

การวิเคราะห์แนวโน้มแฟชั่นโดยใช้เทคนิคการเรียนรู้เชิงลึก

Fashion Trend Analysis using Deep Learning Techniques

บุญยง วงศ์สิริสวัสดิ์*

ดร.ธนภัทร พังคะจิตร**

บทคัดย่อ

ธุรกิจด้านค้าปลีกแฟชั่น (Fashion Retail) เป็นหนึ่งในธุรกิจพื้นฐานโดยสินค้าประเภทเครื่องแต่งกายถือเป็นหนึ่งในสินค้าที่จำเป็นต่อการดำรงชีวิต นอกจากนี้ยังเป็นธุรกิจที่สร้างกำไรในระดับสูงให้แก่ผู้ประกอบการอีกด้วย ดังนั้นการทราบถึงแนวโน้มการแต่งกายของประชากรแยกตามเพศและแยกตามอายุ ย่อมทำให้ผู้ประกอบการสามารถออกแบบผลิตภัณฑ์ได้เหมาะสมกับความต้องการของลูกค้า โดยอาศัยความก้าวหน้าทางด้านเทคโนโลยีทางด้านคอมพิวเตอร์วิชั่นทำให้สามารถเก็บข้อมูลของลูกค้าผ่านกล้องวงจรปิดและนำเสนอข้อมูลเพื่อประโยชน์ในการวิเคราะห์แนวโน้มผ่านทางโปรแกรมระบบรายงานอัจฉริยะในแบบ Real-time ได้

งานสารนิพนธ์ฉบับนี้ได้นำเสนอวิธีการสกัดข้อมูลการแต่งกาย เพศและช่วงอายุของลูกค้าโดยใช้เทคนิคการเรียนรู้เชิงลึก โดยข้อมูลนำเข้าเป็นข้อมูลภาพวิดีโอจากกล้องวงจรปิดภายในร้านค้า เพื่อเพิ่มประสิทธิภาพในการสกัดข้อมูลจึงได้แบ่งเป็น 2 ขั้นตอนได้แก่ขั้นตอนที่ 1 ทำการตรวจหา"คน"ในรูปภาพ โดยใช้ ImageAI และ OpenCV ขั้นตอนที่ 2 นำภาพ"คน"ที่ตรวจหาได้ไปสกัดหารูปหน้า เพื่อทำนายเพศช่วงอายุ อารมณ์ และสกัดการแต่งกายโดยแบ่งเป็นท่อนบน ท่อนล่าง และรองเท้า โดยใช้เทคนิค YOLO หลังจากเสร็จสิ้นกระบวนการสกัดข้อมูลแล้วจะมีการจัดเก็บได้ในฐานข้อมูล NoSQL และมาทำการแสดงผลข้อมูลด้วยภาพและวิเคราะห์โดยใช้ระบบรายงานอัจฉริยะ (Qlik Sense) โดยวิธีการสกัดที่นำเสนอให้ความแม่นยำสูงในการสกัดข้อมูลจากฐานข้อมูลทดสอบ รวมถึงระบบรายงานผลสามารถตอบโต้การใช้งานของระบบ

ABSTRACT

Fashion retail is an essential business involved with necessary things in human life. In addition, a high level of profitability can be generated to entrepreneurs. Therefore, knowing fashion trends can lead to appropriate product design corresponded to the needs of customers. Regarding to advanced technology in computer vision field, it is possible to collect customer data via CCTV. Then, they can be presented in business intelligence system to further trend analysis in real-time.

*นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

**ที่ปรึกษาสารนิพนธ์หลัก

In this work, we present the approach to extract the customer information e.g. gender, age and clothes using deep learning techniques. In order to increase the efficiency of data extraction, we first detect human in the image using ImageAI and OpenCv. Then, age range, gender and emotion can be predicted from human face. In addition, clothes can be extracted by dividing into the top body, bottom body and shoes using YOLO technique. Once finishing extraction process, all information are stored in NoSQL database and visualized by using business intelligence system (Qlik Sense). Our approach offers high accuracy when evaluates on the test dataset. Including a report system that can meet the needs of the system.

บทนำ

จากสถานการณ์ทางเศรษฐกิจที่ผันผวนและความไม่แน่นอนหลายปัจจัยที่ต่อเนื่องมาถึงปี 2020 การสร้างความเชื่อมั่นในการบริโภคสินค้าและบริการ ทำให้ธุรกิจต่างๆ ต้องมีการปรับตัว และความเสี่ยงต่อการถูก Disruption ทั้งจากคู่แข่ง นวัตกรรม เทคโนโลยี หรือแม้แต่จากพฤติกรรมผู้บริโภคที่เปลี่ยนแปลงไป ดังนั้น การศึกษาพฤติกรรมผู้บริโภค โดยการเก็บรวบรวมข้อมูลเกี่ยวกับแนวโน้มแฟชั่น (Fashion trend survey) จึงเป็นหนึ่งในหลักการตลาดที่สำคัญ

ในการเก็บรวบรวมข้อมูลเกี่ยวกับแนวโน้มแฟชั่นในปัจจุบัน สามารถเก็บได้โดยการให้พนักงานทำหน้าที่สำรวจโดยให้ลูกค้ากรอกข้อมูลผ่านแบบสำรวจ หรือ กรอกผ่านแบบสำรวจออนไลน์ที่จัดทำขึ้น ซึ่งปัญหาที่พบคือข้อมูลที่เก็บมาแต่ไม่สามารถใช้งานได้ เช่น ข้อมูลไม่ครบถ้วนและไม่เข้าใจผลิตภัณฑ์ พนักงานต้องใช้ความรู้ลึกในการกรอกข้อมูล เช่นการสำรวจอารมณ์ตั้งแต่ลูกค้าเข้ามาซื้อสินค้า อารมณ์ขณะเลือกซื้อสินค้า จนถึงอารมณ์ขณะชำระเงินซื้อสินค้า ปัญหาในการรวบรวมแบบสำรวจพบว่าจำนวนแบบสำรวจค่อนข้างน้อยและล่าช้า

ปัจจุบันวงการค้าปลีกแฟชั่น เริ่มมีการนำเทคโนโลยีดังกล่าวมาใช้เพิ่มประสิทธิภาพด้านต่างๆ เช่นด้านการบริหารความสัมพันธ์ลูกค้า (Customer Relationship Management : CRM) และกระบวนการรู้จักลูกค้า (Know Your Customer : KYC) เช่นระบบการจดจำใบหน้า การติดตามและวิเคราะห์พฤติกรรมลูกค้า นอกจากนี้ยังสามารถช่วยลดต้นทุนด้านเวลาและทรัพยากรคน โดยสามารถนำมาใช้ตรวจนับสินค้า แบ่งชนิดหรือจำแนกวัตถุ และตรวจสอบสินค้าต่างๆ ได้เป็นต้น

ในงานวิจัยชิ้นนี้เสนอขั้นตอนของการเก็บรวบรวมข้อมูลแนวโน้มแฟชั่น ได้จากการนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) มาสอนให้คอมพิวเตอร์สามารถมองเห็นและแยกแยะวัตถุและบอกความสัมพันธ์ของวัตถุ สามารถเปรียบเทียบกับการเรียนรู้ และแยกแยะสิ่งต่างๆ ที่ได้จากการเรียนรู้ ในช่วงเวลาหนึ่งและสถานที่ต่างๆ ได้โดยใช้การสกัดข้อมูลโดยการการตรวจหา "คน" ในรูปภาพ โดยใช้ ImageAI และ OpenCV และนำภาพ "คน" ที่ตรวจหาได้ไปสกัดหารูปหน้า เพื่อทำนายเพศ ช่วงอายุ อารมณ์ และสกัดการแต่งกายโดยแบ่งเป็น ท่อนบน ท่อน

ล่าง และรองเท้าโดยใช้เทคนิค YOLO หลังจากเสร็จสิ้นกระบวนการสกัดข้อมูลแล้วจะมีการจัดเก็บได้ในฐานข้อมูล NoSQL และมาทำการแสดงผลข้อมูลด้วยภาพและวิเคราะห์โดยใช้ระบบรายงานอัจฉริยะ(Qlik Sense)

วัตถุประสงค์

1. เพื่อนำเสนอวิธีการในการรวบรวมข้อมูลแฟชั่นโดยใช้คอมพิวเตอร์วิชั่น
2. เพื่อนำเสนอวิธีการในการสกัดข้อมูลการแต่งกาย เพศ ช่วงอายุ อารมณ์ของอายุลูกค้าโดยใช้เทคนิคการเรียนรู้เชิงลึก
3. เพื่อนำเสนอการนำข้อมูลมาแสดงผลด้วยแผนภาพข้อมูล โดยใช้ระบบรายงานธุรกิจอัจฉริยะ

ขอบเขตของการวิจัย

การศึกษาวิจัยครั้งนี้ แบ่งการศึกษาออกเป็น 2 ส่วนคือ

1. การรวบรวมข้อมูลการแต่งกายของลูกค้าหรือประชาชนทั่วไป โดยใช้การเรียนรู้เชิงลึกในการตรวจหาคน การตรวจหาใบหน้า เพื่อทำนาย เพศ อายุ และ อารมณ์ และ การตรวจหาการแต่งกาย เพื่อทำนาย ท่อนบน ท่อนล่าง และรองเท้า
2. การแสดงผลข้อมูลด้วยแผนภาพข้อมูลโดยใช้ระบบงานธุรกิจอัจฉริยะ

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถนำข้อมูลที่รวบรวมได้และไปวิเคราะห์ให้เกิดประโยชน์สูงสุดกับงานธุรกิจค้าปลีกแฟชั่น
2. สามารถนำวิธีการที่นำเสนอไปใช้เป็นแนวทางในการใช้และสร้างโมเดลประเภทต่างๆ ให้มีประสิทธิภาพมากยิ่งขึ้น

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจหาแนวโน้มแฟชั่นโดยการนำการเรียนรู้เชิงลึกปัจจุบันมีอยู่จำนวนไม่มาก ซึ่งผู้วิจัยได้ศึกษางานวิจัยที่ได้รับการตีพิมพ์และสรุปได้ดังนี้

Fashionpedia: Ontology, Segmentation, and an Attribute Localization Dataset เป็นงานวิจัยที่สำรวจการแบ่งอินสแตนซ์ (Instance Segmentation) ออกเป็นส่วนๆ ด้วยการจำกัดคุณสมบัติ (Attribute Localization) ซึ่งการรวมแต่ละส่วนของอินสแตนซ์ สามารถใช้การตรวจจับและแบ่งแต่ละอินสแตนซ์ของวัตถุ และได้มีการจัดหมวดหมู่ของแต่แอตทริบิวต์ของภาพแบบละเอียด และงานวิจัยที่นำเสนอ Fashionpedia ประกอบด้วยกัน 2 ส่วนคือ 1) การนิยามและขอบเขตซึ่งสร้างโดยผู้เชี่ยวชาญเฉพาะทางด้านแฟชั่น สามารถแบ่งออกเป็นกลุ่มเสื้อผ้าหลัก 27

กลุ่ม ส่วนประกอบเสื้อผ้าจำนวน 19 ส่วน และองค์ประกอบย่อยๆ อีก 294 คุณลักษณะ 2) ชุดข้อมูลภาพแฟชั่นที่ใช้ในชีวิตประจำวัน และภาพแฟชั่นจากผู้มีชื่อเสียงจากสถานที่ต่างๆ มาสร้างคำอธิบายประกอบร่วมกับการมาสก์แบ่งส่วน (Segmentation masks) และองค์ประกอบต่างๆ ที่เกี่ยวข้องกัน

ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations เป็นงานวิจัยที่เรานำเสนอฐานข้อมูลใหม่ที่เรียกว่า “ModaNet” ซึ่งเป็นฐานข้อมูลขนาดใหญ่ โดยคอลเลกชันของภาพพื้นฐานมาจากชุดข้อมูลของ Paper Doll ซึ่งเป็นชุดข้อมูลภาพแบบสตรีทวิวจำนวน 55,176 ภาพ ที่มีคำอธิบายประกอบรูปแบบหลายเหลี่ยม เพิ่มเติมด้วยภาพสตรีทอีกจำนวน 1 ล้านภาพพร้อมคำอธิบายแบบสั้นๆ เป้าหมายหลักของ ModaNet คือการนำเสนอเทคนิคมาตรฐานในการประเมินความก้าวหน้าในงานด้านคอมพิวเตอร์วิชั่น โดยใช้ข้อมูลขนาดใหญ่ในการทำความเข้าใจด้านแฟชั่น ซึ่งชุดข้อมูลที่มีคำอธิบายแบบหลากหลายสามารถช่วยวัดประสิทธิภาพของอัลกอริทึมที่ล้ำสมัย เช่น Object Detection, Semantic Segmentation, Polygon Prediction ด้านสตรีทแฟชั่นอย่างละเอียด

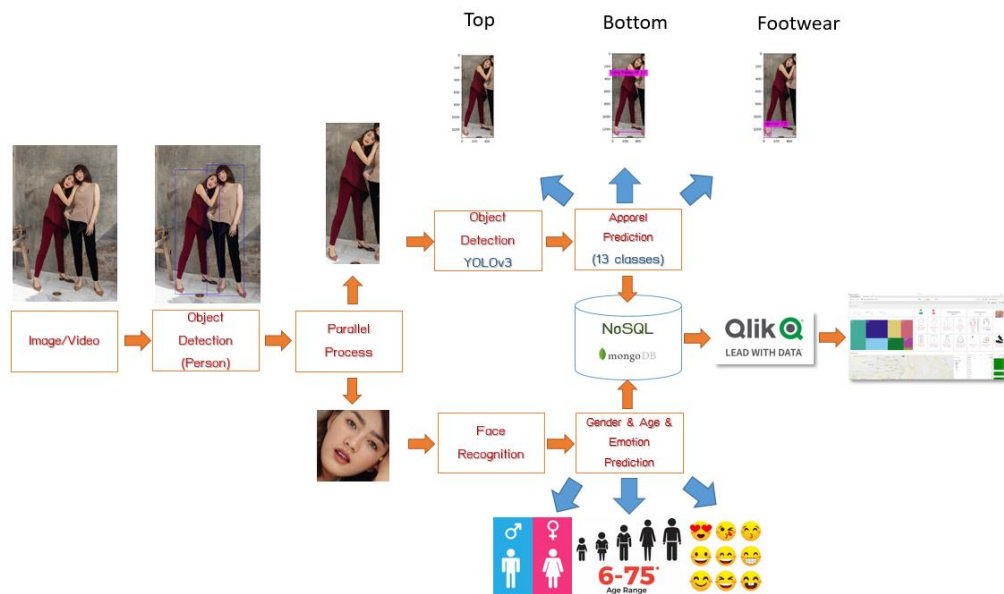
DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images เป็นงานวิจัยที่พัฒนาต่อจาก Deep Fashion โดยมีวัตถุประสงค์ในการแก้ไขปัญหาในด้าน 1) การตรวจจับเสื้อผ้า (Clothes Detection) 2) การคำนวณค่าตำแหน่งเชิง 3 มิติ (Pose Estimation) ของเสื้อผ้า 3) การแบ่งส่วน (Segmentation) ของเสื้อผ้า และ 4) การแก้ไข (Retrieval) ซึ่ง DeepFashion2 มีเสื้อผ้าจำนวน 801,000 รายการ โดยแต่ละรายการมีคำอธิบายประกอบที่หลากหลาย เช่น Style, Scale Viewpoint, Occlusion, Bounding Box, Dense Landmark, Masks นอกจากนี้ยังได้มีชุดข้อมูลเสื้อผ้าแบบ Commercial-Consumer อีก 873,000 คู่ ทำให้คำอธิบายของชุดข้อมูล DeepFashion2 ใหญ่กว่าของ Fashion Global Challenge ถึง 8 เท่า และยังเสนออัลกอริทึมชื่อ Match R-CNN ซึ่งสร้างมาจาก Mask R-CNN เพื่อแก้ปัญหาทั้ง 4 งานที่กล่าวมาข้างต้นในลักษณะแบบครบวงจร โดยใช้การประเมินผลแบบกว้างๆ ด้วยหลักเกณฑ์ที่แตกต่างกันใน DeepFashion2

ระเบียบวิธีวิจัย

งานวิจัยนี้สามารถนำไปประยุกต์ใช้ในงานด้านคอมพิวเตอร์วิชั่น เช่น การจำแนกคนจากวิดีโอ หรือ รูปภาพ การตรวจหาใบหน้า เพื่อใช้ในการทำนายเพศ ช่วงอายุ, อารมณ์ และการจำแนกประเภทการแต่งกาย ท่อนบน ท่อนล่าง รองเท้า โดยมีเป้าหมายเพื่อสอนชุดข้อมูลใหม่และเพิ่มประสิทธิภาพความแม่นยำของโมเดล โดยใช้โครงสร้างเครือข่ายประสาทเทียม (Convolution Neural Network: CNNs) ในการจำแนกประเภทของการแต่งกายเป็นหลัก

แนวทางในการวิจัยนี้ใช้หลักการตรวจหาวัตถุ(Object Detection) จากวิดีโอหรือรูปภาพมาทำการตรวจหาคน(Person) ตรวจหาใบหน้า(Face) การรู้จำใบหน้า(Face Recognition) เพื่อสกัดข้อมูลเพศ (Gender) ช่วงอายุ (Age Ranges) อารมณ์ (Emotion) การตรวจหาการแต่งกาย (Apparel Detection) โดยแบ่งออกเป็น 3 กลุ่มคือ **ท่อนบน (Top Wear)** จำนวน 7 คลาส คือ เสื้อยืด (T-Shirt) เสื้อคอปก (Polo) เสื้อเชิ้ต (Shirt) เสื้อแจ็คเก็ต (Jacket) เสื้อแจ็คเก็ตแบบสวมทับ (Jacket T-Shirt) เสื้อสูท (Suit) กระโปรงยาว (Maxi-Dress) **ท่อนล่าง(Bottom Wear)** จำนวน 4 คลาส คือ กางเกงขายาว (Trousers) กางเกงขาสั้น (Shorts) กระโปรงสั้น (Skirt) กระโปรงทั่วไป (Dress) **รองเท้า(Footwear)** จำนวน 2 คลาส คือ รองเท้าทั่วไป (Normal) รองเท้ากีฬา (Sport)

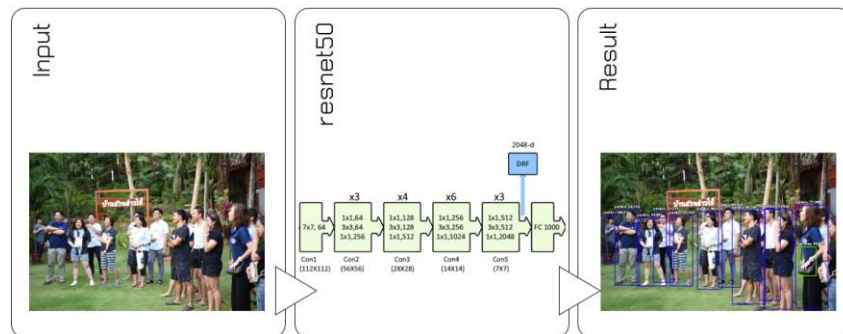
วิธีการดำเนินการเป็นการสกัดข้อมูลจากวัตถุที่ตรวจหาได้ การแสดงกรอบรอบวัตถุ (Bounding Box) และคำอธิบายประกอบ (Annotations) ที่ได้จากโมเดล YOLO v3 จากนั้นจะบันทึกข้อมูลที่สกัดได้ลงในฐานข้อมูลประเภท NoSQL และนำข้อมูลที่ได้มาแสดงเป็นข้อมูลภาพ (Data Visualization) โดยใช้เครื่องมือธุรกิจอัจฉริยะ (Business Intelligent) มาช่วยในการวิเคราะห์ ดังภาพที่ 1



ภาพที่ 1. ภาพรวมระบบ

1. การตรวจหาคน (Person Detection)

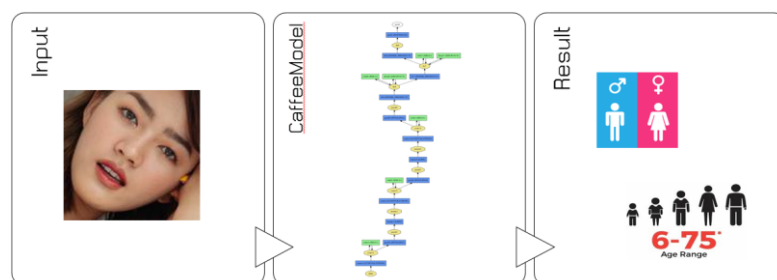
- โมเดลที่นำมาใช้ ImageAI และใช้ Network ที่ชื่อว่า resnet50
- กำหนดค่า minimum_percentage_probability = 50



ภาพที่ 2 ขั้นตอนการตรวจหาคน

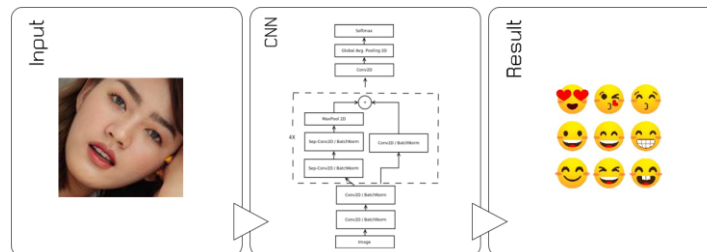
2. การตรวจหาใบหน้า (Face Recognition)

โมเดลที่นำมาใช้ในการทำ Face Detection คือ harrcascade_front_face เป็นการใช้ machine learning ในการเรียนรู้จากภาพเชิงลบและภาพเชิงบวก ในการใช้งานจำเป็นต้องมีไฟล์ที่ได้รับการเรียนและแยกประเภทใบหน้า ก่อนนำไปสกัดข้อมูล ส่วนการสกัดข้อมูลใช้ Caffe เป็น CNN ในการสกัดข้อมูลเฟส โดยการใช้ Network เป็น LeNet-5 ข้อมูลที่ได้จะเป็นเพศชายและหญิง (Male, Female) ในส่วนของการสกัด ช่วงอายุ ใช้ SVM ในการ Classification โดยกำหนดเป็น 8 ช่วง (0-3), (4-7), (8-14), (15-24), (25-37), (38-47), (48-59), (60-100) ดังภาพที่ 3



ภาพที่ 3 ขั้นตอนการสกัดข้อมูลเฟส และช่วงอายุจากใบหน้า

ในส่วนของการสกัด อารมณ์จากใบหน้า ใช้ CNN โดยใช้ชุดข้อมูลมาตรฐาน FER-2013 ข้อมูลที่สกัดได้ กำหนดได้ 7 อารมณ์ (angry, disgust, scared, happy, sad, surprised, neutral)



ภาพที่ 4 ขั้นตอนการสกัดข้อมูลอารมณ์จากใบหน้า

3. การตรวจหาการแต่งกาย (Apparel Detection)

ในการตรวจหาการแต่งกายผู้วิจัย ได้ใช้โอเพนซอร์ส ชื่อว่า DASKNET ซึ่งเป็นแอปพลิเคชันที่พัฒนามาจากการใช้ชุดข้อมูลในการตรวจหาวัตถุชื่อว่า YOLO v3 เป็นโมเดลที่มีความรวดเร็วในการประมวลผลและการแสดงผลแบบเรียลไทม์ รวมทั้งสามารถระบุตำแหน่งของวัตถุที่ตรวจหาได้ในเวลาเดียวกัน ซึ่งผู้วิจัยได้แบ่งการสกัดข้อมูลออกเป็น 3 กลุ่ม ท่อนบน (Top Wear) ท่อนล่าง (Bottom Wear) และ รองเท้า (Footwear) ทั้งหมด 13 คลาส

ตารางที่ 1 ข้อมูลชนิดวัตถุที่สกัดของแต่ละกลุ่ม

กลุ่ม	ชนิดวัตถุที่ตรวจหาได้
ท่อนบน	1) เสื้อยืด 2) เสื้อยืดมีปก 3) เสื้อเชิ้ต 4) เสื้อแจ็กเก็ต 5) เสื้อแจ็กเก็ตแบบสวมทับ 6) เสื้อสูท 7) กระโปรงเต็มตัว
ท่อนล่าง	1) กางเกงขาสั้น 2) กางเกงขาสั้น 3) กระโปรงสั้น 4) กระโปรงทั่วไป
รองเท้า	1) รองเท้าทั่วไป 2) รองเท้ากีฬา

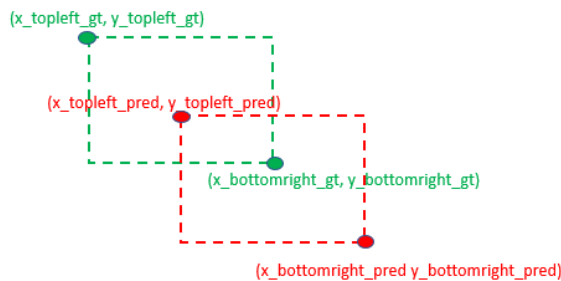
การสร้างแบบจำลองของการแต่งกาย ได้แบ่งการดำเนินการออกเป็น 4 ส่วนคือ 1) การเตรียมข้อมูลในการวิจัย 2) สร้างรูปแบบข้อมูล 3) การนำข้อมูลไปใช้งาน 4) การทดสอบแบบจำลอง

1. การเก็บข้อมูลในการวิจัย (Data Collection)

การเก็บข้อมูลจากผู้วิจัยได้เก็บจากเว็บไซต์ กำหนดขนาดภาพ 480 x 640 พิกเซล มีการแบ่งประเภทของการแต่งกายออกเป็น 13 คลาส กำหนดให้หนึ่งคลาสมีจำนวน 100 ภาพ รวมทั้งสิ้น 1,300 ภาพ หลังจากนั้นทำการแยกข้อมูล 3 กลุ่ม ตามตารางที่ 1 ในการแบ่งชุดข้อมูล ใช้เทคนิค (simple random sampling : SRS) ใช้คอมพิวเตอร์สุ่ม โดยแบ่งเป็น 80:20 เปอร์เซ็น

2. การเตรียมข้อมูลในการวิจัย (Data Preparation)

การเตรียมข้อมูลทางผู้วิจัยได้ใช้โปรแกรม LabelMe ในการทำ Image Annotation ของภาพทั้งหมดให้อยู่ในรูปแบบ ภาพที่ 5

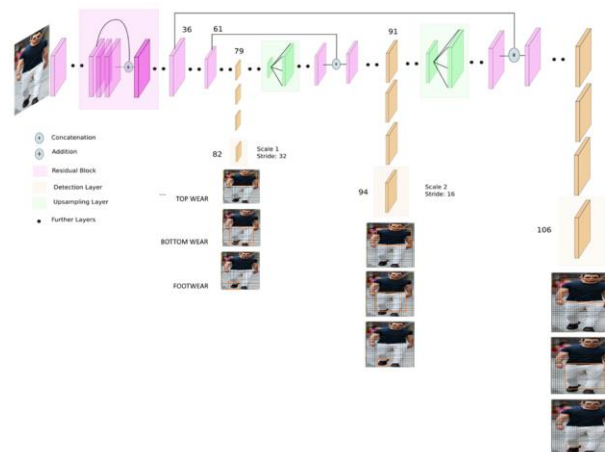


ภาพที่ 5 รูปแบบข้อมูลในการทำ Image Annotation

3. การสร้างโมเดล (Model Training)

ผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็น 3 กลุ่มตามตารางที่ 1. ในการสร้างโมเดลเพิ่มเติมได้ใช้ YOLO v3 และใช้ pretrain โมเดลชื่อว่า darknet53.conv.74.weights เป็น weight เริ่มต้น ดังภาพที่ 6.

- กลุ่มท่อนบน (Top Wear) ทำการสอนโมเดลทั้งหมดจำนวน 8,000 รอบ พบว่า avg. loss อยู่ที่ 0.0552
- กลุ่มท่อนล่าง (Bottom Wear) ทำการสอนโมเดลทั้งหมด 9,000 รอบ พบว่า avg. loss อยู่ที่ 0.0123 และ
- กลุ่มรองเท้า (Footwear) ทำการสอนโมเดลจำนวน 4,000 รอบ พบว่า avg. loss อยู่ที่ 0.1195



ภาพที่ 6 ขั้นตอนการสกัดข้อมูลการแต่งกายที่ได้จากการตรวจหาคน

4. การทดสอบโมเดล (Model Evaluation)

ในการวัดประสิทธิภาพจะเริ่มจากการเลือกชุดของภาพซึ่งประกอบด้วยภาพที่นำมาทดสอบ (Testing Images) และกลุ่มภาพที่ถูกต้อง (Ground Truth Images) ซึ่งหมายถึงกลุ่มของภาพที่คล้ายกันกับภาพที่นำมาทดสอบนั้นไว้ล่วงหน้า ซึ่งการเลือกชุดของภาพแต่ละกลุ่มนี้จัดทำโดยผู้วิจัย โดยใช้ภาพจำนวน 260 ภาพ จากชุดภาพจำนวน 13 คลาสตามตารางที่ 1. ในการวัดประสิทธิภาพใช้การวัดแบบ 1:1 โดยใช้เทคนิค Non-Maximum Suppression: NMS โดยพิจารณาความน่าจะเป็นของแต่ละภาพ ร่วมกับค่า IOU (Intersection Over Union) กับทุกคู่ของคำตอบ เพื่อเลือกผลลัพธ์ที่ดีที่สุดออกมา ในการวัดประสิทธิภาพดังกล่าวได้กำหนดค่า $NMS = 0.6$ และ $IOU = 0.5$



ภาพที่ 7 การแสดงการใช้เทคนิค Non-Maximum Suppression: NMS

T-SHIRT



IoU : 0.8873

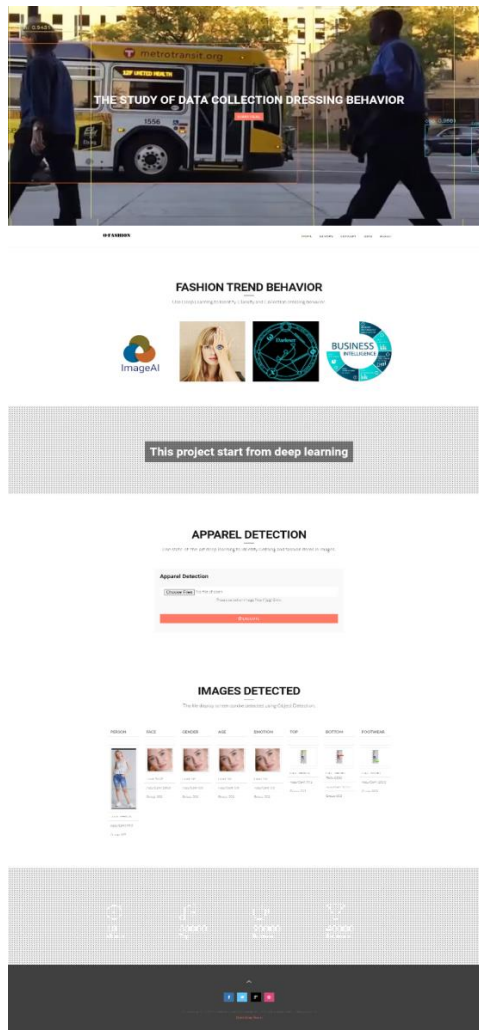


IoU : 0.8672

ภาพที่ 8 ตัวอย่างชุดข้อมูลการตรวจสอบ (Ground Truth)

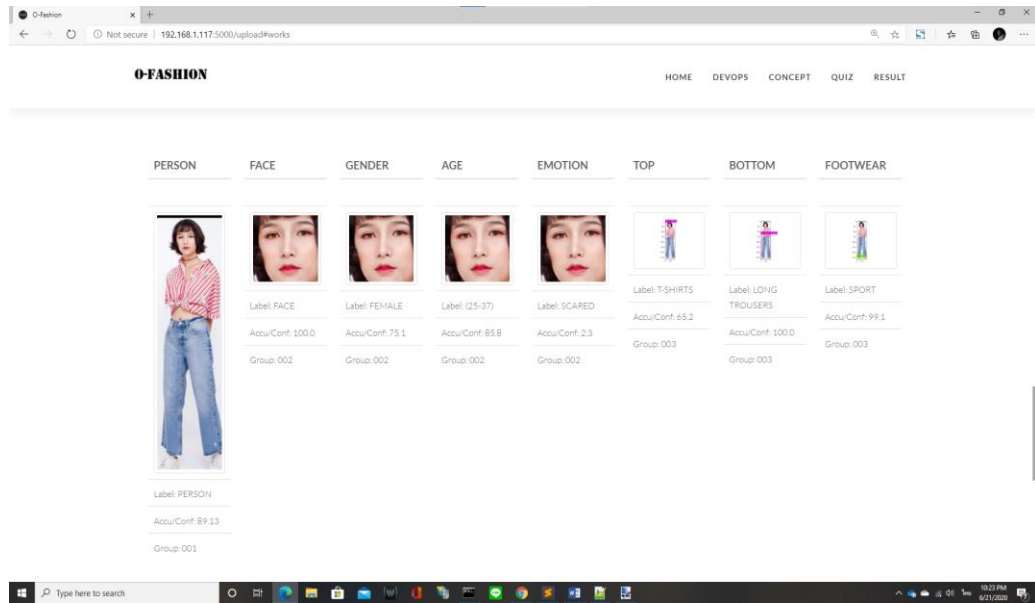
5. การนำมาใช้งาน

ในการนำมาใช้งานทางผู้วิจัยได้จัดสร้างโปรแกรมในรูปแบบของเว็บแอปพลิเคชันชื่อ O-FASHION ทำหน้าที่ในการสกัดข้อมูลจากการแต่งกาย เพศ ช่วงอายุ และอารมณ์จากไฟล์ วีดีโอหรือ ไฟล์ภาพ ซึ่งโปรแกรมที่สร้างขึ้นมีการให้ความรู้เบื้องต้นเกี่ยวกับโมเดล และเครื่องมือต่างๆ ที่นำมาใช้ในงานวิจัยชิ้นนี้ พร้อมทั้งการประมวลผลผลเพื่อแสดงให้เห็นผลลัพธ์ที่ชัดเจน และเข้าใจงานวิจัยชิ้นนี้ทันที ดังภาพที่ 9



ภาพที่ 9 ภาพรวมเว็บแอปพลิเคชัน O-FASHION

การแสดงผลของเว็บแอปพลิเคชัน O-FASHION สามารถแสดงภาพที่ตรวจหาได้เช่น คน หน้าคน และ เสื้อผ้า พร้อมทั้งสกัดข้อมูลการแต่งกาย เพศ ช่วงอายุ และอารมณ์ โดยแอปพลิเคชันสามารถแสดงค่าความเชื่อมั่นของแต่ละส่วนที่โมเดลตรวจหาได้ ดังภาพที่ 10 และตารางที่ 2



ภาพที่ 10 ข้อมูลที่สกัดได้จากแอปพลิเคชัน O-FASHION

ตารางที่ 2 ข้อมูลสรุปที่สกัดได้จากโมเดล

วัตถุที่ตรวจจับได้	Accu/Conf	ผล	ผลการทำนาย
1. คน (PERSON)	89.1	PERSON	TRUE
2. ใบหน้า (FACE)	100.0	FACE	TRUE
3. เพศ (GENDER)	75.1	FEMALE	TRUE
4. อายุ (AGE)	85.8	(25-37)	TRUE
5. อารมณ์ (EMOTION)	2.3	SCARED	FALSE
6. ท่อนบน (TOP WEAR)	65.2	T-SHIRT	TRUE
7. ท่อนล่าง (BOTTOM WEAR)	100.0	LONG TROUSERS	TRUE
8. รองเท้า (FOOTWEAR)	99.1	SPORT	TRUE

การจัดเก็บข้อมูลผู้วิจัยได้ใช้ MongoDB มาเป็นฐานข้อมูล ซึ่งเป็นฐานข้อมูลประเภท NoSQL โดยมีการจัดเก็บ ชื่อภาพ ภาพคน ภาพใบหน้า ข้อมูลเสื้อผ้า เพศ ช่วงอายุ อารมณ์ พิกัดที่จัดเก็บ โดยจัดเก็บในรูปแบบละติจูด (Latitude) และ ลองจิจูด(Longitude) วันที่ เวลา ประเทศ โซน รหัสไปรษณีย์ ซึ่งภาพตรวจหาได้จากโมเดลจะถูกนำไปเข้ารหัส (encoding) ด้วย Standard Format แบบ Base64 ตัวอย่างดังภาพที่ 11

```
{
  "_id" : ObjectId("5ee704ed7bbdc3f44e3eb426"),
  "mcap_id" : "001",
  "tframe_id" : "A18",
  "tframe_x1" : 166,
  "tframe_x2" : 36,
  "tframe_y1" : 489,
  "tframe_y2" : 898,
  "tframe_img_name" : "AA_A18_001_186_56_469_878.jpg",
  "tframe_img_file" : { "$binary" : "Lz1qLzRBQVFTalpKUmdBQkFRQUFBUEFCQUFELzJ3QkRBQW",
    "type" : "Binary" },
  "tframe_lat" : "13.754",
  "tframe_lon" : "100.5014",
  "tcls_accu" : 99.9331,
  "tcls_label" : "person",
  "tcls_model" : "person",
  "upl_date" : ISODate("2020-06-15T12:19:41.352Z"),
  "tframe_postal" : "10100",
  "tframe_city" : "Bangkok",
  "tframe_country" : "TH"
}
```

ภาพที่ 11 ข้อมูลที่จัดเก็บในรูปแบบ NOSQL

การวิเคราะห์ข้อมูล (data analysis & visualization)

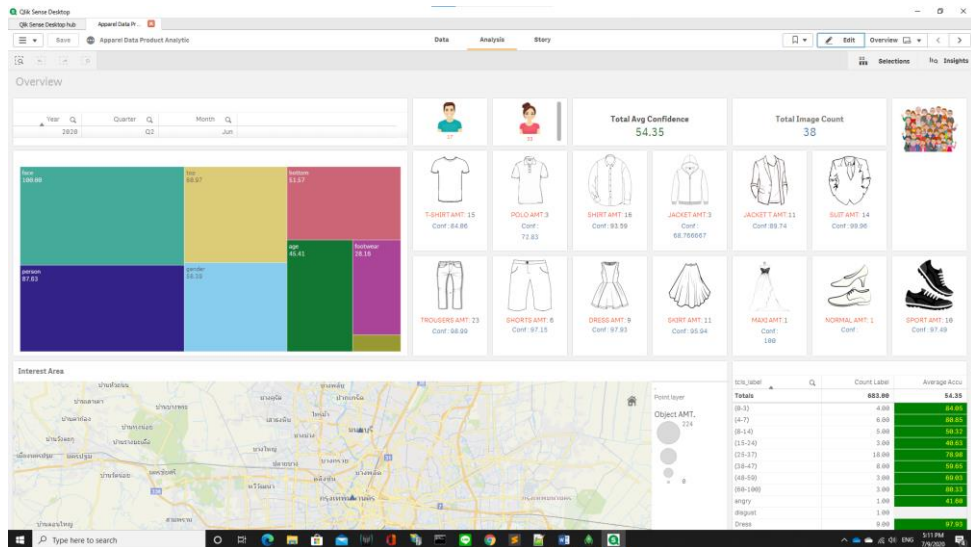
ระบบการวิเคราะห์ข้อมูลจากการทำนายและการแสดงผลเพื่อช่วยในการสนับสนุนการตัดสินใจในการนำข้อมูลไปใช้ จะแสดงให้เห็นขนาดของข้อมูลเพื่อให้ลูกค้าสนใจ และสามารถเจาะจงข้อมูลลงไปถึงระดับต่างๆ ได้ดังนี้

- สามารถระบุช่วงเวลาที่น่าสนใจ (ปี ไตรมาส เดือน)
- สามารถแสดงผลกลุ่มของวัตถุที่เราสนใจ เช่น คน เพศ ช่วงอายุ อารมณ์
- สามารถแสดงผลกลุ่มของเสื้อผ้าที่ตรวจหาได้ เช่น ท่อนบน ท่อนล่าง และรองเท้า
- สามารถแสดงผลวัตถุรายตัวได้ เช่น เสื้อคอกลม เสื้อคอปก เสื้อเชิ้ต เสื้อแจ็กเก็ต

เสื้อแจ็กเก็ตแบบสามทับ เสื้อสูท กางเกงขายาว การเกงขาสั้น กระโปรงทั่วไป กระโปรงสั้น กระโปรงยาว รองเท้าทั่วไป และ รองเท้ากีฬา

- สามารถตรวจสอบพื้นที่บนแผนที่ว่าพบวัตถุที่เก็บอยู่บนฐานข้อมูล

Dashboard ที่จัดทำขึ้นจะแสดงข้อมูล จำนวนภาพ หรือ พื้นที่ที่เก็บข้อมูล จำนวนวัตถุที่ตรวจหาได้ และ ค่าเฉลี่ยความแม่นยำในการทำนายจากโมเดล อีกทั้งยังแสดงให้เห็นการ Classification วัตถุที่เราสนใจ ดังภาพที่ 12

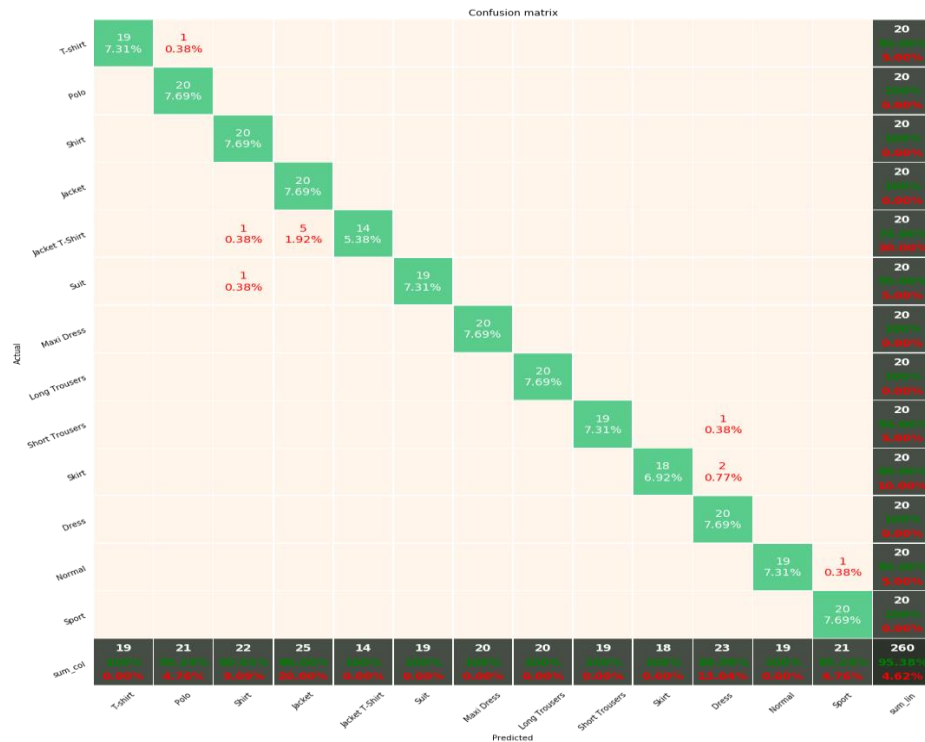


ภาพที่ 12 Dashboard แสดงผลข้อมูลภาพ

ผลการวิจัย

1. การหา Mean Average Precision ด้วยการใช้ Non-Maximum Suppression

ในการประเมินโมเดลสำหรับโมเดลที่เพิ่มเติมเพื่อหาค่าความแม่นยำ ใช้การคำนวณหาค่า Average Precision (AP) หรือ Mean Average Precision (mAP) สำหรับทำนายภาพแบบหลายคลาส (Multiclass) ทางผู้วิจัยได้ใช้อัลกอริทึม (Algorithm) ที่ชื่อว่า Non-Maximum Suppression (NMS) โดยพิจารณาจากค่าความน่าจะเป็นวัตถุนั้น ๆ ร่วมกับ ค่า IOU (Intersection Over Union) กับทุกคู่ของคำตอบ เพื่อเลือกผลลัพธ์ที่ดีที่สุดออกมาจะพบว่าในแต่ละการทดสอบจะมีค่า Intersection over union (IOU) เพียงค่าเดียวเท่านั้นพบว่ามีค่าความถูกต้อง 100% จำนวน 7 คลาส ค่าความถูกต้อง 95% จำนวน 4 คลาส ค่าความถูกต้อง 90% และ 70% อยู่อย่างละ 1 คลาส ดังภาพที่ 13



ภาพที่ 13 ข้อมูลประสิทธิภาพของโมเดลทั้ง 13 คลาส

2. ผลการวัดประสิทธิภาพความถูกต้องของโมเดล

ในการวัดทางผู้วิจัยได้ใช้ชุดข้อมูลภาพจำนวน 260 ภาพ (กำหนด 20 ภาพต่อ 1 ประเภทวัตถุ) ใน 13 ประเภทวัตถุที่แตกต่างกัน ได้แก่ เสื้อยืด (T-Shirt) เสื้อยืดคอปก (Polo) เสื้อเชิ้ต (Shirt) เสื้อแจ็คเก็ต (Jacket) เสื้อแจ็คเก็ตแบบสวมทับ (Jacket T-Shirt) เสื้อสูท (Suit) กระโปรงยาว (Maxi-Dress) กางเกงขายาว (Long Trousers) กางเกงขาสั้น (Short Trousers) กระโปรงสั้น (Skirt) กระโปรงทั่วไป (Dress) รองเท้าทั่วไป (Normal) และ รองเท้ากีฬา (Sport) โดยในแต่ละภาพกำหนดให้คะแนน IOU ต่อตำแหน่งของวัตถุนั้นเมื่อมีค่ามากกว่า หรือเท่ากับ 0.5 จะเป็นตัวอย่างข้อมูลในเชิงบวก (Positive Sample) ในขณะที่ค่าต่ำกว่าให้เป็นข้อมูลในเชิงลบ (Negative Sample) ค่าความแม่นยำ (AP) ของการตรวจหาวัตถุในตารางที่ 3 ผลลัพธ์แสดงให้เห็นถึงประสิทธิภาพในการทำงานของการตรวจหาวัตถุ ว่ามีค่าเฉลี่ยความแม่นยำอยู่ที่ 0.9538 หรือ 95.38% ของวัตถุทุกประเภท

ตารางที่ 3 ข้อมูลแสดงการเปรียบเทียบประสิทธิภาพโมเดล

Accuracy Score : 0.9538461538461539

Report :

	precision	recall	f1-score	support
T-shirt	1.00	0.95	0.97	20
Polo	0.95	1.00	0.98	20
Shirt	0.91	1.00	0.95	20
Jacket	0.80	1.00	0.89	20
Jacket T-Shirt	1.00	0.70	0.82	20
Suit	1.00	0.95	0.97	20
Maxi Dress	1.00	1.00	1.00	20
Long Trousers	1.00	1.00	1.00	20
Short Trousers	1.00	0.95	0.97	20
Skirt	1.00	0.90	0.95	20
Dress	0.87	1.00	0.93	20
Normal	1.00	0.95	0.97	20
Sport	0.95	1.00	0.98	20
accuracy			0.95	260
macro avg	0.96	0.95	0.95	260
weighted avg	0.96	0.95	0.95	260

สรุปและอภิปรายผลการวิจัย

การสอนโมเดลโดยใช้ชุดข้อมูลเพิ่มเติมในการตรวจหาแนวโน้มแฟชั่นโดยใช้ การเรียนรู้เชิงลึก ได้แบ่งชุดข้อมูลออกเป็น 3 ส่วน 1) ท่อนบน 2) ท่อนล่าง 3) รองเท้า โดยใช้ YOLOv3 darknet และ weight มาเป็น pretrain แสดงให้เห็นว่าประสิทธิภาพในการทำงานของการตรวจหาวัตถุ โดยมีค่าเฉลี่ยความแม่นยำอยู่ที่ 0.9538 หรือ 95.38% จากวัตถุทั้งหมดที่เพิ่มเติม

การนำโมเดลที่เกี่ยวข้องและสามารถทำนายผลได้แม่นยำมาใช้ร่วมกัน ซึ่งประกอบด้วย การตรวจหาวัตถุ โดยใช้ pretrain จาก MS COCO มาตรวจหาคน (Person) และ ImageAI มาตรวจหาใบหน้า

การสร้างแอปพลิเคชันมาใช้ในการเก็บข้อมูล และ ผลลัพธ์จากการทำนาย ผู้วิจัยได้ทำการสร้าง Web Application ชื่อ O-FASHION เป็น User Interface (UI) สำหรับใช้ในการอธิบายส่วนงานที่ได้นำเสนอ การเก็บข้อมูลที่ตรวจหาได้ การเก็บผลข้อมูลที่โมเดลทำนายได้ รวมถึงสถานที่ (Latitude, Longitude) วันเวลาในแต่ละช่วง (Times Line) และ สามารถใช้งานได้กับอุปกรณ์หลายประเภท (PC, iPad, Telephone)

การนำเครื่องมือธุรกิจอัจฉริยะ (Business Intelligence : BI) มาใช้ ผู้วิจัยได้นำเครื่องมือ(Tools) ที่ชื่อว่า Qlik Senses มาใช้ในการนำเสนอแผนภาพข้อมูล(Data Visualization) มาช่วยในการเล่าเรื่องด้วยข้อมูล (Data Storytelling) ได้เป็นอย่างดี

ข้อเสนอแนะ

ในการศึกษาด้านแฟชั่นโดยใช้การเรียนรู้เชิงลึก เรื่องของสีของวัตถุที่ตรวจหาได้มีความสำคัญอย่างมาก ผู้วิจัยคิดว่าเราสามารถเพิ่มเรื่องการพิจารณาการกระจายของสีที่อยู่บนวัตถุที่เราตรวจหาได้ โดยใช้การกระจายตัวแบบเกาส์เซียน หรือ การใช้ฮิสโตแกรมสีในแบบจำลองสี HSV (Hue, Saturation, Value) มาใช้ในการเก็บข้อมูลเพิ่มเติม

บรรณานุกรม

Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath

Hariharan, Hartwig Adam, Serge Belongie:“Fashionpedia : Ontology, Segmentation, and an Attribute Localization Dataset, ” arXiv:2004.12276v1 [cs.CV] 26 Apr 2020

Shuai Zheng, M. Hadi Kiapour, Fan Yang Robinson Piramuthu: “ ModaNet: A Large-scale Street Fashion Dataset with Polygon Annotations,” arXiv:1807.01394v4 [cs.CV] 0 Apr 2019

Yuying Ge, Ruimao Zhang, Lingyun Wu, Xiaogang Wang, Xiaoou Tang, and Ping Luo: “DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images,” arXiv:1901.07973v1 [cs.CV] 23 Jan 2019

Pongsate Tangseng, Zhipeng Wu: “Looking at Outfit to Parse Clothing:” arXiv: 1703.01386v1 [cs.CV] 4 Mar 2017

Octavio Arriaga, Paul G. Ploger, Matias Valdenegro: “Real-time Convolutional Neural Networks for Emotion and Gender Classification,” arXiv: 1710.07557, 2017

Ziwei Liu, Ping Luo, Shi Qiu, Xiao gang Wang, Xiaoou Tang: “Deep Fashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations,” (CVPR), 2016, pp. 1096–1104

Kota Yamaguchi, M. Hadi Kiapour, Tamara L. Berg: “Paper Doll Parsing: Retrieving Similar Styles to Parse Clothing Items,” 2013 IEEE International Conference on Computer Vision

Gil Levi and Tal Hassner: “Age and Gender Classification using Convolutional Neural Networks,” IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, 2015