

Hugging Face Computer Vision

컴퓨터 비전

컴퓨터 비전(Computer Vision)

컴퓨터 비전

컴퓨터 비전은 컴퓨터가 이미지나 동영상을 분석하고 해석해 인간처럼 시각적 정보를 이해하고 처리할 수 있는 기술.

이미지 분류(Image Classification) : 주어진 이미지를 사전 정의된 카테고리로 분류하는 작업.

객체 검출(Object detection) : 이미지 내에서 특정 객체의 위치와 종류를 식별하는 작업.

이미지 세분화(Image Segmentation) : 이미지의 각 픽셀을 의미 있는 영역으로 구분하는 작업.

얼굴 인식(Face Recognition) : 이미지나 비디오에서 얼굴을 식별하는 인식하는 작업.

포즈 추정(Pose estimation) : 이미지 내 사람이나 물체의 자세를 파악하는 작업.

OCR(Optical Character Recognition) : 이미지나 스캔된 문서에서 텍스트를 추출하고 디지털 형식으로 변환하는 작업.

이미지 복원(Image restoration) : 손상되거나 열화된 이미지를 원래의 상태로 복원하는 방법.

깊이 추정(Depth estimation) : 2D 이미지로부터 장면의 3D 구조와 깊이 정보를 추출하는 작업

제로샷 이미지 분류 : CLIP

제로샷 이미지 분류(Zero-shot image classification)는 학습 데이터에 존재하지 않는 새로운 Label(Class)에 대해 이미지 분류를 수행할 수 있는 기술을 의미.

제로샷 이미지 분류는 사전 학습된 멀티모달 모델을 활용해 레이블링되지 않는 이미지에 대해서도 분류가 가능하다.

이 기술은 이미지와 텍스트 간 연관성을 학습한 모델을 사용해 주어진 이미지에 대해 가장 잘 부합하는 텍스트 레이블을 찾아내는 방식.

CLIP(Contrastive Language Image Pre-training)

OpenAI에서 2021년에 발표한 멀티모달로 이미지와 텍스트간의 연관성을 학습해 다양한 컴퓨터 비전 및 자연어 처리에서 높은 성능을 발휘하는 모델.

제로샷 객체 검출 : OWLv2

제로샷 객체 검출(Zero-shot object detection)이란 기존에 학습되지 않은 새로운 객체 클래스를 검출할 수 있는 기술.

OWLv2(Vision Transformer for Open-World Localization)

OWL-ViT를 개선한 모델로 OWL은 구글에서 발표한 비전 트랜스포머 기반의 객체 검출 모델이다.