

Article

Remote Sensing Image Scene Classification in Hybrid Classical–Quantum Transferring CNN with Small Samples

Zhouwei Zhang ^{1,2}, Xiaofei Mi ^{1,2} , Jian Yang ^{1,2,*}, Xiangqin Wei ^{1,2} , Yan Liu ^{1,2} , Jian Yan ^{1,2}, Peizhuo Liu ³, Xingfa Gu ^{1,2} and Tao Yu ^{1,2}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhangzw@aircas.ac.cn (Z.Z.); mixf@aircas.ac.cn (X.M.); weixq@aircas.ac.cn (X.W.); liuyan@aircas.ac.cn (Y.L.); yanjian19@mails.ucas.ac.cn (J.Y.); yutao@aircas.ac.cn (T.Y.)

² National Engineering Laboratory for Satellite Remote Sensing Applications, Beijing 100094, China

³ School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China; liupeizhuo@buaa.edu.cn

* Correspondence: yangjian@aircas.ac.cn

Abstract: The scope of this research lies in the combination of pre-trained Convolutional Neural Networks (CNNs) and Quantum Convolutional Neural Networks (QCNN) in application to Remote Sensing Image Scene Classification (RSISC). Deep learning (DL) is improving by leaps and bounds pretrained CNNs in Remote Sensing Image (RSI) analysis, and pre-trained CNNs have shown remarkable performance in remote sensing image scene classification (RSISC). Nonetheless, CNNs training require massive, annotated data as samples. When labeled samples are not sufficient, the most common solution is using pre-trained CNNs with a great deal of natural image datasets (e.g., ImageNet). However, these pre-trained CNNs require a large quantity of labelled data for training, which is often not feasible in RSISC, especially when the target RSIs have different imaging mechanisms from RGB natural images. In this paper, we proposed an improved hybrid classical–quantum transfer learning CNNs composed of classical and quantum elements to classify open-source RSI dataset. The classical part of the model is made up of a ResNet network which extracts useful features from RSI datasets. To further refine the network performance, a tensor quantum circuit is subsequently employed by tuning parameters on near-term quantum processors. We tested our models on the open-source RSI dataset. In our comparative study, we have concluded that the hybrid classical–quantum transferring CNN has achieved better performance than other pre-trained CNNs based RSISC methods with small training samples. Moreover, it has been proven that the proposed algorithm improves the classification accuracy while greatly decreasing the amount of model parameters and the sum of training data.

Keywords: CNN; hybrid classical–quantum neural networks; transfer learning; variational quantum circuit



Citation: Zhang, Z.; Mi, X.; Yang, J.; Wei, X.; Liu, Y.; Yan, J.; Liu, P.; Gu, X.; Yu, T. Remote Sensing Image Scene Classification in Hybrid Classical–Quantum Transferring CNN with Small Samples. *Sensors* **2023**, *23*, 8010. <https://doi.org/10.3390/s23188010>

Academic Editor: Alfred Stein

Received: 29 June 2023

Revised: 4 September 2023

Accepted: 7 September 2023

Published: 21 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent times with the rocketing progress of the earth observation capability and the artificial intelligent technology, the progression and innovation of remote sensing (RS) are more important than ever. RSI data processing is now developing towards RSI the age of big data, which relies on data model by data-driven intelligent analysis [1] (refer to Table 1 for all acronyms that are used throughout the paper). Machine learning (ML) plays an essential role in extracting underlying features or a probability distribution of patterns in RSI datasets, such as damage prediction, target recognition or image classification in hitherto unknown domains [2]. In particular, deep learning (DL) has proven to be a technological innovation and a historic milestone in many fields [3]. It emphasizes neural networks (NNs) involving multiple hidden layers, which use feature representations learned exclusively from the data, instead of handcrafted features that are designed based mainly on domain-specific

knowledge. DL has widely used in various research topics, for example, data analysis tasks, including image segmentation, image target detection, and image classification.

Table 1. List of acronyms.

RSI	Remote Sensing Image
CNN	Convolutional neural network
RSISC	Remote sensing image scene classification
RS	Remote sensing
DL	Deep learning
NN	Neural network
HHL	Harrow Hassidim Lloyd
VQE	Variational quantum eigensolver
QAOA	Quantum approximate optimization algorithm
PQC	Parameterized quantum circuit
NISQ	Noisy intermediate-scale quantum
VQC	Variational quantum circuit
QNN	Quantum neural network
MBS	Many-body system
CQ	Classical–quantum
QC	Quantum–classical
QQ	Quantum–quantum
CC	Classical–classical
QE	Quantum encoding
IQP	Instantaneous quantum polynomial
ILSRC	ImageNet Large Scale Visual Recognition Challenge
AID	Aerial Image dataset
OA	Overall accuracy
BoVW	Bag-of-Visual-Words
SIFT	Scale-invariant feature transform
SVM	Support vector machine

Various deep learning models have been developed with outstanding performance for image classification on RSI datasets in multiple applications. Recently, significant progress has been made in the methods of RSISC. Early studies involving RSISC in CNN were transplanted from natural image data, which led to the increasing of the difficulty of the RSISC task in complex RSIs [4]. Since then, researchers have dedicated to replace hand-engineered features with trainable multilayer networks, and several deep learning models have demonstrated impressive feature representation capability, applicable across a wide range of domains, including RSISC [5]. Deep learning algorithms have opened up an entirely novel frontier of learning algorithms including CNN techniques that have been adopted in the RSISC [6]. CNNs can extract hierarchical and insightful features automatically from the massive size of image data [7].

In recent years, there has been significant progress in RSISC methods using CNNs [2,3,8]. Despite the extraordinary achievement that CNN-based methods have enabled in RSISC, the models of the network construction remain a huge challenge, which is significant for the efficiency of the CNN models [7]. CNN architecture construction requires skill in both DL methodology and professional expertise [9]. In practice, handmade architecture design is complicated and fallible; moreover, the requirement of number of expertise and professionals in both DL and the investigated domain is often unrealistic in practical applications.

In this context, most studies on CNN-based RSISC have utilized networks that are pre-trained on different networks, for instance AlexNet [10], VGGNet [11], GoogleNet [12], and ResNet [13], standing out among numerous DL-based methods due to closely related powerful features extraction performance. Since both assignments involving RSIs data and the pre-trained networks can easily be operated for transferring in RSISC. Although, RSIs differ significantly in terms of spectral resolution, spatial resolution, and radiometric

resolution, etc. [14]. Researchers have used pre-trained CNN models as features extractors to extract high semantic features. These methods have limitation: It has no advantage if significant domain differences exist between source and target datasets.

Fine-tuning CNN-based models is an effective tool for RSISC. However, such approaches still have limitations in three perspectives related to datasets, models, and labels. These limitations are discussed below.

Firstly, the dataset plays an important role in advancing RSISC. The Size-Similarity Matrix of dataset determines the choices of the pre-trained model. This matrix classifies the strategies based on the size of the dataset and its similarity to the dataset in which the pre-trained model was trained. Due to the matrix transformation conducted in the fully connected layers, the dataset must be transformed into a certain fixed size. The most common approach is to crop and interpolate the original image, which inevitably results in a loss of fine information from the original RSIs. Unfortunately, either of these approaches can be detrimental to the performance of RSISC. Therefore, the selection of the pre-trained models depends on the dataset's size and size-similarity. The RSIs dataset must be transformed into a special fixed size [15].

Secondly, from a modeling perspective, most pre-trained models can achieve excellent classification performance. However, these models also have some limitations. The learned feature may be not entirely satisfactory for the properties of target datasets, and pre-trained treatments applied between heterogeneous networks require manual manipulations of layer combinations, depending on type of task. The RSISC task, in particular, requires a massive tuning procedure to achieve optimal classification performance.

Thirdly, many scene images with identical exteriors have different but correlated scene contents, which can lead to the label paradox. Because an RSI contains various objects, such as water, mountains, bare land, buildings, etc., which may be covered in a residential scene as shown in Figure 1, the variance and complex spatial distributions of the scene contents cause the diversity of RSISC. Because of the different varieties in the exteriors of ground objects within the similar semantic information, traditional methods of RSISC often fail to achieve satisfactory classification accuracy. In addition, low scene category separability has been caused by the presence of the same scene contents within different scene categories or high semantic superimposition between different scene types. Another important difficulty of RSISC is the large variance of object scales caused by sensor imaging altitude variation. Furthermore, the label distribution of ground objects in datasets is irregular and often unavailable in the original training dataset, which is typically caused by resembling the sample dataset that has different but correlated labels, as can be seen in Figure 1 [16].

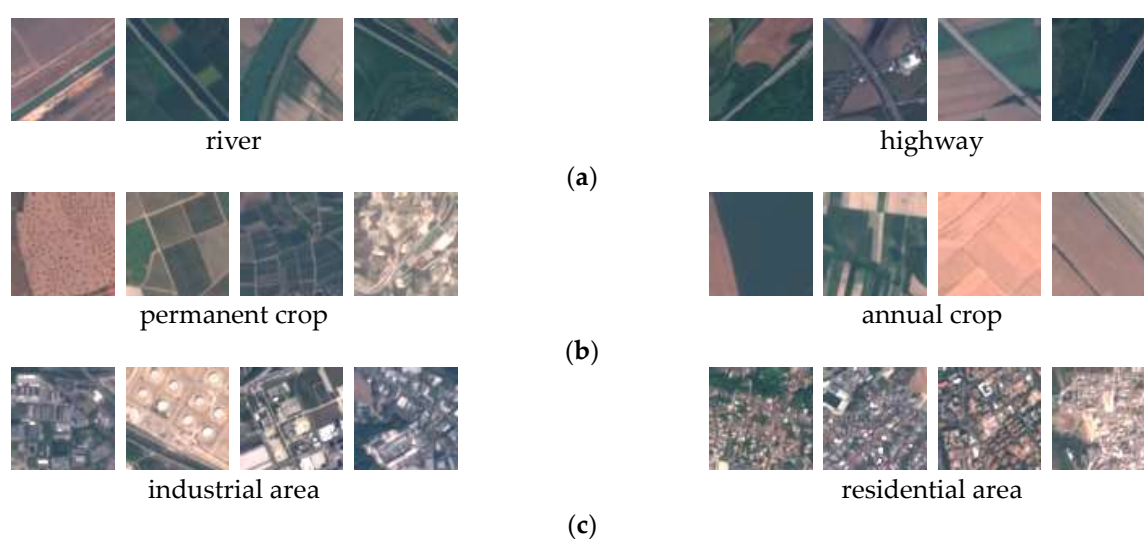


Figure 1. Limitations of RSISC, which include (a) low type separability, (b) complex variance of scene scales, (c) coexistence of multiple objects. The sources are from Eurosat dataset [17].

Both RSISC and quantum computing are emergent techniques that have the potential to transform the research and application of RS. Quantum computing can provide significant advantages in terms of improving the performance of classification techniques. The great potential of quantum computing in ML has been actively investigated. In specific cases, traditional machine learning tasks can be improved with exponential acceleration when it operates on a quantum computer [18,19]. The Harrow–Hassidim–Lloyd (HHL) algorithm is a quantum computing that has been successfully implanted in conventional ML theories and topics, such as data mining, artificial neural networks, computational learning theory, etc., as shown in [20–24]. HHL-based algorithms are based on the quantum phase estimation algorithm, which operates in a high-depth quantum circuit [25]. To bypass this strict requirement of hardware, classical–quantum hybrid algorithms containing a low-depth quantum circuit, for instance, the variational quantum eigensolver (VQE) and the quantum approximate optimization algorithm (QAOA) have been suggested [26,27]. The idea of a hybrid algorithm is to divide the problem into two components, either of which can be operated on a quantum and a classical computer separately. Cai verified the feasibility of a K-means algorithm on a quantum computer [28]. Otgonbaatar proposed a parameterized quantum circuit (PQC) with only 17 quantum bits for classifying a two-label Sentinel RSI dataset. Quantum-based pseudo-labelling for hyperspectral imagery classification is demonstrated by Shaik [22]. Noisy intermediate-scale quantum (NISQ) devices are an advanced quantum computing technology providing solutions for large-scale and complex practical quantum computation, such as solving high-complexity problems or supporting ML algorithms. An NISQ device is termed a variational quantum circuit (VQC) when it has parameterized quantum gates and parameterized quantum circuit. Sometimes, the VQC is a supervised learning algorithm in which quantum neural networks (QNNs) are trained to perform a classification mission.

A transformer-based model is one of the successful approaches of DL and more efficient than other traditional CNN models in a variety of downstream applications. Transformer-based concepts in the context of quantum machine learning consist mainly of three variants: classical–quantum (CQ), quantum–classical (QC), and quantum–quantum (QQ) [23]. The CQ concept can achieve better result in comparison with other transformer-based concept tests and classical–classical concept tests. Therefore, the CQ concept has potential with currently available quantum computers.

The purpose of this paper is to investigate the feasibility and the efficiency of the hybrid classical–quantum transferring CNN model for RSISC. Our study aims to demonstrate that hybrid classical–quantum transferring CNN models can significantly enhance the efficiency and performance of RSISC with small training samples. We focus on the hybrid models scenario, where tensor quantum circuits and CNNs can be jointly trained to achieve RSISC. With the advent of NISQ devices, CQ transfer learning has become particularly attractive, since, with the application of a classical–quantum neural network with large input images, it is possible to efficiently extract highly informative features with a VQC. It is advantageous, since it makes the application of the potential of quantum physics theory—superposition and entanglement, coordinated with the mature methods of classical ML—possible. Accordingly, tensor quantum circuits can be regarded as common quantum feature extractors, which can mimic famous classical networks that are often used as pre-trained networks.

The transfer learning approach in the quantum domain has been largely unexplored, with only very few applications, such as modeling quantum Many-Body Systems (MBSs), associated with classical autoencoders to a quantum Boltzmann machine, and initializing variational quantum networks. In this paper, an innovative, self-designed, and systematic theory, through a concept of tensor quantum circuits and pre-trained networks, is proposed.

We present a model-theoretic approach and provide proof-of-principle examples for practical implementation through numerical simulations. Additionally, we test our model experimentally on physical quantum processors, successfully conducting RSISC with a hybrid CQ system.

The remaining sections of this work are introduced as follows: Section 2 introduces some notations about hybrid classical–quantum networks. Section 3 outlines the architecture of the hybrid classical–quantum transferring CNN. Section 4 reports the evaluation and results of our model. Finally, this paper is concluded in Section 5.

2. Hybrid Classical–Quantum Networks

2.1. Quantum Encoding

The first step is the transformation of the classical image dataset into quantum state; the process denotes the quantum encoding (QE). Most QE methods can be seen as parameterized circuits acting on initial states; the parameters in the parameterized circuits are determined via classical image data. Generally, QE methods can be divided into basis encoding, amplitude encoding, angle encoding, instantaneous quantum polynomial (IQP) encoding, and Hamiltonian evolution ansatz encoding [29–32]. IQP encoding can be achieved more easily than the other QE methods, so it is applied to our proposed algorithm. Then, IQP quantum encoding describes the following quantum state as Equation (1). $R_Y(\bullet)$ denotes Pauli rotation Y gate.

$$\left(\bigotimes_{i=1}^4 R_Y(\pi x_i)\right) |0\rangle^{\otimes 4} = \begin{bmatrix} \cos(\pi x_1) \\ \sin(\pi x_1) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_2) \\ \sin(\pi x_2) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_3) \\ \sin(\pi x_3) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_4) \\ \sin(\pi x_4) \end{bmatrix} \quad (1)$$

2.2. Variational Quantum Circuits

VQCs usually can define a subcircuit, which is a basic circuit architecture where complex VQCs can be constructed through repeating layers. Circuits consists of multiple rotating gates as well as CNOT gates that entangle each qubit with its neighboring qubit. We also need a circuit to encode the classical data onto the quantum state, so that the output of the measurement is related to the input. In this case, we encode the binary input onto the qubits of the corresponding order.

VQCs can define a quantum layer like the classical neural network. Furthermore, arbitrary VQCs can be demonstrated as below:

$$\mathcal{L} = U(w)|x\rangle \quad (2)$$

where w denotes variational parameters, a unitary operation is achieved, acting on the input state $|x\rangle$ of n_q quantum subsystems.

The depth d of a VQC is a superposition of different quantum circuits and matches the product of various parameterized unitary operations with different weights:

$$\mathcal{L} = \mathcal{L}_d \circ \dots \mathcal{L}_2 \circ \mathcal{L}_1 \quad (3)$$

On other hand, we can measure the expectation values of n_q local observables $\hat{z} = [\hat{z}_1, \hat{z}_2, \dots, \hat{z}_n]$ for the extraction of a classical output y . The process is known as a measurement layer, which maps a quantum state to a classical dataset:

$$M := \langle x | \hat{z} | x \rangle \quad (4)$$

The initial quantum layers and the final measurement state in the quantum circuits can be globally written as follows:

$$Q = M \times L \times \varepsilon \quad (5)$$

2.3. Tensor Quantum Circuits

The VQC algorithm is limited to parameters adjustment, using the quantum computer as feedback circuits to adjust the parameters in the parameterized circuits to optimal values. It can only optimize the structure of the fixed circuit; it cannot change its own structure, nor increase the number of entanglement operations, nor change the entanglement charac-

teristics expressed, so the algorithm still means the quantum chip entanglement operation. For complex Hamiltonian encoding, such as correlated systems, this common method does not performs well.

Tensor quantum circuits adopt the classical quantum hybrid algorithm. The classical part adopts the form of the tensor network to process the higher order tensor into the form of multiple lower order tensor compressors. The quantum part adopts VQC to adjust the parameters. Tensor quantum circuits consist of n qubits and consisting of k layers. Parameterized $e^{i\theta X \otimes X}$ gates are between each neighboring qubit in each layer, followed by a series of single-qubit parameterized Z and X rotations, as can be seen in Figure 2.

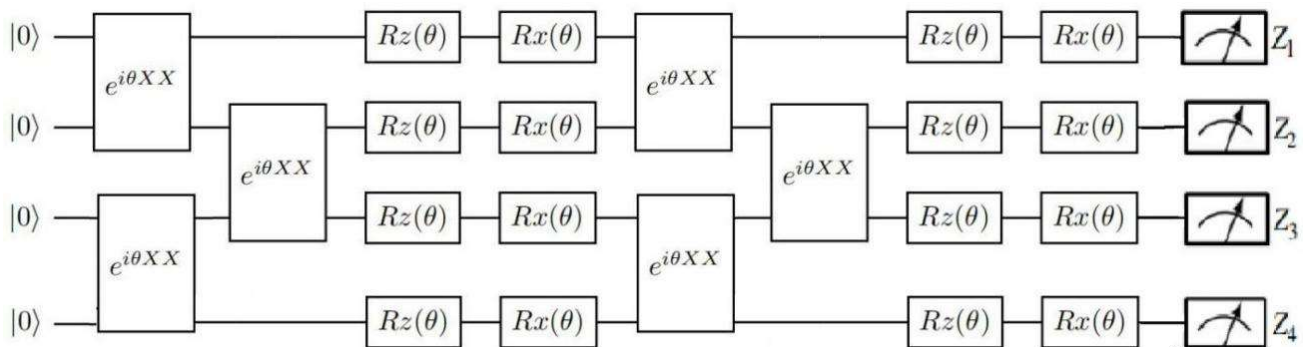


Figure 2. An illustration of tensor quantum circuits.

3. Hybrid Classical–Quantum Transferring CNN

In the following section, we introduce the concept of transferring prior “knowledge” from classical to quantum. A hybrid classical–quantum transferring CNN for RSISC proposal on the basis of two networks A and B is defined in Figure 3.

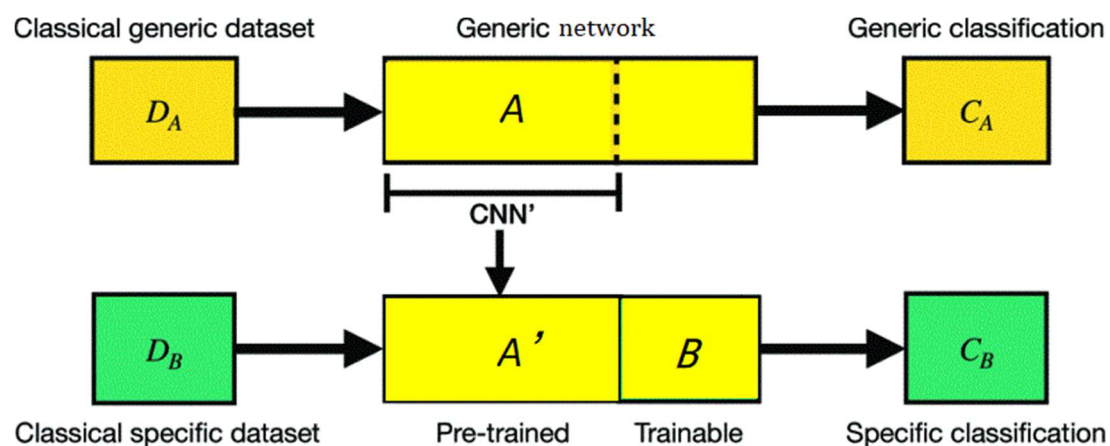


Figure 3. A hybrid classical–quantum transferring CNN for RSISC. ResNet network is utilized as features extraction and tensor quantum circuits are applied as a fully connected layer for RSISC.

A hybrid classical–quantum transferring CNN for RSISC proposal (as shown in Figure 3). The pre-trained network A on a dataset D_A for a task C_A is CNN' . Then, the final layers are removed. The reduced network A' can be taken as feature extractor. A new trainable network B can integrate closely the pre-trained network A' . The parameters of network B can adjust finely on the specific dataset.

D_A = ImageNet: ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with many classes [33].

A = ResNet34: a pre-trained residual neural network [13].

C_A = Image classification.

A' = ResNet34 a residual neural network without the final layer.

D_B = RSI datasets

$B = Q = L_{4 \rightarrow 2} \times Q \times L_{512 \rightarrow 4}$: i.e., 4-qubit tensor quantum circuits and outputs.

C_B = RSISC.

The strategy of feature extraction is much more general than what we needed in this work. In the context for transfer learning, the reduced pre-trained network A' is interpreted as a feature extractor, after removing the final layer of A . A' can produce features that are not problem-specific. In a hybrid structure, network A is classical, and the network B is quantum.

Nowadays, classical hybrid transfer learning is perhaps composed of the efficient and mature tools of DL in the current technological era of ML. It is commonly validated through the successful ML algorithms, especially for RSISC.

The quantum circuit model in Equation (6) demonstrates the concept. We assume quantum circuits of 4 qubits and utilize the following model:

$$\tilde{Q} = L_{4 \rightarrow 2} \times Q \times L_{2 \rightarrow 4} \quad (6)$$

where $L_{2 \rightarrow 4}$ implies residual networks using the activation function, $\tanh \varphi = \tanh$, Q is VQC, and $L_{4 \rightarrow 2}$ denotes residual networks without activation function. The strategy of quantum encoding establishes links between the image dataset input x and its quantum state $|x\rangle$. The chosen embedding map from the classical image dataset input vector can be written as Equation (8):

$$\varepsilon(x) = \bigotimes_{i=1}^4 (R_y(\pi x_i) H) |0\rangle^{\otimes 4} = \begin{bmatrix} \cos(\pi x_1) \\ \sin(\pi x_1) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_2) \\ \sin(\pi x_2) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_3) \\ \sin(\pi x_3) \end{bmatrix} \otimes \begin{bmatrix} \cos(\pi x_4) \\ \sin(\pi x_4) \end{bmatrix} \quad (7)$$

$$L(x) : |x\rangle \rightarrow |y\rangle = K \bigotimes_{i=1}^4 R_y(\pi x_i) |x\rangle \quad (8)$$

where H denotes a single Hadamard gate. The quantum circuits consist of 4 variational circuits. And K indicates an entangling unitary operation of 3 CNOT gates. The model in the red frame demonstrates CNOT gates and 3 rotation gates R_X , R_Y , R_Z . The CNOT gates force quantum entanglement between 2 quantum circuits, allowing the qubits from the circuits to be entangled.

Then, the measurement operators should be projected by measuring the expectation values of 4 observables, locally estimated for each qubit:

$$\hat{z} = [\hat{z}_1, \hat{z}_2, \hat{z}_3, \hat{z}_n] \quad (9)$$

$$\mathbf{M} : z = \langle x | \hat{z} | x \rangle \quad (10)$$

In the final classification stage, the cost function was used as cross entropy and via a LogSoftMax layer. The flowchart of our proposed method is shown in Figure 4, and numerical simulations were conducted using the PyTorch interface and PennyLane software.

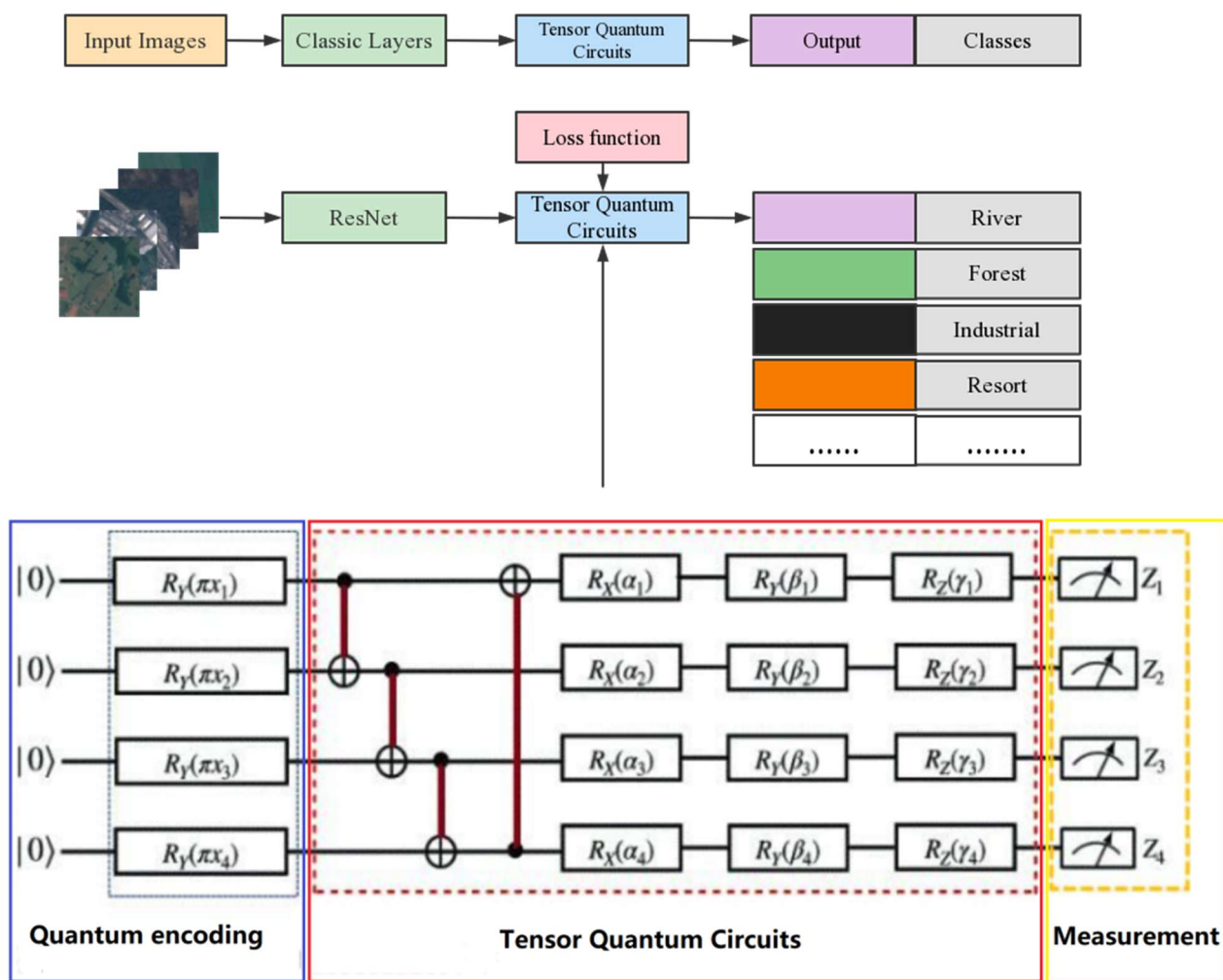


Figure 4. (Top): general scheme for the hybrid classical–quantum transferring CNN for RSISC. (Middle): detailed scheme for classifying the scene image dataset. (Bottom): architecture of the tensor quantum circuit with inputs of 4 qubits. $R_X(\bullet)$, $R_Y(\bullet)$ and $R_Z(\bullet)$ separately denote Pauli rotation X, Y, Z gates.

4. Evaluation an Results

In this section, we assess the hybrid classical–quantum transferring CNN in terms of performance gains for RSISC.

4.1. Data Profile

Our proposed method was evaluated using two challenging RSI datasets: the EuroSat dataset and the Aerial Image dataset (AID) [17,34]. The EuroSat dataset contains images taken from the Sentinel-2 satellite, categorizing the ground objects into 10 distinct land cover categories. The collection includes approximately 27,000 images divided across 10 classes, with patches measuring 64×64 pixels. The data were originally hyperspectral images captured with 13 spectral bands, but we used only RGB channels. The AID dataset includes 10,000 images of size 600×600 pixels, classified into 30 scene classes, with varying numbers of images for each label ranging from 220 to 420. The ground resolution also varies from approximately 8 m to 0.5 m per pixel. Figure 5 show examples of the EuroSAT dataset and Figure 6 shows examples of the AID dataset.

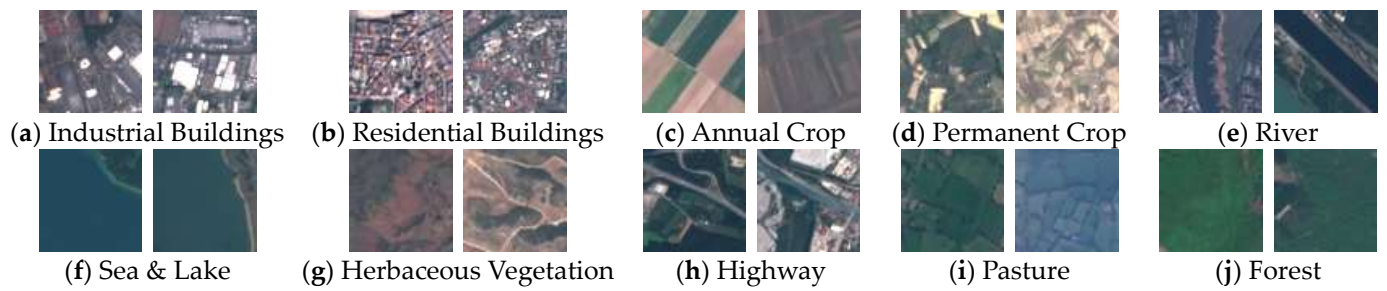


Figure 5. This outline shows all sample images of all 10 categories covered in the EuroSAT dataset. The image size has 64×64 pixels. Each category contains 2000 to 3000 images. In sum, the dataset has 27,000 geo-referenced images [17].

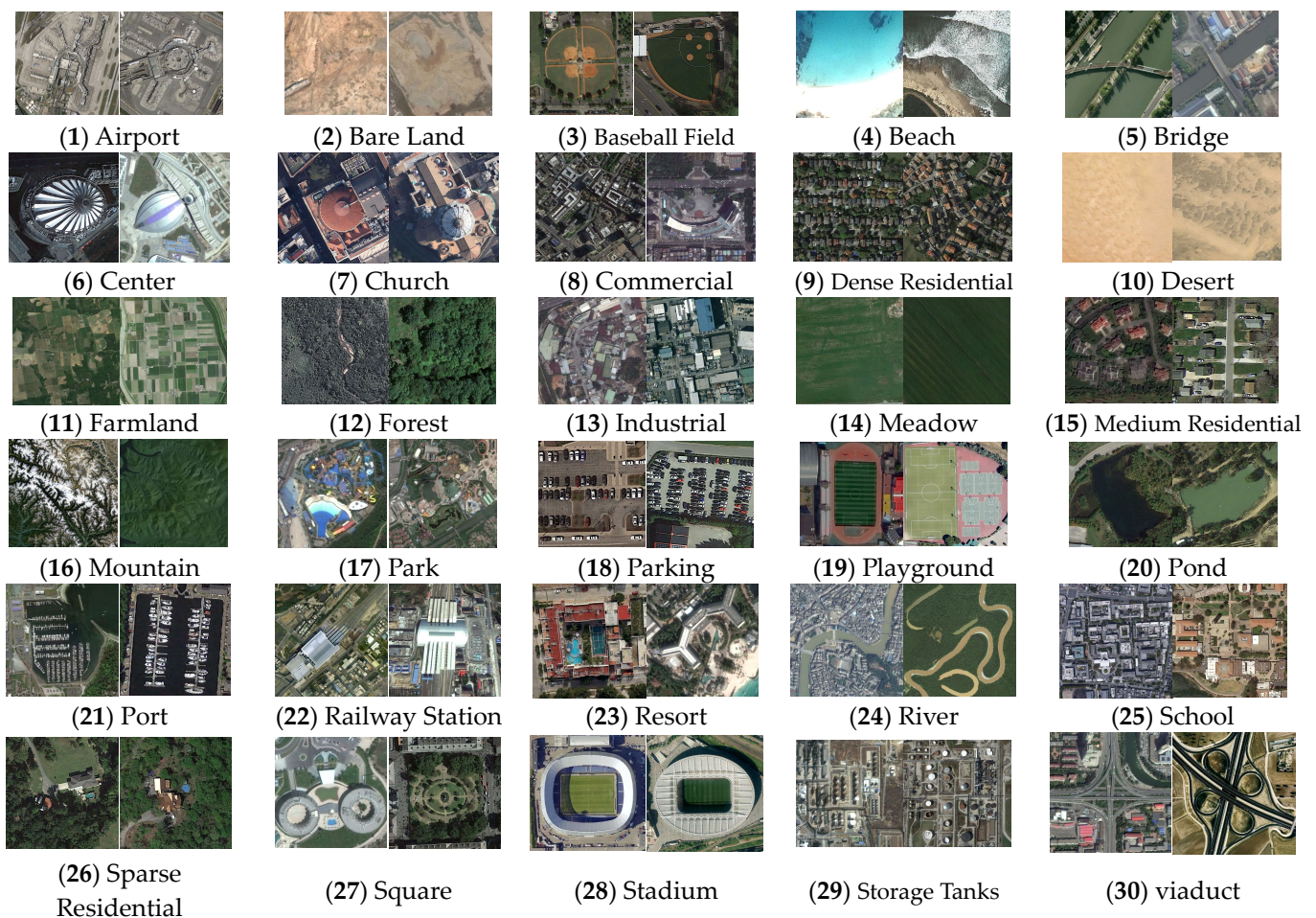


Figure 6. This outline shows all sample images of all 30 categories covered in the AID dataset. The image size has 600×600 pixels. Each category contains 220 to 420 images. In all, the dataset has 10,000 georeferenced images [34].

4.2. Evaluation Criteria

There are two extensively used evaluation criteria in RSISC: overall accuracy (OA) and confusion matrix. OA is an evaluating indicator of the classifier's performance on the whole test data set and is defined as the sum of accurately classified samples divided by the sum of tested samples. It is a commonly used to evaluate the performance of RSISC. The confusion matrix is an informative table used to analyze all the errors and confusions between different classes, generated by comparing the performance of correct and incorrect classification of each single classifier.

The definition of the calculation of OA can be expressed as follows:

$$OA = \frac{\sum_{i=1}^n \sum_{j=1}^k P_{ij}}{T} \quad (11)$$

where P_{ij} is the correct prediction of each single label, and n, k represents the sum of each label and the sum of labels. T is the sum of test dataset.

4.3. Experimental Setup

We split two different RSI datasets in different training/test ratios (10/90 ratio and 20/80 ratio) class-wise. The RGB images vast majority of all aerial and RSI datasets. For performance measurement, we compare with the performance of the proposed algorithm, the Bag-of-Visual-Words (BoVW) approach with scale-invariant feature transform (SIFT) features, a trained support vector machine (SVM) and other deep learning algorithms for accuracy evaluation. Furthermore, we trained a shallow CNN, a ResNet50, and a GoogleNet model on the training dataset [33,35–37]. In addition to the comparison with our proposed model, it is also compared with existing CQ models available in the literature, namely: CNN-QNN end-to-end model [24] and QCNN [38].

In our proposed method, we trained a ResNet34 model and the tensor quantum circuit with the model for 60 epochs over the training dataset, with a quantum depth of 6 and an initial learning rate of $\eta = 0.0006$. The model was validated with respect to the test dataset after each epoch, and overall classification on the datasets was calculated.

We used many commonly used RSISC algorithms for instance, ResNet50 [13], DCNNs [39], AlexNet [40], and VGG-VD16 [10], as criteria to assess the performance improvement of our proposed method. These models were pre-trained on the ImageNet dataset and then fine-tuned to adapt them to the RSIs. For the open-source AID dataset [34], the training:test ratios were set to 20%/80% and 50%/50%.

4.4. Performance Comparison

Furthermore, in order to further verify the efficiency and feasibility of the proposed algorithm in this paper, the proposed algorithm was also compared with other RSISC algorithms for experimental analysis and validated on the validation set. The results show that the proposed algorithm achieved the best classification result (%), as shown in following Algorithm 1 procedures.

Algorithm 1. Comparison with other RSISC algorithms.

1. **INPUT:** RSI dataset as training data, rest RSI dataset as test data.
 2. **OUTPUT:** Generate the predicted category labels.
 3. Prepare a ResNet 34 network with the autoencoder.
 4. Encode the RSI dataset.
 5. **Traning:** a 4 qubit tensor quantum circuit.
 6. Feed the RSI dataset to the tensor quantum circuit.
 7. **Testing:**
 8. Feed the rest RSI dataset to the tensor quantum circuit.
 9. **STOP ALGORITHM.**
-

As can be seen in Table 2, we can analysis the performance of different approaches on the EuroSAT RGB dataset using various training:test splits. We evaluated four frequently used methods, including BovW, CNN, ResNet50, and GoogleNet, as baselines to assess the performance improvement of the hybrid classical–quantum transferring CNN. We used these pre-trained models on the ImageNet dataset and fine-tuned them on the EuroSAT dataset. For ResNet50 and GoogleNet, we replaced the flattened layer, maintaining the original image sizes. The proposed method exhibited better classification accuracy than other methods and significantly reduced model complexity.

Table 2. Overall accuracy (%) of different training:test ratios on the EuroSAT dataset classification (training ratio = 10% and 20%).

Method (%)	10/90	20/80	Number of Parameters
BoVW (SVM, SIFT, k = 10)	54.54	56.13	—
BoVW (SVM, SIFT, k = 100)	63.07	64.80	—
BoVW (SVM, SIFT, k = 500)	65.62	67.26	—
CNN (two layers)	75.88	79.84	50.5 K
ResNet-50	75.06	88.53	25.6 M
GoogLeNet	77.37	90.97	6.8 M
QCNN	93.65	94.72	18 M
CNN-QNN	94.23	95.07	17 M
Ours	95.81	96.62	21.28 M

To quantitatively evaluate the performance of our proposed method, we adopted the OA and confusion matrix as the evaluation metrics. OA is recorded as the number of correctly classified samples divided by the total number of samples. In the confusion matrix, each column denotes the predicted results, and each row denotes the actual ground objects of the class data. It can display the distribution of each class and can be recommended for the analysis of misclassification results between different classes.

Figure 7 present the evaluation of the proposed method on the EuroSAT dataset in a training and a test set (10/90 ratio). The images of river, highway and herbaceous vegetation misclassified easily as others. As can be seen in Table 2, all CNN algorithms surpassed the BoVW method and the classification results of whole deep CNNs performed better than the classification results of shallow CNNs. Nevertheless, the proposed method achieved a classification result of up to 95.81% in a training and a test set (10/90 ratio) for the EuroSAT RSISC. The optimization process is demonstrated in Figure 8. As can be seen in Figure 9, the images of two scene labels are alike, resulting in imperfect classification results compared with other labels.

Table 3 shows the resulting classification accuracies for the best performing DL models CNN models, GoogLeNet, ResNet50, and the proposed method. In experiments with the GoogLeNet, ResNet50, CNN models, and the proposed method, all models were pre-trained on the EuroSAT dataset. For a better comparison of performance of all fine-tuning models, we trained the last layer with a learning rate between 0.01 and 0.0001. Using the proposed method, we achieved a classification accuracy about 17% better than the other pre-trained model, GoogleNet, which had been trained on the EuroSAT dataset in the same training:test dataset ratio setting.

Furthermore, we compared the proposed algorithm with other pre-trained CNN-based classification methods for different training:test ratios on the RSI dataset. As shown in the table below, the most traditional pre-trained CNN models can achieve under 95% for the training ratios of 20%. The hybrid classical–quantum transferring CNN achieved higher accuracy than other pre-trained CNN-based classification methods. In Table 3, the traditional pre-trained models which achieved high classification accuracy use million RSIs from satellite and aerial platforms and cover a huge amount of scenes and objects around the world, for instance, RingMo [41]. The proposed algorithm can also enhance the classification result for the large scale of input images and performs better than all the other algorithms under the training ratio of 20% and 50%, and “*” indicates the best result among all methods. The best classification results can be obtained with the algorithm that achieves 97.33% and 98.82% for the training ratios of 20% and 50%, respectively. Our proposed approach exhibits the performance of the most advanced methods, * indicates the best result among all methods.

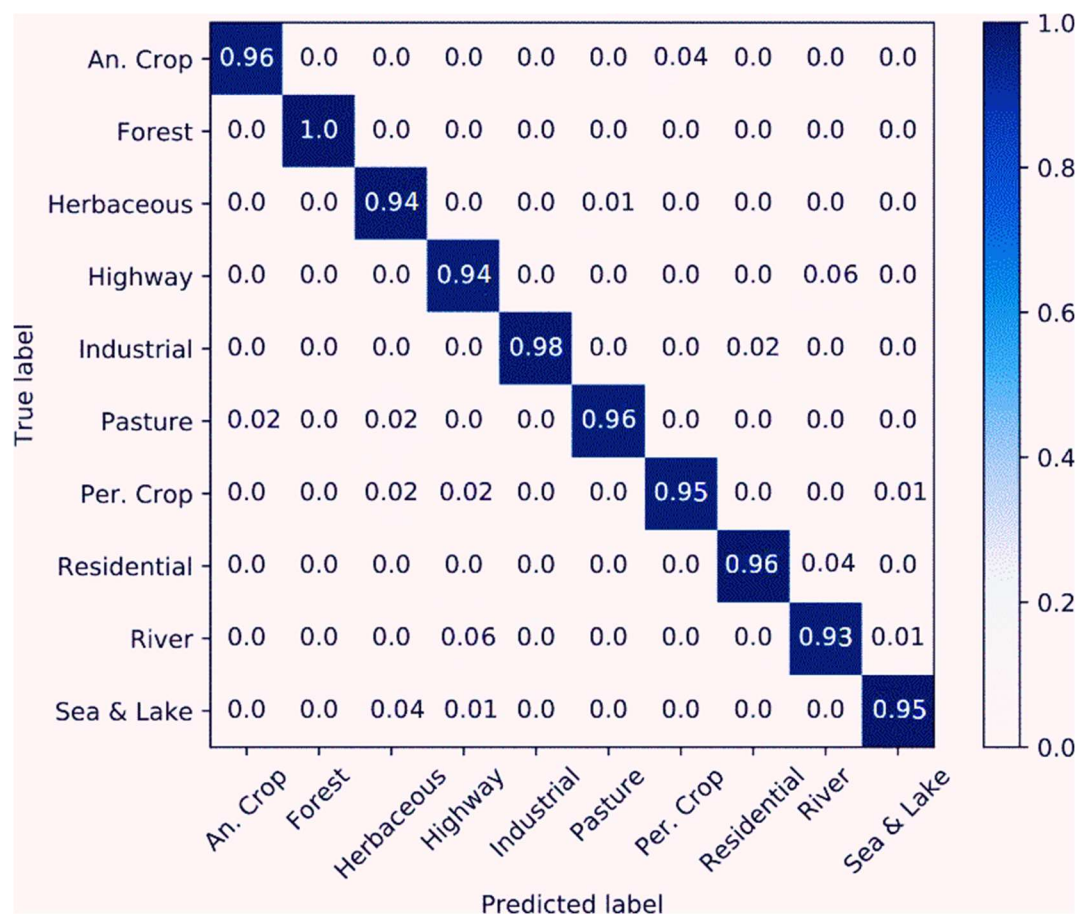


Figure 7. Confusion matrix of the proposed method on the EuroSAT dataset in a training and a test set (10/90 ratio) using RSIs in the RGB color space.

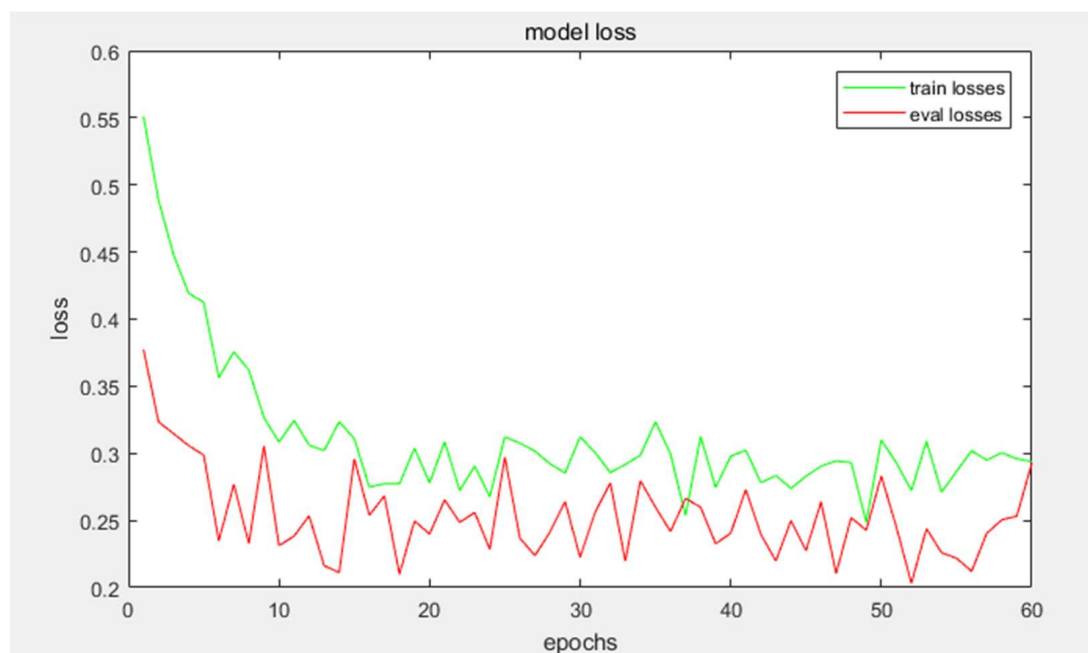


Figure 8. Relation between the loss function and training steps of the proposed method.



Figure 9. The images misclassified as others. (a,b) The images of highways misclassified as rivers. (c,d) The images of rivers misclassified as highways.

Table 3. Contrast of the classification accuracies (%) of different training-test ratios on the AID dataset (training ratio = 20% and 50%).

Method	20/80	50/50
VGG-VD16 [34]	86.59	89.64
GoogLeNet [34]	83.44	86.39
AlexNet + MSCP [42]	88.99	92.36
VGG-VD16 + MSCP [42]	91.52	94.42
AlexNet + SPP [43]	87.44	91.45
RADCNet [44]	88.12	92.53
AlexNet + SAFF [45]	87.51	91.83
VGG-VD16 + SAFF [45]	90.25	93.83
AlexNet + RIR [46]	91.95	94.56
VGG-VD16 + RIR [46]	93.34	95.57
ResNet50 + RIR [46]	94.95	96.48
DCNN [40]	90.82	96.89
CBAM [47]	94.66	96.90
Two-Stream Fusion [48]	92.32	94.58
RTN [49]	92.44	—
GCFs + LOFs [50]	92.48	96.85
CapsNet [51]	91.63	94.74
ARCNet [52]	88.75	93.1
SCCov [53]	93.12	96.1
KFBNet [54]	95.50	97.40
GBNet [55]	92.20	95.48
MG-CAP [56]	93.34	96.12
EAM [57]	94.26	97.06
EAM [57]	93.64	96.62
F ² BRBM [58]	96.05	96.97
MBLANet [59]	95.60	97.14
GRMANet [60]	95.43	97.39
IDCCP [61]	94.80	96.95
MSANet [62]	93.53	96.01
CTNet [63]	96.25	97.70
LSENet [64]	94.41	96.36
DFAGCN [65]	—	94.88
MGML-FENet [66]	96.45	98.60
ESD-MBENet-v1 [67]	96.20	98.85 *
ESD-MBENet-v2 [67]	96.39	98.40
SeCo-ResNet-50 [68]	93.47	95.99
RSP-ViTAEv2-S [69]	96.91	98.22
ISP (ViT) [70]	96.24	97.95
ISSP (ViT) [41]	95.82	97.98
RingMo (ViT-200 W-200 E) [71]	96.54	98.38
ISP (Swin) [70]	96.24	98.03
ISSP (Swin) [72]	96.54	97.95
RingMo (Swin-200 W-200 E) [71]	96.90	98.34
Ours	97.33 *	98.82

We have implemented these networks using a neural network library named Pytorch and a cross-platform Python library for quantum computing named PennyLane [39,73]. The experiments have been operated on a computer graphics workstation equipped with

a single AMD Ryzen Threadripper PRO 3945WX 12-Cores CPU and a single NVIDIA GeForce NVIDIA RTX A5000 24GB GPU.

In this section, we detail various limitations and challenges of this study and how they could potentially be avoided.

1. Lack of access to high-quality, standardized quantum datasets. More work is needed to develop such datasets to properly benchmark our proposed model.
2. Limited types and amount of remote sensing image (RSI) samples currently used. More RSI data from different satellite sources should be collected.
3. More methods for tensor network-based feature extraction should be explored to improve the interpretability and performance.
4. Further research into CQ transferring CNN architectures with sharp priors is needed, as these can help avoid issues like barren plateaus. Architectures like quantum graph neural networks and quantum recurrent neural networks show promise [74].
5. A theory of “quantum geometric deep learning” could help systematically design architectures suited for different quantum datasets, by encoding appropriate symmetries and physical principles.

In summary, the main limitations involve access to high-quality quantum data, limited diversity of current RSI datasets, and the need for better QCNN architectures. Key future work involves developing standardized quantum datasets, collecting more RSI data, exploring tensor network feature extraction, and establishing a theory of quantum geometric DL.

5. Conclusions

Over the past decade, rapid development has provided us with massive remote sensing datasets for intelligent earth observation using RSIs. Nevertheless, the lack of publicly available “big data” of RSIs seriously limits the development of innovative methods, especially traditional DL methods. This work first presents the background of VQC and tensor quantum circuits, and a hybrid classical–quantum transferring CNN applied to RSISC has been proposed. The main advantage of the method in this work is that it allows the input images to be of various shapes and sizes and dispense with the resizing of such images prior to processing. It can hold most key features in RSIs, which is mostly beneficial to conclusively achieving better classification results. In comparison to other deep learning methods with the same number of training samples, the proposed method has achieved amazing scene classification results. Data annotation must be done manually by skilled professionals in the area of RSISC. When a RSI dataset is massive, data annotation can become more complicated because of the huge diversities and variations in RSIs. Most of these models require a large-scale labeled dataset and numerous iterations to train their parameter sets. The experimental results show that our innovative method can achieve satisfied classification results with fewer training samples. This means that in the application of RSISC, the increased workload of data annotation can be reduced manually by skilled professionals. In order to improve classification accuracy, the sum of the CNNs layer has expanded from few layers to hundred layers. Usually, the vast majority of models have huge parameters; moreover, the operation of the CNN models requires an enormity of labeled datasets for model training and powerful equipment with high-performance GPUs for notable performance improvements in operation, which severely limits the development of RSISC methods. However, in comparison with huge CNNs, the proposed method is a compact, lightweight, and efficient RSISC model with fewer parameters. On the other side, the proposed method is extremely expensive and time-consuming, due to the limitations of cross-platform libraries for the differentiable programming of quantum computers. In the future, the method should be improved in quantum circuit simulators with support for automatic differentiation, just-in-time compiling, hardware acceleration, and vectorized parallelism. Especially when the quantum circuit size or the batch dimension is large, the new platform can enable acceleration in quantum circuit simulation.

We should note, however, that our proposed method should be viewed as a new research idea for RSISC, and not be viewed as a replacement for classical pre-trained CNN models. More importantly, a hybrid classical–quantum transferring CNN provides invaluable intuition for constructing a quantum-physics-based DL method. Finally, in terms of future work, we will study process suitability and efficiency assessment in cross-domain RSISC.

Author Contributions: Conceptualization, Z.Z.; methodology, Z.Z.; software, Z.Z. and P.L.; validation, X.M. and Z.Z.; formal analysis, Z.Z.; investigation, Z.Z.; resources, Z.Z.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Z.Z.; visualization, Z.Z.; supervision, X.W. and J.Y. (Jian Yan); project administration, X.M.; funding acquisition, X.G., T.Y., Y.L. and J.Y. (Jian Yang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2019YFE0127300) and the Major Project of High Resolution Earth Observation System (Grant No. 30-Y60B01-9003-22/23).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Tianhai Cheng and Chen Li (Aerospace Information Research Institute, Chinese Academy of Sciences) for their valuable comments and contributions to enhance their letter.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Otgon, S.; Datcun, B.M. Classification of Remot Sensing Images with Parameterized Quantum Gates. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 154–196.
- Zhu, X.X.; Tuia, D. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
- Ma, L.; Liu, Y. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [\[CrossRef\]](#)
- Wang, W.; Chen, Y. Transferring CNN with Adaptive Learning for Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [\[CrossRef\]](#)
- Chen, G.; Wei, J. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [\[CrossRef\]](#)
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *21*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
- Broni-Bediako, C.; Murata, Y.; Mormille, L.G.B. Searching for CNN Architectures for Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4701813. [\[CrossRef\]](#)
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Curran Associates: Red Hook, NY, USA, 2012; pp. 1097–1105.
- Sun, Y.; Xue, B.; Zhang, M.; Yen, G.G. Completely automated CNN architecture design based on blocks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 1242–1254. [\[CrossRef\]](#)
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Proceedings of the Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems (NIPS)*, Stateline, NV, USA, 3–6 December 2012; Volume 25, pp. 1097–1105.
- Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014*; pp. 818–833.
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Richards, J.A.; Jia, X. Sources and characteristics of remote sensing image data. In *Remote Sensing Digital Image Analysis Image Analysis: An Introduction*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 1–38.

15. Coelho, J. Solve Any Image Classification Problem Quickly and Easily. Available online: <https://github.com/pmarcelino/blog> (accessed on 20 April 2023).
16. Cheng, G. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [[CrossRef](#)]
17. Helber, P.; Bischke, B.; Dengel, A. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [[CrossRef](#)]
18. Shor, P.W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **1997**, *26*, 1484.
19. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, UK, 2010.
20. Jeong, Y.; Yu, J. Bulk scanning method of a heavy metal concentration in tailings of a gold mine using SWIR hyperspectral imaging system. *Int. J. Appl. Earth Obs.* **2021**, *102*, 102382. [[CrossRef](#)]
21. Yue, T.; Liu, Y.; Du, Z. Quantum machine learning of eco-environmental surfaces. *Sci. Bull.* **2022**, *67*, 1031–1033. [[CrossRef](#)] [[PubMed](#)]
22. Shaik, R.U.; Unni, A.; Zeng, W. Quantum Based Pseudo-Labeling for Hyperspectral Imagery: A Simple and Efficient Semi-Supervised Learning Method for Machine Learning Classifiers. *Remote Sens.* **2022**, *14*, 5774. [[CrossRef](#)]
23. Mari, A.; Bromley, T.R.; Izaac, J. Transfer learning in hybrid classical-quantum neural networks. *arXiv* **2019**, arXiv:1912.08278. [[CrossRef](#)]
24. Qi, J.; Tejedor, J. Classical-to-Quantum Transfer Learning for Spoken Command Recognition Based on Quantum Neural Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ICASSP 2022, Singapore, 22–27 May 2022.
25. Dunjko, V.; Briegel, H.J. Machine learning & artificial intelligence in the quantum domain: A review of recent progress. *Rep. Prog. Phys.* **2018**, *81*, 074001.
26. Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P.J.; Aspuru-Guzik, A.; O'Brien, J.L. A variational ei-genvalue solver on a photonic quantum processor. *Nat. Commun.* **2014**, *5*, 4213. [[CrossRef](#)]
27. Farhi, E.; Goldstone, J.; Gutmann, S. A Quantum Approximate Optimization Algorithm. *arXiv* **2014**, arXiv:1411.4028.
28. Cai, X.; Li, D.Z.; Liu, X. Experimental Realization of a Quantum Support Vector Machine. *Phys. Rev. Lett.* **2015**, *114*, 140504.
29. Maria, S. Quantum machine learning models are kernel methods. *arXiv* **2021**, arXiv:2101.11020.
30. Havlíček, V.; Córcoles, A.D. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212. [[CrossRef](#)] [[PubMed](#)]
31. Huang, H.-Y.; Broughton, M.; Mohseni, M.; Babbush, R.; Boixo, S.; Neven, H. Power of data in quantum machine learning. *Nat. Commun.* **2021**, *12*, 1–9. [[CrossRef](#)] [[PubMed](#)]
32. Cade, C.; Mineh, L.; Montanaro, A.; Stanisic, S. Strategies for solving the Fermi-Hubbard model on near-term quantum computers. *Phys. Rev. B* **2020**, *102*, 235122. [[CrossRef](#)]
33. Deng, J.; Dong, W.; Socher, R. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
34. Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [[CrossRef](#)]
35. Henderson, M.; Shakya, S.; Pradhan, S.; Cook, T. Quantvolutional neural networks: Powering image recognition with quantum circuits. *Quantum Mach. Intell.* **2020**, *2*, 2. [[CrossRef](#)]
36. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June 2016; pp. 2818–2826.
38. Yang, C.-H.H.; Qi, J.; Chen, S.Y.-C.; Chen, P.; Siniscalchi, S.M.; Ma, X.; Lee, C.-H. Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 6523–6527.
39. Pointer, I. *Programming PyTorch for Deep Learning: Creating and Deploying Deep Learning*; O'Reilly: Sebastopol, CA, USA, 2019.
40. Cheng, G.; Yang, C.; Yao, X.; Guo, L.; Han, J. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2811–2821. [[CrossRef](#)]
41. Xie, Z.; Zhang, Z.; Cao, Y.; Lin, Y.; Bao, J.; Yao, Z.; Dai, Q. Simmim: A simple framework for masked image modeling. *arXiv* **2021**, arXiv:2111.09886.
42. He, N.; Fang, L.; Li, S.; Plaza, A.; Plaza, J. Remote sensing scene classification using multilayer stacked covariance pooling. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6899–6910. [[CrossRef](#)]
43. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sens.* **2017**, *9*, 848. [[CrossRef](#)]
44. Bi, Q.; Qin, K.; Zhang, H.; Li, Z.; Xu, K. RADC-Net: A residual attention based convolution network for aerial scene classification. *Neurocomputing* **2020**, *377*, 345–359. [[CrossRef](#)]

45. Cao, R.; Fang, L.; Lu, T.; He, N. Self-Attention-Based Deep Feature Fusion for Remote Sensing Scene Classification. *IEEE Geosci. Remote. Sens. Lett.* **2021**, *18*, 43–47. [\[CrossRef\]](#)
46. Kunlun, Q.; Chao, Y.; Chuli, H.; Yonglin, S.; Shengyu, S.; Huayi, W. Rotation invariance regularization for remote sensing image scene classification with convolutional neural networks. *Remote Sens.* **2021**, *13*, 569.
47. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
48. Yu, Y.; Liu, F. A two-stream deep fusion framework for high-resolution aerial scene classification. *Comput. Intell. Neurosci.* **2018**, *2018*, 8639367. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Chen, Z.; Wang, S.; Hou, X.; Shao, L.; Dhabhi, A. Recurrent transformer network for remote sensing scene categorisation. *BMVC* **2018**, *266*, 1–11.
50. Zeng, D.; Chen, S.; Chen, B.; Li, S. Improving remote sensing scene classification by integrating global context and local-object features. *Remote Sens.* **2018**, *10*, 734. [\[CrossRef\]](#)
51. Zhang, W.; Tang, P.; Zhao, L. Remote sensing image scene classification using cnn-capsnet. *Remote Sens.* **2019**, *11*, 494. [\[CrossRef\]](#)
52. Wang, Q.; Liu, S.; Chanussot, J.; Li, X. Scene classification with recurrent attention of vhr remote sensing images. *IEEE Trans. Actions Geosci. Remote Sens.* **2018**, *57*, 1155–1167. [\[CrossRef\]](#)
53. He, N.; Fang, L.; Li, S.; Plaza, J.; Plaza, A. Skipconnected covariance network for remote sensing scene classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 1461–1474. [\[CrossRef\]](#)
54. Li, F.; Feng, R.; Han, W.; Wang, L. High-resolution remote sensing image scene classification via key filter bank based on con-volutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8077–8092. [\[CrossRef\]](#)
55. Sun, H.; Li, S.; Zheng, X.; Lu, X. Remote sensing scene classification by gated bidirectional network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 82–96. [\[CrossRef\]](#)
56. Wang, S.; Guan, Y.; Shao, L. Multi-granularity canonical appearance pooling for remote sensing scene classification. *IEEE Trans. Image Process.* **2020**, *29*, 5396–5407. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Zhao, Z.; Li, J.; Luo, Z.; Li, J.; Chen, C. Remote sensing image scene classification based on an enhanced attention module. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1926–1930. [\[CrossRef\]](#)
58. Zhang, X.; An, W.; Sun, J.; Wu, H.; Zhang, W.; Du, Y. Best representation branch model for remote sensing image scene clas-sification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9768–9780. [\[CrossRef\]](#)
59. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [\[CrossRef\]](#)
60. Li, B.; Guo, Y.; Yang, J.; Wang, L.; Wang, Y.; An, W. Gated recurrent multiattention network for vhr remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5606113. [\[CrossRef\]](#)
61. Wang, S.; Ren, Y.; Parr, G.; Guan, Y.; Shao, L. Invariant deep compressible covariance pooling for aerial scene categorization. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6549–6561. [\[CrossRef\]](#)
62. Zhang, G.; Xu, W.; Zhao, W.; Huang, C.; Yk, E.N.; Chen, Y.; Su, J. A multiscale attention network for remote sensing scene images classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9530–9545. [\[CrossRef\]](#)
63. Deng, P.; Xu, K.; Huang, H. When cnns meet vision transformer: A joint framework for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [\[CrossRef\]](#)
64. Bi, Q.; Qin, K.; Zhang, H.; Xia, G.-S. Local semantic enhanced convnet for aerial scene recognition. *IEEE Trans. Image Process.* **2021**, *30*, 6498–6511. [\[CrossRef\]](#)
65. Xu, K.; Huang, H.; Deng, P.; Li, Y. Deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5751–5765. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Zhao, Q.; Lyu, S.; Li, Y.; Ma, Y.; Chen, L. Mgml: Multigranularity multilevel feature ensemble network for remote sensing scene classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 2308–2322. [\[CrossRef\]](#)
67. Zhao, Q.; Ma, Y.; Lyu, S.; Chen, L. Embedded selfdistillation in compact multi-branch ensemble network for remote sensing scene classification. *arXiv* **2021**, arXiv:2104.00222.
68. Ma, O.; Lacoste, A.; Nieto, X.G.-I.; Vazquez, D.; Rodriguez, P. Seasonal contrast: Unsupervised pre-training from uncured remote sensing data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 9414–9423.
69. Wang, D.; Zhang, J.; Du, B.; Xia, G.-S.; Tao, D. An empirical study of remote sensing pretraining. *IEEE Trans. Geo Sci. Remote Sens.* **2022**, *61*, 5608020. [\[CrossRef\]](#)
70. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
71. Sun, X.; Wang, P.; Lu, W.; Zhu, Z.; Lu, X.; He, Q.; Li, J.; Rong, X.; Yang, Z.; Chang, H.; et al. RingMo: A remote sensing foundation model with masked image modeling. *IEEE Trans. Geosci. Remote Sens.* **2022**, *61*, 1–22. [\[CrossRef\]](#)
72. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.

73. Bergholm, V.; Izaac, J.; Schuld, M.; Gogolin, C.; Ahmed, S.; Ajith, V.; Alam, M.S.; Alonso-Linaje, G.; Narayanan, B.A.; Asadi, A.; et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations. *arXiv* **2018**, arXiv:1811.04968.
74. Cerezo, M.; Verdon, G.; Huang, H.Y.; Cincio, L.; Coles, P.J. Challenges and opportunities in quantum machine learning. *Nat. Comput. Sci.* **2022**, *2*, 567–576. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.