

lec01| Introduction to Convolutional Neural Networks for Vissual Recognition



History

Habel&Wisel

포유류의 시각처리 매커니즘 연구

단순한 구조로 받아 들인 후 물체 인식까지 복잡하게 빌드업

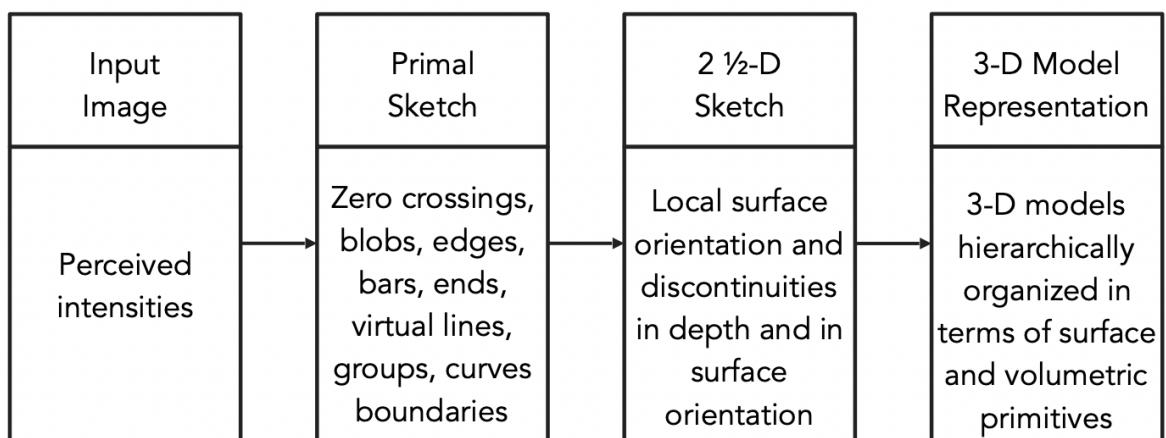
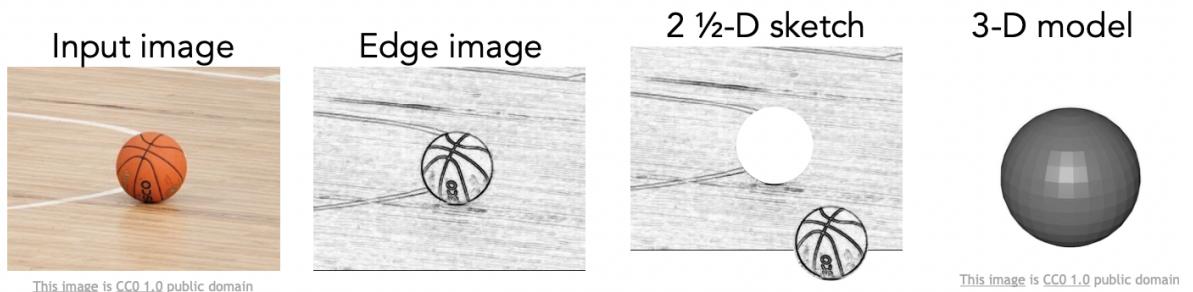
1960s

block world _ Larry Roberts

보이는 것을 인식, 재구성 하는 연구

David Marr

vision연구가 어떠한 알고리즘 방향으로 나아가야하는지 연구



Stages of Visual Representation, David Marr, 1970s

Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 1 - 13

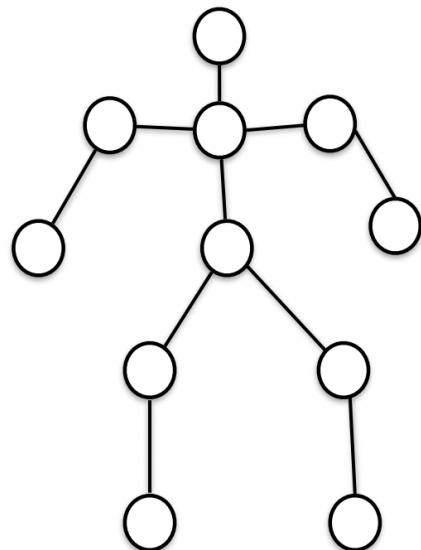
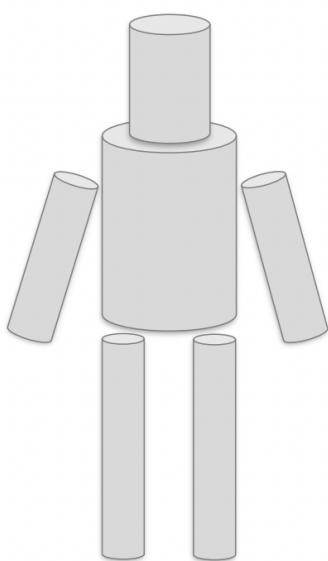
4/4/2017

1970s

simple block world → 실제 물체 인식으로 나아가는가?

1. generalized cylinder : 모든 객체는 간단한 기하학적 물체로 표현 가능
2. pictorial structure: 주요부위 & 관절 위주

- Generalized Cylinder Brooks & Binford, 1979
- Pictorial Structure Fischler and Elschlager, 1973



1980s

단순한 구조로 인식과 표현 : line, edge, straight line을 이용하여 면도기 인식

이때, 객체인식을 먼저하는 것이 아니라 객체 분리를 priority로 삼음

배경, 사람이 속한 픽셀을 분리, graph theory algorithm 사용

1990s

feature based image recognition을 연구. (특징 기반)

객체의 특징 중 일부는 거의 변하지 않난다. ⇒ 특징을 찾아서 다른 객체와 매칭.

SIFT feature를 표지판에서 찾아 추출

이것은 곧 장면 전체 인식에 유용하다. 특징을 뽑아 나열하여 이미지에 대한 단서를 제공할 수 있다.

이미지 내의 여러 부분과 여러 해상도에서 추출한 특징을 하나의 기술자로 표현, sv 알고리즘 사용.

→ 사람인식 연구에 사용했다.

사람의 몸을 어떻게 현실적으로 모델링하는가?

Histogram of gradients, Deformable part models를 이용

2000s

21세기에 들어서면서 사진 품질이 좋아진다.

“object recognition” 이 주요과제

PASCAL visual object challenge(VOC)

20개의 클래스 데이터 셋. 클래스당 수천-수만개의 데이터를 이용해서 알고리즘 테스트.

2007-2012 동안 객체 인식 성능은 꾸준히 증가

Are we ready to recognize every object in the world?

당시 대부분은 기계학습 알고리즘에 해당.

overfitting 문제가 생김

모델입력에는 복잡한 고차원 데이터를 사용, 이를 피팅 하기 위해 많은 파라미터가 필요

충분한 학습데이터가 없을 시에는 overfitting이 빠르게 발생, 일반화 능력 저하

ImageNet

위를 바탕으로 2가지 motivation

- recognize the world of all objects
- overcome the machine learning bottleneck of overfitting

수십억장의 이미지를 다운 → WorldNet이라는 dictionary로 정리

amazon mechanical turk에서 사용하는 clever crowd engineering trick을 도입. (이미지의 정렬, 정제, 레이블등을 제공하는 플랫폼)

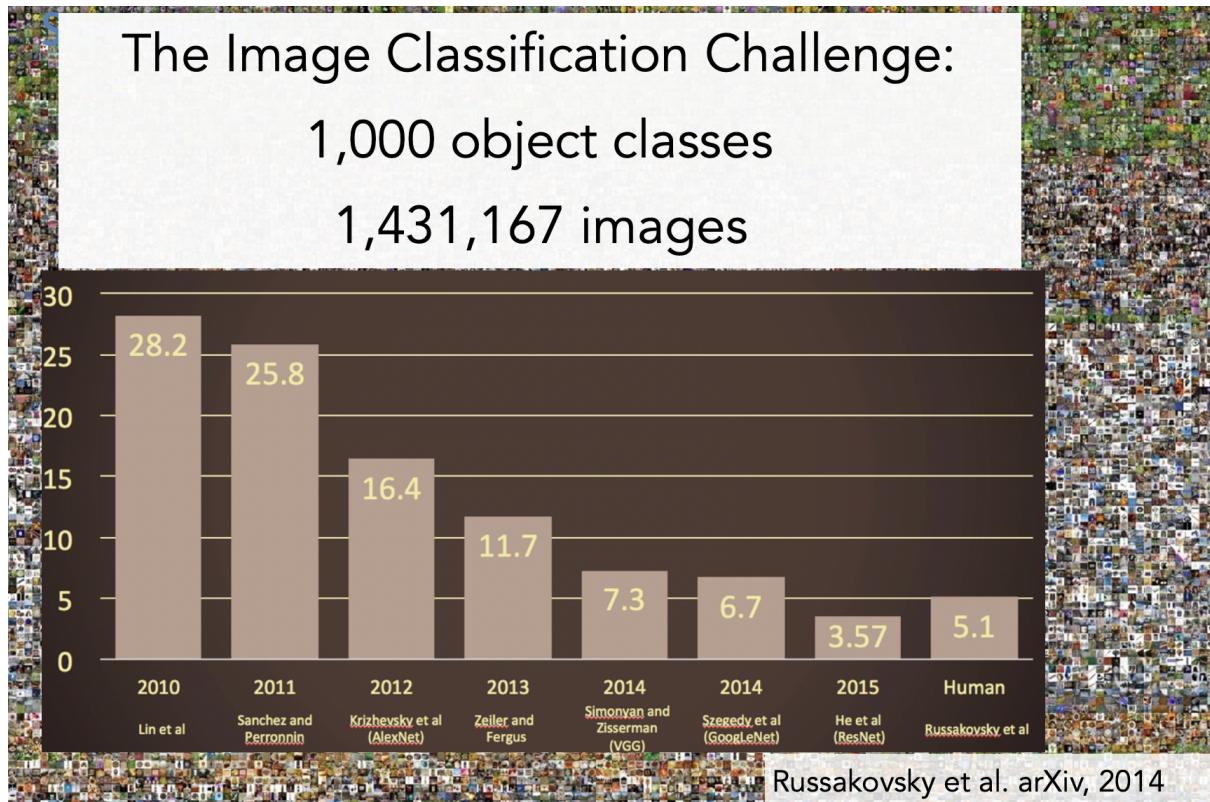
⇒ ImageNet은 15M이상의 이미지, 2.2K의 카테고리 보유 최대 데이터 셋

관점은 어떻게 해당 데이터셋을 이용하여 벤치마킹하는지

→ 2009부터 국제규모 대회 ILSVRC 주최, 이미지 분류 문제를 푸는 알고리즘을 테스트하기 위함

: 예제 이미지, 참가자는 정답후보를 5개 고를 수 있고, 이 중 정답이 있으면 성공

아래 이미지는 오류율



2012년 모델 : Convolutional NN(DP)

CNN이 무엇인지, 법칙, 선례, 최근 동향 등을 해당 강좌에서 다를 예정.

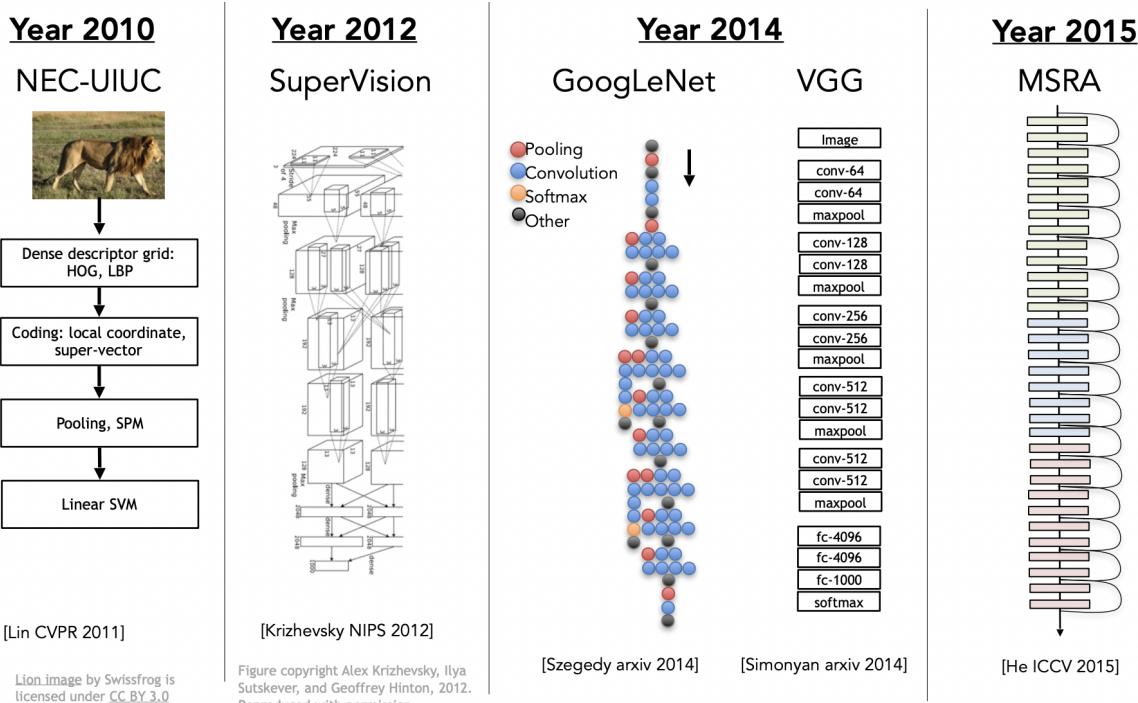
Overview

Image Classification 이 주요 내용

object detection, action classification, image captioning 등을 다룰 예정

image classification이 재사용 될 수 있음.

IMAGENET Large Scale Visual Recognition Challenge



2011을 보면 아직 계층적이다. edge를 뽑고, “불변 특징” 개념도 사용 → 이러한 직관은 cnn에도 영향을 미친다.

2012년을 보면 7 layer conv. nn ~ AlexNet(Supervision) 이후로 계속 우승은 neural network가 가져감.

2015년 까지 layer의 개수는 계속 증가하는 것을 볼 수 있다. 152개에 육박함. 200개 까지 늘리면 성능은 더 좋아진다고 하지만, gpu가 견디지 못함.

CNN의 기초연구

1998년의 해당 연구가 cnn의 기초연구라고 볼 수 있음

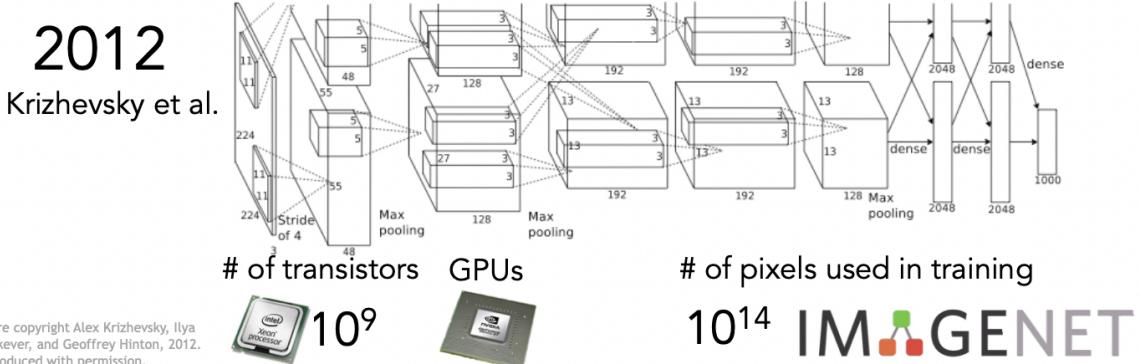
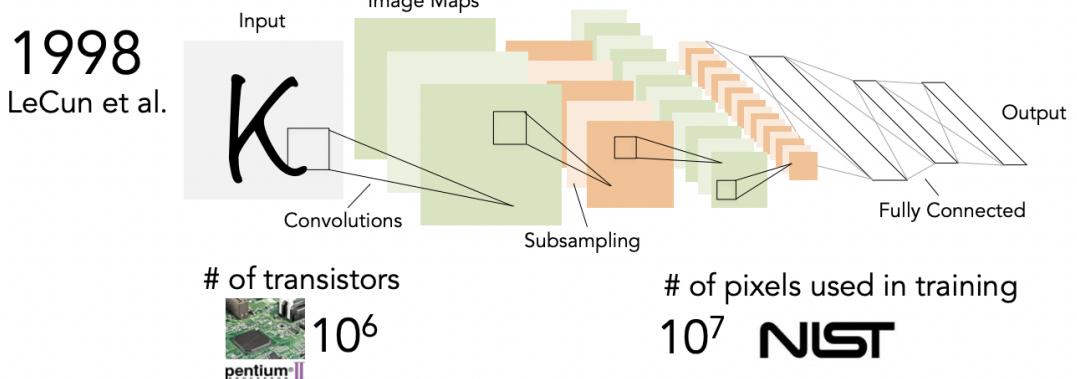


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012.
Reproduced with permission.

Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 1 - 33

4/4/2017

1998~2012 동안 컴퓨터는 더 많은 계산을 빠르게 할 수 있었고, 이게 알고리즘 성능에 영향을 미침

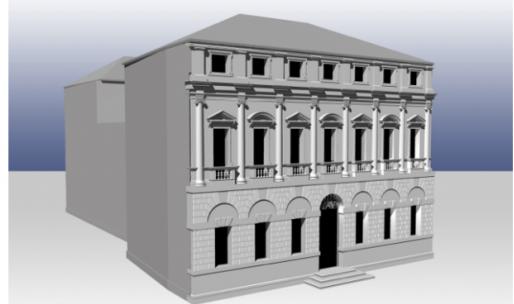
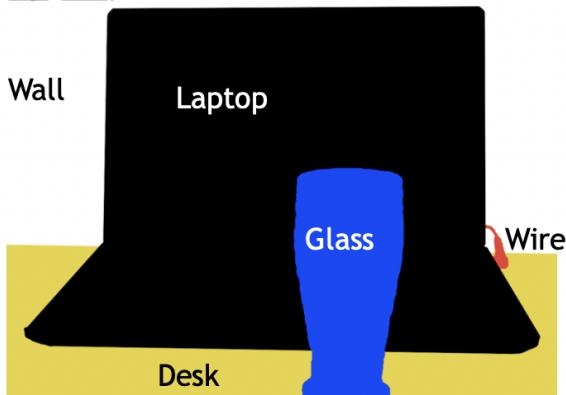
90년대에는 레이블이 새겨진 이미지 데이터 셋을 수집하기 어려웠지만, 지금은 PASCAL, ImageNet과 같이 크고 좋은 퀄리티의 데이터셋을 사용할 수 있다.

두 원인으로 higher capacity model을 만들 수 있음.

컴퓨터 비전의 연구 목적: build machine that can see like people

현존하는 문제들 :

[Image](#) is CC0 1.0 public domain



[Image](#) is GFDL

[Image](#) is CC BY-SA 2.0



Waving

[Image](#) is CC BY-SA 4.0



Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 1 - 35

4/4/2017