

< 응용 논문 >

음향신호를 이용한 모터 기어 박스 모듈의 CNN기반 2단계 불량 검출 기법

최 건 영¹⁾ · 장 일 식²⁾ · 이 영 화²⁾ · 강 현 석³⁾ · 박 구 만⁴⁾

에스엠알 오토모티브 모듈 코리아 선행개발실¹⁾ · 서울과학기술대학교 나노 IT융합대학원 정보통신미디어공학 전공²⁾ · 서울과학기술대학교 일반대학원 IT미디어공학과³⁾ · 서울과학기술대학교 전자IT미디어공학과⁴⁾

Fault Detection of Motor Gear Box Using Two Stage Sound Classification Network

Geonyoung Choi^{*1)} · Il-Sik Chang²⁾ · Younghwa Lee²⁾ · Hyunseok Kang³⁾ · Gooman Park⁴⁾

¹⁾Advanced R&D Department, SMR Automotive Modules Korea Ltd, 116 Jomaru-ro 427beon-gil, Bucheon-si, Gyeonggi 14556, Korea

²⁾Department of Information Technology and Media Engineering, The Graduate School of Nano Design Fusion, Seoul National University of Science and Technology, Seoul 01811, Korea

³⁾Department of Electronic IT Media Engineering, Seoul National University of Science and Technology, Seoul 01811, Korea

⁴⁾Department of IT and Media Engineering, Seoul National University of Science and Technology, Seoul 01811, Korea

(Received 17 December 2021 / Revised 3 January 2022 / Accepted 10 January 2022)

Abstract : A motor gear box is used to control tilting of the mirror and folding of the side mirror wings in a vehicle. Gear box fault detection in the production line is very important because faulty production returns considerable cost when undertaking vehicle maintenance. Fault detection based on acoustic sounds is widely used because it is simple and efficient. Sound-based method offers the advantage of non-destructive inspection, but it also provides limited classification performance. In this article, we propose a two-stage classification algorithm based on CNN. This method detects anomaly in the first stage, then subsequently classifies the faulty class. Mel-spectrogram is used to extract the features of the acoustic motor sounds. Since the proposed method classified the types of defects after determining faulty components, it is expected not only to improve accuracy, but also to reduce the time required to figure out the cause of failure.

Key words : Fault detection(불량진단), CNN(합성곱 신경망), Audio classification(오디오 분류), Mel-spectrogram(멜 스펙트로그램), Motor gear box(모터 기어 박스)

1. 서론

소형 모터 기어 박스는 자동차 여러 부분에 액추에이터로 널리 사용되는 부품이다. 특히 소형 기어박스 모듈은 사이드 미러에서 미러 기울기 제어 및 미러 윙 폴딩 매커니즘 기능을 수행하는 필수 요소이다. 모터 기어박스 같은 작은 부품의 고장 결함은 차량 유지 보수에 의한 비용으로 생산자에게 큰 부담으로 돌아오기 때문에 생산 라인에서의 모듈 단위 불량 검출은 중요하다. 또한 모빌리티에 사용하는 모든 부품은 인간의 생명과 안전에 직결되어 있기 때문에 부품 불량을 정확히 판별하는 것은 완벽한 완성차의 신뢰성 확보를 위한 중요한 단계이다. 모터의 고장을 파악할 수 있는 진단 기법에 대한

다양한 시도와 연구가 이루어지고 있다. 근접 거리에서 음향 센서를 사용, 정상 동작 신호와 인위적으로 고장을 만들어 작동시킨 비정상 신호를 취득하여 RNN(Recurrent Neural Network)신경망을 이용한 고장진단방식¹⁾ 모터의 전류 신호에서 파악한 특정 신호를 KNN(K-Nearest Neighbor)알고리즘에 적용한 고장 진단 방식²⁾ 등 다수의 연구들이 진행되고 있다. 이 중에서 음향신호를 기반으로 하는 불량 판별 기법은 구조를 파괴하지 않고 결함을 판별할 수 있다는 장점이 있으며, 간단하고 효율적이기 때문에 쉽게 사용되고 있다.

현재 제조공정에서는 소음차단 부스내에서 전류, 진동, 청음 테스트를 하면서 동시에 그 자리에서 불량 분

*Corresponding author, E-mail: geonyoung.choi@smr-automotive.com

[†]This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

류를 하는 검수과정을 거치고 있다. 하지만 사람이 직접 판별하는 품질관리에는 다음과 같은 문제점이 예상된다. 작업자의 단순작업에 의한 피로도가 높아져 불량 판별의 실수가 발생할 수 있으며, 작업자의 노하우나 실력 혹은 불량의 유형에 따라 작업 품질 수준에 편차가 생길 수 있다.³⁾ 또한 정확하게 분류를 했다 하더라도 불량을 정상으로, 정상을 불량으로 오분류할 수 있는 확률도 배제할 수 없을뿐더러 1인의 검사자가 아닌 다수의 검사자가 검사작업을 행할 시 그 판별 기준이 검사자마다 다를 수 있으므로 판별의 신뢰성과 작업의 효율면에서도 문제가 발생할 수 있다.⁴⁾ 따라서 본 논문에서는 작업자가 직접 불량을 검수하는 시스템에서 발생할 수 있는 문제점들을 방지하고, 불량 판별의 균일한 결과를 얻기 위해 음향 신호를 기반으로 하는 딥러닝 기반의 불량 검출 기법을 제안한다.

제조 공정에서는 불량이 발생할 확률이 낮기 때문에 실제 생산라인에서 불량 데이터를 구하는 것은 굉장히 어려운 일이다. 따라서 제조 분야에서는 이상치 탐지(Anomaly detection)라는 기법을 이용하여 정상과 불량을 판정하기도 한다.⁵⁾ 정상 데이터와 비정상 데이터가 충분한 경우, 지도 학습을 통해 성능이 뛰어난 모델을 설계하는 것이 가능하지만, 제조 제품의 데이터 특성상 정상 데이터 수와 비교하여 비정상 데이터의 수가 현저히 적고, 비정상 데이터의 유형도 다양하여 특정 짓기 어렵다.⁶⁾ 이러한 이유로 제조 현장에서는 비교적 수집이 쉬운 정상 데이터 만으로 학습을 시키고 정상 제품의 특징만을 추출하여 재구성하는 이상치 탐지를 이용한 시스템을 일반적으로 사용하고 있다. 본 연구에서는 실험을 위해 정상 모터와 불량 모터를 수집하여 제한된 실험 환경에서 동작음을 취득하였으며, 훈련을 위한 데이터를 확보하였다. 따라서 일반 제조공정에서 불량진단에 사용하는 이상치 탐지 알고리즘보다 신뢰성을 높일 수 있는 지도학습 알고리즘인 CNN(Convolution Neural Network)기반 불량 검출 알고리즘을 제안한다.

이미지 분류에 적합한 CNN모델에 오디오 데이터를 사용하기 위해서는 적합한 피처를 선정해야 하는데 오디오 분류모델에서 쓰이고 있는 대표적인 피처인 멜스펙트로그램(Mel-spectrogram)과 MFCC(Mel-frequency cepstral coefficients)를 비교 실험하여 정확도가 높은 멜스펙트로그램을 피처로 사용하였다. 그리고 모델의 과대적합을 방지하고 성능을 향상시키기 위해 진폭(Amplitude)의 변형을 이용한 데이터 증강을 이용하여 데이터 양을 원본 데이터의 5배 증가하여 실험하였으며, 제조 제품의 데이터 특성상 발생하는 정상 데이터와 불량 데이터와의 클래스 불균형 문제의 해결을 위해 데이터 균형화 알

고리즘 중 하나인 언더 샘플링 기법⁷⁾을 사용하였다. 언더 샘플링이란 다수 클래스의 데이터를 소수 데이터 수에 맞도록 다수 클래스의 범위 내에서 데이터를 샘플링하는 방법을 말하며 본 논문에서는 매 에포크 마다 적용하여 클래스의 밸런스를 맞추어 주었다.

오디오 분류에 관한 연구는 국내외로 많은 연구가 이루어지고 있지만, ESC-50,⁸⁾ 또는 UrbanSound8K⁹⁾와 같은 클래스 차이가 분명한 데이터 셋을 사용한 연구가 많다. 하지만 본 연구에서는 기존 연구들의 데이터 유형과는 달리 클래스 차이가 뚜렷하지 않은 모터 기어 박스의 동작음 데이터 셋을 사용한다. 정상과 불량 of 패턴 차이가 아주 미세하거나 불량의 종류가 혼재되어 있는 경우도 있다. 뿐만 아니라 사람이 직접 청각에 의존하여 라벨링을 하였기 때문에 그라운드 트루스(Ground truth)가 확실히 명확하다고 하기에 한계가 있다. 따라서 이러한 데이터로 학습을 한 알고리즘은 하나의 클래스로 예측이 어려울 수 있으며, 기존 연구에 사용되던 데이터셋의 결과와도 차이가 존재한다는 점을 명시한다. 하지만 본 연구는 데이터셋이 가지고 있는 특성을 고려하여 제조 공정에서의 불량 검출 알고리즘 설계에 있어 인공지능 도입의 실현 가능성을 타진할 수 있는 유의미한 결과를 도출하고자 한다.

본 논문에서는 분류의 정확성을 높이기 위해 CNN기반 2단계 불량분류 알고리즘(2-stage fault classification)을 제안한다. 첫번째 단계에서는 정상과 불량을 위한 이진분류(Binary classification)를 수행하고, 두번째 단계에서는 불량 수리에 필요한 원인 분석을 위해 불량을 유형별로 분류하는 다중분류(Multiclass classification)를 진행한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에 사용한 데이터의 특성을 설명하고 데이터의 수집방법과 전처리 과정에 대해 기술한다. 3장에서는 제안한 모델의 구조와 재현 방법에 대해 설명한 후, 4장에서는 각 단계별 실험과 결과 해석을 통해 본 연구의 타당성을 검증한다. 마지막으로 5장에서는 본 연구에 대한 결론을 맺고 향후 연구에 대해 논한다.

2. 데이터 수집 및 전처리

2.1 데이터 특성 및 분류

본 논문에서는 기어 박스의 여러 불량 유형 중 동작음으로 구분이 가능한 3가지 유형의 불량 동작음과 정상 동작음을 포함한 4가지 클래스를 선정하였다. 실험에 사용된 모터 기어 박스는 실제 생산라인에 종사하는 숙련된 검사자가 동작음을 직접 귀로 청취하여 불량 여부를 판별하였으며, Table 1과 같이 불량의 유형에 따라 불

Table 1 Fault number by noise type

Fault No.	Types of noise
1	Shaft mis-align noise
2	Tooth scratch noise
7	Shaft thrust noise

량 식별 번호를 부여하였다. 따라서 앞으로는 각각의 불량 동작음은 불량유형번호로 명명하겠다.

Fig. 1은 정상 동작음과 3가지 클래스의 불량 동작음을 나타낸 웨이브 폼 그래프이다. Fault 1로 분류한 Shaft Mis-align 노이즈는 기어의 회전축이 정밀하게 정렬되어 있지 않아서 발생하는 불량 회전 마찰음이다. ‘윙윙윙’하는 울렁거리는 소리가 나며 Fig. 1(b)와 같이 정상 모터의 동작음 Fig. 1(a)와 거의 차이가 나지 않는다. Fault 2로 분류한 Tooth scratch 노이즈는 톱니기어 사면에 생긴 스크래치에 의해 회전 시 발생하는 주기적인 불량 마찰음이다. 폴딩 또는 언폴딩시 한 번씩만 발생하기도 하며, 톱니기어 이빨 패임 혹은 파열에 의한 경우에는 폴딩과 언폴딩시 모두 발생한다. ‘딱딱딱’하는 주기적인 패턴이 특징이며 Fig. 1(c)와 같이 규칙적인 노이즈 형태가 나타난다. Fault 7로 분류한 Shaft thrust 노이즈는 모터의 샤프트가 모터 벽에 부딪치면서 발생하는 추력음이다. 폴딩 또는 언폴딩 시작 시 ‘딱’하는 단일 소음이 발생하며, Fig. 1(d)와 같은 웨이브 폼 패턴을 보인다. 정상 동작음과 Fault 1 유형의 동작음은 들어서도 구별이 어려울 뿐더러 웨이브 폼 그래프의 형태를 보고서도 쉽게 구분하기가 어렵다.

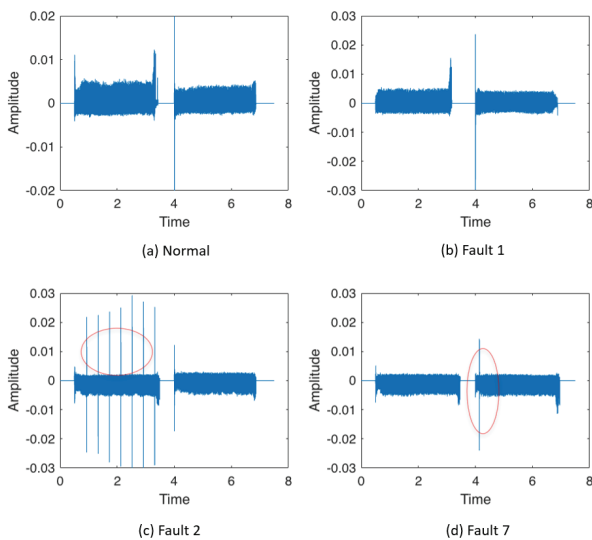


Fig. 1 Waveform graph of motor sounds

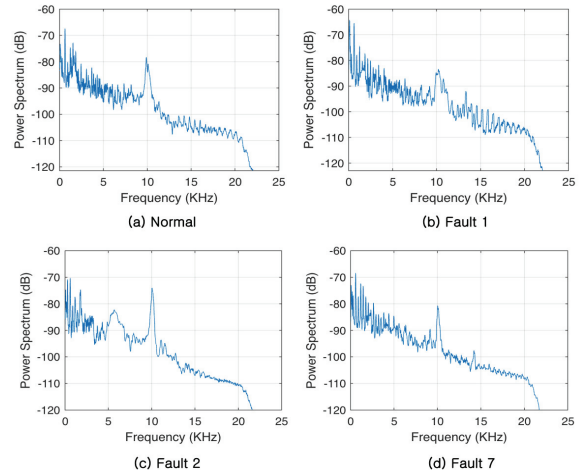


Fig. 2 Power spectrum graph of motor sounds

Fig. 2는 위의 4가지 유형의 모터 동작음에 대한 Power Spectrum 그래프이다. 이들 그래프의 패턴 형태를 보아서도 특히나 정상 동작음과 Fault 1 유형의 동작음은 구분하기에 뚜렷한 특징을 보여 주지 못한다.

2.2 데이터 수집

음향 데이터를 수집할 때는 주변 환경음과 돌발적으로 발생할 수 있는 소음 등을 고려해야한다. 외부 환경음은 동작음 분석에 방해 요인으로 작용할 수 있기 때문이다. 본 연구에서는 정확한 실험 데이터 수집을 위해 방음 설비가 갖춰진 녹음 부스 내에서 녹음을 진행하였다. Fig. 3에서 보는 바와 같이 기어박스와 약 10 cm 거리로 근접하여 마이크를 위치시키고 마이크를 오디오 인터페이스에 연결하여 기어박스 폴딩/언폴딩 동작음을 연속하여 녹음하는 데이터 취득 장치를 구성하였다. 오디오 인터페이스에서 출력되는 오디오신호는 자체 개발한 MATLAB 프로그램을 사용하여 웨이브형식으로 저장하였다. 음향 데이터는 16 bit, 모노형식, 표본주파수는 44.1 KHz로 취득하였다. 모터는 한 번 동작을 완료하는데 3초가 소요되며 접을 때와 펼 때, 그리고 모터 동작의 변환 시간을 포함하여 총 7초간 녹음하였다.

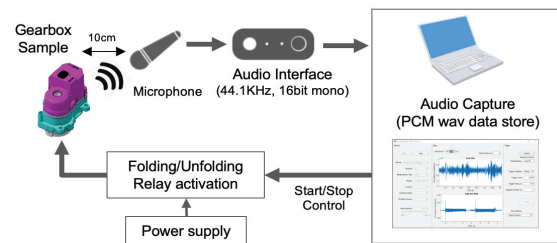


Fig. 3 Experimental apparatus

2.3 데이터 전처리

이미지 처리의 목적으로 개발된 CNN을 오디오 분류에 적용하기 위해서는 데이터 전처리 과정이 필요하다. 전처리 과정에서 추출된 오디오 특징은 CNN의 성능에 큰 영향을 줄 수 있다. 그러나 어떤 특징 추출이 최적의 방법인지에 대한 연구는 많지 않다.^{10,11)} 최근의 오디오 분류의 선행연구들을 살펴보면 MFCC, GFCC(Gammatone Frequency Cepstral Coefficients), CQT(Constant Q Transform) 등의 다양한 피처를 추출하고 학습에 사용하고 있음을 알 수 있다.^{12,13)} 본 실험에서는 분류모델에 적합한 피처를 선정하기 위해 오디오 분류 모델에서 많이 쓰이고 있는 멜 스펙트로그램(a)과 MFCC(b)를 비교하는 실험을 진행하였으며, 각각의 피처별 이미지는 Fig. 4와 같다.

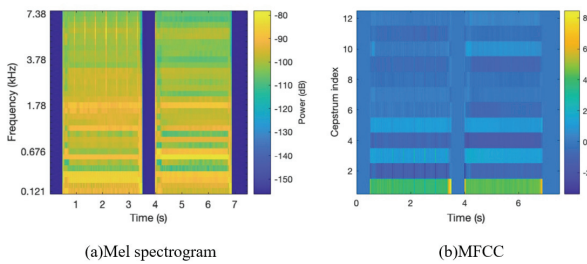


Fig. 4 Images by features

1차원의 음향 신호를 일정한 길이의 프레임으로 나눈 다음, 이 프레임에 대하여 프리에 변환을 적용하여 수평 축에는 시간정보를 나타내고 수직 축에는 주파수 성분의 크기를 데시벨 단위로 표현한 2차원 이미지를 스펙트로그램이라 하는데, 이 스펙트로그램에 멜 필터 뱅크를 적용한 추출 방식이 멜 스펙트로그램이다. 멜 스펙트로그램은 주파수 특성과 시간특성을 동시에 확인할 수 있기 때문에 일반적인 소리 데이터보다 더 효과적으로 분류에 사용될 수 있다. MFCC는 멜 스펙트로그램에 로그를 취하고 역푸리에 변환을 적용하여 주파수 도메인의 정보를 새로운 시간 축으로 바꾼 것을 말한다. 음향 및 음성인식에서 주로 사용되는 특징벡터 추출 방법으로 다른 음향 특징벡터 추출 알고리즘에 비해 연산량 대비 고성능의 특징벡터를 추출하는 것으로 알려져 있다.¹⁴⁾

분류기의 성능에 적합한 피처를 선별하기 위해 데이터 증강을 하지 않은 원본 데이터를 사용하여 피처별 성능을 비교하였다. Table 2는 본 실험에서 선정한 모델로

Table 2 Accuracy by features

	Mel-spectrogram	MFCC
Binary classification	0.765	0.668

실험한 정상 동작음과 3가지 클래스(Fault 1, 2, 7)로 구성된 불량 동작음과의 이진분류에 대한 피처별 정확도를 나타낸 결과이다. Table 2와 같이 정확도 측면에서 멜 스펙트로그램이 MFCC보다 우수한 결과를 나타냄을 확인하였고, 멜 스펙트로그램을 피처로 선정하였다.

2.4 데이터 증강

훈련에 필요한 데이터가 충분하지 않을 경우, 여기에 훈련된 신경망은 적은 양의 훈련데이터에 특화되어 일반화되지 못하고 실제 활용하기 어려워질 수 있다.¹⁵⁾ 따라서 부족한 데이터 개수를 보완하기 위해 데이터 증강 방법을 사용한다. 데이터 증강은 모델을 학습시키기 위해 수집한 데이터를 변형함으로써 다양한 형태의 데이터로 만들어 데이터의 양을 증가시키는 것을 말하며, 증강한 데이터를 이용한 신경망이 더욱 일반화되도록 도와주는 기법이다.¹⁵⁻¹⁷⁾ 영상처리분야에서는 영상의 좌우반전, 회전, 명암대비 등 다양한 데이터 증강 기법들이 시도되고 있으며 그에 따른 신경망의 성능 향상에 대한 연구들이 발표되고 있다.^{17,18)} 음향처리분야에서도 백색잡음 추가, 재생속도의 변환, 그리고 음향의 높낮이를 변환하는 방법 등 다양한 방법들이 활용되고 있다.^{12,15)} 본 실험에서는 학습에 사용되는 동작음의 특성을 고려하여 데이터 증강을 실시해도 데이터가 가진 고유 의미가 달라지지 않는 증강 방법으로 진폭을 변형하는 방법을 선택하였다. 오디오 데이터의 샘플 개수에 일정간격의 5개 점을 찍은 후 기존 진폭에 0.94~1.06의 수치를 랜덤하게 곱하여 곡선을 만들어 준 후 원본 웨이브와 곱해줌으로써 데이터를 증강하였다. 진폭의 변형을 위한 수치 0.94~1.06은 다수의 실험을 통해 청각으로 판단했을 때 최대한 원본 데이터가 손상되지 않는 범위의 값을 선정하였다.

이렇게 함으로써 원본 데이터보다 5배 많은 양의 데이터를 확보하였다. 정상 20,250개, Fault 1 3,000개, Fault 2 3,250개, Fault 7 2,750개로 데이터 증강하였으며, 정상 모터 데이터 수 총 20,250개, 3가지 유형의 불량 모터 데이터 수 총 9,000개로 이진 분류 실험에 사용하였다.

3. 신경망 설계

3.1 백본 네트워크의 비교 분석

먼저 이미지 분류에 뛰어난 성능을 나타내고 있는 백본 네트워크에 실험 데이터를 입력하여 정상과 3종의 불량(Fault 1, 2, 7)을 판별하는 이진 분류 테스트를 진행하였다. 프리 트레이닝 된 Inception,¹⁹⁾ Mobilenet,²⁰⁾ Resnet 50,²¹⁾ VGG 16²²⁾의 4개 모델을 테스트하였으며 결과는 Table 3과 같다. 증강된 데이터셋을 사용하였으며 총 5번의 실험을 통한 평균치를 나타내었다. 일반적인 이미지 분류

Table 3 Accuracy of backbone networks

Backbone networks	Accuracy
Inception	0.816
Mobile net	0.670
Resnet 50	0.715
VGG 16	0.515

에서는 뛰어난 성능을 보였던 네트워크이지만 실험 데이터셋에서는 좋은 결과를 보여주지 못했다. 클래스 구분이 뚜렷하지 않은 동작음이라는 데이터 특성이 오히려 깊고 복잡한 네트워크에서는 큰 성과를 보여줄 수 없다는 결과를 얻을 수 있었다. 따라서 간단하지만 데이터의 특성을 잘 추출하여 분류에 효율적인 망을 고안하게 되었고 CNN 기반의 어텐션 블록을 추가한 네트워크를 디자인하였다.

3.2 CNN 네트워크 구성

본 연구에서 사용한 신경망은 CNN 알고리즘이다. CNN은 합성곱 연산을 사용하여 신경망을 연산하는 것으로 이미지 분류 및 자연어 해석, 영상 분류 등의 다양한 분야에서 사용되고 있으며,²³⁾ CNN을 활용하여 음성을 분류하는 연구도 활발히 진행되고 있다.²⁴⁾ CNN신경망은 주어진 이미지의 일부분에서 특징을 추출하여 피쳐들의 패턴을 파악하는 구조로 이루어지며 인간이 직접 생성한 특징을 통해 예측을 하는 것보다 높은 정확도를 가지는 것으로 알려져 있다.^{25,26)} 본 논문에서는 Fig. 5와 같이 컨볼루션 레이어와 배치 정규화(Batch normalization), ReLu활성화 함수, 2x2 맥스 풀링, 어텐션 블록을 결합한 CNN블록을 하나의 단위로 설정하였다. 자연어 처리분야의 Seq2seq²⁷⁾모델에서 입력 시퀀스가 길어지면 출력 시퀀스의 정확도가 떨어지는 현상을 보정해 주기 위해 고안된 어텐션 매커니즘²⁸⁾을 CNN모델에 적용하였다. 어텐션이라는 것 자체가 어떠한 특성에 대해 집중하는 것을 의미하는데 어텐션 블록을 추가함으로써 이미지 분류의 예측과 연관이 있는 부분을 더 집중하여 학습하도록 하는 효과를 주었다. 어텐션 블록으로 SE(Squeeze and Excitation)²⁹⁾블록 또는 CBAM(Convolution Block

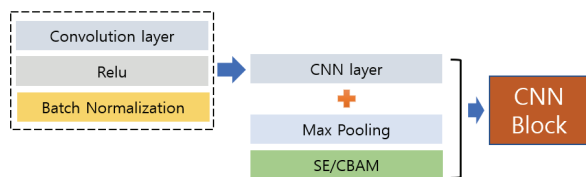


Fig. 5 CNN block

Attention Module)³⁰⁾을 사용하였으며 각각 추가한 것 과 추가하지 않은 것을 비교 분석함으로써 실험에 사용된 데이터셋에 적합한 모델을 찾고자 하였다. 또한 커널 사이즈가 3x3으로 제한된 컨볼루션 레이어는 적층 됨에 따라 비선형성이 증가하게 되는데 이러한 특성이 모델 성능에 어떤 영향을 미칠지 알아보고자 컨볼루션 레이어의 개수와 필터 사이즈를 달리한 4가지 모델로 다양화하여 실험을 진행하였다.

학습에 사용한 CNN블록이 포함된 CNN구조는 Fig. 6과 같다. 커널 사이즈가 3x3인 컨볼루션 연산을 적용해 충분한 정보를 필터링하고 보존할 수 있도록 하였다. 그리고 배치 정규화를 통해 데이터를 정제하여 모델의 과적합을 방지하였으며 ReLu활성화 함수 계층을 통해 비선형성을 추가하였다. 그 후 맥스 풀링 연산을 통해 피쳐 사이즈를 1/2로 줄였다. 맥스 풀링을 거쳐 추출된 피쳐들은 어텐션 블록을 통과한다.

FCL(Fully Connected Layer) 대신 GAP(Global Average Pooling)을 적용하여 파라미터의 증가를 막아주어 모델의 과대적합을 방지하였으며, 텐스 레이어를 거친 후 배치 정규화 과정, 드롭 아웃과 소프트맥스 연산을 통해 최종 분류 결과를 예측하게 된다. 이진 분류인 경우엔

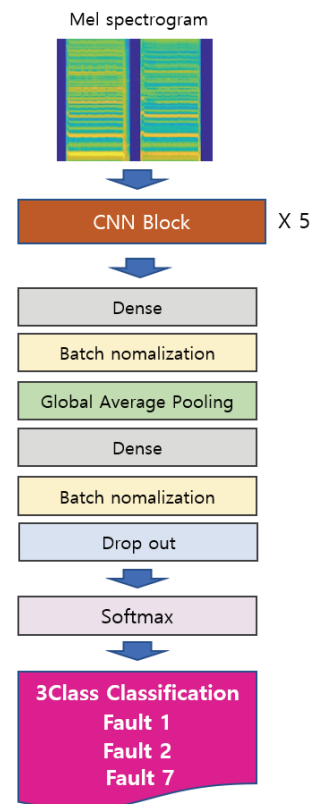


Fig. 6 CNN structure with SE block

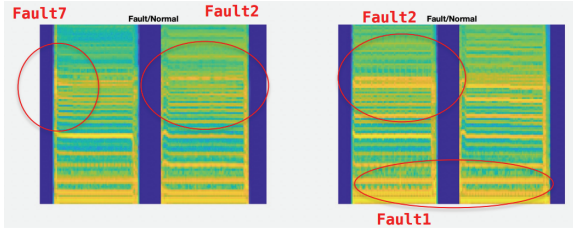


Fig. 7 Mel-spectrogram of mixed faults

Table 4 CNN setting

Input size	216 × 170
Optimizer	Adam optimizer
Batch-size	256
Early stopping patience	20
Epoch	300
Learning rate	0.01
Convolution kernel size	3X3
Max pooling kernel size	2X2
Dense layer	128
Label smoothing	0.1
Dropout rate	0.1

손실함수로 카테고리컬 크로스 엔트로피(Categorical cross entropy)를 사용하였으며, 다중분류인 경우엔 바이너리 크로스 엔트로피(Binary cross entropy)를 사용하였다. Fig. 7에서와 같이 불량 유형이 혼재 되어있거나 불량 유형이 뚜렷하지 않고 모호하게 나타나는 경우가 발생하므로 라벨 스무딩 수치를 0.1로 설정하였으며, 분류 모델에 사용한 그 밖의 수치들은 Table 4와 같다.

4. 실험 및 결과

4.1 이진분류 모델선정

정상과 3가지 클래스(Fault 1, 2, 7)로 구성된 불량과의 이진분류를 진행하였다. CNN블록의 컨볼루션 레이어 개수와 채널 수를 달리한 4가지 모델을 대상으로 어텐션 블록(SE, CBAM)을 추가했을 경우와 그렇지 않은 경우를 비교하여 실험하였다.

증강된 데이터셋을 사용하였으며 총 5번의 실험을 통한 평균치를 구하였다. Table 5와 같이 기본 CNN 네트워크보다 어텐션 블록을 추가했을 경우 정확도가 더 높게 나타났으며, 두 가지 어텐션 블록 중에 SE블록을 사용했을 때 정확도가 더 높은 것으로 나타났다. 또한 컨볼루션 레이어의 깊이가 오히려 얇은 망에서 더 효율적으로 작동하는 것을 확인할 수 있었다. 그리고 컨볼루션 레이어 개수와 채널 수를 달리한 4가지 모델 중 SE블록

Table 5 Accuracy comparison of models

Model	CNN block	Accuracy
No attention	Cov_16, 32, 64, 128, 256	0.683
	Cov_16x2, 32x2, 64x2, 128x2, 256x2	0.675
	Cov_32, 64, 128, 256, 256	0.745
	Cov_32x2, 64x2, 128x2, 256x2, 256x2	0.765
+ SE block	Cov_16, 32, 64, 128, 256	0.908
	Cov_16x2, 32x2, 64x2, 128x2, 256x2	0.841
	Cov_32, 64, 128, 256, 256	0.835
	Cov_32x2, 64x2, 128x2, 256x2, 256x2	0.85
+ CBAM block	Cov_16, 32, 64, 128, 256	0.673
	Cov_16x2, 32x2, 64x2, 128x2, 256x2	0.674
	Cov_32, 64, 128, 256, 256	0.673
	Cov_32x2, 64x2, 128x2, 256x2, 256x2	0.810

Table 6 Accuracy comparison of data sets

Model	Accuracy	
	Original data	Augmented data
Conv_16,32,64,128,256 + SE block	0.765	0.908

이 추가된 Conv_16,32,64,128,256 모델이 가장 성능이 좋은 것으로 나타났다. 따라서 이를 기준 모델로 선정하여 실험을 진행하였다.

선정된 모델을 기준으로 데이터 증강에 따른 성능변화를 알아보았다. Table 6과 같이 약 18.6 % 성능향상을 나타냈다.

4.2 4클래스 다중분류

정상과 3가지 불량 유형(Fault 1, 2, 7)으로 구성된 4가지 클래스의 다중분류를 진행하였다. 5차례의 실험결과를 평균하여 결과를 나타내었다.

Table 7과 같이 4 클래스의 다중 분류의 정확도는 0.608이 나타났다. 4.1절의 Table 4의 실험환경과 동일한

Table 7 Accuracy of 4-classes multiclass classification

Model	Accuracy
Cov_16,32, 64, 128, 256 + SE block	0.608

조건임에도 불구하고 정확도에 있어서 많은 차이를 보이는 것을 확인할 수 있었다.

Fig. 8은 파이썬 프로그램의 사이킷런 라이브러리의 T-SNE(T-Distributed Stochastic Neighbor Embedding)함수를 이용하여 실험 데이터 셋을 시각화한 그림이다. 복잡하고 높은 차원의 데이터를 2차원으로 축소하여 보여줌으로써 데이터의 분포를 이해하는데 도움을 준다. 붉은 색으로 표시된 점은 정상 데이터이며 나머지 녹색과 파란색, 노란색은 각각 Fault 1, 2, 7을 의미한다. 붉은 점으로 표시된 정상 데이터는 좌하단에 군집을 형성하여 비교적 분류가 용이하게 보인다.

Fig. 9는 4가지 다중 분류의 컨퓨전 매트릭스이다. 컨퓨전 매트릭스는 모델의 예측 값과 결과값이 얼마나 일치하는가를 보여주는 표이다. 좌측 상단에서 우측 하단으로 향하는 대각선의 값이 각 클래스 분류의 정확도를 나타낸다. 정상과 Fault 7에 대한 분류는 정확도가 높은 반면 Fault 1과 Fault 2의 경우는 각각 정상으로 오분류하는 수치가 높게 나타나고 있다.

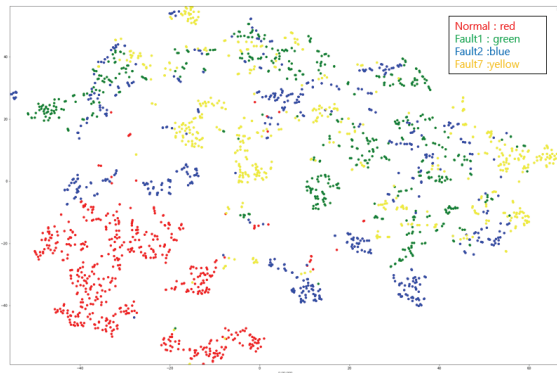


Fig. 8 T-SNE of 4-classes multiclass classification

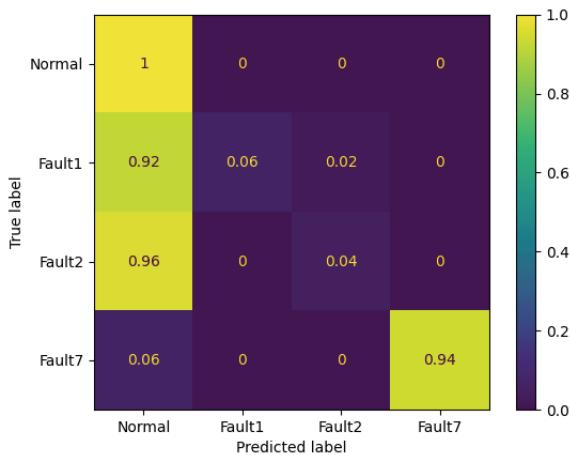


Fig. 9 Confusion matrix of 4-classes multiclass classification

따라서 단계를 분리하여 불량을 검출하는 2단계 불량 검출 알고리즘을 제안하게 되었다.

4.3 2단계 분류모델의 설계

정상과 불량 3가지 종류를 각각 분류하는 문제보다 중요한 것은 정상 데이터와 유형의 종류와 상관없이 불량을 검출해 내는 것이 더욱 중요한 문제이다. 따라서 분류 단계를 2단계로 나누어 진행함으로써 분류 정확도를 높이고자 하였다.

1단계에서는 정상과 불량을 검출하는 이진분류를 진행하고, 2단계에서는 불량만을 모아 3가지 유형으로 분류하는 다중분류를 실시하였다.

정상과 3가지 클래스(Fault 1, 2, 7)로 구성된 불량과의 이진분류의 결과 정확도는 0.908로 나타났으며, 불량 3가지 유형(Fault 1, 2, 7)만으로 구성된 3가지 클래스의 다중분류에서는 Table 8과 같이 정확도는 0.856을 나타내었다.

Table 8 Accuracy of 2-stage classification

Stage	Model	Accuracy
1	Cov_16,32, 64, 128, 256 + SE block	0.908
2		0.856

4.4 실험결과

기존의 이미지 분류에 좋은 성능을 보였던 백본 네트워크와의 비교분석을 통해 동작음 데이터에 적합한 어텐션 블록이 추가된 CNN기반의 분류 모델을 구성할 수 있었다. 또한 시각화 기법을 통한 분석으로 오류의 원인을 찾아내어 불량 검출 모델의 성능향상과 정확도를 높일 수 있는 2단계 불량 검출 알고리즘을 제안하였다. 4클래스 다중분류에서는 60.8%의 정확도를 보였으나, 단계를 분류하여 실험한 결과, 1단계 정상과 불량 이진분류 실험에서는 약 90.8%의 정확도를 보였으며, 2단계 불량 3가지 유형별 다중분류 실험에서는 약 85.6%의 정확도를 나타내었다. 불량 검출을 2단계로 나누어 분류의 정확도를 높일 수 있음을 실험으로 증명하였다.

Fig. 10은 제안한 CNN기반의 불량 검출 모델의 그레드캠³¹⁾을 이용한 시각화 결과를 보여준다. 그레드캠은 모델의 관심영역을 보여주는 딥러닝 시스템 해석방법 중 한가지로 식별에 유효한 관심 영역을 히트맵 형태로 나타낸다. Fig. 10에서와 같이 분류에 사용한 이미지에서 분류 예측에 가장 주요한 역할을 하게 한 이미지 영

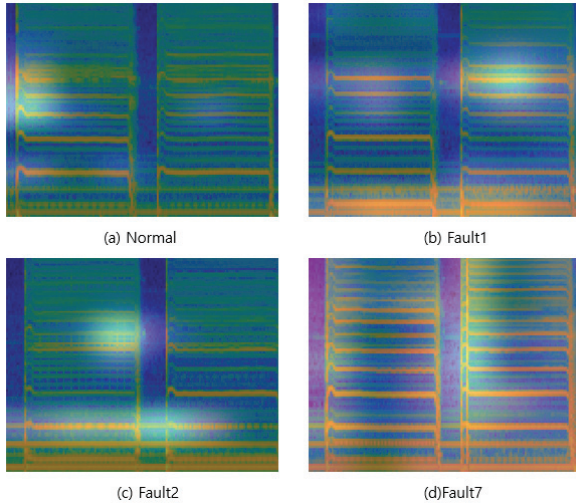


Fig. 10 Grad-CAM visualization result of datasets

역의 정보를 밝은 색으로 표시해 준다. Fault 1의 울렁거림과 Fault 2의 규칙적인 작은 파동, Fault 7의 추력음 발생부분을 밝은 색으로 표시하고 있으며, 이는 모델이 각 데이터의 특징을 잘 판단하고 분석하여 예측하고 있음을 의미한다.

5. 결론

본 논문에서는 모터 기어박스의 음향 신호를 기반으로 하는 CNN알고리즘을 사용한 2단계 딥러닝 불량 검출 방법을 제안하였다. 오디오 피쳐로서 멜 스펙트로그램을 선정하여 사용하였으며, 음향 신호 진폭의 변형을 통한 데이터 증강 기법으로 데이터의 양을 늘여 모델의 과적합을 방지하였다. 또한 언더 샘플링 기법으로 학습 데이터의 특정 클래스로의 편향을 막아 모델의 일반화 성능을 향상시켰으며, 모델의 설계에 있어 각 CNN레이어를 통과 후 어텐션 블록(SE/CBAM)을 삽입하여 채널 또는 공간 정보의 가중치를 부여해 정확도를 높였다.

동작음의 특성상 정상과 불량 패턴 차이가 아주 미세하거나 불량의 종류가 혼재되어 있는 경우, 검수자가 청각을 통해 라벨링 판단을 하기 때문에 잘못 라벨링하는 가능성이 존재한다. 이때 발생할 수 있는 라벨링 오류 데이터를 고려하기 위해 모델 설계시 라벨 스무딩 기법을 적용하였다.

2단계 CNN 기반 불량 검출 모델에서의 1단계에서는 모터 기어 박스의 정상과 불량을 구별하는 이진분류를 진행하였고, 2단계에서는 판별된 불량을 3가지 유형별로 분류하는 다중분류를 수행하였다. 1단계의 이진분류에서는 약 90.8%의 정확도를 나타냈으며, 2단계의 다중분류에서는 약 85.6%의 정확도를 나타내었다.

단지 정확도 측면에서 본다면 실제 공정에서 사용하기엔 아직까지 한계가 있지만 불량 검출에 인공지능을 결합하여 인력 검사에 따른 위험요소를 줄이고 균일한 분류를 할 수 있다는 가능성에서 본 논문의 의미가 있다고 본다.

제안한 방식은 1단계에서 정상과 불량을 판별한 후 2단계에서 불량을 유형별로 분류하였기 때문에 불량 유형별 발생빈도와 원인분석에 유용하게 사용될 수 있을 것이다.

본 논문에서는 실험실에서 취득한 데이터를 바탕으로 설계하였으나 실제 생산 공정에서 사용할 수 있는 실시간 불량 진단 시스템으로 발전시키기 위해서는 잡음과 변수가 많은 실제 공정에서 취득한 현장 데이터를 훈련 데이터로 활용하는 연구가 이어서 진행되어야 한다. 아울러 향후 이어지는 연구에서는 제조공정에서 수집된 모터의 음향신호 뿐 아니라 진동, 전류신호 등 복합 데이터를 활용하여 분류의 정확도를 높이는 후속 연구가 추가로 필요하다.

References

- 1) J. U. Ko, J. H. Jung, M. Y. Kim, H. B. Kong and B. D. Youn, "Noise Robust Fault Diagnosis Technique to Simultaneously Learn Classification and Denoising," Proceedings of the Korean Society of Mechanical Engineers Conference, pp.165-167, 2018.
- 2) S. J. Ham, S. Y. Han, S. G. Kim, H. J. Park, K. J. Park and J. H. Choi, "Fault Diagnosis Study of Train Door Using Motor Current Signal," Journal of Applied Reliability, Vol.19, No.4, pp.353-361, 2019.
- 3) C. Park. "A Study on the Quality Information Management for Preventing Human Errors at Quality of Ship," Ulsan University Graduate School of Automotive Vessel Technology, 2016, UCI:I084:48009-000002213977.
- 4) S. Lee and J. Joh, "Development of a Fault Diagnosis System for Assembly Small Motors Using ANN," Journal of the Korea of Precision Engineering, Vol.18, No.11, pp.124-131, 2001.
- 5) J. Patterson and A. Gibson, Deep Learning: A Practitioner's Approach, O'Reilly Media, Inc., Sebastopol, 2017.
- 6) R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," arXiv preprint arXiv:1901.03407, 2019.
- 7) R. Dubey, J. Zhou, Y. Wang, P. M. Thompson and J. Ye, "Analysis of Sampling Techniques for Imbalanced Data: An n=648 ADNI Study," NeuroImage, Vol.87, pp.220-241, 2014.

- 8) K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in Proceedings of the 23rd Annual ACM Conference on Multimedia, ACM Press, pp.1015-1018, 2015.
- 9) J. Salamon, C. Jacoby and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research," 22nd ACM International Conference on Multimedia (ACM-MM'14), Orlando, FL, USA, pp.1041-1044, 2014.
- 10) V. Boddapati, A. Petef, J. Rasmusson and L. Lundberg, "Classifying Environmental Sounds Using Image Recognition Networks," *Procedia Computer Science*, Vol.112, pp.2048-2056, 2017.
- 11) W. Oh, "Audio Classification Performance of CNN According to Audio Feature Extraction Methods," *Journal of Acoustical Society of Korea*, Vol.38, No.64, 2019.
- 12) J. Sharma, O. -C. Granmo and M. Goodwin, "Environment Sound Classification Using Multiple Feature Channels and Deep Convolutional Neural Networks," arXiv preprint arXiv:1908.11219, 2019.
- 13) X. Li, V. Chebiyyam and K. Kirchhoff, "Multi-stream Network with Temporal Attention for Environmental Sound Classification," arXiv preprint arXiv:1901.08608, 2019.
- 14) N. Dave, "Feature Extraction Methods LPC PLP and MFCC in Speech Recognition," *International Journal for Advance Research in Engineering and Technology*, Vol.1, No.6, pp.1-5, 2013.
- 15) Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," *Nature*, Vol.521, pp.436-444, 2015.
- 16) H. Zhou, Y. Song and H. Shu, "Using Deep Convolutional Neural Network to Classify Urban Sounds," TENCON 2017-2017 IEEE Region 10 Conference, pp.3089-3092, 2017.
- 17) T. Bluche, H. Ney and C. Kermorvant, "Feature Extraction with Convolutional Neural Networks for Handwritten Word Recognition," 2013 12th International Conference on Document Analysis and Recognition, pp.285-289, 2013.
- 18) X. Lu, X. Duan, X. Mao, Y. Li and X. Zhang, "Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection," *Mathematical Problems in Engineering*. Article ID 1376726, 2017.
- 19) C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.1-9, 2015.
- 20) A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.
- 21) K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, 2016.
- 22) K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," ICLR, 2015.
- 23) J. Salamon and J. P. Bello. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *IEEE Signal Processing Letters*, Vol.24, No.3, pp.279-283, 2017.
- 24) E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan and Q. V. Le, "AutoAugment: Learning Augmentation Strategies from Data," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.113-123, 2019.
- 25) E. D. Cubuk, B. Zoph, J. Shlens and Q. V. Le, "Randaugment: Practical Automated Data Augmentation with a Reduced Search Space," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.3008-3017, 2020.
- 26) D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer and B. Lakshminarayanan, "AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty," arXiv Preprint arXiv:1912.02781, 2019.
- 27) I. Sutskever, O. Vinyals and Q. V. Le. "Sequence to Sequence Learning with Neural Networks," *Advances in Neural Information Processing Systems*, pp.3104-3112, 2014.
- 28) D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv Preprint arXiv:1409.0473, 2014.
- 29) J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks," ILSVRC, 2017.
- 30) S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "CBAM: Convolutional Block Attention Module," ECCV, pp.3-19, 2018.
- 31) R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), pp.618-626, 2017.