End-to-end network based on Transformer for automatic detection of COVID-19

Institute of Automation Chinese Academy of Sciences

Intelligent Interaction Team

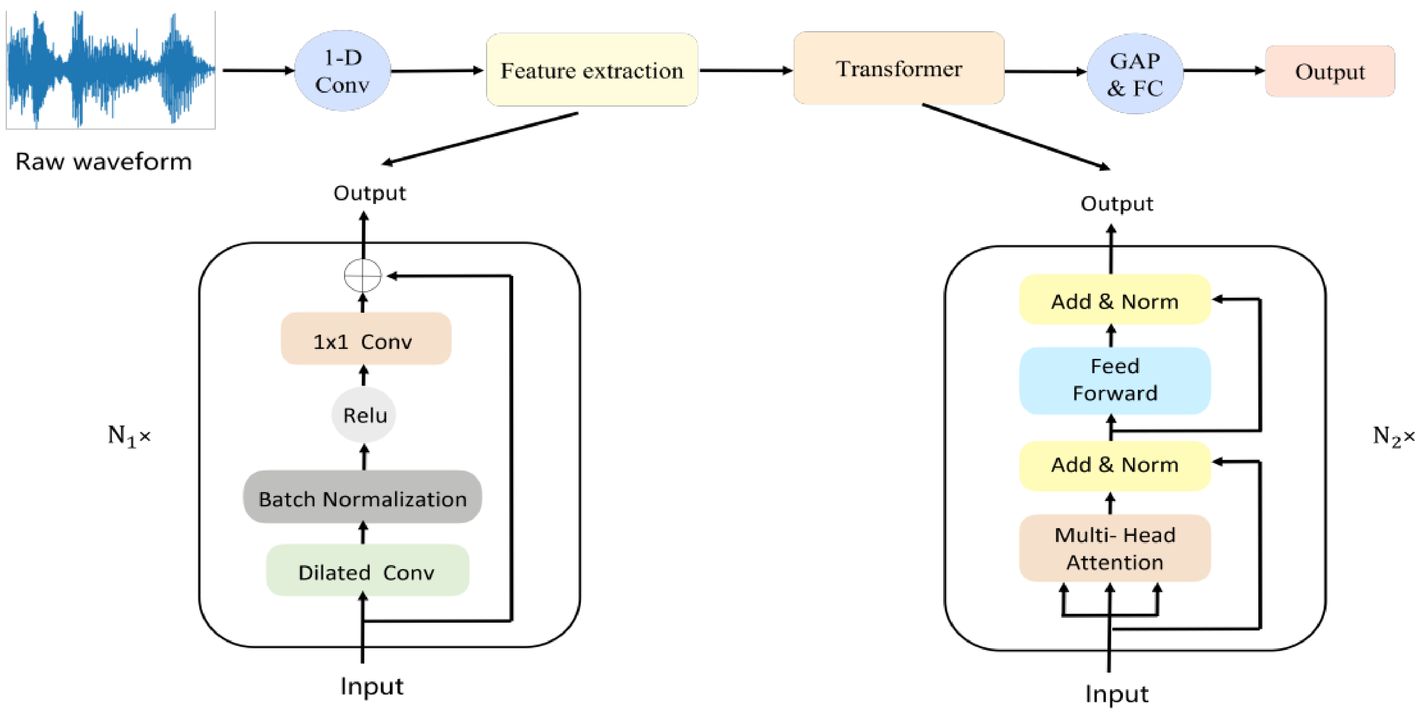
Cong Cai, Bin Liu, Jianhua Tao, Zhengkun Tian, Jiahao Lu, Kexin Wang
cong.cai@nlpr.ia.ac.cn

Introduction

Motivation:
Deep learning and machine learning technology can be used to analyze cough sounds of COVID-19 patients and infer predictions. We propose an end-to-end network based on Transformer for automatic detection of COVID-19.

Contributions:
1. The previous work of detection of COVID-19 is based on spectrogram or handcrafted features. We directly model the raw waveform signals, which can automatic extract features and do end-to-end learning.
2. The dilated convolution architecture can increase the receptive field exponentially to solve the high-dimensional sparse problem.
3. We evaluate the performance of the Transformer architecture in the field of speech signal processing.

Proposed Methods



Our method mainly consists of the feature extraction module and the Transformer.

1. The feature extraction module consists of N dilated convolution blocks. Like temporal convolutional network (TCN), the dilated convolution block includes the dilated convolution, the batch normalization, the ReLU function, the 1D convolution and the residual path.

2. We only use the encoder part of the Transformer. It is based on multi-head attention mechanism.

Experiments and Results

Comparison with Other Features
We compared the MFCC features and the Wav2vec features to compare our methods. Our method surpasses MFCC and achieves the best performance, which shows that it is feasible to classify COVID19 by end-to-end learning features.

Features	AUC
Raw Waveform	72.1%
Wav2vec	78.6%
MFCC	81.5%
Proposed Method	**83.2%**

Transformer Classification Strategy
We verified the class token, global average pooling, and direct use of two full connection layer classification. The ”None” means to directly use two full connected layers for classification.

Metheds		AUC
ResNet-50		82.5%
Transformer	None	81.6%
	Class Token	82.8%
	GAP	**83.2%**

Comparison with Expert
we compare our method with the expert diagnosis in the COUGHVID dataset. Our method outperforms expert diagnosis in detecting COVID-19.

Metheds	Specificity	Sensitivity	AUC
Expert	79%	25%	/
Proposed Method	**87%**	**63%**	**78.4%**

Future work

In the future, we will evaluate our method on other datasets or other task to improve the robustness and universality of our method.