층간 충격원 및 방사소음 예측모델 연구 보고서

12230617 황주영

12201967 최우진

본 연구는 층간 충격 소음 데이터(.wav)를 기반으로 소음의 크기(dB)를 예측하고 소음을 분류하여 파악하는 멀티태스크 모델 개발을 목표로 한다. 초기 베이스라인 모델에서 시작하여, 최신 딥러닝 아키텍처와 데이터 특성을 고려한 고도화된 기법들을 점진적으로 적용하며 성능을 향상시키는 과정을 체계적으로 작성하였다. 최종적으로는 음향 정보뿐만 아니라 파일명에 포함된 물리적 메타 정보(층수, 세기, 거리 등)를 결합하여 예측 정확도를 극대화하는 방안을 제시하고 그 효과를 입증한다.

1. 모델 설계

1.1 데이터 및 전처리

- 데이터셋: 5가지 충격 소음원(book, chair, desk, hammer, lecturestand)으로 구성된 음향 파일(.wav)과 해당 파일의 소음 크기(dB) 라벨로 구성된다.
- 문제 정의: 오디오 파형을 입력받아 소음 크기(dB)를 예측하는 회귀 문제와 소음원을 분류하는 분류 문제를 동시에 해결하는 멀티태스크 학습을 목표로 한다.
- 데이터 분할: 제공된 train_split, val_split, test_split CSV 파일을 기준으로 훈련, 검증, 테스트셋을 구성한다.

특징 추출

- MFCC (초기 CNN 모델): 초기 VGG16 모델에서는 n_mfcc=40으로 설정하여 MFCC를 추출했다.
- Log-Mel Spectrogram : EfficientNet, ConvNeXt 등 모델에서는 n_mels=128로 설정하여 더 풍부한 주파수 정보를 담고 있는 Log-Mel 스펙트로그램을 기본 특징으로 사용했다.
- 입력 데이터 형식: [0, 1] 범위로 정규화한 후, 3채널 이미지(224x224x3) 형태로 변환하여 입력으로 사용했다.

메타 정보 추출 및 활용

- 2F_book_1_3_2.wav와 같은 파일명에서 정규 표현식을 사용하여 층수, 소음원, 세기, 거리, 반복 5가지 메타 정보를 추출했다.
- 추출된 메타 정보 중 floor, intensity, distance, repeat 4가지 수치형 피처를 StandardScaler를 이용해 표준화하여, 값의 범위 차이가 모델 학습에 미치는 영향을 최소화했다.

1.2 모델 아키텍처

모델은 점진적인 성능 개선을 목표로 총 5단계에 걸쳐 개선하였다.

1. CNN (VGG16 기반 멀티태스크 모델) Baseline을 참고한 기본 모델

- ImageNet으로 사전학습된 VGG16 모델의 top 레이어를 제외하고 사용했다.
- 백본의 출력에 GlobalAveragePooling2D를 적용한 후, 여러 Dense 레이어를 거쳐 회귀와 분류를 위한 두 개로 분리하여 구성하였다.
- Fine-tuning을 사용

  Stage 1 (10 epochs): 백본의 가중치는 유지한채 학습

  Stage 2 (136 epochs): 마지막 컨볼루션 블록의 동결을 해제하여 헤드와 함께 미세조정

2. EfficientNet-B0

- VGG16보다 더 효율적인 EfficientNet-B0를 백본으로 교체
- 입력 특징을 MFCC에서 Log-Mel Spectrogram으로 변경하여 더 세밀한 음향 정보를 활용했다.
- MFCC는 소리를 수학적으로 요약하면서 세부 정보가 일부 사라지지만, Log-Mel은 주파수–시간별 변화를 그대로 보존하여 더 세밀한 음향 특징을 제공한다. 특히 층간 충격음처럼 짧고 날카로운 신호는 Log-Mel에서 뚜렷하게 나타나므로, 모델이 이를 학습하여 성능을 높이는 데 도움이 된다.

- VGG16과 유사하게 초기에는 백본을 동결하여 학습하고, 이후 마지막 N개 레이어의 동결을 해제하여 미세조정하는 2-Stage 방식을 사용했다. 또한 EarlyStopping과 ReduceLROnPlateau 콜백을 통해 학습을 안정화했다.

3. 개선한 EfficientNet-B0

- 본질적으로 1채널 데이터임에도 3채널로 복제하여 사용했던 비효율을 개선했습니다. 모델의 첫 번째 컨볼루션 레이어를 1채널 입력을 직접 받도록 수정하여 파라미터 수를 줄이고 정보 왜곡 가능성을 최소화했다.
- Huber Loss: 이상치에 덜 민감하여 안정적인 학습이 가능한 Huber Loss를 회귀 손실 함수로 채택했다
- EMA : 훈련 중 모델의 가중치를 지수이동평균으로 관리하여, 일시적인 그래디언트 변화에 덜 흔들리고 더 안정적인 성능의 모델을 확보했다.
- 회귀 출력 레이어의 편향을 전체 훈련 데이터셋의 dB 평균값으로 초기화했다. 이는 모델이 학습 초반부터 안정적인 범위 내에서 예측을 시작하게 하여 수렴 속도를 높였다.

4. ConvNeXt-Tiny

- Vision Transformer와 결합된 ConvNeXt-Tiny 모델을 사용하였다.
- ConvNeXt의 큰 커널 사이즈가 스펙트로그램의 주파수-시간 패턴을 포착하는 데 유리할 것으로 판단했다.
- 각 스펙트로그램을 개별적으로 Z-score 정규화하여, 녹음 환경이나 소리 크기에 따른 전체적인 편차를 줄이고 모델이 패턴 자체에 집중하도록 했다.
- Time Shifting을 사용하여, 훈련 시 스펙트로그램을 시간축 방향으로 무작위 이동시키는 데이터 증강 기법을 도입하여, 소음 발생 시점 변화에 대한 모델의 안정성을 높였다.

5. ConvNeXt-Tiny + 추가 피쳐 활용

- 별도의 MLP가 표준화된 4가지 메타 피처를 입력받아 메타 임베딩 벡터(m_emb)를 생성한다.
- 음향 특징 벡터와 메타 임베딩 벡터를 결합하여 최종 특징 벡터를 만든다.
- 결합된 특징 벡터를 입력으로 받아, 회귀 헤드는 dB 값을 예측하고 분류 헤드는 소음원을 예측한다.

2. 실험 결과 (Experiments)

각 모델의 테스트셋 기준 최종 성능은 표로 나타내보았다. MAE는 낮을수록, F1-Score는 높을수록 우수한 성능을 의미한다.

단계	모델	핵심 변경 사항	TEST MAE (dB)	TEST F1-score
1	CNN (VGG16)	Baseline (MFCC, 2-Stage FT)	2.592	0.98
2	EfficientNet-B0	Log-Mel 특징, EfficientNet 백본	2.215	0.988
3	개선된 EffNet-B0	1채널 입력, Huber Loss, EMA, 바이어스 초기화	1.684	1
4	ConvNeXt-Tiny	ConvNeXt 백본, 샘플별 정규화, 시간축 증강	1.203	0.997
5	ConvNeXt-Tiny + Meta	메타 정보(층수, 세기, 거리, 반복) 결합	1.144	0.997

학습 곡선: CNN과 EfficientNetB0 초기 모델은 검증 손실에 다소 변동이 있었으나, 이후 개선 모델은 Huber Loss와 EMA 도입 이후 학습 초반부터 매우 안정적으로 손실이 감소하는 경향을 보였다.

3. 성능 분석

본 연구의 핵심은 체계적인 실험 설계를 통해 각 단계의 기술적 변화가 모델 성능에 미치는 영향을 명확히 분석하는 데 있다. 성능 향상은 단일 요인이 아닌, 모델 아키텍처, 학습 전략, 그리고 최종적으로 부가적인 정보 활용을 통해 이루어졌다.

초기 VGG16 모델에서 EfficientNet-B0로 전환

- 입력 특징을 MFCC에서 Log-Mel 스펙트로그램으로 변경하였다. MFCC가 음성의 특징을 요약, 압축하는 데 중점을 둔다면, Log-Mel 스펙트로그램은 충격 소음의

순간적인 주파수 변화와 시간적 패턴을 시각 정보로써 거의 손실 없이 보존하기 때문이다.

- VGG16 대비 EfficientNet-B0는 더 적은 파라미터로 높은 성능을 내는 현대적인 아키텍처이다. 이는 더 풍부해진 Log-Mel 정보로부터 핵심적인 특징을 효과적으로 추출하고, 불필요한 과적합 가능성을 줄여 일반화 성능을 높였다.

개선한 EfficientNet-B0

- 기존 3채널에서 1채널로 직접 처리하도록 모델 구조를 변경하였다. 이는 단순히 파라미터를 줄이는 것을 넘어, 모델이 데이터의 원본 형태를 왜곡 없이 학습하게 하여 초기 수렴 속도와 최종 성능 모두를 개선되었다.
- Huber Loss는 dB 값의 예측이 크게 빗나가는 일부 이상치 데이터에 대한 민감도를 줄여, 전체적인 학습 과정을 안정시켰다. 여기에 EMA기법을 더하여, 훈련 중 발생하는 가중치의 급격한 변화를 부드럽게 만들어 최적점에 더 안정적으로 수렴하도록 하였다.
- 회귀 헤드의 편향을 학습 데이터의 dB 평균값으로 초기화하여, 모델이 탐색을 무작위 지점이 아닌 가장 확률 높은 지점에서 시작하도록 하였다.

ConvNeXt-Tiny

- 샘플별 정규화를 통해 각 소음 데이터의 전체적인 볼륨 차이를 제거하였다. 이를 통해 모델은 소리의 절대적인 크기가 아닌, 충격원 고유의 상대적인 패턴에 집중하여 학습하게 되므로 일반화 성능이 크게 향상된다.
- Time Shifting 데이터 증강을 사용하여 안정화하였다.

ConvNeXt-Tiny + Meta-feature

- 이전 단계까지 모델은 오직 '소리'만 듣고 dB를 예측했다. 하지만 동일한 '책 떨어지는 소리'라도 높은 층에서, 강하게, 가까이서 발생했다면 dB 값은 달라지기 때문에 해당 부분에 대한 정보를 추가하였다.

- 별도의 MLP를 통해 처리된 층수, 세기, 거리 등의 메타 정보를 ConvNeXt가 추출한 음향 특징과 결합함으로, 어떤 조건에서 발생한 소리인가를 종합적으로 판단하게 되어, 예측의 모호성을 해소하고 최종 MAE를 1.144dB까지 낮추는 결정적인 역할을 수행했습니다.
-

4. 결론

본 연구는 층간 충격 소음 예측 모델의 성능을 체계적으로 고도화하는 것을 목표로, 총 5단계에 걸친 점진적 개선 과정을 수행했다.

연구 결과, VGG16 기반의 베이스라인 모델(MAE 2.592dB)에서 시작하여, 최종적으로 물리적 메타 정보를 결합한 ConvNeXt-Tiny 모델을 통해 테스트 MAE 1.144dB라는 높은 예측 정확도를 달성할 수 있었습니다. 이는 초기 모델 대비 오차를 약 55.8% 감소시킨 수치입니다.

1. 학습 전략의 중요성: 단순히 최신 모델을 사용하는 것을 넘어, Log-Mel 스펙트로그램과 같은 풍부한 특징 표현과 1채널 입력 최적화, Huber Loss, EMA 등 학습 전략이 성능 향상의 필수적인 기반임을 확인하였다.
2. 정보 결합: 음향 특징과 물리적 메타 정보의 결합이 예측 성능을 최종적으로 높이는 결정적인 요인이었다.. '층수', '세기', '거리'와 같은 객관적인 물리적 조건은 소리만으로는 파악하기 어려운 미세한 dB 차이를 구분하는 결정적인 단서로 작용했으며, 이를 통해 모델은 단순한 패턴 인식을 넘어 상황을 종합적으로 이해하고 추론할 수 있었다..

결론적으로, 본 연구는 복잡한 현상을 다루는 딥러닝 문제에서 고성능 모델을 개발하기 위해서는 우수한 아키텍처, 안정적인 학습 기술, 그리고 문제의 본질을 담고 있는 컨텍스트 데이터의 활용이 모두 유기적으로 결합되어야 함을 보여주었다.