การจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ

Sign language recognition with recurrent neural network

กวิน เต็งอำนวย

Kavin Tengamnuay

ธัญพิสิษฐ์ พรมมี

Thanphisit Prommee

ธิดารัตน์ โชคชวลิต

Tidarat Chokchawalit

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์ ภาควิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

การจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ

Sign language recognition with recurrent neural network

โดย

กวิน เต็งอำนวย

ธัญพิสิษฐ์ พรมมี

ธิดารัตน์ โชคชวลิต

อาจารย์ที่ปรึกษา

ผศ.ดร.ยุทธนา คิดใจเดียว

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์ ภาควิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชา วิศวกรรมอิเล็กทรอนิกส์

คณะ วิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ

Sign language recognition with recurrent neural network

ผู้จัดทำ	นายกวิน เต็งอำนวย	รหัสนักศึกษา	60010035
	นายธัญพิสิษฐ์ พรมมี	รหัสนักศึกษา	60010457
	นางสาวธิดารัตน์ โชคชวลิต	รหัสนักศึกษา	60010468

ปริญญานิพนธ์นี้ผ่านการตรวจสอบโดยอาจารย์ที่ปรึกษาแล้ว

.....

ผศ.ดร.ยุทธนา คิดใจเดียว

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

หัวข้อปริญญานิพนธ์	การจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ		
นักศึกษา	นายกวิน เต็งอำนวย	รหัสนักศึกษา	60010035
	นายธัญพิสิษฐ์ พรมมี	รหัสนักศึกษา	60010457
	นางสาวธิดารัตน์ โชคชวลิต	รหัสนักศึกษา	60010468
ปริญญา	วิศวกรรมศาสตรบัณฑิต		
ภาควิชา	วิศวกรรมอิเล็กทรอนิกส์		
ปีการศึกษา	2563		
อาจารย์ที่ปรึกษา	ดร.ยุทธนา คิดใจเดียว		

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อวิเคราะห์จำแนกท่าทางภาษามือโดยใช้วิธีทางเทคโนโลยีการประมวลผลโครงข่ายประสาทเทียมแบบวนซ้ำการวิเคราะห์การจำแนกภาษามือนั้นมีความสำคัญเนื่องจากภาษามือเป็นส่วนหนึ่งของการสื่อสารและใช้ในการถ่ายทอดความหมายเมื่อเทียบกับการพูดซึ่งจะทำให้คนปกติสามารถสื่อสารกับผู้ที่มีความผิดปกติทางการได้ยินโดยการทำงานแบ่งเป็น 3 ส่วนคือการเก็บวิดีโอและแปลงเป็นข้อมูล, การเทรนโมเดล, การทำนายซึ่งจากวิดีโอเป็นข้อมูลที่มีความต่อเนื่องเป็นลำดับการเทรนโมเดลจึงต้องใช้โมเดลที่เป็นโครงข่ายประสาทเทียมแบบวนซ้ำและการทำนายจะกำหนดเงื่อนไขเพื่อให้ได้ผลลัพธ์ตามความหมายที่ถูกต้องรวมถึงกำหนดลักษณะของผลลัพธ์ที่ต้องการทั้งข้อความเสียงและภาพแอนิเมชั่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Thesis Title	Sign language recognition with recurrent neural network		
Student	Mr.Kavin Tengamnuay	Student ID	60010035
	Mr.Thanphisit Prommee	Student ID	60010457
	Miss.Tidarat Chokchawalit	Student ID	60010468
Degree	Bachelor of Engineering		
Program	Electronics Engineering		
Year	2020		
Advisor	Asst. Prof. Dr.Yutthana Kidjaidure		

ABSTRACT

This thesis aims to analyze sign language recognition with recurrent neural network. The sign language analysis is important because it is a part of deaf communication and it is used to convey meanings as voice from speech which made hearing impaired people to understand. The processing has 3 part that is convert video to data, model training and prediction. Which the videos have data of sequence. Therefore, model is used recurrent neural network. Condition is made for correct meaning of the prediction results and set output as message, sound and animation.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

ปริญญานิพนธ์เรื่องการจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ สามารถสำเร็จลุลวงไปได้ด้วยดี ทั้งนี้ต้องขอขอบพระคุณ ผศ.ดร.ยุทธนา คิดใจเดียว อาจารย์ที่ปรึกษาโครงงาน ที่คอยให้ความรู้ คำแนะนำ และชี้แนะแนวทางในการปฏิบัติงาน รวมไปถึงการให้คำปรึกษาในเรื่องของทฤษฎี ตลอดจนการตามผลชิ้นงาน ความคืบหน้าในทุกสัปดาห์ ขอขอบพระคุณอาจารย์ภาควิชาอิเล็กทรอนิกส์ท่านอื่นที่มีความเกี่ยวข้อง ซึ่งคอยช่วยแนะนำแก้ไขจุดบกพร่องในชิ้นงาน ทั้งโดยตรงและโดยอ้อมทุกท่าน ขอบคุณนักศึกษาปริญญาโทในภาควิชาอิเล็กทรอนิกส์ทุกท่านที่คอยสนับสนุนในการทดลองและวิธีการทดสอบ รวมไปถึงการเตรียมความพร้อมในการตอบคำถามต่าง ๆ ที่จะต้องเกิดขึ้นในการสอบ

สุดท้ายนี้ต้องขอขอบพระคุณ คุณพ่อและคุณแม่ของข้าพเจ้า ที่อยู่เบื้อหลังความสำเร็จ พร้อมทั้งสนับสนุนในด้านต่าง ๆ ทั้งสถานที่ และสิ่งอำนวยความสะดวกต่าง ๆ ผู้วิจัยหวังเป็นอย่างยิ่งว่าปริญญานิพนธ์ฉบับนี้จะเป็นประโยชน์ไม่มากก็น้อยต่อผู้ที่สนใจในภายภาคหน้า

กวิน เต็งอำนวย
ธัญพิสิษฐ์ พรมมี
ธิดารัตน์ โชคชวลิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของงานวิจัย 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของงานวิจัย 1

1.3 สมมุติฐานของงานวิจัย 1

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในงานวิจัย 2

1.5 ขอบเขตงานวิจัย 2

1.6 ขั้นตอนของงานวิจัย 2

บทที่ 2 ทฤษฎีและเอกสารที่เกี่ยวข้อง 3

2.1 ภาษามือ 3

2.1.1 ประเภทของภาษามือ 3

2.1.2 ลักษณะของภาษามือ 3

2.1.3 โครงสร้างของภาษามือ 4

2.2 โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network) 5

2.3 โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) 9

2.3.1 Backpropagation Through Time 10

2.3.2 เครือข่ายหน่วยความจำระยะสั้นระยะยาว 11

(Long Short Term Memory Network)

2.4 ฟังก์ชันซอฟต์แมกซ์ (Softmax Function) 14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ (ต่อ)

หน้า

2.5 การตรวจจับวัตถุ (Object Detection) 14
2.5.1 Single Shot MultiBox Detector 15
2.6 Mediapipe 15
2.6.1 MediaPipe Hands 15
2.7 Bazel 16
2.8 Library in Python 16
2.8.1 OpenCV 17
2.8.2 Tensorflow 17
2.8.3 Numpy 19
2.8.4 gTTS 19
บทที่ 3 วิธีการดำเนินการวิจัย 20
3.1 หลักการทำงาน 20
3.2 ส่วนของการสร้างระบบตรวจจับมือ (Building Phase) 21
3.3 ส่วนของการเรียนรู้ระบบที่ใช้ในการทำนายภาษามือ (Training Phase) 25
3.4 ส่วนของการทดสอบหรือส่วนของผู้ใช้งาน (Testing Phase) 29
บทที่ 4 ผลการทดลอง 31
4.1 ผลการตรวจจับมือ 31
4.2 ผลการเรียนรู้ของโมเดล LSTM 32
4.3 การทดสอบระบบ 33
4.3.1 การแสดงผลเอาต์พุตของระบบ 33
4.3.2 การทดสอบประสิทธิภาพของระบบ 34
บทที่ 5 สรุปผลการทดลอง 36
5.1 สรุปและวิจารณ์ผลการทดลอง 36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ (ต่อ)

หน้า

บรรณานุกรม..... 37

ภาคผนวก..... 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตารางที่ หน้า

4.1 การทดสอบประสิทธิภาพของระบบ 34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูปที่ หน้า

2.1 โครงสร้างของส่วนประสาท 5

2.2 โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน 6

2.3 อินพุตของตัวกรองหรือเคอร์เนล 7

2.4 ผังคุณลักษณะ 7

2.5 การทำ Padding 7

2.6 การทำ Max Pooling 8

2.7 การทำ Fully Connected 8

2.8 Recurrent neural network 9

2.9 ชั้นประตูลืม 10

2.10 ชั้นประตูทางเข้า 12

2.11 ชั้นประตูแปลงร่าง 12

2.12 ชั้นประตูทางออก 13

2.13 ตัวอย่างของการตรวจจับวัตถุ 15

2.14 รูปตัวอย่างทำงาน tensorflow 18

3.1 Block Diagram แสดงการทำงานของระบบ 20

3.2 Block Diagram แสดงระบบตรวจจับมือ 21

3.3 ข้อมูลวิดีโอท่าภาษามือ 22

3.4 การตรวจจับมือและการสร้างโครงสร้างมือสามมิติ 23

3.5 การสร้างโครงสร้างสามมิติของมือ 23

3.6 พิกัดตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้งหมด 21 จุด 23

3.7 ข้อมูลตำแหน่งบนฝ่ามือและข้อนิ้วมือในไฟล์สกุล.txt 24

3.8 Block Diagram แสดงระบบที่ใช้ในการทำนายภาษามือ 25

3.9 การ train model 27

3.10 ชั้นของโมเดล LSTM 27

3.11 Block Diagram การใช้งานระบบจำแนกภาษามือ 29

4.1 แสดงการพล็อตค่าตำแหน่งคู่อันดับ (x ,y) บนฝ่ามือและข้อนิ้วมือทั้ง 21 จุด 31

4.2 กราฟแสดงค่า loss 32

4.3 กราฟแสดงค่า accuracy 32

4.4 ข้อมูลอินพุตทเป็นวิดีโอท่าทางภาษามือ 33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.5 ผลเอาต์พุตของระบบ 33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

ภาษามือเป็นอวัจนภาษาอย่างหนึ่งซึ่งถูกใช้โดยผู้พิการทางหูเป็นหลัก การสื่อสารจะใช้ลักษณะของมือ แขน และร่างกายเป็นสัญลักษณ์เพื่อช่วยในการสื่อสารความหมายแทนการใช้เสียงพูด เนื่องจากในผู้ทำการวิจัยมีความสนใจในเรื่องของปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งเปรียบเสมือนเครื่องจักรที่มีฟังก์ชัน (Function) ในการทำความเข้าใจเรียนรู้องค์ความรู้ต่าง ๆ อาทิ เช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหาต่าง ๆ โดยถูกออกแบบให้มีระบบทำงานเหมือนกับการทำงานของสมองมนุษย์ ผู้ทำการวิจัยมีจุดประสงค์ที่จะนำมาสร้างและเขียนโปรแกรมที่ใช้ในการวิเคราะห์แยกแยะภาษามือให้ออกมาเป็นคำต่าง ๆ

งานวิจัยชิ้นนี้เป็นการศึกษาวิธีการทำโปรแกรมที่ใช้ในการวิเคราะห์แยกแยะภาษามือให้ออกมาเป็นคำต่าง ๆ แบบอัตโนมัติ โดยโปรแกรมนี้จะช่วยในเรื่องของการสื่อสารกับกลุ่มผู้พิการทางหู ทำให้คนปกติสามารถสื่อสารกับผู้พิการทางหูได้

1.2 ความมุ่งหมายและวัตถุประสงค์ของงานวิจัย

ในการประมวลผลภาพวิดิโอภาษามือนั้นมีการคิดค้นและสร้างเครื่องมือขึ้นมามากมายแต่โดยส่วนใหญ่จะเป็นเทคโนโลยีขั้นสูง กล่าวคือมีการใช้กล้องที่มีคุณภาพสูงและมีราคาแพง จากปัญหาดังกล่าวงานวิจัยนี้จึงมีจุดมุ่งหมายในการประมวลผลภาพโดยใช้กล้องเว็บแคม (Web cam) ซึ่งมีราคาถูกและหาซื้อได้ทั่วไป ผู้ทำการวิจัยได้ใช้โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) และอัลกอริทึม Mediapipe Hand ในการวิเคราะห์และเรียนรู้คำศัพท์ภาษามือไทยในชีวิตประจำวัน 34 คำ

1.3 สมมุติฐานของงานวิจัย

คาดว่าสามารถคิดค้นวิธีการและจัดการขนาดข้อมูลได้เหมาะสม เพื่อให้โปรแกรมสามารถจำแนกและเรียนรู้คำศัพท์ภาษามือไทยได้ โดยใช้โครงข่ายประสาทเทียมแบบวนซ้ำ และมีการใช้อัลกอริทึม Mediapipe Hand ในการตรวจจับมือจากภาพวิดิโอที่บันทึกจากกล้องเว็บแคม เป็นการแก้ไขปัญหาอุปกรณ์ที่มีราคาแพงและลดความซับซ้อนในการประมวลผลจากกล้องสามมิติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในงานวิจัย

งานวิจัยในการสกัดคุณลักษณะ (Feature Extraction) เพื่อนำมาใช้จำแนกตัวอักษรและคำภาษามือนั้นส่วนใหญ่นั้นกระทำโดยใช้กล้องสามมิติหรืออุปกรณ์เสริม เช่น ถุงมือ (Artificial gloves) โดยเฉพาะกับการจำแนกคำส่วนใหญ่เป็นท่าทางเคลื่อนไหวสามมิติและมีสีหน้าเข้ามาเกี่ยวข้อง ในการจำแนกภาษามือหลายๆครั้งจะใช้สีของผิวในการแยกส่วนของมือออกมาก่อนจากนั้นจึงนำมาวิเคราะห์ด้วยการใช้ Canny method ซึ่งผู้ทำการวิจัยพบว่าข้อมูลที่ได้นั้นมีขนาดใหญ่เกินไปและจำเป็นต้องนำไปลดมิติของข้อมูลก่อนซึ่งเป็นการเสียเวลา ผู้ทำวิจัยจึงคิดค้นวิธีการที่สามารถจำลองและจำแนกนิ้วมือจากภาพวิดีโอ และการสกัดคุณลักษณะในการเรียนรู้คำศัพท์ภาษามือไทยจากทิศทางการเคลื่อนที่ของมือและรูปร่างของมือ โดยนำไปวิเคราะห์ผลด้วยโครงข่ายประสาทเทียมแบบวนซ้ำ

1.5 ขอบเขตงานวิจัย

คิดค้นวิธีการสกัดคุณลักษณะของท่าทางมือโดยใช้กล้องเว็บแคมปกติและวิเคราะห์หาคุณลักษณะของท่าทางมือจากเฟรมของภาพซึ่งได้มาจากการบันทึกวิดีโอ โดยการใช้อัลกอริทึม Mediapipe Hand ในการตรวจจับมือ และจำแนกและเรียนรู้คำศัพท์ภาษามือไทย 34 คำ ด้วยโครงข่ายประสาทเทียมแบบวนซ้ำ

1.6 ขั้นตอนของงานวิจัย

เริ่มต้นด้วยการรับวิดีโอจากกล้องเว็บแคมเข้ามา ร่วมกับอัลกอริทึม Mediapipe Hand ที่เป็นการตรวจจับฝ่ามือ ซึ่งจะแปลงวิดีโอในตำแหน่งของมือออกมาเป็นข้อมูล หลังจากนั้นจะนำข้อมูลในแต่ละเฟรม (Frame) ของวิดีโอมาทำการสกัดคุณลักษณะเพื่อนำเข้าไปเรียนรู้โครงข่ายประสาทเทียมแบบวนซ้ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 2
ทฤษฎีและเอกสารที่เกี่ยวข้อง

2.1 ภาษามือ

ภาษามือเป็นอวัจนภาษาอย่างหนึ่ง ที่ประกอบด้วย การสื่อสารด้วยมือ การสื่อสารด้วยร่างกาย และการใช้ริมฝีปากในการสื่อความหมายแทนการใช้เสียงพูด การสื่อสารจะใช้ลักษณะของมือที่ทำเป็นสัญลักษณ์ การเคลื่อนไหวมือ แขน และร่างกาย และการแสดงความรู้สึกทางใบหน้าเพื่อช่วยในการสื่อสารความคิดของผู้สื่อสาร ภาษาสัญลักษณ์ส่วนใหญ่มักใช้ในกลุ่มผู้พิการทางหู

2.1.1 ประเภทของภาษามือ

ภาษามือที่ใช้อยู่ในกลุ่มผู้พิการทางหูมักจะมี 2 ประเภท คือ

ภาษามือธรรมชาติ (Sign language) คือ ภาษาท่าทางที่ผู้พิการทางหูเป็นผู้สร้างขึ้นและใช้ร่วมกันในแต่ละชุมชนหรือในแต่ละชาติ เช่น American Sign language , British Sign language ส่วนมากเป็นท่าเลียนแบบธรรมชาติที่จะช่วยผู้พิการทางหูให้มีพัฒนาการในภาษาประจำชาติเท่าเทียมกับคนปกติ

ภาษามือประดิษฐ์ (Signed) คือ ภาษามือที่ครู ผู้ปกครอง หรือญาติมิตรของผู้พิการทางหูคิดขึ้นแทนภาษาพูดและภาษาเขียนประจำชาติเพื่อให้มีคำใช้ให้เพียงพอในการศึกษาและการสื่อความหมายโดยเฉพาะเรื่องนามธรรม บางครั้งเรียกภาษามือประดิษฐ์นี้ว่า ภาษามือในห้องเรียน หรือภาษามือที่ใช้ในการศึกษาซึ่งเป็นภาษาที่ท่าคำแต่ละคำตามไวยากรณ์ภาษาพูด หรือภาษาเขียนของคนปกติ ภาษามือประดิษฐ์มักจะมีการสะกดนิ้วมือ (Finger Spelling) มาผสมด้วย

2.1.2 ลักษณะของภาษามือ

ภาษามือเป็นภาษาท่าทางซึ่งมีการเคลื่อนไหวของมือเป็นหลักและใช้กิริยาอาการของหน้าตาและร่างกายส่วนหนึ่งเป็นส่วนประกอบช่วยให้เกิดความเข้าใจ ท่าภาษามือที่ผู้พิการทางหูยอมรับจะต้องเป็นท่าที่ทำง่าย สะดวก รวดเร็ว มีความหมายใกล้เคียงธรรมชาติและเหมาะกับหลักสรีระศาสตร์ ท่าภาษามือควรทำอย่างมีจังหวะ คือ มีการเว้นระยะ ไม่ทำท่าทางเร็วจนเกินไป และให้อยู่ในรัศมีที่สายตาสามารถมองเห็นได้ชัดเจน โดยทั่วไปจะทำท่ามือเคลื่อนไหวตั้งแต่ระดับเอวขึ้นมาถึงส่วนบนของร่างกายในท่าทางภาษามือการแสดงสีหน้าและการเคลื่อนไหวของใบหน้า เช่น คิ้ว ปาก เป็นสิ่งสำคัญช่วยให้เข้าใจความหมายในภาษามือชัดเจนยิ่งขึ้น เช่น

- การส่ายศีรษะ หมายถึง การปฏิเสธ
- การขมวดคิ้ว หมายถึง การแสดงความสงสัย
- การเลิกคิ้ว หมายถึง การแสดงคำถามที่ต้องการคำตอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.3 โครงสร้างของภาษามือ

ในภาษาไทย ภาษาอังกฤษ หรือภาษาใดใดก็ตาม จะมีโครงสร้างของภาษา เช่น มีคำ วลี มีประโยคนำมาเรียงกัน เป็นข้อความที่มีความหมาย ภาษามือก็เช่นกันที่การแสดงออกและการเคลื่อนไหว ล้วนมีความหมายทั้งสิ้น เรียกว่า เป็นภาษาที่มีโครงสร้างเช่นกัน ภาษามือประกอบด้วยโครงสร้างดังนี้

2.1.3.1 ท่ามือ (The handshape)

ท่ามือ คือ การทำมือเป็นท่าต่าง ๆ เช่น กำมือ แบมือ กางนิ้ว รวมนิ้ว จีบนิ้ว ฯลฯ แต่ละท่ามีความหมายต่างกันออกไป

2.1.3.2 ตำแหน่งของมือ (The position of the hand)

ตำแหน่งของมือ คือ การวางมือหรือทำท่ามือในตำแหน่งต่าง ๆ ตำแหน่งของมือจะให้ความหมายแตกต่างกัน ถึงแม้ว่าท่ามือจะเป็นท่าเดียวกัน เช่น ใช้นิ้วชี้ชี้ที่หน้าอก หมายถึง “ฉัน” ถ้าชี้ที่ขมับ หมายถึง “รู้” ชี้ออกไปที่คู่สนทนา หมายถึง “เธอ” หรือ “คุณ” เป็นต้น ตำแหน่งที่ทำท่ามือควรจะอยู่ในรัศมีที่สายตาสามารถมองเห็นได้ง่ายและชัดเจน คือ บริเวณศีรษะ และใกล้ใบหน้าไม่ควรต่ำกว่า ระดับเอว คำภาษามือที่แสดงความรู้สึกต่าง ๆ มักจะแสดงท่ามือในตำแหน่งใกล้เคียงกับความหมายของคำนั้น ๆ เช่น

- ท่ามือบริเวณศีรษะ จะเกี่ยวกับความคิด เช่น รู้ฝัน ฉลาด
- ท่ามือบริเวณอก จะเกี่ยวกับความรู้สึก เช่น รักเสียใจขอบคุณ
- ท่ามือบริเวณลำตัว จะเป็นคำทั่ว ๆ ไป เช่น ลูก ซักผ้า รองเท้า

2.1.3.3 การเคลื่อนไหวของมือ (The movement of the hand)

การเคลื่อนไหวของมือ คือ การย้ายมือเปลี่ยนมือ ไปในทิศทางหรือรูปแบบที่ต้องการ ท่ามืออย่างเดียวกัน แต่เคลื่อนไหวไปในทิศทางต่างกัน ความหมายจะแตกต่างกัน เช่น มือทั้งสองตั้งขึ้น หัวแม่มือชิดกัน แล้วเลื่อนออกห่างคือ “เปิด” แต่ถ้าตั้งมือห่างกัน พอสมควรแล้วเลื่อนให้หัวแม่มือชิดกัน หมายถึง “ปิด” การเคลื่อนไหวมือมักจะใช้มือที่ถนัดเป็นมือที่เคลื่อนไหวไม่บังคับว่าจะต้องมือซ้ายหรือมือขวาเสมอไป

2.1.3.4 ทิศทางของฝ่ามือ (The orientation of the palms)

ทิศทางของฝ่ามือ คือ การหันฝ่ามือไปในทิศทางต่างๆ ทิศทางของฝ่ามือเป็นส่วนสำคัญ อย่างหนึ่ง ซึ่งทำให้ท่ามือมีความหมายต่างกัน ท่ามือท่าเดียวกัน ตำแหน่งที่เดียวกัน แต่ทิศทางของฝ่ามือต่างกัน ความหมายจะต่างกัน ตัวอย่างเช่น ตั้งมือขึ้น นิ้วชิดกัน หันฝ่ามือออกยื่นไปข้างหน้า หมายถึง “ของเขา” แต่ถ้าหันฝ่ามือเข้ามาหาตัว หมายถึง “ของฉัน” การทำท่ามือถ้าอยู่ในระดับสายตามองเห็นได้ง่าย เช่น บริเวณใบหน้ามักจะทำมือเดียว เช่น สวย อิ่ม แต่ถ้าระดับต่ำกว่าอกจะทำสองมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เพื่อให้เห็นชัดเจนยิ่งขึ้น เช่น กระโปรง ถ้าทำท่าสองมือจะต้องเคลื่อนมือใดมือหนึ่ง มือที่ถนัดจะเป็นมือที่เคลื่อนไหว มือที่ไม่ถนัดจะทำง่ายกว่า หรืออยู่นิ่ง ๆ เป็นฐาน เพราะมนุษย์ไม่สามารถทำงานพร้อมกัน ทั้งมือซ้ายและมือขวาด้วยท่าทางที่แตกต่างกันในเวลาเดียวกันได้โดยง่าย

2.2 โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network)

โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network) เป็นโมเดล (Model) ทางคณิตศาสตร์ที่จำลองกระบวนการคิดจากสมองมนุษย์ โดยสมองนั้นจะมีหน่วยประมวลผลขนาดเล็กอยู่เยอะมากและเชื่อมโยงกันด้วยโครงข่ายประสาทมากมาย ทำให้มนุษย์เรียนรู้และคิดวิเคราะห์ได้อย่างรวดเร็ว แต่คอมพิวเตอร์นั่นไม่ได้มีโครงข่ายที่ซับซ้อนเหมือนกับสมองของมนุษย์ คอมพิวเตอร์มีหน้าที่เพียงรันโปรแกรมตามคำสั่งเท่านั้น เมื่อให้คอมพิวเตอร์ทำการเรียนรู้อะไรซักอย่างจึงเป็นเรื่องยาก จึงเกิดการจำลองแนวทางการเรียนรู้ของคนไปสู่คอมพิวเตอร์ด้วยโครงข่ายประสาท ส่วนที่เล็กที่สุดคือส่วนประสาทซึ่งทำหน้าที่คำนวณ มีส่วนประกอบดังนี้

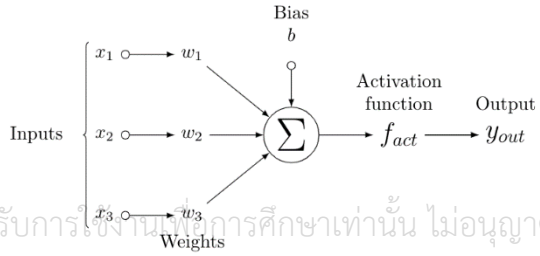
- Input เป็นค่าที่ส่งเข้ามาที่ส่วนประสาท โดยจะมีขาที่เข้ามาได้หลายขา โดยทั่วไปมักจะเท่ากับจำนวนคลาส (Class)

- Weight เป็นการให้น้ำหนักของขาแต่ละที่ส่งเข้ามา โดยมีค่าระหว่าง 0-1 โดยที่ทำการสุ่มขึ้นมา จากนั้นส่วนประสาทเมื่อทำการเรียนรู้เรื่อย ๆ ก็จะเป็นการปรับ Weight เพื่อให้ได้คำตอบที่ใกล้เคียงที่สุด

- Bias คือค่าที่จะช่วยเข้ามาทำให้ค่าที่เข้ามาอยู่ในระหว่าง 0 - 1 ได้ โดยจะเป็นเลขที่สุ่ม และปรับไป เรื่อย ๆ ทุกครั้งที่เรียนรู้

- Output คือผลลัพธ์

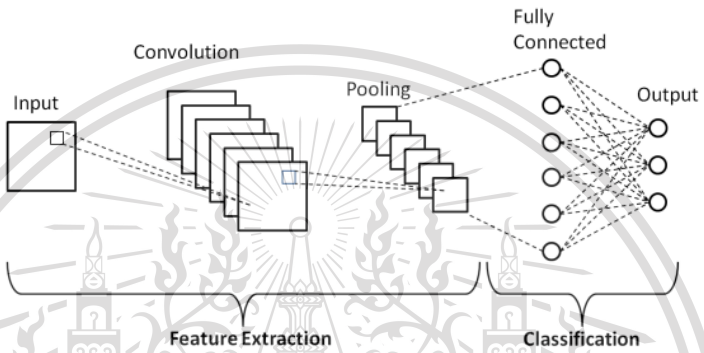
- Back Propagation คือการที่ส่วนประสาทนำค่าผิดพลาดของ Output ที่ได้ กับ Output ที่เรียนรู้ นำไปปรับ Weight และ Bias ให้เกิดผลลัพธ์ที่ถูกต้องตามที่ได้เรียนรู้นั้นเมื่อป้อน Input และ Output เพื่อทำการเรียนรู้ ส่วนประสาทจะทำการบวกตัว Input ด้วย Weight คูณกับค่าที่มาของแต่ละขา จากนั้นบวกค่า Bias แล้วนำไปเข้าฟังก์ชันที่ตัวส่วนประสาทกำหนดไว้ จะได้ Output ออกมา โดยฟังก์ชันก็จะมีหลายแบบขึ้นอยู่รูปแบบที่จะนำไปใช้งาน



รูปที่ 2.1 โครงสร้างของส่วนประสาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นกระบวนการสกัดลักษณะที่สำคัญของภาพออกมา โดยใช้ค่าพิกเซล (Pixel) ซึ่งได้จากข้อมูล อินพุตมีทั้งหมด 3 แชนแนล (Channel) ได้แก่ สีแดง, น้ำเงิน และเขียวสามารถใช้เลข 0 ถึง 255 เพื่อเป็นค่าแทนความเข้มของสี เนื่องจากโครงข่ายประสาทเทียมแบบคอนโวลูชันไม่ค่อยเป็นที่นิยม เพราะต้องใช้ประสิทธิภาพของเครื่องคอมพิวเตอร์อย่างมาก แต่ปัจจุบันมีการพัฒนาหน่วยประมวลผลกราฟฟิก (GPU) ให้รองรับการคำนวณจึงทำให้โครงข่ายประสาทเทียมแบบคอนโวลูชันประสบความสำเร็จในการแก้ปัญหาการจำแจกข้อมูลที่เกี่ยวข้องกับรูปภาพ อีกทั้งในปัจจุบันได้นำอีกแนวคิดไปใช้ต่อยอดมากมาย



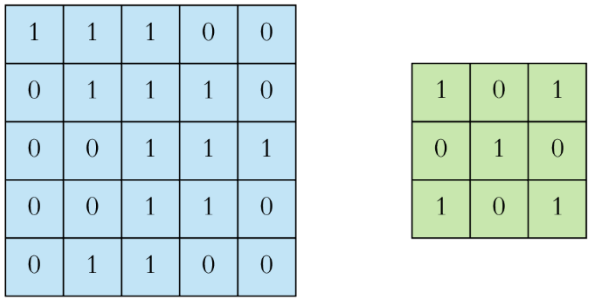
รูปที่ 2.2 โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชันประกอบด้วยดังนี้

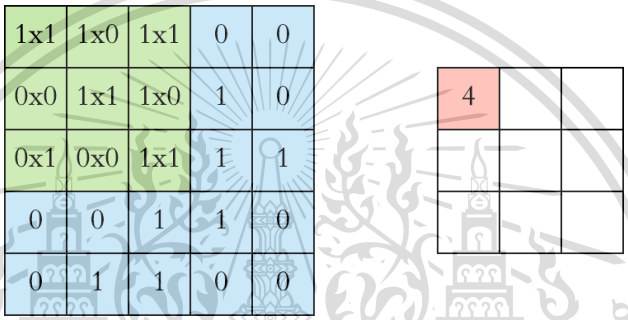
- Convolutional เป็นชั้นหลักของโครงข่ายประสาทเทียมแบบคอนโวลูชันทำหน้าที่รับอินพุตเข้ามาแปลงภาพให้เป็นพิกเซลที่กำหนดให้เป็น 0 – 255 หลังจากนั้นใช้การดำเนินการทางคณิตศาสตร์แปลงเป็นข้อมูล เพื่อหาการสกัดคุณลักษณะ โดยนำมาคูณกับตัวกรองหรือเคอร์เนล (kernel) ที่ทำหน้าที่ดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก โดยปกติตัวกรองหรือเคอร์เนลอันหนึ่งจะดึงคุณลักษณะที่สนใจออกมาหรือที่เรียกว่าผังคุณลักษณะ (Feature Map) ตัวอย่างการหาผังคุณลักษณะหาได้ดังนี้ Feature Map [0,0] = (1x1) + (1x0) + (1x1) + (0x0) + (1x1) + (1x0) + (0x1) + (0x0) + (1x1) = 4

การที่ขยับตัวกรองจะถูกกำหนดด้วยการเลื่อนลำดับ (Stride) ซึ่งสามารถกำหนดค่าของการเลื่อนลำดับให้มากขึ้นได้ ถ้าต้องการให้คำนวณหาคุณลักษณะมีพื้นที่ทับซ้อนกันน้อยลง แต่อย่างไรก็ตามการกำหนดค่าของการเลื่อนลำดับที่มากขึ้นก็ได้จะทำให้ผังคุณลักษณะที่มีขนาดเล็กลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

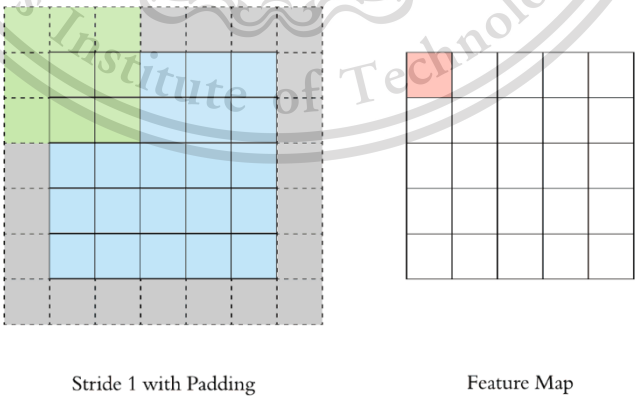


รูปที่ 2.3 อินพุตของตัวกรองหรือเคอร์เนล



รูปที่ 2.4 ผังคุณลักษณะ

- Padding เป็นการกำหนดผังคุณลักษณะให้มีขนาดเท่ากับอินพุตส่วนมากมักจะเติมค่า 0 (สีเทา) หรือค่าต่าง ๆ เข้าไป ในส่วนขงโปรแกรมมักใช้ Padding = same

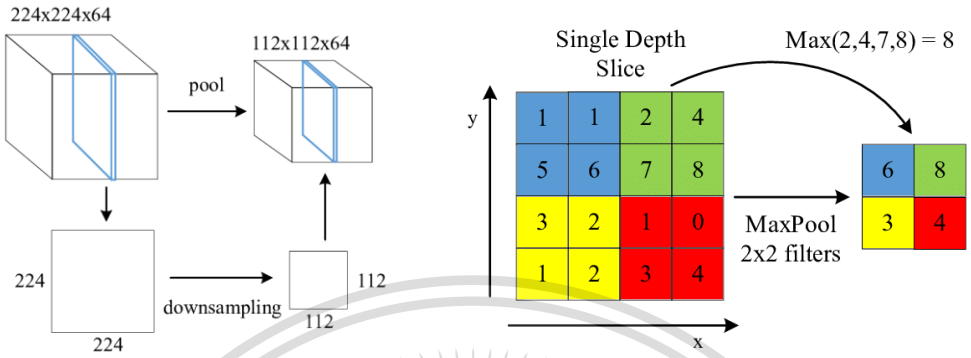


รูปที่ 2.5 การทำ Padding

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

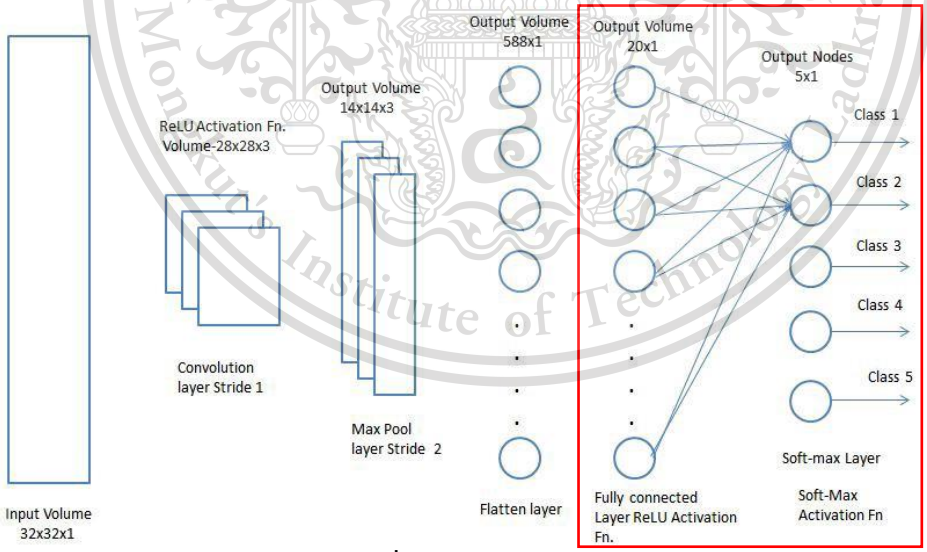
- Pooling เป็นกระบวนการที่ทำหน้าที่ลดขนาดของผังคุณลักษณะที่ได้มาจากการทำโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน มีวัตถุประสงค์ในการลดจำนวนของ

พารามิเตอร์ (Parameter) ที่มีมากเกินไป และลดระยะเวลาในการเรียนรู้ข้อมูล Pooling นั้นมีหลายแบบได้แก่ Max pooling, Min polling, Average polling แต่ที่เป็นที่นิยมใช้คือ Max pooling โดยหาค่ามากที่สุดของแต่ละพิกเซล



รูปที่ 2.6 การทำ Max Pooling

- Fully Connected เป็นชั้นสุดท้ายของการทำโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน ส่วนนี้จะทำหน้าที่นำเอาคุณลักษณะสำคัญไปสร้างเป็นโครงข่ายประสาทเทียมสำหรับการเรียนรู้ จดจำ รูปแบบ และการทำนายประเภท โดยใช้เทคนิคที่ชื่อว่าซอฟต์แมกซ์ (Softmax) เพื่อทำการจำแนกข้อมูล

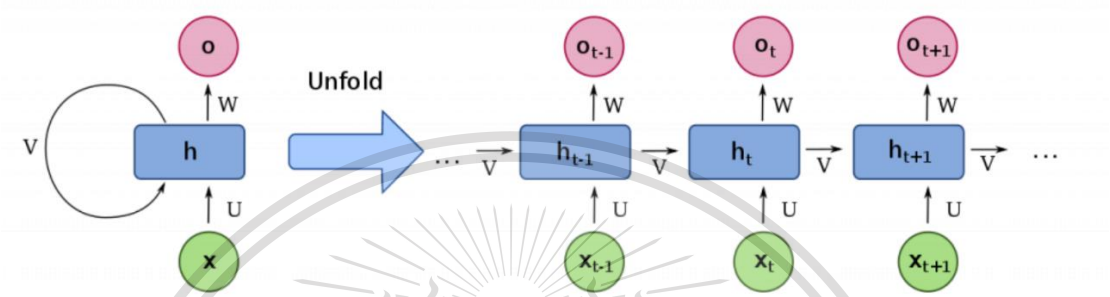


รูปที่ 2.7 การทำ Fully Connected

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.3 โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network)

โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) คือโครงข่ายประสาทเทียมแบบหนึ่ง ซึ่งจะใช้ในการประมวลผลข้อมูลที่มีลำดับ โดยใช้หลักการนำสถานะภายในของโมเดลกลับมาเป็นข้อมูลเข้าใหม่คู่กับข้อมูลเข้าแบบปกติ เรียกว่า สถานะซ่อน (Hidden State) หรือสถานะภายใน (Internal State) ช่วยให้โมเดลรู้จำรูปแบบของลำดับข้อมูลเข้า (Input Sequence) ได้



รูปที่ 2.8 Recurrent neural network

โครงข่ายประสาทเทียมแบบวนซ้ำแบบพื้นฐานในแต่ละโหนด (Node) ของโครงข่ายประสาทเทียมแบบวนซ้ำจะมีข้อมูลเข้าสองอย่าง ได้แก่ ข้อมูลเข้า ณ โหนด นั้น ๆ และผลลัพธ์ที่ได้จากการคำนวณในโหนดก่อนหน้า ซึ่งทั้งสองข้อมูลจะถูกนำมารวมเข้าด้วยกันและออกผลลัพธ์มาเป็นสองทาง คือ ผลลัพธ์ที่ออก ณ โหนดนั้น ๆ และออกเพื่อไปเข้าเป็นข้อมูลขาเข้าในโหนดถัดไป ข้อดีของโครงข่ายประสาทเทียมแบบวนซ้ำ คือ มีการใช้ข้อมูลก่อนหน้าในการทำนายสิ่งที่อาจจะเกิดขึ้นในอนาคต ซึ่งหมายถึงข้อมูลที่เคยเกิดขึ้นในอดีตย่อมส่งผลบทที่ต่อเหตุการณ์ที่จะเกิดขึ้นในอนาคตด้วย โดยโครงข่ายประสาทเทียมแบบวนซ้ำมีสมการดังนี้

$$h_t = f_h(U_h h_{t-1} + W_h x_t + b_h) \quad (2.1)$$

$$y_t = f_y(W_y h_t + b_y) \quad (2.2)$$

โดยที่

f_h คือ activation function ของ hidden layer (เช่น tanh หรือ ReLU หรือ sigmoid function)

f_y คือ activation function ของ output layer (เช่น softmax function)

W_h คือ weight matrix ของ hidden layer

U_h คือ hidden-state-to-hidden-state matrix

จากสมการจะเห็นว่าการคำนวณสถานะซ่อนที่เวลา t (h_t) นั้นจะต้องใช้ 2 ตัวแปรที่สำคัญ คือสถานะซ่อนก่อนหน้า (h_t-1) และข้อมูลเข้า ณ ตอนนั้น (x_t)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น ทั้งนี้ต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.3.1 Backpropagation Through Time

Backpropagation Through Time (BPTT) คือ อัลกอริทึมที่ใช้ในการปรับค่าน้ำหนัก (weights) ในโครงข่ายประสาทเทียมแบบวนซ้ำ โดยการคำนวณความชัน (gradient) ของฟังก์ชั่นการสูญเสีย (loss function) เพื่อช่วยให้โครงข่ายประสาทเทียมสามารถเรียนรู้ข้อมูลได้อย่างถูกต้องและแม่นยำมากขึ้น ในขั้นแรกจะทำการคำนวณไปจนเจอผลลัพธ์ก่อน เรียกกันว่าการส่งผ่านไปข้างหน้า (Forward Pass) หลังคำนวณจนได้ผลลัพธ์ได้แล้ว จึงนำผลลัพธ์ของความผิดพลาดที่ได้มาเพื่อทำการเปลี่ยนแปลงค่าน้ำหนักใหม่ ซึ่งการทำย้อนกลับนี้จะเรียกว่าการส่งผ่านย้อนกลับ (Backward Pass) ในการคำนวณการส่งผ่านไปข้างหน้าจะสนใจ คือ Total Net Input ซึ่งเป็นการคำนวณค่าน้ำหนักคูณกับค่าของแต่ละโหนดและเอาต์พุตที่เป็นการคำนวณคำตอบของโหนดที่ต้องการโดยใช้ฟังก์ชันกระตุ้น (activation function) เข้ามาช่วย

$$net = sigma(weight \times node\ value) + bias \tag{2.3}$$

$$output = activation_function(net) \tag{2.4}$$

เมื่อการคำนวณเสร็จสิ้นจนถึงโหนดของผลลัพธ์ จะสามารถคำนวณค่าความผิดพลาดโดยรวมทั้งหมดของทุกผลลัพธ์ที่ออกมาได้โดยใช้สูตรค่าความผิดพลาดกำลังสอง (Square Error) มาช่วยในการคิดค่าความผิดพลาดโดยรวม ซึ่งจะเรียกค่าความผิดพลาดโดยรวมนี้ว่าความชัน

$$E_{total} = sigma(\left(\frac{1}{2}\right) \times (target - output)^2) \tag{2.5}$$

หลังจากนั้นจะทำการคำนวณย้อนกลับ เพื่อนำค่าความผิดพลาดที่คำนวณมาได้ไปทำการปรับเปลี่ยนค่าน้ำหนัก

$$\frac{\partial \mathrm{E}}{\partial \mathrm{w}} = \frac{\partial \mathrm{E_t}}{\partial \mathrm{o_t}} \frac{\partial \mathrm{o_t}}{\partial \mathrm{h_t}} \frac{\partial \mathrm{h_t}}{\partial \mathrm{h_k}} \frac{\partial \mathrm{h_k}}{\partial \mathrm{w}} \tag{2.6}$$

โดยที่

E คือ ค่าความผิดพลาดโดยรวม

o คือ ผลลัพธ์ที่ได้ในแต่ละ node

h คือ total net input

w คือ น้ำหนัก

$\partial E/\partial o$ คือ การคำนวณว่าค่าความผิดพลาดที่เกิดขึ้นมีผลต่อค่าของผลลัพธ์ที่ได้มากน้อยเพียงใด

$\partial o/\partial h$ คือ การคำนวณความเปลี่ยนแปลงของ total net input ในผลลัพธ์ที่ออกไปในแต่ละโหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

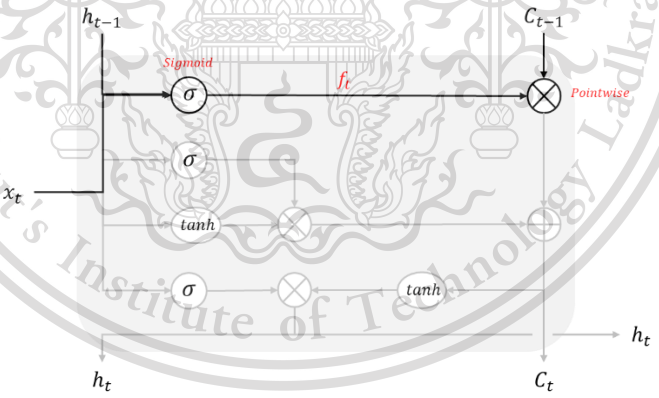
$\partial h/\partial h$ คือ การคำนวณการเปลี่ยนแปลงของ total net input ใน node หนึ่ง ๆ ซึ่งอาจจะส่งผลต่อ total net input ในอีกโหนดหนึ่ง

$\partial h/\partial w$ คือ การคำนวณค่า net input ที่อาจจะส่งผลต่อการเปลี่ยนแปลงในน้ำหนัก

2.3.2 เครือข่ายหน่วยความจำระยะสั้นระยะยาว (Long Short Term Memory Network)

เนื่องจากปัญหาที่เกิดขึ้นในโครงข่ายประสาทเทียมแบบวนซ้ำเกี่ยวกับค่าความชันที่มีค่าน้อยลงจากการทำงานของ back-propagation จึงได้มีการคิดค้นส่วนการเรียนรู้ของเครื่อง (machine learning) ตัวใหม่ที่ใช้หลักการคล้าย ๆ เดิม แต่เปลี่ยนตัวฟังก์ชันด้านในให้มีความเสถียรและมีประสิทธิภาพมากขึ้น ซึ่งนั่นก็คือเครือข่ายหน่วยความจำระยะสั้นระยะยาว (Long Short Term Memory Network) โดยมีจุดเด่น คือ สามารถเลือกได้ว่าข้อมูลไหนที่ควรจะจดจำ ข้อมูลไหนที่ควรจะกำจัดทิ้งออกไป ผ่านการ ‘ลืม’ ของสถานะในโหนดนั้น ๆ ภายในประกอบด้วยฟังก์ชันต่าง ๆ ที่ถูกใช้นำมาประกอบกันเพื่อสร้างฟังก์ชันพิเศษ อันได้แก่ การอ่านข้อมูล การเขียนข้อมูล การอัพเดตข้อมูล และการลืมข้อมูล ซึ่งทำให้การจดจำข้อมูลในแต่ละโหนดได้อย่างคล่องตัวมากขึ้น ฟังก์ชันเหล่านี้เปรียบเสมือนประตูที่คอยควบคุมข้อมูลที่เข้ามาในแต่ละโหนดให้เป็นไปตามทิศทางที่ต้องการ

2.3.2.1 ชั้นประตูลืม (Forget Gate Layer)



รูปที่ 2.9 ชั้นประตูลืม

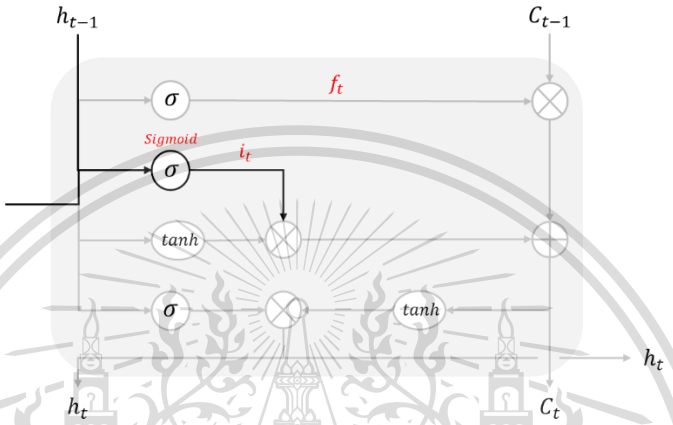
ประตูลืม (Forget gate) คือประตูที่จะใช้เป็นตัวกำหนดว่า ข้อมูลที่เข้ามานั้นสมควรที่จะได้ออกไปทั้งหมดหรือไม่สมควรที่จะมีข้อมูลใด ๆ หลุดรอดออกไป ซึ่งข้อมูลที่จำเป็นต้องใช้ในการตัดสินใจว่าจะเก็บข้อมูลนี้ไว้หรือไม่นั้นก็มาจากข้อมูลขาเข้า ณ โหนดนั้น ๆ รวมกับผลลัพธ์ที่ถูกคำนวณมาจาก โหนดก่อนหน้าผ่านเข้าฟังก์ชันซิกมอยด์ (sigmoid function) และนำไปคูณกับสถานะของโหนดก่อนหน้าโดยใช้สูตรการคูณต่อจุด (Pointwise Multiplication) ในการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.7)$$

ผลลัพธ์ได้ของประตูบานนี้จะอยู่ที่ [0,1] ซึ่งค่า 0 หมายถึงไม่มีข้อมูลใดที่จะสามารถไหลผ่านไปได้ในขณะที่ 1 หมายถึงปล่อยให้ข้อมูลที่เข้ามาไหลผ่านไปได้ทั้งหมด

2.3.2.2 ชั้นประตูทางเข้า (Input Gate Layer)



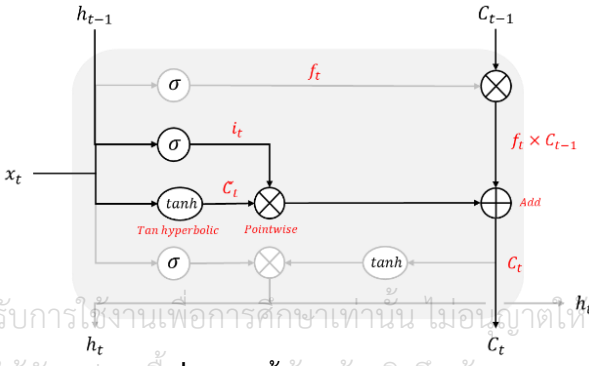
รูปที่ 2.10 ชั้นประตูทางเข้า

ประตูทางเข้า (Input Gate Layer) คือประตูที่ใช้เพื่อเปิดรับข้อมูลที่เข้ามาใหม่เพื่อให้บันทึกลงไปในแต่ละโหนดหรือเรียกว่าการเขียนข้อมูล

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.8)$$

การคำนวณมีวิธีการเดียวกับการคำนวณข้อมูลขาเข้าของประตูลืม

2.3.2.3 ประตูแปลงร่าง (Update Cell State)



รูปที่ 2.11 ชั้นประตูแปลงร่าง

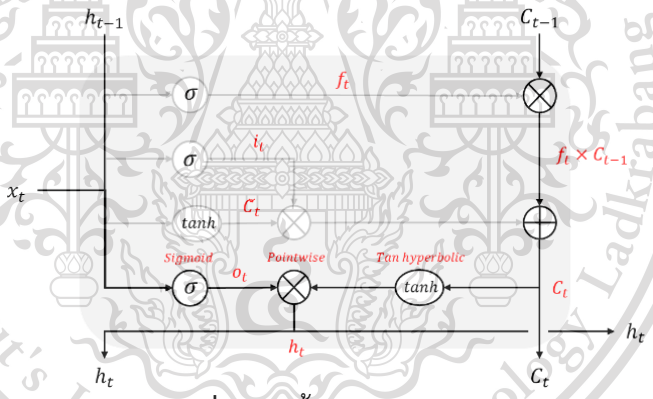
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ระหว่างที่ประตูทางเข้าเปิดประตูเพื่อรับข้อมูลมาจดจำในโหนดของตัวเอง จะมีอีกโหนดหนึ่งที่จะถูกผลักเข้าสู่กระบวนการแปลงโฉมโดยฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (tan hyperbolic) เพื่อให้ได้อีกหนึ่งข้อมูลที่มีความใกล้เคียงกับข้อมูลขาออกจากโหนดนี้ ผลลัพธ์ที่ได้จะเหมือนเป็นผู้ท้าชิงในสนามที่จะถูกนำไปดัดแปลงอีกทีตามสถานะที่ถูกคำนวณมาก่อนหน้าเพื่อส่งเป็นข้อมูลขาออกต่อไป

$$C_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{2.9}$$

เพราะหลังจากคำนวณที่ได้ข้อมูลที่กำลังจะเป็นข้อมูลขาออกแล้ว จะถูกนำไปต่อจุด (pointwise) กับข้อมูลที่ได้จากชั้นประตูทางเข้าและหลังจากได้ผลลัพธ์นี้จะนำไปรวมกับข้อมูลแรกที่ได้จากชั้นประตูลืม เมื่อกระบวนการทั้งหมดเสร็จสิ้นค่าที่ได้ออกมาจะเป็นค่าสถานะใหม่ของโหนดนั้น ๆ ที่ผ่านการอัพเดทสถานะจากโหนดก่อนหน้า

2.3.2.4 ชั้นประตูทางออก (Output Gate Layer)



รูปที่ 2.12 ชั้นประตูทางออก

ชั้นประตูทางออก (Output Gate) คือประตูที่ใช้บอกว่าข้อมูลนี้เตรียมพร้อมที่จะเป็นข้อมูลขาออกแล้วหรือยัง ซึ่งข้อมูลขาออกของแต่ละโหนด ในเครือข่ายหน่วยความจำระยะสั้นระยะยาวนั้นมีด้วยกันทั้งหมด 3 แห่ง แต่มีเพียง 2 ค่า คือ

ค่าที่เกิดจากคำนวณโดยมีการอัพเดทสถานะแล้ว ค่านี้จะถูกส่งต่อทันทีโดยไม่ผ่านฟังก์ชั่นใด ๆ อีกไปให้โหนดถัดไป ค่าสถานะนี้จะดำเนินเป็นเส้นตรงเสมอจากขาเข้าไปจนขาออก โดยจะมีการเปลี่ยนแปลงจากการถูกต่อจุด (Pointwise) ภายในแต่ละโหนด

ค่าของข้อมูลขาเข้าที่ถูกดัดแปลงผ่านการคำนวณในชั้นประตูทางออกค่านี้จะถูกส่งต่อไปเป็นข้อมูลขาเข้าของโหนดถัดไป และยังถูกส่งไปเป็นผลลัพธ์ของโหนดนั้น ๆ อีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

$$O_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o \quad (2.10)$$

หลังจากผ่านการคำนวณค่าซิกมอยด์ (sigmoid) จะถูกนำมาต่อจุด (pointwise) กับค่าสถานะปัจจุบันของโหนดที่ได้ถูกคำนวณมาเรียบร้อยแล้วจากประตูบานอื่นที่ผ่านมา ซึ่งค่าสถานะนั้นจะถูกนำไปเข้าฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ก่อน และจึงนำผลลัพธ์ที่ได้หลังจากเข้าฟังก์ชั่นนั้นแล้วมาต่อจุด (pointwise) กับค่าที่ออกมาจากฟังก์ชันซิกมอยด์

$$h_t = o_t \cdot \tanh(C_t) \quad (2.12)$$

ซึ่งผลลัพธ์ที่ได้จากการกระทำนี้จะถูกแบ่งออกไปเป็นสองข้อมูล คือผลลัพธ์ของโหนดนั้น ๆ และข้อมูลที่จะถูกส่งต่อไปเป็นข้อมูลขาเข้าในโหนดต่อไป

2.4 ฟังก์ชันซอฟต์แมกซ์ (Softmax Function)

ฟังก์ชันซอฟต์แมกซ์ (Softmax Function) หรือ ฟังก์ชันซอฟต์เอดจ์แมกซ์ (Soft ArgMax Function) หรือฟังก์ชันนอร์มัลไลเอกซ์โพเนนเชียล (Normalized Exponential Function) คือฟังก์ชันที่รับอินพุตเป็นเวคเตอร์ (Vector) ของโลจิท (Logit) จำนวนจริง จากนั้นทำการนอร์มัลไล (Normalize) ออกมาเป็นความน่าจะเป็นที่ผลรวมเท่ากับ 1 โดยมีสูตรดังนี้

$$\sigma(z)i = \sigma \frac{e^{z_i}}{\sum j=1^K e^{z_j}} \; for\; i = 1, \dots, K \; and \; z = (z_1, \dots, z_K) \in R^K \quad (2.13)$$

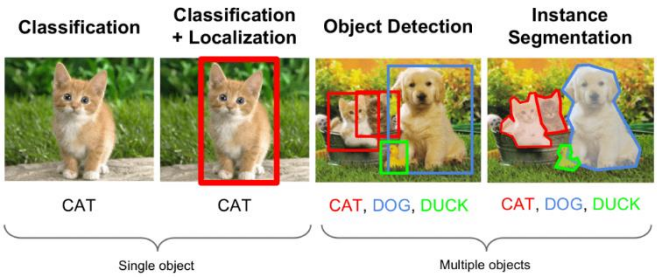
ซอฟต์แมกซ์มักถูกใช้ในชั้นสุดท้ายของโครงข่ายประสาทเพื่อให้เอาต์พุตออกมาเป็นความน่าจะเป็นไปคำนวน Negative Log Likelihood เป็น Cross Entropy Loss

ข้อเสียของซอฟต์แมกซ์คือ เหมาะกับการใช้งานที่คาดหวังเอาต์พุตที่ถูกต้องอันเดียวเท่านั้น เนื่องจากตัวหารต้องค่าผลรวมทั้งหมดทุกครั้ง จึงทำให้มีปัญหาเรื่องประสิทธิภาพ

2.5 การตรวจจับวัตถุ (Object Detection)

การที่จะให้คอมพิวเตอร์มีความสามารถจำแนกวัตถุที่อยู่ในรูปภาพออกมาแล้วทำนายว่าวัตถุเป็นคลาสอะไรต้องอาศัยหลักการของการจำแนกวัตถุ (Object Classification) และการที่จะให้คอมพิวเตอร์หาตำแหน่งของวัตถุที่อยู่ต้องอาศัยหลักการของการตรวจจับวัตถุ (Object Detection) การจำแนกวัตถุสรุปได้ดังนี้ Input Image ⟶ Feature Extractor ⟶ Classifier ⟶ Output Class โดยเป็นการนำข้อมูลเข้า เพื่อสกัดหาคุณลักษณะที่สำคัญทำแล้วจำแนก เพื่อให้ได้เอาต์พุตของคลาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูปที่ 2.13 ตัวอย่างของการตรวจจับวัตถุ

2.5.1 Single Shot MultiBox Detector

Single Shot MultiBox Detector นั้นต้องการเฉพาะข้อมูลนำเข้าของรูปภาพและกล่องข้อมูลที่เป็นความจริง สำหรับวัตถุแต่ละตัวระหว่างสร้างแบบจำลองโดยปกติแล้วจะทำการประเมินตัวอย่างของกล่องตั้งต้นบนหลายๆอัตราส่วนของแต่ละตำแหน่งในแผนที่คุณสมบัติด้วยหลายขนาด (เช่น 8 x 8 และ 4 x 4) ของแต่ละกล่องตั้งต้นและจะทำการคาดเดาขนาดรูปร่างและค่าความมั่นใจของวัตถุทุกหมวดหมู่

2.6 Mediapipe

Mediapipe เป็นแพลตฟอร์ม (Platform) ปัญญาประดิษฐ์แบบโอเพนซอร์ส (Open source) ที่สามารถใช้เป็นไปป์ไลน์ (Pipeline) ตรวจจับและรับรู้ใบหน้า มือ และท่าทางที่มีความซับซ้อน โดยใช้การเร่งความเร็วในการระบุและประมวลผล จึงออกมาเป็นโซลูชันที่แม่นยำและรวดเร็ว

2.6.1 MediaPipe Hands

เป็นการเรียนรู้ของเครื่อง (machine learning) สำหรับการตรวจจับฝ่ามือและนิ้วมือได้อย่างแม่นยำ ซึ่งจะทำการระบุตำแหน่งข้อต่าง ๆ บริเวณมือทั้งหมด 21 จุด ในรูปแบบสามมิติ โดยมีโมเดลอยู่ 2 ส่วน คือ

The palm detector model ทำหน้าที่ในการตรวจจับตำแหน่งฝ่ามือ มีการใช้ Single Shot MultiBox Detector ในการตรวจจับฝ่ามือ ซึ่งสามารถตรวจจับฝ่ามือที่มีหลากหลายขนาดเมื่อเทียบกับขนาดของภาพ และยังสามารถตรวจจับฝ่ามือในขณะที่ถูกบดบังได้อีกด้วย มีความแม่นยำเฉลี่ย 95.7% ในการตรวจจับฝ่ามือ

hand Landmark detector model หลังจากใช้ The palm detector ในการตรวจจับฝ่ามือ จะนำภาพฝ่ามือนั้นมาระบุตำแหน่งบนฝ่ามือและข้อนิ้วมือ 21 จุด และยังทำการลากเส้นเชื่อมแต่ละจุดให้กลายเป็นโครงมือ โดยมีการใช้รูปภาพท่าทางของประมาณ 30000 ภาพมาใช้ในการเรียนรู้เพื่อสร้างโมเดลให้เกิดการทำนายอย่างแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.7 Bazel

Bazel build tools ที่พัฒนาและใช้ภายในองค์กรของ Google ถูกออกแบบด้วยความต้องการที่จะทำการ build software หรือสั่งรันเพื่อทดสอบการทำงานของระบบต่างๆ เป็นไปได้อย่างรวดเร็ว ซึ่งจะต้องตอบสนองต่อแอพพลิเคชั่น (Application) หรือระบบที่มีขนาดใหญ่ได้ด้วย ปัจจุบันทาง Google และทีมพัฒนาได้ผลักดันให้กลายมาเป็นโอเพนซอร์ส

ข้อดีของ Bazel คือ

- High-level build language ออกแบบให้สามารถอ่านและเข้าใจได้ง่าย

- Fast and reliable Bazel เองจะแคช (caches) สิ่งที่ได้ทำงานสำเร็จไปแล้ว และจะสนใจแค่สิ่งที่ได้มีการเปลี่ยนแปลง แล้วจะรีบิวท์ (rebuild) ใหม่แค่ของที่เปลี่ยนเท่านั้น ในกรณีที่อยากได้ความเร็วที่มากขึ้น

- Multi-platform สามารถทำงานได้ทั้งใน Windows, Linux, และ macOS และสามารถบิวท์ไบนารี (build binaries) และดีพอยเอเบิลแพคเกจ (deployable packages) สำหรับหลายแพลตฟอร์ม เช่น desktop server และ mobile ได้ในทีเดียวกัน

- Scalable สามารถ Scale Codebase CI System และสามารถจัดการกับการบิวท์ที่มีจำนวนของซอร์สไฟล์ได้มากกว่า 1 แสนไฟล์อย่างมีประสิทธิภาพ

- Extensible ซัพพอร์ตในหลายภาษา

Bazel จะมีส่วนที่เกี่ยวข้องอยู่ 2 ตัว คือ “WORKSPACE” และ “BUILD”

- WORKSPACE จะเป็นตัวกำหนดว่ากระบวนการบิวท์ของ Bazel ทั้งหมดที่อยู่ภายในไดเรกทอรีโปรเจก (directory project) จะใช้ configuration ผ่านไฟล์ WORKSPACE นี้ทั้งหมด

- BUILD คือไฟล์หลักที่จะกำหนดว่ากระบวนการในการบิวท์เป็นเช่นไร มีการทำงานและผลลัพธ์เป็นแบบไหน

สามารถเรียกใช้ด้วยคำสั่ง “bazel build” ตามด้วย //: และชื่อที่ประกาศไว้ในไฟล์ BUILD เช่น “bazel build //:ProjectRunner”

2.8 Library in Python

ไลบารี (Library) เปรียบเสมือนโปรแกรมสำเร็จรูปที่เก็บฟังก์ชันการทำงานที่เฉพาะทาง ประโยชน์ของไลบารีสำเร็จรูปที่มากับโปรแกรมก็คือผู้ใช้งานไม่จำเป็นที่จะต้องสร้างไลบรารีขึ้นมาใหม่เองทั้งหมด แต่สามารถนำไลบารีที่ได้มีการพัฒนาไว้อยู่แล้วมาใช้งานได้เลย

ไลบารีส่วนใหญ่ของ Python ทั้งที่ถูกพัฒนาโดยผู้พัฒนาหลักและผู้ใช้งานทั่วไปจะถูกสร้างและอัพโหลดขึ้นไปเก็บไว้ในเซิร์ฟเวอร์ (server) ของ Python Package Index (PyPI) หรือไม่ก็เซิร์ฟเวอร์ของ Anaconda Cloud

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.8.1 OpenCV

OpenCV (Open source Computer Vision) เป็นไลบรารีฟังก์ชันการเขียนโปรแกรม โดยส่วนใหญ่จะมุ่งเป้าไปที่การแสดงผลด้วยคอมพิวเตอร์แบบเรียลไทม์ (Real time) โดยตัวอย่างการประยุกต์ใช้งาน OpenCV เช่น ชุดเครื่องมือคุณลักษณะสองมิติและสามมิติ (2D and 3D feature toolkits) การประมาณระยะในขณะเคลื่อนที่ (Egomotion Estimation) ระบบรู้จำใบหน้า (Facial recognition system) ,การจดจำท่าทาง (Gesture recognition) ปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ (Human-Computer interaction; HCI) โดยฟังก์ชันภายใน OpenCV ที่จะใช้ในโปรเจกต์ มีดังนี้

- cv2.imshow เป็นการแสดงภาพบนหน้าต่าง โดยที่ขนาดจะพอดีกับหน้าต่างอัตโนมัติ
- cv2.waitKey เป็นฟังก์ชันหน่วงเวลาเพื่อรับข้อมูลจากคีย์บอร์ด
- cv2.flip เป็นฟังก์ชันการกลับภาพ โดยมีการกลับในแนวนอน (1), การกลับภาพแนวตั้ง (0) และการกลับทั้งแนวนอนและแนวตั้ง (-1)
- cv2.putText เป็นฟังก์ชันการเขียนข้อความลงบนภาพ
- cv2.destroyAllWindows() เป็นฟังก์ชันการปิดทุกหน้าต่างที่ทำการแสดงขึ้น
- cv2.VideoCapture, เป็นฟังก์ชันที่ใช้ในการอ่านไฟล์วิดีโอจากข้อมูลในคอมพิวเตอร์
- cv2.VideoWriter เป็นฟังก์ชันที่ใช้ในการบันทึกวีดีโอ

2.8.2 Tensorflow

Tensorflow คือไลบารีเรียนรู้เชิงลึก (deep learning library) ของ Google โดยทาง Google ได้ใช้การเรียนรู้ของเครื่อง (machine learning) เพิ่มประสิทธิภาพกับผลิตภัณฑ์มากมาย ไม่ว่าจะเป็น เครื่องมือค้นหา (search engine) การแปลภาษา (translation) คำบรรยายภาพ (image captioning) และเครื่องมือช่วยการเสนอแนะ (recommendations) กล่าวคือ Google นำปัญญาประดิษฐ์มาช่วยให้พัฒนาประสบการณ์ของผู้ใช้ ทั้งในแง่ความเร็วของผลลัพธ์และในแง่ผลลัพธ์ที่ถูกต้องแม่นยำมากขึ้น

2.8.2.1 สถาปัตยกรรม Tensorflow

สามารถแบ่งเป็น 3 ส่วน ได้แก่ การเตรียมประมวลผลข้อมูล การสร้างแบบจำลอง และฝึกและประเมินแบบจำลอง

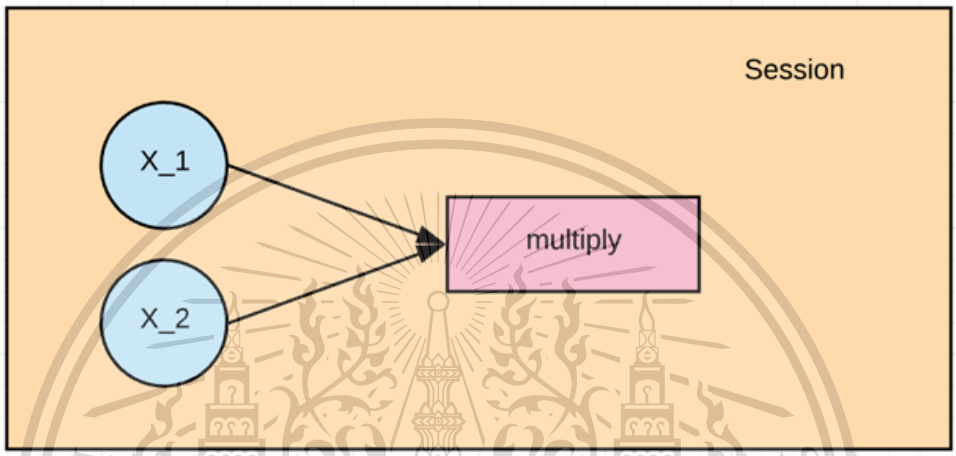
2.8.2.2 ส่วนประกอบของ Tensorflow

ส่วนแรกคือ Tensor การคำนวณทั้งหมดที่เกี่ยวข้องเวกเตอร์และเมทริกซ์หลายมิติ ที่มีข้อมูลจุอยู่หลายหลากชนิด ค่าทั้งหมดในหนึ่ง Tensor จะมีขนาดของข้อมูลแตกต่างกันไปที่เรียกว่ารูปร่าง (shape)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ส่วนถัดมาคือ Graphs โดยจะเป็นตัวรวบรวมและอธิบายชุดการคำนวณทั้งหมดในระหว่างการฝึก สามารถทำงานผ่าน CPUs และ GPUs ได้หลายตัว ทั้งยังทำงานผ่านมือถือได้ ความสามารถในเรื่องการพกพา ทำให้สามารถหยิบใช้งานได้อย่างทันที และสามารถบันทึกกราฟเพื่อดำเนินการต่อในอนาคต การคำนวณทั้งหมดในกราฟเกิดจาก tensor ที่เชื่อมไว้ด้วยกัน

2.8.2.3 ตัวอย่าง Tensorflow



รูปที่ 2.14 รูปตัวอย่างทำงาน tensorflow

จากรูปที่2.14 เป็นการหาผลคูณของ X_1 และ X_2 tensorflow จะทำการสร้างโหนดเพื่อเชื่อมต่อในการดำเนินการแต่ละครั้ง จากภาพคือการคูณ เมื่อกราฟถูกกำหนดแล้ว Tensorflow จึงเริ่มหาผลคูณของ X_1 และ X_2 เมื่อ tensorflow ทำงาน ผลคูณจาก X_1 และ X_2 ก็จะปรากฏออกมาบนหน้าจอแสดงผล

2.8.2.4 Keras

Keras เป็นวิธีเรียกใช้โปรแกรม (Application Programming Interface) โครงข่ายประสาทระดับสูงที่เขียนด้วย Python และสามารถทำงานบน TensorFlow, CNTK หรือ Theano ได้รับการพัฒนาโดยมุ่งเน้นที่การเปิดใช้งานการทดสอบที่รวดเร็ว ความสามารถจากความคิดไปสู่ผลลัพธ์ด้วยความล่าช้าที่อาจเกิดขึ้นน้อยที่สุด โดย Keras สามารถสร้างโครงข่ายประสาทซึ่งรองรับทั้งแบบโครงข่ายประสาทแบบคอนโวลูชันและโครงข่ายประสาทเทียมแบบวนซ้ำ ในการสร้างโมเดลการเรียนรู้ของเครื่องด้วย Keras จะมีวิธีการสร้างโมเดลทั้งหมด 3 แบบ คือ

- Sequential Model จะทำการเรียกแต่ละชั้นที่เพิ่มเข้าไปเรื่อย ๆ แบบตรงไปตรงมา เช่น อินพุตเป็นตัวเลข ชั้นแรกที่เพิ่มเข้าไปจะทำการกำหนดมิติของอินพุตเข้าไป และจำนวน Training Set ชั้นถัดไปคือ Dense หรือ Fully Connected Layer จะทำการเชื่อมต่อชั้นกันทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

- Functional API มีลักษณะคล้ายกับ Sequential Model แต่อินพุตเป็นตัวแปร x ชั้นแรก คือการบอกมิติของอินพุต จากนั้นจะได้ตัวแปรใหม่ เป็นฟังก์ชันสำหรับนำอินพุต มาทำข้อมูลป้อนกลับ (feed forward) ไปเรื่อย ๆ และเอาต์พุตจะเป็นชั้น Dense ปกติ 1 ตัว หรือ function layer หลังจากนั้นจะทำการสร้างโมเดล โดยกำหนดว่าอินพุตและเอาต์พุตคืออะไร

- Model Subclassing จะทำการสร้างคลาส ใหม่ที่สืบทอดมาจาก class keras.Model โดยมี constructor ที่จะทำการเพิ่ม **kwargs เข้าไปในคลาสแม่และสร้างตัวแปรใหม่แบบเดียวกับ functional model ในการสร้างโมเดล จะทำการ override method call จากนั้นจะคล้าย functional model แต่ต่างกันที่ต้องสร้างตัวแปรจากคลาสก่อน

2.8.3 Numpy

NumPy (Numeric Python) เป็นโมดูลส่วนเสริมของ Python ที่มีฟังก์ชันเกี่ยวกับคณิตศาสตร์และการคำนวณต่าง ๆ มาให้ใช้งาน โดยทั่วไปจะเกี่ยวกับการจัดการข้อมูลชุด (Array) ขนาดใหญ่และเมทริกซ์

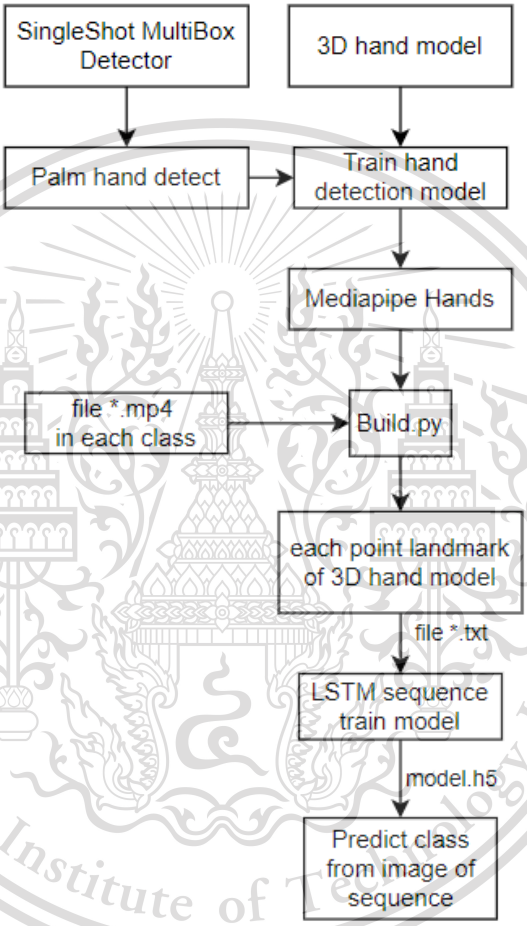
2.8.4 gTTS

gTTS เป็นโมดูลที่ดึง Google TTS (Text-to-Speech) API เข้ามาใช้งาน โดยจะสร้างไฟล์เสียงสังเคราะห์ขึ้นมาในรูปแบบ mp3 ไม่จำกัดความยาว รองรับภาษาต่าง ๆ รวมถึงภาษาไทยด้วย เวลาใช้งานต้องเรียกใช้อินเทอร์เน็ต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3
วิธีการดำเนินการวิจัย

3.1 หลักการทำงาน



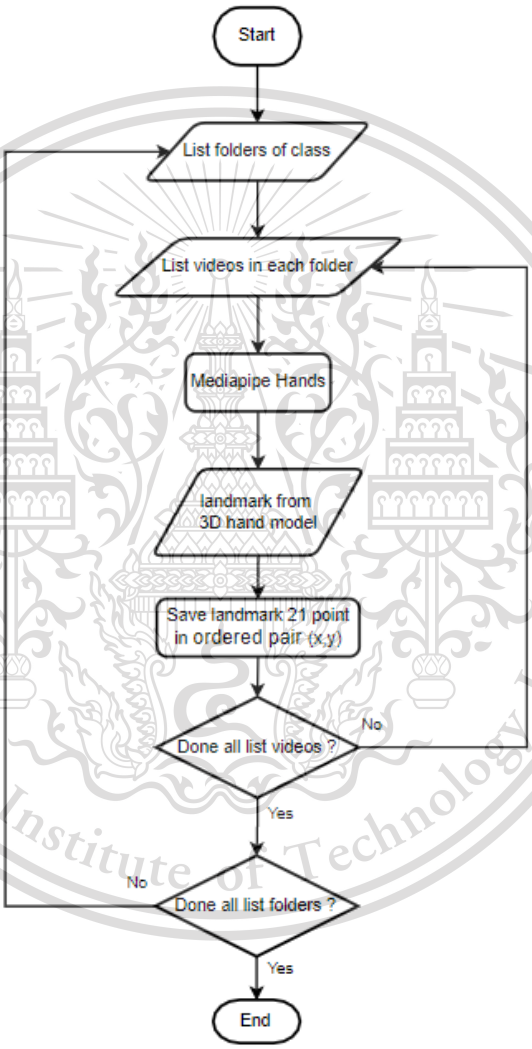
รูปที่ 3.1 Block Diagram แสดงการทำงานของระบบ

จากรูปที่ 3.1 Block Diagram แสดงการทำงานของระบบจำแนกภาษามือด้วยโครงข่ายแบบวนซ้ำ โดยแบ่งเป็น 3 ส่วน คือ ส่วนของการสร้างระบบตรวจจับมือ (Building Phase) ส่วนของการเรียนรู้ระบบที่ใช้ในการทำนายภาษามือ (Training Phase) และส่วนของการทดสอบหรือส่วนของผู้ใช้งาน (Testing Phase) ในส่วนของการสร้างระบบตรวจจับมือ เมื่อรับภาพหรือวิดีโอเข้ามาจะทำการตรวจจับมือ โดยใช้อัลกอริทึมที่ชื่อว่า Mediapipe hands ซึ่งจะใช้ Palm hand detect ในการตรวจจับฝ่ามือที่ใช้หลักการ Single Shot MultiBox Detector (SSD) จากนั้นจะทำการระบุตำแหน่งของข้อนิ้วมือทั้ง 21 จุดจาก โมเดลสำเร็จรูป hand detection และนำตำแหน่งของข้อนิ้วมือมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

จัดเก็บในรูปแบบของไฟล์สกุล.txt ถัดมาในส่วนของการเรียนรู้ระบบที่ใช้ในการทำนาย จะนำข้อมูลที่ได้จัดเก็บในส่วนก่อนหน้านี้มาทำการเรียนรู้เพื่อสร้างโมเดลในการทำนายภาษามือ โดยใช้โมเดล Long Short Term Memory Network (LSTM) ในการเรียนรู้ ทำได้ผลลัพธ์เป็นไฟล์สกุล.h5 ที่ใช้เป็นโมเดลในการทำนายภาษามือในส่วนสุดท้าย คือ ส่วนของการทดสอบหรือส่วนของผู้ใช้งาน

3.2 ส่วนของการสร้างระบบตรวจจับมือ (Building Phase)



รูปที่ 3.2 Block Diagram แสดงระบบตรวจจับมือ

ในขั้นตอนนี้จะเป็นส่วนของการสร้างระบบตรวจจับมือ (Building Phase) ซึ่งเป็นส่วนที่ใช้ในการจัดเก็บข้อมูลเพื่อนำไปใช้ในการสร้างโมเดลทำนายภาษามือ ในขั้นแรกได้ทำการเก็บข้อมูลโดยการถ่ายวิดีโอการทำท่าภาษามือเป็นคำต่าง ๆ โดยที่แต่ละคำจะทำการเก็บเป็นวิดีโออย่างน้อยคำละ 150 วิดีโอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

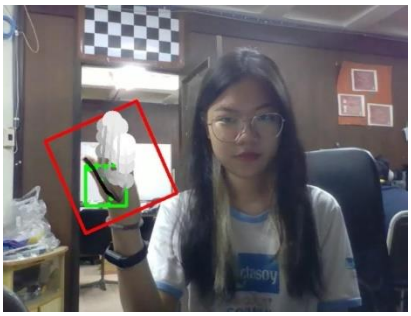
รูปที่ 3.3 ข้อมูลวิดิโอท่าภาษามือ

โปรแกรมที่ 3.1 คำสั่งการนำเข้าวิดิโอท่าภาษามือ เพื่อจัดเก็บค่าตำแหน่งบนฝ่ามือและข้อนิ้วมือ

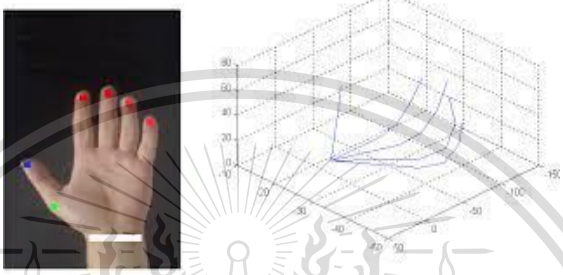
```
cmd = 'bazel-bin\mediapipe\examples\desktop\hand_tracking\hand_tracking_cpu --
calculator_graph_config_file=mediapipe/graphs/hand_tracking/hand_tracking_desktop_liv
e.pbtxt'
inputfilen=' --input_video_path='+input_data_path+word+mp4list
outputfilen=' --output_video_path='+output_data_path+'_'+word+mp4list
cmdret=cmd+inputfilen+outputfilen
os.system(cmdret)
```

จากโคด (code) ด้านบนเป็นคำสั่งในการนำวิดิโอที่ได้รับการจัดเก็บเป็นภาษามือคลาสต่าง ๆ เข้าสู่อัลกอริทึม Mediapipe Hands เพื่อทำการตรวจจับมือและสร้างโมเดลมือสามมิติ โดยมีโมเดลอยู่ 2 ส่วน คือ The palm detector model ที่จะใช้ในการตรวจจับฝ่ามือดังแสดงในรูปที่ 3.4 (ก.) เป็นบริเวณกรอบสีเขียว หลังจากนั้น hand Landmark detector model จะนำโครงสร้างสามมิติของมือที่กำหนดตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้งหมด 21 จุดมากำหนดใส่ในภาพมือในท่าทางต่าง ๆ ดังแสดงในรูปที่ 3.4 (ข.) โดยที่ตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้ง 21 จุด ในแต่ละจุดจะบอกค่าพิกัด x , y และ z โดยที่ค่า x , y และ z จะเป็นค่านอร์มัลไล (normalize) เป็นค่าตั้งแต่ 0 ถึง 1

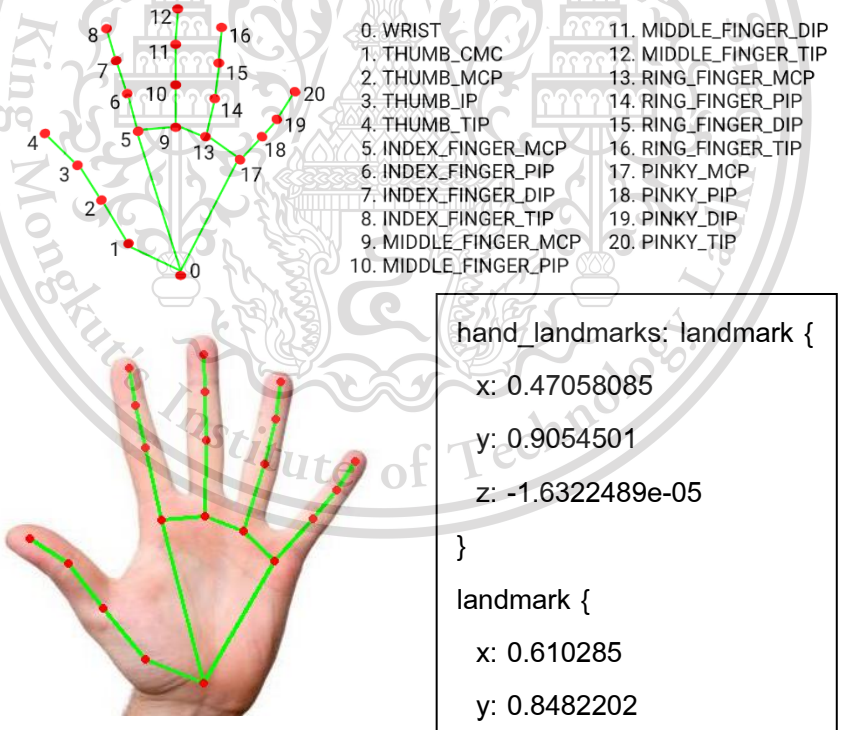
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูปที่ 3.4 การตรวจจับมือและการสร้างโครงสร้างมือสามมิติ



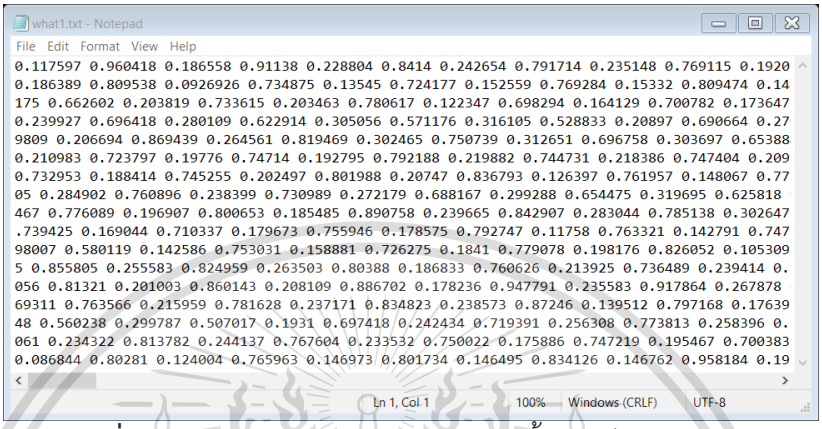
รูปที่ 3.5 การสร้างโครงสร้างสามมิติของมือ



รูปที่ 3.6 พิกัดตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้งหมด 21 จุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

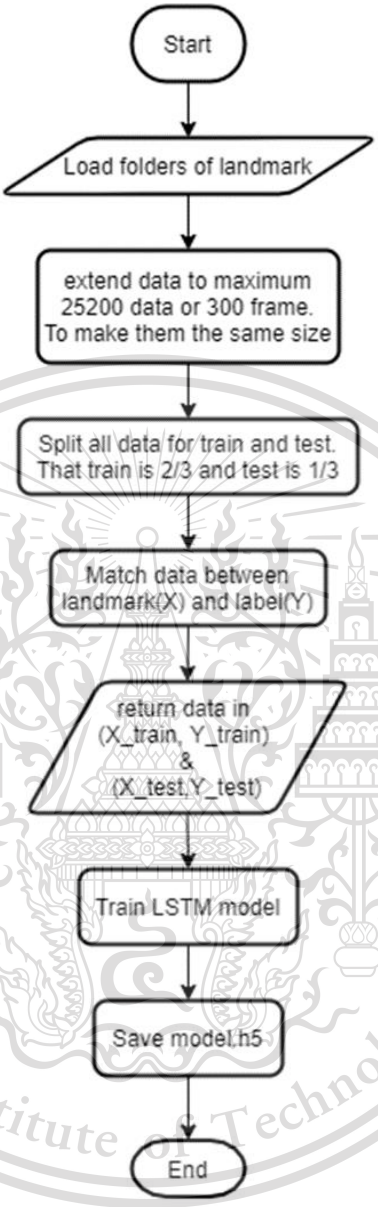
หลังจากนั้นจะทำการนำวิดิโอภาษามือคลาสต่าง ๆ เข้าสู่อัลกอริทึม Mediapipe Hands จะจัดเก็บค่าของตำแหน่งบนฝ่ามือและข้อนิ้วมือในรูปแบบของไฟล์สกุล.txt ดังรูปที่ 3.6 ซึ่งข้อมูลภายในไฟล์สกุล.txt จะแสดงถึงความสัมพันธ์ของการเคลื่อนไหวของมือในท่าทางต่าง ๆ โดยที่จะทำการเก็บค่าตำแหน่งเพียงแค่ค่าในแนวแกน x และแกน y เท่านั้น



รูปที่ 3.7 ข้อมูลตำแหน่งบนฝ่ามือและข้อนิ้วมือในไฟล์สกุล.txt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.3 ส่วนของการเรียนรู้ระบบที่ใช้ในการทำนายภาษามือ (Training Phase)



รูปที่ 3.8 Block Diagram แสดงระบบที่ใช้ในการทำนายภาษามือ

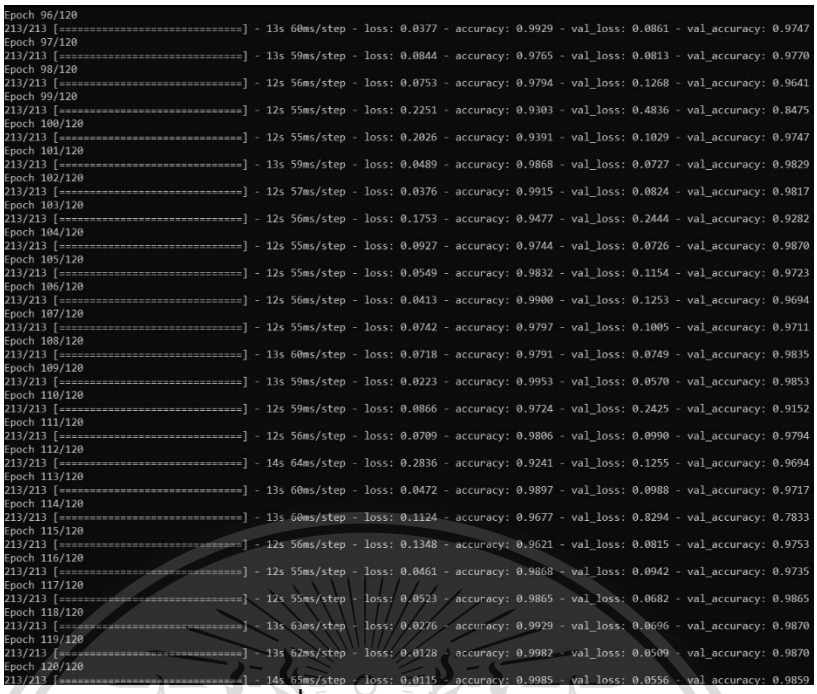
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โปรแกรมที่ 3.2 คำสั่งการนำข้อมูลตำแหน่งบนฝ่ามือและข้อนิ้วมือไปใช้ในการเรียนรู้โมเดล

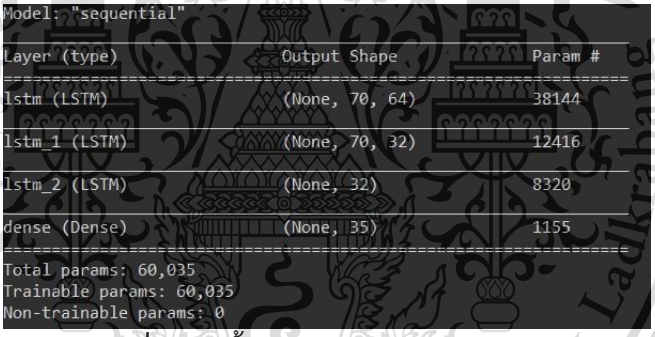
```
for text in textlist:
    if "DS_" in text:
        continue
    textname=dirname+wordname+"/"+text
    numbers=[]
    #print(textname)
    with open(textname,'r') as t:
        numbers = [float(num) for num in t.read().split()]
        for i in range(len(numbers),25200):
            numbers.extend([0.000])
    row=0
    landmark_frame=[]
    for i in range(0,70):
        landmark_frame.extend(numbers[row:row+84])
        row += 84
    landmark_frame=np.array(landmark_frame)
    landmark_frame=list(landmark_frame.reshape(-1,84))
    if (k%3==2):
        XT.append(np.array(landmark_frame))
        YT.append(wordname)
    else:
        X.append(np.array(landmark_frame))
        Y.append(wordname)
    k+=1
```

จากโคดข้างบนเป็นการนำข้อมูลตำแหน่งบนฝ่ามือและข้อนิ้วมือที่ใช้ในการเรียนรู้โมเดล จำเป็นต้องมีขนาดที่เท่ากัน จึงเพิ่มขนาดของวิดีโอให้ในแต่ละวิดีโอมีขนาด 300 เฟรม หลังจากนั้นจะทำการแบ่งข้อมูลที่ใช้เป็นข้อมูลในการเรียนรู้ (Train) และเป็นข้อมูลในการทดสอบ (Test) โดยข้อมูลในการเรียนรู้ (Train) จะนำมาใช้จำนวน 2 ใน 3 ส่วนของข้อมูลทั้งหมดและเป็นข้อมูลในการทดสอบ (Test) จำนวน 1 ใน 3 ส่วนของข้อมูลทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูปที่ 3.9 การ train model



รูปที่ 3.10 ชั้น (Layer) ของโมเดล LSTM

โปรแกรมที่ 3.3 การออกแบบชั้น (Layer) ของเครือข่ายหน่วยความจำระยะสั้นระยะยาว (LSTM)

```
model = Sequential()
model.add(layers.LSTM(64, return_sequences=True, input_shape=(70, 84)))
model.add(layers.LSTM(32, return_sequences=True))
model.add(layers.LSTM(32))
model.add(layers.Dense(label, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam',
metrics=['accuracy'])
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

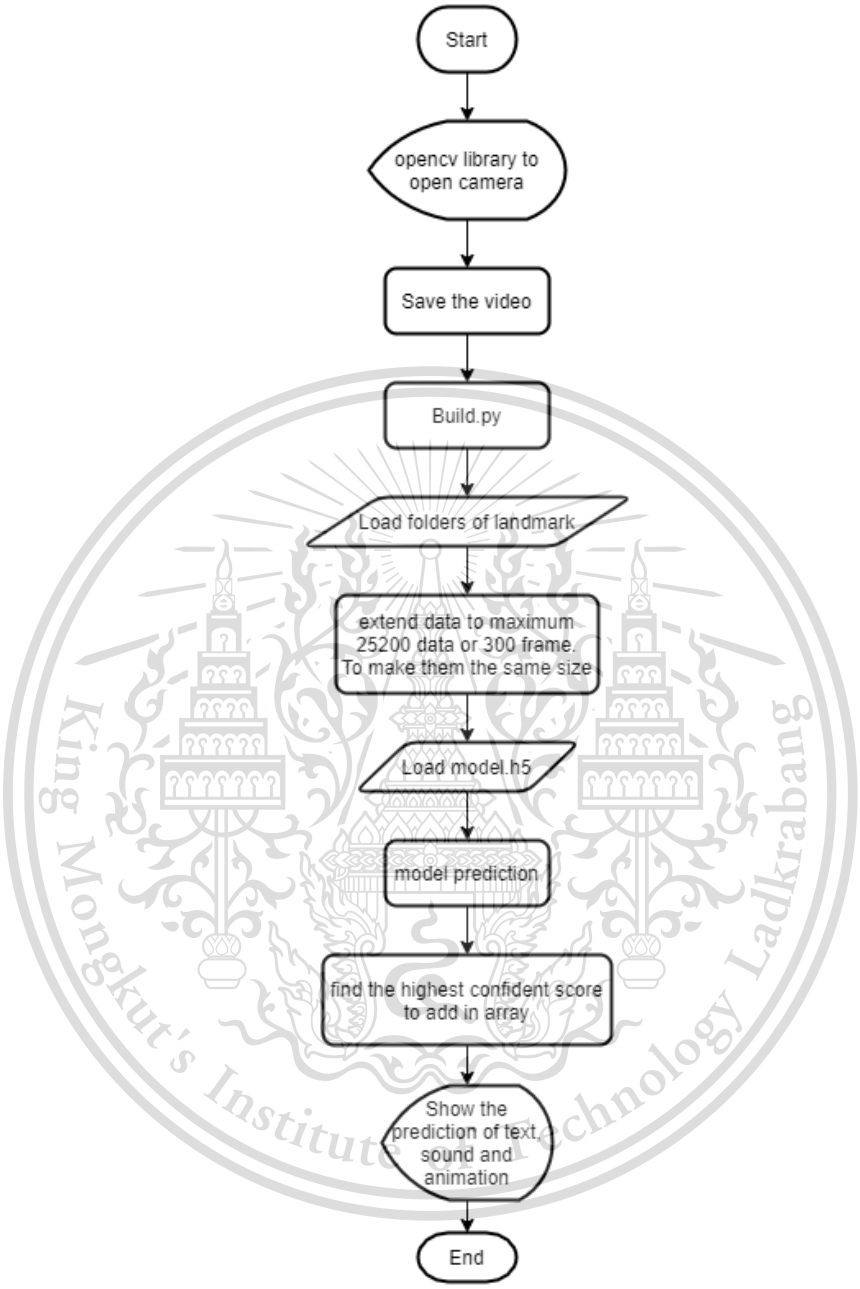
จากโคดด้านบนเป็นเครือข่ายหน่วยความจำระยะสั้นระยะยาวที่จะนำมาสร้างโมเดลในการจำแนกภาษามือที่มีลำดับ (sequential) การเคลื่อนไหวของมือในท่าทางต่าง ๆ โดยโมเดลที่สร้างจะมีจำนวนชั้น (Layer) ทั้งหมด 3 ชั้น และมีชั้นของ Dense 1 ชั้น ที่มีซอฟต์แมกซ์ (softmax) เป็นฟังก์ชันกระตุ้น (activation function) ใช้ categorical_crossentropy มาคำนวณฟังก์ชันการสูญเสีย (loss function) ของฟังก์ชันการจำแนกข้อมูล (classification function) โดยมีสมการดังนี้

$$loss = -\sum_{i=1}^{size}{}^{output} y_i \cdot log\, y_i \tag{3.1}$$

และใช้ Adaptive Moment Estimation (adam) ในการคำนวณค่าน้ำหนักของโมเดลเพื่อให้ค่าสูญเสีย (loss) ที่ต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.4 ส่วนของการทดสอบหรือส่วนของผู้ใช้งาน (Testing Phase)



รูปที่ 3.11 Block Diagram การใช้งานระบบจำแนกภาษามือ

หลังจากสร้างโมเดลที่ใช้ในการจำแนกภาษามือแล้วทำการสร้างระบบสำหรับการใช้งาน จากรูปที่ 3.7 Block Diagram การใช้งานระบบจำแนกภาษามือ เป็นการแสดงการทำงาน จะเริ่มจากการถ่ายและบันทึกวิดีโอการทำท่าทางภาษามือที่ต้องการสื่อสาร จากนั้นวิดีโอที่ถูกบันทึกจะถูกนำเข้าระบบตรวจจับมือ (Building Phase) เพื่อใช้ในการรตรวจจับมือและทำการจัดเก็บค่าของตำแหน่งบนฝ่ามือและข้อนิ้วมือในท่าทางการเคลื่อนไหวต่าง ๆ ในรูปแบบของไฟล์สกุล.txt เมื่อบันทึกค่าของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

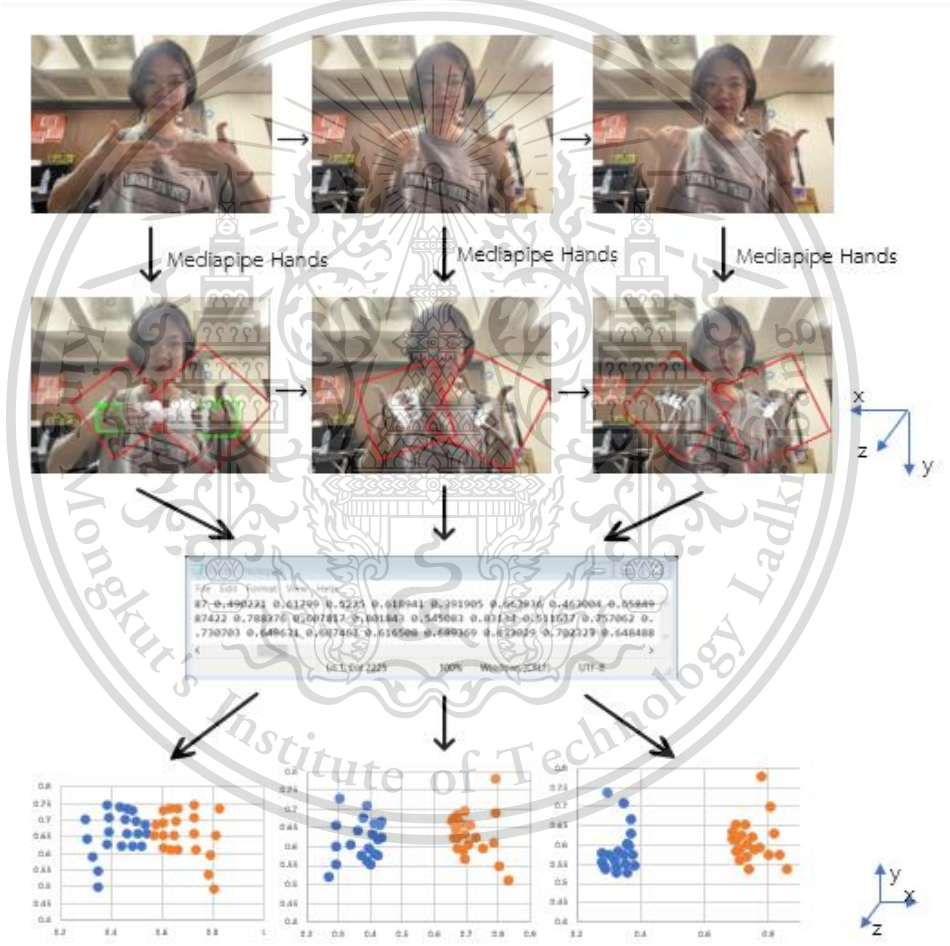
ตำแหน่งบนฝ่ามือและข้อนิ้วมือในรูปแบบของไฟล์สกุล.txt แล้ว จะนำข้อมูลส่วนนี้เข้าสู่โมเดลที่ใช้ในการทำนายภาษามือ โดยโมเดลจะทำการทำนายท่าทางที่แสดงในวิดิโอว่ามีความหมายใกล้เคียงกับคำใดมากที่สุดที่โมเดลได้เรียนรู้มา หลังจากนั้นระบบจะแสดงคำที่ถูกทำนายออกมา พร้อมกับเสียงและแอนิเมชันสื่อถึงคำที่ถูกทำนายออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4
ผลการทดลอง

4.1 ผลการตรวจจับมือ

ในส่วนของการสร้างระบบตรวจจับมือ (Building Phase) ผู้ทำการวิจัยมีเป้าหมายจัดเก็บข้อมูลที่เป็นตำแหน่งคู่อันดับ (x ,y) บนฝ่ามือและข้อนิ้วมือทั้ง 21 จุด โดยจะจัดเก็บข้อมูลในรูปแบบไฟล์สกุล.txt



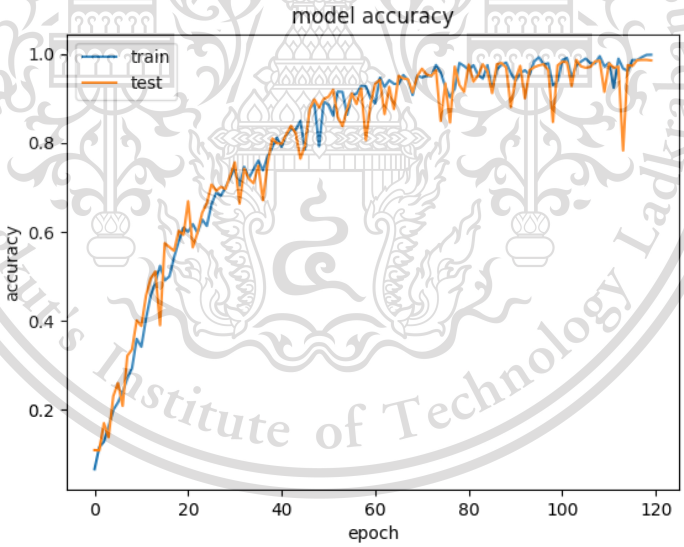
รูปที่ 4.1 แสดงการพล็อตค่าตำแหน่งคู่อันดับ (x ,y) บนฝ่ามือและข้อนิ้วมือทั้ง 21 จุด

จากรูป 4.1 แสดงการพล็อตค่าตำแหน่งคู่อันดับ (x ,y) บนฝ่ามือและข้อนิ้วมือทั้ง 21 จุด จะเห็นได้ว่าค่าตำแหน่งบนฝ่ามือและข้อนิ้วมือที่ผ่านอัลกอริทึม Mediapipe Hands และจัดเก็บอยู่ในรูปแบบไฟล์.txt จะมีความสัมพันธ์กับการทำมือในท่าทางต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.2 ผลการเรียนรู้ของโมเดล LSTM

รูปที่ 4.2 กราฟแสดงค่า loss



รูปที่ 4.3 กราฟแสดงค่า accuracy

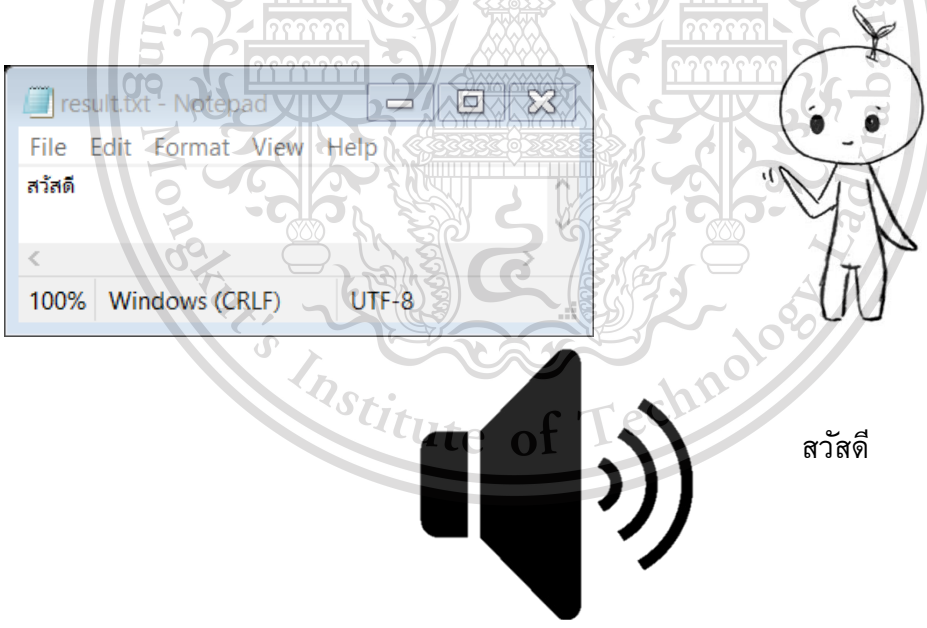
จากรูปที่ 4.2 และรูปที่ 4.3 เป็นกราฟที่แสดงค่า loss และค่า accuracy โดยทำการเรียนรู้โมเดล (Train) จำนวน 120 ครั้ง มีใช้ Adaptive Moment Estimation (adam) เป็นอัลกอริทึม Optimization ในการปรับค่าน้ำหนัก (Weight) และค่า Bias ใช้ categorical_crossentropy มาคำนวณฟังก์ชันการสูญเสีย (loss function) พบว่าจะมีค่า loss ที่ประมาณ 0.08503 และมีค่า accuracy ที่ประมาณ 0.9794

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.3 การทดสอบระบบ

4.3.1 การแสดงผลเอาต์พุตของระบบ

รูปที่ 4.4 ข้อมูลอินพุตทเป็นวิดีโอท่าทางภาษามือ



รูปที่ 4.5 ผลเอาต์พุตของระบบ

ทำการทดสอบการทำงานของระบบ โดยการอัดวิดีโอทำท่าภาษามือดังรูป 4.4 และผลของเอาต์พุตดังแสดงในรูป 4.5 จะมีเอาต์พุตอยู่ 3 ส่วน คือ ส่วนที่แสดงคำแปลของภาษามือ ส่วนที่แสดงแอนิเมชันที่เกี่ยวข้องกับคำแปล และส่วนที่เป็นเสียงพูดของคำที่แปลได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.3.2 การทดสอบประสิทธิภาพของระบบ

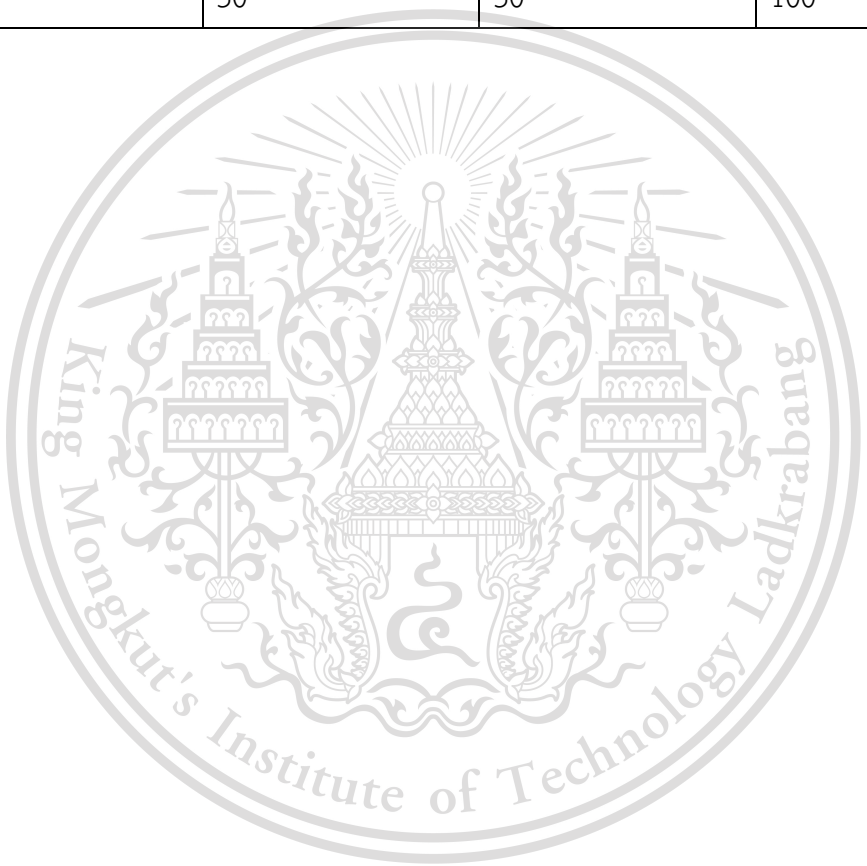
ตารางที่ 4.1 ตารางการทดสอบประสิทธิภาพของระบบ

คำแปลภาษามือ	วิดิโอทั้งหมด	การทำนายถูกต้อง	ร้อยละ (%)
เลข 1	50	49	98
เลข 10	50	44	88
เลข 100	50	49	98
เลข 2	50	42	84
เลข 20	50	50	100
เลข 3	50	50	100
เลข 30	50	44	88
เลข 4	50	49	98
เลข 40	50	49	98
เลข 5	50	50	100
เลข 50	50	48	96
เลข 6	50	48	96
เลข 60	50	37	74
เลข 7	50	44	88
เลข 70	50	48	96
เลข 8	50	46	92
เลข 80	50	48	96
เลข 9	50	50	100
เลข 90	50	49	98
คุณ	50	42	84
ฉัน	50	48	96
ชื่อ	50	40	80
มาก	50	41	82
สบายดี	50	50	100
สวย	50	19	38
สวัสดี	50	50	100
หล่อ	50	29	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ 4.1 (ต่อ) ตารางการทดสอบประสิทธิภาพของระบบ

คำแปลภาษามือ	วิดิโอทั้งหมด	การทำนายถูกต้อง	ร้อยละ (%)
อะไร	50	49	98
อายุ	50	43	86
เข้าใจ	50	31	62
เท่าไหร่	50	45	90
ไม่	50	38	76
ไม่เป็นไร	50	41	82
ไหม	50	50	100



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 5
สรุปผลการทดลอง

5.1 สรุปและวิจารณ์ผลการทดลอง

การวิเคราะห์จำแนกคำภาษาไทยในชีวิตประจำวันของภาษามือได้แบ่งขั้นตอนในการทำงานออกเป็น 3 ส่วน ได้แก่ ส่วนของการสร้างระบบตรวจจับมือ (Building Phase) โดยใช้อัลกอริทึม Mediapipe Hands ที่มีความสามารถในการตรวจจับมือและกำหนดตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้งหมด 21 จุด โดยตำแหน่งแต่ละจุดจะบอกค่าพิกัด x , y และ z ที่เป็นค่านอร์มัลไล (normalize) ตั้งแต่ 0 ถึง 1 ซึ่งทำการเก็บค่าตำแหน่งในแนวแกน x และแกน y ในรูปแบบไฟล์สกุล .txt ที่แสดงความสัมพันธ์ของการเคลื่อนไหวของมือในท่าทางต่าง ๆ แล้วจึงเข้าสู่ส่วนของการเรียนรู้ระบบที่ใช้ในการทำนายภาษามือ (Training Phase) เป็นการนำข้อมูลท่าภาษามือเข้าสู่การเรียนรู้โมเดลเครือข่ายหน่วยความจำระยะสั้นระยะยาว และใช้ categorical_crossentropy มาคำนวณฟังก์ชันการสูญเสีย (loss function) พบว่าค่า loss ที่ประมาณ 0.08503 และมีค่า accuracy ที่ประมาณ 0.9794 ในส่วนของการทดสอบหรือส่วนของผู้ใช้งาน (Testing Phase) จากการทดลองพบว่าโมเดลเครือข่ายหน่วยความจำระยะสั้นระยะยาวสามารถทำนายผลลัพธ์ได้อย่างแม่นยำ มีเปอร์เซ็นต์โดยรวม 80 – 100 % แต่ว่าในบางคลาสมีเปอร์เซ็นต์ความถูกต้องที่ต่ำลงมา เนื่องจากข้อมูลที่ใช้ในการเรียนรู้ของคลาสนั้นไม่ครอบคลุมและมีท่าทางใกล้เคียงกับท่าอื่น ๆ เช่น ในคลาสของคำว่าสวยกับหล่อ

วิดีโอที่นำมาวิเคราะห์นั้นต้องเห็นลักษณะของมือที่ชัดเจน เพื่อให้อัลกอริทึม Mediapipe Hands สามารถใช้วิธีการตรวจจับมือและกำหนดตำแหน่งบนฝ่ามือและข้อนิ้วมือทั้งหมด 21 จุดได้ดี ผู้ทำท่าทางต้องแสดงท่าทางที่ถูกต้อง กล่าวคือต้องไม่บิดมือไปทางซ้ายหรือขวามากเกินไป รวมถึงการไม่ทำฝ่ามือตั้งฉากกับกล้อง ซึ่งจะทำให้ข้อมูลผิดเพี้ยนไปจากสิ่งที่ควรจะเป็น และจากโมเดลไม่สามารถทำนายกท่าทางของมือในระยะกลางหรือระยะไกลจากกล้องได้ เนื่องจากข้อมูลที่นำมาเรียนรู้มีเพียงแค่ในระยะใกล้เท่านั้น ซึ่งในการแก้ปัญหาทำได้โดยเพิ่มข้อมูลที่ใช้ในการเรียนรู้ในระยะกลางและระยะไกลด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

กระทรวงศึกษาธิการ. (2542). **พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2542** กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.

จิตประภา ศรีอ่อน. (2543) **คู่มือการใช้ล่ามภาษามือไทยในห้องเรียน** นครปฐม: วิทยาลัยราชสุดา-มหาวิทยาลัยมหิดล.

ชิตพงษ์ กิตตินราดร (มกราคม 2563) **Neural Network Optimizers** สืบค้นจาก https://guopai.github.io/ml-blog17.html

ศศิธร ทรัพย์วัฒนไพศาล (2554) **การศึกษาผลการใช้บทเรียนมัลติมีเดีย เรื่อง คำศัพท์ภาษามือไทย โดยการเล่านิทานที่มีผลต่อทักษะการใช้ภาษามือของนักศึกษาล่ามภาษามือวิทยาลัยราชสุดา มหาวิทยาลัยมหิดล** (วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยศิลปกร, กรุงเทพฯ) สืบค้นจาก https://bit.ly/33eeK5u

Antonio Gulli & Amita Kapoor (2017) **TensorFlow 1.x Deep Learning Cookbook: Over 90 unique recipes to solve artificial-intelligence driven problems with Python** UK: Packt Publishing Ltd

Antonio Gulli & Sujit Pal (2017) **Deep Learning with Keras** UK: Packt Publishing Ltd

C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, Cl. Chang, MG. Yong, J. Lee & WT. Chang (2019) **Mediapipe: A frame work for building perception pipelines** Retrieved from https://arxiv.org/abs/1906.08172

F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, CL. Chang & M. Grundmann. (2020) **MediaPipe hands: on-device real-time hand tracking** Retrieved from https://arxiv.org/abs/2006.10214

Google Brain (2016) **TensorFlow: A System for Large-Scale Machine Learning** Retrieve From https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Ivan Culjak, David Abram, Tomislav Pribanic, Hrvoje Dzapo & Mario Cifrek (2012) **A brief introduction to OpenCV** Retrieved from https://ieeexplore.ieee.org/document

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

MR. Abid, EM. Petriu & E. Amjadian (2015) **Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar** from https://ieeexplore.ieee.org/document/6894169

Peltarion. (2004) **Categorical crossentropy** Retrieved from https://peltarion.com /knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy

R. Sharma, R. Khapra & N. Dahiya (2020) **Sign Language Gesture Recognition in Sign** Retrieved from https://www.researchgate.net/publication/216679291_Hand_ Gesture_Recognition_for_Sign_Language_A_New_Hybrid_Approach

Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink & Jürgen Schmidhuber (2017) **LSTM: A Search Space Odyssey** Retrieved from https://arxiv.org/abs/ 1503.04069

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu & Alexander C. Berg (2015) **SSD: Single Shot MultiBox Detector** Retrieved from https://arxiv.org/abs/1512.02325

Wojciech Zaremba, Ilya Sutskever & Oriol Vinyals (2015) **Recurrent Neural Network Regularization** Retrieved from https://arxiv.org/abs/1409.2329

W. Liu, Y. Fan, Z. Li & Z. Zhang (2015) **RGBD video based human hand trajectory tracking and gesture recognition system** Retrieved from https://www.hindawi. com/journals/mpe/2015/863732/

Zhaowei, CaiE, Quanfu Fan, Rogerio S. Feris & Nuno Vasconcelos (2016) **A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection** Retrieved from https://arxiv.org/abs/1607.07155

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เว็บไซต์รวบรวมโคดโปรแกรม

https://github.com/kavinteng/Sign-language-recognition

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.