



การระบุตำแหน่งจุดศูนย์กลางรูปร่างตาโดยใช้การเรียนรู้เชิงลึกด้วยการถดถอยเชิงคอนโวลูชันแบบ
หลายความละเอียดและการแปลงไม่เชิงเส้น



โดย
นางสาวพิชชาภา เลิศศิริวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญาโทมหาบัณฑิต
ภาควิชาคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2563

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

การระบุตำแหน่งจุดศูนย์กลางรูปร่างตาโดยใช้การเรียนรู้เชิงลึกด้วยการถดถอยเชิงคอนโวลูชันแบบหลายความละเอียดและการแปลงไม่เชิงเส้น



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญาโทมหาบัณฑิต
ภาควิชาคอมพิวเตอร์
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2563
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

LOCATING THE EYE'S PUPIL CENTER THROUGH DEEP LEARNING USING
MULTISCALE AND NON-LINEARITY CONVOLUTIONAL REGRESSION



A Thesis Submitted in Partial Fulfillment of the Requirements
for Master of Science (INFORMATION TECHNOLOGY AND DIGITAL INNOVATION)

Department of COMPUTER SCIENCE
Graduate School, Silpakorn University

Academic Year 2020

Copyright of Graduate School, Silpakorn University

หัวข้อ	การระบุตำแหน่งจุดศูนย์กลางรูปร่างตาโดยใช้การเรียนรู้เชิงลึกด้วยการถดถอยเชิงคอนโวลูชันแบบหลายความละเอียดและการแปลงไม่เชิงเส้น
โดย	พิชชาภา เลิศศิริวรรณเมธ
สาขาวิชา	เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผนก ก แบบ ก 2 ระดับปริญญาโทมหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ผู้ช่วยศาสตราจารย์ ดร. ภิญโญ แท้ประสาทสิทธิ์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.จุไรรัตน์ นันทานิช)

พิจารณาเห็นชอบโดย

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ศุภวงศ์ ทวีรอบ)

.....อาจารย์ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร.ภิญโญ แท้ประสาทสิทธิ์)

.....ผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร.สุนีย์ พงษ์พินิจภิญโญ)

.....ผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร.รัชดาพร คณาวงษ์)

61318202 : เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญาโทมหาบัณฑิต
คำสำคัญ : การระบุจุดศูนย์กลางระฆังตา การเรียนรู้เชิงลึก การขยายข้อมูล โครงข่ายประสาทเทียม
แบบคอนโวลูชันเชิงลึก

นางสาว พิชชาภา เลิศศิริวรรณ: การระบุตำแหน่งจุดศูนย์กลางระฆังตาโดยใช้การเรียนรู้
เชิงลึกด้วยการถดถอยเชิงคอนโวลูชันแบบหลายความละเอียดและการแปลงไม่เชิงเส้น อาจารย์ที่
ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ ดร. ภิญโญ แพ้ประสาทสิทธิ์

การตรวจหาตำแหน่งระฆังตาเป็นปัญหาที่ได้รับความสนใจอย่างมาก เนื่องจากได้ถูก
นำไปใช้ประโยชน์ในหลาย ๆ ด้าน ตัวอย่างเช่น ในงานวิจัยเกี่ยวกับการปฏิสัมพันธ์ของมนุษย์และ
คอมพิวเตอร์ และการนำมาใช้ทางการแพทย์ อย่างไรก็ตามกระบวนการวิธีระบุตำแหน่งระฆังตาใน
งานวิจัยที่มีมาก่อนหน้าเป็นการศึกษาภายใต้การควบคุมสภาพแวดล้อมที่เปรียบเสมือน
ห้องปฏิบัติการ และสำหรับการนำมาใช้ในกลุ่มประชากรที่มีสิ่วระฆังตาเข้มยังมีความแม่นยำค่อนข้าง
ต่ำ ดังนั้น ผู้วิจัยจึงมีความต้องการที่จะเพิ่มความแม่นยำและความคงทนของวิธีการระบุตำแหน่งจุด
ศูนย์กลางระฆังตา โดยได้นำเสนอตัวแบบโครงข่ายประสาทเทียมที่ออกแบบเพื่อประสิทธิภาพที่ดีขึ้น
ด้วยโครงสร้างที่ถูกพัฒนามาจากโครงสร้างพื้นฐาน อันดับแรกเราเสนอให้ใช้การแปลงไม่เชิงเส้นที่
สูงขึ้นในตัวแบบเพื่อการอนุมานผลลัพธ์ที่แม่นยำขึ้น ต่อมาได้เสนอให้นำคุณลักษณะจากหลายระดับ
ของการทำคอนโวลูชันมาใช้ร่วมกันสำหรับการวิเคราะห์การถดถอยในขั้นตอนสุดท้าย เพื่อเปิดโอกาส
ในการเรียนรู้ของเครื่องด้วยข้อมูลเชิงพื้นที่และการแปลงไม่เชิงเส้นในหลายระดับที่แตกต่างกัน ซึ่งเป็น
การนำไปสู่ตัวแบบการเรียนรู้ที่มีความสามารถในการนำไปประยุกต์ใช้จริงในทางปฏิบัติได้ และ
นอกจากนี้ผู้วิจัยได้ทดลองใช้คอนโวลูชันแบบขยายในลักษณะต่าง ๆ ด้วย ผู้วิจัยประเมินความแม่นยำ
และความคงทนของวิธีการที่ผู้วิจัยนำเสนอผ่านการวิเคราะห์แบบตัดออก โดยใช้ภาพจากฐานข้อมูล
MMU2, BioID และ FFHQ พบว่าตัวแบบที่ผู้วิจัยนำเสนอซึ่งนำคุณลักษณะจากหลายระดับของการ
ทำคอนโวลูชันมาใช้ร่วมกันและใช้การแปลงไม่เชิงเส้นที่สูงขึ้นให้ความแม่นยำและความคงทนที่ดีกว่า
ตัวแบบพื้นฐานอย่างมีนัยสำคัญ และเมื่อเปรียบเทียบกับวิธีการก่อนหน้าด้วยมาตรวัด Wec ตัวแบบที่
นำเสนอสามารถให้ผลลัพธ์ที่ดีกว่าบนฐานข้อมูล BioID

61318202 : Major (INFORMATION TECHNOLOGY AND DIGITAL INNOVATION)

Keyword : pupil localization deep learning data augmentation convolutional neural network

MISS PHITCHAPHA LERTSIRAVARAMETH : LOCATING THE EYE'S PUPIL CENTER THROUGH DEEP LEARNING USING MULTISCALE AND NON-LINEARITY CONVOLUTIONAL REGRESSION THESIS ADVISOR : ASSISTANT PROFESSOR DR. PINYO TAEPRASARTSIIT

Pupil detection has been of wide interest as it can be utilized in numerous applications, such as human-computer interaction and medical diagnosis. However, the recently proposed methods to locate the pupil rely on laboratory conditions and still lack the performance when the surrounding of the pupil is shadowed and comparably dark. Therefore, to further improve the accuracy and the robustness for locating the eye's pupil center, this work presents a deep neural network model that is designed to accomplish more effectively by building on basic neural network structure. We first propose applying more nonlinearity to the network to increase the approximation power. Next, we propose to utilize feature maps from multi-levels of convolutional layers for final regression, so the network can explicitly employ these features together. With the increase in different degrees of nonlinearities and multi spatial information, the generalization capability of the model is increased, which in turn contributes to the model's robustness. Furthermore, we also experiment with dilated convolution layers in diverse settings. To validate the contribution of the multiscale nonlinear features concatenation, ablation study was performed on MMU2, BioID and FFHQ datasets . Overall, the results show better accuracy and robustness when the feature maps are merged together with adding more nonlinearity in the network. Also, the proposed method outperforms previous works in terms of wec accuracy on BioID dataset.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงตามวัตถุประสงค์ทุกประการ ผู้วิจัยขอขอบคุณบุคคลต่าง ๆ ที่ให้คำปรึกษา แนะนำ และช่วยเหลืออย่างดียิ่ง ทั้งด้านวิชาการและด้านการดำเนินงานวิจัย ดังนี้ ขอขอบคุณ อาจารย์ภิญโญ แท้ประสาทสิทธิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้ความรู้และให้คำปรึกษาและคอยผลักดันส่งเสริมการทำวิจัย ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนและเอาใจใส่ ด้วยดีเสมอมา รวมถึงให้คำแนะนำในการดำเนินชีวิตด้วยความเมตตา ขอขอบคุณนายณัฐพงศ์ แต่งจ้อย และนายชนก ปฐมเพทาย ที่ได้ให้ความช่วยเหลือ ให้คำแนะนำ และอำนวยความสะดวกในการใช้เครื่องมือวิจัย จนทำให้วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยดี

สุดท้ายนี้ขอกราบขอบพระคุณอาจารย์ทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้ทำวิจัย ทั้งในอดีตตลอดจนถึงปัจจุบัน และกราบขอพระคุณบิดา มารดา รวมถึงครอบครัวทุกท่านที่ให้การอบรม เลี้ยงดู และส่งเสริมการศึกษาเป็นอย่างดีมาโดยตลอด จนทำให้ผู้วิจัยประสบความสำเร็จในชีวิตเรื่อยมา

พิชชาภา เลิศศิรธรรม



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	3
1.3 สมมติฐานของงานวิจัย.....	4
1.4 ขอบเขตงานวิจัย.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.6 โครงร่างหัวข้อวิทยานิพนธ์.....	5
บทที่ 2 งานวิจัยที่เกี่ยวข้อง.....	6
PupilNet: Convolutional Neural Networks for Robust Pupil Detection.....	7
CNN-Based Pupil Center Detection for Wearable Gaze Estimation System.....	9
DeepVOG : Open-source pupil segmentation and gaze estimation in neuroscience using deep learning.....	14
Pupil Localisation and Eye Centre Estimation Using Machine Learning and Computer Vision.....	16
บทที่ 3 ทฤษฎีที่เกี่ยวข้อง.....	21
ดวงตาและโครงสร้างทางกายวิภาค.....	21

การเรียนรู้ของเครื่อง (Machine Learning)	22
การเรียนรู้เชิงลึก (Deep Learning).....	23
โครงข่ายประสาทเทียม (Neural network)	24
โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural network - CNN).....	25
การขยายข้อมูล (Data Augmentation).....	25
การปรับปรุงภาพด้วยวิธีการปรับเท่าฮิสโทแกรม (Histogram Equalization)	25
ความคงทน (Robustness).....	26
ความแม่นยำ (Accuracy).....	26
บทที่ 4 วิธีการดำเนินงานวิจัย	27
4.1 ข้อมูลที่ใช้ในงานวิจัย.....	27
4.2 แนวความคิดและขั้นตอนการทดลอง	29
4.2.1 การเตรียมข้อมูล	29
4.2.2 การเรียนรู้เชิงลึกของเครื่องและการทดสอบการเรียนรู้	33
4.2.3 การวัดประสิทธิภาพ	38
บทที่ 5 ผลการดำเนินงานวิจัย.....	41
การทดลองที่ 1 การวิเคราะห์การถดถอยที่มีระดับของการแปลงไม่เชิงเส้นที่เพิ่มขึ้น	41
การทดลองที่ 2 การวิเคราะห์การถดถอยที่นำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมา ประมวลผลร่วมกัน	46
การทดลองที่ 3 การวิเคราะห์การถดถอยที่มีระดับของการแปลงไม่เชิงเส้นที่เพิ่มขึ้นพร้อมกับการ นำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกัน	51
การทดลองที่ 4 การวิเคราะห์การถดถอยโดยการใช้ชั้นคอนโวลูชันแบบขยาย.....	62
การทดลองที่ 5 การเปรียบเทียบประสิทธิภาพของตัวแบบกับงานวิจัยอื่น ๆ ที่เกี่ยวข้องบน ฐานข้อมูล BiOD.....	65
การทดลองที่ 6 การทดลองฝึกตัวแบบการเรียนรู้กับรูปภาพที่ไม่มีพื้นหลังสีดำ	66
บทที่ 6 สรุปผลการทดลองและข้อเสนอแนะ	67

รายการอ้างอิง 71

ประวัติผู้เขียน 74



สารบัญตาราง

หน้า

ตารางที่ 2.1 รายละเอียดโครงสร้างของตัวแบบในแต่ละขั้นตอนวิธี	10
ตารางที่ 2.2 ผลการประเมินตัวแบบสำหรับจำแนกประเภทภาพตามสถานะตาของวิธี A และ B	12
ตารางที่ 2.3 ความแตกต่างของพื้นที่รูปร่างตาที่ตัวแบบแยกได้และพื้นที่จริง	16
ตารางที่ 2.4 ประสิทธิภาพของวิธีการที่นำเสนอโดยใช้มาตรวัด Wec และ Bec	19
ตารางที่ 2.5 ตารางเปรียบเทียบประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนสัมบูรณ์ สัมประสิทธิ์การกำหนด ระยะทางแบบยุคลิด และ มาตรวัด SED	20
ตารางที่ 4.1 ตารางแสดงโครงสร้างของการนำคุณลักษณะจากชั้นต่าง ๆ มาประมวลผลร่วมกันในชั้นการเชื่อมต่อแบบสมบูรณ์	36
ตารางที่ 5.1 ตารางเปรียบเทียบประสิทธิภาพของ SimpleCNN และ DoubleCNN	42
ตารางที่ 5.2 ตารางเปรียบเทียบประสิทธิภาพของ SimpleCNN, ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3	46
ตารางที่ 5.3 ตารางเปรียบเทียบประสิทธิภาพของ DoubleConcatCNNv1, DoubleConcatCNNv2 ,DoubleConcatCNNv3 และตัวแบบที่นำเสนอก่อนหน้า	52
ตารางที่ 5.4 ผลการทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ DoubleConcatCNNv1 กับ DoubleConcatCNNv2	54
ตารางที่ 5.5 ผลการทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ DoubleConcatCNNv2 กับ DoubleConcatCNNv3	54
ตารางที่ 5.6 ตารางแสดงโครงสร้างของการนำคุณลักษณะจากชั้นต่าง ๆ มาประมวลผลร่วมกันในชั้นการเชื่อมต่อแบบสมบูรณ์	63
ตารางที่ 5.7 ตารางเปรียบเทียบประสิทธิภาพของ AtrousDoubleConcatCNN และ DoubleConcatCNN	64
ตารางที่ 5.8 ตารางแสดงการเปรียบเทียบประสิทธิภาพของ DoubleConcatCNNv3 กับงานวิจัยที่เกี่ยวข้อง	65

ตารางที่ 5.9 ประสิทธิภาพของตัวแบบ SimpleCNN, DoubleConcatCNNv2 และ DoubleConcatCNNv3 บนชุดข้อมูลฝึก BioID ที่มีวิธีการเตรียมข้อมูลต่างกัน (แบบมีพื้นหลังสีดำ และแบบไม่มีพื้นหลังสีดำ) 66



สารบัญภาพ

หน้า

ภาพที่ 1.1 ค่าการดูดกลืนแสงของฮีโมโกลบินและออกซีฮีโมโกลบิน.....	1
ภาพที่ 1.2 องค์ประกอบของตา : ขนตา (Eyelid), รูม่านตา (Pupil), Sclera(ตาขาว), Iris(ม่านตา) ..	2
ภาพที่ 1.3 ลักษณะตาทั่วไปของคนเอเชีย	3
ภาพที่ 2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้ระบุตำแหน่งจุดศูนย์กลาง.....	7
ภาพที่ 2.2 ภาพรวมของระบบที่ใช้ในงานวิจัย (ที่มาของภาพ Fuhl และคณะ [8]).....	8
ภาพที่ 2.3 Detection Rate ของวิธีการที่ผู้วิจัยนำเสนอเปรียบเทียบกับวิธีการซึ่งเป็น state-of-the-art ทั้งหมด 4 วิธี (ExCuSe, SET, Starburs และ Swirski).....	9
ภาพที่ 2.4 ภาพรวมของระบบในวิธีการ A และ B (ที่มาของภาพ Chinsatit และคณะ [9]).....	10
ภาพที่ 2.5 ตัวอย่างภาพที่ทำนายผิดพลาด	12
ภาพที่ 2.6 ค่าความคลาดเคลื่อนเฉลี่ยของแต่ละวิธีการ	13
ภาพที่ 2.7 อัตราความสำเร็จของวิธีที่ผู้วิจัยนำเสนอ โดยค่าในแกน y คือร้อยละของภาพที่มีความคลาดเคลื่อนน้อยกว่าค่าที่สัมพันธ์กันในแกน x เมื่อ x คือระยะห่างในช่วง 0-12 พิกเซล	14
ภาพที่ 2.8 สถาปัตยกรรมที่ใช้ในงานวิจัยนี้ (ที่มาของภาพ Yiu และคณะ [10]).....	15
ภาพที่ 2.9 ผลลัพธ์ที่ผิดพลาดจากภาพที่มีอุปสรรค (ที่มาของภาพ Yiu และคณะ [10])	16
ภาพที่ 2.10 ภาพรวมของกระบวนการที่ใช้ในงานวิจัย (ที่มาของภาพ Khan และคณะ [13]).....	17
ภาพที่ 2.11 การทำคอนโวลูชันตามแนวแกนนอนและการทำคอนโวลูชันตามแนวแกนตั้ง.....	18
ภาพที่ 3.1 เทคนิคการเรียนรู้ของเครื่อง	22
ภาพที่ 4.1 ตัวอย่างภาพในชุดข้อมูล MPIIGaze	27
ภาพที่ 4.2 ตัวอย่างภาพในชุดข้อมูล Bioid	27
ภาพที่ 4.3 ตัวอย่างภาพในชุดข้อมูล MMU2.....	28
ภาพที่ 4.4 ตัวอย่างภาพในชุดข้อมูล Flickr-Faces-HQ-68000 (FFHQ).....	28
ภาพที่ 4.5 การกระจายของตำแหน่งจุดศูนย์กลางรูม่านตาในชุดข้อมูล MPII.....	29

ภาพที่ 4.6 ตัวอย่างภาพข้อมูลฝึกที่ผ่านการขยายข้อมูล 29

ภาพที่ 4.7 ตัวอย่างภาพตากรรมที่จุดกึ่งกลางตาอยู่ตรงกลางระหว่างพิกเซล..... 30

ภาพที่ 4.8 ขั้นตอนการเตรียมข้อมูลสำหรับฐานข้อมูล BioID และ FFHQ 31

ภาพที่ 4.9 ภาพหลังการเตรียมข้อมูลของชุดข้อมูลทดสอบทั้งสามฐานข้อมูล 32

ภาพที่ 4.10 การกระจายตัวของผลเฉลี่ยในชุดข้อมูลทดสอบ ก) ฐานข้อมูล MMU2 ข) ฐานข้อมูล BioID ค) ฐานข้อมูล FFHQ 33

ภาพที่ 4.11 ตัวอย่างภาพในชุดข้อมูลทดสอบหลังจากผ่านการปรับเท่าฮิสโทแกรม..... 33

ภาพที่ 4.12 โครงสร้างพื้นฐาน (SimpleCNN)..... 35

ภาพที่ 4.13 โครงสร้างของตัวแบบ DoubleConcatCNN ที่นำเสนอ..... 38

ภาพที่ 5.1 แผนภาพเศษเหลือของตัวแบบ SimpleCNN ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y..... 44

ภาพที่ 5.2 แผนภาพเศษเหลือของตัวแบบ DoubleCNN ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y..... 45

ภาพที่ 5.3 แผนภาพเศษเหลือของตัวแบบ ConcatCNNv1 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y..... 48

ภาพที่ 5.4 แผนภาพเศษเหลือของตัวแบบ ConcatCNNv2 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพ

ภาพที่ 6.1 แผนภูมิแท่งแสดงอัตราความสำเร็จของตัวแบบที่นำเสนอในการทดลองที่ 1-4 เมื่อทดสอบ
บนฐานข้อมูล MMU2, BioID และ FFHQ..... 67

ภาพที่ 6.2 แผนภาพกล่องแสดงการกระจายของระยะทางคลาดเคลื่อนเฉลี่ยเมื่อทดลองซ้ำ 5 รอบ 68

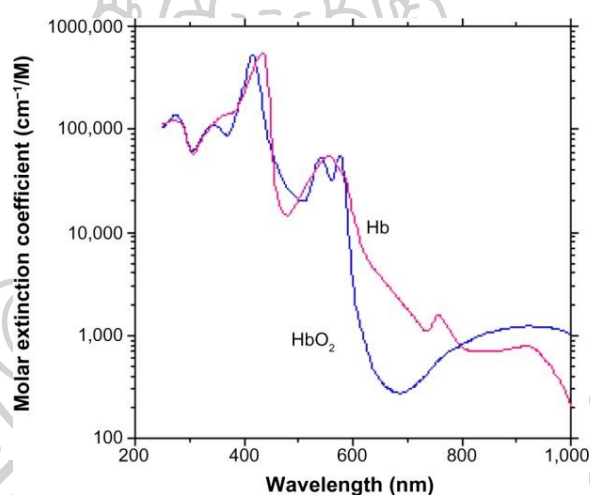


บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

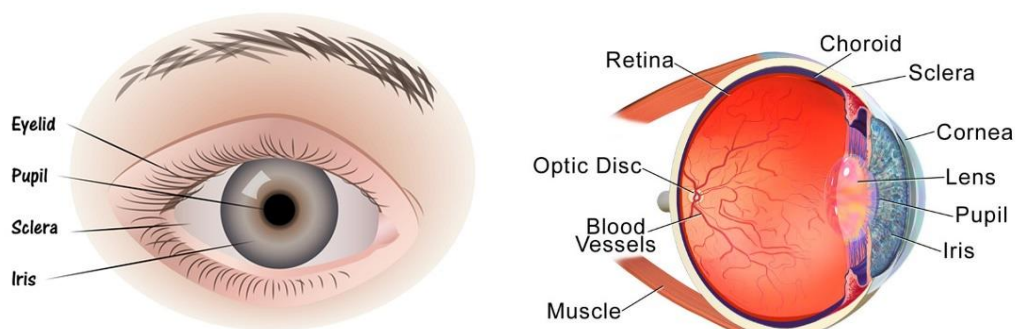
ปัจจุบันมีงานวิจัยจำนวนมากเกี่ยวข้องกับการตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกาย ซึ่งหลักการตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกายมีพื้นฐานมาจากการดูดซับคลื่นแสงที่แตกต่างกันของฮีโมโกลบินที่จับกับออกซิเจน (Oxyhemoglobin, HbO₂) และฮีโมโกลบินที่ไม่จับกับออกซิเจน (Hb) ดังภาพที่ 1.1 ทำให้เกิดความแตกต่างของคลื่นแสงช่วงที่ไม่มีเลือดไหลผ่านและช่วงที่มีเลือดไหลผ่านในหลอดเลือดแดง จึงส่งผลต่อการเปลี่ยนแปลงของสีผิวที่สัมพันธ์กับการบีบและคลายตัวของหัวใจ [1] ดังนั้นการประมวลผลภาพจากวิดีโอเฟรมเพื่อวิเคราะห์การเปลี่ยนแปลงของสีที่สะท้อนออกมาจากฮีโมโกลบินอย่างละเอียดจึงเป็นวิธีหนึ่งที่สามารถตรวจวัดสัญญาณชีพจรได้โดยไม่มีการสัมผัสร่างกาย



ภาพที่ 1.1 ค่าการดูดกลืนแสงของฮีโมโกลบินและออกซิฮีโมโกลบิน
(ที่มาของภาพ omlc.org/spectra/hemoglobin/index.html)

อย่างไรก็ตามแม้จะมีการศึกษาและวิจัยอย่างแพร่หลายเกี่ยวกับการตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกาย การนำไปใช้งานในสภาพแวดล้อมทั่วไปก็ยังคงไม่เป็นที่นิยมมากนัก เนื่องจากมีความแม่นยำต่ำเมื่อใช้กับผิวสีเข้มและสัตว์ซึ่งมีขนปกคลุมตามร่างกาย จากปัญหานี้จึงทำให้ผู้วิจัยต้องการหาพื้นที่อื่นที่สามารถลดข้อจำกัดดังกล่าวลงแต่ยังคงมีปริมาณเส้นเลือดมากเพียงพอที่จะสะท้อนสีของฮีโมโกลบินออกมาได้มาทดลอง ซึ่งผู้วิจัยพบว่าด้านหลังบริเวณรูม่านตามีเส้นเลือดจำนวนมากอยู่ดังแสดงในภาพที่ 1.2 และงานวิจัยของ Flammer และคณะ [2] ยังแสดงให้เห็นว่าข้อมูลจากเส้นเลือดหลังบริเวณรูม่านตาสามารถใช้ประเมินความเสี่ยงของการเป็นโรคกลุ่มที่เกิดจากหลอดเลือดหัวใจและหลอดเลือดได้ ดังนั้นผู้วิจัยจึงมีสมมติฐานว่ารูม่านตาจะสามารถสะท้อนการ

เปลี่ยนแปลงของสีจากฮีโมโกลบินซึ่งไหลเวียนอยู่ในเส้นเลือดออกมาได้ เพื่อทดสอบสมมติฐานจึงจำเป็นต้องระบุตำแหน่งจุดศูนย์กลางรูม่านตาในภาพและประมวลผลจากสีบริเวณนั้น แต่การที่รูม่านตามีขนาดเล็ก การนำมาใช้งานจึงต้องอาศัยวิธีระบุตำแหน่งจุดศูนย์กลางรูม่านตาที่มีความแม่นยำสูง



ภาพที่ 1.2 องค์ประกอบของตา : ขนตา (Eyelid), รูม่านตา (Pupil), Sclera(ตาขาว), Iris(ม่านตา) (ที่มาของภาพ news-medical.net [3] และ wikipremed.com [4])

สำหรับวิธีระบุตำแหน่งจุดศูนย์กลางรูม่านตา สามารถแบ่งได้เป็น 2 ประเภท ประเภทแรกคือ Geometry-based ขั้นตอนสำหรับการระบุตำแหน่งจุดศูนย์กลางรูม่านตาประเภท Geometry-based จะมีลำดับวิธีที่ถูกกำหนดแนวทางไว้ชัดเจน มีหลักการทั่วไปคือการหาเส้นรอบรูม่านตาในภาพ ซึ่งโดยมากมักตรวจจับเส้นโค้งหรือวงกลมโดยใช้เทคนิค Hough Circle Transform หรือนิยามตัวแบบทางคณิตศาสตร์ชนิดที่เป็นสมการอิลิปติก (Elliptic equation) ขึ้นมา เพื่อตรวจจับรูม่านตาโดยเฉพาะ อย่างไรก็ตามการหาขอบรูม่านตาที่ถูกต้องสมบูรณ์ได้นั้นไม่ใช่เรื่องง่าย โดยเฉพาะการหาขอบของวัตถุในภาพที่มีคุณภาพต่ำ

ประเภทที่สองคือ Machine learning-based หรือประเภทที่ใช้การเรียนรู้ของเครื่อง เป็นการสร้างระบบที่ทำให้คอมพิวเตอร์สามารถเรียนรู้ได้จากชุดข้อมูลตัวอย่าง แทนการเขียนเป็นชุดคำสั่งให้ปฏิบัติตามขั้นตอนที่กำหนด การเรียนรู้ของเครื่องจะเรียนรู้ว่าข้อมูลเข้า (ภาพดวงตา) และข้อมูลออก (จุดศูนย์กลางรูม่านตา) มีความเกี่ยวข้องกันอย่างไร และระบบจะทำการวิเคราะห์ถึงความสัมพันธ์ของข้อมูลเพื่อที่จะผลิตผลลัพธ์ที่แม่นยำออกมา ซึ่งโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolution Neural Network: CNN) [5] เป็นรูปแบบหนึ่งของการเรียนรู้ของเครื่องที่ปัจจุบันนิยมใช้อย่างแพร่หลาย เนื่องจากเป็นวิธีที่มีศักยภาพสูงในการจัดการปัญหาทางด้านคอมพิวเตอร์วิทัศน์และยังสามารถตอบโจทย์งานได้หลายด้าน

อย่างไรก็ตามการระบุจุดศูนย์กลางรูม่านตาในงานวิจัยที่เกี่ยวข้องก่อนหน้านี้ ยังขาดความแม่นยำเมื่อนำมาใช้กับภาพตาที่มีรูม่านตาขนาดใหญ่และม่านตาสีเข้มอย่างเช่นในภาพที่ 4 ทำให้ไม่สามารถนำเครื่องมือวิจัยที่มีอยู่มาปรับใช้กับกลุ่มประชากรส่วนใหญ่ของทวีปเอเชียได้อย่างมี

ประสิทธิภาพ เนื่องจากโดยเทคนิคภาพดวงตาในลักษณะนี้จะมีความทำลายมากกว่า จากการศึกษาที่ความเข้มแสงจากจุดบริเวณม่านตาไปถึงจุดบริเวณรูม่านตามีความแตกต่างน้อย ขอบของรูม่านตาจึงถูกตรวจจับได้ยากขึ้นการหาจุดศูนย์กลางจึงมักมีความคลื่อนเพิ่มขึ้นตามไปด้วย



ภาพที่ 1.3 ลักษณะตาทั่วไปของคนเอเชีย
(ที่มาของภาพ huffpost.com)

ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะพัฒนาวิธีการหาจุดศูนย์กลางรูม่านตา เพื่อเพิ่มความแม่นยำและความคงทน (Robustness) โดยเฉพาะอย่างยิ่งกับภาพตาที่มีม่านตาสีเข้มเช่นเดียวกับสีของรูม่านตา โดยผู้วิจัยใช้การประมวลผลภาพด้วยเทคโนโลยีการเรียนรู้เชิงลึกผ่านโครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งภายในงานวิจัยได้ศึกษาและวิเคราะห์วิธีการเพิ่มความสามารถของโครงข่ายประสาทเทียมนำเสนอวิธีในการสร้างโครงข่ายประสาทเทียมเพื่อระบุจุดศูนย์กลางรูม่านตา และเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีที่นำเสนอกับวิธีอื่น ๆ ในงานวิจัยที่เกี่ยวข้องโดยใช้ข้อมูลสาธารณะที่ได้มาตรฐานในการทดสอบ

1.2 วัตถุประสงค์ของงานวิจัย

ในปัจจุบันมีตัวแบบการเรียนรู้เชิงลึกมากมายทำหน้าที่ระบุตำแหน่งจุดศูนย์กลางรูม่านตา แต่ยังคงมีปัญหาความแม่นยำกับรูปภาพดวงตาที่มีขนาดรูม่านตาใหญ่และมีม่านตาสีเข้ม ซึ่งเป็นลักษณะทั่วไปของคนเอเชีย ผู้วิจัยจึงต้องการเพิ่มความแม่นยำและความคงทนในการระบุตำแหน่งจุดศูนย์กลางรูม่านตาของการฝึกสอนตัวแบบการเรียนรู้เชิงลึก โดยมุ่งเน้นไปที่ภาพที่มีความทำลายได้แก่ ภาพดวงตาที่มีม่านตาและรูม่านตาสีเข้มใกล้เคียงกัน ภาพที่มีการสะท้อนของสภาพแวดล้อมภายนอกเข้ามาบริเวณดวงตา และภาพที่ไม่อยู่ในสภาพแวดล้อมจำกัดโดยการใช้แสงอินฟราเรด

1.3 สมมติฐานของงานวิจัย

ตัวแบบการเรียนรู้เชิงลึกซึ่งใช้การถดถอยเชิงคอนโวลูชันจะสามารถระบุตำแหน่งจุดกึ่งกลางรูม่านตาได้อย่างแม่นยำจากการแยกคุณลักษณะของพื้นที่ย่อยในภาพดวงตา ยิ่งไปกว่านั้นคุณลักษณะแบบไม่เชิงเส้นที่แยกได้ในชั้นคอนโวลูชันแต่ละชั้นให้ข้อมูลสำหรับการอนุมานตำแหน่งรูม่านตาที่แตกต่างกัน การนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันจึงน่าจะให้ผลลัพธ์ที่ดีกว่าการเลือกนำผลลัพธ์จากชั้นใดเพียงชั้นเดียวมาใช้งาน และผู้วิจัยคาดการณ์ว่าการแปลงไม่เชิงเส้นที่สูงขึ้นจากการเชื่อมต่อชั้นคอนโวลูชันและฟังก์ชันการกระตุ้นไม่เชิงเส้น (Non-linear activation function) จะส่งผลต่อความแม่นยำของตัวแบบ ซึ่งสามารถทดสอบและวิเคราะห์ได้จากการเปรียบเทียบความแม่นยำจากตัวแบบที่มีจำนวนชั้นที่แตกต่างกันไป

1.4 ขอบเขตงานวิจัย

1. งานวิจัยนี้จะมุ่งเน้นการพัฒนาวิธีการระบุจุดศูนย์กลางรูม่านตาให้มีความแม่นยำและความคงทนเพิ่มขึ้น
2. ข้อมูลที่ใช้ในงานวิจัยเป็นภาพถ่ายที่ไม่ถูกควบคุมโดยการใช้แสงแสงอินฟราเรด โดยได้รวบรวมภาพจากฐานข้อมูลสาธารณะ ได้แก่ MPIIGaze, BioID, MMU2 และ Flickr-Faces-HQ-68000 (FFHQ) และครอปตัดภาพบริเวณดวงตาให้มีขนาด 32 x 64 พิกเซล
3. ภาพที่ใช้ฝึกและทดสอบตัวแบบทั้งหมดเป็นภาพที่ถูกตัดพื้นที่ให้เหลือเพียงบริเวณดวงตาเท่านั้น แต่ไม่จำเป็นที่ดวงตาต้องอยู่ตรงกลางภาพ
4. ตัวแบบการเรียนรู้เชิงลึกที่ได้จากงานวิจัยนี้ ใช้ระบุตำแหน่งจุดศูนย์กลางรูม่านตา คือ พิกัดในแนวแกนนอนและพิกัดในแนวแกนตั้ง

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ตัวแบบสามารถระบุตำแหน่งจุดศูนย์กลางรูม่านตาได้อย่างแม่นยำแม้ในกรณีที่มีสีเข้มและภาพไม่อยู่ในสภาพแวดล้อมที่จำกัดโดยการใช้แสงอินฟราเรด และนำไปสู่การประมวลผลภาพจากวิดีโอเฟรมบริเวณดวงตาเพื่อวัตถุประสงค์ชันิจจริงโดยไม่มีการสัมผัสร่างกายได้อย่างแม่นยำ
2. สามารถนำตัวแบบไปประยุกต์ใช้กับงานอื่นที่เกี่ยวข้องกับการระบุจุดศูนย์กลางรูม่านตาได้ เช่น งานเกี่ยวกับการหาดำแหน่งการเพ่งมอง (eyegaze estimation)
3. สามารถนำองค์ความรู้ที่ได้จะการทดลองเปรียบเทียบสถาปัตยกรรมที่แตกต่างกัน ไปปรับใช้และต่อยอดในการออกแบบสถาปัตยกรรมการเรียนรู้เชิงลึกในอนาคตได้

1.6 โครงร่างหัวข้อวิทยานิพนธ์

ลำดับหัวข้องานวิจัยจะเป็นไปดังนี้ คือ บทที่ 2 จะกล่าวถึงงานวิจัยที่เกี่ยวข้อง บทที่ 3 กล่าวถึงทฤษฎีที่เกี่ยวข้อง บทที่ 4 วิธีดำเนินงานวิจัย บทที่ 5 ผลการดำเนินงานวิจัย และบทที่ 6 สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ



บทที่ 2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับโครงร่างวิทยานิพนธ์ฉบับนี้ได้แบ่งเป็น 2 กลุ่มได้แก่ งานวิจัยเกี่ยวกับการตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกาย และงานวิจัยเกี่ยวกับการระบุตำแหน่งจุดศูนย์กลางรูม่านตาโดยใช้ตัวแบบการเรียนรู้เชิงลึก สำหรับกลุ่มแรกพบว่างานวิจัยในอดีตพยายามตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกายด้วยการประมวลผลภาพจากวิดีโอเฟรม เพื่อลดความทรมานและความเจ็บปวด หรือแม้แต่ความกดดันที่คนและสัตว์ได้รับจากการติดตั้งเครื่องมือและอุปกรณ์สำหรับตรวจวัดสัญญาณชีพจร งานวิจัยของ Poh และคณะ [6] ใช้ภาพจากวิดีโอเฟรมบริเวณใบหน้ามาแยกแหล่งกำเนิดที่มองไม่เห็น (Blind Source Separation) ด้วยวิธี Independence Component Analysis (ICA) และวิเคราะห์การเปลี่ยนแปลงความถี่ของสัญญาณด้วย Fast-Fourier Transform (FFT) เพื่อวัดค่าสัญญาณชีพ, Unakafov และคณะ [7] เสนอวิธีตรวจวัดชีพจรของสัตว์ในตระกูลลิงพิจารณาจากวิดีโอเฟรมภาพใบหน้าลิงโดยใช้หลักการ imaging photoplethysmography (iPPG) ประมวลผลสัญญาณภาพจากวิดีโอเฟรมโดยคำนวณจากความแตกต่างของสัญญาณสีเขียวและแดง (Green-Red difference) เพื่อหาสัญญาณที่มีการเปลี่ยนแปลงที่สอดคล้องกับสัญญาณชีพจรและวิเคราะห์การเปลี่ยนแปลงความถี่ของสัญญาณด้วย FFT อย่างไรก็ตามงานดังกล่าวยังมีความแม่นยำน้อยเมื่อนำไปใช้กับคนที่มีผิวสีเข้มหรือสัตว์ที่มีขนปกคลุมเป็นจำนวนมาก จึงเป็นที่มาของการหาพื้นที่บริเวณอื่นของร่างกายเพื่อลดข้อจำกัดและขยายขอบเขตของการตรวจวัดสัญญาณชีพจรโดยไม่สัมผัสร่างกาย

สำหรับงานวิจัยกลุ่มที่สองซึ่งเกี่ยวข้องกับการระบุตำแหน่งจุดศูนย์กลางรูม่านตา สามารถแบ่งได้เป็น 2 ประเภท ดังที่กล่าวไว้ก่อนหน้านี้คือ Geometry-based และ Machine learning-based งานวิจัยที่เกี่ยวข้องในกลุ่ม Geometry-based นั้นต้องอาศัยความสัมพันธ์ทางเรขาคณิตระหว่างองค์ประกอบต่าง ๆ ภายในภาพเพื่อตรวจจับวัตถุ ซึ่งในกระบวนการตรวจจับนั้นต้องการความละเอียดของภาพและอัตราส่วนของแสงต่อสัญญาณรบกวนค่อนข้างสูงและเพื่อให้ได้มาซึ่งความแม่นยำ นอกจากนี้วิธีนี้ยังมีข้อจำกัดอยู่ที่เมื่อแสงเงาของสภาวะแวดล้อมเปลี่ยนแปลง ประสิทธิภาพของการตรวจจับจะเปลี่ยนตามไปด้วย จึงจำเป็นต้องปรับแต่งค่าต่าง ๆ ให้มีความเหมาะสมตามลักษณะสภาพแวดล้อมและข้อมูลอยู่เสมอ ทำให้ไม่เหมาะสมกับการนำมาใช้ในสภาพแวดล้อมที่หลากหลาย ดังนั้นงานวิจัยนี้จึงมุ่งเน้นไปที่วิธีในประเภท Machine learning-based ซึ่งมีความสามารถในการรวบรวมความรู้ผ่านกระบวนการเรียนรู้ (Learning process) โดยความรู้เหล่านี้จะถูกจัดเก็บอยู่ในโครงข่ายในรูปแบบค่าน้ำหนัก ทำให้สามารถนำไปทำนายหรือประมาณความสัมพันธ์ของข้อมูลชุดอื่น ๆ ที่อยู่ในรูปแบบเดียวกันได้ โดยค่าน้ำหนักจะเปลี่ยนแปลงไปตามสถาปัตยกรรมของโครงข่ายประสาทและชุดข้อมูลที่ฝึก ซึ่งงานวิจัยแต่ละชิ้นจะมีโครงสร้างโครงข่าย

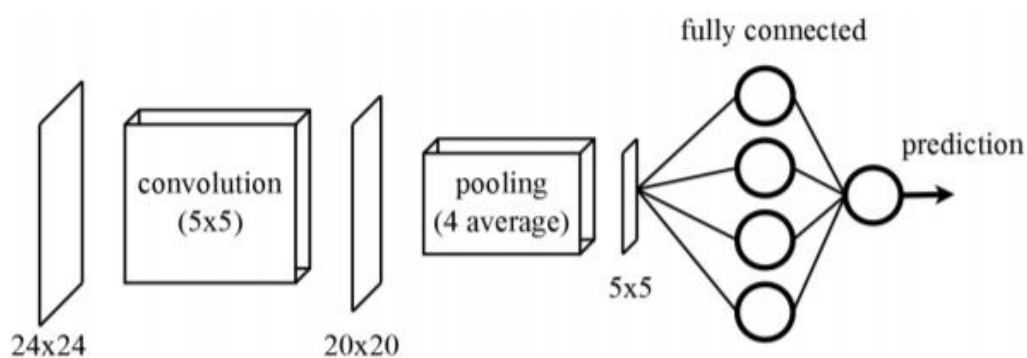
ประสาทเทียมที่แตกต่างกันรวมทั้งข้อมูลที่ได้รับเข้ามาในตัวแบบ และในบทนี้ได้ยกตัวอย่างงานวิจัยที่เกี่ยวข้องไว้ดังนี้

PupilNet: Convolutional Neural Networks for Robust Pupil Detection

ในปี 2016 Fuhr และคณะ และคณะนำเสนอการตรวจจับรูม่านตาด้วยการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน 2 ตัวแบบ ตัวแบบแรกมีจุดประสงค์เพื่อระบุตำแหน่งจุดศูนย์กลางรูม่านตาแบบหยาบ โดยพิจารณาจากภาพที่ถูกลดความละเอียดลงเพื่อลดต้นทุนการคำนวณและลดผลกระทบจากสัญญาณรบกวนที่อาจเกิดขึ้นจากสาเหตุต่าง ๆ เช่น การเปลี่ยนแปลงความสว่าง การสะท้อนของแสง การบดบังที่เกิดจากส่วนประกอบอื่น ๆ ของตา เป็นต้น ตัวแบบที่สองรับตำแหน่งจุดศูนย์กลางรูม่านตาแบบหยาบที่ได้จากตัวแบบแรกมาเพื่อช่วยในการระบุตำแหน่งจุดศูนย์กลางรูม่านตาอย่างละเอียดอีกครั้งด้วยการพิจารณาจากความละเอียดของภาพต้นฉบับ

การระบุตำแหน่งจุดศูนย์กลางรูม่านตาแบบหยาบ

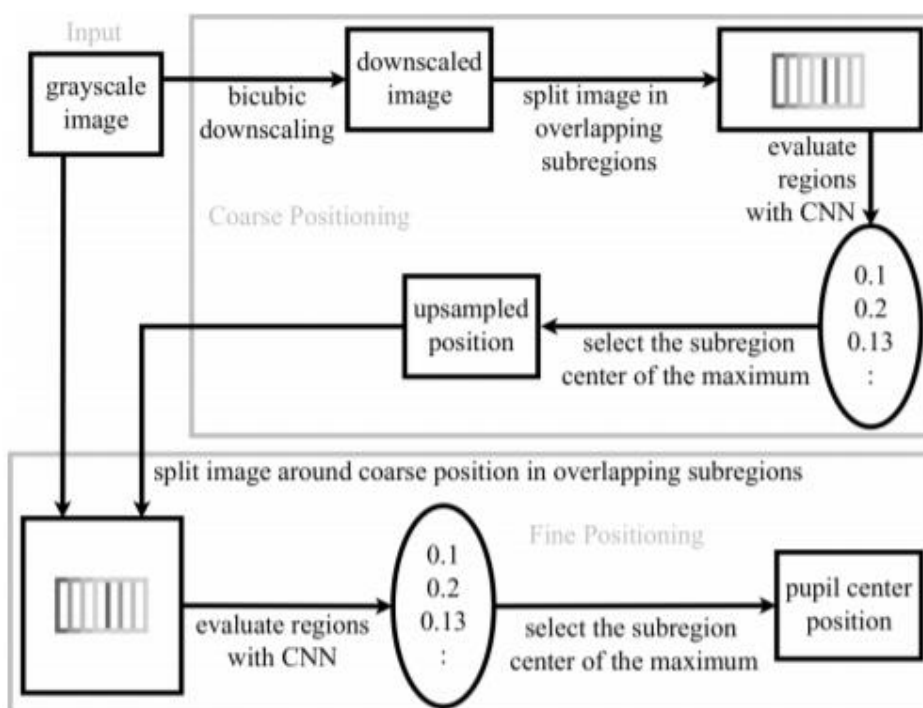
งานวิจัยนี้เลือกใช้วิธีการลดความละเอียดของภาพลงด้วยวิธีไบคิวบิก (Bicubic) ทำให้ภาพเดิมซึ่งมีความละเอียด 384×288 พิกเซล มีความละเอียด 96×72 พิกเซล ต่อมาแบ่งภาพเป็นรูปย่อย ๆ ขนาด 24×24 พิกเซล เพื่อให้มั่นใจว่าครอบคลุมบริเวณรูม่านตา ภาพย่อยแต่ละภาพถูกแบ่งโดยเว้นระยะห่าง 1 พิกเซล และจะถูกนำไปคำนวณค่าความเชื่อมั่นของการมีรูม่านตาอยู่ในภาพด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชันซึ่งประกอบไปด้วย ชั้นคอนโวลูชัน (Convolutional Layer) ขนาดตัวกรอง (Filter size) เท่ากับ 5×5 , ชั้นพูลลิงแบบค่าเฉลี่ย (Average mean pooling Layer) ขนาด 4×4 และชั้นการเชื่อมต่อเต็มรูปแบบ (Fully Connected Layer) ดังแสดงในภาพที่ 4 ซึ่งจุดศูนย์กลางของภาพย่อยที่มีค่าความเชื่อมั่นสูงที่สุดจะถูกพิจารณาเป็นจุดศูนย์กลางรูม่านตาแบบหยาบ



ภาพที่ 2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้ระบุตำแหน่งจุดศูนย์กลางรูม่านตาแบบหยาบ (ที่มาของภาพ Fuhr และคณะ [8])

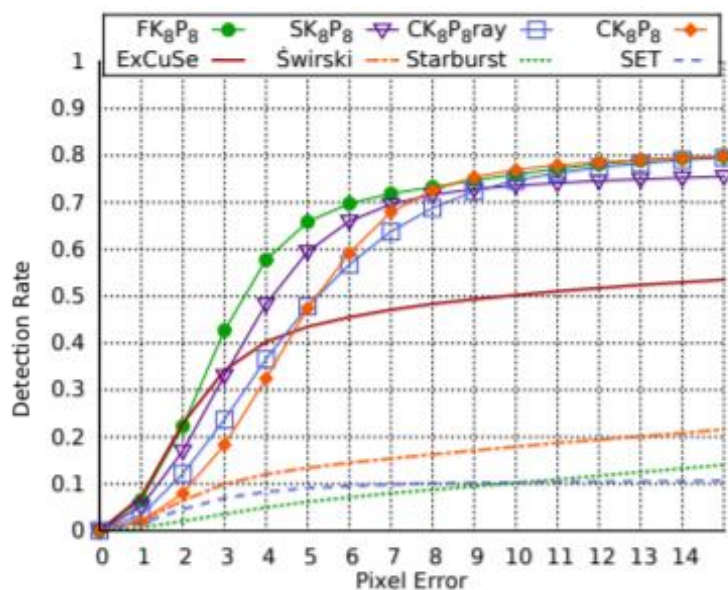
การระบุตำแหน่งจุดศูนย์กลางรูม่านตาแบบละเอียด

ตัวแบบที่สองซึ่งใช้ระบุตำแหน่งจุดศูนย์กลางรูม่านตาแบบละเอียดมีโครงสร้างคล้ายกับโครงสร้างของตัวแบบแรกต่างกันที่ขนาด เนื่องจากตัวแบบที่สองพิจารณาภาพย่อยที่ความละเอียดต้นฉบับ โดยภาพย่อยที่อยู่บริเวณตำแหน่งจุดศูนย์กลางรูม่านตาแบบหยาบจะถูกส่งไปหาค่าความเชื่อมั่นต่อในโครงข่ายคอนโวลูชันตัวแบบที่สอง ในทำนองเดียวกันจุดศูนย์กลางของภาพย่อยที่มีความเชื่อมั่นสูงที่สุดคือตำแหน่งจุดศูนย์กลางรูม่านตาแบบละเอียด



ภาพที่ 2.2 ภาพรวมของระบบที่ใช้ในงานวิจัย (ที่มาของภาพ Fuhl และคณะ [8])

Fuhl และคณะ ประเมินสมรรถนะของวิธีการด้วยค่าเฉลี่ยของอัตราการตรวจพบภาพที่มีระยะความคลาดเคลื่อนของตำแหน่งที่ทำนายได้และตำแหน่งจริงห่างกันน้อยกว่าค่าที่สอดคล้องกับค่าในแกน x (Detection Rate) สังเกตว่าวิธีการที่ผู้วิจัยได้นำเสนอ (FK8P8, SK8P8, CK8P8ray, CK8P8) มีประสิทธิภาพมากกว่าวิธีการซึ่งเป็น state-of-the-art (ExCuSe, SET, Starburs และ Swirski) อย่างชัดเจน



ภาพที่ 2.3 Detection Rate ของวิธีการที่ผู้วิจัยนำเสนอเปรียบเทียบกับวิธีการซึ่งเป็น state-of-the-art ทั้งหมด 4 วิธี (ExCuSe, SET, Starburs และ Swirski)
(ที่มาของภาพ Fuhl และคณะ [8])

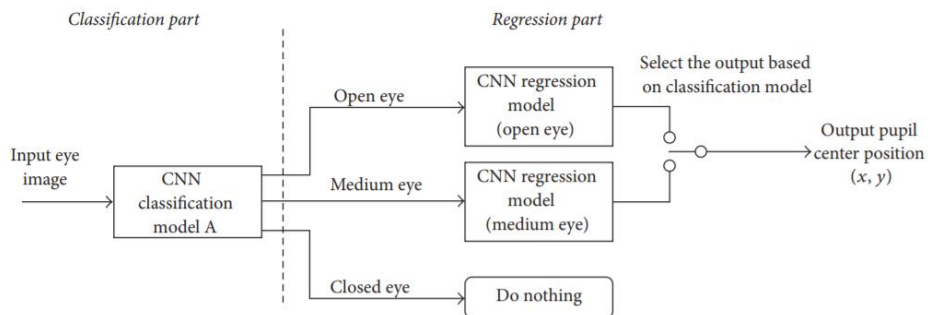
อย่างไรก็ตามวิธีการที่ผู้วิจัยได้นำเสนอยังมี Detection Rate ไม่สูงมากนัก โดยถ้าพิจารณา ค่าความคลาดเคลื่อนที่ยอมรับได้คือค่าที่น้อยกว่า 5 พิกเซลตามที่คุณผู้วิจัยได้นำเสนอ วิธีที่ดีที่สุดมี Detection Rate เพียง 68%

CNN-Based Pupil Center Detection for Wearable Gaze Estimation System

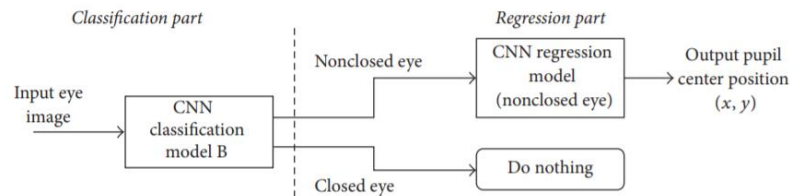
ในปี 2017 Chinsatit และคณะ [9] นำเสนอวิธีระบุตำแหน่งจุดศูนย์กลางรูม่านตาเพื่อทำนายตำแหน่งการเพ่งมองโดยใช้ภาพอินฟราเรดของดวงตาเป็นข้อมูลฝึก งานวิจัยนี้เลือกใช้ตัวแบบการเรียนรู้เชิงลึก 2 ตัวแบบ ตัวแบบแรกมีจุดประสงค์เพื่อจำแนกประเภทภาพตามสถานะของตา (Classification part) และตัวแบบที่สองใช้การวิเคราะห์การถดถอยเพื่อระบุตำแหน่งจุดศูนย์กลางรูม่านตา (Regression part) โดยผู้วิจัยทำการทดลอง 2 วิธีการ ได้แก่

วิธีการ A : ตัวแบบแรกจะจำแนกภาพตาตามสถานะของตาเป็น 3 ประเภท ได้แก่ 1. ภาพที่ปรากฏรูม่านตาชัดเจน (Open eye) 2. ภาพที่ปรากฏรูม่านตาไม่ชัดเจน (Medium eye) 3. ภาพที่ไม่ปรากฏรูม่านตา (Closed eye) เพื่อกรองประเภทภาพที่ไม่ปรากฏรูม่านตาออกแล้วส่งภาพที่เหลือ (ภาพที่ปรากฏรูม่านตาชัดเจนและภาพที่ปรากฏรูม่านตาไม่ชัดเจน) ไปยังตัวแบบที่สองเพื่อระบุตำแหน่งจุดศูนย์กลางรูม่านตาตามสถานะที่แบ่งไว้ในตัวแบบแรก

วิธีการ B : ตัวแบบแรกจะจำแนกภาพตาตามสถานะของตาเป็น 2 ประเภท ได้แก่ 1. ภาพที่ปรากฏรูม่านตา (Nonclosed eye) 2. ภาพที่ไม่ปรากฏรูม่านตา (Closed eye) เพื่อกรองประเภทภาพที่ไม่ปรากฏรูม่านตาออกแล้วส่งภาพที่ปรากฏรูม่านตาไปยังตัวแบบที่สองเพื่อระบุตำแหน่งจุดศูนย์กลางรูม่านตา



(a) The proposed method A



(b) The proposed method B

ภาพที่ 2.4 ภาพรวมของระบบในวิธีการ A และ B (ที่มาของภาพ Chinsatit และคณะ [9])

รายละเอียดโครงสร้างของตัวแบบทั้งสองแสดงในตารางที่ 2.1

ตารางที่ 2.1 รายละเอียดโครงสร้างของตัวแบบในแต่ละขั้นตอนวิธี

Name	Item	Conv1	Conv2	Conv3	Conv4	Conv5	Full1	Full2	Out
Classification model A	Channel	48	128	192	192	128	1024	1024	3 classes
	Filter size	11 × 11	5 × 5	3 × 3	3 × 3	3 × 3	—	—	—
	Pooling size	2 × 2	2 × 2	2 × 2	2 × 2	3 × 3	—	—	—
	Normalization	yes	—	—	—	—	Yes	Yes	—
	Dropout	—	—	—	—	—	Yes	Yes	—
Classification model B	Channel	48	128	192	192	128	1024	1024	2 classes
	Filter size	11 × 11	5 × 5	3 × 3	3 × 3	3 × 3	—	—	—
	Pooling size	2 × 2	2 × 2	2 × 2	2 × 2	3 × 3	—	—	—
	Normalization	yes	—	—	—	—	Yes	Yes	—
	Dropout	—	—	—	—	—	Yes	Yes	—
Regression model	Channel	96	256	512	512	512	4096	4096	2 reg.
	Filter size	7 × 7	5 × 5	3 × 3	3 × 3	3 × 3	—	—	—
	Pooling size	3 × 3	2 × 2	—	—	3 × 3	—	—	—
	Normalization	yes	—	—	—	—	Yes	Yes	—
	Dropout	—	—	—	—	—	Yes	Yes	—

ฟังก์ชันต้นทุน (Cost function) ที่ใช้ในตัวอย่างสำหรับจำแนกประเภท คือ

$$\text{Mean squared error} = \frac{\sum_{i=1}^{N_o} (o_i - b_i)^2}{N_o} \quad \text{สมการที่ 2.1}$$

โดย o_i คือ ค่าที่ตัวแบบทำนายได้ของภาพ i , b_i คือ ค่าจริงของภาพ i และ N_o คือ จำนวนผลลัพธ์ทั้งหมด

ตัวแบบถดถอย (Regression model) ที่ใช้ระบุตำแหน่งจุดศูนย์กลางรูปร่างประกอบไปด้วย ชั้นคอนโวลูชัน 5 ชั้น ชั้นการเชื่อมต่อเต็มรูปแบบ 3 ชั้น และใช้การดรอปเอาต์ (Dropout), พูลลิง (Pooling) และนอร์มัลไลเซชัน (Normalization) ตามที่แสดงในตารางที่ 2.1 โดยในชั้นซ่อนตัวทั้งหมดใช้ฟังก์ชันเรกทิไฟเออร์เชิงเส้น (Rectified Linear Unit; ReLU) เป็นฟังก์ชันกระตุ้น (Activation Function) และใช้วิธีการนอร์มัลไลเซชันดังนี้

$$x'_k = \frac{x_k}{\sqrt{\sum_{i=1}^{N_i} x_i^2}} \quad \text{สมการที่ 2.2}$$

โดยที่ k คือตำแหน่งโหนดของข้อมูลเข้า, x_k คือ ข้อมูลเข้าที่โหนด k , x'_k คือ ผลลัพธ์จากการนอร์มัลไลเซชันที่โหนด k และ N_i คือจำนวนสมาชิกของข้อมูลในแต่ละชั้น

ข้อมูลที่น่ามาใช้ฝึกตัวแบบ ถูกบันทึกโดยใช้กล้องอินฟราเรด โดยผู้วิจัยเก็บภาพขณะกระพริบตา เพื่อให้ภาพดวงตาในแต่ละสถานะมีจำนวนใกล้เคียงกัน ทั้งนี้กล้องอินฟราเรดจะช่วยให้ภาพดวงตาที่เก็บได้มีความแตกต่างของสีบริเวณรูม่านตาและม่านตามากขึ้น

การประเมินผลตัวแบบผู้วิจัยเลือกใช้วิธี leave one out เพื่อวิเคราะห์ซ้ำ (Cross validation) และใช้ตัวแบบที่ผ่านการฝึกมาแล้วจากฐานข้อมูล ImageNet เพื่อหลีกเลี่ยงเหตุการณ์ที่ตัวแบบอิงกับข้อมูลเรียนรู้มากเกินไป (Overfitting) ผลการประเมินตัวแบบสำหรับจำแนกประเภทภาพตามสถานะตาของวิธีการ A และ B เป็นไปตามตารางที่ 2.2

ตารางที่ 2.2 ผลการประเมินตัวแบบสำหรับจำแนกประเภทภาพตามสถานะตาของวิธี A และ B

(a) Classification result of method A				
	Open	Predict Medium	Closed	Accuracy
Actual Open	5465	1013	48	83.64%
Medium	784	4595	855	73.67%
Closed	0	707	6133	89.66%

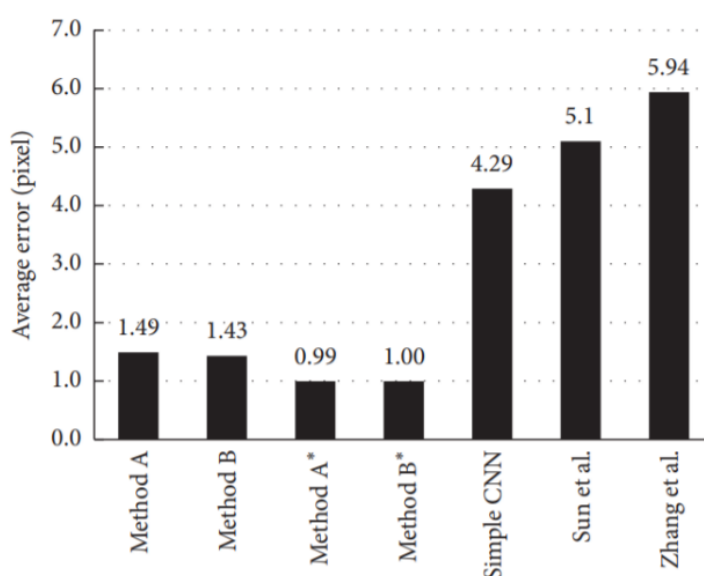
(b) Classification result of method B				
	Nonclosed	Predict Closed		Accuracy
Actual Nonclosed	6596	244		96.43%
Closed	776	6064		88.65%

จากตารางพบว่าการจำแนกภาพตาที่อยู่ในสถานะปรากฏรูม่านตาไม่ชัดเจนมีประสิทธิภาพต่ำกว่าการจำแนกภาพตาในสถานะอื่น ซึ่งมีความสมเหตุสมผลเนื่องจากเป็นสถานะที่ยากต่อการแยกแยะ สังเกตได้จากภาพที่ 2.5 อย่างไรก็ตามจากผลการทดลองยังถือว่าทั้งตัวแบบที่ใช้ในวิธีการ A และ B มีประสิทธิภาพสูงในการคัดกรองภาพตาที่อยู่ในสถานะที่ไม่ปรากฏรูม่านตาออก



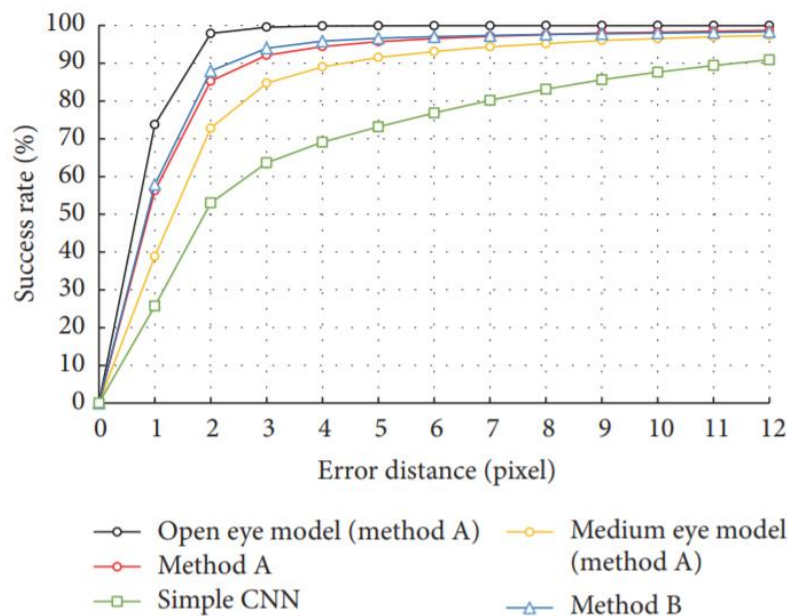
ภาพที่ 2.5 ตัวอย่างภาพที่ทำนายผิดพลาด

การประเมินผลตัวแบบที่ใช้สำหรับระบุตำแหน่งจุดศูนย์กลางรูمانةตา ผู้วิจัยเปรียบเทียบความแม่นยำของตัวแบบในวิธีการ A, B, A* และ B* โดยที่วิธีการ A* และ B* ผู้วิจัยให้นิยามว่าเป็นวิธีการเดียวกับ A และ B ในกรณีที่ขั้นตอนจำแนกประเภทภาพตามสถานะของตามีความถูกต้อง 100 เปอร์เซ็นต์ นอกจากนี้ได้เปรียบเทียบวิธีการระบุจุดศูนย์กลางรูمانةตาเมื่อไม่มีการจำแนกประเภทตามสถานะของตาด้วย (Simple CNN) และเพื่อเปรียบเทียบกับวิธีที่มีมาก่อนหน้า ผู้วิจัยนำวิธีระบุตำแหน่งจุดศูนย์กลางรูمانةตาของ Sun และคณะ และ Zhang และคณะ มาแสดงผลร่วมด้วยในแผนภูมิภาพที่ 2.6



ภาพที่ 2.6 ค่าความคลาดเคลื่อนเฉลี่ยของแต่ละวิธีการ

จากภาพที่ 2.6 แสดงให้เห็นอย่างชัดเจนว่าวิธีการระบุจุดศูนย์กลางรูمانةตาที่มีการจำแนกประเภทภาพตามสถานะของตาก่อน มีความแม่นยำมากกว่าวิธีที่ไม่มีการจำแนกประเภทภาพตามสถานะของตา และยังมีค่าความคลาดเคลื่อนน้อยกว่าวิธีของ Sun และคณะ และ Zhang และคณะ ด้วย และภาพที่ 2.7 แสดงอัตราความสำเร็จของวิธีที่ผู้วิจัยนำเสนอ (Success rate)



ภาพที่ 2.7 อัตราความสำเร็จของวิธีที่ผู้วิจัยนำเสนอ โดยค่าในแกน y คือร้อยละของภาพที่มีความคลาดเคลื่อนน้อยกว่าค่าที่สัมพันธ์กันในแกน x เมื่อ x คือระยะห่างในช่วง 0-12 พิกเซล

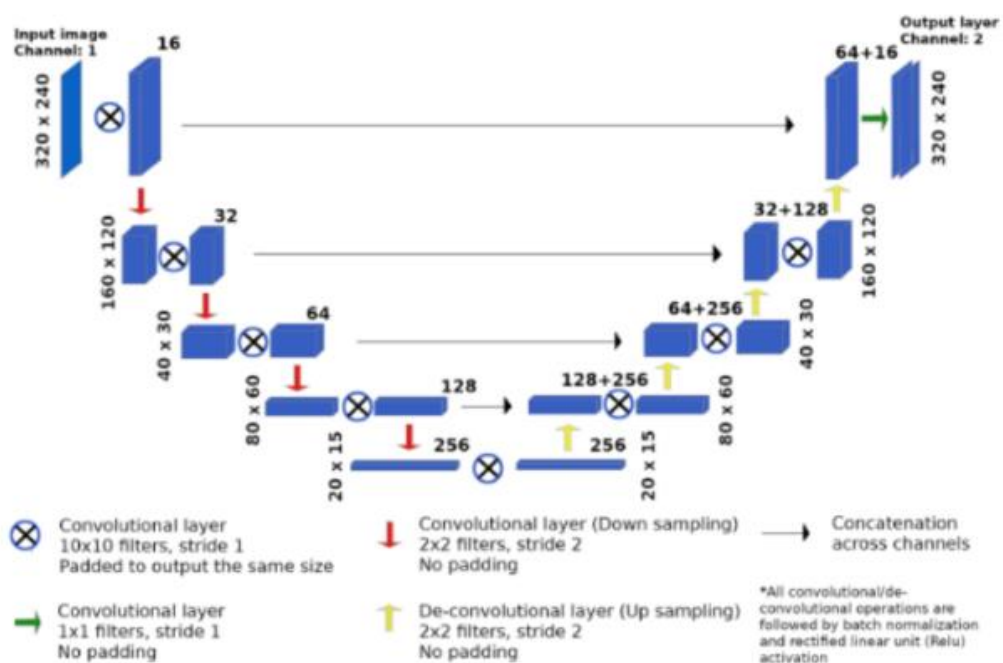
จะเห็นว่าวิธีการที่ดีที่สุดที่ผู้วิจัยนำเสนอให้ความแม่นยำที่สูงมาก อย่างไรก็ตามภาพดวงตาที่ผู้วิจัยนำมาทดลองมีส่วนสำคัญที่ทำให้การระบุตำแหน่งจุดศูนย์กลางรูม่านตาสามารถทำได้ง่ายขึ้น เนื่องจากภาพดวงตาที่ถ่ายด้วยกล้องอินฟราเรด สีของรูม่านตาและม่านตาจะแตกต่างกันชัดเจนมากขึ้น

DeepVOG : Open-source pupil segmentation and gaze estimation in neuroscience using deep learning

ในปี 2019 Yiu และคณะ [10] นำเสนอโครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับแยกพื้นที่รูม่านตาโดยได้รับแรงบันดาลใจจากสถาปัตยกรรม U-net [11] ซึ่งเป็นวิธีการทำ semantic segmentation ที่พิจารณาภาพแบบหยาบและละเอียดประกอบกัน ผู้วิจัยตั้งใจใช้ประโยชน์จากการแยกพื้นที่รูม่านตาเพื่อระบุจุดศูนย์กลางรูม่านตา โดยวาดวงกลมล้อมพื้นที่รูม่านตาที่แยกได้และใช้จุดศูนย์กลางของวงกลมเพื่อระบุจุดศูนย์กลางรูม่านตา นอกจากนี้ผู้วิจัยได้ใช้ขั้นตอนวิธีในงานของ Świrski and Dodgson, 2013 [12] มาทำนายตำแหน่งการเพ่งมองด้วย

สำหรับโครงสร้างตัวแบบการเรียนรู้เชิงลึกที่ผู้วิจัยเลือกใช้แสดงในภาพที่ 2.8 ประกอบไปด้วยส่วนการเข้ารหัส 4 ชั้น และถอดรหัส 4 ชั้น การเข้ารหัสแต่ละชั้นประกอบไปด้วยชั้นคอนโวลูชัน

ขนาด 10×10 ด้วยช่วงก้าว (Stride) เท่ากับ 1, การนอร์มัลไลเซชัน, ฟังก์ชันเรกทิไฟเออร์เชิงเส้น (ReLU), ชั้นคอนโวลูชัน ขนาด 2×2 ด้วยช่วงก้าวเท่ากับ 2, การนอร์มัลไลเซชันและฟังก์ชันเรกทิไฟเออร์เชิงเส้น สำหรับการถอดรหัสเพื่อเพิ่มขนาดของข้อมูลและสอดแทรกข้อมูลในชั้นก่อนหน้า แต่ละชั้นประกอบไปด้วย ชั้นคอนโวลูชัน ขนาด 2×2 ด้วยช่วงก้าวเท่ากับ 2, การนอร์มัลไลเซชัน, ฟังก์ชันเรกทิไฟเออร์เชิง, ชั้นคอนโวลูชัน ขนาด 10×10 ด้วยช่วงก้าวเท่ากับ 1, การนอร์มัลไลเซชัน และฟังก์ชันเรกทิไฟเออร์เชิงเส้น หลังจากผ่านขั้นตอนการเข้ารหัสและถอดรหัสแล้วจะผ่านชั้นคอนโวลูชันขนาด 1×1 ด้วยช่วงก้าวเท่ากับ 1 และชั้นข้อมูลออก 2 ชั้นซึ่งมีขนาดเท่ากับชั้นข้อมูลเข้า



ภาพที่ 2.8 สถาปัตยกรรมที่ใช้ในงานวิจัยนี้ (ที่มาของภาพ Yiu และคณะ [10])

การฝึกสอนใช้ภาพทั้งหมด 3,946 ภาพ ซึ่งได้มาจากวิดีโอความละเอียด 320×240 พิกเซล ถ่ายในท้อง MRI ที่สภาวะแสงน้อยและแสงปกติ โดยมีการขยายข้อมูล (Data augmentation) ด้วยวิธีการหมุน ยืด ขยาย กลับด้าน และใช้ข้อมูลตรวจสอบจำนวน 2,002 ภาพ การทดสอบความแม่นยำของตัวแบบใช้ภาพจากชุดข้อมูลเดียวกันและใช้ภาพจากฐานข้อมูลอื่น ๆ ด้วยเพื่อทดสอบความคงทน

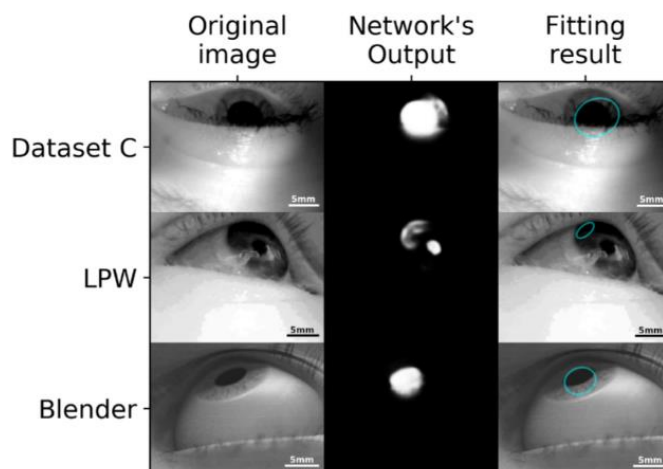
ความคลาดเคลื่อนของพื้นที่รูปร่างที่ตัวแบบแยกได้และพื้นที่จริงในตารางที่ 2.3 ถูกประเมินด้วยค่ามัธยฐานของดัชนีความเหมือนของไดซ์ (Dice's similarity coefficient), ระยะทางแบบยูคลิด (Euclidean distance) และวิธีวัดระยะทางแบบเฮาส์ดอร์ฟฟ์ (Hausdorff distance) ซึ่ง

หากพื้นที่ที่แยกได้มีความคล้ายกับพื้นที่จริงมากค่าที่คำนวณได้จะมีค่าเข้าใกล้ศูนย์ จากตารางจะเห็นว่าความแม่นยำของการระบุจุดศูนย์กลางรูม่านตามีความแม่นยำสูงทั้งกับข้อมูลที่คล้ายชุดฝึกและข้อมูลจากแหล่งข้อมูลอื่น ๆ โดยมาตรวัดแบบยุคลิดวัดระยะทางคลาดเคลื่อนจากจุดศูนย์กลางของพื้นที่รูม่านตาที่แยกได้และจุดศูนย์กลางของพื้นที่จริงเฉลี่ยได้น้อยกว่า 2 พิกเซล

ตารางที่ 2.3 ความแตกต่างของพื้นที่รูม่านตาที่ตัวแบบแยกได้และพื้นที่จริง

Datasets	Dice's coefficient	Euclidean distance (px)	Hausdorff distance (px)
Dataset A & B (1892)	0.966 (0.948-0.976)	1.0 (0.6-1.6)	2.8 (2.0-4.0)
Dataset C (323)	0.97 (0.958-0.978)	0.9 (0.6-1.3)	2.8 (2.0-3.2)
Delhi (763)	0.978 (0.971-0.983)	0.8 (0.5-1.2)	2.8 (2.0-3.0)
LPW (466)	0.938 (0.914-0.957)	0.9 (0.6-1.4)	3.2 (2.2-4.0)
MMU Iris (167)	0.958 (0.947-0.968)	1.0 (0.6-1.4)	2.2 (2.1-3.0)
Blender (361)	0.965 (0.901-0.982)	1.8 (1.3-2.4)	3.6 (2.0-8.2)

อย่างไรก็ตามแม้ว่าค่าความแม่นยำของการระบุจุดศูนย์กลางรูม่านตาให้ผลลัพธ์ที่น่าพึงพอใจ แต่การแยกพื้นที่รูม่านตาขนาดใหญ่หรือรูม่านตาซึ่งถูกล้อมรอบด้วยบริเวณที่มีแสงสว่างน้อย พบว่าจะเกิดความคลาดเคลื่อนดังแสดงในภาพที่ 2.9 ส่งผลให้กระบวนการหาจุดศูนย์กลางรูม่านตามีความแม่นยำน้อยลงไปด้วย

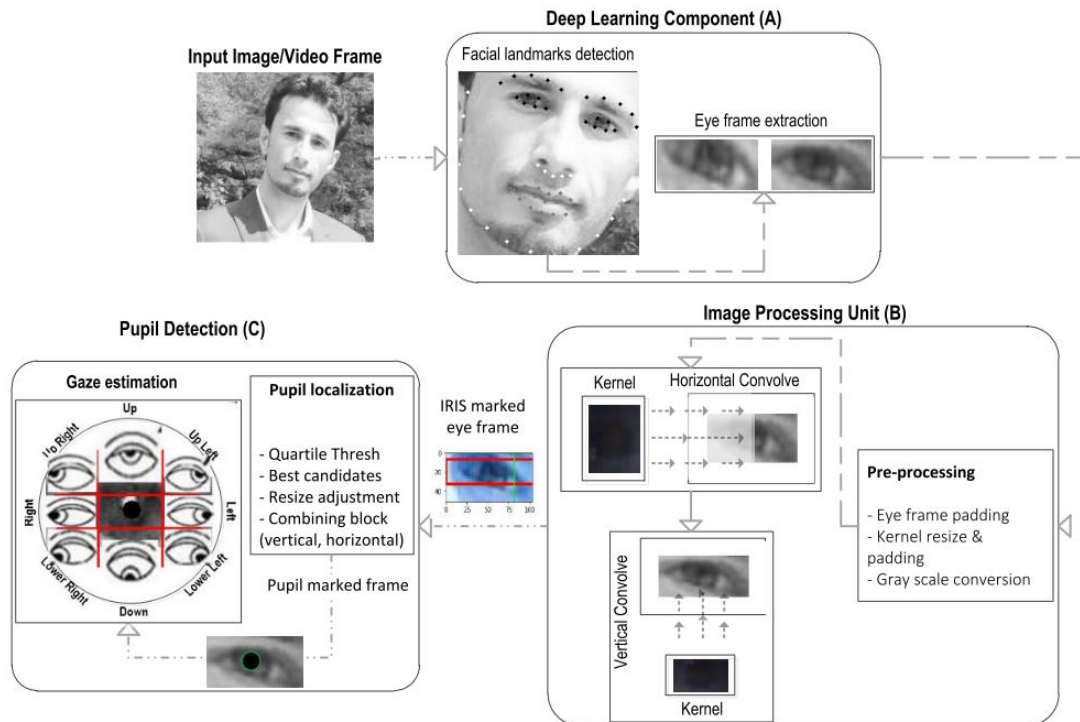


ภาพที่ 2.9 ผลลัพธ์ที่ผิดพลาดจากภาพที่มีอุปสรรค (ที่มาของภาพ Yiu และคณะ [10])

Pupil Localisation and Eye Centre Estimation Using Machine Learning and Computer Vision

ในปี 2020 Khan และคณะ [13] นำเสนอวิธีการระบุตำแหน่งรูม่านตาสำหรับภาพที่มีความละเอียดต่ำ โดยใช้กระบวนการเรียนรู้เชิงลึกร่วมกับวิธีการประมวลผลภาพ ภาพรวมของกระบวนการที่ใช้ในงานวิจัยแสดงดังภาพที่ 2.10 โดยมีขั้นตอนหลัก ๆ 3 ขั้นตอน คือ A) สกัดบริเวณของตาจาก

ภาพใบหน้าโดยใช้กระบวนการเรียนรู้เชิงลึก B) ระบุตำแหน่งม่านตาโดยวิธีการประมวลผลภาพ C) ระบุตำแหน่งรูม่านตา

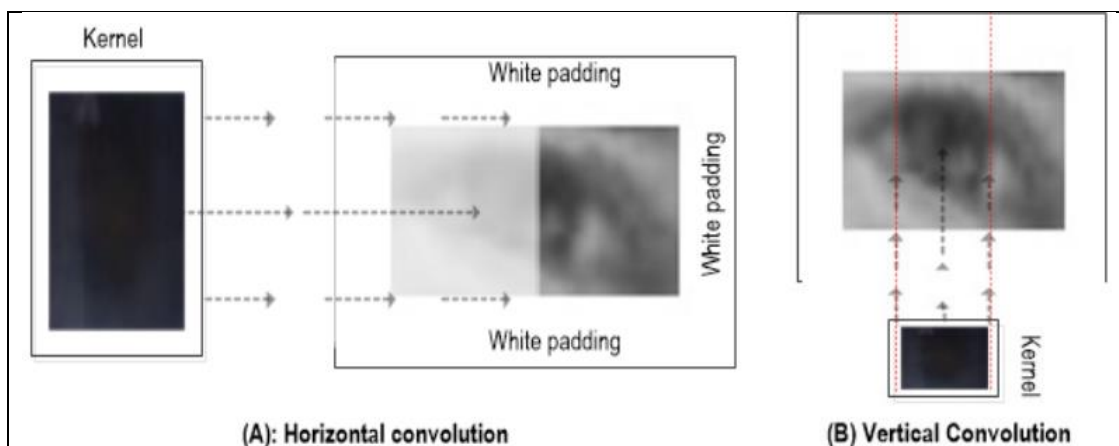


ภาพที่ 2.10 ภาพรวมของกระบวนการที่ใช้ในงานวิจัย (ที่มาของภาพ Khan และคณะ [13])

สำหรับขั้นตอนการสกัดบริเวณดวงตาจากภาพใบหน้า งานวิจัยได้ใช้ตัวแบบการเรียนรู้เชิงลึกตรวจจับอวัยวะต่าง ๆ บนใบหน้า ในไลบรารี Dlib โดยผลลัพธ์คือพิกัดตำแหน่งของอวัยวะต่าง ๆ บนใบหน้าจำนวน 68 จุด ซึ่งรวมไปถึงจุดตำแหน่งของม่านตา เปลือกตาด้วย ดังนั้นจึงสามารถครอบตัดบริเวณดวงตาได้อย่างง่ายดาย ในขั้นตอนนี้นอกจากจะช่วยกำจัดองค์ประกอบอื่น ๆ บนใบหน้าแล้วยังเป็นขั้นตอนสำคัญในการกำจัดสิ่งรบกวนที่สามารถทำให้อัลกอริทึมสับสนว่าเป็นดวงตาด้วย เช่น บริเวณคิ้ว เป็นต้น

ขั้นตอนการระบุตำแหน่งม่านตา ผู้วิจัยใช้คอนโวลูชันฟังก์ชันในการกรองภาพเพื่อหาส่วนที่เหมาะสมที่สุดที่จะเป็นม่านตาในภาพ โดยผู้วิจัยสร้างเคอร์เนลที่ปรับเปลี่ยนได้ (custom kernel) ประโยชน์จากการใช้เคอร์เนลที่ปรับเปลี่ยนได้แทนที่จะใช้เคอร์เนลแบบดั้งเดิมที่มีสีดำคือ เคอร์เนลจะสามารถนำไปใช้ได้ในกรณีทั่วไปมากกว่า โดยจะสามารถแทนม่านตาของกลุ่มประชากรที่หลากหลาย รวมถึงไปถึงการมีลักษณะทางสัณฐานวิทยาที่หลากหลายกว่า นอกจากนี้ถ้าเคอร์เนลมีขนาดใหญ่ไปก็จะนำไปสู่การได้ส่วนภาพที่ผิดพลาดได้ และถ้าเล็กไปก็อาจจะได้มาซึ่งส่วนภาพที่มาจากสิ่งรบกวน การ

ใช้เคอร์เนลที่ปรับเปลี่ยนได้จึงได้ใช้มาเพื่อแก้ปัญหาต่าง ๆ นี้ โดยขนาดของเคอร์เนลจะถูกกำหนดด้วยขนาดของภาพข้อมูลเข้า (บริเวณดวงตาที่ถูกครอบตัดมาแล้ว) นอกจากนี้ภาพดวงตาที่ถูกครอบตัดมา จะถูกเติมด้วยสีขาวบริเวณขอบภาพ เพื่อขยายขนาดให้เพียงพอสำหรับการทำคอนโวลูชันด้วย



ภาพที่ 2.11 การทำคอนโวลูชันตามแนวแกนนอนและการทำคอนโวลูชันตามแนวแกนตั้ง

(ที่มาของภาพ Khan และคณะ [13])

โดยทั่วไปการทำคอนโวลูชัน เคอร์เนลจะถูกเลื่อนทั้งตามแนวแกนนอนและแนวแกนตั้งพร้อม ๆ กัน แต่ในงานวิจัยนี้จะทำคอนโวลูชันตามแนวนอนและแนวตั้งตามภาพที่ 2.11 จากสมมติฐานว่ารูม่านตาจะเป็นส่วนในภาพที่มีมืดที่สุด และมันตามีลักษณะเป็นทรงกลม เคอร์เนลตามแนวนอน จะถูกปรับขนาดความสูงตามความสูงของภาพตาที่ถูกครอบตัด (crop) มา โดยมีความกว้างเท่ากับ 0.4 เท่าของความกว้างของภาพตา หลังจากทำคอนโวลูชันตามแนวนอนจะได้ตำแหน่งมุมบนซ้ายในแนวแกน x ที่เหมาะสมที่สุดที่จะเป็นตำแหน่งม่านตา มา จากนั้นจึงทำคอนโวลูชันตามแนวตั้งต่อ โดยความสูงของเคอร์เนลจะถูกกำหนดให้เท่ากับ 0.4 เท่าของความสูงของภาพตา และความกว้างมีขนาดเท่ากับ ความกว้างของเคอร์เนลในแนวนอน หลังจากทำคอนโวลูชันจะได้ค่าในแนวแกน y ของตำแหน่งมุมบนซ้ายของบริเวณม่านตา ซึ่งในขั้นตอนนี้จะทำให้ได้บริเวณม่านตาที่มีขนาดเท่ากับเคอร์เนลในแนวแกนตั้ง ดังนั้นในบริเวณม่านตาซึ่งเป็นบริเวณสีเหลี่ยมนี้จึงสามารถคำนวณบริเวณกึ่งกลางเพื่อมาอนุมานเป็นตำแหน่งจุดศูนย์กลางรูม่านตาได้

การทดสอบในงานวิจัยนี้ได้เลือกใช้มาตรวัดทั่วไป เช่น ระยะทางแบบยุคลิด สัมประสิทธิ์การกำหนด (R^2) ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) และยังใช้มาตรวัด Wec และ Bec เพื่อเปรียบเทียบกับงานวิจัยอื่น ๆ ที่เกี่ยวข้องด้วย ยิ่งไปกว่านั้นผู้วิจัยยังได้นำเสนอมาตรวัด S_{ED} เพื่อเป็นมาตรวัดมาตรฐานสำหรับงานที่เกี่ยวข้องกับการระบุจุดศูนย์กลางรูม่านตาด้วย

สำหรับมาตรวัด Wec และ Bec ได้ถูกนำเสนอใน [14] และใช้งานอย่างแพร่หลาย ซึ่งมีสูตรการคำนวณดังสมการที่ 2.3 มาตรวัด Wec ใช้ค่ามากที่สุดระหว่างระยะทางคลาดเคลื่อนที่ได้จากการระบุจุดศูนย์กลางตาของตาซ้ายและระยะทางคลาดเคลื่อนที่ได้จากการระบุจุดศูนย์กลางตาของตาขวา มาแปลงให้เป็นค่ามาตรฐานโดยนอร์มอลไลซ์ด้วยระยะทางระหว่างจุดศูนย์กลางรูม่านตาซ้ายและจุดศูนย์กลางรูม่านตาขวา ส่วนการคำนวณมาตรวัด Bec สามารถทำได้ในทำนองเดียวกัน เพียงแต่เปลี่ยนจากการใช้ค่ามากที่สุดเป็นค่าน้อยสุด ดังนั้นค่าของมาตรวัด Wec และ Bec นั้นยังเข้าใกล้ศูนย์หมายความว่ายิ่งแม่นยำ

$$Wec = \frac{\max(\|\tilde{C}_l - C_l\|, \|\tilde{C}_r - C_r\|)}{\|C_l - C_r\|} \quad \text{สมการที่ 2.3}$$

$$Bec = \frac{\min(\|\tilde{C}_l - C_l\|, \|\tilde{C}_r - C_r\|)}{\|C_l - C_r\|} \quad \text{สมการที่ 2.4}$$

เมื่อ C_l, C_r คือค่าจริงของจุดศูนย์กลางของรูม่านตาซ้ายและขวา และ \tilde{C}_l, \tilde{C}_r คือค่าที่ทำนายได้ของจุดศูนย์กลางของรูม่านตาซ้ายและขวา

และสำหรับมาตรวัด S_{ED} ผู้วิจัยได้นำเสนอขึ้นเป็นครั้งแรกเพื่อเป็นมาตรวัดมาตรฐานแบบที่ไม่ขึ้นกับตาอีกข้างบนใบหน้า สามารถคำนวณได้จาก

$$S_{ED} = \frac{\sqrt{(Cx_e - Cx_a)^2 + (Cy_e - Cy_a)^2}}{\sqrt{(x_{min} - x_{max})^2 + (y_{min} - y_{max})^2}} \times 100$$

เมื่อ Cx_e และ Cx_a แทนค่าที่ทำนายได้และค่าจริงตามแนวแกน x และ Cy_e และ Cy_a แทนค่าที่ทำนายได้และค่าจริงตามแนวแกน y และ x_{min}, y_{min} คือ พิกัดจุดของมุมที่ใกล้ที่สุดของกรอบตา และ x_{max}, y_{max} คือ พิกัดจุดของมุมที่ไกลที่สุดของกรอบตา

ผลลัพธ์ของงานวิจัยนี้แสดงให้เห็นดังตารางที่ ตารางที่ 2.4 และตารางที่ 2.5

ตารางที่ 2.4 ประสิทธิภาพของวิธีการที่นำเสนอโดยใช้มาตรวัด Wec และ Bec

Dataset	Wec (%)		Bec (%)	
	Error ≤ 0.05	Error ≤ 0.1	Error ≤ 0.05	Error ≤ 0.1
BIO-ID	94.5	100	98.34	100
Talking-Face	97.10	100	99.7	100
GI4E	95.05	100	98.71	100

ตารางที่ 2.5 ตารางเปรียบเทียบประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนสัมบูรณ์ สัมประสิทธิ์การกำหนด ระยะทางแบบยุคลิด และ มาตรฐาน S_{ED}

Dataset	$\mu x_a-x_e $	$\mu y_a-y_e $	R^2_x	R^2_y	ED(c_a, c_e)	%ED(c_a, c_e)
BIO-ID	1.04	0.57	0.993	0.998	1.43	3.98
Talking-Face	1.23	0.97	0.990	0.956	1.96	2.49
GI4E	1.32	0.71	0.996	0.999	1.70	3.87

จากตารางที่ 2.4 ความแม่นยำที่ดีที่สุดที่ทำได้คือร้อยละ 97.10 เมื่อความผิดพลาดมาตรฐาน $Wec \leq 0.05$ ซึ่งมีประสิทธิภาพดีกว่างานที่นำเสนอล่าสุด [15] ที่ทำได้ร้อยละ 89.59 บนฐานข้อมูลเดียวกัน จากผลการทดลองการที่ประสิทธิภาพบนฐานข้อมูล talking-face ดีกว่าบนฐานข้อมูลอื่น ๆ นั้นคาดว่าเป็นเพราะภาพในฐานข้อมูลนี้มีความท้าทายน้อยกว่าเนื่องจากถูกบันทึกมาจากบุคคลเดียว ทำให้ไม่มีความหลากหลายในแง่ของรูปร่างสัณฐานของดวงตา สังเกตว่างานวิจัยจะเน้นพูดถึงค่า Wec มากกว่าค่า Bec เนื่องการจะได้ค่า Wec ที่ต่ำกว่ามีความท้าทายมากกว่าค่า Bec สำหรับประสิทธิภาพบนฐานข้อมูล BioID ร้อยละ 94.5 มีค่า $Wec \leq 0.05$ ซึ่งเมื่อเปรียบเทียบกับงานที่เกี่ยวข้องก่อนหน้าพบว่า [15][16] มีอัตราความสำเร็จร้อยละ 81 และร้อยละ 84 ตามลำดับ นอกจากนี้บนฐานข้อมูล GI4E วิธีที่ผู้วิจัยนำเสนอให้ประสิทธิภาพสูงกว่างานวิจัยที่เกี่ยวข้องล่าสุดโดยให้อัตราความสำเร็จร้อยละ 95 ในขณะที่งานวิจัยล่าสุดทำได้เพียงร้อยละ 93.9

สำหรับการใช้มาตรฐานที่ผู้วิจัยนำเสนอ S_{ED} ให้ผลลัพธ์ไปในแนวทางเดียวกันดูได้จากตารางที่ 2.5 ในคอลัมน์สุดท้าย แต่จะเห็นได้ถึงความไม่สอดคล้องเมื่อดูจากมาตรฐานที่เหลือ ทั้งนี้เนื่องจากมาตรฐานอื่น ๆ มีค่าขึ้นกับขนาดของภาพและขนาดของภาพในแต่ละฐานข้อมูลมีความแตกต่างกัน จึงไม่สามารถนำมาเปรียบเทียบกันได้โดยตรง อย่างไรก็ตามจะเห็นได้ว่างานวิจัยนี้มีประสิทธิภาพที่สูงมากเนื่องจากสามารถให้ความแม่นยำและความคงทนสูง แต่สำหรับวิธีการที่ใช้ได้สร้างขึ้นมาได้ สมมติฐานว่ารูปร่างตาจะเป็นส่วนที่ดำที่สุดและม่านตาเป็นทรงกลม ซึ่งทำให้อาจเกิดความคลาดเคลื่อนได้สูงหากนำไปใช้กับภาพที่มีองศาการมองที่ผิดแปลกไป หรือแม้แต่การที่ม่านตาสีอ่อนก็อาจจะทำให้เกิดความคลาดเคลื่อนได้ในขั้นตอนของการตรวจหาตำแหน่งม่านตา เนื่องจากความเข้มของสีไม่ได้เป็นไปตามสมมติฐานของงาน

บทที่ 3

ทฤษฎีที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาทฤษฎีที่เกี่ยวข้องในเรื่องต่าง ๆ เพื่อเป็นพื้นฐานและแนวทางในการดำเนินงาน ดังนี้ ดวงตาและองค์ประกอบของดวงตา, การเรียนรู้ของเครื่อง, การเรียนรู้เชิงลึก, โครงข่ายประสาทเทียมแบบคอนโวลูชัน และการเรียนรู้ของโครงข่ายประสาทเทียม, ดวงตาและโครงสร้างทางกายวิภาค

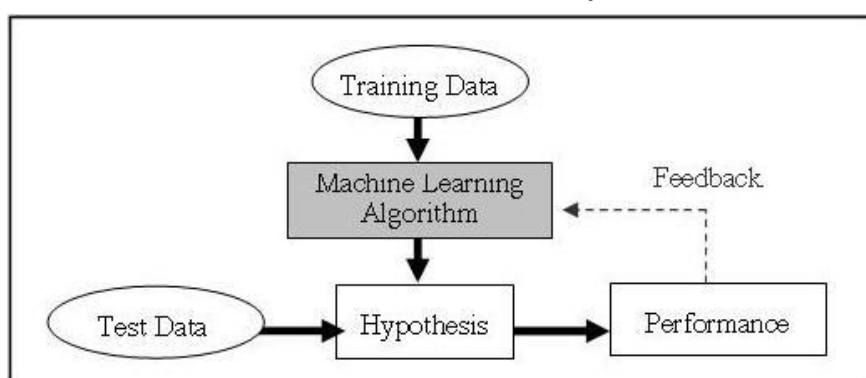
ดวงตาเป็นอวัยวะสำคัญสำหรับการมองเห็น โดยองค์ประกอบหลักๆ สามารถแบ่งได้ดังนี้

1. กระจกตา (Cornea) เป็นผิวชั้นนอกสุด มีลักษณะใสคล้ายกระจกที่สามารถสะท้อนแสงได้
2. ม่านตา (Iris) เป็นชั้นถัดจากกระจกตา เป็นเนื้อเยื่อชนิดหนึ่งของลูกตา มีลวดลายและสีที่แตกต่างกันขึ้นกับบุคคล อาจเป็นสีเขียว สีฟ้า สีน้ำตาลหรือสีดำ ม่านตาสามารถปรับหดและขยายได้ เพื่อให้รูม่านตาเล็กลงหรือขยายใหญ่ขึ้น เพื่อรับแสงให้พอเหมาะในการเห็นภาพ
3. รูม่านตา (Pupil) อยู่ตรงกลางของม่านตา มีลักษณะเป็นวงกลมเล็ก ๆ และมีสีที่เข้มกว่าบริเวณอื่น
4. เลนส์แก้วตา (Lens) อยู่ด้านหลังม่านตาและรูม่านตา ลักษณะเหมือนเลนส์นูน กลม ใสมีเยื่อบาง ๆ เป็นเปลือกหุ้มโดยรอบ
5. จอประสาทตา (Retina) อยู่ด้านหลังของลูกตาประกอบด้วยเส้นประสาทที่มีความละเอียดสูง
6. คอโรยด์ (Choroid) คอโรยด์เป็นชั้นที่แทรกอยู่ระหว่างจอประสาทผนังลูกตา (ตาขาว) ด้านหลัง ประกอบด้วยเส้นเลือดจำนวนมาก เพื่อหล่อเลี้ยงจอประสาทตา
7. ตาขาว (Sclera) คือส่วนที่เป็นเปลือกหรือผนังหุ้มลูกตา ลูกตาทั้งลูกประกอบด้วยส่วนของตาขาวนี้ ยกเว้นส่วนหน้าที่เป็นกระจกตาเท่านั้น ตาขาวหรือผนังลูกตามีสีขาวทึบหนา ยึดหยุ่นได้เล็กน้อย
8. กล้ามเนื้อตา เกาะติดอยู่บนตาขาวด้านนอก ใต้เยื่อตา

Andrew Iwach รองศาสตราจารย์คลินิกทางด้านจักษุวิทยาจาก University of California San Francisco สหรัฐอเมริกา กล่าวว่า ดวงตาเป็นเพียงพื้นที่แห่งเดียวในร่างกายที่สามารถมองเห็นเส้นประสาท เส้นโลหิตแดง และเส้นโลหิตดำได้ โดยที่ไม่ต้องทำการผ่าหรือตัดเนื้อเยื่อขึ้นส่วนใด ๆ ออกเลยและกระบวนการเกิดโรคใด ๆ ที่สามารถสังเกตเห็นได้ในดวงตานั้นก็อาจจะเกิดขึ้นกับส่วนอื่น ๆ ในร่างกายได้ด้วยเช่นกัน[17]

การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องเป็นการสอนอัลกอริทึมให้เรียนรู้ทำความเข้าใจและตัดสินใจได้ด้วยตัวเองจากข้อมูลที่ป้อนให้ แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์ โดยกระบวนการเรียนรู้ของเครื่องเกี่ยวข้องกับหลักสถิติศาสตร์และคณิตศาสตร์อย่างมาก โดยเฉพาะสาขาการหาค่าเหมาะสมที่สุด เนื่องจากการศึกษาการวิเคราะห์ข้อมูลเพื่อทำนาย การเรียนรู้ของเครื่องสามารถนำไปประยุกต์ใช้งานได้หลากหลาย ไม่ว่าจะเป็นการกรองอีเมลขยะ การรู้จำตัวอักษร เครื่องมือค้นหาหรือแม้แต่คอมพิวเตอร์วิทัศน์ เทคนิคการเรียนรู้ของเครื่องแสดงดังภาพที่ 3.1



ภาพที่ 3.1 เทคนิคการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง สามารถแบ่งโดยกว้าง ๆ ได้เป็น 3 ประเภท ตามประเภทของ "ข้อมูลฝึก" หรือ "ข้อมูลขาเข้า" ได้ดังนี้

- การเรียนรู้แบบมีผู้สอน (supervised learning) – ข้อมูลตัวอย่างและผลลัพธ์ที่ผู้สอนต้องการจะถูกป้อนเข้าสู่คอมพิวเตอร์ เป้าหมายคือการสร้างกฎทั่วไปที่สามารถเชื่อมโยงข้อมูลขาเข้ากับขาออกได้
- การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) – ไม่มีการทำผลเฉลย (label) ใด ๆ โดยจะให้คอมพิวเตอร์หาโครงสร้างของข้อมูลขาเข้าเอง
- การเรียนรู้แบบเสริมกำลัง (reinforcement learning) - คอมพิวเตอร์มีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนไปตลอดเวลาโดยคอมพิวเตอร์จะต้องทำงานบางอย่าง (เช่น ขับรถ) โดยที่ไม่มีผู้สอนคอยบอกอย่างจริงจังว่าวิธีการที่ทำอยู่นั้นเข้าใกล้เป้าหมายแล้วหรือไม่ ตัวอย่างเช่น การเรียนรู้เพื่อเล่นเกม

นอกจากนี้ การเรียนรู้ของเครื่องยังสามารถแบ่งประเภทของ "งาน" ได้ตาม "ข้อมูลขาออก" จากระบบที่เครื่องจักรได้เรียนรู้แล้ว เป็นหลายประเภทดังนี้

- การจำแนกประเภทข้อมูล (classification) - ข้อมูลขาเข้าถูกจำแนกออกเป็นหลายประเภท (class) เพื่อให้เครื่องเรียนรู้ จากนั้นตัวแบบจะสามารถกำหนดประเภทให้กับข้อมูลใหม่ที่ไม่เคยเห็นมาก่อนได้ โดยปกติแล้วจะทำโดยวิธีการเรียนรู้แบบมีผู้สอน ตัวอย่างของการแบ่งประเภทข้อมูลได้แก่ การกรองอีเมลขยะ โดยอีเมลจะถูกแบ่งเป็นประเภท"ขยะ"และ"ไม่ใช่ขยะ"
 - การวิเคราะห์การถดถอย (regression) - ใช้หลักการเกี่ยวกับการแบ่งประเภทข้อมูล แต่ข้อมูลขาออกเป็นลักษณะต่อเนื่อง
 - การแบ่งกลุ่มข้อมูล (clustering) - เป้าหมายคือการแบ่งข้อมูลขาเข้าเป็นกลุ่มๆ โดยอัลกอริทึมจะไม่ทราบกลุ่มดังกล่าวล่วงหน้า (ไม่เหมือนกับการแบ่งประเภทข้อมูล) โดยปกติแล้วมักเป็นการเรียนรู้แบบไม่มีผู้สอน
 - การประเมินความหนาแน่น (density estimation) - เป็นการหาการกระจายของข้อมูลในมิติบางมิติ
 - การลดขนาดของมิติ (dimensionality reduction) - เป็นการเชื่อมโยงข้อมูลหลายมิติไปสู่ปรภูมิที่มีมิติต่ำกว่า
- (<https://th.wikipedia.org/wiki/การเรียนรู้ของเครื่อง>)

การเรียนรู้เชิงลึก (Deep learning)

การเรียนรู้เชิงลึกเป็นสาขาหนึ่งของการเรียนรู้ของเครื่อง พื้นฐานของการเรียนรู้เชิงลึกคือ อัลกอริทึมที่พยายามจะสร้างแบบจำลองเพื่อแทนความหมายของข้อมูลในระดับสูงโดยการสร้างสถาปัตยกรรมข้อมูลขึ้นมาที่มีโครงสร้างประกอบไปด้วยชั้นของข้อมูลหลายชั้น แต่ละชั้นเกิดจากการแปลงไม่เชิงเส้น

การเรียนรู้เชิงลึก อาจมองได้ว่าเป็นวิธีการหนึ่งของการเรียนรู้ของเครื่องที่พยายามเรียนรู้วิธีการแทนข้อมูลอย่างมีประสิทธิภาพ ตัวอย่างเช่น รูปภาพภาพหนึ่ง สามารถแทนได้เป็นเวกเตอร์ของความสว่างต่อพิกเซล หรือมองในระดับสูงขึ้นไปเป็นเซตของขอบของวัตถุต่าง ๆ หรือมองว่าเป็นพื้นที่ของรูปร่างใด ๆ ก็ได้ การแทนความหมายดังกล่าวจะทำให้การเรียนรู้ที่จะทำงานต่าง ๆ ทำได้ง่ายขึ้น ไม่ว่าจะเป็นการรู้จำใบหน้าหรือการรู้จำการแสดงออกทางสีหน้า การเรียนรู้เชิงลึกถือว่าเป็นวิธีการที่มีศักยภาพสูงในการจัดการกับฟีเจอร์ (features) สำหรับการเรียนรู้แบบไม่มีผู้สอนหรือการเรียนรู้แบบกึ่งมีผู้สอน

นักวิจัยด้านการเรียนรู้ของเครื่องได้เสนอสถาปัตยกรรมการเรียนรู้หลายแบบบนหลักการของการเรียนรู้เชิงลึกนี้ ได้แก่ โครงข่ายประสาทเทียมแบบลึก (Deep Artificial Neural Networks) โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks) โครงข่ายความเชื่อแบบลึก (Deep Belief Networks) และโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) ซึ่งมีการนำมาใช้งานอย่างแพร่หลายในทางคอมพิวเตอร์วิทัศน์ การรู้จำเสียงพูด การประมวลผลภาษาธรรมชาติ การรู้จำเสียง และชีวสารสนเทศศาสตร์ (<https://th.wikipedia.org/wiki/การเรียนรู้เชิงลึก>)

โครงข่ายประสาทเทียม (Neural network)

โครงข่ายประสาทเทียมเปรียบเสมือนฟังก์ชันทางคณิตศาสตร์ ที่ถูกจำลองให้มีลักษณะการทำงานคล้ายกับโครงข่ายประสาทแท้ โดยโครงข่ายประสาทเทียมจะหาความสัมพันธ์ของข้อมูลนำเข้า และข้อมูลส่งออกจากการประมวลผลโดยคำนวณแบบเชื่อมต่อกันของหน่วยประมวลผลย่อยหลาย ๆ หน่วย จนเป็นโครงข่ายที่ประมวลผลและทำงานร่วมกันได้ตามวัตถุประสงค์ของงานที่สนใจ ในการสร้างโครงข่ายประสาทเทียมประกอบไปด้วย 5 องค์ประกอบหลัก ได้แก่

1. ข้อมูลนำเข้า
2. ข้อมูลส่งออก
3. ค่าน้ำหนัก ซึ่งค่าน้ำหนักจะปรับเปลี่ยนผ่านกระบวนการเรียนรู้ โดยขึ้นอยู่กับข้อมูลและสถาปัตยกรรมของโครงข่ายที่ใช้ เมื่อนำข้อมูลเข้าส่งไปในโครงข่ายจะทำให้ได้ค่าผิดพลาดระหว่างค่าผลเฉลยและผลลัพธ์ที่ได้จากโครงข่าย กระบวนการเรียนรู้จะนำค่าความผิดพลาดมาใช้ในการปรับค่าน้ำหนักเพื่อให้ได้ผลลัพธ์ที่ใกล้เคียงกับค่าเป้าหมายมากที่สุด เมื่อเสร็จสิ้นกระบวนการเรียนรู้ค่าน้ำหนักจะถูกเก็บเป็นค่าคงที่เพื่อนำไปใช้งาน ทั้งนี้ค่าน้ำหนักยังสามารถนำไปฝึกฝนผ่านกระบวนการเรียนรู้ต่อได้
4. ฟังก์ชันผลรวม เป็นผลรวมของผลคูณของข้อมูลนำเข้าและค่าน้ำหนัก
5. และฟังก์ชันการแปลง เป็นฟังก์ชันที่ใช้ในการคำนวณค่าผลรวมเป็นค่าผลลัพธ์ ฟังก์ชันการแปลงจะเป็นฟังก์ชันไม่เชิงเส้นเพื่อทำให้โครงข่ายมีความสามารถประมาณค่าบนชุดข้อมูลที่มีความซับซ้อนได้

(อำภา สารศิริ (2559) เทคนิคการเรียนรู้พื้นฐานโครงข่ายประสาทเทียม. แหล่งที่มา :

<http://www.mut.ac.th/research-detail-92>)

สถาปัตยกรรมของโครงข่ายประสาทเทียมนั้นเป็นสิ่งสำคัญอย่างยิ่งในการสร้างโครงข่ายประสาทให้มีความเหมาะสมกับชุดข้อมูลและวัตถุประสงค์ของการทำงาน ซึ่งงานด้านคอมพิวเตอร์วิทัศน์นิยมใช้สถาปัตยกรรมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural network - CNN)

สถาปัตยกรรมแบบคอนโวลูชันเป็นโครงข่ายประสาทเทียมรูปแบบหนึ่งที่พิจารณาข้อมูลเป็นกลุ่มย่อย ๆ แล้วนำมาประกอบกัน โดยใช้ตัวกรอง (filter) มาทำคอนโวลูชันกับข้อมูล ทำให้สถาปัตยกรรมนี้มีความเหมาะสมที่จะนำไปใช้งานกับรูปภาพ เนื่องจากข้อมูลรูปภาพประกอบไปด้วยจุดภาพหรือพิกเซลมากมาย การที่มนุษย์จะสามารถรู้จำวัตถุใด ๆ หรือภาพใด ๆ ได้ ล้วนแต่เป็นการมองหลายพิกเซลประกอบกัน หากพิจารณาเพียงพิกเซลใดพิกเซลหนึ่งจึงเป็นไปได้ยากที่จะรู้ความหมาย เช่นเดียวกับกับคอมพิวเตอร์ การที่โครงข่ายพิจารณาภาพเป็นพื้นที่เล็ก ๆ จึงสามารถสกัดข้อมูลบางอย่างออกมาได้ดีกว่า เช่น คุณลักษณะสีของพื้นที่นั้น ๆ ลายเส้น หรือการตัดกันของสี

การขยายข้อมูล (Data Augmentation)

การขยายข้อมูลเป็นการเพิ่มจำนวนข้อมูลให้มากขึ้น เนื่องจากตามหลักการยังมีข้อมูลมาก การเรียนรู้ของเครื่องก็จะมีประสิทธิภาพที่ดีขึ้น การขยายข้อมูลสามารถทำได้หลายวิธี แต่ควรต้องคำนึงถึงความเหมาะสมกับข้อมูลที่จะนำมาใช้งานจริงด้วย ในการทำงานกับรูปภาพอาจนำรูปภาพที่มีอยู่เดิมมากลับซ้าย กลับขวา หมุน ทำให้ภาพเบลอ เลื่อนภาพ หรือครอบตัดภาพ เป็นต้น ดังนั้น ภาพต้นฉบับหนึ่งภาพจึงสามารถขยายได้เป็นหลาย ๆ ภาพที่แตกต่างกัน นอกจากนี้เมื่อข้อมูลมีความหลากหลายมากขึ้น จะสามารถลดปัญหาที่ตัวแบบหรือตัวแบบทำงานได้ไม่แม่นยำมากับข้อมูลรูปภาพที่ใช้ในการฝึกแต่ทำงานได้ไม่แม่นยำในข้อมูลจริงได้

การปรับปรุงภาพด้วยวิธีการปรับเท่าฮิสโทแกรม (Histogram Equalization)

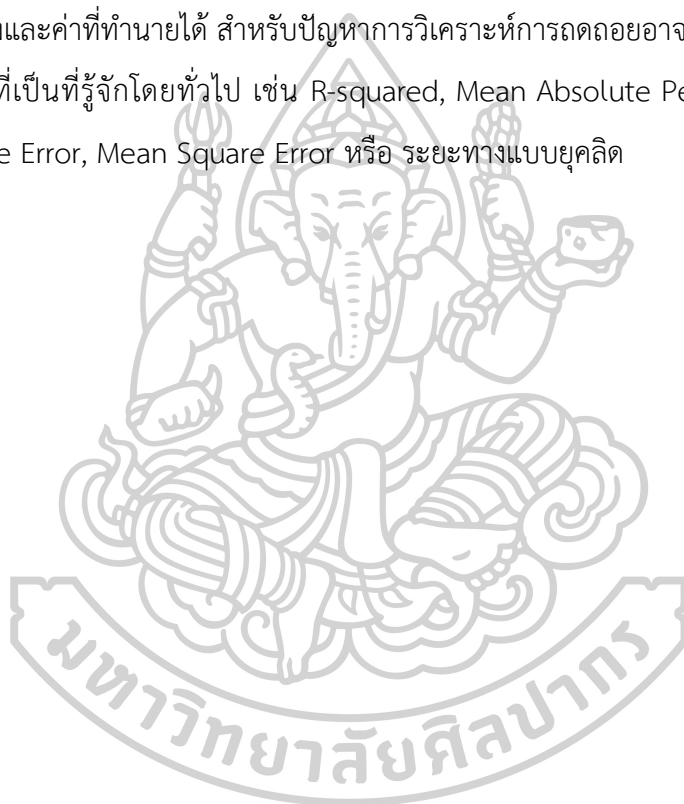
การปรับปรุงภาพด้วยวิธีการปรับเท่าฮิสโทแกรม หรือ การทำ Histogram Equalization เป็นการปรับปรุงภาพเพื่อให้มีการกระจายของพิกเซลหรือจุดภาพในแต่ละค่าความเข้มสีให้มีจำนวนใกล้เคียงกัน โดยอาศัยหลักการทางสถิติกระจายความน่าจะเป็นของค่าระดับสีเทา การปรับเท่าฮิสโทแกรมในงานวิจัยนี้ใช้ฟังก์ชัน `cv2.equalizeHist()` ที่มีอยู่ในแพ็คเกจ `opencv` ซึ่งเป็นแพ็คเกจที่เหมาะสมสำหรับใช้งานกับข้อมูลรูปภาพโดยเฉพาะ

ความคงทน (Robustness)

ความคงทนเป็นคุณลักษณะที่แสดงความสามารถของวิธีการที่นำเสนอว่าผลการทดสอบจะไม่เปลี่ยนแปลงหรือไม่ส่งผลต่อความถูกต้องแม่นยำอย่างมีนัยสำคัญเมื่อมีการเปลี่ยนแปลงสถานะเล็กน้อยจากสถานะปกติ

ความแม่นยำ (Accuracy)

ความแม่นยำคือค่าที่ใช้ประเมินผลลัพธ์ของวิธีการ สามารถดูได้จากค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่ทำนายได้ สำหรับปัญหาการวิเคราะห์การถดถอยอาจวัดความคลาดเคลื่อนได้ด้วยมาตรวัดที่เป็นที่รู้จักโดยทั่วไป เช่น R-squared, Mean Absolute Percentage Error, Root Mean Square Error, Mean Square Error หรือ ระยะทางแบบยุคลิด



บทที่ 4

วิธีการดำเนินงานวิจัย

ในบทนี้จะกล่าวถึงข้อมูลที่ใช้ในงานวิจัย แนวความคิด และขั้นตอนการทดลอง

4.1 ข้อมูลที่ใช้ในงานวิจัย

ผู้วิจัยได้เลือกใช้ชุดข้อมูลสาธารณะ MPIIGaze เพื่อฝึกฝนตัวแบบการเรียนรู้เชิงลึกเนื่องจากภาพในฐานข้อมูลถูกถ่ายโดยอยู่ในสภาวะแวดล้อมที่ไม่ถูกควบคุมโดยการใช้แสงอินฟราเรด นอกจากนี้ภาพตาส่วนใหญ่ยังมีลักษณะที่คล้ายกับคนเอเชีย ซึ่งข้อมูลที่ผู้วิจัยนำมาใช้ผ่านการปรับแต่งให้เหลือเพียงบริเวณดวงตาเรียบร้อยแล้ว และภาพดวงตามีขนาด 32×64 พิกเซล มีจำนวนทั้งหมด 34,099 ภาพ อย่างไรก็ตามผู้วิจัยเลือกมาใช้งานเพียง 2,500 ภาพ โดยวิธีการสุ่ม ตัวอย่างภาพในชุดข้อมูล MPII แสดงให้เห็นดังภาพที่ 4.1



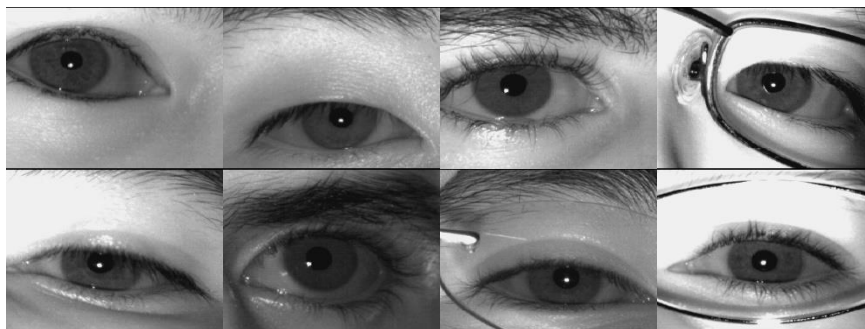
ภาพที่ 4.1 ตัวอย่างภาพในชุดข้อมูล MPIIGaze

สำหรับข้อมูลทดสอบได้ใช้ทั้งหมด 3 ชุดข้อมูลได้แก่ BioID, MMU2 และ Flickr-Faces-HQ (FFHQ) โดยชุดข้อมูล BioID ประกอบไปด้วยภาพใบหน้าขนาด 384×286 พิกเซล ทั้งหมด 1,520 ภาพ มีความหลากหลายของการเอียงของศีรษะ ระดับแสงสว่างและอารมณ์บนใบหน้า ซึ่งในการนำมาเป็นข้อมูลทดสอบ ภาพใบหน้าจากฐานข้อมูล BioID จะถูกตัดพื้นที่เหลือเพียงบริเวณดวงตาเท่านั้น



ภาพที่ 4.2 ตัวอย่างภาพในชุดข้อมูล BioID

ชุดข้อมูล MMU2 ประกอบไปด้วยภาพตาขนาด 320 x 238 พิกเซล ทั้งหมด 995 ภาพ ซึ่งถูกบันทึกจากอาสาสมัครในเอเชีย, ตะวันออกกลาง, แอฟริกา และยุโรป โดยระดับแสงสว่างในแต่ละภาพมีความหลากหลายและแตกต่างกันชัดเจน มีการปรากฏของสัญญาณรบกวนต่าง ๆ ในภาพ ตัวอย่างเช่น จุดในดวงตาที่เกิดจากการสะท้อนของแสง การบดบังที่เกิดจากขนตาหรือผมดังแสดงในภาพที่ 4.3



ภาพที่ 4.3 ตัวอย่างภาพในชุดข้อมูล MMU2

ชุดข้อมูล Flickr-Faces-HQ-68000 (FFHQ) ประกอบไปด้วยภาพใบหน้าขนาดใหญ่ 1024 x 1024 พิกเซล ทั้งหมด 1,000 ภาพ มีความหลากหลายทั้งในแง่ของอายุ เชื้อชาติและพื้นหลังของภาพ โดยภาพในชุดข้อมูลนี้ถูกรวบรวมมาจากเว็บไซต์ Flickr ซึ่งในขั้นตอนการเตรียมข้อมูลเพื่อใช้เป็นข้อมูลทดสอบ ภาพใบหน้าจากรฐานข้อมูล FFHQ จะถูกตัดพื้นที่เหลือเพียงบริเวณดวงตา เช่นเดียวกับฐานข้อมูล BiOid



ภาพที่ 4.4 ตัวอย่างภาพในชุดข้อมูล Flickr-Faces-HQ-68000 (FFHQ)

4.2 แนวความคิดและขั้นตอนการทดลอง

ผู้วิจัยได้วางแผนขั้นตอนการทดลองไว้ 3 ขั้นตอนโดยมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล

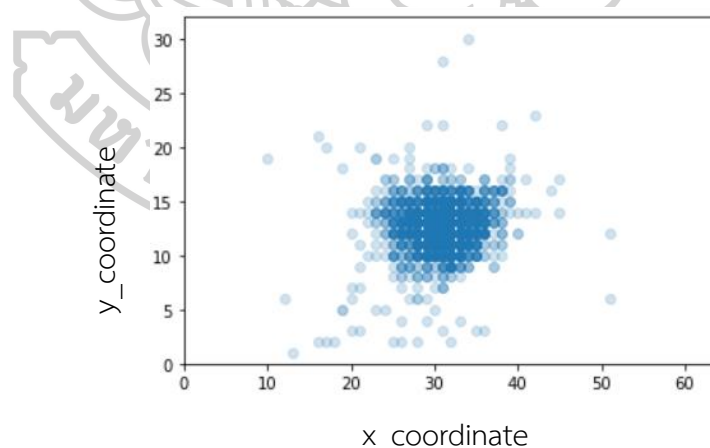
ขั้นตอนที่ 2 การเรียนรู้เชิงลึกของเครื่อง

ขั้นตอนที่ 3 การวัดประสิทธิภาพ

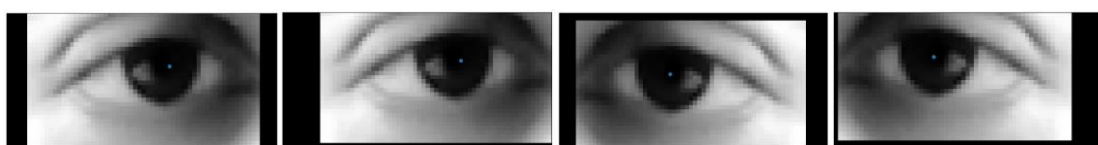
4.2.1 การเตรียมข้อมูล

การเตรียมข้อมูลฝึก จากที่กล่าวมาจะเห็นว่าข้อมูลที่น่ามาใช้ฝึกตัวแบบมีความหลากหลายในแง่มุมมองรูปร่าง ลักษณะของดวงตา รวมถึงองศาการมอง แต่จุดศูนย์กลางดวงตาของภาพในชุดข้อมูลจะมีความเอนเอียง (Bias) อยู่ที่บริเวณใจกลางของภาพ โดยจากการสำรวจข้อมูลฝึกสอน ตำแหน่งจุดศูนย์กลางรูม่านตาซึ่งเป็นผลเฉลยมีการกระจายดังภาพที่ 4.5 ซึ่งมีค่าเฉลี่ยของคู่อันดับของจุดศูนย์กลางรูม่านตาคือ (31, 13) และมีค่าการเบี่ยงเบนเท่ากับ 3.24 และ 2.17 ตามแนวแกน x และแนวแกน y

ดังนั้นผู้วิจัยจึงได้ทำการขยายข้อมูล (Data Augmentation) เพื่อเพิ่มจำนวนภาพดวงตาที่ใช้ในกระบวนการเรียนรู้เชิงลึกและเพิ่มการกระจายของตำแหน่งจุดศูนย์กลางรูม่านตาในข้อมูลฝึก โดยใช้ 2 วิธีคือ 1) กลับภาพในทิศทางแนวนอน 2) เลื่อนภาพตามทิศทางแนวนอนและทิศทางแนวตั้งไม่เกิน 10 เปอร์เซ็นต์ของความกว้างในแต่ละทิศทาง ตัวอย่างภาพข้อมูลฝึกที่ผ่านการขยายข้อมูลแสดงให้เห็นดังภาพที่ 4.6



ภาพที่ 4.5 การกระจายของตำแหน่งจุดศูนย์กลางรูม่านตาในชุดข้อมูล MPII



ภาพที่ 4.6 ตัวอย่างภาพข้อมูลฝึกที่ผ่านการขยายข้อมูล

สำหรับผลเฉลยของข้อมูล (Ground Truth) ถูกเตรียมโดยการระบุตำแหน่งจุดกึ่งกลางรูمانةตาในภาพดวงตาแต่ละภาพด้วยวิธีแมนนวลผ่าน <https://www.makesense.ai/> โดยบันทึกตำแหน่งคู่อันดับ (x, y) บนระบบคาร์ทีเซียน (Cartesian Coordinate System)

จากที่ข้อมูลฝึกมีขนาดค่อนข้างเล็ก การระบุตำแหน่งผลเฉลยบนภาพจึงมีความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนได้สูง อาจด้วยสาเหตุที่เกิดจากความผิดพลาดของตัวผู้ระบุผลเฉลยเอง หรือเป็นความคลาดเคลื่อนที่เกิดด้วยความจำเป็น เช่น หากจุดกึ่งกลางรูمانةตาอยู่บริเวณกึ่งกลางของพิกเซลทั้งสี่ดังภาพที่ 4.7 การระบุผลเฉลยจะจำเป็นต้องระบุผลเฉลยที่พิกเซลใดพิกเซลหนึ่งในสี่พิกเซลนี้ ทำให้ผลเฉลยที่ได้ไม่ตรงกับตำแหน่งจุดศูนย์กลางรูمانةตาจริงในภาพ และส่งผลให้ค่าคลาดเคลื่อนที่คำนวณในขั้นตอนสุดท้ายอาจสูงกว่าความเป็นจริง เพื่อแก้ปัญหาดังกล่าวผู้วิจัยจึงเพิ่มความละเอียดภาพขณะสร้างผลเฉลย โดยขยายขนาดภาพในชุดข้อมูลฝึกจาก 32×64 พิกเซล เป็น 150×300 พิกเซล เมื่อระบุผลเฉลยบนภาพขนาดใหญ่เรียบร้อยแล้วจึงสเกลค่าผลเฉลยกลับไปให้ขอบเขตของความละเอียด 32×64 พิกเซล อีกครั้ง โดยการสเกลค่ากลับไปเป็นเลขทศนิยมจะทำให้ข้อมูลผลเฉลยมีความละเอียดและตรงกับความเป็นจริงมากขึ้น

ภาพที่ 4.7 ตัวอย่างภาพตากรณีจุดกึ่งกลางตาอยู่ตรงกลางระหว่างพิกเซล

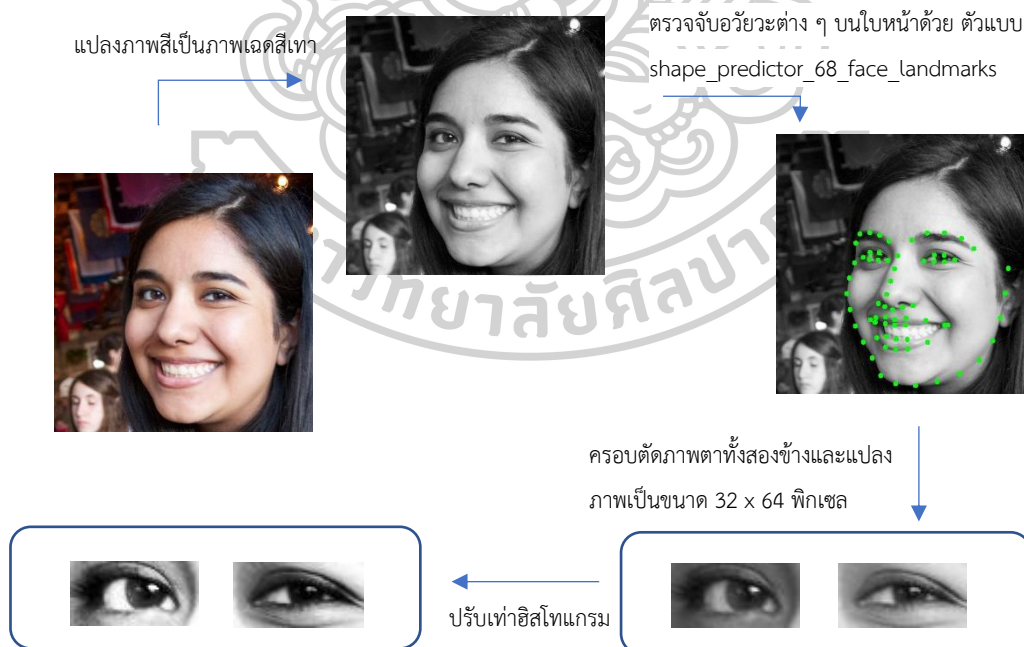
การเตรียมข้อมูลทดสอบ ภาพต้นฉบับทั้งหมดที่เป็นภาพสีจะถูกเปลี่ยนให้อยู่ในรูปแบบของภาพเฉดสีเทา (gray-scale) สำหรับภาพในฐานข้อมูล BioID และ Flickr-Faces-HQ-68000 เป็นภาพใบหน้า เพื่อให้การครอบตัดภาพดวงตาเป็นไปโดยง่ายจึงได้ใช้ตัวแบบตรวจจับอวัยวะต่าง ๆ บนใบหน้าชื่อ shape_predictor_68_face_landmarks ในไลบรารี Dlib มาช่วยตรวจจับบริเวณดวงตาดังภาพที่ 4.8 เมื่อทราบตำแหน่งของดวงตา ต่อไปจึงครอบตัดภาพบริเวณตาทั้งสองข้างโดยกำหนดอัตราส่วนในการครอบตัดให้มีอัตราส่วนความกว้างต่อความยาวเท่ากับ 1:2 จากนั้นจึงปรับความละเอียดภาพให้มีขนาดเท่ากับความละเอียดของภาพชุดข้อมูลฝึก สำหรับข้อมูลทดสอบในฐานข้อมูล MMU2 ได้ถูกเตรียมให้มีความละเอียดภาพเท่ากับความละเอียดของภาพชุดข้อมูลฝึกเช่นเดียวกัน โดยคงอัตราส่วนของภาพเดิมไว้ด้วยการเติมสีดำ ตัวอย่างภาพที่ชุดข้อมูลทดสอบทั้งสามฐานข้อมูล

หลังจากผ่านการเตรียมข้อมูลแสดงให้เห็นในภาพที่ 4.9 และการกระจายของผลเฉลี่ยในแต่ละชุดข้อมูลฝึกแสดงให้เห็นดังภาพที่ 4.10

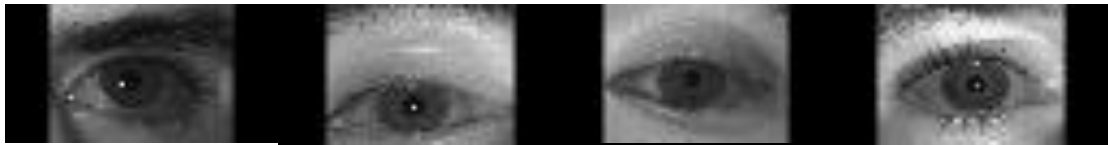
บนฐานข้อมูล MMU2 มีค่าเฉลี่ยของตำแหน่งจุดศูนย์กลางรูม่านตาอยู่ที่ตำแหน่ง (31, 16) และมีค่าการเบี่ยงเบนของตำแหน่ง x และ y เท่ากับ 5.04 และ 3.47 บนฐานข้อมูล BioID มีค่าเฉลี่ยของตำแหน่งจุดศูนย์กลางรูม่านตาอยู่ที่ตำแหน่ง (33, 14) และมีค่าการเบี่ยงเบนของตำแหน่ง x และ y เท่ากับ 2.60 และ 1.56 บนฐานข้อมูล FFHQ มีค่าเฉลี่ยของตำแหน่งจุดศูนย์กลางรูม่านตาอยู่ที่ตำแหน่ง (32, 14) และมีค่าการเบี่ยงเบนของตำแหน่ง x และ y เท่ากับ 5.13 และ 1.82

นอกจากนี้ผู้วิจัยได้ปรับปรุงคุณภาพภาพในข้อมูลชุดทดสอบด้วยวิธีการปรับเท่าฮิสโทแกรมเพื่อให้ภาพมีความคมชัดมากขึ้นด้วย ตัวอย่างภาพชุดข้อมูลทดสอบทั้งสามฐานข้อมูลหลังจากการปรับเท่าฮิสโทแกรมแสดงในภาพที่ 4.11

งานวิจัยนี้เน้นไปที่การหาจุดศูนย์กลางรูม่านตาในขณะที่ตาอยู่ในสภาวะเปิดหรือภาพที่ปรากฏรูม่านตาเท่านั้น ในกระบวนการทดสอบผู้วิจัยจึงได้คัดเลือกภาพตาที่ไม่ปรากฏรูม่านตาออกไปทำให้ชุดข้อมูลทดสอบมีภาพตาที่ได้จากฐานข้อมูล BioID จำนวน 2,436 ภาพ และฐานข้อมูล FFHQ จำนวน 1,706 ภาพ



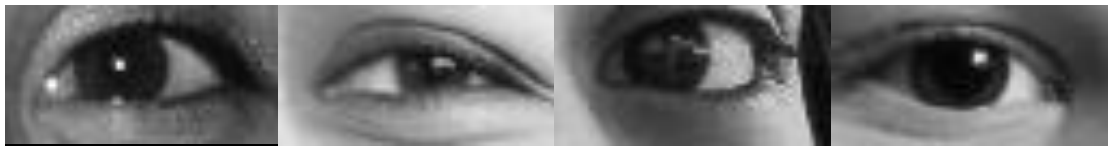
ภาพที่ 4.8 ขั้นตอนการเตรียมข้อมูลสำหรับฐานข้อมูล BioID และ FFHQ



ก) ฐานข้อมูล MMU2

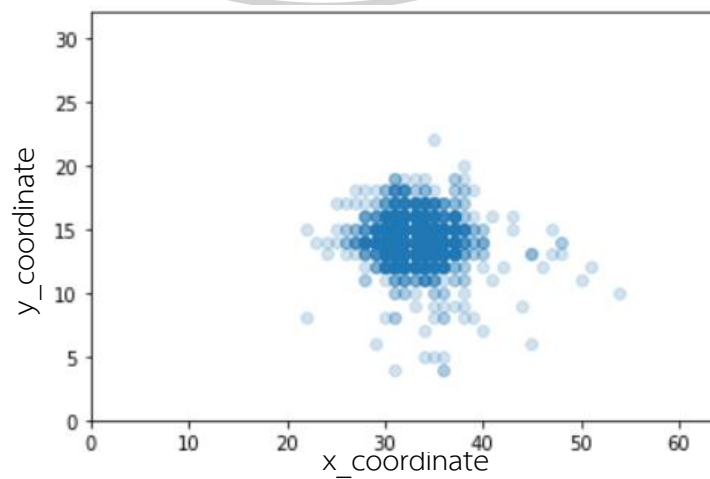
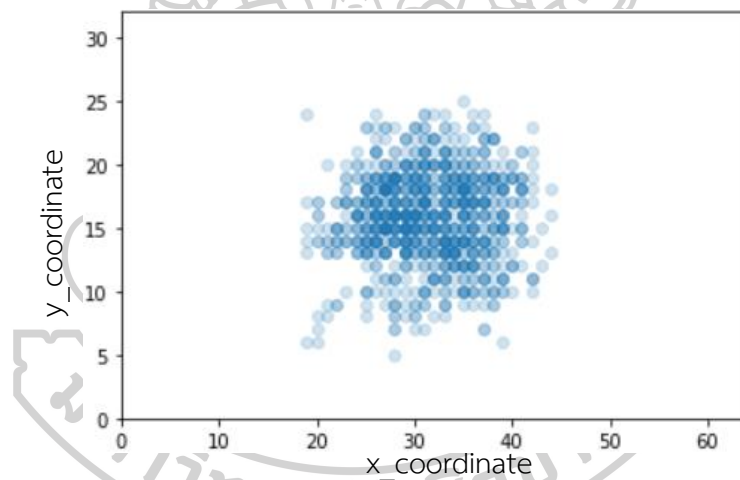


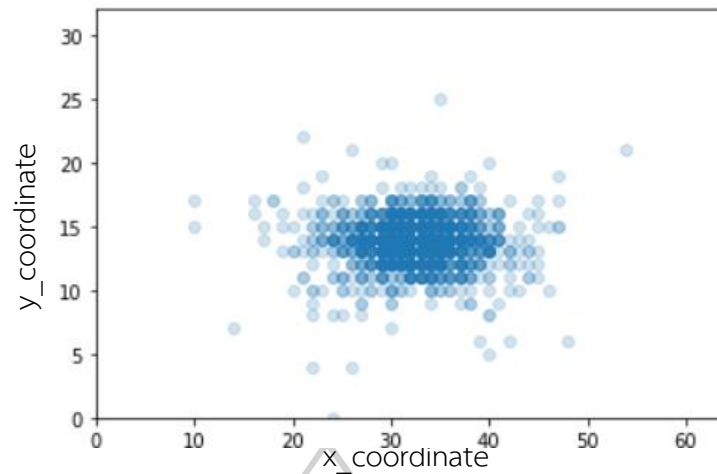
ข) ฐานข้อมูล BiID



8) ฐานข้อมูล FFHQ

ภาพที่ 4.9 ภาพหลังการเตรียมข้อมูลของชุดข้อมูลทดสอบทั้งสามฐานข้อมูล





ค) ฐานข้อมูล FFHQ

ภาพที่ 4.10 การกระจายตัวของผลเฉลี่ยในชุดข้อมูลทดสอบ ก) ฐานข้อมูล MMU2 ข) ฐานข้อมูล

BioID ค) ฐานข้อมูล FFHQ



ก) ฐานข้อมูล MMU2



ข) ฐานข้อมูล BioID



8) ฐานข้อมูล FFHQ

ภาพที่ 4.11 ตัวอย่างภาพในชุดข้อมูลทดสอบหลังจากผ่านการปรับเท่าฮิสโทแกรม

4.2.2 การเรียนรู้เชิงลึกของเครื่องและการทดสอบการเรียนรู้

ในขั้นตอนการออกแบบโครงข่ายประสาทเทียมเพื่อวิเคราะห์การถดถอยของตำแหน่งจุดศูนย์กลางรูม่านตานั้นไม่มีขั้นตอนและรูปแบบเฉพาะ ต้องมีการทดลองและปรับเปลี่ยนโครงสร้างตามความเหมาะสมของงานที่แตกต่างกันไป อย่างไรก็ตามผู้วิจัยได้เริ่มต้นจากการใช้โครงสร้างพื้นฐานที่มี

ขนาดเล็กก่อน และปรับเปลี่ยนคุณลักษณะในโครงข่ายบางอย่างตามสมมติฐานเพื่อความแม่นยำและความคงทนที่ดีขึ้น

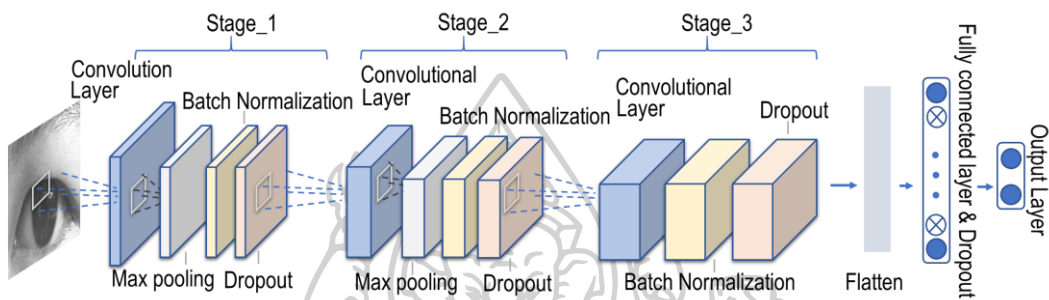
การระบุจุดศูนย์กลางรูปร่างตาด้วยความแม่นยำระดับพิกเซลเป็นเรื่องที่ต้องต่อสู้กับหลายปัจจัย เช่น ความไม่แน่นอนของตำแหน่งตาในภาพ ซึ่งอาจเกิดจากการครอบตัดภาพหรือการที่ตาในภาพเหลื่อมมองไปยังตำแหน่ง, ทิศ ที่แตกต่างกัน ความหลากหลายของลักษณะดวงตา/ขนาดของตา การบดบังบริเวณลูกตาจากองค์ประกอบอื่น ๆ ของตา รวมไปถึงข้อจำกัดเรื่องขนาดของข้อมูลเข้าที่มีขนาดค่อนข้างเล็ก

สิ่งต่าง ๆ เหล่านี้เป็นอุปสรรคในการวิเคราะห์การถดถอยในระดับพิกเซลอย่างปฏิเสธไม่ได้ ดังนั้นเทคนิคและโครงสร้างพื้นฐานโดยทั่วไปอาจให้ผลลัพธ์ที่ไม่เป็นที่พึงพอใจเท่าที่ควร ซึ่งในบทความนี้จะนำเสนอวิธีการในการระบุจุดศูนย์กลางรูปร่างตาโดยใช้ลักษณะโครงสร้างพื้นฐานที่เป็นที่นิยมทั่วไป โดยถือเป็นตัวแบบเริ่มต้นในการวิจัยในครั้งนี้ และจะนำเสนอคุณลักษณะเพิ่มเติมตามสมมติฐานต่อไป โครงสร้างเริ่มต้นหรือโครงสร้างพื้นฐานที่นำมาใช้ได้ทำการแยกคุณลักษณะตามลำดับชั้น

(hierarchical feature extraction) ทั้งหมด 3 ชั้น (Stage) ดังภาพที่ 4.12 แต่ละชั้นประกอบไปด้วยโครงสร้างดังต่อไปนี้

1. ชั้นคอนโวลูชันที่เชื่อมต่อกับฟังก์ชันเรกทิไฟเออร์เชิงเส้น (ReLU) เป็นกระบวนการที่นำตัวกรองขนาด 3×3 มาประมวลผลรูปภาพ เพื่อเน้นคุณสมบัติที่สำคัญบางอย่างในภาพ โดยผลลัพธ์ที่ได้จากกระบวนการนี้คือคุณลักษณะหรือ Feature maps จำนวน 32 Feature maps ในชั้นที่ 1 และจำนวน 64 และ 128 Feature maps ในชั้นต่อไป
2. ชั้นพูลลิงแบบค่ามากที่สุด (Max pooling) เป็นกระบวนการที่นำหน้าต่างขนาด 2×2 มาลดขนาดสัดส่วนภาพ เพื่อสกัดให้เหลือเพียงลักษณะเด่นโดยพิจารณาเลือกค่าที่สุดในหน้าต่าง เมื่อดำเนินการเลื่อนหน้าต่างไปยังพื้นที่ต่าง ๆ ในภาพ โดยในชั้นแรกขนาดของ Feature maps จะถูกลดลงเหลือ 16×32 และในชั้นที่สองจะลดลงเหลือ 8×16 โดยภายในโครงข่ายนี้ได้ยกเว้นชั้นพูลลิงแบบค่าที่สุดในชั้นสุดท้ายไว้ เพื่อป้องกันการเสียรายละเอียดมากเกินไปเนื่องจากข้อมูลฝึกที่ใช้มีขนาดค่อนข้างเล็ก
3. ชั้นนอร์มัลไลเซชัน เป็นกระบวนการที่ใช้ปรับค่าข้อมูลในแต่ละชั้นให้เป็นมาตรฐาน เทคนิคนี้จะช่วยลดปัญหา Vanishing gradient ลง และยังให้ผลในการ Regularize ด้วย
4. ชั้นดรอปเอาต์ที่ระดับ 10 เปอร์เซ็นต์ เพื่อลดปัญหา Overfitting

จากขั้นสุดท้ายจะได้คุณลักษณะเด่นของภาพข้อมูลเข้าในระดับสูง ซึ่งจะนำไปเชื่อมต่อกับชั้นเชื่อมแบบเต็มรูปแบบ ประกอบไปด้วยโหนด 100 โหนด เชื่อมกับฟังก์ชันซิกมอยด์ ส่วนการวิเคราะห์การถดถอยในชั้นข้อมูลออกใช้ฟังก์ชันการกระตุ้นเป็นฟังก์ชันเชิงเส้นเพื่อให้ได้ข้อมูลออกเป็นค่าที่มีลักษณะต่อเนื่องโดยในชั้นนี้ประกอบไปด้วย 2 โหนด ซึ่งสอดคล้องกับคู่อันดับ x, y ของตำแหน่งจุดศูนย์กลางรูมาตา และฟังก์ชันต้นทุน (Loss Function) สำหรับประเมินความแม่นยำของตัวแบบได้ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared error; MSE)



ภาพที่ 4.12 โครงสร้างพื้นฐาน (SimpleCNN)

จากโครงสร้างพื้นฐานที่ใช้ฟังก์ชันเรกติไฟเออร์เชิงเส้นเป็นฟังก์ชันกระตุ้น อาจทำให้ความเป็น nonlinear ของตัวแบบมีค่อนข้างน้อย ผู้วิจัยจึงได้นำเสนอการเพิ่มชั้นคอนโวลูชันที่เชื่อมต่อกับฟังก์ชันเรกติไฟเออร์เชิงเส้นเข้าไปในแต่ละชั้น เพื่อเพิ่มระดับความซับซ้อนของการแปลงไม่เชิงเส้น (degree of nonlinearity transform) ให้กับโครงข่าย ซึ่งน่าจะตอบโจทย์การวิเคราะห์การถดถอยได้ดีขึ้น

สำหรับในโครงสร้างพื้นฐานการคำนวณในชั้นการเชื่อมต่อเต็มรูปแบบได้ใช้คุณลักษณะแบบไม่เชิงเส้นที่แยกได้ในชั้นคอนโวลูชันสุดท้ายเพื่อการประมวลผลในชั้นการเชื่อมต่อแบบสมบูรณ์ ซึ่งการใช้คุณลักษณะในขั้นสุดท้ายจะมีการรู้จำวัตถุในระดับที่สูง ทำให้สามารถรู้ตำแหน่งคร่าว ๆ ของดวงตาได้ดีขึ้น แต่ความแม่นยำของการระบุตำแหน่งอาจจะไม่สามารถทำได้ดีมาก เนื่องจากคุณลักษณะในขั้นสุดท้ายเกิดจากการย่อพื้นที่ลงไปเรื่อย ๆ ทำให้คุณลักษณะในขั้นสุดท้ายมีขนาดเล็กที่สุด และในงานวิจัยนี้ขนาดข้อมูลเข้ามีขนาดเล็กอยู่แล้วด้วย จึงอาจทำให้คุณลักษณะที่ในขั้นสุดท้ายไม่สามารถนำมาวิเคราะห์ตำแหน่งของจุดศูนย์กลางที่ต้องการความแม่นยำระดับพิกเซลได้ เพื่อไม่ให้ข้อมูลบางอย่างถูกละเลยและเป็นการเปิดโอกาสให้เครื่องสามารถนำข้อมูลต่าง ๆ ในแง่ของตำแหน่งคร่าว ๆ และตำแหน่งที่แม่นยำมาทำการอนุมานให้เป็นผลลัพธ์ที่ดีได้ ผู้วิจัยจึงคิดว่ามีความจำเป็นที่จะต้องดึง

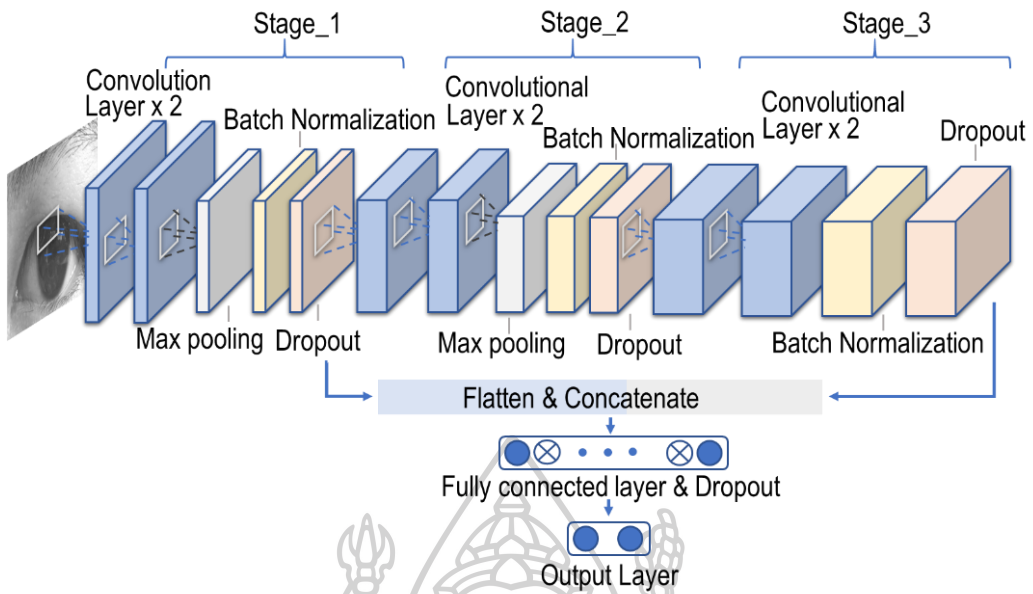
คุณลักษณะที่แยกได้ในชั้นก่อนหน้ามาประมวลผลร่วมกันด้วย โดยในการทดลองได้ออกแบบการนำคุณลักษณะในชั้นต่าง ๆ มารวมกันในรูปแบบที่แตกต่างกันออกไป

ก่อนที่จะนำเสนอรูปแบบการนำมารวมกันของคุณลักษณะในชั้นต่าง ๆ เพื่อให้เกิดความเข้าใจที่ตรงกัน ผู้วิจัยขออนุญาตชื่อที่ใช้เรียกคุณลักษณะเพิ่มเติมที่กล่าวถึงก่อนหน้าก่อน โดยต่อไปนี้โครงสร้างของการถดถอยเชิงคอนโวลูชันที่มีการใช้คอนโวลูชัน 2 ชั้นในแต่ละชั้น จะถูกเรียกว่า ‘DoubleCNN’ ดังนั้นหากไม่มีคำว่า ‘Double’ ในชื่อของตัวแบบหมายความว่า ตัวแบบมีโครงสร้างที่ใช้คอนโวลูชันเพียงชั้นเดียวในแต่ละชั้นคอนโวลูชัน และโครงสร้างของการถดถอยเชิงคอนโวลูชันที่นำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันจะถูกเรียกว่า ‘ConcatCNN’ เช่นเดียวกันหากไม่มีคำว่า ‘Concat’ ในชื่อตัวแบบหมายความว่า ตัวแบบใช้คุณลักษณะที่แยกได้ในชั้นคอนโวลูชันสุดท้ายเท่านั้นในการนำไปคำนวณที่ชั้นการเชื่อมต่อเต็มรูปแบบ

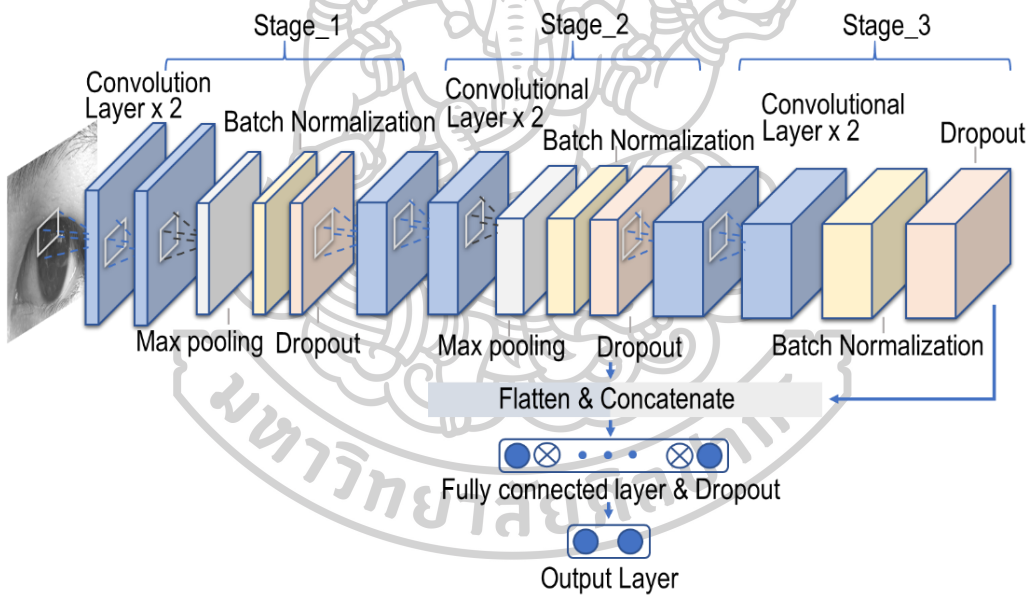
สำหรับการนำมารวมกันของคุณลักษณะจากชั้นคอนโวลูชันในชั้นต่าง ๆ ได้ออกแบบให้มีลักษณะที่แตกต่างกันดังตารางที่ 4.1 และเพื่อแสดงให้เห็นถึงรายละเอียดโครงสร้างของตัวแบบที่นำเสนออย่างชัดเจนสามารถดูโครงสร้างของทั้งตัวแบบได้จากภาพที่ 4.13

ตารางที่ 4.1 ตารางแสดงโครงสร้างของการนำคุณลักษณะจากชั้นต่าง ๆ มาประมวลผลร่วมกันในชั้นการเชื่อมต่อแบบสมบูรณ์

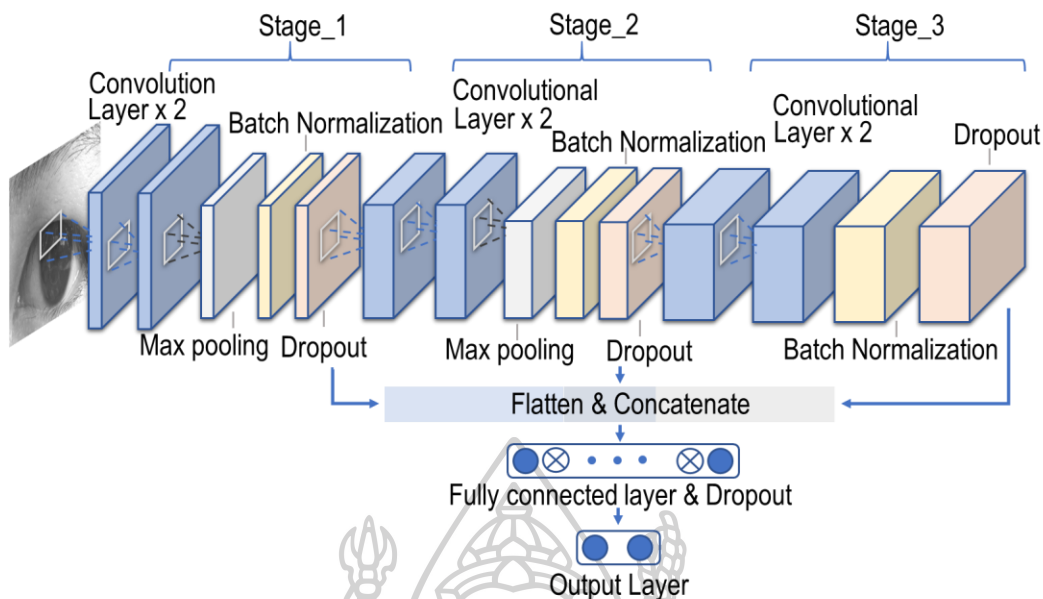
	ชั้นการเชื่อมต่อแบบสมบูรณ์ประมวลผลจาก		
	คุณลักษณะจากชั้นที่ 1	คุณลักษณะจากชั้นที่ 2	คุณลักษณะจากชั้นที่ 3
ConcatCNNv1	✓		✓
ConcatCNNv2		✓	✓
ConcatCNNv3	✓	✓	✓



ก) โครงสร้างของ DoubleConcatCNNv1



ข) โครงสร้างของ DoubleConcatCNNv2



ค) โครงสร้างของ DoubleConcatCNNv3

ภาพที่ 4.13 โครงสร้างของตัวแบบ DoubleConcatCNN ที่นำเสนอ

ในการทดสอบสมมติฐานว่าการเพิ่มคุณลักษณะ Concat และ Double นั้นช่วยให้ตัวแบบมีความแม่นยำและความคงทนมากขึ้น ผู้วิจัยได้ทดลองเปรียบเทียบประสิทธิภาพของตัวแบบผ่านการวิเคราะห์แบบตัดออก (Ablation Analysis) โดยเริ่มจากการทดสอบตัวแบบพื้นฐานหรือ SimpleCNN และค่อย ๆ เพิ่มคุณลักษณะ Concat และ Double เข้าไปเพื่อให้เห็นความแตกต่างของประสิทธิภาพที่เกิดจากแต่ละคุณลักษณะอย่างชัดเจน

4.2.3 การวัดประสิทธิภาพ

ในงานวิจัยนี้ได้แบ่งการวัดประสิทธิภาพออกเป็น 4 ขั้นตอน ดังรายละเอียดด้านล่าง โดยการประเมินประสิทธิภาพของตัวแบบจะถูกดำเนินการฝึกฝนและทดสอบตัวแบบผ่านหน่วยประมวลผลกลาง Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz การ์ดแสดงผล NVIDIA GeForce GTX 1070 3 GB GDDR5 ซึ่งในกระบวนการเรียนรู้เชิงลึกได้ใช้วิธีปรับปรุงค่าน้ำหนักของเส้นเชื่อมในโครงข่ายประสาทเทียมด้วย Adam optimizer และใช้อัตราการเรียนรู้เท่ากับ 0.001

4.2.3.1 การใช้เทคนิค 5-Fold Cross Validation สลับเวียนข้อมูลที่ใช้ในการฝึกฝนและการตรวจสอบ โดยในที่นี้ได้นำชุดข้อมูลฝึกทั้งหมดมาแบ่งเป็น 5 ส่วนเท่า ๆ กัน เพื่อทำการทดสอบทั้งหมด 5 รอบ ในรอบแรกให้ข้อมูลส่วนแรกเป็นข้อมูลตรวจสอบ ที่เหลือเป็นข้อมูลสำหรับฝึกสอนตัวแบบ รอบที่สองนำข้อมูลส่วนที่สองเป็นข้อมูลตรวจสอบ และใช้ส่วนที่ 1 3 ถึง 5 สำหรับฝึกสอน และ

ทำวนจนครบตามจำนวน 5 รอบ เพื่อเลือกตัวแบบที่มีค่า validation loss ต่ำที่สุด มาเปรียบเทียบกับกันในกลุ่มข้อมูลทดสอบ (Testing data)

4.2.3.2 การใช้มาตรวัดระยะทางแบบยูคลิเดียน (Euclidean distance) วัดความแม่นยำของตัวแบบสำหรับการระบุตำแหน่งจุดศูนย์กลางรูปร่างตามชุดข้อมูลทดสอบ โดยมาตรวัดระยะทางแบบยูคลิเดียนเป็นมาตรวัดมาตรฐานที่นิยมใช้ในงานวิเคราะห์การถดถอย สามารถอธิบายความคลาดเคลื่อนได้อย่างเรียบง่ายและชัดเจนจากการวัดระยะทางระหว่างจุดศูนย์กลางรูปร่างจริงและจุดศูนย์กลางที่ทำนายได้ดังสมการที่ 4.1

$$d = \sqrt{(x - \tilde{x})^2 + (y - \tilde{y})^2} \quad \text{สมการที่ 4.1}$$

เมื่อ (x, y) คือ พิกัดของจุดศูนย์กลางรูปร่างจริง และ (\tilde{x}, \tilde{y}) คือ พิกัดของจุดศูนย์กลางรูปร่างที่ทำนายได้

ในการวิเคราะห์ความแม่นยำของตัวแบบจะพิจารณาจากค่าเฉลี่ยของความผิดพลาดเชิงตำแหน่งที่ได้จากมาตรวัดระยะทางแบบยูคลิเดียนและอัตราความสำเร็จหรือ success rate ของการระบุจุดศูนย์กลางเฉลี่ยที่มีความผิดพลาดน้อยกว่า 2 พิกเซล ซึ่งสำหรับระยะทาง 2 พิกเซลเป็นระยะที่งานวิจัยนี้กำหนดให้เป็นความคลาดเคลื่อนที่ยอมรับได้ของการระบุตำแหน่งรูปร่าง เนื่องจากในขั้นตอนของการสร้างผลเฉลยให้กับรูปภาพนั้นอาจมีความคลาดเคลื่อนได้ในระยะ 1-2 พิกเซล ดังนั้นจึงถือว่าการทดสอบที่มีความคลาดเคลื่อนในระยะ 2 พิกเซลเป็นสิ่งที่ยอมรับได้และไม่กระทบกับการใช้งานในจุดประสงค์ที่ตั้งไว้

นอกจากนี้เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่นำเสนอกับงานก่อนหน้า ได้นำมาตรวัด Wec และ Bec (ดูสมการที่ 2.3 และสมการที่ 2.4) มาใช้ด้วย เนื่องข้อจำกัดของการใช้มาตรวัดระยะทางแบบยูคลิเดียนที่สามารถนำไปเปรียบเทียบได้เมื่อสเกลการวัดของแต่ละตัวแปรเท่ากันเท่านั้น ในปัญหาการระบุจุดศูนย์กลางรูปร่างตามสเกลของมาตรวัดยูคลิเดียนจะมีค่าแตกต่างกันไปตามขนาดและความละเอียดของรูปภาพ ดังนั้นผู้วิจัยจึงเลือกใช้มาตรวัด Wec และ Bec มาช่วยในการเปรียบเทียบประสิทธิภาพ

4.2.3.3 การวิเคราะห์ความเหมาะสมของตัวแบบโดยใช้แผนภาพเศษเหลือ คือการนำเศษเหลือ (Residual) หรือค่าความแตกต่างระหว่างค่าจริงและค่าที่ทำนายได้จากตัวแบบ มาสร้างเป็นแผนภาพกระจายซึ่งสามารถทำได้หลายรูปแบบ ภายในงานวิจัยนี้ใช้ 1.) แผนภาพการกระจายระหว่างเศษเหลือและค่าที่ทำนายได้จากตัวแบบตามพิกัดในแนวแกน x และแนวแกน y 2.) แผนภาพ

การกระจายระหว่างเศษเหลือและค่าจริงตามพิกัดในแนวแกน x และ y 3.) ฮิสโทแกรมของเศษเหลือในแนวแกน x และ y

4.2.3.4 การวิเคราะห์ความคงทนของตัวแบบที่นำเสนอ โดยจะพิจารณาใน 2 ระดับคือ

4.2.3.4.1 ระดับความน่าเชื่อถือของเครื่องมือหรือตัวแบบที่ใช้ เป็นการตรวจสอบความแปรปรวนของตัวแบบโดยพิจารณาจากค่าสัมประสิทธิ์ความแปรผัน (coefficient of variation, CV) ของค่าเฉลี่ยค่าคลาดเคลื่อนที่ได้จากการทำซ้ำ 5 รอบในแต่ละการทดลอง เพื่อสร้างความเชื่อมั่นในผลการวิเคราะห์ในระดับต่อไปว่าความแตกต่างที่เกิดขึ้นเกิดจากตัวแปรที่ศึกษาจริงไม่ได้เกิดจากความแปรปรวนของตัวแบบเอง โดยถ้าค่าสัมประสิทธิ์ความแปรผันมีค่าน้อย แสดงว่าการวัดซ้ำมีความคงเส้นคงวาหรือความเที่ยงตรงสูง ตามมาตรฐานการวิเคราะห์ค่าสัมประสิทธิ์ความแปรผันน้อยกว่าร้อยละ 5 ถือว่ายอมรับได้

4.2.3.4.2 ระดับการเรียนรู้ ในขั้นตอนนี้จะพิจารณาส่วนเบี่ยงเบนมาตรฐาน (standard deviation, SD) ของค่าคลาดเคลื่อนเฉลี่ยตามการแปรเปลี่ยนของชุดข้อมูลทดสอบ ในที่นี้จะหาส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนเฉลี่ยที่ได้จากฐานข้อมูล MMU2, BioID และ FFHQ หากส่วนเบี่ยงเบนมาตรฐานมีค่าน้อย หมายความว่า การเปลี่ยนแปลงข้อมูลเข้าที่อยู่ภายใต้ภาพจากสามฐานข้อมูลนี้ ไม่สามารถทำให้ความแม่นยำที่ได้เปลี่ยนแปลงไปหรือไม่ส่งผลต่อความถูกต้องแม่นยำอย่างมีนัยสำคัญ ดังนั้นเมื่อนำไปใช้กับภาพดวงตาอื่น ๆ ที่อาจมีลักษณะบางประการที่เปลี่ยนไป ความแม่นยำก็น่าจะไม่เปลี่ยนไปจากเดิมมาก หรือเรียกได้ว่ามีความคงทนนั่นเอง



บทที่ 5

ผลการดำเนินงานวิจัย

บทนี้จะกล่าวถึงผลการทดลองเพื่อประเมินประสิทธิภาพของวิธีการและเทคนิคที่ใช้ในการระบุจุดศูนย์กลางรูปร่างตามที่ได้กล่าวไว้ในบทที่ 4 โดยได้แบ่งการแสดงผลการทดลองออกเป็น 6 การทดลอง เพื่อประเมินความแม่นยำและพฤติกรรมของตัวแบบในแง่มุมที่แตกต่างกัน โดยการทดลองที่ 1, 2, 3, 4 จะเป็นเรื่องของการทดสอบความแม่นยำในการระบุตำแหน่งจุดศูนย์กลางรูปร่างตามของตัวแบบต่าง ๆ ที่ศึกษาในงานวิจัยนี้ เพื่อประเมินผลกระทบต่อความแม่นยำของวิธีการที่นำเสนอผ่านการวิเคราะห์แบบตัดออก ซึ่งวิธีการที่งานนี้แนะนำและวิเคราะห์มีอยู่สามส่วนคือ (ก) การซ้อนชั้นคอนโวลูชันที่มากขึ้น (ข) การนำคุณลักษณะจากหลายชั้นมาประมวลรวมกันเพื่อระบุตำแหน่งรูปร่างตามและ (ค) การใช้คอนโวลูชันแบบขยาย (atrous convolution) ทั้งนี้ผลลัพธ์ที่นำเสนอในการวิเคราะห์แบบตัดออกนี้ คัดมาเฉพาะตัวแบบที่มีศักยภาพสูงมาเป็นฐานในการวิเคราะห์ ไม่ได้แนะนำเสนอทุกการจัดกลุ่ม (combination) ที่เป็นไปได้เพื่อไม่ให้ข้อมูลในภาพและตารางมีมากเกินไป และเพื่อให้ผู้อ่านได้เห็นอิทธิพลของเทคนิคที่กำลังศึกษาได้ง่ายขึ้น ส่วนในการทดลองที่จัดทำขึ้นมาก่อนหน้ามีการทดลองจัดกลุ่มตัวแบบอีกเป็นจำนวนมาก ทว่าสมรรถนะของตัวแบบเหล่านั้นด้อยกว่าตัวแบบที่คัดเลือกมานี้ จึงได้ตัดออกไปจากผลลัพธ์ในการทดลอง สำหรับการทดลองที่ 5 เป็นการเปรียบเทียบความแม่นยำของตัวแบบที่ดีที่สุดที่งานวิจัยนี้ได้นำเสนอกับงานวิจัยที่เกี่ยวข้องและการทดลองที่ 6 เป็นการทดสอบพฤติกรรมของตัวแบบเพิ่มเติมเมื่อฝึกสอนตัวแบบด้วยภาพที่มีลักษณะแตกต่างออกไป

การทดลองที่ 1 การวิเคราะห์การถดถอยที่มีระดับของการแปลงไม่เชิงเส้นที่เพิ่มขึ้น

ในการทดลองนี้มีวัตถุประสงค์ในการพิสูจน์ว่าการเพิ่มระดับของการแปลงไม่เชิงเส้นโดยการเพิ่มชั้นคอนโวลูชันที่เชื่อมต่อกับฟังก์ชันเรกติไฟเออร์เชิงเส้นส่งผลต่อความแม่นยำและความคงทนของการวิเคราะห์การถดถอยจุดศูนย์กลางรูปร่างตามอย่างมีนัยสำคัญ จึงได้เปรียบเทียบตัวแบบการเรียนรู้ที่ได้จากโครงสร้าง SimpleCNN กับโครงสร้าง DoubleCNN โดยนำมาทดสอบกับชุดข้อมูลสามชุด ได้แก่ MMU2, BioID และ FFHQ

ผลการทดสอบแสดงเป็นข้อมูลตัวเลขดังตารางที่ 5.1 ได้แก่ อัตราความสำเร็จ (% Success) ของการระบุจุดศูนย์กลางเฉลี่ยที่มีความผิดพลาดน้อยกว่า 2 พิกเซล, ค่าเฉลี่ยของความผิดพลาดเชิงตำแหน่ง (Average Euclidean Error, Avg Euc Err) ตามที่นิยามไว้ในสมการที่ 4.1, ค่าสัมประสิทธิ์

ความแปรผันของค่าเฉลี่ยค่าคลาดเคลื่อนที่ได้จากการทำซ้ำ 5 รอบ (CV) และส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนเฉลี่ยตามการแปรเปลี่ยนของชุดข้อมูลทดสอบ (SD)

ตารางที่ 5.1 ตารางเปรียบเทียบประสิทธิภาพของ SimpleCNN และ DoubleCNN

Model name	MMU2			BioID			FFHQ			SD
	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	
SimpleCNN	62.77	2.57	5.00	93.57	1.27	4.50	74.77	2.09	2.78	0.66
DoubleCNN	83.86	1.65	15.53	92.69	1.32	5.24	83.25	1.75	3.35	0.23

จากผลลัพธ์ของการทดลองข้างต้นจะพบว่าการเพิ่มระดับของการแปลงไม่เชิงเส้นเข้าไปในตัวแบบมีผลกับความแม่นยำของการระบุศูนย์กลางรูปร่าง โดยที่แนวโน้มที่จะทำให้มีความแม่นยำเพิ่มขึ้นโดยเฉพาะอย่างยิ่งกับฐานข้อมูล MMU2 ที่ทำให้อัตราความสำเร็จของการระบุศูนย์กลางรูปร่างที่อยู่ในระดับความคลาดเคลื่อนที่ยอมรับได้เพิ่มขึ้นเป็นร้อยละ 83.86 จากเดิมอยู่ที่ร้อยละ 62.77 และจากค่าความคลาดเคลื่อนเฉลี่ย 2.57 พิกเซลลดลงเหลือ 1.65 พิกเซล สำหรับบนฐานข้อมูล FFHQ อัตราความสำเร็จของการระบุศูนย์กลางรูปร่างที่อยู่ในระดับความคลาดเคลื่อนที่ยอมรับได้เพิ่มขึ้นเป็นร้อยละ 83.25 จากเดิมอยู่ที่ร้อยละ 74.77 และจากค่าความคลาดเคลื่อนเฉลี่ย 2.09 พิกเซลลดลงเหลือ 1.75 พิกเซล แต่สำหรับบนฐานข้อมูล BioID นั้นยังไม่เกิดการเปลี่ยนแปลงที่เห็นได้ชัด โดย SimpleCNN และ DoubleCNN ให้อัตราความสำเร็จร้อยละ 93.57 และร้อยละ 92.69 และความคลาดเคลื่อนเฉลี่ย 1.27 และ 1.32 พิกเซล อย่างไรก็ตามทั้งสองตัวแบบมีความแม่นยำที่ค่อนข้างสูงบนฐานข้อมูล BioID

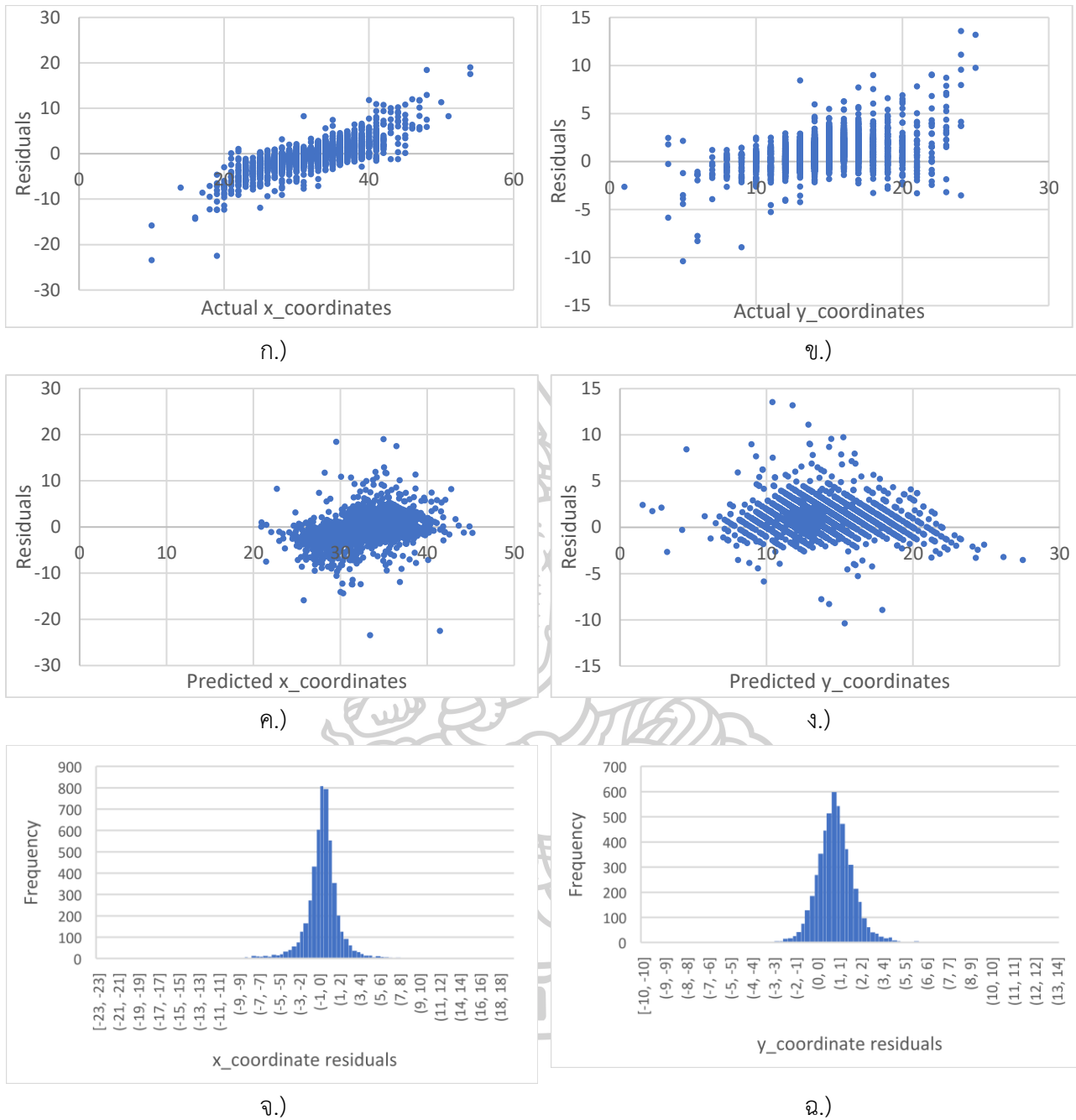
ผลการทดสอบความคงทนการจากการฝึกตัวแบบซ้ำ 5 รอบในแต่ละทดลองแสดงให้เห็นว่าความแปรปรวนเมื่อใช้ DoubleCNN มีค่ามากกว่า SimpleCNN สำหรับบนฐานข้อมูล MMU2 จะเห็นได้ชัดเจนว่ามีความแปรปรวนเพิ่มขึ้นสูงสุดโดยมีค่าสัมประสิทธิ์ความแปรผันเท่ากับร้อยละ 15.53 แสดงว่าตัวแบบ DoubleCNN มีความไม่เที่ยงตรงบนฐานข้อมูล MMU2 แต่สำหรับฐานข้อมูล

BioID และ FFHQ แม้ความแปรปรวนจะเพิ่มสูงมากขึ้นแต่ค่าสัมประสิทธิ์ความแปรผันยังอยู่ในระดับที่พอรับได้

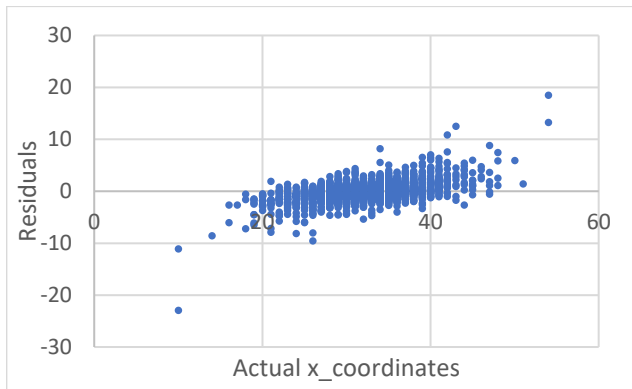
สำหรับการกระจายของค่าความคลาดเคลื่อนเฉลี่ยบนฐานข้อมูลทั้งสามพบว่า DoubleCNN มีการกระจายที่น้อยกว่าจากส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.23 อย่างไรก็ตามยังไม่สามารถสรุปได้ว่าความคงทนของ DoubleCNN นั้นดีกว่า SimpleCNN เนื่องจากความน่าเชื่อถือของตัวแบบ DoubleCNN ต่ำกว่าตัวแบบ SimpleCNN และบางส่วนมีค่าเกินเกณฑ์ที่ยอมรับได้โดยทั่วไปด้วย

นอกจากการวิเคราะห์ด้วยตัวเลขข้างต้นได้นำแผนภาพส่วนเหลือมาพิจารณาด้วยร่วมด้วยดังแผนภาพที่ 5.1 และภาพที่ 5.2 เนื่องจากแผนภาพเศษเหลือสามารถบ่งบอกถึงความเอนเอียงไปจนถึงความผิดปกติในตัวแบบได้ดีกว่าข้อมูลตัวเลข สำหรับการใช้ตัวแบบ SimpleCNN ผลการตรวจสอบพบว่าลักษณะการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x และ y มีแนวโน้มเพิ่มมากขึ้น โดยเมื่อ x มีค่าต่ำเศษเหลือจะมีค่าเป็นลบ และเมื่อ x มีค่าสูงเศษเหลือจะมีค่าเป็นบวก นอกจากนี้แผนภาพการกระจายยังมีลักษณะเป็นรูปปากแตรเมื่อค่า y อยู่ที่บริเวณขอบด้วย แสดงว่ายิ่งค่า x และ y มีค่าสูงมากหรือต่ำมากหรือใกล้เคียงบริเวณขอบ ภาพมีแนวโน้มจะทำนายด้วยความคลาดเคลื่อนที่มากขึ้น โดยตัวแบบจะทำนายเอนเอียงไปยังตำแหน่งจุดศูนย์กลางภาพ และจากการตรวจสอบการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x และแกน y พบว่าเศษเหลือมีการกระจายอย่างสม่ำเสมอ ๆ ค่าศูนย์ แสดงให้เห็นถึงความแปรปรวนของค่าเศษเหลือคงที่และเป็นอิสระกับค่าที่ทำนายได้ สุดท้ายแผนภาพฮิสโตแกรมแสดงให้เห็นว่าการกระจายของค่าคลาดเคลื่อนมีลักษณะของการแจกแจงแบบปกติ

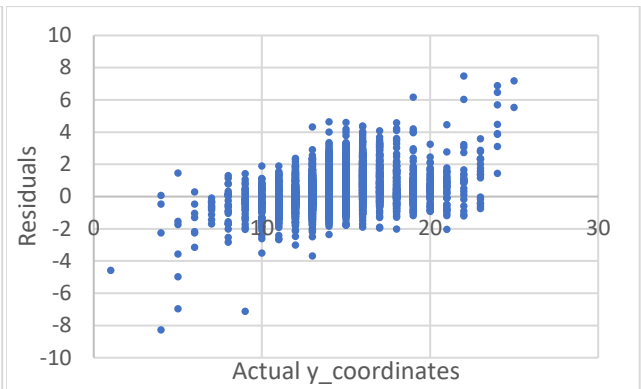
สำหรับการใช้ตัวแบบ DoubleCNN ผลการตรวจสอบพบว่าลักษณะการกระจายเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x และ y มีลักษณะกระจายแบบสม่ำเสมออยู่ภายในแนวนอนรอบ ๆ ค่าศูนย์ แต่ยังคงมีแนวโน้มที่เมื่อ x และ y มีค่าสูงมากหรือต่ำมากหรือใกล้เคียงบริเวณขอบภาพ เศษเหลือจะมีค่ามากขึ้นหรือน้อยลงตามไปด้วย ซึ่งแสดงถึงว่าตัวแบบมีความเอนเอียงเล็กน้อย และจากการตรวจสอบการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x และ y พบว่าเศษเหลือมีการกระจายอย่างสม่ำเสมอ ๆ ค่าศูนย์ แสดงให้เห็นถึงความแปรปรวนของค่าเศษเหลือคงที่และเป็นอิสระกับค่าที่ทำนายได้ สุดท้ายแผนภาพฮิสโตแกรมแสดงให้เห็นว่าการกระจายของค่าคลาดเคลื่อนมีลักษณะของการแจกแจงแบบปกติ



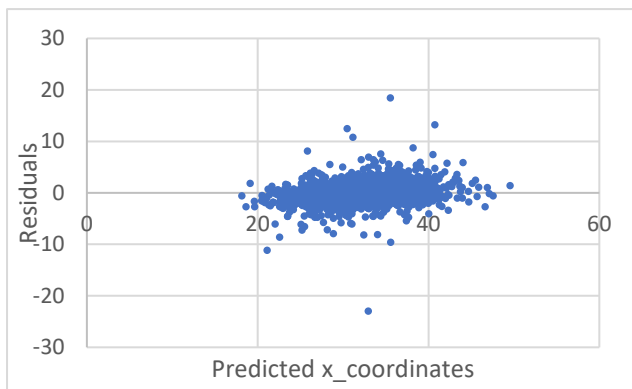
ภาพที่ 5.1 แผนภาพเศษเหลือของตัวแบบ SimpleCNN ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



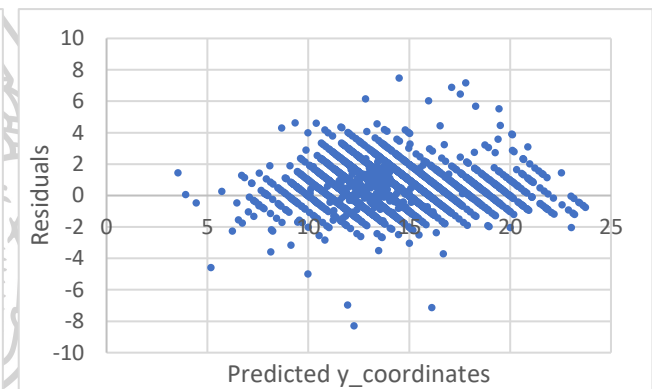
ก.)



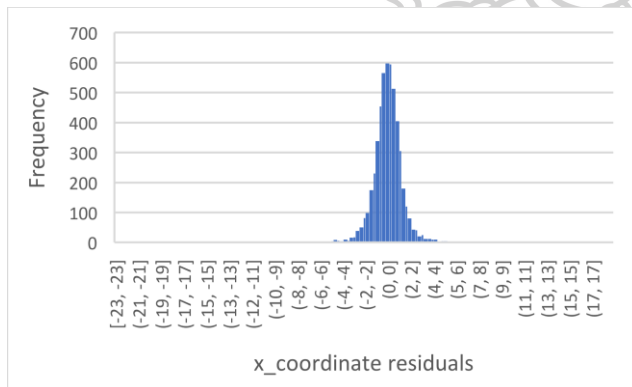
ข.)



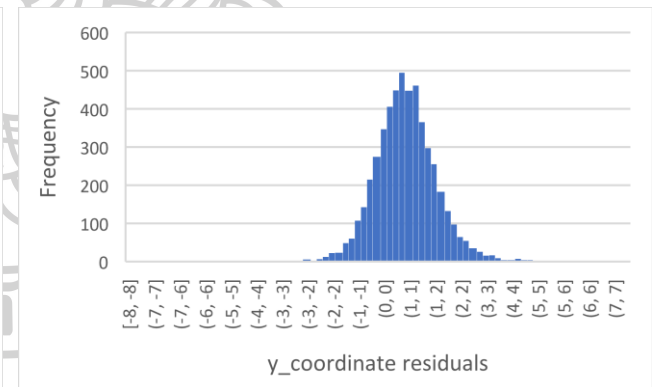
ค.)



ง.)



จ.)



ฉ.)

ภาพที่ 5.2 แผนภาพเศษเหลือของตัวแบบ DoubleCNN ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y

การทดลองที่ 2 การวิเคราะห์การถดถอยที่นำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกัน

การทดลองนี้มีจุดประสงค์เพื่อพิสูจน์ว่าการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันส่งผลต่อความแม่นยำและความคงทนของการวิเคราะห์การถดถอยจุดศูนย์กลางรูปร่างตามที่มีนัยสำคัญ ผู้วิจัยจึงได้ทดสอบตัวแบบพื้นฐาน SimpleCNN เปรียบเทียบกับตัวแบบที่มีรายละเอียดโครงสร้างดังตารางที่ 4.1 ซึ่งออกแบบให้มีการนำมารวมกันของคุณลักษณะไว้แตกต่างกัน เพื่อทดสอบผลกระทบของการนำมารวมกันจากชั้นต่าง ๆ ผลการทดสอบแสดงในตารางที่ 5.2 ซึ่งมีการวัดสมรรถนะเป็นไปในลักษณะเดียวกับกับตารางที่ 5.1

ตารางที่ 5.2 ตารางเปรียบเทียบประสิทธิภาพของ SimpleCNN, ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3

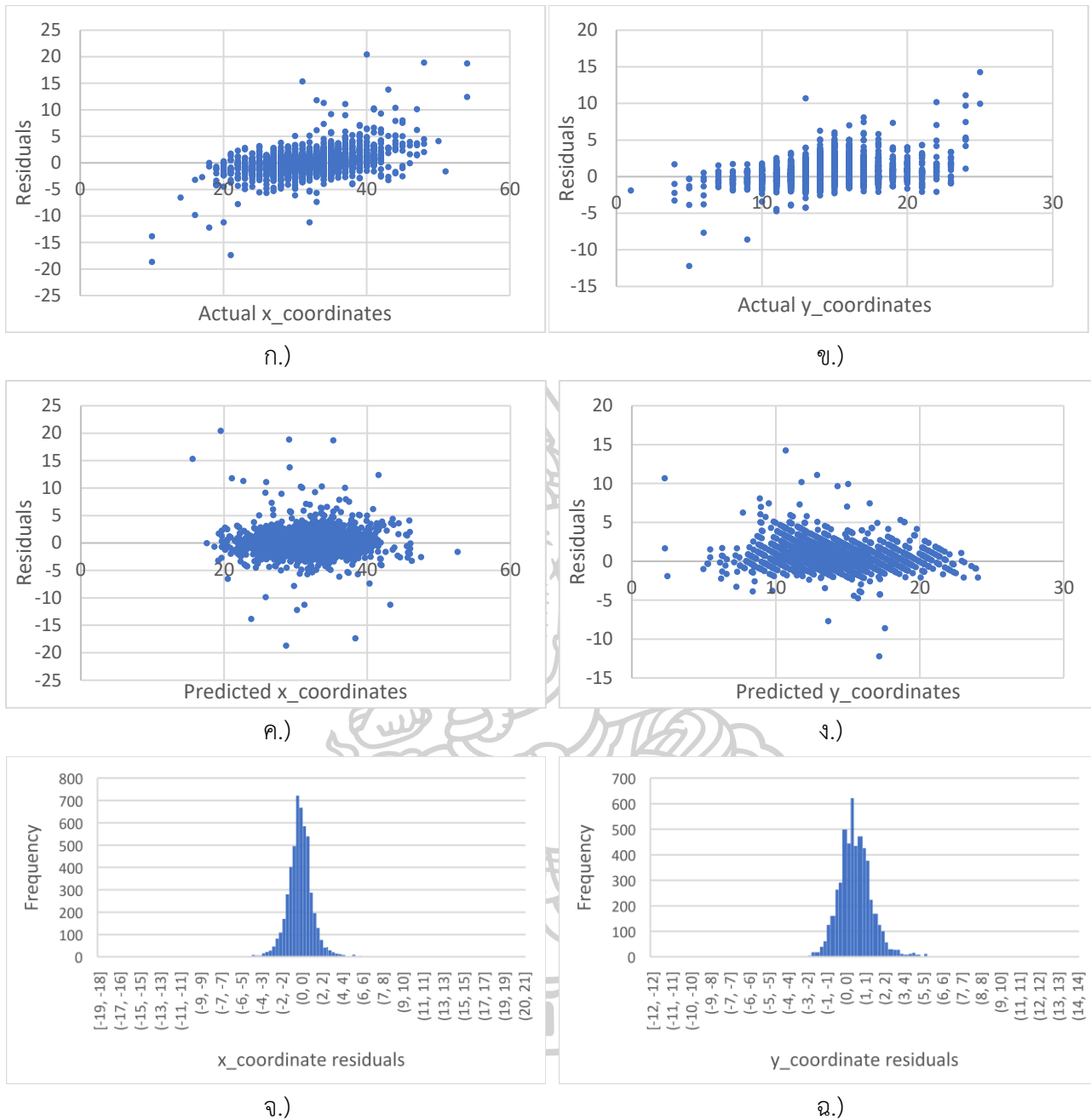
Model name	MMU2			BioID			FFHQ			
	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	SD
SimpleCNN	62.77	2.57	5.00	93.57	1.27	4.50	74.77	2.09	2.78	0.66
ConcatCNNv1	84.48	1.67	8.15	92.16	1.35	3.96	78.75	1.91	3.74	0.28
ConcatCNNv2	83.38	1.73	7.13	92.59	1.34	5.48	80.80	1.85	3.12	0.27
ConcatCNNv3	82.93	1.70	4.98	92.60	1.37	3.20	79.41	1.90	3.60	0.27

จากผลการทดลองสังเกตได้ว่าตัวแบบ ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3 มีประสิทธิภาพใกล้เคียงกัน จึงไม่สามารถสรุปได้ว่าชั้นใดที่ส่งผลต่อประสิทธิภาพของตัวแบบมากกว่ากันอย่างมีนัยสำคัญ แต่ทั้งสามตัวแบบได้แสดงให้เห็นว่าสามารถทำให้ความแม่นยำเพิ่มขึ้นได้โดยบนฐานข้อมูล MMU2 และ FFHQ ให้อัตราความสำเร็จที่สูงกว่า SimpleCNN รวมถึงให้ค่าคลาดเคลื่อนเฉลี่ยที่ลดลงด้วย การนำคุณลักษณะจากชั้นก่อนหน้ามาประมวลผลร่วมกันในชั้นการ

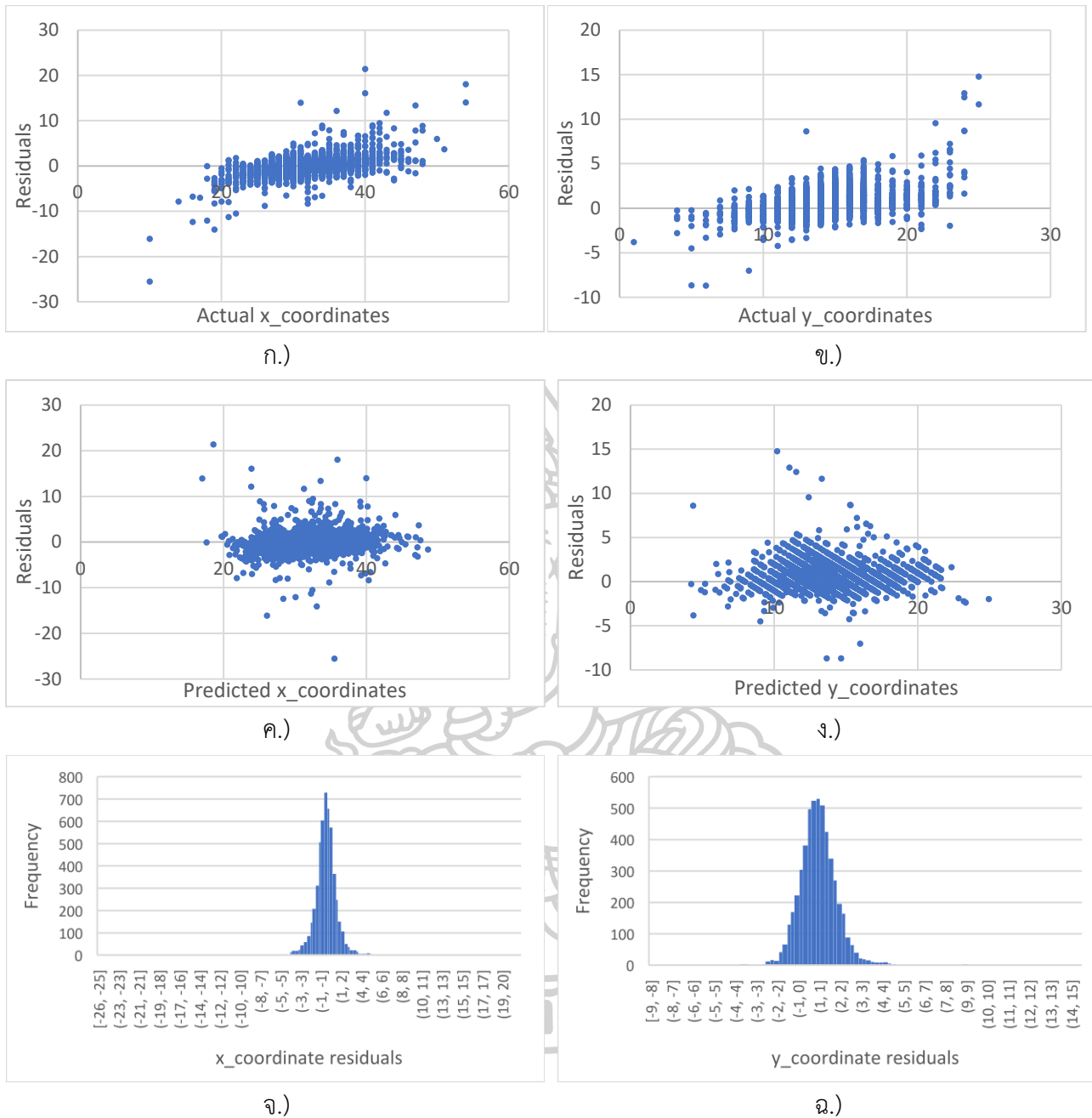
เชื่อมต่อแบบสมบูร์นจึงมีแนวโน้มที่จะส่งผลให้ความแม่นยำของตัวแบบดีขึ้น แต่จะสังเกตว่าการใช้เทคนิคที่เพิ่มขึ้นในสองการทดลองที่ผ่านมายังไม่สามารถทำให้ความแม่นยำของตัวแบบเปลี่ยนแปลงไปอย่างมีนัยสำคัญได้บนฐานข้อมูล BioID

จากตารางยังแสดงให้เห็นว่าตัวแบบ ConcatCNNv3 มีความคงทนที่เพิ่มขึ้นจาก SimpleCNN เนื่องจากส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนเฉลี่ยลดลงอย่างเห็นได้ชัดและค่าสัมประสิทธิ์ความแปรผันถือว่าอยู่ในเกณฑ์ที่ยอมรับได้ แต่สำหรับ ConcatCNNv1 และ ConcatCNNv2 แม้มีแนวโน้มว่าจะให้ความคงทนที่เพิ่มมากขึ้นจากการที่ส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนลดลง แต่ยังไม่สามารถสรุปได้แน่ชัดเนื่องจากมีค่าสัมประสิทธิ์ความแปรผันของตัวแบบที่เพิ่มขึ้นเกินร้อยละ 5 ในบางฐานข้อมูล ความน่าเชื่อถือของตัวแบบจึงน้อยลง

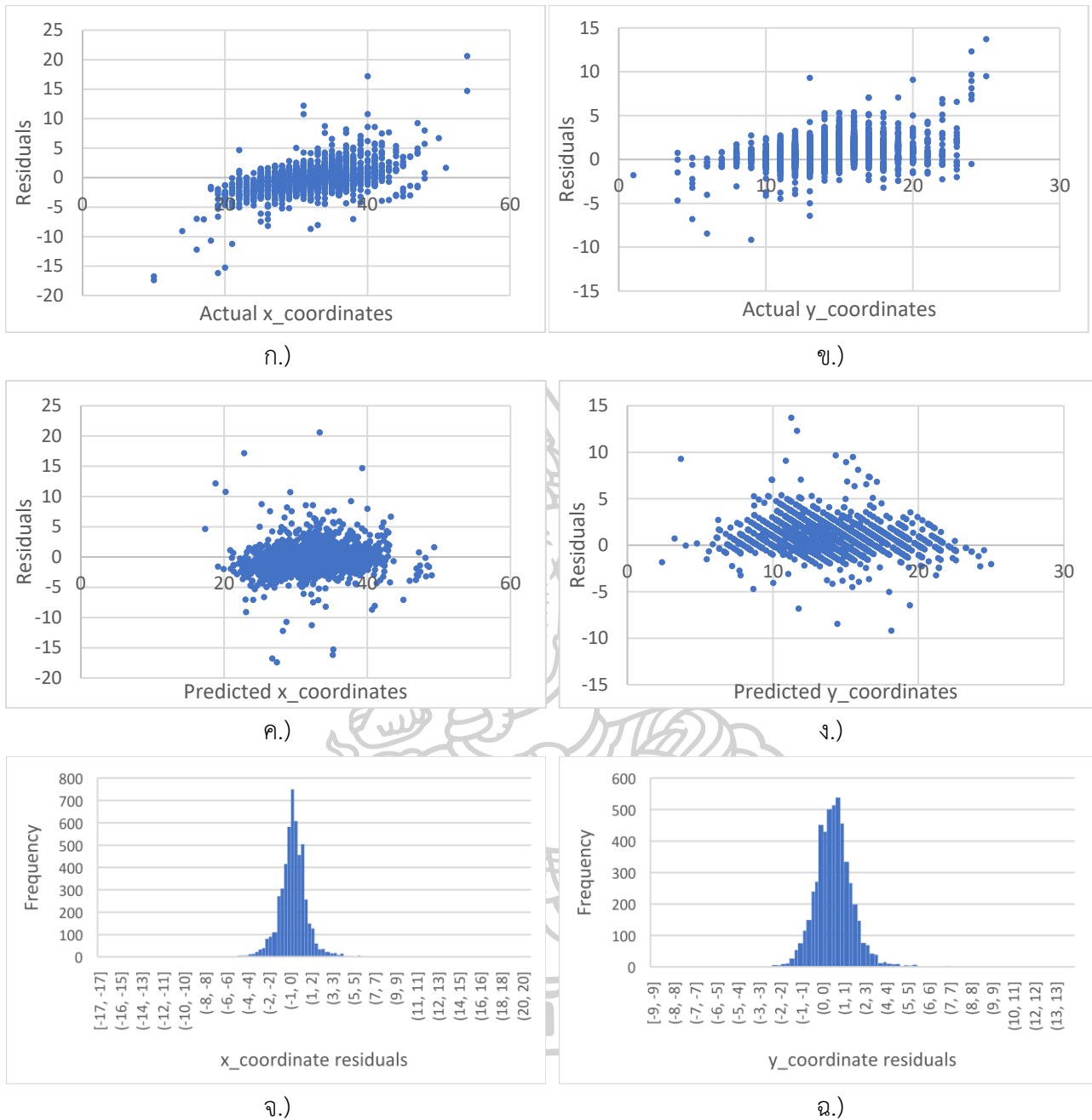
ต่อมาได้ศึกษาความเหมาะสมของตัวแบบเพิ่มเติมโดยพิจารณาจากแผนภาพเศษเหลือดังแสดงในภาพที่ 5.3, ภาพที่ 5.4 และภาพที่ 5.5 พบว่าแผนภาพเศษเหลือของ ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3 มีลักษณะที่เป็นไปในทิศทางเดียวกัน คือ ลักษณะการกระจายเศษเหลือกับพิกัดผลเฉลยในแนวแกน x และ y มีลักษณะกระจายแบบสุ่มอยู่ภายในแนวนอนรอบ ๆ ค่าศูนย์ แต่ยังคงมีแนวโน้มที่เมื่อค่า x และ y มีค่าสูงมากหรือต่ำมากหรือใกล้เคียงบริเวณขอบภาพเศษเหลือจะมีค่ามากขึ้นหรือน้อยลงตามไปด้วย ซึ่งแสดงถึงว่าตัวแบบมีความเอนเอียงเล็กน้อย และจากการตรวจสอบการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x และ y พบว่าเศษเหลือมีการกระจายอย่างสุ่มรอบ ๆ ค่าศูนย์ แสดงให้เห็นถึงความแปรปรวนของค่าเศษเหลือคงที่และเป็นอิสระกับค่าที่ทำนายได้ สุดท้ายแผนภาพฮิสโทแกรมแสดงให้เห็นว่าการกระจายของค่าคลาดเคลื่อนมีลักษณะของการแจกแจงแบบปกติ



ภาพที่ 5.3 แผนภาพเศษเหลือของตัวแบบ ConcatCNNv1 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



ภาพที่ 5.4 แผนภาพเศษเหลือของตัวแบบ ConcatCNNv2 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



ภาพที่ 5.5 แผนภาพเศษเหลือของตัวแบบ ConcatCNNv3 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y

การทดลองที่ 3 การวิเคราะห์การถดถอยที่มีระดับของการแปลงไม่เชิงเส้นที่เพิ่มขึ้นพร้อมกับการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกัน

สำหรับการทดลองนี้ผู้วิจัยนำเทคนิคที่ใช้ในการทดลองที่ 1 และ 2 มาใช้พร้อมกัน โดยปรับแต่งตัวแบบ ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3 ด้วยการเพิ่มชั้นคอนโวลูชันที่เชื่อมต่อกับฟังก์ชันแรกทีไฟเออร์เชิงเส้นในแต่ละชั้น ทำให้ในแต่ละชั้นมีชั้นคอนโวลูชัน 2 ชั้น จึงเรียกตัวแบบหลังการปรับแต่งด้วยชื่อ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 ตามลำดับ โดยในการทดลองนี้ได้ทดสอบและเปรียบเทียบกับตัวแบบจากการทดลองก่อนหน้า บนฐานข้อมูล MMU2, BioID และ FFHQ ผลการทดลองได้แสดงให้เห็นดังตารางที่ 5.3

จากผลการทดลองสังเกตได้ว่าตัวแบบ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 สามารถให้ความแม่นยำที่สูงที่สุดบนฐานข้อมูล FFHQ, BioID และ MMU2 ตามลำดับ ซึ่งจากความแม่นยำที่ทั้งสามตัวแบบนี้ทำได้ทำให้สามารถสรุปได้ว่า การใช้ระดับของการแปลงไม่เชิงเส้นที่เพิ่มขึ้นพร้อมกับการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันให้ความแม่นยำที่ดีกว่าการใช้เทคนิคทั้งสองแยกกัน

จากตารางยังแสดงให้เห็นว่าตัวแบบ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 มีความเที่ยงตรงของตัวแบบนี้ลดลงเมื่อเปรียบเทียบกับตัวแบบ ConcatCNNv1, ConcatCNNv2 และ ConcatCNNv3 แต่ในขณะเดียวกันส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนก็ลดลงด้วย

ตารางที่ 5.3 ตารางเปรียบเทียบประสิทธิภาพของ DoubleConcatCNNv1, DoubleConcatCNNv2, DoubleConcatCNNv3 และตัวแบบที่นำเสนอก่อนหน้านี้

Model name	MMU2			BioID			FFHQ			SD
	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	
SimpleCNN	62.77	2.57	5.00	93.57	1.27	4.50	74.77	2.09	2.78	0.66
DoubleCNN	83.86	1.65	15.53	92.69	1.32	5.24	83.25	1.75	3.35	0.23
ConcatCNN v1	84.48	1.67	8.15	92.16	1.35	3.96	78.75	1.91	3.74	0.28
ConcatCNN v2	83.38	1.73	7.13	92.59	1.34	5.48	80.80	1.85	3.12	0.27
ConcatCNN v3	82.93	1.70	4.98	92.60	1.37	3.20	79.41	1.90	3.60	0.27
DoubleCon- catCNNv1	80.72	1.75	8.83	92.51	1.31	5.51	84.95	1.69	3.39	0.24
DoubleCon- catCNNv2	89.43	1.44	12.68	93.82	1.24	3.07	82.39	1.73	5.05	0.25
DoubleCon- catCNNv3	90.37	1.39	5.80	94.00	1.25	5.02	83.68	1.73	4.64	0.24

ต่อมาได้ศึกษาความเหมาะสมของตัวแบบเพิ่มเติมโดยพิจารณาจากแผนภาพเศษเหลือดังแสดงในภาพที่ 5.6, ภาพที่ 5.7 และภาพที่ 5.8 พบว่าแผนภาพเศษเหลือของ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 มีลักษณะที่เป็นไปในทิศทางเดียวกัน คือ ลักษณะการกระจายเศษเหลือกับพิกัดผลเฉลยในแนวแกน x และ y มีลักษณะกระจายแบบสุ่มอยู่ภายในแนวนอนรอบ ๆ ค่าศูนย์ แต่ยังคงมีแนวโน้มที่เมื่อค่า x และ y มีค่าสูงมาก

หรือต่ำมากหรือใกล้เคียงบริเวณขอบภาพ เศษเหลือจะมีค่ามากขึ้นหรือน้อยลงตามไปด้วย ซึ่งแสดงถึงว่าตัวแบบมีความเอนเอียงเล็กน้อย และจากการตรวจสอบการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x และ y พบว่าเศษเหลือมีการกระจายอย่างสุ่มรอบ ๆ ค่าศูนย์ แสดงให้เห็นถึงความแปรปรวนของค่าเศษเหลือคงที่และเป็นอิสระกับค่าที่ทำนายได้ สุดท้ายแผนภาพฮิสโทแกรมแสดงให้เห็นว่าการกระจายของค่าคลาดเคลื่อนมีลักษณะของการแจกแจงแบบปกติ

จากผลการทดลองที่ยังไม่สามารถสรุปได้ถึงความแตกต่างของ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 อย่างมีนัยสำคัญ จึงได้ทำการวิเคราะห์เพิ่มเติมในส่วนนี้ โดยใช้การทดสอบที (t-test) แบบจับคู่หรือ Paired t-test ซึ่งเป็นสถิติที่ใช้ในการทดสอบความแตกต่างของข้อมูล 2 ข้อมูลของกลุ่มตัวอย่างเพียงกลุ่มเดียว สำหรับงานนี้จะนำมาทดสอบความแตกต่างของค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv1, DoubleConcatCNNv2 และ DoubleConcatCNNv3 บนข้อมูลชุดทดสอบ ซึ่งมีจำนวนรวมกันทั้งสิ้น 5,137 ภาพ ในที่นี้จะแบ่งการทดสอบเป็น 2 การทดสอบ

- 1.) ทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ DoubleConcatCNNv1 กับ DoubleConcatCNNv2 โดยมีสมมุติฐานหลัก (Null Hypothesis: H_0) คือ ค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv1 และ DoubleConcatCNNv2 ไม่มีความแตกต่างกัน และสมมุติฐานรอง (Alternative Hypothesis: H_a) คือ ค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv1 สูงกว่าค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv2
- 2.) ทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ DoubleConcatCNNv2 กับ DoubleConcatCNNv3 โดยมีสมมุติฐานหลัก (Null Hypothesis: H_0) คือ ค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv2 และ DoubleConcatCNNv3 ไม่มีความแตกต่างกัน และสมมุติฐานรอง (Alternative Hypothesis: H_a) คือ ค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv2 สูงกว่าค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv3

ตารางที่ 5.4 ผลการทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ
DoubleConcatCNNv1 กับ DoubleConcatCNNv2

	DoubleConcatCNNv1	DoubleConcatCNNv2
Mean	1.516883734	1.441638159
Variance	1.400550709	1.189171827
Observations	5137	5137
Pearson Correlation	0.696182783	
Hypothesized Mean Difference	0	
df	5136	
t Stat	6.056880219	
P(T<=t) one-tail	7.43E-10	
t Critical one-tail	1.645150365	
P(T<=t) two-tail	1.48659E-09	
t Critical two-tail	1.960425982	

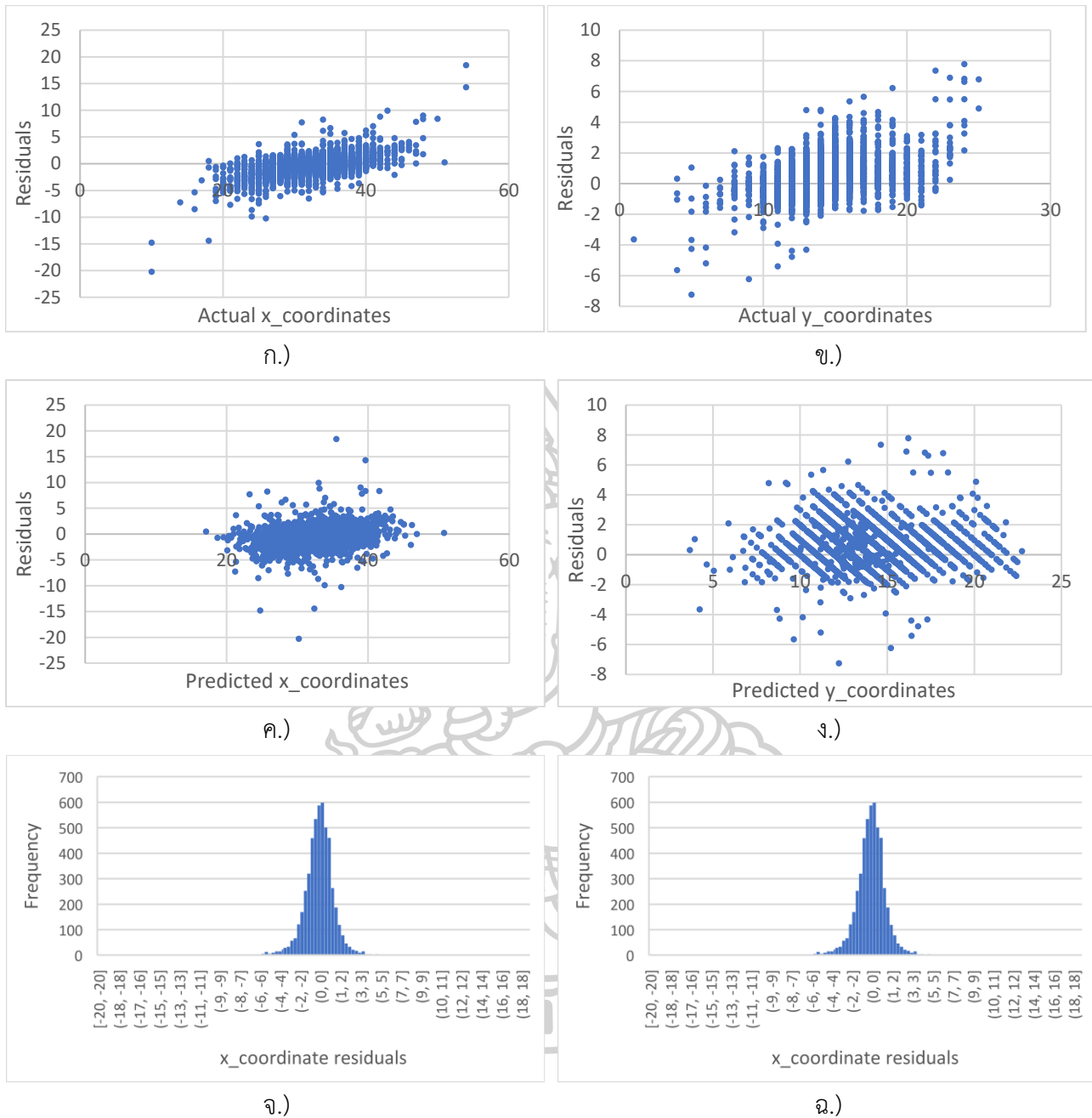
ตารางที่ 5.5 ผลการทดสอบ Paired t-test ระหว่างค่าคลาดเคลื่อนที่ได้จากตัวแบบ
DoubleConcatCNNv2 กับ DoubleConcatCNNv3

	DoubleConcatCNNv2	DoubleConcatCNNv3
Mean	1.441638159	1.397052807
Variance	1.189171827	1.810075912
Observations	5137	5137
Pearson Correlation	0.703972412	
Hypothesized Mean Difference	0	
df	5136	
t Stat	3.307247142	
P(T<=t) one-tail	0.000474327	
t Critical one-tail	1.645150365	
P(T<=t) two-tail	0.000948655	
t Critical two-tail	1.960425982	

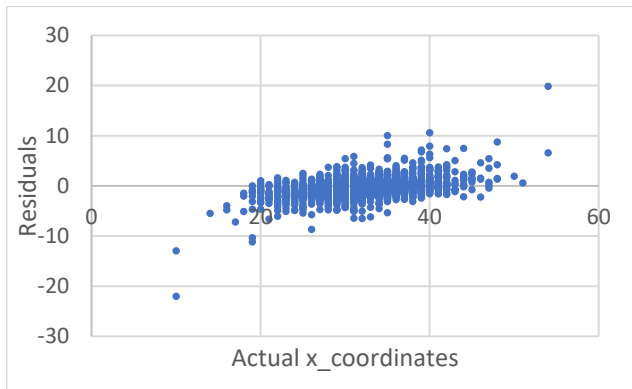
จากตารางผลการทดสอบ Paired t-test ทั้งสอง จะเห็นว่า ค่า t-Stat มากกว่าค่า t-Critical one-tail และ P-value น้อยกว่า 0.05 จึงสรุปได้ว่ามีหลักฐานเพียงพอที่จะปฏิเสธสมมุติฐานหลัก และยอมรับสมมุติฐานรอง นั่นคือ ค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv1 สูงกว่าค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv2 และค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv2 สูงกว่าค่าคลาดเคลื่อนเฉลี่ยที่ได้จากตัวแบบ DoubleConcatCNNv3 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

จากผลการทดลองข้างต้นจึงสรุปว่าตัวแบบมีประสิทธิภาพดีที่สุดคือตัวแบบ DoubleConcatCNNv3 นอกจากนี้สำหรับเวลาที่ใช้ในการประมวลผลพบว่าใช้เวลาเฉลี่ย 0.03 วินาทีต่อภาพ และเวลาที่ใช้ในการประมวลผลเมื่อรวมเวลาการครอปตัดข้อมูลรูปภาพให้เหลือเพียงบริเวณตรงตาพบใช้เวลา 0.05 วินาทีต่อภาพ ซึ่งเมื่อจับเวลาจากการใช้ตัวแบบ SimpleCNN ที่เป็นโครงสร้างการเรียนรู้เชิงลึกที่มีความซับซ้อนไม่มาก ก็ได้ให้ผลเช่นเดียวกัน ทำให้สามารถสรุปได้ว่าการเพิ่ม hyper-parameter ให้กับโครงสร้างด้วยวิธีการที่นำเสนอไม่ทำให้เสียค่าใช้จ่ายในการคำนวณอย่างมีนัยสำคัญ

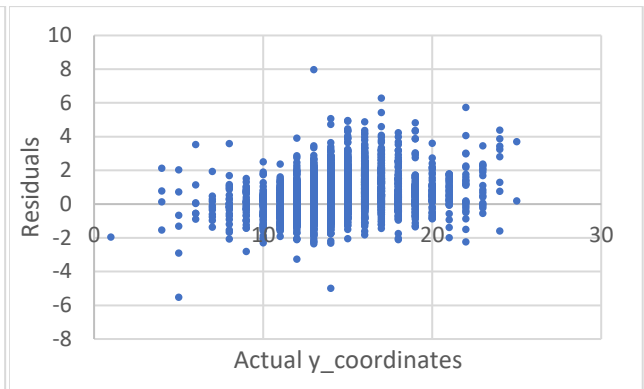
ในการทดลองนี้ผู้วิจัยได้แสดงตัวอย่างผลลัพธ์จากการระบุจุดศูนย์กลางรูม่านตาด้วยตัวแบบ SimpleCNN และ DoubleConcatCNNv3 ดังภาพที่ 5.9, ภาพที่ 5.10 และภาพที่ 5.11 โดยที่จุดสีเขียวคือค่าจริงและจุดสีน้ำเงินคือค่าที่ทำนายได้



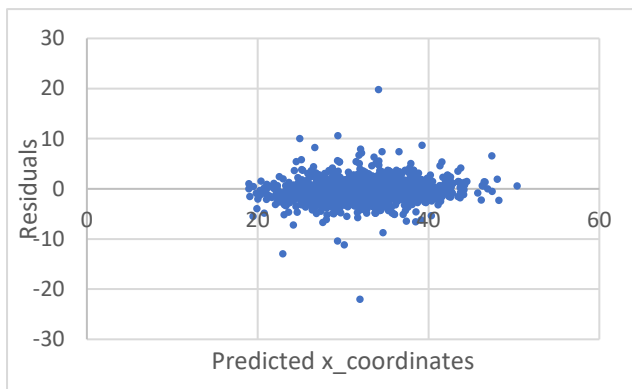
ภาพที่ 5.6 แผนภาพเศษเหลือของตัวแบบ DoubleConcatCNNv1 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



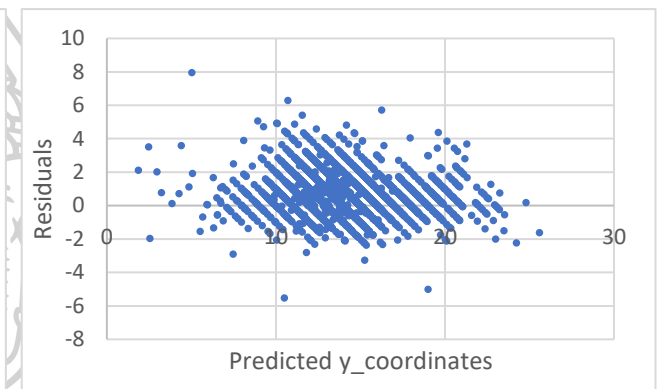
ก.)



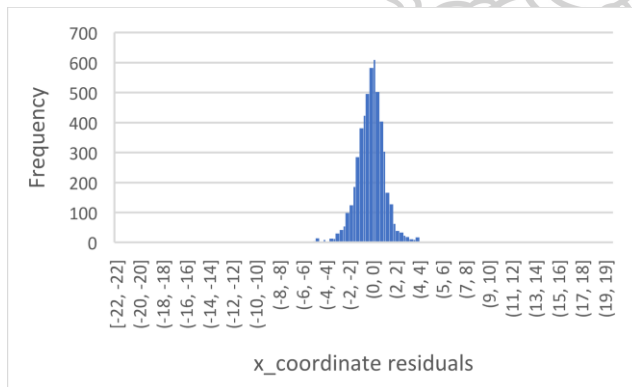
ข.)



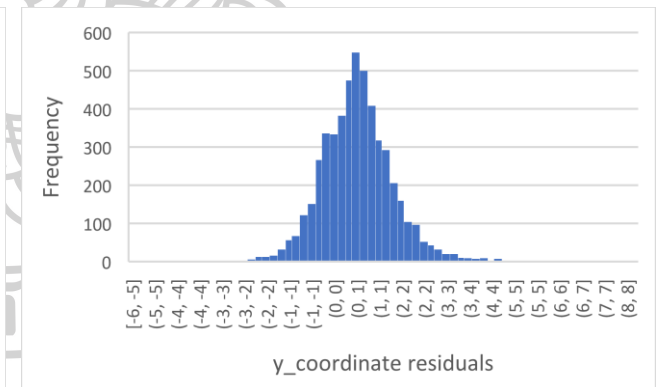
ค.)



ง.)

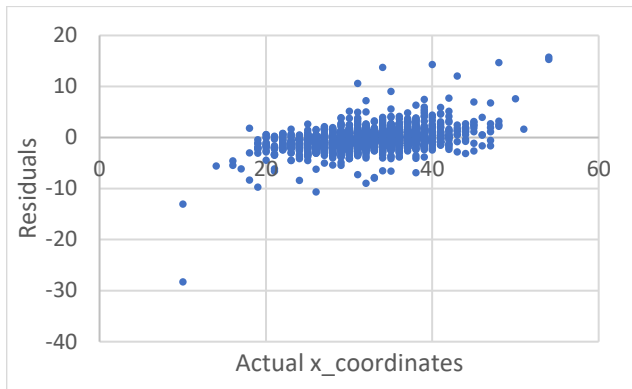


จ.)

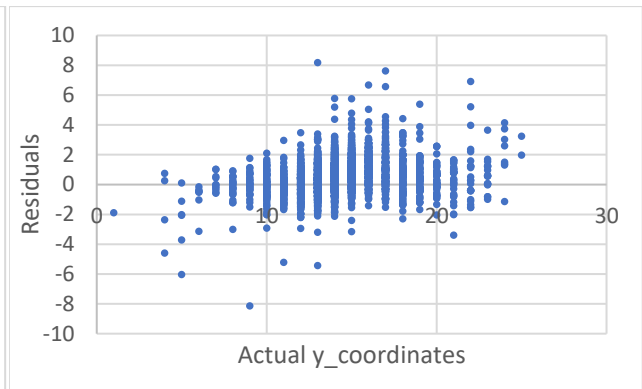


ฉ.)

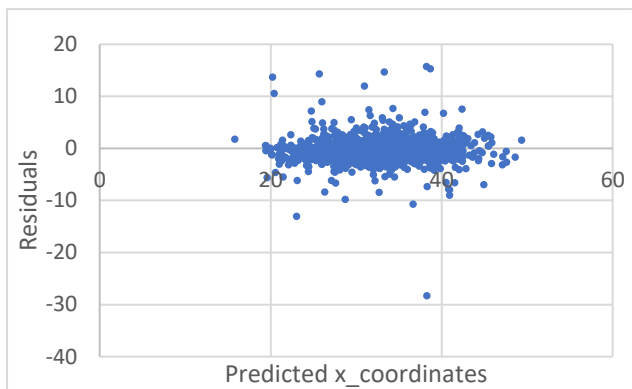
ภาพที่ 5.7 แผนภาพเศษเหลือของตัวแบบ DoubleConcatCNNv2 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



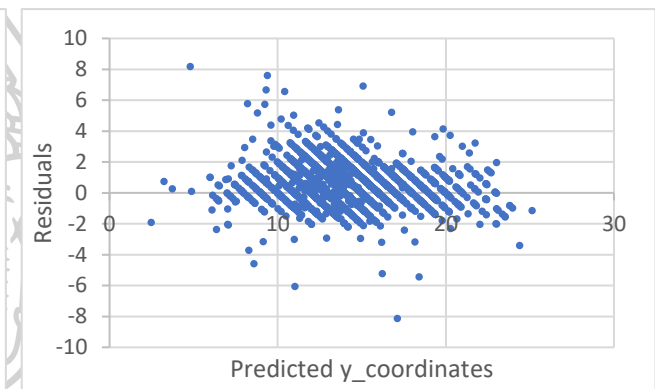
ก.)



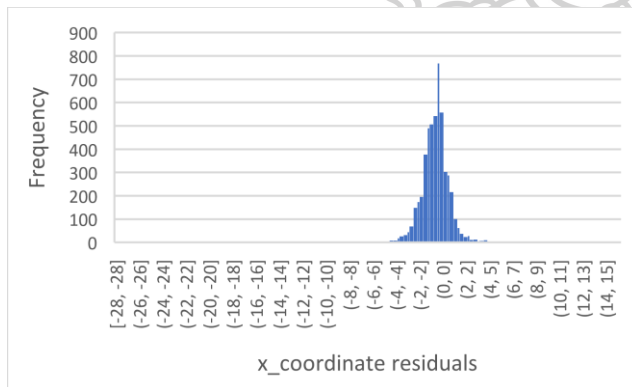
ข.)



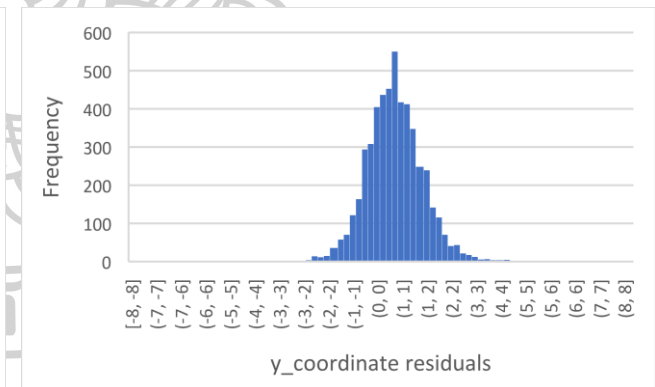
ค.)



ง.)

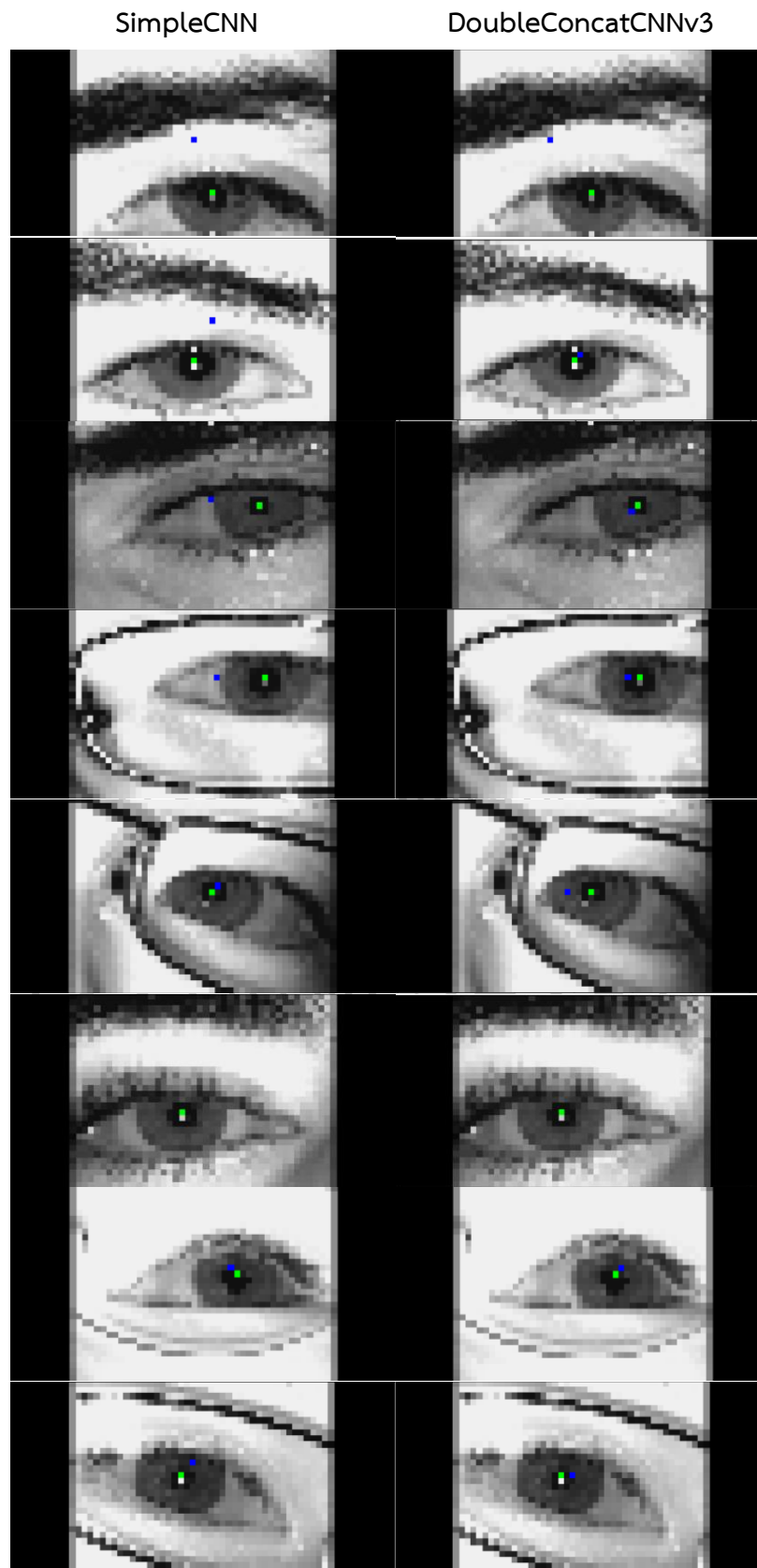


จ.)

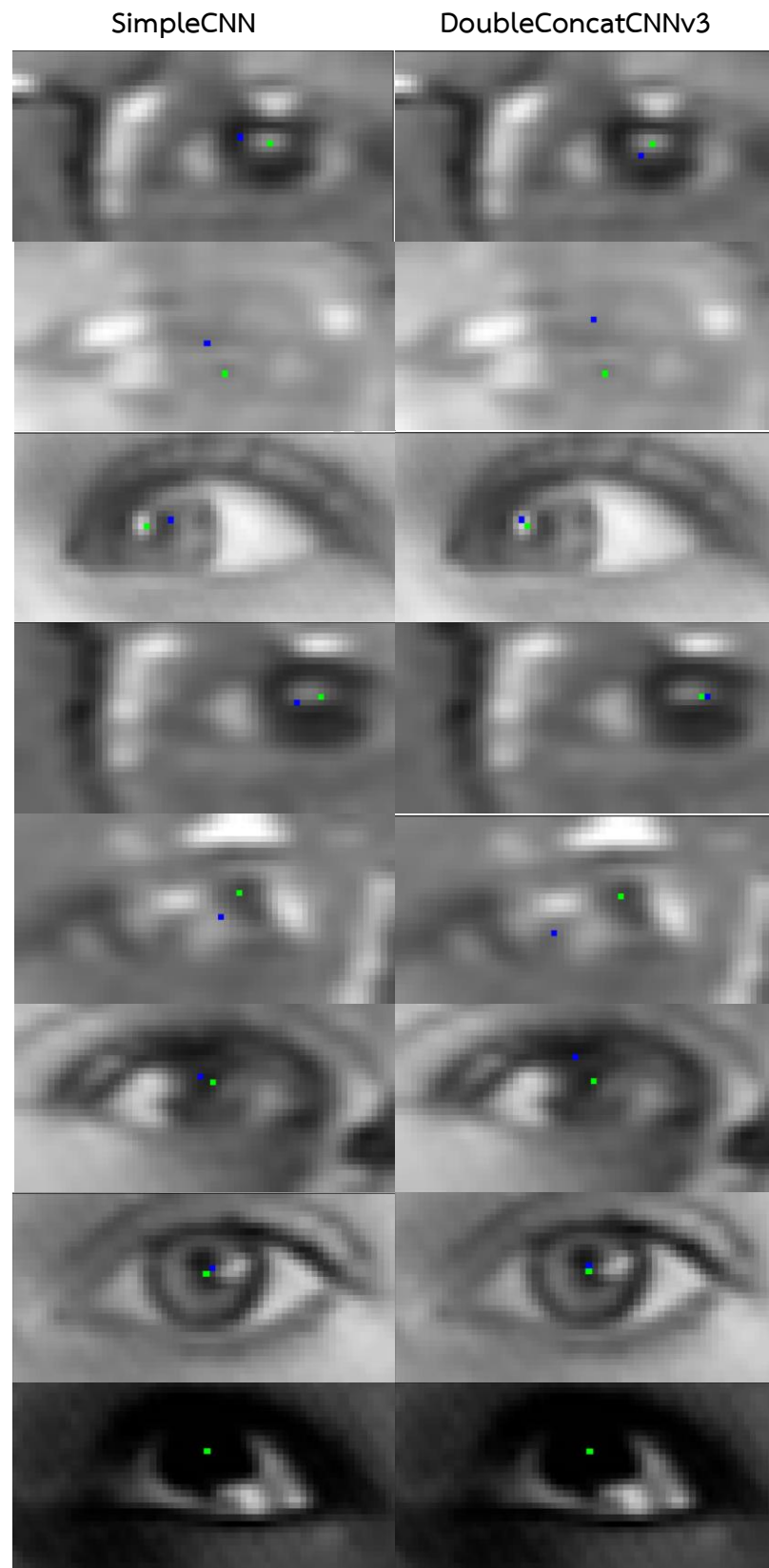


ฉ.)

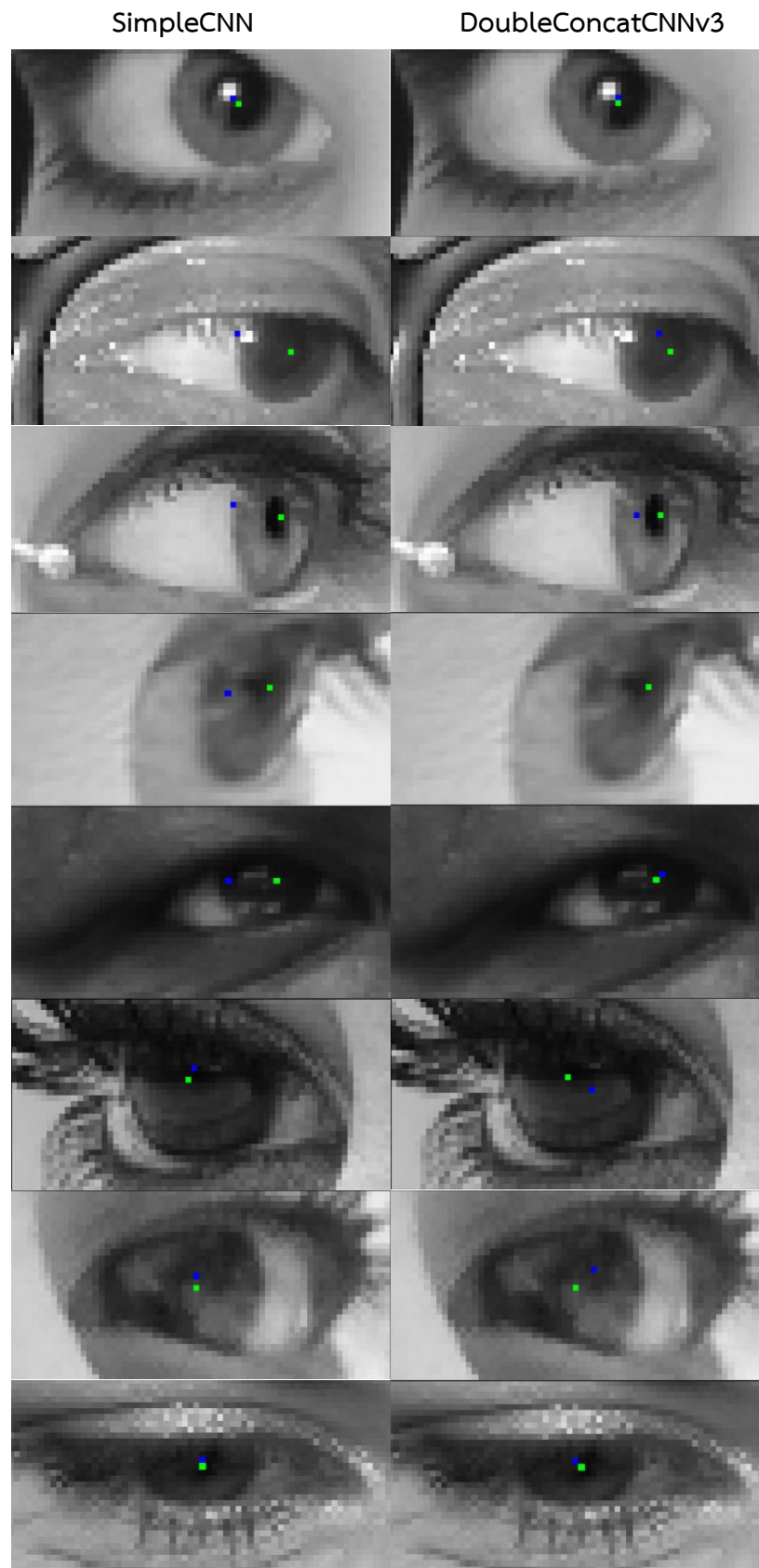
ภาพที่ 5.8 แผนภาพเศษเหลือของตัวแบบ DoubleConcatCNNv3 ก.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน x ข.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดผลเฉลี่ยในแนวแกน y ค.) แผนภาพการกระจายระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน x ง.) แผนภาพการกระจายระหว่างระหว่างเศษเหลือกับพิกัดที่ทำนายได้ในแนวแกน y จ.) ฮิสโทแกรมเศษเหลือในแนวแกน x ฉ.) ฮิสโทแกรมเศษเหลือในแนวแกน y



ภาพที่ 5.9 ตัวอย่างผลลัพธ์จาก SimpleCNN และ DoubleConcatCNNv3 บนฐานข้อมูล MMU2



ภาพที่ 5.10 ตัวอย่างผลลัพธ์จาก SimpleCNN และ DoubleConcatCNNv3 บนฐานข้อมูล BioID



ภาพที่ 5.11 ตัวอย่างผลลัพธ์จาก SimpleCNN และ DoubleConcatCNNv3 บนฐานข้อมูล FFHQ

จากการเปรียบเทียบแสดงให้เห็นว่า DoubleConcatCNNv3 ยังคงมีความแปรปรวนอยู่กับภาพที่มีองค์ประกอบอื่นรวมอยู่ในภาพ เช่น คิ้ว โดยเฉพาะคิ้วที่มีลักษณะหนา ความแปรปรวนจะเห็นได้ชัดบนฐานข้อมูล MMU2 เนื่องจากเป็นภาพที่เตรียมข้อมูลภาพผ่านการสเกลภาพและเติมพื้นที่หลังเท่านั้น ไม่ได้ผ่านการครอปตัด จึงทำให้อาจมีองค์ประกอบอื่น ๆ ปะปนอยู่

สำหรับภาพประเภทที่มีผลเฉลยอยู่ห่างจากจุดศูนย์กลางภาพไปมาก ๆ เห็นได้ชัดว่า DoubleConcatCNNv3 ทำได้ดีกว่า ส่วน SimpleCNN จะมีแนวโน้มที่จะทำนายผลลัพธ์ออกมาอยู่บริเวณจุดศูนย์กลางแม้ว่าจุดศูนย์กลางรูปร่างจะอยู่บริเวณอื่น และสำหรับภาพตาที่มีลักษณะสีม่านตาอ่อน DoubleConcatCNNv3 สามารถทำได้ดีเช่นเดียวกับภาพที่มีลักษณะสีม่านตาเข้ม

การทดลองที่ 4 การวิเคราะห์การถดถอยโดยการใช้ชั้นคอนโวลูชันแบบขยาย

จากปัญหาความแปรปรวนเมื่อมีองค์ประกอบอื่นของใบหน้าอยู่ในภาพข้อมูลเข้า การทดลองนี้ต้องการทดสอบเพิ่มเติมว่าการขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันเพื่อรวบรวมบริบทที่ใหญ่ขึ้นจะสามารถเพิ่มประสิทธิภาพการวิเคราะห์การถดถอยของจุดศูนย์กลางรูปร่างตาได้หรือไม่ จึงได้ทดสอบโดยนำตัวแบบจากการทดลองก่อนหน้านี้ได้แก่ DoubleConcatCNNv2 และ DoubleConcatCNNv3 มาปรับแต่งเพิ่มเติม การขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันผู้วิจัยใช้การกำหนดค่า dilation rate เท่ากับ 2 หรือเรียกอีกอย่างว่าเป็นการใช้ Atrous Convolution และได้ปรับแต่งจากคอนโวลูชันชั้นแรกไปยังชั้นถัด ๆ ไป ตามตารางที่ 5.6 ผลการทดลองแสดงดังตารางที่ 5.7

ตารางที่ 5.6 ตารางแสดงโครงสร้างของการนำคุณลักษณะจากชั้นต่าง ๆ มาประมวลผลร่วมกันในชั้นการเชื่อมต่อแบบสมบูรณ์

	ตัวแบบพื้นฐาน	ปรับแต่งด้วยคอนโวลูชันแบบขยายใน		
		คอนโวลูชัน ชั้นที่ 1	คอนโวลูชัน ชั้นที่ 2	คอนโวลูชัน ชั้นที่ 3
AtrousDoubleConcatCNNv2.1 (ADCv2.1)	DoubleConcatCNNv2	✓		✓
AtrousDoubleConcatCNNv2.2 (ADCv2.2)	DoubleConcatCNNv2		✓	✓
AtrousDoubleConcatCNNv2.3 (ADCv2.3)	DoubleConcatCNNv2	✓	✓	✓
AtrousDoubleConcatCNNv3.1 (ADCv3.1)	DoubleConcatCNNv3	✓		✓
AtrousDoubleConcatCNNv3.2 (ADCv3.2)	DoubleConcatCNNv3		✓	✓
AtrousDoubleConcatCNNv3.3 (ADCv3.3)	DoubleConcatCNNv3	✓	✓	✓



ตารางที่ 5.7 ตารางเปรียบเทียบประสิทธิภาพของ AtrousDoubleConcatCNN และ DoubleConcatCNN

Model name	MMU2			BioID			FFHQ			SD
	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	% Success	Avg Euc Err	CV	
DoubleConcatCNNv2	89.43	1.44	12.68	93.82	1.24	3.07	82.39	1.73	5.05	0.25
DoubleConcatCNNv3	90.37	1.39	5.80	94.00	1.25	5.02	83.68	1.73	4.64	0.24
ADCv2.1	87.66	1.50	7.15	91.54	1.37	5.09	82.17	1.81	3.11	0.22
ADCv2.2	86.95	1.57	6.13	93.42	1.25	3.72	82.12	1.80	3.77	0.27
ADCv2.3	80.50	1.86	11.43	92.91	1.30	2.63	82.65	1.80	2.69	0.31
ADCv3.1	89.13	1.43	5.52	92.42	1.32	6.92	81.93	1.75	3.22	0.22
ADCv3.2	86.79	1.57	9.20	93.75	1.84	5.63	81.04	1.25	3.72	0.29
ADCv3.3	85.33	1.66	4.55	93.72	1.26	2.14	81.21	1.85	2.65	0.30

จากตารางพบว่า AtrousDoubleConcatCNNv3.2 ให้ค่าคลาดเคลื่อนเฉลี่ยที่ต่ำที่สุดบนฐานข้อมูล FFHQ แต่ในภาพรวมตัวแบบ DoubleConcatCNNv3 ยังมีประสิทธิภาพที่ดีกว่าทั้งในด้านความแม่นยำและความคงทน สำหรับตัวแบบที่มีการขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันตัวอื่น ๆ พบว่ายังมีประสิทธิภาพที่ต่ำกว่าตัวแบบพื้นฐาน DoubleConcatCNNv2 และ DoubleConcatCNNv3

จากการสังเกตค่าสัมประสิทธิ์ความแปรผันของตัวแบบพบว่าการขยายมุมมองในชั้นคอนโวลูชันมีแนวโน้มที่จะทำให้ความคงเส้นคงวาของตัวแบบเพิ่มมากขึ้นบนฐานข้อมูล FFHQ และ MMU2 ในขณะที่เดียวกันก็มีแนวโน้มที่จะทำให้ความคงเส้นคงวาของตัวแบบต่ำลงเช่นกันบนฐานข้อมูล BioID นอกจากนี้สังเกตได้ว่าบนฐานข้อมูล MMU2 ยังมีการขยายมุมมองการมองเห็นในชั้นคอนโวลูชันที่มากขึ้น ยิ่งส่งผลให้ความแม่นยำต่ำลง

การทดลองที่ 5 การเปรียบเทียบประสิทธิภาพของตัวแบบกับงานวิจัยอื่น ๆ ที่เกี่ยวข้องบนฐานข้อมูล BioID

การทดลองนี้มีจุดประสงค์เพื่อเปรียบเทียบตัวแบบที่ดีที่สุดที่ผู้วิจัยนำเสนอกับงานวิจัยที่เกี่ยวข้อง โดยจากผลการทดลองข้างต้นผู้วิจัยได้นำตัวแบบ DoubleConcatCNNv3 มาทดสอบบนฐานข้อมูล BioID อีกครั้งและใช้มาตรวัด Wec และ Bec เพื่อเปรียบเทียบกับงานวิจัยก่อนหน้า โดยจะเน้นไปที่มาตรวัด Wec เป็นหลัก ผลการทดลองแสดงให้เห็นตามตารางที่ 5.8

ตารางที่ 5.8 ตารางแสดงการเปรียบเทียบประสิทธิภาพของ DoubleConcatCNNv3 กับงานวิจัยที่เกี่ยวข้อง

methods	Wec \leq 0.05	Wec \leq 0.1
George, A., 2016 [18]	85.8	94.3
Chen, S., 2015 [19]	88.7	95.2
Khan, W., 2020 [13]	94.5	100
Ahmed, N., 2021 [20]	91.7	97.9
Gou, C., 2019 [21]	92.3	99.1
DoubleConcatCNNv3	97.59	99.74

จากผลการทดลองพบว่าวิธีการที่นำเสนอในภาพรวมให้ประสิทธิภาพที่ดีกว่าวิธีการในงานวิจัยที่เกี่ยวข้องก่อนหน้าโดยร้อยละ 97.59 มีค่า Wec \leq 0.05 และร้อยละ 99.74 มีค่า Wec \leq 0.1 มีเพียงงานวิจัยของคุณ Khan และคณะซึ่งมีประสิทธิภาพที่ดีกว่าเมื่อใช้เกณฑ์ค่า Wec \leq 0.1 โดยให้อัตราความสำเร็จถึงร้อยละ 100 นอกจากนี้ตัวแบบที่ผู้วิจัยนำเสนอร้อยละ 99.82 มีค่า Bec \leq 0.05 และร้อยละ 100 มีค่า Bec \leq 0.1

การทดลองที่ 6 การทดลองฝึกตัวแบบการเรียนรู้กับรูปภาพที่ไม่มีพื้นหลังสีดำ

เนื่องจากภาพจากฐานข้อมูล MPII ที่ใช้เป็นข้อมูลฝึก เป็นภาพที่ถูกครอบตัดมาเฉพาะพื้นที่บริเวณดวงตา ในขั้นตอนการเตรียมข้อมูลจึงทำให้เกิดพื้นหลังที่มีสีดำขึ้น ซึ่งผู้วิจัยคิดว่าภาพที่มีพื้นหลังสีดำอาจส่งผลให้การเรียนรู้มีประสิทธิภาพที่ลดลง ในการทดลองนี้จึงมีจุดประสงค์เพื่อทดสอบว่าการฝึกฝนตัวแบบด้วยภาพที่มีพื้นหลังสีดำส่งผลต่อการวิเคราะห์การถดถอยอย่างมีนัยสำคัญ ในการทดลองผู้วิจัยได้ใช้ข้อมูลฝึกเป็นฐานข้อมูล BioID แทน และเปรียบเทียบประสิทธิภาพเมื่อเตรียมข้อมูลฝึกแบบมีพื้นหลังสีดำ กับเมื่อเตรียมข้อมูลฝึกแบบไม่มีพื้นหลังสีดำ

โดยในการเตรียมข้อมูลแบบไม่มีพื้นหลังสีดำผู้วิจัยใช้การครอบตัดบริเวณดวงตาจากภาพใบหน้าซึ่งเป็นภาพต้นฉบับบนฐานข้อมูล BioID ซึ่งในการทดลองนี้ผู้วิจัยได้ควบคุมตำแหน่งของผลเฉลยสำหรับชุดข้อมูลฝึกให้มีลักษณะการกระจายเหมือนกับชุดข้อมูลฝึก MPII ที่ใช้ในการทดลองก่อนหน้าด้วย ผลการทดลองแสดงดังตารางที่ 5.9

ตารางที่ 5.9 ประสิทธิภาพของตัวแบบ SimpleCNN, DoubleConcatCNNv2 และ DoubleConcatCNNv3 บนชุดข้อมูลฝึก BioID ที่มีวิธีการเตรียมข้อมูลต่างกัน (แบบมีพื้นหลังสีดำ และแบบไม่มีพื้นหลังสีดำ)

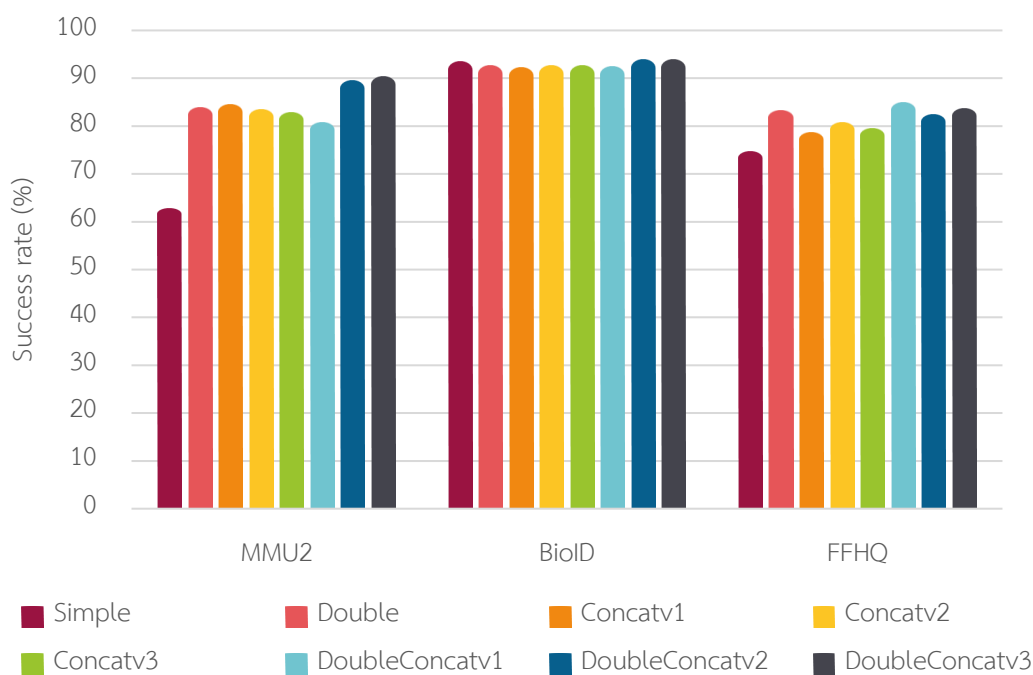
	ประสิทธิภาพเมื่อฝึกตัวแบบ ด้วยฐานข้อมูลBioID แบบมี พื้นฐานสีดำ		ประสิทธิภาพเมื่อฝึกตัวแบบ ด้วยฐานข้อมูล BioID แบบ ไม่มีพื้นหลังสีดำ	
	MMU2	FFHQ	MMU2	FFHQ
SimpleCNN	2.80	1.74	2.79	1.58
DoubleConcatCNNv2	1.96	1.54	2.42	1.58
DoubleConcatCNNv3	1.92	1.58	4.89	1.46

จากผลการทดลองพบว่าประสิทธิภาพของตัวแบบที่ได้รับการฝึกฝนจากภาพที่มีพื้นหลังสีดำ ให้ผลที่ดีกว่าบนฐานข้อมูล MMU2 และประสิทธิภาพของตัวแบบที่ได้รับการฝึกฝนจากภาพที่ไม่มีพื้นหลังสีดำให้ผลที่ดีกว่าเมื่อทดสอบบนฐานข้อมูล FFHQ

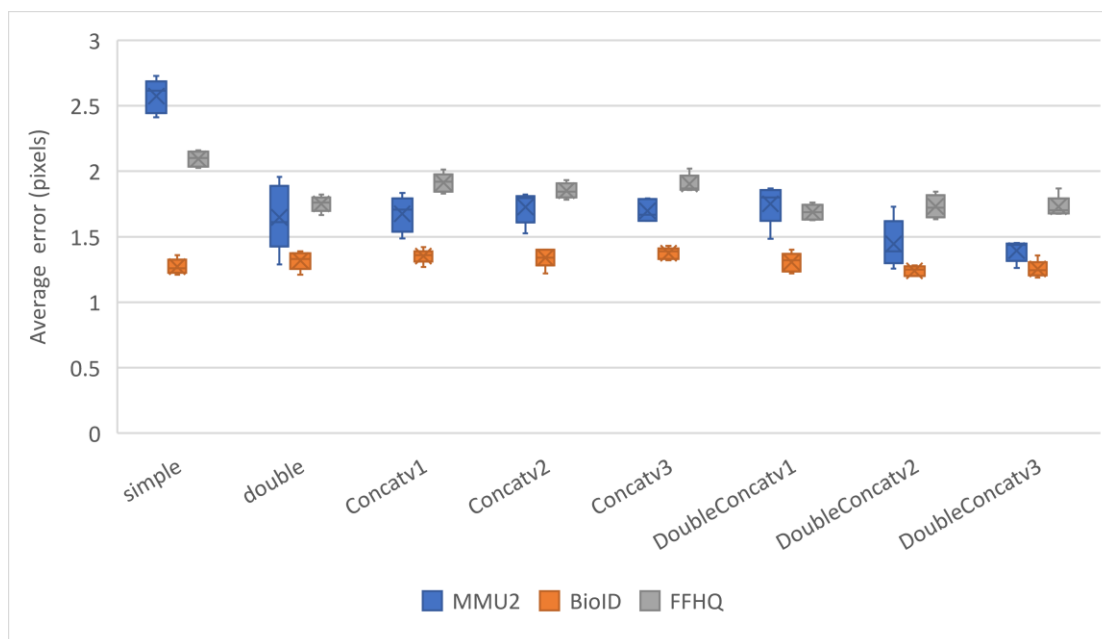
บทที่ 6

สรุปผลการทดลองและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอการปรับปรุงความแม่นยำและความคงทนของการวิเคราะห์การถดถอยของรูปร่างตาเพื่อให้ตอบโจทย์ของกลุ่มประชากรที่มีลักษณะม่านตาสีเข้มแบบที่ไม่ต้องการพึ่งพากล้องอินฟราเรดและการจัดแสงเพิ่มเติม โดยได้เลือกใช้ฐานข้อมูลในการทดสอบที่มีความท้าทายและหลากหลาย สำหรับวิธีการระบุจุดศูนย์กลางรูปร่างตาที่นำเสนอผู้วิจัยใช้เทคนิคการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันพร้อมกับการเพิ่มระดับการแปลงไม่เชิงเส้นในตัวแบบการเรียนรู้ ซึ่งผู้วิจัยได้ทำการทดสอบประสิทธิภาพของแต่ละเทคนิคที่ใช้ผ่านการวิเคราะห์แบบตัดออก โดยได้สรุปผลการทดลองในการทดลองที่ 1 – 3 เป็นแผนภาพเพื่อแสดงผลลัพธ์ของข้อมูลในรูปแบบที่เข้าใจง่ายขึ้นดังภาพที่ 6.1 และภาพที่ 6.2



ภาพที่ 6.1 แผนภูมิแท่งแสดงอัตราความสำเร็จของตัวแบบที่นำเสนอในการทดลองที่ 1-4 เมื่อทดสอบบนฐานข้อมูล MMU2, BioID และ FFHQ



ภาพที่ 6.2 แผนภาพกล่องแสดงการกระจายของระยะทางคลาดเคลื่อนเฉลี่ยเมื่อทดลองซ้ำ 5 รอบ

จากผลการทดลองที่ได้พบว่าการเพิ่มระดับการแปลงไม่เชิงเส้นจากการเพิ่มชั้นคอนโวลูชันในแต่ละชั้นทำให้ความคลาดเคลื่อนลดลงอย่างมีนัยสำคัญ รวมถึงมีแนวโน้มทำให้ตัวแบบมีความคงทนเพิ่มขึ้นด้วย และสำหรับเทคนิคการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันให้ผลลัพธ์ไปในแนวทางเดียวกัน คือสามารถทำให้ตัวแบบมีความแม่นยำและความคงทนเพิ่มขึ้น และเมื่อนำทั้งสองเทคนิคมาใช้ร่วมกันยิ่งทำให้ประสิทธิภาพของตัวแบบเพิ่มมากขึ้น

โดยการนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันจากทั้งสามชั้นคอนโวลูชันใน DoubleConcatCNN3 มีประสิทธิภาพดีที่สุด ซึ่งอาจสรุปได้ว่าการเพิ่มระดับการแปลงไม่เชิงเส้นเข้าไปในโครงสร้างทำให้ตัวแบบมีความสามารถในการแปลงไม่เชิงเส้นที่มากพอ โดยตัวแบบที่ได้อาจลักษณะที่เป็นเหมือน quadratic function มากขึ้น หรืออาจมีลักษณะเป็น cubic function ซึ่งทำให้ตัวแบบมีการอนุมานค่าที่ดีมาก

นอกจากนี้การนำคุณลักษณะจากชั้นคอนโวลูชันหลายชั้นมาประมวลผลร่วมกันอาจทำให้สามารถคำนึงถึงความละเอียดในระดับต่าง ๆ ที่เหมาะกับภาพนั้น ๆ ได้ดีกว่า เพราะจากข้อมูลเข้าจะพบว่าระดับการซูม (zoom level) จะไม่เท่ากันในแต่ละภาพ โดยบางที่ตาอยู่ไกลบางที่ตาอยู่ใกล้ ทำให้มีความละเอียดที่หลากหลาย การที่จะเลือกคุณลักษณะในชั้นสุดท้ายไปประมวลผลในชั้นการเชื่อมต่อแบบสมบูรณ์เพียงอย่างเดียวแล้วหวังว่าจะตอบโจทย์ได้ตลอดอาจจะเป็นไปได้ยาก โดยเฉพาะ

การทำนายที่แม่นยำในระดับพิกเซล การเพิ่มคุณลักษณะนี้เข้าไปจึงสามารถแก้ปัญหาเรื่องของการรวม ทำให้ level scale factor ของการพิจารณาทำได้ดีกว่าเดิม จึงทำให้ DoubleConcatv3 มีประสิทธิภาพที่ดีกว่าในสถานะทั่วไป

ทั้งนี้ยังมีเหตุผลบางประการที่ทำให้ไม่สามารถเห็นการเปลี่ยนแปลงที่เกิดขึ้นกับฐานข้อมูล BioID ได้มากนัก ซึ่งผู้วิจัยคาดว่าอาจเป็นเพราะบนฐานข้อมูล BioID มีลักษณะการกระจายของผลเฉลยที่เกาะกลุ่มอยู่บริเวณจุดศูนย์กลางของภาพมากเกินไป ทำให้ไม่ว่าจะใช้ตัวแบบใด ๆ จึงสามารถให้ผลที่ดีได้ไม่ต่างกัน โดยเฉพาะกับตัวแบบพื้นฐานที่มีความเอนเอียงไปกับภาพที่มีจุดศูนย์กลางรูมาวนตาอยู่บริเวณจุดศูนย์กลางภาพด้วย แต่สำหรับฐานข้อมูล MMU2 และฐานข้อมูล FFHQ นั้นแตกต่างออกไปจากการกระจายของตำแหน่งจุดศูนย์กลางรูมาวนตาในภาพที่สูงกว่า จึงสามารถสังเกตเห็นการเปลี่ยนแปลงได้ชัดเจนกว่า

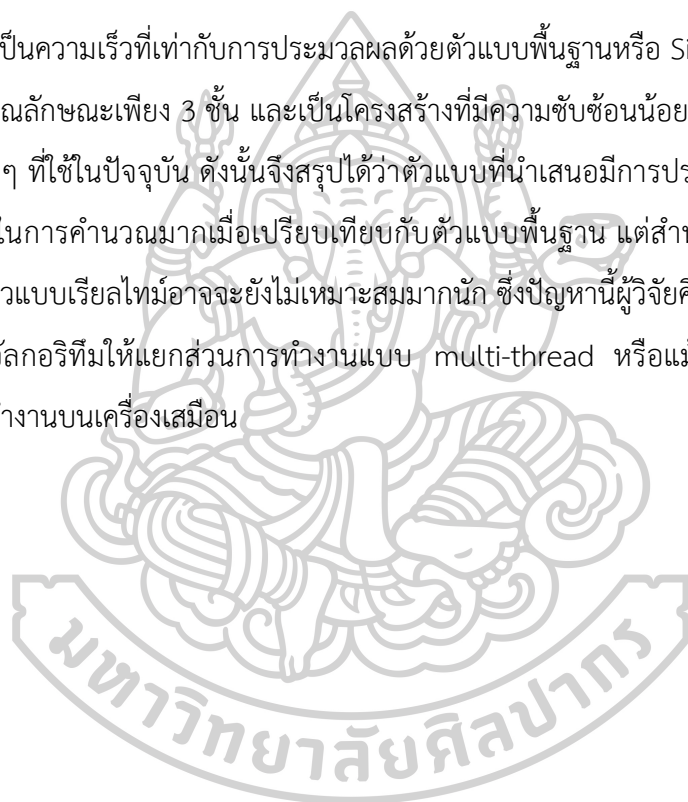
นอกจากนี้ผู้วิจัยได้ทดสอบการขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันด้วย ซึ่งผลการทดลองพบว่า การขยายขอบเขตการมองเห็นในคอนโวลูชันเหมาะสมสำหรับบางฐานข้อมูลและไม่เหมาะสมกับบางฐานข้อมูล ซึ่งสังเกตว่าการขยายขอบเขตการมองเห็นในคอนโวลูชันไม่เหมาะสมกับภาพที่เสกของดวงตาในภาพที่มีขนาดเล็ก และอาจสรุปได้ว่าการขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันไม่เหมาะสมกับปัญหาการระบุจุดศูนย์กลางรูมาวนตาในงานวิจัยนี้ แม้บนบางฐานข้อมูลการขยายขอบเขตการมองเห็นในชั้นคอนโวลูชันจะช่วยให้ตัวแบบมีความคงเส้นคงวามากขึ้น แต่เนื่องจากภาพข้อมูลเข้ามีขนาดเล็ก การนำวิธีนี้มาใช้จึงไม่สามารถช่วยในเรื่องความแม่นยำได้มากนัก

การทดลองต่อมาได้เปรียบเทียบตัวแบบที่ดีที่สุดที่ผู้วิจัยนำเสนอ DoubleConCatCNNv3 กับงานวิจัยอื่น ๆ ที่เกี่ยวข้อง บนฐานข้อมูล BioID ซึ่งเป็นฐานข้อมูลที่มีความท้าทายและเป็นที่ยอมรับในการนำมาใช้ทดสอบ พบว่า DoubleConCatCNNv3 ให้ผลลัพธ์ที่น่าประทับใจ มีความแม่นยำกว่างานก่อนหน้าจากการใช้มาตรวัดที่เป็นมาตรฐาน Wec นอกจากนี้จะเห็นว่าในการทดสอบได้มีภาพที่มีลักษณะมาวนตาสีอ่อนปะปนอยู่ด้วยซึ่งตัวแบบที่ผู้วิจัยนำเสนอก็สามารถทำนายจุดศูนย์กลางรูมาวนตาได้แม่นยำเช่นกัน ทำให้สามารถสรุปได้ว่า แม้ในงานวิจัยนี้จะต้องรับมือกับปัญหาที่ซับซ้อนกว่าในแง่ที่ภาพไม่ได้มีการจัดแสงเพื่อให้ได้ภาพตาที่มีความสมบูรณ์แบบขององค์ประกอบต่าง ๆ ของตา แต่การนำไปใช้กับภาพตาประเภทดังกล่าวก็ยังสามารถให้ความแม่นยำที่ดีได้

สำหรับการทดลองสุดท้ายซึ่งทดสอบประสิทธิภาพของตัวแบบ เมื่อฝึกด้วยรูปภาพที่มีพื้นหลังสีดำเปรียบเทียบกับฝึกด้วยรูปภาพที่ไม่มีพื้นหลังสีดำ พบว่าตัวแบบจะให้ประสิทธิภาพดีเมื่อภาพ

ที่นำมาทดสอบเป็นไปในลักษณะเดียวกัน ซึ่งสำหรับปัญหาการวิเคราะห์หาจุดศูนย์กลางรูม่านตา การนำไปใช้งานจริงมีความเป็นไปได้สูงที่ภาพข้อมูลเข้าจะเป็นภาพใบหน้า ทำให้ต้องถูกนำไปครอบตัดอีกทีเพื่อตัดองค์ประกอบอื่น ๆ บนใบหน้าออกไปให้เหลือเพียงแค่บริเวณดวงตา ดังนั้นในอนาคตจึงควรใช้การฝึกฝนตัวแบบโดยใช้ข้อมูลฝึกที่ไม่มีภาพพื้นหลังเป็นสีดำ

สำหรับในเรื่องความเร็วที่ใช้ในการประมวลผลพบว่าตัวแบบ DoubleConcatCNNv3 ที่มีการเพิ่มเทคนิคที่นำเสนอสองเทคนิค ใช้เวลาเท่ากับ 0.03 วินาทีต่อภาพ และเมื่อรวมเวลาการเตรียมข้อมูลด้วยวิธีครอบตัดภาพจากภาพใบหน้าพบว่าใช้เวลาเท่ากับ 0.05 วินาทีต่อภาพ ซึ่งจากการทดลองพบว่าเป็นความเร็วที่เท่ากับการประมวลผลด้วยตัวแบบพื้นฐานหรือ SimpleCNN ที่มีความลึกของการแยกคุณลักษณะเพียง 3 ชั้น และเป็นโครงสร้างที่มีความซับซ้อนน้อยมากเมื่อเปรียบเทียบกับโครงสร้างอื่น ๆ ที่ใช้ในปัจจุบัน ดังนั้นจึงสรุปได้ว่าตัวแบบที่นำเสนอมีการประมวลผลที่เร็วและไม่ได้เสียค่าใช้จ่ายในการคำนวณมากเมื่อเปรียบเทียบกับตัวแบบพื้นฐาน แต่สำหรับการนำไปใช้งานกับภาพเคลื่อนไหวแบบเรียลไทม์อาจจะยังไม่เหมาะสมมากนัก ซึ่งปัญหานี้ผู้วิจัยคิดว่าสามารถปรับปรุงได้โดยการปรับอัลกอริทึมให้แยกส่วนการทำงานแบบ multi-thread หรือแม้แต่การปรับการทำงานบน GPU ให้ทำงานบนเครื่องเสมือน



รายการอ้างอิง

1. Nitzan, M., A. Romem, and R. Koppel, *Pulse oximetry: Fundamentals and technology update*. 2014, Medical Devices: Evidence and Research. p. 231-239.
2. Josef Flammer, K.K., Rosa M. Bruno, Agostino Viridis, Andreas J. Flammer, and Stefano Taddei, *The eye and the heart*. Notes and Queries, 1861. **s2-XII(297)**: p. 190-190.
3. HH Patel, M.P. *Anatomy of the Human Eye*. 2019-12-05]; Available from: <https://www.news-medical.net/health/Anatomy-of-the-Human-Eye.aspx>.
4. *Vision - Structures of the Eye*. 2019-12-05]; Available from: https://www.wikipremed.com/mcat_course_psychology.php?module=2§ion=3.
5. Javadi, A.H., et al., *Set: A pupil detection method using sinusoidal approximation*. Frontiers in Neuroengineering, 2015. **8**(APR).
6. Poh, M.-Z., D.J. McDuff, and R.W. Picard, *Non-contact, automated cardiac pulse measurements using video imaging and blind source separation*. Optics Express, 2010. **18**(10): p. 10762.
7. Unakafov, A.M., et al., *Using imaging photoplethysmography for heart rate estimation in non-human primates*. PLoS ONE, 2018. **13**(8).
8. Fuhl, W., et al., *PupilNet: Convolutional Neural Networks for Robust Pupil Detection*. ArXiv, 2016.
9. Chinsatit, W. and T. Saitoh, *CNN-Based Pupil Center Detection for Wearable Gaze Estimation System*. Applied Computational Intelligence and Soft Computing, 2017.
10. Yiu, Y.H., et al., *DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning*. Journal of Neuroscience Methods, 2019. **324**.
11. Ronneberger, O., P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in

- Bioinformatics), 2015. **9351**: p. 234-241.
12. Świrski, L. and N.A. Dodgson, *A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting*. Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI), 2013.
 13. Khan, W., et al., *Pupil Localisation and Eye Centre Estimation Using Machine Learning and Computer Vision*. Sensors, 2020. **20**(13): p. 3785.
 14. Jesorsky, O., K.J. Kirchberg, and R.W. Frischholz, *Robust face detection using the Hausdorff distance*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2001. **2091**: p. 90-95.
 15. Borza, D., R. Itu, and R. Danescu, *In the eye of the deceiver: Analyzing eye movements as a cue to deception*. Journal of Imaging, 2018. **4**(10): p. 120.
 16. Valenti, R. and T. Gevers. *Accurate eye center location and tracking using isophote curvature*. in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008. IEEE.
 17. Holzman, A.E. *What do your eyes say about your health?* 2014; Available from: <https://www.drholzman.com/what-do-your-eyes-say-140878/>.
 18. George, A. and A. Routray, *Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images*. IET Computer Vision, 2016. **10**(7): p. 660-669.
 19. Chen, S. and C. Liu, *Eye detection using discriminatory Haar features and a new efficient SVM*. Image and Vision Computing, 2015. **33**: p. 68-77.
 20. Ahmed, N.Y., *Real-time accurate eye center localization for low-resolution grayscale images*. Journal of Real-Time Image Processing, 2021. **18**(1): p. 193-220.
 21. Gou, C., et al., *Cascade learning from adversarial synthetic images for accurate pupil detection*. Pattern Recognition, 2019. **88**: p. 584-594.



ประวัติผู้เขียน

ชื่อ-สกุล	พิชชาภา เลิศศิริวรรณ
วัน เดือน ปี เกิด	16 พฤษภาคม 2539
สถานที่เกิด	อำเภอเมืองนครปฐม จังหวัดนครปฐม
วุฒิการศึกษา	สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ มหาวิทยาลัยศิลปากร พ.ศ.2560
ที่อยู่ปัจจุบัน	61 ซ.พุทธรักษา2 ถ.พุทธรักษา ตำบลสนามจันทร์ อำเภอเมืองนครปฐม จังหวัดนครปฐม 73000

