BirdCLEF 2024 Report

1. 대회

텍스트, 과일, 스크린샷, 음식이(가) 표시된 사진

자동 생성된 설명

* 1. 대회명 : BirdCLEF 2024
  2. 대회 소개
     1. ‘Kaggle’의 ‘BirdCLEF 2024’는 인도 남서부 해안을 따라 뻗어 있는 웨스턴 가츠 산맥에 서식하는 조류의 음성을 통해서 조류를 분류(식별 혹은 탐지)하고 예측하는 대회이다.
  3. 대회 규칙 – 제약 사항
     1. Inference note book(submission)
        1. ‘Kaggle’ 노트북만을 사용하야여 한다.
        2. Inference, Test(by secret test data set) : CPU만을 사용하여야 한다.
        3. Inference, Test(by secret test data set) : 런타임이 2시간을 넘을 수 없다.
        4. Inference, Test(by secret test data set) : 인터넷 사용이 불가능하다.
        5. Inference, Test(by secret test data set) : pre-trained 모델의 사용은 가능하다.

즉 훈련과 검증은 GPU로 하되, CPU로만 테스트셋과 스코어링셋을 진행해야한다.

* 1. 음성 분류?
     1. 음성분류의 역사
        1. AudioClassification은 텍스트 분류와 마찬가지로 입력 데이터에 클래스 레이블 출력을 할당한다. 다만 텍스트 대신 오디오 형이 있다는점이 차이점이다.
        2. 초기 음성인식은 음성 데이터의 패턴 분류로부터 시작하였다. HMM(Hidden Markov Models)이 도입되면서 음성인식이 발전하게 되었고 이후 GMM-HMM(Gaussian Mixture Model-Hidden Markov Model)이 나오면서 발전하기 시작하였다. 이 모델은 초기 음성 인식 시스템에 효과적이었다. 이후 딥러닝이 나오고 딥러닝 기반 모델들이 개발되기 시작하면서 음성 분류에 비약적인 발전이 있었고 많은 딥러닝 모델(CNN, RNN, Transformer등)들이 음성인식 모델을 대체하게 되었다. 현재는 Trasformer, Pre-trained Language Models 등 다양한 모델들이 활발히 개발되고 연구되고 있다.
     2. 딥러닝 음성분류의 대표적인 모델
        1. CNN : 본디 음성데이터를 다루기 위한 딥러닝 모델이었으나, 음성 데이터를 이미지로 변환하여 사용하면 높은 정확도를 자랑한다. 음성 데이터를 시간-주파수 형태로 처리. Spectrogram과 같은 음성 입력데이터를 받아 음성의 특징을 추출하고 분류한다.
           1. AlexNet
           2. ResNet
           3. EfficientNet
        2. RNN : 시퀀스 데이터를 처리하는데 특화된 딥러닝 모델이다. 음성 데이터 자체도 시퀀스 데이터이기 때문에 음성데이터 처리에도 효과적이다.

1. 데이터
   1. 데이터 개요
      1. 인도 남서부 해안지역인 웨스턴 가츠 산맥에 서식하는 조류의 음성 데이터이다.
   2. 데이터설명

Xeno-canto라는 전 세계의 새소리를 수집하고 공유하는 사이트에서 제공하는 데이터를 기반으로 metadata가 이루어져 있다.

* + 1. primary\_label : 새의 종
    2. secondary\_labels
    3. type : 새의 울음 / 노래 구분
    4. latitude, longitude : 녹음이 기록된 장소
    5. scientific\_name : 학명
    6. common\_name : 향명
    7. author : 녹음자
    8. licnese : 라이센스
    9. rating : 녹음의 품질 (0 ~ 5)
    10. url : xeno-canto에서 해당 새 소리의 url
    11. filename : 파일 이름 (XC~.ogg)
    12. filepath : 캐글데이터셋 내 파일 위치
    13. target : 종 구분을 숫자로표시 (0 ~ 182)
    14. xc\_id : xeno-canto 아이디
  1. 데이터 수
     1. 총 데이터 수 : 24459개
        1. Traindataset : Validation dataset = 19567 : 4892
     2. 레이블의 수 : 182(종)
  2. 예시
     1. Asian Brwon Flycatcher (쇠솔딱새)

새, 야외, 명금류, 하늘이(가) 표시된 사진

자동 생성된 설명

파형과 스펙토그램

스크린샷, 라인, 텍스트, 그래프이(가) 표시된 사진

자동 생성된 설명

* + 1. Marsh Sandpiper (쇠청다리도요)

새, 수생 조류, 물, 도요과이(가) 표시된 사진

자동 생성된 설명

파형과 스펙토그램

스크린샷, 그래프, 텍스트, 라인이(가) 표시된 사진

자동 생성된 설명

* + 1. Zitting Cisticola (개개비사촌)

새, 야외, 참새, 잔디이(가) 표시된 사진

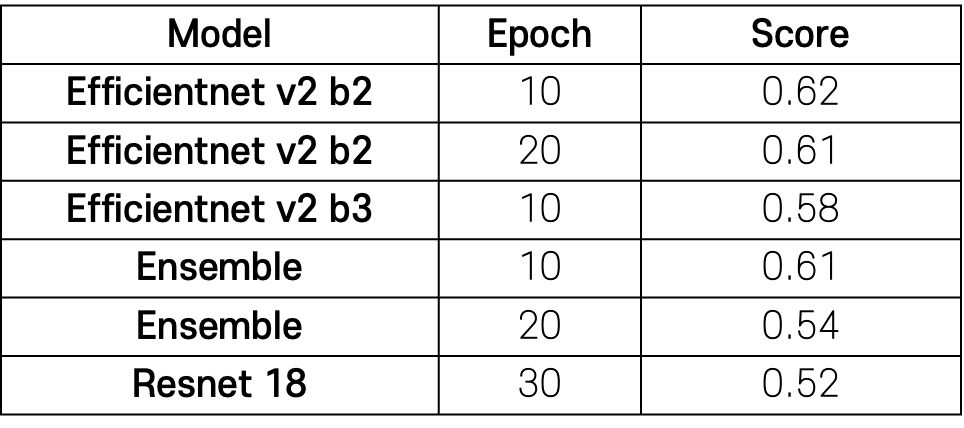
자동 생성된 설명

파형과 스펙토그램

스크린샷, 텍스트, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

1. 전처리
2. 모델링
3. 결론
   1. 모델별 성능
      1. 사용 모델별 성능(표)



* + 1. 사용 모델별 성능(그래프)

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

* 1. 성능에 대한 고찰 및 개선점
     1. Pretrained의 유무
        1. Efficientnet v2 b2(Epoch – 10) - imagenet vs Efficientnet v2 b3(Epoch – 10)

Efficientnet v2의 2가지를 비교해보면 b3의 parameter수가 b2에 비해 약 300만개 많음에도 불구하고 성능은 역으로 0.04 감소하는 결과가 나타났다.

모델이 대규모 이미지셋을 통해서 사전학습을 진행하는것은 곳 인간이 시각세계에서 데이터를 학습하여 일반화하는것과 동일한 원리이다. 즉 다양한 상황에서 일반화된 모델로서 기능하며 성능의 향상을 기대할 수 있기때문에 사전학습의 유무가 성능에 많은 영향을 미쳤다.

* + 1. Epoch에 대하여
       1. Efficientnet b2와 앙상블 기법은 총 2가지의 에포크(10, 20)를 가지고 학습을 진행하였다. Efficientnet에서 에포크의 수가 올라갔는데도 불구하고 성능은 역으로 낮아진 것을 보아 과적합이 발생했다는것을 알 수 있다.(앙상블 기법 또한 개별모델의 과적합을 줄이기 위해 사용하였으나 이 또한 낮아졌다) 대회 기한이 짧고 훈련하는데 많은 시간이 소모되었기에, 시간이 충분하다면 Earlystopping을 활용하여 과적합이 되기 직전의 최적점을 확인하고, 추가적인 스코어링을 진행해보는것도 좋을 것 같다.
    2. 모델의 크기
       1. 파라미터 수 와 성능

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

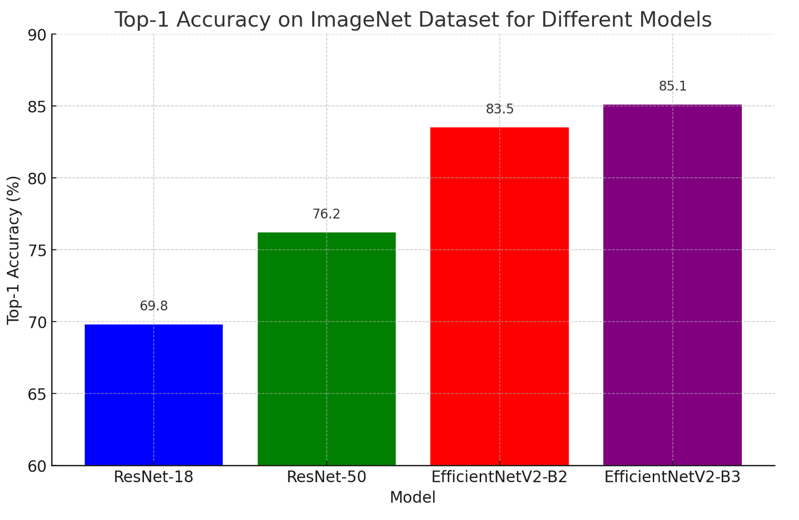
자동 생성된 설명

텍스트, 폰트, 번호, 스크린샷이(가) 표시된 사진

자동 생성된 설명

도표, 스크린샷, 라인, 직사각형이(가) 표시된 사진

자동 생성된 설명



* + 1. 다양한 전처리 기법
       1. 이번 공모전에서는 mixup과 Frequency, Time Masking만을 이용하여 데이터 증강을 진행하였다.
       2. 추후 시간이 된다면 노이즈를 추가하거나, Shifting(원본데이터 좌우이동), Streching(빠르기 조절)등을 활용하여 추가적인 데이터 증강을 진행 한 후 스코어링을 진행한다면 더 좋은 결과를 가져올 수 있을것이다.
    2. Metric 조정
       1. 이번 공모전에서서 Scoring metrics는 sklearn에서 제공하는 roc-auc 스코어링 방법활용한 메트릭스이었으나, 우리팀은 편의상 auc메트릭스만을 가지고 구현을 진행하였다. 후에 가능하다면 roc-auc를 구현 후 다양한 모델에 적용해보는 것도 좋을 것 같다.