การพัฒนาแบบจำลองรู้จำใบหน้าแมวด้วยการเรียนรู้เชิงลึก

นายภัทรพล ศรีรักษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาการสารสนเทศมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2565

THE DEVELOPMENT OF A CAT FACE RECOGNITION MODEL USING DEEP LEARNING

PATTARAPOL SRIRAK

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Information Science in Information Technology
Suranaree University of Technology
Academic Year 2022

การพัฒนาแบบจำลองรู้จำใบหน้าแมวด้วยการเรียนรู้เชิงลึก

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

(รองศาสตราจารย์ ดร.ศิรปัฐช์ บุญครอง)
ประธานกรรมการ

(รองศาสตราจารย์ ดร.ธรา อั่งสกุล)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

(อาจารย์ ดร.ธรรมศักดิ์ เธียรนิเวศน์)
กรรมการ

(รองศาสตราจารย์ ดร.ยุพาพร รักสกุลพิวัฒน์)
รองอธิการบดีฝ่ายวิชาการและประกันคุณภาพ

(รองศาสตราจารย์ ดร.ธรา อั่งสกุล)
คณบดีสำนักวิชาศาสตร์และศิลป์ดิจิทัล

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภัทรพล ศรีรักษ์ : การพัฒนาแบบจำลองรู้จำใบหน้าแมวด้วยการเรียนรู้เชิงลึก
(THE DEVELOPMENT OF A CAT FACE RECOGNITION MODEL USING DEEP LEARNING)
อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.ธรา อั่งสกุล, 59 หน้า.

คำสำคัญ: การรู้จำใบหน้าแมว/การประมวลผลภาพ/การเรียนรู้เชิงลึก

ปัจจุบันจำนวนสัตว์เลี้ยงในประเทศไทยมีจำนวนสูงมากขึ้นในทุกปี ปัญหาที่เกิดขึ้นกับการเลี้ยงสัตว์คือการที่สัตว์เลี้ยงสูญหายหรือพลัดหลง และแมวหนึ่งในสัตว์เลี้ยงที่มีความนิยมมีอัตราพลัดหลงแล้วกลับมาเพียง 64% และอัตราของการเสียชีวิตระหว่างสูญหายของแมวยังสูงกว่าสุนัขถึง 2 เท่า การระบุตัวตนแมวอาจช่วยทำให้ระบุตำแหน่ง หรือให้ข้อมูลเกี่ยวกับแมวที่พลัดหลงได้ ซึ่งวิธีการระบุตัวตนถูกจำแนกออกเป็น 3 ประเภท ได้แก่ การระบุตัวตนแบบ การระบุตัวตนแบบกึ่งถาวร และการระบุตัวตนแบบถาวร ซึ่งแต่ละวิธีจะมีข้อดี ข้อเสีย และความเสี่ยงที่แตกต่างกัน จึงได้มีการนำเสนอวิธีการระบุตัวตนด้วยชีวมาตร ซึ่งงานวิจัยชิ้นนี้ได้เลือกใช้วิธีการระบุตัวตนด้วยใบหน้า ซึ่งเป็นการประยุกต์ใช้วิธีจากการระบุตัวตนด้วยใบหน้าของมนุษย์ และได้นำเสนอวิธีการรู้จำใบหน้าแมวโดยประกอบไปด้วย 3 ขั้นตอน 1) ขั้นตอนการตรวจจับใบหน้า จะเป็นการใช้แบบจำลองเพื่อตรวจจับใบหน้าแมวในภาพโดยใช้สถาปัตยกรรม YOLOv5Face ซึ่งเป็นสถาปัตยกรรมที่ดัดแปลงมาจาก YOLO ให้สามารถระบุจุดให้ที่ความสนใจในภาพได้ จึงสามารถตรวจจับใบหน้าและตำแหน่งของหู ตา และจมูกได้ 2) ขั้นตอนตรวจสอบใบหน้า จะใช้แบบจำลอง EfficientNetV2S ที่ถูกฝึกฝนไว้ล่วงหน้ากับสถาปัตยกรรมแบบแฝดสาม เพื่อเรียนรู้คุณลักษณะจากภาพใบหน้าคู่เหมือนของแมวตัวเดียวกัน และคู่ต่างของแมวคนละตัว และ 3) ขั้นตอนการระบุตัวตน เป็นกระบวนการที่จะนำคุณลักษณะที่สกัดได้จากใบหน้าแมวมาใช้เพื่อจัดเรียงหรือเปรียบเทียบข้อมูลชุดอื่น ๆ โดยใช้วิธีการ K-NN เพื่อหาจำนวนใบหน้าคล้ายคลึง และระบุว่าเป็นแมวตัวใดในฐานข้อมูล ซึ่งผลลัพธ์ของงานวิจัยชิ้นนี้แบบจำลองสำหรับการตรวจจับใบหน้ามีค่า mAP อยู่ที่ 0.9974 และขั้นตอนการระบุตัวตนแบบ 5 อันดับแรก มีความแม่นยำสูงสุดอยู่ที่ 89% ซึ่งสามารถนำขั้นตอนวิธีการพัฒนาแบบจำลองไปประยุกต์ใช้ในงานการรู้จำใบหน้าสำหรับสัตว์เลี้ยงประเภทอื่น หรือเพื่อประยุกต์ใช้ในการขึ้นทะเบียนสัตว์เลี้ยง การแพทย์สัตว์ และสุขภาพสัตว์เลี้ยง

สำนักวิชาศาสตร์และศิลป์ดิจิทัล
ปีการศึกษา 2565

ลายมือชื่อนักศึกษา _____
ลายมือชื่ออาจารย์ที่ปรึกษา _____

PATTARAPOL SRIRAK : THE DEVELOPMENT OF A CAT FACE RECOGNITION MODEL USING DEEP LEARNING. THESIS ADVISOR : ASSOC. PROF. THARA ANGSKUN, Ph.D., 59 PP.

Keyword: Cat Face Recognition/Image Processing/Deep Learning

The increasing number of pets in Thailand, particularly cats, has led to a growing concern regarding lost or missing animals. With a 64% rate of lost cats and a high mortality rate among them, cat identification methods to determine their location and provide information about lost cats. This research introduces a deep learning-based cat identification method consisting of three main steps: 1) Detection, using the YOLOv5 Face architecture to recognize the face and the position of ears, eyes, and nose; 2) Verification, utilizing a triplet neural network architecture with Triplet Loss Function and hard batch method for the training, and using Global Orthogonal Regularization; and 3) Identification, implementing FAISS and K-NN to find the number of similar faces and identify the cat within the database. The results demonstrate that the face detection model achieves a mAP of 0.9974, and the top 5 identification algorithm exhibits an accuracy rate of 89%, indicating the effectiveness of the proposed method for cat identification in various applications.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Institute of Digital Arts and Science
Academic Year 2022

Student's Signature _____
Advisor's Signature _____

กิตติกรรมประกาศ

ในวิทยานิพนธ์เล่มนี้สำเร็จด้วยดี ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธรา อั่งสกุล ผู้ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ ที่ได้กรุณาให้คำปรึกษา คำแนะนำ ให้กำลังใจ และให้ความช่วยเหลือด้วยความเมตตาเสมอมา ทั้งในด้านวิชาการและด้านการ ดำเนินการวิจัย

ที่จะต้องขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ศิรปัฐช์ บุญครอง ประธานกรรมการในการสอบวิทยานิพนธ์ อาจารย์ ดร.ธรรมศักดิ์ เธียรนิเวศน์ ที่ได้กรุณาเสียสละเวลาอันมีค่าในการพิจารณา และให้คำแนะนำในการแก้ไขปรับปรุงวิทยานิพนธ์ เสนอแนวทาง ความรู้ และให้ข้อเสนอแนะ ที่เป็นประโยชน์ต่อการวิจัยในครั้งนี้เป็นอย่างยิ่ง

สุดความสามารถในวันนี้ต้อง ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.จิติมนต์ อั่งสกุล อาจารย์ ดร.อรรคพล วงศ์กอบลาภ ผู้ช่วยศาสตราจารย์ ดร.พิชญสินี กิจวัฒนาถาวร และคณาจารย์ทุกท่าน ที่ได้ช่วยเหลือในการทำงาน ให้คำปรึกษา ความรู้ และประสบการณ์ พี่เก่ที่ช่วยดูแลจัดตารางงานผู้ช่วยสอนให้มีเวลาได้ศึกษาหาความรู้ พี่นาที่ดูแลและคอยตามเอกสารในทุกภาคการศึกษา ตั้งแต่ในสาขาวิชาเทคโนโลยีสารสนเทศ จนกระทั่งเป็นสำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี จนสามารถทำการวิจัยในครั้งนี้ได้จนเสร็จสิ้น

ก็ต้องขอขอบคุณ ขอบคุณ เพื่อน ๆ พี่ ๆ และน้อง ๆ ทุกท่านในตลอดระยะยาวนานที่ผ่านมาที่คอยช่วยเหลือ ทั้งด้านการเรียน การทำงาน การใช้ชีวิต และเป็นกำลังใจที่ดีเสมอมาให้แก่กัน

จบท้ายนี้ขอกราบขอบพระคุณทุกคนในครอบครัว ที่คอยให้กำลังใจ ขอบคุณเจ๊าะแจ๊ะ จู้จี้ จิ๋ว และพ่อแมวจี๋ยที่แม้ไม่ได้อยู่แล้ว ก็ยังคงเป็นแรงบันดาลใจและแรงผลักดันในการทำวิทยานิพนธ์จนสำเร็จลงได้ด้วยดี ความดีอย่างใดที่เกิดจากวิทยานิพนธ์เล่มนี้ ข้าพเจ้าขอมอบให้กับ บิดา มารดา และครูอาจารย์ที่เคารพทุกท่าน ที่ได้ให้วิชาความรู้ และเป็น กำลังใจให้ข้าพเจ้าเสมอมาจนสำเร็จการศึกษา

ภัทรพล ศรีรักษ์

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย) ..... ก
บทคัดย่อ (ภาษาอังกฤษ) ..... ข
กิตติกรรมประกาศ ..... ค
สารบัญ ..... ง
สารบัญตาราง ..... ฉ
สารบัญรูป ..... ช
บทที่ **1**
1 บทนำ **1**
1.1 ที่มาและความสำคัญ 1
1.2 วัตถุประสงค์งานวิจัย 4
1.3 วัตถุประสงค์งานวิจัย 4
1.4 ข้อตกลงเบื้องต้น 4
1.5 ขอบเขตการวิจัย 4
1.6 ประโยชน์ที่คาดว่าจะได้รับ 5
1.7 คำอธิบายศัพท์ 5
2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง **7**
2.1 แนวคิดเกี่ยวกับการเรียนรู้ของเครื่อง 7
2.2 แนวคิดเกี่ยวกับการเรียนรู้เชิงลึก 9
2.3 แนวคิดและกระบวนการถ่ายโอนความรู้ 12
2.4 แนวคิดและกระบวนการเกี่ยวกับการตรวจจับวัตถุ 13
2.5 กระบวนการยืนยันใบหน้า 26
2.6 กระบวนการระบุใบหน้า 34
2.7 งานวิจัยที่เกี่ยวข้องกับการรู้จำสัตว์เลี้ยง 36
3 วิธีดำเนินการวิจัย **40**
3.1 วิธีวิจัย 40
3.2 ชุดข้อมูลที่ใช้ในงานวิจัย 43
3.3 กระบวนการตรวจจับใบหน้า 44
3.4 กระบวนการตรวจสอบใบหน้า 44
3.5 กระบวนการระบุตัวตน 47
3.6 กระบวนการวัดผล 47

สารบัญ (ต่อ)

หน้า

4 การทดสอบและอภิปรายผล **49**

4.1 การทดสอบการตรวจจับใบหน้า 49

4.2 ชุดข้อมูลที่ใช้ในการทดสอบการระบุตัวตน 50

4.3 การทดสอบการระบุตัวตน 51

4.4 อภิปรายผล 53

5 สรุปและข้อเสนอแนะ **54**

5.1 ผลการวิจัย 54

5.2 การประยุกต์ผลการวิจัย 54

5.3 ปัญหาและข้อเสนอแนะ 55

รายการอ้างอิง **56**

ประวัติผู้เขียน **59**

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ **หน้า**

4.1 ตารางแสดงค่าความแม่นยำของแบบจำลองเมื่อปรับ M และ k 52

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญรูป

รูปที่ **หน้า**

1.1 รูปแบบการระบุตัวตนสัตว์เลี้ยง 2
2.1 ภาพอธิบายกระบวนการเรียนรู้แบบมีผู้สอน 8
2.2 ภาพอธิบายกระบวนการเรียนรู้แบบไม่มีผู้สอน 8
2.3 ตัวอย่างการเรียนรู้แบบเสริมกำลัง 9
2.4 ลักษณะการทำงานภายในนิวรอน 10
2.5 ส่วนประกอบของนิวรอล ในโครงข่ายประสาทเทียม 10
2.6 ประเภทของการเรียนรู้เชิงลึก 11
2.7 ภาพตัวอย่างการนำ Transfer Learning มาใช้กับงานใหม่ 12
2.8 ขั้นตอนการทำฮาร์ (Harr-Like) 13
2.9 ตัวอย่างการทำฮาร์ (Haar-Like) 14
2.10 ขั้นตอนการหาคุณลักษณะเฉพาะในการตรวจจับวัตถุด้วยฮอก (HOG) 14
2.11 ตัวอย่างการทำงานของฮอก (HOG: Histrograms of Oriented Gradient) 15
2.12 ขั้นตอนของซีเอ็นเอ็นเพื่อทำการคัดแยกข้อมูล 16
2.13 ตัวอย่างการทำคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากขนาด 2×2 17
2.14 ตัวอย่างการทำพูลลิ่งด้วยค่าสูงสุด 18
2.15 ตัวอย่างการทำพูลลิ่งด้วยค่าเฉลี่ย 18
2.16 ตัวอย่างการทำพูลลิ่งด้วย L2-norm 19
2.17 ตัวอย่างขั้นตอนการทำอาร์-ซีเอ็นเอ็น 21
2.18 ตัวอย่างขั้นตอนการทำฟาสต์เตอร์อาร์-ซีเอ็นเอ็น 22
2.19 โครงสร้างการทำงานของโยโล 22
2.20 ตัวอย่างการทำงานของโยโล 23
2.21 ภาพตัวอย่างผลลัพธ์การตรวจจับใบหน้ามนุษย์ของ YOLOv5-Face 24
2.22 โครงสร้างการทำงานของเอสเอสดี 24
2.23 ตัวอย่างขั้นตอนการแปลงภาพใบหน้าเป็น Embedding 26
2.24 ตัวอย่างการแสดง Embedding Space ในรูปแบบ 3 มิติ 27
2.25 การสกัดคุณลักษณะด้วย Eigenface 27
2.26 ภาพตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วย Eigenface 28
2.27 ภาพขั้นตอนการสกัดคุณลักษณะด้วย LBP 29
2.28 ภาพตัวอย่างการทำงานของโมเดล Siamese Network 29
2.29 ภาพตัวอย่างการแสดง Embedding Space ในรูปแบบ 3 มิติ 31
2.30 ภาพตัวอย่างวิธีการปรับระยะของ Triplet Loss 32
2.31 ตัวอย่างระยะห่างของชุดข้อมูลแต่ละประเภท 33

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

2.32 ภาพจำลองการใช้ GOR เพื่อเพิ่มการกระจายตัวของข้อมูล 34
2.33 การสร้างเส้นแบ่งข้อมูลของเอสวีเอ็ม (SVM) 34
2.34 การทำงานของเคเอ็นเอ็น 35
2.35 ภาพอธิบายกระบวนทำงานของ FAISS 36
2.36 ภาพเปรียบเทียบวิธีการรู้จำใบหน้าแมวจากการทบทวนวรรณกรรม 39
3.1 ภาพตัวอย่างขั้นตอนการรู้จำใบหน้าแมว 42
3.2 ตัวอย่างข้อมูลภาพแมวจาก Kaggle 43
3.3 ภาพตัวอย่างใบหน้าแมวที่ถูกนำเสนอในงานของ Lin และ Kou 43
3.4 ภาพตัวอย่างผลลัพธ์การระบุตำแหน่งใบหน้าแมวและการปรับแต่งภาพ 44
3.5 ภาพตัวอย่างการฝึกฝนใบหน้าแมวด้วย Triplet Network 45
3.6 ภาพแสดงกระบวนสกัดคุณลักษณะและระบุใบหน้าแมว 46
3.7 ภาพตัวอย่างการเปรียบเทียบระหว่าง Hard Triplet 46
3.8 ตารางการวัดประสิทธิภาพของแบบจำลอง 47
4.1 กราฟแสดงผล mAP จากการฝึกฝนแบบจำลองสำหรับการตรวจจับใบหน้าแมว 49
4.2 ภาพตัวอย่างการตรวจจับใบหน้าแมวโดยการระบุจุดสำคัญบนใบหน้า 50
4.3 ตัวอย่างผลลัพธ์ใบหน้าแมวในชุดข้อมูลสำหรับการระบุตัวตน 50
4.4 การกระจายตัวของข้อมูลใบหน้าแมวที่ถูกประมวลผลและลดมิติให้อยู่ในลักษณะ 3 มิติ 51
4.5 ภาพกราฟแสดงค่า Loss ในการฝึกฝนชุดข้อมูลในการจำแนกใบหน้า 51
4.6 ภาพตัวอย่างการทำนายเมื่อ M=1 และ K=5 52
4.7 ภาพผลลัพธ์ค่า Confusion Matrix ขนาด 24x24 ของแมวจำนวน 24 คลาส 52

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

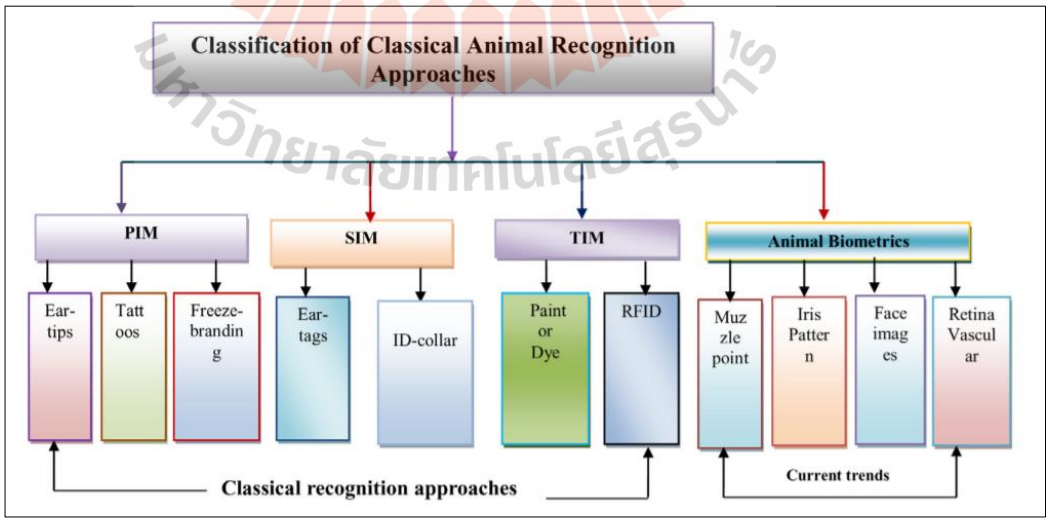
ปัจจุบันประเทศไทยมีอัตราจำนวนผู้เลี้ยงสัตว์เลี้ยงสูงมากขึ้น เนื่องจากรูปแบบการใช้ชีวิตของคนที่ครองโสดมากขึ้น คู่แต่งงานที่ไม่มีบุตร สังคมผู้สูงอายุ รวมถึงสื่อโซเชียลและกลุ่มของคนรักสัตว์ที่มีการแชร์ภาพพฤติกรรมของสุนัข แมว และสัตว์เลี้ยงอื่น ๆ บนโลกออนไลน์มากขึ้น ผู้บริโภคมีกำลังซื้อทำให้หันมานิยมเลี้ยงสัตว์เพิ่มขึ้น และการเลี้ยงสัตว์เลี้ยงเพื่อการนันทนาการมีมากขึ้น เนื่องจากการนันทนาการจัดเป็นศาสตร์แขนงหนึ่งซึ่งช่วยให้เกิดความผ่อนคลาย รวมทั้งยังเข้ามามีส่วนร่วมในการลดความตึงเครียดของสังคมได้อีกด้วย เพราะประโยชน์และคุณค่าของนันทนาการนั้นจัดเป็นความต้องการขั้นพื้นฐานของมนุษย์ (Butler, 1959) นอกจากนี้อัตราการเกิดในไทยในประเทศไทยในช่วงปี 2560-2564 ยังคงมีแนวโน้มลดลง (สำนักงานสถิติแห่งชาติ, 2565) และในทางกันข้ามอัตราการเลี้ยงสัตว์ยังคงสูงขึ้นเรื่อย ๆ (ฐานข้อมูลเพื่อการขึ้นทะเบียนสุนัข-แมว, 2564) สะท้อนให้เห็นเทรนด์การเลี้ยงสัตว์เพื่อเป็นลูก (Pet Parent) ที่เติบโตในประเทศไทย 2560-2564 อัตราการเกิดภายในประเทศไทยลดลงอย่างต่อเนื่อง (สำนักงานสถิติแห่งชาติ, 2565) แต่การเลี้ยงสัตว์เลี้ยงมีอัตราเติบโตมากขึ้นอย่างเนื่อง จาก 13 ล้านตัวในปี 2560 และมีจำนวน 14.5 ล้านตัวในปี 2563 (ฐานข้อมูลเพื่อการขึ้นทะเบียนสุนัข-แมว, 2564) ซึ่งในปี 2564 ธุรกิจสินค้าและบริการเกี่ยวเนื่องกับสัตว์เลี้ยงมีมูลค่า 3,954 ล้านบาท และธุรกิจอาหารสัตว์เลี้ยงมีมูลค่า 40,638 ล้านบาทและยังคาดว่าจะเติบโตเฉลี่ยปีละ 8.4% (Compound Annual Growth Rate: CAGR) มาอยู่ที่ 66,748 ล้านบาทในปี 2569 (วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล, 2566)

ปัญหาที่เกิดขึ้นกับการเลี้ยงสัตว์คือการที่สัตว์เลี้ยงสูญหายหรือพลัดหลง และด้วยจำนวนสัตว์เลี้ยงที่ขยายจำนวนเพิ่มมากขึ้นทำให้ในประเทศไทยมีจำนวนสัตว์เลี้ยงจำนวนถึง 10.4 ล้านตัว เป็นสัตว์จรจัด 1.2 ล้านตัว (สำนักงานปศุสัตว์, 2561) ด้วยจำนวนสัตว์ที่มากขึ้นจึงเป็นเรื่องยากที่ผู้พบเจอจะสามารถแยกได้ว่าเป็นสัตว์ที่พบนั้นเป็นสัตว์เลี้ยงที่พลัดหลงหรือเป็นสัตว์จรจัด จึงไม่สามารถที่จะหาแจ้งหรือส่งคืนไปยังเจ้าของได้

แมวเป็นหนึ่งในสัตว์เลี้ยงที่มีความนิยมสูง สถิติประชากรผู้เลี้ยงแมวทั่วโลกว่า ประเทศไทยมีประชากรผู้เลี้ยงแมวอยู่ที่อันดับ 8 (33%) ของโลก (Statista, 2017) และเมื่อเปรียบเทียบสัดส่วนการสัตว์เลี้ยงในประเทศได้แก่ สุนัข 40.4% แมว 37.1% และกลุ่มสัตว์เลี้ยง Exotic 22.6% (วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล, 2565) แมวจึงอยู่ในสัตว์เลี้ยงกลุ่มใหญ่ในประเทศไทย นอกจากนี้แมวเป็นสัตว์ที่รักอิสระ ชอบเดินเตร็ดเตร่ และมีความระมัดระวังตัว ทำให้ยากต่อการพบเห็นและนำแมวที่หายไปกลับไปหาเจ้าของอีกครั้ง นอกจากนี้แมวเมื่อพลัดหลงแล้วมีอัตราการกลับมาเพียง 64% เท่านั้น แตกต่างจากสุนัขที่สูงถึง 93% และอัตราของการเสียชีวิตระหว่างสูญหายของแมวยังสูงกว่าสุนัขถึง 2 เท่า (lostpetresearch, 2020) การระบุตัวตนแมวอาจช่วยทำให้ระบุตำแหน่ง หรือให้ข้อมูลเกี่ยวกับแมวที่พลัดหลงได้

ปัจจุบัน วิธีการระบุตัวตนของสัตว์เลี้ยงสามารถทำได้โดยการใช้วัตถุเพื่อระบุตัวตน ซึ่งสามารถจำแนกออกเป็น 3 ประเภท (ดังรูปที่ 1.1) ได้แก่ 1) การระบุตัวตนแบบชั่วคราว (Temporary Recognition Method: TRM) 2) การระบุตัวตนแบบกึ่งถาวร (Semi-Permanent Recognition Method: SRM) และ 3) การระบุตัวตนแบบถาวร (Permanent Recognition Method: PRM) โดยการระบุตัวตนแบบชั่วคราว เช่น การห้อยแท็กบีแอลอี (BLE: Bluetooth Low Energy) และการย้อมสี การระบุตัวตนแบบกึ่งถาวร เช่น การติดเบอร์ที่หูและการสวมปลอกคอ ส่วนการระบุตัวตนแบบถาวร เช่น การประทับเย็น การสักหู การตัดใบหู หรือการฝังไมโครชิป ซึ่งวิธีการเหล่านี้ค่อนข้างรุนแรงและมีความเสี่ยงต่อการติดเชื้อและยากต่อการดูแล ในขณะที่การใส่ปลอกคอหรือการห้อยแท็กบีแอลอีนั้นมีค่าใช้จ่ายและมีโอกาสสูญหายได้ง่ายเนื่องจากเป็นอุปกรณ์สวมใส่ (Kumar, 2016)

มีหลายงานวิจัยที่พยายามแก้ไขปัญหาในการระบุตัวตนของสัตว์โดยใช้ชีวมาตรโดยระบุตัวตนจากร่างกายของสัตว์เอง เพื่อช่วยลดค่าใช้จ่ายของอุปกรณ์และยังลดความรุนแรงที่ใช้ในการระบุตัวตนอีกด้วย (Kumar, 2014) ซึ่งวิธีที่ใช้ชีวมาตรในการระบุตัวตนทั้งในปศุสัตว์และสัตว์เลี้ยงมีวิธีการที่หลากหลายเช่น การระบุลักษณะสำคัญ (Tharwat, 2014) ลักษณะรูปแบบ (Kumar, 2018) ลักษณะใบหน้า (Tzu-Yuan Lin, 2018) รูปแบบม่านตา (Gonzales-Barron et al., 2008) ซึ่งจากวิธีการใช้ชีวมาตรระบุตัวตนแต่ละวิธีนั้นมี ข้อจำกัดที่แตกต่างกันเช่น หากเป็นระบุตัวตนโดยลักษณะภาพรวมโดยใช้ ลักษณะทางกายภาพ สีขนหรือสีผิว ก็อาจเกิดความผิดเพี้ยนจากลักษณะท่าทางหรือระยะห่าง หากเป็นลักษณะที่มีรายละเอียดสูงเช่น เส้นเลือดดวงตา ลายกล้ามเนื้อ รูปแบบดวงตา ก็จำเป็นต้องใช้ภาพถ่ายระยะใกล้ซึ่งเหมาะกับปศุสัตว์ หรือสัตว์เลี้ยงที่อยู่ในพื้นที่ปิดมากกว่า งานชิ้นนี้จึงเลือกการระบุตัวตนโดยใบหน้าที่สามารถตรวจจับได้ง่าย และมีรายละเอียดที่สามารถใช้ในการหาลักษณะเพื่อระบุตัวตน



รูปที่ 1.1 รูปแบบการระบุตัวตนสัตว์เลี้ยง

เทคโนโลยีรู้จำใบหน้าคือเทคโนโลยีที่ใช้เพื่อระบุตัวตน โดยจะทำการตรวจสอบใบหน้าโดยค้นหาใบหน้าจากภาพหรือวีดีโอ จากนั้นจะนำภาพใบหน้าที่ได้ผ่านกระบวนการวิเคราะห์เพื่อหาลักษณะของใบหน้าเพื่อตรวจสอบว่าใบตรงกับข้อมูลที่มีอยู่หรือไม่ ทำให้สามารถระบุตัวตนได้ ซึ่งระบบรู้จำใบหน้าประสบความสำเร็จอย่างมากในการใช้งานเพื่อระบุตัวตนมนุษย์ (Schroff et al., 2015) ซึ่งได้นำเสนอวิธีการระบุตัวตน งานวิจัยชิ้นนี้ได้ศึกษาวิธีการระบุตัวตนสัตว์เลี้ยงจากใบหน้า โดยศึกษาจากใบหน้าแมวเพื่อค้นหาวิธีการที่จะระบุใบหน้าและค้นหาลักษณะบนใบหน้าเพื่อระบุตัวตน

การรู้จำใบหน้าของแมวเป็นงานวิจัยที่ค่อนข้างใหม่ซึ่งมีจุดมุ่งหมายเพื่อพัฒนาอัลกอริทึมและระบบการเรียนรู้ของเครื่อง ที่สามารถระบุและจำแนกแมวแต่ละตัวตามลักษณะใบหน้าซึ่งจะนำไปประยุกต์ใช้งานได้หลากหลาย เช่นระบุแมวที่หายไปเพื่อส่งกลับคืนเจ้าของหรือการติดตามการเคลื่อนไหวของแมวเพื่อการอนุรักษ์และลงทะเบียนสัตว์เลี้ยง

ปัจจุบันการศึกษาขั้นตอนวิธี แบบจำลองและเครื่องมือรู้จำใบหน้าแมว ยังมีความท้าทายจากความหลากหลายของลักษณะใบหน้าและรูปลักษณ์ของแมว ในการพัฒนาวิธีการเรียนรู้เพื่อระบุใบหน้า และขั้นตอนการจำแนกคุณลักษณะ เพื่อให้ได้ซึ่งผลลัพธ์ด้านประสิทธิภาพและความเร็วที่ดีที่สุด ซึ่งแต่ละงานมีวิธีการแตกต่างกันหลากหลายวิธี เช่น Cat Face Recognition Based on MFCC and GMM (Y. Fan, C. Yang, and C. Chen, 2021) ที่ศึกษาคุณลักษณะของแมวโดยใช้ สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepstrum coefficient, MFCC) สำหรับการวิเคราะห์คุณลักษณะและ (Gaussian mixture modelm, GMM) หรือ Cat face recognition using Siamese network (Li, H., & Zhang, W.,2022) ที่ใช้วิธีการเปรียบเทียบภาพด้วย Siamese Network ที่ใช้โครงขายประสาทเทียมเพื่อเปรียบเทียบผลลัพธ์ หรืองาน Cat face recognition using deep learning ของ Tzu-Yuan Lin, 2018 ใช้การระบุ Feature แต่ละส่วนของใบหน้าเช่น ตา หู จมูก และหน้าผาก เพื่อนำมาเปรียบเทียบโดยการใช้ Faster R-CNN งาน Pet Cat Face Verification and Identification (Adam Klein, 2019) ที่นำเสนอการตรวจระบุตัวตนแบบ 2 ขั้นตอน ที่แบ่งเป็นขั้นตอนการตรวจจับและระบุตัวตนโดยใช้เทคนิคการเปรียบเทียบแฝดสาม

อย่างไรก็ตามวิธีการในแต่ละงานยังมีข้อจำกัดที่แตกต่างกันเช่น การตรวจจับใบหน้าของแมวเมื่ออยู่ร่วมกับสภาพแวดล้อม (Kumar 2018) หรือการระบุตัวตนของแมวที่มีลักษณะคลุมเครือเช่น แมวดำที่ลักษณะใบหน้าของ สีจมูก หน้าผาก เป็นสีเดียวกัน ทำให้ระบุตัวตนได้ยาก (Tzu-Yuan Lin 2018) การระบุใบหน้าแมวได้เพียงตัวจากภาพ จึงมีงานวิจัยที่เพิ่มความสามารถในการรู้จำใบหน้าแมวออกมาอย่างต่อเนื่อง

งานวิจัยนี้จึงนำเสนอวิธีการพัฒนาแบบจำลองรู้จำใบหน้าแมวแบบ 2 ขั้นตอนดังนี้ 1.ขั้นตอนการตรวจจับใบหน้า จะใช้สถาปัตยกรรมของแบบจำลองตรวจจับวัตถุ โยโล เวอร์ชันที่ 5-เฟส (YoLoV5-Face: You Only Look Once Version5 - Face) ซึ่งเป็นสถาปัตยกรรม ที่ปรับแต่ง โยโล เวอร์ชั่นที่ 5 (YoLoV5: You Only Look Once Version5) โดย YoLoV5 จะเป็นการตรวจจับวัตถุ

แบบขั้นตอนเดียว แต่ YoLoV5-Face จะถูกปรับแต่งให้สามารถระบุจุดสำคัญบนใบหน้าได้ โดยใช้แบบจำลองที่ถูกฝึกฝนมาแล้ว(pre-trained model) มาปรับแต่งและใช้ในการระบุบริเวณใบหน้าและจุดสำคัญบนใบหน้าแมว และจัดเรียงอยู่ในลักษณะใบหน้าตรงเพื่อที่จะนำไปเปรียบเทียบกับข้อมูลที่มีในลักษณะเดียวกัน 2.ขั้นตอนการระบุตัวตน จะใช้โครงสร้างแบบ Triplet Neural Network โดยใช้การถ่ายโอนความรู้ (Transfer Learning) จากแบบจำลอง EfficientNetV2s และสกัดคุณลักษณะของภาพออกเป็น Embedding Vector และใช้อัลกอลิทึม และใช้ไลบรารี่ FAISS เพื่อแบ่งกลุ่มข้อมูลของใบหน้า เพื่อใช้ในการระบุตัวตนของสัตว์เลี้ยงซึ่งจะช่วยให้สามารถรู้จำและระบุตัวตนแมวได้ เพื่อเป็นแนวทางการระบุตัวตนสัตว์เลี้ยง และลดค่าใช้จ่ายของอุปกรณ์ ลดความเสี่ยงด้านการสูญหายของวัตถุ ช่วยลดความเสี่ยงในการปฏิสัมพันธ์เชิงร่างกายแก่สัตว์เลี้ยงที่เกิดขึ้นได้ในการระบุตัวตนแบบดั้งเดิมและยังมุ่งเน้นการพัฒนาเพื่อเพิ่มความแม่นยำและเพิ่มความสามารถในการรู้จำใบหน้าแมว

1.2 เป้าหมายงานวิจัย

เพื่อออกแบบและพัฒนาแบบจำลองการรู้จำใบหน้าแมวเพื่อระบุตัวตนโดยใช้เทคนิคการเรียนรู้เชิงลึก

1.3 วัตถุประสงค์งานวิจัย

1.3.1 แบบจำลองสามารถค้นหาใบหน้าแมวได้เมื่ออยู่ในสภาพแวดล้อมร่วมกับวัตถุอื่น ได้ถูกต้องอย่างน้อยร้อยละ 80

1.3.2 แบบจำลองสามารถระบุตัวตนด้วยหน้าแมวที่ตรวจจับเพื่อระบุตัวตนได้อย่างถูกต้องอย่างน้อยร้อยละ 80

1.4 ข้อตกลงเบื้องต้น

1.4.1 แบบจำลองที่พัฒนาขึ้นจะเรียนรู้จากการจดจำใบหน้าเท่านั้น

1.4.2 แบบจำลองที่พัฒนาขึ้นจะทำงานได้บนระบบที่สนับสนุนภาษาไพธอน (Python) เท่านั้น

1.5 ขอบเขตการวิจัย

1.5.1 งานวิจัยนี้มุ่งเน้นการพัฒนาขั้นตอนวิธีเพื่อสร้างแบบจำลองที่ใช้ค้นหาใบหน้าของแมวเมื่ออยู่ในสภาพแวดล้อมร่วมกับวัตถุอื่น ๆ ได้ และสามารถรู้จำใบหน้าแมวที่ค้นพบ และระบุตัวตนได้อย่างถูกต้อง

1.5.2 โดยระบบที่พัฒนาขึ้น จะประกอบไปด้วย 2 ส่วนคือ ส่วนแรกคือส่วนของการค้นหาใบหน้าของแมว เพื่อที่จะนำใบหน้าที่พบไปปรับเป็นขนาดที่ต้องการและนำไปใช้ในการ วิเคราะห์ต่อไป ส่วนที่สองเป็นส่วนของการนำใบหน้าแมวไปหาองค์ประกอบด้วยการเรียนรู้ เชิงลึก จากนั้นระบบจะนำข้อมูลที่ได้เปรียบเทียบกับข้อมูลในแบบจำลองว่ามีใบหน้ามีความคล้ายคลึงกับใบหน้าในฐานข้อมูลหรือไม่

มหาวิทยาลัยเทคโนโลยีสุรนารี

1.5.3 ใบหน้าของแมวที่จะใช้เพื่อระบุตัวตนจะต้องเห็นภาพใบหน้าและมีองค์ประกอบใบหน้าครบถ้วนชัดเจน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 ได้แบบจำลองสามารถตรวจจับใบหน้าแมวได้ เมื่ออยู่ในสภาพแวดล้อมร่วมกับวัตถุอื่น ๆ ได้

1.6.2 ได้แบบจำลองสามารถแยกแยะใบหน้าแมวที่ตรวจจับ และระบุตัวตนได้อย่างถูกต้อง

1.7 คำอธิบายศัพท์

1.7.1 เทคนิคการเรียนรู้เชิงลึก หมายถึงเทคนิคซึ่งเป็นส่วนของการเรียนรู้ของเครื่อง (Machine Learning) โดยใช้การทำงานแบบโครงข่ายประสาทเทียม (Neural Network) เป็นการฝึกฝนคอมพิวเตอร์ให้สามารถทำงานได้เหมือนมนุษย์ ที่มีการเรียนรู้หลายลำดับชั้นเพื่อทำการค้นหาลักษณะของข้อมูล (Goodfellow et al., 2016). ตัวอย่างเช่นการระบุตัวเลขหรือตัวอักษรหรือใบหน้า ซึ่งในงานวิจัยนี้เทคนิคการเรียนรู้เชิงลึกจะใช้เพื่อจดจำใบหน้าแมว

1.7.2 แบบจำลอง หมายถึงชุดของการประมวลผลรูปแบบการเรียนรู้ของเครื่อง ซึ่งเป็นตัวแทนทางคณิตศาสตร์ของกระบวนการในการสร้างรูปแบบการเรียนรู้ของเครื่อง โดยจะต้องใช้ข้อมูลในการฝึกกับขั้นตอนวิธีการเรียนรู้ของเครื่อง เพื่อเรียนรู้วิธีในการทำงานหรือประมวลผล

1.7.3 โครงข่ายประสาทเทียม หมายถึงแบบจำลองทางคอมพิวเตอร์ใช้สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบมีการเชื่อมต่อกัน โดยพัฒนาตามแนวคิดโครงข่ายไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (neurons) และ จุดประสานประสาท (synapses) ซึ่งทำให้เกิดข่ายงานประสาทที่เกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน (ณัฐโชติ พรหมฤทธิ์, สัจจาภรณ์ ไวจรรยา, 2564)

1.7.4 การระบุตัวตน หมายถึงกระบวนการตรวจสอบเอกลักษณ์หรือลักษณะของบุคคลหรือสัตว์ ว่าเป็นผู้ที่อ้างถึงนั้นจริง ๆ หรือเป็นระบุได้ว่าผู้ที่ถูกตรวจสอบนั้นคือใคร

1.7.5 การรู้จำใบหน้า หมายถึง การระบุตัวตนของบุคคลหรือสัตว์ จากประสบการณ์ข้อมูลหรือความรู้ก่อนหน้า โดยใช้การจำลักษณะสำคัญหรือลักษณะของข้อมูล แล้วจึงเปรียบเทียบว่ามีความตรงกันหรือใกล้เคียงกันเพื่อระบุตัวตน

1.7.6 โยโล (YOLO: You Only Look Once) เป็นแบบจำลองสำหรับการตรวจจับวัตถุแบบเรียลไทม์ ที่มีความโดดเด่นเรื่องความเร็วและความถูกต้อง

1.7.7 การถ่ายโอนความรู้ (Transfer Learning) เป็นเทคนิคในการเรียนรู้ของเครื่องโดยนำแบบจำลองที่พัฒนาขึ้นสำหรับงานหนึ่งมาใช้ซ้ำ เป็นจุดเริ่มต้นสำหรับแบบจำลองในงานที่สอง แนวคิดเบื้องหลังการเรียนรู้แบบโอนย้ายคือแบบจำลองที่ได้รับการฝึกฝนในชุดข้อมูลขนาดใหญ่สามารถใช้เพื่อแยกคุณสมบัติที่เป็นประโยชน์สำหรับงานอื่นแต่เกี่ยวข้องกัน วิธีนี้สามารถประหยัดเวลาและทรัพยากรเมื่อเทียบกับการฝึกแบบจำลองตั้งแต่ต้น (ณัฐโชติ พรหมฤทธิ์, สัจจาภรณ์ ไวจรรยา, 2564)

1.7.8 แบบจำลองที่ฝึกไว้ล่วงหน้า (Pretrained Model) คือแบบจำลองที่สร้างขึ้นเพื่อแก้ปัญหาที่คล้ายกัน โดยสามารถใช้แบบจำลองที่ได้รับการฝึกอบรมจากปัญหาอื่นเป็นจุดเริ่มต้นสำหรับแบบจำลองใหม่ที่ได้รับการปรับแต่งอย่างละเอียดเพื่อให้รู้จักชุดของคลาสใหม่ สิ่งนี้สามารถประหยัดเวลาและทรัพยากรได้มากเมื่อเทียบกับการฝึกแบบจำลองใหม่

1.7.9 โครงข่ายประสาทแบบแฝดสาม (Triplet Neural Network) เป็นประเภทของโครงข่ายประสาทเทียมที่ออกแบบมาเฉพาะสำหรับงานที่เกี่ยวข้องกับการ embedding และการเปรียบเทียบความเหมือนของจุดข้อมูล โดยจะเรียนรู้คุณสมบัติการฝังตัวของจุดข้อมูล เพื่อให้ระยะห่างระหว่างจุดที่คล้ายคลึงกันลดลง ในขณะที่ระยะห่างระหว่างจุดที่ต่างกันจะถูกขยายให้ใหญ่สุด สิ่งนี้ทำได้โดยการฝึกอบรมเครือข่ายโดยใช้จุดข้อมูลสามส่วน ซึ่งประกอบด้วยจุดยึด จุดบวก (คล้ายกับจุดยึด) และจุดลบ (แตกต่างจากจุดยึด) เครือข่ายได้รับการฝึกฝนให้ลดระยะห่างระหว่างจุดยึดและจุดบวกให้น้อยที่สุด ในขณะที่เพิ่มระยะห่างระหว่างจุดยึดและจุดลบให้สูงสุด (Shrestha, 2019)

1.7.10 Embedding เทคนิคในแมชชีนเลิร์นนิงที่แมปข้อมูลมิติสูง เช่น ข้อความหรือรูปภาพในพื้นที่มิติที่ต่ำกว่า ซึ่งมักเรียกว่าพื้นที่ฝังหรือพื้นที่คุณลักษณะ พื้นที่ฝังจะรักษาข้อมูลความคล้ายคลึงกันระหว่างอินพุตต่าง ๆ โดยการจับคู่อินพุตที่คล้ายกันให้อยู่ใกล้กันและอินพุตที่แตกต่างกันซึ่งอยู่ห่างกัน

1.7.11 Loss Function เป็นองค์ประกอบพื้นฐานของการเรียนรู้ของเครื่อง โดย Loss Function เปรียบเสมือนเป้าหมายในการฝึกฝนแบบจำลอง หากแบบจำลองฝึกฝนและสามารถทำนายได้อย่างค่า Loss ก็จะต่ำลงหากทำนายผิดพลาดค่า Loss ก็จะสูงขึ้น นอกจากนี้ค่า Loss Function ยังมีหลายประเภท ขึ้นอยู่กับลักษณะของงานที่จะนำมาใช้

1.7.12 EfficientNetV2S เป็นหนึ่งในโมเดลเครือข่ายประสาทเทียมที่พัฒนาเพื่อใช้ในการจำแนกภาพ โดยมีการปรับปรุงจากรุ่น EfficientNet เดิม เพื่อเพิ่มความเร็วในการฝึกสอนโมเดลและลดขนาดโมเดลให้มีประสิทธิภาพมากขึ้นในการประมวลผลภาพ EfficientNetV2S ใช้วิธีการปรับขนาดแบบสหสัมพันธ์ (Compound Scaling) ที่ปรับขนาดทั้งความลึก ความกว้าง และความละเอียดของภาพพร้อมกัน ทำให้โมเดลนี้มีความสมดุลระหว่างความแม่นยำและประสิทธิภาพการคำนวณ จึงเหมาะสำหรับการใช้งานในอุปกรณ์ที่มีทรัพยากรจำกัดหรือในงานที่ต้องการประสิทธิภาพสูง (Tan & Le, 2021; Liu et al., 2022).

บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยได้ทบทวนแนวคิดทฤษฎีต่าง ๆ ที่เกี่ยวข้องกับการพัฒนาระบบรู้จำใบหน้าสัตว์เลี้ยงโดยใช้การเรียนรู้เชิงลึก ในการศึกษาครั้งนี้ ผู้วิจัยได้ทบทวนแนวคิดทฤษฎีต่าง ๆ ที่เกี่ยวข้องกับการพัฒนาระบบรู้จำใบหน้าสัตว์เลี้ยงโดยใช้การเรียนรู้เชิงลึก เพื่อนำมาใช้เป็นแนวทางในการกำหนดเนื้อหาที่จะสร้างเครื่องมือในงานวิจัย โดยมีความรู้และทฤษฎีที่เกี่ยวข้องดังนี้

2.1 แนวคิดเกี่ยวกับการเรียนรู้ของเครื่อง
2.2 แนวคิดเกี่ยวกับการเรียนรู้เชิงลึก
2.3 แนวคิดและกระบวนการถ่ายโอนความรู้
2.4 แนวคิดและกระบวนการเกี่ยวกับการตรวจจับวัตถุ
2.5 กระบวนการยืนยันใบหน้า
2.6 กระบวนการระบุใบหน้า
2.7 งานวิจัยที่เกี่ยวข้องกับการรู้จำสัตว์เลี้ยง

2.1 แนวคิดเกี่ยวกับการเรียนรู้ของเครื่อง

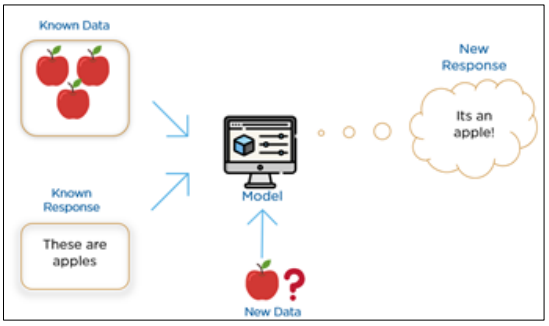
การเรียนรู้ของเครื่อง (Machine Learning) คือระบบที่สามารถเรียนรู้ได้จากข้อมูลด้วยตนเอง ซึ่งสามารถเรียนรู้โดยมีหรือไม่มีผลลัพธ์ก็ได้ เป็นเทคนิคที่ถูกนำมาใช้ในการแก้ปัญหาที่มีความซับซ้อนหรือปัญหาที่อาศัยประสบการณ์ของมนุษย์เพื่อเข้าใจบริบท เช่น การวิเคราะห์ข้อความที่มนุษย์สามารถแยกแยะข้อความได้ว่า เป็นเชิงบวก เชิงลบ หรือเป็นกลาง (สำนักงานส่งเสริมเศรษฐกิจดิจิทัล, 2563) การรู้จักวัตถุที่มนุษย์สามารถมองและรับรู้ได้ว่าสิ่งของตรงหน้าคืออะไร ซึ่งคอมพิวเตอร์สามารถเรียนรู้และเข้าใจสิ่งต่าง ๆ เหมือนมนุษย์ได้ แต่ต้องอาศัยเวลาและข้อมูลจำนวนมากในการเรียนรู้ เพื่อสร้างประสบการณ์ในการแยกแยะข้อมูลเหมือนกับมนุษย์ (Goodfellow et al., 2016).

2.1.1 ประเภทการเรียนรู้ของเครื่อง

ประเภทของการเรียนรู้ของเครื่องสามารถแบ่งออกเป็น 3 ประเภท ซึ่งแต่ละเภทจะมีความเหมาะสมกับงานที่แตกต่างกันคือ

1) การเรียนรู้โดยมีผู้สอน (Supervised Learning)

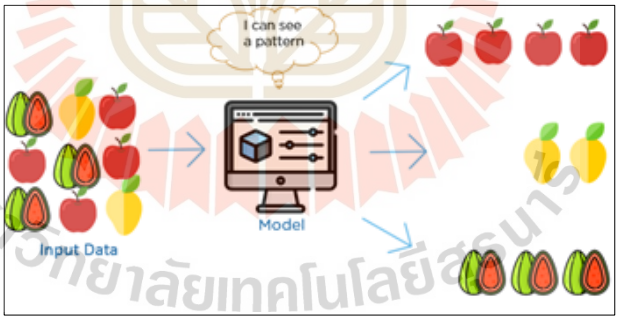
เป็นกระบวนการที่โมเดลจะได้รับการฝึกฝนโดยใช้ข้อมูลที่มีคำตอบหรือเป้าหมายที่ถูกต้อง (labels) ซึ่งเป็นที่มาของคำว่า "การสอน" ในกระบวนการนี้ ข้อมูลนำเข้า (input) และข้อมูลเป้าหมาย (target) จะถูกใช้ร่วมกันเพื่อให้โมเดลสามารถเรียนรู้ความสัมพันธ์ระหว่างสองส่วนนี้ได้ เมื่อโมเดลเรียนรู้จากข้อมูลเหล่านี้แล้ว มันจะสามารถทำนายหรือทำการตัดสินใจในอนาคตได้ ตัวอย่างเช่น การใช้โมเดลในการทำนายราคาบ้านจากข้อมูลต่าง ๆ เช่น พื้นที่บ้าน จำนวนห้องนอน และอายุของบ้าน เป็นต้น โมเดลจะพยายามค้นหาความสัมพันธ์ระหว่างคุณสมบัติเหล่านี้และราคาบ้านเพื่อใช้ในการทำนายราคาบ้านอื่น ๆ ที่ไม่มีข้อมูลเป้าหมาย



รูปที่ 2.1 ภาพอธิบายกระบวนการเรียนรู้แบบมีผู้สอน

2) การเรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning)

เป็นกระบวนการที่โมเดลถูกฝึกสอนโดยไม่มีข้อมูลที่มีป้ายกำกับ (labels) หรือคำตอบที่ถูกต้อง ในการเรียนรู้รูปแบบนี้ โมเดลจะพยายามค้นหาความสัมพันธ์และโครงสร้างภายในข้อมูลเอง โดยไม่ต้องอิงกับเป้าหมายที่ชัดเจน ตัวอย่างของการเรียนรู้แบบไม่มีผู้สอน ได้แก่ การจัดกลุ่มข้อมูล (Clustering) ที่โมเดลพยายามแบ่งข้อมูลออกเป็นกลุ่ม ๆ ที่มีความคล้ายคลึงกัน, การลดมิติของข้อมูล (Dimensionality Reduction) ที่ช่วยลดจำนวนคุณลักษณะในชุดข้อมูลเพื่อให้สามารถวิเคราะห์ได้ง่ายขึ้น, และการค้นหารูปแบบ (Pattern Discovery) ที่ใช้ในการตรวจสอบและค้นหารูปแบบที่ซ่อนอยู่ในข้อมูล กระบวนการเหล่านี้ช่วยให้โมเดลสามารถทำความเข้าใจข้อมูลที่ซับซ้อนและค้นพบข้อมูลใหม่ๆ ได้แม้จะไม่มีคำตอบที่ชัดเจนในการฝึกสอน



รูปที่ 2.2 ภาพอธิบายกระบวนการเรียนรู้แบบไม่มีผู้สอน

3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

เป็นกระบวนการที่โมเดลหรือเอเยนต์ (agent) เรียนรู้จากการลองผิดลองถูก โดยมีแนวคิดหลักคือการเลือกกระทำสิ่งที่ให้ผลตอบแทน (reward) สูงที่สุด ในกระบวนการนี้ เอเยนต์จะทำการทดลองกระทำต่าง ๆ ในสภาพแวดล้อมที่กำหนดไว้ โดยทุกการกระทำจะมีผลลัพธ์ซึ่งอาจเป็นผลดีหรือผลเสีย และเอเยนต์จะเรียนรู้จากผลลัพธ์เหล่านั้นเพื่อปรับปรุงการตัดสินใจในอนาคตผ่านการสะสมประสบการณ์จากสถานการณ์ในอดีตหรือจากระบบจำลอง (simulation) เอเยนต์จะพัฒนานโยบายการตัดสินใจ (policy) ที่ช่วยให้สามารถตัดสินใจได้อย่างเหมาะสมในสถานการณ์ต่าง ๆ ที่อาจเกิดขึ้น เพื่อให้ได้รับผลตอบแทนสูงสุดในระยะยาว

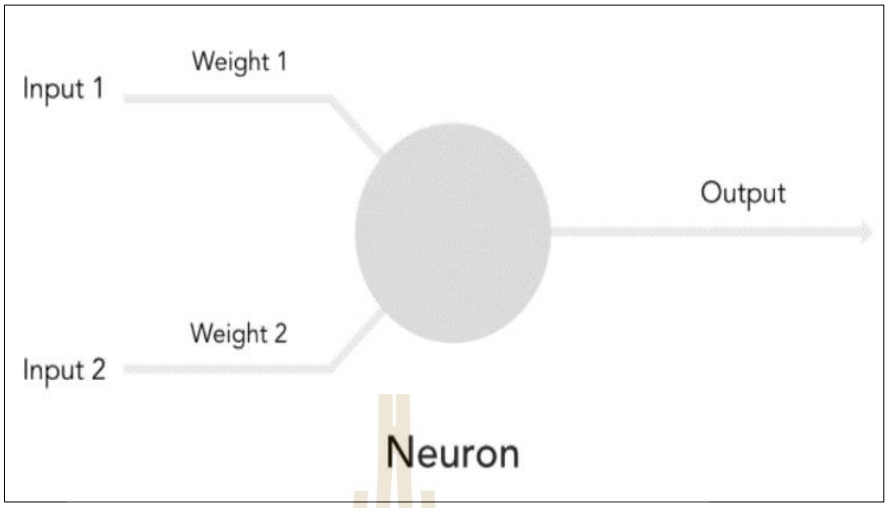
รูปที่ 2.3 ตัวอย่างการเรียนรู้แบบเสริมกำลัง

2.2 แนวคิดเกี่ยวกับการเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep learning) เป็นรูปแบบหนึ่งของการเรียนรู้ของเครื่อง ซึ่งอยู่บนพื้นฐานของระเบียบวิธีที่จะสร้างรูปแบบแนวคิดระดับสูงให้กับข้อมูลโดยใช้การประมวลผลบนโครงสร้างซับซ้อน โดยการเรียนรู้เชิงลึกมีพื้นฐานมาจากโครงข่ายประสาทเทียม (ณัฐโชติ พรหมฤทธิ์, สัจจาภรณ์ ไวจรรยา, 2564)

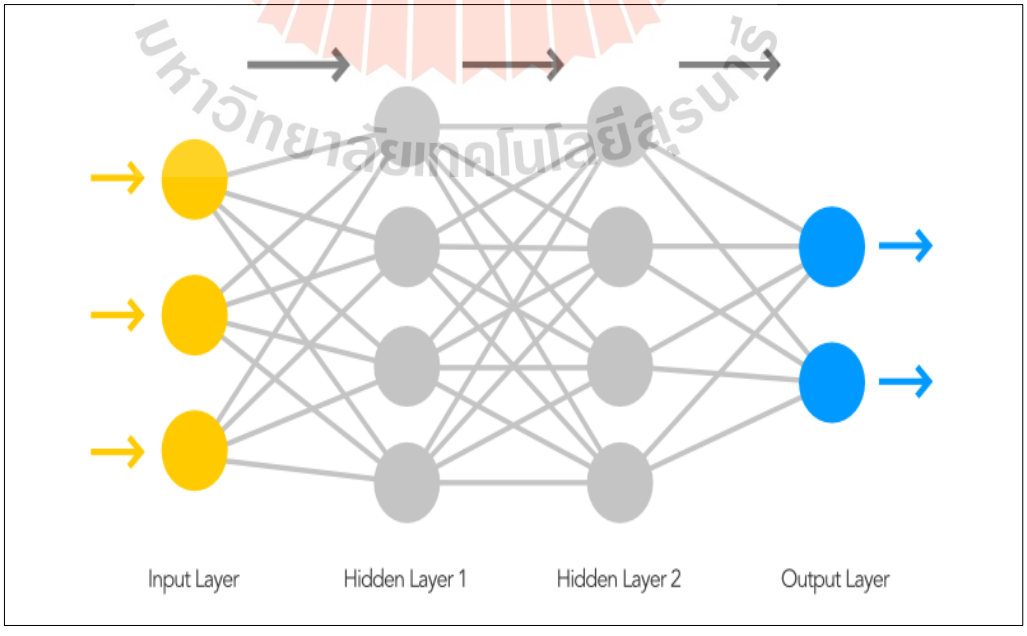
โครงข่ายประสาทเทียม (Neural Network) เป็นการจำลองมาจากสมองของมนุษย์ โดยการทำงานของสมองมนุษย์นั้นจะมีหน่วยประมวลผลขนาดเล็กจำนวนมากและเชื่อมโยงกันด้วยโครงข่ายประสาท ทำให้สามารถเรียนรู้และคิดวิเคราะห์ได้ จึงเกิดการจำลองแนวทางการเรียนรู้ของมนุษย์ไปสู่คอมพิวเตอร์ด้วยโครงข่ายประสาทเทียม โดยมีส่วนประกอบที่เล็กที่สุดคือ นิวรอน (Neuron) ซึ่งทำหน้าที่คำนวณค่านำเข้า (Input) ที่เข้ามา เพื่อให้ได้ผลลัพธ์ออกไป โดยมีส่วนประกอบสำคัญ ดังแสดงในรูปที่ 2.4 ดังนี้

- **ค่านำเข้า** (Input) หรือค่าที่ส่งเข้ามาที่นิวรอน โดยจะมีขาที่เข้ามาได้หลายขา ขึ้นอยู่กับการสร้าง
- **ค่าน้ำหนัก (Weight)** เป็นการให้น้ำหนักของขาแต่ละขาที่ส่งเข้ามา โดยมีค่าระหว่าง 0-1 เมื่อเริ่มต้นจะเป็นการสุ่มค่าขึ้นมา จากนั้นตัวนิวรอนเมื่อเรียนรู้เรื่อย ๆ ก็จะปรับน้ำหนักตัวเองให้ได้คำตอบที่ใกล้เคียงที่สุด
- **ค่าเอนเอียง (Bias)** คือ ค่าที่จะช่วยเข้ามาทำให้ค่าที่เข้ามาอยู่ในระหว่าง 0 - 1 ได้ โดยจะเป็นเลขสุ่มและปรับไปเรื่อย ๆ ทุกครั้งที่เรียนรู้
- **ค่าส่งออก (Output)** คือ ผลลัพธ์
- **การแพร่ย้อนกลับ** (Back Propagation) คือ การที่นิวรอนนำค่าที่ผิดพลาดของผลลัพธ์ที่ได้ กับ ผลลัพธ์ที่เรียนรู้ นำไปปรับน้ำหนักและค่าเอนเอียง ให้เกิดผลลัพธ์ที่ถูกต้องตามที่ได้เรียนรู้มา



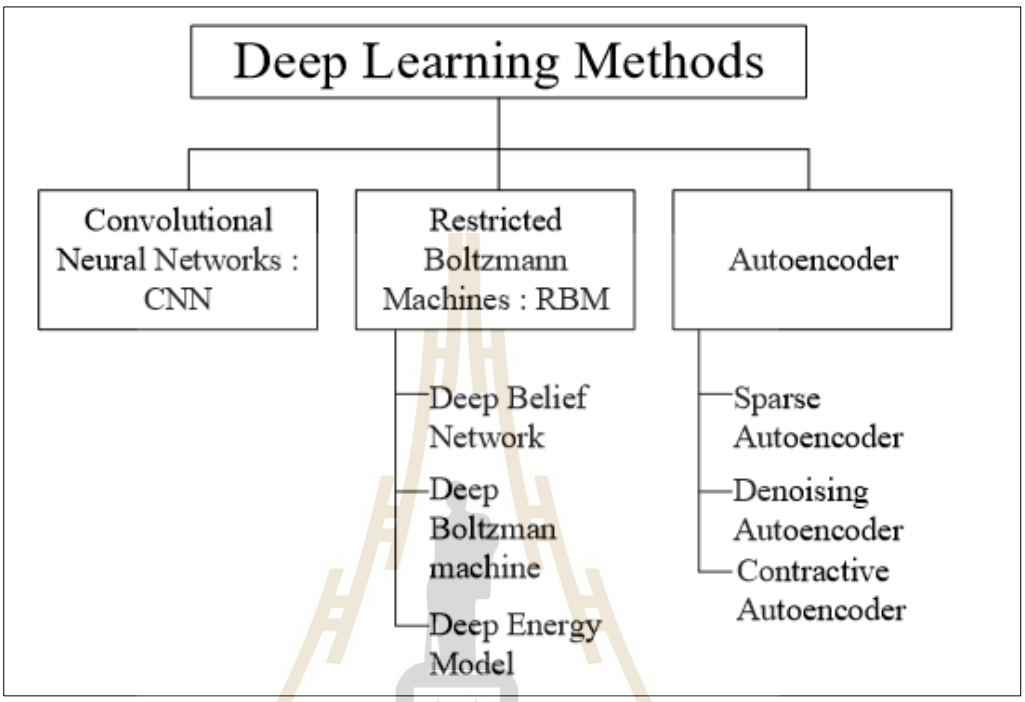
รูปที่ 2.4 ลักษณะการทำงานภายในนิวรอน

การใช้งานโครงข่ายประสาทเทียมนั้นจะใช้ลักษณะของนิวรอนหลายตัวเพื่อประมวลผลร่วมกัน โดยจะมีการส่งค่าต่อเนื่องกัน (Feed-Forward Neural Network) โดยจะแบ่งนิวรอนเป็นกลุ่ม ๆ โดยแต่ละกลุ่มจะเรียกเป็นลำดับชั้นโดยข้อมูลที่เข้ามาจะไหลไปในทิศทางเดียว ไม่ไหลย้อนกลับ จากลำดับชั้นนึงสู่อีก ลำดับชั้นนึงเพื่อให้ได้ผลลัพธ์ แม้ไม่ได้ข้อตกลงหรือการระบุคำจำกัดความที่ชัดเจนเกี่ยวกับจำนวนชั้น แต่ก็ได้มีการกล่าวว่าเมื่อลำดับชั้นที่มีมากกว่า 3 ชั้นขึ้นไป หรือ มีเส้นทางการคำนวณเครดิต (Credit Assignment Paths) มากกว่า 10 เส้น ก็จะเรียกได้ว่าเป็น การเรียนรู้เชิงลึก ดังแสดงในรูปที่ 2.5



รูปที่ 2.5 ส่วนประกอบของนิวรอล ในโครงข่ายประสาทเทียม

รูปแบบของการเรียนรู้เชิงลึก การเรียนรู้เชิงลึกสามารถแบ่งออกได้หลายประเภทดังแสดงในรูปที่ 2.6

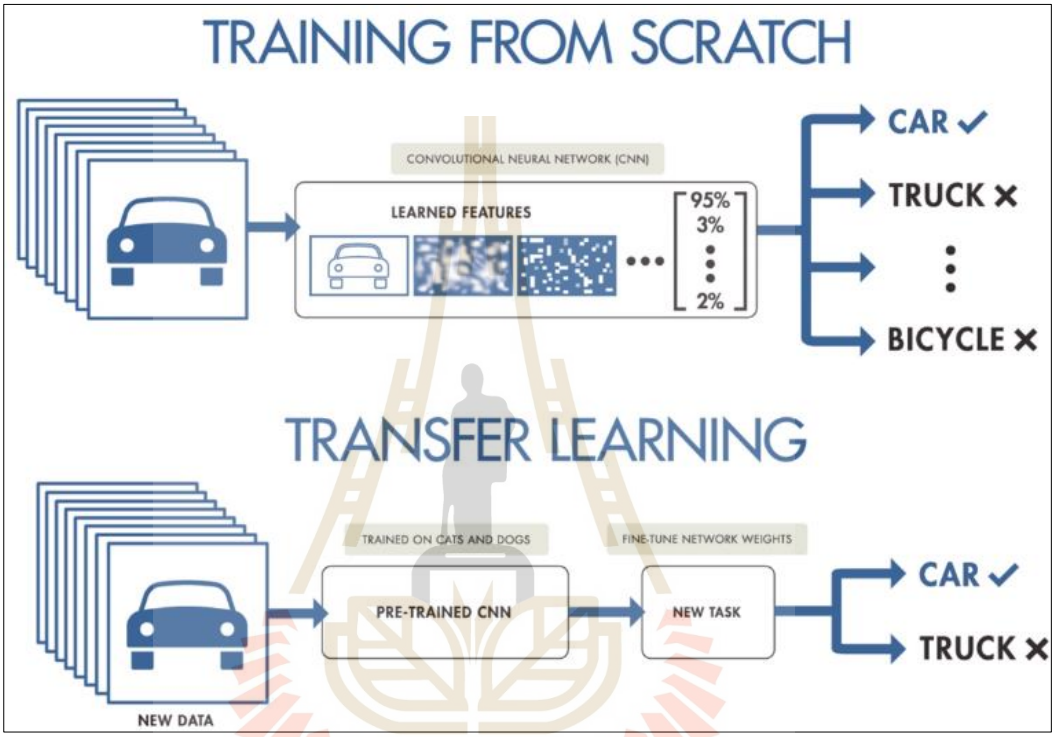


รูปที่ 2.6 ประเภทของการเรียนรู้เชิงลึก

ประเภทของการเรียนรู้เชิงลึกแต่ละวิธีจะมีความเหมาะสมกับการดำเนินงานที่แตกต่างกัน เช่น โครงข่ายประสาทเทียม (CNN) เหมาะกับงานวิจัยการรู้จำวัตถุ โครงข่ายความเชื่อแบบลึก(DBN) เหมาะกับงานประเภทการรู้จำตัวอักษร เครื่องจักรโบลทซ์มันน์เชิงลึก (DBM) และ การเข้ารหัสอัตโนมัติ (auto-encoders) เหมาะกับงานวิจัยประเภทการจำแนกข้อมูลตัวอักษร การใช้งานนั้นขึ้นอยู่กับความเหมาะสมและจุดมุ่งหมายในการดำเนินงาน โดยในงานวิจัยนี้เลือกใช้การเรียนรู้เชิงลึกแบบซีเอ็นเอ็น (CNN: Convolutional Neural Networks) เนื่องจากต้องการใช้รู้จำวัตถุสำหรับเรียนรู้การตรวจจับใบหน้าแมวซึ่งมีลักษณะเดียวกันกับวัตถุ

2.3 แนวคิดและกระบวนการถ่ายโอนความรู้

การถ่ายโอนความรู้เป็นเทคนิคการเรียนรู้ของเครื่องที่จะใช้ น้ำหนัก (Weight) ของแบบจำลองที่ผ่านการฝึกฝนบนชุดข้อมูลที่มีขนาดใหญ่มาแล้ว มาปรับแต่งเพื่อใช้กับชุดข้อมูลที่มีขนาดเล็ก ทำให้สามารถสร้างแบบจำลองที่มีประสิทธิภาพสูง โดยไม่ต้องสิ้นเปลืองเวลาและทรัพยากรในการฝึกฝนแบบจำลองใหม่ตั้งแต่เริ่มแรก (Andrew Ng, 2016) ดังแสดงในรูปที่ 2.7



รูปที่ 2.7 ภาพตัวอย่างการนำ Transfer Learning มาใช้กับงานใหม่

ทั้งนี้งานวิจัยเกี่ยวกับตรวจจับใบหน้าแมวและการรู้จำใบหน้าแมว สามารถใช้การถ่ายโอนความรู้จากแบบจำลองที่เรียนรู้เกี่ยวกับภาพมาแล้วได้เช่น ImageNet (Deng et al., 2016) ซึ่งเป็นแบบจำลองที่ถูกออกแบบมาใช้ในการตรวจจับวัตถุ ซึ่งฝึกฝนมาจากชุดรูปภาพขนาดใหญ่กว่า 14 ล้านภาพ ทำให้ง่ายต่อการนำมาประยุกต์ใช้กับแบบจำลองที่ใช้งานเกี่ยวกับภาพ เนื่องจากแบบจำลองมีการเรียนรู้เกี่ยวกับคุณลักษณะของภาพอยู่แล้ว

2.4 แนวคิดและกระบวนการเกี่ยวกับการตรวจจับวัตถุ

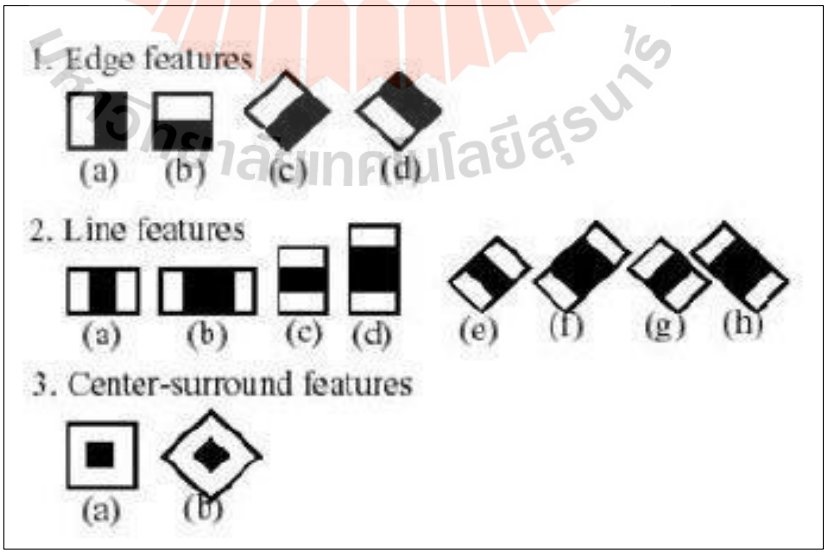
กระบวนการตรวจจับวัตถุ เป็นการระบุวัตถุที่สนใจภายในรูปภาพหรือวิดีโอ โดยจะวาดกรอบรอบ ๆ วัตถุในภาพ เช่น รถยนต์ อาคาร หรือสัตว์ และจัดประเภทเป็นหมวดหมู่ที่กำหนดไว้ล่วงหน้า การตรวจจับวัตถุสามารถนำมาประยุกต์ใช้ได้หลากหลาย และงานหนึ่งที่สำคัญก็คือการตรวจจับใบหน้า ซึ่งมีจะถูกนำมาใช้เพื่อระบุและแยกใบหน้าออกจากรูปภาพ การตรวจจับใบหน้าเป็นขั้นตอนสำคัญในการจดจำใบหน้า โดยมีหน้าที่ในการระบุตัวบุคคลตามลักษณะใบหน้า การตรวจจับใบหน้าสามารถทำได้โดยใช้ขึ้นตอนวิธีการตรวจจับวัตถุต่าง ๆ โดยจะมีทั้งแบบดั้งเดิม (Traditional) ที่ใช้การระบุคุณลักษณะไว้ล่วงหน้า (Feature-Based) และแบบใช้การเรียนรู้เชิงลึก (deep learning-based)

2.4.1 กระบวนตรวจจับวัตถุแบบดั้งเดิม

เป็นกระบวนตรวจจับวัตถุโดยใช้คุณลักษณะที่สร้างขึ้นด้วยมือ หมายถึงกระบวนการของการออกแบบคุณลักษณะด้วยตนเอง เช่น ขอบ มุม หรือฮิสโตแกรมของค่าสี เพื่อระบุวัตถุในรูปภาพ คุณลักษณะจะถูกป้อนเข้าสู่อัลกอริทึมการตรวจจับวัตถุ เพื่อระบุตำแหน่งและจำแนกวัตถุ

1) วิธีการแบบฮาร์ (Haar-Like Feature)

วิธีการแบบฮาร์ (Viola and Jones, 2001) เป็นวิธีการตรวจจับและตีความจากวัตถุภายในภาพด้วยหลักการของคลื่นขนาดเล็กแบบฮาร์ (Haar Wavelet) สำหรับการสร้างรูปสี่เหลี่ยม เพื่อนำมาใช้เป็นตัวกำหนดคุณลักษณะพิเศษของแต่ละคุณลักษณะ (Feature) รูปที่ 2.9 แสดงถึงผลต่างระหว่างพื้นที่ส่วนที่เป็นสีขาวและพื้นที่ส่วนที่เป็นสีดำ ซึ่งแต่ละคุณลักษณะสามารถเปลี่ยนแปลงขนาดและตำแหน่งได้ คุณลักษณะที่ถูกกำหนดมานั้น ถูกนำไปใช้สำหรับการตรวจจับหาลักษณะบนภาพแบบต่าง ๆ เช่น เพื่อใช้ในการค้นหาบริเวณที่เป็นส่วนของใบหน้ามนุษย์ (Viola and Jones, 2001) ดังแสดงตัวอย่างในรูปที่ 2.8 และ 2.9



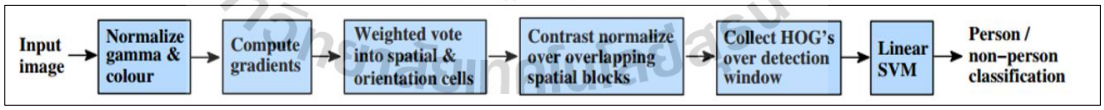
รูปที่ 2.8 ขั้นตอนการทำฮาร์ (Harr-Like)

รูปที่ 2.9 ตัวอย่างการทำฮาร์ (Haar-Like)

หลักการพื้นฐานของการตรวจจับใบหน้าของไวโอลา (Viola) และโจนส์ (Jones) คือ การนำภาพที่ ต้องการตรวจหามาแบ่งเป็นภาพย่อย ๆ จากภาพย่อยดังกล่าวจะถูกพิจารณาเป็นภาพรับเข้า (Input Image) ของกระบวนการตรวจหาใบหน้า ซึ่งเทคนิคโดยทั่ว ๆ ไป ที่พบเห็นในการตรวจหาบริเวณที่ เป็นใบหน้า ทำได้โดยปรับขนาดของภาพที่รับเข้าจาก ภาพที่มีความแตกต่างกันหลาย ๆ แบบและจากนั้นจึงใช้ตัวตรวจจับ (Detector) ที่มีขนาดคงที่มาใช้ในการค้นหาวัตถุ ซึ่งวิธีการตรวจจับแบบนี้มักมีข้อเสียคือระยะเวลาในการคำนวณนั้นไม่คงที่

2) วิธีการแบบฮอก (HOG: Histogram of Oriented Gradient)

เป็นวิธีการสร้างภาพรวมของการหาค่าความถี่ของทิศทางตามค่าเกรเดียนท์ (HOG) มีทั้งหมดหกขั้นตอน ดังแสดงในรูปที่ 2.10

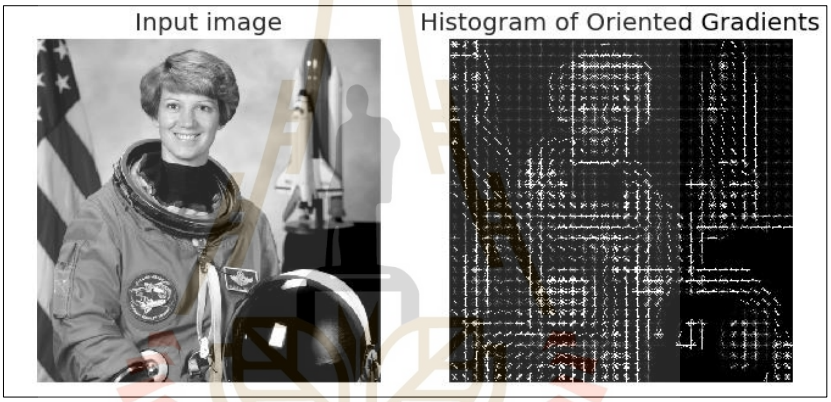


รูปที่ 2.10 ขั้นตอนการหาคุณลักษณะเฉพาะในการตรวจจับวัตถุด้วยฮอก (HOG)

จากภาพที่ 2.11 ดาลัลและทริก (Dalal and Triggs, 2005) กล่าวว่า หน้าต่างในการตรวจจับจะเรียง ต่อกันเป็นแผ่นที่ทับซ้อนกัน ซึ่งใช้ค่าความถี่ตามทิศทางแกรเดียนท์มาเป็นคุณสมบัติในการจำแนกคุณลักษณะเฉพาะ การรวมตัวกันของเวกเตอร์จะใช้ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) จำแนกประเภทที่เป็นวัตถุและไม่ใช่วัตถุ หน้าต่างในการตรวจจับจะถูกกราดทุกตำแหน่งของภาพ วิธีการดังกล่าวมุ่งเน้นไปที่ขั้นตอนในการสกัดคุณลักษณะเฉพาะของภาพ (Dalal, N., & Triggs, B., 2005) โดยเมื่อรับภาพเข้ามาในระบบ ขั้นตอนแรก จะนอร์มอไลซ์ค่าความสว่าง (Gamma) และสี (Colour) ขั้นตอนที่สอง จะคำนวณหาค่าเกรเดียนท์ ขั้นตอนที่สาม จะโหวตหาค่าถ่วงน้ำหนัก

ของระยะห่างที่สอดคล้องกันของเซลล์ (Cells) ขั้นตอนที่สี่ จะนอร์มอไลซ์พื้นที่ที่ซับซ้อนกันของบล๊อค (Block) ขั้นตอนที่ห้า จะรวบรวมหน้าต่างของฮอก (HOG) สำหรับการตรวจจับทั้งหมด ขั้นตอนที่หก จะจำแนกคุณลักษณะเฉพาะด้วยซัพพอร์ตเวกเตอร์แมชชีนแบบ เชิงเส้น (Linear)

ค่าความถี่ของทิศทางตามค่าเกรเดียนท์ เป็นวิธีการดึงคุณลักษณะเฉพาะของวัตถุ ซึ่งสามารถดึงรูปร่างภายในภาพโดยใช้การกระจายตัวของความเข้มเกรเดียนท์หรือทิศทางของเส้นขอบ การดึงคุณลักษณะของฮอก(HOG) จะกระทำโดยการแบ่งภาพออกเป็นส่วนย่อยขนาดเล็กหรือ เรียกว่า เซลล์ (Cells) โดยแต่ละเซลล์จะรวบรวมฮีสโตรแกรมของทิศทางเกรเดียนท์หรือทิศทางของขอบ ภายในเซลล์ที่มีขนาดหนึ่งมิติ (1-D) โดยที่จะมีการรวมฮีสโตรแกรมนั้นเข้าด้วยกัน เพื่อแสดงถึงคุณลักษณะเฉพาะของวัตถุที่สนใจ เพื่อให้มีประสิทธิภาพของความถูกต้องเพิ่มมากขึ้น สามารถนำฮีสโตรแกรมมานอร์มอลไลซ์ด้วยการคำนวณตัวชี้วัดของค่าความเข้มทั่วทั้งพื้นที่ขนาดใหญ่ของภาพหรือ เรียกอีกอย่างว่าบล๊อค (Block) (ชุมพล เสนาพันธ์ และ คำรณ สุนัต, 2556)



รูปที่ 2.11 ตัวอย่างการทำงานของฮอก (HOG: Histrograms of Oriented Gradient)

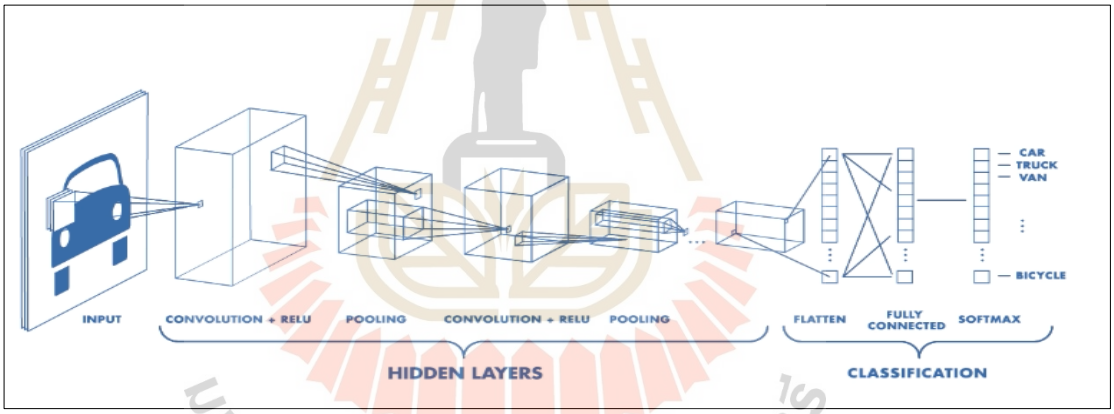
แม้ว่าการระบุคุณลักษณะแบบดั้งเดิมจะเข้าใจได้ง่ายและใช้ทรัพยากรในการประมวลผลน้อย แต่ก็ต้องใช้เวลาและบุคคลที่ความเข้าใจในตัวข้อมูล และยังสามารถเกิดอคติต่อการระบุคุณลักษณะข้อมูลได้ ทำให้การใช้การเรียนรู้เชิงลึกเป็นที่นิยมมากขึ้น แม้จะต้องใช้ทรัพยากรในการประมวลผลสูงแต่ก็สามารถเรียนรู้คุณลักษณะได้เอง และสามารถจัดการกับข้อมูลที่ซับซ้อนและมีขนาดใหญ่ได้

2.4.2 กระบวนการตรวจจับวัตถุโดยใช้การเรียนรู้เชิงลึก

การตรวจจับวัตถุการเรียนรู้เชิงลึกเป็นเทคนิคเกี่ยวกับการมองเห็นของคอมพิวเตอร์ที่ใช้โครงข่ายประสาทเทียมเชิงลึกเพื่อตรวจจับและจำแนกวัตถุภายในภาพหรือวีดิโอ โดยทั่วไปจะเรียนรู้ที่จะจดจำรูปแบบในข้อมูลภาพและสร้างการคาดการณ์สำหรับตำแหน่งของวัตถุและชื่อคลาส วิธีการนี้มีประสิทธิภาพดีกว่าวิธีการตรวจจับวัตถุแบบดั้งเดิมที่ต้องสร้างคุณลักษณะขึ้นมาด้วยตัวเอง และยังถูกนำมาใช้อย่างแพร่หลาย เช่น รถยนต์ที่ขับเคลื่อนด้วยตัวเอง ระบบตรวจจับวัตถุ

2.4.2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

โครงข่ายประสาทเทียมแบบคอนโวลูชัน หรือ ซีเอ็นเอ็น (CNN: Convolutional Neural Network) เป็นสถาปัตยกรรมโครงข่ายประสาทเทียมประเภทการเรียนรู้เชิงลึกที่ใช้สำหรับการวิเคราะห์ภาพและวิดีโอ โดยจะใช้การดำเนินการทางคณิตศาสตร์ที่เรียกว่า convolutions กับข้อมูลนำเข้า เพื่อสร้างผังคุณลักษณะที่สรุปข้อมูลสำคัญในอินพุต โดยมีจุดเด่นคือสามารถดึงลักษณะและจำแนกข้อมูลได้โดยตรง จึงทำให้เหมาะสำหรับการรู้จำภาพ โดยซีเอ็นเอ็นเป็นโครงข่ายใยประสาทแบบหลายชั้น (Multilayer perceptron) โดยเลอคัน (LeCun) นำเสนอโครงสร้างซีเอ็นเอ็นแบบวิธีการแพร่ค่าความผิดพลาดย้อนกลับ (Back-propagation) สำหรับการรู้จำตัวเลขด้วยลายมือ ทำให้ซีเอ็นเอ็นรูปแบบนี้เป็นที่นิยมใช้จนถึงปัจจุบัน เนื่องจากโครงข่ายของระบบนี้ไม่ได้ถูกสอนแบบชั้นต่อชั้น แต่จะใช้การปรับค่าน้ำหนักเพื่อลดฟังก์ชันความคลาดเคลื่อน ด้วยวิธีการแพร่ค่าความผิดพลาดย้อนกลับ นอกจากนี้ลักษณะเด่นของข้อมูลจะถูกดึงออกมาโดยอัตโนมัติ ด้วยการเรียนรู้หน้ากากคอนโวลูชันที่เกิดจากค่าน้ำหนักของแผนภาพคุณลักษณะเด่น (Feature maps) โดยภาพโครงสร้างโดยทั่วไปของซีเอ็นเอ็นแสดงดังรูปที่ 2.12



รูปที่ 2.12 ขั้นตอนของซีเอ็นเอ็นเพื่อทำการคัดแยกข้อมูล

โครงสร้างโดยทั่วไปของซีเอ็นเอ็น ประกอบด้วย ข้อมูลอินพุต (INPUT) ชั้นคอนโวลูชัน(CONV) ชั้นพูลลิ่ง (POOL) ชั้นลักษณะเด่น (FC) และชั้นเอาต์พุต (OUTPUT) ซึ่งแต่ละส่วนมี การทำงานที่แตกต่างกันดังซึ่งสามารถอธิบายได้ดังต่อไปนี้

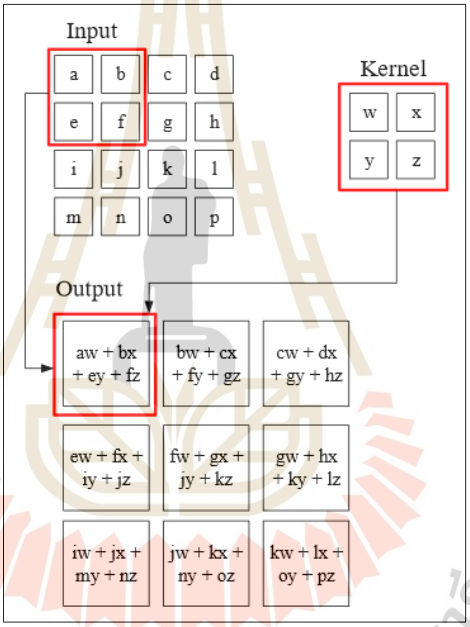
1) **ชั้นอินพุต (Input layer)**
ชั้นนี้เป็นการนำเข้าภาพเข้าสู่ระบบ

2) **ชั้นคอนโวลูชัน (Convolutional layer)**
ในขั้นตอนนี้คือนำภาพอินพุตมาคอนโวลูชันกับหน้ากาก โดยที่ค่าสัมประสิทธิ์ในหน้ากากนั้นได้ จากการสุ่มขึ้นมา โดยฟังก์ชันที่นิยมใช้ในการชั้นคอนโวลูชันจะมีทั้งหมด 3 ฟังก์ชันได้แก่

- Sigmoid function
- Tan hyperbolic function (tanh function)
- Rectified Linear Unit (ReLU function)

จากฟังก์ชันกระตุ้นทั้ง 3 แบบที่ได้กล่าวไปนั้น ReLU function ถูกนำมาใช้งานกับซีเอ็นเอ็นมากกว่าฟังก์ชันอื่น เนื่องจากลดโอกาสที่บางข้อมูลจะสูญหาย โดยที่หากข้อมูลมีค่ามากกว่า 0 จะ เลือกมาเป็นเอาท์พุต แต่สำหรับกรณีข้อมูลมีค่าน้อยกว่า 0 จะให้ค่าเอาท์พุตเป็น 0

ตัวอย่างการทำคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากขนาด 2×2 แสดงดังรูปที่ 2.13



รูปที่ 2.13 ตัวอย่างการทำคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากขนาด 2×2

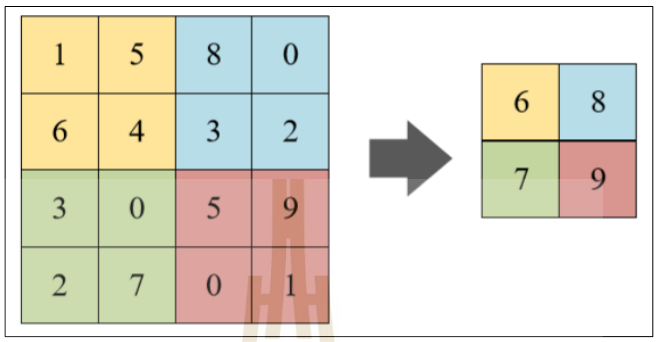
ตัวอย่างการคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากขนาด 2×2 ทำให้ได้เอาท์พุตขนาด 3×3 พิกเซล ซึ่งเอาท์พุตเกิดจากการนำค่าสัมประสิทธิ์ของหน้ากากไปคูณ กับค่าของอินพุตจากนั้นนำทุกค่าที่ได้บวกกันเพื่อเป็นเอาท์พุต จากตัวอย่างจะเห็นว่าหน้ากากมี สัมประสิทธิ์ w x y และ z ไปคอนโวลูชันกับอินพุต a b e และ f ได้เอาท์พุตเป็น aw+bx+ey+fz จากนั้นเลื่อนหน้ากากไปยังข้อมูลอินพุตถัดไป และทำกระบวนการเดิมจนครบขนาดของอินพุต โดย เอาท์พุตจะมีขนาดลดลงเนื่องจากกระบวนการนี้จะสูญเสียขอบของภาพไป

3) ชั้นพูลลิ่ง (Pooling layer)

ในขั้นตอนนี้เป็นการลดขนาดของข้อมูลลงโดยวิธีการทำพูลลิ่งที่นิยมนั้นมี 3 แบบด้วยกัน ดังต่อไปนี้

(1) พูลลิ่งด้วยค่าสูงสุด (Max pooling)

การทำพูลลิ่งด้วยค่าสูงสุดนั้นจะเลือกค่าสูงสุดในแต่ละบล็อคที่หน้ากากพูลลิ่งครอบอยู่บนข้อมูลอินพุตมาเป็นคำตอบ ตัวอย่างการทำงานของการทำพูลลิ่งด้วยค่าสูงสุดแสดงได้ดังรูปที่ 2.14

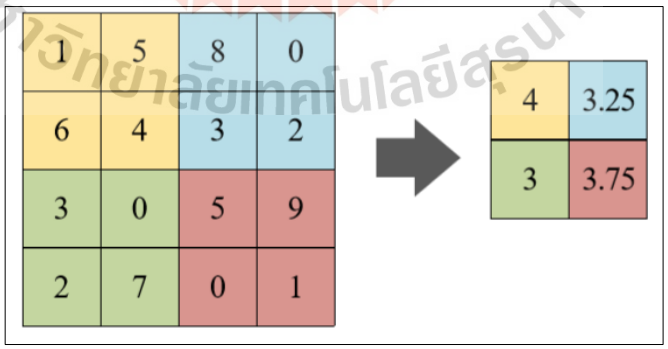


รูปที่ 2.14 ตัวอย่างการทำพูลลิ่งด้วยค่าสูงสุด

รูปที่ 2.14 การทำงานของพูลลิ่งด้วยค่าสูงสุดระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากพูลลิ่งขนาด 2×2 จะมีขั้นตอนการทำงานคือ การนำหน้ากากขนาด 2×2 ไปครอบที่อินพุต จากนั้นจะเลือก ค่าสูงสุดออกมา ซึ่งจะได้เอาท์พุตเป็น 6 8 7 และ 9 ตามลำดับ โดยขนาดของเอาท์พุตจะมีขนาดลดลง ครึ่งหนึ่งของขนาดอินพุต

(2) การพูลลิ่งด้วยค่าเฉลี่ย (Average pooling)

พูลลิ่งด้วยค่าเฉลี่ยจะคล้ายกับการทำพูลลิ่งด้วยค่าสูงสุด มีส่วนที่แตกต่างกันคือจะคำนวณ ค่าเฉลี่ยในแต่ละบล็อคแทนที่การเลือกค่าสูงสุด ตัวอย่างการทำงานแสดงได้ดังรูปที่ 2.15



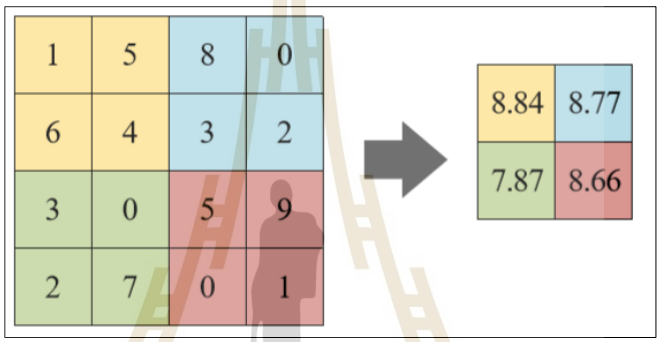
รูปที่ 2.15 ตัวอย่างการทำพูลลิ่งด้วยค่าเฉลี่ย

มหาวิทยาลัยเทคโนโลยีสุรนารี

รูปที่ 2.15 การทำงานของพูลลิ่งด้วยค่าเฉลี่ยระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากพูลลิ่งขนาด 2×2 ขั้นตอนการทำงานคือ นำหน้ากากขนาด 2×2 ไปครอบที่อินพุต จากนั้นจะเลือกหาค่าเฉลี่ยในแต่ละกรอบที่ครอบออกมา ซึ่งจะได้เอาท์พุตเป็น 4 3.25 3 และ 2.75 ตามลำดับ ซึ่งจะเห็น ว่าค่าของเอาท์พุตที่ได้จะมีค่าที่ใกล้เคียงกัน โดยขนาดของเอาท์พุตจะมีขนาดลดลงครึ่งหนึ่งของขนาดอินพุต

(3) การพูลลิ่งด้วย $L_2 - norm$

วิธีการนี้จะเป็นการหาค่า $L_2 - norm$ ของแต่ละบล็อกมาเป็นเอาท์พุต ตัวอย่างการทำงาน แสดงได้ดังรูปที่ 2.16



รูปที่ 2.16 ตัวอย่างการทำพูลลิ่งด้วย $L_2 - norm$

จากวิธีการพูลลิ่งทั้ง 3 วิธีกับอินพุตรูปแบบเดียวกันก็จะให้ค่าเอาท์พุตที่แตกต่างกัน โดยการทำพูลลิ่งด้วยค่าสูงสุดจะให้ข้อมูลในชั้นเอาท์พุตมีความแตกต่างกัน แต่สำหรับการทำพูลลิ่งด้วยค่าเฉลี่ย และพูลลิ่งด้วย $L_2 - norm$ จะให้ค่าของข้อมูลเอาท์พุตที่ใกล้เคียงกัน ดังนั้นวิธีการที่นิยมใช้กันคือ การทำพูลลิ่งด้วยค่าสูงสุด เนื่องจากมีขั้นตอนการทำงานที่ง่ายกว่า และได้ค่าที่มีความแตกต่างกัน มากกว่าวิธีการอื่น ซึ่งจะส่งผลถึงลักษณะของข้อมูลที่ได้ด้วย

4) ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully Connected layer)

เมื่อผ่านกระบวนการทำงานของชั้นคอนโวลูชัน ชั้นพูลลิ่งมาแล้วต่อมาจะนำเอาท์พุตจาก ชั้นเหล่านี้เข้าสู่ชั้นลักษณะเด่น เพื่อเตรียมตัวเข้าสู่ชั้นจำแนกต่อไป โดยชั้นนี้จะเชื่อมโยงนำเอาเซลล์ประสาทจากชั้นก่อนหน้า (ชั้นคอนโวลูชัน และชั้นพูลลิ่ง) มาเชื่อมโยงกันกับเซลล์ประสาททุกเซลล์ที่มี โดยจะเชื่อมต่อกันเป็น 1 มิติ ดังนั้นจะไม่สามารถมีชั้นคอนโวลูชันหลังจากชั้นนี้ได้อีก โดยชั้นนี้จะทำเชื่อมต่อไปยังชั้นเอาท์พุต

5) ชั้นเอาท์พุต (Output layer)

เป็นชั้นของการจำแนกข้อมูล โดยจำนวนเซลล์ประสาทในชั้นนี้จะเท่ากับจำนวนคน ที่ใช้ในการรู้จำภาพใบหน้าบุคคล

ซึ่งซีเอ็นเอ็นยังมีปัญหาในการตรวจจับวัตถุ เนื่องจากระยะเวลาในการประมวลผลที่ใช้เวลาค่อนข้างนานในการประมวลผลทั้งภาพเพื่อค้นหาวัตถุ และเมื่อสิ่งที่ต้องการหาอาจไม่มีปรากฏในภาพ หรืออาจมีวัตถุเดียวปรากฏขึ้นมาพร้อมกันทีเดียวหลายอัน โดยแต่ละอันอยู่ต่างมุมภายในภาพ และอาจมีขนาดเล็กใหญ่แตกต่างกันไป ซึ่งอาจทำให้ไม่พบวัตถุที่ต้องหาการหาแม้จะมีอยู่ในภาพ ทำให้เกิดการพัฒนาต่อยอดเพื่อแก้ปัญหาจนเกิดเป็นสถาปัตยกรรมต่างๆ โดยสามารถแบ่งได้เป็นทั้งหมด 2 ประเภท

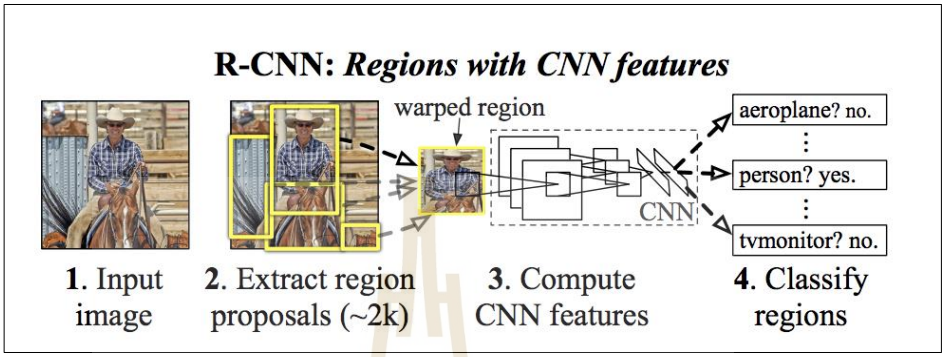
1) One-Stage Object Detection ถูกออกแบบมาเพื่อให้สามารถตรวจจับวัตถุในการส่งต่อข้อมูลเพียงครั้งเดียวในโครงข่ายประสาทเทียม แบบจำลองแบบขั้นตอนเดียวมีประสิทธิภาพสูงและเหมาะสำหรับการใช้งานแบบเรียลไทม์ที่มีความเร็วเป็นปัจจัยสำคัญ แต่แบบจำลองแบบขั้นตอนเดียวมักมีความแม่นยำน้อยกว่าแบบจำลองแบบสองขั้นตอนและอาจมีปัญหาในการตรวจจับวัตถุขนาดเล็กหรือถูกบดบัง โดยสถาปัตยกรรมแบบจำลองตัวอย่างเช่น ตระกูล YOLO หรือ SSD เป็นต้น

2) Two-Stage Object Detection ถูกออกแบบมาเพื่อให้สามารถตรวจจับวัตถุด้วยสองขั้นตอน โดยขั้นตอนแรกจะใช้ RPN (Region Proposal Network) เพื่อสร้างชุด Region Proposal เพื่อหาบริเวณที่คาดว่าจะมีวัตถุ และขั้นตอนที่สองจะใช้ classifier เพื่อจำแนกว่าแต่ละบริเวณที่ถูกนำเสนอโดย Region Proposal มีวัตถุอยู่หรือไม่ โดยแบบจำลองสองขั้นตอนมีความแม่นยำสูง แต่ใช้ทรัพยากรมากกว่า และใช้เวลามากกว่าแบบจำลองขั้นตอนเดียวในการคำนวณ โดยสถาปัตยกรรมแบบจำลองตัวอย่างเช่น ตระกูล R-CNN

2.4.2.2 อาร์-ซีเอ็นเอ็น (R-CNN: Region with CNN)

อาร์-ซีเอ็นเอ็น เป็นวิธีการตรวจจับวัตถุโดยพัฒนาต่อเนื่องมาจากโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยเกอร์ชิคและคณะ (Girshick, et. al.) ได้นำเสนอวิธีการเพิ่มความเร็วและความแม่นยำในการตรวจจับวัตถุของการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยใช้รีเจียน พร็อพโพซอล (Region Proposal) คาดเดาบริเวณที่มีวัตถุจากนั้นจึงนำเข้าข้อมูลภาพไปประมวลผลด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชันเพื่อวิเคราะห์ข้อมูลและได้ผลลัพธ์ออกมา โดยรีเจียน พร็อพโพซอล เป็นวิธีการจำแนกพื้นที่เพื่อคาดเดาบริเวณที่มีวัตถุโดยรีเจียน พร็อพโพซอล ทำหน้าที่เสนออาร์โอไอ (ROI: region of interest) เป็นกรอบสี่เหลี่ยมมุมฉากหลากไซส์ หลายตำแหน่งขึ้นมา จำนวน 2000 อันต่อ 1 ภาพอินพุต และภาพอินพุตต้นฉบับจะถูกแบ่งด้วยอาร์โอไอ ก่อนจะถูกส่งต่อไปให้ซีเอ็นเอ็นค้นหาคุณลักษณะของภาพ ซึ่งจะได้ฟีเจอร์เวคเตอร์ (feature vector) ขนาด 4096 ช่อง แล้วจึงใช้เอสวีเอ็ม (SVM) เพื่อแยกประเภท (classification) และวิเคราะห์ส่วนที่ตัด (crop) ออกมาว่ามีวัตถุที่ต้องการอยู่ในประเภท (class) ใดที่เราให้ความสนใจหรือไม่ ซึ่งการทำอาร์โอไอทำให้ไม่ว่าภาพภาพหนึ่งจะมีวัตถุที่ต้องการค้นหาอยู่กี่คลาส หรือแต่ละวัตถุจะอยู่ในมุมไหนของภาพ หรือมีขนาดใหญ่เล็กแค่ไหนก็จะสามารถสร้างอาร์โอไอครอบและตัด (crop) เฉพาะส่วนของวัตถุนั้น ๆ เพื่อส่งภาพที่ตัด (crop) ไปให้ซีเอ็นเอ็น (CNN) และเอสวีเอ็ม (SVM) ช่วยกันทำการแยกรูป (image classifier) ต่อได้

ข้อเสียของอาร์-ซีเอ็นเอ็น คือ การทำอาร์โอไอมีจำนวนเยอะมาก เนื่องจากใน 1 ภาพมี 2000 อาร์โอไอทำให้ต้องใช้เวลาในการประมวลผลมาก และการแยกคุณลักษณะด้วยวิธีซีเอ็นเอ็น (CNN feature extraction) ของแต่ละอาร์โอไอก็ทำงานซ้ำซ้อนเนื่องจากภาพที่ตัดมีบริเวณพื้นที่ทับซ้อนกัน ตัวอย่างขั้นตอนการทำอาร์-ซีเอ็นเอ็นแสดงดังรูปที่ 2.17



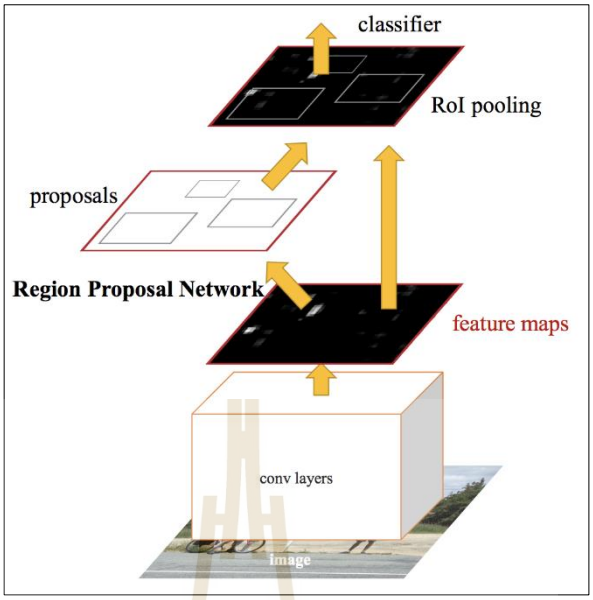
รูปที่ 2.17 ตัวอย่างขั้นตอนการทำอาร์-ซีเอ็นเอ็น

2.4.2.3 ฟาส์ตอาร์-ซีเอ็นเอ็น (Fast R-CNN: Fast Region-based Convolutional Network)

ฟาส์ตอาร์-ซีเอ็นเอ็นเป็นวิธีการตรวจจับวัตถุที่ถูกพัฒนาต่อยอดมาจากอาร์-ซีเอ็นเอ็น โดย เกอร์ชิคและคณะ (Girshick, et. al.) ได้นำเสนอขึ้นมาเพื่อแก้ปัญหาของอาร์-ซีเอ็นเอ็น โดยเพิ่มอาร์โอไอพูลลิ่ง เพื่อทำการสกัดคุณลักษณะ (feature extraction) ของภาพด้วยซีเอ็นเอ็นกับอินพุตทั้งภาพเพียงรอบเดียว จากนั้นจึงนำอาร์โอไอบริเวณที่ต้องการนำไปสกัดคุณลักษณะเฉพาะในบริเวณนั้น ทำให้ลดการทำซีเอ็นเอ็น 2000 ครั้งต่อภาพ เหลือทำซีเอ็นเอ็นเพียงครั้งเดียวและนำคุณลักษณะในแต่ละอาร์โอไอ ไปใช้เพื่อประมวลผลกับวิธีการแยกประเภท (classifier) ต่อไป เพื่อแยกแยะ ซึ่งการทำฟาสต์ อาร์-ซีเอ็นเอ็น ทำให้สามารถเพิ่มความให้กับการค้นหาวัตถุเพิ่มขึ้นถึง 25 เท่า

2.4.2.4 ฟาสต์เตอร์อาร์-ซีเอ็นเอ็น (Faster RCNN)

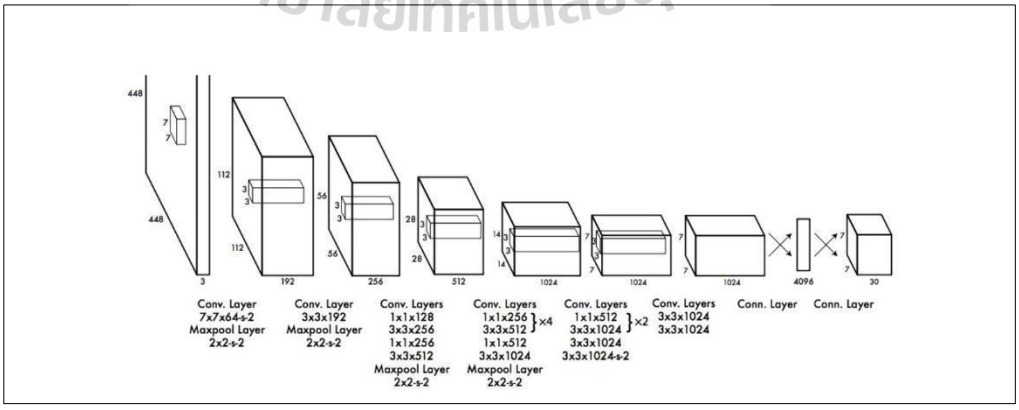
ฟาสต์เตอร์อาร์-ซีเอ็นเอ็นเป็นวิธีการซึ่งพัฒนาต่อยอดมาจากฟาสต์อาร์-ซีเอ็นเอ็น โดยปรับปรุงรีเจียน พร็อพโพซอล (Region proposals) ที่ใช้สำหรับหาอาร์โอไอ (ROI) ซึ่งเป็นส่วนที่ช้าที่สุดและเป็นคอขวดการคำนวณแล้ว นอกจากนี้การทำอาร์โอไอแบบเดิมยังเป็นกำหนดไว้ล่วงหน้า ที่ไม่สามารถเรียนรู้อะไรจากภาพได้ ฟาสต์เตอร์อาร์-ซีเอ็นเอ็น จึงได้เพิ่มส่วนเครือข่ายรีเจียนพร็อพโพซอล (Region proposal network: RPN) แยกออกมาจากซีเอ็นเอ็น โดยนำคุณลักษณะภาพที่สกัดแล้ว มาใช้ร่วมกับโครงข่ายประสาทเทียมเพื่อเรียนรู้ลักษะของภาพ และมาใช้คาดเดาว่าควรสร้างอาร์โอไอ ขึ้นที่ตรงไหนในภาพบ้าง โดยการใช้วิธีการแบ่งส่วนภายในภาพ (segmentation) เพื่อหาส่วนที่มีความน่าจะมีวัตถุในบริเวณนั้น และนำบริเวณที่มีลักษณะคล้ายกันมารวมกัน และสุดท้ายจึงสร้างเป็นกรอบภาพออกมาเป็นลำดับชั้น ดังแสดงในรูปที่ 2.18



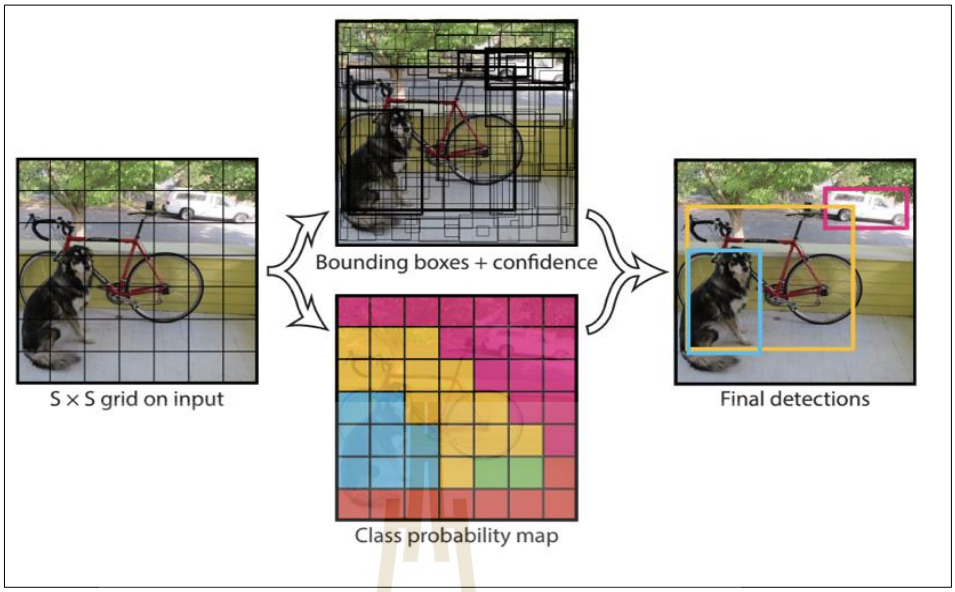
รูปที่ 2.18 ตัวอย่างขั้นตอนการทำฟาสต์เตอร์อาร์-ซีเอ็นเอ็น

2.4.2.5 โยโล (YOLO: You only Look Once)

โยโลเป็นสถาปัตยกรรมการตรวจจับวัตถุแบบขั้นตอนเดียวที่ได้รับความนิยมและมีประสิทธิภาพซึ่งใช้กันอย่างแพร่หลาย โดยถูกออกแบบมาเพื่อตรวจจับวัตถุแบบเรียลไทม์ได้อย่างรวดเร็วและแม่นยำ โดยแนวความคิดการตรวจจับวัตถุของโยโลจะแบ่งภาพอินพุตออกเป็นตารางและคาดคะเนวัตถุภายในแต่ละช่องต่างจากฟาสต์เตอร์อาร์ซีเอ็นเอ็นที่คาดเดาบริเวณที่จะให้ความสนใจแล้วจึงส่งไปคัดแยกวัตถุ โยโลโดยจะทำนายความน่าจะเป็นของ คลาส ตำแหน่งกล่องรอบวัตถุ(Bounding Box) และระดับความเชื่อมั่นของวัตถุ และใช้ Non-Max Suppression (NMS) เพื่อลบ Bouding Box ที่ทับซ้อนกัน ดังแสดงในรูปที่ 2.19 และ 2.20



รูปที่ 2.19 โครงสร้างการทำงานของโยโล



รูปที่ 2.20 ตัวอย่างการทำงานของโยโล

โดยจากแนวคิดดั้งเดิม YOLO ได้มีการพัฒนาและถูกนำเสนอออกมาอีกหลายเวอร์ชัน โดยแต่ละเวอร์ชันก็ได้รับการปรับปรุงและเปลี่ยนแปลง

YOLOv2 (Redmon et al., 2016) ได้นำฟีเจอร์เช่นการทำ anchor boxes, multi-scale predictions, and batch normalization layers. เพื่อเพิ่มประสิทธิภาพในการทำงานทำให้มีความแม่นยำที่มากขึ้น และทำให้สามารถตรวจจับวัตถุขนาดเล็กหรือถูกบดบังและตรวจจับวัตถุหลายชิ้นในเซลล์กริดเดียวได้

YOLOv3: (Redmon et al., 2018) ได้มีการปรับปรุงให้โครงข่ายมีจำนวนชั้นที่มากขึ้น ปรับปรุง anchor boxes เพื่อเพิ่มประสิทธิภาพ ปรับเปลี่ยน loss function ทำให้ให้มีประสิทธิภาพการทำงานมากขึ้น ลดการเกิด false positive ให้น้อยลง นอกจากใน YOLOv3 ได้มีการพัฒนา YOLOv3 PyTorch โดย Glenn Jocher ที่ได้ดัดแปลงจากเดิมที่ YOLO ทำงานบน DarkNet Framework ซึ่งใช้ภาษา C ให้สามารถทำงานบน PyTorch โดยใช้ภาษา Python ได้ ทำให้ YOLO ในเวอร์ชันนี้เป็นที่นิยมมากขึ้น

YOLOv4: (Bochkovskiy, 2021) มีการปรับปรุงและนำฟีเจอร์ใหม่ ๆมาใช้เพื่อเพิ่มประสิทธิภาพ เช่น Mosaic data augmentation, DropBlock regularization, and CIoU loss function. เพื่อสร้างความหลากหลายให้กับข้อมูลและทำให้มีความแม่นยำมากขึ้น

YOLOv5: (Jocher, 2021) YOLOv5 พัฒนาต่อเนื่องมาจาก YOLOv3 PyTorch แต่ได้รับการพัฒนาในแนวทางเดียวกับ YOLOv4 และรายละเอียดการพัฒนาหลักมีดังนี้

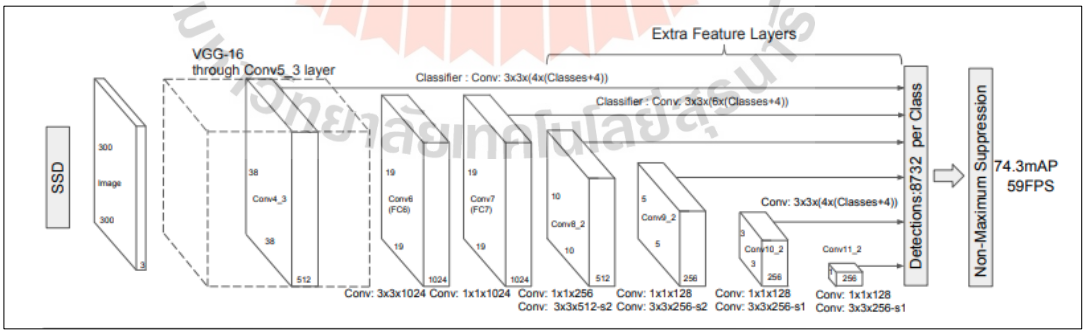
- ปรับปรุง PANet ที่ใช้ในส่วนหัว ทำให้ลดจำนวนพารามิเตอร์ลง และเพิ่มความเร็วในการทำนายและเพิ่มความแม่นยำ
- นำ FP16 เข้ามาใช้
- ปรับปรุง CSP เพื่อเพิ่มความเร็ว ลดขนาด และเพิ่มความแม่นยำ

YOLOv5-Face: (Qi et al., 2021) เป็นรุ่นพัฒนาของสถาปัตยกรรม YOLOv5 ในงานวิจัย ซึ่งออกแบบมาสำหรับการตรวจจับใบหน้าโดยเฉพาะ โดยเพิ่มส่วนของ Regression ซึ่งทำให้สามารถใช้เพื่อระบุจุดสำคัญบนใบหน้าได้ ดังแสดงในรูปที่ 2.21

รูปที่ 2.21 ภาพตัวอย่างผลลัพธ์การตรวจจับใบหน้ามนุษย์ของ YOLOv5-Face

2.4.2.6 เอสเอ็สดี (SSD: Single Shot Detector)

เป็นการพัฒนารูปแบบการตรวจจับวัตถุต่อยอดจากโยโล ซึ่งได้ใช้ซีเอ็นเอ็นในการเรียนรู้เช่นเดียวกันกับโยโลแต่เสนอให้ทายกล่องที่มีวัตถุต่าง ๆ จากหลายขนาด แทนที่จะทายขนาดเดียวแบบโยโล วิธีนี้ช่วยให้รับมือกับภาพที่มีวัตถุขนาด ต่างกันมากได้ โดยจะใช้โครงข่ายประสาทเทียมแบบวนซ้ำแบบ feedforward เพื่อทำนายประเภทวัตถุและตำแหน่งในภาพพร้อมกัน



รูปที่ 2.22 โครงสร้างการทำงานของเอสเอสดี

RCNN, YOLO และ SSD เป็นสถาปัตยกรรมการตรวจจับวัตถุที่ได้รับความนิยมและถูกใช้อย่างแพร่หลายในด้านการมองเห็นคอมพิวเตอร์และการเรียนรู้เชิงลึก โดยแต่ละสถาปัตยกรรมนี้มีจุดแข็งและจุดอ่อนของตัวเอง และการเลือกสถาปัตยกรรมที่จะใช้งานขึ้นอยู่กับความต้องการของงานนั้น ๆ

Faster RCNN เป็นแบบจำลองการตรวจจับวัตถุแบบสองขั้นตอนที่ขึ้นชื่อในด้านความแม่นยำและประสิทธิภาพสูง แบบจำลองนี้ใช้เครือข่ายข้อเสนอภูมิภาค (RPN) เพื่อสร้างข้อเสนอภูมิภาค จากนั้นจัดประเภทข้อเสนอแต่ละภูมิภาคเป็นวัตถุหรือพื้นหลัง RCNN ที่เร็วขึ้นมีความแม่นยำสูง แต่อาจมีราคาสูงในการคำนวณ ทำให้ไม่เหมาะกับการใช้งานแบบเรียลไทม์

YOLO (You Only Look Once) เป็นแบบจำลองการตรวจจับวัตถุแบบช็อตเดียวที่ออกแบบมาเพื่อดำเนินการตรวจจับวัตถุแบบเรียลไทม์ แบบจำลองนี้ใช้โครงข่ายประสาทเทียมแบบเส้นเดียวเพื่อคาดการณ์วัตถุในภาพ และมีประสิทธิภาพสูง ทำให้เหมาะสำหรับการใช้งานแบบเรียลไทม์ แม้ว่า YOLO จะไม่แม่นยำเท่ากับ Faster RCNN แต่ก็เร็วกว่าและมีประสิทธิภาพมากกว่า ทำให้เป็นตัวเลือกที่ดีสำหรับแอปพลิเคชันที่ต้องการความเร็วและประสิทธิภาพเป็นสำคัญ (Redmon et al., 2016)

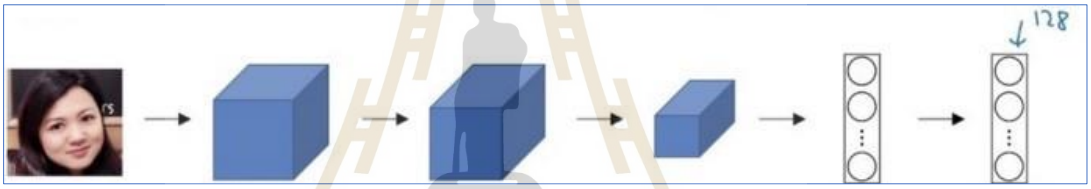
SSD (Single Shot MultiBox Detector) เป็นแบบจำลองการตรวจจับวัตถุแบบช็อตเดียวที่ออกแบบมาเพื่อดำเนินการตรวจจับวัตถุหลายสเกลภายในภาพเดียว แบบจำลองนี้ใช้ตัวแยกฟีเจอร์หลายสเกลเพื่อตรวจจับวัตถุในสเกลหลายขนาด และเป็นที่ทราบกันดีว่ามีประสิทธิภาพแบบเรียลไทม์ จึงเหมาะอย่างยิ่งสำหรับการใช้งานในกล้องวงจรปิดและแอพพลิเคชั่นยานยนต์อัตโนมัติ (Liu et al., 2016)

โดยสรุป เมื่อเลือกระหว่าง Faster RCNN, YOLO และ SSD สิ่งสำคัญคือต้องพิจารณาความต้องการในการใช้งาน Faster RCNN เป็นแบบจำลองที่มีความแม่นยำสูงซึ่งเหมาะสำหรับงานที่ความแม่นยำเป็นสิ่งสำคัญ ในขณะที่ YOLO เป็นแบบจำลองที่รวดเร็วและมีประสิทธิภาพซึ่งเหมาะสำหรับการใช้งานแบบเรียลไทม์ SSD เหมาะอย่างยิ่งสำหรับงานที่ต้องการการตรวจจับวัตถุหลายขนาด การเลือกรุ่นที่จะใช้จะขึ้นอยู่กับข้อกำหนดเฉพาะของงาน และการแลกเปลี่ยนระหว่างความแม่นยำ ความเร็ว และประสิทธิภาพที่ต้องการ ซึ่งในงานวิจัยชิ้นได้เลือกใช้ YOLO เนื่องจากในขั้นตอนการตรวจจับผู้วิจัยได้ให้ความสำคัญกับด้านความเร็ว ซึ่งเมื่อเปรียบเทียบระหว่าง Faster RCNN แล้ว YOLO จะมีประสิทธิภาพสูงกว่าในการตรวจจับวัตถุแบบเรียลไทม์ และเมื่อเปรียบเทียบกันระหว่าง YOLO และ SSD ที่เป็นสถาปัตยกรรมที่ตรวจจับวัตถุแบบขั้นตอนเดียวเช่นเดียวกัน และมีโดดเด่นในด้านความเร็วเหมือนกัน แต่สถาปัตยกรรมของ YOLOv5-Face ได้มีจุดที่สามารถทำนายได้ทั้งตำแหน่งของใบหน้าและจุดสำคัญบนใบหน้า ซึ่งจะถูกใช้ปรับตำแหน่งของภาพใบหน้าให้อยู่ในลักษณะหน้าตรงซึ่งมีผลต่อกระบวนการรู้จำใบหน้าในขั้นตอนถัดไป (Deng et al., 2009).

2.5 กระบวนการยืนยันใบหน้า

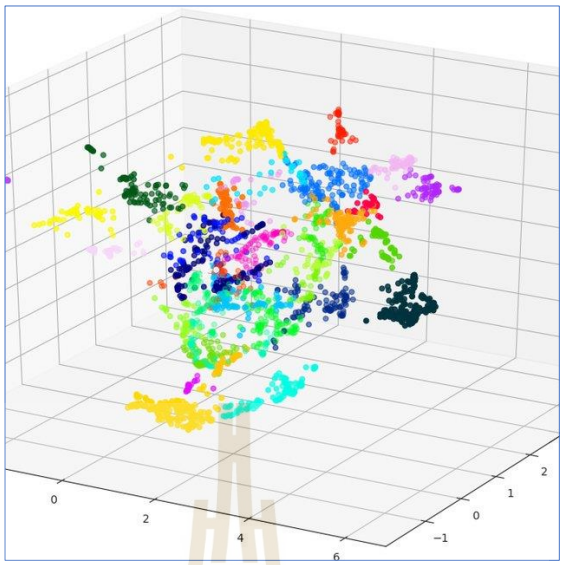
กระบวนการยืนยันใบหน้าเป็นกระบวนการที่สำคัญในการตรวจสอบว่าภาพใบหน้าสองภาพเป็นของบุคคลคนเดียวกันหรือไม่ โดยกระบวนการนี้ถือเป็นส่วนสำคัญในระบบการจดจำใบหน้า ซึ่งรวมถึงการเปรียบเทียบใบหน้ากับฐานข้อมูลของใบหน้าที่รู้จักเพื่อระบุตัวตนของบุคคล วิธีการยืนยันใบหน้านี้สามารถทำได้โดยการวัดค่าความคล้ายคลึงกันระหว่างใบหน้าสองใบหน้าและตัดสินว่าเป็นของบุคคลเดียวกันหรือไม่

การเปรียบเทียบใบหน้านั้นมีหลากหลายวิธี โดยทั่วไปแบ่งออกเป็นสองวิธีหลักคือ Feature-Based และ Deep Learning-Based วิธีการแบบ Feature-Based จะมุ่งเน้นที่การแยกลักษณะเฉพาะของใบหน้าที่สร้างขึ้นด้วยมือ เช่น ตา จมูก และปาก แล้วใช้คุณสมบัติเหล่านี้เพื่อเปรียบเทียบใบหน้า ในขณะที่วิธีการที่ใช้การเรียนรู้เชิงลึก (Deep Learning-Based) จะอาศัยเครือข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks: CNN) เพื่อเรียนรู้และสกัดคุณลักษณะเชิงลึกของใบหน้า และย่อขนาดมิติของภาพให้เหลือข้อมูลที่สำคัญ เรียกกระบวนการนี้ว่าการฝัง (Embedding) ดังรูปที่ 2.23 ซึ่งจะใช้เพื่อเปรียบเทียบใบหน้า



รูปที่ 2.23 ตัวอย่างขั้นตอนการแปลงภาพใบหน้าเป็น Embedding

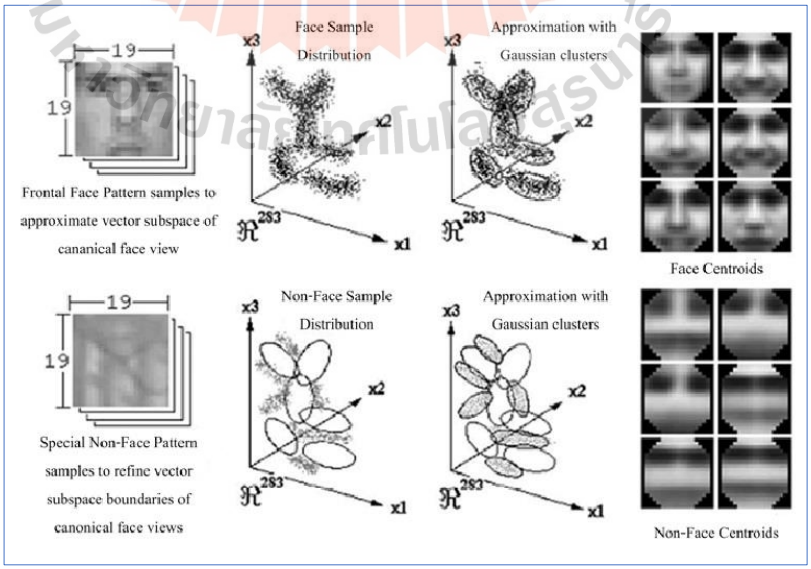
วิธีการทั่วไปในการเปรียบเทียบใบหน้าที่ใช้กันอย่างแพร่หลายคือการวัดระยะทางระหว่างข้อมูล เช่น ระยะทางแบบยุคลิด (Euclidean distance) หรือระยะทางแบบคอไซน์ (Cosine distance) เพื่อวัดความคล้ายคลึงกันระหว่าง embeddings ที่ได้จากการฝัง หากระยะห่างระหว่าง embeddings ต่ำกว่าเกณฑ์ที่กำหนดไว้ ระบบจะถือว่าใบหน้าทั้งสองเป็นของบุคคลคนเดียวกัน นอกจากนี้ยังมีการพัฒนาเทคนิคการเปรียบเทียบอื่น ๆ ที่ช่วยเพิ่มความแม่นยำและประสิทธิภาพในการยืนยันใบหน้าในสภาพแวดล้อมต่าง ๆ



รูปที่ 2.24 ตัวอย่างการแสดง Embedding Space ในรูปแบบ 3 มิติ

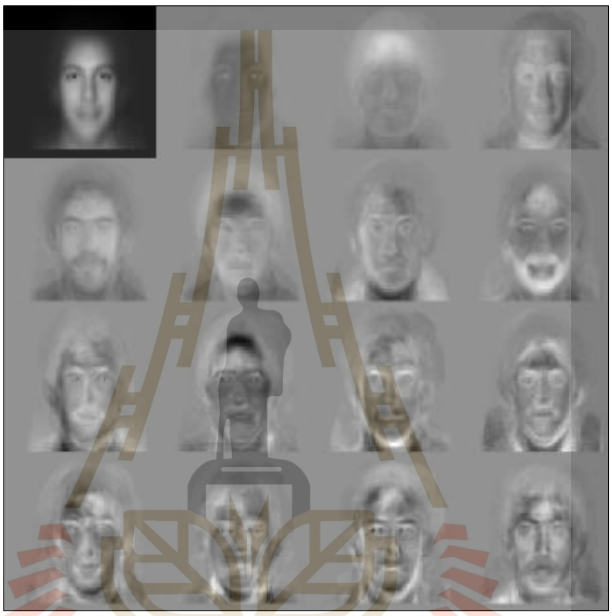
2.5.1 กระบวนสกัดคุณลักษณะใบหน้าแบบ Feature-Based Eigenface

ไอเกนเฟส (Eigenface) ถูกนำเสนอโดย Sirovich และ Kirby ไอเกนเฟส (Eigenface) ถูกนำเสนอโดย Sirovich และ Kirby ในปี ค.ศ. 1987 และถูกนำมาใช้กันอย่างแพร่หลายในระบบจดจำใบหน้าในยุคแรก ๆ เป็นวิธีการดั้งเดิมในการจดจำใบหน้า ซึ่งใช้การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PCA) เพื่อปรับภาพใบหน้าไปยังให้มีมิติที่น้อยกว่า โดยที่ภาพใบหน้าจะแสดงด้วยชุดของเวกเตอร์ลักษณะเฉพาะหรือลักษณะเฉพาะ



รูปที่ 2.25 การสกัดคุณลักษณะด้วย Eigenface

โดยเทคนิคในการพัฒนาระบบรู้จำใบหน้าก็จะทำได้โดยการแปลงภาพถ่ายใบหน้าบุคคลสองมิติไปเป็นเวกเตอร์หนึ่งมิติเก็บไว้ในฐานข้อมูล และเมื่อต้องการนำรูปภาพใบหน้าบุคคลที่สนใจมาเปรียบเทียบก็จะทำการแปลงภาพใบหน้านั้นเป็นเวกเตอร์หนึ่งมิติด้วย แล้วนำเวกเตอร์ไปเปรียบเทียบกับภาพในฐานข้อมูลเพื่อหาผลลัพธ์ ยกตัวอย่างการเปรียบเทียบใบหน้าที่ 1 เมื่อเปรียบเทียบกับเบสิคเซต มีความเหมือนกับ Eigenface 1 10% เหมือนกับ Eigenface 2 55% เหมือนกับ Eigenface 3 ติดลบ 3% เมื่อนำใบหน้าที่ 2 มาเปรียบเทียบ แล้วได้สัดส่วนของ % ในทิศทางเดียวกันนี้ ก็ถือว่า หน้าที่ 1 กับ หน้าที่ 2 นั้นเป็นหน้าเดียวกัน

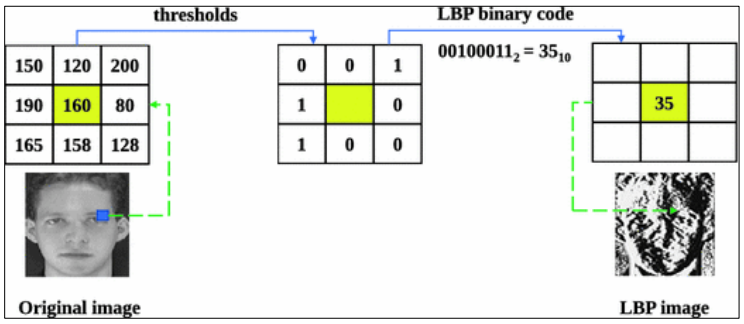


รูปที่ 2.26 ภาพตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วย Eigenface

Local Binary Patterns (LBP)

Local Binary Patterns (LBP) เป็นการวิเคราะห์พื้นผิวของรูปภาพ ซึ่งเป็นวิธีการสกัดคุณลักษณะข้อมูลแบบดั้งเดิมที่เป็นที่นิยมโดยเฉพาะในงานด้านการรู้จำใบหน้า โดยถูกนำเสนอครั้งแรกโดย Ojala et. al. ในปี 1996 ซึ่ง LBP

โดยขั้นตอนการทำงานของ LBP คือจะแบ่งรูปภาพออกเป็นช่องเล็ก ๆ แล้วเปรียบเทียบค่าความเข้มสีของพิเซลและเปรียบเทียบกับพิกเซลในบริเวณรอบ ๆ หากพิกเซลที่อยู่โดยรอบมีค่าความเข้มที่มากกว่าก็จะให้ค่าเป็น 1 และน้อยกว่าเป็น 0 จากนั้นจึงเอาค่าไบนารี่ของพิกเซลโดยรอบมาเรียงต่อกันเป็นเลขฐานสองและแปลงกับเป็นเลขฐานสิบ เพื่อใช้ลดมิติและอธิบายพื้นผิวของภาพ ดังแสดงในรูปที่ 2.27



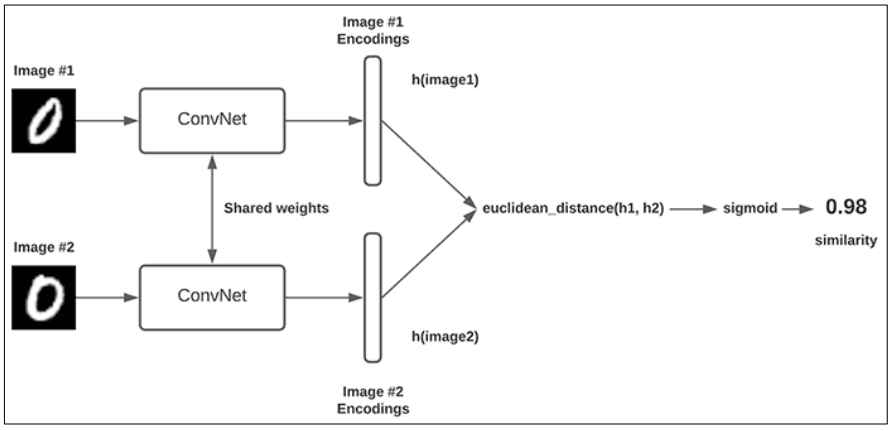
รูปที่2.27 ภาพขั้นตอนการสกัดคุณลักษณะด้วย LBP

2.5.2 กระบวนสกัดคุณลักษณะใบหน้าแบบ Deep Learning-Based Siamese Network

Siamese Network เป็นประเภทสถาปัตยกรรมของ neural network ซึ่งถูกนำเสนอโดย Bromley et al. in their 1994 ในเรื่อง "Signature Verification Using a Siamese Time Delay Neural Network" โดยสถาปัตยกรรมนี้ถูกออกแบบมาเพื่อเปรียบเทียบความเหมือนระหว่าง 2 ข้อมูลนำเข้า ซึ่ง Siamese Network เป็นที่นิยมและถูกประยุกต์เพื่อนำมาใช้งานอย่างหลากหลายในการเรียนรู้ความเหมือนและหนึ่งในงานที่ถูกนำมาใช้ก็คือการรู้จำใบหน้า

Siamese Network จะมี 2 เครือข่ายย่อยภายในเพื่อประมวลผลข้อมูลที่เข้ามาจาก 2 input โดยทั้งสองเครือข่ายจะใช้ค่าน้ำหนักเดียวกัน เพื่อให้ทั้งสองเครือข่ายย่อยนั้นมีการประมวลผลที่เหมือนกัน เมื่อข้อมูลผ่านการประมวลผลแล้วจะได้ผลลัพธ์ออกมาเป็น Embedding ของข้อมูล จากนั้นจะถูกนำไปเปรียบเทียบความเหมือนโดยใช้ Euclidian Distance เพื่อหาระยะห่างข้อมูลทั้งสองตัวากระยะทางน้อยนั้นแสดงว่าข้อมูลมีความเหมือนกัน

จุดเด่นของ Siamese Network เมื่อนำมาในงานการรู้จำใบหน้า แบบจำลองจะเรียนรู้จากความเหมือนของข้อมูล ทำให้สามารถสกัดคุณลักษณะเฉพาะของใบหน้าที่เหมือนกันได้ เมื่อฝึกฝนแบบจำลองโดยส่งภายของบุคคลเดียวกันทั้งสองภาพเข้าไปในแบบจำลอง ผลลัพธ์ที่ได้ออกมาก็จะมีค่าใกล้กันมากขึ้น



รูปที่ 2.28 ภาพตัวอย่างการทำงานของโมเดล Siamese Network

Triplet Neural Network

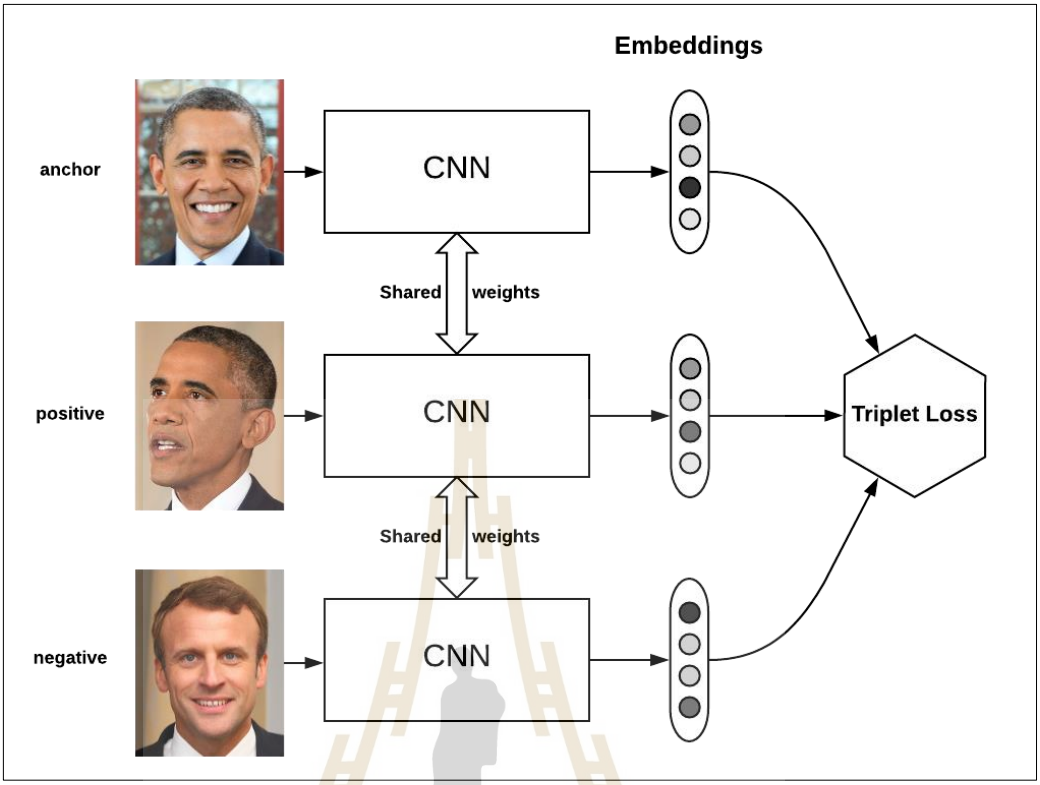
Triplet Neural Network ถูกนำเสนอในชื่อโปรเจค FaceNet ของ Google ในปี 2015 ซึ่ง FaceNet เป็น deep learning-based ในการรู้จำใบหน้าที่ได้รับความนิยมเป็นอย่างมาก โดยสถาปัตยกรรมนี้ถูกออกแบบมาเพื่อเปรียบเทียบระหว่าง 3 ข้อมูลนำเข้า เพื่อเรียนรู้ความเหมือนและความต่างของข้อมูล

Triplet Neural Network แตกต่างจาก Siamese Network โดยจะไม่มีเครือข่ายย่อยภายใน แต่จะมีเลเยอร์ในการรับข้อมูลทั้ง 3 ชุด ซึ่งประกอบไปด้วย ภาพหลัก(Anchor), ภาพเหมือน(positive), ภาพต่าง(negative) ก่อนที่จะถูกส่งไปประมวลผลในเครือข่ายเดียวกันเพื่อประมวลผลข้อมูลที่เข้ามาจาก 3 input เมื่อข้อมูลผ่านการประมวลผลแล้วจะได้ผลลัพธ์ออกมาเป็น Embedding ของข้อมูล จากนั้นจะถูกนำไปประมวลผลใน Loss Function เช่น Triplet Loss ที่จะคำนวน Euclidian Distance ของข้อมูลและปรับการกระจายตัวของข้อมูล โดยภาพใบหน้าของบุคคลเดียวกันจะมีระยะทางที่สั้น และภาพของบุคคลคนละคนกันนั้นจะต้องมีค่าระยะทางห่างออกจากภาพหลัก เพื่อจำแนกความเหมือนและต่างของใบหน้า

จุดเด่นของ Triplet Neural Network เมื่อนำมาใช้ในงานการรู้จำใบหน้าแบบจำลองจะเรียนรู้จากความเหมือนและความต่างของข้อมูล ทำให้แบบจำลองสามารถเรียนรู้และสกัดคุณลักษณะของใบหน้าที่เหมือนกันและจุดที่ต่างกันได้ ต่างจาก Siamese Network ที่เรียนรู้จากความเหมือนของคู่ข้อมูลเพียงอย่างเดียว

ในงานวิจัยชิ้นนี้ได้ให้ความสนใจกับเทคโนโลยีการเรียนรู้เชิงลึก จึงเลือกใช้เครือข่ายแบบ Triplet Neural Network เพื่อเรียนภาพคู่เหมือนและคู่ต่างของภาพ ซึ่งได้มีงานที่มาช่วยแก้ไขและเพิ่มประสิทธิภาพในการเรียนรู้ความแตกต่าง ดังแสดงตัวอย่างในรูปที่ 2.29

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 2.29 ภาพตัวอย่างการแสดง Embedding Space ในรูปแบบ 3 มิติ

2.5.3 กระบวนการเรียนรู้คุณลักษณะ Triplet Loss

Triplet Loss หรือฟังก์ชันค่า Loss แบบแฝดสาม เป็นฟังก์ชันที่นิยมใช้คู่กับ Triplet Neural Network ซึ่งถูกนำเสนอครั้งแรกใน FaceNet ของ Google ในปี 2015 ซึ่งถูกนำเสมอเพื่อมาใช้ในงานรู้จำใบหน้ามนุษย์

โดยหลักการคือเมื่อมีภาพใบหน้าเข้ามาภายในแบบจำลองและถูกแปลงเป็นเวกเตอร์ Triplet Loss จะคำนวณระยะทางระหว่าง 3 ใบหน้า โดยใบหน้าบุคคลคนเดียวกันต้องมีระยะทางที่ใกล้กัน และใบหน้าของบุคคลอื่นจะต้องระยะทางที่ไกลมากออกจากกัน โดยภาพใบหน้าหลักจะเรียกว่าภาพ Anchor ภาพใบหน้าบุคคลคนเดียวกันจะเรียกว่าภาพ Positive และภาพบุคคลอื่นจะถูกเรียกว่าภาพ Negative

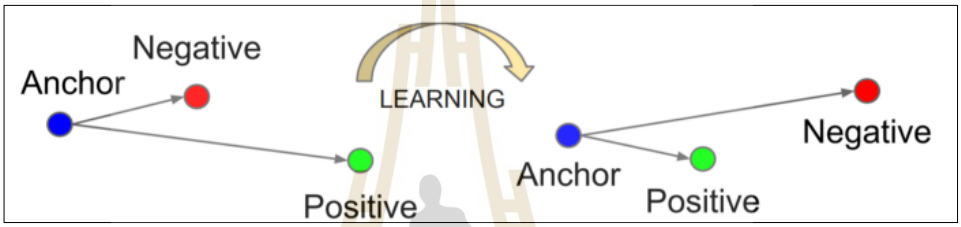
ซึ่งในการคำนวนระยะทาง Triplet Loss สามารถใช้ได้หลากหลายวิธีเช่นแบบ L2 หรือ Euclidean หรือวิธีอื่น ๆ ที่เป็นมาตราวัดในลักษณะของระยะห่าง โดยค่า Loss จะถูกคำนวนด้วยวิธีการดังนี้

$$L=\max(d(a,p)-d(a,n)+\text{margin},0)$$

โดยมีการแทนค่าดังนี้

L แทน ค่า Loss
d แทน ฟังก์ชันคำนวณระยะทาง
a แทน ภาพหลักหรือ anchor
p แทน ภาพเหมือนหรือ positive
n แทน ภาพต่างหรือ negative

margin เป็นค่า hyper parameter ที่สามารถปรับแต่งได้เพื่อปรับระยะที่เหมาะให้กับข้อมูล ให้มีความกระจายตัวต่อกัน



รูปที่ 2.30 ภาพตัวอย่างวิธีการปรับระยะของ Triplet Loss

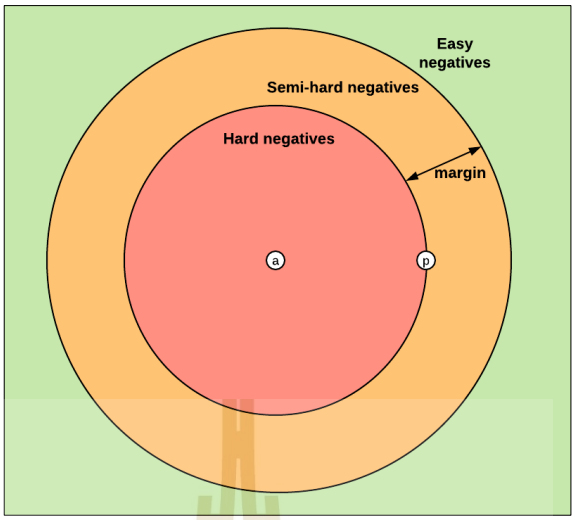
Triplet Mining

Triplet Mining เป็นเทคนิคอย่างหนึ่งในกระบวนการ Triplet Loss เนื่องจากโดยหลักการของ Triplet Network และ Triplet Loss นั้นเป็นเพียงกระบวนในการคำนวนข้อมูลจากภาพที่ถูกนำเข้ามาเท่านั้น ซึ่งโดยปกติแล้วภาพจะถูกนำเข้ามาด้วยการสุ่ม ดังนั้นชุดข้อมูลแบบแฝดสาม (Triplet) จะถูกแบบออกเป็น 3 ประเภท ดังแสดงในรูปที่ 2.31 ดังนี้

1) easy triplets เป็นชุดข้อมูลที่ให้ค่า loss เป็น 0 เนื่องจาก d(a,p)+margin < d(a,n) หรือค่าระยะทางระหว่างภาพคู่เหมือนใกล้กว่าภาพคู่ต่างอยู่แล้วจึงไม่เกิดผลในการจำแนกข้อมูล

2) hard triplets เป็นชุดข้อมูลที่คู่ต่างมีค่าระยะทางใกล้กว่าภาพคู่เหมือน หรือ d(a,n)<d(a,p) จึงทำให้เกิดการเรียนรู้ว่าภาพบางภาพแม้คล้ายกันแต่ไม่ใช่ข้อมูลที่ถูกต้อง

3) semi-hard triplets เป็นชุดข้อมูลที่เหมือนกับชุดข้อมูล easy triplet แต่เมื่อถูกคำนวนด้วยค่า margin แล้วมีระยะทางมากกว่าคู่ต่าง ทำให้สามารถนำมาใช้เรียนรู้ได้ หรือ d(a,p)<d(a,n)<d(a,p)+margin



รูปที่ 2.31 ตัวอย่างระยะห่างของชุดข้อมูลแต่ละประเภท

ซึ่งจะเห็นได้ว่าชุดข้อมูลที่สุ่มเข้ามาบางครั้งอาจไม่มีผลกับการฝึกฝนแบบจำลองเลย ทำให้เกิดความเสียเปล่าจากทั้งเวลาและทรัพยากร Hermans et al จึงได้เสนอ Triplet Mining ที่ได้พัฒนาการเลือกชุดข้อมูลเพื่อเพิ่มประสิทธิภาพในการคำนวนค่า Loss โดยเสนอมีการใช้ขนาดชุดการประมวลผล B = PK โดยที่ P แทนจำนวนของคลาสหรือบุคคลที่ต้องสุ่มขึ้นมาและ K คือจำนวนของภาพในแต่ละคลาส และไปเลือกชุดข้อมูลขึ้นมาโดยแบ่งเป็นสองแบบดังนี้

1) batch all เป็นกระบวนการเลือกชุดข้อมูลในกลุ่ม hard และ semi hard เพื่อปรับค่า loss ในขั้นที่ไม่สูงมาก

2) batch hard เป็นกระบวนการในแต่ละรอบจะหารูปที่ยากที่สุดใน batch หรือค่าภาพคู่เหมือนที่มีระยะทางมากที่สุด และภาพคู่ต่างที่มีระยะทางใกล้ที่สุด

โดยจากวิธีการทั้งสอง Herman ยังได้ระบุอีกว่าวิธี Batch Hard นี้เป็นวิธีการที่ดีที่สุดในการเรียนรู้แบบ Triplet Loss

Global Orthogonal Regularization

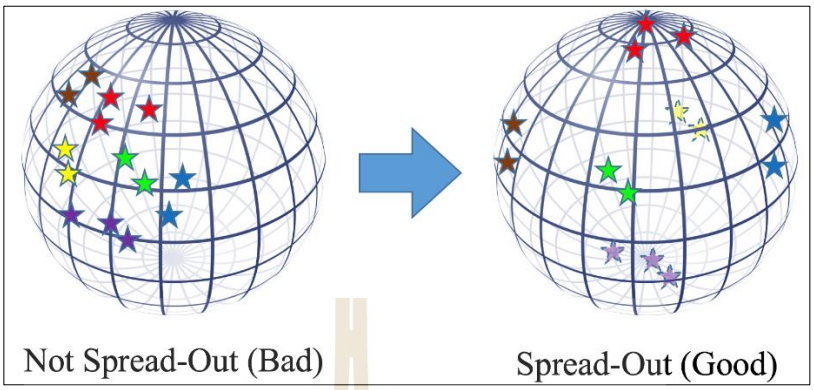
Global Orthogonal Regularization (GOR) เป็นกระบวนการปรับแต่งให้แบบจำลองมีประสิทธิภาพในการทำนาย โดยวิธีการนี้ถูกนำเสนอโดย Zhang et al. ซึ่งจุดประสงค์หลักของวิธีการนี้เพิ่มประสิทธิภาพในการจายตัวของข้อมูล โดยเสนอวิธีการเพิ่มประสิทธิภาพให้กับการใช้ฟังก์ชัน Pair-wise Loss และ Triplet Loss โดยใช้สมการดังนี้

$$\ell_{gor} = M_1^2 + \max(0, M_2 - \frac{1}{d})$$

$$M_1 = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x_i})^T f(\mathbf{x_i}^-)$$

$$M_2 = \frac{1}{N}\sum_{i=1}^{N} \left(f(\mathbf{x_i})^T f(\mathbf{x_i}^-)\right)^2$$

โดยใช้การเพิ่มโทษหรือค่าระยะทางให้กับชุดข้อมูลชุดต่าง ทำให้มีระยะทางออกห่างจากภาพมากขึ้นกว่าปกติ ดังแสดงตัวอย่างในรูปที่ 2.32



รูปที่ 2.32 ภาพจำลองการใช้ GOR เพื่อเพิ่มการกระจายตัวของข้อมูล

2.6 กระบวนการระบุใบหน้า

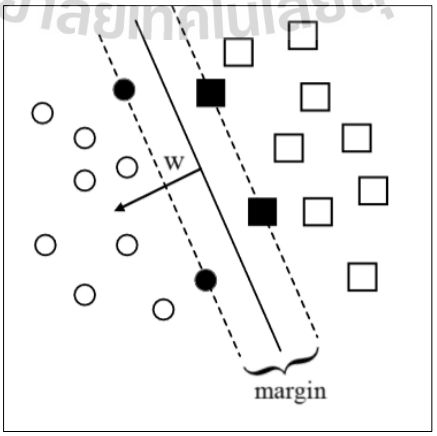
กระบวนการระบุใบหน้าเป็นกระบวนการที่จะนำฟีเจอร์ที่ได้จากการสกัดคุณลักษณะบนใบหน้ามาใช้เพื่อจัดเรียงหรือเปรียบเทียบข้อมูลชุดอื่นๆ เพื่อที่จะระบุตัวตนของแมว ซึ่งผู้วิจัยได้ทบทวนวรรณกรรมในหัวข้อต่าง ๆ ดังนี้

เอสวีเอ็ม (SVM)

เอสวีเอ็ม (SVM: Support Vector Machines) เป็นกระบวนการทางคณิตศาสตร์ซึ่งสามารถสร้างระนาบที่เหมาะสมที่สุดในการแบ่งกลุ่ม โดยจะพยายามสร้างเส้นแบ่งระหว่างข้อมูลให้มีระยะห่างระหว่างขอบของข้อมูลทั้งสองกลุ่มมากที่สุด

2.6.1 แอลเอสวีเอ็ม (Linear Support Vector Machines)

สำหรับการแก้ปัญหาแบบเป็นเชิงเส้นนั้น มิติของข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยใช้ระนาบการตัดสินใจ ดังแสดงในตัวอย่างรูป 2.32



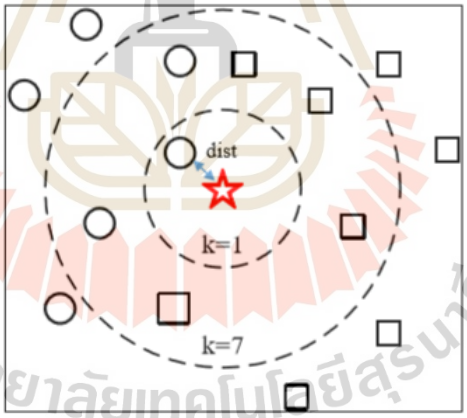
รูปที่ 2.33 การสร้างเส้นแบ่งข้อมูลของเอสวีเอ็ม (SVM)

วิธีของซัพพอร์ตเวคเตอร์นั้นจะพยายามหาระยะจากเส้นแบ่งถึงขอบของข้อมูลให้มากที่สุด โดยที่ค่ามาร์จินน้อยที่สุดเพื่อให้ได้เส้นแบ่งที่เหมาะสมที่สุด ซึ่งจุดที่อยู่บนเส้นขอบจะถูกเรียกว่า ซัพพอร์ตเวคเตอร์ จากรูปที่ 2.33 คือวงกลมสีดำและสี่เหลี่ยมสีดำที่อยู่บนเส้นประ

2.6.2 เอ็นแอลเอสวีเอ็ม (Non-linear Support Vector Machines) เอ็นแอลเอสวีเอ็มให้การแบ่งกลุ่มของข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (non-linear dataset) จำเป็นต้องแบ่งกลุ่มข้อมูลตัวอย่างไปสู่ปริภูมิมิติที่สูงขึ้น (higher dimensional space) หรือ ฟีเจอร์สเปซ (feature space) โดยการแปลงดังกล่าวต้องทำกับฟังก์ชันที่ไม่เป็นเชิงเส้น โดยสมการที่ใช้ในการคำนวณเพื่อแบ่งกลุ่มข้อมูลที่ไม่เป็นเชิงเส้น

เคเอ็นเอ็น (kNN: k-Nearest Neighbor)

เคเอ็นเอ็น (kNN: k-Nearest Neighbor) จะจำแนกข้อมูลโดยหาระยะระหว่างข้อมูลอินพุตที่สนใจกับข้อมูลที่อยู่ฐานข้อมูล จากนั้นนำระยะที่ห่างได้มาเรียงลำดับตัวที่ใกล้ที่สุด k ตัว แล้วดูว่าระยะห่างทั้ง k ตัวว่าข้อมูลใดมีค่าน้อยที่สุด ระบบก็จะระบุได้ว่าอินพุตนั้นตรงกันกับข้อมูลในฐานข้อมูลดังกล่าวโดยการคำนวณระยะความใกล้นั้นจะใช้การหาระยะทางแบบระยะยูคลิด (Euclidean distance) สำหรับการวัดระยะทางระหว่างสองวัตถุ ถ้าวัตถุห่างกันมากแสดงว่ามีความคล้ายน้อย ถ้าวัตถุห่างกันน้อยแสดงว่ามีความคล้ายกันมาก

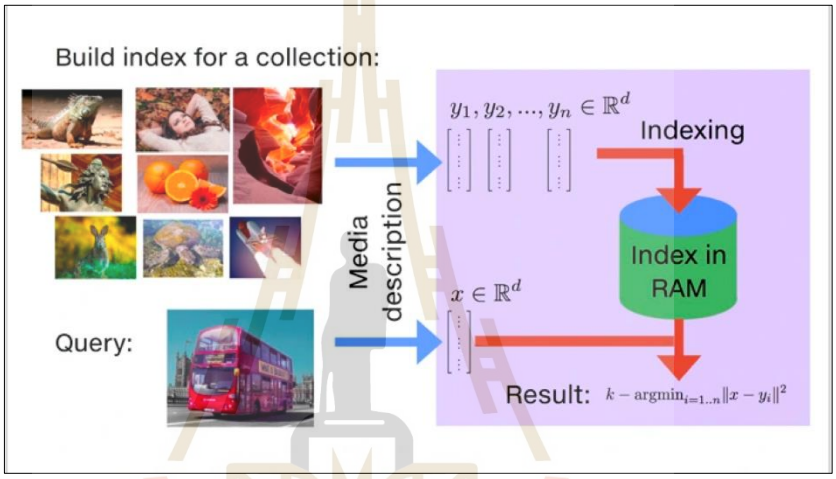


รูปที่ 2.34 การทำงานของเคเอ็นเอ็น

จากรูปที่ 2.34 จะเห็นว่าหากต้องการจำแนกว่าดาวอยู่ในกลุ่มข้อมูลใดระหว่างกลุ่มวงกลม หรือกลุ่มสี่เหลี่ยม สามารถจำแนกได้ด้วยเคเอ็นเอ็น (kNN) โดยขั้นตอนแรกเป็นการคำนวณระยะห่างระหว่างดาวกับข้อมูลทุกจุดที่อยู่ใกล้เคียง จากนั้นหากพิจารณา k=1 ก็จะสนใจเพียงสมาชิก 1 ตัวที่อยู่ใกล้กับตัวที่สนใจมากที่สุด โดยในรูปนี้จะตัดสินว่าเป็น ข้อมูลวงกลม แต่หากพิจารณา k=7 ก็จะ สนใจสมาชิกที่อยู่รอบข้าง 7 ตัว โดยพบว่ามีข้อมูลวงกลม 3 ตัว และข้อมูลสี่เหลี่ยม 4 ตัว ดังนั้นระบบจะตัดสินว่าดาวจัดอยู่ในกลุ่มเดียวกันกับข้อมูลสี่เหลี่ยม จะเห็นว่าที่ค่า k แตกต่างกัน ส่งผลถึงคำตอบที่ได้แตกต่างกันด้วย

FAISS (Facebook AI Similarity Search)

Facebook AI Similarity Search (FAISS) เป็นไลบรารีโอเพ่นซอร์สที่ถูกพัฒนาโดย Facebook สำหรับการค้นหาตามความคล้ายคลึงกันของข้อมูล FAISS ถูกนำมาใช้ในงานวิจัยชิ้นนี้ ด้วยจุดเด่นคือด้านความเร็วและความสามารถในการจัดกลุ่มของข้อมูลในรูปแบบ Vector โดยในงานวิจัยชิ้นนี้ จะแบบจำลองในขั้นตอน Verification สร้างเวกเตอร์จากใบหน้าแมวในฐานข้อมูล และใช้ FAISS เพื่อสร้าง Index ของใบหน้าแมวเก็บไว้ใน RAM ในการเรียกใช้งาน เมื่อมีการส่งใบหน้าแมวเข้ามาใหม่ จะประมวลผลด้วยแบบจำลอง ก่อนจะนำเวกเตอร์ที่ได้มาเปรียบเทียบกับใบหน้าที่ถูกทำ Index ไว้ใน RAM ดังแสดงกระบวนการทำงานในรูปที่ 2.35



รูปที่ 2.35 ภาพอธิบายกระบวนทำงานของ FAISS

2.7 งานวิจัยที่เกี่ยวข้องกับการรู้จำสัตว์เลี้ยง

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าปัจจุบันมีการศึกษาเกี่ยวกับระบุตัวตนของสัตว์ทั้งในประเภทของสัตว์เลี้ยงและปศุสัตว์ โดยในหลายงานวิจัยได้กล่าวถึงความเป็นสมาร์ทซิตี้หรือการเปลี่ยนผ่านเข้าสู่ยุคดิจิทัล (Kumar and Singh, 2018; Moreira et al, 2019) ในด้านของปศุสัตว์การระบุตัวตนสัตว์ก็จะสามารถทำให้ระบุที่มาของอาหาร ควบคุมดูแลโรคได้ (Tharwat et al., 2014; Awad, 2013; Kumar et al., 2018) ทำให้ผู้บริโภคสามารถเชื่อใจวัตถุดิบที่รับประทานได้มากขึ้น ในด้านของสัตว์เลี้ยงด้วยจำนวนของผู้เลี้ยงสัตว์ทั่วโลกมีแนวโน้มที่มากขึ้นเรื่อย ๆ (ฐานข้อมูลเพื่อการขึ้นทะเบียนสุนัข-แมว, 2564; วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล, 2566; สำนักงานปศุสัตว์, 2561; Statista, 2017) สิ่งที่มักเกิดขึ้นกับผู้เลี้ยงสัตว์ คือ การพลัดหลงกับสัตว์เลี้ยงหรือหลุดหายไป ทำให้มีการศึกษาเกี่ยววิธีการที่จะระบุตัวตนของสัตว์ว่าจะใช้วิธีการใด เนื่องจากวิธีการที่มีอยู่ส่วนใหญ่ปัจจุบันยังมีปัญหาด้านความรุนแรง การสูญหาย หรือต้นทุนของวัสดุ เป็นต้น ทำให้มีงานวิจัยนำเสนอกับวิธีการใหม่ ๆ ในการระบุตัวตนของสัตว์เลี้ยง ผู้วิจัยได้แบ่งงานวิจัยเกี่ยวข้องที่ได้ทบทวนวรรณกรรมมาออกเป็นสามด้าน ได้แก่ 1) ด้านวิธีการระบุตัวตนของสัตว์เลี้ยงในปัจจุบัน เพื่อศึกษา

เกี่ยวกับปัญหาและวิธีการในการระบุตัวตนของสัตว์เลี้ยงในปัจจุบันเพื่อให้ทราบถึงข้อดีข้อเสียและวิธีการที่จะสามารถพัฒนาต่อได้อนาคต 2) ด้านการระบุตัวตนของปศุสัตว์ เพื่อศึกษาเกี่ยวกับวิธีการและเทคโนโลยีที่ใช้ในปศุสัตว์ ข้อจำกัด และรูปแบบการพัฒนา เพื่อนำมาประยุกต์ใช้ภายในงานวิจัย 3) ด้านการระบุตัวตนสัตว์เลี้ยง เพื่อศึกษาเกี่ยวกับงานวิจัยที่มีลักษณะเดียวกันว่ามีการใช้วิธีการใดในการระบุตัวตน ซึ่งทำให้ทราบถึงข้อดี ข้อเสีย และข้อจำกัดในแต่ละวิธีการ ซึ่งจะถูกนำมาต่อยอดการพัฒนาในงานวิจัยชิ้นนี้

จากการศึกษางานวิจัยด้านวิธิการระบุตัวตนสัตว์เลี้ยงในปัจจุบันพบว่าวิธีการระบุตัวตนสัตว์แบบเดิมแบ่งออกเป็น 3 ประเภทได้แก่ การระบุตัวตนแบบถาวร การระบุตัวตนแบบกึ่งถาวร และการระบุตัวตนแบบชั่วคราว ซึ่งตัวอย่างของการระบุตัวตนแต่ละแบบได้แก่ การระบุตัวตนแบบถาวร เช่น การสักใบหู การฝังไมโครชิป การเจาะหูหรือการตัดหู การประทับร้อน การประทับเย็น ตัวอย่างการระบุตัวตนแบบกึ่งถาวรเช่น การติดป้ายที่หู การใส่ปลอกคอ เป็นต้น ซึ่งจะพบว่าวิธีการแบบถาวรและกึ่งถาวรนั้นเป็นวิธีการที่รุนแรงกับร่างกายของสัตว์ ต้องใช้ผู้เชี่ยวชาญหรือผู้ที่มีประสบการณ์ และมีความเสี่ยงในการติดเชื้อขึ้นกับสัตว์ ตัวอย่างการระบุตัวตนแบบชั่วคราว เช่น การย้อมสี การแขวนป้าย RFID วิธีการแบบชั่วคราวจะมีการปฏิสัมพันธ์กับร่างกายของสัตว์น้อยกว่าเนื่องจากเป็นการสวมใส่อุปกรณ์ทำให้ไม่มีการใช้ความรุนแรง แต่ก็เป็นวิธีที่สามารถสลาย หรือสูญหายได้ ปัญหาเรื่องการระบุตัวตนจึงเป็นสิ่งที่ท้าทายในงานวิจัยด้านนี้ ซึ่งได้มีการนำเสนอวิธีการแก้ปัญหาโดยใช้ ชีวมาตรสัตว์ โดยแบ่งออกเป็น 4 วิธีดังนี้ 1)การระบุจากจุดบนจมูก 2)การระบุจากรูปแบบดวงตา 3)การระบุจากใบหน้า 4)การระบุจากหลอดเลือดจอประสาทตา (Kumar, 2018) ซึ่งวิธีการระบุตัวตนโดยใช้ชีวมาตรสัตว์จะสามารถลด ความรุนแรง ความเสี่ยงที่จะเกิดโรค และโอกาสสูญหายของอุปกรณ์ได้

แต่อย่างไรก็ตามการระบุตัวตนสัตว์เลี้ยงด้วยชีวมาตรสามารถทำได้หลายวิธีแต่ละวิธีนั้นก็มีข้อจำกัดที่แตกต่างกัน ในงานวิจัยชิ้นนี้ผู้วิจัยได้ให้ความสนใจกับการพัฒนาการรู้จำด้วยใบหน้า ซึ่งเมื่อศึกษาจากงานวิจัยที่เกี่ยวข้องแล้วในกระบวนการสัตว์เลี้ยงจะถูกแบ่งออกเป็น 3 ขั้นตอนคือ 1) การระบุใบหน้า 2) การระบุคุณลักษณะ 3) การระบุตัวตน ซึ่งเมื่อเปรียบเทียบในแต่ละขั้นตอนจะมีรายละเอียดดังนี้

ขั้นตอนการระบุใบหน้า เป็นขั้นตอนแรกเริ่มข้อการรู้จำใบหน้า เพื่อที่จะค้นหาใบหน้าของสิ่งที่เราสนใจก่อนจะนำมาสกัดคุณลักษณะ ซึ่งการระบุใบหน้าจะถูกออกมาเป็น 2 แบบคือ Feature-Based และ Deep Learning-Based ซึ่งแบบ Feature-Based เช่น ในงานของมอเรราห์และคณะ (Moreira et. al., 2019) ได้ใช้ EigenFace และ Fisher Face ในการตรวจจับใบหน้าสุนัข งานของรวนและคณะ (Ruan et. al., 2014) เลือกใช้ Deformable Parts Model (DPM) ซึ่งมีจุดเด่นคือความรวดเร็วและเข้าใจได้ง่าย แต่ก็มีข้อจำกัด เช่น แบบจำลองเพื่อสกัดคุณลักษณะไม่มีความยืดหยุ่นทำให้ใช้งานต่อได้ยาก หรือต้องใช้มนุษย์เป็นผู้กำหนดคุณลักษณะด้วยตัวเองทำให้ต้องใช้เวลา และอาจทำให้เกิดความลำเอียงในการกำหนดคุณลักษณะได้ ทั้งนี้ในระยะหลังวิธีการแบบ Deep Learning Based นั้นก็เป็นได้รับความนิยมมากขึ้นด้วยปัจจัยทางด้านฮาร์ดแวร์ที่มีประสิทธิภาพมากขึ้นและซอฟต์แวร์ที่มีมีการพัฒนาเครื่องมือใหม่ ๆ ให้สามารถใช้งานได้ง่ายขึ้นโดยในงานของยูนและคณะ (Yoon et al., 2016) ได้เลือกใช้ ResNet-like หรือลินและคู (Lin and Kuo, 2018) ได้ใช้ Faster R-CNN แบบสองขั้นตอนในการค้นหาบริเวณที่น่าสนใจและใบหน้า อดัม ไคล์น (Adam

Klein, 2019) เลือกใช้ YOLOv3 ในการตรวจจับใบหน้า ซึ่งวิธีการแบบ Deep Learning มีข้อดีสามารถให้คอมพิวเตอร์เป็นผู้ค้นคุณลักษณะข้อมูลได้ มีความยืดหยุ่นและสะดวก สามารถนำไปใช้งานต่อในงานใกล้เคียงได้แต่ก็ต้องใช้พลังในการประมวลผล ซึ่งโดยสรุปแล้วแต่ละวิธีนั้นก็จะมีข้อดีและข้อเสียที่แตกต่างกัน ซึ่งในงานวิจัยชิ้นนี้จึงเลือกใช้ YOLOv5-Face เนื่องจากเป็นวิธีการแบบการเรียนรู้เชิงลึก และมีความสามารถในการตรวจจับใบหน้าได้ในหลากหลายสภาพแวดล้อม มีการทำงานที่รวดเร็วและสามารถระบุจุดสำคัญบนใบหน้าได้

ขั้นตอนการสกัดคุณลักษณะ เป็นขั้นตอนที่คอมพิวเตอร์จะเรียนรู้ลักษณะของภาพ เช่นเดียวกับขั้นตอนระบุใบหน้าการสกัดคุณลักษณะทำได้ 2 แบบคือ Feature-Based และ Deep Learning-Based ซึ่งจากการทบทวนวรรณกรรมแต่ละงานวิจัย จะมีการเลือกใช้วิธีที่แตกต่างกัน อาทิ วิธีการแบบ Feature-Based ซึ่ง คุมาร์และซิง (Kumar, Singh, 2014) ได้นำเสนอวิธีการสกัดคุณลักษณะแบบ Fisher Linear Projection and Preservation (FLPP) ที่เป็นกระบวนสกัดคุณลักษณะของใบหน้าสุนัขพร้อมกระจายชุดข้อมูลระหว่างคลาส งานวิจัยของ หลินและคู (Lin and Kuo, 2018) ที่ถึงแม้จะเลือกใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันในการระบุตัวตน แต่เลือกใช้ EigenFace เพื่อสกัดคุณลักษณะบนใบหน้าแมวจากชิ้นส่วนใบหน้า และในด้านการใช้ Deep Learning Based ไลและคณะ (Lai et. al., 2019) ได้ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันสถาปัตยกรรม Faster R-CNN ในการเรียนรู้คุณลักษณะของภาพโดยนำกระบวนการแบบ Coarse-to-fine โดยใช้แบบจำลองสองตัว แบบจำลองตัวแรกเพื่อทำนายพันธุ์ของสุนัขและ แบบจำลองตัวที่สองเพื่อสกัดคุณลักษณะของใบหน้าเพื่อระบุตัวตน งานวิจัยของมูเก้ตและคณะ (Mougeot et. al., 2019) แม้ไม่มีขั้นตอนของการระบุใบหน้าแต่ใช้การสกัดคุณลักษณะจากภาพใบหน้าสุนัขที่ผ่านการประมวลผลมาแล้วด้วยการใช้ โครงข่ายประสาทเทียมแบบแฝดสามคู่กับฟังก์ชัน Triplet Loss เพื่อเรียนรู้คุณลักษณะของใบหน้า เช่นเดียวกับงานของ อดัม ไคล์น (Adam Klein, 2019) และยูนและคณะ(Yoon et. al., 2021) ซึ่งงานวิจัยชิ้นนี้ให้ความสนใจการเรียนรู้เชิงลึกจึงเลือกใช้วิธีการแบบ Deep Learning Based โดยเลือกใช้โครงข่ายประสาทเทียมแบบแฝดสามกับฟังก์ Triplet Loss มาใช้ในการพัฒนา เนื่องจากเป็นกระบวนการที่ได้ความนิยมและได้รับการพัฒนาวิธีการอย่างต่อเนื่อง จึงง่ายต่อการพัฒนาและต่อยอดทั้งในปัจจุบันและอนาคต

ขั้นตอนการระบุตัวตนเป็นขั้นตอนที่จะนำข้อมูลที่ถูกสกัดคุณลักษณะมารวบรวมเอาไว้ เพื่อเปรียบเทียบกับภาพใบหน้าเป้าหมายที่ถูกนำเข้ามาเพื่อระบุตัวตน ซึ่งกระบวนการที่นิยมนั้นมีหลากหลายวิธีดังนี้ Support Vector Machine (SVM) ซึ่งถูกใช้ในงานวิจัยของ หลินและคู (Lin and Kuo, 2018) เพื่อเปรียบเทียบค่าคุณลักษณะโดยใช้การลากผ่านของค่าข้อมูล หรือ K-Nearest Neighbor(K-NN) ที่ถูกใช้ในงานของ ของมูเก้ตและคณะ (Mougeot et. al., 2019) อดัม ไคล์น (Adam Klein, 2019) และยูนและคณะ(Yoon et. al., 2021) ที่ใช้การเปรียบเทียบค่าจากข้อมูลในตำแหน่งโดยรอบ โดยจุดเด่นคือสามารถปรับค่าข้อมูลนำเข้าแล้วจำนวนเพื่อนบ้านใกล้เคียงได้ทำให้มักถูกนำมาวัดผลการทำนายแบบ 1 อันดับจนถึง 5 อันดับได้

การศึกษางานวิจัยด้านวิธีการระบุตัวตนสัตว์เลี้ยงในแง่ผลภาพรวมของวิธีการมีรายละเอียดดังนี้ เชงและคณะ (Chen et al., 2016) ได้นำเสนอวิธีการระบุตัวตนแมวโดยใช้ลักษณะของจมูก เนื่องจากลักษณะจมูกของสัตว์ส่วนใหญ่จะมีลักษณะเหมือนลายนิ้วของคน เช่นเดียวกับปศุสัตว์ โดยได้มีการทดลองใช้รูปจมูกแมวกว่า 700 รูป จากแมว 70 ตัว เพื่อระบุตัวตนของแมว แต่อย่างไรก็ตาม

งานวิจัยดังกล่าวไม่ได้ระบุผลการทดลองที่เป็นตัวเลขไว้อย่างชัดเจน หลิวและคณะ (Liu et al., 2012) ได้ใช้การตรวจจับใบหน้าสุนัขเพื่อคัดแยกสายพันธุ์โดยใช้สไลดิงวินโดวส์ (sliding window) เพื่อระบุจมูกและตาเพื่ออ้างอิงไปยังจุดที่เป็นหูและส่วนประกอบที่เหลือ จากนั้นจึงวิเคราะห์ฮีสโตแกรม (Histogram) และใช้ เอสไอเอฟที (SIFT: Scale Invariant Feature Transform) ในการวิเคราะห์เพื่อหาจุดเด่นของรูปภาพที่ได้รับเข้ามา ซึ่งมีความแม่นยำ 67% จากสุนัข 133 สายพันธุ์ จำนวน 8,351 รูปภาพ และในงานวิจัยของ ลินและคู (Lin and Kuo, 2018) พัฒนาระบบรู้จำใบหน้าแมวโดยใช้การเรียนรู้เชิงลึกแบบฟาสต์เตอร์อาร์-ซีเอ็นเอ็น (Faster R-CNN) โดยศึกษาจากรูปภาพของแมวจำนวน 150 ตัว ซึ่งในการพัฒนา ผู้วิจัยได้แบ่งแบบจำลองออกเป็นสองส่วน คือ ส่วนหรับการตรวจจับใบหน้า และส่วนสำหรับการระบุส่วนประกอบบนใบหน้า ได้แก่ หน้าผาก ตา จมูก และปาก จากนั้นจึงใช้เอสวีเอ็ม (SVM) ในการวิเคราะห์ลักษณะของส่วนประกอบเพื่อทำการคัดแยกแมวแต่ละตัว ซึ่งในการทดลองสามารถระบุตัวตน โดยมีความแม่นยำภาพ 94.1% แต่อย่างไรก็ตามยังมีข้อจำกัดด้านสภาพแสงและแมวบางลักษณะ เช่น แมวดำ ที่ไม่สามารถแยกคุณลักษณะเฉพาะออกมาเพื่อคัดแยกได้

ผู้วิจัย	ชื่องานวิจัย	รูปแบบ	การระบุใบหน้า						การสกัดคุณลักษณะ										กระบวนการระบุใบหน้า					
			กระบวนการ	ระบุตำแหน่งใบหน้า	วาดแบบกล่องใบหน้า	ระบุจุดสำคัญบนใบหน้า	ปรับตำแหน่งใบหน้า	ระบุสายพันธุ์	จำนวนชั้น	Eigen Face	HOG	Local Binary Pattern	Fisher Linear Projection and Preservation (FLPP)	Convolutional Neural Network (CNN)	Triplet Neural Network	Siamese Network	GDR	Triplet Mining	SVM	KNN	RANSAC	LDA	FAISS	CNN (Prediction)
Adam Klein	Pet Cat face Verification and Identification	Deep Learning Based	YOLO	x	x				1						x		x	x		x				
Lin, Kou	Cat face recognition using deep learning	Mix Based	Faster RCNN	x		x			2	x									x					
Lai et. al.	Dog Identification using Soft Biometrics and Neural Networks	Deep Learning Based	Faster RCNN	x		x	x	x	2					x										x
Mougeot et al.	A Deep Learning Approach for Dog Face Verification and Recognition	Deep Learning Based													x			x						
Moreira et. al.	Where is My Puppy? Retrive Lost Dogs by Facial Features	Mix Based								x		x		x					x					
Kumar, Singh	Monitoring of pet animal in smart cities using animal biometrics	Feature Based											x						x			x		
Chen et. al.	Locality Constrained Sparse Representation for Cat Recognition	Feature Based									x								x					
Ruan et. al.	Detection of user-registered dog faces	Feature Based	DPM	x		x			1		x													
งานวิจัยชิ้นนี้		Deep Learning Based	YOLOv5-Face	x	x	x	x		1						x		x	x		x			x	

รูปที่ 2.36 ภาพเปรียบเทียบวิธีการรู้จำใบหน้าแมวจากการทบทวนวรรณกรรม

จากการทบทวนรวรรณกรรมผู้วิจัยพบว่าในการพัฒนาระบบรู้จำใบหน้าแมวโดยใช้การเรียนรู้เชิงลึก จะใช้การระบุตัวตนแบบชีวมาตรโดยการรู้จำใบหน้าซึ่งจะช่วยลดความรุนแรง ลดการใช้อุปกรณ์ และช่วยป้องกันการสูญหายของอุปกรณ์ในการระบุตัวตน

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยชิ้นนี้มีจุดมุ่งหมายเพื่อออกแบบและพัฒนาระบบรู้จำใบหน้าแมวโดยใช้การเรียนรู้เชิงลึก โดยมุ่งเน้นความแม่นยำในการระบุตัวตนจากใบหน้าและลักษณะของสัตว์เลี้ยงได้โดยมีรายละเอียดดังนี้

3.1 วิธีวิจัย

3.2 ชุดข้อมูลที่ใช้ในงานวิจัย

3.3 การวัดประสิทธิภาพของระบบ

3.1 วิธีวิจัย

งานวิจัยชิ้นนี้เป็นงานวิจัยเชิงพัฒนา (Research and Development) โดยในหัวข้อนี้ได้กล่าวถึงวิธีที่ใช้วิจัยวงจรพัฒนาระบบ (System Development Life Cycle: SDLC) มาเป็นแนวทางเพื่อลดความซ้ำซ้อน ของขั้นตอนการปฏิบัติงานและข้อผิดพลาดในแต่ละขั้นตอน อีกทั้งยังช่วยให้ทำงานเป็นระบบแยกส่วนการทำงานได้ชัดเจน ซึ่งแบ่งได้เป็น 5 ส่วนดังนี้

1. วางแผน
2. วิเคราะห์
3. ออกแบบระบบ
4. พัฒนาระบบ
5. ดูแลรักษาระบบ

วิธีดำเนินการวิจัยเป็นไปตามลำดับดังนี้

3.1.1 วางแผน

จากการศึกษาพบว่าปัจจุบันสังคมประเทศไทยปัจจุบันทิศทางจำนวนผู้เลี้ยงสัตว์เลี้ยงมีมากขึ้นเนื่องจากรูปแบบการใช้ชีวิต ซึ่งตัวเลขของภาคธุรกิจเป็นตัวหลักฐานที่แสดงถึงการขยายตัวของผู้เลี้ยงสัตว์เลี้ยง ปัญหาที่มักเกิดขึ้นกับผู้เลี้ยงสัตว์เลี้ยงก็คือสัตว์เลี้ยงพลัดหลงและไม่ได้กลับมา เนื่องจากผู้ที่พบเจอไม่สามารถแยกได้ว่าเป็นสัตว์ที่พบนั้นเป็นสัตว์เลี้ยงที่พลัดหลงหรือเป็นสัตว์จรจัด ด้วยจำนวนสัตว์จรจัดที่เยอะ ทำให้ยากที่จะแยกแยะสัตว์เลี้ยงพบเห็นได้ และวิธีการระบุตัวตนส่วนใหญ่ที่มีในปัจจุบันนั้น หากเป็นวิธีการถาวรก็มักต้องใช้ผู้เชี่ยวชาญและมีความเสี่ยงในการดำเนินงานเช่นการผ่าตัดฝังไมโครชิป ส่วนวิธีการแบบชั่วคราวนั้นก็มีโอกาสที่อุปกรณ์จะสูญหายได้ ทำให้ผู้วิจัยเห็นว่าเทคโนโลยีการรู้จำใบหน้าจะทำให้สามารถระบุตัวตนสัตว์เลี้ยงโดยใช้ชีวมาตรได้ ซึ่งจะช่วยลดความเสี่ยงในการระบุตัวตนแบบถาวรเช่นการผ่าตัด และช่วยลดค่าอุปกรณ์ในการระบุตัวตนแบบชั่วคราว จึงเกิดเป็นแนวคิดของการพัฒนาแบบจำลองในงานวิจัยนี้

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.1.2 วิเคราะห์

ปัญหาที่เกิดกับสัตว์เลี้ยงก็คือ สัตว์เลี้ยงพลัดหลงและไม่ได้กลับมา โดยเปรียบเทียบจากสถิติแล้วจำนวนสุนัขและแมวมีอัตราการหายไปใกล้เคียงกัน แต่อัตราในการกลับมาของสุนัขสูงกว่าแมวอยู่ 5 เท่า แมวมีอัตราการหายไปและเสียชีวิตสูง โดยแม้จะใช้วิธีการตามหาเช่น การสอบถามบริเวรใกล้เคียง การติดป้ายประกาศ หรือการโพสต์ตามหาบนสื่อสังคมออนไลน์เช่นเดียวกัน ซึ่งสาเหตุส่วนหนึ่งเกิดจากผู้ที่พบเจอไม่สามารถแยกได้ว่าเป็นสัตว์ที่พบนั้นเป็นสัตว์เลี้ยงที่พลัดหลงหรือเป็นสัตว์จรจัด ซึ่งการระบุตัวตนสัตว์เลี้ยงได้ก็จะทำให้สามารถระบุตัวตนของผู้ดูแลได้

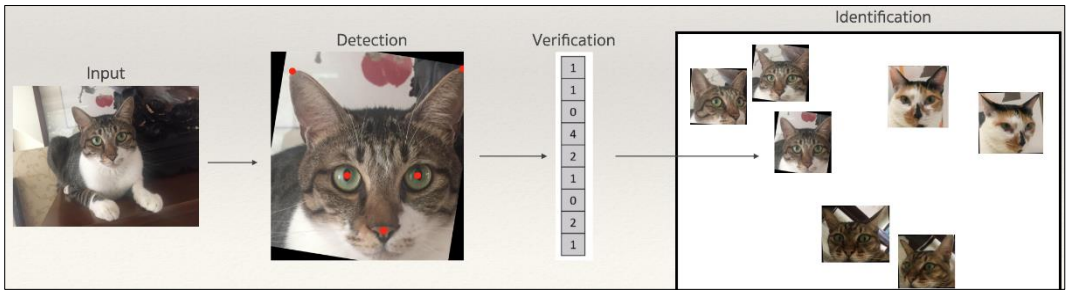
ปัจจุบันวิธีการระบุตัวตนของสัตว์เลี้ยงนั้นสามารถทำได้โดยใช้วัตถุเพื่อระบุตัวตน โดยจำแนกออกเป็น 3 ประเภทได้แก่ การระบุตัวตนแบบ, แบบกึ่งถาวร, แบบถาวร ซึ่งแต่ละวิธีนั้นก็มีข้อดีและข้อเสียที่แตกต่างกัน งานวิจัยชิ้นนี้จึงเสนอวิธีการระบุตัวตนแบบชีวมาตรจะสามารถช่วยเพิ่มทางเลือกให้ผู้ที่เลี้ยงสัตว์เลี้ยงได้

แม้การระบุตัวตนแบบชีวิตมาตรจะเป็นที่ประสบความสำเร็จในมนุษย์แต่ก็ยังมีข้อจำกัดในสัตว์เลี้ยงเช่น จำนวนข้อมูลที่พร้อมใช้งานได้บนอินเตอร์เน็ตนั้นมีน้อย การเลือกใช้คุณลักษณะที่จะนำมาใช้ระบุตัวตน สภาพแวดล้อมที่แตกต่างกันเช่น แสง สี ท่าทาง ซึ่งยังทำให้เกิดการระบุตัวตนผิดพลาด ทำให้เกิดงานวิจัยที่นำเสนอวิธีต่างๆที่จะนำมาใช้เพื่อระบุตัวตน ซึ่งถือว่าเป็นความท้าทายในศึกษาวิธีการจะใช้ระบุตัวตนของสัตว์ด้วยชีวมาตร

3.1.3 ออกแบบระบบ

การออกแบบและพัฒนาระบบรู้จำใบหน้าสัตว์เลี้ยงโดยใช้การเรียนรู้เชิงลึก โดยจากการทบทวนวรรณกรรมพบว่าขั้นตอนในการรู้จำใบหน้าประกอบไปด้วย 3 ขั้นตอน ได้แก่ 1) การตรวจจับใบหน้า ค้นหาใบหน้าของแมวภายในภาพซึ่งมีความท้าทายคือจำนวนใบหน้าของแมวในภาพด้วยสภาพแวดล้อมที่แตกต่างกัน ระยะห่าง แสงและเงา ที่อาจเป็นปัจจัยที่ทำให้ตรวจจับใบหน้าไม่พบ รวมถึง 2) การเปรียบเทียบใบหน้า เป็นกระบวนการที่ต้องใช้การสกัดคุณลักษณะโดยใช้การฝึกฝนแบบจำลองให้สามารถแยกคุณลักษณะของแมวแต่ละด้วย โดยให้ค่าของเวกเตอร์นั้นมีการเกาะตัวในคลาสสเดียวกัน และกระจายจากคลาสอื่น ๆ มากที่สุด 3) การระบุตัวตน เมื่อสามารถสร้างชุดเวกเตอร์ของข้อมูลได้แล้ว จะต้องใช้กระบวนการในการสร้างพื้นที่การฝังตัวของข้อมูล(Embedding Space) เพื่อใช้สำหรับการนำข้อมูลที่มีในฐานข้อมูลไปเปรียบเทียบค่าระยะทางกับใบหน้าเป้าหมายที่ต้องการระบุตัวตน

ภาพตัวอย่างขั้นตอนการรู้จำใบหน้าแมวดังรูป



รูปที่ 3.1 ภาพตัวอย่างขั้นตอนการรู้จำใบหน้าแมว

ทั้งนี้หลังจากสกัดคุณลักษณะใบหน้าแมวได้แล้ว ใบหน้าของแมวจะถูกเปรียบเทียบโดยการนำเวกเตอร์มาจัดวางใน embedding space เผื่อใช้ในการเปรียบเทียบระยะด้วยวิธีการ KNN เพื่อระบุตัวตนแมว

3.1.4 การพัฒนาระบบ

1) พัฒนาผ่านคลาวด์แพลตฟอร์ม Google Colab Pro+
2) ภาษาและเฟรมเวิร์คโปรแกรมป ระยุกต์สำหรับใช้พัฒนา
   - ไพธอน (Python)
   - PyTorch
   - TensorFlow
   - Keras

3.1.5 การดูแลรักษาระบบ

การดูแลรักษาและปรับปรุงระบบรู้จำใบหน้าสัตว์เลี้ยงโดยใช้การเรียนรู้เชิงลึกจะต้องทำการสร้างดัชนีข้อมูลใหม่ด้วย FAISS ใหม่เสมอเมื่อมีการนำเข้าข้อมูลภาพใบหน้าของแมวตัวใหม่ๆมาเก็บไว้ในระบบ นอกจากนี้สามารถฝึกให้แบบจำลองที่ใช้งานเรียนรู้จากข้อมูลเดิมที่มีอยู่ทำให้มีความแม่นยำมากกว่าเดิม หรือสามารถนำไปปรับใช้โดยใช้การส่งการเรียนรู้ (Transfer learning) เพื่อที่จะนำแบบจำลองเดิมไปประยุกต์ใช้กับข้อมูลที่เปลี่ยนไปหรือข้อมูลประเภทอื่น

3.2 ชุดข้อมูลที่ใช้ในงานวิจัย

ชุดข้อมูลในงานวิจัยชิ้นนี้จะใช้ข้อมูล 2 ชุด ชุดแรกจะนำมาจากแคกเกิล (Kaggle) ซึ่งประกอบไปด้วยข้อมูลใบหน้าของแมว 9,993 รูปภาพ โดยแบ่งเป็น 7,284 ภาพสำหรับการฝึกฝน และ 2,709 ภาพสำหรับการทดสอบ ตัวอย่างข้อมูลดังรูปที่ 3.2 เพื่อใช้สำหรับการฝึกแบบจำลองให้สามารถตรวจจับใบหน้าของแมวได้ สำหรับข้อมูลที่ใช้สำหรับการฝึกเพื่อระบุตัวตนของแมวภายในงานวิจัย ชุดข้อมูลชุดที่สองผู้วิจัยได้ใช้ข้อมูลใบหน้าของแมวจากในงานวิจัยของ Lin และ Kou ที่ได้รวบรวมใบหน้าแมวกว่า 519 ตัว จากสถานรับเลี้ยงสัตว์ซึ่งแมวแต่ละตัวจะมีภาพไม่ต่ำกว่า 10 ภาพ มุมที่แตกต่าง โดยจาก 13,536 ภาพ เมื่อคัดแยกรูปภาพที่มีความละเอียดต่ำ, มีภาพน้อยกว่า 5 รูป, ภาพซ้ำ, ภาพเบลอ และไม่สมบูรณ์อื่นๆ มีภาพแมวคงเหลือ 396 ตัว โดยแบ่งเป็นชุดสำหรับฝึกฝนจำนวน 328 ตัว 8,125 ภาพ ชุดสำหรับการตรวจสอบ 68 ตัว 2,258 ภาพ เพื่อนำมาใช้ฝึกและทดสอบแบบจำลอง เพื่อให้สามารถรู้จำใบหน้าแมวแต่ละตัวได้และทดสอบความแม่นยำในการทำนายผล

รูปที่ 3.2 ตัวอย่างข้อมูลภาพแมวจาก Kaggle

รูปที่ 3.3 ภาพตัวอย่างใบหน้าแมวที่ถูกนำเสนอในงานของ Lin และ Kou

3.3 กระบวนการตรวจจับใบหน้า

กระบวนตรวจจับใบหน้าในงานวิจัยชิ้นนี้เริ่มแรกได้ใช้การตรวจจับใบหน้าโดยใช้ pre-trained โมเดล โดยวิธีการแรกใช้ Haar Cascade Classifier และ FHOG-based SVM ผลลัพธ์นั้นมีความผิดพลาดสูง จึงได้เปลี่ยนทดลองกับวิธีการ Deep Learning โดยใช้ CNN ซึ่งมีความแม่นยำกว่าสองวิธีการแรก แต่มีข้อจำกัดคือใช้เวลาประมวลผลนานและสามารถค้นหาได้เพียง 1 ใบหน้าต่อภาพ ผู้วิจัยจึงได้ปรับเปลี่ยนการพัฒนาเพื่อใช้งาน Faster R-CNN เช่นเดียวกับงานวิจัยของ Lin และ Kou ซึ่งมีข้อจำกัดที่สถาปัตยกรรมถูกออกแบบมาเพื่อให้ระบุ Bounding Box ของใบหน้าเท่านั้น หากต้องการระบุจุดสำคัญบนใบหน้าจะต้องพัฒนาแบบจำลองเพิ่มสำหรับการประมวลผลซึ่งจะต้องใช้เวลาและทรัพยากรในการประมวลผลเพิ่มขึ้น

ผู้วิจัยจึงเลือกใช้สถาปัตยกรรมตระกูล YOLO ที่มีจุดเด่นในเรื่องความเร็วและความแม่นยำแต่เช่นเดียวกับ Faster R-CNN แบบจำลองได้ถูกออกแบบมาเพื่อใช้สำหรับการค้นหาบริเวณที่น่าจะมีวัตถุและสร้าง Bounding Box ขึ้นมารอบวัตถุเท่านั้น ผู้วิจัยจึงเลือกใช้ YOLOv5-Face ในเวอร์ชันย่อยที่ได้รับการปรับปรุงให้สามารถใช้การระบุจุดสำคัญบนใบหน้าได้

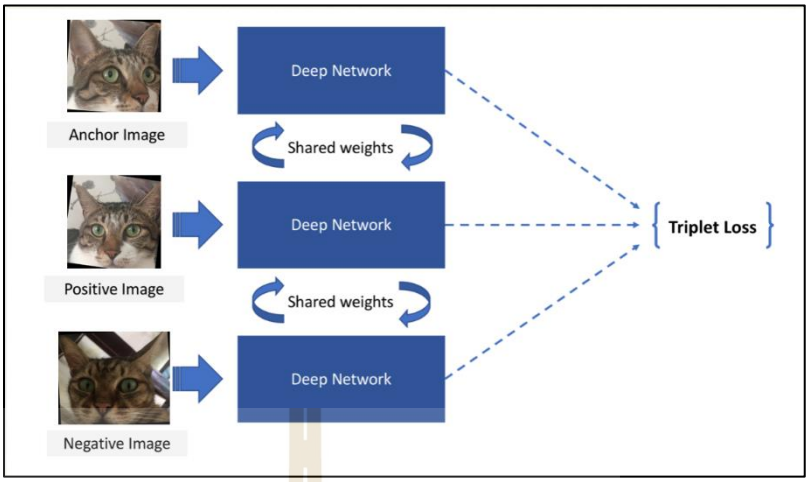
รูปที่ 3.4 ภาพตัวอย่างผลลัพธ์การระบุตำแหน่งใบหน้าแมวและการปรับแต่งภาพ

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.4 กระบวนการตรวจสอบใบหน้า

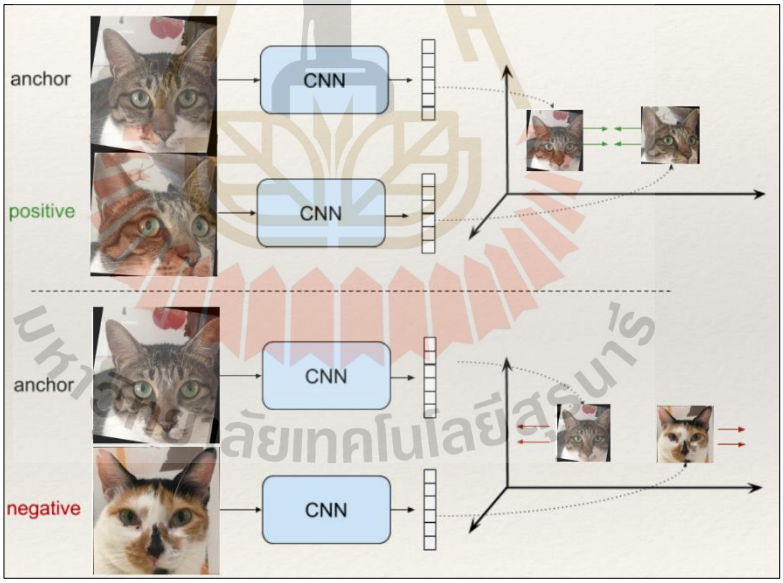
แบบจำลองที่ใช้ในขั้นตอนตรวจสอบใบหน้านั้นจะพัฒนาในรูปแบบโครงข่ายประสาทเทียมแบบแฝดสาม (Triplet Neural Network) ซึ่งเป็นโมเดลสามารถรับข้อมูลนำเข้าทั้งหมด 3 ชุด และมีการแชร์น้ำหนัก (Weight) กันภายในแบบจำลอง เพื่อให้ภาพเดียวกันได้ผลลัพธ์ออกมาเหมือนกัน

โดยในกระบวนการนี้ งานวิจัยชิ้นนี้ใช้การถ่ายโอนความรู้ (Transfer Learning) เข้ามาช่วยเพื่อให้ประหยัดเวลาและทรัพยากรในการพัฒนา โดยแบบจำลองที่เลือกใช้จะเป็นแบบจำลอง EfficientNetV2S ซึ่งโมเดลที่ถูกฝึกฝนไว้ล่วงหน้า (Pre-trained Model) โดยจะไม่ใช้เลเยอร์ชั้นสุดท้าย แต่จะใช้เพียงเลเยอร์รองสุดท้ายเพราะต้องการผลลัพธ์ข้อมูลในลักษณะของเวกเตอร์ โดยจะเชื่อมเลเยอร์ต่อท้ายเพื่อปรับผลลัพธ์ให้เป็นเวกเตอร์ขนาด 128 จากนั้นผลลัพธ์ที่จะถูกใช้ฟังก์ชัน L2 เพื่อ Normalize ข้อมูล



รูปที่ 3.5 ภาพตัวอย่างการฝึกฝนใบหน้าแมวด้วย Triplet Network

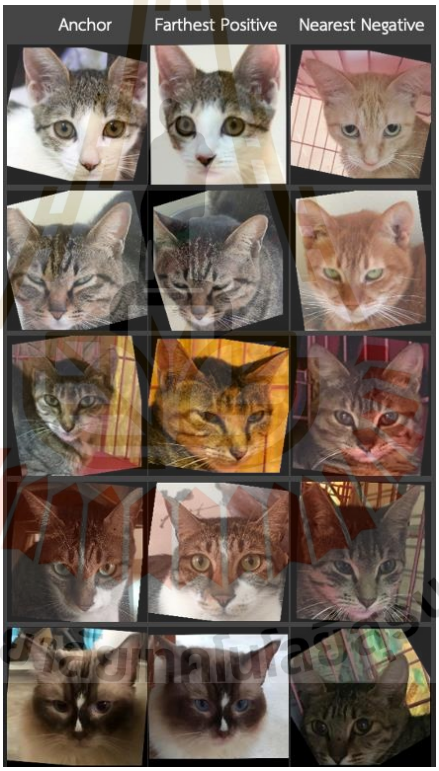
และเมื่อได้ผลลัพธ์แบบเวกเตอร์ก็จะถูกนำไปใช้ในฟังก์ชัน Triplet Loss เพื่อภาพใบหน้าของแมวตัวเดียวกันอยู่ในกลุ่มเดียวกัน และแมวที่แตกต่างกันจะกระจายตัวจากกัน



รูปที่ 3.6 ภาพแสดงกระบวนสกัดคุณลักษณะและระบุใบหน้าแมว

อย่างไรก็ตามการใช้ Triplet Neural Network และ Triplet Loss ยังไม่สามารถสร้างการกระจายตัวของข้อมูลได้ดีเพียงพอ เพราะบางครั้งลักษณะใบหน้าแมวในคลาสอื่น ก็อาจจะมีค่าระยะทางที่ใกล้กว่าใบหน้าแมวในคลาสเป้าหมาย จึงได้มีนำเสนอวิธีการในการที่จะช่วยเลือกข้อมูลและการปรับค่า Loss ของข้อมูลเพื่อให้แบบจำลองเรียนรู้คุณลักษณะที่เหมือนและแตกต่างกันมากขึ้น

โดยกระบวนการ Hard Batch ที่ถูกนำเสนอโดย Hermans et al คือการทำ Triplet Mining ซึ่งจะช่วยทำให้แบบจำลองได้เรียนรู้ภาพของคลาสที่มีความใกล้เคียงกัน โดยเมื่อนำ Triplet Mining แบบ Hard Batch มาใช้ ในกระบวนคัดเลือกรูปภาพแต่ละรอบจะทำการสุ่มคลาสจากจำนวน BatchSize/4 คลาส จากนั้นหนึ่งคลาสจากที่สุ่มมาจะถูกเลือกเพื่อเป็นภาพ Anchor และภาพ Positive คลาสที่เหลือจะถือว่าเป็นภาพ Negative ในการเลือกภาพ จะทำการคำนวณระยะทางด้วยวิธีการ Euclidian Distance ระหว่างภาพ Anchor กับทุกภาพที่เหลือในคลาสและจะเลือกภาพที่มีระยะทางสูงสุด (Farthest Positive) เพื่อใช้หาภาพแมวตัวเดียวกันที่มีความแตกต่างกันมากที่สุด จากนั้นคลาสที่เหลือที่เป็น Negative จะถูกสุ่มเลือกมาและหาภาพที่มีความคล้ายกับภาพ Anchor ที่สุดโดยการคำนวณหาระยะทางใกล้ที่สุด (Nearest Negative) จากนั้นค่าจะนำค่าที่ได้ไปคำนวนและปรับค่า Loss ทำให้ข้อมูลมีการกระจายตัวที่ดียิ่งขึ้น



รูปที่ 3.7 ภาพตัวอย่างการเปรียบเทียบระหว่าง Hard Triplet

นอกจากนี้ภายใน Triplet Loss Function ได้มีการนำ GOR ที่ถูกเสนอโดย Zhang et. al. มาใช้เพื่อทำ Regularization โดยจะเพิ่มโทษให้คู่ข้อมูล Anchor, Negative ให้มีระยะห่างออกจากกันมากขึ้น เพื่อเพิ่มการกระจายตัวของข้อมูลเพื่อไม่ให้เกิดการ over-fitting

3.5 กระบวนการระบุตัวตน

เมื่อได้ผลลัพธ์เป็นคุณลักษณะของใบหน้าในรูปแบบเวกเตอร์แล้ว ในขั้นการระบุตัวตนจะใช้ Facebook AI Similarity Search (FAISS) เพื่อค้นหาตามความคล้ายคลึงกัน ถูกนำมาใช้ในงานวิจัยชิ้นนี้ ด้วยจุดเด่นคือด้านความเร็วและความสามารถในการจัดกลุ่มของข้อมูลในรูปแบบ Vector โดยในงานวิจัยชิ้นนี้ จะแบบจำลองในขั้นตอน Verification สร้างเวกเตอร์จากใบหน้าแมวในฐานข้อมูล และใช้ FAISS เพื่อสร้าง Index ของใบหน้าแมวเก็บไว้ใน RAM ในการเรียกใช้งาน เมื่อมีการส่งใบหน้าแมวเข้ามาใหม่ จะประมวลผลด้วยแบบจำลอง ก่อนจะนำเวกเตอร์ที่ได้มาเปรียบเทียบกับใบหน้าที่ถูกทำ Index ไว้ใน RAM ด้วยวิธีการ KNN ซึ่งจะถูกใช้ในการทำนายใบหน้าจำนวน M รูป ซึ่งค่า K ถูกใช้ในกระบวนการทดสอบ 1,3,4 และ 5

3.6 กระบวนการวัดผล

ในงานวิจัยชิ้นนี้จะวัดประสิทธิภาพของระบบโดยวิเคราะห์จากผลการตรวจจับใบหน้า และผลการทำนายของแบบจำลองว่าสามารถทำได้ถูกต้องหรือไม่ ใช้การวัดประสิทธิภาพของแบบจำลอง (Two-class prediction) ที่แสดงในรูปที่ 3.8 ซึ่งผลลัพธ์ที่ได้จะเป็นคลาส Positive หรือ Negative เพื่อนำไปหาค่าความแม่นยำของแบบจำลองโดยจะหาได้จากค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) โดยใช้สัญลักษณ์ดังนี้

1) ค่าทีพี (True Positive (TP)) คือ สิ่งที่ทาย (prediction) ว่ามันมีจริงและเฉลย (Ground truth) บอกว่ามีจริง

2) ค่าทีเอ็น (True Negative (TN)) คือ สิ่งที่ทายว่ามันไม่มีจริง และเฉลยก็บอกว่ามันไม่มีจริง

3) ค่าเอฟพี (False Positive (FP)) คือ สิ่งที่ทายว่ามันมีจริง แต่เฉลยก็บอกว่ามันไม่จริง

4) ค่าเอฟเอ็น (False Negative (FN)) คือ สิ่งที่ทายว่ามันมีไม่จริง แต่เฉลยก็บอกว่ามันมีจริง

		Predicted classes	
		Yes	No
Actual class	Yes	True Positive	False Negative
	No	False Positive	True Negative

รูปที่ 3.8 ตารางการวัดประสิทธิภาพของแบบจำลอง

ค่าความแม่นยำ (Precision) คือ อัตราการทายถูกต่อการทายทั้งหมด โดย n คือจำนวนที่แบบจำลองพยายามทายทั้งหมด เช่น ทายมา 2 ตัว โอกาสก็มีถูกทั้งหมด ก็ ทีพี (TP) = 2 หรือ ทายถูก 1 ทาย ผิด 1, ทีพี (TP) =1 เอฟพี (FP) =1 หรือ ทายผิดหมดเลย ก็เป็นได้

$$\text{precision} = \frac{tp}{tp + fp} = \frac{tp}{n}$$

ส่วนความระลึก (Recall)l คือ จำนวน ที่ทายถูกต่อจำนวนที่มีจริง (Ground Truth) ทั้งหมด

$$\text{recall} = \frac{tp}{tp + fn}$$

ซึ่งค่าความแม่นยำเฉลี่ย (AP: Average Precision) คำนวนได้จริงพื้นที่ใต้กราฟของค่าความแม่นยำ (Precision) กับค่าความระลึก (Recall) โดยจะมีค่าอยู่ระหว่าง 0 ถึง 1 ถ้าเป็น 1 ซึ่งการหาพื้นที่ใต้กราฟจะใช้สมการดังนี้

$$\text{AP} = \sum_{k=1}^{N} p(k)\Delta r(k)$$

โดยที่ N คือ จำนวน รูปทั้งหมดที่เราทายออกมาทั้งถูกและไม่ถูก ส่วน P(k) คือ ค่าความแม่นยำ (Precision) ที่ตำแหน่ง ที่เกินขีดแบ่ง (threshold) ที่ค่า k นั้น ๆ ส่วน delta r(k) คือ การเปลี่ยนแปลงของ ค่าความระลึก (recall) ระหว่าง k-1 และ k ซึ่งเป็นการคำนวนหาค่าความแม่นยำเฉลี่ย (AP) นี้ จะหาเฉพาะ เพียง 1 คลาส เท่านั้น ซึ่งหลังจากเราหาค่าในทุก ๆ คลาส แล้วจากนั้น นำค่าเอพีมาทำการเฉลี่ย จึงจะได้เป็นค่าเอ็มเอพี (mean Average Precision : mAP) ของแบบจำลองซึ่งแทนได้ดังสมการนี้

$$mAP = \frac{AP}{no.of\ queries}$$

ซึ่งงานวิจัยชิ้นนี้จะนำค่าเอ็มพีจากค้นหาใบหน้าและระบุตัวตนของแมวได้อย่างถูกได้ เมื่ออยู่ในสภาพแวดล้อมร่วมกับวัตถุอื่น ๆ มาใช้วัดความแม่นยำของแบบจำลอง เพื่อให้บรรลุตามสมมุติฐานที่ตั้งไว้

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4
การทดสอบและอภิปรายผล

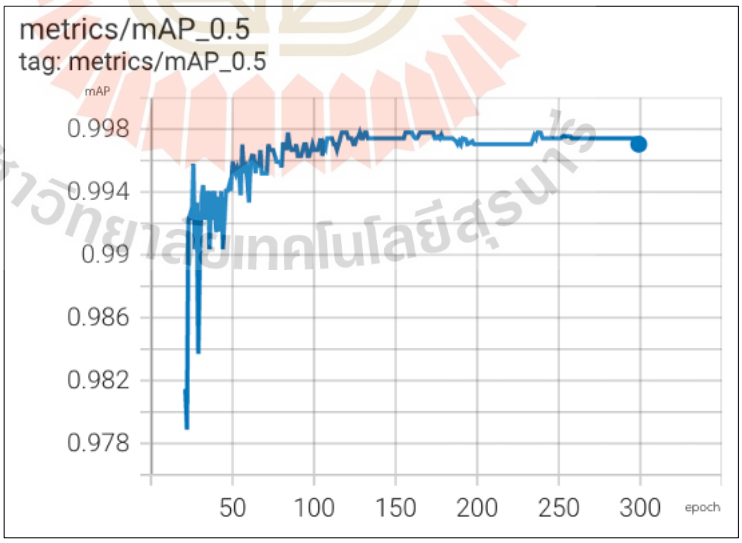
ในการทดสอบประสิทธิภาพของแบบจำลองจะแบ่งออกเป็น 2 ส่วนคือ การวัดประสิทธิภาพการตรวจจับใบหน้า และวัดประสิทธิภาพของการระบุตัวตน

โดยในส่วนของการตรวจใบหน้าจะมุ่งเน้นการตรวจหาใบหน้าแมวในภาพที่มีสภาพแวดล้อมที่ตรวจจับได้ยาก โดยต้องสามารถตรวจจับได้หลายใบหน้าพร้อมกันและสามารถที่จะระบุจุดสำคัญบนใบหน้าได้ ได้แก่ ปลายหูซ้าย ปลายหูขวา ตาซ้าย ตาขวา และจมูก เพื่อให้สามารถที่จะนำภาพไปจัดเตรียมและประมวลผลในการตรวจจับได้ใบหน้าได้

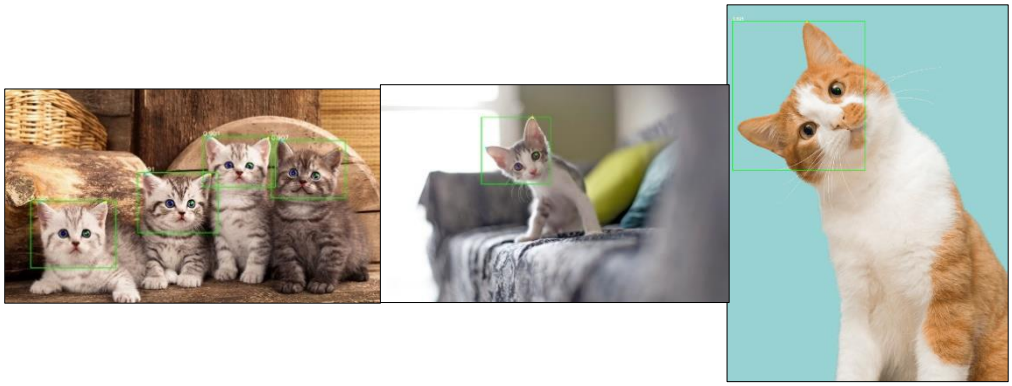
การตรวจการระบุตัวตนจะแสดงผลลัพธ์ของการระบุตัวตนบนใบหน้าต่างๆ โดยจะมุ่งเน้นการปรับแต่งค่า Hyper Parameter เพื่อให้ได้ผลลัพธ์ในการตรวจจับใบหน้าสูงสุด

4.1 การทดสอบการตรวจจับใบหน้า

แบบจำลองสำหรับการตรวจจับใบหน้าได้ถูกฝึกฝนเป็นจำนวน 300 epoch ซึ่งมีค่า mAP อยู่ที่ 0.9974 โดยข้อมูลใบหน้าแมวจากหลากหลายขนาดถูกนำเข้า เพื่อตรวจจับใบหน้าโดยมีค่า Threshold 0.5 พบว่าใบหน้าแมวสามารถตรวจจับได้แม้อยู่ในต่างมุม สภาพแวดล้อมหรือมีการบดบัง นอกจากนี้ยังสามาถระบุจุดสำคัญบนใบหน้าของแมวแต่ละตัวได้



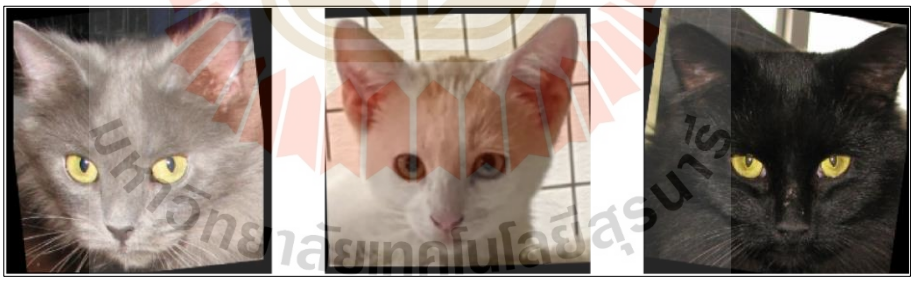
รูปที่ 4.1 กราฟแสดงผล mAP จากการฝึกฝนแบบจำลองสำหรับการตรวจจับใบหน้าแมว



รูปที่ 4.2 ภาพตัวอย่างการตรวจจับใบหน้าแมวโดยการระบุจุดสำคัญบนใบหน้า

4.2 ชุดข้อมูลที่ใช้ในการทดสอบการระบุตัวตน

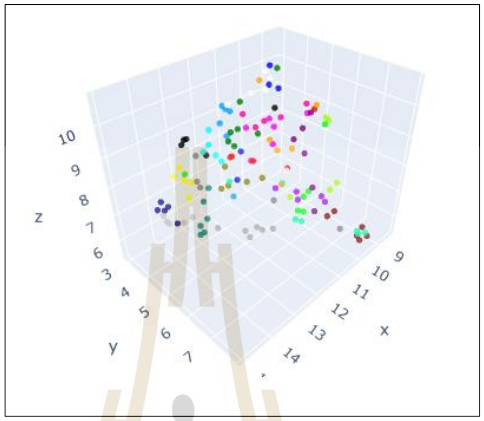
ในการทดสอบการระบุตัวตน ใช้ข้อมูลชุดทดสอบจากงานวิจัยของ Lin และ Kou โดยเป็นภาพใบหน้าของแมว 69 ตัว จำนวน 2,258 ภาพ ประกอบไปด้วยภาพของแมวที่มีทั้งลักษณะที่คล้ายกันและแตกต่างกัน ใบหน้าของแมวถูกนำเข้าในแบบจำลองตรวจจับใบหน้าเพื่อให้ได้เฉพาะส่วนใบหน้าของแมว และถูกตำแหน่งของใบหน้าให้อยู่ในลักษณะหน้าตรง โดยให้ดวงตาอยู่ในระนาบเดียวกันเพื่อให้ข้อมูลใบหน้าของแมวอยู่ในลักษณะเดียวกัน ตำแหน่งภาพที่หลุดจากเฟรมใบหน้าแมวจะถูกแทนที่ด้วยสีดำ จากนั้นปรับขนาดภาพมีขนาด 224x224 เพื่อใช้ในการประมวลผลต่อไป



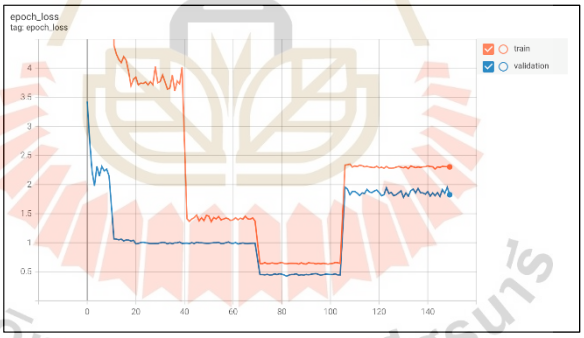
รูปที่ 4.3 ตัวอย่างผลลัพธ์ใบหน้าแมวในชุดข้อมูลสำหรับการระบุตัวตน

4.3 การทดสอบการระบุตัวตน

กระบวนการทดสอบการรู้จำใบหน้า ภาพใบหน้าแมวทั้งหมดจะถูกนำเข้าและประมวลผลด้วย Verification Model และได้ผลลัพธ์ออกเป็นเวกเตอร์ของข้อมูลขนาด 128 มิติ โดยก่อนการประมวลผลภาพใบหน้าแมวที่ต้องการค้นหาจะถูกสุ่มจากชุดข้อมูล และใบหน้าที่เหลือจะถูกนำไปประมวลผลและสร้างฐานข้อมูลด้วย FAISS เพื่อใช้สำหรับการทำนายคลาสที่ถูกต้องของข้อมูล



รูปที่ 4.4 การกระจายตัวของข้อมูลใบหน้าแมวที่ถูกประมวลผลและลดมิติให้อยู่ในลักษณะ 3 มิติ

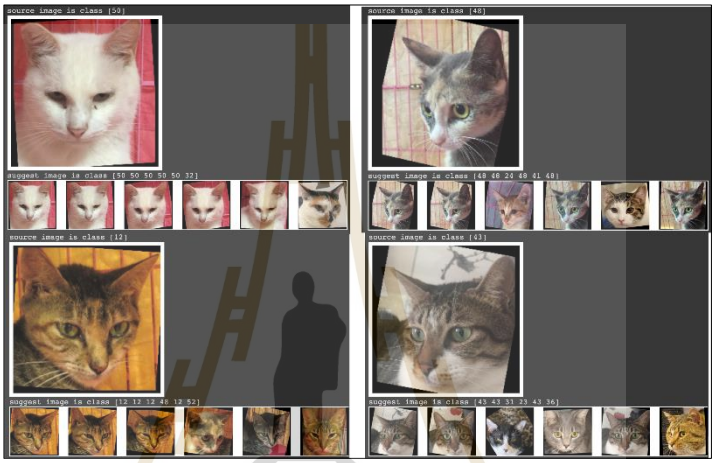


รูปที่ 4.5 ภาพกราฟแสดงค่า Loss ในการฝึกฝนชุดข้อมูลในการจำแนกใบหน้า

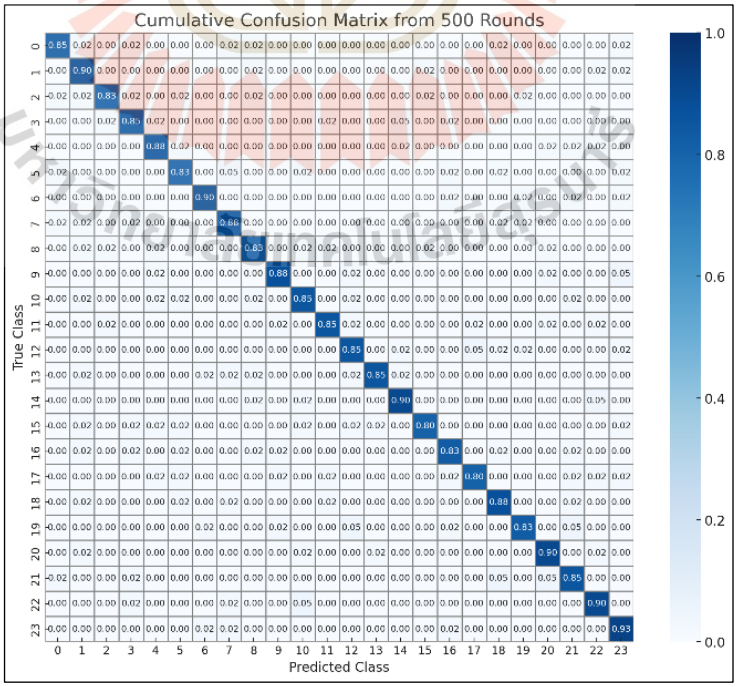
ในการวัดผลงานวิจัยชิ้นนี้ใช้วิธีการเดียวกับกระบวนการของ Meugot et al ที่จะทดสอบในแต่ละรอบ จะทำการสุ่มคลาสของแมวออกมาจำนวน c คลาส และจำนวนรูปของแมวที่ถูกนำมาประมวลผลถูกแทนด้วย Nc และให้จำนวน M แทนค่าจำนวนใบหน้าของแมวต้องการค้นหา เช่น M=2 จะเป็นการนำใบหน้าของแมวที่ต้องการค้นหา 2 รูปโดยเชื่อมระหว่างรูปทั้งสองเวกเตอร์ให้กลายเป็นเวกเตอร์เดียว มาใช้ค้นหาข้อมูลที่เหลือในฐานข้อมูล และ k แทนค่าเพื่อนบ้านใน Nearest Neighbor เป็นจำนวนใบหน้าใกล้เคียงที่ต้องการค้นหา ซึ่งในการทดสอบจะทำซ้ำทั้งหมดจำนวน 500 รอบ โดยแต่ละรอบจะสุ่มคลาสจากชุดข้อมูลและวัดจากค่าความถูกต้องใน 100 ครั้ง และนำเสนอค่าความแม่นยำเฉลี่ย ค่าสูงสุด ค่าต่ำสุด โดยที่ M=1 และ K=1 จะเป็นตัวแทนของการเรียนรู้แบบ one-shot และสำหรับค่าอื่นๆค่า k=M+1 เสมอ

ตารางที่ 4.1 ตารางแสดงค่าความแม่นยำของแบบจำลองเมื่อปรับ M และ k

		Identification Accuracy		
M	k	Average	Min	Max
1	1	75.9%	71%	80%
2	3	85.2%	78%	89%
3	4	86.8%	84%	89%
4	5	82.8%	78%	87%



รูปที่ 4.6 ภาพตัวอย่างการทำนายเมื่อ M=1 และ K=5



รูปที่ 4.7 ภาพผลลัพธ์ค่า Confusion Matrix ขนาด 24x24 ของแมวจำนวน 24 คลาส

Confusion Matrix แสดงผลลัพธ์การทำนายของทั้ง 24 คลาส โดยแต่ละแถวของเมทริกซ์แทนคลาสจริง ในขณะที่แต่ละคอลัมน์แทนคลาสที่ทำนาย ค่าที่ตำแหน่ง (i, j) ในเมทริกซ์แสดงถึงจำนวนตัวอย่างที่เป็นคลาส i และทำนายว่าเป็นคลาส j หาก i=1 และ j=1 คือค่าที่ตำแหน่ง (1,1) ในเมทริกซ์ ค่านี้แสดงถึงจำนวนของ True Positives (TP) สำหรับคลาสที่ 1 หมายความว่า โมเดลทำนายว่าข้อมูลเป็นคลาสที่ 1 และความจริงคือตัวอย่างเหล่านั้นอยู่ในคลาสที่ 1 เช่นกัน ยิ่งค่าที่ตำแหน่ง (1,1) มีค่าสูงเท่าไรก็แสดงว่าโมเดลทำนายคลาสที่ 1 อย่างแม่นยำมากขึ้น

4.4 อภิปรายผล

จากการทดสอบรู้จำใบหน้าโดยแยกออกเป็นสองส่วนคือกระบวนการตรวจจับใบหน้า และการบวนการระบุตัวตน ผลลัพธ์ของการตรวจจับใบหน้าออกมาเป็นที่น่าพึงพอใจโดยสามารถระบุใบหน้าได้แม้อยู่ในสภาพแวดล้อมที่แตกต่างและสามารถตรวจจับใบหน้าที่มีหลากหลายขนาดและหลากหลายใบหน้าได้ภายในภาพเดียวกัน แต่อย่างไรก็ตามด้วยค่า Threshold ที่ค่อนข้างต่ำอาจทำให้เกิดปัญหาการตรวจพบใบหน้าที่ผิดพลาดได้

ในส่วนของกระบวนการระบุตัวตนผลลัพธ์สามารถระบุตัวตนได้แต่ขั้น one-shot recognition แต่ค่าความแม่นยำยังต่ำกว่าวัตถุประสงค์ในการศึกษาครั้งนี้ และเมื่อเพิ่มจำนวนภาพที่ใช้สำหรับการค้นก็ยิ่งเพิ่มความแม่นยำมากขึ้น แต่เมื่อค่า M=4 และ K=5 ค่าความแม่นยำกลับน้อยลง อาจตีความได้ว่าเมื่อจำนวนภาพที่ใช้ค้นหามากขึ้นอาจทำให้เกิดความแปรปรวนของข้อมูลหรือข้อมูลในฐานข้อมูลนั้นยังไม่มีการกระจายตัวที่ดีพอเพื่อต้องค้นหาในจำนวนที่มากขึ้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 5
สรุปและข้อเสนอแนะ

จากการศึกษาและพัฒนาแบบจำลองสำหรับการระบุตัวตนแมวด้วยใบหน้าโดยใช้การเรียนรู้เชิงลึก ผู้วิจัยข้อสรุปผลใน 3 หัวข้อดังนี้

5.1 ผลการวิจัย

งานวิจัยครั้งนี้ได้เสนอแบบจำลองในการรู้จำใบหน้าแมวโดยกระบวนการรู้จำใบหน้าถูกแบ่งออกเป็น 2 ส่วนคือส่วนของแบบจำลองที่ใช้ในการรู้จำใบหน้าและแบบจำลองสกัดคุณลักษณะเพื่อเปรียบเทียบใบหน้าและนำมาระบุตัวตน

โดยในส่วนของการตรวจใบหน้าจะมุ่งเน้นการตรวจหาใบหน้าแมวในภาพที่มีสภาพแวดล้อมที่ตรวจจับได้ยาก โดยต้องสามารถตรวจจับได้หลายใบหน้าพร้อมกัน และสามารถที่จะระบุจุดสำคัญบนใบหน้าได้ ได้แก่ ปลายหูซ้าย ปลายหูขวา ตาซ้าย ตาขวา และจมูก เพื่อให้สามารถที่จะนำภาพไปจัดเตรียมและประมวลผลในการตรวจจับได้ใบหน้าได้

ในส่วนของกระบวนการระบุตัวตนนั้น แบบจำลองสามารถสกัดคุณลักษณะและระบุตัวตนได้ตั้งแต่ขั้น one-shot recognition แต่ค่าความแม่นยำนั้นยังต่ำกว่าจะนำไปประยุกต์ใช้จริง และเมื่อเพิ่มจำนวนภาพที่ใช้สำหรับการค้นก็ยิ่งเพิ่มความแม่นยำมากขึ้น

โดยในระหว่างการวิจัยผู้วิจัยพบว่าด้วยเทคโนโลยีการระบุสิ่งของหรือการตรวจจับใบหน้ามนุษย์ประสบความสำเร็จค่อนข้างมากในการใช้งานทำให้แบบจำลองมีประสิทธิภาพได้ถึง 86.8% การพัฒนาแบบจำลองสำหรับตรวจจับใบหน้าแมวในสามารถใช้การถ่ายโอนความรู้ได้โดยประหยัดทรัพยากรและมีความแม่นยำสูง

5.2 การประยุกต์ผลการวิจัย

แบบจำลองการรู้จำใบหน้าแมวสามารถนำไปประยุกต์ใช้งานได้หลากหลายเช่น ในหัวข้อเมืองอัจฉริยะสามารถนำแบบจำลองนี้ไปใช้ในการพัฒนาระบบกล้องวงจรปิดที่สามารถติดตามสัตว์เลี้ยงได้ หากมีการพลัดหลงไปยังพื้นสาธารณะ ระบบขึ้นทะเบียนสัตว์เลี้ยงที่จะสามารถเก็บข้อมูลของเจ้าของและสัตว์เลี้ยงได้ และยังสามารถนำไปใช้ในการเก็บข้อมูลทางการแพทย์ของสัตว์ได้อีกด้วย

นอกจากการประยุกต์ใช้ทางตรงแล้ว งานวิจัยชิ้นนี้ยังสามารถนำไปประยุกต์ใช้ทางอ้อมได้ใช้การศึกษาวิธีการคุณลักษณะสัตว์เลี้ยงหรือสัตว์ป่า การนำแบบจำลองไปใช้ในลักษณะการถ่ายโอนความรู้ในการรู้จำสัตว์ชนิดอื่นๆ โดยเฉพาะในหัวข้อการติดตามสัตว์ป่า การระบุตัวตนและติดตามได้ถือเป็นปัจจัยสำคัญอย่างหนึ่ง

5.3 ปัญหาและข้อเสนอแนะ

ในขั้นตอนการเปรียบเทียบใบหน้านั้นมีความแตกต่างจากกระบวนการตรวจจับใบหน้า นอกจากสถาปัตยกรรมในการฝึกฝนแล้ว ผู้วิจัยพบว่า ปัจจัยสำคัญที่ส่งผลต่อการพัฒนาคือการเลือกใช้ข้อมูล ด้วยกระบวนการระบุใบหน้ามีความละเอียดกว่าการตรวจจับใบหน้า ซึ่งใบหน้าแมวแม้จะอยู่ในเงื่อนไขที่แตกต่างกัน เช่น สภาพแสง ความเบลอ ตำแหน่ง แต่ก็ยังมีลักษณะร่วมที่เหมือนกัน ทำให้สามารถหาคุณลักษณะเพื่อตรวจจับได้ง่ายเพียงต้องมีข้อมูลที่เป็นตัวแทนที่ดีและมีจำนวนมากพอ แตกต่างการระบุตัวตนด้วยใบหน้าที่จะต้องมีรายละเอียดภายในภาพ และไม่ถูกบดบัง เช่น แบบจำลองอาจทำนายว่าแมวที่ลักษณะแตกต่างกัน แต่หันหน้าไปในทิศทางเดียวกันนั้นเป็นแมวตัวเดียวกันเพียงเพราะมีคุณลักษณะในการหันที่เหมือนกัน

ในส่วนของข้อเสนอแนะหรือแนวทางการทำงานวิจัยต่อไปนั้น ได้มีแนวทางในการพัฒนาดังนี้

1) การเพิ่มข้อกำหนดรูปแบบของข้อมูลใบหน้าแมว เช่นการเก็บภาพใบหน้าในลักษณะการหันรอบใบหน้าด้าน ซึ่งอาจใช้การสกัดคุณลักษณะจากวีดีโอที่นำมาแยกเฟรมได้

2) การใช้ Segmentation เพื่อตรวจจับใบหน้าโดยมองข้ามพื้นหลังเพื่อลดความคลาดเคลื่อนของข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี

รายการอ้างอิง

ณัฐโชติ พรหมฤทธิ์, สัจจาภรณ์ ไวจรรยา. (2564). **Fundamental of DEEP LEARNING in Practice**. กรุงเทพฯ: สำนักพิมพ์อินโฟเพรส.

ปริญญา สงวนสัตย์. (2562). **Artificial Intelligence with Machine Learning: AI สร้างได้ด้วยแมชชีนเลิร์นนิ่ง.** กรุงเทพฯ: สำนักพิมพ์อินโฟเพรส.

สำนักงานส่งเสริมเศรษฐกิจดิจิทัล. (2563). **Machine Learning: สิ่งใกล้ตัวแห่งโลกยุคใหม่.** (ออนไลน์). เข้าถึงได้จาก: https://www.depa.or.th/th/article-view/article11-2563

Adam Klein. (2019). **Pet Cat Face Verification and Identification.** Stanford University, CS230. (online). Retrieved from https://cs230.stanford.edu/projects_fall_2019/reports/26251543.pdf

Alaa Tharwat, Tarek Gaber, Aboul Ella Hassanien, Hassan A. Hassanien, & Mohamed F. Tolba. (2014). **Cattle Identification Using Muzzle Print Images Based on Texture Features Approach.** In Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA), Springer, pp. 217-227.

Ali Ismail Awad, Hossam M. Zawbaa, Hamdi A. Mahmoud, Eman Hany Hassan Abdel Nabi, Rabie Hassan Fayed, & Aboul Ella Hassanien. (2013). **A Robust Cattle Identification Scheme Using Muzzle Print Images.** In Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, pp. 529-534.

Andrew Ng. (2016). **Nuts and bolts of building AI applications using deep learning.** [Video file]. Retrieved from https://www.youtube.com/watch?v=F1ka6a13S9I

Bohan Yoon, Hyeonji So, & Jongtae Rhee. (2021). **A Methodology for Utilizing Vector Space to Improve the Performance of a Dog Face Identification Model.** Applied Sciences, 11(5), 2074.

Delong Qi, Weijun Tan, Qi Yao, & Jingfeng Liu. (2022). **YOLO5Face: Why Reinventing a Face Detector.** Computer Vision – ECCV 2022 Workshops, Springer, pp. 320-332.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). **ImageNet: A Large-Scale Hierarchical Image Database.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248-255.

Elad Hoffer, & Nir Ailon. (2015). **Deep Metric Learning Using Triplet Network.** In Proceedings of the Similarity-Based Pattern Recognition Workshop, SIMBAD 2015, Springer, pp. 84-92.

Enosh Shrestha. (2019). **Triplet Loss and Siamese Neural Networks.** (online). Retrieved from https://medium.com/@enoshshr/triplet-loss-and-siamese-neural-networks-5d363fdeba9b.

Ethem Alpaydin. (2010). **Introduction to Machine Learning (2nd ed.).** Cambridge, MA: MIT Press.

Florian Schroff, Dmitry Kalenichenko, & James Philbin. (2015**). FaceNet: A Unified Embedding for Face Recognition and Clustering.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823.

Guillaume Mougeot, Dewei Li, & Shuai Jia. (2019). **A Deep Learning Approach for Dog Face Verification and Recognition.** In Proceedings of the International Conference on Neural Information Processing (ICONIP 2019), Springer, pp. 357-366.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. (2016). **Deep Residual Learning for Image Recognition.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

Alexander Hermans, Lucas Beyer, & Bastian Leibe. (2017). **In Defense of the Triplet Loss for Person Re-Identification.** (online). Retrieved from https://arxiv.org/abs/1703.07737.

Ian Goodfellow, Yoshua Bengio, & Aaron Courville. (2021). **Deep Learning.** Cambridge, MA: MIT Press.

Jiongxin Liu, Angjoo Kanazawa, David Jacobs, & Peter Belhumeur. (2012). **Dog Breed Classification Using Part Localization.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1720-1727.

Joseph Redmon, Santosh Divvala, Ross Girshick, & Ali Farhadi. (2016). **You Only Look Once: Unified, Real-Time Object Detection.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788.

Kenneth Lai, Xinyuan Tu, & Svetlana Yanushkevich. (2019). **Dog Identification Using Soft Biometrics and Neural Network.** In Proceedings of the International Joint Conference on Biometrics (IJCB 2019), pp. 1-8.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, & Saining Xie. (2022). **A ConvNet for the 2020s.** (online). Retrieved from https://arxiv.org/abs/2201.03545.

Ross Girshick, Jeff Donahue, Trevor Darrell, & Jitendra Malik. (2013). **Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.** In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587.

Rohith Gandhi. (2018). **R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms.** (online). Retrieved from https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

Santosh Kumar & Sanjay Kumar Singh. (2014). Biometric Recognition for Pet Animal. **Journal of Software Engineering and Applications**, 7, 470-482.

Santosh Kumar & Sanjay Kumar Singh. (2018). **Monitoring of Pet Animal in Smart Cities Using Animal Biometrics.** Future Generation Computer Systems, 83, 553-563.

Santosh Kumar, Amit Pandey, K. Sai Ram Satwik, Sunil Kumar, Sanjay Kumar Singh, & Amit Kumar Singh. (2018**). Deep Learning Framework for Recognition of Cattle Using Muzzle Point Image Pattern.** Measurement, 116, 305-315.

Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. (2015). **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.** In Proceedings of the IEEE Conference on Neural Information Processing Systems (NeurIPS 2015), pp. 91-99.

Mingxing Tan & Quoc V. Le. (2021). **EfficientNetV2: Smaller Models and Faster Training.** In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), PMLR 139:10096-10106.

Thierry Pinheiro Moreira, Mauricio Lisboa Perez, Rafael de Oliveira Werneck, & Eduardo Valle. (2016). **Where Is My Puppy? Retrieving Lost Dogs by Facial Features.** Multimedia Tools and Applications, 76(14), 15325-15340.

Tzu-Yuan Lin & Yan-Fu Kuo. (2018). **Cat Face Recognition Using Deep Learning.** Presented at the 2018 ASABE Annual International Meeting, Detroit, Michigan, July 29 - August 1, 2018.

Xu Zhang, Felix X. Yu, Sanjiv Kumar, & Shih-Fu Chang. (2017). **Learning Spread-Out Local Feature Descriptors.** In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017), Venice, Italy, pp. 4595-4603.

Yu Fan, Chih-Chung Yang, & Chin-Ta Chen. (2021). **Cat Face Recognition Based on MFCC and GMM.** Presented at the 2021 6th International Conference on Image, Vision and Computing (ICIVC), Qingdao, China, July 23-25, 2021.

Yu-Chen Chen, Shintami C. Hidayati, Wen-Huang Cheng, Min-Chun Hu, & Kai-Lung Hua. (2016). **Locality Constrained Sparse Representation for Cat Recognition.** In Proceedings of the 22nd International Conference on MultiMedia Modeling (MMM 2016), Miami, USA, January 4-6, 2016.

Zhiwei Ruan, Guijin Wang, Jing-Hao Xue, Xinggang Lin, & Yong Jiang. (2014). **Detection of User-Registered Dog Faces.** In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, pp. 3632-3640.

ประวัติย่อ

ชื่อ – สกุล : นายภัทรพล ศรีรักษ์

ประวัติการศึกษา

ปี 2567	กำลังศึกษา มหาบัณฑิต สำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี
ปี 2558	หลักสูตรซอฟต์แวร์วิสาหกิจ สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี

ประสบการณ์การทำงาน:

- ผู้ช่วยสอนและวิจัย, มหาวิทยาลัยเทคโนโลยีสุรนารี, 2559 - 2566
- เจ้าหน้าที่วิเคราะห์ระบบคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี, 2566 - ปัจจุบัน

ทักษะและความสามารถ:

- การเขียนโปรแกรมภาษา Python, Java
- การวิเคราะห์ข้อมูล (Data Analysis)
- การพัฒนาเว็บแอปพลิเคชัน (Web Application Development)
- การใช้งานเครื่องมือ DevOps เช่น Jenkins, Docker

ความสนใจเฉพาะทาง:

- การเรียนรู้ของเครื่อง (Machine Learning)
- การพัฒนาและปรับปรุงกระบวนการ CI/CD เพื่อเพิ่มประสิทธิภาพการพัฒนาและการส่งมอบซอฟต์แวร์

มหาวิทยาลัยเทคโนโลยีสุรนารี