

โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกสำหรับการจำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ

Deep Convolutional Neural Networks for Plant Recognition in the Natural Environment

จักรินทร์ สุกุณแสน¹, โอลาริก สุรินตะ²

Jakkarin Sanuksan¹, Olarik Surinta²

Received: 15 May 2018 ; Revised : 20 August 2018 ; Accepted: 5 November 2018

บทคัดย่อ

งานวิจัยฉบับนี้ได้ศึกษาโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก (Deep Convolutional Neural Network: Deep CNN) เพื่อใช้จำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ โดยมีจุดประสงค์เพื่อเปรียบเทียบ โครงข่ายประสาทเทียมแบบคอนโวลูชัน จำนวน 4 โครงสร้าง ประกอบด้วย LeNet-5, AlexNet, GoogLeNet และ VGGNet ข้อมูลพรรณไม้ที่นำมาใช้ในการทดสอบ มีจำนวนทั้งสิ้น 3 ชุดข้อมูลคือ PNE, 102 Flower และ Folio ทั้งนี้ชุดข้อมูล PNE และ 102 Flower เป็นรูปภาพที่อยู่ในสิ่งแวดล้อมทางธรรมชาติทำให้มีพื้นหลังที่ซับซ้อน สำหรับข้อมูลชุด Folio เป็นรูปภาพใบไม้ที่ถ่ายในห้องทดลองโดยกำหนดให้พื้นหลังของภาพเป็นสีขาว จากผลเปรียบเทียบระหว่างโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกโดยใช้โครงสร้าง GoogLeNet และ VGGNet พบว่าโครงสร้างแบบ GoogLeNet มีประสิทธิภาพสูงที่สุดในชุดข้อมูล PNE และ 102 Flower และยังใช้เวลาในการเรียนรู้ที่เร็วกว่าเมื่อเทียบกับโครงสร้างแบบ VGGNet โดยใช้จำนวนรอบในการเรียนรู้ 10,000 รอบ แต่ทั้งนี้โครงสร้างแบบ VGGNet มีอัตราความถูกต้องสูงที่สุดในชุดข้อมูล Folio ซึ่งเป็นรูปภาพที่ถ่ายในห้องทดลองที่มีพื้นหลังเป็นสีขาว และใช้จำนวนรอบในการเรียนรู้เพียง 1,000 รอบ จึงสรุปได้ว่าหากต้องการที่จะสร้างโมเดลเพื่อนำไปใช้ในการจำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติควรจะใช้โครงสร้างแบบ GoogLeNet

คำสำคัญ: การรู้จำพรรณไม้ การเรียนรู้เชิงลึก โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก โครงสร้างแบบ AlexNet โครงสร้างแบบ GoogLeNet โครงสร้างแบบ VGGNet

Abstract

This paper examines a deep convolutional neural network (Deep CNN) for plant recognition in the natural environment. The primary objective was to compare 4 CNN architectures including LeNet-5, AlexNet, GoogLeNet, and VGGNet on three plant datasets; PNE, 102 Flower, and Folio. The images in the PNE and 102 Flower dataset include a complicated background because they were taken in a natural environment. On the other hand, the images in the Folio dataset are only leaf images that were taken in a laboratory environment using a white background. The comparison of deep CNN using GoogLeNet and VGGNet Architecture show that GoogLeNet outperformed while working on the PNE and 102 Flower dataset when using a training time with iterations of 10,000 epochs. GoogLeNet

¹ นิสิตปริญญาโท ห้องปฏิบัติการมัลติเอเจนต์ ระบบอัจฉริยะ และการจำลองสถานการณ์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150

² อาจารย์ ห้องปฏิบัติการมัลติเอเจนต์ ระบบอัจฉริยะ และการจำลองสถานการณ์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150

¹ Master Student, Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand. E-mail: jakkarin.san@msu.ac.th

² Lecturer, Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand. E-mail: olarik.s@msu.ac.th

also faster than the VGGNet architecture. However, the experiment showed that the VGGNet architecture outperforms the other CNN architectures on the Folio dataset and used only 1,000 epochs for training. In our experiment, we can create a model from the deep CNN using GoogleNet architecture, and this is because it showed better results with the plant images that were taken in the natural environment.

Keywords: Plant Recognition, Deep Learning, Deep Convolutional Neural Network, AlexNet Architecture, GoogLeNet Architecture, VGGNet Architecture

บทนำ

โครงข่ายประสาทเทียม (Neural Network) เป็นโครงข่ายที่จำลองหลักการการทำงานของโครงข่ายประสาทที่อยู่ในสมองของมนุษย์ โดยมีการแบ่งลักษณะของโครงสร้างการทำงานออกเป็นชั้น (Layer) การทำงานของโครงข่ายประสาทเทียมประกอบด้วย ชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นแสดงผลลัพธ์ (Output Layer) ซึ่งจำนวนหน่วย (Unit หรือ Node) ของชั้นแสดงผลลัพธ์ขึ้นอยู่กับจำนวนหมวดหมู่ (Category) ที่ทำการจำแนกประเภท (Classification) เช่น หากต้องการจำแนกประเภทข้อมูลตัวเลขอารบิก จำนวนของหมวดหมู่จะมีค่าเป็น 10 หมวด ซึ่งหมายถึงค่าความน่าจะเป็น (Probability) ที่จะเป็นคำตอบของตัวเลข 0-9 ดังนั้น หากค่าความน่าจะเป็นที่สูงที่สุดอยู่ในหน่วยใดก็ถือว่าเป็นคำตอบที่ได้จากการคาดการณ์หรือการพยากรณ์ (Predict)

นักวิจัยได้ปรับปรุงโครงข่ายให้มีจำนวนชั้นซ่อนเพิ่มขึ้น เนื่องจากจะช่วยในการคำนวณให้มีประสิทธิภาพมากขึ้น แต่ในทางกลับกัน จะส่งผลให้กระบวนการในการเรียนรู้ยาวนานขึ้น ซึ่งวิธีนี้เรียกว่า การเรียนรู้เชิงลึก (Deep Learning)¹⁻³ โดย LeCun et al.^{1,2} ได้นำเสนอโครงข่ายที่มีลักษณะการทำงานแบบคอนโวลูชัน (Convolutional Networks) โครงข่ายลักษณะนี้จึงถูกเรียกว่า โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) โครงสร้าง (Architecture) ถูกกำหนดไว้ 5 ชั้น โครงสร้างนี้มีชื่อเรียกว่า LeNet-5 โดยโครงข่ายประสาทเทียมแบบคอนโวลูชัน จะตัดขั้นตอนในส่วนของการหาคุณลักษณะพิเศษ (Feature Extraction) ทำให้นักวิจัยไม่จำเป็นต้องใช้วิธีการหาคุณลักษณะพิเศษแบบเดิม เช่นการใช้รูปแบบ (Pattern) การใช้รูปร่าง (Shape) และการใช้สี (Colour) แต่ทั้งนี้จะใช้ขั้นตอนในการคอนโวลูชันเป็นขั้นตอนหนึ่งในการคำนวณหาคุณลักษณะพิเศษ การทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชันจะทำงานควบคู่กับชั้นพหุคูณเพื่อช่วยลดขนาดของเมทริกซ์ (Matrix) ที่ใช้ในการคำนวณลง ดังนั้น การเรียนรู้เชิงลึกที่ใช้โครงสร้างแบบ CNN จึงเหมาะสมกับข้อมูลประเภทรูปภาพ ทำให้นักวิจัยสามารถนำการเรียนรู้เชิงลึกไปประยุกต์ใช้ในงานด้านการค้นหา

(Detection)⁴ we propose a Spatially Constrained Convolutional Neural Network (SC-CNN การจำแนกประเภท⁵ contribute to a better understanding of the relationship between environmental factors and healthy crops, reduce the labor costs for farmers and increase the operation speed and accuracy. Implementing machine learning methods such as deep neural networks on agricultural data has gained immense attention in recent years. One of the most important problems is automatic classification of plant species based on their types. Automatic plant type identification process could offer a great help for application of pesticides, fertilization and harvesting of different species on-time in order to improve the production processes of food and drug industries. In this paper, we propose a Convolutional Neural Network (CNN และการจัดกลุ่ม (Clustering)⁶ รูปภาพ

จากงานวิจัยของ Liu et al.⁷ ได้นำเสนอวิธีการจำแนกประเภทดอกไม้ (Flower Classification) โดยใช้วิธี CNN ที่ปรับเปลี่ยนโครงสร้างให้มีจำนวน 8 ชั้น (Layer) ซึ่งมีชั้นคอนโวลูชัน (Convolutional Layer) จำนวน 5 ชั้น และ ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) จำนวน 3 ชั้น และใช้ฟังก์ชันซอฟต์แวร์แมกซ์ (Softmax Function) ในการคำนวณหาผลลัพธ์ โดยได้ทดสอบกับชุดข้อมูลดอกไม้ที่มีจำนวนทั้งสิ้น 63,442 รูปภาพ แบ่งออกเป็นดอกไม้ 79 สายพันธุ์ และยังได้ทดสอบกับข้อมูลชุด Oxford 103 Flower Species ที่มีจำนวนทั้งสิ้น 103 สายพันธุ์ ประกอบด้วยรูปภาพดอกไม้ทั้งสิ้น 8,189 รูปภาพ แต่ละสายพันธุ์จะมีรูปภาพของดอกไม้ตั้งแต่ 40 ถึง 200 รูปภาพ จากการทดลองพบว่าวิธีการ CNN โดยกำหนดให้มีจำนวนชั้นทั้งสิ้น 8 ชั้น มีอัตราความถูกต้อง 84.02% จากนั้นจึงได้ทดสอบกับข้อมูลชุด Large-scale 79 Flower Species ซึ่งมีดอกไม้ทั้งสิ้น 79 สายพันธุ์ มีจำนวนรูปภาพทั้งสิ้น 52,775 รูปภาพ ในการทดสอบได้แบ่งรูปภาพจำนวน 47,500 เป็นข้อมูลชุดเรียนรู้ และ 5,275 รูปภาพเป็นข้อมูลชุดทดสอบ พบว่าวิธี CNN ที่นำเสนอมีอัตราความถูกต้อง 76.54% เมื่อเทียบ

กับวิธี CNN แบบ LeNet-5 ที่มีอัตราความถูกต้อง 70.12% ซึ่งมีความแม่นยำเพิ่มขึ้นถึง 6%

Pawara *et al.*⁸ ได้เปรียบเทียบวิธีการหาคุณลักษณะพิเศษแบบเฉพาะพื้นที่ (Local Descriptor) และวิธี Bag of Word (BOW) กับวิธี Deep CNNs สำหรับจำแนกพรรณไม้ (Plant Recognition) โดยชุดข้อมูลที่ใช้ในการทดสอบประกอบด้วย AgrilPlant (10 สายพันธุ์ 3,000 รูปภาพ) LeafSnap (184 สายพันธุ์ 7,719 รูปภาพ) และ Folio (32 สายพันธุ์ 640 รูปภาพ) โดยรูปภาพที่ใช้ในการทดลองถูกปรับให้มีขนาด 256x256 พิกเซล วิธี Cross Validation ถูกนำมาใช้ในการแบ่งชุดข้อมูล โดยกำหนดให้ k มีค่าเท่ากับ 5 เพื่อแบ่งข้อมูลเป็น 80% สำหรับเรียนรู้ และ 20% สำหรับทดสอบ โดยทดสอบด้วยวิธี Histogram of Oriented Gradients (HOG) และ HOG-BOW เพื่อสร้างคุณลักษณะพิเศษ จากนั้นจึงนำคุณลักษณะพิเศษที่ได้ไปทำการเรียนรู้ด้วยวิธีการคำนวณเพื่อนบ้านใกล้ที่สุด k ตำแหน่ง (k-Nearest Neighbors: KNN), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และเพอร์เซปตรอนแบบหลายชั้น (Multi Layer Perceptron: MLP) จากการทดลองวิธี HOG-BOW ร่วมกับ MLP ให้ประสิทธิภาพสูงสุดในข้อมูลชุด LeafSnap และวิธี HOG-BOW ร่วมกับ SVM ให้ประสิทธิภาพสูงสุดในข้อมูลชุด AgrilPlant และ Folio นอกจากนี้ยังได้ทดสอบวิธี Deep CNNs โดยใช้โครงสร้างแบบ AlexNet และ GoogLeNet จากการทดลองพบว่า โครงสร้างแบบ AlexNet มีอัตราความถูกต้องสูงสุด 97.67% สำหรับข้อมูลชุด Folio และ GoogLeNet มีอัตราความถูกต้องสูงสุดในข้อมูลชุด AgrilPlant และ LeafSnap โดยมีความถูกต้องที่ 98.33% และ 97.66% ตามลำดับ

Reyes *et al.*⁹ นำเสนอวิธี Deep CNNs เพื่อจำแนกประเภทของพืช โดยใช้ชุดข้อมูล LifeCLEF 2015 ในการทดสอบ งานวิจัยนี้ใช้วิธีการถ่ายโอนความรู้ (Transfer Learning) ซึ่งใช้ชุดข้อมูล ILSVRC2012 ซึ่งประกอบด้วยรูปภาพ 1.2 ล้านรูป และมีหมวดหมู่มากถึง 1,000 หมวดในการเรียนรู้เพื่อสร้างโมเดล จากนั้นจึงนำค่าน้ำหนัก (Weight) ที่ได้มาเรียนรู้อีกครั้งกับข้อมูล LifeCLEF 2015 วิธีนี้ทำให้ลดระยะเวลาในการเรียนรู้ ชุดข้อมูล LifeCLEF 2015 ที่ใช้ในการเรียนรู้มีจำนวน 91,759 รูปภาพ โดยเป็นพันธุ์ไม้จำนวน 1,000 สายพันธุ์ ภายในชุดข้อมูลได้แบ่งแยกรูปภาพของพืช โดยแบ่งออกเป็น 7 มุมมอง ประกอบด้วย ภาพถ่ายกิ่ง ดอก ก้านดอก ผล ใบไม้ กลุ่มของใบไม้ และภาพรวมของต้นไม้ โดยได้นำมุมมองทั้ง 7 มุมไปทดสอบ ผลการทดสอบพบว่ากลุ่มภาพถ่ายดอกมีอัตราความแม่นยำสูงที่สุดประมาณ 65% โดยรวมแล้วมีความแม่นยำเฉลี่ยเท่ากับ 48.6%

ในงานวิจัย^{10,11} โครงข่ายประสาทเทียมเชิงลึก (Deep Neural Networks) ถูกนำมาใช้เพื่อจำแนกพืชที่เป็นโรค (Plant Disease Recognition) โดย Sladojevic *et al.*¹⁰ ได้นำรูปภาพใบไม้ (Leaf) มาเพื่อทดลอง โดยมีรูปตัวอย่างใบไม้ที่เป็นโรคจำนวน 13 ประเภท และรวมกับใบไม้ที่อยู่ในสภาพปกติอีก 2 ประเภท รวมเป็น 15 ประเภท ในการทดลองใช้เทคนิคในการเพิ่มข้อมูลในการเรียนรู้ (Augmented Data) ด้วยวิธีการประมวลผลภาพ (Image Processing) ประกอบด้วย การนำภาพไปบิดให้ผิดรูป (Distortion) และการหมุนภาพ (Rotation) เป็นต้น ข้อมูลที่ใช้ในการทดสอบมีจำนวนทั้งสิ้น 4,483 รูปภาพ และเมื่อนำไปผ่านวิธี Augmented จะได้รูปภาพเพิ่มขึ้นเป็น 30,880 รูปภาพ จากการทดลองพบว่า มีอัตราความถูกต้อง 96.3%

งานวิจัย¹¹ วิธีการเรียนรู้เชิงลึกถูกนำมาช่วยในการวินิจฉัยโรคพืชจากภาพถ่าย ซึ่งในการทดลองใช้ชุดข้อมูล PlantVillage ที่มีจำนวน 54,306 รูปภาพจากพืชทั้งสิ้น 14 ชนิด โดยระบุชนิดของโรคไว้ทั้งสิ้น 26 โรค โดยรูปภาพได้ถูกปรับให้มีขนาด 256x256 พิกเซล และใช้วิธีการเรียนรู้เชิงลึกที่มีโครงสร้างแบบ AlexNet และ GoogLeNet โดยทำการทดสอบในสองรูปแบบประกอบด้วย เริ่มต้นเรียนรู้จากศูนย์ (Training from Scratch) และเรียนรู้แบบวิธีการถ่ายโอนความรู้ในการทดสอบประสิทธิภาพของการเรียนรู้เชิงลึกใช้วิธี Cross Validation เพื่อแบ่งข้อมูลออกเป็นข้อมูลชุดเรียนรู้ 80, 60, 50, 40 และ 20% ตามลำดับ จากการทดลองพบว่าเมื่อแบ่งข้อมูลออกเป็นชุดเรียนรู้จำนวน 80% และชุดทดสอบจำนวน 20% โครงสร้างแบบ GoogLeNet ที่ใช้การเรียนรู้แบบ วิธีการถ่ายโอนความรู้มีอัตราความถูกต้องสูงที่สุดคือ 99.34%

งานวิจัยของ Sun *et al.*¹² นำเสนอวิธีการจำแนกพรรณพืช (Plant Identification) ที่เป็นรูปถ่ายที่อยู่ในสภาวะแวดล้อมทางธรรมชาติโดยเป็นพันธุ์ไม้โดยข้อมูลชุด BJFU100 ประกอบไปด้วยพันธุ์ไม้จำนวน 100 ชนิด ซึ่งเป็นรูปภาพดอกไม้ทั้งสิ้น 10,000 รูปภาพ โดยใช้วิธี CNN ที่มีโครงสร้างแบบ ResNet โดยทดลองจำนวนชั้นตั้งแต่ 18, 26, 34 และ 50 ชั้น ในการทดลองได้แบ่งข้อมูลชุดเรียนรู้จำนวน 80% และชุดทดสอบ 20% จากการทดลองปรากฏว่าโครงสร้างแบบ ResNet ที่มีจำนวน 26 ชั้น มีความถูกต้อง 91.78% และเมื่อนำไปทดสอบกับชุดข้อมูล Flavia มีอัตราความถูกต้อง 99.65%

งานวิจัยที่เกี่ยวข้องที่ได้กล่าวมาข้างต้น สามารถนำวิธีการเรียนรู้เชิงลึกไปใช้เพื่อจำแนกรูปภาพพรรณไม้ ทั้งที่จำแนกจากดอกไม้ (Flower) ใบไม้ (Leaf) และภาพถ่ายดอกไม้ที่อยู่ในสภาวะแวดล้อมทางธรรมชาติ และยังสามารถนำไป

ช่วยในการวินิจฉัยโรคที่เกิดขึ้นกับพรรณไม้ได้

งานวิจัยฉบับนี้ได้ศึกษาโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกซึ่งเปรียบเทียบกับโครงสร้างที่มีความแตกต่างกันจำนวน 3 รูปแบบ ประกอบด้วย AlexNet, GoogLeNet และ VGGNet^{2,13-15} เพื่อนำมาใช้สำหรับจำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ (Plant Classification in Natural Environment) อีกทั้งยังได้นำเสนอชุดข้อมูล PNE ที่เก็บรวบรวมพรรณไม้ทั้งสิ้น 10 ชนิด โดยทุกรูปเป็นภาพถ่ายดอกไม้ที่อยู่ในสภาวะแวดล้อมทางธรรมชาติซึ่งยากต่อการจำแนก เนื่องจากในรูปภาพอาจมีทั้งรูปดอกไม้ ใบไม้ และอาจรวมถึงรูปองค์ประกอบอื่นเช่น ก้าน หรือพื้นหลังที่เป็นวัตถุอื่น เป็นต้น

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

โดยทั่วไปแล้ว โครงสร้างของโครงข่ายประสาทเทียมประกอบด้วย ชั้นนำเข้า (Input Layer), ชั้นซ่อน (Hidden Layer) และชั้นแสดงผลลัพธ์ (Output Layer) หรือเรียกว่า Multi-Layer Perceptron (MLP) แต่งานวิจัยที่นำเสนอโดย LeCun et al.² ได้เพิ่มการคำนวณที่มีลักษณะแบบคอนโวลูชัน (Convolutional) เข้าไปในโครงข่าย จึงเรียกว่าโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) โดยโครงข่าย CNN ประกอบด้วยชั้นคอนโวลูชันชั้นพูลลิ่ง และชั้นเชื่อมโยงแบบสมบรูณ์ ซึ่งชั้นเชื่อมโยงแบบสมบรูณ์ ก็คือชั้นซ่อน และชั้นแสดงผลลัพธ์ที่ปรากฏในโครงข่ายประสาทเทียม

ทั้งนี้ โครงข่าย CNN สามารถทำได้ทั้งการสกัดคุณลักษณะพิเศษของรูปภาพ และการจำแนกประเภท² ทำให้เป็นจุดเด่นของการเรียนรู้ด้วยของโครงข่าย CNN ซึ่งมีความแตกต่างกับวิธีการเรียนรู้เครื่องจักร (Machine Learning) ทั่วไปที่จะทำหน้าที่เพียงจำแนกประเภทข้อมูลหรือจัดกลุ่มข้อมูลเพียงเท่านั้น

ชั้นคอนโวลูชัน (Convolutional Layer)

คอนโวลูชันเป็นการคำนวณเพื่อหาผลลัพธ์ของ นิวรอน (Neuron) ที่เชื่อมต่อ (Connected) มาจากพื้นที่ส่วนย่อย (Local Region) ของรูปภาพ โดยรูปภาพจะถูกกำหนดให้เป็นสี่เหลี่ยมจัตุรัสขนาด $W \times W \times M$ โดยที่ W คือความสูงและความกว้างของรูปภาพ และ M คือมิติของรูปภาพ เช่น ภาพสี RGB จะถูกกำหนดให้ $M = 3$

การคำนวณในชั้นคอนโวลูชัน พื้นที่ส่วนย่อยจะถูกนำมาคำนวณแบบ dot product กับเคอร์เนล (Kernel) ที่มี

ขนาด $H \times H \times K$ โดยเคอร์เนลต้องมีขนาดเล็กกว่าขนาดของรูปภาพ และ K คือจำนวนของเคอร์เนลที่ต้องการคำนวณผลลัพธ์ที่ได้จากการคอนโวลูชัน เรียกว่า Feature Map การคำนวณ คอนโวลูชัน เพื่อหา Feature Map แสดงดัง Equation (1)

$$u_{ijm} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p,j+q,k} h_{pqkm} + b_{ijm} \tag{1}$$

โดยที่ x คือ รูปภาพที่ใช้ในการคำนวณ
 h คือ เคอร์เนลขนาด $H \times H$
 b คือ Bias

Figure 1 แสดงให้เห็นถึงวิธีการคอนโวลูชัน ระหว่าง Local Region และเคอร์เนล

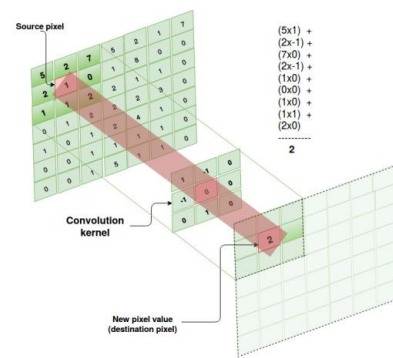


Figure 1 The illustration of the convolutional Layer. The dot production calculation between sub-region of the image and convolution kernel.

ชั้นพูลลิ่ง (Pooling Layer)

โดยทั่วไปแล้วชั้นพูลลิ่ง เป็นชั้นที่คั่นกลางระหว่างชั้นคอนโวลูชัน โดยมีไว้เพื่อลดขนาด (Downsample) ของ Feature Map ให้เล็กลง โดยใช้ฟังก์ชันค่าเฉลี่ย ฟังก์ชันค่าต่ำสุด และฟังก์ชันค่าสูงสุดในการคำนวณ หากเลือกใช้ฟังก์ชันค่าสูงสุดในการคำนวณ ดังนั้น ชั้นนี้เรียกว่า Max Pooling นั้นหมายถึง Feature Map จะถูกแบ่งออกเป็น Local Region จากนั้น Local Region จะถูกแบ่งออกเป็น $p \times p$ ส่วน p โดยจะกำหนดให้อยู่ในช่วงระหว่าง 2 และ 5 ดังนั้น ค่าที่มากที่สุดในแต่ละ Pool จะถูกเลือกเพื่อนำมาเป็นตัวแทน จากนั้นระบบจะเลื่อน (Stride) ไปยัง Local Region ต่อไปโดยจะเลื่อนไปเรื่อย ๆ เช่น ครึ่งละ 2 พิกเซล ไปจนกระทั่งจุดสุดท้ายของ Feature Map การคำนวณ Max Pooling แสดงดัง Figure 2

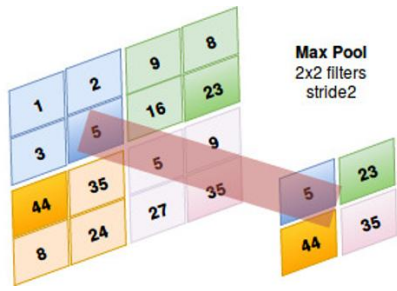


Figure 2 The illustration of max pooling with 2 x 2 kernel and stride 2.

ชั้นเชื่อมต่อแบบสมบูรณ์ (Fully-Connected Layer)

ชั้นเชื่อมต่อแบบสมบูรณ์ เป็นชั้นที่ทำหน้าที่ในการจำแนกประเภทของวัตถุ (Object) ซึ่งทุก Neuron ที่อยู่ในชั้นเชื่อมต่อแบบสมบูรณ์จะถูกเชื่อมโยงกับชั้นคอนโวลูชัน และชั้นพูลลิ่ง อย่างสมบูรณ์ ซึ่งก็คือโครงสร้างของนิวรอลเน็ตเวิร์ก (Neural Network) โดยจำนวนผลลัพธ์คือจำนวนของกลุ่มที่ต้องการจำแนก ซึ่งคำนวณด้วยการหาค่าความน่าจะเป็น (Probability) ด้วยฟังก์ชันซอฟต์แมกซ์ แสดงดังสมการ 2

$$f_i(x) = \frac{\exp(x_i)}{\sum_j^n \exp(x_j)} \quad (2)$$

โดยที่คือผลลัพธ์ของแต่ละโหนด

โครงสร้างแบบ LeNet-5 (LeNet-5 Architecture)

ปี ค.ศ. 1989 นักวิจัย LeCun *et al.*¹ ได้นำเสนอโครงข่าย CNN โดยมีจุดประสงค์เพื่อการรู้จำตัวเลขรหัสไปรษณีย์ ในการทดลอง รูปภาพของตัวเลขจะถูกปรับให้มีขนาด 16x16 พิกเซล (256 Input Unit) โครงสร้างถูกกำหนดให้มีชั้นซ่อนจำนวน 3 ชั้น (H1-H3) โดยชั้น H1 ประกอบด้วยเคอร์เนลขนาด 5x5 จำนวน 12 เคอร์เนล เมื่อผ่านการคำนวณจำนวนของ Hidden Unit จะมีขนาด 768 (12x64) และในชั้น H2 มีจำนวนทั้งสิ้น 12 เคอร์เนล แต่ละเคอร์เนล มีขนาด 5x5x8 และเมื่อผ่านการคำนวณชั้น H2 จะมีชั้นซ่อนขนาด 192 (12x16) โหนด จากนั้นโหนดในชั้นซ่อน ทั้งหมดจะถูกส่งไปยังชั้น H3 ซึ่งเป็นชั้นของการเชื่อมต่อแบบสมบูรณ์ที่กำหนดให้ชั้นซ่อนมีจำนวน 30 โหนด และสุดท้ายส่งออกไปยังชั้นแสดงผลลัพธ์ที่กำหนดให้เป็น 10 โหนด เพื่อให้สอดคล้องกับตัวเลข 0-9 โดยในชั้นเชื่อมต่อแบบสมบูรณ์เป็นการทำงานแบบโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back-propagation Neural Network)

ปี ค.ศ. 1998 ในงานวิจัย² ได้นำเสนอโครงข่าย CNN ที่กำหนดโครงสร้างแบบ LeNet-5 โดยโครงสร้างประกอบไปด้วยชั้นคอนโวลูชันจำนวน 2 ชั้น และชั้นพูลลิ่ง จำนวน 2 ชั้น และชั้นสุดท้ายคือชั้นเชื่อมต่อแบบสมบูรณ์ เพื่อใช้สำหรับการจำแนกวัตถุ ดังนั้น โครงสร้าง LeNet-5² จึงมีความแตกต่างกับโครงสร้างที่นำเสนอในงานวิจัย¹ คือการเพิ่มชั้นพูลลิ่งลงไปเพื่อเป็นการลดขนาดของ Feature Map โครงสร้างแบบ LeNet-5 แสดงดัง Figure 3

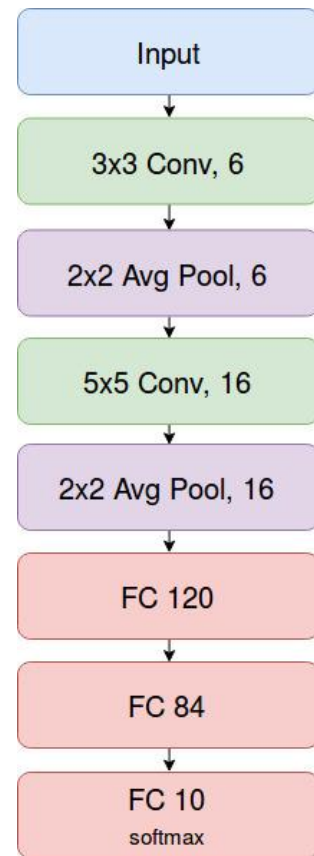


Figure 3 The architecture of LeNet-5.

โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก (Architecture of the Deep Convolutional Neural Networks)

โครงสร้างของโครงข่ายแบบ CNN สามารถเพิ่มจำนวนของชั้นคอนโวลูชัน ได้ไม่จำกัดจึงทำให้เกิดโครงสร้างลักษณะใหม่เกิดขึ้น เช่น AlexNet, GoogLeNet และ VGGNet เป็นต้น

โครงสร้างแบบ AlexNet (AlexNet Architecture)

ปี ค.ศ. 2012 นักวิจัย Krizhevsky *et al.*¹³ นำเสนอ Deep CNN ที่มีจำนวนทั้งสิ้น 8 ชั้น โดยโครงสร้างนี้เรียกว่า AlexNet ประกอบไปด้วยชั้นคอนโวลูชัน และชั้นเชื่อมต่อแบบ

สมบูรณ์ จำนวน 5 และ 3 ชั้น ตามลำดับ โดยผลลัพธ์ของชั้นเชื่อมโยงแบบสมบูรณ์ ถูกกำหนดให้มีจำนวน 100 โหนด เนื่องจากในงานวิจัยได้ใช้ชุดข้อมูล ImageNet LSVRC-2010 ในการทดสอบ ข้อมูลมีจำนวนทั้งสิ้น 1.2 ล้านรูปภาพ และแบ่งออกเป็น 1,000 กลุ่ม ในขั้นสุดท้ายของชั้นเชื่อมโยงแบบสมบูรณ์ ใช้วิธีซอร์ฟแม็กซ์เพื่อคำนวณหาผลลัพธ์

รูปภาพที่ใช้ในการทดลองจะถูกเปลี่ยนให้มีขนาด 224x224x3 พิกเซล ดังนั้น ชั้นคอนโวลูชัน (Conv{1}) ในชั้นแรกถูกกำหนดให้มี 96 เคอร์เนลโดยที่แต่ละเคอร์เนลมีขนาด 11x11x3 ในการคำนวณจะเลื่อน (Stride) เคอร์เนลครั้งละ 4 พิกเซล ผลลัพธ์จากการคอนโวลูชันจะมีขนาด 55x55x96 ชั้นที่สองของ ชั้นคอนโวลูชัน (Conv{2}) ถูกกำหนดให้มี 256 เคอร์เนลโดยที่แต่ละเคอร์เนลมีขนาด 5x5x48 ผลลัพธ์จากการคอนโวลูชันจะมีขนาด 27x27x256หลังจากการคำนวณ Conv{1, 2} ข้อมูลจะถูกส่งไปชั้นพูลลิ่งเพื่อทำการ Normalized และหาค่าสูงสุดในแต่ละ Pool (Max Pooling) โดยที่แต่ละ Pool จะถูกแบ่งเป็น 2x2 ส่วน

ในคอนโวลูชันชั้นที่สาม จำนวนของเคอร์เนลถูกกำหนดให้มี 384 เคอร์เนล โดยที่แต่ละเคอร์เนลมีขนาด 3x3x256 ผลลัพธ์จากการคอนโวลูชันจะมีขนาด 13x13x256 โดยชั้นที่สี่ ถูกกำหนดให้มี 384 เคอร์เนล โดยที่แต่ละเคอร์เนล มีขนาด 3x3x192 ผลลัพธ์จากการคอนโวลูชัน จะมีขนาด 13x13x384 ชั้นที่ห้า มีจำนวน 256 เคอร์เนล โดยแต่ละเคอร์เนลมีขนาด 3x3x192 สุดท้ายผลลัพธ์จากการคอนโวลูชัน มีขนาด 13x13x256 โดยที่ Conv{1-5} ใช้ฟังก์ชัน Rectified Linear Units (ReLU) เป็นฟังก์ชันกระตุ้น (Activation Function)

สุดท้ายแล้ว ชั้นเชื่อมโยงแบบสมบูรณ์ถูกแบ่งออกเป็น 2 ชั้น และกำหนดให้มีจำนวนชั้นละ 4,096 โหนด (Node) โครงสร้าง AlexNet แสดงดัง Figure 4

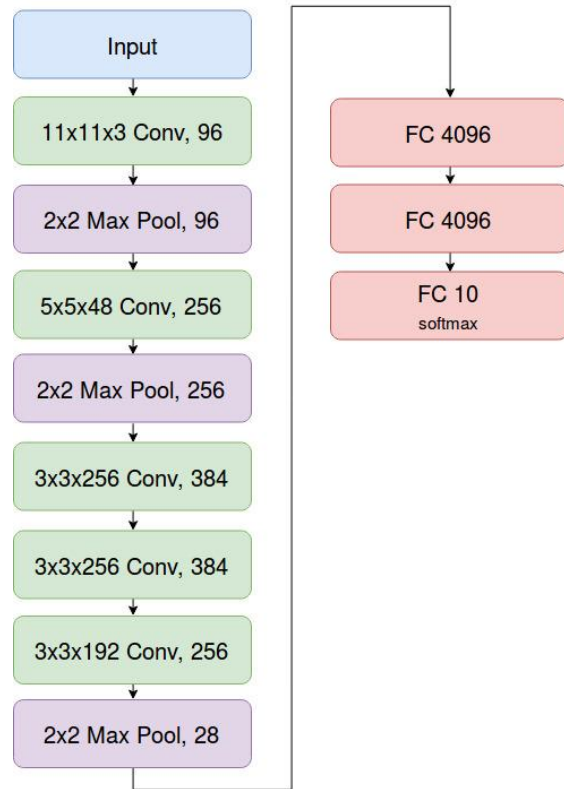


Figure 4 The AlexNet architecture.

โครงสร้างแบบ GoogLeNet (GoogLeNet Architecture)

โครงสร้างแบบ GoogLeNet¹⁴ ถูกนำเสนอในปี ค.ศ. 2014 โดยใช้โครงสร้าง Inception (Inception Architecture) เพื่อให้ CNN มีความลึกและกว้าง อีกทั้งยังไม่ทำให้การประมวลผลช้าลง ในแต่ละ Inception ประกอบไปด้วยการคำนวณคอนโวลูชันที่ใช้เคอร์เนล ขนาด 1x1, 3x3 และ 5x5 ซึ่งเป็นการประมวลผลแบบขนาน (Parallel) อีกทั้งยังทำให้จำนวนมิติของข้อมูลลดลง (Dimension Reduction) (แสดงดัง Figure 5) โครงสร้างแบบ GoogLeNet ถูกออกแบบให้มีโครงสร้างแบบ Inception จำนวน 9 ชั้น จากจำนวนทั้งสิ้น 22 ชั้น และใช้ฟังก์ชัน ReLU เพื่อเป็นฟังก์ชันกระตุ้น และชั้นเชื่อมโยงแบบสมบูรณ์กำหนดให้มีจำนวน 1024 โหนด ก่อนจะส่งต่อไปยังชั้นแสดงผล โครงสร้างแบบ GoogLeNet แสดงดัง Figure 6

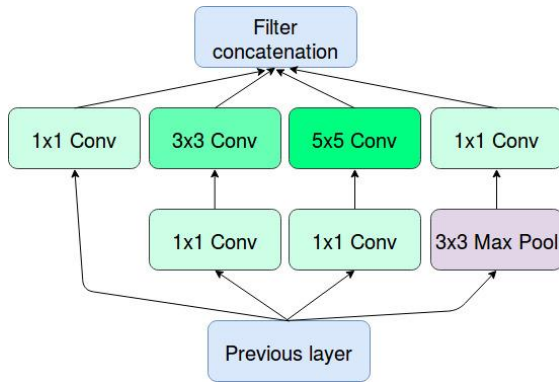


Figure 5 Inception Structure

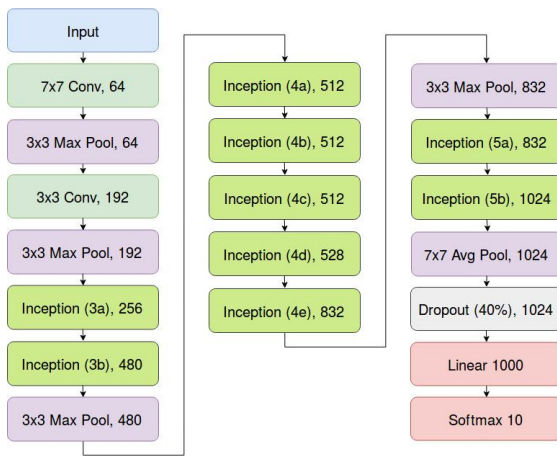


Figure 6 The googLeNet architecture.

โครงสร้างแบบ VGGNet (VGGNet Architecture)

ในปี ค.ศ. 2015 งานวิจัย¹⁵ ได้นำเสนอโครงสร้างแบบ VGGNet ซึ่งเป็นโครงสร้างแบบ Deep CNN เนื่องจากมีความลึกของชั้นจำนวน 16 โดยใช้ Kernel ขนาด 3x3 ในการ Convolution

โครงสร้างของ VGGNet-16 ประกอบด้วย ชั้นคอนโวลูชัน (Conv) จำนวน 5 กลุ่ม โดย Conv กลุ่มที่ 1 หรือ Conv1 ประกอบด้วย Conv1 {1, 2} และมีจำนวน เคอร์เนล ชั้นละ 64 เคอร์เนล ใน Conv2 ประกอบด้วย Conv2{1, 2} มีจำนวนเคอร์เนลชั้นละ 128 เคอร์เนล ใน Conv3 ประกอบด้วย Conv3 {1, 2, 3} และมีจำนวน Kernel ชั้นละ 256 เคอร์เนล ใน Conv4 ประกอบด้วย Conv4{1, 2, 3} และมีจำนวน เคอร์เนล ชั้นละ 512 เคอร์เนล และชั้น Conv5 ประกอบด้วย Conv5{1, 2, 3} ในแต่ละชั้นมีจำนวน เคอร์เนล ชั้นละ 512 เคอร์เนล โดยที่ใน Conv1-5 จะตามด้วยชั้นแมกซ์พูลลิ่ง (Max Pooling) จากนั้นข้อมูลจะส่งไปยัง ชั้นเชื่อมโยงแบบสมบูรณ์ ที่มีจำนวนโหนด 4,096 และ 4,096 โหนด และในชั้นสุดท้ายของชั้นเชื่อมโยงแบบสมบูรณ์ ถูกกำหนดให้มีจำนวน 1,000 โหนด โดยคำนวณหา

ผลลัพธ์ด้วยวิธีซอฟต์แวร์ โครงสร้างแบบ VGGNet-16 แสดงดัง Figure 7

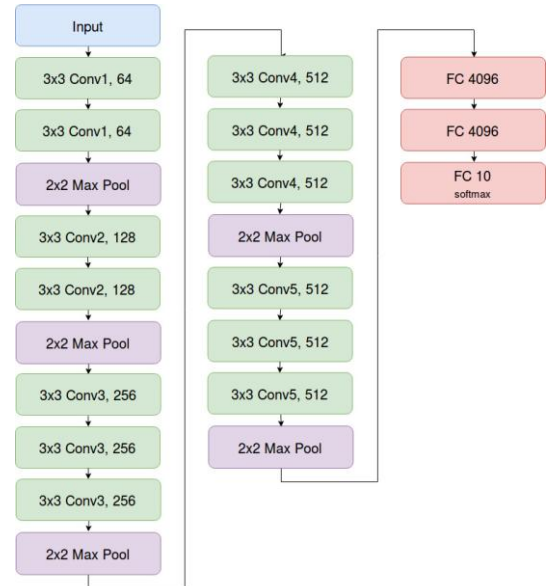


Figure 7 Architecture of VGGNet16.

ข้อมูลพรรณไม้ (Plant Datasets)

ข้อมูลพรรณไม้ที่ใช้ในงานวิจัยนี้มีทั้งสิ้น 3 ชุดข้อมูล ประกอบด้วยชุดข้อมูล Plants in Natural Environment (PNE), Folio และ 102 Flower

ชุดข้อมูลพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ (Plants in Natural Environment Dataset: PNE Dataset)

ชุดข้อมูล PNE ที่นำเสนอในงานวิจัยฉบับนี้ประกอบด้วยพรรณไม้จำนวน 10 ชนิด ชนิดละ 300 รูปภาพ ชนิดของพรรณไม้ที่อยู่ในชุดข้อมูล PNE แสดงดัง Table 1 และตัวอย่างของพรรณไม้แสดงดัง Figure 8 รูปภาพพรรณไม้เป็นรูปภาพที่อยู่ในสิ่งแวดล้อมทางธรรมชาติที่เก็บรวบรวมจากสวนไม้หอมและบริเวณรอบมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี วิทยาเขตสุรินทร์¹⁶ โดยรูปภาพบางส่วนก็รวบรวมมาจากอินเทอร์เน็ต สุดท้ายแล้ว รูปภาพพรรณไม้ทุกรูปถูกนำมาคัดกรองและยืนยันความถูกต้องของสายพันธุ์ (Specie)

ความท้าทายในการจำแนกประเภทรูปภาพพรรณไม้ในชุดข้อมูล PNE ได้แก่ ความแตกต่างของพรรณไม้ชนิดเดียวกัน แต่มีสี (Color) ที่แตกต่างกัน เช่น ดอกกุหลาบหิน มีทั้งดอกสีชมพูอ่อน ชมพู และแดง (Figure 8 แถวที่ 1) ทั้งนี้มีดอกไม้จำนวน 5 ประเภทที่เป็นดอกไม้สีขาว ประกอบด้วย ดอกนางแย้ม พุทราชาติ มะลิลา โมกพวง สีสาวดี (Figure 8 แถวที่ 2, 4-7) ดอกพุทราชาติและมะลิลา (Figure แถวที่ 4, 5)

ที่รูปของดอกไม้ไม่ได้เป็นจุดเด่นของภาพ และดอกสำหลัา บัตตานี ที่แสดงใน Figure 8 แถวที่ 10 แสดงถึงรูปร่าง (Shape) ที่มีความแตกต่างกัน โดยมีทั้งดอกตูม และดอกที่ กำลังบาน

Table 1 List of Thai Plants in the PNE Dataset.

Class No.	Plant Name	Thai Plant Name
1	Fairy Rose	กุหลาบหนู
2	Glory Tree	นางแย้ม
3	Antigonía	พวงชมพู
4	Jasminum Auriculatum	พุทธรชาด
5	Arabian Jasmine	มะลิลา
6	Wild Water Plum	โมกพวง
7	Plumeria	ลีลาวดี
8	Rangoon Creeper	เล็บมือนาง
9	Cape of Good Hope	วาสนา
10	Dwarf Ylang-Ylang Shrub	สำหลัา บัตตานี

ชุดข้อมูล 102 Flower (102 Flower Dataset)

ชุดข้อมูล 102 Flower¹⁷ เป็นรูปภาพของจำนวน 102 สายพันธุ์ โดยแต่ละสายพันธุ์ถูกเก็บรวบรวมตั้งแต่ 40-250 รูปภาพ รวมทั้งสิ้น 8,189 รูปภาพ โดย 102 Flower Dataset เป็นรูปภาพดอกไม้ที่พบทั่วไปในประเทศอังกฤษ (แสดงดัง Figure 9) และถูกนำไปใช้ในการจำแนกประเภทดอกไม้ (Flower Classification) โดยการจำแนกประเภทนั้น ใช้วิธีการ SIFT, HOG และ HSV Color Space เพื่อหาคุณลักษณะพิเศษ และใช้วิธี SVM ในการจำแนกประเภทดอกไม้



Figure 8 Images from the PNE dataset. Each row shows 3 images from the same category.



Figure 9 Sample images from 102 Flower dataset.

ชุดข้อมูลFolio (Folio Dataset)

Folio Dataset¹⁸ ประกอบด้วยใบไม้ทั้งสิ้น 32 ชนิด ซึ่งเก็บรวบรวมชนิดละ 20 ใบ รวมแล้วมีรูปภาพใบไม้ทั้งสิ้น 640 รูปภาพ ที่เก็บรวบรวมจากบริเวณ University of Mauritius รูปภาพใบไม้ถูกถ่ายโดยกำหนดให้มีพื้นหลังเป็นฉากสีขาว โดย Folio Dataset ถูกเก็บรวบรวมเพื่อนำรูปภาพใบไม้ไปหาคุณลักษณะพิเศษ โดยใช้ลักษณะรูปร่างของใบ (Shape Feature) และ ฮิสโตแกรมสี (Color Histogram) และใช้ kNN เพื่อจำประเภทใบไม้ (Plant Leaf Recognition) ตัวอย่างของ Folio Dataset แสดงใน Figure 10

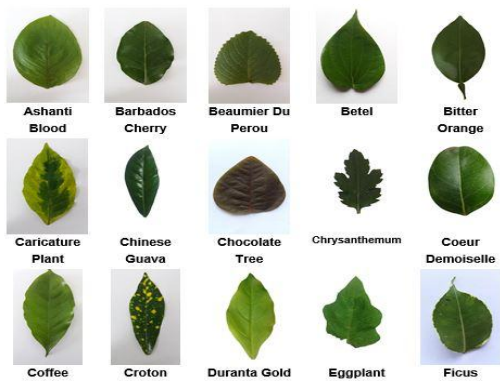


Figure 10 Sample images from Folio dataset.

ผลการทดลองและอภิปรายผล (Result and Discussion)

งานวิจัยฉบับนี้ รูปภาพที่ใช้ในการวิจัยถูกปรับให้มีขนาด 256x256 พิกเซล โดยชุดข้อมูลถูกแบ่งด้วยสัดส่วน 80:20 เพื่อใช้เป็นข้อมูลชุดเรียนรู้ (Training Set) และข้อมูลชุดทดสอบ (Test Set) ตามลำดับ ใช้วิธี Stratified Cross-Validation¹⁹ กำหนดให้มีจำนวน 5-fold โดยใช้ค่าความถูกต้อง (Accuracy) และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) สำหรับการประเมินประสิทธิภาพการทดลอง โดยใช้อัตราความถูกต้องแบบ Top-1 (Top-1 Accuracy) และใช้ GPU รุ่น GeForce GTX 1060 แรม 6GB ในการทดลอง

ชุดข้อมูลที่ใช้ในการทดลองจำนวน 3 ชุด ประกอบด้วย ชุดข้อมูล PNE, 102 Flower และ Folio (รายละเอียดของชุดข้อมูลแสดงดัง Table 2) ถูกนำมาใช้เพื่อทดสอบประสิทธิภาพของวิธีการเรียนรู้เชิงลึก (Deep Learning) ด้วยโครงข่ายประสาทเทียมแบบ

คอนโวลูชัน (Convolutional Neural Network: CNN) โดยโครงสร้าง (Architecture) ที่นำมาเพื่อเปรียบเทียบประกอบด้วย LeNet-5, AlexNet, GoogLeNet และ VGGNet^{2,13-15}

Table 2 Overview of plant datasets.

Dataset	Category	Image	Training	Test
PNE	10	3,000	2400	600
Folio	32	637	510	127
102 Flower	102	8,189	6,564	1,625

การประเมินผลการทดลองในชุดข้อมูล PNE (PNE Dataset Evaluation)

สำหรับการทดสอบกับชุดข้อมูล PNE ที่ประกอบไปด้วยสมุนไพรมะนาว 10 ชนิด และมีรูปภาพสมุนไพรมะนาวทั้งสิ้น 3,000 รูปภาพ โดยใช้วิธี Deep CNN พบว่าวิธี CNN ที่ใช้โครงสร้างแบบ GoogLeNet ที่กำหนดให้เรียนรู้จำนวน 10,000 รอบ (Epoch) มีอัตราความถูกต้อง 99.17% ซึ่งสูงที่สุด ตามด้วย VGGNet (เรียนรู้จำนวน 1,000 และ 10,000 รอบ) มีอัตราการเรียนรู้ 97.17 และ 97.33% ตามลำดับ แต่ทั้งนี้หากใช้การเรียนรู้จำนวน 10,000 รอบ พบว่า VGGNet และ AlexNet มีอัตราความถูกต้องเท่ากันที่ 97.33% แต่ VGGNet มีค่าความเบี่ยงเบนมาตรฐานที่ต่ำกว่า แสดงว่าจากการทดลองหลายรอบ VGGNet ให้ผลความถูกต้องที่ไม่เกิดการกระจายเมื่อเทียบกับ AlexNet

หากเปรียบเทียบความเร็วในการประมวลผลของ Deep CNN ในการทดสอบด้วยการเรียนรู้จำนวน 10,000 รอบ โครงสร้างแบบ VGGNet ใช้เวลาในการประมวลผลนานที่สุดประมาณ 6 วัน 7 ชั่วโมง (~6.7d) ตามด้วย GoogLeNet, LeNet-5 และ AlexNet เวลาในการประมวลผล ~1.9d, ~14h และ ~6.5h ตามลำดับ ผลการทดลอง PNE Dataset แสดงดัง Table 3

Table 3 Test accuracy comparison of CNN architectures on PNE dataset.

CNN Architectures	Number of Epochs / Training Time			
	1,000	Time	10,000	Time
LeNet-5	88.50±0.14	~1.3h	89.17±2.13	~14h
AlexNet	96.67±0.23	~1.5h	97.33±2.18	~6.5h
GoogLeNet	99.04±0.11	~4.3h	99.17±1.75	~1.9d
VGGNet	97.17±0.64	~20h	97.33±0.85	~6.7d

การประเมินผลการทดลองในชุดข้อมูล 102 Flower (102 Flower Dataset Evaluation)

การเปรียบเทียบระหว่างวิธี Deep CNN และวิธีที่รวมกันระหว่าง HSV+SIFT+HOG¹⁷ โดยนำมาทดสอบกับชุดข้อมูล 102 Flower พบว่าวิธี Deep CNN ที่มีโครงสร้างแบบ GoogLeNet (เรียนรู้จำนวน 10,000 รอบ) มีอัตราความถูกต้องสูงที่สุด โดยมีความถูกต้อง 78.89% ซึ่งสูงกว่าวิธี HSV+SIFT+HOG ที่มีความถูกต้อง 72.8%

หากเปรียบเทียบเฉพาะ Deep CNN พบว่าจำนวนในการเรียนรู้ 10,000 รอบให้ผลการทดลองที่สูงกว่าจำนวน 1,000 รอบ โดยโครงสร้างที่ให้อัตราความถูกต้องสูงสุดเรียงตามลำดับคือ GoogLeNet, AlexNet, VGGNet และ LeNet-5 โดยมีอัตราความ ถูกต้อง 78.89, 66.38, 64.12 และ 31.42% ตามลำดับ ผลลัพธ์ที่ได้จากการทดลองแสดงดัง Table 4

Table 4 The accuracy of CNN architectures on 102 Flower dataset.

CNN Architectures	Number of Epochs / Training Time			
	1,000	Time	10,000	Time
LeNet-5	33.19±0.27	~4.4h	31.42±1.12	~1.16d
AlexNet	66.06±1.74	~6.3h	66.38±2.46	~1.6d
GoogLeNet	74.98±1.59	~1.3d	78.89±0.69	~4.4d
VGGNet	62.60±0.87	~1.19d	64.12±2.74	~17d

การประเมินผลการทดลองในชุดข้อมูล Folio (Folio Dataset Evaluation)

ในงานวิจัย¹⁸ ใช้วิธีการหาคุณลักษณะพิเศษจากรูปร่างของใบไม้ (Shape Feature) และใช้ ฮีสโตแกรมของสี (Color Histogram) ร่วมกับ KNN เพื่อรู้จำใบไม้จากชุดข้อมูล Folio จากงานวิจัยระบุว่าวิธีดังกล่าวมีอัตราความถูกต้อง 87.3% ซึ่งมีอัตราความถูกต้องสูงกว่าวิธี CNN ที่ใช้โครงสร้างแบบ LeNet-5 และ AlexNet โดยโครงสร้างแบบ LeNet-5 และ AlexNet มีอัตราความถูกต้อง 84.25 และ 85.04% เมื่อทำการเรียนรู้จำนวน 1,000 รอบ ในทางกลับกัน กรณีที่เรียนรู้ด้วยจำนวนรอบที่เพิ่มขึ้น เมื่อทดสอบที่ 10,000 รอบ อัตราความถูกต้องกลับลดลง โดยโครงสร้างแบบ LeNet-5 มีความถูกต้อง 74.02% และโครงสร้างแบบ AlexNet มีความถูกต้อง 73.23%

เมื่อนำผลการทดลองจากงานวิจัย¹⁸ มาเปรียบเทียบกับโครงสร้างแบบ GoogLeNet กลับพบว่ามีอัตราความถูกต้องที่ใกล้เคียงกัน โดย GoogLeNet มีอัตราความถูกต้อง 87.40% ซึ่งสูงกว่าเพียง 0.1% แต่ทั้งนี้โครงสร้างแบบ VGG-

Net เป็นวิธี Deep CNN ที่มีความถูกต้องสูงที่สุดจากการทดลองทั้งหมด โดยมีความถูกต้องถึง 91.85% แต่ต้องใช้เวลาในการเรียนรู้นานประมาณ 3 ชั่วโมง 50 นาที (~3.5h) จากการทดลองแสดงให้เห็นว่าโครงสร้างแบบ VGGNet ให้ผลการทดลองที่ดีกับชุดข้อมูลที่มีจำนวนข้อมูลไม่มาก ชุดข้อมูล Folio มีจำนวนรูปภาพใบไม้เพียง 637 รูป ผลการทดลองแสดงดัง Table 5

Table 5 Test accuracy comparison of CNN architectures on Folio dataset.

CNN Architectures	Number of Epochs / Training Time			
	1,000	Time	10,000	Time
LeNet-5	84.25±0.51	~20m	74.02±0.95	~2.5h
AlexNet	85.04±0.48	~30m	73.23±0.81	~1.4h
GoogLeNet	87.40±1.54	~3.5h	82.68±2.41	~7.2h
VGGNet	91.85±3.31	~3.5h	87.49±1.54	~1.8d

สรุปผลการทดลอง (Conclusions)

งานวิจัยฉบับนี้ นำเสนอวิธีการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก (Deep Learning) โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks) ซึ่งประกอบไปด้วยโครงสร้างแบบ LeNet-5, AlexNet, GoogLeNet และ VGGNet เพื่อใช้สำหรับการจำแนกพรรณไม้ ซึ่งได้นำไปทดสอบกับชุดข้อมูลพรรณไม้ทั้งสิ้น 3 ชุด ได้แก่ PNE, Folio และ 102 Flower Dataset

งานวิจัยฉบับนี้ได้เลือกใช้ NVIDIA Deep Learning GPU Training System (DIGITS) เป็นเครื่องมือในการทดสอบการเรียนรู้เชิงลึก ทำให้สามารถประมวลผลได้บนหน่วยประมวลผลด้านกราฟฟิก (Graphics Processing Unit: GPU) และได้เลือกใช้ GPU รุ่น GeForce GTX 1060 ที่มีขนาดหน่วยความจำ 6GB และมีจำนวน NVIDIA CUDA Cores จำนวน 1152 Core

จากผลการทดลองพบว่าโครงสร้างแบบ VGGNet ให้ผลการทดลองสูงที่สุดกับชุดข้อมูล Folio ที่มีข้อมูลขนาดเล็กเพียง 637 ชุดข้อมูล และใช้จำนวนรอบในการเรียนรู้เพียง 1,000 รอบ สำหรับโครงสร้างแบบ GoogLeNet นั้นมีผลการทดลองสูงที่สุดเมื่อทดสอบกับชุดข้อมูล PNE และ 102 Flower ที่มีจำนวนข้อมูลในการเรียนรู้ 3,000 และ 8,189 รูปภาพ ตามลำดับ ซึ่งเป็นข้อมูลจำนวนมาก แต่ต้องใช้จำนวนรอบในการเรียนรู้มากถึง 10,000 รอบ

หากเปรียบเทียบกันระหว่างโครงสร้างแบบ GoogLeNet และ VGGNet ที่มีจำนวนของชั้น (Layer) ที่ใกล้เคียงกันปรากฏว่าโครงสร้างแบบ VGGNet ใช้เวลาในการเรียนรู้นานกว่าโครงสร้างแบบ GoogLeNet แต่ในทางกลับกันเมื่อเปรียบเทียบอัตราความถูกต้องพบว่าโครงสร้างแบบ VGGNet มีอัตราความถูกต้องที่ต่ำกว่าโครงสร้างแบบ GoogLeNet เมื่อทดสอบกับข้อมูลชุด 102 Flower และ PNE ซึ่งข้อมูลทั้งสองชุดนั้นเป็นรูปภาพดอกไม้และพืชสมุนไพร และรูปภาพแต่ละรูปเป็นรูปภาพที่อยู่ในสิ่งแวดล้อมทางธรรมชาติที่มีพื้นหลังที่ซับซ้อน แต่เมื่อทดสอบกับชุดข้อมูล Folio ที่เป็นรูปภาพใบไม้ที่ถ่ายในห้องทดลองโดยกำหนดให้พื้นหลังเป็นสีขาวปรากฏว่าโครงสร้างแบบ VGGNet มีอัตราความถูกต้องสูงที่สุด

จากผลของการวิจัยสรุปได้ว่าโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก โดยใช้โครงสร้างแบบ GoogLeNet เหมาะสำหรับนำไปใช้เพื่อจำแนกพรรณไม้ที่อยู่ในสิ่งแวดล้อมทางธรรมชาติ เนื่องจากมีอัตราความถูกต้องที่สูง และยังช่วยลดระยะเวลาในการเรียนรู้

วิธีการเรียนรู้เชิงลึกจำเป็นต้องใช้ชุดข้อมูลรูปภาพจำนวนมากในการเรียนรู้ ทั้งนี้เพื่อทำให้มีอัตราการรู้จำเพิ่มสูงขึ้น งานวิจัยครั้งถัดไป จะนำเทคนิคการเพิ่มจำนวนของข้อมูลชุดเรียนรู้ที่เรียกว่า Data Augmentation ทำให้มีข้อมูลสำหรับการเรียนรู้เพิ่มขึ้น และใช้วิธีการถ่ายโอนความรู้ (Transfer Learning) ซึ่งจะช่วยลดระยะเวลาในการเรียนรู้และอาจส่งผลให้มีประสิทธิภาพในการเรียนรู้เพิ่มขึ้น^{8,20}

กิตติกรรมประกาศ

รับทุนสนับสนุนการทำวิจัย และการศึกษาจากมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี วิทยาเขตสุรินทร์ และขอบคุณผู้ช่วยศาสตราจารย์พัชรภาพรณ แสงโยจารย์ นักพฤกษศาสตร์ ที่ช่วยคัดกรองและยืนยันความถูกต้องของสายพันธุ์จากรูปภาพพรรณไม้ทุกรูป

เอกสารอ้างอิง

1. Le Cun Y, Matan O, Boser B, Denker JS, Henderson D, Howard RE, et al. Handwritten Zip Code Recognition with Multilayer Networks. In: 10th International Conference on Pattern Recognition. 1990. pp. 35–40.
2. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. In: Proceedings of the IEEE. 1998. pp. 2278–2324.
3. Ciresan D, Meier U, Schmidhuber J. Multi-column Deep Neural Networks for Image Classification. In:

IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2012. pp. 3642–9.

4. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Trans Med Imaging. 2016;35(5):1196–206.
5. Yalcin H, Razavi S. Plant Classification using Convolutional Neural Networks. In: International Conference on Agro-Geoinformatics. 2016. pp. 1–5.
6. Li Y, Hong H, Fang T. Hierarchical Segmentation of Remote Sensing Images by Unsupervised Deep Learning Features. In: International Symposium on Computational Intelligence and Design (ISCID). 2017. pp. 448–53.
7. Liu Y, Zhou D, Tang F, Meng Y, Dong W. Flower Classification via Convolutional Neural Network. In: IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications. 2017. pp. 110–6.
8. Pawara P, Okafor E, Surinta O, Schomaker L, Wiering M. Comparing Local Descriptors and Bags of Visual Words to Deep Convolutional Neural Networks for Plant Recognition. In: 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM). 2017. pp. 479–86.
9. Reyes AK, Caicedo JC, Camargo JE. Fine-tuning deep convolutional networks for plant recognition. In: Conference on Conference and Labs of the Evaluation forum (CLEF). 2015. pp. 1–9.
10. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Comput Intell Neurosci. 2016:1–11.
11. Mohanty SP, Hughes DP, Salathé M. Using Deep Learning for Image-Based Plant Disease Detection. Front Plant Sci. 2016(7):1–10.
12. Sun Y, Liu Y, Wang G, Zhang H. Deep Learning for Plant Identification in Natural Environment. Comput Intell Neurosci. 2017:1–7.
13. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Net-

- works. In: *Advances In Neural Information Processing Systems*. 2012. pp. 1–9.
14. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. pp. 1–9.
 15. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *International Conference on Learning Representations*. 2015. pp. 1–10.
 16. จินตนา รุ่งเรือง และวสา วงศ์สุขแสง. รายงานผลการดำเนินงานโครงการอนุรักษ์พันธุกรรมพืชอันเนื่องมาจากพระราชดำริ เรื่องการดูแลสวนสมุนไพรและสวนไม้หอม. *สุรินทร์*; 2558.
 17. Nilsback ME, Zisserman A. Automated Flower Classification over a Large Number of Classes. In: *Proceedings - 6th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*. 2008. p. 722–9.
 18. Munisami T, Ramsurn M, Kishnah S, Pudaruth S. Plant Leaf Recognition Using Shape Features and Colour Histogram with K-nearest Neighbour Classifiers. *Procedia Comput Sci*. 2015(58):740–7.
 19. ศักดิ์ เพิ่มพรรษา นิตยา เกิดประสพ และกิตติศักดิ์ เกิดประสพ. การประยุกต์เทคนิคแปลงข้อความแบบเวกเตอร์ และการกระจายแบบเบย์ เพื่อทำนายการบุกรุกเครือข่ายคอมพิวเตอร์. รายงานสืบเนื่องจากการประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 6. รายงานสืบเนื่องจากการประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 6; 2558. pp. 75–84.
 20. Pawara P, Okafor E, Schomaker L, Wiering MA. Data Augmentation for Plant Classification Data Augmentation for Plant Classification. In: *18th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*. 2017. pp. 615–26.