

การสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึกสำหรับการค้นหารูปภาพย้อนกลับ

จักรินทร์ สันติรัตนภักดี^{1*} ศุภกฤษฎี นวัตกรรมกุล²

Received : January 10, 2025

Revised : August 6, 2025

Accepted : August 27, 2025

บทคัดย่อ

การสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึกสำหรับการค้นหารูปภาพย้อนกลับ มีวัตถุประสงค์เพื่อพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก และเพื่อประเมินความแม่นยำของผลการค้นหารูปภาพย้อนกลับ งานวิจัยนี้ใช้โมเดล ResNet50 ที่ผ่านการถ่ายโอนความรู้จากการฝึกฝนล่วงหน้า และปรับแต่งโมเดลให้กลายเป็นตัวสกัดคุณลักษณะบนชุดข้อมูลรูปภาพอาหารไทยจำนวน 20 รายการที่ได้รับความนิยมในระดับนานาชาติ เพื่อสร้างเป็นชุดข้อมูลแทนความหมายของภาพสำหรับนำไปเปรียบเทียบกับรูปภาพค้นหาด้วยการวัดค่าความคล้ายคลึงโคไซน์ ส่งผลให้สามารถค้นหารูปภาพที่ต้องการได้อย่างรวดเร็วและแม่นยำโดยไม่จำเป็นต้องใช้ข้อมูลป้ายกำกับ ผลการประเมินความแม่นยำของการค้นหารูปภาพย้อนกลับ โดยเฉพาะ 3 อันดับแรกมีความแม่นยำสูงถึงร้อยละ 80 เมื่อเพิ่มจำนวนผลลัพธ์เป็น 5 และ 10 รูปภาพ ความแม่นยำยังคงอยู่ในระดับที่ดี ผลลัพธ์ดังกล่าวสอดคล้องกับพฤติกรรมของผู้ใช้ที่มักสนใจเพียงผลลัพธ์ในอันดับต้น ๆ เท่านั้น อย่างไรก็ตาม ความคล้ายคลึงของภาพที่มีลักษณะใกล้เคียงกัน ตลอดจนความแปรปรวนของมุมมอง ขนาด และสภาพแสง รวมถึงความซับซ้อนของพื้นหลังยังส่งผลให้การจดจำคุณลักษณะเฉพาะมีความคลาดเคลื่อนได้ ผลผลิตจากงานวิจัยนี้คือการประเมินประสิทธิภาพของโมเดลในสถานการณ์จริง อันจะเป็นแนวทางในการค้นหาภาพในระบบพาณิชย์อิเล็กทรอนิกส์ หรือการระบุภาพซ้ำในสื่อออนไลน์ เป็นต้น

คำสำคัญ: การค้นหารูปภาพ การค้นหารูปภาพย้อนกลับ การเรียนรู้เชิงลึก การสกัดคุณลักษณะ

¹ อาจารย์ประจำหลักสูตรบริหารธุรกิจบัณฑิต สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยวังษวลิตกุล
อีเมล: chakkarin_san@vu.ac.th

² ผู้ช่วยศาสตราจารย์ ดร. สำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี อีเมล: suphakit@sut.ac.th

* ผู้มีพันธหลักอีเมล: chakkarin_san@vu.ac.th

IMAGE FEATURE EXTRACTION BY DEEP LEARNING MODELS FOR REVERSE IMAGE SEARCH

Chakkarin Santirattanaphakdi^{1*} Suphakit Niwattanakul²**Abstract**

Image feature extraction by deep learning models for reverse image search was aimed to develop a model of the image feature extraction employing deep learning models and evaluate the precise results of reverse image search. This research utilized ResNet50 model, which has been pre-trained transfer learning then well-tuned as the feature extractor for a dataset of 20 Thai-food-image categories that have been internationally popular. The processes mentioned were constructed as a dataset representing a semantic image for comparing with search images using cosine similarity measurement. These processes enabled the fast and accurate image retrieval without the need for labeled data. The research resulted showed the precise evaluation of the reverse image search, especially for the first three results achieving an 80-percent precision. When increasing the number of retrieved results to 5 and 10 images, the precision was at a good level. These results aligned with users' behavior, who have typically focused only on the top-ranked results. However, similarity among visually alike images, variations in viewpoint, scale, illumination, including background clutter affected errors in recognizing distinctive features. The outcome of this research was an evaluation of the performance of the model employing in the real-situation scenarios, which could serve as a guideline for developing image retrieval systems in e-commerce or duplicating image identification for online media.

Keywords: Image retrieval, Reverse image search, Deep learning, Feature extraction

¹ Lecturer, Bachelor of Business Administration Program in Digital Business Technology, Faculty of Business Administration, Vongchavalitkul University, e-mail: chakkarin_san@vu.ac.th

² Associate Prof. Dr., Institute of Digital Arts and Science, Suranaree University of Technology, e-mail: suphakit@sut.ac.th

* Corresponding author, e-mail: chakkarin_san@vu.ac.th

บทนำ

ยุคดิจิทัลที่เทคโนโลยีการถ่ายภาพพัฒนาอย่างรวดเร็ว ส่งผลให้ปริมาณของรูปภาพเพิ่มขึ้นอย่างต่อเนื่อง โดยในปีค.ศ. 2024 คาดการณ์ว่าจะมีรูปภาพทั่วโลกมากถึง 1.94 ล้านล้านภาพ ตัวเลขดังกล่าวสะท้อนถึงพฤติกรรมการถ่ายภาพในชีวิตประจำวันที่เพิ่มขึ้นอย่างชัดเจน ซึ่งได้รับแรงผลักดันจากเทคโนโลยีกล้องถ่ายรูปของสมาร์ทโฟนที่สามารถถ่ายภาพที่มีคุณภาพสูงได้ในทุกสถานการณ์ ขณะเดียวกันยังเข้าถึงง่ายและราคาย่อมเยา คาดการณ์ว่าจำนวนรูปภาพจะเพิ่มขึ้นอย่างต่อเนื่องเป็น 2 ล้านล้านภาพในปีค.ศ. 2025 และจะเพิ่มขึ้นอย่างก้าวกระโดดเป็น 3 ล้านล้านภาพภายในปีค.ศ. 2030 (Broz, 2024) จากปัจจัยสนับสนุนจากหลายประการ เช่น เทคโนโลยีการจัดเก็บรูปภาพบนคลาวด์ และการใช้โซเชียลมีเดียที่แพร่หลาย

รูปภาพไม่ได้เป็นเพียงแค่สื่อกลางในการเก็บความทรงจำส่วนบุคคลเท่านั้น แต่ยังกลายเป็นเครื่องมือสำคัญในการสร้างเนื้อหาออนไลน์ที่มีประสิทธิภาพ อย่างไรก็ตาม รูปภาพในรูปแบบข้อมูลดิบ (Raw data) มักมีขนาดใหญ่และซับซ้อน ส่งผลให้การจัดเก็บและการประมวลผลมีต้นทุนสูง ทั้งในแง่ของทรัพยากรทางคอมพิวเตอร์และเวลาในการประมวลผล โดยการค้นหาภาพในฐานข้อมูลขนาดใหญ่ยังเป็นหนึ่งในอุปสรรคสำคัญในการนำรูปภาพมาใช้ประโยชน์ เนื่องจากรูปภาพอาจไม่ได้รับการจัดหมวดหมู่ที่ชัดเจนหรือไม่มีการระบุป้ายกำกับที่เหมาะสม ส่งผลให้การค้นหาภาพที่ตรงกับความต้องการเป็นเรื่องยากและใช้เวลาประมวลผลมากขึ้น (Goodfellow et al., 2016) ดังนั้นการใช้เทคนิคการสกัดคุณลักษณะ (Feature extraction) จึงมีความสำคัญในการแยกคุณลักษณะที่สำคัญของภาพออกจากข้อมูลดิบ เพื่อลดมิติความซับซ้อน และทำให้การค้นหาหรือการนำไปใช้งานต่าง ๆ เป็นไปได้อย่างรวดเร็วและมีประสิทธิภาพมากยิ่งขึ้น

ปัจจุบันการใช้โมเดลการเรียนรู้เชิงลึก (Deep learning) เพื่อการสกัดคุณลักษณะภาพได้รับการยอมรับว่ามีประสิทธิภาพสูงในการแก้ไขปัญหาดังกล่าว โดยเฉพาะกับภาพที่มีความซับซ้อนสูงและต้องการการประมวลผลที่รวดเร็วและแม่นยำ จากการนำโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ซึ่งมีความสามารถในการสกัดคุณลักษณะภาพโดยอัตโนมัติในหลายระดับ ตั้งแต่คุณลักษณะระดับต่ำ (Low-level feature) เช่น สี พื้นผิว และรูปร่าง ไปจนถึงคุณลักษณะระดับสูง (High-level feature) ที่สามารถอธิบายวัตถุหรือองค์ประกอบที่อยู่ในภาพ (LeCun et al., 2015) ซึ่งทำให้โมเดลสามารถเรียนรู้จากข้อมูลและสกัดคุณลักษณะที่เป็นประโยชน์สำหรับงานที่เฉพาะเจาะจง เนื่องจากสามารถเรียนรู้ความสัมพันธ์กันของคุณลักษณะในภาพได้โดยไม่ต้องอาศัยผู้เชี่ยวชาญในการคัดเลือกคุณลักษณะ ตลอดจนสามารถลดความซับซ้อนของมิติข้อมูลจำนวนมากให้เหลือเฉพาะคุณลักษณะที่สำคัญ จึงมักถูกนำไปใช้ในการวิเคราะห์และประมวลผลภาพ เช่น การจำแนกรูปภาพ (Image classification) การตรวจจับวัตถุ (Object detection) และการแปลความหมายภาพ (Image segmentation) ตลอดจนการค้นหารูปภาพ (Image retrieval) ซึ่งเป็นกระบวนการที่ช่วยให้สามารถค้นหารูปภาพที่มีความคล้ายคลึงหรือสอดคล้องกับ

รูปภาพค้นหา (Image query) ที่ผู้ใช้ส่งเข้ามาได้อย่างแม่นยำ (Mishra et al., 2020) จากการสกัดคุณลักษณะจากภาพ แล้วเปรียบเทียบความคล้ายคลึงกันของคุณลักษณะในภาพบนฐานข้อมูลภาพขนาดใหญ่ โดยไม่จำเป็นต้องพึ่งพาคำค้นหา (Keyword) หรือข้อความที่ผู้ใช้ป้อนเข้ามา (Razavian et al., 2014) ดังนั้นจึงสามารถจัดการกับความซับซ้อนของข้อมูลภาพที่หลากหลาย และไม่เป็นระเบียบได้ดีกว่าการใช้วิธีการค้นหาภาพแบบดั้งเดิม ทำให้สามารถค้นหารูปภาพที่มีลักษณะคล้ายคลึงกันได้อย่างมีประสิทธิภาพ และสามารถนำไปประยุกต์ใช้ในหลากหลายบริบท

หนึ่งในรูปแบบการค้นหารูปภาพที่มีการใช้งานเพิ่มมากขึ้นในปัจจุบันคือการค้นหาภาพย้อนกลับ (Reverse image search) โดยเฉพาะการค้นหาภาพที่คล้ายคลึงกันจากรูปภาพค้นหาบนฐานข้อมูลขนาดใหญ่ เทคโนโลยีนี้ได้รับความนิยมอย่างมากในการค้าปลีกออนไลน์ การค้นหาภาพทางการแพทย์ หรือการวิจัยทางวิทยาศาสตร์ (Deng et al., 2009) โดยเฉพาะการค้นหารูปภาพอาหารไทยที่มีความคล้ายคลึงกันนับว่าเป็นประเด็นที่ได้รับความสนใจในงานวิจัยด้านปัญญาประดิษฐ์และคอมพิวเตอร์วิทัศน์ (Computer vision) เนื่องจากอาหารไทยเป็นตัวแทนของวัฒนธรรมและเอกลักษณ์ประจำชาติที่มีความหลากหลายและรายละเอียดเฉพาะตัว ดังนั้นอาหารไทยจึงไม่ได้เป็นเพียงอาหารประจำชาติ แต่ยังมีบทบาทสำคัญในการส่งเสริมการท่องเที่ยวทั้งทางตรงและทางอ้อม โดยเฉพาะในการเสริมสร้างภาพลักษณ์ของประเทศไทยในระดับสากล และสะท้อนถึงความหลากหลายและเอกลักษณ์ทางวัฒนธรรม รวมถึงภูมิปัญญาท้องถิ่นที่ทำให้อาหารไทยได้รับการยอมรับและชื่นชมในระดับโลก อย่างไรก็ตาม กระบวนการค้นหารูปภาพที่มีลักษณะคล้ายคลึงกันนั้นเผชิญกับความท้าทายหลายด้านที่ต้องได้รับการพิจารณา เนื่องจากความหลากหลายของรายการอาหาร และเอกลักษณ์เฉพาะตัวของอาหารแต่ละรายการ เช่น ผัดไทย และผัดซีอิ้วมีความคล้ายคลึงกันในบางแง่มุม เช่น เส้นที่ใช้หรือวิธีการปรุงอาหาร แต่ยังคงมีความแตกต่างในรายละเอียด เช่น ชนิดของซอส ผักที่ใช้และการตกแต่งบนจานที่เป็นเสน่ห์ที่ทำให้อาหารไทยได้รับความนิยมในระดับนานาชาติ (Tourism Authority of Thailand, 2021) นอกจากนั้นบริบทของภาพถ่ายอาหารที่มักมีพื้นหลังและองค์ประกอบอื่น ๆ เช่น จาน ช้อน หรือเครื่องเคียง ซึ่งอาจรบกวนการระบุลักษณะสำคัญของอาหารในภาพ อีกทั้งสภาพแสงในภาพ เช่น แสงธรรมชาติ หรือแสงประดิษฐ์ อาจส่งผลต่อประสิทธิภาพของโมเดลที่ใช้ประมวลผลภาพ โดยเฉพาะในกรณีที่แสงเปลี่ยนแปลงอย่างมากระหว่างภาพถ่าย

อย่างไรก็ตาม การค้นหารูปภาพย้อนกลับโดยทั่วไปจะมุ่งที่ความคล้ายคลึงเพียงแค่อัตลักษณ์ของรูปภาพที่ตรงกับรูปภาพค้นหาเท่านั้น หากแต่พฤติกรรมของผู้ใช้มักจะสนใจบริบทของผลลัพธ์จากการค้นหามากกว่า (Kiros et al., 2014) เช่นเดียวกับความสนใจของผู้ใช้ที่มักสนใจผลลัพธ์ในอันดับต้น ๆ เท่านั้น และไม่ต้องการผลลัพธ์จำนวนมากเกินไปอาจทำให้เสียเวลาในการคัดเลือกรูปภาพที่ตรงกับความต้องการได้ ความหลากหลายเหล่านี้ทำให้การพัฒนาโมเดลเพื่อค้นหาภาพที่คล้ายคลึงกันจำเป็นต้องมีความแม่นยำสูงในการดึงคุณลักษณะสำคัญที่ใช้แยกแยะรายละเอียดต่าง ๆ เพื่อนำไปประยุกต์ใช้ในแอปพลิเคชันแนะนำรายการอาหาร เพื่อให้

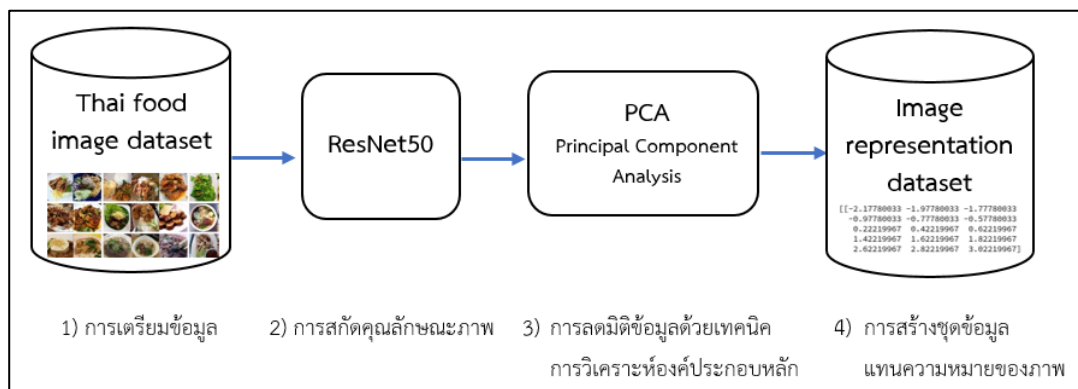
สอดคล้องกับแนวคิดการท่องเที่ยวเชิงอาหาร (Gastronomy tourism) ที่มุ่งเน้นการสัมผัสประสบการณ์ด้านอาหารและวัฒนธรรมการกินของแต่ละท้องถิ่น ซึ่งถือเป็นหนึ่งในซอฟต์แวร์ (Soft power) ที่แข็งแกร่งของประเทศไทย งานวิจัยนี้จึงมุ่งพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก และประเมินความแม่นยำของผลการค้นหาภาพย้อนกลับบนชุดข้อมูลรูปภาพอาหารไทยจำนวน 20 รายการที่ได้รับความนิยมในระดับนานาชาติ อันจะช่วยค้นหาภาพที่ตรงกับความต้องการจากฐานข้อมูลขนาดใหญ่อย่างมีประสิทธิภาพด้วยการใช้รูปภาพค้นหาที่เชื่อมโยงกับอาหารแต่ละรายการ อีกทั้งยังเป็นการเพิ่มประสบการณ์ที่ดีในการสนับสนุนความต้องการของผู้ใช้ด้วยการให้ข้อมูลอาหารไทยที่ถือเป็นหนึ่งในเอกลักษณ์ทางวัฒนธรรมของประเทศไทย

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก
2. เพื่อประเมินความแม่นยำของผลการค้นหาภาพย้อนกลับ

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงประยุกต์เพื่อพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึกสำหรับการค้นหาภาพย้อนกลับด้วยการเขียนโปรแกรมภาษาไพทอน (Python) บน Google Colaboratory แบบมีค่าใช้จ่ายที่เพิ่มทรัพยากรการประมวลผลที่เร็วขึ้น หน่วยความจำขนาดใหญ่ขึ้น และเวลาการเชื่อมต่อที่ยาวนานขึ้น เหมาะสำหรับการฝึกโมเดลการเรียนรู้ของเครื่องที่ต้องใช้ทรัพยากรการประมวลผลประกอบด้วย 4 ขั้นตอน ได้แก่ การเตรียมข้อมูล การสกัดคุณลักษณะภาพ การลดมิติข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก และการสร้างชุดข้อมูลแทนความหมายของภาพ ดังภาพที่ 1



ภาพที่ 1 ขั้นตอนการดำเนินการพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก

1) การเตรียมชุดข้อมูล (Dataset preparation) ที่จำเป็นสำหรับการวิเคราะห์รูปภาพในการพัฒนาตัวแบบการสกัดคุณลักษณะ และการค้นหารูปภาพแบบย้อนกลับบนชุดข้อมูล FoodDudy (GemmyTheGeek, 2021) ได้รับการพัฒนาขึ้นเพื่อสนับสนุนความต้องการของนักท่องเที่ยวและผู้บริโภคที่ไม่คุ้นเคยกับอาหารไทยด้วยการให้ข้อมูลเกี่ยวกับระดับความเผ็ด และส่วนประกอบสำคัญในอาหารแต่ละจาน ซึ่งมีประโยชน์ต่อผู้ที่มีข้อจำกัดด้านอาหาร เช่น การแพ้สารอาหารบางชนิด ชุดข้อมูลประกอบด้วยภาพอาหารไทย 48 รายการ จำนวนมากกว่า 14,000 ภาพ รวบรวมข้อมูลเริ่มต้นจากการค้นหาภาพผ่านแพลตฟอร์มออนไลน์ เช่น Google และ Bing โดยใช้คำค้นที่ครอบคลุมชื่ออาหารที่ได้รับความนิยมในระดับนานาชาติ เช่น Tom Yum, Pad Thai, และ Green Curry โดยผ่านการคัดกรองความเหมาะสมของภาพเพื่อให้ได้ชุดข้อมูลที่สมบูรณ์ และผ่านการตรวจสอบคุณภาพความคมชัดให้เป็นไปตามมาตรฐาน เพื่อสนับสนุนการสร้างโมเดลปัญญาประดิษฐ์ที่เกี่ยวข้องกับอาหารไทยในบริบทต่าง ๆ

งานวิจัยนี้คัดเลือกรายการอาหารไทยแบบเฉพาะเจาะจง 20 รายการจากจำนวนทั้งหมด 40 อันดับที่ได้รับการแนะนำเป็นอาหารไทยที่ไม่ควรพลาด (40 Thai foods we can't live without) (Wiens, 2017) จากเว็บไซต์ CNN Travel ได้แก่ 1) ต้มยำ 2) แกงส้ม 3) แกงเขียวหวาน 4) พะแนง 5) แกงมัสมั่น 6) แกงจืด 7) ไข่เจียว 8) หมูย่าง 9) ข้าวขาหมู 10) ข้าวหมกไก่ 11) ข้าวมันไก่ 12) ลาบ 13) ผัดกะเพรา 14) ผัดเม็ดมะม่วง 15) ส้มตำ 16) ผัดผักบุ้ง 17) ผัดซีอิ้ว 18) ผัดไทย 19) ข้าวเหนียวมะม่วง และ 20) ส้มขี้ลิงทอด ดังตารางที่ 1 โดยสุ่มรูปภาพอาหารจากชุดข้อมูล FoodDudy รายการละ 100 รูปภาพ รวมทั้งสิ้น 2,000 รูปภาพ มาสร้างเป็นชุดข้อมูลรายการอาหารแบบไม่มีป้ายกำกับข้อมูล (Unlabelled data) เพื่อเน้นการพัฒนาาระบบที่สามารถค้นหาหรือเปรียบเทียบความคล้ายคลึงระหว่างรูปภาพได้ทันทีโดยไม่ต้องใช้ป้ายกำกับเพื่อระบุคลาส ก่อนจะปรับภาพภาพทั้งหมดให้ขนาดเท่ากันเป็น 224*224 พิกเซล ซึ่งเป็นขนาดมาตรฐานที่นิยมใช้ เนื่องจากภาพที่มีขนาดไม่เท่ากันจะส่งผลให้การประมวลผลข้อมูลขาดความเป็นระบบ และเพิ่มภาระในการคำนวณของโมเดล

ตารางที่ 1 รายการอาหารไทยจำนวน 20 รายการ และตัวอย่างรูปภาพจากชุดข้อมูล FoodDudy

รายการอาหาร	ตัวอย่างรูปภาพอาหาร					
	1	2	3	4	...	100
1) ต้มยำ (Spicy soup)					...	
2) แกงส้ม (Sour soup)					...	

ตารางที่ 1 รายการอาหารไทยจำนวน 20 รายการ และตัวอย่างรูปภาพจากชุดข้อมูล FoodDudy (ต่อ)

รายการอาหาร	ตัวอย่างรูปภาพอาหาร					
	1	2	3	4	...	100
3) แกงเขียวหวาน (Green curry)					...	
4) พะแนง (Panang curry)					...	
5) แกงมัสมั่น (Massaman curry)					...	
6) แกงจืด (Soup)					...	
7) ไข่เจียว (Omelet)					...	
8) หมูย่าง (Grilled pork skewers)					...	
9) ข้าวขาหมู (Thai pork leg stew)					...	
10) ข้าวหมกไก่ (Thai chicken biryani)					...	
11) ข้าวมันไก่ (Hainanese chicken rice)					...	
12) ลาบ (Spicy minced salad)					...	
13) ผัดกะเพรา (Stir fried Thai basil)					...	
14) ผัดเม็ดมะม่วง (Braised chestnut)					...	

ตารางที่ 1 รายการอาหารไทยจำนวน 20 รายการ และตัวอย่างรูปภาพจากชุดข้อมูล FoodDudy (ต่อ)

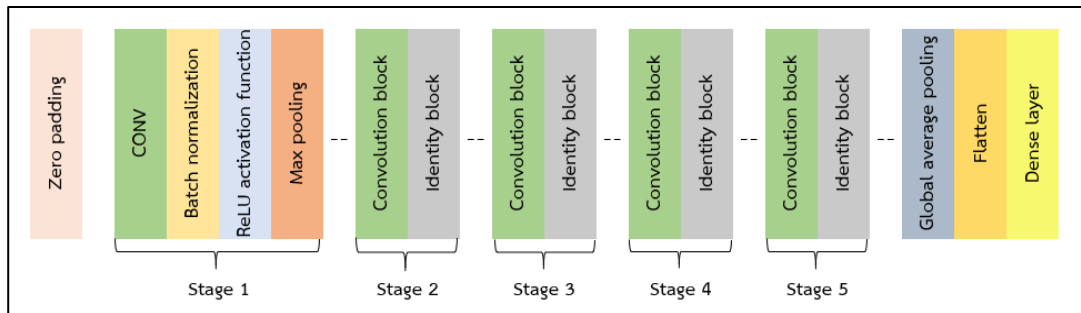
รายการอาหาร	ตัวอย่างรูปภาพอาหาร					
	1	2	3	4	...	100
15) ส้มตำ (Green papaya salad)					...	
16) ผัดผักบุ้ง (Stir-fried morning glory)					...	
17) ผัดซีอิ้ว (Fried noodle in soy sauce)					...	
18) ผัดไทย (Thai stir-fried noodles)					...	
19) ข้าวเหนียวมะม่วง (Mango sticky rice)					...	
20) สังขยาฟักทอง (Egg custard in pumpkin)					...	

2) การสกัดคุณลักษณะภาพ (Image feature extraction) ด้วยโมเดลที่ได้รับการฝึกฝนล่วงหน้า (Pre-trained model) เพื่อนำมาใช้พัฒนาโมเดลสำหรับการจำแนกรูปภาพอาหารไทยด้วยความสามารถในการถ่ายโอนความรู้ (Transfer learning) จากการฝึกฝนบนชุดข้อมูลขนาดใหญ่ ร่วมกับการปรับแต่งโมเดลเพิ่มเติม (Fine-tuning) ให้เหมาะสมกับชุดข้อมูลด้วยการปรับค่าน้ำหนักบางส่วนของโมเดล หรือฝึกฝนเฉพาะบางชั้นของโครงข่าย ทำให้สามารถใช้งานได้อย่างมีประสิทธิภาพแม้ในสภาพแวดล้อมที่แตกต่างจากชุดข้อมูลเดิม ปัจจุบันมีหลายโมเดลที่ได้รับการฝึกฝนล่วงหน้าที่น่าสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน มาพัฒนาเพื่อจำแนกรูปภาพ ในเบื้องต้นคณะผู้วิจัยได้คัดเลือกโมเดลเพื่อทดสอบกับชุดข้อมูลรูปภาพคน สัตว์ สิ่งของ และสถานที่ที่ที่คณะผู้วิจัยรวบรวมเอง จำนวน 4,512 ภาพ โดยทำการประเมินจนครบ 3 รอบ แบบผลลัพธ์เป็นอิสระต่อกัน แล้วนำผลการประเมินมาหาค่าเฉลี่ยเพื่อให้ความคลาดเคลื่อนลดน้อยลงจนถึงในระดับที่ยอมรับได้ ผลการประเมิน พบว่า ทั้ง 5 โมเดล ประกอบด้วย ResNet50, VGG16, MobileNetV3, DenseNet121 และ SqueezeNet นั้นมีความถูกต้อง (Accuracy) ในการจำแนกข้อมูลเท่ากับร้อยละ 81.06, 81.03, 80.03, 81.05 และ 79.01 ตามลำดับ ถือว่าอยู่ในระดับที่สูงและค่าความถูกต้องทั้งหมดใกล้เคียงกัน

อย่างไรก็ดี แต่ละโมเดลนั้นมีความแตกต่างกันอย่างชัดเจน ทั้งในด้านความเร็วและขนาดของโมเดล ดังนั้นในงานวิจัยนี้จึงเลือกใช้โมเดล ResNet50 เป็นโครงสร้างพื้นฐาน (Baseline model) เนื่องจากมีประสิทธิภาพสูงในการประมวลผลภาพด้วยจำนวนพารามิเตอร์ประมาณ 25 ล้านตัว ซึ่งน้อยกว่าโมเดลอื่นในรุ่นเดียวกันอย่าง VGG-16 แต่ให้ผลลัพธ์ที่มีความแม่นยำสูงกว่าในการทดสอบประสิทธิภาพ นอกจากนี้ยังเป็นสถาปัตยกรรมที่เหมาะสมสำหรับการถ่ายโอนความรู้ เนื่องจากมีน้ำหนักที่ผ่านการฝึกไว้ล่วงหน้า (Pre-trained weights) ให้เลือกใช้งานหลากหลาย สามารถนำไปปรับใช้กับชุดข้อมูลใหม่ได้อย่างรวดเร็วและแม่นยำ โดยไม่จำเป็นต้องฝึกฝนโมเดลใหม่ตั้งแต่ต้น ด้วยความสามารถในการเรียนรู้คุณลักษณะที่ซับซ้อนทำให้ ResNet-50 เป็นหนึ่งในโมเดลพื้นฐานที่ได้รับการยอมรับอย่างกว้างขวางในวงการปัญญาประดิษฐ์และการเรียนรู้ของเครื่อง

โมเดล ResNet50 (Residual Network 50 layers) (He et al., 2016) เป็นโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ผ่านการถ่ายโอนความรู้จากฝึกฝนล่วงหน้าบนชุดข้อมูล ImageNet โดยออกแบบสถาปัตยกรรมเพื่อแก้ปัญหาที่เกิดขึ้นเมื่อเพิ่มจำนวนเลเยอร์ในโครงข่ายที่มีความลึกสูง เนื่องจากในอดีตการเพิ่มความลึกของชั้นโครงข่ายถูกคาดหวังว่าจะช่วยให้โมเดลสามารถเรียนรู้ลักษณะเฉพาะของข้อมูลได้ละเอียดและแม่นยำมากขึ้น แต่กลับพบปัญหาสำคัญ คือความแม่นยำของโมเดลลดลงเมื่อความลึกเพิ่มขึ้น (Degradation problem) และปัญหาค่า Gradient ที่ลดลงจนแทบไม่สามารถอัปเดตพารามิเตอร์ในเลเยอร์แรก ๆ ของโครงข่ายได้ (Vanishing gradient problem) จึงเป็นที่มาของการนำแนวคิด Residual learning มาเพื่อแก้ไขข้อจำกัดดังกล่าว โดยเปลี่ยนเป้าหมายของโครงข่ายจากการเรียนรู้ฟังก์ชันเป้าหมายแบบตรงไปตรงมา (Direct mapping) มาเป็นการเรียนรู้ส่วนต่างของฟังก์ชันเป้าหมาย (Residual function) แทน เพื่อลดความซับซ้อนของกระบวนการเรียนรู้ เนื่องจากฟังก์ชันที่โครงข่ายต้องเรียนรู้มีความง่ายและตรงไปตรงมามากกว่าการเรียนรู้ฟังก์ชันเป้าหมายในภาพรวมทั้งหมด

การทำงานภายในโครงข่าย ResNet50 บนแนวคิด Residual learning ถูกนำมาใช้ผ่านโครงสร้างที่เรียกว่า Residual Block ซึ่งเป็นส่วนประกอบหลักของโมเดล ในแต่ละ Residual block ข้อมูลอินพุตจะถูกส่งผ่านเลเยอร์ย่อยต่าง ๆ เช่น การทำคอนโวลูชัน การปรับแต่งด้วย Batch normalization และการใช้ฟังก์ชันกระตุ้น ReLU เพื่อสร้างฟังก์ชันที่ต้องเรียนรู้ หลังจากนั้นฟังก์ชันที่ได้จะถูกบวกกับข้อมูลอินพุตเดิมผ่านกระบวนการ Skip connection ซึ่งช่วยให้ข้อมูลสามารถไหลผ่านไปยังเลเยอร์ถัดไปได้ โดยวิธีการนี้ไม่เพียงแต่ช่วยลดการสูญเสียข้อมูลระหว่างการส่งผ่านเลเยอร์เท่านั้น แต่ยังช่วยป้องกันปัญหาค่า Gradient ที่ค่อย ๆ หายไปในโครงข่ายที่มีความลึกมาก ซึ่งอาจทำให้โมเดลไม่สามารถเรียนรู้ได้อย่างมีประสิทธิภาพ โดยโครงสร้างของสถาปัตยกรรม ResNet50 ถูกออกแบบมาอย่างเป็นระบบ ประกอบด้วย 5 ส่วน ดังภาพที่ 2



ภาพที่ 2 โครงสร้างของสถาปัตยกรรม ResNet50

จากภาพที่ 2 โครงสร้างของสถาปัตยกรรม ResNet50 ซึ่งแต่ละส่วนมีบทบาทสำคัญในการสกัดคุณลักษณะจากข้อมูลภาพ เพื่อใช้ในการวิเคราะห์และจำแนกประเภทภาพ มีรายละเอียดดังนี้

ส่วนที่ 1 การประมวลผลเบื้องต้น (Initial processing) โดยในช่วงแรกข้อมูลภาพจะถูกประมวลผลผ่านขั้นตอนที่สำคัญ ได้แก่ Zero padding เพื่อขยายขนาดของภาพให้เหมาะสมสำหรับการคำนวณในเลเยอร์ถัดไป ตามด้วยเลเยอร์ Convolution ซึ่งใช้ตัวกรอง (Filters) จำนวน 64 ตัวที่มีขนาด 7×7 เพื่อสกัดคุณลักษณะพื้นฐาน เช่น ขอบเขต (Edges) และพื้นผิว (Textures) ของวัตถุ หลังจากนั้นจะใช้ Batch normalization เพื่อปรับค่าข้อมูลให้เหมาะสมสำหรับการเรียนรู้ และใช้ ReLU activation เพื่อเพิ่มความไม่เชิงเส้น (Non-linearity) ของโมเดล ต่อด้วยการใช้พูลลิงแบบมากที่สุด (Max pooling) เพื่อย่อขนาดเชิงพื้นที่ของข้อมูลภาพ โดยยังคงคุณลักษณะที่สำคัญไว้สำหรับกระบวนการสกัดคุณลักษณะในระดับที่สูงขึ้นในช่วงถัดไป

ส่วนที่ 2 การสกัดคุณลักษณะเบื้องต้น (Feature extraction on early layers) โดยใช้โครงสร้าง Residual blocks จำนวน 3 บล็อก แต่ละบล็อกประกอบด้วยเลเยอร์ Convolution ขนาด 3×3 และ Skip connection เพื่อเป็นกลไกการเชื่อมต่อแบบข้ามเลเยอร์ที่สามารถช่วยลดปัญหา Gradient vanishing และเพิ่มประสิทธิภาพการเรียนรู้คุณลักษณะที่มีความลึกมากขึ้น เช่น รูปร่าง (Shapes) และสี (Color) ของวัตถุ

ส่วนที่ 3 การสกัดคุณลักษณะระดับกลาง (Feature extraction on mid-level layers) ที่ประกอบด้วย Residual blocks จำนวน 4 บล็อก ซึ่งมีการเพิ่มจำนวนตัวกรองในเลเยอร์ Convolution เพื่อเพิ่มความสามารถในการจับข้อมูลที่ซับซ้อนขึ้น คุณลักษณะที่ได้ในช่วงนี้เริ่มรวมข้อมูลเชิงโครงสร้างและเชิงบริบท (Contextual information) เช่น ลวดลาย (Patterns) และคุณลักษณะเชิงพื้นที่ (Spatial feature) ซึ่งมีความสำคัญต่อการวิเคราะห์และประมวลผลโครงสร้างภาพในระดับที่สูงขึ้น

ส่วนที่ 4 การสกัดคุณลักษณะเชิงลึก (Feature extraction on deep layers) จาก Residual blocks จำนวน 6 บล็อก เพื่อเรียนรู้คุณลักษณะในระดับที่ลึกและเป็นนามธรรมมากขึ้น เช่น การแยกรูปแบบของวัตถุเฉพาะเจาะจง และบริบทเชิงพื้นที่ของวัตถุ โดยคุณลักษณะในช่วงนี้สามารถบ่งบอกถึงความแตกต่าง

ระหว่างประเภทของวัตถุในรูปภาพได้อย่างชัดเจนจึงเหมาะสำหรับการนำไปใช้ในงานวิเคราะห์ภาพที่ต้องการความแม่นยำสูง เช่น การค้นหารูปภาพ และการจำแนกรูปภาพที่ซับซ้อน

ส่วนที่ 5 การสกัดคุณลักษณะขั้นสุดท้าย (Final feature extraction) ด้วย Residual blocks จำนวน 3 บล็อก โดยฟีเจอร์ที่ได้จากเลเยอร์สุดท้ายจะถูกย่อขนาดเชิงพื้นที่ด้วย Global average pooling ซึ่งช่วยลดความซับซ้อนของข้อมูล แต่ยังคงรักษาค่าที่สำคัญของคุณลักษณะของแต่ละตัวกรองไว้ จากนั้นข้อมูลจะถูกแปลงให้อยู่ในรูปเวกเตอร์ผ่าน Flatten layer และส่งผ่านไปยัง Dense layer สำหรับการจำแนกภาพในชุดข้อมูล ImageNet ที่มีจำนวนคลาสเท่ากัน

กระบวนการสกัดคุณลักษณะโดยนำเข้าสู่ข้อมูล (Input data) รูปภาพที่ผ่านกระบวนการเตรียมชุดข้อมูลเข้าสู่ตัวสกัดคุณลักษณะด้วยการปรับแต่ง (Fine-tuning) โมเดล ResNet50 ให้กลายเป็นตัวสกัดคุณลักษณะที่มุ่งเน้นการดึง Feature map จากภาพ โดยปรับแต่งโครงสร้างโมเดล ในส่วนที่ 5 จากการตัด Dense layer ซึ่งเป็นเลเยอร์สุดท้ายที่ทำหน้าที่จำแนกรูปภาพออกจากโมเดล แต่ยังคงรักษา Global average pooling layer ไว้ เนื่องจากมีบทบาทสำคัญในการลดมิติของ Feature map แปลงเป็นเวกเตอร์คุณลักษณะ (Feature vector) ซึ่งเป็นตัวแทนของข้อมูลที่สำคัญจากภาพที่สามารถนำไปประยุกต์ใช้ในการค้นหารูปภาพโดยไม่ต้องพึ่งพาผลลัพธ์จากการจำแนกประเภทภาพ อีกทั้งการตัด Dense layer ออก ยังช่วยเพิ่มความยืดหยุ่นให้กับโมเดลในการนำเวกเตอร์คุณลักษณะไปใช้วิเคราะห์เชิงลึกจากลักษณะเฉพาะของภาพ เช่น การค้นหารูปภาพที่มีลักษณะคล้ายคลึงกัน หรือการวิเคราะห์ลักษณะเชิงพื้นที่ของภาพในแง่มุมต่าง ๆ

3) การลดมิติข้อมูลด้วยเทคนิคการวิเคราะห์องค์ประกอบหลัก (Dimensionality reduction with PCA) เนื่องจากข้อมูลที่มีจำนวนมิติสูง (High-dimensional data) ซึ่งมักจะเกิดขึ้นจากการสกัดคุณลักษณะซึ่งเวกเตอร์ที่มีจำนวนมิติสูงนั้นอาจทำให้การประมวลผลข้อมูลมีความซับซ้อนและใช้ทรัพยากรมากเกินไป อีกทั้งยังอาจส่งผลกระทบต่อประสิทธิภาพของตัวแบบ (Model performance) (Keogh & Mueen 2017) ดังนั้นการลดมิติข้อมูลจึงเป็นขั้นตอนที่สำคัญในการปรับปรุงประสิทธิภาพของกระบวนการวิเคราะห์ข้อมูล งานวิจัยนี้ใช้การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) (Jolliffe, 2002) ซึ่งเป็นหนึ่งในเทคนิคที่ได้รับความนิยมมาใช้ในการลดมิติข้อมูลจากการแปลงข้อมูลจากมิติที่สูงให้เหลือต่ำลง โดยไม่สูญเสียข้อมูลที่สำคัญมากนัก เทคนิคนี้มีพื้นฐานจากแนวคิดในการค้นหาแกนมิติใหม่ที่เรียกว่าองค์ประกอบหลัก (Principal components) ซึ่งเป็นแกนที่สามารถอธิบายความแปรปรวนของข้อมูลได้มากที่สุด โดยการหาค่าความแปรปรวนสูงสุดของข้อมูลในแต่ละมิตินั้นการวิเคราะห์องค์ประกอบหลักจะสร้างการแปลงเชิงเส้นที่ช่วยลดจำนวนมิติของข้อมูลลงในลำดับที่สำคัญที่สุด

กระบวนการวิเคราะห์องค์ประกอบหลัก จะเริ่มต้นด้วยการสร้างอ็อบเจกต์องค์ประกอบหลัก ซึ่งเป็นเครื่องมือที่ใช้ในการคำนวณและระบุองค์ประกอบหลักของข้อมูลในชุดข้อมูลที่มีมิติสูง โดยการวิเคราะห์

องค์ประกอบหลักจะทำการหาค่าความแปรปรวนสูงสุดในข้อมูล และกำหนดแกนมิติใหม่ที่สุดคล้อยกับความแปรปรวนเหล่านั้น หลังจากนั้นข้อบกพร่องขององค์ประกอบหลักจะถูกปรับแต่งชุดข้อมูลที่ได้จากการสกัดคุณลักษณะ เพื่อหาค่าขององค์ประกอบหลักที่สามารถอธิบายความแปรปรวนของข้อมูลได้มากที่สุด จึงไม่เพียงแต่จะช่วยลดมิติของข้อมูลให้เหลือเพียงมิติที่สำคัญที่สุด แต่ยังช่วยในการปรับปรุงประสิทธิภาพของการประมวลผลข้อมูลในขั้นตอนถัดไป เช่น การสร้างตัวแบบหรือการวิเคราะห์ข้อมูลเพิ่มเติม เนื่องจากการลดมิติข้อมูลจะช่วยลดความซับซ้อนของข้อมูล และลดปัญหาที่อาจเกิดจากการที่จำนวนมิติสูงเกินไป เช่น ปัญหาความแปรปรวนของข้อมูล (Variance) ที่ไม่สามารถอธิบายได้ทั้งหมด

งานวิจัยนี้กำหนดจำนวนองค์ประกอบหลักเท่ากับ 300 เพื่อควบคุมความสมดุลระหว่างการลดมิติข้อมูลและการรักษาความแปรปรวนที่สำคัญในชุดข้อมูลที่ใช้ในการวิเคราะห์หรือฝึกฝนโมเดล (Model training) ดังนั้นการเลือกจำนวนองค์ประกอบหลักที่เหมาะสมเป็นกระบวนการที่ต้องพิจารณาจากหลายปัจจัย เช่น จำนวนมิติที่ต้องการลดลง และสัดส่วนของความแปรปรวนที่ต้องการคงไว้ในข้อมูลที่ถูกแปลงแล้ว เพื่อตอบสนองความต้องการในการคงข้อมูลสำคัญไว้และเพิ่มประสิทธิภาพในการฝึกตัวแบบต่อไป

4) การสร้างชุดข้อมูลแทนความหมายของภาพ (Create the image representations dataset) ถือเป็นกระบวนการที่สำคัญในการจัดการและเก็บรักษาข้อมูลที่ได้จากการประมวลผลภาพ ซึ่งเป็นการแปลงข้อมูลจากรูปภาพให้กลายเป็นข้อมูลเชิงตัวเลขที่สามารถนำมาประมวลผลและใช้งานต่อไปได้อย่างมีประสิทธิภาพ ประกอบด้วยข้อมูลสำคัญที่ได้จากขั้นตอนต่าง ๆ ในการประมวลผลภาพ เช่น เวกเตอร์คุณลักษณะที่ลดมิติแล้ว (Dimensionality-reduced feature vectors) ซึ่งได้จากกระบวนการลดมิติข้อมูลด้วยการวิเคราะห์องค์ประกอบหลัก เพื่อคัดเลือกและเก็บข้อมูลที่สำคัญที่สุดจากชุดข้อมูลที่มีมิติสูง รวมทั้งข้อบกพร่องขององค์ประกอบหลักที่ใช้ในการลดมิติข้อมูล ซึ่งเป็นส่วนสำคัญในการแปลงข้อมูลภาพให้เป็นเวกเตอร์คุณลักษณะที่เหมาะสมสำหรับการนำไปใช้งานต่อไป นอกจากนี้ยังมีการบันทึกเส้นทางจัดเก็บของรูปภาพแต่ละไฟล์ ซึ่งทำให้สามารถติดตามและเข้าถึงภาพต้นฉบับได้ทันที

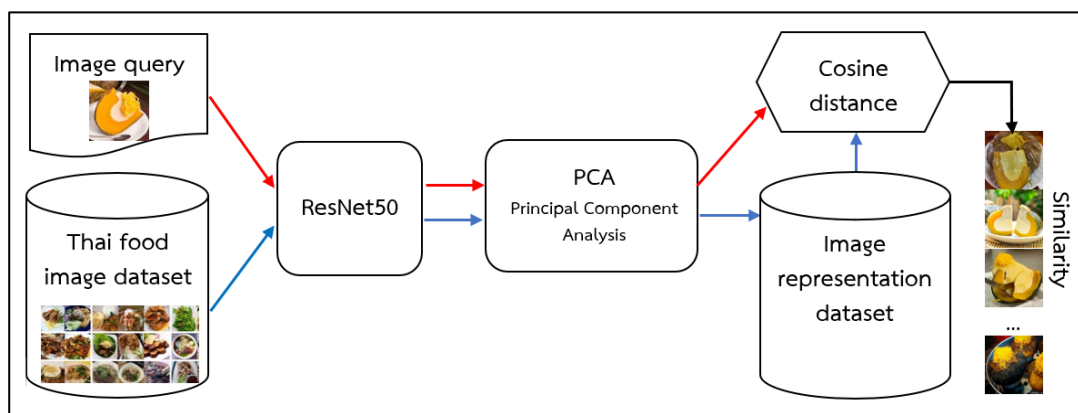
การจัดเก็บข้อมูลในลักษณะนี้จะช่วยให้ชุดข้อมูลแทนความหมายของภาพทำหน้าที่เป็นคลังข้อมูลที่สามารถนำไปใช้ในการค้นหารูปภาพ เนื่องจากข้อมูลที่เก็บรวบรวมในเชิงตัวเลขนั้นสามารถเปรียบเทียบความคล้ายคลึง (Similarity) ของภาพต่าง ๆ ได้อย่างรวดเร็วและไม่จำเป็นต้องทำการประมวลผลภาพต้นฉบับซ้ำในแต่ละครั้ง ตลอดจนช่วยให้การบริหารจัดการข้อมูลเป็นไปได้อย่างมีประสิทธิภาพ อันเป็นส่วนสำคัญในการพัฒนาแอปพลิเคชันต่อไปในอนาคต

ผลการวิจัยและอภิปรายผล

การสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึกสำหรับการค้นหารูปภาพย้อนกลับ แบ่งผลการวิจัยและอภิปรายผลตามวัตถุประสงค์ ได้แก่ 1) เพื่อพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก และ 2) เพื่อประเมินความแม่นยำของผลการค้นหารูปภาพย้อนกลับ มีรายละเอียดดังนี้

1) ผลการพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดลการเรียนรู้เชิงลึก เพื่อสกัดคุณลักษณะรูปภาพบนชุดข้อมูลรูปภาพอาหารไทยที่ได้รับความนิยมจากนานาชาติจำนวน 20 รายการด้วยโมเดล ResNet50 ที่ผ่านการถ่ายโอนความรู้จากการฝึกฝนบนชุดข้อมูลขนาดใหญ่ ร่วมกับการปรับแต่งโมเดลเพิ่มเติมให้เหมาะสมกับชุดข้อมูลด้วยการปรับค่าน้ำหนักบางส่วน of โมเดลร่วมกับฝึกฝนเฉพาะบางชั้นของโครงข่าย ก่อนจะแปลงเป็น Feature map ซึ่งประกอบด้วยเวกเตอร์เชิงตัวเลข ขนาด 2,048 องค์ประกอบที่เป็นผลลัพธ์จากกระบวนการสกัดคุณลักษณะ โดยเวกเตอร์นี้ได้จากเลเยอร์ Global average pooling ซึ่งทำหน้าที่สรุปข้อมูลจากพีเจอร์เมทริกซ์ขนาดใหญ่ที่ได้จากเลเยอร์ Convolutional ก่อนหน้านั้นให้อยู่ในรูปแบบของเวกเตอร์ขนาด 1 มิติที่แสดงลักษณะสำคัญของภาพ เช่น รูปร่าง รูปทรง สี หรือพื้นผิวที่โมเดลสามารถเรียนรู้และตรวจจับได้ในการฝึกฝน โดยคุณลักษณะเหล่านี้เป็นข้อมูลที่สำคัญในเชิงบริบทและโครงสร้างของภาพต้นฉบับ

อย่างไรก็ดี ความซับซ้อนของมิติข้อมูลดังกล่าวอาจส่งผลกระทบต่อประสิทธิภาพในการประมวลผล ซึ่งอาจนำไปสู่ปัญหาคำสาปแห่งมิติ (Curse of dimensionality) ดังนั้นการวิเคราะห์องค์ประกอบหลักซึ่งช่วยแปลงข้อมูลจากมิติที่สูงให้เหลือมิติที่ต่ำลง โดยยังคงรักษาความแปรปรวนที่สำคัญในข้อมูลไว้ วิธีการนี้ไม่เพียงช่วยลดความซับซ้อนของข้อมูล แต่ยังช่วยเพิ่มประสิทธิภาพในการประมวลผล งานวิจัยนี้มุ่งเน้นไปที่องค์ประกอบที่สำคัญที่สุดจำนวน 300 องค์ประกอบ โดยตัดข้อมูลที่มีผลกระทบน้อยหรือที่เป็นสัญญาณรบกวนออกไปก่อน จะนำไปสร้างเป็นชุดข้อมูลแทนความหมายของภาพ เพื่อเป็นตัวแบบการค้นหารูปภาพย้อนกลับ ดังภาพที่ 3



ภาพที่ 3 กระบวนการค้นหารูปภาพย้อนกลับ

จากภาพที่ 3 กระบวนการค้นหาภาพย้อนกลับเริ่มต้นด้วยการนำภาพค้นหา (Query image) เข้าสู่ตัวแบบ โดยผ่านขั้นตอนการประมวลผลล่วงหน้า (Pre-processing) เพื่อปรับขนาดภาพเป็น 224*224 พิกเซลที่เหมาะสมกับการสกัดคุณลักษณะรูปภาพในโมเดล ResNet50 จากนั้นภาพจะถูกลดมิติข้อมูลด้วยการวิเคราะห์องค์ประกอบหลักแล้วจะนำ Feature map ของภาพค้นหาไปเปรียบเทียบกับ Feature map ที่ถูกจัดเก็บไว้ในชุดข้อมูลแทนความหมายของภาพ ซึ่งเป็นคลังข้อมูลของภาพที่ผ่านการประมวลผลล่วงหน้า และสกัดคุณลักษณะมาแล้ว การเปรียบเทียบนี้ใช้การคำนวณค่าความคล้ายคลึงโคไซน์ (Cosine similarity) (Manning et al., 2008) เป็นมาตรวัดทางคณิตศาสตร์ที่ใช้เพื่อวัดความแตกต่างหรือความคล้ายคลึงระหว่างเวกเตอร์สองตัวในพื้นที่เวกเตอร์เชิงมิติ ดังสมการที่ 1 ตัวอย่างการคำนวณระยะทางโคไซน์จากระยะห่างระหว่างเวกเตอร์ A และเวกเตอร์ B

$$\text{Cosine similarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (1)$$

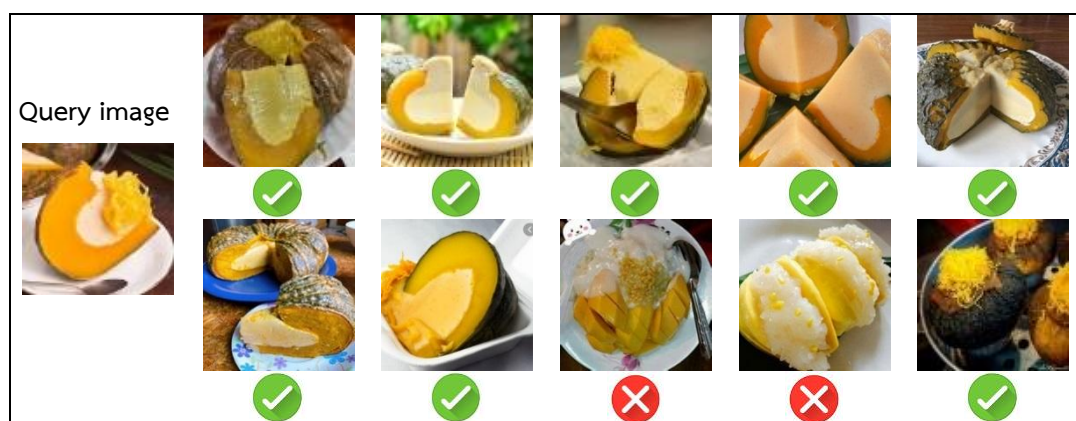
กำหนดให้ $A \cdot B$ คือ ผลคูณเชิงจุด (Dot product) ของเวกเตอร์
 $\|A\|$ และ $\|B\|$ คือ ขนาดของเวกเตอร์ A และเวกเตอร์ B ซึ่งคำนวณจากผลรวมของค่ากำลังสองของแต่ละองค์ประกอบในเวกเตอร์
 Cosine similarity คือ ค่าความคล้ายคลึงโคไซน์ ซึ่งมีค่าอยู่ในช่วง $[-1, 1]$

จากสมการที่ 1 การแปลผลค่าความคล้ายคลึงเชิงโคไซน์ หากมีค่าใกล้ 1 หมายถึงเวกเตอร์ A และเวกเตอร์ B มีทิศทางที่คล้ายคลึงกันมาก ในขณะที่ค่าที่เข้าใกล้ 0 หมายถึงเวกเตอร์ไม่มีความสัมพันธ์เชิงทิศทาง และค่าที่ใกล้ -1 หมายถึงเวกเตอร์มีทิศทางตรงข้ามกัน ดังนั้นหากใช้ในการเปรียบเทียบความแตกต่างระหว่างเวกเตอร์ จะคำนวณได้ด้วยระยะทางโคไซน์ (Cosine distance) ดังสมการที่ 2

$$\text{Cosine distance} = 1 - \text{cosine_similarity} \quad (2)$$

จากสมการที่ 2 จะเห็นได้ว่าระยะทางโคไซน์มีค่าอยู่ในช่วง $[0, 2]$ ดังนั้นเมื่อค่าความคล้ายคลึงโคไซน์มีค่าใกล้ 1 จะส่งผลให้ระยะทางโคไซน์มีค่าใกล้ 0 ในทางกลับกัน เมื่อค่าความคล้ายคลึงโคไซน์มีค่าใกล้ 0 จะส่งผลให้ระยะทางโคไซน์มีค่าใกล้ 1 และเมื่อค่าความคล้ายคลึงโคไซน์มีค่าใกล้ -1 จะส่งผลให้ระยะทางโคไซน์มีค่าใกล้ 2 จะเห็นได้ว่าระยะทางโคไซน์ไม่ได้สนใจค่าขนาด (Magnitude) ของเวกเตอร์ แต่จะเน้นวัดมุมระหว่างเวกเตอร์ทั้งสอง เพื่อพิจารณาทิศทางเชิงสัมพัทธ์ (Relative orientation) ของเวกเตอร์ในพื้นที่นั้นเพื่อวัดค่าความ

แตกต่างเชิงเวกเตอร์ระหว่างภาพค้นหาและภาพในฐานข้อมูล ยิ่งค่าระยะทางน้อยเท่าใด แสดงว่าภาพในฐานข้อมูลมีความคล้ายคลึงกับภาพค้นหามากเท่านั้น ก่อนจะจัดอันดับภาพในฐานข้อมูล (Image ranking) ตามค่าความคล้ายคลึงเรียงลำดับจากมากไปหาน้อยสำหรับการนำเสนอเป็นผลการค้นหารูปภาพแก่ผู้ใช้ในรูปแบบแผงภาพ (Image grid) ร่วมกับค่าความคล้ายคลึงหรือรายละเอียดของภาพนั้น ๆ ดังภาพที่ 4 ตัวอย่างรูปภาพค้นหาและผลการค้นหารูปภาพย้อนกลับจำนวน 10 รูปภาพที่ผ่านการประเมินความถูกต้องจากผู้เชี่ยวชาญ ดังนั้นจึงช่วยให้สามารถค้นหาภาพที่ต้องการได้อย่างรวดเร็วและแม่นยำ โดยไม่จำเป็นต้องมีข้อมูลป้ายกำกับในภาพค้นหา



ภาพที่ 4 ตัวอย่างรูปภาพค้นหา และผลการค้นหารูปภาพย้อนกลับจำนวน 10 รูปภาพที่ผ่านการประเมินความถูกต้องจากผู้เชี่ยวชาญ

2) ผลประเมินความแม่นยำของผลการค้นหาภาพย้อนกลับ โดยนำโมเดลต้นแบบทดสอบโดยใช้รูปภาพค้นหาในโดเมนอาหารไทย 20 รายการจากผู้ใช้ด้วยรูปภาพค้นหาที่ตัวแบบไม่เคยเห็นมาก่อน (Unseen dataset) รายการอาหารละ 1 รูปภาพ รวมทั้งสิ้น 20 รูปภาพ เนื่องจากเป็นตัวชี้วัดความสามารถของโมเดลในการทำงานกับข้อมูลใหม่ที่มีความหลากหลาย และสะท้อนถึงประสิทธิภาพของโมเดลในสถานการณ์จริงที่มีผู้ใช้งานด้วยรูปภาพในบริบทที่แตกต่างกัน เช่น รูปภาพที่มีความแตกต่างด้านระยะถ่าย มุมมอง แสงสว่าง ตลอดจนพื้นหลัง จะช่วยทดสอบความยืดหยุ่นของโมเดลในการจัดการกับความแปรปรวนของข้อมูล เช่น ภาพถ่ายในระยะที่ต่างกัน อาจทำให้รายละเอียดของรูปภาพเปลี่ยนไป หรือมุมมองจากด้านบนหรือด้านข้างอาจส่งผลต่อลักษณะเด่นของอาหารแต่ละรายการ นอกจากนี้สีภาพแสงที่ต่างกัน เช่น แสงธรรมชาติ แสงจากหลอดไฟ หรือการมีเงาในภาพ ยังเป็นปัจจัยที่เพิ่มความซับซ้อนในการสกัดคุณลักษณะ

ดังนั้นเพื่อเป็นการสะท้อนถึงความพร้อมของโมเดลต่อการใช้งานในวงกว้าง โดยเฉพาะในบริบทที่ผู้ใช้งานแต่ละคนมีแหล่งข้อมูลและเงื่อนไขที่แตกต่างกัน เนื่องจากหากโมเดลสามารถรักษาความแม่นยำในระดับสูงได้กับข้อมูลที่ไม่เคยเห็นมาก่อน จะช่วยเพิ่มศักยภาพในการนำไปใช้งานจริง การทดสอบนี้จึงถือเป็นขั้นตอนสำคัญที่ช่วยให้มั่นใจได้ว่าโมเดลสามารถตอบโจทย์ในสถานการณ์จริงได้อย่างมีประสิทธิภาพ และสอดคล้องกับความต้องการของผู้ใช้งาน จึงเป็นที่มาของการให้ผู้เชี่ยวชาญเข้ามามีบทบาทในการประเมินความแม่นยำ เพื่อเพิ่มความน่าเชื่อถือและลดข้อผิดพลาดที่อาจเกิดขึ้นจากกระบวนการอัตโนมัติหรือข้อจำกัดของตัวแบบในการสกัดคุณลักษณะเฉพาะทาง

การประเมินจากผู้เชี่ยวชาญหลายคนยังช่วยลดความคลาดเคลื่อนของการประเมินที่อาจเกิดขึ้นจากการพิจารณาโดยผู้เชี่ยวชาญเพียงท่านเดียว ทำให้ผลการประเมินมีความน่าเชื่อถือมากขึ้น งานวิจัยนี้กำหนดจำนวนผู้เชี่ยวชาญ 4 ท่านทำการประเมินผลด้วยรูปภาพค้นหา 1 รูปภาพต่อ 1 รายการอาหาร ท่านละ 5 รายการ หากผลการค้นหารูปภาพใดถูกต้องเท่ากับ 1 คะแนน ในทางตรงกันข้าม หากไม่ถูกต้องเท่ากับ 0 คะแนน เมื่อทำการประเมินแล้วจะนำคะแนนทั้งหมดมาคำนวณเป็นค่าความแม่นยำ (Manning et al., 2008) เพื่อประเมินประสิทธิภาพของตัวแบบดังสมการที่ 3 จำนวนผลลัพธ์ที่ถูกต้องหารด้วยจำนวนผลลัพธ์ที่ค้นหาทั้งหมดที่จำนวน k อันดับตามการแสดงผลการค้นหาเป็น 3, 5, 10, 20, 50 และ 100 รูปภาพต่อรายการ

$$\text{Precision@}k = \frac{\text{จำนวนผลลัพธ์ที่ถูกต้องใน } k \text{ อันดับ}}{k} \quad (3)$$

งานวิจัยนี้ให้ความสำคัญค่าความคลาดเคลื่อน (error) ระหว่างค่าที่วัดได้กับค่าที่แท้จริง ถ้าค่าที่วัดได้ใกล้เคียงกับค่าจริงมากแสดงว่าการวัดนั้นมีความถูกต้องสูง (เรวัต แสงสุริยงค์, 2565) โดยการวัดทุกครั้งมักมีค่าความคลาดเคลื่อนเชิงสถิติ (Statistical error) เกิดขึ้นเสมอ ดังนั้นเมื่อครบรอบจะทำการประเมินอีกครั้งจนครบ 3 รอบด้วยรูปภาพค้นหาเดิมแบบผลลัพธ์เป็นอิสระต่อกัน แล้วนำผลการประเมินมาหาค่าเฉลี่ยเพื่อให้ความคลาดเคลื่อนลดน้อยลงจนถึงในระดับที่ยอมรับได้

ผลการประเมินจากผู้เชี่ยวชาญ พบว่า ค่าความแม่นยำจากการค้นหารูปภาพที่ k อันดับตามจำนวนการแสดงผลรูปภาพอาหารจากการค้นหาเป็น 3, 5, 10, 20, 50 และ 100 รูปภาพต่อรายการ คิดเป็นค่าความแม่นยำเฉลี่ยเท่ากับ 80%, 76%, 75%, 69%, 62% และ 45% ตามลำดับ ดังตารางที่ 2 โดยค่าความแม่นยำจะลดลงเมื่อจำนวนผลลัพธ์ที่เพิ่มขึ้น เมื่อแปลความหมายค่าความแม่นยำ พบว่า ผลลัพธ์จากการค้นหารูปภาพจำนวน 3 ลำดับอยู่ในระดับดีมาก ตามด้วยผลลัพธ์จากการค้นหารูปภาพจำนวน 5 และ 10 ลำดับอยู่ในระดับดี แต่เมื่อผลลัพธ์จากการค้นหามีจำนวนมากขึ้นเป็น 20 และ 50 รูปภาพ ค่าความแม่นยำจะลดลงอยู่ในระดับพอใช้ และเมื่อจำนวนผลลัพธ์เท่ากับ 100 รูปภavnั้นตัวแบบจำเป็นต้องได้รับการปรับปรุงต่อไปในอนาคต

ตารางที่ 2 ค่าความแม่นยำเฉลี่ยจากการค้นหารูปภาพที่ k อันดับตามจำนวนการแสดงผลรูปภาพจากการค้นหา เป็น 3, 5, 10, 20, 50 และ 100 รูปภาพต่อรายการ

ลำดับรายการอาหาร	Precision					
	@3	@5	@10	@20	@50	@100
1) ต้มยำ	1.00	1.00	0.90	0.85	0.70	0.50
2) แกงส้ม	0.67	0.80	0.80	0.75	0.72	0.55
3) แกงเขียวหวาน	1.00	0.80	0.70	0.65	0.60	0.40
4) พะแนง	0.33	0.60	0.70	0.60	0.55	0.35
5) แกงมัสมั่น	0.67	0.60	0.60	0.55	0.50	0.30
6) แกงจืด	1.00	0.80	0.80	0.70	0.60	0.45
7) ไช้เจียว	0.67	0.60	0.70	0.65	0.55	0.40
8) หมูย่าง	1.00	1.00	0.90	0.80	0.70	0.55
9) ข้าวขาหมู	1.00	0.80	0.80	0.75	0.72	0.60
10) ข้าวหมกไก่	0.67	0.60	0.70	0.70	0.65	0.50
11) ข้าวมันไก่	0.67	0.60	0.60	0.55	0.50	0.35
12) ลาบ	1.00	0.80	0.70	0.65	0.60	0.40
13) ผัดกะเพรา	1.00	1.00	0.90	0.85	0.70	0.50
14) ผัดเผ็ดมะม่วง	0.67	0.80	0.80	0.75	0.72	0.55
15) ส้มตำ	1.00	0.80	0.70	0.65	0.60	0.45
16) ผัดผักบุ้ง	0.33	0.60	0.70	0.60	0.55	0.35
17) ผัดซีอิ้ว	0.67	0.60	0.60	0.55	0.50	0.30
18) ผัดไทย	1.00	0.80	0.80	0.70	0.60	0.45
19) ข้าวเหนียวมะม่วง	0.67	0.60	0.70	0.65	0.55	0.40
20) ส้มขี้ฟ้าทอด	1.00	1.00	0.90	0.80	0.70	0.55
ค่าเฉลี่ย	0.80	0.76	0.75	0.69	0.62	0.45

จากตารางที่ 2 เมื่อพิจารณาในรายละเอียด พบว่า ค่าความแม่นยำของการค้นหารูปภาพอาหารไทย นั้นขึ้นอยู่กับปัจจัยหลายประการที่เกี่ยวข้องกับคุณลักษณะของภาพ และประสิทธิภาพของโมเดลที่ใช้ ในการประมวลผลภายใต้สถานการณ์จริงที่มีผู้ใช้งานด้วยรูปภาพในบริบทที่แตกต่างกันมีรายละเอียดดังนี้

1) ความชัดเจนของคุณลักษณะเฉพาะในภาพอาหาร จะเห็นได้จากรายการอาหารที่มีค่าความแม่นยำ สูงอย่างต้มยำ และส้มขี้ฟ้าทอด แสดงถึงความสามารถของตัวแบบในการสกัดคุณลักษณะเด่นได้อย่างแม่นยำ คุณลักษณะเหล่านี้อาจรวมถึงสี สัน รูปทรงของส่วนประกอบในจาน เช่น ต้มยำมักมีสีแดงจากน้ำแกงที่ชัดเจน

ในขณะที่สังขยาฟักทองนอกจากสีเหลืองที่ชัดเจนแล้วยังมีรูปทรงเด่นชัดระหว่างชั้นฟักทองและเนื้อสังขยาในทางกลับกัน รายการอาหารที่มีความคล้ายคลึงกันมากอย่างพะแนง และแกงมัสมั่นนั้นอาจมีน้ำแกงที่มีสีคล้ายคลึงกัน และอยู่ภาชนะรูปทรงที่ใกล้เคียงกัน จึงอาจส่งผลต่อประสิทธิภาพในการวิเคราะห์ความคล้ายคลึง

2) บริบทของภาพที่หลากหลาย (Context of images) เช่น อาหารที่มักจัดอยู่ในภาชนะลักษณะที่ใกล้เคียงกัน รูปแบบการตกแต่งจาน หรือองค์ประกอบเสริมในภาพ อาจทำให้การสกัดคุณลักษณะสำคัญในภาพมีความคลาดเคลื่อน เช่น รายการอาหารประเภทข้าวที่ในภาพมักมีช้อนส้อมเป็นองค์ประกอบ เช่นเดียวกับรายการอาหารประเภทเส้นที่มักมีตะเกียบปรากฏในภาพเป็นองค์ประกอบ

3) ความแปรปรวนของรูปภาพ (Variance of image) เนื่องจากรูปภาพอาหารนั้นมีบริบทแตกต่างจากรูปภาพในโดเมนอื่น ดังนั้นอาจประสบกับปัญหาความแปรปรวนของรูปภาพ (Gonzalez & Woods, 2008) ในหลายกรณี ประกอบด้วย

3.1) ความแปรปรวนของมุมมอง (View-point variance) เป็นผลจากการถ่ายภาพอาหารในมุมมองต่าง ๆ เช่น การหมุนหรือเปลี่ยนทิศทางของภาชนะในหลายมิติ อาจทำให้การสกัดคุณลักษณะไม่สามารถระบุวัตถุได้อย่างแม่นยำในทุกมุมมอง เนื่องจากตำแหน่งหรือทิศทางภายในภาพจะถูกลดความสำคัญทันที

3.2) ความแปรปรวนของขนาด (Scale variance) ซึ่งเกิดจากปัจจัยต่าง ๆ เช่น ระยะทางระหว่างวัตถุและกล้องถ่ายภาพ หรือการขยายเข้าขยายออกของเลนส์ ส่งผลให้ขนาดของวัตถุในภาพแตกต่างกัน

3.3) สภาพแสง (Illumination conditions) โดยรูปภาพที่ถ่ายในสภาพแสงต่างกัน เช่น แสงจ้า แสงน้อย หรือแสงจากมุมที่ไม่เหมาะสม อาจทำให้การสกัดลักษณะเด่นในภาพมีความคลาดเคลื่อน

3.4) พื้นหลังที่ซับซ้อน (Cluttered background) จากการจัดภาชนะ และการตกแต่งภาพอาหารด้วยองค์ประกอบอื่นที่ไม่เกี่ยวข้องโดยใช้สีหรือลวดลายที่คล้ายกับอาหาร หรือมีวัตถุอื่นที่ไม่เกี่ยวข้องจำนวนมากในภาพอาจรบกวนการจดจำลักษณะเฉพาะของวัตถุเป้าหมาย

3.5) ความแปรปรวนภายในคลาส (Intra-class variation) เกิดจากความหลากหลายที่ถูกจัดอยู่ในคลาสเดียวกัน เช่น รายการอาหารเดียวกันแต่ถูกปรุงขึ้นจากวัตถุดิบหลักที่แตกต่างกัน ซึ่งส่งผลต่อคุณลักษณะสำคัญของอาหารแต่ละรายการ

ตารางที่ 3 ตัวอย่างรูปภาพที่มีปัญหาความแปรปรวนของรูปภาพ

Context Of images	View-point variance	Scale variance	Illumination conditions	Cluttered background	Intra-class variation
					

จากตารางที่ 3 ปัญหาความแปรปรวนของรูปภาพในการประมวลผล แบ่งออกได้เป็น 3 ประเด็น ได้แก่ 1) ความแปรปรวนที่รบกวนการตรวจจับคุณลักษณะสำคัญจากสถานะแสง และสิ่งรบกวนในฉาก ซึ่งอาจลดความเด่นชัดขององค์ประกอบที่สำคัญของอาหาร โดยบางภาพถูกถ่ายภายใต้แสงสลัวหรือเงามืดส่งผลให้สีของอาหารผิดเพี้ยนและพื้นผิวดูไม่ชัดเจน เช่น แกงที่มีน้ำข้นอาจถูกถ่ายในภาวะแสงน้อยอาจทำให้สีดูทึบเกินไปและพื้นผิวดูไม่ชัดเจน ตรงกันข้าม ภาพอาหารที่ถูกถ่ายในแสงที่จัดจ้าเกินไป ส่งผลให้สีของอาหารซีดจางหรือสว่างเกินจริง เช่น น้ำแกงที่ควรมีสีจัดจ้านกลับดูซีดจาง นอกจากนั้นภาพที่มีองค์ประกอบอื่นนอกเหนือจากอาหารบริเวณด้านหลังหรือรอบข้างถูกตกแต่งด้วยวัตถุอื่นหรือมีรายละเอียดแวดล้อมที่ไม่เกี่ยวข้องจำนวนมาก เช่น ผ้าเช็ดปาก ช้อนส้อม แก้วน้ำ หรือแม้แต่บุคคลในพื้นที่หลัง ตลอดจนฉากหลังที่มีลวดลายสะดุดตา ซึ่งเบี่ยงเบนความสนใจของตัวแบบไปจากอาหารซึ่งเป็นจุดหลักภายในภาพ และอาจทำให้เข้าใจผิดว่าคุณลักษณะแวดล้อมเป็นส่วนหนึ่งของอาหารนั้น 2) ความแปรปรวนที่ส่งผลให้ตัวแบบจดจำคุณลักษณะผิดพลาดจากการเปลี่ยนแปลงมุมมองและขนาดของวัตถุ โดยภาพที่ถ่ายจากมุมบน มุมข้าง หรือมุมเอียงอาจทำให้โครงสร้างและการจัดวางองค์ประกอบของอาหารเปลี่ยนแปลงไปตามภาชนะ เช่น อาหารประเภทแกงเมื่อถ่ายจากด้านบนของภาชนะอาจเห็นแค่พื้นผิวของน้ำแกง แต่เมื่อถ่ายจากด้านข้างจะเห็นความลึกและลักษณะภาชนะชัดเจนยิ่งขึ้น ตรงกันข้ามกับอาหารประเภทเส้นที่ภาพจากมุมบนจะเห็นรายละเอียดทั้งหมด แต่หากถ่ายจากด้านข้างจะมองเห็นเส้นซ้อนทับกันโดยเห็นเครื่องเคียงด้านบนไม่ชัดเจน นอกจากนั้นยังมีบางภาพที่มีการถ่ายจากระยะใกล้มากเกินไปจนเห็นเฉพาะบางส่วนของอาหาร ในขณะที่บางภาพถ่ายจากระยะไกลจนทำให้อาหารดูมีขนาดเล็กและขาดรายละเอียด ความแปรปรวนนี้ทำให้ตัวแบบอาจไม่สามารถเรียนรู้คุณลักษณะเชิงโครงสร้างที่แท้จริงของ

อาหารได้อย่างถูกต้อง 3) ความหลากหลายของรูปภาพระหว่างคลาส (Intra-class variation) และความคล้ายคลึงกันของรูปภาพภายในคลาสเดียวกัน (Inter-class similarity) ซึ่งส่งผลต่อความสามารถของตัวแบบในการจำแนกภาพที่มีลักษณะใกล้เคียงกัน เช่น อาหารประเภทผัด บางภาพมีสีเข้มเพราะใช้ซอสเข้มข้น ขณะที่บางภาพใช้ผักสดและมีสีเขียวสดใส ขณะที่แกงชนิดเดียวกันอาจมีน้ำแกงสีเข้มในบางภาพและสีอ่อนในอีกภาพ จากวัตถุดิบหรือกรรมวิธีปรุงที่แตกต่างกันตามบริบทของผู้ประกอบอาหาร ทำให้แม้อยู่ในคลาสเดียวกันก็มีลักษณะภาพที่แตกต่างกันมาก นอกจากนี้ การตกแต่งจานด้วยเครื่องเคียงหลากสี หรือมีข้าวสวยร่วมในจาน ซึ่งสร้างความแตกต่างจากภาพอื่นแม้จะเป็นอาหารรายการเดียวกัน ความแปรปรวนภายในคลาสเช่นนี้อาจทำให้ตัวแบบเกิดการจดจำคุณลักษณะร่วมที่ไม่สอดคล้องกัน และส่งผลต่อความสามารถในการจำแนก ดังนั้นการจัดการกับความแปรปรวนเหล่านี้ด้วยการเพิ่มข้อมูล (Data augmentation) หรือการสร้างชุดข้อมูลที่มีความหลากหลาย ตลอดจนการเลือกตัวแบบที่มีความสามารถในการเรียนรู้แบบไม่ขึ้นกับมุมมองหรือบริบท จึงมีความสำคัญต่อการพัฒนาตัวแบบที่มีความแม่นยำในการจำแนกภาพอาหาร

สรุป

งานวิจัยนี้นำเสนอการพัฒนาตัวแบบการสกัดคุณลักษณะภาพด้วยโมเดล ResNet50 บนชุดข้อมูลรูปภาพอาหารไทยที่ได้รับความนิยมจากนิตยสารอาหาร 20 รายการ ซึ่งจะได้ผลลัพธ์เป็นเวกเตอร์จำนวน 2,048 องค์ประกอบ จากนั้นใช้เทคนิคการวิเคราะห์องค์ประกอบหลักช่วยลดมิติของข้อมูลเหลือองค์ประกอบที่สำคัญที่สุด 300 องค์ประกอบ ก่อนจะนำไปสร้างชุดข้อมูลแทนความหมายของภาพสำหรับนำไปเปรียบเทียบกับรูปภาพค้นหาการคำนวณค่าความคล้ายคลึงโคไซน์เป็นผลลัพธ์จากการค้นหาเรียงจากมากไปน้อย ดังนั้นจึงช่วยให้ผู้ใช้งานสามารถค้นหาภาพที่ต้องการได้อย่างรวดเร็วและแม่นยำ โดยไม่จำเป็นต้องมีข้อมูลป้ายกำกับในภาพค้นหา จึงสามารถนำไปประยุกต์ใช้งานต่างๆ ที่เกี่ยวข้องกับการค้นหารูปภาพบนฐานข้อมูลขนาดใหญ่

ผลการประเมินความแม่นยำเฉลี่ยของผลการค้นหาภาพย้อนกลับโดยผู้เชี่ยวชาญด้วยการทดสอบกับรูปภาพค้นหาที่ตัวแบบไม่เคยเห็นมาก่อนจำนวน 20 รูปภาพ พบว่า ค่าความแม่นยำเฉลี่ยจากการทดสอบ 3 ครั้งอยู่ในระดับที่น่าพึงพอใจ โดยเฉพาะ 3 อันดับแรกที่มีค่าความแม่นยำสูงถึง 0.80 และเมื่อเพิ่มจำนวนผลลัพธ์จากการค้นหาเป็น 5 และ 10 รูปภาพนั้นค่าความแม่นยำลดลงเล็กน้อยเท่ากับ 0.76 และ 0.75 ตามลำดับ ผลการประเมินดังกล่าวสอดคล้องกับพฤติกรรมของผู้ใช้ที่ต้องการเพียงผลลัพธ์ที่ตนเองสนใจเท่านั้น ผลลัพธ์จำนวนมากจากการค้นหาทั้งที่ตรงและไม่ตรงกับความต้องการ อาจทำให้ผู้ใช้เสียเวลาคัดเลือกรูปภาพที่ตรงกับความต้องการอย่างแท้จริง เนื่องจากเป็นไปได้น้อยมากที่ผู้ใช้จะเลือกดูรูปภาพจากผลการค้นหาทั้งหมด จึงเป็นการเสียโอกาสสำหรับผู้ใช้หากมีรูปภาพบางส่วนที่ไม่ถูกเรียกดู โดยผู้ใช้จะเรียกดูเอกสารไม่เกิน 5 รายการที่ค้นหา (Jansen & Spink, 2006) ทั้งนี้มีเพียงผู้ใช้ร้อยละ 28.3 เท่านั้นที่เรียกดูเพียง 2-3 รายการ อีกทั้งร้อยละ

55 จะเรียกคุณผลลัพธ์เพียง 1 รายการต่อ 1 การค้นหา ดังนั้นการค้นหารูปภาพจึงต้องมุ่งเพิ่มค่าความแม่นยำในผลการค้นหาในลำดับต้น ๆ มากกว่าการปรับปรุงเพื่อให้สามารถจัดการกับผลลัพธ์ที่มีจำนวนมากได้ดีขึ้น

อย่างไรก็ดี การค้นหาภาพอาหารไทยมีความท้าทายจากปัจจัยหลายประการ โดยเฉพาะความชัดเจนของคุณลักษณะเฉพาะในอาหารบางชนิดที่มีลักษณะคล้ายคลึงกัน ตลอดจนบริบทของภาพที่หลากหลาย อาจทำให้โมเดลมีความคลาดเคลื่อนในการสกัดคุณลักษณะ นอกจากนี้ความแปรปรวนของรูปภาพยังส่งผลต่อการสกัดคุณลักษณะ เช่น มุมมอง ขนาด สภาพแสง และความซับซ้อนของพื้นหลังจากวัตถุอื่น ๆ ที่ไม่เกี่ยวข้อง ซึ่งอาจรบกวนการจดจำลักษณะเฉพาะของอาหารได้ ผลการวิจัยนี้จึงเป็นขั้นตอนสำคัญในการประเมินประสิทธิภาพของโมเดลในสถานการณ์จริงและช่วยให้เห็นถึงจุดที่จำเป็นต้องพัฒนา เพื่อให้สามารถใช้งานได้จริงในบริบทที่หลากหลาย ไม่ว่าจะเป็นลักษณะของภาพ ความคลาดเคลื่อน หรือเงื่อนไขการใช้งานในแต่ละสภาพแวดล้อม นอกจากนี้ ยังเป็นแนวทางการวิจัยในอนาคต เพื่อปรับปรุงความแม่นยำให้ตอบสนองต่อความต้องการของผู้ใช้งาน และส่งเสริมการค้นหารูปภาพย้อนกลับในเชิงพาณิชย์หรือเชิงอุตสาหกรรมต่อไป

ข้อเสนอแนะ

การเพิ่มประสิทธิภาพในการค้นหารูปภาพย้อนกลับนั้นการใช้ชุดข้อมูลที่มีความหลากหลายและครอบคลุมลักษณะของอาหารไทยแต่ละรายการควรเป็นลำดับแรกในการพัฒนา ตลอดจนการปรับปรุงกระบวนการสกัดคุณลักษณะด้วยเทคนิคใหม่ ๆ เช่น การใช้ Attention mechanism เพื่อโฟกัสไปยังส่วนสำคัญของภาพ ร่วมกับการใช้เทคนิค Multi-scale feature extraction จะช่วยให้โมเดลให้เข้าใจลักษณะเฉพาะที่ละเอียดขึ้น และการใช้ Metric learning เพื่อเพิ่มความสามารถของระบบในการแยกแยะความคล้ายคลึงอย่างละเอียด จะช่วยยกระดับผลลัพธ์ให้ตอบโจทย์การใช้งานในสถานการณ์จริงได้ดียิ่งขึ้น อย่างไรก็ตามงานวิจัยในครั้งต่อไป ควรเปรียบเทียบประสิทธิภาพของแต่ละโมเดลในเชิงลึกมากยิ่งขึ้น โดยเฉพาะการใช้ค่าความถูกต้อง (Accuracy) ค่าการเรียกคืน (Recall) ค่าประสิทธิภาพโดยรวม (F-measure) และค่าเฉลี่ยความแม่นยำเฉลี่ย (Mean Average Precision: mAP) ซึ่งใช้ประเมินประสิทธิภาพการค้นหารูปภาพย้อนกลับได้อย่างครอบคลุมในทุกมิติ

เอกสารอ้างอิง

เรวัต แสงสุริยงค์. (2565). ความเสี่ยงของการเกิดความคลาดเคลื่อนในการวิจัยเชิงปริมาณด้านสังคมวิทยา.

วารสารวิชาการมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา, 30(1), 158-185.

Broz, M. (2024). Photo statistics: How many photos are taken every day?. Retrieved from <https://photutorial.com/photos-statistics/>

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). **ImageNet: A large-scale hierarchical image database**. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). Miami, FL, USA: IEEE.
- GemmyTheGeek. (2021). **Overcome your fear for Thai food with Foody Dudy**. Retrieved from <https://gemmythegeek.medium.com/overcome-your-fear-for-thai-food-with-foody-dudy-7ff6d7702b22>
- Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing* (3rd ed.). Pearson.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). **Deep residual learning for image recognition**. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). Las Vegas, NV, USA: IEEE.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information processing & management*, 42(1), 248-263
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.
- Keogh, E., & Mueen, A. (2017). **Curse of dimensionality**. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 314–315). Boston: Springer.
- Kiros, R., Salakhutdinov, R., & Zemel, R. (2014). Unifying visual-semantic embeddings with multimodal neural language models. *arXiv preprint arXiv:1411.2539*.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). **Introduction to information retrieval**. Cambridge University Press.
- Mishra, A., Kumar, A., Yadav, A., Ansari, A., & Gupta, N. (2020). Image Search Engine. *International Research Journal of Engineering and Technology (IRJET)*, 7(5), 2841-2844.
- Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). **CNN features off-the-shelf: An astounding baseline for recognition**. In 2014 IEEE conference on computer vision and pattern recognition workshops (pp. 806-813)
- Tourism Authority of Thailand. (2021). *Eat Like You Wanna Live GASTRONOMY FOR ALL*. Tourism Authority of Thailand

Wiens, M. (2017). **40 Thai foods we can't live without**. Retrieved from
<https://edition.cnn.com/travel/article/bangkok-food-thai-dishes/index.html>