

Lecture1) Introduction to Convolutional Neural Networks for Visual Recognition

- 동물들에게 시각이 중요함을 설명함
- 사람의 뉴런의 50%가 시각과 관련이 되어있다.
- 시각과 관련한 기계, 센서들도 우선적으로 개발되기 시작했다.(카메라)
- 사람의 뇌는 엷지를 받아들이는 간단한 세포에서 시작하여, 더 복잡한 정보를 처리하는 세포를 통해 시각 정보를 처리한다.

History

- **1966** "The Summer Vision Project"
- **1970s** "David Marr"
 - 이미지를 3D 모델로 인식하기까지의 과정을 설명
 - input image → primal sketch(점, 선) → 2.5D Sketch(레이어, depth) → 3D Model Representation(표면, 볼륨)
 - 어떻게 시각 정보를 분해할 수 있을까?
- **1970s** Cylinder, Pictorial Structure
 - 사물을 나타내는 방법에 대해 고민하기 시작함
 - 모든 물체는 간단한 기하학적 원시요소로 이루어졌다.
 - 복잡성을 낮추고, 간단한 모양으로 나타내고자 함
- **1980s** 선과 모서리로 이루어진, 직선의 조합으로 물체를 나타내고자함
- Object detection task가 복잡하다면, image segmentation을 먼저 수행
 - Image segmentation: 의미있는 구역의 픽셀들을 그룹화하는 것
 - 그래프 이론을 적용하기도 함
- ML 기술의 발전: SVM, boosting, 그래프 모델 ⇒ 신경망의 기초가 되었다.
- **2000s** 실시간 face detection
- **1999** object detection 기술의 발전, SIFT(object recognition)
 - 물체의 이동
 - 물체의 중요한 특징이 비슷한 물체를 찾으려 한다.
- **2006** Spatial Pyramid Matching
 - wave를 활용하여, 어떤 scene인지 알아맞출 수 있도록 단서를 제공
 - 이미지를 설명하기 위해, 여러 다른 해상도(resolution)의 이미지를 함께 사용, SVM 알고리즘 사용
- **2009** Human Recognition, HoG(Histogram of Gradients)
 - = deformable part models
 - 더 자연스러운 사람의 모형을 찾을 수 있게 되었음
- **2000s** Object Recognition → important problem
 - benchmark dataset을 갖추기 시작함(평가가 가능해짐)
 - ex) PASCAL Visual Object Challenge
 - 20 classes 분류
 - object detection 성능이 점점 좋아짐
 - ex) IMAGENET

- 어떤 형태의 모델로든, 분류 성능을 높이면 된다.
- 복잡한 시각 데이터, high dim input, 적은 train dataset → 오버피팅 속도가 빨라서 generalization 문제가 있었음
- 따라서 이를 해결하기 위해 아주 큰 데이터셋을 만들고자 함
- 22000개의 클래스, 1500만 ~ 4000만개의 이미지

How to Benchmark?

- 2009 ImageNet 팀이 Challenge를 기획
 - Image Classification Challenge
 - 1000 object classes
 - 140만개 images
- 2012년, ImageNet Challenge에서 25% 미만의 error rate 달성
 - by Convolutional Neural Network Model
 - Deep Learning, Neural Network의 잠재력을 보여줌

Visual Recognition Problems

- Image Classification: 전체 이미지를 하나의 카테고리로 분류
- Object Detection: bounding box를 통해 개체의 위치와 종류를 파악, 배경과 개체들을 구분함
- Image Captioning: 자연어로 이미지를 설명함

ImageNet Challenge prized Models

CNN은 Object Recognition에서 매우 중요함

- **2011** 여전히 계층적인 구조를 가진다.
- **2012** AlexNet(=SuperVision)
- **2014** GoogleNet, VGG
- **2015** MSRA(Microsoft)

이후로 GPU 성능이 더 좋아지며, 더 복잡한 네트워크를 사용할 수 있게 되었다.

Convolutional Neural Network Algorithm History

2012년도에 처음 만들어진 것이 아니다.

- 1998 LeCun, 숫자를 인식하는 CNN을 만들어냄
 - 픽셀을 통해, 어떤 글자인지 나타냄
 - 2012 AlexNet보다 더 간단한 구조였다.
 - 왜 2012년도가 되어서야 유명해졌는가?
- CNN 발전의 이유
 - 컴퓨팅파워 개선(트랜지스터의 개수 상승, GPU 등)
 - 데이터의 수 증가(대량의 labeled dataset 제공 - PASCAL, ImageNet)