

Hybrid Quantum Deep Learning With Superpixel Encoding for Earth Observation Data Classification

Fan Fan¹, Yilei Shi, *Member, IEEE*, Tobias Guggemos, and Xiao Xiang Zhu², *Fellow, IEEE*

Abstract—Earth observation (EO) has inevitably entered the Big Data era. The computational challenge associated with analyzing large EO data using sophisticated deep learning models has become a significant bottleneck. To address this challenge, there has been a growing interest in exploring quantum computing as a potential solution. However, the process of encoding EO data into quantum states for analysis potentially undermines the efficiency advantages gained from quantum computing. This article introduces a hybrid quantum deep learning model that effectively encodes and analyzes EO data for classification tasks. The proposed model uses an efficient encoding approach called superpixel encoding, which reduces the quantum resources required for large image representation by incorporating the concept of superpixels. To validate the effectiveness of our model, we conducted evaluations on multiple EO benchmarks, including Overhead-MNIST, So2Sat LCZ42, and SAT-6 datasets. In addition, we studied the impacts of different interaction gates and measurements on classification performance to guide model optimization. The experimental results suggest the validity of our model for accurate classification of EO data. Our models and code are available on <https://github.com/zhu-qlab/SEQNN>.

Index Terms—Image classification, quantum circuit, quantum image encoding, quantum machine learning (QML), superpixel.

I. INTRODUCTION

IMAGE classification contributes significantly to deriving important information from Earth observation (EO) data, providing valuable insights into various applications, such as urban planning, environmental monitoring, and resource management. The utilization of machine learning techniques in the EO field enables automatic analysis and interpretation of large amounts of EO data. In recent years, numerous contributions have been made [1], [2], [3]. To accurately handle the complexity and richness of EO data, large machine learning models and complex training strategies have been proposed.

Received 12 January 2024; revised 30 September 2024; accepted 3 December 2024. Date of publication 1 January 2025; date of current version 4 June 2025. The work of Xiao Xiang Zhu was supported by Munich Center for Machine Learning. (*Corresponding author: Xiao Xiang Zhu.*)

Fan Fan is with the Chair of Data Science in Earth Observation (SiPEO), Technical University of Munich (TUM), 80333 Munich, Germany, and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany.

Yilei Shi is with the School of Engineering and Design, TUM, 80333 Munich, Germany.

Tobias Guggemos is with IMF, DLR, 82234 Weßling, Germany, and also with the Faculty of Physics, Vienna Center for Quantum Science and Technology (VCQ), University of Vienna, 1090 Vienna, Austria.

Xiao Xiang Zhu is with SiPEO, TUM, 80333 Munich, Germany, and also with Munich Center for Machine Learning, 80333 Munich, Germany (e-mail: xiaoxiang.zhu.ieee@gmail.com).

Digital Object Identifier 10.1109/TNNLS.2024.3518108

For instance, Li et al. [4] introduced a co-training procedure to combine knowledge from different tasks; Chen et al. [5] used the state-space model to enhance the representational capacity and efficiency of EO data analysis; Kang et al. [6] proposed a 2-D spectral representation for hyperspectral image classification, and Duan et al. [7] exploited a self-supervised learning method to train a deep neural network from unlabeled hyperspectral data. However, training such models often requires high computational power. In addition, with the advancements in remote sensing technologies, EO has entered the Big Data realm. As a result, there is a growing demand for computation power, which has become a significant bottleneck for efficient and accurate analysis of EO data with classical machines. To tackle this challenge, quantum computing might be one of the solutions in the future since it is expected to accelerate computationally expensive tasks [8].

However, in the Noisy Intermediate-Scale Quantum (NISQ) era [9], quantum machines are constrained by a limited number of qubits and quantum gates, and they lack full fault tolerance. Despite these limitations, researchers have been actively investigating the potential of quantum computing for various tasks [10], [11], [12].

Regarding image classification with quantum machine learning (QML), the quantum models typically involve two main steps: encoding the image data into quantum states and performing the neural network for analysis and classification. The image encoding step is preliminary and challenging, given its requirement for a substantial allocation of quantum resources, which could impact the overall computation speed-up achieved in the subsequent steps. To tackle this, most proposed algorithms have taken a hybrid approach, using quantum algorithms either for feature extraction only from the local area [13] or for final prediction based on the high-level global features obtained from classical algorithms [14].

Therefore, the validity of using quantum computing for image comprehension remains uncertain, and this study aims to make a contribution in this regard. However, the available encoding techniques are still insufficient for exploring the full potential of QML for image understanding in the NISQ era, particularly for EO images, which generally contain large amounts of spatial and spectral information.

To fill this research gap, we present a novel hybrid quantum neural network (SEQNN) that introduces a new approach for image representation, superpixel encoding, aimed at achieving accurate classification of EO images while using reduced quantum resources.

We evaluated our model's validity using three EO benchmarks, namely, Overhead-MNIST [15], So2Sat LCZ42 [16], and SAT-6 [17]. The experimental results show that our model can outperform more complex baseline models in terms of classification accuracy, indicating that our model can effectively extract features from EO images. Furthermore, we experimentally examined the impacts of different interaction gates, measurements, and structures of our model on classification performance. Notably, we adopted two different approaches to assess the significance of the features measured from different bases. These investigations could contribute to enhancing the understanding of quantum model design for image classification tasks.

Regarding the efficiency of our model for image classification, the utilization of superpixel encoding allows encoding EO images into quantum states with reduced quantum resources compared with conventional encoding techniques, such as FRQI [18] and MCQI [19]. In addition, the number of operations required for convolutional transformation in our model's feature extraction part remains independent of the input's spatial dimensions. Besides that, the experiments show that despite having fewer trainable parameters, our model can surpass more complex models. Those factors contribute to efficiency advantages of our model for EO data classification.

In summary, the main contributions of this work include the below.

- 1) We propose a new hybrid quantum-classical neural network (SEQNN) that can effectively extract features using quantum circuits for accurate EO data classification.
- 2) We introduce a hybrid encoding technique, superpixel encoding, to efficiently represent classical images in the quantum domain while reducing quantum resources.
- 3) We investigate the impacts of quantum gates, measurements, and structure of our model through various approaches, enhancing the understanding of quantum model design for image classification.

This article is organized as follows. In Section II, we briefly introduce the background knowledge about quantum computing and QML. Section III provides an overview of related work for image encoding and classification with quantum computing. Section IV presents our proposed hybrid neural network, and Section V describes the experimental evaluation of our model's effectiveness. In Section VI, we analyze the efficiency of our model and discuss the optimization of the quantum components for classification. Finally, Section VII presents the conclusion and outlines future research directions.

II. BACKGROUND

In this section, we briefly introduce some fundamental concepts of quantum computing and QML. Interested readers may refer to the work [20] for more detailed information.

A. Quantum Computing

Quantum computing is a computational methodology that leverages quantum mechanics for performing computational

tasks. A qubit is the basic unit for quantum computing, analogous to a bit in classical computing. However, unlike classical bits, a quantum state can be a linear combination of basis states (e.g., $|0\rangle$ and $|1\rangle$), a property known as superposition. In addition, qubits can be entangled, meaning the state of one qubit can affect the state of the other once they are entangled. For information processing, it is necessary to manipulate qubits and change their states. This is achieved using quantum gates, which are analogous to classical logic gates, but the operation of the quantum gate must be unitary. The quantum circuit model is one of the most widely used approaches in quantum computing, using a sequence of quantum gates to transform qubits from the initial state to the final states for computational purposes.

To conduct quantum computing with the quantum circuit model, there are generally three steps.

1) *Data Encoding*: The first step of quantum computing is to prepare the data into a quantum state as the input for quantum algorithms. When analyzing images using quantum computing, it is essential to represent the information contained in the image as a quantum state. Several methods for image encoding have been proposed so far. Section III will provide further details.

2) *Quantum State Transformation*: Transforming the quantum state from the encoding state to the desired state is crucial in quantum computing. To achieve this, a quantum circuit is constructed consisting of a sequence of quantum gates that manipulate qubits and transform their states through operations such as entanglement and rotation.

3) *Measurement*: This step occurs at the end of the quantum circuit, converting the output of the computation from the quantum domain into a classical state. The resulting classical data represent the solution to the computational task.

B. Quantum Machine Learning

When applying quantum computing to machine learning or deep learning tasks, a parameterized quantum circuit (PQC) is widely used. As a hybrid framework, the quantum circuit, which defines the structure of the QML model, includes a set of gates with parameters that are optimized during the training process on a classical computer. The evolution and measurement of the quantum states are then performed on the quantum machine.

III. RELATED WORK

In recent years, there has been an increasing amount of literature focusing on QML for image classification despite the current constraints of quantum hardware. This section provides an overview of related studies, covering two aspects: techniques for encoding classical data into the quantum state and studies applying QML for image classification, both in general and in the EO domain. Note that unless otherwise specified, the classical algorithms in this article refer to methods that do not involve quantum computing.

A. Image Encoding for Quantum Computing

To process classical data with a quantum computer, the data have to be represented as a quantum state, and there are two

basic encoding approaches: basis embedding and amplitude embedding. The former adopts a computational basis state of a quantum system to represent each input, which requires the input data to be in the form of binary strings. Regarding amplitude embedding, it associates the input data with the amplitude of a quantum state, which requires the input data to be normalized.

When considering the encoding of a classical image into a quantum state, there have been various techniques that can be categorized based on whether the encoded output can faithfully retrieve the original image.

1) *Retrievable Encoding*: The methods grouped in this category enable the retrieval of the original image from the obtained quantum state after encoding. There are three classic methods aiming to encode gray-scale images: Qubit Lattice [21] assigns a single qubit to one pixel's value; Real Ket [22] and FRQI [18] use superposition to encode the spatial coordinates and the amplitude to represent the corresponding pixel values. Besides, new approaches have been developed to address different challenges in image encoding, and interested readers may refer to the study [23]. However, it is still challenging to encode images efficiently for practical applications with current methods due to the high requirement for quantum resources.

2) *Approximation Encoding*: Besides encoding the exact input data, several studies, e.g., [24], [25], [26], attempted to prepare a quantum state with reduced quantum resources to approximate the retrievable encoding output.

3) *Feature Map Encoding*: These approaches transform the input image into quantum Hilbert spaces as a feature map. The output of the encoding process cannot be used to obtain the original image, but it serves as a representation of the input image for analysis purposes. The quantum circuits for mapping can be fully dependent on the input, such as [27] or be parameterized by the trainable parameters along with the input, such as [28] and [29]. Regarding mapping images into a quantum feature space, studies [30], [31] first split images into patches and use a small number of qubits to encode the image patch by patch.

4) *Classical Algorithm-Assisted Encoding*: In addition to the approaches introduced above, researchers have also attempted to adopt classical deep learning techniques to reduce the images' features for the enhancement of data encoding efficiency, such as [32], [33], and [34].

B. QML for Image Classification

To classify classical images with quantum computing, various studies have been conducted. Based on the roles that quantum computing plays in classification, we categorize these studies into three groups.

1) *Pure QML Model*: This type of model aims to exclusively use quantum computing without relying on classical components to classify images.

To date, there are a limited number of studies on pure QML models for image classification due to the limitations of quantum hardware in the NISQ era. Nevertheless, notable contributions, exemplified by [35] and [36], have also been made for image classification.

2) *Quantum-Classical Model*: This type of hybrid model uses quantum computing to process and transform low-level features and classical computing to perform classification based on the transformed features.

For this type of model, quantum computing can be used to process the given image patch by patch for local feature extraction. The obtained features from the entire image are subsequently used by classical algorithms for further feature extraction and classification. For example, Henderson et al. [37] used random circuits, and Riaz et al. [38] used the strongly entangled circuits for non-trainable encoding to locally transform the input image with the sliding window technique, and the resulting features are fed as input to classical deep learning models for further feature extraction and final classification. In contrast, Matic et al. [39], Chen et al. [40], and Wang et al. [41] proposed trainable embedding circuits for patch encoding, and the measured features from the entire image were then used for classification with a dense layer. In addition, there are proposed models that encode the entire image into a quantum state for feature extraction, followed by the usage of classical algorithms for final classification. For instance, the study [42] adopts FRQI [18] to represent images as quantum states. Other studies, such as [43] and [44], encode individual pixel values into quantum states using one qubit per pixel. However, due to the computational power constraints at the time, the input images had to be downsized in these studies, which could negatively influence the classification performance for practical applications due to information loss.

In the field of EO, Chalumuri et al. [13] presented a hybrid model in which quantum computing conducts the rowwise processing of gray-scaled EO images for feature extraction, and the obtained features are then used as input to a classical fully connected deep neural network for classification. Sebastianelli et al. [45] used a quantum convolution method to preprocess multidimensional EO data for classification. In another study, Majji et al. [46] used quantum computing to fuse the SAR data and optical data in a pixelwise manner for classification with a CNN model. Their experiments demonstrated improved performance compared with a model only using SAR data for classification.

3) *Classical-Quantum Model*: Compared with the hybrid models described above, this type of model uses classical machine learning or deep learning algorithms to extract high-level features from images, whereas quantum computing is primarily adopted to transform these features or perform the final classification accordingly.

For instance, to classify EO data, Gawron and Lewiński [14] used principal component analysis, Zaidenberg et al. [47] adopted a CNN model, Chang et al. [48] applied a convolutional autoencoder (CAE), and Otgonbaatar and Datcu [49] used a VGG16 with a CAE to extract high-level features as inputs for their proposed quantum algorithms. When dealing with multicategory classification, Sebastianelli et al. [50] introduced a model with three parts: a classical CNN for feature reduction, a quantum circuit for feature transformation, and a classical dense layer for final prediction.

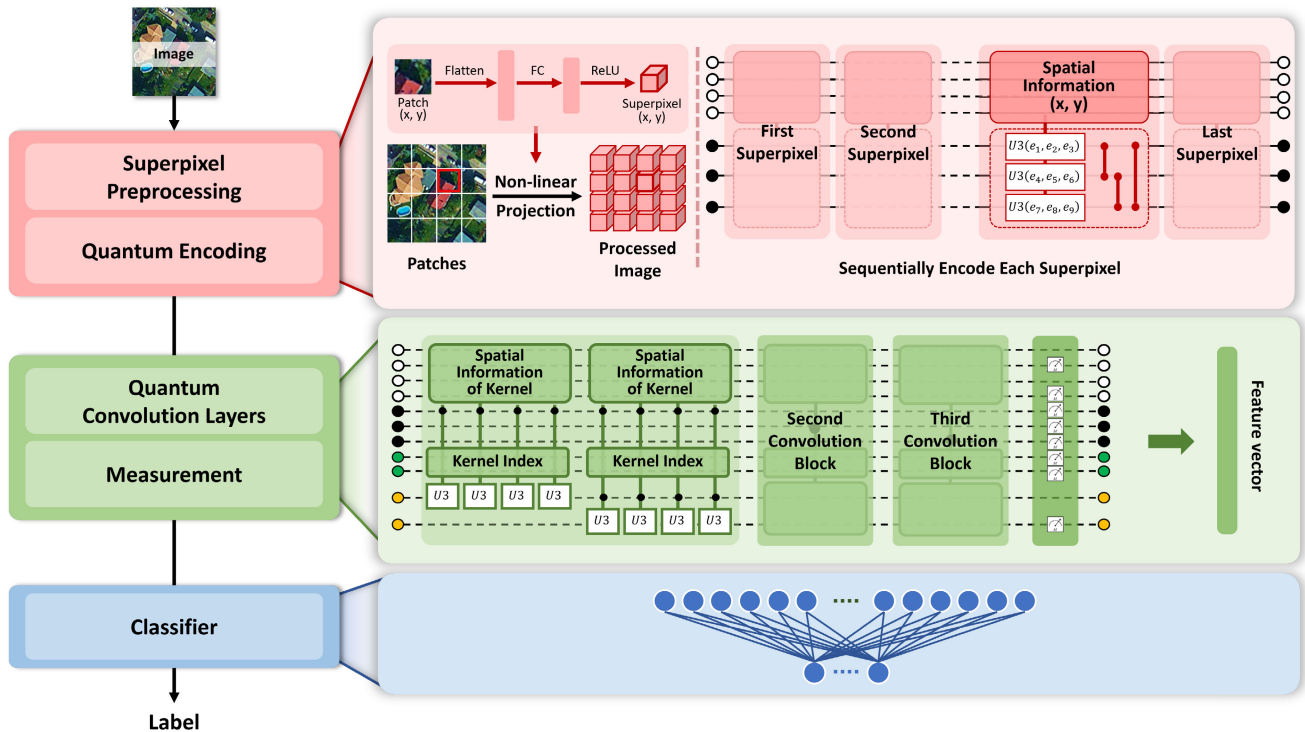


Fig. 1. Structure of the SEQNN model: 1) superpixel preprocessing: the original input image is split into 4×4 patches in the illustrated example; each patch is nonlinearly projected to one superpixel represented by nine features; the processed image's spatial dimension is 4 and each superpixel has nine elements; 2) quantum encoding: the obtained superpixels are individually encoded into a quantum circuit; color indicates the quantum register (four white qubits for q_l to represent the spatial indexing and three black qubits for q_e to encode the nine elements); U3 represents U3 gate and e_i indicates i th element for the corresponding superpixel; the red lines connecting each pair of qubits in q_e indicate the two-qubit gates; 3) quantum convolution layers: these layers aim to conduct convolutional operations in the quantum domain to extract high-level features from the encoded image; color indicates the quantum registers (white for q_l and black for q_e from superpixel encoding step; green for q_k , and yellow for q_r for feature extraction); U3 represents U3 gate and dot markers indicate the controlled state $|1\rangle$; 4) measurement: it generates a feature vector in the classical state derived from the generated quantum feature maps; and 5) classifier: one classical dense layer with a softmax activation function is used as one example for final classification.

IV. METHODOLOGY

In this study, we introduce a hybrid quantum neural network (SEQNN) designed to enhance classification performance in EO data analysis while minimizing quantum resource usage. Fig. 1 illustrates the structure of our model, which comprises three components: superpixel encoding, feature extraction, and classification. A detailed description of each component is provided below.

A. Superpixel Encoding

Image encoding plays a crucial role in image classification with quantum computing. As discussed in Section III, various techniques have been investigated and developed. In this article, we propose a new image encoding technique called superpixel encoding, aimed at effectively representing images as quantum states while reducing the required quantum resources. Fig. 1 provides an overview of this process, which is divided into two tasks as follows.

1) *Superpixel Preprocessing*: In our encoding method, superpixels are designed to capture the information contained within several pixels in the same region of the input image. To represent the image using superpixels, inspired by [51], we first split the image into fixed-size patches, with each patch corresponding to one superpixel. Unlike the model proposed in [51], we preserve the spatial relationships of the superpixels

in two dimensions for further analysis, as illustrated in Fig. 1. Therefore, positional embedding is not required.

To reduce the number of features required for encoding and enable the integration of nonlinearity for the further process, we use a trainable nonlinear projection for all the patches. To be more specific, all the features within a patch are first flattened, and a fully connected layer with trainable parameters is used for their transformation. This fully connected layer incorporates the ReLU activation function afterward, setting negative values to zero. This approach introduces nonlinearity into the transformation and also reduces the number of quantum gates needed for encoding, as feature value zero allows the corresponding gates to be omitted. In the end, this transformation yields an element vector for the corresponding superpixel.

In this manner, we represent the input image with a reduced number of superpixels, each associated with an element vector. To be more specific, let us consider an image denoted as $I \in \mathbb{R}^{N \times N \times C}$, where N is the spatial resolution, C is the number of channels, and \mathbb{R} denotes the space of real-valued numbers. In this example, the given image is divided into patches of size $P \times P$, and each patch is transformed into a superpixel representation with E elements via the nonlinear projection. As a result, the input image can be represented as $I' \in \mathbb{R}^{(N/P) \times (N/P) \times E}$, in which the spatial dimensions are reduced from N to (N/P) , and each superpixel requires E

elements to represent. The choice of patch size P and the number of elements E should be adapted based on the specific requirements of the tasks and the available computational resources.

For encoding, the superpixels' element vectors and their corresponding locations in the processed image I' will be represented with a quantum state.

2) *Quantum Encoding Processing*: To efficiently encode the processed image I' into a quantum state for high-level feature extraction using quantum computing, our method leverages the quantum superposition property.

Specifically, it requires two quantum registers. The first register, denoted as q_l , is used to encode the locations of superpixels, while the second register, denoted as q_e , stores the elements of the superpixels. Due to quantum superposition, $2\log_2(N/P)$ qubits from the q_l register are sufficient to represent all possible (x, y) -coordinates of superpixels. To encode the elements of each superpixel in the second register, we use the U3 gate, which is capable of rotating the qubit with three Euler angles and, accordingly, encodes three different elements. Therefore, $E/3$ qubits from the register q_e are enough. To establish a mapping between the elements of the superpixels and their location information, we apply the controlled U3 gates, and the spatial coordinates decide the controlling states. For each superpixel, once all the E elements have been uploaded with the controlled U3 gates, we proceed to apply two-qubit gates between each pair of qubits in q_e . This type of two-qubit interaction configuration is named all-to-all configuration, and as indicated in [52], the usage of two-qubit gates in this configuration could generally improve the expressibility and entangling capability of the quantum circuits. In this way, we upload the processed image I' into a quantum state $|I'\rangle$.

Regarding the two-qubit gates between qubits in q_e , there are various options, such as CZ, XX, YY, ZZ, CX, SWAP, and ISWAP gates. Each of those gates offers unique operations. The CZ gate, for instance, phases $|11\rangle$ state while leaving others unchanged. The XX gate applies the Pauli-X gate to both the qubits simultaneously, whereas the YY and ZZ gates apply the Pauli-Y and Pauli-Z, respectively, to both the qubits. The CX gate swaps $|11\rangle$ and $|10\rangle$ states, whereas the SWAP gate swaps $|01\rangle$ and $|10\rangle$ states. As for the ISWAP gate, it not only swaps $|01\rangle$ and $|10\rangle$ states but also adds a relative phase. The study of the impacts of these gates on the image classification performance of our model is detailed in Section V-A.

B. Feature Extraction

To extract high-level features from the encoded data $|I'\rangle$, our model uses a sequence of quantum convolution layers that produce quantum feature maps, denoted as $|f_m\rangle$. To obtain the derived feature values from these feature maps, a measurement layer at the final stage of feature extraction is used.

1) *Quantum Convolution Layers*: Convolutional operations play an important role in feature extraction in deep learning. The quantum convolution layers in our model perform the convolution operation within the quantum domain, with the quantum circuit based on the QC-CNN model [42]. To achieve

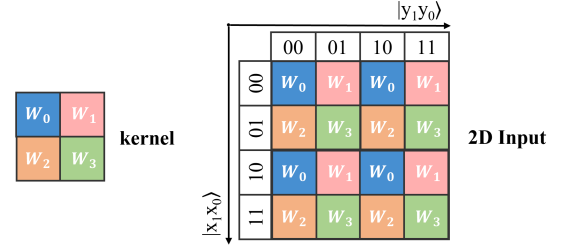


Fig. 2. Illustration of the convolutional operation in our model using a 2×2 -sized kernel with weight values from W_0 to W_3 ; the quantum states $|x_1, x_0\rangle$ and $|y_1, y_0\rangle$ represent the two dimensions of the input image, which has a spatial size of 4×4 .

rapid dimension reduction for feature maps without the need for pooling layers, we match the convolution stride with the kernel size, which is adapted to the input image's dimensions. In this article, we demonstrate our model using 2×2 kernels for clarity and illustrative purposes.

As illustrated in Fig. 1, the quantum convolution layers in our model are organized into convolution blocks, with each block corresponding to a qubit in the q_e register for feature extraction. Within each block, multiple quantum convolution layers can be applied to produce feature maps. For each convolution layer, multiple kernels can be used, each indicated by a state of the q_k register. A single kernel requires four U3 gates with trainable parameters applied to the register q_r , corresponding to the specified kernel size of 2×2 .

Fig. 2 shows the convolution operation in our model, clarifying why only four gates, equivalent to four operations, are sufficient to apply this kernel across the entire input for convolution transformation. Specifically, to use a kernel on a 2-D input with convolution operations, the elementwise products between the kernel and all the input regions must be computed. For the example in Fig. 2, the values of the 2-D input colored the same should be transformed with the same weight. To identify the coordinates of this 4×4 -sized input, four qubits $|x_0, x_1, y_0, y_1\rangle$ from the register q_l are required. Notably, as shown in Fig. 2, the state of $|x_0, y_0\rangle$ from q_l can be used to determine the coordinates of the superpixels that should be transformed with the same weight, particularly, $|00\rangle$ for the superpixels with the weight W_0 , $|01\rangle$ for W_1 , $|10\rangle$ for W_2 , and $|11\rangle$ for W_3 . Thus, for the given kernel, only four gates are sufficient for the transformation regardless of the input's size, which provides efficiency advantages for feature extraction.

Ultimately, after a series of convolution blocks, the quantum feature maps $|f_m\rangle$ with the required dimensions are generated. The high-level features within $|f_m\rangle$ can then be obtained through measurements.

2) *Measurement*: To perform measurement on the generated quantum feature maps, a set of measurement operators need to be identified. For each operator M , an expectation value of the quantum state $|f_m\rangle$ is obtained following (1). The expectation values construct a feature vector for each input image for further classification

$$E(M) = \langle f_m | M | f_m \rangle. \quad (1)$$

To be more specific, the operators are defined according to the chosen basis for measurement, such as the X-basis, Y-basis, or Z-basis. When considering a particular basis formed by states $|b\rangle$, one operator M_i for the desired state $|b_i\rangle$ can be identified as $|b_i\rangle\langle b_i|$. The expectation value $E(M_i)$ is the i th feature in the formed feature vector.

In our model, various measurement bases can be used on the quantum feature maps to obtain distinct feature vectors, and the features measured from different basis can also be concatenated to form a larger feature vector. However, note that increasing the number of measured basis will significantly increase the number of runs of the quantum circuits. Therefore, the number of features comes at the cost of the overall efficiency of our model.

For this study, we focus on three basic measurement bases: X-basis, Y-basis, and Z-basis. We prepared the corresponding operators using Pauli-X, Pauli-Y, and Pauli-Z operators. Regarding the impacts of different measurements on the classification performance of our model, including single-basis measurement versus multiple-basis measurement, as well as the comparison between the X-basis, Y-basis, and Z-basis measurements, we conducted experiments, which are described in Section V-B.

C. Classification

Given the obtained feature vector for the input image, a classifier is used for the final classification. In this study, we use a classical dense layer as an example, in which all the neurons are fully connected, and each neuron encodes one feature value from the feature vector. In the end, a softmax activation function is used to yield a probability distribution for multicategory classification tasks.

V. EXPERIMENTS

To examine the impact of different interaction gates and measurements in the SEQNN model, as well as to validate its quantum circuit for EO classification tasks, we conducted a series of experiments using various EO benchmarks. Furthermore, we compared our model with different deep learning models to provide a comprehensive evaluation of our model's performance.

Data Preparation: In our experiments, we investigated our model's structure and evaluated its performance using three EO benchmarks, namely, Overhead-MNIST [15], So2Sat LCZ42 [16], and SAT-6 [17]. However, to conduct the experiments efficiently, we made adjustments to the size of the datasets mentioned above by narrowing our focus to specific categories or reducing the labeled samples within each benchmark. The details of our experimental data are provided below.

Overhead-MNIST [15] consists of gray-scaled overhead view images (28×28) with a balanced ratio of ten kinds of entities. In this study, we set five kinds of entities ("car," "ship," "plane," "harbor," and "parking lot") as our target classes. For our experiments, we randomly selected 15% samples from the available 4443 training images to create the validation dataset, and the remaining samples were used to form the training dataset. As for testing, we used all 637 testing

images from these target classes in its test dataset. All the images were enlarged to 32×32 by padding with zeros for the experiments.

So2Sat LCZ42 [16] has around half a million Sentinel-1 and Sentinel-2 image patches with the size of 32×32 from 42 cities with 17 LCZ labels. In our experiments, we focused on the annotated Sentinel-2 patches from 3 German cities (Berlin, Munich, and Cologne) and used four bands (red, green, blue, and near-infrared) for classification evaluation. To simplify the experiments, we combined the LCZ labels into five semantic classes following the land cover classification scheme [53]. For each semantic class, we randomly sampled 2000 patches from these three cities, and the selected patches were further divided into three sets: 60% for training, 20% for validation, and 20% for testing, with no overlap in between.

SAT-6 [17] comprises over 0.4 million images, each with a size of 28×28 pixels, covering six land cover classes. The images in the dataset have four channels. For our experiments, we randomly selected 4200 samples from the training data for the training dataset and another 1200 samples for the validation dataset. Regarding the test dataset, we randomly sampled 200 images in each class from the test data. Each sample was also resized to 32×32 by padding with zeros. Eventually, we prepared three balanced datasets, each containing 4200, 1200, and 1200 samples, respectively.

Model Preparation: In our experiments, we mainly evaluated our model for image classification using a total of 12 qubits. Among these, 9 qubits were used for the superpixel encoding to process 64 superpixels, with each consisting of nine elements after preprocessing from a 4×4 -sized patch. The remaining qubits included one qubit to indicate the index of the applied kernels and the resulting feature maps, while two qubits were used to encode the values of the generated feature maps following the quantum convolution operation.

Experimental Setting: We trained the aforementioned model for 200 epochs with a learning rate of 0.01, using a batch size of 50. Each training was repeated three times, and we compared and discussed the averaged classification accuracy of the trained model with the lowest validation loss value.

To train models that incorporate quantum computing, we used a noiseless simulator provided by the TensorFlow Quantum platform [54]. This simulator produces analytic results, enabling us to verify the validity of our model without accounting for the noise effects.

A. Analysis of the Two-Qubit Gates for Superpixel Encoding

To study the effects of the two-qubit gates in our image classification model, we trained and compared our model with and without the inclusion of these gates. In this experiment, we focused on the following gates: CZ, XX, YY, ZZ, CX, SWAP, and ISWAP. Note that these two-qubit gates are not trainable in our models. Thus, by comparing the test accuracy achieved by the trained models with or without different gates, we can analyze the impacts of these gates on our model's classification performance.

To mitigate the potential influence of varying measurement strategies on performance, we incorporated all the features

TABLE I

CLASSIFICATION PERFORMANCE COMPARISON BETWEEN THE MODEL WITH DIFFERENT TWO-QUBIT GATES

Data	Gate	Training Acc.	Validation Acc.	Test Acc.
Overhead-MNIST	\	0.958 ± 0.011	0.918 ± 0.008	0.906 ± 0.007
	CZ	0.948 ± 0.007	0.913 ± 0.012	0.923 ± 0.013
	XX	0.950 ± 0.012	0.916 ± 0.000	0.902 ± 0.009
	YY	0.944 ± 0.015	0.918 ± 0.013	0.902 ± 0.005
	ZZ	0.944 ± 0.009	0.913 ± 0.007	0.902 ± 0.008
	CX	0.927 ± 0.012	0.901 ± 0.008	0.901 ± 0.010
	SWAP	0.945 ± 0.009	0.913 ± 0.001	0.909 ± 0.004
	ISWAP	0.939 ± 0.007	0.906 ± 0.003	0.904 ± 0.020
	So2Sat LCZ42	\	0.931 ± 0.007	0.904 ± 0.009
CZ		0.929 ± 0.008	0.909 ± 0.006	0.905 ± 0.005
XX		0.935 ± 0.008	0.907 ± 0.000	0.903 ± 0.003
YY		0.937 ± 0.027	0.908 ± 0.019	0.900 ± 0.015
ZZ		0.932 ± 0.015	0.907 ± 0.009	0.898 ± 0.007
CX		0.920 ± 0.007	0.887 ± 0.011	0.883 ± 0.017
SWAP		0.929 ± 0.016	0.909 ± 0.011	0.901 ± 0.009
ISWAP		0.925 ± 0.042	0.897 ± 0.027	0.888 ± 0.033
SAT-6		\	0.970 ± 0.010	0.951 ± 0.012
	CZ	0.976 ± 0.008	0.954 ± 0.005	0.951 ± 0.001
	XX	0.971 ± 0.009	0.958 ± 0.003	0.951 ± 0.004
	YY	0.967 ± 0.010	0.946 ± 0.010	0.947 ± 0.005
	ZZ	0.976 ± 0.005	0.954 ± 0.006	0.944 ± 0.008
	CX	0.969 ± 0.007	0.944 ± 0.002	0.937 ± 0.001
	SWAP	0.977 ± 0.007	0.953 ± 0.000	0.946 ± 0.001
	ISWAP	0.976 ± 0.003	0.952 ± 0.006	0.945 ± 0.004

the highest average test accuracy is in bold

measured from the X-basis, Y-basis, and Z-basis into a feature vector for the final classifier.

Table I displays the classification performance of our trained models on the experimental datasets. The experimental results show that using two-qubit gates in superpixel encoding can enhance classification accuracy. Moreover, the test accuracy-based classification performance indicates that the CZ gates exhibit advantages when used in image classification tasks, but the differences between different gates are quite marginal. In the following experiments, unless otherwise stated, we adopt the CZ gates in our model for evaluation.

B. Analysis of the Measurement for Feature Extraction

To construct a feature vector for the final classification of an input image, we can apply the measurements to the generated quantum feature maps from various bases, such as X-basis, Y-basis, and Z-basis. This enables us to obtain distinct features. For the final classification, we can also concatenate the measured features from multiple bases into a larger feature vector. However, note that the increase in measured features results in an additional cost in terms of the number of runs of quantum circuits and the trainable parameters of the subsequent classifier. Therefore, it is imperative to investigate the significance of the features from different measurements in relation to the classification performance since it allows us to optimize the measurement process of our model and enhance its overall efficiency.

To this end, we first used two approaches, one based on fine-tuning and the other based on integrated gradients (IGs), to evaluate the features obtained from different bases

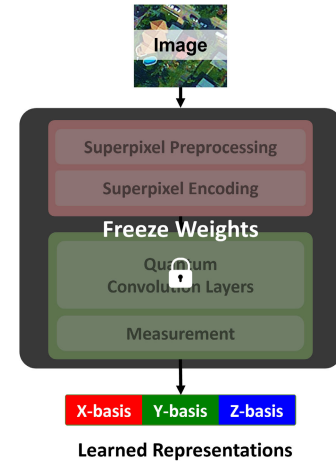


Fig. 3. Learned representations from the pre-trained SEQNN model.

for classification purposes, with the aim of optimizing the measurement process. Subsequently, a model incorporating the optimized measurement was trained, whose performance was compared with that of models using features from different bases to verify the validity of the optimized measurement strategy.

1) *Fine-Tuning-Based Approach*: This method requires a pre-trained SEQNN model that uses the features from all three different bases (X-basis, Y-basis, and Z-basis) for classification. It is worth noting that this pre-trained model adopted the CZ gates. As shown in Fig. 3, the pre-trained model was used to generate learned representations for each input image. Specifically, the parameters within the trained model were frozen, and the dense layer responsible for the final classification was removed. The resulting model produced learned representations for input images, which were subsequently categorized into three groups based on the measured basis. Then, for each group of features, we trained a new model having only one dense layer for classification. In the end, the newly trained models' performances were compared with the pre-trained model to evaluate how representative the measured features were from different bases.

Table II presents the classification accuracy achieved by the retrained models for each dataset, along with a comparison of their performance regarding test accuracy with the pre-trained model. As indicated in the table, the retrained models using all the learned representations (192 features) from input images can achieve comparable classification performance as the pre-trained models. Regarding the models using a subset of the learned representations (64 features), those using the features from the X-basis for classification exhibit superior performance compared with the ones using features from either Y-basis or Z-basis. When compared with the pre-trained model, a decrease in classification accuracy can be observed, which is expected. The reason for that is the pre-trained model that outputs the learned representations was optimized to yield the best features from three bases for classification rather than the features from one single basis.

2) *IG-Based Approach*: The technique, IGs, was introduced in [55], aiming to attribute the predictions of a deep network

TABLE II
COMPARISON OF CLASSIFICATION PERFORMANCE BETWEEN THE RETRAINED MODELS WITH DIFFERENT MEASURED FEATURES

Data	Model	Features / Representations	Training Acc.	Validation Acc.	Test Acc.	Performance Diff.
Overhead-MNIST	Pre-trained	192	0.948 ± 0.007	0.913 ± 0.012	0.923 ± 0.013	\
	XYZ-bases	192	0.949 ± 0.002	0.919 ± 0.005	0.925 ± 0.005	0.002 \uparrow
	X-basis	64	0.851 ± 0.002	0.850 ± 0.005	0.859 ± 0.002	0.064 \downarrow
	Y-basis	64	0.669 ± 0.007	0.671 ± 0.017	0.649 ± 0.015	0.274 \downarrow
	Z-basis	64	0.841 ± 0.001	0.842 ± 0.003	0.824 ± 0.002	0.099 \downarrow
So2Sat LCZ42	Pre-trained	192	0.929 ± 0.008	0.909 ± 0.006	0.905 ± 0.005	\
	XYZ-bases	192	0.925 ± 0.002	0.905 ± 0.001	0.910 ± 0.001	0.005 \uparrow
	X-basis	64	0.909 ± 0.001	0.899 ± 0.001	0.896 ± 0.001	0.009 \downarrow
	Y-basis	64	0.836 ± 0.001	0.839 ± 0.001	0.835 ± 0.004	0.070 \downarrow
	Z-basis	64	0.776 ± 0.001	0.773 ± 0.001	0.766 ± 0.001	0.139 \downarrow
SAT-6	Pre-trained	192	0.976 ± 0.008	0.954 ± 0.005	0.951 ± 0.001	\
	XYZ-bases	192	0.976 ± 0.002	0.959 ± 0.001	0.957 ± 0.002	0.006 \uparrow
	X-basis	64	0.957 ± 0.001	0.941 ± 0.003	0.949 ± 0.001	0.002 \downarrow
	Y-basis	64	0.909 ± 0.002	0.905 ± 0.000	0.887 ± 0.002	0.064 \downarrow
	Z-basis	64	0.915 ± 0.001	0.910 ± 0.000	0.902 ± 0.002	0.049 \downarrow

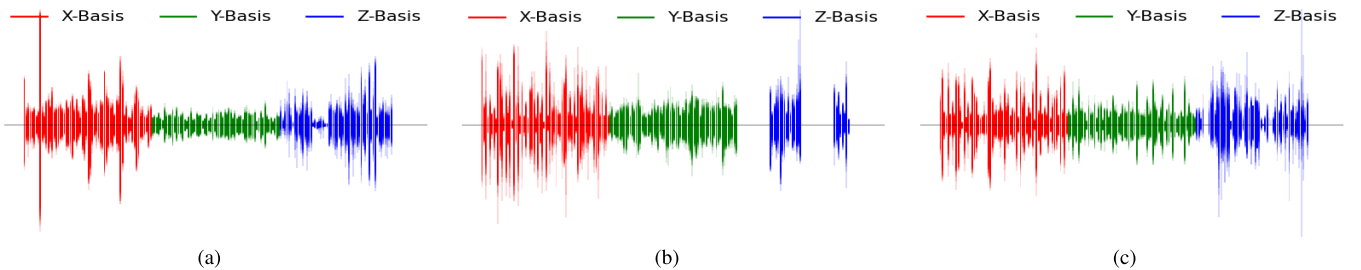


Fig. 4. Visualization of the IGs regarding the training datasets: red for the X-basis, green for the Y-basis, and blue for the Z-basis. (a) Overhead-MNIST. (b) So2Sat LCZ42. (c) SAT-6.

to its input features. For this approach, we also focused on the learned representations obtained from a pretrained SEQNN model performing measurements on three bases, as also depicted in Fig. 3. Accordingly, we established a baseline vector of the same size as the learned representations, where all values were set to zero. For each training sample, we evenly interpolated 240 vectors between the baseline vector and its learned representation, and these vectors were grouped sequentially into batches of size 32. For each batch of the vectors, we calculated the gradients to quantify the changes in the predictions of the final classifier from the pretrained model due to the changes in the learned presentations. The obtained gradients were integrated across the batches and scaled with respect to the learned representations. The IGs provide an indication of the significance level of the features in the learned representations regarding the final classifier’s prediction.

Fig. 4 visualizes the IGs for the trained model with the best performance from the repeated experiments. The features measured from different bases are visually distinguished by color: red for the X-basis, green for the Y-basis, and blue for the Z-basis. The height of each line represents the significance level of the corresponding feature in predicting an arbitrary sample, while the saturation of the color indicates the overall significance level of that feature over all the samples. As depicted in the figure, the features measured from the X-basis are more significant compared with the features from other bases for the final classifier’s prediction.

The findings through the above two approaches align with each other, providing further support for the notion that the X-basis features are more representative for prediction.

To evaluate the effectiveness of solely using X-basis features, we carried out experiments to compare the classification performance of the SEQNN model with the features from all three bases and the one using X-basis features. The results, as shown in Table III, demonstrate that the performance of the model only using X-basis features is comparable to that of the model using the features from all three bases (with the largest difference in test accuracy being 0.01). However, the number of trainable parameters is significantly reduced. Thus, in the following experiments, we mainly used the X-basis features in our model for further evaluation.

C. Analysis of the Structure for Superpixel Encoding

To evaluate the influence of structural modifications on our model for superpixel encoding, we trained two models with different numbers of qubits for superpixel encoding and compared their impacts on classification accuracy. Based on the findings from the previous experiments, the trained models used the CZ-gate and X-basis measurement for classification. In addition, the same structure was adopted for feature extraction and final classification in each trained model to prevent their effects on classification performance.

Table IV presents the experimental results, highlighting the findings related to superpixel encoding. As shown in the

TABLE III
COMPARISON OF CLASSIFICATION PERFORMANCE WITH AND WITHOUT OPTIMIZED MEASUREMENT PROCESS

Data	Model	Parameters	Measured Features	Training Acc.	Validation Acc.	Test Acc.	Performance Diff.
Overhead-MNIST	w/o Optimization	1262	192	0.948 ± 0.007	0.913 ± 0.012	0.923 ± 0.013	\
	w/ Optimization	622	64	0.933 ± 0.008	0.917 ± 0.022	0.913 ± 0.004	0.010 ↓
So2Sat LCZ42	w/o Optimization	1694	192	0.929 ± 0.008	0.909 ± 0.006	0.905 ± 0.005	\
	w/ Optimization	1054	64	0.940 ± 0.009	0.918 ± 0.009	0.914 ± 0.004	0.009 ↑
SAT-6	w/o Optimization	1887	192	0.976 ± 0.008	0.954 ± 0.005	0.951 ± 0.001	\
	w/ Optimization	1119	64	0.977 ± 0.009	0.955 ± 0.005	0.952 ± 0.004	0.001 ↑

TABLE IV
CLASSIFICATION PERFORMANCE OF THE MODELS WITH DIFFERENT STRUCTURES FOR SUPERPIXEL ENCODING

Data	Qubits	Elements / Superpixel	Training Acc.	Validation Acc.	Test Acc.	Performance Diff.
Overhead-MNIST	11	6	0.919 ± 0.004	0.907 ± 0.003	0.898 ± 0.006	\
	12	9	0.933 ± 0.008	0.917 ± 0.022	0.913 ± 0.004	0.015 ↑
So2Sat LCZ42	11	6	0.919 ± 0.005	0.905 ± 0.005	0.903 ± 0.007	\
	12	9	0.940 ± 0.009	0.918 ± 0.009	0.914 ± 0.004	0.011 ↑
SAT-6	11	6	0.963 ± 0.005	0.944 ± 0.006	0.944 ± 0.004	\
	12	9	0.977 ± 0.009	0.955 ± 0.005	0.952 ± 0.004	0.008 ↑

TABLE V
CLASSIFICATION PERFORMANCE OF MODELS WITH QUANTUM CIRCUITS FOR FEATURE EXTRACTION COMPARED WITH THEIR CLASSICAL COUNTERPARTS

Data	Model	Number of Parameters			Training Acc.	Validation Acc.	Test Acc.
		Superpixel Preprocessing	Feature Extraction	Classifier			
Overhead-MNIST	CNN*	153	218	325	0.923 ± 0.006	0.902 ± 0.010	0.886 ± 0.015
	CNN†	153	925	325	0.959 ± 0.008	0.916 ± 0.003	0.907 ± 0.009
	ours	153	144	325	0.933 ± 0.008	0.917 ± 0.022	0.913 ± 0.004
So2Sat LCZ42	CNN*	585	218	325	0.908 ± 0.015	0.886 ± 0.016	0.882 ± 0.002
	CNN†	585	925	325	0.936 ± 0.013	0.904 ± 0.004	0.900 ± 0.006
	ours	585	144	325	0.940 ± 0.009	0.918 ± 0.009	0.914 ± 0.004
SAT-6	CNN*	585	218	390	0.943 ± 0.010	0.924 ± 0.005	0.920 ± 0.002
	CNN†	585	925	390	0.971 ± 0.009	0.948 ± 0.008	0.942 ± 0.007
	ours	585	144	390	0.977 ± 0.009	0.955 ± 0.005	0.952 ± 0.004

table, with one additional qubit for encoding, the model can encode more elements per superpixel (9 versus 6), resulting in improved classification accuracy.

D. Analysis of the Validity of the Quantum Circuit

To evaluate the effectiveness of the quantum component of our model for feature extraction from the processed image after superpixel preprocessing, we compared our model's performance with two CNN-based models (CNN* and CNN† in Table V). Like our model, these models use superpixel preprocessing to transform the input images. The processed images are then fed into a CNN with two convolutional layers for feature extraction, followed by a dense layer for final classification. Thus, these competitors share the same pipeline as our model, except that our model encodes the processed images into the quantum states and performs convolutional operations in the quantum domain for feature extraction.

The difference between CNN* and CNN† lies in the number of kernels in the first convolutional layer. Specifically,

CNN* applies two kernels to ensure that the number of parameters in the convolutional layers is comparable to our model, while CNN† uses nine kernels in the first convolution layer. In the second convolutional layer, both the models use 16 kernels to ensure the extraction of the same number of features. This setup allows us to compare our model's quantum component with its classical counterpart by evaluating the classification performance, thereby investigating its validity.

Table V presents the experimental results. As shown, CNN*, CNN†, and our model have the same amount of trainable parameters for superpixel preprocessing and final classification. However, for feature extraction, our model has only 144 trainable parameters, whereas CNN* and CNN† have 218 and 925, respectively. Despite this difference, our quantum circuit with 144 parameters outperforms both classical CNN models with more parameters (218 and 925), demonstrating the validity of our quantum algorithm for encoding and feature extraction.

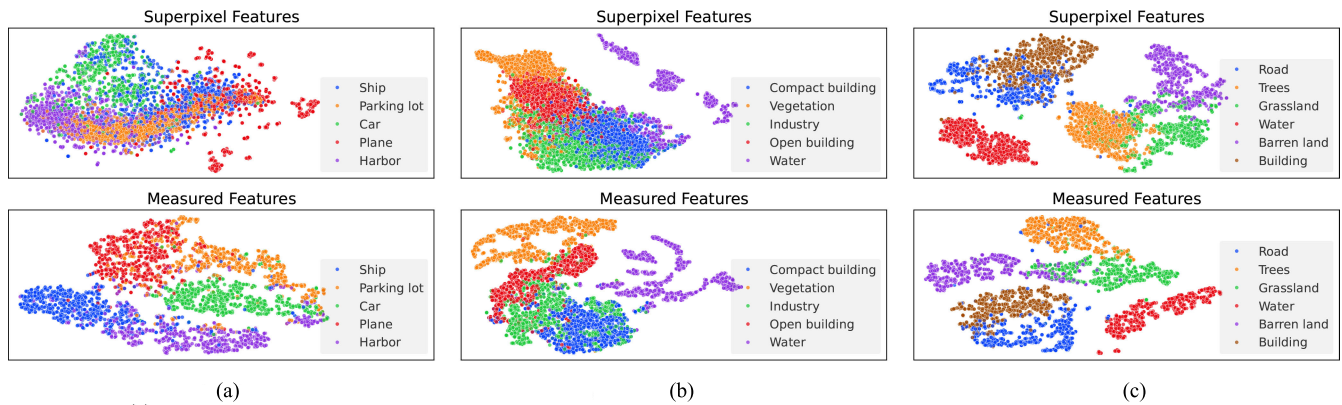


Fig. 5. Visualization of the features from the training samples at different stages of our model using t-SNE: dots represent the samples and colors indicate the corresponding categories. (a) Overhead-MNIST. (b) So2Sat LCZ42. (c) SAT-6.

Fig. 5 visualizes the features from the training samples at two stages of our model (i.e., the processed image for quantum encoding, and the feature vector obtained from our quantum circuit) using t-distributed stochastic neighbor embedding (t-SNE) [59]. In the visualization, each dot represents a sample, and the colors indicate different categories. As shown in the figure, the features extracted by our quantum circuit result in better separation of samples from different categories compared with using the processed image directly, suggesting that our quantum circuit effectively enhances feature extraction for classification.

E. Analysis of the Classification Performance

We compared our model's performance with various competitors to assess its validity for EO data classification tasks. Building upon the previous experimental conclusions in this study, our model with CZ-gates and the measurement in the X-basis was trained and compared for each dataset.

For comparison, we used four classical deep neural networks: CNN, ResNet [56], DenseNet [57], and ViT [51]. These models are commonly used in EO data classification studies and operate on principles similar to our model for image understanding. Specifically, CNN, ResNet, and DenseNet use convolutional operations for feature extraction, similar to our model, while ViT splits images into patches as our model for further analysis. To ensure a fair comparison, we simplified the structures of these models to minimize differences in the number of trainable parameters. The CNN model included four convolutional layers and one dense layer for classification. In the ResNet model, we used two residual blocks where the convolutional layers were configured with four filters. The DenseNet architecture started with one convolutional layer, followed by four dense blocks with transition layers in between. Regarding the ViT model, the patch size was set to 4×4 , and each patch was projected to a vector with four features. Two transformer blocks with a two-headed attention mechanism processed these patches for feature extraction. Moreover, for evaluation purposes, three quantum models, QNN [37], QCNN [50], and HQCNN [58], were also incorporated into the study. The QNN model used a quantum circuit for patchwise processing and a CNN for prediction.

For multispectral images, the quantum circuit processed each spectral band individually, with the output passed to the CNN for final classification. The QCNN and HQCNN models began with a classical CNN to extract 8 and 12 features, respectively. These features were then processed by a quantum algorithm to generate 16 final features, which were subsequently fed into a classical dense layer for classification. It is important to note that the quantum circuits of these models differ.

These competitors were also trained for 200 epochs but with their default learning rates, and each experiment was repeated three times for comparison. Table VI presents the classification accuracy achieved by different models. As indicated in the table, our model with fewer trainable parameters can outperform the competitors with a higher parameter count in terms of test accuracy. This result suggests that our model can extract critical features more efficiently, thereby validating its suitability for EO data classification tasks.

VI. DISCUSSION

This study introduces a hybrid quantum-classical deep learning model with superpixel encoding for EO data classification. Our model demonstrates the capability to process and analyze images using reduced quantum resources during the classification process. In this section, we will interpret our findings from the following aspects and explore the broader implications of our results.

A. Model's Efficiency

The efficiency of the deep learning models that integrate quantum computing is crucial for evaluation, given that the primary expectation of quantum computing is to accelerate analysis. However, considering the current limitations of available quantum machines, we evaluated our model with simulators and analyzed its efficiency from three aspects.

1) *Number of Qubits*: The number of required qubits in QML models is one important criterion, especially in the NISQ era. In comparison to other image encoding techniques, such as MCQI [19], which is known for its qubit efficiency in image encoding, our proposed encoding method has the potential to reduce the required number of qubits.

TABLE VI
CLASSIFICATION PERFORMANCE COMPARISON BETWEEN THE PROPOSED MODEL AND OTHERS

Data	Model	Parameter	Training Acc.	Validation Acc.	Test Acc.
Overhead-MNIST	CNN	1035	0.920 ± 0.028	0.888 ± 0.036	0.871 ± 0.051
	ResNet[56]	1043	0.947 ± 0.006	0.897 ± 0.010	0.884 ± 0.012
	DenseNet[57]	979	0.951 ± 0.010	0.912 ± 0.013	0.904 ± 0.038
	ViT[51]	1321	0.931 ± 0.009	0.910 ± 0.006	0.904 ± 0.003
	QCNN[50]	1065	0.941 ± 0.013	0.918 ± 0.012	0.896 ± 0.019
	HQCNN[58]	873	0.941 ± 0.018	0.912 ± 0.018	0.902 ± 0.023
	QNN[37]	1203	0.917 ± 0.034	0.900 ± 0.010	0.884 ± 0.016
	Ours	622	0.933 ± 0.008	0.917 ± 0.022	0.913 ± 0.004
So2Sat LCZ42	CNN	1223	0.937 ± 0.011	0.911 ± 0.005	0.907 ± 0.006
	ResNet[56]	1619	0.947 ± 0.004	0.919 ± 0.006	0.914 ± 0.004
	DenseNet[57]	1365	0.960 ± 0.012	0.928 ± 0.004	0.918 ± 0.006
	ViT[51]	1665	0.945 ± 0.008	0.914 ± 0.006	0.910 ± 0.007
	QCNN[50]	1601	0.941 ± 0.009	0.909 ± 0.010	0.907 ± 0.008
	HQCNN[58]	1969	0.916 ± 0.007	0.897 ± 0.003	0.898 ± 0.005
	QNN[37]	1709	0.932 ± 0.008	0.896 ± 0.004	0.892 ± 0.008
	Ours	1054	0.940 ± 0.009	0.918 ± 0.009	0.914 ± 0.004
SAT-6	CNN	1288	0.959 ± 0.012	0.938 ± 0.007	0.936 ± 0.004
	ResNet[56]	1652	0.976 ± 0.005	0.941 ± 0.010	0.939 ± 0.001
	DenseNet[57]	1398	0.980 ± 0.006	0.958 ± 0.005	0.949 ± 0.008
	ViT[51]	1794	0.965 ± 0.006	0.944 ± 0.008	0.940 ± 0.005
	QCNN[50]	1618	0.971 ± 0.008	0.942 ± 0.008	0.940 ± 0.012
	HQCNN[58]	1986	0.953 ± 0.009	0.936 ± 0.008	0.925 ± 0.002
	QNN[37]	1766	0.968 ± 0.006	0.948 ± 0.009	0.944 ± 0.001
	Ours	1119	0.977 ± 0.009	0.955 ± 0.005	0.952 ± 0.004

Note: the highest average accuracy is in bold

TABLE VII
QUANTUM RESOURCES NEEDED FOR IMAGE ENCODING

Encoding Method	Qubits	Quantum Gates
Ours	$2 \log_2(\frac{N}{P}) + \frac{E}{3}$	$\frac{EN^2}{3P^2}$ U3 gates with $2 \log_2(\frac{N}{P})$ controllers
		$\frac{N^2E(E-3)}{18P^2}$ two-qubit gates
		$2 \log_2(\frac{N}{P})$ H gates
MCQI [19]	$\log_2(CN^2) + 1$	CN^2 RY gates with $\log_2(CN^2)$ controllers
		$\log_2(CN^2)$ H gates

N: input image's spatial dimension; C: input image's channel; P: patch's size for superpixel; E: elements for superpixel

Specifically, as shown in Table VII, when given an image with a size of $N \times N \times C$, where N represents the spatial dimension and C represents the spectral information, the number of qubits required to represent this image using MCQI [19] is $\log_2(CN^2) + 1$. However, when our superpixel encoding is applied, the input image can be represented as $(N/P) \times (N/P) \times E$, where P denotes the patch size used for superpixel identification, and E represents the number of elements within each superpixel. In total, $2 \log_2(N/P) + (E/3)$ qubits are needed. Thus, when E is smaller than the threshold $3 \log_2(2CP^2)$, our approach could further reduce the number of qubits required for encoding compared with MCQI. Note that each patch from the original image used for identifying superpixels consists of CP^2 values, and based on our previous experience, a smaller number of elements for each superpixel is usually sufficient (e.g., in our experiment, as introduced in Section V, E was set as 9, while the threshold for four-band

images was determined to be 21). Therefore, our encoding technique has the potential to generally reduce the number of qubits required for encoding, and we could modify the patch size P and the element number E for each superpixel based on the tasks and available computational resources.

2) *Number of Gates*: Our proposed hybrid model uses a sequence of quantum gates for superpixel encoding and feature extraction, and each gate corresponds to one operation during the classification process. Thus, the number of gates is commonly used as a metric to evaluate the model's efficiency.

For superpixel encoding, after preprocessing, the processed image's spatial dimension is reduced compared with the input image. As outlined in Section IV-A, our model applies controlled U3 gates for quantum encoding, and the number of controller qubits depends on the spatial dimension of the input. By effectively reducing it, fewer controllers are needed for these U3 gates, thereby reducing the number of elementary

gates needed for decomposition. Thus, when compared with conventional image encoding approaches, such as MCQI [19], our method demands fewer quantum gates to represent images in quantum states.

To be more specific, as indicated in Table VII, given the image with a size of $N \times N \times C$ (where N represents the spatial dimension and C represents the spectral information), MCQI requires CN^2 RY gates with $\log_2(CN^2)$ controllers and $\log_2(CN^2)$ H gates. In contrast, our encoding technique allows us to represent the image with $(EN^2/3P^2)$ U3 gates with $2\log_2(N/P)$ controllers, $((N^2E(E-3))/18P^2)$ two-qubit gates, and $2\log_2(N/P)$ H gates, in which P denotes the patch size and E denotes the number of elements per superpixel. Note that the number of multicontrolled gates dominates image encoding compared with H gates and two-qubit gates. Typically, multicontrolled gates are more complex than those gates because they require a greater number of elementary gates for construction. When comparing the quantity of multicontrolled gates, the ratio between our approach and MCQI is given by $(E/3CP^2)$. In our experiment in Section V, our encoding method requires only around 4.69% multicontrolled gates that are needed to encode four-band images using MCQI. In addition, the number of controllers required for encoding is reduced from $\log_2(CN^2)$ to $2\log_2(N/P)$. Therefore, the count of elementary gates for encoding can be further diminished.

Moreover, as elaborated in Section IV-B, in our model, the number of gates used for feature extraction using quantum convolution layers remains independent of the spatial dimension of the input. In contrast to classical convolution layers, the operation increases quadratically with the spatial size due to the sliding window mechanism. Thus, fewer operators are needed in our model for feature extraction.

3) *Number of Trained Parameters*: To train a model, we should optimize all its trainable parameters during the training process. The model with more trainable parameters generally requires a longer training time.

According to the experimental results in Table VI, our model, despite having fewer trainable parameters, achieves higher test accuracy compared with other competitors with more parameters. Notably, Table V highlights the advantage of the quantum component in our model with respect to the number of parameters. These experimental results indicate that our model demonstrates superior training efficiency.

Thus, based on the above analysis, our model is able to encode EO images with reduced quantum gates and qubits compared with other quantum models using conventional image encoding methods, e.g., MCQI [19]. When compared with classical models that use convolutional operations for feature extraction, our model requires fewer operations. For classification, our model can be trained by optimizing a small number of parameters while achieving accurate classification results. These factors indicate our model's efficiency advantages for EO data classification tasks.

B. Model's Structure

In our study, we conducted an experimental investigation into the effects of various gates and measurements on clas-

sification performance, aiming to provide guidance for our model's optimization.

1) *Gate Selection*: In our experiments, we focused on the effects of different two-qubit gates on superpixel encoding and their impact on classification performance. Specifically, we have examined seven types of gates, including CZ, XX, YY, ZZ, CX, SWAP, and ISWAP gates. Note that these two-qubit gates have no trainable parameters in our model, so the number of parameters of the models with different gates is equal.

As the experimental results shown in Table I, the selection of tested two-qubit gates can lead to different classification performances. However, it is important to note that the observed performance differences were moderate, with the largest test accuracy difference being 0.021. The utilization of CZ gates in superpixel encoding generally provides advantages for our model to accurately classify EO data.

2) *Measurement Selection*: For our model, the measurement aims to yield classical features from generated quantum feature maps. Given a quantum feature map $|f_m\rangle$, we could measure it from any arbitrary basis and obtain unique features for the final classifier to conduct prediction. Therefore, the choice of measurement is significant for our model's validity for feature extraction.

In this study, the primary focus was to investigate three basic bases, i.e., X-basis, Y-basis, and Z-basis. As outlined in Section V-B, the selection of the basis for measurement can have a substantial impact on the validity of feature extraction in our model and, accordingly, classification performance. As demonstrated in Table II and Fig. 4, the features measured from the X-basis exhibit greater significance for final classification compared with those obtained from other bases. The experimental results in Table III indicate that the performance of the model using X-basis features is comparable to the one using features from multiple bases, which validates the effectiveness of X-basis measurement in our model for EO data classification.

VII. CONCLUSION AND FUTURE WORK

In this study, we introduce and evaluate a novel hybrid quantum-classical neural network with superpixel encoding for the classification of EO data. To gain a comprehensive understanding of our model's optimization, we experimentally investigated the impacts of different gates, measurements, and the structure of our model on its classification performance. The experimental results indicate our model can leverage quantum computing to effectively extract features from EO data for classification and outperform other deep learning models with higher complexity, as evidenced by its superior classification accuracy.

Furthermore, our proposed superpixel encoding allows us to represent images in the quantum domain with fewer quantum resources, which enables the investigation of the potential in QML for image comprehension, especially within the constraints of the NISQ era. By achieving efficiency benefits in terms of qubits, quantum gates, and trainable parameters, our model might contribute to addressing challenges in the remote sensing domain, particularly in EO image classification tasks.

Regardless, future research could continue to explore the following directions: 1) exploring more suitable encoding approaches for EO data of different modalities, such as synthetic aperture radar (SAR); 2) investigating further optimization of the measurement basis, expanding beyond the three basic bases; and 3) studying the potential of pure QML models for EO data understanding.

REFERENCES

- [1] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [2] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.
- [3] Q. Wang, W. Huang, Z. Xiong, and X. Li, "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1414–1428, Apr. 2022.
- [4] Q. Li, Y. Chen, X. He, and L. Huang, "Co-training transformer for remote sensing image classification, segmentation, and detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5606218.
- [5] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, and Z. Shi, "RSMamba: Remote sensing image classification with state space model," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [6] X. Kang, Y. Zhu, P. Duan, and S. Li, "Two-dimensional spectral representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5502809.
- [7] P. Duan, Z. Xie, X. Kang, and S. Li, "Self-supervised learning-based oil spill detection of hyperspectral images," *Sci. China Technol. Sci.*, vol. 65, no. 4, pp. 793–801, Apr. 2022.
- [8] D. Ristè et al., "Demonstration of quantum advantage in machine learning," *Npj Quantum Inf.*, vol. 3, no. 1, p. 16, Apr. 2017.
- [9] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.
- [10] R. Zhang, L. Jiao, L. Li, X. Liu, F. Liu, and S. Yang, "A quantum evolutionary learning tracker for video," *IEEE Trans. Evol. Comput.*, vol. 28, no. 2, pp. 418–431, Apr. 2024.
- [11] Y. Li, M. Tian, G. Liu, C. Peng, and L. Jiao, "Quantum optimization and quantum learning: A survey," *IEEE Access*, vol. 8, pp. 23568–23593, 2020.
- [12] Y. Li, R. Liu, X. Hao, R. Shang, P. Zhao, and L. Jiao, "EQNAS: Evolutionary quantum neural architecture search for image classification," *Neural Netw.*, vol. 168, pp. 471–483, Nov. 2023.
- [13] A. Chalumuri, R. Kune, S. Kannan, and B. S. Manoj, "Quantum-enhanced deep neural network architecture for image scene classification," *Quantum Inf. Process.*, vol. 20, no. 11, p. 381, Nov. 2021.
- [14] P. Gawron and S. Lewiński, "Multi-spectral image classification with quantum neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 3513–3516.
- [15] D. Noever and S. E. M. Noever, "Overhead MNIST: A benchmark satellite dataset," 2021, *arXiv:2102.04266*.
- [16] X. X. Zhu et al., "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 3, pp. 76–89, Sep. 2020.
- [17] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "DeepSat: A learning framework for satellite imagery," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2015, pp. 1–10.
- [18] P. Q. Le, F. Dong, and K. Hirota, "A flexible representation of quantum images for polynomial preparation, image compression, and processing operations," *Quantum Inf. Process.*, vol. 10, no. 1, pp. 63–84, Feb. 2011.
- [19] B. Sun, A. M. Iliyasa, F. Yan, F. Dong, and K. Hirota, "An RGB multi-channel representation for images on quantum computers," *J. Adv. Comput. Intell. Inform.*, vol. 17, no. 3, pp. 404–417, May 2013.
- [20] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [21] S. E. Venegas-Andraca and S. Bose, "Storing, processing, and retrieving an image using quantum mechanics," *Proc. SPIE*, vol. 5105, pp. 137–147, Aug. 2003.
- [22] J. I. Latorre, "Image compression and entanglement," 2005, *arXiv:0510031*.
- [23] J. Su, X. Guo, C. Liu, and L. Li, "A new trend of quantum image representations," *IEEE Access*, vol. 8, pp. 214520–214537, 2020.
- [24] S.-R. Wu, C.-T. Li, and H.-C. Cheng, "Efficient data loading with quantum autoencoder," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
- [25] B. Duan and C.-Y. Hsieh, "Hamiltonian-based data loading with shallow quantum circuits," *Phys. Rev. A*, vol. 106, no. 5, Nov. 2022, Art. no. 052422.
- [26] K. Shen, B. Jobst, E. Shishenina, and F. Pollmann, "Classification of the fashion-MNIST dataset on a quantum computer," 2024, *arXiv:2403.02405*.
- [27] V. Havlíček et al., "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.
- [28] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, "Quantum embeddings for machine learning," 2020, *arXiv:2001.03622*.
- [29] N. Thumwanit, C. Lortaraprasert, and R. Raymond, "Invited: Trainable discrete feature embeddings for quantum machine learning," in *Proc. 58th ACM/IEEE Design Autom. Conf. (DAC)*, Dec. 2021, pp. 1352–1355.
- [30] C. Onuoha, J. Flaherty, T. Nguyen, and T. C. Thang, "Data scanning methods for quantum-classical interface in quantum neural networks," in *Proc. IEEE Int. Conf. Consum. Electronics-Asia (ICCE-Asia)*, Oct. 2022, pp. 1–4.
- [31] P. Easom-McCaldin, A. Bouridane, A. Belatreche, R. Jiang, and S. Al-Maadeed, "Efficient quantum image classification using single qubit encoding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1472–1486, Feb. 2024.
- [32] T. Hur, L. Kim, and D. K. Park, "Quantum convolutional neural network for classical data classification," *Quantum Mach. Intell.*, vol. 4, no. 1, p. 3, Jun. 2022.
- [33] A. Senokosov, A. Sedykh, A. Saginalieva, B. Kyriacou, and A. Melnikov, "Quantum machine learning for image classification," 2023, *arXiv:2304.09224*.
- [34] T. Hur, I. F. Araujo, and D. K. Park, "Neural quantum embedding: Pushing the limits of quantum supervised learning," *Phys. Rev. A*, vol. 110, no. 2, Aug. 2024, Art. no. 022411.
- [35] Y. Li, R.-G. Zhou, R. Xu, J. Luo, and W. Hu, "A quantum deep convolutional neural network for image recognition," *Quantum Sci. Technol.*, vol. 5, no. 4, Jul. 2020, Art. no. 044003.
- [36] S. J. Wei, Y. H. Chen, Z. R. Zhou, and G. L. Long, "A quantum convolutional neural network on NISQ devices," *AAPPS Bull.*, vol. 32, no. 1, pp. 1–11, 2022.
- [37] M. Henderson, S. Shakya, S. Pradhan, and T. Cook, "Quantum neural networks: Powering image recognition with quantum circuits," *Quantum Mach. Intell.*, vol. 2, no. 1, p. 2, Jun. 2020.
- [38] F. Riaz, S. Abdulla, H. Suzuki, S. Ganguly, R. C. Deo, and S. Hopkins, "Accurate image multi-class classification neural network model with quantum entanglement approach," *Sensors*, vol. 23, no. 5, p. 2753, Mar. 2023.
- [39] A. Matic, M. Monnet, J. M. Lorenz, B. Schachtner, and T. Messerer, "Quantum-classical convolutional neural networks in radiological image classification," in *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, Sep. 2022, pp. 56–66.
- [40] S. Y.-C. Chen, T.-C. Wei, C. Zhang, H. Yu, and S. Yoo, "Quantum convolutional neural networks for high energy physics data analysis," *Phys. Rev. Res.*, vol. 4, no. 1, Mar. 2022, Art. no. 013231.
- [41] A. Wang, J. Hu, S. Zhang, and L. Li, "Shallow hybrid quantum-classical convolutional neural network model for image classification," *Quantum Inf. Process.*, vol. 23, no. 1, p. 17, Jan. 2024.
- [42] F. Fan, Y. Shi, T. Guggemos, and X. X. Zhu, "Hybrid quantum-classical convolutional neural network model for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 12, pp. 18145–18159, Dec. 2024.
- [43] W. Li, P.-C. Chu, G.-Z. Liu, Y.-B. Tian, T.-H. Qiu, and S.-M. Wang, "An image classification algorithm based on hybrid quantum classical convolutional neural network," *Quantum Eng.*, vol. 2022, pp. 1–9, Jul. 2022.
- [44] Y. Zeng, H. Wang, J. He, Q. Huang, and S. Chang, "A multi-classification hybrid quantum neural network using an all-qubit multi-observable measurement strategy," *Entropy*, vol. 24, no. 3, p. 394, Mar. 2022.
- [45] A. Sebastianelli et al., "Quantum4EO: Empowering Earth observation by means of quantum neural networks," 2024, *arXiv:2407.17108*.

- [46] S. R. Majji, A. Chalumuri, R. Kune, and B. S. Manoj, "Quantum processing in fusion of SAR and optical images for deep learning: A data-centric approach," *IEEE Access*, vol. 10, pp. 73743–73757, 2022.
- [47] D. A. Zaidenberg, A. Sebastianelli, D. Spiller, B. Le Saux, and S. L. Ullo, "Advantages and bottlenecks of quantum machine learning for remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 5680–5683.
- [48] S. Y. Chang, B. L. Saux, S. Vallecorsa, and M. Grossi, "Quantum convolutional circuits for Earth observation image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 4907–4910.
- [49] S. Otgonbaatar and M. Datcu, "Classification of remote sensing images with parameterized quantum gates," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [50] A. Sebastianelli, D. A. Zaidenberg, D. Spiller, B. Le Saux, and S. Ullo, "On circuit-based hybrid quantum neural networks for remote sensing imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 565–580, 2021.
- [51] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [52] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Adv. Quantum Technol.*, vol. 2, no. 12, Dec. 2019, Art. no. 1900070.
- [53] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 151–162, Aug. 2019.
- [54] M. Broughton et al., "TensorFlow quantum: A software framework for quantum machine learning," 2020, *arXiv:2003.02989*.
- [55] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.
- [56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 4700–4708.
- [58] A. Sebastianelli, M. P. D. Rosso, S. L. Ullo, and P. Gamba, "On quantum hyperparameters selection in hybrid classifiers for Earth observation data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [59] L. Van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.



Fan Fan received the Bachelor of Industrial Design degree from Xi'an Jiaotong University, Xi'an, China, in 2016, and the Master of Human-Computer Interaction degree from Bauhaus University Weimar, Weimar, Germany, in 2020. He is currently pursuing the Ph.D. degree with the Chair of Data Science in Earth Observation, Technical University of Munich, Munich, Germany.

He has been with German Aerospace Center, Weßling, Germany, as a Research Assistant, since November 2020. His research interests focus on quantum algorithms in EO data analysis, including quantum machine learning and quantum optimization.



Yilei Shi (Member, IEEE) received the Dipl.-Ing. degree in mechanical engineering and the Dr.-Ing. degree in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2010 and 2019, respectively.

He is currently a Senior Scientist with the Chair of Remote Sensing Technology, TUM. His research interests include fast solver and parallel computing for large-scale problems, high-performance computing and computational intelligence, advanced methods on synthetic aperture radar (SAR) and InSAR processing, machine learning, and deep learning for a variety of data sources, such as SAR, optical images, and medical images, and partial differential equation (PDE)-related numerical modeling and computing.



Tobias Guggemos received the master's degree in computer science from Ludwig-Maximilians-Universität München (LMU), Munich, Germany, in 2014, and the Ph.D. degree in network security and cryptography from LMU in 2020, under the supervision of Prof. Kranzlmüller.

He has organized and held introductory and practical courses on quantum computing at Ludwig Maximilian University of Munich (LMU), Munich, Germany, since 2016, where he actively changed his research toward quantum computing. He joined German Aerospace Center, Weßling, Germany, in May 2021, for researching quantum algorithms in Earth sciences and recently moved to the University of Vienna, Vienna, Austria, in June 2022, for a post-doctoral position at the Christian Doppler Laboratory for Photonic Quantum Computer of Prof. Philip Walter at the Faculty of Physics.



Xiao Xiang Zhu (Fellow, IEEE) received the master's (M.Sc.) degree, the doctor of engineering (Dr.-Ing.) degree, and the "Habilitation" degree in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is the Chair Professor for Data Science in earth observation at TUM, and was the Founding Head of the Department "EO Data Science" at the Remote Sensing Technology Institute, German Aerospace Center (DLR). Since May 2020, she has been the PI and the Director of the international future laboratory AI4EO, Munich. Since October 2020, she has also been serves as the Director of the Munich Data Science Institute (MDSI), TUM. He was a Guest Scientist or Visiting Professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan, and University of California, Los Angeles, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently a Visiting AI professor at ESA's Phi-laboratory. Her main research interests are remote sensing and earth observation, signal processing, machine learning and data science, with their applications in tackling societal grand challenges, e.g. global urbanization, UN's SDGs, and climate change.