

การศึกษาแบบจำลองการเรียนรู้การตรวจจับใบหน้าและ  
นับจำนวนคนเพื่อประเมินความสนใจ  
ผ่านเทคนิคการเรียนรู้เชิงลึก

FACE DETECTION AND PEOPLE COUNTING LEARNING  
MODEL STUDY TO ASSESS ATTENTION  
THROUGH DEEP LEARNING TECHNIQUES



สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)  
ภาควิชาสถิติ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FACE DETECTION AND PEOPLE COUNTING LEARNING  
MODEL STUDY TO ASSESS ATTENTION THROUGH  
DEEP LEARNING TECHNIQUES



A COOPERATIVE EDUCATION SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENT FOR  
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)  
DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
ACADEMIC YEAR 2022

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา

การศึกษาแบบจำลองการเรียนรู้การตรวจจับใบหน้าและ  
นับจำนวนคนเพื่อประเมินความสนใจผ่านเทคนิคการเรียนรู้เชิงลึก  
Face Detection and People Counting Learning Model  
Study to Assess Attention Through Deep Learning  
Techniques

ชื่อนักศึกษา

นางสาวศิริณญา เทวกุล รหัสนักศึกษา 62050837

ปริญญา

วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)

ภาควิชา

สถิติ

ปีการศึกษา

2565

อาจารย์ที่ปรึกษา

ดร. สกุนา ศรีอินมัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้  
สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)  
ประจำปีการศึกษา 2565

|   |   |
|---|---|
| คณะกรรมการสอบ                                     | ลายมือชื่อ  |
| ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์<br>ประธานกรรมการ       |   |
| คุณวิศิษฐ์ กิจชัยนุกูล<br>กรรมการ                 |  |
| ดร. สกุนา ศรีอินมัย<br>กรรมการและอาจารย์ที่ปรึกษา |  |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|                  |  |
|------------------|--|
| หัวข้อสหกิจศึกษา | การศึกษาแบบจำลองการเรียนรู้การตรวจจับใบหน้าและ<br>นับจำนวนคนเพื่อประเมินความสนใจผ่านเทคนิคการเรียนรู้เชิงลึก |
| ชื่อนักศึกษา     | นางสาวศิรินญา เทวกุล รหัสนักศึกษา 62050837   |
| ปริญญา           | วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)  |
| ภาควิชา          | สถิติ  |
| คณะ              | วิทยาศาสตร์  |
| มหาวิทยาลัย      | สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)  |
| ปีการศึกษา       | 2565   |
| อาจารย์ที่ปรึกษา | ดร. สกฤณา ศรีอินมัย  |

### บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาเทคโนโลยีที่ใช้ในการตรวจจับบุคคล นับจำนวนบุคคล และประเมินความสนใจของแต่ละบุคคล เพื่อเปรียบเทียบประสิทธิภาพระหว่างมุมมองในการตรวจจับ โดยผู้วิจัยได้ศึกษาและใช้แบบจำลองโยโลเวอร์ชัน 7 (YOLOv7) ในการตรวจจับและนับจำนวนบุคคล และใช้แบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ในการตรวจจับจุดสังเกตบนใบหน้าและวัดองศาระหว่างใบหน้าและกล้องเพื่อประเมินความสนใจ ในการตรวจจับและนับจำนวนบุคคลได้แบ่งข้อมูลทดสอบออกเป็น 2 ชุด คือ ชุดข้อมูล A1 และชุดข้อมูล B1 ข้อมูลแต่ละชุดประกอบด้วยวิดีโอในมุมมองแนวระนาบ และมุมมองแนวเหนือศีรษะ ผลการวิเคราะห์คือ ในมุมมองแนวเหนือศีรษะได้ค่าความถูกต้อง 72.84% มุมมองแนวระนาบได้ค่าความถูกต้อง 65.67% ซึ่งค่าความแม่นยำ ค่าความระลึก และค่าวัดประสิทธิภาพของมุมมองแนวเหนือศีรษะมีค่าสูงกว่าในมุมมองแนวระนาบ จึงสรุปได้ว่าแบบจำลอง YOLOv7 เหมาะกับการตรวจจับและนับจำนวนบุคคลในมุมมองแนวเหนือศีรษะมากกว่ามุมมองแนวระนาบ และได้ค่า AUC = 0.80 คือ แบบจำลองสามารถทำนายและจำแนกบุคคลได้ในระดับดี ถัดมาแบบจำลองในการตรวจจับและวัดองศาของใบหน้าและกล้องเพื่อประเมินความสนใจ จะทำการทดสอบด้วยข้อมูล 2 ชุด คือข้อมูลชุด A2 และข้อมูลชุด B2 โดยจะทดสอบในมุมมองแนวระนาบด้วยมุมในการหันใบหน้า 15 องศา ได้ค่าความถูกต้อง 55% และได้ค่า AUC = 0.76 คือ แบบจำลองสามารถทำนายได้ในระดับดี จากผลการวัดประสิทธิภาพต่าง ๆ ของแบบจำลองในงานวิจัยครั้งนี้สามารถนำไปเป็นต้นแบบหรือเป็นแนวทางในการนับจำนวนคนและการตรวจจับความสนใจของลูกค้าที่มีต่อป้ายโฆษณาดิจิทัลขนาดเล็กได้

**คำสำคัญ :** โยโล, มีเดียไปป์ , การตรวจจับบุคคล , การตรวจจับจุดสังเกตบนใบหน้า, ป้ายโฆษณาดิจิทัล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|                      |  |
|----------------------|--|
| <b>Title</b>         | Face Detection and People Counting Learning Model Study to Assess Attention Through Deep Learning Techniques |
| <b>Students</b>      | Miss Sirinya Thewakul Student ID 62050837  |
| <b>Degree</b>        | Bachelor of Science (Applied Statistics)   |
| <b>Department</b>    | Statistics   |
| <b>Faculty</b>       | Science  |
| <b>University</b>    | King Mongkut's Institute of Technology Ladkrabang (KMITL)  |
| <b>Academic Year</b> | 2022   |
| <b>Advisor</b>       | Dr. Sakuna Sriarnomai  |

### Abstract

The purposes of the research were to study and compare the performance of different technologies used to detect people, count the number of people, and assess individual interests. The researcher used the YOLOv7 model for detecting and counting people. The MediaPipe Face Mesh model is used for facial landmarks detection and measuring the angle between face and camera to assess attention. Two sets of test data were used: data set A1 and data set B1. Each data set included videos from both horizontal perspective and overhead perspective. The results showed that the overhead perspective had an average accuracy of 72.84%, while the horizontal perspective had an accuracy of 65.67%. The precision, recall, and f1-score of the overhead perspective were also higher than those of the horizontal perspective. Therefore, the YOLOv7 model was found to be better suited for detecting people in overhead perspective compared to horizontal perspective. The AUC value of 0.80 indicated that the model could predict and classify individuals to a good extent. Additionally, the model was used to detect and measure the angle between the face and the camera to assess attention. This was tested using data set A2 and data set B2, which were both in horizontal perspective with a face angle of 15 degrees. The accuracy for this test was found to be 55%, and the AUC value was 0.76, suggests that the model could predict to a good extent. The various performance measurements obtained in this research can be used as an approach for counting the number of people and detecting customer interest in small digital signage.

**Keywords :** YOLO, MediaPipe, Person detection, Facial Landmarks Detection, Digital

เอกสารนี้ Signage ที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากความกรุณา และความเมตตาจากอาจารย์ที่ปรึกษาสหกิจศึกษา ดร. สกุนดา ศรีอินมัย ที่ได้สละเวลาอันมีค่า ให้คำปรึกษา แนวคิด ข้อคิดเห็นต่างๆ และแนวทางในการวิจัย อันเป็นประโยชน์อย่างยิ่ง อีกทั้งยังช่วยแก้ปัญหาที่เกิดขึ้นระหว่างการทำดำเนินงาน ตลอดจนตรวจทานความถูกต้อง แก้ไขข้อบกพร่องต่างๆ จนสหกิจศึกษาฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยซาบซึ้งและกราบขอบคุณเป็นอย่างสูง ณ โอกาสนี้

ผู้วิจัยขอกราบขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.พรพิมล ชัยวุฒิศักดิ์ ที่ให้เกียรติเป็นประธานกรรมการสอบสหกิจฉบับนี้ กรุณาให้คำปรึกษา คำแนะนำ และชี้ให้เห็นถึงข้อบกพร่องต่างๆ จนทำให้สหกิจศึกษาฉบับนี้เสร็จสมบูรณ์มากยิ่งขึ้น

กราบขอบคุณ คุณวิศิษฐ์ กิจชัยนุกูล ที่ให้ความเมตตาในการสนับสนุนในด้านการศึกษา และในการทำงานวิจัยฉบับนี้ อีกทั้งสละเวลาในการตรวจสอบและให้คำปรึกษาเกี่ยวกับการทำงานตลอดการทำสหกิจศึกษา และต้องขอขอบคุณทางบริษัทที่ให้ความอนุเคราะห์สถานที่ และข้อมูลที่ใช้ในการศึกษาในการทำผลวิจัยของงานวิจัยฉบับนี้

กราบขอบคุณคณาจารย์ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่ได้มอบวิชาความรู้และประสบการณ์ในการทำวิจัยฉบับนี้ พร้อมทั้งให้คำแนะนำ คำปรึกษา และช่วยเหลือผู้ทำวิจัยในด้านต่างๆ ตั้งแต่เริ่มต้นจนจบในการทำวิจัย

และสุดท้ายนี้ขอกราบขอบคุณบิดา มารดา และครอบครัวที่ให้การสนับสนุนในด้านการศึกษาเล่าเรียน ตลอดจนคอยให้กำลังใจและให้ความช่วยเหลือเสมอมา รวมถึงขอบคุณเพื่อนนักศึกษาที่คอยให้กำลังใจและให้ความช่วยเหลือในการทำสหกิจศึกษาฉบับนี้

ศิริณญา เทวกุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

|  | หน้า     |
|--|----------|
| บทคัดย่อภาษาไทย.....   | ก        |
| บทคัดย่อภาษาอังกฤษ.....  | ข        |
| กิตติกรรมประกาศ .....  | ค        |
| สารบัญ.....  | ง        |
| สารบัญตาราง.....   | ช        |
| สารบัญรูป .....  | ฉ        |
| <b>บทที่ 1 บทนำ.....</b>   | <b>1</b> |
| 1.1 ความเป็นมาและความสำคัญของปัญหา.....                                      | 1        |
| 1.2 วัตถุประสงค์ของงานวิจัย .....  | 2        |
| 1.3 ขอบเขตของงานวิจัย .....  | 3        |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ .....  | 3        |
| 1.5 นิยามศัพท์เฉพาะ .....  | 4        |
| <b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....</b>                           | <b>7</b> |
| 2.1 แนวคิดและทฤษฎีเกี่ยวกับการตรวจจับ .....                                  | 7        |
| 2.1.1 การตรวจจับบุคคล (Person Detection) .....                               | 7        |
| 2.1.2 การตรวจจับใบหน้า (Face Detect).....                                    | 7        |
| 2.1.3 การตรวจจับจุดสังเกตบนใบหน้า (Facial Landmarks Detection) .....         | 8        |
| 2.2 แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองที่ใช้ในการตรวจจับบุคคล .....             | 8        |
| 2.2.1 โยโล่ (YOLO).....  | 8        |
| 2.2.2 โยโล่ เวอร์ชัน 7 (YOLOv7).....   | 9        |
| 2.3 แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองที่ใช้ในการตรวจจับจุดสังเกตบนใบหน้า ..... | 13       |
| 2.3.1 ไลบรารี มีเดียไปป์ (Library MediaPipe).....                            | 13       |
| 2.3.2 มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh).....                          | 15       |
| 2.3.3 การประมาณท่าศีรษะ(Head pose estimation).....                           | 18       |
| 2.4 ทฤษฎีของป้ายโฆษณาดิจิทัล (Digital Signage).....                          | 19       |
| 2.5 เทคโนโลยีและภาษาโปรแกรมที่เกี่ยวข้อง .....                               | 20       |
| 2.6 ทฤษฎีเกี่ยวกับการวัดประสิทธิภาพของแบบจำลอง .....                         | 21       |

เอกสารนี้เป็นเอกสารที่ 2.6.1 เมทริกซ์ความสับสน (Confusion Matrix) วนกลับไปหน้า 21 ให้ประโยชน์ด้านภาพ  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|   |           |
|---|-----------|
| 2.6.2 เมทริกซ์ความสับสนสำหรับการจำแนกประเภทหลายชั้น (Confusion Matrix for Multiclass Classification)..... | 23        |
| 2.6.3 Receiver Operating Characteristic Curve (ROC Curve).....  | 27        |
| 2.7 งานวิจัยที่เกี่ยวข้อง .....   | 28        |
| <b>บทที่ 3 วิธีการดำเนินงานวิจัย.....</b>   | <b>31</b> |
| 3.1 ประชากรและกลุ่มตัวอย่าง .....   | 31        |
| 3.1.1 ประชากรในการวิจัย.....  | 31        |
| 3.1.2 กลุ่มตัวอย่างที่ใช้ในการศึกษา .....   | 32        |
| 3.2 เครื่องมือที่ใช้ในงานวิจัย .....  | 33        |
| 3.2.1 การตรวจจับและนับจำนวนบุคคล .....  | 34        |
| 3.2.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 34        |
| 3.3 ขั้นตอนการเตรียมเครื่องมือในการตรวจจับ.....   | 34        |
| 3.3.1 ภาษาไพธอน 3.9.13.....   | 34        |
| 3.3.2 การตรวจจับและนับจำนวนบุคคล .....  | 34        |
| 3.3.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 34        |
| 3.4 ขั้นตอนการดำเนินงานวิจัย.....   | 35        |
| 3.4.1 การตรวจจับและนับจำนวนบุคคล .....  | 35        |
| 3.4.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 38        |
| 3.5 การเก็บรวบรวมข้อมูล .....   | 40        |
| 3.5.1 การตรวจจับและนับจำนวนบุคคล .....  | 40        |
| 3.5.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 41        |
| 3.6 ขั้นตอนการเตรียมข้อมูลและการวิเคราะห์ข้อมูล.....  | 42        |
| 3.6.1 การตรวจจับและนับจำนวนบุคคล .....  | 42        |
| 3.6.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 43        |
| 3.7 การทดสอบแบบจำลอง .....  | 46        |
| 3.7.1 การตรวจจับและนับจำนวนบุคคล .....  | 46        |
| 3.7.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 48        |
| 3.8 การประเมินประสิทธิภาพของแบบจำลอง .....  | 51        |
| 3.8.1 การตรวจจับและนับจำนวนบุคคล .....  | 51        |
| 3.8.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ .....                                 | 51        |
| <b>บทที่ 4 ผลการวิจัยและการอภิปรายผล .....</b>  | <b>52</b> |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ฉบับเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 4.1 ผลการทดสอบประสิทธิภาพของแบบจำลอง .....

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|   |           |
|---|-----------|
| 4.1.1 การทดสอบประสิทธิภาพของแบบจำลองโยโล เวอร์ชัน 7.....                          | 52        |
| 4.1.2 การทดสอบประสิทธิภาพของแบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh)..... | 60        |
| 4.2 อภิปรายผล.....  | 67        |
| <b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....</b>                                   | <b>68</b> |
| 5.1 สรุปผลการวิจัย .....  | 68        |
| 5.1.1 การตรวจจับและนับจำนวนบุคคล.....   | 68        |
| 5.1.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ....          | 69        |
| 5.2 ข้อจำกัด.....   | 70        |
| 5.3 ข้อเสนอแนะ.....   | 71        |
| เอกสารอ้างอิง.....  | 72        |
| ภาคผนวก.....  | 76        |
| ภาคผนวก ก.....  | 77        |



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

| ตารางที่  | หน้า |
|---|------|
| 2.1 เกณฑ์การตัดสินใจทั่วไปของค่าประมาณพื้นที่ใต้โค้ง AUC.....                                       | 28   |
| 3.1 จำนวนเฟรมที่ใช้ทดสอบในชุดข้อมูล A1.....   | 33   |
| 3.2 จำนวนเฟรมทั้งหมดและจำนวนเฟรมสุ่มขึ้นมาที่ใช้ทดสอบในชุดข้อมูล B1.....                            | 33   |
| 3.3 ดัชนีของจุดสังเกตบนใบหน้าที่เราสนใจ ซึ่งกระจายบนใบหน้าเท่า ๆ กัน.....                           | 45   |
| 4.1 เปรียบเทียบผลการวัดประสิทธิภาพในมุมมองแนวระนาบและแนวเหนือศีรษะ.....                             | 58   |
| 4.2 การใช้ทรัพยากรของเครื่องในขณะที่ทำการทดสอบแบบจำลอง YOLOv7.....                                  | 60   |
| 4.3 ค่าการประเมินประสิทธิภาพในการหันของใบหน้าด้วยมุม 10 15 และ 20 องศา.....                         | 65   |
| 4.4 ค่าการประเมินประสิทธิภาพในการตรวจจับและวัดองศาระหว่างใบหน้าที่กับกล้องเพื่อประเมินความสนใจ..... | 65   |
| ก.1 คำสั่งที่ใช้ในการแตกวิดีโอเป็นเฟรมภาพ.....  | 77   |
| ก.2 คำสั่งที่ใช้ในการสุ่มภาพขึ้นมาในทุกๆ 5 วินาที.....  | 77   |
| ก.3 คำสั่งที่ใช้ในการตรวจจับและนับจำนวนบุคคลที่เดินผ่าน.....  | 78   |
| ก.4 คำสั่งที่ใช้ในการตรวจจับและวัดองศาระหว่างใบหน้าที่กับกล้องเพื่อประเมินความสนใจ.....             | 85   |
| ก.5 คำสั่งที่ใช้ในการสร้างเมตริกซ์ความสับสนแบบ Multiclass.....                                      | 90   |
| ก.6 คำสั่งที่ใช้ในการสร้างกราฟ ROC แบบ Multiclass.....  | 90   |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป

| รูปที่  | หน้า |
|---|------|
| 2.1 สถาปัตยกรรมของโยโล่.....  | 8    |
| 2.2 โครงสร้างของโยโล่ที่เป็นชั้นแบบโครงข่ายประสาทเทียม.....   | 9    |
| 2.3 E-ELAN และงานก่อนหน้าเกี่ยวกับประสิทธิภาพของเลเยอร์ (Layer) สูงสุด.....                         | 11   |
| 2.4 กราฟเปรียบเทียบประสิทธิภาพโยโล่ เวอร์ชัน 7 (YOLOv7) กับเครื่องมืออื่นๆ.....                     | 11   |
| 2.5 สถาปัตยกรรมของโยโล่ เวอร์ชัน 7.....   | 12   |
| 2.6 เปรียบเทียบโยโล่ เวอร์ชัน 7 (YOLOv7) กับโมเดลพื้นฐานอื่นๆ.....                                  | 12   |
| 2.7 ชุดเครื่องมือมีเดียไปป์ (MediaPipe).....  | 14   |
| 2.8 มีเดียไปป์ เฟซ เมช 468 จุด.....   | 16   |
| 2.9 ตัวอย่างภาพที่ตรวจจับ 468 จุด.....  | 17   |
| 2.10 ตัวอย่างภาพที่ตรวจจับ 6 จุด.....   | 17   |
| 2.11 ลักษณะทิศทางของศีรษะในมุม Pitch Yaw Roll.....  | 18   |
| 2.12 ตัวอย่างผลลัพธ์การหันซ้าย-ขวา และมองตรง.....   | 19   |
| 2.13 เมทริกซ์ความสับสน.....   | 22   |
| 2.14 ตัวอย่างการกำหนด FP ของคลาสที่สนใจ.....  | 23   |
| 2.15 ตัวอย่างการกำหนด FN ของคลาสที่สนใจ(ต่อ).....   | 24   |
| 2.16 ตัวอย่างการกำหนด TP.....   | 24   |
| 2.17 ตัวอย่างการกำหนด TN ของคลาสที่สนใจ(ต่อ).....   | 25   |
| 2.18 ตัวอย่างการกำหนดค่าที่นำไปคิด Precision ของคลาสที่สนใจ(ต่อ).....                               | 25   |
| 2.19 ตัวอย่างการกำหนดค่าที่นำไปคิด Recall ของคลาสที่สนใจ(ต่อ).....                                  | 26   |
| 2.20 ตัวอย่างการแสดงค่าในแต่ละคลาส และค่าเฉลี่ยรวม.....   | 26   |
| 2.21 ตัวอย่างค่า AUC ในแต่ละกรณี.....   | 27   |
| 3.1 ตัวอย่าง Anaconda Navigator.....  | 35   |
| 3.2 ขั้นตอนการดำเนินงานวิจัยหัวข้อการตรวจจับและนับจำนวนบุคคลที่เดินผ่าน.....                        | 37   |
| 3.3 ขั้นตอนการดำเนินงานวิจัยหัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้อง<br>เพื่อดูความสนใจ..... | 39   |
| 3.4 ตัวอย่างเฟรมที่ถ่ายในมุมมองแนวเหนือศีรษะ.....   | 40   |
| 3.5 ตัวอย่างเฟรมที่ถ่ายในมุมมองแนวระนาบ.....  | 41   |
| 3.6 ตัวอย่างป้ายโฆษณาดิจิทัลขนาดเล็ก.....   | 41   |
| 3.7 แสดงการเคลื่อนไหวในวินาทีที่ 1.....   | 42   |
| 3.8 แสดงการเคลื่อนไหวในวินาทีที่ 2.....   | 42   |

## สารบัญรูป (ต่อ)

| รูปที่  | หน้า |
|---|------|
| 3.9 แสดงการเคลื่อนไหวในวินาทีที่ 3.....                                   | 42   |
| 3.10 แสดงการเคลื่อนไหวในวินาทีที่ 4.....                                  | 42   |
| 3.11 แสดงการเคลื่อนไหวในวินาทีที่ 5.....                                  | 42   |
| 3.12 แสดงไลบรารี (Library) นำเข้า และกำหนดพารามิเตอร์.....                | 43   |
| 3.13 เลือก มีเดียไปป์ เฟซ เมช 6 จุด.....                                  | 44   |
| 3.14 แสดงการกำหนดจุดสังเกตบนใบหน้าในโปรแกรม.....                          | 45   |
| 3.15 แสดงการกำหนด p1.....   | 45   |
| 3.16 แสดงการสร้าง list สำหรับตรวจจับใบหน้า.....                           | 45   |
| 3.17 แสดงการกำหนดตัวแปรที่เกี่ยวข้อง.....                                 | 45   |
| 3.18 ตัวอย่างเฟรมที่ตรวจจับด้วย YOLOv7 แบบมุมมองแนวเหนือศีรษะ.....        | 46   |
| 3.19 ตัวอย่างเฟรมที่ตรวจจับด้วย YOLOv7 แบบมุมมองแนวระนาบ.....             | 47   |
| 3.20 ตัวอย่างคำสั่งในการการตรวจจับและนับจำนวนบุคคล.....                   | 47   |
| 3.21 ตัวอย่างผลการทดสอบการตรวจจับและนับจำนวนคน.....                       | 47   |
| 3.22 แสดงการลบ frame count ที่ต่ำกว่า 150 frame count.....                | 48   |
| 3.23 แสดงการเพิ่มใบหน้าที่ยังไม่เคยตรวจพบมาก่อน.....                      | 48   |
| 3.24 แสดงการติดตามใบหน้าด้วยฟังก์ชัน math.dist().....                     | 49   |
| 3.25 แสดงการกำหนดองค์ประกอบในการหันของใบหน้า.....                         | 50   |
| 4.1 เมตริกซ์ความสับสนของชุดข้อมูล A1 ในมุมมองแนวระนาบ.....                | 53   |
| 4.2 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวระนาบ.....      | 53   |
| 4.3 เมตริกซ์ความสับสนของชุดข้อมูล B1 ในมุมมองแนวระนาบ.....                | 54   |
| 4.4 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวระนาบ.....      | 55   |
| 4.5 เมตริกซ์ความสับสนของชุดข้อมูล A1 ในมุมมองแนวเหนือศีรษะ.....           | 56   |
| 4.6 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวเหนือศีรษะ..... | 56   |
| 4.7 เมตริกซ์ความสับสนของชุดข้อมูล B1 ในมุมมองแนวเหนือศีรษะ.....           | 57   |
| 4.8 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวเหนือศีรษะ..... | 57   |
| 4.9 กราฟ ROC แสดงค่า AUC โดยรวมในมุมมองแนวระนาบ.....                      | 59   |
| 4.10 กราฟ ROC แสดงค่า AUC โดยรวมในมุมมองแนวเหนือศีรษะ.....                | 59   |
| 4.11 เมตริกซ์ความสับสนของชุดข้อมูล A2.....                                | 61   |
| 4.12 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A2.....                      | 61   |
| 4.13 เมตริกซ์ความสับสนของชุดข้อมูล B2 (10 องค์).....                      | 62   |

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป (ต่อ)

| รูปที่   | หน้า |
|--|------|
| 4.14 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (10 องศา)..... | 62   |
| 4.15 เมทริกซ์ความสับสนของชุดข้อมูล B2 (15 องศา).....           | 63   |
| 4.16 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (15 องศา)..... | 63   |
| 4.17 เมทริกซ์ความสับสนของชุดข้อมูล B2 (20 องศา).....           | 64   |
| 4.18 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (20 องศา)..... | 64   |
| 4.19 ตัวอย่างการหันใบหน้าด้วยมุม 15 องศา.....                  | 66   |
| 4.20 กราฟ ROC แสดงค่า AUC โดยรวมของแบบจำลอง.....               | 67   |



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการโฆษณาามีหลากหลายรูปแบบ เช่น ป้ายโฆษณาไวเนล ป้ายโฆษณาดิจิทัล ใบบลิว โทรทัศน์ และสื่อสังคมออนไลน์ ป้ายโฆษณาเป็นสื่อที่มีความสำคัญมากในการประชาสัมพันธ์ ข้อดีของป้ายโฆษณาแบบดิจิทัล สามารถรองรับสื่อได้หลากหลายประเภท ทั้งข้อความ ภาพกราฟิก วิดีโอ และเนื้อหาอื่น ๆ สามารถปรับแต่งป้ายโฆษณาดิจิทัลได้เกือบทุกแง่มุม ตั้งแต่ข้อความที่ใช้ไปจนถึงรูปภาพ แต่ก็มีข้อเสียคือ การนำไปติดตั้งในตำแหน่งที่ไม่เหมาะสมจะทำให้ได้รับความสนใจน้อย ซึ่งจะมีผลต่อยอดขาย เป็นต้น จากผลการวิจัยพบว่าป้ายดิจิทัลได้รับการพิสูจน์แล้วว่าสามารถนำไปใช้ได้หลากหลาย และมีส่วนช่วยอย่างมากในการเพิ่มยอดขายและปรับปรุงภาพลักษณ์ของบริษัทและแบรนด์สินค้า (Kim et al., 2013)

การเปลี่ยนแปลงวิถีชีวิตในยุคดิจิทัลของมนุษย์นั้น เริ่มขึ้นในช่วงปลายศตวรรษที่ 20 ซึ่งเทคโนโลยีมีการพัฒนาอย่างรวดเร็ว อย่างไรก็ตามการก้าวหน้าด้านเทคโนโลยีของการสื่อสารจะเป็นสิ่งสนับสนุนให้การโฆษณาสินค้าหรือบริการมีความน่าสนใจมากยิ่งขึ้น (Bleier and Eisenbeiss, 2015) ซึ่งการโฆษณามีผลต่อธุรกิจต่าง ๆ อาทิ การเพิ่มยอดขาย ช่วยเพิ่มโอกาสในการขายสินค้าและบริการ ทำให้ผู้บริโภครู้จักสินค้าและบริการของธุรกิจเรามากขึ้น เกิดความคิดเห็นที่ดีกับสินค้าหรือบริการของเรา ซึ่งจะเป็นปัจจัยสำคัญในการตัดสินใจซื้อสินค้าและบริการของลูกค้า มีผลต่อการสร้างความรู้สึกต่อสินค้าและบริการ การโฆษณาช่วยสร้างความรู้สึกและความต้องการต่อสินค้าและบริการของธุรกิจเรา โดยการนำเสนอสิ่งที่น่าสนใจและดึงดูดใจลูกค้า ทำให้ลูกค้าเกิดความสนใจและติดตามสินค้าและบริการของเราได้ง่ายขึ้น มีผลต่อการสร้างความน่าเชื่อถือ การโฆษณาช่วยสร้างความน่าเชื่อถือในตลาดและยกระดับภาพลักษณ์ของธุรกิจ โดยการนำเสนอผลิตภัณฑ์และบริการที่มีคุณภาพและมีความน่าเชื่อถือจะช่วยสร้างความไว้วางใจ ซึ่งเป็นส่วนหนึ่งในการตัดสินใจให้กับลูกค้าในการเลือกซื้อสินค้าและบริการของเรา และมีผลต่อการช่วยเปิดตลาดใหม่ การโฆษณาช่วยเปิดตลาดใหม่สำหรับธุรกิจ โดยการเน้นกลุ่มลูกค้าที่ไม่เคยเข้าถึงมาก่อนหรือสร้างความต้องการใหม่ ๆ ในตลาดซึ่งช่วยเพิ่มโอกาสในการขายสินค้าและบริการให้กับธุรกิจของเราได้มากขึ้น ดังนั้น การโฆษณาช่วยให้ธุรกิจของเรามีการแข่งขันได้ดีกว่าธุรกิจอื่นในตลาด เนื่องจากลูกค้าจะรู้จักและมีความคิดเห็นต่อธุรกิจของเรามากขึ้น และมีโอกาสที่จะเลือกซื้อสินค้าและบริการของเราแทนการใช้งานสินค้าและบริการของคู่แข่ง โดยการโฆษณาช่วยให้ลูกค้ารับรู้ข้อมูลคุณลักษณะของสินค้านั้นได้ ดังนั้น การโฆษณาส่วนมากมักจะอาศัยภาพถ่ายหรือวิดีโอเพื่อประกอบการโฆษณา เนื่องจากสามารถสื่อสารให้คนดูเข้าใจและรับรู้สินค้าได้ดียิ่งขึ้น (จารุณี และคณะ, 2565)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปัจจุบันเทคโนโลยีได้เข้ามามีบทบาทในชีวิตประจำวันเป็นอย่างมาก รวมถึงสิ่งอำนวยความสะดวกและการใช้เทคโนโลยีในการวิเคราะห์ด้านต่าง ๆ เช่น วิเคราะห์เกี่ยวกับการตรวจจับบุคคล การตรวจจับใบหน้า การตรวจจับจุดสังเกตบนใบหน้า และการตรวจจับท่าทางการเคลื่อนไหว โดยการใช้เทคโนโลยีเข้ามาเกี่ยวข้องในการวิเคราะห์ คือ การตรวจจับกลุ่มผู้คนที่เดินผ่าน หรือการตรวจจับบุคคล (Person Detection) เป็นกระบวนการที่ใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) โดยใช้คอมพิวเตอร์ในการวิเคราะห์ภาพ เพื่อตรวจจับและแยกแยะว่าภาพนั้นมีมนุษย์อยู่หรือไม่มีมนุษย์อยู่ ซึ่งเป็นเทคโนโลยีที่ใช้ระบบคอมพิวเตอร์และเซ็นเซอร์ในการปรากฏของบุคคลในภาพหรือวิดีโอ เป็นหนึ่งในหัวข้อสำคัญในการวิเคราะห์ภาพดิจิทัล เพื่อให้สามารถระบุและติดตามการเคลื่อนไหวหรือติดตามพฤติกรรมของมนุษย์ได้

โดยการทำป้ายโฆษณาดิจิทัลเกิดปัญหาต่อทางธุรกิจของบริษัทผลิตเครื่องดื่มแห่งหนึ่ง คือ ความไม่แน่นอนในผลตอบรับ ซึ่งการทำป้ายโฆษณาไม่ได้รับรองผลตอบรับที่แน่นอน และอาจทำให้ธุรกิจสูญเสียเวลา และค่าใช้จ่ายได้ เนื่องจากในการจัดทำป้ายโฆษณาย่อมมีค่าใช้จ่ายในการติดตั้ง เช่น ในการติดตั้งป้ายโฆษณาดิจิทัลตามร้านสะดวกซื้อ ด้วยเหตุนี้ จึงควรเลือกติดตั้งป้ายโฆษณาดิจิทัลในที่ที่มีคนเดินผ่านเป็นจำนวนมาก และควรติดตั้งในมุมมองที่ผู้คนสนใจเป็นจำนวนมากด้วยเช่นกัน ซึ่งเราจำเป็นต้องทราบจำนวนผู้คนที่เดินผ่าน ในงานวิจัยนี้ทางผู้จัดทำจึงสนใจที่จะทำหัวข้อดังต่อไปนี้ คือ 1. การตรวจจับและนับจำนวนบุคคล 2. การตรวจจับและวัดองศาระหว่างใบหน้าและกล้อง เพื่อประเมินความสนใจ โดยอัลกอริทึม (Algorithm) ที่จะนำมาใช้ในการตรวจจับบุคคล (Person Detection) คือ โยโล เวอร์ชัน 7 (YOLOv7) ซึ่งเป็นเทคโนโลยีการตรวจจับวัตถุที่ใช้การเรียนรู้เชิงลึก (Deep Learning) เพื่อตรวจจับบุคคลในภาพหรือวิดีโอ โดยโยโล เวอร์ชัน 7 สามารถตรวจจับและประมวลผลได้อย่างรวดเร็ว และมีความแม่นยำสูง โดยแบบจำลองจะเรียนรู้และประมวลผลภาพทั้งหมดในกรอบเดียวกันและมีการใช้แบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ซึ่งเป็นหนึ่งในแบบจำลองที่อยู่ในไลบรารีมีเดียไปป์ (Library MediaPipe) แบบจำลองใช้ในการตรวจจับจุดสังเกตบนใบหน้า (Facial Landmarks Detection) เพื่อนำไปใช้ในการตรวจจับความสนใจของผู้คนที่เดินผ่านในบริเวณที่กล้องบันทึก

## 1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อศึกษาเทคโนโลยีที่ใช้ในการตรวจจับและนับจำนวนบุคคลกล้องที่บันทึก (Person Detection) ด้วยแบบจำลองโยโล เวอร์ชัน 7 (YOLOv7)
  - 2) เพื่อเปรียบเทียบประสิทธิภาพการตรวจจับและนับจำนวนบุคคลด้วยแบบจำลองโยโล เวอร์ชัน 7 (YOLOv7) ในมุมมองแนวระนาบและมุมมองแนวเหนือศีรษะ
  - 3) เพื่อศึกษาและวิเคราะห์ลักษณะการหันใบหน้าของแต่ละบุคคลโดยวัดองศาจากการหัน เพื่อ
- เอกสารนี้ ประเมินและนับจำนวนคนที่มีความสนใจต่อกล้อง ด้วยแบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe ไม่ว่าการ Face Mesh) อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 1.3 ขอบเขตของงานวิจัย

#### 1) ขอบเขตด้านข้อมูล

##### 1.1) ชุดข้อมูลทดสอบในการตรวจจับบุคคล

เป็นวิดีโอในมุมมองที่สนใจ ซึ่งแต่ละชุดข้อมูลจะประกอบด้วย มุมมองด้านบนเหนือศีรษะ (คล้ายกับกล้องวงจรปิด) และมุมมองในแนวระนาบ ข้อมูลนี้เป็นข้อมูลที่ทำการบินที่วิดีโอด้วยตนเอง โดยแบ่งเป็น 2 ชุด คือ ชุดข้อมูล A1 และชุดข้อมูล B1 ในชุดข้อมูล A1 มีจำนวน 150 เฟรม/มุมมอง และชุดข้อมูล B1 มีจำนวน 180 เฟรม/มุมมอง

##### 1.2) ชุดข้อมูลทดสอบในการตรวจจับความสนใจ

เป็นวิดีโอในมุมมองแนวระนาบ จะแบ่งข้อมูลเป็น 2 ชุด คือ ชุดข้อมูล A2 และชุดข้อมูล B2 ในชุดข้อมูล A1 เป็นข้อมูลที่ทำการบินที่ด้วยตนเอง และชุดข้อมูล B2 เป็นชุดที่เก็บข้อมูล โดยทำการบันทึกคลิกวิดีโอด้วยตนเอง 8 วิดีโอ และนำข้อมูลมาจากอินเทอร์เน็ต 2 วิดีโอ ซึ่งแต่ละชุดข้อมูลประกอบด้วยวิดีโอจำนวน 10 วิดีโอ

#### 2) ขอบเขตด้านมุมมองที่พิจารณา

ในการตรวจจับและนับจำนวนบุคคลจะพิจารณามุมมองแนวระนาบกับมุมมองแนวเหนือศีรษะและนำมาเปรียบเทียบกัน แต่ในการตรวจจับความสนใจจะพิจารณาเพียงมุมมองแนวระนาบเท่านั้น

#### 3) ขอบเขตด้านระยะในการตรวจจับ

ในการตรวจจับและนับจำนวนบุคคลจะพิจารณาคนที่เดินอยู่ในวิดีโอในระยะที่ห่างจากกล้องประมาณ 0.5 – 10 เมตร ส่วนในการตรวจจับความสนใจจะพิจารณาในระยะที่แบบจำลองสามารถตรวจจับได้คือประมาณ 1 - 2 เมตร

#### 4) ขอบเขตด้านป้ายโฆษณาดิจิทัล

เป็นป้ายโฆษณาดิจิทัลขนาดเล็กที่อยู่บนชั้นวางของตามร้านค้าปลีกหรือร้านสะดวกซื้อ โดยมีขนาดจอยู่ที่ประมาณ 7-10 นิ้ว

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) สามารถนำแบบจำลองที่ได้มาตรวจนับจำนวนบุคคลที่เดินผ่าน และสามารถนำไปปรับใช้ในงานที่เกี่ยวข้องได้

2) สามารถนำแบบจำลองที่ได้ไปวิเคราะห์เพื่อประเมินความสนใจของผู้ที่เดินกล้อง และเป็นต้นแบบในการปรับปรุงและพัฒนาเพื่อใช้งานต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.5 นิยามศัพท์เฉพาะ

- 1) **การเรียนรู้เชิงลึก (Deep Learning)** คือการพัฒนาเทคโนโลยีคอมพิวเตอร์ให้สามารถเลียนแบบการทำงานของมนุษย์ เป็นเทคนิคการเรียนรู้ของเครื่อง (Machine Learning : ML) ที่สามารถสร้างแบบจำลองประมวลผลข้อมูลขนาดใหญ่และซับซ้อนได้ โดยใช้โครงข่ายประสาทเทียม (Neural Networks) ซึ่งจะประกอบด้วยชั้นเลเยอร์ (Layer) ต่าง ๆ ที่มีการเชื่อมต่อกันเพื่อประมวลผลข้อมูล การเรียนรู้เชิงลึกนี้มีความสามารถในการเรียนรู้และจำแนกข้อมูลได้ดีกว่าเทคนิคการเรียนรู้อื่นๆ เนื่องจากสามารถสกัดลักษณะที่ซับซ้อนและเชิงลึกของข้อมูลได้ และยังสามารถจำแนกวัตถุหรือภาพที่ซับซ้อนได้อย่างมีประสิทธิภาพ และยังสามารถนำไปประยุกต์ใช้ในหลายๆ ด้าน เช่น การจำแนกภาพ (Image Classification) การตรวจจับวัตถุ (Object Detection) การแปลภาษา (Machine Translation) การจัดหมวดหมู่ของข้อมูล (Data Categorization) และอื่นๆ อีกมากมาย (นพรุจ และณัฐภูมิ, 2563)
- 2) **โยโล (You Only Look Once หรือ YOLO)** เป็นแบบจำลองการเรียนรู้เชิงลึก (Deep Learning) ที่นิยมใช้ในการตรวจจับวัตถุในภาพ โดยมีแนวคิดในการออกแบบที่แตกต่างกับแบบจำลองอื่นๆ ซึ่งเน้นความเร็วและความแม่นยำในการตรวจจับ โดยใช้หลักการของการเรียนรู้เชิงลึก ในการเรียนรู้ลักษณะของวัตถุและพิกัดตำแหน่งของวัตถุบนภาพได้ โดยรวดเร็วและแม่นยำ เมื่อแบบจำลองพบวัตถุ จะทำการทำนายประเภทของวัตถุ และค่าความน่าจะเป็นว่าวัตถุนั้นๆ อยู่ในแต่ละช่องเท่าไร โดยประโยชน์ของโยโล (YOLO) คือสามารถใช้ในการตรวจจับวัตถุในเวลาเร็วและสามารถทำงานได้บนอุปกรณ์ที่มีความจำกัดด้านความสามารถ เช่น อุปกรณ์เคลื่อนที่ หรือระบบสมองประสาทเทียม (Neural Network) ที่มีขนาดเล็ก (Wang et al., 2022)
- 3) **คลาส (Class)** เป็นตัวกำหนดคุณสมบัติที่ถูกสร้างขึ้น ซึ่งสามารถใช้ในการแยกประเภทของวัตถุได้ โดยในงานวิจัยฉบับนี้คลาสหมายถึง จำนวนคนที่อยู่ในเฟรม เช่น มีคนอยู่ในเฟรม 2 คนคือเป็นคลาส 2 ในค่าจริง หรือแบบจำลองสามารถทำนายได้ 2 คน คือเป็นคลาส 2 ในค่าพยากรณ์ เป็นต้น (ไพชยนต์, 2557)
- 4) **การตรวจจับวัตถุ (Object Detection)** คือ กระบวนการที่จับตัวแบบวัตถุที่อยู่ในภาพหรือวิดีโอ โดยมีการหาพิกัดของวัตถุ (Bounding Box) และแปลงเป็นประเภทของวัตถุ เช่น รถยนต์ คน หรือสัตว์ โดยใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) ซึ่งมีโครงข่ายประสาทเทียมแบบสังวัตนาการ (CNN) เป็นอัลกอริทึม (Algorithm) หลักในการทำงาน การตรวจจับวัตถุมีการประยุกต์ใช้งานกับหลายสาขาอุตสาหกรรม เช่น รถยนต์ไร้คนขับ (Autonomous Vehicles) การวิเคราะห์ภาพทางการแพทย์ การตรวจสอบความปลอดภัยในระบบกล้องวงจรปิด และการตรวจจับวัตถุในการทำงานของหุ่นยนต์ เป็นต้น (ธนะพงษ์, 2566) ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี หากมีการนำเอกสารนี้ไปเผยแพร่โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย

- 5) **การตรวจจับบุคคล (Person Detection)** เป็นส่วนหนึ่งของการตรวจจับวัตถุ (Object detection) โดยทั่วไปแล้วการตรวจจับวัตถุจะมีหลายคลาส (Class) ในสิ่งที่ต้องการตรวจจับ เช่น มนุษย์ รถยนต์ จักรยาน สัตว์ เป็นต้น แต่ในกรณีการตรวจจับบุคคล จะเน้นไปที่การตรวจจับและติดตามบุคคลที่เคลื่อนไหวโดยเฉพาะ มักใช้แบบจำลองโครงข่ายประสาทเทียมแบบสังวัตนาการ (CNN) เนื่องจาก CNN มีความสามารถในการดูและวิเคราะห์ภาพได้ดี โดยการตรวจจับบุคคลนั้นจะมีเครื่องมือที่ใช้กันอย่างแพร่หลายคือ โยโล (YOLO) โดยใช้การจำแนกและตรวจจับวัตถุในภาพโดยรวม และหาตำแหน่งของบุคคล โดยสามารถทำได้ด้วยความแม่นยำสูงและความเร็วในการประมวลผลภาพได้อย่างมีประสิทธิภาพ (Huang and Lin, 2020)
- 6) **ชุดข้อมูลฝึกฝน (Training Dataset)** คือชุดข้อมูลที่ใช้สำหรับการสอนแบบจำลองเรียนรู้เชิงลึก เพื่อให้แบบจำลองเรียนรู้เข้าใจและสามารถทำนายผลลัพธ์ได้ถูกต้องตามที่ต้องการ ซึ่งชุดข้อมูลฝึกฝนนั้นเป็นข้อมูลที่มีการประมวลผลและจัดกลุ่มให้เหมาะสมต่อการใช้งาน โดยส่วนใหญ่จะมาจากการรวบรวมข้อมูลจากหลายแหล่ง (ปีณณา, 2563)
- 7) **ชุดข้อมูลทดสอบ (Testing Dataset)** หมายถึงชุดข้อมูลที่ใช้สำหรับการทดสอบ (Testing) แบบจำลอง (Model) หรือโปรแกรมคอมพิวเตอร์ เพื่อวัดประสิทธิภาพและความแม่นยำของแบบจำลองหรือโปรแกรมกับข้อมูลที่ไม่เคยเห็น (Unseen Data) ก่อนนำแบบจำลองหรือโปรแกรมไปใช้งานจริง โดยจะมีข้อมูลในชุดข้อมูลทดสอบ (Testing Dataset) แยกจากชุดข้อมูลสำหรับการฝึก (Training Dataset) ซึ่งแบบจำลองได้เรียนรู้และปรับปรุงตนเอง ซึ่งชุดข้อมูลทดสอบ จะไม่มีอยู่ในชุดข้อมูลสำหรับการฝึกแบบจำลอง เพื่อให้เป็นการทดสอบแบบจำลองโดยที่ไม่มีความลำเอียงใดๆ ที่เกิดจากการฝึกแบบจำลองด้วยชุดข้อมูลเดียวกัน (ปีณณา, 2563)
- 8) **อัลกอริทึม (Algorithm)** หมายถึง ขั้นตอนวิธีการแก้ไขปัญหาหรือการทำงานที่เรียงลำดับกันเป็นขั้นตอน โดยอาจประกอบไปด้วยขั้นตอนต่างๆ เช่น การรับข้อมูลเข้า (Input) , การประมวลผล (Processing), และการส่งผลลัพธ์ออก (Output) เพื่อให้ผู้ใช้งานเข้าใจและทำงานตามขั้นตอนได้ง่ายขึ้น โดยอัลกอริทึม (Algorithm) เป็นส่วนสำคัญของการเขียนโปรแกรมคอมพิวเตอร์ เนื่องจากโปรแกรมคอมพิวเตอร์ต้องใช้อัลกอริทึม (Algorithm) เพื่อทำงานตามเป้าหมายที่ต้องการ และต้องผ่านการทดสอบและปรับปรุงอัลกอริทึม (Algorithm) เพื่อให้โปรแกรมทำงานได้อย่างถูกต้องและมีประสิทธิภาพสูงสุด (ปีณณา, 2563)
- 9) **มีเดียไปป์ (MediaPipe)** เป็นเครื่องมือและไลบรารี (Library) ที่พัฒนาโดยกูเกิ้ล (Google) ซึ่งมีประโยชน์ในการประมวลผลภาพและวิดีโอ โดยมีการรวมเทคโนโลยีต่างๆ เช่น การเรียนรู้ของเครื่อง และเทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) เข้าด้วยกัน

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์ที่จัดทำขึ้นเพื่อการศึกษาเท่านั้น เมื่อผู้จัดทำเห็นสมควรจะปรับปรุงแก้ไขเนื้อหา  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อให้ผู้ใช้สามารถสร้างแอปพลิเคชันที่มีการประมวลผลภาพและวิดีโอได้อย่างง่ายดาย (Bora et al, 2023)

- 10) **มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh)** เป็นหนึ่งในแบบจำลองที่ถูกพัฒนาโดยมีเดียไปป์และได้ถูกฝึกฝนด้วยข้อมูลแบบจำลองใบหน้า 3 มิติ และ จุดสังเกตบนใบหน้า 2 มิติ โดยใช้ข้อมูลอบรมมากกว่า 33,000 รูปภาพของใบหน้าผู้ชายและผู้หญิง และทำการสกัดจุดสังเกตบนใบหน้าในแต่ละใบหน้าจำนวน 468 จุด โดยมีการปรับปรุงแบบจำลองด้วยการใช้การขยายข้อมูล (Data Augmentation) ซึ่งเป็นการเพิ่มจำนวนข้อมูลให้มีจำนวนมากขึ้น เพื่อเพิ่มประสิทธิภาพในการทำนาย (Muhi et al, 2022)
- 11) **การตรวจจับใบหน้า (Face Detect)** เป็นกระบวนการค้นหาและตรวจสอบว่าภาพหรือวิดีโอที่มีใบหน้าของมนุษย์หรือไม่ โดยการตรวจจับใบหน้าจะต้องเรียกใช้เทคโนโลยีการแยกแยะวัตถุ และการเรียนรู้ของเครื่อง ที่ได้รับการฝึกสอนด้วยข้อมูลภาพของใบหน้ามนุษย์จริง ซึ่งการตรวจจับใบหน้าเป็นขั้นตอนแรกที่ต้องผ่านเพื่อให้สามารถประมวลผลภาพและวิดีโอที่เกี่ยวข้องได้อย่างเหมาะสมต่อไปในขั้นต่อไป (Zhang et al, 2019)
- 12) **การตรวจจับจุดสังเกตบนใบหน้า (Facial Landmarks Detection)** เป็นกระบวนการที่ใช้แบบจำลองและเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เพื่อตรวจจับและประมวลผลแต่ละจุดบนใบหน้าของมนุษย์ โดยมีจุดบนใบหน้าเหล่านี้เรียกว่าจุดสังเกต โดยแต่ละจุดสังเกตจะมีตำแหน่งของพิกัดบนภาพ (x, y) ที่ตรงกับจุดบนใบหน้าของมนุษย์ เช่น จุดบนปลายคิ้ว จุดบนปลายจมูก (Radmehr et al, 2021)
- 13) **ป้ายโฆษณาดิจิทัล (Digital Signage)** หมายถึง สื่อประชาสัมพันธ์ที่มีทั้งภาพและเสียง แสดงข่าวสาร การบริการและกิจกรรมต่างๆ ผ่านหน้าจอคอมพิวเตอร์ในระบบดิจิทัลและประมวลผลด้วยระบบคอมพิวเตอร์ (Cahyadi et al, 2012)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยมีความสนใจในการตรวจจับบุคคลและตรวจจับความสนใจของผู้คน โดยเครื่องมือที่นำมาศึกษาคือ โยโล่ (YOLO หรือ You Only Look Once) และไลบรารี มีเดียไปป์ (Library MediaPipe) ซึ่งใช้แบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ในการตรวจจับจุดสังเกตบนใบหน้า รายละเอียด ทฤษฎี และงานวิจัยที่เกี่ยวข้องของแต่ละวิธีแสดงดังต่อไปนี้

#### 2.1 แนวคิดและทฤษฎีเกี่ยวกับการตรวจจับ

การตรวจจับสามารถตรวจจับได้หลากหลายวิธี ตามวัตถุประสงค์ของแต่ละบุคคล เช่น การตรวจจับบุคคล การตรวจจับใบหน้า การตรวจจับจุดสังเกตบนใบหน้า รายละเอียดดังนี้

##### 2.1.1 การตรวจจับบุคคล (Person Detection)

หมายถึง การตรวจหาคนในภาพหรือวิดีโอ โดยทั่วไปจะใช้หลักการการตรวจจับวัตถุ (Object detection) ในการหาพื้นที่ที่มีคนปรากฏอยู่ในภาพ หรือใช้วิธีการตรวจจับคนจากลักษณะรูปร่างของร่างกาย (Body Detection) แล้วนำมาวิเคราะห์เพื่อตรวจสอบว่ามีคนหรือไม่ในภาพ เมื่อพบบุคคลในภาพ จะทำการสร้างกรอบรอบตัวเพื่อบ่งบอกว่า ในกรอบนั้นคือบุคคลที่ตรวจจับได้ มักใช้เทคโนโลยีการประมวลผลภาพและการเรียนรู้เชิงลึก (Deep Learning) ใช้การค้นหาตำแหน่งและขนาดของบุคคลในภาพหรือวิดีโอ ซึ่งมักจะใช้งานในการตรวจจับมนุษย์เพื่อใช้ในการประเมินผู้สอนระบบรักษาความปลอดภัย การตรวจจับการกระทำของมนุษย์ และงานอื่น ๆ เพื่อจำแนกในภาพมีคนอยู่หรือไม่ โดยส่วนใหญ่จะใช้แบบจำลองซึ่งถูกฝึกสอนด้วยข้อมูลภาพของคนและไม่มีกรจำกัดจำนวนคนที่ต้องการตรวจจับ (Huang and Lin, 2020)

##### 2.1.2 การตรวจจับใบหน้า (Face Detect)

การตรวจจับใบหน้าเป็นปัญหาพื้นฐานและสำคัญต่อการมองเห็นของคอมพิวเตอร์ (Computer Vision) ซึ่งหมายถึงการตรวจจับหน้าในภาพหรือวิดีโอ โดยใช้วิธีการตรวจจับวัตถุ (Object Detection) หรือการตรวจจับวัตถุโดยใช้วิธีการตรวจจับวัตถุโดยใช้แบบจำลองสำหรับตรวจจับหน้า (Face Detection Model) เพื่อหาพื้นที่ที่มีหน้าปรากฏอยู่ในภาพ โดยแบบจำลองสำหรับตรวจจับหน้ามักจะใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) เพื่อให้มีความแม่นยำสูงในการตรวจจับหน้า ในบางกรณีการตรวจจับใบหน้า อาจนำไปใช้ในการระบุตัวตนของบุคคลด้วยการรู้จำใบหน้า (Face Recognition) ซึ่งเป็นกระบวนการระบุตัวตนจากใบหน้าของบุคคลโดยใช้การเปรียบเทียบลักษณะหรือคุณลักษณะทางชีวภาพของใบหน้ากับฐานข้อมูลที่มีอยู่ (Zhang et al, เอกสารนี้ 2019) สารที่ส่งวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.1.3 การตรวจจับจุดสังเกตบนใบหน้า (Facial Landmarks Detection)

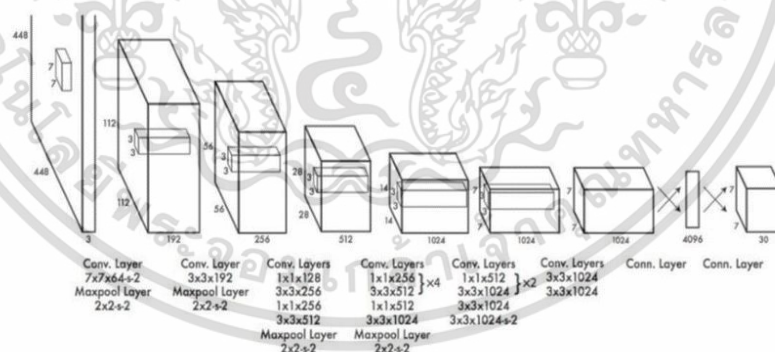
ถูกพัฒนาขึ้นมาเพื่อใช้ในการตรวจจับจุดสังเกตบนใบหน้า เช่น ปลายแหลมของตา ปลายจมูก ปาก ซึ่งจุดเหล่านี้เป็นจุดที่มีความสำคัญสำหรับการวิเคราะห์และสกัดลักษณะบนใบหน้า เช่น การวัดความยาวและความกว้างของตา การวัดระยะห่างระหว่างจุดบนใบหน้า เพื่อใช้ในการสกัดลักษณะบนใบหน้า เช่น การหาความแม่นยำในการวัดความสูงของจมูก ระยะห่างระหว่างขอบปาก และจมูก ซึ่งจะต่างจากการตรวจจับใบหน้าที่ตรวจจับแค่ขอบเขตของใบหน้าเท่านั้น (Radmehr et al, 2021)

ดังนั้น จุดแตกต่างระหว่างการตรวจจับบุคคล กับ การตรวจจับใบหน้า คือ การตรวจจับบุคคลจะตรวจจับและระบุตำแหน่งของคนที่ปรากฏอยู่ในภาพทั้งหมด ส่วนการตรวจจับใบหน้า จะค้นหาและตรวจสอบว่าภาพหรือวิดีโอนั้นมีใบหน้าหรือไม่ และส่วนการตรวจจับจุดสังเกตบนใบหน้า จะตรวจจับจุดต่างๆ บนใบหน้าเท่านั้น เช่น จุดบนจมูก จุดบนดวงตา จุดบนปาก โดยทั้งหมดเป็นการตรวจจับที่แตกต่างกันแต่ก็สามารถทำงานร่วมกันได้

## 2.2 แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองที่ใช้ในการตรวจจับบุคคล

ในการตรวจจับบุคคล (Person Detection) มีหลายอัลกอริทึม (Algorithm) ที่สามารถใช้ได้ในการตรวจจับนี้ เช่น Haar Cascade, SSD, YOLO โดยในที่นี้จะแสดงรายละเอียดเกี่ยวกับโวล (YOLO) ที่นำมาศึกษาในงานวิจัยนี้

### 2.2.1 โวล (YOLO)

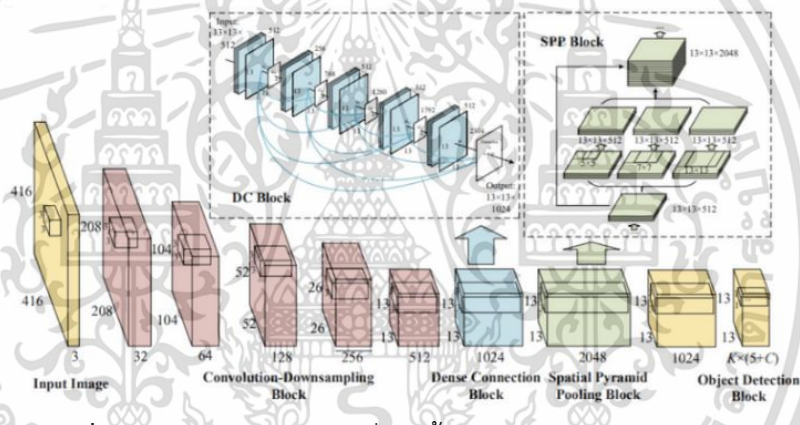


รูปที่ 2.1 สถาปัตยกรรมของโวล

ที่มา : Vaidya et al., (2023)

แบบจำลองโวล (YOLO) เป็นโครงสร้างโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network : CNN) ที่ผ่านการออกแบบและฝึกสอนมาแล้วเพื่อใช้งานสำหรับการตรวจจับวัตถุ (Object Detection) โดยโครงข่ายประสาทเทียมแบบสังวัตนาการ (CNN) เป็นวิธีหนึ่งในการเรียนรู้แบบการเรียนรู้เชิงลึก (Deep Learning) เป็นการจำลองการไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มองเห็นของมนุษย์ที่สามารถแยกแยะคุณลักษณะ (Feature) ของวัตถุที่มองเห็น ที่ได้รับความนิยมในการประมวลผลภาพ (Image Processing) เนื่องจากสามารถจดจำลักษณะของภาพได้ดี และมีความสามารถในการแยกแยะวัตถุที่อยู่ในภาพได้อย่างมีประสิทธิภาพ (นพรุจ และณัฐวุฒิ, 2563) แบบจำลองโยโล่มีความสามารถสูงทั้งด้านความเร็วในการคำนวณและความแม่นยำ และง่ายต่อการฝึกสอนต่อ (Transfer Learning) จึงทำให้โครงสร้างของโยโล่ เป็นที่นิยมอย่างมาก โดยเป็นแบบจำลองการตรวจจับวัตถุตามเวลาจริงที่ทันสมัยซึ่งได้รับการฝึกฝนในชุดข้อมูล MS COCO รูปที่ 2.1 แสดงสถาปัตยกรรมของแบบจำลองโยโล่ ตรงข้ามกับเทคนิคอื่นๆ ซึ่งต้องใช้การสแกนภาพสองครั้ง โดยโยโล่เป็นเครื่องตรวจจับแบบขั้นตอนเดียว กล่าวคือ จะดูภาพอินพุต (Input) เพียงครั้งเดียว แบบจำลองโยโล่ทำงานโดยแบ่งภาพออกเป็นตาราง  $S \times S$  จากนั้นกำหนดความน่าจะเป็นสำหรับกรอบขอบเขตที่จะเกิดขึ้น เนื่องจากลักษณะขั้นตอนเดียวของโยโล่ จึงตรวจจับวัตถุในภาพได้รวดเร็วมาก (จันธิรัก และคณะ, 2565)



รูปที่ 2.2 โครงสร้างของโยโล่ที่เป็นขั้นแบบโครงข่ายประสาทเทียม  
ที่มา : Huang et al, (2020)

ในรูปที่ 2.2 ขั้นแบบโครงข่ายประสาทเทียมแต่ละชั้นจะมีการทำงานหลายกระบวนการกริด (grid) จะเพิ่มขึ้นตั้งแต่ 3 ไปจนได้ผลลัพธ์ที่ดี

### 2.2.2 โยโล่ เวอร์ชัน 7 (YOLOv7)

โยโล่ เวอร์ชัน 7 เป็นเครื่องตรวจจับวัตถุแบบเรียลไทม์แบบขั้นตอนเดียว ตรวจจับวัตถุตามเวลาจริงที่เร็วและแม่นยำสูง โดยมีจุดมุ่งหมายเพื่อปรับปรุงการตรวจจับวัตถุโดยการปรับปรุงตามฟังก์ชันการสูญเสีย และวิธีการฝึกอบรมที่มีประสิทธิภาพ โยโล่ เวอร์ชัน 7 ใช้เครือข่ายการรวมเลเยอร์แบบขยายที่มีประสิทธิภาพ (Extended Efficient Layer Aggregation Network หรือ E-ELAN)

ด้วยการสับเปลี่ยน รวม และขยายจำนวนสมาชิกสำหรับการรวมขั้นสุดท้าย E-ELAN ช่วยให้การไล่ระดับสีของเครือข่ายสั้นลง ซึ่งจะเป็นการเพิ่มประสิทธิภาพ นอกจากนี้ E-ELAN ยังได้รับการปรับให้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เหมาะสมสำหรับการใช้งานหน่วยความจำที่ต่ำลง การปรับขนาดแบบผสมสำหรับทั้งความลึกและความกว้างสำหรับการปรับขนาดแบบจำลองตามการต่อข้อมูลในรุ่นนี้ การเพิ่มความลึกของภาพส่งผลให้ความกว้างของภาพเพิ่มขึ้นด้วยปัจจัยเดียวกัน สิ่งนี้ทำให้ได้โครงสร้างที่เหมาะสมยิ่งขึ้นของแบบจำลองที่ปรับขนาด (Wang et al., 2022)

โพลี เวอร์ชัน 7 ปรับปรุงความเร็วและความแม่นยำโดยการปฏิรูปสถาปัตยกรรมหลายอย่าง เช่นเดียวกับ YOLOv4 โดย YOLOv7 จะไม่ใช้ข้อมูลที่ฝึกไว้ล่วงหน้าของ ImageNet แต่แบบจำลองได้รับการฝึกฝนโดยใช้ชุดข้อมูล COCO ทั้งหมด

การปรับพารามิเตอร์แบบจำลองใหม่เป็นวิธีปฏิบัติของการรวมแบบจำลองการคำนวณหลายแบบจำลองในขั้นตอนการอนุมานเพื่อเร่งเวลาการอนุมานในโพลี เวอร์ชัน 7 มีการใช้เทคนิค "เครือข่ายการรวมเลเยอร์แบบขยายที่มีประสิทธิภาพ" หรือ E-ELAN เพื่อแสดงความสามารถนี้

สถาปัตยกรรม YOLOv7 (Wang et al., 2022)

- เครือข่ายการรวมเลเยอร์แบบขยายที่มีประสิทธิภาพ (Extended Efficient Layer Aggregation Network หรือ E-ELAN)
- การปรับขนาดโมเดลที่ใช้การเชื่อมต่อแบบต่อเนื่อง (Concatenation based Models) จะช่วยให้แบบจำลองทำงานได้ดีกว่าแบบจำลองเดิม เนื่องจากมีประสิทธิภาพในการตรวจจับวัตถุและการติดตามวัตถุในภาพมากขึ้น และเพื่อให้เหมาะกับการใช้งานร่วมกับแบบจำลองอื่นๆ

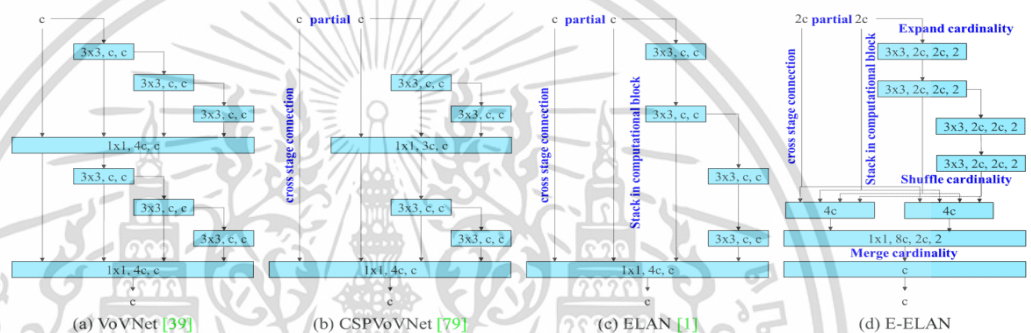
การปฏิรูปสถาปัตยกรรม

เครือข่ายการรวมเลเยอร์แบบขยายที่มีประสิทธิภาพ (E-ELAN) เป็นชั้น CNN (Convolutional Neural Network) ที่ออกแบบมาเพื่อช่วยในการสกัดภาพและเพิ่มประสิทธิภาพในการตรวจจับวัตถุ โดยมีการรวมชั้นการถ่ายโอนรูปแบบระหว่างเลเยอร์ เพื่อลดการคำนวณและประหยัดเวลาในการสร้างแบบจำลอง ทำให้แบบจำลองมีความเร็วและประสิทธิภาพการทำงานที่ดีขึ้น โดย E-ELAN จะทำการสกัดลักษณะเด่น (Features) ของภาพด้วยการเข้าร่วมเลเยอร์ ที่แตกต่างกันในระดับประสิทธิภาพและความลึกต่าง ๆ และใช้เทคนิคการรวมแบบชั้นความลึก (Aggregation) ในการรวมลักษณะเด่นจากเลเยอร์ ที่ต่างกันเข้าด้วยกันเพื่อสร้างฟีเจอร์สำหรับการตรวจจับวัตถุ โดย E-ELAN มีลักษณะเด่นดังต่อไปนี้ (Wang et al., 2022)

- ค่าเข้าถึงหน่วยความจำ: E-ELAN มีความเร็วและใช้พื้นที่หน่วยความจำน้อยกว่าแบบจำลองอื่นๆ ทำให้สามารถนำไปใช้งานได้ง่ายและเหมาะสมกับการใช้งานในอุตสาหกรรมจริง
- อัตราส่วนช่องสัญญาณ I/O: E-ELAN มีช่องสัญญาณน้อยกว่าแบบจำลองอื่นๆ ทำให้ช่วยลดการเกิดความล่าช้าในการส่งข้อมูล

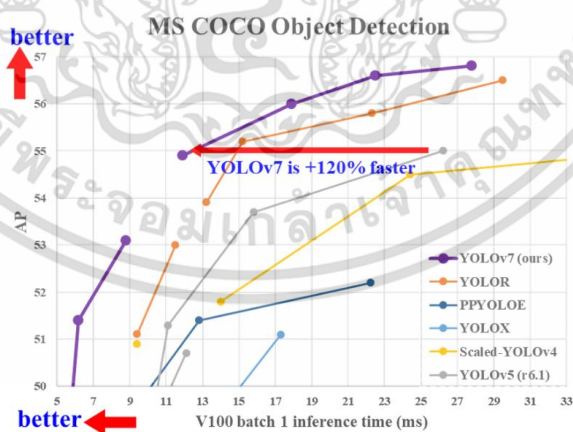
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การดำเนินการอย่างชาญฉลาดขององค์ประกอบ: E-ELAN ใช้วิธีการรวมชั้นโดยใช้การรวมชั้นเป็นประเภทการรวมชั้นอย่างมีประสิทธิภาพ (Efficient Layer Aggregation) ทำให้การสร้างแบบจำลองที่มีประสิทธิภาพสูงและความแม่นยำสูง
- การเปิดใช้งาน: E-ELAN เป็นแบบจำลองที่สามารถเปิดใช้งานได้ง่าย เนื่องจากมีการเขียนโปรแกรมในภาษาไพธอน (Python) และใช้ไลบรารี TensorFlow 2.0 ในการฝึกฝนแบบจำลอง
- เส้นทางไล่ระดับสี: E-ELAN ใช้เส้นทางไล่ระดับสีในการระบุวัตถุในภาพ ซึ่งช่วยให้การตรวจจับวัตถุมีประสิทธิภาพและความแม่นยำสูงขึ้น



รูปที่ 2.3 E-ELAN และงานก่อนหน้าเกี่ยวกับประสิทธิภาพของเลเยอร์สูงสุด

ที่มา : Wang et al., (2022)



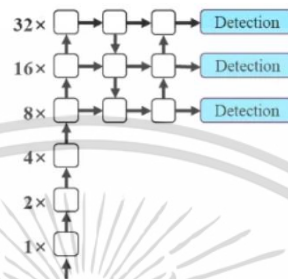
รูปที่ 2.4 กราฟเปรียบเทียบประสิทธิภาพโพลี เวอร์ชัน 7 กับเครื่องมืออื่นๆ

ที่มา : Wang et al., (2022)

เมื่อเทียบกับรุ่น Cascade-Mask R-CNN ที่ประสิทธิภาพดีที่สุดโพลี เวอร์ชัน 7 นั้นให้ความแม่นยำสูงขึ้น 2% ที่ความเร็วการอนุมานที่เพิ่มขึ้นอย่างมาก สิ่งนี้น่าประหลาดใจเนื่องจากเวอร์ชันใหม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Version) R-CNN ดังกล่าวใช้สถาปัตยกรรมแบบหลายขั้นตอน ซึ่งก่อนหน้านี้มีความแม่นยำในการตรวจจับสูงกว่าสถาปัตยกรรมเครื่องตรวจจับแบบขั้นตอนเดียวอย่างมาก

โยโล เวอร์ชัน 7 มีประสิทธิภาพเหนือกว่า YOLOR, YOLOX, Scaled-YOLOv4, YOLOv5, DETR, ViT Adapter-B และอัลกอริธึมการตรวจจับวัตถุอื่นๆ อีกมากมายในด้านความเร็วและความแม่นยำ ดังรูปที่ 2.4



รูปที่ 2.5 สถาปัตยกรรมของโยโล เวอร์ชัน 7

ที่มา : Wang et al., (2022)

| Model                | #Param. | FLOPs  | Size | $AP^{val}$ | $AP_{50}^{val}$ | $AP_{75}^{val}$ | $AP_S^{val}$ | $AP_M^{val}$ | $AP_L^{val}$ |
|----------------------|---------|--------|------|------------|-----------------|-----------------|--------------|--------------|--------------|
| YOLOv4 [3]           | 64.4M   | 142.8G | 640  | 49.7%      | 68.2%           | 54.3%           | 32.9%        | 54.8%        | 63.7%        |
| YOLOR-u5 (r6.1) [81] | 46.5M   | 109.1G | 640  | 50.2%      | 68.7%           | 54.6%           | 33.2%        | 55.5%        | 63.7%        |
| YOLOv4-CSP [79]      | 52.9M   | 120.4G | 640  | 50.3%      | 68.6%           | 54.9%           | 34.2%        | 55.6%        | 65.1%        |
| YOLOR-CSP [81]       | 52.9M   | 120.4G | 640  | 50.8%      | 69.5%           | 55.3%           | 33.7%        | 56.0%        | 65.4%        |
| YOLOv7               | 36.9M   | 104.7G | 640  | 51.2%      | 69.7%           | 55.5%           | 35.2%        | 56.0%        | 66.7%        |
| improvement          | -43%    | -15%   | -    | +0.4       | +0.2            | +0.2            | +1.5         | =            | +1.3         |

รูปที่ 2.6 เปรียบเทียบโยโล เวอร์ชัน 7 กับแบบจำลองพื้นฐานอื่นๆ

ที่มา : Wang et al., (2022)

โยโล เวอร์ชัน 7 มีโครงสร้างประกอบด้วย 3 ส่วนหลักๆ คือ (Chen et al., 2022)

- เครือข่ายหลัก (Backbone Network) : เป็นแบบจำลองโครงข่ายประสาทเทียมแบบสังวัตนาการ (CNN) ที่มีหน้าที่ซับซ้อนในการเรียนรู้และสกัดลักษณะของภาพ โดยแบบจำลอง โยโล เวอร์ชัน 7 นี้ใช้ EfficientNetv2 ซึ่งเป็นแบบจำลอง CNN ที่มีความแม่นยำและความเร็วในการคำนวณที่ดี
- เครือข่ายคอ (Neck Network) : เป็นแบบจำลองที่รวมลักษณะเฉพาะของข้อมูลเครือข่ายหลักเข้าด้วยกันเพื่อสกัดลักษณะของวัตถุอย่างละเอียด โดยแบบจำลองโยโล เวอร์ชัน 7 นี้ใช้ เครือข่ายการรวมเลเยอร์ที่มีประสิทธิภาพ (ELAN) เพื่อรวมลักษณะเฉพาะของข้อมูลจากเครือข่ายหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เครือข่ายหัว (Head Network) : เป็นแบบจำลองที่ใช้ในการทำนายว่าวัตถุที่ตรวจจับอยู่คืออะไร โดยแบบจำลองโวล์ เวอร์ชัน 7 นี้ใช้เลเยอร์ การตรวจจับ เพื่อสร้างพิกัดของวัตถุ (Bounding Box) และทำนายคลาส (Class) ของวัตถุ

## 2.3 แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองที่ใช้ในการตรวจจับจุดสังเกตบนใบหน้า

ในการตรวจจับจุดสังเกตบนใบหน้า (Facial Landmark Detection) มีหลายอัลกอริทึม (Algorithm) ที่สามารถใช้ได้ในการตรวจจับนี้ เช่น Dlib, MTCNN, MediaPipe โดยในที่นี้จะแสดงรายละเอียดเกี่ยวกับมีเดียไปป์ที่นำมาศึกษาในงานวิจัยนี้

### 2.3.1 ไลบรารี มีเดียไปป์ (Library MediaPipe)

มีเดียไปป์พัฒนาโดยกูเกิ้ล (Google) ใช้เฟรมเวิร์ก (Frame Work) TensorFlow เป็นโครงการโอเพ่นซอร์ส (Open Source) โดยใช้การผสมผสานระหว่างการมองเห็นด้วยคอมพิวเตอร์ (Computer Vision) และเทคนิคการเรียนรู้ของเครื่อง เพื่อช่วยให้นักพัฒนาสามารถสร้างไปป์ไลน์ (Pipeline) (Lugaresi et al, 2019) การเรียนรู้ของเครื่องแบบหลายแบบจำลองที่ซับซ้อนสำหรับงานต่างๆ รวมถึงการตรวจจับวัตถุ การติดตาม การตรวจจับใบหน้า การประมาณท่าทาง เป็นต้น และมีการใช้งานอย่างแพร่หลายในการพัฒนาแอปพลิเคชันที่เกี่ยวข้องกับการประมวลผลภาพและวิดีโอ โดยมีความยืดหยุ่นและสามารถปรับเปลี่ยนได้ตามความต้องการของผู้ใช้งาน ซึ่งมีมีเดียไปป์ เขียนด้วย C++ และไพธอน (Bora et al, 2023) เป็นไลบรารีด้านการตรวจจับวัตถุ โดยภายในไลบรารีจะรวบรวมอัลกอริทึม (Algorithm) จากนักวิจัยอื่น ๆ เอาไว้เพื่อให้สามารถเรียกใช้งานได้สะดวก ซึ่งมีสมรรถนะที่ดีในการตรวจจับหน้าใหม่ ๆ ที่ไม่ได้ถูกฝึกไว้ก่อน รวมถึงการติดตามใบหน้าที่รวดเร็ว และมีการใช้งานง่ายและสามารถตรวจจับหน้าได้รวดเร็ว และยังมีความสามารถในการตรวจจับวัตถุต่าง ๆ ในภาพและวิดีโอ อาทิเช่น การตรวจจับใบหน้า (Facial Detection), การตรวจจับสายตา (Eye Tracking), การตรวจจับมือ (Hand Tracking), การติดตามการเคลื่อนไหวของร่างกาย (Pose Estimation) และอื่น ๆ ซึ่งสามารถนำไปใช้ในการพัฒนาแอปพลิเคชันต่าง ๆ เช่น การควบคุมเกมด้วยการเคลื่อนไหวของร่างกาย, การสร้างเอฟเฟกต์ (Effect) ในวิดีโอ, หรือการตรวจจับการอ่านและการแปลภาษามือ โดยมีมีเดียไปป์เป็นเฟรมเวิร์ก (Frame Work) ที่ถูกออกแบบมาเพื่อสร้างและประมวลผลข้อมูลอนุกรมเวลา เช่น วิดีโอและเสียง โดยสามารถทำงานบนแพลตฟอร์มต่างๆ เช่น เดสก์ท็อป/เซิร์ฟเวอร์, Android, iOS และอุปกรณ์ฝังตัว เช่น Raspberry Pi และ Jetson Nano (Atila, 2022) โดยมีมีเดียไปป์เป็นเครื่องมือที่มีความสามารถในการใช้ประโยชน์จากแนวคิดและเทคโนโลยีของเฟรมเวิร์กการเรียนรู้ (Machine Learning) เพื่อให้ผู้พัฒนาสามารถสร้างและพัฒนาผลิตภัณฑ์และบริการที่ปฏิวัติวงการที่เราใช้ในชีวิตประจำวัน นอกจากนี้มีมีเดียไปป์ยังเน้นให้ความสำคัญกับประสิทธิภาพการทำงานและการใช้ทรัพยากร ซึ่งสามารถทำงานได้ไม่ว่าจะเป็นในอุปกรณ์ที่มีขนาดใหญ่หรือแม้แต่อุปกรณ์ IoT ที่มีขนาดเล็ก ทำให้เฟรมเวิร์กนี้เป็นทางเลือกที่น่าสนใจไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งยังมีให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



สามารถนำไปประยุกต์ใช้งานในสถานการณ์และอุปกรณ์ต่างๆ ได้อย่างหลากหลาย และมีความสามารถในการตรวจจับใบหน้าที่ดี (Lugaresi et al, 2019)

### 2.3.2 มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh)

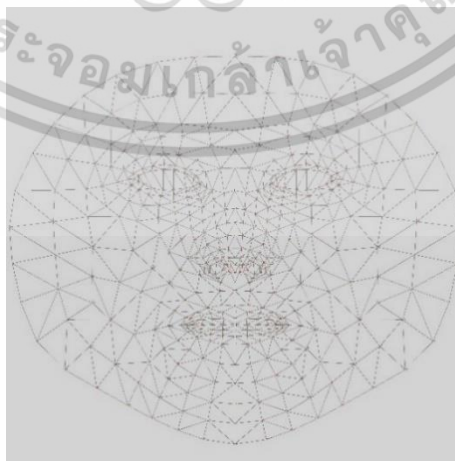
เป็นแบบจำลองตรวจจับจุดบนใบหน้า (Facial landmark detection) ที่พัฒนาโดยกูเกิล (Google) ในส่วนของไลบรารี มีเดียไปป์ (Library MediaPipe) โดยใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) และแบบจำลองปัญญาประดิษฐ์ (Artificial Intelligence) โดยสามารถตรวจจับได้ด้วยวิธีการสกัดจุดสังเกตบนใบหน้า ด้วยวิธีการเรียนรู้ของเครื่อง เพื่อระบุจุดบนใบหน้าของมนุษย์ด้วยความแม่นยำสูง โดยจุดบนใบหน้านี้สามารถใช้ในการตรวจจับและวิเคราะห์พฤติกรรมของใบหน้าได้ เช่น การเคลื่อนไหวของปาก การสัมผัสของมือกับใบหน้า หรือการระบายอารมณ์ที่แสดงผ่านใบหน้า เป็นต้น (Radmehr et al, 2021)

มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) เป็นแบบจำลองในมีเดียไปป์ที่ใช้สำหรับตรวจจับและสร้างกริด (grid) สามมิติบนหน้าของมนุษย์ ซึ่งสามารถใช้สำหรับการทำความเข้าใจรูปหน้าของมนุษย์ได้ โดยจะใช้การเรียนรู้ของเครื่อง และการมองเห็นด้วยคอมพิวเตอร์ (Computer Vision) ในการวิเคราะห์ภาพและสร้างกริดสามมิติที่เชื่อมต่อกับจุดบนหน้า เช่น จุดบนปลายจมูก จุดบนแก้ม จุดบนคิ้ว ฯลฯ โดยจุดบนใบหน้าที่ตรวจจับได้มีจำนวนทั้งหมด 468 จุด ซึ่งเป็นจุดที่คำนวณขึ้นมาจากแบบจำลองทางคณิตศาสตร์ โดยจุดบนใบหน้านี้จะสามารถใช้ในการสร้างแบบจำลองของหน้าเพื่อนำไปใช้งานต่อได้ เช่น ใช้ในการตรวจจับความคล้ายคลึงในการรู้สึกของมนุษย์ หรือใช้ในการปรับแต่งหน้าให้เป็นรูปทรงต่างๆ เช่น ปรับทรงหน้า เป็นต้น (Gomez et al, 2021)

แบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) นั้นใช้โครงข่ายประสาทเทียม (Neural Network) แบบโครงข่ายประสาทเทียมแบบส่งข้อมูลไปข้างหน้า (Feedforward Neural Network หรือ FNN) ที่ประกอบด้วยชั้นเชื่อมต่อเต็มรูปแบบ (Fully Connected Layers) หรือ ชั้นเชื่อมต่อแบบแน่น (Dense Layers) จำนวน 5 ชั้น โดยมีฟังก์ชันเปิดใช้งาน (Activation Function) เป็น Rectified Linear Unit (ReLU) ในชั้นซ่อน และใช้ชั้นสุดท้ายเป็นชั้นผลลัพธ์ ซึ่งมีจำนวนโหนด (Node) ในชั้นนี้เท่ากับ 468 ซึ่งจะใช้ในการสกัดจุดสังเกตบนใบหน้าของรูปภาพใหม่ที่ไม่เคยเห็นมาก่อน แบบจำลองการเรียนรู้เชิงลึก (Deep learning) ที่ใช้ในมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) นั้นถูกพัฒนาขึ้นด้วย TensorFlow แบบจำลองได้ถูกฝึกฝนด้วยข้อมูลแบบจำลองใบหน้า 3 มิติ และ จุดสังเกตบนใบหน้า 2 มิติ โดยใช้ข้อมูลอบรมมากกว่า 33,000 ภาพ รูปภาพของใบหน้าผู้ชายและผู้หญิง และทำการสกัดจุดสังเกตบนใบหน้าในแต่ละใบหน้าจำนวน 468 จุด (Gulati) โดยมีการปรับปรุงแบบจำลองด้วยการใช้การขยายข้อมูล (Data augmentation) เพื่อลดโอกาสเกิด Overfitting นอกจากนี้แบบจำลอง มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) จะใช้ CNN แล้วยังใช้การทำงานแบบการคอนโวลูชันแยกตามความลึกแบบแยกชั้น (Depthwise Separable Convolution) เพื่อเพิ่มประสิทธิภาพและความถูกต้องของการระบุจุดสังเกตบนใบหน้าการคอนโวลูชัน

ชั้นแยกตามชั้น (DSC) เป็นส่วนหนึ่งของสถาปัตยกรรมของแบบจำลอง เพื่อทำการสกัดคุณลักษณะของภาพใบหน้า และควบคุมแบบจำลองให้มีประสิทธิภาพสูงและใช้ทรัพยากรน้อยลง โดย DSC เป็นการนำเอาคอนโวลูชัน (Convolution) แบบการคอนโวลูชันแยกตามความลึก (Depthwise Convolution) และการคอนโวลูชันเพิ่มความลึก (Pointwise Convolution) มาใช้ร่วมกันในการทำคอนโวลูชัน (Convolution) แต่ละชั้น โดยการคอนโวลูชันแยกตามความลึก (Depthwise Convolution) จะทำการคำนวณค่าฟังก์ชันลักษณะ (Feature map) ของแต่ละช่องแยกตามแชนเนล (channel) ของรูปภาพ และการคอนโวลูชันเพิ่มความลึก (Pointwise Convolution) จะทำการส่งผลลัพธ์จากการคอนโวลูชันแยกตามความลึก (Depthwise Convolution) มาเปลี่ยนแปลงมิติของฟังก์ชันลักษณะ (Feature map) เพื่อให้สามารถนำไปใช้ในขั้นต่อไปได้ โดย DSC จะช่วยลดจำนวนพารามิเตอร์และเพิ่มประสิทธิภาพของแบบจำลองได้ เนื่องจากใช้ทรัพยากรน้อยลงในการคำนวณเมื่อเทียบกับคอนโวลูชัน (Convolution) แบบธรรมดา และมีการลดการ Overfitting ของแบบจำลองได้ (Atila, 2022)

สรุปได้ว่ามีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ใช้แบบจำลองการเรียนรู้เชิงลึกที่พัฒนาขึ้นเองโดยใช้ convolutional neural network (CNN) ในการตรวจจับและสกัดลักษณะของใบหน้า (Face Landmark Detection) และทำการสร้างตาข่าย (Mesh) บนใบหน้าจากจุดสังเกตที่ตรวจจับได้ โดยมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ได้ถูกฝึกด้วยข้อมูลใบหน้ามากมายทั้งในด้านข้อมูลแบบจำลองใบหน้า 3 มิติ และ จุดสังเกตบนใบหน้า 2 มิติ และได้ทำการปรับปรุงแบบจำลองด้วยการขยายข้อมูล (Data augmentation) และใช้โครงข่ายประสาทเทียมแบบส่งข้อมูลไปข้างหน้า (FNN) ในการสกัดจุดสังเกตบนใบหน้าของรูปภาพใหม่ที่ไม่เคยเห็นมาก่อน ด้วยการตรวจจับจุดบนใบหน้านี้ สามารถนำข้อมูลจุดบนใบหน้าเหล่านี้มาใช้ในการวิเคราะห์การเคลื่อนไหวของใบหน้าหรือรูปร่างของใบหน้าได้



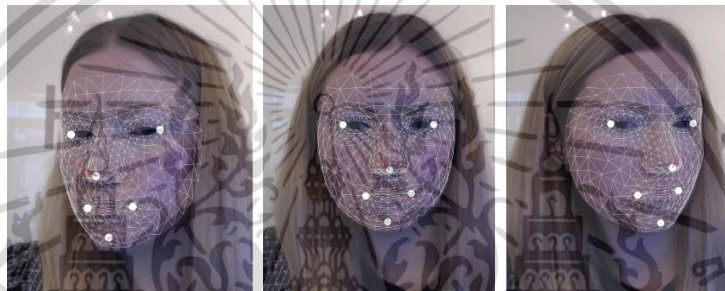
รูปที่ 2.8 มีเดียไปป์ เฟซ เมช 468 จุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ที่มา : Khalid, (2021)  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.9 ตัวอย่างใบหน้าที่ตรวจจับ 468 จุด

ที่มา : Gulati, (2022)



รูปที่ 2.10 ตัวอย่างใบหน้าที่ตรวจจับ 6 จุด

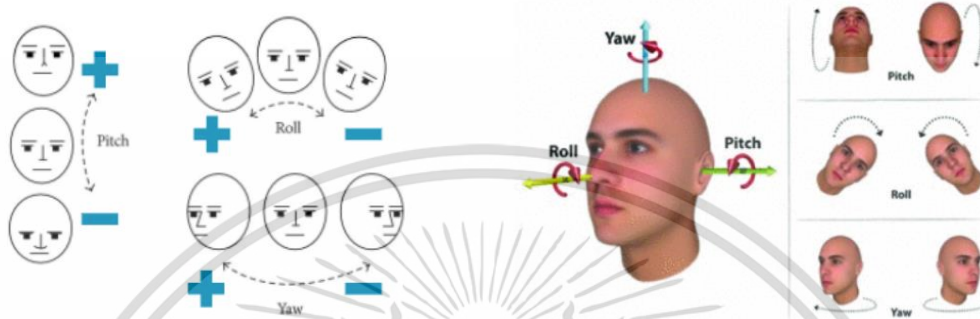
ที่มา : Thierfelder, (2022)

ประสิทธิภาพของเครื่องตรวจจับจุดสังเกตบนใบหน้าได้รับการปรับปรุงอย่างมากในปีที่ผ่านมา ตัวตรวจจับจุดสังเกตจากไลบรารีมีเดียไปป์ (Library MediaPipe) ที่ประเมินจุดสังเกตใบหน้า 3 มิติ 468 จุด แบบเรียลไทม์ (Real time) จุดสังเกต 468 จุดประกอบด้วยข้อมูลของบริเวณใบหน้าที่แตกต่างกัน เช่น แก้ม หน้าผาก ปาก ดวงตา ฯลฯ โดยแบบจำลองนี้จะตรวจจับจุดบนหน้าโดยอิงจากคุณลักษณะที่เป็นหน้าตาอย่างเช่น ร่องรอยของหน้า และโครงสร้างของหน้า ซึ่งต่างจากการตรวจจับใบหน้าที่ตรวจจับแค่ขอบเขตของใบหน้าเท่านั้น ก่อนการทำงานของแบบจำลองตาข่ายใบหน้า (Face Mesh) นี้ จะมีการใช้แบบจำลองที่ชื่อว่า การตรวจจับใบหน้าด้วยมีเดียไปป์ (MediaPipe Face Detection) ในการตรวจจับใบหน้าที่ก่อน เพื่อส่งให้ตาข่ายใบหน้า (Face Mesh) ทำงานต่อไป โดยตาข่ายใบหน้า (Face Mesh) จะทำงานโดยการประมวลผลภาพ 2 มิติของภาพจากกล้องที่จับหน้าของมนุษย์ แล้วนำไปแปลงเป็นข้อมูล 3 มิติเพื่อทำการสกัดคุณลักษณะและตำแหน่งของแต่ละจุดบนหน้า โดยใช้ CNN ที่มีสถาปัตยกรรมการเรียนรู้แบบมีผู้สอน (Supervised) โดยใช้ชุดข้อมูลในการเรียนรู้ และปรับแต่งเพื่อให้แบบจำลองสามารถตรวจจับและสกัดคุณลักษณะของแต่ละจุดบนหน้าได้อย่างแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.3.3 การประมาณท่าศีรษะ (Head pose estimation)

การประมาณท่าศีรษะด้วยองศาการหันหน้าคือการใช้เทคโนโลยีคอมพิวเตอร์ในการตรวจจับและประมาณการท่าต่างๆ ของศีรษะของมนุษย์ จะใช้องศาเป็นหน่วยในการวัด โดยองศาการหันหน้าจะบอกถึงการเคลื่อนไหวของศีรษะและมุมมองของผู้ใช้ รวมทั้งบ่งบอกถึงความสนใจได้อีกด้วย (Radmehr, 2021)



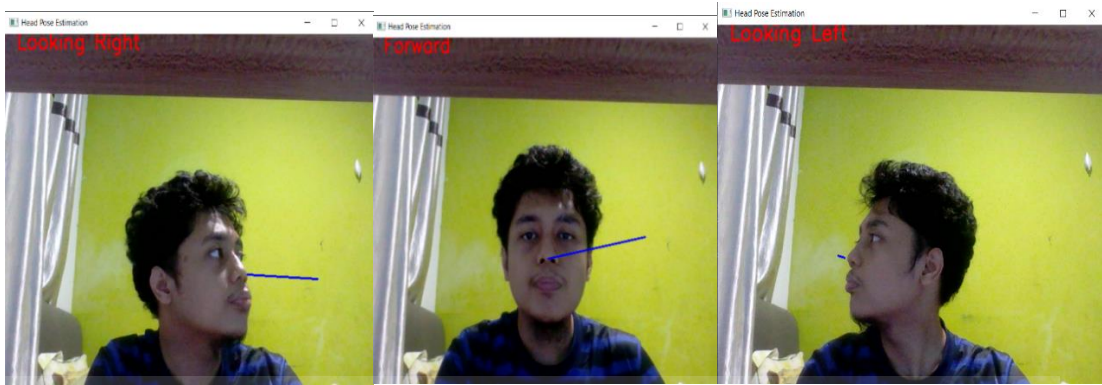
รูปที่ 2.11 ลักษณะทิศทางของศีรษะในมุม Pitch Yaw Roll  
ที่มา : Radmehr et al.(2021)

ในการมองเห็นด้วยคอมพิวเตอร์ (Computer Vision) การประมาณท่าศีรษะเป็นกระบวนการเพื่ออนุมานการวางแนวของศีรษะมนุษย์ในพื้นที่สามมิติจากภาพดิจิทัลสองมิติ ท่าทางของศีรษะสามารถแสดงได้ด้วยมุมการหมุนในปริภูมิสามมิติ มุมการหมุนสามารถแสดงท่าทางของศีรษะได้อย่างแม่นยำในทุกสถานะ ซึ่งประกอบด้วยทิศทางแกน Y (Yaw) ทิศทางแกน X (Pitch) และทิศทางแกน Z (Roll) ดังแสดงในรูปที่ 2.11 ภายใต้องศาที่เหมาะสมคือ  $-90^{\circ} \sim 90^{\circ}$  (Song et al, 2021)

จากรูปที่ 2.11 Pitch เป็นการหันหน้าแบบก้ม-เงย ถ้าเงยหน้า ค่าองศาที่ได้จะเป็นบวก ถ้าก้มหน้าค่าที่ได้จะเป็นลบ ส่วน Yaw เป็นการหันแบบซ้าย-ขวา ถ้าหันขวาค่าจะเป็นบวก ถ้าหันซ้ายค่าที่ได้จะเป็นลบ ซึ่งในงานวิจัยนี้จะพิจารณาค่า Pitch ในแกน X และค่า Yaw ในแกน Y

ตามที่กล่าวไว้ก่อนหน้านี้มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ให้ความสามารถในการตรวจจับใบหน้าและจุดสังเกตบนใบหน้า 468 จุดกระจายอยู่ทั่วใบหน้า พร้อมกับตำแหน่ง (พิกัด x และ y สำหรับแต่ละจุดสังเกตที่ตรวจพบ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.12 ตัวอย่างผลลัพธ์การหันซ้าย-ขวา และมองตรง  
ที่มา : Khalid, (2021)

## 2.4 ทฤษฎีของป้ายโฆษณาดิจิทัล (Digital Signage)

เป็นนวัตกรรมของสื่อป้ายโฆษณาในรูปแบบใหม่สำหรับยุคดิจิทัลที่จะเข้ามาทดแทนสื่อสิ่งพิมพ์และป้ายโฆษณาแบบเดิม โดยสามารถสื่อสารโต้ตอบกับผู้ใช้งาน อีกทั้งยังช่วยเสริมสร้างจินตนาการด้วยเทคโนโลยีภาพและเสียง ช่วยให้ลูกค้าตัดสินใจต่อสินค้าได้มากขึ้น โดยความสามารถในการแสดงผลของป้ายโฆษณาดิจิทัล คือ สามารถแสดงหน้าจอออกเป็นส่วนๆ เพื่อแสดงผลข้อมูลในรูปแบบที่หลากหลายในเวลาเดียวกัน ไม่ว่าจะเป็นข้อมูลในรูปแบบ รูปภาพ ภาพเคลื่อนไหว และอื่นๆ (Cahyadi et al, 2012) **ข้อดีจากการใช้ป้ายโฆษณาดิจิทัล (Digital Signage) ในธุรกิจ**

1. **พื้นที่ไม่จำกัด:** ป้ายดิจิทัลช่วยให้เรามีพื้นที่ไม่จำกัด เราจึงสามารถแสดงเนื้อหาได้มากขึ้นเพื่อดึงดูดผู้ชม สร้างประสบการณ์ให้กับผู้ใช้งานและลูกค้าได้ดียิ่งขึ้น การมองเห็นที่ชัดเจน เป็นเหตุผลที่ว่าทำไม Digital Signage จึงสามารถดึงดูดผู้คนได้มากกว่าสื่อโฆษณาอื่นๆ ซึ่งไม่เพียงแต่นำเสนอเนื้อหาที่น่าสนใจแก่ลูกค้าเท่านั้น แต่ยังดึงดูดความสนใจของลูกค้าอีกด้วย จึงทำให้เป็นที่ต้องการมากขึ้น
2. **การโต้ตอบที่ดีขึ้น:** เนื่องจากเราสามารถแสดงเนื้อหาได้มากขึ้น จึงมีการโต้ตอบที่ดีขึ้น โดยใช้เนื้อหาที่มีมิติเดียวที่สมบูรณ์ยิ่งขึ้น สร้างความโดดเด่น การมีเทคโนโลยีระดับ High-End ช่วยทำให้คุณดูโดดเด่นกว่าคนอื่น ๆ และช่วยให้คุณได้เปรียบเหนือคู่แข่ง โดยผู้ที่เห็นหน้าจอของป้ายดิจิทัลในช่วง 30 วันที่ผ่านมา มักจะจำเนื้อหาได้ และลูกค้าส่วนใหญ่รู้สึกสนใจที่จะติดตามแบรนด์นั้น
3. **เนื้อหาที่ให้ข้อมูล:** เนื่องจากเรามีพื้นที่ไม่จำกัด เราจึงสามารถแสดงข้อมูลที่ละเอียดมากขึ้นได้ สามารถควบคุมเนื้อหาได้ง่ายตามที่ต้องการ ข้อดีอย่างหนึ่งของการมีป้ายดิจิทัลคือ สามารถควบคุมข้อมูลที่จะแสดงได้ ช่วยให้สามารถแสดงเนื้อหา ทำให้พวกเขาได้รับประสบการณ์ที่ดีและเกิดความภักดี (Loyalty) ซึ่งเป็นผลดีต่อธุรกิจในระยะยาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ผลผลิตที่เพิ่มขึ้น: ด้วยการกดเพียงไม่กี่ครั้ง เราก็สามารถแสดงเนื้อหาใหม่ได้ทันที ด้านการลดต้นทุน นอกจากการปรับปรุงการเชื่อมต่อระหว่างลูกค้ากับทางบริษัทแล้วป้ายโฆษณาดิจิทัล (Digital Signage) ยังช่วยประหยัดเงินได้มาก เพราะช่วยลดค่าใช้จ่ายในการพิมพ์โปสเตอร์แบบเก่า และคาดว่าโฆษณาในรูปแบบนี้น่าจะเลิกใช้ในอีกไม่นานบ่อยครั้งที่การแสดงผลข้อมูลเกี่ยวกับผลิตภัณฑ์หรือกิจกรรมให้กับลูกค้าขึ้นอยู่กับเวลาป้ายโฆษณาดิจิทัล สามารถอัปเดตได้ง่าย ไม่เพิ่มต้นทุน และค่าเสียโอกาส หากไม่ได้แสดงป้ายที่พิมพ์ออกมาได้ทันเวลา ซึ่งการใช้ป้ายดิจิทัลเป็นวิธีที่ชาญฉลาดและช่วยเพิ่มพื้นที่จุดโฆษณาได้อีกด้วย

## 2.5 เทคโนโลยีและภาษาโปรแกรมที่เกี่ยวข้อง

ปัจจุบันมีภาษาเขียนโปรแกรมจำนวนมากมายไพธอนเป็นหนึ่งในภาษาเขียนโปรแกรมที่ใช้กันมากที่สุดในโลก และอยู่ในอันดับที่หนึ่งของการจัดอันดับภาษาเขียนโปรแกรมของปี 2019 (ศุภกร, 2562) และปัจจุบันไพธอนเป็นหนึ่งในภาษาสมัยใหม่ที่นิยม ซึ่งเป็นภาษาการเขียนโปรแกรมที่ใช้อย่างแพร่หลายในเว็บแอปพลิเคชัน การพัฒนาซอฟต์แวร์ วิทยาศาสตร์ข้อมูล และการเรียนรู้ของเครื่อง นักพัฒนาใช้ไพธอน เนื่องจากมีประสิทธิภาพ เรียนรู้ง่าย และสามารถทำงานบนแพลตฟอร์มต่างๆ ได้มากมาย

### คุณสมบัติไพธอน

1. ภาษาที่แปลผลแล้ว ไพธอนเป็นภาษาที่แปลผลแล้ว ซึ่งหมายความว่าสามารถเรียกใช้โค้ดที่ละบรรทัดได้โดยตรง หากมีข้อผิดพลาดในโค้ดโปรแกรม ก็จะหยุดทำงานทันที  
ดังนั้นโปรแกรมเมอร์จึงสามารถค้นหาข้อผิดพลาดในโค้ดได้อย่างรวดเร็ว
2. ภาษาที่ใช้งานง่าย ไพธอน ใช้คำที่เหมือนในภาษาอังกฤษ ซึ่งแตกต่างจากภาษาการเขียนโปรแกรมอื่นๆ เนื่องจากไพธอน ไม่ใช้วงเล็บปีกกา แต่จะการใช้การเยื้องแทน
3. ภาษาที่ระบุประเภทแบบไดนามิก โปรแกรมเมอร์ไม่ต้องระบุประเภทตัวแปรเมื่อเขียนโค้ด (Code) เนื่องจากไพธอนจะกำหนดไว้ที่รันไทม์ (Run Time) ด้วยเหตุนี้จึงสามารถเขียนโปรแกรมไพธอนได้รวดเร็วขึ้น
4. ภาษาระดับสูง ไพธอนมีความใกล้เคียงกับภาษามนุษย์มากกว่าภาษาการเขียนโปรแกรมอื่นๆ ดังนั้นโปรแกรมเมอร์จึงไม่ต้องกังวลกับฟังก์ชันการทำงานพื้นฐานต่างๆ เช่น สถาปัตยกรรมและการจัดการหน่วยความจำ
5. ภาษาเชิงอ็อบเจกต์ ไพธอนถือว่าทุกสิ่งเป็นวัตถุ แต่ก็ยังรองรับการเขียนโปรแกรมประเภทอื่นๆ ด้วย เช่น การเขียนโปรแกรมเชิงโครงสร้างและเชิงฟังก์ชัน

เอกสารนี้เป็นเอกสารลิขสิทธิ์พื้นฐานในไพธอนที่เกี่ยวข้อง (อรพิน, 2564) อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. NumPy เป็นไลบรารียอดนิยมที่นักพัฒนาใช้เพื่อสร้างและจัดการอาร์เรย์ จัดการรูปร่างเชิงตรรกะ และดำเนินการคำนวณพีชคณิตเชิงเส้นได้อย่างง่ายดาย โดย NumPy รองรับการทำงานร่วมกับภาษาต่างๆ มากมาย เช่น C และ C++
2. OpenCV-Python เป็นไลบรารีที่นักพัฒนาใช้ในการประมวลผลรูปภาพสำหรับการประยุกต์ใช้เชิงคอมพิวเตอร์วิทัศน์ ซึ่งมีฟังก์ชันมากมายสำหรับงานประมวลผลภาพ เช่น การอ่านและเขียนรูปภาพพร้อมกัน, การสร้างสภาพแวดล้อม 3 มิติจาก 2 มิติ และการบันทึกภาพและวิเคราะห์ภาพจากวิดีโอ ส่วนใหญ่ใช้สำหรับการประมวลผลภาพหรือวิดีโอ และการวิเคราะห์รวมถึงการตรวจจับวัตถุ การตรวจจับใบหน้า ฯลฯ
3. PyTorch เป็นไลบรารีการเรียนรู้ของเครื่องที่ใหญ่ที่สุดที่ปรับการคำนวณเทนเซอร์ให้เหมาะสม มี API มากมายสำหรับการคำนวณเทนเซอร์พร้อมการเร่งความเร็ว GPU ที่แข็งแกร่ง นอกจากนี้ยังช่วยแก้ปัญหาแอปพลิเคชันที่เกี่ยวข้องกับโครงข่ายประสาทเทียม ซึ่งใช้สำหรับการสร้างและการจัดการกับแบบจำลองที่ใช้ในการตรวจจับวัตถุ
4. Matplotlib นักพัฒนาซอฟต์แวร์ใช้ Matplotlib เพื่อลงจุดข้อมูลในกราฟสองมิติและสามมิติ คุณภาพสูง ซึ่งมักใช้ในงานทางวิทยาศาสตร์ Matplotlib ช่วยให้คุณสามารถแสดงข้อมูลเป็นภาพโดยแสดงผลในแผนภูมิต่างๆ เช่น แผนภูมิแท่งและแผนภูมิเส้น คุณยังสามารถลงจุดแผนภูมิได้หลายรายการพร้อมกัน และกราฟสามารถใช้งานได้ในทุกแพลตฟอร์มอีกด้วย
5. Math มีฟังก์ชันและคลาสที่ใช้ในการดำเนินการทางคณิตศาสตร์พื้นฐาน ซึ่งมีฟังก์ชันที่มีอยู่ใน Math เช่น sqrt (square root), sin (sine), cos (cosine), log (logarithm) และอื่น ๆ เพื่อช่วยในการคำนวณค่าทางคณิตศาสตร์ต่าง ๆ ในโปรแกรมไพธอน
6. MediaPipe เป็นเครื่องมือสำหรับการพัฒนาและประมวลผลภาพและวิดีโอในเวลาเรียลไทม์ มีความสามารถในการตรวจจับและติดตามวัตถุหรือส่วนของวัตถุในภาพหรือวิดีโอ ซึ่ง MediaPipe มีฟีเจอร์ที่หลากหลายเช่นการตรวจจับใบหน้า (Face Detection), การตรวจจับจุดสังเกตบนใบหน้า (Face Landmarks Detection), การติดตามมือ (Hand Tracking), การตรวจจับร่างกาย (Pose Detection) และอื่น ๆ โดย MediaPipe มีอินเทอร์เฟซที่ใช้งานง่ายและสามารถใช้ได้กับภาษาไพธอนเพื่อการพัฒนาแอปพลิเคชันที่เกี่ยวข้องกับการประมวลผลภาพและวิดีโอ

## 2.6 ทฤษฎีเกี่ยวกับการวัดประสิทธิภาพของแบบจำลอง

### 2.6.1 เมทริกซ์ความสับสน (Confusion Matrix)

เป็นเครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย หรือ Prediction ซึ่งเป็นตัวเอกสารนี้ ประเมินประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมแบบการจัดกลุ่มหรือจัดประเภทไม่ว่าการ (Classification) (กรินทร์, 2563) แสดงตัวอย่างดังรูปที่ 2.13 เจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## Confusion Matrix

|                        | Actually Positive (1) | Actually Negative (0) |
|------------------------|-----------------------|-----------------------|
| Predicted Positive (1) | True Positives (TPs)  | False Positives (FPs) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs)  |

รูปที่ 2.13 เมทริกซ์ความสับสน

ที่มา : ภาคกร, (2562)

บวกจริง (True Positive หรือ TP) = สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นจริง ในกรณีทำนายว่าจริง และสิ่งที่เกิดขึ้นก็คือจริง

ลบจริง (True Negative หรือ TN) = สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นในกรณี ทำนายว่าไม่จริง และสิ่งที่เกิดขึ้นก็คือ ไม่จริง

บวกเท็จ (False Positive หรือ FP) = สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้นคือทำนายว่า จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง

ลบเท็จ (False Negative หรือ FN) = สิ่งที่ทำนายไม่ตรงกับที่ที่เกิดขึ้นจริง คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้น คือ จริง

โดยค่า TP, TN, FP, FN ในตารางจะแทนด้วยค่าความถี่ จากเมทริกซ์ความสับสน สามารถคำนวณการประเมินประสิทธิภาพของการทำนายแบบจำลอง ในรูปแบบค่าต่าง ๆ ได้หลายค่า (ศรีบุญญา, 2563) ได้แก่

1. ค่าความถูกต้อง (Accuracy) เป็นการวัดค่าของแบบจำลองที่ทำนายถูกก็เปอร์เซ็นต์จากจำนวนการทำนายทั้งหมดในภาพรวม จะมีค่าอยู่ระหว่าง 0-1 ถ้าค่าเข้าใกล้ 1 แสดงว่าแบบจำลองทำนายผลได้ดีมาก

$$\text{ค่าความถูกต้อง} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.1)$$

หรือกล่าวได้ว่า

$$\text{ค่าความถูกต้อง} = \frac{\text{ผลรวมของตัวเลขบนเส้นทแยงมุมในตารางเมทริกซ์ความสับสน}}{\text{ผลรวมทั้งหมดในตารางเมทริกซ์ความสับสน}} \quad (2.2)$$

2. ค่าความแม่นยำ (Precision) เป็นค่าที่แบบจำลองทำนายเป็นคลาสที่สนใจ และถูกต้องต่อค่าที่แบบจำลองทำนายว่าเป็นคลาสที่สนใจทั้งถูกและผิด

เอกสารนี้เป็นเอกสารที่สงวนค่าความแม่นยำงานเพื่อการศึกษา  $\frac{TP}{TP+FP}$  อนุญาตให้นำไปใช้ประโยชน์ด้าน (2.3) ำ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ค่าความระลึก (Recall) หรือ True Positive Rate (TPR) เป็นค่าที่แบบจำลองทำนายเป็นคลาสที่กำลังพิจารณาและถูกต้องต่อคลาสที่สนใจทั้งหมด

$$\text{ค่าความระลึก} = \frac{TP}{TP+FN} \quad (2.4)$$

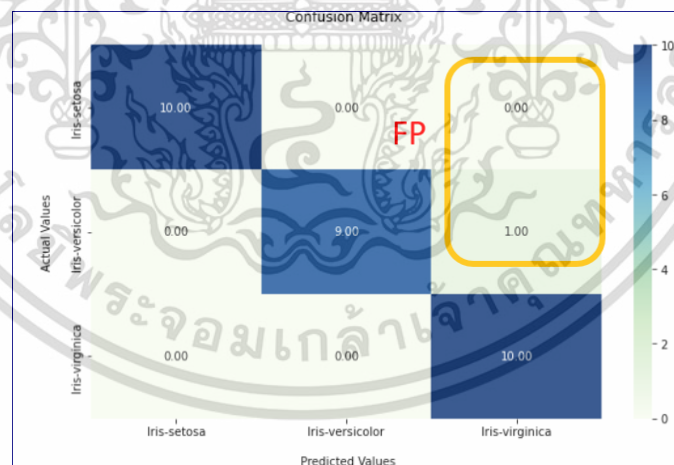
4. ค่าวัดประสิทธิภาพ (F1-score) ค่าเฉลี่ยถ่วงน้ำหนักระหว่างค่าความแม่นยำ (Precision) และ ค่าความระลึก (Recall) เป็นการวัดความแม่นยำและค่าความระลึกของแบบจำลองไปพร้อม ๆ กัน

$$\text{ค่าวัดประสิทธิภาพ} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2.5)$$

โดยค่าวัดประสิทธิภาพ ที่ได้จากการคำนวณจะมีค่าอยู่ระหว่าง 0 ถึง 1 ถ้าค่าที่คำนวณได้เข้าใกล้ 1 หมายความว่า การให้ผลในการจำแนกมีประสิทธิภาพสูง และถ้าค่าคำนวณได้เข้าใกล้ 0 หมายความว่า การให้ผลการจำแนกมีประสิทธิภาพต่ำ

### 2.6.2 เมทริกซ์ความสับสนสำหรับการจำแนกประเภทหลายชั้น (Confusion Matrix for Multiclass Classification)

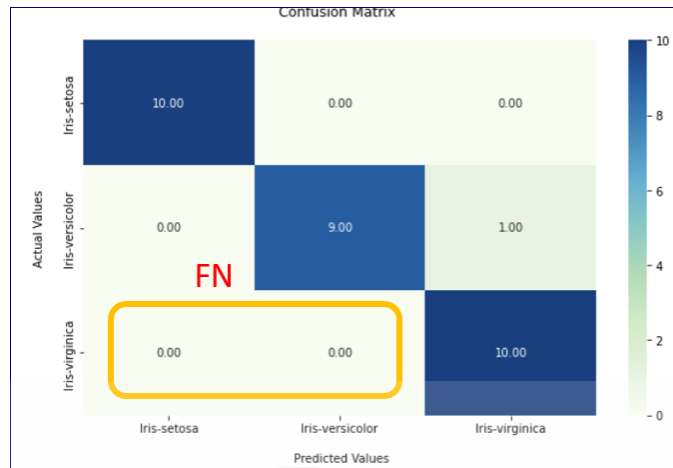
เมทริกซ์ความสับสนสำหรับการจำแนกประเภทหลายชั้น คือเมทริกซ์  $n \times n$  ที่ใช้ในการวัดประสิทธิภาพของแบบจำลองการจำแนกประเภท แสดงการกำหนดค่า TP TN FN FP ดังต่อไปนี้ (Grandini et al, 2020)



รูปที่ 2.14 ตัวอย่างการกำหนด FP ของคลาสที่สนใจ

ที่มา : Chelliah, (2022)

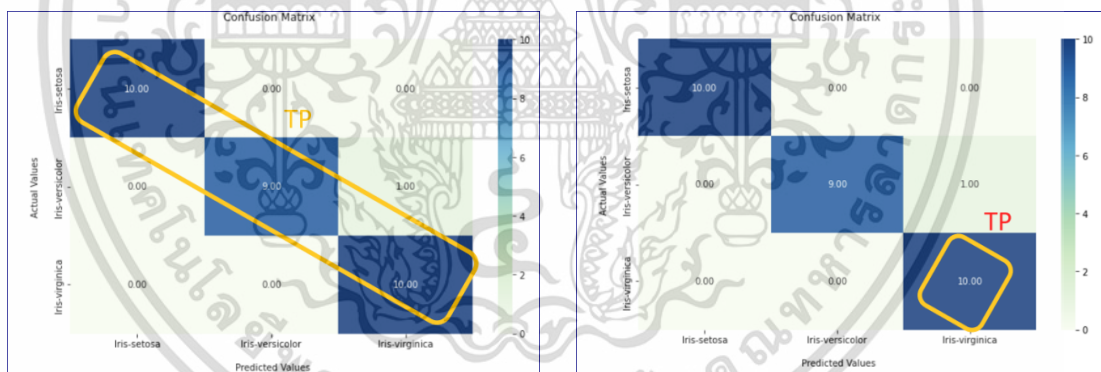
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.15 ตัวอย่างการกำหนด FN ของคลาสที่สนใจ(ต่อ)

ที่มา : Chelliah, (2022)

จากรูปที่ 2.14 เป็นการกำหนด FP ซึ่งหมายถึง สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้นจริง โดยทำนายว่าเป็นคลาสที่สนใจ และสิ่งที่เกิดขึ้นจริงไม่ใช่คลาสที่สนใจ และจากรูปที่ 2.15 เป็นการกำหนด FN หมายถึง สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น โดยทำนายเป็นคลาสอื่นที่ไม่ได้สนใจแต่สิ่งที่เกิดขึ้นจริงเป็นคลาสที่สนใจ



ก.

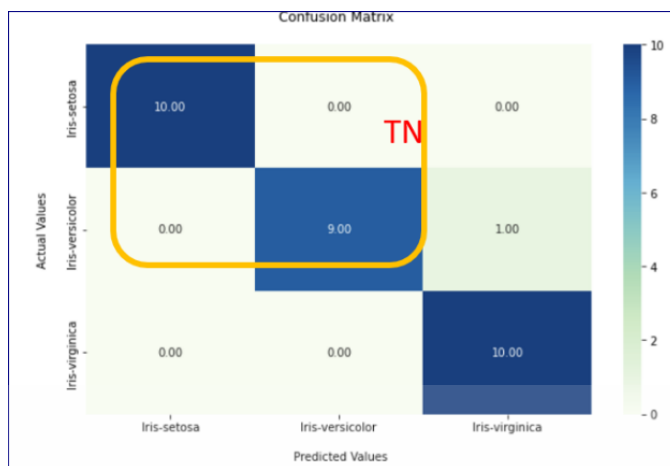
ข.

รูปที่ 2.16 การกำหนดค่า TP

ก. ตัวอย่างการกำหนด TP ของคลาสทั้งหมด ข. ตัวอย่างการกำหนด TP ของคลาสที่สนใจ

ที่มา : Chelliah, (2022)

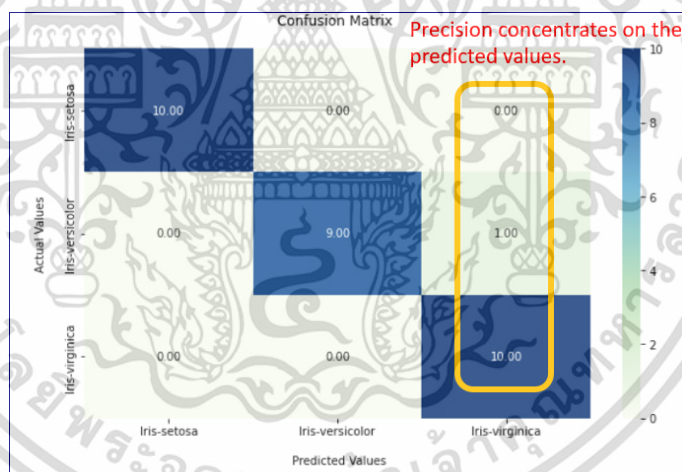
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.17 ตัวอย่างการกำหนด TN ของคลาสที่สนใจ(ต่อ)

ที่มา : Chelliah, (2022)

จากรูปที่ 2.16 กำหนด TP โดยที่ TP หมายถึง สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นจริง โดยทำนายว่าเป็นคลาสที่สนใจ และสิ่งที่เกิดขึ้นจริงก็เป็นคลาสที่สนใจเช่นกัน และจากรูปที่ 2.17 เป็นการกำหนด TN หมายถึง สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น โดยทำนายเป็นคลาสอื่นที่ไม่ได้สนใจและสิ่งที่เกิดขึ้นจริงก็เป็นคลาสอื่นที่ไม่ได้สนใจ ซึ่งเป็นค่าที่นอกเหนือจากการกำหนด TP FP และ FN ไปแล้ว

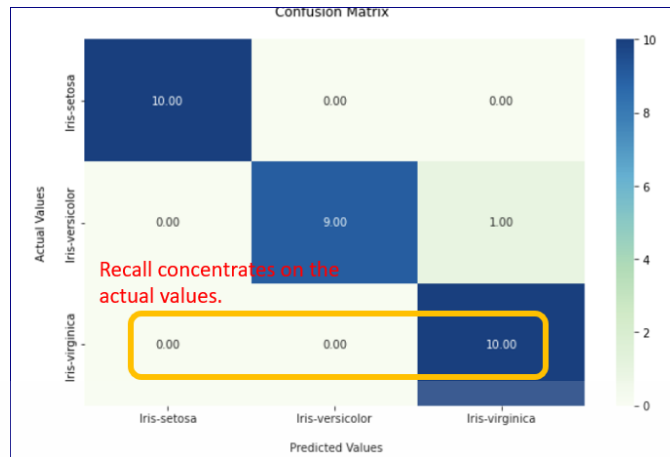


รูปที่ 2.18 ตัวอย่างการกำหนดค่าที่นำไปคิด Precision ของคลาสที่สนใจ(ต่อ)

ที่มา : Chelliah, (2022)

จากรูปที่ 2.18 เป็นการคิดค่าความแม่นยำของแต่ละคลาส ซึ่งคิดจากค่าแนวทแยงในคลาสที่สนใจหารด้วยผลรวมของทุกค่าที่อยู่ในแนวตั้งของคลาสนั้น ๆ เช่น  $10/11 = 0.909 = 0.91$  ค่าความแม่นยำของคลาสนี้มีค่าเท่ากับ 0.91

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.19 ตัวอย่างการกำหนดค่าที่นำไปคิด Recall ของคลาสที่สนใจ(ต่อ)

ที่มา : Chelliah, (2022)

จากรูปที่ 2.19 เป็นการคิดค่าความระลึกของแต่ละคลาส ซึ่งคิดจากค่าแนวทแยงในคลาสที่สนใจหารด้วยผลรวมทุกค่าที่อยู่ในแนวนอนของคลาสนั้น ๆ เช่น  $10/10 = 1$  ค่าความระลึกของคลาสนี้มีค่าเท่ากับ 1

|                 | precision | recall | f1-score | support |
|-----------------|-----------|--------|----------|---------|
| Iris-setosa     | 1.00      | 1.00   | 1.00     | 10      |
| Iris-versicolor | 1.00      | 0.90   | 0.95     | 10      |
| Iris-virginica  | 0.91      | 1.00   | 0.95     | 10      |
| accuracy        |           |        | 0.97     | 30      |
| macro avg       | 0.97      | 0.97   | 0.97     | 30      |
| weighted avg    | 0.97      | 0.97   | 0.97     | 30      |

รูปที่ 2.20 ตัวอย่างการแสดงค่าในแต่ละคลาส และค่าเฉลี่ยรวม

ที่มา : Chelliah, (2022)

Macro avg เป็นการคำนวณโดยใช้ค่าเฉลี่ยเลขคณิตต่อคลาสทั้งหมด โดยไม่คำนึงถึงขนาดของคลาส เช่น ค่าเฉลี่ยของ Recall =  $(1+0.90+1)/3 = 0.966 = 0.97$

Weighted avg เป็นค่าเฉลี่ยถ่วงน้ำหนักคำนวณโดยใช้ค่าเฉลี่ยของคะแนนต่อคลาสโดยพิจารณาจากสัดส่วนของการสนับสนุน (Support) แต่ละคลาส โดยให้แต่ละคลาสมีความสำคัญเท่ากัน เช่น

$$\text{สัดส่วนของแต่ละคลาส} = 10/30 = 0.333$$

$$\text{ค่าเฉลี่ยถ่วงน้ำหนักของ Recall} = (1*0.333)+(0.9*0.333)+(1*0.333) = 0.9657 = 0.97$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.6.3 Receiver Operating Characteristic Curve (ROC Curve)

ROC (Receiver Operating Characteristic) Curve เป็นเส้นที่ใช้วัดถึงประสิทธิภาพของโมเดลแบบ Classification ว่าสามารถทำนายประเด็นที่สนใจได้อย่างแม่นยำได้ดีเท่าไร เส้นกราฟจะเริ่มจาก (0,0) และจบที่ (1,1) เสมอ โดย True Positive Rate (TPR) หรือ ค่าความอ่อนไหว (Sensitive) หรือค่าความระลึก (Recall) คือ สัดส่วนความถูกต้องของผลการทำนายที่เราสนใจ (True Positive) และ False Positive Rate (FPR) หรือค่าความจำเพาะ (Specificity) ก็คือ สัดส่วนความถูกต้องของผลการทำนายที่เราไม่ได้สนใจ (True Negative) แสดงสูตร FPR และ TPR ดังต่อไปนี้

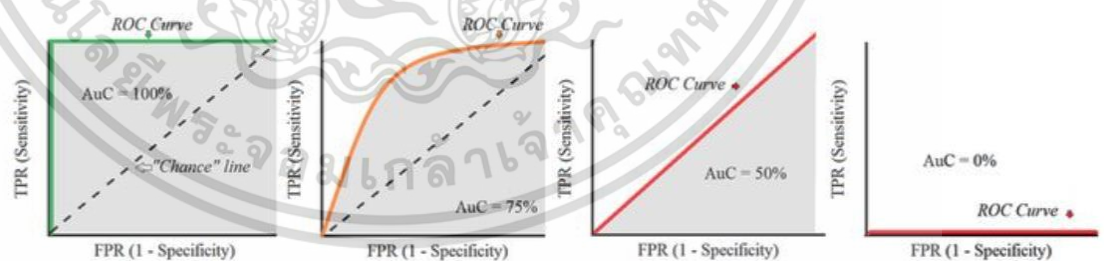
$$\text{อัตราผลบวกปลอม (False Positive Rate)} = \frac{FP}{FP+TN} \quad (2.6)$$

หรือ  $1 - \text{Specificity}$

$$\text{อัตราผลบวกจริง (True Positive Rate)} = \frac{TP}{TP+FN} \quad (2.7)$$

หรือ Sensitivity

พื้นที่ใต้เส้นกราฟ (Area Under The Curve หรือ AUC) ซึ่งค่า AUC นี้จะได้มาจากกราฟ ROC พื้นที่ใต้กราฟมีการพล็อตระหว่าง TPR (True Positive Rate) และ FPR (False Positive Rate) ซึ่งถ้า AUC มีค่าสูงก็หมายความว่าแบบจำลองของเรามีประสิทธิภาพมาก โดยค่า AUC มีค่าอยู่ระหว่าง 0-1 ยิ่งเข้าใกล้ 1 แปลว่า แบบจำลองทำงานได้ดีมาก โดยทั่วไปกราฟยิ่งชิดด้านบนมาก ยิ่งแปลว่าทำนายได้ดี แต่หากกราฟเป็นเส้นตรงจะแสดงว่าแบบจำลองทำนายได้ไม่ดี เกณฑ์การวัดผลของค่า AUC แสดงดังตารางที่ 2.1 (พงษ์เดช และภัทรินทร์, 2564)



รูปที่ 2.21 ตัวอย่างค่า AUC ในแต่ละกรณี

รูปที่ : Glen, (2019)

จากรูปที่ 2.21 จะเห็นได้ว่าถ้าค่าของ Sensitivity และ 1-Specificity มีค่าสูงเส้นกราฟจะโค้งเข้าหามุมบนทางด้านซ้ายมือ หมายถึง โมเดลสามารถทำนายได้ดีมากดังเส้นสีเขียว เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 เกณฑ์การตัดสินใจทั่วไปของค่าประมาณพื้นที่ใต้โค้ง AUC (พงษ์เดช และภัทรนันท์, 2564)

| ค่า AUC      | ความหมายของการจำแนกได้ถูกต้อง |
|--------------|-------------------------------|
| มากกว่า 0.90 | โมเดลทำงานได้ดีมาก            |
| 0.75 – 0.90  | โมเดลทำงานได้ดี               |
| 0.50 – 0.75  | โมเดลทำงานได้ค่อนข้างต่ำ      |
| ต่ำกว่า 0.5  | ไม่สามารถนำมาใช้ได้           |

จากตารางที่ 2.1 พบว่าหากพื้นที่ใต้โค้งมีค่า AUC มากกว่า 0.90 หมายความว่าแบบจำลองมีความแม่นยำสูงสามารถจำแนกข้อมูลได้อย่างถูกต้อง แต่หากค่า AUC ต่ำกว่า 0.50 แบบจำลองจะมีความแม่นยำในการจำแนกข้อมูลต่ำจึงไม่สามารถนำไปใช้งานได้

## 2.7 งานวิจัยที่เกี่ยวข้อง

Gerka (2018) งานวิจัยนี้มีวัตถุประสงค์เพื่อการตรวจจับจำนวนคนในบริเวณเตียงเพื่อเพิ่มความปลอดภัยให้กับผู้ใช้เครื่องช่วยหายใจ ในการศึกษาประเมินผลนี้ ใช้ตัวแยกประเภท Random Forest และบรรลุอัตราการตรวจจับ 84% และ 96% สำหรับสองสถานการณ์ที่แตกต่างกัน

Bhuiyan (2022) งานวิจัยนี้มีวัตถุประสงค์เพื่อทบทวนแนวทาง เทคนิค และกรอบการทำงานต่างๆ ในการวิเคราะห์ฝูงชนในการติดตามและเฝ้าระวังผู้คน โดยเฉพาะในการวิเคราะห์ฝูงชนที่หนาแน่นตามวิดีโอเฝ้าระวังในพิธีฮัจญ์

Muhi, Farhat and Frikha (2022) งานวิจัยนี้มีวัตถุประสงค์เพื่อการจดจำใบหน้าที่สวมหน้ากากที่มีประสิทธิภาพ โดยผลการวิเคราะห์พบว่า Media pipe Landmark ส่งคืนความแม่นยำในการจดจำที่ดีกว่า Dlib ในทุกสถานการณ์ และ Media pipe Landmark ได้รับความแม่นยำในการทดสอบ 98.3% ดังนั้น Media pipe Landmark จึงให้ความแม่นยำที่ดีกว่าสำหรับการจดจำใบหน้าที่สวมหน้ากาก

Radmehr, Asgari and Masouleh (2021) งานวิจัยนี้มีวัตถุประสงค์เพื่อทดลองเกี่ยวกับการเลียนแบบท่าศีรษะและตาของมนุษย์โดยใช้หุ่นยนต์คูกุชานแบบ 3-DOF Agile Eye ร่วมกับ ROS และ MediaPipe Framework ได้มีการนำเสนอการตั้งค่าเชิงกลซึ่งรวมถึงการรวมกันของกลไกคูกุชานทรงกลมแบบ 3-DOF เพื่อเลียนแบบการเคลื่อนไหวของศีรษะและดวงตาของมนุษย์ ด้วยเหตุนี้ไลบรารี มีเดียไปป์ (Library MediaPipe) ซึ่งเปิดตัวโดย Google จึงถูกใช้เพื่อค้นหาท่าทางของศีรษะได้อย่างแม่นยำ

Smith and Foxcroft (2009) งานวิจัยนี้มีวัตถุประสงค์เพื่อประเมินความสัมพันธ์ระหว่างการโฆษณาเครื่องดื่มแอลกอฮอล์ และพฤติกรรมการดื่มที่เป็นผลตามมามีในคนหนุ่มสาว การศึกษาไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเมินโฆษณาเครื่องดื่มแอลกอฮอล์และความเสี่ยงด้านการตลาดที่หลากหลาย รวมถึงสื่อสิ่งพิมพ์ และวิทยุกระจายเสียง การดื่มที่วัดได้โดยใช้การวัดผลลัพ์ที่หลากหลาย ประเมินผู้ดื่มและผู้ไม่ดื่ม แยกกัน ผู้ไม่ดื่มมีแนวโน้มที่จะกลายเป็นผู้ดื่มอย่างมีนัยสำคัญ เมื่อติดตามผลโดยเปิดรับโฆษณา เครื่องดื่มแอลกอฮอล์มากขึ้น

Vaidya et al. (2023) งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโซลูชันสำหรับการตรวจจับโรคพืช โดยใช้แบบจำลองในตระกูล YOLO ของแบบจำลองการตรวจจับวัตถุ YOLOv7 แบบจำลองที่ได้รับการฝึกสอนในชุดข้อมูล PlantDoc แสดงให้เห็นถึงความก้าวหน้าอย่างมากเหนือแบบจำลองที่ใช้ก่อนหน้านี้จากตระกูล YOLO สำหรับปัญหาเดียวกัน เมื่อเทียบกับรุ่นที่มีอยู่แล้ว Precision 72.8%, Recall 68.5%, and mAP of 71% โดยรวมแล้ว YOLOv7 ได้รับการพิสูจน์แล้วว่าดีกว่ารุ่นตรวจจับวัตถุรุ่นก่อนๆ ที่วัดโดยเมตริกประสิทธิภาพ ซึ่งหมายถึงความแม่นยำเฉลี่ย (mAP) นอกจากนี้ยังตรวจจับวัตถุได้เร็วกว่ารุ่นก่อนหน้าอย่างมาก

ธนะพงษ์ และคณะ. (2566) งานวิจัยนี้ได้ทำการตรวจสอบการตรวจจับอารมณ์บนใบหน้าของผู้สูงอายุไทยโดยใช้ YOLOv7 ประสิทธิภาพของ YOLOv7 ในการตรวจจับอารมณ์บนใบหน้าผู้สูงอายุชาวไทยเมื่อเปรียบเทียบกับ Faster R-CNN และ SSD จากผลการทดลองพบว่า YOLOv7 ให้ประสิทธิภาพที่ดีที่สุดของ mAP ด้วยค่า 95%

Soma and Waddenkery. (2022). งานวิจัยนี้มีวัตถุประสงค์ในการพัฒนาโซลูชันที่เหมาะสมที่สุดในการตรวจจับคั้นหาวัตถุหลายชิ้น ประสิทธิภาพในการตรวจจับจะเพิ่มขึ้นโดยการปรับค่าเกณฑ์ความเชื่อมั่น สำหรับชุดข้อมูลรูปภาพ 100 ชุด ที่ค่าเกณฑ์ความเชื่อมั่น 0.3 อัลกอริทึม (Algorithm) มีอัตราการจำแนกที่ดีที่สุดที่ 91% เช่นเดียวกับชุดข้อมูลรูปภาพ 350 ชุด ที่ค่าเกณฑ์ความเชื่อมั่นคือ 0.3 อัตราการจำแนกที่ดีที่สุดคือ 80% จากงานวิจัยนี้สรุปได้ว่า 0.3 เป็นค่าเกณฑ์ความเชื่อมั่นที่เหมาะสมที่สุด

จากงานวิจัยของ Vaidya et al. และงานวิจัยของธนะพงษ์ และคณะ ทำให้ทางผู้วิจัยสนใจในอัลกอริทึม (Algorithm) YOLOv7 ในการตรวจจับบุคคล เนื่องจากให้ผลการทดลองและประสิทธิภาพที่ค่อนข้างดี และแม่นยำ และจากงานวิจัยของ Radmehr, Asgari and Masouleh ทำให้ทางผู้วิจัยสนใจศึกษาและเลือกไลบรารีมีเดียไปป์ (Library MediaPipe) มาใช้ในการตรวจจับจุดสังเกตบนใบหน้าเพื่อประเมินความสนใจ ซึ่งมีความแม่นยำในการค้นหาท่าทางของศีรษะได้เป็นอย่างดี และจาก Song et al. กล่าวว่ามุมการหันใบหน้าสามารถแสดงท่าทางของศีรษะได้อย่างแม่นยำในทุกสถานะ โดยแบบจำลองที่ตรวจจับจุดสังเกตบนใบหน้าคือแบบจำลองมีเดียไปป์ เฟซ เมช เป็นหนึ่งในแบบจำลองที่อยู่ในไลบรารี มีเดียไปป์ ซึ่งส่วนใหญ่นำไปใช้ในการตรวจจับความรู้สึกบนใบหน้า การตรวจจับจุดบริเวณรอบดวงตาเพื่อตรวจจับการผลอหลับขณะขับรถ หรือการอ่านริมฝีปาก เป็นต้น ซึ่งในงานวิจัยนี้ได้ใช้แบบจำลองมีเดียไปป์ เฟซ เมชในการตรวจจับความสนใจ ผู้วิจัยได้กำหนดองศาในการหันใบหน้าเพื่อดูความสนใจที่ 3 ระดับ คือ 10 15 และ 20 องศา โดยป้ายโฆษณาดิจิทัลขนาด

เล็กในที่นี้มีขนาดประมาณ 7-10 นิ้ว โดยขนาดหน้าจอ 10 นิ้ว จะมีความยาวและความกว้าง

โดยประมาณอยู่ที่ 27.2\*16.6 เซนติเมตร และจากงานวิจัยของ Soma and Waddenkery. ที่ได้ทดสอบค่าเกณฑ์ความเชื่อมั่นและได้ผลลัพธ์ว่าค่าเกณฑ์ 0.3 เป็นค่าเกณฑ์ความเชื่อมั่นที่เหมาะสมที่สุด ทางผู้วิจัยจึงกำหนดค่าเกณฑ์ความเชื่อมั่นในการตรวจจذبอยู่ที่ 0.3 ในการตรวจจذبบุคคลโดยใช้แบบจำลองโยโล เวอร์ชัน 7



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### วิธีการดำเนินงานวิจัย

ในงานวิจัยครั้งนี้ผู้วิจัยได้ประยุกต์ใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้ว เพื่อทำการทดสอบ การตรวจจับและนับจำนวนบุคคล พร้อมทั้งประเมินความสนใจโดยใช้จุดสังเกตบนใบหน้า ในงานวิจัย นี้จะแบ่งออกเป็น 2 หัวข้อ

1. การตรวจจับและนับจำนวนบุคคล โดยใช้แบบจำลองโยโล เวอร์ชัน 7 (YOLOv7)
2. การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ใช้แบบจำลอง มีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ซึ่งเป็นแบบจำลองหนึ่งในไลบรารี มีเดียไปป์ (Library MediaPipe)

โดยรายละเอียดของวิธีการดำเนินงานวิจัยสามารถแบ่งเป็น 5 หัวข้อย่อย คือ ประชากรและ กลุ่มตัวอย่าง เครื่องมือที่ใช้ในงานวิจัย ขั้นตอนการเตรียมเครื่องมือในการตรวจจับ ขั้นตอนการ ดำเนินงานวิจัย การเก็บรวบรวมข้อมูล ขั้นตอนการเตรียมข้อมูลและการวิเคราะห์ข้อมูล การทดสอบ แบบจำลอง และการประเมินประสิทธิภาพของแบบจำลอง แสดงดังต่อไปนี้

#### 3.1 ประชากรและกลุ่มตัวอย่าง

##### 3.1.1 ประชากรในการวิจัย

ประชากรในหัวข้อการตรวจจับและนับจำนวนบุคคลเป็นรูปที่ได้จากการนำวิดีโอมาแตก เป็นเฟรม ซึ่งวิดีโอเป็นการบันทึกกลุ่มคนที่สัญจรผ่านในขณะที่บันทึกวิดีโอ มีความยาววิดีโอละ 5 นาที โดยข้อมูลชุด A1 มีจำนวนวิดีโอ 1 วิดีโอ/มุมมอง เป็นชุดข้อมูลที่ได้ทำการแตกเฟรมและสุ่มขึ้น มาแล้ว เพื่อนำมาทดสอบเปรียบเทียบประสิทธิภาพ แต่ข้อมูลชุด B1 เป็นชุดข้อมูลที่ได้ทำการแตก เฟรมและสุ่มขึ้นมาทดสอบเอง มีจำนวน 3 วิดีโอ/มุมมอง ประชากรในการตรวจจับและวัดองศา ระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ เป็นการตรวจจับแบบสตรีมวิดีโอ ในกรณีนี้ ระบบ ตรวจจับจะทำงานบนวิดีโอที่เป็นลำดับของเฟรมภาพ โดยทำการตรวจจับคนต่อเนื่องตามเวลา จะมี การเลื่อนเฟรมภาพไปเรื่อย ๆ ในอัตราเฟรมที่ตามมากับวิดีโอ เพื่อตรวจจับและติดตามการเคลื่อนไหว ของคนในวิดีโอ แสดงรายละเอียดต่อไปนี้

1. การตรวจจับและนับจำนวนบุคคล แบ่งข้อมูลออกเป็น 2 ชุด

ข้อมูลชุด A1 : จำนวนเฟรมที่ได้จากการแตกเฟรมจากวิดีโอ โดยแบ่งเป็น 2 มุมมอง คือ มุมมองแนวระนาบ และมุมมองแนวเหนือศีรษะ จำนวนมุมมองละ 500 เฟรม

ข้อมูลชุด B1 : จำนวนเฟรมที่ได้จากการแตกเฟรมจากวิดีโอ โดยแบ่งเป็น 2 มุมมอง คือ มุมมองแนวระนาบ และมุมมองแนวเหนือศีรษะ จำนวนมุมมองละ 900 เฟรม ซึ่ง มาจากวิดีโอ 3 วิดีโอ/มุมมอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ แบ่ง

ข้อมูลออกเป็น 2 ชุด

ข้อมูลชุด A2 : กลุ่มคนที่อยู่ในวิดีโอที่ได้ทำการบันทึกวิดีโอเอง จำนวนทั้งหมด 10 วิดีโอ

ข้อมูลชุด B2 : กลุ่มคนที่อยู่ในวิดีโอทั้งหมด 10 วิดีโอ ซึ่งประกอบด้วย ข้อมูลที่บันทึกวิดีโอเอง 8 วิดีโอ และนำเข้าจากอินเทอร์เน็ต 2 วิดีโอ ในวิดีโอที่ได้บันทึกด้วยตนเอง จะมีความเร็วในการบันทึกภาพ 30 เฟรม/วินาที

### 3.1.2 กลุ่มตัวอย่างที่ใช้ในการศึกษา

กลุ่มตัวอย่างที่เลือกมาวิเคราะห์ในการตรวจจับและนับจำนวนบุคคลเป็นเฟรมที่ได้จากการนำวิดีโอมาแตกเป็นเฟรมรูปภาพ โดยข้อมูลชุด A1 เป็นชุดข้อมูลที่ได้ทำการแตกเฟรมและสุ่มขึ้นมาแล้ว เพื่อนำมาทดสอบเปรียบเทียบประสิทธิภาพ แต่ข้อมูลชุด B1 เป็นชุดข้อมูลที่ได้ทำการแตกเฟรมและสุ่มขึ้นมาทดสอบเอง สำหรับกลุ่มตัวอย่างที่เลือกมาวิเคราะห์ในการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ คือกลุ่มคนที่อยู่ในวิดีโอที่มีระยะห่างจากกล้องประมาณ 1-2 เมตร

#### 1. หัวข้อการตรวจจับและนับจำนวนบุคคล

ข้อมูลชุด A1 : เฟรมที่ถูกสุ่มขึ้นมาใช้ในการทดสอบ 150 เฟรม/มุมมอง

ข้อมูลชุด B1 : เฟรมที่สุ่มเฟรมขึ้นมาทั้งหมด 300 เฟรม/วิดีโอ สุ่มมา 60 เฟรม/วิดีโอด้วยวิธีการสุ่มตัวอย่างแบบไม่ใช้ความน่าจะเป็น ในที่นี้ใช้ 6 วิดีโอ จะได้กลุ่มตัวอย่างที่ใช้ทดสอบ 180 เฟรม/มุมมอง แสดงดังตารางที่ 3.1 และ 3.2

#### 2. หัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

ในหัวข้อนี้จะใช้

ข้อมูลชุด A2 : กลุ่มคนที่อยู่ในวิดีโอที่อยู่ห่างจากกล้องประมาณ 1-2 เมตร หรือกลุ่มคนที่สนใจกล้อง มีวิดีโอจำนวน 10 วิดีโอ

ข้อมูลชุด B2 : กลุ่มคนที่อยู่ในวิดีโอที่อยู่ห่างจากกล้องประมาณ 1-2 เมตร หรือกลุ่มคนที่สนใจกล้อง มีวิดีโอจำนวน 10 วิดีโอ

ตารางที่ 3.1 จำนวนเฟรมที่ใช้ทดสอบในชุดข้อมูล A1

| ชุดข้อมูล | มุมมอง        | จำนวนวิดีโอ | ความยาว (นาท) | จำนวนเฟรมทั้งหมด | จำนวนเฟรมที่สุ่ม |
|-----------|---------------|-------------|---------------|------------------|------------------|
| A1        | แนวระนาบ      | 1           | 5             | 500              | 150              |
|           | แนวเหนือศีรษะ | 1           | 5             | 500              | 150              |

ตารางที่ 3.2 จำนวนเฟรมทั้งหมดและจำนวนเฟรมสุ่มขึ้นมาที่ใช้ทดสอบในชุดข้อมูล B1

| ชุดข้อมูล | มุมมอง        | วิดีโอ  | ความยาว (นาท) | จำนวนเฟรมทั้งหมด | จำนวนเฟรมที่สุ่ม |
|-----------|---------------|---------|---------------|------------------|------------------|
| B1        | แนวระนาบ      | วิดีโอ1 | 5             | 300              | 60               |
|           |               | วิดีโอ2 | 5             | 300              | 60               |
|           |               | วิดีโอ3 | 5             | 300              | 60               |
|           |               | รวม     |               | 900              | 180              |
|           | แนวเหนือศีรษะ | วิดีโอ4 | 5             | 300              | 60               |
|           |               | วิดีโอ5 | 5             | 300              | 60               |
|           |               | วิดีโอ6 | 5             | 300              | 60               |
|           |               | รวม     |               | 900              | 180              |

จากตารางที่ 3.1 ชุดข้อมูล A1 มีจำนวนเฟรมทั้งหมด 500 เฟรม/มุมมอง สุ่มจำนวนเฟรมที่ใช้ในการทดสอบ 150 เฟรม/มุมมอง และจากตารางที่ 3.2 ชุดข้อมูล B1 มีจำนวนเฟรมทั้งหมด 900 เฟรม/มุมมอง สุ่มขึ้นมาใช้ในการทดสอบ 180 เฟรม/มุมมอง

### 3.2 เครื่องมือที่ใช้ในงานวิจัย

ในงานวิจัยใช้ภาษาไพธอนในการเขียนโปรแกรมในทั้ง 2 หัวข้อ เนื่องจากเป็นภาษาที่มีโครงสร้างที่เข้าใจง่าย การฝึกเขียนโปรแกรมด้วยไพธอนจึงทำได้ไม่ยากและใช้เวลาไม่นาน โดยแต่ละหัวข้อจะแสดงเครื่องมือและแบบจำลองที่นำมาใช้ในการทดสอบ ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.1 การตรวจจับและนับจำนวนบุคคล

- Google Colaboratory (Google Colab) ใช้สำหรับเขียนและรันโค้ดไพธอน ในเว็บเบราว์เซอร์โดยไม่ต้องติดตั้งในเครื่องของเราเอง โดยใช้พื้นที่บนคลาวด์และมีการใช้งานของระบบการประมวลผลแบบแบ่งเป็นเซลล์ (Cell) แต่ละเซลล์สามารถรันโค้ดไพธอนได้แยกกัน ซึ่งสามารถเพิ่มหรือลบเซลล์ได้ตามต้องการ

- YOLOv7 เป็นแบบจำลองในการตรวจจับวัตถุต่าง ๆ ได้อย่างถูกต้องและแม่นยำ โดยนำแบบจำลองนี้มาใช้ในการตรวจจับและนับจำนวนบุคคล

### 3.2.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

- Jupyter Notebook เป็นแอปพลิเคชันเว็บเบราว์เซอร์ที่ใช้ในการเขียนและเรียกใช้งานโปรแกรมไพธอน (และภาษาโปรแกรมอื่น ๆ) ในรูปแบบของเซลล์ (Cells) สามารถรวมการเขียนโค้ดพร้อมกับการแสดงผลเป็นเอกสารที่เรียบง่ายและสื่อสารได้ง่าย ๆ ซึ่งใช้

- MediaPipe Face Mesh เป็นแบบจำลองในการตรวจจับจุดสังเกตบนใบหน้า ซึ่งเป็นหนึ่งในแบบจำลองที่อยู่ในไลบรารีมีเดียไปป์ นำแบบจำลองนี้มาใช้ในการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

## 3.3 ขั้นตอนการเตรียมเครื่องมือในการตรวจจับ

### 3.3.1 ภาษาไพธอน 3.9.13 สามารถดาวน์โหลดและติดตั้งเวอร์ชันที่สนใจได้จาก

<https://www.python.org/downloads/>

### 3.3.2 การตรวจจับและนับจำนวนบุคคล

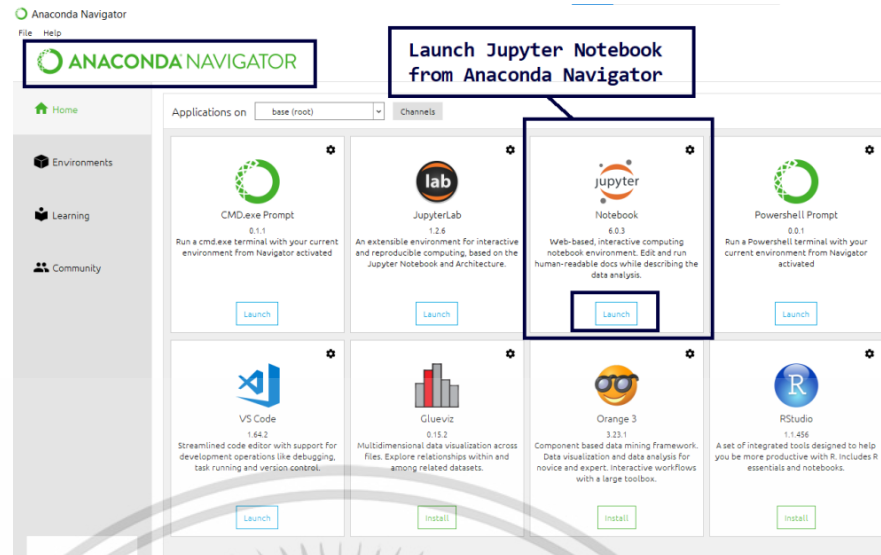
ในการตรวจจับและนับจำนวนบุคคลจะทำการรันโปรแกรมด้วย Google Colaboratory โดยใช้แบบจำลองโยโล เวอร์ชัน 7 ที่ได้ผ่านการฝึกสอนแล้ว โดยในการนำแบบจำลองนี้มาใช้ต้องทำการดาวน์โหลดรูปแบบไฟล์ที่ใช้เก็บค่าพารามิเตอร์ (Parameters) หรือน้ำหนัก (Weights) ของแบบจำลองก่อน ซึ่งไฟล์นี้จะถูกนำไปใช้ในกระบวนการดาวน์โหลดและการนำแบบจำลองไปใช้งานในการตรวจจับ โดยไฟล์ดังกล่าวสามารถดาวน์โหลดได้ตาม URL ด้านล่างนี้

<https://github.com/WongKinYiu/yolov7/releases/download/v0.1/yolov7.pt>

### 3.3.3 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

ในการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้อง ทำการรันโปรแกรมด้วย Jupyter Notebook โดย Jupyter Notebook สามารถเรียกใช้งานได้จาก Anaconda Navigator ซึ่ง Anaconda เป็นแพลตฟอร์มในการรวบรวมโปรแกรมต่าง ๆ รวมไปถึงคำสั่งต่าง ๆ เพื่อใช้ในการรันโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 ตัวอย่าง Anaconda Navigator

สามารถดาวน์โหลด Anaconda Navigator ได้จาก

<https://www.anaconda.com/download>

และสามารถดูขั้นตอนการดาวน์โหลดอย่างละเอียดได้จาก

<https://www.youtube.com/watch?v=ZEVDBSuVXpk&t=168s>

สำหรับการนำเข้าหรือเรียกใช้ไลบรารีมีเดียไปป์เพื่อนำมาตรวจจับจุดสังเกตบนใบหน้า สามารถดูขั้นตอนได้ในหัวข้อ 3.7.2

### 3.4 ขั้นตอนการดำเนินงานวิจัย

แบ่งเป็น 2 หัวข้อที่ทำการตรวจจับ ได้แก่

#### 3.4.1 การตรวจจับและนับจำนวนบุคคล

ระบบตรวจจับบุคคลด้วยวิธีโพลี เวอร์ชัน 7 เป็นระบบที่ตรวจจับมนุษย์ที่ตรวจเจอในแต่ละเฟรม และทำการนับจำนวนคนที่สามารถตรวจจับได้ตามเงื่อนไขที่กำหนด ทำการทดสอบบนเว็บเบราว์เซอร์ Google Colab โดยมีรายละเอียดดังนี้

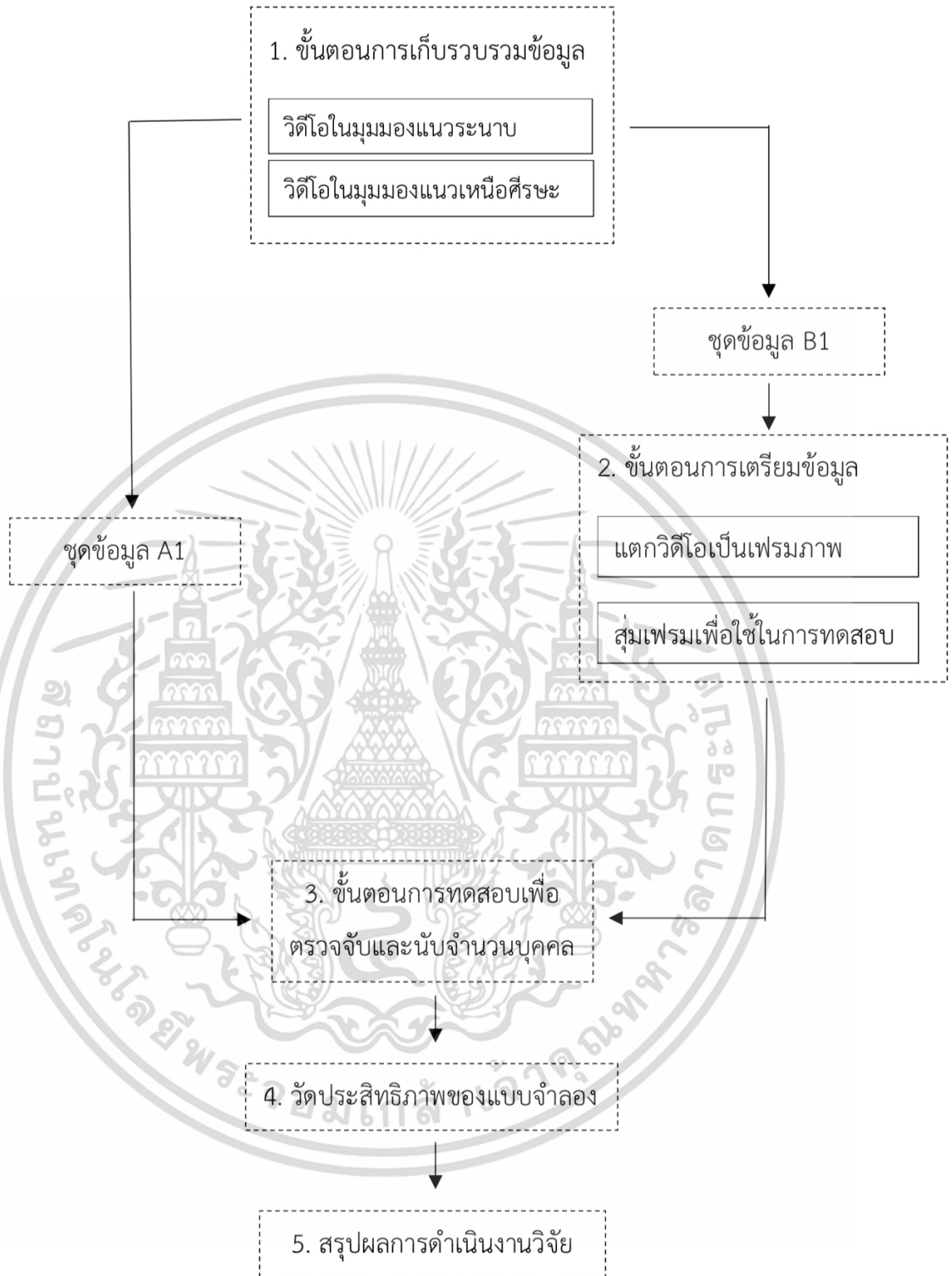
1. การเก็บรวบรวมข้อมูล เป็นขั้นตอนในการรวบรวมข้อมูลเพื่อนำมาทดสอบ หรือเป็นขั้นตอนในการเตรียมวิดีโอในมุมมองที่สนใจ ซึ่งมีข้อมูลในการทดสอบ 2 ชุด คือ ข้อมูลชุด A1 และข้อมูลชุด B1 โดยแต่ละชุดจะแบ่งเป็น 2 มุมมอง คือ 1. มุมมองระนาบ 2. มุมมองเหนือศีรษะ (แสดงในหัวข้อ 3.5.1)
2. การเตรียมข้อมูล เป็นการเตรียมความพร้อมของข้อมูลก่อนนำข้อมูลไปทดสอบ (แสดงในหัวข้อ 3.6.1)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์โดยมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย

- 2.2 สุ่มเฟรมขึ้นมาเพื่อทำการทดสอบ จากทั้งหมด 900 เฟรม/มุมมอง สุ่มขึ้นมา 180 เฟรม/มุมมอง
3. การทดสอบเพื่อตรวจจับและนับจำนวนบุคคล ใช้แบบจำลองโยโล เวอร์ชัน 7 ที่ผ่านการฝึกสอนแล้ว ในการตรวจจับและนับจำนวนบุคคล
4. วัดประสิทธิภาพของแบบจำลอง ทำการวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าวัดประสิทธิภาพ (F1-score) และพิจารณาจากกราฟ ROC
5. สรุปผลการดำเนินงานวิจัย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 ขั้นตอนการดำเนินงานวิจัยหัวข้อการตรวจจับและนับจำนวนบุคคล

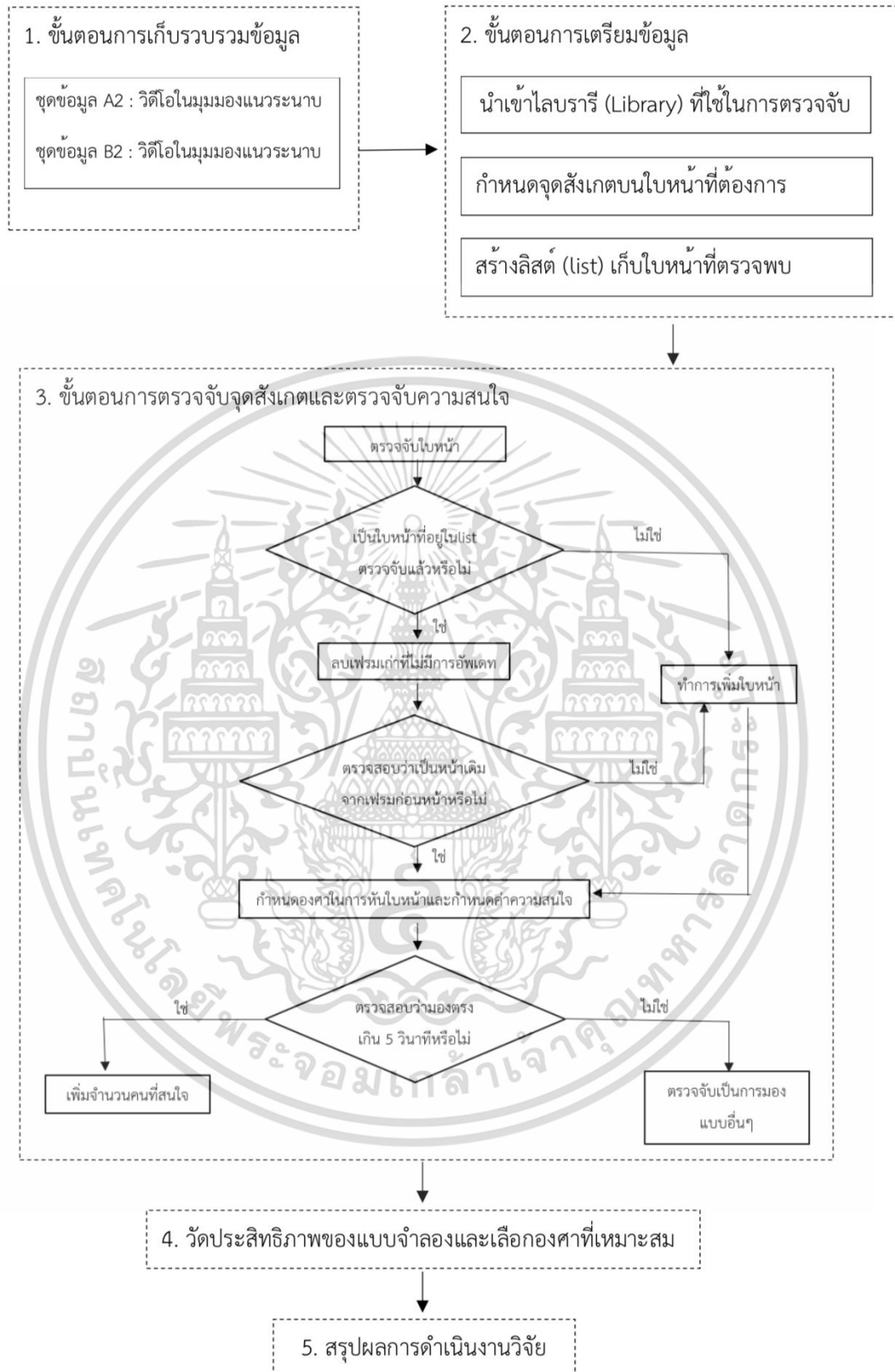
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.4.2 การตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

ระบบตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ เป็นระบบที่ตรวจจับจุดสังเกตบนใบหน้าด้วยแบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) โดยจะตรวจจับใบหน้าคนต่อเนื่องตามเวลา และนับจำนวนคนที่สนใจ ตามเงื่อนไขที่กำหนด ทำการทดสอบบนโปรแกรม Jupyter Notebook โดยมีขั้นตอนการดำเนินงานดังนี้

1. การเก็บรวบรวมข้อมูล เป็นขั้นตอนในการรวบรวมข้อมูลเพื่อนำมาทดสอบ หรือเป็นขั้นตอนในการเตรียมวิดีโอในมุมมองที่สนใจ ซึ่งมีข้อมูลในการทดสอบ 2 ชุด คือ ข้อมูลชุด A2 และข้อมูลชุด B2 โดยจะทดสอบในมุมมองแนวระนาบเพียงมุมมองเดียว (แสดงในหัวข้อ 3.5.2)
2. การเตรียมข้อมูล เป็นการเตรียมความพร้อมของข้อมูลก่อนนำข้อมูลไปทดสอบ (แสดงรายละเอียดใน 3.6.2)
  - 2.1 นำเข้าไลบรารี (Library) ที่ใช้ในการตรวจจับ
  - 2.2 กำหนดจุดสังเกตบนใบหน้าที่ต้องการตรวจจับ
  - 2.3 สร้างลิสต์ (list) เก็บใบหน้าที่ตรวจพบ
3. ขั้นตอนการตรวจจับจุดสังเกตและความสนใจ ใช้แบบจำลองมีเดียไปป์ เฟซ เมช ในการตรวจจับจุดสังเกตบนใบหน้า และกำหนดองศาในการหันใบหน้าที่ระดับ 10 15 และ 20 องศา เพื่อประเมินความสนใจ โดยใช้ไลบรารี `Math.dist()` ในการติดตามใบหน้าเพื่อดูการเคลื่อนไหว
4. วัดประสิทธิภาพของแบบจำลองและเลือกองศาที่เหมาะสมในการนำไปใช้เพื่อประเมินความสนใจ โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าวัดประสิทธิภาพ (F1-score) และพิจารณาจากกราฟ ROC
5. สรุปผลการดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 ขั้นตอนการดำเนินงานวิจัยหวัข้อการตรวจจับและวัดองศาหว่างใบหน้กากบกล้อง  
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อประเมินความสนใจ ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.5 การเก็บรวบรวมข้อมูล

เป็นการรวบรวมข้อมูลที่จะนำมาทำการทดสอบ ในที่นี้ผู้วิจัยได้แบ่งเป็น 2 หัวข้อ คือ 1.การตรวจจับและนับจำนวนบุคคล 2.การตรวจจับและวัดองศาห่าหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ในแต่ละหัวข้อประกอบด้วยชุดข้อมูลทดสอบ A1 และ B1 และข้อมูลทดสอบ A2 และ B2 ตามลำดับ โดยในการตรวจจับและนับจำนวนบุคคลจะมีทั้งข้อมูลในมุมมองแนวระนาบ และมุมมองแนวเหนือศีรษะ และในการตรวจจับและวัดองศาห่าหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ จะมีเพียงข้อมูลในแนวระนาบเท่านั้น แสดงรายละเอียดดังนี้

#### 3.5.1 การตรวจจับและนับจำนวนบุคคล

1. ชุดข้อมูลทดสอบ A1 เป็นชุดข้อมูลที่บันทึกด้วยตนเอง ซึ่งจะประกอบด้วย 2 มุมมอง ดังที่กล่าวข้างต้น โดยมีมุมมองละ 1 วิดีโอ ความยาววิดีโอละ 5 นาที โดยข้อมูลชุดนี้ทำการบันทึกในพื้นที่ห้างสรรพสินค้าแห่งหนึ่งในบริเวณบริเวณบันไดเลื่อน และประตูทางเข้าของห้างสรรพสินค้า แสดงดังรูปที่ 3.4(ก) และ 3.5(ก) ตามลำดับ

2. ชุดข้อมูลทดสอบ B1 ได้จากวิดีโอที่บันทึกเองในมุมมองที่สนใจ แบ่งเป็น 2 มุมมอง

- มุมมองแนวเหนือศีรษะ (ในกรณีนี้ถ่ายคนที่ยืนอยู่บนบันไดเลื่อน โดยผู้บันทึกวิดีโอ ยืนอยู่ด้านบนบันไดเลื่อน) โดยความยาวของวิดีโอคือ 5 นาที/วิดีโอ ในมุมมองนี้ทำการบันทึกทั้งหมด 3 วิดีโอ โดยพื้นที่ในการบันทึกคือ บริเวณบันไดเลื่อนของสถานีรถไฟฟ้าแห่งหนึ่ง แสดงดังรูปที่ 3.4(ข)



(ก) ข้อมูลชุดทดสอบ A1

(ข) ข้อมูลชุดทดสอบ B1

รูปที่ 3.4 ตัวอย่างเฟรมที่ถ่ายในมุมมองแนวเหนือศีรษะ

- มุมมองแนวระนาบ (ในกรณีนี้ถ่ายวิดีโอโดยยืนแนวเดียวกับผู้ที่ถูกถ่าย) โดยความยาวของวิดีโอคือ 5 นาที/วิดีโอ และทำการบันทึกทั้งหมด 3 วิดีโอ เช่นเดียวกัน โดยพื้นที่ในการบันทึกของข้อมูลชุด B1 ในแนวระนาบ คือ บริเวณทางเดินเข้า-ออกของสถานีรถไฟฟ้าแห่งหนึ่ง แสดงดังรูปที่ 3.5(ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก) ข้อมูลชุดทดสอบ A1

(ข) ข้อมูลชุดทดสอบ B1

รูปที่ 3.5 ตัวอย่างเฟรมที่ถ่ายในมุมมองแนวระนาบ

### 3.5.2 การตรวจจับและวัดองศาห่างไบหน้ากับกัองเพื่อประเมินความสนใจ

1. ชุดข้อมูลทดสอบ A2 เป็นชุดข้อมูลที่บันทึกเอง ซึ่งจะประกอบด้วยข้อมูลทดสอบจำนวน 10 วิดีโอ
2. ชุดข้อมูลทดสอบ B2 ได้จากชุดข้อมูลที่บันทึกด้วยตนเองและนำเข้าจากอินเทอร์เน็ตบางส่วน โดยวิดีโอที่ได้ทำการบันทึกด้วยตนเองจำนวน 8 วิดีโอ ซึ่งบันทึกด้วยกล้องโทรศัพท์ เช่น ให้บุคคลเดินและหยุดยืนหน้ากล้อง หันซ้าย-ขวา เป็นต้น และวิดีโอที่นำเข้าจากอินเทอร์เน็ตจำนวน 2 วิดีโอ รวมทั้งหมด 10 วิดีโอ เช่นเดียวกับชุดข้อมูล A2



รูปที่ 3.6 ตัวอย่างป้ายโฆษณาดิจิทัลขนาดเล็ก

โดยการตรวจจับและวัดองศาห่างไบหน้ากับกัองเพื่อประเมินความสนใจใช้กับป้ายโฆษณาดิจิทัลขนาดเล็ก ดังรูปที่ 3.6 ซึ่งจะมีขนาดของจออยู่ที่ประมาณ 7 – 10 นิ้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.6 ขั้นตอนการเตรียมข้อมูลและการวิเคราะห์ข้อมูล

เป็นการเตรียมความพร้อมของข้อมูลที่จะนำมาทำการทดสอบ โดยในแต่ละหัวข้อประกอบด้วย ชุดข้อมูลทดสอบ A1 B1 และข้อมูลทดสอบ A2 B2 ตามลำดับ โดยในแต่ละชุดข้อมูลทดสอบนี้ แสดง ขั้นตอนการเตรียมข้อมูล ดังนี้

#### 3.6.1 การตรวจจับและนับจำนวนบุคคล (ชุดข้อมูลที่ผ่านการฝึกสอนแล้วของโยโล่ เวอร์ชัน 7)

1. เมื่อได้ข้อมูลทดสอบชุด B1 มาแล้ว นำวิดีโอที่ได้มาทำการแตกเฟรมให้เป็นรูปภาพ โดยกำหนดการแตกเฟรมเป็น 1 fps (1เฟรม/วินาที) คือ 1 นาทีได้ 60 เฟรม ซึ่งใน 1 วิดีโอมีความยาว 5 นาที จะได้เฟรมทั้งหมด 300 เฟรม

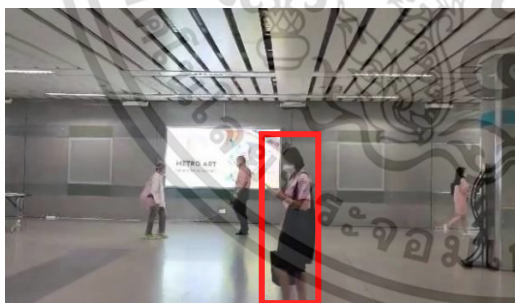
2. สุ่มเฟรมขึ้นมาเพื่อทำการทดสอบ ในที่นี้ใช้วิธีการสุ่มตัวอย่างแบบเจาะจง (Purposive Sampling) ซึ่งเป็นการสุ่มตัวอย่างแบบไม่ใช้ความน่าจะเป็น โดยการนับไปครั้งละ 5 เฟรมหรือนับไป 5 วินาที แล้วสุ่มขึ้นมา จะได้จำนวนเฟรมที่นำไปทดสอบ 60 เฟรม/วิดีโอ



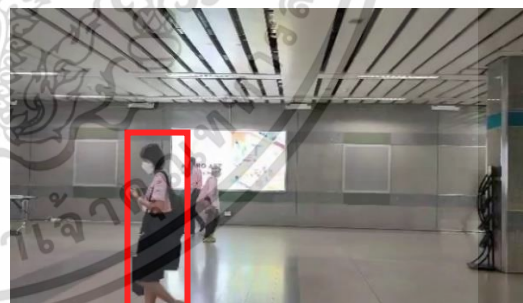
รูปที่ 3.7 แสดงการเคลื่อนไหวในวินาทีที่ 1



รูปที่ 3.8 แสดงการเคลื่อนไหวในวินาทีที่ 2



รูปที่ 3.9 แสดงการเคลื่อนไหวในวินาทีที่ 3



รูปที่ 3.10 แสดงการเคลื่อนไหวในวินาทีที่ 4



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับนำไปใช้งานเพื่อการศึกษานั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้รูปที่ 3.11 แสดงการเคลื่อนไหวในวินาทีที่ 5

เอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.7 – 3.11 จะเห็นได้ว่าบุคคลหนึ่งใช้เวลาประมาณ 5 วินาที ในการเดินเข้า-ออก จากเฟรมกล้อง และเพื่อลดการนับคนแบบซ้ำซ้อน จึงใช้การสุ่มโดยการนับครั้งละ 5 เฟรม

### 3.6.2 การตรวจจับและวัดองศาห่าหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

#### 1. นำเข้าไลบรารี (Library) ที่ใช้ในการตรวจจับ

นำเข้า Library MediaPipe และ library อื่น ๆ พร้อมทั้งกำหนดพารามิเตอร์ต่างๆ ที่จำเป็นในการวิเคราะห์ ดังนี้

```
import cv2
import mediapipe as mp
import numpy as np
import math

mp_face_mesh = mp.solutions.face_mesh
face_mesh = mp_face_mesh.FaceMesh(min_detection_confidence=0.5, min_tracking_confidence=0.5,max_num_faces=50)
```

#### รูปที่ 3.12 แสดงไลบรารี (Library) ที่นำเข้า และกำหนดพารามิเตอร์

ไลบรารี (Library) ที่เกี่ยวข้อง

- Library OpenCV

เป็นไลบรารีที่ใช้ในการทำงานกับภาพและวิดีโอ ซึ่งมีฟังก์ชันที่สามารถอ่านและแสดงภาพได้ นอกจากนี้ยังมีฟังก์ชันที่เกี่ยวข้องกับการประมวลผลภาพ เช่น การแปลงสี การตัดภาพ และอื่น ๆ

- Library MediaPipe

เป็นไลบรารีที่พัฒนาขึ้นโดย Google ที่มุ่งเน้นการวิเคราะห์และการประมวลผลสื่อในเวลาเรียลไทม์ ในที่นี้ใช้ในการตรวจจับจุดสังเกตบนใบหน้า โดยมีคลาส mp.FaceMesh() เพื่อสร้างแบบจำลองตาข่ายใบหน้า (Face Mesh) โดยสามารถกำหนดพารามิเตอร์ได้ดังนี้

- min\_detection\_confidence ค่าความเชื่อมั่นขั้นต่ำ ([0.0, 1.0]) จากแบบจำลองการตรวจจับใบหน้าเพื่อให้ถือว่าการตรวจจับสำเร็จ ค่าความเชื่อมั่นเป็น 0.5

- min\_tracking\_confidence ค่าความเชื่อมั่นขั้นต่ำ ([0.0, 1.0]) จากแบบจำลองการติดตามจุดสังเกตสำหรับจุดสังเกตใบหน้าที่จะถือว่าการติดตามสำเร็จ มิฉะนั้น ระบบจะเรียกใช้การตรวจจับใบหน้าโดยอัตโนมัติในภาพอินพุต (Input) ถัดไป การตั้งค่าเป็นค่าที่สูงกว่าจะช่วยเพิ่มความแม่นยำของกระบวนการได้ โดยที่การตรวจจับใบหน้าจะทำงานในทุกภาพ ค่าเริ่มต้นเป็น 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- `max_num_faces` ใช้ในการกำหนดจำนวนสูงสุดของใบหน้าที่ถูกตรวจจับได้ใช้ในการจำกัดจำนวนใบหน้าที่จะตรวจจับในแต่ละเฟรมของวิดีโอหรือภาพเพื่อเพิ่มประสิทธิภาพ และความเร็วในการประมวลผล เมื่อค่า `max_num_faces` ถูกตั้งค่าเป็นจำนวนใด จำนวนหน้าที่มากกว่าจำนวนที่กำหนดจะไม่ถูกตรวจจับและส่งผลลัพธ์กลับมา ในที่นี้กำหนดจำนวนใบหน้าที่สามารถตรวจจับได้มากที่สุดถึง 50 ใบหน้า

- ◉ Library Numpy

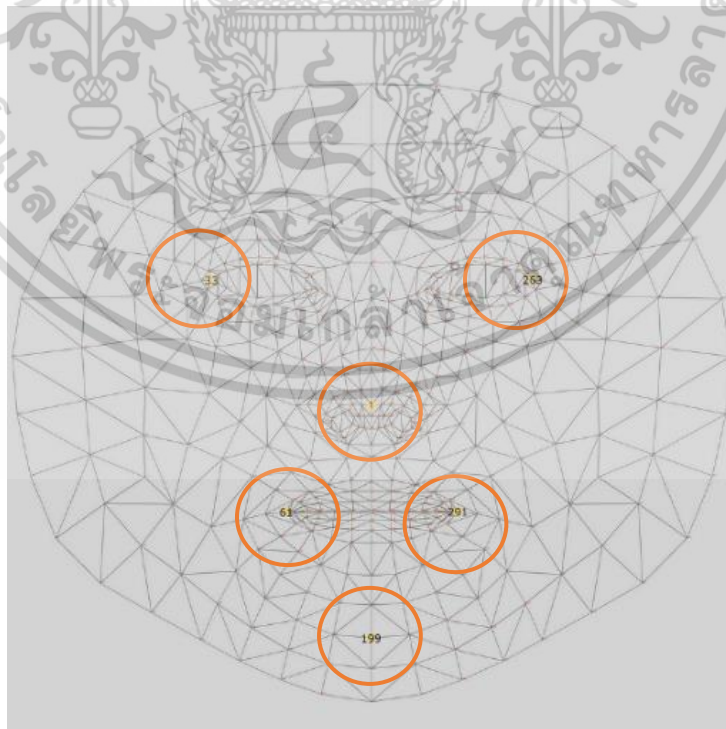
เป็นไลบรารีที่ใช้ในการทำงานกับอาร์เรย์หลายมิติ (multidimensional arrays) ในที่นี้สามารถใช้ในการจัดการข้อมูลเชิงตารางที่ได้จากการตรวจจับจุดสังเกตบนใบหน้า

- ◉ Library Math

เป็นโมดูลที่ให้ฟังก์ชันทางคณิตศาสตร์เพื่อการคำนวณทางคณิตศาสตร์ เช่น ฟังก์ชันคำนวณค่าสแควร์รูท ค่า `sin`, `cos`, `tan` และอื่น ๆ ในที่นี้ใช้สำหรับการติดตามใบหน้า (Facial Tracking)

## 2. กำหนดจุดสังเกตบนใบหน้าที่ต้องการตรวจจับ

ในแบบจำลองที่เสนอ จะมีการเลือกจุดสังเกตหลักเพียง 6 จุดจากจุดสังเกตที่พบทั้งหมด 468 จุด แสดงดังรูปที่ 3.13 และตารางที่ 3.3 อธิบายดัชนีจุดสังเกตมีเดียไปป์ที่สอดคล้องกัน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ **รูปที่ 3.13** เลือก มีเดียไปป์ เฟซ เมช 6 จุดหน้าไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา : Khalid, (2021)เจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 ดัชนีของจุดสังเกตบนใบหน้าที่เราสนใจ ซึ่งกระจายบนใบหน้าเท่า ๆ กัน

| ดัชนีจุดสังเกต (idx) | คำอธิบาย       |
|----------------------|----------------|
| 1                    | จมูก           |
| 33                   | หางตาฝั่งขวา   |
| 263                  | หางตาฝั่งซ้าย  |
| 61                   | มุมปากฝั่งขวา  |
| 291                  | มุมปากฝั่งซ้าย |
| 199                  | คาง            |

```
if idx == 33 or idx == 263 or idx == 1 or idx == 61 or idx == 291 or idx == 199:
    if idx == 1:
```

รูปที่ 3.14 แสดงการกำหนดจุดสังเกตบนใบหน้าในโปรแกรม

โดยกำหนดจุดสังเกตบนใบหน้า จุดที่ 1 หรือ จุดจมูก เป็นจุด p1 ซึ่งกำหนดไว้เพื่อนำไปใช้ในการติดตามใบหน้าว่าเป็นใบหน้าเดิมหรือไม่ ดังรูปที่ 3.15

```
p1 = (int(nose_2d[0]), int(nose_2d[1]))
```

รูปที่ 3.15 แสดงการกำหนด p1

3. สร้างลิสต์ (list) ไว้เก็บใบหน้าที่ตรวจจับได้ และกำหนดตัวแปรที่เกี่ยวข้อง

```
face_id = []
```

รูปที่ 3.16 แสดงการสร้าง list สำหรับตรวจจับใบหน้า

```
count = 0
frame_count = 0
Attention = 0
```

รูปที่ 3.17 แสดงการกำหนดตัวแปรที่เกี่ยวข้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 3.17 แสดงการกำหนดตัวแปรที่เกี่ยวข้อง ดังนี้

- count คือ ตัวแปรที่เก็บจำนวนเฟรมในแต่ละลักษณะการมอง เช่น เมื่อเริ่มมองตรงตัวแปร count จะเริ่มนับว่ามองตรงผ่านไปกี่เฟรมแล้ว
- frame\_count คือ ตัวแปรไว้สำหรับเก็บจำนวนเฟรมโดยเริ่มนับตั้งแต่เริ่มต้นวิดีโอจนจบวิดีโอ
- Attention คือ ตัวแปรสำหรับเก็บจำนวนคนที่สนใจ

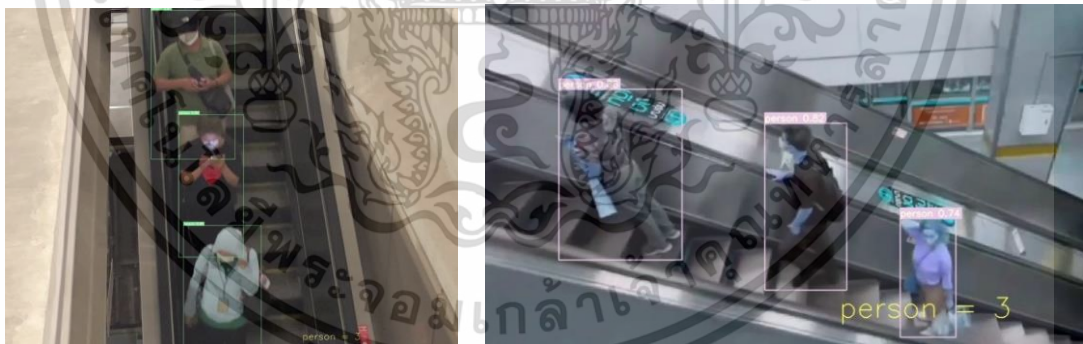
โดยวิดีโอในหัวข้อนี้ มีความเร็วในการบันทึก 30 เฟรม/วินาที

### 3.7 การทดสอบแบบจำลอง

#### 3.7.1 การตรวจจับและนับจำนวนบุคคล

ใช้แบบจำลองโยโล เวอร์ชัน 7 โดยกำหนดค่า `conf_thres=0.3`

Confidence Threshold คือ เกณฑ์ความเชื่อมั่น หรือค่าที่กำหนดให้เป็นค่าความเชื่อถือ (Confidence) ขั้นต่ำของวัตถุที่ตรวจจับในแบบจำลองโยโล เวอร์ชัน 7 มีค่าตั้งแต่ 0 – 1 ในกรณีนี้ ค่า `conf_thres = 0.3` หมายถึงว่า เฉพาะคนที่มีความเชื่อถือ มากกว่าหรือเท่ากับ 0.3 เท่านั้นจึงจะถูกตรวจพบและถูกตีความว่าเป็นวัตถุในภาพ วัตถุที่มีความเชื่อถือต่ำกว่า 0.3 จะไม่ถูกตรวจพบในกระบวนการตรวจจับและจะไม่ถูกนับ เพื่อป้องกันการนับคนที่ยืนซ้อน และเห็นเพียงแค่เล็กน้อยเท่านั้น



(ก) ชุดข้อมูลทดสอบ A1

(ข) ชุดข้อมูลทดสอบ B1

รูปที่ 3.18 ตัวอย่างเฟรมที่ตรวจจับด้วย YOLOv7 แบบมุมมองแนวเหนือศีรษะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก) ชุดข้อมูลทดสอบ A1

(ข) ชุดข้อมูลทดสอบ B1

รูปที่ 3.19 ตัวอย่างเฟรมที่ตรวจจับด้วย YOLOv7 มุมมองแนวระนาบ

```
!python /content/drive/MyDrive/YOLOv7-object-detect-count-blur-crop/detect_and_counting_total.py
--weights yolov7.pt --conf 0.3 --source inference/sampling/sam --classes 0
```

รูปที่ 3.20 ตัวอย่างคำสั่งในการการตรวจจับและนับจำนวนบุคคล

- YOLOv7-object-detect-count-blur-crop/detect\_and\_counting\_total.py เป็นโค้ดที่ใช้สำหรับการตรวจจับและนับจำนวนคน
- weights yolov7.pt นำเข้าไฟล์ .pt จากที่ได้ดาวน์โหลดมาก่อนหน้าในหัวข้อ 3.3.2
- conf 0.3 เป็นการกำหนดค่าความเชื่อถือ เฉพาะคนที่มีความเชื่อถือมากกว่า หรือเท่ากับ 0.3 เท่านั้นจึงจะถูกตรวจพบว่าเป็นวัตถุในภาพ วัตถุที่มีค่าความเชื่อถือว่าต่ำกว่า 0.3 จะไม่ถูกตรวจพบในกระบวนการตรวจจับและจะไม่ถูกนับ
- source inference/sampling/sam
- classes 0 เป็นการบอกว่าจะโฟกัสที่คลาส 0 คลาสเดียว ซึ่งคลาส 0 คือ มนุษย์ (Balon, 2022)หมายความว่า จะตรวจจับมนุษย์เพียงอย่างเดียว

```
6 persons, Done. (16.0ms) Inference, (1.4ms) NMS
The image with the result is saved in: runs/detect/exp36/140.jpg
Total counted objects by every frame: 266
Total objects in the current frame: 6
6 persons, Done. (16.0ms) Inference, (1.5ms) NMS
The image with the result is saved in: runs/detect/exp36/141.jpg
Total counted objects by every frame: 272
Total objects in the current frame: 6
6 persons, Done. (16.0ms) Inference, (1.3ms) NMS
The image with the result is saved in: runs/detect/exp36/142.jpg
Total counted objects by every frame: 277
Total objects in the current frame: 5
5 persons, Done. (16.0ms) Inference, (1.8ms) NMS
The image with the result is saved in: runs/detect/exp36/143.jpg
Total counted objects by every frame: 281
Total objects in the current frame: 4
4 persons, Done. (16.0ms) Inference, (1.5ms) NMS
The image with the result is saved in: runs/detect/exp36/144.jpg
Total counted objects by every frame: 286
Total objects in the current frame: 5
5 persons, Done. (16.0ms) Inference, (1.5ms) NMS
The image with the result is saved in: runs/detect/exp36/145.jpg
Total counted objects by every frame: 298
Total objects in the current frame: 12
```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ โดยสถาบันวิจัยระบบสารสนเทศ  
รูปที่ 3.21 ตัวอย่างผลการทดสอบการตรวจจับและนับจำนวนคน โยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยสามารถศึกษาเพิ่มเติมได้จาก : <https://github.com/Zeeshan1/YOLOv7-object-detect-count-blur-crop>

### 3.7.2 การตรวจจับและวัดองศาห่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

จากการเตรียมข้อมูลในหัวข้อ 3.6.2 ถัดมาเป็นการทดสอบแบบจำลอง ดังนี้

#### 1. ตรวจจับใบหน้าในแต่ละเฟรม

ถ้าพบใบหน้าที่อยู่ในลิสต์ (list) แล้ว อันดับแรกจะทำการลบเฟรมเก่าออก การทำงานคือ ถ้าไม่มีการอัปเดต frame count จะทำการลบ frame count ที่เก่ากว่า 150 frame count ให้เหลือแต่ frame count ภายใน 150 frame count เพื่อลดการซ้ำซ้อนของข้อมูล

```
if face_id:
    for frame in face_id:
        if frame_count - frame["frame_count"] > 150 :
            face_id.remove(frame)
```

รูปที่ 3.22 แสดงการลบ frame count ที่เก่ากว่า 150 frame count

ถ้าพบใบหน้าในเฟรม ขั้นตอนถัดไป คือ การติดตามใบหน้า เป็นตามขั้นตอนที่ 6 ถ้าไม่พบใบหน้าจะทำการเพิ่มใบหน้าเข้าไปในลิสต์ (list) เพื่อเก็บไว้ตรวจสอบกับเฟรมถัดไปว่าเป็นใบหน้าเดียวกันกับใบหน้าก่อนหน้าหรือไม่

```
else:
    face_id.append({"p1":p1,"frame_count":frame_count,"count":0,"Attention":False})
```

รูปที่ 3.23 แสดงการเพิ่มใบหน้าที่ยังไม่เคยตรวจพบมาก่อน

โดย append() เป็นเมธอด (Method) ในภาษาไพธอนที่ใช้สำหรับเพิ่มข้อมูลลงในสมาชิกของลิสต์ (List) หรือในตำแหน่งสุดท้ายของอาร์เรย์ (Array) โดยการใช้ .append() จะทำให้ข้อมูลถูกเพิ่มเข้าไปในตำแหน่งสุดท้ายของลิสต์หรืออาร์เรย์ที่กำหนด ในที่นี้จะทำการเพิ่มดังต่อไปนี้

- จุดตรวจจับที่จมูก (p1)
- frame count
- count = 0 คือ จำนวนเฟรมในแต่ละลักษณะการมอง โดยตั้งให้เริ่มต้นการนับที่ 0
- Attention = False ตั้งเป็นค่าเริ่มต้น (Default) หมายถึง ใบหน้าที่ยังไม่เคยถูกตรวจจับว่าสนใจมาก่อน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. การติดตามใบหน้าที่เคลื่อนไหว (Facial Tracking)

เมื่อพบใบหน้าในเฟรม ถัดมาคือการติดตามใบหน้า (Facial Tracking) เป็นการหาตำแหน่งของใบหน้าในแต่ละเฟรม สำหรับในวิดีโอจะประกอบด้วยเฟรมหลายเฟรมมาต่อกัน จึงมีขั้นตอนเพื่อตรวจสอบว่าเป็นใบหน้าเดิมจากเฟรมก่อนหน้าหรือไม่ โดยใช้ `math.dist()` เป็นเมธอด (method) ในการติดตามใบหน้า โดยใช้จุดตรงมุม(จุดp1) ในการติดตาม โดยการทำงานคือ จะส่งกลับระยะทางแบบยูคลิดระหว่างจุดสองจุด (p และ q) โดยที่ p และ q เป็นพิกัดของจุดมุมนั้น ๆ

กำหนดค่าพิกัด x และ y ของจุด P และ Q แล้วคำนวณหาระยะห่างแบบยูคลิด

(Euclidean) ระหว่างจุด P และ Q โดยใช้ `math.dist([Px, Py], [Qx, Qy])`

พารามิเตอร์:

$P_x, P_y$ : ลำดับหรือการวนซ้ำของพิกัดที่แสดงถึงจุดแรก

$Q_x, Q_y$ : ลำดับหรือพิกัดที่วนซ้ำได้ซึ่งเป็นตัวแทนของจุดที่สอง

ส่งค่ากลับ (return) : ระยะทางแบบยูคลิดที่คำนวณได้ระหว่างจุดที่กำหนดให้

$$d = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2}$$

โดยกำหนดว่า ถ้ารัศมีของจุดมุมจุดเก่า และจุดมุมจุดใหม่ ห่างกันไม่เกิน 50 พิกเซล

(pixel) จะทำการอัปเดต (update) จุดมุม(p1) และอัปเดต frame count ซึ่งการทำงานคือ จะนำจุดมุม(p1) และ frame count ในปัจจุบันแทนที่ค่าเดิม แต่ถ้ารัศมีของจุดมุมจุดเก่า และจุดมุมจุดใหม่ ห่างกันเกิน 50 พิกเซล จะถือว่าเป็นอีกคนหนึ่ง และจะทำการเพิ่มใบหน้าเข้าไปในลิสต์ (list) ดังที่กล่าวไว้ข้างต้น ในการติดตามใบหน้าด้วยฟังก์ชัน `math.dist()` แสดงโค้ดดังรูปที่ 3.24

```
for fid in face_id:
    if math.dist(fid["p1"],p1) < 50:
        fid["p1"] = p1
        fid["frame_count"] = frame_count
        break
    else:
        index+=1
if index >= len(face_id):
    face_id.append({"p1":p1,"frame_count":frame_count,"count":0,"Attention":False})
```

รูปที่ 3.24 แสดงการติดตามใบหน้าด้วยฟังก์ชัน `math.dist()`

ซึ่งจะมีเกณฑ์ในการทดสอบดังนี้

เอกสารนี้เป็นเอกสารที่ 3. เกณฑ์ในการกำหนดดวงศการหันของใบหน้า นั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้วิจัยได้กำหนดองศาในการทดสอบไว้ 3 ระดับคือ 10 15 และ 20 องศา โดยจะแสดงการกำหนด 10 องศา ดังนี้

หันหน้าไปทางซ้าย > 10 องศา จะให้แสดงคำว่า “Looking Left”

หันหน้าไปทางขวา > 10 องศา จะให้แสดงคำว่า “Looking Right”

เงยหน้า > 10 องศา จะให้แสดงคำว่า “Looking Up”

ก้มหน้า > 10 องศา จะให้แสดงคำว่า “Looking Down”

ถ้านอกเหนือจากเงื่อนไขดังกล่าว จะขึ้นคำว่า “Forward” คือ มองตรง

กำหนดให้มีการนับโดยตัวแปร count เมื่อขึ้นคำว่า “Forward”

#### 4. เกณฑ์ในการกำหนดจำนวนวินาทีของความสนใจ

เมื่อขึ้นคำว่า “Forward” ตัวแปร count จะนับเฟรมไปและเมื่อเกิน 150 เฟรมหรือ 5 วินาที จะทำการบวกเพิ่มจำนวนคนที่สนใจ ในตัวแปร Attention โดยมีการตรวจสอบว่าถ้าบุคคลนั้นยังไม่เคยถูกเพิ่มในตัวแปร Attention มาก่อน ก็จะต้องเพิ่มบุคคลนั้นเข้าไปในจำนวนคนที่สนใจ (Attention) ดังแสดงในรูปที่ 3.21

```

if y < -10:
    text = "Looking Left"
    face_id[index]["count"] = 0
elif y > 10:
    text = "Looking Right"
    face_id[index]["count"] = 0
elif x < -10:
    text = "Looking Down"
    face_id[index]["count"] = 0
elif x > 10:
    text = "Looking Up"
    face_id[index]["count"] = 0
else:
    if face_id[index]["count"] > 150:
        text = "Attention"
        if not face_id[index]["Attention"]:
            Attention+=1
            face_id[index]["Attention"] = True
    else:
        text = "Forward"
        face_id[index]["count"]+=1

```

รูปที่ 3.25 แสดงการกำหนดองศาในการหันของใบหน้า

ในการกำหนดความสนใจที่ 5 วินาที เนื่องจากในกรณีศึกษาโฆษณาที่มีความยาวประมาณ 10-30 วินาที ทางผู้วิจัยสนใจเลือกวิดีโอที่สั้นที่สุดคือ 10 วินาที และคิดว่าความสนใจอย่างน้อยใน 10 วินาที ถ้าบุคคลมองอย่างน้อยครึ่งหนึ่งของวิดีโอจะพิจารณาว่าบุคคลนั้นมีความสนใจในวิดีโอ จึงเป็นเหตุผลในการกำหนดว่าเมื่อมองตรงเกิน 5 วินาทีจะให้ตรวจจับว่าบุคคลนั้นเป็นคนที่สนใจ

เมื่อทดสอบ 10 องศาเรียบร้อยแล้วจะทำการทดสอบที่ 15 องศา และ 20 องศา ตามลำดับ เอกสารนี้เช่นเดียวกับการทดสอบ 10 องศาข้างต้น เพื่อเลือกองศาที่เหมาะสมที่สุดในการนำไปใช้เพื่อประเมินว่าไม่ว่าการความสนใจของบุคคลนั้นมีให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.8 การประเมินประสิทธิภาพของแบบจำลอง

เมทริกซ์ความสับสน (Confusion Matrix) เป็นหนึ่งในวิธีที่ใช้ในการประเมินประสิทธิภาพของระบบหรือแบบจำลองทางสถิติ มักนิยมใช้ในงานที่เกี่ยวข้องกับการจำแนกหรือการทำนาย โดยพิจารณาจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าวัดประสิทธิภาพ (F1-score) และพิจารณาจากกราฟ ROC ในการพิจารณาค่าดังกล่าว ในที่นี้จะสนใจค่าความถูกต้อง (Accuracy) ของแบบจำลอง เนื่องจากต้องการเปรียบเทียบกับเครื่องมือการตรวจจับบุคคลที่ใช้อยู่ในปัจจุบันของบริษัทการศึกษา โดยค่าอื่นหามาเพื่อประกอบการตัดสินใจว่าแบบจำลองสามารถทำนายได้ดีหรือไม่ดี

ในเมทริกซ์ความสับสนเป็นตารางการวัดประสิทธิภาพของระบบโดยการจำแนกประเภท ซึ่งจะประกอบด้วยค่าพยากรณ์ (Predicted Value) ที่แบบจำลองทำนาย และค่าจริง (True Value) ในงานวิจัยนี้มีการกำหนดการนับค่าจริง ดังนี้

**3.8.1 การตรวจจับและนับจำนวนบุคคล** แบ่งการทดสอบเป็น 2 มุมมอง คือ มุมมองแนวระนาบ และมุมมองแนวเหนือศีรษะ

- ในการตรวจจับและนับบุคคล กรณีมุมมองแนวเหนือศีรษะ จะโฟกัสที่ส่วนบนเป็นต้นไป ถ้าเห็นเหนือเอวขึ้นมาเต็ม ๆ จะนับเป็น 1 คน
- ในการตรวจจับและนับบุคคล กรณีมุมมองแนวระนาบ จะนับเป็น 1 คน ก็ต่อเมื่อเห็นคนเต็มทั้งตัว โดยไม่โดนบดบัง

**3.8.2 การตรวจจับและวัดองศาห้วงใบหน้ากับกล้องเพื่อประเมินความสนใจ**

- ในการตรวจจับความสนใจจะนับจำนวนคนที่สนใจ ก็ต่อเมื่อมองตรงนานเกิน 5 วินาที
- ในการตรวจจับความสนใจจะนับจำนวนคนที่สนใจ เมื่อหันใบหน้าในองศาที่กำหนดเท่านั้น โดยได้กำหนดองศาในการหันใบหน้าเพื่อดูความสนใจที่ 3 ระดับ คือ 10 15 และ 20 องศา และทำการทดสอบทั้ง 3 ระดับ เพื่อดูว่าการหันใบหน้าที่ด้วยองศาใดแทนความสนใจได้ดีที่สุด และนำองศาที่นั้นไปวัดผลประสิทธิภาพแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### ผลการวิจัยและการอภิปรายผล

จากการวิเคราะห์ข้อมูลในการศึกษาเกี่ยวกับแบบจำลองการตรวจจับบุคคล นับจำนวนบุคคล ผู้วิจัยเลือกใช้แบบจำลองโยโล เวอร์ชัน 7 (YOLOv7) ที่ได้ผ่านการฝึกสอนมาแล้ว และการศึกษาเกี่ยวกับการประเมินความสนใจของผู้คนที่เดินผ่านกล้อง ผู้วิจัยเลือกใช้แบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ซึ่งเป็นหนึ่งในแบบจำลองของไลบรารีมีเดียไปป์ (Library MediaPipe) ทั้ง 2 หัวข้อได้แบ่งข้อมูลทดสอบออกเป็น 2 ชุด คือข้อมูลชุด A และข้อมูลชุด B ข้อมูลที่นำมาทำการทดสอบเป็นข้อมูลที่บันทึกวิดีโอด้วยตนเอง ซึ่งในแต่ละชุดข้อมูล ประกอบด้วยมุมมองแนวเหนือศีรษะ และมุมมองแนวระนาบ แสดงผลการทดสอบประสิทธิภาพของแต่ละแบบจำลอง ดังนี้

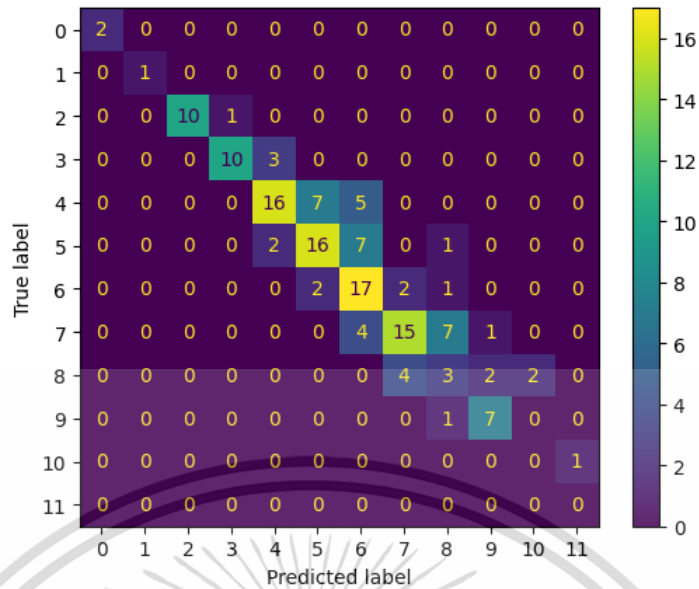
#### 4.1 ผลการทดสอบประสิทธิภาพของแบบจำลอง

ในงานวิจัยนี้ได้ทำการทดสอบประสิทธิภาพโดยหาค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าวัดประสิทธิภาพ จากค่า TN TP FN FP ในเมตริกซ์ความสับสนสำหรับการจำแนกประเภทหลายชั้น โดยแบ่งออกเป็น 2 แบบจำลอง ดังนี้

**4.1.1 การทดสอบประสิทธิภาพของแบบจำลองโยโล เวอร์ชัน 7** การตรวจจับและนับจำนวนบุคคล ได้แบ่งข้อมูลที่ใช้ในการทดสอบออกเป็น 2 ชุด และแต่ละชุดจำแนกเป็น 2 มุมมอง

1. มุมมองแนวระนาบ ได้ทำการทดสอบประสิทธิภาพ โดยมีข้อมูลในการทดสอบ 2 ชุด
    - ข้อมูลชุด A1 มีจำนวนเฟรมที่ใช้ในการทดสอบ 150 เฟรม โดยในเมตริกซ์ความสับสนของข้อมูลชุดนี้ จะแบ่งคลาสได้เป็น 12 คลาส คือ 0 1 2 3 4 5 6 7 เป็นต้น
      - คลาส 0 หมายถึง จำนวนคนที่ตรวจจับได้ 0 คน
      - คลาส 1 หมายถึง จำนวนคนที่ตรวจจับได้ 1 คน
      - คลาส 2 หมายถึง จำนวนคนที่ตรวจจับได้ 2 คน เป็นต้น
    - True label หมายถึง ค่าจริงที่นับได้
    - Predicted label หมายถึง ค่าพยากรณ์ที่แบบจำลองสามารถพยากรณ์ได้
- โดยค่าในตารางจะเป็นค่าความถี่ แสดงดังรูปที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 เมทริกซ์ความสับสนของชุดข้อมูล A1 ในมุมมองแนวระนาบ

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 1.0000    | 1.0000 | 1.0000   | 2       |
| 1            | 1.0000    | 1.0000 | 1.0000   | 1       |
| 2            | 1.0000    | 0.9091 | 0.9524   | 11      |
| 3            | 0.9091    | 0.7692 | 0.8333   | 13      |
| 4            | 0.7619    | 0.5714 | 0.6531   | 28      |
| 5            | 0.6400    | 0.6154 | 0.6275   | 26      |
| 6            | 0.5152    | 0.7727 | 0.6182   | 22      |
| 7            | 0.7143    | 0.5556 | 0.6250   | 27      |
| 8            | 0.2308    | 0.2727 | 0.2500   | 11      |
| 9            | 0.7000    | 0.8750 | 0.7778   | 8       |
| 10           | 0.0000    | 0.0000 | 0.0000   | 1       |
| 12           | 0.0000    | 0.0000 | 0.0000   | 0       |
| accuracy     |           |        | 0.6467   | 150     |
| macro avg    | 0.6226    | 0.6118 | 0.6114   | 150     |
| weighted avg | 0.6837    | 0.6467 | 0.6557   | 150     |

รูปที่ 4.2 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวระนาบ

จากรูปที่ 4.2 ค่าความถูกต้อง ค่าเฉลี่ยความแม่นยำ ค่าเฉลี่ยค่าความระลึก และค่าเฉลี่ยค่าวัดประสิทธิภาพ จากการคำนวณค่าเฉลี่ยแบบถ่วงน้ำหนัก (weighted avg) คือให้น้ำหนักในแต่ละคลาสเท่ากัน ซึ่งจะได้ค่าการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวระนาบ ดังนี้

ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดเท่ากับ 64.67%

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 68.37%

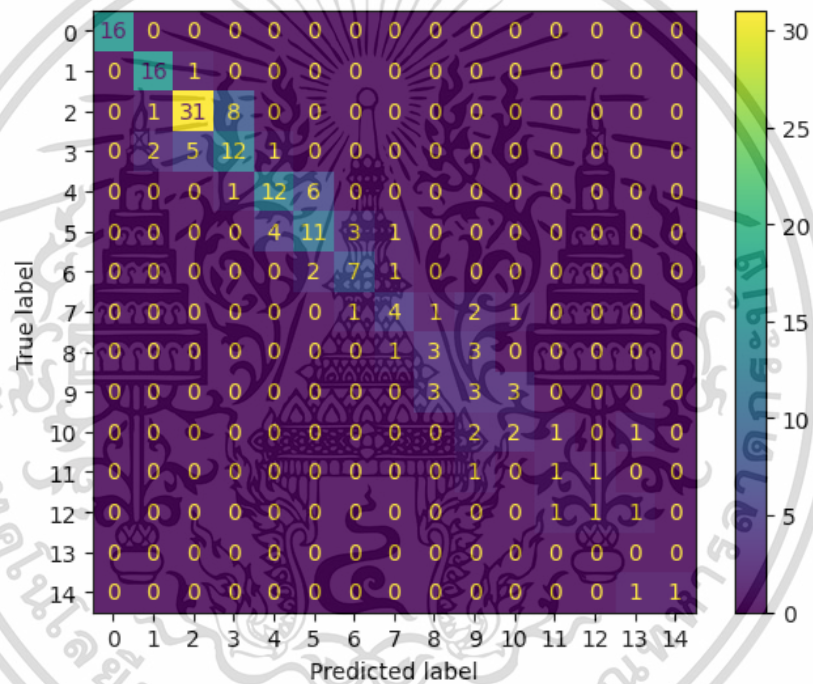
ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 64.67%

เอกสารนี้เป็นเอกสารที่สงวนไว้ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 65.57%

และจะเห็นได้ว่าค่าความแม่นยำในคลาสที่ 0 1 และ 2 มีค่าเท่ากับ 1.00 ซึ่งหมายความว่าแบบจำลองนี้สามารถทำนายคลาสที่มีคนอยู่ 0 1 และ 2 คนได้ดีที่สุด คือ 100% และสามารถตรวจจับคลาสที่มีคนอยู่ 0 และ 1 คนได้ดีที่สุดคือ มีค่าความระลึกเท่ากับ 1 หรือ 100% และตรวจจับคลาสที่มีคนอยู่ 2 คนได้รองลงมา โดยมีค่าความระลึกเท่ากับ 90.91%

- ข้อมูลชุด B1 มีจำนวนเฟรมที่ใช้ในการทดสอบ 180 เฟรม โดยในเมทริกซ์ความสับสนของข้อมูลชุดนี้ จะแบ่งคลาสได้เป็น 15 คลาส คือ คลาสที่ 0 ถึง 14 แสดงในรูปที่ 4.3



รูปที่ 4.3 เมทริกซ์ความสับสนของชุดข้อมูล B1 ในมุมมองแนวระนาบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 1 0          | 1.0000    | 1.0000 | 1.0000   | 16      |
| 1 1          | 0.8421    | 0.9412 | 0.8889   | 17      |
| 1 2          | 0.8378    | 0.7750 | 0.8052   | 40      |
| 1 3          | 0.5714    | 0.6000 | 0.5854   | 20      |
| 1 4          | 0.7059    | 0.6316 | 0.6667   | 19      |
| 1 5          | 0.5789    | 0.5789 | 0.5789   | 19      |
| 1 6          | 0.6364    | 0.7000 | 0.6667   | 10      |
| 1 7          | 0.5714    | 0.4444 | 0.5000   | 9       |
| 1 8          | 0.4286    | 0.4286 | 0.4286   | 7       |
| 1 9          | 0.2727    | 0.3333 | 0.3000   | 9       |
| 1 10         | 0.3333    | 0.3333 | 0.3333   | 6       |
| 1 11         | 0.3333    | 0.3333 | 0.3333   | 3       |
| 1 12         | 0.5000    | 0.3333 | 0.4000   | 3       |
| 1 13         | 0.0000    | 0.0000 | 0.0000   | 0       |
| 1 14         | 1.0000    | 0.5000 | 0.6667   | 2       |
| accuracy     |           |        | 0.6667   | 180     |
| macro avg    | 0.5741    | 0.5289 | 0.5436   | 180     |
| weighted avg | 0.6841    | 0.6667 | 0.6727   | 180     |

รูปที่ 4.4 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวระนาบ

จากรูปที่ 4.4 จะได้ค่าการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวระนาบ ดังนี้

ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดเท่ากับ 66.67%

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 68.41%

ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 66.67%

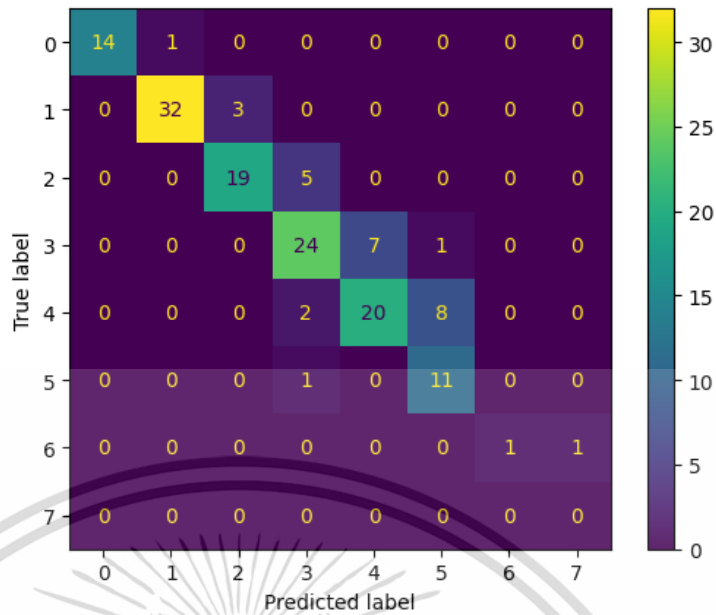
ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 67.27%

และจะเห็นได้ว่าค่าความแม่นยำในคลาสที่ 0 และ 14 มีค่าเท่ากับ 1.00 ซึ่งหมายความว่าแบบจำลองนี้สามารถทำนายคลาสที่มีคนอยู่ 0 และ 14 คนได้ดีที่สุด คือ 100% แต่ในคลาสที่ 14 อาจจะไม่บอกไม่ได้ว่าดีที่สุดในแง่ของในคลาสที่ 14 มีเพียง 1 เฟรมเท่านั้นที่มีคนอยู่ 14 คน และสามารถตรวจจับคลาสที่มีคนอยู่ 0 คนได้ดีที่สุดคือ มีค่าความระลึกเท่ากับ 1 หรือ 100% และตรวจจับคลาสที่มีคนอยู่ 1 คนได้รองลงมา โดยมีค่าความระลึกเท่ากับ 94.21%

2. มุมมองแนวเหนือศีรษะ ทำการทดสอบประสิทธิภาพโดยมีข้อมูลในการทดสอบ 2 ชุด

- ข้อมูลชุด A1 มีจำนวนเฟรมที่ใช้ในการทดสอบ 150 เฟรม โดยในเมทริกซ์ความสับสนของข้อมูลชุดนี้ จะแบ่งคลาสได้เป็น 8 คลาส คือ 0 1 2 3 4 5 6 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.5 เมทริกซ์ความสับสนของชุดข้อมูล A1 ในมุมมองแนวเหนือศีรษะ

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 1.0000    | 0.9333 | 0.9655   | 15      |
| 1            | 0.9697    | 0.9143 | 0.9412   | 35      |
| 2            | 0.8636    | 0.7917 | 0.8261   | 24      |
| 3            | 0.7500    | 0.7500 | 0.7500   | 32      |
| 4            | 0.7407    | 0.6667 | 0.7018   | 30      |
| 5            | 0.5500    | 0.9167 | 0.6875   | 12      |
| 6            | 1.0000    | 0.5000 | 0.6667   | 2       |
| 7            | 0.0000    | 0.0000 | 0.0000   | 0       |
| accuracy     |           |        | 0.8067   | 150     |
| macro avg    | 0.7343    | 0.6841 | 0.6923   | 150     |
| weighted avg | 0.8299    | 0.8067 | 0.8126   | 150     |

รูปที่ 4.6 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวเหนือศีรษะ

จากรูปที่ 4.6 จะได้ค่าการวัดประสิทธิภาพของชุดข้อมูล A1 ในมุมมองแนวเหนือศีรษะ ดังนี้

ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดเท่ากับ 80.67%

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 82.99%

ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 80.67%

ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 81.26%

และจะเห็นได้ว่าค่าความแม่นยำในคลาสที่ 0 และ 6 มีค่าเท่ากับ 1.00 ซึ่งหมายความว่า

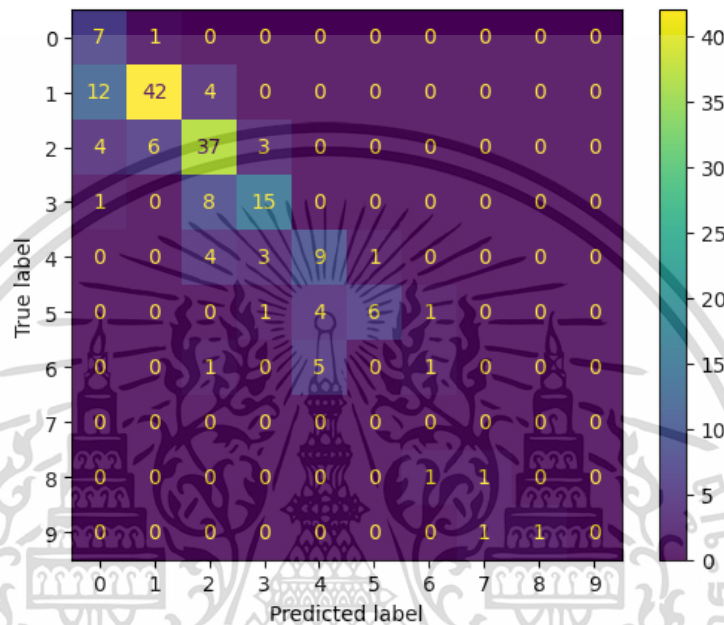
แบบจำลองนี้สามารถทำนายคลาสที่มีคนอยู่ 0 และ 6 คนได้ดีที่สุด คือ 100% แต่ในคลาสที่ 6 อาจจะ

บอกไม่ได้ว่าดีที่สุดใน เนื่องจากในคลาสที่ 6 มีเพียง 1 เฟรมเท่านั้นที่มีคนอยู่ 6 คน และในค่าความระลึก

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถตรวจจับคลาสที่มีคนอยู่ 0 5 และ 1 คนได้ดีในระดับใกล้เคียงกันคือ มีค่าความระลึกเท่ากับ 93.33% 91.67% 91.43% ตามลำดับ

- 2.2 ข้อมูลชุด B1 มีจำนวนเฟรมที่ใช้ในการทดสอบ 180 เฟรม โดยในเมทริกซ์ความสับสนของข้อมูลชุดนี้ จะแบ่งคลาสได้เป็น 10 คลาส คือ 0 ถึง 9 แสดงในตารางที่ 4.8



รูปที่ 4.7 เมทริกซ์ความสับสนของชุดข้อมูล B1 ในมุมมองแนวเหนือศีรษะ

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.2917    | 0.8750 | 0.4375   | 8       |
| 1            | 0.8571    | 0.7241 | 0.7850   | 58      |
| 2            | 0.6852    | 0.7400 | 0.7115   | 50      |
| 3            | 0.6818    | 0.6250 | 0.6522   | 24      |
| 4            | 0.5000    | 0.5294 | 0.5143   | 17      |
| 5            | 0.8571    | 0.5000 | 0.6316   | 12      |
| 6            | 0.3333    | 0.1429 | 0.2000   | 7       |
| 7            | 0.0000    | 0.0000 | 0.0000   | 0       |
| 8            | 0.0000    | 0.0000 | 0.0000   | 2       |
| 9            | 0.0000    | 0.0000 | 0.0000   | 2       |
| accuracy     |           |        | 0.6500   | 180     |
| macro avg    | 0.4206    | 0.4136 | 0.3932   | 180     |
| weighted avg | 0.6877    | 0.6500 | 0.6555   | 180     |

รูปที่ 4.8 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวเหนือศีรษะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.8 จะได้ค่าการวัดประสิทธิภาพของชุดข้อมูล B1 ในมุมมองแนวเหนือศีรษะ ดังนี้  
 ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดเท่ากับ 65.00%  
 ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 68.77%  
 ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 65.00%  
 ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 65.55%

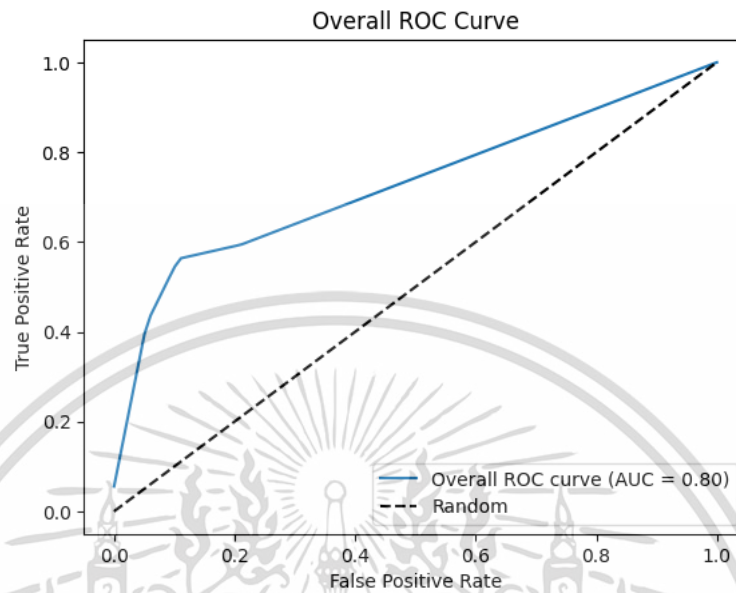
และจะเห็นได้ว่าค่าความแม่นยำในคลาสที่ 1 และ 5 มีค่าเท่ากับ 85.71% ซึ่งหมายความว่าแบบจำลองนี้สามารถทำนายคลาสที่มีคนอยู่ 1 และ 5 คนได้ดีที่สุด รองลงมาคือคลาสที่ 2 และ 3 ได้ค่า 68.52% และ 68.18% ตามลำดับ และในค่าความระลึกสามารถตรวจจับคลาสที่มีคนอยู่ 0 คนได้ดีที่สุดด้วยค่า 87.50% รองลงมาคือคลาสที่ 2 และ 1 คือสามารถตรวจจับคลาสที่มีคนอยู่ได้ 74.00% และ 72.41% ตามลำดับ

ตารางที่ 4.1 เปรียบเทียบผลการวัดประสิทธิภาพในมุมมองแนวระนาบและแนวเหนือศีรษะ

| มุมมอง        | ชุดข้อมูล | ค่าความแม่นยำ | ค่าความระลึก | ค่าวัดประสิทธิภาพ | ค่าความถูกต้อง |
|---------------|-----------|---------------|--------------|-------------------|----------------|
| แนวระนาบ      | A1        | 0.6837        | 0.6467       | 0.6557            | 0.6467         |
|               | B1        | 0.6841        | 0.6667       | 0.6727            | 0.6667         |
|               | ค่าเฉลี่ย | 0.6839        | 0.6566       | 0.6642            | 0.6567         |
| แนวเหนือศีรษะ | A1        | 0.8299        | 0.8067       | 0.8126            | 0.8067         |
|               | B1        | 0.6877        | 0.6500       | 0.6555            | 0.6500         |
|               | ค่าเฉลี่ย | 0.7588        | 0.7284       | 0.7341            | 0.7284         |

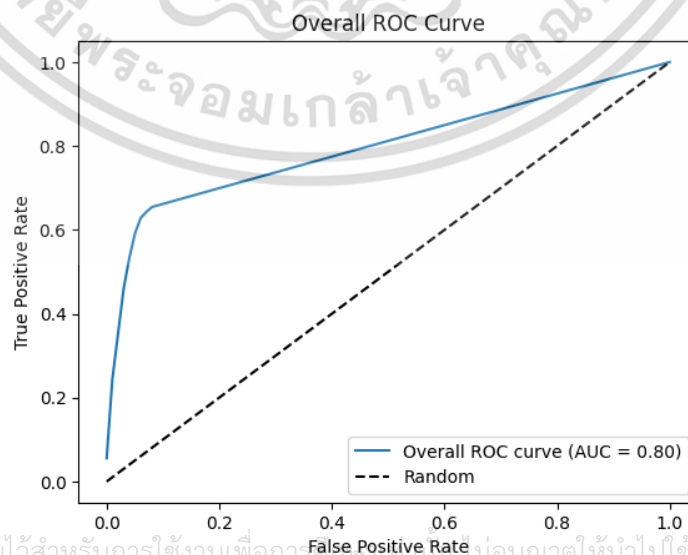
จากตารางที่ 4.1 แสดงค่าความแม่นยำ ค่าความระลึก ค่าวัดประสิทธิภาพ และค่าความถูกต้อง ในมุมมองแนวระนาบและแนวเหนือศีรษะดังนี้ ค่าความถูกต้อง ในมุมมองแนวระนาบได้ค่า 65.67% และในมุมมองแนวเหนือศีรษะได้ค่าความถูกต้อง อยู่ที่ 72.84% ซึ่งเป็นการบอกว่าแบบจำลองสามารถทำนายมุมมองแนวเหนือศีรษะได้ถูก 72.84% ซึ่งมากกว่ามุมมองในแนวระนาบ ถัดมาในมุมมองแนวระนาบได้ค่าความแม่นยำอยู่ที่ 68.39% มุมมองแนวเหนือศีรษะได้ค่าความแม่นยำ 75.88% หมายความว่า แบบจำลองสามารถทำนายว่าเป็นคลาสที่สนใจได้ 75.88% ซึ่งมากกว่าในมุมมองแนวระนาบ ส่วนค่าความระลึกในมุมมองแนวเหนือศีรษะได้ค่า 72.84% มากกว่าค่าในมุมมองแนวระนาบที่ได้ 65.67% หมายความว่า เมื่อเรากำลังสนใจคลาส ๆ หนึ่งอยู่ และแบบจำลองสามารถทำนายถูกว่าเป็นคลาสนั้น ๆ โดยในมุมมองแนวเหนือศีรษะมีประสิทธิภาพมากกว่ามุมมองแนวระนาบ ส่วนค่าวัดประสิทธิภาพเป็นค่าเฉลี่ยถ่วงน้ำหนักระหว่างค่าความแม่นยำ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และค่าความระลึก ซึ่งให้ผลเช่นเดียวกับก่อนหน้าคือ แบบจำลองมีประสิทธิภาพที่ดีกว่าเมื่อใช้ในมุมมองแนวเหนือศีรษะ



รูปที่ 4.9 กราฟ ROC แสดงค่า AUC โดยรวมในมุมมองแนวระนาบ

จากรูปที่ 4.9 แสดงกราฟ ROC Curve จะเห็นได้ว่า กราฟมีลักษณะโค้งในช่วงต้น และเข้าใกล้เส้นทแยงมุมเล็กน้อย ซึ่งเส้นทแยงมุมคือเส้นที่แสดงถึงการทำนายแบบสุ่ม ถ้าเส้นกราฟต่ำกว่าเส้นทแยงมุม หรือค่า AUC ต่ำกว่า 0.5 แบบจำลองจะไม่สามารถนำไปใช้ได้ แต่ในที่นี้ค่า AUC = 0.80 ซึ่งอยู่ในเกณฑ์ที่แบบจำลองมีประสิทธิภาพการทำนายในระดับดี นั่นคือ แบบจำลองสามารถทำนายและจำแนกบุคคลได้ดี



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น รูปที่ 4.10 กราฟ ROC แสดงค่า AUC โดยรวมในมุมมองแนวเหนือศีรษะ

จากรูปที่ 4.10 แสดงกราฟ ROC Curve จะเห็นได้ว่า กราฟมีลักษณะโค้งในช่วงต้น และเข้าใกล้เส้นทแยงมุมเล็กน้อย ในที่นี้ค่า AUC = 0.80 โดยอยู่ในเกณฑ์ที่แบบจำลองมีประสิทธิภาพในการทำนายระดับดี นั่นคือ แบบจำลองสามารถทำนายและจำแนกบุคคลได้ดี ซึ่งในมุมมองแนวเหนือศีรษะนี้ได้ค่า AUC เท่ากันกับค่าในมุมมองแนวระนาบในข้างต้น

ดังนั้น จึงสรุปได้ว่าในมุมมองแนวเหนือศีรษะ มีค่าสูงกว่าทุก ๆ ค่าในมุมมองแนวระนาบ แสดงให้เห็นว่าแบบจำลองนี้เหมาะกับการตรวจจับบุคคลในมุมมองแนวเหนือศีรษะมากกว่ามุมมองแนวระนาบ

จากค่า AUC ในกราฟ ROC ของมุมมองแนวระนาบและมุมมองแนวเหนือศีรษะที่ได้ AUC = 0.8 ทั้งสองกราฟ แสดงให้เห็นว่าแบบจำลองสามารถทำนายได้อยู่ในเกณฑ์ที่ดี คือสามารถทำนายและจำแนกบุคคลได้ในระดับดี

ตารางที่ 4.2 การใช้ทรัพยากรของเครื่องในขณะที่ทำการทดสอบแบบจำลอง YOLOv7

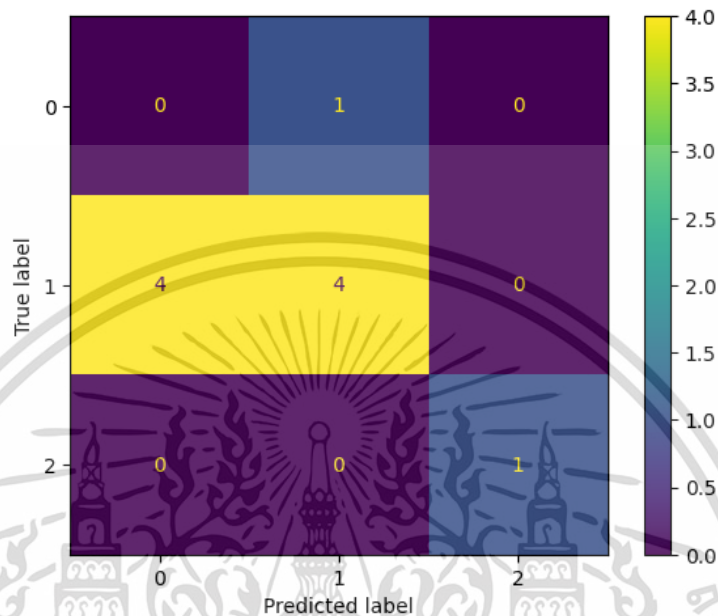
| มุมมอง        | ชุดข้อมูล | วิดีโอ   | ความยาว (นาที) | จำนวนเฟรม | RAM ของระบบ (GB) | RAM ของ GPU(GB) | DISK (GB) | ระยะเวลาในการรัน (วินาที) |
|---------------|-----------|----------|----------------|-----------|------------------|-----------------|-----------|---------------------------|
| แนวระนาบ      | A1        | วิดีโอ 1 | 5              | 150       | 4.5              | 1.4             | 24.9      | 16.62                     |
|               | B1        | วิดีโอ 1 | 5              | 60        | 4.0              | 1.4             | 24.4      | 4.45                      |
|               |           | วิดีโอ 2 | 5              | 60        | 3.8              | 1.1             | 24.4      | 2.8                       |
|               |           | วิดีโอ 3 | 5              | 60        | 4.0              | 1.4             | 24.4      | 2.54                      |
| แนวเหนือศีรษะ | A1        | วิดีโอ 1 | 5              | 150       | 4.6              | 1.4             | 25.0      | 17.58                     |
|               | B1        | วิดีโอ 1 | 5              | 60        | 4.0              | 1.4             | 24.4      | 3.98                      |
|               |           | วิดีโอ 2 | 5              | 60        | 3.7              | 1.0             | 24.4      | 3.07                      |
|               |           | วิดีโอ 3 | 5              | 60        | 4.0              | 1.4             | 24.4      | 2.6                       |

จากตารางที่ 4.2 แสดงให้เห็นว่าเมื่อจำนวนเฟรมมากขึ้น RAM ของระบบ และระยะเวลาในการรันจะเพิ่มขึ้นตามไปด้วย แต่ถ้าจำนวนเฟรมที่นำมาทดสอบมีจำนวนไม่แตกต่างกัน การใช้ทรัพยากรของเครื่องก็ไม่แตกต่างกัน

#### 4.1.2 การทดสอบประสิทธิภาพของแบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh)

โดยการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ได้แบ่งข้อมูลที่ใช้ในการทดสอบออกเป็น 2 ชุด คือ ชุดข้อมูล A2 และชุดข้อมูล B2 แต่ละชุดประกอบด้วยวิดีโอเอกสารนี้จำนวน 10 วิดีโอ ทำการทดสอบและประเมินผลประสิทธิภาพของแบบจำลองได้ดังต่อไปนี้ด้านการคำนวณว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ชุดข้อมูล A2 ประกอบด้วยวิดีโอที่ใช้ในการทดสอบ 10 วิดีโอ ซึ่งเป็นวิดีโอที่ผู้วิจัยทำการบันทึกเอง โดยทำการทดสอบการค้นใบหน้าด้วยองศา 10 15 และ 20 องศา ผลที่ได้มีค่าเท่ากันทั้ง 3 องศา รูปที่ 4.11 แสดงเมทริกซ์ความสับสนของชุดข้อมูล A2



รูปที่ 4.11 เมทริกซ์ความสับสนของชุดข้อมูล A2

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 1            | 0.8000    | 0.5000 | 0.6154   | 8       |
| 2            | 1.0000    | 1.0000 | 1.0000   | 1       |
| accuracy     |           |        | 0.5000   | 10      |
| macro avg    | 0.6000    | 0.5000 | 0.5385   | 10      |
| weighted avg | 0.7400    | 0.5000 | 0.5923   | 10      |

รูปที่ 4.12 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล A2

จากรูปที่ 4.12 เนื่องจากผลลัพธ์ในการทำนาย ได้ออกมาเหมือนกันทั้ง 3 ขนาดองศาที่ทำ การทดสอบ จึงได้ค่าการวัดประสิทธิภาพของชุดข้อมูล A2 ดังนี้

ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมด คิดจาก 10 กรณี ซึ่งแบบจำลอง สามารถทำนายถูก 5 กรณี คิดเป็น 50.00%

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 74.00%

ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 50.00%

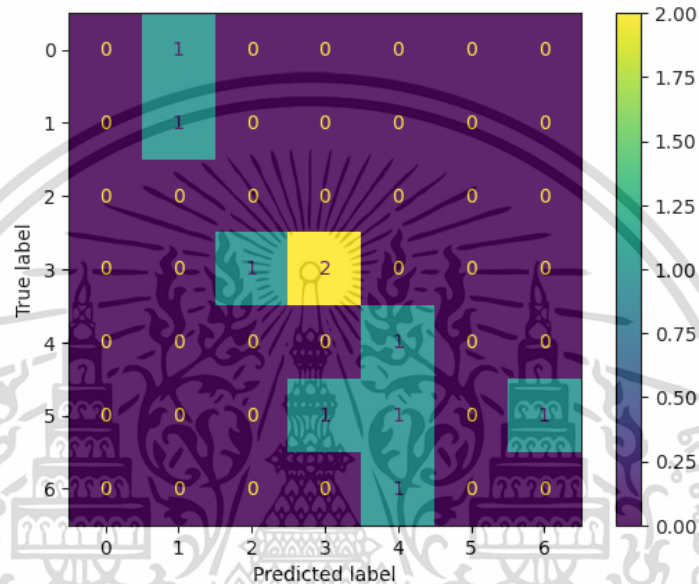
ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 59.23%

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการวิจัยเท่านั้น การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ชุดข้อมูล B2 ประกอบด้วยวิดีโอที่ใช้ในการทดสอบ 10 วิดีโอ ซึ่งเป็นวิดีโอที่ผู้วิจัยได้ทำการบันทึกเอง 8 วิดีโอ ร่วมกับข้อมูลที่ไม่ได้ทำการบันทึกเอง 2 วิดีโอ โดยทำการทดสอบการหันใบหน้า 3 ขนาดองศา คือ 10 15 และ 20 องศา ได้ผลดังนี้

- 10 องศา



รูปที่ 4.13 เมทริกซ์ความสัมพันธ์ของชุดข้อมูล B2 (10 องศา)

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 1            | 0.5000    | 1.0000 | 0.6667   | 1       |
| 2            | 0.0000    | 0.0000 | 0.0000   | 0       |
| 3            | 0.6667    | 0.6667 | 0.6667   | 3       |
| 4            | 0.3333    | 1.0000 | 0.5000   | 1       |
| 5            | 0.0000    | 0.0000 | 0.0000   | 3       |
| 6            | 0.0000    | 0.0000 | 0.0000   | 1       |
| accuracy     |           |        | 0.4000   | 10      |
| macro avg    | 0.2143    | 0.3810 | 0.2619   | 10      |
| weighted avg | 0.2833    | 0.4000 | 0.3167   | 10      |

รูปที่ 4.14 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (10 องศา)

จากรูปที่ 4.14 ได้ค่าการวัดประสิทธิภาพของชุดข้อมูล B2 ในการหันใบหน้า 10 องศา ดังนี้

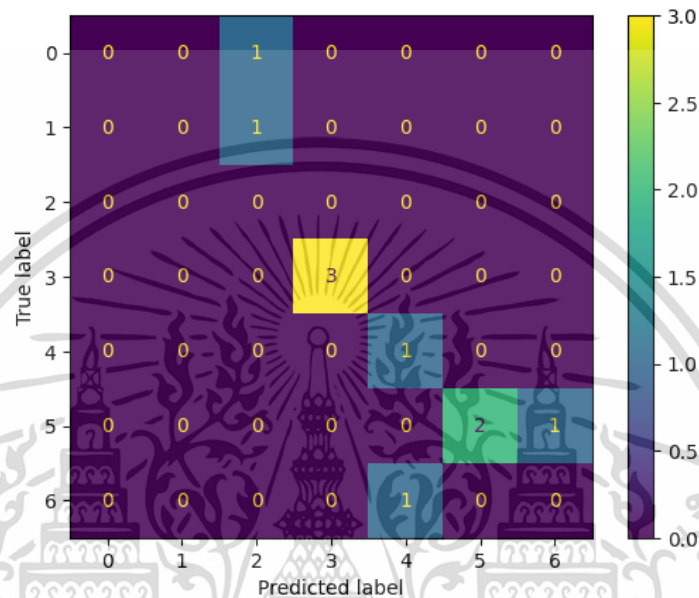
ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดคิดจาก 10 กรณี ซึ่งแบบจำลองเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า สามารถทำนายถูก 4 กรณี คิดเป็น 40.00% ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 28.33%

ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 40.00%

ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 31.67%

- 15 องศา



รูปที่ 4.15 เมทริกซ์ความสับสนของชุดข้อมูล B2 (15 องศา)

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 1            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 2            | 0.0000    | 0.0000 | 0.0000   | 0       |
| 3            | 1.0000    | 1.0000 | 1.0000   | 3       |
| 4            | 0.5000    | 1.0000 | 0.6667   | 1       |
| 5            | 1.0000    | 0.6667 | 0.8000   | 3       |
| 6            | 0.0000    | 0.0000 | 0.0000   | 1       |
| accuracy     |           |        | 0.6000   | 10      |
| macro avg    | 0.3571    | 0.3810 | 0.3524   | 10      |
| weighted avg | 0.6500    | 0.6000 | 0.6067   | 10      |

รูปที่ 4.16 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (15 องศา)

จากรูปที่ 4.16 ได้ค่าการวัดประสิทธิภาพของชุดข้อมูล B2 ในการหันใบหน้า 15 องศา ดังนี้

ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดคิดจาก 10 กรณี ซึ่งแบบจำลอง

สามารถทำนายถูก 6 กรณี คิดเป็น 60.00%

เอกสารนี้เป็นเอกสารทบทวนเว็บไซต์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

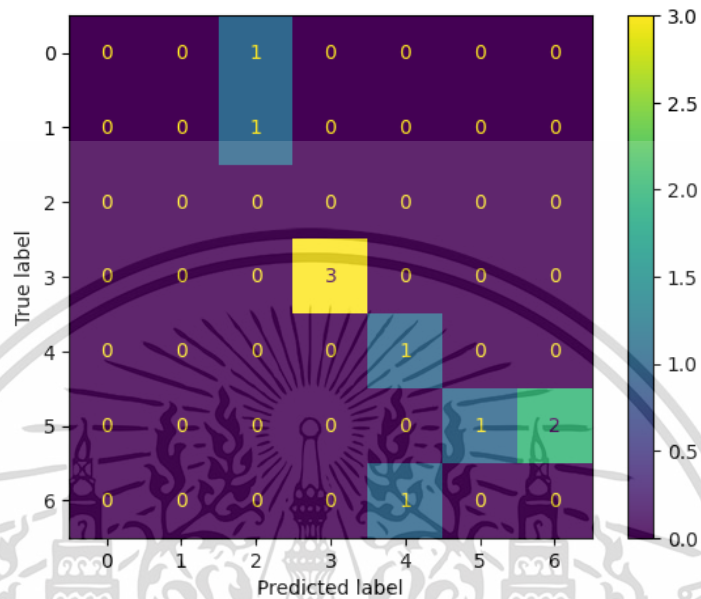
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่มีเหตุใดแบบสงวนเนื้อหา และต้องยังต้องแจ้งเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 65.00%

ค่าเฉลี่ยค่าความระลึก (Average of Recall) เท่ากับ 60.00%

ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 60.67%

• 20 องศา



รูปที่ 4.17 เมทริกซ์ความสับสนของชุดข้อมูล B2 (20 องศา)

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 1            | 0.0000    | 0.0000 | 0.0000   | 1       |
| 2            | 0.0000    | 0.0000 | 0.0000   | 0       |
| 3            | 1.0000    | 1.0000 | 1.0000   | 3       |
| 4            | 0.5000    | 1.0000 | 0.6667   | 1       |
| 5            | 1.0000    | 0.3333 | 0.5000   | 3       |
| 6            | 0.0000    | 0.0000 | 0.0000   | 1       |
| accuracy     |           |        | 0.5000   | 10      |
| macro avg    | 0.3571    | 0.3333 | 0.3095   | 10      |
| weighted avg | 0.6500    | 0.5000 | 0.5167   | 10      |

รูปที่ 4.18 ผลการคำนวณการวัดประสิทธิภาพของชุดข้อมูล B2 (20 องศา)

จากรูปที่ 4.18 ได้ค่าการวัดประสิทธิภาพของชุดข้อมูล B2 ในการหันใบหน้า 20 องศา ดังนี้  
ค่าความถูกต้อง (Accuracy) ของคลาสทั้งหมดคิดจาก 10 กรณี ซึ่งแบบจำลองสามารถทำนายถูก 5 กรณี คิดเป็น 50.00%

ค่าเฉลี่ยค่าความแม่นยำ (Average of Precision) เท่ากับ 65.00%

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ (Average of Recall) เท่ากับ 50.00% ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยค่าวัดประสิทธิภาพ (Average of F1-score) เท่ากับ 51.67%

ตารางที่ 4.3 ค่าการประเมินประสิทธิภาพในการหั่นของใบหน้าด้วยมุม 10 15 และ 20 องศา

| ชุดข้อมูล | องศาที่ใช้ทดสอบ | ค่าความแม่นยำ | ค่าความระลึก | ค่าวัดประสิทธิภาพ | ค่าความถูกต้อง |
|-----------|-----------------|---------------|--------------|-------------------|----------------|
| A2        | 10              | 0.7400        | 0.5000       | 0.5923            | 0.5000         |
|           | 15              |               |              |                   |                |
|           | 20              |               |              |                   |                |
| B2        | 10              | 0.2833        | 0.4000       | 0.3167            | 0.4000         |
|           | 15              | 0.6500        | 0.6000       | 0.6067            | 0.6000         |
|           | 20              | 0.6500        | 0.5000       | 0.5167            | 0.5000         |

จากตารางที่ 4.3 แสดงให้เห็นว่าข้อมูลในชุด A2 ได้ค่าเท่าเดิมในทั้ง 3 องศาที่ใช้ทดสอบ คือ ไม่มีการเปลี่ยนแปลงในผลการทำนายเมื่อมุมในการหั่นใบหน้าถูกเปลี่ยน แต่ข้อมูลชุด B2 ในองศาการหั่นใบหน้าที่ 15 องศา ได้ค่าความแม่นยำ ค่าความระลึก ค่าวัดประสิทธิภาพ และค่าความถูกต้อง ที่มากกว่าองศาในการหั่นใบหน้าที่ 10 องศา และ 20 องศา แสดงให้เห็นว่ามุมที่เหมาะสมในการหั่นใบหน้าเพื่อประเมินความสนใจ คือ 15 องศา

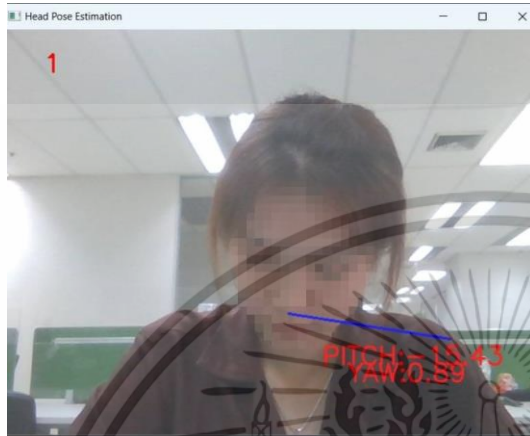
ตารางที่ 4.4 ค่าการประเมินประสิทธิภาพในการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

| ชุดข้อมูล | องศาที่ใช้ทดสอบ | ค่าความแม่นยำ | ค่าความระลึก | ค่าวัดประสิทธิภาพ | ค่าความถูกต้อง |
|-----------|-----------------|---------------|--------------|-------------------|----------------|
| A2        | 15              | 0.7400        | 0.5000       | 0.5923            | 0.5000         |
| B2        | 15              | 0.6500        | 0.6000       | 0.6067            | 0.6000         |
| ค่าเฉลี่ย |                 | 0.6950        | 0.5500       | 0.5995            | 0.5500         |

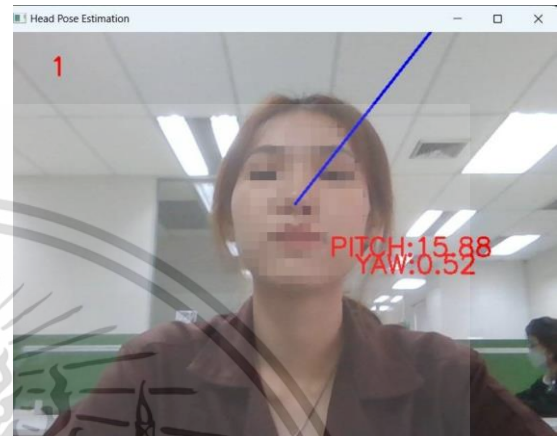
จากตารางที่ 4.4 แสดงค่าความแม่นยำ ค่าความระลึก ค่าวัดประสิทธิภาพ และค่าความถูกต้อง ของข้อมูล ดังนี้ ค่าความถูกต้อง ของคลาสทั้งหมดคิดจาก 20 กรณี ซึ่งแบบจำลองสามารถทำนายถูกโดยคิดเป็น 55.00% ซึ่งแปลผลได้ว่าแบบจำลองสามารถทำนายได้ถูกต้อง 55.00% ถัดมา ค่าความแม่นยำอยู่ที่ 69.50% หมายความว่า แบบจำลองสามารถทำนายว่าเป็นคลาสที่สนใจได้ 69.50% ส่วนค่าความระลึกได้ค่า 55.00% หมายความว่า เมื่อเรากำลังสนใจคลาส ใดๆ หนึ่งอยู่ และไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองสามารถทำนายถูกกว่าเป็นคลาสนั้น ๆ 55.00% และสุดท้ายค่าวัดประสิทธิภาพเป็นค่าเฉลี่ยถ่วงน้ำหนักระหว่างค่าความแม่นยำและค่าความระลึกลับ ซึ่งให้ผลว่า แบบจำลองมีประสิทธิภาพในการตรวจจับอยู่ที่ 59.95%

ประเมินความสนใจ คือ 15 องศา



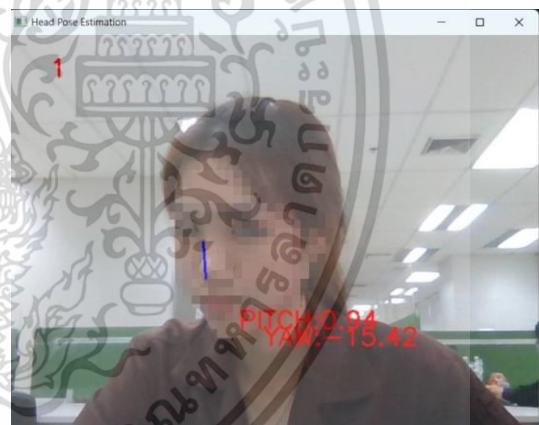
(ก) ก้มหน้า 15 องศา



(ข) เงยหน้า 15 องศา



(ค) หันขวา 15 องศา



(ง) หันซ้าย 15 องศา

รูปที่ 4.19 ตัวอย่างการหันใบหน้าด้วยมุม 15 องศา

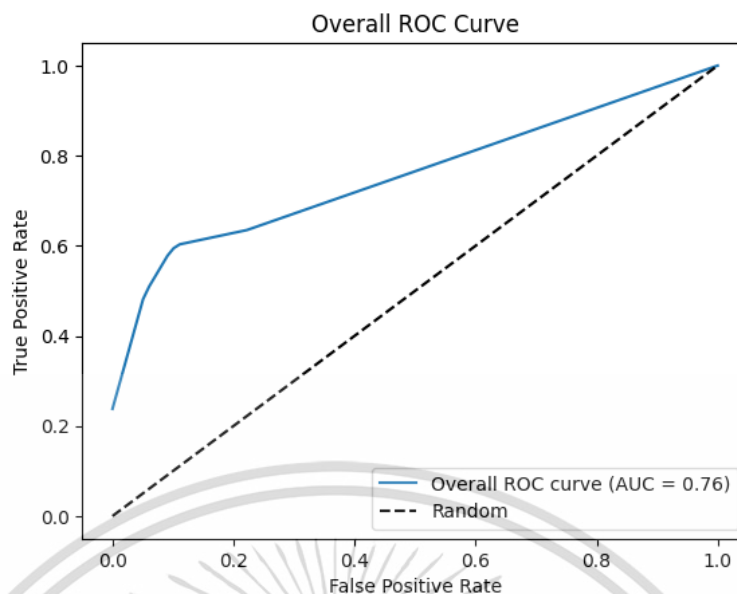
จากรูปที่ 4.19 โดย Pitch คือ การหันใบหน้าในแกน X คือการก้มและเงย ส่วน Yaw คือการหันใบหน้าในแกน Y หันซ้าย และขวาโดย (ก) มุมก้มจะมีค่าติดลบ

(ข) มุมเงยจะมีค่าเป็นบวก

(ค) หันขวาจะมีค่าเป็นบวก

(ง) หันซ้ายจะมีค่าติดลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 กราฟ ROC แสดงค่า AUC โดยรวมของแบบจำลอง

แสดงกราฟ ROC Curve จะเห็นได้ว่า กราฟมีลักษณะโค้งในช่วงต้น และเข้าใกล้เส้นทแยงมุมตั้งแต่ช่วงกลางถึงปลาย ในที่นี้ค่า AUC = 0.76 โดยอยู่ในเกณฑ์ที่แบบจำลองมีประสิทธิภาพในการทำนายระดับดี นั่นคือ แบบจำลองสามารถทำนายและจำแนกบุคคลได้ดี

## 4.2 อภิปรายผล

จากงานวิจัยในอดีตที่ผ่านมา งานวิจัยที่เกี่ยวข้องได้แก่ งานวิจัยของ Al-Nuimi and Mohammed et al. (2021) ได้ทำการวิจัยเกี่ยวกับการประมาณทิศทางใบหน้าตามจุดสังเกตของ Mediapipe โดยใช้ฟังก์ชันตรีโกณมิติเพื่อประเมินทิศทางของใบหน้าโดยใช้จุดสังเกตที่ตรวจพบ จุดสังเกตเหล่านี้ถูกตรวจพบด้วยวิธีแบบจำลอง 3 มิติของ Mideapipe ในบรรดาจุดสังเกตที่ตรวจพบ จุดหลายจุดจะถูกเลือกเป็นจุดสังเกต เพื่อนำไปใช้ในการตรวจจับในการโกงข้อสอบ โดย วิธี Mediapipe เป็นเครื่องมือตรวจหาใบหน้าที่มีความแม่นยำสูง ซึ่งการคาดคะเนจุดสังเกตมีความแม่นยำประมาณ 100% การคาดคะเนที่แม่นยำนี้ให้จุดสังเกตที่แม่นยำซึ่งสามารถนำไปใช้ในการประมาณท่าศีรษะได้ ซึ่งงานวิจัยดังกล่าวใช้วิธีตรวจจับจุดสังเกตบนใบหน้าวิธีเดียวกัน แต่ใช้การประมาณทิศทางใบหน้าด้วยวิธีที่ต่างกันคือในงานวิจัยของ Al-Nuimi and Mohammed ใช้ฟังก์ชันตรีโกณมิติแต่งานวิจัยฉบับนี้ใช้อองศาที่กำหนดขึ้นคือ 10 15 และ 20 องศา โดยทำการทดสอบและเลือกองศาในการหันใบหน้าที่เหมาะสมที่สุดคือ 15 องศา และจากงานวิจัยของ Radmehr, A. et al. (2021) ได้กล่าวไว้ว่าไลบรารี มีเดียไปป์ (Library MediaPipe) สามารถค้นหาท่าทางของศีรษะได้อย่างแม่นยำ ผลการวิเคราะห์งานวิจัยนี้ ในหัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่โดยไม่ได้รับอนุญาต  
 ที่ได้ค่าความถูกต้อง เพียงแค่ 0.55 หรือ 55% อาจจะยังไม่สอดคล้องกับงานวิจัยที่อ้างอิงข้างต้น  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



## 2) มุมมองแนวเหนือศีรษะ

ชุดข้อมูล A1 : ประกอบด้วยเฟรมที่ใช้ทดสอบ จำนวน 150 เฟรม จากค่าความแม่นยำ (Precision) สามารถทำนายคลาสที่มีคนอยู่ 0 และ 6 คนได้ดีที่สุด คือ 100% แต่ในคลาสที่ 6 อาจจะบอกไม่ได้ว่าดีที่สุดใน เนื่องจากในคลาสที่ 6 มีเพียง 1 เฟรมเท่านั้นที่มีคนอยู่ 6 คน

ชุดข้อมูล B1 : ประกอบด้วยเฟรมจำนวน 180 เฟรมที่ใช้ในการทดสอบ จากค่าความแม่นยำ (Precision) ในคลาสที่ 1 และ 5 มีค่าเท่ากับ 85.71% ซึ่งหมายความว่าแบบจำลองนี้สามารถทำนายคลาสที่มีคนอยู่ 1 และ 5 คนได้ดีที่สุด

และจากกราฟ ROC เมื่อได้ทำการรวมชุดข้อมูล A และ B เข้าด้วยกันพบว่าได้ค่า AUC = 0.80 เช่นเดียวกับมุมมองในแนวระนาบ ซึ่งอยู่ในเกณฑ์ที่แบบจำลองสามารถทำนายและจำแนกบุคคลได้ในระดับดี

โดยจะพิจารณาค่าความถูกต้อง (Accuracy) โดยรวมเป็นหลัก แต่เนื่องจากค่าอื่น เช่น ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าวัดประสิทธิภาพ (F1-score) ให้ผลลัพธ์ในทางเดียวกันคือ ค่าในมุมมองแนวเหนือศีรษะมีค่าที่สูงกว่าค่าในมุมมองแนวระนาบทุกกรณี จึงสรุปและตอบวัตถุประสงค์ข้อ 1 และ 2 ได้ว่ามุมมองแนวเหนือศีรษะสามารถตรวจจับและนับคนโดยมีประสิทธิภาพที่สูงกว่าในมุมมองแนวระนาบ ซึ่งแสดงว่าแบบจำลองสามารถตรวจจับชุดข้อมูลในมุมมองแนวเหนือศีรษะได้ดีกว่ามุมมองแนวระนาบ และจากค่า AUC ในกราฟ ROC ของมุมมองแนวระนาบและมุมมองแนวเหนือศีรษะได้ค่า AUC ที่เท่ากัน แสดงให้เห็นว่าแบบจำลองสามารถทำนายและจำแนกบุคคลได้ในระดับดี

### 5.1.2 การตรวจจับและวัดองศาห้วงใบหน้ากับกล้องเพื่อประเมินความสนใจ

จากวัตถุประสงค์ในข้อ 3) เพื่อศึกษาและวิเคราะห์ลักษณะการหันใบหน้าของแต่ละบุคคลโดยวัดองศาจากการหัน เพื่อประเมินและนับจำนวนคนที่มีความสนใจต่อกล้อง ด้วยแบบจำลองมีเดียไปป์ เฟซ เมช (MediaPipe Face Mesh) ข้อมูลที่ใช้ทดสอบแบ่งเป็น 2 ชุดข้อมูล คือ ชุดข้อมูล A2 และชุดข้อมูล B2 ได้ทำการทดสอบชุดข้อมูลละ 10 วิดีโอ โดยได้ทำการทดสอบกับองศาในการหันใบหน้าที 10 องศา 15 องศา และ 20 องศา พบว่าองศาที่มีความเหมาะสมกับข้อมูลชุดนี้คือ 15 องศา เนื่องจากให้ค่าการวัดประสิทธิภาพที่สูงที่สุด จึงใช้เกณฑ์การหันใบหน้าที 15 องศา และเมื่อนำข้อมูลทั้ง 2 ชุดมาวัดประสิทธิภาพพบว่า ได้ค่าความถูกต้อง (Accuracy) 55% ซึ่งหมายความว่าแบบจำลองสามารถทำนายได้ถูกต้อง 55.00% จากกราฟ ROC ได้ค่า AUC = 0.76 ซึ่งอยู่ในเกณฑ์ที่แบบจำลองมีประสิทธิภาพในการทำนายระดับดี นั่นคือ แบบจำลองสามารถทำนายและจำแนกบุคคลได้ดี เนื่องจากงานวิจัยนี้เป็นงานวิจัยต้นแบบ เพื่อทดสอบว่าการวัดองศาในการหันใบหน้าสามารถตรวจจับความสนใจต่อป้ายโฆษณาดิจิทัลได้หรือไม่ ซึ่งผลที่ได้อาจจะยังไม่ได้เป็นค่าที่ได้มาตรฐานพอที่จะนำไปใช้ได้จริง งานวิจัยนี้จึงเป็นแนวทางในการพิจารณาอีกวิธีหนึ่งเพื่อนำไปศึกษาและพัฒนาต่อไป

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี หากมีผู้ใดนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาต  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการวิเคราะห์โดยพิจารณาค่าความถูกต้อง (Accuracy) วิธีการตรวจจับและนับจำนวนบุคคลโดยใช้แบบจำลองโยโล เวอร์ชัน 7 ผลลัพธ์ของแต่ละมุมมองคือแบบจำลองสามารถทำนายมุมมองแนวระนาบได้ถูก 65.67% และมุมมองแนวเหนือศีรษะได้ถูก 72.84% และวิธีการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจโดยใช้มีเดียไปป์ แบบจำลองสามารถทำนายได้ถูก 55.00% ดังนั้นแบบจำลองในงานวิจัยครั้งนี้อาจจะยังไม่ได้เป็นค่าที่ได้มาตรฐานมากพอที่จะนำไปใช้ได้จริง แต่สามารถนำไปเป็นต้นแบบเพื่อวิเคราะห์การตรวจจับและนับจำนวนคนที่เดินผ่านได้ ในกรณีต้องการทราบจำนวนคนที่เดินผ่าน และใช้วิธีการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ในกรณีต้องการทราบคนที่สนใจ

## 5.2 ข้อจำกัด

1. ในแบบจำลองโยโล เวอร์ชัน 7 เป็นการนำแบบจำลองมาทดสอบกับชุดข้อมูลที่บันทึกเองโดยไม่ได้มีการปรับแก้โค้ด ซึ่งใช้เพื่อเปรียบเทียบผลลัพธ์ในมุมมองแนวระนาบและมุมมองแนวเหนือศีรษะ
2. ในขนาดของป้ายที่ได้กล่าวไว้ในงานวิจัยคือ 7-10 นิ้ว ซึ่งขนาดป้ายอาจมีการปรับเปลี่ยนได้ตามความต้องการหรือตามการใช้งาน
3. ในหัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจแบบจำลองของเรามีข้อจำกัดคือ สามารถตรวจจับจุดสังเกตบนใบหน้าได้ในระยะประมาณ 2 เมตร สำหรับคนที่สนใจแต่อยู่ไกลเกิน 2 เมตร จะไม่สามารถตรวจจับได้ ซึ่งสามารถนำแบบจำลองนี้ไปพัฒนาต่อเพื่อการตรวจจับให้ได้ระยะที่ไกลมากขึ้นกว่าเดิม
4. ในหัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ถ้าตรวจจับใบหน้าไม่พบก็จะไม่สามารถเข้าอัลกอริทึม (Algorithm) ในการตรวจสอบความสนใจได้
5. การกำหนดองศาการหันใบหน้าที่ 10 15 20 องศาที่เหมาะสมคือ 15 องศา หน้านี้เป็นเพียงต้นแบบในการทดสอบว่าแบบจำลองสามารถตรวจจับคนที่สนใจได้หรือไม่ได้ ซึ่งในกรณีนำไปใช้งานจริงอาจต้องขึ้นอยู่กับขนาดจอโฆษณาที่นำไปติดตั้ง
6. การกำหนดเกณฑ์การนับว่าบุคคลนั้นสนใจคือ เมื่อมองตรงเกิน 5 วินาที ซึ่งในการนำไปใช้งานจริงทางผู้พัฒนาอาจจะไม่ได้สนใจในระยะเวลา 5 วินาที อาจสนใจเพียงแค่หันมามอง 1 วินาทีในระยะเวลาที่ใช้จึงเพื่อทดสอบในงานวิจัยนี้เท่านั้น
7. ป้ายโฆษณาดังกล่าวเป็นป้ายโฆษณาแบบไมโคร คอมพิวเตอร์ (Micro Computer) เป็นป้ายที่มีองค์ประกอบหลักอยู่ที่คอมพิวเตอร์ เป็นศูนย์กลางในการควบคุมและจัดการป้ายโฆษณา อาจมีข้อจำกัดด้านเทคโนโลยีหรือด้านการใช้ทรัพยากรของเครื่อง ทำให้ไม่สามารถใช้ไลบรารีหรือแบบจำลองบางตัวได้
8. ทั้ง 2 หัวข้อ วิดีโอที่ใช้ในการทดสอบเป็นวิดีโอที่มีการตั้งค่า (Setting) ขึ้นมาเอง

เอกสารนี้เป็นเอกสารที่จะสร้างแบบจำลองบางตัวอาจจะต้องใช้เวลาในการฝึกสอนหรือทดสอบที่นานกว่า  
ไม่ว่าการอุปกรณ์ที่ใช้ในการวิเคราะห์ต้องมีความพร้อมและมีประสิทธิภาพที่สูงกว่านี้การทุกครั้งที่มีการนำไปใช้

### 5.3 ข้อเสนอแนะ

1. หัวข้อการตรวจจับและนับจำนวนบุคคล การใช้แบบจำลองโพลีโวลเวอร์ชัน 7 เกิดข้อผิดพลาดในการตรวจจับคือ ในบางเฟรมที่คนใส่เสื้อสีเดียวกับพื้นหลัง ทำให้การตรวจจับและนับคนผิดพลาดไป จึงควรทำการฝึกตัวแบบด้วยตนเองเพื่อให้มีความถูกต้องที่แม่นยำมากขึ้น

2. หัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจสามารถนำไปปรับใช้ในการตรวจจับเพศ อายุ ของผู้คนที่เดินผ่าน เพื่อเพิ่มประสิทธิภาพการใช้งานของแบบจำลองให้ตรงเป้าหมายของผู้ที่ต้องการศึกษามากยิ่งขึ้นและเพื่อให้สามารถจัดทำป้ายโฆษณาดิจิทัลที่เหมาะสมหรือเจาะจงกับกลุ่มผู้คนที่เดินผ่านได้มากที่สุด

3. หัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ในกรณีนำไปใช้จริง อาจต้องมีการปรับให้เหมาะสมกับการใช้งาน เช่น กรณีในการกำหนดจำนวนวินาทีที่สนใจ และกรณีในการกำหนดองศาในการหันของใบหน้า เป็นต้น

4. หัวข้อการตรวจจับและวัดองศาระหว่างใบหน้ากับกล้องเพื่อประเมินความสนใจ ได้กำหนดองศาในการหันใบหน้าคือ 10 15 และ 20 องศา และได้เลือกมุมที่เหมาะสมคือ 15 องศา แต่อาจจะไม่ครอบคลุมคนที่อยู่ชิดขอบของเฟรมกล้อง ซึ่งอาจจะไม่สามารถใช้มุมการหันใบหน้า 15 องศาได้ งานวิจัยนี้จึงเป็นแนวทางในการนำไปพัฒนาต่อให้มีการตรวจจับให้มีประสิทธิภาพมากขึ้นกว่าเดิม

5. ในปัจจุบันบริษัทกรณีศึกษามีป้ายโฆษณาดิจิทัลอยู่แล้ว แต่ใช้ไลบรารีตัวเดิม ซึ่งในงานวิจัยนี้อาจเป็นทางเลือกในการนำแบบจำลองตัวใหม่ เช่น แบบจำลองโพลีโวลเวอร์ชัน 7 ไปปรับใช้งานได้

6. ทั้ง 2 หัวข้อ ควรทำการเปรียบเทียบกับแบบจำลองตัวอื่นด้วย ไม่ว่าจะเป็นแบบจำลอง Faster R-CNN, SSD, Haar Cascade หรือตัวแบบอื่น ๆ ที่ได้รับความนิยม เป็นต้น เพื่อที่จะเลือกตัวแบบจำลองที่ให้ความแม่นยำที่มากที่สุด

7. เนื่องจากข้อมูลที่ใช้เป็นข้อมูลทดสอบมี 2 ชุด ซึ่งอาจจะหาชุดข้อมูลทดสอบมาเพิ่ม เพื่อให้มีชุดข้อมูลที่หลากหลายมากขึ้น จะช่วยเพิ่มให้แบบจำลองมีการเรียนรู้จากชุดข้อมูลที่หลากหลาย

8. ในงานวิจัยนี้เป็นเพียงต้นแบบเพื่อเป็นขั้นตอนถัดไปในการพัฒนาหรือการนำไปประยุกต์ใช้ในการตรวจจับคนและตรวจจับความสนใจของคน โดยสามารถปรับระยะในการตรวจจับหรือปรับเปลี่ยนอัลกอริทึม (Algorithm) เป็นตัวอื่นที่สนใจ เพื่อให้แบบจำลองสามารถทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

9. เนื่องจากป้ายโฆษณาดิจิทัลดังกล่าวเป็นจอขนาดเล็ก ถ้าต้องการนำไปปรับใช้ในหน้าจอขนาดใหญ่ ควรฝึกสอนข้อมูลที่อยู่มุมไกลมากขึ้นกว่าเดิม เพื่อให้สามารถตรวจจับคนที่อยู่ในระยะไกลกว่าเดิมได้

10. ควรมีการพิจารณาการใช้ทรัพยากรของเครื่องด้วย เนื่องจากการนำไปใช้งานจริงอาจไม่รองรับแบบจำลองหรือไลบรารีนั้น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง

- กรินทร์ โกสิยานุรักษ์. 2563. การประมาณระดับการสั่นสะเทือนจากการเสียมคูลของโบลเวอร์ โดยใช้โครงข่ายประสาทเทียมร่วมกับระบบควบคุมแบบเซอร์โวและตัวสังเกต. วิศวกรรมศาสตรมหาบัณฑิตสาขาวิชาวิศวกรรมเมคคาทรอนิกส์. มหาวิทยาลัยเทคโนโลยีสุรนารี
- จันธิรัก น้ำใส, ศรศักดิ์ ทาวงษ์, ภาณุ ดงทอง, ธนะพัฒน์ เชี่ยวชาญวัฒนา และศรุต อัครเรืองสุข. 2565. ระบบตรวจจับด้วยวิธีการ YOLO ในกรณีตัวอย่างงูเข้าที่พักอาศัย. วารสารวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏอุดรธานี. 1: 59-72.
- จารุณี เจริญรส, กฤษณา เลขยันต์, แนน บุญศิริ และ อลิษา สีสด. 2565. การศึกษาทิศทางของแสงที่มีความเหมาะสมกับการถ่ายภาพเพื่อการโฆษณาขนมไทย. วารสารนวัตกรรมสื่อและการสื่อสาร. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี. 1: 25-47
- ธนพงษ์ ขจรตันติชัย. 2566. Facial Emotion Detection for Thai Elderly People using YOLOv7. 2023 15th International Conference on Knowledge and Smart Technology (KST), Phuket, Thailand, 2023. pp. 1-4.
- นพรุจ พัฒนสาร และณัฐภูมิ ศรีวิบูลย์. 2563. การประมวลผลภาพสำหรับการจำแนกคุณภาพมะม่วงพันธุ์โชคอนันต์. Journal of Information Science and Technology. 10: 24-29.
- ปัทมา ฤกษ์สิกุล. 2563. Machine Learning – คำศัพท์ที่ควรรู้. [online]. <https://twinsynergy.co.th/machine-learning-คำศัพท์ที่ควรรู้/>.
- พงษ์เดช สารการ และภัทรนันท์ หมั่นพลศรี. 2564. จุดตัดที่เหมาะสมสำหรับการวิเคราะห์เส้นโค้ง Receiver Operating Characteristic(ROC) ในการพัฒนาเครื่องมือนวัตกรรมทางสุขภาพ: กรณีตัวอย่างโดยใช้โปรแกรม STATA. วารสารไทยโภชนาการ. 16: 93-108
- ไพชยนต์ คงไชย. 2557. การพัฒนาขั้นตอนวิธีเพื่อจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์แบบคลุมเครือที่กะทัดรัด. วิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์. มหาวิทยาลัยเทคโนโลยีสุรนารี
- ภากร กัทชลี. 2565. Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย ใน Machine learning. [online]. <https://shorturl.asia/4dXLQ>.
- ศรัณญา นาทองห่อ. 2563. การสรุปใจความสำคัญของข้อความแบบสัทสำหรับข่าวท่องเที่ยวภาษาไทย. วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์. จุฬาลงกรณ์มหาวิทยาลัย
- ศุภกร รักใหม่. 2562. ไพธอนโครงสร้างโมเลกุลและกราฟที่เบียร์. Thai Journal of Physics. 3: 79-86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 อรุณ ประวัตินิสุทธิ. 2564. Python สำหรับงาน Data Science Data Visualization และ  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้นแต่มีเหตุที่แสดงโดยชัดแจ้งและต้องอย่างองเองใจของเอกสารทุกครั้งที่มีการนำไปใช้

**Machine Learning.** พิมพ์ครั้งที่ 1. กรุงเทพฯ : โปรวิชั่น

- Al-Nuimi, A. M. and Mohammed, G. J. 2021. **Face Direction Estimation based on Mediapipe Landmarks.** 2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM), Mosul, Iraq, 2021, pp. 185-190
- Atila, U and Sabaz, F. 2022. **Turkish lip-reading using Bi-LSTM and deep learning models.** Engineering Science and Technology, an International Journal. 1-10.
- Balon, T, KnapikBoguslaw, M and Cyganek, C. 2022. **New Thermal Automotive Dataset for Object Detection.** Conference on ComputerScience and Intelligence Systems. pp.43-48
- Bleier, A and Eisenbeiss, M. (2015). **The importance of trust for personalized online advertising.** Journal of Retailing,91, 390-490.
- Bhuiyan, M. R, Abdullah, J, Hashim, N. et al. 2022. **Video analytics using deep learning for crowd analysis: a review.** Multimed Tools Appl 81, 27895–27922 (2022).
- Bora, J, Dehingia, S, Boruah, A, Chetia, A, A. and Gogoi, D. 2023. **Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning.** Elsevier : 1384-1393
- Cahyadi, E, Bandung, Y. and Wiguna, A. 2012. "Application of digital signage for tourism information and communication system," 2012 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA), Denpasar-Bali, Indonesia, 2012, pp. 72-75
- Chelliah, I. 2022. **Confusion Matrix for Multiclass Classification.** [online]. <https://medium.com/mlearning-ai/confusion-matrix-for-multiclass-classification-f25ed7173e66>
- Chen, J, Baia, S, Wan, G. and Li, Y. 2023. **Research on YOLOv7-based defect detection method for automotive running lights.** SYSTEMS SCIENCE & CONTROL ENGINEERING: AN OPEN ACCESS JOURNAL. 1: 1-11.
- Gerka ,A. 2018. "Detecting the Number of Persons in the Bed Area to Enhance the Safety of Artificially Ventilated Persons," 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), Ostrava, Czech Republic, 2018, pp. 1-6.
- Glen, S. 2019. **ROC Curve Explained in One Picture.** [online]. <https://www.datasciencecentral.com/roc-curve-explained-in-one-picture/>

- Gomez, L.F. et al. 2021. "Improving Parkinson Detection using Dynamic Features from Evoked Expressions in Video," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 1562-1570.
- Grandini, M, Bagli, E. and Visani, G. 2020. **Metrics for Multi-Class Classification: an Overview.** arXiv preprint arXiv:2008.05756.
- Gulati, A, P. 2022. **Facial Landmarks Detection Using Mediapipe Library.** [online]. <https://www.analyticsvidhya.com/blog/2022/03/facial-landmarks-detection-using-mediapipe-library/>
- Huang, H and Lin, S. 2020. **WiDet:Wi-Fibaseddevice-freepassive person detection with deep convolutional neural networks.** The International Journal for the Computer and Telecommunications Industry : 357-366.
- Huang, Z, Wang, J, Fu, X, Yu, T, Guo, Y. and Wang, R. 2020. **DC-SPP-YOLO: Dense Connection and spatial pyramid pooling based YOLO for object detection.** College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
- Khalid, I. A. 2021. **Head Pose Estimation using Python.** [online]. <https://towardsdatascience.com/head-pose-estimation-using-python-d165d3541600>
- Kim, Y., Lee, J., and Han, H. 2013. **A Study on the Method of Digital Signage in the Fashion Store VMD - Focusing on Global SPA Brands In Korea.** Korean Institute of Interior Design Journal. 6: 220-232
- Kukil. 2022. **Introduction to MediaPipe.** [online]. <https://learnopencv.com/introduction-to-MediaPipe/>
- Lugaresi, C, Tang, J, Nash, H, McClanahan, C, Uboweja, E, Hays, M, Zhang, F, Chang, C, Yong, M, Lee, J, Chang, W, Hua, W, Georg, M and Grundmann, M. 2019. **MediaPipe: A Framework for Building Perception.** arXiv preprint arXiv:1906.08172
- Muhi, A, Farhat, M. and Frikha, M. 2022. "**Transfer Learning for Robust Masked Face Recognition,**" 2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sfax, Tunisia, pp. 1-5.
- Python Software Foundation. 2023. **Math-Mathematical functions.** [online]. <https://docs.python.org/3/library/math.html#math.dist>.

- Imitation of the Human Head-and-Eye Pose Using the 3-DOF Agile Eye Parallel Robot with ROS and MediaPipe Framework," 2021 9th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, Islamic Republic of, 2021, pp. 472-478.
- Smith, L.A. and Foxcroft, D.R. 2009. **The effect of alcohol advertising, marketing and portrayal on drinking behaviour in young people: systematic review of prospective cohort studies.** BMC Public Health 9, 51 (2009).
- Soma, S and Waddenkery, N. 2022. "Machine-Learning Object Detection and Recognition For Surveillance System using YoloV3," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-5
- Song, H, Lu, X, Liu, X, Zhu, X and Wang, H. 2021. "Head Pose Estimation of Stroke Patients Based on Depth Residual Network," 2021 International Conference on Networking Systems of AI (INSAI), Shanghai, China : 244-248
- Thierfelder, S. 2022. **Head Pose Estimation with MediaPipe and OpenCV in Javascript.** [online]. <https://medium.com/@susanne.thierfelder/head-pose-estimation-with-mediapipe-and-opencv-in-javascript-c87980df3acb>
- Vaidya, S, Kavthekar, S. and Joshi, A. 2023. "Leveraging YOLOv7 for Plant Disease Detection," 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) , Kottayam, India, 2023, pp. 1-6.
- Wang, C. Y, Bochkovskiy, A and Mark Liao, H. Y. 2022. **YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.** [online]. <https://arxiv.org/pdf/2207.02696.pdf>
- Zhang, Y, Ding, M, Bai, Y. and Ghanem, B. 2019. **Detecting small faces in the wild based on generative adversarial network and contextual information.** Pattern Recognition: 74-86.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ภาคผนวก ก.

ตารางที่ ก.1 คำสั่งที่ใช้ในการแตกวิดีโอเป็นเฟรมภาพ

```

import cv2
import os
listing = os.listdir(r'/ path สำหรับเก็บวิดีโอ')
count = 1
for vid in listing:
    vid = r"/ path สำหรับเก็บวิดีโอ /"+vid
    vidcap = cv2.VideoCapture(vid)
    def getFrame(sec):
        vidcap.set(cv2.CAP_PROP_POS_MSEC,sec*1000)
        hasFrames,image = vidcap.read()
        if hasFrames:
            cv2.imwrite("/path สำหรับเก็บเฟรมภาพทั้งหมด/" str(count)+".jpg",image) # Save
frame as JPG file
        return hasFrames
    sec = 0
    frameRate = 1 # Change this number to 1 for each 1 second
    success = getFrame(sec)
    while success:
        count = count + 1
        sec = sec + frameRate
        sec = round(sec, 2)
        success = getFrame(sec)

```

ตารางที่ ก.2 คำสั่งที่ใช้ในการสุ่มภาพขึ้นมาในทุกๆ 5 วินาที

```

import os
import cv2
import numpy as np
from PIL import Image
# Set folder path and number of frames to sample

```

เอกสารนี้เป็นเอกสารทบทวนใจสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่าในรูปแบบใดๆ ทั้งสิ้น อนึ่งทางเว็บไซต์ฉบับนี้จัดทำขึ้นโดยอัตโนมัติจากเอกสารทุกครั้งที่มีการแก้ไข

```

folder_path = '/path สำหรับเก็บเฟรมภาพทั้งหมด'
# Get list of all image file names in folder
image_file_names = os.listdir(folder_path)
# Sort image file names
image_file_names.sort()
# Select frames with a gap of 5 frames
selected_image_file_names = []
for i in range(0, len(image_file_names), 5):
    selected_image_file_names.append(image_file_names[i])
# Initialize count variable
count = 0
for image_name in selected_image_file_names:
    # Read image
    img_path = os.path.join(folder_path, image_name)
    img = Image.open(img_path)
    # Convert image to numpy array
    img_np = np.array(img)
    count += 1
cv2.imwrite("/path สำหรับเก็บเฟรมที่สุ่มขึ้นมาแล้ว/"+str(count)+".jpg", img_np)

```

### ตารางที่ ก.3 คำสั่งที่ใช้ในการตรวจจับและนับจำนวนบุคคล

```

import argparse
import time
from pathlib import Path
import cv2
import torch
import torch.backends.cudnn as cudnn
from numpy import random
from models.experimental import attempt_load
from utils.datasets import LoadStreams, LoadImages
from utils.general import check_img_size, check_requirements, check_imshow,
non_max_suppression, apply_classifier, \

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ไม่ว่ากรณีใดๆ ห้ามเผยแพร่หรือทำซ้ำโดยไม่ได้รับอนุญาต และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

scale_coords, xyxy2xywh, strip_optimizer, set_logging, increment_path
from utils.plots import plot_one_box
from utils.torch_utils import select_device, load_classifier, time_synchronized,
TracedModel
total_last=[]
def count(founded_classes,im0):
    model_values=[]
    aligns=im0.shape
    align_bottom=aligns[0]
    align_right=(aligns[1]/1.7 )
    for i, (k, v) in enumerate(founded_classes.items()):
        a=f"{k} = {v}"
        model_values.append(v)
        align_bottom=align_bottom-35
        cv2.putText(im0, str(a), (int(align_right),align_bottom),
cv2.FONT_HERSHEY_SIMPLEX, 1,(45,255,255),1,cv2.LINE_AA)
def total(founded_classes,im0,total_last):
    ab=im0.shape
    right=(ab[1]/1.7 )
    total_last.append(sum(founded_classes.values()))
    print("Total counted objects by every frame:",sum(total_last))
    print("Total objects in the current frame:",total_last[-1])
def detect(save_img=False):
    source, weights, view_img, save_txt, imgsiz, trace = opt.source, opt.weights,
opt.view_img, opt.save_txt, opt.img_size, not opt.no_trace
    save_img = not opt.nosave and not source.endswith('.txt') # save inference
images
    webcam = source.isnumeric() or source.endswith('.txt') or
source.lower().startswith(
        ('rtsp://', 'rtmp://', 'http://', 'https://'))
    # Directories
    save_dir = Path(increment_path(Path(opt.project) / opt.name,
exist_ok=opt.exist_ok)) # increment run

```

```

(save_dir / 'labels' if save_txt else save_dir).mkdir(parents=True, exist_ok=True)
# make dir
# Initialize
set_logging()
device = select_device(opt.device)
half = device.type != 'cpu' # half precision only supported on CUDA
# Load model
model = attempt_load(weights, map_location=device) # load FP32 model
stride = int(model.stride.max()) # model stride
imgsz = check_img_size(imgsz, s=stride) # check img_size
if trace:
    model = TracedModel(model, device, opt.img_size)
if half:
    model.half() # to FP16
# Second-stage classifier
classify = False
if classify:
    modelc = load_classifier(name='resnet101', n=2) # initialize
    modelc.load_state_dict(torch.load('weights/resnet101.pt',
map_location=device)['model']).to(device).eval()
# Set Dataloader
vid_path, vid_writer = None, None
if webcam:
    view_img = check_imshow()
    cudnn.benchmark = True # set True to speed up constant image size
inference
    dataset = LoadStreams(source, img_size=imgsz, stride=stride)
else:
    dataset = LoadImages(source, img_size=imgsz, stride=stride)
# Get names and colors
names = model.module.names if hasattr(model, 'module') else model.names
colors = [[random.randint(0, 255) for _ in range(3)] for _ in names]
# Run inference

```

```

if device.type != 'cpu':
    model(torch.zeros(1, 3, imgsz,
imgsz).to(device).type_as(next(model.parameters())) # run once
    old_img_w = old_img_h = imgsz
    old_img_b = 1
    t0 = time.time()
    for path, img, im0s, vid_cap in dataset:
        img = torch.from_numpy(img).to(device)
        img = img.half() if half else img.float() # uint8 to fp16/32
        img /= 255.0 # 0 - 255 to 0.0 - 1.0
        if img.ndimension() == 3:
            img = img.unsqueeze(0)
        # Warmup
        if device.type != 'cpu' and (old_img_b != img.shape[0] or old_img_h !=
img.shape[2] or old_img_w != img.shape[3]):
            old_img_b = img.shape[0]
            old_img_h = img.shape[2]
            old_img_w = img.shape[3]
            for i in range(3):
                model(img, augment=opt.augment)[0]
        # Inference
        t1 = time_synchronized()
        pred = model(img, augment=opt.augment)[0]
        t2 = time_synchronized()
        # Apply NMS
        pred = non_max_suppression(pred, opt.conf_thres, opt.iou_thres,
classes=opt.classes, agnostic=opt.agnostic_nms)
        t3 = time_synchronized()
        # Apply Classifier
        if classify:
            pred = apply_classifier(pred, modelc, img, im0s)
        # Process detections
        for i, det in enumerate(pred): # detections per image

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของสำนักงานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้นผู้ที่มีเหตุอันสมควร และต้องขออนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

if webcam: # batch_size >= 1
    p, s, im0, frame = path[i], '%g: ' % i, im0s[i].copy(), dataset.count
else:
    p, s, im0, frame = path, "", im0s, getattr(dataset, 'frame', 0)
p = Path(p) # to Path
save_path = str(save_dir / p.name) # img.jpg
txt_path = str(save_dir / 'labels' / p.stem) + (" if dataset.mode == 'image'
else f_{frame}') # img.txt

gn = torch.tensor(im0.shape)[[1, 0, 1, 0]] # normalization gain whwh
if len(det):
    # Rescale boxes from img_size to im0 size
    det[:, :4] = scale_coors(img.shape[2:], det[:, :4], im0.shape).round()
    founded_classes={} # Creating a dict to storage our detected items
    # Print results
    for c in det[:, -1].unique():
        n = (det[:, -1] == c).sum() # detections per class
        class_index=int(c)
        count_of_object=int(n)
        founded_classes[names[class_index]]=int(n)
        s += f'{n} {names[int(c)]}' * (n > 1), " # add to string
        count(founded_classes=founded_classes,im0=im0) # Applying
counter function
        total(founded_classes,im0,total_last)
    # Write results
    for *xyxy, conf, cls in reversed(det):
        if save_txt: # Write to file
            xywh = (xyxy2xywh(torch.tensor(xyxy).view(1, 4)) / gn).view(-
1).tolist() # normalized xywh
            line = (cls, *xywh, conf) if opt.save_conf else (cls, *xywh) # label
format
            with open(txt_path + '.txt', 'a') as f:
                f.write('%g ' * len(line).rstrip() % line + '\n')
        if save_img or view_img: # Add bbox to image

```

```

        label = f'{names[int(cls)]} {conf:.2f}'
        plot_one_box(xyxy, im0, label=label, color=colors[int(cls)],
line_thickness=1)

    # Print time (inference + NMS)
    print(f'{s}Done. ({(1E3 * (t2 - t1)):.1f}ms) Inference, ({(1E3 * (t3 - t2)):.1f}ms)
NMS')

    # Stream results
    if view_img:
        cv2.imshow(str(p), im0)
        cv2.waitKey(1) # 1 millisecond

    # Save results (image with detections)
    if save_img:
        if dataset.mode == 'image':
            cv2.imwrite(save_path, im0)
            print(f" The image with the result is saved in: {save_path}")
        else: # 'video' or 'stream'
            if vid_path != save_path: # new video
                vid_path = save_path
            if isinstance(vid_writer, cv2.VideoWriter):
                vid_writer.release() # release previous video writer
            if vid_cap: # video
                fps = vid_cap.get(cv2.CAP_PROP_FPS)
                w = int(vid_cap.get(cv2.CAP_PROP_FRAME_WIDTH))
                h = int(vid_cap.get(cv2.CAP_PROP_FRAME_HEIGHT))
            else: # stream
                fps, w, h = 30, im0.shape[1], im0.shape[0]
                save_path += '.mp4'
            vid_writer = cv2.VideoWriter(save_path,
cv2.VideoWriter_fourcc(*'mp4v'), fps, (w, h))
            vid_writer.write(im0)

    if save_txt or save_img:

```

```

        s = f"\n{len(list(save_dir.glob('labels/*.txt')))} labels saved to {save_dir /
'labels'}" if save_txt else "
```

```

    #print(f"Results saved to {save_dir}{s}")
    print(f'Done. ({time.time() - t0:.3f}s)')
if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument('--weights', nargs='+', type=str, default='yolov7.pt',
help='model.pt path(s)')
    parser.add_argument('--source', type=str, default='inference/images',
help='source') # file/folder, 0 for webcam
    parser.add_argument('--img-size', type=int, default=640, help='inference size
(pixels)')
    parser.add_argument('--conf-thres', type=float, default=0.25, help='object
confidence threshold')
    parser.add_argument('--iou-thres', type=float, default=0.45, help='IOU threshold
for NMS')
    parser.add_argument('--device', default="", help='cuda device, i.e. 0 or 0,1,2,3 or
cpu')
    parser.add_argument('--view-img', action='store_true', help='display results')
    parser.add_argument('--save-txt', action='store_true', help='save results to *.txt')
    parser.add_argument('--save-conf', action='store_true', help='save confidences in
--save-txt labels')
    parser.add_argument('--nosave', action='store_true', help='do not save
images/videos')
    parser.add_argument('--classes', nargs='+', type=int, help='filter by class: --class
0, or --class 0 2 3')
    parser.add_argument('--agnostic-nms', action='store_true', help='class-agnostic
NMS')
    parser.add_argument('--augment', action='store_true', help='augmented
inference')
    parser.add_argument('--update', action='store_true', help='update all models')
    parser.add_argument('--project', default='runs/detect', help='save results to
project/name')
    parser.add_argument('--name', default='exp', help='save results to
project/name')

```

```

parser.add_argument('--exist-ok', action='store_true', help='existing project/name
ok, do not increment')
parser.add_argument('--no-trace', action='store_true', help='don't trace model')
opt = parser.parse_args()
print(opt)
#check_requirements(exclude=('pycocotools', 'thop'))
with torch.no_grad():
    if opt.update: # update all models (to fix SourceChangeWarning)
        for opt.weights in ['yolov7.pt']:
            detect()

```

ตารางที่ ก.4 คำสั่งที่ใช้ในการตรวจจับและวัดองศาห่างไบหน้ากับกัองเพื่อประเมินความสนใจ

```

import cv2
import MediaPipe as mp
import numpy as np
import math
mp_face_mesh = mp.solutions.face_mesh
face_mesh = mp_face_mesh.FaceMesh(min_detection_confidence=0.5,
min_tracking_confidence=0.5,max_num_faces=20)
text = ""
cap = cv2.VideoCapture('HH.mp4')
count = 0
frame_count = 0
Attention = 0
face_id = []
while cap.isOpened():
    ret, image = cap.read()
    # if frame is read correctly ret is True
    if ret:
        # เพิ่มจำนวนเฟรม
        frame_count += 1

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งถ้าผมให้ตแบบลงเนื้อหา และตยั้งยั้งอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรนำไปใช้

```

else:
    # ถ้าอ่านไม่ได้แล้วออกจาก loop
    break

#image = cv2.rotate(image, cv2.ROTATE_90_COUNTERCLOCKWISE)
# Flip the image horizontally for a later selfie-view display
# Also convert the color space from BGR to RGB
#image = cv2.cvtColor(cv2.flip(image, 1), cv2.COLOR_BGR2RGB)
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB) #แปลงภาพจากสเกลสี BGR ให้
เป็นสเกลสี RGB

# To improve performance
image.flags.writeable = False
# Get the result
results = face_mesh.process(image)
# To improve performance
image.flags.writeable = True
# Convert the color space from RGB to BGR
image = cv2.cvtColor(image, cv2.COLOR_RGB2BGR)
img_h, img_w, img_c = image.shape
#cv2.putText(image, str(frame_count), (img_w-100, 50),
cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255), 2)
    cv2.putText(image, "Attention =" +str(Attention), (50, 350),
cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 255, 0), 2)
    face_3d = []
    face_2d = []
    if results.multi_face_landmarks:
        cv2.putText(image, str(len(results.multi_face_landmarks)), (50, 50),
cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255), 2)
        for face_landmarks in results.multi_face_landmarks:
            #print(len(results.multi_face_landmarks))
            for idx, lm in enumerate(face_landmarks.landmark):

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือกรรมสิทธิ์ในเนื้อหาบางส่วน โดยผู้จัดทำขึ้นเพื่อประโยชน์ทางการศึกษาเท่านั้น ไม่สามารถ  
 ให้นำไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของลิขสิทธิ์ หรือใช้เพื่อการค้าโดยไม่ได้รับอนุญาตจากเจ้าของลิขสิทธิ์ หากมีผู้พบเห็น  
 ไม่ถูกต้อง กรุณาแจ้งให้ทราบเพื่อดำเนินการปรับปรุงแก้ไข

```

if idx == 1:
    nose_2d = (lm.x * img_w, lm.y * img_h)
    nose_3d = (lm.x * img_w, lm.y * img_h, lm.z * 8000)
x, y = int(lm.x * img_w), int(lm.y * img_h)
# Get the 2D Coordinates
face_2d.append([x, y])
# Get the 3D Coordinates
face_3d.append([x, y, lm.z])
# Convert it to the NumPy array
face_2d = np.array(face_2d, dtype=np.float64)
# Convert it to the NumPy array
face_3d = np.array(face_3d, dtype=np.float64)
# The camera matrix
focal_length = 1 * img_w
cam_matrix = np.array([[focal_length, 0, img_h / 2],
                       [0, focal_length, img_w / 2],
                       [0, 0, 1]])
# The Distance Matrix
dist_matrix = np.zeros((4, 1), dtype=np.float64)
# Solve PnP
ret, rot_vec, trans_vec = cv2.solvePnP(face_3d, face_2d, cam_matrix,
dist_matrix)
# Get rotational matrix
rmat, jac = cv2.Rodrigues(rot_vec)
# Get angles
angles, mtxR, mtxQ, Qx, Qy, Qz = cv2.RQDecomp3x3(rmat)
# Get the y rotation degree
x = round(angles[0] * 360,2)
y = round(angles[1] * 360,2)
# print(y)
face_3d = []
face_2d = []
# Display the nose direction

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้นแต่มีเหตุที่เห็นสมควรและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

nose_3d_projection, jacobian = cv2.projectPoints(nose_3d, rot_vec,
trans_vec, cam_matrix, dist_matrix)

p1 = (int(nose_2d[0]), int(nose_2d[1]))
p2 = (int(nose_3d_projection[0][0][0]), int(nose_3d_projection[0][0][1]))
cv2.line(image, p1, p2, (255, 0, 0), 2)
index = 0 #เป็นตัวบอกตำแหน่งของ face_id คนนั้นๆ
#cv2.putText(image,str(index), p1, cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 255,
0), 2) แสดง index บนจุด p1

if face_id:
    for frame in face_id:
        if frame_count - frame["frame_count"] > 150 : #ลบเฟรมเก่า -> 5วินาที
            face_id.remove(frame)
            #fid["p1"][0]-50 < p1[0] & p1[0] < fid["p1"][0]+50: #range(+100) เช็คว่า
เป็นหน้าเก่าใหม่
            for fid in face_id:
                if math.dist(fid["p1"],p1) < 50: #จุดเก่าและใหม่ห่างกันไม่เกิน50 ก็จะได้ว่าเป็นคนเดียวกัน
                    fid["p1"] = p1 #เก่า = ใหม่ (เก่าถูกแทนที่ด้วยใหม่)(เอาค่าฝั่งขวาไปแทนฝั่งซ้าย) แทนที่จุดเก่าด้วยจุดใหม่
                    fid["frame_count"] = frame_count #นับเฟรมเก่าให้กลายเป็นเฟรมใหม่
                    break
            else:
                index+=1
            if index >= len(face_id):
face_id.append({"p1":p1,"frame_count":frame_count,"count":0,"Attention":False})
            else:
face_id.append({"p1":p1,"frame_count":frame_count,"count":0,"Attention":False})
            #face_id[index]["count"] #เรียกใช้ count ของตัวนั้นๆ
            # See where the user's head tilting
            if y < -10:
                text = "Looking Left"
                face_id[index]["count"] = 0

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้น ผู้ที่เห็นชอบเห็นชอบและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

elif y > 10:
    text = "Looking Right"
    face_id[index]["count"] = 0
elif x < -10:
    text = "Looking Down"
    face_id[index]["count"] = 0
elif x > 10:
    text = "Looking Up"
    face_id[index]["count"] = 0
else:
    if face_id[index]["count"] > 150:
        text = "Attention"
        if not face_id[index]["Attention"]: #เช็คว่เคยสนใจมาก่อนไหม
            Attention+=1
            face_id[index]["Attention"] = True #บันทึกว่หน้านั้นได้ถูกนับไปแล้
(เคยสนใจแล้)
        else:
            text = "Forward"
            face_id[index]["count"]+=1
    # Add the text on the image
    #cv2.putText(image, text, p1, cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255),
2)
    cv2.putText(image, str(index+1)+text+str(face_id[index]["count"]), (50,
70*(index+1)), cv2.FONT_HERSHEY_SIMPLEX, 0.8, (0, 0, 255), 2)
    #cv2.putText(image, "PITCH:" + str(x), (p1[0]+40,p1[1]+60),
cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255), 2)
    #cv2.putText(image, "YAW:" + str(y), (p1[0]+70,p1[1]+80),
cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255), 2)
else:
    count = 0
    cv2.imshow('Head Pose Estimation', image)
    if cv2.waitKey(2) == ord('q'):
        break

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
cap.release()
cv2.destroyAllWindows()
```

#### ตารางที่ ก.5 คำสั่งที่ใช้ในการสร้างเมทริกซ์ความสับสนแบบ Multiclass

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import metrics
from sklearn.metrics import classification_report
# ข้อมูลผลลัพธ์ที่ทำนาย
predicted = np.array([6,3,4,2,2,2,3,1,1,4])
actual = np.array([5,3,5,2,3,3,5,0,1,4])
confusion_matrix = metrics.confusion_matrix(actual, predicted)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix)
cm_display.plot()
plt.show()
print(classification_report(actual, predicted, digits=4, zero_division=0))
#zero_division เปลี่ยนค่า precision ที่ไม่มีการทำนายเป็น 0
```

#### ตารางที่ ก.6 คำสั่งที่ใช้ในการสร้างกราฟ ROC แบบ Multiclass

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, auc
from sklearn.preprocessing import label_binarize
# ข้อมูลผลลัพธ์ที่ทำนาย (ตัวอย่างเลขคลาส)
predicted = np.array([6, 3, 5, 3, 3, 4, 5, 2,
2, 4, 1, 1, 0, 0, 0, 0, 1, 1, 1,
2])
actual = np.array([5, 3, 5, 3, 3, 6, 5, 0, 1,
4, 1, 1, 1, 1, 1, 1, 1, 0, 1, 2])
# แปลงค่าผลลัพธ์เป็นรูปแบบ One-hot Encoding
predicted_onehot = label_binarize(predicted, classes=np.unique(predicted))
actual_onehot = label_binarize(actual, classes=np.unique(actual))
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้เฉพาะที่วิทยาลัยเท่านั้น ไม่ควรเผยแพร่ไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

n_classes = predicted_onehot.shape[1]
mean_fpr = np.linspace(0, 1, 100) # FPR ที่ใช้ในการสร้าง ROC curve
mean_tpr = 0.0 # TPR เฉลี่ย
mean_auc = 0.0 # AUC เฉลี่ย
# คำนวณ ROC curve และ AUC สำหรับแต่ละคลาสแยกตาม binary classification
for i in range(n_classes):
    fpr, tpr, _ = roc_curve(actual_onehot[:, i], predicted_onehot[:, i])
    roc_auc = auc(fpr, tpr)
    # ประมวลผลในแต่ละคลาส
    mean_tpr += np.interp(mean_fpr, fpr, tpr)
    mean_auc += roc_auc
# คำนวณค่าเฉลี่ยของ TPR และ AUC สำหรับทุกคลาส
mean_tpr /= n_classes
mean_auc /= n_classes
# สร้าง overall ROC curve
overall_fpr = mean_fpr
overall_tpr = mean_tpr
overall_auc = mean_auc
# พล็อต overall ROC curve
plt.plot(overall_fpr, overall_tpr, label='Overall ROC curve (AUC = %0.2f)' %
overall_auc)
plt.plot([0, 1], [0, 1], 'k--', label='Random')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Overall ROC Curve')
plt.legend(loc='lower right')
plt.show()

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



งานทะเบียนคณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
คำรับรองเล่มสหกิจศึกษา

วันที่ 9 เดือน กรกฎาคม พ.ศ 2566

ข้าพเจ้า นางสาวศิริณญา เทวกุล รหัสประจำตัว 62050837

นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ ขอรับรองว่าโครงการ  
สหกิจศึกษา เรื่อง

ชื่อภาษาไทย การศึกษาแบบจำลองการเรียนรู้การตรวจจับใบหน้าและนับจำนวนคนเพื่อ  
ประเมินความสนใจผ่านเทคนิคการเรียนรู้เชิงลึก

ชื่อภาษาอังกฤษ FACE DETECTION AND PEOPLE COUNTING LEARNING MODEL STUDY  
TO ASSESS ATTENTION THROUGH DEEP LEARNING TECHNIQUES

ปีการศึกษา 2565

เป็นผลงานวิจัยที่ได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อน  
เรียบร้อยแล้ว และได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่ม  
สหกิจศึกษาระดับสมบูรณ์แล้ว  
โปรแกรมอักขรวิสุทธิ 0.36%

ลงชื่อ.....ศิริณญา เทวกุล.....

(นางสาวศิริณญา เทวกุล)

นักศึกษา

ข้าพเจ้า ดร. สุกุณา ศรีโนมัย อาจารย์ที่ปรึกษาสหกิจศึกษา ได้ตรวจสอบสหกิจศึกษาของนักศึกษา  
ข้างต้นแล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็น  
หลักฐาน

ลงชื่อ.....สุกุน.....

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้