กลไกจุดสนใจแบบเน็ตเวิร์กละเอียดสำหรับการจำแนกประเภทของรูปภาพอาหาร

น.ส.วศิณี นุชศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Attentional Fine-Grained Network for Food Image Categorization

Miss Vasinee Nussiri

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2019

หัวข้อวิทยานิพนธ์	กลไกจุดสนใจแบบเน็ตเวิร์กละเอียดสำหรับการจำแนกประเภทของรูปภาพอาหาร
โดย	น.ส.วศิณี นุชศิริ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล)

..... กรรมการ
(อาจารย์ ดร.ดวงดาว วิชาดากุล)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ธนภัทร ฆังคะจิตร)

วศิณี นุชศิริ : กลไกจุดสนใจแบบเน็ตเวิร์กละเอียดสำหรับการจำแนกประเภทของรูปภาพอาหาร. (Attentional Fine-Grained Network for Food Image Categorization) อ.ที่ปรึกษาหลัก : ผศ. ดร.พีรพล เวทีกูล

ในปัจจุบัน มีจำนวนรูปภาพอาหารมากมายที่ถูกอัพโหลดผ่านเครือข่ายสังคม โดยรูปภาพส่วนหนึ่งไม่ได้รับการระบุป้ายชื่ออาหาร การใช้แอปพลิเคชันสำหรับการจำแนกประเภทรูปภาพของอาหาร สามารถช่วยระบุป้ายชื่อ และจัดจำแนกประเภทของรูปภาพอาหารเหล่านั้นได้ ปัญหาของงานจำแนกประเภทของรูปภาพอาหาร จัดเป็นงานที่ค่อนข้างมีความซับซ้อน เนื่องจากจำนวนของประเภทอาหารมีมากกว่าหนึ่งร้อยประเภท และอาหารบางประเภทยังมีลักษณะที่แตกต่างกันเล็กน้อย ไม่ว่าจะเป็นประเภทของส่วนผสม หรือลักษณะการจัดวางจาน ซึ่งปัญหาเหล่านี้นำไปสู่งานที่เรียกว่า งานจำแนกประเภทรูปภาพแบบละเอียด (Fine-grained Image Classification) ในปัจจุบันแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ (Bilinear Convolutional Neural Networks หรือ B-CNN) ถูกนำมาใช้ในการจำแนกประเภทของรูปภาพอาหาร เนื่องจากแบบจำลองนี้มีความแม่นยำในการจำแนกประเภทของรูปภาพสูง และสามารถสกัดลักษณะของรูปภาพออกมาอย่างหลากหลาย เพื่อโฟกัสรายละเอียดของอาหารในแต่ละประเภท แต่เนื่องจากคุณลักษณะของรูปภาพที่ถูกสกัดมานั้น บางลักษณะอาจจะไม่ได้มีความสำคัญต่อรูปภาพนั้น ๆ ด้วยเหตุผลดังกล่าว งานวิจัยนี้จึงได้นำเสนอกลไกจุดสนใจ (Attention Mechanism) มาสกัดลักษณะที่จำเพาะของรูปภาพอาหารในแต่ละประเภท อีกทั้งงานวิจัยนี้เลือกคอนโวลูชันเน็ตเวิร์กที่มีประสิทธิภาพในการจำแนกประเภทของรูปภาพดีกว่าคอนโวลูชันเน็ตเวิร์กแบบอื่น ๆ ในปัจจุบัน คือ อินเซ็บชันเวอร์ชันสาม และ อินเซ็บชันเรสเน็ตเวอร์ชันสอง (Inception-Resnet-v2 หรือ In-res-v2) มาเป็นตัวสกัดลักษณะของรูปภาพ โดยงานวิจัยนี้ได้ทำการทดลองกับชุดข้อมูลเชิงรูปภาพจาก Wongnai ซึ่งเป็นแอปพลิเคชันสำหรับการอัปโหลดรูปภาพอาหาร โดยผลการทดลองพบว่าแบบจำลองที่ได้นำเสนอ มีประสิทธิภาพในการจำแนกประเภทของรูปภาพอาหารได้อย่างถูกต้องแม่นยำมากขึ้นเมื่อเปรียบเทียบกับแบบจำลองอื่น ๆ

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2562 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6071026721 : MAJOR COMPUTER SCIENCE
KEYWORD:

Vasinee Nussiri : Attentional Fine-Grained Network for Food Image Categorization. Advisor: Asst. Prof. PEERAPON VATEEKUL

Nowadays, many food images are posted on various social network platforms without identification labels. An automatic food categorization application would greatly help to identify and classify food categories. Food categorization is a complex problem since the number of category types can be more than one hundred. Many kinds of food are similar with only subtle differences in taste and presentation and this can lead to a problem called “fine-grained issue”. Recently, a bilinear model was employed which showed good accuracy and generated excessive features to capture details among different food categories, albeit with limited performance. Diverse food categories require disparate sets of features. Here, an attention mechanism was applied to capture suitable features and specifically identify each food category. Furthermore, the performance of a bilinear backbone was also enhanced by applying Inception in correlation with Inception-ResNet-v2 and Inception-v3 networks. The experiment was conducted on the Wongnai dataset containing various images that were separated into 83 classes. Results showed that our attentional model outperformed the traditional bilinear model.

Field of Study: Computer Science Student's Signature .....
Academic Year: 2019 Advisor's Signature .....

กิตติกรรมประกาศ

การที่วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีนั้น นอกจากการทำงานของตัวผู้วิจัยแล้ว ยังมีบุคคลท่านอื่นที่เป็นส่วนสำคัญที่ได้ให้ความช่วยเหลือในการจัดทำวิทยานิพนธ์ฉบับนี้ขึ้นมา ผู้วิจัยรู้สึกทราบซึ้งในความกรุณาเหล่านี้เป็นอย่างมากจึงใคร่ขอใช้เนื้อหาในส่วนกิตติกรรมประกาศของวิทยานิพนธ์ฉบับนี้แสดงความขอบพระคุณเป็นอย่างสูงมา ณ ที่นี้

ขอขอบพระคุณอาจารย์ที่ปรึกษา ผศ. ดร. พีรพล เวทีกูล ที่ได้อุทิศเวลาอันมีค่า ให้คำปรึกษาแนะนำรวมทั้งแนวคิดในการทำวิทยานิพนธ์ ตลอดจนเอาใจใส่ตรวจแก้ไขข้อบกพร่องต่างๆ ทำให้วิทยานิพนธ์มีความสมบูรณ์และคุณค่ายิ่งขึ้น อีกทั้งนอกจากจะคอยให้คำปรึกษาทั้งด้านวิชาการ งานวิจัย และตรวจทานวิทยานิพนธ์แล้วยังคอยให้คำปรึกษาด้านทักษะการใช้ชีวิต และยังเป็นตัวอย่างที่ดีให้แก่ผู้วิจัยในด้านการทำงานและการแก้ไขปัญหามาโดยตลอด

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ ซึ่งประกอบไปด้วย ศ. ดร. บุญเสริม กิจสิริกุล อ.ดร. ดวงดาว วิชาดากุล และ ดร. ธนภัทร ฆังคะจิตร ที่ได้กรุณาให้เกียรติเป็นคณะกรรมการ รวมทั้งให้คำปรึกษาและข้อเสนอแนะอันเป็นประโยชน์อย่างมากต่อการทำวิจัยและวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณครอบครัวและบุคคลอันเป็นที่รักของผู้วิจัยที่ให้การสนับสนุนในทุก ๆ ด้าน และคอยให้กำลังใจตลอดระยะเวลาในการดำเนินการทำงานวิจัยนี้

สุดท้ายนี้ คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องสักการะบูชาพระคุณบิดามารดา ครูอาจารย์ ตลอดจนผู้มีพระคุณทุกท่านที่ให้การอบรมสั่งสอนประสิทธิ์ประสาทวิชา ส่งเสริมสนับสนุนและชี้แนะแนวทางการศึกษาจนทำให้ผู้วิจัยก้าวมาสู่ความสำเร็จในครั้งนี้

วศิณี นุชศิริ

สารบัญ

หน้า

สารบัญรูปภาพ

สารบัญตาราง

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

การเพิ่มขึ้นอย่างรวดเร็วของรูปภาพที่ถูกอัปโหลดหรือถูกแชร์ในปัจจุบัน เกิดจากอิทธิผลของเครือข่ายสังคม โดยรูปภาพอาหารจัดเป็นหนึ่งในประเภทของรูปภาพที่ได้รับความนิยม และมีจำนวนครั้งในการถูกอัปโหลดจากผู้ใช้งานบ่อยบนเครือข่ายสังคม เพื่อบอกเล่าชีวิตประจำวัน หรือเพื่อแนะนำอาหารที่พวกเขาถูกใจให้แก่ผู้อื่น รูปภาพอาหารบางรูปได้รับการระบุป้ายชื่ออาหารว่าเป็นอาหารประเภทใด แต่สำหรับบางรูปยังไม่ได้รับการระบุ รูปภาพที่ไม่ได้รับการระบุป้ายชื่ออาหารก่อให้เกิดความยากลำบากในการค้นหาข้อมูลสำหรับรูปภาพนั้น ๆ ต่อผู้ใช้ ดังนั้น การจำแนกประเภทรูปภาพจึงมีประโยชน์ต่อการค้นหาข้อมูลเกี่ยวกับรูปภาพอาหาร และสามารถช่วยให้ผู้ใช้สามารถจำแนกรูปภาพอาหารได้อย่างถูกต้อง รูปภาพอาหารแต่ละประเภทมักจะมีลักษณะที่คล้ายคลึงกัน ตัวอย่างเช่น ในบางประเภทมีรูปร่างที่แตกต่างกันเล็กน้อย มีสีที่ใกล้เคียงกัน หรือมีการจัดแต่งจานที่คล้ายกัน และในอาหารบางประเภท ยังมีส่วนผสมที่เหมือนกันอีกด้วย สิ่งเหล่านี้ทำให้การจำแนกประเภทของรูปภาพอาหารเป็นงานที่มีความท้าทายและมีความซับซ้อน

นักวิจัยหลายคนได้นำเสนองานวิจัยที่หลากหลาย และแตกต่างกันสำหรับการจำแนกประเภทของรูปภาพอาหาร ในงานวิจัย [1] นำเสนอแบบจำลองที่มีแนวคิดของกูเกิลเน็ต (GoogLeNet) [2] ซึ่งมีพื้นฐานมาจากนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน (Convolutional Neural Network) เป็นตัวสกัดคุณลักษณะของรูปภาพ เพื่อที่จะใช้ในการสอนแบบจำลองสำหรับการจำแนกประเภทของรูปภาพอาหาร โดยใช้ชุดข้อมูลสาธารณะที่มีอยู่แล้ว งานวิจัย [3] ได้นำเสนอ อินเซ็บชันเรสเน็ต (Inception-Resnet) และเทคนิคการปรับจูน เพื่อพัฒนาประสิทธิภาพของการจำแนกประเภทของรูปภาพอาหาร ผลการทดลองพบว่า อินเซ็บชันเรสเน็ต มีประสิทธิภาพดีกว่ากูเกิลเน็ต เนื่องจากแบบจำลอง อินเซ็บชันเรสเน็ต มีการเพิ่มการเชื่อมต่อทางลัดในการส่งข้อมูลคุณลักษณะไปยังชั้นที่ต่ำกว่า จึงทำให้การจำแนกประเภทง่ายขึ้น และสามารถช่วยแก้ปัญหาการเกิด แวนิชชิงกราเดียน (Vanishing Gradient) จึงช่วยเพิ่มความแม่นยำให้กับแบบจำลอง

งานวิจัยที่อ้างอิงข้างต้นยังให้ประสิทธิภาพในการจำแนกประเภทของรูปภาพของอาหารอย่างแม่นยำไม่มากนัก เนื่องจากรูปภาพอาหารจัดเป็นรูปภาพที่มีความคล้ายคลึงกันค่อนข้างสูง เลยส่งผลให้ค่าความแม่นยำในแต่ละประเภทมีความแม่นยำค่อนข้างต่ำ และยังมีการจำแนกประเภทของอาหารผิดประเภทอยู่บ่อยครั้ง ดังนั้น งานจำแนกประเภทรูปภาพแบบละเอียด (Fine-grained Image Classification) จึงเป็นงานที่สามารถแก้ไขปัญหาดังกล่าวได้ เนื่องจากงานจำแนกประเภทรูปภาพ

แบบละเอียด มุ่งเน้นไปยังการคัดแยกรูปภาพระหว่างประเภทที่ยากต่อการจำแนก หรือมีลักษณะของภาพที่มีความแตกต่างกันเพียงเล็กน้อย นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ (Bilinear Convolutional Neural Networks หรือ B-CNN) [4] เป็นแบบจำลองที่ได้รับความนิยมในงานจำแนกประเภทรูปภาพอาหารแบบละเอียด โดยแบบจำลองถูกสร้างมาจากสองนิวรอลเน็ตเวิร์กคอนโวลูชัน ซึ่งคุณลักษณะที่ถูกสกัดจากนิวรอลเน็ตเวิร์กคอนโวลูชันทั้งสอง จะถูกรวมกันด้วยการคูณแบบการคูณภายนอกตามแต่ละตำแหน่งของรูปภาพของสองเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะ ดังนั้นนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่จึงสามารถจับตำแหน่งของลักษณะที่มีปฏิสัมพันธ์กันในลักษณะคงที่ ซึ่งลักษณะดังกล่าวทำให้นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่มีประโยชน์ต่อการแก้ไขปัญหาของงานจำแนกประเภทรูปภาพอาหารแบบละเอียด งานวิจัย [5] ประยุกต์นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ โดยนำโครงข่ายเชิงลึกมาเป็นตัวสกัดคุณลักษณะ ผลการทดลองพบว่านิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มี อินเซ็บชันเวอร์ชันสาม (Inception-v3 หรือ In-v3) [6] และ อินเซ็บชันเวอร์ชันสี่ (Inception-v4 หรือ In-v4) [7] เป็นตัวสกัดคุณลักษณะที่มีประสิทธิภาพดีกว่าแบบจำลองอื่น ๆ

กลไกจุดสนใจ (Attention Mechanism) [8] ถูกนำมาประยุกต์ใช้กับการเรียนรู้เชิงลึก (Deep Learning) เพื่อที่จะพัฒนาผลการทดลองของงานจำแนกประเภทรูปภาพ เนื่องจาก กลไกจุดสนใจจะมุ่งเน้นไปยังการสกัดคุณลักษณะสำคัญของรูปภาพ โดยคัดเลือกลักษณะที่มีความจำเพาะต่อรูปภาพนั้น ๆ ท่ามกลางคุณลักษณะที่หลากหลายของรูปภาพ เพื่อที่จะช่วยทำให้การจำแนกประเภทรูปภาพมีความถูกต้องแม่นยำมากขึ้น

งานวิจัยนี้จะนำแนวคิดของการจำแนกประเภทรูปภาพแบบละเอียด เพื่อที่จะใช้ในการจำแนกประเภทของรูปภาพอาหารอย่างแม่นยำ โดยนำนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่มาใช้ในงานวิจัย เนื่องจากเป็นแบบจำลองที่มีความนิยม และให้ความแม่นยำสูงในการจำแนกรูปภาพแบบละเอียดมากกว่าแบบจำลองอื่น ๆ โดยงานวิจัยนี้เลือกคอนโวลูชันเน็ตเวิร์กที่มีประสิทธิภาพในการจำแนกประเภทของรูปภาพดีกว่าคอนโวลูชันเน็ตเวิร์กแบบอื่น ๆ ในปัจจุบัน คือ อินเซ็บชันเวอร์ชันสาม และ อินเซ็บชันเรสเน็ตเวอร์ชันสอง (Inception-Resnet-v2 หรือ In-res-v2) มาเป็นตัวสกัดลักษณะของรูปภาพ แต่เนื่องจากคุณลักษณะของรูปภาพที่ถูกสกัดมานั้น มีความหลากหลาย แต่คุณลักษณะบางคุณลักษณะอาจจะไม่ได้มีความจำเป็นหรือสำคัญต่อรูปภาพนั้น ๆ ด้วยเหตุผลดังกล่าว งานวิจัยนี้จึงนำกลไกจุดสนใจมาจับลักษณะที่จำเพาะของรูปภาพของแต่ละประเภทอาหารนั้น ๆ อีกทั้งยังได้ทดลองกับข้อมูลนำเข้าเชิงรูปภาพที่มีขนาดต่างกัน ระหว่างขนาดรูปภาพที่มีความละเอียดต่ำและรูปภาพที่มีความละเอียดสูงกว่า เพื่อวิเคราะห์ความสำคัญของความละเอียดรูปภาพ ที่มีผลต่อประสิทธิภาพของงานจำแนกประเภทรูปภาพ โดยได้ทำการทดลองกับชุดข้อมูลเชิงรูปภาพ จาก Wongnai ซึ่งเป็นแอปพลิเคชันสำหรับการอัปโหลดรูปภาพอาหาร

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อนำเสนอวิธีการจำแนกประเภทของรูปภาพอาหาร โดยใช้แบบจำลองการเรียนรู้เชิงลึกที่สามารถพิจารณาความสำคัญของแต่ละคุณลักษณะที่แตกต่างกันในแต่ละประเภทอาหาร โดยมุ่งเน้นการปรับปรุงแบบจำลองเชิงลึกเพื่อให้สามารถจำแนกประเภทอาหารที่มีความคล้ายคลึงกันให้เกิดความแม่นยำมากที่สุด

1.3 ขอบเขตการวิจัย

1. งานวิจัยฉบับนี้ครอบคลุมเฉพาะการจำแนกประเภทของรูปภาพอาหารจากแอปพลิเคชัน Wongnai เท่านั้น
2. เปรียบเทียบประสิทธิภาพของแบบจำลอง โดยใช้ค่าเฉลี่ยมาตรฐานเอฟวัน (F1)
3. ทำการทดลองโดยใช้กลไกจุดสนใจหลายแบบ เพื่อหาแบบจำลองที่ดีที่สุด สำหรับการจำแนกประเภทของรูปภาพ

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถเพิ่มประสิทธิภาพในการจำแนกประเภทของรูปภาพอาหารได้อย่างแม่นยำ
2. สามารถประหยัดเวลาในการระบุป้ายชื่ออาหาร
3. สามารถนำกรอบงานวิจัยนี้ไปประยุกต์ใช้กับข้อมูลรูปภาพอื่น ๆ ที่มีลักษณะคล้ายกันได้

1.5 ขั้นตอนการดำเนินงาน

1. ศึกษาเกี่ยวกับนิวรอลเน็ตเวิร์ก และการเรียนรู้เชิงลึกที่ใช้ในการจำแนกประเภทรูปภาพ
2. เก็บรวบรวมข้อมูล และออกแบบการทดลอง
3. สร้างวิธีการทดลอง พัฒนาแบบจำลอง และเก็บผลการทดลอง
4. สรุปผลการทดลองทั้งหมด
5. สอบหัวข้อวิทยานิพนธ์
6. ทำการทดลองตามสิ่งที่นำเสนอ
7. ปรับค่าพารามิเตอร์ของแบบจำลอง เพื่อให้ได้ประสิทธิภาพดีที่สุด
8. เขียนบทความเพื่อตีพิมพ์ผลงานทางวิชาการ
9. สรุปผล และเรียบเรียงวิทยานิพนธ์
10. สอบวิทยานิพนธ์

1.6 ผลงานวิจัยที่ตีพิมพ์

“Food Image Categorization Using Attentional Bilinear Model” โดย วศิณี นุชศิริ และ พีรพล เวทีกูล ในงานประชุมวิชาการ “2019 - The 11th International Conference on Information Technology and Electrical Engineering (ICITEE 2019)” ซึ่งจัดขึ้น ณ โรงแรมฮอลิเดย์อินน์ จังหวัดชลบุรี ประเทศไทย ระหว่างวันที่ 10-11 ตุลาคม 2562

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

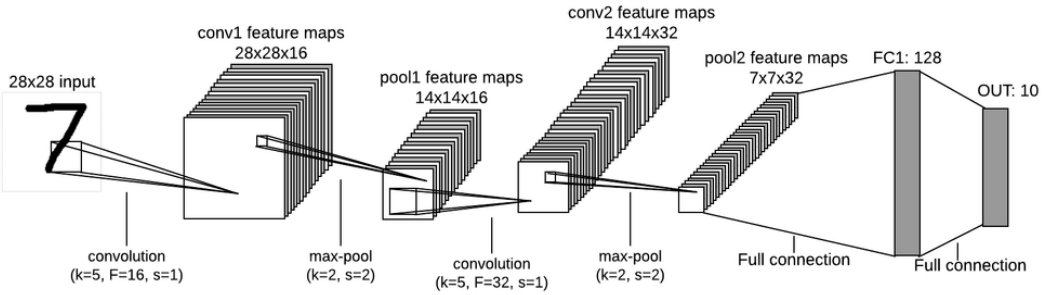
ทฤษฎีที่เกี่ยวข้องกับงานวิจัยนี้แบ่งออกได้เป็น 3 หัวข้อ ได้แก่ นิวรอลเน็ตเวิร์กเชิงลึก นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ และกลไกจุดสนใจ

2.1 นิวรอลเน็ตเวิร์กเชิงลึก (Deep Neural Network)

คือนิวรอลเน็ตเวิร์กที่มีชั้นซ่อน (Hidden Layer) จำนวนหลาย ๆ ชั้น ตัวอย่างของนิวรอลเน็ตเวิร์กเชิงลึกเช่น เน็ตเวิร์กความเชื่อเชิงลึก (Deep Belief Network หรือ DBN) นิวรอลเน็ตเวิร์กแบบวกกลับ (Recurrent Neural Network หรือ RNN) หน่วยความระยะสั้นแบบยาว (Long-Short Term Memory หรือ LSTM) นิวรอลเน็ตเวิร์กคอนโวลูชัน (Convolutional Neural Network หรือ CNN) จุดเด่นของนิวรอลเน็ตเวิร์กที่มีชั้นซ่อน คือความสามารถในการเรียนรู้คุณลักษณะ (Feature) ที่ซ่อนอยู่ในข้อมูลรับเข้า จึงมีความแตกต่างกับนิวรอลเน็ตเวิร์กทั่วไปที่จะต้องสกัดข้อมูลตัวแทนก่อน โดยรายละเอียดของนิวรอลเน็ตเวิร์กเชิงลึกที่ใช้ในงานวิจัยนี้ มีดังต่อไปนี้

2.1.1 นิวรอลเน็ตเวิร์กคอนโวลูชัน (Convolutional Neural Network หรือ CNN)

เป็นนิวรอลเน็ตเวิร์กเชิงลึกที่มีจุดเริ่มต้นมาจากการรู้จำภาพตัวอักษร โดยจะแปลงรูปภาพเป็นเมตริกซ์แล้วจึงนำเข้านิวรอลเน็ตเวิร์ก จากนั้นจะใช้ตัวกรอง (filter) เพื่อสร้างเป็นฟีเจอร์ใหม่ (Feature Map) เพื่อใช้เป็นข้อมูลนำเข้าของชั้นถัดไป โครงสร้างของนิวรอลเน็ตเวิร์กคอนโวลูชัน ซึ่งเกิดจากการนำชั้นหลายๆประเภทมาประกอบเข้าด้วยกัน โครงสร้างพื้นฐานของนิวรอลเน็ตเวิร์กคอนโวลูชันแสดงดังรูปที่ 1



รูปที่ 1 โครงสร้างนิวรอลเน็ตเวิร์กคอนโวลูชัน

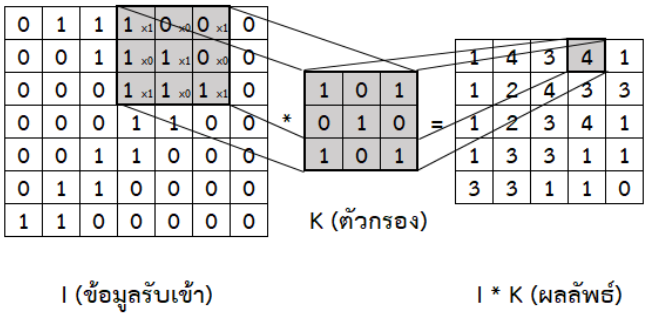
(ที่มา: https://www.easy-tensorflow.com/tf-tutorials/convolutional-neural-nets-cnns/ Accessed: August 20, 2019)

โดยข้อมูลรายละเอียดของชั้นต่าง ๆ ในนิวรอลเน็ตเวิร์กคอนโวลูชันมีองค์ประกอบที่ต้องพิจารณาดังต่อไปนี้

1) ชั้นคอนโวลูชัน (Convolutional Layer)

หน้าที่ของชั้นคอนโวลูชัน คือหาฟีเจอร์ของกลุ่มข้อมูลที่อยู่ใกล้ ๆ กัน โดยใช้ผลคูณเชิงสเกลาร์ (dot product) ของเมทริกซ์กับตัวกรอง (filter) โดยทุก ๆ การทำคอนโวลูชันของข้อมูลนำเข้า จะใช้ค่าน้ำหนักของตัวกรองร่วมกัน จำนวนผลลัพธ์ที่ได้จากชั้นคอนโวลูชันจะเท่ากับจำนวนของตัวกรองที่ใช้ ซึ่งหลังจากการทำคอนโวลูชันจะมีการใช้ฟังก์ชันกระตุ้น เพื่อนำผลลัพธ์ที่ได้ส่งต่อไปเป็นข้อมูลนำเข้าสำหรับเน็ตเวิร์กชั้นต่อไป โดยตัวอย่างการทำคอนโวลูชันแสดงดังรูปที่ 2 กำหนดให้ข้อมูลรับเข้าแทนด้วยเมทริกซ์ และตัวกรองแทนด้วยเมทริกซ์ ซึ่งมีขนาด ผลลัพธ์ของการทำคอนโวลูชัน สามารถคำนวณได้จากสมการ (1)

$$(I * K)_{xy} = \sum_{i=1}^{h}\sum_{j=1}^{w} K_{ij} \cdot I_{x+i-1,y+j-1} \tag{1}$$



รูปที่ 2 ตัวอย่างการทำคอนโวลูชัน

(รูปนี้ดัดแปลงจาก: http://2017g10ceimageprocessing.blogspot.com/2017/11/blog-post_29.html/ Accessed: December 18, 2019)

ในชั้นคอนโวลูชัน มีองค์ประกอบที่ต้องพิจารณาดังต่อไปนี้

1.1) ขนาดของตัวกรอง (Filter Size) คือความกว้างและความสูงของตัวกรองที่จะนำมาใช้ในการทำคอนโวลูชัน (ค่า w และ h ในสมการที่ 1) โดยในตัวอย่างการทำคอนโวลูชันในรูปที่ 2 ใช้ตัวกรองที่มีขนาด 3x3

1.2) ชนิดของการทาคอนโวลูชัน (Convolution Type)

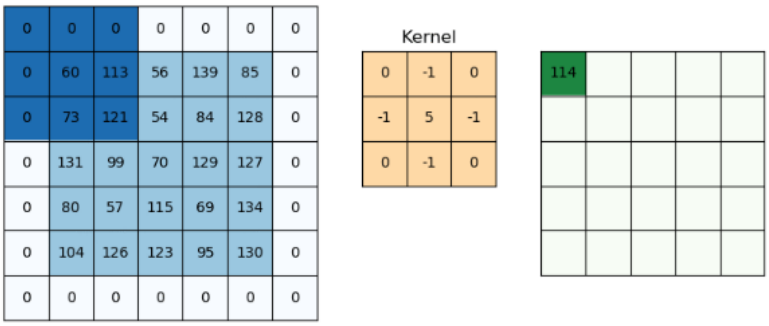
- คอนโวลูชันแบบแคบ (Narrow Convolution)

คอนโวลูชันที่ถูกนำมาใช้ส่วนใหญ่เป็นการทำคอนโวลูชันแบบแคบ ซึ่งตัวกรองที่นำมาจะไม่มีการทำเลยขอบเมตริกซ์ของข้อมูลนำเข้า โดยผลลัพธ์ที่ได้จากข้อมูลนำเข้าที่มีขนาด

NxN กับตัวกรองที่มีขนาด MxM จะได้เมตริกซ์ขนาด (N-M+1) x (N-M+1) แสดงตัวอย่างดังรูปที่ .22

- คอนโวลูชันแบบกว้าง (Wide Convolution)

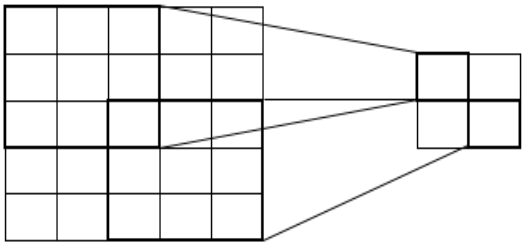
เป็นการทำคอนโวลูชันที่มีการกระทำเลยขอบเมตริกซ์ของข้อมูลนำเข้า โดยในพื้นที่ส่วนที่เกินขอบเมทริกซ์จะถูกทำการเสริมเติม (Padding) ออกไปด้วยค่า 0 ผลลัพธ์ที่ได้จากข้อมูลนำเข้าที่มีขนาด NxN กับตัวกรองที่มีขนาด MxM จะได้เมตริกซ์ขนาด (N+M-1) x (N+M-1) จุดเด่นของการทำคอนโวลูชันแบบกว้างคือ เพื่อป้องกันการสูญเสียข้อมูลตรงบริเวณขอบของข้อมูลนำเข้า แสดงตัวอย่างดังรูปที่ 3



รูปที่ 3 การทำคอนโวลูชันแบบกว้าง

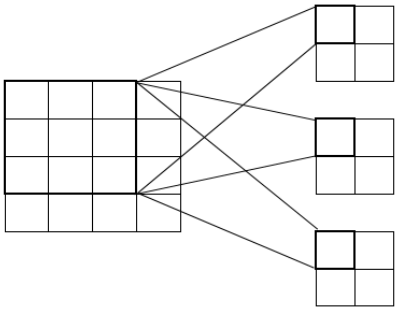
(ที่มา: https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional-layers/ Accessed: August 20, 2019)

1.3) ขนาดของการก้าวข้าม (Stride Size) คือจำนวนช่องของข้อมูลรับเข้า ที่จะเลื่อนไปเมื่อทำ การหาผลลัพธ์ของการคอนโวลูชันในแต่ละช่อง โดยทั่วไปมักจะใช้ขนาดของการก้าวข้ามเป็น 1 ตัวอย่างการทำคอนโวลูชันที่มีขนาดของการก้าวข้ามเป็น 1 แสดงในรูปที่ 2 ลักษณะของการทำคอนโวลูชันที่มีขนาดของการก้าวข้ามเป็น 2 แสดงตัวอย่างดังรูปที่ 4



รูปที่ 4 การทำคอนโวลูชันขนาด 5x5 ตัวกรองขนาด 3x3 และมีขนาดของการก้าวข้ามเป็น 2 (รูปนี้ดัดแปลงจาก: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2/ Accessed: December 18, 2019)

1.4) จำนวนตัวกรอง (Number of Filters) คือจำนวนของตัวกรองที่ใช้ในแต่ละชั้นของการทำคอนโวลูชันซึ่งสามารถมีตัวกรองได้มากกว่าหนึ่งตัว และน้ำหนักของตัวกรองแต่ละตัวอาจจะมีความแตกต่างกันได้ ซึ่งการกำหนดจำนวนตัวกรองในชั้นคอนโวลูชันใด ๆ จะเป็นการกำหนดจำนวนช่องสัญญาณ (Channel) ของข้อมูลนำเข้าสำหรับชั้นถัดไป แสดงตัวอย่างดังรูปที่ 5



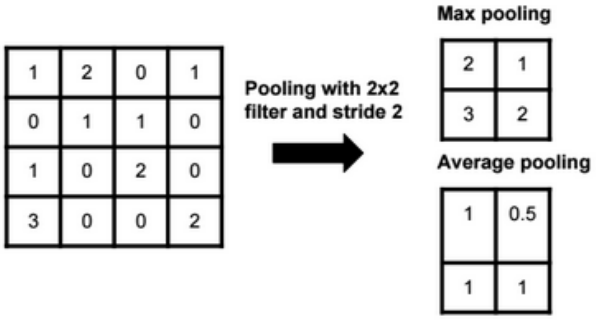
รูปที่ 5 การทำคอนโวลูชันโดยมีจำนวนตัวกรองเท่ากับ 3

(รูปนี้ดัดแปลงจาก: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2/ Accessed: December 18, 2019)

1.5) จำนวนช่องสัญญาณ (Channel) หรือความลึกของข้อมูลรับเข้า โดยจำนวนช่องสัญญาณอาจจะมีค่ามากกว่าหนึ่งค่าได้ เช่น ในการวิจัยทางด้านรูปภาพที่มีการใช้ช่องสัญญาณ 3 ช่องแทนค่าของแม่สี 3 สี หรืออาจจะเกิดจากจำนวนของตัวกรองในชั้นคอนโวลูชันก่อนหน้า

2) ชั้นการรวม (Pooling Layer)

หน้าที่ของชั้นการรวมคือ ลดขนาดของข้อมูลให้เหลือเฉพาะข้อมูลที่มีความสำคัญเท่านั้น ซึ่งนิยมนำมาต่อจากชั้นคอนโวลูชัน โดยทั่วไปใช้การเลือกข้อมูลที่มีค่ามากที่สุด (Max Pooling) หรือค่าเฉลี่ย (Average Pooling) มาจากแต่ละช่วงของเมตริกซ์เพื่อสร้างเป็นเมทริกซ์ใหม่ที่มีขนาดเล็กลง ดังรูปที่ 6



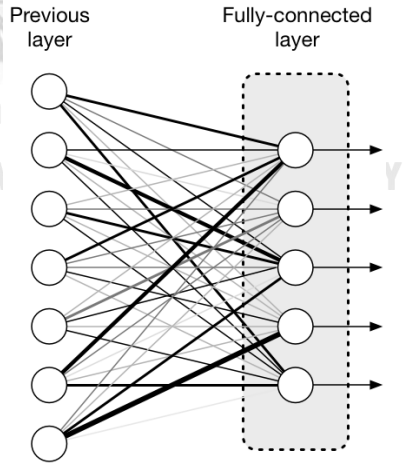
รูปที่ 6 ตัวอย่างชั้นการรวมโดยค่าที่มากที่สุดและค่าเฉลี่ย

(ที่มา http://sqlml.azurewebsites.net/2017/09/12/convolutional-neural-network/

Accessed: August 22, 2019)

3) ชั้นการเชื่อมโยงเต็มรูปแบบ (Fully Connected Layer)

เป็นการเชื่อมโยงเต็มรูปแบบ ซึ่งเป็นชั้นสุดท้ายของนิวรอลเน็ตเวิร์กคอนโวลูชัน โดยอยู่หลังจากชั้นคอนโวลูชันและชั้นการรวม โครงสร้างของชั้นการเชื่อมโยงเต็มรูปแบบแสดงดังรูปที่ 7 ประกอบด้วยชั้นย่อย ๆ ที่มีเพอร์เซ็ปตรอนอยู่จำนวนหนึ่ง โดยที่เพอร์เซ็ปตรอนทุกตัวจะมีเส้นเชื่อมกันเพอร์เซ็ปตรอนทุกตัวในชั้นก่อนหน้าและชั้นถัดไป ทำให้สามารถทำการคำนวณการป้อนไปข้างหน้าและการแพร่กระจายย้อนกลับได้ด้วยวิธีการปกติได้ ชั้นการเชื่อมโยงเต็มรูปแบบแสดงดังรูปที่ 7



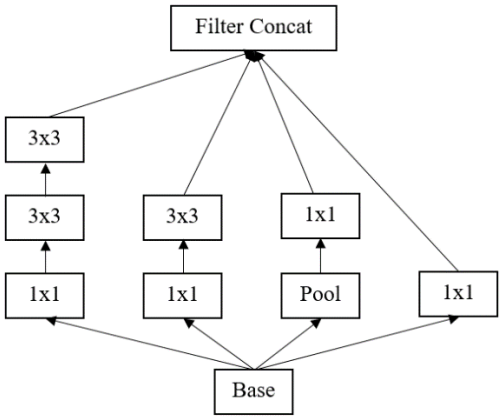
รูปที่ 7 ชั้นการเชื่อมโยงเต็มรูปแบบ

(ที่มา https://mc.ai/fully-connected-layer-with-dynamic-input-shape/

Accessed: August 22, 2019)

2.1.2 อินเซ็บชันเน็ตเวิร์ก (Inception Network)

โครงสร้างของอินเซ็บชันเน็ตเวิร์กถูกเสนอในปี 2015 ด้วยชื่อ GoogLeNet โดยอินเซ็บชันถูกพิจารณาให้เป็นเทคนิคที่ทันสมัยสำหรับโครงสร้างของการเรียนรู้เชิงลึก เพื่อนำมาใช้ในงานจำแนกประเภทรูปภาพหรือใช้ในการแก้ปัญหาการตรวจจับ แนวคิดหลักของอินเซ็บชันคือ การลดแบบจำลองที่มีขนาดใหญ่หรือมีความลึก ให้มีขนาดที่เล็กลง โดยลดชั้นของเน็ตเวิร์กหรือลดจำนวนของพารามิเตอร์ แต่ประสิทธิภาพของแบบจำลองไม่ได้ลดตาม และยังสามารถลดเวลาในการคำนวนอีกด้วย ซึ่งวิธีการนี้ เรียกว่า การแยกตัวประกอบ (Factorization) โดยอินเซ็บชันเวอร์ชันสามถูกพัฒนามาด้วยการนำวิธีการดังกล่าวมาใช้ โครงสร้างของอินเซ็บชันเวอร์ชันสาม แสดงได้ดังรูปที่ 8

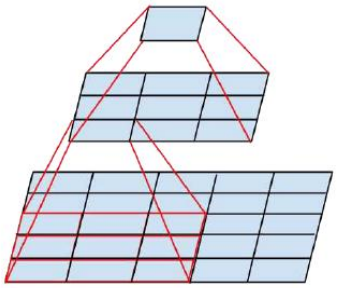


รูปที่ 8 โครงสร้างของอินเซ็บชันเวอร์ชันสาม (ที่มา: รูปที่ 5 ของ [6])

วิธีการแยกตัวประกอบ สามารถแยกย่อยเป็นสองวิธี ดังนี้

1) การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่เล็กลง

ตัวอย่างเช่น หนึ่ง 5x5 คอนโวลูชัน ถูกแทนที่ด้วยสอง 3x3 คอนโวลูชัน ซึ่งการแทนดังกล่าวสามารถลดจำนวนพารามิเตอร์ได้ถึง 28 เปอร์เซ็นต์ การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่เล็กลง แสดงได้ดังรูปที่ 9

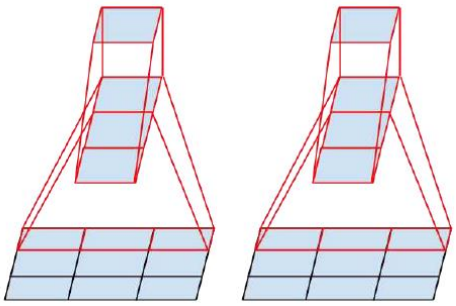


รูปที่ 9 การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่เล็กลง

(ที่มา https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c/ Accessed: August 22, 2019)

2) การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่ไม่สมมาตร

ตัวอย่างเช่น หนึ่ง 3x3 คอนโวลูชัน ถูกแทนที่ด้วย หนึ่ง 3x1 คอนโวลูชันที่ตามด้วย หนึ่ง 1x3 คอนโวลูชัน ซึ่งการแทนดังกล่าว สามารถลดจำนวนพารามิเตอร์ได้ถึง 33 เปอร์เซ็นต์ การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่ไม่สมมาตร แสดงได้ดังรูปที่ 10



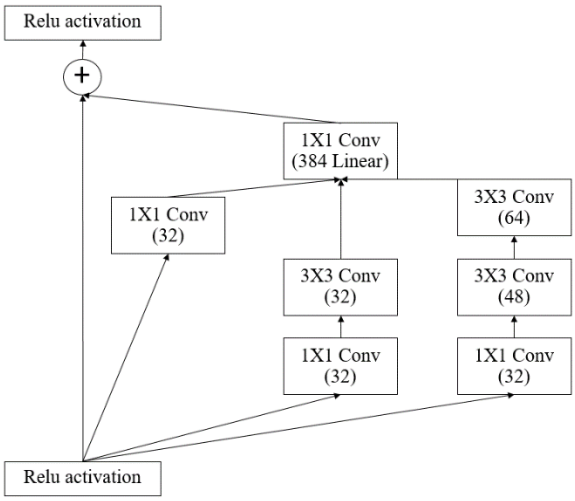
รูปที่ 10 การแยกตัวประกอบให้กลายเป็นคอนโวลูชันที่ไม่สมมาตร

(ที่มา https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c/ Accessed: August 22, 2019)

2.1.3 อินเซ็บชันเรสเน็ต (Inception-Resnet)

อินเซ็บชันเรสเน็ต [7] เป็นการรวมแนวคิดระหว่างอินเซ็บชันเน็ตเวิร์ก [6] และเรสเน็ต (ResNet) [9] โดยมีการเพิ่มการเชื่อมต่อทางลัด (shortcut connection) เพื่อข้ามชั้นคอนโวลูชันที่ไม่จำเป็นไป โดยที่ค่าถ่วงน้ำหนักบนชั้นเหล่านั้นก็จะถูกสอนให้เข้าใกล้ 0 ส่วนข้อมูลที่จำเป็นก็ยังสามารถไหลผ่านการเชื่อมต่อทางลัดไปยังชั้นต่อไปได้ แนวคิดแบบนี้ช่วยแก้ปัญหาการเกิด แวนิชชิง กราเดียน อินเซ็บชันเรสเน็ตเป็นแบบจำลองที่ประหยัดเวลาการคำนวน และยังมีประสิทธิภาพ

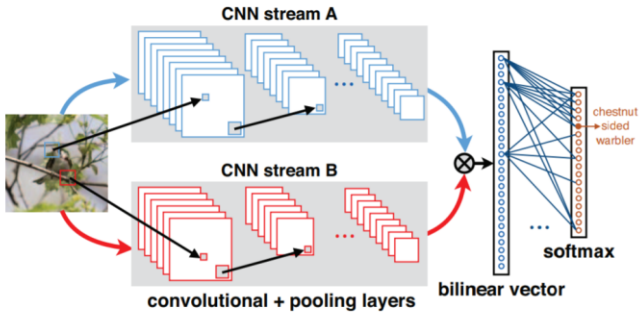
มากกว่าอินเซ็บชันเวอร์ชันสี่อีกด้วย อ้างอิงจากงานวิจัย [7] โครงสร้างของอินเซ็บเรสเน็ตชันเวอร์ชันสอง แสดงได้ดังรูปที่ 11



รูปที่ 11 โครงสร้างของอินเซ็บชันเรสเน็ตเวอร์ชันสอง (ที่มา: รูปที่ 16 ของ [6])

2.2 นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ (Bilinear Convolutional Neural Networks หรือ B-CNN)

นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ [10] คือแบบจำลองที่อาศัยสองนิวรอลเน็ตเวิร์กคอนโวลูชันเป็นตัวสกัดคุณลักษณะสำหรับการจำแนกประเภทของรูปภาพเพื่อทำให้เกิดความสัมพันธ์แบบคู่ เป็นผลให้แบบจำลองนี้สามารถเพิ่มคุณลักษณะของภาพได้มากขึ้นจากแบบจำลองอื่น ๆ โดยแนวคิดของนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่นิยมนำมาประยุกต์ใช้สำหรับงานจำแนกประเภทรูปภาพอย่างละเอียด โดยมีโครงสร้างแสดงดังรูปที่ 12



รูปที่ 12 โครงสร้างนิวรอลเน็ตเวิร์กคอนโวลูชันเชิงเส้นคู่ (ที่มา: รูปที่ 1 ของ [6])

โดยนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่มีองค์ประกอบดังสมการนี้

$$\mathcal{B} = (f_A, f_B, P, C) \tag{2}$$

กำหนดให้ f_A, f_B แทนฟังก์ชันแสดงคุณลักษณะ

P แทนฟังก์ชันชั้นการรวม

C แทนฟังก์ชันการจำแนกประเภท

ฟังก์ชันแสดงคุณลักษณะ มีสมการดังนี้

$$f : \mathcal{L} \times \mathcal{I} \rightarrow \mathbb{R}^{K \times D} \tag{3}$$

กำหนดให้ รูปภาพ $l \in \mathcal{L}$ และ ตำแหน่ง $I \in \mathcal{I}$

ผลลัพธ์ของคุณลักษณะมีขนาด $K \times D$

โดยคุณลักษณะที่ถูกสกัดจากนิวรอลเน็ตเวิร์กคอนโวลูชันทั้งสองจะถูกรวมกันในแต่ละตำแหน่ง ด้วยวิธีการคูณแบบการคูณภายนอก (Outer Product) ตามสมการดังนี้

$$\text{bilinear}(l, I, f_A, f_B) = f_A(l, I)^T f_B(l, I) \tag{4}$$

ซึ่งทั้งฟังก์ชัน f_A และ f_B ต้องมีมิติของคุณลักษณะที่เท่ากัน K มิติ เพื่อความสอดคล้องกัน ฟังก์ชันการรวม P รวมคุณลักษณะของทุกตำแหน่งในรูปภาพ โดยมีสมการดังนี้

$$\Phi(I) = \sum_{l \in \mathcal{L}} \text{bilinear}(l, I, f_A, f_B) = \sum_{l \in \mathcal{L}} f_A(l, I)^T f_B(l, I) \tag{5}$$

2.3 กลไกจุดสนใจ (Attention Mechanism หรือ Att)

หลักกการของกลไกจุดสนใจ [8] จะสร้างค่าความสนใจ (Attention) ให้กับสมาชิกแต่ละตัวในลำดับข้อมูล แล้วนำค่าที่ได้คูณกลับไปหาสมาชิกตัวนั้น ๆ เพื่อสร้างเวกเตอร์ผลลัพธ์ที่จะนำไปใช้ในนิวรอลเน็ตเวิร์กต่อไป ค่าความสนใจนี้จึงเปรียบเสมือนค่าน้ำหนักถ่วงข้อมูลแต่ละตัวในลำดับข้อมูลนั้น ถ้าข้อมูลไหนควรสนใจมาก ค่าความสนใจก็จะมากตามไปด้วย ซึ่งหากข้อมูลใดถูกให้ค่าความสนใจมาก ข้อมูลนั้นก็จะเป็นข้อมูลที่จำเป็น และสามารถเพิ่มประสิทธิภาพให้แก่แบบจำลองได้ โดยกลไกจุดสนใจสำหรับงานจำแนกประเภทของรูปภาพมีหลายแบบ ดังต่อไปนี้

2.3.1 กลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อน (Soft Attention)

กลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อนจะให้ค่าน้ำหนักแก่คุณลักษณะที่มีความสำคัญด้วยค่าที่สูง และให้ค่าน้ำหนักแก่คุณลักษณะที่ถูกไม่ค่อยมีความสำคัญด้วยค่าที่ต่ำลงมา โดยพิจารณาค่าคุณลักษณะในแต่ละตำแหน่งของภาพ โดยโครงสร้างของกลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อน

สำหรับการจำแนกประเภทของรูปภาพแสดงได้ดังรูปที่ 13 และสมการแสดงการสร้างค่าความสนใจอย่างอ่อนของรูปภาพแสดงได้ดังนี้

$$x = \sum_{i=1}^{KxK} l_i X_i \tag{6}$$

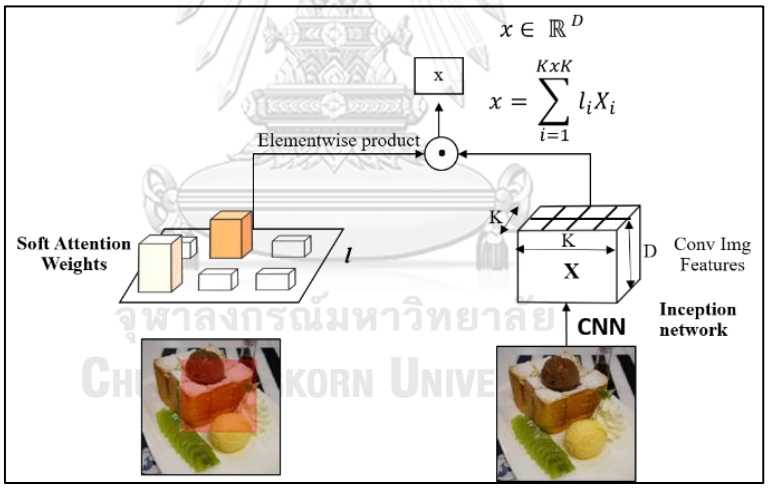
$$l_i = Softmax(Att(X_i)) \tag{7}$$

กำหนดให้ x คือ เวกเตอร์ผลลัพธ์

X_i คือ คุณลักษณะของรูปภาพที่ถูกสกัดออกมาในแต่ละตำแหน่งของรูปภาพต้นฉบับ โดยรูปภาพมีขนาด KxK มิติ

l_i คือ ค่าความน่าจะเป็นของน้ำหนักของคุณลักษณะในแต่ละตำแหน่งที่ถูกคำนวนด้วยฟังก์ชันค่าสูงสุดอย่างอ่อน (Softmax Function)

Att คือ ฟังก์ชันสำหรับคำนวณค่าความสนใจ



รูปที่ 13 โครงสร้างของกลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อนสำหรับการจำแนกประเภทของรูปภาพ (ดัดแปลงมาจาก: รูปที่ 2 ของ [11])

2.3.2 กลไกจุดสนใจที่ให้ค่าความสนใจอย่างหนัก (Hard Attention)

กลไกจุดสนใจที่ให้ค่าความสนใจอย่างหนักมีโครงสร้างคล้ายกับกลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อน แต่จะเลือกให้ค่าน้ำหนักแก่คุณลักษณะที่มีความสำคัญเพียงตำแหน่งเดียวด้วยค่า 1 และให้ค่าน้ำหนักแก่คุณลักษณะในตำแหน่งอื่นที่ไม่ถูกเลือกด้วยค่า 0 โดยโครงสร้างของกลไกจุดสนใจ

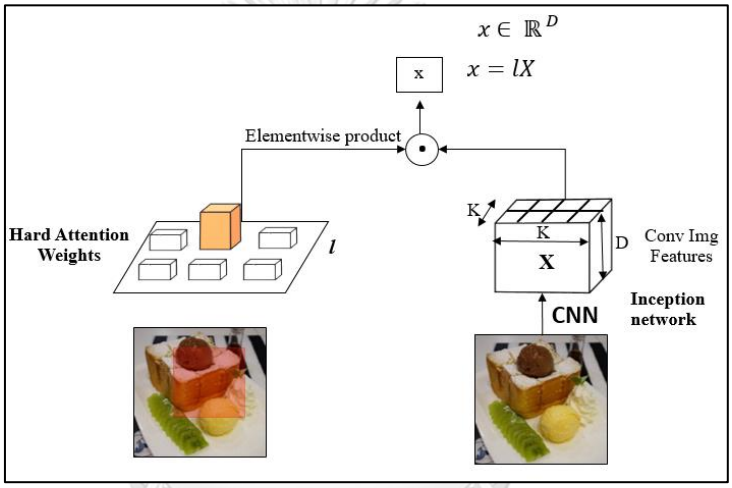
ที่ให้ค่าความสนใจอย่างหนักสำหรับการจำแนกประเภทของรูปภาพแสดงได้ดังรูปที่ 14 และสมการแสดงการสร้างค่าความสนใจอย่างหนักของรูปภาพแสดงได้ดังสมการที่ 8

$$x = lX \tag{8}$$

กำหนดให้ x คือ เวกเตอร์ผลลัพธ์

X คือ คุณลักษณะของรูปภาพที่ถูกสกัดออกมาในตำแหน่งที่มีความสำคัญมากที่สุด

l คือ ค่าน้ำหนักของรูปภาพในตำแหน่งที่มีความสำคัญมากที่สุด โดยมีค่าเป็น 1



รูปที่ 14 โครงสร้างของกลไกจุดสนใจที่ให้ค่าความสนใจอย่างหนักสำหรับการจำแนกประเภทของรูปภาพ (ดัดแปลงมาจาก: รูปที่ 2 ของ [11])

2.3.3 กลไกจุดสนใจที่ให้ค่าความสนใจตกค้าง (Residual Attention)

กลไกจุดสนใจที่ให้ค่าความสนใจตกค้างถูกสร้างขึ้นจากการซ้อนกันของกลไกจุดสนใจในหลายส่วน แต่ละส่วนของกลไกจุดสนใจจะถูกแยกออกเป็นสองสาขา คือ สาขาหน้ากาก (Mark Branch) และสาขาสายผ่าน (Trunk Branch) โดยสาขาสายผ่านเป็นส่วนที่อยู่ด้านบนของกลไกจุดสนใจ ที่จะทำการประมวลผลคุณลักษณะ ทำการกระตุ้นบล็อกของเรสเน็ต หรือบล็อกอื่น ๆ ก่อน ด้วยข้อมูลนำเข้า x และข้อมูลส่งออก T(x) อีกทั้งยังสามารถนำไปประยุกต์ใช้กับโครงสร้างเน็ตเวิร์กอื่น ๆ ที่เป็นที่นิยมได้อีกด้วย สาขาสายผ่านเป็นสาขาที่มีโครงสร้างจากบนลงล่างเพื่อเรียนรู้หน้ากากขนาดเดียวกัน M(x) โดย M(x) เปรียบเสมือนตัวควบคุมนิวรอลเน็ตเวิร์กหลัก ข้อมูลส่งออกของกลไกจุดสนใจแทนด้วย H ซึ่งมีสมการดังต่อไปนี้

$$H_{i,c}(x) = M_{i,c}(x) * T_{i,c}(x) \quad (9)$$

โดย i คือตำแหน่งของรูปภาพ และ c คือดัชนีของช่องโดยเริ่มจาก 1 ถึง C

กลไกจุดสนใจในสาขาหน้ากากสามารถเลือกคุณสมบัติในระหว่างการอนุมานไปข้างหน้า และสามารถปรับเกรเดียนในระหว่างการแพร่กระจายย้อนกลับ ดังนั้นกลไกจุดสนใจในสาขาหน้ากากสามารถทนต่อข้อมูลรบกวน และสามารถป้องกันเกรเดียนที่ไม่เหมาะสมจากข้อมูลรบกวน เพื่อที่จะสามารถปรับพารามิเตอร์ของสาขาสายผ่าน โดยเกรเดียนของสาขาหน้ากากสำหรับข้อมูลนำเข้ามีสมการดังนี้

$$\frac{\partial M(x,\theta)T(x,\emptyset)}{\partial \emptyset} = M(x,\theta)\frac{\partial T(x,\emptyset)}{\partial \emptyset} \quad (10)$$

กำหนดให้ θ แทนพารามิเตอร์ของสาขาหน้ากาก

∂ แทนพารามิเตอร์ของสาขาสายผ่าน

2.3.4 กลไกจุดสนใจที่ให้ค่าความสนใจแบบหลายหัว (Multi-head Attention)

กลไกจุดสนใจแบบหลายหัว [8] เป็นกลไกจุดสนใจที่โฟกัสไปที่การสกัดหาคุณลักษณะเฉพาะที่ในหลาย ๆ หน่วยข้อมูล ณ ตำแหน่งที่แตกต่างกัน แตกต่างจากการปรับค่าจุดสนใจเชิงสเกลาร์ (Scaled Dot-Product Attention) ที่นำ คีย์ (Key หรือ k) มาคูณกันเชิงสเกลาร์ (dot product หรือ •)กับ คิวรี (Query หรือ q) เพื่อหาตำแหน่งที่ควรจะให้ความสนใจ และนำค่าแวลลู (Value หรือ v) ที่ได้มาใช้ ข้อจำกัดของการปรับค่าจุดสนใจเชิงสเกลาร์คือ ค่าตัวแทนที่ได้จากการคำนวณนั้นอาจไม่ใช่ค่าที่เหมาะสมเพื่อนำไปใช้ในกลไกจุดสนใจ และค่าของในแต่ละหน่วยข้อมูลจะมีปฏิสัมพันธ์กันได้เพียงแบบเดียว ซึ่งไม่เพียงพอสำหรับงานทั่วไป ที่หน่วยต่างๆ ของข้อมูลสามารถมีความสัมพันธ์กันได้หลายแบบ เพื่อพัฒนาจุดบกพร่องที่กล่าวมานั้น จึงมีการนำค่าน้ำหนักมาใช้ โดยจะทำการแบ่ง k, q, v ออกเป็นหลาย ๆ ชุด แล้วทำการคูณค่าน้ำหนักที่แตกต่างกันกลับเข้าไปใน k, q, v ในแต่ละชุด เพื่อที่สามารถหาค่าตัวแทนได้อย่างเหมาะสมมากขึ้น จากนั้นจึงคำนวนค่าความสนใจในแต่ละชุดก่อนที่จะรวมทุกชุดเข้าด้วยกันเป็นข้อมูลเพื่อส่งไปประมวลผลต่อไป โดยค่าความสนใจชุดนึงจะเรียกว่าหัว และกลไกจุดสนใจที่ให้ค่าความสนใจแบบหลายหัว คือการนำทุกหัวมารวมกัน โดยโครงสร้างของกลไกจุดสนใจที่ให้ค่าความสนใจแบบหลายหัวแสดงได้ดังรูปที่ 15 และสมการแสดงการสร้างค่าความสนใจแบบหลายหัวแสดงได้ดังสมการต่อไปนี้

$$score(h_i) = softmax(\frac{q_{i \bullet k_i^T}}{\sqrt{d_k}}) \quad (11)$$

$$C_i = \sum_{i=1}^{n} v_i score(h_i) \quad (12)$$

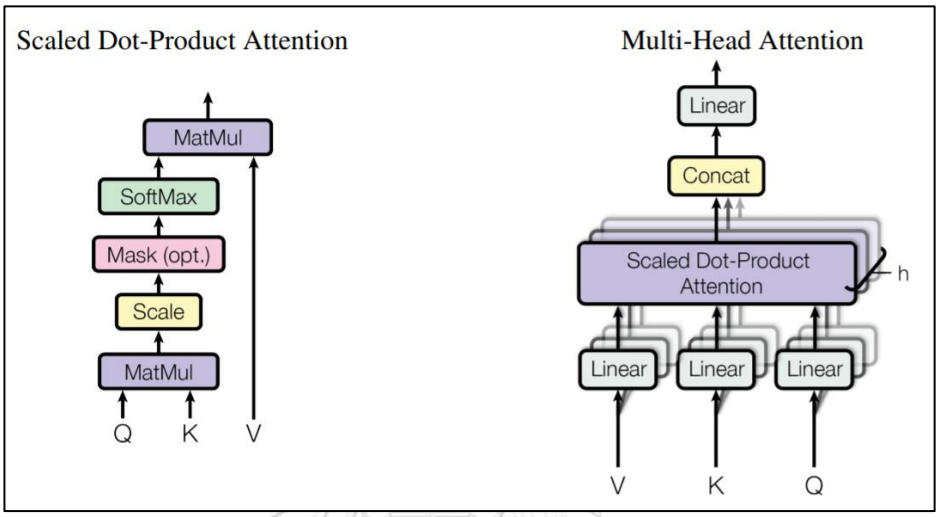
$$h_{out} = [c_1 ; c_2 ; \dots ; c_h] \quad (13)$$

กำหนดให้ d_k คือ มิติของแต่ละค่า k, q, v

$score(h_i)$ คือ ค่าความสนใจของแต่ละหัว

C_i คือ ค่าน้ำหนักของแต่ละแวลลู

h_{out} คือ ผลลัพธ์ของกลไกจุดสนใจแบบหลายหัว



รูปที่ 15 โครงสร้างของกลไกจุดสนใจที่ให้ค่าความสนใจแบบหลายหัว สำหรับการจำแนกประเภทของรูปภาพ

[ที่มา https://towardsdatascience.com/attention-and-its-different-forms-7fc3674d14dc Accessed: November 20, 2019]

บทที่ 3
งานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ ซึ่งคืองานวิจัยกลไกจุดสนใจแบบเน็ตเวิร์กละเอียด สำหรับการจำแนกประเภทของรูปภาพอาหาร โดยงานวิจัยแต่ละงานจะมีโครงสร้างแบบจำลอง และวัตถุประสงค์ที่แตกต่างกัน โดยในหัวข้อนี้จะแบ่งงานวิจัยออกเป็น 2 กลุ่ม ได้แก่ (1) แบบจำลองการจำแนกประเภทของรูปภาพแบบเน็ตเวิร์กละเอียดโดยใช้การเรียนรู้เชิงลึก (2) แบบจำลองการจำแนกประเภทของรูปภาพแบบเน็ตเวิร์กละเอียดโดยใช้นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ ในงานวิจัยนี้ได้เลือกใช้แบบจำลองการจำแนกประเภทของรูปภาพแบบเน็ตเวิร์กละเอียดโดยใช้นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ เนื่องจากเป็นแบบจำลองที่ได้รับความนิยมและให้ประสิทธิภาพที่แม่นยำมากกว่า ในงานของการจำแนกประเภทของรูปภาพอาหาร

3.1 แบบจำลองการจำแนกประเภทของรูปภาพแบบเน็ตเวิร์กละเอียดโดยใช้การเรียนรู้เชิงลึก

งานวิจัยของ Singla และคณะ [1] ได้เสนอแบบจำลองคอนโวลูชันเน็ตเวิร์กที่ใช้การเรียนรู้แบบส่งต่อ คือ กูเกิลเน็ต หรืออินเซ็บชันเวอร์ชันสาม โดยการวิจัยถูกแบ่งออกเป็นสองส่วน คือการจำแนกรูปภาพที่เป็นภาพอาหาร และรูปภาพที่ไม่ได้เป็นภาพอาหาร และการจำแนกประเภทของรูปภาพอาหารบนชุดข้อมูลรูปภาพอาหารสาธารณะ ผลการทดลองพบว่าการทดลองในส่วนของการจำแนกรูปภาพที่เป็นภาพอาหาร และรูปภาพที่ไม่ได้เป็นภาพอาหาร มีความแม่นยำที่สูง แต่การจำแนกประเภทของรูปภาพอาหาร มีความแม่นยำไม่สูงมาก เนื่องจากรูปภาพอาหารบางประเภทมีส่วนผสมของอาหารประเภทอื่น ๆ รวมอยู่ด้วย ทำให้ภาพนั้นเกิดความซับซ้อนในการจำแนกประเภทของรูปภาพอาหารที่มีความคล้ายคลึงกัน

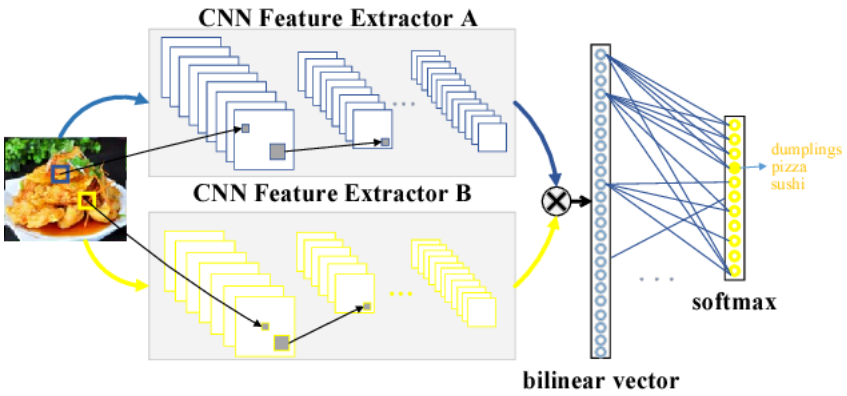
3.2 แบบจำลองการจำแนกประเภทของรูปภาพแบบเน็ตเวิร์กละเอียดโดยใช้นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่

นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ คือแบบจำลองที่ทันสมัย สำหรับงานการจำแนกประเภทของรูปภาพอย่างละเอียด ในปี 2543 Tenenbaum และ Freeman [12] ได้นำเสนอนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่เป็นครั้งแรก เพื่อจำลองสองปัจจัยของรูปภาพ โดยแบ่งเป็น สไตล์และเนื้อหา ในแนวคิดที่ว่าระบบการรับรู้พื้นฐานจะสามารถจดจำสิ่งต่าง ๆ ได้จากปัจจัยสองสิ่งที่แตกต่างกัน การใช้แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ สามารถทำให้ชุดข้อมูลเรียนรู้ เรียนรู้ถึงอิทธิพลของสไตล์ และเนื้อหา ว่าสามารถถูกแยกออกจากกันได้อย่างมีประสิทธิภาพ

นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่แบบพื้นฐานสำหรับการจำแนกประเภทของรูปภาพถูกเสนอโดย Lin และคณะ [10] แบบจำลองในงานวิจัยดังกล่าวได้นำสองคอนโวลูชันเน็ตเวิร์ก ทั้งแบบ

M-Net [13] และD-Net [14] มาคูณกันแบบการคูณภายนอกบนชุดข้อมูลสาธารณะ เพื่อสร้างมิติของคุณลักษณะรูปภาพที่หลากหลายที่สามารถนำไปคำนวณค่าเกรเดียนของโดเมน ต่อมาในปี 2560 Lin และคณะ [15] ได้เพิ่มเมทริกซ์การทำนอร์มอลไลซ์เพื่อเพิ่มความแม่นยำจากงานวิจัยก่อนหน้า แต่อย่างไรก็ตามนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่แบบดั้งเดิม ก็ยังมีปัญหาในชั้นการรวม ที่มีขนาดของคุณลักษณะใหญ่เกินไป จากปัญหาดังกล่าวทำให้มีงานวิจัยหลายงาน พยายามแก้ปัญหา Kong และคณะ [16] ได้นำเสนอชั้นรวมเชิงคู่ระดับต่ำ (Low-Rank Bilinear Pooling หรือ LRBP) โดยใช้โมดูลเคอร์เนลเพื่อลดขนาดคุณลักษณะและปรับปรุงการคำนวณ Gao และคณะ [17] เสนอชั้นรวมเชิงคู่ที่กระชับ (Compact Bilinear Polling หรือ CBP) เพื่อลดขนาดของมิติ ด้วยการประยุกต์การฉายภาพร่างเทนเซอร์(Tensor Sketch Projection) และการฉายแบบสุ่ม (Random Maclaurin projection) Yin และคณะ [18] เสนอวิธีใหม่ของเคอร์เนลเชิงลึกเพื่อจัดการการปฏิสัมพันธ์ของคุณลักษณะลำดับสูงกว่า ในขณะที่ Cai และคณะ [19] ได้เสนอตัวทำนายพหุนามเพื่อใช้ประโยชน์จากค่าสถิติที่มีลำดับสูงขึ้น Yu และคณะ [20] เสนอชั้นรวมเชิงคู่แบบลำดับชั้น โดยรวมการปฏิสัมพันธ์ระหว่างชั้น และการเลือกคุณลักษณะการเรียนรู้ เพื่อรวมคุณลักษณะหลายชั้น สำหรับการปรับปรุงการรู้จำอย่างละเอียด

แนวคิดของนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ถูกนำมาใช้ในงานการจำแนกประเภทของรูปภาพในหลายสาขา ตัวอย่างเช่น นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ประสบความสำเร็จในงานการรู้จำใบหน้า [21] การระบุตัวตนของบุคคล [22] และการจำแนกภาพทางจุลพยาธิวิทยา [23] นอกจากนั้นแบบจำลองดังกล่าวยังถูกนำไปใช้ในงานวิจัยการจำแนกประเภทของรูปภาพอาหาร [5] อีกด้วย โดยในงานวิจัยนี้ ได้ปรับปรุงแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่แบบดั้งเดิม โดยการใช้คอนโวลูชันเน็ตเวิร์กแบบสอนก่อน บนชุดข้อมูล FOOD-101 UECFOOD-100 และ UECFOOD-256 เป็นตัวสกัดคุณลักษณะ ผลการทดลองพบว่าการนำคอนโวลูชันเน็ตเวิร์กแบบสอนก่อนเข้ามาเป็นตัวสกัดคุณลักษณะ ทำให้แบบจำลองมีประสิทธิภาพดีขึ้นในการจำแนกประเภทของรูปภาพอาหาร โครงสร้างของแบบจำลองแสดงดังรูปที่ 16



รูปที่ 16 นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่

(อ้างอิงจาก Fig.1 ใน [5])

3.3 ประเด็นที่พบจากงานวิจัยก่อนหน้าและสิ่งที่นำมาปรับปรุงในงานวิจัยนี้

ในหัวข้อนี้จะเป็นการสรุปประเด็นที่พบจากงานวิจัยก่อนหน้าทั้งสองกลุ่ม รวมทั้งแนวทางในการพัฒนาต่อ โดยมีรายละเอียด ดังต่อไปนี้

3.3.1 แบบจำลองคอนโวลูชันเน็ตเวิร์ก

สำหรับงานวิจัยที่ใช้แบบจำลองคอนโวลูชันเน็ตเวิร์ก การเลือกใช้เน็ตเวิร์กเพื่อสกัดคุณลักษณะของภาพถือว่าเป็นสิ่งสำคัญที่จะทำให้แบบจำลองสามารถจำแนกประเภทรูปภาพได้อย่างแม่นยำ โดยในงานวิจัยที่เกี่ยวข้องนั้น ได้เลือกแบบจำลองอินเซ็บชันเน็ตเวิร์กเวอร์ชันสาม ซึ่งเป็นเน็ตเวิร์กที่ใช้การเรียนรู้แบบส่งต่อ ผลการทดลองพบว่า แบบจำลองสามารถจำแนกประเภทรูปภาพที่เป็นอาหาร และรูปภาพที่ไม่ใช่อาหารด้วยความแม่นยำสูง แต่สำหรับรูปภาพที่เป็นประเภทอาหาร กลับได้ความแม่นยำที่ไม่สูงมาก เนื่องจากรูปภาพอาหารในแต่ละประเภท มีความคล้ายคลึงกันสูงทำให้ยากต่อการจำแนก และแบบจำลองที่ใช้ ไม่ได้นำคุณลักษณะที่ได้จากแต่ละประเภทอาหารมาทำการเรียนรู้อย่างมีความสัมพันธ์กัน จึงทำให้แบบจำลองยังไม่สามารถจำแนกรูปภาพที่มีลักษณะใกล้เคียงกันได้ดีเท่าที่ควร

แนวทางการพัฒนาแบบจำลองที่ได้จากงานวิจัยดังกล่าวคือ การปรับปรุงเน็ตเวิร์กที่ใช้สกัดคุณลักษะ ให้เป็นเน็ตเวิร์กที่มีประสิทธิภาพสูงในการจำแนกประเภทรูปภาพมากขึ้น ด้วยการใช้เน็ตเวิร์กที่ทันสมัยในปัจจุบัน และทำการปรับปรุงแบบจำลองให้สามารถจำแนกรูปภาพที่มีลักษณะใกล้เคียงให้ดีขึ้น ด้วยการใช้แบบจำลองที่เป็นนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่

3.3.2 แบบจำลองคอนโวลูชันเน็ตเวิร์กเชิงเส้นคู่

สำหรับงานวิจัยที่ใช้แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ สามารถสกัดคุณลักษณะของรูปภาพอาหารได้มากขึ้น และสามารถจำแนกประเภทของรูปภาพอาหารที่มีลักษณะที่ใกล้เคียงกันได้ดี เนื่องจากตัวแบบจำลองได้มุ่งเน้นไปที่การหาความสัมพันธ์ของคุณลักษณะในแต่ละตำแหน่ง จึงทำให้แบบจำลองสามารถสร้างคุณลักษณะที่หลากหลายมากขึ้นกว่าแบบจำลองก่อน ๆ แต่เนื่องจากคุณลักษณะที่สกัดได้มีปริมาณมาก ซึ่งบางคุณลักษณะอาจไม่จำเป็นต่อการจำแนกประเภทรูปภาพนั้น ๆ จึงทำให้คุณลักษณะที่สกัดมานั้น ยังไม่ใช้คุณลักษณะที่จำเพาะ อีกทั้งเน็ตเวิร์กที่ใช้ในส่วนการสกัดคุณลักษณะ ยังไม่ใช่เน็ตเวิร์กที่ทันสมัยเท่าที่ควร และมีความลึกของชั้นเน็ตเวิร์กมาก จึงทำให้เสียเวลาในการประมวลผล

แนวทางการพัฒนาแบบจำลองที่ได้จากงานวิจัยดังกล่าวคือ การนำกลไกจุดสนใจมาใช้เพื่อให้แบบจำลองสามารถคัดเลือกเฉพาะคุณลักษณะที่เป็นจุดสังเกตุของรูปภาพอาหารแต่ละประเภท และปรับปรุงเน็ตเวิร์กสำหรับการสกัดคุณลักษณะด้วยวิธีการต่าง ๆ เพื่อให้แบบจำลองที่ได้ มีประสิทธิภาพที่ดีขึ้น

บทที่ 4
แนวคิดในการดำเนินงาน และแบบจำลองที่นำเสนอ

ในบทนี้จะทำการสร้างแบบจำลองเพื่อจำแนกประเภทของรูปภาพอาหาร โดยได้นำแนวคิดของนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ถูกเสนอโดย Lin และคณะ [10] ซึ่งเป็นแบบจำลองที่นิยมนำมาใช้สำหรับงานการจำแนกประเภทของรูปภาพอย่างละเอียด งานวิจัยนี้ได้ปรับปรุงนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ โดยการใช้การรวมกันของคอนโวลูชันเน็ตเวิร์กที่มีประสิทธิภาพดีในงานการจำแนกประเภทของรูปภาพ เป็นตัวสกัดคุณลักษณะของรูปภาพ และได้เพิ่มกลไลจุดสนใจร่วมด้วย เพื่อสกัดคุณลักษณะที่สำคัญและเป็นลักษณะเด่นของรูปภาพนั้น ๆ โดยวิธีการที่นำเสนอถูกแบ่งออกเป็น 2 ส่วน ประกอบด้วย (1) อธิบายรายละเอียดและการจัดเตรียมชุดข้อมูลรูปภาพอาหารที่ใช้ในงานวิจัยนี้ (2) แสดงวิธีการที่ใช้ในการจำแนกประเภทของรูปภาพอาหารด้วยแบบจำลองที่งานวิจัยนี้นำเสนอ

4.1 การเตรียมข้อมูล

ในหัวข้อนี้จะกล่าวถึงการเตรียมข้อมูลเพื่อนำไปใช้เป็นข้อมูลรับเข้าของแบบจำลองที่ใช้จำแนกประเภทของรูปภาพอาหาร โดยชุดข้อมูลที่ถูกนำมาใช้ในงานวิจัยนี้ คือชุดข้อมูลประเภทรูปภาพจากเว็บไซต์และแอปพลิเคชัน Wongnai ซึ่งเป็นแอปพลิเคชันสำหรับการค้นหาอาหารอันดับ 1 ของไทยที่มีการรวบรวมข้อมูลเกี่ยวกับอาหารมากที่สุด ในปัจจุบัน Wongnai เป็นผู้นำตลาดระบบรีวิวร้านอาหารในประเทศไทย โดยมีฐานข้อมูลมากกว่า 230,000 ร้านทั่วประเทศไทย ตัวอย่างชุดข้อมูลแสดงในรูปที่ 17

รูปที่ 17 รูปภาพตัวอย่างชุดข้อมูลจาก Wongnai

ชุดข้อมูลที่นำมานั้น มีจำนวนรูปภาพทั้งหมด 144,164 รูป ขนาดรูปภาพจากชุดข้อมูลมีขนาดความกว้าง ระหว่าง 700-1000 พิกเซล (Pixel) และมีขนาดความสูงระหว่าง 1000-1200 พิกเซล โดยได้เลือกรูปภาพที่ถูกอัปโหลดในช่วง 5 ปีที่ผ่านมาถึงปีปัจจุบัน ที่อยู่ในช่วงปี พ.ศ. 2558-2562 ซึ่งชุดข้อมูลนี้มีทั้งรูปภาพอาหาร ขนม และเครื่องดื่ม ที่ถูกแบ่งเป็นประเภทย่อย ๆ อีกหลากหลายประเภท โดยรายละเอียดของการจัดเตรียมข้อมูลมีดังต่อไปนี้

4.1.1 การปรับความละเอียดของรูปภาพ

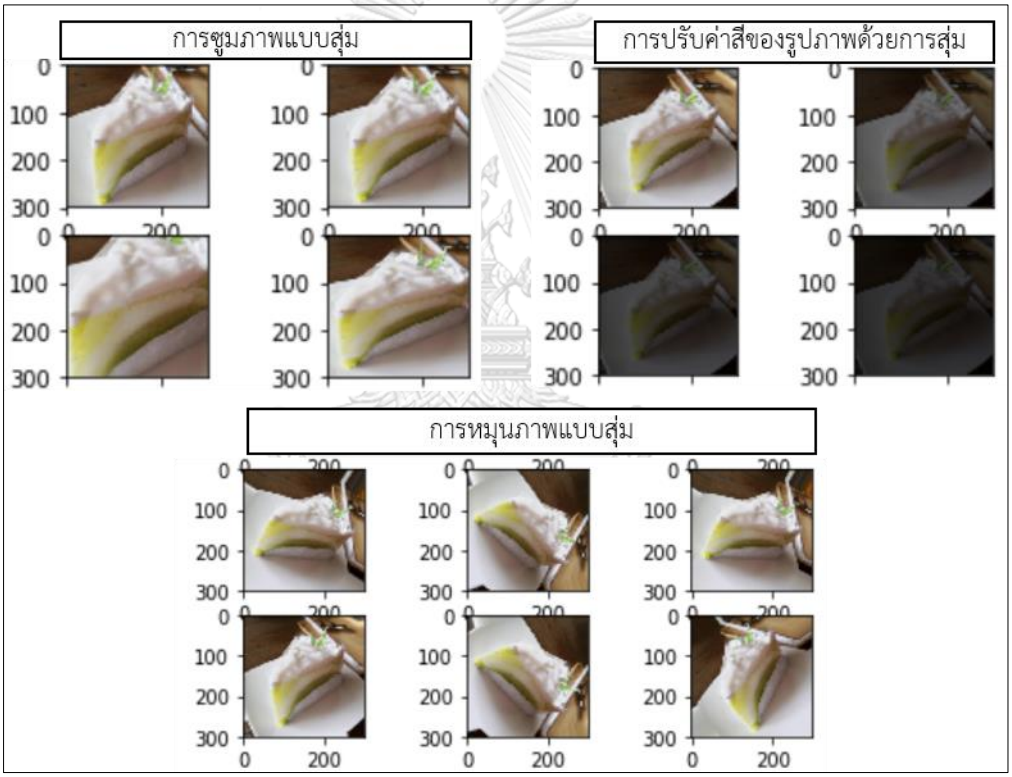
เนื่องจากชุดข้อมูลประเภทรูปภาพจากเว็บไซต์และแอปพลิเคชัน Wongnai ที่ได้รับมานั้น มีความละเอียดของรูปภาพที่แตกต่างกัน ดังนั้นจึงต้องมีการปรับขนาดของภาพให้มีความละเอียดเท่ากันก่อนนำไปใช้ในการทดลอง โดยการปรับความละเอียดของรูปภาพในงานวิจัยใช้วิธีการที่เรียกว่า Lanczos Resampling [24] ซึ่งเป็นวิธีการสำหรับปรับขนาดของภาพที่สามารถรักษาคุณภาพของภาพ รวมถึงความคมชัดได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ได้ทำการปรับความละเอียดให้มีขนาดที่เล็กลง (Downsampling) โดยได้ทำการปรับความละเอียดของภาพออกเป็นสองชุด เพื่อใช้ทำการทดลองเปรียบเทียบเรื่องผลของความละเอียดของรูปภาพ ที่ส่งผลต่อการจำแนกประเภทของรูปภาพอาหาร ในชุดที่ 1 จะใช้ภาพที่มีความละเอียด 299x299 พิกเซล เพื่อให้ขนาดของภาพสอดคล้องกับขนาดของชุดข้อมูลที่ถูกสอนมาแล้วในคอนโวลูชันเน็ตเวิร์กที่นำมาใช้ และในชุดที่สอง จะใช้ภาพที่มีความละเอียด 450x450 พิกเซล

4.1.2 การประมวลผลก่อน

ในส่วนของการประมวลผลก่อนนั้น ได้นำแนวคิดของการสร้างภาพใหม่ โดยการดัดแปลงภาพเดิมที่มี (Data Augmentation) ซึ่งวิธีดังกล่าว เป็นการช่วยเพิ่มจำนวนรูปภาพให้มีหลายมุมมอง ทำให้แบบจำลองสามารถจดจำลักษณะที่สำคัญของรูปภาพนั้น ๆ ได้ และยังเป็นการเพิ่มประสิทธิภาพในการเรียนรู้ของแบบจำลองอีกด้วย โดยในงานวิจัยนี้ได้เลือกวิธีการดังต่อไปนี้ในการดัดแปลงภาพเดิมที่มี ตัวอย่างการดัดแปลงภาพเดิมที่มีแสดงในรูปที่ 18

ตัวอย่างวิธีการของการดัดแปลงภาพเดิมที่มี

- พลิกและบิดเบือนภาพด้วยการสุ่ม
- เลื่อนภาพในแนวนอนและแนวตั้งด้วยการสุ่ม
- ครอบตัดภาพด้วยการสุ่ม
- ตั้งค่าช่วงการเปลี่ยนแชนเนลด้วยการสุ่ม
- ปรับค่าสีของรูปภาพด้วยการสุ่ม
- ซูมภาพแบบสุ่ม
- หมุนภาพแบบสุ่ม

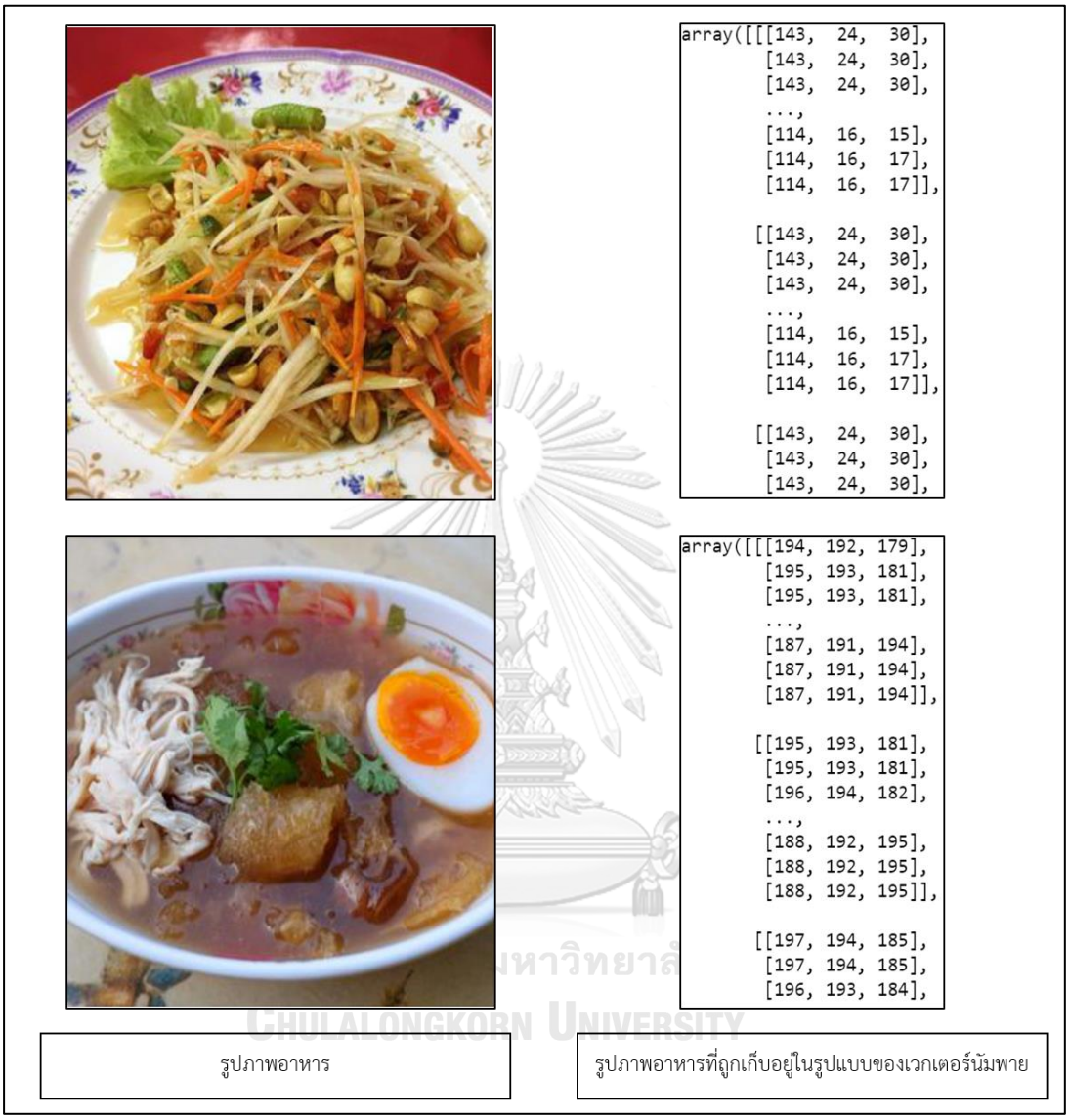


รูปที่ 18 ตัวอย่างการดัดแปลงภาพเดิมที่มี

4.1.2 การแปลงชุดข้อมูลรูปภาพอาหารให้เก็บอยู่ในรูปเวกเตอร์นัมพาย

สำหรับการแปลงชุดข้อมูลรูปภาพอาหารให้อยู่ในรูปแบบเวกเตอร์นัมพาย ซึ่งเป็นหนึ่งในไลบรารี่ของภาษาไพธอนสำหรับใช้คำนวณทางคณิตศาสตร์ มีวัตถุประสงค์เพื่อใช้ในการลดเวลาในการเปิดไฟล์เมื่อนำชุดข้อมูลเข้าเน็ตเวิร์กสำหรับฝึกสอน อีกทั้งยังสามารถลดขั้นตอนแบบเดิมซึ่งเป็น

การเปิดภาพทีละภาพสำหรับเวลาส่งไปฝึกสอนที่เน็ตเวิร์ก ตัวอย่างของการแปลงรูปภาพอาหารให้อยู่ในรูปเวกเตอร์นัมพาย แสดงในรูปที่ 19

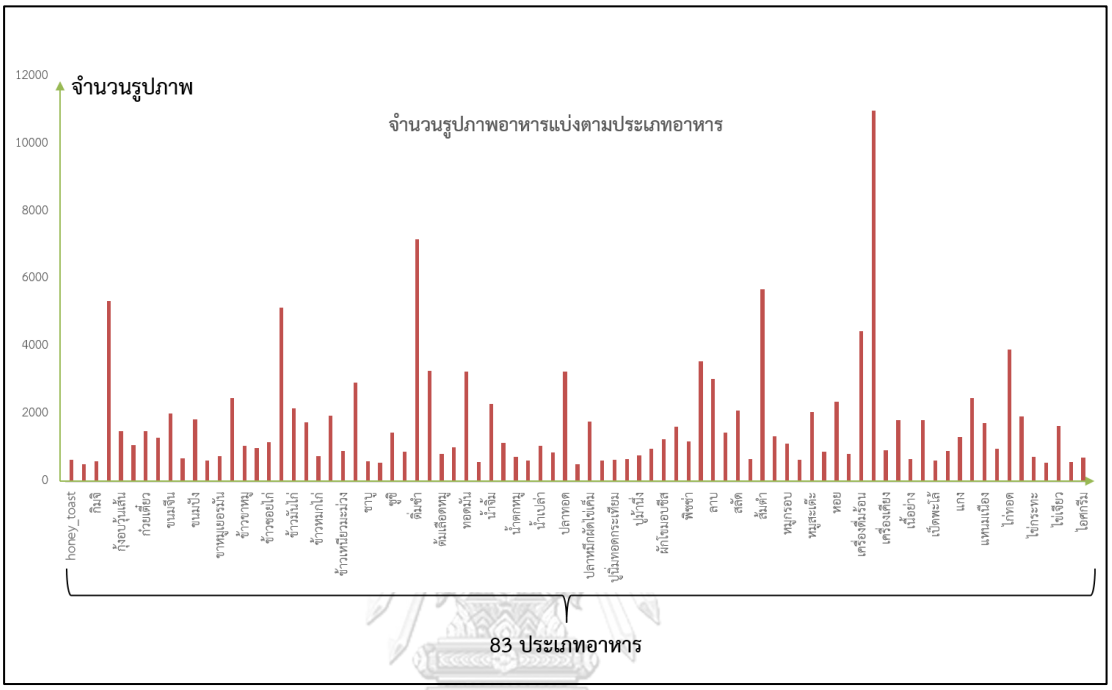


รูปที่ 19 ตัวอย่างของการแปลงรูปภาพอาหารให้อยู่ในรูปเวกเตอร์นัมพาย

4.1.3 การจัดการกับข้อมูลที่ไม่สมดุล

เนื่องจากในชุดข้อมูลสำหรับทำการฝึกสอน มีจำนวนรูปภาพในแต่ละประเภทของอาหารไม่เท่ากัน อาจทำให้แบบจำลองไม่สามารถทำการสอนได้อย่างมีประสิทธิภาพเท่าที่ควร โดยแบบจำลองอาจมีความลำเอียงต่อประเภทอาหารที่มีจำนวนรูปภาพน้อย เพื่อแก้ไขปัญหาข้อมูลที่ไม่สมดุล จึงต้องทำการคำนวณเพื่อประเมินค่าน้ำหนักของข้อมูลในแต่ละประเภท และใช้พารามิเตอร์การปรับค่าน้ำหนัก (Class Weight) ในขณะที่ทำการสอนแบบจำลอง ด้วยการรวมน้ำหนักของรูปภาพอาหารแต่

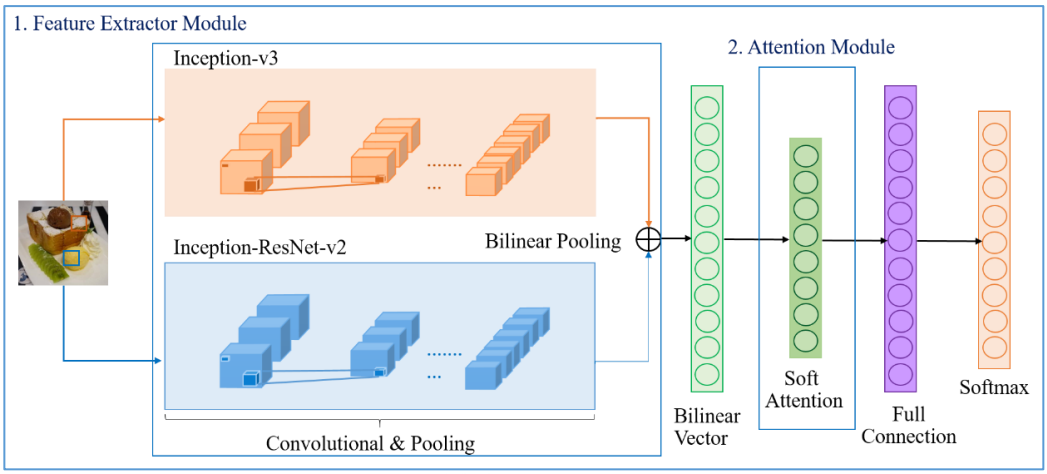
ละประเภทเข้ากับฟังก์ชันต้นทุน (Cost Function) โดยเลือกให้น้ำหนักที่สูงขึ้นกับประเภทอาหารที่มีรูปภาพน้อยกว่า และลดน้ำหนักให้กับประเภทอาหารที่มีรูปภาพมากกว่า กราฟแท่งแสดงจำนวนรูปภาพอาหารแบ่งตามประเภทอาหารของชุดข้อมูลสำหรับทำการฝึกสอน แสดงได้ดังรูปที่ 20



รูปที่ 20 กราฟแท่งแสดงจำนวนรูปภาพอาหารแบ่งตามประเภทอาหาร ของชุดข้อมูลสำหรับทำการฝึกสอน

4.2 แบบจำลองสำหรับการจำแนกประเภทของรูปภาพอาหาร

งานวิจัยนี้ได้ทำการปรับปรุงแบบจำลองที่นำเสนอในงานวิจัย [5] โดยทำการปรับปรุงนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ แบบจำลองในงานวิจัยนี้จะถูกแบ่งออกเป็นสองส่วน คือ (1) ส่วนการสกัดคุณลักษณะของรูปภาพ (2) และส่วนของกลไกจุดสนใจ โครงสร้างของแบบจำลองที่นำเสนอมีชื่อว่า แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีกลไกจุดสนใจ แบบจำลองที่นำเสนอแสดงดังรูปที่ 21



รูปที่ 21 แบบจำลองที่นำเสนอ
(แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีกลไกจุดสนใจ)

4.2.1 ส่วนการสกัดคุณลักษณะ

ตัวสกัดคุณลักษณะของรูปภาพนับว่าเป็นส่วนที่สำคัญในงานจำแนกประเภทรูปภาพ เนื่องจากคุณลักษณะของรูปภาพที่แตกต่างกันสามารถส่งผลต่อการจำแนกประเภทของรูปภาพของแบบจำลองได้ โดยงานวิจัยนี้ได้ปรับปรุงตัวสกัดคุณลักษณะของรูปภาพให้เป็นคอนโวลูชันเน็ตเวิร์กที่มีประสิทธิภาพมากขึ้นสำหรับงานการจำแนกประเภทของรูปภาพ ซึ่งคอนโวลูชันเน็ตเวิร์กดังกล่าวเป็นเน็ตเวิร์กที่ใช้การเรียนรู้แบบส่งต่อ (Transfer Learning) ซึ่งเป็นการนำองค์ความรู้จากสิ่งที่เคยถูกแก้ปัญหาได้แล้วมาในงานก่อนหน้าไปใช้ในโจทย์ที่ใกล้เคียงกัน ในงานวิจัยนี้ ได้ใช้คอนโวลูชันเน็ตเวิร์กสองตัวที่แตกต่างกันทั้งเน็ตเวิร์กชั้นบน และเน็ตเวิร์กชั้นล่างตามรูปที่ 21 มาเป็นตัวสกัดคุณลักษณะของโครงสร้างนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ โดยคอนโวลูชันเน็ตเวิร์กทั้งสองตัวจะถูกรวมกันในแต่ละตำแหน่ง ด้วยวิธีการคูณแบบการคูณภายนอก การรวมกันด้วยวิธีนี้ส่งผลให้คุณลักษณะที่ถูกสกัดได้จากแต่ละเน็ตเวิร์กมีความสัมพันธ์กันในทุกตำแหน่ง ส่งผลให้แบบจำลองนี้สามารถสกัดคุณลักษณะของรูปภาพได้อย่างแตกต่างและหลากหลาย

4.2.2 ส่วนของกลไกจุดสนใจ

งานวิจัยนี้ ได้นำโครงสร้างของกลไกจุดสนใจมารวมกับนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ โดยกลไกจุดสนใจจะถูกนำมาต่อหลังเวกเตอร์ของคุณลักษณะที่ถูกสกัดมาจากการรวมกันของสองคอนโวลูชันเน็ตเวิร์ก เนื่องจากคุณลักษณะที่ถูกสกัดมานั้นมีหลากหลาย และคุณลักษณะบางตัวอาจไม่จำเป็นต่อรูปภาพอาหารบางประเภท การเพิ่มกลไกจุดสนใจจึงมีความสำคัญในการมุ่งเน้นการ

คัดเลือกคุณลักษณะที่มีความเด่นชัดต่อการจำแนกรูปภาพนั้น ๆ โดยงานวิจัยนี้ จะเริ่มทดลองจากกลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อนก่อน จากนั้นจึงนำไปเปรียบเทียบกับกลไกจุดสนใจแบบอื่น

บทที่ 5
การเตรียมการทดลอง และวิธีการวัดผล

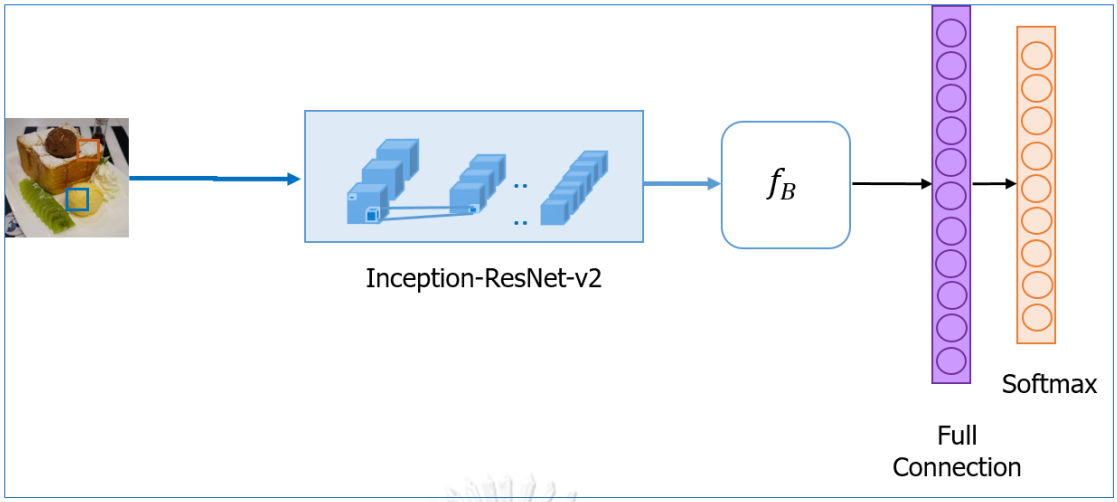
หัวข้อนี้จะกล่าวถึงแบบจำลองอ้างอิง ที่ถูกสร้างขึ้นเพื่อนำมาเปรียบเทียบประสิทธิภาพกับแบบจำลองที่นำเสนอในงานวิจัยนี้ ระบบที่ใช้ในการทดลอง และวิธีการวัดผล โดยแบบจำลองที่ Chen และคณะ [5] นำเสนอ ไม่ได้ถูกนำมาเปรียบเทียบประสิทธิภาพ เนื่องจากในปัจจุบันมีคอนโวลูชันเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะ คืออินเซ็บชันเรสเนส ที่สามารถให้ประสิทธิภาพได้ดีกว่าคอนโวลูชันเน็ตเวิร์กของงานวิจัยดังกล่าว อ้างอิงจากงานวิจัย [7] ดังนั้นแบบจำลองที่ใช้ในการเปรียบเทียบประสิทธิภาพมีดังต่อไปนี้

5.1 แบบจำลองอ้างอิง เพื่อใช้ในการเปรียบเทียบประสิทธิภาพ

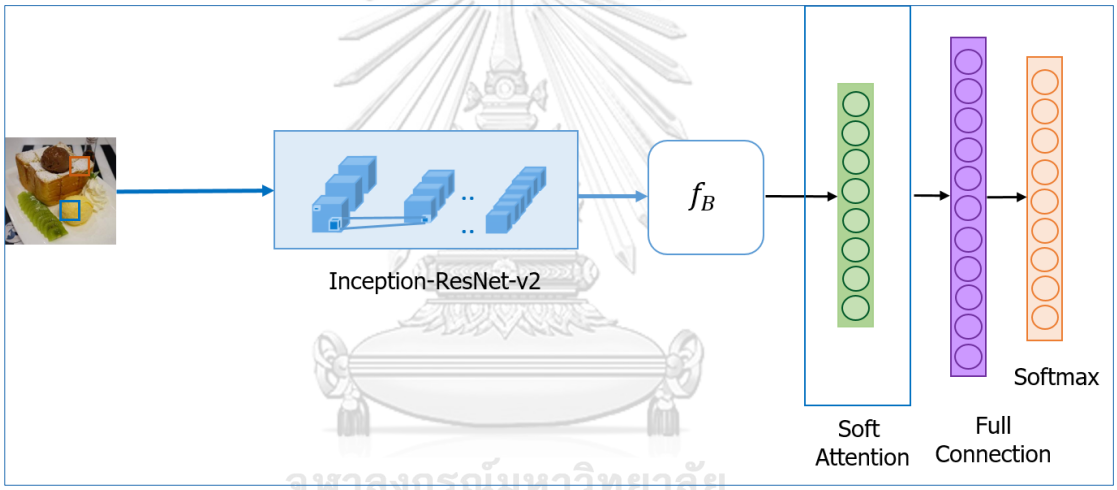
งานวิจัยนี้จะทำการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ไม่มีกลไกจุดสนใจ และแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีกลไกจุดสนใจ บนชุดข้อมูลประเภทรูปภาพอาหารของ Wongnai ที่ถูกแบ่งออกเป็น 83 ประเภทอาหาร แบบจำลองอ้างอิงที่ถูกนำมาใช้เพื่อเปรียบเทียบกับแบบจำลองที่นำเสนอ แบ่งออกเป็น 2 โครงสร้าง ดังนี้ (1) แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน (2) แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ไม่มีกลไกจุดสนใจ ซึ่งในส่วนของแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ไม่มีกลไกจุดสนใจนั้น ได้ทำการเปรียบเทียบคอนโวลูชันเน็ตเวิร์กที่ทำหน้าที่เป็นตัวสกัดคุณลักษณะโดยเปรียบเทียบระหว่างเน็ตเวิร์กที่เหมือนกัน และเน็ตเวิร์กที่แตกต่างกัน เพื่อหาแบบจำลองที่ดีที่สุดในการนำไปเปรียบเทียบกับแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีกลไกจุดสนใจ

5.1.1 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน (Convolutional Neural Networks หรือ CNN)

คือแบบจำลองที่นำคอนโวลูชันเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะของรูปภาพมาใช้เพียงแค่ตัวเดียว นั่นคืออินเซ็บชันเน็ตเวิร์กเวอร์ชันสอง โดยในการสร้างแบบจำลองอ้างอิงนั้นจะเริ่มทำการทดลองตั้งแต่แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันที่ยังไม่มีกลไกจุดสนใจ และแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันที่ต่อด้วยกลไกจุดสนใจ โครงสร้างของแบบจำลองทั้งสอง แสดงดังรูปที่ 22 และ 23 ตามลำดับ



รูปที่ 22 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน



รูปที่ 23 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันที่มีกลไกจุดสนใจ

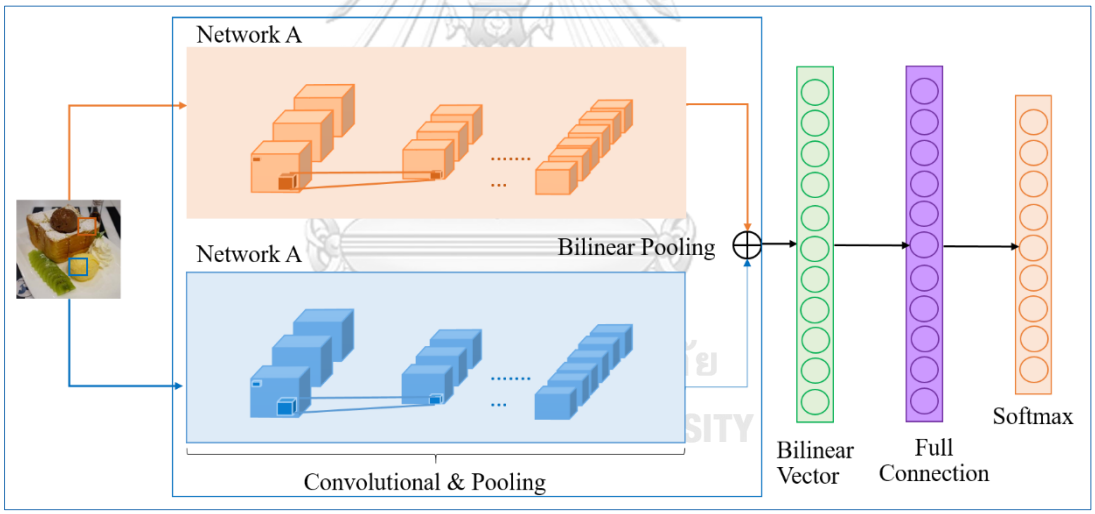
5.1.2 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ (Bilinear Convolutional Neural Networks หรือ B-CNN)

คือแบบจำลองที่ถูกดัดแปลงมาจากแบบจำลองในหัวข้อ 5.1.1 โดยได้เพิ่มคอนโวลูชันเน็ตเวิร์กมาอีกหนึ่งเน็ตเวิร์ก เรียกว่าแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ การเพิ่มคอนโวลูชันเน็ตเวิร์กมาเนื่องจากต้องการทดสอบประสิทธิภาพของแบบนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ ว่าการที่สามารถสกัดหาคุณลักษณะที่หลากหลายของรูปภาพได้มากขึ้น จะสามารถแก้ปัญหาการจำแนกประเภทของรูปภาพแบบละเอียดได้มากกว่าแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันโดยทั่วไปมากน้อยเพียงใด

แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ไม่มีกลไกจุดสนใจนั้น ได้ถูกแบ่งเป็นสองแบบจำลองย่อย เพื่อทำการเปรียบเทียบคอนโวลูชันเน็ตเวิร์กที่ทำหน้าที่เป็นตัวสกัดคุณลักษณะโดยเปรียบเทียบระหว่างเน็ตเวิร์กที่เหมือนกัน และเน็ตเวิร์กที่แตกต่างกัน เพื่อหาแบบจำลองที่ดีที่สุดก่อนที่จะนำแบบจำลองในหัวข้อนี้ไปเปรียบเทียบกับแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีกลไกจุดสนใจ

5.1.2.1 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีคอนโวลูชันเน็ตเวิร์กที่เหมือนกัน (B-CNN [A, A])

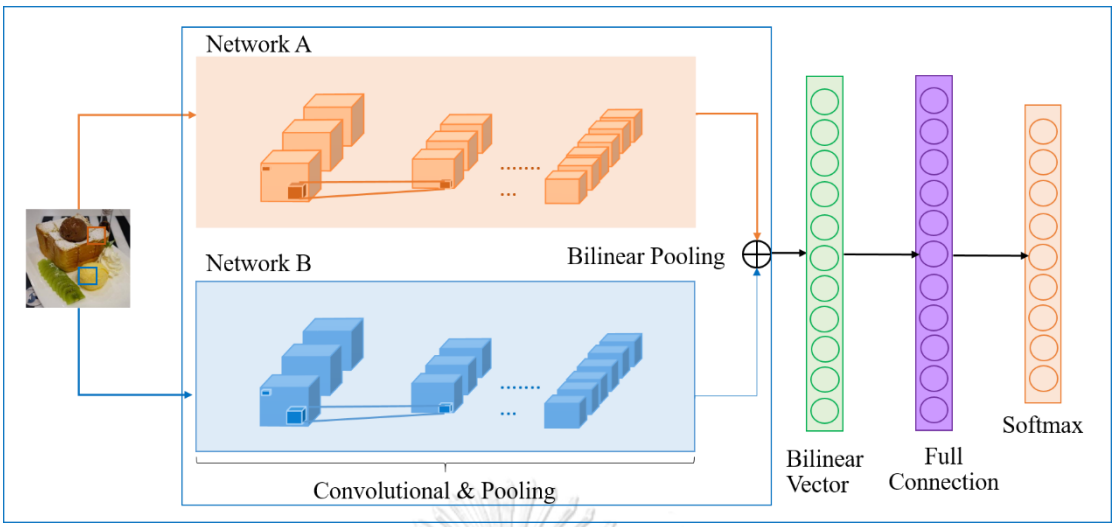
คือแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ใช้คอนโวลูชันเน็ตเวิร์กตัวเดียวกันในการสกัดคุณลักษณะของรูปภาพ ในการวิจัยนี้ใช้อินเซ็บชันเน็ตเวิร์กเวอร์ชันสาม และ อินเซ็บชันเรสเน็ตเวอร์ชันสอง มาเป็นคอนโวลูชันเน็ตเวิร์กที่ใช้ในการเปรียบเทียบกัน โครงสร้างของแบบจำลองแสดงดังรูปที่ 24



รูปที่ 24 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีคอนโวลูชันเน็ตเวิร์กที่เหมือนกัน

5.1.2.2 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน (B-CNN [A, B])

คือแบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่ใช้คอนโวลูชันเน็ตเวิร์กที่แตกต่างกันในการสกัดคุณลักษณะของรูปภาพ ในการวิจัยนี้ใช้การรวมกันของอินเซ็บชันเน็ตเวิร์กเวอร์ชันสาม และ อินเซ็บชันเรสเน็ตเวอร์ชันสอง โครงสร้างของแบบจำลองแสดงดังรูปที่ 25



รูปที่ 25 แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นคู่ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน

5.2 ระบบที่ใช้ในการทดลอง

ในส่วนนี้จะอธิบายถึงสภาพแวดล้อมที่ใช้ในการทดลอง ได้แก่ การแบ่งชุดข้อมูล และวิธีการสอนนิวรอลเน็ตเวิร์กเชิงลึก

5.2.1 การแบ่งชุดข้อมูล

ในงานวิจัยนี้ใช้ข้อมูลรูปภาพจาก Wongnai ตั้งแต่ปี 2558 ถึงปี 2562 รวมเป็นระยะเวลา 5 ปี โดยทำการจำแนกประเภทของรูปภาพอาหารเป็น 83 ประเภท และได้แบ่งชุดข้อมูลย่อยออกเป็น 3 ชุดซึ่งมีสัดส่วนของรูปภาพอาหารในแต่ละประเภทเท่า ๆ กัน ได้แก่ ชุดข้อมูลสำหรับทำการฝึกสอน จำนวน 115,496 ภาพ ชุดข้อมูลสำหรับตรวจสอบจำนวน 14,290 ภาพ และชุดข้อมูลสำหรับทำการทดสอบจำนวน 14,378 ภาพ ในการทดลองแบบจำลอง ตารางสรุปสถิติการจัดแบ่งข้อมูลในแต่ละชุดแสดงในตารางที่ 1

ตารางที่ 1 สถิติการจัดแบ่งข้อมูลในแต่ละชุด

ชุดข้อมูล	จำนวนประเภทอาหาร	จำนวนรูปภาพ	%	เฉลี่ยรูปภาพต่อประเภทอาหาร
ชุดข้อมูลสำหรับทำการฝึกสอน	83	115,496	80	1,392
ชุดข้อมูลสำหรับตรวจสอบ		14,290	10	173
ชุดข้อมูลสำหรับทำการทดสอบ		14,378	10	173

5.2.2 วิธีการสอนนิวรอลเน็ตเวิร์ก

ในส่วนนี้จะอธิบายถึงการตั้งค่าพารามิเตอร์สำหรับการสอนนิวรอลเน็ตเวิร์ก โดยในแต่ละแบบจำลอง จะมีการตั้งค่าพารามิเตอร์บางค่าที่ต่างกันตามความเหมาะสม เพื่อให้แบบจำลองมีการเรียนรู้อย่างมีประสิทธิภาพ ซึ่งในการทดลองจะเลือกค่าน้ำหนักจากรอบการเรียนรู้ (Epoch) บนชุดข้อมูลสำหรับตรวจสอบ ที่มีค่าความสูญเสีย (Loss) ที่ต่ำสุด ไปใช้ในชุดข้อมูลทดสอบ ตารางที่ 2 เป็นการแสดงรายละเอียดของค่าพารามิเตอร์ที่ใช้ในแต่ละแบบจำลอง

พารามิเตอร์ที่ใช้สอนนิวรอลเน็ตเวิร์กมีดังนี้

- ขนาดชุดข้อมูล (Batch Size)
- อัตราการเรียนรู้ (Learning Rate) ซึ่งเป็นค่าที่บ่งบอกว่าการเรียนรู้ในแต่ละรอบจะมีการเปลี่ยนแปลงน้ำหนักด้วยอัตราส่วนของผลต่างของผลลัพธ์ไปมากเท่าใด
- ค่าที่เหมาะสมที่สุด (Optimization) เป็นค่าที่ใช้เพื่อลดค่าของฟังก์ต้นทุนให้มีค่าน้อยที่สุด โดยใช้วิธีปรับปรุงน้ำหนักของเส้นเชื่อมในนิวรอลเน็ตเวิร์ก ในการทดลองนี้ ทุกแบบจำลองใช้ Adam เป็นตัวช่วยปรับค่าเกรเดียนในการเรียนรู้
- ค่าสัญญาณตกหาย (Dropout) เป็นค่าที่ใช้เพื่อป้องกันการยึดติดกับชุดข้อมูลฝึกสอน หรือ Overfitting
- ฟังก์ชันต้นทุน/วัตถุประสงค์ ในการทดลองนี้ ทุกแบบจำลองใช้ฟังก์ชันต้นทุนประเภท Cross-entropy

ตารางที่ 2 ค่าพารามิเตอร์ที่ใช้ในแต่ละแบบจำลอง

แบบจำลอง	Batch size	Learning Rate	Epoch ที่เลือก
In-res-v2 + Soft Attention	200	0.002	70
B-CNN [In-v3, In-v3]	200	0.001	64
B-CNN [In-res-v2, In-res-v2]	164	0.002	74
B-CNN [In-v3, In-res-v2]	164	0.001	83
B-CNN [In-v3, In-res-v2] + Residual Attention	164	0.001	60
B-CNN [In-v3, In-res-v2] + Multi-head Attention	164	0.001	100
B-CNN [In-v3, In-res-v2] + Soft Attention	164	0.001	80

5.3 การวัดผล

การวัดประสิทธิภาพการจำแนกแบบหลายคลาส (Multi-class Classification) สามารถแสดงได้ดังนี้

5.3.1 คอนฟิวชันเมทริกซ์ (Confusion Matrix)

คือ เมทริกซ์ที่แสดงผลของการจำแนกโดยแจกแจงจำนวนที่จำแนกได้ตามคลาส ดังตัวอย่างในตารางที่ 3 ซึ่งแสดงการจาแนกข้อมูลเป็น 83 คลาส โดยค่าแต่ละแถวแสดงจำนวนข้อมูลที่มีคลาสนั้นเป็นคำตอบที่ถูกต้อง ส่วน ค่าในแต่ละหลักแสดงจำนวนข้อมูลที่ทำนายได้ในคลาสนั้น กำหนดให้สำหรับคลาสใด ๆ

(1) TP คือ จำนวนข้อมูลที่ทำนายได้คลาสซึ่งกำลังสนใจอยู่และทำนายถูก (True Positive)

(2) FP คือ จำนวนข้อมูลที่ทำนายได้คลาสซึ่งกำลังสนใจอยู่และทำนายผิด (False Positive)

(3) TN คือ จำนวนข้อมูลที่ทำนายได้คลาสซึ่งไม่ได้สนใจอยู่และทำนายถูก (True Negative)

(4) FN คือ จำนวนข้อมูลที่ทำนายได้คลาสซึ่งไม่ได้สนใจอยู่และทำนายผิด (False Negative)

ตารางที่ 3 ตัวอย่างคอนฟิวชันเมทริกซ์ของการจำแนกแบบหลายคลาส

		คลาสที่ทำนาย		
		$C_{0\ldots}C_{k-1}$	C_k	$C_{k+1\ldots}C_n$
คลาสจริง	$C_{k+1\ldots}C_n$	TN	FP	TN
	C_k	FN	TP	FN
	$C_{0\ldots}C_{k-1}$	TN	FP	TN

5.3.2 ตัววัดประสิทธิภาพจำแนกตามคลาส

โดยทั่วไปตัววัดประสิทธิภาพที่นิยมใช้กันในงานวิจัยมีอยู่ 4 ค่า ดังนี้

ค่าความเที่ยง (Precision หรือ P) เป็นการวัดความแม่นยำของแบบจำลองโดยการพิจารณาแยกทีละคลาส ตัวอย่างเช่น การวัดว่าแบบจำลองทำนายว่าคำตอบที่เป็นบวกถูกต้องเท่าไหร่จากผลการทำนายคลาสบวกทั้งหมดเท่าไหร่

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

ค่าความระลึก (Recall หรือ R) เป็นการวัดความถูกต้องของแบบจำลองโดยการพิจารณาแยกทีละคลาส ตัวอย่างเช่น การวัดว่าผลการทำนายคลาสบวกความถูกต้องเท่าไหร่เมื่อเทียบกับคลาสบวกจริงทั้งหมด

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

ค่าเอฟวัน (F1 score หรือ F1) เป็นการวัดความเที่ยงและความระลึกของแบบจำลองไปพร้อม ๆ กันโดยคำนวณได้จาก สมการดังต่อไปนี้

$$F1 = \frac{2\times Recall\times Precision}{Recall+Precision} \tag{13}$$

ค่าความแม่นยำ (Accuracy หรือ Acc) เป็นการวัดความแม่นยาของแบบจำลองโดยรวม กล่าวคือ แบบจำลอง ทำนายถูกกี่ครั้งจากจำนวนการทำนายทั้งหมด

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{14}$$

ค่าเฉลี่ยจุลภาค (Micro-average หรือ Micro-avg) เป็นการนำค่าข้อมูลการนับจากทุกรูปแบบคำตอบมารวมกัน แล้วนำผลรวมนั้นไปหาค่าตัวชี้วัดโดยตรง

กำหนดให้ M แทนจำนวนคลาสที่แตกต่างกันบนชุดข้อมูล

ค่าเฉลี่ยจุลภาคของเอฟวัน แสดงได้ตามสมการดังต่อไปนี้

$$Precision_{micro} = \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M}(TP_i+FP_i)} \tag{15}$$

$$Recall_{micro} = \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M}(TP_i+FN_i)} \tag{16}$$

$$F1_{micro} = 2\times\frac{Precision_{micro}\times Recall_{micro}}{Precision_{micro}+Recall_{micro}} \tag{17}$$

ค่าเฉลี่ยมหภาค (Macro-average หรือ Macro-avg) เป็นการนำค่าตัวชี้วัดแต่ละตัวมาเฉลี่ยตามจำนวนรูปแบบคำตอบ

กำหนดให้ M แทนจำนวนคลาสที่แตกต่างกันบนชุดข้อมูล

ค่าเฉลี่ยมหภาคของเอฟวัน แสดงได้ตามสมการดังต่อไปนี้

$$Precision_{macro} = \frac{1}{M}\times\sum_{i=1}^{M}\frac{TP_i}{TP_i+FP_i} \tag{18}$$

$$Recall_{macro} = \frac{1}{M}\times\sum_{i=1}^{M}\frac{TP_i}{TP_i+FN_i} \tag{19}$$

$$F1_{macro} = 2 \times \frac{Precision_{macro} \times Recall_{macro}}{Precision_{macro} + Recall_{macro}} \quad (20)$$

บทที่ 6
ผลการทดลอง

ในส่วนของผลการทดลอง แบบจำลองที่วิทยานิพนธ์นี้ได้นำเสนอ จะถูกนำไปเปรียบเทียบกับแบบจำลองในหัวข้อที่ 5.1 โดยแบบจำลองจะถูกฝึกสอนโดยใช้ชุดข้อมูลสำหรับทำการฝึกสอน หลังจากนั้นจึงเลือกแบบจำลองที่มีความแม่นยำสูงสุดเพื่อทำการทดสอบบนชุดข้อมูลสำหรับตรวจสอบ แล้วจึงทำการวัดประสิทธิภาพของแบบจำลองแต่ละประเภทบนข้อมูลทดสอบ ในวิทยานิพนธ์ฉบับนี้ จะทำการเปรียบเทียบผลการทดลองออกเป็น 3 ส่วนหลัก ตามลำดับของการทดลอง ได้แก่ (1) ประสิทธิภาพของ B-CNN (2) ประสิทธิภาพของ CNN เมื่อนำมาเปรียบเทียบกับ B-CNN (3) ผลการทดลองโดยรวม และการอภิปรายผล

6.1 ประสิทธิภาพของ B-CNN

ในหัวข้อนี้ จะเป็นการแสดงประสิทธิภาพของแบบจำลองที่นำเสนอ หรือแบบจำลอง B-CNN ที่มีผลต่อการจำแนกประเภทของรูปภาพของอาหาร โดยแบ่งการทำการทดลองย่อย ๆ ตามการปรับรูปแบบของปัจจัยที่มีผลต่อแบบจำลอง ได้แก่ เน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะ การนำกลไกจุดสนใจมาใช้ร่วมกับ B-CNN และความละเอียดของรูปภาพที่มีผลต่อแบบจำลอง และนำผลการทดลองมาเปรียบเทียบกัน เพื่อให้ได้ B-CNN ที่มีประสิทธิภาพมากที่สุด

6.1.1 ส่วนของการสกัดคุณลักษณะ

การทดลองในหัวข้อนี้ จะเป็นการเปรียบเทียบคอนโวลูชันเน็ตเวิร์กแบบต่าง ๆ เพื่อหาคอนโวลูชันเน็ตเวิร์กที่มีประสิทธิภาพมากที่สุดมาใช้เป็นส่วนของการสกัดคุณลักษณะของแบบจำลอง B-CNN โดยทำการเปรียบเทียบส่วนของการสกัดคุณลักษณะที่ใช้คอนโวลูชันเน็ตเวิร์กเหมือนกัน และส่วนของการสกัดคุณลักษณะที่ใช้คอนโวลูชันเน็ตเวิร์กต่างกัน

จากผลการทดลองในตารางที่ 4 ซึ่งเป็นการเปรียบเทียบ B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่เหมือนกัน และ B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน พบว่า B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน ให้ผลเอฟวันสำหรับค่าเฉลี่ยมหภาคที่มากกว่าแบบจำลองอื่น และสามารถทำนายประเภทของรูปภาพอาหารได้อย่างแม่นยำ โดยมีสถิติค่าเอฟวันในแต่ละประเภทอาหารมากกว่าแบบจำลองประเภทอื่น ดังแสดงในตารางที่ 5 สาเหตุเกิดจากการใช้คอนโวลูชันเน็ตเวิร์กที่ต่างกัน ทำให้การเกิดคุณลักษณะที่หลากหลายเนื่องจากในแต่ละเน็ตเวิร์กถูกให้ค่าน้ำหนักที่ต่างกัน จึงสามารถสร้างคุณลักษณะของรูปภาพได้มากกว่า ประสิทธิภาพของแบบจำลองจึงดีกว่า

ตารางที่ 4 ผลการทดลองเปรียบเทียบระหว่างง B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่เหมือนกัน และ B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
B-CNN [In-v3, In-res-v2]	F1	**71.15**	75.18	**75.18**
	P	77.30	75.18	
	R	69.10	75.18	
B-CNN [In-res-v2, In-res-v2]	F1	70.20	**77.12**	77.12
	P	76.63	77.12	
	R	69.29	77.12	
B-CNN [In-v3, In-v3]	F1	67.37	75.37	75.37
	P	75.00	75.37	
	R	66.61	75.37	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 5 การเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวัน ของ B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่เหมือนกัน และ B-CNN ที่มีคอนโวลูชันเน็ตเวิร์กที่ต่างกัน โดยมีหน่วยเป็นจำนวนประเภทอาหาร

แบบจำลอง	ชนะ	เสมอ	แพ้
B-CNN [In-v3, In-res-v2] vs B-CNN [In-res-v2, In-res-v2]	46	0	37
B-CNN [In-v3, In-res-v2] vs B-CNN [In-v3, In-v3]	51	0	32

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.1.2 ส่วนของกลไกจุดสนใจ

ในหัวข้อนี้ ได้นำเอาแบบจำลองภาพที่มีประสิทธิดีที่สุดจากหัวข้อที่ 6.1.1 นั่นคือ B-CNN [In-v3, In-res-v2] มาทำการทดลองต่อโดยการเพิ่มกลไกจุดสนใจเข้ามา โดยในการทดลองนี้ได้ใช้กลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อน เพื่อเปรียบเทียบความสำคัญของกลไกจุดสนใจที่มีผลต่องานจำแนกประเภทของรูปภาพอาหาร โดยผลการทดลองในตารางที่ 6 เป็นการเปรียบเทียบระหว่างแบบจำลองที่ไม่มีกลไกจุดสนใจ และแบบจำลองที่มีกลไกจุดสนใจ พบว่าแบบจำลองที่มีกลไกจุดสนใจสามารถเพิ่มค่าเอฟวัน ค่าความเที่ยง และค่าความระลึกสำหรับค่าเฉลี่ยมหภาคและค่าเฉลี่ยจุลภาคได้เฉลี่ย 14 เปอร์เซ็นต์ เนื่องจากการเพิ่มกลไกจุดสนใจ สามารถดึงคุณลักษณะที่สำคัญของรูปภาพนั้น ๆ ออกมาได้ ทำให้ผลการทดลองมีความแม่นยำกว่าแบบจำลองอื่น ๆ จากตารางที่ 7

แสดงการเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวัน ของแบบจำลองแต่ละประเภท พบว่าแบบจำลองที่มีกลไกจุดสนใจสามารถชนะแบบจำลองอื่น ๆ ได้มากถึง 99 เปอร์เซ็นต์ โดยตารางที่ 8 แสดงให้เห็นว่าแบบจำลองนี้สามารถจำแนกประเภทของรูปภาพอาหารในแต่ละประเภทได้อย่างแม่นยำ เมื่อวัดจาก 10 อันดับแรก ที่มีค่าเอฟวันมากกว่า 90 เปอร์เซ็นต์ ขึ้นไป ซึ่งมากกว่าแบบจำลองที่ไม่มีกลไกจุดสนใจโดยเฉลี่ยประมาณ 11 เปอร์เซ็นต์

ตารางที่ 6 ผลการทดลองเปรียบเทียบระหว่างแบบจำลองที่ไม่มีกลไกจุดสนใจ และแบบจำลองที่มีกลไกจุดสนใจ

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
B-CNN [In-v3, In-res-v2] +Soft Att	F1	**86.55**	**90.08**	**90.08**
	P	86.54	90.08	
	R	86.82	90.08	
B-CNN [In-v3, In-res-v2]	F1	71.15	75.18	75.18
	P	77.30	75.18	
	R	69.10	75.18	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 7 การเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวันของแบบจำลองที่ไม่มีกลไกจุดสนใจ กับแบบจำลองที่มีกลไกจุดสนใจ หน่วยเป็นจำนวนประเภทอาหาร

แบบจำลอง	ชนะ	เสมอ	แพ้
B-CNN [In-v3, In-res-v2] +Att vs B-CNN [In-v3, In-res-v2]	82	0	1

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 8 ค่าเอฟวันของประเภทอาหารที่มีค่าสูงสุด 10 อันดับแรกของแบบจำลองที่ไม่มีกลไกจุดสนใจ เทียบกับแบบจำลองที่มีกลไกจุดสนใจ

ประเภทอาหาร	F1 (%)	
	B-CNN [In-v3, In-res-v2] + Att	B-CNN [In-v3, In-res-v2]
กุ้งอบวุ้นเส้น	97.96	87.82
ปลาทอด	97.77	87.85
ทอดมัน	96.56	86.84
ส้มตำ	96.35	87.73
ข้าวซอยไก่	95.95	86.76
ผัดไทกุ้งสด	95.91	90.06
ข้าวผัด	95.80	88.22
ใบเหลียงผัดไข่	95.76	66.23
ปูม้านึ่ง	95.36	83.82
กุ้ง	95.26	87.15

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.1.3 ประเภทของกลไกจุดสนใจ

ในหัวข้อนี้ เป็นการเปรียบเทียบ B-CNN ที่มีกลไกจุดสนใจ โดยใช้ประเภทของกลไกจุดสนใจที่แตกต่างกัน โดยผลการทดลองพบว่าง B-CNN ที่เพิ่มกลไกจุดสนใจที่ให้ค่าความสนใจอย่างอ่อน ให้ประสิทธิภาพในการทดลองมากที่สุด ดังตารางที่ 9

ตารางที่ 9 ผลการทดลองเปรียบเทียบของ B-CNN ที่มีกลไกจุดสนใจ โดยใช้ประเภทของกลไกจุดสนใจที่แตกต่างกัน

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
B-CNN [In-v3, In-res-v2] + Soft Att	F1	**86.55**	**90.00**	**90.00**
	P	86.54	90.08	
	R	86.82	90.08	
B-CNN [In-v3, In-res-v2] + Multi-head Att	F1	79.13	79.86	79.86
	P	73.92	79.86	
	R	74.57	79.86	
B-CNN [In-v3, In-res-v2] + Residual Att	F1	66.58	72.17	72.17
	P	75.17	72.17	
	R	65.61	72.17	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.1.4 ความละเอียดของภาพ

ในหัวข้อนี้ เป็นการเปรียบเทียบความสามารถในการจำแนกประเภทของรูปภาพอาหาร โดยมีปัจจัยเป็นความละเอียดของรูปภาพที่ใช้ โดยได้นำแบบจำลองที่ดีที่สุดจากหัวข้อที่ 6.1.3 นั่นคือ B-CNN [In-v3, In-res-v2] + Soft Att ที่ใช้ข้อมูลนำเข้าเป็นรูปภาพอาหาร ซึ่งมีความละเอียดของภาพ 299x299 พิกเซล มาทำการเปรียบเทียบกับแบบจำลองเดียวกัน แต่ใช้ข้อมูลนำเข้าเป็นรูปภาพอาหารที่มีความละเอียดของภาพที่มากขึ้น (Higher Resolution หรือ HR) ซึ่งมีความละเอียดของภาพเพิ่มขึ้น 1.5 เท่า หรือ 450x450 พิกเซล

จากตารางที่ 10 แบบจำลองที่ใช้ข้อมูลนำเข้าเป็นรูปภาพอาหาร ซึ่งมีความละเอียดของภาพเพิ่มขึ้น 1.5 เท่า มีค่าเอฟวันสำหรับค่าเฉลี่ยมหภาค และค่าเฉลี่ยจุลภาคมากกว่าแบบจำลองที่ใช้ข้อมูลนำเข้าเป็นรูปภาพที่มีความละเอียดน้อยกว่า และสามารถทำนายประเภทของรูปภาพอาหารได้อย่างแม่นยำ โดยมีสถิติค่าเอฟวันในแต่ละประเภทอาหารมากกว่าแบบจำลองประเภทอื่น ดังแสดงในตารางที่ 11 เนื่องจากการใช้รูปภาพที่มีความละเอียดมากขึ้น ทำให้คอนโวลูชันเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะ สามารถมองเห็นคุณลักษณะได้ชัดขึ้น และไม่สูญเสียข้อมูลที่สำคัญไป

ตารางที่ 10 ผลการทดลองเปรียบเทียบระหว่างแบบจำลองที่ใช้ความละเอียดของข้อมูลนำเข้าต่างกัน

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
B-CNN [In-v3, In-res-v2] +Soft Att (HR)	F1	**87.72**	**90.51**	**90.51**
	P	87.34	90.51	
	R	87.37	90.51	
B-CNN [In-v3, In-res-v2] +Soft Att	F1	86.55	90.08	90.08
	P	86.54	90.08	
	R	86.82	90.08	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 11 การเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวันของแบบจำลองที่ใช้ความละเอียดของข้อมูลนำเข้าต่างกัน โดยมีหน่วยเป็นจำนวนประเภทอาหาร

แบบจำลอง	ชนะ	เสมอ	แพ้
B-CNN [In-v3, In-res-v2] +Soft Att (HR) vs B-CNN [In-v3, In-res-v2] +Soft Att	46	0	37

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.2 ประสิทธิภาพของ CNN เมื่อนำมาเปรียบเทียบกับ B-CNN

ในหัวข้อนี้ เป็นการเปรียบเทียบ B-CNN กับ CNN เพื่อแสดงให้เห็นถึงประสิทธิภาพของการนำคอนโวลูชันเน็ตเวิร์ก 2 เน็ตเวิร์กมาใช้ร่วมกันได้อย่างชัดเจนขึ้น โดยแบ่งหัวข้อในการเปรียบเทียบออกเป็น 2 หัวข้อย่อย ได้แก่ การเปรียบเทียบระหว่าง CNN ที่ไม่มีกลไกจุดสนใจ เทียบกับ B-CNN ที่ไม่มีกลไกจุดสนใจ และการเปรียบเทียบระหว่าง CNN ที่มีกลไกจุดสนใจ และ B-CNN ที่มีกลไกจุดสนใจ

6.2.1 เปรียบเทียบ CNN ที่ไม่มีกลไกจุดสนใจ และ B-CNN ที่ไม่มีกลไกจุดสนใจ

จากผลการทดลองในตารางที่ 12 พบว่า B-CNN ให้ผลเอฟวันสำหรับค่าเฉลี่ยมหภาค และค่าเฉลี่ยจุลภาคมากกว่าแบบจำลองอื่นที่ใช้คอนโวลูชันแบบคอนโวลูชันเชิงเส้นเพียงเน็ตเวิร์กเดียว โดย B-CNN สามารถทำนายประเภทของรูปภาพอาหารได้อย่างแม่นยำ และมีสถิติค่าเอฟวันในแต่ละประเภทอาหารมากกว่าแบบจำลองประเภทอื่น ดังแสดงในตารางที่ 13 สาเหตุเกิดจากการใช้คอนโวลูชันเน็ตเวิร์ก 2 เน็ตเวิร์กมาคูณกันแบบการคูณภายนอก ทำให้แบบจำลองเกิดการสร้างคุณลักษณะที่หลากหลายมากกว่าการใช้เน็ตเวิร์กเพียง 1 เน็ตเวิร์ก ส่งผลให้ในแต่ละประเภทของอาหาร สามารถสร้างคุณลักษณะของตนเองได้มากขึ้น แบบจำลองจึงสามารถเรียนรู้ และแยกคุณลักษณะของประเภทของอาหารที่มีลักษณะที่คล้ายกันได้ดีขึ้น ส่งผลให้ B-CNN มีประสิทธิภาพที่ดีกว่า

ตารางที่ 12 ผลการทดลองเปรียบเทียบระหว่าง CNN ที่ไม่มีกลไกจุดสนใจ และ B-CNN ที่ไม่มีกลไกจุดสนใจ

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
In-res-v2	F1	62.80	69.24	69.24
	P	73.21	69.24	
	R	61.91	69.24	
B-CNN [In-res-v2, In-res-v2]	F1	**70.20**	**77.12**	**77.12**
	P	76.63	77.12	
	R	69.29	77.12	
B-CNN [In-v3, In-v3]	F1	67.37	75.37	75.37
	P	75.00	75.37	
	R	66.61	75.37	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 13 การเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวันของ CNN ที่ไม่มีกลไกจุดสนใจ และ B-CNN ที่ไม่มีกลไกจุดสนใจเมื่อทดสอบด้วยชุดข้อมูลทดสอบโดยมีหน่วยเป็นจำนวนประเภทอาหาร

แบบจำลอง	ชนะ	เสมอ	แพ้
B-CNN [In-res-v2, In-res-v2] vs In-res-v2	62	0	21
B-CNN [In-v3, In-v3] vs In-res-v2	50	0	33

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.2.2 เปรียบเทียบ CNN ที่มีกลไกจุดสนใจ และ B-CNN ที่มีกลไกจุดสนใจ

จากผลการทดลองในตารางที่ 14 พบว่า B-CNN ที่มีกลไกจุดสนใจ ให้ผลเอฟวันสำหรับค่าเฉลี่ยมหภาค และค่าเฉลี่ยจุลภาคมากกว่า CNN ที่มีกลไกจุดสนใจ โดย B-CNN สามารถทำนายประเภทของรูปภาพอาหารได้อย่างแม่นยำ และมีสถิติค่าเอฟวันในแต่ละประเภทอาหารมากกว่าแบบจำลองประเภทอื่น ดังแสดงในตารางที่ 15 สาเหตุเกิดจากการใช้คอนโวลูชันเน็ตเวิร์ก 2 เน็ตเวิร์กมาคูณกันแบบการคูณภายนอกและการเพิ่มกลไกจุสนใจลงไปในแบบจำลอง ทำให้แบบจำลองเกิดการสร้างคุณลักษณะที่หลากหลายมากกว่าการใช้เน็ตเวิร์กเพียง 1 เน็ตเวิร์ก และยังสามารถสกัดคุณลักษณะที่สำคัญและเหมาะสมกับประเภทอาหารนั้น ๆ อีกด้วย ส่งผลให้ B-CNN ที่มีกลไกจุดสนใจ มีประสิทธิภาพที่ดีกว่า CNN ที่มีกลไกจุดสนใจ

ตารางที่ 14 ผลการทดลองเปรียบเทียบระหว่าง CNN ที่มีกลไกจุดสนใจ และ B-CNN ที่มีกลไกจุสนใจ

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
In-res-v2 + Soft Att	F1	67.10	72.62	72.62
	P	73.92	72.62	
	R	66.57	72.62	
B-CNN [In-v3, In-res-v2] + Soft Att	F1	86.55	90.08	**90.08**
	P	86.54	90.08	
	R	86.82	90.08	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 15 การเปรียบเทียบการทำนายประเภทของรูปภาพอาหาร โดยเปรียบเทียบจากค่าเอฟวัน ของ CNN ที่มีกลไกจุดสนใจ และ B-CNN ที่มีกลไกจุดสนใจ โดยมีหน่วยเป็นจำนวนประเภทอาหาร

แบบจำลอง	ชนะ	เสมอ	แพ้
B-CNN [In-v3, In-res-v2]+Soft Att vs In-res-v2 + Soft Att	81	0	2

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

6.3 ผลการทดลองโดยรวม และการอภิปรายผล

หัวข้อนี้เป็นการสรุปผล ระหว่างแบบจำลองที่นำเสนอ หรือ B-CNN ที่มีกลไกจุดสนใจ ที่ใช้ข้อมูลนำเข้าซึ่งมีความละเอียดของภาพเพิ่มขึ้น 1.5 เท่า (B-CNN [In-v3, In-res-v2]+Soft Attention (HR)) เทียบกับแบบจำลองอื่น ๆ ทั้งหมด และวิเคราะห์ผลการทดลองอย่างละเอียด โดยใช้ประเภทอาหารที่มีค่าเอฟวันต่ำสุด 5 อันดับสุดท้ายจากแบบจำลอง In-res-v2 + Soft Att เป็นเกณฑ์การเปรียบเทียบ เพื่อแสดงให้เห็นถึงภาพรวมของการทดลองในงานจำแนกประเภทของรูปภาพอาหาร ด้วยแบบจำลองการเรียนรู้เชิงลึกในแบบต่าง ๆ

จากผลการทดลองในตารางที่ 16 แสดงให้เห็นถึงลำดับประสิทธิภาพของแบบจำลองในแบบต่าง ๆ ที่มีผลต่องานจำแนกประเภทของรูปภาพอาหาร เริ่มตั้งแต่แบบจำลองนิวรอลเน็ตเวิร์กแบบคอนโวลูชัน ไปจนถึงแบบจำลองที่วิทยานิพนธ์นี้ได้นำเสนอ คือ แบบจำลอง B-CNN [In-v3, In-res-v2]+Soft Att (HR) ซึ่งผลการทดลองแสดงให้เห็นว่า แบบจำลองที่นำเสนอให้ค่าเอฟวันมากที่สุด เนื่องจากการใช้ B-CNN ที่มีกลไกจุดสนใจ ร่วมกับข้อมูลนำเข้าที่มีความละเอียดของรูปภาพสูงมีผลต่องานจำแนกภาพที่มีความคล้ายคลึงกัน เพราะแบบจำลองสามารถสกัดหาคุณลักษณะที่สำคัญจากภาพนั้น ๆ ได้มาก และละเอียดขึ้น โดยที่ไม่สูญเสียคุณลักษณะบางอย่างไป

ตารางที่ 16 ผลการทดลองโดยรวม

แบบจำลอง	ตัววัด	Macro-avg (%)	Micro-avg (%)	Acc (%)
B-CNN [In-v3, In-res-v2] + Soft Att (HR)	F1	**87.72**	**90.51**	**90.51**
	P	87.34	90.51	
	R	87.37	90.51	
B-CNN [In-v3, In-res-v2] + Soft Att	F1	86.55	90.08	90.08
	P	86.54	90.08	
	R	86.82	90.08	
B-CNN [In-v3, In-res-v2] + Multi-head Att	F1	79.13	79.86	79.86
	P	73.92	79.86	
	R	74.57	79.86	
B-CNN [In-v3, In-res-v2] + Residual Att	F1	66.58	72.17	72.17
	P	75.17	72.17	
	R	65.61	72.17	
B-CNN [In-v3, In-res-v2]	F1	71.15	75.18	75.18
	P	77.30	75.18	
	R	69.10	75.18	
B-CNN [In-res-v2, In-res-v2]	F1	70.20	77.12	77.12
	P	76.63	77.12	
	R	69.29	77.12	
B-CNN [In-v3, In-v3]	F1	67.37	75.37	75.37
	P	75.00	75.37	
	R	66.61	75.37	
In-res-v2 + Soft Att	F1	67.10	72.62	72.62
	P	73.92	72.62	
	R	66.57	72.62	

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

จากผลการทดลองในตารางที่ 17 แสดงให้เห็นว่า แบบจำลองที่นำเสนอสามารถปรับปรุงค่าเอฟวันของประเภทอาหารที่มีค่าต่ำสุด 5 อันดับสุดท้ายจากแบบจำลอง In-res-v2 + Soft Att โดยมีค่าเอฟวันเพิ่มขึ้นโดยเฉลี่ย 179 เปอร์เซ็นต์

ตารางที่ 17 ค่าเอฟวันของประเภทอาหารที่มีค่าต่ำสุด 5 อันดับสุดท้ายของแบบจำลอง In-res-v2 +Soft Attention เทียบกับแบบจำลองที่นำเสนอ

ประเภทอาหาร	**ค่าเอฟวัน**		
	In-res-v2 +Soft Attention	B-CNN [In-v3, In-res-v2] + Soft Attention	**B-CNN [In-v3, In-res-v2] + Soft Attention (IR)**
ก๋วยเตี๋ยว	5.26%	85.61%	87.42%
ขนมหวาน	11.59%	46.43%	57.45%
ข้าวขาหมู	14.63%	83.72%	83.14%
ข้าวหมกไก่	15.38%	81.60%	81.54%
เนื้อย่าง	22.15%	53.23%	55.24%

* ตัวหนาคือแบบจำลองที่มีประสิทธิภาพดีที่สุด

ตารางที่ 18 และตารางที่ 19 ได้ทำการยกตัวอย่างคอนฟิวชันเมทริกซ์ของประเภทอาหารที่มีความคล้ายคลึงกันจาก 5 อันดับสุดท้าย ได้แก่ ขนมหวาน-เค้ก และ กิมจิ-เครื่องเคียง ของแบบจำลอง In-res-v2 +Soft Att และ แบบจำลอง B-CNN [In-v3, In-res-v2] + Soft Att (HR) ตามลำดับ เพื่อให้เห็นถึงประสิทธิภาพของ B-CNN ที่มีกลไกจุดสนใจ และประสิทธิภาพของความละเอียดภาพที่มีผลต่อแบบจำลอง

ผลจากตารางที่ 19 แสดงให้เห็นว่า แบบจำลองที่นำเสนอสามารถทำนายป้ายชื่ออาหารได้แม่นยำมากขึ้น และสามารถลดผลการทายผิดจากประเภทของรูปภาพอาหารที่มีลักษณะที่คล้ายคลึงกัน ตัวอย่างเช่น รูปภาพอาหารที่มีป้ายชื่อว่า ขนมหวาน ที่มักถูกทายผิดเป็นเค้ก ถูกทายเป็นเค้กได้อย่างถูกต้องมากขึ้นเมื่อใช้แบบจำลองที่นำเสนอ ตัวอย่างของรูปภาพระหว่างประเภทอาหารที่มีความคล้ายคลึงกันถูกแสดงดังรูปที่ 26 และการวิเคราะห์ลักษณะของภาพที่มีความคล้ายคลึงกันในแต่ละประเภทอาหาร แสดงดังตารางที่ 20

ตารางที่ 18 คอนฟิวชันเมทริกซ์ของแบบจำลอง In-res-v2 +Soft Att

		ป้ายชื่อที่ทำนายได้			
		กิมจิ	**ขนมหวาน**	**เค้ก**	**เครื่องเคียง**
ป้ายชื่อที่ถูกต้อง	กิมจิ	16	0	0	3
	ขนมหวาน	0	9	16	1
	เค้ก	0	13	112	0
	เครื่องเคียง	13	4	0	23

ตารางที่ 19 คอนฟิวชันเมทริกซ์ของแบบจำลอง B-CNN [In-v3, In-res-v2] + Soft Att (HR)

		ป้ายชื่อที่ทำนายได้			
		กิมจิ	**ขนมหวาน**	**เค้ก**	**เครื่องเคียง**
ป้ายชื่อที่ถูกต้อง	กิมจิ	58	0	0	4
	ขนมหวาน	0	23	7	0
	เค้ก	0	6	168	0
	เครื่องเคียง	4	1	0	40

รูปที่ 26 ตัวอย่างของรูปภาพระหว่างประเภทอาหารที่มีความคล้ายคลึงกัน

ตารางที่ 20 สรุปการวิเคราะห์ลักษณะของภาพที่มีความคล้ายคลึงกันในแต่ละประเภทอาหาร

		ประเภทของรูปภาพอาหารที่มีความคล้ายคลึงกัน				
		เนื้อย่าง-คอหมูย่าง	ขนมหวาน-เค้ก	กิมจิ-เครื่องเคียง	ซูชิ-ติ่มซำ	ขาหมูเยอรมัน-ไก่ย่าง
ประเภทของลักษณะ	สี	อาหารทั้งสองประเภทมีโทนสีที่คล้ายคลึงกันในโทนน้ำตาลเข้ม	อาหารทั้งสองประเภทมีโทนสีที่หลากหลายเนื่องจากประกอบด้วยวัตถุดิบหลายประเภท โดยสีส่วนใหญ่จะเป็นสีขาว สีแดง และสีเหลือง	อาหารทั้งสองประเภทมีโทนสีที่คล้ายคลึงกันในโทนส้ม-แดง	อาหารทั้งสองประเภทมีโทนสีที่หลากหลาย เนื่องจากตัววัตถุดิบทั้งในซูชิ และติ่มซำ ประกอบด้วยผักที่มีสีเขียว และตัวเนื้อสัตว์ที่มีโทนสีส้ม / น้ำตาล	มีโทนสีที่คล้ายคลึงกัน ส้ม/น้ำตาลเข้ม
	วัตถุดิบ	วัตถุดิบเป็นเนื้อหมูเหมือนกัน ประกอบด้วยเนื้อสีขาว และหนังสีน้ำตาลเข้ม	วัตถุดิบส่วนมากเป็นขนมปัง ผลไม้ และ ครีมหรือแยม	วัตถุดิบส่วนมากเป็นผักที่มีชิ้นเล็ก ๆ และมีน้ำจิ้มราดบนวัตถุดิบเหล่านั้น	วัตถุดิบที่คล้ายกันคือผักที่มีสีเขียว และเนื้อหมูหรือเนื้อปลาที่มีสีน้ำตาลอ่อน	ประกอบไปด้วยเนื้อหมูและไก่ ที่มีส่วนที่เป็นเนื้อ หนัง และกระดูก เหมือนกัน
	รูปทรง	รูปทรงของประเภทอาหารทั้งสองมีรูปทรงเป็นทรงรี ที่แบนและยาว อีกทั้งยังมีลักษณะของการถูกหั่นในรูปแบบที่ใกล้เคียงกัน	รูปทรงของประเภทอาหารทั้งสองมีลักษณะเป็นชิ้นเล็ก ๆ หลายชิ้นแยกออกจากกัน โดยส่วนมากมีรูปทรงกลม และเป็นชิ้นสามเหลี่ยม	รูปทรงของประเภทอาหารทั้งสองมีลักษณะเป็นชิ้นเล็ก ๆ หลายชิ้นแยกออกจากกัน	รูปทรงของประเภทอาหารทั้งสองมีลักษณะเป็นชิ้นเล็ก ๆ หลายชิ้นแยกออกจากกัน โดยส่วนมากมีรูปทรงกระบอก และทรงกลม	ขาหมู และสะโพกไก่มีรูปทรงที่คล้ายกัน อีกทั้งยังมีลักษณะของการถูกหั่นในรูปแบบที่ใกล้เคียงกัน
	ภาชนะ	ภาชนะส่วนใหญ่เป็นจาน/ชาม	ภาชนะส่วนใหญ่เป็นจาน/ชาม บางรูปประกอบด้วยภาชนะหลายชิ้นที่เหมือนกัน	ภาชนะส่วนใหญ่เป็นจาน/ชาม ขนาดเล็ก บางรูปประกอบด้วยภาชนะหลายชิ้น	ภาชนะส่วนใหญ่เป็นจาน/ชาม ขนาดเล็ก บางรูปประกอบด้วยภาชนะหลายชิ้นที่เหมือนกัน	ภาชนะส่วนใหญ่เป็นจาน/ชาม บางรูปมีถ้วยน้ำจิ้มวางข้างภาชนะหลัก

บทที่ 7
สรุปผลการวิจัยและแนวทางการวิจัยในขั้นถัดไป

7.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้ ได้นำเสนอแนวคิดและแบบจำลองที่ใช้จำแนกประเภทของรูปภาพอาหาร โดยใช้ข้อมูลรูปภาพจาก Wongnai ซึ่งเป็นแอปพลิเคชันสำหรับการอัปโหลดรูปภาพอาหาร โดยได้ทำการทดลองด้วยวิธีการจำแนกประเภทรูปภาพแบบละเอียด ที่มุ่งเน้นไปยังการคัดแยกรูปภาพระหว่างประเภทที่ยากต่อการจำแนก อันเนื่องมาจากการมีลักษณะที่คล้ายคลึงกัน ให้ได้ผลแม่นยำมากที่สุด

แบบจำลองของการจำแนกประเภทของรูปภาพอาหารที่วิทยานิพนธ์ฉบับนี้ได้เสนอ คือ B-CNN ซึ่งจากผลการทดลองพบว่า B-CNN ให้ผลลัพธ์ในการจำแนกประเภทของรูปภาพอาหารแม่นยำกว่าคอนโวลูชันเน็ตเวิร์กแบบทั่วไป เนื่องจากการคูณกันแบบการคูณภายนอกของสองเน็ตเวิร์กสามารถสกัดหาคุณลักษณะได้หลากหลาย และทำให้คุณลักษณะที่ได้มีความสัมพันธ์กันในทุกตำแหน่ง นอกจากการใช้นิวรอลเน็ตเวิร์กแบบคอนโวลูชันเชิงเส้นเพื่อสกัดหาคุณลักษณะของภาพแล้วนั้น แบบจำลองได้เสนอการใช้กลไกจุดสนใจ เพื่อทำการคัดเลือกคุณลักษณะที่คัดเลือกคุณลักษณะที่มีความเด่นชัดต่อการจำแนกรูปภาพนั้น ๆ อีกทั้งยังได้ทดลองเรื่องความละเอียดของภาพที่ส่งผลต่อการจำแนกประเภทของรูปภาพอาหาร ซึ่งผลการทดลองพบว่า การเพิ่มทั้งความละเอียดของภาพ และการเพิ่มกลไกจุดสนใจลงไปในแบบจำลอง ทำให้ประสิทธิภาพในการจำแนกประเภทของรูปภาพอาหารมีความแม่นยำมากขึ้นกว่าแบบจำลองตั้งต้น

สำหรับการเปรียบเทียบประสิทธิภาพของแบบจำลอง งานวิจัยนี้ได้เปรียบเทียบการทดลองออกเป็น 4 ส่วน ได้แก่ เปรียบเทียบระหว่าง B-CNN และ CNN เปรียบเทียบคอนโวลูชันเน็ตเวิร์กที่ใช้เป็นตัวสกัดคุณลักษณะใน B-CNN เปรียบเทียบประเภทของกลไกจุดสนใจที่ส่งผลต่อ B-CNN และเปรียบเทียบขนาดภาพของข้อมูลนำเข้าที่ส่งผลต่อประสิทธิภาพของแบบจำลอง ผลการทำลองพบว่าการนำขนาดภาพของข้อมูลนำเข้าที่ใหญ่ขึ้น ไปใช้กับ B-CNN โดยมีคอนโวลูชันเน็ตเวิร์กที่ต่างกันเป็นตัวสกัดคุณลักษณะ มาต่อด้วยการนำกลไกจุดสนใจประเภทที่ให้ค่าความสนใจแบบอ่อน สามารถช่วยเพิ่มประสิทธิภาพในการจำแนกประเภทของรูปภาพอาหารที่มีลักษณะคล้ายคลึงกันได้อย่างแม่นยำมากกว่าแบบจำลองตั้งต้น เนื่องจากแบบจำลองสามารถสกัดหาคุณลักษณะที่มีความสำคัญต่อประเภทอาหารนั้น ๆ จากคุณลักษณะที่หลากหลาย อีกทั้งแบบจำลองยังสามารถหาคุณลักษณะได้อย่างละเอียด และชัดเจนมากขึ้นจากความละเอียดของรูปภาพที่มากขึ้น

7.2 แนวทางการวิจัยถัดไป

สำหรับแนวทางในการวิจัยในอนาคต สามารถแบ่งได้ ดังต่อไปนี้

1. การนำข้อมูลเชิงตัวอักษรไปใช้ในแบบจำลอง

เนื่องจากในปัจจุบันการอัปโหลดภาพของผู้ใช้ มักถูกอัปโหลดพร้อมกับข้อความที่บรรยายเกี่ยวกับภาพดังกล่าว ไม่ว่าจะเป็นการอธิบายถึงลักษณะของภาพนั้น ๆ หรือบรรยายเกี่ยวกับความรู้สึกที่มีต่อภาพ แอปพลิเคชันเกี่ยวกับอาหารก็ถือเป็นแพลตฟอร์มหนึ่ง ที่ผู้ใช้สามารถอัปโหลดทั้งรูปภาพ และข้อความเข้าไปพร้อมกัน โดยข้อความเหล่านั้นทำให้ผู้ใช้คนอื่นที่เข้ามาค้นหารูปภาพอาหารเหล่านั้น สามารถทราบรายละเอียดเกี่ยวกับอาหารประเภทนั้น ๆ ได้มากขึ้นกว่าการเห็นเฉพาะแค่รูปถ่าย ไม่ว่าจะเป็น รสชาติของอาหาร ส่วนผสม รวมไปถึง ลักษณะของอาหาร ดังนั้นการนำข้อมูลรับเข้าเชิงตัวอักษรมาใช้ร่วมกับข้อมูลรับเข้าเชิงรูปภาพ ก็เป็นอีกแนวทางหนึ่งที่อาจจะสามารถช่วยเพิ่มประสิทธิภาพของงานจำแนกประเภทรูปภาพอาหาร เนื่องจากแบบจำลองจะสามารถสกัดคุณลักษณะของอาหารแต่ละประเภทได้มากขึ้น โดยเฉพาะอาหารบางประเภทที่มีลักษณะคล้ายคลึงกัน แบบจำลองจะสามารถคัดแยกเฉพาะคุณลักษณะที่จำเพาะต่ออาหารประเภทนั้น ๆ ได้ละเอียดกว่าการใช้ข้อมูลเชิงรูปภาพเพียงอย่างเดียว

2. การปรับปรุงคอนโวลูชันเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะ

นอกจากคอนโวลูชันเน็ตเวิร์กที่เป็นตัวสกัดคุณลักษณะที่ถูกนำมาใช้เป็นส่วนหนึ่งของแบบจำลองในวิทยานิพนธ์ชิ้นนี้นั้น ยังมีคอนโวลูชันเน็ตเวิร์กที่ถูกพัฒนาขึ้นมาใหม่ เพื่อช่วยในการจำแนกประเภทของรูปภาพได้อย่างมีประสิทธิภาพมากขึ้น หนึ่งในนั้นคือ EfficientNet ที่เป็นแบบจำลองที่กำลังเป็นที่นิยมในปัจจุบัน เนื่องจากเป็นแบบจำลองที่มีการปรับความกว้าง และความลึกของแบบจำลองไปพร้อม ๆ กันได้อย่างเหมาะสม นอกจากนั้นแบบจำลองนี้ยังมีจำนวนพารามิเตอร์ที่ลดลงจากแบบจำลองเดิมอีกด้วย โดยผู้เขียนได้ลองนำแบบจำลองดังกล่าวมาใช้ และพบว่าแบบจำลองนี้มีแนวโน้มที่จะให้ประสิทธิภาพที่ดีขึ้นหากนำมาพัฒนาต่อ

รายการอ้างอิง

[1] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained googlenet model," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 2016: ACM, pp. 3-11.

[2] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[3] Q. Yu, D. Mao, and J. Wang, "Deep Learning Based Food Recognition," ed, 2016.

[4] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Bilinear classifiers for visual recognition," in *Advances in neural information processing systems*, 2009, pp. 1482-1490.

[5] H. Chen, J. Wang, Q. Qi, Y. Li, and H. Sun, "Bilinear cnn models for food recognition," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2017: IEEE, pp. 1-6.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[8] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998-6008.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[10] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear cnn models for fine-grained visual recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1449-1457.

[11] S. Yan, J. S. Smith, W. Lu, and B. Zhang, "Hierarchical Multi-scale Attention Networks for action recognition," *Signal Processing: Image Communication,* vol. 61, pp. 73-84, 2018.

[12] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural computation,* vol. 12, no. 6, pp. 1247-1283, 2000.

[13] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531,* 2014.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[15] T.-Y. Lin and S. Maji, "Improved bilinear pooling with cnns," *arXiv preprint arXiv:1707.06772,* 2017.

[16] S. Kong and C. Fowlkes, "Low-rank bilinear pooling for fine-grained classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 365-374.

[17] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact bilinear pooling," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 317-326.

[18] Y. Cui, F. Zhou, J. Wang, X. Liu, Y. Lin, and S. Belongie, "Kernel pooling for convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2921-2930.

[19] S. Cai, W. Zuo, and L. Zhang, "Higher-order integration of hierarchical convolutional activations for fine-grained visual categorization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 511-520.

[20] C. Yu, X. Zhao, Q. Zheng, P. Zhang, and X. You, "Hierarchical bilinear pooling for fine-grained visual recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 574-589.

[21] A. R. Chowdhury, T.-Y. Lin, S. Maji, and E. Learned-Miller, "One-to-many face recognition with bilinear cnns," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016: IEEE, pp. 1-9.

[22] E. Ustinova, Y. Ganin, and V. Lempitsky, "Multi-region bilinear convolutional neural networks for person re-identification," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017: IEEE, pp. 1-6.

[23] C. Wang, J. Shi, Q. Zhang, and S. Ying, "Histopathological image classification with bilinear convolutional neural networks," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017: IEEE, pp. 4050-4053.

[24] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of applied meteorology,* vol. 18, no. 8, pp. 1016-1022, 1979.

บรรณานุกรม

ภาคผนวก

ตารางที่ 21 สถิติการจัดแบ่งประเภทชุดข้อมูลของอาหารแต่ละประเภท

ประเภทอาหาร	จำนวนรูปภาพ	ประเภทชุดข้อมูล		
		ฝึกสอน	ตรวจสอบ	ทดสอบ
honey_toast	652	518	68	66
กระเพาะปลา	520	425	42	53
กิมจิ	609	471	74	64
กุ้ง	5347	4258	494	595
กุ้งอบวุ้นเส้น	1502	1201	129	172
ก๋วยจั๊บ	1076	856	104	116
ก๋วยเตี๋ยว	1501	1206	150	145
ก๋วยเตี๋ยวต้มยำ	1295	1076	116	103
ขนมจีน	2015	1601	217	197
ขนมถ้วย	691	554	70	67
ขนมปัง	1849	1467	205	177
ขนมหวาน	627	479	85	63
ขาหมูเยอรมัน	746	589	73	84
ข้าว	2478	2024	223	231
ข้าวขาหมู	1072	886	97	89
ข้าวคลุกกะปิ	992	798	109	85
ข้าวซอยไก่	1168	920	97	151
ข้าวผัด	5168	4189	468	511
ข้าวมันไก่	2163	1745	234	184
ข้าวราดแกง	1750	1408	137	205
ข้าวหมกไก่	765	646	48	71
ข้าวหมูแดงหมูกรอบ	1958	1598	160	200
ข้าวเหนียวมะม่วง	904	725	70	109
คอหมูย่าง	2932	2317	299	316

ชาบู	598	474	88	36
ซุปเห็ด	556	455	63	38
ซูชิ	1446	1162	171	113
ตับหวาน	884	725	86	73
ติ่มซำ	7192	5797	688	707
ต้มยำ	3296	2722	242	332
ต้มเลือดหมู	811	653	79	79
ต้มแซ่บกระดูกอ่อน	1011	820	104	87
ทอดมัน	3256	2642	236	378
ทาโกะยากิ	582	446	50	86
น้ำจิ้ม	2294	1800	258	236
น้ำซุป	1155	911	140	104
น้ำตกหมู	742	594	66	82
น้ำพริกไข่ปู	617	492	41	84
น้ำเปล่า	1058	845	115	98
บะหมี่แห้ง	868	706	89	73
ปลาทอด	3262	2617	239	406
ปลานึ่ง	509	404	34	71
ปลาหมึกผัดไข่เค็ม	1782	1430	130	222
ปอเปี๊ยะทอด	626	513	44	69
ปูนิ่มทอดกระเทียม	650	528	50	72
ปูผัดผงกะหรี่	664	546	51	67
ปูม้านึ่ง	773	637	61	75
ผักสด	972	806	84	82
ผักโขมอบชีส	1262	1007	149	106
ผัดไทกุ้งสด	1618	1271	179	168
พิซซ่า	1196	898	189	109
ยำ	3561	2852	310	399
ลาบ	3048	2477	288	283
สปาเก็ตตี้	1453	1171	166	116

สลัด	2103	1609	268	226
สเต็กหมู	666	523	83	60
ส้มตำ	5699	4617	521	561
ส้มตำข้าวโพด	1336	1089	127	120
หมูกรอบ	1122	891	107	124
หมูมะนาว	656	527	66	63
หมูสะเต๊ะ	2070	1650	148	272
หมูแดดเดียว	894	718	86	90
หอย	2379	1863	256	260
ออส่วน	820	635	66	119
เครื่องดื่มร้อน	4465	3521	587	357
เครื่องดื่มเย็น	10996	8812	1215	969
เครื่องเคียง	935	776	87	72
เค้ก	1822	1414	195	213
เนื้อย่าง	670	558	52	60
เป็ด	1834	1428	187	219
เป็ดพะโล้	624	481	41	102
เย็นตาโฟ	918	765	79	74
แกง	1327	1089	106	132
แซลมอน	2486	1879	438	169
แหนมเนือง	1728	1397	127	204
ใบเหลียงผัดไข่	968	778	108	82
ไก่ทอด	3927	3151	431	345
ไก่ย่าง	1934	1592	163	179
ไข่กระทะ	745	596	104	45
ไข่ตุ๋น	562	415	75	72
ไข่เจียว	1659	1359	132	168
ไส้อั่ว	578	450	71	57
ไอศกรีม	718	554	105	59
รวม	144,163	115,495	14,290	14,378

ตารางที่ 22 ผลการทดลองอย่างละเอียดของแบบจำลองที่นำเสนอ

Classes	f1-score	precision	recall	support
honey toast	90.73%	84.89%	95.45%	66
กระเพาะปลา	88.49%	84.76%	92.57%	53
กิมจิ	89.55%	88.57%	92.63%	64
กุ้ง	96.43%	98.33%	94.61%	595
กุ้งอบวุ้นเส้น	97.04%	100.80%	97.35%	172
ก๋วยจั๊บ	84.11%	83.65%	78.72%	116
ก๋วยเตี๋ยว	82.42%	83.56%	81.31%	145
ก๋วยเตี๋ยวต้มยำ	79.55%	83.72%	75.79%	103
ขนมจีน	84.12%	83.50%	84.74%	197
ขนมถ้วย	94.31%	97.24%	91.55%	67
ขนมปัง	82.00%	86.81%	77.71%	177
ขนมหวาน	57.45%	65.29%	59.52%	63
ขาหมูเยอรมัน	86.34%	87.37%	85.33%	84
ข้าว	93.21%	90.26%	96.37%	231
ข้าวขาหมู	81.82%	75.75%	68.29%	89
ข้าวคลุกกะปิ	90.76%	86.95%	94.94%	85
ข้าวซอยไก่	94.96%	101.25%	89.42%	151
ข้าวผัด	97.00%	95.38%	98.67%	511
ข้าวมันไก่	88.33%	87.19%	89.50%	184
ข้าวราดแกง	73.29%	74.36%	72.24%	205
ข้าวหมกไก่	84.54%	96.55%	75.24%	71
ข้าวหมูแดงหมูกรอบ	91.17%	91.85%	90.50%	200
ข้าวเหนียวมะม่วง	92.91%	97.00%	89.16%	109
คอหมูย่าง	91.70%	91.84%	91.56%	316
ซาบู	58.10%	52.00%	65.89%	36
ซุปเห็ด	77.61%	72.45%	83.58%	38

ซูชิ	85.53%	78.47%	94.04%	113
ตับหวาน	91.19%	90.00%	92.41%	73
ติ่มซำ	95.26%	86.76%	97.90%	707
ต้มยำ	90.59%	89.17%	92.06%	332
ต้มเลือดหมู	82.70%	77.00%	89.34%	79
ต้มแซ่บกระดูกอ่อน	79.46%	68.51%	77.52%	87
ทอดมัน	95.92%	93.90%	98.03%	378
ทาโกะยากิ	92.12%	98.05%	86.88%	86
น้ำจิ้ม	84.11%	80.91%	87.59%	236
น้ำซุป	80.10%	79.36%	80.85%	104
น้ำตกหมู	78.71%	89.50%	70.29%	82
น้ำพริกไข่ปู	89.36%	86.44%	92.48%	84
น้ำเปล่า	89.83%	93.21%	86.69%	98
บะหมี่แห้ง	90.16%	86.81%	93.78%	73
ปลาทอด	98.10%	100.20%	96.09%	406
ปลานึ่ง	89.59%	92.91%	86.51%	71
ปลาหมึกผัดไข่เค็ม	94.77%	98.14%	91.64%	222
ปอเปี๊ยะทอด	84.43%	71.22%	90.41%	69
ปูนิ่มทอดกระเทียม	84.89%	80.75%	89.50%	72
ปูผัดผงกะหรี่	84.26%	91.47%	85.12%	67
ปูม้านึ่ง	97.30%	97.95%	96.67%	75
ผักสด	78.62%	83.94%	73.95%	82
ผักโขมอบชีส	95.33%	96.23%	94.45%	106
ผัดไทกุ้งสด	96.52%	93.62%	99.62%	168
พิซซ่า	90.48%	90.89%	90.07%	109
ยำ	89.74%	86.30%	93.48%	399
ลาบ	93.50%	90.20%	97.05%	283
สปาเก็ตตี้	93.06%	91.92%	94.24%	116
สลัด	84.82%	79.82%	90.50%	226
สเต็กหมู	80.50%	91.36%	72.00%	60

ส้มตำ	99.28%	100.90%	97.72%	561
ส้มตำข้าวโพด	98.30%	97.12%	99.50%	120
หมูกรอบ	83.97%	85.33%	82.65%	124
หมูมะนาว	82.34%	89.04%	76.60%	63
หมูสะเต๊ะ	96.14%	98.89%	93.54%	272
หมูแดดเดียว	90.24%	95.75%	85.33%	90
หอย	97.07%	99.57%	94.69%	260
ออส่วน	90.14%	90.89%	89.39%	119
เครื่องดื่มร้อน	91.39%	87.14%	91.64%	357
เครื่องดื่มเย็น	96.27%	94.63%	97.98%	969
เครื่องเคียง	58.06%	63.67%	63.39%	72
เค้ก	86.56%	77.58%	85.57%	213
เนื้อย่าง	45.24%	49.06%	42.00%	60
เป็ด	88.13%	94.19%	82.82%	219
เป็ดพะโล้	82.21%	90.24%	85.53%	102
เย็นตาโฟ	90.31%	87.00%	93.89%	74
แกง	89.94%	92.40%	87.61%	132
แซลมอน	93.45%	93.18%	93.72%	169
แหนมเนือง	90.73%	88.85%	92.69%	204
ใบเหลียงผัดไข่	97.12%	97.12%	97.12%	82
ไก่ทอด	86.73%	86.24%	87.22%	345
ไก่ย่าง	85.55%	80.43%	91.39%	179
ไข่กระทะ	93.67%	88.27%	99.78%	45
ไข่ตุ๋น	80.46%	89.93%	72.83%	72
ไข่เจียว	94.22%	91.39%	97.24%	168
ไส้อั่ว	90.50%	91.29%	89.72%	57
ไอศกรีม	77.19%	69.57%	86.75%	59
macro avg	87.72%	87.34%	87.37%	14378
micro avg	90.51%	90.51%	90.51%	14378
weighted avg	90.35%	90.51%	90.51%	14378

ประวัติผู้เขียน

ชื่อ-สกุล	วศิณี นุชศิริ
วัน เดือน ปี เกิด	27 กรกฎาคม 2537
สถานที่เกิด	เชียงใหม่
วุฒิการศึกษา	วท.บ. (เกียรตินิยมอันดับสอง) คณะวิทยาศาสตร์ ภาควิชาคณิตศาสตร์ และวิทยาการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย (พ.ศ. 2556 - 2560)
ที่อยู่ปัจจุบัน	131/596 ไอดีโอ ท่าพระ อินเตอร์เชนจ์ ถนน เพชรเกษม แขวง วัดท่าพระ เขต บางกอกใหญ่ กรุงเทพมหานคร 10600
ผลงานตีพิมพ์	V. Nussiri and P. Vateekul, "Food Image Categorization Using Attentional Bilinear Model," in 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), 2019: IEEE, pp. 1-6. P. Pugsee, V. Nussiri, and W. Kittirungruang, "Opinion Mining for Skin Care Products on Twitter," in International Conference on Soft Computing in Data Science, 2018: Springer, pp. 261-271.