

포렌식을 위한 MFCC 기반의 오디오 음질 분류 방법

서영은*¹, 안상근*², 이해연**

MFCC-based Audio Quality Classification Method for Forensics

Young-Eun Seo*¹, Sang-Geun Ahn*², and Hae-Yeoun Lee**

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00242116)

요약

통신, 엔터테인먼트, 보안 등 다양한 분야에서 오디오의 활용이 증가함에 따라서, 그 음질을 평가하여 응용하는 것이 중요해지고 있다. 특히, 오디오 포렌식 분야에서 오디오의 고음질과 저음질을 신속하게 판별하는 기술은 불법적인 녹취 데이터를 분류하는데 기여한다. 본 논문에서는 변형된 DenseNet 기반의 음질 분류 모델을 통해 오디오의 음질을 판단하는 방법을 제안한다. 3초 단위 오디오를 입력받아, MFCC(Mel-Frequency Cepstral Coefficients)를 이용해 2차원 신호 특징을 추출하였고, 변형된 DenseNet 모델에 적용하여 음질 판단에 활용하였다. 국내외 데이터셋과 자체적으로 수집한 오디오를 사용하여 실험을 수행하였다. CNN 모델을 사용하였을 때 86.7% 정확도를 보였으나, 제안한 변형된 DenseNet 모델을 사용하였을 때 94% 정확도를 달성하였다.

Abstract

As the use of audio increases in various fields such as communication, entertainment, and security, it is becoming more important to evaluate accurately its quality and apply. In particular, the technology to quickly determine the high and low quality of audio can contribute to classify illegal recording data in the field of audio forensics. In this paper, we propose a method of determining the audio quality using a modified DenseNet-based audio quality classification model. When audio in units of 3 seconds was input, the 2D feature of the audio was extracted using Mel-Frequency Cepstral Coefficients(MFCC) and applied to the modified DenseNet model to determine audio quality. Experiments were performed using domestic and foreign datasets and audio collected in-house. Although the accuracy of the CNN model was 86.7%, the proposed modified DenseNet model achieved the accuracy of 94%.

Keywords

audio quality classification, MFCC, CNN, modified DenseNet, forensics

* 금오공과대학교 컴퓨터소프트웨어공학과 학사과정
- ORCID¹: <https://orcid.org/0009-0003-8310-5034>
- ORCID²: <https://orcid.org/0009-0003-3839-7242>
** 금오공과대학교 컴퓨터소프트웨어공학과 교수(교신저자)
- ORCID: <http://orcid.org/0000-0002-6081-1492>

· Received: Oct. 04, 2023, Revised: Oct. 13, 2023, Accepted: Oct. 16, 2023
· Corresponding Author: Hae-Yeoun Lee
Dept. of Computer Software Engineering, Kumoh National Institute of Technology, Korea
Tel.: +82-54-458-7548, Email: haeyeoun.lee@kumoh.ac.kr

I. 서 론

음성 기술의 급격한 발전과 스마트 기기의 보급으로 오디오는 우리 일상생활에서 중요한 역할을 하고 있다. 통신, 엔터테인먼트, 보안 등 다양한 분야에서 오디오의 활용이 증가함에 따라, 그 음질을 정확하게 평가하고 응용하는 것이 중요해지고 있다. 특히 오디오를 불법적으로 사용하여 사회적 문제가 빈번히 발생하고 있어서 방대한 양의 오디오에서 고음질 및 저음질 여부를 판단하여 분류하는 것은 오디오 포렌식을 위해서 필요한 인적, 물적 및 시간적 자원을 줄여주기 때문에 필수적이다.

오디오의 음질을 판단하는 과정은 오디오 분석의 핵심 부분으로 음성 인식, 음악 분류, 보안 시스템 등 다양한 응용 분야에서 중요한 선행 과정으로 작용한다. 그러나, 오디오의 음질 판단은 주관적인 요소가 많아 신뢰성과 일관성에 어려움을 겪고 있다. 이를 해결하기 위해 다양한 딥러닝 기술을 활용한 연구가 진행되고 있다.

본 논문에서는 오디오의 MFCC 특징을 변형된 DenseNet 기반의 음질 분류 모델에 적용하여 오디오에 대하여 고음질 및 저음질을 높은 정확도로 판단하는 방법을 제안한다. 오디오의 2차원 신호 특징을 추출하기 위해 MFCC를 이용하였고, 추출된 특징을 DenseNet201을 변형한 음질 분류 모델을 설계하고, 적용하여 학습과 검증을 수행하였다. 또한, 학습 및 검증에 사용되지 않은 테스트 데이터를 사용해 성능을 평가하였고 CNN 방법과 성능을 비교하였다.

본 논문의 2절에서는 음질 분류와 관련한 관련 연구들을 제시하고, 3절에서는 오디오 신호에서 MFCC 특징을 추출하는 기법과 DenseNet을 변형한 음질 분류 모델의 구조 등을 자세하게 기술한다. 4절에서는 실험 결과를 제시하고, 5절에서는 결론 및 향후 연구 계획에 관해 기술한다.

II. 관련 연구

음성 인식, 음성 제어, 음성 부호화 등 분야에서 음성 활동 감지 연구는 핵심적인 역할을 담당하고

있다. 이러한 분야에서 음성 활동 감지는 정확한 오디오 신호 분석과 처리의 기반이 되며, 오디오 신호 내에서 실제 음성 부분을 식별하는 중요한 과정이다. 최근에는 딥러닝을 활용하여 음성 활동 감지 정확도를 높이기 위한 방법들이 제안되고 있다.

Y. Kang, J. Lee는 무작위 딥러닝을 이용하여 스펙트로그램과 시계열 데이터를 분류했다. 무작위 학습 방법으로 RVFL, deep RVFL, ROCKET을 선택하고 결과를 CNN, LSTM과 비교했다. CNN 스펙트로그램 분류 정확도는 평균 82.0%, ROCKET 분류 성능은 83.7%, LSTM 가변 길이 시계열 데이터의 분류 정확도는 81.8%, ROCKET 분류 성능은 85.7%였다[1].

X. Huang et al.은 차량 내부 소음의 음질 예측 및 개선을 위해 잡음 시간-주파수 이미지를 입력으로 사용하는 정규화된 심층 CNN 모델인 TF-CNN 모델을 개발했다. T-CNN 및 F-CNN 모델은 각각 93.5% 및 94.1% 정확도였으나, TF-CNN 모델은 97.0%의 정확도를 달성하여 두 모델보다 성능이 좋았다[2].

A. Abbaskhah et al.은 MFCC 특징을 추출하여 영아의 울음을 분류했다. 5개 클래스로 분류하도록 설계된 SVM, MLP, CNN 분류기를 사용하여 각각 성능을 평가했다. CNN 분류기가 최적의 조건에서 92.1%로 가장 높은 정확도를 보였다[3].

M. Borksy et al.은 다중 분류기를 사용하여 음성 음질 분류를 위한 다양한 특징 유형의 성능을 조사하였고, GMM, SVM, RF 및 DNN 분류기를 사용하여 전체 COVAREP 기능 벡터, 하위 집합 및 MFCC의 성능을 평가했다. COVAREP을 사용했을 때 정확도는 SVM, RF, GMM 및 DNN에 대해 각각 78.0%, 79.8%, 68.1% 및 77.0%를 보였고, MFCC를 사용했을 때 정확도는 SVM, RF, GMM 및 DNN에 대하여 각각 78.4%, 79.5%, 56.3% 및 76.7%에 도달했다[4].

J. Lee et al.은 MFCC 특징을 추출하여 MobileNet에 적용하여 오디오에서 사람 음성 구간을 추출하는 연구를 수행하였다. 세그먼트 단위로 93.9% 정확도, 전체 오디오에 대해 88.1% 정확도를 달성했다[5].

III. 제안하는 MFCC 특징을 활용한 변형된 DenseNet 기반의 음질 분류 방법

제안하는 방법에서는 3초 단위의 오디오로부터 MFCC를 추출하고 MFCC 스펙트로그램 이미지로 시각화한다. 이를 DenseNet201 기반의 변형된 음질 분류 모델의 입력으로 사용하여 오디오를 고음질과 저음질로 분류한다. 이러한 전체 과정을 그림 1에 도시하였다.

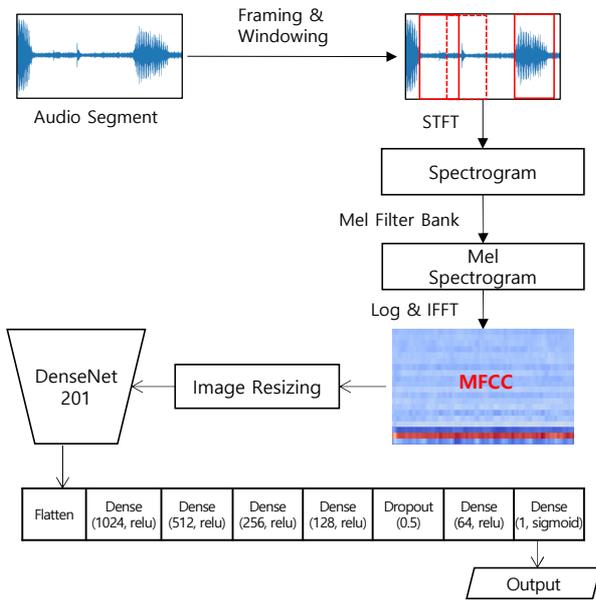


그림 1. 오디오 음질 분류 과정
Fig. 1. Audio quality classification workflow

3.1 MFCC 특징 추출

MFCC는 음성 및 오디오 신호 처리에서 널리 사용되는 특징 추출 기술이다. 본 연구에서는 주어진 오디오로부터 MFCC를 추출하고 이를 2차원 이미지로 시각화하여 변형된 DenseNet 기반의 음질 분류 모델을 통하여 음질 분류에 활용한다.

그림 1에 도시된 것과 같이 MFCC 특징 추출 과정은 다음과 같다. 먼저, 입력된 오디오 신호를 작은 프레임으로 나눈다. 그 후, 각 프레임에 대해 STFT(Short-Time Fourier Transform)를 수행하여 주파수 영역으로 신호를 변환한다. 다음으로, Mel 스케일의 필터 बैं크를 적용하여 주파수 영역을 Mel 주파수 스케일로 변환하며, 이후에는 선형 스케일에서 로그 스케일로 변화한다. 그 다음에 주파수 영역에서 적

절한 계수를 추출하기 위해 DCT(Discrete Cosine Transform)를 적용하며, DCT 결과로부터 일부 계수를 선택하여 MFCC 특징 벡터를 생성한다[6].

이와 같은 MFCC 추출 과정을 통해 오디오 세그먼트는 변형된 DenseNet 모델의 입력 데이터가 되고, 학습, 검증, 평가 데이터셋 모두에 MFCC 추출 과정을 적용하였다.

3.2 변형된 DenseNet 기반의 음질 분류 모델

본 연구에서는 음질 분류를 위해 DenseNet201 아키텍처를 활용하여 그림 2과 같이 변형한 딥러닝 모델을 구축하였다.

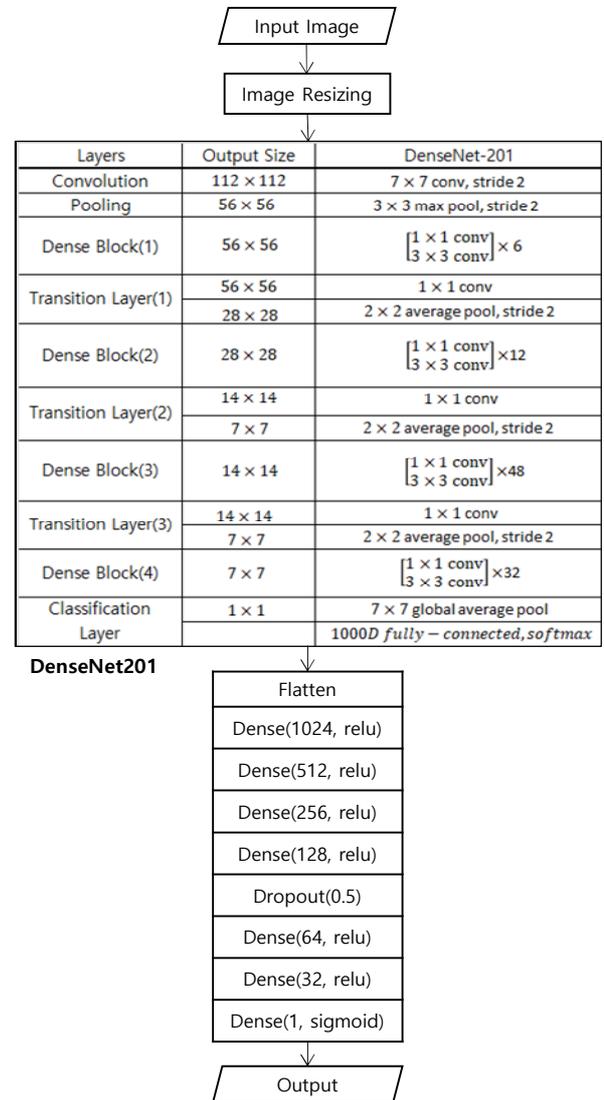


그림 2. 변형된 DenseNet201 음성 분류 모델 구조
Fig. 2. Structure of modified DenseNet201 audio classification model

DenseNet201은 심층 CNN의 한 종류로서, 201개의 계층으로 구성되어 있다. Dense Convolutional Network라는 개념을 기반으로 하며, Convolution Layer, Pooling Layer, Transition Layer, Dense Block의 조합으로 이루어져 있다.

Convolution 계층은 입력 이미지에서 특징을 추출한다. 이미지에서 작은 부분을 슬라이딩하면서 각 지역의 특징을 감지한다. Pooling 계층은 특징 맵의 크기를 줄인다. 이는 계산 비용을 줄이고, 모델의 파라미터 수를 줄이는 데 도움이 되며, 주로 Max Pooling을 사용하여 각 영역에서 가장 큰 값을 선택하여 정보를 압축한다. Dense 블록은 여러 Convolution 계층과 병목 계층으로 구성된다. 밀집한 연결로 인해 그래디언트 소실 문제가 감소하고, 모델이 더욱 풍부한 기능을 학습할 수 있다. Dense 블록 내에서 특징 추출 및 변환이 이루어진다. Transition 계층은 특징 맵의 차원을 줄인다. 이전 밀집 블록의 출력을 받아 차원을 줄인 다음, 다음 밀집 블록으로 전달한다. Classification 계층은 분류 작업을 수행하며, 주어진 입력 이미지가 어떤 클래스에 속하는지를 예측한다[7].

음질 분류를 위해 설계한 변형된 DenseNet 모델에서 입력 이미지의 크기는 224×224 로 설정하였고, 과적합 방지를 위해 모델에 50% 드롭아웃 비율을 가지는 무작위 드롭아웃 계층을 추가하였다. 기존 DenseNet201 모델에 512, 256, 128, 64, 32개의 뉴런을 가진 다섯 개의 완전 연결 계층을 추가하였다. 각 계층을 ReLU(Rectified Linear Unit) 활성화 함수를 통해 비선형성을 부여하였으며, 이를 통해 고차원 특징의 추상화와 표현 능력을 강화하였다. 최종 출력 계층은 하나의 뉴런을 갖고 시그모이드 활성화 함수를 적용하여 이진 분류를 수행한다.

IV. 실험 결과 및 분석

실험은 11th Gen Intel(R) Core(TM) i9-11900 @ 2.50GHz CPU 및 64GB RAM, NVidia GeForce RTX 3090 GPU 환경에서 진행하였다. 모델 학습 파라미터로 학습률은 $10e-4$ 로 설정하였다. 손실 함수는 이진 크로스 엔트로피를 사용하고, 최적화 함수는

Adam Optimizer를 사용하였다. 학습은 최대 100 epoch로 진행했다. 제안하는 변형된 DenseNet 기반 음질 분류 방법과 범용적인 CNN 기반 방법 모두 동일한 데이터셋을 학습, 검증, 평가에 사용하였다.

4.1 실험 데이터셋

실험을 위한 데이터셋은 Google Audio Set, Korean Conversational Speech Corpus[8], OpenSLR Seoul Corpus[9], AI Hub 한국인 대화 음성[10], IRMAS Dataset[11], GTZAN Dataset[12], Urban Sound Dataset[13], ESC Dataset[14] 등 공개 데이터셋과 YouTube 영상, 스마트폰, 녹음기 등으로 직접 수집한 데이터를 사용해 구축하였다. 고음질(HQ) 데이터는 영화, 음악, 동물 소리, 악기 소리, 자연 소리, 사물 소리, TV 소리, Youtube 음성 등이 포함되었고, 저음질(LQ) 데이터는 직접 수집한 사용자 음성 녹음, 전화 통화 녹음 등이 포함되었다.

실험에서 사용한 학습 데이터 및 검증 데이터, 평가 데이터의 비율은 0.8:0.1:0.1로 표 1에 데이터 수를 나타내었다. 전체 데이터 28,000개 중 25,200개의 데이터를 학습 및 검증에 사용하였고, 2,800개의 학습 및 검증에 사용되지 않은 데이터를 독립적인 평가에 사용하였다.

표 1. 실험에 사용된 데이터 수

Table 1. Number of data used for experiments

Class	Training	Validation	Testing	Total
HQ	11,200	1,400	1,400	14,000
LQ	11,200	1,400	1,400	14,000

4.2 제안하는 음질 분류 방법의 학습

제안하는 변형된 DenseNet 기반의 음질 분류 방법은 Tensorflow와 Keras를 사용하여 DenseNet-201 아키텍처를 기반으로 구현하였다.

그림 3에는 제안하는 방법에 대해 표 1에서 설명한 학습 및 검증 데이터를 사용해 훈련하는 과정에서 학습 정확도 및 손실 경향을 epoch 단위로 나타낸 그래프를 도시하였다.

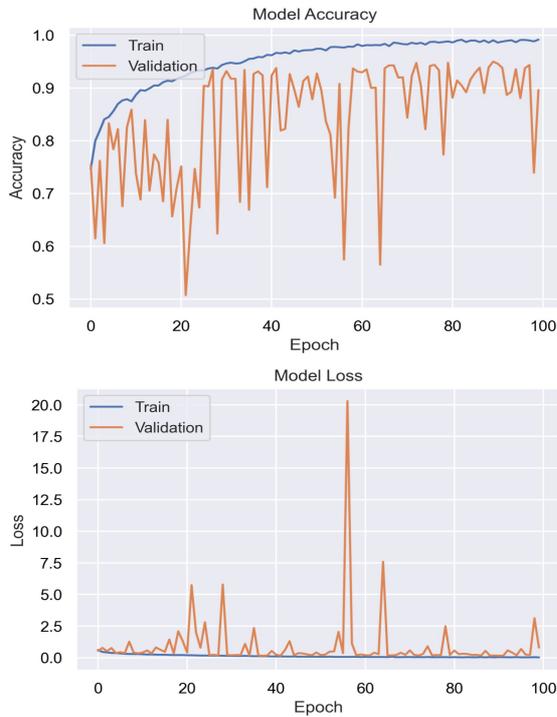


그림 3. 제안하는 방법의 학습 정확도 및 손실 경향
Fig. 3. Learning accuracy and loss trends of the proposed method

제안하는 방법은 초기 epoch부터 안정적으로 학습되었으며, 학습 손실은 epoch 진행에 따라 점진적으로 감소하는 양상을 보였다. 그러나, 일부 epoch에서 검증 정확도가 급격히 낮게 나타났고, 특히 57 epoch에서 검증 손실이 비정상적으로 높게 나타났다. 이러한 현상의 원인으로서는 과적합, 학습률 설정, 데이터 불균형, 모델의 복잡성 등 여러 가지 요소가 고려된다. 학습률 조정, 데이터 재샘플링, 모델 구조 간소화와 같은 기술적인 조치를 통해 이 문제를 해결할 수 있을 것으로 예상된다.

검증 데이터를 기준으로 훈련 과정에서의 경향을 분석해 변형된 DenseNet 모델에서 정확도가 제일 높은 최적 epoch에서 성능을 표 2에 나타내었다. 90 epoch에서 학습 정확도 99.0%, 검증 정확도 95%로 제일 높은 성능을 보였다.

표 2. 제안하는 방법의 최적 epoch에서의 성능
Table 2. Performance at optimal epoch of the proposed method

Epoch #90	Accuracy	Loss
Train	99	0.03
Validation	95	0.18

4.3 CNN 기반의 음질 분류 방법과 학습

제안하는 방법과 성능 비교를 위하여 범용적인 CNN 모델 기반의 음질 분류 방법을 수립하였고, 제안하는 방법과 동일 데이터를 사용하여 학습과 검증을 수행하였다. CNN 모델은 그림 4와 같이 6개의 합성곱 계층과 Pooling 계층을 순차적으로 거쳐서, 배치 정규화를 통한 안정화 및 Flatten 계층을 통해 2차원 이미지 데이터를 1차원 벡터로 변환한다. 이어서 완전 연결 계층(Fully-connected Layer)을 통해 고차원 특성을 학습하고, Dropout 계층을 활용하여 모델을 정규화하여 과적합을 예방한다.

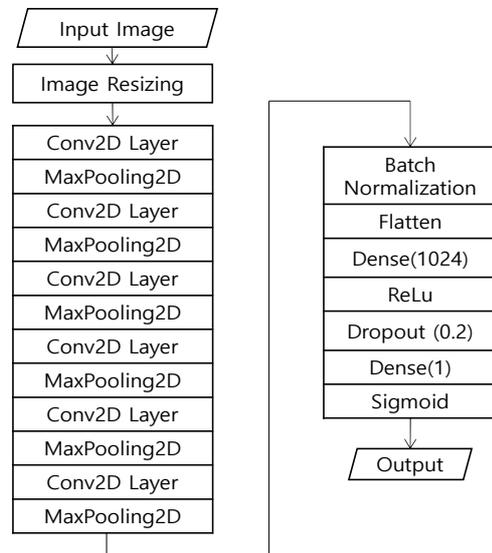


그림 4. CNN 모델 구조
Fig. 4. Structure of CNN model

그림 5에는 CNN 기반의 음질 분류 방법에 대해 학습 및 검증 데이터를 사용해 훈련하는 과정에서 학습 정확도 및 손실 경향을 epoch 단위로 나타낸 그래프를 도시하였다.

CNN 방법의 학습 과정은 초기 epoch부터 안정적으로 진행되었으며, epoch이 증가함에 따라 학습 정확도가 향상되는 경향을 보였다. 그러나, 몇몇 epoch에서 검증 정확도와 검증 손실의 변동이 관찰되었다. 이러한 변동은 주로 과적합, 데이터 불균형, 모델 복잡성 문제 등으로 해석할 수 있으며, 학습률이나 드롭아웃 조정 등으로 이러한 문제를 해결할 수 있을 것으로 보인다.

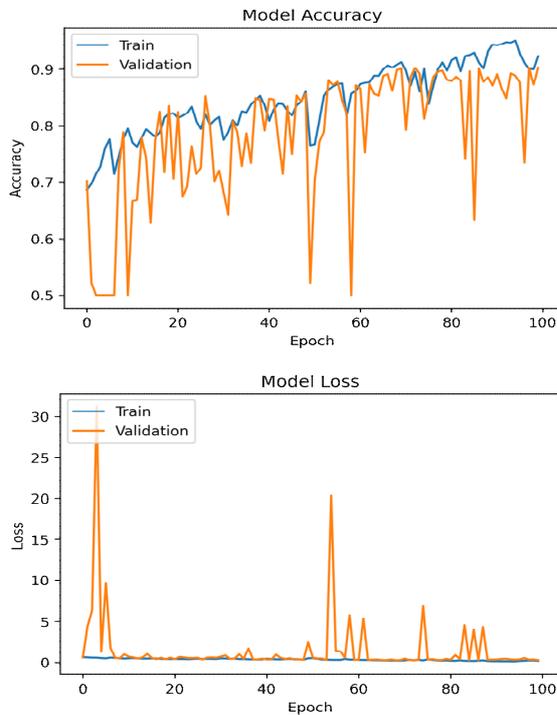


그림 5. CNN 방법의 학습 정확도 및 손실 경향
Fig. 5. Learning accuracy and loss trends of CNN method

검증 데이터를 기준으로 훈련 과정에서의 경향을 분석해 CNN 모델에서 정확도가 제일 높은 최적 epoch에서의 성능을 표 3에 나타내었다. 100 Epoch에서 학습 정확도 92.3%. 검증 정확도 90.3%로 제일 높은 성능을 보였다.

표 3. CNN 방법의 최적 epoch에서의 성능
Table 3. Performance at optimal epoch of CNN method

Epoch #100	Accuracy	Loss
Train	92.3	0.19
Validation	90.3	0.27

4.4 제안하는 음질 분류 방법과 CNN 기반 음질 분류 방법의 성능 비교

표 1에서 설명한 학습과 검증에 사용하지 않은 평가 데이터를 사용하여 제안하는 방법과 범용적인 CNN 방법에 대한 성능을 평가하였다.

그림 6 및 그림 7에 제안하는 방법과 CNN 방법의 평가 데이터에 대한 정확도 분석 결과를 혼동행렬로 나타내었다. 제안하는 방법은 94.0% 정확도, CNN 방법은 86.7% 정확도를 보여서, 제안하는 방

법의 성능이 더 좋음을 알 수 있다. 또한, 두 방법 모두 저음질(LQ) 오디오를 고음질(HQ) 오디오로 분류하는 오류가 HQ 오디오를 LQ 오디오로 분류하는 오류보다 큰 것을 알 수 있다.

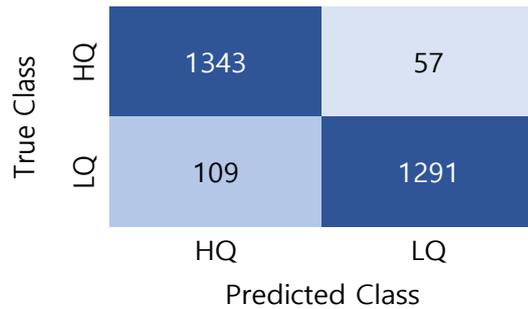


그림 6. 제안하는 방법의 혼동행렬
Fig. 6. Confusion matrix of the proposed method

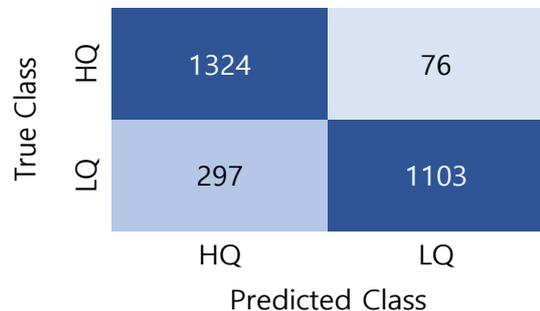


그림 7. CNN 방법의 혼동행렬
Fig. 7. Confusion matrix of CNN method

제안하는 방법에서 HQ가 LQ로 판단된 57개 데이터를 분석해본 결과, 끊기는 소리, 울리는 소리, 무음 또는 작은 소리 같은 판단하기 힘든 소리가 포함되어 있었다. LQ이 HQ로 판단된 297개 데이터를 분석해본 결과, 모두 직접 녹음한 음성 데이터 및 통화 녹음 데이터였다. 주로 음낮이의 변화가 크게 없거나 말소리가 크게 잘 들리는 음성 데이터였기 때문에, 고음질로 판단되었다고 추정된다.

V. 결론 및 향후 연구 방향

본 논문에서는 공개된 데이터셋과 자체적으로 구축한 데이터셋을 사용해 오디오에 대해서 MFCC 특징을 추출한 후에 변형된 DenseNet 음질 분류 모델에 적용하여 고음질 또는 저음질로 분류하는 방법을 제안하였다.

또한 실험을 통해서 제안하는 방법이 오디오의 음질을 정확하게 분류할 수 있는 높은 성능을 갖고 있음을 보였다.

이와 같은 제안하는 방법을 이용하여 오디오의 정확한 음질 평가를 통해, 오디오 포렌식 분야에서 불법적인 녹취 데이터를 판단하는 것에 활용할 수 있다.

향후 연구 방향으로는 본 연구는 포렌식 관점에서 음성에 중요성을 두고 있기에 오디오 내에서 사람의 음성에 초점을 두어서 알고리즘을 개발하고 성능을 향상할 계획이다. 또한, MobileNet과 같은 CNN 변형 모델[15]이나 음성 시계열 특성을 활용한 LSTM-GAN 모델[16]과 같은 딥러닝 모델의 적용을 통하여 성능 개선을 모색할 예정이다. 이러한 개선을 통해 음성 구간의 정확한 분류와 음질 평가를 통하여 포렌식 분야에 활용성을 더욱 향상시킬 수 있을 것으로 기대한다.

References

- [1] Y. Kang and J. Lee, "Randomized learning-based classification of sound quality using spectrogram image and time-series data: A practical perspective", *Engineering Applications of Artificial Intelligence*, Vol. 120, pp. 105867, Apr. 2023. <https://doi.org/10.1016/j.engappai.2023.105867>.
- [2] X. Huang, H. Huang, J. Wu, M. Yang, and W. Ding, "Sound quality prediction and improving of vehicle interior noise based on deep convolutional neural networks", *Expert Systems with Applications*, Vol. 160, pp. 113657, Dec. 2020. <https://doi.org/10.1016/j.eswa.2020.113657>.
- [3] A. Abbaskhah, H. Sedighi, and H. Marvi, "Infant cry classification by MFCC feature extraction with MLP and CNN structures", *Biomedical Signal Processing and Control*, Vol 86, No. Part B, pp. 105261, Sep. 2023. <https://doi.org/10.1016/j.bspc.2023.105261>.
- [4] M. Borsky, D. D. Mehta, J. H. V. Stan, and J. Gudnason, "Modal and non-modal voice quality classification using acoustic and electroglottographic features", *IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 25, No. 12, pp. 2281-2291, Dec. 2017. <https://doi.org/10.1109/taslp.2017.2759002>.
- [5] J. Lee, H. Han, Y. Choi, and H. Lee, "Voice Activity Segment Audio Deduction Method using MobileNet", *Journal of KIIT*, Vol. 20, No. 3, pp. 1-8, Mar. 2022. <https://doi.org/10.14801/jkiit.2022.20.3.1>.
- [6] S. Lee, C. Lee, S. Jeong, and J. Kim, "Method of Speech Feature Parameter Extraction Using Modified-MFCC", *Proc. of the IEEK Conference*, pp. 269-272, Jun. 2001.
- [7] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks", *Proc. of the IEEE conference on computer vision and pattern recognition*, Honolulu Hawaii, pp. 4700-4708, Jul. 2017. <https://doi.org/10.48550/arXiv.1608.06993>.
- [8] Magic Data "ASR-KCSC: a Korean Conversational Speech Corpus", <https://magichub.com/datasets/korean-conversational-speech-corpus/> [accessed: Oct. 13, 2023]
- [9] W. Yun, K. Yoon, S. Park, J. Lee, S. Cho, D. Kang, and J. Kim, "The Korean Corpus of Spontaneous Speech", *Phonetics and Speech Sciences*, Vol. 7, No. 2, pp. 103-109, Jul. 2015. <https://doi.org/10.13064/KSSS.2015.7.2.103>.
- [10] AI Hub, "Korean Conversation Voice", <https://aihub.or.kr/aidata/7968> [accessed: Oct. 13, 2023]
- [11] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, "A Comparison of Sound Segregation Techniques for Predominant Instrument Recognition in Musical Audio Signals", *Proc. of the ISMIR*, pp. 559-564, Oct. 2012. <https://doi.org/10.5281/zenodo.1416076>.
- [12] Kaggle, "GTZAN Dataset - Music Genre Classification", <https://www.kaggle.com/datasets/>

andradaolteanu/gtzan-dataset-music-genre-classification
[accessed: Oct. 13, 2023]

- [13] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research", Proc. of the ACM MM, New York, United States, pp. 1041-1044, Nov. 2014. <https://doi.org/10.1145/2647868.2655045>.
- [14] K. J. Piczak, "ESC: dataset for environmental sound classification", Proc. of the 23rd ACM MM, New York, United States, pp. 1015-1018, Oct. 2015. <https://doi.org/10.1145/2733373.2806390>.
- [15] G.-J. Park, S.-H. Choi, and K.-S. Kim, "Implementation of Urinalysis Service Application based on MobileNetV3", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 23, No. 4, pp. 41-46, Aug. 2023. <https://doi.org/10.7236/JIIBC.2023.23.4.41>.
- [16] H. Jeong and H.-J. Kim, "A Dynamic Correction Technique of Time-Series Data using Anomaly Detection Model based on LSTM-GAN", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 23, No. 2, pp. 103-111, Apr. 2023. <https://doi.org/10.7236/JIIBC.2023.23.2.103>.

저자소개

서 영 은 (Young-Eun Seo)



2021년 3월 ~ 현재 :
금오공과대학교
컴퓨터소프트웨어공학과
학사과정
관심분야 : 컴퓨터비전, 딥러닝

안 상 근 (Sang-Geun Ahn)



2018년 3월 ~ 현재 :
금오공과대학교
컴퓨터소프트웨어공학과
학사과정
관심분야 : 컴퓨터보안, 영상처리,
딥러닝

이 해 연 (Hae-Yeoun Lee)



1997년 : 성균관대학교 정보공학과
(학사)
1999년 : KAIST 전산학과
(공학석사)
2006년 : KAIST 전자전산학과
(공학박사)
2008년 ~ 현재 : 금오공과대학교

컴퓨터소프트웨어공학과 교수

관심분야 : Digital Forensics, Image Processing,
Computer Vision, Internet of Things