Applying Variational Circuits in Deep Learning Architectures for Improving Discriminative Power of Speaker Identification Embeddings

Raphael Blankson[1]

Evgeniy Pavlovskiy, PhD[2]

[1] [2]Department of Mathematics and Mechanics
Novosibirsk State University

Overview

Intro

Background Of Study

- **Variational Circuits:** are parameter-dependent quantum circuits that can be optimized by a classical computer with regard to a given objective [1]
- **Embeddings:** are functions that maps classical data into quantum states that can be acted upon by quantum operations for quantum information processing.

Background of Study

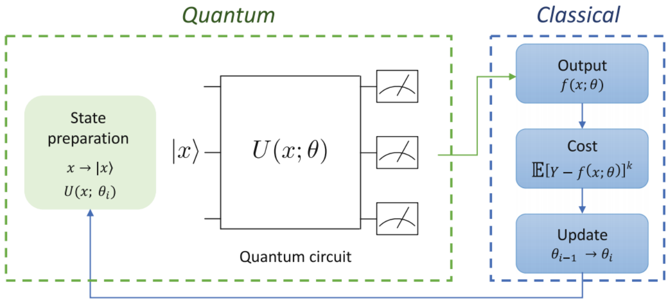


Fig. 1. Scheme of a hybrid quantum-classical algorithm for supervised learning. The quantum variational circuit is depicted in green, while the classical component is represented in blue. (Color figure online)

Problem Statement

- Modern dataset is constantly increasing and classical Machine learning algorithms in the near future will suffer computational bottlenecks, thus the pressure to innovate machine learning is fast increasing everyday [2].
- At the time of writing, there's no quantum application to audio data.

Purpose of the research

- Investigate the influence of quantum variational circuits on the generalization power of deep learning architectures (ResNet18) on speaker embeddings.
- We investigate using photonic quantum circuits.

Limitation of Study

- Limited number of qumodes.
- No perfect quantum hardware available, thus we have to make use of Near term quantum computers that have limited qumodes.
- The current photonic quantum computer available in the world is not yet ready for machine learning tasks.
- The overhead of simulating quantum algorithms on classical devices is very high and slow.
- Coming up with quantum algorithms that outperform classical computers is very difficult since the laws of physics restricts our access to information stored in quantum systems [4].

Definition of Terms

- Qubits or Qumodes or Wires
- Gates
- Circuits
- Fourier transform
- FFT - Fast Fourier Transform
- MFCC - Mel Frequency Cepstral Coefficients
- STFT - Short Time Fourier Transform
- Spectogram

Audio img

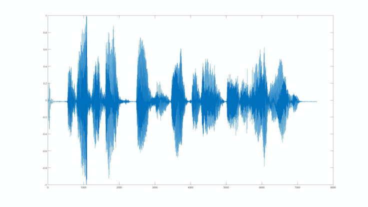


Figure 1: Raw audio waveform

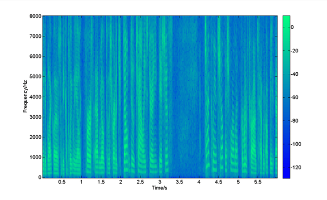


Figure 2: Spectogram

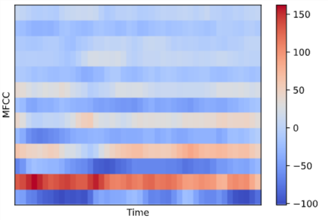


Figure 3: MFCC

Review

Transfer Learning Hybrid Neural Network [6]

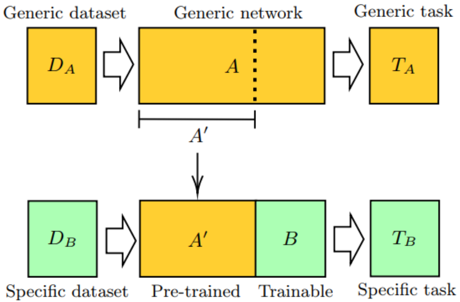


FIG. 1. General representation of the *transfer learning* method, where each of the neural networks A and B can be either classical or quantum. Network A is pre-trained on a dataset D_A and for a task T_A. A reduced network A', obtained by removing some of the final layers of A, is used as a fixed feature extractor. The second network B, usually much smaller than A', is optimized on the specific dataset D_B and for the specific task T_B.

Continuous Variable Neural Networks [3]

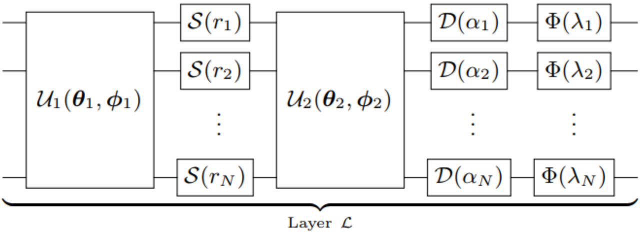
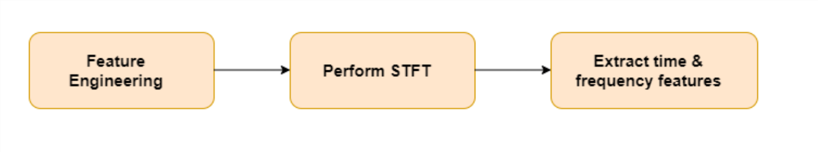


FIG. 1. The circuit structure for a single layer of a CV quantum neural network: an interferometer, local squeeze gates, a second interferometer, local displacements, and finally local non-Gaussian gates. The first four components carry out an affine transformation, followed by a final nonlinear transformation.

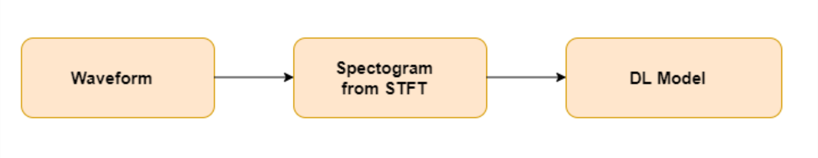
Deep Learning for Audio Signal Signal Processing [5]

- Deep learning has enabled practical applications in signal processing often outperforming traditional signal processing techniques on a large scale.
- Raw audio samples form a one-dimensional time series signal, which is different from images (2-D).
- Unlike images, audio signals have to be studied sequentially and in chronological order (audio-specific solutions).

Traditional ML Preprocessing pipeline for Audio data



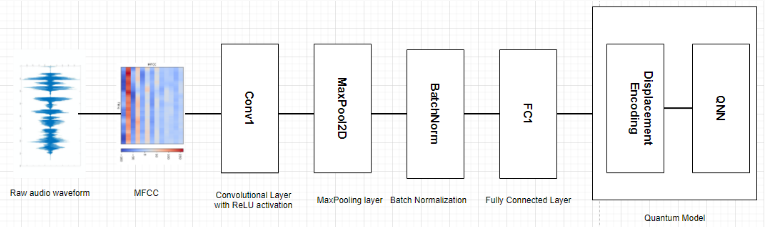
Deep Learning pipeline for Audio data



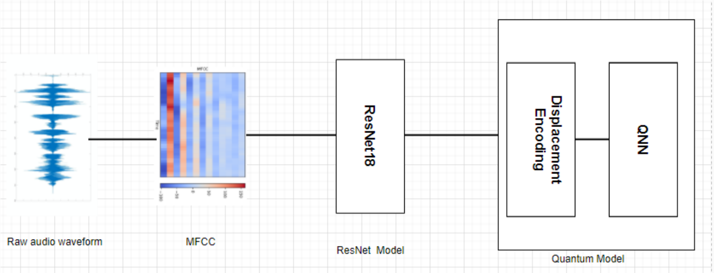
Methodology

Architectural Design

- CNN + QNN:



- **ResNet18 + QNN:**
  - Up-Sample the audio to required input of the RESNET (3, 224, 224).
  - Replace the output to the required number of classes of the speaker dataset.



Experiment and Results

Experimental Environment

- Librosa was used to preprocess audio data.
- Experiments were conducted using Pytorch-Lightning.
- The quantum experiments were run using pennylane on the strawberryfields simulator.
- Future experiments could include running on actual strawberryfields hardware.

Encoding Methodology

- The best method to encode data into a Quantum circuit is still an open research question.
- Data was encoded using the `Displacement` encoding.

Datasets

- **Research Experiment(Audio) Data**
  - Experiment with Speaker Dataset from kaggle originally containing **5** famous speakers with duration **1** second and sample rate **16kHz**
  - 2 speakers Nelson Mandela and Benjamin Netanyau were selected for the experiment.
  - **25%** of data for test and **20%** of remaining for validation

Results

Table 1: Initial Results

ResNet18	CNN + QNN	ResNet18 + QNN
98%	92%	[74%]

Latest Updates

Latest Updates - Mixup

- Mixed two audio data using process called mix-up (create a form of superposition)

$$\tilde{x} = \lambda x_i + (1-\lambda) x_j, \qquad \text{where } x_i, x_j \text{ are raw input vectors}$$
$$\tilde{y} = \lambda y_i + (1-\lambda) y_j, \qquad \text{where } y_i, y_j \text{ are one-hot label encodings}$$

Latest Updates

- Accuracy improved in the quantum models

Table 2: Results

Exp	CNN + QNN	ResNet18 + QNN
Validation	98%	99%
Test	100%	100%

Latest Updates

- Paper presented at ICDSA 2021 and proceedings will be published in SCOPUS Indexed Springer Book Series

Conclusion

Conclusion

- Quantum advantage in prediction accuracy is still possible -more investigations are needed to justify this claim.
- Eventhough there is still a lot of work to do, quantum machine learning remains a very promising emerging field of research.

Key References

- Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M. Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, Keri McKiernan, Johannes Jakob Meyer, Zeyue Niu, Antal Száva, and Nathan Killoran **PennyLane: Automatic differentiation of hybrid quantum-classical computations**. 2018. arXiv:1811.04968.
- Martin Hilbert and Priscilla Lopez, **The world's technological capacity to store, communicate, and compute information,** Science, 332(6025):60-65,2011.
- Killoran, Nathan & Bromley, Thomas & Arrazola, Juan & Schuld, Maria & Quesada, Nicolas & Lloyd, Seth. (2019) **Continuous-variable quantum neural networks.**. Physical Review Research. 1. 10.1103/PhysRevResearch.1.033063.

Key References

- Maria Schuld, Ilya Sinayskiy, Peter Shadbolt, and Francesco Petuccione, (2014), **An introduction to quantum machine learning,** Contemporary Physics, 1336-5812.
- Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schluter, Shuo-yiin Chang, Tara Sainath, (2019) **Deep Learning for Audio Signal Processing,** Journal of Selected Topics of Signal Processing, Vol.13,No.2,May 2019, PP.206-219.
- Mari, Andrea & Bromley, Thomas & Izaac, Josh & Schuld, Maria & Killoran, Nathan. (2019). **Transfer learning in hybrid classical-quantum neural networks.**
- Strawberry fields documentation, "*https://strawberryfields.readthedocs.io*"

Key References

- Homayoon Beigi, **Fundamentals of Speaker Recognition,** ISBN 987-0-387-77591-3, Springer, 2011.
- M. Mclaren, Y. Lei, L. Ferrer, **Advances in deep neural network approaches to speaker recognition,** in Proc. of ICASSP, 2015, pp.4814-1418
- Kaggle speaker dataset **www.kaggle.com/kongaevans/speaker-recognition dataset.**
- Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, et al, (2014), "*A variational eigenvalue solver on a photonic quantum processor,*" in Nature Communications 5, 4213.

Thanks

Questions?