ประสิทธิภาพการรู้จำภาษามือไทยโดยใช้เครือข่ายประสาทเทียมแบบคอนโวลูชัน 2 มิติ

เอกบดินทร์ เกตุขาว*

[1]สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏจันทรเกษม, กรุงเทพฯ
*ผู้รับผิดชอบบทความ: email eakbodin.g@chandra.ac.th

บทคัดย่อ

ในบทความนี้ ได้เสนอการใช้ 2D Convolutional Neural Networks สำหรับการรู้จำลักษณะท่าทางภาษามือไทย โดยได้ฝึกเครือข่ายเชิงลึกแบบ end-to-end สำหรับการรับรู้ท่าทางแบบต่อเนื่อง ซึ่งเครือข่ายที่ใช้คือ 2D convolutions เพื่อดึงข้อมูลคุณลักษณะที่เกี่ยวข้องกับลักษณะท่าทางของภาษามือไทย โดยได้ทำการออกแบบโมเดล 2D CNN และการสร้างชุดข้อมูลลักษณะท่าทางของภาษามือไทย ซึ่งใช้ 2D convolution และ pooling layers ช่วยในการเรียนรู้ความแตกต่างของข้อมูล โดยได้ใช้ลักษณะท่าทางภาษามือไทย จำนวน 3 ท่าทาง คือ สวัสดี รัก และไม่สบาย ทำการเลือกภาพท่าทางละ 1000 ภาพรวม 3000 ภาพในการเรียนรู้และใช้ท่าทางละ 100 ภาพรวมเป็น 300 ในการทดสอบ โดยในการทดลองได้แสดงให้เห็นว่าโมเดลที่ได้ออกแบบสามารถช่วยเพิ่มประสิทธิภาพการรับรู้ท่าทางอย่างมาก ได้ค่าความแม่นยำ (Accuracy) เป็น 0.93 และได้ค่าการสูญเสีย (Loss) เป็น 0. 27 ที่ได้รับจาก 2D CNNs ใช้เวลาในการเรียนรู้ทั้งหมด 1 ชั่วโมง 15 นาที 46 วินาที

คำสำคัญ: เครือข่ายประสาทเทียมแบบคอนโวลูชัน 2 มิติ, การรู้จำลักษณะท่าทางภาษามือ, การเรียนรู้เชิงลึก

The Performance of Thai Sign Language Recognition Using 2D Convolutional Neural Networks

Eakbodin Gedkhaw[1*]

[1]Department of Business Computer, Faculty of Management Science, Chandrakasem Rajabhat University, Bangkok

*corresponding author: email eakbodin.g@chandra.ac.th

Abstract

In this paper propose 2D Convolutional Neural Networks in Thai sign language recognition. Network was train by end-to-end for continuous gesture recognition. 2D convolutions was choose to extract features related to the gestures of Thai sign language. The design of the 2D CNN model and the generate of Thai sign language gesture data set, using 2D convolution and pooling layers, was used to train differentiation of the data. In experiment use 3 Thai sign language gestures: "Hello", "Love" and "Sick", 1000 images of each gesture, total 3000 images were selected in training and used 100 images of each gesture, total of 300 images were selected to test. The experiment shown that the designed model can greatly enhance the gesture perception, accuracy value is 0.93 and loss value is 0.27 obtained from 2D CNNs. The total learn time is 1 hr 15 min 46 sec.

Keywords: 2D Convolution Neural Network, Sign Language Recognition, Deep Learning

1. บทนำ

ลักษณะท่าทางภาษามือไทยเป็นรูปแบบธรรมชาติที่ใช้ในการสื่อสารของมนุษย์ที่มีความผิดปกติทางด้านการได้ยิน เพื่อให้ผู้พิการทางการได้ยินสามารถใช้มือในการสื่อความหมายและถ่ายทอดอารมณ์แทนการพูดหรือการสื่อสาร [1] สำหรับการสื่อสารโดยใช้ภาษามือไทยจะใช้ลักษณะท่าทางของมือที่ทำเป็นสัญลักษณ์ มีการเคลื่อนไหวมือ แขนและร่างกาย และการแสดงความรู้สึกทางใบหน้าเพื่อช่วยในการสื่อสารความคิดของผู้ที่ต้องการสื่อสาร แต่ข้อจำกัดทางด้านการสื่อสารระหว่างผู้พิการทางการได้ยินกับบุคคลทั่วไปยังมีอยู่มาก มีเพียงผู้พิการเองกับผู้ที่สนใจศึกษา หรือทำงานที่เกี่ยวข้องกับผู้พิการทางการได้ยินเท่านั้น จึงทำให้ยังคงจำกัดในการสื่อสารกับผู้พิการทางการได้ยินด้วยภาษามือ โดยในประเทศไทยนั้นพบว่าผู้ที่มีความบกพร่องทางการได้ยิน จำนวน 278,550 คน คิดเป็นร้อยละ 17.78 จากจำนวนผู้พิการทั้งหมด 1,568,847 คน

ปัจจุบันความก้าวหน้าทางด้าน Computer vision ได้พัฒนาไปอย่างต่อเนื่อง โดยนักวิจัยได้ร่วมกันพัฒนาเทคโนโลยีให้เข้ามามีส่วนช่วยในการอำนวยความสะดวกให้กับผู้พิการหรือช่วยให้มนุษย์สามารถปฏิสัมพันธ์กับคอมพิวเตอร์ได้ (Human Computer Interaction) [2] ซึ่งการรู้จำ (Recognition) ก็เป็นอีกหนึ่งวิธีการในการให้คอมพิวเตอร์สามารถเรียนรู้พฤติกรรมของ

มนุษย์ได้ และความท้าทายที่ต้องการให้คอมพิวเตอร์สามารถเข้าใจภาษาธรรมชาติมากขึ้น ทำให้เป็นข้อดีในการนำระบบการรู้จำเข้ามามีส่วนช่วย เพื่อให้ผู้ที่มีความบกพร่องทางการได้ยินสามารถที่จะทำการติดต่อสื่อสารกับบุคคลโดยทั่วไปได้

จากการศึกษาพบว่ามีงานวิจัยเกี่ยวกับการรู้จำลักษณะท่าทางของภาษามือนั้นแบ่งเป็นสองวิธีที่นิยมใช้ คือ ใช้เทคนิควิธีการแบบ Vision base และเทคนิควิธีการแบบ Sensor base ซึ่งข้อดีของการใช้ Vision base คือ ไม่จำเป็นต้องมีอุปกรณ์ที่ซับซ้อนมาก เน้นการคำนวณเป็นหลัก อุปกรณ์ที่ใช้ เช่น กล้องดิจิทัล Kinect Sensor หรือ Leap Motion เป็นต้น โดยข้อมูลจะได้จากกล้องดิจิทัล จากนั้นนำข้อมูลไปทำการคำนวณที่ซับซ้อน มีการปรับปรุงภาพในการประมวลผลเบื้องต้น เพื่อทำการคัดแยกคุณลักษณะเฉพาะที่ต้องใช้ในกระบวนการรู้จำ ส่วนการใช้ Sensor base จะเป็นการใช้อุปกรณ์เซ็นเซอร์เข้ามาใช้งานตั้งแต่ขั้นตอนการรับข้อมูลดิบที่ส่งตรงจากอุปกรณ์เซ็นเซอร์ เช่น การใช้เซ็นเซอร์ Data Glove [3] หรือการใช้สัญญาณ Surface EMG [4] เป็นต้น เพื่อนำข้อมูลดิบที่ได้ไปประมวลในกระบวนการรู้จำต่อไป

สำหรับเทคนิคในการรู้จำ สามารถแบ่งกลุ่มได้เป็น 5 กลุ่ม [5] คือ เทคนิคแบบใช้ขั้นตอนวิธีในการคำนวณทางตรรกะในการเรียนรู้ เช่น การใช้ Decision trees เป็นต้น เทคนิคแบบใช้ขั้นตอนวิธีในการรับรู้เพื่อการเรียนรู้ เป็นการเรียนรู้โดยใช้ปัญญาประดิษฐ์ เช่น การใช้ Artificial Neural Networks (ANN) เป็นต้น เทคนิคแบบใช้ขั้นตอนวิธีทางสถิติสำหรับการเรียนรู้ เช่น การใช้ k-nearest neighbors (kNN) เป็นต้น เทคนิคแบบใช้ Support Vector Machines ซึ่งจะใช้ Support Vector Machines ในการเรียนรู้เพื่อการรู้จำหรือใช้ร่วมกับ Hidden Makov Model (HMM) ในการรู้จำได้อย่างมีประสิทธิภาพ และสุดท้าย เทคนิคแบบใช้การรู้จำเชิงลึก อาทิเช่น การใช้ Convolutional Neural Networks (CNNs) และการใช้ Recurrent Neural Network (RNN) เป็นต้น

จากการศึกษาพบว่าปัจจุบันการรู้จำลักษณะท่าทางของภาษามือไทยยังมีข้อจำกัด คือยังไม่มีงานวิจัยที่สามารถนำมาพัฒนาเป็นนวัตกรรมที่สามารถทำการแปลภาษามือให้คนปกติสามารถเข้าใจได้ ในการแปลลักษณะท่าทางมาเป็นข้อความหรือคำพูด อีกทั้งทางด้านของภาษาและการรู้จำลักษณะท่าทางของภาษามือด้วยคอมพิวเตอร์ที่จะทำให้เทคโนโลยีสามารถเข้าใจภาษาหรือสื่อสารได้ยังมีความท้าทายอย่างยิ่งในปัจจุบัน ประกอบกับวิธีการเรียนรู้แบบลึกได้กลายเป็นทางออกที่เป็นไปได้สำหรับการรับรู้ท่าทาง ในช่วงไม่กี่ปีที่ผ่านมาการใช้เครือข่าย Neural Network (CNN) แบบ Convolutional ได้ประสบความสำเร็จในการรับมือกับความท้าทายในการรับรู้ท่าทาง [6] ดังนั้นบทความนี้เราเสนอการใช้ 2D Convolutional Neural Networks สำหรับการรู้จำลักษณะท่าทางภาษามือไทยที่เป็นอิสระของผู้ใช้แบบต่อเนื่อง เพื่อหาประสิทธิภาพดีที่สุดในการทดลองต่างๆ เพื่อประเมินผลของการเรียนรู้เชิงลึกโดยใช้แบบจำลองที่ได้รับการฝึกล่วงหน้าในชุดข้อมูลที่แตกต่างกัน

โดยส่วนที่เหลือของบทความนี้ ได้จัดลำดับขั้นตอนการดำเนินงานงานดังต่อไปนี้ ทำการอธิบายแนวคิดและทฤษฎีเกี่ยวกับการรู้จำลักษณะท่าทางของภาษามือไทยและงานวิจัยที่เกี่ยวข้องในส่วนที่ 2 จากนั้นทำการแสดงผลกระบวนการในการดำเนินงานและผลการประเมินประสิทธิภาพของการทดลองในส่วนที่ 3 และส่วนที่ 4 สุดท้ายทำการสรุปผลการทดลองในส่วนที่ 5

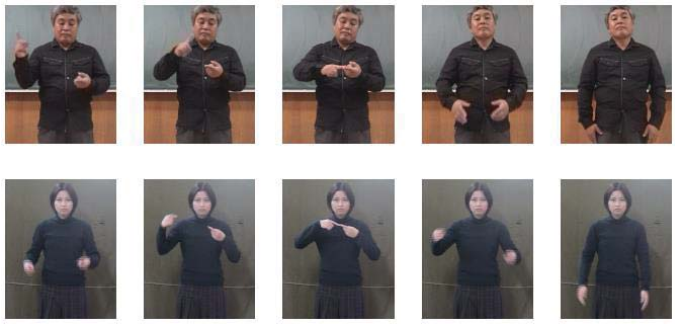
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การรู้จำลักษณะท่าทางของภาษามือไทย (Thai Sign Language Recognition)

ภาษามือไทยเป็นภาษาที่ใช้ลักษณะท่าทางของมือในการสื่อสารของผู้ที่มีความบกพร่องทางการได้ยิน [7] เพื่อการติดต่อสื่อสารกับผู้อื่นของคนไทย เพื่อใช้สื่อสารของคนหูหนวกหรือผู้ที่มีความบกพร่องทางการได้ยินที่ต่างจากการสื่อสารของบุคคลทั่วไปที่มีความเป็นปกติ ซึ่งการเรียนรู้ภาษาของคนปกติจะรับรู้ภาษาโดยอาศัยการพูดและการรับฟังในการสนทนาระหว่างกัน แต่คนหูหนวกหรือผู้ที่มีความบกพร่องทางการได้ยินนั้นไม่สามารถที่จะใช้โสตประสาทเหล่านั้นได้ ดังนั้นภาษามือไทยจึงเป็นภาษาใน

การสื่อสารที่สำคัญอย่างยิ่งของคนหูหนวกหรือผู้ที่มีความบกพร่องทางการได้ยิน โดยภาษามือจะใช้ลักษณะท่าทางของมือ การเคลื่อนไหวของมือ ตำแหน่งของท่ามือ สีหน้า และกิริยาท่าทางประกอบในการสื่อความหมายแทนการถ่ายทอดเป็นคำพูด ซึ่งเป็นคำเฉพาะของคนไทยที่ขึ้นอยู่กับองค์ประกอบด้านขนบธรรมเนียมประเพณีวัฒนธรรมและลักษณะภูมิศาสตร์ของประเทศไทย

สำหรับภาษามือนั้นมีองค์ประกอบหลักๆ ที่ประกอบด้วย 5 องค์ประกอบที่สำคัญ ได้แก่ รูปมือ (Hand Shape) ตำแหน่งของมือ (Location) การเคลื่อนไหวของมือ (Movement) ทิศทางของฝ่ามือ (Palm Orientation) และการแสดงออกทางสีหน้า (Facial Expression) แสดงได้ดังภาพที่ 1



ภาพที่ 1 การใช้ภาษามือสำหรับสื่อสาร
ที่มา: T. Matsuo, Y. Shirai and N. Shimada, 2008

การรู้จำลักษณะท่าทางของภาษามือ คือ การทำให้คอมพิวเตอร์สามารถรับรู้ความหมายของลักษณะท่าทางที่มนุษย์จะสามารถตอบโต้กับคอมพิวเตอร์ได้โดยใช้ภาษาธรรมชาติ มีนักวิจัยได้ทำการศึกษาค้นคว้าวิธีการต่างๆ เพื่อหาวิธีการหรือกระบวนการในการให้คอมพิวเตอร์สามารถปฏิสัมพันธ์กับมนุษย์ได้ อาทิเช่น การใช้ Support Vector Machine ในการรู้จำการสะกดคำของนิ้วมือข้างเดียวแบบคงที่ของภาษามืออเมริกัน การรู้จำท่าทางมือแบบไดนามิกจากรูปแบบมือหมุนวน [9] การรู้จำตัวอักษรภาษาอเมริกันโดยใช้ Microsoft Kinect [10] เป็นต้น ซึ่งในบทความนี้จะเน้นไปที่การรู้จำลักษณะท่าทางของภาษามือของผู้ที่มีความบกพร่องทางการได้ยิน ในการรู้จำโดยทั่วไปนั้นในขั้นตอนแรกของการรู้จำลักษณะท่าทางเป็นขั้นตอนการประมวลผลข้อมูลเบื้องต้น (Preprocessing) โดยจะทำการนำเข้าข้อมูลลักษณะท่าทางเข้าสู่ระบบ จากนั้นข้อมูลที่ได้ก็จะเข้าสู่ขั้นตอนของการคัดแยกคุณลักษณะของท่าทางมือที่สำคัญ (Gesture Feature) เพื่อคัดเลือกคุณลักษณะที่มีลักษณะพิเศษและเป็นเอกลักษณ์เฉพาะได้อย่างชัดเจน ว่าเป็นท่าทางที่ให้ความหมายอะไร เพื่อนำมาใช้ในการฝึกในระบบ จากนั้นก็จะนำข้อมูลที่ได้ส่งต่อไปยังขั้นตอนการแยกแยะคุณลักษณะ (Classification) เพื่อทำการรู้จำลักษณะท่าทางของมือ

2.2 CONVOLUTIONAL NEURAL NETWORKS

โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นโครงข่ายประสาทเทียมที่นิยมมากในการรู้จำรูปภาพและ โครงข่ายประสาทเทียมนี้จำลองมาจากการทำงานของสมองมนุษย์ [11] โดยโครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นโครงข่ายประสาทเทียมที่มีการรู้จำแบบต้องมีการสอน (Supervised Learning) โครงข่ายประสาทเทียมจึงจะรู้จำลักษณะเฉพาะของวัตถุต่างๆได้ โครงข่ายประสาทเทียมแบบคอนโวลูชันมีส่วนประกอบหลายชั้น ทำงานร่วมกันเป็นโครงข่ายใหญ่ ข้อมูลนำเข้าจะใช้ค่าพิกเซลโดยตรง โดยไม่ต้องผ่านกระบวนการหาฟีเจอร์ โดยโครงข่ายประสาทเทียมแบบคอนโวลูชันใช้หาฟีเจอร์แบบอัตโนมัติได้ ในแต่ละชั้น (Layer) ของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

สำหรับ 2D Convolutional Neural Networks ข้อมูลอินพุตจะถูกรวมเข้ากับ Kernels แบบ 2D โดยการคอนโวลูชันที่เกิดขึ้นโดยการคำนวณผลรวมของผลคูณระหว่างข้อมูลอินพุตและเคอร์เนล ซึ่งเคอร์เนลจะครอบคลุมเหนือข้อมูลอินพุตเพื่อ

ครอบคลุมมิติเชิงพื้นที่เต็มรูปแบบ คุณสมบัติการคอนโวเลทที่ผ่านฟังก์ชั่นการเปิดใช้งาน (activation function)เพื่อแนะนำโมเดลแบบไม่เชิงเส้น ใน 2D Convolution ค่า activation ที่ตำแหน่งเชิงพื้นที่ (x, y) ใน j ที่แมปคุณลักษณะของเลเยอร์ i แสดง โดยค่า v x, y i, j ถูกสร้างโดยใช้ สมการต่อไปนี้ [12]

$$v_{i,j}^{x,y} = \emptyset\left(b_{i,j} + \sum_{\tau=1}^{d_{l-1}} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j,\tau}^{\sigma,\rho} \times v_{i-1,\tau}^{x+\sigma,y+\rho}\right) \quad (1)$$

โดยที่ ϕ คือ activation function ส่วน $b_{i,j}$ คือพารามิเตอร์ที่มีความเอนเอียง (bias) สำหรับ j ที่แมปคุณลักษณะของเลเยอร์ i, ค่า d_{l-1}คือจำนวนของแมปคุณลักษณะในเลเยอร์ (l - 1) และความลึกของเคอร์เนล $w_{i,j}$ สำหรับ j ที่แมปคุณลักษณะของเลเยอร์ i ค่า 2γ + 1 คือความกว้างของเคอร์เนล ค่า 2δ + 1 คือความสูงของเคอร์เนลและ $w_{i,j}$ คือค่าของน้ำหนักพารามิเตอร์ j ที่แมปคุณลักษณะของเลเยอร์ i

3. ขั้นตอนวิธีการดำเนินการวิจัย

ในบทความนี้ทำการทดลองเพื่อหาประสิทธิภาพในการรู้จำโดยใช้ 2D Convolutional Neural Networks สำหรับการรู้จำลักษณะท่าทางภาษามือไทยที่เป็นอิสระของผู้ใช้แบบต่อเนื่อง โดยผู้วิจัยได้ทำการออกแบบการทดลองโดยใช้เครื่องมือ คือ หน่วยประมวลผลข้อมูล (CPU) เป็น Intel(R) Core(TM) i5-4200M 2.50 GHz ระบบปฏิบัติการ Windows 10 Home Single Language 64 bit หน่วยความจำหลัก (RAM) 8.00 GB และเลือกใช้ Anaconda3.0, และภาษา Python สำหรับการประมวลผลหาประสิทธิภาพในการรู้จำโดยใช้ 2D Convolutional Neural Networks สำหรับการรู้จำลักษณะท่าทางภาษามือไทยที่เป็นอิสระของผู้ใช้แบบต่อเนื่อง โดยมีขั้นตอนการดำเนินการเริ่มต้นจากการเตรียมชุดข้อมูล (Data Set) ภาพลักษณะท่าทางของภาษามือไทย ผู้วิจัยได้เลือกใช้ลักษณะท่าทางภาษามือไทย 3 ท่าทาง คือท่ามือ “สวัสดี”, ท่ามือ “รัก” และท่ามือ “ไม่สบาย” เพื่อใช้ในการฝึกจำนวน 3,000 ภาพ(ท่าทางละ 1,000 ภาพ) สำหรับการเรียนรู้ และอีก 300 ภาพ (ท่าทางละ 100 ภาพ) ในการทดสอบ โดยเป็นภาพที่ได้จากกล้องดิจิทัลที่มีขนาดของภาพเป็น 100*78 ผ่านการแปลงให้อยู่ในรูประดับสีเทาโดยใช้การกรองคุณลักษณะท่าทางแบบ Threshold ซึ่งภาพลักษณะท่าทางภาษามือไทยที่ใช้ทดลองเป็นชุดข้อมูลที่ผู้วิจัยได้สร้างขึ้นเองด้วยการถ่ายลักษณะท่าทางภาษามือไทยต่างๆ ที่เป็นภาพฝึก 3,000 ภาพและภาพทดสอบอีก 300 ภาพ โดยใช้บุคคลแสดงท่าทางเพียงคนเดียว แสดงได้ดังภาพที่ 2

ภาพที่ 2 ท่าทางภาษามือไทยที่ใช้ในการฝึกและทดสอบ

เมื่อได้ชุดข้อมูลภาพลักษณะท่าทางภาษามือไทยเรียบร้อยแล้ว จากนั้นนำชุดข้อมูลที่ได้มาทำการสกัดคุณลักษณะ (feature extraction) ของภาพท่าทางต่างๆ โดยการแปลงให้อยู่ในภาพแบบระดับสีเทาและใช้การกรองคุณลักษณะท่าทางแบบ Threshold ทำให้ได้ภาพคุณลักษณะที่ต้องการ แสดงได้ดังภาพที่ 3 เพื่อใช้ในการเรียนรู้ของระบบรู้จำต่อไป

ภาพที่ 3 คุณลักษณะท่าทางภาษามือไทยที่ใช้ทดลอง

จากนั้นผู้วิจัยได้ทำการออกแบบเครือข่ายประสาทแบบ Convolutional โดยมีรายละเอียดดังนี้

1. ชั้น Input data ใช้ shape [None, 78, 100, 1]

2. ชั้นที่ซ่อนแรกคือชั้นแบบ Convolutional เรียกว่า Convolution2D เลเยอร์มี แมปคุณลักษณะเป็น 32 ซึ่งมีขนาด 3 × 3 และใช้ฟังก์ชัน activation แบบ 'relu'

3. ถัดไปทำการกำหนดเลเยอร์การรวมที่ใช้เวลาสูงสุดที่เรียกว่า MaxPooling2D มีการกำหนดค่าด้วยขนาดพูล 2 × 2

4. ชั้นที่ซ่อนที่สองใช้ Convolution2D เลเยอร์ มี แมปเป็น 64 ซึ่งมีขนาด 3 × 3 และใช้ฟังก์ชัน activation แบบ 'relu'

5. ถัดไปทำการกำหนดเลเยอร์ MaxPooling2D มีการกำหนดค่าด้วยขนาดพูล 2 × 2

6. ชั้นที่ซ่อนที่สามใช้ Convolution2D เลเยอร์ มี แมปคุณลักษณะเป็น 128 ซึ่งมีขนาด 3 × 3 และใช้ฟังก์ชัน activation แบบ 'relu'

7. ถัดไปทำการกำหนดเลเยอร์ MaxPooling2D มีการกำหนดค่าด้วยขนาดพูล 2 × 2

8. ชั้นที่ซ่อนที่สี่ใช้ Convolution2D เลเยอร์ มี แมปคุณลักษณะเป็น 64 ซึ่งมีขนาด 3 × 3 และใช้ฟังก์ชัน activation แบบ 'relu'

9. ถัดไปทำการกำหนดเลเยอร์ MaxPooling2D มีการกำหนดค่าด้วยขนาดพูล 2 × 2

10. ชั้นที่ซ่อนที่ห้าใช้ Convolution2D เลเยอร์ มี แมปคุณลักษณะเป็น 32 ซึ่งมีขนาด 3 × 3 และใช้ฟังก์ชัน activation แบบ 'relu'

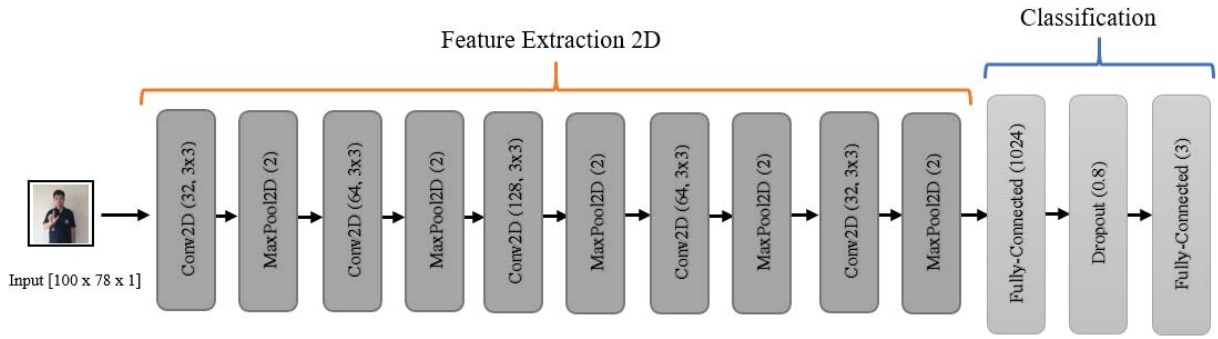
11. ถัดไปทำการกำหนดเลเยอร์ MaxPooling2D มีการกำหนดค่าด้วยขนาดพูล 2 × 2

12. ชั้นถัดไปเป็น fully_connected เลเยอร์ ชั้นเชื่อมต่ออย่างเต็มที่กับ 1024 เซลล์และใช้ฟังก์ชัน activation แบบ 'relu'

13. ชั้นถัดไปคือชั้นการจัดระเบียบโดยใช้ dropout เรียกว่า Dropout มีการกำหนดค่าให้สุ่มเลือก 80% ของเซลล์ประสาทในเลเยอร์

14. สุดท้ายเลเยอร์เอาท์พุทเป็น fully_connected เลเยอร์ ชั้นเชื่อมต่ออย่างเต็มที่กับ 3 คลาสและใช้ฟังก์ชัน activation แบบ 'softmax' เพื่อสร้างการคาดการณ์ที่น่าจะเป็นไปได้สำหรับแต่ละคลาส

โดยแบบจำลองได้รับการฝึกโดยใช้การสูญเสียลอการิทึมและอัลกอริธึมการไล่ระดับสีแบบ ADAM ใช้ learning_rate=0.001 ทำการประเมินแบบจำลองด้วย perceptron หลายชั้น เป็น CNN ที่มีแบบจำลองมีขนาดพอดีกับ epochs 50 รอบ ที่มีการอัปเดต batch ทุกๆ 100 ภาพ สามารถแสดงรายละเอียดได้ดังภาพที่ 4



ภาพที่ 4 สถาปัตยกรรมโมเดล 2DCNN ที่ใช้ในการทดลอง

จากนั้นขั้นตอนสุดท้ายเป็นการทดลองโดยทำการทดสอบประสิทธิภาพความถูกต้องในการรู้จำลักษณะท่าทางภาษามือไทย เพื่อทดสอบโมเดลที่ได้ทำการออกแบบ โดยการใช้ชุดข้อมูลลักษณะท่าทางของภาษามือไทย เพื่อทำการแยกแยะคุณลักษณะท่าทางภาษามือไทยที่มีความหมายที่ถูกต้อง จากนั้นทำการวิเคราะห์ผลและคำนวณประสิทธิภาพในการทำงานของโมเดลและสรุปผลการทดลอง

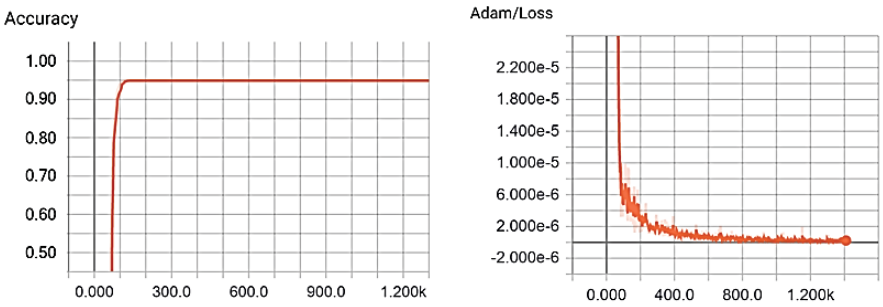
การวัดประสิทธิภาพเพื่อเปรียบเทียบความแม่นยำในงานวิจัยนี้ใช้การวัดค่าความแม่นยำ (Accuracy) เป็นค่าที่ได้จากวิธีการทดลองเพื่อแยกแยะคุณลักษณะท่าทางของภาษามือไทย (Classification) โดยคิดเป็นค่าร้อยละ (%) ใช้สูตรการคำนวณดังนี้

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

โดยที่ TP คือ ค่าที่แยกแยะคุณลักษณะได้ถูกต้องเชิงบวก
TN คือ ค่าที่แยกแยะคุณลักษณะได้ถูกต้องเชิงลบ
FP คือ ค่าที่แยกแยะคุณลักษณะได้ผิดพลาดเชิงบวก
FN คือ ค่าที่แยกแยะคุณลักษณะได้ผิดพลาดเชิงลบ

4. ผลการทดลอง

ในส่วนนี้จะแสดงผลจากการรู้จำลักษณะท่าทางของภาษามือโดยใช้การเรียนรู้เชิงลึก ซึ่งใช้อัลกอริทึม 2D Convolutional Neural Networks ร่วมกับการใช้งานกล้องดิจิทัล ที่ให้ภาพ RGB และทดลองกับท่าทางต่างๆ จำนวน 3 ท่าทาง คือ ท่าทาง “สวัสดี” ท่าทาง “รัก” และท่าทาง “ไม่สบาย” โดยการทดลองแสดงประสิทธิภาพของอัลกอริทึมในการแยกแยะเพื่อการรู้จำลักษณะท่าทางของภาษามือไทย โดยสามารถผลการออกแบบโมเดล 2D Convolutional Neural Networks โดยใช้ TensorBoard ที่ผู้วิจัยได้ออกแบบเพื่อการเรียนรู้ ซึ่งผลจากการเรียนรู้ ได้ค่าความแม่นยำ (Accuracy) เป็น 0.93 และได้ค่าการสูญเสีย (Loss) เป็น 0. 27 ใช้เวลาในการเรียนรู้ทั้งหมด 1 ชั่วโมง 15 นาที 46 วินาที สามารถแสดงผลการทดลอง ได้ดังภาพที่ 5 ซึ่งเป็นผลลัพธ์จากการประมวลที่ได้จาก TensorBoard



ภาพที่ 5 ผลค่าความแม่นยำและการสูญเสีย

จากภาพที่ 5 แสดงให้เห็นว่าผลของข้อมูลเกี่ยวกับความถูกต้องตามการฝึกที่สามารถมองเห็นสะท้อนข้อมูลการฝึกในการปรับปรุงที่สำคัญในประสิทธิภาพการรับรู้ในการทำซ้ำทั้งหมดตามที่คาดไว้สำหรับการรับรู้ท่าทาง ซึ่งเมื่อมีการเรียนรู้เพิ่มขึ้นประสิทธิภาพความถูกต้องในการรู้จำก็จะมีเพิ่มมากขึ้น โดยเลเยอร์มีผลอย่างมากต่อการเพิ่มประสิทธิภาพน้ำหนักในเครือข่ายแบบลึกและการปรับปรุงแผนที่ความลึกที่ดีขึ้นจากลำดับความลึกและแยกแยะโดยใช้ 2D Convolutional Neural Networks เมื่อเมื่อทำการฝึกวิธีการเรียนรู้อย่างสมบูรณ์ และจะดียิ่งขึ้นเมื่อมีข้อมูลการฝึกมากขึ้น โดยชั้น fully connected ในการจำแนกสามารถแยกแยะได้อย่างมีประสิทธิภาพแม้ว่าลักษณะท่าทางของชุดข้อมูลที่ผ่านการฝึกจะมีความแตกต่างกันอย่างมากก็ตาม การเลือกใช้ weight อย่างรอบคอบก็จะสามารถปรับปรุงประสิทธิภาพได้เป็นอย่างมากด้วยเช่นกัน โดยแสดงค่าความแม่นยำ (Accuracy) และค่าการสูญเสีย (Loss) ในการทดสอบสร้างแบบจำลองจะเห็นว่าการเรียนรู้ในแต่ละรอบ (Round) มีค่าความแม่นยำที่แตกต่างกัน โดยผลการทดลองสร้างแบบจำลองได้ค่าความแม่นยำสูงสุดคือ 93.00% แสดงถึงประสิทธิภาพในการจำแนกลักษณะท่าทางของภาษามือไทย โดยใช้วิธี Deep Learning ด้วยอัลกอริทึม 2DCNN ซึ่งเป็นการประยุกต์ใช้ลักษณะท่าทางของภาษาไทยและนำเทคนิควิธีการ Deep Learning โดยใช้ 2DCNN มาเรียนรู้ภาษามือไทย เพื่อพัฒนาต่อยอดไปเป็นนวัตกรรมสำหรับการแปลลักษณะท่าทางของภาษามือออกมาเป็นคำพูดหรือตัวอักษรแบบเรียลไทม์ต่อไป

5. สรุปผลการทดลองและการอภิปรายผล

ในการศึกษาครั้งนี้ ได้เสนอการใช้ 2D Convolutional Neural Networks (2D CNNs) กับปัญหาการรับรู้ท่าทางของภาษามือไทย โดยทำการหาประสิทธิภาพความถูกต้องในการรู้จำลักษณะท่าทางของภาษามือไทย โดยสิ่งที่เสนอสำหรับงานนี้ คือ การออกแบบโมเดล 2D CNN และการสร้างชุดข้อมูลลักษณะท่าทางของภาษามือไทย ซึ่งใช้ 2D convolution และ pooling layers ช่วยในการเรียนรู้ความแตกต่างของข้อมูล ในการทดลองได้แสดงให้เห็นว่าโมเดลที่ได้ออกแบบสามารถช่วยเพิ่มประสิทธิภาพการรับรู้ท่าทางอย่างมาก ได้ค่าความแม่นยำ (Accuracy) เป็น 0.93 และได้ค่าการสูญเสีย (Loss) เป็น 0. 27 ที่ได้รับจาก 2D CNNs ใช้เวลาในการเรียนรู้เพื่อการเรียนรู้ทั้งหมด 1 ชั่วโมง 15 นาที 46 วินาที และจะดีขึ้นเมื่อมีข้อมูลการฝึกมากขึ้น สำหรับชั้น fully connected ในการจำแนกถึงแม้ว่าลักษณะท่าทางของชุดข้อมูลที่ผ่านการฝึกจะแตกต่างกันก็ตาม ถ้า layers ต้องถูกเตรียมใช้งานตั้งแต่เริ่มต้นการเลือกใช้ weight อย่างรอบคอบก็จะสามารถปรับปรุงประสิทธิภาพได้เป็นอย่างมาก และการใช้วิธี Deep Learning ด้วยอัลกอริทึม 2DCNN ในการประมวลผลภาพเพื่อสร้างแบบจำลอง โดยผลการสร้างแบบจำลองแสดงค่าความแม่นยำสูงสุดคือ 93.00% แสดงถึงประสิทธิภาพในการสร้างแบบจำลองเพื่อจำแนกคุณลักษณะท่าทางของภาษามือไทยด้วยวิธี Deep Learning สอดคล้องกับงานวิจัย [6] ที่แสดงให้เห็นว่าการประยุกต์ใช้วิธี Deep Learning ด้วยอัลกอริทึม CNN มีประสิทธิภาพในการจำแนกข้อมูลประเภทที่ไม่ได้มีโครงสร้างเป็นรูปแบบเฉพาะตัว (Unstructured Data) อย่างเช่น รูปภาพ

(Image) ซึ่งสามารถสกัดคุณลักษณะเด่นจากรูปภาพได้อย่างมีประสิทธิภาพ สำหรับงานในอนาคตควรมีการเปรียบเทียบเทคนิคเพิ่มเติม เช่น เปรียบเทียบกับอัลกอริทึมอื่นๆ พัฒนาท่าทางเพิ่มขึ้น ตลอดจนพัฒนาในรูปแบบ3D CNN และทำการใช้วิธีการคัดแยกคุณลักษณะที่มีประสิทธิภาพเพิ่มขึ้น เพื่อทำการทดสอบประสิทธิภาพให้ครอบคลุมในทุกๆ ด้าน และนำไปประยุกต์ใช้งานจริงแบบเรียลไทม์

เอกสารอ้างอิง (References)

[1] A. Kiani Sarkaleh, F. Poorahangaryan, B. Zanj and A. Karami, **A Neural Network based system for Persian sign language recognition**, 2009 IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, 2009, pp. 145-149.

[2] M. M. Islam, S. Siddiqua and J. Afnan, **Real time Hand Gesture Recognition using different algorithms based on American Sign Language**, 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Dhaka, 2017, pp. 1-6.

[3] J. E. López-Noriega, M. I. Fernández-Valladares and V. Uc-Cetina, **Glove-based sign language recognition solution to assist communication for deaf users**, 2014 11th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Campeche, 2014, pp. 1-6.

[4] C. Savur and F. Sahin, **Real-Time American Sign Language Recognition System Using Surface EMG Signal**, 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, 2015, pp. 497-502.

[5] M. Seymour and M. Tsoeu, **A mobile application for South African Sign Language (SASL) recognition**, AFRICON 2015, Addis Ababa, 2015, pp. 1-5.

[6] L. Pigou, S. Dieleman, P. Kindermans and B. Schrauwen, **Sign Language Recognition Using Convolutional Neural Networks**, In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham.

[7] H. Wang, X. Chai, Y. Zhou and X. Chen, **Fast sign language recognition benefited from low rank approximation**, 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, 2015, pp. 1-6.

[8] T. Matsuo, Y. Shirai and N. Shimada, **Automatic generation of HMM topology for sign language recognition**, 2008 19th International Conference on Pattern Recognition, Tampa, FL, 2008, pp. 1-4.

[9] H. G. Doan, H. Vu and T. H. Tran, **Dynamic hand gesture recognition from cyclical hand pattern**, 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, 2017, pp. 97-100.

[10] Cao Dong, M. C. Leu and Z. Yin, **American Sign Language alphabet recognition using Microsoft Kinect**, 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015, pp. 44-52.

[11] C. Teerapat and P. Suebskul, **BUILDING DETECTION FROM STREET-SIDE IMAGES IN RURAL AREA**, Thesis, Chulalongkorn University, 2014.

[12] S. K. Roy, G. Krishna, S. R. Dubey, B. B. Chaudhuri, **HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification**, Computer Vision and Pattern Recognition, Published in IEEE Geoscience and Remote Sensing Letters, 2019