

1. According to our research, a machine learning algorithm has an accuracy of 80%. However, when we applied the algorithm to practical use, the accuracy was below the expected accuracy. Describe two or more causes to this.

Ans)

먼저 첫번째 원인으로는, 실제 데이터 중 일부가 out of distribution인 경우입니다. 실제 상황에 머신러닝알고리즘을 적용할 때, 입력 데이터의 오류입니다. 예를 들어, 강수량 예측 문제에서 기압계의 데이터가 입력되어야 하지만 태양의 온도에 대한 데이터를 입력한 경우 이는 무의미한 입력 데이터로 간주되고, 예측한 정확도보다 떨어진 결과를 만들어낼 것입니다.

두번째 원인으로는, 실제 상황에서 적용되는 test data의 불확실성입니다. Cats vs dogs 예제를 예로 들면, test data에서 blurred image가 있습니다. 카메라 렌즈의 오류 또는 센서의 오류로 인해 test 하려는 이미지의 불확실성이 높아지는 경우입니다. 이에 대한 알고리즘의 정확도를 높이기 위해 더 많은 라벨 이미지와 학습데이터를 통해 알고리즘을 학습시킬 수 있습니다.

2. Please suggest and describe two industry applicable Deep Learning projects you would carry out using a dataset of aerial images.

Ans)

첫번째로, 인공위성 이미지를 이용한 해양생물의 경로 및 위치 추적입니다. 예를 들면, 전세계적으로 멸종 위기에 있거나 개체수 관리가 필요한 해양생물의 개체수 파악은 중요한 문제입니다. 고래의 경우 위성 이미지를 이용하여 고래를 자동으로 감지하여 고래의 개체수를 평가하고 보존 조치를 취할 수 있습니다. 일반적으로 고래 탐지에 사용되는 목격 조사, 음향 조사는 비용과 시간적으로 비효율적인 측면이 있기 때문에 딥러닝 기법을 이용하여 고해상도 이미지를 통해 감지하는 방법이 더 효과적일 것입니다.

두번째로, 건물 설치를 위한 공간 추출에 사용될 수 있습니다. 고해상도 항공 이미지를 이용하여 건물 설치를 위한 적합한 공간을 분석할 수 있습니다. 딥러닝 기법을 이용하여 건설하려는 건물의 크기 및 형태를 고려하여 적절한 공간을 제시하는 것입니다. 직접 사람이 공간을 탐색하는 것보다 시각적으로 향상된 건물 객체를 생성할 수 있고, 항공 이미지를 이용하여 더욱 간편하고 정확하게 분석할 수 있을 것입니다.

3. You have a large data set consisting of high resolution aerial orthophotos. Your objective is to create an API that detects small objects within an orthophoto (e.g. Trees, Cars, People, etc). Please explain how you would create a Deep Learning Pipeline by elaborating on how you would approach the following steps. (No more than 300 words in total)

-Data Preprocessing/Labeling: 주어진 이미지 데이터가 최대한 같은 포맷을 갖추도록 하는 과정인 데이터 전처리과정, 어떤 데이터가 어떤 결과를 가리키는지 패턴을 입력하는 과정인 라벨링이다.

-Model Selection: 데이터 전처리과정과 라벨링 과정을 거친 후, 어떤 목적으로 이미지를 학습시킬 지에 따라 적합한 모델을 선별하는 과정을 거친다. Forward 또는 backward 방법을 통해 최적의 모델을 선정한다.

-Model Training: 선별된 모델 알고리즘을 통해 학습을 진행한다.

-Model Optimization/Hyperparameter Tuning: 딥러닝 모델의 학습을 수행하기 위하여 사전에 설정해야 하는 값인 하이퍼파라미터의 최적값을 분석하고 조절하는 과정입니다.

-Model Hosting/Deployment/Management: 학습시킨 모델을 최종적으로 확인하고 이를 배포하는 과정입니다. 기계학습 모델 배포를 위해 좋은 환경을 가지고 있는 다양한 플랫폼을 탐색하고 조건에 맞는 플랫폼을 선정하여 배포 및 관리를 지속적으로 진행합니다.