บทความวิจัย

Computer Science Department
Faculty of Informatics, Mahasarakham University

แอปพลิชันรวมสินค้าไลฟ์สดด้วยการประมวลผลทางภาพ

Application Live Shop With Image Processing

กฤตเมธ บัวสิงห์, กิตติศักดิ์ มนพรมมา, ดร.พัฒนพงษ์ ชมพูวิเศษ
63011212023@msu.ac.th, 63011212071@msu.ac.th, phatthanaphong.c@msu.ac.th

บทคัดย่อ

ในปัจจุบันเทคโนโลยีในชีวิตประจำวันมากขึ้น เช่น การซื้อของออนไลน์ และ การไลฟ์สดขายของและงานการตรวจจับเป้าหมายที่อยู่ในภาพมีความพัฒนาขึ้นอย่างต่อเนื่องทางซึ่งการที่จะซื้อของออนไลน์ในไลฟ์สดนั้นต้องใช้เวลาในการซื้อ โดยในงานนี้จะใช้หลักการ การเรียนรู้เชิงลึก มาฝึกอมรบโมเดลเพื่อตรวจจับเป้าหมาย ชื่อว่า Yolov5 และหลังจากการตรวจจับได้จะใช้ Tesseract OCR ในการแปลงตัวอักษรในเป้าหมายที่ตรวจได้มาแสดงผลใน mobile application

คำสำคัญ Yolov5, Tesseract OCR,การเรียนรู้เชิงลึก

1.บทนำ

ปัจจุบันเทคโนโลยีอิเล็กทรอนิกส์ มีความสำคัญในการดำรงชีวิตของผู้คนในปัจจุบันมากขึ้น ผู้คนมากมายได้ทำการใช้เทคโนโลยีในการค้าขายของออนไลน์ผ่านการไลฟ์และได้เกิดการพัฒนาของ ปัญญาประดิษฐ์ (AI : Artificial Intelligence) คือเครื่องจักรหรือเทคโนโลยีที่มีฟังก์ชันทีมีความสามารถในการทำความเข้าใจเรียนรู้องค์ความรู้ต่างๆ อาทิเช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหาต่างๆ เครื่องมือที่มีความสามารถเหล่านี้ก็ถือว่าเป็นปัญญาประดิษฐ์ เพราะฉะนั้นจึงสามารถกล่าวได้ว่า AI ถือกำเนิดขึ้นเมื่อเครื่องจักรมีความสามารถที่จะเรียนรู้นั่นเอง

ปัจจุบันการที่จะขายพัสดุออนไลน์ในการไลฟ์สดนั้น ใช้เวลาที่ค่อนข้างมากที่จะต้องให้ลูกค้าเข้ามารับชมเพื่อเลือกสิ้นค้าทีละชิ้น และทำให้ลูกค้าบางรายที่ไม่สะดวกมารับชมไม่ได้เลือกซื้อสินค้าที่ต้องการจึงทำให้พวกเราได้ทำการพัฒนาแอปพลิชั่นรวมสินค้าไลฟ์สดขึ้นมาเพื่อแก้ไขปัญหาดังกล่าว

ดังนั้นเพื่อให้เกิดความสะดวกสบายพวกเราจึงได้พัฒนาแอปพลิเคชั่นเพื่อตอบโจทย์การใช้งานของแม่ค้าและลูกค้า ก็คือแอปพลิชั่นรวมสินค้าไลฟ์สดซึ่งจะทำการรวมเอาสินค้าต่างๆในไลฟ์สดของแม่ค้าแต่ละคนมารวมไว้ในแอปพลิเคชั่นเดียว โดยการใช้ image processing ในการตรวจจับข้อมูลสินค้าในไลฟ์สด และนำข้อมูลมาเก็บในฐานข้อมูลและนำไปแสดงผลในแอปพลิเคชั่น และเมื่อเลือกสินค้าที่ต้องการจะสามารถดูข้อมูลของสินค้าและไปสู่หน้าไลฟ์สดของแม่ค้าได้ เพื่อให้ลูกค้าได้เลือกสินค้ากันอย่างสะดวกและแม่ค้าก็ไม่เสียลูกค้าที่ไม่สะดวกรับชม

2.ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 Flutter [1] คือ Framework ที่ใช้สร้าง UI สำหรับ mobile application ที่สามารถทำงานได้ทั้ง IOS และ Android ในเวลาเดียวกันโดยที่ใช้ source code ตัวเดียวกัน โดยภาษาที่ใช้ใน Flutter นั้นจะเป็นภาษา dart ซึ่งถูกพัฒนาโดย Google และ ยังเป็น open source ที่สามารถใช้งานได้ฟรี

2.1.2 Firebase เป็นหนึ่งในผลิตภัณฑ์ของ Google โดย Firebase [2]คือ Platform ที่รวบรวมเครื่องมือต่าง ๆ สำหรับการจัดการในส่วนของการเก็บข้อมูล ซึ่งทำให้สามารถ สร้าง Mobile Application ได้อย่างม

ประสิทธิภาพ และยังลดเวลาและค่าใช้จ่ายของการทำ Server side หรือการวิเคราะห์ข้อมูลให้อีกด้วย โดยมีทั้งเครื่องมือที่ฟรี และเครื่องมีที่มีค่าใช้จ่าย Firebase มีบริการให้ใช้หลายอย่าง สามารถแบ่งเป็นหมวดหมู่ดังนี้

2.1.3 การประมวลผลภาพ (Image Processing) [3]เป็นการนำภาพมาประมวลผลหรือคิดคำนวณด้วย คอมพิวเตอร์แล้วใช้กรรมวิธีใดๆมากระทำกับข้อมูลภาพเพื่อให้ได้ภาพที่มีคุณสมบัติตามต้องการทั้งในเชิงคุณภาพและปริมาณ โดยมีขั้นตอนต่างๆที่สำคัญ คือ การทำให้ภาพมีความคมชัดมากขึ้น การกำจัดสัญญาณรบกวนออกจากภาพ การแบ่งส่วนของวัตถุที่สนใจออกมาจากภาพ

2.1.4 OCR Optical Character Recognition หรือ "การรู้จำอักขระด้วยแสง" เป็นเทคโนโลยีที่แปลงเอกสาร รูปภาพ หรือแม้แต่ ลายเซ็นอิเล็กทรอนิกส์ ให้อยู่ในรูปแบบตัวอักษร หรือ ข้อความ (Plain Text) ไฟล์ดิจิทัลที่สามารถสืบค้นได้ด้วยคำค้นหา

2.1.5 Object Detection [5]การตรวจจับวัตถุ คือ เทคโนโลยีในทางคอมพิวเตอร์หลักการที่เกี่ยวกับ Computer Vision และ Image Processing ที่ใช้ในงาน AI ตรวจจับวัตถุชนิดที่กำหนด เช่น มนุษย์ รถยนต์ อาคาร ที่อยู่ในรูปภาพ หรือวิดีโองาน Object Detection การตรวจจับวัตถุในรูปภาพสามารถเจาะลึกลงไปได้อีกหลายแขนง เช่น การทำ ตรวจจับหน้าคน ตรวจจับคนเดินถนน สามารถประยุกต์ใช้ได้หลากหลาย เช่น ใช้ในงานรักษาความปลอดภัย และรถยนต์ไร้คนขับ เป็นต้น แบบดั้งเดิมสามารถทำงานได้ผลลัพธ์ที่ดี ในบางสถานการณ์ในสภาพแวดล้อมที่แปรปรวนระบบตรวจจับใบหน้าในกล้องดิจิตอล

2.1.6 YOLO (You Only Look Once) เป็นอัลกอริทึม [6]ที่นำแนวความคิดของการทำนายตำแหน่งและขนาดของ กล่องจากความน่าจะเป็นที่กล่องนั้น จะเป็นกรอบล้อม วัตถุ แต่สิ่งที่ YOLO มีความสามารถและความเร็ว เหนือกว่าอัลกอริทึมอื่น เช่น Faster R-CNN ที่มีการ ทำงานในลักษณะที่จะทำนายตำแหน่งของกรอบล้อม วัตถุและค่อยนำวัตถุในกล่องนั้น ไปผ่านแบบจำลองเพื่อ ทำนายวัตถุในกล่อง แต่ YOLO นั้นจะทำนายทั้งกรอบ ล้อมวัตถุ และความน่าจะเป็นของวัตถุบางส่วนที่อยู่ใน กรอบออกมาพร้อมกัน

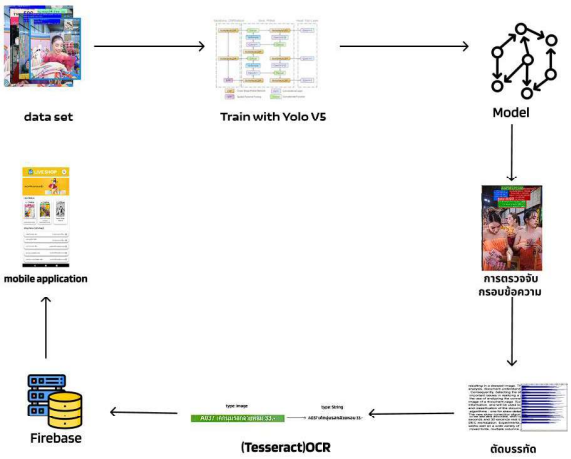
2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1Safety Helmet Detection Based on YOLOv5 [7]

เป็นงานวิจัยของ Fangbo Zhou, Huailin Zhao , Zhen Nie มหาวิทยาลัย Shanghai Institute of Technology [7]ที่พัฒนาระบบตรวจสอบความปลอดภัยของคนงาน ซึ่งระบบจะทำการตรวจจับภาพของคนงาน และ จะแสดงกรอบที่มีคำอธิบาย โดยได้นำรูปภาพจากอินเทอร์เน็ตจำนวน 6045 ภาพมาทำการทดสอบ โดยศีรษะของคนงานที่ไม่สวมหมวกนิรภัยจะมีคำอธิบายว่า"Alarm" และสำหรับศีรษะของคนงานที่สวมหมวกนิรภัยจะมีคำอธิบายว่า“Helmet” โดยใช้ algorithm ที่ทันสมัยที่สุดในการตรวจสอบ algorithm งานตรวจจับวัตถุถูกใช้อย่างแพร่หลายในความเป็นจริง เป้าหมายของการตรวจจับคือการค้นหาวัตถุทั้งหมดที่น่าสนใจในภาพ ซึ่งจะมี 2 งานย่อยคือการกำหนดหมวดหมู่และการระบุตำแหน่งของวัตถุ แม้ว่าอัลกอริธึมการตรวจจับอัลกอริธึม การเปลี่ยนแปลง ของ สภาพอากาศ การกระจายตัวของคนงานที่ไม่สม่ำเสมอและหมวกกันน็อคแบบต่างๆ ทำให้ความแม่นยำนั้นรับประกันได้ยากว่ามันสามารถ

ทำงานได้ถูกต้อง จึงได้นำตัว YOLOv5 มาใช้ซึ่งมีความเร็วและความแม่นยำสูง

3.ขั้นตอนการดำเนินงาน



ภาพประกอบที่ 1 ภาพรวมของระบบ

การเก็บข้อมูล เก็บข้อมูลโดยการบันทึกคลิปวิดีโอ แล้วนำวิดีโอมาเปิดและบันทึกเป็นภาพนิ่ง

เตรียมข้อมูลในการ training โดยบันทึกรูปภาพจาก Video ไลฟ์สดที่เก็บมาได้ โดยจะทำการบันทึกเฉพาะช่วงที่มีการเปลี่ยนแปลงของข้อความบนหน้าจอไลฟ์สดดังโดยจะได้จาก ร้าน FIRST SHOP V2 จำนวน 33 ภาพ จากร้าน KANYA SHOP ขายถูกทุกอย่าง จำนวน 28 ภาพ จากร้าน มหัศจรรย์"วันของAuuM" จำนวน 33 ภาพ รวมกันทั้งหมด 92 รูปภาพ

การวาดภาพผลเฉลย ทำการวาดภาพผลเฉลยรูปภาพผ่าน Roboflow ผลเฉลยในรูปภาพที่เตรียมมาทั้งหมด โดยจะจัดหมวดหมู่ในที่ที่สนใจเป็น “box”และในพื้นที่ที่ไม่สนใจเป็น “no” ในส่วนนี้จะนำข้อมูลไปทำการ Generate รูปภาพเพิ่มเพื่อเพิ่มจำนวนของข้อมูลที่จะทำไป train จะทำการ Generate ทั้งหมด 3 ครั้ง โดยจะได้รูปภาพจากการ Generate ทั้งหมด 558 ภาพ

เทรนข้อมูลด้วย yolov5 ในส่วนนี้จะเป็นการอธิบายถึงโครงสร้างของ yolov5 และการเทรนข้อมูลโดยจะมีขั้นตอนการทำงานแบ่งเป็นส่วนหลักๆดังนี้

ในส่วนที่ 1 จะเป็นสำหรับการคัดกรองภาพเพื่อดึงลักษณะเด่นของรูปภาพออกมาโดยจะมีขั้นตอนการทำงานดังนี้

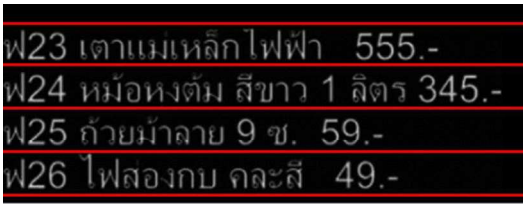
1.Convolution Layer
2.Rectified Linear Unit (ReLU)
3.ReLU leaky
4.Pooling layer
5.Flatten
6.Fully Connected
7.Softmax

ในส่วนที่ 2 เป็นส่วนที่ทำหน้าที่ Detect Object โดยใช้การสร้าง bounding box โดยการคำนวณพิกัดของ object ที่ Convolution Layer เพื่อทำนายความน่าจะเป็นของแต่ละ Object Class และ bounding โดยจะผ่านกระบวนการดังต่อไปนี้

1.box coordinates
2.Bounding Box
3.anchor
4.Intersection over Union
5.Non-Max Suppression

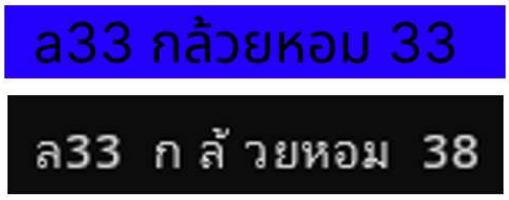
การตัดบรรทัดด้วยเทคนิค projection profile ในส่วนของงานในส่วนนี้จะเป็นการนำรูปภาพจากที่ทำการ detect มาได้มาทำการแปลงเป็นภาพ binary และทำการใช้เทคนิค horizontal projection profile มาทำการแบ่งบรรทัดเพื่อทำให้สินค้าแยกชิ้นกันอย่างชัดเจน

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2 ผลลัพท์การแยกบรรทัด

ขั้นตอนการเรียกใช้ Tesseract OCR เพื่อทำการนำข้อมูลที่ได้จากการตัดบรรทัดมาแปลงเป็นตัวอักษรเพื่อนนำเข้าไปเก็บใน database



ภาพประกอบที่ 3 ตัวอย่างการทำ OCR

ในส่วนของ mobile application จะใช้ Flutter ในการสร้าง application ที่ทำการล็อกอินด้วย facebookได้และเชื่อมต่อกับ firebase และดึงข้อมูลจาก filebase มาแสดงผลได้และเพื่อลดความผิดพลาดในการแสดงผลข้อมูลจะมีรูปภาพของร้านค้าในขนาดที่ทำการไลฟ์สดขึ้นโชว์ในหน้ารายละเอียดสินค้าด้วย

ในส่วนของการจัดการข้อมูลจะใช้ Cloud Firestore จัดเก็บข้อมูล โดยใช้ฐานข้อมูล NoSQL

ในส่วนของการวัดประสิทธิภาพการจะวัดการ Detect และการ OCR

1.การวัดประสิทธิภาพการ Detect วัดด้วย Mean Average Precision (mAP)

2.การวัดประสิทธิภาพการ OCR วัดด้วย CER(Character Error Rate)

5.สรุปและอภิปรายผลการทดลอง

ในการตรวจจับกรอบของข้อความทาง ผู้จัดทำได้เลือกใช้สถาปัตยกรรม Yolov5 ในการฝึกอบรมข้อมูล โดยในการทดลองนี้จะใช้ภาพในการฝึกอบรมโมเดล โดยจะแบ่งภาพออกเป็น 3 ส่วนได้แก่ Training จำนวน 558 รูปภาพ Validation จำนวน 17 รูปภาพ และ Testing จำนวน 13 รูปภาพ โดยรวมแล้วจะใช้รูปภาพทั้งหมด 588 รูปภาพ โดยโมเดลจะมีประสิทธิภาพดังนี้

ตารางที่ 1 ประสิทธิภาพของโมเดล

Class	Images	Instances	P	R	F1 score
box	558	594	1	0.997	0.9985
No box	558	558	0.997	1	0.9985

จากผลการทดสอบนี้, โมเดลแสดงความแม่นยำสูง (Precision) และความครอบคลุมสูง (Recall) สำหรับทั้งคลาส "box" และคลาส "False" ด้วยค่า F1 score ที่สูงถึง 0.9985 ซึ่งแสดงถึงประสิทธิภาพดีในการตรวจจับหรือการจำแนกวัตถุในทั้งสองคลาส

ในการประเมินประสิทธิภาพ ของ Tesseract OCR จะใช้ CERในการประเมินโดยจากผลการทดลองข้างต้นพบว่าประสิทธิภาพในการทำนายผล Model โดยมีค่า CER ต่ำที่สุดอยู่ที่ 0.0% และค่า CER สูงที่สุดอยู่ที่ 44.26% ค่าเฉลี่ย Character Error Rate (CER) ที่นำมาทดสอบ อยู่ที่ 8.731%

5.สรุปและอภิปรายผลการทดลอง

หลังจากพัฒนาโมเดลและทดสอบโมเดลการตรวจจับกรอบของข้อความโดยใช้รูปภาพจากร้านทั้งสิ้น 92 รูปภาพ แบ่งออกเป็น Training จำนวน 62 รูปภาพเพื่อจะทำการหาตำแหน่งของกรอบข้อความในไลฟ์สดและ ตรวจจับไม่ได้จำนวน 2 รูปภาพดังนั้นโมเดลนี้มีความแม่นยำประมาณ 93.33% ใน

Computer Science Department
Faculty of Informatics, Mahasarakham University

การตรวจจับกรอบข้อความในภาพทั้ง 30 ภาพ ที่ทดสอบและเมื่อนำข้อมูลรหัสสินค้า ชื่อสินค้า ราคาสินค้า ที่ตรวจจับมาได้มาแปลงจากภาพเป็นข้อความภาษาไทยด้วย Tesseract OCR โดยเมื่อผลทำนายและผลเฉลยมา ทดสอบประสิทธิภาพจะมีค่า CER ต่ำที่สุดอยู่ที่ 0.0% และค่า CER สูงที่สุดอยู่ที่ 44.26% ค่าเฉลี่ยของ CER ที่นำมาทดสอบจะอยู่ที่ 8.731% ของร้านค้าได้มีการฝึกอบรมโมเดล Validation จำนวน 17 รูปภาพ และ Testing จำนวน 13 รูปและเนื่องจากมีข้อมูล Training set จำนวนไม่มากจึงนำไปทำ Data Augmentation เพื่อเพิ่มจำนวน Training set เป็น 558 รูปภาพ เพื่อนำมาอบรมโมเดลเพื่อตรวจจับกรอบของข้อความ และเมื่อทำการทดสอบประสิทธิภาพของโมเดลพบว่า ภาพรวมของโมเดล โมเดลสามารถตรวจจับรูปภาพได้ถูกต้อง 28 รูปภาพ จากรูปภาพทั้งหมด 30

6.เอกสารอ้างอิง

1 Hizoka, "Flutter," [Online]. Available: https://shorturl.asia/zPKJZ. [Accessed 15 07 2022].

2 J. Saengow, "Firebase," [Online]. Available: https://shorturl.asia/g6H0J. [Accessed 07 16 2022].

3 N. Jesadapatrakul, "Image Processing," [Online]. Available: https://shorturl.asia/4LirK. [Accessed 10 06 2022].

4 NUMKINGSTON, "Optical Character Recognition," [Online]. Available: https://shorturl.asia/nmq7h. [Accessed 12 07 2022].

5 S. KANOKTIPSATHARPORN. [Online]. Available: https://shorturl.asia/twFLQ. [Accessed 14 07 2022].

6 S. D. R. G. a. A. F. J. Redmon, "ieeexplore," [Online]. Available: https://ieeexplore.ieee.org/document/7780460. [Accessed 14 07 2022].

7 H. Z. a. Z. N. F. Zhou, "Safety Helmet Detection Based on YOLOv5," [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9362711. [Accessed 07 15 2022].

8 I. Kunakorntum, "medium," [Online]. Available: http://surl.li/ctfuc. [Accessed 15 07 2022].

9 E. BARRETT, "fortune," [Online]. Available: https://fortune.com/2018/10/28/in-china-facial-recognition-tech-is-watching-you/. [Accessed 14 07 2022].

Computer Science Department
Faculty of Informatics, Mahasarakham University