Urban Land Cover Classification with Efficient Hybrid Quantum Machine Learning Model

Fan Fan
Data Science in Earth Observation
Technical University of Munich (TUM)
Munich, Germany
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Wessling, Germany
fan.fan@(tum, dlr).de

Yilei Shi
School of Engineering and Design
Technical University of Munich (TUM)
Munich, Germany
yilei.shi@tum.de

Xiao Xiang Zhu
Data Science in Earth Observation
Technical University of Munich (TUM)
Munich, Germany
Munich Center for Machine Learning
Munich, Germany
xiaoxiang.zhu@tum.de

***Abstract*—Urban land cover classification aims to derive crucial information from earth observation data and categorize it into specific land uses. To achieve accurate classification, sophisticated machine learning models trained with large earth observation data are employed, but the required computation power has become a bottleneck. Quantum computing might tackle this challenge in the future. However, representing images into quantum states for analysis with quantum computing is challenging due to the high demand for quantum resources. To tackle this challenge, we propose a hybrid quantum neural network that can effectively represent and classify remote sensing imagery with reduced quantum resources. Our model was evaluated on the Local Climate Zone (LCZ)-based land cover classification task using the TensorFlow Quantum platform, and the experimental results indicate its validity for accurate urban land cover classification.**

I. Introduction

In the Earth Observation (EO) field, urban land cover classification is of great importance in tracking changes on the Earth's surface and providing insights into human settlement for sustainable urbanization. However, advances in remote sensing technologies are propelling earth observation into the Big Data era. To tackle the complexity and richness of EO data for precise information extraction, large machine learning models have been employed. As a result, an increasing demand for computation power is observed, posing a bottleneck. Quantum computing might contribute in this regard in the future since its primary expectation is to speed up computationally expensive tasks.

Quantum Machine Learning (QML) integrates quantum computing and machine learning. Regarding image classification tasks, numerous contributions, exemplified by [1]–[3], have been made to the field. It is crucial to note that current quantum devices support only a small number of qubits and quantum gates. Thus, many quantum algorithms for image classification are either employed to extract local features [4] or to make final classification based on the global features extracted from classical deep learning models [5]. For land cover classification with QML, many of the proposed models reply on classical deep learning techniques to significantly reduce features from EO data for quantum computing, as demonstrated by [6], [7]. While this approach helps reduce the necessary quantum resources for classification purposes, the effectiveness of QML for image understanding remains uncertain, as its performance could heavily depend on the classical algorithms employed in these hybrid models.

Some other proposed QML models focus on encoding and analyzing the entire input image with quantum computing for classification, such as [8], [9], but in these studies, the input image was downsized for evaluation due to the high demand for quantum resources in image encoding with methods such as FRQI [10] and MCQI [11]. As a result, those models can only be evaluated on a small scale, and their practical performance remains unclear. To tackle this challenge, the study [12] employs a preparation circuit to approximate the quantum state for image representation. Nevertheless, the available quantum resources in the preparation circuit generally affect the validity of the approximation. Thus, the existing encoding methods are insufficient to investigate the effectiveness of QML for image understanding.

In this study, we are motivated to fill this research gap, and we seek to investigate a hybrid quantum neural network for image understanding, which can effectively represent and analyze EO imagery with reduced quantum resources for land cover classification. Our model's performance was evaluated using the TensorFlow Quantum platform [13] with the So2Sat LCZ42 [14] dataset. The experimental results suggest that our model can extract effective features from Sentinel-2 data for urban land cover classification purposes.

Additionally, gate complexity analysis indicates that our model can represent input images as quantum states using fewer quantum resources than MCQI [11]. Regarding operations for feature extraction, our model capitalizes quantum parallelism to accelerate the convolution transformation by simultaneously processing all desired quantum states. Furthermore, in experiments, our model, despite having fewer trainable parameters, outperforms more complex models. These findings underscore the efficiency of our model in classifying images.

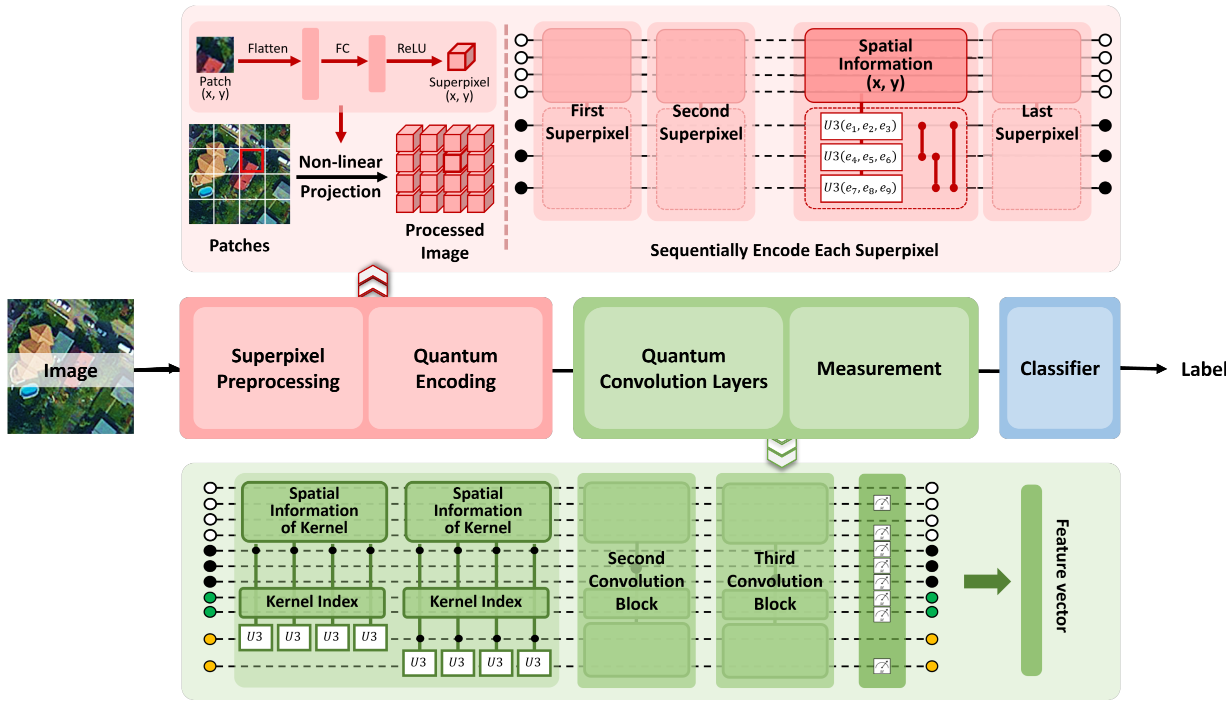


Fig. 1. The framework of our model: 1) superpixel preprocessing: the input image is split into patches, and each patch is non-linearly projected to one superpixel with multiple features; 2) quantum encoding: the obtained superpixels are individually encoded into a quantum circuit; 3) quantum convolution layers: these layers aim to conduct convolutional operations in the quantum domain to extract high-level features from the encoded image; 4) measurement: it yields a feature vector in the classical state derived from the generated quantum feature maps; 5) classifier: one classical dense layer with a softmax activation function is employed for final classification

The paper is structured as follows: Section II introduces basic quantum computing concepts, and Section III details the structures of our proposed models. Section IV describes the experiments for our model's evaluation and analyzes the results. Section V discusses the efficiency of our model for image representation and classification. In the end, Section VI concludes the work.

II. Background

Quantum computing explores the principles of quantum mechanics for computation purposes. Qubit is the foundational unit in quantum computing. Unlike a classic bit, a qubit's state is represented as a linear combination of basis states (e.g., $|0\rangle$, $|1\rangle$). Besides that, qubits can be entangled, and the states of entangled qubits are correlated. To manipulate qubits and transform their states for computation, a series of quantum gates is applied. The quantum gate is an analogous concept to the logic gate in classical computers, and it can convert the input state into another via rotation, entanglement, etc. Interested readers may refer to the work [15] for more detailed information.

To leverage quantum computing for machine learning or deep learning tasks, a hybrid approach, Parameterized Quantum Circuit (PQC), is widely used. Specifically, these quantum circuits are constructed with fixed quantum gates, and their parameters can be optimized for the given task during the training process using classical optimization techniques. The evolution of the quantum states and measurements is conducted on a quantum machine.

III. Methodology

This study introduces a new hybrid quantum neural network designed to represent and classify EO imagery efficiently in terms of the necessary quantum resources. As illustrated in Fig. 1, our proposed model comprises three components: Data Encoding, Feature Extraction, and Classification, each consisting of multiple steps. In total, there are five steps: *A) Superpixel Preprocessing*, *B) Quantum Encoding*, *C) Quantum Convolution Layers*, *D) Measurement*, and *E) Classifier*. We will explain them individually in the following.

A. Superpixel Preprocessing

Our model utilizes superpixels to represent the input image, and each superpixel aims to represent the information from the same image region. As shown in Fig. 1, the input image is partitioned into patches, each contributing to generating features for the corresponding superpixel. This process transforms the input image into superpixels while maintaining the spatial relationship.

To effectively reduce features for each superpixel without causing significant information loss in each patch, we apply a trainable non-linear projection across all the patches. To be more specific, given one patch, its features are first flattened, and a fully connected layer with trainable parameters is adopted to reduce the number of features for the corresponding superpixel. In the end, a ReLU activation function introduces non-linearity into the process. With this approach, the input image is represented by superpixels, each having an element vector.

The patch size and the size of the element vector for superpixels should be adapted based on the tasks and the available quantum resources. Specifically, given an image denoted as $N \times N \times C$, where N represents the spatial resolution and C represents the number of bands, when identifying the patch size P and the number of elements E in the element vector, the processed image can be represented as $\frac{N}{P} \times \frac{N}{P} \times E$. Enlarging the patch size or reducing the elements for each superpixel will lead to a reduction of the necessary quantum gates. A detailed discussion can be found in Section V.

B. Quantum Encoding

To represent the processed image into a quantum state with a small number of qubits, we leverage quantum superposition and entanglement for image encoding. Two quantum registers, q_l and q_e, depicted in white and black, respectively, in Fig. 1, are employed. The former indicates the location of the superpixel, and the latter encodes the features of the superpixel. These two registers are entangled to map the location of the superpixel and its features.

Specifically, to represent a processed image of size $\frac{N}{P} \times \frac{N}{P} \times E$, it requires $2\log_2(\frac{N}{P}) + \frac{E}{3}$ qubits. Exploiting the superposition property, the register q_l with $2\log_2(\frac{N}{P})$ qubits using Hadamard gates is able to indicate all possible locations. Thus, for each superpixel, there exists one identical state in q_l corresponding to its location. To encode its features, $\frac{E}{3}$ qubits in q_e are employed, each using a U3 gate to encode three features. The applied U3 gate is controlled based on the state of q_l to establish the entanglement between q_l and q_e. To enhance the expressibility of the quantum circuit for encoding, inspired by the study [16], we incorporate CZ gates between each pair of qubits in q_e. In this manner, we sequentially encode all the superpixels and represent the image in a quantum state for further feature extraction.

C. Quantum Convolution Layers

The quantum convolution layers are supposed to execute convolutional operations in the quantum domain to extract high-level features from the encoded image and yield quantum feature maps. The corresponding quantum circuit for this purpose follows the principles of the QC-CNN model [17], where the convolutional stride and kernel size are kept consistent for rapid dimension reduction without the need for pooling layers.

Our model contains multiple convolution blocks for feature extraction, each corresponding to one qubit in q_e. Each convolution block can have multiple convolution layers to generate feature maps with desired dimensions, and within each convolution layer, several kernels can be employed.

As depicted in Fig. 1, in the quantum circuit for convolutional operations, the qubits colored in green belong to the quantum register q_k, indicating the indices of the kernels and the corresponding feature maps. The qubits colored in yellow,

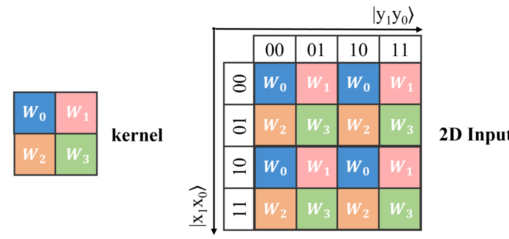


Fig. 2. Illustration of the convolutional operation in our model using a 2×2-sized kernel with weight values from W_0 to W_3

on the other hand, are from the quantum register q_r, encoding the values of feature maps. In this figure, the model with 2×2-sized kernels is exemplified for clarification. Thus, four U3 gates with trainable parameters are used as one kernel for feature extraction, each indicating one weight of the kernel.

Note that to apply a 2×2-sized kernel over the entire image for convolutional transformation in the quantum domain, only four U3 gates, each equivalent to a single operation, would be sufficient. Specifically, for the example in Fig. 2, the pixels of the input colored the same should be transformed with the same weight for convolutional transformation. To identify the pixels of the same color, only the states of two qubits from q_l (denoted as $|x_0\rangle$ and $|y_0\rangle$ in this case) need to be considered. Thus, to apply this kernel across the entire input, four U3 gates controlled by these states are adequate for this operation, regardless of the input size. However, the required operations for classical convolutional transformation scale quadratically with the input size. This observation suggests that our model has an efficiency advantage in feature extraction compared to its classical counterparts.

D. Measurement

The measurement translates the obtained quantum features $|f_m\rangle$ to classical feature values and forms a feature vector for the final classification. Therefore, the qubits encoding feature maps are of particular interest.

Additionally, a set of measurement operators needs to be identified. Our model measures features in the X-basis formed by states $|b\rangle$. For each state of interest $|b_i\rangle$, we define an operator M_i as $|b_i\rangle\langle b_i|$. Consequently, the expectation value $E(M_i)$ regarding the quantum state $|f_m\rangle$ can be obtained following $\langle f_m| M_i |f_m\rangle$. This value will be considered as the i-th feature in the feature vector for subsequent classification.

E. Classifier

Given the feature vector obtained from the measurement, we apply a classical classifier for the final classification. It is worth noting that the classifier can be customized based on task requirements. Here, we use a single dense layer as one example for the final classification. Each neuron encodes one feature value from the feature vector, and all the neurons are fully connected. In the end, a softmax activation function is applied to generate a probability distribution for multi-category classification tasks.

IV. Experiment

To evaluate the validity of our model for urban land cover classification, we used the So2Sat LCZ42 [14] dataset and compared our model's performance against various deep learning models. Since our experiments aim to verify our model's effectiveness, we used a noiseless simulator from the TensorFlow Quantum platform [13] to train our model.

For the experimental data, So2Sat LCZ42 has around half a million Sentinel-2 image patches, each sized at 32×32 pixels and annotated into 17 LCZ labels. Due to the constraints of computation power for quantum simulation, we reduced the size of the dataset in our experiments, narrowing our focus to three German cities: Munich, Berlin, and Cologne. For each patch from these cities, only four bands (Red, Green, Blue, and Near-infrared) were used for classification. Moreover, following the land cover classification scheme in [18], we defined 5 semantic classes from 13 dominant LCZ labels in these cities. Table I provides an overview of those semantic classes and their corresponding LCZ labels. To train the models, we randomly selected $10,000$ patches with a balanced ratio regarding the semantic classes from the target cities, using 6000, 2000, and 2000 samples for training, evaluation, and testing, respectively.

As for the model, we set the patch size as 4, and each patch has 9 elements. Thus, the size of the image after superpixel preprocessing was $8 \times 8 \times 9$. Accordingly, 9 qubits were required to represent the processed image. As for feature extraction, we applied one qubit for kernels and two qubits for feature values. In total, 12 qubits were required for our model in the experiments. Regarding competitors, we selected CNN and ViT [19], in which CNN is a convolution-based model for feature extraction, and ViT tokenizes patches of the input image for further analysis. These models operate on principles similar to our model for image understanding. To ensure a fair comparison, we modified these models' structures to make the differences in trainable parameter numbers between them insignificant. Specifically, the CNN model consisted of 4 convolution layers with 4, 6, 8 and 16 kernels, respectively, followed by one dense layer. The ViT model comprised two transformer blocks, each with a 2-headed attention module for feature extraction.

To train the above-mentioned models, we used the categorical cross entropy as the loss function, and each model was trained for 200 epochs. Our model utilized a learning rate of 0.01, while the others were trained using the default setting. The batch size was set as 50 in the experiments. Each training was conducted three times, and the averaged accuracy of the trained model with the lowest validation loss value was compared and discussed.

Table II presents the experimental results. As the table indicates, our model with fewer trainable parameters outperforms others in terms of test accuracy. Fig. 3 shows the confusion matrices on the test data for the trained models with the lowest validation loss. As indicated in the figure, our model can classify the classes 'open built-up area' and 'vegetation' more accurately than others, although the differences in classification accuracy for each class are marginal. Moreover, Fig. 4 displays the LCZ maps produced by our trained models with the lowest validation loss in the repeated experiments. These maps were generated by applying a sliding window approach to cloud-removed Sentinel-2 images from Google Earth Engine [20]. As depicted in the figure, our models are able to classify land covers based on Sentinel-2 data and display reasonable urban structures.

TABLE I
Derived Urban Land Cover Classes

Label	Semantic Class	LCZ
1	Compact built-up area	1, 2, 3
2	Open built-up area	4, 5, 6
3	Large Low-rise, Heavy industry	8, 10
4	Vegetation	A, B, C, D
5	Water	G

Regarding the comparison of the LCZ maps generated by other models, to the best of the authors' knowledge, the true LCZ labels for these cities are unavailable. Therefore, as illustrated in Fig. 5, we created LCZ reference maps for these cities. These maps were generated based on the classification results of three trained models (ours, CNN, and ViT). For each pixel, we used the majority of the classified labels as references. In cases where all the classification results are different, the corresponding area is annotated without references. As shown in the figure, less than 3% of areas (colored in white) in these cities lack references. In other regions, our model achieves an average consistency of 93.29% with the reference, compared to 90.93% for the CNN model and 90.99% for the ViT model. However, it is important to note that the reference maps may differ from the ground truth.

The experimental results suggest that our model can extract important features from Sentinel-2 imagery for urban land cover classification, utilizing a reduced number of trainable parameters. This indicates our model's effectiveness in feature extraction. However, it is important to note that our model was trained using a noiseless simulator in the experiments, and its performance with noisy quantum devices could be compromised due to the noise effects.

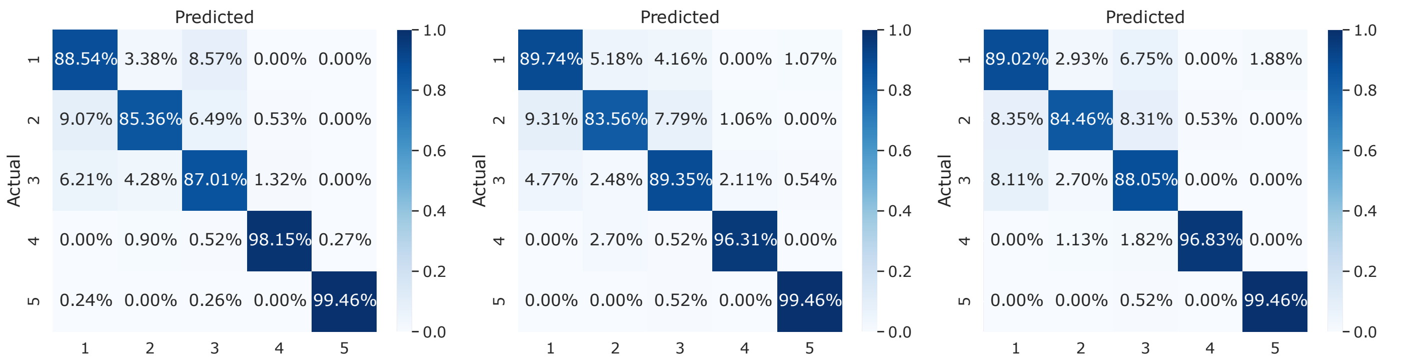
V. Discussion

Besides the efficiency gained from requiring fewer parameters for training, our model also realizes efficiency advantages through a reduction in the necessary quantum resources for EO data representation and classification.

Regarding QML for image classification, the number of required qubits and quantum gates both play crucial roles in the model's efficiency, particularly for image representation. To date, various image representation techniques have been proposed, as discussed in [21]. MCQI [11] is one of the most popular techniques for multi-spectral image representation, capable of significantly reducing the number of qubits for image encoding by exploiting superposition and entanglement

TABLE II
THE CLASSIFICATION PERFORMANCE COMPARISON BETWEEN THE PROPOSED MODEL AND OTHERS

Data	Model	Parameter	Training Acc.	Validation Acc.	Test Acc.
So2Sat LCZ42	Ours	1054	0.940 ± 0.009	0.918 ± 0.009	0.914 ± 0.004
	CNN	1223	0.937 ± 0.011	0.911 ± 0.005	0.907 ± 0.006
	ViT [19]	1665	0.945 ± 0.008	0.914 ± 0.006	0.910 ± 0.007



(a) Ours (b) CNN (c) ViT

Fig. 3. Confusion matrices on the test data for the trained models with lowest validation loss: (a) Ours with the overall accuracy 0.914; (b) CNN with the overall accuracy 0.914; (c) ViT with the overall accuracy 0.913; Classes 1–5 are compact built-up area, open built-up area, large low-rise and heavy industry, vegetation, and water, respectively

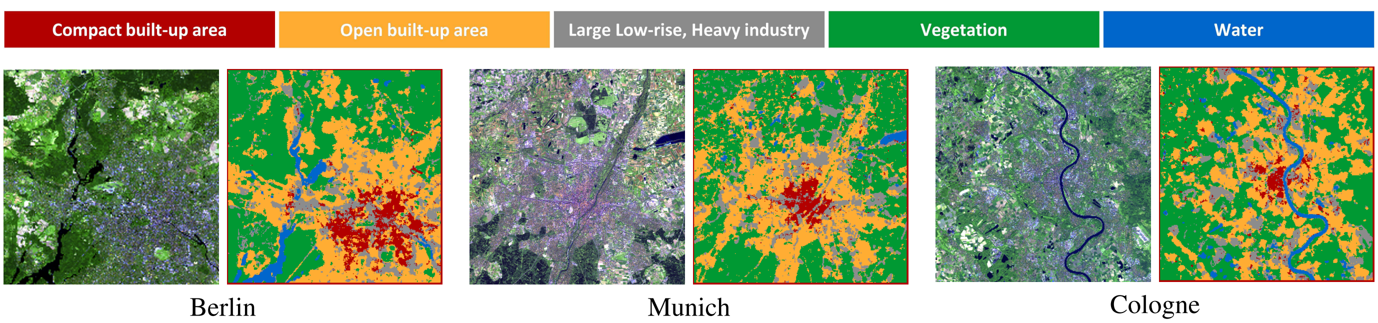


Fig. 4. Land cover maps for Berlin, Munich, and Cologne produced by our trained model with the lowest validation loss: each map covers an area of 30×30 km^2 with the GSD of 10 m; the false-colored cloud removed Sentinel-2 Google Earth images for the corresponding cities

properties. To encode a given image, denoted as $N \times N \times C$, where N represents the spatial resolution and C represents the number of bands, MCQI requires $\log_2(CN^2) + 1$ qubits, $\log_2(CN^2)$ Hadamard gates and CN^2 RY gates controlled by $\log_2(CN^2)$ qubits for image representation.

In comparison, with our approach, when the patch size is set as P, and the number of elements for each superpixel is set as E, our model needs $2\log_2(\frac{N}{P}) + \frac{E}{3}$ qubits, $2\log_2(\frac{N}{P})$ Hadamard gates, $\frac{N^2E(E-3)}{18P^2}$ CZ gates and $\frac{EN^2}{3P^2}$ U3 gates with $2\log_2(\frac{N}{P})$ controllers. Thus, when E is smaller than $3\log_2(2CP^2)$, our approach could further reduce qubits for image representation compared to MCQI. Regarding the required quantum gates, we mainly compare the multi-controlled gates. When E is smaller than $3CP^2$, the count of multi-controlled gates can decrease. Furthermore, the reduction in the number of controllers from $\log_2(CN^2)$ to $2\log_2(\frac{N}{P})$ leads to an additional reduction in elementary quantum gates for building the required quantum circuits.

Specifically, in our experiments, as introduced in Section IV, our model requires 9 qubits to represent an image with dimensions $32 \times 32 \times 4$, whereas MCQI requires 13 qubits. In terms of multi-controlled gates, our model employs 192 U3 gates with 6 controllers, while MCQI requires 4096 RY gates with 12 controllers. Thus, our approach can significantly reduce the number of qubits and quantum gates required for image encoding, leading to efficient image representation. The detailed comparison of quantum resources for EO data representation can be found in Table III.

In terms of feature extraction in our model, as detailed in Section III, the quantum gates required for convolutional transformation over the entire image remain independent of the input's spatial resolution. In contrast, the operations needed for classical convolutional transformation exhibit quadratic scaling with the input size.

As a result, our analysis suggests that our hybrid quantum neural network can efficiently represent and classify images, considering both the number of trainable parameters and the required quantum resources.

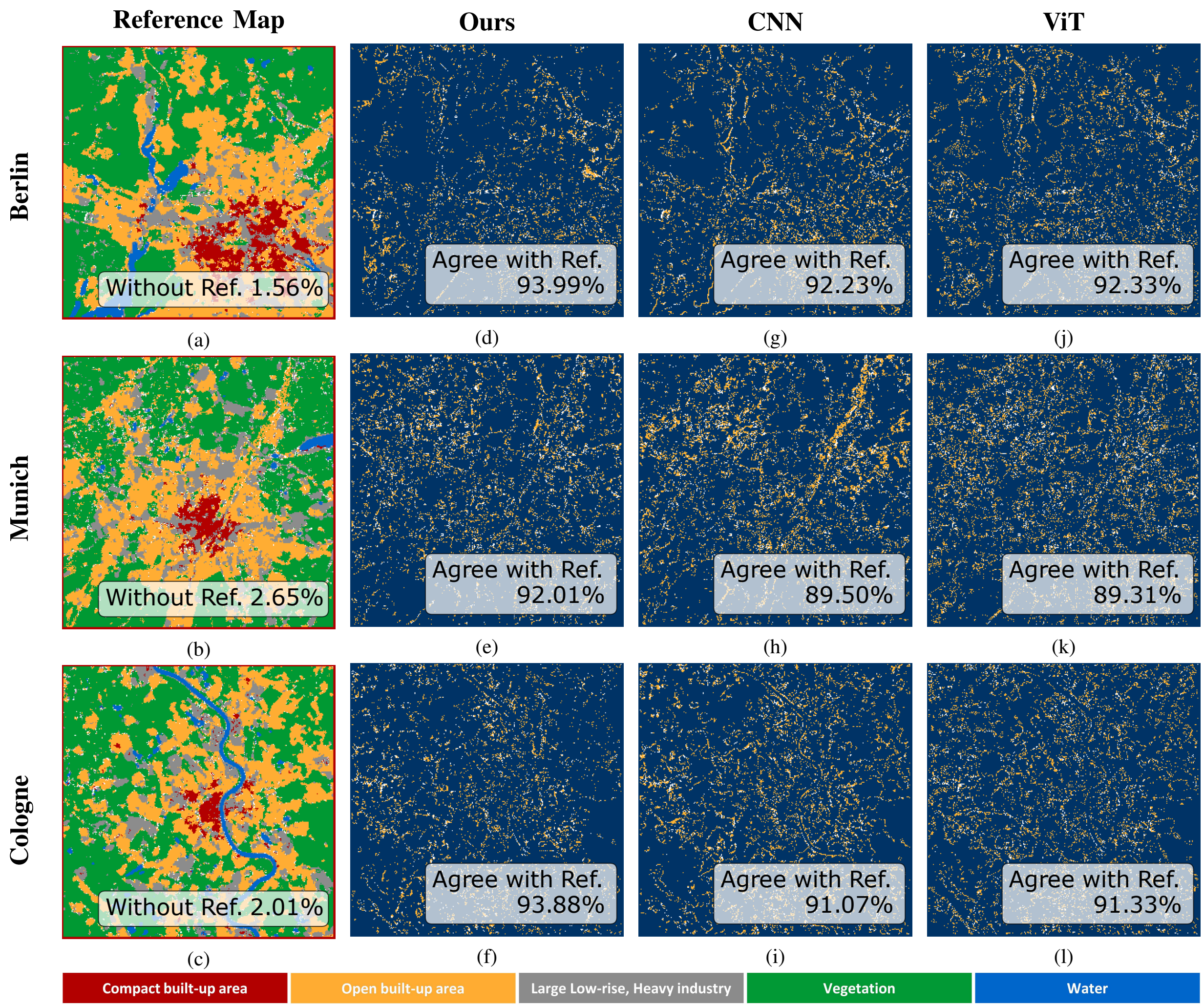


Fig. 5. (a-c) Land cover reference maps generated using a majority vote approach based on the classification results from the best-performing trained models; (d-f) Differences in the classification results between our model and the reference map; (g-i) Differences in the classification results between the CNN model and the reference map; (j-l) Differences in the classification results between the ViT model and the reference map; For the maps in (a-c): pixels in white for the inconsistent results and in other colors for the consistent results; For the maps in (d-l): pixels in white for areas without references, pixels in blue and yellow for the results that agree and disagree with the references, respectively.

TABLE III
QUANTUM RESOURCES NEEDED FOR EO DATA ENCODING

Method	Ours	MCQI [11]
Qubits	9	13
Gates	1) 192 U3 gates with 6 controllers; 2) 6 Hadamard gates; 3) 192 CZ gates	1) 4096 RY gates with 12 controllers; 2) 12 Hadamard gates

VI. CONCLUSION

In this study, we introduce an efficient hybrid quantum machine learning model for the representation and analysis of earth observation data. The proposed model can extract important features from EO imagery for urban land cover classification tasks with reduced trainable parameters. In addition to that, in terms of the required quantum resources for EO data representation, our model demonstrates a significant reduction compared to the MCQI technique.

Regardless, future research could continue to explore the following directions: 1) investigating the potentials of QML for Synthetic Aperture Radar (SAR) analysis; 2) exploring the transferability and generalizability of QML for land cover classification tasks.

REFERENCES

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

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

[3] P. Gawron and S. Lewiński, "Multi-spectral image classification with quantum neural network," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 3513–3516.

[4] A. Matic, M. Monnet, J. M. Lorenz, B. Schachtner, and T. Messerer, "Quantum-classical convolutional neural networks in radiological image

classification," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2022, pp. 56–66.

[5] A. Senokosov, A. Sedykh, A. Sagingalieva, and A. Melnikov, "Quantum machine learning for image classification," *arXiv preprint arXiv:2304.09224*, 2023.

[6] S. Otgonbaatar and M. Datcu, "Classification of remote sensing images with parameterized quantum gates," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[7] A. Sebastianelli, D. A. Zaidenberg, D. Spiller, B. Le Saux, and S. L. Ullo, "On circuit-based hybrid quantum neural networks for remote sensing imagery classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 565–580, 2021.

[8] 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, 2022.

[9] F. Fan, Y. Shi, and X. X. Zhu, "Urban land cover classification from sentinel-2 images with quantum-classical network," in *2023 Joint Urban Remote Sensing Event (JURSE)*. IEEE, 2023, pp. 1–4.

[10] P. Q. Le, F. Dong, and K. Hirota, "A flexible representation of quantum images for polynomial preparation, image compression, and processing operations," *Quantum Information Processing*, vol. 10, pp. 63–84, 2011.

[11] B. Sun, A. Iliyasu, F. Yan, F. Dong, and K. Hirota, "An rgb multi-channel representation for images on quantum computers," *J. Adv. Comput. Intell. Intell. Inform*, vol. 17, no. 3, 2013.

[12] Y. Lü, Q. Gao, J. Lü, M. Ogorzałek, and J. Zheng, "A quantum convolutional neural network for image classification," in *2021 40th Chinese Control Conference (CCC)*, 2021, pp. 6329–6334.

[13] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. H. Yoo, S. V. Isakov, P. Massey, R. Halavati, M. Y. Niu, A. Zlokapa *et al.*, "Tensorflow quantum: A software framework for quantum machine learning," *arXiv preprint arXiv:2003.02989*, 2020.

[14] X. X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Haberle, Y. Hua, R. Huang *et al.*, "So2sat lcz42: A benchmark data set for the classification of global local climate zones [software and data sets]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 3, pp. 76–89, 2020.

[15] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.

[16] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.

[17] F. Fan, Y. Shi, T. Guggemos, and X. X. Zhu, "Hybrid quantum-classical convolutional neural network model for image classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[18] 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 Journal of Photogrammetry and Remote Sensing*, vol. 154, pp. 151–162, 2019.

[19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[20] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.

[21] J. Su, X. Guo, C. Liu, and L. Li, "A new trend of quantum image representations," *IEEE Access*, vol. 8, pp. 214 520–214 537, 2020.