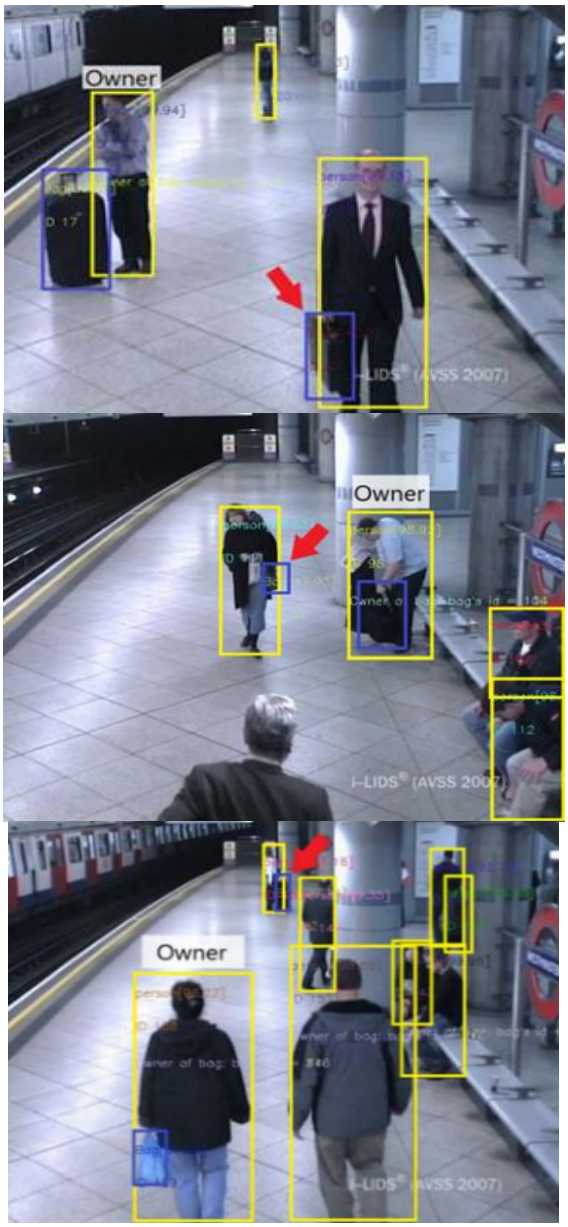
ตารางที่ 3 : ประสิทธิภาพการตรวจจับวัตถุแต่ละชนิด

LEVEL	CATEGORY	Accuracy	Precision	Recall
AVSS AB Easy	People	100%	100%	100%
	Luggage	55%	95.6%	56.4%
AVSS AB Medium	People	93.3%	100%	93.3%
	Luggage	27.4%	100%	27.4%
AVSS AB Hard	People	93.8%	100%	93.8%
	Luggage	30.8%	100%	30.8%
Overall	People	94.6%	100%	94.6%
	Luggage	35.0%	98.2%	35.9%

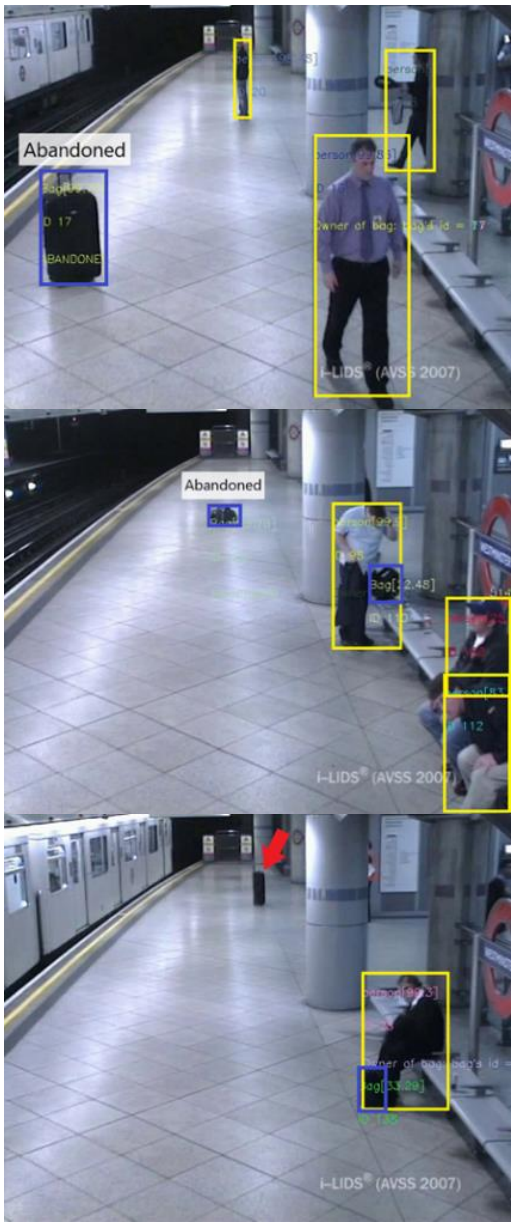
รูปที่ 10 : ตัวอย่างผลของการตรวจจับวัตถุแต่ละชนิด

4.5) ผลการทดสอบประสิทธิภาพการระบุความเป็นเจ้าของ

การทดสอบในส่วนนี้เป็นการทดสอบผลการระบุความเป็นเจ้าของแยกตามความยากง่าย 3 กรณี จากตารางที่ 4 และ 5 แสดงให้เห็นว่ากรณีการตรวจจับวัตถุที่ง่าย จะได้ค่าความแม่นยำ (Precision) 86.6% และค่าความครบถ้วน (Recall) 76.5% ซึ่งถือว่ามีทั้งความแม่นยำและความครบถ้วน โดยกรณีที่มีความยากมากขึ้นจะได้ค่าความถูกต้องที่ลดลง ความผิดพลาดจะเกิดจากการที่ไม่สามารถตรวจจับกระเป๋าได้ โดยมีสาเหตุจากกระเป๋าถูกบัง มีขนาดเล็ก หรือมีรูปร่างที่คล้ายกับวัตถุอื่น รูปที่ 11 เป็นตัวอย่างภาพของผลการระบุความเป็นเจ้าของ โดยที่แสดงผลข้อความ Owner (เจ้าของ) และจุดที่ลูกศรสีแดง คือ ตัวอย่างกรณีที่ไม่สามารถระบุความเป็นเจ้าของ



รูปที่ 11 : ตัวอย่างผลของการระบุความเป็นเจ้าของ



รูปที่ 12 : ตัวอย่างผลของการตรวจจับการละทิ้งกระเป๋า

ตารางที่ 4 : ผลการระบุความเป็นเจ้าของ

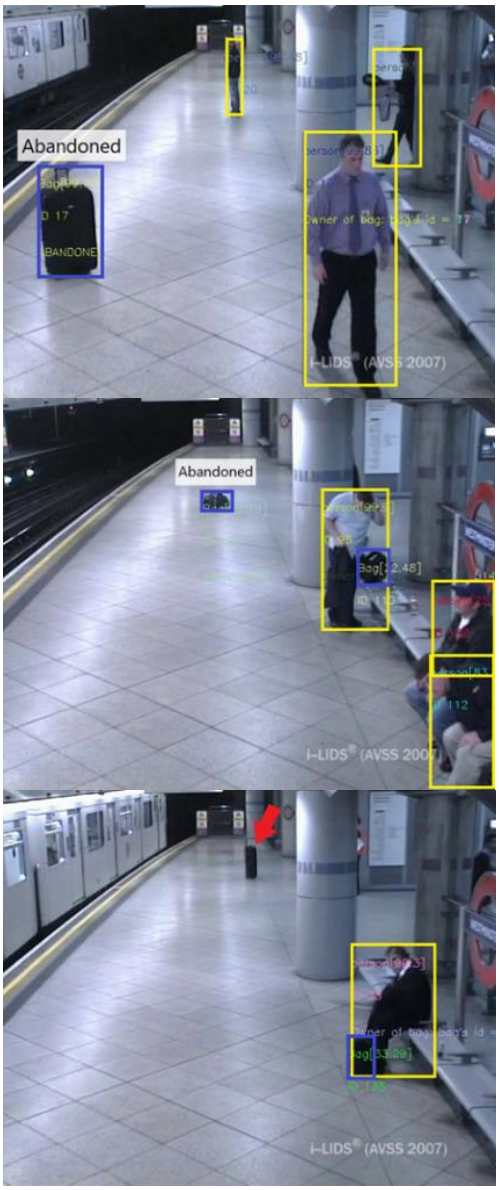
LEVEL	TP	FP	TN	FN
AVSS AB Easy	13	2	2	4
AVSS AB Medium	18	4	4	8
AVSS AB Hard	20	9	25	17
Overall	51	15	31	29

ตารางที่ 5 : ประสิทธิภาพการระบุความเป็นเจ้าของ

LEVEL	Accuracy	Precision	Recall
AVSS AB Easy	71.4%	86.6%	76.5%
AVSS AB Medium	64.7%	81.8%	69.2%
AVSS AB Hard	63.4%	69.0%	54.1%
Overall	65.1%	77.2%	63.8%

4.6) ผลการทดสอบประสิทธิภาพการระบุการละทิ้งกระเป๋า

การทดสอบในส่วนนี้เป็นการทดสอบผลการระบุการละทิ้งกระเป๋าว่าเกิดเหตุการณ์คนเดินถือกระเป๋าแล้ววางทิ้งไว้หรือไม่ โดยในวิดีโอทดสอบแบ่งเป็น 3 กรณี จากตารางที่ 6 จะเห็นว่าสามารถระบุการละทิ้งกระเป๋าได้ในสองกรณี แต่ไม่สามารถระบุการละทิ้งในกรณีที่ยากที่สุดได้ เนื่องจากกระเป๋าถูกวางอยู่ในระยะที่ค่อนข้างไกล รูปที่ 12 แสดงตัวอย่างผลของการตรวจจับการละทิ้งกระเป๋า โดยที่แสดงผลข้อความ Abandoned (ละทิ้งกระเป๋า) โดยลูกศรสีแดงแสดงกรณีที่ไม่สามารถระบุการละทิ้งกระเป๋า



รูปที่ 12 : ตัวอย่างผลของการตรวจจับการละทิ้งกระเป๋า

ตารางที่ 6 : ผลการตรวจจับการละทิ้งกระเป๋า

LEVEL	Our Method	Baseline [19]
AVSS AB Easy	✓	✓
AVSS AB Medium	✓	✓
AVSS AB Hard	×	×
Overall	66.6%	66.6%

หมายเหตุ (✓ คือ ถูกต้อง, × คือ พลาด)

5) สรุป

บทความนี้ได้นำเสนอวิธีการตรวจจับสัมภาระที่ปราศจากเจ้าของซึ่งวิเคราะห์จากภาพคนและสัมภาระที่เป็นกระเป๋าชนิดต่าง ๆ ประกอบด้วย กระเป๋าสะพาย (Backpack) กระเป๋าถือ (Handbag) และกระเป๋าเดินทาง (Suitcase) โดยสามารถประมวลผลได้แบบเรียลไทม์ และมีประสิทธิภาพความถูกต้องสูง การใช้เทคนิควิธีโครงข่ายประสาทเทียมและการประมวลผลภาพมาประยุกต์เพื่อให้สามารถช่วยตรวจจับเหตุการณ์ที่เสี่ยงต่อการทิ้งสัมภาระ ไม่ว่าผู้ทิ้งจะตั้งใจหรือไม่ก็ตาม หลังจากที่ตรวจจับได้แล้ว จะแจ้งเตือนไปยังผู้ที่มีส่วนเกี่ยวข้อง เพื่อให้ผู้ดูแลสามารถเข้าไปจัดการได้อย่างรวดเร็ว ส่วนสำคัญของงานวิจัยนี้เป็นการระบุความเป็นเจ้าของสัมภาระและการระบุสัมภาระที่ปราศจากผู้ครอบครองโดยใช้การวิเคราะห์ความสัมพันธ์ในเชิงตำแหน่งและการเคลื่อนที่ของคนและกระเป๋าร่วมกับการติดตามวัตถุ เพื่อให้สามารถประมวลผลได้เร็วสำหรับการต่อยอดเพื่อการนำไปใช้งานในอนาคต

กิตติกรรมประกาศ

ผู้เขียนขอขอบคุณโครงการสร้างปัญญาวิทย์ ผลิตนักเทคโน (Young Scientist and Technologist Programme: YSTP) ของสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ที่ให้ทุนสนับสนุนการวิจัยปริญญานิพนธ์ (Senior Project)

REFERENCES

[1] E. Luna, J. C. San Miguel, D. Ortego, and J. M. Martínez, "Abandoned object detection in video-surveillance: survey and comparison," *Sensors*, vol. 18, no. 12, pp. 1–32, Dec. 2018.

[2] F. Porikli, "Detection of temporarily static regions by processing video at different frame rates," in *Proc. IEEE Int. Conf. Adv. Video Signal-based Surveillance*, London, UK, Sep. 5–7, 2007, pp. 236–241.

[3] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP J. Adv. Signal Process.*, vol. 2008, Oct. 2007, Art. no. 197875 (2007).

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," *29th IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, Jun. 27–30, 2016, pp. 779–788.

[5] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conf. Comput. Vision Pattern Recognit.*, Columbus, OH, USA, Jun. 23–28, 2014, pp. 580–587.

[6] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, Jul. 21–26, 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[7] T. Bouwmans, F. Porikli, B. Höferlin, and A. Vacavant, Background Modeling and Foreground Detection for Video Surveillance, Boca Raton, FL, USA: CRC Press, 2014.

[8] A. Borji, M. Cheng, H. Jiang and J. Li, " Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.

[9] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A Benchmark dataset and evaluation methodology for video object segmentation," in *29th IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, Jun. 27–30, 2016, pp. 724–732.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: arXiv:1704.04861.

[11] S. Smeureanu and R. T. Ionescu, "Real-time deep learning method for abandoned luggage detection in video," in *Proc. 26th Eur. Signal Process. Conf. EUSIPCO*, Rome, Italy, Sep. 3–7, 2018, pp. 1775–1779.

[12] X. Xie, C. Wang, S. Chen, G. Shi, and Z. Zhao, "Real-time illegal parking detection system based on deep learning," in *Int. Conf. Deep Learn. Technol.*, Chengdu, China, Jun. 2–4, 2017, pp. 23–27.

[13] C.-Y. Lin, K. Muchtar, and C.-H. Yeh, "Robust techniques for abandoned and removed object detection based on Markov random field," *J. Vis. Commun. Image Represent.*, vol. 39, pp. 181–195, Aug. 2016.

[14] I. Dahi, M. C. E. Mezouar, N. Taleb, and M. Elbahri, "An edge-based method for effective abandoned luggage detection in complex surveillance videos," *Comput. Vis. Image Underst.*, vol. 158, pp. 141–151, May 2017.

[15] J. Kim and D. Kim, "Accurate abandoned and removed object classification using hierarchical finite state machine," *Image Vis. Comput.*, vol. 44, pp. 1–14, Dec. 2015.

[16] T. Lin *et al.*, " Microsoft COCO: Common Objects in Context," 2014. [Online]. Available: arXiv:1405.0312.

[17] A. Kuznetsova *et al.*, "The Open Images Dataset V4 Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale," *Int. J. Comput. Vis.*, vol. 128, pp. 1956–1981, Mar.2020.

[18] *i-Lids dataset for AVSS 2007*, Sep. 2007. [Online]. Available: http://www.eecs.qmul.ac.uk/~andrea /avss2007_d.html

[19] P. L. Venetianer, Z. Zhang, W. Yin, and A. J. Lipton, "Stationary target detection using the objectvideo surveillance system," in *IEEE Conf. Adv. Video Signal-based Surveillance*, London, UK, Sep. 5–7, 2007, pp. 242–247.