Lecture 1

수강생이 exponentially 증가하고 있다. 17년 8월 11일 확인결과 730명이 수강중이다.

세상에선 정말 많은 양의 데이터가 비주얼 형식으로 저장되어 있지만, 이런 데이터는 처리하는데 몇 가지 어려움이 있다. 인터넷을 이용하면 비주얼 데이터가 쏟아질 정도로 많지만, 사실 이들의 구성이나 내용에 대해 이해하는 과정은 충분하지 어려운 편이다.

이 이상의 인트로는 그냥 생략하도록 한다.

컴퓨터 비전의 역사

고생물학을 보면 생물의 종이 폭발적으로 증가한 순간이 있고 이를 생물학적 빅뱅이라고 부르는데, 오스트리아인 동물학자가 굉장히 설득력 있는 설을 제시한다. 화석을 통해 추정컨데 이는 동물이 ‘눈’이라는 기관을 갖게 된 시점에서 생물들의 진화 전쟁의 게임판도가 급격하게 변화했다는 것이다. 포식자가 먹잇감을 쫓고, 동물들은 이로부터 도망치기 위해 빠르게 진화해야 했다는 것이다. 즉 비젼은 생물에게 있어서 생존하고 기능하고 서로 소통하는 등 생명의 유지에 정말 중요한 역할을 수행한 것이다.

그렇다면 사람의 비전은 어떻게 발전되었는가? 처음에는 바늘 구멍을 이용한 카메라부터 시작하여 동물 실험을 통해 동물의 비젼 프로세스 이해까지 이어졌다. 고양이의 시각 시스템에서 중요한 역할을 하는 세포는 윤곽선의 이동을 감지하는 세포 등을 기반으로 한다. 즉 윤곽선 등의 기초적인 정보부터 시작해서 다른 정보들을 조합하고 처리하여 복잡한 비주얼 데이터를 이해하게 되는 것이다.

처음 컴퓨터 비젼에 관한 연구로 larry robert의 블록을 단훈화 하고 이를 다시 복원해내는 과정에 관한 연구가 첫번째 컴퓨터 비젼에 관한 박사 학위논문이었다. 이후 MIT의 summer project 라는게 수행이 됐는데, 한 여름기간동안 컴퓨터 비젼의 벌크를 완성하겠다는 야심찬 프로젝트를 진행하였다. 이 외에도 70년 후반에 쓰여진 책 등등 비젼에 관한 아이디얼 한 차원에서의 논의가 이루어졌다.

이후 70년대를 지나며 그냥 단순화된 블럭 세계를 넘어서 실제 세계의 오브젝트를 인식하고 표현하는 문제에 대해 생각하게 된다. 이 때 나온 것이 Generalized Cylinder, Pictorial Structure이다. 모든 물체들을 원초적인 모형으로 표현할 수 있다는 것인데, 가령 사람을 간단한 실린더 모형으로 표현한다거나 관절들과 그 사이의 거리들을 이용한 그래프로 표현하는 것 등이다.

오브젝트 인식

Normalized Cut의 경우 각 이미지의 의미있는 부분을 서로 연결하고 그룹을 만들어 내어 물체를 인식하는 방법이며 image segmentation 이라고 불린다. 그리고 이어지는 중요한 연구로는 얼굴 인식이 있는데, 이때부터 확률론을 기반으로 한 머신러닝이 입지를 갖게 된다. 그 중 실시간으로 동작이 가능할 정도의 알고리즘이 2001년에 연구되었고 논문 출판 후 5년만에 실제 디지털 카메라에 구현되게 된다.

어느 시점부터인가 연구그룹 내에서 스스로에게 드는 질문이 생겼는데, ‘이제 지구상의 모든, 혹은 대부분의 오브젝트들을 인식할 정도로 준비가 되었는가? 였다.’ 머신러닝 관련 메소드들, 그래피컬 모델이던 서포터벡터 머신이나 에이다부스트 모델의 경우 트레이닝 과정에서 오버피팅이 되는 경항이 있다. 그리고 다른 목잡한 문제로는 비주얼 데이터 자체가 차원이 너무 높아서 매우 복잡하다는 점이다. 그리고 데이터의 수가 부족하면 오버피팅의 문제를 피할 수가 없다. 따라서 세상에 있는 수많은 데이터들을 모아서 최대한 많은 데이터들을 이용하여 학습을 진행하고 싶었다. 해서 나온게 ImageNet 인듯 하다.

이 데이터셋을 이용해서 이미지넷 챌린지를 열었는데, 2012년에 전년대비 에러율을 큰 폭으로 떨어뜨리며 괄목할 만한 성과를 냈다. 이 시점이 이미지 인식에서 CNN 기반 혹은 딥러닝 모델이 엄청난 효율을 보인 다는 것을 알게 된 시점이고 관련된 분야- 음성인식이나 텍스트 처리 등의 다양한 영역에서 주목을 받게 되었다.

이런 이미지 분류 문제는 세팅에 대한 요구사항이 엄격하지 않으므로 일반적인 분야에 범용적으로 사용할 수 있다. 따라서 산업 등등 많은 분야에 응용될 수 있다는 장점이 있다.

그리고 이 수업에선 다른 비주얼 인식 문제들에 대해서도 다루게 될 것인데, 오브젝트 인식과 이미지 캡셔닝 등을 포함한다. 이미지에서 각 오브젝트들이 어디에 위치해 있는지 박스치고 표시해주는 과정이다.

근데 이런 연구가 98년도에 이미 있었는데, 십수년이 지난 지금에서 갑자기 주목받는 이유는 컴퓨팅 장비의 발전 덕분이다. 컴퓨터의 성능이 해를 넘기며 큰 폭으로 증가했으며 특히 그래픽 데이터를 병렬연선으로 다루는데 특화된 유닛의 개발이 이어지면서 CNN을 이용한 연구가 매우 수월해졌다. 따라서 예전 연구방식과 같은 접근 방법으로 그저 스케일을 키우기만 해도 상당한 퍼포먼스를 확보할 수 있는 경향이 있다. 또한 파스칼 이미지넷 등 90년도에 사용할 수 있었던 데이터셋과는 비교할 수 없을 정도의 대형 데이터셋이 등장하면서 학습의 퀄리티가 증가한 것 역시 영향력이 있다.

하지만 여전히 사람의 시각 시스템이 현재의 비젼 시스템과 같이 오브젝트에 박스를 치고 레이블링을 하지 않는 다는 점에서 컴퓨터 비젼 분야는 많은 챌린지들을 갖고 있다. 관련된 연구로는 앞서 언급된 image segmentation 을 기반으로 한 디텍션이나 혹은 사물 이미지를 이용하여 3D 모델링을 통해 오브젝트를 역으로 구현해 내는 등의 아이디어들이 연구된 적이 있고 이들이 남은 챌린지의 일부이다. 혹은 이미지에 표시된 물체들간의 관계를 통해 트리를 만들어 접근하는 방식도 있다.

Andrej Karpathy는 살아있다.