Entanglement structure detection via computer vision

Rui Li[©],[1] Junling Du,[2] Zheng Qin,[1] Shikun Zhang,[1] Chunxiao Du,[1] Yang Zhou,[1,3,*] and Zhisong Xiao[1,4]

[1]*School of Physics, Beihang University, Beijing 100191, China*
[2]*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*
[3]*Research Institute for Frontier Science, Beihang University, Beijing 100191, China*
[4]*School of Instrument Science and Opto-Electronics Engineering,*
Beijing Information Science and Technology University, Beijing 100192, China

Quantum entanglement plays a pivotal role in various quantum information processing tasks. However, a universal and effective way to detect entanglement structures is still lacking, especially for high-dimensional and multipartite quantum systems. Noticing the mathematical similarities between the common representations of many-body quantum states and the data structures of images, we are inspired to employ advanced computer vision technologies for data analysis. In this work, we propose a hybrid convolutional neural network–transformer model for both the classification of Greenberger-Horne-Zeilinger and W states and the detection of various entanglement structures. By leveraging the feature-extraction capabilities of convolutional neural networks and the powerful modeling abilities of transformers, we not only can effectively reduce the time and computational resources required for the training process but can also obtain high detection accuracies. Through numerical simulation and physical verification, it is confirmed that our hybrid model is more effective than traditional techniques and thus offers a powerful tool for characterizing multipartite entanglement structures.

I. INTRODUCTION

Quantum entanglement underlies the unique characteristics and advantages of quantum systems over classical systems [1–3]. It plays a vital role in various branches of quantum information [4,5] such as quantum communication [6], quantum cryptography [7], and quantum computing [8,9]. An improper amount of entanglement, incorrect structure, or pattern of a multipartite entangled state could all affect the overall efficiency of given quantum computation tasks or the performance of given quantum protocols. As a result, knowledge of entangled states [2,10] is necessary, which is to say, it is essential to determine the entanglement type, entanglement separability, and entanglement depth [7] of these states. For low-dimensional systems, a necessary and sufficient condition exists for the separability of bipartite systems: the Peres-Horodecki criterion based on the positivity-of-partial-transpose criterion [11,12]. However, for higher-dimensional systems, or more parties, the characterization of the set of positive maps is an NP-hard problem. Various theoretical and experimental tools [13–15] have been proposed to analyze and detect multipartite quantum entanglement. One of the most common approaches is the entanglement witness [16–20], which can distinguish a specific entangled state from separable ones.

However, existing entanglement detection methods lack universality [19]. For example, one cannot determine the entanglement witness operator for all possible entangled states and an arbitrary number of partitions. Even worse is the lack of a general description of all entangled states in multipartite quantum systems. Furthermore, the biggest challenge for analytic or numeric approaches is that as the qubit number increases, both the number of possible quantum states and the number of entanglement structures increase exponentially. Consequently, researchers have turned to machine learning techniques for help [21].

At first, most related studies adopted supervised training methods [22–25], achieving high accuracy for specific tasks. However, the process of labeling a large number of quantum states is very time-consuming and almost impossible for multiqubit systems. Thus, semisupervised [26,27] and unsupervised learning [28] techniques that can predict a large number of unlabeled quantum states from a small number of labeled states have been developed. The problem with these methods is they can offer only limited precision and require further experimental validation. Considering the structural resemblance between the mathematical representation of quantum states and classical image data, we are inspired to utilize established computer vision techniques for entanglement structure detection, which are known for their reduced resource consumption and computational efficiency.

In this study, a hybrid convolutional neural network–transform model is proposed for the detection of multipartite entanglement. This network structure was initially developed for image processing and is particularly adept at handling large quantities of data. Therefore, a deep convolutional neural network (CNN) [29,30] can effectively identify local features and alter data dimensions through convolution operations, while the self-attention mechanism of the

---

*Contact author: yangzhou9103@buaa.edu.cn

012448-1

transformer [31–33] can capture long-distance dependences. Thus, the trained neural network can identify the key features in different descriptions of entanglement structures, precisely delineating the boundaries between completely independent samples and various entangled samples. This brings about a major advantage: the same neural network can be applied to both the classification of entanglement states (Greenberger-Horne-Zeilinger (GHZ) [34] and *W* states [35]) and the detection of specific entanglement structures (entanglement separability and entanglement depth). As far as we know, the current machine learning method can handle only a single specific task, such as determining whether a particular type of entanglement exists or detecting the separability and depth of some special multipartite entangled states. Different from it, our hybrid model demonstrates exceptional performance in accomplishing both classification and detection tasks with low time and computational cost. Numerical examples, ranging from 3- to 10-qubit systems, demonstrate that our network achieves an average classification accuracy exceeding 99.57% and a 95% accuracy in detecting entanglement structures.

The structure of this paper is organized as follows. In Sec. II, we introduce the definition of GHZ and *W* entangled states used for data generation and the concept of using entanglement separability and entanglement depth as descriptors of the entanglement structure. In Sec. III, we describe the process of dataset preparation for our analysis and introduce a hybrid CNN-transformer model that captures the inherent properties of quantum states to classify and detect entangled states. In Sec. IV, we quantify the performance of our trained neural network through experiments with numerical simulations and real quantum devices. Finally, this paper is concluded in Sec. V.

II. PRELIMINARIES

A. GHZ and *W* entanglement states

Separable and entangled states. For two quantum systems *A* and *B*, their states are represented by the state vectors $|\Psi_A\rangle$ and $|\Psi_B\rangle$ in their Hilbert spaces H_A and H_B. The joint quantum state of the two systems is represented by the state vector $|\Psi_{AB}\rangle$ in the Hilbert space $H_A \otimes H_B$. If $|\Psi_{AB}\rangle$ can be written as a product state, it is a separable state; otherwise, it is an entangled state [36].

Generalized to an *N*-qubit case, the quantum state is usually represented by a density matrix:

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i. \tag{1}$$

Here, p_i is the probability that $0 \leqslant p_i \leqslant 1$ and $\sum_i p_i = 1$, and $\hat{\rho}_i^k$ is the pure-state density matrix of each subsystem. If this density matrix ρ can be expressed as a convex combination of multiple product states,

$$\hat{\rho} = \sum_l p_i \hat{\rho}_i^1 \otimes \hat{\rho}_i^2 \cdots \otimes \hat{\rho}_i^N, \tag{2}$$

the corresponding state is considered separable; otherwise, it is entangled.

For quantum systems with three or more particles, there are primarily two interesting classes of entangled states: GHZ

and *W* states. These two classes of entangled states can be partitioned into two disjoint categories through stochastic local operations and classical communication (SLOCC) [37].

GHZ states. GHZ states are a special type of multipartite entangled state in which the entanglement between all particles is global. In an *N*-particle system, the GHZ state can be expressed as

$$|\Psi_{\text{GHZ}}\rangle = \frac{1}{\sqrt{2}} \left(|1\rangle^{\otimes N} + |0\rangle^{\otimes N}\right). \tag{3}$$

W *states.* W states are another type of multipart entangled state characterized by the fact that even if one particle is lost after the measurement, the remaining particles remain entangled, unlike the GHZ state, which becomes completely separable. In an *N*-particle system, the *W* state can be expressed as

$$|\Psi_W\rangle = \frac{1}{\sqrt{N}} \left(|10\cdots0\rangle + |010\cdots0\rangle + \cdots + |0\cdots01\rangle\right). \tag{4}$$

The *W* state is locally indistinguishable from other states in its equivalence class under local unitary transformations [38]. SLOCC can distinguish between these two classes of entangled states because they exhibit different invariances under local operations and classical communication. This classification is significant for quantum information processing tasks, such as quantum computation and quantum communication, because it helps us to understand the application and limitations of entanglement resources in these tasks.

B. Entanglement structure

The entanglement structure of quantum states refers to entanglement separability and entanglement depth (or entanglement producibility). This can be analogous to the combinatorial problem of an *N*-body system. There are 2^{N-1} types of partitions, and $(N + 1)$ kinds of split methods to decompose an *N*-body system into multiple subsystems $\Lambda = \{\Lambda_1, \Lambda_2, \ldots, \Lambda_k\}$, $k \leqslant N$. Determining the exact analytic equation is very difficult, but recursive and dynamic programming methods can be used to calculate the integer division of *n* particles. In general, we can use the Young diagram [39] to directly show the partition. For example, a four-body system has eight types of partitions and five types of split methods. These partitions include $4, 1 \otimes 3$, $2 \otimes 2$, $1 \otimes 1 \otimes 2$, and $1 \otimes 1 \otimes 1 \otimes 1$. This can be represented by the Young diagram shown in Fig. 1(a).

In the Young diagram, each row represents the number of entanglements contained in a subsystem, with the number of entanglements in the subsystems decreasing from top to bottom. Therefore, the width of the top row represents the maximum number of entangled particles, and the number of rows represents the number of separable subsystems.

k-*separable (h-inseparable).* k-separability (h-inseparability) is an indicator used to represent the number of separable subsystems in a multiparticle quantum system. A quantum state of *N* particles is said to be *k*-separable if it can be divided into *k* groups $(1 \leqslant k \leqslant N)$ such that there is no entanglement between these groups, although particles within each group may be entangled with each
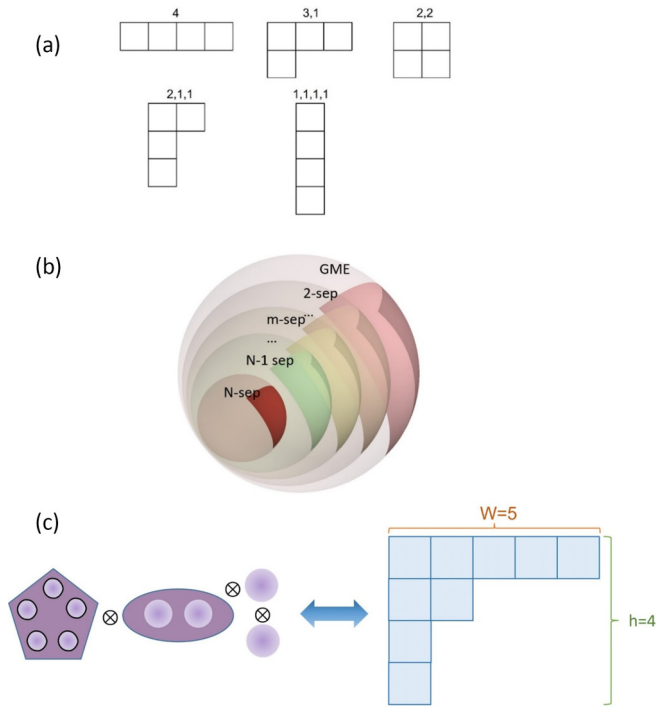
FIG. 1. Young-diagram representation and entanglement structure analysis of multibody systems. (a) All possible Young diagrams for a four-particle system. (b) The separability hierarchy of N-qubit states. (c) Characterization of the entanglement structure of a nine-body system using Young diagrams, which can be divided into $h = 4$ subsystems, with an entanglement depth of $w = 5$, corresponding to the height h and width w of the Young diagram.

other. Mathematically, a k-separable pure state $\rho_{k\text{-sep}}$ can be expressed as

$$\rho_{k\text{-sep}} = \rho_{A1} \otimes \rho_{A2} \cdots \otimes \rho_{Ak}, \qquad (5)$$

indicating that the N-body system is divided into k-separable subsystems. A mixed state is called k-separable if it is a convex combination of k-separable pure states. If a quantum state is not k-separable, it is considered h-inseparable, implying that the system cannot be divided into h-separable subsystems.

As shown in Fig. 1(b), an N-qubit state is considered N-separable (N-sep) if it can be fully decomposed into N independent quantum subsystems, implying that the entire quantum system exhibits no entanglement. On the contrary, if a multiqubit state cannot be represented as a convex combination of any separable states, it exhibits genuine multipartite entanglement (GME) [40]. GME represents an entanglement structure involving the entire multiqubit system and cannot be described by decomposing it into smaller subsystems' entanglement structures, indicating that the quantum state exhibits strong quantum correlations among all its subsystems.

k-*producible (entanglement depth* w*).* k-producibility is an indicator used to represent the maximum degree of entanglement in the system, which can also be referred to as the entanglement depth w. A k-producible pure state $\rho_{k\text{-pro}}$ can be written as

$$\rho_{k\text{-pro}} = \rho_{A1} \otimes \rho_{A2} \cdots \otimes \rho_{Am} m \qquad (6)$$

where ρ_{Am} denotes the state with the maximum number of k entangled particles. If a quantum state is not k-producible, its entanglement depth is at least $k + 1$, indicating that there are at least $k + 1$ particles entangled together in the system.

The larger the value of entanglement depth w or the smaller the value of k-separability is, the greater the entanglement in the multipartite system is. This concept is intuitively illustrated in Fig. 1(c).

III. METHODS

A. Dataset preparation

For this study, we developed a comprehensive and diverse dataset to train and evaluate the performance of our proposed computer vision techniques for quantum entanglement detection. The dataset comprises 200 000 density matrices, 100 000 for GHZ states and 100 000 for W states. To ensure a robust and representative dataset, we employed a systematic approach to generate a wide range of entangled structures, spanning from GME to completely separable states.

Our dataset creation process involved several key steps to ensure the inclusion of all possible subsystem configurations and entanglement depths. As a starting point, we considered a preset number of particles, denoted as N, which served as the foundation for our subsequent steps. We then performed 10 000 random integer decompositions on this preset number N to obtain a variety of potential subsystem sizes k_i. The number of integers generated in each decomposition, termed h, corresponds to the number of separable subsystems in the N-partite system. The largest integer among the k_i values represents the entanglement depth w, which is the maximum number of particles entangled together within the entire system. By controlling the number of subsystems h and the entanglement depth w through the integer decomposition process, we were able to precisely engineer the entanglement structure of each quantum state in our dataset, especially for complex scenarios involving a large number of particles.

Next, we carefully selected quantum states for each subsystem based on their physical significance and the number of particles involved. For single-particle subsystems ($k_i = 1$), we chose arbitrary points on the Bloch sphere and computed their density matrices. In the case of two-particle subsystems ($k_i = 2$), we randomly selected one of the four Bell states and calculated its density matrix. For subsystems with three or more particles ($k_i \geqslant 3$), we derived their density matrices using the GHZ- and W-state definitions, Eqs. (3) and (4), as these states are particularly important within their respective subsystems. Finally, we combined these various states following the formula in Eq. (2) to construct our comprehensive dataset.

To further enhance the dataset's robustness and reduce redundancy, we implemented a data augmentation technique inspired by the quantum circuit [41]. This involved rotating the density matrices generated for each subsystem within a range of 0 to $\pi/10$, effectively mitigating overfitting and increasing the dataset's diversity and representativeness.

To accurately capture the composition and entanglement structures of the quantum states, we developed a precise

labeling scheme. The labels, represented as strings, encapsulate the subsystems: "One" for a single-particle subsystem in a pure state, "Bell" for a two-particle subsystem in a Bell state, and "GHZ$_k$" for a k-particle subsystem in a GHZ state. By combining these terms with hyphens, we created comprehensive labels such as GHZ$_6$, Bell-One-GHZ$_3$, and One-Bell-GHZ$_3$, which effectively illustrate the multifaceted nature of quantum entanglement in complex multiparticle systems.

This approach to dataset generation ensures a comprehensive representation of entangled states while maintaining the rigor and objectivity necessary for academic research. By systematically exploring the full range of entangled structures and incorporating randomization into the probability coefficients, our dataset serves as a robust foundation for evaluating the performance of our machine learning and computer vision techniques for quantum entanglement detection.

B. Hybrid CNN-transformer model

In our experiments, we pretested with a traditional fully connected neural network (FNN) [42] and found that the results were excellent when the number of particles was small, but when the number of particles was greater than seven, the accuracy dropped significantly. This is because the FNN cannot extract local features, and it cannot capture the local spatial information of the input data. This problem can be solved by a CNN, which can effectively identify local features through a convolution operation, making it superior for image processing and other artificial-intelligence fields.

In essence, the FNN needs to flatten the input data into a one-dimensional vector, thus losing spatial information, whereas the CNN can accept the input data of the original shape, retain the spatial structure information, and help extract more efficient features. Through convolution operations, CNN can efficiently extract local features while reducing the number of parameters. The addition of convolutional layers makes CNN translationally invariant and thus achieves remarkable success in image processing, computer vision, and other fields. However, CNNs still have limitations in terms of capturing long-distance dependences. With the development of computer vision technology, a vision transformer (ViT) [31,33], which is the latest neural network development direction, aims to make full use of the advantages of transformer structure. Transformers have made breakthroughs in the field of natural language processing, mainly because their self-attention mechanism can capture long-distance dependences. ViT divides an image into small pieces and then processes these small pieces using a transformer to capture global features. ViT has surpassed CNN in many image-recognition tasks and has become the current frontier technology in the field of computer vision.

We adopted the perspective that quantum states can be regarded as analogous to images in the context of their mathematical representations. Drawing upon this conceptual similarity, we leverage advanced computer vision techniques to analyze these quantum states, effectively transforming the problem of entanglement detection into a task that parallels image analysis. This innovative approach facilitates the exploration of entanglement structure detection using
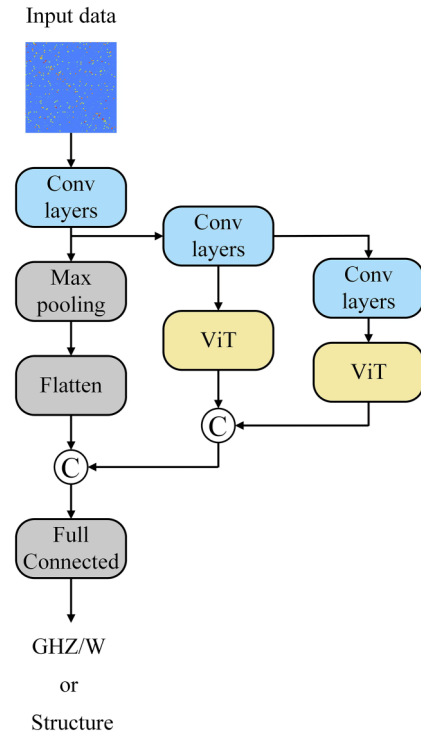

FIG. 2. Hybrid CNN-transformer model.

well-established methods in the field of computer vision. Inspired by the paper "An image is worth 16x16 words: Transformers for image recognition at scale" [33], we employ CNNs and transformers to extract local and global features, respectively. The dimensions of the density matrices vary with the number of particles. When the particle count is high, the density matrices become larger, necessitating more time and computational resources for the transformer. To address this issue, we first applied a convolutional layer to extract features from the density matrices while altering their dimensions.

For particle numbers $n = 3, 4, 5, 6$, and 7, the density matrix dimensions were relatively small, eliminating the need for dimensionality reduction. However, for $n = 8, 9$, and 10, the density matrix dimensions were larger, and we applied two-dimensional convolutions to reduce their dimensions to 128×128. This process minimizes redundant features and decreases the computational load for the subsequent transformer calculations.

The network structures for different numbers of particles are shown in Fig. 2. The convolutional layers ("Conv layers") initially act as feature detectors, scanning the input data to highlight patterns that are akin to edges or textures in an image. These layers are critical for discerning local variations within the quantum data, which are essential clues to the nature of entanglement. The "Max pooling" layer serves to distill information by focusing on the most prominent features detected by the convolutional layers, effectively reducing the dimensionality and computational complexity of the data. The "Flatten" operation then converts this condensed feature map into a one-dimensional array, setting the stage for a deeper analysis.
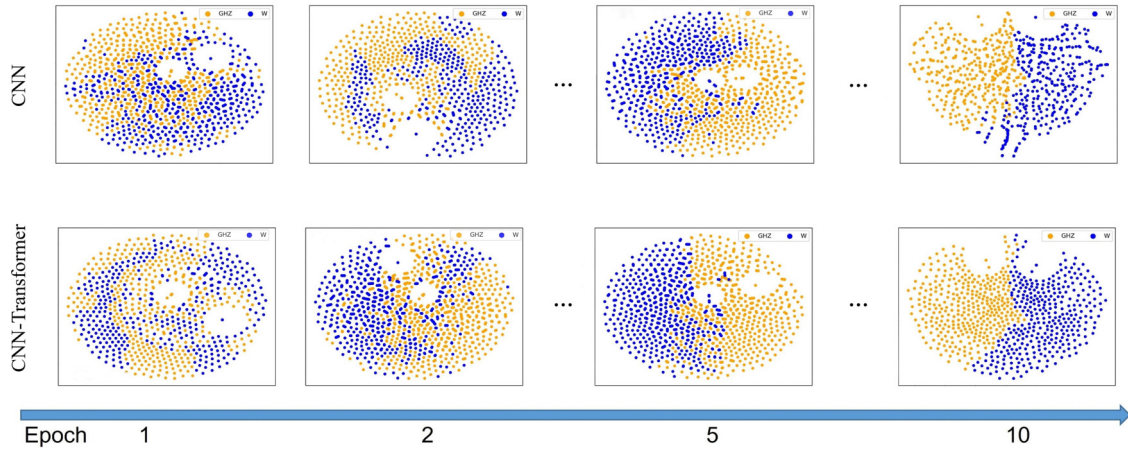
FIG. 3. During the training process, the evolution of feature vectors for 10 000 entangled samples of five-qubit states.

The ViT model includes a patch-embedding process and position embedding. The patch-embedding process divides these feature maps into smaller, manageable pieces, or patches, similar to breaking an image into segments. For different numbers of particles ($n = 3$ to 10), we vary the patch size accordingly (from 2×2 to 32×32), optimizing the model's ability to process quantum states of varying complexity. Position embeddings are added to these patches to retain information about the relative or absolute position of the features within the original quantum state, which is crucial since the spatial relationship can hold significant quantum information.

The circle labeled "C" in Fig. 2 represents the concatenation process. It combines the features extracted from separate pathways in a network. This step is essential for merging different types of information processed by the network (both local and global features) into a comprehensive feature set.

Finally, the fully connected layers, a multilayer perceptron [43], take this richly processed information and determine the specific type of entanglement structure. This process mirrors the way we classify images based on a detailed understanding of their content, learned through both local and global observations.

IV. NUMERICAL RESULT

In this section, we present our numerical results, including the evaluation metrics used to assess the performance of the model. By implementing the CNN and CNN-transformer architectures, we showcase the successful classification of GHZ or W states and the detection of entanglement structures. Furthermore, we assess the robustness of the model under various noisy settings, providing valuable insights for optimizing future quantum state classification tasks in real experiments.

A. Classification of GHZ- and W-class states

In our study, we initially focused on classifying GHZ and W states. We generated mixed states in the dataset, which comprised both GHZ and W states. Through the implementation of both CNN and CNN-transformer models, we could accurately and effectively discern whether the quantum states

of 3- to 11-particle systems contained GHZ or W states. By fine-tuning the model parameters, we attained a 100% accuracy rate in distinguishing between these quantum states.

We used t-distributed stochastic-neighbor-embedding (t-SNE) [44,45] plots to analyze our data. Figure 3 shows t-SNE plots for the case of $N = 5$. It can be seen that as training progressed the clustering features were more and more clear. In early training, the GHZ- and W-state data points overlapped or were disorganized, indicating that the model had not fully learned their differences. In later epochs, the GHZ- and W-state data points separated and formed compact clusters, indicating that the model captured the high-dimensional data structure. This allows it to effectively differentiate between these two quantum states. As training continued, the clusters became more distinct, indicating improved classification performance.

Compared with the CNN model, the CNN-transformer showed better classification on the t-SNE plot. It achieved clear separation and compactness of data point clusters in fewer epochs, learning distinguishing features more quickly. The CNN-transformer model also had fewer outlier points, suggesting better handling of potential anomalies or noise. This difference in performance is due to the CNN-transformer model combining the CNN's local perception and the transformer's global perception, making it better at capturing complex data correlations and context.

In summary, the CNN-transformer model outperformed the CNN model when classifying the GHZ and W states. The t-SNE plots exhibit faster learning and better classification performance.

B. Detecting entanglement structure

Detecting entanglement structures, particularly in our massive dataset, is a challenging multiclassification problem. As the number of particles increases, the number of classifications exponentially increases, making the problem more complex. To address this issue effectively, we need to design a powerful model that can capture the entanglement features of different particle numbers.

As mentioned earlier, the CNN-transformer model demonstrates superior performance in classifying the GHZ and W
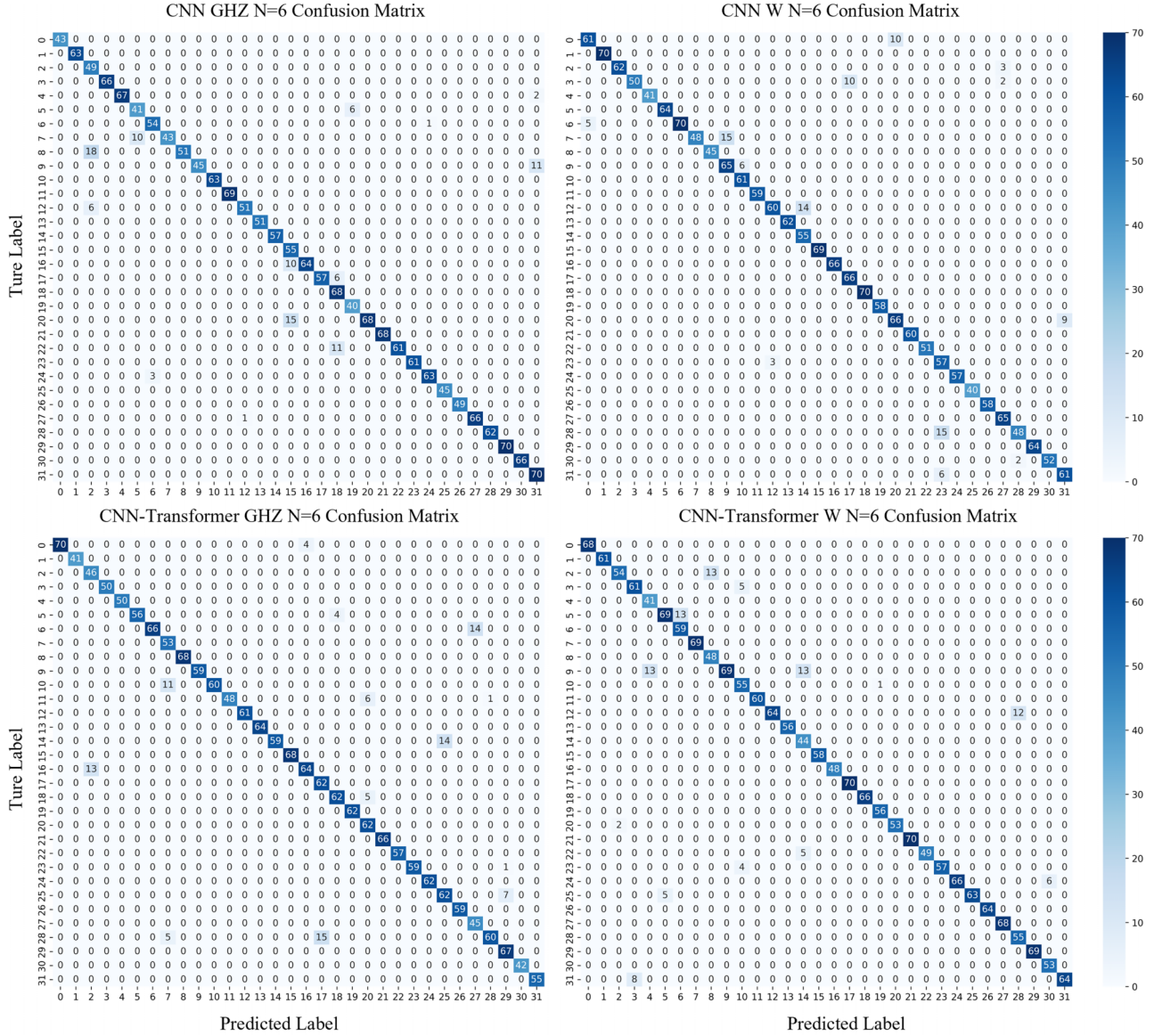
FIG. 4. The confusion matrices for six-particle entanglement structures. In each confusion matrix, the horizontal axis (x axis) represents the predicted class labels, while the vertical axis (y axis) represents the true class labels. Each cell in the matrix contains a count of instances where the model predicted a particular class label (x axis) for the instances of the true class label (y axis). There are a total of 32 entanglement structures for six particles, resulting in label sequences from 0 to 31. The values in the cells are represented by varying shades of blue, with larger values corresponding to darker shades and smaller values appearing lighter, almost white.

states. Its advantage lies in the combination of the CNN's local perception capabilities and the transformer's global perception capabilities. This also allows the CNN-transformer model to perform better in complex multiclassification problems. As the number of particles and classifications grows exponentially, we should further optimize the architecture of the CNN-transformer model to maintain its high performance. Our approach involves increasing the number of layers of the model and adjusting its hyperparameters to enhance its expressive power.

To validate the performance of the CNN-transformer model's performance in handling this multiclassification problem, we adopted confusion matrices [46] as visualization tools. By examining the confusion matrices, we can observe the superior performance of the CNN-transformer model in detecting entanglement structures of the GHZ and *W* states more intuitively. Figure 4 shows the case with $N = 6$. Figure 4 showcases four confusion matrices: the top left corner represents the CNN model for the GHZ class with an accuracy of 92.25%, the top right corner represents the CNN model for the *W* class with an accuracy of 94.32%, the bottom left corner represents the CNN-transformer model for the GHZ class with an accuracy of 94.25%, and the bottom right corner represents the CNN-transformer model for the *W* class with an accuracy of 94.81%. It can be noticed that the main diagonal elements (i.e., true positives and true negatives) of the CNN-transformer model are larger than those of the CNN model, suggesting that the CNN-transformer model has an advantage in predicting the correct number of GHZ- and *W*-state samples. This observation indicates that the

CNN-transformer model has higher accuracy in distinguishing these two quantum states. In addition, we find that the off-diagonal elements (i.e., false positives and false negatives) of the CNN-transformer model are smaller compared to those of the CNN model.

This performance difference can be attributed to the CNN-transformer model's combination of the CNN's local perception capabilities and the transformer's global perception capabilities, making it more advantageous in capturing correlations and contextual information in complex data. Therefore, when dealing with GHZ- and W-state classification tasks, the CNN-transformer model achieves better performance than the pure CNN model.

As the particle number increases, the entanglement structures become more complex, and the density matrices exhibit higher dimensions, increasing the computational complexity. We anticipate that the accuracy and precision of our CNN and CNN-transformer models may gradually decrease with an increase in the particle number. However, with complex optimization, the maximum particle number could potentially reach 15 while still achieving an accuracy greater than 75%. The scalability of our methods depends on the efficiency of feature extraction and the balance between local-information preservation and computational-complexity reduction. Further experiments are required to validate our results.

C. Physical verification of entanglement detection

In practical applications, different types of noise make the precise determination of the entanglement structure more challenging. It is crucial to note that datasets formed by real experimental states may exhibit different distributions compared with random test datasets. This discrepancy is a primary challenge in real-world applications of machine learning, particularly when obtaining labels for test data is difficult or expensive, such as in biomedicine [47], material science [48], and physical science [49].

Fortunately, our hybrid CNN-transformer model can address this issue. In the following, we test the performance of our model with real nonidealities in noisy intermediate-scale quantum (NISQ) devices [50]. OriginQ's quantum computing platform is favored for its accessibility and the maturity of its software development tool kit, QPANDA [51]. All subsequent physical experiments were conducted on a six-qubit superconducting quantum computer, OriginQ Wuyuan No. 2 [52]. Owing to the limitations of the current quantum computing capabilities, our data were restricted to three, four, and five particles.

A total of 4000 data samples were prepared as the test set in our pretrained model. The preparation of the dataset is illustrated in Fig. 5, using the case of three particles as an example. Figures 5(a) and 5(b) show the quantum circuits for the generation of the three-particle GHZ state and W state. Figures 5(c) and 5(d) depict the density matrices numerically generated by the definitions of the GHZ and W state, whereas Figs. 5(e) and 5(f) present the density matrices generated through the real quantum devices.

We first used the model to classify quantum states, achieving an experimental classification accuracy of 100%.

Subsequently, we classified the entanglement structures. The CNN-transformer model achieved an accuracy of 99.27% for three particles, 98.85% for four particles, and 97.36% for five particles. The experimental results demonstrated the reliability and robustness of our model.

However, it should be noted that in the NISQ device, as the number of particles increases, fidelity inevitably becomes lower and lower. In the quantum computer we used, fidelity drops to 0.6158 for five particles. We noticed that the elements on the diagonal in the real quantum computer were significantly larger than those in other positions. Therefore, we added white noise to test the performance of our model more comprehensively.

Quantum states with white noise were prepared for 6–10 particles:

$$\rho_{ng} = p|\Psi_{\text{GHZ}}\rangle\langle\Psi_{\text{GHZ}}| + \frac{(1-p)I_n}{2^n}, \quad p \in [0, 1], \quad (7)$$

$$\rho_{nw} = p|\Psi_W\rangle\langle\Psi_W| + \frac{(1-p)I_n}{2^n}, \quad p \in [0, 1], \quad (8)$$

where $|\Psi_{\text{GHZ}}\rangle$ is as given in Eq. (3) and $|\Psi_W\rangle$ is as given in Eq. (4).

After generating the dataset, we first classified these states into GHZ and W classes. The classification performance is demonstrated from two perspectives, as shown in Fig. 6(a). The first perspective focuses on accuracy. It shows that accuracy remains at 100% without noise, even as the number of particles increases. Meanwhile, in the presence of noise, we maintain relatively good results despite a slight decrease in detection accuracy. For five and seven particles, both the CNN and CNN-transformer models achieve an accuracy of 100%. However, as the number of particles increases to nine, the accuracy of the CNN model drops to 93.45%, while the CNN-transformer model maintains a higher accuracy of 95.83%. This observation demonstrates that despite the presence of noise, our models still perform well in handling complex systems with increasing particle numbers. Furthermore, the loss-function values for training and validation converge to a low value, reinforcing the effectiveness and robustness of our models. The second perspective is from the required epoch, emphasizing that the exponential increase in data size with the increase in particle number leads to more training iterations. However, the growth of iterations remains linear, and the epoch termination condition employs early stopping to prevent model overfitting. Consequently, training stops automatically when there is no performance improvement. It is also noteworthy that, even as the particle number increases to 11, the number of training iterations remains below 100. This reflects a considerable level of efficiency and speed which is particularly important for large-scale systems. We utilize tenfold cross validation to evaluate the reliability of our data.

Figure 6(b) displays the loss values of the CNN and CNN-transformer models during the training epochs for quantum systems with $n = 5$, 7, and 9 . The curves demonstrate a steady decrease in loss, indicating learning and improvement in detecting entanglement structures as the training
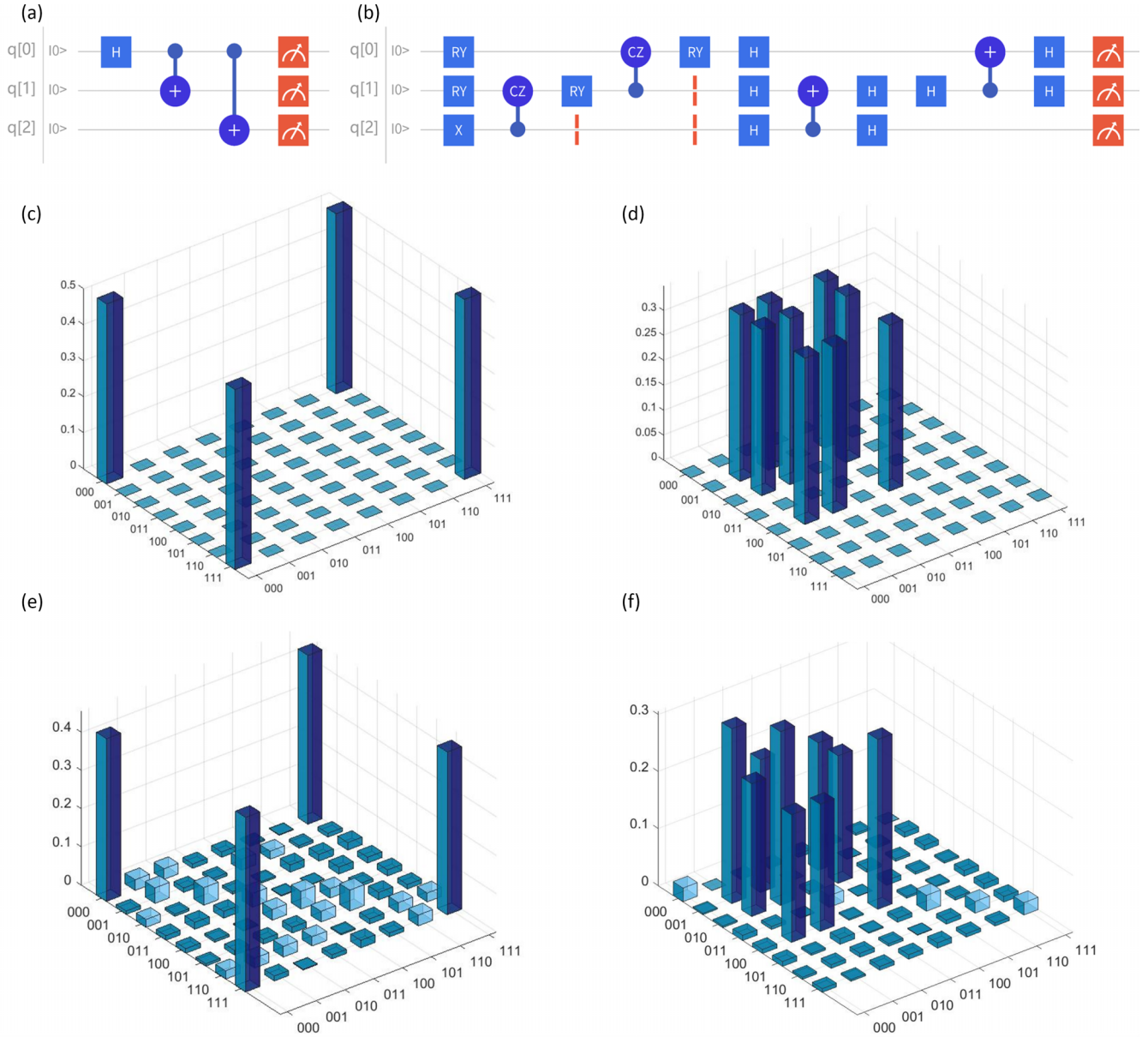
FIG. 5. (a) The circuit employed for generating the GHZ state on OriginQ hardware. The circuit comprises a sequence of operations executed on three qubits, namely, q[0], q[1], and q[2]. The operations involve Hadamard gates (H) and controlled-NOT (CNOT) operations. (b) The circuit utilized for generating the W state on OriginQ hardware. This circuit consists of a series of operations applied to three qubits, q[0], q[1], and q[2]. The operations include RY (Z rotations), Hadamard gates (H), CNOT operations, and controlled-Z (CZ) gates. Moreover, barrier operations are employed to ensure the proper execution order on the hardware. (c) and (d) The ideal density matrices for the GHZ and W states, respectively. (e) and (f) The experimentally obtained density matrices prepared using the OriginQ Wuyuan No. 2 hardware exhibited fidelity of 0.8187 and 0.7578 for the GHZ and W states, respectively.

progresses. The CNN-transformer models exhibit lower loss values compared to CNNs alone, especially in early epochs, suggesting faster convergence. Despite the increasing complexity with higher quantum bit numbers, CNN-transformer models maintain lower losses, highlighting their potential for handling larger quantum systems. The reduced number of epochs required for convergence also implies that our models can be trained more efficiently, achieving optimal performance in a shorter period.

Next, we use noisy data to detect entangled structures with particle numbers from 6 to 10. Figures 7(a) and 7(b) show the

accuracies of the two models with noise. It is worth noting that our method achieves 95.42% accuracy for 6 particles and 90.23% accuracy for 10 particles, which is slightly lower than the 93.52% accuracy for the noise-free cases. Although the accuracy was somewhat affected, it is still acceptable considering the small error bars. This indicates that our method exhibits good stability and robustness under different particle numbers and noise levels. To address the issue of decreased accuracy, we can construct more complex model structures, adjust the model hyperparameters, and adopt more advanced training strategies. In addition, we can enhance the ability of
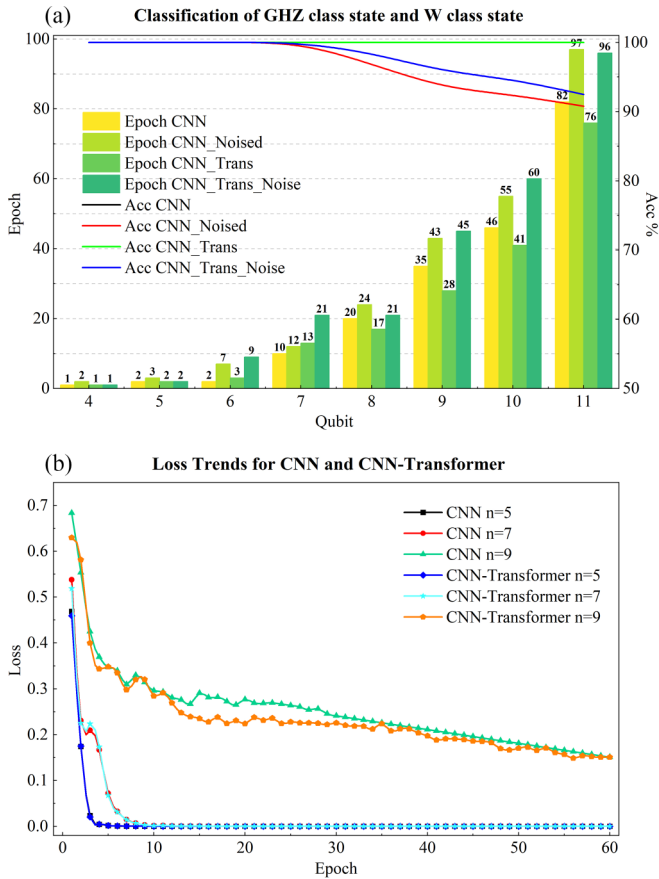
FIG. 6. (a) Classification of a GHZ-class state and a *W*-class state. The bar chart and line chart use different colors to represent the same model. For the bar chart: yellow shows CNN, light green shows CNN_Noised, mint green shows CNN-Trans, and dark green shows CNN-Trans_NOISE. For the line chart, black shows CNN, red shows CNN_Noised, green shows CNN-Trans, and blue shows CNN-Trans_Noise. The right *y* axis indicates the accuracy of these four models at various particle numbers, while the left *y* axis denotes the number of iterations required to achieve high accuracy in machine learning training. (b) The changes in the loss for both CNN and CNN-transformer models during training with five, seven, and nine particles as the number of epochs increases.

this model by increasing the diversity and size of the dataset to accomplish more tasks. This will help the model to perform better when dealing with various situations encountered in real-world applications.

V. CONCLUSION

This article discussed the application of computer vision methods to identify the entanglement types and structures of quantum states simultaneously. Computer vision technologies, particularly the CNN and transformer models, are inherently adept at handling and analyzing data in regular-sized matrices quickly and accurately. Therefore, compared with other data-driven methods for entanglement detection, our hybrid CNN-transformer model has the following advantages.
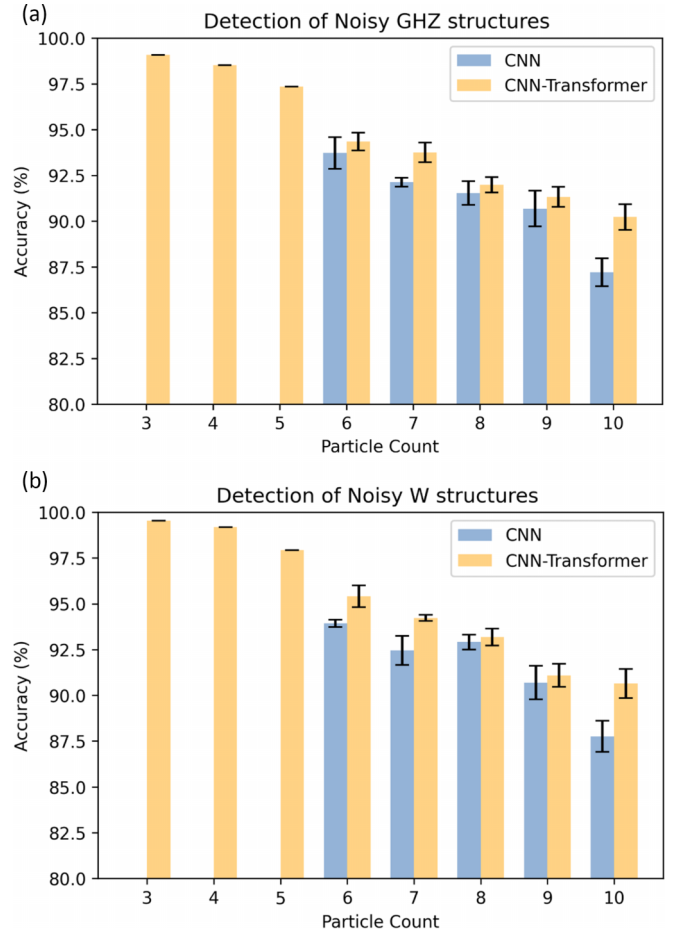
FIG. 7. (a) The accuracy of detecting GHZ entanglement structures with noise. (b) The accuracy of detecting *W* entanglement structures with noise.

First, based on the same structure, our model can address a broader range of problems. In this work, we used the same trained neural network for both the classification of GHZ and *W* states and the detection of specific entanglement structures. If we were to expand the dataset, it would also allow for the classification of cat, Gaussian, and Gottesman-Kitaev-Preskill (GKP) states. Additionally, through detailed analysis of quantum states, the model can detect more entanglement structures, which has significant importance in fields such as quantum communication and quantum computing. This demonstrates the great practicality of our model.

Second, our approach exhibits significant accuracy and efficacy. Numerical examples with 3–10 particles show that even with the presence of noise, our model achieves an average accuracy of 98.32% for entanglement state classification and an average accuracy of 95% for entanglement structure detection. Moreover, our model offers a remarkable balance between high performance and rapid data processing. The rapid decrease in the loss value demonstrates the effective convergence of our method. Thus, our approach can effectively reduce time and computational costs and has the potential for application to large-scale systems.

In conclusion, benefiting from the powerful data processing capabilities of computer vision, we obtained high accuracy

for both the classification of entangled states and detection of various entanglement structures with low time and computational costs. Here, we just considered the cases of 3–10 particles because of the limitation of the current quantum state tomography [53,54] rather than the capability of our model. In our next work [55], we adopt an innovative approach to overcome this limitation and apply our model to systems with a large number of qubits.

[1] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935).

[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[3] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[4] C. H. Bennett and D. P. DiVincenzo, Quantum information and computation, Nature (London) **404**, 247 (2000).

[5] G. Jaeger, *Entanglement, information, and the interpretation of quantum mechanics* (Springer, Berlin, 2009).

[6] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Nonlocality and communication complexity, Rev. Mod. Phys. **82**, 665 (2010).

[7] A. K. Ekert, Quantum cryptography based on Bell's theorem, Phys. Rev. Lett. **67**, 661 (1991).

[8] A. Galindo and M. A. Martín-Delgado, Information and computation: Classical and quantum aspects, Rev. Mod. Phys. **74**, 347 (2002).

[9] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2001), Vol. 2.

[10] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, Y.-C. Liang, X. Ma, Y.-A. Chen, and J.-W. Pan, Entanglement structure: Entanglement partitioning in multipartite systems and its experimental detection using optimizable witnesses, Phys. Rev. X **8**, 021072 (2018).

[11] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[12] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Separability-entanglement classifier via machine learning, Phys. Rev. A **98**, 012315 (2018).

[13] K. Chen and L.-A. Wu, Test for entanglement using physically observable witness operators and positive maps, Phys. Rev. A **69**, 022312 (2004).

[14] V. Gulati, Arvind, and K. Dorai, Classification and measurement of multipartite entanglement by reconstruction of correlation tensors on an NMR quantum processor, Eur. Phys. J. D **76**, 194 (2022).

[15] G. Sarbicki, G. Scala, and D. Chruściński, Family of multipartite separability criteria based on a correlation tensor, Phys. Rev. A **101**, 012341 (2020).

[16] R. Horodecki, M. Horodecki, and P. Horodecki, Teleportation, Bell's inequalities and inseparability, Phys. Lett. A **222**, 21 (1996).

[17] B. M. Terhal, Bell inequalities and the separability criterion, Phys. Lett. A **271**, 319 (2000).

[18] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki, Optimization of entanglement witnesses, Phys. Rev. A **62**, 052310 (2000).

[19] O. Gühne and G. Tóth, Entanglement detection, Phys. Rep. **474**, 1 (2009).

[20] D. Chruściński and G. Sarbicki, Entanglement witnesses: Construction, analysis and classification, J. Phys. A **47**, 483001 (2014).

[21] P. J. Cavalcanti, G. Scala, A. Mandarino, and C. Lupo, Information theoretical perspective on the method of entanglement witnesses, arXiv:2308.07744.

[22] C. Chen, C. Ren, H. Lin, and H. Lu, Entanglement structure detection via machine learning, Quantum Sci. Technol. **6**, 035017 (2021).

[23] A. C. B. Greenwood, L. T. H. Wu, E. Y. Zhu, B. T. Kirby, and L. Qian, Machine-learning-derived entanglement witnesses, Phys. Rev. Appl. **19**, 034058 (2023).

[24] Z. Chen, X. Lin, and Z. Wei, Certifying unknown genuine multipartite entanglement by neural networks, Quantum Sci. Technol. **8**, 035029 (2023).

[25] P.-H. Qiu, X.-G. Chen, and Y.-W. Shi, Detecting entanglement with deep quantum neural networks, IEEE Access **7**, 94310 (2019).

[26] L. Zhang, Z. Chen, and S.-M. Fei, Entanglement verification with deep semisupervised machine learning, Phys. Rev. A **108**, 022427 (2023).

[27] Y.-J. Luo, J.-M. Liu, and C. Zhang, Detecting genuine multipartite entanglement via machine learning, Phys. Rev. A **108**, 052424 (2023).

[28] Y. Chen, Y. Pan, G. Zhang, and S. Cheng, Detecting quantum entanglement with unsupervised learning, Quantum Sci. Technol. **7**, 015005 (2022).

[29] J. Gu *et al.*, Recent advances in convolutional neural networks, Pattern Recognit. **77**, 354 (2018).

[30] A. Vedaldi and K. Lenc, Matconvnet: Convolutional neural networks for Matlab, in *Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia* (ACM, New York, NY, 2015), pp. 689–692.

[31] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, Piscataway, NJ, 2021), pp. 10012–10022.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need,

in *31st Conference on Neural Information Processing Systems* (Association of Computational Machinery, 2017).

[33] A. Dosovitskiy *et al.*, An image is worth 16x16 words: Transformers for image recognition at scale, arXiv:2010.11929.

[34] M. Żukowski, A. Zeilinger, M. Horne, and H. Weinfurter, Quest for GHZ states, Acta Phys. Pol. A **93**, 187 (1998).

[35] M. Eibl, N. Kiesel, M. Bourennane, C. Kurtsiefer, and H. Weinfurter, Experimental realization of a three-qubit entangled *W* state, Phys. Rev. Lett. **92**, 077901 (2004).

[36] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, Phys. Rev. A **40**, 4277 (1989).

[37] D. Li, X. Li, H. Huang, and X. Li, Stochastic local operations and classical communication invariant and the residual entanglement for *n* qubits, Phys. Rev. A **76**, 032304 (2007).

[38] Z.-C. Zhang, G.-J. Tian, and T.-Q. Cao, Strong quantum nonlocality for multipartite entangled states, Quantum Inf. Process. **20**, 1 (2021).

[39] Z. Ren, W. Li, A. Smerzi, and M. Gessner, Metrological detection of multipartite entanglement from young diagrams, Phys. Rev. Lett. **126**, 080502 (2021).

[40] G. Tóth and O. Gühne, Detecting genuine multipartite entanglement with two local measurements, Phys. Rev. Lett. **94**, 060501 (2005).

[41] Y. Zhu, Y.-D. Wu, G. Bai, D.-S. Wang, Y. Wang, and G. Chiribella, Flexible learning of quantum states with generative query neural networks, Nat. Commun. **13**, 6222 (2022).

[42] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, Convolutional, long short-term memory, fully connected deep neural networks, in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Piscataway, NJ, 2015), pp. 4580–4584.

[43] M. W. Gardner and S. Dorling, Artificial neural networks (the multilayer perceptron)–A review of applications in the atmospheric sciences, Atmos. Environ. **32**, 2627 (1998).

[44] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets, Nat. Commun. **10**, 5415 (2019).

[45] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).

[46] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, Confusion matrix-based feature selection, in *Proceedings of the 22nd Midwest Artificial Intelligence and Cognitive Science Conference (MAICS)*, Cincinnati, OH, USA (2011), https://ceur-ws.org/Vol-710/paper37.pdf.

[47] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, Applications of deep learning in biomedicine, Mol. Pharmaceutics **13**, 1445 (2016).

[48] W. D. Callisterand and D. Rethwisch, *Materials Science and Engineering: An Introduction*, 7th ed. (John Wiley & Sons, New York, 2007).

[49] G. Flato *et al.*, *Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Cambridge University Press, Cambridge, England, 2013), Chap. 9.

[50] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[51] QPANDA, https://qpanda-tutorial.readthedocs.io/zh/latest/index.html.

[52] OriginQ quantum computing, https://originqc.com.cn/index.html?lang=en_US.

[53] A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography, Rev. Mod. Phys. **81**, 299 (2009).

[54] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, npj Quantum Inf. **6**, 20 (2020).

[55] R. Li, Z. Qin, S. Zhang, C. Du, Y. Zhou, and Z. Xiao, Entanglement structure detection with low cost (unpublished).