

Cs231n Lecture 1 Summary

Introduction to Convolutional Neural Networks for Visual Recognition

2170051 송여진

1. 컴퓨터비전이란?

컴퓨터 비전(Computer vision)의 중요성은 갈수록 커지고 있다. 인터넷 트래픽의 80%는 비디오 데이터인데, 이같은 결과는 인터넷의 데이터 중 대부분이 시각 데이터라는 사실을 보여준다. 따라서 시각 데이터를 잘 활용할 수 있는 알고리즘을 개발하는 것이 중요하다. 또한 시각 데이터는 해석하기가 굉장히 어렵고 관련 분야도 생물학, 물리학, 심리학 등 다양하다. 이 수업에서는 알고리즘 중 CNN을 주로 배울 것이다.

2. 컴퓨터비전의 역사

생물에게 비전(시각)은 굉장히 중요한 요소이다. 5억 4천만년 전 생물들에게 최초로 눈(eyes)이 생겨났고, 이에 따라 생물종은 폭발적으로 증가하였다. 비전은 가장 큰 감각 체계이며 우리의 생존, 일, 여가 등 많은 부분을 담당한다.

인간이 만든 공학적 비전은 카메라라고 할 수 있는데, 카메라의 역사는 1600년대부터 시작된다. 르네상스 시대의 최초의 카메라인 *obscura*는 생물학적으로 발전한 초기의 눈과 굉장히 유사하다. 빛을 모아주는 구멍이 있고 카메라 뒤의 평평한 면이 정보를 모아서 이미지를 투영한다.

그동안 생물학자들은 비전의 매커니즘을 연구하기 시작한다. 포유류의 시각 처리 매커니즘을 알아내기 위해 고양이의 뇌로 실험을 하였는데, 일차 시각 피질의 뉴런들은 '경계(edges)'가 움직이면 이에 반응하는 세포들이 가장 중요하다는 사실을 알아내었다. 컴퓨터 비전의 역사는 1960년대 초반에 시작된다. 'Block World'라는 컴퓨터 비전 분야 최초의 박사 논문은 우리 눈에 보이는 사물들을 기하학적 모양으로 단순화시켰다.

1966년에는 MIT에서 중요한 여름 프로젝트가 진행되었는데, 'The Summer Vision Project'이다. 이 프로젝트의 목표는 시각 시스템의 전반을 구현하기 위해서 참가자들을 효율적으로 이용하는 것이었다.

David Marr라는 MIT의 비전 과학자는 70년대에 중요한 책을 저술한다. 이 책은 컴퓨터가 비전을 인식하기 위해 어떠한 방식으로 알고리즘을 개발해야 하는지 다룬 책이다. 그는 이 책에서 우리가 눈으로 받아들인 이미지를 최종적인 full-3D 표현으로 만들려면 몇 개의 단계를 거쳐야 한다고 주장했다.

첫번째는 'primal sketch'라는 단계인데, 주로 경계, 선, 커브가 표현된다. 그리고 다음 단계에서는 2.5D sketch라고 하여 표면 정보, 깊이 정보, 레이어, 불연속 점 등이 표현된다. 마지막으로 이것들을 한데 모아서 3D 모델이 되는 것이다. 이런 방식은 실제로 수십 년 간 컴퓨터 비전 분야를 지배하는 방식이었다.

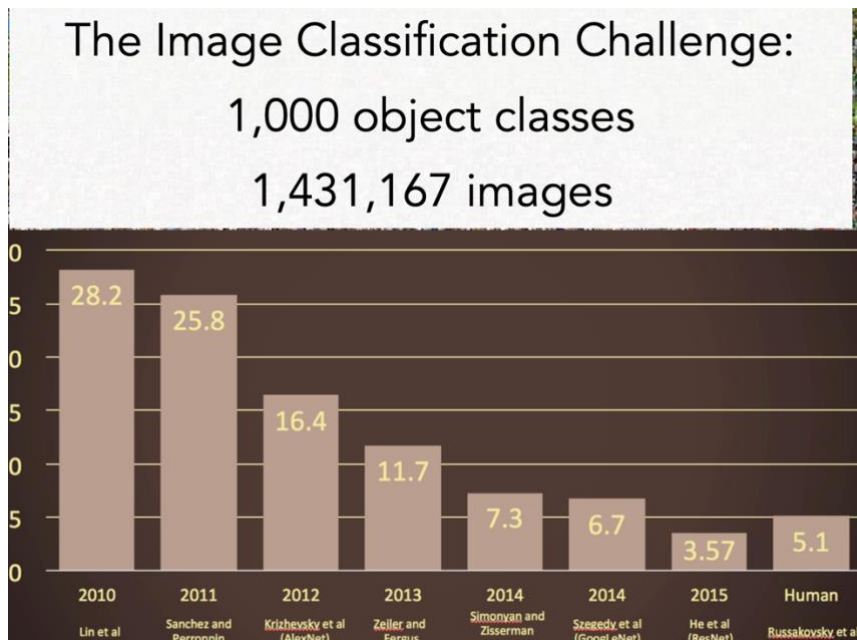
70년대에는 또 다른 중요한 연구가 시작되었는데, 어떻게 하면 단순한 블록 세계를 뛰어넘어 실제 세계를 표현할 수 있을까? 라는 질문이 탄생했다. 이 당시에는 인터넷이 잘 보급되지 않았고 수집할 수 있는 데이터도 굉장히 적었다. 그래서 'generalized cylinder' 라는 객체를 원통형 모양으로 바꿔서 단순화시키거나 'pictorial structure' 라는 객체를 점과 선으로 잇는 방식으로 단순화시켜서 표현하였다.

하지만 이러한 방식은 큰 효과를 나타내지 못했고, 연구자들은 객체 전체를 인식하기 보다는 객체 분할(segmentation)을 시도해보는 방식으로 바꾸었다. 이를 '영상분할(Image Segmentation)'이라고 부르는데, 이 연구는 영상분할 문제를 해결하기 위해서 그래프 이론을 도입하였다.

90년대에서 2010년도까지의 시대를 풍미했던 컴퓨터 비전 알고리즘은 '특징 기반 객체인식 알고리즘' 이었다. 이 시기에 나온 가장 유명한 알고리즘이 SIFT feature 인데, 객체에서 변하지 않는 중요한 부분을 찾아 그 특징을 다른 객체에 매칭시키는 방식이었다.

사람들은 컴퓨터비전에서 객체인식 기술을 측정하기 위해 PASCAL Visual Object Challenge (VOC) 를 개최하였는데, 정확도는 점점 증가하였지만 기계학습 알고리즘이 training 과정에서 Overfit 을 하는 것 같다는 사실을 발견하였다. 이 문제의 원인은 시각 데이터가 너무 복잡하다는 것 때문이었는데, 학습 데이터가 부족할수록 overfitting 이 빠르게 발생하였다.

이 문제를 해결하기 위해 ImageNet 프로젝트가 시작되었다. 인터넷에서 수십억 장의 이미지를 찾아서 클래스별로 나누고 데이터셋을 만들었다. 이 ImageNet 팀은 2009년부터 국제 규모의 대회를 개최하였는데 ILSVRC 이다. 이 대회의 목적은 이미지 분류 문제를 푸는 알고리즘을 테스트하기 위함이었다.



이 사진으로 보면 2012년에 전년도 대비 유의미하게 오류율이 줄어든 것을 알 수 있는데, 2012년에 CNN 이 등장했기 때문이다. CNN 은 Deep Learning 이라고도 하며 수업에서 중점적으로 다룰 내용이다.

3. 컴퓨터 비전의 목표

컴퓨터 비전은 많이 발전했지만 그래도 사람의 시각(Vision)보다는 부족한 점이 많다. 그리고 단순히 사진의 객체를 인식한다는 것을 넘어 그 안에 담긴 의미까지 파악할 수 있게 되는 것이 최종적인 목표일 것이다.