ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448120

EXPLOITING A QUANTUM MULTIPLE KERNEL LEARNING APPROACH FOR LOW-RESOURCE SPOKEN COMMAND RECOGNITION

Xianyan Fu[1], *Xiao-Lei Zhang*[1*], *Chao-Han Huck Yang*[2], *Jun Qi*[3]

[1] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China
[2] Georgia Institute of Technology, Atlanta, USA
[3] Hong Kong Baptist University, Hong Kong SAR

ABSTRACT

We propose a theoretical analysis of quantum projection learning (QPL) that employs multiple kernels, highlighting its advantages through representation error analysis. Building upon previous studies that utilized a single quantum kernel-based method, we further investigate a quantum projection framework that incorporates multiple Gaussian kernels for low-resource spoken command recognition. Our empirical results align with our theoretical insights, suggesting that methods based on multiple kernels can further enhance the performance of QPL. By leveraging the quantum-to-classical projected output embeddings, we integrate this with a prototypical network for acoustic modeling. When evaluated using Arabic, Chuvash, Irish, and Lithuanian low-resource speech from CommonVoice, our proposed method surpasses the recurrent neural network and single kernel-based classifier baselines by an average of +5.28%.

Index Terms— quantum kernel projection, multiple kernel learning, low-resource speech classification

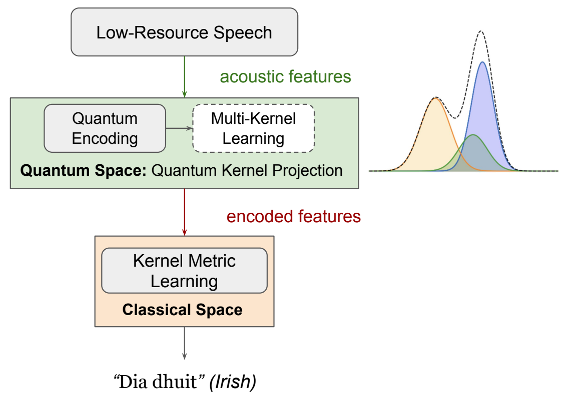

Fig. 1: **Computing diagram of quantum kernel learning**. Quantum kernel projection occurs on near-term quantum processing units (QPU) or simulators (e.g., CPU or TPU).

1. INTRODUCTION

Quantum machine learning (QML) [1–3] is a subfield of machine learning that harnesses the principles and capabilities of quantum computing to perform certain types of computations more efficiently than classical computers [4–6]. QML sits at the intersection of quantum computing and machine learning, and it aims to leverage quantum computing's advantages to solve machine learning problems more efficiently, such as quantum speech signal processing [7, 8] and quantum natural language processing [9]. Moreover, the development of quantum algorithms like quantum convolutional neural network [10] and quantum reinforcement learning [11, 12] holds promise for improving the performance of classical deep learning models. Qi *et al.* [7, 13] and Chen *et al.* [8] attempt to exploit hybrid quantum-classical neural networks, such as quantum convolutional neural network (QCNN) [10], for spoken command recognition and achieve competitive experimental results compared to classical machine learning counterparts. To further develop the use of QML algorithms for spoken command recognition, this work focuses on a new quantum projection learning (QPL) method with multiple Gaussian kernels.

Quantum kernel learning (QKL) [14] is an exciting area of research within QML. It focuses on leveraging quantum computing to enhance the efficiency and capabilities of kernel methods, which are fundamental in classical machine learning. By harnessing the unique properties of quantum computing, such as quantum parallelism and quantum entanglement, researchers aim to develop quantum kernels that can potentially outperform their classical counterparts in various machine learning tasks. For the spoken command recognition task, Chen *et al.* [15] initially investigates the deployment of QKL for spoken command recognition, where even better empirical performance has been attained than classical machine learning approaches.

In this work, we exploit a quantum multiple kernel learning (QMKL) and employ it for the application of spoken command recognition. QMKL is an extension of QKL to deal with the combination of multiple kernel functions in quantum computing for machine learning tasks. QMKL takes advantage of quantum computing's capabilities to efficiently compute combinations of multiple kernels, which can be challenging and computationally expensive in classical settings. By harnessing quantum algorithms and quantum data representations[1], QMKL aims to enhance the effectiveness of combining these

*Xiao-Lei Zhang is the corresponding author.

[1] A tutorial and guidelines for QKL can be found in https://www.tensorflow. org/quantum/tutorials/quantum_data

kernels to improve machine learning models. Additionally, the basic idea behind QMKL is to represent multiple kernels using quantum states and perform quantum operations to combine them effectively. This can lead to more expressive and powerful feature representations on machine learning tasks, potentially resulting in better predictive performance. Inspired by the quantum advantages of QMKL, in this work, we focus on using the QMKL approach for spoken command recognition.

2. QUANTUM KERNEL LEARNING

Quantum kernel learning (QKL) refers to a quantum version of kernel methods. A kernel method transforms each input to another vector in a high-dimensional vector space, which is known as the reproducing kernel Hilbert space (RKHS). The kernel method tries to learn a linear function in RKHS. Since the dimension of RKHS could be infinite, it enables the kernel method to own a powerful representation capability. Furthermore, a kernel trick is employed to allow for an efficient computation of the inner product between these high-dimensional vectors. Quantum kernel learning considers using the computation of kernel functions built upon quantum computers.

More specifically, given a quantum circuit map Φ and a training dataset $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$, for two feature data $\mathbf{x}_i$ and $\mathbf{x}_j$, a quantum kernel function k owns the form as:

$$k(\mathbf{x}_i, \mathbf{x}_j) := |\langle \Phi(\mathbf{x}_i) | \Phi(\mathbf{x}_j) \rangle|. \tag{1}$$

Through the representer theorem [16], we can express the classified function f as:

$$f(\mathbf{x}) = \sum_{n=1}^{N} \beta_n k(\mathbf{x}_n, \mathbf{x}), \tag{2}$$

where $\beta_i \in \mathbb{R}$ is an element of the N-dimensional vector $\boldsymbol{\beta}$ that requires to be estimated using an additional classical supervised learning method like support vector machine. The quantum kernel function k can be constructed by using a randomized quantum circuit as shown in Figure 1.

3. QUANTUM MULTIPLE KERNEL LEARNING

3.1. Quantum Multiple Kernel Learning Methodology

In recent years, multiple kernel learning (MKL) has been used to boost the expressive power of kernel machines. The idea of MKL is to use a combination of kernel functions instead of a single kernel function for learning. Similarly, QMKL applies multiple quantum kernel functions to construct the kernels, which is a linear combination of M different quantum kernels k_m. The mathematical form of quantum multiple kernel function $\hat{k}$ can be expressed as:

$$\hat{k}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{M} k_m(\mathbf{x}_i, \mathbf{x}_j; \theta_m), \tag{3}$$

where θ_m parameterizes each quantum kernel function.

Furthermore, the quantum multiple kernel function $\hat{k}$ can be generated by using the QKL architecture as shown in Figure 1 for k times. Then, the updated classified function f is

$$f(\mathbf{x}) = \sum_{n=1}^{N} \beta_n \hat{k}(\mathbf{x}_n, \mathbf{x}). \tag{4}$$

3.2. Theoretical Understanding of QMKL

Given the number of qubits Q and a prediction target function $h(\mathbf{x})$, we show that QMKL demonstrates a theoretical advantage by providing an upper bound as:

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{D}} |h(\mathbf{x}) - f(\mathbf{x})| \leq \mathcal{O}\left(\frac{2^Q}{\sqrt{N}M}\right), \tag{5}$$

where for a given arbitrary error ϵ, the amount of training data $N \propto \frac{2^{2Q}}{\epsilon^2}$. Compared to the approximation upper bound in [3], we attain a lower upper bound as Eq. (5) which is inversely proportional to the number of quantum kernels M.

3.3. QMKL for Spoken Command Recognition

Our theoretical results as Eq. (5) suggest that the enhancement of expressive power for a QKL is to employ multiple quantum kernels. Particularly, the estimation of quantum kernels in QMKL is not explicitly shown in the training process, while allowing for the quantum kernels to be fine-tuned in each training stage to attain better representation capability.

To associate with the the task of spoken command recognition, we use an angle encoding method to transform classical spoken data into quantum state features which have been discussed in [3, 15]. More specifically, we leverage the technique of quantum multiple kernels to map acoustic features like MFCC into latent feature embeddings by using quantum measurement projections. Furthermore, the generated feature embeddings are fed into a prototype network supervised by a kernel metric loss function. The optimization of QMKL parameters is related to the enhancement of the feature representation by maximizing the inter-class distances and also minimizing the inter-class similarities in the latent feature space, which facilities the subsequent kernel SVM for classification. In particular, we utilize three different Gaussian kernels, each of which is responsible for mapping the input features into high-dimensional space.

4. EXPERIMENTS AND RESULT ANALYSIS

4.1. Dataset

Our experiments of spoken command recognition are built upon the open Google Command Dataset [17]. The dataset consists of totally $11,165$ training examples and $6,500$ test data, respectively, which cover ten command classes like 'down', 'up', and etc. Moreover, white background noises are added to the original speech samples to simulate real scenarios. In particular, four low-resource language subsets, including Lithuanian [18], Arabic [19], and two other two languages from [20], are considered in the dataset to compare the performance of our methods on the low-resource tasks for spoken command recognition.

4.2. Experimental Setups

We first employ the tools Librosa [21] and Keras [22] to transform the original audios into MFCC features based on 60 bands and 1024-point FFT computation. Besides, we conduct our experiments on a classical GPU to simulate the empirical results of quantum machine learning models, which accounts for the fact that lower accuracies for the quantum baseline models could be probably attained in our experiments.

Our baseline models comprise recurrent neural network (RNN) with attention mechanism [23], QCNN combined with RNN (QCNN-RNN) [8], and the QKL approach [15]. To build up our QMKL architecture, we utilize the combinatorial quantum kernel projection in the stage of quantum data encoding, which maps the extracted MFCC features into high-dimensional quantum states in the Hilbert space. Furthermore, the metric learning method is applied to capture the encoded features from the post-quantum kernel projection. Furthermore, by taking the quantum states as discriminative features, we could take advantage of the classical SVM for multi-class classification of spoken command recognition, where a prototype loss function [24] is used to maximize an inter-class distance of the latent feature space.

Besides, Eq. (5) provides a mathematical analysis on approximating quantum multiple kernels using different combinations of kernel multiplications. However, the selection of kernel parameters in the composite kernel is still a challenging problem. In our experiments, we employed a non-universal quantum computing model known as Deterministic Quantum Computation with One Quantum Bit (DQC1) [25] to estimate the trace of the unitary operator, i.e., the composite kernel function, without the need for an explicit computation of each kernel. This approach also allows for the adjustment of composite kernel parameters to achieve the enhancement of expressiveness [26].

To obtain the optimal parameter set for the model, we allocated 80% of the data samples for training and 20% for testing. To mitigate the complexity of parameter optimization in QMKL, we utilized 50% of the training datasets. By employing various random seeds, the dataset is divided into 20 training and testing set instances. These twenty different splits were used for hyperparameter tuning. Subsequently, we computed the average and standard deviation of classification accuracy over 100 iterations of training and testing to acquire the set.

4.3. Experimental Results

4.3.1. English Spoken Command Recognition

Table 1 shows the empirical results of the proposed QMKL and the different classification baseline systems. Inspired by the classification models as discussed in 4.2, we employed cross-validation to mitigate the impact of data insufficiency, aiming to train models and attain a more competitively robust average accuracy.

As shown in Table 1, in comparison of the average classification accuracy on English speech commands across different training set sizes, our proposed QMKL exhibits higher accuracy than the results in the work [15], where a Gaussian-QKL is built with a single kernel. Moreover, our proposed QMKL consistently attains the best empirical performance, achieving the highest accuracy increase of 9.8% on $1k$ dataset.

Besides, as the amount of training data is progressively decremented to below $1k$, the experimental performance of QMKL witnesses a significant performance improvement. The baseline results of QMKL not only outperform QKL, but it also attains better performance than the deep learning methods based on RNNs and the hybrid quantum-classical algorithm QCNN-RNN in terms of average accuracy.

Furthermore, in the presence of limited training data, compared to the existing models such as RNNs, hybrid quantum RNN models, and QKL models, our proposed QMKL achieves more stable and superior validation accuracy.

Table 1: Experimental results of spoken command recognition in the context of different sizes of training data.

Number of training Utterances	500	1k	5k	11k
RNN [23]	38.6	57.5	84.9	**95.1**
QCNN-RNN [8]	42.2	63.8	79.6	93.2
Gaussian-QKL (reproduced)	45.4	67.6	82.8	93.4
QMKL (proposed)	**51.6**	**77.4**	**85.6**	94.0

Table 2: Experimental results of spoken command recognition on the tasks of four low-resource languages: Arabic, Chuvash, Irish, and Lithuanian.

Language	Utterances	Classes	Model	Accuracy(%)
Arabic (ar)	1600	16	RNN	66.5
			QCNN-RNN	67.1
			QKL [15]	69.1
			QMKL (proposed)	**70.5**
Chuvash (cv)	706	10	RNN [23]	17.6
			QCNN-RNN [8]	27.4
			QKL [15]	40.6
			QMKL (proposed)	**51.3**
Irish (ir)	1200	10	RNN [23]	43.3
			QCNN-RNN [8]	45.8
			QKL [15]	58.9
			QMKL (proposed)	**60.4**
lithuanian (it)	489	15	RNN [23]	45.1
			QCNN-RNN [8]	44.4
			QKL [15]	57.7
			QMKL (proposed)	**65.2**

4.3.2. An Visualization of the Classification Based on Multi-kernel and Single-kernel

Figure **??** illustrates the visualized impact of QKL and QMKL on the classification performance of the spoken command recognition. We utilized the Arabic dataset containing 1600 utterances distributed across 16 distinct classes. Initially, acoustical features undergo processing through quantum single-kernel and multi-kernel methods, resulting in latent feature embeddings based on the output measurement projections. These features are subsequently taken as inputs into a prototype network, with the prototype loss function as the optimization

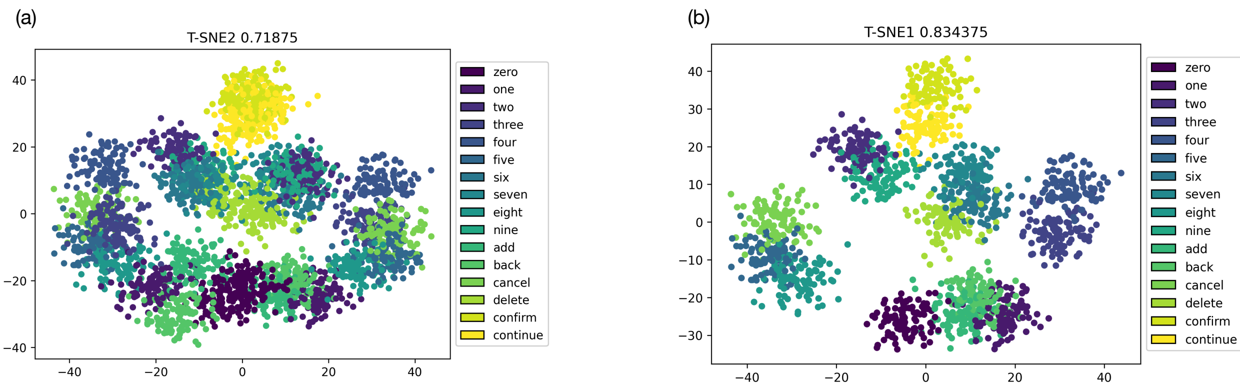

Fig. 2: **A visualization of the classification performance of QKL and QMKL for spoken command recognition.**. **a** A single quantum kernel is used; **b** Two quantum kernels are employ. The QMKL exhibits more discriminative capabilities among the classes. T-SNE is abbreviated as t-distributed Stochastic Neighbor Embedding, which is an unsupervised, non-linear technique primarily used for visualizing high-dimensional data.

objective during the training of kernel SVM. At this stage, we gather the features following the fully connected layer and employ T-SNE for visualizing the feature data in high-dimensional space. This process enables us to observe the low-dimensional manifolds even within the high-dimensional space.

From Figure 1, it is discernible that the latent feature embeddings collected after multi-kernel quantum processing exhibit a more compact representation on the low-dimensional plane compared to those derived from quantum single-kernel processing. The boundaries between different classes become more distinct, thereby substantially reducing category overlap, which suggests a diminished clustering tendency among data categories. Furthermore, we can interpret this as an indication that the Cluster Distances of feature points, post-multi-kernel processing, can be obviously reduced, which corresponds to an improved feature representation. Consequently, this can facilitate subsequent classification tasks in learning meaningful features with higher average classification accuracy.

4.3.3. Low-resource Spoken Command Recognition

Table 2 presents the classification average accuracies for the four low-resource languages. As shown in Table 2, it is evident that the classification task for the low-resource language 'cv' presents the most formidable challenge, with all models exhibiting notably low accuracy. Following this, the classification difficulty for languages 'ir' and 'it' ranks slightly lower than that of 'cv', with marginal disparities in accuracy.

When the number of utterances falls below 1000, we observe that RNN exhibits notably lower accuracy on 'cv' and 'it', at 17.6% and 45.1%,Conversely, QKL demonstrates relatively superior performance compared to RNNs and the hybrid quantum-classical model QCNN-RNN.

Additionally,in cases of limited training data for these four languages, our proposed MQKL consistently achieves higher accuracy than RNNs and QCNN-RNN.It surpasses the performance of the single-kernel QKL baseline, particularly excelling in the case of 'cv', where MQKL exhibits a remarkable 10.7% performance improvement over QKL.

4.4. Disscussion

To investigate the effectiveness of our quantum machine learning approach, we considered two classification tasks: the English-spoken command dataset and four low-resource spoken language datasets. Our simulation results suggest that QMKL surpasses a single quantum kernel.QMKL improved the average test accuracy by 9.8% and 10.7% on the English spoken command dataset and low resource spoken classification dataset, respectively.Thus, the related experimental results highlight the advantages of QMKL under both sufficient and limited training data, and they corroborate our theoretical analysis in Eq. (5).

5. CONCLUSION

In this work, we exploit the quantum multiple kernel learning as an approach for classifying low-resource spoken commands.Our method emphasizes three key points: (1) We provide a QMKL framewor for spoken command recognition, along with a theoretical upper bound to analyze its advantage over QKL when multiple kernels are used; (2) In low-resource spoken command recognition system, our proposed QMKL exhibits better experimental performance than both QKL and the existing models like RNN, QCNN-RNN and QKL; (3) Despite limited training data for languages,our proposed QMKL even achieves better empirical performance than the baseline results of the existing models.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation of China (NSFC) under Grant 62176211, and in part by the Project of the Science, Technology, and Innovation Commission of Shenzhen Municipality, China under Grant JCYJ20210324143006016 and JSGG20210802152546026.

7. REFERENCES

[1] M Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J Coles, "Challenges and Opportunities in Quantum Machine Learning," *Nature Computational Science*, vol. 2, no. 9, pp. 567–576, 2022.

[2] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, "The Power of Quantum Neural Networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

[3] Hsin-Yuan Huang et al., "Power of Data in Quantum Machine Learning," *Nature Communications*, vol. 12, no. 1, pp. 1–9, 2021.

[4] John Preskill, "Quantum Computing in the NISQ Era and Beyond," *Quantum*, vol. 2, pp. 79, 2018.

[5] Jun Qi, Chao-Han Huck Yang, Pin-Yu Chen, and Min-Hsiu Hsieh, "Theoretical Error Performance Analysis for Variational Quantum Circuit Based Functional Regression," *npj Quantum Information*, vol. 9, no. 1, pp. 4, 2023.

[6] Jun Qi, Xiao-Lei Zhang, and Javier Tejedor, "Optimizing Quantum Federated Learning Based on Federated Quantum Natural Gradient Descent," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2023, pp. 1–5.

[7] Jun Qi and Javier Tejedor, "Classical-to-Quantum Transfer Learning for Spoken Command Recognition Based on Quantum Neural Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 8627–8631.

[8] Chao-Han Huck Yang, Jun Qi, Samuel Yen-Chi Chen, Pin-Yu Chen, Sabato Marco Siniscalchi, Xiaoli Ma, and Chin-Hui Lee, "Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 6523–6527.

[9] Chao-Han Huck Yang, Jun Qi, Samuel Yen-Chi Chen, Yu Tsao, and Pin-Yu Chen, "When BERT Meets Quantum Temporal Convolution Learning for Text Classification in Heterogeneous Computing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022.

[10] Iris Cong, Soonwon Choi, and Mikhail D Lukin, "Quantum Convolutional Neural Networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 2019.

[11] Mateusz Ostaszewski, Lea M Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko, "Reinforcement Learning for Optimization of Variational Quantum Circuit Architectures," *arXiv preprint arXiv:2103.16089*, 2021.

[12] Samuel Yen-Chi Chen, Chao-Han Huck Yang, Jun Qi, Pin-Yu Chen, Xiaoli Ma, and Hsi-Sheng Goan, "Variational Quantum Circuits for Deep Reinforcement Learning," *IEEE Access*, vol. 8, pp. 141007–141024, 2020.

[13] Jun Qi, Chao-Han Huck Yang, and Pin-Yu Chen, "QTN-VQC: An End-to-End Learning Framework for Quantum Neural Networks," in *NeurIPS 2021 Workshop on Quantum Tensor Networks in Machine Learning*, 2022.

[14] Carsten Blank, Daniel K Park, June-Koo Kevin Rhee, and Francesco Petruccione, "Quantum Classifier with Tailored Quantum Kernel," *npj Quantum Information*, vol. 6, no. 1, pp. 41, 2020.

[15] Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Tara N Sainath, Sabato Marco Siniscalchi, and Chin-Hui Lee, "A Quantum Kernel Learning Approach to Acoustic Modeling for Spoken Command Recognition," in *Prof. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[16] Bernhard Schölkopf and Alexander J Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT press, 2002.

[17] P. Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," in *arXiv preprint*, 2018, p. arXiv:1804.03209.

[18] Aliaksei Kolesau and Dmitrij Sesok, "Unsupervised Pretraining for Voice Activation," in *Applied Sciences*, 2020.

[19] L. Benamer and O. Alkishriwo, "Database for Arabic Speech Commands Recognition," in *Proc. CEST*, 2020.

[20] R. Ardila, M. Branson, K. Davis, M. Kohler, J. Meyer, R. Morais M. Henretty, L. Saunders, F. Tyers, and G. Weber, "Common Voice: A Massively-multilingual Speech Corpus," *Proceedings of the 12th Language Resources and Evaluation Conference*, vol. pp. 42184222, 2020.

[21] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "Librosa: Audio and Music Signal Analysis in Python," *Proceedings of the 14th Python in Science Conference*, vol. 8, pp. pp. 1825, 2015.

[22] K. Choi, D. Joo, and J. Kim, "Kapre: On-GPU Audio Preprocessing Layers for A Quick Implementation of Deep Neural Network Models with Keras," *Machine Learning for Music Discovery Workshop, ICML,*, 2017.

[23] D. C. De Andrade, S. Leo, M. L. D. S. Viana, and C. Bernkopf, "A Neural Attention Model for Speech Command Recognition," in *arXiv preprint*, 2018.

[24] Jaesung Huh, Minjae Lee, Heesoo Heo, Seongkyu Mun, and Joon Son Chung, "Metric learning for keyword spotting," in *IEEE Spoken Language Technology Workshop*. IEEE, 2021, pp. 133–140.

[25] Emanuel Knill and Raymond Laflamme, "Power of One Bit of Quantum Information," *Physical Review Letters*, vol. 81, no. 25, pp. 5672, 1998.

[26] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao, "Expressive Power of Parametrized Quantum Circuits," *Physical Review Research*, vol. 2, no. 3, pp. 033125, 2020.