

จากวิธีดั้งเดิมสู่ปัญญาประดิษฐ์



สมหญิง ไทยนิมิต

การจำแนกภาพ

จากวิธีดั้งเดิมสู่ปัญญาประดิษฐ์

Image Classification

From Traditional Approach to Modern AI

สมหญิง ไทยนิมิต

พิมพ์ครั้งแรก กรกฎาคม พุทธศักราช 2568

จำนวนหน้า 354 หน้า ราคา 499 บาท

ลิขสิทธิ์ของ รองศาสตราจารย์สมหญิง ไทยนิมิต

ห้ามคัดลอกเนื้อหาก่อนได้รับอนุญาต

ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

สมหญิง ไทยนิมิต.

การจำแนกภาพ: จากวิธีดั้งเดิมสู่ปัญญาประดิษฐ์

Image Classification: From Traditional Approach to Modern AI

กรุงเทพฯ : [ม.ป.พ.], 2025.

354 หน้า.

1. ปัญญาประดิษฐ์ -- โปรแกรมคอมพิวเตอร์. 2. ปัญญาประดิษฐ์ -- การประมวลผลข้อมูล.
I. ชื่อเรื่อง.

006.3

ISBN e-book 978-616-626-259-9

จัดพิมพ์โดย สมหญิง ไทยนิมิต

ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กทม.

พิมพ์ที่ โรงพิมพ์ดิจิทัล Hi-Speed Laser Print

50 ซอย สุวรรณวาทกิจ ลาดยาว จตุจักร กทม. 10900

โทรศัพท์ 0633299879

อีเมล hispeedlaserprint2014@gmail.com

คำนำ

การจำแนกภาพเป็นกระบวนการสำคัญในงานประมวลผลภาพและคอมพิวเตอร์วิทัศน์ ที่สามารถนำไปประยุกต์ใช้งานได้หลากหลายทั้งในอุตสาหกรรมและชีวิตประจำวัน เทคนิคการจำแนกภาพแบบวิธีดั้งเดิมอาศัยการสกัดคุณลักษณะจากภาพ ขั้นตอนการสกัดใช้ความเข้าใจภาพและงานที่ต้องนำการจำแนกไปประยุกต์ใช้ จากนั้นนำคุณลักษณะที่สกัดได้ผ่านขั้นตอนการจำแนก ขั้นตอนทั้งสองมีการเรียนรู้ที่แยกจากกัน ส่วนเทคนิคการจำแนกภาพในปัจจุบันนิยมใช้โครงข่ายคอนโวลูชันเชิงลึกที่สามารถออกแบบให้การเรียนรู้ของขั้นตอนการสกัดคุณลักษณะและขั้นตอนการจำแนกเป็นการเรียนรู้แบบตั้งแต่ต้นจนจบ การเรียนรู้ของทั้งสองขั้นตอนไม่แยกขาดจากกัน ช่วยเพิ่มประสิทธิภาพการเรียนรู้ แต่การทำความเข้าใจกลไกการทำงานรวมทั้งแนวคิดการออกแบบโครงข่ายเชิงลึกสำหรับผู้เรียนเริ่มต้นทำได้ยาก การเรียนรู้เทคนิคการจำแนกภาพแบบดั้งเดิมจึงยังสำคัญ ช่วยผู้เรียนเข้าใจแนวคิดการออกแบบและการทำงานของโครงข่ายเชิงลึกได้ง่ายขึ้น เนื่องจากแนวคิดในการออกแบบของโครงข่ายภายในโครงข่ายมีรากฐานมาจากเทคนิคแบบดั้งเดิม หนังสือเล่มนี้แบ่งเนื้อหาออกเป็นสี่ส่วนดังนี้

ความรู้พื้นฐานการประมวลผลภาพ ประกอบด้วยบทที่ 1 2 และ 3 หนังสือเริ่มต้นด้วยการแนะนำคำนิยามจำกัดความของศัพท์ที่เกี่ยวข้องกับการจำแนกภาพ เนื่องจากการจำแนกภาพเป็นขบวนการสำคัญในงานอื่น ๆ เช่น งานแบ่งส่วนภาพ งานตรวจจับภาพ ทำให้มักเกิดความสับสน บทที่ 1 ครอบคลุมนิยาม บอกความเหมือนและแตกต่างของศัพท์ที่เกี่ยวข้องเหล่านี้ รวมทั้งอธิบายภาพรวมใหญ่ของขั้นตอนการจำแนกภาพ ความแตกต่างระหว่างเทคนิคการจำแนกภาพแบบดั้งเดิมและแบบโครงข่ายคอนโวลูชัน ในบทที่ 2 กล่าวถึงภาพดิจิทัลทั้งภาพสีและภาพเทา การแปลงภาพระหว่างภาพสีและภาพเทา อธิบายคุณสมบัติพื้นฐานที่สำคัญของภาพคือฮิสโทแกรมและการปรับเปลี่ยนฮิสโทแกรมภาพ การตัดแบ่งภาพโดยอาศัยฮิสโทแกรม รวมทั้งครอบคลุมกลไกสำคัญ การทำคอนโวลูชันซึ่งเป็นกลไกหลักในการกรองสัญญาณจากภาพและยังเป็นกลไกหลักของโครงข่ายคอนโวลูชันอีกด้วย บทที่ 3 เริ่มต้นด้วยการอธิบายคุณลักษณะพื้นฐานสำคัญของภาพ ได้แก่ สี ขอบภาพ พื้นผิว คุณลักษณะเหล่านี้ถูกนำไปใช้ทั้งในการปรับปรุงคุณภาพภาพ และใช้เป็นคุณลักษณะสำคัญการจำแนกภาพและวัตถุภายในภาพ คุณลักษณะเหล่านี้ถูกนำมาต่อยอดเป็นเวกเตอร์อธิบายคุณลักษณะที่ใช้แทนภาพหรือวัตถุได้ดีขึ้น ทนทานต่อสัญญาณรบกวนต่าง ๆ ได้ดีขึ้น

การจำแนกภาพแบบดั้งเดิม แยกขบวนการได้เป็นการสกัดคุณลักษณะและตัวจำแนก การสกัดคุณลักษณะครอบคลุมเนื้อหาในบทที่ 3 บทที่ 4 อธิบายตัวจำแนกที่นิยมใช้ในเทคนิคการจำแนกแบบดั้งเดิมและการเรียนรู้ของตัวจำแนก ได้แก่ โครงข่ายประสาทเทียม ชัฟฟลอร์ดเวกเตอร์แมชชีน เพื่อนบ้านใกล้สุด K ตัว โดยโครงข่ายประสาทเทียมถูกนำไปใช้ทั้งในงานจำแนกภาพแบบดั้งเดิมและใช้เป็นส่วนหนึ่งในโครงข่ายเชิงลึก ตัวอย่างงานจำแนกภาพแบบดั้งเดิมอธิบายในหัวข้อ 4.6 7.6 และ 8.1

การจำแนกภาพแบบโครงข่ายคอนโวลูชัน ประกอบด้วยเนื้อหา บทที่ 5 และ 6 และ 7 บทที่ 5 ครอบคลุมแนวคิด การออกแบบและตัวอย่างโครงข่ายคอนโวลูชันที่นิยมใช้ การเรียนรู้ และอัลกอริทึมหาค่าเหมาะที่สุดที่นิยมใช้ในโครงข่ายคอนโวลูชัน บทที่ 6 อธิบายมาตรวัด การแบ่งข้อมูลตัวอย่าง เพื่อให้การวัดผลมีประสิทธิภาพ รวมทั้งอธิบายวิธีแก้ปัญหาโครงข่ายเรียนรู้เกิน และเทคนิคการส่งผ่านความรู้ที่นำโครงข่ายที่ผ่านการเรียนรู้มาก่อน มาประยุกต์ใช้กับงานใหม่ที่แตกต่างจากงานเดิม

การประยุกต์ใช้งาน การประยุกต์ใช้งานจำแนกภาพที่สำคัญงานหนึ่ง คือใช้ในงานตรวจจับวัตถุ เนื้อหาบทที่ 8 อธิบายแนวคิดการออกแบบ การประยุกต์ใช้การจำแนกในงานตรวจจับวัตถุ ฐานข้อมูลสาธารณะที่นิยมใช้ การปรับปรุงฟังก์ชันสูญเสียให้เหมาะกับงานตรวจจับวัตถุ รวมทั้งโครงข่ายคอนโวลูชันที่นิยมใช้ในการตรวจจับวัตถุ ส่วนบทที่ 9 ต่อยอดแนวคิดงานจำแนกภาพเป็นการวัดความคล้ายของภาพผ่านเวกเตอร์เอ็มเบดดิ้ง ซึ่งมีแนวคิดคล้ายกับเวกเตอร์คำอธิบายคุณลักษณะ แตกต่างกันในส่วนการเน้นให้โครงข่ายเรียนรู้ให้เอ็มเบดดิ้งของภาพเดียวกันหรือของวัตถุเดียวกันถูกส่งต่อไปยังปริภูมิเวกเตอร์บริเวณใกล้เคียงกัน นอกจากนั้นยังต้องอยู่ห่างจากเวกเตอร์เอ็มเบดดิ้งของวัตถุต่างชนิดกัน เนื้อหาในบทที่ 8 และ 9 แสดงให้เห็นถึงการนำงานจำแนกภาพไปประยุกต์ใช้งานในลักษณะที่แตกต่างกัน ทำให้ผู้อ่านได้เห็นแนวทางการออกแบบโครงข่ายคอนโวลูชันและฟังก์ชันสูญเสียสำหรับการนำการจำแนกภาพไปประยุกต์ใช้งานต่าง ๆ

หนังสือเล่มนี้อธิบายพื้นฐานงานจำแนกภาพ ซึ่งเป็นงานที่มีการพัฒนาอย่างต่อเนื่อง และการพัฒนาในปัจจุบันเป็นไปอย่างรวดเร็วโดยเฉพาะในส่วนเทคนิคการจำแนกภาพแบบโครงข่ายเชิงลึก หนังสือเรียบเรียงขึ้นเพื่อใช้ประกอบการเรียนการสอนวิชาการประมวลผลภาพดิจิทัล การวิเคราะห์ภาพและการรู้จำ การเรียนรู้ของคอมพิวเตอร์สำหรับการประยุกต์ใช้งานด้านภาพ เหมาะใช้สอนกับนิสิตวิศวกรรมไฟฟ้าชั้นปี 4 หรือนิสิตปริญญาโท โดยผู้เขียนได้จัดทำโปรแกรมภาษาไพธอนสำหรับให้นิสิตเรียนรู้งานจำแนกภาพในภาคปฏิบัติ นอกจากนั้นหากผู้อ่านพบข้อผิดพลาดหรือมีข้อเสนอแนะเพิ่มเติมในการปรับปรุงหนังสือ กรุณาแจ้งมาที่ sying.thai@gmail.com ผู้เขียนยินดีรับฟังและแลกเปลี่ยนความคิดเห็นเพื่อยกระดับคุณภาพของหนังสือให้ดียิ่งขึ้น

ท้ายสุดผู้เขียนขอขอบคุณนักวิจัยร่วม นิสิต ผู้ทรงคุณวุฒิ และผู้อ่านทุกท่านที่ให้คำวิจารณ์ การติชม และข้อเสนอแนะที่ช่วยในการปรับปรุงงานนี้ ขอขอบคุณอาจารย์ ผู้ร่วมงาน และผู้สร้างสรรค์ที่ได้รวบรวมและถ่ายทอดองค์ความรู้ รวมถึงสร้างแรงบันดาลใจที่ช่วยเติมเต็มสิ่งแวดลอมการเรียนรู้ให้เต็มไปด้วยชีวิตชีวา ทั้งนี้ทำให้ผู้เขียนได้เรียนรู้ศักยภาพที่ไม่มีที่สิ้นสุดของมนุษย์ สุดท้าย ขอขอบคุณครอบครัว และเพื่อน ๆ สำหรับ ความเกื้อหนุน และการสนับสนุนอย่างต่อเนื่อง ซึ่งความอบอุ่นหลากหลายเหล่านี้ คือแรงผลักดันสำคัญที่ช่วยให้ผู้เขียนก้าวเดินได้อย่างมั่นคง

สมหญิง ไทยนิมิต

สารบัญ

คำนำ	I
สารบัญ	III
สารบัญตาราง	XVII
บทที่ 1: บทนำ	1
1.1 นิยามศัพท์ที่เกี่ยวข้องกับการจำแนก	2
1.2 การประยุกต์ใช้งาน	5
1.3 ความท้าทาย	6
1.4 ภาพรวม	9
1.4 กรณีศึกษา: ระบบคัดแยกผักผลไม้แบบอัตโนมัติ	15
บทที่ 2: ภาพดิจิทัลและการประมวลผลภาพ	21
2.1 ภาพดิจิทัล	21
2.2 ชนิดของภาพ	24
2.3 โมเดลสี	27
2.4 การแปลงภาพสีเป็นภาพเทา	30
2.5 การแปลงโมเดลสี	31
2.6 ฮิสโทแกรม	32
2.7 การแปลงระดับเทา	33
2.8 ฮิสโทแกรมอีควอไลเซชัน	34
2.9 การตัดแบ่ง	37
2.10 การตัดแบ่งสี	38
2.11 การกรองและคอนโวลูชัน	40
2.12 สังเกตแบบแยกกันได้	49
2.13 สหสัมพันธ์	51
2.13 สหสัมพันธ์	52
บทที่ 3: การสกัดคุณลักษณะและคำอธิบายภาพ	59
3.1 พื้นฐานการสกัดคุณลักษณะและการแทนภาพ	59
3.2 คุณลักษณะและการสกัดคุณลักษณะ	63

3.3 คำอธิบายคุณลักษณะ	89
3.4 การเปรียบเทียบสเกลคุณลักษณะ	96
3.5 ข้อจำกัดและการพัฒนาต่อ	97
บทที่ 4: ตัวจำแนก	103
4.1 ประเภทตัวจำแนก	104
4.2 ภาพรวมการเรียนรู้แบบมีผู้สอน	106
4.3 โครงข่ายประสาทเทียม	107
4.4 ซัพพอร์ตเวกเตอร์แมชชีน	130
4.5 เพื่อนบ้านใกล้สุด K ตัว	136
4.6 กรณีสึกษา: การจำแนกตัวอักษรบนป้ายถนน	139
บทที่ 5: โครงข่ายประสาทเทียมแบบคอนโวลูชัน	149
5.1 ชั้นพื้นฐานในโครงข่ายคอนโวลูชัน	150
5.2 ตัวอย่างสถาปัตยกรรมโครงข่ายคอนโวลูชัน	157
5.3 จำนวนพารามิเตอร์ที่ต้องเรียนรู้	171
5.4 การกำหนดอัตราการเรียนรู้แบบต่าง ๆ	173
5.5 อัลกอริทึมหาค่าเหมาะสมขั้นสูง	176
บทที่ 6: การวัดและเพิ่มประสิทธิภาพโมเดล	187
6.1 ตัวชี้วัดประสิทธิภาพการจำแนกข้อมูล	187
6.2 การเตรียมและแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพของโมเดล	194
6.3 การวิเคราะห์ปัญหาการเรียนรู้ไม่พอและท่องจำชุดฝึก	198
6.4 แนวทางเพิ่มประสิทธิภาพการเรียนรู้ของเครื่อง	200
6.5 การทำแบตช์นอร์มัลไลเซชัน	202
6.6 การกำหนดค่าพารามิเตอร์ตั้งต้น	204
6.7 แนวทางการแก้ปัญหาท่องจำชุดฝึก	207
6.8 หยุดล่วงหน้า	214
6.9 การส่งผ่านความรู้	214
บทที่ 7: การจำแนกภาพ	225
7.1 การจำแนกภาพและประเภทงานจำแนกภาพ	225
7.2 ฟังก์ชันสูญเสียสำหรับงานจำแนกข้อมูล	228
7.3 การจำแนกข้อมูลทวิภาคด้วยการถดถอยโลจิสติก	234

7.4 การจำแนกข้อมูลหลายคลาสด้วยการถดถอยโลจิสติก	239
7.5 การจำแนกโรคต่อหินจากภาพถ่ายจอประสาทตา	240
7.6 การจำแนกโรคต่อหินด้วยคุณลักษณะแบบดั้งเดิม	246
7.7 การจำแนกภาพด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน	249
บทที่ 8: การตรวจจับสนุ	261
8.1 การตรวจจับสนุแบบดั้งเดิม	262
8.2 การตรวจจับสนุด้วยโครงข่ายคอนโวลูชัน	264
8.3. ฐานข้อมูล	270
8.4 การจำแนกประเภทอัลกอริทึมตรวจจับสนุ	273
8.5 โครงข่ายคอนโวลูชัน R-CNN	274
8.6 โครงข่าย FAST R-CNN และ FASTER R-CNN	276
8.7 โครงข่าย YOLO	280
8.8 การประยุกต์ใช้งานตรวจจับสนุ	292
บทที่ 9: การเรียนรู้ความคล้ายและอิมเมจเอ็มเบดดิ้ง	301
9.1 อิมเบดดิ้ง	303
9.2 ขั้นตอนการสร้างอิมเมจเอ็มเบดดิ้งด้วยโครงข่ายคอนโวลูชัน	304
9.3 ฟังก์ชันสูญเสีย	306
9.4 สถาปัตยกรรม	309
9.5 การประยุกต์ใช้งาน	314
ภาคผนวก ก.	327

สารบัญรูป

รูปที่ 1. 1 (ก) การจำแนกภาพ (IMAGE CLASSIFICATION) (ข) การตรวจจับวัตถุ (OBJECT DETECTION) (ค) การแบ่งส่วนภาพเชิงความหมาย (SEMANTIC SEGMENTATION) (ง) การแบ่งส่วนภาพเชิงรายตัวอย่าง (INSTANCE SEGMENTATION).....	3
รูปที่ 1. 2 ปัญหาและความท้าทายที่พบบ่อยในงานจำแนกภาพและตรวจจับวัตถุ.....	7
รูปที่ 1. 3 แสดงภาพแม่แบบหรือเทมเพลตและภาพอินพุตที่ใช้ในการจับคู่เทมเพลต	9
รูปที่ 1. 4 บล็อกไดอะแกรมการจำแนกข้อมูลแบบดั้งเดิม ซึ่งคุณลักษณะที่สกัดออกแบบเบื้องต้นโดยมนุษย์.....	10
รูปที่ 1. 5 ชั้นของนิเวศในโครงข่ายประสาทเทียม NN	11
รูปที่ 1. 6 การเรียนรู้แบบมีผู้ฝึกสอน (SUPERVISED LEARNING)	12
รูปที่ 1. 7 บล็อกไดอะแกรมการจำแนกภาพแบบโครงข่ายคอนโวลูชันเชิงลึก	15
รูปที่ 2. 1 การตรวจจับภาพดิจิทัล (IMAGE ACQUISITION).....	21
รูปที่ 2. 2 การอ้างอิงภาพดิจิทัลที่แต่ละพิกเซลเก็บค่าความเข้มแสงมีค่าระหว่าง 0 ถึง 255 จัดเก็บด้วยข้อมูลอาร์เรย์ 2 มิติ ขนาดเท่ากับ จำนวนแถว x จำนวนหลัก	22
รูปที่ 2. 3 พิกเซลเพื่อนบ้านรอบจุด (R, C) (ก) แบบ 4- จุดเพื่อนบ้าน และ (ข) แบบ 8- จุดเพื่อนบ้าน... ..	24
รูปที่ 2. 4 ภาพไบนารี ค่าความเข้มแสงแต่ละจุดภาพมีค่าได้สองค่า คือ 0 หรือ 1	24
รูปที่ 2. 5 ภาพสีจริง จัดเก็บค่าความเข้มแสงแต่ละองค์ประกอบภาพ เช่น ภาพสี RGB ประกอบไปด้วยองค์ประกอบภาพสีแดง เขียว และน้ำเงิน.....	25
รูปที่ 2. 6 ภาพสีอินเด็กซ์ แต่ละพิกเซลเก็บค่าดัชนีของตารางสี (COLOR MAP) ที่ใช้.....	26
รูปที่ 2. 7 การอ่านภาพ ด้วยไลบรารี OPENCV ไพธอน.....	26
รูปที่ 2. 8 (ก) การผสมสีแบบบวกของแสง (ข) การผสมสีแบบลบของผงสี.....	27
รูปที่ 2. 9 โมเดลสี RGB แทนได้ด้วยจุดบนสีเหลี่ยมลูกบาศก์ขนาด 1 หน่วย.....	28
รูปที่ 2. 10 (ก) ภาพต้นฉบับ (ข) ตารางแจกแจงความถี่ (ค) ฮิสโทแกรมภาพ	32
รูปที่ 2. 11 (ก) ภาพต้นฉบับ (ข) ภาพเทา (ค) ภาพลบของภาพเทา (ง) ภาพที่ได้จากการรวมภาพลบขององค์ประกอบภาพสี RGB (ค) ภาพลบที่เกิดจากการทำภาพลบเฉพาะกับองค์ประกอบแสง (V) ของภาพสี HSV ของภาพต้นฉบับแล้วนำมาผสมรวมกับองค์ประกอบ S และ V เดิม	34
รูปที่ 2. 12 การทำฮิสโทแกรมฮิสโตแกรม	36
รูปที่ 2. 13 (ก) ภาพต้นฉบับ และฮิสโทแกรมของภาพ (ข) ภาพผลลัพธ์ที่ได้จากการทำ ฮิสโทแกรมฮิสโตแกรมกับภาพต้นฉบับและฮิสโทแกรมของภาพที่ค่าความสว่างมีการกระจายตัวที่	

ใกล้เคียงเส้นตรงมากขึ้น ทำให้เห็นลายภายในม่านตาชัดเจนขึ้น (ภาพถ่ายต้นฉบับได้จาก ฐานข้อมูลสาธารณะ [3]).....	36
รูปที่ 2. 14 (ก) ภาพถ่ายตา (ข) ฮิสโทแกรมภาพ (ค) ผลการตัดแบ่งที่ค่าขีดแบ่ง 80 (พิมพ์ซ้ำจาก [2])	38
รูปที่ 2. 15 (ก) ภาพต้นฉบับ (ข) ตัวอย่างสูตรสำหรับหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของสีที่ สนใจ (ค) ผลการตัดแยกสีองค์ประกอบ H ในโดเมน HSV (ง) ผลการตัดแยกสีที่ละองค์ประกอบ สี โมเดล RGB (พิมพ์ซ้ำจาก [2])	40
รูปที่ 2. 16 การอ้างอิงตำแหน่งค่าสัมประสิทธิ์ของตัวกรองขนาด 3×3 และ 5×5	41
รูปที่ 2. 17 การทำคอนโวลูชัน	42
รูปที่ 2. 18 การเสริมเต็มภาพแบบ (ก) ZERO PADDING (ข) REPLICATION PADDING (ค) SYMMETRY PADDING (ง) CIRCULAR PADDING	43
รูปที่ 2. 19 กราฟด้านซ้ายแสดงค่าความเข้มแสงของทุกจุดภาพในแถวที่ 10 ของภาพตา กราฟ ด้านขวาแสดงค่าความเข้มแสงของทุกจุดภาพในแถวที่ 90 ของภาพตา.....	44
รูปที่ 2. 20 ตัวกรองค่าเฉลี่ยขนาด 3×3	45
รูปที่ 2. 21 การกรองสัญญาณความถี่ต่ำด้วยตัวกรองค่าเฉลี่ยขนาด 3×3 และ 9×9	46
รูปที่ 2. 22 ฟังก์ชันเกาส์เซียน (ซ้าย) $\sigma = 3$ (ขวา) $\sigma = 6$	46
รูปที่ 2. 23 ตัวกรองเกาส์เซียนขนาด 3×3 และ 5×5	47
รูปที่ 2. 24 การกรองสัญญาณความถี่ต่ำด้วยตัวกรองเกาส์เซียนมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 และ 3.....	47
รูปที่ 2. 25 (ซ้าย) ภาพต้นฉบับที่มีสัญญาณรบกวนประเภทเกลือและพริกไทย (กลาง) ภาพหลังผ่านตัว กรองมัธยฐาน (ขวา) ภาพหลังผ่านตัวกรองเกาส์เซียน.....	48
รูปที่ 2. 26 ตัวกรองขอบภาพแบบต่าง ๆ (ที่มา: [2]).....	48
รูปที่ 2. 27 (ซ้าย) ฟังก์ชัน LOG (ขวา) ตัวกรอง LOG ขนาด 5×5	50
รูปที่ 2. 28 บล็อกไดอะแกรมการกรองด้วยตัวกรอง LOG แบบแยกได้	50
รูปที่ 2. 29 เปรียบเทียบการทำคอนโวลูชันกับคอร์เรเลชัน	51
รูปที่ 2. 30 การตรวจจับตัวอักษร N สีขาวด้วยวิธีจับคู่เทมเพลต	53
รูปที่ 3. 1 คุณลักษณะแบบครอบคลุมและแบบเฉพาะที่.....	61
รูปที่ 3. 2 คุณลักษณะที่ใช้ในการจำแนกข้อมูลควรเป็นคุณลักษณะที่มีการเกาะกลุ่มกัน (มีค่าความ แปรปรวนต่ำ) และแต่ละกลุ่มไม่มีการทับซ้อนกัน ตามแผนภาพด้านซ้ายมือ ส่วนแผนภาพ ด้านขวาแสดงคุณลักษณะที่มีการกระจายตัวทับซ้อนกันระหว่างคลาส ทำให้เกิดความผิดพลาด ในการจำแนกข้อมูล	63

รูปที่ 3.3 แผนผังการกระจายค่าสีผิวมนุษย์เชื้อชาติต่าง ๆ สีชาติ (ภาพจาก [8] อนุญาตภายใต้สิทธิ์ CC BY 4.0).....	64
รูปที่ 3.4 ขอบภาพชนิดต่าง ๆ.....	68
รูปที่ 3.5 ตัวกรองเกรเดียนต์ชนิดต่าง ๆ.....	70
รูปที่ 3.6 ภาพต้นฉบับและภาพขอบที่ได้จากตัวกรอง SOBEL ขนาด 3×3	70
รูปที่ 3.7 ภาพขอบที่ได้จากตัวกรองโซเบล (g) ไม่ผ่านการตัดแบ่ง (ข) ผ่านการตัดแบ่งด้วยค่าขีดแบ่งเท่ากับ 20 (ค) ผ่านการตัดแบ่งด้วยค่าขีดแบ่งเท่ากับ 70	71
รูปที่ 3.8 แสดงค่าสัมประสิทธิ์ตัวกรองลาปลาเซียนแบบ 1 และ 2 มิติ	72
รูปที่ 3.9 อนุพันธ์อันดับหนึ่งและสองของสัญญาณขอบขาขึ้น.....	73
รูปที่ 3.10 ตัวกรองขอบภาพแบบต่าง ๆ	73
รูปที่ 3.11 สเปกตรัมและมุมที่ได้จากการแปลงภาพด้วยฟูเรียร์.....	77
รูปที่ 3.12 ตัวอย่างสัญญาณเวฟเลทที่นิยมใช้เป็นสัญญาณฐานหลักในการแยกองค์ประกอบสัญญาณใด ๆ	78
รูปที่ 3.13 เวฟเลทที่สเกลและตำแหน่งต่างๆ กัน.....	79
รูปที่ 3.14 ขั้นตอนการวิเคราะห์และสังเคราะห์สัญญาณแบบฟิลเตอร์แบงก์.....	80
รูปที่ 3.15 บล็อกไดอะแกรมแสดงขั้นตอนการวิเคราะห์และสังเคราะห์สัญญาณด้วยตัวกรองสองช่องสัญญาณแบบสองระดับ	81
รูปที่ 3.16 การแปลงเวฟเลทสัญญาณภาพ 2 มิติ.....	82
รูปที่ 3.17 การแปลงกลับเวฟเลทสัญญาณภาพ 2 มิติ.....	83
รูปที่ 3.18 (ก) ภาพต้นฉบับ (ข) การจัดเก็บภาพ (ค) ภาพผลลัพธ์จากการแปลงเวฟเลทสองระดับ (ภาพผลลัพธ์ผ่านการปรับความเปรียบต่างเพื่อการแสดงผล).....	83
รูปที่ 3.19 เมทริกซ์ระดับเทา	85
รูปที่ 3.20 ตัวอย่างการหาค่า LBP โดยใช้หน้ากากสี่เหลี่ยม	87
รูปที่ 3.21 หน้ากากวงกลมที่นิยมใช้ในการหาค่า LBP.....	88
รูปที่ 3.22 ฮิสโตแกรมทิศทางค่าเกรเดียนต์.....	90
รูปที่ 3.23 ขนาดของภาพมีผลทำให้มุมภาพสกัดได้เป็นเส้นขอบหรือเส้นโค้งเมื่อวัตถุมีขนาดใหญ่ขึ้น	92
รูปที่ 3.24 ภาพในปริภูมิขนาด แบบ 4 ระดับ 5 สเกล	92

รูปที่ 3. 25 ตัวอย่างภาพผลลัพธ์ที่ได้จากการหา DIFFERENCE OF GAUSSIAN ในปริภูมิขนาด โดยภาพในแต่ละ OCTAVE เกิดจากนำภาพไปคอนโวลูชันกับตัวกรองเกาส์เซียน ภาพ DOG เกิดจากการนำภาพที่ผ่านตัวกรองเกาส์เซียนที่มีสเกลติดกันมาลบกัน	93
รูปที่ 3. 26 การสร้างคำอธิบายคุณลักษณะจากจุดสำคัญใน SIFT [25]	94
รูปที่ 3. 27 แนวคิดการสร้างฮีโแกรมสำหรับจำแนกภาพด้วย BoVW	95
รูปที่ 4. 1 ขั้นตอนการจำแนกภาพ	103
รูปที่ 4. 2 แบบจำลองการเรียนรู้ด้วยเครื่อง ได้จากการเรียนรู้หาฟังก์ชันความสัมพันธ์ระหว่างข้อมูลตัวอย่าง ในกรณีนี้ได้สมการเส้นตรงสีฟ้าเป็นเส้นตัดสินใจจำแนกกระหว่างสองคลาส	104
รูปที่ 4. 3 การเรียนรู้แบบมีผู้สอน (SUPERVISED LEARNING)	106
รูปที่ 4. 4 (ก) โครงสร้างเซลล์ประสาทในสมองมนุษย์ (ข) แบบจำลองนิวรอนในโครงข่ายประสาทเทียม	108
รูปที่ 4. 5 โครงสร้างโครงข่ายประสาทเทียมประกอบด้วยชั้นอินพุต ชั้นซ่อน และชั้นเอาต์พุต	109
รูปที่ 4. 6 (ก) โครงข่ายแบบไปข้างหน้า (ข) โครงข่ายแบบป้อนกลับ	110
รูปที่ 4. 7 (ก) เพอร์เซปตรอน (ข) ฟังก์ชันกระตุ้นขั้นบันได (ขวา) เพอร์เซปตรอนเรียนรู้ผ่านชุดฝึกสอนเพื่อหาขอบเขตเส้นตัดสินใจ สำหรับจำแนกคลาสแบบไบนารี ถ้า $z > 0$ จำแนกเป็นคลาสบวก ไม่เช่นนั้นเป็นคลาสลบ	111
รูปที่ 4. 8 ตัวอย่างข้อมูล (ก) จำแนกแบบเชิงเส้น (LINEAR SEPARABLE) และ (ข) จำแนกแบบไม่เชิงเส้น (NON-LINEAR SEPARABLE)	112
รูปที่ 4. 9 โครงข่ายประสาทเทียม MLP ที่มีชั้นซ่อน 2 ชั้น	113
รูปที่ 4. 10 (ก) โครงข่ายประสาทเทียมแบบส่งไปข้างหน้ามีชั้นซ่อน 1 ชั้น (ข) ฟังก์ชันกระตุ้นรีลู	114
รูปที่ 4. 11 แสดงพัฒนาการของเส้นขอบเขตตัดสินใจในโครงข่ายแบบง่ายที่มีชั้นซ่อน 1 ชั้นมีนิวรอน 3 ตัว โดยใช้ฟังก์ชันกระตุ้นมีเรลู (ภาพดัดแปลงจาก [3])	115
รูปที่ 4. 12 แสดงการเพิ่มจำนวนเอาต์พุตโหนดในโครงข่ายแบบง่าย โดยแต่ละเอาต์พุตโหนดสร้างเส้นขอบเขตตัดสินใจที่แตกต่างกัน	116
รูปที่ 4. 13 แสดงการเพิ่มจำนวนอินพุตโหนดในโครงข่ายแบบง่าย ทำให้พื้นที่การตัดสินใจมีความซับซ้อนมากขึ้น เส้นขนานแสดงความชันของระนาบเกินในแต่ละบริเวณ (ภาพดัดแปลงจาก [3])	116
รูปที่ 4. 14 ฟังก์ชันกระตุ้น (สีฟ้า) และอนุพันธ์ของฟังก์ชันกระตุ้น (สีเหลือง)	118

รูปที่ 4. 15 ฟังก์ชันกระตุ้นซิกมอยด์และซอฟต์แวร์แมกซ์ให้ผลลัพธ์ค่าระหว่าง $[0,1]$ ข้อแตกต่างคือผลรวมค่าในเวกเตอร์เอาต์พุตของฟังก์ชันซอฟต์แวร์แมกซ์เท่ากับ 1 ในขณะที่ผลรวมค่าในเวกเตอร์เอาต์พุตของฟังก์ชันซิกมอยด์ไม่จำเป็นต้องมีค่าเท่ากับ 1	121
รูปที่ 4. 16 กราฟแสดงการลู่เข้าหาจุดต่ำสุดของฟังก์ชันสูญเสียค่าเฉลี่ยความผิดพลาดกำลังสอง..	123
รูปที่ 4. 17 การหาอนุพันธ์ย่อยด้วยกฎห่วงโซ่	125
รูปที่ 4. 18 (ก) การลู่เข้าหาจุดต่ำสุดสัมพัทธ์และจุดต่ำสุดสัมบูรณ์ (ข) ปัญหาจุดอานม้า.....	127
รูปที่ 4. 19 กำหนดค่าอัตราเรียนรู้มากเกินไปอาจทำให้เกิดปัญหาโมเดลไม่สามารถลู่เข้าหาจุดเหมาะสมได้	128
รูปที่ 4. 20 (ก) เส้นปะแสดงเส้นตัดสินใจหรือระนาบเกินที่ใช้จำแนกข้อมูล ซึ่งมีได้หลายเส้น (ข) SVM หาระนาบเกินที่เหมาะสมที่สุด (เส้น H_0) ซึ่งเป็นเส้นที่มีระยะขอบมากที่สุด (เส้นสีเขียว) อีกนัยหนึ่งคือห่างจากซัพพอร์ตเวกเตอร์มากที่สุด.....	131
รูปที่ 4. 21 ตัวจำแนก SVM แบบเชิงเส้นกรณีระยะขอบอ่อน (SOFT MARGIN) ที่รองรับการจำแนกผิดพลาดด้วยตัวแปรสแลค (SLACK)	133
รูปที่ 4. 22 ฟังก์ชันเคอร์เนลทำหน้าที่ส่งต่อข้อมูลไปยังปริภูมิที่สูงขึ้นทำให้สามารถจำแนกข้อมูลในปริภูมิไม่เชิงเส้นด้วยตัวจำแนกแบบเชิงเส้นได้.....	134
รูปที่ 4. 23 การจำแนกข้อมูลแบบหลายคลาสด้วย SVM (ก) แบบหนึ่ง-ต่อ-หนึ่ง (ข) และหนึ่ง-ต่อ-ที่เหลือ โดยสังเกตว่า ในวิธีหนึ่ง-ต่อ-หนึ่ง เส้นตัดสินใจพิจารณาเฉพาะข้อมูลสองคลาสในแต่ละคู่โดยไม่สนใจข้อมูลในคลาสอื่นๆ.....	135
รูปที่ 4. 24 เพื่อนบ้านใกล้สุด K-NN ถ้ากำหนด K เท่ากับ 1 ข้อมูลใหม่จะจัดอยู่ในคลาสสีแดง ถ้า K เท่ากับ 3 ข้อมูลใหม่จะจัดอยู่ในคลาสสีน้ำเงิน.....	137
รูปที่ 4. 25 แผนผังการไหลการตรวจจับและจำแนกตัวอักษรบนป้ายถนน [27]	139
รูปที่ 4. 26 การตรวจจับป้ายถนน (ก) ภาพอินพุต (ข) ภาพเอาต์พุตที่ตรวจจับได้.....	140
รูปที่ 4. 27 ผลการตัดแบ่งตัวอักษรบนป้ายถนนด้วยเทคนิค MSER [25]	141
รูปที่ 5. 1 โครงข่ายประสาทเทียมคอนโวลูชัน (CNNs) สถาปัตยกรรม LENET-5 [1] (สร้างรูปโครงข่ายด้วยโปรแกรม NN_SVG [3])	149
รูปที่ 5. 2 ภาพอินพุตหลังการเสริมเต็มด้วยศูนย์และตำแหน่งการเลื่อนหน้าจากตัวกรองในการทำคอนโวลูชันเมื่อกำหนดให้ $s = 1$ (แถวบน) และ $s = 2$ (แถวล่าง).....	152
รูปที่ 5. 3 การหาค่าสูงสุดและหาค่าเฉลี่ยด้วยหน้าต่างขนาด 3×3 และตัวอย่าง $s = 3$	154
รูปที่ 5. 4 ขนาดข้อมูลในชั้นหาค่าขึ้นอยู่กับค่าตัวอย่างและวิธีที่ใช้หาค่า	155

รูปที่ 5.5 กระบวนการแฟลทเทน (FLATTEN) ทำหน้าที่แปลงข้อมูลเมทริกซ์เป็นเวกเตอร์แถวเพื่อ ป้อนเข้าสู่ชั้นเดนซี้.....	156
รูปที่ 5.6 โมดูล INCEPTION (n) แบบพื้นฐาน (ข) แบบลดมิติ (DIMENSION REDUCTION)	163
รูปที่ 5.7 ตัวดำเนินการคอนโวลูชัน 1×1 พิจารณาพิกเซลในแนวลึก (แนวองค์ประกอบ) ของภาพ ผลลัพธ์ที่ได้จากตัวดำเนินการจะเป็นภาพที่มีจำนวนแถวและจำนวนหลักเท่ากับภาพอินพุต แต่ มีความลึกเท่ากับ 1	164
รูปที่ 5.8 (ก) สถาปัตยกรรม GOOGLENET หรือ INCEPTION V1 (ข) โมดูลตัวจำแนกเสริมในชั้น INCEPTION 4[A] และ INCEPTION 4[D] ใช้เฉพาะช่วงฝึกสอนโครงข่าย	165
รูปที่ 5.9 (ก) การเชื่อมต่อแบบทั่วไป (ข) การเชื่อมต่อแบบข้าม.....	167
รูปที่ 5.10 การออกแบบสถาปัตยกรรมแบบต่างๆ ที่เลือกใช้ชั้นสำหรับการสกัดคุณลักษณะที่แตกต่าง กัน (ด้านซ้าย) สถาปัตยกรรมแบบทั่วไป ที่ส่วนสกัดคุณลักษณะจะใช้ชั้นคอนโวลูชันเรียงต่อ สลับกับชั้นดีคค่า (รูปกลาง) สถาปัตยกรรมแบบอินเซปชัน ใช้โมดูลอินเซปชันเรียงต่อสลับกับชั้น ดีคค่า (รูปขวา) สถาปัตยกรรมเรสซิดวล ใช้บล็อกเรสซิดวลเรียงต่อกันไปในส่วนสกัดคุณลักษณะ โดยไม่นิยมใช้ชั้นดีคค่าสลับวางระหว่างบล็อกเรสซิดวล	167
รูปที่ 5.11 (ก) บล็อกเรสซิดวล (RESIDUAL BLOCK) (ข) บล็อกเรสซิดวลบอตเทิลเนค (RESIDUAL BOTTLENECK BLOCKS).....	168
รูปที่ 5.12 กราฟหมายกำหนดการอัตราเรียนรู้แบบต่าง ๆ.....	174
รูปที่ 6.1 ผลการทดลองโมเดลจำแนกแบบทวิภาคที่กำหนดค่าขีดแบ่งต่างกัน	188
รูปที่ 6.2 เมทริกซ์ความสับสน (ก) – (ข) งานจำแนกแบบทวิภาค (ค) งานจำแนกแบบหลายคลาส ..	189
รูปที่ 6.3 ตารางเมทริกซ์สับสนงานจำแนกข้อมูลแบบสามคลาส (ก) เมื่อลดรูปลงเป็นเมทริกซ์สับสน จำแนกสองคลาสระหว่างคลาสที่สนใจและคลาสอื่น ๆ จะได้เมทริกซ์สับสนเมื่อสนใจคลาส A ตามรูป (ข) และ เมทริกซ์สับสนเมื่อสนใจคลาส B ในรูป (ค).....	192
รูปที่ 6.4 (ซ้าย) กราฟ ROC และ (ขวา) พื้นที่ใต้กราฟ ROC	193
รูปที่ 6.5 การประเมินแบบไขว้ $K = 5$ กลุ่ม โดยแบ่งข้อมูลออกเป็น 5 กลุ่ม แต่ละรอบใช้ $K-1$ กลุ่ม สำหรับฝึกสอนและ 1 กลุ่มสำหรับประเมิน วนจนครบทุกกลุ่ม	197
รูปที่ 6.6 เส้นขอบเขตการตัดสินใจกรณีท้องจำชุดฝึก เรียนรู้ไม่พอ และกรณีครอบคลุมน้อยทั่วไป....	199
รูปที่ 6.7 ตัวอย่างกราฟเรียนรู้กรณีท้องจำชุดฝึก เรียนรู้ไม่พอ และกรณีครอบคลุมน้อยทั่วไป.....	200
รูปที่ 6.8 กราฟความสัมพันธ์ระหว่างค่าความสูญเสียกับจำนวนนิรอนในชั้นซ่อนของโครงข่าย ประสาทเทียมแบบตื้น	201
รูปที่ 6.9 แมตริกซ์นอร์มัลไลเซชันทำให้การกระจายตัวของข้อมูลอยู่ในรูปแบบเดียวกัน	203

รูปที่ 6. 10 เทคนิคเสริมข้อมูลด้วยวิธีประมวลผลภาพแบบต่าง ๆ.....	209
รูปที่ 6. 11 โครงข่าย GAN	210
รูปที่ 6. 12 แสดงแนวคิดหลักของเทคนิคครอปเอาต์.....	211
รูปที่ 6. 13 การส่งผ่านความรู้แบบใช้โมเดลฟรีเทรนเป็นตัวสกัดคุณลักษณะ	217
รูปที่ 6. 14 การส่งผ่านความรู้แบบปรับจูน	218
รูปที่ 7. 1 งานจำแนกภาพแบ่งตามจำนวนคลาสหรือเลเบลที่กำหนดให้ภาพ	226
รูปที่ 7. 2 กราฟค่าสูญเสียอิงจ้.....	231
รูปที่ 7. 3 ขั้นตอนการถอดถอยโลจิสติก	235
รูปที่ 7. 4 ฟังก์ชันสูญเสียครอสเอนโทปี.....	236
รูปที่ 7. 5 การหาค่าเกรเดียนต์การถอดถอยโลจิสติกด้วยการส่งต่อไปข้างหน้า การแพร่กระจาย ย้อนกลับและกฎห่วงโซ่.....	237
รูปที่ 7. 6 (ซ้าย) การถอดถอยโลจิสติกแบบทวิภาค (ขวา) การถอดถอยโลจิสติกแบบหลายคลาส	239
รูปที่ 7. 7 อัตราส่วนด้วยต่อจัน (CDR) ที่ใช้ประเมินโรคต้อหิน (ภาพตาจากฐานข้อมูลสาธารณะ DRIONS-DB [5]).....	242
รูปที่ 7. 8 ขั้นตอนหลักการจำแนกโรคต้อหินด้วยภาพถ่ายจอประสาทตา.....	242
รูปที่ 7. 9 ตัวอย่างภาพตาจากฐานข้อมูลสาธารณะต่าง ๆ.....	243
รูปที่ 7. 10 ภาพรวมแผนผังการไหลขั้นตอนการจำแนกโรคต้อหิน.....	244
รูปที่ 7. 11 แสดงตัวอย่างขั้นตอนการจำแนกโรคต้อหินโดยอาศัยคุณลักษณะทางกายภาพของข้อ ประสาทตาและตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนนำเสนอโดย [16].....	246
รูปที่ 7. 12 แสดงตัวอย่างแผนผังการไหลการจำแนกโรคต้อหินด้วยการผสมคุณลักษณะสาม คุณลักษณะ [18].....	248
รูปที่ 7. 13 โครงข่ายคอนโวลูชันสำหรับจำแนกโรคต้อหิน [19].....	250
รูปที่ 7. 14 การใช้โมเดลฟรีเทรนโครงข่ายคอนโวลูชันในการทำหน้าที่สกัดคุณลักษณะ [23].....	251
รูปที่ 8. 1 (ก) การตรวจจับวัตถุ (ข) การแบ่งข้อมูลภาพ.....	261
รูปที่ 8. 2 อัลกอริทึมตรวจจับบุคคลด้วยคำอธิบายภาพ HOG และตัวจำแนก SVM [2].....	263
รูปที่ 8. 3 (ซ้าย) บาวนดิ้งบ็อกซ์ที่ได้จากขั้นตอนสร้างพื้นที่นำเสนอซึ่งมีหลายกล่อง (ขวา) ผลลัพธ์ที่ได้ หลังจากขั้นตอนขจัดกล่องที่ซ้ำซ้อน	266
รูปที่ 8. 4 ค่า IOU (INTERSECTION OVER UNION)	268
รูปที่ 8. 5 กราฟ PRECISION-RECALL (PR) (ซ้าย) กราฟ PR ของวัตถุ 1 คลาสเมื่อกำหนด IOU ค่าเดียว (ขวา) เมื่อกำหนดค่าขีดแบ่ง IOU สูงขึ้น ค่าความแม่นยำเฉลี่ย (AP) ต่ำลง	269

รูปที่ 8. 6 แสดงตัวอย่างผลเฉลยภาพของฐานข้อมูล PASCAL VOC ปี 2012 [3].....	271
รูปที่ 8. 7 ตัวอย่างผลเฉลยการจำแนกภาพของฐานข้อมูล COCO ปี 2017 [6] (ใช้ภาพภายใต้ลิขสิทธิ์ ATtribution SHAREALike License) [5]	272
รูปที่ 8. 8 ตัวอย่างผลเฉลยการตรวจจับวัตถุของฐานข้อมูล OPEN IMAGES V7 [8].....	273
รูปที่ 8. 9 ภาพรวมอัลกอริทึมตรวจจับวัตถุแบบตอนเดียว (ภาพบน) เทียบกับขั้นตอนการตรวจจับวัตถุ แบบสองตอน (ภาพล่าง)	274
รูปที่ 8. 10 ภาพรวมอัลกอริทึมตรวจจับวัตถุ R-CNN.....	276
รูปที่ 8. 11 ภาพรวมอัลกอริทึมตรวจจับวัตถุ FAST R-CNN.....	276
รูปที่ 8. 12 อัลกอริทึมตรวจจับวัตถุ FASTER R-CNN	278
รูปที่ 8. 13 เน็ตเวิร์คเสนอพื้นที่ (REGION PROPOSAL NETWORK) AND PLACEMENTS OF ANCHORS	279
รูปที่ 8. 14 ภาพรวมสถาปัตยกรรม YOLOV1.....	281
รูปที่ 8. 15 บล็อก CONV* และ RESIDUAL ใน DARKNET-53.....	285
รูปที่ 8. 16 สถาปัตยกรรม YOLOV3 เพิ่มประสิทธิภาพการตรวจจับวัตถุขนาดเล็กด้วยการเพิ่มความลึก ของโครงข่ายสกัดคุณลักษณะและเพิ่มส่วนทำนายแบบหลายรายละเอียด (MULTI-SCALE PREDICTIONS) ในระบบตรวจจับวัตถุ.....	287
รูปที่ 8. 17 สถาปัตยกรรมอัลกอริทึมตรวจจับวัตถุ แบ่งฟังก์ชันการทำงานออกเป็นสามส่วนหลัก ได้แก่ แบล็คโบน คอ และ หัว.....	287
รูปที่ 8. 18 โมเดลสกัดคุณลักษณะแบบหลายรายละเอียด (ก) FPN: รวบรวมคุณลักษณะระดับ P3 ถึง P7 จากบนลงล่าง (ข) PANET ต่อยอดจาก FPN โดยเพิ่มการรวมคุณลักษณะจากล่างขึ้นบน (ค) NAS-FPN เริ่มมีการเชื่อมโยงคุณลักษณะแบบข้ามระดับ เพื่อให้โมเดลค้นหาคุณลักษณะพิเศษ รวมทั้งออกแบบให้มีการวางบล็อกซ้ำได้ (บล็อกสีม่วง) (ง) BIFPN ปรับแต่งโมเดลให้คุณลักษณะ มีการข้ามขึ้นจากชั้นอินพุตไปยังชั้นเอาต์พุต	289
รูปที่ 8. 19 สถาปัตยกรรมตัวตรวจจับวัตถุ EFFICIENTDET ที่ใช้โมเดล EFFICIENTNET เป็นแบล็คโบนและ ใช้ BIFPN ในส่วนคอ และคอนโวลูชันในส่วนหัว	289
รูปที่ 8. 20 ตัวตรวจจับวัตถุ YOLOV4 [31]	291
รูปที่ 8. 21 สถาปัตยกรรม YOLOV8 [33]	292
รูปที่ 8. 22 แสดงภาพรวมของสถาปัตยกรรม EFFICIENTNET-YOLOV5 [33]	293
รูปที่ 9. 1 การรู้จำบุคคลด้วยใบหน้าโดยใช้การเรียนรู้ความคล้ายและเอ็มเบดดิ้ง.....	302
รูปที่ 9. 2 เอ็มเบดดิ้งของข้อมูลที่มีความหมายใกล้เคียงกันจะถูกฉายลงบนปริภูมิเวกเตอร์ในตำแหน่งที่ใกล้ กันกว่าเอ็มเบดดิ้งของข้อมูลที่มีความหมายแตกต่างกัน.....	303

รูปที่ 9.3 การเรียนรู้เพื่อสร้างอิมเมจเอ็มเบดดิ้งด้วยโครงข่ายคอนโวลูชันและฟังก์ชันสูญเสีย.....	305
รูปที่ 9.4 การส่งต่อเวกเตอร์คุณลักษณะลงบนปริภูมิเวกเตอร์ (ด้านซ้าย) เมื่อใช้ฟังก์ชันสูญเสียจำแนกข้อมูลทั่วไป (ด้านขวา) เมื่อใช้ฟังก์ชันสูญเสียกับงานสร้างเอ็มเบดดิ้ง	306
รูปที่ 9.5 การเรียนรู้ทริปเลตใช้ภาพสามภาพ การเรียนรู้มีวัตถุประสงค์ส่งต่อเอ็มเบดดิ้งของคลาสเดียวกันให้อยู่ใกล้กันและผลึกเอ็มเบดดิ้งของภาพต่างคลาสนอกห่างกันมากกว่าระยะห่างระหว่างภาพในคลาสเดียวกัน	307
รูปที่ 9.6 โครงข่ายเฟสเน็ตประกอบด้วยการรับภาพแบบแบตช์ ตามด้วยโครงข่ายเชิงลึกคอนโวลูชัน ชั้น L2 นอร์มัลไลเซชันที่ให้ผลลัพธ์เป็นเอ็มเบดดิ้งเวกเตอร์ และตามด้วยการเรียนรู้ทริปเลตในช่วงฝึกสอน	309
รูปที่ 9.7 สถาปัตยกรรมแฝดสยาม	313
รูปที่ 9.8 ขั้นตอนการสืบค้นข้อมูลด้วยภาพ (IMAGE RETRIEVAL)	317
รูปที่ 9.9 ตัวอย่างเฟรมภาพจากกล้องถ่ายวิดีโอพร้อมกรอบตรวจจับสุนัขที่ได้จากโมเดล YOLOv3 รูปเล็กด้านขวาแสดงตัวอย่างภาพที่ผ่านการตัดภาพ	318
รูปที่ 9.10 เอ็มเบดดิ้งของสุนัขจริงที่สร้างได้จากการส่งผ่านความรู้แบบต่าง ๆ [19]	319
รูปที่ 9.11 การหาค่าเฉลี่ยความแม่นยำในงานสืบค้นข้อมูล	321
รูปที่ 9.12 ค่าเฉลี่ยของค่าเฉลี่ยความแม่นยำที่ k ของโมเดลทั้งสี่ [19]	322

สารบัญตาราง

ตารางที่ 2. 1 ฟังก์ชันที่นิยมใช้ในการแปลงภาพเทาและวัตถุประสงค์ในการแปลงภาพ	35
ตารางที่ 3. 1 ตารางแจกแจงค่าเวกเตอร์สี.....	65
ตารางที่ 3. 2 สมการหาค่าคุณลักษณะจากเมตริกซ์ระดับเทา	86
ตารางที่ 4. 1 เปรียบเทียบประสิทธิภาพฟังก์ชันกระตุ้นแบบต่าง ๆ [7]	122
ตารางที่ 4. 2 ฟังก์ชันวัดระยะทางแบบต่าง ๆ.....	138
ตารางที่ 4. 3 ประสิทธิภาพการตัดแบ่งตัวอักษร [25]	141
ตารางที่ 4. 4 เปรียบเทียบประสิทธิภาพการจำแนกตัวอักษรบนป้ายถนน.....	142
ตารางที่ 5. 1 รายละเอียดโครงข่ายคอนโวลูชัน LENET-5	158
ตารางที่ 5. 2 รายละเอียดโครงข่ายคอนโวลูชัน ALEXNET	160
ตารางที่ 5. 3 รายละเอียดโครงข่ายคอนโวลูชันตระกูล VGGNET	162
ตารางที่ 5. 4 สถาปัตยกรรมโครงข่าย RESNET ที่ความลึกต่าง ๆ [16].....	169
ตารางที่ 5. 5 เปรียบเทียบประสิทธิภาพโครงข่าย EFFICIENTNET กับโครงข่ายอื่น.....	170
ตารางที่ 5. 6 เปรียบเทียบแนวคิดในการออกแบบ	171
ตารางที่ 5. 7 จำนวนพารามิเตอร์ของโครงข่ายคอนโวลูชัน.....	172
ตารางที่ 5. 8 เปรียบเทียบข้อดีข้อเสียของกำหนดการอัตราเรียนรู้รูปแบบต่าง ๆ.....	175
ตารางที่ 6. 1 เกณฑ์การประเมินประสิทธิภาพโมเดลด้วยค่า AUC	193
ตารางที่ 6. 2 รหัสสั้นย่อต่าจำแนกภาพออกเป็นสี่หมวดหมู่ { นก, สุนัข, แมว, ห่าน}	195
ตารางที่ 6. 3 แสดงรายละเอียดภาพในฐานข้อมูลทั้งหก	219
ตารางที่ 6. 4 เปรียบเทียบประสิทธิภาพโมเดลต่าง ๆ กับงานจำแนกภาพ	220
ตารางที่ 7. 1 ค่าผลบวกเชิงเส้นและค่าหลังผ่านฟังก์ชันกระตุ้นซิกมอยด์	235
ตารางที่ 7. 2 เปรียบเทียบการถดถอยโลจิสติกและโครงข่ายคอนโวลูชัน	241
ตารางที่ 7. 3 ฐานข้อมูลภาพตามมาตรฐานที่นิยมใช้	243
ตารางที่ 7. 4 ค่าความไวและความจำเพาะของระบบจำแนกโรคต้อหินด้วยโครงข่ายเชิงลึก RESNET50 [24]	254
ตารางที่ 8. 1 ฐานข้อมูลที่นิยมใช้ในการวัดประสิทธิภาพระบบตรวจจับวัตถุ.....	271
ตารางที่ 8. 2 สถาปัตยกรรม YOLOv1	282
ตารางที่ 8. 3 สถาปัตยกรรม DARKNET-53 ใน YOLOv3 [13]	286
ตารางที่ 8. 4 โมเดลที่นิยมใช้เป็นแบคโบนในงานตรวจจับวัตถุ [15]	288

ตารางที่ 9. 1 โครงข่าย NN1 [3]	311
ตารางที่ 9. 2 ผลของขนาดเอ็มเบดดิ้งต่อประสิทธิภาพโครงข่ายเฟสเน็ต	312
ตารางที่ 9. 3 ผลเปรียบเทียบประสิทธิภาพโครงข่ายเฟสเน็ตเมื่อเลือกใช้โครงข่ายเชิงลึกแบบต่าง ๆ	312
ตารางที่ 9. 4 ข้อมูลนำเข้าพร้อมผลเฉลยโครงข่ายแผดสยาย	314
ตารางที่ 9. 5 จำนวนภาพที่ใช้ทดสอบและประเมินโมเดล	315
ตารางที่ 9. 6 อัตราการปฏิเสธที่ผิดพลาดของระบบยืนยันบุคคลด้วยใบหน้า [10] เมื่อกำหนดค่าอัตรา การยอมรับที่ผิดพลาด ที่ 5, 7.5 และ 10 เปอร์เซนต์	316
ตารางที่ 9. 7 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเอ็มเบดดิ้งของชุดจริงที่สร้างด้วยโมเดล เฟสเน็ตที่มีการส่งผ่านและไม่ส่งผ่านความรู้	319
ตารางที่ 9. 8 ผลเปรียบเทียบประสิทธิภาพโมเดลแบบต่าง ๆ	320

บทที่ 1:

บทนำ

การติดต่อสื่อสารแลกเปลี่ยนข้อมูลในยุคปัจจุบันนิยมทำผ่านสื่อออนไลน์ ทางระบบอินเทอร์เน็ต ซึ่งทำให้หลายรูปแบบทั้งการส่งข้อความข่าวสาร เสียง ภาพ และวิดีโอ ผ่านอุปกรณ์สื่อสารทั้งแบบมีสายและไร้สาย ได้แก่ สมาร์ทโฟน คอมพิวเตอร์ โน้ตบุ๊ก แท็บเล็ต และอุปกรณ์ IoT เทคโนโลยีเหล่านี้ทำให้การรับส่งข้อมูลทำได้ง่าย สะดวกและรวดเร็ว นอกจากนั้นยังมีแพลตฟอร์มออนไลน์ที่ช่วยกระจายข้อมูลในลักษณะเครือข่ายได้ ทำให้มีการส่งผ่านข้อมูลดิจิทัลในปริมาณมากและมีข้อมูลใหม่ที่หลากหลายถูกจัดเก็บเพิ่มขึ้นตลอดเวลา ทั้งข้อมูลชนิดข้อความ (Texts) ภาพนิ่ง (Images) ภาพเคลื่อนไหวหรือวิดีโอ (Videos) และคลิปเสียง (Voices) ข้อมูลจำนวนมากเหล่านี้ถูกจัดเก็บไว้ทั้งในเครื่องคอมพิวเตอร์ และในระบบคลาวด์ (Cloud storage) ช่วยให้ผู้ใช้งานอินเทอร์เน็ตสามารถเข้าถึงและใช้ประโยชน์จากข้อมูลที่ถูกจัดเก็บและได้รับอนุญาตให้ใช้งานได้ในปริมาณมากขึ้นมาก

ปัจจุบันการจัดเก็บและจัดทำชุดข้อมูลขนาดใหญ่ หรือข้อมูลมหัต (Big data) ทำได้ง่ายขึ้นและมีต้นทุนต่ำลง ข้อมูลขนาดใหญ่เหล่านี้มีประโยชน์อย่างมากเมื่อนำไปผ่านการประมวลผลวิเคราะห์ เรียนรู้หาความเชื่อมโยงระหว่างข้อมูล การจัดกลุ่มข้อมูล เพื่อเรียนรู้หาข้อจำกัดหรือความเฉพาะเจาะจงของข้อมูลในแต่ละกลุ่ม ความรู้ที่ได้จากฐานข้อมูลขนาดใหญ่นี้สามารถนำไปใช้ประโยชน์ได้มากและหลากหลาย เช่น ใช้ทำความเข้าใจพฤติกรรมการซื้อขายของลูกค้าแต่ละกลุ่มได้ ใช้คาดการณ์แนวโน้มที่เกิดขึ้นหรือพยากรณ์โรค ทำให้บริษัทสามารถตอบสนองและให้บริการลูกค้าในแต่ละกลุ่มได้ดีและมีประสิทธิภาพเพิ่มขึ้น เป็นต้น

เพื่อให้มนุษย์สามารถใช้ประโยชน์จากข้อมูลจำนวนมากได้อย่างมีประสิทธิภาพ จึงจำเป็นต้องพัฒนาเทคโนโลยีที่เกี่ยวข้อง ได้แก่ เทคโนโลยีการประมวลผลภาพ (Digital image processing) คอมพิวเตอร์วิทัศน์ (Computer vision) การจำแนกข้อมูลภาพ (Image classification) การเรียนรู้ของเครื่อง (Machine learning) และด้านปัญญาประดิษฐ์ (Artificial intelligence) ให้มีความก้าวหน้า มีประสิทธิภาพ เพื่อให้เครื่องจักรสามารถสืบค้น เรียนรู้จากข้อมูลจำนวนมากที่มีอยู่ได้อย่างอัตโนมัติและมีประสิทธิภาพ สามารถวิเคราะห์ สรุป ดึงข้อมูลสำคัญ หาโมเดลความสัมพันธ์และค่าตัวแปรที่น่าสนใจจากข้อมูลที่มีอยู่จำนวนมากได้อย่างถูกต้อง แม่นยำ รวมทั้งสามารถนำเสนอข้อมูลหรือสารสนเทศต่าง ๆ ได้ตามต้องการ ทำให้สามารถใช้งานข้อมูลจำนวนมากได้อย่างเกิดประโยชน์สูงสุด

1.1 นิยามศัพท์ที่เกี่ยวข้องกับการจำแนก

ในช่วงการพัฒนาเทคโนโลยีดิจิทัลขั้นต้น เทคโนโลยีต่าง ๆ ที่กล่าวไว้ก่อนหน้านี้ เช่น การประมวลผลภาพ การเรียนรู้ด้วยเครื่อง มีการวิจัยพัฒนาที่ค่อนข้างเป็นอิสระต่อกัน โดยมีเทคโนโลยีการประมวลผลภาพเป็นพื้นฐาน การพัฒนามีวัตถุประสงค์หลักร่วมกัน คือ ทำให้เครื่องคอมพิวเตอร์หรือเครื่องจักรมีความสามารถใกล้เคียงกับความสามารถของมนุษย์ ทั้งในด้าน การมองเห็น การวิเคราะห์ตีความ และทำความเข้าใจภาพที่มองเห็น หรือเสียงที่ยิน โดยมนุษย์สามารถรู้จำ แบ่งหมวดหมู่ จำแนกประเภทของวัตถุได้ดี สามารถเชื่อมโยงความสัมพันธ์ระหว่างสิ่งของต่าง ๆ ที่ปรากฏภายในภาพได้ มนุษย์สามารถตีความ และเข้าใจความหมายต่าง ๆ ภายในรูปภาพได้อย่างรวดเร็ว เช่น นับจำนวนคนที่ปรากฏภายในภาพและฐานข้อมูลภาพจากแหล่งภาพต่าง ๆ ทั้งยังสามารถระบุกิจกรรมที่เกิดขึ้นภายในภาพได้ แต่สำหรับเครื่องจักรคอมพิวเตอร์ กระบวนการจำแนก รู้จำ ตีความภาพหรือหาวัตถุภายในภาพ ทำได้ยากและยังให้ประสิทธิภาพไม่ดี โดยเฉพาะเมื่อประมวลผลกับภาพที่มีความซับซ้อน มีวัตถุจำนวนมากภายในภาพ ทำให้นักวิจัยด้านการจำแนกภาพ การรู้จำวัตถุมีมาอย่างต่อเนื่อง โดยผลงานวิจัยที่คอมพิวเตอร์เริ่มให้ประสิทธิภาพการจำแนกรู้จำภาพแบบก้าวกระโดดเกิดขึ้นเมื่อมีการนำเทคโนโลยีการเรียนรู้เชิงลึก (Deep learning) มาประยุกต์ใช้ทำความเข้าใจภาพเพื่อความเข้าใจในการพัฒนาระบบ ผู้อ่านควรทำความเข้าใจคำศัพท์ที่เกี่ยวข้องดังต่อไปนี้

การรู้จำวัตถุ (Object recognition) เป็นศาสตร์ที่เครื่องจักรคอมพิวเตอร์ทำการอ่านข้อมูลภาพหรือภาพถ่ายวิดีโอ แล้วประมวลผล วิเคราะห์ภาพ เพื่อรู้จำวัตถุที่ปรากฏภายในภาพ การรู้จำวัตถุเป็นคำที่ใช้ในความหมายรวมๆ โดยอาจหมายถึงการจำแนกภาพ (Image classification) หรืออาจหมายถึงชุดขั้นตอนที่ประกอบด้วย การตรวจหาวัตถุ (Object localization) และการจำแนกวัตถุ

โดยในรายละเอียด สามารถแบ่งลักษณะงานที่เกี่ยวข้องกับการรู้จำวัตถุ ออกได้เป็นสามลักษณะใหญ่ๆ คือ การจำแนกข้อมูล การแบ่งส่วนภาพ และการตรวจจำวัตถุ

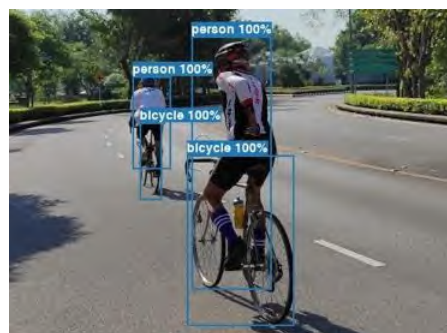
การจำแนกข้อมูล (Classification) เป็นศาสตร์ที่คอมพิวเตอร์ประมวลผลข้อมูลนำเข้า เช่น ภาพ เสียง ข้อความ แล้วตัดสินใจจัดหมวดหมู่หรือประเภทให้กับข้อมูล โดยกำหนดเรียกประเภทของข้อมูลว่า **คลาส (Class)** เช่น งานจำแนกเพศจากเสียง (Voice gender classification) ระบบอ่านไฟล์เสียงแล้วตัดสินใจว่าเสียงที่ได้ยินเป็นเสียงของคลาสไหนระหว่างคลาสเสียงผู้หญิงหรือคลาสเสียงผู้ชาย ระบบจำแนกอารมณ์จากภาพถ่าย (Emotion classification) ระบบอ่านภาพใบหน้าแล้วตัดสินใจว่าใบหน้าไหนที่เห็นจัดอยู่ในคลาสใดระหว่างคลาส ความสุข เศร้า โกรธ หรือเฉยๆ เป็นต้น ในกรณีที่มีข้อมูลนำเข้าเป็นภาพ นิยมเรียกขบวนการนี้ว่า **การจำแนกภาพ** และถ้าข้อมูลนำเข้าเป็นข้อความ นิยมเรียกว่า **การจำแนกข้อความ (Text classification)** ซึ่งงานจำแนกทั้งสองมีการใช้เทคนิคทั้งที่

เหมือนกันและแตกต่างกัน เนื้อหาในหนังสือเล่มนี้ครอบคลุมเนื้อหาที่เกี่ยวข้องกับข้อมูลภาพ ซึ่งเป็นข้อมูลสองมิติ

ขั้นตอนการจำแนกภาพเริ่มต้นด้วยการอ่านภาพนำเข้า ประมวลผลเพื่อสกัดหาคุณลักษณะสำคัญภายในภาพ ตัดสินใจระบุประเภทว่าเป็นภาพในคลาสใด เช่น แมว ม้า สุนัข เป็นต้น โดยผลลัพธ์ที่ได้จากการจำแนกภาพ อาจระบุเป็นคลาสของวัตถุภายในภาพอย่างเดียว หรือระบุคลาสของภาพพร้อมแสดงค่าความน่าจะเป็นที่ภาพนั้นเป็นภาพของคลาสที่ระบุ การจำแนกภาพโดยทั่วไปให้ความสนใจกับวัตถุที่เด่นหนึ่งวัตถุภายในภาพเป็นสำคัญ ดังแสดงในภาพที่ 1.1 (ก) ซึ่งจำแนกภาพว่าจัดอยู่ในคลาส ดอกไม้ เป็นต้น โดยผลลัพธ์ที่ได้จากการจำแนกภาพไม่มีการระบุตำแหน่งของวัตถุภายในภาพ



(ก)



(ข)



(ค)



(ง)

รูปที่ 1.1 (ก) การจำแนกภาพ (Image classification) (ข) การตรวจจับวัตถุ (Object detection) (ค) การแบ่งส่วนภาพเชิงความหมาย (Semantic segmentation) (ง) การแบ่งส่วนภาพเชิงรายตัวอย่าง (Instance segmentation)

การตรวจจับวัตถุ (Object detection) เป็นเทคนิคตรวจหาวัตถุที่ปรากฏภายในภาพ ให้ผลลัพธ์เป็นประเภทของวัตถุแต่ละวัตถุที่พบภายในภาพ พร้อมทั้งระบุตำแหน่งของวัตถุ การระบุตำแหน่งนิยมทำด้วยการตีกรอบสี่เหลี่ยมครอบวัตถุ กรอบนี้เรียกว่า กล่องขอบเขตหรือบาวนิงบ็อกซ์ (Bounding box: Bbox) และอัลกอริทึมระบุพิกัดและขนาดของกล่องขอบเขตที่หาได้ โดยสามารถระบุพิกัดกล่องได้หลายวิธี เช่น ระบุจุดเริ่มต้นจุดมุมบนซ้ายของกล่องขอบเขต ความกว้าง และความสูง ของกล่อง (x, y, w, h) หรือระบุพิกัดจุดศูนย์กลาง ความกว้าง และความสูง ของกล่องขอบเขต (cx, cy, w, h)

เป็นต้น ภาพที่ 1.1(ข) แสดงตัวอย่างผลลัพธ์ที่ได้จากการตรวจจับวัตถุ โดยกรอบสีฟ้า ระบุว่า เป็นวัตถุคลาส bicycle มีความมั่นใจในการจำแนกคลาส 100% เป็นต้น

เทคนิคการตรวจจับวัตถุมีความคล้ายคลึงกับเทคนิคการบอกตำแหน่งวัตถุ (Object localization) บ่อยครั้งพบการใช้คำสองคำนี้ทดแทนกัน คำศัพท์ทั้งสองมีความแตกต่างกันเพียงเล็กน้อย นั่นคือ การบอกตำแหน่งวัตถุมักมีการเจาะจงวัตถุที่ต้องการหาตำแหน่ง เช่น การบอกตำแหน่งตาข่ายในภาพ ระบบค้นหาพื้นที่ตาข่ายในภาพ หลังจากพบพื้นที่ตาข่าย ระบบอาจประมาณขอบเขตตาข่ายด้วยวงกลมหรือวงรี และระบุจุดพิกัดของเส้นวงรอบตาข่ายได้ประมาณไว้ เช่น ระบุพิกัดจุดศูนย์กลางตาข่าย และรัศมี กรณีประมาณขอบเขตตาข่ายด้วยวงกลม หรือ ระบุ 0 ในกรณีไม่พบตาข่ายในภาพ ในขณะที่การตรวจจับวัตถุ ค้นหาพร้อมทีกรอบกล่องขอบเขตของวัตถุทุกวัตถุที่ตรวจพบภายในภาพ

การแบ่งส่วนภาพ (Image segmentation) การแบ่งส่วนภาพจัดเป็นเทคนิคการจำแนกภาพในระดับพิกเซล การแบ่งส่วนภาพหลังอ่านภาพนำเข้าแล้ว ระบบตัดสินใจจำแนกพิกเซลภายในภาพทีละพิกเซลว่าแต่ละพิกเซลเป็นพิกเซลของวัตถุใดในคลาส โดยนิยมแสดงผลลัพธ์การแบ่งส่วนภาพด้วยค่าสี เช่น พิกเซลสีม่วงหมายถึงพิกเซลคลาสแมว สีเขียวเป็นพิกเซลคลาสสุนัข เป็นต้น

การแบ่งส่วนภาพแบ่งการทำงานออกเป็นสองรูปแบบ ได้แก่ การแบ่งส่วนภาพเชิงความหมาย (Semantic segmentation) และการแบ่งส่วนภาพเชิงรายตัวอย่าง (Instance segmentation)

การแบ่งส่วนภาพเชิงความหมาย จำแนกพิกเซลภายในภาพเป็นวัตถุคลาสใหญ่ ๆ เช่น คลาสคน รถ สุนัข โดยไม่แบ่งแยกวัตถุที่พบว่ายู่ในคลาสเดียวกันแต่เป็นวัตถุคนละชิ้น เช่น ตรวจพบพิกเซลของสุนัขสองตัวภายในภาพ ผลการแบ่งส่วนภาพเชิงความหมาย แทนพิกเซลของสุนัขทั้งสองด้วยสีเดียวกัน และระบุคลาสเป็น “Dog” โดยไม่แบ่งแยกว่าเป็นสุนัขต่างตัว ดังแสดงตัวอย่างในรูปที่ 1.1 (ค)

การแบ่งภาพเชิงรายตัวอย่าง ค้นหาพิกเซลที่เป็นส่วนหนึ่งของวัตถุทุกวัตถุภายในภาพ และพิจารณาแยกคลาสของวัตถุแต่ละชิ้นแตกต่างกัน เช่น รูป 1.1(ง) ตรวจพบพิกเซลของสุนัข 2 ตัว อัลกอริทึมระบุคลาสของสุนัขตัวแรกเป็นคลาส “Dog1” แทนด้วยสีชมพู และสุนัขตัวที่ 2 เป็นคลาส “Dog2” แทนด้วยสีฟ้า

ข้อแตกต่างที่สำคัญระหว่างการแบ่งส่วนภาพและการตรวจจับภาพ คือวิธีการแสดงผลของวัตถุที่ตรวจพบหรือแบ่งส่วนได้ ผลลัพธ์การแบ่งส่วนภาพไม่จำเป็นการแบ่งส่วนภาพเชิงความหมายหรือเชิงรายตัวอย่าง เส้นกำหนดขอบเขตของวัตถุที่แบ่งส่วนได้จะใกล้เคียงกับเส้นรอบรูปร่างของวัตถุ เส้นขอบเขตมีความโค้งงอไปตามรูปร่างจริงของวัตถุที่แบ่งส่วนได้ ในขณะที่การตรวจจับวัตถุขอบเขตของวัตถุที่ตรวจจับได้เป็นกรอบสี่เหลี่ยมล้อมรอบวัตถุที่ตรวจพบ

1.2 การประยุกต์ใช้งาน

เนื่องจากความก้าวหน้าในการพัฒนาเทคโนโลยีด้านปัญญาประดิษฐ์ การเรียนรู้ด้วยเครื่องบวกกับความก้าวหน้าด้านคอมพิวเตอร์ และเซ็นเซอร์ที่เกี่ยวข้องกับประสาทสัมผัสของมนุษย์ เช่น เซ็นเซอร์ดมกลิ่น เซ็นเซอร์รับเสียงและภาพ ทำให้มีการนำเทคนิคการจำแนกภาพ การตรวจจับวัตถุ การแบ่งส่วนภาพ และการรู้จำภาพ ไปประยุกต์ใช้งานอย่างแพร่หลายในอุตสาหกรรมต่าง ๆ ยกตัวอย่าง เช่น

อุตสาหกรรมยานยนต์ มีการพัฒนาระบบขับเคลื่อนอัตโนมัติไร้คนขับ (Unmanned vehicle) ซึ่งระบบต้องทำการตรวจจับวัตถุ จำแนก และรู้จำวัตถุภายในภาพ เพื่อให้ยานยนต์สามารถหลบหลีกเครื่องกีดขวาง ตรวจจับป้ายบนถนน [1] รู้จำและปฏิบัติตามป้ายเครื่องหมายจราจร เพื่อให้ยานยนต์สามารถขับเคลื่อนไปยังสถานที่เป้าหมายได้อย่างถูกต้องและปลอดภัย ในกรณียานยนต์มีคนขับ มีการพัฒนาใช้เทคนิคตรวจจับดวงตา ทำการแยกส่วนภาพหาพื้นที่ตาดำตาขาว เพื่อใช้เป็นค่าบ่งชี้การหลับตาและระดับความง่วงของคนขับ (Drowsiness classification) นอกจากนี้ยังมีการนำเทคนิคการจำแนกภาพไปประยุกต์ใช้ในการตรวจจับผู้ขับขี่ฝ่าฝืนกฎจราจร เช่น ตรวจจับผู้ขับขี่ยานยนต์ที่ฝ่าฝืนขับข้ามเส้นขาวที่แบบอัตโนมัติ เป็นต้น

อุตสาหกรรมทางการแพทย์ เทคนิคที่ถูกประยุกต์ใช้งานอย่างกว้างขวางในวงการอุตสาหกรรมทางการแพทย์ คือ เทคนิคการแบ่งส่วนภาพ เช่น การแยกส่วนภาพถ่ายสมองมนุษย์ เพื่อแบ่งส่วนพิกเซลก้อนเนื้อออกจากพิกเซลส่วนอื่น ๆ ของสมอง ผลการแบ่งส่วนภาพช่วยให้การผ่าตัดทางไกลทำได้รวดเร็ว และง่ายขึ้น การแบ่งส่วนภาพจอบประสาทตาจากภาพถ่ายตา เพื่อใช้วิเคราะห์ค่าพารามิเตอร์ที่สำคัญสำหรับการวิเคราะห์และจำแนกโรคตา (Eye disease classification) เช่น จำแนกภาพถ่ายผู้ป่วยโรคเบาหวาน และต้อหิน เป็นต้น โดยทั้งเทคนิคการแบ่งส่วนภาพ การตรวจจับวัตถุ และจำแนกภาพ เป็นเครื่องมือสำคัญในการพัฒนาระบบการวินิจฉัยโรคโดยใช้คอมพิวเตอร์ช่วย (Computer-Aided Diagnosis: CAD) ซึ่งในปัจจุบันได้มีการพัฒนาระบบ CAD สำหรับช่วยวินิจฉัยโรคหลายโรค ได้แก่ อัลไซเมอร์ วินิจฉัยหูก้อนในปอด ช่วยตรวจสอบโรคมะเร็งเต้านม [2] ช่วยตรวจหาโรคหลอดเลือดหัวใจตีบ เป็นต้น

อุตสาหกรรมรักษาความปลอดภัย ปัจจุบันมีการนำเอาเทคโนโลยีด้านการตรวจจับ รู้จำบุคคลและวัตถุไปประยุกต์ใช้ในงานรักษาความปลอดภัยอย่างกว้างขวาง เช่น พัฒนาระบบตรวจจับและรู้จำใบหน้ามนุษย์ [3] ระบบระบุยืนยันตัวตนด้วยลายม่านตาและลายนิ้วมือ สำหรับความปลอดภัยในการเข้าออกอาคารสถานที่ต่าง ๆ โดยในปัจจุบันกล้องวงจรปิดมีราคาถูกลงและสามารถประมวลผลแบบเรียลไทม์ได้ จึงนิยมติดตั้งทั้งภายในและภายนอกอาคาร จึงนิยมพัฒนาระบบตรวจจับและติดตามวัตถุและคนที่ปรากฏภายในภาพกล้องวงจรปิด ตัวอย่าง เช่น พัฒนาเทคนิคสำหรับวิเคราะห์ จำแนก

สิ่งผิดปกติแปลกปลอม หรือจำแนกิริยาท่าทาง (Action recognition) ของบุคคลที่ปรากฏภายในภาพ เพื่อวิเคราะห์หาสิ่งผิดปกติ และแจ้งเตือนอันตรายที่อาจเกิดขึ้นได้อย่างอัตโนมัติ

อุตสาหกรรมหุ่นยนต์และการผลิต การตรวจจับ แบ่งส่วน จำแนกและรู้จำภาพ เป็นส่วนสนับสนุนที่สำคัญในการพัฒนาหุ่นยนต์ ช่วยเพิ่มประสิทธิภาพการทำงานของหุ่นยนต์ให้สามารถมองเห็นและทำงานตามคำสั่งได้ เช่น การแบ่งส่วนภาพ และการจำแนกภาพเพื่อหารอยแตกในชิ้นงานโลหะหรือในแผ่นวงจรอิเล็กทรอนิกส์ การตรวจจับภาพในหุ่นยนต์นับและบรรจุชิ้นงานหรือวัตถุในสายการผลิต เช่น การบรรจุเม็ดยา ในหุ่นยนต์ค้นหาผู้ประสบภัยแผ่นดินไหวหรือตึกถล่ม หุ่นยนต์เหล่านี้ต้องทำการตรวจจับจำแนกคนและวัตถุที่ต้องค้นหาในระหว่างปฏิบัติการ นอกจากนี้ ในปัจจุบันมีความต้องการใช้งานหุ่นยนต์ในภาคงานบริการต่าง ๆ เพิ่มขึ้นเป็นจำนวนมาก เช่น หุ่นยนต์ช่วยเหลือผู้ป่วยหรือผู้สูงอายุ ที่หุ่นยนต์ต้องมีความสามารถในการมองเห็น จำแนกผู้ป่วยและวัตถุต่าง ๆ ภายในห้องได้ สามารถเคลื่อนที่ภายในห้องได้อย่างปลอดภัย รวมทั้งยังต้องสามารถรู้จำเสียง เพื่อให้หุ่นยนต์สามารถโต้ตอบและให้บริการผู้ป่วยได้อย่างอัตโนมัติ

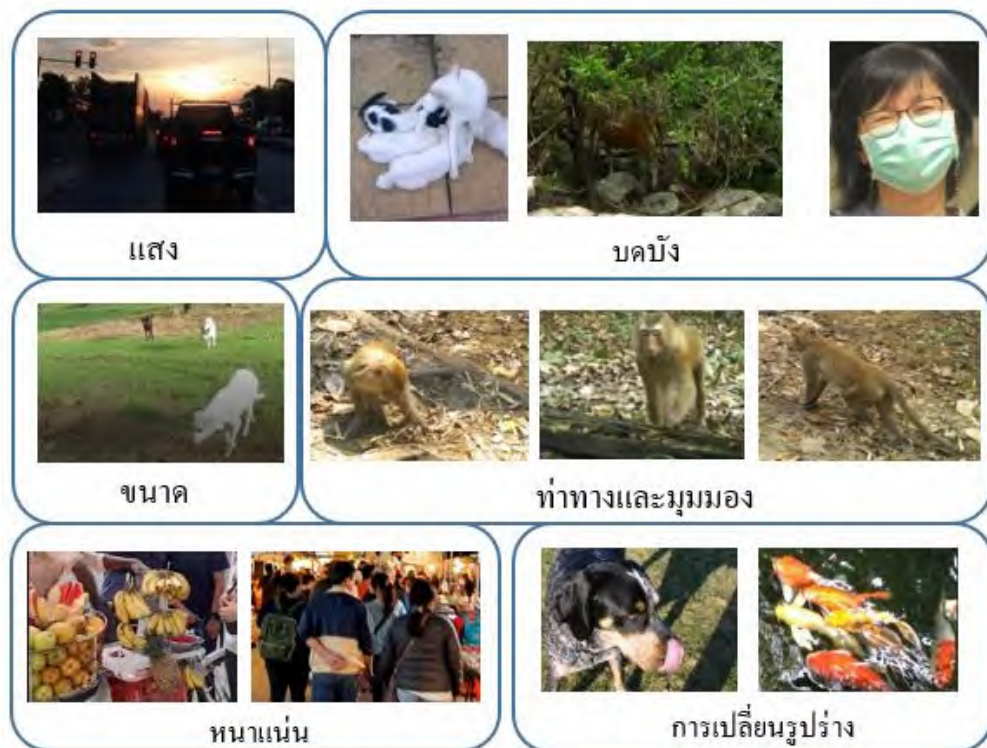
อุตสาหกรรมเกษตร เนื่องจากปัญหาในการขาดแคลนแรงงาน รวมทั้งปัญหาโรคพืชและสัตว์ที่เกิดจากเชื้อโรคที่มีคนเป็นพาหะสำคัญ ทำให้มีความต้องการใช้งานระบบที่เกี่ยวข้องกับงานการเกษตรแบบอัตโนมัติ รวมทั้งมีความต้องการระบบเกษตรอัจฉริยะเพิ่มขึ้น ตัวอย่าง เช่น ระบบตรวจจับและจำแนกมะเขือเทศ เพื่อการให้น้ำแบบอัตโนมัติ การรดน้ำนักสุกรหรือวัวแบบอัตโนมัติ ระบบจำแนกโรคใบพืช ระบบติดตามพฤติกรรมภายในคอก ระบบจำแนกพื้นที่ปลูกข้าวจากภาพถ่ายดาวเทียม และระบบตรวจสอบรอยร้าวในไข่ [4] เป็นต้น

1.3 ความท้าทาย

จากการที่วิทยาการด้านการจำแนกภาพและวัตถุถูกนำไปประยุกต์ใช้งานอย่างกว้างขวางในหลากหลายอุตสาหกรรม ทำให้เกิดความต้องการระบบจำแนกภาพ รู้จำวัตถุที่มีประสิทธิภาพสูง มีความแม่นยำในการจำแนกภาพ และตรวจจับวัตถุต่าง ๆ ภายในภาพได้มากกว่าหนึ่งวัตถุ เป็นระบบที่สามารถปรับใช้ได้กับภาพถ่ายจากพื้นที่ ๆ มีสภาพแวดล้อมที่แตกต่างกันมาก ยกตัวอย่างเช่น ระบบยานยนต์ไร้คนขับควรใช้งานได้ดีตลอดวันทั้งในเวลากลางวัน ย่ำรุ่งและอื่น ๆ ทำให้การพัฒนาเทคโนโลยีในวิทยาการเหล่านี้มีความท้าทายเป็นอย่างมากตามปัจจัยที่หลากหลาย ปัญหาความท้าทายที่พบย่อยแสดงภาพตัวอย่างในรูปที่ 1.2

ความแตกต่างของแสงภายในภาพ (Illumination) การบันทึกภาพถ่ายทั้งภาพฟิล์มและภาพดิจิทัลค่าที่บันทึกมีความสัมพันธ์เกี่ยวข้องกับค่าความเข้มแสงที่สะท้อนออกมาจากผิววัตถุ แสงที่ตกกระทบวัตถุ และแสงพื้นหลังรอบ ๆ วัตถุ ค่าความเข้มแสงที่บันทึกไว้มีผลต่อการประมวลผลภาพด้วยเครื่องคอมพิวเตอร์อย่างมาก ตัวอย่างจากการทดลองรู้จำใบหน้าคนในงานวิจัย [5] พบว่าคุณลักษณะที่

คอมพิวเตอร์เรียนรู้จากภาพถ่ายใบหน้าของคน ๆ เดียวกัน ภายใต้แสงที่แตกต่างกัน เช่น ภาพถ่ายในที่แสงสลัวใกล้ค่ำ และถ่ายกลางแจ้ง มีความแตกต่างจากคุณลักษณะที่เรียนรู้ได้จากภาพถ่ายใบหน้าของบุคคลอื่นที่ถ่ายในพื้นที่ ๆ มีลักษณะแสงเดียวกันมากกว่ามาก ทำให้บ่อยครั้งระบบจำแนกใบหน้าบุคคลทำงานผิดพลาด ชนิดของแสง ระดับความเข้มของแสง มุมและลักษณะของแสงสว่างที่ตกกระทบวัตถุจึงเป็นปัญหาท้าทายที่สำคัญในงานประมวลผลภาพดิจิทัล



รูปที่ 1. 2 ปัญหาและความท้าทายที่พบบ่อยในงานจำแนกภาพและตรวจจับวัตถุ

ขนาดของวัตถุ (Scale) ขนาดของวัตถุภายในภาพเปลี่ยนแปลงไปตามระยะห่างระหว่างวัตถุกับเลนส์กล้อง ภาพถ่ายของวัตถุที่มีการเคลื่อนที่ หรือการถ่ายภาพในสถานที่จริงที่บ่อยครั้งที่กล้องมีการเคลื่อนที่ ทำให้ได้ภาพของวัตถุเดิมในขนาดต่าง ๆ กัน ระบบรู้จำ จำแนกวัตถุและภาพที่ดีจึงควรออกแบบรองรับขนาดที่เปลี่ยนไปของวัตถุ โดยทั่วไประบบมักมีข้อจำกัด ไม่สามารถรู้จำวัตถุที่มีขนาดเล็กมาก ๆ ได้

มุมมอง หรือท่าทางของวัตถุ (Pose and viewpoint) มุมมองภาพถ่ายของวัตถุขึ้นหนึ่ง ๆ สามารถมีความแตกต่างกันอย่างมากได้ ตัวอย่างเช่น ภาพถ่ายหน้าตรงและด้านข้างของใบหน้าคนเดียวกัน สามารถมีลักษณะที่ต่างกันอย่างมาก ทำให้การพัฒนาระบบจำแนกบุคคลด้วยใบหน้าสำหรับใช้ในสถานที่เปิด เช่น สนามกีฬา ทำได้ยากกว่าการจำแนกบุคคลด้วยใบหน้าในอาคารสถานที่ทำงานที่มีการตั้งกล้องระยะคงที่และถ่ายภาพหน้าคนด้านตรงเป็นสำคัญ

การเปลี่ยนรูปร่าง (Deformation) วัตถุและสิ่งมีชีวิตหลายชนิด เช่น ดอกไม้ แมว สามารถเปลี่ยนแปลงรูปร่างของตัวเองได้ ทำให้ต้องพัฒนาระบบที่รองรับการเปลี่ยนแปลงรูปร่างของวัตถุที่เราสนใจ เช่น ในระบบจำแนกสุนัข ระบบต้องรู้จำสุนัขตัวนั้นในอริยาบถต่าง ๆ กันได้ทั้งในเวลานอนหลับ หาว เหา และ ขู่ ซึ่งอริยาบถเหล่านี้ส่งผลให้ภาพถ่ายของสุนัขที่ได้มีรูปร่างที่แตกต่างกันมาก โดยทั่วไปการเปลี่ยนแปลงรูปร่างไม่เป็นปัญหาสำหรับมนุษย์ในการรู้จำสุนัขตัวนั้น ๆ แต่เป็นปัญหาสำคัญสำหรับเครื่องจักรและคอมพิวเตอร์ในการรู้จำวัตถุที่มีการเปลี่ยนแปลงรูปร่าง เนื่องจากคุณลักษณะสำคัญที่สกัดได้จากภาพของสุนัขตัวหนึ่งในอริยาบถต่าง ๆ มีความแตกต่างกันมาก และที่สำคัญคือค่าความแตกต่างของคุณลักษณะที่สกัดได้จากสุนัขตัวเดียวกันมีค่ามากกว่าความแตกต่างเมื่อเทียบกับคุณลักษณะที่สกัดได้จากสุนัขตัวอื่น ๆ ส่งผลให้เกิดความผิดพลาดในการรู้จำหรือจำแนกภาพนั้น ๆ

ความหลากหลายของชนิดวัตถุที่ต้องการจำแนก ในการพัฒนาเทคโนโลยีการรู้จำภาพและวัตถุในเบื้องต้น การจำแนก ตรวจสอบ วัตถุมีความเจาะจง เช่น ตรวจสอบใบหน้า ตรวจสอบสุนัข แต่ปัจจุบันนักวิชาการและนักวิจัยมีความสนใจต้องการพัฒนาระบบจำแนกภาพที่มีศักยภาพใกล้เคียงมนุษย์มากขึ้น เป็นระบบที่สามารถตรวจสอบและจำแนกวัตถุทั่วไปที่ปรากฏภายในภาพ โดยไม่ต้องเจาะจงระบุชนิดของวัตถุที่ต้องการค้นหา นั่นคือ เมื่อภาพถูกนำเข้าไปกับระบบจำแนกและตรวจสอบภาพ ระบบต้องสามารถตรวจสอบวัตถุทุกชนิดที่อยู่ในภาพได้ ส่งผลให้นักวิจัยต้องรวบรวมฐานข้อมูลภาพขนาดใหญ่ สำหรับใช้ทำวิจัยด้านการจำแนกรู้จำภาพ โดยฐานข้อมูลเหล่านี้ประกอบไปด้วยภาพของวัตถุหลายหมื่นชนิด เช่น ฐานข้อมูล ImageNet [6] ที่ประกอบด้วยภาพถ่ายของวัตถุ 22,000 ชนิด มีภาพในฐานข้อมูลมากกว่าล้านภาพ นอกจากนี้ ยังมีการเผยแพร่ภาพ วิดีโอ และฐานข้อมูลภาพจำนวนมากบนอินเทอร์เน็ต ทำให้สามารถพัฒนางานวิจัยจากการตรวจสอบวัตถุเฉพาะ ไปสู่การตรวจสอบวัตถุได้หลากหลายชนิดมากขึ้น ฐานข้อมูลขนาดใหญ่ที่เปิดสาธารณะให้ดาวน์โหลดมาใช้งานได้ฟรี ช่วยให้เครื่องจักรทำความเข้าใจหาความสัมพันธ์ระหว่างข้อมูลที่มนุษย์ใช้สื่อสารระหว่างกันได้ดีขึ้น ส่งผลให้สามารถวิเคราะห์และประยุกต์ใช้ข้อมูลแบบอัตโนมัติให้ตรงตามความต้องการของผู้ใช้มากขึ้น

ความผันแปรของลักษณะของวัตถุภายในคลาสเดียวกัน (Intra-class variance) วัตถุในคลาสเดียวกัน สามารถมีรูปร่างลักษณะแตกต่างกันมากได้ เช่น รถยนต์ รองเท้า เสื้อผ้า อาคาร ที่มาในหลากหลายรูปแบบ ทำให้ต้องหาลักษณะเด่นที่สามารถใช้เป็นตัวแทนของวัตถุในคลาสเดียวกันและเป็นลักษณะเด่นที่แตกต่างจากลักษณะเด่นของคลาสอื่น ๆ ซึ่งทำได้ยาก

การบดบัง (Occlusion) ภาพแต่ละภาพที่เรามองเห็นในแต่ละวันมักประกอบไปด้วยวัตถุหลายวัตถุจัดวางในลักษณะที่มีความซับซ้อน วัตถุส่วนใหญ่มักถูกบดบังด้วยวัตถุอื่น ๆ เช่น คนเดินด้านหน้า บังคนที่เดินตามมาข้างหลัง โดยอาจบดบังเพียงบางส่วนหรือบดบังทั้งหมดก็ได้ หรือสัตว์ต่าง ๆ ที่อยู่ ในกรงถูกบดบังด้วยโครงสร้างกรง หรือกิ่งไม้ เป็นต้น

นอกจากนั้นวัตถุหลายประเภทมักถูกบดบังโดยธรรมชาติ เช่น ม่านตามนุษย์ที่ถูกบดบังด้วยขอบตา และขนตา การบดบังเหล่านี้เมื่อมองด้วยตามนุษย์ สมองของมนุษย์สามารถคาดเดาและเพิ่มเติม ข้อมูลภาพส่วนที่ขาดหายไปบางส่วนจากการบดบังได้อย่างอัตโนมัติ ส่งผลให้มนุษย์สามารถรู้จำและ จำแนกวัตถุที่ถูกบดบังได้ดี มนุษย์สามารถระบุคนคุ้นเคยที่เดินตามหลังผู้อื่นมาแม้ได้เห็นเพียงเสี้ยว หนึ่งของคนนั้น แต่ข้อมูลที่หายไปจากการถูกบดบัง เป็นอุปสรรคที่สำคัญของเครื่องจักรในการเรียนรู้ จำแนก และระบุขอบเขตของวัตถุ ระบบรู้จำวัตถุหลายระบบมักมีขั้นตอนประมวลผลก่อน (Pre-processing) เพื่อกำจัดสิ่งบดบัง เช่น การตัดขนตาและขอบตาออกจากส่วนม่านตาก่อนส่งภาพที่ได้ ให้กับขั้นตอนการรู้จำม่านตา [7]

จำนวนวัตถุหรือความหนาแน่นของวัตถุภายในภาพ (Clustered objects) ภาพที่ประกอบไปด้วย วัตถุจำนวนมาก มักเกิดปัญหาการทับซ้อน บดบังกันของวัตถุ และบ่อยครั้งประกอบไปด้วยวัตถุที่มี ความคล้ายคลึง ทำให้เกิดความสับสน ยากต่อการแยกประเภทของวัตถุ

1.4 ภาพรวม

จากความท้าทายข้างต้น การพัฒนาระบบรู้จำวัตถุให้มีประสิทธิภาพใกล้เคียงมนุษย์ทำได้ยาก งานวิจัยในระยะแรกเน้นการจำแนกภาพผ่านการตรวจจับวัตถุที่ต้องการ โดยนิยามหาความเหมือน ระหว่างภาพวัตถุต้นแบบที่ต้องการ เรียกว่า เทมเพลต หรือแม่แบบ (Template) กับภาพนำเข้าหรือ ภาพอินพุต การหาทำโดยเลื่อนภาพเทมเพลตไปบนภาพอินพุต และหาค่าความเหมือนระหว่างภาพ เทมเพลตและภาพอินพุตส่วนที่อยู่ด้านล่างเทมเพลต จากนั้นเลื่อนภาพเทมเพลตไปเรื่อย ๆ จน ครอบคลุมครบทุกพื้นที่ในภาพอินพุต วิธีนี้เรียกว่า การจับคู่เทมเพลต (Template matching) รูปที่ 1.3 แสดงตัวอย่างภาพแม่แบบและภาพอินพุตที่ต้องการทำการค้นหาภาพแม่แบบ ข้อดีของวิธีจับคู่ ภาพแม่แบบ คือ ประมวลผลได้ง่าย สะดวก ใช้งานได้กับการตรวจหาวัตถุที่มีความเฉพาะเจาะจง มีภาพ ต้นแบบ และวัตถุที่ต้องการค้นหาไม่ถูกบดบัง ข้อเสียของวิธีนี้ คือไม่สามารถค้นหาวัตถุที่ถูกบดบัง หรือวัตถุที่มีขนาด และแสงที่แตกต่างจากภาพเทมเพลตมาก ๆ ได้



(ก) เทมเพลต



(ข) ภาพอินพุต

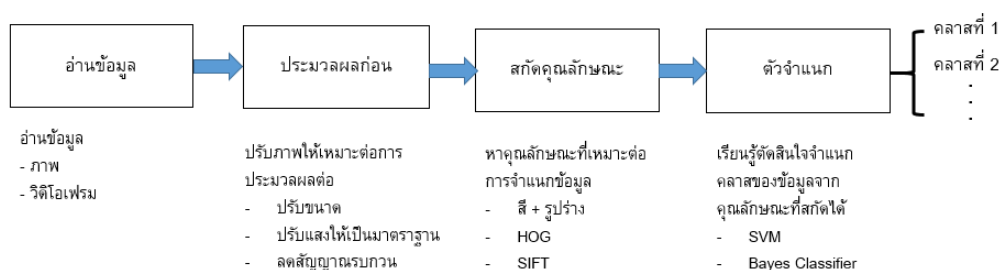
รูปที่ 1. 3 แสดงภาพแม่แบบหรือเทมเพลตและภาพอินพุตที่ใช้ในการจับคู่เทมเพลต

1.3.1 การจำแนกภาพแบบดั้งเดิม

นอกจากวิธีจับคู่แม่แบบแล้ว งานวิจัยด้านการรู้จำจำแนกวัตถุและภาพในช่วงต้น ๆ นิยมประมวลผลภาพอินพุตเพื่อสกัดหาคุณลักษณะหรือฟีเจอร์ (Feature) ที่สำคัญของวัตถุที่ต้องการจำแนก โดยต้องเป็นคุณลักษณะที่แตกต่างจากวัตถุอื่น ๆ ภายในภาพ ตัวอย่างเช่น การใช้สี (Color) เป็นคุณลักษณะในการจำแนกพริกแดงออกจากพริกเขียว เป็นต้น คุณลักษณะที่ใช้ในการจำแนกวัตถุขึ้นอยู่กับชนิดของวัตถุและภาพ คุณลักษณะที่นิยมใช้ในยุคเริ่มต้นเป็นคุณลักษณะแบบง่าย ได้แก่ สี ขอบภาพ (Edge) มุม (Corner) และค่าสัมประสิทธิ์ที่ได้จากการแปลงรูปแบบต่าง ๆ (Image transformation) การจำแนกภาพมีขั้นตอนการทำงานแสดงในรูปที่ 1.4 ซึ่งประกอบด้วยขั้นตอนหลักสามขั้นตอน ได้แก่

1. การประมวลผลก่อน (Pre-processing)
2. การสกัดคุณลักษณะ (Feature extraction) และ
3. ขั้นตอนจำแนก (Classification)

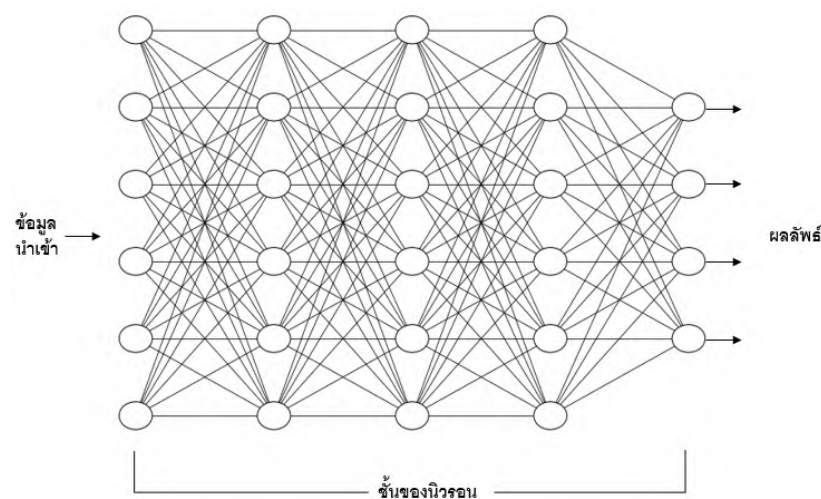
การทำงานเริ่มต้นจากการอ่านภาพอินพุต นำภาพที่รับเข้ามาผ่านขั้นตอนประมวลผลก่อน เพื่อปรับปรุงหรือปรับเปลี่ยนภาพให้เหมาะกับการประมวลผลในขั้นตอนถัดไป เช่น ทำการลดสัญญาณรบกวนภายในภาพ เป็นต้น จากนั้นนำภาพที่ผ่านการปรับปรุงเข้าขั้นตอนสกัดคุณลักษณะ ซึ่งทำหน้าที่สกัดคุณลักษณะเด่นภายในภาพอินพุต คุณลักษณะเด่นที่สกัดได้ถูกส่งเข้าขั้นตอนจำแนก เพื่อตัดสินใจจำแนกคลาสของภาพอินพุต โดยช่วงต้น ๆ ของงานวิจัย ขั้นตอนการจำแนกมักใช้วิธีการแบบง่าย เช่น เทคนิคขีดแบ่ง (Thresholding) ถ้าคุณลักษณะที่สกัดได้เกินค่าที่กำหนด ระบบตัดสินใจจำแนกเป็นภาพคลาส 1 ไม่เช่นนั้นเป็นภาพคลาส 0 เป็นต้น ต่อมาได้มีการนำศาสตร์การเรียนรู้ด้วยเครื่อง (Machine learning) มาประยุกต์ใช้งานทั้งในส่วนการสกัดคุณลักษณะเด่นและส่วนจำแนกข้อมูล ซึ่งตัวจำแนกข้อมูลที่นิยมใช้ ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และตัวจำแนกเบย์ (Bayes classifier) เป็นต้น



รูปที่ 1.4 บล็อกไดอะแกรมการจำแนกข้อมูลแบบดั้งเดิม ซึ่งคุณลักษณะที่สกัดออกแบบเบื้องต้นโดยมนุษย์

เทคนิคการจำแนกภาพที่ใช้วิธีการสกัดคุณลักษณะ และคุณลักษณะที่ใช้ในการจำแนก ถูกออกแบบเบื้องต้นโดยผู้พัฒนาระบบจำแนก เช่น ใช้สีและรูปร่างในการจำแนกภาพส้มและกล้วย เป็นต้น เทคนิค

การจำแนกในกลุ่มนี้ถูกจัดเป็น การจำแนกภาพแบบดั้งเดิม (Traditional image classification approach) หรือการจำแนกภาพแบบออกแบบด้วยมือ หรือวิศวกรรมลักษณะ (Handcrafted หรือ feature engineer approach) ซึ่งมีข้อเสียสำคัญ คือ ผู้ออกแบบต้องทำความเข้าใจภาพและวัตถุที่ต้องการจำแนก นอกจากนั้นยังมีข้อจำกัด เป็นวิธีที่ไม่เหมาะต่อการจำแนกภาพที่มีวัตถุหลากหลายภายในภาพ เนื่องจากบ่อยครั้งที่วัตถุต่างชนิดกันมีคุณลักษณะสำคัญรวมกัน เช่น การจำแนกพริกเขียวในถาดที่มีผักและผลไม้หลากหลายชนิด ถ้าออกแบบให้ใช้สีเขียวในการจำแนกพริก พื้นที่ ๆ ได้จากขั้นตอนสกัดสีเขียวมีหลายพื้นที่ ซึ่งอาจเป็นพื้นที่ของพริก ผักใบเขียว และผลไม้เปลือกเขียว ทำให้ต้องเพิ่มขั้นตอนอื่น เพื่อตัดวัตถุที่ไม่ต้องการทิ้ง หรือต้องสกัดคุณลักษณะสำคัญเพิ่มเติม เพื่อใช้คุณลักษณะร่วมกันหลายคุณลักษณะในการจำแนกแต่ละวัตถุภายในภาพ โดยทั่วไปคุณลักษณะที่สกัดได้เพียงคุณลักษณะเดียวมักไม่ดีพอที่ใช้แยกวัตถุหลายประเภทได้พร้อมกัน จึงได้มีการนำเทคนิคการเรียนรู้ด้วยเครื่องมาเรียนรู้คุณลักษณะสำคัญหรือหาความสัมพันธ์ภายในของพิกเซลของชุดภาพนำเข้าอย่างอัตโนมัติ

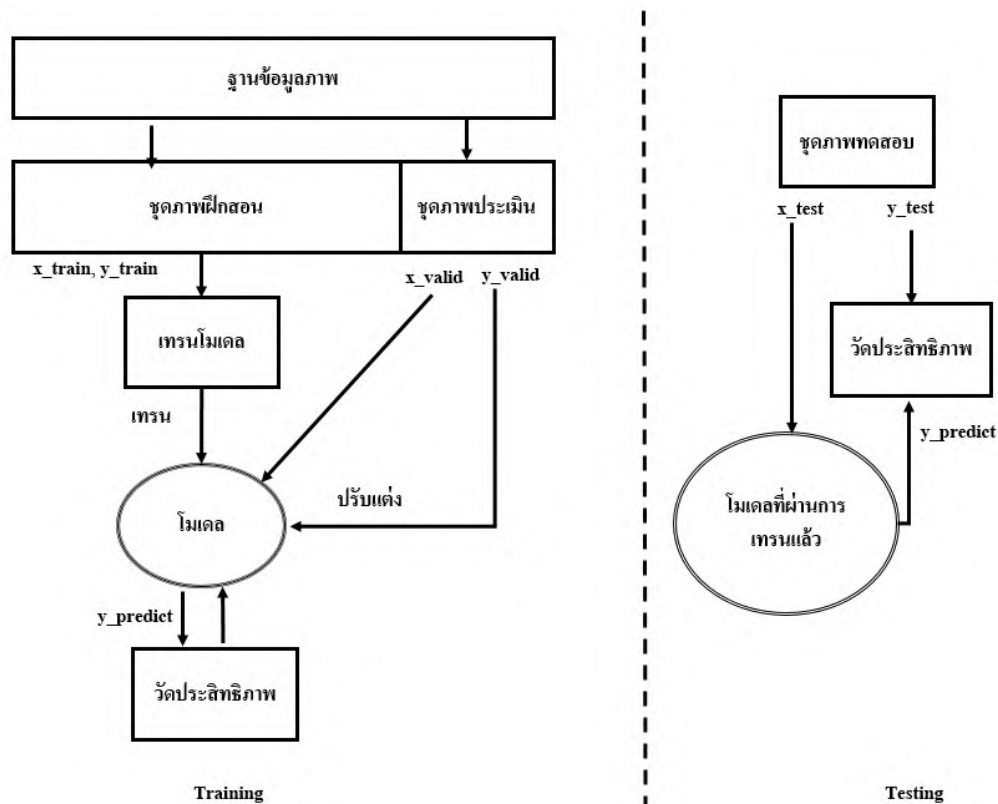


รูปที่ 1.5 ชั้นของนิวรอนในโครงข่ายประสาทเทียม NN

เริ่มต้นได้มีการนำเทคโนโลยีโครงข่ายประสาทเทียม (Artificial Neural Network: NN) ที่สามารถเรียนรู้ความสัมพันธ์ของข้อมูลนำเข้าได้อย่างอัตโนมัติ โดยนักวิจัยได้พัฒนาโครงข่ายประสาทเทียมที่อาจมีแนวคิดเริ่มต้นจากการจำลองเซลล์ประสาท หรือเรียกว่านิวรอน (Neurons) และกลไกการทำงาน การเชื่อมโยงระหว่างนิวรอนในสมองมนุษย์ สถาปัตยกรรมโครงข่ายประสาทเทียมแสดงตัวอย่างในรูปที่ 1.5 ซึ่งประกอบไปด้วยชั้นของนิวรอนหลายชั้นที่มีการเชื่อมต่อส่งข้อมูลระหว่างชั้นเพื่อให้นิวรอนในโครงข่ายเรียนรู้ความสัมพันธ์ของชุดข้อมูลนำเข้าผ่านชั้นต่าง ๆ ของนิวรอน

ในช่วงต้น โครงข่ายประสาทเทียมและเทคนิคการเรียนรู้ด้วยเครื่องอื่น ๆ ถูกนำมาประยุกต์ใช้ในการสกัดคุณลักษณะ โดยเทคนิคเหล่านี้ สร้างโมเดลซึ่งมักอยู่ในรูปสมการคณิตศาสตร์สำหรับการรู้จำ

จำแนกวัตถุต่าง ๆ โดยการสร้างโมเดลอาจสร้างโดยให้โครงข่ายการเรียนรู้สร้างโมเดลจากชุดตัวอย่างข้อมูล โดยการเรียนรู้ทั้งแบบมีผู้ฝึกสอน (Supervised learning) และไม่มีผู้ฝึกสอน (Unsupervised learning)



รูปที่ 1.6 การเรียนรู้แบบมีผู้ฝึกสอน (Supervised learning)

เนื้อหาในหนังสือเล่มนี้เน้นไปที่โครงข่ายประสาทเทียมที่มีการเรียนรู้แบบมีผู้ฝึกสอน ซึ่งผู้พัฒนาโครงข่ายต้องเตรียมฐานข้อมูลภาพ เริ่มต้นด้วยข้อมูลชุดฝึกสอน (Training set) สำหรับให้โครงข่ายใช้เรียนรู้ โดยชุดฝึกสอนประกอบไปด้วยข้อมูลและผลเฉลย (Label) ของข้อมูลนั้น ๆ เช่น ภาพแมวพร้อมผลเฉลยคลาสแมว (คลาส 0) ภาพสุนัขพร้อมผลเฉลยคลาสสุนัข (คลาส 1) เป็นต้น โมเดลจะผ่านการเรียนรู้เพื่อปรับพารามิเตอร์ต่าง ๆ ของโมเดลจากข้อมูลตัวอย่างในชุดฝึกสอน ในช่วงนี้เรียกว่าช่วงฝึกสอนหรือเทรน (Training) โมเดล โมเดลที่ได้หลังการฝึกสอนแต่ละครั้งถูกวัดประสิทธิภาพกับข้อมูลชุดประเมินผล (Validation set) โดยทั่วไปโมเดลจะผ่านการฝึกสอนหลายรอบเพื่อให้ได้โมเดลที่ให้ประสิทธิภาพดีที่สุดหรือได้ตามต้องการ โมเดลที่ได้นี้ถูกนำไปจำแนกข้อมูลกับข้อมูลชุดใหม่ที่โมเดลไม่เคยเจอมาก่อน เรียกข้อมูลชุดนี้ว่า ชุดทดสอบ (Test set) โดยเรียกการทำงานในส่วนทดสอบนี้ว่า การทำนาย (Predict) หรือการอนุมาน (Inference) หรือการทดสอบ (Testing) ภาพในชุดทดสอบเป็นภาพคนละชุดกับภาพในชุดฝึกสอนและชุดประเมิน เพื่อทดสอบศักยภาพของโมเดลในการปรับตัวให้กับภาพในงานจริง รูปที่ 1.6 แสดงขั้นตอนการจำแนกภาพด้วยการเรียนรู้แบบมีผู้ฝึกสอน

เนื่องจากโมเดลที่ได้อาศัยการเรียนรู้หาความสัมพันธ์และรูปแบบ (Pattern) จากชุดข้อมูลตัวอย่างในฐานข้อมูลชุดฝึกสอน ภาพที่ใช้ในการฝึกสอนจึงมีความสำคัญและส่งผลต่อประสิทธิภาพของโมเดลที่สร้างขึ้นโดยตรง ตัวอย่าง เช่น กรณีสร้างโมเดลจำแนกตัวเลข 0-9 ถ้าภาพในฐานข้อมูลชุดฝึกเป็นภาพเลข 0-9 จากการพิมพ์โดยไม่มีชุดภาพตัวอย่างตัวเลขจากการเขียนลายมือจริงเลย โมเดลที่ได้จากการฝึกสอน เรียนรู้รูปแบบและจำแนกตัวเลขจากการพิมพ์เพียงอย่างเดียว การนำโมเดลไปจำแนกตัวเลขจากภาพลายมือจริงจะให้ค่าผิดพลาดที่สูง ในการฝึกสอนจำนวนของข้อมูลตัวอย่างจึงควรมีจำนวนมากพอสำหรับให้โมเดลเรียนรู้ และควรครอบคลุมชนิดของวัตถุที่ต้องการจำแนก ในกรณีที่ต้องการสร้างโมเดลที่สามารถจำแนกภาพหรือวัตถุได้หลากหลายประเภท ชนิดและจำนวนภาพในฐานข้อมูลชุดฝึกสอนต้องประกอบไปด้วยภาพของวัตถุคลาสต่าง ๆ จำนวนมาก ทำให้พบปัญหาในการประยุกต์ใช้โครงข่ายประสาทเทียมในยุคต้น ๆ เนื่องจากขาดฐานข้อมูลขนาดใหญ่สำหรับฝึกสอนโมเดล และการฝึกสอนโมเดลใช้เวลานาน ทำให้ประสิทธิภาพการจำแนกข้อมูลของระบบที่พัฒนาขึ้นยังไม่สูงเพียงพอที่จะนำไปประยุกต์ใช้ในงานจริงได้อย่างกว้างขวาง

1.3.2 การจำแนกภาพโดยใช้โครงข่ายคอนโวลูชัน

การพัฒนางานวิจัยด้านการรู้จำและจำแนกภาพเริ่มได้รับความสนใจแบบก้าวกระโดด เมื่อ Lecun และคณะ [8] ได้นำเสนอโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional neural network) หรือเรียกสั้น ๆ ว่าโครงข่าย CNN สำหรับจำแนกตัวเลขลายมือเขียน โครงข่าย CNN เป็นโครงข่ายประสาทเทียมประเภทหนึ่ง ประกอบไปด้วยชั้นคอนโวลูชันเรียงต่อกัน ดังแสดงในรูปที่ 1.7 ชั้นคอนโวลูชันใช้ตัวดำเนินการคอนโวลูชันและตัวกรอง (Filters) ซึ่งมีลักษณะการทำงานคล้ายการมองเห็นของมนุษย์ ที่แบ่งภาพที่มองเห็นออกเป็นพื้นที่ย่อย จากนั้นเรียนรู้คุณลักษณะภายในพื้นที่ย่อย ทำการเชื่อมโยงคุณลักษณะของพื้นที่ย่อยที่อยู่ใกล้กัน คุณลักษณะที่สกัดด้วยชั้นคอนโวลูชันจึงเหมาะต่องานประมวลผลภาพและงานจำแนกวัตถุภายในภาพ โดยเรียกคุณลักษณะที่ได้ในแต่ละชั้นคอนโวลูชันว่า **แผนที่คุณลักษณะ** (Feature maps) ซึ่งแผนที่คุณลักษณะในแต่ละชั้นถูกเชื่อมโยงผ่านชั้นต่าง ๆ ภายในโครงข่ายก่อนถูกส่งต่อไปยังชั้นจำแนก เพื่อตัดสินใจว่าสิ่งที่มองเห็นเป็นวัตถุคลาสไหน

ที่มวิจัย [8] นำเสนอโครงข่ายคอนโวลูชันชื่อ LeNet ที่สร้างโมเดลจำแนกตัวเลขลายมือเขียนจากภาพฐานข้อมูลที่ประกอบด้วยภาพตัวเลขลายมือเขียน 7,291 ภาพ และตัวเลขพิมพ์ 2,549 ภาพ จากการทดลองพบว่าการฝึกสอนโครงข่าย CNN ทำได้ง่าย มีจำนวนพารามิเตอร์ที่โมเดลต้องเรียนรู้้น้อยกว่าโครงข่าย NN แบบเดิม ทำให้เหมาะต่อการนำไปประยุกต์ใช้กับฐานข้อมูลที่มีขนาดใหญ่ขึ้น นอกจากนั้นเมื่อเปรียบเทียบประสิทธิภาพ พบว่าโครงข่าย CNN จำแนกตัวเลขได้ดีกว่าโครงข่ายที่ได้พัฒนามาก่อนหน้าแบบมีนัยสำคัญ ทำให้นักวิจัยในลำดับต่อมาให้ความสนใจกับการพัฒนาโครงข่ายคอนโวลูชันสำหรับงานจำแนกและตรวจจับวัตถุเพิ่มขึ้น และจัดกลุ่มเทคนิคที่ใช้โครงข่าย

CNN ในการสกัดคุณลักษณะเป็น การจำแนกภาพโดยใช้โครงข่ายคอนโวลูชัน (CNN-based Image classification approach)

โครงข่ายคอนโวลูชันจัดเป็นโครงข่ายประสาทเทียมชนิดหนึ่งที่มีการประมวลผลด้วยตัวกรองและการทำคอนโวลูชัน การอ้างอิงโครงข่ายประสาทเทียมในหนังสือเล่มนี้หมายถึงโครงข่ายประสาทเทียมแบบดั้งเดิมและแบบคอนโวลูชัน และจากงานวิจัยในลำดับต่อมา [9-11] พบว่าประสิทธิภาพในการเรียนรู้โมเดลโครงข่ายประสาทเทียมขึ้นอยู่กับสถาปัตยกรรมของโครงข่ายที่ออกแบบ โดยความลึก (จำนวนชั้น) และความกว้าง (จำนวนนิวรอนสูงสุดในแต่ละชั้น) ของโครงข่ายมีผลต่อประสิทธิภาพในการจำแนกข้อมูลแต่การพัฒนาโครงข่ายประสาทเทียมขนาดใหญ่ ต้องการทรัพยากรในการประมวลผลสูงมาก การฝึกสอนใช้หน่วยความจำจำนวนมากและใช้เวลานาน รวมทั้งต้องสร้างฐานข้อมูลภาพขนาดใหญ่มารองรับการพัฒนาในงานวิจัยด้านนี้

ฐานข้อมูลขนาดใหญ่ที่เป็นที่รู้จัก ได้แก่ ฐานข้อมูล CIFAR-10/100 [12], PASCAL VOC [13], และ ImageNet [6] โดยในฐานข้อมูล CIFAR-10 ประกอบไปด้วยภาพสี่ของวัตถุจำนวน 10 คลาส ได้แก่ เครื่องบิน รถยนต์ นก แมว หมา กวาง กบ ม้า เรือ รถบรรทุก วัตถุแต่ละคลาสมีภาพ 6,000 ภาพ แบ่งเป็นภาพชุดฝึกสอน 5,000 ภาพและภาพชุดทดสอบ 1,000 ภาพ ส่วนฐานข้อมูล CIFAR-100 มีภาพของวัตถุ 100 คลาส ๆ ละ 600 ภาพ (500 ภาพสำหรับฝึกสอนและ 100 ภาพสำหรับทดสอบ) PASCAL VOC เป็นฐานข้อมูลภาพจำนวน 19,737 ภาพ ฐานข้อมูลจัดทำขึ้นเพื่อใช้ในงานจำแนกตรวจจับและแบ่งส่วนภาพของวัตถุ 20 ชนิด และต่อมาได้มีการสร้างฐานข้อมูล ImageNet ซึ่งเป็นฐานข้อมูลเก็บภาพตัวอย่างของวัตถุหลายชนิด โดยเริ่มต้นเก็บภาพของวัตถุ 1,000 คลาส จนถึงปัจจุบันที่มีภาพของวัตถุต่าง ๆ มากถึง 22,000 คลาส โดยมีภาพตัวอย่างพร้อมผลเฉลยระบุคลาสของวัตถุในภาพมากกว่า 14 ล้านภาพ โดยงานวิจัยทั่วไปนิยมใช้ฐานข้อมูลย่อยชื่อ ImageNet-1K หรือ ILSVRC2017 ในการสร้างและทดสอบโมเดล ฐานข้อมูลย่อยประกอบไปด้วยภาพของวัตถุ 1,000 คลาส มีภาพในชุดฝึกสอน 1,281,167 ภาพ ชุดประเมินผล 50,000 ภาพ และชุดภาพทดสอบ 100,000 ภาพ [14]

การพัฒนาโครงข่ายคอนโวลูชันได้รับความสนใจอย่างก้าวกระโดดอีกครั้ง เมื่อมีการนำโครงข่ายคอนโวลูชันไปทำงานบนหน่วยประมวลผลด้านกราฟิก (Graphic Processing Unit: GPU) ที่ช่วยเพิ่มประสิทธิภาพในการคำนวณด้านกราฟิก โดยตัวดำเนินการหนึ่งในนั้น คือ การทำคอนโวลูชัน ส่งผลให้นักวิจัยสามารถออกแบบและทดสอบโครงข่าย CNN ที่มีความลึกและมีสถาปัตยกรรมที่ซับซ้อนเพิ่มขึ้นได้ การพัฒนาที่รวดเร็วนี้ส่งผลให้ได้ระบบและโมเดลที่มีประสิทธิภาพสูงขึ้น สามารถจำแนกและตรวจจับวัตถุได้หลากหลายประเภทมากขึ้น งานวิจัยในช่วงหลัง ๆ จึงเน้นไปที่การพัฒนาโครงข่ายคอนโวลูชัน ที่มีการต่อชั้นหลาย ๆ ชั้น สถาปัตยกรรมของโครงข่ายมีความลึกเพิ่มขึ้น เกิดการพัฒนาศาสตร์ในสาขาการเรียนรู้เชิงลึกโดยการเรียนรู้เชิงลึกเป็นการเรียนรู้ของโครงข่ายประสาทเทียมที่มี