

การเรียนรู้เชิงลึกโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับ
ระบบรู้จำภาพใบหน้าบุคคล

DEEP LEARNING USING CONVOLUTIONAL NEURAL NETWORKS
FOR FACE RECOGNITION



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาเพื่อขอรับปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2559

KMITL-2016-EN-M-010-097

การเรียนรู้เชิงลึกโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับ
ระบบรู้จำภาพใบหน้าบุคคล

DEEP LEARNING USING CONVOLUTIONAL NEURAL NETWORKS
FOR FACE RECOGNITION



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโทรคมนาคม
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2559
KMITL-2016-EN-M-010-097

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DEEP LEARNING USING CONVOLUTIONAL NEURAL NETWORKS
FOR FACE RECOGNITION



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN TELECOMMUNICATION ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2016
KMITL-2016-EN-M-010-097

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2016

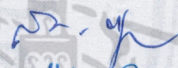


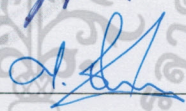
FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

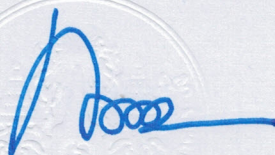
หัวข้อวิทยานิพนธ์ การเรียนรู้เชิงลึกโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับระบบรู้จำ
ภาพใบหน้าบุคคล
Thesis Title Deep Learning using Convolutional Neural Networks for Face Recognition
นักศึกษา นางสาวสุทิพา เผือกผ่อง
รหัสประจำตัว 57601138
ปริญญา วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา วิศวกรรมโทรคมนาคม
อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.ดร.ศรวัฒน์ ชิวปรีชา
หมายเลขวิทยานิพนธ์ KMITL-2016-EN-M-010-097

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
ผศ.ดร.สิรภาพ	ผู้ประกาย	
ผศ.ดร.มนตรี	ค่าเงิน	
รศ.ดร.มนตรี	ศิริปรัชญานันท์	
รศ.ดร.ยุทธพงษ์	รังสรรค์เสรี	
ผศ.ดร.ศรวัฒน์	ชิวปรีชา	

วัน / เดือน / ปี ที่สอบ วันพฤหัสบดีที่ 14 กรกฎาคม พ.ศ. 2559 เวลา 13.00-15.00 น.
สถานที่สอบ ณ อาคารเฉลิมพระเกียรติ ห้อง HM-303

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิศวกรรมศาสตร์ รับรองแล้ว



(รองศาสตราจารย์ ดร. คมสัน มาลีสี)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น
ฉบับตี คณะวิศวกรรมศาสตร์
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างวันที่ 14 กรกฎาคม พ.ศ. 2559 นำไปใช้

หัวข้อวิทยานิพนธ์	การเรียนรู้เชิงลึกโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน สำหรับระบบรู้จำภาพใบหน้าบุคคล
นักศึกษา	นางสาวสุทิพา เผือกผ่อง
รหัสประจำตัว	57601138
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมโทรคมนาคม
พ.ศ.	2559
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.ศรวัฒน์ ชิวปรีชา

บทคัดย่อ

ปัจจุบันโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) เป็นระบบรู้จำภาพใบหน้าที่มีประสิทธิภาพใกล้เคียงกับมนุษย์ อย่างไรก็ตามวิธีการนี้ต้องใช้ข้อมูลจำนวนมากในการสอน ซึ่งเป็นเรื่องยากหากนำไปใช้งานกับการประยุกต์ทางด้านความปลอดภัย วิทยานิพนธ์นี้นำเสนอระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันด้วยวิธีการถ่ายโอนการเรียนรู้ (TL-CNN) สำหรับฐานข้อมูลขนาดเล็ก ในขั้นตอนแรก CNN จะถูกสอนด้วยฐานข้อมูลขนาดใหญ่ จากนั้นชั้นจำแนกจะถูกแทนที่ด้วยชั้นใหม่และถูกสอนด้วยภาพใหม่ซึ่งเป็นฐานข้อมูลขนาดเล็ก วิธีการนี้ถูกเรียกว่า "การถ่ายโอนการเรียนรู้" ในวิทยานิพนธ์นี้ ฐานข้อมูลขนาดเล็ก 2 ฐานข้อมูล ซึ่งประกอบไปด้วยฐานข้อมูล 9 คน และฐานข้อมูล 109 คน ถูกนำมาใช้ศึกษาประสิทธิภาพของระบบรู้จำใบหน้าที่แตกต่างกัน นอกจากนี้ อิทธิพลของอัตราการเรียนรู้ต่อประสิทธิภาพการรู้จำภาพใบหน้าบุคคลถูกนำมาศึกษาเช่นกัน ผลการทดลองเบื้องต้นพบว่าความถูกต้องของ CNN อยู่ในระดับที่ยอมรับได้ (ร้อยละ 80) นอกจากนี้ ผลการทดลองพบว่า TL-CNN ที่มีอัตราการเรียนรู้เท่ากับ 0.001 แสดงความถูกต้องที่ดีที่สุดเมื่อเปรียบเทียบกับ PCA สำหรับการดึงลักษณะ และ k-Nearest Neighbors (kNN) และ Support Vector Machine (SVM) สำหรับการจำแนก

Thesis	Deep Learning using Convolutional Neural Networks for Face Recognition
Student	Miss Suthipa Pauekpong
Student ID.	57601138
Degree	Master of Engineering
Program	Telecommunication Engineering
Year	2016
Thesis Advisor	Asst. Prof. Dr. Sorawat Chivapreecha

ABSTRACT

Nowadays, convolution neural networks (CNN) performs a closed-to-human face recognition system. However, a very large amount of image resource is required for training, which is too difficult to apply for security application. This thesis presents the face recognition system using transfer learning convolutional neural networks (TL-CNN) for small dataset. In the first step, CNN is trained with a large dataset of face images. Then, the classification layers are replaced with the new layers and are retrained with new subject images, which is a small dataset. This method is called "transfer learning". In this thesis, 2 private small datasets, including 9 persons and 109 persons datasets, is adopted to study the performance of the different face recognition systems. Furthermore, the effect of learning rate on face recognition performance is also investigated. The preliminary results show that the accuracy of CNN is in an acceptable level (80%). Further, the results show that TL-CNN with learning rate of 0.001 exhibits the best accuracy as compared to PCA for feature extraction and k-nearest neighbor (kNN) and support vector machine (SVM) for classification.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต่อ!อ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากข้าพเจ้าได้รับความกรุณา ความเมตตา จาก ผศ.ดร.ศรวัฒน์ ชิวปรีชา ที่คอยให้ความช่วยเหลือ ดูแลเอาใจใส่ แนะนำ ชี้แนะข้อบกพร่องและ ร่วมแก้ไขปัญหา ติดตามความก้าวหน้าของงานวิจัย รวมทั้งฝึกฝนให้ข้าพเจ้ามีทักษะทางด้าน การคิด การอ่าน การเขียนและการนำเสนอผลงานวิชาการ ซึ่งเป็นประโยชน์อย่างมากในการพัฒนาตนเอง

ข้าพเจ้าขอขอบคุณ ดร.อิทธิพันธ์ เมธเศรษฐ์ นักวิจัยจากเนคเทค ที่คอยให้ความช่วยเหลือ ช่วยแก้ปัญหาที่เกิดขึ้น ตลอดจนพี่ๆในแลป Image Technology Laboratory (IMG) จากศูนย์ เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ที่ร่วมกันแก้ปัญหาที่เกิดขึ้น อีกทั้งดูแลข้าพเจ้า เป็นอย่างดีตลอดการทำวิจัยที่เนคเทค

ข้าพเจ้าขอขอบคุณ ดร.พนชิต กิตติปัญญางาม บริษัท สำนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) ที่คอยดูแล ให้คำปรึกษาเกี่ยวกับงานวิจัยมาตลอดมา

ข้าพเจ้าขอขอบคุณนายบริวัฒน์ พิจารณา และนายเอกราช บำรุงไทยชัยชาญที่คอย ช่วยเหลือรวมทั้งให้คำปรึกษายามเมื่อมีปัญหา และร่วมแก้ปัญหาไปด้วยกัน

ขอขอบพระคุณทุนสนับสนุนการศึกษาจาก “โครงการพัฒนาศักยภาพบุคคลากรเพื่อการวิจัย และพัฒนา สำหรับภาคอุตสาหกรรม (NUI-RC) สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.)” สัญญาทุนเลขที่ NUI-RC01-57-022 สำหรับการศึกษาในระดับมหาบัณฑิต

ขอขอบพระคุณทุนสนับสนุนการทำวิจัยบางส่วนจาก “โครงการ True lab” เลขที่สัญญา TUC/CM2/2014-0978/5

ขอขอบคุณเพื่อนร่วมงานในห้องปฏิบัติการภาควิชาโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่านที่ได้ให้คำปรึกษา แนะนำ และ ถ่ายทอดเทคโนโลยีความรู้ให้แก่ข้าพเจ้า

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาและครอบครัวของข้าพเจ้าที่ได้ให้ชีวิตและ โอกาสทางการศึกษา คอยเป็นกำลังใจและให้ความห่วงใยเสมอมา ตลอดจนคุณครูและอาจารย์ทุกท่านที่กรุณาประสิทธิ์ประสาทวิชาความรู้อันเป็นประโยชน์แก่ข้าพเจ้า

คุณค่าและประโยชน์อันพึงมาจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบแด่ผู้มีพระคุณทุกท่าน

สุทิพา เผือกผ่อง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 ขอบเขตของการศึกษา.....	2
1.4 รายละเอียดของวิทยานิพนธ์.....	2
บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง.....	4
2.1 การทบทวนงานวิจัยที่เกี่ยวข้อง (Literature Reviews).....	4
2.2 หลักการของระบบรู้จำภาพใบหน้าบุคคล.....	7
2.2.1 การดึงลักษณะเด่นของข้อมูล.....	7
2.2.2 การจำแนกข้อมูลภาพใบหน้า.....	10
2.3 การเรียนรู้เชิงลึก.....	15
2.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN).....	16
บทที่ 3 โครงข่ายประสาทเทียมแบบคอนโวลูชันกับระบบรู้จำภาพใบหน้าบุคคลที่น่าเสนอ.....	31
3.1 ระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน.....	31
3.1.1 การออกแบบระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน.....	31
3.1.2 การทำงานของ CNN.....	35
3.1.3 CNN แบบถ่ายโอนการเรียนรู้ (Transfer learning CNN).....	36
3.2 ฐานข้อมูลที่ใช้การทดลอง.....	38
3.2.1 ฐานข้อมูลภาพใบหน้าขนาดเล็ก 9 คน.....	39
3.2.2 ฐานข้อมูลภาพใบหน้าขนาดเล็ก 109 คน.....	40
3.2.3 ฐานข้อมูลขนาดใหญ่ (Celebrity face).....	41

สารบัญ (ต่อ)

	หน้า
3.3 ขั้นตอนการทดลอง.....	43
บทที่ 4 การทดลองและการวิเคราะห์ผลการทดลอง.....	44
4.1 การจัดการฐานข้อมูลในการทดลอง.....	44
4.2 การรู้จำภาพใบหน้าบุคคลโดยใช้ CNN.....	45
4.2.1 การทดลองการใช้ CNN กับฐานข้อมูลขนาดเล็ก (9 และ 109 คน) โดยตรง.....	46
4.2.2 การทดลองการใช้ CNN แบบถ่ายโอนการเรียนรู้ (TL-CNN).....	48
4.3 การรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป.....	51
4.3.1 การรู้จำภาพใบหน้าบุคคลโดยใช้ PCA และ kNN.....	52
4.3.2 การรู้จำภาพใบหน้าบุคคลโดยใช้ PCA และ SVM.....	54
4.4 การเปรียบเทียบและวิเคราะห์ผลการทดลอง.....	63
4.5 การประยุกต์ใช้งานระบบรู้จำภาพใบหน้าบุคคล.....	65
4.5.1 ระบบเข้าสู่ประตูโดยใช้ภาพใบหน้า.....	65
4.5.2 แอปพลิเคชันยืนยันตัวบุคคลบนโทรศัพท์สมาร์ทโฟน.....	68
บทที่ 5 สรุปงานวิจัยและข้อเสนอแนะ.....	71
5.1 สรุปเนื้อหาวิทยานิพนธ์.....	71
5.2 ข้อเสนอแนะและแนวทางการพัฒนาต่อ.....	72
เอกสารอ้างอิง.....	73
ภาคผนวก ก.....	76
ภาคผนวก ข.....	82
ประวัติผู้เขียน.....	104

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ข้อมูลแต่ละบุคคล.....	12
3.1 รายละเอียดแต่ละชั้นของโครงสร้าง.....	34
3.2 ปริมาณข้อมูลในฐานข้อมูล.....	43
4.1 ฐานข้อมูลในการทดสอบ.....	44
4.2 ผลเปรียบเทียบค่าความถูกต้องของการทดลอง CNN และ TL-CNN.....	51
4.3 ผลการทดลองโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ kNN ในการจำแนก.....	54
4.4 ผลการทดลองการใช้ SVM ในการจำแนกและมีเส้นแบ่งแบบเป็นเส้นตรง.....	57
4.5 ค่าความถูกต้องในการทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนกเส้นแบ่งแบบเป็น RBF.....	62
4.6 ค่าความถูกต้องในการทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก.....	62
4.7 เปรียบเทียบค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลของวิธีการทั่วไป และการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน.....	63
4.8 เวลาที่ใช้ในการดำเนินการระบบรู้จำภาพใบหน้าบุคคล.....	64

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 การรู้จำภาพใบหน้าบุคคลโดยทั่วไป.....	7
2.2 แปลงภาพให้เป็นเวกเตอร์.....	8
2.3 การทำงานของ kNN.....	11
2.4 การสร้างเส้นแบ่งของ SVM.....	13
2.5 ภาพรวมของการเรียนรู้เชิงลึก.....	15
2.6 การรู้จำภาพใบหน้าบุคคลโดยใช้ CNN.....	16
2.7 โครงสร้างโดยทั่วไปของ CNN.....	17
2.8 ผลของ Sigmoid function	18
2.9 ผลของ tanh function.....	18
2.10 ผลของ ReLU function.....	19
2.11 ตัวอย่างการทำงานคอนโวลูชัน.....	20
2.12 การทำพูลลิ่งด้วยค่าสูงสุด.....	21
2.13 การทำพูลลิ่งด้วยค่าเฉลี่ย.....	21
2.14 การทำพูลลิ่งด้วย $L_2 - norm$	22
2.15 โครงสร้าง LeNet-1.....	23
2.16 การแพร่แบบย้อนกลับด้วยแผนภาพคุณลักษณะเด่นจากการทำคอนโวลูชัน.....	29
2.17 การแพร่แบบย้อนกลับด้วยแผนภาพคุณลักษณะที่ได้จากการทำพูลลิ่ง.....	30
3.1 โครงสร้างของระบบ CNN.....	32
3.2 ค่าความถูกต้องสำหรับจำนวนเซลล์ประสาทที่แตกต่างกัน.....	34
3.3 โครงสร้าง CNN สำหรับฐานข้อมูล 9 คน.....	35
3.4 โครงสร้าง CNN สำหรับฐานข้อมูล 109 คน.....	35
3.5 CNN แบบถ่ายโอนการเรียนรู้.....	36
3.6 โครงสร้าง CNN กับฐานข้อมูลขนาดใหญ่.....	37
3.7 โครงสร้าง TL-CNN สำหรับฐานข้อมูล 9 คน.....	38
3.8 โครงสร้าง TL-CNN สำหรับฐานข้อมูล 109 คน.....	38
3.9 ตัวอย่างภาพใบหน้าที่ใช้ในการทดลอง.....	39
3.10 ตัวอย่างภาพถ่ายจากฐานข้อมูล 9 คน.....	39
3.11 ตัวอย่างภาพฐานข้อมูลภาพใบหน้า 9 คน.....	40
3.12 ตัวอย่างภาพในฐานข้อมูลภาพใบหน้า 109 คน.....	40
3.13 ภาพใบหน้าในอิริยาบถต่างๆ.....	41

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.14 ภาพตัวอย่างของ Bill Gate.....	42
3.15 ภาพตัวอย่างของ Taylor Swift	42
4.1 ภาพรวมโครงสร้าง CNN ที่ใช้ในการทดลอง.....	45
4.2 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูล 9 คน.....	46
4.3 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูล 109 คน.....	47
4.4 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูลขนาดใหญ่.....	48
4.5 ค่าความถูกต้องของการทดลอง TL-CNN กับฐานข้อมูล 9 คน.....	49
4.6 ค่าความถูกต้องของการทดลอง TL-CNN กับฐานข้อมูล 109 คน.....	50
4.7 ระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป.....	52
4.8 การทำงานของ PCA ในการดึงลักษณะเด่น และใช้ kNN ในการจำแนก.....	52
4.9 การทำงานของ PCA ในการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่และ kNN จำแนก.....	53
4.10 การทดลองการใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก.....	55
4.11 การทดลองการใช้ PCA ในการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่และใช้ SVM จำแนก.....	55
4.12 ค่าความถูกต้องของการทดลอง PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนกเมื่อ เส้นแบ่งแบบ RBF กับฐานข้อมูล 9 คน.....	58
4.13 ค่าความถูกต้องของการทดลอง PCA (Celeb)+ SVM เมื่อเส้นแบ่ง RBF กับฐานข้อมูล 9 คน	59
4.14 ค่าความถูกต้องของการทดลอง PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนกเมื่อ เส้นแบ่งแบบ RBF กับฐานข้อมูล 109 คน.....	60
4.15 ค่าความถูกต้องของการทดลอง PCA (Celeb) + SVM เมื่อเส้นแบ่ง RBF กับฐานข้อมูล 109 คน.....	61
4.16 ภาพรวมการติดตั้งระบบเข้าประตูโดยใช้ภาพใบหน้าบุคคล.....	66
4.17 ภาพรวมการทำงานของระบบผ่านเข้าประตูโดยใช้ภาพใบหน้าบุคคล.....	67
4.18 ภาพรวมของระบบแอปพลิเคชันยืนยันตัวตนบุคคล.....	68
4.19 แผนผังการทำงานของแอปพลิเคชัน.....	69
4.20 ขั้นตอนการทำงานของแอปพลิเคชัน.....	69

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ระบบรู้จำภาพใบหน้า ถือว่าเป็นหนึ่งในระบบที่ใช้ในการพิสูจน์ยืนยันตัวตนบุคคลโดยใช้คุณลักษณะจำเพาะทางสรีระ (Biometric) โดยระบบรู้จำภาพใบหน้าจะทำงานโดยการเปรียบเทียบภาพใบหน้าจากภาพถ่ายดิจิทัลหรือภาพจากกล้องวิดีโอของบุคคลที่เราสนใจกับฐานข้อมูลภาพใบหน้าที่มีอยู่ และเมื่อเปรียบเทียบเสร็จก็จะแสดงผลภาพใบหน้าที่อยู่ในฐานข้อมูลที่มีใบหน้าตรงกันกับภาพที่นำมาเปรียบเทียบออกมา ระบบรู้จำภาพใบหน้านั้นได้ถูกพัฒนาอย่างต่อเนื่องเป็นเวลามากกว่าสิบปีมาแล้ว เนื่องจากเป็นระบบที่ได้รับความสนใจมาจากนักวิชาการหลายสาขาวิชาจึงทำให้ระบบรู้จำภาพใบหน้าที่มีผู้คนสนใจศึกษาและพัฒนากันอย่างมากมาย จนทำให้มีการพัฒนาอัลกอริทึมในการทำงานของระบบออกมามากมายหลายรูปแบบแตกต่างกันไป ซึ่งการพัฒนาอัลกอริทึมก็แตกต่างกันไปตามยุคสมัยด้วย อันเนื่องมาจากปัจจัยด้านองค์ความรู้และเทคโนโลยีของอุปกรณ์ต่างๆ ที่พัฒนาขึ้นให้มีความเหมาะสมที่จะนำมาใช้ในระบบ จึงจำเป็นต้องออกแบบอัลกอริทึมใหม่ให้เหมาะสมกับอุปกรณ์ใหม่ๆ ด้วย

ในปัจจุบันระบบรู้จำภาพใบหน้าได้มีการพัฒนาไปอย่างมาก ทำให้ระบบรู้จำภาพใบหน้าที่มีความน่าเชื่อถือเพิ่มมากขึ้น จนมีการนำระบบรู้จำภาพใบหน้ามาใช้ประโยชน์กันอย่างแพร่หลายในต่างประเทศ ซึ่งหลายๆ ประเทศได้มีการนำระบบรู้จำภาพใบหน้ามาติดตั้งในสนามบิน ใช้สำหรับยืนยันตัวบุคคลร่วมกันกับการทำงานของเจ้าหน้าที่ เพื่อป้องกันคนร้ายหลบหนีเข้าหรือออกนอกประเทศ หรือการใช้ในการยืนยันตัวเองเพื่อเข้าระบบหรือผ่านเข้าประตู แทนการใช้ลายนิ้วมือหรือคีย์การ์ด ซึ่งการนำไปประยุกต์ใช้งานในลักษณะนี้ จำนวนบุคคลที่อยู่ในฐานข้อมูลมีไม่มาก ดังนั้นจึงจำเป็นต้องหาวิธีการจำแนกข้อมูลที่เหมาะสมกับจำนวนคนในฐานข้อมูลประมาณ 10 – 100 คน เพื่อให้การนำไปประยุกต์ใช้กับการยืนยันตัวบุคคลก่อนเข้าห้องมีประสิทธิภาพสูงสุด ซึ่งวิธีที่สนใจจะนำมาประยุกต์ใช้คือ โครงข่ายประสาทเทียมแบบคอนโวลูชัน

1.2 ความมุ่งหมายและวัตถุประสงค์ของวิทยานิพนธ์

1. เพื่อศึกษาการทำงานของระบบรู้จำภาพใบหน้าบุคคลโดยใช้ Convolutional Neural Networks (CNN)
2. เพื่อศึกษาการประยุกต์ใช้ CNN ในการรู้จำภาพใบหน้าบุคคลกับฐานข้อมูลขนาดเล็ก
3. เพื่อออกแบบระบบรู้จำภาพใบหน้าบุคคลโดยใช้ CNN กับฐานข้อมูลขนาดเล็ก
4. เพื่อเพิ่มประสิทธิภาพการทำงานของระบบรู้จำภาพใบหน้าบุคคลโดยใช้ CNN ด้วยวิธีการถ่ายโอนการเรียนรู้กับฐานข้อมูลขนาดเล็ก
5. เปรียบเทียบประสิทธิภาพในการรู้จำภาพใบหน้าบุคคลของ CNN กับวิธีการอื่น

1.3 ขอบเขตของวิทยานิพนธ์

1. ใช้ระบบรู้จำภาพใบหน้าบุคคลด้วยวิธี CNN กับฐานข้อมูล 9 คน และ 109 คน
2. ใช้ CNN แบบถ่ายโอนการเรียนรู้เพื่อเพิ่มประสิทธิภาพระบบรู้จำภาพใบหน้าบุคคลกับฐานข้อมูล 9 คน และ 109 คน
3. เปรียบเทียบค่าความถูกต้องของระบบรู้จำภาพใบหน้าบุคคลด้วยวิธี CNN กับวิธีการอื่น (ใช้ PCA ในการดึงลักษณะเด่น และใช้ kNN หรือ SVM ในการจำแนก)
4. เปรียบเทียบเวลาที่ใช้ในการดำเนินงานของระบบรู้จำภาพใบหน้าบุคคลด้วยวิธี CNN กับวิธีการอื่น (ใช้ PCA ในการดึงลักษณะเด่น และใช้ kNN หรือ SVM ในการจำแนก)

1.4 รายละเอียดของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 5 ส่วนดังต่อไปนี้

บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหาซึ่งจะกล่าวถึงการทำงานของระบบรู้จำภาพใบหน้าบุคคล ตัวอย่างการนำระบบรู้จำภาพใบหน้าบุคคลไปประยุกต์ใช้งาน และมูลเหตุจูงใจในการทำวิทยานิพนธ์นี้ และจะกล่าวถึงจุดประสงค์ของวิทยานิพนธ์ ขอบเขตของวิทยานิพนธ์ สุดท้ายจะอธิบายรายละเอียดในแต่ละส่วนของเล่มวิทยานิพนธ์

บทที่ 2 กล่าวถึงงานวิจัยก่อนหน้าที่เกี่ยวกับระบบรู้จำภาพใบหน้าบุคคลตั้งแต่อดีต จนถึงปัจจุบัน และหลักการที่ใช้ระบบรู้จำภาพใบหน้าบุคคลโดยทั่วไปซึ่งได้แก่การใช้ PCA ในการดึงลักษณะเด่น การใช้ kNN และ SVM ในการจำแนก รวมทั้งหลักการที่ใช้ในระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 กล่าวถึงโครงข่ายประสาทเทียมแบบคอนโวลูชันกับระบบรู้จำภาพใบหน้าบุคคลที่นำเสนอ ซึ่งประกอบด้วยรายละเอียดการออกแบบโครงสร้างของ CNN การใช้งาน CNN และ CNN แบบถ่ายโอนการเรียนรู้ ฐานข้อมูลที่ใช้ในวิทยานิพนธ์นี้ รวมไปถึงคอมพิวเตอร์ในการดำเนินงาน และโปรแกรมที่ใช้

บทที่ 4 การออกแบบการทดลอง การวิเคราะห์ผลการทดลอง ซึ่งประกอบด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันด้วยวิธีการถ่ายโอนการเรียนรู้ และวิธีการรู้จำภาพใบหน้าบุคคลโดยทั่วไปซึ่งประกอบด้วยการใช้ PCA ในการดึงลักษณะเด่นและการใช้ kNN กับ SVM ในการจำแนก การวิเคราะห์ผลการทดลองในแต่ละส่วนอีกทั้งการเปรียบเทียบผลการทดลองระหว่างวิธีการทั่วไป กับโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยจะพิจารณาทั้งค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล และเวลาที่ใช้ในการทำงาน สุดท้ายจะกล่าวถึงการประยุกต์ใช้งานระบบรู้จำภาพใบหน้าบุคคล

บทที่ 5 บทสรุปของวิทยานิพนธ์และข้อเสนอแนะ รวมทั้งแนวทางการนำไปพัฒนาต่อ

ทฤษฎีหลักการและงานวิจัยที่เกี่ยวข้อง

ในบทที่ 2 นี้จะกล่าวถึงการพัฒนากระบวนจำภาพใบหน้าบุคคลเริ่มตั้งแต่ตั้งแต่การพัฒนา จนกระทั่งเป็นวิธีการที่นิยมใช้ในปัจจุบัน ถูกกล่าวถึงในหัวข้อการทบทวนงานวิจัยที่เกี่ยวข้อง จากนั้น จะอธิบายหลักการที่ใช้ในกระบวนจำภาพใบหน้าบุคคลโดยจะแบ่งเป็นกระบวนจำภาพใบหน้าบุคคล โดยทั่วไป และกระบวนจำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งกระบวน จำภาพใบหน้าบุคคลโดยทั่วไปนั้นจะแบ่งออกเป็นการดึงลักษณะเด่นโดยใช้การวิเคราะห์ องค์ประกอบหลัก (Principal Component Analysis: PCA) และการจำแนกภาพใบหน้าบุคคลโดยใช้ การพิจารณาค่าใกล้เคียง k ตัว (k-Nearest Neighbor: kNN) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

2.1 การทบทวนงานวิจัยที่เกี่ยวข้อง (Literature Reviews)

กระบวนจำภาพใบหน้าบุคคลนั้นถูกพัฒนาอย่างต่อเนื่องเป็นเวลานาน ปัจจุบันมีการนำองค์ ความรู้ใหม่ๆ มาพัฒนาระบบจำภาพใบหน้าบุคคลให้มีความแม่นยำเพิ่มมากขึ้น ซึ่งมีงานวิจัยก่อน หน้าที่ศึกษาเกี่ยวกับกระบวนจำภาพใบหน้าบุคคลดังจะกล่าวต่อไป

งานวิจัยเกี่ยวกับกระบวนจำภาพใบหน้าบุคคลเริ่มต้นครั้งแรกเมื่อปี 1888 โดย Galton และ คณะ [1] ได้ออกแบบระบบเพื่อเก็บลักษณะภาพใบหน้าแต่ละบุคคลภายใต้ชื่อ Profile ID เพื่อทำการ ระบุตัวบุคคลจากภาพใบหน้า ต่อมาในช่วงปี 1960 Bledsoe [2] ได้ออกแบบกระบวนจำภาพใบหน้า บุคคลแบบกึ่งอัตโนมัติขึ้นมา โดยระบบนี้จะใช้คนในการบอกตำแหน่งต่างๆบนภาพใบหน้า จากนั้นจะ ใช้คอมพิวเตอร์ในการคำนวณอัตราส่วนบนภาพใบหน้าเพื่อทำการจำแนกบุคคล ในปี 1971 Goldstein [3] ได้เก็บลักษณะบนภาพใบหน้า 21 ลักษณะ จากนั้นใช้การจำแนกรูปแบบ (Pattern Classification) ในการระบุตัวบุคคล จากนั้นในปี 1973 Fischler และ Elschanger [4] ได้นำเสนอ ระบบจำภาพใบหน้าบุคคลแบบอัตโนมัติขึ้นมา โดยการใช้ local template matching และ global measure ในการหาลักษณะเด่นโดยไม่ต้องใช้คนในการบอกตำแหน่ง และในปีเดียวกัน Kenade [5] ได้ดึงลักษณะเด่น 16 อย่างบนภาพใบหน้าออกมา และมีความถูกต้องในการรู้จำภาพใบหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

45 – 75% จะเห็นว่างานวิจัยในช่วงเริ่มต้นของระบบรู้จำภาพใบหน้าบุคคลนั้นจะมุ่งเน้นไปยังการหาวิธีในการดึงลักษณะเด่นของภาพใบหน้า

ต่อมาในปี 1987 Sirovich [6] ได้นำวิธีการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) มาใช้ในการดึงลักษณะเด่นจากภาพใบหน้าเป็นครั้งแรก หลังจากนั้นไม่นานในปี 1991 Turk และ Pentland [7] ได้นำ PCA มาใช้ในการดึงลักษณะเด่นของภาพใบหน้าเช่นกัน แต่ในครั้งนี้นักวิจัยเรียกวิธีการที่นำ PCA มาใช้กับภาพใบหน้าว่า Eigenface หรือภาพใบหน้าไอเกน นอกจากนี้พวกเขายังใช้ Nearest Neighbor ในการจำแนกภาพใบหน้าบุคคลอีกด้วย เรียกว่างานวิจัยนี้เป็นจุดเริ่มต้นความนิยมในการศึกษาและวิจัยเกี่ยวกับระบบรู้จำภาพใบหน้าบุคคล ซึ่งวิธีการนี้ได้รับความนิยมเป็นอย่างมากจนถึงปัจจุบัน ซึ่งในปีเดียวกัน Golomb [8] ได้นำโครงข่ายประสาทเทียม (Neural Network) มาใช้ในการจำแนกเพศจากภาพใบหน้า โดยระบบนี้สามารถจำแนกได้และมีค่าความผิดพลาดเพียง 8% เท่านั้น ต่อมาในปี 1997 Lawrence [9] ใช้ Self-organizing map (SOM) ในการดึงลักษณะเด่นจากนั้นใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันในการจำแนกภาพใบหน้าบุคคล ในขณะที่โครงข่ายประสาทเทียมแบบคอนโวลูชันยังไม่ได้รับความนิยมมากเนื่องจากมีความซับซ้อนในการทำงาน อีกทั้งงานวิจัยนี้ยังใช้ประโยชน์จากโครงข่ายประสาทเทียมแบบคอนโวลูชันไม่เต็มที่ และในปี 1999 Philips [10] ได้นำซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) มาใช้ในการรู้จำภาพใบหน้าบุคคลโดยใช้ PCA ในการดึงลักษณะเด่นของภาพใบหน้าก่อน จากนั้นจึงใช้ SVM ในการจำแนก ซึ่งงานวิจัยในช่วงนี้จะนิยมใช้ฐานข้อมูลภาพใบหน้ามาตรฐาน ORL (Oracle Research Laboratory) [11] ซึ่งมีจำนวนคนในฐานข้อมูล 40 คน แต่ละคนมีแสงที่ตกกระทบบนภาพใบหน้าในมุมต่างๆ รวมถึงอิริยาบถของใบหน้าที่แตกต่างกันด้วย ซึ่งจะถ่ายบนภาพพื้นหลังที่เหมือนกัน โดยค่าความถูกต้องของระบบรู้จำภาพใบหน้าในช่วงนี้จะอยู่ในช่วง 80 - 90%

หลังจากนั้นงานวิจัยเกี่ยวกับระบบรู้จำภาพใบหน้าบุคคลถูกนำเสนอต่อมาเรื่อยๆ แต่ยังใช้วิธีการเดิม เช่น ในปี 2010 Agarwal และคณะ [12] ใช้ PCA และโครงข่ายประสาทเทียม เพื่อศึกษาอิทธิพลของจำนวนภาพใบหน้าไอเกนที่มีต่อค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล โดยผลการทดลองพบว่าจำนวนภาพใบหน้าไอเกน 50 ภาพให้ค่าความถูกต้องในการรู้จำภาพใบหน้าสูงสุด นอกจากนี้ระบบที่ช่วยให้ประสิทธิภาพในการรู้จำภาพใบหน้าสูงสุดเมื่อเปรียบเทียบกับวิธี K-Means และ Fuzzy Ant with Fuzzy C-means ในปีเดียวกัน Kim และ Kang [13] เปรียบเทียบประสิทธิภาพของการรู้จำภาพใบหน้าบุคคลด้วย PCA และโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back-propagation) จากการทดลองพบว่าระบบที่นำเสนอมีค่าความถูกต้องสูงสุดเมื่อเปรียบเทียบกับ

ระยะห่างยูคลิดีส (Euclidean distance) และระยะห่างมหาลาโนบิส (Mahalanobis distance) เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

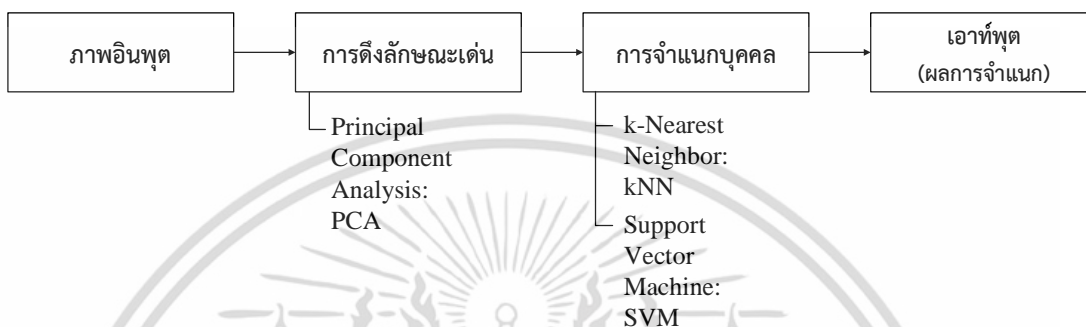
ในปี 2012 Li และ Chen [14] ใช้ PCA และ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ในการรู้จำภาพใบหน้าบุคคล โดยในการทดลองจะเปรียบเทียบระบบที่นำเสนอกับการใช้ PCA เพียงอย่างเดียว ผลการทดลองพบว่าการใช้ PCA ร่วมกับ SVM ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลสูงกว่าการใช้ PCA อย่างเดียว

จนกระทั่งในช่วงปี 2010 การเรียนรู้เชิงลึก (Deep Learning) ถูกนำเสนอขึ้นมา ซึ่งการเรียนรู้เชิงลึกนี้มีหลายวิธีการซึ่งแต่ละวิธีก็จะเหมาะกับการทำงานที่แตกต่างกัน โดยการเรียนรู้เชิงลึกนี้ในช่วงแรกนำเสนอระบบเกี่ยวกับการรู้จำภาพใบหน้าบุคคลและการรู้จำวัตถุ ซึ่งวิธีหนึ่งที่นิยมใช้คือโครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งโครงสร้างถูกนำเสนอในปี 1990 โดย LeCun [15] ซึ่งได้นำเสนอโครงสร้าง LeNet-1 เพื่อใช้งานกับระบบรู้จำตัวเลข และสำหรับระบบรู้จำภาพใบหน้าบุคคลนั้นม้งานวิจัยที่เป็นที่รู้จักในปี 2014 Sun และคณะ [16] ศึกษาการรู้จำภาพใบหน้าบุคคลกับฐานข้อมูลขนาดใหญ่ (10,000 คน) โดยใช้การเรียนรู้เชิงลึกวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งงานวิจัยนี้เน้นไปที่ขนาดของฐานข้อมูลที่ใช้ในการรู้จำภาพใบหน้า ในงานวิจัยนี้ใช้ภาพในฐานข้อมูลประมาณ 200,000 ภาพ และเปรียบเทียบวิธีที่นำเสนอกับวิธีการอื่น โดยใช้ฐานข้อมูลขนาดใหญ่ ผลการทดลองพบว่า วิธีการที่นำเสนอให้ค่าความถูกต้องในการรู้จำภาพใบหน้าที่สูงกว่าวิธีอื่น และในปี 2015 Parkhi และคณะ [17] ได้ศึกษาระบบรู้จำภาพใบหน้าบุคคลที่เหมาะสมกับฐานข้อมูลขนาดใหญ่ (200,000 ภาพ) ที่ให้ค่าความผิดพลาดน้อย และค่าความถูกต้องมากกว่า 90% ซึ่งในงานวิจัยนี้ใช้การเรียนรู้เชิงลึกแบบโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยผลการทดลองที่ได้พบว่า ค่าความถูกต้องสูงกว่าวิธีอื่น

จะเห็นว่างานวิจัยที่เกี่ยวข้องกับระบบรู้จำภาพใบหน้าบุคคลที่ผ่านมานั้นถูกพัฒนาอย่างต่อเนื่องเพื่อให้เหมาะกับการนำไปประยุกต์ใช้ และสำหรับงานวิจัยที่ใช้การเรียนรู้เชิงลึกโดยส่วนมากจะทำกับฐานข้อมูลขนาดใหญ่ [16-19] ผู้วิจัยเห็นว่าหากลองนำการเรียนรู้เชิงลึกมาปรับใช้กับฐานข้อมูลขนาดเล็ก (10 – 100 คน) อาจทำให้เพิ่มประสิทธิภาพได้มากกว่าการใช้ PCA ในการดึงลักษณะเด่น และวิธีการจำแนกแบบต่างๆได้

2.2 หลักการของระบบรู้จำภาพใบหน้าบุคคลแบบทั่วไป

ระบบรู้จำภาพใบหน้าบุคคลนั้นถูกพัฒนาขึ้นมาอย่างต่อเนื่อง และในวิทยานิพนธ์นี้จะกล่าวถึงระบบรู้จำภาพใบหน้าบุคคลใน 2 ลักษณะได้แก่ ระบบรู้จำภาพใบหน้าบุคคลแบบทั่วไป และระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยที่ระบบรู้จำภาพใบหน้าบุคคลแบบทั่วไปนั้นจะประกอบด้วยส่วนต่างๆดังแสดงในรูปที่ 2.1



รูปที่ 2.1 การรู้จำภาพใบหน้าบุคคลโดยทั่วไป

จากรูปที่ 2.1 จะเห็นว่าภาพอินพุตจะผ่านกระบวนการดึงลักษณะเด่น ซึ่งสามารถทำได้หลายวิธี โดยในที่นี้จะยกตัวอย่างการใช้ PCA ในการดึงลักษณะเด่น จากนั้นนำลักษณะเด่นที่ได้เข้าสู่กระบวนการจำแนกบุคคลด้วยเทคนิคการจำแนกต่างๆ เช่น kNN หรือ SVM และสุดท้ายเอาต์พุตที่ได้เป็นผลของการจำแนกจะถูกแสดง

การรู้จำภาพใบหน้าบุคคลโดยทั่วไปนั้นจำเป็นต้องมี 2 ขั้นตอนหลักในการดำเนินงานได้แก่ การดึงลักษณะเด่น และการจำแนก ซึ่งในวิทยานิพนธ์นี้จะกล่าวถึงการดึงลักษณะเด่นโดยใช้ PCA และการจำแนกโดยใช้ kNN และ SVM

2.2.1 การดึงลักษณะเด่นของข้อมูล

วิธีการทั่วไปที่นิยมใช้ในการดึงลักษณะเด่นของข้อมูลได้แก่ การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis :PCA) ซึ่งเป็นเทคนิคในดึงลักษณะเด่นในของภาพอินพุตเพื่อสร้างเวกเตอร์ใหม่ที่อยู่ในปริภูมิย่อย (Subspace) ที่มีมิติน้อยกว่าเวกเตอร์เดิม

การนำ PCA มาใช้ในการพัฒนาระบบรู้จำภาพใบหน้าก็จะทำได้โดยการแปลงภาพถ่ายใบหน้าบุคคลสองมิติไปเป็นเวกเตอร์หนึ่งมิติ จากนั้นจะดึงลักษณะเด่นด้วย PCA และเก็บไว้ในฐานข้อมูล และเมื่อต้องการนำรูปภาพใบหน้าบุคคลที่สนใจมาเปรียบเทียบกับก็จะทำการแปลงภาพใบหน้าเป็นเวกเตอร์หนึ่งมิติด้วย จากนั้นจึงผ่านกระบวนการ PCA เช่นกันแล้วจึงนำเวกเตอร์ไปเปรียบเทียบกับภาพในฐานข้อมูลเพื่อหาผลลัพธ์ โดยในงานวิจัยนี้จะใช้วิธีการรู้จำภาพใบหน้ามนุษย์โดยใช้ PCA ในการศึกษาเกี่ยวกับประสิทธิภาพวิธีการที่ใช้ในระบบรู้จำใบหน้าจะให้ความสนใจกับ

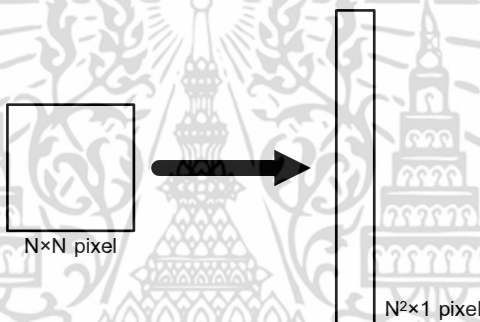
ระเบียบวิธีการ PCA เป็นพิเศษเนื่องจากว่าเป็นระเบียบวิธีการที่มีความซับซ้อนน้อย การนำ PCA มาใช้ในการแทนภาพใบหน้ามนุษย์เป็นครั้งแรกนำเสนอโดย Sirovich และคณะ [6] และต่อมา Turk และ Pentland [7] ได้นำ PCA มาใช้ในการรู้จำภาพใบหน้าโดยมีการกำหนดชื่อเรียกวิธีนี้ขึ้นมาว่า Eigenface หรือภาพใบหน้าไอเกน โดยมีขั้นตอนดังต่อไปนี้

1) การเตรียมข้อมูล

เตรียมภาพในฐานข้อมูล จำนวน M ภาพ โดยแต่ละภาพมีขนาด $N \times N$ พิกเซล แปลงให้เป็นเวกเตอร์ที่มีขนาด $N^2 \times 1$ โดยเวกเตอร์ของภาพเก็บในเมตริกซ์ S ดังแสดงในสมการที่ 2.1

$$S = \{\Gamma_1, \Gamma_2, \dots, \Gamma_M\} \quad (2.1)$$

โดย Γ_i คือ เวกเตอร์ของภาพ ซึ่งมีขนาด $N^2 \times 1$



รูปที่ 2.2 แปลงภาพให้เป็นเวกเตอร์

2) หาภาพใบหน้าเฉลี่ยของฐานข้อมูล (Ψ)

คำนวณภาพใบหน้าเฉลี่ยจากภาพทุกภาพที่มีในฐานข้อมูล สามารถหาได้ดังสมการที่ 2.2

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (2.2)$$

3) หาความแตกต่าง (Φ_i)

หาความแตกต่างระหว่างแต่ละภาพในฐานข้อมูล กับภาพใบหน้าเฉลี่ยของฐานข้อมูล แสดงได้ดังสมการที่ 2.3

$$\Phi_i = \Gamma_i - \Psi \quad (2.3)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) คำนวณโควาเลี่ยนเมตริกซ์ (Covariance matrix) หรือเมตริกซ์ความแปรปรวนร่วม

(C)

การหาโควาเลี่ยนเมตริกซ์สามารถแสดงได้ดังสมการที่ 2.4

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = \mathbf{A} \mathbf{A}^T \quad (2.4)$$

โดย $\mathbf{A} = [\Phi_1, \Phi_2, \dots, \Phi_M]$

5) หาค่าไอเกน (Eigen value: λ_i) และ ไอเกนเวคเตอร์ (Eigen vector: \mathbf{v}_i)

กำหนดให้ $\mathbf{L} = \mathbf{A}^T \mathbf{A}$ จากนั้นหาค่าไอเกน และไอเกนเวคเตอร์ จาก \mathbf{L} ดังแสดงในสมการที่ 2.5

$$\mathbf{L} \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (2.5)$$

จากสมการที่ 2.19 คูณ \mathbf{A} เข้าไปทั้ง 2 ข้างของสมการจะได้ดังสมการที่ 2.6

$$\mathbf{A} \mathbf{L} \mathbf{v}_i = \mathbf{A} \lambda_i \mathbf{v}_i \quad (2.6)$$

แทนค่า $\mathbf{L} = \mathbf{A}^T \mathbf{A}$ จะได้ดังสมการที่ 2.7

$$\mathbf{A} \mathbf{A}^T \mathbf{A} \mathbf{v}_i = \mathbf{A} \lambda_i \mathbf{v}_i \quad (2.7)$$

จากสมการที่ 2.21 นำ \mathbf{A} ทหารทั้ง 2 ข้างของสมการจะได้ดังสมการที่ 2.8

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (2.8)$$

โดยที่ λ_i เป็นค่าไอเกน

\mathbf{v}_i เป็นไอเกนเวคเตอร์

6) โดยค่าไอเกนที่ได้นั้นจะเรียงจากมากไปน้อย จากนั้นจึงเลือกไอเกนเวคเตอร์ที่มีค่าไอเกนสูงสุดมา k ตัว

7) หาภาพใบหน้าไอเกน (\mathbf{U})

หาภาพใบหน้าไอเกนได้จากสมการที่ 2.9

$$\mathbf{U} = \mathbf{A} \mathbf{V} \quad (2.9)$$

โดยที่ $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8) หาค่าสัมประสิทธิ์ของภาพใบหน้าในฐานข้อมูล (Ω) โดยการโปรเจกชันลงบนภาพใบหน้าไอเกน แสดงได้ดังสมการที่ 2.10

$$\Omega = U^T A \quad (2.10)$$

โดยที่ $\Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix}$ และ ω_k คือ สัมประสิทธิ์ของภาพใบหน้าในฐานข้อมูลภาพที่ k

9) เมื่อมีภาพที่ต้องการทดสอบเข้ามา นำมาแปลงเป็นเวกเตอร์ (T) จากนั้นลบด้วยค่าเฉลี่ยของฐานข้อมูล แล้วจึงคูณด้วยภาพใบหน้าไอเกน ผลลัพธ์ที่ได้คือค่าสัมประสิทธิ์ของภาพทดสอบ (Ω_T) แสดงได้ดังสมการที่ 2.11

$$\Omega_T = U [T - \Psi] \quad (2.11)$$

โดยที่ Ω_T คือ สัมประสิทธิ์ของภาพใบหน้าที่น่ามาทดสอบ

เมื่อภาพใบหน้าที่ต้องการทดสอบผ่านกระบวนการดึงลักษณะเด่นเรียบร้อยแล้ว ในส่วนต่อมาจะกล่าวถึงการจำแนกภาพใบหน้าจากลักษณะเด่นโดยใช้วิธีการจำแนกแบบต่างๆ

2.2.2 การจำแนกข้อมูลภาพใบหน้า

การจำแนกข้อมูลเป็นขั้นตอนที่สำคัญในการรู้จำภาพใบหน้าบุคคลเพื่อระบุว่าเป็นภาพที่น่ามาทดสอบนั้นตรงกันกับใครในฐานข้อมูล โดยวิธีที่ใช้ในวิทยานิพนธ์นี้มี 2 วิธีได้แก่ kNN และ SVM

1) k-Nearest Neighbor (kNN)

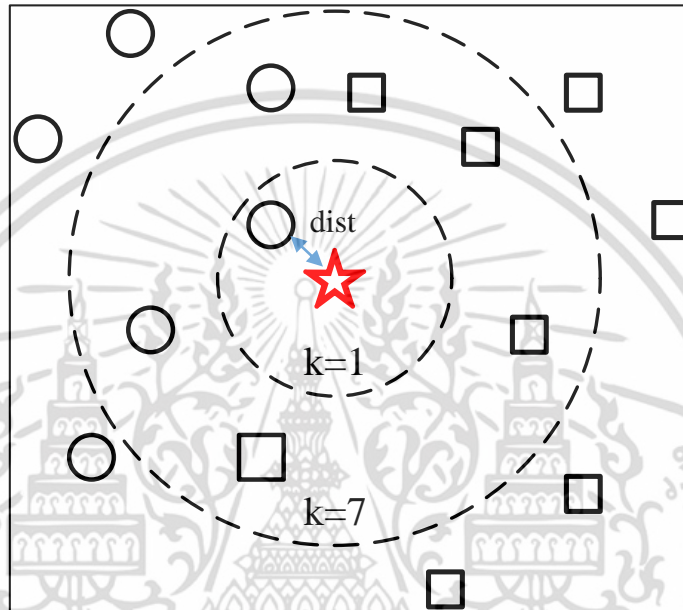
k-Nearest Neighbor หรือ kNN จะจำแนกข้อมูลโดยหาระยะระหว่างข้อมูลอินพุตที่สนใจกับข้อมูลที่อยู่ฐานข้อมูล จากนั้นหาระยะที่ห่างได้มาเรียงลำดับตัวที่ใกล้ที่สุด k ตัว จากนั้นจะดูว่าระยะห่างทั้ง k ตัวว่าข้อมูลใดมีค่าน้อยที่สุด ระบบก็จะระบุได้ว่าอินพุตนั้นตรงกันกับข้อมูลในฐานข้อมูลดังกล่าว โดยการคำนวณระยะความใกล้นั้นจะการใช้การหาระยะทางแบบ Euclidean distance คือการวัดระยะทางระหว่างสองวัตถุถ้าวัตถุห่างกันมากแสดงว่ามีความคล้ายน้อย ถ้าวัตถุห่างกันน้อยแสดงว่ามีความคล้ายกันมาก การคำนวณแสดงดังสมการที่ 2.12

$$\text{dist}(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.12)$$

โดยค่า p_i แทนค่าจากฐานข้อมูล

q_i แทนค่าของอินพุตที่ต้องการทดสอบ

ภาพอธิบายการทำงานของ kNN สามารถอธิบายได้ดังรูปที่ 2.3



รูปที่ 2.3 การทำงานของ kNN

จากรูปที่ 2.3 จะเห็นว่าหากต้องการจำแนกว่าดาวอยู่ในกลุ่มข้อมูลใดระหว่างกลุ่มวงกลม หรือกลุ่มสี่เหลี่ยม สามารถจำแนกได้ด้วย kNN โดยขั้นตอนแรกเป็นการคำนวณระยะห่างระหว่างดาว กับข้อมูลทุกจุดที่อยู่ใกล้เคียง จากนั้นหากพิจารณา $k=1$ ก็จะสนใจเพียงสมาชิก 1 ตัวที่อยู่ใกล้กับตัวที่สนใจมากที่สุด โดยในรูปนี้จะตัดสินว่าเป็น ข้อมูลวงกลม แต่หากพิจารณา $k=7$ ก็จะสนใจสมาชิกที่อยู่รอบข้าง 7 ตัว โดยพบว่าข้อมูลวงกลม 3 ตัว และข้อมูลสี่เหลี่ยม 4 ตัว ดังนั้นระบบจะตัดสินว่าดาวจัดอยู่ในกลุ่มเดียวกันกับข้อมูลสี่เหลี่ยม จะเห็นว่าที่ค่า k แตกต่างกัน ส่งผลถึงคำตอบที่ได้แตกต่างกันด้วย

และเพื่อเพิ่มความเข้าใจให้มากขึ้นได้ยกตัวอย่างการใช้งาน kNN : ตารางที่ 2.1 แสดงข้อมูลส่วนบุคคล 4 คนได้แก่ ส่วนสูง น้ำหนัก รอบเอว รอบอก และเพศ จงหาว่าบุคคลที่มีลักษณะต่อไปนี้นี้เป็นเพศอะไร เมื่อกำหนดให้ส่วนสูง = 170 นิ้ว น้ำหนัก = 60 กิโลกรัม รอบเอว = 30 นิ้ว และรอบอก = 34 นิ้ว โดยให้หากรณีที่ $k=1$ และ $k=3$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ข้อมูลแต่ละบุคคล

No.	Height (cm)	Weight (kg)	Waistline (inch)	Chest (inch)	Gender
1	165	60	32	37	F
2	175	75	33	43	M
3	166	50	30	34	M
4	155	50	28	32	F

วิธีทำ

คำนวณ Euclidean distance ตามสมการที่ 2.26 ระหว่างข้อมูลที่สนใจ กับข้อมูลในฐานข้อมูลทั้ง 4 คนได้ผลลัพธ์ดังต่อไปนี้

$$dist1 = \sqrt{(165-170)^2 + (60-60)^2 + (32-30)^2 + (37-34)^2} = \sqrt{38} = 6.1644$$

$$dist2 = \sqrt{(175-170)^2 + (75-60)^2 + (33-30)^2 + (43-34)^2} = \sqrt{340} = 18.4391$$

$$dist3 = \sqrt{(166-170)^2 + (50-60)^2 + (30-30)^2 + (34-34)^2} = \sqrt{116} = 10.7703$$

$$dist4 = \sqrt{(155-170)^2 + (50-60)^2 + (28-30)^2 + (32-34)^2} = \sqrt{333} = 18.2483$$

เรียงค่า k จะได้ว่า $dist1 < dist3 < dist4 < dist2$

กรณี $k = 1$ จะเลือกคำตอบมา 1 คำตอบได้แก่ $dist1$ ซึ่งมีค่าน้อยที่สุด

ดังนั้นคำตอบกรณี $k=1$ เนื่องจากค่าระยะห่างที่น้อยที่สุดคือค่าที่นำไปเปรียบเทียบกับบุคคลที่ 1 จึงสรุปว่าบุคคลนั้นเป็น เพศหญิง (F) เช่นเดียวกับบุคคลที่ 1

กรณี $k=3$ จะเลือกคำตอบมา 3 คำตอบเรียงจากน้อยไปมากจะได้ว่า

$dist1$ ซึ่งก็คือ เพศหญิง (F)

$dist3$ ซึ่งก็คือ เพศชาย (M)

$dist4$ ซึ่งก็คือ เพศหญิง (F)

จะเห็นว่าคำตอบที่ได้เป็นเพศหญิง 2 คำตอบ และเพศชาย 1 คำตอบ โดยระบบจะนำคำตอบที่ถูกต้องมากที่สุดเป็นคำตอบของระบบ โดยในกรณีนี้ มีคำตอบเป็นเพศหญิง 2 คำตอบ

ดังนั้นคำตอบกรณี $k=3$ จะได้ว่าบุคคลที่ 5 เป็นเพศหญิง (F)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) การจำแนกข้อมูลภาพใบหน้าด้วย SVM

Support Vector Machines (SVM) [20 - 21] เป็นกระบวนการทางคณิตศาสตร์ซึ่งสามารถสร้างระนาบที่เหมาะสมที่สุดในการแบ่งกลุ่ม โดยจะพยายามสร้างเส้นแบ่งระหว่างข้อมูลให้มีระยะห่างระหว่างขอบของข้อมูลทั้งสองกลุ่มมากที่สุด

2.1) Linear Support Vector Machines

สำหรับการแก้ปัญหาแบบเป็นเชิงเส้นนั้น มิติของข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยใช้ระนาบการตัดสินใจดังสมการที่ 2.13

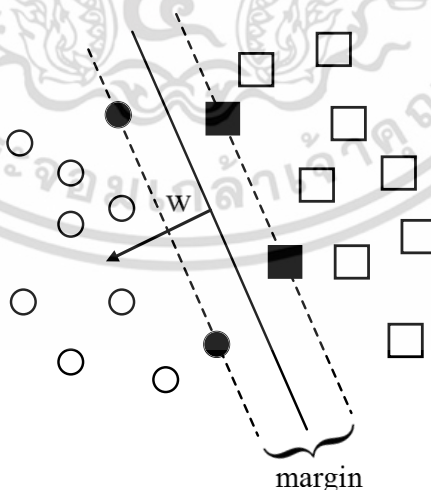
$$f(x) = \text{sign}((w \cdot x) + b) \quad (2.13)$$

โดยที่ $\{(x_i, y_i), \dots, (x_l, y_l)\}$ ($x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}$) ซึ่ง $(x_i, y_i), \dots, (x_l, y_l)$ เป็นตัวอย่างที่ใช้สำหรับสอน l คือจำนวนข้อมูลตัวอย่าง n คือจำนวนมิติข้อมูลเข้า และ y คือผลลัพธ์มีค่าเป็น +1 หรือ -1

w เป็นค่าน้ำหนัก

b คือค่าไบแอสสมการซึ่งใช้สำหรับจำแนกข้อมูล

ระยะระหว่างขอบของข้อมูลหรือ มาร์จิน (Margin) ระหว่างข้อมูลที่เป็นบวก และข้อมูลที่เป็นลบถูกกำหนดโดย 2 ระนาบได้แก่ $x \cdot w + b = \pm 1$ แสดงได้ดังรูปที่ 2.4 ซึ่งจะเห็นว่าไม่มีจุดที่อยู่ระหว่างเส้นแบ่งทั้งสอง และความกว้างระหว่างเส้นแบ่งถึงขอบของข้อมูลมีระยะ $2/\|w\|$



รูปที่ 2.4 การสร้างเส้นแบ่งของ SVM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งวิธีของซัพพอร์ตเวกเตอร์นั้นจะพยายามหาระยะจากเส้นแบ่งถึงขอบของข้อมูลให้มากที่สุด โดยที่ค่ามาร์จินน้อยที่สุด เพื่อให้ได้เส้นแบ่งที่เหมาะสมที่สุด ซึ่งจุดที่อยู่บนเส้นขอบจะถูกเรียกว่า ซัพพอร์ตเวกเตอร์ จากรูปที่ 2.5 คือวงกลมสีดำและสี่เหลี่ยมสีดำที่อยู่บนเส้นประ

หากไม่สามารถจำแนกข้อมูลได้ถูกต้องทั้งหมด จำเป็นต้องมีตัวแปรเพื่อใช้ยอมรับค่าความผิดพลาด ซึ่งจะเพิ่มตัวแปร สแลค (ξ_i) เข้าไปซึ่งสามารถแสดงได้ดังสมการที่ 2.14

$$\begin{aligned} \mathbf{w}^T \cdot \mathbf{x} + b &\geq +1 - \xi_i; y_i = +1 \\ \mathbf{w}^T \cdot \mathbf{x} + b &\geq -1 + \xi_i; y_i = -1 \end{aligned} \quad (2.14)$$

โดยที่ $\xi_i > 0$

ทำให้โครงสร้างของ SVM ประกอบด้วย 2 ส่วนคือ การเพิ่มระยะขอบให้มากที่สุด และการแก้ปัญหาด้วยการลดข้อผิดพลาดให้ต่ำที่สุด โดยสมการการหาค่ามาร์จินที่น้อยที่สุด แสดงได้ดังสมการที่ 2.15

$$\text{Margin} = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (2.15)$$

โดย C เป็นค่าที่กำหนดความสมดุลระหว่างการให้ความสำคัญระยะขอบมากที่สุด หรือให้ความสำคัญค่าความผิดพลาดที่ต้องการให้ต่ำ ซึ่งถ้า C มีค่ามาก หมายถึงไม่ยอมให้มีคลาสอื่นอยู่ระหว่างเส้นขอบได้ แต่ถ้า C มีค่าน้อยจะยอมให้มีคลาสอื่นอยู่ระหว่างเส้นขอบได้บ้างหรือการยอมให้มีค่าความผิดพลาดเกิดขึ้นได้

จากนั้นจะแก้ปัญหาตัวแปรแบบเซตคู่ด้วยฟังก์ชันลากรองจ์ (Lagrangian) จะทำให้ได้ข้อจำกัดที่ดีที่สุด ได้ดังสมการที่ 2.16

$$f(x) = \text{sign} \left(\sum_{j=1}^N w_j \alpha_j^* K(x, x_j) + b \right) \quad (2.16)$$

โดยที่ α_j^* เป็นค่าสัมประสิทธิ์ของข้อมูลที่ j

ซึ่งจะช่วยจำแนกข้อมูลที่ไม่เป็นฟังก์ชันการจำแนกแบบเชิงเส้น ซึ่ง $K(x, x_j)$ เป็น Kernel function ซึ่งจะอธิบายในหัวข้อถัดไป

2.2) Non-linear Support Vector Machines

เพื่อให้การแบ่งกลุ่มของข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (non-linear dataset)

จำเป็นต้องแบ่งกลุ่มข้อมูลตัวอย่างไปสู่ปริภูมิมิติที่สูงขึ้น (higher dimensional space) หรือเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

feature space โดยการแปลงดังกล่าวต้องทำกับฟังก์ชันที่ไม่เป็นเชิงเส้น โดยสมการที่ใช้ในการคำนวณเพื่อแบ่งกลุ่มข้อมูลที่ไม่เป็นเชิงเส้นได้ดังนี้

Polynomial kernel แสดงได้ดังสมการที่ 2.17

$$K(\mathbf{x}, \mathbf{x}_j) = (\mathbf{x} \cdot \mathbf{x}_j + 1)^n \quad (2.17)$$

Radial Basis Function kernel (RBF) แสดงได้ดังสมการที่ 2.18

$$K(\mathbf{x}, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x} - \mathbf{x}_j\|^2} \quad (2.18)$$

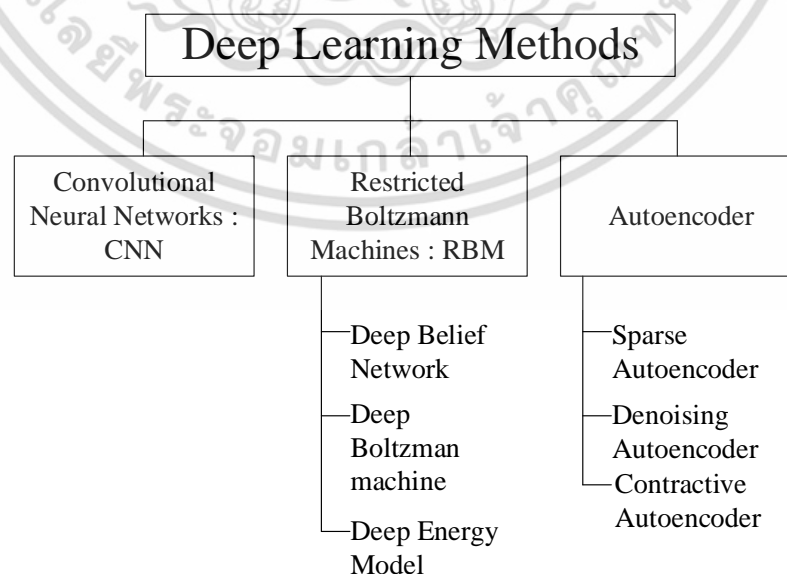
Sigmoid kernel แสดงได้ดังสมการที่ 2.19

$$K(\mathbf{x}, \mathbf{x}_j) = \tanh(K(\mathbf{x} \cdot \mathbf{x}_j) - \delta) \quad (2.19)$$

ซึ่งในวิทยานิพนธ์นี้จะใช้ kernel ในการทดสอบ 2 kernel ด้วยกัน ได้แก่ แบบเป็นเส้นตรง และแบบ RBF

2.3 การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep learning) เป็นรูปแบบหนึ่งของ Machine learning ซึ่งอยู่บนพื้นฐานของระเบียบวิธีที่จะสร้างรูปแบบแนวคิดระดับสูงให้กับข้อมูล โดยการใช้การประมวลผลบนโครงสร้างซับซ้อน การเรียนรู้เชิงลึกมีหลายรูปแบบดังแสดงในรูปที่ 2.5



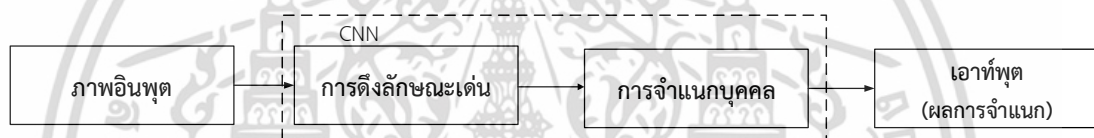
รูปที่ 2.5 ภาพรวมของการเรียนรู้เชิงลึก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.5 การเรียนรู้เชิงลึกสามารถแบ่งออกได้หลายประเภท ดังต่อไปนี้ Convolutional Neural Networks (CNN) Restricted Boltzmann Machine และ Auto encoder ซึ่งแต่ละวิธีก็จะเหมาะกับการดำเนินงานที่แตกต่างกันเช่น CNN เหมาะกับงานวิจัยการรู้จำวัตถุ Deep Belief Network (DBN) [22] เหมาะกับงานประเภทการรู้จำตัวอักษร Deep Boltzmann Machine (DBM) [22] และ Stacked (de-noising) auto-encoders [23] เหมาะกับงานวิจัยประเภทการจำแนกข้อมูลตัวอักษร เป็นต้น ซึ่งการใช้งานนั้นขึ้นอยู่กับความเหมาะสมและจุดมุ่งหมายในการดำเนินงาน โดยในวิทยานิพนธ์นี้เลือกใช้การเรียนรู้เชิงลึกแบบ Convolutional Neural Networks

2.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN)

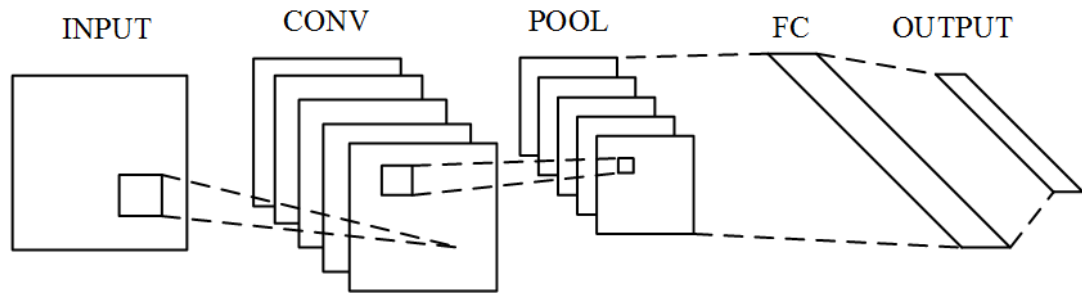
โครงข่ายประสาทเทียมแบบคอนโวลูชันแตกต่างกับระบบรู้จำภาพใบหน้าบุคคลโดยใช้ทั่วไปคือการรวมเอาส่วนของการดึงลักษณะเด่นและส่วนของการจำแนกเอาไว้ด้วยกันดังแสดงได้รูปที่ 2.6



รูปที่ 2.6 การรู้จำภาพใบหน้าบุคคลโดยใช้ CNN

จากรูปที่ 2.6 การรู้จำภาพใบหน้าบุคคลโดยใช้ CNN นั้นจะรวมเอาส่วนของการดึงลักษณะเด่นและการจำแนกเข้าไว้ในวิธีการเดียว ทำให้ไม่จำเป็นต้องหาวิธีการดึงลักษณะเด่นอื่นๆ หรือวิธีการจำแนกอื่นมาใช้ โดยการดึงลักษณะเด่นนั้นจะมีวิธีการย่อภายในคือการทำคอนโวลูชัน

Convolutional Neural Networks (CNN) [24] หรือ โครงข่ายประสาทเทียมแบบคอนโวลูชันนั้นมีจุดเด่นคือสามารถดึงลักษณะเด่นและจำแนกข้อมูลได้โดยตรง เหมาะสำหรับการรู้จำภาพ โดย CNN เป็นโครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer perceptron) โดย LeCun [15] เป็นคนแรกที่นำเสนอโครงสร้าง CNN แบบวิธีการแพร่ค่าความผิดพลาดย้อนกลับ (Back-propagation) สำหรับการรู้จำตัวเลขด้วยลายมือ ทำให้ CNN รูปแบบนี้เป็นที่นิยมใช้จนถึงปัจจุบัน เนื่องจากโครงข่ายของระบบนี้ไม่ได้ถูกสอนแบบขั้นต่อขั้น แต่จะทำการปรับค่าน้ำหนักเพื่อลดฟังก์ชันความคลาดเคลื่อนด้วยวิธีการแพร่ค่าความผิดพลาดย้อนกลับ นอกจากนี้ลักษณะเด่นของข้อมูลจะถูกดึงออกมาโดยอัตโนมัติด้วยการเรียนรู้หน้าจากคอนโวลูชันที่เกิดจากค่าน้ำหนักของแผนภาพคุณลักษณะเด่น (Feature maps) โดยภาพโครงสร้างโดยทั่วไปของ CNN แสดงได้ดังรูปที่ 2.7



รูปที่ 2.7 โครงสร้างโดยทั่วไปของ CNN

รูปที่ 2.7 แสดงโครงสร้างโดยทั่วไปของ CNN ซึ่งประกอบด้วย ข้อมูลอินพุต (INPUT) ชั้นคอนโวลูชัน (CONV) ชั้นพูลลิ่ง (POOL) ชั้นลักษณะเด่น (FC) และชั้นเอาต์พุต (OUTPUT) ซึ่งแต่ละส่วนมีการทำงานที่แตกต่างกันดังซึ่งสามารถอธิบายได้ดังต่อไปนี้

1) ชั้นอินพุต (Input layer)

ชั้นนี้เป็นการนำเอาภาพอินพุตสู่ระบบ

2) ชั้นคอนโวลูชัน (Convolutional layer)

ในชั้นตอนนี้คือนำภาพอินพุตมาคอนโวลูชันกับหน้ากาก โดยที่ค่าสัมประสิทธิ์ในหน้ากานั้นได้จากการสุ่มขึ้นมา ซึ่งการทำคอนโวลูชันสำหรับเอาต์พุตที่ j ในชั้น l มีสมการในดำเนินการแสดงได้ดัง (2.20)

$$y_j^{(l)}(x, y) = \phi^{(l)} \left(\sum_{i \in l} \sum_{(u, v) \in K} w_{ji}^{(l)}(u, v) y_i^{(l-1)}(x+u, y+v) + b_j^{(l)} \right) \quad (2.20)$$

โดยที่ $K = \{(u, v) \in \mathbb{N}^2 \mid 0 \leq u \leq s_x \text{ and } 0 \leq v \leq s_y\}$ และ (s_x, s_y) เป็นขนาดของหน้ากากคอนโวลูชัน

$w_{ji}^{(l)}(u, v)$ เป็นหน้ากากที่ใช้ในการคอนโวลูชันในชั้น l

$y_i^{(l-1)}(x+u, y+v)$ เป็นอินพุต

$b_j^{(l)}$ เป็นค่าไบแอส

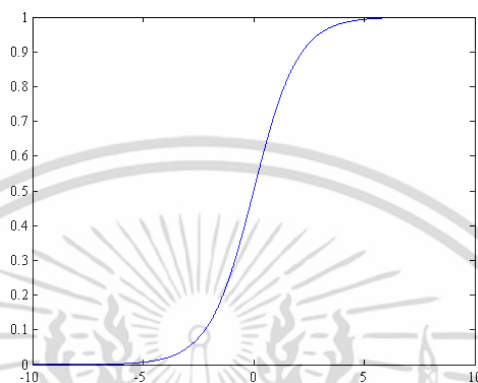
$\phi^{(l)}$ เป็นฟังก์ชันกระตุ้น (Activation function)

โดยฟังก์ชันกระตุ้นที่นิยมใช้กันมี 3 ชนิดด้วยกันได้แก่

1. Sigmoid function

สมการในของ Sigmoid function แสดงได้ดังสมการที่ 2.21 และผลการทำงานของ Sigmoid function แสดงได้ดังรูปที่ 2.8

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.21)$$



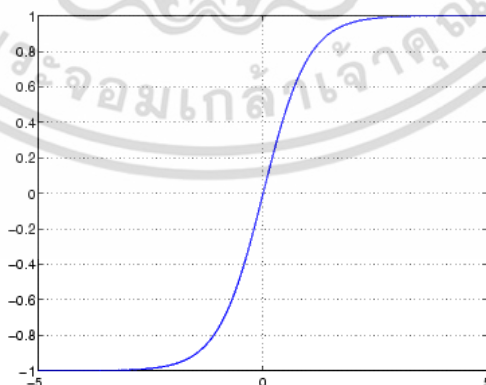
รูปที่ 2.8 ผลของ Sigmoid function

รูปที่ 2.8 จะเห็นว่าค่าในแกน X มีค่าอยู่ในช่วง -10 – 10 และแกน Y จะมีค่าตั้งแต่ 0 – 1

2. Tan hyperbolic function (tanh function)

สมการในการดำเนินการของ tanh function แสดงได้ดังสมการที่ 2.22 และผลการดำเนินการแสดงได้ดังรูปที่ 2.9

$$f(x) = \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.22)$$



รูปที่ 2.9 ผลของ tanh function

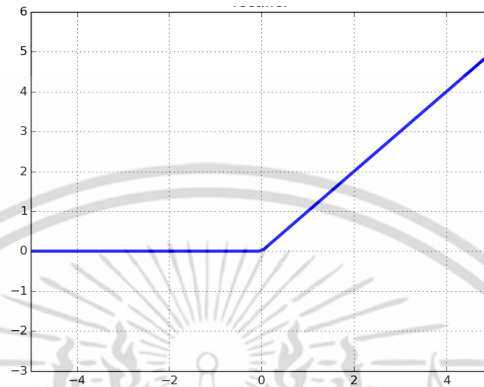
รูปที่ 2.10 จะเห็นว่าค่าในแกน X มีค่าอยู่ในช่วง -5 – 5 และแกน Y จะมีค่าตั้งแต่ -1 – 1 ซึ่ง

แตกต่างจาก Sigmoid function ที่มีค่าในแกน Y อยู่ในช่วงบวกเท่านั้น เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. Rectified Linear Unit (ReLU function)

สมการในการดำเนินการของ ReLU function แสดงได้ดังสมการที่ 2.23 และผลการดำเนินการแสดงได้ดังรูปที่ 2.10

$$f(x) = \max(0, x) \quad (2.23)$$

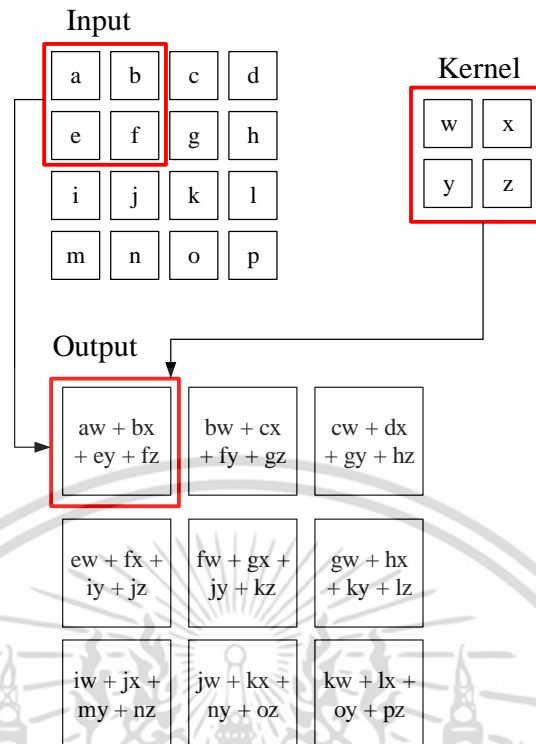


รูปที่ 2.10 ผลของ ReLU function

รูปที่ 2.11 จะเห็นว่าเมื่อค่าในแกน X มีค่าติดลบ ค่าในแกน Y จะมีค่าเป็น 0 และเมื่อค่าในแกน X เปลี่ยนแปลงเข้าสู่ค่าบวกค่าในแกน Y จะเท่ากับค่าในแกน X นั้น

จากฟังก์ชันกระตุ้นทั้ง 3 แบบที่ได้กล่าวไปนั้น ReLU function ถูกนำมาใช้งานกับ CNN มากกว่าฟังก์ชันอื่น เนื่องจากโอกาสที่บางข้อมูลจะสูญหาย โดยที่หากข้อมูลมีค่ามากกว่า 0 จะเลือกมาเป็นเอาต์พุต แต่สำหรับกรณีข้อมูลมีค่าน้อยกว่า 0 จะให้ค่าเอาต์พุตเป็น 0

และตัวอย่างการทำคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากอกขนาด 2×2 สามารถแสดงได้ดังรูปที่ 2.11



รูปที่ 2.11 ตัวอย่างการทำงานคอนโวลูชัน

รูปที่ 2.11 เป็นตัวอย่างการคอนโวลูชันระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากอกขนาด 2×2 ทำให้ได้เอาต์พุตขนาด 3×3 พิกเซล ซึ่งเอาต์พุตเกิดจากการนำค่าสัมประสิทธิ์ของหน้ากอกไปคูณกับค่าของอินพุตจากนั้นนำทุกค่าที่ได้บวกกันเพื่อเป็นเอาต์พุต จากตัวอย่างจะเห็นว่า หน้ากอกมีสัมประสิทธิ์ w, x, y และ z ไปคอนโวลูชันกับอินพุต a, b, e และ f ได้เอาต์พุตเป็น $aw+bx+ey+fz$ จากนั้นเลื่อนหน้ากอกไปยังข้อมูลอินพุตถัดไป และทำกระบวนการเดิมจนครบขนาดของอินพุต โดยเอาต์พุตจะมีขนาดลดลงเนื่องจากกระบวนการนี้จะสูญเสียขอบของภาพไป

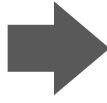
3) ชั้นพูลลิ่ง (Pooling layer)

ในชั้นตอนนี้เป็นการลดขนาดของข้อมูลลงโดยวิธีการทำพูลลิ่งที่นิยมนั้นมี 3 แบบด้วยกันดังต่อไปนี้

1. พูลลิ่งด้วยค่าสูงสุด (Max pooling)

การทำพูลลิ่งด้วยค่าสูงสุดนั้นจะเลือกค่าสูงสุดในแต่ละบล็อกที่หน้ากอกพูลลิ่งครอบอยู่ข้อมูลอินพุตมาเป็นคำตอบ ตัวอย่างการทำงานของการทำพูลลิ่งด้วยค่าสูงสุดแสดงได้ดังรูปที่ 2.12

1	5	8	0
6	4	3	2
3	0	5	9
2	7	0	1



6	8
7	9


รูปที่ 2.12 การทำพูลลิงด้วยค่าสูงสุด

รูปที่ 2.12 การทำงานของพูลลิงด้วยค่าสูงสุดระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากลูปลิงขนาด 2×2 ขั้นตอนการทำงานคือ นำหน้ากากลูปลิงขนาด 2×2 ไปครอบที่อินพุต จากนั้นจะเลือกค่าสูงสุดออกมา ซึ่งจะได้เอาท์พุตเป็น 6 8 7 และ 9 ตามลำดับ โดยขนาดของเอาท์พุตจะมีขนาดลดลงครึ่งหนึ่งของขนาดอินพุต

2. การพูลลิงด้วยค่าเฉลี่ย (Average pooling)

พูลลิงด้วยค่าเฉลี่ยจะคล้ายกับการทำพูลลิงด้วยค่าสูงสุด มีส่วนที่แตกต่างกันคือจะคำนวณค่าเฉลี่ยในแต่ละบล็อก แทนที่การเลือกค่าสูงสุด ตัวอย่างการทำงานแสดงได้ดังรูปที่ 2.13

1	5	8	0
6	4	3	2
3	0	5	9
2	7	0	1



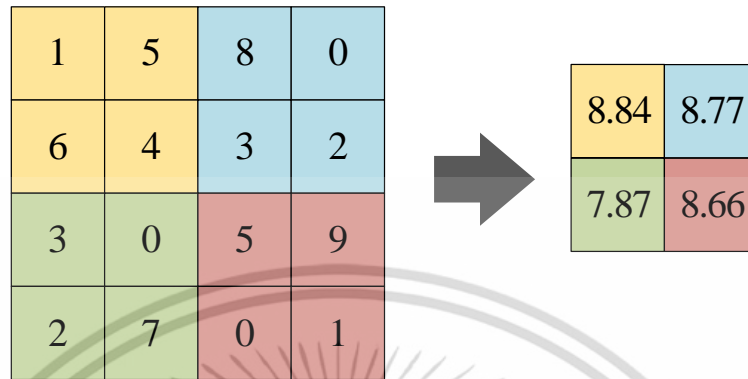
4	3.25
3	3.75

รูปที่ 2.13 การทำพูลลิงด้วยค่าเฉลี่ย

รูปที่ 2.13 การทำงานของพูลลิงด้วยค่าเฉลี่ยระหว่างอินพุตขนาด 4×4 พิกเซล กับหน้ากากลูปลิงขนาด 2×2 ขั้นตอนการทำงานคือ นำหน้ากากลูปลิงขนาด 2×2 ไปครอบที่อินพุต จากนั้นจะเลือกหาค่าเฉลี่ยในแต่ละกรอบที่ครอบออกมา ซึ่งจะได้เอาท์พุตเป็น 4 3.25 3 และ 3.75 ตามลำดับ ซึ่งจะเห็นว่าค่าของเอาท์พุตที่ได้จะมีค่าที่ใกล้เคียงกัน โดยขนาดของเอาท์พุตจะมีขนาดลดลงครึ่งหนึ่งของขนาดอินพุต

3. การพูลลิงด้วย L_2 -norm (L_2 -norm pooling)

วิธีการนี้จะเป็นการหาค่า L_2 -norm ของแต่ละบล็อกมาเป็นเอาต์พุต ตัวอย่างการทำงาน แสดงได้ดังรูปที่ 2.14



รูปที่ 2.14 การทำพูลลิงด้วย L_2 -norm

รูปที่ 2.14 การทำงานของพูลลิงด้วย L_2 -norm ระหว่างอินพุตขนาด 4×4 พิกเซล กับ หน้ากากพูลลิงขนาด 2×2 ขั้นตอนการทำงานคือ นำหน้ากขนาด 2×2 ไปครอบที่อินพุต จากนั้น คำนวณค่า L_2 -norm ในแต่ละกรอบที่ครอบออกมา โดยการคำนวณจะนำอินพุตในกรอบแต่ละตัว ยกกำลังสอง จากนั้นนำทั้งหมดมาบวกกันแล้วจึงหารากที่สองของผลลัพธ์นั้น ซึ่งจะได้เอาต์พุตเป็น 8.84 8.77 7.87 และ 8.66 ตามลำดับ จะเห็นว่าค่าที่ได้ออกมาเป็นเอาต์พุตนั้นมีค่าใกล้เคียงกัน โดยขนาดของเอาต์พุตจะมีขนาดลดลงครึ่งหนึ่งของขนาดอินพุต

จากวิธีการพูลลิงทั้ง 3 วิธีกับอินพุตรูปแบบเดียวกันก็จะให้ค่าเอาต์พุตที่แตกต่างกัน โดยการพูลลิงด้วยค่าสูงสุดจะให้ข้อมูลในชั้นเอาต์พุตมีความแตกต่างกัน แต่สำหรับการพูลลิงด้วยค่าเฉลี่ย และพูลลิงด้วย L_2 -norm จะให้ค่าของข้อมูลเอาต์พุตที่ใกล้เคียงกัน ดังนั้นวิธีการที่นิยมใช้กันคือ การพูลลิงด้วยค่าสูงสุด เนื่องจากมีขั้นตอนการทำงานที่ง่ายกว่า และได้ค่าที่มีความแตกต่างกัน มากกว่าวิธีการอื่น ซึ่งจะส่งผลถึงลักษณะของข้อมูลที่ได้ด้วย

4) ชั้นลักษณะเด่น (Fully-Connected layer)

เมื่อผ่านกระบวนการทำงานของชั้นคอนโวลูชัน ชั้นพูลลิงมาแล้วต่อมาจะนำเอาต์พุตจาก ชั้นเหล่านี้เข้าสู่ชั้นลักษณะเด่น เพื่อนเตรียมตัวเข้าสู่ชั้นจำแนกต่อไป โดยชั้นนี้จะเชื่อมโยงจะนำเอา เซลล์ประสาทจากชั้นก่อนหน้า (ชั้นคอนโวลูชัน และชั้นพูลลิง) มาเชื่อมโยงกันกับเซลล์ประสาททุก เซลล์ที่มี โดยจะเชื่อมต่อกันเป็น 1 มิติ ดังนั้นจะไม่สามารถมีชั้นคอนโวลูชันหลังจากชั้นนี้ได้อีก โดยชั้น นี้จะทำเชื่อมต่อไปยังชั้นเอาต์พุต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) ชั้นเอาต์พุต (Output layer)

โดยชั้นนี้จะเป็นชั้นของการจำแนกข้อมูล โดยจำนวนเซลล์ประสาทในชั้นนี้จะเท่ากับจำนวนคนที่ใช้ในการรู้จำภาพใบหน้าบุคคล ซึ่งสมการที่ใช้ในการจำแนกแสดงได้สมการที่ 2.24

$$y_i = \frac{\exp(y'_i)}{\sum_{j=1}^n \exp(y'_j)} \quad (2.24)$$

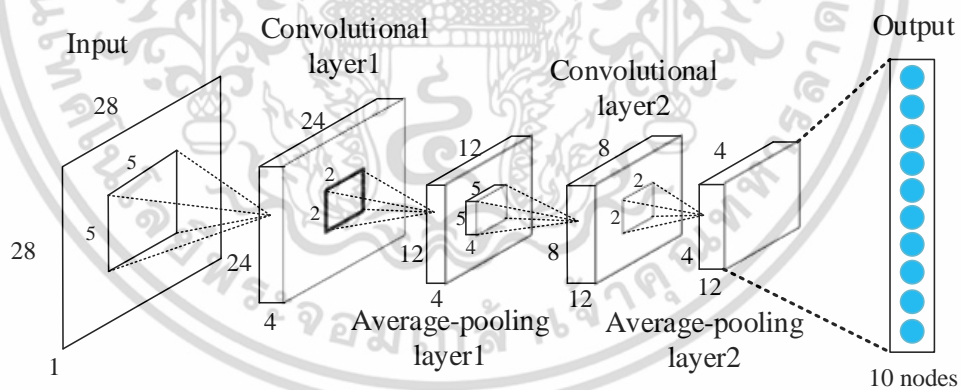
เมื่อ $y'_j = \sum_{i=1}^m x_i \cdot w_{i,j} + b_j$ โดยที่ m เป็นจำนวนเซลล์ประสาทในชั้นลักษณะเด่น

x_i เป็นอินพุตของเซลล์ประสาท j

y_j เป็นเอาต์พุต

n เป็นจำนวนบุคคลที่ต้องการจำแนก

ในส่วนต่อมาจะกล่าวถึงตัวอย่างการทำงานของ CNN โดยโครงสร้างที่ใช้เป็นโครงสร้าง LeNet-1 ของ LeCun [15] สำหรับการรู้จำตัวเลขด้วยลายมือ โดยภาพที่นำมาใช้นั้นเป็นภาพตัวเลขที่เขียนด้วยลายมือ และจะมีการจำแนกตัวเลขทั้งหมด 10 ตัว ได้แก่ 0 ถึง 10 ซึ่งโครงสร้างแสดงได้ดังรูปที่ 2.15



รูปที่ 2.15 โครงสร้าง LeNet-1

รูปที่ 2.15 โครงสร้างของ LeNet-1 ประกอบด้วย ชั้นอินพุต 1 ชั้น ชั้นคอนโวลูชัน 2 ชั้น ชั้นพูลลิง 2 ชั้น และชั้นพูลลิงชั้นที่ 2 เชื่อมต่อเข้าไปยังชั้นเอาต์พุต ซึ่งมีขั้นตอนการดำเนินงานดังต่อไปนี้

1) อินพุต ($y^{(0)}$) มีขนาด 28×28 พิกเซล ซึ่งเป็นภาพโทนสีเทาทำให้มิติของข้อมูลเป็น 1 ซึ่งจะนำไปทำคอนโวลูชันกับหน้ากากขนาด 5×5 ทำให้มีค่าน้ำหนัก ($w_{j_0}^{(1)}(u, v)$) 25 ค่า โดยใช้หน้ากากคอนโวลูชันจำนวน 4 หน้ากาก โดยการทำงานในชั้นนี้แสดงได้ดังสมการที่ 2.25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$y_j^{(1)}(x, y) = \phi^{(1)} \left(\sum_{(u,v) \in K} w_{j0}^{(1)}(u, v) y^{(0)}(x+u, y+v) + b_j^{(1)} \right) \quad (2.25)$$

โดยที่ $K = \{(u, v) \in \mathbb{N}^2 \mid 0 \leq u \leq 5 \text{ and } 0 \leq v \leq 5\}$

$y_j^{(1)}(x, y)$ เป็นเอาต์พุตของชั้นคอนโวลูชันชั้นที่ 1

$\phi^{(1)}$ ฟังก์ชันกระตุ้น

$y^{(0)}(x+u, y+v)$ เป็นข้อมูลอินพุต

$w_{j0}^{(1)}(u, v)$ หน้ากากคอนโวลูชัน

$b_j^{(1)}$ ค่าไบแอส

โดยเอาต์พุตที่ได้จากการนำอินพุตขนาด 28×28 พิกเซล คอนโวลูชันกับหน้ากากขนาด 5×5 จำนวน 4 หน้ากาก ทำให้ได้เอาต์พุตจากการคอนโวลูชัน 24×24 จำนวน 4 ชั้น

2) นำเอาต์พุตจากชั้นคอนโวลูชันชั้นที่ 1 ไปลดขนาดโดยใช้พูลลิ่งด้วยค่าเฉลี่ย ทำให้ขนาดของข้อมูลลดลงครึ่งหนึ่งทำให้ขนาดของเอาต์พุตที่ได้จากการทำพูลลิ่งคือ 12×12 จำนวน 4 ชั้น

3) จากนั้นนำเอาต์พุตจากชั้นพูลลิ่งชั้นที่ 1 ไปคอนโวลูชันกับหน้ากากขนาด 3×3 จำนวน 12 หน้ากาก โดยมีสมการในการดำเนินการดังสมการที่ 2.26

$$y_j^{(3)}(x, y) = \phi^{(3)} \left(\sum_{(u,v) \in K} w_{j2}^{(3)}(u, v) y^{(2)}(x+u, y+v) + b_j^{(3)} \right) \quad (2.26)$$

โดยที่ $K = \{(u, v) \in \mathbb{N}^2 \mid 0 \leq u \leq 3 \text{ and } 0 \leq v \leq 3\}$

$y_j^{(3)}(x, y)$ เป็นเอาต์พุตของชั้นคอนโวลูชันชั้นที่ 2

$\phi^{(3)}$ ฟังก์ชันกระตุ้น

$y^{(2)}(x+u, y+v)$ เป็นข้อมูลอินพุต

$w_{j2}^{(3)}(u, v)$ หน้ากากคอนโวลูชัน

$b_j^{(3)}$ ค่าไบแอส

โดยเอาต์พุตที่ได้จากการนำอินพุตขนาด 12×12 พิกเซล คอนโวลูชันกับหน้ากากขนาด 5×5 จำนวน 12 หน้ากาก ทำให้ได้เอาต์พุตจากการคอนโวลูชัน 8×8 จำนวน 12 ชั้น

4) นำเอาต์พุตจากชั้นคอนโวลูชันชั้นที่ 2 ไปลดขนาดโดยใช้พูลลิ่งด้วยค่าเฉลี่ย ทำให้ขนาดของข้อมูลลดลงครึ่งหนึ่งทำให้ขนาดของเอาต์พุตที่ได้จากการทำพูลลิ่งคือ 6×6 จำนวน 12 ชั้น

5) ขั้นสุดท้ายคือการนำเอาต์พุตจากชั้นพูลลิ่งชั้นที่ 2 นำมาต่อกันเพื่อเป็นชั้นเอาต์พุตของระบบโดยมีจำนวนเซลล์ประสาทที่ชั้นเอาต์พุต 10 เซลล์ประสาท ซึ่งเอาต์พุตจะมีค่า +1 เมื่อคำตอบถูกต้อง และ -1 เมื่อคำตอบผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนต่อมาเป็นการนำหาค่าความผิดพลาดจากนั้นจะนำค่าความผิดพลาดที่ได้แพร่ย้อนกลับ (Back-propagation) เพื่อปรับค่าน้ำหนักในชั้นก่อนหน้า

6) หาค่าเกรเดียนต์ที่ชั้นเอาต์พุต

การหาค่าเกรเดียนต์ของเซลล์ประสาท k ที่ชั้นเอาต์พุตได้ดังสมการที่ 2.27

$$\delta_k^{(5)} = e_k \phi' \left(\sum_j w_{kj}^{(5)} y_j^{(4)} \right) \quad (2.27)$$

โดยที่ $\delta_k^{(5)}$ เป็นค่าเกรเดียนต์ที่ชั้นเอาต์พุต

$e_k = o_k - t_k$ เป็นค่าความผิดพลาดที่เซลล์ประสาท k

$w_{kj}^{(5)}$ เป็นค่าน้ำหนักของชั้นเอาต์พุต

$y_j^{(4)}$ เป็นข้อมูลที่ชั้นพูลลิงชั้นที่ 2

ϕ' เป็นฟังก์ชันกระตุ้น

จากนั้นนำมาหาค่าน้ำหนักที่เปลี่ยนแปลงจากสมการที่ 2.28

$$\Delta w_{kj}^{(5)} = -\lambda \delta_k^{(5)} y_j^{(4)} \quad (2.28)$$

โดยที่ $\Delta w_{kj}^{(5)}$ ค่าน้ำหนักที่เปลี่ยนแปลงของชั้นเอาต์พุต

λ เป็นอัตราการเรียนรู้

$y_j^{(4)}$ เป็นข้อมูลที่ชั้นพูลลิงชั้นที่ 2

ในขั้นตอนการเรียนรู้เป็นการเรียนรู้แบบคำนวณซ้ำ โดยในแต่ละรอบของการเรียนรู้ นั้น จะปรับค่าน้ำหนักดังสมการที่ 2.29

$$w_{kj}^{(5)}(p+1) = w_{kj}^{(5)}(p) + \Delta w_{kj}^{(5)} \quad (2.29)$$

โดยที่ $w_{kj}^{(5)}(p)$ เป็นค่าน้ำหนักจากเซลล์ประสาท j ถึงเซลล์ประสาท k ในชั้นเอาต์พุต ในรอบการคำนวณ p

$w_{kj}^{(5)}(p+1)$ เป็นค่าน้ำหนักจากเซลล์ประสาท j ถึงเซลล์ประสาท k ในชั้นเอาต์พุต ในรอบการคำนวณ $p+1$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) หาค่าเกรเดียนต์ที่ชั้นพูลลิงชั้นที่ 2

$$\delta_j^{(4)}(x, y) = \sum_{k=1}^K \sum_{(x,y)} \delta_k^{(5)} w_{kj}^{(5)}(x, y) \quad (2.30)$$

โดยที่ $\delta_j^{(4)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นพูลลิงชั้นที่ 2

K เป็นจำนวนเซลล์ประสาทที่ชั้นเอาท์พุท

$\delta_k^{(5)}$ เป็นค่าเกรเดียนต์ที่ชั้นเอาท์พุท

$w_{kj}^{(5)}(x, y)$ เป็นค่าน้ำหนักที่ชั้นเอาท์พุท

จากนั้นนำมาหาค่าน้ำหนักที่เปลี่ยนแปลงจากสมการที่ 2.31

$$\Delta w_{ji}^{(4)} = -\lambda \sum_{(x,y)} \delta_j^{(4)}(x, y) \sum_{m=1}^2 \sum_{n=1}^2 y_i^3(2x+m, 2y+n) \quad (2.31)$$

โดยที่ $\Delta w_{ji}^{(4)}$ ค่าน้ำหนักที่เปลี่ยนแปลงของชั้นพูลลิงชั้นที่ 2

λ เป็นอัตราการเรียนรู้

$y_i^3(2x+m, 2y+n)$ เป็นข้อมูลที่ชั้นคอนโวลูชันชั้นที่ 2

ในขั้นตอนการเรียนรู้เป็นการเรียนรู้แบบคำนวณซ้ำ โดยในแต่ละรอบของการเรียนรู้ นั้น จะปรับค่าน้ำหนักดังสมการที่ 2.32

$$w_{ji}^{(4)}(p+1) = w_{ji}^{(4)}(p) + \Delta w_{ji}^{(4)} \quad (2.32)$$

โดยที่ $w_{ji}^{(4)}(p)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นพูลลิงชั้นที่ 2 ในรอบการคำนวณ p

$w_{ji}^{(4)}(p+1)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นพูลลิงชั้นที่ 2 ในรอบการคำนวณ $p+1$

8) หาค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 2

$$\delta_j^{(3)}(x, y) = \sum_{k \in K_s} \delta_k^4(\lfloor x/2 \rfloor, \lfloor y/2 \rfloor) \cdot w_{kj}^4 \quad (2.33)$$

โดยที่ $\delta_j^{(3)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 2

K_s เป็นแผนภาพคุณลักษณะเด่นลักษณะจากการทำพูลลิ่งชั้นที่ 2

$\delta_k^4(\lfloor x/2 \rfloor, \lfloor y/2 \rfloor)$ เป็นค่าเกรเดียนต์ที่ชั้นพูลลิ่งชั้นที่ 2

w_{kj}^4 เป็นค่าน้ำหนักที่ชั้นพูลลิ่งชั้นที่ 2

จากนั้นนำมาหาค่าน้ำหนักที่เปลี่ยนแปลงจากสมการที่ 2.34

$$\Delta w_{ji}^{(3)}(u, v) = -\lambda \sum_{(x, y)} \left(\delta_j^{(3)}(x, y) y_i^2(x+u, u+v) \right) \quad (2.34)$$

โดยที่ $\Delta w_{ji}^{(3)}(u, v)$ ค่าน้ำหนักที่เปลี่ยนแปลงของชั้นคอนโวลูชันชั้นที่ 2 โดยที่ (u, v) เป็นขนาดของหน้าต่างคอนโวลูชันชั้นที่ 2

λ เป็นอัตราการเรียนรู้

$\delta_j^{(3)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 2

$y_i^2(x+u, u+v)$ เป็นข้อมูลที่ชั้นพูลลิ่งชั้นที่ 1

ในขั้นตอนการเรียนรู้เป็นการเรียนรู้แบบคำนวณซ้ำ โดยในแต่ละรอบของการเรียนรู้นั้น จะปรับค่าน้ำหนักดังสมการที่ 2.35

$$w_{ji}^{(3)}(p+1) = w_{ji}^{(3)}(p) + \Delta w_{ji}^{(3)} \quad (2.35)$$

โดยที่ $w_{ji}^{(3)}(p)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นคอนโวลูชันชั้นที่ 2 ในรอบการคำนวณ p

$w_{ji}^{(3)}(p+1)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นคอนโวลูชันชั้นที่ 2 ในรอบการคำนวณ $p+1$

9) หาค่าเกรเดียนต์ที่ชั้นพูลลิงชั้นที่ 1

$$\delta_j^{(2)}(x, y) = \sum_{k \in K_c} \sum_{(u, v)} \delta_k^{(3)}(x, y) \cdot w_{kj}^3(u, v) \quad (2.36)$$

โดยที่ $\delta_j^{(2)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นพูลลิงชั้นที่ 1

K_c เป็นแผนภาพคุณลักษณะเด่นที่ได้จากการทำคอนโวลูชันชั้นที่ 2

$\delta_k^{(3)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 2

$w_{kj}^3(u, v)$ เป็นค่าน้ำหนักที่ชั้นคอนโวลูชันชั้นที่ 2

จากนั้นนำมาหาค่าน้ำหนักที่เปลี่ยนแปลงจากสมการที่ 2.37

$$\Delta w_{ji}^{(2)}(u, v) = -\lambda \sum_{(x, y)} \delta_j^{(2)}(x, y) \sum_{m=1}^2 \sum_{n=1}^2 y_i^1(2x+m, 2y+n) \quad (2.37)$$

โดยที่ $\Delta w_{ji}^{(2)}(u, v)$ ค่าน้ำหนักที่เปลี่ยนแปลงของชั้นพูลลิงชั้นที่ 1

λ เป็นอัตราการเรียนรู้

$\delta_j^{(2)}(x, y)$ เป็นชั้นเกรเดียนต์ที่ชั้นพูลลิงชั้นที่ 1

$y_i^1(2x+m, 2y+n)$ เป็นข้อมูลที่ชั้นคอนโวลูชันชั้นที่ 1

ในขั้นตอนการเรียนรู้เป็นการเรียนรู้แบบคำนวณซ้ำ โดยในแต่ละรอบของการเรียนรู้นั้น จะปรับค่าน้ำหนักดังสมการที่ 2.38

$$w_{ji}^{(2)}(p+1) = w_{ji}^{(2)}(p) + \Delta w_{ji}^{(2)} \quad (2.38)$$

โดยที่ $w_{ji}^{(2)}(p)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นพูลลิงชั้นที่ 1 ในรอบการคำนวณ p

$w_{ji}^{(2)}(p+1)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นพูลลิงชั้นที่ 1 ในรอบการคำนวณ $p+1$

10) หาค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 1

$$\delta_j^{(1)}(x, y) = \sum_{k \in K_s} \delta_k^2(\lfloor x/2 \rfloor, \lfloor y/2 \rfloor) \cdot w_{kj}^2 \quad (2.39)$$

โดยที่ $\delta_j^{(1)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 1

K_s เป็นแผนภาพคุณลักษณะเด่นจากการทำพูลลิงในชั้นที่ 1

$\delta_k^2(\lfloor x/2 \rfloor, \lfloor y/2 \rfloor)$ เป็นค่าเกรเดียนต์ที่ชั้นพูลลิงในชั้นที่ 1

w_{kj}^2 เป็นค่าน้ำหนักที่ชั้นพูลลิงในชั้นที่ 1

จากนั้นนำมาหาค่าน้ำหนักที่เปลี่ยนแปลงจากสมการที่ 2.40

$$\Delta w_{ji}^{(1)}(u, v) = -\lambda \sum_{(x, y)} \left(\delta_j^{(1)}(x, y) y_i^0(x+u, u+v) \right) \quad (2.40)$$

โดยที่ $\Delta w_{ji}^{(1)}(u, v)$ ค่าน้ำหนักที่เปลี่ยนแปลงของชั้นคอนโวลูชัน 1 โดยที่ (u, v) เป็นขนาดของหน้ากากคอนโวลูชัน 1

λ เป็นอัตราการเรียนรู้

$\delta_j^{(1)}(x, y)$ เป็นค่าเกรเดียนต์ที่ชั้นคอนโวลูชันชั้นที่ 1

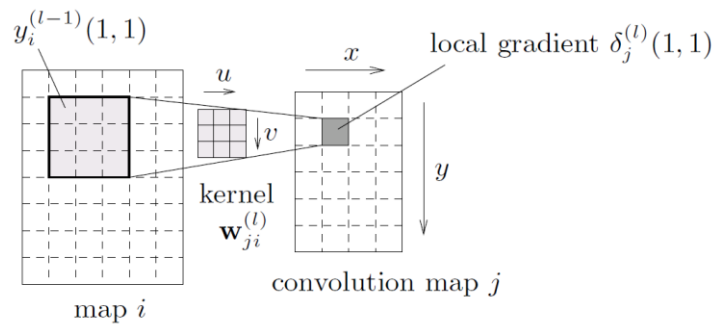
$y_i^0(x+u, u+v)$ เป็นข้อมูลที่อินพุต

ในขั้นตอนการเรียนรู้เป็นการเรียนรู้แบบคำนวณซ้ำ โดยในแต่ละรอบของการเรียนรู้นั้น จะปรับค่าน้ำหนักดังสมการที่ 2.41

$$w_{ji}^{(1)}(p+1) = w_{ji}^{(1)}(p) + \Delta w_{ji}^{(1)} \quad (2.41)$$

โดยที่ $w_{ji}^{(1)}(p)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นคอนโวลูชันชั้นที่ 1 ในรอบการคำนวณ p

$w_{ji}^{(1)}(p+1)$ เป็นค่าน้ำหนักจากเซลล์ประสาท i ถึงเซลล์ประสาท j ในชั้นคอนโวลูชันชั้นที่ 1 ในรอบการคำนวณ $p+1$

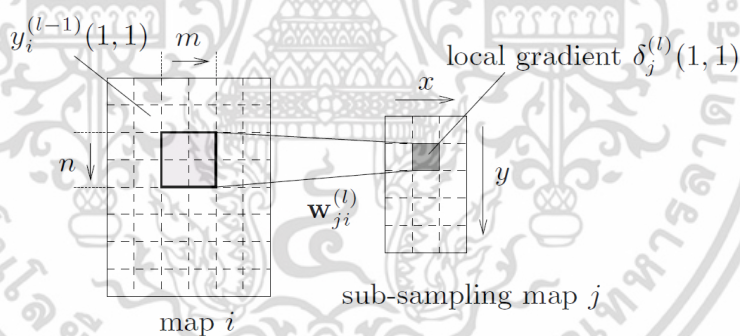


รูปที่ 2.16 การแพร่แบบย้อนกลับด้วยแผนภาพคุณลักษณะเด่นจากการทำคอนโวลูชัน [25]

รูปที่ 2.16 แสดงการแพร่ค่าความผิดพลาดกลับด้วยแผนภาพคุณลักษณะเด่น โดยแผนภาพคุณลักษณะ i คอนโวลูชันกับหน้ากอกที่มีขนาด (u, v) ได้เอาท์พุตเป็นแผนภาพคุณลักษณะเด่นจากการทำคอนโวลูชัน j ที่มีขนาด (x, y)

สำหรับค่าไบแอส $b_j^{(l)}$ ของแผนภาพคุณลักษณะเด่นจากการทำคอนโวลูชัน j ในชั้น l จะถูกปรับได้ตามสมการที่ 2.42

$$\Delta b_j^{(l)} = -\lambda \sum_{(x,y)} (\delta_j^{(l)}(x,y)) \quad (2.42)$$



รูปที่ 2.17 การแพร่แบบย้อนกลับด้วยแผนภาพคุณลักษณะที่ได้จากการทำพูลลิง [25]

จากรูปที่ 2.17 แสดงการแพร่ค่าความผิดพลาดกลับด้วยแผนภาพคุณลักษณะเด่น โดยแผนภาพคุณลักษณะเด่น i ทำพูลลิงด้วยหน้ากอกขนาด (s_x, s_y) ได้เอาท์พุตเป็นแผนภาพคุณลักษณะที่ได้จากการทำพูลลิง j ขนาด (x, y) และการปรับค่าไบแอสแสดงได้ดังสมการที่ 2.42 เช่นเดียวกับคอนโวลูชัน

หากพิจารณาภาพรวมการทำงานของ CNN พบว่าการทำงานคล้ายกับการทำงานของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ แต่แตกต่างกันที่การคำนวณของชั้นคอนโวลูชันและชั้นพูลลิง และเมื่อมีการคำนวณค่าความผิดพลาด จะนำไปปรับทั้งค่าน้ำหนัก และค่าสัมประสิทธิ์ในหน้ากอกคอนโวลูชันด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

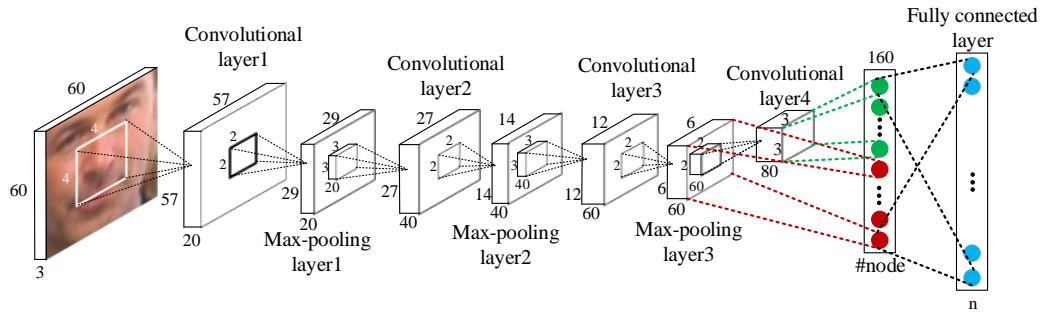
โครงข่ายประสาทเทียมแบบคอนโวลูชันกับระบบรู้จำภาพใบหน้าบุคคลที่น่าเสนอ

ในบทนี้จะกล่าวถึงการออกแบบของระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน การทำงานของระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน และโครงข่ายประสาทเทียมแบบคอนโวลูชันด้วยวิธีการถ่ายโอนการเรียนรู้ จากนั้นจะกล่าวถึงฐานข้อมูลที่ใช้ในการทดลอง และขั้นตอนการทดลอง

3.1 ระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

ระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นวิธีที่สามารถดึงลักษณะเด่นของข้อมูล และการจำแนกข้อมูลได้ในระบบเดียว ซึ่งเป็นลักษณะเด่นของระบบรู้จำภาพใบหน้าบุคคลประเภทนี้ ดังนั้นในวิทยานิพนธ์นี้จึงนำเสนอระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นหลักในการประมวลผล

3.1.1 การออกแบบระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน การออกแบบระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันมีโครงสร้างพื้นฐานมาจากงานวิจัยของ Sun และคณะ [16] เนื่องจากค่าความถูกต้องของระบบที่อ้างอิงนี้ให้ค่าความถูกต้องสูง ซึ่งความแตกต่างกันระหว่างระบบที่อ้างอิงและระบบที่ใช้คือขนาดของภาพอินพุตที่เข้าสู่ระบบ โดยภาพที่ใช้ในวิทยานิพนธ์นี้มีขนาด 60×60 พิกเซล ซึ่งมีขนาดใหญ่กว่าภาพในงานวิจัยของ Sun และคณะ [16] ที่มีภาพขนาด 31×39 พิกเซล ส่งผลให้ขนาดของข้อมูลในแต่ละชั้นแตกต่างกัน โดยโครงสร้าง CNN ที่ใช้ในวิทยานิพนธ์นี้แสดงได้ดังรูปที่ 3.1



รูปที่ 3.1 โครงสร้างของระบบ CNN

จากรูปที่ 3.1 สามารถอธิบายโครงสร้างและการทำงานของ CNN ได้ดังต่อไปนี้

- ภาพอินพุตเป็นภาพสีขนาด 60×60 พิกเซล จึงทำให้มิติในชั้นอินพุตเป็น $60 \times 60 \times 3$ โดยตัวเลขตัวสุดท้ายแสดงถึงจำนวนชั้นของภาพอินพุต เช่นในวิทยานิพนธ์นี้ใช้ภาพสี RGB จะมีจำนวนชั้นเป็น 3 เป็นต้น

- ในชั้นคอนโวลูชันชั้นที่ 1 (Convolutional layer 1) เป็นการนำภาพอินพุตไปคอนโวลูชันกับหน้ากากขนาด 4×4 จำนวน 20 หน้ากาก ซึ่งการออกแบบขนาดของหน้ากากนั้นจะพิจารณาขนาดของภาพอินพุต โดยจะเลือกขนาดของหน้ากากให้มีขนาดเล็กกว่าครึ่งหนึ่งของด้านที่สั้นที่สุดของอินพุต ซึ่งในการสอนครั้งแรกนี้ค่าสัมประสิทธิ์ของหน้ากากได้จากการสุ่ม ทำให้ได้ภาพเอาท์พุตเป็นแผนภาพคุณลักษณะเด่นที่ได้จากการทำคอนโวลูชันขนาด $57 \times 57 \times 20$ โดยสาเหตุที่ขนาดของภาพลดลงจาก 60×60 พิกเซล เหลือ 57×57 พิกเซล เนื่องจากการสูญเสียขอบของภาพไปจากการทำคอนโวลูชัน ซึ่งสมการการทำคอนโวลูชันสามารถแสดงได้ดังสมการที่ 3.1

$$y_j^{(1)}(x, y) = \phi^{(1)} \left(\sum_{i \in I} \sum_{(u, v) \in K} w_{ji}^{(1)}(u, v) y_i^{(0)}(x+u, y+v) + b_j^{(1)} \right) \quad (3.1)$$

โดยที่ $y_i^{(0)}(x+u, y+v)$ เป็นภาพอินพุต

$\phi^{(1)}$ เป็นฟังก์ชันกระตุ้น ซึ่งในวิทยานิพนธ์นี้เลือกใช้แบบ ReLU เนื่องจากลดโอกาสที่บางข้อมูลจะสูญหาย โดยที่หากข้อมูลมีค่ามากกว่า 0 จะเลือกข้อมูลนั้นมาเป็นเอาท์พุต แต่สำหรับกรณีข้อมูลมีค่าน้อยกว่า 0 จะทำให้ค่าเอาท์พุตเป็น 0

$w_{ji}^{(1)}(u, v)$ เป็นค่าสัมประสิทธิ์ในหน้ากากคอนโวลูชัน ซึ่งในชั้นนี้จะมีขนาด 4×4 จำนวน 20 หน้ากาก

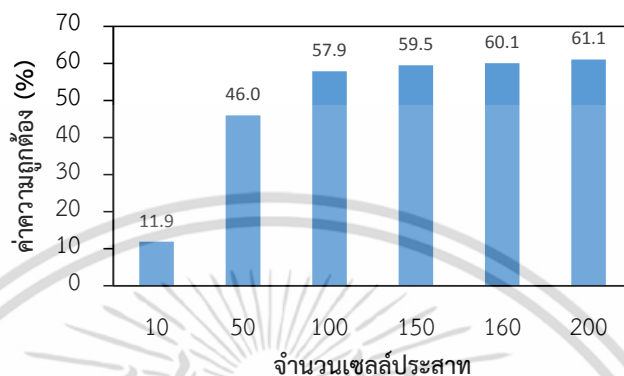
$b_j^{(1)}$ เป็นค่าไบแอส ซึ่งในวิทยานิพนธ์นี้ ค่าไบแอสจะมีค่าเท่ากับ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- จากนั้นเข้าสู่ชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 1 (Max-pooling layer 1) โดยใช้หน้ากากลขนาด 2×2 กับเอาต์พุตที่ได้จากชั้นคอนโวลูชันชั้นที่ 1 เพื่อให้ขนาดของภาพลดลงครึ่งหนึ่ง โดยได้แผนภาพคุณลักษณะเด่นที่ได้จากการทำพูลลิ่งขนาด $29 \times 29 \times 20$ ซึ่งเหตุผลที่เลือกใช้การทำพูลลิ่งด้วยค่าสูงสุดนั้น เพื่อให้เอาต์พุตจากการทำพูลลิ่งยังคงมีลักษณะเด่นไว้
- ต่อมาเข้าสู่ชั้นคอนโวลูชันชั้นที่ 2 (Convolutional layer 2) ด้วยการนำภาพเอาต์พุตจากชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 1 ไปคอนโวลูชันกับหน้ากากลขนาด 3×3 จำนวน 40 หน้ากากล ซึ่งจะได้ผลลัพธ์เป็นแผนภาพคุณลักษณะเด่นที่ได้จากการทำคอนโวลูชันขนาด $27 \times 27 \times 40$
- ต่อมาเป็นชั้นพูลลิ่งด้วยค่าสูงสุด 2 (Max-pooling layer 2) โดยใช้หน้ากากลขนาด 2×2 กับเอาต์พุตที่ได้จากชั้นคอนโวลูชัน 2 ทำให้ได้แผนภาพคุณลักษณะเด่นที่ได้จากการทำพูลลิ่งขนาด $14 \times 14 \times 40$
- เข้าสู่ชั้นคอนโวลูชันชั้นที่ 3 (Convolutional layer 3) ด้วยการนำภาพเอาต์พุตจากชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 2 ไปคอนโวลูชันกับหน้ากากลขนาด 3×3 จำนวน 60 หน้ากากล ได้แผนภาพคุณลักษณะที่ได้จากการทำคอนโวลูชันขนาด $12 \times 12 \times 60$
- ทำพูลลิ่งด้วยค่าสูงสุดชั้นที่ 3 (Max-pooling layer 3) โดยใช้หน้ากากลขนาด 2×2 กับเอาต์พุตที่ได้จากชั้นคอนโวลูชันชั้นที่ 3 ทำให้ได้แผนภาพคุณลักษณะเด่นที่ได้จากการทำพูลลิ่งขนาด $6 \times 6 \times 60$
- ในชั้นคอนโวลูชันชั้นที่ 4 (Convolutional layer 4) จะนำภาพเอาต์พุตจากชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 3 ไปคอนโวลูชันกับหน้ากากลขนาด 2×2 จำนวน 80 หน้ากากล ซึ่งได้ผลลัพธ์เป็นแผนภาพคุณลักษณะเด่นที่ได้จากการทำคอนโวลูชันขนาด $3 \times 3 \times 80$
- จากนั้นจะนำเอาต์พุตที่ได้จากชั้นคอนโวลูชันชั้นที่ 4 และชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 3 มาต่อกันเพื่อนำไปเป็นชั้นจำแนก ซึ่งการนำชั้นคอนโวลูชันชั้นที่ 4 และชั้นพูลลิ่งด้วยค่าสูงสุดชั้นที่ 3 มาต่อกันทำให้ได้ชั้นลักษณะเด่น ในชั้นนี้จะประกอบไปด้วยข้อมูลที่มีความละเอียด 2 ความละเอียด ได้แก่ $3 \times 3 \times 80$ และ $6 \times 6 \times 60$ ตามลำดับ ซึ่งเอาต์พุตที่ได้จากชั้นนี้คือชั้นคุณลักษณะเด่นที่มีจำนวนเซลล์ประสาทเท่ากับ 160 เซลล์ประสาท โดยจะนำข้อมูล $3 \times 3 \times 80$ และ $6 \times 6 \times 60$ มาต่อกัน และทำการเปลี่ยนขนาดให้เป็น 160 เซลล์ประสาท ซึ่งมีค่าเท่ากับงานวิจัยของ Sun และคณะ [16] แต่อย่างไรก็ตามเพื่อให้มั่นใจว่าจำนวนเซลล์ประสาทที่เลือกใช้มีความเหมาะสมกับระบบที่ใช้ในวิทยานิพนธ์นี้ ดังนั้นในเบื้องต้นจึงศึกษาอิทธิพลของจำนวนเซลล์ประสาทต่อค่าความถูกต้องเพื่อพิสูจน์ว่าจำนวนเซลล์ประสาทที่เลือกใช้มีความเหมาะสมต่อระบบที่นำเสนอจริง ในการศึกษาเบื้องต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นี้จะทดสอบเฉพาะกับฐานข้อมูลขนาดใหญ่ โดยกำหนดจำนวนเซลล์ประสาทเป็น 10 50 100 150 160 และ 200 ซึ่งค่าความถูกต้องที่ได้จากจำนวนเซลล์ประสาทที่ต่างกันสามารถแสดงได้ดังรูปที่ 3.2 นอกจากนี้รายละเอียดของมิติในแต่ละชั้นของโครงสร้างที่ใช้ในวิทยานิพนธ์นี้สามารถสรุปได้ดังตารางที่ 3.1



รูปที่ 3.2 ค่าความถูกต้องสำหรับจำนวนเซลล์ประสาทที่ต่างกัน

จากรูปที่ 3.2 พบว่าค่าความถูกต้องเพิ่มขึ้นเมื่อจำนวนเซลล์ประสาทเพิ่มขึ้น แต่อย่างไรก็ตามค่าความถูกต้องที่ได้จากจำนวนเซลล์ประสาทมากกว่า 100 มีค่าเพิ่มขึ้นเพียงเล็กน้อย หรือกล่าวอีกนัยหนึ่งว่าค่าความถูกต้องที่ได้จากจำนวนเซลล์ประสาทมากกว่า 100 มีค่าเพิ่มขึ้นอย่างไม่มีนัยสำคัญ ดังนั้นจึงสามารถสรุปได้ว่าจำนวนเซลล์ประสาทที่เลือกใช้ (160 เซลล์ประสาท) สามารถนำไปใช้กับการทดลองอื่นได้โดยจะไม่ส่งผลต่อค่าความถูกต้องที่ได้

ตารางที่ 3.1 รายละเอียดแต่ละชั้นของโครงสร้าง

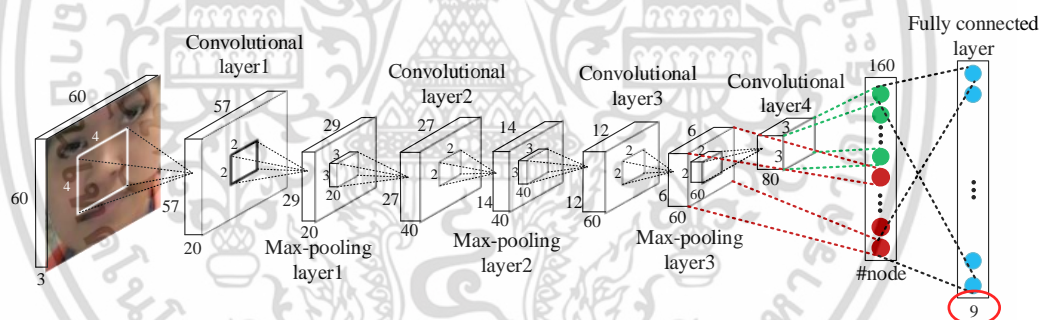
ชั้น	มิติ
Training data	60×60×3
Convolution layer1	57×57×20
Max-pooling layer1	29×29×20
Convolution layer2	27×27×40
Max-pooling layer2	14×14×40
Convolution layer3	12×12×60
Max-pooling layer3	6×6×60
Convolution layer4	3×3×80
#node	160
Output	n

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

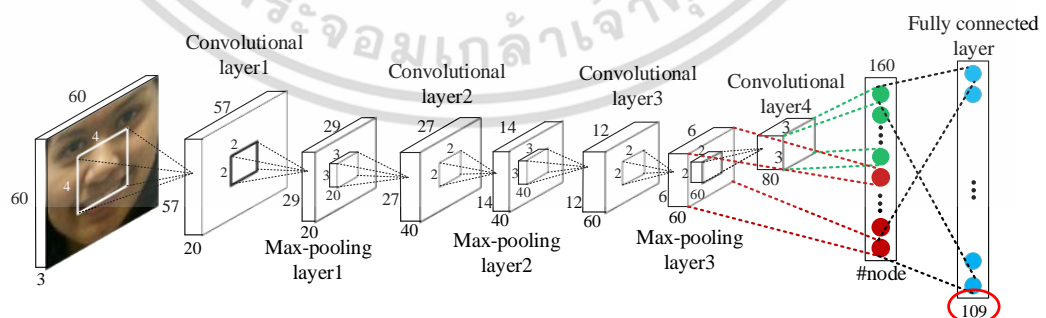
ตารางที่ 3.1 อธิบายมิติของข้อมูลที่อยู่ในแต่ละชั้นของโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้ในวิทยานิพนธ์นี้ โดยขนาดของหน้ากากในแต่ละชั้นจะส่งผลถึงขนาดของเอาต์พุตที่ได้ในแต่ละชั้นด้วย และจำนวนชั้นคอนโวลูชันไม่สามารถเพิ่มจากนี้ได้อีก เนื่องจากหากเพิ่มอีกชั้นเอาต์พุตที่ได้จะไม่มีลักษณะเด่นใดๆหลงเหลืออีก ดังนั้นในวิทยานิพนธ์นี้จึงใช้จำนวนชั้นคอนโวลูชัน 4 ชั้น และสำหรับชั้นเอาต์พุตจะมีจำนวนมิติของข้อมูลเป็น n คือ ตามจำนวนบุคคลในฐานข้อมูลนั้นๆ เช่น ถ้าฐานข้อมูลขนาด 9 คน จะต้องใช้ $n=9$ เป็นต้น

3.1.2 การทำงานของ CNN

การทำงานของ CNN ในวิทยานิพนธ์นี้จะใช้โครงสร้างในรูปที่ 3.1 ซึ่งได้มีการอธิบายส่วนต่างๆไปแล้วในหัวข้อก่อนหน้า โดยจะมีการปรับเปลี่ยนจำนวนเอาต์พุตให้เหมาะสมกับขนาดของฐานข้อมูลที่ใช้ ซึ่งจะทดลอง CNN กับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยตรง โครงสร้าง CNN เมื่อนำไปใช้กับฐานข้อมูล 9 คน แสดงได้ดังรูปที่ 3.3 และเมื่อนำไปใช้กับฐานข้อมูล 109 คน โครงสร้าง CNN ที่ใช้จะแสดงได้ดังรูปที่ 3.4



รูปที่ 3.3 โครงสร้าง CNN สำหรับฐานข้อมูล 9 คน



รูปที่ 3.4 โครงสร้าง CNN สำหรับฐานข้อมูล 109 คน

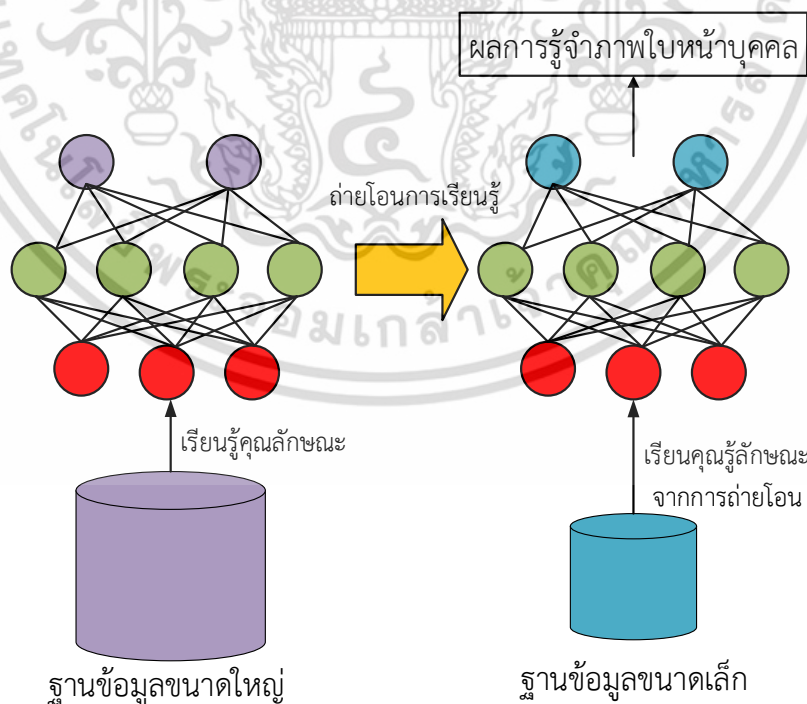
รูปที่ 3.3 และ 3.4 จะเห็นว่าโครงสร้าง CNN มีส่วนที่เหมือนกันดังต่อไปนี้ ภาพอินพุตที่ใช้เป็นภาพสีขนาด 60×60 พิกเซล จำนวนชั้นคอนโวลูชัน 4 ชั้น พูลถึง 3 ชั้น และชั้นลักษณะเด่นจะใช้เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอาต์พุตจากชั้นคอนโวลูชันชั้นที่ 4 และชั้นพูลลิ่งชั้นที่ 3 มาต่อกัน โดยขนาดหน้ากากในแต่ละชั้นก็ใช้จำนวนเท่ากันด้วย แต่มีส่วนที่แตกต่างกันคือจำนวนเซลล์ประสาทที่ชั้นเอาต์พุตซึ่งปรับตามจำนวนบุคคลในฐานข้อมูล เช่นกรณีนำไปใช้กับฐานข้อมูล 9 คน ชั้นเอาต์พุตจะใช้ 9 เซลล์ประสาท และหากนำไปใช้กับฐานข้อมูล 109 คน ชั้นเอาต์พุตจะมี 109 เซลล์ประสาท

ซึ่งโดยทั่วไปแล้ว CNN จะถูกนำมาใช้กับฐานข้อมูลขนาดใหญ่ แต่ในวิทยานิพนธ์นี้สนใจการใช้งาน CNN กับฐานข้อมูลขนาดเล็ก ทำให้การทำงานของ CNN เป็นไปอย่างไม่เต็มประสิทธิภาพ ดังนั้นในวิทยานิพนธ์นี้จึงนำวิธี CNN แบบถ่ายโอนการเรียนรู้ (Transfer learning CNN: TL-CNN) มาประยุกต์ใช้งาน ซึ่งจะอธิบายในส่วนถัดไป

3.1.3 CNN แบบถ่ายโอนการเรียนรู้ (Transfer Learning CNN: TL-CNN)

โดยปกติแล้วโครงข่ายประสาทเทียมแบบคอนโวลูชันจะเหมาะกับการนำไปใช้กับฐานข้อมูลขนาดใหญ่ซึ่งจะทำให้เกิดการเรียนรู้ลักษณะได้ดี แต่หากนำมาใช้กับฐานข้อมูลขนาดเล็กอาจให้ประสิทธิภาพที่ไม่ดีเท่ากับการใช้กับฐานข้อมูลขนาดใหญ่ เนื่องจากจำนวนข้อมูลที่ใช้ในการเรียนรู้มีไม่มากพอที่จะทำให้โครงสร้าง CNN เกิดการเรียนรู้คุณลักษณะได้ จึงมีการนำ CNN แบบถ่ายโอนการเรียนรู้มาประยุกต์ใช้ [26] ซึ่งการทำงานของวิธีนี้สามารถแสดงได้ดังรูปที่ 3.5

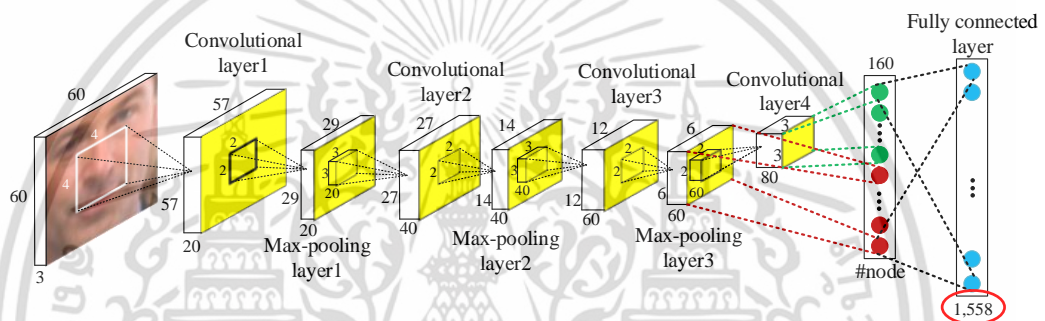


รูปที่ 3.5 CNN แบบถ่ายโอนการเรียนรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.5 การทำงานของ CNN แบบถ่ายโอนการเรียนรู้มีขั้นตอนการทำงานเริ่มต้นจากการสอนด้วยฐานข้อมูลขนาดใหญ่ก่อน เพื่อให้ระบบสามารถเรียนรู้ลักษณะของฐานข้อมูลขนาดใหญ่ จากนั้นจะเปลี่ยนชั้นจำแนกและฐานข้อมูลเป็นฐานข้อมูลขนาดเล็ก แล้วทำการสอนอีกครั้ง ในการสอนครั้งนี้การเรียนรู้ลักษณะจะเริ่มต้นจากลักษณะที่ได้จากการสอนด้วยฐานข้อมูลขนาดใหญ่ ซึ่งโครงสร้างที่ใช้สอนกับฐานข้อมูลขนาดใหญ่และฐานข้อมูลขนาดเล็กจะต้องใช้โครงสร้างแบบเดียวกัน ทำให้ผลสุดท้ายในการทำงานของระบบนี้ดีเทียบเท่ากับการใช้ฐานข้อมูลขนาดใหญ่ แม้ว่าในความจริงนั้นจะใช้เพียงฐานข้อมูลขนาดเล็กเท่านั้น

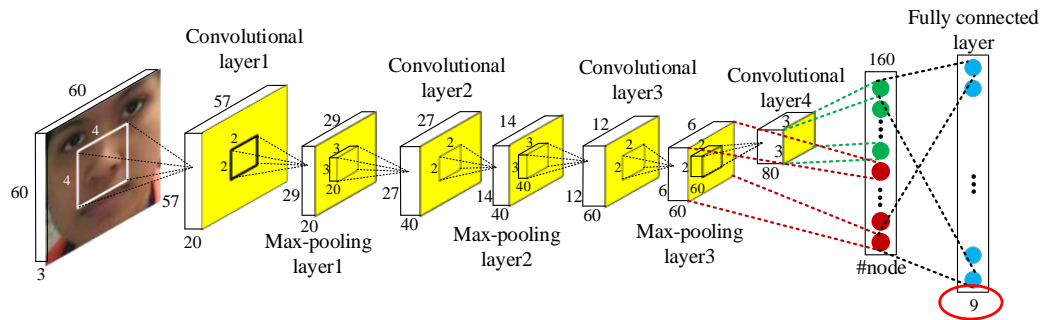
ขั้นตอนแรกคือการใช้ CNN กับฐานข้อมูลขนาดใหญ่ก่อน เพื่อให้โครงสร้าง CNN เกิดการเรียนรู้คุณลักษณะจากฐานข้อมูลขนาดใหญ่ โดยในการทดลองจะใช้โครงสร้างในรูปที่ 3.6



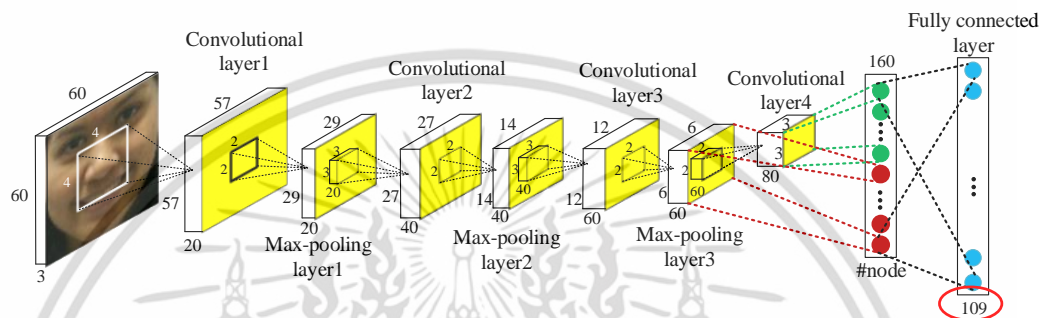
รูปที่ 3.6 โครงสร้าง CNN กับฐานข้อมูลขนาดใหญ่

รูปที่ 3.6 เป็นการนำ CNN ไปใช้กับฐานข้อมูลขนาดใหญ่ ซึ่งโครงสร้างจะเหมือนกับโครงสร้าง CNN ที่ใช้ในรูปที่ 3.1 แต่มีการปรับจำนวนเซลล์ประสาทในชั้นเอาต์พุตเป็น 1,558 เซลล์ประสาท ซึ่งจะสอดคล้องกับจำนวนบุคคลในฐานข้อมูลขนาดใหญ่ที่มี 1,558 คน โดยการใช้ CNN กับฐานข้อมูลใหญ่นั้นจะทำให้เกิดการเรียนรู้คุณลักษณะได้ เนื่องจากข้อมูลมีจำนวนมากพอที่ทำให้โครงสร้าง CNN เกิดการเรียนรู้ได้

จากนั้นจะเปลี่ยนชั้นจำแนก และฐานข้อมูลที่ใช้ เป็นสำหรับการนำไปใช้กับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยลักษณะที่ได้จากการเรียนรู้ครั้งแรกนั้นจะถูกนำไปใช้เป็นค่าคุณลักษณะเริ่มต้น (ค่าน้ำหนักและค่าสัมประสิทธิ์ในหน้ากาคอนโวลูชันเริ่มต้น) ในการเรียนรู้ครั้งใหม่ โดยโครงสร้าง CNN ที่จะนำไปใช้กับฐานข้อมูลขนาดเล็ก (9 และ 109 คน) จะแสดงได้ดังรูปที่ 3.7 และ 3.8 ตามลำดับ



รูปที่ 3.7 โครงสร้าง TL-CNN สำหรับฐานข้อมูล 9 คน



รูปที่ 3.8 โครงสร้าง TL-CNN สำหรับฐานข้อมูล 109 คน

รูปที่ 3.7 และ 3.8 แสดงโครงสร้าง TL-CNN สำหรับฐานข้อมูล 9 คน และ 109 คน ตามลำดับ ซึ่งจะเห็นว่าโครงสร้างเหมือนกันการใช้ CNN กับฐานข้อมูลขนาดเล็กโดยตรงในรูปที่ 3.3 และ 3.4 แต่จะแตกต่างกันตรงที่ค่าลักษณะเริ่มต้นจะใช้ค่าลักษณะที่ได้จากการสอน CNN กับฐานข้อมูลขนาดใหญ่ ซึ่งจะได้โครงสร้างที่มีการเรียนรู้คุณลักษณะไปแล้ว จากนั้นเมื่อนำมาใช้กับฐานข้อมูล 9 คนและ 109 คน ก็จะเกิดการเรียนรู้คุณลักษณะจากฐานข้อมูล 9 คนและ 109 คน อีกครั้ง โดยในการเรียนรู้คุณลักษณะครั้งนี้เพื่อนำไปใช้ในการจำแนกบุคคล

3.2 ฐานข้อมูลที่ใช้การทดลอง

ฐานข้อมูลที่ใช้ในการทดลองประกอบด้วยฐานข้อมูล 3 ฐานข้อมูล ได้แก่ ฐานข้อมูลขนาดเล็ก จำนวน 2 ฐานข้อมูล (ฐานข้อมูล 9 คน และฐานข้อมูล 109 คน) และฐานข้อมูลขนาดใหญ่ (1,558 คน) ซึ่งฐานข้อมูลขนาดเล็กนี้เป็นฐานข้อมูลที่จัดเตรียมขึ้นเอง สำหรับภาพที่ใช้ในการทดลองทั้งหมดนี้เป็นภาพที่ตัดเฉพาะบริเวณที่เป็นภาพใบหน้า ตัวอย่างภาพใบหน้าแสดงได้ดังรูปที่ 3.9



รูปที่ 3.9 ตัวอย่างภาพใบหน้าที่ใช้ในการทดลอง

จากรูปที่ 3.9 เป็นภาพที่ตัดเอาเฉพาะส่วนของใบหน้าเพื่อนำไปใช้ในการทดลอง พบว่าภาพใบหน้านี้จะประกอบด้วย ดวงตา 2 ข้าง จมูก และปาก ซึ่งสามารถสังเกตได้อย่างชัดเจน รายละเอียดของฐานข้อมูลทั้ง 3 ฐานข้อมูลสามารถอธิบายได้ดังต่อไปนี้

3.2.1 ฐานข้อมูลภาพใบหน้าขนาดเล็ก 9 คน

ฐานข้อมูลชุดนี้เป็นฐานข้อมูลที่จัดเตรียมขึ้นจากสมาชิกห้องปฏิบัติการวิจัยการออกแบบระบบประมวลผลสัญญาณดิจิทัล ห้อง T-217 ศึกษาคณะวิศวกรรมโทรคมนาคม โดยมีสมาชิกทั้งหมด 9 คน แต่ละคนมีการจัดเก็บมากกว่า 1 ภาพ ภาพในฐานข้อมูลนี้มี 542 ภาพ ซึ่งภาพในฐานข้อมูลนี้ได้จากกล้องดิจิทัลที่ติดอยู่หน้าห้อง T-217 การจัดเก็บฐานข้อมูลจัดเตรียมขึ้นระหว่างวันที่ 16 พฤศจิกายน 2558 ถึง 15 มกราคม 2559 โดยตัวอย่างภาพถ่ายจากฐานข้อมูล 9 คน แสดงได้ดังรูปที่ 3.10

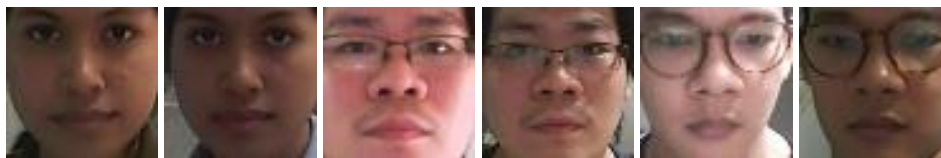


รูปที่ 3.10 ตัวอย่างภาพถ่ายจากฐานข้อมูล 9 คน

รูปที่ 3.10 เป็นภาพที่ได้จากการเก็บข้อมูลจากกล้องดิจิทัลที่ติดหน้าห้องในเวลาที่แตกต่างกัน จะเห็นว่าทั้ง 2 ภาพมีพื้นหลังเหมือนกัน แต่มีแสงที่ตกกระทบบนภาพใบหน้าที่แตกต่างกัน กล่าวคือภาพด้านขวาจะมีความสว่างมากกว่าภาพด้านซ้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนของภาพที่ใช้ประมวลผลนั้นจะผ่านกระบวนการตรวจจับภาพใบหน้าและการตัดภาพ ใบหน้าเฉพาะส่วนที่ต้องการ ตัวอย่างภาพใบหน้าฐานข้อมูล 9 คนสามารถแสดงได้ดังรูปที่ 3.11



รูปที่ 3.11 ตัวอย่างภาพฐานข้อมูลภาพใบหน้า 9 คน

รูปที่ 3.11 แสดงภาพตัวอย่างของบุคคล 3 คน โดยภาพที่นำมาแสดงจะมีคนละ 2 ภาพ พบว่าทุกภาพเป็นภาพถ่ายใบหน้าตรง สามารถมองเห็นดวงตา จมูก และปากได้อย่างชัดเจน จากภาพ ทั้ง 6 ภาพนี้จะเห็นว่าสภาพแสงในแต่ละภาพนั้นแตกต่างกัน แม้จะเป็นบุคคลคนเดียวกัน กล่าวคือ ภาพถ่ายในวันและเวลาที่แตกต่างกัน จะได้ภาพที่มีแสงตกกระทบบนภาพใบหน้าที่แตกต่างกัน

3.2.2 ฐานข้อมูลภาพใบหน้าขนาดเล็ก 109 คน

ฐานข้อมูลนี้เป็นฐานข้อมูลที่จัดเตรียมขึ้นมาจากนักศึกษาคณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ประกอบด้วยบุคคล 109 คน ซึ่งการจัดเก็บ ฐานข้อมูลนั้นเป็นการถ่ายภาพด้วยกล้องดิจิทัล การถ่ายภาพไม่มีการควบคุมสภาพแวดล้อม แสง และสถานที่ถ่ายภาพ โดยระยะเวลาที่ใช้ในการจัดเก็บฐานข้อมูลคือตั้งแต่วันที่ 20 พฤศจิกายน ถึง 11 ธันวาคม 2556 มีภาพในฐานข้อมูลทั้งหมด 874 ภาพ ตัวอย่างภาพในฐานข้อมูลภาพใบหน้า 109 คน แสดงได้ดังรูปที่ 3.12



รูปที่ 3.12 ตัวอย่างภาพในฐานข้อมูลภาพใบหน้า 109 คน

รูปที่ 3.12 เป็นภาพตัวอย่างจากฐานข้อมูล 109 คน จะเห็นว่าพื้นหลังในการถ่ายภาพทั้ง 2 ภาพนี้แตกต่างกัน สภาพแสงที่ตกกระทบบนภาพใบหน้าที่แตกต่างกันด้วย นอกจากนี้ภาพใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฐานข้อมูลนี้ยังมีอิริยาบถของภาพใบหน้าที่แตกต่างกันดังแสดงในรูปที่ 3.13 โดยภาพที่แสดงนี้เป็นภาพที่ผ่านขั้นตอนการตรวจจับภาพใบหน้าแล้ว



รูปที่ 3.13 ภาพใบหน้าในอิริยาบถต่างๆ

รูปที่ 3.13 เป็นภาพที่แสดงอิริยาบถที่แตกต่างกันของใบหน้าในฐานข้อมูล 109 คน เรียงจากซ้ายไปขวาประกอบด้วย ภาพใบหน้าปกติ ขมวดคิ้ว ยิ้ม หลับตา อ้าปาก และ แลบลิ้น ตามลำดับ ซึ่งภาพใบหน้าของแต่ละบุคคลจะมีอิริยาบถเช่นนี้เหมือนกัน

3.2.3 ฐานข้อมูลขนาดใหญ่ (Celebrity face)

ฐานข้อมูลขนาดใหญ่ หรือ MSRA-CFW: Data Set of Celebrity Faces on the Web [27-28] เรียกอย่างย่อว่า Celebrity face เป็นฐานข้อมูลที่รวบรวมภาพใบหน้าบุคคลที่มีชื่อเสียงทั่วโลกจากเว็บไซต์ต่างๆ เช่น Angelina Joie, Bill Gates และ Paris Hilton เป็นต้น ซึ่งเป็นฐานข้อมูลขนาดใหญ่ที่ถูกใช้อย่างแพร่หลาย [16, 18] และเป็นที่ยอมรับ ภาพที่ปรากฏในฐานข้อมูลนี้เป็นภาพที่มีความหลากหลายทั้งอิริยาบถ สถานที่ในการถ่ายทำ และสภาพแวดล้อม ทำให้ฐานข้อมูลนี้มีความหลากหลายและเป็นธรรมชาติมากที่สุด โดยฐานข้อมูลจะมีไฟล์เดอร์ของแต่ละบุคคลโดยในแต่ละไฟล์เดอร์นั้นประกอบด้วย

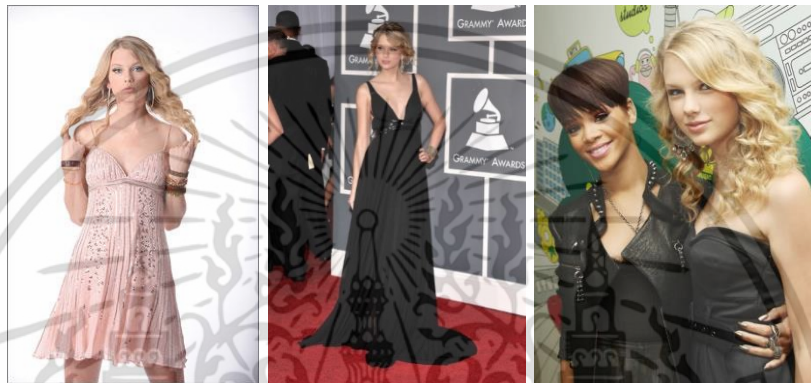
- 1) ภาพตัวอย่างขนาดเล็กของบุคคลนั้น
- 2) ไฟล์ข้อมูลของภาพใบหน้า (info.txt) ซึ่งจะแสดงลิงค์ของภาพต้นฉบับ หรือ

Original web images (OWI) ของภาพตัวอย่าง

ภาพที่ใช้ในการทดลองนี้เป็นภาพสี่ที่ตัดเฉพาะส่วนของภาพใบหน้าตัวอย่าง โดยตัวอย่างภาพในฐานข้อมูลนี้แสดงได้ดังรูปที่ 3.14 เป็นภาพของ Bill Gate และรูปที่ 3.15 เป็นภาพของ Taylor Swift



รูปที่ 3.14 ภาพตัวอย่างของ Bill Gate



รูปที่ 3.15 ภาพตัวอย่างของ Taylor Swift

จากรูปที่ 3.14 จะเห็นว่าภาพของ Bill Gate ในฐานะข้อมูลนี้มีหลายอิริยาบถ และเป็นภาพที่ถ่ายแบบใกล้เคียงใบหน้า และถ่ายแบบเต็มตัว นอกจากนี้ยังถ่ายในเหตุการณ์ที่ต่างกันอย่างอื่นด้วย และในรูปที่ 3.15 ภาพของ Taylor Swift เป็นภาพที่เห็นเต็มตัว ทำให้ภาพใบหน้าที่ได้มีขนาดเล็กลง และบางภาพยังมีบุคคลมากกว่า 1 บุคคล โดยภาพที่มีใบหน้าขนาดเล็ก หรือภาพที่มีบุคคลมากกว่า 1 บุคคล จะทำให้ไม่สามารถตรวจจับภาพใบหน้าได้

ในฐานะข้อมูลนี้มีภาพทั้งหมด 202,792 ภาพ แต่เมื่อถึงขั้นตอนการตรวจจับภาพใบหน้า มีหลายภาพที่ไม่สามารถตรวจจับได้ ซึ่งมีสาเหตุมาจาก ภาพมีขนาดเล็กไป ไม่สามารถหาบริเวณใบหน้าตา หรือปากได้ หรือแม้กระทั่งการโหลดภาพจากลิงค์ภาพที่ได้มา บางลิงค์เป็นลิงค์ที่หมดอายุไม่สามารถโหลดภาพได้ ทำให้มีภาพจำนวนมากที่ต้องสูญเสียไปในขั้นตอนนี้ สุดท้ายแล้วภาพที่พร้อมสำหรับการนำไปทดลองในส่วนถัดไปมีทั้งหมด 103,182 ภาพ จากบุคคลทั้งหมด 1,558 คน ซึ่งแต่ละคนมีอย่างน้อยคนละ 1 ภาพ

จากรฐานข้อมูลทั้ง 3 ฐานข้อมูลที่ได้กล่าวข้างต้นสามารถสรุปจำนวนภาพใบหน้าและจำนวนคนในฐานข้อมูลได้ดังตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ปริมาณข้อมูลในฐานข้อมูล

ชื่อฐานข้อมูล	จำนวนภาพทั้งหมด	จำนวนคนทั้งหมด
ฐานข้อมูลขนาดเล็ก 9 คน	542	9
ฐานข้อมูลขนาดเล็ก 109 คน	874	109
ฐานข้อมูลขนาดใหญ่	103,182	1,558

3.3 ขั้นตอนการทดลอง

ในวิทยานิพนธ์นี้ศึกษาการนำระบบรู้จำภาพใบหน้าบุคคลไปประยุกต์ใช้งานกับฐานข้อมูลขนาดเล็ก โดยในขั้นตอนการทดลองนั้นมีลำดับดังต่อไปนี้

- 1) ระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งจะทดสอบใน 2 ลักษณะคือ การใช้ CNN กับฐานข้อมูลขนาดเล็กโดยตรง และการใช้ CNN แบบถ่ายโอนการเรียนรู้
- 2) ระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป ซึ่งในงานวิจัยนี้จะทดลองด้วยการใช้ PCA ในการดึงลักษณะเด่น จากนั้นใช้ kNN และ SVM ในการจำแนกภาพใบหน้าบุคคล
- 3) การเปรียบเทียบวิธีการรู้จำภาพใบหน้าบุคคลโดยทั่วไป กับโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โดยในการทดลองจะใช้ฐานข้อมูลขนาดเล็ก 2 ฐานข้อมูลได้แก่ ฐานข้อมูล 9 คน และ 109 คน ในการทดลองหาประสิทธิภาพของระบบ และใช้ฐานข้อมูลขนาดใหญ่ในการดึงลักษณะเด่นของข้อมูลเท่านั้น ไม่ได้ใช้เพื่อการทดสอบวัดประสิทธิภาพ

บทที่ 4

การทดลองและการวิเคราะห์ผลการทดลอง

ในบทนี้กล่าวถึงการทดลองระบบรู้จำภาพใบหน้าบุคคลด้วย CNN และระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป ตามขั้นตอนการทดลองที่ได้กล่าวไปในบทที่ 3 โดยในส่วนแรกนั้นจะใช้ CNN กับฐานข้อมูลโดยตรง การทดลองด้วย CNN แบบถ่ายโอนการเรียนรู้ จากนั้นจึงทดสอบโดยใช้ PCA ในการดึงลักษณะเด่น และใช้ kNN และ SVM ในการจำแนก การทดลองทั้งหมดจะศึกษาเกี่ยวกับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) และสุดท้ายเป็นการเปรียบเทียบผลการทดลองของระบบรู้จำภาพใบหน้าบุคคลที่ทำการศึกษาทั้งหมดนอกจากนี้ยังกล่าวถึงการประยุกต์ใช้งานระบบรู้จำภาพใบหน้าบุคคลซึ่งในวิทยานิพนธ์นี้ยกตัวอย่างการประยุกต์ใช้งาน 2 ตัวอย่าง ประกอบด้วย ระบบเข้าประตูโดยใช้ภาพใบหน้า และแอปพลิเคชันยืนยันตัวตนบุคคลบนโทรศัพท์สมาร์ทโฟน

4.1 การจัดการฐานข้อมูลในการทดลอง

เนื่องจากการทดลองนี้ใช้ฐานข้อมูลทั้งหมด 3 ชุดด้วยกันได้แก่ ฐานข้อมูลขนาดใหญ่ (Celebrity face) ฐานข้อมูลขนาดเล็ก 9 คน และ 109 คน ตามลำดับ จึงจำเป็นต้องมีการกำหนดขนาดของฐานข้อมูลที่ใช้ในการสอนและทดสอบ โดยการเลือกจำนวนภาพที่ใช้ในการสอนและทดสอบนั้น เลือกโดย 80% ของจำนวนภาพทั้งหมดของแต่ละคน จะเป็นภาพที่ใช้ในการสอน และอีก 20% ของจำนวนภาพแต่ละคนจะใช้ในการทดสอบ ซึ่งสามารถสรุปจำนวนได้ดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ฐานข้อมูลในการทดสอบ

ฐานข้อมูล	จำนวนบุคคล	ภาพทั้งหมด	ชุดสอน	ชุดทดสอบ
ฐานข้อมูลขนาดใหญ่	1,558	103,182	91,034	12,148
ฐานข้อมูลขนาดเล็ก 9 คน	9	542	431	111
ฐานข้อมูลขนาดเล็ก 109 คน	109	874	654	220

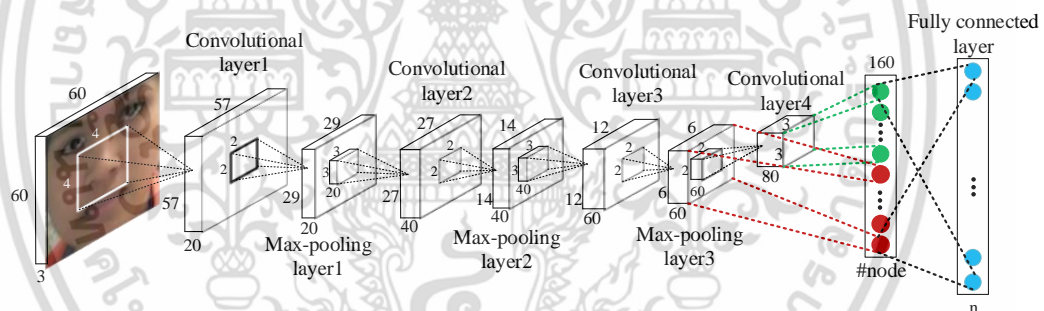
จากตารางที่ 4.1 ฐานข้อมูลขนาดใหญ่มีจำนวนบุคคลในฐานข้อมูลทั้งหมด 1,558 คน มีภาพทั้งหมด 103,182 ภาพ แบ่งออกเป็นชุดสอน 91,034 ภาพ และชุดทดสอบ 12,148 ภาพ ต่อมาฐานข้อมูลขนาดเล็ก 9 คน มีจำนวนบุคคลในฐานข้อมูลทั้งหมด 9 คน มีภาพทั้งหมด 542 ภาพ แบ่งเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ออกเป็นชุดสอน 431 ภาพ และชุดทดสอบ 111 ภาพ และสุดท้ายฐานข้อมูลขนาดเล็ก 109 คน มีจำนวนบุคคลในฐานข้อมูลทั้งหมด 109 คน มีภาพทั้งหมด 874 ภาพ แบ่งออกเป็นชุดสอน 654 ภาพ และชุดทดสอบ 220 ภาพ

4.2 การรู้จำภาพใบหน้าบุคคลโดยใช้ CNN

การทำงานของระบบรู้จำภาพใบหน้าบุคคลโดยใช้วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) จะใช้ Caffe [29] ซึ่งเป็นเครื่องมือในการทำงานของ CNN ซึ่งพัฒนาโดย UC Berkeley โดยจะใช้คอมพิวเตอร์ที่มีการ์ดจอช่วยในการประมวลผล โดยการ์ดจอนี้จำเป็นจะต้องมี CUDA ซึ่งเป็นการ์ดจอของบริษัท NVIDIA ที่จะช่วยในการประมวลผลให้มีความเร็วเพิ่มขึ้น จากการใช้ CPU ในการทำงานเพียงอย่างเดียว ซึ่งภาพรวมโครงสร้าง CNN ที่ใช้ในการทดลองนี้แสดงได้ดังรูปที่

4.1



รูปที่ 4.1 ภาพรวมโครงสร้าง CNN ที่ใช้ในการทดลอง

จากรูปที่ 4.1 โครงสร้าง CNN ที่ใช้ในการทดลองการรู้จำภาพใบหน้าบุคคล โดยภาพที่ใช้ในการทดลองเป็นภาพสีขนาด 60x60 พิกเซล ซึ่งจำนวนเซลล์ประสาทในชั้นสุดท้าย (n) จะขึ้นอยู่กับจำนวนบุคคลในฐานข้อมูลที่ใช้ โดยหากใช้กับฐานข้อมูล 9 คน $n=9$ และสำหรับฐานข้อมูล 109 คน จะใช้ $n=109$ ซึ่งในการทดลองมีการกำหนดค่าดังต่อไปนี้

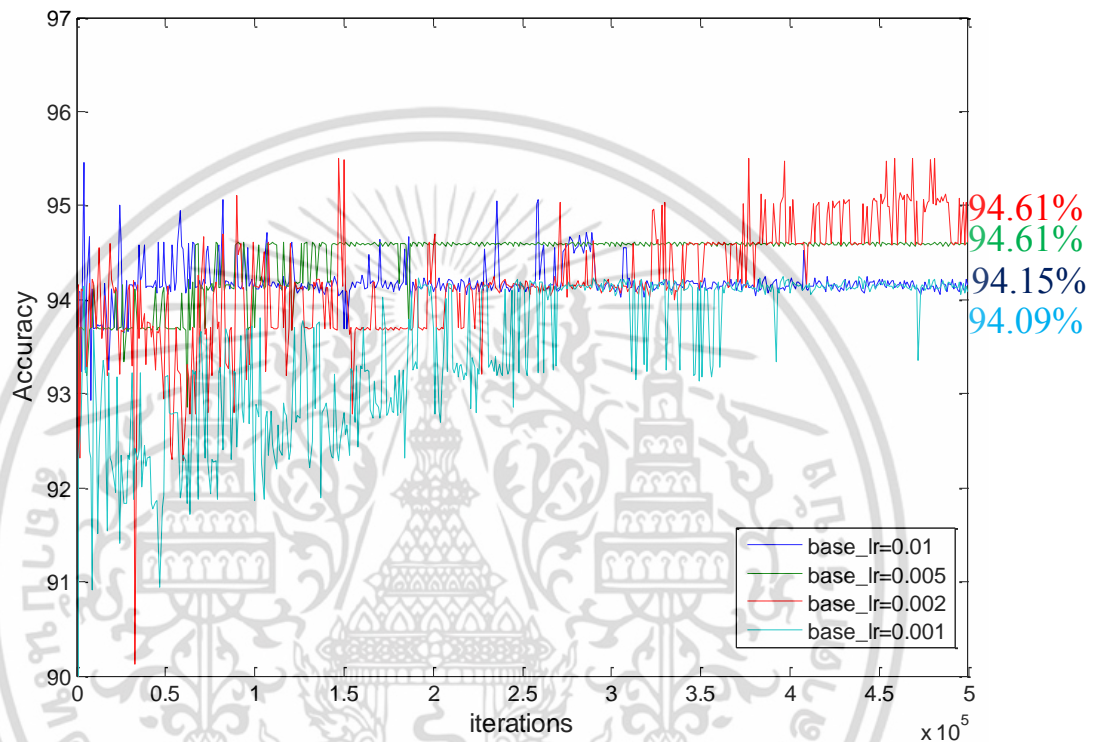
- จำนวนรอบในการสอน 500,000 รอบ
- มีการทดสอบเมื่อการสอนผ่านไปทุกๆ 1,000 รอบ

การทดลองในส่วนนี้แบ่งออกเป็น 2 ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1 การทดลองการใช้ CNN กับฐานข้อมูลขนาดเล็ก (9 และ 109 คน) โดยตรง

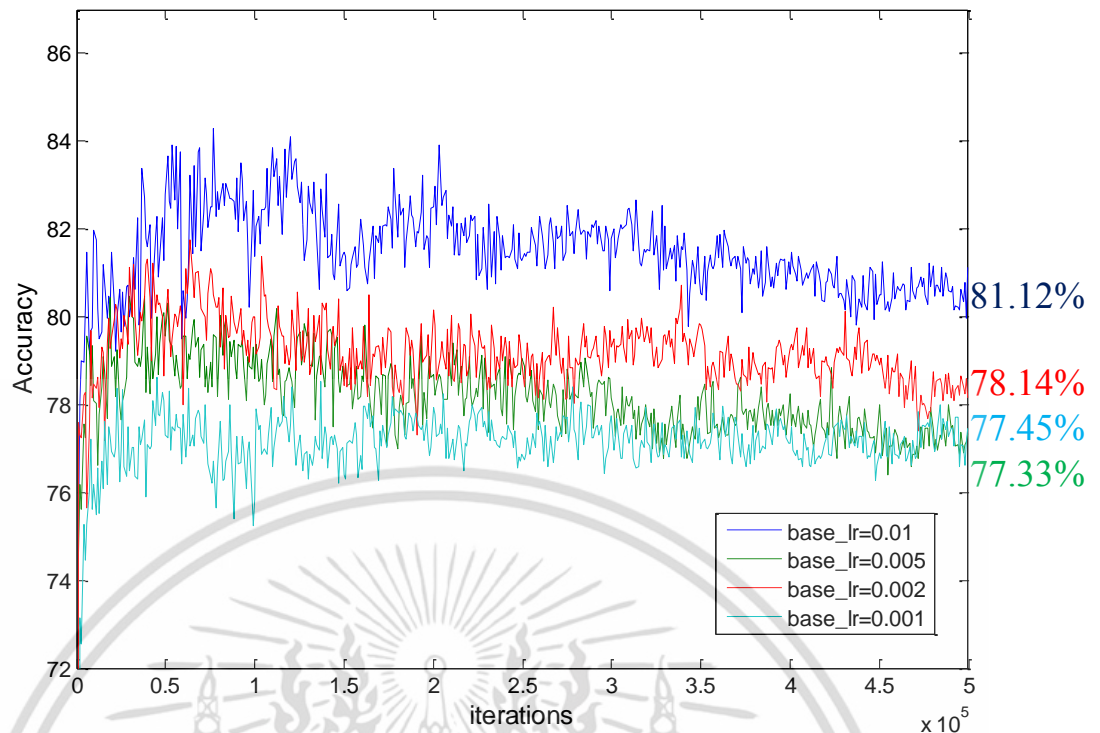
ในการทดลองนี้จะใช้โครงสร้างในรูปที่ 4.1 โดยมีการกำหนดค่า n ให้เท่ากับจำนวนบุคคลในฐานข้อมูล ในการทดลองนี้จำทดลองกับฐานข้อมูลขนาดเล็ก 9 คน และ 109 คน ซึ่งจะโครงสร้างจะมีค่า $n=9$ และ $n=109$ ตามลำดับ โดยมีการปรับค่าอัตราการเรียนรู้เริ่มต้น ซึ่งค่าอัตราการเรียนรู้นั้นจะมีการปรับค่าทุกๆ 10,000 รอบในการสอน โดยในการทดลองนี้จะมีค่าอัตราการเรียนรู้เริ่มต้นเป็น 0.01 0.005 0.002 และ 0.001 ตามลำดับ ผลการทดลองกับฐานข้อมูล 9 คนแสดงได้ในรูปที่ 4.2



รูปที่ 4.2 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูล 9 คน

จากรูปที่ 4.2 กราฟแสดงความสัมพันธ์ระหว่างจำนวนรอบที่ใช้สอนและค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล จะเห็นว่าการปรับค่าอัตราการเรียนรู้เริ่มต้นที่ค่าต่างๆ ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลใกล้เคียงกันเมื่อมีจำนวนรอบในการสอน 500,000 รอบ โดยค่าความถูกต้องสูงสุดคือ 94.61 %

ต่อมาเป็นการนำไป CNN ไปใช้กับฐานข้อมูล 109 คน ซึ่งปรับค่าอัตราการเรียนรู้เริ่มต้นเช่นเดียวกับการทดลองก่อนหน้านี้ ซึ่งผลการทดลองแสดงได้ดังรูปที่ 4.3



รูปที่ 4.3 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูล 109 คน

จากรูปที่ 4.3 แสดงความสัมพันธ์ระหว่างจำนวนรอบที่ใช้สอนและค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล จะเห็นว่าการปรับค่าอัตราการเรียนรู้เริ่มต้นที่ค่าต่างๆ ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลต่างกันเล็กน้อย โดยค่าอัตราการเรียนรู้เริ่มต้น 0.005 ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าสูงสุด 81.12 % และเมื่อค่าอัตราการเรียนรู้เริ่มต้นมีการเปลี่ยนแปลง ค่าความถูกต้องที่ได้ก็จะแตกต่างกันด้วย

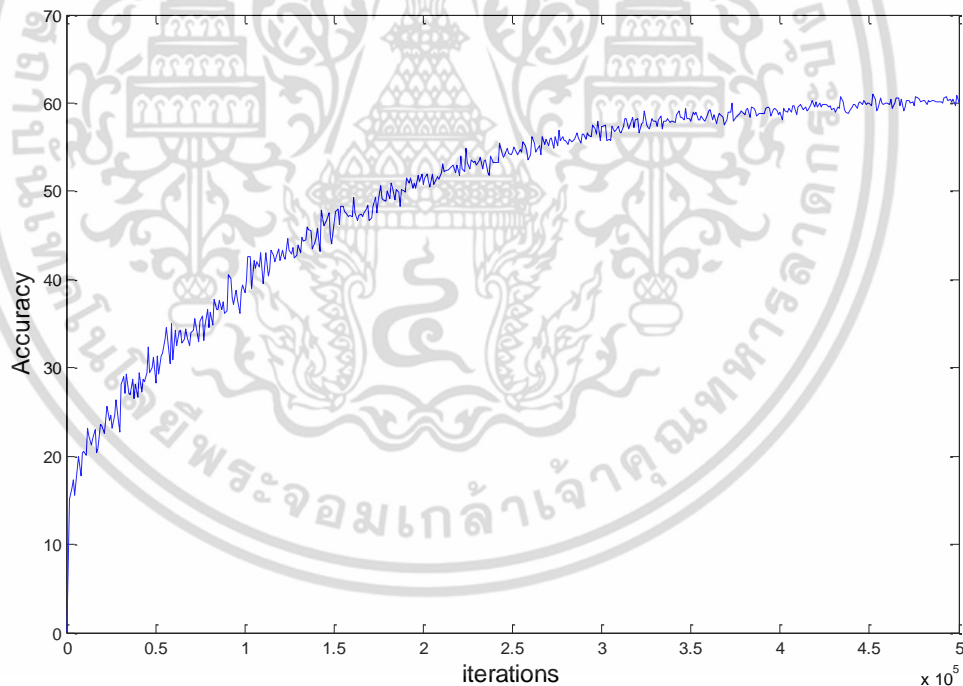
จากการทดลองการใช้งาน CNN กับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยตรงพบว่า ฐานข้อมูล 9 คน มีค่าความถูกต้องสูงสุดในการรู้จำภาพใบหน้า 94.61 % และฐานข้อมูล 109 คน มีความความถูกต้องสูงสุดในการรู้จำภาพใบหน้า 81.12 % เมื่อจำนวนบุคคลในฐานข้อมูลเพิ่มขึ้นจะส่งผลให้ค่าความถูกต้องมีค่าลดลงเมื่อใช้ระบบรู้จำภาพใบหน้าบุคคลแบบเดียวกัน และสำหรับฐานข้อมูล 9 คนนั้น มีค่าความถูกต้องสูงสุดถึง 94.61 % เนื่องจากลักษณะของภาพในฐานข้อมูลที่มีความใกล้เคียงกัน แตกต่างกันเพียงสภาพแสงที่ตกกระทบบนภาพใบหน้าเท่านั้น แต่ฐานข้อมูล 109 คนนั้นแตกต่างกันทั้งสภาพแสง และสภาพแวดล้อมรอบข้าง นอกจากนี้อิริยาบถของภาพใบหน้ายังมีลักษณะที่แตกต่างกัน 6 ลักษณะ ทำให้ภาพในฐานข้อมูลมีความแตกต่างกันมาก เหล่านี้จึงเป็นสาเหตุที่สำคัญทำให้ค่าความถูกต้องของฐานข้อมูลนี้มีค่าเท่ากับ 81.12 % เท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลอง CNN กับฐานข้อมูลขนาดเล็กโดยตรงพบว่า ค่าความถูกต้องของระบบมีค่าสูงอยู่ในระดับที่ยอมรับได้ แต่อย่างไรก็ตามประสิทธิภาพของ CNN โดยทั่วไปนั้นสามารถให้ค่าความถูกต้องที่สูงกว่า 90 % ดังนั้นจึงได้ทดลองใช้ CNN แบบถ่ายโอนการเรียนรู้กับฐานข้อมูลขนาดเล็กทั้ง 2 ฐานข้อมูล เพื่อเพิ่มประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลกับฐานข้อมูลขนาดเล็กให้มีประสิทธิภาพสูงขึ้น

4.2.2 การทดลองการใช้ CNN แบบถ่ายโอนการเรียนรู้ (TL-CNN)

สำหรับระบบ TL-CNN นี้ เริ่มทดลองโดยการใช้ CNN กับฐานข้อมูลขนาดใหญ่ (Celebrity face) ก่อน จากนั้นเปลี่ยนชั้นจำแนกและฐานข้อมูลเพื่อนำไปประยุกต์ใช้กับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) ในการทดลองนี้จะใช้โครงสร้างเช่นเดียวกับ CNN ดังแสดงในรูปที่ 4.1 ค่าคุณลักษณะเด่นเริ่มต้น (ค่าน้ำหนัก และค่าสัมประสิทธิ์ในหน้ากาคอนโวลูชัน) จะใช้จากค่าคุณลักษณะเด่นที่ได้จากการสอนด้วยฐานข้อมูลขนาดใหญ่ โดยผลการทดลองเบื้องต้นกับฐานข้อมูลขนาดใหญ่แสดงได้ดังรูปที่ 4.4



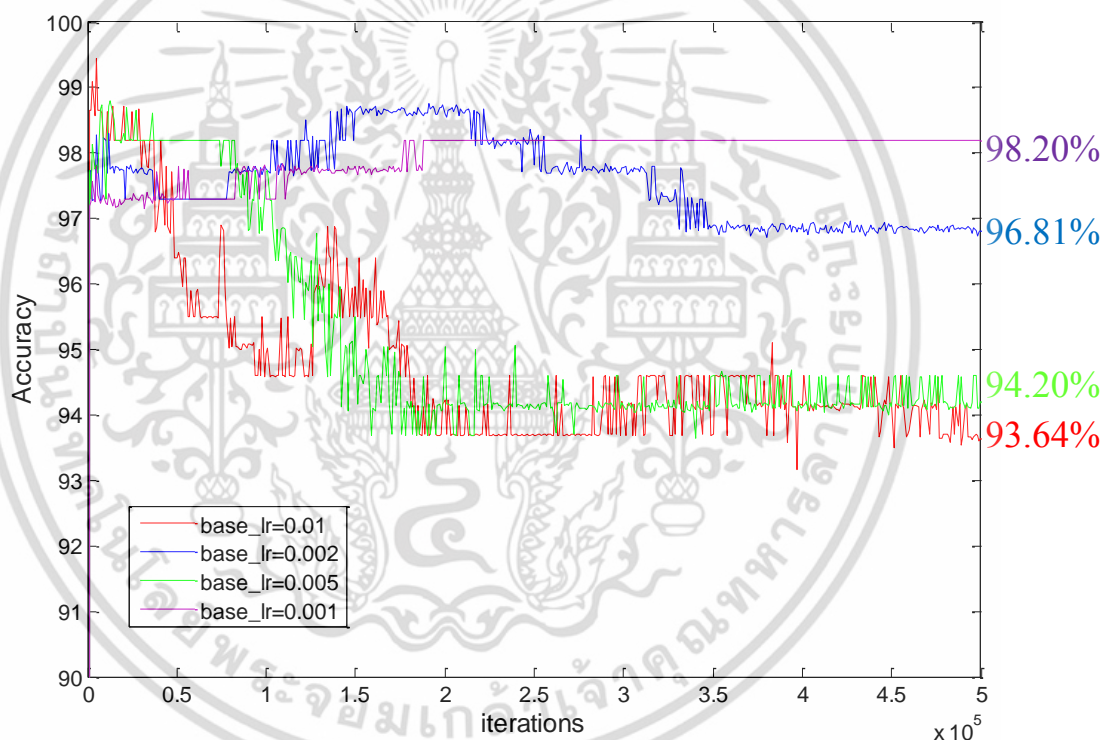
รูปที่ 4.4 ค่าความถูกต้องของการทดลอง CNN กับฐานข้อมูลขนาดใหญ่

รูปที่ 4.4 แสดงความสัมพันธ์ระหว่างจำนวนรอบที่ใช้สอนและค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล ผลการทดลอง CNN กับฐานข้อมูลขนาดใหญ่และมีค่าอัตราการเรียนรู้เริ่มต้นเป็น 0.01 พบว่ามีค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลเท่ากับ 60.06 % โดยการทดลองนี้ทำเพื่อนำค่า

เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

น้ำหนัก และค่าสัมประสิทธิ์ในหน้ากาคอนโวลูชันที่ได้จากการเรียนรู้ด้วยฐานข้อมูลขนาดใหญ่ไปเป็นค่าน้ำหนักและค่าสัมประสิทธิ์ในหน้ากาคอนโวลูชันเริ่มต้นให้กับการใช้ CNN กับฐานข้อมูลขนาดเล็ก เพื่อเพิ่มค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลของฐานข้อมูลขนาดเล็กเนื่องจากค่าน้ำหนักที่ได้จากการเรียนรู้กับฐานข้อมูลขนาดเล็กเพียงอย่างเดียวไม่เพียงพอที่จะทำให้ระบบรู้จำใบหน้าบุคคลมีประสิทธิภาพสูง

จากนั้นจะทำการเปลี่ยนชั้นการจำแนกจากฐานข้อมูลขนาดใหญ่เป็นฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) และค่าน้ำหนักเริ่มต้นในการสอนฐานข้อมูลขนาดเล็กได้มาจากการสอนด้วยฐานข้อมูลขนาดใหญ่ในการทดลองก่อนหน้า โดยปรับค่าอัตราการเรียนรู้เริ่มต้นในการทดลองแต่ละครั้งเป็น 0.01 0.005 0.002 และ 0.001 ตามลำดับ เช่นเดียวกับการทดลอง CNN กับฐานข้อมูลขนาดเล็กโดยตรง ซึ่งผลการทดลองสำหรับฐานข้อมูล 9 คนแสดงได้ดังรูปที่ 4.5



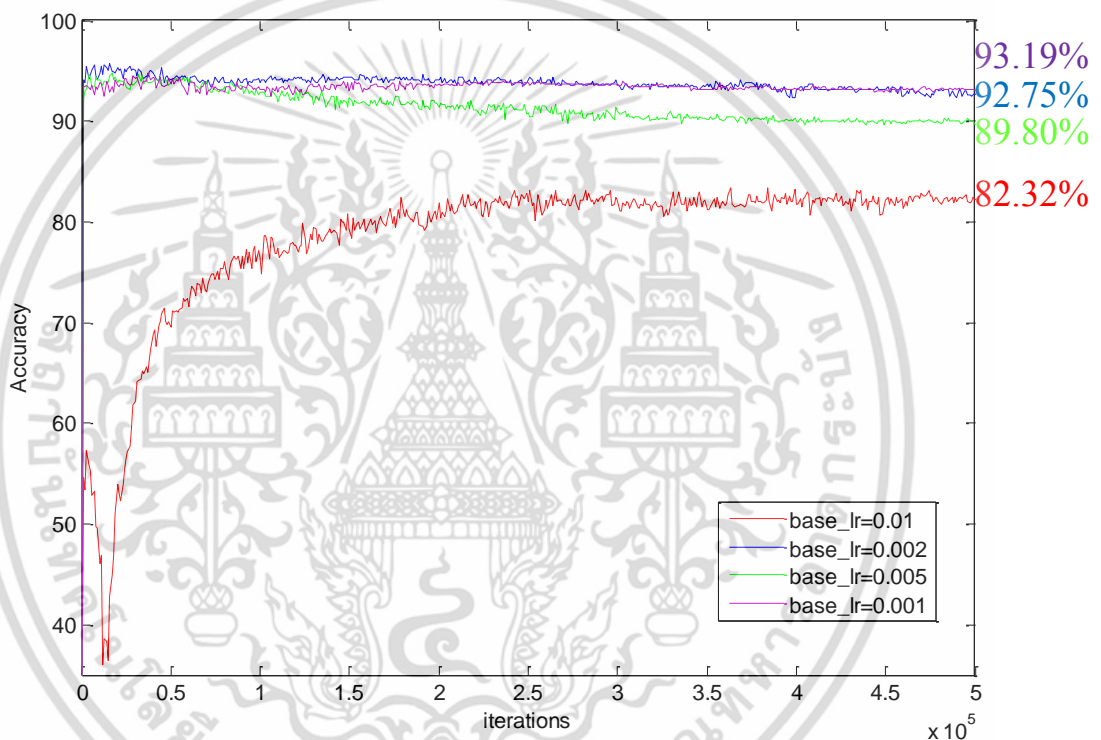
รูปที่ 4.5 ค่าความถูกต้องของการทดลอง TL-CNN กับฐานข้อมูล 9 คน

รูปที่ 4.5 แสดงความสัมพันธ์ระหว่างจำนวนรอบที่ใช้สอนและค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล ผลการทดลอง CNN แบบถ่ายโอนการเรียนรู้กับฐานข้อมูล 9 คน พบว่าเมื่อค่าอัตราการเรียนรู้เริ่มต้นลดลงส่งผลให้ค่าความถูกต้องในการรู้จำภาพใบหน้าเพิ่มขึ้น โดยค่าอัตราการเรียนรู้เริ่มต้น 0.001 มีค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลสูงสุดเท่ากับ 98.20 % เมื่อพิจารณาสมการที่ 2.28 พบว่าเมื่อปรับค่าอัตราการเรียนรู้จะส่งผลต่อการปรับค่าน้ำหนักในการเรียนรู้ โดยหาก

ค่าอัตราการเรียนรู้เริ่มต้นมีค่าน้อย จะทำให้การปรับค่าน้ำหนักในแต่ละรอบที่ใช้สอนจะมีค่าน้อยตามเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้วย การปรับค่าน้ำหนักที่ละน้อยนี้จะส่งผลให้การคำนวณมีเสถียรภาพ ซึ่งสามารถสังเกตได้จากการแกว่งของค่าความถูกต้องของระบบรู้จำภาพใบหน้าบุคคล สำหรับอัตราการเรียนรู้เริ่มต้นเท่ากับ 0.001 จะมีการแกว่งของค่าความถูกต้องของระบบรู้จำภาพใบหน้าบุคคลน้อยสุดดังแสดงในรูปที่ 4.5 หรือกล่าวอีกนัยหนึ่งว่ามีเสถียรภาพของการคำนวณสูง นอกจากนี้การปรับค่าน้ำหนักที่ละน้อยทำให้ระบบเรียนรู้ที่ละน้อย และอาจจะทำให้ค่าความถูกต้องของระบบรู้จำภาพใบหน้าบุคคลมีค่าสูงขึ้นด้วย

ต่อมาเป็นการทดลอง CNN แบบถ่ายโอนการเรียนรู้กับฐานข้อมูล 109 คน โดยมีการปรับค่าอัตราการเรียนรู้เริ่มต้นในการทดลองแต่ละครั้งเป็น 0.01 0.005 0.002 และ 0.001 ตามลำดับ ผลการทดลองสำหรับฐานข้อมูล 109 คนแสดงได้ดังรูปที่ 4.6



รูปที่ 4.6 ค่าความถูกต้องของการทดลอง TL-CNN กับฐานข้อมูล 109 คน

จากรูปที่ 4.6 กราฟแสดงความสัมพันธ์ระหว่างจำนวนรอบที่ใช้สอนและค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคล ผลการทดลอง CNN แบบถ่ายโอนการเรียนรู้กับฐานข้อมูล 109 คน พบว่าเมื่อค่าอัตราการเรียนรู้เริ่มต้นลดลงส่งผลให้ค่าความถูกต้องในการรู้จำภาพใบหน้าเพิ่มขึ้น โดยค่าอัตราการเรียนรู้เริ่มต้น 0.001 มีค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลสูงสุดเท่ากับ 93.19 % ทั้งนี้ผลการทดลองนี้ยังสอดคล้องกับผลการทดลองข้างต้นที่ว่า เมื่ออัตราการเรียนรู้เริ่มต้นมีค่าลดลงจะทำให้การคำนวณมีเสถียรภาพสูง นอกจากนี้ยังยืนยันว่าเมื่ออัตราการเรียนรู้เริ่มต้นมีค่าน้อยจะทำให้ระบบมีประสิทธิภาพสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทดลอง CNN กับฐานข้อมูลขนาดเล็กโดยตรง และ CNN แบบถ่ายโอนการเรียนรู้ผลการทดลองเปรียบเทียบสามารถสรุปได้ดังตารางที่ 4.2

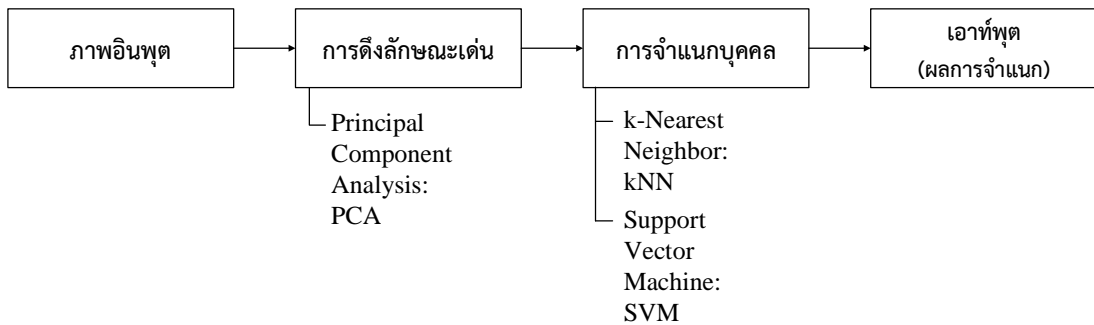
ตารางที่ 4.2 ผลเปรียบเทียบค่าความถูกต้องของการทดลอง CNN และ TL-CNN

รูปแบบการทดสอบ	ค่าความถูกต้อง	
	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน
CNN	94.61%	81.12%
TL-CNN	98.20%	93.19%

จากตารางที่ 4.2 พบว่าการรู้จำภาพใบหน้าบุคคลโดยใช้ TL-CNN ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าสูงกว่าการใช้ CNN กับฐานข้อมูลขนาดเล็กโดยตรง ทั้งกรณีฐานข้อมูล 9 คน และ 109 คน เนื่องจากค่าน้ำหนักที่ได้จากฐานข้อมูลขนาดใหญ่ซึ่งมีจำนวนข้อมูลในฐานข้อมูลมากพอที่จะทำให้ CNN เกิดการเรียนรู้คุณลักษณะได้ และเมื่อนำค่าน้ำหนักนั้นมาเป็นค่าน้ำหนักเริ่มต้นให้กับการทดลอง CNN กับฐานข้อมูลขนาดเล็กจึงทำให้สามารถเพิ่มประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลได้ในส่วนของการปรับอัตราการเรียนรู้เริ่มต้นของ TL-CNN นั้น พบว่าผลการทดลองให้ผลที่สอดคล้องกันระหว่างฐานข้อมูล 9 คน และ 109 คน กล่าวคือเมื่อค่าอัตราการเรียนรู้เริ่มต้นมีค่าน้อยทำให้มีเสถียรภาพในการคำนวณสูงและทำให้ความถูกต้องของการรู้จำภาพใบหน้าบุคคลมีค่าเพิ่มขึ้นด้วย ทั้งกรณีของ CNN และ TL-CNN นอกจากนี้คุณภาพของภาพในฐานข้อมูลเป็นอีกหนึ่งปัจจัยที่ส่งผลต่อประสิทธิภาพของการรู้จำภาพใบหน้าบุคคล

4.3 การรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป

การรู้จำภาพใบหน้าบุคคลโดยทั่วไปแล้วจะมีการทำงานของระบบดังนี้เริ่มต้นจากภาพอินพุตจากนั้นจะเข้าสู่กระบวนการดึงลักษณะเด่น และกระบวนการจำแนกบุคคลตามลำดับ สุดท้ายจะได้เอาท์พุตซึ่งเป็นคำตอบที่ได้จากการทำงานของระบบรู้จำภาพใบหน้าบุคคล โดยการทำงานของระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไปนั้นจะพัฒนาโปรแกรมโดยใช้ภาษา c++ ร่วมกันกับไลบรารี openCV ในการพัฒนาโปรแกรมขึ้นมา ซึ่งภาพรวมการทำงานของระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไปแสดงได้ดังรูปที่ 4.7



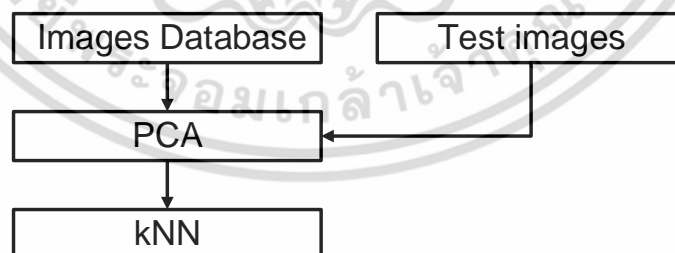
รูปที่ 4.7 ระบบรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป

จากรูปที่ 4.7 วิธีการดึงลักษณะเด่นที่ใช้ในวิทยานิพนธ์นี้คือ PCA และการจำแนกที่ใช้มี 2 วิธีด้วยกันได้แก่ kNN และ SVM โดยการทดลองการรู้จำภาพใบหน้าบุคคลในส่วนนี้แบ่งออกได้เป็น 2 การทดลอง วิธีการแรกคือ การใช้ PCA ในการดึงลักษณะเด่น และใช้ kNN ในการจำแนก และวิธีการที่สองคือ การใช้ PCA ในการดึงลักษณะเด่นเช่นกัน แต่ใช้ SVM ในการจำแนก โดยฐานข้อมูลที่ใช้ในการทดลองคือฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) ในการหาประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลซึ่งการทดลองแสดงได้ดังต่อไปนี้

4.3.1 การรู้จำภาพใบหน้าบุคคลโดยใช้ PCA และ kNN

การรู้จำภาพใบหน้าบุคคลโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ kNN ในการจำแนก จะแบ่งย่อยออกเป็น 2 การทดลองดังต่อไปนี้

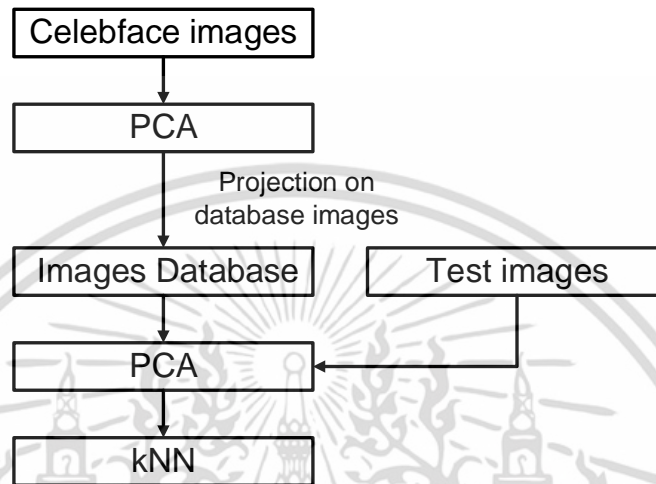
1) การใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยตรงจากนั้นใช้ kNN ในการจำแนกภาพใบหน้าบุคคล (PCA + kNN) แผนภาพการทำงานแสดงได้ดังรูปที่ 4.8



รูปที่ 4.8 การทำงานของ PCA ในการดึงลักษณะเด่น และใช้ kNN ในการจำแนก

จากรูปที่ 4.8 เป็นการนำ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็ก (9 และ 109 คน) จากนั้นเมื่อมีภาพที่ต้องการทดลองก็จะผ่านกระบวนการ PCA เช่นเดียวกัน และใช้ kNN ในการจำแนกภาพใบหน้าบุคคล

2) การใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อน จากนั้นจึงนำลักษณะเด่นที่ได้จากฐานข้อมูลขนาดใหญ่ไปลดขนาดลงบนฐานข้อมูลขนาดเล็กและดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็กอีกครั้ง แล้วจึงใช้ kNN ในการจำแนกภาพใบหน้าบุคคล (PCA (Celeb) + kNN) แผนภาพการทำงานแสดงได้ดังรูปที่ 4.9



รูปที่ 4.9 การทำงานของ PCA ในการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่และ kNN จำแนก

จากรูปที่ 4.9 เป็นการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อน ซึ่งฐานข้อมูลขนาดใหญ่ในวิทยานิพนธ์นี้ใช้ฐานข้อมูลภาพใบหน้าบุคคลที่มีชื่อเสียง (Celebrity face) จากนั้นนำไปโปรเจกต์ลงบนฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) จากนั้นใช้ PCA ในการดึงลักษณะเด่นอีกครั้ง และเมื่อมีภาพทดสอบเข้ามา ก็จะผ่านกระบวนการ PCA และเข้าสู่ขั้นตอนการจำแนกด้วย kNN ซึ่งการทดลองในส่วนนี้จัดทำขึ้นเพื่อให้การทดลองมีความคล้ายกันกับ TL-CNN โดยมีการดึงลักษณะเด่นจากฐานข้อมูลใหญ่ก่อน จากนั้นจึงนำไปใช้งานกับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน)

การทดลองโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ kNN ในการจำแนกนั้นจะทดลองกับฐานข้อมูล 9 คน และ 109 คน ตามลำดับ นอกจากนี้มีการปรับค่า k ในกระบวนการ kNN เพื่อหาค่า k ที่เหมาะสมกับการทำงานของระบบรู้จำภาพใบหน้าบุคคล โดยค่า k ที่ใช้ในการทดลองได้แก่ 1 2 3 4 และ 5 ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลแสดงได้ดังตารางที่ 4.3

ตารางที่ 4.3 ผลการทดลองโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ kNN ในการจำแนก

k	ค่าความถูกต้อง			
	PCA + kNN		PCA (Celeb) + kNN	
	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน
1	82.88%	68.18%	87.39%	67.73%
2	86.49%	65.91%	84.68%	64.09%
3	85.59%	63.64%	84.68%	60.91%
4	77.48%	61.82%	81.98%	59.09%
5	80.18%	61.36%	79.28%	61.36%

จากตารางที่ 4.3 พบว่าค่า k ในกระบวนการ kNN สำหรับการทดลองในส่วนแรก (PCA + kNN) ค่าความถูกต้องสูงสุดที่ได้จากฐานข้อมูล 9 คน และ 109 คน มีค่าเท่ากับ 86.49 % (k = 2) และ 68.18 % (k = 1) ตามลำดับและค่าความถูกต้องของการทดลองในส่วนที่สอง (PCA (Celeb) + kNN) ของฐานข้อมูล 9 คน และ 109 คน มีค่าเท่ากับ 87.39 % (k = 1) และ 67.73 % (k = 1) ตามลำดับเนื่องจากกระบวนการ kNN เป็นวิธีการที่สนใจข้อมูลที่อยู่รอบข้างข้อมูลที่นำมาทดสอบ ดังนั้นค่า k จึงส่งผลต่อค่าความถูกต้องของระบบรู้จำภาพใบหน้าบุคคล โดยผลการทดลอง PCA + kNN กับฐานข้อมูล 9 คนมีค่าความถูกต้องสูงสุดเมื่อค่า k = 2 โดยจะสนใจสมาชิกที่อยู่รอบข้างข้อมูลทดสอบ 2 ตัว ให้ค่าความถูกต้องที่สูงกว่า แต่สำหรับกรณีอื่นนั้นค่า k=1 ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าสูงสุด เนื่องจากพิจารณาสมาชิกที่อยู่รอบตัวที่สนใจเพียงตัวเดียวเท่านั้น

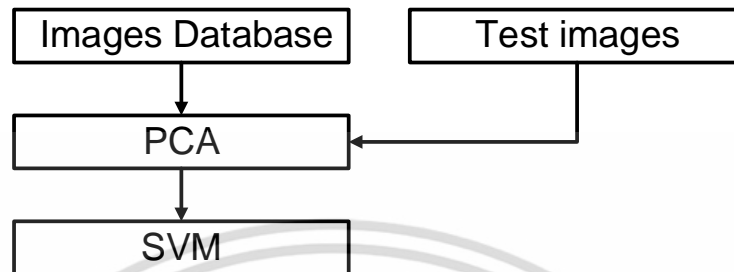
จากผลการทดลองสามารถสรุปได้ว่า ค่าความถูกต้องที่ได้จากการดึงลักษณะเด่นจากฐานข้อมูลโดยตรงหรือ การดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อน ไม่มีความแตกต่างอย่างมีนัยสำคัญ หรือกล่าวอีกนัยหนึ่งว่าการใช้ฐานข้อมูลขนาดใหญ่ในการดึงลักษณะเด่นโดยใช้ PCA ก่อนการนำมาใช้กับฐานข้อมูลขนาดเล็กนั้น ไม่ส่งผลต่อค่าความถูกต้องอย่างมีนัยสำคัญ

4.3.2 การรู้จำภาพใบหน้าบุคคลโดยใช้ PCA และ SVM

การทำงานของระบบรู้จำภาพใบหน้าบุคคลโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก สามารถแบ่งออกเป็น 2 การทดลองดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

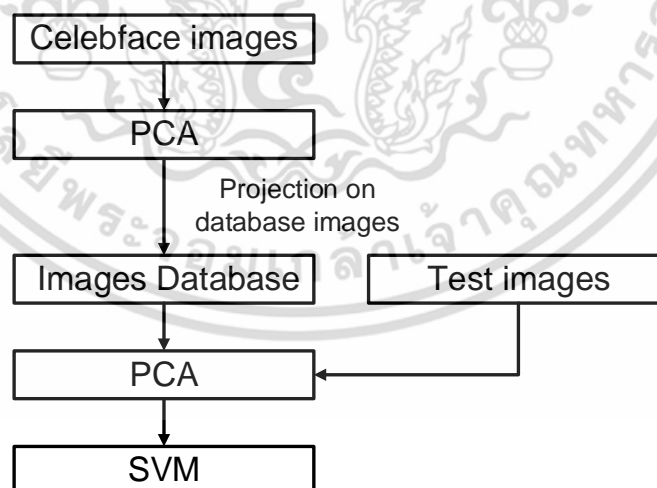
1) การใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยตรง จากนั้นใช้ SVM ในการจำแนกภาพใบหน้าบุคคล (PCA + SVM) แผนภาพการทำงานแสดงได้ดังรูปที่ 4.10



รูปที่ 4.10 การทดลองการใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก

จากรูปที่ 4.10 สามารถอธิบายการทำงานได้ดังต่อไปนี้ ภาพในฐานะข้อมูลผ่านกระบวนการดึงลักษณะเด่นโดยใช้ PCA จากนั้นเมื่อมีภาพที่ต้องการทดสอบจะผ่านกระบวนการ PCA เช่นเดียวกัน และเข้าสู่ขั้นตอนการจำแนกภาพใบหน้าบุคคลโดยใช้ SVM

2) การใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อน จากนั้นจึงนำไปดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็ก แล้วจึงใช้ SVM ในการจำแนกภาพใบหน้าบุคคล (PCA (Celeb) + SVM) แผนภาพการทำงานแสดงได้ดังรูปที่ 4.11



รูปที่ 4.11 การทดลองการใช้ PCA ในการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่และใช้ SVM จำแนก

จากรูปที่ 4.11 จะมีการดึงลักษณะเด่น 2 ครั้ง ครั้งแรกคือการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่โดยใช้ PCA ก่อน จากนั้นนำลักษณะเด่นที่ได้ไปแจกแจงบนฐานข้อมูลขนาดเล็ก เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(9 คน และ 109 คน) และดึงลักษณะเด่นด้วย PCA อีกครั้ง เมื่อมีภาพที่ต้องการทดสอบก็จะผ่านกระบวนการ PCA เช่นเดียวกัน และเข้าสู่ขั้นตอนการจำแนกภาพใบหน้าบุคคลโดยใช้ SVM

การทดลองโดยใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนกนั้นจะทดลองกับฐานข้อมูล 9 คน และ 109 คน ตามลำดับ และมีการศึกษาผลของเส้นแบ่ง (Hyper plane) ต่อค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลโดยทดลอง 2 ลักษณะได้แก่ แบบเส้นตรง (Linear) และแบบ RBF (Radial basis function) โดยเส้นแบ่งแบบเส้นตรงนั้นจะมีการปรับค่า C ในช่วง $1 \times 10^{-2} - 1 \times 10^{10}$ และเส้นแบ่งแบบ RBF มีการปรับค่า 2 ตัวแปรคือ C และ γ โดยปรับในช่วง $1 \times 10^{-2} - 1 \times 10^{10}$ และ $1 \times 10^{-9} - 1 \times 10^3$ ตามลำดับ

โดยสาเหตุของการเลือกศึกษาอิทธิพลของประเภทของเส้นแบ่งนั้นเนื่องจากการจะดูผลของความเป็นเชิงเส้นและความไม่เป็นเชิงเส้นของเส้นแบ่ง โดยเส้นแบ่งแบบเส้นตรง จะเป็นเชิงเส้น ($K(x, x_j) = x^T x_j$) และเส้นแบ่งแบบ RBF จะไม่เป็นเชิงเส้น ($K(x, x_j) = e^{-\gamma \|x - x_j\|^2}$) สำหรับการเลือกศึกษาอิทธิพลของค่าคงที่ C นั้นเนื่องจากค่า C จะส่งผลถึงระยะมาร์จิน ดังแสดงในสมการที่ 2.15 โดยถ้าค่า C มีค่ามากจะทำให้ข้อมูลไม่สามารถอยู่ระหว่างระยะมาร์จินได้ ถ้าค่า C มีค่าน้อยส่งผลให้ข้อมูลบางจุดสามารถอยู่ระหว่างมาร์จินได้ ในส่วนสุดท้ายการเลือกศึกษาอิทธิพลของค่าคงที่ γ เพราะการปรับค่า γ จะส่งผลต่อการวางตัวของเส้นแบ่ง

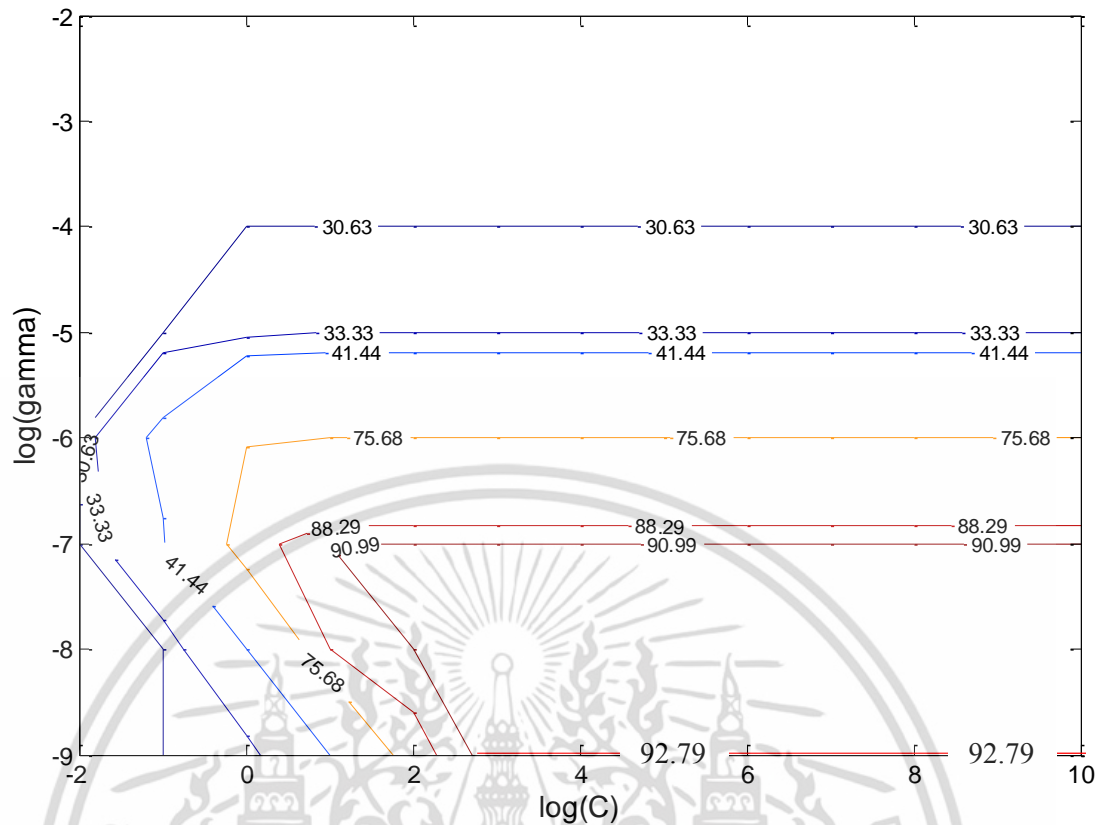
ผลการทดลองระบบรู้จำภาพใบหน้าบุคคลกับฐานข้อมูลเล็ก (9 คน และ 109 คน) โดยใช้เส้นแบ่งแบบเป็นเส้นตรง โดยที่ค่าความถูกต้องของการทดลองในส่วนนี้สามารถแสดงได้ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการทดลองการใช้ SVM ในการจำแนกและมีเส้นแบ่งแบบเป็นเส้นตรง

C	ค่าความถูกต้อง			
	ฐานข้อมูล 9 คน		ฐานข้อมูล 109 คน	
	PCA + SVM	PCA (Celeb) + SVM	PCA + SVM	PCA (Celeb) + SVM
1.00E-02	92.79 %	91.89 %	71.36 %	70.00 %
1.00E-01	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+00	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+01	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+02	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+03	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+04	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+05	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+06	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+07	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+08	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+09	92.79 %	91.89 %	71.36 %	70.00 %
1.00E+10	92.79 %	91.89 %	71.36 %	70.00 %
1.00E-02	92.79 %	91.89 %	71.36 %	70.00 %

จากตารางที่ 4.4 พบว่า ในเบื้องต้นการปรับค่า C ในช่วงที่ศึกษาไม่ส่งผลต่อค่าความถูกต้อง โดยการทดลองทั้ง 2 แบบ ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าใกล้เคียงกัน โดยการใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลโดยตรงให้ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลสูงกว่าการใช้ PCA ดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อนเพียงเล็กน้อย

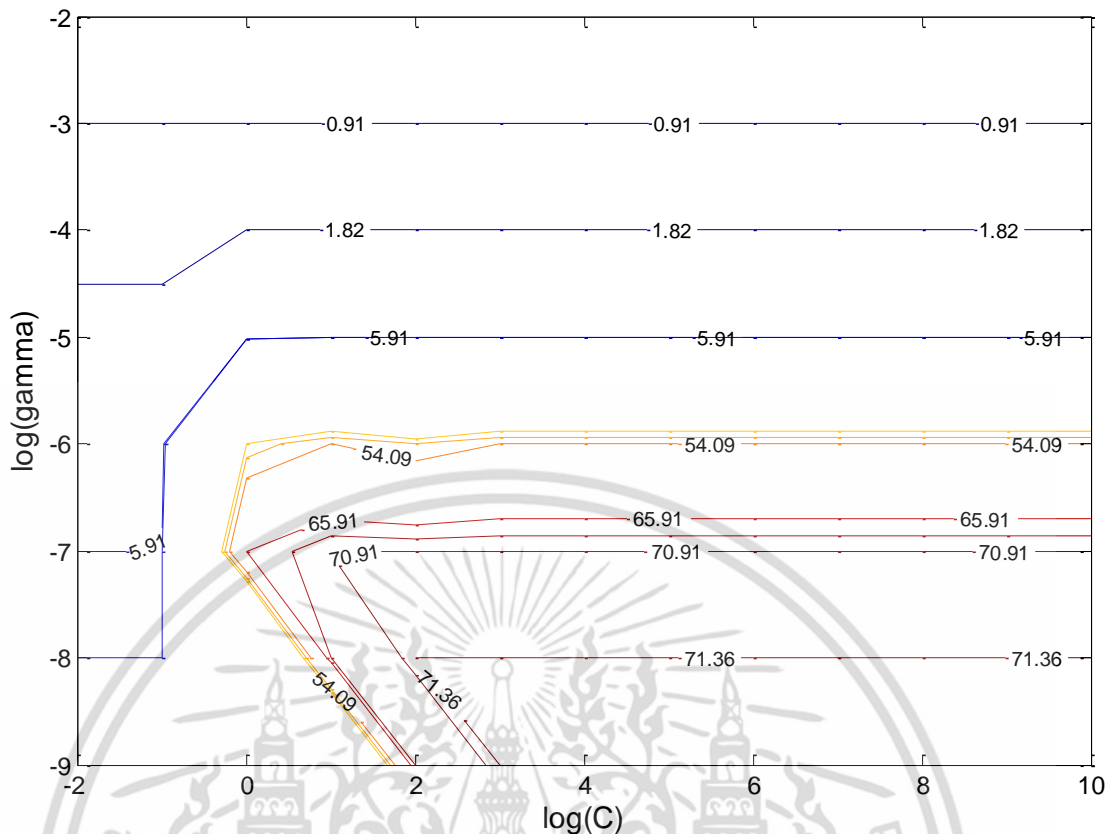
และการทดลองการรู้จำภาพใบหน้าบุคคลโดยใช้เส้นแบ่งแบบ RBF และการปรับค่า C และ γ นั้นสามารถแสดงได้ดังนี้การทดลอง PCA + SVM โดยใช้เส้นแบ่งแบบ RBF กับฐานข้อมูล 9 คน ซึ่งการปรับค่า C และ γ สามารถแสดงได้ในรูปที่ 4.12



รูปที่ 4.12 ค่าความถูกต้องของการทดลอง PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก เมื่อเส้นแบ่งแบบ RBF กับฐานข้อมูล 9 คน

รูปที่ 4.12 แสดงแผนภาพความสัมพันธ์ระหว่าง $\log(C)$ และ $\log(\gamma)$ โดยมีค่าอยู่ในช่วง -2 ถึง 10 และ -9 ถึง 3 ตามลำดับ โดยตัวเลขที่ปรากฏตรงกลางคือค่าความถูกต้องเมื่อค่า $\log(C)$ และ $\log(\gamma)$ เป็นค่าแตกต่างกัน จากแผนภาพจะเห็นว่า ที่ $\log(C)$ มีค่าตั้งแต่ 3 ถึง 10 และค่า $\log(\gamma)$ เป็น -9 ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าสูงที่สุดคือ 92.79 % และเมื่อค่า C และ γ เปลี่ยนไป ส่งผลต่อค่าความถูกต้องที่เปลี่ยนแปลงด้วยเช่นกัน

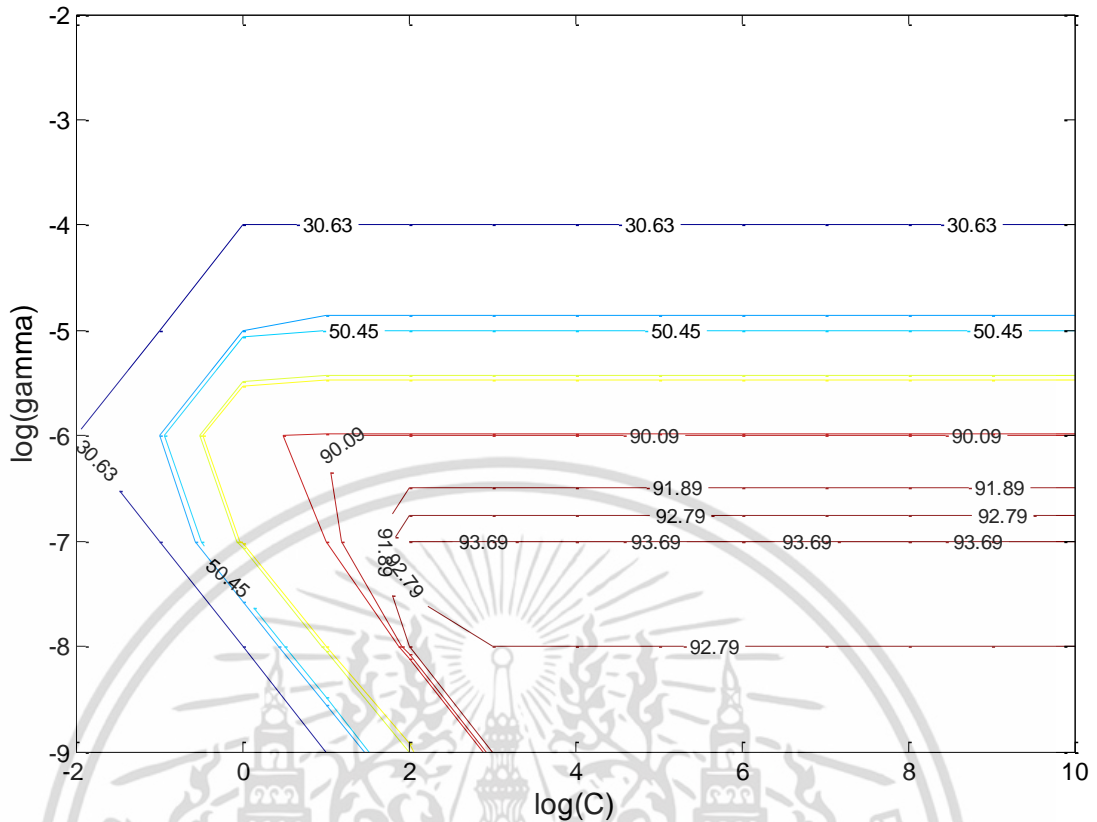
การทดลอง PCA(Celeb)+ SVM โดยใช้เส้นแบ่งแบบ RBF กับฐานข้อมูล 9 คน โดยการปรับค่า C และ γ สามารถแสดงได้ในรูปที่ 4.13



รูปที่ 4.13 ค่าความถูกต้องของการทดลอง PCA (Celeb)+ SVM เมื่อเส้นแบ่ง RBF กับฐานข้อมูล 9 คน

รูปที่ 4.13 แผนภาพแสดงความสัมพันธ์ระหว่าง $\log(C)$ และ $\log(\gamma)$ โดยมีค่าอยู่ในช่วง -2 ถึง 10 และ -9 ถึง 3 ตามลำดับ โดยตัวเลขที่ปรากฏตรงกลางคือค่าความถูกต้องเมื่อค่า $\log(C)$ และ $\log(\gamma)$ เป็นค่าแตกต่างกันพบว่าค่า $\log(C)$ มีค่าในช่วง 2 ถึง 10 เมื่อ $\log(\gamma)$ มีค่า -8 และ $\log(C)$ มีค่า 3 ถึง 10 เมื่อ $\log(\gamma)$ มีค่า -9 จะให้ผลการรู้จำภาพใบหน้าบุคคลมีค่าความถูกต้องสูงสุด 71.36 %

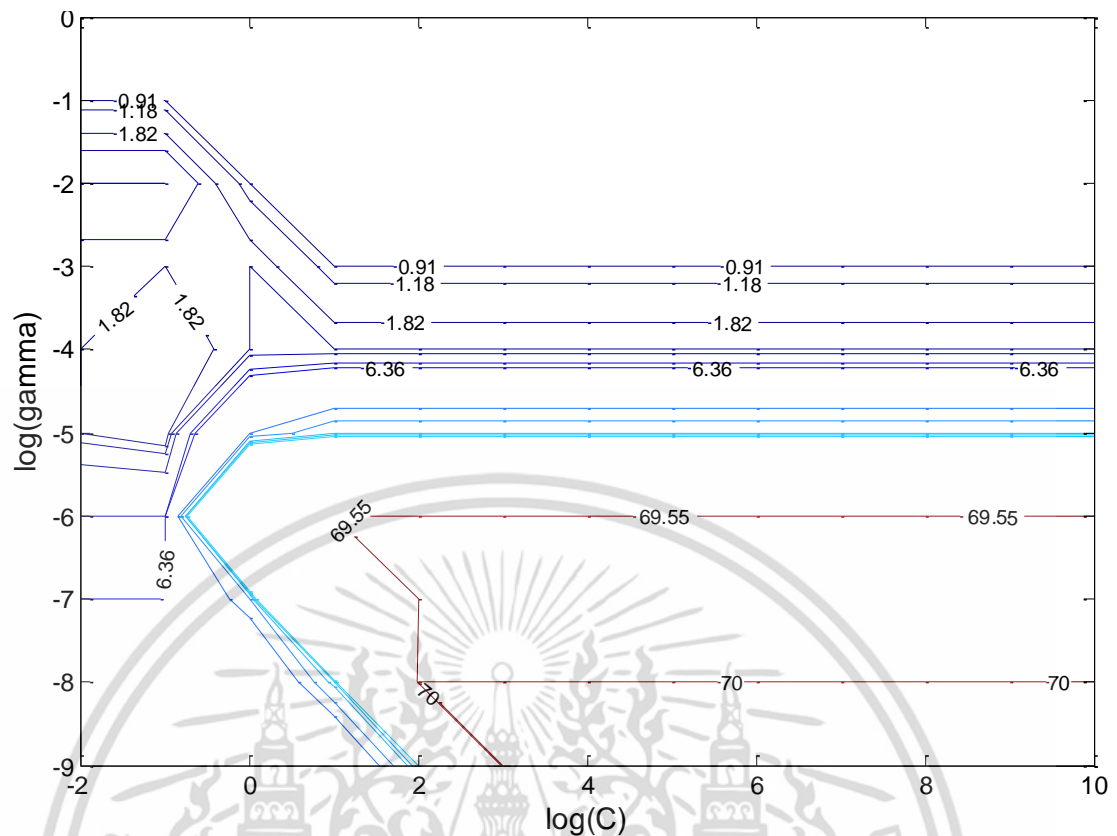
การทดลอง PCA+ SVM โดยใช้เส้นแบ่งแบบ RBF กับฐานข้อมูล 109 คน โดยการปรับค่า C และ γ สามารถแสดงได้ในรูปที่ 4.14



รูปที่ 4.14 ค่าความถูกต้องของการทดลอง PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก เมื่อเส้นแบ่งแบบ RBF กับฐานข้อมูล 109 คน

รูปที่ 4.14 แสดงแผนภาพความสัมพันธ์ระหว่าง $\log(C)$ และ $\log(\gamma)$ โดยมีค่าอยู่ในช่วง -2 ถึง 10 และ -9 ถึง 3 ตามลำดับ โดยตัวเลขที่ปรากฏตรงกลางคือค่าความถูกต้องเมื่อค่า $\log(C)$ และ $\log(\gamma)$ เป็นค่าแตกต่างกันพบว่าค่า $\log(C)$ มีค่าในช่วง 2 ถึง 10 เมื่อ $\log(\gamma)$ มีค่า -7 จะให้ผลการรู้จำภาพใบหน้าบุคคลมีค่าความถูกต้องสูงสุด 93.69 %

และสุดท้ายการทดลอง PCA(Celeb) + SVM โดยใช้เส้นแบ่งแบบ RBF กับฐานข้อมูล 109 คน โดยการปรับค่า C และ γ สามารถแสดงได้ในรูปที่ 4.15



รูปที่ 4.15 ค่าความถูกต้องของการทดลอง PCA (Celeb) + SVM เมื่อเส้นแบ่ง RBF กับฐานข้อมูล 109 คน

รูปที่ 4.15 แสดงแผนภาพความสัมพันธ์ระหว่าง $\log(C)$ และ $\log(\text{gamma})$ โดยมีค่าอยู่ในช่วง -2 ถึง 10 และ -9 ถึง 3 ตามลำดับ โดยตัวเลขที่ปรากฏตรงกลางคือค่าความถูกต้องเมื่อค่า $\log(C)$ และ $\log(\text{gamma})$ เป็นค่าแตกต่างกัน พบว่าค่า $\log(C)$ มีค่าในช่วง 2 ถึง 10 เมื่อ $\log(\text{gamma})$ มีค่า -8 และ $\log(C)$ มีค่าในช่วง 3 ถึง 10 เมื่อ $\log(\text{gamma})$ มีค่า -9 จะให้ผลการรู้จำภาพใบหน้าบุคคลมีค่าความถูกต้องสูงสุด 70%

ผลการทดลองของการใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนกโดยที่เส้นแบ่งแบบ RBF ดังแสดงในรูปที่ 4.6 ถึง 4.9 สามารถสรุปได้ดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 ค่าความถูกต้องในการทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก
เส้นแบ่งแบบเป็น RBF

ค่าความถูกต้อง			
ฐานข้อมูล 9 คน		ฐานข้อมูล 109 คน	
PCA + SVM	PCA (Celeb) + SVM	PCA + SVM	PCA (Celeb) + SVM
92.79%	93.69%	71.36%	70.00%

จากตารางที่ 4.5 พบว่า การดึงลักษณะเด่นด้วย PCA กับฐานข้อมูลขนาดเล็กโดยตรง และการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ มีค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลใกล้เคียงกัน ซึ่งการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อนจากนั้นนำไปใช้กับขนาดเล็กมีค่าความถูกต้องสูงสุดในการรู้จำภาพใบหน้าบุคคล 93.69 % สำหรับฐานข้อมูล 9 คนและสำหรับฐานข้อมูล 109 คน การดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็กโดยตรงให้ค่าความถูกต้องสูงสุด 71.36 %

จากผลการทดลองการดึงลักษณะเด่นด้วย PCA และจำแนกบุคคลโดยใช้ SVM โดยใช้เส้นแบ่งแบบเส้นตรงและเส้นแบ่งแบบRBF สามารถสรุปได้ดังตารางที่ 4.6

ตารางที่ 4.6 ค่าความถูกต้องในการทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก

ประเภท ของเส้น แบ่ง	ค่าความถูกต้อง			
	PCA + SVM		PCA (Celeb) + SVM	
	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน
เส้นตรง	92.79%	71.36%	91.89%	70%
RBF	92.79%	71.36%	93.69%	70%

จากตารางที่ 4.6 ผลการเปรียบเทียบประเภทของเส้นแบ่งที่ส่งผลต่อค่าความถูกต้องในการทดลองการจำแนกด้วย SVM กับฐานข้อมูล 9 คน และ 109 คน พบว่ามีค่าเท่ากันยกเว้นกรณี การดึงลักษณะเด่นจากฐานข้อมูล 9 คนโดยตรง โดยเส้นแบ่งแบบ RBF มีค่าความถูกต้องสูงกว่าแบบเป็นเส้นตรง

จากผลการทดลองสามารถสรุปได้ว่า ค่าความถูกต้องที่ได้จากการดึงลักษณะเด่นจากฐานข้อมูลโดยตรงหรือ การดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อน ไม่มีความแตกต่างอย่างมีนัยสำคัญ หรือกล่าวอีกนัยหนึ่งว่าการใช้ฐานข้อมูลขนาดใหญ่ในการดึงลักษณะเด่นโดยใช้ PCA ก่อน

การนำมาใช้กับฐานข้อมูลขนาดเล็กนั้น ไม่ส่งผลต่อค่าความถูกต้องอย่างมีนัยสำคัญ
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การเปรียบเทียบและวิเคราะห์ผลการทดลอง

ในส่วนนี้จะแสดงการเปรียบเทียบผลการทดลองที่ได้จากระบบรู้จำภาพใบหน้าบุคคลวิธีที่แตกต่างกันดังที่ได้แสดงมาข้างต้น ในส่วนแรกจะแสดงผลการเปรียบเทียบประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลที่แตกต่างกันดังแสดงในตารางที่ 4.7 ในส่วนที่สองจะพิจารณาเวลาที่ใช้ในการทดลองของแต่ละระบบ ซึ่งสามารถแสดงได้ดังตารางที่ 4.8

ตารางที่ 4.7 เปรียบเทียบค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลของวิธีการทั่วไป และการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

ระบบ	ฐานข้อมูล 9 คน	ฐานข้อมูล 109 คน
PCA + kNN	86.49%	68.18%
PCA(Celeb) + kNN	87.39%	67.73%
PCA + SVM	92.79%	71.36%
PCA(Celeb) + SVM	93.69%	70.00%
CNN	94.61%	81.12%
TL-CNN	98.20%	93.19%

จากตารางที่ 4.7 พบว่าค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลของฐานข้อมูล 9 คน มีค่าสูงกว่าฐานข้อมูล 109 คน ในทุกระบบรู้จำภาพใบหน้าบุคคลที่นำมาศึกษา ซึ่งมีสาเหตุหลักมาจาก (1) ฐานข้อมูล 9 คน มีการควบคุมสภาวะในการถ่ายภาพ เช่น กำหนดสถานที่ กำหนดให้เป็นภาพหน้าตรงปกติ เป็นต้น ถึงแม้จะมีการควบคุมสภาวะในการถ่ายภาพแล้ว แต่สิ่งที่ไม่สามารถควบคุมได้คือ สภาวะแสงขณะบันทึกภาพ เนื่องจากบริเวณที่บันทึกภาพมีแสงจากธรรมชาติที่เข้ามาทางหน้าต่าง ซึ่งจะส่งผลต่อภาพที่ได้จากวันเวลาที่แตกต่างกัน (2) ภาพในฐานข้อมูล 109 คน เป็นภาพใบหน้าที่มีอิริยาบถของภาพใบหน้าแตกต่างกัน 6 ลักษณะ และไม่มีการควบคุมสภาวะของการถ่ายภาพ (3) ภาพของแต่ละบุคคลในฐานข้อมูล 9 คน มีจำนวนมากกว่าฐานข้อมูล 109 คน ซึ่งทำให้การเรียนรู้คุณลักษณะของฐานข้อมูล 9 คน ดีกว่าฐานข้อมูล 109 คน

เมื่อพิจารณาเปรียบเทียบระบบรู้จำภาพใบหน้าบุคคลทั่วไปพบว่า การจำแนกด้วยวิธี SVM มีประสิทธิภาพสูงกว่าวิธี kNN เนื่องจากระบบ kNN พิจารณาเฉพาะสมาชิกที่อยู่รอบข้างของข้อมูลที่สนใจเท่านั้น แต่วิธี SVM จะใช้การสร้างเส้นแบ่งเพื่อจำแนกข้อมูล ต่อมาเมื่อพิจารณาการดึงลักษณะ

เด่นด้วยวิธี PCA ของทั้งสองวิธีการจำแนก พบว่าการดึงลักษณะเด่นโดยใช้ฐานข้อมูลขนาดใหญ่จะให้ค่าความถูกต้องแตกต่างกับการดึงลักษณะเด่นโดยใช้ฐานข้อมูลขนาดเล็กโดยตรงอย่างไม่มีนัยสำคัญ

เมื่อเปรียบเทียบการรู้จำภาพใบหน้าบุคคลด้วยระบบทั้งหมดพบว่าระบบ TL-CNN ให้ค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลสูงสุด เนื่องจากการดึงลักษณะเด่นและการจำแนกของระบบ CNN สามารถปรับค่าน้ำหนักในการสอนโดยใช้ค่าความผิดพลาดในการรู้จำภาพใบหน้าบุคคลทำให้ระบบนี้มีประสิทธิภาพในการดึงลักษณะเด่นและการจำแนกสูงกว่าระบบอื่น นอกจากนี้ระบบ TL-CNN ใช้ฐานข้อมูลขนาดใหญ่ในการดึงลักษณะเด่นก่อนจะนำลักษณะเด่นมาใช้กับฐานข้อมูลขนาดเล็ก ค่าน้ำหนักที่ใช้จึงได้มาจากบุคคลจำนวนมากซึ่งสามารถเพิ่มประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลได้

ต่อมาจะกล่าวถึงเวลาที่ใช้ในการดำเนินงาน โดยการใช้ PCA ในการดึงลักษณะเด่น และใช้ kNN และ SVM ในการจำแนกนั้นจะจับเวลาในส่วนของการเตรียมการตั้งแต่เริ่มต้นอ่านภาพจากฐานข้อมูล จนกระทั่งได้ลักษณะเด่นพร้อมต่อการนำไปจำแนก และเวลาในส่วนการทดสอบนั้นเริ่มต้นจากการอ่านไฟล์ภาพที่ใช้ทดสอบจนกระทั่งได้ผลลัพธ์ออกมา และสำหรับวิธีการ CNN และ TL-CNN นั้นเวลาที่ใช้ในการเตรียมการเริ่มต้นตั้งแต่การเตรียมโครงสร้างในการทำงาน จนกระทั่งการสอนครบ 500,000 รอบ โดยรายละเอียดเวลาในการดำเนินงานนั้นแสดงได้ดังตารางที่ 4.8

ตารางที่ 4.8 เวลาที่ใช้ในการดำเนินการระบบรู้จำภาพใบหน้าบุคคล

ระบบ	เวลา (วินาที)			
	ฐานข้อมูล 9 คน		ฐานข้อมูล 109 คน	
	เตรียมการ	ทดสอบ	เตรียมการ	ทดสอบ
PCA + kNN	12.502	0.813	42.599	2.455
PCA (celeb) + kNN	354.59	0.852	364.89	2.532
PCA + SVM	27.493	1.537	59.077	4.396
PCA (celeb) + SVM	917.362	1.565	1,122.95	4.567
CNN	30,136.54	0.0601	30,141.57	0.0604
TL-CNN	30,174.88*	0.0605	30,247.73*	0.0607

* เวลาเตรียมการไม่รวมเวลาของการสอนด้วยข้อมูลขนาดใหญ่

จากตารางที่ 4.8 เวลาในการเตรียมการในการทดลอง TL-CNN เป็นเวลาที่ใช้กับฐานข้อมูลขนาดเล็กเท่านั้น ยังไม่รวมเวลาในการสอนด้วยฐานข้อมูลขนาดใหญ่ซึ่งใช้เวลา 31,439.48 วินาที

และสาเหตุที่เวลาที่ใช้ในการเตรียมการของ CNN และ TL-CNN ใกล้เคียงกันนั้นเนื่องจากใช้จำนวนเอกสารนี้เป็นเอกสารที่ส่งวันละทีละใบการเขียนเพื่อการศึกษาเท่านั้น เมื่ออยู่ให้เห็นระเบียบของหน้ากระดาษไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รอบในการสอน 500,000 รอบเท่านั้น หากพิจารณาโดยภาพรวมพบว่าเวลาของการเตรียมการและการทดสอบของฐานข้อมูล 9 คน มีค่าน้อยกว่าฐานข้อมูล 109 คน เนื่องจากจำนวนรูปของฐานข้อมูล 109 คน มีมากกว่าฐานข้อมูล 9 คน นอกจากนี้พบว่าเวลาของการทดสอบมีค่าน้อยกว่าเวลาของการเตรียมการอย่างมีนัยสำคัญ ซึ่งมีสาเหตุมาจาก (1) รูปของการทดสอบมีจำนวนน้อยกว่าการเตรียมการ (2) ในขั้นตอนการทดสอบไม่มีการดึงลักษณะเด่นของรูปภาพ กล่าวคือลักษณะเด่นของรูปภาพได้มาจากขั้นตอนการเตรียมการ เมื่อมีภาพเข้ามาทดสอบลักษณะเด่นที่ได้จากการเตรียมการจะถูกโปรเจคไปบนภาพทดสอบเพื่อจำแนกบุคคลต่อไป ทั้งนี้เมื่อพิจารณาถึงเวลาของการเตรียมการ พบว่าเวลาเตรียมการของ CNN มีค่ามากที่สุด (มากกว่าระบบ PCA + kNN ประมาณ 2,000 เท่า) ถึงแม้ว่าระบบ CNN จะมีการ์ดจอช่วยประมวลผลก็ตาม และเมื่อเรียงเวลาของการเตรียมการจากน้อยไปมากพบว่าระบบ PCA + kNN มีค่าน้อยที่สุด และเพิ่มขึ้นเป็น PCA + SVM และ CNN ตามลำดับ

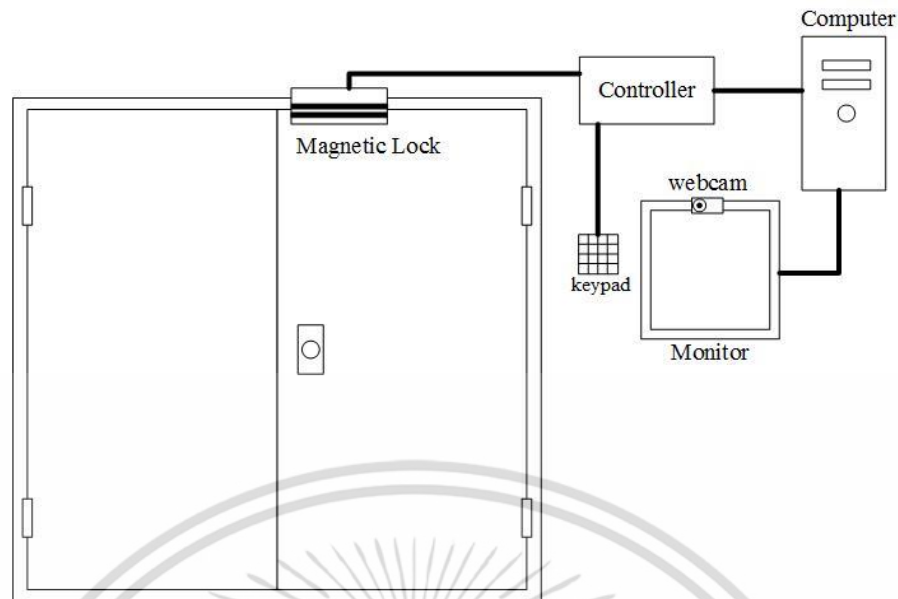
สำหรับเวลาของการทดสอบ พบว่าระบบ CNN มีค่าน้อยที่สุด (น้อยกว่า PCA + SVM ประมาณ 25 เท่า) และเมื่อเรียงเวลาของการทดสอบจากน้อยไปมากพบว่าระบบ CNN มีค่าน้อยที่สุด และเพิ่มขึ้นเป็น PCA + kNN และ PCA + SVM ตามลำดับ จะเห็นได้ว่า ถึงแม้ว่าระบบ CNN จะมีเวลาในการทดสอบต่ำที่สุด แต่อย่างไรก็ตามในขั้นตอนของการเตรียมการนั้น ระบบ CNN ใช้เวลาในการเตรียมการสูงมากซึ่งสามารถพิจารณาเป็นข้อเสียเปรียบของระบบ CNN ได้ แต่ถ้ามองในมุมของประสิทธิภาพและเวลาในการทดสอบแล้ว ระบบ CNN เป็นระบบที่มีความน่าเชื่อถือและน่าสนใจที่จะนำไปพัฒนาและประยุกต์ใช้ในงานต่อไป

4.5 การประยุกต์ใช้งานระบบรู้จำภาพใบหน้าบุคคล

ในส่วนนี้จะกล่าวถึงการนำระบบรู้จำภาพใบหน้าบุคคลไปประยุกต์ใช้ใน 2 ลักษณะด้วยกัน ได้แก่

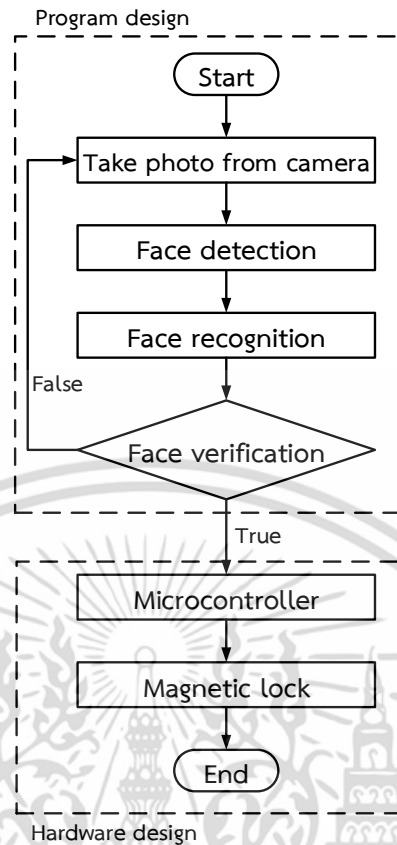
4.5.1 ระบบผ่านเข้าประตูโดยใช้ภาพใบหน้าบุคคล

ระบบผ่านเข้าประตูโดยใช้ภาพใบหน้าบุคคลนี้มีวัตถุประสงค์เพื่อใช้ภาพใบหน้าบุคคลในการยืนยันตัวบุคคลสำหรับการผ่านเข้าห้องหรือผ่านเข้าระบบ อาจนำไปประยุกต์ใช้ร่วมกับการกวดรหัส หรือกุญแจ หรืออาจใช้ระบบนี้เพียงอย่างเดียว ซึ่งภาพรวมของการติดตั้งอุปกรณ์แสดงได้ดังรูปที่ 4.16



รูปที่ 4.16 ภาพรวมการติดตั้งระบบเข้าประตูโดยใช้ภาพใบหน้าบุคคล

จากรูปที่ 4.16 อุปกรณ์ที่จำเป็นต้องใช้มีดังต่อไปนี้ กลอนแม่เหล็กไฟฟ้าสำหรับล็อกประตู คอมพิวเตอร์สำหรับประมวลผล หน้าจอสำหรับแสดงผล กล้องสำหรับถ่ายภาพ กล้องควบคุมสำหรับการควบคุมเปิดปิดประตูตามคำสั่งจากโปรแกรม และสุดท้ายปุ่มกดสำหรับการกรรหัสผ่านเข้าประตู สำหรับกรณีที่ไม่สามารถตรวจจับภาพใบหน้าได้ หรือ กรณีที่บุคคลนั้นมีภาพใบหน้าเก็บไว้ในฐานข้อมูล ซึ่งถือได้ว่าเป็นระบบสำรอง นอกเหนือจากการใช้ภาพใบหน้าบุคคลในการผ่านเข้าประตู โดยการทำงานของระบบนี้สามารถแสดงได้ดังรูปที่ 4.17



รูปที่ 4.17 ภาพรวมการทำงานของระบบผ่านเข้าประตูโดยใช้ภาพใบหน้าบุคคล

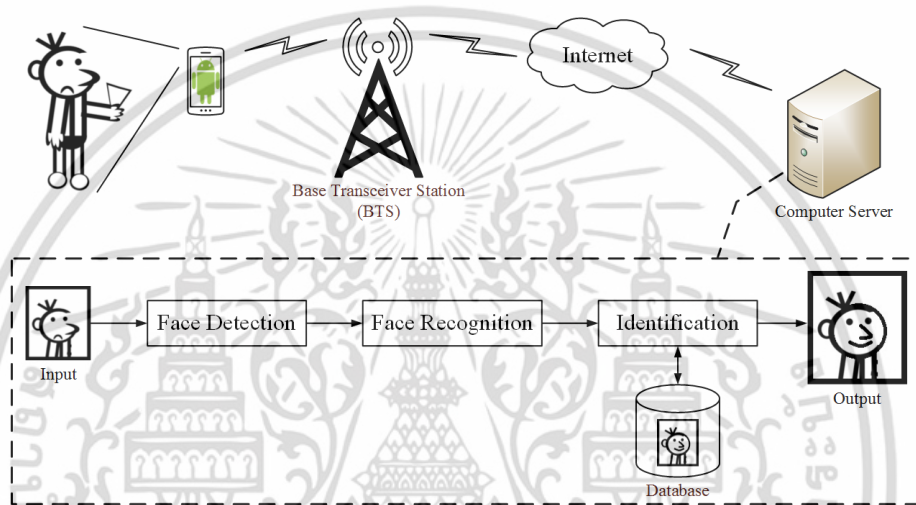
จากรูปที่ 4.17 สามารถแบ่งการทำงานของระบบเข้าประตูออกเป็น 2 ส่วนด้วยกันได้แก่ ส่วนการทำงานของโปรแกรม และส่วนการควบคุมอุปกรณ์ต่างๆ ซึ่งมีกระบวนการทำงานดังต่อไปนี้ เมื่อมีบุคคลต้องการเข้าประตูจะต้องยืนหน้ากล้องและถ่ายภาพ จากนั้นภาพถ่ายจะเข้าสู่กระบวนการตรวจจับภาพใบหน้าบุคคล จากนั้นเข้าสู่การรู้จำภาพใบหน้าบุคคล หากตรงกันกับบุคคลที่อยู่ในฐานข้อมูล ระบบก็จะอนุญาตให้ผ่านเข้าประตู โดยจะไปสั่งไมโครคอนโทรลเลอร์ให้เปิดประตูแบบอัตโนมัติ แต่หากไม่ตรงกันกับบุคคลในฐานข้อมูลระบบก็จะย้อนกลับไปเพื่อรับภาพถ่ายเข้ามาใหม่

จากที่ได้กล่าวมาข้างต้น เมื่อนำระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันมาประยุกต์ใช้งานกับระบบผ่านเข้าประตูโดยใช้ภาพใบหน้า ก็จะทำให้ระบบผ่านเข้าประตูโดยใช้ภาพใบหน้ามีประสิทธิภาพเพิ่มขึ้น และยังใช้เวลาในการทดสอบน้อย แต่อย่างไรก็ตามระบบ CNN นี้มีข้อเสียคือใช้เวลาในการเตรียมการมาก และถ้ามีการเพิ่มสมาชิกเพื่อผ่านเข้าประตู จึงจำเป็นที่จะต้องสอนระบบใหม่อีกครั้ง โดยสรุปแล้วเมื่อนำโครงข่ายประสาทเทียมแบบคอนโวลูชันมาประยุกต์ใช้งานกับระบบผ่านเข้าประตูโดยใช้ภาพใบหน้าบุคคลนั้นจะทำให้มีประสิทธิภาพเพิ่มขึ้น ยกเว้นเวลาในการสอนที่ต้องใช้เวลามากขึ้นตามไปด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

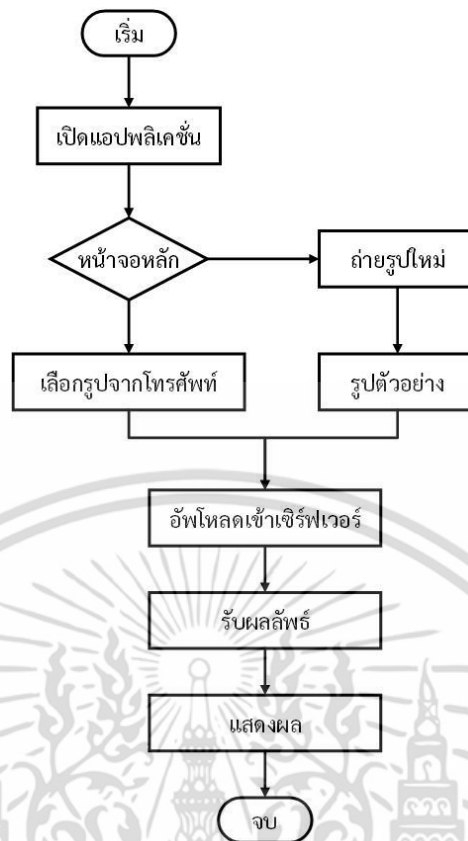
4.6.2 แอปพลิเคชันยืนยันตัวตนบุคคลบนโทรศัพท์มือถือ

การทำงานของแอปพลิเคชันบนโทรศัพท์มือถือสำหรับการยืนยันตัวตนนั้นเริ่มต้นจากถ่ายภาพบุคคลจากแอปพลิเคชันบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์ จากนั้นภาพที่ได้จะถูกส่งผ่านเครือข่ายอินเทอร์เน็ตไปยังเซิร์ฟเวอร์ โดยการทำงานของระบบนั้นจะอยู่ในเซิร์ฟเวอร์ทั้งหมด เมื่อภายในเซิร์ฟเวอร์ทำการประมวลผลเสร็จแล้ว ก็จะส่งกลับมายังโทรศัพท์มือถือเพื่อแสดงผลยังส่วนของผู้ใช้งานต่อไป ซึ่งภาพรวมการทำงานของระบบนี้สามารถแสดงได้ดังรูปที่ 4.18



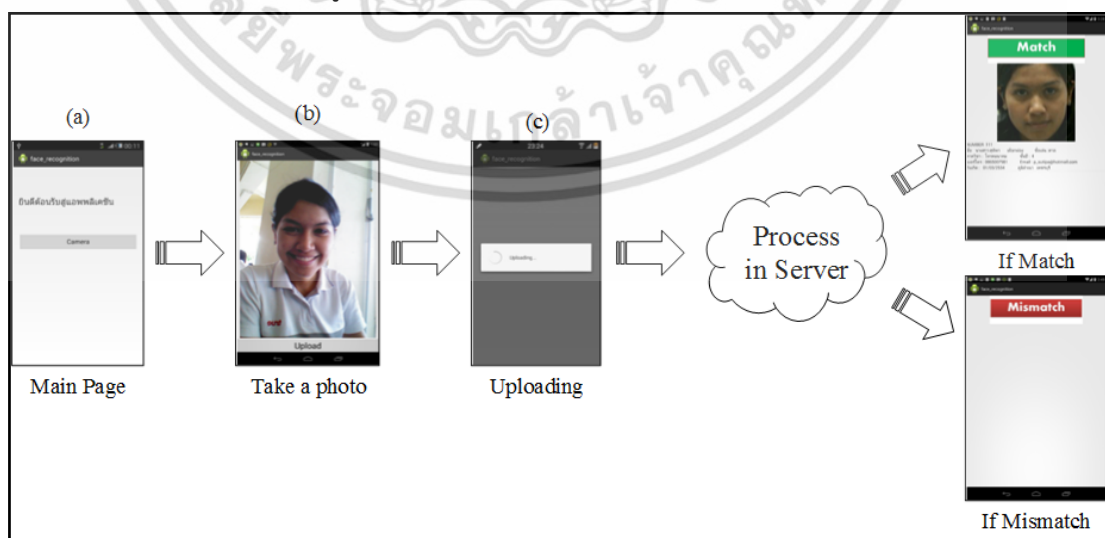
รูปที่ 4.18 ภาพรวมของระบบแอปพลิเคชันยืนยันตัวตนบุคคล

การทำงานของแอปพลิเคชัน จะถ่ายรูปและอัปโหลดขึ้นไปบนเซิร์ฟเวอร์ และเชื่อมต่ออีกครั้งเพื่อรับผลลัพธ์หากถูกต้องจะส่งข้อมูลของบุคคลมาแสดงด้วย แผนผังแสดงการทำงานของแอปพลิเคชันแสดงได้ดังรูปที่ 4.19 และขั้นตอนการทำงานของแอปพลิเคชันแสดงได้ดังรูปที่ 4.20



รูปที่ 4.19 แผนผังการทำงานของแอปพลิเคชัน

จากรูปที่ 4.19 ขั้นตอนการทำงานสามารถอธิบายได้ดังนี้ เริ่มต้นจากการเปิดแอปพลิเคชันขึ้นมา จากนั้นเมื่อเข้าสู่หน้าจอหลักจะมี 2 ทางเลือกด้วยกันได้แก่ การเลือกภาพถ่ายจากคลังภาพในโทรศัพท์หรือ การเลือกเปิดกล้องเพื่อถ่ายภาพใหม่ เมื่อเลือกรูปภาพได้แล้วกด Upload เพื่ออัปโหลดภาพเข้าสู่แอปพลิเคชัน จากนั้นรอรับผลลัพธ์กลับมาเพื่อแสดงผลยังหน้าจอของผู้ใช้งาน โดยภาพขั้นตอนการทำงานแสดงได้ดังรูปที่ 4.20



รูปที่ 4.20 ขั้นตอนการทำงานของแอปพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.20 เมื่อเข้ามายังแอปพลิเคชันจะพบกับหน้าหลัก (a) จากนั้นจะเลือกเปิดกล้องเพื่อถ่ายภาพ เมื่อถ่ายภาพแล้วจะเข้าสู่หน้า Preview เพื่อดูภาพที่เพิ่งถ่ายไป และจะมีปุ่มให้กดเพื่ออัปโหลดเข้าสู่เซิร์ฟเวอร์ (b) ในหน้า Uploading จะเป็นหน้าที่รอระหว่างที่ภาพถูกส่งไปประมวลผลในเซิร์ฟเวอร์ (c) จากนั้นเมื่อเซิร์ฟเวอร์ประมวลผลเสร็จจะส่งกลับมาแสดงผลยังหน้าจอ โดยหากภาพที่ถ่ายตรงกับภาพในฐานข้อมูลจะขึ้นว่า Match และจะแสดงข้อมูลขึ้นมา แต่หากไม่ตรงกันกับบุคคลในฐานข้อมูลจะขึ้นว่า Mismatch

สำหรับแอปพลิเคชันยืนยันตัวตนบุคคลบนสมาร์ตโฟนสามารถนำไปใช้งานได้หลายลักษณะเช่น การยืนยันตัวตนบุคคลสูญหาย กรณีนำไปใช้ในการติดตามบุคคลสูญหาย การใช้งานในการตรวจสอบประวัติอาชญากรรม เป็นต้น ระบบโครงข่ายประสาทเทียมแบบคอนโวลูชันสามารถนำไปประยุกต์ใช้งานกับแอปพลิเคชันยืนยันตัวตนบุคคลบนสมาร์ตโฟนได้ เนื่องจากระบบโครงข่ายประสาทเทียมแบบคอนโวลูชันให้ประสิทธิภาพการรู้จำภาพใบหน้าสูงเมื่อฐานข้อมูลมีขนาดใหญ่และใช้เวลาในการทดสอบน้อยกว่าเมื่อเปรียบเทียบกับระบบอื่น แต่อย่างไรก็ตามเวลาในการเตรียมการสำหรับระบบโครงข่ายประสาทเทียมแบบคอนโวลูชันมีค่ามากกว่าระบบอื่น

บทที่ 5

สรุปงานวิจัยและข้อเสนอแนะ

ในบทที่ 5 นี้จะกล่าวถึงการสรุปเนื้อหาของวิทยานิพนธ์ และสรุปผลการอภิปรายการทดลอง รวมทั้งข้อเสนอแนะและแนวทางการนำระบบรู้จำภาพใบหน้าบุคคลโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

5.1 สรุปเนื้อหาวิทยานิพนธ์

วิทยานิพนธ์นี้นำเสนอการรู้จำภาพใบหน้าบุคคลด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks) สำหรับฐานข้อมูลขนาดเล็ก (9 คน และ 109 คน) โดยเปรียบเทียบวิธีการรู้จำภาพใบหน้าบุคคลแบบทั่วไปกับวิธี CNN และยังสามารถทำการทดลองใช้ CNN แบบถ่ายโอนการเรียนรู้เพื่อเพิ่มประสิทธิภาพการทำงานของระบบรู้จำภาพใบหน้าบุคคล ซึ่งสามารถสรุปได้ดังต่อไปนี้

1) สำหรับการจำแนกด้วยวิธี kNN และวิธี SVM โดยใช้ PCA ในการดึงลักษณะเด่นจากฐานข้อมูลขนาดเล็กโดยตรงและการดึงลักษณะเด่นจากฐานข้อมูลขนาดใหญ่ก่อนมีค่าความถูกต้องในการรู้จำภาพใบหน้าบุคคลที่ใกล้เคียงกัน

2) หากเรียงลำดับประสิทธิภาพในการทำงานของระบบรู้จำภาพใบหน้าบุคคลจากมากไปน้อยพบว่า TL-CNN > CNN > PCA + SVM (\approx PCA (Celeb) + SVM) > PCA + kNN (\approx PCA (Celeb) + kNN)

3) เมื่อเปรียบเทียบการรู้จำภาพใบหน้าบุคคลด้วยวิธี CNN และ CNN แบบถ่ายโอนการเรียนรู้ (TL-CNN) พบว่า TL-CNN ให้ค่าความถูกต้องสูงกว่า เนื่องจากค่าน้ำหนักที่ได้จากฐานข้อมูลขนาดใหญ่ซึ่งมีภาพในฐานข้อมูลจำนวนมากจึงทำให้ประสิทธิภาพของการเรียนรู้คุณลักษณะของ CNN เพิ่มขึ้น และเมื่อนำค่าน้ำหนักที่ได้จากฐานข้อมูลขนาดใหญ่ไปกำหนดเป็นค่าน้ำหนักเริ่มต้นของฐานข้อมูลขนาดเล็กจึงช่วยเพิ่มประสิทธิภาพของระบบรู้จำภาพใบหน้าบุคคลได้ เพื่อให้การคำนวณมีความถูกต้องและเสถียรภาพในการคำนวณสูง อัตราการเรียนรู้เริ่มต้นจะต้องมีค่าน้อย นอกจากนี้คุณภาพของภาพในฐานข้อมูลเป็นอีกหนึ่งปัจจัยที่ส่งผลต่อประสิทธิภาพของการรู้จำภาพใบหน้าบุคคล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) เมื่อพิจารณาเวลาในการทำงานของระบบรู้จำภาพใบหน้าบุคคลพบว่า การใช้ CNN และ TL-CNN ใช้เวลาในการสอนมากที่สุด และใช้เวลาในการทดสอบน้อยที่สุดเมื่อเทียบกับวิธีการรู้จำภาพใบหน้าบุคคลอื่น

5.2 ข้อเสนอแนะและแนวทางการพัฒนาต่อ

การเพิ่มประสิทธิภาพการรู้จำภาพใบหน้าบุคคลด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันนั้น อาจทำได้โดยการปรับปรุงโครงสร้าง โดยการเพิ่มหรือลดจำนวนชั้นในโครงสร้าง หรือการเลือกชุดของฐานข้อมูลที่ใช้ในการฝึกฝนจนได้มีความเหมาะสม โดยอาจจะเลือกใช้ฐานข้อมูลภาพใบหน้าของคนเอเชีย เพื่อให้เหมาะกับการนำไปประยุกต์ใช้มากขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] F. Galton. “Personal Identification” **Scientific American Supplement**, August 1888. pp. 1053
- [2] W. W. Bledsoe. “The model method in facial recognition” **Technical report**, 1964.
- [3] A. Golstein, L. Harmon, and A. Lest. “Identification of human faces” **Proc. IEEE**, 1971. pp. 59:748–760
- [4] M. Fischler and R. Elschlager. “The representation and matching of pictorial structures” **IEEE Trans. on Computers**, 1973. pp. C-22(1):67 – 92
- [5] T. Kenade. “**Picture Processing System by Computer Complex and Recognition of Human Faces**” PhD Thesis of Kyoto University, 1973
- [6] L. Sirovich and M. Kirby. “Low-dimensional procedure for the characterization of human faces” **Journal of the Optical Society of America A - Optics, Image Science and Vision**, March 1987. pp. 4(3):519–524
- [7] M. Turk and A. Pentland. “Eigenfaces for Recognition” **Journal of Cognitive Neuroscience**, vol. 3, no 1, 1991. pp. 71-86
- [8] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. “A Neural Network Identifies Sex from Human Faces” **NIPS**, 1990. pp. 2
- [9] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. “Face recognition: A convolutional neural network approach” **IEEE Trans. on Neural Networks**, 1997. pp. 8:98–113
- [10] P. J. Phillips. “Support vector machines applied to face recognition” **Advances in Neural Information Processing Systems**, 1999.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [11] F. Samaria and A. Harter. “Parameterisation of a Stochastic Model for Human Face Identification” **Proc. 2nd IEEE Workshop on Applications of Computer Vision**, December 1994.
- [12] M. Agarwal, H. Agrawal, N. Jain, and M. Kumar. “Face Recognition using Principle Component Analysis, Eigenface and Neural Network” **Proc. International Conference on Signal Acquisition and Processing (ICSAP)**, February 2010. pp .310 – 314
- [13] J. M. Kim and M. Kang. “A Study of Face Recognition using the PCA and Error Back-Propagation” **Proc. International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)**, August 2010. pp. 241 – 244
- [14] Xianwei and C. Guolong. “Face Recognition Based on PCA and SVM” **Proc. Symposium on Photonics and Optoelectronics (SOP)**, May 2012. pp. 1 – 4
- [15] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. “Handwritten digit recognition with a back-propagation network” **Proc. Advances in Neural Information Processing Systems**, 1990. pp. 396-404
- [16] Y. Sun, X. Wang, and X. Tang. “Deep learning face representation from predicting 10,000 classes” **Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 2014. pp. 1891–1898
- [17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. “Deep Face Recognition” **Proc. of the British Machine Vision Conference**, 2015. pp. 41.1-41.12
- [18] Y. Sun, X. Wang and X. Tang. “Hybrid Deep Learning for Face Verificaion” **Proc. IEEE International Conference on Computer Vision (ICCV)**, 2013. pp. 1489-1496
- [19] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales. “When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition” [Online]. Available: arxiv.org/abs/1504.02351, 2015.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [20] C. Cortes and V. Vapnik. “Support vector networks” **Machine Learning**, vol. 20, no 3, 1995. pp. 273–297
- [21] informy. “**Support vector machines**”. [Online]. Available: <http://informy.tumblr.com/post/17937920903/support-vector-machines>. 2016.
- [22] Z. P. Fu, Y. N. ZhANG, and H. Y. Hou. “Survey of deep learning in face recognition” **Proc. IEEE International Conference on Orange Technologies**, 2014. pp. 5-8
- [23] C. Sri, N. Pathirage, L. Li, W. Liu, and M. Zhang. “Stacked Face De-noising Auto Encoders For Expression-Robust Face Recognition” **Proc. International Conference on Digital Image Computing: Techniques and Applications (DICTA)**, 2015. pp. 1-7
- [24] A. Gibiansky. “**Convolutional Neural Networks**”. [Online]. Available: <http://andrew.gibiansky.com/blog/machine-learning/convolutional-neural-networks/>. 2016.
- [25] S. Duffner. “**Face Image Analysis With Convolutional Neural Networks**” Doctoral degree of Faculty of Applied Sciences. 2007.
- [26] G. Xavier, B. Antoine, and B. Yoshua. “Domain adaptation for large-scale sentiment classification: A deep learning approach” **Proc. The 28th International Conference on Machine Learning**, 2011.
- [27] X. Zhang, L. Zhang, X. J. Wang, H. Y. Shum. “Finding Celebrities in Billions of Webpages” **IEEE Transaction on Multimedia**, vol.14, no 4, 2012. pp. 995-1007
- [28] T. Ahonen, A. Hadid, and M. Pietikainen. “Face Recognition with Local Binary Patterns” **Proc. European Conference on Computer Vision**, 2004. pp. 469-481
- [29] Jia et al. “**Caffe: Convolutional Architecture for Fast Feature Embedding**” [Online]. Available: arXiv preprint arXiv:1408.5093, 2014.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Proceedings of ISMAC 2015



An Implementation of Face Recognition System Applications

S. Pauekpong*, S. Chivapreecha†, I. Methasate‡, P. Kittipanya-ngam§

* †Department of Telecommunication Engineering, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

‡National Electronics and Computer Technology Center
National Science and Technology Development Agency, Pathumthani 12120, Thailand

§Innovation Center True Corporation, Bangkok 10130, Thailand

*s7601138@kmitl.ac.th, †sorawat@telecom.kmitl.ac.th, ‡ithipan.methasate@nectec.or.th, §panachit@gmail.com

Abstract—This paper presents an implementation of face recognition system applications. The face recognition system consists of face detection using Viola and Jones method, face recognition engine using Principal Component Analysis: PCA, and identification using minimum Euclidean distance among faces in database. However normalized cross correlation will be calculated and used to be criteria to verify an unknown face which is not appear in database. Moreover, two applications of face recognition system are proposed: 1. Face access control and 2. Face verification using smartphone via internet. First application, face access control, the door will be opened when an input image that taken from camera is one of person who their face images are in database. For second application, users can use an application on smartphone to take photo of target person and then automatically send to the server via internet. At server side, face image of target person will be processed using face recognition system and database. Finally, send back information to smartphone if target person is matched with someone in database.

Keywords—face recognition system; face access control; biometric

I. INTRODUCTION

Nowadays the authentication technologies have been applied to the various applications. One of the popular method is a face recognition, which is widely used as the airport security system, identification system, missing persons searching, criminal arresting, etc. Generally, this technology has been widely developed as an application on Android smartphone. Because, today, these smartphones become the necessary communication devices, which are easier to access the face recognition.

The face recognition research can be divided into the several interest issues. The first interest subject is the problem solving of image illumination by using histogram equalization [1]. The second interest topic is the feature extraction via principal component analysis (PCA) [2], which is the basic technique to obtain the dominant feature of the data. The last issue is the applications of face recognition, for example, the face recognition implementation for client server mobile application using PCA [3], the application in microcontroller [4], etc.

Although, the face recognition on smartphone is continuously developed, the processor and memory of the smartphone

are still limited. Face recognition system requires an effective processor and a large volume of the memory, which lead to the difficulty in smartphone application development. Therefore, in this work, the face recognition processing, including face detection, face recognition, and identification process, is done by the server. While, the smartphone is only adopted to take a photo and send the photo to the server. Finally, the server send the result back to the smartphone, which is only employed to display the result.

The face recognition of the smartphone application is very useful in many applications, such as the missing persons finding, the offender tracking, etc. In addition, this system can be also adopted to check the criminal background. Further, this application is not only valid for the mobile phone but also the personal computer with internet connection.

According to the importance of face recognition, in this paper, the application of face recognition, including face access control and face verification by using internet smartphone, were studied. The system was mainly developed by using the C++ language. Further, the C language and JAVA were also conducted to develop the microcontroller and Android application, respectively.

The rest of this paper is organized as the follows: background of face detection, face recognition, face identification, and normalized cross correlation are introduced in section II. Section III represents the proposed system of this work. The experimental results are shown in Section IV. Finally, the conclusions are drawn in Section V.

II. BACKGROUND

Our methods are fundamental algorithms, including face detection, face recognition, identification using Euclidean distance, and unknown face verification. These algorithms are explained below.

A. Face Detection

Face detection is the technique to detect face region in an image. The Viola and Jones [5] method is employed to this system. The flow diagram of this method is shown in Fig. 1. This technique combine three different procedures. The image is scanned by using integral image method for the several time.

The filter is rectangle and can be divided into 2 parts, including black and white. This filter is known as Haar-Like Feature method. The face region is considered by using AdaBoost cascade classifier. This method detect the face by rejecting the non-face region.



Fig. 1. The flow diagram of Viola and Jones face detection.

B. Face Recognition

There are many popular face recognition method, such as principal component analysis (PCA), linear discriminant analysis (LDA or Fisherfaces) and elastic bunch graph matching (EBGM), etc. In this paper, the principal component analysis or Eigenface is adopted as the face recognition method. This is the basic method for face recognition, which contain the following steps:

1. Collect M face images with the size of $N \times N$ pixels store in set S as defined in (1)

$$S = \{\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M\} \quad (1)$$

where Γ_n is each image that transformed into a vector. ($n = 1, 2, \dots, M$)

2. Find mean image (Ψ) as defined in (2)

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \quad (2)$$

3. Find difference (Φ) between input image and mean image (Ψ) as defined in (3)

$$\Phi_i = \Gamma_i - \Psi \quad (3)$$

4. Calculate the covariance matrix (C) as defined in (4)

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = \mathbf{A} \mathbf{A}^T \quad (4)$$

5. Define $L = \mathbf{A}^T \mathbf{A}$, find eigenvalue and eigenvector of L as (5)

$$L \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (5)$$

where \mathbf{v}_i is eigenvector and λ_i is eigenvalue
 Then, multiply by \mathbf{A} both sides of the equation
 $\mathbf{A} L \mathbf{v}_i = \mathbf{A} \lambda_i \mathbf{v}_i$ or $\mathbf{A} \mathbf{A}^T \mathbf{A} \mathbf{v}_i = \mathbf{A} \lambda_i \mathbf{v}_i$
 Substitute $C = \mathbf{A} \mathbf{A}^T$, that is $C \mathbf{A} \mathbf{v}_i = \mathbf{A} \lambda_i \mathbf{v}_i$

6. Choose eigenvector which has top eigenvalues k values
 7. Find eigenface (\mathbf{U}) as defined in (6)

$$\mathbf{U} = \mathbf{A} \mathbf{V} \quad (6)$$

where \mathbf{V} is matrix of eigenvector $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k]$

8. Find coefficient of image in the database (Ω) via projection face image into eigenface (\mathbf{U}) as defined in (7)

$$\Omega = \mathbf{U}^T \mathbf{A} \quad (7)$$

$$\text{where } \Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix}$$

9. Get the test images (Γ), which is the same sizes of images in the database subtract with mean image (Ψ) multiplied by the Eigenfaces. The result is the coefficient of the test image (Ω_p) as defined in (8)

$$\Omega_p = \mathbf{U} [\Gamma - \Psi] \quad (8)$$

$$\text{where } \Omega_p = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix}$$

C. Identification using Eucliden distance

This process measures Euclidean distance to obtain the minimum distance between test image and images in database as defined in (9)

$$\epsilon_k = \|\Omega_p - \Omega\| \quad (9)$$

where $1 \leq k \leq M$

D. Unknown face verification

If the unknown face image is specified by the face image in the training set, this can be considered as an error of the system. then, the threshold calculated by normalize cross correlation (NCC) is conducted to prevent this error. If the value of NCC is less than 0.75, the system decide the test image as an unknown face image.

$$NCC(I_1, I_2) = \frac{\sum_{i=1}^M [I_1(x_i) - \bar{I}_1] [I_2(x_i) - \bar{I}_2]}{\sqrt{\sum_{i=1}^M [I_1(x_i) - \bar{I}_1]^2 [I_2(x_i) - \bar{I}_2]^2}} \quad (10)$$

where $I_1(x_i)$ is intensity of test image.

$I_2(x_i)$ is intensity of database face image.

\bar{I}_1 is mean intensity of I_1 ($\bar{I}_1 = \frac{1}{N} \sum_{i=1}^M I_1(x_i)$)

\bar{I}_2 is mean intensity of I_2 ($\bar{I}_2 = \frac{1}{N} \sum_{i=1}^M I_2(x_i)$)

III. PROPOSED SYSTEM

This paper proposes applications of face recognition in 2 ways, including face verification via mobile application and face access control.

The face verification via mobile application, its procedure is shown as the following. First, the image is captured by using mobile application, then the image is transferred to the server via internet. The image is evaluated by face recognition system in the server. Finally, the result is transferred from the server to mobile phone via internet. The overview of the system is shown in Fig. 2.

For face access control, first, the webcam or camera, which is installed in front of the room, capture an image of person, who want to access the room. Then, the image is identified by

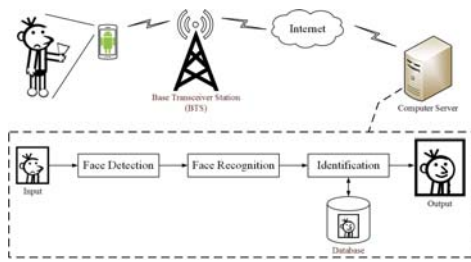


Fig. 2. The overview of mobile application for verification.

face detection, face recognition, identification and unknown verification. If that person is available in the database, the program will send data to the microcontroller to open the door automatically. In contrast, if that person is unavailable in the database, the access is denied and message is displayed on the monitor and the door is close. The overview of face access control system is shown in Fig. 3.

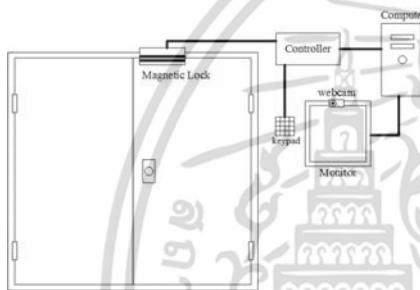


Fig. 3. Face access control system.

A. Design of face recognition system

Face recognition system is shown as the following steps. First, Viola and Jones method is adopted to perform the face detection by cropping the face image from the photo. Then, for face recognition process, this face image is recognized by using PCA algorithm. Further, in identification process, this face image is comparing with the face images in the database. The image with the minimum Euclidean distance is considered as the result. Finally, the normalized cross correlation (NCC) is adopted to check the person who in test image is available in the database or not.

B. Design of mobile application

For mobile application, first, the photo is taken by the application. Then, the photo is transferred to the server via internet. The photo is identified on the server by using face recognition system. Finally, the mobile receive the result from the server and display. If the photo match the person in the database, the server will send the personal information to the mobile phone. However, if the result is mismatch, the mismatch message will be displayed on the mobile phone. These processes can be summarized as shown in Fig. 4.

In Fig. 4, the main page as shown in Fig. 4(a) is used to open the camera and take a photo. The photo is previewed as

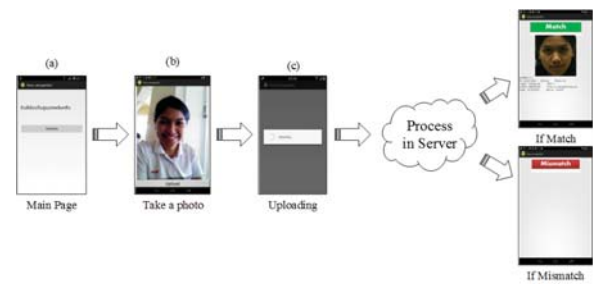


Fig. 4. Working process of the mobile application.

shown in Fig. 4(b). In the uploading page as depicted in Fig. 4(c), the photo is uploaded to the server. Then, the photo is identified by face recognition system in the server. Finally, the server send the result back to the mobile phone.

C. Design of face access control system

The face access control system is divided into 2 parts, including program design and hardware design. The program design is face recognition system as mentioned in subsection A. For hardware design, if program sent data to microcontroller, the magnetic lock is controlled to open the door automatically. The flow diagram show in Fig. 5.

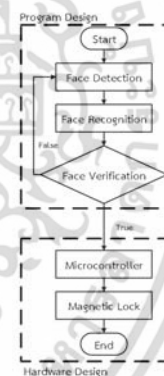


Fig. 5. The process of face access control system.

The required equipments for the hardware design can be shown as the follows: (i) microcontroller using Arduino UNO R3, (ii) Relay, (iii) Keypad, and (iv) Magnetic lock. The device connectivity is shown in Fig. 6. If the person in the photo is in the database, the program will send the data to microcontroller. Then the microcontroller controls the relay to cut off an electricity supply of magnetic lock. Finally, the door is open.

D. Experimental setup

In order to test an accuracy of the face recognition system, the different number of images in the database and different face actions were conducted. The number of face images in the database were 10, 20, 50, 80, and 100 images, respectively. Moreover, the different action of face images were normal face, smile face, close eye, open mouth and other action. The

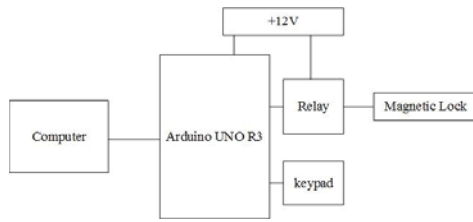


Fig. 6. Device connectivity.

example of face action are shown in Fig. 7. The database is the normal face of the students in faculty of engineering, King Mongkut's Institute of Technology Ladkrabang.



Fig. 7. Face image in difference actions.

IV. EXPERIMENTAL RESULTS

A. The system performance

From Fig. 8 shows the comparison between the number of images and accuracy percentage for different face actions. The results revealed that the normal face exhibit the highest accuracy percentage as compared to other actions because the face action in database is normal face. These results can be implied that the accuracy is depended on the face action in database. Further, the accuracy is found to decrease with increasing number of face images.

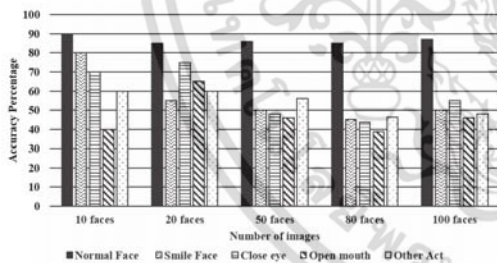


Fig. 8. Comparison of the data in a bar chart with accuracy percentage.

B. Result of uploading photo from application to server

From Fig. 9 represents the photo uploading process obtained by Wireshark, which indicate the connectivity action during uploading the photo. It can be explained as shown in the following steps

(1) The smartphone uploads the photo to the server via HTTP protocol with POST command as shown in Fig. 9. While, the photo is recognized on the server the mobile screen is represent in Fig. 10(a)

(2) When the face recognition process is complete, the smartphone sent the GET command to receive the result from the server.

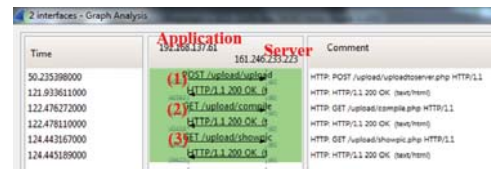


Fig. 9. Time flow while sniffing package

(3) If the result is match, the smartphone send the GET command to the server. Then, the server response to the command by sending the photo and personal information to smartphone as shown in Fig. 10(b). In contrast, if the result is mismatch, the mismatch message is occurred on the mobile screen as depicted in Fig. 10(c).



Fig. 10. (a) Uploading screen, (b) Match screen, (c) Mismatch screen

V. CONCLUSIONS

This paper proposes face recognition system and its applications, including mobile application for face verification and face access control system. The experimental results revealed that the accuracy is found to decrease with increasing the number of face. For different face actions, the highest accuracy is obtained by the normal face for the different number of images because the face action in database is normal face. In other words, the accuracy is depended on the face action in database. Both applications can work well properly.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the NSTDA University Industry Research Collaboration (NUI-RC-P44-22-57-022M). Moreover, we also acknowledge the partial financial support from True Lab (TU/CM2/2014-0978/5), Innovation Center True Corporation.

REFERENCES

- [1] D. H. Liuc, K. M. Lam, and L. S. Shen, Illumination invariant face recognition. *Pattern Recognition*, 38:1705-1716,2005.
- [2] M. Turk and A. Pentland, Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1)-70-86, 1991.
- [3] E. Kremic, A. Subasi, and K. Hajdarevic, Face Recognition Implementation for Client Server Mobile Application using PCA. In *Proc. ITI*, 2012.
- [4] H. Hassan, R. A. Baker, and A. T. F. Mokhtar, Face Recognition Based on Auto-Switching Magnetic Door Lock System Using Microcontroller. In *Proc. ICSET*, 2012.
- [5] P. Viola and M. J. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, *Computer Vision and Pattern Recognition*, 511-518, 2001.

ภาคผนวก ข
โปรแกรมที่ใช้ในวิทยานิพนธ์ การติดตั้ง Caffe และ Source code



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมที่เกี่ยวข้องในงานวิจัย

โปรแกรมที่เกี่ยวข้องในงานวิจัยนี้จะแบ่งออกเป็น 2 ส่วนได้แก่ ระบบรู้จำภาพใบหน้าบุคคล โดยใช้ CNN และการรู้จำภาพใบหน้าบุคคลด้วยวิธีการทั่วไป ซึ่งโปรแกรมที่ใช้ในการทดลองทั้ง 2 ส่วนนั้นมีการดำเนินงานที่แตกต่างกัน ซึ่งสามารถแสดงได้ดังต่อไปนี้

- Caffe [29]

ในส่วนของการทดลองการทำงานของ CNN นั้นจะทำบนระบบปฏิบัติการ Ubuntu ซึ่งใช้เครื่องมือชื่อ Caffe ซึ่งพัฒนาโดย Berkeley Vision and Learning Center (BVLC) ซึ่งการทำงานของ Caffe นั้นจะใช้การ์ดจอซึ่งมี CUDA ในการช่วยประมวลผล เพื่อให้การประมวลผลใช้เวลาลดลงเมื่อเทียบกับการใช้ CPU ในการทำงาน

- Microsoft Visual Studio 2012

Microsoft Visual Studio 2012 เป็นเครื่องมือที่พัฒนาโดยไมโครซอฟท์ทำงานบนระบบปฏิบัติการไมโครซอฟท์วินโดวส์ ซึ่งช่วยในการพัฒนาโปรแกรมคอมพิวเตอร์ โดยในงานวิจัยนี้จะใช้ภาษา c++ เป็นหลักในดำเนินการของโปรแกรม อีกทั้งยังใช้ openCV ซึ่งเป็นไลบรารีที่ช่วยในการประมวลผลเกี่ยวกับคอมพิวเตอร์วิทัศน์ (Computer vision) ซึ่งทั้งหมดนี้จะช่วยในการพัฒนาโปรแกรม โดยการทำงานในส่วนนี้จะใช้ CPU ในการทำงานเพียงอย่างเดียว ซึ่งการทดลองที่พัฒนาด้วยภาษา c++ รวมกันกับ openCV มีดังต่อไปนี้

1. การทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ kNN ในการจำแนก
2. การทดลองใช้ PCA ในการดึงลักษณะเด่นและใช้ SVM ในการจำแนก

Caffe

Caffe เป็น deep learning framework ซึ่งถูกพัฒนาโดย Berkeley Vision and Learning Center (BVLC) โดยเหตุผลที่เลือก Caffe มาใช้ในการดำเนินงาน

- โครงสร้างของ Caffe เอื้อต่อการนำไปประยุกต์ใช้ หรือการพัฒนานวัตกรรม โดย โมเดล และประสิทธิภาพการทำงานถูกกำหนดโดยการปรับตั้งค่าแทนการเขียนโปรแกรมแบบซับซ้อน (hard-coding) สามารถเลือกการทำงานได้ระหว่างการทำงานโดยอาศัย CPU หรือ GPU
- มีนักพัฒนาร่วมมือกันเพื่อให้ Caffe พัฒนาได้อย่างรวดเร็ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ความเร็วในการทำงานของ Caffe เหมาะสำหรับการทำการทดลองในการวิจัย หรือการนำไปประยุกต์ใช้ในอุตสาหกรรม
- มีกลุ่มนักพัฒนาที่ช่วยกันแบ่งปันสิ่งที่ได้ค้นพบใหม่ๆและการร่วมกันแก้ปัญหาที่พบเจอ

การติดตั้ง Caffe

โดยระบบปฏิบัติการที่แนะนำสำหรับการติดตั้ง Caffe นั้นได้แก่ Ubuntu, OS X โดยในงานวิจัยนี้ได้ติดตั้ง Caffe บนระบบปฏิบัติการ Ubuntu 14.04 LTS

1) อัปเดตและอัปเดตคอมพิวเตอร์

```
>> sudo apt-get update && sudo apt-get upgrade
>> sudo apt-get install build-essential
```

2) ติดตั้ง CUDA และ cuDNN (กรณีที่คอมพิวเตอร์ไม่มีการ์ดจอ NVIDIA ให้ข้ามไปขั้นตอนที่ 3)

```
>> wget http://developer.download.nvidia.com/compute/cuda/7_0/Prod/local_installers/cuda_7.0.28_linux.run
>> chmod +x cuda_7.0.28_linux.run && mkdir nvidia_installers
>> ./cuda_7.0.28_linux.run -extract=`pwd`/nvidia_installers
>> cd nvidia_installers
>> sudo ./NVIDIA-Linux-x86_64-346.46.run
>> sudo modprobe nvidia
>> sudo apt-get install build-essential
>> sudo ./cuda-linux64-rel-7.0.28-19326674.run
>> sudo ./cuda-samples-linux-7.0.28-19326674.run
```

อัปเดต path variables ในไฟล์ ~/.bashrc โดยการเพิ่มคำสั่งต่อไปนี้

```
>> export PATH=$PATH:/usr/local/cuda-7.0/bin
>> export LD_LIBRARY_PATH=:/usr/local/cuda-7.0/lib64
```

หลังจาก save ไฟล์ .bashrc แล้วให้พิมพ์คำสั่ง

```
>> source ~/.bashrc
```

ติดตั้ง cuDNN

```
>> tar -zxf cudnn-6.5-linux-x64-v2.tgz
>> cd cudnn-6.5-linux-x64-v2
>> sudo cp lib* /usr/local/cuda/lib64/
>> sudo cp cudnn.h /usr/local/cuda/include/
```

3) ติดตั้ง library และโปรแกรมต่างๆที่จำเป็นสำหรับการทำงานของ caffe

```
>> sudo apt-get install -y libprotobuf-dev libleveldb-dev
libsnappy-dev libopencv-dev libboost-all-dev libhdf5-serial-
dev protobuf-compiler gfortran libjpeg62 libfreeimage-dev
libatlas-base-dev git python-dev python-pip libgoogle-glog-dev
libbz2-dev libxml2-dev libxslt-dev libffi-dev libssl-dev
libgflags-dev liblmdb-dev python-yaml python-numpy
libopenblas-dev
```

รันติดตั้ง pillow

```
>> sudo easy_install pillow
```

ทำการคัดลอกโฟลเดอร์ caffe มายังคอมพิวเตอร์ที่ต้องการลงโปรแกรม

```
>> cd ~
>> git clone https://github.com/BVLC/caffe.git
```

ติดตั้ง Caffe

```
>> cd caffe
>> cat python/requirements.txt | xargs -L 1 sudo pip install
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คัดลอก Makefile และอัปเดตการตั้งค่า

```
>> cp Makefile.config.example Makefile.config
>> nano Makefile.config
```

สามารถตั้งค่าต่างๆได้โดยการลบ # ออกหน้าบรรทัดที่ต้องการดังต่อไปนี้

1. USE_CUDNN := 1 เพื่อสร้าง Caffe ใช้เฉพาะ GPU ในการประมวลผล (มีการ์ดจอ NVIDIA)
2. CPU_ONLY := 1 เพื่อสร้าง Caffe ใช้เฉพาะ CPU ในการประมวลผล
3. เปลี่ยนค่าของ BLAS := ATLAS เป็น BLAS := open เพื่อที่จะกระจายการประมวลผลไปทุกแกนประมวลผล (Cores) ของคอมพิวเตอร์

4) สร้าง Caffe

ให้ตั้งค่า X = จำนวน core ของ CPU ที่เรามี

```
>> make pycaffe -jX
>> make all -jX
>> make test -jX
```

5) ทดสอบการทำงานของ Caffe

ทดสอบโดยการสอนด้วยตัวอย่างของชุดข้อมูล mnist ด้วยคำสั่งต่อไปนี้

```
>> ./data/mnist/get_mnist.sh
>> ./examples/mnist/create_mnist.sh
>> ./examples/mnist/train_lenet.sh
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของ CNN ด้วย Caffe บนฐานข้อมูลที่ต้องการ

การทำงานในขั้นตอนนี้จำเป็นจะต้องมีไฟล์ภาพ (.jpg) ที่แบ่งเป็นโพลเดอร์สำหรับสอน และทดสอบเตรียมไว้แล้ว ซึ่งมีขั้นตอนในการดำเนินงานดังต่อไปนี้

1. การแปลงภาพจากไฟล์ .jpg เป็นฟอร์แมตของไฟล์ฐานข้อมูล lmdb

โดยในขั้นตอนนี้ภาพทั้งหมดที่จะนำมาใช้ต้องเป็นไฟล์ .jpg และแยกโพลเดอร์ระหว่างภาพที่ใช้สอน และทดสอบ โดยในโพลเดอร์นั้นจะประกอบด้วย โพลเดอร์ภาพของแต่ละคนและไฟล์ label ซึ่งอยู่ในรูปแบบ train.txt หรือ test.txt ซึ่ง label นี้จะเริ่มต้นตั้งแต่ 0 ถึงจำนวนคนในฐานข้อมูล -1

อ้างอิงจากไฟล์ `examples/imagenet/create_imagenet.sh`

โดยจะเปลี่ยนข้อมูลบางอย่างในไฟล์ `create_imagenet.sh` ดังต่อไปนี้

โพลเดอร์ที่อยู่ของไฟล์ lmdb

```
-EXAMPLE = /path/to/subfolder/under/examples/to/store/lmdb
file
```

โพลเดอร์ที่เก็บไฟล์ฐานข้อมูล

```
-DATA = /path/to/folder/containing/train.txt & val.txt
```

โพลเดอร์ย่อยที่เก็บไฟล์สำหรับสอน

```
-TRAIN_DATA_ROOT=/path/to/folder/omitted/in/train.txt
```

โพลเดอร์ย่อยที่เก็บไฟล์สำหรับสอน

```
-TEST_DATA_ROOT=/path/to/folder/omitted/in/val.txt
```

เปลี่ยนรายละเอียดเป็นที่อยู่ไฟล์ที่ต้องการ (สำหรับสอน และ ทดสอบ)

```
-GLOG...$EXAMPLE/..._train(val)_lmdb (rename the database for
training and testing)
```

หากเป็นภาพสีให้เพิ่มคำว่า `- -gray` หลังจาก `- -shuffle` \

2. แก้ไขโครงสร้าง CNN เพื่อนำมาใช้กับไฟล์ฐานข้อมูลที่ต้องการ

อ้างอิงจากไฟล์ `examples/mnist/lenet.prototxt`

`examples/mnist/lenet_train_test.prototxt`

`examples/mnist/lenet_solver.prototxt`

`examples/mnist/train_lenet.sh`

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภายในไฟล์ `lenet.prototxt` และ `lenet_train_test.prototxt` จะทำการแก้ไขดังต่อไปนี้

1. `input_dim: (batchsize, #channels, imagesize1,imagesize2)`
2. `inner_product_param: num_output: #labels`
3. “data” layer: `phase: TRAIN/TEST source: “...”`(path to training/testing lmdb database)

ภายในไฟล์ `lenet_solver.prototxt` จะทำการแก้ไขดังต่อไปนี้

1. `test_iter: #test batches (the total test images evaluated in TRAIN phase = test_iter*TEST batch size)`
2. `base_lr: initial learning rate (default:.01, change to a smaller number if getting NAN loss in training)`
3. `snapshot_prefix: “...”`(filepath to snapshot saved)
4. `snapshot: # of iterations per snapshot`
5. `max_iter: # of iterations`

ภายในไฟล์ `train_lenet.sh` จะทำการแก้ไขดังต่อไปนี้

```
-solver = *_solver.prototxt
```

3. Train model รันคำสั่งนี้ในหน้าต่าง Terminal

```
>> ./examples/.../train_*.sh
```

4. Test model โดยการเปลี่ยนที่อยู่เป็น โพลเดอร์ของ caffe

```
>> cd caffe_root
```

จากนั้นรันคำสั่ง

```
>> ./build/tools/caffe test -model /path/to/*_train_test.proto
txt -weights /path/to/trained/caffe/caffemodel -iterations #te
st images/batch_size
```

Source Code

examples/celeb/create_celeb.sh

เป็นโปรแกรมสำหรับการแปลงข้อมูลภาพ jpg เป็นไฟล์ฐานข้อมูล lmdb มีการกำหนดที่อยู่ของไฟล์และการกำหนดต่างๆ ดังแสดงในโปรแกรมต่อไปนี้

```
#!/usr/bin/sh
# Create the imagenet lmdb inputs
# N.B. set the path to the imagenet train + val data dirs
EXAMPLE=~/.caffe/examples/celebface_lenet
#DATA=~/.caffe/data/data_celeb
DATA=~/.caffe/data/3_Folder_colour
TOOLS=~/.caffe/build/tools
#TRAIN_DATA_ROOT=~/.caffe/data/data_celeb/train/
#TEST_DATA_ROOT=~/.caffe/data/data_celeb/test/
TRAIN_DATA_ROOT=~/.caffe/data/3_Folder_colour/train/
TEST_DATA_ROOT=~/.caffe/data/3_Folder_colour/test/
# Set RESIZE=true to resize the images to 256x256. Leave as false if
images have
# already been resized using another tool.
RESIZE=false
if $RESIZE; then
    RESIZE_HEIGHT=256
    RESIZE_WIDTH=256
else
    RESIZE_HEIGHT=0
    RESIZE_WIDTH=0
fi
if [ ! -d "$TRAIN_DATA_ROOT" ]; then
    echo "Error: TRAIN_DATA_ROOT is not a path to a directory:
$TRAIN_DATA_ROOT"
    echo "Set the TRAIN_DATA_ROOT variable in create_imagenet.sh to
the path" \
        "where the ImageNet training data is stored."
    exit 1
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

fi

if [ ! -d "$TEST_DATA_ROOT" ]; then
    echo "Error: TEST_DATA_ROOT is not a path to a directory:
$TEST_DATA_ROOT"

    echo "Set the TEST_DATA_ROOT variable in create_imagenet.sh to the
path" \
        "where the ImageNet validation data is stored."
    exit 1
fi

echo "Creating train lmdb..."
GLOG_logtostderr=1 $TOOLS/convert_imageset \

    $TRAIN_DATA_ROOT \
    $DATA/train.txt \
    $EXAMPLE/celebface_train_lmdb
echo "Creating validate lmdb..."
GLOG_logtostderr=1 $TOOLS/convert_imageset \

    $TEST_DATA_ROOT \
    $DATA/test.txt \
    $EXAMPLE/celebface_test_lmdb
echo "Done."

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

examples/celeb/make_celeb_mean.sh

โปรแกรมนี้จะใช้สำหรับการสร้างไฟล์ค่าเฉลี่ยของฐานข้อมูล มีการกำหนดที่อยู่ไฟล์และค่าต่างๆดังต่อไปนี้

```
#!/usr/bin/env sh
# Compute the mean image from the imagenet training lmdb
# N.B. this is available in data/ilsvrc12

EXAMPLE=examples/celebface_lenet
DATA=data/3_Folder_colour
TOOLS=build/tools

$TOOLS/compute_image_mean $EXAMPLE/celebface_train_lmdb \
  $DATA/celebface_mean.binaryproto

echo "Done."
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

examples/celeb/celebface lenet train test MoMo.prototxt

โปรแกรมนี้เป็นโครงสร้างของ CNN ในการทำงาน ซึ่งแบ่งออกเป็นชั้นต่าง และมีการกำหนดพารามิเตอร์ดังแสดงในโปรแกรมดังต่อไปนี้

```
##### DATA Layer #####
name: "face_train_val"
layer {
  top: "data_1"
  top: "label_1"
  name: "data_1"
  type: "Data"
  data_param {
    source: "examples/celebface_lenet/celebface_train_lmdb"
    backend:LMDB
    batch_size: 128
  }
  transform_param {
    mean_file: "data/3_Folder_colour/celebface_mean.binaryproto"
    mirror: true
  }
  include: { phase: TRAIN }
}

layer {
  top: "data_1"
  top: "label_1"
  name: "data_1"
  type: "Data"
  data_param {
    source: "examples/celebface_lenet/celebface_validate_lmdb"
    backend:LMDB
    batch_size: 128
  }
  transform_param {
    mean_file: "data/3_Folder_colour/celebface_mean.binaryproto"
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    mirror: true
  }
  include: {
    phase: TEST
  }
}

```

```
##### CONV NET 1 #####
```

```

layer {
  name: "conv1_1"
  type: "Convolution"
  bottom: "data_1"
  top: "conv1_1"
  param {
    name: "conv1_w"
    lr_mult: 1
    decay_mult: 1
  }
  param {
    name: "conv1_b"
    lr_mult: 2
    decay_mult: 0
  }
  convolution_param {
    num_output: 20
    kernel_size: 4
    stride: 1
    weight_filler {
      type: "gaussian"
      std: 0.01
    }
    bias_filler {
      type: "constant"
      value: 0
    }
  }
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    }
  }
  layer {
    name: "relu1_1"
    type: "ReLU"
    bottom: "conv1_1"
    top: "conv1_1"
  }
  layer {
    name: "norm1_1"
    type: "LRN"
    bottom: "conv1_1"
    top: "norm1_1"
    lrn_param {
      local_size: 5
      alpha: 0.0001
      beta: 0.75
    }
  }
  layer {
    name: "pool1_1"
    type: "Pooling"
    bottom: "norm1_1"
    top: "pool1_1"
    pooling_param {
      pool: MAX
      kernel_size: 2
      stride: 2
    }
  }
  layer {
    name: "conv2_1"
    type: "Convolution"
    bottom: "pool1_1"
    top: "conv2_1"
  }

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

param {
  name: "conv2_w"
  lr_mult: 1
  decay_mult: 1
}
param {
  name: "conv2_b"
  lr_mult: 2
  decay_mult: 0
}
convolution_param {
  num_output: 40
  kernel_size: 3
  group: 2
  weight_filler {
    type: "gaussian"
    std: 0.01
  }
  bias_filler {
    type: "constant"
    value: 0.1
  }
}
}
}
layer {
  name: "relu2_1"
  type: "ReLU"
  bottom: "conv2_1"
  top: "conv2_1"
}
layer {
  name: "norm2_1"
  type: "LRN"
  bottom: "conv2_1"

```

เอกสารนี้เป็นเอกสารที่ สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

top: "norm2_1"
lrn_param {
  local_size: 5
  alpha: 0.0001
  beta: 0.75
}
}
layer {
  name: "pool2_1"
  type: "Pooling"
  bottom: "norm2_1"
  top: "pool2_1"
  pooling_param {
    pool: MAX
    kernel_size: 2
    stride: 2
  }
}
layer {
  name: "conv3_1"
  type: "Convolution"
  bottom: "pool2_1"
  top: "conv3_1"
  param {
    name: "conv3_w"
    lr_mult: 1
    decay_mult: 1
  }
  param {
    name: "conv3_b"
    lr_mult: 2
    decay_mult: 0
  }
  convolution_param {
    num_output: 60

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

kernel_size: 3
weight_filler {
  type: "gaussian"
  std: 0.01
}
bias_filler {
  type: "constant"
  value: 0
}
}
}

layer {
  name: "pool3_1"
  type: "Pooling"
  bottom: "conv3_1"
  top: "pool3_1"
  pooling_param {
    pool: MAX
    kernel_size: 2
    stride: 2
  }
}

layer {
  name: "conv4_1"
  type: "Convolution"
  bottom: "pool3_1"
  top: "conv4_1"
  param {
    name: "conv4_w"
    lr_mult: 1
    decay_mult: 1
  }
  param {
    name: "conv4_b"

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    lr_mult: 2
    decay_mult: 0
  }
  convolution_param {
    num_output: 80
    kernel_size: 2
    stride: 2
    weight_filler {
      type: "gaussian"
      std: 0.01
    }
    bias_filler {
      type: "constant"
      value: 0.1
    }
  }
}
}
layer{
  name:"flatten_pool3_1"
  type:"Flatten"
  bottom:"pool3_1"
  top:"flatten_pool3_1"
}
layer{
  name:"flatten_conv4_1"
  type:"Flatten"
  bottom:"conv4_1"
  top:"flatten_conv4_1"
}
layer{
  name:"contact_conv"
  type:"Concat"
  bottom:"flatten_conv4_1"
  bottom:"flatten_pool3_1"
}

```

เอกสารนี้เป็นเอกสารที่ สงวนลิขสิทธิ์ไว้สำหรับการทำงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    top:"contact_conv"
  }
  layer {
    name: "deepid_1"
    type: "InnerProduct"
    bottom: "contact_conv"
    top: "deepid_1"
    param {
      name: "fc6_w"
      lr_mult: 1
      decay_mult: 1
    }
    param {
      name: "fc6_b"
      lr_mult: 2
      decay_mult: 0
    }
    inner_product_param {
      num_output: 160
      weight_filler {
        type: "gaussian"
        std: 0.005
      }
      bias_filler {
        type: "constant"
        value: 0.1
      }
    }
  }
}

```

```

}
layer {
  name: "relu6_1"
  type: "ReLU"
  bottom: "deepid_1"
  top: "deepid_1"

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

}
layer {
  name: "drop6_1"
  type: "Dropout"
  bottom: "deepid_1"
  top: "deepid_1"
  dropout_param {
    dropout_ratio: 0.5
  }
}
}

```

```

layer {
  name: "fc8_1"
  type: "InnerProduct"
  bottom: "deepid_1"
  top: "fc8_1"
  param {
    name: "fc8_w"
    lr_mult: 1
    decay_mult: 1
  }
  param {
    name: "fc8_b"
    lr_mult: 2
    decay_mult: 0
  }
  inner_product_param {
    num_output: 1558
    weight_filler {
      type: "gaussian"
      std: 0.01
    }
    bias_filler {
      type: "constant"

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    }
  }
}

layer {
  name: "accuracy_1"
  type: "Accuracy"
  bottom: "fc8_1"
  bottom: "label_1"
  top: "accuracy_1"
  include: { phase: TEST }
}
layer {
  name: "loss_1"
  type: "SoftmaxWithLoss"
  bottom: "fc8_1"
  bottom: "label_1"
  top: "loss_1"
  #loss_weight: 0.5
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

examples/celeb/celebface lenet solver MoMo.prototxt

โปรแกรมนี้สำหรับตั้งค่าพารามิเตอร์ในการสอนและทดสอบ

```
# The train/test net protocol buffer definition
net:
"examples/celebface_lenet/celebface_lenet_train_test_MoMo.prototxt"
# test_iter specifies how many forward passes the test should carry
out.
# In the case of MNIST, we have test batch size 100 and 100 test
iterations,
# covering the full 10,000 testing images.
test_iter: 100
# Carry out testing every 500 training iterations.
test_interval: 1000
# The base learning rate, momentum and the weight decay of the
network.
base_lr: 0.001
momentum: 0.9
weight_decay: 0.0005
# The learning rate policy
lr_policy: "step"
gamma: 0.95
stepsize: 10000
power: 0.75
# Display every 100 iterations
display: 100
# The maximum number of iterations
max_iter: 500000
# snapshot intermediate results
snapshot: 50000
snapshot_prefix: "examples/celebface_lenet/celebface_MoMo"
# solver mode: CPU or GPU
solver_mode: GPU
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

examples/celeb/train_celebface MoMo.sh

ไฟล์นี้จะเป็นไฟล์สำหรับการสั่งรัน โดยการกำหนดค่าต่างๆ

```
#!/usr/bin/env sh
```

```
./build/tools/caffe train --  
solver=examples/celebface_lenet/celebface_lenet_solver_MoMo.prototxt  
#--snapshot=examples/celebface_lenet/celebface_MoMo_iter_500000_5  
.solverstate
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

