사전 학습된 딥러닝 모델의 Mel-Spectrogram 기반 기침 탐지를 위한 Attention 기법에 따른 성능 분석

박창준[O], 김인기[*], 김범준[*], 전영훈[**], 곽정환(교신저자)[**]
[O]한국교통대학교 AI로봇공학과,
[*]한국교통대학교 교통에너지융합학과,
[**]한국교통대학교 소프트웨어학과
e-mail: cjp128@a.ut.ac.kr[O], cv2@kakao.com[*], zhskfkseh@gmail.com[*],
0hoon.jeon@gmail.com[**], jgwak@ut.ac.kr[**]

Attention Modules for Improving Cough Detection Performance based on Mel-Spectrogram

Changjoon Park[O], Inki Kim[*], Beomjun Kim[*], Younghoon Jeon[**], Jeonghwan Gwak(Corresponding Author)[**]
[O]Dept. of Ai·Robotics Engineering, Korea National University of Transportation,
[*]Dept. of IT·Energy Convergence, Korea National University of Transportation,
[**]Dept. of Software, Korea National University of Transportation

● 요 약 ●

호흡기 관련 전염병의 주된 증상인 기침은 공기 중에 감염된 병원균을 퍼트리며 비감염자가 해당 병원균에 노출된 경우 높은 확률로 해당 전염병에 감염될 위험이 있다. 또한 사람들이 많이 모이는 공공장소 및 실내 공간에서의 기침 탐지 및 조치는 전염병의 대규모 유행을 예방할 수 있는 효율적인 방법이다. 따라서 본 논문에서는 탐지해야 하는 기침 소리 및 일상생활 속 발생할 수 있는 기침과 유사한 배경 소리 들을 Mel-Spectrogram으로 변환한 후 시각화된 특징을 CNN 모델에 학습시켜 기침 탐지를 진행하며, 일반적으로 사용되는 사전 학습된 CNN 모델에 제안된 Attention 모듈의 적용이 기침 탐지 성능 향상에 도움이 됨을 입증하였다.

키워드: 기침 탐지(Cough Detection), 어텐션(Attention), 멜 스펙트로그램(Mel-Spectrogram), Bottleneck Attention Module, 사전 학습(Pre-trained)

I. Introduction

기침은 사람이 본인의 몸을 방어하기 위한 중요한 보호 작용으로 기관지와 연결된 어떠한 경로를 통해 이물질이 들어오는 것을 막으며, 기관지 내에 존재하는 분비물 혹은 해로운 물질을 배출하는 역할을 한다. 또한 기침을 통해 폐렴, 폐섬유증 등 폐 질환을 비롯한 기관지의 위험신호로도 발생하는데, 그 예로 최근 유행했던 COIVD-19의 대표적인 증상이다. 이처럼 신체를 방어하기 위한 행동이지만 전염성 바이러스를 가진 사람의 기침으로 인해 튀어 나간 분비물이 공중으로 퍼져 다른 사람들을 전염시킬 위험이 존재한다. 실제로 COVID-19 팬데믹 이후 폭발적으로 감염되는 사례는 줄어들고 있으나 전염성이 매우 높은 COVID-19 및 독감의 전염 위험성은 아직도 크게 자리 잡고 있으며, 전염 초기 국가적인 의료인력의 부재에 큰 어려움을 가지고 있었다. 이러한 문제점들을 해결할 수 있는 시스템 중 하나인 지능형 알고리즘 기반 질병 분류 시스템은 비대면 진단, 의료인력 부족 현상 해소, 비용 절감 등의 장점을 내세운다. 따라서 음성 기반 기침 탐지 지능형 알고리즘은 사람들이 많이 모이는 공공장소나 실내의 여러 장소에서의 기침 탐지는 전염병의 감염, 비 감염 여부를 판별할 수 있는 토대가 될 수 있으며, 기침 탐지 시 적절한 조치를 통해 비감염자의 전염병 감염을 예방할 수 있는 가장 효율적인 방법이며, 비대면 진단이 가능한 시스템으로 진단 중 전염의 위험성을 가지지 않으며, 개인 진단 및 특정한 상황 혹은 장소에서의 대규모 판별이 가능하다. 본 논문에서는 상업 시설 및 일반 환경에서 기침 탐지를 수행하는 CNN 기반 분류 모델들을 비교 및 분석하고, Attention 기법을 적용하여 성능 향상을 통한 강인한 모델 설계 방안을 제시한다. 실험을 위해 기침 외 다른 클래스가 포함된 음성 데이터셋을 Mel-Spectrogram으로 변환 후 이미지화하여 ImageNet을 기반으로 사전 학습된 CNN 모델들의 학습 및 성능 평가를 진행하였다. 학습된

CNN 모델 중 F1-Score을 기준으로 Top-3 모델을 선정하고 Attention 기법을 이용한 추가 실험을 진행한다. Attention 기법으로 Bottleneck Attention Module(BAM)[1], Convolutional Block Attention Module(CBAM)[2], Channel Attention[3], Spatial Attention[2]을 이용하였으며, 성능 변화를 분석하고 기침 탐지를 위한 CNN 모델에 효과적인 Attention 기법을 확인한다.

II. Related Works

1. RNN 기반 Cough Detection 기법

시계열 음성 신호의 예측을 위한 딥러닝 모델은 일반적으로 Recurrent Neural Network(RNN)를 기반으로 한 모델을 통해 진행되며, 기침 중 COVID19에 걸린 사람의 기침 탐지를 위한 RNN 모델 연구[4]를 통해 COVID19 감염자의 기침 탐지를 진행하기 위해 기존 음성 신호뿐만 아니라 Spectrogram, Mel-Spectrogram, Mel-Frequency Cepstral Coefficient(MFCC), MFCC의 Velocity Map, MFCC의 Acceleration Map 등 음성 신호를 다양하게 변환하여 해당 특징 추출 이후 적합한 특징을 선별해 학습에 입력으로 사용함을 알 수 있다. 이후 RNN 모델이 출력한 값을 이용해 분류기를 학습시켜 기침 탐지가 가능함을 알 수 있다. 제안하는 기법은 시계열 신호를 모델의 입력에 바로 사용하는 것이 아닌 해당 데이터를 변환 및 이미지화를 진행한 후 기침의 특징이 Fig. 1.과 같이 시각적으로 나타나는 부분을 판별한다.

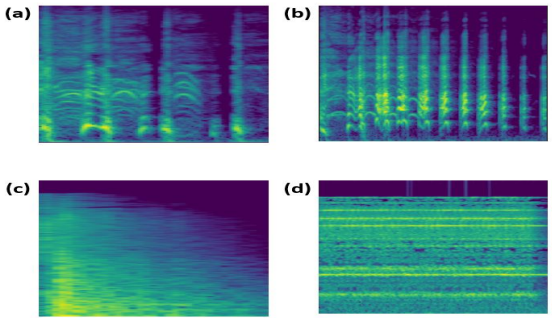

Fig. 1. (a) Cough, (b) Laughter, (c) Doorslam, (d) Phone, each class Mel-Spectrogram samples

2. CNN 기반 Cough Detection 기법

Convolutional Neural Network(CNN)를 기반으로 한 기침 탐지 모델 연구인 Deep learning based cough detection camera using enhanced feature 이경태 외 5명[5]에서 수행된 실험은 하나의 차원을 가지고 있는 시계열 음성 데이터를 각 모델에 적합한 특징으로 변환한 후 사전 학습된 CNN 모델에 입력으로 사용하여 분류기를 학습시킨다. 본 논문은 해당 과정 이후 각 모델의 기침 탐지 성능을 분석해 순위를 정해 Top-3로 선정된 모델에 Attention 기법을 적용한 결과를 확인한다.

III. Proposed Method

1. Environment Setting

제안하는 모델의 평가를 위한 실험 환경은 Table 1과 같으며, 모델의 최적화를 위해 Optimizer로 Adam, Loss는 Binary Cross Entropy, 하이퍼파라미터로 epoch = 100, learning rage = 0.0001, batch_size = 64를 사용하였다.

Table 1. System Environment

Item	Value
CPU	Intel i9-10900K
Memory Size	128GB
GPU	NVIDIA RTX 3090
Python	3.8.4
Pytorch	1.11.0
TorchVision	0.11.0
Cuda	11.3

2. Cough Dataset for Cough Detection

데이터셋은 Dcase2016-2[6]과 Coswara[7]의 dataset을 사용하였다. Dcase2016-2는 각 클래스당 20개의 데이터로 구성되어 있으며, 총 10개의 클래스 중 Cough, Clearthroat 클래스는 학습 데이터의 Cough 라벨로, Doorslam, Drawer, Keyboard, Knock, Laughter, Pageturn, Phone, Speech 클래스는 학습 데이터의 Backgrounds 라벨로 지정하였다. 해당 데이터는 불균형 하게 구성 되어 있기에 과적합을 어느정도 해소 하기 위한 목적으로 Coswara 데이터셋에서 Dcase2016-2의 Cough와 Clearthroat 클래스와 비슷한 양상을 보이는 데이터를 50개 선별하여 추가 한 후 학습 데이터의 Cough 라벨을 Augmentation 하여 최종 90개의 Cough, 90개의 증강된 Cough, 180개의 Backgrounds 데이터를 구성하여 데이터셋의 불균형성을 어느정도 해소한 후 학습을 진행하였다. Test 데이터로는 Dcase2016-2에서 제공하는 Test 데이터를 이용했으며, 두 개 이상의 클래스가 한 번에 나타나는 음성 데이터는 제외한 Cough 201개, Background 997개의 샘플을 이용하여 실험을 진행하였다.

3. Evaluation of Pre-trained Models

Pytorch에서 제공하는 ImageNet으로 사전 학습된 CNN 모델을 이용하여 기침 탐지 데이터셋을 학습시킨 후 성능 분석 및 Top-3을 선정하였다. 모델의 선정 기준은 Table 2의 Test Accuracy 순위를 토대로 EfficientNet B0, ResNet-152, ShuffleNet을 Top-3 모델로 선정하였다.

Table 2. Comparison of test result for CNN Models

	Accuracy	Loss
EfficientNet_B0	**92.7318**	**0.0127**
ResNet-152	**92.0635**	**0.4819**
ShuffleNet	**91.7293**	**0.1753**
Inception_V3	91.1445	0.0086
DenseNet-121	90.9774	0.2960
MobileNet	89.6408	0.0645
ResNet-50	89.5572	0.0021
EfficientNet_B4	89.1395	0.8014
EfficientNet_B3	89.1395	0.0772
EfficientNet_B2	88.7218	0.0577
ResNet-101	87.9699	0.0000
VGG-19	87.4687	0.0000
DenseNet-169	87.3016	0.0281
ResNext-101	83.8764	0.5085
EfficientNet_b1	83.7093	0.1227
Average	89.01	0.22

4. Attention Module-based Models

Top-3 모델인 EfficientNet B0, ResNet-152, ShuffleNet 모델에 BAM, CBAM, Channel Attentionn, Spatial Attention을 적용하였다. Attention 기법은 Figs. 2-4와 같이 CNN 모델의 내부의 각 특징 추출 블록 뒤에 추가하였다.

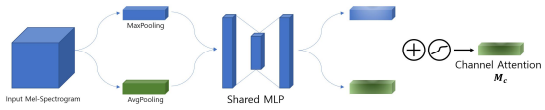

Fig. 2. Channel Attention Module Architecture

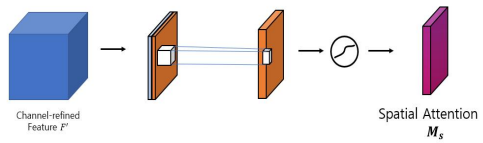

Fig. 3. Spatial Attention Module Architecture

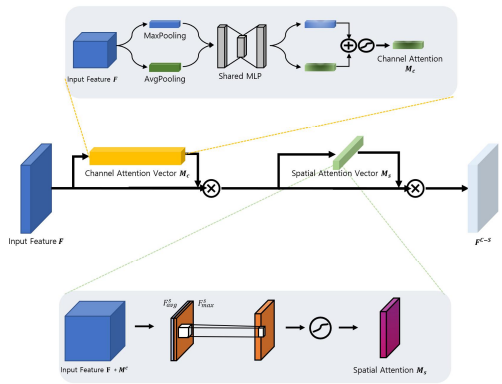

Fig. 4. Applied Attention Module

5. Result of Top-3 Models with Attention

Table 3은 Attention 모듈을 추가한 딥러닝 모델들의 성능을 비교한 표이다.

Table 3. Comparison of test result for our proposed model with CNN Models

	Accuracy	F1 Score
BAM_ResNet-152	**94.4862**	**0.9668**
CBAM_ShuffleNet	**92.8989**	**0.9584**
EfficientNet_B0(SE)	**92.7318**	**0.9552**
BAM_ShuffleNet	91.9799	0.9522
ShuffleNet	91.7293	0.9498
Iinception_V3	91.1445	0.9480
SE_ShuffleNet	88.7218	0.9307
Spatial_ShuffleNet	86.4662	0.9229
Spatial_ResNet-152	84.1270	0.9098
SE_ResNEt-152	83.0409	0.8970
ResNet-152	81.5372	0.8823
SE_ResNEt-152	78.1955	0.8588
CBAM_ResNet-152	69.5906	0.7962
CBAM_EfficientNet_B0	66.5831	0.7758
Spatial_EfficientNet_B0	63.7427	0.7407
CBAM_ResNet-152	58.6466	0.6946
BAM_EfficientNet_B0	41.0192	0.4660
Average	79.8000	0.8700

ResNet-152 모델에 BAM 모듈을 추가하여 학습을 진행 하였을 시 Test_Accuracy가 92.7318에서 94.4862로 증가하였고, Test_f1 최고 점수를 기록하였다.

Fig. 5.는 ResNet-152 모델의 혼동행렬과 BAM 모듈을 적용한 ResNet-152 모델의 혼동행렬을 비교한 그림이며 Attention 모듈을 적용한 이후 탐지한 기침 데이터가 123개에서 169개로 증가 하였으며, 이를 통해 BAM Attention 모듈을 적용한 ResNet-152 모델은 Attention 모듈을 적용하지 않은 모델보다 더욱 기침 탐지에 강인함을 시사한다.

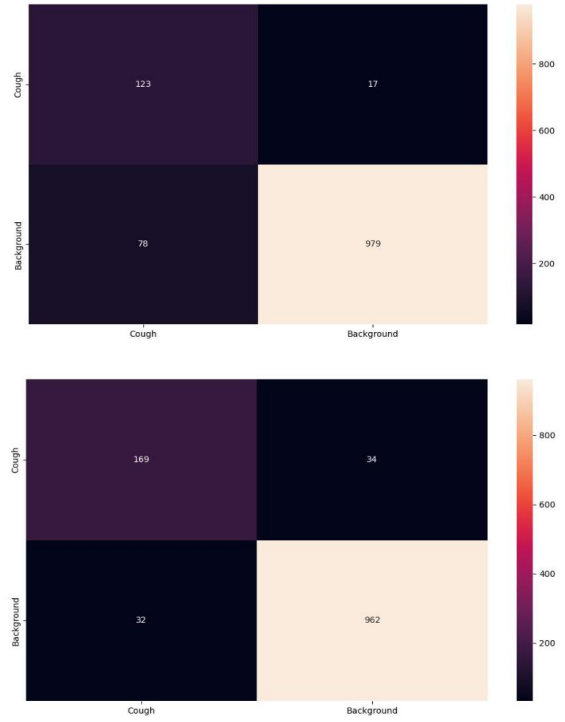

Fig. 5. Confusion Matrix of ResNet-152 and BAM_ResNet-152

IV. Conclusions

본 논문에서는 Mel-Spectrogram으로 시각화된 기침 정보를 강인하게 분류할 수 있도록 Attention 기법이 추가된 CNN 모델을 제안하였다. 불균형한 데이터셋을 이용한 학습으로 인해 사전 학습된 CNN 모델에 Attention 기법을 Top-3 모델에 추가 한 결과 ShuffleNet 모델의 경우 Test Accuracy가 증가했지만, Background 특징에 더 강인한 모습을 보이며 오히려 기침 탐지 성능은 저하 되는 모습을 보였다, 기본적으로 SENet의 Attention 모듈이 내장 되어있는 EfficientNet_B0 모델과 BAM Module을 적용한 ResNet-152 모델은 전체적인 데이터에 집중하여 기침 탐지에 좋은 성능을 냄을 입증했다. 특히, ResNet-152 모델은 Attention 모듈을 추가한 후의 기침 탐지 성능이 매우 높아지며 이는 적절한 Attention 기법을 모델에 추가 함으로써 기침 탐지 성능 향상의 잠재성을 보여주었다. 추후 연구에서는 다양한 Attention 기법을 섞은 Mixed Attention을 이용하여 더욱 강인한 기침 탐지 모델을 설계할 예정이다.

ACKNOWLEDGEMENT

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education [2021RIS-001(1345341783)]

REFERENCES

[1] J.Park, S. Woo, J.Lee, Y., and I.Kweon, "BAM: Bottleneck Attention Module", British Machine Vision Conference (BMVC), 2018

[2] S. Woo, J.Park, J.Lee, Y., and I.Kweon, "Cbam: Convolutional Block Attention Module", European Conference on Computer Vision (ECCV), pp. 3-19, 2018

[3] J. Hu, L. Shen, G. Sun, “Squeeze-and-Excitation Networks”, In Proceedings of The IEEE Conference on Computer Vision And Pattern Recognition, pp. 7132-7141, 2018

[4] A. Hassan, I. Shahin, and M. B. Alsabek “covid-19 Detection System Using Recurrent Neural Networks”, Institute of Electrical and Electronics Engineers (IEEE), pp. 1-5, November 2020

[5] G. Lee, H. Nam, S. Kim, S. Chio, Y. Kim and Y. Park, “Deep Learning Based Cough Detection Camera Using Enhanced Features”, Expert Systems with Applications, Vol. 206-117811, November, 2022

[6] K. Koutini, H. Eghbal-zadeh, M. Dorfer and G. widmer, “The Receptive Field as a Regularizer in Deep Convolutional Neural Networks for Acoustic Scene Classification”, European Signal Processing Conference (EUSIPCO), pp. 1-5, 2019

[7] D. Bhattacharya, D. Dutta, N. K. Sharme, S. R. Chetupalli, P. More, S. Ganapathy, C. C, S. Nori, S. K. K, S. Gonuguntla and M. Alagesan, “Coswara: A Website Application Enabling COVID-19 Screening By Analysing Respiratory Sound Samples And Health Symptoms”