Hindawi
Quantum Engineering
Volume 2023, Article ID 2842217, 11 pages
https://doi.org/10.1155/2023/2842217

Research Article

Robustness and Explainability of Image Classification Based on QCNN

Guoming Chen,[1] **Shun Long**,[2] **Zeduo Yuan**,[2] **Wanyi Li**,[1] **and Junfeng Peng**[1]

[1]*School of Computer Science, Guangdong University of Education, Guangzhou, Guangdong 510303, China*
[2]*Department of Computer Science, Jinan University, Guangzhou, Guangdong 510632, China*

Correspondence should be addressed to Guoming Chen; isscdz@mail.sysu.edu.cn

Received 24 September 2022; Revised 30 October 2022; Accepted 17 December 2022; Published 17 January 2023

Academic Editor: YuBo Sheng

In this paper, we propose a multiscale entanglement renormalization ansatz (MERA) feature extraction method based on a novel quantum convolutional neural network (QCNN) for binary scanning tunneling microscopy (STM) image classification. We design QCNN quantum circuits for state preparation, quantum convolution, and quantum pooling in the TensorFlow quantum framework and compare the performance of QCNN classifier and two hybrid quantum-classical QCNN models. Adversarial attacks are considered as a type of interpretable method to evaluate the robustness of QCNN models. The similarity between the pixels of image bitplane slicing and Ising phase transition opens up new ways for exploring classification performance enhancement by QCNN classifiers. Classification performance of different bitplanes of QCNN also shows that they can robustly resist adversarial attacks such as FGSM, CW, JSMA, and DEEPFOOL.

1. Introduction

In recent years, many researchers focused on the interplay between machine learning and quantum physics and investigated if quantum technology can help to improve traditional machine learning algorithms such as supervised learning, principal component analysis, and other dimension reduction algorithms [1–4]. Cong et al. [5] proposed a quantum circuit-based convolutional neural network (CNN) which can accurately recognize quantum states. Kossaifi and Bulat [6] parameterized the global CNN with a single higher-order tensor. Tensor methods have the potential to parameterize network structure representations in a compact manner. Via imposing a low rank structure on the tensor, it can regularize the network, reduce the number of parameters, and obtain higher accuracy and compression. Henderson et al. [7] proposed a quantum convolution layer with a number of random quantum circuits for feature extraction in image classification. Broughton and Verdon [8] proposed a software framework for quantum machine learning where quantum circuits for supervised learning in classification is make up of a sequence of quantum gates. Quantum circuits play an important role in machine learning [9–12]. More works of development in QCNN are still be carried out. Henderson et al. [7] evaluated CNNs, QCNN and CNNs with additional nonlinearity models on the MNIST dataset. They showed that the QCNN model had better accuracy as well as faster convergence when compared to the purely classical CNNs. Wei et al. [13] proposed a QCNN model for recognition of handwritten numbers and simulated three types of image filtering, smoothing, sharpening, and edge detection. It could reduce the computing complexity compared with CNNs. Chen et al. [14] proposed a QCNN model for the classification of high energy physics events. It demonstrated an advantage of learning faster than the CNNs. Li et al. [15] proposed a quantum-classical hybrid processing model inspired by the variation quantum algorithms on the MNIST and GTSRB datasets and verified the feasibility and validity when compared with CNNs.

However, the complex QCNN with features as a black box cannot explain its internal mechanism well, so verifying the robustness is crucial toward the trustworthy QCNN. In order for QCNN to be trusted, it needs to reliably explain why QCNN makes certain predictions. Traditional robust and explainable methods [16, 17] on classifier has the limitations of not revealing the intrinsic mechanism where general image bitplane slicing has similar phase transition pattern to that of the Ising models, and the adversarial attacked image has minor visual differences but the bitplanes of attacked image show significant differences. Our scheme is explainable that the accuracies of QCNN models with local construct features on some bitplanes also reveal the corresponding differences under different attacks with different level of attack intensity. With the arrival of noisy intermediate-scale quantum (NISQ) era, to solve these issues, an explainable anti-adversarial attacks image classification scheme is proposed in this paper. First, we simulate adversarial attacks and inject the perturbations on the input data. Then, the bitplane slicing and a feature extraction method are both employed to the novel QCNNs. Finally, the robustness of classification performance evaluation are utilized to evaluate the model interpretation that some bitplanes give over approximation of robust accuracy while other bitplanes give under approximation of robust accuracy. Adversarial attacks [18–21] inject imperceptible perturbations to images and lead to deterioration of performance in deep image classifiers, it raise security concerns of image classification. The typical adversarial attacks include FGSM, JSMA, CW, DEEPFOOL, etc. Attackers can inject perturbations to specific objects, or inject imperceptible noise to the background, or inject perturbations to the whole image, the strength of the attack depends on specific parameters. Adversarial attacks are considered as a type of interpretable method, that is, the classification results of normal samples and adversarial samples can be analyzed and reasoned by different features, which can assist scientists to design a more appropriate structure of network. Adversarial attacks and different features are applied to study the robustness of interpretations for our QCNNs. The contribution of this paper are as follows: (1) We propose a MERA feature extraction method on our new designed quantum convolutional neural networks (QCNNs) for STM image classification. (2) We discover image bitplane slicing has similar phase transition pattern as that of the Ising model and explore the correlation between this pattern and the classification performance enhancement by QCNN classifiers. (3) For robustness and explainability, the classification performance of different bitplanes of QCNN also shows that they can robustly resist adversarial attacks such as FGSM, CW, JSMA and DEEPFOOL.

2. Local MERA Feature Construction

To improve the transparency of QCNNs, the proposed explanations provide the local construct features and the global QCNN framework to enhance the understanding of classifiers. Building on the recent interest in tensor networks for machine learning [22, 23], tensor networks have been a tool for the analysis of quantum many-body systems, it encodes the coefficients of the state wave function, ensembles of microstates and is superior to dimension reduction. Tensor networks can be interpreted as part of linear classifiers operating in exponentially high dimensional spaces to be useful in image analysis application and measure the scale of particles/pixels with degrees of granulation. This granularity can be distilled and encoded into a global QCNN.

Multiscale entanglement renormalization ansatz (MERA) and discrete wavelet transformations have a similar multiscale representation. Tensor networks can be used for physical states classification and simulating entangled correlated systems. Such correlation states can be simulated with multilevel analysis for extracting local features. Hallam and Grant [24] proposed a method for tensorizing neural networks by way of approximating scale invariant quantum states. They employed MERA as a replacement for the fully connected layers in a convolutional neural network on the CIFAR datasets. The proposed method provides great compression for the same level of accuracy and great accuracy for the same level of compression. MERA is a powerful tool to study phase transition, critical phenomena and strong coupling problems. In deep learning, people have observed that deep neural networks have the ability to extract features layer by layer. Inspired by the fact that general image bitplane slicing has a similar phase transition pattern to that of the Ising model. With granularity at different scales, we can explore the distinguishing ability of generated features and convert the coarse-grained Ising phase/state classification into fine-grained (pixel-level) image classification.

The scale-invariant MERA provides an efficient way to extract scaling operators. Unitary gates with reflection symmetry in MERA quantum circuits are scale representation of quantum many-body wave function which structurally similar to mappings of convolutional networks and MERA [25, 26] can encode correlations between different scales for data compression. Equation (1) is a 2×2 unitary matrix represented as U_{sw} with some reflections.

$$U_{sw} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{1}$$

Equation (2) is a 3×3 unitary matrix with one parameter of reflection symmetric matrices $v(\theta)$.

$$v(\theta) = \exp\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & \theta & 0 \\ -\theta & 0 & -\theta \\ 0 & \theta & 0 \end{bmatrix}\right). \tag{2}$$

It can also be denoted as follows:

$$v(\theta) = \frac{1}{2}\begin{bmatrix} \cos(\theta)+1 & \sqrt{2}\sin(\theta) & \cos(\theta)-1 \\ -\sqrt{2}\sin(\theta) & 2\cos(\theta) & -\sqrt{2}\sin(\theta) \\ \cos(\theta)-1 & \sqrt{2}\sin(\theta) & \cos(\theta)+1 \end{bmatrix}. \tag{3}$$

The unitary circuit is formed by 3×3 reflection symmetric matrices $v(\theta)$ with the swap gate U_{sw} and a symmetric transforms parameter dilation factor three. The representation of Figure 1 is a scale layer unitary circuit with three rotation angles θ as given below.

V is a $3\,N \times 3\,N$ matrix with parameter N in a scale layer unitary circuit, the decomposition form is,

$$V = V_3 U_{sw} V_2 U_{sw} V_1. \tag{4}$$

V_k is direct sum of N matrices $v(\theta_k)$ from (3),

$$Vk = v(\theta_k) \oplus v(\theta_k) \oplus v(\theta_k) \oplus v(\theta_k) \ldots . \tag{5}$$

The scale layer unitary circuit has parameters θ_1, θ_2, and θ_3. Right, left of an edge-centered, and site-centered symmetric sequence are entered into this circuit. The multiscale circuit which encodes the images and its output are then be chosen to yield the ten output features. There is a connection between MERA quantum circuits and discrete wavelet transforms. We describe how MERA quantum circuits can be exploited to develop a new feature extraction method; the process is similar to features extraction from the wavelet transformation of the given image. From MERA circuits, we can distill features from an image and they can be integrated into the QCNN.

3. Quantum Convolutional Neural Network

Quantum convolutional neural network (QCNN) [27] can recognize specific features of quantum states. It is significant to study the combination of local features and the global QCNN circuit structure and that of the bidirectional contributions. Different from the previous work and recent advances, we study the information distil ability on scale patterns as local correlation features to be integrated into global QCNNs and spread into the entire unitary evolution system instead of parameterizing the whole network with single tensors. As many-body wave-functions are structurally similar to mappings of convolutional networks, we analyze the transformation classification pattern of the physical state/phase into the learning of traditional image classification. It is critical to prepare quantum initial states where the higher entangled state correspond to higher separated weight function. With entangled state, the QCNN would have more expressive power than its classical counterpart.

$$U|\psi_0\rangle = U \sum_{i=0}^{2^n-1} \alpha_i \Big| i\rangle = \sum_{i=0}^{2^n-1} \beta_i \Big| i\rangle. \tag{6}$$

In a quantum system, an initial state $|\psi_0\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle, \alpha \in \mathbb{C}$ where $\{\langle i, i = 0, 2, \ldots, 2^n - 1\}$ denotes a set of bases in the Hilbert space, $\alpha_i, \beta_i \in \mathbb{C}$. The QCNN applies the unitary transformation U on it. Quantum circuit operates quantum bits form by quantum logic gates which are building blocks of quantum circuits. They are combined to form a global quantum circuit, and the whole quantum circuit is a large unitary matrix. It is critical to find a good set of parameters for the quantum circuit like activation function in the network. The information can be distilled through MERA based local features according to different data distributions.

The QCNN architectures for image classification task are illustrated in Figure 2. In this architecture, the first layer is quantum cluster state prepare layer which is shown in Figure 3. Where H gate is applied to any of its qubits indicates an excitation and CZ gate is applied to any of the two adjacent qubits to get the highly entangled state.

The second layer is the input layer where the encoded MERA features are distilled as the rotation angles θ of single-qubit RX, RY, RZ gates and the rotation angles θ of the two qubit XX, YY, ZZ gates. The transformation of encoded features parameters enter into the parameterized unitary circuit where XX is supposed to be tensor product of X with X with rotation angles θ. Equations (7)–(9) define RX, RY, RZ gates in the circuit, when the parameters enter into the input layer, and they decide the rotation angle around the X, Y and Z axis in Bloch sphere. The gradient of the QCNN is relatively smooth, so local MERA features are vital to adjusting the gradient and exploring the correlation between scale dimension reduced features and the gradient. The cluster state prepare layer and input layer are added to the quantum circuit in order.

$$R_x(\theta) \equiv e^{-i(\theta/2)X} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} X = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \tag{7}$$

$$R_y(\theta) \equiv e^{-i(\theta/2)Y} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} Y = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \tag{8}$$

$$R_z(\theta) \equiv e^{-i(\theta/2)Z} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} Z = \begin{bmatrix} e^{-(i\theta/2)} & 0 \\ 0 & e^{(i\theta/2)} \end{bmatrix}. \tag{9}$$

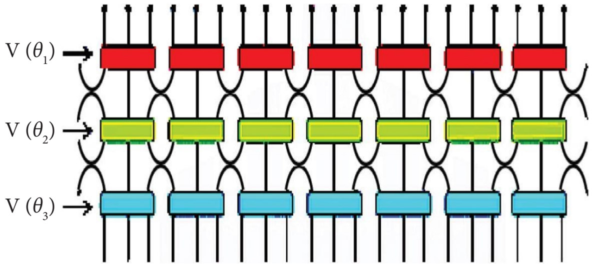


Figure 1: Scale layer unitary circuit.

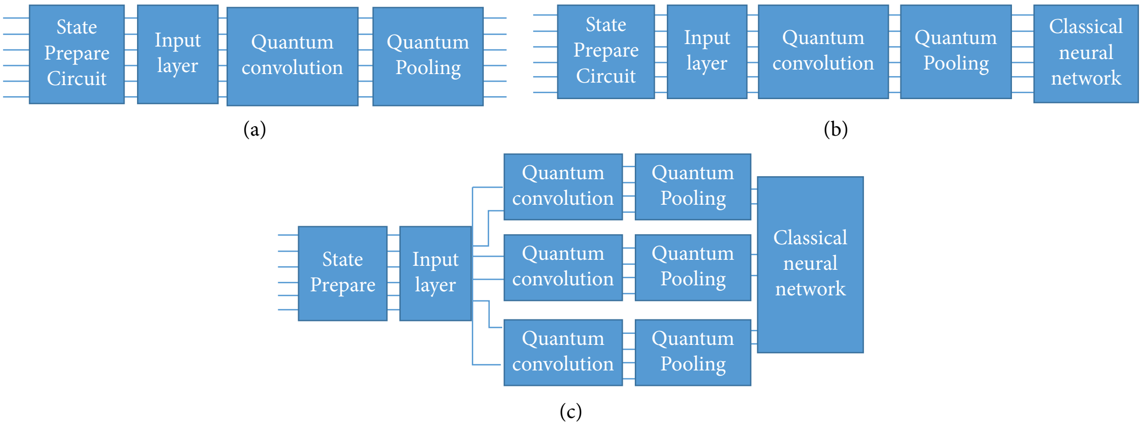


Figure 2: QCNN and hybrid QCNNs with classical architecture. (a) QCNN, (b) hybrid QCNN, and (c) hybrid QCNN with multiple quantum.

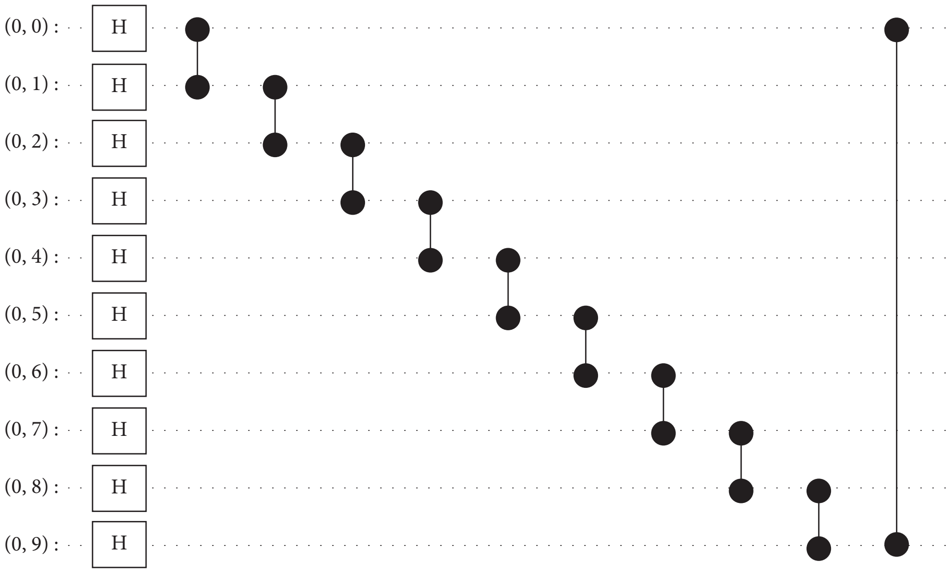


Figure 3: State prepare circuit.

One and two qubit parameterized unitary matrices construct the convolution and pooling layers. The third layer is quantum convolution layer. Figure 4 depicts RX, RY, RZ and XX, YY, ZZ gates in quantum convolution layer that can be constructed by a cascade of two-qubit parameterized unitary with pairs of adjacent qubits. The last layer is quantum pooling layer. Figure 5 depicts RX, RY, RZ and CNOT gates in quantum pooling layer. CNOT gates are used to control entanglement. Two arbitrary qubit unitary make a parameterized pooling from two qubits to one qubit unitary circuit. The quantum pooling layer pools half of the qubits by two-qubit pooling. The pooling layer output the qubits where the label 1 assigned one state while −1 assigned the other state. The pooling layer is followed by the repeated measurement observable Z on state $|\phi\rangle$ which is denoted as $\langle Z\rangle_{|\phi\rangle} \equiv \langle\phi|Z|\phi\rangle = |\alpha|^2 - |\beta|^2$ where $\langle Z\rangle \in [-1, 1]$. In QCNN architecture, image pixels are not suitable features entering into quantum circuit for classification, we have made MERA features as parameters of QCNN.

Figure 2(a) depicts a QCNN architecture is constructed by cluster state prepare layer, input layer, convolution, pooling and measurement layers. Figure 2(b) illustrates the hybrid QCNN model which combines a classical neural network with a single quantum convolution and pooling layer. Figure 2(c) illustrates the hybrid QCNN with multiple quantum which combines multiple quantum convolutions and pooling layer with a classical neural network. In recent years, more and more researchers concern about how to improve the performance of the deep network, it should not only pay attention to the depth of the network, an opposite direction of neural network by expanding the width instead of increasing the depth, called broad learning system should be worthy of attention. The difference between hybrid QCNNs and hybrid QCNNs with multiple quantum is that the width of the hybrid QCNNs with multiple quantum is expanded wider than the hybrid QCNNs. In some of the bitplanes, the difference of their performance will increase when compared to the original image, which mean that there is still difference in their anti-adversarial attack ability.

4. Experimental Results

In order to verify the effectiveness and demonstrate the interpretation of our proposed QCNN classifier based on MERA features. We implement three sets of experiment in an environment of TensorFlow-quantum 0.3.0 and cirq 0.8.0. The first experiment includes a dataset of 7589 scanning tunneling microscopy (STM) images [28], labeled as acquired either with a good or bad probe. STM images including 1761 images of good probe and good image which are labeled as class 1. 5828 images of bad ones, with an imperfect acquisition (e.g., inadequate sample region or coarse in sample, noisy image without probe sample contact; blurry images with dull probe, replicated images with multiple-feature probe; artifact with contaminated probe), which are labeled as class 2. In these experiments, the MERA features and Box-counting fractal features [29] were normalized to $[-\pi, \pi]$ as parameters of rotation angles in RX, RY, RZ gates and XX, YY, and ZZ gates and then feed into parameterized quantum circuits. This combination of encoded local scale features and QCNN better demonstrates the multiscale nature of data distribution.

A performance comparison of the three models: QCNN, hybrid QCNN, and hybrid QCNN with multiple quantum layers (horizontal expansion of hybrid QCNN by increasing the width) suggests that improvement have been achieved via the proposed features. Parts of the STM images are shown as Figure 6(a), 60% of these images are also randomly selected for training and the remaining 40% for validation. In Figure 6(b), QCNN, Hybrid QCNN, and Hybrid QCNN with multiple quantum layers have achieved accuracies of about 75%–97%, the convergence rate improvement of pure QCNN has much room for improvement when compared with the other two quantum classical hybrid models. Better convergence is expected for all three QCNN models, particularly the pure QCNN one.

MERA and fractal features have similarities in multiscale image analysis and representation methods; it is important to study the characteristics of images at various scales. By multiscale decomposition, the image information is distilled, which triggers improvement of the QCNN performance. Comparisons have been made between the MERA and Box-counting fractal features in our QCNN model and each feature shows some advantage respectively. Table 1 shows the classification performance comparison between the MERA and Box-counting fractal features. The accuracies of boxcounting fractal outperforms MERA 98.44% vs 95.31% and they must be accompanied by that the accuracies of some high-order bitplanes of MERA outperform those of boxcounting fractal features.

Ising model [30] can depicts the phase transition of ferromagnetic materials. When heated over some temperature threshold, the system loses its magnetism temporarily until cooled down to that threshold. The transition between magnetic and non-magnetic phases is called phase transition. The Monte Carlo method and Ising model-based metropolis algorithm are used to generate images which is shown in Figure 7(b). Granularity distribution with different scale can be used in classification, accompanied by different position distribution of the pixel values. Figure 7(a) depicts that the image bitplane decomposition has a similar phase transition to that of the Ising model. It is particularly important to study the relationship between phase transition and classification performance and gives an interpretation of why it achieve better performance.

Network structure and parameter adjustment help to improve performance. It suggests that the phase transition of the original image is universal. Especially, this phase transition is more likely to have a strong correlation with the classification performance. In the second set of experiments, image pixels are treated as physical particles. To demonstrate the effectiveness of our QCNN models. We first used the Monte Carlo method and Ising method-based Metropolis algorithm to simulate 10000 images with 100×100 pixels which indicate ten different scale levels of granularity. They are shown in Figure 7(b). For these 10000 Ising scale images, a pre-defined number of clustering has been followed in order to fix the number of categories to 10. The [0-0.1], (0.1-0.2], (0.2-0.3], (0.3-0.4], (0.4-0.5], (0.5-0.6], (0.6-0.7], (0.7-0.8], (0.8-0.9], (0.9-1] region have been labeled with 1 to 10, each corresponding to a category representing one of the ten different scale of granularity from top to bottom in Figure 7(b), i.e., from fine-grained to coarse-grained image granularity. QCNN built on various scale granules can make use of granularity and scale for classification. We divided ten different scale Ising images into five groups [0-0.1] and (0.9-1], (0.1-0.2] and (0.8-0.9], (0.2-0.3] and (0.7-0.8], (0.3-0.4] and (0.6-0.7], and (0.4-0.5] and (0.5-0.6] and performed binary classification, respectively. QCNN, hybrid QCNN, and hybrid QCNN with multiple quantum layers with different network structures have achieved accuracies of about 60%–97% on groups [0-0.1] and (0.9-1]. Accuracies of

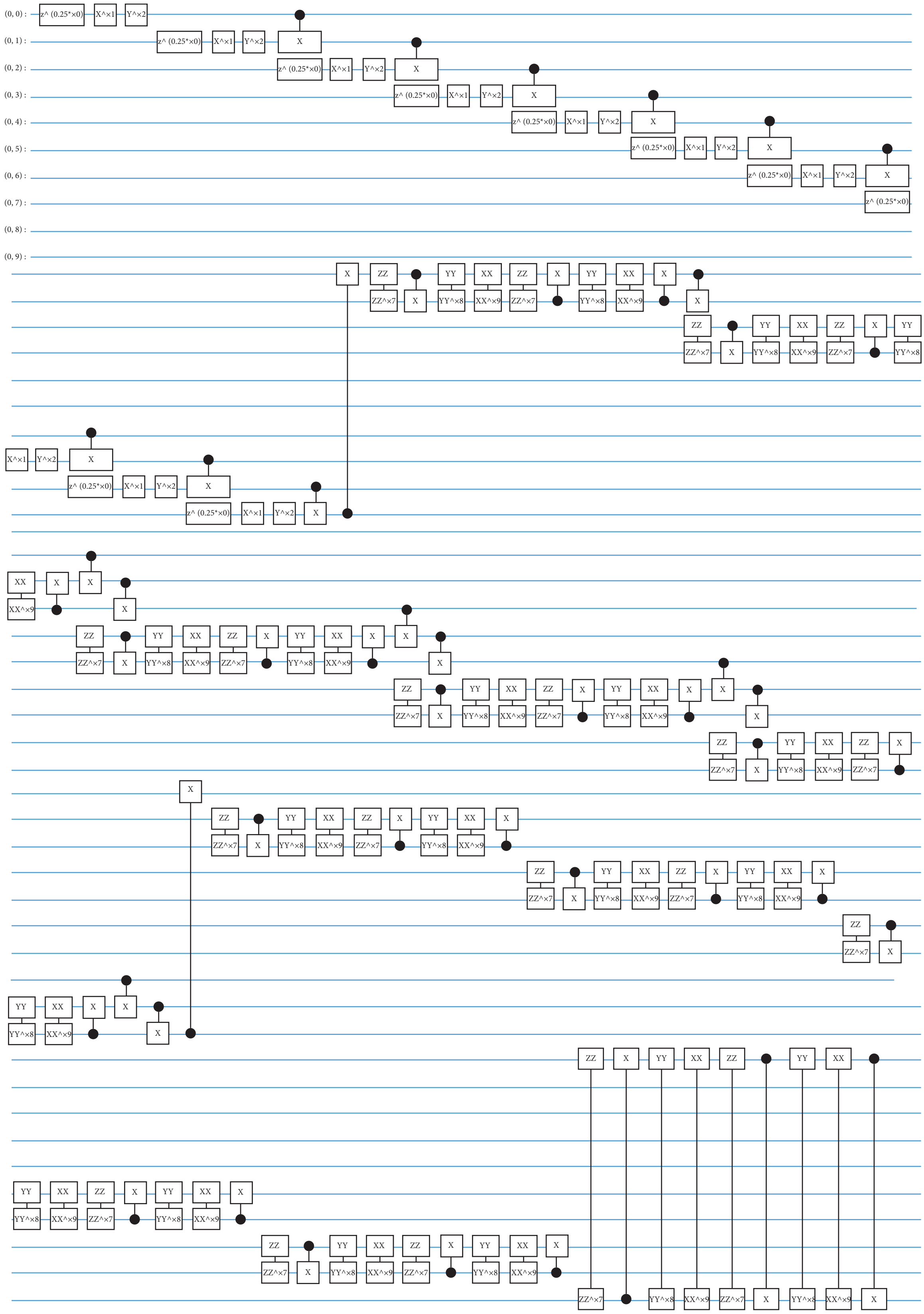


Figure 4: Quantum convolution.

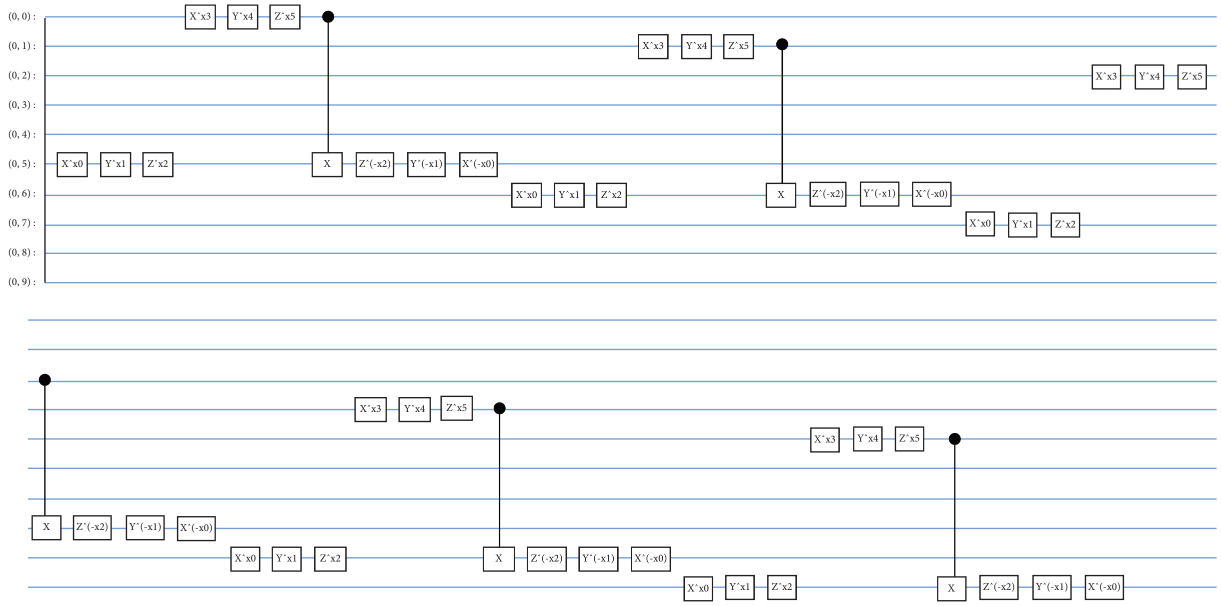
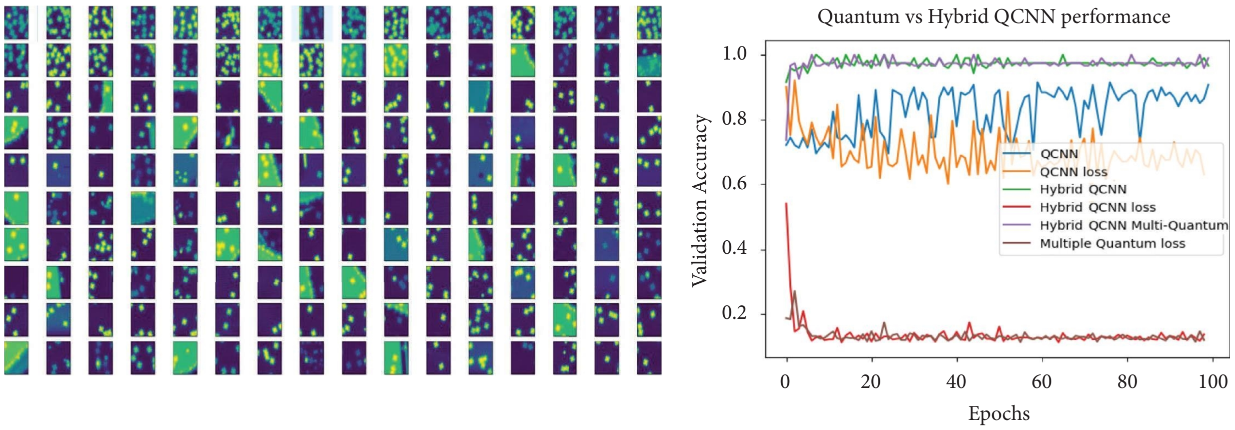


FIGURE 5: Quantum pooling.



(a)

(b)

FIGURE 6: Binary classification on STM images. (a) The STM dataset and (b) QCNN recognition accuracy.

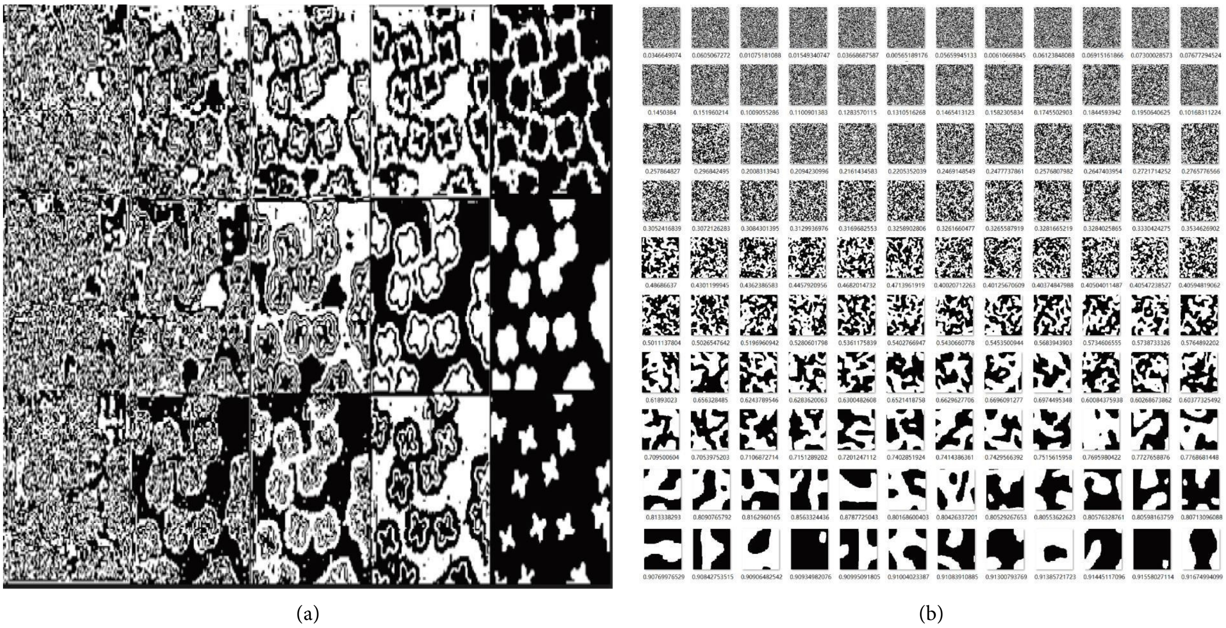
about 70%–97% have been achieved on groups (0.1-0.2] and (0.8-0.9]. Accuracies of about 74%–95% have been achieved on groups (0.2-0.3] and (0.7-0.8]. Accuracies of about 65%–97% have been achieved on groups (0.3-0.4] and (0.6-0.7], accuracies of about 75%–97% have been achieved on groups (0.3-0.4] and (0.6-0.7]. They help to explore the relation between the quantization of pixels of the image and quantum particles.

In the third set of experiments, we evaluate the robustness and explainability of QCNNs by exploring the bitplanes and their antiadversarial attacks classification performance. We simulate the fast gradient sign method (FGSM) attack on STM to generate adversarial samples, which exploits the maximum direction of gradient changes in the network to inject perturbation noise to make the model deteriorate under the attack. Figure 8 depicts the original image and the adversarial sample images generated by FGSM adversarial attack with different strengths on STM. The attacked image has minor visual differences but the bitplanes patterns under attacks show significant differences and distortions. The first column from top to bottom is the original image, the associated 8th, 7th, 6th, and 5th bitplanes of the original image; the second column is attacked image by FGSM attack with the attack strength parameter eps which is set to 12/255 and its associated sliced bit-planes; the third column is the adversarial sample image when eps parameter is set to 16/255 and its associated bit-planes; the fourth column is the adversarial sample image when eps parameter is set to 24/255 and its associated bitplane; the fifth column is the adversarial sample image when eps parameter is set to 32/255 and its associated bitplane.

Figure 8 shows the original image and its adversarial sample by FGSM attack and its associated bitplane changes. The visualization of the bitplanes changes helps in

Table 1: Experimental results of different bitplanes on MERA and boxcounting features.

Feature	MERA accuracy	Box accuracy	MERA loss	Box loss
Original	0.9531	0.9844	0.2526	0.0837
Bitplane1	0.7246	0.7326	0.2353	0.3046
Bitplane2	0.7244	0.7355	0.2839	0.2101
Bitplane3	0.7203	0.7364	0.2892	0.2299
Bitplane4	0.7348	0.7552	0.2486	0.2869
Bitplane5	0.9688	0.8906	0.2118	0.7642
Bitplane6	0.9844	0.9062	0.2044	0.2981
Bitplane7	0.9844	0.8281	0.6144	0.7873
Bitplane8	0.9531	0.9821	0.7535	0.6259



(a)

(b)

Figure 7: The similar phase transition in image. (a) Image bitplanes slicing, and (b) multiscale Ising model.

interpreting the feasibility of anti-adversarial attack abilities. Table 2 shows comparisons of classification accuracy of different bit-planes against FGSM attacks. Different bitplanes have different anti-adversarial attack abilities, and disturbances in some bitplanes will be suppressed. Classification accuracy of classifier performance comparison under a different intensity of FGSM attack has been made between the original image and that of each bitplanes under FGSM attacks. Table 2 depicts that when eps parameter is set to 32/255 in FGSM attack, the accuracy rate is 98.44% in the seventh bitplane; when eps parameter is set to 24/255, the accuracy rate is 96.88% in the sixth plane; when eps parameter is set to 16/255, the accuracy rate reaches 92.19% in the eighth bitplane; when eps parameter is set to 12/255, the accuracy rate is 98.44% in the eighth bitplane. The experimental results show that bitplane slicing can help identify the true class of adversarial samples and show good classification performance against attacks.

We simulate four typical adversarial attacks: FGSM, CW, JSMA and DEEPFOOL. The following experiments are security evaluation on our QCNN which can resist FGSM, CW, JSMA, and DEEPFOOL adversarial attack. Figure 9 depicts different attacks on the background of the target image. The upper images are the original image and the FGSM, CW, JSMA, DEEPFOOL attacked image, and the lower images are the corresponding perturbations attack noise. The color bar indicates the strength of the attack which is also shown in Figure 9. An interesting observation is that some bit-planes can help classifier to improve accuracy. Table 3 depicts that in FGSM adversarial attack, the eighth bitplane yields the best accuracy 95.31%; in CW adversarial attack, the sixth bitplane yields the best accuracy 92.19%; in JSMA adversarial attack, the fifth bit-plane yields the best accuracy 87.50%; in DEEPFOOL adversarial attack, the sixth bitplane yields the best accuracy 93.75%. The experimental results show that bitplanes slicing of QCNN can accurately identify the true class of adversarial samples and show good classification performance against different attacks. Image bitplane slicing has a similar pattern to that of the Ising phase transition. There is research significance to explore the correlation between the chaotic nature of image and the classification/clustering models where the pixels of the image and the Ising chaology particles share similar patterns.

In the feature extraction section, the time complexity of the boxcounting algorithm is shown to be $O(\mathrm{nlogn})$,

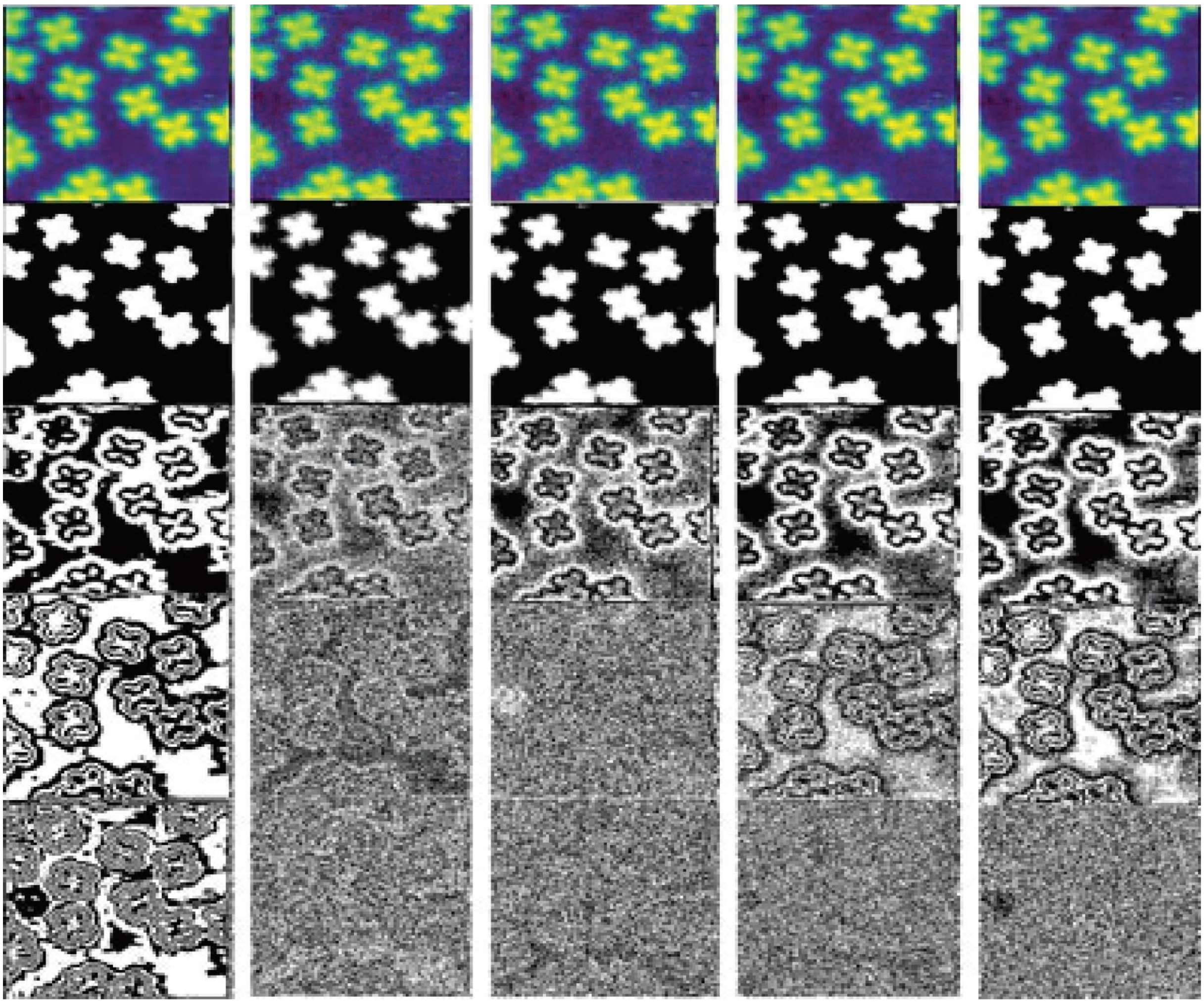


Figure 8: FGSM adversarial attack image.

Table 2: QCNN accuracy against FGSM adversarial attacks.

Parameter	32/255	24/255	16/255	12/255
FGSMs	0.9062	0.9219	0.9487	0.9844
Bitplane1	0.7394	0.7439	0.7426	0.7541
Bitplane2	0.7278	0.7297	0.7384	0.7641
Bitplane3	0.7386	0.7344	0.7528	0.7921
Bitplane4	0.7473	0.7469	0.7531	0.7853
Bitplane5	0.9062	0.8281	0.8594	0.8750
Bitplane6	0.8906	0.9688	0.8906	0.8281
Bitplane7	0.9844	0.9665	0.8750	0.9688
Bitplane8	0.9375	0.9531	0.9219	0.9844

where n is the number of considered points. The time complexity of MERA is shown to be $O(n^{6\log\chi} \cdot \chi^4)$, where χ is a refinement parameter for bond dimensions. We will demonstrate that for the neural computation with an input size of $n = 2^k$, the operator used in quantum implementation is $O(\log^2 n)$, while it is $O(n)$ on classical computers. We adopt the widely used time-space product complexity as the cost complexity, for the quantum implementation, the time complexity circuit depth is $O(d \cdot \log^2 n)$, where d is the number of layers, while the space complexity (i.e., qubit numbers) is $O(\log n)$. The time-space complexity is $O(d \cdot \log^3 n)$, the hybrid quantum-classical complexity can still lower than $O(d \cdot n^2)$ on the classical computer.

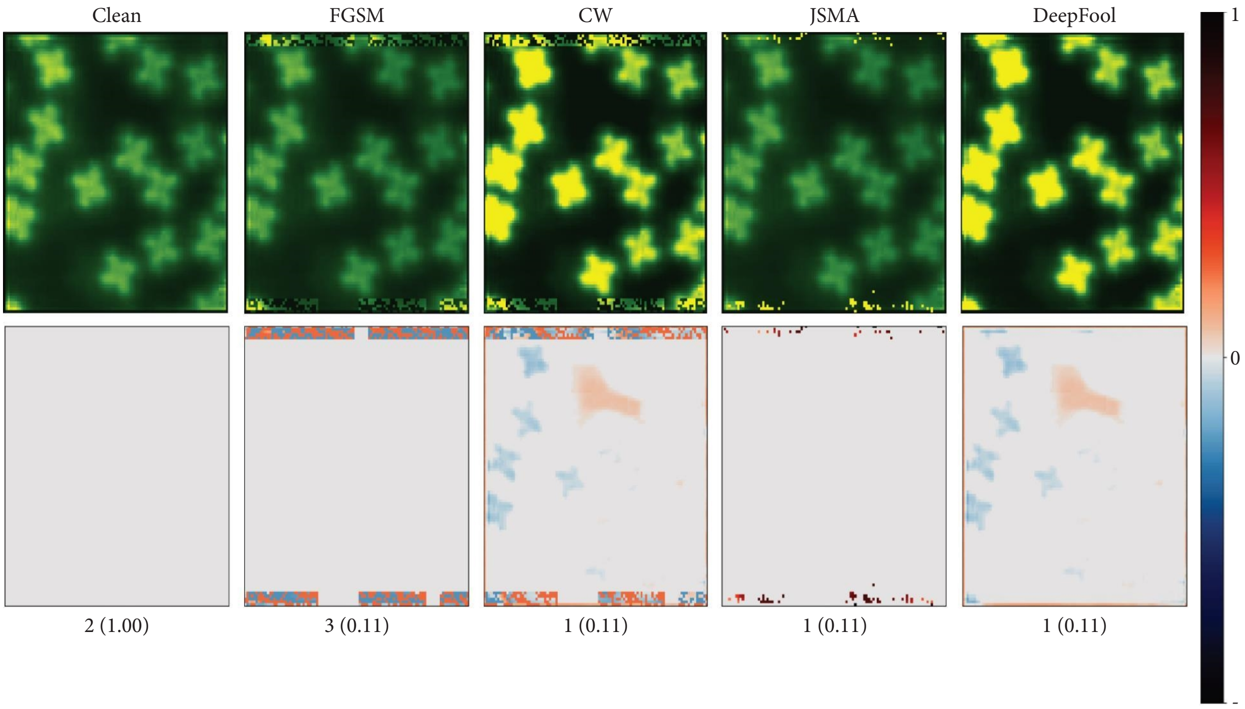


Figure 9: Different kinds of adversarial attacked image and their associated perturbation.

Table 3: QCNN accuracy against different adversarial attacks.

Types	FGSM	CW	JSMA	DeepFool
Attacks	0.8906	0.8906	0.8438	0.8750
Bitplane1	0.7498	0.7465	0.7599	0.7364
Bitplane2	0.7236	0.7283	0.7681	0.7826
Bitplane3	0.7942	0.7563	0.8013	0.7169
Bitplane4	0.8045	0.7954	0.7842	0.7765
Bitplane5	0.9219	0.8112	0.8750	0.8906
Bitplane6	0.8906	0.9219	0.7188	0.9375
Bitplane7	0.9375	0.9062	0.8728	0.9062
Bitplane8	0.9531	0.7656	0.8594	0.9219

5. Conclusions

In this paper we propose a multiscale entanglement renormalization ansatz (MERA) features extraction method based on a novel quantum convolutional neural network (QCNN) for binary scanning tunneling microscopy (STM) image classification. We design QCNN's quantum circuits for state preparation, quantum convolution, and quantum pooling in the TensorFlow quantum framework and compare the performance of QCNN classifier and two hybrid quantum-classical QCNN models. We also reveal the intrinsic mechanism where general image bitplane slicing has a similar pattern to that of the Ising phase transition, and the adversarial attacked images have minor visual differences but the bitplanes of attacked image show significant differences. Our scheme can robustly resist adversarial attacks and it is explainable that the classification performance of different bitplanes of QCNN also shows the corresponding differences under FGSM, CW, JSMA, and DEEPFOOL attacks with different levels of attack intensity.

Data Availability

The data used to support the findings of this study are available at https://alex-krull.github.io/stm-data.html.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the financial support from National Key R&D Program of China (2019YFC0120102), Natural Science Foundation of Guangdong Province (No. 2018A 0303130169, No. 2020A151501212), Opening Project of Guangdong Province Key Laboratory of Big Data Analysis and Processing at the Sun Yat-Sen University (No. 201902), Key laboratory of Industrial Equipment Quality Big Data, MIIT (No. 2021-IEQBD-03), and Educational Science Planning Project of Guangdong Province (No. 2022GXJK073).

References

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, pp. 195–202, 2017.

[2] V. Havlíček, A. D. Córcoles, K. Temme et al., "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, pp. 209–212, 2019.

[3] A. W. Harrow and A. Montanaro, "Quantum computational supremacy," *Nature*, vol. 549, no. 7671, pp. 203–209, 2017.

[4] E. Tang, "A quantum-inspired classical algorithm for recommendation systems," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, Phoenix AZ USA, June 2019.

[5] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 2019.

[6] J. Kossaifi and A. Bulat, "T-net: Parametrizing fully convolutional nets with a single high-order tensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[7] M. Henderson, S. Shakya, S. Pradhan, and T. Cook, "Quanvolutional neural networks: powering image recognition with quantum circuits," *Quantum Machine Intelligence*, vol. 2, no. 1, p. 2, 2020.

[8] M. Broughton and G. Verdon, "Tensorflow quantum: a software framework for quantum machine learning," 2020, https://arxiv.org/abs/2003.02989.

[9] M. Plesch and C. Brukner, "Quantum-state preparation with universal gate decompositions," *Physical Review A*, vol. 83, no. 3, pp. 032302–032320, 2011.

[10] A. M. Childs and N. Wiebe, "Hamiltonian simulation using linear combinations of unitary operations," 2012, https://arxiv.org/abs/1202.5822.

[11] R. Iten, R. Colbeck, I. Kukuljan, J. Home, and M. Christandl, "Quantum circuits for isometries," *Physical Review A*, vol. 93, no. 3, Article ID 032318, 318 pages, 2016.

[12] J. G. Liu and L. Wang, "Differentiable learning of quantum circuit born machines," *Physical Review A*, vol. 98, no. 6, Article ID 062324, 324 pages, 2018.

[13] S. Wei, Y. Chen, Z. Zhou, and G. Long, "A quantum convolutional neural network on nisq devices," *AAPPS Bulletin*, vol. 32, no. 1, p. 2, 2022.

[14] S. Y. C. Chen, T. C. Wei, C. Zhang, H. Yu, and S. Yoo, "Quantum convolutional neural networks for high energy physics data analysis," *Physical Review Research*, vol. 4, no. 1, Article ID 013231, 2022.

[15] Y. C. Li, R. G. Zhou, R. Xu, J. Luo, and W. Hu, "A quantum deep convolutional neural network for image recognition," *Quantum Science and Technology*, vol. 5, no. 4, Article ID 044003, 2020.

[16] A. Datta and M. Fredrikson, "Machine learning explainability and robustness: connected at the hip," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Singapore, August 2021.

[17] G. Ras, N. Xie, M. Van Gerven, and D. Doran, "Explainable deep learning: a field guide for the uninitiated," *Journal of Artificial Intelligence Research*, vol. 73, pp. 329–397, 2022.

[18] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[19] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision a survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[20] J. W. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828–841, 2019.

[21] X. Y. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, 2019.

[22] S. Y. Yang and M. Wang, "Deep sparse tensor filtering network for synthetic aperture radar images classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 3919–3924, 2018.

[23] E. M. Stoudenmire and D. J. Schwab, "Supervised learning with quantum-inspired tensor networks," 2016, https://arxiv.org/abs/1605.05775.

[24] A. Hallam and E. Grant, "Compact neural networks based on the multiscale entanglement renormalization ansatz," 2017, https://arxiv.org/abs/1711.03357.

[25] G. Evenbly and S. R. White, "Entanglement renormalization and wavelets," *Physical Review Letters*, vol. 116, no. 14, Article ID 140403, 2016.

[26] J. Haegeman, B. Swingle, M. Walter, J. Cotler, G. Evenbly, and V. B. Scholz, "Rigorous free-fermion entanglement renormalization from wavelet theory," *Physical Review X*, vol. 8, no. 1, Article ID 011003, 2018.

[27] G. Chen and Q. Chen, *Quantum Convolutional Neural Network for Image Classification"*, Pattern Analysis and Applications, Newyork, NY, USA, 2022.

[28] A. Krull, P. Hirsch, C. Rother, A. Schiffrin, and C. Krull, "Artificial-intelligence-driven scanning probe microscopy," *Communications on Physics*, vol. 3, no. 1, p. 54, 2020.

[29] C. Panigrahy and A. Seal, "Is box-height really a issue in differential box counting based fractal dimension?" in *Proceedings of the International Conference on Information Technology*, Delhi, India, June 2019.

[30] B. Coyle, D. Mills, V. Danos, and E. Kashefi, "The born supremacy: quantum advantage and training of an Ising born machine," *Npj Quantum Information*, vol. 6, no. 1, p. 60, 2020.