

인공지능은 물체를 어떻게 이해할까?

물체 탐지 (object detection)는 위치 추정 (x, y, w, h)과 분류 이 두가지를 동시에 하는 작업입니다. 위치추정은 x, y 좌표와 그 상자의 너비 w 와 높이 h 를 예측해야되서 네가지 숫자를 나타내고 알고리즘에 따라서 AI가 정보를 얼마나 확실한 지에 따라서 분류를 위한 물체 종류의 갯수에 비례하는 confidence score 분류 점수를 나타낼 수 있습니다. 이 두가지를 동시에 해야되기 때문에 다중테스트 학습 multitask learning 이라고 하고 다양한 다중 손실함수들을 합쳐서 사용합니다.

입력 이미지를 받으면 딥러닝으로 특성값을 추출한 다음에 위치를 추정합니다. 어떠한 영역이 이 사진 속에서 물체인지 바운딩 박스로 표현하면 Localization (x, y, w, h) 위치추정으로 바운딩 박스의 맨 왼쪽, 위의 좌표 x, y 하고 너비 w , 높이 h 로 표현이 가능합니다. 물체의 위치를 예측하고 추정하고 나서 이 안에 있는 물체가 무엇인지 분류(classification)를 합니다. 위치추정은 숫자로 되어 있기 때문에 회기 분석을 해야됩니다. 그래서 학습할때 L2 손실함수를 사용합니다. 분류 문제에서는 classification 에서 사용하는 softmax 를 적용한 다음에 cross entropy 손실함수를 사용합니다. 다 간이 다중 태스크 함수 multitask learning 하고 다중 손실 함수들을 합쳐서 이 모델을 정의합니다.

ImageNet 같은 큰 데이터 세트에서 미리 학습된 기저 네트워크를 사용해야 좋은 특성 값을 뽑을 수가 있습니다. 한 사진 많은 물체가 있고 그것을 파악하는데 사진의 부분마다 후보박스를 그려 살펴보면 너무나 많은 후보박스들로 복잡도가 높아집니다. 그래서 선택적 검색을 합니다. 선택적 검색은 박스안에 물체가 있는지를 판단합니다. 물체가 없다면 무시하고 물체가 있다고 하면 박스를 다시 정밀하게 분석합니다.

물체가 있는 바운딩 박스의 갯수를 줄이기 위해 선택적 검색을 하고 얼마나 박스들이 잘 겹쳤는지, 후보 박스의 물체가 충분히 들어 있는지 또는 예측한 물체 탐지의 위치 추정이 우리 정답이랑 충분히 겹쳤는지 측정을 하기 위해서 IoU(intersection over union)을 통해서 측정합니다. 어떠한 박스 안에 물체가 있는지를 IoU intersection over union 으로 판단을 할 수가 있습니다. 합집합분에 교집합을 하면 비교가 가능합니다. 이 방법으로 어떠한 박스안에 물체가 있다 없다를 판별하고 이것의 기준으로 예측한 물체 탐지가 맞았는지도 판단이 가능합니다. 중복 탐지 결과 제거도 할 수가 있습니다. Non-max suppression(NMS) 기법을 통해서 IoU score 가 높으면 심각하게 겹친 같은 클래스 탐지 결과로 중복 박스를 제거를 할 수가 있습니다. 중복된 탐지 결과들을 제거하기 위해서 NMS 을 통해서 한 물체당 하나의 위치 탐지 결과를 남도록 후처리 합니다.

물체 탐지 기법들은 다단계 탐지 기법: RCNN, Fast-RCNN, Faster-RCNN, mask RCNN 이 있었고 한번만에 하는 single stage object detection 기법은 SSD, YOLO 가 있습니다. 선택적 검색을 컴퓨터 비전 방식 말고 Reason proposal network RPN 이 제안이 되었고 RPN 딥러닝 기반 선택적 검색합니다. 각 관심 지역별로 고정된 크기 특성 값을 꺼내기 위해서 ROI Pooling 활용 지역별 특성값 동일 차원 벡터로 변형합니다. 최신 물체 탐지 기법들은 고정 박스들을 사용하지 않는 기법들이 제한되고 있고 그를 위해서 DETR, 트랜스포머 어텐션을 이용한 기법 Facebook AI research 에서 CenterNe 이 있습니다. 고정박스가 없는 방법들의 좋은 점은 때에 따라서 NMS 중복된 결과 제거하는 것도 제거하는 후처리 과정도 제거할수 있습니다.

디텍트론을 통해서 다중 태스크 물체 이해를 하고 물체를 탐지하는 것만 아니라 물체들이 무엇을 하는지 혹은 사람들이 어떤 포즈를 갖고 있고 어떤 부분이 물체고 어떤 픽셀이 물체가 아닌지 세그멘테이션을 통해서 알 수가 있습니다.

물체간의 상호 작용 탐지하는 기법들은 다단계 탐지 기법: InteractNet, Message-passing Network 부터 단단계 탐지 기법: UnionDet, HOTR, IPNet 이 있습니다. 사전 지식을 이용하면 아직까지는 데이터 셋이 크지 않기 때문에 처음 보는 상호 작용이나 아주 몇번 학습되지 못한 상호 작용도 잘 탐지 할 수가 있습니다.언어로 한 방법도 있었고 미리 사전 학습을 데이터로부터 직접 학습하는 경우도 있습니다. 관계들이 너무 많기 때문에 아주 많은 관계들 레이블링을 할 수가 없어서 잘 조절해서 해결을 하면 더 좋은 사람과 물체 상호작용을 예측하는 모델을 만들수 있습니다. 오브젝트 디텍션 물체 탐지에서도 고정박스(Anchor box)가 미리 정해 놓은 후보 박스가 없는 방법들이 있듯이 지금 이 상호 작용에서도 그런 박스가 미리 없어도 되는 기법들이 많이 제안되고 있습니다. 예를 들어 발표될 HOTR (CVPR'21)가 있습니다.

물체 탐지로 인공지능 자동차도 나올수가 있습니다. 기계학습(Machine Learning)은 경험적인 데이터를 바탕으로 지식을 자동으로 습득하여 스스로 성능을 향상하고 데이터를 기반으로 모델을 자동으로 생성합니다. 딥러닝도 기계학습의 한 방법입니다. 딥러닝은 인간의 신경망 원리를 이용해서 Deep Neural Network 을 이용한 기계학습 방법입니다.

초창기에 Multilayer Perceptron 으로 input layer 와 output layer 사이에 hidden layer 가 증가할수록 분류력이 좋아지지만 효율성의 문제가 발생할 수도 있습니다. 자율주행에서 이것은 자동차 앞에 있는 물건들을 판별하는데 사용할 수가 있습니다. 만약 앞에 도로가 있고 앞에 보이는 가장 가까운 물체가 있거나 없거나 그리고 있다면 얼마나 멀리 있는지 판별하는게 중요할 것입니다. 이 물체가 생물인지 아닌지 혹은 자동차인지 아닌지를 이 Multilayer Perceptron 을 하면 문제가 많이 생기게 되어서 딥러닝으로 CNN, RNN, Deep Reinforcement Learning 으로 더욱 더 제대로 된 사물 인지 정보를 할 수가 있습니다. 인공지능은 앞에 사물들을 인지하여 스스로 안 부딪치면서 주행할수록 알고리즘을 만들 수가 있게 됩니다.