TNI Journal of Engineering and Technology
Vol.7 No.1 January - June 2019

การปรับปรุงประสิทธิภาพในการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันโดยใช้เทคนิคการเพิ่มภาพ

Improving the Performance of an Image Classification with Convolutional Neural Network Model by Using Image Augmentations Technique

พิมพา ชีวาประกอบกิจ
Pimpa Cheewaprakobkit

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยนานาชาติเอเชีย-แปซิฟิก
Faculty of Information Technology, Asia-Pacific International University
pimpa@apiu.edu

รับต้นฉบับ:11 พฤศจิกายน 2561; รับบทความฉบับแก้ไข: 14 พฤษภาคม 2562 ; ตอบรับบทความ: 16 พฤษภาคม 2562
เผยแพร่ออนไลน์: 28 มิถุนายน 2562

บทคัดย่อ

การจำแนกภาพเป็นหนึ่งในความท้าทายสำหรับมนุษย์และคอมพิวเตอร์ เพราะเป็นงานที่ต้องอาศัยการวิเคราะห์ นักวิจัยพยายามที่จะแก้ปัญหานี้ในหลายๆ ด้าน โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional neural network: CNN) ได้ถูกนำมาใช้กันอย่างแพร่หลายในการรับรู้ภาพการจำแนกภาพและการตรวจจับวัตถุ ในกระบวนการของ CNN โมเดล จะมีการสอนให้เครื่องเรียนรู้และทดสอบ บทความนี้มีวัตถุประสงค์เพื่อปรับปรุงประสิทธิภาพในการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันโดยใช้เทคนิคการเพิ่มภาพ และเปรียบเทียบค่าความแม่นยำของเทคนิคการสร้างภาพเทียมแบบต่างๆ ข้อมูลที่ใช้ในการทดลองนี้รวบรวมจาก Canada Institute for Advanced Research (CIFAR-10) ซึ่งมีทั้งหมด 60,000 ภาพ แต่ละภาพมีขนาด 32x32 พิกเซล แบ่งเป็น 10 หมวด ในรอบแรกของการฝึกสอนให้คอมพิวเตอร์เรียนรู้ จะใช้วิธี 10- fold Cross Validation ในการแบ่งข้อมูล 50,000 ภาพ สำหรับการฝึกสอนและ 10,000 ภาพเพื่อทดสอบ ในรอบที่ 2 ทำการสุ่มภาพจากทุกหมวดมาเพื่อสร้างภาพเทียม โดยจะสุ่มใช้เทคนิคการปรับสีของภาพ การหมุนภาพ การย่อ ขยายภาพ หรือการกลับด้านของภาพ อย่างใดอย่างหนึ่งในแต่ละภาพ จนได้ภาพใหม่จำนวน 10,000 ภาพ รวมกับของเดิม 50,000 ภาพ รวมทั้งสิน 60,000 ภาพ ทำการฝึกสอนใหม่ 300 รอบ ผลการทดลองพบว่าการใช้เทคนิคการเพิ่มภาพ ด้วยการสร้างภาพเทียมจะช่วยให้ประสิทธิภาพในการจำแนกภาพแม่นยำสูงขึ้นจาก 84.79% เป็น 87.57%

คำสำคัญ: โครงข่ายประสาทเทียมแบบคอนโวลูชัน, การจำแนกภาพ, เทคนิคการเพิ่มภาพ

Abstract

Image classification is one of the challenges for both humans and computers because it is a critical task. Researchers try to solve this task in many ways. In neural networks, Convolutional Neural Network (CNN) has been widely used for images recognition, images classifications, and objects detections. The process of CNN model includes training and testing. This paper aims to improve the performance of an image classification with Convolutional Neural Network model by using image augmentation technique and compare the accuracy of each technique. The dataset that we used in this paper was collected from Canadian Institute for Advanced Research (CIFAR-10) dataset which consists of a total of 60,000 images. The color images are of size 32x32 in 10 different classes. In the first stage, the 10- fold cross validation method is used to divide 50,000 images for training 300 rounds and 10,000 images for testing. In the second stage, in order to create artificial images, 10,000 images were randomly-selected from all categories and applied the image augmentation technique such as color adjustment, rotation, zoom, or flip image to the image. This image augmentations technique were randomly selected only one technique and applied to one image. The training images were 60,000 images in total; 50,000 images from the first stage plus 10,000 new images. Then the computer trained the images for 300 rounds. The evaluation shows that after using image augmentations technique by creating artificial images, the accuracy of Convolutional Neural Network model performs better from 84.79% up to 87.57%

Keywords: Convolutional Neural Network, Image Classification, image augmentations technique

1) บทนำ

การจำแนกภาพเป็นหัวข้อสำคัญ ที่ได้รับความสนใจอย่างมากในช่วงทศวรรษที่ผ่านมา เนื่องจากการจำแนกภาพที่มีลักษณะซับซ้อนส่วนใหญ่ต้องอาศัยคนในการวิเคราะห์เพื่อจำแนกภาพ การพัฒนาโมเดลในการจำแนกภาพนี้มีวัตถุประสงค์เพื่อจัดประเภทของภาพตามเนื้อหา นักวิจัยส่วนใหญ่อาศัยการกำหนดคุณลักษณะของภาพตามหมวดต่างๆ ด้วยตนเองก่อน [1] หลังจากนั้นจึงให้คอมพิวเตอร์เป็นตัวช่วยในการแยกประเภทของภาพตามลักษณะที่กำหนดไว้ อย่างไรก็ตามเมื่อภาพมีจำนวนมากก็จะเป็นปัญหาที่ยากเกินไปที่จะค้นหาคุณสมบัติจากภาพเหล่านั้น นี่คือหนึ่งในเหตุผลที่โมเดลโครงข่ายประสาทเทียมเข้ามามีบทบาทในไม่กี่ปีที่ผ่านมา

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) ได้ถูกนำมาใช้ในการแก้ปัญหาจัดหมวดหมู่ (Classification) และจำแนกภาพที่มีประสิทธิภาพใกล้เคียงกับมนุษย์ ลักษณะเด่นของ CNN คือในขั้นตอนของการฝึกสอน (train) จะทำการคัดเลือกคุณลักษณะเอง และทำซ้ำไปเรื่อยๆ จนกว่าจะได้คุณลักษณะที่มีความสัมพันธ์กับผลลัพธ์มากที่สุด ซึ่งแตกต่างจาก Machine learning แบบเดิม ที่มนุษย์เป็นคนกำหนดคุณลักษณะ (features) ให้กับเครื่องเพื่อฝึกสอน ซึ่งมีจุดอ่อนคือ ถ้าคุณลักษณะของภาพมีความซับซ้อน จะยากต่อการกำหนดคุณลักษณะจะทำให้คอมพิวเตอร์เรียนรู้ได้ไม่ดี และผลลัพธ์ที่ได้ไม่มีประสิทธิภาพเท่าที่ควร ทั้งนี้ประสิทธิภาพในการทำนายของ CNN ยังขึ้นอยู่กับจำนวนและลักษณะของภาพที่ใช้ในการฝึกสอนให้คอมพิวเตอร์เรียนรู้ ซึ่งจำเป็นต้องใช้ภาพจำนวนมากในการฝึกสอน ดังนั้นการกำหนดจำนวนภาพและลักษณะของภาพจึงเป็นสิ่งสำคัญ แต่ CNN มีข้อเสียคือใช้ทรัพยากรและเวลาในการประมวลผลมาก

นาย Tanner และ Wong [2] ได้พัฒนาแบบจำลองในการฝึกสอนการเรียนรู้เชิงลึกให้กับคอมพิวเตอร์ด้วยการเพิ่มรูปภาพ เพื่อให้ข้อมูลมีเพียงพอและเกิดประสิทธิภาพในการจำแนกภาพ โดยใช้เทคนิคที่เรียกว่า PCA (principal component analysis) เพื่อวิเคราะห์องค์ประกอบหลักและจับลักษณะสำคัญของภาพ PCA นี้จะทำการสร้างตัวแปรใหม่ขึ้นมาที่เรียกว่า Component เพื่อลดจำนวนตัวแปรลง แต่ยังคงสามารถอธิบายลักษณะและข้อมูลให้ได้มากที่สุด PCA เหมาะกับงานพวก image processing ที่ข้อมูลมีจำนวนตัวแปรมากๆ ต่อมา Hinton และคณะ [3] ได้นำเสนออัลกอริทึมการเรียนรู้อย่างที่มีประสิทธิภาพ คือ Contrastive Divergence (CD) ซึ่งแต่ละเลเยอร์ได้รับการฝึกสอนเป็นชั้น ๆ จากการเรียนรู้อย่างลึกซึ้ง จึงสามารถเป็นตัวแทนของลักษณะลำดับชั้นได้ โดยจะใช้หลายชั้นและค่าน้ำหนักที่สอดคล้องกัน อย่างไรก็ตามเมื่อมิติการป้อนข้อมูลมีขนาดใหญ่เกินไปจึงต้องใช้เวลานานในการฝึกสอนให้คอมพิวเตอร์เรียนรู้

ในการนำเทคนิคการจำแนกภาพไปประยุกต์ใช้กับงานจะเห็นได้มากมาย ตัวอย่างเช่น Zafrulla, Brashear, Starner, Hamilton และ Presti [4] ได้ศึกษาความแตกต่างของภาษามือ ที่ใช้ในสหรัฐอเมริกาและที่ใช้ในประเทศอังกฤษ และอีกท่านคือนักวิศวกรชาวญี่ปุ่น ชื่อ Makoto Koike [5] ต้องการคัดแยกแตงกวาจากสวนของเขาออกเป็น 9 เกรด ถ้าหากใช้คนเรียนรู้ความแตกต่างของแตงกวาแต่ละเกรด เช่น สี ความมันวาวของผิว จำนวนหนามและขนาดของหนาม ความโค้งงอของลูกแตงกวา ซึ่งต้องใช้เวลาเป็นเดือนๆ ในการเรียนรู้สำหรับการแยกแตงกวาในแต่ละเกรด ดังนั้นเขาจึง สร้างแบบจำลอง Machine Learning เพื่อให้เครื่องคอมพิวเตอร์เรียนรู้คุณสมบัติเชิงภาพของแตงกวา โดยเขาถ่ายภาพแตงกวาเกรดต่างๆ จำนวน 7,000 ภาพ ขนาด 80x80 พิกเซลโดยใช้การประมวลผลแบบ Deep Learning เนื่องจากข้อมูลมีความซับซ้อน และผลการทดสอบในการแยกแยะมีความถูกต้อง 95% แต่เมื่อลองใช้จริงกลับลดลงเหลือ 70% ซึ่งเขาตั้งสมมติฐานว่าสาเหตุน่าจะเกิดจากขนาดภาพที่เล็กเกินไป ทำให้รายละเอียดของหนาม และลักษณะเชิงภาพอื่นๆ หายไป หรือจำนวนภาพที่น้อยเกินไป ในงานวิจัยของ Hussain และคณะ [6] ได้ทำการเปรียบเทียบกลยุทธ์ต่างๆ ที่ใช้ในการเพิ่มรูปภาพและประสิทธิภาพในการจำแนกพบว่า การลดสิ่งรบกวนในภาพ (Gaussian filters) มีประสิทธิภาพในการจำแนกภาพแม่นยำ 88% ซึ่งสูงกว่าการพลิกกลับด้านของภาพ (flips) ที่มีความแม่นยำเพียง 84% แต่ในทางตรงกันข้ามการจำแนกข้อมูลที่เป็นเสียง การเพิ่มเสียงรบกวนให้มากขึ้น จะทำให้ประสิทธิภาพในการจำแนกข้อมูลประเภทเสียงจะลดลง มีความแม่นยำในการจำแนกเพียง 66% เท่านั้น

การวิจัยนี้มีวัตถุประสงค์เพื่อปรับปรุงประสิทธิภาพในการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันโดยใช้เทคนิคการเพิ่มภาพ โดยในการทดลองจะทำการเปรียบเทียบค่าความแม่นยำในการจำแนกภาพที่ได้จากการใช้เทคนิคการเพิ่มภาพกับความแม่นยำในการจำแนกภาพแบบปกติที่ไม่ได้ใช้เทคนิคการเพิ่มภาพ และเปรียบเทียบค่าความแม่นยำของเทคนิคการสร้างภาพเทียมแบบต่างๆ เช่น การปรับสีของภาพ การหมุนภาพ การย่อ ขยายภาพ หรือการกลับด้านของภาพ ซึ่งจะเป็นประโยชน์กับงานที่มีจำนวนรูปภาพอย่างจำกัด หรือการหารูปภาพเพิ่มทำได้ยาก และต้องการลดค่าใช้จ่ายและเวลาจากการเก็บภาพตัวอย่างจำนวนมาก

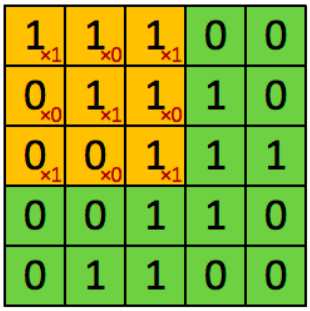
2) ทฤษฎี

2.1) โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

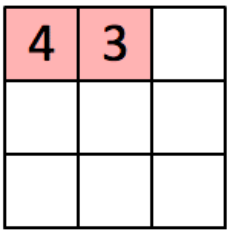
โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN) เป็นรูปแบบ architecture หนึ่งของ feed-forward neural networks จัดเป็นการเรียนรู้เชิงลึก (Deep learning) เช่นกัน [7] โดยจะจำลองการมองเห็นของมนุษย์ในพื้นที่ย่อยๆ จากการแยกแยะคุณลักษณะเชิงภาพ เช่น สี ลายเส้น และอื่นๆ จากนั้นนำมาผสานกันเพื่อทำนายว่าภาพนั้นคือภาพอะไร ในการทำงานของ Convolutional Neural Network (CNN) จะมี 4 กระบวนการคือ

1. คอนโวลูชัน (Convolution) จุดประสงค์ของการทำคอนโวลูชันเพื่อต้องการระบุคุณลักษณะที่สำคัญและเกี่ยวข้องกับภาพ โดยขั้นตอนนี้จะทำการสร้าง Sliding window (Filter) มาสแกนรูปภาพเพื่อแยกองค์ประกอบต่างๆ เช่น รูปทรงของเส้นขอบ สี โดยปกติภาพจะมีสีหลักๆ 3 สี คือ สีแดง สีน้ำเงิน และสีเขียว แบ่งเป็น 3 แชนแนล (Channel) ซึ่งแต่ละพิกเซลสามารถแทนค่าด้วยตัวเลขเพื่อบอกความเข้มของสี ตั้งแต่ 0-255 ในการทำภาพขาวดำจะใช้เพียง 1 แชนแนล สีดำแทนด้วยเลข 1 และสีขาวแทนด้วยเลข 0 จากนั้นจะใช้ filter เป็นตัวกรอง จากรูปที่ 1 (เมทริกซ์สีเหลือง ขนาด 3x3) ทำการ convolution กับภาพขาวดำ เพื่อเก็บค่าไว้ในเมทริกซ์ชุดใหม่ ที่เรียกว่า

คอนโวลฟีเจอร์ (Convolved Feature) หรือ ฟีเจอร์แมพ (feature map) ดังรูป ที่ 2



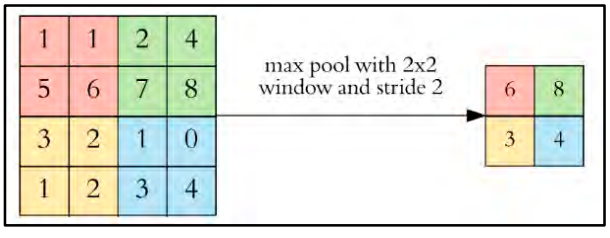
รูปที่ 1: การทำงานของตัวกรองค่า (Filter) บนภาพขาวดำ



รูปที่ 2: เมทริกซ์ชุดใหม่หรือฟีเจอร์แมพ (feature map)

2. Rectified Linear Unit หรือ ReLU เป็น activation function จุดประสงค์ของการทำ ReLU เพื่อแปลงเมทริกซ์ฟีเจอร์แมพ จากข้อ 1 ให้อยู่ในรูปของ nonlinear คือไม่มีลักษณะเป็นเชิงเส้น และช่วยแก้ปัญหาการไล่ระดับสีที่หายไป (gradient vanishing) โดยจะทำการแทนที่ค่าในพิกเซลที่เป็นลบด้วย 0 หลังจากทำ ReLU แล้วค่าที่ได้จะเป็นค่าบวกเท่านั้น ส่วนใหญ่นิยมใช้ Softmax function เพื่อความง่ายในการคำนวณและประสิทธิภาพของผลลัพธ์

3. การพูลลิ่ง (Pooling) เป็นการลดมิติของฟีเจอร์แมพให้มีขนาดเล็กลงแต่ยังคงรักษารายละเอียด ข้อมูลสำคัญของภาพไว้ การพูลลิ่งมีประโยชน์ในเรื่องความไวในการคำนวณ และแก้ปัญหา overfitting การพูลลิ่งสามารถจำแนกเป็นประเภทต่างๆ ได้เช่น พูลลิ่งด้วยค่าสูงสุด (Max Pooling) ค่าเฉลี่ย (Mean Pooling) ดังแสดงตัวอย่างในรูปที่ 3 การพูลลิ่งด้วยค่าสูงสุด (Max Pooling) [8]



รูปที่ 3: การพูลลิ่งด้วยค่าสูงสุด(Max Pooling)

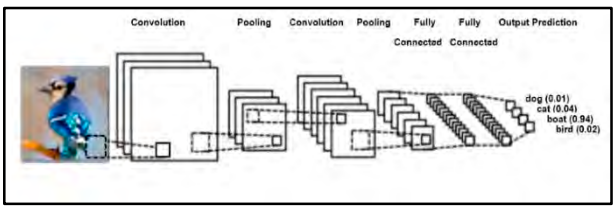
4. การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer) จากกระบวนการทั้ง 3 ขั้นตอน ที่กล่าวไว้ข้างต้น คือ

1) คอนโวลูชัน (Convolution) ทำการระบุคุณลักษณะที่สำคัญๆ และเกี่ยวข้องกับภาพ

2) Rectified Linear Unit (ReLU) ทำการแปลงเมทริกซ์ชุดใหม่หรือฟีเจอร์แมพ

3) การพูลลิ่ง (Pooling) ทำการลดมิติของฟีเจอร์แมพให้มีขนาดเล็กลงแต่ยังคงรักษารายละเอียด ข้อมูลสำคัญของภาพไว้

ในขั้นตอนนี้จะเป็นการทำซ้ำตั้งแต่ขั้นตอนของคอนโวลูชัน (Convolution) ถึงการพูลลิ่ง (Pooling) จนกว่าจะเกิดการเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์ (Fully Connected Layer) เราสามารถอธิบายภาพรวมของขั้นตอนที่ 1 คอนโวลูชัน(Convolution) และขั้นตอนที่ 2 Rectified Linear Unit (ReLU) ไว้ด้วยกันได้ ดังนั้นในภาพที่ 4 จึงเห็นเพียงขั้นตอนของคอนโวลูชัน(Convolution) Pooling และ Fully Connected จุดประสงค์ของขั้นตอนนี้เพื่อจะนำลักษณะเด่นๆ ที่สำคัญของรูปที่ได้จากกระบวนการก่อนหน้ามาทำการสร้างเป็น Neural Network สำหรับการเรียนรู้และทำนายประเภทของรูปภาพ โดยจะทำการคัดกรองรูปที่รับเข้ามาและจัดให้อยู่ในรูปของคลาส (Classes) และผลลัพธ์ที่ได้จะแสดงค่าความมั่นใจ (Confident) ในทำนายประเภทของรูปภาพ ดังตัวอย่างในรูปที่ 4



รูปที่ 4: การเชื่อมต่อกันของแต่ละชั้นอย่างสมบูรณ์โดยแสดงคลาสที่ทำนายและค่าความมั่นใจ

2.2) การเพิ่มภาพ (Image Augmentation)

การสร้างโมเดลเพื่อให้คอมพิวเตอร์ช่วยในการวิเคราะห์ ทำนาย หรือจำแนกข้อมูลนั้นจำเป็นต้องอาศัยฐานข้อมูลเพื่อฝึกสอน (Training) ให้คอมพิวเตอร์เรียนรู้ การมีข้อมูลที่มากเพียงพอจะทำให้คอมพิวเตอร์มีประสิทธิภาพในการวิเคราะห์ได้ถูกต้องแม่นยำ แต่เนื่องจากบางครั้งข้อมูลที่มีอยู่นั้นมีจำนวนจำกัดดังนั้นเทคนิคการเพิ่มภาพ (image augmentation) จึงเข้ามาช่วยในการแก้ปัญหาดังกล่าว [9] ได้กล่าวถึงวัตถุประสงค์ของเทคนิคการเพิ่มภาพไว้ว่า เป็นการสร้างภาพแบบอัตโนมัติ หรือเรียกได้ว่าเป็นการสร้างข้อมูลเทียมเพื่อขยายชุดข้อมูลเพื่อนำไปใช้ในการเรียนรู้ของเครื่องคอมพิวเตอร์ โดยเฉพาะการเรียนรู้แบบเชิงลึก เช่น โครงข่ายประสาทเทียม (Neural Network) [6] ได้อธิบายถึงการเพิ่มภาพ สามารถทำได้หลายวิธี เช่น การบิดหมุนภาพ พลิกกลับด้านในมุมต่างๆ ยืดภาพ หดภาพ เปลี่ยนสี การตัดบางส่วนของภาพออก ย่อ ขยาย และเทคนิคที่น่าสนใจอีกอย่างหนึ่งก็คือ การลดความชัดเจนของภาพ ทำให้ภาพเพี้ยนไปจากความเป็นจริง [10]

3) วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการทดลองนี้รวบรวมจาก Canada Institute for Advanced Research (CIFAR-10) ซึ่งมีทั้งหมด 60,000 ภาพ แต่ละภาพมีขนาด 32x32 พิกเซล แบ่งเป็น 10 หมวด ได้แก่ เครื่องบิน รถยนต์ นก แมว กวาง สุนัข กบ ม้า เรือ และรถบรรทุก

รูปที่ 5: ตัวอย่างภาพในแต่ละหมวด

เครื่องมือที่ใช้ในการวิจัยนี้คือ Keras และ Tensorflow โดยจะทำการสร้างโมเดลโครงข่ายประสาทแบบคอนโวลูชัน และทำการวิเคราะห์จำแนกรูปภาพว่าอยู่ในหมวดใด

การพัฒนาแอปพลิเคชันด้วย TensorFlow ซึ่งเป็นไลบรารีสำหรับใช้พัฒนา Machine learning เขียนด้วยภาษาไพทอน (Python) รองรับการจำแนกข้อมูลด้วยโครงข่ายประสาทแบบคอนโวลูชัน ส่วน Keras เป็น Deep-learning library ในภาษาไพทอน (Python) ที่ใช้ สำหรับบิดเบือนรูปภาพให้ต่างไปจากภาพเดิม โดยสามารถรันบน CPU และเพิ่มประสิทธิภาพโดยการรันบน GPU ได้ การทำงานของ Keras ไม่สามารถทำงานโดดๆ ได้ ทำได้เพียงแค่เป็น Front-end โดยมี Tensorflow ทำหน้าที่เป็น Back-end

สำหรับการติดตั้ง Keras และ Tensorflow ก่อนอื่นจำเป็นจะต้องติดตั้งเครื่องมือเหล่านี้ก่อน [11]

1) Visual Studio 2017 Community Edition ใช้สำหรับ Compile CUDA library

2) Anaconda (64-bit) with Python 3.6 แนะนำใช้ Python เวอร์ชัน 3 เพราะใหม่และเสถียรกว่าเวอร์ชัน 2

3) CUDA V8 (64-bit) ใช้สำหรับ คำนวนบน GPU เช่น การคูณ Matrix parallel computing

4) cuDNN v6 สำหรับ CUDA 8.0 ใช้ประมวลผล Deep Neural Net (Convolution Neural Networks)

5) Keras เป็น Deep-learning library ในภาษาไพทอน (Python)

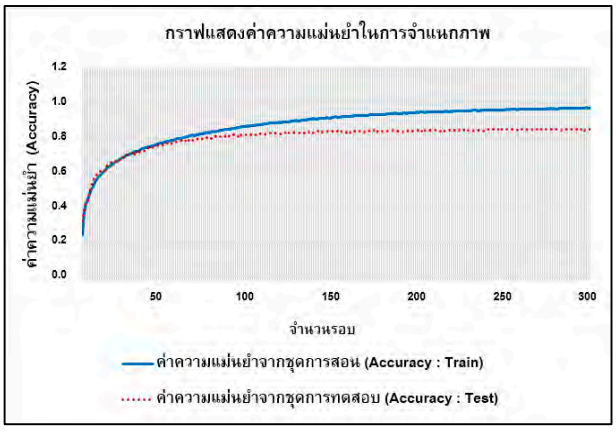
6) Tensorflow-gpu (1.4) เป็น Back-end สำหรับทำ Deep Neural Network

จากข้อมูลภาพที่มีทั้งหมด 60,000 ภาพ จะใช้ 50,000 ภาพในการฝึกสอนให้คอมพิวเตอร์เรียนรู้ และอีก 10,000 ภาพเพื่อใช้ทดสอบความถูกต้องของโมเดล โดยใช้เทคนิค 10-fold Cross Validation ในการทดลองได้กำหนดให้คอมพิวเตอร์สุ่มภาพมาจำนวน 50,000 ภาพ จาก 10 หมวดหมู่ เพื่อฝึกสอนให้คอมพิวเตอร์เรียนรู้ เป็นจำนวน 300 รอบ แล้วจึงนำภาพที่เหลืออีก 10,000 ภาพมาทดสอบความถูกต้องในการจำแนก ในขั้นตอนถัดไปจะใช้เทคนิคการเพิ่มภาพ ซึ่งจะทำการสุ่มภาพจากทุกหมวดมา 10,000 ภาพ และทำการสุ่มใช้เทคนิคการปรับสีของภาพ การหมุนภาพ การย่อ ขยายภาพ หรือการกลับด้านของภาพอย่างใดอย่างหนึ่ง ในแต่ละภาพ จนได้ภาพใหม่จำนวน 10,000 ภาพ แล้วจึงฝึกให้คอมพิวเตอร์เรียนรู้ใหม่ 300 รอบ แล้วจึงทำการเปรียบเทียบค่าความแม่นยำในการจำแนกภาพก่อนและหลังการใช้เทคนิคการเพิ่ม

4) ผลการทดลอง

ในการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันโดยกำหนดให้เครื่องคอมพิวเตอร์เรียนรู้จำนวน 300 รอบ และเปรียบเทียบค่าความแม่นยำในการจำแนกภาพก่อนและหลังการใช้เทคนิคการเพิ่มภาพ

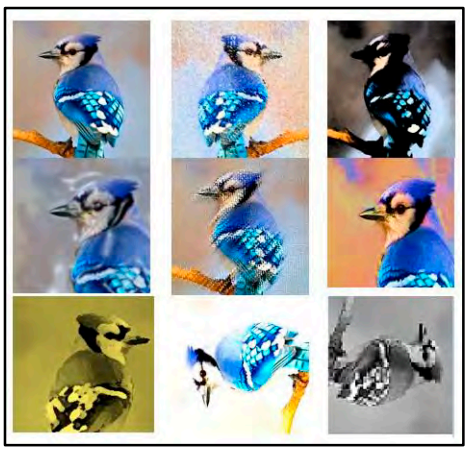
การจำแนกภาพก่อนใช้เทคนิคการเพิ่มภาพ จำนวนรูปที่ใช้ในการฝึกสอน 50,000 ภาพ จำนวนเรียนรู้ 300 รอบ หากเปรียบเทียบกับจำนวนรอบในการฝึกสอนให้คอมพิวเตอร์เรียนรู้กับค่าความแม่นยำในการจำแนกภาพ จะพบว่าเมื่อเพิ่มจำนวนรอบในการฝึกสอนมากขึ้น ความแม่นยำในการจำแนกภาพก็จะสูงขึ้นตามไปด้วย ดังรูปที่ 6 จากกราฟจะเห็นได้ว่าในรอบที่ 250-300 ค่าความแม่นยำในการทำนายเริ่มคงที่ มีการเปลี่ยนแปลงเพียงเล็กน้อย จึงหยุดการฝึกสอนที่ 300 รอบ เส้นทึบแสดงค่าความแม่นยำจากชุดการสอน มีค่า 96.90% ซึ่งมากกว่าค่าความแม่นยำที่ได้จากชุดทดสอบที่มีค่า 84.79%



รูปที่ 6: กราฟแสดงค่าความแม่นยำในการจำแนกภาพของการฝึกสอนและการทดสอบ ก่อนการใช้เทคนิคการเพิ่มภาพ

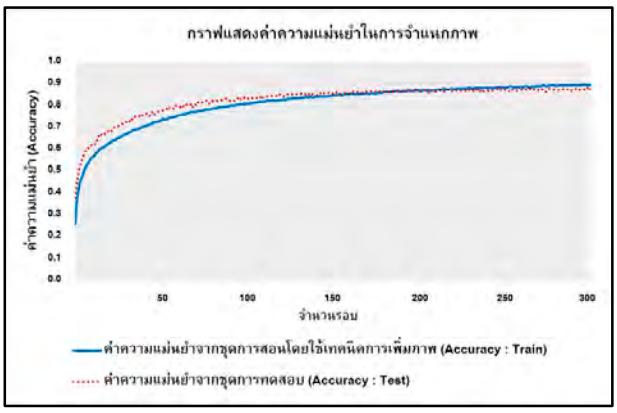
ในการปรับปรุงประสิทธิภาพของโมเดลในการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันนั้น จะใช้เทคนิคการเพิ่มภาพ เหตุผลหลักที่ใช้เทคนิคการเพิ่มภาพนี้เพื่อลดปัญหา Overfitting ซึ่งเป็นเหตุการณ์ที่โมเดลทำนายได้ดีมากในขั้นตอนของการสอน (Training data) แต่เมื่อนำโมเดลนั้นไปทำงานกับข้อมูลที่ต้องการทดสอบ (Testing data) ประสิทธิภาพในการทำนายต่ำมาก ซึ่งเป็นข้อมูลที่โมเดลไม่เคยเห็นมาก่อน และปัญหาจากการที่มีข้อมูลไม่เพียงพอในการฝึกสอนให้คอมพิวเตอร์เรียนรู้

การใช้เทคนิคการเพิ่มภาพ โดยการสร้างภาพเทียมจากเทคนิคการปรับสีของภาพ หมุนภาพ ย่อ ขยายภาพ กลับด้านของภาพ ดังแสดงตัวอย่างในรูปที่ 7



รูปที่ 7: ตัวอย่างภาพที่ได้จากการสร้างขึ้นมาใหม่ (generate image augmentation)

ในการทดลองได้ทำการสุ่มภาพจากทุกหมวดมา 10,000 ภาพ และทำการสุ่มใช้เทคนิคการปรับสีของภาพ การหมุนภาพ การย่อ ขยายภาพ หรือการกลับด้านของภาพ อย่างใดอย่างหนึ่ง ในแต่ละภาพ จนได้ภาพใหม่เพิ่มขึ้นจากเดิมจำนวน 10,000 ภาพ เมื่อรวมกับของเดิมอีก 50,000 ภาพที่ใช้สำหรับฝึกสอนในครั้งแรก รวมเป็น 60,000 ภาพ แล้วจึงฝึกให้คอมพิวเตอร์เรียนรู้ใหม่ 300 รอบ ผลลัพธ์ที่ได้ดังแสดงในรูปที่ 8



รูปที่ 8: กราฟแสดงค่าความแม่นยำในการจำแนกภาพหลังจากการใช้เทคนิคการเพิ่มภาพ

จากรูปที่ 8 กราฟแสดงค่าความแม่นยำในการจำแนกภาพ หลังจากการใช้เทคนิคการเพิ่มภาพ เส้นทึบแสดงค่าความแม่นยำจากชุดฝึกสอน เมื่อใช้เทคนิคการเพิ่มภาพมีค่า 89.17% และเส้นประแสดงค่าความแม่นยำจากชุดทดสอบ 87.57% จะพบว่าเมื่อมีการใช้เทคนิคการเพิ่มภาพแล้วค่าความแม่นยำในการจำแนกภาพจากชุดทดสอบมีค่าสูงใกล้เคียงกับค่าที่ได้จากชุดฝึกสอน

ตารางที่ 1 ทำการเปรียบเทียบค่าความแม่นยำในการจำแนกภาพก่อนและหลังการใช้เทคนิคการเพิ่มภาพ ที่จำนวนการเรียนรู้ 100, 200 และ 300 รอบ จะพบว่าเทคนิคการเพิ่มภาพช่วยให้ประสิทธิภาพในการจำแนกภาพหรือค่าในการทำนายถูกต้องสูงขึ้นจาก 84.79% เป็น 87.57% เพิ่มขึ้น 2.78%

ตารางที่ 1 เปรียบเทียบค่าความแม่นยำในการจำแนกภาพก่อนและหลังการใช้เทคนิคการเพิ่มภาพ

จำนวนรอบการเรียนรู้	ความแม่นยำในการจำแนกภาพจากชุดทดสอบก่อนการใช้เทคนิคการเพิ่มภาพ (%)	ความแม่นยำในการจำแนกภาพจากชุดทดสอบหลังการใช้เทคนิคการเพิ่มภาพ (%)
100	81.96	83.76
200	83.03	86.50
300	84.79	87.57

หากทำการเปรียบเทียบค่าความแม่นยำในการจำแนกภาพ ตามเทคนิคต่างๆ ของการสร้างภาพเทียม สามารถแสดงได้ดังตารางที่ 2

ตารางที่ 2 แสดงการเปรียบเทียบค่าความแม่นยำในการจำแนกภาพด้วยเทคนิคต่างๆ

เทคนิคการสร้างภาพเทียม	ค่าความแม่นยำในการจำแนกภาพ
การปรับสีของภาพ	85.26%
การหมุนภาพ	83.49%
การย่อ หรือ ขยายภาพ	86.84%
การกลับด้านของภาพ	83.71%

จากตารางที่ 2 จะพบว่า เทคนิคการสร้างภาพเทียมด้วยวิธีการย่อหรือขยายภาพ มีค่าความแม่นยำในการจำแนกภาพสูงที่สุดคือ 86.84% รองลงมาคือ การปรับสีของภาพ 85.26% การกลับด้านของภาพ 83.71% และการหมุนภาพ 83.49% ตามลำดับ การที่เทคนิคการย่อหรือขยายภาพ มีค่าความแม่นยำในการจำแนกภาพสูงที่สุด อาจเป็นเพราะการย่อภาพทำให้ความชัดเจนและความละเอียดของภาพที่ได้ลดลง ส่วนการขยายภาพทำให้ความชัดเจนของภาพที่ได้เพิ่มขึ้น จึงส่งผลต่อการเรียนรู้ของเครื่อง เมื่อนำภาพจากชุดทดสอบไปเปรียบเทียบกับภาพตัวอย่างจากชุดฝึกสอน ทำให้พบภาพที่มีความใกล้เคียงกันมากขึ้น จึงทำให้ค่าความแม่นยำในการจำแนกภาพถูกต้องสูง

จากผลการทดลองจะพบว่าเทคนิคการกลับด้านของภาพ ในงานวิจัยนี้คือ 83.71% ซึ่งสอดคล้องกับงานวิจัยของ Hussain และคณะ [9] ที่ได้ทำการทดลองเทคนิคการเพิ่มภาพด้วยการกลับด้านของภาพได้ค่าความแม่นยำ 84%

5) สรุปผลและข้อเสนอแนะ

ในการพัฒนาปรับปรุงประสิทธิภาพการจำแนกภาพด้วยโครงข่ายประสาทแบบคอนโวลูชันโดยใช้เทคนิคการเพิ่มภาพ จะช่วยให้การฝึกสอนคอมพิวเตอร์ให้เรียนรู้มีประสิทธิภาพมากยิ่งขึ้น และจำนวนรอบในการฝึกสอนซ้ำๆ ก็มีส่วนสำคัญเช่นกัน ยิ่งมีจำนวนรูปภาพมาก และจำนวนรอบในการฝึกสูง ความแม่นยำในการจำแนกรูปภาพก็สูงขึ้นด้วย แต่ไม่สามารถระบุได้อย่างชัดเจนว่า ควรใช้กี่ภาพ และจำนวนรอบ

ในการฝึกสอนควรเป็นเท่าใด เพราะประสิทธิภาพในการทำนายขึ้นอยู่กับชนิดของภาพ ที่มีความซับซ้อนแตกต่างกัน เช่น จำแนกภาพใบหน้าของคน กับจำแนกภาพวัตถุ สิ่งของ ซึ่งการจำแนกภาพใบหน้าคนที่มีความใกล้เคียงกันมากและมีความซับซ้อนมาก ควรมีจำนวนภาพที่มากและรอบในการฝึกสอนที่สูง แต่หากต้องการจำแนกว่าภาพนี้เป็นรถยนต์ หรือ เรือ ซึ่งความซับซ้อนไม่มาก จำนวนภาพและรอบในการเรียนรู้อาจจะไม่จำเป็นต้องมากเท่าการจำแนกภาพใบหน้าคน สำหรับจำนวนรอบที่ใช้ในการฝึกสอน สามารถเปรียบเทียบค่าความแม่นยำในการจำแนกภาพกับจำนวนรอบในการฝึกสอน หากค่าความแม่นยำในการจำแนกถึงจุดที่สูงในระดับที่พอใจและค่าเริ่มคงที่ ก็สามารถหยุดการฝึกสอนได้

ในการปรับปรุงประสิทธิภาพของโมเดล CNN ในอนาคต สามารถทำได้โดย 1) การเพิ่มจำนวนรอบในการฝึกสอน (training) ให้มากขึ้น ซึ่งในการทดลองนี้ได้ ทำการ training เพียง 300 รอบเท่านั้น 2) ใช้เทคนิคการเพิ่มจำนวนภาพให้มากยิ่งขึ้น ข้อควรคำนึงถึงคือการเพิ่มรูปภาพให้มีจำนวนมากขึ้น เวลาที่ใช้ในการฝึกสอน (training) ก็จะนานตามไปด้วยเช่นกัน

REFERENCES

[1] N. Dalal, and B. Triggs, “HISTOGRAMS OF ORIENTED GRADIENTS FOR HUMAN DETECTION,” *IEEE Computer Society Conference On Computer Vision And Pattern Recognition*, vol. 1, pp. 886–893, 2005.

[2] M. A. Tanner, and W. H. Wong, “The Calculation of Posterior Distributions By Data Augmentation,” *Journal Of The American Statistical Association*, vol. 82 No. 398, pp.528–540, 1987.

[3] G. E. Hinton, S. Osindero, and Y.W. Teh, “A FAST LEARNING ALGORITHM FOR DEEP BELIEF NETS,” *NEURAL COMPUTATION*, vol.18, no.7, pp.1527 –1554, 2006.

[4] Z. Zafrulla, H. Brashear, n T. Starner, H. Hamilton, and P. Presti, “American Sign Language Recognition With the Kinect,” *ICMI '11 Proceedings of the 13th international conference on multimodal interfaces*, pp. 279-286, 2011.

[5] T. Siriborvornratanakul, “Five questions with Deep Learning: Automatic cucumber sorting system from pictures @Cucumber Farm in Japan (Part 2/2).”[Online]. Available: https://mgronline.com/daily/detail/9590000091327 [Accessed: 20-Dec-2018].

[6] Z. Hussain, F. Gimenez, D. Yi and D. Rubin. “Differential Data Augmentation Techniques for Medical Imaging Classification Tasks,” *AMIA Annual Symposium Proceedings Archive,* pp. 979-984, 2018.

[7] C. Niyomthum, “What is CNN?.” [Online]. Available: https://medium.com/@thebear19/neural-network-101-cnn-with-tensorflow-fd5d515e979b. [Accessed: 04-Oct-2018].

[8] Mc. Ai, “Let's see how CNN Thinks!!!.” [Online]. Available: https://mc.ai/มาลองดูวิธีการคิดของ-cnn-ก/. [Accessed: 10-Nov-2018].

[9] M. D. Bloice, C. Stocker, and A. Holzinger, “Augmentor: An Image Augmentation Library for Machine Learning,” *The Journal of Open Source Software*, vol.2. pp. 1-5, 2017.

[10] C. Shorten, “Image Augmentation Examples in Python.” [Online]. Available: https://towardsdatascience.com/image-augmentation-examples-in-python-d552c26f2873. [Accessed: 10-Nov-2018].

[11] P. Bee, “How to install Keras + Tensorflow (GPU version) on windows 10.” [Online]. Available: https://medium.com/boobeejung/วิธีการติดตั้ง-keras-tensorflow-gpu-version-บน-windows-10-บทความนี้เขียนขึ้นเพื่อหลายๆ-คนที่กำ-e8e3f5105baa. [Accessed: 04-Oct-2018].

[12] K. Jarrett, K. Kavukcuoglu, M. Ranzato and Y. LeCun, “What Is The Best Multi-Stage Architecture for Object Recognition?,” in *2009 IEEE 12th International Conference On Computer Vision*, pp. 2146–2153, 2009.