บทที่ 2
ทฤษฎีและวิจัยงานที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การประมวลผลภาพ (Images Processing)

2.1.1.1 ภาพ (Images)

เป็นข้อมูลรูปแบบหนึ่งที่มีในระบบคอมพิวเตอร์ ซึ่งจะแสดงให้เห็นในรูปแบบดิจิทัล ปกติแล้วภาพในระบบดิจิทัลจะเป็นข้อมูลขนาดเล็ก ๆ ที่เรียงกันในภาพเรียกว่า พิกเซล(Pixel) โดยจะมีค่าส่องสว่าง (Intensity) ที่ต่างกัน ดังนั้นตอนที่พิกเซลทีมีค่าส่องสว่างมาเรียงต่อ ๆ กัน เป็นจำนวนมาก ๆ จะทำให้เราสามารถมองเห็นภาพได้ในระบบคอมพิวเตอร์ การประมวลผลภาพดิจิทัล (Digital Image Processing) เป็นการประมวลผล ข้อมูลภาพที่ได้จาก กล้องถ่ายรูป และตัวรับรู้ (Sensors) เพื่อทำการปรับปรุง ปรับเปลี่ยนคุณสมบัติ ของภาพในการนำไปใช้ในโปรแกรมประยุกต์ตามที่ต้องการ (Applications) ตัวอย่างเช่น การตรวจจับใบหน้าคนในภาพ (face detection) การตรวจจับรอยร้าของเลือดในตา (Extrusion detection in retina images) เป็นต้น

2.1.1.2 ภาพดิจิทัล

ภาพดิจิทัลจะแสดงข้อมูลที่มีความสัมพันธ์เชิงพื้นที่ที่ข้อมูลจะจะบอกถึงความส่องสว่าง และ สีของระบบข้อมูลแบบ 2 มิติ (2-D discrete space) ภาพ (m = 1,2,3,...., M : n = 1,2,3,...,N) ซึ่งได้จากการแสดงข้อมูลภาพ I(x ,y) ซึ่งจะเกิดกับประเภทของเซนเซอร์บางประเภทเท่านั้น เช่น กล้องแบบ CDD (Charge Coupled Device) โดยส่วนมากค่าที่ได้จากค่าเฉลี่ยสัญญาณโดยรอบ ของจุดที่สนใจ ตัวแปร m และ n จะใช้อ้างแทน แถว และ คอลัมน์ของภาพ จุดแต่ละจุดของภาพ (พิกเซล หรือ Pixel) จะใช้ข้อมูลแบบอาเรย์ 2 มิติ ในการประมวลผลภาพจากระบบคอมพิวเตอร์ภาพดิจิทัล หรือ Digital Images คือ อาเรย์หรือ เมตริกของจุดภาพหรือพิกเซลในภาพ (โดยพิกเซลมีรูปร่างเป็นสี่เหลี่ยมจัตุรัส)ที่เรียงตัวกันในรูปแบบของแถว (Rows) และหลัก หรือ คอลัมน์ (columns) พิกเซลที่อยู่ในภาพนั้นจะมีค่าความส่องสว่าง (Intensity) ที่อยู่ในช่วงที่สามารถแสดงผลได้ตัวอย่างเช่น ภาพระดับเทา (Geyscale Images) แบบ 8 บิต 1 พิกเซลจะแทนด้วยค่าความส่องสว่างที่แทนค่าด้วย 8 บิต

2.1.1.3 สีของภาพ

โดยทั่วไปแล้วช่องสีจะมีอย่างน้อง 1 ช่องสี ซึ่งจะเป็นค่ากำหนดความส่องสว่างและค่าสี ที่ตำแหน่งใดตำแหน่งหนึ่งของภาพ I(m ,n) เช่นในกรณีที่ภาพมีช่องสีเพียงช่องเดียว ค่าสีของแต่ละพิกเซลจะเป็นจำนวนเต็มบวกเพียงค่าเดียว จะแทนค่าของระดับสัญญาณที่ตำแหน่งนั้น ๆ ในภาพ การนำค่าสีที่แทนค่าด้วยตัวเลข ไปใช้กับอุปกรณ์เพื่อแสดงภาพนั้นเรียกว่า

Computer Science Department
Faculty of Informatics Mahasarakham University

Color-Map ซึ่ง Color-Map นั้นจะกำหนดเฉดสีให้กับตัวเลขที่แสดงระดับค่าสี แล้วจะทำให้เราสามารถมองเห็นภาพเป็นสีต่าง ๆ ได้ ซึ่งภาพที่มีช่องสีช่องเดียว Color-Map ที่นิยมนำมาใช้ในการประมวลผล ระดับเทา (Greyscale) โดยภาพที่มี Color-Map เป็นระดับเทานั้น เราจะเรียกภาพนั้นว่าภาพระดับเทา (Greyscale images)

2.1.1.4 ภาพแบบ RGB (RGB image)

คือ RGB Image หรือ True Color Image เป็นรูปที่เก็บโดยใช้อาร์เรย์ 3 มิติ ขนาด m x n x 3โดยที่ m คือความยาว และ n คือความกว้างของภาพในหน่วยพิกเซล ส่วนมิติสุดท้ายนั้นในแต่ละมิติจะเก็บค่าสีแยกกัน คือสีแดง (Red) สีเขียว(Green) และสีน้ำเงิน (Blue)

2.1.1.5 ระบบสี RGB

ภาพ RGB หรือที่เรียกว่า ภาพแบบสีจริง (True color) จะมี 3สี ได้แก่ แดงเขียวและน้ำเงินตามลำดับ เรียกว่าเป็นการเก็บข้อมูลแบบ 3มิติ 1 พิกเซลในภาพนั้นจะมีองค์ประกอบของสีทั้งหมด 3 องค์ประกอบ ซึ่งในระบบคอมพิวเตอร์สีที่จะแสดงที่ของภาพที่เกิดจากการ (Map-Color) ของทั้ง 3 องค์ประกอบสีและจะเห็นได้ว่า 1 พิกเซล ในภาพนั้นต้องใช้ความละเอียดบิตขนาด 24 บิต ในการแสดงค่าสีทำให้ภาพที่มีองค์ประกอบ หรือ ช่องสี3 ช่องสีจะถูกเรียกว่าภาพแบบ 24 บิต ซึ่งในการสร้างสีจะเกิดจากการผสมแม่สีแสงแบบบวก (Additive Color Mixing) ของทั้ง 3 ช่องสี (ซึ่งแสดงในรูปแบบแม่สีแสง) ตัวอย่างเช่น สีขาวจะเกิดจากการผสมระหว่างช่องสีแดงในปริมาณที่เท่ากัน และมักจะแทนค่าด้วย(255,255,255) ซึ่งหมายถึง ช่องสีแดงจะมีค่าสี255 ช่องเขียวจะมีค่าสี 255 และช่องสีน้ำเงินจะมีค่าสี 255 ตามลำดับ

2.1.1.6 การดำเนินการสัณฐานของภาพ (Morphology Operation)

การดำเนินการปรับสัณฐานของภาพเป็นกระบวนการเพื่อปรับเปลี่ยนรูปส่วนของวัตถุที่อยู่ในภาพ โดยการเปลี่ยนรูปร่างนี้จะอาศัยการประมวลผลทางคณิตศาสตร์โดยการใช้การคำนวณด้วยกลุ่มของฟังก์ชันแบบเชิงเส้น โดยทั่วไปการดำเนินการ Morphology นั้นจะใช้กับภาพที่เป็นภาพขาวดำ (ภาพแบบไบนารี) โดยทั่วไปการดำเนินการ Morphology นั้นจะเป็นการเปลี่ยนแปลงรูปร่างของวัตถุในภาพตามรูปแบบของต้นแบบ (Templates หรือ Structural Elements) ซึ่งต้นแบบนี้จะถูกนำไปประมวลผลกับภาพในทุก ๆ ตำแหน่ง (ตำแหน่งของพิกเซล) เพื่อดำเนินการเปลี่ยนแปลงข้อมูลกับพิกเซลข้างเคียงตามที่กำหนดในต้นแบบ ในหัวข้อนี้จะอธิบายรายละเอียดเกี่ยวกับการดำเนินการ Morphology ได้แก่การขยายรูปร่าง การลดรูปร่าง การทำ Openingและ Closing

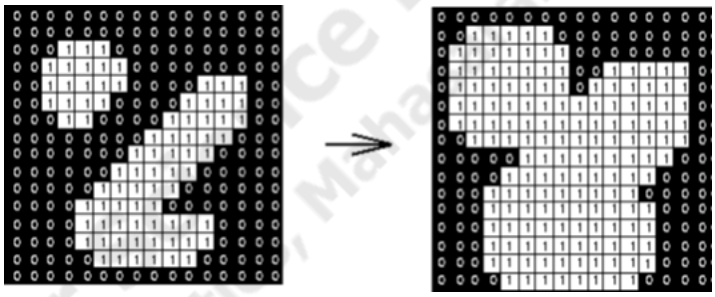
Computer Science Department Faculty of Informatics Mahasarakham University

2.1.1.7 ต้นแบบ (Structural Element)

ต้นแบบ H (Structural Elements หรือ Templates) จะเป็นข้อมูลภาพแบบไบนารีแบบหนึ่งที่สามารถใช้ในการกำหนดรูปแบบของต้นแบบตามที่ต้องการ ผ่านการกำหนดค่าของข้อมูลข้างเคียงของจุดกลาง (Original หรือ Spot) ในภายในต้นแบบ

2.1.1.8 การขยายส่วนของวัตถุในภาพ (Image Dilation)

การขยาย (Dilation) จะพิจารณาข้อมูลภาพซึ่งเป็นภาพขาว-ดำ เป็นการขยายภาพให้ใหญ่ขึ้น เพื่อเพิ่มสีให้กับวัตถุที่แสดงผลในขั้นตอนสุดท้าย ซึ่งการขยายวัตถุจะทาได้โดยการกำหนดส่วนประกอบโครงสร้าง (Structuring Element) และนำส่วนประกอบโครงสร้างไปกราดบนข้อมูลภาพตามลำดับตลอดทั้งภาพ โดยเมื่อจุดเริ่มต้นของส่วนประกอบโครงสร้างหรือจุดกำเนิดตรงกับตำแหน่งข้อมูลภาพที่เท่ากับ 1 จะทำการยูเนียนส่วนประกอบโครงสร้าง เข้ากับข้อมูลภาพ

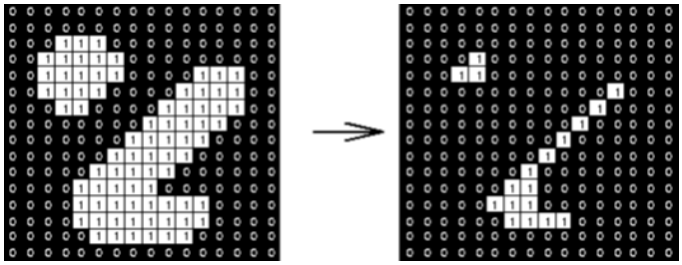


ภาพประกอบที่ 2.1 แสดงผลของการขยายส่วนของวัตถุในภาพ (Image Dilation)
ที่มา : [6] Nick Efford. Digital Image Processing: A Practical Introduction Using Java. PearsonEducation,2000[Online].
www.cs.princeton.edu/~pshilane/class/mosai/

2.1.1.9 การกัดกร่อนวัตถุภาพ (Image Erosion)

การกร่อนขนาด (Erosion) เป็นการกร่อนขนาดบริเวณขอบของวัตถุ ซึ่งการกร่อนมีวิธีคล้ายกับการขยายคือ สร้างส่วนประกอบโครงร่างขึ้นมาแล้วนำไปกราดตามข้อมูลภาพ โดยจะเลื่อนไปทุกตำแหน่งเปรียบเทียบกับข้อมูลภาพ ถ้าข้อมูลมีค่าเหมือนกับส่วนประกอบโครงร่างจะทำการกำหนดค่าข้อมูลภาพที่ตรงกับตำแหน่งที่ตรงกับจุดเริ่มต้นหรือ จุดกำเนิดของส่วนประกอบโครงร่างให้เท่ากับ 1 ดังภาพประกอบที่ 2.2

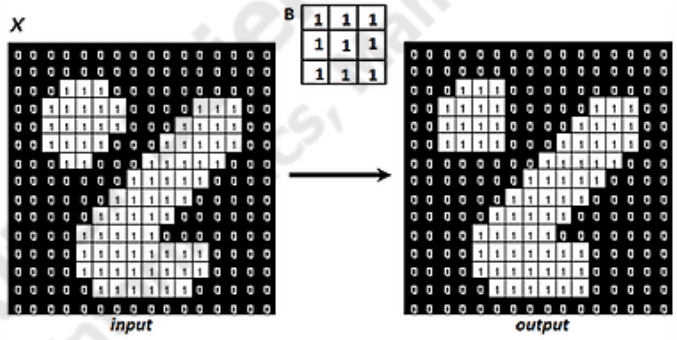
Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 2.2 แสดงผลของการกัดกร่อนวัตถุภาพ (image erosion)
ที่มา : [6] Nick Efford. Digital Image Processing: A Practical Introduction Using Java. PearsonEducation,2000[Online].www.cs.princeton.edu/~pshilane/class/mosai/

2.1.1.10 การทำ Opening

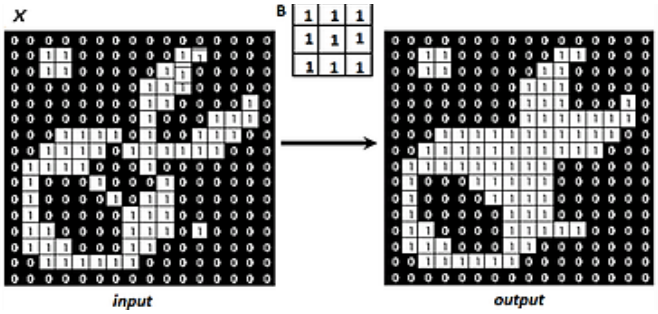
Opening นิยามของ opening ง่ายๆ คือเอา Image มา Erode แล้วค่อย Dilate ใช้ในการลบ Noise (เพราะว่า Noise หายไปตอน Erode แต่ขนาดของวัตถุเล็กลงก็เอาคืนด้วยการ Dilate) ใช้ในการลบขอบที่ยื่นๆ ของวัตถุด้วย



ภาพประกอบที่ 2.3 แสดงผลของการเปิดโดยใช้องค์ประกอบโครงสร้างสี่เหลี่ยม 3x3
ที่มา : [6] Nick Efford. Digital Image Processing: A Practical Introduction Using Java. PearsonEducation,2000[Online].www.cs.princeton.edu/~pshilane/class/mosai/

2.1.1.11 การทำ closing

Closing ตรงข้างกับ Opening คือการนำ Image มา Dilate แล้ว ค่อย Erode ใช้ในการลบ Small Holes (หายไปตอบ Dilate แล้วลดขนาดวัตถุที่บวม ขึ้นมาด้วย Erode) สามารถใช้ในการ เชื่อมวัตถุที่แยกจากกัน (เพราะ Noise



ภาพประกอบที่ 2.3 แสดงผลของการเปิดโดยใช้องค์ประกอบโครงสร้างสี่เหลี่ยม 3x3(ต่อ)
ที่มา : [6] Nick Efford. Digital Image Processing: A Practical Introduction Using Java. PearsonEducation,2000[Online].www.cs.princeton.edu/~pshilane/class/mosai/

2.1.1.12 การหาส่วนรอบวัตถุ (Boundary Extraction)

เราสามารถหาส่วนรอบวัตถุได้ด้วยการทำ Erosion ภาพด้วยต้นแบบที่มีขนาดเล็กๆ จากนั้นก็ทำการนำภาพก่อนมา Erosion และทำการลบด้วยผลลัพธ์ที่ได้จากการทำ Erosion ดังนั้น

$$I_b = I - (I \ominus H)$$

สำหรับภาพไบนารี I และต้นแบบ H ส่วนของเส้นรอบวัตถุ *A* (Ib) สามารถหาได้จาก

2.1.1.13 การเติมส่วนพิกเซล (Region Filling)

โดยปกติภาพแบบไบนารีจะได้จากการแบ่งส่วนภาพโดยการใช้ค่าขีดแบ่ง (Threshold) จากภาพระดับเทาหรือภาพสีโดยการดำเนินการดังกล่าวจะทำให้เกิดส่วนภาพที่เรียกว่า หลุม (Holes)

2.1.2 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) เป็นเทคนิคที่ได้รับความนิยมเป็นอย่างมาก เนื่องจากการผลลัพธ์ที่ได้จากการทำงานของ CNN ในการนำมาใช้กับการจำแนกของภาพได้ผลลัพธ์ที่ดีและเป็นที่น่าพอใจ นอกจากนั้นยังมีการทำเอาเทคนิคของ CNN มาใช้ในแอพพลิเคชันในหลากหลายแอพพลิเคชันข้อดีของ CNN เมื่อเทียบกับเทคนิคเดิมสำหรับการรู้จำวัตถุในภาพ คือ การตรวจจับคุณลักษณะที่สำคัญโดยอัตโนมัติโดยไม่ต้องให้มนุษย์ควบคุม และมีประสิทธิภาพในเชิง

Computer Science Department
Faculty of Informatics, Mahasarakham University

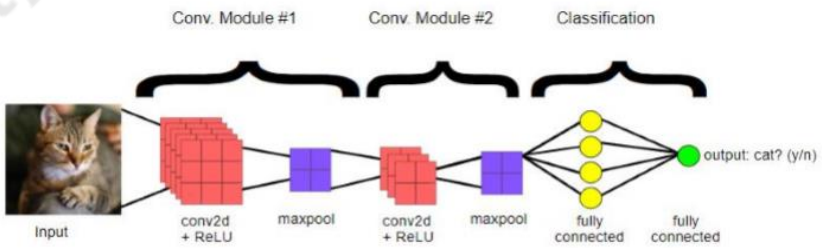
คำนวณ ใช้ convolution พิเศษและรวมการดำเนินงานและดำเนินการแชร์พารามิเตอร์ ซึ่งทำให้ซีเอ็นเอ็นสามารถทำงานบนอุปกรณ์ใด ๆ ทำให้น่าสนใจอย่างกว้างขวาง

Convolutional Neural Networks หรือ (CNN) ได้รับการการพัฒนาแนวความคิดมาจาก Multiple Layer Perceptron (MLPs) [3] ซึ่งเป็นการทำงานที่อยู่บนวิธีการของโครงข่ายประสามเทียม (Artificial Neural Network) หลักการทำงานของ CNN จะทำงานในรูปแบบของเทคนิคการ Optimization โดยจะมีการนำเข้าข้อมูลที่เรียกว่า ชุดฝึกฝน และ ผลเฉลยของชุดฝึกฝน CNN จะประกอบไปด้วยพารามิเตอร์ ซึ่งเรียกว่า พารามิเตอร์ของโมเดล จากนั้น CNN จะทำการ optimize พารามิเตอร์เหล่านี้โดยการลดข้อผิดพลาดระหว่างข้อมูลนำเข้าและผลเฉลย ที่เรียกว่า “loss” โดยระบบของ CNN จะทำให้ loss ที่เกิดขึ้นในระบบน้อยที่สุด โดยการปรับค่าพารามิเตอร์ของโมเดล ชุดพารามิเตอร์ที่ดีที่สุด (ที่ทำให้เกิด loss น้อยที่สุด) จะถูกนำมาใช้เป็นโมเดล เพื่อนำมาใช้งานต่อไป องค์ประกอบของ CNN ประกอบไปด้วย องค์ประกอบดังต่อไปนี้

2.1.2.1 Input Layer ข้อมูลรูปภาพใบหน้าที่นวนเข้ามาไปสกัดเอกลักษณ์เพื่อที่จะนำไปจำแนกประเภท

2.1.2.2 The Hidden layers (ส่วนที่สกัดเอกลักษณ์) ส่วนการดึงคุณลักษณะของข้อมูล โดยในส่วนนี้จะดำเนินการ convolutions ซึ่งสามารถกำหนดความละเอียด (ชั้น) ของการ convolutions และจากนั้นจะมีการรวมการดำเนินการ ที่เรียกว่า pooling ซึ่งจะถูกนำมาสร้างเป็นเอกลักษณ์เพื่อใช้ในขั้นตอนการจำแนกต่อไป

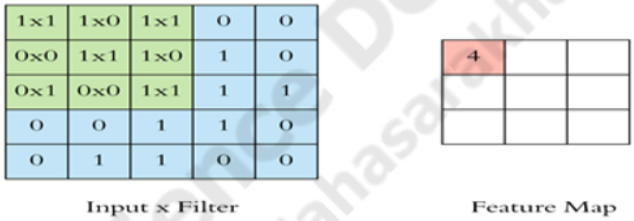
2.1.2.3 ส่วนการจำแนกประเภท ซึ่งโดยทั่วไปจะใช้โครงข่ายประสาทเทียม (fully connected neural network) โดยจะทำหน้าที่นำข้อมูลจากขั้นตอนที่ 2.1.2.2 มาทำการประเมินเพื่อหากลุ่มข้อมูลของข้อมูลนำเข้า โดยจะการลดค่า loss ที่เกิดขึ้นในระบบทั้งหมด โครงสร้างของ CNN สามารถแสดงดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.4 แสดงโมเดล CNN ทั้งหมดที่มีสถาปัตยกรรมที่คล้ายกัน

ที่มา : [7] จักรกฤษณ์ ประดุจชนม์. (2019). เรียนรู้และทำความเข้าใจเรื่อง Convolutional Neural Network (CNN). (ออนไลน์). https://www.glurgeek.com/education/ml-cnn/. [11 สิงหาคม2563]

จากภาพประกอบที่ 2.4 สามารถสรุปการทำงานได้ดังต่อไปนี้ เมื่อนำภาพเข้าประมวล convolution + pooling operations จะมีการทำงานซึ่งภาพจะถูก convolute ด้วย mask ที่มีการกำหนดขึ้น จากนั้นจะมีการรวมผลลัพธ์เพื่อเป็น output สำหรับการ convolution จำนวนของการทำ convolution จะถูกกำหนดก่อนการทำงานซึ่งจำนวนของ convolution นี้จะเรียกว่า hidden layers ในระบบ สำหรับการสร้าง CNN จะต้องคำนึงถึงองค์ประกอบดังต่อไปนี้



ภาพประกอบที่ 2.5 ตัวอย่างการดำเนินการในรูปแบบ 2D โดยใช้ตัวกรอง 2x2

ที่มา : [7] จักรกฤษณ์ ประดุจชนม์. (2019). เรียนรู้และทำความเข้าใจเรื่อง Convolutional Neural Network (CNN). (ออนไลน์). https://www.glurgeek.com/education/ml-cnn/. [11 สิงหาคม2563]

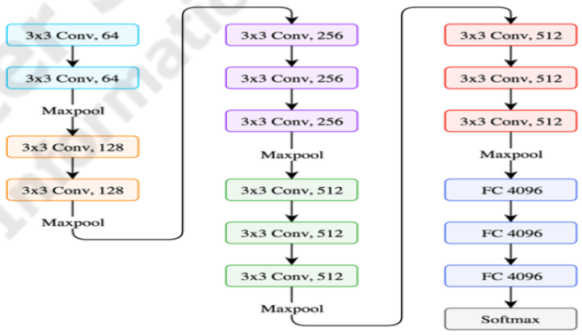
2.1.2.4 Convolution Layer โครงสร้างหลักของ CNN คือการคำนวณทางคณิตศาสตร์เพื่อทำการ Convolute ข้อมูล 2 ข้อมูล ได้แก่ ข้อมูลภาพและข้อมูล Mask จากนั้นภาพจะถูกกรองด้วย Mask โดยการประมวลผล Convolution จากนั้นจะทำการเลื่อนตัวกรองนี้ไปไปยังทุกตำแหน่งในภาพ (พิกเซล) ตัวอย่างของมูลภาพและ Mask แสดงดังภาพประกอบที่ 8

2.1.2.5 Pooling การพูลลิ่ง ช่วยลดมิติของฟีเจอร์แมพลงแต่ยังคงรักษาข้อมูลสำคัญไว้การพูลลิ่งสามารถจำแนกเป็นประเภทต่างได้เช่น พูลลิ่งด้วยค่าสูงสุด (Max Pooling), ค่าเฉลี่ย (Average Pooling), ผลรวม การพูลลิ่ง ทำให้ผลลัพธ์ที่ได้มามีขนาดเล็กและจัดการได้ง่ายขึ้น นอกจากนี้ยังลด จำนวนพารามิเตอร์และการคำนวณที่เกินจำเป็นในโครงข่าย

Computer Science Department Faculty of Informatics, Mahasarakham University

2.1.2.6 Training การฝึกฝนจะมีส่วนอย่างมากที่จะทำให้ CNN สามารถแสดงผลลัพธ์ที่แม่นจำ โดยการฝึกฝนจะมีพารามิเตอร์ที่จะต้องกำหนดอยู่ 2 ค่า คือ Leading Rate และ Epoch Learning Rate จะเป็นค่าหน่วยข้อมูลทีถูกกำหนดขึ้นเพื่อใช้ในการปรับปรุงค่าของพารามิเตอร์ของ CNN ในแต่ละรอบ ในขณะที่ Epoch จะเป็นตัวกำหนดรอบที่ CNN จะทำงานมากที่สุด

จะเห็นว่าการทำงานของ CNN จะเป็นรูปแบบของการทำงานวนซ้ำเพื่อ update ค่าของพารามิเตอร์ เพื่อให้ได้ค่า loss ที่น้อยที่สุด ดังนั้นการทำงานของ CNN จะขึ้นอยู่กับการออกแบบ ชุดของ layers หรือ สถาปัตยกรรมของ CNN ตัวอย่าง CNN สำหรับทำงานตัวไปได้แก่ VGG Model VGG เป็นเครือข่ายประสาทเทียมแบบ CNN ที่พัฒนาจากนักวิจัยที่ Oxford Geometry Group ซึ่งเป็นชื่อ VGG โดยสถาปัตยกรรมถูกพัฒนาเพื่อใช้ในการจำแนกข้อมูลภาพจากฐานข้อมูล ImageNet โดยประสิทธิภาพของการทำงานในการจำแนกข้อมูลภาพจะมีอัตราความผิดพลาดอยู่ที่ 7.3% เป็นชุดข้อมูลภาพที่มีเนื้อหาครอบคลุมมากที่สุดและมีการแข่งขันทุกปีซึ่งนักวิจัยจากทั่วโลกแข่งขันกัน สถาปัตยกรรมซีเอ็นเอ็นที่มีชื่อเสียงทั้งหมดเปิดตัวในการแข่งขันครั้งนี้ในบรรดาโมเดล CNN ที่มีประสิทธิภาพดีที่สุด VGG โดดเด่นด้วยความเรียบง่าย ตัวอย่างของสถาปัตยกรรมของ VGG แสดงดังภาพประกอบที่ 2.7



ภาพประกอบที่ 2.6 VGG Layer

ที่มา : [7] จักรกฤษณ์ ประดุจชนม์. (2019). เรียนรู้และทำความเข้าใจเรื่อง Convolutional Neural Network (CNN). (ออนไลน์). https://www.glurgeek.com/education/ml-cnn/. [11 สิงหาคม2563]

VGG เป็น CNN ที่ประกอบไปด้วย layer ที่ใช้ในการทำงานของเครือข่ายประสาทเทียมทั้งหมด 16 ชั้นโดยไม่นับชั้นของ Maxpool และ SoftMax ในตอนท้าย นอกจากนี้ยังเรียกว่า

Computer Science Department
Faculty of Informatics, Mahasarakham University

VGG16 ซึ่งจะเป็นสถาปัตยกรรมเป็นสถาปัตยกรรมที่เราได้ทำงานร่วมกับข้างต้น การเรียงซ้อนกันแบบเรียงซ้อนกัน + เลเยอร์รวมกันตามด้วย ANN ที่เชื่อมต่อกันอย่างเต็มที่ สำหรับ CNN เราสามารถแสดงผลลัพธ์ในส่วนประกอบต่าง ๆ ได้ สิ่งนี้จะทำให้เรามองลึกเข้าไปในผลงานภายในของตนและช่วยให้เราเข้าใจได้ดียิ่งขึ้น ตัวอย่างเช่น Feature Maps ในแต่ละชั้นของ Convolution Layers แสดงดัง ภาพประกอบที่ 2.8

ภาพประกอบที่ 2.7 แสดง Feature maps ของแต่ละชั้นใน convolution layers
ที่มา : [7] จักรกฤษณ์ ประดุจชนม์. (2019). เรียนรู้และทำความเข้าใจเรื่อง Convolutional Neural Network (CNN). (ออนไลน์). https://www.glurgeek.com/education/ml-cnn/. [11 สิงหาคม2563]

2.1.3 เครื่องมือที่ใช้ในการพัฒนา

2.1.3.1 Hardware

1. Computer
2. กล้องวีดีโอ web cam

2.1.3.2 Software

1. Python 3.7.0
2. MySQL
3. JavaScript
4. PHP 7.3.0 ORC2

2.1.4 รายละเอียดโปรแกรมที่จะพัฒนา (Software Specification)

1) Input/output Specification

Input: ภาพวีดีโอที่ได้จากกล้อง Web Cam

Output: ข้อมูลของลูกค้าจากการรู้จำใบหน้ามีดังนี้

1.1 รายการสินค้าหรือบริการที่ลูกค้าเคยซื้อหรือใช้

1.2 ข้อมูลสถิตการซื้อสิ้นค้าหรือบริการที่ลูกค้ามาใช้

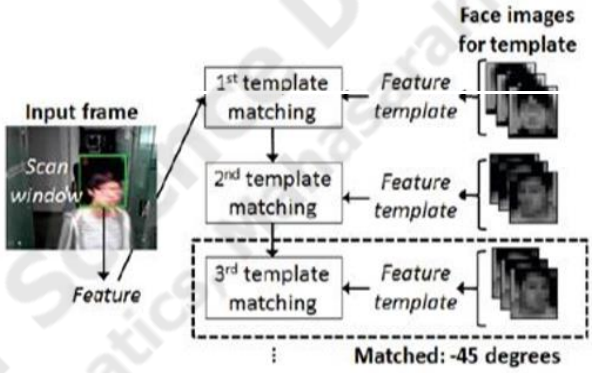
Computer Science Department
Faculty of Informatics, Mahasarakham University

2) Functional Specification

ในแต่ละฟังก์ชันของโปรแกรมประยุกต์นี้มีวิธีและการออกแบบวิธีการคิดและประมวลผลดังนี้

2.1 การตรวจจับใบหน้าจากภาพวีดีโอ ภาพนำเข้าจะเป็นภาพจากกล้องวีดีโอที่มีใบหน้าของลูกค้า โดยจะทำการติดตั้งกล้องไว้บริเวณเค้าท์เตอร์ของร้านค้า ขั้นตอนการตรวจจับใบหน้าในภาพมีดังต่อไปนี้

2.1.1 ใช้เทคนิคการใช้ต้นแบบ โดยภาพจะแบ่งออกเป็นส่วนๆ เรียกว่า Window จากนั้นแต่ละ Window จะทำการนำมาเปรียบเทียบกับต้นแบบของใบหน้า (Template) จากนั้นจะทำการคำนวณระยะทางของแต่ละ Window กับ Template ซึ่ง Template จะใช้ในรูปแบบหลายระดับ ภาพรวมของเทคนิคแสดงดัง ภาพประกอบที่ 2.9



ภาพประกอบที่ 2.8 แสดงขั้นตอนการตรวจจับใบหน้าในภาพ

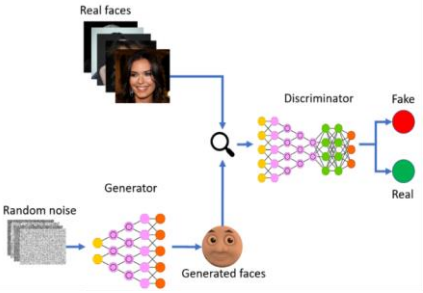
ที่มา : [8] Andrea Missinato. Generative adversarial networks: when the machine learnung is a game. (ออนไลน์). https://www.spindox.it/en/blog/generative-adversarial-neural-networks/. [2 กันยายน 2563]

2.1.2 ปรับปรุงผลการตรวจจับใบหน้า หลังจากที่มีการตรวจจับใบหน้าโดยใช้ Template Matching แล้วจะได้กลุ่มของ Window ที่คาดว่าจะเป็นใบหน้าในภาพ โดยกลุ่มภาพนี้อาจจะมี Window ที่ไม่ใช่ใบหน้า ดังนั้นขั้นตอนนี้จะทำการลบภาพในกลุ่ม ดังกล่าวที่ไม่ใช่ใบหน้าออก ซึ่งจะใช้ข้อมูลบริบทด้วยเทคนิค Igenface เข้ามาช่วย

2.1.5 รู้จำใบหน้า

ในขั้นตอนนี้จะทำการรู้จำใบหน้าที่มีการตรวจจับในขั้นตอนที่ การรู้จำใบหน้าจะใช้วิธีการเรียนรู้เชิงลึก (Deep Learning) โดยใช้เทคนิค Generative Adversarial Networks (GAN)

เทคนิคนี้จะทำการสกัดเอกลักษณ์ของใบหน้า ที่มีความเป็นเอกลักษณ์ของจุดในภาพ จะมีการปรับน้ำหนัก (Weights) ของโครงข่าย



ภาพประกอบที่ 2.9 แสดงการทำงานของ GAN ในการสกัดเอกลักษณ์ของใบหน้า

ที่มา : [8] Andrea Missinato. Generative adversarial networks: when the machine learnung is a game. (ออนไลน์). https://www.spindox.it/en/blog/generative-adversarial-neural-networks/. [2 กันยายน 2563]

น้ำหนักของโครงข่ายจะนำมาใช้ในการสร้างเอกลักษณ์ของใบหน้า จากนั้นจะนำข้อมูลใบหน้ามาเปรียบเทียบกับข้อมูลหน้าในฐานข้อมูลโดยใช้การคำนวณระยะทางระหว่างเอกลักษณ์ของภาพและภาพที่อยู่ในฐานข้อมูลจากนั้นทำการตั้งค่าขีดแบ่งเพื่อกำหนดว่าเป็นหน้าของลูกค้าที่มีในฐานข้อมูลหรือลูกค้าใหม่

2.1.6 การทำนายช่วงอายุและเพศ

ขั้นตอนนี้จะทำการทำการพยากรณ์ข้อมูลจากภาพที่ใช้ข้อมูลของใบหน้าลูกค้าเป็นจุดเริ่มต้นภาพใบหน้าของลูกค้าที่ตรวจจับได้ในขั้นตอนที่ 1 จะทำการขยายให้ใหญ่ขึ้นเพื่อให้สามารถเห็นส่วนอื่น ๆ ของลูกค้า เช่น ส่วนของผม และ ลำตัว จากนั้นภาพจะถูกส่งเข้าไปใน CNN (ที่อธิบายในหัวข้อที่ผ่านมา) เพื่อทำการพยากรณ์ช่วงอายุและเพศ โดยจะทำการสร้างตัวแบบ 2 ตัวแบบ ได้แก่

2.1.6.1 ตัวแบบสำหรับพยากรณ์ช่วงอายุ

2.1.6.2 ตัวแบบสำหรับพยากรณ์เพศ

2.1.7 การบันทึกข้อมูลลูกค้า

เมื่อได้ข้อมูลลูกค้าแล้วขั้นตอนต่อไปจะเป็นการบันทึกข้อมูลลงในฐานข้อมูล โดยจะแบ่งการบันทึกข้อมูลดังนี้

2.4.1. หากเป็นลูกค้าใหม่จะทำการการบันทึกข้อมูล record ใหม่ลงในฐานข้อมูล

Computer Science Department Faculty of Informatics Mahasarakham University

2.4.2. หากเป็นลูกค้าเก่าจะทำการบันทึกข้อมูลร้านการสินค้าที่ลูกค้าใช้หรือซื้อ

2.1.8 การแสดงข้อมูลลูกค้า

เมื่อลูกค้าเก่าเข้ามาใช้งานในระบบจะทำการแสดงรายการสินค้าและบริการที่ลูกค้าเคยเข้ามา ใช้งานหรือบริการในร้าน โดยจะทำการดึงข้อมูลจากฐานข้อมูลมาแสดง

2.1.9 การแสดงข้อมูลการใช้สินค้าบริการของร้าน

กระบวนการนี้จะเป็นการนำข้อมูลจากฐานข้อมูลที่เกี่ยวกับลูกค้า ได้แก่ อายุ เพศ และสินค้า มาแสดง

2.2 งานวิจัยที่เกี่ยวข้อง

Face Recognition คือ ระบบการรู้จดจำใบหน้า เป็นอัลกอริทึมหนึ่งที่ใช้งานกันอย่างแพร่หลายในการยืนยันบุคคล ไม่ว่าจะเป็นการล็อกอินเข้าใช้งานมือถือ จนไปถึงหน่วยงานความมั่นคงที่ใช้ Face Recognition

Face Recognition จึงเป็นสิ่งที่นักพัฒนาปัญญาประดิษฐ์ให้ความสนใจ เพราะ Face Recognition เป็นส่วนหนึ่ง ของ CV เพื่อให้คอมพิวเตอร์รับรู้และรู้จักสิ่งที่มนุษย์มองเห็น ไม่ว่าจะเป็นบุคคล สัตว์ สิ่งของ ในภาษา Python เราสามารถทำ Face Recognition ได้มานานแล้วโดยใช้ OpenCV และ ML ทำระบบตรวจจับใบหน้าด้วย OpenCV กับภาษา Python ปัญหาหลายอย่างในการทำฐานข้อมูล ซึ่งยุ่งยากเกินไป Python ได้มีนักพัฒนา ได้พัฒนาโมดูลที่ช่วยให้ทำ Face Recognition ได้ง่าย ๆ ไม่กี่คำสั่ง โดยอาศัย dlib ซึ่งเป็น machine learning ในการช่วยพัฒนาโมดูลนี้มีชื่อว่า face recognition โมดูล face recognition เป็นโมดูลที่ช่วยทำให้ Face Recognition เป็นเรื่องง่าย ๆ โดยมีความสามารถหลายอย่าง ไม่ว่าจะเป็น Face Recognition , ตกแต่งหน้าตาในรูปภาพ เป็นต้น และยังสามารถนำไปทำ Face Recognition แบบ Realtime ได้อีกด้วย

- ใช้ MIT License สามารถนำไปใช้พัฒนาในโปรแกรมเพื่อการค้าได้
- รองรับทั้ง Python 2 และ Python 3 [2]

ปฏิวัติ อิงคสันตติกุล [3] ภาควิชาฟิสิกส์ประยุกต์ คณะวิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยเชียงใหม่ เป็นโครงงาน ที่จัดทำขึ้นเพื่อการพัฒนาระบบรู้จำใบหน้าบุคคลซึ่ง ประกอบไปด้วยการค้นหาตำแหน่งของภาพใบหน้าและการรู้จำภาพหน้า ตรงการค้นหาตำแหน่งใบหน้าทำโดยค้นหาโครงหน้าด้วยการใช้วงรีไปวัดความคล้ายกับโครงหน้าของภาพใบหน้าการ ค้นหา ตำแหน่งตาดำอาศัยการปรับค่าความสว่างของภาพจนกระทั่งบริเวณกลางๆของภาพใบหน้าเหลือแต่ภาพจุด

Computer Science Department
Faculty of Informatics, Mahasarakham University

ของตาดำจากนั้น จึงใช้ค่ามาตรฐานสำหรับใบหน้าคนเพื่อคำนวณตำแหน่งของจมูกและปากต่อไป ส่วนการรู้จำภาพใบหน้าได้ใช้ค่าพารามิเตอร์ที่ได้จากโครงหน้าและตำแหน่งของตาดำปากและจมูก นำไปเปรียบเทียบกับการวิเคราะห์องค์ประกอบหลักของภาพใบหน้า และการวิเคราะห์ฟาสฟูเรียร์ ทรานฟอร์มเพื่อให้สามารถจำแนกได้ว่าใบหน้านั้น ๆ มีอยู่ในฐานข้อมูลเดิมหรือไม่

กฤติกา ศรีพงศ์สุข ณัฏฐา ปัญญาพูนตระกูล และธนาวุฒิ โชติชนาภิบาล [4] ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ เป็นโครงงานที่จัดทำขึ้นมาเพื่อใช้ในรู้จำใบหน้าบุคคลที่ทำงานบน คอมพิวเตอร์ส่วนบุคคลที่มีกล้องเว็บแคมต่ออยู่เทคนิคการรู้จำที่ใช้ในระบบนี้คือเทคนิค Eigen face ซึ่งการทำงานของระบบ สามารถแบ่งได้เป็น 2 ส่วน คือ ขั้นตอนการเรียนรู้ซึ่งจะนำภาพใบหน้าของบุคคลที่ต้องการจะรู้จำมาทำการวิเคราะห์องค์ประกอบหลักและขั้นตอนการรู้จำซึ่งจะวิเคราะห์ภาพใบหน้าทดสอบของบุคคลหนึ่งๆ เพื่อหาว่าตรงกับภาพใบหน้าใดที่ได้ เก็บไว้ในขั้นตอนการเรียนรู้หรือไม่

ทฤษฎี Eigen face (ไอเกนเฟซ) ไอเกนเฟซ (en: Eigenface) เป็นชื่อเรียกเซตของ ไอเกนเวกเตอร์ ซึ่งใช้ในระบบ การรู้จำใบหน้า ตัวกลุ่มของเวกเตอร์นี้ พัฒนาขึ้นโดย สิโรวิชย์ และ เคอร์บี้ ในปีพ.ศ. 2530 ถูกนำมาใช้แยกแยะลักษณะใน หน้ามนุษย์เป็นครั้งแรกโดย แมทธิว เตริก และ อเล็กซ์ เพนท์แลนด์รูปแบบการใช้งานทำโดยเปรียบเทียบลักษณะของภาพกับ เวกเตอร์ในเบสิคเซต ไอเกนเฟซ เป็นชื่อที่รู้จักกันดีกว่า ไอเกนพิกเจอร์ แต่โดยพื้นฐานล้วนมาจาก วิธีการของ สิโรวิชย์ และ เคอร์บี้ ในทฤษฎีเรื่องการวิเคราะห์ส่วนประกอบ หรือที่เรียกโดยย่อว่า PCA ยกตัวอย่างเช่น หน้าที่ 1 เมื่อเปรียบเทียบกับเบสิค เซต มีความเหมือนกับ ไอเกนเฟซ 1-10เปอร์เซนต์ เหมือนกับ ไอเกนเฟซ 2-55 เปอร์เซนต์ เหมือนกับ ไอเกนเฟซ 3 ติดลบ 3 เปอร์เซนต์ เมื่อนำหน้าที่ 2 มาเปรียบเทียบ แล้วได้สัดส่วนของเปอร์เซ็นต์ในทิศทางเดียวกันนี้ ก็ถือว่า หน้าที่ 1 กับ หน้าที่ 2 นั้นเป็นหน้าเดียวกัน ยกตัวอย่างว่ามนุษย์เราทำ face recognition อย่างไร แล้วค่อยเปรียบเทียบว่าเราจะใช้คอมพ์ทำ face recognition ได้อย่างไร สำหรับมนุษย์เราเวลาเราจะหาว่าคนที่เราพึ่งเจอในบริษัท PAKGON ชื่ออะไร (สมมุติว่าเรามีภาพของ คนๆนั้น และสมมุติให้ภาพคนๆนั้นเป็น X0) เราก็จะนำภาพคนๆนั้นมาเทียบกับสมุดบัญชีรายชื่อบริษัท PAKGON ที่มีทั้งชื่อ และหน้าคน (สมมุติว่า มีพนักงานอยู่ 100 คนในฐานข้อมูล สมมุติให้เป็น X1, X2, ..., X100) สำหรับมนุษย์เรา เราก็จะนำ X0 มาเทียบกับ X1 แล้วเก็บความเหมือนและความต่างไว้ในใจ แล้วเราก็นำ X0 มาเทียบกับ X2 ถ้าความเหมือนและต่างน้อยกว่า X0 และ X1 เราก็ เก็บ X0 และ X2 ไว้ในใจ แล้วทำ X0 และ X3 ต่อไป เราทำไปจนถึงคนที่ 100 เราก็จะได้ภาพที่ใกล้เคียง กับ X0 มากที่สุด (สมมุติว่าเป็น X10) และสรุปได้ว่า X0 คือ X10 วิธีการทำในคอมพิวเตอร์ก็คล้ายๆกัน เพียงแต่ เวลา เปรียบเทียบระหว่าง X0

Computer Science Department
Faculty of Informatics, Mahasarakham University

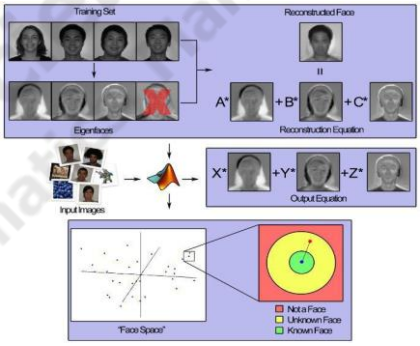
และ X1 แทนที่จะใช้เม็ดสี(pixels) นักวิจัยได้คิดค้นตัวแปรใหม่ขึ้นมาชื่อ Eigen face ซึ่งมีความเสถียรมากกว่าและเร็วกว่าการใช้เม็ดสี สำหรับรายละเอียดแต่ละ step จะเป็นดังนี้

- Align face ก็จะคล้ายๆกับเราเอากระดาษ A4 หลายๆแผ่นมาเรียงซ้อนกัน แล้วเราก็จับมันเขย่าๆ เพื่อให้มันอยู่ในแนวเดียวกัน เราจะได้เปรียบเทียบเอกสารได้ง่ายขึ้น (สมมุติว่าเอกสารหนึ่งชิ้นคือกระดาษ A4 หนึ่งแผ่น)

- คำนวณ dot product ระหว่าง Eigen-vector กับข้อมูลหน้าคน Dot product วิธีก็ง่ายๆ เช่น ถ้าเรามี vector อยู่ สอง vector เช่น (3, 1, 2) และ (2, 4, 6) วิธีคำนวณ dot product ก็จะเป็น 3×2 บวก 1×4 บวน 2×6 ได้เป็น 6 + 4 + 12 = 22 ในตัวอย่างข้างต้น เราสามารถเปรียบ (3, 1, 2) เป็น หน้าคน และ (2, 4, 6) เป็น Eigen-vector

- หา Euclidean distance เราใช้หลักการคล้ายๆกันระหว่างหน้าคนที่ทดสอบ กับ หน้าคนในฐานข้อมูล

- โปรแกรมจะคืนชื่อคนที่มีระยะทาง Euclidean distance น้อยที่สุด



ภาพประกอบที่ 2.10 การจดจำใบหน้าโดยใช้ Eigen faces

ที่มา : [5] A. S. Tolba, A. H. El-Baz, and A. A. El-Harby. (2005). Face recognition: a literature review. International Journal of Signal Processing, 2(2), (88-103).

Computer Science Department
Faculty of Informatics Mahasarakham University