การศึกษาวิธีการตรวจจับการยกมือโดยเฟรมเวิร์กมีเดียไปป์และแบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็น

A Study of Hand-Raising Detection Method Using MediaPipe Framework and CNN Deep Learning Model

พิรุฬห์พิชญ์ ลือโฮ้ง[1], เตชค์ฐสิณป์ เพียซ้าย[2], พิมผกา ประเสริฐศิลป์[3]

[1]สาขาวิชาวิทยาศาสตร์และเทคโนโลยี, มหาวิทยาลัยสุโขทัยธรรมาธิราช, Phirunphit@gmail.com

[2]สาขาวิชาวิทยาศาสตร์และเทคโนโลยี, มหาวิทยาลัยสุโขทัยธรรมาธิราช, Tejtasin.phi@stou.ac.th

[3]สาขาวิชาวิทยาศาสตร์และเทคโนโลยี, มหาวิทยาลัยสุโขทัยธรรมาธิราช, Pimpaka.pra@stou.ac.th

บทคัดย่อ

บทความวิจัยนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาขั้นตอนวิธีการตรวจจับการยกมือด้วยเฟรมเวิร์กมีเดียไปป์และแบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็น และ 2) เพื่อประเมินประสิทธิภาพของขั้นตอนวิธีการที่ได้พัฒนา

ขั้นตอนพัฒนาวิธีการตรวจจับการยกมือที่ตำแหน่งมืออยู่เหนือศีรษะประกอบด้วย 1) การใช้ภาพอินพุตที่มีการยกมือและไม่ยกมือจำนวน 70 ภาพ 2) ภาพอินพุตจะถูกประมวลผลด้วยเฟรมเวิร์กมีเดียไปป์เพื่อระบุตำแหน่งมือที่ยกและตำแหน่งใบหน้าของผู้ยกมือ 3) ภาพที่ระบุตำแหน่งมือและใบหน้าจะถูกฝึกสอนและทดลองด้วยแบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็นเพื่อทำนายการยกมือที่อยู่เหนือศีรษะและ 4) ระบุผลการทำนายของการยกมือเหนือศีรษะ การพัฒนาวิธีการที่นำเสนอใช้ภาษาไพธอน คลังชุดคำสั่งเฟรมเวิร์กมีเดียไปป์ คลังชุดคำสั่งโอเพนซีวี และคลังชุดคำสั่ง TensorFlow

ผลการวิจัยพบว่า 1) วิธีการที่ได้พัฒนาใช้เฟรมเวิร์กมีเดียไปป์สำหรับทำการประมวลผลก่อนและใช้แบบจำลองการเรียนรู้เชิงลึกแบบ CNN สำหรับการทำนายการตรวจจับการยกมือที่มีโครงสร้างเลเยอร์ซ่อนแบบ dense layer 2 เลเยอร์ (64 และ 32 หน่วย พร้อมฟังก์ชันกระตุ้น ReLU) และเอาท์พุตเลเยอร์ 1 หน่วยพร้อมฟังก์ชันกระตุ้น Sigmoid 2) ผลการประเมินได้ประสิทธิภาพความแม่นยำในการตรวจจับการยกมือเหนือศีรษะร้อยละ 96.20

คำหลัก: การตรวจจับการยกมือ, เฟรมเวิร์กมีเดียไปป์, แบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็น

Abstract

The objectives of this research are: 1) to develop the steps of a hand-raising detection method with the MediaPipe framework and CNN deep learning model, and 2) to evaluate the performance of the steps of the developed method.

The develop steps of hand-raising detection method that hand position above a head consist of 1) using input images that are 70 images of hand-raising and no hand-raising, 2) processing the input images with the MediaPipe framework for identifying the raised hand position and face position of raised hand person, 3) identifying the images of the hand and face positions use for training and testing the CNN deep learning model, and 4) identifying a predicted result of hand-raising above the head. The development of the proposed method uses the Python language, the MediaPipe framework library, the OpenCV library, and the TensorFlow library.

The research results show that: 1) The developed method uses the MediaPipe framework for preprocessing and the CNN deep learning model for the prediction of hand-raising detection. The CNN structure has a dense two-layer hidden layer (64 and 32 units with the ReLU function) and one output layer with the Sigmoid function. 2) The evaluation results show a precision of 96.20% in detecting hand-raising above the head.

Keywords: Hand-Raising Detection, MediaPipe framework, CNN deep learning model

ความเป็นมาและความสำคัญของปัญหา

การตรวจจับพฤติกรรมความสนใจต่อกิจกรรมการเรียนรู้ของผู้เรียนทุกวัยสามารถช่วยทำให้ผู้สอนทราบถึงความสนใจของผู้เรียนได้ การยกมือของผู้เรียนถือเป็นพฤติกรรมที่แสดงออกรูปแบบหนึ่งที่แสดงถึงความสนใจในกิจกรรมได้ โดยการยกมือขึ้นเหนือศีรษะจะแสดงออกถึงความสนใจมากที่สุดของผู้ยกมือเมื่อเปรียบเทียบกับผู้ยกมือที่ยกมือไม่ได้อยู่เหนือศีรษะ ซึ่งการยกมือนั้นผู้สอนหรือผู้สังเกตพฤติกรรมจะสามารถทราบจากใบหน้าได้ว่าใครเป็นผู้ยกมือ ดังนั้นหากมีเครื่องมือที่สามารถช่วยรายงานหรือเก็บข้อมูลการยกมือของผู้เรียนได้จะช่วยทำให้ผู้สอนหรือระบบตรวจจับจะสามารถนำข้อมูลที่ได้มาวิเคราะห์พฤติกรรมหรือผลกระทบต่างๆ ที่สะท้อนต่อการแสดงออกด้วยการยกมือขึ้นเหนือศีรษะพร้อมภาพใบหน้าของผู้ยกมือได้ เช่น ความสนใจหรือความตั้งใจยกมือของผู้นั้นเมื่อต้องการสอบถามหรือต้องการแสดงคิดเห็น เป็นต้น และระบบตรวจจับการยกมืออาจถือได้ว่าเป็นนวัตกรรมเกี่ยวกับด้านการเรียนรู้และการมีส่วนร่วม เช่น การมีส่วนร่วมในชั้นเรียนที่เป็นรูปแบบหนึ่งได้ เป็นต้น การศึกษาเพื่อพัฒนา

วิธีการตรวจจับการยกมือขึ้นในลักษณะเหนือศีรษะและให้รู้ว่าเป็นใครยกมือให้เป็นแบบอัตโนมัติสามารถใช้เทคนิคต่างๆ ได้ เช่น การประยุกต์การเรียนรู้ของเครื่องหรือการเรียนรู้เชิงลึกเพื่อจดจำท่าทางการยกมือเป็นต้น งานวิจัยที่เกี่ยวข้องกับการตรวจจับท่าทางของมือของผู้ยกมือมีหลายงานวิจัย ได้แก่

กาญดา ทรัพย์เย็น และคณะ (2559) ได้พัฒนาซอฟต์แวร์สำหรับการทดสอบภาษามือแบบสะกดนิ้วมือเพื่อช่วยให้บุคคลที่มีความบกพร่องทางการได้ยินและการพูดได้สามารถสื่อสารทางภาษาอังกฤษได้ โดยผู้พัฒนาได้ใช้ภาษา C++ พัฒนาร่วมกับคลังชุดคำสั่ง OpenCV เพื่อประมวลผลท่าทางหรือสัญลักษณ์ภาษามือแปลให้เป็นข้อความตัวอักษร

Zhou, H., Jiang, F., & Shen, R. (2018) ได้เสนออัลกอริทึมการรู้จำท่าทางของการยกมือจากภาพของผู้ยกมือที่อยู่ในห้องเรียนด้วยหลักการการรู้จำวัตถุและการประมาณท่าทางโดยใช้อัลกอริทึม R-FCN ที่เป็นเทคนิคการเรียนรู้เชิงลึกของ Convolutional Neural Network (CNN) รูปแบบหนึ่ง

ธนกฤต เอี่ยมเต็ง และ ธนากรสุวรรณกรกิจ (2563) ได้ใช้วิธีการประมวลผลภาพเพื่อตรวจจับการยื่นมือจับสิ่งของของผู้ป่วยที่แขนไม่สามารถเคลื่อนที่ได้อย่างอิสระโดยใช้กล้อง 2 กล้องเป็นอุปกรณ์ตรวจจับร่างกายส่วนของท่อนแขน

พิศณุ คูมีชัย และคณะ (2564) ได้ประยุกต์ใช้อัลกอริทึม 3 วิธีสำหรับการรู้จำสัญญาณมือทางทหาร ได้แก่ วิธีโครงข่ายประสาทเทียม วิธี K-Nearest Neighbors (K-NN) และวิธี template matching ซึ่งตรวจจับสัญญาณผ่านการขยับมือหรือนิ้วในลักษณะต่างๆ กัน

ในส่วนของงานวิจัยที่เกี่ยวข้องกับการตรวจจับภาพใบหน้า ได้แก่

ศานต์สัมพันธ์ จิรกุลชัยวงศ์ และ จีระยุทธ เวทย์วีระพงศ์ (2562) ได้เสนอการปรับปรุงวิธีการรู้จำภาพใบหน้าโดยใช้เทคนิค eigenface การปรับปรุงใช้หลักการของ K-means ในการปรับปรุงฐานข้อมูลสำหรับการแบ่งกลุ่มภาพและปรับปรุงการเปรียบเทียบภาพทดสอบสำหรับการระบุตัวบุคคลด้วยการวัดระยะทาง

เอกรัตน์ สุขสุคนธ์ (2564) ได้ใช้คลังชุดคำสั่งสำเร็จรูปสำหรับการรู้จำใบหน้าของนักศึกษาเปรียบเทียบกับภาพใบหน้าต้นฉบับเพื่อลงทะเบียนการเข้าเรียน

กิตติธัช ตันมา และคณะ (2566) ได้เสนอวิธีการรู้จำใบหน้าบุคคลด้วยหลักการวิเคราะห์ภาพใบหน้าโดยรวมและภาพใบหน้าเฉพาะส่วนโดยใช้เทคนิคการเรียนรู้ของเครื่อง 3 เทคนิคได้แก่ เทคนิคการถดถอยลอจิสติก เทคนิคซัพพอร์ตเวกเตอร์แมทชีน และเทคนิคป่าสุ่ม แล้วทำการโหวตแต่ละค่าผลลัพธ์เป็นตัวแทนของระบบ

จากแต่ละงานวิจัยที่เกี่ยวข้องพบว่าหากต้องการตรวจจับว่าใครเป็นผู้ยกมือจำเป็นต้องทำการตรวจจับตำแหน่งของท่าทางมือที่ยกขึ้นเหนือศีรษะและตำแหน่งของภาพใบหน้ารวมกันโดยอาจประยุกต์ใช้คลังชุดคำสั่ง (library) ของโอเพนซีวี (OpenCV) และเทคนิคการเรียนรู้เชิงลึก เช่น CNN เป็นต้น มาร่วมกันช่วยพัฒนาขั้นตอนสำหรับวิธีการตรวจจับได้ อย่างไรก็ตาม การใช้เทคนิคการเรียนรู้เชิงลึก

แบบ CNN จะมีข้อจำกัดในเรื่องของจำนวนภาพข้อมูลที่นำมาใช้สร้างแบบจำลองที่มากพอเนื่องจากวิธีการเรียนรู้ของ CNN จะจดจำแพทเทิร์น (pattern) ของวัตถุหรือคนที่ยกมือต้องผ่านกระบวนลดมิติของภาพหรือความละเอียด (resolution) ภาพเพื่อสร้างแบบจำลองสำหรับการทำนาย ดังนั้นการป้องกันการลดผลกระทบดังกล่าวของภาพข้อมูลการยกมืออาจต้องหาตัวแทน pattern ของคนยกมือในภาพเพื่อไม่ให้กระทบต่อกระบวนการในโครงสร้างของ CNN และอาจช่วยลดจำนวนของภาพข้อมูลอินพุตที่นำมาใช้ฝึกสอนเพื่อสร้างแบบจำลองได้ เฟรมเวิร์กมีเดียไปป์ (MediaPipe) เป็นเครื่องมือสำเร็จรูปเครื่องมือหนึ่งที่ช่วยในการตรวจจับท่าทางในภาพ เช่น มือหรือใบหน้าได้ เป็นต้น โดยภาพข้อมูลที่ผ่านกระบวนการของเฟรมเวิร์ก MediaPipe จะถูกตรวจจับและทำเครื่องหมายเป็นป้ายกำกับท่าทางมือหรือใบหน้าไว้บนภาพที่สามารถถือเป็นตัวแทน pattern ในภาพได้ ดังนั้นในบทความนี้จึงนำเสนอการศึกษาการใช้เฟรมเวิร์กมีเดียไปป์และแบบจำลองการเรียนรู้เชิงลึกแบบ CNN เพื่อพัฒนาเป็นวิธีการตรวจจับการยกมือขึ้นเหนือศีรษะ

วัตถุประสงค์

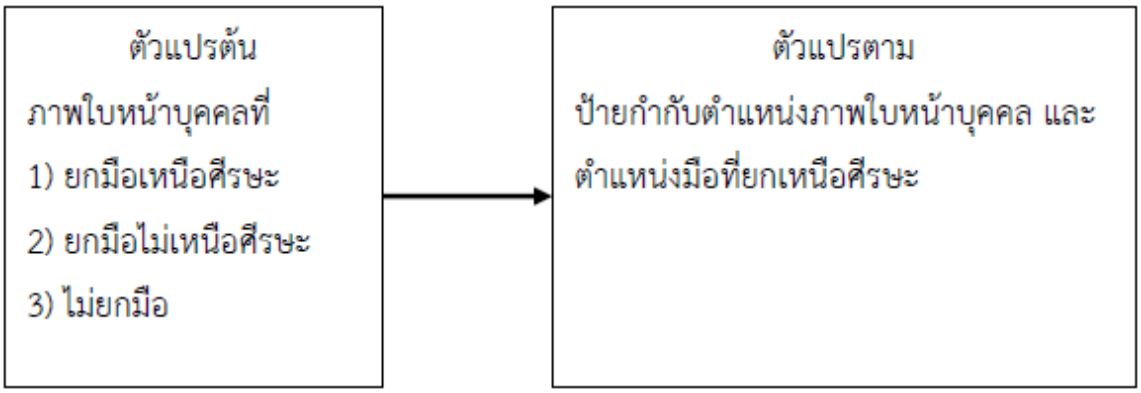
1. เพื่อพัฒนาขั้นตอนวิธีในการตรวจจับการยกมือด้วยเฟรมเวิร์กมีเดียไปป์และแบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็น
2. เพื่อประเมินประสิทธิภาพของขั้นตอนวิธีในการตรวจจับการยกมือด้วยเฟรมเวิร์กมีเดียไปป์และแบบจำลองการเรียนรู้เชิงลึกแบบซีเอ็นเอ็น

ประโยชน์ที่คาดว่าจะได้รับ

ได้ขั้นตอนวิธีการสำหรับต้นแบบการตรวจจับการยกมือเพื่อประยุกต์ใช้วิเคราะห์พฤติกรรมการยกมือที่ผู้ยกมือมีความสนใจต่อกิจกรรม เช่น กิจกรรมการเรียนการสอนในลักษณะการมีส่วนร่วมในชั้นเรียนได้ เป็นต้น

กรอบแนวคิด

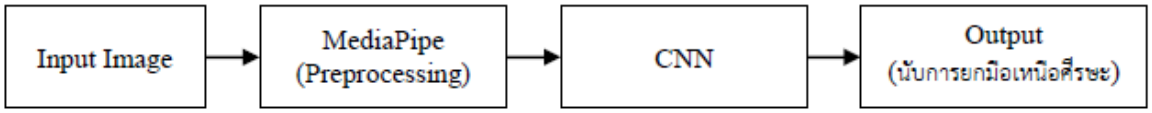
การตรวจจับการยกมือในงานวิจัยนี้ ตำแหน่งของมือที่ยกต้องอยู่เหนือตำแหน่งของศีรษะ โดยตำแหน่งของศีรษะจะถูกตรวจจับโดยใช้ทั้งภาพใบหน้า ตัวแปรใช้พิจารณาดังภาพ 1



ภาพ 1 กรอบแนวคิดที่พิจารณา

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงทดลองเพื่อศึกษาพัฒนาและประเมินประสิทธิภาพของวิธีการตรวจจับการยกมือขึ้นเหนือศีรษะด้วยเฟรมเวิร์ก MediaPipe และแบบจำลองการเรียนรู้เชิงลึกแบบ CNN โดยเฟรมเวิร์ก MediaPipe ทำหน้าที่ประมวลผลก่อน (pre-processing) และแบบจำลอง CNN ทำหน้าที่สร้างแบบจำลองและทำนายผล โดยมีสมมติฐานว่า วิธีการที่นำเสนอสามารถให้ประสิทธิภาพค่าความแม่นยำ (precision) ที่สูง ขั้นตอนของวิธีการที่นำเสนอดังภาพ 2



ภาพ 2 ขั้นตอนการประมวลผลของวิธีการที่นำเสนอ

ขั้นตอนการประมวลผล ประกอบด้วย

1. การใช้ภาพอินพุต (input image) 50 ภาพสำหรับการฝึกสอนเพื่อสร้างแบบจำลองสำหรับการทำนายประกอบด้วย ภาพคนหนึ่งคนต่อภาพที่ยกมือเหนือศีรษะ 30 ภาพ และที่ไม่ยกมือจำนวน 20 ภาพ และการใช้ภาพอินพุต 70 ภาพสำหรับการทดสอบเพื่อทำนายตำแหน่งของคนที่ยกมือเหนือศีรษะ
2. ภาพอินพุตทั้งหมดจะถูกประมวลผลก่อนด้วยเฟรมเวิร์ก MediaPipe เพื่อตรวจจับและทำป้ายกำกับตำแหน่งของมือที่ยกและตำแหน่งภาพใบหน้าของคนที่ยกมือ
3. ภาพที่ถูกทำป้ายกำกับจะถูกป้อนให้กับแบบจำลอง CNN เพื่อสร้างแบบจำลองหรือทำนาย โดยโครงสร้างของ CNN ประกอบด้วย เลเยอร์ซ่อน 2 เลเยอร์ และเลเยอร์เอาท์พุต 1 เลเยอร์
4. ภาพเอาท์พุต (output) เป็นภาพที่ถูกประมวลผลและทำป้ายกำกับเป็นตำแหน่งของมือและตำแหน่งภาพใบหน้าของผู้ที่ยกมือเหนือศีรษะ

5. การประเมินประสิทธิภาพของวิธีการที่นำเสนอจะแสดงตาราง confusion matrix และคำนวณค่าตามสมการ (1)-(4)

Confusion matrix	Actual positive	Actual negative
Predicted positive	TP (True positive) (ผลทำนายและผลจริงยกมือเหนือศีรษะ)	FP (False positive) (ผลทำนายยกมือเหนือศีรษะ ผลจริงไม่ได้ยกมือเหนือศีรษะ)
Predicted negative	FN (False negative) (ผลทำนายไม่ได้ยกมือเหนือศีรษะ ผลจริงยกมือเหนือศีรษะ)	TN (True negative) (ผลทำนายและผลจริงไม่ได้ยกมือเหนือศีรษะ)

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

โดยผลจริงเป็นผลจากผู้เชี่ยวชาญพิจารณาตามเงื่อนไขของการยกมือที่ตำแหน่งของมืออยู่เหนือศีรษะของผู้ยกมือคนนั้นหรือเห็นใบหน้าของผู้ยกมือเหนือศีรษะ สมการ 1) แสดงถึงความแม่นยำ (precision) ที่ผลทำนายตอบตรงกับผลจริง สมการ 2) แสดงถึงการเรียกคืน (recall) ที่ผลทำนายตอบครบกับผลจริง สมการ 3) หรือ F1-Score แสดงถึงการวัดความสามารถของแบบจำลอง และสมการ 4) แสดงถึงความถูกต้อง (accuracy) ของการทำนายที่ทำนายถูกทั้งหมดเทียบกับจำนวนข้อมูลทั้งหมด

เครื่องมือที่ใช้ในการวิจัย ประกอบด้วย 1) โปรแกรมเขียนภาษาไพธอน Visual Studio Code เวอร์ชัน 1.86, 2) ภาษาไพธอน เวอร์ชัน 3.10.13, 3) คลังชุดคำสั่งเฟรมเวิร์ก MediaPipe เวอร์ชัน 0.10.9, 4) คลังชุดคำสั่ง OpenCV เวอร์ชัน 4.9.0.80, 5) คลังชุดคำสั่ง TensorFlow/Keras และโครงสร้างของแบบจำลองการเรียนรู้เชิงลึกแบบ CNN ประกอบด้วย เลเยอร์ที่ 1 แบบ Dense (64 หน่วยและฟังก์ชันกระตุ้น ReLU) และเลเยอร์ที่ 2 แบบ Dense (32 หน่วยและฟังก์ชันกระตุ้น ReLU) และสำหรับเอาท์พุตเลเยอร์แบบ Dense (1 หน่วยและฟังก์ชันกระตุ้น Sigmoid) ขนาดภาพอินพุตกำหนดขนาด 250x250 พิกเซล จำนวน epoch เท่ากับ 15 และขนาด batch size เท่ากับ 16

ผลการวิจัย

สำหรับการทดลองได้กำหนดโครงสร้างเลเยอร์ของโครงข่ายประสาทแบบคอนโวลูชันสำหรับใช้ตรวจจับการยกมือขึ้นเหนือศีรษะประกอบด้วยเลเยอร์ที่ 1 แบบ Dense (64 หน่วยและฟังก์ชันกระตุ้น ReLU) และเลเยอร์ที่ 2 แบบ Dense (32 หน่วยและฟังก์ชันกระตุ้น ReLU) และสำหรับเอาท์พุตเลเยอร์แบบ Dense (1 หน่วยและฟังก์ชันกระตุ้น Sigmoid) โดยภาพอินพุตที่ป้อนเข้าระบบ CNN เป็นภาพที่ได้จากการตรวจจับตำแหน่งของท่าทางมือและตำแหน่งของใบหน้าจาก MediaPipe ดังภาพ 3 และภาพอินพุตจำนวน 70 ภาพที่ประกอบด้วยภาพการยกมือเหนือศีรษะและยกมือไม่อยู่เหนือศีรษะถูกใช้ในการฝึกสอนเพื่อสร้างแบบจำลองสำหรับ CNN ในการตรวจจับแยกประเภทการยกมือ

ภาพ 3 ภาพฝึกสอน

ภาพ 4 ผลลัพธ์ตัวอย่างจาก MediaPipe

ผลการทดลองกับภาพชุดข้อมูลทดสอบ 70 ภาพแสดงดังตาราง 1

ตาราง 1 ผลการทดลองกับภาพชุดข้อมูลทดสอบจำนวน 70 ภาพ

No.1, ผู้เชี่ยวชาญ: 3 TP=3, TN=0, FP=0, FN=0	No.2, ผู้เชี่ยวชาญ: 2 TP=2, TN=0, FP=0, FN=0	No.3, ผู้เชี่ยวชาญ: 1 TP=1, TN=0, FP=0, FN=0	No.4, ผู้เชี่ยวชาญ: 2 TP=2, TN=0, FP=0, FN=0
No.5, ผู้เชี่ยวชาญ: 2 TP=2, TN=0, FP=0, FN=0	No.6, ผู้เชี่ยวชาญ: 3 TP=2, TN=0, FP=0, FN=1	No.7, ผู้เชี่ยวชาญ: 0 TP=0, TN=0, FP=0, FN=0	No.8, ผู้เชี่ยวชาญ: 3 TP=1, TN=0, FP=0, FN=2
No.9, ผู้เชี่ยวชาญ: 0 TP=0, TN=1, FP=0, FN=0	No.10, ผู้เชี่ยวชาญ: 1 TP=0, TN=0, FP=0, FN=1	No.11, ผู้เชี่ยวชาญ: 4 TP=2, TN=0, FP=0, FN=2	No.12, ผู้เชี่ยวชาญ: 4 TP=1, TN=0, FP=0, FN=3
No.13, ผู้เชี่ยวชาญ: 0 TP=0, TN=0, FP=0, FN=0	No.14, ผู้เชี่ยวชาญ: 4 TP=4, TN=0, FP=0, FN=0	No.15, ผู้เชี่ยวชาญ: 1 TP=1, TN=0, FP=0, FN=0	No.16, ผู้เชี่ยวชาญ: 3 TP=3, TN=0, FP=0, FN=0
No.17, ผู้เชี่ยวชาญ: 1 TP=1, TN=0, FP=0, FN=0	No.18, ผู้เชี่ยวชาญ: 4 TP=0, TN=0, FP=0, FN=4	No.19, ผู้เชี่ยวชาญ: 0 TP=0, TN=0, FP=0, FN=0	No.20, ผู้เชี่ยวชาญ: 0 TP=0, TN=0, FP=0, FN=0

No.21, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.22, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.23, ผู้เชี่ยวชาญ: 3
TP=1, TN=0, FP=0, FN=2
No.24, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
No.25, ผู้เชี่ยวชาญ: 3
TP=3, TN=0, FP=0, FN=0
No.26, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
Education
No.27, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.28, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
No.29, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
No.30, ผู้เชี่ยวชาญ: 3
TP=1, TN=0, FP=0, FN=2
No.31, ผู้เชี่ยวชาญ: 2
TP=1, TN=0, FP=0, FN=1
No.32, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.33, ผู้เชี่ยวชาญ: 3
TP=0, TN=0, FP=0, FN=3
No.34, ผู้เชี่ยวชาญ: 3
TP=2, TN=0, FP=0, FN=1
No.35, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.36, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=2
No.37, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.38, ผู้เชี่ยวชาญ: 2
TP=1, TN=0, FP=0, FN=1
No.39, ผู้เชี่ยวชาญ: 0
TP=0, TN=0, FP=0, FN=0
No.40, ผู้เชี่ยวชาญ: 3
TP=2, TN=0, FP=0, FN=1
No.41, ผู้เชี่ยวชาญ: 3
TP=0, TN=0, FP=0, FN=3
No.42, ผู้เชี่ยวชาญ: 6
TP=6, TN=0, FP=0, FN=0
No.43, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
No.44, ผู้เชี่ยวชาญ: 1
TP=1, TN=0, FP=0, FN=0
No.45, ผู้เชี่ยวชาญ: 3
No.46, ผู้เชี่ยวชาญ: 0
No.47, ผู้เชี่ยวชาญ: 1
No.48, ผู้เชี่ยวชาญ: 0

และสำหรับการทดลองกับข้อมูลภาพอินพุตชุดเดียวกันโดยไม่มีขั้นตอนการประมวลผลก่อนด้วยเฟรมเวิร์กมีเดียไปป์และป้อนประมวลผลโดยตรงกับแบบจำลองการเรียนรู้เชิงลึกแบบ CNN อย่างเดียว จะได้ค่าเปรียบเทียบกับวิธีการที่ได้นำเสนอในตาราง confusion matrix

จากตาราง 1 สามารถแสดงค่า confusion matrix ได้ดังนี้

Confusion matrix	วิธีการแบบ MediaPipe + CNN		วิธีการแบบ CNN	
	ผลจริงที่ยกมือเหนือศีรษะ	ผลจริงที่ไม่ได้ยกมือเหนือศีรษะ	ผลจริงที่ยกมือเหนือศีรษะ	ผลจริงที่ไม่ได้ยกมือเหนือศีรษะ
ผลทำนายที่ยกมือเหนือศีรษะ	TP=76	FP=3	TP=92	FP=88
ผลทำนายที่ไม่ได้ยกมือเหนือศีรษะ	FN=47	TN=0	FN=31	TN=0

ในแต่ละค่าของทั้งสองวิธีในตาราง confusion matrix จะถูกนำมาคำนวณค่าประสิทธิภาพตามสมการ (1)-(4) และแสดงค่าคำนวณดังตาราง 2

ตาราง 2 การเปรียบเทียบประสิทธิภาพ

ประสิทธิภาพ	วิธีการแบบ MediaPipe + CNN	วิธีการแบบ CNN
Precision	96.20%	49.20%
Recall	61.79%	74.80%
F1-Score	75.25%	59.36%
Accuracy	60.32%	43.60%

จากตารางเปรียบเทียบแสดงให้เห็นว่า ค่า precision=96.20% ของวิธีการที่นำเสนอจะให้ความแม่นยำที่ผลทำนายตรงกับผลจริงสูงกว่าวิธีการที่ใช้เฉพาะแบบจำลอง CNN สำหรับทำการตรวจจับเพียงอย่างเดียวแม้จะมีค่า recall ที่สูงกว่าแต่เพราะจากผลจากการใช้ pattern กับแบบจำลอง CNN ในการสร้างแบบจำลองอาจทำนายการตรวจจับในภาพที่ไม่ใช่บริเวณของผู้ยกมือจริงได้ทำให้ความถูกต้อง (accuracy) จึงต่ำกว่าวิธีการที่นำเสนอแม้จะมีค่าความถูกต้องไม่สูงเนื่องจากการใช้จำนวนภาพยกมือที่มีจำนวนไม่มากสำหรับการฝึกสอนเพื่อสร้างแบบจำลอง CNN

สรุปผลการวิจัย

จากตาราง 1 พบว่า วิธีการที่ได้พัฒนามีประสิทธิภาพความแม่นยำในการตรวจจับการยกมือขึ้นเหนือศีรษะในระดับดีมากโดยมีค่าร้อยละ 96.20 และแสดงให้เห็นว่าการประยุกต์ใช้เฟรมเวิร์ก MediaPipe ร่วมกับแบบจำลองการเรียนรู้เชิงลึกแบบ CNN มีโอกาสช่วยทำให้โมเดลสามารถจำแนกการยกมือเหนือศีรษะได้อย่างมีประสิทธิภาพได้

อภิปรายผล

จากการวิจัยพบว่าวิธีการตรวจจับการยกมือขึ้นเหนือศีรษะที่พัฒนามีค่าความแม่นยำถึงร้อยละ 96.20 ซึ่งสามารถอภิปรายประเด็นสำคัญได้ดังนี้ การประยุกต์ใช้เฟรมเวิร์ก MediaPipe ในการตรวจจับตำแหน่งมือที่ยกและใบหน้าจะเป็นการทำเครื่องหมายและป้ายกำกับไว้เพื่อใช้เป็นตัวแทนลักษณะเด่น (feature) ของการยกมือสำหรับฝึกสอนและทดสอบแบบจำลอง CNN แทนการใช้ภาพ pattern โดยตรง เนื่องจากเครื่องหมายที่กำกับบนมือและกรอบใบหน้าด้วยเฟรมเวิร์กมีเดียไปป์จะช่วยลดผลกระทบจากการลดมิติภาพที่ส่งผลต่อความละเอียดของภาพได้อย่างดีเพราะเครื่องหมายยังคงระบุตำแหน่งของมือและใบหน้าได้ อย่างไรก็ตาม ยังมีข้อสังเกตจากจำนวนภาพข้อมูลที่ใช้สร้างแบบจำลอง CNN มีปริมาณจำนวนภาพสำหรับฝึกสอนน้อยทำให้ความถูกต้องของการทำนายไม่สูงมากนัก

ข้อเสนอแนะ

แม้ความแม่นยำของวิธีการที่นำเสนอจะสูงแต่ความถูกต้องขึ้นอยู่ปริมาณของจำนวนภาพยกมือหรือไม่ยกมือเพื่อนำมาฝึกสอนสร้างแบบจำลองและทดสอบการทำนาย ดังนั้นควรเพิ่มจำนวนภาพสำหรับการฝึกสอนและใช้ทดสอบให้มากขึ้น และสภาพแวดล้อมในวิธีที่นำเสนอเหมาะกับการใช้ในห้องเรียน

เอกสารอ้างอิง

กาญดา ทรัพย์เย็น, วรัญญา สมานทรัพย์, และจารุวรรณ สุระเสียง. (2559). ซอฟต์แวร์เพื่อการประมวลภาษามือ. *วารสารวิชาการ มทร.สุวรรณภูมิ ฉบับมนุษยศาสตร์และสังคมศาสตร์,* 4(2), 159-170.

Zhou, H., Jiang, F., & Shen, R. (2018). Who Are Raising Their Hands? Hand-Raiser Seeking Based on Object Detection and Pose Estimation. Proceedings of the 10th Asian Conference on Machine Learning. *Proceedings of Machine Learning Research.* 95:470-485 Available from https://proceedings.mlr.press/v95/zhou18a.html.

ธนกฤต เอี่ยมเต็ง, และธนากร สุวรรณกรกิจ. (2563). *การตรวจจับการยื่นมือจับสิ่งของโดยใช้กล้อง.* [ปริญญานิพนธ์วิศวกรรมศาสตรบัณฑิต]. ห้องสมุดคณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

พิศณุ คูมีชัย, ภาณุกร วัฒนจัง, นฐกร มุสิกะจินดา, วรุตม์ ขำญาติ, และดนุภัทร ศักดิ์. (2564). ระบบรู้จำสัญญาณมือทางทหารด้วยการประมวลผลภาพ. *วารสารวิทยาศาสตร์ เทคโนโลยีและนวัตกรรม,* 2(5), 46-55.

ศานต์สัมพันธ์ จิรกุลชัยวงศ์ , ดร.จีระยุทธ เวทย์วีระพงศ. (2562). *การปรับปรุงการรู้จำภาพใบหน้าของวิธีไอเกนเฟสมุมมอง.* การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 20. มหาวิทยาลัยขอนแก่น.

เอกรัตน์ สุขสุคนธ์. (2564). ระบบลงทะเบียนใบหน้าและตรวจสอบนักศึกษาเข้าห้องเรียนด้วยการประมวลผลภาพร่วมกับไลบรารีการรู้จำใบหน้า. *วารสารวิทยาการสารสนเทศและวิศวกรรมศาสตร์,* 11(2), 43-52.

กิตติธัช ตันมา, เตชค์ฐสิณป์ เพียซ้าย, และณัฐพร เห็นเจริญเลิศ. (2566). *วิธีการรู้จำใบหน้าบุคคลด้วยภาพใบหน้าโดยรวมและเฉพาะส่วน.* ใน รายงานสืบเนื่องจากการประชุมวิชาการและการนำเสนอผลงานวิจัยระดับชาติ ครั้งที่ 10 และระดับนานาชาติครั้งที่ 8 (น. 1060-1073). มหาวิทยาลัยภาคตะวันออกเฉียงเหนือ.

MediaPipe. (2025, December 19). In *Wikipedia.* https://en.wikipedia.org/wiki/MediaPipe.