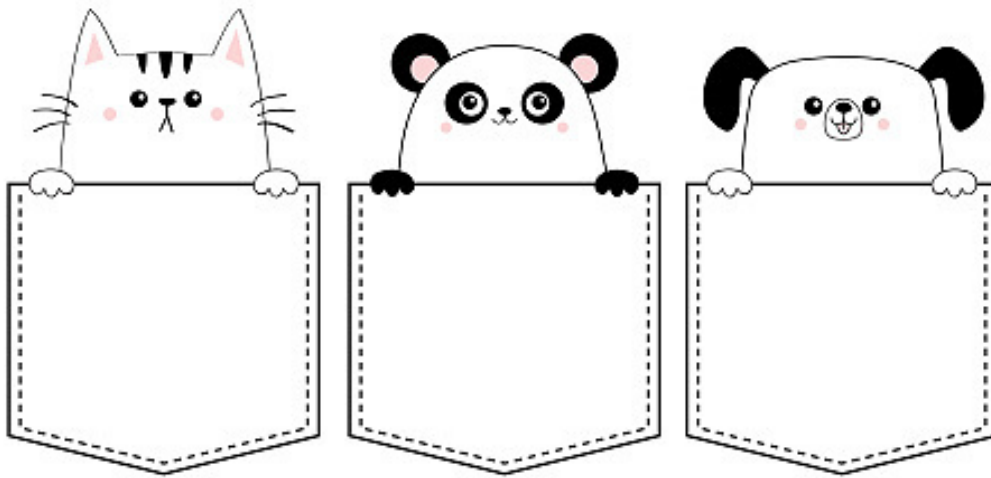




Animal Image Dataset Classification

☰ Kaggle URL	Animal Image Dataset(DOG, CAT and PANDA). Kaggle
👤 팀원	④ 윤재은 준혁 준혁 박권 권소정 경배 경배 김이 이성진
▼ 팀번호	2팀
🕒 생성일	@July 11, 2022 9:09 AM



1. 프로젝트 개요



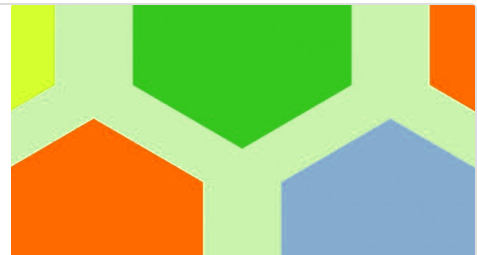
Animal Image Dataset(DOG, CAT and PANDA)을 활용하여
CNN Image Classification Model 학습하기.

Animal Image Dataset(DOG, CAT and PANDA)

Animal Image Dataset(DOG, CAT and PANDA)

Dataset for Image Classification Practice

[k https://www.kaggle.com/datasets/ashishsaxena2209/animal-image-datasetdog-cat-and-panda](https://www.kaggle.com/datasets/ashishsaxena2209/animal-image-datasetdog-cat-and-panda)



2. 분석 데이터 셋

animals

| cats: "cats_0001.jpg ~ cats_1000.jpg (1,000 images)

| dogs: "dogs_0001.jpg ~ dogs_1000.jpg (1,000 images)

| panda: "panda_0001.jpg ~ panda_1000.jpg (1,000 images)

※ All images consist of RGB 3 Channels and have different sizes!!

3. Image Load & Preprocessing

Image Load & Preprocessing

Aa 이름	🕒 생성일	☰ 태그
윤재은	@July 14, 2022 1:18 PM	
권소정	@July 14, 2022 1:18 PM	
김경배	@July 14, 2022 1:18 PM	
박준혁	@July 14, 2022 1:18 PM	
이성진	@July 14, 2022 1:18 PM	

4. 모델 구현

모델 구현

Aa 이름	🕒 생성일	☰ 태그
윤재은	@July 14, 2022 1:18 PM	
권소정	@July 14, 2022 1:18 PM	
김경배	@July 14, 2022 1:18 PM	
박준혁	@July 14, 2022 1:18 PM	
이성진	@July 14, 2022 1:18 PM	

5. 프로젝트 회고

a. 프로젝트를 진행하며 어려웠던 점



윤재은

val_accuracy가 낮게 나와서 이미지 사이즈 바꾸기, 레이어 더 쌓기, 컴파일 파라미터 변경 등 여러가지를 적용하여 모델 개선을 해보고 싶었는데 시간문제로 시도를 많이 못해본 점이 아쉽다.



권소정

backend ai에서 진행했음에도 처음에 커널이 자꾸 죽어서 프로젝트 실행조차 어려웠는데, 다른 팀원 분이 숫자를 줄이고 reshape한 이미지를 저장한 뒤 불러와서 다시 해보라고 알려주셔서 다행히 프로젝트를 진행할 수 있었다..!) ㅎㅎ 모델링과 데이터 증강을 여러가지 방법으로 시도해보았는데 정확도가 낮게 나온 점이 너무 아쉬웠다.. 더 많이 해보고 싶었는데 자꾸 커널이 죽어서 여러가지 시도를 해보기에 어려움이 있었다.
사실 이미지를 불러오는 과정부터 애를 먹었다.. ㅎㅎㅜㅜ 아직 CNN 자체가 많이 어렵게 다가온다.. 연습을 더 많이 해봐야겠다.



김경배

과대적합과 성능을 해결하는 것이 쉽지 않은 과정이었다. 모델링과 증강을 통해 개선해보았지만 적은 이미지 수가 아쉬워 이미지를 추가적으로 구하여 표본을 늘렸다.
하지만 학습 속도도 문제지만 backend ai에서 진행해도 터졌다.

프로젝트 진행 시간 = 기다리는 시간



박준혁

우선 개발 환경을 만드는 것부터 쉽지 않았다. 로컬에서 작업하기엔 Image의 용량이나 CNN Model의 Parameter를 감당하기 어려울 거 같아, Google Colab에서 작업을 진행하였는데, 데이터를 다운받고, Google Drive에 Zip을 풀고, 또 Amount 하는데 꽤 시간이 많이 걸렸다.

GPU, RAM 등 HW적으로 아쉬움이 있었다. Image를 Resizing할 때 데이터 손실을 최소화하려면 최대한 큰 크기로 Resizing 해야 하는데, 메모리 용량 때문에 64x64로 자르는 것이 최선이였다. 또한, Image가 1,000개 밖에 없어, 과대적합이 일어나는 경향을 보여 데이터 증강을 시도했는데 이 역시 용량 문제로 다소 한정적인 증강만 가능했다.



이성진

시간이 부족한데 용량이 큰 이미지 데이터를 다루는 것이 어려웠다. CNN 개념에 대해서 나름 복습해보면서 실습을 해보려했는데 개념과 달리 코드를 적용하는 게 쉽지 않았다.

b. 프로젝트를 통해 배운 내용과 의견



윤재은

드디어(이제서야..) CNN 전체적인 흐름에 대해 깨달았다!!! 역시 직접 쳐보고 고민하는 과정을 겪어야 한다. 또한 다른 조원들의 결과물을 보며 함수적용하는 코드라든가, 튜닝하는 법 등을 배울 수 있었다.



권소정

처음에는 프로젝트 시작부터 막막했는데 다른 팀원 분이 결과물을 공유해주셔서 코드를 따라해보면서 전체적인 흐름을 어느정도 익힐 수 있었다. 혼자서 층을 쌓아보고 데이터 증강을 해봤을 때는 정확도가 다른 분들에 비해 낮게 나온 것 같지만, 데이터증강의 parameter를 좀 더 살펴보고 전체적인 흐름을 익혔다는 사실에 나름 만족했다!

모델 성능을 높이는 부분은 아직까지 감이 안 잡히긴 한다.. 그래도 이번 프로젝트는 다른 분들의 결과물을 보며 많이 배울 수 있는 프로젝트였다.)



김경배

loss, accuracy 그래프를 보며 성능을 높이면서 과대적합을 해결하기 위한 방법을 다양하게 실습할 수 있던 프로젝트였다. 왜 딥러닝 랩실에 좋은 컴퓨터가 있는지 몸으로 느꼈다.



박준혁

과대적합을 해결할 다양한 테크닉을 더 알아야 할 것 같다. 보통 모델을 복잡하게 짜고, 튜닝을 통해 과대적합을 해결하는 방향으로 모델을 개발하기 때문에 어떻게 Overfitting을 해결하고, 일반화를 강화하느냐가 모델의 정확도 성패에 열쇠라고 할 수 있을 것 같다.



이성진

CNN의 개념에 대해 복습해보면서 개념은 어느정도 잡힌거 같다. 모델을 적용하는 시간이 부족해서 좀더 실습해보면서 코드에 대한 이해도도 높이고 싶다!!

c. 프로젝트 만족도

Aa 이름	# Completed	# Goal	Σ 만족도
-------	-------------	--------	-------

Aa 이름	# Completed	# Goal	Σ 만족도
<u>윤재은</u>	7	10	*****🏠*** 70%
<u>권소정</u>	6	10	*****🏠*** 60%
<u>김경배</u>	8	10	*****🏠** 80%
<u>박준혁</u>	8	10	*****🏠** 80%
<u>이성진</u>	7	10	*****🏠*** 70%