

การตรวจจับมังคุดด้วยการเรียนรู้เชิงลึก Mangosteen Detection Using Deep Learning

ราตรี คำโมง (Ratree Kummong)*

Received: February 16, 2021

Revised: April 28, 2021

Accepted: May 31, 2021

* ผู้นิพนธ์ประสานงาน: ราตรี คำโมง (Ratree Kummong) อีเมล: ratree@uru.ac.th

บทคัดย่อ

เทคโนโลยีด้านหุ่นยนต์ได้รับความสนใจและพัฒนาสูงขึ้นอย่างรวดเร็ว มีการพัฒนาหุ่นยนต์มาใช้เก็บผลไม้แทนมนุษย์ หุ่นยนต์นี้อาศัยกล้องถ่ายภาพและคอมพิวเตอร์ประมวลผลตรวจจับภาพเพื่อจำแนกตำแหน่งและผลไม้ที่เหมาะสมสำหรับควบคุมแขนกลในการเก็บผลไม้ เทคนิคการตรวจจับภาพจึงเป็นเทคโนโลยีที่สำคัญในการพัฒนาเทคโนโลยีการมองเห็นของหุ่นยนต์ มังคุดเป็นผลไม้ที่ประเทศไทยส่งออกเป็นอันดับต้น ๆ ของโลกมีลำต้นสูงเก็บเกี่ยวลำบาก การตรวจจับมังคุดที่อยู่บนต้นต้องจำแนกมังคุดอ่อน มังคุดแก่หรือสุกที่อยู่บนกิ่งและอาจถูกบังด้วยใบ งานวิจัยนี้จึงนำเสนอการตรวจจับมังคุดที่อยู่บนต้นที่ได้จากภาพถ่าย ด้วยเทคนิคการเรียนรู้เชิงลึกโครงข่ายประสาทคอนโวลูชันแนล เปรียบเทียบกับฟาสเตอร์อาร์ชีเอ็นเอ็น โดยขั้นตอนการวิจัยมีดังนี้ 1) การเตรียมภาพ 2) การฝึกสอนการเรียนรู้เชิงลึกตรวจจับมังคุดแก่หรือสุก และ 3) การทดสอบตรวจจับมังคุดแก่หรือสุก ประเมินประสิทธิภาพโมเดล ด้วยคอนฟิวชันเมทริกซ์ (Confusion Matrix) ผลการทดลองพบว่าแบบจำลองโครงข่ายประสาทคอนโวลูชันแนลสามารถจำแนกมังคุดแก่หรือสุกได้ค่าความถูกต้องร้อยละ 50.62 ค่าความเที่ยงร้อยละ 26.42 ค่าความครบถ้วนร้อยละ 93.33 และ เอฟวันสกอร์ ร้อยละ 41.18 แบบจำลองฟาสเตอร์อาร์ชีเอ็นเอ็นสามารถจำแนกมังคุดแก่หรือสุกได้ค่าความถูกต้องร้อยละ 90.22 ค่าความเที่ยงร้อยละ 63.16 ค่าความครบถ้วนร้อยละ 48.98 และ เอฟวันสกอร์ ร้อยละ 55.17

คำสำคัญ: มังคุด การตรวจจับมังคุด การเรียนรู้เชิงลึก

Abstract

Robot technology has developed more and more rapidly. Robot were developed to harvest fruit instead of humans. Robot uses a camera and computer to process images, detect fruit and location on branches to control a mechanical arm to pick fruit. Image detection technique is therefore an important technology use to development of vision robot technology. Thailand's mangosteen export is one of the top in the world. A Mangosteen is a tall trunk which makes it difficult to harvest, to detect mangosteen on plants, must be classify is young or ripe, lies on the branches and may be covered by the leaves. This research presents the detection of mangosteen on plants by image using Convolutional Neural Network (CNN) compare with Faster R-CNN deep learning framework. The research methods are as follows 1) image preparation 2) training deep learning to mature or ripe mangosteen detection 3) testing to mature or ripe mangosteen detection. The performance of detection model employs the confusion matrix. The results show that can mature or ripe mangosteen detection, the model of CNN is an accuracy of 50.62%, precision of 26.42%, recall of 93.33% and F1-Score of 41.18%, the Faster R-CNN is an accuracy of 90.22%, precision of 63.16%, recall of 48.98% and F1-Score of 55.17%.

Keywords: Mangosteen, Mangosteen Detection, Deep Learning.

* มหาวิทยาลัยราชภัฏอุดรดิตถ์

* Uttaradit Rajabhat University

1. บทนำ

ในปัจจุบันเทคโนโลยีด้านหุ่นยนต์ได้รับความสนใจและพัฒนามากขึ้นอย่างรวดเร็ว มีการนำมาใช้ในอุตสาหกรรมต่าง ๆ และส่งเสริมให้มีการคิดค้นเทคนิคใหม่ ๆ เพื่อให้หุ่นยนต์สามารถทำงานได้ใกล้เคียงกับมนุษย์มากที่สุด

หุ่นยนต์เริ่มมีบทบาทมากขึ้นในด้านการเกษตร เช่น หุ่นยนต์เก็บผลส้ม พัฒนาโดยบริษัทเอเนอร์จี้ เทคโนโลยี [1] อาศัยคอมพิวเตอร์และกล้องถ่ายภาพสีในการจำแนกตำแหน่งและผลส้มที่เหมาะสมสำหรับควบคุมแขนกลในการเก็บแต่ละลูก และหุ่นยนต์เก็บผลสตอร์วเบอร์รี่ เป็นผลงานของนักวิจัยชาวญี่ปุ่นสถาบันวิจัย IAM-BRAIN [2] โปรแกรมที่ถูกเขียนในตัวหุ่นยนต์จะทำการวิเคราะห์ภาพที่เก็บได้ด้วยหลักการประมวลผลภาพ (Image Processing) เพื่อวิเคราะห์หาว่าผลสตอร์วเบอร์รี่นั้นสุกพอดีตามที่ต้องการหรือไม่ ถ้าสุกโดยที่ผลมีสีแดงมากกว่า 80 เปอร์เซ็นต์ เครื่องก็จะทำการตัดที่ก้านแล้วเก็บผลสตอร์วเบอร์รี่ไว้ในถัง การทำงานของหุ่นยนต์เก็บผลไม้ มีการใช้เทคนิคการประมวลผลภาพเป็นปัจจัยหลักในการหาคุณลักษณะเป้าหมายที่ต้องการก่อนทำการเก็บผลไม้ ประกอบกับเทคโนโลยีการเรียนรู้เชิงลึกซึ่งเป็นเทคโนโลยีที่นิยมในปัจจุบันที่จะทำให้คอมพิวเตอร์เรียนรู้ได้ด้วยเหตุนี้ การวิจัยในส่วนของการจำแนกรูปภาพโดยใช้การเรียนรู้เชิงลึกจึงมีความสำคัญ

มัจจุเป็นผลไม้ที่ประเทศไทยส่งออกมากเป็นอันดับต้น ๆ ของโลก ไปสู่ในหลาย ๆ ประเทศทั่วโลกเช่น สหรัฐอเมริกา จีน ญี่ปุ่น และเนเธอร์แลนด์ ข้อมูลปี พ.ศ. 2562 พบว่ามีพื้นที่เก็บเกี่ยวมัจจุของประเทศไทยประมาณ 429,316 ไร่ [3] มีการส่งออก 409,255 ตัน มีมูลค่า 16,727.14 ล้านบาท [4] มัจจุถือเป็นผลไม้ที่โดดเด่น และมีชื่อเสียงของประเทศไทย ด้วยรูปลักษณะและสีส้มที่สวยงาม มีเนื้อนุ่ม รสหวาน จนได้รับการขนานนามว่าราชินีแห่งผลไม้ มัจจุเป็นไม้ยืนต้นขนาดกลางถึงขนาดใหญ่ ลำต้นทรงกลม สูงประมาณ 7-25 เมตร

เนื่องจากมัจจุเป็นผลไม้ที่มีลำต้นสูงจึงมีปัญหาในการเก็บเกี่ยวซึ่งมนุษย์ไม่สามารถเอื้อมมือไปเก็บผลมัจจุได้โดยตรง ดังนั้นงานวิจัยนี้จึงนำเสนอการตรวจจับมัจจุที่อยู่บนต้นเพื่อเป็นประโยชน์ต่อการมองเห็นของหุ่นยนต์ในการจำแนกมัจจุอ่อน มัจจุแก่หรือสุกที่อยู่บนกิ่งและอาจถูกบังด้วยใบเพื่อการเก็บเกี่ยวจากต้นแทนมนุษย์

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่ามีงานวิจัยเกี่ยวกับ

การตรวจจับภาพมัจจุหลายงานวิจัย แต่ไม่มีงานวิจัยใดที่มีการตรวจจับมัจจุที่อยู่บนต้น ได้แก่ งานวิจัยของ อซิช่าห์ (Azizah) และคณะ [5] ได้ประยุกต์ใช้การเรียนรู้เชิงลึกแบบโครงข่ายประสาทคอนโวลูชันแนล (Convolutional Neural Network หรือ CNN) ในการคัดแยกมัจจุที่ดี (Fine) และมีตำหนิ (Defect) สามารถตรวจจับมัจจุที่มีตำหนิได้ถูกต้องถึงร้อยละ 97 บาร์โกติ (Bargoti) และคณะ [6] ได้ตรวจจับผลไม้ในสวน ได้แก่ มะม่วง อัลมอนต์ และแอปเปิ้ล โดยใช้โครงข่ายตรวจจับวัตถุที่ล้ำสมัยอย่างฟาสเตอร์อาร์ซีเอ็นเอ็น (Faster R-CNN) วัดประสิทธิภาพการตรวจจับด้วยเอฟวันสกอร์ (F1-Score) มากกว่า 0.9 สำหรับมะม่วงและแอปเปิ้ล เมอส์สัน (Muresan) และคณะ [7] ได้ทดลองตรวจจับผลไม้และผักรูปทรงไข่ จำนวน 82,213 ภาพจากผลไม้และผัก 120 ชนิดด้วยการเรียนรู้เชิงลึกโครงข่ายประสาทคอนโวลูชันแนล สามารถตรวจจับได้ถูกต้องร้อยละ 95.23

จากการศึกษาโครงข่ายตรวจจับวัตถุ (Objection Detection Framework) ที่นิยมใช้คือ โยโล (YOLO หรือ You Only Look Once) เอสเอสดี (SSD หรือ Single-Shot Multibox) และฟาสเตอร์อาร์ซีเอ็นเอ็น แต่โยโลและ เอสเอสดี มีข้อด้อยในการตรวจจับวัตถุขนาดเล็กจึงมีความแม่นยำในการตรวจจับวัตถุที่น้อยกว่า ฟาสเตอร์อาร์ซีเอ็นเอ็น [8]

งานวิจัยนี้จึงนำเสนอการตรวจจับมัจจุที่อยู่บนต้นที่ได้จากภาพถ่าย ด้วยโครงข่ายตรวจจับการเรียนรู้เชิงลึกโครงข่ายประสาทคอนโวลูชันแนลเปรียบเทียบกับฟาสเตอร์อาร์ซีเอ็นเอ็นที่มีความสามารถเลือกบริเวณที่น่าจะเป็นวัตถุเพื่อนำมาสกัดคุณลักษณะ จากนั้นจึงแยกประเภทจนกระทั่งทราบถึงผลลัพธ์หรือวัตถุที่ตรวจจับ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการดำเนินงานวิจัยนี้ได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

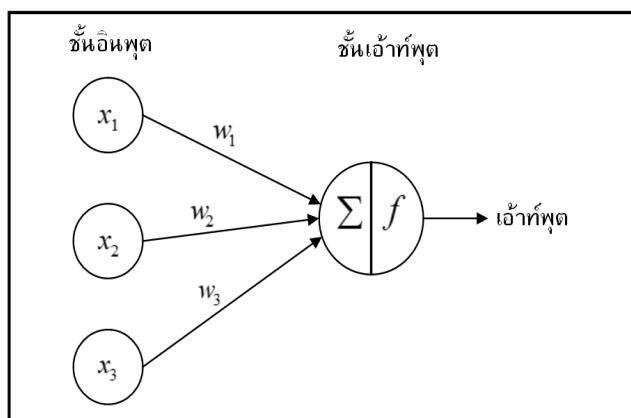
2.1 โครงข่ายประสาทเทียม (Artificial Neural Networks)

โครงข่ายประสาทเทียมเริ่มพัฒนาในปี ค.ศ. 1943 โดย วอร์เรน แมคคูลล็อก (Warren McCulloch) และ วอลเตอร์พิตต์ (Walter Pitts) [9] ได้คิดริเริ่มรูปแบบของโครงข่ายประสาทเทียมขึ้นซึ่งได้กลายเป็นจุดกำเนิดของการจัดรูปแบบคณิตศาสตร์ของโครงข่ายประสาทเทียม ที่จำลองการทำงานของเซลล์สมองมนุษย์ด้วยคอมพิวเตอร์ สามารถเรียนรู้และฝึกฝนได้เหมือน

มนุษย์ในการเรียนรู้ การจดจำแบบรูป (Pattern Recognition) และการอุปมาความรู้อื่น (Knowledge deduction)

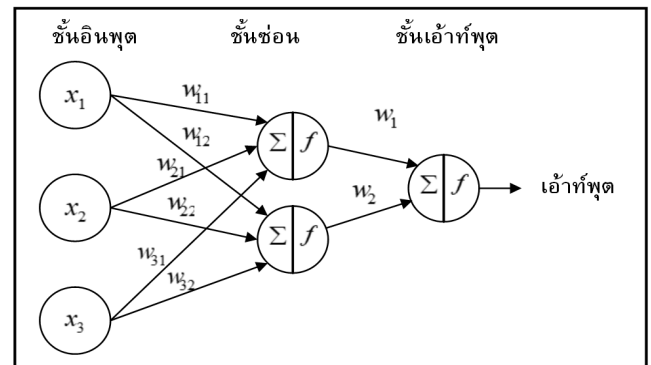
สมองมนุษย์ประกอบด้วย เซลล์ประสาท (Neuron) โยประสาทนำเข้า (Dendrite) และเส้นประสาทนำออก (Axon) โดยโยประสาทนำเข้าจะมีหน้าที่รับสัญญาณไฟฟ้าเคมีซึ่งส่งมาจากเซลล์ประสาทและส่งต่อไปยังเซลล์ประสาทที่อยู่ใกล้เคียง เมื่อเซลล์ประสาทได้รับสัญญาณไฟฟ้าเคมีเข้ามาถ้ามีค่าเกินค่าค่าหนึ่ง เซลล์ประสาทจะถูกกระตุ้นและส่งสัญญาณไปทางเส้นประสาทนำออกต่อไป เมื่อจำลองเป็นโครงข่ายประสาทเทียมจะให้โหนด (Node) ทำหน้าที่เลียนแบบเซลล์ประสาท แต่ละโหนดรับค่านำเข้าได้หลายค่า ผลการกระตุ้นหรือส่งค่าออกได้มาจากการคำนวณโดยอาศัยฟังก์ชันการกระตุ้น (Activation Function) กับผลรวมเชิงเส้นแบบถ่วงน้ำหนักของค่านำเข้า นักวิจัยหลายท่านได้แบ่งโครงข่ายประสาทเทียมออกเป็นสองแบบกว้าง ๆ คือ โครงข่ายชั้นเดียว (Single layer) และโครงข่ายหลายชั้น (Multi layer)

โครงข่ายประสาทเทียมชั้นเดียวเป็นโครงข่ายที่มีเพียงชั้นอินพุต (Input Layer) และชั้นเอาต์พุต (Output Layer) เท่านั้น โดยชั้นอินพุตทำหน้าที่รับข้อมูลอินพุตแล้วรวมทุกข้อมูลอินพุตและค่าน้ำหนักของแต่ละจุดที่เชื่อมโยงกันส่งไปยังชั้นเอาต์พุตเพื่อมาคำนวณทางคณิตศาสตร์ด้วยฟังก์ชันกระตุ้น ลักษณะโครงข่ายประสาทเทียมชั้นเดียวดังภาพที่ 1



ภาพที่ 1 ลักษณะโครงข่ายประสาทเทียมชั้นเดียว

โครงข่ายประสาทเทียมแบบหลายชั้น เป็นโครงข่ายที่มีชั้นซ่อนตั้งแต่ 1 ชั้นขึ้นไป เพื่อให้สามารถแก้ปัญหาที่มีความซับซ้อน ลักษณะโครงข่ายประสาทเทียมหลายชั้นดังภาพที่ 2



ภาพที่ 2 ลักษณะโครงข่ายประสาทเทียมหลายชั้น

2.2 การเรียนรู้เชิงลึก (Deep Learning)

การเรียนรู้เชิงลึกเป็นชนิดหนึ่งการเรียนรู้ด้วยตัวเองของเครื่องจักร (Machine Learning) ที่มีการคำนวณแบบไม่เชิงเส้น (nonlinear) หลายชั้น การเรียนรู้เชิงลึกเป็นโครงข่ายประสาทเทียมขนาดใหญ่ การเรียนรู้เชิงลึกได้รับความนิยมในการรู้จำภาพ โดยเป็นการใช้โครงข่ายประสาทเทียมประเภท โครงข่ายประสาทคอนโวลูชันแนล

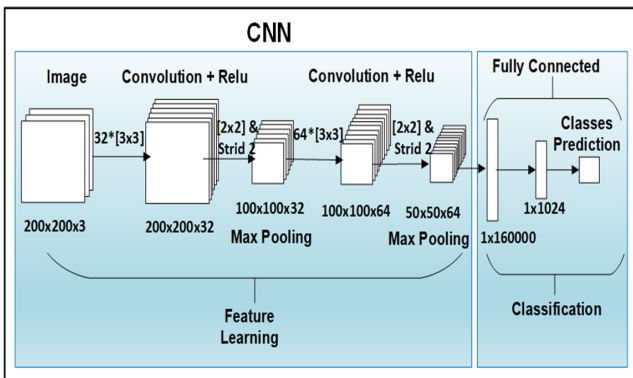
2.2.1 โครงข่ายประสาทคอนโวลูชันแนล โครงข่ายประสาทคอนโวลูชันแนล [10], [11] ใช้ในการเรียนรู้ด้วยตัวเองของเครื่องจักร นำมาประยุกต์ใช้กันอย่างแพร่หลายในการตรวจจับและจดจำรูปภาพ [12], [13] มีสถาปัตยกรรมเช่นเดียวกับโครงข่ายประสาทเทียมหลายชั้น เช่นเดียวกับเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron หรือ MLP) จึงเป็นการเรียนรู้เชิงลึกชนิดหนึ่ง

โครงข่ายประสาทคอนโวลูชันแนล ประกอบด้วย ชั้นอินพุต และชั้นเอาต์พุต รวมด้วยชั้นซ่อนหลายชั้น (Multiple Hidden Layer) การคำนวณในเลเยอร์ใช้หลักการคำนวณทางคณิตศาสตร์แบบสังวัตนาการ (Convolution) โครงสร้างของโครงข่ายประสาทคอนโวลูชันแนล ตามภาพที่ 3 ในชั้นซ่อนประกอบด้วย

1. ชั้นคอนโวลูชัน (Convolution) เป็นชั้นที่ประยุกต์ด้วยการนำตัวกรองหลายตัวมาปฏิบัติการทางคณิตศาสตร์แบบสังวัตนาการกับข้อมูลนำเข้ารูปภาพ ตัวกรองที่นำมาใช้เช่นขนาด 3x3 เป็นต้น แต่ละส่วนของรูปภาพ ชุดของการปฏิบัติการทางคณิตศาสตร์แปลงเป็นตัวเลขเชิงเดี่ยวส่งต่อไปยังฟังก์ชันการกระตุ้น (Activation) จนได้ผลลัพธ์ออกมาฟังก์ชันกระตุ้นที่ใช้คือ เรคตีไฟด์ลิเนียร์ยูนิต (Rectified Linear Units หรือ ReLU) ผลลัพธ์ที่ได้เรียกว่าฟีเจอร์แมป (Feature Map)

2. ขั้นพูลลิง (Pooling) อยู่ถัดจากชั้นคอนโวลูชันทำหน้าที่ลดขนาด (Downsample) ของรูปภาพผลลัพธ์จากชั้นคอนโวลูชันเพื่อลดขนาดพีเจอร์แม็บ ให้เล็กลงทำให้ประมวลผลได้เร็วกว่าโดยใช้ฟังก์ชันค่าเฉลี่ย ฟังก์ชันค่าต่ำสุด และฟังก์ชันค่าสูงสุดในการคำนวณ ส่วนมากนิยมใช้ฟังก์ชันค่าสูงสุดในการคำนวณเรียกขั้นนี้ว่าแมกซ์พูลลิง (Max Pooling)

3. ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected) ทำหน้าที่ในการจำแนกวัตถุจากคุณลักษณะที่แยกออกมาหลังจากลดขนาดในขั้นพูลลิง ในขั้นนี้ทุก ๆ โหนดถูกเชื่อมต่อกับทุก ๆ โหนดในชั้นก่อนหน้านี้



ภาพที่ 3 โครงสร้างโครงข่ายประสาทคอนโวลูชันแนล

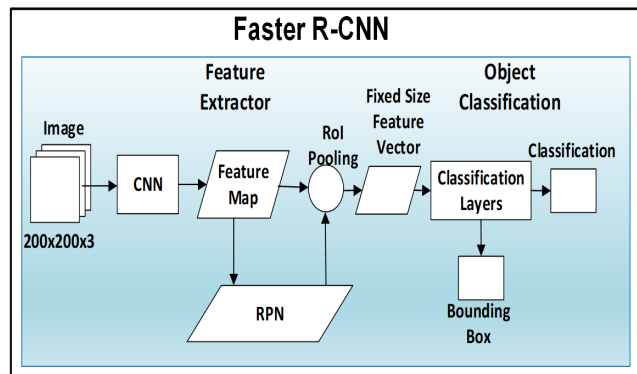
2.2.2 ฟาสเตอร์อาร์ซีเอ็นเอ็น

ฟาสเตอร์อาร์ซีเอ็นเอ็น สร้างและพัฒนาโดย เร็น (Ren) และคณะ [14] เป็นวิธีการเรียนรู้เชิงลึกของเครื่องบนพื้นฐานโครงข่ายประสาทคอนโวลูชันแนล พัฒนามาจากซีเอ็นเอ็น อาร์ซีเอ็นเอ็น (R-CNN) และฟาสเตอร์อาร์ซีเอ็นเอ็น (Fast R-CNN) ที่นิยมใช้ในการตรวจจับและรู้จำวัตถุ (Object Detection and Recognition) ที่ปรับปรุงข้อจำกัดหลายอย่าง เพื่อเพิ่มความแม่นยำและความรวดเร็วในการทำนาย โครงสร้างของฟาสเตอร์อาร์ซีเอ็นเอ็น ดังภาพที่ 4 ประกอบด้วย

1. ส่วนฐานเป็นโครงข่ายประสาทคอนโวลูชันแนล ทำหน้าที่นำเข้าข้อมูลรูปภาพมาสกัดคุณลักษณะเด่นจนได้พีเจอร์แม็บ แล้วส่งออกไป 2 ทาง ทางที่ 1 ส่งไปส่วนเสนอบริเวณ (Region Proposal Network หรือ RPN) และทางที่ 2 ส่งไปรอค่าจากส่วนเสนอบริเวณ ก่อนการ โรลพูลลิง (RoI Pooling หรือ Region of Interest Pooling)

2. ส่วนเสนอบริเวณ ใช้สำหรับตรวจจับที่สกัดเอาบริเวณที่น่าจะเป็นวัตถุจากพีเจอร์แม็บ แล้วส่งค่าบริเวณ (Region) ที่ได้สู่กระบวนการ RoI Pooling

3. ส่วนโรลพูลลิงทำหน้าที่จำแนกประเภท (Classification) ด้วยการนำค่าบริเวณ (Region) จาก ส่วนเสนอบริเวณมาพูลลิงกับพีเจอร์แม็บ ของทางที่ 2 เพื่อสร้างพีเจอร์เวกเตอร์ (Feature Vector) ที่มีขนาดคงที่ส่งไปชั้นเชื่อมโยงสมบูรณ์ (Fully Connected Layers) ทำการคำนวณผลลัพธ์แล้วแสดงการทำนายออกมาเป็น 2 ลักษณะ คือจำแนกประเภท และกรอบรอบบริเวณ (Bounding Box) วัตถุ เพื่อตอบว่า มีวัตถุใดอยู่บริเวณใดของภาพ



ภาพที่ 4 โครงสร้างของฟาสเตอร์อาร์ซีเอ็นเอ็น

2.3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยหลายงานวิจัยไม่พบงานวิจัยที่เกี่ยวกับการจดจำมัจจุสำหรับการเก็บเกี่ยวด้วยโครงข่ายประสาทเทียม การเรียนรู้เชิงลึก แต่มีงานวิจัยที่เกี่ยวกับการจดจำผลไม้ อื่น ๆ สำหรับการเก็บเกี่ยวด้วยโครงข่ายประสาทเทียมการเรียนรู้เชิงลึกดังต่อไปนี้

อซีซาร์ห์ (Azizah) และคณะ [5] ได้นำโครงข่ายประสาทเทียม การเรียนรู้เชิงลึกมาประยุกต์ใช้ในการคัดแยกมังคุดที่ดีและมีตำหนิ เมอร์สัน (Muresan) และคณะ [7] ได้ทดลองตรวจจับผลไม้และผักรูปทรงไข่ โจว (Zhou) และคณะ [15] ตรวจจับและคัดแยกมะเขือเทศพันธุ์ต่าง ๆ จากการศึกษาวิจัยที่ใช้ฟาสเตอร์อาร์ซีเอ็นเอ็นในการตรวจจับผลไม้พบหลายงานวิจัย ได้แก่ งานวิจัยของ ฟู่ (Fu) และคณะ [16] ตรวจจับผลไม้กีว งานวิจัยของ ไม (Mai) และคณะ [17] คัดแยกผลไม้ขนาดเล็ก และงานวิจัยของ บาร์โกติ (Bargoti) และคณะ [18] ได้ตรวจจับผลไม้ในสวน ได้แก่ มะม่วง อัลมอนต์ และแอปเปิ้ล

3. วิธีดำเนินการวิจัย

3.1 การเตรียมข้อมูล

รูปภาพที่ใช้ในการวิจัยเพื่อการพิจารณาการตรวจจับภาพมั่งคุดและจดจำภาพมั่งคุดแก่หรือสุกโดยใช้การเรียนรู้เชิงลึกที่นำมาทดลองได้มาจากการถ่ายภาพมั่งคุดจากสวนศรีเพชร ตำบลแม่พูล อำเภอลับแล จังหวัดอุตรดิตถ์ กำหนดดังนี้ เป็นภาพสีอาร์สบี (RGB) ชนิดเจเป็ก (jpeg) จำนวนรูปภาพทั้งหมดที่ใช้วิจัยทั้งหมด 420 ภาพ แบ่งออกเป็นดังต่อไปนี้

1) ชุดที่ 1 รูปภาพชุดฝึกสอน (Training Data) 320 ภาพ ความละเอียด 250x250 พิกเซล ภาพละ 1 ลูก ประกอบด้วย ลูกมั่งคุดดิบบนกิ่ง 100 ภาพ ลูกมั่งคุดแก่หรือสุกบนกิ่ง 100 ภาพ ลูกมั่งคุดดิบใบบัง 30 ภาพ ลูกมั่งคุดแก่หรือสุกใบบัง 30 ภาพ ลูกมั่งคุดดิบกิ่งบัง 30 ภาพ ลูกมั่งคุดแก่หรือสุกกิ่งบัง 30 ภาพ

2) ชุดที่ 2 รูปภาพชุดทดสอบ (Testing Data) 100 ภาพ ความละเอียด 1024x682 พิกเซล ประกอบด้วย รูปภาพมั่งคุดที่ถ่ายบนต้นที่มีทั้งลูกมั่งคุดใบและกิ่ง 80 ภาพ รูปภาพมะปรางที่ถ่ายบนต้นที่มีทั้งลูกมะปรางใบและกิ่ง 20 ภาพ แบ่งเป็นข้อมูลทดสอบ 10 ชุด แต่ละชุดมีภาพมั่งคุด 8 ภาพ มะปราง 2 ภาพ

3.2 เครื่องมือที่ใช้ในการวิจัย

1) ฮาร์ดแวร์ ใช้เครื่องไมโครคอมพิวเตอร์ หน่วยประมวลผลกลางคอร์ไอห้า (CORE i5) 3.4 กิกะเฮิร์ต หน่วยความจำหลัก 8 กิกะไบต์ ฮาร์ดดิสก์เอสเอสดี 256 กิกะไบต์

2) ซอฟต์แวร์ ใช้อนาคอนดา (Anaconda) ในการจัดการสภาพแวดล้อมต่าง ๆ เพื่อใช้ในการทดสอบฟาสเตอร์อาร์ซีเอ็นเอ็นที่เขียนด้วยโปรแกรมภาษาไพทอนรุ่น 3.7 (Python 3.7) เขียนคำสั่งและทดสอบโปรแกรมภายใต้ อะตอม (ATOM) ทั้งหมดนี้เป็นซอฟต์แวร์ที่ไม่มีลิขสิทธิ์

3.3 การแยกคุณลักษณะมั่งคุดแก่หรือสุก

พิจารณาจากเส้นขอบลูกมั่งคุด พื้นสีผิวมั่งคุด มั่งคุดแก่หรือสุกพื้นที่สีผิวผลมั่งคุดมีสีแดง แดงอมม่วง ม่วงเข้มไปจนถึงม่วงดำ ดังภาพที่ 5

3.4 การทดลอง

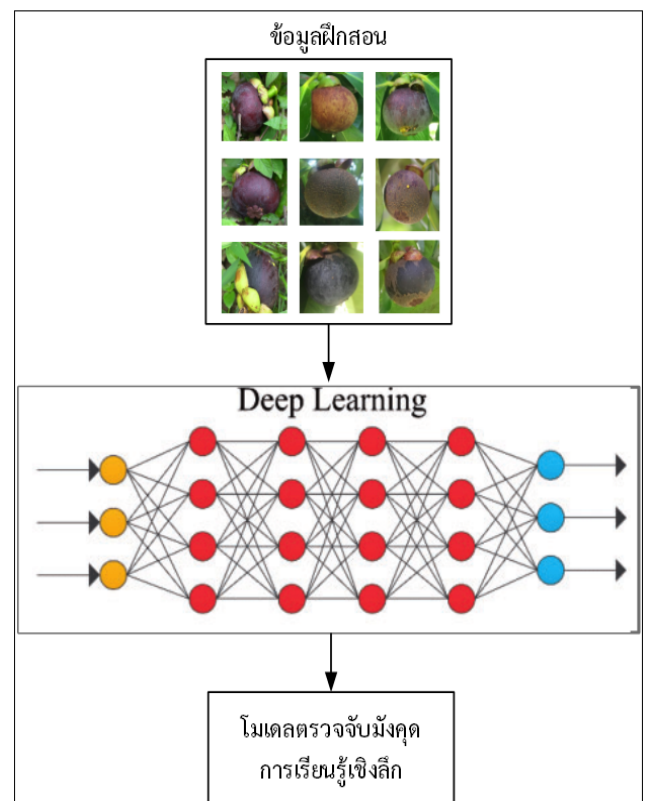
งานวิจัยนี้เป็นการเรียนรู้เชิงลึกคอนโวลูชันแนล เปรียบเทียบกับฟาสเตอร์อาร์ซีเอ็นเอ็น มาตรวจจับและตัดแยกมั่งคุดแก่หรือสุก ในการทดลองมี 2 ขั้นตอนคือ ขั้นตอนการฝึกสอน (Training Phase) และขั้นตอนการทดสอบ (Testing Phase)

1) ขั้นตอนการฝึกสอน นำรูปภาพชุดฝึกสอน 320 ภาพ นำมาใช้กัน 10 ชุด (10-Fold Cross Validation) ด้วยการสุ่ม

แบบเจาะจงให้แต่ละชุดมีภาพครบทุกแบบ ๆ ละเท่า ๆ กัน ไปสอนโมเดลด้วยอัลกอริทึมคอนโวลูชันแนล 10 รอบ (Epochs) และฟาสเตอร์อาร์ซีเอ็นเอ็น อีกจำนวน 20,000 รอบ ดังภาพที่ 6 จนกระทั่งได้โมเดลตรวจจับและตัดแยกมั่งคุดแก่หรือสุก อัลกอริทึมละ 10 โมเดล

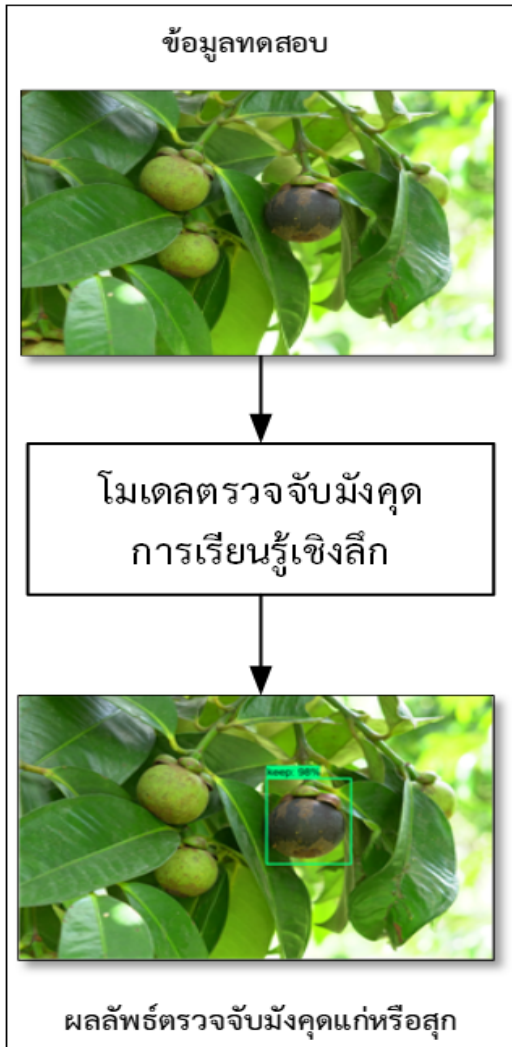


ภาพที่ 5 มั่งคุดแก่หรือสุก



ภาพที่ 6 กระบวนการฝึกสอนโมเดลตรวจจับมั่งคุด

2) ขั้นตอนการทดสอบ นำโมเดลตรวจจับมั่งคุดด้วยคอนโวลูชันแนล 10 โมเดล และฟาสเตอร์อาร์ชีเอ็นเอ็น 10 โมเดล แต่ละโมเดลนำไปทดสอบตรวจจับมั่งคุดแก่หรือสุกกับรูปภาพชุดทดสอบที่ละชุดที่แบ่งไว้ทั้ง 10 ชุด แบบไม่ซ้ำกัน ดังภาพที่ 7



ภาพที่ 7 การทดสอบโมเดลตรวจจับมั่งคุด

3.5 การวัดประสิทธิภาพผลการตรวจจับมั่งคุด

1) เครื่องมือวัดการฝึกสอนโครงข่าย จากภาพที่ 6 ในกระบวนการฝึกสอนเพื่อให้โครงข่ายประสาทเทียมได้เรียนรู้ได้ใช้แบคพรอพาเกชัน (Back propagation) ในการปรับปรุงค่าน้ำหนัก (Weigh) โดยเริ่มจากการหาความผิดพลาด (Error) ด้วยความผิดพลาดกำลังสองเฉลี่ย (Mean Square Error หรือ MSE) ที่คำนวณได้จากเอาท์พุตของโครงข่ายประสาทเทียม นำมาเปรียบเทียบกับเฉลย (Target)

เมื่อได้ค่าความผิดพลาดก็จะแพร่ค่าความผิดพลาดที่ได้กลับไปยังค่าน้ำหนักต่าง ๆ ซึ่งมีสูตรความผิดพลาดกำลังสองเฉลี่ย ดังสมการที่ 1

$$E_{total} = \sum \frac{1}{2}(target - output)^2 \quad (1)$$

โดยที่ target คือเวกเตอร์ภาพเฉลย และ output เป็นเวกเตอร์ที่โครงข่ายประสาทเทียมที่ได้ทำนายไว้

2) เครื่องมือวัดการทดสอบโครงข่ายที่นำมาวัดประสิทธิภาพโมเดลตรวจจับมั่งคุดคือคอนฟิวชันเมทริกซ์ (Confusion Matrix) เพื่อประเมินผลการตรวจจับมั่งคุด (Prediction) เทียบกับค่าจริง (Actual) คอนฟิวชันเมทริกซ์ของการจำแนกข้อมูลประเภท 2 คลาส ได้แก่ คลาสมั่งคุดแก่หรือสุก แทนด้วย $Y = 0$ และคลาสที่ไม่ใช่มั่งคุดสุกหรือแก่ แทนด้วย $Y = 1$ ตามตารางที่ 1

ตารางที่ 1 คอนฟิวชันเมทริกซ์

		ผลตรวจจับ	
		$Y = 0$	$Y = 1$
ค่าจริง	$Y = 0$	TP	FN
	$Y = 1$	FP	TN

TP (True Positive) คือ โมเดลทำนายว่า $Y = 0$ และคนบอกว่า $Y = 0$

TN (True Negative) คือ โมเดลทำนายว่า $Y = 1$ และคนบอกว่า $Y = 1$

FP (False Positive) คือ โมเดลทำนายว่า $Y = 0$ และคนบอกว่า $Y = 1$

FN (False Negative) คือ โมเดลทำนายว่า $Y = 1$ และคนบอกว่า $Y = 0$

ค่าความถูกต้อง (Accuracy) ตามที่สมการที่ 2

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

ค่าความเที่ยง (Precision) เป็นค่าที่บอกว่าโมเดลตรวจจับมั่งคุดแก่หรือสุกได้ถูกต้องเท่าไรดังสมการที่ 3

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

ค่าความครบถ้วน (Recall) เป็นค่าที่บอกว่าโมเดลตรวจจับมัลแวร์หรือสื่อกี่ตัวถูกต้องเป็นอัตราส่วนเท่าไรของมัลแวร์หรือสื่อกี่ตัวทั้งหมด ดังสมการที่ 4

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

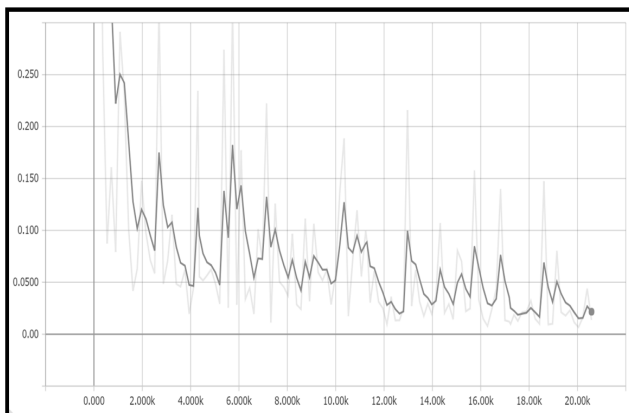
เอฟวันสกอร์คือค่าประสิทธิภาพโดยรวมโดยการนำค่าความเที่ยงกับค่าความครบถ้วนมาหาค่าเฉลี่ย ซึ่งค่าสูง ๆ แสดงว่าโมเดลมีประสิทธิภาพดี มีสูตรคำนวณตามสมการที่ 5

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

4. ผลการทดลอง

4.1 ผลการฝึกสอน

ผลการฝึกสอนที่ได้คือค่าความผิดพลาดที่เกิดจากการตรวจจับมัลแวร์โดยใช้โครงข่ายประสาทเทียมฟาสเตอร์อาร์ซีเอ็นเอ็น ได้ผลการฝึกสอนของการตรวจจับมัลแวร์ดังภาพที่ 8 แกนแนวนอนคือค่าของจำนวนครั้งที่ป้อนรูปภาพเข้าไปให้โครงข่ายประสาทเทียมได้ฝึกสอน และแกนแนวตั้งคือค่าความผิดพลาดที่ได้จากการฝึกสอน เส้นสีส้มจาง ๆ หมายถึงลักษณะของเส้นจริงก่อนการถูกปรับให้เรียบ และเส้นสีส้มเข้มหมายถึงลักษณะของเส้นที่ถูกปรับให้เรียบเพื่อดูลักษณะของกราฟได้ง่ายขึ้น เมื่อนำข้อมูลมาวิเคราะห์ในรูปแบบกราฟแล้วจะเห็นว่าค่าความผิดพลาดในการฝึกสอนของโครงข่ายประสาทเทียมฟาสเตอร์อาร์ซีเอ็นเอ็นนั้นค่าความผิดพลาดที่เกิดขึ้นมากที่สุดจะอยู่ในขั้นตอนช่วงเริ่มต้น ค่าผิดพลาดกำลังสองเฉลี่ยมีค่าเท่ากับ 2.0107×10^{-5} และค่าความผิดพลาดที่เกิดขึ้นน้อยที่สุดจะอยู่ในขั้นตอนช่วงท้าย ค่ากำลังสองเฉลี่ย มีค่าเท่ากับ 0.0137×10^{-5}



ภาพที่ 8 กราฟแสดงค่าความผิดพลาดของการฝึกสอนโครงข่ายประสาทเทียม

4.2 ผลการทดสอบ

นำทั้ง 2 โมเดลไปตรวจจับภาพชุดทดสอบ จำนวน 100 ภาพ ประกอบด้วย

- 1) ภาพมัลแวร์ 80 ภาพ ที่มีลูกติดอย่างเดียว 62 ภาพ ลูกสุกอย่างเดียว 1 ภาพ มีทั้งดิบและสุก 17 ภาพ และทั้ง 80 ภาพ มีลูกติดจำนวน 332 ลูก ลูกแก่หรือสุก 24 ลูก
- 2) ภาพมะพร้าว 20 ภาพ ที่มีลูกติดอย่างเดียว 10 ภาพ ลูกสุกอย่างเดียว 9 ภาพ มีทั้งดิบและสุก 1 ภาพ และทั้ง 20 ภาพ มีมะพร้าวดิบ 17 ลูก มะพร้าวสุก 25 ลูก

โมเดลคอนโวลูชันแนลทำนายผลตรวจจับเป็นภาพ ดังนั้นภาพที่มีทั้งลูกติดและสุกที่นำมาทดสอบจึงไม่สามารถนำมาประเมินประสิทธิภาพได้ แต่โมเดลฟาสเตอร์อาร์ซีเอ็นเอ็นทำนายผลตรวจจับเป็นลูก ผลการทดสอบทั้ง 2 โมเดลได้ผลดังตารางคอนฟิวชันเมทริกซ์ ในตารางที่ 2 ผลการประเมินค่าความถูกต้อง ความเที่ยง ความครบถ้วน และเอฟวันสกอร์ดังในตารางที่ 3

ตารางที่ 2 คอนฟิวชันเมทริกซ์ผลการตรวจจับ

		ผลทำนาย			
		CNN		Faster R-CNN	
		Y = 0	Y = 1	Y = 0	Y = 1
ค่าจริง	Y = 0	14	1	24	25
	Y = 1	39	27	14	335

ตารางที่ 3 ค่าความถูกต้อง ความเที่ยง ความครบถ้วน และเอฟวันสกอร์

	CNN		Faster R-CNN	
	ค่า	ร้อยละ	ค่า	ร้อยละ
ความถูกต้อง	0.5062	50.62	0.9020	90.2
ความเที่ยง	0.2642	26.42	0.6316	63.16
ความครบถ้วน	0.9333	93.33	0.4898	48.98
เอฟวันสกอร์	0.4118	41.18	0.5517	55.17

5. สรุป

งานวิจัยนี้นำเสนอการตรวจจับมั่งคุดที่อยู่บนต้นไม้ในสวนที่ได้จากภาพถ่ายด้วยโครงร่างตรวจจับการเรียนรู้เชิงลึกคอนโวลูชันแนลเปรียบเทียบกับฟาสเตอร์อาร์ชีเอ็นเอ็นประกอบด้วย 2 ขั้นตอน ขั้นตอนแรกขั้นตอนการฝึกสอนนำภาพถ่ายมั่งคุดจำนวน 320 ไปฝึกสอนโมเดลด้วยอัลกอริทึมคอนโวลูชันแนล จำนวน 10 รอบ และการเรียนรู้เชิงลึกฟาสเตอร์อาร์ชีเอ็นเอ็น จำนวน 20,000 รอบ จนได้โมเดลตรวจจับมั่งคุด ขั้นตอนสุดท้ายเป็นขั้นตอนการทดสอบนำโมเดลไปตรวจจับมั่งคุดแก่หรือสุกกับรูปภาพที่ถ่ายมาจากต้นไม้ในสวนที่มีทั้งลูกมั่งคุดใบและกิ่ง จำนวน 80 ภาพที่มีลูกดิบอย่างเดียว 62 ภาพ ลูกสุกอย่างเดียว 1 ภาพ มีทั้งดิบและสุก 17 ภาพ และทั้ง 80 ภาพมีลูกดิบจำนวน 332 ลูก ลูกแก่หรือสุก 24 ลูก ภาพมะปราง 20 ภาพ ที่มีลูกดิบอย่างเดียว 10 ภาพ ลูกสุกอย่างเดียว 9 ภาพ มีทั้งดิบและสุก 1 ภาพ และทั้ง 20 ภาพ มีมะปรางดิบ 17 ลูก มะปรางสุก 25 ลูก

ประเมินผลด้วยคอนฟิวชันเมตริกซ์ ได้ผลการประเมินดังนี้ แบบจำลองโครงข่ายประสาทคอนโวลูชันแนลสามารถจำแนกมั่งคุดแก่หรือสุก ได้ค่าความถูกต้องร้อยละ 50.62 ค่าความเที่ยงร้อยละ 26.42 ค่าความครบถ้วนร้อยละ 93.33 และ เอฟวันสกอร์ ร้อยละ 41.18 แบบจำลองฟาสเตอร์อาร์ชีเอ็นเอ็น สามารถจำแนกมั่งคุดแก่หรือสุก ได้ค่าความถูกต้องร้อยละ 90.22 ค่าความเที่ยงร้อยละ 63.16 ค่าความครบถ้วนร้อยละ 48.98 และ เอฟวันสกอร์ ร้อยละ 55.17

ข้อเสนอแนะในการวิจัย รูปภาพที่มีคุณสมบัติเกินขอบเขตที่โปรแกรมได้กำหนดไว้ จะทำให้ประสิทธิภาพในการแปลงค่าพารามิเตอร์ลดลง ทำให้เกิดความผิดพลาดแยกแยะรูปภาพ รวมถึงความละเอียดของภาพมีน้อยเกินกว่าจะใช้ในการคัดแยกคุณสมบัติของรูปภาพได้ รวมถึงรูปภาพวัตถุที่มีสีใกล้เคียงกับสีผิวมั่งคุดและมีเส้นโค้งกลมเหมือนผลมั่งคุด จะทำให้เกิดความผิดพลาดในการแยกแยะมั่งคุดในภาพได้ และความผิดพลาดที่อาจเกิดจากตำแหน่งของรูปภาพ เช่น ภาพที่ถ่ายมั่งคุดในระยะไกล และความผิดพลาดที่เกิดจากการส่องสว่าง และการสะท้อนของแสงที่มีในรูปภาพ ซึ่งจะส่งผลให้รูปภาพนั้นมีสีผิวมั่งคุดที่จางมากเกินไป หรือมีสีผิวที่เข้มมากเกินไป

สำหรับแนวทางการลดค่าความผิดพลาดที่เกิดขึ้น

สามารถดำเนินการได้สองวิธีการ ดังนี้

- 1) เพิ่มจำนวนรอบฝึกสอนโครงข่ายประสาทเทียม
- 2) เพิ่มข้อมูลรูปภาพในการฝึกสอนโครงข่ายประสาทเทียม

6. กิตติกรรมประกาศ

ขอขอบคุณ คุณวิญญู ศรีเพชร เจ้าของสวนศรีเพชร ตำบลหัวดง อำเภอลับแล จังหวัดอุตรดิตถ์ ที่ได้อนุญาตให้เข้าไปบันทึกภาพมั่งคุดภายในสวน และขอขอบคุณสำนักงานกองทุนส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกว.) ที่ได้สนับสนุนทุนในการดำเนินการวิจัย

7. เอกสารอ้างอิง

- [1] มติชน, “หุ่นยนต์” เกษตรกรในอนาคต. สืบค้นจาก : https://www.matichon.co.th/lifestyle/news_387076. สืบค้นเมื่อ 1 มิถุนายน 2562.
- [2] วิชาการดอทคอม, การพัฒนาเทคโนโลยีหุ่นยนต์เก็บสตรอว์เบอร์รี่. สืบค้นจาก: <http://www.vcharkarn.com/vnews/153453>. สืบค้นเมื่อ 2 มิถุนายน .
- [3] สำนักงานเศรษฐกิจการเกษตร, ข้อมูลเศรษฐกิจการเกษตร. สืบค้นจาก: <https://www.moac.go.th> สืบค้นเมื่อ 20 มกราคม 2563.
- [4] แนวหน้า, โลกธุรกิจ. สืบค้นจาก: <https://www.naewna.com/business/513651>. สืบค้นเมื่อ 1 กันยายน 2563.
- [5] L. M. Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, “Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection.” *IEEE International Conference on Control System, Computing and Engineering (ICCSC)*, Penang, Malaysia, pp. 242-246, 2017.
- [6] S. Bargoti, and J. Underwood. “Deep Fruit Detection in Orchards.” *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, pp. 5164-5172, 2017.
- [7] H. Mureşan, and M. Oltean. “Fruit recognition from images using deep learning.” *An International Scientific Journal of Sapientia Hungarian University of Transylvania*, Vol. 10, No. 1, pp. 26-42, June, 2018.

- [8] S. A. Sanchez, H. J. Romero, and A. D. Morales. "A review: Comparison of performance metrics of pretrained models for object detection using the Tensor Flow framework." *IOP Conference Series: Materials Science and Engineering, Cartagena, Colombia*, pp. 1-15, 2019
- [9] W. S. McCulloch, and W. Pitts, "A logical calculus of the ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115–137, 1943.
- [10] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. "Backpropagation Applied to Handwritten Zip Code Recognition." *Neural Computation*, Vol. 1, No. 4, pp. 541-551, December, 1989.
- [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient- Based Learning Applied to Document Recognition." in *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, November, 1998.
- [12] M. Subhi, and S. H. Ali, "A Deep Convolutional Neural Network for Food Detection and Recognition." *IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Sarawak, Malaysia, pp. 284-287, 2018.
- [13] M. M. Rahman, S. M. Islam, R. Sassi, and M. Aktaruzzaman. "Convolutional neural networks performance comparison for handwritten Bengali numerals recognition." *SN Applied Sciences*, November, 2019.
- [14] R. S. Ren, K. He, R. Girshick, and J. Sun, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. Available Online at <https://arxiv.org/abs/1506.01497>, accessed on 1 December, 2019.
- [15] Y. Zhou, T. Xu, W. Zheng, and H. Deng. "Classification and recognition approaches of tomato main organs based on DCNN." *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, Vol. 33, No. 15, pp. 219-226, August, 2017
- [16] L. Fu, Y. Feng, Y. Majeed, and X. Zhang. "Kiwifruit detection in field images using Faster R-CNN with ZFNet" *IFAC-PapersOnLine*, Vol. 51, Issue 17, pp. 45-50. 2018
- [17] X. Mai, H. Zhang, M. and Q. -. Meng, "Faster R-CNN with Classifier Fusion for Small Fruit Detection." *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia, pp. 7166-7172, 2018.
- [18] S. Bargoti, and J. Underwood, "Deep Fruit Detection in Orchards." *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, pp. 3626-3633, 2017.