2018년 2학기 인공지능 최종 과제

HUMPBACK WHALE IDENTIFICATION CHALLENGE

# 주제 설명

멸종해가는 고래종의 보호를 위해 과학자들은 사진 감시 시스템을 사용해 해양활동을 감시하는데 고래의 꼬리 모양과 독특한 표식을 토대로 분석하여 고래 종을 분류 하고 고래 꼬리의 움직임을 자세히 기록합니다. 이러한 작업들은 40년간 과학자들의 수작업으로 수행 하였으나 이는 한계가 있습니다. 이 Challenge에서는 고래 종을 식별하는 프로그램을 만들어야 합니다. 연구 기관 및 공공기관에서 수집한 이미지로 구성된 Happy Whale 데이터 베이스를 통하여 분석을 진행합니다.

# 조team 설명

조 이름 KMAI\_MATRIX

학번\_이름 20143043\_김윤성

수행한 업무 분장 100% 혼자

# 적용한 기계 학습 상세 및 구현 방법

1. 세부 설명

딥러닝 구현 라이브러리인 파이토치를 이용. 주어진 train데이터로 모델을 학습시켜 test 데이터들의 ID를 예측하는 과제입니다.

ResNet – 잔류학습이용, 중간에 지름길을 만들어 신호감쇠 해결한 모델, 전역평균 풀링을 사용하여 매개변수의 수를 줄였습니다.

1. 구현 환경, 구현 방법

윈도우7, 파이썬 3.7버전, PyTorch 1.0, cuda 9.0, 주피터노트북

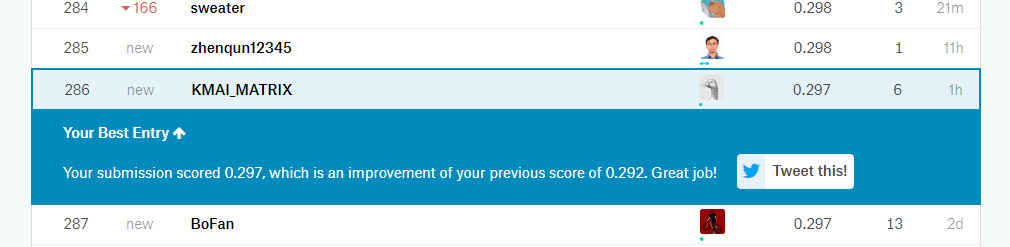
Competition에 open되어 있는 커널들을 참고하여 작성 했습니다. ResNet34 모델을 채용하여 구현하였습니다.

1. 코드 상세 설명 (반드시, 관련 코드 pdf와 함께 압축해서 첨부할 것)

코드는 세 개의 파일로 정의 되어 있습니다. my\_utils에는 학습에 필요한 여러 메소드와 클래스들이 구현되어있습니다. KaggleProject파일은 메인 파일로 학습의 전체적인 과정이 코드로 담겨 있습니다.

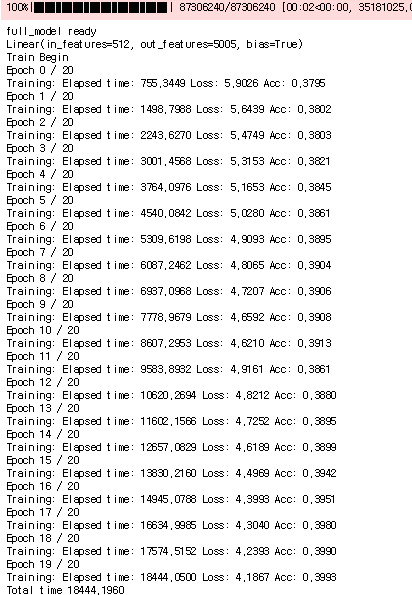
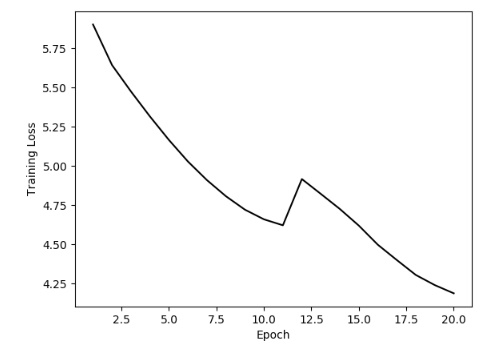
정의된 Whale데이터 로더에 train 이미지들과 라벨이 표시된 csv파일을 싣고 pretrained된 resnet34모델과 함께training합니다. 손실함수로 교차엔트로피를 사용하고 optimizer로는 처음에 아담을 선택했으나 loss값이 너무 크고 정확도가 좋지않아 스토캐스틱 경사하강법을 채택했습니다. 훈련이 끝난 모델은 저장해두었다가 test 데이터를 predict하는 작업을 수행합니다.

# 결과



# 결론

처음에 주제 선정을 할 때, protein competition을 점수를 더 쳐주겠다고 하셨지만 너무 큰 데이터양에 겁을 먹었고 또한 딥러닝과 캐글 초보자이기 때문에 완료된 competition인 Humpback whale을 택했습니다. 커널이나 소스코드들이 많이 오픈이 되어 있다고 해서 몇 개를 보았지만 제가 사용하는 PyTorch가 아닌 케라스나 텐서플로우가 주를 이루고 있어서 굉장히 애를 먹었습니다. 도저히 온전한 제 힘만으로는 코드를 작성하기 어려워 오픈된 커널들을 가져오거나 참고하여 코드를 구현하게 되었습니다. 테스트 하던 중 제 노트북과 데스크탑으로는 cuda지원이 되지않아cpu를 활용했지만 너무나도 느려서 하루를 꼬박 켜놓아야 했습니다. 심지어 구현한 코드에 문제가 있었던 것인지 정확도가 5프로가 채 나오지 않는 참사가 발생하기도 했습니다. 인공지능은 지원이 빵빵한 곳에서 해야 된다는 세미나 연사님 들의 말씀이 떠오르던 순간이었습니다. 하는 수 없이 PC방에서 3~4일간 최종 프로젝트 마무리작업을 하게 되었는데 점수가 제 예상보다 좋지는 않지만 그래도 내가 짠 코드가 돌아는 가는구나 라는 뿌듯함을 느껴봤습니다. 제가 생각하기에 점수가 잘 안나온 이유는 subset을 활용한 pretrain의 부재와 모델의 학습률을 적용하는데 약간 문제가 있다고 짐작하고 있습니다. 사실 기말고사 및 다른 프로젝트들 일정에 치여 문제점을 모두 고쳐나가지는 못하였으나 교수님 말씀대로 방학이라는 시간을 이용해서 이번학기 인공지능에 한발자국 내딛은 발걸음을 조금 더 전진시켜보고 싶습니다. 인공지능 수업에서 배운 내용들을 캐글이라는 좋은 기회로 마무리 하게 되어서 굉장히 보람찬 한 학기 였다는 생각이 듭니다. 감사합니다.



Ps. 교수님 Humpback whale competition 중간에 데이터셋이 바뀌고 (1GB->6GB) 상금도 업데이트되어 새로 오픈됬습니다ㅠ.