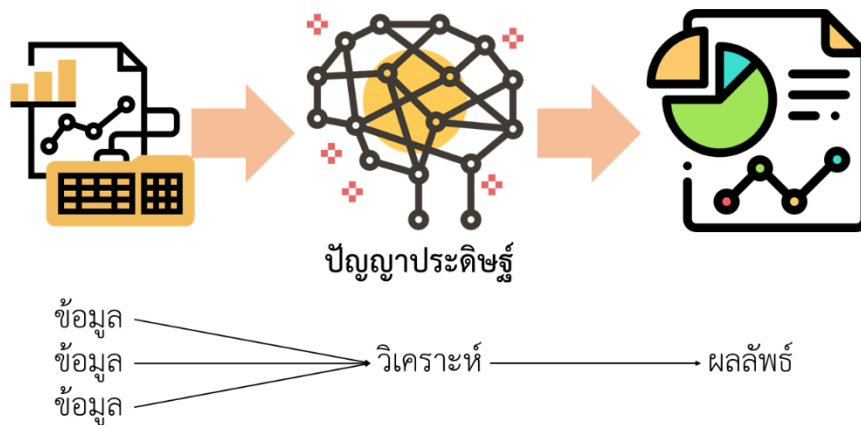


บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการจัดการทำวิทยานิพนธ์ในด้านปัญญาประดิษฐ์และการเรียนรู้ของเครื่อง มีความจำเป็นอย่างยิ่งที่ผู้จัดทำจะต้องทราบถึงหลักการทางเทคโนโลยีและระบบงานที่เกี่ยวข้อง เพื่อใช้เป็นแนวทางและหลักอ้างอิงในการศึกษา เพื่อให้การจัดทำวิจัยเกิดขึ้นได้ ทางผู้จัดทำจึงได้ทำการศึกษาหลักการทางทฤษฎีและระบบงานที่เกี่ยวข้องดังนี้

2.1 เทคโนโลยีที่ใช้ในการวิจัย



ภาพประกอบที่ 2-1 อธิบายการทำงานของปัญญาประดิษฐ์

2.1.1 AI (Artificial Intelligence)

ปัญญาประดิษฐ์คือสิ่งที่มนุษย์สร้างขึ้นมาเพื่อให้สามารถคิดและทำอะไรบางอย่างเองแทนมนุษย์ได้โดยไม่ต้องมีมนุษย์มาสั่ง ไม่จำเป็นต้องป้อนคำสั่งหรือโปรแกรมใด ๆ ให้กับปัญญาประดิษฐ์ การทำงานที่เกิดจากการเรียนรู้และวิเคราะห์ข้อมูลและได้ผลลัพธ์ออกมาเองปัญญาประดิษฐ์คือสิ่งที่พยายามทำให้คอมพิวเตอร์เหมือนมนุษย์ทำสิ่งต่าง ๆ ผ่านการเรียนรู้เหมือนมนุษย์ การเรียนรู้จึงเป็นสิ่งสำคัญสำหรับปัญญาประดิษฐ์ [1]

2.1.1.1 Acting Humanly การกระทำคล้ายมนุษย์

สื่อสารกับมนุษย์ได้ด้วยภาษาที่มนุษย์ใช้ เช่น ภาษาอังกฤษ เป็นการประมวลผลภาษาธรรมชาติ (natural language processing) เช่น การใช้เสียงสั่งให้คอมพิวเตอร์พิมพ์เอกสาร

- มีประสาทรับสัมผัสคล้ายมนุษย์ โดยสามารถมองเห็นและรับภาพได้โดยใช้อุปกรณ์รับสัญญาณภาพ (sensor)

- Machine learning เรียนรู้ด้วยตัวเองโดยสามารถตรวจจับรูปแบบการเกิดของเหตุการณ์ใด ๆ แล้วปรับตัวสู่สิ่งแวดล้อมที่เปลี่ยนไปได้

2.1.1.2 Thinking Humanly การคิดคล้ายมนุษย์

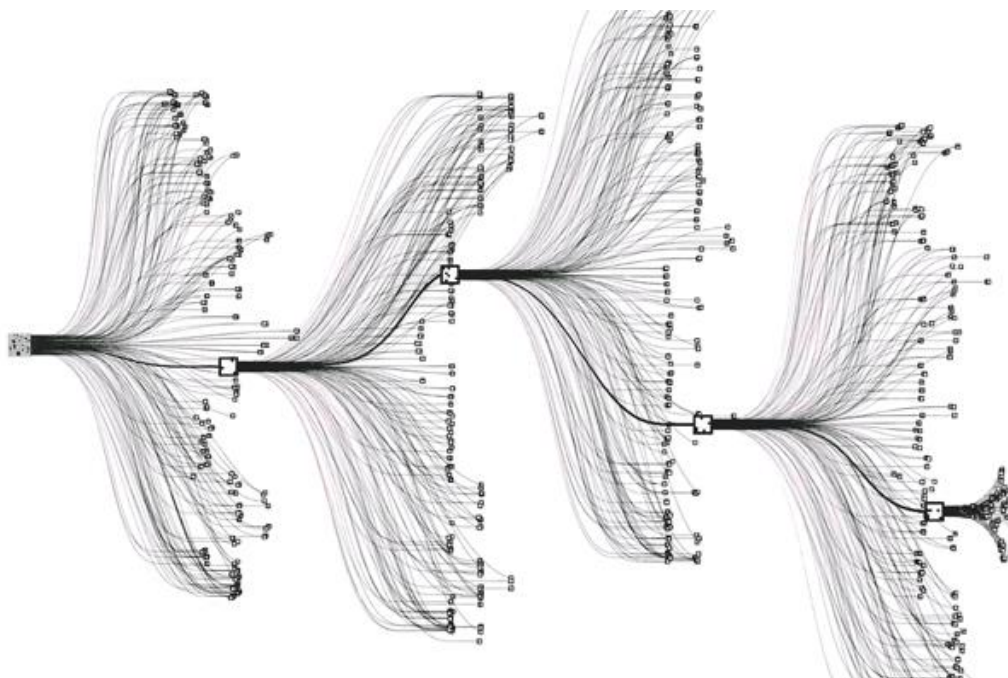
ก่อนที่จะทำให้เครื่องคิดอย่างมนุษย์ได้ ต้องรู้ก่อนว่ามนุษย์มีกระบวนการคิดอย่างไร ซึ่งการวิเคราะห์ลักษณะการคิดของมนุษย์เป็นศาสตร์ด้าน cognitive science เช่น ศึกษาโครงสร้างสามมิติของเซลล์สมอง การแลกเปลี่ยนประจุไฟฟ้าระหว่างเซลล์สมอง วิเคราะห์การเปลี่ยนแปลงทางเคมีไฟฟ้าในร่างกายระหว่างการคิด

2.1.1.3 Thinking rationally คิดอย่างมีเหตุผล

คิดอย่างมีเหตุผล หรือคิดถูกต้อง โดยใช้หลักตรรกศาสตร์ในการคิดหาคำตอบอย่างมีเหตุผล

2.1.1.4 Acting rationally กระทำอย่างมีเหตุผล

กระทำอย่างมีเหตุผล โดยจะตอบสนองต่อการกระทำแต่ละแบบโดยการนำเอาสภาพแวดล้อมเข้ามาวิเคราะห์ร่วมด้วย



ภาพประกอบที่ 2-2 รูปแบบการทำงานของ Deepmind's AlphaGo

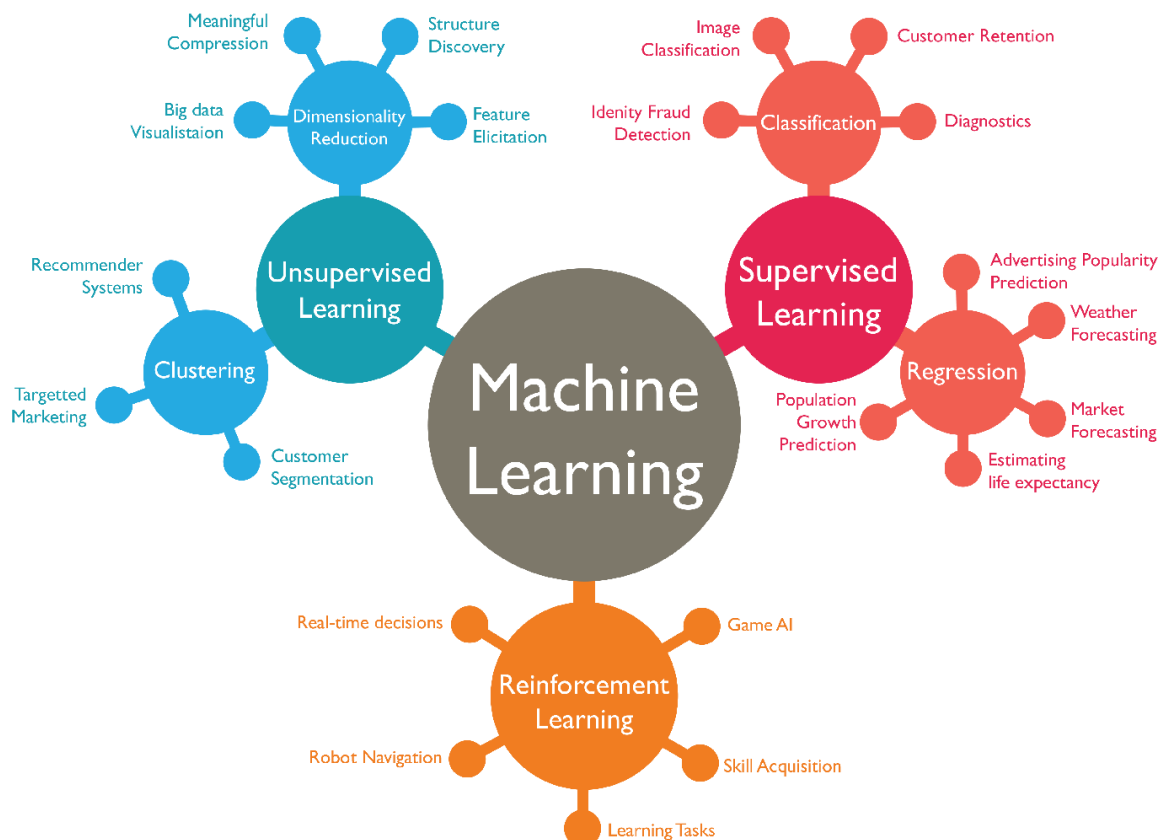
ตัวอย่างล่าสุดคือ AI ของ Google (Deepmind's AlphaGo) สามารถทำงานโดยการรับรู้และแยกแยะภาพและวิเคราะห์ค่าต่าง ๆ ว่าคือภาพหรือค่านั้นหมายถึงอะไรได้ นอกจากความสามารถในการ

แยกแยะและการรับรู้แล้ว ยังมีความสามารถในการทำงานโดยการเรียนรู้ ซึ่งแผนกวิจัยของ Google ได้แสดงให้เห็นถึงการเรียนรู้ของหุ่นยนต์ในการเล่นเกมส์ที่ไม่เคยเล่นมาก่อน และ AI ก็ใช้เวลาไม่นานในการเรียนรู้การเล่นเกมนั้น จนในที่สุด AI ก็มีกลยุทธ์ในการเล่นเกมส์ต่าง ๆ จนสามารถเอาชนะได้อย่างรวดเร็ว [2]

2.1.2 Machine Learning

การทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง โดยใช้ ข้อมูล Machine Learning แบ่งได้ 3 แบบ [3]

- 1) Supervised Learning เรียนรู้โดยมี data มาสอน
- 2) Unsupervised Learning เรียนรู้โดยไม่มี data สอน
- 3) Reinforcement Learning เรียนรู้ตามสภาพแวดล้อม



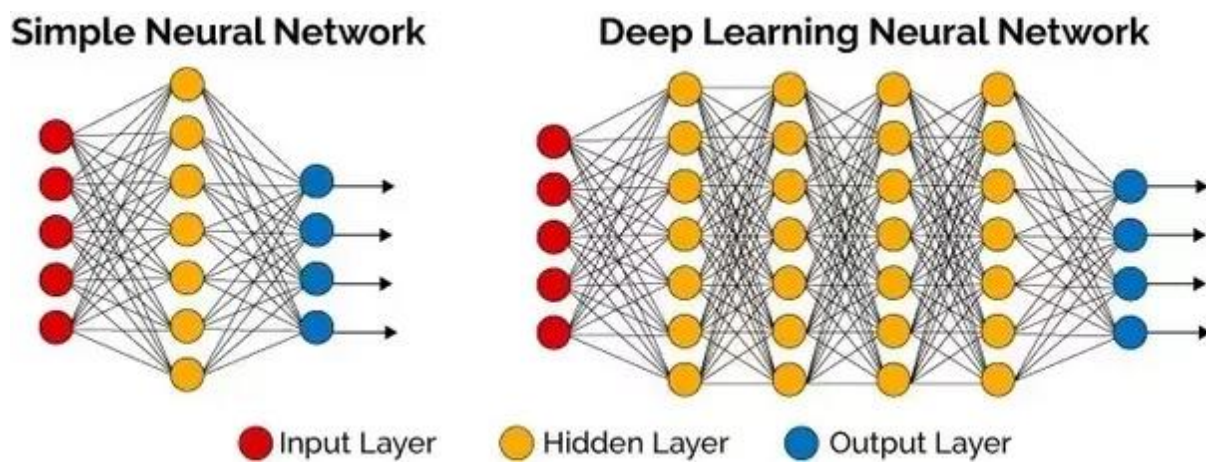
ภาพประกอบที่ 2-3 Machine Learning

Supervised Learning

คือการเรียนรู้ โดยมี data มาสอน ตัวอย่างเช่น เด็กชายต้องไปสอบแยกแยะประเภทหมาแมวเราจึงชี้ให้เด็กชาย รู้จักแมว ชี้ให้รู้จักหมา หลายๆครั้งจนเด็กจำได้แล้วจึงอุ้มแมวมาถามเด็กว่า นี้อะไร? เด็กก็จะแยกแยะได้ ว่าเป็นแมว คอมพิวเตอร์ก็ต้องเรียนรู้เช่นเดียวกัน

2.1.3 Deep Learning

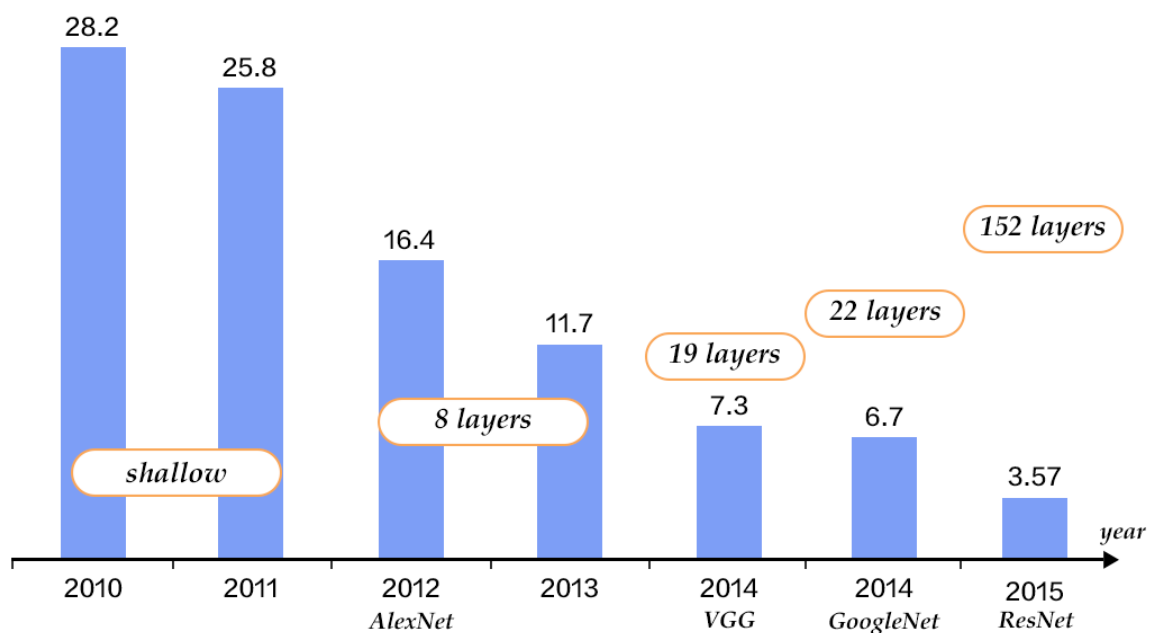
Deep Learning เป็น Machine Learning ชนิดหนึ่ง ที่ใช้การเรียนรู้ของข้อมูลโดยอาศัยความเข้าใจเป็นลำดับขั้น เพื่อให้ model เรียนรู้ความสัมพันธ์ที่มีความซับซ้อนได้ ซึ่งสามารถทำได้โดยการใช้ Deep neural network (Neural Network ที่มี hidden layer จำนวนมาก) ดังรูปที่ 2-4[4]



ภาพประกอบที่ 2-4 Neural Network และ Deep Learning

2.1.3.1 ความเป็นมาของ Deep Learning

Deep Learning เป็นที่รู้จักจากงานวิจัยที่ชื่อว่า ImageNet Classification with Deep Convolutional Neural Networks (AlexNet) ที่ได้ทำการ Classify รูปภาพ 1,200,000 รูปที่อยู่ใน Imagenet (Database ของรูปภาพที่จัดทำโดย Stanford) Imagenet จะให้นักวิจัยด้าน Machine Learning ส่งอัลกอริทึม เข้ามาแข่งขันว่าใครจะ Classify รูปภาพได้มีความผิดพลาดน้อยที่สุดและในปี 2012 Deep Learning ได้ทำ Error rate ได้น้อยที่สุด โดยได้ Error rate ที่ 15.3 % ห่างจากอันดับ 2 ถึง 10.9 % โดยอันดับ 2 ทำได้เพียง 26.2 % เลยทำให้นักวิจัยได้หันมาให้ความสนใจกับ Deep Learning

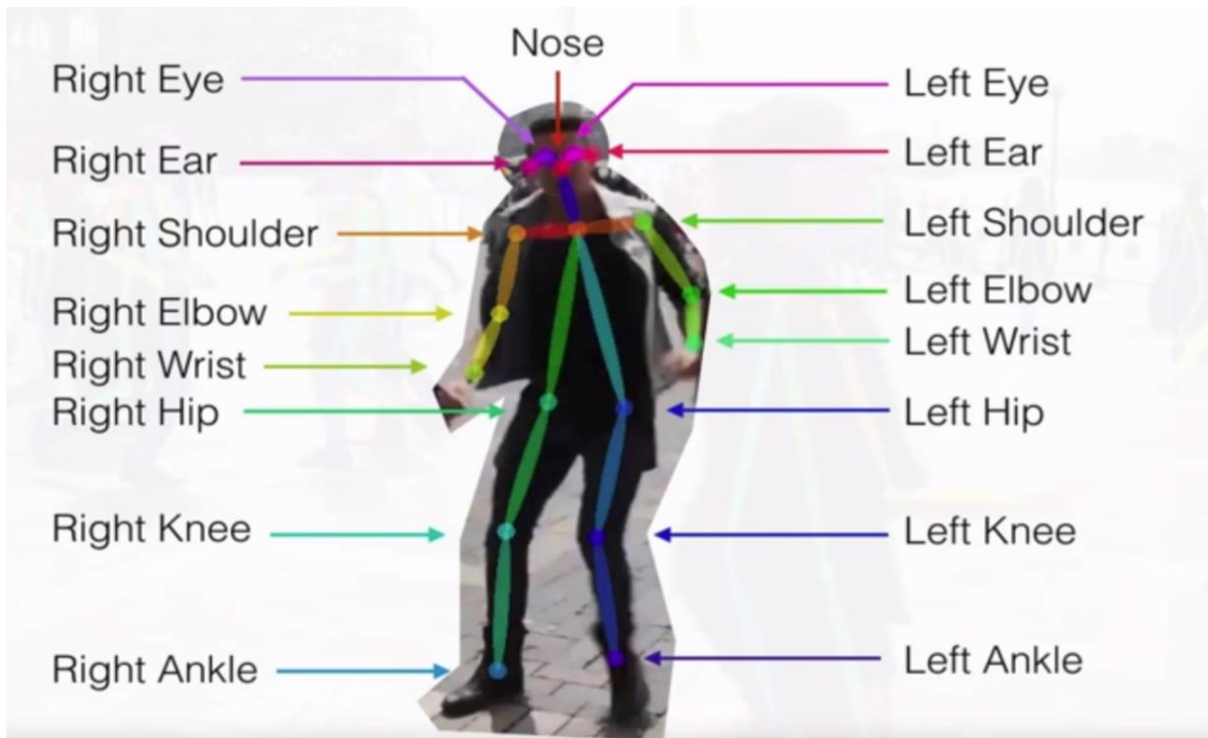


ภาพประกอบที่ 2-5 แสดง Error rate จากการแข่งขัน Imagenet

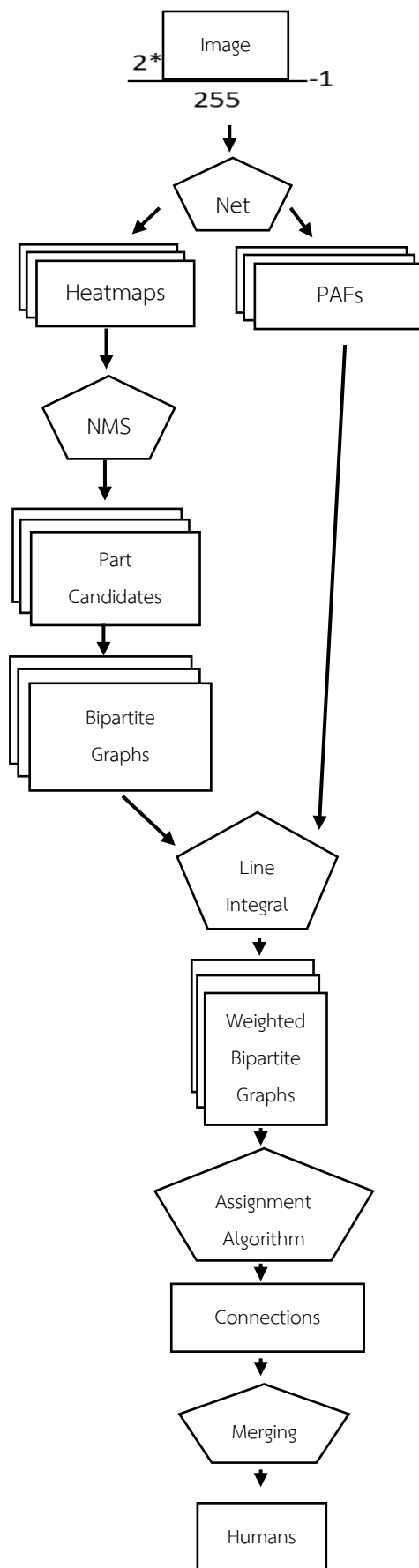
จากภาพประกอบที่ 2-5 ตั้งแต่ปี 2012 ที่มีการนำ Deep Learning มาใช้ Classify รูปภาพในการแข่งขันของ Imagenet ทำให้การแข่งขันปีหลังจากนั้นเป็นต้นมา Deep Learning จึงเป็นอัลกอริทึมหลักที่นักวิจัยเลือกใช้และเมื่อ Deep Learning ทำให้ Error rate ลดลงจนต่ำกว่า 10% จนได้ Error rate ที่นักวิจัยยอมรับว่าสามารถนำมาใช้งานจริงได้ บริษัทเทคโนโลยีต่าง ๆ จึงได้นำไปพัฒนาต่อยอดใช้กับงานด้านต่าง ๆ มากมาย

2.1.4 Openpose

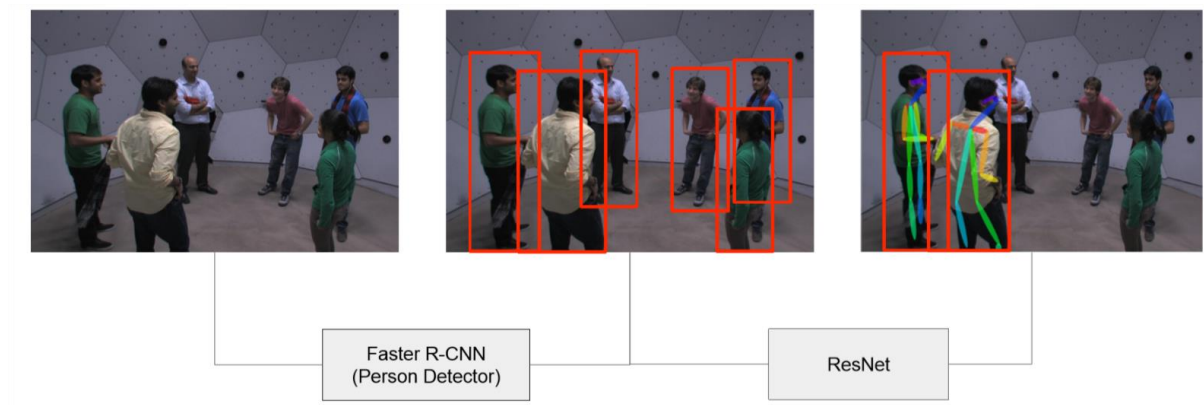
Openpose เป็น open-source library ที่ใช้ในการตรวจจับ key point หลายคนแบบเรียลไทม์ (real-time multi-person key point detection) พัฒนาโดย Gines Hidalgo, Zhe Cao, Tomas Simon, Shih-En Wei, Hanbyul Joo and Yaser Sheikh. ช่วยให้สามารถตรวจจับ 18 จุดสำคัญของร่างกายจากภาพถ่าย วิดีโอ และ Webcam



ภาพประกอบที่ 2-6 การทำงานของ Openpose

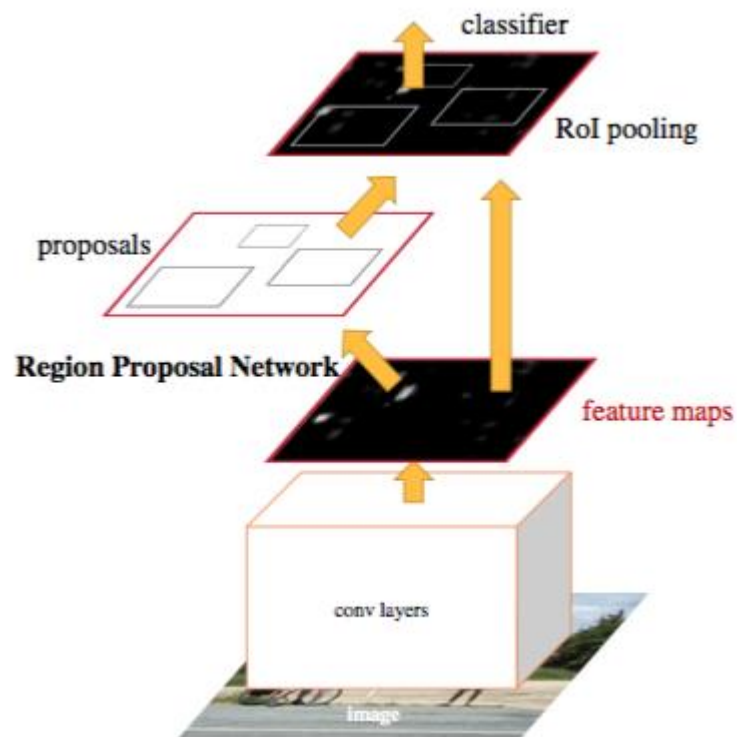


ภาพประกอบที่ 2-7 ขั้นตอนการทำงานของ Openpose



ภาพประกอบที่ 2-8 Top-down Approach: Person Detection + Pose Estimation

2.1.4.1 Faster R-CNN

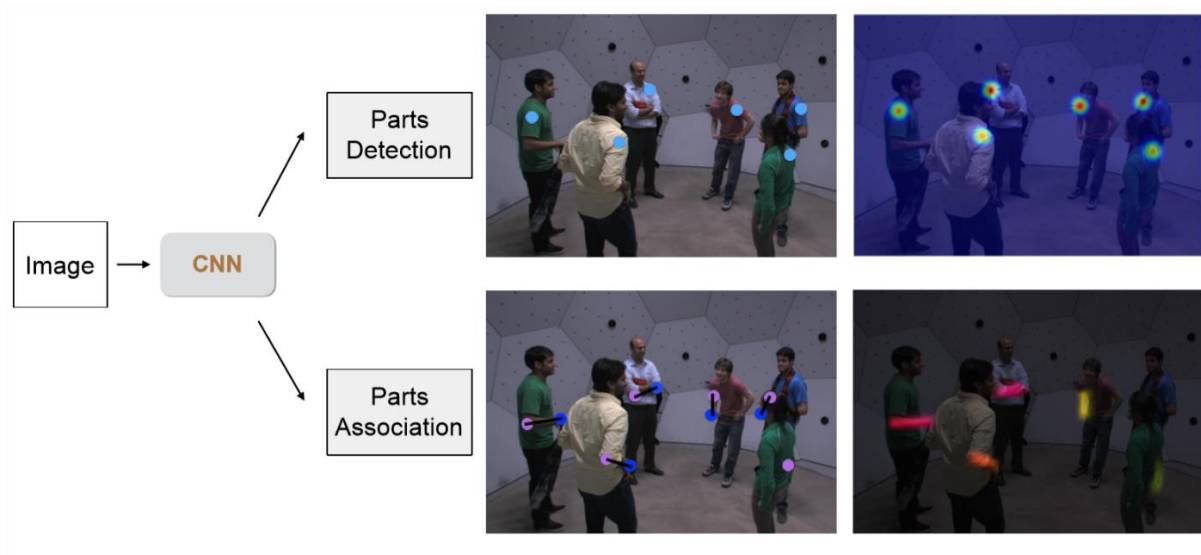


ภาพประกอบที่ 2-9 Faster R-CNN

Faster R-CNN คือ แนวความคิดที่ใช้จำแนกและตรวจจับวัตถุ ที่ได้รับการพัฒนาต่อยอดมาจาก R-CNN และ fast-R-CNN Selective Search ที่ช้าและ train ไม่ได้ Faster-RCNN เลยออกมาเพื่อพัฒนาจุดตรงนี้โดยรวม region proposal เข้าไป train เป็นส่วนหนึ่งของ CNN[5]

2.1.4.2 ResNet (Residual Network)

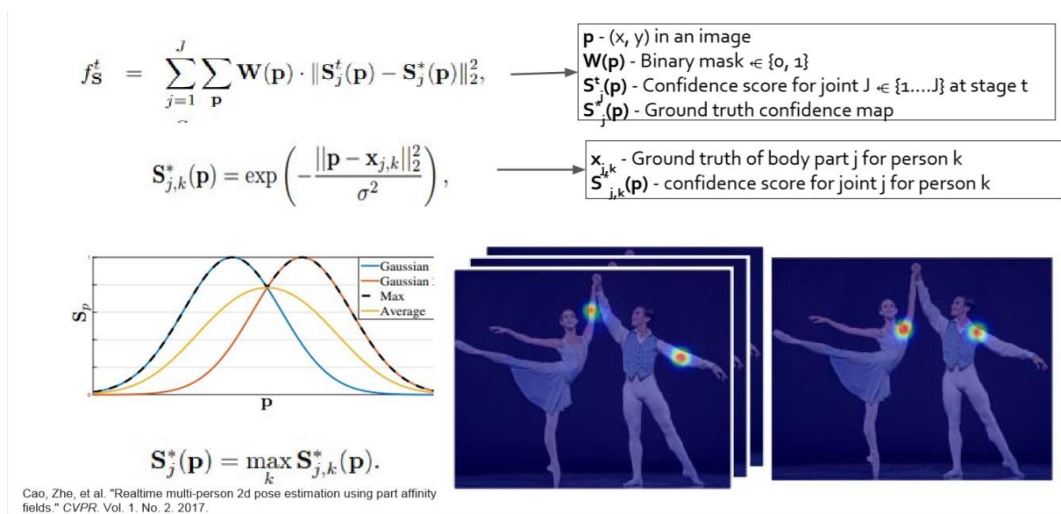
ResNet คือวิธีแก้ปัญหาเมื่อ CNN ที่มีหลายชั้น เมื่อชั้นไหนไม่จำเป็นจริง ๆ ค่าถ่วงน้ำหนักบนชั้นเหล่านั้นก็จะถูก train ให้เข้าใกล้ 0 ไปเอง ข้อมูลที่จำเป็นก็ยังสามารถ flow ผ่าน skip connection ไปชั้นต่อไปได้ นี่คือการมาของชื่อ Residual Network หรือ ResNet



ภาพประกอบที่ 2-10 Bottom-up Approach: Parts Detection and Parts Association

2.1.4.3 Parts Detection and Parts Association

ใช้อัลกอริทึม เสนอบริเวณที่น่าจะมีวัตถุอยู่ ถูกสร้างขึ้นโดยใช้ความรู้จากการรับรู้โดยการมองเห็นของมนุษย์ โดยใช้ feature ต่าง ๆ เช่น สี edge gradient การกระจายตัวของสี หรือ feature อื่น ๆ เพื่อทำการเสนอว่า ตรงไหนน่าจะมีวัตถุอยู่ [6] [7]



ภาพประกอบที่ 2-11 Parts Detection

$L = (L_1, L_2, \dots, L_C)$, L_i is a vector field - for each limb (c)

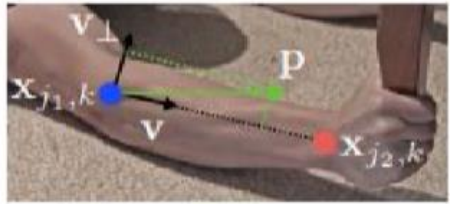
$$f_L^t = \sum_{c=1}^C \sum_p W(p) \cdot \|L_c^t(p) - L_c^*(p)\|_2^2,$$

$$L_{c,k}^*(p) = \begin{cases} v & \text{if } p \text{ on limb } c, k \\ 0 & \text{otherwise.} \end{cases}$$

$0 \leq v \cdot (p - x_{j_1,k}) \leq l_{c,k}$ and $|v_{\perp} \cdot (p - x_{j_1,k})| \leq \sigma_l$,

$$L_c^*(p) = \frac{1}{n_c(p)} \sum_k L_{c,k}^*(p),$$

v - normalized unit vector along $x_{j_1,k}x_{j_2,k}$
 $l_{c,k}$ - distance between $x_{j_1,k}x_{j_2,k}$
 σ_l - limb width



$n_c(p)$ - No of non-zero vectors at point p for all k people

Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." CVPR. Vol. 1. No. 2. 2017.

ภาพประกอบที่ 2-12 Part Affinity Fields for Part Association

2.1.5 ภาษา python

Python คือชื่อภาษาที่ใช้ในการเขียนโปรแกรมภาษาหนึ่ง ซึ่งถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux , Windows NT, Windows 2000, Windows XP หรือแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาตัว นี้เป็น Open Source เหมือนอย่าง PHP ทำให้ทุกคนสามารถที่จะนำ Python มาพัฒนาโปรแกรมของเราได้ฟรีๆโดยไม่ต้องเสียค่าใช้จ่าย และความเป็น Open Source ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ Python มีความสามารถสูงขึ้น และใช้งานได้ครบครันกับทุกลักษณะงานโค้ดของ Python ถูกสร้างขึ้นมาจากภาษาซี การประมวลผลจะทำในแบบอินเทอร์พรีเตอร์ คือจะประมวลผลไปทีละบรรทัดและปฏิบัติตามคำสั่งที่ได้รับ Python เวอร์ชันแรกคือ เวอร์ชัน 0.9.0 ออกมาเมื่อปี 2533 และเวอร์ชันปัจจุบันคือ 3.5 [8]

2.1.6 Weight training

weight training คือการออกกำลังกายโดยใช้น้ำหนัก ไม่ว่าจะเป็นใช้ อุปกรณ์ หรือใช้น้ำหนักจากตัวผู้ฝึกเองโดยการออกกำลังกายด้วยน้ำหนักจำเป็นต้องมีการฝึกท่าที่ถูกต้องเพื่อไม่ให้เกิดการบาดเจ็บ และช่วยให้พัฒนาได้ดีที่สุดโดยการออกกำลังกายด้วยน้ำหนักจุดประสงค์หลักจะเน้นไปที่การพัฒนากล้ามเนื้อ เมื่อเรามีกล้ามเนื้อที่ดี เราจะสามารถต่อยอดไปกีฬาชนิดอื่นได้ง่าย [9]

2.2 งานวิจัยที่เกี่ยวข้อง

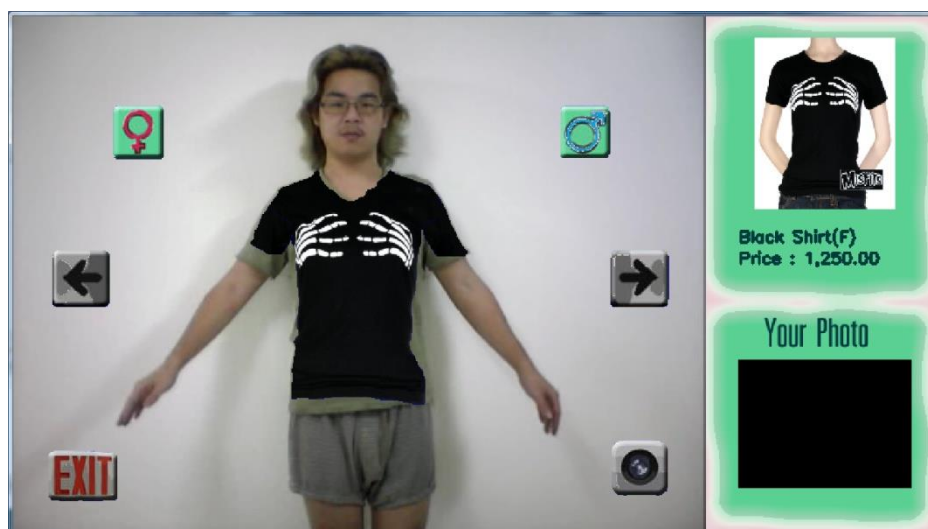
2.2.1 Virtual Fitting Room Using Augmented Reality

โครงการห้องลองเสื้อเสมือนโดยใช้ออกเมนต์เตดเรียลลิตี้ เป็นโปรแกรมจำลองการลองเสื้อที่พัฒนาขึ้นเพื่อแก้ไขปัญหาของการลองเสื้อ ที่สามารถนำไปใช้งานได้ง่าย ซึ่งผลการทดสอบความผิดพลาดของการประมาณตำแหน่งศีรษะอยู่ที่ 24.76 pixelต่อภาพ การประมาณตำแหน่งลำตัวอยู่ที่ 30.89 pixel ต่อภาพ และการประมาณทิศทางของแขนท่อนบนอยู่ที่ 13.46 องศาต่อภาพ ซึ่งถือว่ายังไม่เป็นที่น่าพึงพอใจ ยังสามารถพัฒนาเพิ่มต่อไปได้ โดยยังพบปัญหาในกรณีที่มีความหลากหลายของสภาพแสงและท่าทางของผู้ใช้งานที่เกินข้อจำกัดของโปรแกรม แต่เมื่อทดสอบดูคุณภาพจากการใช้งานจริงแล้ว ความผิดพลาดที่เกิดขึ้นยังถือว่ายอมรับได้ คุณภาพของภาพและผลการทำงาน ถือว่าอยู่ในเกณฑ์ที่สามารถนำไปใช้งานได้ และสามารถให้ผู้ที่สนใจนำไปใช้งานได้จริง [10]

นักประดิษฐ์ นาย ดนัยนันท์ เก่าเงิน, นาย ณัชนนท์ วงษ์วีไล

สถานที่วิจัย จุฬาลงกรณ์มหาวิทยาลัย

ตีพิมพ์เมื่อ 2553



ภาพประกอบที่ 2-13 โครงการห้องลองเสื้อเสมือนโดยใช้ออกเมนต์เตดเรียลลิตี้

2.2.2 Through-Wall Human Pose Estimation Using Radio Signals

เป็นงานวิจัยเกี่ยวกับการใช้ AI ในการวิเคราะห์ท่าทางของมนุษย์ และ คาดเดาท่าทางเมื่อมีสิ่งกีดขวาง หรือก็คือการใช้ AI วิเคราะห์ท่าทางเมื่อ เราอยู่หลังกำแพง [11]

นักประดิษฐ์ Mingmin Zhao, Tianhong Li, Mohammad Abu, Alsheikh Yonglong Tian, Hang Zhao, Dina Katabi, Antonio Torralba

สถานที่วิจัย MIT CSAIL

ตีพิมพ์เมื่อ 2018



ภาพประกอบที่ 2-14 การคาดเดาท่าทางมนุษย์เมื่อมีสิ่งกีดขวาง

การอ้างอิง

- [1] G-able. (2018). G-able. เรียกใช้เมื่อ 20 กันยายน 2561 จาก G-able: <https://www.g-able.com/thinking/artifact-intelligence/>
- [2] Phyblas.hinaboshi. (18 สิงหาคม 2561). Phyblas.hinaboshi. เรียกใช้เมื่อ 21 กันยายน 2561 จาก Phyblas.hinaboshi: จาก <https://phyblas.hinaboshi.com/panyapradit>
- [3] Vithan Minaphinant. (28 กุมภาพันธ์ 2561). blog.finnomena. เรียกใช้เมื่อ 23 กันยายน 2561 จาก blog.finnomena: <https://blog.finnomena.com/machine-learning-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-fa8bf6663c07>
- [4] Suphan Fayong. (2561). <http://codeonthehill.com>. เรียกใช้เมื่อ 25 พฤศจิกายน 2561 จาก <http://codeonthehill.com>: <http://codeonthehill.com/machine-learning-3-deep-learning/>
- [4] Sanparith Marukatat. (13 กรกฎาคม 2560). medium. เข้าถึงได้จาก medium: <https://medium.com/@sanparithmarukatat/%E0%B8%AA%E0%B8%99%E0%B8%B8%E0%B8%81%E0%B8%81%E0%B8%B1%E0%B8%9A-neural-network-2-11a7194ed236>
- [5] Tomas Simon, Shih-En Wei, Yaser Sheikh Zhe Cao. (2560). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Pittsburgh, เพนซิลเวเนีย: Carnegie Mellon University.
- [6] Sven Jens Jorgensen Miguel Arduengo. (2560). ROS Wrapper for Real-Time Multi-Person Pose Estimation with a Single Camera. csic.
- [7] Saixiii. (27 เมษายน 2560). Saixiii. เรียกใช้เมื่อ 2561 กันยายน 20 จาก Saixiii: <https://saixiii.com/python-programming/>
- [8] Daniel Bubnis. (1 มกราคม 2561). healthline. เรียกใช้เมื่อ 26 กันยายน 2561 จาก healthline: <https://www.healthline.com/health/fitness-exercise-weight-training>
- [9] ณัชนนท์ วงษ์วีไล ดนัยนันท์ เก่าเงิน. (2553). Virtual Fitting Room Using Augmented Reality. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- [10] Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, Dina Katabi Mingmin Zhao. (2561). Through-Wall Human Pose Estimation Using Radio Signals. Cambridge: MIT CSAIL.