UST 과학기술연합대학원대학교
UNIVERSITY OF SCIENCE & TECHNOLOGY

www.ust.ac.kr

14장

CNN 소개

이 홍 석 (hsyi@kisti.re.kr)
한국과학기술정보연구원 슈퍼컴퓨팅응용센터

프로젝트: 도시 교통 소리 분류를 위한 진단기

이홍석 (hsyi@kisti.re.kr)

합성곱 신경망을 이용한 도심지 교통 소리 분류

이홍석O 오현정 부이 칵 남 조지호
한국과학기술정보연구원 슈퍼컴퓨팅본부
hsyi@kisti.re.kr guswjd3927@naver.com hoainam.bk2012@gmail.com jhcho@kisti.re.kr

Urban Traffic Sound Classification Using Convolutional Neural Networks

Hongsuk YiO Hyeonjeong Oh Khac-Hoai Nam Bui Ji-ho Cho
Korea Institute of Science and Technology Information, Supercomputing Center
Korea Advanced Institute of Science and Technology

요 약

기존 교통 분야의 딥러닝을 활용한 연구들은 실시간 데이터를 시간 구간 단위의 단일한 평균속도 특성 값으로 모델을 훈련하고 교통상황을 예측하였다. 본 논문에서는 도심지 주요 혼잡 도로에서 발생하는 교통상황 소리의 특성을 이용하여, 교통상황을 분류할 수 있는 합성곱 신경망(Convolutional neural network, CNN) 모델을 제안한다. 제안하는 모델은 교통상황 영상에서 뽑아낸 교통 소리를 로그-스케일의 멜-스펙트로스램(Mel-scaled spectrogram)로 특성 추출된 이미지를 입력층으로 사용하였다. 또한, 교통 흐름 시간을 고려하여 세 개의 클래스를 구성된 출력층을 구성하였다. 계산 결과는 교통상황 분류의 정확도는 94.13%를 얻었다. 이 결과는 도심지 교통혼잡을 해결하기 위해서 교통소음을 사용하는 것이 적절하다는 것을 의미한다.

표 1 클래스 별 데이터 셋트 구성

클래스	시간	ClassID	개수
오전첨두	7시~10시	1	4919
비첨두	13시~15시	2	1800
오후첨두	17시~20시	3	2396
전체 개수			9115

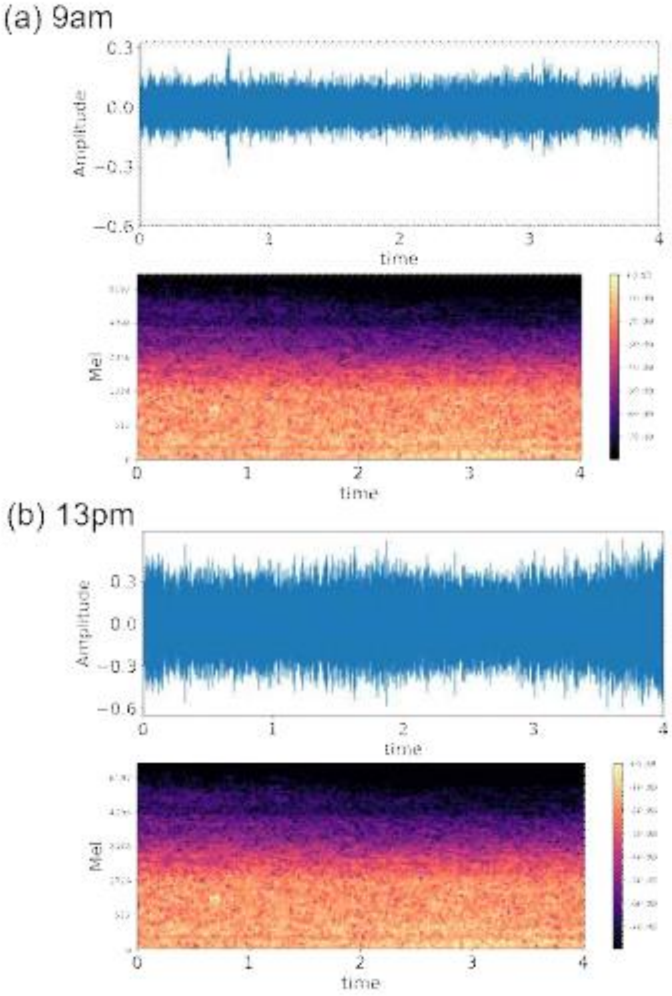


그림 1 (a) 음성 신호와 Log-Mel

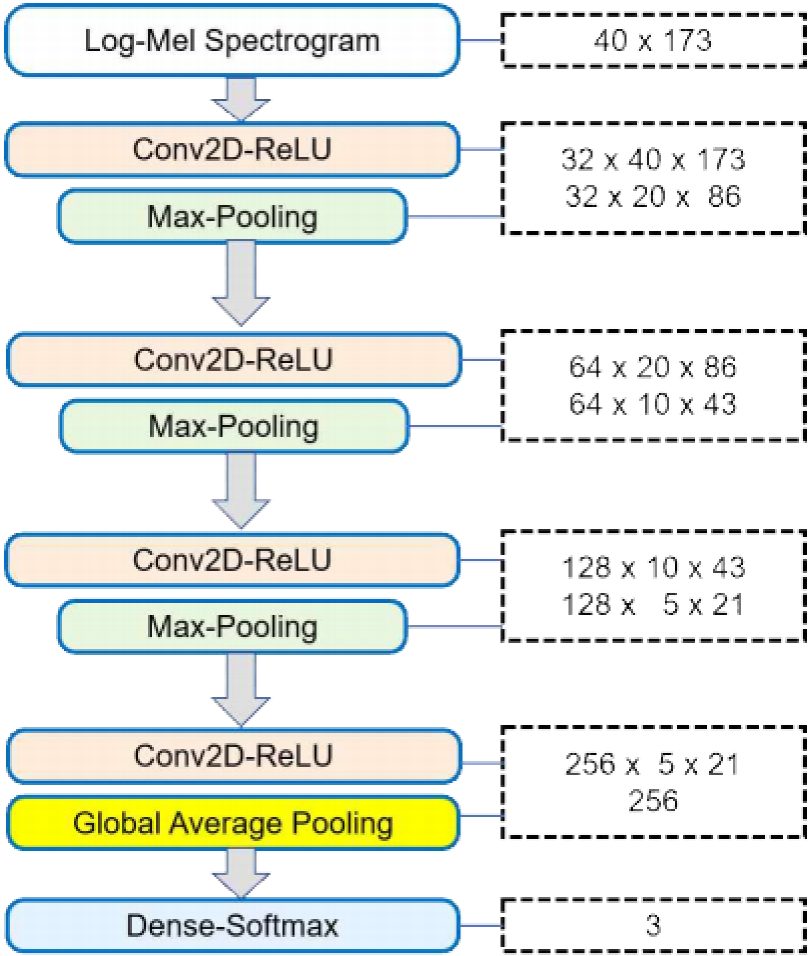


그림 2 도심지 교통소리 기반 CNN 모델

표3 CNN 모델의 정확도와 손실.

Data Set	ClassID	Label	Accuracy(%)
Log-Mel Spectrogrm	1	오전첨두	95.37
	2	비첨두	95.32
	3	오후첨두	91.97

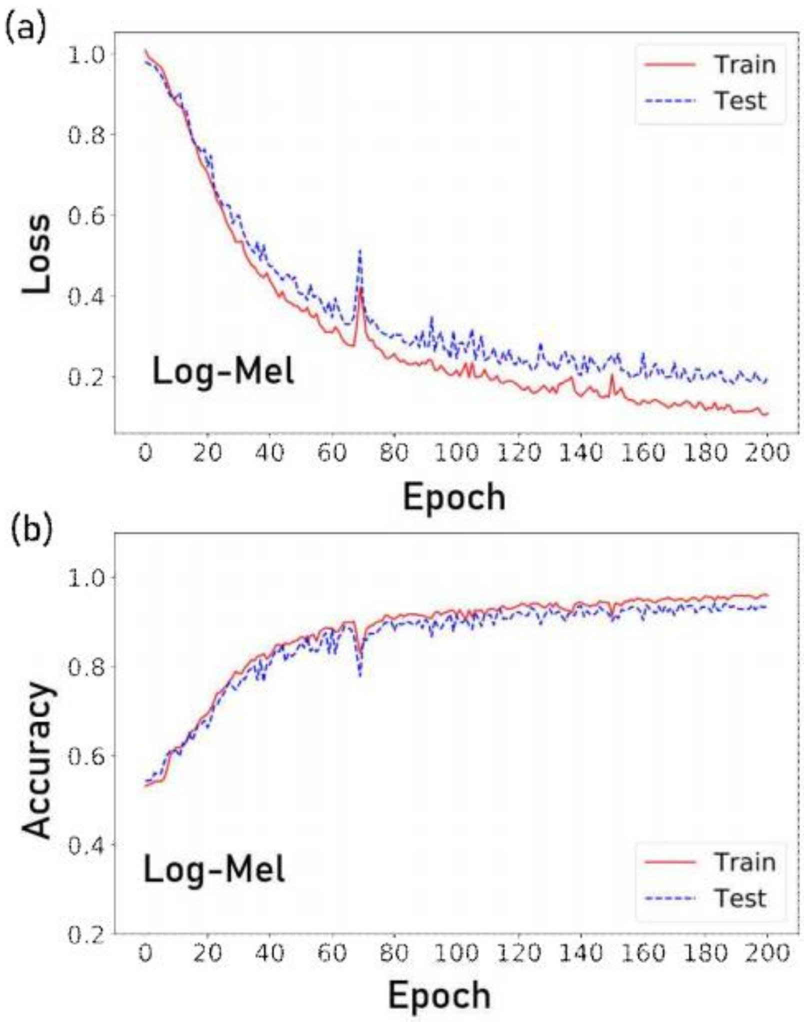


그림 3 Log-Mel의 (a)손실 함수와 (b) 정확도

Traffic Condition Classification using Road Sound Datasets

KHAC-HOAI NAM BUI, HYEONJEONG OH, AND HONGSUK YI
Korea Institute of Science and Technology Information

Corresponding author: Hongsuk Yi (e-mail: hsyi@kisti.re.kr).

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government (MSIT) (No.2018-0-00494, Development of deep learning-based urban traffic congestion prediction and signal control solution system) and Korea Institute of Science and Technology Information(KISTI) grant funded by the Korea government (MSIT) (K-20-L02-C09-S01).

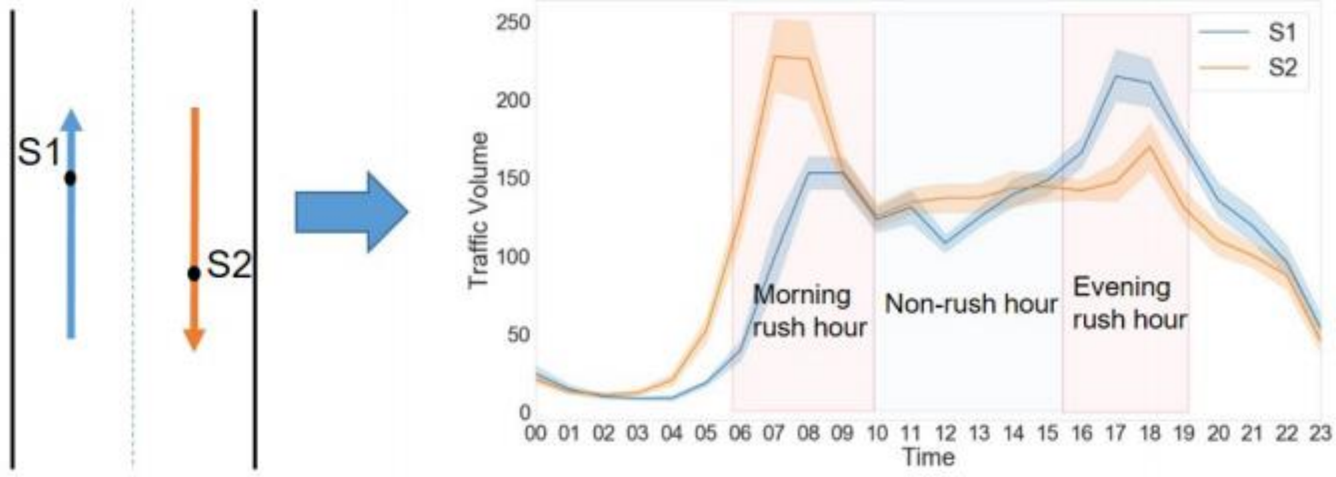


FIGURE 2. An example of measuring traffic volume at an asymmetric road using VDS.

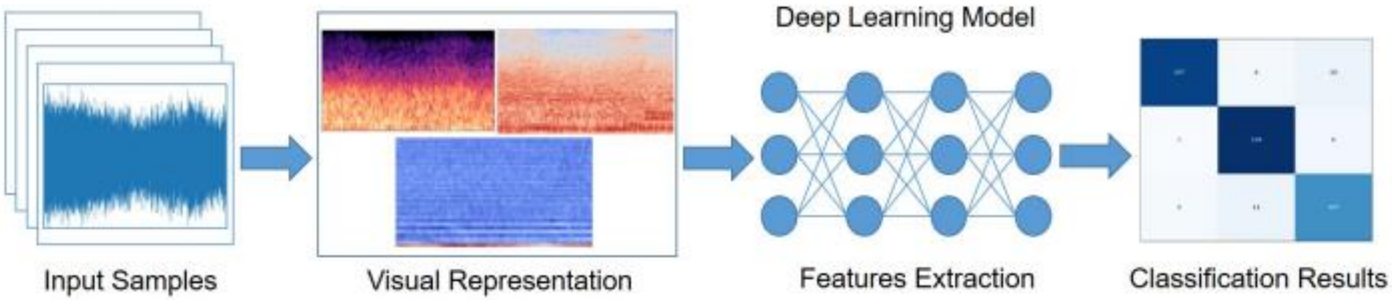


FIGURE 3. The main process of the urban sound classification problem.

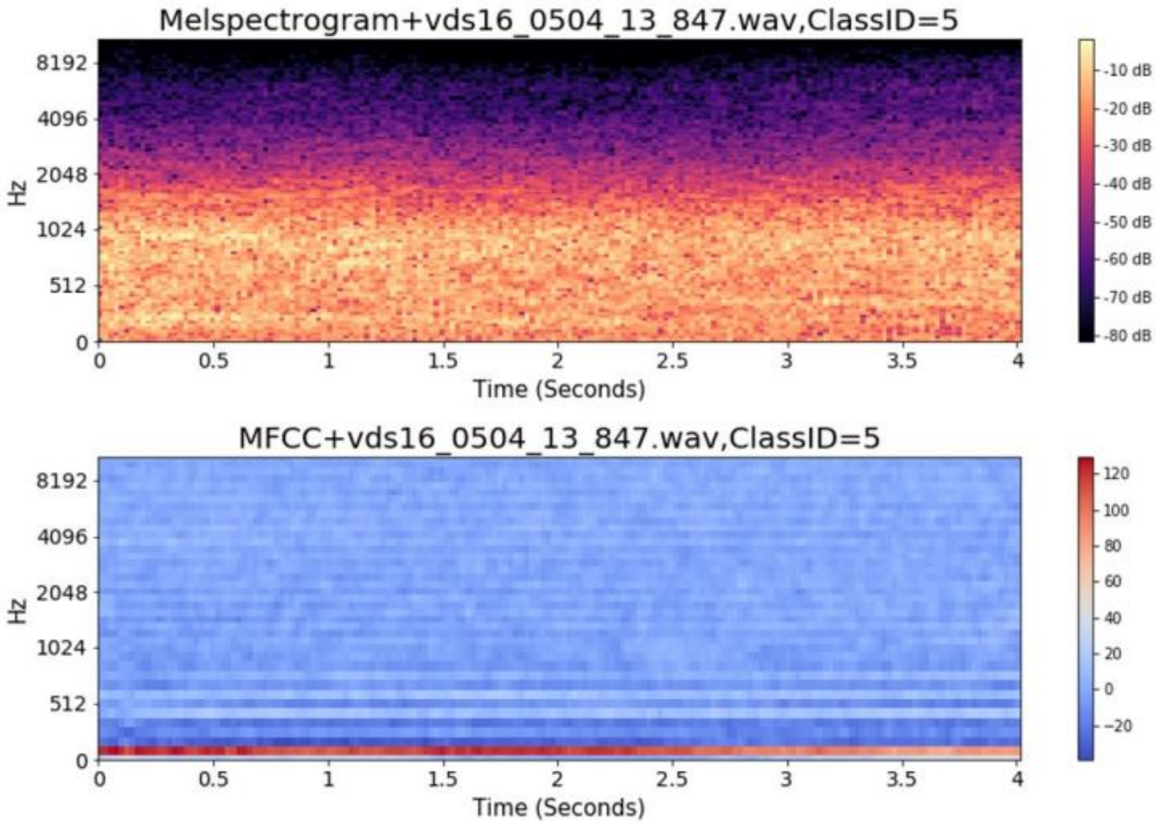


FIGURE 4. The output images by using Mel-Spectrogram and MFCC.

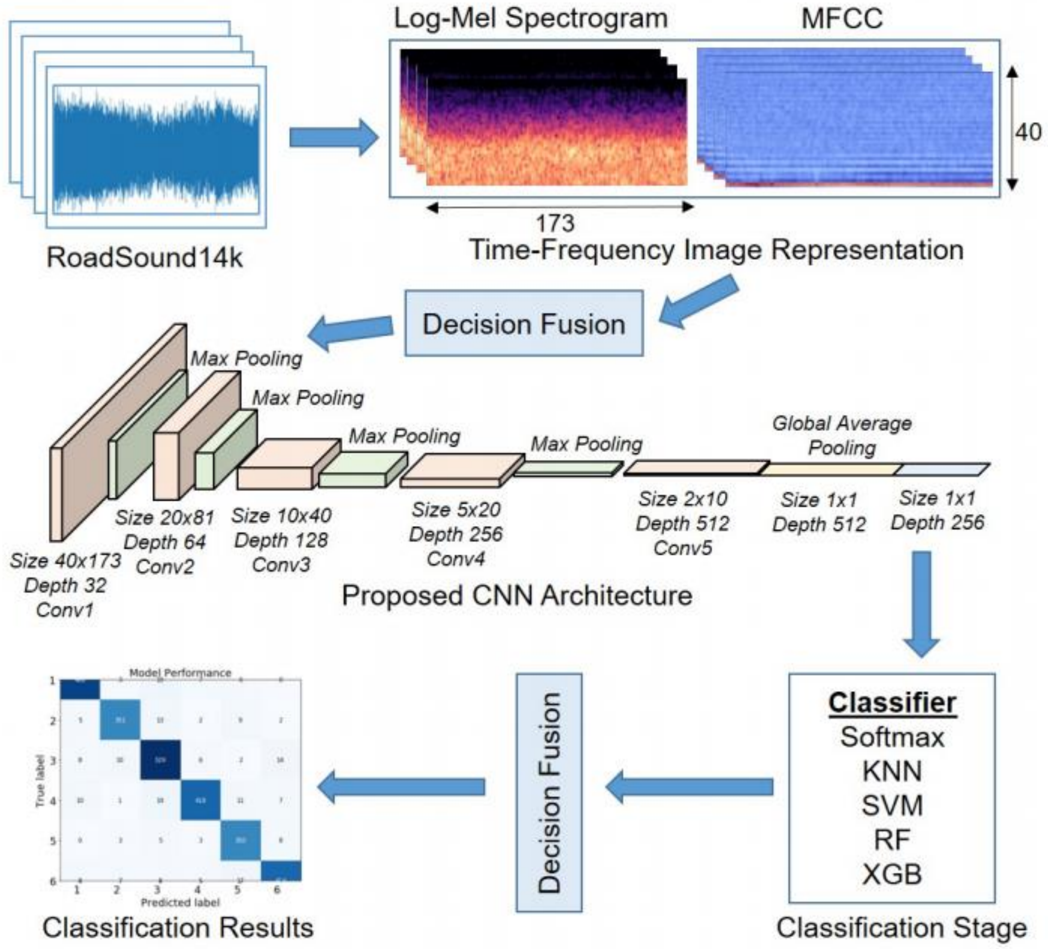


FIGURE 5. System architecture for the RSDC problem.

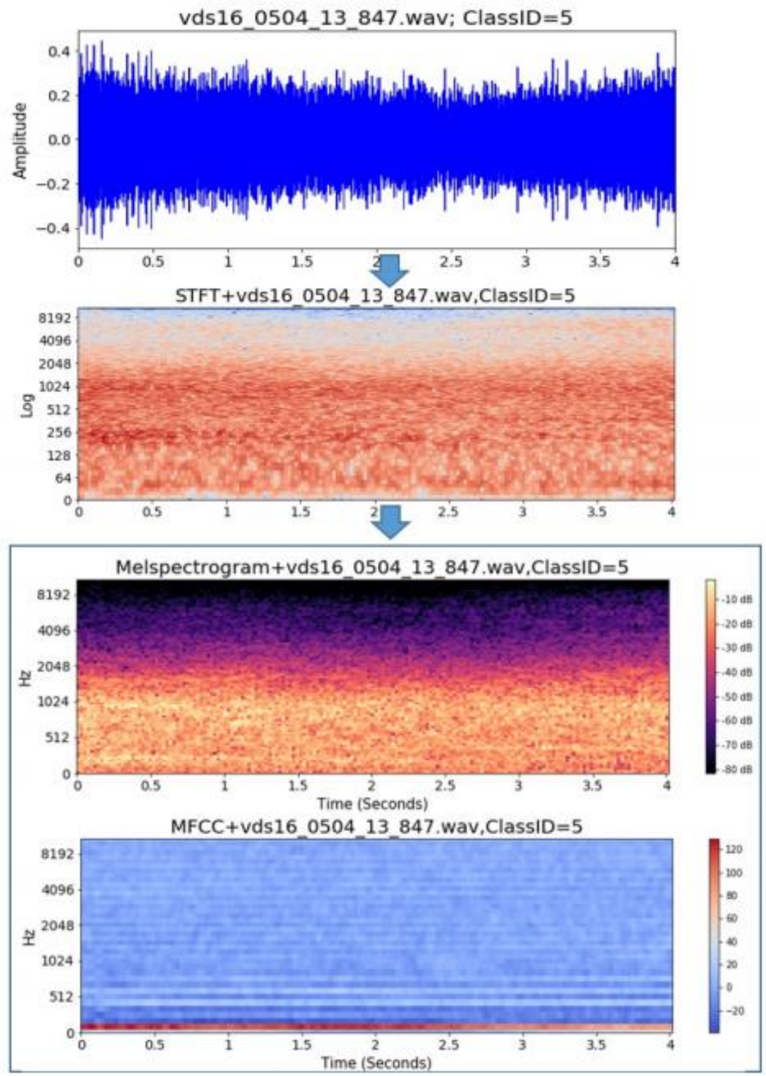


FIGURE 6. The results of visual representation process using log-Mel spectrogram and MFCC methods.

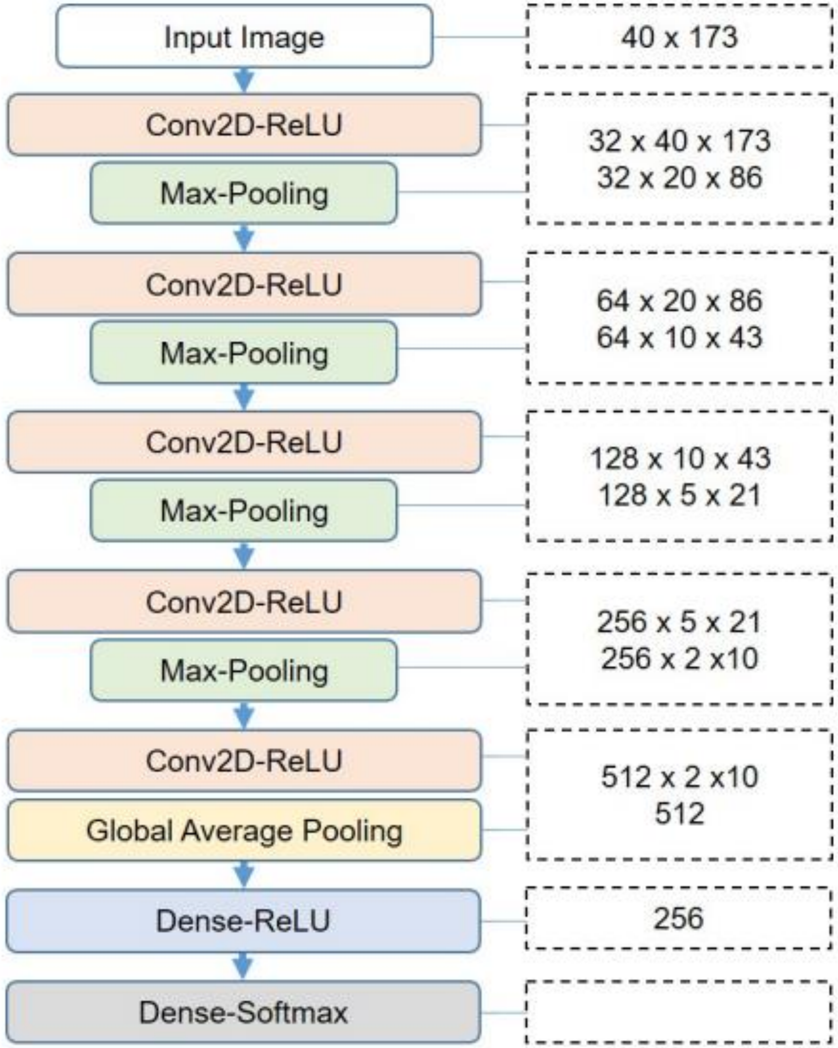


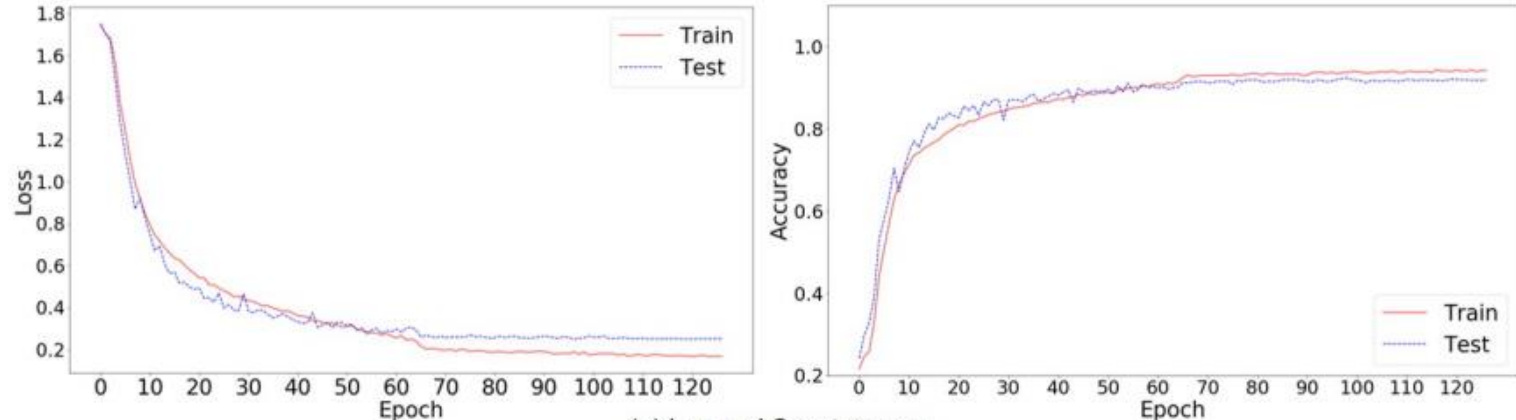
FIGURE 7. Proposed CNN architecture with Softmax for the classification.

$$Accurancy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

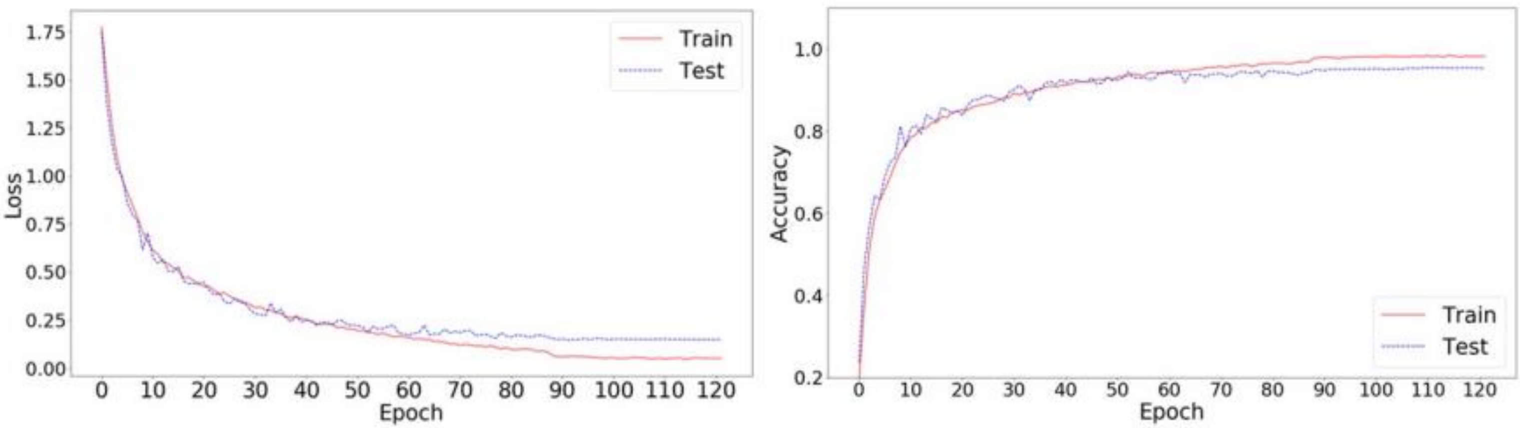
$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Prescision + Recall} \tag{10}$$



(a) Log-mel Spectrogram



(b) MFCC

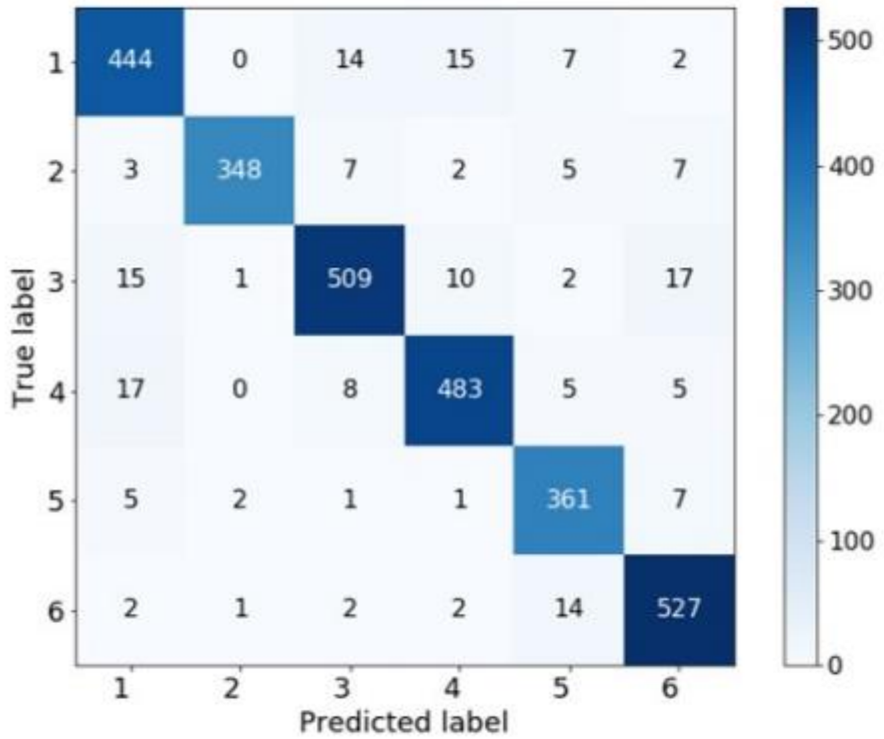


FIGURE 9. The prediction results using the proposed CNN model with softmax for the classification.

TABLE 2. Comparison results among CNN architectures

Evaluation Critical	AlexNet	VGG-16	VGG-19	Proposed CNN
Accuracy	0.9273	0.9277	0.9281	0.9372
Precision	0.9288	0.9284	0.9284	0.9378
Recall	0.9274	0.9277	0.9281	0.9372
F1-Score	0.9276	0.9278	0.9282	0.9373
Time(Sec)	398	830	1022	274

TABLE 3. Accuracy results (%) of different algorithms for the classification stage

ClassID	Softmax	KNN	SVM	RF	XGBoot
1	92.12	**92.53**	92.32	92.12	91.91
2	93.55	**95.43**	95.16	95.16	94.89
3	91.88	**93.86**	93.14	92.78	92.06
4	**93.24**	92.08	92.47	92.08	92.08
5	**95.75**	94.16	94.96	94.96	95.22
6	**96.17**	95.25	95.25	95.44	94.89
Average	93.72	**93.83**	93.79	93.65	93.37
Time(Sec)	-	638	45	333	**16**

TABLE 4. Classification results for the RSDC problem of the proposed CNN-KNN model

ClassID	Precision	Recall	F1-Score	Support
1	0.92	0.93	0.92	482
2	0.97	0.95	0.96	372
3	0.93	0.94	0.93	554
4	0.95	0.92	0.94	518
5	0.92	0.94	0.93	377
6	0.95	0.95	0.95	548
micro avg	0.94	0.94	0.94	2851
macro avg	0.63	0.63	0.63	2851
weight avg	0.94	0.94	0.94	2851

UST 과학기술연합대학원대학교
UNIVERSITY OF SCIENCE & TECHNOLOGY

Thank You!

www.ust.ac.kr