

유위드 3팀 프로젝트 결과보고서

■ 프로젝트 명칭 : 동물 이미지 분류를 위한 딥러닝 모델 개발

■ 프로젝트 목적

- 프로젝트 목적은 AI 모델 개발을 전반적으로 실습하며 실제 협업에서 프로젝트가 진행되는 방식을 배우는 데 있음. 단순히 알고리즘을 테스트하는 것이 아니라, 공식 문서를 찾아 이해하고 이를 내 코드에 적절하게 적용하는 과정을 통해 협업 시 필요한 기술과 문제 해결 능력을 향상시키는 것이 목표임.

■ 프로젝트 개요

깃허브(GitHub)에 대한 학습

- 버전 관리를 위한 Git의 기본 명령어를 익히고, GitHub 플랫폼을 활용하여 프로젝트를 효율적으로 관리하는 방법을 학습함.

VSCode를 활용한 개발 환경 설정

- Visual Studio Code(VSCode)를 활용하여 Python 가상 환경을 설정하고, 편리한 코딩을 위한 확장 프로그램을 활용하며, 통합된 개발 환경에서 프로젝트를 관리하는 방법을 익힘.

파이썬 기본 학습 및 라이브러리 활용

- 파이썬의 기초 문법을 익히고, 데이터 처리 및 분석에 필요한 라이브러리(Numpy, Pandas 등)를 학습하며, 이를 실제 데이터 전처리와 분석 작업에 활용함.

모델 개발의 전반적인 과정 학습

- 데이터 수집, 전처리, 학습 및 검증, 그리고 모델 저장 및 배포에 이르는 머신러닝/딥러닝 모델 개발의 전체 과정을 이해하고 실습함.

딥러닝에 대한 구체적인 공부

- Convolutional Neural Network(CNN) 등 딥러닝 알고리즘의 기본 원리를 학습하며, TensorFlow와 Keras를 활용하여 실제 이미지 분류 모델을 구현하고 학습시킴.

■ 프로젝트 참가자

성명
이준서
조정민
김지안
김민지
조수연

■ 프로젝트 세부시행 결과

데이터셋

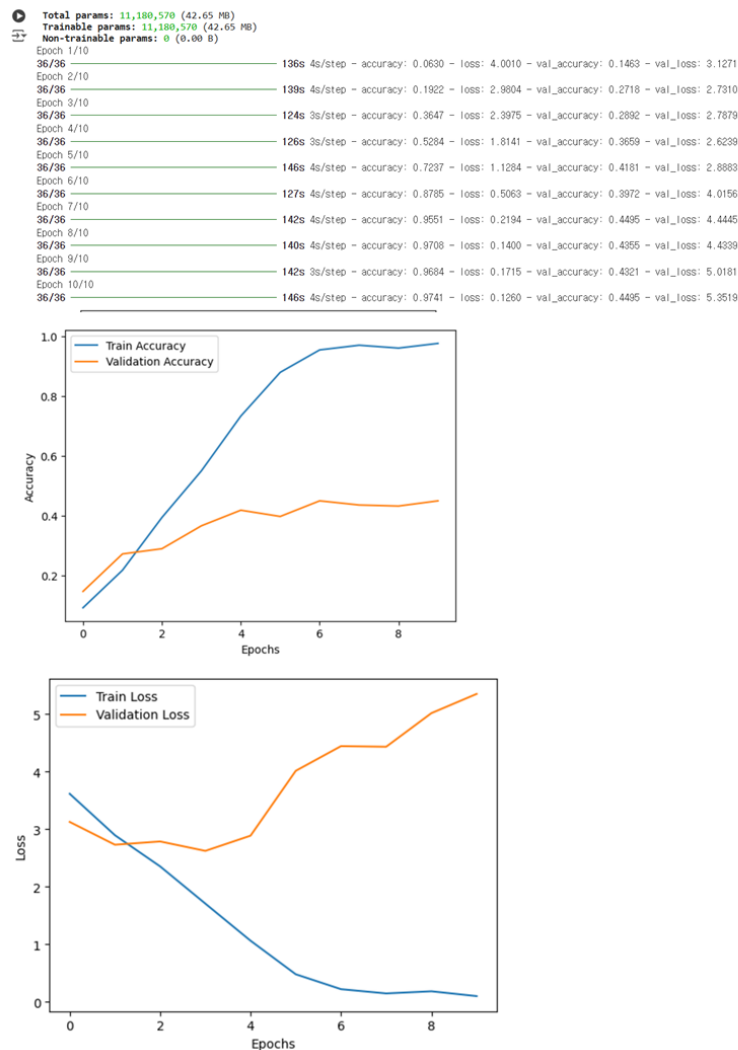
- 동물 이미지 분류라는 문제에 대해 Kaggle에서 ¹데이터셋을 가져옴.
- 90종의 동물로 분류된 5400개의 이미지를 224*224로 리사이즈하고, 훈련/검증 8:2의 비율로 데이터를 분리함.

모델 학습

- 이미지 데이터라는 특수성으로 인해 CNN 라이브러리를 사용하여 딥러닝 신경망 학습을 시킴.
- 3번에 걸친 컨볼루션-풀링으로 이미지 특징을 추출하는 모델을 설정함.
- Batch size=32, Epoch=10 으로 데이터셋을 학습시킴

학습 결과

- 위의 모델을 통해 에포크-정확도, 에포크-손실값 의 두가지 그래프로 모델을 시각화함.
- GPU 성능의 문제로 인해 Google Colab을 사용하였고 더 적은 데이터셋을 사용함.



¹ <https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals>

- 데이터셋이 늘어날수록 정확해지는 CNN의 특성상, Google Collab 내에서 돌릴수 있는 데이터셋의 부족으로 인해 모델의 정확도가 떨어짐. 그러나 epoch의 값에 따라 정확도가 우상향하는 경향을 보이는 걸로 보아 충분한 이미지 학습을 통해 좋은 모델로 발전할 가능성을 봄.
- 실제 훈련데이터와 검증데이터의 차이의 원인 중 하나로 동물들의 종류가 90개나 되었다는것도 한가지의 문제요소가 될 수 있을 것이라 판단됨. 부족한 데이터셋에 많은 동물들의 데이터를 학습시키다보니 각각의 동물들의 특징을 잡는 리소스가 부족하지 않았나 하는 생각이 듦. ‘강아지인지, 고양이인지, 아니면 다른동물인지’ 와 같은 데이터셋을 가지고 모델을 만들었다면 특징을 충분히 잡을정도의 이미지 데이터를 학습시킬 수 있어 더욱 좋은 결과를 얻었을 것이라 여겨짐.