

Lecture 1 Introduction

Computer vision 분야는 시각 데이터(visual data)를 다루는 분야 (이미지, 영상 등)

인터넷 상에는 엄청난 양의 시각 데이터가 존재하고 이런 픽셀 데이터는 우리가 내용을 알아내기 힘든 암흑 물질과도 같은 존재이다.

➔ 이번 강의에서는 신경망 모델을 이용한 시각적 인식을 배운다.

5억 4천만년 전 이후의 화석에서는 엄청난 양의 종이 갑자기 발견됨 (분화(생물의 다양화))

-> why? 빙하기의 도래

최초로 삼엽충이 "눈"을 가졌다. 눈을 가지면서 먹이 관계가 형성되었고 이를 통해 분화 및 천적 등의 관계가 발생했고 이것이 바로 비전의 시작이었다. (생물학적 비전)

카메라의 발전 : 우리는 이제 어딜 가든 카메라, 센서를 볼 수 있음 (인구 수보다 많음) (공학적 비전)

"포유류의 시각적 처리 메커니즘은 뭘까?" 라는 의문, 어떤 자극을 줘야 시각 피질의 뉴런들이 반응할까? -> 생각보다 단순하다. 경계가 움직이면 시각 피질의 뉴런들이 여기에 반응

우리는 눈으로 받아들인 "이미지"를 "최종적인 full 3D 표현"으로 만드는 알고리즘을 공부해야함

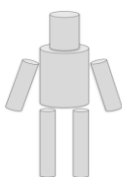
시각적 표현의 단계

이미지 -> 경계 -> 표면 -> 3D

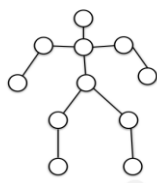
객체 인식에 대한 고민 (어떻게 해야 장난감 같은 단순한 블록 체계를 뛰어넘어서 실제 세계를 인식하고 표현할 수 있을까?)

➔ 모든 객체는 단순한 기하학적 형태로 표현 가능

• Generalized Cylinder
Brooks & Binford, 1979



• Pictorial Structure
Fischler and Elschlager, 1973



Generalized cylinder & pictorial structure(복잡한 객체의 단순화)

객체 분할(segmentation)

이미지의 각 픽셀을 의미 있는 방향으로 군집화 (ex. 배경인 픽셀과 사람이 속한 픽셀을 구분)

객체 인식 (풀어야 할 문제!!)

객체에서 중요한 특징들을 추출하여 이를 식별하고 이미지와 매칭 (이미지 전체 매칭보다 간편)

: 이미지에 존재하는 "특징"을 사용하면서 컴퓨터 비전은 한 단계 도약

ImageNet(객체인식 challenge 대회)에서 우승한 알고리즘 => convolutional neural network(CNN)

(한국말로 번역하면 합성곱 신경망..딥러닝이라고도 부름)

CNN 모델은 컴퓨터 비전 분야의 진보를 이뤄냄

CNN 모델이 2012년에 하루 아침에 나온 것은 아니다 -> 90년대부터 연구 이루어짐

근데 이제 빛을 보기 시작하게 된 건 컴퓨터의 CPU 및 GPU 성능이 좋아졌기 때문 (GPU는 아주 강력한 병렬 처리가 가능한 데 계산 집약적인 CNN 모델을 고속으로 처리하는 데 안성 맞춤이다)

이 수업에서 다루는 주제

1. Image Classification (알고리즘이 이미지 한 장을 보고 몇 개의 고정된 카테고리에서 정답을 고른다)
2. Object detection (이미지가 주어지면 고양이가 어디에 있는지 네모박스를 객체의 위치에 정확히 그린다)
3. Image Captioning (이미지가 주어지면 이미지를 묘사하는 적절한 문장 생성한다)

Deep learning에서는 잘 레이블된 dataset과 연산량이 중요하다.

이미지의 내용을 풍부하고 깊게 이해하는 것이 컴퓨터 비전 분야가 진정으로 추구하는 방향 (이미지는 정말 많은 정보가 담겨있기 때문에)

CNN 알고리즘을 python을 이용해 구현할 예정이다 (tensorflow 같은 도구도 이용)