

Classical-to-quantum convolutional neural network transfer learning

Juhyeon Kim^a, Joonsuk Huh^{a,b,c}, Daniel K. Park^{d,e,*}

^a SKKU Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon, Republic of Korea

^b Department of Chemistry, Sungkyunkwan University, Suwon, Republic of Korea

^c Institute of Quantum Biophysics, Sungkyunkwan University, Suwon, Republic of Korea

^d Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea

^e Department of Statistics and Data Science, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Communicated by J. Wu

Keywords:

Quantum computing

Quantum machine learning

Quantum convolutional neural network

Transfer learning

ABSTRACT

Machine learning using quantum convolutional neural networks (QCNNs) has demonstrated success in both quantum and classical data classification. In previous studies, QCNNs attained a higher classification accuracy than their classical counterparts under the same training conditions in the few-parameter regime. However, the general performance of large-scale quantum models is difficult to examine because of the limited size of quantum circuits, which can be reliably implemented in the near future. We propose transfer learning as an effective strategy for utilizing small QCNNs in the noisy intermediate-scale quantum era to the full extent. In the classical-to-quantum transfer learning framework, a QCNN can solve complex classification problems without requiring a large-scale quantum circuit by utilizing a pre-trained classical convolutional neural network (CNN). We perform numerical simulations of QCNN models with various sets of quantum convolution and pooling operations for MNIST data classification under transfer learning, in which a classical CNN is trained with Fashion-MNIST data. The results show that transfer learning from classical to quantum CNN performs considerably better than purely classical transfer learning models under similar training conditions.

1. Introduction

Machine learning (ML) with a parameterized quantum circuit (PQC) is a promising approach for improving existing methods beyond classical capabilities [1–7]. This is a classical-quantum hybrid algorithm in which the cost function and its corresponding gradient are computed using quantum circuits [8,9] and the model parameters are updated classically. Such hybrid ML models are particularly advantageous when cost function minimization is difficult to perform classically [4,10,11]. These models optimize the quantum gate parameters under the given experimental setup, and hence can be robust to systematic errors. Furthermore, they are less prone to decoherence because iterative computation can be exploited to reduce the quantum circuit depth. Thus, the hybrid algorithm has the potential to achieve quantum advantage in solving various problems in the noisy intermediate-scale quantum (NISQ)¹ era [12,13].

A critical challenge in the utilization of PQC for solving real-world problems is the barren plateau phenomenon in the optimization landscape, which makes training the quantum model that samples from the

Haar measure difficult as the number of qubits increases [14]. One way to avoid the barren plateau is to adopt a hierarchical structure [15,16], in which the number of qubits decreases exponentially with quantum circuit depth, such as in quantum convolutional neural networks (QCNNs) [4]. The hierarchical structure is interesting from a theoretical perspective because of its close connection to tensor networks [15,17]. Moreover, the shallow depth of a QCNN, which grows logarithmically with the number of input qubits, makes it well suited for NISQ computing. In addition, an information-theoretic analysis shows that the QCNN architecture can help reduce the generalization error [18], which is one of the central goals of machine learning. All these factors motivate the application of QCNN for machine learning. QCNNs have been shown to be useful for solving both quantum [4,19] and classical [20] problems despite their restricted structure with a shallow-depth quantum circuit. In Ref. [20], for binary classification with the MNIST [21] and Fashion-MNIST [22] datasets, QCNN yielded higher classification accuracy than the classical convolutional neural network (CNN) when only 51 or fewer parameters were used to construct these models. The best-known

* Corresponding author at: Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea.

E-mail addresses: joonsukhuh@gmail.com (J. Huh), dkd.park@yonsei.ac.kr (D.K. Park).

¹ NISQ refers to the domain of quantum computing where the number of quantum processors that can be manipulated reliably is limited due to noise, yet holds the potential to surpasses the classical capabilities to a certain extent. As NISQ technology becomes increasingly accessible, the discovery of its real-world applications has become crucially important.

classical CNN-based classifiers for the same datasets typically employ millions of parameters. However, the size of the quantum circuits that can be implemented with current quantum devices is too small to incorporate such a large number of parameters. Therefore, two important issues remain. The first is to verify whether a QCNN can continue to outperform its classical counterpart as the number of trainable model parameters increases. The second is to utilize small QCNNs that can be realized in the near future to the full extent, so that a quantum advantage can be achieved in solving practical problems. The latter is the main focus of this work.

An ML problem for which the quantum advantage in the few-parameter regime can be exploited is transfer learning (TL) [23–26]. TL aims to utilize what has been learned in one setting to improve generalization in another setting that is independent of the former. TL can be applied to classical-quantum hybrid networks such that the parameters learned for a classical model are transferred to training a quantum model or vice versa [27]. In the classical-to-quantum (C2Q) TL scheme, the number of qubits increases with the number of output nodes (or features) of the pre-trained classical neural network. This indicates that the transferred part of a classical neural network should have a small number of output nodes to find applications in the NISQ era. For example, using a pre-trained feedforward neural network with a large number of nodes throughout the layers would not be well suited for near-term hybrid TL. By contrast, building a TL model with a classical and quantum CNN is viable because the number of features in the CNN progressively decreases via subsampling (i.e., pooling), and the QCNN has already exhibited an advantage with a small number of input qubits.

Motivated by the aforementioned observations, we propose a TL framework for classical-to-quantum convolutional neural networks (C2Q-CNNs). Unlike previous works, C2Q-CNN transfers knowledge from a pre-trained (source) classical CNN to a quantum CNN, thereby preserving the benefits of quantum CNN. Our method avoids the need for classical data dimensionality reduction, commonly required in existing methods, as the classical CNN serves as the source for TL. Additionally, we introduce new ansatzes for both quantum pooling and convolutional operations, enriching the model selection in the QCNN. To evaluate the performance of C2Q-CNN, we conduct numerical simulations on the MNIST data classification task using PennyLane [28]. The classical CNN is pre-trained on the Fashion-MNIST dataset. The simulations assess the classification accuracy under different quantum convolution and pooling operations and compare C2Q-CNN with various classical-to-classical CNN (C2C-CNN) TL schemes. The results show that C2Q-CNN outperforms C2C-CNN with respect to classification accuracy under similar training conditions. Furthermore, the new quantum pooling operation developed in this work is more effective in demonstrating the quantum advantage.

The remainder of this paper is organized as follows. Section 2 reviews QCNNs and TL. This section also introduces the generalization of the pooling operation of the QCNN. Section 3 explains the general framework for C2Q-CNN TL. The simulation results are presented in Section 4. MNIST data classification was performed with a CNN pre-trained for Fashion-MNIST data, and the performance of the C2Q-CNN models was compared with that of various C2C-CNN models. The conclusions and outlook are presented in Section 5.

2. Preliminaries

2.1. Quantum convolutional neural network

Quantum convolutional neural networks are parameterized quantum circuits with unique structures inspired by classical CNNs [4, 20]. In general, QCNNs follow two basic principles of classical CNNs: translational invariance of convolutional operations and dimensionality reduction via pooling. However, QCNNs differ from classical CNNs in several aspects. First, the data are defined in quantum Hilbert space,

which grows exponentially with the number of qubits. Consequently, the quantum convolutional operation is not an inner product, as in the classical case, but a unitary transformation of a state vector, which is a linear map that transforms a vector to a vector, whereas a classical convolution operation is a linear map that transforms a vector to a scalar. The pooling in QCNN traces out half of the qubits, similar to the pooling in the CNN that subsamples the feature space. Typically, the pooling layer includes parameterized two-qubit controlled-unitary gates, and the control qubits are traced out after the gate operations. Without loss of generality, we refer to the structure of a parameterized unitary operator for either convolution or pooling as ansatz. The cost function of a model with given parameters is defined with an expectation value of some observable with respect to the final quantum state obtained after repeating quantum convolutional and pooling operations. The QCNN is trained by updating the model parameters to minimize the cost function until a pre-determined convergence condition is met. The general concept of a QCNN is illustrated in Fig. 1(a). An example of a circuit with eight input qubits is shown in (b). The depth of the QCNN circuit after repeating the convolution and pooling until one qubit remains is $O(\log N)$, where N is the number of input qubits. This shallow depth allows the QCNN to perform well on quantum hardware that will be developed in the near future.

The quantum convolution and pooling operations can be parameterized in many ways. The convolution ansatzes evaluated in this study are illustrated in Fig. 2. Among them, circuits (b) to (j) are the nine ansatzes previously tested in Ref. [20]. These ansatzes are motivated by past studies. For instance, circuit (b) is a parameterized quantum circuit that was used to train a tree tensor network (TTN) [15]. The four-qubit parameterized quantum circuits analyzed by Sim et al. [29] were modified to two-qubit circuits to serve as building blocks for the convolution layer, resulting in circuits (c), (d), (e), (f), (h), and (i). Circuits (h) and (i) are two-qubit versions of the circuits with the best expressibility, while circuit (c) is of the best entangling circuit. Circuits (d), (e), and (f) represent a good balance of expressibility and entangling capability. Circuit (g) is used for the two-body variational quantum eigensolver entangler [30] and can generate arbitrary $SO(4)$ gates [31], making it a suitable candidate for building convolution layer in QCNN. Circuit (j) is a parameterized arbitrary $SU(4)$ gate [19,32] capable of performing arbitrary two-qubit unitary operations. Because the convolutional operations act on two qubits, parameterized $SU(4)$ operations provides the most general ansatz. In this study, we introduce two new convolutional ansatzes, (a) and (k), to our benchmark. The former aims to study the classification capability of a QCNN when only pooling operations are trained. The latter is inspired by the generalized pooling operation described in the following paragraph, with an $SU(2)$ gate applied to a control qubit to split the subspaces in an arbitrary superposition.

The pooling ansatzes used in previous studies were simple single-qubit-controlled rotations followed by tracing out the control qubit. For example, in Ref. [20], a pooling operation in the following form was used:

$$\text{Tr}_A \left[(|1\rangle\langle 1|_A \otimes R_z(\theta_1)_B + |0\rangle\langle 0|_A \otimes R_x(\theta_2)_B) \rho_{AB} U_p^\dagger \right], \quad (1)$$

where $\text{Tr}_A(\cdot)$ represents a partial trace over subsystem A , $R_i(\theta)$ is the rotation around the i axis of the Bloch sphere by an angle of θ , θ_1 and θ_2 are the free parameters, ρ_{AB} is a two-qubit state subject to pooling, and U_p^\dagger is the conjugate transpose of the unitary gate for pooling. The pooling operation in Eq. (1) is referred to as ZX pooling. In addition to ZX pooling, generalized pooling is introduced as

$$\text{Tr}_A \left[(|1\rangle\langle 1|_A \otimes U(\theta_1, \phi_2, \lambda_3)_B + |0\rangle\langle 0|_A \otimes U(\theta_4, \phi_5, \lambda_6)_B) \rho_{AB} U_p^\dagger \right]. \quad (2)$$

Here, $U(\theta, \phi, \lambda) = R_z(\phi)R_x(-\pi/2)R_z(\theta)R_x(\pi/2)R_z(\lambda)$ and can implement any unitary operator in $SU(2)$. Again, U_p^\dagger is the conjugate transpose of the corresponding unitary gate for pooling. The unitary gates used in ZX pooling and generalized pooling are compared in Fig. 3.

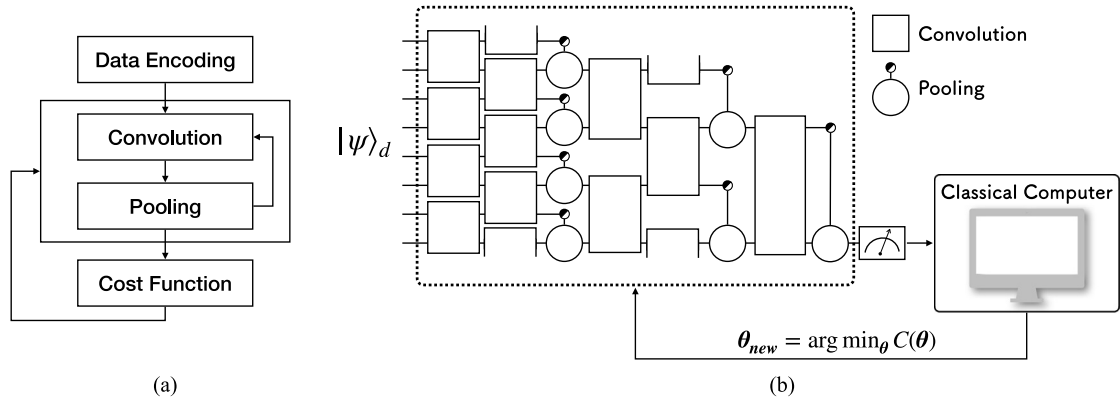


Fig. 1. (a) Schematics of the QCNN algorithm with (b) an example for eight input qubits. Given a quantum state, $|\psi\rangle_d$, which encodes classical data, the quantum circuit comprises two parts: convolutional filters (rectangles) and pooling (circles). The convolutional filter and pooling use parameterized quantum gates. Three layers of convolution-pooling pairs are presented in this example. In each layer, the convolutional filter applies the identical two-qubit ansatz to the nearest neighbor qubits in a translationally invariant manner. The quantum convolutional operations in the QCNN circuits are designed to meet the closed boundary condition, as indicated by the open-ended gates in the figure, ensuring the top and bottom qubits in each layer are connected. Pooling operations within the layer are identical to each other, but differ from convolutional filters. The pooling operation is represented as a controlled unitary transformation, and the half-filled circle on the control qubit indicates that different unitary gates can be applied to each subspace of the control qubit. The measurement outcome of the quantum circuit is used to calculate the user-defined cost function. A classical computer is used to compute the new set of parameters based on the gradient, and the quantum circuit parameters are updated for the subsequent round. The optimization process is iterated until pre-selected conditions are met.

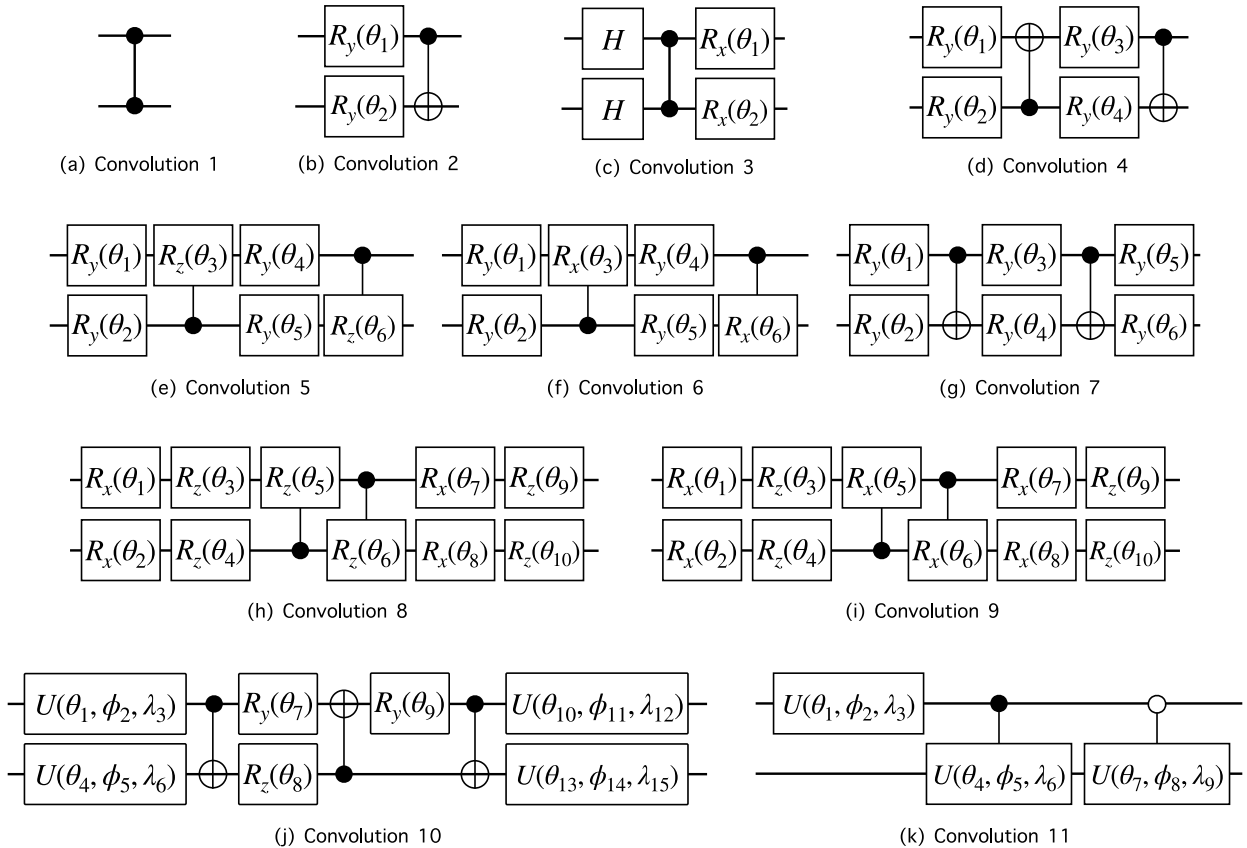


Fig. 2. Parameterized quantum circuits used in the convolutional layer. The convolutional circuits from (b) to (j) are adapted from Ref. [20], whereas (a) and (k) are the new convolutional circuits tested in this study. $R_i(\theta)$ is the rotation around the i -axis of the Bloch sphere by an angle of θ , and H is the Hadamard gate. $U(\theta, \phi, \lambda)$ is an arbitrary single-qubit gate, which can be expressed as $U(\theta, \phi, \lambda) = R_z(\phi)R_x(-\pi/2)R_z(\theta)R_x(\pi/2)R_z(\lambda)$. $U(\theta, \phi, \lambda)$ can implement any unitary operation in $SU(2)$. As (j) can express an arbitrary two-qubit unitary gate, we test it without any parameterized gates for pooling in addition to ZX pooling and generalized pooling. For (k), we do not apply parameterized gates for pooling. In these cases, pooling simply traces out the top qubit after convolution.

2.2. Transfer learning

Transferring the knowledge accumulated from one task to another is a typical intelligent behavior that human learners always experience. TL refers to the application of this concept in ML. Specifically, TL aims to improve the training of a new ML model by utilizing a reference (or

source) ML model that is pre-trained for a different but related task with a different dataset [23–26]. Transfer learning encompasses three main categories: inductive transfer learning (ITL), transductive transfer learning (TTL), and unsupervised transfer learning (UTL) [24,33]. ITL applies when label information is available for the target domain, while TTL applies when label information is only available for the source

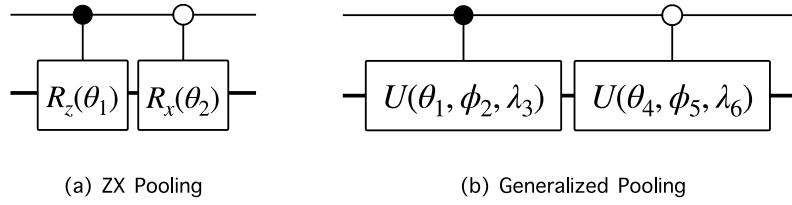


Fig. 3. Parameterized quantum gates used in the pooling layer. The pooling circuit (a) is adapted from Ref. [20], and (b) is the generalized pooling method introduced in this work. Generalized pooling applies two arbitrary single-qubit unitary gate rotations, $U(\theta_1, \phi_2, \lambda_3)$ and $U(\theta_4, \phi_5, \lambda_6)$, which are activated when the control qubit is 1 (filled circle) or 0 (open circle), respectively. The control (first) qubit is traced out after the gate operations to reduce the dimensions. The single-qubit unitary gate is defined as $U(\theta, \phi, \lambda) = R_z(\phi)R_x(-\pi/2)R_z(\theta)R_x(\pi/2)R_z(\lambda)$, and it can implement any unitary in $SU(2)$. The thinner horizontal line (top qubit) indicates the qubit that is being traced out after gate operations.

domain. UTL, on the other hand, applies when label information is unavailable for both the source and target domains. In this study, we have chosen to focus exclusively on ITL to ensure simplicity and clarity in our explanations and demonstrations. Henceforth, when we refer to transfer learning, it pertains specifically to ITL. Detailed information regarding our numerical simulations will be presented later in the manuscript.

TL is known to be particularly useful for training a deep learning model that takes a long time owing to the large amount of data, especially if the features extracted in early layers are generic across various datasets. In such cases, starting from a pre-trained network such that only a portion of the model parameters is fine-tuned for a particular task can be more practical than training the entire network from scratch. For example, suppose that a neural network is trained with data A to solve task A and finds the set of parameters (i.e., weights and biases) $\mathbf{w}_A \in \mathbb{R}^{N_A}$. To solve task B given dataset B , the neural network is not trained from scratch, as this may require vast computational resources. Instead, the parameters associated with some of the earlier layers of the reference neural network are used as a set of fixed parameters for the new neural network that is subjected to solving task B with data B . In other words, some elements of the parameters for this new learning problem, denoted by $\mathbf{w}_B \in \mathbb{R}^{N_B}$, are identical to those of \mathbf{w}_A . Hence the number of parameters subject to new optimization is less than N_B . The successful application of TL can improve training performance by starting from a higher training accuracy, achieving a faster rate of accuracy improvement, and converging to a higher asymptotic training accuracy [34].

The aforementioned observations imply that TL is also beneficial when the amount of data available is insufficient to work with or extremely small to build a good model. Because processing big data in the NISQ era will be challenging, working with small amounts of data through TL is a promising strategy for near-term quantum ML. The target ML model subjected to fresh training (i.e., fine-tuning) in TL typically has a much smaller number of parameters than the pre-trained model. This and the success of QCNN in the few-parameter regime together promote the development of the classical-to-quantum CNN transfer learning.

3. Classical-to-quantum transfer learning

An extension of TL to quantum ML was proposed, and its general concept was formulated in Ref. [27]. Although the performances of the quantum models were not compared with those of their classical counterparts, three different scenarios of quantum TL, namely C2Q, quantum-to-classical, and quantum-to-quantum, were shown to be feasible. Among these three possible scenarios, we focus on C2Q TL as mentioned in Section 1, because we aim to utilize QCNNs in the few-parameter regime to the full extent. Sufficient reduction of the data dimensionality (i.e., the number of attributes or features) by classical learning would ensure that the size of a quantum circuit subject to training is sufficiently small for implementation with NISQ devices. The dimensionality reduction technique is also necessary to simplify

expensive quantum state preparation routines to represent classical data in a quantum state [35–43].

C2Q TL has been utilized for image data classification [27,44] and spoken command recognition [45]. These works serve as proof of principle for the general idea and present interesting examples to motivate further investigations and benchmarks. The parameterized quantum circuits therein are vulnerable to the barren plateau problem, because they follow the basic structure of a fully connected feedforward neural network with the same number of input and output qubits. Moreover, these studies used classical neural networks to significantly reduce the number of data features to only four or eight. This means that most of the feature extraction is performed classically; hence, the necessity of the quantum part is unclear. These studies encode the reduced data onto a quantum circuit using simple single-qubit rotations, also known as qubit encoding [15,42], which makes the number of model parameters grow polynomially with the number of data features. In contrast, the number of model parameters in our ML algorithm scales logarithmically with the number of input qubits. Furthermore, all of these works use only one type of ansatz based on repetitive applications of single-qubit rotations and controlled-NOT gates. Finally, the performance of C2Q TL was not compared with that of the C2C version in any of these studies. Because the pre-trained classical neural network performs a significant dimensionality reduction (and hence feature extraction), the absence of a direct comparison with C2C TL raises the question of whether the quantum model achieves any advantage over its classical counterparts.

In this study, we present a classical-to-quantum transfer learning framework with QCNN. Our framework facilitates the transfer of knowledge from a pre-trained classical CNN to a quantum CNN, leveraging the unique advantages offered by QCNNs. The adoption of QCNNs as the target model holds crucial importance for several reasons. Firstly, QCNNs possess the capability to circumvent the barren plateau effect, a critical bottleneck encountered during the training of quantum neural networks. This property of QCNNs addresses a major challenge in quantum machine learning and enhances the training process. Furthermore, previous research has demonstrated the advantages of QCNNs over their classical counterparts, particularly in the few-parameter regime, along with their good generalization capabilities. As a result, fine-tuning a machine learning model using a QCNN is expected to yield enhanced classification performance compared to fine-tuning with a traditional CNN. To illustrate the practical implementation of our C2Q TL framework, we provide an example of transfer learning using C2Q-CNN, which serves as the basis for our benchmark studies. The schematic representation of this process can be found in Fig. 4, showcasing the application of our proposed framework.

The general model is flexible with the choice of data encoding, which loads classical data features to a quantum state $|\psi\rangle_d$, and ansatz, the quantum circuit model subject to training. We performed extensive benchmarking over the various ansatzes presented in Section 2.1 to classify MNIST data using a classical model pre-trained with Fashion-MNIST data. Finally, we compared the classification accuracies of C2Q and various C2C models. The C2Q models performed noticeably better than all C2C models tested in this study under similar training conditions. More details on the simulation and results are presented in the following section.

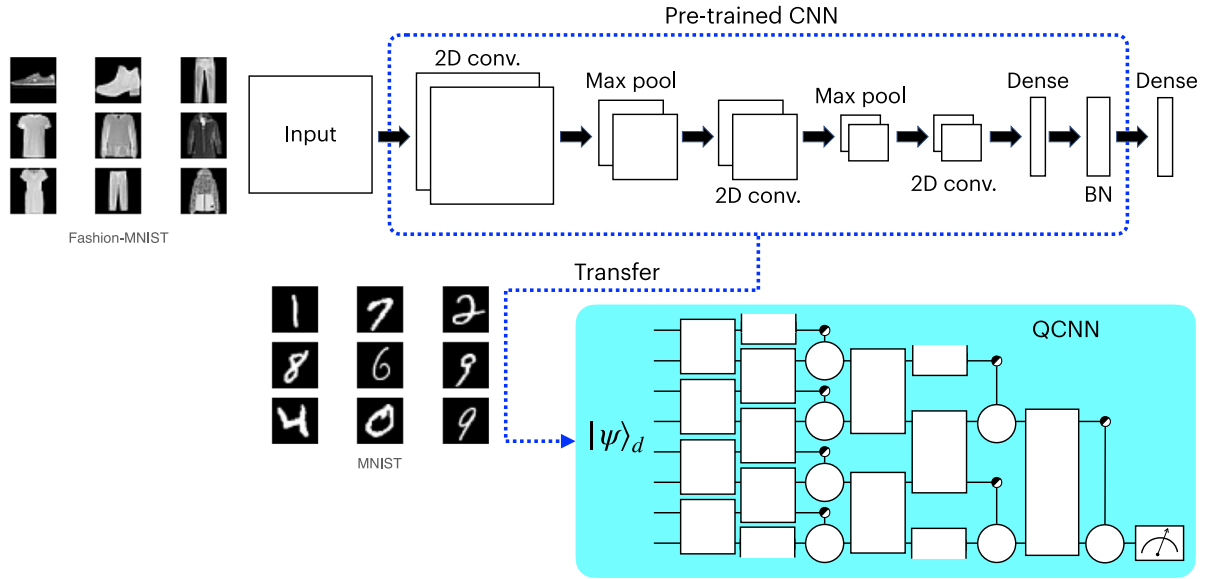


Fig. 4. An example of classical-to-quantum convolutional neural network transfer learning simulated in this work for benchmarking and comparison to purely classical models. A source CNN is trained on the Fashion-MNIST dataset. Then, the transfer learning trains a QCNN for MNIST data classification by utilizing the earlier layers of the pre-trained CNN for feature extraction. The source CNN contains convolution (conv.), pooling, dense and batch normalization (BN) layers.

4. Simulation results

To demonstrate the advantage of C2Q-CNN, we performed classical simulations of binary classification using PennyLane [28]. The benchmark was performed using two standard image datasets, MNIST and Fashion-MNIST, which were accessed through Keras [46]. Examples of the datasets are shown on the left in Fig. 4. Note that both datasets have 28×28 features and 10 classes. Among the 10 classes of MNIST data, we performed three independent binary classification tasks aimed at distinguishing between 0 and 1, between 2 and 3, and between 8 and 9. To represent classical data as a quantum state in a QCNN, the classical data must be encoded into a quantum state. The number of data features that can be encoded in N qubits ranges from N to 2^N depending on the choice of the encoding method [15,40,42,43,47]. Among many options, we used amplitude encoding to represent as many features as possible (see Appendix A). All C2Q-CNN simulations were performed with eight input qubits, to which the amplitude encoding loads 256 features. Quantum feature maps that encode only eight features, such as qubit encoding, are not considered because they require extreme dimensionality reduction on the classical end, which may dominate the classification result as described in the previous section. Furthermore, amplitude encoding was shown to work well with QCNNs for classical data classification [20].

The source classical CNN model, depicted in Fig. 4, was trained on 60,000 Fashion-MNIST data for multinomial classification. The TL is implemented by replacing the final dense layer of the pre-trained CNN with a QCNN. The pre-trained CNN is utilized as the source model to perform binary classification on the MNIST dataset. The source model takes the MNIST data as input and outputs 256 values using the parameters pre-trained on the Fashion-MNIST dataset. These output values are then encoded into 8 input qubits of the QCNN, which is fine-tuned using 10,000 MNIST data samples. The fine-tuning process adjusts the parameters of the QCNN to optimize its performance on the binary classification task. The trainable parameters in the QCNN were optimized by minimizing the cross-entropy cost function with the Adam optimizer [48] using PennyLane [28]. The number of MNIST test data was approximately 2000.

C2Q transfer models can be split into three sets based on different pooling variations. The first set uses ZX pooling with convolution circuits (a)–(j), as shown in Fig. 2. The second set includes the generalized pooling circuit and convolution circuits (a)–(j), as shown in

Fig. 2. Finally, we constructed transfer models without parameterized quantum gates in pooling layers. We refer to this pooling strategy that merely traces out one of the qubits as trivial pooling. Trivial pooling is tested with convolution circuits (j) and (k), as shown in Fig. 2.

To compare C2Q-CNN TL classification results with its classical counterparts, 1D and 2D CNN C2C TL models were constructed with a similar number of trainable parameters as C2Q-CNN models. The 1D CNN model was composed of a 1D convolution layer and 1D max pooling layer with 64 trainable parameters. Similarly, the 2D CNN model was composed of a 2D convolution layer and 2D max pooling layer with 76 trainable parameters. The CNNs subjected to fine-tuning for the MNIST data use the cross-entropy cost function with the Adam optimizer, as in the C2Q-CNN case. These classical CNN architectures are built using Keras [46]. Detailed descriptions of C2C models are provided in Appendix B. The training process used mini-batch gradient descent with a batch size of 50 and a learning rate of 0.01. We also fixed the number of training iterations at 200 for the C2Q TL and C2C TL models. The other training conditions were kept the same in the C2Q and C2C transfer models to make the comparison as fair as possible.

The TL classification results are shown in Fig. 5. Each bar represents the C2Q classification accuracy averaged over ten randomly initialized parameters. Different bars along the x-axis represent different convolutions, labeled according to Fig. 2. The unfilled, filled, and hatched bars represent the results of ZX pooling, generalized pooling, and trivial pooling, respectively. The blue dashed line and green solid line represent the results of the C2C TL using 1D and 2D CNN architectures, respectively.

The 0 vs. 1 classification results are shown in Fig. 5(a). For ZX and generalized pooling, most of the average test accuracies were greater than 95% and 98%, respectively. The test accuracy with ZX pooling with convolution 2, 4, 7, 8, 9, and 10 ansatz and generalized pooling with all convolution ansatz is greater than that of 1D and 2D C2C TL. The accuracy with trivial pooling with convolution 10 ansatz was also higher than that of C2C TL. In addition, the accuracy of trivial pooling with convolution ansatz 10 is higher than that of trivial pooling with convolution ansatz 11. This can be attributed to the fact that the ansatz 10 has more parameters and is capable of expressing an arbitrary $SU(4)$ including the ansatz 11. The test accuracy of generalized pooling is greater than that of ZX pooling when the convolution ansatz is the same. This can be inferred from the fact that generalized pooling can be trained to learn ZX pooling.

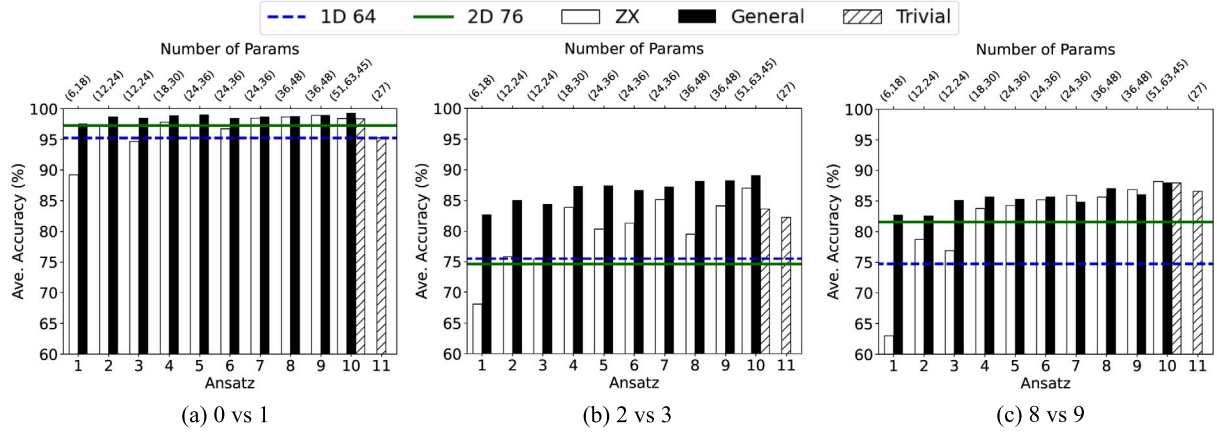


Fig. 5. Summary of the classification results with PennyLane simulations (quantum part) and Keras (classical part). Each bar represents the classification test accuracy of C2Q TL averaged over 10 instances given by the random initialization of parameters. The different bars along the x-axis indicate that the results are for different convolution ansatz, labeled according to Fig. 2. The unfilled, filled, and hatched bars represent the results of ZX pooling, generalized pooling, and trivial pooling, respectively. The number of trainable model parameters for each case is shown at the top of the x-axis. The horizontal lines represent the results of the C2C TL with 1D and 2D CNN architectures. The number of trainable model parameters for each case is provided in the legend.

The 2 vs. 3 classification results are shown in Fig. 5(b). Most of the ZX pooling average accuracies were between 70% and 90%, and most of the generalized pooling average accuracies were between 85% and 90%. The test accuracies in (b) are lower than those of (a) because 2 vs. 3 image classifications are more difficult than 0 vs. 1 image classifications. All C2Q TL classification accuracies are higher than the C2C classification accuracies except for ZX pooling with convolution 1 and 3, which use a much smaller number of parameters than the purely classical TL. The test accuracy of generalized pooling is greater than that of ZX pooling when the convolution is the same. In addition, the accuracy of trivial pooling with convolution ansatz 10 is higher than that with ansatz 11. These results are consistent with those obtained from the 0 vs. 1 classification, providing further evidence that the ansatz with more model parameters and improved expressibility is favored. This suggests that the use of more complex models, such as the generalized pooling, is advantageous in solving classification problems, particularly in comparison to less expressive models such as ZX pooling.

The 8 vs. 9 classification results are shown in Fig. 5(c). Most of the ZX pooling average accuracies were between 70% and 90%, and most of the generalized pooling average accuracies were between 85% and 90%. The test accuracies in (c) are lower than those in (a) because 8 vs. 9 image classification is more difficult than 0 vs. 1 image classification, but the accuracies in (c) are similar to those in (b). All C2Q TL classification accuracies are higher than the C2C TL classification accuracies except for ZX pooling with convolution 1, 2, and 3, which use a much smaller number of parameters than the purely classical TL. As before, trivial pooling with convolution ansatz 10 has higher accuracy than trivial pooling with ansatz 11. the accuracy of generalized pooling is consistently higher than that of ZX pooling when the underlying convolution is identical. However, there are instances where the performance of ZX pooling is comparable to that of generalized pooling, such as with the use of convolution ansatz 7. Nevertheless, a Welch's t-test analysis [49] revealed that these results are not statistically significant. Based on these findings, we can conclude that generalized pooling is a more favorable approach for solving all of the classification problems tested, compared to ZX pooling.

The results in Fig. 5(a), (b), and (c) show a QCNN's tendency to perform better when the convolution circuits have a larger number of trainable parameters. However, simply increasing the number of trainable parameters does not always guarantee to improve the test accuracy because it is affected by various conditions, such as statistical error and quantum gate arrangement. For example, ZX pooling with convolution ansatz 5, 6, 7 in (b) has the same number of trainable

parameters, but their average accuracies are different. In the current study, overfitting was not observed as the number of model parameters was much smaller than the number of data, but it is important to keep this issue in mind when designing and implementing larger models in the future.

In summary, generalized pooling mostly produces higher classification accuracy than ZX pooling with the same convolution circuit. This is as expected since the generalized pooling has more model parameters. Moreover, it can be reduced to ZX pooling under appropriate parameter selection. The accuracy of all ZX pooling, generalized pooling, and trivial pooling circuits tends to be higher when the convolution circuits have a larger number of gate parameters. Although C2Q models have fewer trainable parameters than C2C models, most C2Q models outperform C2C models.

To validate our findings, we conducted a Welch's t-test analysis [49] to determine the statistical significance of the improved classification results obtained by the C2Q models. Our results show that ZX pooling with convolution ansatz 9 and generalized pooling with convolution ansatz 4, 5, 9, and 10 have a statistically significant quantum advantage over both 1D and 2D classical models for all classification problems, despite having a smaller number of model parameters. Further details on the statistical analysis can be found in Appendix C. These findings underscore the potential of quantum-enhanced machine learning models in solving complex classification tasks, even with limited model resources.

The underlying source of the quantum advantage in quantum computing remains an open question. However, it is speculated that the advantage is related to certain properties of quantum computing that have no classical equivalent. The first property is the ability of quantum measurements to discriminate non-orthogonal states, which enables quantum computers to capture subtle differences in data that are not captured by classical computers. The second property is the ability of quantum convolutional operations to create entanglement among all qubits through the use of two-qubit gates between nearest neighbors, which allows for the capture of non-local correlations. In addition, the ability of a quantum computer to store N -dimensional data in $\lceil \log_2(N) \rceil$ qubits, and the ability of the QCNN to classify M -qubit quantum states using only $O(\log(M))$ parameters, make it possible to construct an extremely compact machine learning model.

5. Conclusion

In this study, we proposed a classical-to-quantum CNN (C2Q-CNN), a transfer learning (TL) model that uses some layers of a pre-trained

CNN as a starting point for a quantum CNN (QCNN). The QCNN constitutes an extremely compact machine learning (ML) model because the number of trainable parameters grows logarithmically with the number of initial qubits [4] and is promising because of the absence of barren plateaus [16] and generalization capabilities [18]. Supervised learning with a QCNN has also demonstrated classification performance superior to that of its classical counterparts under similar training conditions for a number of canonical datasets [20]. C2Q-CNN TL provides an approach to utilize the advantages of QCNN in the few-parameter regime to the full extent. Moreover, the proposed method is suitable for implementation in quantum hardware expected to be developed in the near future because it is robust to systematic errors and can be implemented with a shallow-depth quantum circuit. Therefore, C2Q-CNN TL is a strong candidate for practical applications of NISQ computing in ML with a quantum advantage.

To demonstrate the quantum advantage of C2Q-CNN, we conducted a comparative study between two classical-to-classical (C2C) transfer learning (TL) models and C2Q TL models. The C2C and C2Q TL models shared the same pre-trained CNN, with the C2C TL models having slightly more parameters than the C2Q TL models. The pre-training was performed on the Fashion-MNIST dataset for multinomial classification. Then the target model replaced the final dense layer of the source model and was trained for three independent binary classification tasks using the MNIST data. Our simulation results, obtained using PennyLane and Keras, revealed that the C2Q models consistently achieved higher classification accuracy compared to the C2C models, despite having fewer trainable parameters. These results highlight the potential of quantum-enhanced transfer learning in improving the performance of machine learning models. It is important to note that while our simulation utilized a source CNN specifically designed for this study, the C2Q-CNN TL framework is compatible with other existing CNNs, such as VGGNet [50], ResNet [51], and DenseNet [52]. This compatibility enhances the versatility of our approach, enabling researchers to leverage established CNN designs within the quantum-enhanced transfer learning paradigm.

The potential future research directions are as follows. First, the reason behind the quantum advantage demonstrated by C2Q-CNN remains unclear. Although rigorous analysis is lacking, we speculate that this advantage is related to the ability of a quantum measurement to discriminate non-orthogonal states, for which a classical analog does not exist. Moreover, verifying whether the quantum advantage would continue to hold as the number of trainable parameters increases and for other datasets would be interesting. To increase the number of model parameters for a fixed number of features and input qubits, one may consider generalizing the QCNN model to utilize multiple channels, as in many classical CNN models. Note that, in the TL tested in our experiment, the final dense layer was replaced with a model subjected to fine-tuning, while the entire convolutional part was frozen. Testing the various depths of frozen layers would be an interesting topic for future research. For example, freezing a smaller number of layers to use the features of an earlier layer of the convolutional stage can be beneficial when the new dataset is small and significantly different from the source. The focus of this study was on inductive TL, for which both the source and new datasets were labeled. Exploring the potential of leveraging quantum techniques in other TL scenarios, such as self-taught, unsupervised, and transductive TL [24], is a promising direction for future research. Furthermore, extending the C2Q TL approach to address other machine learning problems, such as semi-supervised learning [53,54] and one-class classification [55,56], poses an open challenge for future research.

CRedit authorship contribution statement

Juhyeon Kim: Performed simulations, Analyzed and discussed the results, Contributed to the writing of manuscript. **Joonsuk Huh:** Analyzed and discussed the results, Contributed to the writing of manuscript. **Daniel K. Park:** Conceived the main framework, Designed the experiments, Analyzed and discussed the results, Contributed to the writing of manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the Yonsei University Research Fund of 2022 (2022-22-0124), the National Research Foundation of Korea, South Korea (Grant Nos. 2021M3H3A1038085, 2019M3E4A1079666, 2022M3E4A1074591, 2022M3H3A1063074, and 2021M3E4A1038308), and the KIST Institutional Program (2E32241-23-010).

Appendix A. Encoding classical data to a quantum state

The first step in applying quantum ML to a classical dataset is to transform the classical data into a quantum state. Without loss of generality, we consider the classical data given as an N -dimensional real vector $\vec{x} \in \mathbb{R}^N$. Several encoding methods exist to achieve this, such as algorithms that require a quantum circuit with $O(N)$ width and $O(1)$ depth and algorithms that require $O(\log(N))$ width and $O(\text{poly}(N))$ depth [40–43]. Among the various encoding methods explored previously [20], we observed that amplitude encoding performs best in most cases.

A.1. Amplitude encoding

Amplitude encoding encodes classical data into the probability amplitude of each computational quantum state. Amplitude encoding transforms $x = (x_1, \dots, x_N)^T$ of dimension $N = 2^n$ classical data into an n -qubit quantum state $|\psi(x)\rangle$ as follows:

$$U(x) : x \in \mathbb{R}^N \rightarrow |\psi(x)\rangle = \frac{1}{\|\vec{x}\|} \sum_{i=1}^N x_i |i\rangle, \quad (\text{A.1})$$

where $|i\rangle$ denotes the i th computational basis state. Amplitude encoding can optimize the number of parameters on the $O(\log(N))$ scale. However, the quantum circuit depth of amplitude encoding typically increases with $O(\text{poly}(N))$.

A.2. Qubit encoding

Qubit encoding uses a constant quantum circuit depth, while using N qubits. Qubit encoding rescales classical data to x_i which lies between 0 and π , and then inputs x_i into a single qubit as $|\psi(x)\rangle = \cos(\frac{x_i}{2})|0\rangle + \sin(\frac{x_i}{2})|1\rangle$ for $i = 1, \dots, N$. Therefore, qubit encoding transforms $x = (x_1, \dots, x_N)^T$ into N qubits as

$$U(x) : x \in \mathbb{R}^N \rightarrow |\psi(x)\rangle = \bigotimes_{i=1}^N \left(\cos\left(\frac{x_i}{2}\right)|0\rangle + \sin\left(\frac{x_i}{2}\right)|1\rangle \right) \quad (\text{A.2})$$

where $x_i \in [0, \pi)$ for all i . This unitary operator $U(x)$ can be expressed as the tensor product of a single qubit unitary operator $U(x) = \bigotimes_{j=1}^N U_{x_j}$, where

$$U_{x_j} = e^{-i\frac{x_j}{2}\sigma_y} = \begin{bmatrix} \cos\left(\frac{x_j}{2}\right) & -\sin\left(\frac{x_j}{2}\right) \\ \sin\left(\frac{x_j}{2}\right) & \cos\left(\frac{x_j}{2}\right) \end{bmatrix}. \quad (\text{A.3})$$

Table C.1ZX pooling with the 0 vs. 1 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|--------|--------|--------|--------|--------|--------|---------------|---------------|---------------|---------------|
| 1D 64 p -value | 0.0312 | 0.1289 | 0.7327 | 0.0793 | 0.1795 | 0.2957 | 0.0316 | 0.0240 | 0.0166 | 0.0422 |
| 2D 76 p -value | 0.0047 | 0.8761 | 0.0323 | 0.5240 | 0.9725 | 0.5402 | 0.1348 | 0.0888 | 0.0353 | 0.2441 |

Table C.2ZX pooling with the 2 vs. 3 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|--------|--------|--------|---------------|--------|---------------|---------------|--------|---------------|---------------|
| 1D 64 p -value | 0.0768 | 0.9352 | 0.9927 | 0.0497 | 0.2332 | 0.1700 | 0.0304 | 0.3209 | 0.0446 | 0.0119 |
| 2D 76 p -value | 0.0289 | 0.6896 | 0.7954 | 0.0047 | 0.0579 | 0.0375 | 0.0023 | 0.0985 | 0.0040 | 0.0006 |

Table C.3ZX pooling with the 8 vs. 9 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|--------|--------|--------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1D 64 p -value | 0.0008 | 0.2373 | 0.4708 | 0.0035 | 0.0024 | 0.0012 | 0.0007 | 0.0010 | 0.0004 | 0.0002 |
| 2D 76 p -value | 0.0000 | 0.2806 | 0.0495 | 0.1591 | 0.0904 | 0.0204 | 0.0117 | 0.0085 | 0.0015 | 0.0002 |

Table C.4Generalized pooling with the 0 vs. 1 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|--------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1D 64 p -value | 0.1079 | 0.0215 | 0.0302 | 0.0169 | 0.0139 | 0.0297 | 0.0217 | 0.0206 | 0.0151 | 0.0103 |
| 2D 76 p -value | 0.7661 | 0.0535 | 0.1171 | 0.0308 | 0.0197 | 0.1026 | 0.0598 | 0.0609 | 0.0268 | 0.0107 |

Table C.5Generalized pooling with the 2 vs. 3 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1D 64 p -value | 0.0921 | 0.0290 | 0.0389 | 0.0106 | 0.0099 | 0.0138 | 0.0109 | 0.0074 | 0.0068 | 0.0048 |
| 2D 76 p -value | 0.0129 | 0.0021 | 0.0034 | 0.0006 | 0.0006 | 0.0008 | 0.0006 | 0.0004 | 0.0003 | 0.0002 |

Table C.6Generalized pooling with the 8 vs. 9 classification p -value results.

| Convolution circuit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1D 64 p -value | 0.0080 | 0.0079 | 0.0013 | 0.0009 | 0.0012 | 0.0010 | 0.0016 | 0.0004 | 0.0007 | 0.0002 |
| 2D 76 p -value | 0.4616 | 0.4574 | 0.0202 | 0.0085 | 0.0261 | 0.0062 | 0.0279 | 0.0010 | 0.0038 | 0.0002 |

Appendix B. Classical neural network

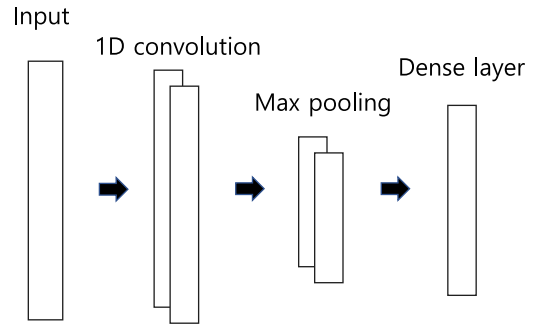
We devised classical convolutional neural networks to compare C2C TL and C2Q TL under the same training conditions. In particular, we assigned the similar number of model parameters to C2C TL and C2Q TL. We created 1D and 2D CNN models for the C2C TL using Keras [46]. We used the same pre-trained CNN model with C2Q TL, which is introduced in Fig. 4. To compare the test accuracies of C2C and C2Q transfer models under the same conditions, the learning rate, batch size, and number of iterations were fixed, and the Adam optimizer was used.

B.1. 1D CNN

The structure of the 1D CNN model is illustrated in Fig. B.6. The CNN takes 256 features produced by the source (pre-trained) CNN as input, and passes them on to 1D convolution and max pooling layers. The output feature size is reduced to 28 at the end of the max pooling layer. Finally, a dense layer is applied to reduce the size of the output features to two for binary classification. The total number of trainable parameters is 64.

B.2. 2D CNN

The structure of the 2D CNN model is shown in Fig. B.7. The CNN takes 16×16 data, which is reshaped from 256 features produced by the source (pre-trained) CNN as input. This two-dimensional data is passed on to the 2D convolution and max pooling layer twice, and the output features are reduced to eight. Finally, a dense layer is applied to

**Fig. B.6.** 1D CNN model.

reduce the size of the output features to two for binary classification. The total number of trainable parameters is 76.

Appendix C. Welch's t-test

The Welch's t-test [49] is a widely-utilized method for assessing the equality of means between two populations with unequal variances. In our study, the Welch's t-test was implemented using SciPy [57] to obtain p -values between the C2Q TL model and the C2C TL models. In accordance with standard statistical practice [58,59], a p -value less than α , where α is typically set to 0.05, is considered to indicate

Table C.7
Trivial pooling p -value results.

| Classification | 0 vs 1 | | 2 vs 3 | | 8 vs 9 | |
|------------------|---------------|--------|---------------|---------------|---------------|---------------|
| Convolution | 10 | 11 | 10 | 11 | 10 | 11 |
| 1D 64 p -value | 0.0340 | 0.9678 | 0.0602 | 0.1097 | 0.0002 | 0.0005 |
| 2D 76 p -value | 0.1305 | 0.1192 | 0.0065 | 0.0173 | 0.0003 | 0.0048 |

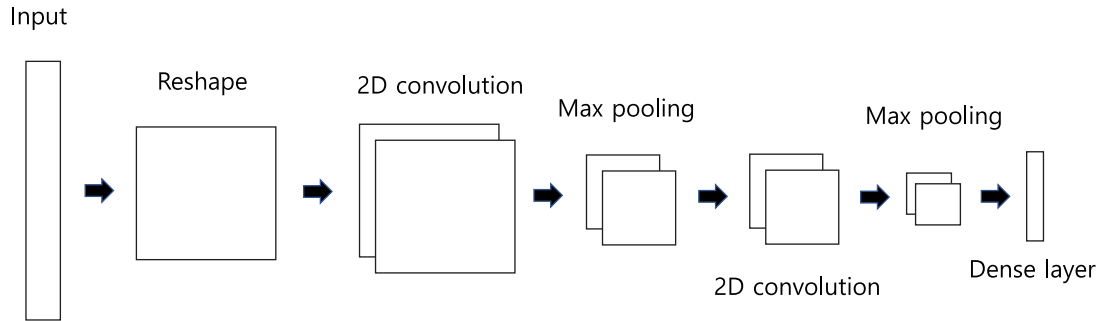


Fig. B.7. 2D CNN model.

statistical significance. Based on the obtained p -values, we conclude that if a C2Q TL model demonstrates statistical significance compared to the C2C TL models and achieves a higher accuracy, it can be inferred to possess a meaningful advantage (quantum advantage). The p -value results were organized into groups based on the ZX pooling, generalized pooling, and trivial pooling approaches, as previously discussed in the paper. For each pooling type, the p -value results were further grouped by the type of learning problem, which in this case was binary classification for 0 and 1, 2 and 3, and 8 and 9. The results of ZX pooling are presented in Tables C.1, C.2, and C.3 for classifying between 0 and 1, 2 and 3, and 8 and 9, respectively. The results of generalized pooling are presented in Tables C.4, C.5, and C.6 in the same order. Finally, the results of trivial pooling are listed in Table C.7. We highlighted in bold any statistically significant p -values where the corresponding C2Q model exhibited higher accuracy than the C2C model.

References

- [1] Jonathan Romero, Jonathan P. Olson, Alan Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, *Quantum Sci. Technol.* 2 (4) (2017) 045001.
- [2] K. Mitarai, M. Negoro, M. Kitagawa, K. Fujii, Quantum circuit learning, *Phys. Rev. A* 98 (2018) 032309.
- [3] Marcello Benedetti, Erika Lloyd, Stefan Sack, Mattia Fiorentini, Parameterized quantum circuits as machine learning models, *Quantum Sci. Technol.* 4 (4) (2019) 043001.
- [4] Iris Cong, Soonwon Choi, Mikhail D. Lukin, Quantum convolutional neural networks, *Nat. Phys.* 15 (12) (2019) 1273–1278.
- [5] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, Patrick J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* 3 (9) (2021) 625–644.
- [6] S. Mangini, F. Tacchino, D. Gerace, D. Bajoni, C. Macchiavello, Quantum computing models for artificial neural networks, *Europhys. Lett.* 134 (1) (2021) 10002.
- [7] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, Dacheng Tao, Learnability of quantum neural networks, *PRX Quantum* 2 (2021) 040337.
- [8] Jun Li, Xiaodong Yang, Xinhua Peng, Chang-Pu Sun, Hybrid quantum–classical approach to quantum optimal control, *Phys. Rev. Lett.* 118 (2017) 150503.
- [9] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, Nathan Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* 99 (2019) 032331.
- [10] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alan Aspuru-Guzik, Jeremy L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nature Commun.* 5 (1) (2014) 4213.
- [11] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, Alan Aspuru-Guzik, The theory of variational hybrid quantum–classical algorithms, *New J. Phys.* 18 (2) (2016) 023023.
- [12] John Preskill, Quantum computing in the NISQ era and beyond, *Quantum* 2 (2018) 79.
- [13] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermann Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, Alan Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Modern Phys.* 94 (2022) 015004.
- [14] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, Hartmut Neven, Barren plateaus in quantum neural network training landscapes, *Nature Commun.* 9 (1) (2018) 4812.
- [15] Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, Simone Severini, Hierarchical quantum classifiers, *npj Quantum Inf.* 4 (1) (2018) 65.
- [16] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, Patrick J. Coles, Absence of barren plateaus in quantum convolutional neural networks, *Phys. Rev. X* 11 (2021) 041011.
- [17] Rui Huang, Xiaoqing Tan, Qingshan Xu, Variational quantum tensor networks classifiers, *Neurocomputing* 452 (2021) 89–98.
- [18] Leonardo Banchi, Jason Pereira, Stefano Pirandola, Generalization in quantum machine learning: A quantum information standpoint, *PRX Quantum* 2 (2021) 040321.
- [19] Ian MacCormack, Conor Delaney, Alexey Galda, Nidhi Aggarwal, Prineha Narang, Branching quantum convolutional neural networks, *Phys. Rev. Res.* 4 (2022) 013117.
- [20] Tak Hur, Leeseok Kim, Daniel K. Park, Quantum convolutional neural network for classical data classification, *Quantum Mach. Intell.* 4 (1) (2022) 3.
- [21] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [22] Han Xiao, Kashif Rasul, Roland Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.
- [23] Stevo Bozinovski, Reminder of the first paper on transfer learning in neural networks, *Informatica (Slovenia)* 44 (1976) 2020.
- [24] Sinno Jialin Pan, Qiang Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359.
- [25] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, Chunfang Liu, A survey on deep transfer learning, in: Věra Kůrková, Yannis Manolopoulos, Barbara Hammer, Lazaros Iliadis, Ilias Maglogiannis (Eds.), *Artificial Neural Networks and Machine Learning, ICANN 2018*, Springer International Publishing, Cham, 2018, pp. 270–279.
- [26] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016, <http://www.deeplearningbook.org>.
- [27] Andrea Mari, Thomas R. Bromley, Josh Izaac, Maria Schuld, Nathan Killoran, Transfer learning in hybrid classical-quantum neural networks, *Quantum* 4 (2020) 340.
- [28] 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, Nathan Killoran, PennyLane: Automatic differentiation of hybrid quantum–classical computations, 2020.
- [29] Sukin Sim, Peter D. Johnson, Alan Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum–classical algorithms, *Adv. Quantum Technol.* 2 (12) (2019) 1900070.
- [30] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, Todd J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, *Phys. Rev. Lett.* 122 (23) (2019) 230401.

- [31] Hai-Rui Wei, Yao-Min Di, Decomposition of orthogonal matrix and synthesis of two-qubit and three-qubit orthogonal gates, *Quantum Inf. Comput.* 12 (3–4) (2012) 262–270.
- [32] Farrokh Vatan, Colin Williams, Optimal quantum circuits for general two-qubit gates, *Phys. Rev. A* 69 (3) (2004) 032315.
- [33] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, Qing He, A comprehensive survey on transfer learning, *Proc. IEEE* 109 (1) (2021) 43–76.
- [34] Emilio Soria Olivas, Jose David Martin Guerrero, Marcelino Martinez Sober, *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques - 2 Volumes*, Information Science Reference - Imprint of: IGI Publishing, Hershey, PA, 2009.
- [35] Gui-Lu Long, Yang Sun, Efficient scheme for initializing a quantum register with an arbitrary superposed state, *Phys. Rev. A* 64 (2001) 014303.
- [36] Andrei N. Soklakov, Rüdiger Schack, Efficient state preparation for a register of quantum bits, *Phys. Rev. A* 73 (2006) 012307.
- [37] Michele Mosca, Phillip Kaye, Quantum networks for generating arbitrary quantum states, in: *Optical Fiber Communication Conference and International Conference on Quantum Information*, Optical Society of America, 2001, p. PB28.
- [38] Martin Plesch, Časlav Brukner, Quantum-state preparation with universal gate decompositions, *Phys. Rev. A* 83 (2011) 032302.
- [39] Mikko Möttönen, Juha J. Vartiainen, Ville Bergholm, Martti M. Salomaa, Transformation of quantum states using uniformly controlled rotations, *Quantum Inf. Comput.* 5 (6) (2005) 467–473.
- [40] Israel F. Araujo, Daniel K. Park, Francesco Petruccione, Adenilton J. da Silva, A divide-and-conquer algorithm for quantum state preparation, *Sci. Rep.* 11 (1) (2021) 6329.
- [41] T.M.L. Veras, I.C.S. De Araujo, K.D. Park, A.J. da Silva, Circuit-based quantum random access memory for classical data with continuous amplitudes, *IEEE Trans. Comput.* (2020) 1.
- [42] Ryan LaRose, Brian Coyle, Robust data encodings for quantum classifiers, *Phys. Rev. A* 102 (2020) 032420.
- [43] Israel F. Araujo, Daniel K. Park, Teresa B. Ludermir, Wilson R. Oliveira, Francesco Petruccione, Adenilton J. da Silva, Configurable sublinear circuits for quantum state preparation, 2021, arXiv preprint arXiv:2108.10182.
- [44] Harshit Mogalapalli, Mahesh Abburi, B. Nithya, Surya Kiran Vamsi Bandreddi, Classical-Quantum transfer learning for image classification, *SN Comput. Sci.* 3 (1) (2022) 20.
- [45] Jun Qi, Javier Tejedor, Classical-to-quantum transfer learning for spoken command recognition based on quantum neural networks, in: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2022*, pp. 8627–8631.
- [46] François Chollet, et al., Keras, 2015, <https://keras.io>.
- [47] Vojtech Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, Jay M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature* 567 (7747) (2019) 209–212.
- [48] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
- [49] B.L. Welch, The generalization of 'student's' problem when several different population variances are involved, *Biometrika* 34 (1–2) (1947) 28–35.
- [50] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
- [51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016*, pp. 770–778.
- [52] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017*, pp. 2261–2269.
- [53] Oliver Chapelle, Bernhard Schölkopf, Alexander Zien, *Semi-Supervised Learning*, MIT Press, Cambridge, 2006.
- [54] Wandong Zhang, Q.M. Jonathan Wu, Yimin Yang, Semisupervised manifold regularization via a subnetwork-based representation learning model, *IEEE Trans. Cybern.* PP (99) (2022) 1–14.
- [55] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, Marius Kloft, Deep one-class classification, in: *International Conference on Machine Learning, PMLR, 2018*, pp. 4393–4402.
- [56] Wandong Zhang, Q.M. Jonathan Wu, W.G. Will Zhao, Haojin Deng, Yimin Yang, Hierarchical one-class model with subnetwork for representation learning and outlier detection, *IEEE Trans. Cybern.* PP (99) (2022) 1–14.
- [57] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C.J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nature Methods* 17 (2020) 261–272.
- [58] Valen E. Johnson, Revised standards for statistical evidence, *Proc. Natl. Acad. Sci.* 110 (48) (2013) 19313–19317.
- [59] Martin Krzywinski, Naomi Altman, Significance, P values and t-tests, *Nat. Methods* 10 (11) (2013) 1041–1042.

Juhyeon Kim received the B.S. degree in department of physics from Sungkyunkwan University (SKKU) in 2021. He is master's student in advanced institute of nano technology in SKKU. His research interests include quantum algorithms, machine learning, and bioinformatics, especially on quantum algorithms for machine learning.

Joonsuk Huh is an associate professor of the chemistry department of Sungkyunkwan University (SKKU), and he is visiting Xanadu Quantum Technologies Inc. for his sabbatical year of 2023. He received his Ph.D. in theoretical physics in 2011 from the University of Frankfurt, where he developed computational methods for the classical simulation of vibronic spectra. He then joined Harvard University as a postdoc in 2011. In 2015, Huh and his co-workers linked molecular vibronic spectra to quantum sampling. His M-Qudit lab at SKKU develops quantum algorithms for chemistry, bioinformatics, machine learning, and matrix functions.

Daniel K. Park is an Assistant Professor at Yonsei University in Korea, where he works on quantum information processing and machine learning. Before joining Yonsei University in 2022, Daniel was a research professor at Sungkyunkwan University for approximately 1 year and at KAIST for approximately 2 years, where he also worked as a post-doctoral researcher for 3 years. Daniel obtained his Ph.D. degree in Physics-Quantum Information in 2015 at the University of Waterloo and Institute for Quantum Computing.