Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Speech Recognition

C.-H. Huck Yang[1], Jun Qi[1], Samuel Y.-C. Chen[2], Pin-Yu Chen[3], Sabato Marco Siniscalchi[1,4,5], Xiaoli Ma[1], Chin-Hui Lee[1]

[1]Georgia Institute of Technology, [2]Brookhaven National Laboratory, [3]IBM Research, USA; [4]University of Enna, Italy; [5]NTNU, Norway

Overview

We propose a new decentralized feature extraction approach in federated learning to address privacy-preservation issues for speech command recognition with variational quantum circuits.

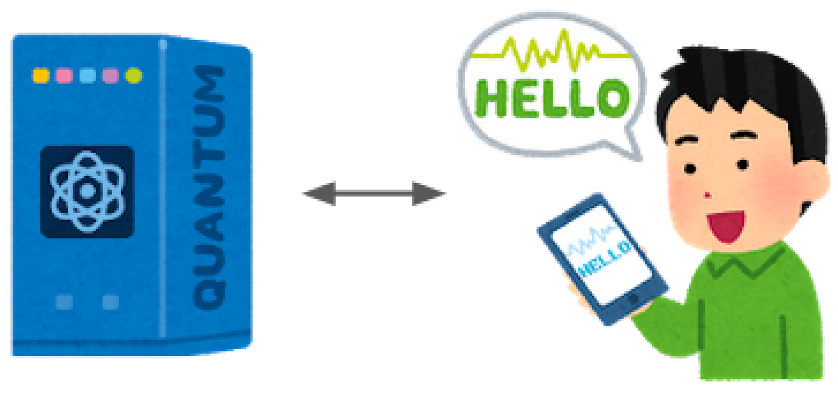
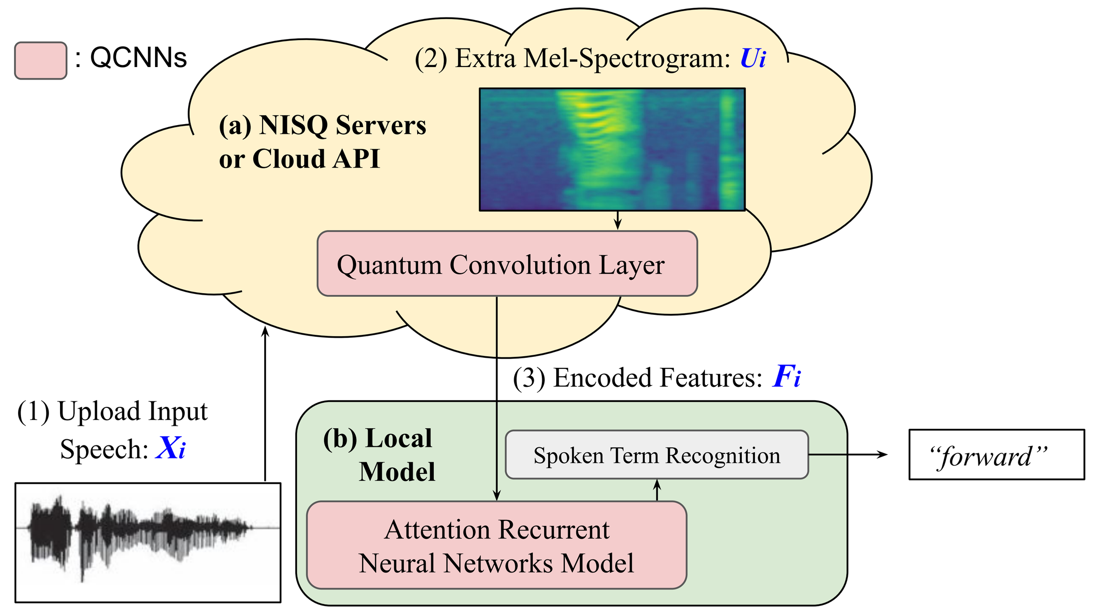


Figure 1: Quantum Computing & Speech Recognition

Keywords: Acoustic Modeling, Quantum Machine Learning, Automatic Speech Recognition, and Vertical Federated Learning.

Architecture

Proposed quantum machine learning for acoustic modeling (QML-AM) architecture in a vertical federated learning progress including:



- (a) a quantum convolution layer on Noisy Intermediate-Scale Quantum (NISQ) servers or cloud API.
- (b) a local model for speech processing.

Quantum Convolution (Quanv)

- 2D Mel-spectrogram input vectors are chunked into several 2×2 patches, and the n^{th} patch is fed into quantum circuit; encoded into intial quantum states, $\mathbf{I_x}[n] = e(\mathbf{u}_i[n])$.
- The initial quantum states go through the quantum circuit with the operator $q(\cdot)$, and generate $O_x[n] = q(\mathbf{I}_x[n])$.
- The outputs after applying the quantum circuit are necessarily measured by projecting the qubits onto a set of quantum state basis that spans all of the possible quantum states and quantum operations. Thus we get the desired output value, $f_{\mathbf{x},n} = d(\mathbf{O_x}[n])$.

More details refer to our demonstration.

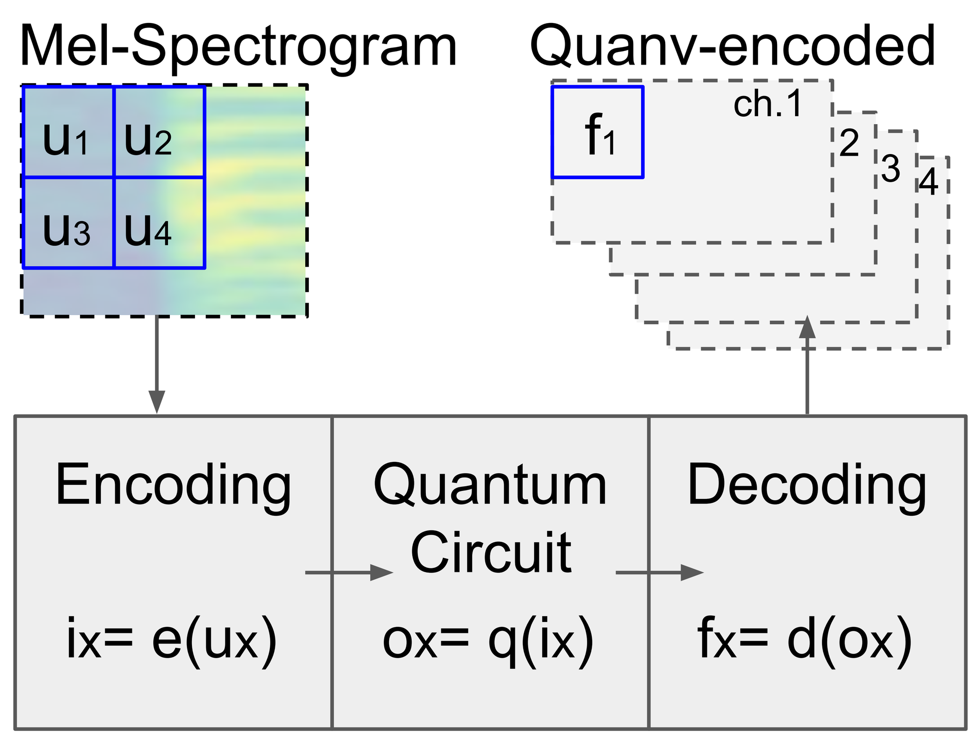


Figure 2: QCNN Computing Process.

Performance

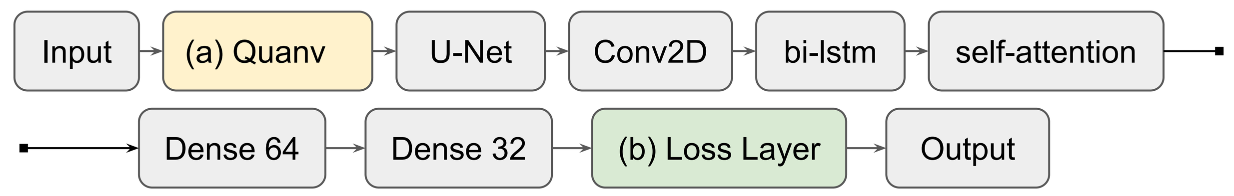


Figure 3: Proposed QCNN architecture for ASR tasks.

The loss layer could further be replaced by connectionist temporal classification (CTC) loss for character-level ASR in our Demonstration.

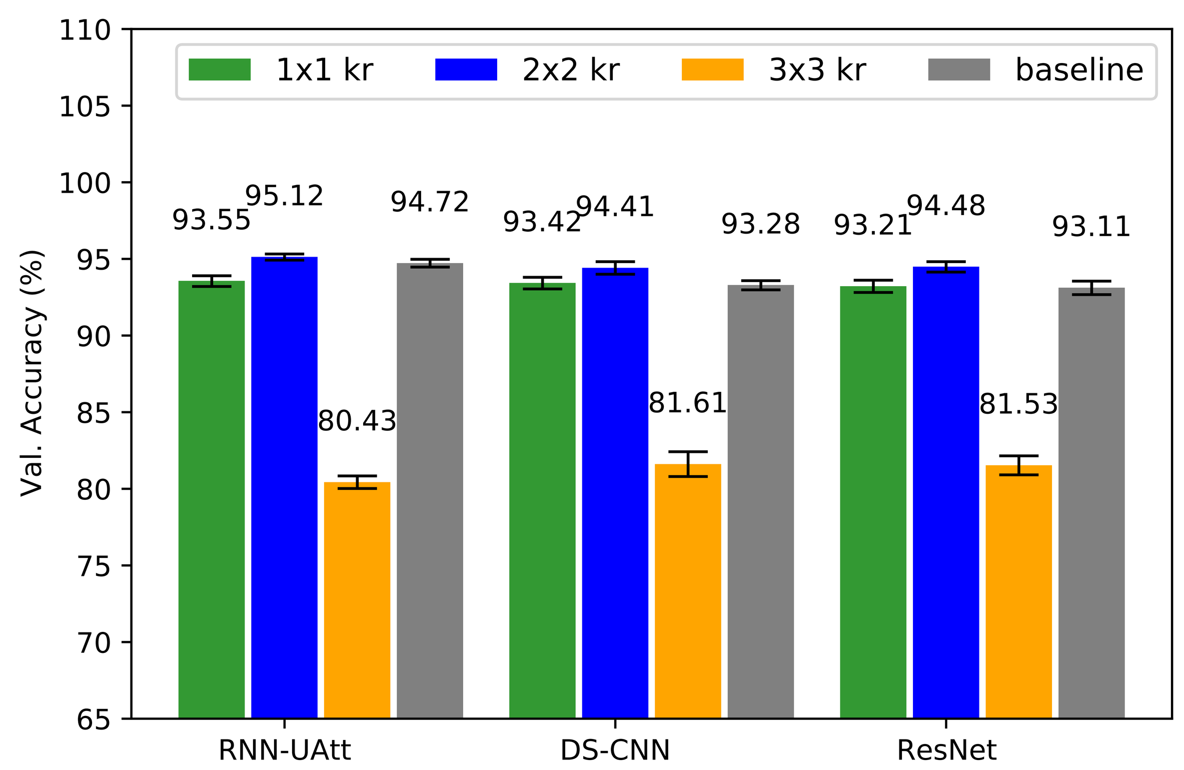


Figure 4: Performance studies of different quantum kernel size (dubbed kr) with DNN acoustic models for designing QCNN models.

QCNNs with the 2×2 kernel show better accuracy and lower deviations than all other models tested.

Model	Acc. (↑)	Para. (↓) / Memory
(1) RNN_{Att}	94.2	170,915
Conv + (1)	94.32	174,975
Quanv + (1)	94.75	174,955 + 4 (qubits)
(2) RNN_{UAtt}	94.72	176,535
Conv + (2)	94.74	180,595
Quanv + (2)	**95.12**	180,575 + 4 (qubits)

Table 1: Comparisons of spoken-term recognition on Google Commands dataset with the noise setting for classification accuracy (Acc). The additional convolution (conv) and quantum convolution (quanv) layer have the same 2×2 kernel size.

ICASSP Demo Show & Tell Session

IJCAI Tutorial QNN for Speech & NLP

QCNN Encoded Features

We visualize the different acoustic features extracted from CNN and QCNN encoders.

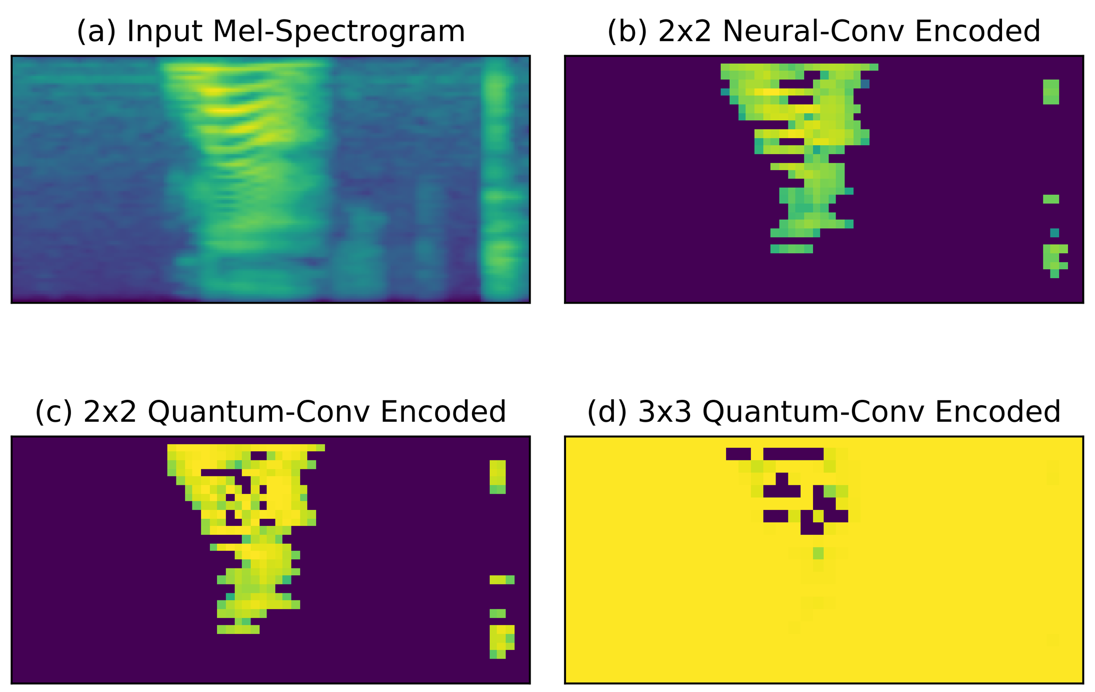


Figure 5: Visualization of the encoded features from different types of convolution layers. The audio transcription is "yes" of the input speech.

The IBM Qiskit quantum computing tool is used to deploy and simulate the 4 and 9 qubits QCNN.

Neural Saliency Analysis

Class activation mapping (CAM) evaluates different neural acoustic models on Mel-features.

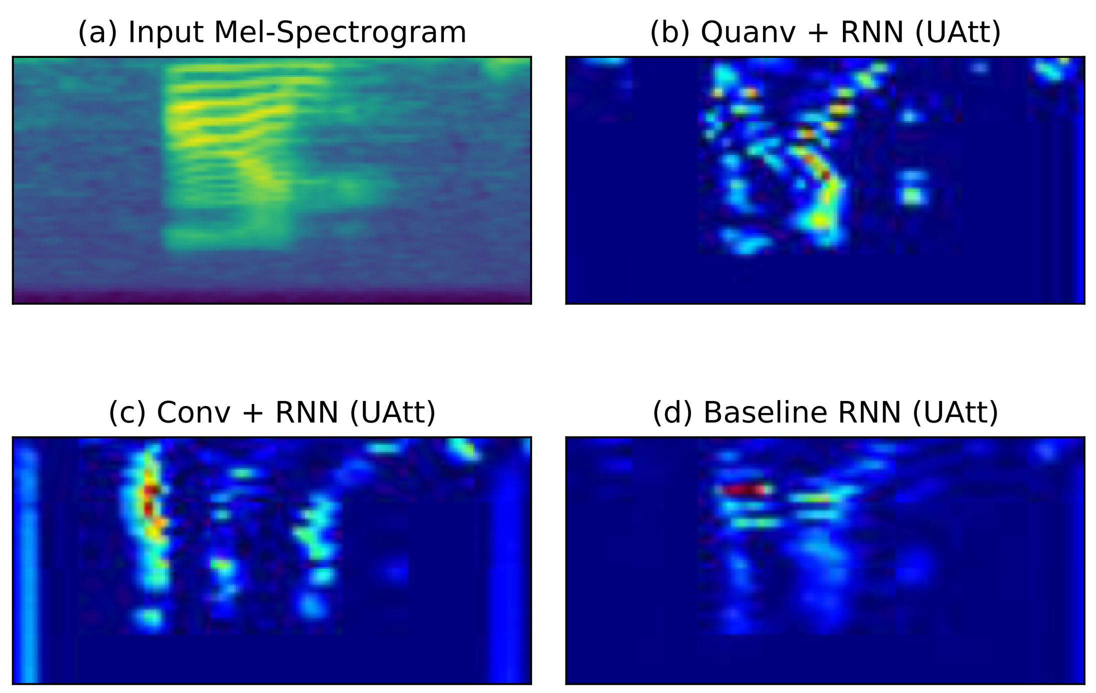


Figure 6: (a) Mel features with command of "on"; (b) a 2×2 quantum convolution layer followed by RNN_{UAtt}; (c) a well-trained 2×2 CNN layer followed by RNN_{UAtt}, and (d) baseline RNN_{UAtt}.

QCNN identifies related low-frequency patterns.

Reference

[1] Yang et al. Decentralizing feature extraction with quantum convolutional neural network for automatic speech recognition. *IEEE ICASSP*, 2021.

Conclusion and Future Research

The proposed QCNN models show competitive recognition results for spoken-term recognition on quantum machines and simulators.

Our future work includes incorporating QCNN into continuous ASR and applying statistical privacy measurements.

More Information

Code `https://github.com/huckiyang/QuantumSpeech-QCNN`

Email `huckiyang@gatech.edu`