ISSN 3027-6357 (Print)
ISSN 3027-6365 (Online)

วารสารวิทยาศาสตร์และเทคโนโลยี Homepage : *http://li01.tci-thaijo.org/index.php/tstj*
ปีที่ 32 ฉบับที่ 2 (มีนาคม-เมษายน 2567) หน้า [99-114]

การตรวจจับใบหน้าบุคคลจาก 3 สภาวะแวดล้อมด้วย YOLOv4

Detecting Human Faces in Three Different Environmental Conditions using YOLOv4

พิศณุ คูมีชัย*
กองวิชาวิศวกรรมศาสตร์ ฝ่ายศึกษาโรงเรียนนายเรือ
สมุทรปราการ 10270
Pisanu Kumeechai*
Department of Engineering, Education Branch, Royal Thai Naval Academy,
Samut Prakan 10270

Received 5 September 2023; Received in revised 10 December 2023; Accepted 15 January 2024

บทคัดย่อ

งานวิจัยนี้มุ่งเน้นการตรวจจับใบหน้าบุคคลโดยใช้โมเดล YOLOv4 และเปรียบเทียบกับอัลกอริทึมอื่น ๆ ที่ใช้ในงานวิจัยตรวจจับใบหน้า โดยใช้ชุดข้อมูลรูปภาพจำนวนมากจาก 3 สภาวะแวดล้อมที่แตกต่างกันอย่างมีนัยสำคัญ คือภายในอาคารที่มีแสงน้อย, ภายในอาคารที่มีแสงสว่างเพียงพอ, และภายนอกอาคารในช่วงเวลาปกติที่มีแสงสว่าง เพื่อวัดประสิทธิภาพของแต่ละอัลกอริทึมในการตรวจจับใบหน้ากับอีก 4 อัลกอริทึมที่ถูกใช้ในการเปรียบเทียบคือ Viola-Jones Face Detection Algorithm, DeepFace, FaceNet, และ MTCNN ผลการทดลองพบว่า YOLOv4 วัดประสิทธิภาพด้วย confusion matrix มีค่าเฉลี่ยในการตรวจจับภาพใบหน้าจากทั้ง 3 สภาวะแวดล้อมด้วย precision ที่ 0.96, recall ที่ 0.93 และ f1-score ที่ 0.94 ดังนั้น YOLOv4 เป็นอัลกอริทึมที่มีประสิทธิภาพในการตรวจจับใบหน้าบุคคลที่ดีที่สุดในงานวิจัยนี้

คำสำคัญ: การตรวจจับใบหน้าบุคคล; YOLOv4; สภาวะแวดล้อมที่แตกต่างกัน; Confusion matrix

Abstract

This research focuses on detecting human faces using the YOLOv4 model and compares it with other algorithms used in face detection studies. The dataset consists of a large number of images taken in three significantly different environmental conditions: indoors with low light, indoors with sufficient light, and outdoors during normal daylight hours. The performance of each algorithm in detecting human faces is measured and compared with four other algorithms: Viola-Jones Face

***ผู้รับผิดชอบบทความ:** Pisanu41984198@hotmail.com **doi:** 10.14456/tstj.2024.16

Detection Algorithm, DeepFace, FaceNet, and MTCNN. The results of the experiment show that the YOLOv4 model outperforms the other algorithms in all three environmental conditions. According to confusion matrix evolution, the average precision for detecting human faces using the YOLOv4 model is 0.96, with a recall of 0.93 and an f1-score of 0.94. Therefore, the YOLOv4 algorithm is the most effective algorithm for detecting human faces in this study.

Keywords: Detecting human faces; YOLOv4; Different environmental; Confusion matrix

1. บทนำ

การตรวจจับใบหน้าบุคคลจากการประมวลผลภาพดิจิทัลในปัจจุบันมีประโยชน์มากมาย อาทิเช่น การรักษาความปลอดภัย การตรวจจับใบหน้าสามารถใช้เพื่อรักษาความปลอดภัยโดยตรวจจับใบหน้าของบุคคลที่เข้าถึงพื้นที่เฉพาะ หรือตรวจสอบการเข้าถึงระบบความปลอดภัยในองค์กร หรือระบบการเข้าถึงข้อมูลสำคัญอื่น ๆ การตรวจจับใบหน้าสามารถใช้เพื่อควบคุมการเข้าถึงของบุคคลในบริเวณที่สำคัญที่มีการติดตั้งระบบการตรวจจับใบหน้า ซึ่งช่วยให้เพิ่มความปลอดภัยและป้องกันการเข้าถึงที่ไม่ได้รับอนุญาตหรือการโจมตีระบบ การจัดการและควบคุมบุคคล การตรวจจับใบหน้าสามารถใช้เพื่อจัดการและควบคุมการเข้าถึงของบุคคลในสถานที่ต่าง ๆ เช่น ในสนามบิน ห้างสรรพสินค้า หรือโรงงานที่ต้องการตรวจสอบการเข้า-ออกของบุคคล และการตรวจสอบเวลาเข้า-ออก การตรวจจับใบหน้าสามารถใช้เพื่อตรวจสอบเวลาเข้า-ออกของพนักงานหรือบุคคลที่เข้าใช้บริการ ซึ่งช่วยให้บริษัทหรือองค์กรตรวจสอบเวลาการทำงานของพนักงานได้อย่างถูกต้องและสะดวกสบาย แต่ในการตรวจจับภาพใบหน้าบุคคลด้วยการประมวลผลภาพดิจิทัลในปัจจุบันก็ยังมีปัญหาหลายอย่าง ซึ่งสามารถพบเจอได้ดังนี้ ปัญหาความแม่นยำของการตรวจจับภาพใบหน้าบุคคลที่มีการเปลี่ยนแปลงของความเข้มแสง และความเป็นไปได้ที่ภาพอาจจะมีการเลือกใช้แสงสว่างที่ไม่เหมาะสม ปัญหาในการตรวจจับภาพใบหน้าบุคคลที่มีการเปลี่ยนแปลงของมุมมอง ซึ่งอาจจะทำให้การตรวจจับภาพใบหน้าไม่สามารถทำได้ด้วยความแม่นยำสูง เนื่องจากการเปลี่ยนแปลงของมุมมองอาจทำให้ใบหน้าของบุคคลอยู่ในแนวตั้งหรือแนวนอนที่แตกต่างกัน ปัญหาในการตรวจจับภาพใบหน้าบุคคลที่มีการสวมหน้ากากอนามัยจะทำให้มีการปกปิดส่วนของใบหน้าที่สำคัญ เช่น ปากและจมูก ซึ่งอาจทำให้การตรวจจับภาพใบหน้าไม่สามารถทำได้ด้วยความแม่นยำสูงและ ปัญหาในการตรวจจับภาพใบหน้าบุคคลที่มีการเปลี่ยนแปลงของลักษณะทางชีวภาพ เช่น อายุ การขึ้นรูปผม การเปลี่ยนแปลงของน้ำหนัก ซึ่งอาจทำให้การตรวจจับภาพใบหน้าไม่สามารถทำได้ด้วยความแม่นยำสูง ในการตรวจจับภาพใบหน้าบุคคลเป็นหัวข้อที่มีการวิจัยอย่างกว้างขวางในด้านคอมพิวเตอร์วิทยาและการประมวลผลภาพดิจิทัลหลายงานที่น่าสนใจอาทิเช่น งานวิจัยที่เกี่ยวกับการตรวจจับใบหน้าบุคคลในสภาพแวดล้อมที่หลากหลายและท้าทาย เช่น แสงต่ำ หรือหลายใบหน้าที่อยู่ในระยะใกล้ ๆ กัน โดยใช้โมเดล RetinaFace ที่พัฒนาขึ้นเพื่อการตรวจจับใบหน้าโดยที่ไม่ต้องใช้ขั้นตอนของการสกัดคุณสมบัติ (feature extraction) ก่อนจะนำไปตรวจจับในขั้นตอนถัดไป ซึ่งปัญหาที่เกิดคือปัญหาเรื่องความแม่นยำในการตรวจจับภาพใบหน้าบุคคลโดยใช้โมเดลเรียนรู้เชิงลึก เช่น YOLOv3 และ RetinaFace [1] ปัญหาในการตรวจจับภาพใบหน้าบุคคลที่มีการสวมหน้ากากอนามัย ซึ่งงานวิจัยที่ใช้โมเดลแบบเดียวกับ YOLOv3 ซึ่งใช้ร่วมกับ Feature Pyramid Network (FPN) ในการทำให้โมเดลสามารถตรวจจับวัตถุขนาดเล็กได้ดีขึ้น และใช้ Darknet-53 เป็นโครงข่ายหลักในการเรียนรู้ ในการตรวจจับใบหน้าที่สวมหน้ากาก โดยแนบป้ายกำกับให้กับชุดข้อมูลเพื่อช่วยให้โมเดลสามารถตรวจจับใบหน้าที่สวมหน้ากากได้ และทดสอบการทำงานของโมเดลด้วยชุดข้อมูลที่เก็บภาพใบหน้าที่สวมหน้ากาก และชุดข้อมูลที่เก็บภาพใบหน้าที่ไม่สวมหน้ากาก ผลการทดสอบแสดงให้เห็นว่าโมเดลที่นำเสนอในงานวิจัยนี้สามารถตรวจจับใบหน้าที่สวมหน้ากากได้ดีกว่าโมเดลที่ไม่ได้นำป้ายกำกับข้อมูลสวมหน้ากากในการเรียนรู้ เพื่อใช้ตรวจจับภาพใบหน้าบุคคลที่มีการสวมหน้ากากอนามัย [2]

จากปัญหาในการตรวจจับภาพใบหน้าที่กล่าวไปข้างต้นก็มีผู้ทำการวิจัยเพื่อพัฒนาและเพิ่มประสิทธิภาพในหลายด้าน อาทิเช่น การตรวจจับ แยกและรู้จำใบหน้าด้วยการประมวลผลภาพดิจิทัลโดยเริ่มต้นด้วยการตรวจจับใบหน้าจากภาพ แล้วนำไปแยกและสกัดคุณสมบัติของใบหน้า เช่น สีผิวหน้า รูปร่างหน้า และความหนาแน่นของข้อมูล จากนั้นใช้เทคนิคการเข้ารหัสใบหน้าเพื่อสร้างรูปแบบหน้าตาและนำไปใช้ในการรู้จำใบหน้า รวมถึง

แนวโน้มของการพัฒนาระบบตรวจจับ แยกและรู้จำใบหน้าในอนาคต โดยเน้นไปที่การใช้งานและประโยชน์ของแต่ละเทคนิคในการแก้ไขปัญหาที่เกิดขึ้นในปัจจุบันและในอนาคตของการตรวจจับ แยกและรู้จำใบหน้า [3] มีการใช้ Deep Learning-based Face Detection เช่น Convolutional Neural Network (CNN) และ Deep Residual Network (ResNet) เพื่อให้สามารถตรวจจับและระบุใบหน้าได้อย่างแม่นยำ มีการระบุตำแหน่งของแต่ละส่วนของใบหน้า มีการทำซ้ำการตรวจจับใบหน้าและการแบ่งกลุ่มของใบหน้า [4] ต่อมามีการพัฒนา Deep Learning-based Face Detection และ Recognition โดยใช้โครงข่ายประสาทเทียม (Neural Network) และโมเดล Deep Learning ต่าง ๆ เช่น CNN, RNN, LSTM, และ GAN เพื่อตรวจจับและรู้จำใบหน้า [5] มีการรวบรวมและวิเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจจับและรู้จำใบหน้าที่แสดงถึงเทคนิคต่าง ๆ ที่ใช้ในการตรวจจับและรู้จำใบหน้า เช่น Haar cascade classifier, Local binary patterns (LBP), Scale-invariant feature transform (SIFT), Speeded-up robust features (SURF), และ Histogram of Oriented Gradients (HOG) และอื่น ๆ [6] การวิเคราะห์งานวิจัยที่เกี่ยวข้องกับการรู้จำใบหน้า ที่ได้แสดงถึงเทคนิคต่าง ๆ ที่ใช้ในการรู้จำใบหน้า โดยใช้โครงข่ายประสาทเทียม (Neural Network) และโมเดล Deep Learning ต่าง ๆ เช่น CNN, RNN, LSTM, และ GAN เพื่อนำมาประยุกต์ใช้ในงานการรู้จำใบหน้า [7] นอกจากนั้นก็ยังมีเทคนิคต่าง ๆ ที่ใช้ในการรู้จำใบหน้า เช่น Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Kernel PCA, และ Local Binary Pattern (LBP) และอื่น ๆ ในการรู้จำใบหน้าในระบบความปลอดภัย การรู้จำใบหน้าในการจดจำและระบุตัวตน [8]

ในงานวิจัยเก่ามีงานวิจัยหลายตัวที่ได้ศึกษาข้อมูลเพิ่มเติมเกี่ยวกับการพัฒนาเทคโนโลยีการตรวจจับใบหน้าบุคคล อาทิเช่นงานวิจัย Zhou, Y. และคณะ [9] ที่ได้เน้นไปที่การตรวจจับคุณลักษณะเด่นของรูปหน้า (Facial Landmark Detection) โดยกล่าวถึงการพัฒนาวิธีการตรวจจับจุดเด่นบนใบหน้าโดยใช้ Deep Learning และการใช้ข้อมูลรูปภาพจำนวนมาก (Large Scale Image Data) เพื่อเพิ่มประสิทธิภาพในการประมวลผลการเรียนรู้ Taigman, Y. และคณะ [10] ได้เสนอวิธีการ DeepFace ซึ่งเป็นระบบการตรวจจับใบหน้าโดยใช้ Deep Learning ทำให้ระบบตรวจจับใบหน้าของเครื่องจักรมีความแม่นยำใกล้เคียงกับมนุษย์ ซึ่งทำให้ระบบเหล่านี้สามารถใช้งานในสถานการณ์ต่าง ๆ ที่ต้องใช้ความแม่นยำสูง เช่น การยืนยันตัวตนด้วยใบหน้าเพื่อเข้าใช้งานระบบหรือสถานที่สำคัญต่าง ๆ ซึ่งมีความแม่นยำในการตรวจจับที่สูงถึง 97.35% โดยใช้ชุดข้อมูลจำนวนมากถึง 4 ล้านรูปภาพ Zhang, Z. และคณะ [11] ได้เสนอการใช้ Deep Multi-Task Learning (DMTL) ในการตรวจจับจุดเด่นบนใบหน้า (Facial Landmark Detection) โดยใช้ Convolutional Neural Network (CNN) และ Multi-Task Learning (MTL) ในการฝึกสอนโมเดล ผลการทดลองแสดงให้เห็นว่า DMTL สามารถให้ความแม่นยำในการตรวจจับจุดเด่นบนใบหน้าได้สูงกว่าวิธีการทั่วไปโดยมีค่าความคลาดเคลื่อนเฉลี่ยเพียง 3.77 พิกเซล Xie, S. และคณะ [12] ได้นำเสนอ Aggregated Residual Transformations (ResNeXt) ซึ่งเป็นโมเดล CNN ที่ใช้ในการฝึกสอนและพัฒนาระบบการตรวจจับใบหน้า โมเดลนี้สร้างจากการรวมเครื่องมือของ Residual Networks (ResNets) และ Inception Networks เข้าด้วยกัน เพื่อสร้างโมเดลที่มีประสิทธิภาพและความแม่นยำในการตรวจจับวัตถุ ผลการทดลองแสดงให้เห็นว่าโมเดล ResNeXt มีประสิทธิภาพดีกว่าโมเดลอื่น ๆ โดย ResNeXt มีความแม่นยำในการตรวจจับของใบหน้าถึง 96.4% ในภาพที่มีความซับซ้อนสูง Liu, W. และคณะ [13] ได้เสนอ Single Shot Multibox Detector (SSD) ซึ่งเป็นแบบจำลองการตรวจจับวัตถุที่มีประสิทธิภาพและความเร็วสูง โดยทำงานได้ในเวลาเดียวกันทั้งในการตรวจ

จับและการจำแนกหมวดหมู่ของวัตถุ โมเดล SSD ใช้เทคโนโลยี Convolutional Neural Network (CNN) เพื่อสกัดลักษณะของวัตถุและใช้ Multibox layer เพื่อทำนายตำแหน่งและหมวดหมู่ของวัตถุ ผลการทดลองมีอัตราการตรวจจับถูกต้องที่สูงถึง 74.3% Wu, X. และคณะ [14] เสนอชุดฝึกโมเดล Light CNN ซึ่งเป็นโมเดล CNN ที่มีความเร็วและประสิทธิภาพสูงในการสร้างการแทนที่ใบหน้าที่ดีคือสามารถระบุตัวตนของบุคคลได้อย่างถูกต้อง และควรมีความเสถียรไม่เปลี่ยนแปลงตามสภาพแวดล้อม โดยใช้ข้อมูลที่มีป้ายกำกับไม่ถูกต้องหรือผิดพลาด (noisy labels) ในการฝึกสอน โมเดล Light CNN มีความเร็วสูงกว่าโมเดล CNN อื่น ๆ มีความแม่นยำสูงถึง 98.06% ในชุดข้อมูล LFW (Labeled Faces in the Wild) ที่มีความซับซ้อนสูง Dang, H. และChan, C. S. [15] ได้เสนอการใช้เทคโนโลยี Deep Learning และ Multi-Task Learning (MTL) ในการพัฒนาระบบการตรวจจับใบหน้าที่มีประสิทธิภาพสูงในเงื่อนไขการตรวจจับที่ไม่จำกัด ใช้ Convolutional Neural Network (CNN) ในการฝึกสอนระบบการตรวจจับใบหน้าและการวิเคราะห์การถดถอยเชิงเส้น (linear regression) ในการตรวจจับพิกัดของจุดบนใบหน้า ผลการทดลองแสดงให้เห็นว่าการใช้ Learning (MTL) สามารถเพิ่มประสิทธิภาพของการตรวจจับใบหน้าได้โดยมีค่าความคลาดเคลื่อนเฉลี่ยเพียง 4.2% Li, Y. และคณะ [16] เสนอโมเดล Attention-based Deep Convolutional Neural Network (ADCNN) สำหรับการตรวจจับและกำหนดตำแหน่งของจุด landmark บนใบหน้า โดยใช้การจัดการ attention mechanism เพื่อช่วยในการคัดกรองคุณลักษณะที่สำคัญของภาพ ผลการทดลองแสดงให้เห็นว่าโมเดล ADCNN มีประสิทธิภาพในการตรวจจับและกำหนดตำแหน่งของจุด landmark บนใบหน้า มีความแม่นยำสูงถึง 98.21% Bochkovskiy, A และคณะ [17] ได้เสนอ YOLOv4 เป็นโมเดลการตรวจจับวัตถุที่มีความเร็วและประสิทธิภาพสูง โดยใช้ชุดข้อมูล COCO ในการฝึกสอน โมเดล YOLOv4 มีความแม่นยำสูงถึง 43.5% ในการตรวจจับวัตถุทั่วไป และ 65.7% ในการตรวจจับวัตถุขนาดเล็ก โดยสามารถตรวจจับวัตถุจากกล้องที่มีได้ถึง 65 FPS เมื่อประมวลผลบน GPU Wang, X. และคณะ [18] เสนอการใช้ YOLOv4 ในการตรวจจับใบหน้าของมนุษย์ โดยใช้เทคนิคการปรับแต่งขนาดของ anchor boxes เพื่อช่วยในการตรวจจับใบหน้าที่มีขนาดแตกต่างกันได้ดีขึ้น ผลการทดลองแสดงให้เห็นว่าโมเดลที่ใช้เทคนิคนี้มีประสิทธิภาพในการตรวจจับใบหน้าของมนุษย์ โดยใช้ชุดข้อมูล WIDER FACE และ FDDB ในการทดสอบ Kumar, M. และคณะ [19] เสนอการเปรียบเทียบประสิทธิภาพของ YOLOv4 และ RetinaNet ในการตรวจจับใบหน้า ผลการทดลองแสดงให้เห็นว่าโมเดล YOLOv4 มีความแม่นยำสูงกว่า RetinaNet ในการตรวจจับใบหน้า โดยมีค่า F1-score ที่สูงกว่า 98% ในการตรวจจับใบหน้าทั้งในภาพนิ่งและวิดีโอ โดยโมเดล YOLOv4 มีความเร็วในการทำงานที่สูงกว่า RetinaNet ด้วย โดยมีความเร็วในการตรวจจับวัตถุได้ถึง 38 FPS โดยทั่วไป ผลการทดสอบแสดงว่า YOLOv4 เป็นโมเดลที่มีประสิทธิภาพและความเร็วในการตรวจจับใบหน้าที่ดีกว่า RetinaNet

วัตถุประสงค์ของงานวิจัยคือตรวจจับใบหน้าบุคคลด้วยอัลกอริทึม YOLOv4 อันเป็นอัลกอริทึมหลักในงานวิจัยนี้แล้ว เปรียบเทียบกับอีก 4 อัลกอริทึมอันประกอบด้วย Viola-Jones, YOLOv4, DeepFace, FaceNet และMTCNN เพื่อหาประสิทธิภาพที่ดีที่สุดในการทำงาน อัลกอริทึมหลักที่นำมาทดสอบคือ YOLOv4 (You Only Look Once version 4) เป็นอัลกอริทึมที่ใช้เทคโนโลยี Deep Learning และ Object Detection Algorithm ในการตรวจจับวัตถุต่าง ๆ ได้อย่างรวดเร็วและมีความแม่นยำสูง และสามารถตรวจจับใบหน้าได้ด้วยความแม่นยำสูงในระยะเวลาอันสั้น อัลกอริทึมนี้เหมาะสำหรับการใช้งานในระบบตรวจจับและติดตามใบหน้าในเวลาจริง [18] ตามที่แสดงดัง (Figure 1)

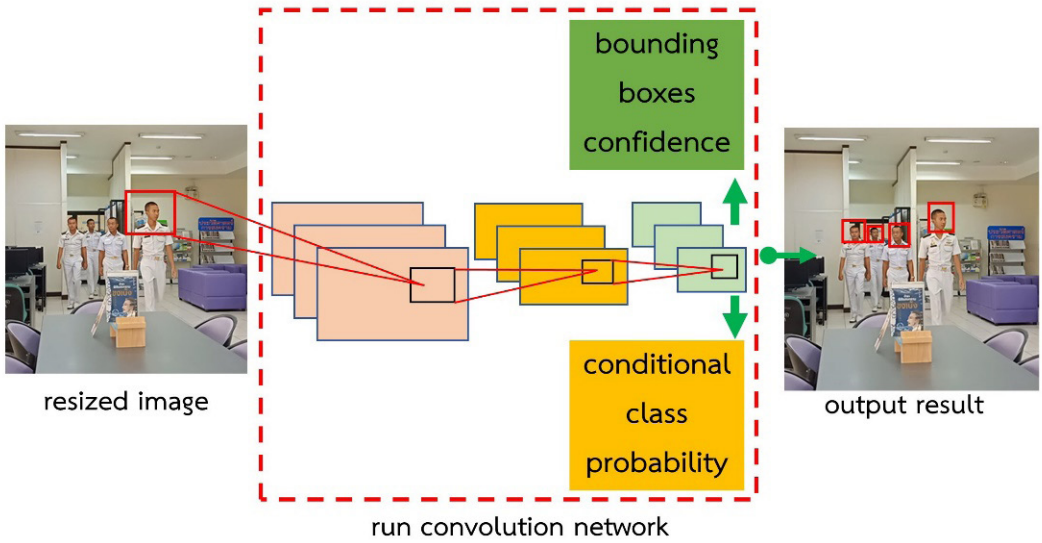

Figure 1 The working principle of YOLO

2. ขอบเขตงานวิจัย

วัตถุประสงค์ของงานวิจัยคือตรวจจับใบหน้าบุคคลด้วยอัลกอริทึม YOLOv4 อันเป็นอัลกอริทึมหลักในงานวิจัยนี้ เปรียบเทียบกับอีก 4 อัลกอริทึมอันประกอบด้วย 1) Viola-Jones Face Detection Algorithm, 2) DeepFace, 3) FaceNet, และ 4) MTCNN เพื่อหาประสิทธิภาพที่ดีที่สุดในการทำงานอุปกรณ์ที่ใช้กล้องดิจิทัลขนาดไฟล์ภาพ 640 คูณ 840 อัตราเฟรม 30 เฟรมต่อวินาที ชุดข้อมูลรูปภาพทำการตรวจจับภาพใบหน้าจาก 3 สภาวะแวดล้อมประกอบไปด้วย 1) ภายในอาคารที่มีแสงน้อย 2) ภายในอาคารที่มีแสงสว่างเพียงพอ และ 3) นอกอาคารในช่วงเวลาปกติที่มีแสงสว่าง จากภาพรวม 1,000 ภาพ แบ่งเป็นตัวอย่างการฝึก 800 ภาพ และตัวอย่างทดสอบ 100 ภาพ

3. ทฤษฎีที่นำเสนอ

3.1 ทฤษฎีของ YOLOv4

YOLOv4 [20] เป็นอัลกอริทึมสำหรับการตรวจจับวัตถุที่พัฒนาโดยอาศัยการใช้งาน Deep Learning โดยเฉพาะโมเดล CNN (Convolutional Neural Networks) โดยโมเดลของ YOLOv4 นั้นถูกพัฒนาขึ้นโดยอาศัยสถาปัตยกรรมของ darknet เพื่อเพิ่มประสิทธิภาพในการคำนวณ และช่วยให้การตรวจจับวัตถุเกิดขึ้นได้อย่างรวดเร็วและแม่นยำ

อินเทอร์เฟซของ YOLO มีเลเยอร์แบบบิด 24 ชั้น ตามด้วยเลเยอร์ที่เชื่อมต่อกันทั้งหมด 2 ชั้น มันก็แค่ใช้ 1 × 1 เลเยอร์การลดตามด้วยเลเยอร์คอนโวลูชั่น 3 × 3 ฝึกโครงข่ายประสาทเทียมด้วย 9 เลเยอร์แบบหมุนวนแทนที่จะเป็น 24 และตัวกรองน้อยกว่าในเลเยอร์เหล่านั้น ทิ้งปริมาตรของเครือข่ายทั้งหมดไว้ พารามิเตอร์การฝึกอบรมและการทดสอบระหว่าง YOLO และ Fast YOLO จะเหมือนกัน YOLO ได้รับการปรับให้เหมาะสมสำหรับข้อผิดพลาดผลรวมกำลังสองภายในเอาต์พุตของแบบจำลองของเรา การทำนายความเชื่อมั่นสำหรับกล่องที่ไม่มีวัตถุ YOLO ใช้พารามิเตอร์สองตัวคือ ค่า λcoord และ λnoobj ถึงบรรลุเป้าหมายนี้ YOLO ตั้งค่า λcoord = 5 และ λnoobj = .5

3.2 วิธีการตรวจจับใบหน้าของ DeepFace

Convolutional Neural Networks (CNN) [21] เป็นโมเดล Deep Learning ที่ถูกพัฒนาขึ้นเพื่อใช้

ในการประมวลผลภาพ โดยใช้ตัวกรอง (filter) ที่มีขนาดเล็ก ๆ ผ่านภาพเพื่อสกัดลักษณะที่สำคัญของภาพ เป็นโมเดลประมวลผลข้อมูลที่ใช้ในการวิเคราะห์ภาพ โดยมีความสามารถในการเรียนรู้ลักษณะพื้นฐานของภาพอย่างที่ไม่จำเป็นต้องระบุลักษณะภาพไว้ล่วงหน้า โมเดล CNN จะประกอบด้วยชั้นของการคำนวณหลักที่สำคัญที่เรียกว่า Convolutional Layer

Convolutional Layer จะนำ Filter ที่มีขนาดเล็ก ๆ มาคอนโวลูชันกับภาพเพื่อดึงข้อมูลลักษณะของภาพ โดย Filter จะเป็นเมทริกซ์ขนาดเล็ก ซึ่งจะถูกนำไปคูณกับพื้นที่ของภาพที่เรียกว่า receptive field เพื่อสกัดลักษณะพื้นฐานของภาพ เมื่อ Filter ผ่านทุกส่วนของพื้นที่ในภาพแล้ว จะได้ Feature Map ที่เป็นรูปแบบของข้อมูลที่ถูกสกัดมาจากภาพ สำหรับสมการของ Convolutional Layer สามารถเขียนได้ตามสมการที่ (1)

$$y(i,j) = \sum_m \sum_n x(i+m, j+n) w(m,n) \quad (1)$$

โดยที่

$y(i,j)$ คือ ผลลัพธ์จากการคำนวณของ Convolutional Layer ที่ตำแหน่ง (i,j)

$x(i+m, j+n)$ คือ ภาพ ณ พิกัด $i+m, j+n$ นำมาคำนวณด้วยกระบวนการคอนโวลูชัน

$w(m,n)$ คือ ค่าของเมทริกซ์ Filter ที่ตำแหน่ง m,n

3.3 วิธีการตรวจจับใบหน้าของ Viola-Jones Paul viola และ Michael J. Jones

Viola-Jones Paul viola และ Michael J. Jones [22] เทคนิคการตรวจจับใบหน้าด้วยวิธีนี้ ได้นำเสนอเทคนิคการตรวจจับใบหน้าที่มีความเร็วและมีความถูกต้องในการตรวจจับสูงในปี 2001 เทคนิคการตรวจจับใบหน้าของ Viola-Jones เป็นเทคนิคที่ได้รับการยอมรับและรู้จักในงานวิจัยเรื่องการตรวจจับใบหน้าเป็นอย่างมาก โดยเทคนิคการตรวจจับ ใบหน้า Viola-Jones นี้สามารถแบ่งออกเป็น 3 ขั้นตอน การคำนวณรูปแบบการจำลองด้วย Integral Image การค้นหารูปแบบการจำลองด้วย Ada-boost และการรวมตัวจำแนกกลุ่มแบบต่อเรียง (Cascaded Classifier) โดยแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้ หลักการพื้นฐานของเทคนิค การตรวจจับใบหน้าของ Viola-Jones คือ การนำภาพที่ต้องการตรวจหาใบหน้ามาแบ่งเป็นภาพย่อย (Sub-window) จากนั้นภาพย่อยดังกล่าวจะถูกพิจารณาเป็นภาพอินพุตของกระบวนการตรวจหาใบหน้าเทคนิคทั่วไป ในการตรวจหาใบหน้าจะทำการปรับขนาดของภาพอินพุตแตกต่างกันหลาย ๆ ขนาด และใช้ตัวตรวจจับ (Detector) ที่มีขนาดคงที่ค้นหาวัตถุ ข้อเสียของการตรวจจับใบหน้าแบบนี้ คือ ระยะเวลาในการคำนวณไม่คงที่ ดังนั้น Viola-Jones จึงเสนอเทคนิคการตรวจจับใบหน้าใหม่โดยใช้การจำลองรูปแบบ Haar-like เป็นตัวตรวจจับทำการปรับขนาดของตัวตรวจจับแทนการปรับขนาดภาพอินพุต และใช้ตัวตรวจจับทำการตรวจจับใบหน้าหลาย ๆ รอบ โดยแต่ละรอบใช้ขนาดของ ตัวตรวจจับแตกต่างกัน ซึ่งเมื่อทำการเปรียบเทียบกับวิธีเดิมพบว่า เวลาที่ใช้ในการคำนวณไม่ได้ แตกต่างกันมาก แต่ใช้เวลาในการคำนวณการตรวจจับภาพใบหน้าแต่ละรอบมีค่าคงที่ แม้ขนาดของตัวตรวจจับจะแตกต่างกันก็ตาม เทคนิคที่ Viola-Jones นำเสนอนี้จะทำการคำนวณการจำลอง รูปแบบ Haar-like จาก “การตรวจจับใบหน้าด้วยวิธีการพื้นฐานของการจำลองรูปแบบ Haar-like

3.4 วิธีการตรวจจับใบหน้าของ MTCNN

MTCNN [23-24] เป็นเฟรมเวิร์กงานหลายงานแบบเรียงซ้อนเชิงลึก ซึ่งใช้ประโยชน์จากความสัมพันธ์โดยธรรมชาติระหว่างการตรวจจับและการจัดตำแหน่งเพื่อเพิ่มประสิทธิภาพ กรอบของ MTCNN ใช้ประโยชน์จากสถาปัตยกรรมแบบเรียงซ้อนด้วยสามขั้นตอนของเครือข่าย Convolution แบบลึกที่ออกแบบอย่างระมัดระวัง เพื่อทำนายใบหน้าและตำแหน่งสถานที่สำคัญในลักษณะหยาบถึงละเอียด

ภาพรวมของ MTCNN แสดงใน (Figure 2) ให้รูปภาพในตอนแรกเราจะปรับขนาดให้เป็นขนาดต่าง ๆ เพื่อสร้างปิรามิดรูปภาพซึ่งเป็นอินพุตของเฟรมเวิร์กแบบเรียงซ้อนสามขั้นตอนต่อไปนี้

ขั้นที่ 1 ของอัลกอริทึม MTCNN เรียกว่า (P-Net) ซึ่งจะใช้ (CNN) เพื่อตรวจจับใบหน้าในภาพโดยประมาณ โดย (P-Net) จะสร้างกรอบที่อาจเป็นไปได้ที่จะเป็นรูปหน้า ซึ่งจะมีลักษณะดังนี้ มีขนาดและรูปร่างที่แตกต่างกัน อาจซ้อนทับกัน ขั้นตอนนี้ใช้เวกเตอร์การถดถอยกรอบขอบเขตเพื่อปรับกรอบที่อาจเป็นไปได้ที่จะเป็นรูปหน้าให้แม่นยำยิ่งขึ้น เวกเตอร์การถดถอยนี้ประกอบด้วยข้อมูลเกี่ยวกับตำแหน่งและขนาดของกรอบ หลังจากปรับกรอบผู้สมัครให้แม่นยำยิ่งขึ้นแล้ว ขั้นตอนนี้จะใช้ (NMS) เพื่อกำจัดกรอบที่ซ้อนทับกันออก

ขั้นที่ 2 จะตรวจสอบกรอบที่อาจเป็นไปได้ที่จะเป็นรูปหน้าทั้งหมดอีกครั้งโดยใช้ CNN ที่เรียกว่า Refine Network (R-Net) โดยจะตรวจสอบว่ากรอบเหล่านี้เป็นรูปหน้าจริงหรือไม่ หากใช่ R-Net จะปรับกรอบให้แม่นยำยิ่งขึ้น นอกจากนี้ R-Net จะใช้ NMS เพื่อกำจัดกรอบที่ซ้อนทับกันออก

ขั้นที่ 3 เรียกว่า O-Net ซึ่งจะใช้ CNN เพื่อตรวจจับใบหน้าในภาพได้อย่างแม่นยำยิ่งขึ้น O-Net จะทำงานคล้ายกับ R-Net แต่จะมีความแม่นยำมากกว่า R-Net เล็กน้อย O-Net จะพิจารณาปัจจัยเพิ่มเติม เช่น ตำแหน่งของจุดสังเกตใบหน้า จุดสังเกตใบหน้าคือจุดสำคัญบนใบหน้า เช่น มุมตา มุมปาก และจมูก จุดสังเกตใบหน้าเหล่านี้สามารถช่วยระบุใบหน้าได้อย่างแม่นยำยิ่งขึ้น O-Net จะส่งออกตำแหน่งของจุดสังเกตใบหน้าห้า

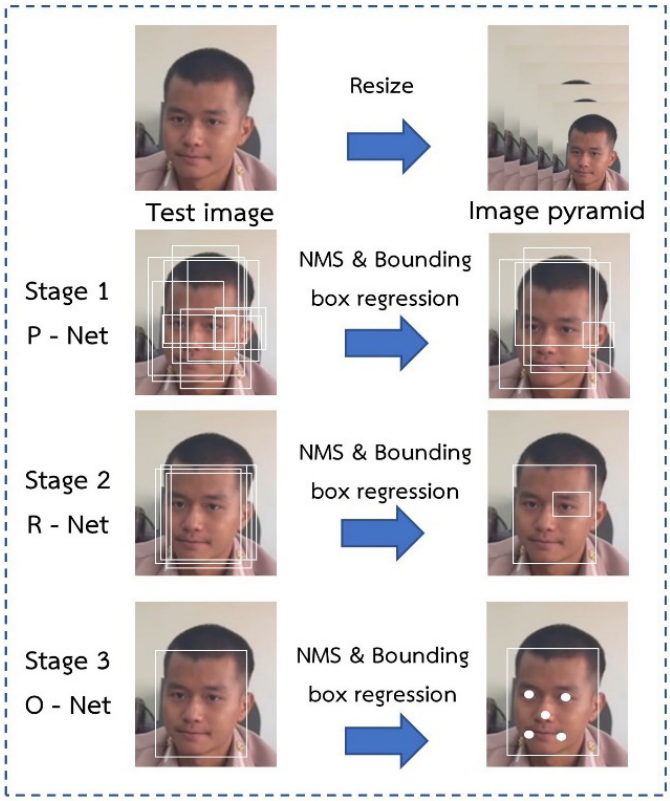

Figure 2 The working principle of YOLO Overview of the cascading framework MTCNN, which includes a deep convolutional multi-step, three-step model.

ตำแหน่ง ตำแหน่งเหล่านี้จะใช้ในการปรับกรอบใบหน้าให้แม่นยำยิ่งขึ้น

3.5 วิธีการตรวจจับใบหน้าของ Facenet

Facenet สะท้อนลักษณะใบหน้าของภาพเป็นหลายมิติ และคำนวณปริภูมิแบบยุคลิด ระยะห่างในการตัดสินใบหน้ามนุษย์ [25] เนื่องจากภาพของบุคคลคนเดียวกันจะมียุคลิดที่สั้นกว่า ระยะห่างจากอวกาศ Facenet ใช้สิ่งนี้เพื่อทำการจดจำใบหน้า สุดท้าย Facenet จะให้มิติ 128 เวกเตอร์ลักษณะเฉพาะของใบหน้ามนุษย์ ทั้ง MTCNN และ Facenet ใช้หลักการของโครงข่ายประสาทเทียม เมื่อเปรียบเทียบกับอัลกอริธึม CNN เดียว MTCNN จะทำให้อัลกอริธึมซับซ้อนและทำให้การประมวลผล ของภาพได้แม่นยำยิ่งขึ้น นอกจากนี้โดยการแบ่งกระบวนการจดจำใบหน้าออกเป็น 2 ขั้นตอนคือ การทำนายอัลกอริธึมระบบใหม่นี้ทำงานได้ดีกว่าระบบอื่น

3.6 การวัดประสิทธิภาพด้วย confusion matrix

Confusion matrix [26] เป็นเครื่องมือสำหรับวัดประสิทธิภาพของระบบการตรวจจับภาพใบหน้าในงานวิจัยนี้ โดยจะแบ่งผลการตรวจออกเป็น 4 ช่วง ได้แก่ 1) True Positive (TP), False Positive (FP), False Negative (FN) และ True Negative (TN)

Confusion matrix เป็นเครื่องมือสำหรับวัดประสิทธิภาพของระบบการจำแนก โดยเฉพาะการจำแนกแบบ binary classification ซึ่งแบ่งผลการจำแนกออกเป็น 4 ส่วนตามที่นำเสนอดัง (Table 1)

เมื่อคำนวณได้ค่า TP, FP, FN, และ TN แล้ว สามารถใช้ค่าเหล่านี้ในการคำนวณค่าประสิทธิภาพของระบบ โดยมีสูตรสำคัญดังนี้

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1\ score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

โดยที่

Precision คือ ความแม่นยำในการพยากรณ์ Positive

Recall คือ ความแม่นยำในการตรวจพบ Positive

F1 score คือ เป็นค่าเฉลี่ยความแม่นยำของ Precision และ Recall เป็นวิธีการรวมความสามารถของความแม่นยำในการพยากรณ์

4. ผลการวิจัย

ภาพในงานวิจัยนี้ถ่ายด้วยกล้องดิจิทัลที่มีความละเอียดภาพ 640 x 840 พิกเซล อัตราเฟรม 30 เฟรมต่อวินาที ชุดข้อมูลรูปภาพภาพรวม 1,000 ภาพเพื่อทำการตรวจจับภาพใบหน้าใน 3 สภาวะที่ประกอบไปด้วย 1) ภายในอาคารที่มีแสงน้อย 2) ภายในอาคารที่มีแสงสว่างเพียงพอ และ 3) นอกอาคารในช่วงเวลาปกติที่มีแสงสว่าง แบ่งเป็นตัวอย่างภาพในการฝึก 800 ภาพ และภาพตัวอย่างในการทดสอบ 100 ภาพ ในงานวิจัยนี้

Table 1 เครื่องมือประเมินผลลัพธ์ของการทำนาย (confusion matrix)

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

วัดประสิทธิภาพด้วย confusion matrix การตรวจจับใบหน้าบุคคลด้วยอัลกอริทึม YOLOv4 เปรียบเทียบกับอีก 4 อัลกอริทึม 1) Viola-Jones Face Detection Algorithm 2) DeepFace 3) FaceNet และ 4) MTCNN จาก 3 สภาวะแวดล้อมประกอบไปด้วย 1) ภายในอาคารที่มีแสงน้อย 2) ภายในอาคารที่มีแสงสว่างเพียงพอ และ 3) นอกอาคารในช่วงเวลาปกติที่มีแสงสว่าง

จากการตรวจจับใบหน้าบุคคลด้วยอัลกอริทึม YOLOv4 ทำการเปรียบเทียบกับอีก 4 อัลกอริทึมอันประกอบด้วย 1) Viola-Jones Face Detection Algorithm 2) DeepFace, 3) FaceNet, และ 4) MTCNN เพื่อหาประสิทธิภาพที่ดีที่สุดในการตรวจจับใบหน้า จาก 3 สภาวะแวดล้อมประกอบไปด้วย 1) ภายในอาคารที่มีแสงน้อย โดยมีค่าความเข้มแสงน้อยกว่า 100 ลักซ์ 2) ภายในอาคารที่มีแสงสว่างเพียงพอ โดยมีค่าความเข้มแสงประมาณ 300 ลักซ์ และ 3) นอกอาคารในช่วงเวลาปกติที่มีแสงสว่าง โดยมีค่าความเข้มแสงประมาณ 500 ลักซ์ สามารถสรุปผลการวิจัยได้ตาม (Table 2-4) ตามลำดับดังนี้

Table 2 การประเมินประสิทธิภาพภายในสภาพแวดล้อมภายในอาคารที่มีแสงน้อย (Evaluating performance within a low-light indoor environment)

วิธีการ	ขนาดภาพ	Precision	Recall	F1-score
Viola-Jones	416 × 416 (fixed)	0.71	0.65	0.60
Viola-Jones	608 × 608 (fixed)	0.83	0.72	0.77
DeepFace	416 × 416 (fixed)	0.89	0.77	0.74
DeepFace	608 × 608 (fixed)	0.92	0.81	0.86
FaceNet	416 × 416 (fixed)	0.62	0.66	0.78
FaceNet	608 × 608 (fixed)	0.88	0.77	0.82
MTCNN	416 × 416 (fixed)	0.79	0.60	0.73
MTCNN	608 × 608 (fixed)	0.89	0.78	0.83
YOLOv4	416 × 416 (fixed)	0.90	0.82	0.84
YOLOv4	608 × 608 (fixed)	0.94	0.89	0.91

Table 3 การประเมินประสิทธิภาพภายในสภาพแวดล้อมภายในอาคารที่มีแสงสว่างเพียงพอ (Evaluating performance within a well-lit indoor environment)

วิธีการ	ขนาดภาพ	Precision	Recall	F1-score
Viola-Jones	416 × 416 (fixed)	0.85	0.61	0.69
Viola-Jones	608 × 608 (fixed)	0.91	0.82	0.86
DeepFace	416 × 416 (fixed)	0.89	0.80	0.82
DeepFace	608 × 608 (fixed)	0.95	0.88	0.91
FaceNet	416 × 416 (fixed)	0.88	0.71	0.76
FaceNet	608 × 608 (fixed)	0.93	0.86	0.89
MTCNN	416 × 416 (fixed)	0.83	0.76	0.77
MTCNN	608 × 608 (fixed)	0.92	0.87	0.89
YOLOv4	416 × 416 (fixed)	0.91	0.89	0.90
YOLOv4	608 × 608 (fixed)	0.96	0.95	0.95

Table 4 การประเมินประสิทธิภาพของแสงสว่างภายนอกอาคารในช่วงเวลากลางวันปกติ (Evaluating the efficiency of outdoor lighting during normal daylight hours)

วิธีการ	ขนาดภาพ	Precision	Recall	F1-score
Viola-Jones	416 × 416 (fixed)	0.87	0.74	0.69
Viola-Jones	608 × 608 (fixed)	0.94	0.81	0.87
DeepFace	416 × 416 (fixed)	0.90	0.82	0.86
DeepFace	608 × 608 (fixed)	0.97	0.87	0.92
FaceNet	416 × 416 (fixed)	0.90	0.81	0.87
FaceNet	608 × 608 (fixed)	0.95	0.84	0.89
MTCNN	416 × 416 (fixed)	0.86	0.71	0.82
MTCNN	608 × 608 (fixed)	0.96	0.86	0.91
YOLOv4	416 × 416 (fixed)	0.92	0.88	0.91
YOLOv4	608 × 608 (fixed)	0.98	0.94	0.96

จากการประเมินประสิทธิภาพของแต่ละอัลกอริทึมในการตรวจจับใบหน้าบุคคลภายในอาคารที่มีแสงน้อยได้ โดย YOLOv4 ได้ค่า precision, recall และ F1-score ที่ดีกว่าอัลกอริทึมอื่น ๆ ที่เทียบเท่ากันทั้งหมดในส่วนนี้ โดย YOLOv4 มีค่า precision ที่ 0.94 และ recall ที่ 0.89 และมีค่า f1-score ที่ 0.91 โดยเฉลี่ย ซึ่งสูงกว่าแต่ละอัลกอริทึมอื่น ๆ ที่เทียบเท่ากันทั้งหมดในส่วนนี้ ดังนั้น YOLOv4 จึงเป็นอัลกอริทึมที่เหมาะสมที่สุดสำหรับการตรวจจับใบหน้าบุคคลภายในอาคารที่มีแสงน้อยในการประเมินประสิทธิภาพ

จากการประเมินประสิทธิภาพของแต่ละอัลกอริทึมในการตรวจจับใบหน้าบุคคลภายในอาคารที่มีแสงสว่างเพียงพอได้โดย YOLOv4 ได้ค่า precision, recall และ f1-score ที่ดีกว่าอัลกอริทึมอื่น ๆ ที่เปรียบเทียบเท่ากันทั้งหมดในส่วนนี้ โดย YOLOv4 มีค่า precision ที่ 0.96 และ recall ที่ 0.95 และมีค่า f1-score ที่ 0.95 โดยเฉลี่ย ซึ่งสูงกว่าแต่ละอัลกอริทึมอื่น ๆ ที่เปรียบเทียบเท่ากันทั้งหมดในส่วนนี้ ดังนั้น YOLOv4 จึงเป็นอัลกอริทึมที่เหมาะสมที่สุดสำหรับการตรวจจับใบหน้าบุคคลภายในอาคารที่มีแสงสว่างเพียงพอในการประเมินประสิทธิภาพ

จากการประเมินประสิทธิภาพของแต่ละอัลกอริทึมในการตรวจจับใบหน้าบุคคลนอกอาคารในช่วงเวลาปกติที่มีแสงสว่างได้ โดย YOLOv4 ได้ค่า precision, recall และ f1-score ที่ดีกว่าอัลกอริทึมอื่น ๆ ที่เปรียบเทียบเท่ากันทั้งหมดในส่วนนี้ โดย YOLOv4 มีค่า precision ที่ 0.98 และ recall ที่ 0.94 และมีค่า f1-score ที่ 0.96 โดยเฉลี่ย ซึ่งสูงกว่าแต่ละอัลกอริทึมอื่น ๆ ที่เทียบเท่ากันทั้งหมดในส่วนนี้ ดังนั้น YOLOv4 จึงเป็นอัลกอริทึมที่เหมาะสมที่สุดสำหรับการตรวจจับใบหน้าบุคคลนอกอาคารในช่วงเวลาปกติที่มีแสงสว่างในการประเมินประสิทธิภาพ

สามารถแสดงตัวอย่างการตรวจจับภาพใบหน้าทั้ง 3 สภาวะแวดล้อมใน (Figure 3-5) ตามลำดับ

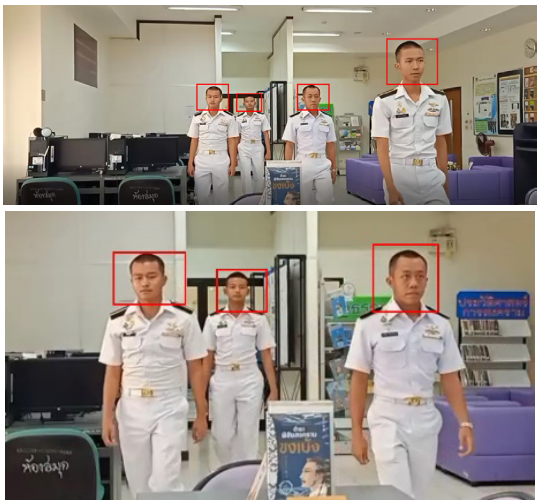

Figure 3 Detecting faces inside a well-lit building

Figure 4 Detecting faces inside a dimly lit building

Figure 5 Detecting individual faces outside the building during regular hours

5. สรุป

สาเหตุที่ YOLOv4 มีประสิทธิภาพสูงสุดอยู่ในทุกสภาวะแวดล้อมเนื่องจากมีความสามารถในการตรวจจับวัตถุได้อย่างแม่นยำและรวดเร็วมากกว่าอัลกอริทึมอื่น ๆ ที่ถูกเปรียบเทียบด้วย โดยเฉพาะในการตรวจจับใบหน้าบุคคลซึ่งเป็นงานที่ซับซ้อนและต้องการความแม่นยำสูง Viola-Jones Face Detection Algorithm และ MTCNN เป็นอัลกอริทึมที่ใช้วิธีการตรวจจับแบบ Cascade Classifier ซึ่งมีประสิทธิภาพในการตรวจจับใบหน้าไม่ได้สูงเท่ากับ YOLOv4 และการตรวจจับอาจจะยากขึ้นในสภาวะที่มีแสงน้อยหรือภายในอาคารที่มีแสงสว่างเพียงพอ DeepFace และ FaceNet เป็นอัลกอริทึมที่ใช้ Deep Learning ในการสกัดลักษณะหน้าใบบุคคลเพื่อใช้ในการตรวจจับ แต่อย่างไรก็ตามการสกัดลักษณะหน้าใบบุคคลอาจไม่เพียงพอในบางกรณีและอาจไม่สามารถรับมือกับ

ถึงแม้ว่าเทคโนโลยีการตรวจจับใบหน้าด้วย YOLOv4 เป็นเทคโนโลยีที่มีประสิทธิภาพสูง แต่ก็ยังมีข้อเสียบางอย่าง เช่น การตรวจจับบางครั้งอาจเกิดข้อผิดพลาดได้ เพราะตัวอัลกอริทึม YOLOv4 อาจไม่สามารถตรวจจับใบหน้าบุคคลได้อย่างแม่นยำเสมอได้ โดยเฉพาะในสภาวะที่มีแสงน้อยหรือภายในอาคารที่มีแสงสว่างเพียงพอไม่ได้ ซึ่งอาจจะส่งผลให้มีข้อผิดพลาดในการตรวจจับหรือไม่ตรวจจับใบหน้าเลย YOLOv4 ต้องการทรัพยากรคอมพิวเตอร์และโปรแกรมที่มีความซับซ้อนสูง ซึ่งอาจทำให้ต้องใช้งบประมาณสูงกว่าการใช้งานโปรแกรมตรวจจับใบหน้าอื่น ๆ และอาจไม่สามารถทำงานได้บนระบบที่มีข้อจำกัด เช่น ระบบฮาร์ดแวร์ที่มีประสิทธิภาพต่ำหรือมีพื้นที่จำกัดเป็นต้น

งานวิจัยนี้แสดงให้เห็นว่าโมเดล YOLOv4 มีประสิทธิภาพในการตรวจจับใบหน้าบุคคลที่ดีที่สุดในทั้ง 3 สภาวะแวดล้อมที่แตกต่างกัน ซึ่งแสดงให้เห็นว่า YOLOv4 เป็นอัลกอริทึมที่มีประสิทธิภาพและมีความยืดหยุ่นสูง ซึ่งช่วยให้นักวิจัยสามารถเข้าใจถึงข้อดีและข้อเสียของอัลกอริทึมต่าง ๆ ได้ดียิ่งขึ้น และยังสามารถนำไปประยุกต์ใช้งานได้ในหลากหลายสภาพแวดล้อม

6. ข้อเสนอแนะ

การใช้งาน YOLOv4 ในการตรวจจับใบหน้าเป็นเทคโนโลยีที่มีความสามารถสูงและได้รับความนิยมอย่างกว้างขวาง ดังนั้น ข้อเสนอแนะในการใช้งาน YOLOv4 ในการตรวจจับภาพใบหน้า และแนวทางในการพัฒนาสามารถสรุปได้ดังนี้

1) การตั้งค่าพารามิเตอร์ การตั้งค่าพารามิเตอร์ของ YOLOv4 เช่น ขนาดภาพ, จำนวนรอบการฝึกฝน, และค่าความแม่นยำต่าง ๆ จะส่งผลต่อประสิทธิภาพของการตรวจจับ ดังนั้น การตั้งค่าพารามิเตอร์ให้เหมาะสมจะช่วยเพิ่มประสิทธิภาพให้กับการตรวจจับภาพใบหน้า

2) การทดสอบและปรับปรุง การทดสอบและปรับปรุง YOLOv4 เพื่อให้มีประสิทธิภาพการตรวจจับใบหน้าที่ดีขึ้น โดยการใช้ชุดข้อมูลทดสอบที่หลากหลายและการปรับแต่งพารามิเตอร์ต่าง ๆ จะช่วยให้ได้ผลลัพธ์ที่ดีขึ้น

3) การพัฒนาเทคโนโลยีเพิ่มเติม การพัฒนาเทคโนโลยีเพิ่มเติมที่สามารถช่วยเพิ่มประสิทธิภาพในการตรวจจับใบหน้า โดยเช่นการใช้ Deep Learning และ Computer Vision เพื่อสกัดลักษณะของใบหน้าอย่างละเอียดและมีประสิทธิภาพสูงในการตรวจจับ

4) การปรับแต่งและตรวจสอบข้อมูล การตรวจสอบความถูกต้องของข้อมูลก่อนนำเข้าในโมเดล เช่น การตรวจสอบชื่อและตำแหน่งของไฟล์ภาพ การตรวจสอบความสมดุลของข้อมูล และการตรวจสอบการเข้ารหัสของข้อมูล เป็นสิ่งสำคัญที่ช่วยให้มีประสิทธิภาพการตรวจจับที่ดี การเพิ่มข้อมูลเพิ่มเติมเพื่อเพิ่มความหลากหลายและความถูกต้องของข้อมูล เช่น การเพิ่มภาพจากมุมมองต่าง ๆ และแบบจำลองใหม่ เป็นสิ่งที่ช่วยเพิ่มประสิทธิภาพการตรวจจับได้

7. กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยดี ขอขอบคุณกองวิชาวิศวกรรมศาสตร์ ฝ่ายศึกษาโรงเรียนนายเรือ กองทัพเรือที่ช่วยเอื้อเฟื้อทั้งอุปกรณ์และสถานที่ในการวิจัยจนงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

8. References

[1] Deng, D., Guo, J., & Yuxiang, Z. (2020). RetinaFace: Single-stage Dense Face Localisation in the Wild. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, 5203-5212.

[2] Qin, Z., Zhang, Y., Huang, K., & Zhao, D. (2020). Towards Mask Face Detection with A Convolutional Neural Network. Applied Sciences, 10(17), 5945.

[3] Arumugam, P., & Natarajan, N. (2017). A Survey of Face Detection, Extraction and Recognition. International Journal of Computer Applications, 163(7), 1-6.

[4] Ranjan, R., Sankaranarayanan, S., & Castillo, C. D. (2020). Deep Learning-based Face Detection: A Comprehensive Survey. ACM Computing Surveys (CSUR), 53(6), 1-36.

[5] Le, N. N. H., & Bui, T. D. (2021). Face Detection and Recognition using Deep Learning: A Survey. Journal of Ambient Intelligence and Humanized Computing, 12(2), 1735-1757.

[6] Sarfraz, M. S., & Alahakoon, R. (2018). Face Detection and Recognition Techniques: A Survey. Journal of King Saud University-Computer and Information Sciences, 30(4), 428-444.

[7] Ahmed, S. M., Abd El-Latif, S. M., Aly, S. S., & ElAlfy, A. S. (2020). A Comprehensive Survey of Face Recognition Techniques. IEEE Access, 8, 158557-158576.

[8] Goyal, P., Bhatia, K., & Singh, R. (2017). Face Recognition Techniques: A Comprehensive Survey. International Journal of Computer Applications, 160(1), 18-25.

[9] Zhou, Y., Liao, S., & Li, S. Z. (2018). Recent progress in facial landmark detection. arXiv preprint arXiv:1806.01893.

[10] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

[11] Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2015). Facial landmark detection by deep multi-task learning. In European

conference on computer vision (pp. 94-108).

[12] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2016). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37).

[14] Wu, X., He, R., Sun, Z., & Tan, T. (2018). A light CNN for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security, 13(11), 2884-2896.

[15] Dang, H., & Chan, C. S. (2019). Unconstrained face detection using deep learning and multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

[16] Li, Y., Li, F., Li, S., Li, X., & Chen, J. (2020). A novel attention-based deep convolutional neural network for facial landmark detection. Cognitive Computation, 12(3), 449-459.

[17] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

[18] Wang, X., Liu, D., & Yu, X. (2021). Adaptive object detection for human faces using YOLOv4. Multimedia Tools and Applications, 80(16), 24807-24822.

[19] Kumar, M., Singh, M., Verma, A. K., & Chandra, S. (2021). Performance evaluation of YOLOv4 and RetinaNet for face detection. Journal of Ambient Intelligence and Humanized Computing, 12(10), 11503-11512.

[20] 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).

[21] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[22] Maghraby M.AbdallaO.Enany,Hybrid Face Detection System using Combination of Viola - Jones Method and Skin Detection, International Journal of Computer Applications (0975-8887) Volume 71-No.6, May 2013

[23] Li Zhihua, Zhang Jianyu & Wei Zhongcheng. (2022). Design on face recognition system based on MTCNN and Facenet. Modern Electronic Technology (04), 139-143.

[24] Zhang Ying. (2022). Study on improved face recognition algorithm based on Facenet. Electronic Technology (03), 25-27.

[25] Schroff, F., Kalenichenko, D. & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[26] Wang, J., & Wang, L. (2013). Face Recognition: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 43(3), 401-422.