วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 34 ฉบับที่ 2 เม.ย.–มิ.ย. 2567
The Journal of KMUTNB., Vol. 34, No. 2, Apr.–Jun. 2024

บทความวิจัย

การวิเคราะห์อารมณ์จากใบหน้ามนุษย์แบบเวลาจริงโดยใช้เทคนิคการเรียนรู้เชิงลึกบนอุปกรณ์สมองกลฝังตัวต้นทุนต่ำ

ชูพันธุ์ รัตนโภคา
ภาควิชาเทคโนโลยีวิศวกรรมอิเล็กทรอนิกส์ วิทยาลัยเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

นราธิป แสงซ้าย*
ภาควิชาเทคโนโลยีวิศวกรรมเครื่องกล วิทยาลัยเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 2555 2000 ต่อ 6435 อีเมล: naratip.s@cit.kmutnb.ac.th DOI: 10.14416/j.kmutnb.2022.11.001
รับเมื่อ 2 กุมภาพันธ์ 2565 แก้ไขเมื่อ 24 มีนาคม 2565 ตอบรับเมื่อ 29 มีนาคม 2565 เผยแพร่ออนไลน์ 1 พฤศจิกายน 2565
© 2022 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

บทคัดย่อ

อารมณ์เป็นการแสดงออกทางพื้นฐานอย่างหนึ่งของมนุษย์ ซึ่งมีการแสดงออกในหลากหลายรูปแบบ เช่น จากเสียง ใบหน้า หรือท่าทาง ในการพัฒนาระบบที่มีการปฏิสัมพันธ์กับมนุษย์นั้น การรับรู้ถึงอารมณ์ที่มนุษย์ตอบสนองกับระบบกลับมานั้น จึงเป็นเรื่องที่สำคัญเป็นอย่างยิ่ง บทความนี้นำเสนอการออกแบบและพัฒนาแบบจำลองการเรียนรู้เชิงลึก YOLOv4-tiny และ YOLOv5s เพื่อวิเคราะห์อารมณ์จากใบหน้ามนุษย์ โดยแบบจำลองทำงานบนอุปกรณ์สมองกลฝังตัวต้นทุนต่ำ Jetson Nano ที่ติดตั้งกล้องในตัว จากนั้นภาพเคลื่อนไหวที่รับเข้ามาจากกล้องจะถูกตรวจจับใบหน้าแบบเวลาจริง เพื่อตีกรอบรอบใบหน้าและแสดงผลการวิเคราะห์อารมณ์ของใบหน้านั้น ซึ่งแบบจำลองสามารถจำแนกอารมณ์ทั้งหมดออกเป็น 7 อารมณ์ ได้แก่ โกรธ รังเกียจ กลัว ดีใจ เสียใจ ตกใจ และเฉยๆ โดยใช้ชุดข้อมูลภาพ RAF-DB ในการฝึกฝนและทดสอบแบบจำลอง ผลจากการประเมินประสิทธิภาพด้านความถูกต้องของแบบจำลอง พบว่า โดยภาพรวมแบบจำลอง YOLOv5s ให้ผลการทำงานที่ดีกว่า YOLOv4-tiny โดยค่า F1 Score ของแบบจำลอง YOLOv5s ได้คะแนน 0.806 ในขณะที่แบบจำลอง YOLOv4-tiny ได้คะแนน 0.774 สำหรับประสิทธิภาพด้านความเร็วในการประมวลผลพบว่า แบบจำลอง YOLOv5 สามารถแสดงวิดีโอได้ประมาณ 11 FPS ในขณะที่ YOLOv4-tiny สามารถแสดงผลวิดีโอได้ประมาณ 10.5 FPS

คำสำคัญ: การเรียนรู้เชิงลึก การตรวจจับวัตถุ การจำแนกอารมณ์ โยโล เจ็ทสันนาโน

การอ้างอิงบทความ: ชูพันธุ์ รัตนโภคา และ นราธิป แสงซ้าย, "การวิเคราะห์อารมณ์จากใบหน้ามนุษย์แบบเวลาจริงโดยใช้เทคนิคการเรียนรู้เชิงลึกบนอุปกรณ์สมองกลฝังตัวต้นทุนต่ำ," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 34, ฉบับที่ 2, หน้า 1–12, เลขที่บทความ 242-195788, เม.ย.–มิ.ย. 2567.

Research Article

A Real-Time Analysis of Human Face Emotions Using Deep Learning Techniques on Low-Cost Embedded Devices

Choopan Rattanapoka
Department of Electronics Engineering Technology, College of Industrial Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Naratip Sangsai*
Department of Mechanical Engineering Technology, College of Industrial Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

* Corresponding Author, Tel. 0 2555 2000 Ext. 6435, E–mail: naratip.s@cit.kmutnb.ac.th DOI: 10.14416/j.kmutnb.2022.11.001
Received 2 February 2022; Revised 24 March 2022; Accepted 29 March 2022; Published online: 1 November 2022
© 2022 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

Abstract

Emotions are one of the most fundamental human expressions, which can be expressed in a variety of ways, such as by voice, face, or gestures. In the development of systems that interact with humans, it is extremely important to recognize how human emotions react to the system. This article presents the design and development of the YOLOv4-tiny and YOLOv5s deep learning models to analyze emotions from human faces. The model runs on a low-cost embedded device, the Jetson Nano, equipped with a built-in camera. The motion picture received from the camera is identified by detecting and framing faces, then displayed the emotional analysis of the faces in real-time. The model can categorize 7 emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. The RAF-DB image dataset was used to train and test the models. Upon the evaluation, we found that the YOLOv5s model performed better than the YOLOv4-tiny model in terms of accuracy, with an F1 score of 0.806 compared to 0.774 for the YOLOv4-tiny model. In terms of processing speed, the YOLOv5 model can display video at roughly 11 FPS, while the YOLOv4-tiny can display video at approximately 10.5 FPS.

Keywords: Deep Learning, Object Detection, Emotion Classification, YOLO, Jetson Nano

Please cite this article as: C. Rattanapoka and N. Sangsai, "A real-time analysis of human face emotions using deep learning techniques on low-cost embedded devices," *The Journal of KMUTNB*, vol. 34, no. 2, pp. 1–12, ID. 242-195788, Apr.–Jun. 2024 (in Thai).

1. บทนำ

อารมณ์เป็นการแสดงออกทางพื้นฐานอย่างหนึ่งของมนุษย์ อารมณ์ของมนุษย์นั้นมีการแสดงออกในหลากหลายรูปแบบ เช่น จากเสียง ใบหน้า หรือท่าทาง การพัฒนาระบบที่มีการปฏิสัมพันธ์กับมนุษย์ การรับรู้ถึงอารมณ์ที่มนุษย์ตอบสนองกับระบบกลับมานั้นจึงเป็นเรื่องที่สำคัญเป็นอย่างยิ่งรวมถึงในปัจจุบันที่มีการนำเอาปัญญาประดิษฐ์เข้ามาประยุกต์ใช้ในชีวิตประจำวันมากขึ้น การที่บริการต่างๆ สามารถรับรู้ถึงอารมณ์ของผู้รับบริการจะช่วยให้การปรับปรุงบริการทำได้อย่างมีประสิทธิภาพมากขึ้น โดยบทความนี้มุ่งเน้นที่การวิเคราะห์อารมณ์บนใบหน้ามนุษย์ ตัวอย่างระบบที่สามารถนำการวิเคราะห์อารมณ์บนใบหน้าของมนุษย์ไปใช้ได้ เช่น ระบบวิเคราะห์การทำงานของพนักงานตามเคาน์เตอร์ โดยสามารถนำเอาอารมณ์ของผู้รับบริการมาประเมินการให้บริการของพนักงานคนนั้นได้ หรือแม้แต่ระบบที่ใช้สำหรับวิเคราะห์อารมณ์ของนักศึกษาที่กำลังเรียนอยู่เพื่อประเมินการสอนของผู้สอน

การแสดงอารมณ์ออกทางใบหน้านั้น Freitas-Magalhães ได้กล่าวไว้ใน [1] อารมณ์พื้นฐานที่มนุษย์ที่แสดงออกทางใบหน้ามีด้วยกัน 7 อารมณ์ ได้แก่ ดีใจ (Joy) รังเกียจ (Disgust) โกรธ (Anger) ดูถูก (Contempt) กลัว (Fear) ตกใจ (Surprise) และเศร้า (Sadness) อย่างไรก็ตาม Disgust และ Contempt ค่อนข้างมีความคล้ายกัน คือ ความไม่ชอบโดย Disgust จะใช้กับการไม่ชอบสิ่งของ เช่น อาหารเน่าเสีย ในขณะที่ Contempt คือการไม่ชอบลักษณะนิสัยของผู้อื่น

ปัจจุบันการเรียนรู้เชิงลึกได้รับความนิยมเป็นอย่างมากเนื่องด้วยประสิทธิภาพของเครื่องคอมพิวเตอร์ในปัจจุบันที่มีมากขึ้น และผลลัพธ์ที่ได้จากการใช้งานแบบจำลองการเรียนรู้เชิงลึกมีค่าที่ดี การเรียนรู้เป็นเรียนรู้จากข้อมูลที่มีปริมาณมากๆ เพื่อให้แบบจำลองสามารถสกัดคุณลักษณะเด่นจากข้อมูลได้แบบอัตโนมัติ แบบจำลองการเรียนรู้เชิงลึกนั้นมีหลายประเภท แต่ประเภทของแบบจำลองการเรียนรู้เชิงลึกที่ได้รับความนิยมเป็นอย่างมาก คือ แบบจำลองเชิงลึกในงานรู้จำภาพ โดยเป็นการใช้โครงข่ายประสาทเทียมประเภท Convolutional Neural Network (CNN)

ซึ่งตัวอย่างงานวิจัยที่มีการนำเทคนิคการเรียนรู้เชิงลึกมาใช้ในการตรวจจับอารมณ์บนใบหน้าของมนุษย์ เช่น งานวิจัยของ Chen และคณะ [2] ได้เปรียบเทียบระหว่างโครงข่ายประสาทเทียมแบบตื้นที่ทำงานร่วมกับ Support Vector Machine (SVM) กับแบบจำลองการเรียนรู้เชิงลึก AlexNet [3] ที่ผ่านการฝึกจากชุดข้อมูลภาพ ImageNet โดยไม่มีการฝึกแบบละเอียดเพิ่มเติมและแบบจำลองการเรียนรู้เชิงลึก AlexNet ที่มีการฝึกแบบละเอียดเพิ่มเติม โดยแบบจำลองมีการใช้ชุดข้อมูลภาพจาก 2 แหล่ง เข้ามาฝึก คือ 1) ArtPhoto [4] ซึ่งมีหมวดหมู่อารมณ์ทั้งหมด 8 อารมณ์ ได้แก่ โกรธ รังเกียจ กลัว เศร้า ความสนุก กลัวเกรง ความพึงพอใจ และความตื่นเต้น และ 2) FlickrEmotion ที่ผู้วิจัยสร้างขึ้นมาเอง ผลจากการทดลองพบว่า แบบจำลองการเรียนรู้เชิงลึกที่มีการฝึกแบบละเอียดให้ประสิทธิภาพความแม่นยำดีที่สุดอยู่ที่ประมาณ 80%

งานวิจัยของ Mehendale [5] ได้นำเสนอการใช้ CNN จำนวน 2 โครงข่าย ทำงานร่วมกันเพื่อตรวจจับอารมณ์บนใบหน้าจากวิดีโอ โดยแบบจำลองสามารถจำแนกอารมณ์ได้ทั้ง 5 อารมณ์ ได้แก่ โกรธ เศร้า มีความสุข กลัว และ ตกใจ และมีความแม่นยำอยู่ที่ 78% ถึง 96%

โดยปกติแล้วการใช้งานแบบจำลองการเรียนรู้เชิงลึกจำเป็นต้องใช้หน่วยประมวลผลกราฟิก (GPU) เพื่อให้การประมวลผลเป็นไปอย่างรวดเร็ว ในช่วงไม่กี่ปีที่ผ่านมาที่แบบจำลองการเรียนรู้เชิงลึก และอินเทอร์เน็ตของสรรพสิ่ง (Internet of things; IoT) ได้รับความนิยมมากขึ้น ทางบริษัท Nvidia จึงได้ผลิต และจำหน่ายบอร์ดสมองกลฝังตัวที่มีขนาดเล็ก และมีหน่วยประมวลผลกราฟิกในตัว อีกทั้งยังมีราคาที่ไม่สูงมากนัก ซึ่งบอร์ดนี้มีชื่อว่า Jetson Nano

งานวิจัยที่ได้นำบอร์ด Jetson Nano มาใช้งานเกี่ยวกับการประมวลผลแบบจำลองการเรียนรู้เชิงลึกแบบเวลาจริง เช่น งานวิจัยของ Vijitkunsawat และ Chantngarm [6] เสนอการเปรียบเทียบแบบจำลองการเรียนรู้ 3 ประเภท คือ Support Vector Machine (SVM), Artificial Neural Network – Multilayer Perceptron (ANN-MLP) และ Convolutional Neural Network – Long Short Term

Memory (CNN-LSTM) ในการควบคุมการเคลื่อนที่ของรถจำลองแบบอัตโนมัติ ซึ่งใช้บอร์ด Jetson Nano เพื่อประมวลผลแบบเวลาจริง ผลจากการทดสอบพบว่า แบบจำลองการเรียนรู้ CNN-LSTM มีความแม่นยำมากที่สุด

งานวิจัยของ Inthanon และ Mungsing [7] เสนอการตรวจจับอาการง่วงนอนจากใบหน้าของมนุษย์ เพื่อแจ้งเตือนและป้องกันการหลับในของผู้ขับขี่ยานพาหนะ โดยการจับจุดสำคัญ (Landmarks) ทั้งหมด 68 จุด บนใบหน้ามนุษย์ด้วยไลบรารี DLib การทดสอบอัลกอริทึมบนบอร์ด Jetson Nano พบว่า มีความเร็วในการประมวลผลที่ดี และอัลกอริทึมที่พัฒนาขึ้นมีความแม่นยำอยู่ที่ 80%

งานวิจัย Chavan และคณะ [8] เสนอการจำแนกสายพันธุ์พืช 15 สายพันธุ์ ด้วยการใช้กล้องที่เชื่อมต่อกับบอร์ด Jetson Nano เพื่อประมวลผลแบบจำลองการเรียนรู้เชิงลึกแบบ CNN ทั้งหมด 3 สถาปัตยกรรม คือ AlexNet, ResNet50 และ MobileNetv2 โดยผลการทดสอบพบว่า AlexNet ให้ความแม่นยำมากที่สุดอยู่ที่ 72%

ดังนั้นงานวิจัยนี้นำมีวัตถุประสงค์เพื่อเสนอการออกแบบและพัฒนาแบบจำลองการเรียนรู้เชิงลึกเพื่อวิเคราะห์อารมณ์จากใบหน้ามนุษย์แบบเวลาจริง และประเมินประสิทธิภาพของแบบจำลองทั้งด้านความแม่นยำและความเร็วในการประมวลผล โดยทดสอบการใช้งานจริงของแบบจำลองบน บอร์ด Jetson Nano ซึ่งเป็นอุปกรณ์สมองกลฝังตัวต้นทุนต่ำ พร้อมทั้งติดตั้งกล้องในตัว ในการจับภาพเคลื่อนไหว เพื่อตีกรอบใบหน้าและแสดงผลการวิเคราะห์อารมณ์ของใบหน้านั้น โดยแบบจำลองจะจำแนกอารมณ์ทั้งหมดออกเป็น 7 อารมณ์ ได้แก่ โกรธ รังเกียจ กลัว ดีใจ เสียใจ ตกใจ และเฉยๆ

2. วัสดุ อุปกรณ์และวิธีการวิจัย

2.1 บอร์ด Nvidia Jetson Nano

Nvidia Jetson Nano [9] คือบอร์ดประมวลผลขนาดเล็กซึ่งพัฒนาโดยบริษัท Nvidia ที่มีประสิทธิภาพสูงสำหรับนำไปใช้กับงานประมวลผลด้านปัญญาประดิษฐ์ในระบบสมองกลฝังตัว หรืออินเทอร์เน็ตของสรรพสิ่ง โดยบอร์ดรุ่นที่มีประสิทธิภาพในระดับที่เหมาะสมและราคาไม่สูงมาก คือ บอร์ด Jetson Nano รุ่นที่มีหน่วยความจำขนาด 4 GB ซึ่งมีราคาบอร์ดอยู่ที่ประมาณ 3,500 บาท

บอร์ด Jetson Nano มีคุณสมบัติทางเทคนิคคร่าวๆ ได้แก่ หน่วยประมวลผลกราฟิก Nvidia Maxwell ที่มีจำนวน Nvidia CUDA core จำนวน 128 คอร์ มีหน่วยประมวลผลกลางเป็น ARM Cortex-A57 MPCore ที่มีคอร์ประมวลผล 4 คอร์ และมีหน่วยความจำแบบ 64-bit LPDDR4 ขนาด 4 GB

การใช้งาน Jetson Nano จะต้องเตรียมการ์ด microSD โดยขนาดที่แนะนำคือ 32 GB จากนั้น จึงดาวน์โหลดและติดตั้ง Jetson Nano Developer Kit SD Card Image [10] หรือ JetPack ลงบนการ์ด microSD ซึ่งภายใน JetPack จะประกอบด้วยระบบปฏิบัติการ Ubuntu ไดร์เวอร์สำหรับ Jetson (L4T) และไลบรารีสำหรับใช้งาน CUDA ร่วมกับการพัฒนาแบบจำลองการเรียนรู้เชิงลึก เช่น TensorRT, cuDNN, CUDA Toolkit, Multimedia API และไลบรารี Computer Vision เป็นต้น

2.2 อัลกอริทึม You Only Look Once (YOLO)

YOLO [11] เป็นแบบจำลองการเรียนรู้เชิงลึกสำหรับการตรวจจับวัตถุและทำนายวัตถุ โดยใช้แนวคิดที่นำ CNN เพียงเครือข่ายเดียว สำหรับการตรวจจับและตีกรอบวัตถุพร้อมกับทำนายผลลัพธ์ที่เป็นความน่าจะเป็นของวัตถุในกรอบนั้นพร้อมๆ กัน ซึ่งส่งผลให้ YOLO มีประสิทธิภาพด้านความเร็วในการประมวลผลสูงมาก จากนั้นได้มีการพัฒนาอย่างต่อเนื่อง และมีการเผยแพร่ YOLO เวอร์ชัน 2 (YOLOv2) [12] และ YOLO เวอร์ชัน 3 (YOLOv3) ตามลำดับ [13]

จากนั้นผู้พัฒนาหลักของ YOLO ได้ประกาศเลิกการพัฒนา YOLO ต่อ แต่ชุมชนผู้วิจัยด้าน Computer Vision ได้ร่วมกันพัฒนา YOLOv4 และเผยแพร่โปรแกรม พร้อมกับเผยแพร่บทความวิจัย [14] รวมถึงมีแบบจำลองรุ่นที่ใช้งานบนอุปกรณ์ที่มีทรัพยากรที่จำกัด ที่เรียกว่า YOLOv4-tiny ที่บีบอัดเครือข่ายให้มีขนาดเล็กลง ทำให้มีแบบจำลองมีการ

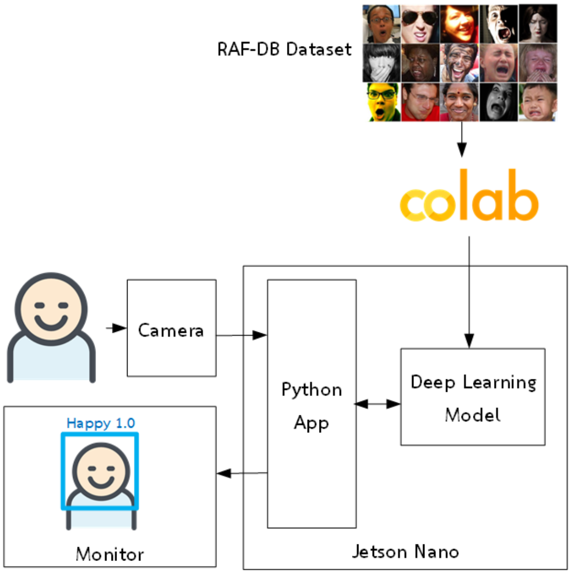
ทำงานที่เร็วขึ้น แต่แลกกับความแม่นยำที่ลดลง

หลังจากการเผยแพร่ของ YOLOv4 ประมาณ 2 เดือน Glenn Jocher ผู้พัฒนาได้ประกาศเผยแพร่โปรแกรม YOLOv5 แต่เขาไม่ได้เขียนบทความวิชาการแต่อย่างไร YOLOv5 ที่พัฒนาขึ้นนี้ แตกต่างจาก YOLO ต้นฉบับทุกเวอร์ชัน เนื่องจากใช้ PyTorch แทนที่การใช้ไลบรารี Darknet แต่อย่างไรก็ตาม YOLOv5 มีโครงสร้างของสถาปัตยกรรมเครือข่ายพื้นฐานเหมือนกับ YOLOv4 ทั้งนี้ได้มีข้อถกเถียงกันว่าแบบจำลองนี้จะสามารถใช้ชื่อ YOLOv5 ได้หรือไม่ ในเมื่อไม่ได้เป็นการพัฒนาต่อจาก YOLOv4 แต่เป็นการตั้งชื่อให้เหมือนเป็นรุ่นใหม่กว่าเพียงเท่านั้น ทั้งนี้ YOLOv5 ได้เผยแพร่แบบจำลองที่ฝึกมาให้แล้ว หลักๆ อยู่ 5 ขนาด คือ YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l และ YOLOv5x ซึ่งมีจำนวนพารามิเตอร์ในแบบจำลองจากน้อยไปมาก ตามลำดับ

ในงานวิจัยนี้จึงได้เลือกใช้แบบจำลอง YOLOv4-tiny และ YOLOv5s ซึ่งเป็นแบบจำลองที่มีขนาดเล็ก ที่ให้ความเร็วในการประมวลผลที่สูง และมีความแม่นยำในระดับที่ยอมรับได้

2.3 การออกแบบภาพรวมการทำงานของระบบ

ภาพรวมการทำงานของระบบวิเคราะห์อารมณ์จากภาพใบหน้าของมนุษย์แสดงดังรูปที่ 1 โดยเริ่มจากนำข้อมูลภาพและฉลาก (Label) จากชุดข้อมูลภาพ RAF-DB (Real-world Affective Faces Database) [15], [16] มาแปลงรูปแบบให้เป็นรูปแบบที่แบบจำลองการเรียนรู้เชิงลึกแต่ละแบบต้องการ ต่อมาจึงนำข้อมูลภาพและฉลากนั้นไปใช้สำหรับการสร้าง ฝึกฝน และทดสอบแบบจำลองบน Google Collaboratory (Google Colab) ซึ่งเป็นบริการคลาวด์ของ Google ที่ให้ผู้ใช้งานสามารถใช้งาน CPU และ GPU เพื่อประมวลผลได้ฟรีในเวลาที่จำกัดในแต่ละวัน ซึ่งเมื่อฝึกฝนแบบจำลองเสร็จเรียบร้อยแล้ว จะได้ผลลัพธ์เป็นไฟล์ค่าน้ำหนัก (ค่าของพารามิเตอร์) ของแบบจำลองอยู่ที่ Google Colab จากนั้นจึงดาวน์โหลดและนำไฟล์นี้ไปติดตั้งลงในบอร์ด Jetson Nano ที่มีการเชื่อมต่อกับกล้อง และจอภาพ



รูปที่ 1 ภาพรวมการทำงานของระบบ

ในการใช้งานระบบ การทำงานหลักอยู่ที่แอปพลิเคชัน Python ที่พัฒนาขึ้น ที่มีหน้าที่ดึงภาพจากกล้อง และส่งผ่านภาพนั้นไปยังแบบจำลองการเรียนรู้เชิงลึก เพื่อประมวลผลอารมณ์ของใบหน้าในภาพนั้นออกมา ผลลัพธ์ของแบบจำลองการเรียนรู้เชิงลึกจะได้พิกัดของใบหน้ามนุษย์ในภาพ และอารมณ์ของใบหน้าในภาพนั้น พร้อมทั้งค่าความมั่นใจของการทำนายโดยมีค่าตั้งแต่ 0.0–1.0 จากนั้นแอปพลิเคชัน Python นำค่าที่ได้เหล่านี้มาแสดงผลร่วมกับภาพที่กล้องจับได้ ออกบนหน้าจอ

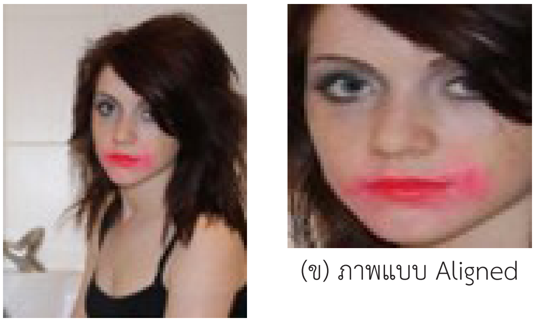
2.4 การเตรียมชุดข้อมูลสำหรับฝึกสอนแบบจำลอง

ในการฝึกแบบจำลองการเรียนรู้ทั้ง YOLOv4-tiny และ YOLOv5s ในงานวิจัยนี้ ได้ใช้ชุดข้อมูลภาพของ RAF-DB ที่ถือว่าเป็นชุดข้อมูลภาพที่เกี่ยวกับการแสดงอารมณ์บนใบหน้ามนุษย์ที่ค่อนข้างใหญ่ ซึ่งแบ่งภาพออกเป็นหมวดหมู่อารมณ์ทั้งหมด 7 อารมณ์ ได้แก่ ตกใจ กลัว รังเกียจ ดีใจ เสียใจ โกรธ และเฉยๆ รวมทั้งหมด 15,339 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับฝึกสอน 12,271 ภาพ และชุดข้อมูลสำหรับทดสอบแบบจำลอง 3,068 ภาพ ตัวอย่างภาพของแต่ละหมวดหมู่อารมณ์แสดงดังรูปที่ 2

(ก) ตกใจ (ข) กลัว (ค) รังเกียจ (ง) เสียใจ

(จ) ดีใจ (ฉ) โกรธ (ช) เฉยๆ

รูปที่ 2 ตัวอย่างภาพในชุดข้อมูลภาพ RAF-DB



(ก) ภาพแบบ Original (ข) ภาพแบบ Aligned

รูปที่ 3 ตัวอย่างรูปแบบภาพในชุดข้อมูลภาพ RAF-DB

อีกทั้งข้อมูลภาพภายในมี 2 รูปแบบ คือ Original และ Aligned โดยชุดข้อมูลภาพที่เป็น Original จะเป็นภาพมุมมองแบบปกติที่พบเห็นกันทั่วไป ในขณะที่ชุดข้อมูลภาพแบบ Aligned นั้น จะเป็นภาพที่เหลือเฉพาะใบหน้าของคนและมีการปรับมุมของภาพให้ตั้งตรงมากที่สุด ดังรูปที่ 3 แสดงความแตกต่างของชุดข้อมูลภาพแบบ Original และ Aligned ในงานวิจัยนี้ได้เลือกใช้ชุดข้อมูลภาพแบบ Original เนื่องจากต้องการให้แบบจำลองเรียนรู้ใบหน้าคนพร้อมกับฉากหลัง เพื่อให้ตอนนำไปใช้งานจริงแบบจำลองจะได้สามารถแยกใบหน้าคนที่จะทำการวิเคราะห์อารมณ์ออกจากฉากหลังของภาพได้โดยอัตโนมัติ

สิ่งที่ RAF-DB ให้มาพร้อมกับชุดข้อมูลภาพก็คือไฟล์เอกสารชื่อ list_partition_label.txt ซึ่งก็คือฉลาก และไฟล์ในไดเรกทอรี boundingbox ที่เก็บค่าตำแหน่งกรอบรอบใบหน้าของมนุษย์ภายในภาพ

เนื่องจากรูปแบบข้อมูลสำหรับฝึกแบบจำลองของ YOLOv4 นั้น มีรูปแบบต่างจากที่ RAF-DB เตรียมมาให้ โดยต้องเตรียมไฟล์ train.txt และ test.txt ที่ภายในมีชื่อไฟล์ที่จะใช้สำหรับฝึกและทดสอบแบบจำลองตามลำดับ ต้องปรับแก้ไฟล์ yolov4-tiny-custom.cfg ให้เหมาะสมกับจำนวนคลาสของวัตถุที่แบบจำลองจะสามารถจำแนกได้ พร้อมทั้งสร้างไฟล์ obj.names และ obj.data สำหรับเก็บชื่อคลาสของแบบจำลอง และการตั้งค่าภาพรวมของแบบจำลองตามลำดับ อีกทั้งต้องคำนวณตำแหน่งของกรอบรอบใบหน้ามนุษย์ใหม่ เนื่องจากค่าพิกัดที่ YOLOv4 ต้องการกับที่ RAF-DB ให้มานั้นมีรูปแบบที่ต่างกัน

สำหรับ YOLOv5 นั้น มีโครงสร้างของชุดข้อมูลที่ต่างจาก YOLOv4 โดยจะใช้ไฟล์ data.yaml ในการเก็บข้อมูลการตั้งค่ารวมของแบบจำลองแทนที่ไฟล์ obj.names และ obj.data และใช้ไดเรกทอรีชื่อ train และ test ที่มีรูปภาพสำหรับฝึกและทดสอบแบบจำลองแทนที่จะใช้ไฟล์ train.txt และ test.txt เหมือนกับ YOLOv4

2.5 การสร้างและฝึกแบบจำลอง

สำหรับแบบจำลอง YOLOv4 นั้น เริ่มต้นการสร้างแบบจำลองโดยการโคลนโปรแกรมต้นฉบับ YOLOv4-tiny จาก Github ของ Alexey มายัง Google Colab จากนั้นปรับแต่งไฟล์ชื่อ Makefile เพื่อให้ไลบรารีนั้นรองรับกับการทำงานบน GPU และใช้คำสั่ง make เพื่อคอมไพล์

เพื่อให้การฝึกแบบจำลองมีความแม่นยำและใช้เวลาในการฝึกน้อยลง ทาง Alexey ได้เตรียมไฟล์น้ำหนักของแบบจำลองที่ฝึกไว้แล้วให้ผู้พัฒนาแบบจำลองคนอื่นสามารถดาวน์โหลดมาใช้เป็นน้ำหนักตั้งต้นของแบบจำลอง เพื่อทำการ Transfer Learning ได้ ซึ่งในที่นี้ใช้ไฟล์น้ำหนักแบบจำลอง YOLOv4-tiny ชื่อ yolov4-tiny.conv.29

จากนั้นเรียกใช้คำสั่งตามรูปที่ 4 เพื่อฝึกแบบจำลอง YOLOv4-tiny โดยจะได้ผลลัพธ์จากการใช้งานคำสั่งนี้

```
!./darknet detector train data/obj.data
cfg/yolov4-tiny-custom.cfg
yolov4-tiny.conv.29 -dont_show -map
```

รูปที่ 4 คำสั่งสำหรับฝึกแบบจำลอง YOLOv4

```
!python train.py --img 416 --batch 64
--epochs 100
--data /content/yolov5/data.yaml
--weights yolov5s.pt
--cache
--project '/content/drive/yolov5'
--name 'exp1'
```

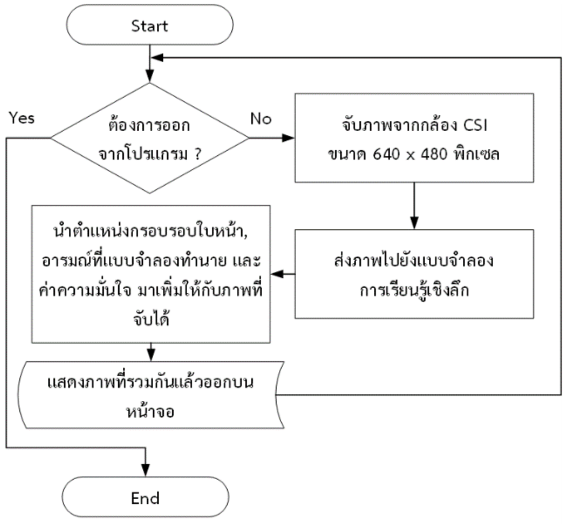
รูปที่ 5 คำสั่งสำหรับฝึกแบบจำลอง YOLOv5

เป็นไฟล์น้ำหนัก 2 ไฟล์ คือ ไฟล์น้ำหนักที่ให้ความแม่นยำมากที่สุดในชุดข้อมูลทดสอบ และไฟล์น้ำหนักของการฝึกแบบจำลองรอบสุดท้าย

สำหรับแบบจำลอง YOLOv5 นั้น เริ่มต้นการสร้างแบบจำลองโดยการโคลนโปรแกรมต้นฉบับ YOLOv5 จาก Github ของ Ultralytics มายัง Google Colab หลังจากนำชุดข้อมูลภาพและไฟล์ที่เกี่ยวข้องที่ได้เตรียมไว้ก่อนนี้ เข้าไปใน Google Colab และเรียกใช้คำสั่งตามรูปที่ 5 เพื่อฝึกแบบจำลอง YOLOv5s ซึ่งค่าน้ำหนักที่ถูกฝึกมาก่อนแล้วของ YOLOv5s จะถูกดาวน์โหลดมาตั้งต้นให้โดยอัตโนมัติ สำหรับการทำ Transfer Learning และหลังจากการฝึกแบบจำลองจะได้ผลลัพธ์ เป็นไฟล์น้ำหนัก 2 ไฟล์ คือ ไฟล์น้ำหนักที่ให้ความแม่นยำมากที่สุดในชุดข้อมูลทดสอบ และไฟล์น้ำหนักของการฝึกแบบจำลองรอบสุดท้าย

2.6 ขั้นตอนการทำงานของโปรแกรมวิเคราะห์อารมณ์จากใบหน้าของมนุษย์

แอปพลิเคชันสำหรับวิเคราะห์อารมณ์จากใบหน้ามนุษย์แบบเวลาจริงพัฒนาด้วยภาษา Python ผู้วิจัยได้พัฒนาไว้ 2 โปรแกรม คือโปรแกรมที่ทำงานร่วมกับแบบจำลอง YOLOv4-tiny และโปรแกรมที่ทำงานร่วมกับแบบจำลอง YOLOv5s โปรแกรมทั้ง 2 โปรแกรม ถูกติดตั้งลงบนบอร์ด



รูปที่ 6 แผนผังการทำงานของโปรแกรม

Jetson Nano ที่ติดตั้งกล้องประเภท Serial ที่เรียกว่า CSI Camera เชื่อมต่อกับบอร์ดผ่านทางพอร์ตกล้องโดยเฉพาะ ซึ่งขั้นตอนการทำงานของโปรแกรม แสดงดังรูปที่ 6

เมื่อเรียกโปรแกรมขึ้นมาทำงาน โปรแกรมจะเชื่อมต่อกับกล้องแบบ CSI ผ่าน Gstreamer เพื่อจับภาพจากกล้องทีละเฟรม โดยมีการตั้งค่าให้กล้องทำงานที่ความละเอียด 640 x 480 พิกเซล จากนั้น นำภาพที่จับได้ส่งไปยังแบบจำลองการเรียนรู้เชิงลึก ซึ่งแบบจำลองการเรียนรู้จะคืนตำแหน่งกรอบรอบใบหน้าที่พบในภาพ ทำนายอารมณ์ของใบหน้าที่พบ และความมั่นใจของการทำนายออกมา ซึ่งโปรแกรมก็จะนำข้อมูลที่แบบจำลองการเรียนรู้เชิงลึกคืนมา เพิ่มเข้าไปกับภาพที่จับได้ แล้วแสดงออกบนหน้าจอ หลังจากนั้นโปรแกรมก็วนจับภาพต่อไปของกล้องเข้ามาประมวลผลต่อไปเรื่อยๆ จนกว่าจะออกจากโปรแกรม

2.7 วิธีการประเมินผลแบบจำลอง

วิธีการประเมินผลแบบจำลองประกอบด้วยกัน 2 รูปแบบ ได้แก่ การประเมินประสิทธิภาพด้านความแม่นยำของแบบจำลอง และการประเมินประสิทธิภาพด้านความเร็วในการทำงานของแบบจำลอง เพื่อเลือกแบบจำลองที่เหมาะสมที่สุดสำหรับทำงานบนบอร์ด Jetson Nano

2.7.1 การประเมินประสิทธิภาพด้านความแม่นยำ

การตรวจสอบความถูกต้องของแบบจำลองสำหรับตรวจจับวัตถุนั้นมีค่าที่สำคัญ 1 ค่า เรียกว่า Intersection over Union (IoU) ซึ่งเป็นการคำนวณสัดส่วนของพื้นที่กรอบรอบวัตถุของค่าความเป็นจริงและที่แบบจำลอง ตีกรอบซ้อนทับกัน หารด้วยพื้นที่ที่กรอบรอบวัตถุของค่าความเป็นจริงและที่แบบจำลองตีกรอบรวมกัน โดยค่า IoU ที่เหมาะสมเพื่อพิจารณาว่าแบบจำลองสามารถทำนายได้อย่างถูกต้องควรมีค่าตั้งแต่ 0.5 ขึ้นไป

ทั้งนี้ การประเมินประสิทธิภาพด้านความถูกต้องของแบบจำลองนั้น จะมีค่าที่สำคัญอยู่ 4 ค่า ได้แก่ Precision, Recall, Accuracy และ F1 score และสำหรับแบบจำลองที่จำแนกหลายคลาสนั้น ปกติจะคำนวณทั้ง 4 ค่านี้ โดยการใช้ Confusion Matrix เข้ามาช่วยในการคำนวณ

2.7.2 การประเมินประสิทธิภาพด้านความเร็ว

นอกเหนือจากการประเมินประสิทธิภาพด้านความแม่นยำของแบบจำลองการเรียนรู้เชิงลึกแล้ว สิ่งที่จำเป็นต้องคำนึงถึงอีกอย่างก็คือ ประสิทธิภาพด้านความเร็วของแบบจำลอง เนื่องจากอุปกรณ์ที่ใช้งานแบบจำลองนี้เป็นอุปกรณ์ที่มีทรัพยากรจำกัด ซึ่งก็คือ บอร์ด Jetson Nano ดังนั้นถึงแม้ว่าแบบจำลองจะมีความแม่นยำสูง แต่หากมีการประมวลผลที่นานมากเกินไป ก็จะไม่เหมาะสมกับวัตถุประสงค์ของงานวิจัยที่ต้องการวิเคราะห์อารมณ์บนใบหน้ามนุษย์แบบเวลาจริง ดังนั้นเพื่อประเมินประสิทธิภาพด้านความเร็วจึงใช้จำนวนอัตราเฟรมของวิดีโอที่แสดงภาพที่กล้องจับได้หลังที่ส่งภาพไปประมวลผลแล้ว โดยมีหน่วยเป็น จำนวนเฟรมต่อวินาที (Frames per second; FPS)

3. ผลการทดลอง

3.1 ผลทดสอบประสิทธิภาพของแบบจำลอง

การทดสอบประสิทธิภาพของแบบจำลอง ได้ทำการทดสอบด้วยชุดข้อมูลทดสอบ ซึ่งประกอบด้วยรูปภาพอารมณ์ทั้งหมด 3,068 ภาพ จากนั้นนำรูปภาพในชุดข้อมูลทดสอบทั้งหมดผ่านเข้าแบบจำลองทั้ง YOLOv4-tiny และ YOLOv5s เพื่อตรวจจับและแปรผลอารมณ์ของใบหน้ามนุษย์จากรูปภาพในชุดข้อมูลทดสอบ โดยกำหนดค่า IoU Threshold เป็น 0.50 และค่า Confidence Threshold เป็น 0.25 ซึ่งเมื่อนำผลลัพธ์จากการทำนายของแบบจำลอง YOLOv4-tiny และ YOLOv5s มาสร้างเป็น Confusion Matrix ได้ดังรูปที่ 7 และรูปที่ 8 ตามลำดับ โดยคอลัมน์ “Not Detect” หมายถึงแบบจำลองไม่สามารถหาใบหน้ามนุษย์ในภาพนั้นได้ เมื่อนำค่าจาก Confusion Matrix ของทั้ง 2 แบบจำลองมาคำนวณค่า Precision, Recall, Accuracy และ F1 score ของแต่ละคลาสอารมณ์ และค่าเฉลี่ยรวมของแบบจำลอง ได้ผลลัพธ์ดังตารางที่ 1

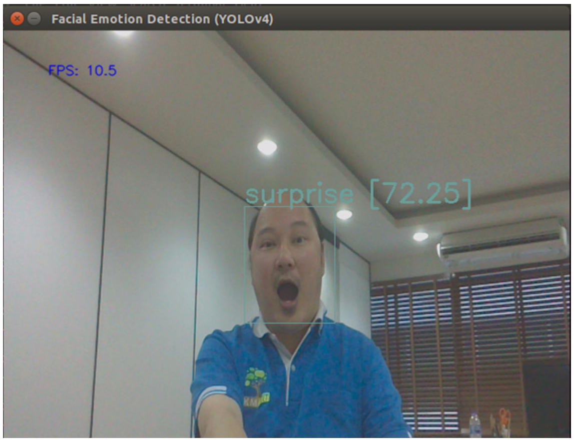
จากผลการประเมินประสิทธิภาพด้านความถูกต้องของแบบจำลอง พบว่า โดยภาพรวมแบบจำลอง YOLOv5s ให้ผลการทำงานที่ดีกว่า YOLOv4-tiny ซึ่งจำนวนภาพที่แบบจำลอง YOLOv4-tiny ไม่สามารถจับภาพใบหน้าได้ทั้งหมด 28 ภาพ ในขณะที่แบบจำลอง YOLOv5s ไม่สามารถตรวจจับ

		Predicted							
		Surprise	Fear	Disgust	Happiness	Sadness	Anger	Neural	Not Detect
Ground-Truth	Surprise	279	8	3	7	9	2	17	4
	Fear	11	43	0	4	8	3	4	1
	Disgust	4	0	78	22	19	11	24	2
	Happiness	5	0	6	1110	14	4	40	6
	Sadness	5	3	9	16	398	3	39	5
	Anger	2	5	9	10	10	116	8	2
	Neutral	16	0	7	29	39	3	578	8

รูปที่ 7 Confusion Matrix ของแบบจำลอง YOLOv4-tiny

		Predicted							
		Surprise	Fear	Disgust	Happiness	Sadness	Anger	Neural	Not Detect
Ground-Truth	Surprise	285	5	8	5	7	2	17	0
	Fear	5	45	3	4	11	2	4	0
	Disgust	2	0	93	12	21	10	21	1
	Happiness	3	0	10	1123	11	5	33	0
	Sadness	3	2	9	9	425	1	27	2
	Anger	3	2	11	5	3	128	8	2
	Neutral	13	0	10	23	60	3	570	1

รูปที่ 8 Confusion Matrix ของแบบจำลอง YOLOv5s



(ก) ผลการจับอารมณ์ตกใจ

(ข) ผลการจับอารมณ์ดีใจ

รูปที่ 9 ตัวอย่างผลการใช้งานจริงของแบบจำลอง YOLOv4-tiny บนบอร์ด Jetson Nano

ตารางที่ 1 ประสิทธิภาพด้านความถูกต้องของแบบจำลอง YOLOv4-tiny และ YOLOv5s

อารมณ์	แบบจำลอง YOLOv4-tiny				แบบจำลอง YOLOv5s			
	Precision	Recall	Accuracy	F1 score	Precision	Recall	Accuracy	F1 score
Surprise	0.866	0.848	0.970	0.857	0.910	0.866	0.976	0.886
Fear	0.729	0.581	0.985	0.647	0.833	0.608	0.988	0.703
Disgust	0.696	0.488	0.962	0.574	0.646	0.581	0.961	0.612
Happiness	0.927	0.937	0.947	0.932	0.951	0.948	0.961	0.949
Sadness	0.801	0.832	0.942	0.816	0.790	0.889	0.946	0.837
Anger	0.817	0.716	0.977	0.763	0.848	0.790	0.981	0.818
Neutral	0.814	0.850	0.924	0.832	0.838	0.838	0.928	0.838
เฉลี่ย	**0.807**	**0.750**	**0.958**	**0.774**	**0.831**	**0.789**	**0.963**	**0.806**

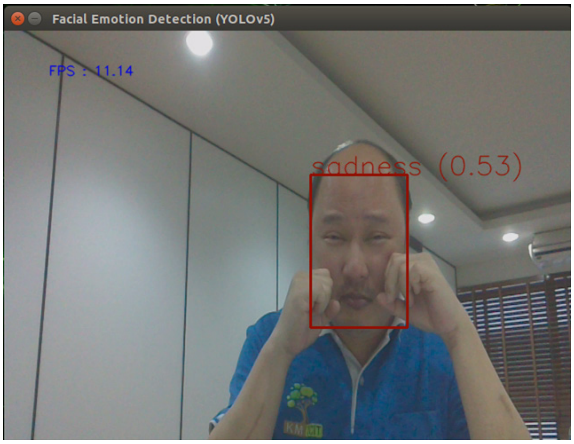
ใบหน้าได้เพียงแค่ 6 ภาพ และเมื่อพิจารณาค่า F1 score จะพบว่า แบบจำลอง YOLOv5s ได้คะแนน 0.806 ในขณะที่แบบจำลอง YOLOv4-tiny ได้คะแนน 0.774

3.2 ผลการทำงานบนบอร์ด Jetson Nano

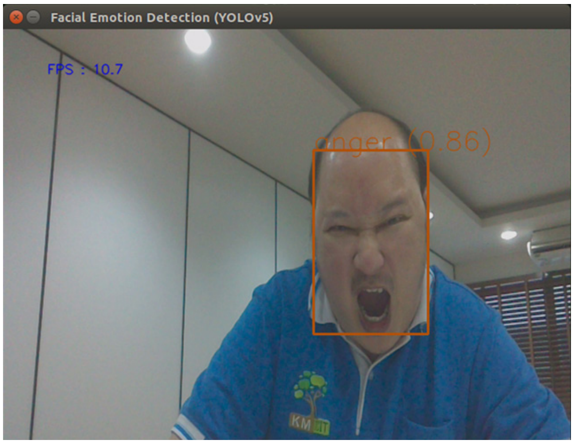
จากการนำแบบจำลองที่ฝึกเสร็จเรียบร้อยแล้ว ไปติดตั้งในบอร์ด Jetson Nano ทำงานร่วมกับโปรแกรม Python ที่พัฒนาขึ้นเพื่อจับภาพของกล้อง แล้วส่งต่อไปยังแบบจำลองเพื่อวิเคราะห์อารมณ์ของมนุษย์แบบเวลาจริง โดยโปรแกรมแสดงวิดีโอที่กล้องจับได้ พร้อมทั้งตีกรอบใบหน้ามนุษย์ที่ตรวจจับอารมณ์ได้ และบอกค่าความเชื่อมั่นในการทำนาย รวมถึงแสดงค่า FPS ของโปรแกรมในขณะทำงาน

สำหรับผลการทดสอบแบบจำลอง YOLOv4-tiny พบว่าโปรแกรมสามารถทำงานด้วยความเร็วในการแสดงผลที่ประมาณ 10.5 FPS และสามารถทำนายผลอารมณ์ตกใจ ดีใจ โกรธ และเฉยๆ ได้ทั้งหมด 4 อารมณ์ ผู้วิจัยไม่สามารถแสดงอารมณ์บนใบหน้าเพื่อให้แบบจำลองตรวจจับอารมณ์กลัว รังเกียจ และเสียใจได้ โดยรูปที่ 9 แสดงตัวอย่างผลการทำงานของโปรแกรมที่ใช้แบบจำลอง YOLOv4-tiny ที่ตรวจจับอารมณ์ตกใจและอารมณ์ดีใจ

ในขณะที่ผลการทดสอบแบบจำลอง YOLOv5s พบว่าโปรแกรมสามารถทำงานด้วยความเร็วในการแสดงผลที่



(ก) ผลการจับอารมณ์เสียใจ



(ข) ผลการจับอารมณ์โกรธ

รูปที่ 10 ตัวอย่างผลการใช้งานจริงของแบบจำลอง YOLOv5s บนบอร์ด Jetson Nano

ประมาณ 11 FPS ซึ่งเร็วกว่าการทำงานของ YOLOv4-tiny เล็กน้อย และสามารถทำนายผลอารมณ์ตกใจ รังเกียจ ดีใจ โกรธ เสียใจ และเฉยๆ ได้ทั้งหมด 6 อารมณ์ แต่อย่างไรก็ตาม ผู้วิจัยไม่สามารถแสดงอารมณ์บนใบหน้าเพื่อให้แบบจำลอง YOLOv5s ตรวจจับอารมณ์กลัวได้ รูปที่ 10 แสดงตัวอย่างผลการทำงานของโปรแกรมที่ใช้แบบจำลอง YOLOv5s ที่ตรวจจับอารมณ์เสียใจ และอารมณ์โกรธ

4. อภิปรายผลและสรุป

งานวิจัยนี้ได้เสนอการวิเคราะห์อารมณ์จากใบหน้ามนุษย์แบบเวลาจริง โดยได้ทดสอบแบบจำลองในการตรวจจับวัตถุ 2 แบบจำลอง คือ แบบจำลอง YOLOv4-tiny และแบบจำลอง YOLOv5s จากการทดลองพบว่า แบบจำลอง YOLOv5 ให้ประสิทธิภาพด้านความแม่นยำที่สูงกว่า YOLOv4-tiny โดยมีค่าแบบจำลอง YOLOv5s มีค่า F1 Score เฉลี่ยอยู่ที่ 0.806 ในขณะที่แบบจำลอง YOLOv4-tiny มีค่า F1 Score เฉลี่ยอยู่ที่ 0.774 และเมื่อนำแบบจำลองที่ได้ฝึกฝนเรียบร้อยแล้ว ไปติดตั้งบนบอร์ด Jetson Nano พบว่าแบบจำลอง YOLOv5s มีความเร็วในการทำงานมากกว่าแบบจำลอง YOLOv4-tiny เล็กน้อย โดยโปรแกรมที่ใช้แบบจำลอง YOLOv5s มีอัตราการแสดงผลเฟรมต่อวินาทีของวิดีโอแบบเวลาจริงที่จับจากกล้องอยู่ที่ 11 FPS ในขณะที่โปรแกรมที่ใช้แบบจำลอง YOLOv4-tiny มีอัตราการแสดงผลเฟรมต่อวินาทีของวิดีโอแบบเวลาจริงที่จับจากกล้องอยู่ที่ 10.5 FPS

จากการทดลองยังพบอีกว่า ในการใช้งานจริงหากไม่แสดงอารมณ์ทางใบหน้าอย่างชัดเจน แบบจำลองทั้ง 2 แบบจำลอง จะไม่สามารถวิเคราะห์อารมณ์ได้อย่างถูกต้อง หรือจับไม่ได้เลย โดยแบบจำลอง YOLOv4-tiny สามารถจับอารมณ์หลักๆ ได้ 4 อารมณ์ ในขณะที่แบบจำลอง YOLOv5s สามารถจับอารมณ์ได้ 6 อารมณ์ ซึ่งอาจจะมีสาเหตุมาจากชุดภาพในการฝึกแบบจำลองของ RAF-DB ที่ไม่สมดุล โดยเฉพาะอย่างยิ่งชุดภาพฝึกแบบจำลองสำหรับอารมณ์กลัวที่มีชุดภาพในการฝึกแบบจำลองเพียง 281 ภาพ ในขณะที่ชุดภาพฝึกแบบจำลองสำหรับอารมณ์ดีใจมีถึง 4,772 ภาพ อีกทั้งบางใบหน้านั้นสามารถแยกอารมณ์แบบชัดเจนได้ยาก เช่น อารมณ์เสียใจและอารมณ์เฉย เนื่องจากใบหน้าเฉยๆ บางครั้งก็ดูเหมือนเสียใจ

จากการทดสอบแบบจำลองทั้งประสิทธิภาพด้านความแม่นยำ และด้านความเร็วของทั้ง 2 แบบจำลอง จึงสามารถสรุปได้ว่า แบบจำลอง YOLOv5s ให้ประสิทธิภาพที่ดีกว่า YOLOv4-tiny ทั้ง 2 ด้าน และสามารถทำงานบนบอร์ด Jetson Nano ที่เป็นอุปกรณ์สมองกลฝังตัวต้นทุนต่ำได้เป็นอย่างดี

เพื่อปรับปรุงประสิทธิภาพของระบบวิเคราะห์อารมณ์จากใบหน้ามนุษยน์แบบเวลาจริงบนอุปกรณ์สมองกลฝังตัวต้นทุนต่ำให้ดีขึ้นไปอีกนั้น จำเป็นต้องเพิ่มชุดข้อมูลฝึกให้มากขึ้นและมีความสมดุลขแงข้แมูล รวมถึงต้องทำความสะอาดชุดข้อมูลฝึกให้ดี เนื่องจากในชุดข้อมูลภาพ RAF-DB เองนั้นก็ยังพบหลายภาพที่มีใบหน้าไม่ชัดเจนแต่มีการติดฉลากอารมณ์ไว้ ซึ่งทำให้รูปภาพเหล่านี้มีผลกับการฝึกแบบจำลอง

สำหรับบอร์ด Jetson Nano นั้น เนื่องจากผู้ผลิตคือ Nvidia ทางผู้ผลิตได้นำเสนอ NVIDIA Deepstream SDK สำหรับการจับวิดีโอที่มีความเร็วมากขึ้น และ TensorRT สำหรับการแปลงแบบจำลอง TensorFlow ให้เล็กลงและทำงานได้เร็วขึ้น ซึ่งงานวิจัยในอนาคตควรประยุกต์ใช้งานทั้ง Deepstream และ TensorRT บนแบบจำลองที่ทำงานบนบอร์ด Jetson Nano เพื่อให้ได้ใช้ศักยภาพของบอร์ด Jetson Nano อย่างสูงสุด

5. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากวิทยาลัยเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ตามสัญญาเลขที่ Res-CIT0609/2021

เอกสารอ้างอิง

[1] A. Freitas-Magalhães, “Facial expression of emotion,” in *Encyclopedia of Human Behavior (Second Edition)*, V. S. Ramachandran, Ed. San Diego: Academic Press, 2012, pp. 173–183.

[2] M. Chen, L. Zhang, and J. P. Allebach, “Learning deep features for image emotion classification,” in *2015 IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 4491–4495.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[4] J. Machajdik and A. Hanbury, “Affective image classification using features inspired by psychology and art theory,” in *Proceedings of the 18th ACM International Conference on Multimedia*, 2010, pp. 83–92.

[5] N. Mehendale, “Facial emotion recognition using convolutional neural networks (FERC),” *SN Applied Sciences*, vol. 2, no. 3, pp. 446, 2020,

[6] W. Vijitkunsawat and P. Chantngarm, “Comparison of machine learning algorithm’s on self-driving car navigation using nvidia jetson nano,” presented at the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June, 2020 (in Thai).

[7] P. Inthanon and S. Mungsing, “Detection of drowsiness from facial images in real-time video media using Nvidia Jetson nano,” presented at the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June, 2020 (in Thai).

[8] S. Chavan, J. Ford, X. Yu, and J. Saniie, “Plant species image recognition using artificial intelligence on jetson nano computational platform,” presented at the 2021 IEEE International Conference on Electro Information Technology (EIT), Mt. Pleasant, MI, USA, 14–15 May, 2021.

[9] Jetson Nano., (2021, Nov.). NVIDIA Developer. [Online]. Available: https://developer.nvidia.com/embedded/jetson -nano

[10] Jetson Nano Developer Kit., (2021, Nov.). NVIDIA Developer. https://developer.nvidia.com/embedded/jetson-nano-developer-kit

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv:1506.02640 [cs]*, 2022.

[12] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," arXiv:1612.08242 [cs], 2022.

[13] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767 [cs]*, 2022.

[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv:2004.10934 [cs, eess]*, 2021.

[15] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2584–2593.

[16] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.