Lung Sound Classification using lightweight architectures

Vasu Mehra

co-advised by

Prof Prasanta Kumar Ghosh and Prof Siddharth Barman

Indian Institute of Science, Bangalore

June 11, 2024

Introduction

- Lung Sound Analysis is extremely important for adventitious sound and disease detection
- But at the same time, reducing the model size and number of parameters is also essential

Timeline

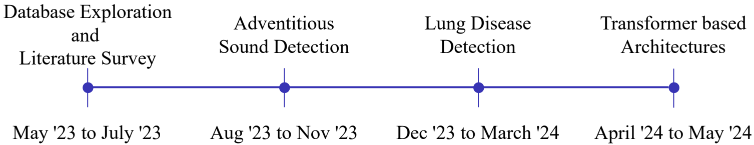
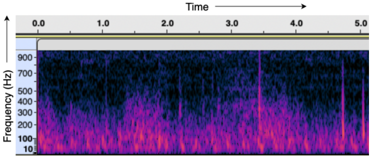


Figure 1: Timeline of the project work

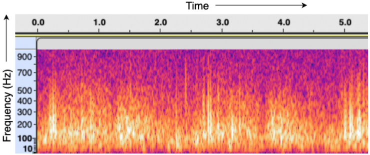
Problem Statement

- Analysis of lung sounds is essential for detection of any abnormality present in the lungs
- Abnormality is usually indicated by an adventitious sound, which might indicate the presence of a disease
- We also need to decrease the model parameters in check to ensure that the model size does not blow up

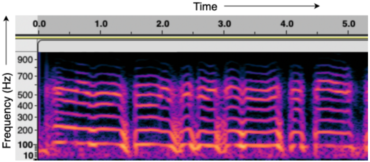
Understanding Lung Sounds



(a) Normal Lung Sound



(b) Crackles



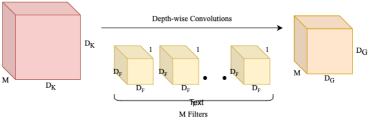
(c) Wheezes

Figure 2: Spectral representation of different types of lung sounds

Deep Learning Models

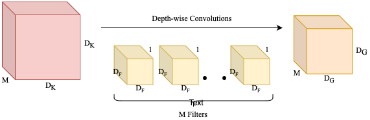
- Depthwise Separable Convolutional Networks
- Transformer architecture based on Facebook's `wav2vec2-base`

Depthwise Separable Convolutional Networks

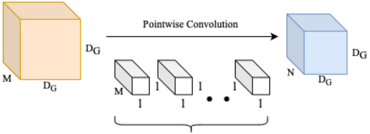


(a) Depthwise Convolutions

Depthwise Separable Convolutional Networks



(a) Depthwise Convolutions



(b) Pointwise Convolutions

Figure 3: Steps involved in Depthwise Separable Convolutional Networks

Depthwise Separable Convolutional Networks

- Comparison between the number of parameters in regular CNN and Depthwise Separable CNN(DSC)

$$
\begin{aligned}
\frac{\text{no. of params in regular CNN}}{\text{no. of params in DSC}} &= \frac{D_k^2 \times M \times F}{D_k^2 \times M + M \times F} \\
&= \frac{D_k^2 \times M \times F}{D_k^2 \times M + M \times F} \\
&= \frac{D_k^2 \times F}{D_k^2 + F}
\end{aligned}
$$

Depthwise Separable Convolutional Networks

- Comparison between the number of parameters in regular CNN and Depthwise Separable CNN(DSC)

$$\frac{\text{no. of params in regular CNN}}{\text{no. of params in DSC}} = \frac{D_k^2 \times M \times F}{D_k^2 \times M + M \times F}$$
$$= \frac{D_k^2 \times M \times F}{D_k^2 \times M + M \times F}$$
$$= \frac{D_k^2 \times F}{D_k^2 + F}$$

- For a 28x28x3 image and 64 filters, the ratio of the number of parameters is $50176/848 \approx 59.17$!

Wav2Vec 2.0 Architecture

- We use a classification head with the original Wav2Vec 2.0 architecture for lung sound classification.

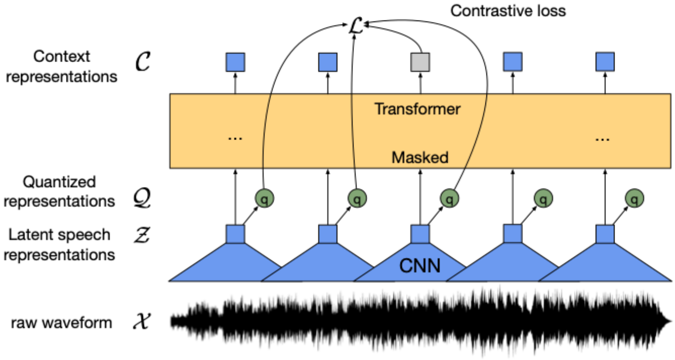


Figure 4: Facebook's `Wav2Vec 2.0` architecture for speech recognition

Transformer Pruning

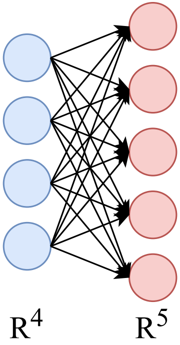
Transformer Pruning

- We introduce bottleneck layers into the feed-forward networks of each encoder block.

भारतीय विज्ञान संस्थान

Transformer Pruning

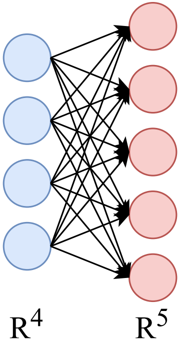
- We introduce bottleneck layers into the feed-forward networks of each encoder block.



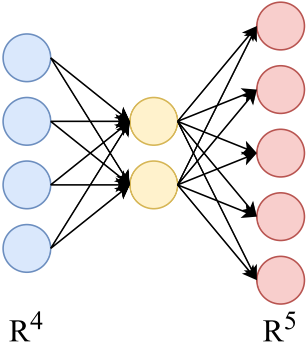
(a) before using bottleneck layer

Transformer Pruning

- We introduce bottleneck layers into the feed-forward networks of each encoder block.



(a) before using bottleneck layer



(b) after using bottleneck layer

Figure 5: Bottleneck layer to restrict the number of parameters

Task 1: Adventitious Sound Detection

Architecture	Accuracy	Senstivity	Params
MNRNN[et.18]	65.6 %	62.1 %	
CNN[NP20]	**81.3%**	**82.6%**	4.9M
DSC[1]	78.7 %	79.5 %	500.5 K
`wav2vec2-base`	29.4 %	24.5 %	94.6 M
wav2vec2-architecture[2]	38.9 %	29.6 %	59.8 M

[1] Depthwise Separable CNN(our work)

[2] our best wav2vec2 architecture with 512 neurons in bottleneck layer

[3] our best wav2vec2 architecture with 256 neurons in bottleneck layer

Results and Discussion

Task 1: Adventitious Sound Detection

Architecture	Accuracy	Senstivity	Params
MNRNN[et.18]	65.6 %	62.1 %	
CNN[NP20]	**81.3%**	**82.6%**	4.9M
DSC[1]	78.7 %	79.5 %	500.5 K
`wav2vec2-base`	29.4 %	24.5 %	94.6 M
wav2vec2-architecture[2]	38.9 %	29.6 %	59.8 M

Task 2: Lung Disease Detection

Architecture	Accuracy	Senstivity	Params
ResNet+NL[ea20b]	64.2 %	63.2 %	
CNN-RNN[ea20a]	66.38%	48.6 %	
DSC	**85.9%**	**74.3%**	**12.6K**
wav2vec2-architecture[3]	81.6 %	65.7 %	44.1 M

[1] Depthwise Separable CNN(our work)

[2] our best wav2vec2 architecture with 512 neurons in bottleneck layer

[3] our best wav2vec2 architecture with 256 neurons in bottleneck layer

Conclusion and Future Work

Conclusion

- Depthwise Separable Convolution gives good results for disease classification task.
- Adventitious Sound Detection task requires more feature learning and parameters than disease classification.
- The bottleneck layer successfully learns the representations for decomposition

Future Work

- More data should be collected and annotated by experts.
- Pretraining of our wav2vec2 architecture should be completed till cnvergence.
- Research to make wav2vec2 independent of source sampling rate and training on more useful features like mel-spectrogram, and MFCC.

References I

J. Acharya et. al., *Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning*, 2020.

Y. Ma et. al, *Lungrn+nl: an improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation, interspeech*, 2020.

Kirill Kochetov et.al, *Noise Masking Recurrent Neural Network for Respiratory Sound Classification*, 2018.

Truc Nguyen and Franz Pernkopf, *Lung sound classification using snapshot ensemble of convolutional neural networks*, 2020.