

Introduction to Convolutional Neural Networks

for Visual Recognition

Computer Vision

: 시각적 데이터들을 효과적으로 이해 및 분석해서, 컴퓨터로 하여금 인간의 시각적인 인식능력을 가질 수 있도록 알고리즘을 개발한 것

History of Computer Vision

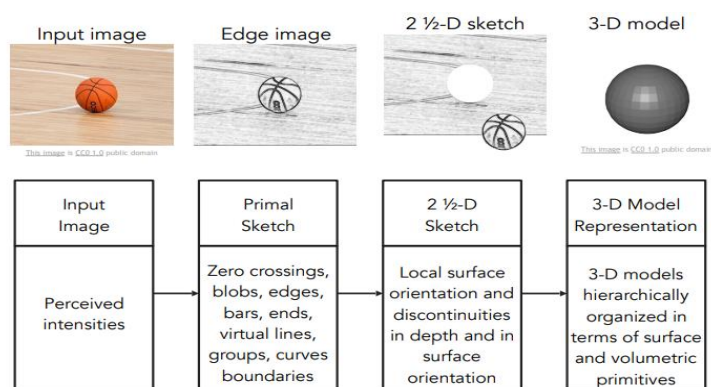
1959 Hubel & Wiesel

Hubel & Wiesel은 "포유류의 시각적 처리 메커니즘"에 대해 알고 싶었고, 이를 위해 고양이 뇌에 전극을 꽂아 어떠한 자극을 주어야 1차 시각 피질의 뉴런들이 격렬하게 반응하는지 관찰하였습니다. 1차 시각 피질에 존재하는 다양한 종류의 세포 중 가장 중요한 세포는 edges가 움직이면 반응하는 단순한 세포들입니다. 이들은 시각 처리가 처음에는 단순한 구조로 시작되어 그 정보가 layer를 거치며 실제 input을 인지할 수 있을 때까지 점차 복잡해진다고 결론을 내렸습니다.

1960대 초 Larry Roberts의 Block World

사물을 기하학적 모양으로 단순화 후 재구성

1970년대 David Marr



Stages of Visual Representation, David Marr, 1970s

눈으로 받아들인 이미지(input)을 3D 표현으로 만들려면 몇 단계 과정을 거쳐야 함

객체 분할(이미지의 각 픽셀을 의미 있는 방향으로 군집화)에 초점을 맞추기 시작

1999/2000년대

기계학습

SVM, Boosting, Graphical models, 초기 Neural Network

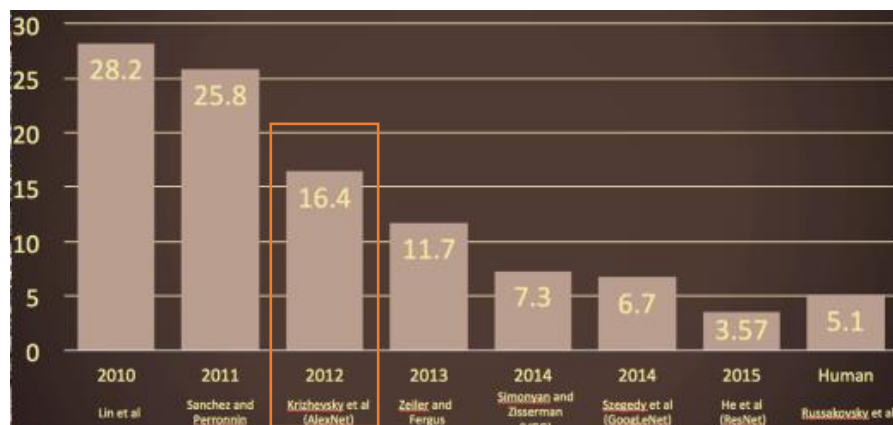
Paul Viola, Michael Jones: AdaBoost를 이용한 실시간 얼굴인식 성공(2001)

특정기반 객체인식 알고리즘(90년대 후~2010)(ex) David Lowe의 SIFT feature)

: 객체의 특징 중 일부는 변화에 더 강인하고 불변하다는 점 발견하여 이미지 전체를 매칭하기 보다는 이와 같은 중요한 특징들을 찾아내고 그 특징들을 다른 객체에 매칭시키기 시작.

ImageNet

: 가장 큰 데이터셋을 만들어 세상의 모든 객체를 인식하고 Overfitting 극복하려 함



2012년 CNN

CS231n Overview

Image Classification

: 한 장의 이미지를 보며 몇 개의 고정된 카테고리 안에서 답을 고르는 것

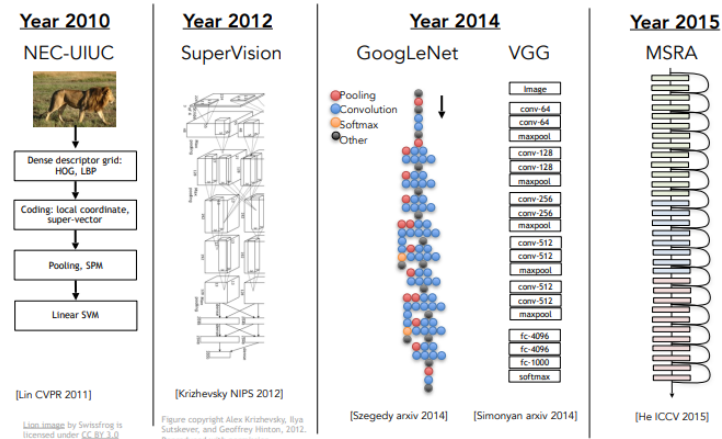
Object Detection

: object 가 있는 위치에 네모 박스를 그릴 수 있어야 함

Image Captioning

: 이미지가 입력으로 주어지면 이미지를 묘사하는 적절한 문장을 생성

CNN



2011 년, Lin et al 의 알고리즘

: 계층적(hierarchical), edges 를 뽑고 "불변 특징"

2012, Alex Krizhevsky & Ilya Sutskever

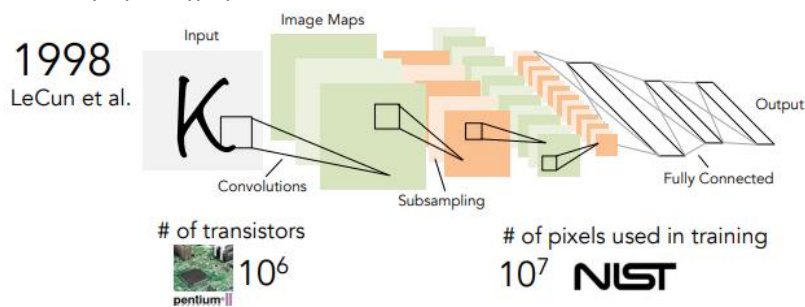
: AlexNet(7-Layer Convolutional neural network)

2014, GoogleNet

2015, MSRA

1998, Jan LeCun & Bell Labs

: 이미지를 입력으로 받아서 숫자와 문자를 인식할 수 있는 CNN 을 구축
AlexNet 과 구조 유사



raw pixel 을 입력으로 받아 여러 Convolution Layer Layer 를 거치고 Sub-Sampling, Fully Connected Layer 를 거칩니다.

90 년대의 LeNet 아키텍처를 공유하기 때문에 2012 년의 CNN 아키텍처는 서로 비슷합니다.

연산량과 데이터의 증가에 따라 CNN 발전