การใช้โครงข่ายประสาทเทียมแบบคอนโวลูชั่นภาษากายของคน สำหรับการควบคุมหุ่นยนต์ขับเคลื่อนสองล้อ

Using the Convolution Neural Network to Predict Human Body Languages for Two Wheel Drive Robot Control

ดำรงค์ศักดิ์ กิจเดช* และ วีระพันธ์ ด้วงทองสุข

สาขาวิชาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเอเชียอาคเนย์
19/1 ถนนเพชรเกษม แขวงหนองค้างพลู เขตหนองแขม กรุงเทพฯ 10160
ผู้นิพนธ์ประสานงาน : dumrongsakk@sau.ac.th, 02-8074500-27 ต่อ 404, 302

วันที่รับบทความ: 12 พฤษภาคม 2565 / วันที่แก้ไขบทความ: 15 มิถุนายน 2565 / วันที่ตอบรับการตีพิมพ์: 16 มิถุนายน 2565

บทคัดย่อ ปัจจุบันมีการใช้ปัญญาประดิษฐ์มาช่วยงานในหลายเรื่องและบ่อยครั้งที่จำเป็นต้องใช้กล้องร่วมกับปัญญาประดิษฐ์เพื่อให้เป้าหมายลุล่วงไปได้ โดยทั่วไปหุ่นยนต์จะใช้การควบคุมอัตโนมัติหรือการควบคุมผ่านตัวควบคุม อย่างไรก็ตามความอิสระและเป็นธรรมชาติค่อนข้างน้อย งานวิจัยนี้จึงได้นำเสนอการใช้โครงขายประสาทเทียมแบบคอนโวลูชั่นทำนายภาษากายของคนสำหรับควบคุมหุ่นยนต์ขับเคลื่อนสองล้อเพื่อเพิ่มความอิสระในการควบคุม หลักการทำงานคือกล้องจะทำการรับภาพและส่งไปยังคอมพิวเตอร์ ท่าทางที่ประมาณได้จะถูกส่งเป็นสัญญาณไร้สายจากคอมพิวเตอร์ไปยังหุ่นยนต์เพื่อให้ทำงานตามคำสั่ง การทดลองจะใช้ภาพจากกล้องหลายชนิดในการเก็บภาพสำหรับทำการฝึกปัญญาประดิษฐ์ เมื่อได้ค่าน้ำหนักที่ใช้ในการทดลองแล้ว ระบบจะถูกปรับปรุงเพื่อให้สามารถทำงานแบบเวลาจริง ผลที่ได้คือใช้เวลาในการประมวลผลแต่ละภาพโดยเฉลี่ย 0.05 วินาที ค่าความถูกต้องอยู่ที่ 80 เปอร์เซ็นต์

คำสำคัญ **:** โครงข่ายประสาทเทียม, คอนโวลูชั่น, การควบคุมหุ่นยนต์, ภาษากาย, ปัญญาประดิษฐ์

Abstract Currently, artificial intelligent (AI) has been used to contribute in many applications. In several time, a camera is used together with the AI to achieve the working target. In general, the robots are controlled by means of automatic control or remote control system which the degrees of freedom are quite small. Thus, this research presented a convolutional neural network (YOLOv5) to predict the human body languages in order to control the robots. Convolution neural network can be used to predict the human postures. After that, all signals were sent to the robots via wireless computer system for working on order. For experiments, the images from several cameras were used for the AI training. After that, the system was approved for the real-time operation. The results indicated that the each prediction time was averaged about 0.05 second and the accuracy of 80 percent.

Keywords: Neural Networks, Convolution, Robot Control, Body language, AI

1. บทนำ

ปัจจุบันกล้องถูกนำมาใช้งานร่วมกับหุ่นยนต์มากขึ้นดังใน [1] และ [2] เนื่องจากต้องการความยืดหยุ่นในการหยิบจับวัตถุและการหาวัตถุ ในการควบคุมหุ่นยนต์โดยปกติจะใช้การควบคุมแบบอัตโนมัติหรือการควบคุมแบบใช้ตัวควบคุม (Remote Control) อย่างไรก็เมื่อต้องการสั่งงานด้วยตัวเองนั่นคือการใช้ Remote Control ความยืดหยุ่นในการสั่งงานนั้นค่อนข้างน้อยและหากเป็นผู้ใช้งานใหม่จะทำการควบคุมได้ยาก

ปัญญาประดิษฐ์ถูกพัฒนาและออกแบบมาในหลายรูปแบบหากต้องการรู้แค่ชนิดและตำแหน่งที่วัตถุอยู่จะเป็นกลุ่มของ YOLO (You Only Look Once) โดยมีตั้งแต่ YOLOv1 – YOLOv5 [3 - 7] ซึ่งแต่ละเวอร์ชั่นจะมีข้อดีและข้อเสียแตกต่างกันออกไปจึงได้มีการนำเวอร์ชั่นต่างๆมาเปรียบเทียบกันอยู่เสมอ เช่นการเปรียบเทียบกันระหว่าง YOLOv3 และ YOLOv5 ดังแสดงใน [3] นอกจากนี้ยังมีการดัดแปลงเพื่อเพิ่มความเร็วหรือผลลัพธ์ที่ได้อย่างเช่นการดัดแปลง YOLOv4 เป็น YOLOv4 tiny ใน [7] อีกกลุ่มของปัญญาประดิษฐ์ที่ถูกพัฒนามาควบคู่กันคือ R- CNN (region- based convolutional neural networks) โดยมีตั้งแต่ R-CNN, Mask R-CNN และ Fast R-CNN เป็นต้น โดยเทคนิคนี้เป็นการหาตำแหน่ง ชนิดและขอบของวัตถุที่สนใจ ซึ่งการเตรียมข้อมูลในเวลามากกว่าและใช้เวลาในการ Training ข้อมูลมากกว่า ดังแสดงใน [8] และยังมีการเปรียบเทียบข้ามกันระหว่างปัญญาประดิษฐ์ 2 แบบนี้ด้วยดังแสดงใน [5]

การใช้งานหุ่นยนต์เคลื่อนที่ (Mobile Robot) [9 - 11] แบบขับเคลื่อนสองล้อมีอย่างต่อเนื่อง ลักษณะของหุ่นยนต์ประเภทนี้คือใช้เพียงแค่สองล้อในการขับเคลื่อนและมีล้อประคองเพื่อไม่ให้หุ่นยนต์ล้ม โดยงานวิจัยส่วนมากจะมีล้อประคองตั้งแต่ 1 – 4 ล้อ จำนวนล้อขึ้นอยู่กับการวางตำแหน่งของน้ำหนักและการวางตำแหน่งล้อ หุ่นยนต์ประเภทนี้ถูกนำไปใช้ในภารกิจต่าง ๆ เช่น การสำรวจ [11], การเคลื่อนยนต์วัตถุ เป็นต้น ซึ่งหากเป็นงานประเภทที่เป็นการทำงานแบบอัตโนมัติจำเป็นต้องมีเซ็นเซอร์ติดที่บริเวณล้อทั้งสองข้างเพื่อทำการระบุตำแหน่งของหุ่นยนต์ [9] ซึ่งต้องทำการเขียนโปรแกรมแบบวงปิดและใช้การแปลงกลับ (Inverse Kinematic) ซึ่งเพิ่มความซับซ้อนและค่าใช้จ่ายให้ระบบเป็นอย่างมาก

งานวิจัยนี้จึงได้นำเสนอการนำปัญญาประดิษฐ์เข้ามาช่วยในการประมาณท่าทางของคน ซึ่งคล้ายกับการใช้มือโบกรถจอดในช่องจอดรถโดยไม่ได้มีการพูดคุยโดยใช้เพียงแค่ท่าทางในการบอก หลักการทำงานคือ ผู้ควบคุมทำการแสดงท่าทางด้านหน้ากล้อง จากนั้นภาพที่ได้จะถูกส่งเข้าไปประมวลผลโดยใช้ YOLOv5 การประมวลผลทั้งหมดทำงานบนคอมพิวเตอร์จากนั้นคำสั่งจะถูกส่งเป็นสัญญาณแบบไร้สายไปยังหุ่นยนต์เพื่อให้หุ่นยนต์ทำงานตามคำสั่ง

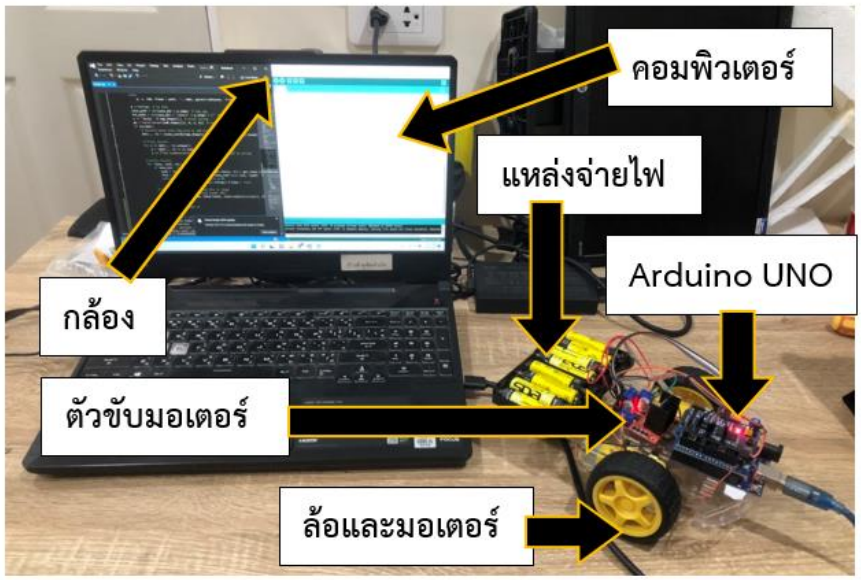
เนื่องจากการทำนายภาษากายของคนไม่สามารถควบคุมสภาพแวดล้อมได้เหมือนการประมาณท่าทางวัตถุในอุตสาหกรรมที่สามารถคุมสีของพื้นหลังและแสงที่ตกกระทบได้ทำให้สามารถใช้การประมวลผลภาพได้ (Image Processing) จึงจำเป็นต้องใช้ปัญญาประดิษฐ์เข้ามาแก้ปัญหาส่วนของสภาพแวดล้อมที่ไม่แน่นอน และจากการทบทวนวรรณกรรมพบว่าการใช้ปัญญาประดิษฐ์ในการทำนายภาษากายของคน หากต้องการการทำงานแบบเวลาจริงใช้การประมวลผลน้อยควรใช้เป็นปัญญาประดิษฐ์แบบ YOLO ซึ่ง YOLOv5 เป็นปัญญาประดิษฐ์ที่ใหม่ที่สุดในฝั่งของ YOLO โดยใช้เวลาในการฝึกข้อมูลน้อยที่สุด มีความแน่นยำมากที่สุด และมีความเร็วในการประมาณมากที่สุด

2. อุปกรณ์การทดลอง

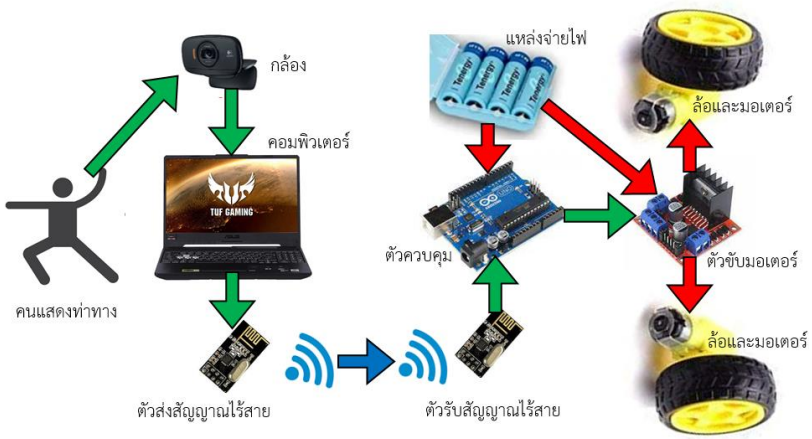
อุปกรณ์ที่ใช้ในการทดลองมีส่วนประกอบหลักคือ กล้อง Webcam สำหรับการเก็บภาพที่จะใช้ในการฝึกข้อมูลและการหาตำแหน่งและชนิดของวัตถุแบบเรียลไทม์ซึ่งจะถูกติดอยู่ที่ตัวคอมพิวเตอร์ ส่วนต่อมาคือคอมพิวเตอร์ Asus รุ่น Tuf gaming F15 ใช้ในการรับภาพ

จากกล้องแล้วนำมาประมวลผลโดยใช้ปัญญาประดิษฐ์ชนิด YOLOv5 ส่วนถัดมาคือ NRF24L01 โมดูลไร้สายติดตั้งไว้ 2 จุดคือที่คอมพิวเตอร์และหุ่นยนต์

ส่วนต่อมาคือชุดหุ่นยนต์ขับเคลื่อนสองล้อขนาดเล็ก ภายในมีมอเตอร์กระแสตรงขนาดเล็ก 2 ตัว ล้อขับเคลื่อน 2 ล้อ ล้อหมุนอิสระ 1 ล้อ บอร์ด Arduino UNO ส่วนของตัวขับมอเตอร์เป็น L298N Driver ส่วนแหล่งจ่ายไฟใช้เป็นถ่าน AA ขนาด 1.2 โวลต์ 8 ก้อน โดยที่ภาพรวมโครงสร้างของหุ่นยนต์ถูกแสดงในรูปที่ 1 และการเชื่อมต่อระบบแสดงในรูปที่ 2 โดยที่กล้องจะทำการเก็บภาพท่าทางของคนและส่งไปยังคอมพิวเตอร์ โดยภายในจะมีส่วนของการรับภาพส่งไปยังปัญญาประดิษฐ์ จากนั้นผลที่ได้จะถูกประมวลผลและแปลงค่าส่งเป็นสัญญาณไร้สายไปยังตัวควบคุมที่หุ่นยนต์ โดยที่ตัวหุ่นยนต์จะมีแหล่งจ่ายไฟเป็นแบตเตอรี่ จากนั้นหุ่นยนต์จะส่งสัญญาณไปที่ตัวขับเคลื่อนมอเตอร์ เพื่อทำให้มอเตอร์หมุนในทิศทางที่กำหนด



รูปที่ 1 ส่วนประกอบอุปกรณ์ที่ใช้ในการทดลอง



รูปที่ 2 การเชื่อมต่อของระบบและหลักการทำงาน (ลูกศรสีเขียวคือสัญญาณ และลูกศรสีแดงคือไฟเลี้ยงระบบ)

3. ระเบียบวิธีวิจัย

ระเบียบวิธีวิจัยจะถูกแบ่งออกเป็นสองส่วนคือ ส่วนของการเตรียมข้อมูลและการฝึกข้อมูล และส่วนของขั้นตอนในการทดลอง

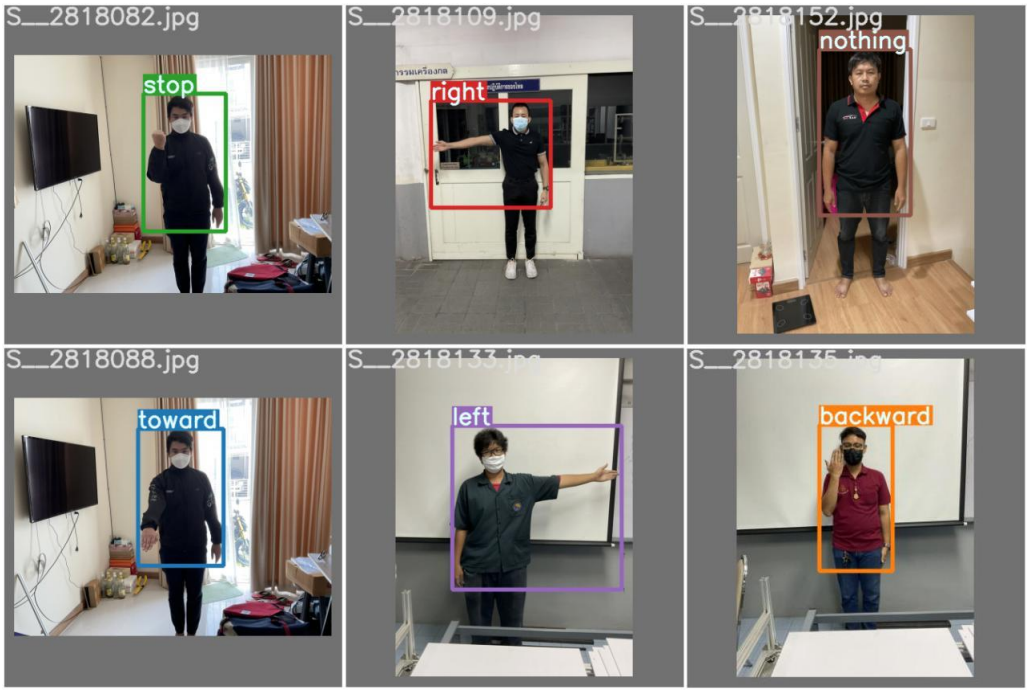
3.1 การเตรียมข้อมูล (Dataset) และการฝึกข้อมูล (Data Training)

งานวิจัยนี้ใช้ปัญญาประดิษฐ์ชนิด YOLOv5 (You Only Look Once Version 5) ในการทำนายตำแหน่งและท่าทางของคนซึ่งเป็นลักษณะของภาษากาย โดยที่ YOLOv5 มีพื้นฐานมาจากเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชั่น (Convolution Neural Networks) โครงสร้างของ YOLOv5 ถูกแสดงดังรูปที่ 3 จากรูปจะเห็นว่าโครงสร้างมีความซับซ้อนกว่าโครงข่ายประสาทเทียมแบบคอนโวลูชั่นมาก ปัจจุบันเป็นที่นิยมมากในการนำมาใช้งานร่วมกับกล้อง เนื่องจากโครงข่ายประสาทเทียมชนิดนี้การฝึกข้อมูลใช้เวลาน้อยมากเมื่อเทียบกับโครงข่ายประสาทเทียมชนิดอื่น และยังสามารถปรับแต่งค่าต่าง ๆ ได้โดยง่าย โครงข่ายประสาทเทียมนั้นคือการจำลองโครงข่ายประสาทในสมองของมนุษย์นั่นเองซึ่งทำให้สามารถตัดสินใจในงานที่มีความซับซ้อนได้และมีความเร็วในการประมวลผลสูง โดยที่ความแม่นยำขึ้นอยู่กับจำนวนของภาพตัวอย่างและความหลากหลายของภาพตัวอย่างที่นำมาฝึกปัญญาประดิษฐ์

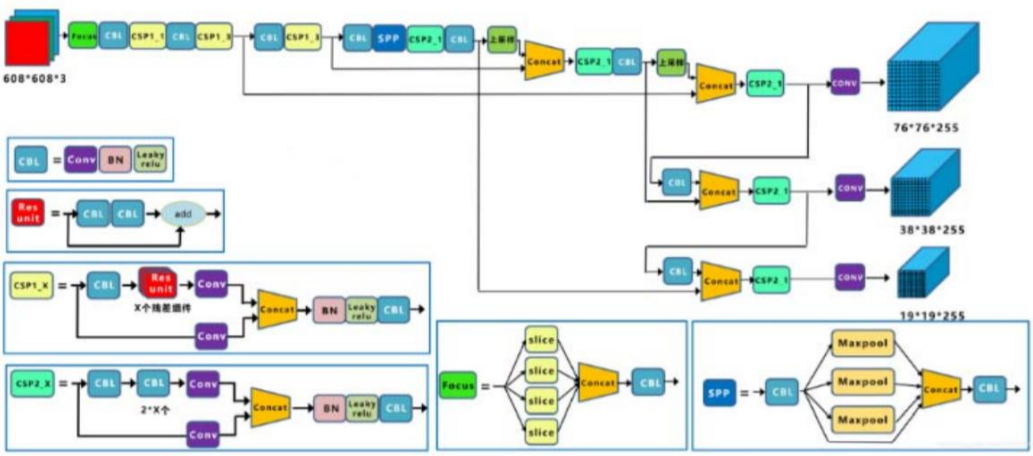
ภาพที่ใช้ในการฝึกปัญญาประดิษฐ์จะใช้ภาพที่ได้จากกล้อง Webcam ใช้ภาพตัวอย่างทั้งหมด 122 ภาพ คนที่เข้าร่วมการทดลอง 7 คน สถานที่เก็บภาพ 7 สถานที่ และมีทั้งหมด 6 ท่าทาง นั่นคือ เดินหน้า (Toward), ถอยหลัง (Backward), หยุด (Stop), เลี้ยวขวา (Right), เลี้ยวซ้าย (Left) และ ไม่ต้องทำอะไร (Nothing) ท่าทางต่างๆ ถูกแสดงดังรูปที่ 3 โดยภาพทั้งหมดเป็นภาพที่ใช้ในการฝึกปัญญาประดิษฐ์ (Training) 70 เปอร์เซ็นต์ ภาพที่ใช้ในการประเมินผลการฝึก (Valid) 10 เปอร์เซ็นต์ และภาพที่ใช้ในการทดสอบ หลังจากทำการฝึกปัญญาประดิษฐ์เรียบร้อยแล้วอีก 20 เปอร์เซ็นต์ ค่าตัวแปรต่าง ๆ ที่ใช้ในการฝึกปัญญาประดิษฐ์มีดังนี้ใช้อัตราการเรียนรู้ (Learning Rate) อยู่ที่ 0.0003 ใช้ Batch size เป็น 16 จำนวนรอบในการฝึก (Epoch) ขนาดภาพที่ใช้ในการฝึกคือ 416x416 pixel อยู่ที่ 600 รอบ

โครงข่ายประสาทเทียมส่วนของแบบจำลองที่ใช้ในการทดลองเป็น YOLOv5 แบบ L โดยที่ YOLOv5 มีโครงข่ายประสาทเทียมทั้งหมด 4 แบบ คือ S, M, L และ X ดังแสดงในรูปที่ 5 ยิ่งโครงข่ายมีความซับซ้อนจะยิ่งต้องใช้พื้นที่ในการจัดเก็บข้อมูลมากขึ้น แต่จะทำให้ระบบที่มีความซับซ้อนมากมีความแม่นยำขึ้นด้วย ภาพโครงข่ายประสาทเทียม YOLOv5 ส่วนของคอนโวลูชั่นถูกแสดงในรูปที่ 4 โดยส่วนนี้จะเป็นการผสมกันหลายชั้นของ Convolution Layer, BN (Batch Normalization) Layer และ Leaky ReLU (Rectified Linear Unit) เพื่อช่วยให้โครงข่ายประสาทเทียมสามารถทำนายท่าทางได้ดียิ่งขึ้น โดยที่ Convolution Layer คือการนำภาพมาใส่ตัวกรอง (Filter) แบบต่างๆ ส่วน Batch Normalization จะช่วยให้แต่ละ Layer ในโครงข่ายประสาทเทียมสามารถเรียนรู้ได้ด้วยตัวเองอย่างอิสระและลดการผูกติดกับ Layer อื่นๆ ส่วน Learky ReLU คือการเปลี่ยนค่าต่างๆให้เป็นค่าบวกทั้งหมดเนื่องจากการกรองภาพโดยใช้ Convolution Layer จะทำให้ค่าที่ได้บางค่ามีค่าติดลบทำให้ไม่สามารถแสดงภาพได้

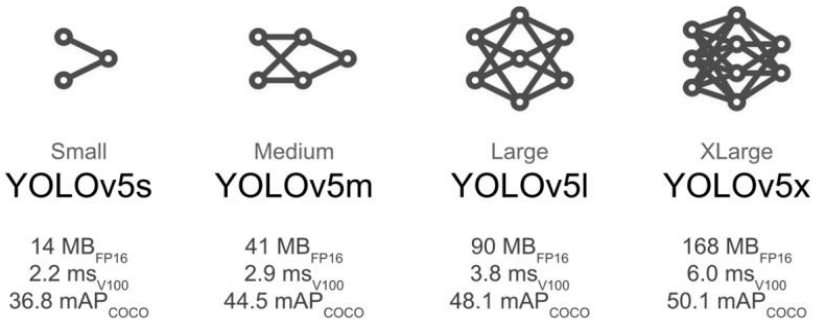
การฝึกปัญญาประดิษฐ์จำเป็นต้องใช้คอมพิวเตอร์ที่มีการ์ดจอที่มีความสามารถในการแสดงผลสูง สำหรับช่วยในการฝึก โดยงานวิจัยนี้ใช้การคำนวณบน Google Collab ซึ่งเป็นการประมวลผลแบบออนไลน์ โดยใช้การ์ดจอ NVIDIA Tesla T4 และเวลาในการฝึกข้อมูลทั้งหมด 1 ชั่วโมง 18 นาที 1 วินาที หลังจากทำการฝึกเสร็จจะได้แบบจำลอง (Model) สำหรับการทำนายท่าทางแบบเวลาจริง



รูปที่ 3 ท่าทางทั้งหมดที่ใช้ในการทดลอง



รูปที่ 4 โครงสร้างโครงข่ายประสาทเทียม YOLOv5 [12]



รูปที่ 5 แสดงขนาดต่างๆของแบบจำลอง YOLOv5 [13]

3.2 ขั้นตอนในการทดลอง

หลังจากทำการฝึกปัญญาประดิษฐ์โดยใช้การ์ดจอผ่าน Google Collab จนได้ค่าน้ำหนัก (Weight) มาแล้ว จากนั้นนำค่าน้ำหนักที่ได้มาใช้กับโปรแกรมส่วนของการทำนายท่าทาง ภาพรวมการทำงานถูกแสดงดังรูปที่ 3 โดยเริ่มจากกล้องทำการรับภาพท่าทางที่คนทำอยู่ขณะนั้นและส่งไปให้ YOLOv5 ซึ่งทำงานผ่านโค๊ด Python ถ้าตรวจจับพบท่าทางที่กำหนดไว้ในการฝึกตัว YOLOv5 จะทำการระบุว่าเป็นท่าอะไรและอยู่ตำแหน่งไหนของภาพ จากนั้นจะทำการส่งท่าที่พบออกไปเป็น String ไปยังบอร์ด Arduino UNO ที่อยู่กับคอมพิวเตอร์จากนั้นบอร์ดจะทำการส่งสัญญาณแบบไร้สายไปยังตัวรับสัญญาณไร้สายอีกตัวหนึ่ง เมื่อบอร์ด Arduino UNO ที่อยู่ที่ตัวหุ่นยนต์ได้รับสัญญาณแล้วจะทำการแปลงค่าจาก String เป็น Integer และทำการสั่งมอเตอร์ผ่านทางตัวขับเคลื่อนมอเตอร์ ให้หุ่นยนต์สามารถเคลื่อนที่ตามท่าทางที่สั่งงานไว้ผ่านท่าทาง โดยที่ลักษณะการทำงานจะเป็นไปตามรูปที่ 6 ซึ่งเป็นการสั่งงานให้หุ่นยนต์เคลื่อนที่ตามท่าทางที่กำหนด

ขั้นตอนการเก็บผลการทดลองแบ่งออกเป็น 2 ส่วนหลักๆ ส่วนแรกคือการทดลองแบบรับภาพที่มีอยู่แล้วและทำการทำนายท่าทางโดยไม่ได้นำไปควบคุมหุ่นยนต์ ส่วนที่สองเป็นการรับภาพจากกล้องโดยตรงจากนั้นทำการทำนายและส่งผลที่ได้ไปควบคุมหุ่นยนต์ทันที

สำหรับส่วนแรกมีการเก็บผลสองแบบ แบบแรกคือการเก็บผลในระหว่างการฝึกข้อมูลโดยผลจะได้จากการใช้ภาพที่มีอยู่ในการฝึกและการใช้ภาพที่ไม่ได้มีอยู่ในการฝึก โดยผลที่ได้จะถูกแสดงเป็นกราฟแกน y แสดงเป็นเปอร์เซ็นต์มีค่าระหว่าง 0 – 1 และแกน x แสดงเป็นจำนวนครั้งที่ทำการฝึกข้อมูล โดยมีค่าที่แสดงดังนี้ Box, Objectness, Classification, Precision, Recall, Val Box, Val Objectness, Val Classification และ mAP (mean Average Precision) จากนั้นนำค่า Precision และค่า Recall มาเทียบกับค่า Confidence ด้วย แบบที่สองคือการนำภาพที่ไม่ได้อยู่ในการฝึกข้อมูลและการประเมินผลข้อมูล แต่ยังคงเป็นภาพของคนที่ใช้ในการฝึกข้อมูลอยู่โดยการสุ่มภาพมาทั้งหมด 10 ภาพ จากนั้นทำการทำนายแล้วเทียบกับเฉลยและเวลาที่ใช้

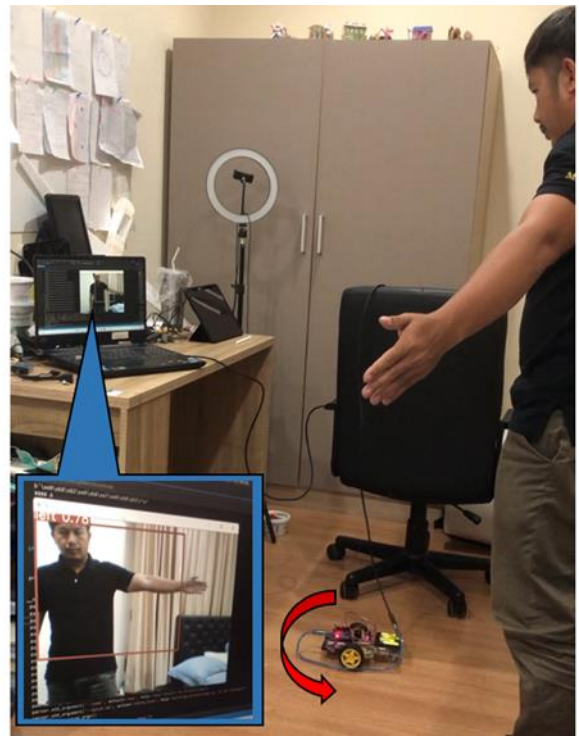
ส่วนที่สองคือการทดสอบแบบเวลาจริงโดยใช้ผู้ทดสอบสองคนคือคนที่มีภาพอยู่ในการฝึกข้อมูลและอีกคนไม่มีภาพอยู่ในการฝึกข้อมูลทำการทดสอบท่าละ 3 ครั้ง ทุกครั้งจะทำการเปลี่ยนเสื้อเป็นสีที่มีในการฝึกข้อมูล 2 ชุด และสีที่ไม่มีในการฝึกข้อมูลอีก 1 ชุด ทั้งสองคน ผลที่ได้คือท่าที่ทำนายได้เทียบกับท่าที่ทำจริงและจับเวลา

$$\text{precision} = \frac{tp}{tp + fp} \quad (1)$$

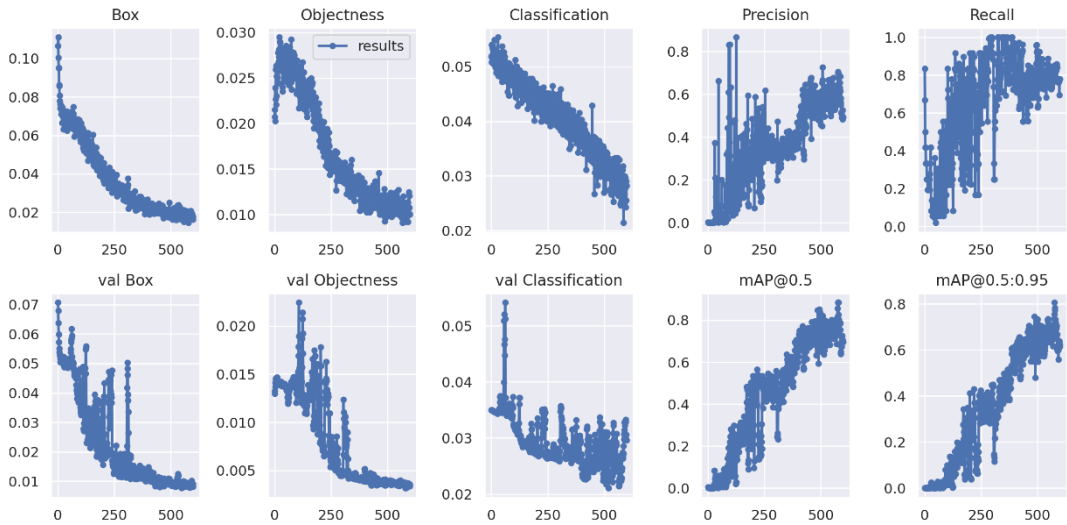
$$\text{recall} = \frac{tp}{tp + fn} \quad (2)$$

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1} \quad (3)$$

จากสมการด้านบน tp คือทำนายว่าใช่และใช่จริง ส่วน fp คือทำนายว่าใช่และไม่ใช่จริงและ fn คือทำนายว่าไม่ใช่แต่ความจริงใช่ ค่า Precision คือค่าความแม่นยำสามารถหาได้จากสมการที่ (1) ค่า Recall คือค่าความถูกต้องสามารถหาได้จากสมการที่ (2) และค่า F_1 Score คือค่าเฉลี่ยของ Precision และ Recall ดังแสดงในสมการที่ (3) ส่วน Confidence คือค่าความเชื่อมั่น สมมติว่า ตัวปัญญาประดิษฐ์ทำนายท่าทางว่าคือเลี้ยวขวา โดยมีความเชื่อมั่นว่าคือขวาถึง 60% แต่ค่าความเชื่อมั่นไม่ได้เป็นการรับประกันความถูกต้อง



รูปที่ 6 การทดลองสั่งงานหุ่นยนต์ผ่านท่าทาง



รูปที่ 7 ผลจากการฝึกปัญญาประดิษฐ์

4. ผลการทดลอง

ตารางที่ 1 แสดงผลการทดสอบภาพที่ไม่ได้นำไปฝึกปัญญาประดิษฐ์และไม่ได้นำไปประเมินผล ซึ่งเป็นการทดลองโดยการนำภาพมาเข้าโปรแกรมโดยตรงไม่ผ่านกล้อง โดยตารางประกอบด้วยท่าทางของภาพ ท่าทางที่ทำนายได้ เวลาที่ใช้ในแต่ละภาพ และค่าเฉลี่ยเวลา โดยสามารถตรวจจับได้ 100 เปอร์เซ็นต์ ความถูกต้องอยู่ที่ 50 เปอร์เซ็นต์ เวลาที่ใช้เฉลี่ยต่อภาพคือ 0.0384 วินาที

รูปที่ 7 แสดงผลที่ได้จากการฝึกปัญญาประดิษฐ์ซึ่งไม่ได้นำภาพอื่นมาทำการทดสอบและไม่ได้ทำงานแบบเรียลไทม์ โดยเราจะพิจารณาผลการฝึกข้อมูลในรอบที่ 500 ซึ่งแกน y แสดงเป็นเปอร์เซ็นต์มีค่าระหว่าง 0 – 1 และแกน x แสดงเป็นจำนวนครั้งที่ทำการฝึกข้อมูล กราฟ Box แสดงค่าความผิดพลาดในการตีกรอบท่าทางที่ทำนายได้ซึ่งอยู่ที่ 2 เปอร์เซ็นต์ กราฟ Objectness แสดงค่าความไม่ปกติของการทำนายท่าทางซึ่งอยู่ที่ 10 เปอร์เซ็นต์ กราฟ Classification แสดงค่าความผิดพลาดในการทำนายท่าทางซึ่งอยู่ที่ 2.5 เปอร์เซ็นต์ กราฟ Precision คือค่าความแม่นยำอยู่ที่ 60 เปอร์เซ็นต์ Recall คือค่าความถูกต้องอยู่ที่ 80 เปอร์เซ็นต์ ส่วน Val Box, Val Objectness และ Val Classification คือการนำภาพที่ไม่ได้นำมาฝึกปัญญาประดิษฐ์มาทดสอบกับแบบจำลองที่ทำการฝึกแล้ว ซึ่งมีค่าผิดพลาดอยู่ที่ 1 เปอร์เซ็นต์, 0.5 เปอร์เซ็นต์ และ 2.5 เปอร์เซ็นต์ ตามลำดับ กราฟ mAP ที่ความเชื่อมั่นเท่ากับ 0.5 แสดงค่าเฉลี่ยความแม่นยำของทุก Class ของวัตถุที่ทำการทำนาย มีค่าเท่ากับ 80 เปอร์เซ็นต์ และที่ค่าความเชื่อมั่นระหว่าง 0.5 – 0.95 อยู่ที่ 75 เปอร์เซ็นต์

จากตารางที่ 2 แสดงการเก็บผลการทดลองแบบเรียลไทม์ผ่านกล้อง โดยที่ตารางส่วนที่แรเงาแสดงผลการทดลองของผู้ร่วมทดลองที่ไม่ได้มีอยู่ในการฝึกข้อมูล โดยที่ทำการทดลองสามครั้ง ครั้งที่1 ใส่เสื้อสีเหลืองที่ไม่ได้นำมาใช้ในการฝึกข้อมูล ครั้งที่ 2 และ 3 เป็นเสื้อสีเลือดหมูและเทา ตามลำดับ ซึ่งเป็นสีของเสื้อที่ใช้ในการฝึกข้อมูล ส่วนตารางส่วนที่ไม่ได้แรเงาแสดงผลการทดลองของผู้ร่วมทดลองที่มีอยู่ในการฝึกข้อมูล โดยที่ทำการทดลองสามครั้ง ครั้งที่1 ใส่เสื้อฟ้าที่ไม่ได้นำมาใช้ในการฝึกข้อมูล ครั้งที่ 2 และ 3 เป็นเสื้อสีเลือดหมูและเทา ตามลำดับ ซึ่งเป็นสีของเสื้อที่ใช้ในการฝึก จากการทดลองพบว่าการใช้สีของเสื้อที่ไม่ได้มีอยู่ในการฝึกข้อมูลร่วมกับผู้ร่วมการทดลองที่ไม่ได้มีอยู่ในการฝึกข้อมูล ทำให้ความสามารถในการทำนายลดลงอย่างมาก อยู่ที่ 50 เปอร์เซ็นต์ และท่าที่สามารถทำนายได้ถูกอยู่ที่ 33.33 เปอร์เซ็นต์ เวลาที่ใช้เฉลี่ย 1.1217 วินาที ในทางกลับกันการใช้สีของเสื้อที่มีอยู่ในการฝึกข้อมูลร่วมกับผู้ร่วมการทดลองที่มีอยู่ในการฝึกข้อมูล ทำให้ความสามารถในการทำนายเพิ่มขึ้นอย่างมาก โดยสามารถทำนายได้ 100 เปอร์เซ็นต์ และความถูกต้องอยู่ที่ 66.66 เปอร์เซ็นต์ เวลาที่ใช้เฉลี่ย 1.1227 วินาที

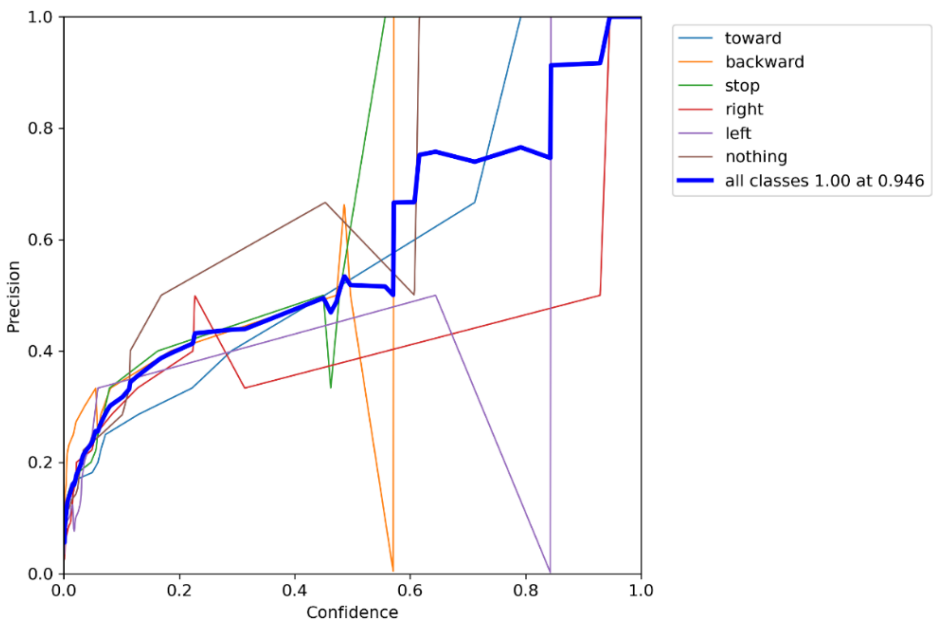
รูปที่ 8 แสดงค่า Precision เทียบกับค่าความเชื่อมั่น (Confidence) ผลคือยิ่งมีความเชื่อมั่นมากยิ่งมีความแม่นยำมากขึ้นตาม และรูปที่ 9 แสดงค่า F_1 เทียบกับค่าความเชื่อมั่น ค่า F_1 สูงสุดคือ 0.6 ที่ค่าความเชื่อมั่น 0.55

ตารางที่ 1 แสดงเวลาที่ใช้ในการหาตำแหน่งและจำแนกท่าทางคนโดยการสุ่มภาพจากภาพที่ไม่ได้ใช้ในการฝึก

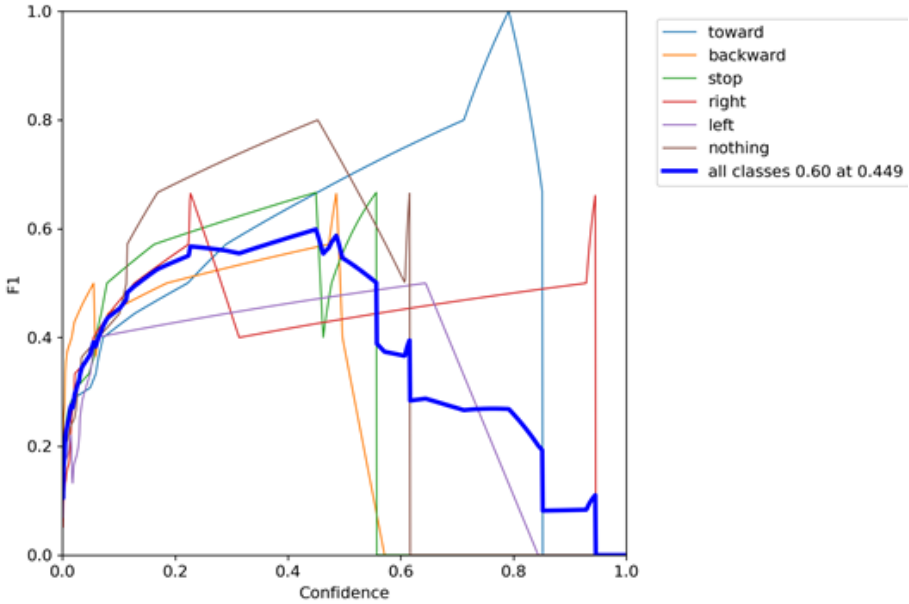
ภาพที่	ท่าทางของภาพ	ท่าทางที่ตรวจจับได้	เวลาที่ใช้
1	Nothing	Toward	0.054s
2	Right	Right	0.054s
3	Right	Right	0.028s
4	Left	Left	0.027s
5	Toward	Toward	0.029s
6	Nothing	Toward	0.054s
7	Stop	Backward	0.051s
8	Left	Right	0.027s
9	Nothing	Toward	0.027s
10	Backward	Backward	0.033s
เวลาที่ใช้เฉลี่ย			0.0384

ตารางที่ 2 เวลาที่ใช้ในการหาตำแหน่งและจำแนกท่าทางคนโดยใช้ผู้ร่วมการทดลอง 2 คน (ส่วนที่แรเงาคือผู้ ร่วมการทดลองที่ไม่ได้มีภาพอยู่ในการฝึกข้อมูล และส่วนที่ไม่ได้แรเงาคือผู้ร่วมการทดลองที่มีภาพอยู่ในการฝึกข้อมูล)

ท่าทางจริง	ทำนายได้ ครั้งที่ 1	เวลาที่ใช้	ทำนายได้ ครั้งที่ 2	เวลาที่ใช้	ทำนายได้ ครั้งที่ 3	เวลาที่ใช้
Nothing	-	-	Toward	1.125	Toward	1.120
Toward	-	-	Toward	1.126	Toward	1.124
Backward	-	-	Backward	1.121	Backward	1.128
Stop	Right	1.117	Backward	1.128	Backward	1.117
Right	Left	1.121	Left	1.121	Right	1.129
Left	Left	1.127	Left	1.121	Left	1.122
เวลาเฉลี่ย		1.1217		1.1237		1.123
Nothing	-	-	Toward	1.120	Toward	1.113
Toward	Toward	1.118	Toward	1.117	Toward	1.125
Backward	-	-	Backward	1.125	Backward	1.124
Stop	Backward	1.126	-	-	Backward	1.125
Right	Right	1.126	Right	1.117	Right	1.124
Left	Left	1.114	Left	1.112	Left	1.125
เวลาเฉลี่ย		1.121		1.1182		1.1227



รูปที่ 8 ค่าความแม่นยำเทียบกับค่าความเชื่อมั่นของแต่ละท่าทาง



รูปที่ 9 ค่าเฉลี่ยของความแม่นยำและความถูกต้องเทียบกับค่าความเชื่อมั่นของแต่ละท่าทาง

5. สรุปผลการทดลอง

ผลการทดลองในตารางที่ 1 เป็นการตัดปัญหาเรื่องสัญญาณรบกวนและความไม่แน่นอนของภาพ ทำให้ง่ายต่อการวิเคราะห์ แต่ภาพที่ใช้ในการทดลองยังคงเป็นผู้ร่วมการทดลองที่ใช้ในการฝึกข้อมูล ความถูกต้องอยู่ที่ 50 เปอร์เซ็นต์ ถือว่ายังต่ำอยู่ ส่วนตารางที่ 2 ซึ่งเป็นการทดลองแบบเรียลไทม์ จากผลการทดลองพบว่ายิ่งใช้ผู้ร่วมการทดลองที่มากขึ้นและมีความหลากหลายของชุดที่มากจะช่วยให้ความถูกต้องและแม่นยำมีค่ามากขึ้น และการทำนายที่ผิดพลาดเกินจากท่าทางมีความคล้ายกันอย่างมากนั่นคือท่า Nothing และ Toward หากมองจากทางด้านหน้าจะพบว่ามีความแตกต่างน้อยมาก อีกท่าที่มีความคล้ายกันคือ Stop และ Backward ต่างกันแค่แบมือกับกำมือเท่านั้น ส่วนรูปที่ 7 ค่า Precision และ Recall มีการแกว่งไปมาอย่างมากแสดงถึงความไม่แน่นอนของสองค่านี้ ซึ่งเกิดจากการทายท่าทางที่ผิดพลาดของท่าทางสองคู่ด้านบน สุดท้ายคือรูปที่ 9 แสดงให้เห็นว่าค่าความเชื่อมั่นมากไม่ได้แสดงถึงความถูกต้องเสมอไป

จากสรุปผลการทดลองข้างต้นสามารถนำมาสรุปเป็นผลการทดลองรวมดังนี้ คือจำเป็นต้องเพิ่มภาพที่ใช้ในการฝึกข้อมูลให้มากกว่านี้และใช้ผู้ร่วมการทดลองที่มากขึ้น รวมไปถึงจำนวนครั้งในการฝึกข้อมูล เพื่อให้สามารถทำนายท่าทางได้อย่างครอบคลุมกับทุกคนอย่างมีประสิทธิภาพ ซึ่งการเตรียมข้อมูลในการฝึกข้อมูลต้องใช้เวลาที่มาก โดยจะมีการเพิ่มชุดข้อมูลที่มากขึ้นในงานวิจัยต่อไป

6. กิตติกรรมประกาศ

ขอขอบคุณสาขาวิชาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเอเชียอาคเนย์ ที่เอื้อเฟื้อสถานที่ในการทดลอง

7. เอกสารอ้างอิง

[1] Kijdech D. (2020). Weed Classification by Using Convolution Neural Network for Studying and Weed Eliminate Robot, The 7th SAU National Interdisciplinary Conference 2020, 29-30 May 2020.

[2] Kijdech D. and Vongbunyong S. (2021). Artificial Intelligence in Localization and Classification of Cactus for Automatic Watering Works, The 8th SAU National Interdisciplinary Conference 2021, 4-5 June 2021.

[3] Kuznetsova A., Maleva T. and Soloviev V., (2020). "Detecting Apples in Orchards Using YOLOv3 and YOLOv5 in General and Close- Up Images," Advances in Neural Networks – ISNN, pp. 233-243.

[4] Yang G., Feng W., Jin J., Lei Q., Li X., Gui G. and Wang W., (2020). "Face Mask Recognition System with YOLOV5 Based on Image Recognition," IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, pp. 1398-1404.

[5] Benjdira B., Khursheed T., Koubaa A., Ammar A. and Ouni K., (2019). "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," Proceedings of the 1st International Conference on Unmanned Vehicle Systems (UVS), Muscat, Oman, 5-7 February.

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

[7] Jiang Z., Zhao L., Li S. and Jia Y., (2020). "Real-time object detection method based on improved YOLOv4- tiny," Computer Vision and Pattern Recognition.

[8] He K., Gkioxari G., Dollar P. and Girshick R. (2021). Mask R-CNN, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969

[9] Myint C., and Nu Win N., (2016). "Position and Velocity control for Two- Wheel Differential Drive Mobile Robot," International Journal of Science, Engineering and Technology Research (IJSETR) Volume 5, Issue 9, September 2016.

[10] Mondada F., Franzi E. and Ienne P., (2005) "Mobile robot miniaturisation: A tool for investigation in control algorithms," Experimental Robotics III, pp. 501–513.

[11] Desouza G. N. and Kak A. C., (2002). "Vision for mobile robot navigation: a survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(2), pp. 237–267.

[12] Yang G., Feng W., Jin J., Lei Q., Li X., Gui G. and Wang W., "Face Mask Recognition System with YOLOV5 Based on Image Recognition," IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 2020, pp. 1398-1404.

[13] Dlužnevskij D., StefanoviČ P. and Ramanauskaite S., (2021). "Investigation of YOLOv5 Efficiency in iPhone Supported Systems," Baltic Journal of Modern Computing , 9(3), pp. 333-344.

ประวัติผู้เขียนบทความ

นาย ดำรงค์ศักดิ์ กิจเดช
การศึกษา : วิศวกรรมศาสตร-มหาบัณฑิต (วิศวกรรมเครื่องกล)
งานวิจัยที่สนใจ : แขนกลอัตโนมัติ, หุ่นยนต์สำรวจ, ระบบวิชั่น, หุ่นยนต์คล้ายมนุษย์, ปัญญาประดิษฐ์

รศ.ดร. วีระพันธ์ ด้วงทองสุข
การศึกษา : วิศวกรรมศาสตร-ดุษฎีบัณฑิต สาขาเครื่องกล วิศวกรรมเครื่องกล
งานวิจัยที่สนใจ : การเพิ่มสมรรถนะการถ่ายเทความร้อนของอุปกรณ์ทางความร้อน, การไหลสองสถานะ, ของไหลนาโน, การไหลในช่องทางการไหลขนาดเล็ก