Report Date: 07/08/2022
To: ematson@purdue.edu, ahsmith@purdue.edu, lhiday@purdue.edu, and lee3450@purdue.edu
From: What is today's lunch?

- Ilmun Ku(mun90505@hufs.ac.kr)
- Seungyeon Roh(shtmddus99@konkuk.ac.kr)
- Gyeongyeong Kim(kky57389@sunmoon.ac.kr)
- Charles Taylor(taylo869@purdue.edu)

Summary

The methodology draft has been written. The methodology includes data augmentation methods and deep learning(DL) technologies. Data augmentation methods are such as time stretching and pitch scaling. DL technologies are namely Convolutional Neural Network(CNN) and Recurrent Neural Network(RNN).

What 'What is today's lunch?' completed this week:

• Methodology draft has been written.

B. Data Augmentation

Data augmentation methods are pivotal for the smooth and continuous improvement of audio classification performance while utilizing these DL algorithms. In this study, time-stretching, pitch **scaling**, time masking, and frequency masking are practiced to make augmentation for audio feature data. Time stretching and Pitch scaling are raw audio augmentation technology whereas time masking and frequency masking is spectrogram augmentation technology that treats augmentation as a visual problem, rather than an audio problem.
Time stretching is proven to be useful in enhancing accuracy for LSTM-based RNN using raw audio data [1]. It changes the speed and duration of sound without affecting the pitch or frequency of sound. Pitch scaling is recognized to be advantageous when exercised in advancing CNN accuracy [2] . It changes the pitch without affecting the speed of the data. Time masking and frequency masking are widely known to improve the network performance without the extra arrangement for the network or hyperparameter. [3]. They help the network to be robust against deformation and prevent overfitting by presenting corrupted data on purpose.

C. Deep Learning

CNN has demonstrated itself very effective not only for image classification but also has been shown to produce promising results for audio classification [1]. The details of CNN structure that is utilized in this study are summarized in Table 1. Keras of Tensorflow was employed to implement CNN model.

Layer	Stride	Out dim	Activation function	Kernel Size	Rate
Conv2D	(2,2)	32	ReLU	3 x 3	
Max Pooling 2D	(2,2)			2 x 2	
Batch Normalization	(2,2)				
Conv2D	(2,2)	32	ReLU	3 x 3	

Max Pooling 2D	(2,2)				
Batch Normalization	(2,2)				
Conv2D	(2,2)	32	ReLU	2 x 2	
Max Pooling 2D	(2,2)				
Batch Normalization	(2,2)				
Flatten		1			
Dense		64	ReLU		
Dropout					0.3
Dense(Output)		3	Softmax		

Table 1

RNN is proper to process sequential data such as text or sound data. However, RNN has its own downside which is gradient exploding and gradient vanishing [6]. To improve this issue, Long Short Term Memory(LSTM) networks were applied. In LSTM layers, self-recurrent weights make the cell in the memory block retain previous information [7] [8]. In CNN and RNN, dropout rate is set to 0.3. All default setting was practiced for all other parameters.

Layer	Stride	Out dim	Activation function	Kernel Size	Rate
LSTM	(2,2)	64	tanh	3 x 3	
LSTM	(2,2)	64	tanh	3 x 3	
Dense	(2,2)	64	ReLU	2 x 2	
Dropout					0.3
Dense(Output)		3	Softmax		

Table 2

• Methodology draft has been written.

Things to do by next week

- Going on a data collection trip with Mia.

Problems or challenges:

- Further Data collection especially noise of environment would be conducted in the future.

References

[1] RASHID, Khandakar M.; LOUIS, Joseph. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 2019, 42: 100944.

[2] ESA, M. F. M., et al. Learning convolution neural network with shift pitching based data augmentation for vibration analysis. In: *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. p. 012086.

[3] PARK, Daniel S., et al. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*, 2019.

[5] Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., ... & Wilson, K. (2017, March). CNN architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 131-135). IEEE.

[6] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *arXiv preprint arXiv:1211.5063*, 2012.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] Bae, S. H., Choi, I. K., & Kim, N. S. (2016, September). Acoustic Scene Classification Using Parallel Combination of LSTM and CNN. In *DCASE* (pp. 11-15).