

การประมวลผลภาพสำหรับระบบชำระเงินด้วยตนเอง โดยใช้การตรวจจับวัตถุ และการรู้จำหมวดหมู่ย่อย

Self-checkout system using vision-based object detection and subcategory recognition

ณิชารีย์ ดำรงศักดิ์¹ และ จูติรัตน์ ศิริบรรณรัตน์กุล^{1*}

Nicharee Dumrongsak¹ and Thitirat Siriborvornratanakul^{1*}

¹คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ 148 ถนนเสรีไทย แขวงคลองจั่น เขตบางกะปิ กรุงเทพมหานคร ประเทศไทย

¹Graduate School of Applied Statistics, National Institute of Development Administration (NIDA), 148 SeriThai Rd., Khlong Chan, Bang Kapi, Bangkok Thailand

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนากรอบแนวคิดสำหรับระบบชำระเงินด้วยตนเอง ที่สามารถตรวจจับและบันทึกสินค้าทั้งหมดได้โดยอาศัยภาพจากกล้องที่ติดตั้งไว้ ณ จุดชำระเงิน ด้วยการนำแบบจำลองการเรียนรู้เชิงลึกมาใช้ในการตรวจจับสินค้า ร่วมกับเทคนิคของคอมพิวเตอร์วิทัศน์ที่มีความแม่นยำ สามารถแยกแยะสินค้าที่มีลักษณะคล้ายคลึงกัน หรือกลุ่มสินค้าที่มีหมวดหมู่ย่อย (sub category) และมีความยืดหยุ่นในการรองรับสินค้าใหม่ได้โดยไม่ต้องทำการฝึกสอนแบบจำลองใหม่ ด้วยกลไกการสกัดคุณลักษณะเฉพาะ และการจับคู่คุณลักษณะกับภาพอ้างอิง ผู้วิจัยได้ทำการออกแบบกรอบแนวคิดโดยใช้แบบจำลอง YOLO ในการตรวจจับและระบุรูปทรงบรรจุภัณฑ์ของสินค้า พร้อมกำหนดตำแหน่งขอบเขต (Bounding boxes) ของสินค้าที่ตรวจพบในภาพ แล้วนำภาพในขอบเขตของสินค้าที่ตรวจพบไประบุลักษณะเฉพาะของภาพ (feature point) ด้วยอัลกอริทึม SIFT, ORB และ BRISK และจับคู่ (feature matching) กับภาพต้นแบบของสินค้าเพื่อทำนายชนิดของสินค้า จากการทดลองใช้กรอบแนวคิดนี้บนชุดข้อมูลของสินค้าอุปโภคบริโภคในประเทศไทยที่ผู้วิจัยเก็บข้อมูลมาพบว่า กรอบแนวคิดนี้สามารถนำมาใช้ในการตรวจจับและบันทึกสินค้าในระบบชำระเงินด้วยตนเองได้ โดยแบบจำลอง YOLO สามารถทำนายรูปทรงและระบุตำแหน่งของสินค้าด้วยค่า mAP เฉลี่ยอยู่ที่ 0.727 และอัลกอริทึมสำหรับระบุลักษณะเฉพาะของภาพสินค้าที่ดีที่สุดคือ SIFT โดยมีความแม่นยำในการทำนายชนิดของสินค้าถูกต้อง 82.96% กล่าวโดยสรุป จากการทดลองของงานวิจัยนี้ พบว่า การนำเทคนิคการสกัดคุณลักษณะเฉพาะ และการจับคู่คุณลักษณะกับภาพอ้างอิงอย่าง SIFT สามารถช่วยเพิ่มประสิทธิภาพของแบบจำลอง YOLO ได้เนื่องจาก SIFT ช่วยให้สามารถทำนายชนิดของสินค้าได้แม่นยำ แม้ว่าการระบุขอบเขตของสินค้านั้นจะไม่ครอบคลุมพื้นที่ทั้งหมดในภาพของตัวสินค้าก็ตาม แต่ข้อจำกัดที่พบจากการทดลอง คือ การระบุลักษณะเฉพาะของภาพด้วยอัลกอริทึม SIFT และจับคู่กับภาพต้นแบบของสินค้านั้น จะมีความถูกต้องในการทำนายต่ำ เมื่อทำนายสินค้าที่มีลักษณะเป็นรูปทรงขวด ซึ่งคาดว่าเป็นผลจากปัจจัยเรื่องของแสง และเงาที่สะท้อนจากขวด ดังนั้น ควรมีการควบคุมสิ่งแวดล้อมบริเวณจุดชำระเงิน เพื่อให้ภาพสินค้าที่จะนำไปประมวลผลมีความคมชัด ไม่มีสิ่งกีดขวาง และปราศจากแสง และเงารบกวน

คำสำคัญ: ระบบชำระเงินด้วยตนเอง, แบบจำลองการเรียนรู้เชิงลึก, สินค้าที่มีหมวดหมู่ย่อย, ลักษณะเฉพาะของภาพ

ABSTRACT

The purpose of this research was to study and develop a conceptual framework for self-checkout systems to use only cameras at the checkout station to detect goods from customers. This framework is designed to use deep learning models and computer vision to classify similar products or identify sub-category products with similar appearances but different sizes. The framework is flexible to use with no retraining whenever a new product or package is introduced by using feature point extraction and feature point matching. The researcher uses the YOLO model to detect and identify the shape of the products and also defines the bounding boxes in the detected image, and then defines the feature point of the product using feature point algorithms (e.g. SIFT, ORB, and BRISK) and feature matching with reference images. The results from the experiment on the Thai products dataset that was collected by the researcher found that this framework can be applied to effectively detect and recognize retail products in a self-checkout system. The YOLO model can predict the shape and position of the product with an average mAP of 0.727, and the best feature point algorithm for product classification from this experiment is SIFT, which has an accuracy of 82.96%. In summary, SIFT can improve the accuracy of YOLO. Because of SIFT, the system can accurately predict the product class, although the YOLO bounding box does not cover an

entire region occupied by a product. However, this experiment has a limitation because SIFT has low accuracy when predicting products in bottle shape. This is expected to be a result of the light and shadows reflected from the bottles. Therefore, the environment at the checkout area should be controlled. The product images should be clear, have no obstructions, and be free from disturbing light and shadows.

KEYWORDS: Self-checkout system, Deep learning, Sub category, Feature point

*Corresponding Author: thitirat@as.nida.ac.th

Received: 14/06/2023; Revised: 29/02/2024; Accepted: 19/06/2024

1. บทนำ

ร้านค้าปลีก (retail store) เป็นแหล่งจำหน่ายสินค้าอุปโภคบริโภคในชีวิตประจำวันของประชาชนทุกเพศทุกวัย ในประเทศไทยมีร้านค้าปลีกอยู่หลากหลายรูปแบบแยกตามขนาดของร้านและประเภทสินค้าที่วางจำหน่ายแตกต่างกัน เพื่อรองรับผู้บริโภคกลุ่มต่างๆ อาทิเช่น ซูเปอร์มาร์เก็ต เป็นร้านค้าปลีกขนาดใหญ่ที่เน้นสินค้าประเภทอาหารและของใช้จำเป็นในครัวเรือน ส่วนร้านสะดวกซื้อ จะมีขนาดของร้านที่เล็ก เน้นขายสินค้าอุปโภคบริโภคที่ใช้ในชีวิตประจำวัน เป็นต้น ดังนั้น ร้านค้าปลีกแต่ละประเภทจึงให้ประสบการณ์ในการเดินเลือกซื้อสินค้าที่แตกต่างกัน แต่เมื่อพิจารณาถึงกระบวนการชำระสินค้าในร้านค้าปลีกทุกประเภทพบว่าผู้บริโภคจะต้องนำสินค้ามาชำระเงินที่เคาน์เตอร์คิดเงิน และจะมีพนักงานบันทึกสินค้าที่ละชิ้น เพื่อคำนวณค่าสินค้าทั้งหมดพร้อมรับชำระเงิน ทว่าในสถานการณ์ที่มีผู้มาใช้บริการมาก กระบวนการดังกล่าวจะกลายเป็นจุดคอขวดที่สร้างระยะเวลาในการรอคิวให้กับผู้บริโภค หากร้านค้ามีจำนวนเคาน์เตอร์คิดเงินและพนักงานรองรับไม่เพียงพอ ซึ่งการให้บริการที่ล่าช้านี้ส่งผลต่อการตัดสินใจใช้บริการร้านค้าปลีก (Chaiyasoonthorn & Suksa-ngiam, 2011) จึงเป็นที่มาให้ระบบชำระเงินด้วยตนเอง (self-checkout system) ถูกพัฒนาขึ้นเพื่อนำมาใช้แก้ปัญหาข้างต้น และเป็นการปรับปรุงประสบการณ์ของผู้บริโภคให้สอดคล้องกับวิถีชีวิตของผู้คนในยุคสมัยใหม่ ที่สามารถเรียนรู้การใช้งานเทคโนโลยีต่างๆ ด้วยตนเองได้มากยิ่งขึ้น

ระบบชำระเงินด้วยตนเอง (self-checkout system) เป็นระบบที่พัฒนาขึ้นมาเพื่อให้ผู้บริโภคสามารถทำการบันทึกสินค้าด้วยตนเองและระบบจะทำการคำนวณค่าสินค้าทั้งหมดพร้อมรับชำระเงิน โดยที่ไม่จำเป็นต้องมีพนักงานมาประจำเครื่องเพื่อทำการให้ ซึ่งปัจจุบันในยุคที่เทคโนโลยีมีความก้าวหน้ามาก ระบบชำระเงินด้วย

ตนเองได้ถูกนิยมนำมาพัฒนาด้วยการเรียนรู้เชิงลึก (deep learning) สำหรับใช้ในการตรวจจับสินค้า (object detection) ซึ่งการใช้การเรียนรู้เชิงลึก จะช่วยให้ระบบสามารถตรวจจับและบันทึกสินค้าทั้งหมดได้เองโดยอาศัยภาพจากกล้องที่ติดตั้งไว้ ผู้บริโภคเพียงนำสินค้าทั้งหมดมาวางในพื้นที่ที่กำหนด และระบบประมวลผลแล้วชำระเงิน ดังนั้น วิธีนี้จึงเป็นวิธีที่สะดวกสบายต่อผู้บริโภค และสามารถช่วยลดเวลาในการทำรายการได้อีกด้วย

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจจับสินค้าด้วยการเรียนรู้เชิงลึก พบว่า มีงานวิจัยในต่างประเทศมากมายที่ใช้แบบจำลองของการเรียนรู้เชิงลึกแตกต่างกัน แต่แบบจำลองที่ให้ประสิทธิภาพดีในปัจจุบัน จำเป็นที่จะต้องเรียนรู้ใหม่ทุกครั้งเมื่อมีสินค้าชนิดใหม่เข้ามา จึงกลายเป็นปัญหาสำคัญเมื่อนำมาใช้ในธุรกิจค้าปลีกที่มักเกิดสินค้าใหม่หรือเปลี่ยนแปลงรวดเร็วของผลิตภัณฑ์แทบทุกวัน ซึ่งการรวบรวมข้อมูลของสินค้าใหม่และการเรียนรู้ของแบบจำลองในแต่ละครั้งจะต้องใช้เวลานาน ก่อให้เกิดปัญหาเรื่องความยืดหยุ่นในการนำไปใช้งานจริง (Wei et al., 2020) นอกจากนี้การใช้แบบจำลองของการเรียนรู้เชิงลึกในการตรวจจับสินค้า มักพบปัญหาเรื่องของการแยกแยะสินค้าที่มีลักษณะคล้ายคลึงกัน หรือกลุ่มสินค้าที่มีหมวดหมู่ย่อย (sub category) เช่น สินค้ายี่ห้อเดียวกัน แต่มีรสชาติแตกต่างกัน หรือมีขนาดแตกต่างกัน ซึ่งสินค้ากลุ่มนี้จะมีรูปลักษณ์และรายละเอียดของผลิตภัณฑ์ที่แตกต่างกันเพียงเล็กน้อยจนทำให้แบบจำลองทำนายสินค้าผิดพลาดไปรวมถึงเรื่องของสินค้าที่มีความใกล้เคียงกันและการวางทับซ้อนกันของสินค้า มักจะเป็นปัญหาสำหรับการใช้แบบจำลองของการเรียนรู้เชิงลึกในการตรวจจับสินค้าในปัจจุบัน (Hsia et al., 2022) ดังนั้นจากปัญหาของการนำแบบจำลองการเรียนรู้เชิงลึกมาใช้ในการตรวจจับสินค้าของระบบชำระเงินด้วยตนเองที่กล่าวมาข้างต้นนี้ทำให้ผู้วิจัยมีความสนใจที่จะ

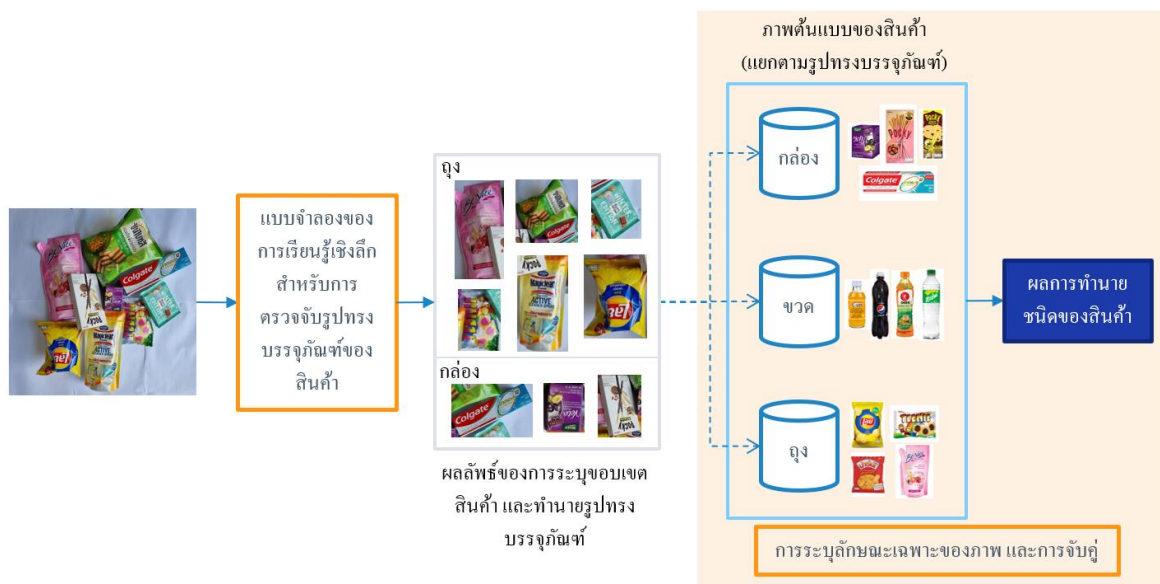
ทำการค้นคว้าวิจัยถึงความเป็นไปได้ในการพัฒนารอบแนวคิดที่จะสามารถแก้ปัญหาดังกล่าว โดยอาศัยการผสมผสานของแบบจำลองการเรียนรู้เชิงลึกที่มีความรวดเร็วในการตรวจจับ ร่วมกับเทคนิคของคอมพิวเตอร์วิทัศน์ที่มีความแม่นยำในการแยกแยะสินค้ามีลักษณะคล้ายคลึงกัน และมีความยืดหยุ่นในการรองรับสินค้าใหม่ได้โดยไม่ต้องทำการฝึกสอนแบบจำลองใหม่ทุกครั้งที่มีการเพิ่มสินค้าใหม่

วัตถุประสงค์และวิธีการทดลอง

1. กรอบแนวคิด

งานวิจัยนี้ได้แรงบันดาลใจในการสร้างกรอบแนวคิดมาจากการสังเกตรูปทรงบรรจุภัณฑ์ของสินค้าที่ขายในร้านค้าปลีก ผู้วิจัยได้เล็งเห็นว่าสินค้าส่วนใหญ่มีกรูปร่างของบรรจุภัณฑ์ที่ใกล้เคียงกัน อีกทั้งบรรจุภัณฑ์ยัง

เป็นส่วนใหญ่ไม่ได้ถูกเปลี่ยนแปลงบ่อยนัก เมื่อเทียบกับผลผลิตของผลิตภัณฑ์ โดยสามารถจัดแบ่งเป็นกลุ่มของรูปทรงได้ อาทิเช่น ขวด กล่อง กระป๋อง ถุง ด้วย เป็นต้น ผู้วิจัยจึงอาศัยแบบจำลองของการเรียนรู้เชิงลึกในการคัดแยกสินค้าทั้งหมดจากรูปทรงบรรจุภัณฑ์ เพื่อกำหนดพื้นที่ที่สนใจ (region of interest : ROI) ที่คาดว่าจะเป็นบริเวณของสินค้าแต่ละชิ้น จากนั้น ในการจำแนกและจดจำเพื่อบันทึกสินค้าแต่ละชิ้น จะอาศัยเทคนิคของการระบุลักษณะเฉพาะของภาพ (feature point) และการจับคู่ (feature matching) ในการเปรียบเทียบบริเวณของสินค้าแต่ละชิ้นที่ได้จากแบบจำลองที่กล่าวมาข้างต้น เทียบกับภาพต้นแบบของสินค้า ที่จะเป็นภาพที่แสดงผลผลิตของผลิตภัณฑ์ โดยภาพต้นแบบจะมาจากฐานข้อมูลสินค้าที่เก็บภาพของสินค้าทุกชิ้นที่ขายในร้านค้าปลีก โดยแบ่งแยกตามกลุ่มของรูปทรงบรรจุภัณฑ์ เพื่อให้่ายต่อการนำมาเปรียบเทียบ ดังนั้น กรอบแนวคิดนี้จึงมีภาพรวมของการทำงานดังในรูปที่ 1



รูปที่ 1 ภาพกรอบแนวคิดของงานวิจัย

2. การเตรียมข้อมูล

ในการฝึกสอนแบบจำลองของการเรียนรู้เชิงลึก เพื่อให้สามารถคัดแยกสินค้าจากรูปทรงบรรจุภัณฑ์ได้ จะต้องใช้ชุดข้อมูลรูปภาพสินค้าที่จัดชำระเงิน งานวิจัยนี้จึงเลือกใช้ชุดข้อมูล RPC dataset (Wei et al., 2019) ซึ่งเป็นชุดข้อมูลสำหรับสนับสนุนงานวิจัยที่เกี่ยวกับการตรวจจับและการรับรู้สินค้าในจุดชำระเงิน ซึ่งประกอบด้วยภาพทั้งหมด 83,739 ภาพ จากสินค้าทั้งหมด 200 ชนิด (class) โดยเป็นภาพของสินค้าแต่ละชนิดแบบ 360 องศา จำนวน 53,739 และเป็นภาพถ่ายสินค้าที่จัดชำระเงิน

30,000 ภาพ ซึ่งภาพถ่ายสินค้าที่จัดชำระเงินของชุดข้อมูลนี้ได้จำลองตามสถานการณ์การชำระเงินอย่างสมจริง สินค้าจะถูกสุ่มเลือกมาวางรวมกันอย่างอิสระ มีความยุ่งเหยิง และซับซ้อน ผู้วิจัยได้นำสินค้าทั้งหมดของ RPC dataset มาจัดกลุ่มตามรูปทรงของบรรจุภัณฑ์ สามารถแบ่งออกมาได้ทั้งหมด 6 กลุ่ม ได้แก่ ถุง กล่อง ขวด กระป๋อง ถ้วย และรูปทรงที่ไม่มีรูปแบบ เมื่อพิจารณาจำนวนสินค้าในแต่ละกลุ่มพบว่า สินค้าที่มีรูปทรงของบรรจุภัณฑ์เป็นถุงมีจำนวนมากที่สุด ในชุดข้อมูลนี้ รองลงมาคือกล่องและขวดตามลำดับ ดังแสดงในรูปที่ 2 โดยสัดส่วนของ 3

อันดับแรกคิดเป็นสัดส่วนจะครอบคลุมถึง 76% ของชุดข้อมูลนี้ ดังนั้น ผู้วิจัยจึงเลือกรูปทรงบรรจุภัณฑ์ที่เป็นถุง กล่อง และขวด สำหรับการทดสอบสมมติฐานเบื้องต้นใน

การทำแบบจำลองของการเรียนรู้เชิงลึก เพื่อคัดแยกสินค้าจากรูปทรงบรรจุภัณฑ์

รูปทรงของบรรจุภัณฑ์	จำนวนสินค้า	จำนวน annotations	ตัวอย่างสินค้า
ถุง	75	113,893	
กล่อง	55	78,746	
ขวด	22	32,075	
กระป๋อง	23	31,916	
รูปทรงที่ไม่มีรูปแบบ	15	21,561	
ถ้วย	10	16,142	
จำนวนรวม	200	294,333	

รูปที่ 2 จำนวนของสินค้าใน RPC dataset ที่จัดกลุ่มตามรูปทรงบรรจุภัณฑ์ และตัวอย่างสินค้าในแต่ละกลุ่ม

3. การเก็บข้อมูลสำหรับทดสอบ

เพื่อให้แบบจำลองของการเรียนรู้เชิงลึกมีความยืดหยุ่นกับชุดข้อมูลที่หลากหลาย สามารถประยุกต์ (generalize) กับสินค้าที่ไม่เคยเห็นได้ดี ผู้วิจัยจึงได้ทำการสร้างชุดข้อมูลจากสินค้าอุปโภคบริโภคในประเทศไทยจำนวน 61 ชนิด (TH-products dataset) โดยแบ่งเป็นสินค้าที่มีบรรจุภัณฑ์รูปทรงถุง จำนวน 22 ชนิด สินค้าที่มีบรรจุภัณฑ์รูปทรงกล่อง จำนวน 17 ชนิด และสินค้าที่มีบรรจุภัณฑ์รูปทรงขวด จำนวน 22 ชนิด ซึ่งจากสินค้าทั้งหมดจะมี 39 ชนิด ที่เป็นสินค้าประเภทหมวดหมู่ย่อย ชุดข้อมูลนี้จะประกอบด้วยภาพต้นแบบของแต่ละสินค้า ซึ่งเป็นภาพจากอินเทอร์เน็ต หรือภาพเรนเดอร์ (render) ของสินค้า มีอย่างน้อย 1-3 ภาพต่อชนิด และกำหนดชื่อของภาพด้วยหมายเลข class ของสินค้า เพื่อใช้อ้างอิงในตอนทำนายผลจากการระบุลักษณะเฉพาะของภาพ และการจับคู่ โดยภาพต้นแบบทั้งหมดในชุดข้อมูลนี้มีจำนวน 95 ภาพ แสดงดังรูปที่ 3 และมีภาพถ่ายสินค้าที่จุดขายเงินจำนวน 120 ภาพ โดยถ่ายด้วยกล้องมือถือ Samsung

galaxy S21 ultra ความละเอียด 108 MP จากหลังของภาพถ่ายสินค้าที่จุดขายเงินจะเป็นกระดาษสีขาว ขนาด 45x45 เซนติเมตร เพื่อให้สินค้าแยกจากฉากหลังอย่างชัดเจน แต่ละภาพจะสุ่มเลือกสินค้า 2-11 ชิ้น และจัดวางอย่างเป็นอิสระ ส่วนการจัดเก็บข้อมูลของสินค้าประเภทหมวดหมู่ย่อย สินค้าจะถูกบันทึกเป็นคนละ class กัน เช่น เลย์ร็อค มี 3 ขนาด จะแยกเป็น 3 class ดังแสดงในรูปที่ 4



รูปที่ 3 ภาพต้นแบบของทุกสินค้าใน TH-products dataset

Class	สินค้า	ภาพต้นแบบ
0	Lays Classic Potato Salt 73g.	
1	Lays Rock Potato Salt 73g.	
2	Lays Rock Potato Salt 48g.	
3	Lays Rock Potato Salt 27g.	

รูปที่ 4 ตัวอย่างการเก็บข้อมูลของสินค้าประเภทหมวดหมู่ย่อย

4. แบบจำลองของการเรียนรู้เชิงลึกสำหรับการตรวจจับรูปทรงบรรจุภัณฑ์ของสินค้า

จากการศึกษางานวิจัยที่เกี่ยวข้องกับระบบชำระเงินด้วยตนเองพบว่าแบบจำลองของการเรียนรู้เชิงลึกที่นิยมนำมาใช้ในการตรวจจับสินค้า และจดจำประเภทของสินค้า พร้อมระบุตำแหน่งขอบเขต (Bounding boxes) ของสินค้าที่ตรวจพบในภาพ คือ แบบจำลอง YOLO เนื่องจากเป็นแบบจำลองที่มีจุดเด่นในเรื่องของความเร็วในการตรวจจับสินค้าเมื่อเทียบกับแบบจำลองอื่นๆ (Liu et al., 2021; James et al., 2021; Shoman et al., 2022) ซึ่งสาเหตุที่ทำให้ YOLO ตรวจจับได้เร็วนั้นคือ YOLO มีหลักการทำงานแบบ 1 ขั้นตอน (one-stage object detector) ซึ่งสามารถทำนายขอบเขตและคำนวณความน่าจะเป็นของชนิดสินค้าได้ภายในขั้นตอนการประมวลผลเดียวกัน (Redmon et al., 2016) ผู้วิจัยจึงเลือกใช้แบบจำลอง YOLO ในการตรวจจับและระบุรูปทรงบรรจุภัณฑ์ของสินค้า ซึ่งปัจจุบัน YOLO ได้ถูกพัฒนาถึงเวอร์ชันที่ 5 แล้ว (YOLOv5) จุดเด่นของ YOLOv5 คือสามารถใช้งานด้วย PyTorch และมีความเร็วในการประมวลผลที่รวดเร็วขึ้น ซึ่งยังคงประสิทธิภาพที่ใกล้เคียงกับ YOLO v4 ผู้วิจัยจึงเลือกใช้แบบจำลอง YOLOv5 โดยนำแบบจำลองมาจาก <https://github.com/ultralytics/yolov5> และผู้วิจัยเลือกทำการฝึกสอนแบบจำลอง YOLOv5 ขนาด medium (YOLOv5m) เนื่องจากมีสัดส่วนของความเร็วในการประมวลผลเทียบกับความถูกต้องอยู่ในระดับกลางเมื่อเทียบจากขนาดทั้งหมด ซึ่งผู้วิจัยคิดว่าเป็นขนาดที่เหมาะสมกับงานวิจัยนี้ ที่ต้องการความสมดุลในเรื่องของความเร็วและความถูกต้อง ผู้วิจัยจึงได้นำแบบจำลอง YOLOv5m มาฝึกสอนด้วยข้อมูล RPC ที่จัดเตรียมไว้สำหรับบรรจุภัณฑ์รูปทรงถุง กล่อง และขวด โดยแบ่งชุด

ข้อมูลออกเป็นสัดส่วน 70-15-15 เพื่อฝึกสอน และทดสอบ แล้วทำการฝึกสอนทั้งแบบจำลอง (ไม่มีการ freezing layer) โดยใช้ทรัพยากรของ Google Colaboratory ในการฝึกสอนแบบจำลองทั้งหมด ผู้วิจัยได้ใช้ pretrained weights บน COCO dataset ที่ทางผู้พัฒนา YOLO จัดเตรียมไว้ และทำการย่อขนาดภาพก่อนเริ่มฝึกสอนเป็นขนาด 512x512 pixel พร้อมทั้งกำหนดรายละเอียดในการฝึกสอนด้วย $lr = 0.01$, $weight_decay = 0.0005$ และใช้ optimizer SGD จากนั้นทำการฝึกสอนทุกเลเยอร์ของแบบจำลองด้วยจำนวน 16 batch 120 epochs

การวัดผลของแบบจำลองจะใช้ตัวชี้วัด 3 ค่า ได้แก่ precision, recall และ mAP@.5 ซึ่งคำนวณตามหลักการของตัวชี้วัดประสิทธิภาพสำหรับแบบจำลองการตรวจจับวัตถุ (Padilla et al., 2020) โดยผลลัพธ์ที่ได้จากการฝึกสอนแบบจำลอง YOLOv5m ดังแสดงในตารางที่ 1 จะพบว่า ผลลัพธ์บนข้อมูลชุดทดสอบมีค่าความแม่นยำในการทำนายรูปทรง และระบุตำแหน่งของสินค้าด้วยค่า mAP เฉลี่ยอยู่ที่ 0.995 โดยเมื่อพิจารณาค่า precision ที่บ่งบอกถึงความแม่นยำของแบบจำลองในการทำนายรูปทรง จะพบว่า แบบจำลองนี้ มีความแม่นยำในการทำนายรูปทรงกล่อง และถุง มากกว่ารูปทรงขวด ส่วนค่า recall ที่แสดงถึงความสามารถในการตรวจจับ พบว่าแบบจำลองนี้สามารถทำได้ดีกับทั้ง 3 รูปทรง อย่างไรก็ตาม ในส่วนของผลลัพธ์ที่มีค่าเบี่ยงเบนมาตรฐานเท่ากับศูนย์นั้น แสดงให้เห็นว่าผลลัพธ์ของแบบจำลองที่ได้ทำการฝึกสอนจำนวน 3 รอบ ให้ผลลัพธ์ที่แตกต่างกันน้อยมาก และมีค่าความแม่นยำสูง เป็นผลมาจากการที่ข้อมูลในชุดทดสอบเป็นข้อมูลสินค้าใน distribution เดียวกันกับข้อมูลชุดฝึกสอน และมีสินค้าอยู่เพียง 200 ชนิด จึงทำให้แบบจำลองสามารถเรียนรู้สินค้าได้อย่างครบถ้วน

ตารางที่ 1 ผลลัพธ์จากการทำนายรูปทรงบรรจุภัณฑ์ และระบุตำแหน่งของสินค้า บน RPC dataset ชุดทดสอบ โดยฝึกสอนแบบจำลอง 3 รอบ (ด้วย parameter เดียวกัน) แล้วนำมาคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน

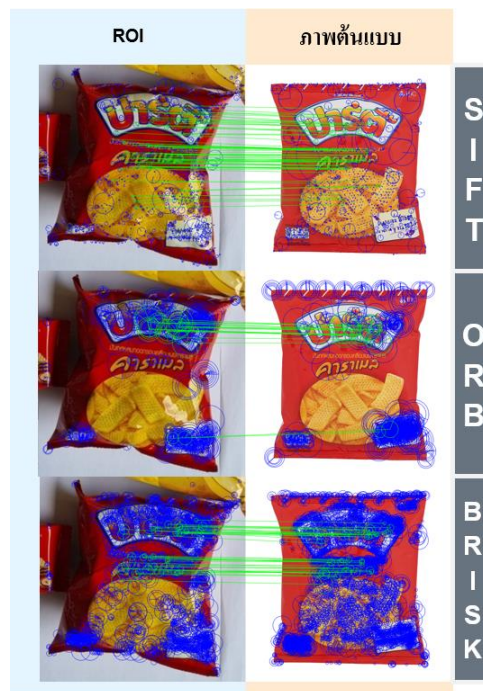
Class	Precision	Recall	mAP@.5
กล่อง	1.000 ± 0.001	0.997 ± 0.001	0.995 ± 0.000
ขวด	0.998 ± 0.001	0.999 ± 0.001	0.995 ± 0.000
ถุง	1.000 ± 0.000	0.998 ± 0.002	0.995 ± 0.000
เฉลี่ย	0.999 ± 0.001	0.998 ± 0.000	0.995 ± 0.000

5. การระบุลักษณะเฉพาะของภาพ และการจับคู่

ด้วยวิวัฒนาการของเทคโนโลยีทำให้งานวิจัยในด้านคอมพิวเตอร์วิทัศน์ถูกพัฒนาอย่างต่อเนื่อง อัลกอริทึมในการระบุลักษณะเฉพาะของภาพก็เป็นหนึ่งในชิ้นงานสำคัญที่ผู้วิจัยให้ความสำคัญ เนื่องจากสามารถนำมาใช้ประโยชน์ในงานประมวลผลภาพได้หลากหลาย อัลกอริทึมที่โด่งดัง และเป็นที่ยอมรับแรกถูกพัฒนาและเสนอโดย Lowe ในปี 1999 ได้นำเสนออัลกอริทึม Scale Invariant Feature Transform algorithm (SIFT) (Lowe, 1999) ซึ่งคุณสมบัติเด่นของ SIFT คือ การระบุลักษณะเฉพาะของภาพ โดยไม่ขึ้นกับขนาด หรือทิศทางของวัตถุในภาพ ทำให้สามารถนำไปใช้งานได้สะดวก และมีความแม่นยำ สืบเนื่องมาจากหลักการในการทำ Keypoint Detection ของ SIFT นั้น จะสร้าง Scale-Space ของภาพด้วยการทำ Gaussian Blur กับภาพต้นแบบหลายๆ ครั้ง และหาความแตกต่างระหว่างภาพ Gaussian Blur ที่อยู่ใกล้ติดกันด้วย Difference of Gaussian (DoG) สิ่งนี้ช่วยให้ SIFT สามารถค้นหาคุณสมบัติของภาพโดยไม่ขึ้นกับขนาดของภาพได้ จากนั้นทำ Keypoint Localization เพื่อคัดกรองจำนวน feature point ที่ได้ ด้วยการปรับแต่งตำแหน่งของจุดสำคัญที่เป็นไปได้ และกำจัด keypoint ที่ไม่จำเป็นออกไป ถัดมาในปี 2011 ได้มีอัลกอริทึมที่พัฒนาและเสนอขึ้นมาเพื่อปรับปรุงในเรื่องของความเร็วการประมวลผล โดยอัลกอริทึมแรกถูกเสนอโดย Rublee และคณะ คือ ORB (Rublee et al., 2011) ซึ่ง ORB เป็นการพัฒนาโดยผสมผสานระหว่าง FAST keypoint detector กับ BRIEF descriptor เข้าด้วยกัน และ

อีกอัลกอริทึม คือ BRISK (Leutenegger et al., 2011) ที่ถูกเสนอโดย Leutenegger และคณะ โดยจุดเด่นของ BRISK คือความเร็ว และมีต้นทุนในการคำนวณที่ต่ำกว่า SIFT ซึ่งทั้ง 2 อัลกอริทึมมีหลักการในการทำ Keypoint Detection ด้วยการใช้ FAST detector ในการหาจุดสำคัญของภาพ เพื่อระบุจุดที่อาจจะเป็นมุมและขอบ จากการเปลี่ยนแปลงของความเข้มของพิกเซลโดยรอบ

ในงานวิจัยนี้ผู้วิจัยได้เลือกอัลกอริทึม SIFT, ORB และ BRISK ในการนำมาระบุลักษณะเฉพาะของภาพสินค้าและภาพต้นแบบ แล้วนำลักษณะเฉพาะที่ได้จากทั้ง 2 ภาพมาทำการจับคู่ด้วยวิธี FLANN based Matcher (FLANN) วิธีการดังกล่าวจะอาศัยหลักการ Approximate Nearest Neighbors ทำให้มีความรวดเร็วในการประมวลผล โดยในการจับคู่และทำนายผล ผู้วิจัยจะพิจารณาจำนวนคู่ของจุดที่มีลักษณะเฉพาะตรงกัน โดยจะทำนายผล class ของสินค้า จากภาพต้นแบบที่มีจำนวนคู่ของจุดที่มีลักษณะเฉพาะตรงกันเป็นจำนวนมากที่สุด และใช้ชื่อไฟล์ของภาพต้นแบบอ้างอิงหมายเลข class ในการทำนายผล การระบุลักษณะเฉพาะของภาพในแต่ละอัลกอริทึม และการจับคู่ของจุดด้วยวิธี FLANN แสดงดังในรูปที่ 5



รูปที่ 5 การระบุลักษณะเฉพาะ (วงกลมสีน้ำเงิน) ของภาพในแต่ละอัลกอริทึม และการจับคู่ด้วยวิธี FLANN (เส้นสีเขียว)

ในส่วนของการพิจารณาความถูกต้องของการระบุลักษณะเฉพาะของภาพ และการจับคู่ ผู้วิจัยจะพิจารณาจากการทำนาย class ของสินค้า และวัดผลจากการคำนวณค่าความถูกต้อง (Classification Accuracy)

6. ผลการทดลอง

งานวิจัยนี้แบ่งการทดลองออกเป็น 3 ขั้นตอน ได้แก่ 1) การทดลองแบบจำลองของการเรียนรู้เชิงลึกสำหรับการตรวจจับรูปทรงบรรจุภัณฑ์ของสินค้า 2) การทดลองอัลกอริทึมสำหรับระบุลักษณะเฉพาะของภาพสินค้า และการจับคู่ภาพต้นแบบกับภาพสินค้าที่จุดขายเงิน 3) การทำงานร่วมกันของการตรวจจับรูปทรงบรรจุภัณฑ์ และการระบุลักษณะเฉพาะของภาพ พร้อมการจับคู่โดยการทดลองทั้งหมดจะทดสอบด้วยชุดข้อมูล TH-products dataset

ผลจากการใช้แบบจำลอง YOLOv5m ที่ฝึกสอนด้วย RPC dataset เมื่อนำมาทดสอบบนชุดข้อมูลใหม่ ที่ไม่เคยเห็นมาก่อน ผลลัพธ์ที่ได้พบว่าความแม่นยำในการทำนายรูปทรงและระบุตำแหน่งของสินค้าลดลง ค่า mAP เฉลี่ยอยู่ที่ 0.759 ดังแสดงในตารางที่ 2 รูปทรงที่ทำนายผิดพลาดมากที่สุด คือ บรรจุภัณฑ์รูปทรงถุง รองลงมาคือ กล่อง และขวด ตามลำดับ แต่ถ้าพิจารณาค่า recall ที่แสดงถึงความสามารถในการตรวจจับ จะพบว่าแบบจำลองนี้สามารถตรวจจับรูปทรงขวดได้เพียง 66% ของชุดข้อมูล นอกจากนี้ จากผลการทดลองผู้วิจัยคิดว่าสาเหตุที่ค่า mAP ลดลงนั้น เกิดจากการฝึกสอนแบบจำลองด้วยชุดข้อมูลที่ยังครอบคลุมรูปทรงบรรจุภัณฑ์ของสินค้าไม่มากพอ ทำให้เมื่อนำแบบจำลองดังกล่าวมาใช้บนชุดข้อมูลใหม่ที่มีรูปทรงบรรจุภัณฑ์แตกต่างจากชุดข้อมูลฝึกสอนจึงทำนายผิดพลาดไป การแก้ไขสามารถทำได้โดยการเพิ่มรูปทรงของบรรจุภัณฑ์ในแต่ละ class บนชุดข้อมูลฝึกสอนให้มีความหลากหลายมากขึ้น

ตารางที่ 2 ผลลัพธ์จากการทำนายรูปทรงบรรจุภัณฑ์และระบุตำแหน่งของสินค้าบน TH-products dataset โดยฝึกสอนแบบจำลอง 3 รอบ (ด้วย parameter เดียวกัน) แล้วนำมาคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน

Class	Precision	Recall	mAP@.5
กล่อง	0.792 ± 0.071	0.577 ± 0.182	0.695 ± 0.100
ขวด	0.950 ± 0.018	0.642 ± 0.047	0.797 ± 0.033
ถุง	0.708 ± 0.178	0.646 ± 0.095	0.688 ± 0.077
เฉลี่ย	0.817 ± 0.078	0.622 ± 0.102	0.727 ± 0.054

ถัดมาในส่วนของการทดลองการทดลอง อัลกอริทึมสำหรับระบุลักษณะเฉพาะของภาพสินค้าและการจับคู่ภาพต้นแบบกับภาพสินค้าที่จุดชำระเงิน ผู้วิจัยทำการทดลองโดยกำหนดให้ข้อมูลที่ได้จากแบบจำลองของการเรียนรู้เชิงลึกถูกต้องสมบูรณ์ 100% ผลจากการทดลองใช้อัลกอริทึมที่แตกต่างกันในการระบุลักษณะเฉพาะของภาพสินค้าและภาพต้นแบบ แล้วนำมาจับคู่เพื่อทำนายชนิดของสินค้า โดยวัดผลด้วยค่าความถูกต้อง (classification accuracy) ที่คำนวณมาจากสัดส่วนของจำนวนผลลัพธ์ที่ทำนายถูกต้องเทียบกับจำนวนของการทำนายทั้งหมด แล้วนำมาคิดเป็นเปอร์เซ็นต์ ซึ่งผลการทดลองดังแสดงในตารางที่ 3 พบว่า การใช้ SIFT ให้ความแม่นยำในการทำนายชนิดของสินค้าถูกต้องสูงสุดถึง

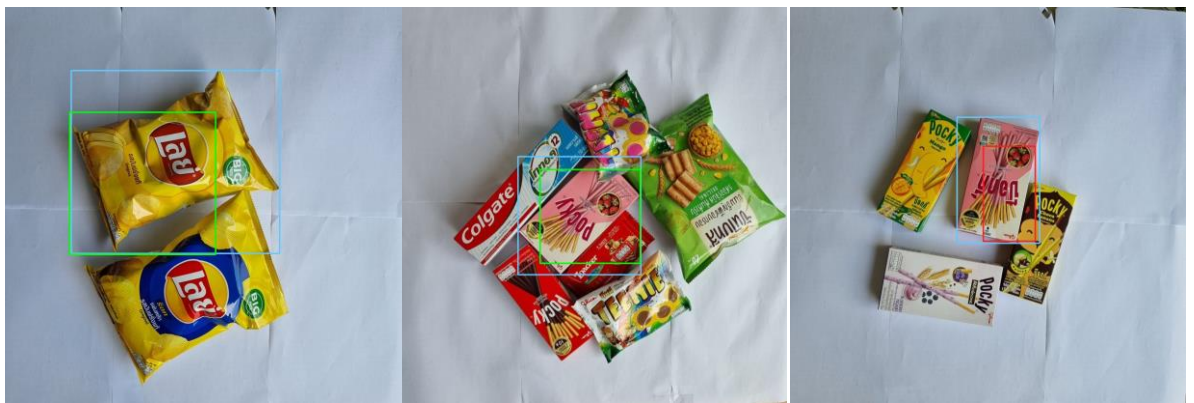
82.96% รองลงมาคือ BRISK 72.91% และ ORB 57.41% ตามลำดับ การที่ SIFT สามารถทำนายได้แม่นยำที่สุด เนื่องจากคุณสมบัติของ SIFT สามารถรองรับการระบุคุณลักษณะของภาพได้โดยไม่ขึ้นกับขนาด หรือทิศทางของสินค้า ในขณะที่ BRISK และ ORB สามารถทำนายสินค้านามบัตรรูปทรงกล่องได้แม่นยำที่สุด เนื่องจากการระบุคุณลักษณะของภาพของสองอัลกอริทึมนี้จะเน้นตรงจุดที่เป็นมุมและขอบ นอกจากนี้ จากผลการทดลองยังชี้ให้เห็นว่ารูปทรงที่มีค่าความถูกต้องในการทำนายต่ำทั้ง 3 อัลกอริทึม คือรูปทรงขวด อย่างไรก็ตามในภาพรวมของทั้ง 3 รูปทรง การใช้ SIFT ก็ยังคงให้ค่าความถูกต้องอยู่ในเกณฑ์ที่ดีและยอมรับได้ ผู้วิจัยจึงเลือก SIFT เพื่อนำไปทดลองต่อในการทดลองที่ 3

ตารางที่ 3 ผลลัพธ์ค่าความถูกต้อง (classification accuracy) จากการใช้อัลกอริทึมต่างๆ ในการระบุลักษณะเฉพาะของภาพสินค้า และการจับคู่ บน TH-products dataset

รูปทรง	SIFT	ORB	BRISK
กล่อง	96.09%	77.65%	92.18%
ขวด	69.18%	42.77%	42.14%
ถุง	82.33%	52.21%	78.71%
เฉลี่ย	82.96%	57.41%	72.91%

การทดลองสุดท้ายผู้วิจัยได้นำผลลัพธ์ที่ได้จากการทดลองที่ 1 ซึ่งระบุรูปทรงของสินค้า และตำแหน่งขอบเขต (Bounding boxes) ของสินค้าแต่ละชิ้นที่มีอยู่ในภาพจุดชำระเงิน นำมาส่งต่อเป็นข้อมูลให้กับการระบุลักษณะเฉพาะของภาพสินค้าด้วย SIFT โดยไม่มีการคัดกรองผลลัพธ์ของตำแหน่งขอบเขตที่ได้จากแบบจำลอง YOLO ส่งผลให้ถ้าหากผลลัพธ์ในการทดลองที่ 1 มีการระบุขอบเขตของสินค้าผิดพลาดไป ผลการทำนาย class ของสินค้าด้วย SIFT ก็จะผิดพลาดด้วยเช่นกัน ดังนั้นผู้วิจัยจึงพิจารณาผลลัพธ์ 2 แบบ ได้แก่ การพิจารณาค่าความถูกต้องในการทำนายชนิดของสินค้า จากค่า mAP ที่ $IOU > 50\%$ และแบบที่สองคือการพิจารณาเฉพาะค่าความ

ถูกต้องในการทำนายชนิดของสินค้า (ไม่สนใจค่า IOU) ผลลัพธ์ในส่วนแรก queพิจารณาจากค่า mAP จะอยู่ที่ 44.73% ซึ่งเมื่อดูจากผลลัพธ์ที่ได้จะพบว่าผลการทดลองนี้ มีความแม่นยำในการระบุขอบเขตของสินค้าได้ปานกลาง และการทำนายชนิดของสินค้านั้นอยู่ในเกณฑ์ที่ต่ำ แต่หากพิจารณาเฉพาะการทำนาย class ของสินค้าจะพบว่า SIFT ทำนายออกมามีค่าความถูกต้องเพิ่มขึ้น อยู่ที่ 51.44% ผลลัพธ์ดังแสดงในตารางที่ 4 นั้นแสดงให้เห็นว่า SIFT สามารถทำนายและให้ผลลัพธ์ที่ถูกต้องได้ถึงแม้ว่าการระบุขอบเขตของสินค้านั้นจะครอบคลุมพื้นที่ไม่เต็มสินค้า ดังแสดงตัวอย่างในรูปที่ 6



รูปที่ 6 ภาพตัวอย่างของผลลัพธ์ที่ทำนายถูก แม้การระบุขอบเขตของสินค้าจะครอบคลุมพื้นที่ไม่เต็มสินค้า โดยกรอบสีฟ้า หมายถึง พื้นที่ของสินค้าทั้งหมด กรอบสีเขียว หมายถึง ผลลัพธ์ที่ทำนายถูกและครอบคลุมพื้นที่สินค้าเกิน 50% และกรอบสีแดง หมายถึง ผลลัพธ์ที่ทำนายถูกและครอบคลุมพื้นที่ไม่ถึง 50% ของสินค้า

ตารางที่ 4 ผลลัพธ์ค่าความถูกต้อง (classification accuracy) จากการพิจารณาการทำนายชนิดของสินค้าด้วย SIFT โดยไม่สนใจค่า IOU

รูปทรง	ค่าความถูกต้องในการทำนายชนิดของสินค้า (ไม่สนใจค่า IOU)
กล่อง	51.79%
ขวด	38.30%
ถุง	64.25%
เฉลี่ย	51.44%

ในแง่ของเวลาการประมวลผลและตรวจจับสินค้าจากภาพที่จุดชำระเงิน ที่มีขนาด 512x512 pixel พบว่าความเร็วในการประมวลผลและตรวจจับ เฉลี่ยอยู่ที่ 39 วินาทีต่อรูป ซึ่งความเร็วจะขึ้นอยู่กับจำนวนของสินค้าที่วางอยู่ในจุดชำระเงิน ยิ่งมีจำนวนสินค้ามากก็จะใช้เวลามาก เมื่อพิจารณาเวลาในการประมวลผลของแต่ละส่วนแยกกัน จะพบว่า ส่วนแรกที่ใช้แบบจำลอง YOLO ในการระบุรูปทรงของสินค้า และตำแหน่งขอบเขต (Bounding boxes) ของสินค้าแต่ละชิ้นที่มีอยู่ในภาพจุดชำระเงิน จะใช้เวลาเฉลี่ยอยู่ที่ 0.015 วินาที และในส่วนที่สองการระบุลักษณะเฉพาะของภาพสินค้าด้วย SIFT นั้น จะใช้เวลาเฉลี่ยอยู่ที่ 0.375 วินาที ทั้งนี้ เวลาในส่วนที่สองนั้นจะขึ้นอยู่กับจำนวนภาพต้นแบบที่มีอยู่ในฐานข้อมูลด้วย ยังมีจำนวนมากก็จะใช้เวลาในการประมวลผล class ของสินค้านาน

7. สรุปผลการทดลอง

จากผลการทดลองทั้งหมดที่กล่าวมาข้างต้นสามารถสรุปได้ว่ากรอบแนวคิดที่ผู้วิจัยพัฒนาโดยอาศัยการผสมผสานของแบบจำลองการเรียนรู้เชิงลึกที่มีความรวดเร็วในการตรวจจับ ซึ่งผู้วิจัยเลือกใช้แบบจำลอง YOLOv5 มาผสมผสานร่วมกับเทคนิคของคอมพิวเตอร์วิทัศน์อย่าง SIFT ในการระบุลักษณะเฉพาะของภาพสินค้าสามารถนำมาใช้ในการตรวจจับและบันทึกสินค้าในระบบชำระเงินด้วยตนเองได้ โดยข้อดีของกรอบแนวคิดนี้ คือความยืดหยุ่นในการรองรับสินค้าชนิดใหม่ที่เพิ่มขึ้นมาหรือสินค้าเดิมที่มีการเปลี่ยนแปลงลวดลายบรรจุภัณฑ์ กรอบแนวคิดนี้จะสามารถทำงานต่อเนื่องได้ทันที เพียงแค่เพิ่มภาพต้นแบบของสินค้าใหม่เข้ามาในฐานข้อมูล ไม่จำเป็นต้องทำการฝึกสอนแบบจำลองใหม่ทุกครั้ง ซึ่งผู้วิจัยเสนอแนะให้เก็บภาพต้นแบบอย่างน้อย 2 ภาพต่อชนิดของสินค้า เพื่อให้ครอบคลุมทั้งด้านหน้าและด้านหลังของบรรจุภัณฑ์สินค้า นอกจากนี้กรอบแนวคิดนี้ยังสามารถรองรับสินค้าประเภทหมวดหมู่ย่อยได้ ทั้งกลุ่มที่มีลวดลายใกล้เคียงกัน และกลุ่มที่มีลวดลายเหมือนกันแต่ขนาดแตกต่างกัน

ในส่วนของการใช้อัลกอริทึม SIFT สำหรับระบุลักษณะเฉพาะของภาพสินค้า และการจับคู่ งานวิจัยนี้พบว่า วิธีการดังกล่าวสามารถช่วยเพิ่มประสิทธิภาพของ

แบบจำลอง YOLO ได้ เนื่องจากการทำ Keypoint Detection ของ SIFT นั้น สร้าง Scale-Space ของภาพด้วยการทำ Gaussian Blur กับภาพต้นแบบหลายๆครั้ง และหาความแตกต่างระหว่างภาพ Gaussian Blur ที่อยู่ในลำดับติดกันด้วย Difference of Gaussian (DoG) จึงช่วยให้ SIFT สามารถค้นหาคุณสมบัติของภาพโดยไม่ขึ้นกับขนาดและทิศทางของภาพได้ และยังคงมีความแม่นยำในการทำนายชนิดของสินค้า แม้ว่าการระบุขอบเขตของสินค้านั้นจะไม่ครอบคลุมพื้นที่ทั้งหมดในภาพของตัวสินค้า อย่างไรก็ตามข้อจำกัดที่ผู้วิจัยพบจากการทดลองคือ การควบคุมสิ่งแวดล้อมบริเวณจุดชำระเงิน เนื่องจากภาพที่จะสามารถนำไปประมวลผลได้นั้น ควรที่จะต้องมีความคมชัด ไม่มีสิ่งกีดขวาง และปราศจากแสงและเงารบกวน นอกจากนี้ จากผลการทดลองแสดงให้เห็นว่าสินค้ากลุ่มรูปทรงขวดนั้นมีผลลัพธ์ในการทำนายที่ไม่ดีนักเนื่องจากแบบจำลอง YOLO ที่ผู้วิจัยฝึกสอนมาไม่สามารถตรวจจับรูปทรงขวดได้ดีนัก จึงส่งผลให้สินค้ารูปทรงขวดไม่ถูกระบุรูปทรงของสินค้า และตำแหน่งขอบเขต เพื่อส่งต่อมาประมวลผลการระบุลักษณะเฉพาะของภาพสินค้า และการจับคู่ แล้วทำนาย class ของสินค้า ดังนั้น วิธีการแก้ไขคือควรปรับปรุงแบบจำลองให้มีความแม่นยำมากขึ้นก่อน เพื่อให้แบบจำลองสามารถตรวจจับสินค้าในจุดชำระเงินได้อย่างครบถ้วน

อย่างไรก็ตามงานวิจัยนี้ยังมีจุดที่สามารถปรับปรุงและพัฒนาประสิทธิภาพของกรอบแนวคิดที่นำเสนอได้ โดยปรับเปลี่ยนในส่วน of แบบจำลองการเรียนรู้เชิงลึก ไปใช้แบบจำลองในกลุ่ม Segmentation แทนการใช้ Object detection เพื่อให้สามารถระบุขอบเขตของสินค้าได้ตรงตามความเป็นจริง และไม่ติดสินค้าอื่นๆที่ใกล้เคียงกัน ทำให้การนำบริเวณดังกล่าวไปทำนายชนิดของสินค้ามีความแม่นยำมากยิ่งขึ้น นอกจากนี้ ยังสามารถต่อยอดงานวิจัยนี้ได้ด้วยการคัดเลือกอัลกอริทึมสำหรับระบุลักษณะเฉพาะของภาพสินค้า และการจับคู่ ที่มีประสิทธิภาพมากกว่าการใช้ SIFT และ FLANN ทั้งในด้านความเร็ว และความแม่นยำ เพื่อเพิ่มประสิทธิภาพของกรอบแนวคิดนี้ ทั้งนี้ในส่วนของการประเมินประโยชน์และความคุ้มค่าของระบบสำหรับการนำไปประยุกต์ใช้จริงนั้น การนำระบบไปทำการทดลองในร้านค้าที่สินค้ามีการใช้การสแกนบาร์โค้ดอยู่เดิมแล้วเปรียบเทียบกับร้านค้าที่

สินค้าไม่มีการใช้บาร์โค้ด จะช่วยให้ผู้วิจัยสามารถเปรียบเทียบความคุ้มค่าและคุณประโยชน์ของระบบลักษณะนี้ในบริบทที่แตกต่างกันได้ดียิ่งขึ้น

เอกสารอ้างอิง

- Chaiyasoonthorn, W., & Suksa-ngiam, W. (2011). Factors influencing store patronage: A study of modern retailers in Bangkok Thailand. *International Journal of Trade, Economics and Finance*, 2(6), 520. (in Thai)
- Hsia, C. H., Chang, T. H. W., Chiang, C. Y., & Chan, H. T. (2022). Mask R-CNN with new data augmentation features for smart detection of retail products. *Applied Sciences*, 12(6), 2902.
- James, N., Antony, N. T., Shaji, S. P., Baby, S., & Annakutty, J. (2021). Automated Checkout for Stores: A Computer Vision Approach. *Revista GEINTEC-Gestao Inovacao e Tecnologias*, 11(3), 1830-1841.
- Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011, November). BRISK: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*. pp. 2548-2555.
- Liu, H., Li, Y., & Liu, D. (2021, July). Object detection and recognition system based on computer vision analysis. In *Journal of Physics: Conference Series* (Vol. 1976, No. 1, p. 012024). IOP Publishing.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*. pp. 1150-1157.
- Padilla, R., Netto, S. L., & Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*. pp. 237-242.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 779-788.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision*. pp. 2564-2571.
- Shoman, M., Aboah, A., Morehead, A., Duan, Y., Daud, A., & Adu-Gyamfi, Y. (2022). A Region-Based Deep Learning Approach to Automated Retail Checkout. *arXiv preprint arXiv:2204.08584*.
- Wei, X. S., Cui, Q., Yang, L., Wang, P., & Liu, L. (2019). RPC: A large-scale retail product checkout dataset. *arXiv preprint arXiv:1901.07249*.
- Wei, Y., Tran, S., Xu, S., Kang, B., & Springer, M. (2020). Deep learning for retail product recognition: Challenges and techniques. *Computational intelligence and neuroscience*, 2020(1), 8875910.